

# **Oracle® Big Data Discovery**

Data Exploration and Analysis Guide

Version 1.0.0 • Revision A • March 2015

# Copyright and disclaimer

Copyright © 2003, 2015, Oracle and/or its affiliates. All rights reserved.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners. UNIX is a registered trademark of The Open Group.

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

**U.S. GOVERNMENT END USERS:** Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

This software or hardware and documentation may provide access to or information on content, products and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.

# Table of Contents

<b>Copyright and disclaimer .....</b>	<b>2</b>
<b>Preface .....</b>	<b>13</b>
About this guide .....	13
Who should use this guide .....	13
Conventions used in this document .....	13
Contacting Oracle Customer Support .....	14
 <b>Part I: Introduction to Big Data Discovery</b>	
<b>Chapter 1: Overview .....</b>	<b>16</b>
Value of Using Big Data Discovery .....	16
Your tasks within Big Data Discovery .....	17
About the Workflow in Big Data Discovery .....	18
Find .....	19
Profile and Enrich Data .....	21
Explore .....	21
Transform .....	22
Discover: Analyze and Obtain Results .....	23
Filter, Use Guided Navigation, and Search .....	24
Understand the Data Flow .....	24
 <b>Part II: Logging in and Managing Your Account</b>	
<b>Chapter 2: Getting Access to Big Data Discovery .....</b>	<b>27</b>
Options for access to Big Data Discovery .....	27
Logging in to Big Data Discovery .....	27
<b>Chapter 3: Managing Your Big Data Discovery Account .....</b>	<b>29</b>
Changing your Big Data Discovery password .....	30
Selecting the locale to use for Big Data Discovery .....	31
About locales .....	31
Configuring the locale for your user account .....	32
Selecting a locale from the user menu .....	32
<b>Chapter 4: Using Big Data Discovery on an iPad .....</b>	<b>34</b>
 <b>Part III: Searching, Filtering, and Navigating in Big Data Discovery</b>	
<b>Chapter 5: Using Refinements for Filtering .....</b>	<b>36</b>
About using refinements for filtering .....	36

Using a value list to select refinements . . . . .	37
Using a range filter to select refinements . . . . .	41
Using date/time values for refinement . . . . .	44
About implicit refinements . . . . .	46
<b>Chapter 6: Using the Big Data Discovery Search Feature . . . . .</b>	<b>48</b>
About the Big Data Discovery search feature . . . . .	48
Displaying and using the results of a search . . . . .	48
Notes about keyword searches . . . . .	56
<b>Chapter 7: Managing Selected Refinements . . . . .</b>	<b>58</b>
About the Selected Refinements panel . . . . .	58
Removing refinements from the Selected Refinements panel . . . . .	60
<b>Part IV: Finding and Browsing Available Data Sets and Projects</b>	
<b>Chapter 8: About the Catalog . . . . .</b>	<b>62</b>
<b>Chapter 9: Filtering and Searching Data Sets and Projects . . . . .</b>	<b>63</b>
<b>Chapter 10: Browsing Information About Data Sets or Projects . . . . .</b>	<b>65</b>
<b>Chapter 11: About Tags . . . . .</b>	<b>67</b>
Adding and removing tags for catalog items . . . . .	67
<b>Part V: Creating, Exploring, and Deleting Data Sets</b>	
<b>Chapter 12: Creating a New Data Set . . . . .</b>	<b>69</b>
Loading data from a file . . . . .	69
<b>Chapter 13: Exploring a Data Set . . . . .</b>	<b>71</b>
About exploring a data set . . . . .	71
Selecting the data to explore . . . . .	71
Switching Explore between tile view and table view . . . . .	72
Filtering and sorting the attributes in Explore . . . . .	72
Working with an individual attribute in Explore . . . . .	74
Using the scratchpad . . . . .	75
<b>Chapter 14: Deleting Data Sets . . . . .</b>	<b>80</b>
Deleting a data set . . . . .	80
<b>Part VI: Creating and Configuring Projects</b>	
<b>Chapter 15: Creating a project . . . . .</b>	<b>82</b>
About projects . . . . .	82
Creating a project using the Catalog . . . . .	82
Creating a project from within a data set . . . . .	83

<b>Chapter 16: Managing Project Pages</b> .....	<b>84</b>
Adding a page to a project .....	84
Renaming a page .....	84
Changing the layout of a page .....	85
Changing the data set binding for a page .....	85
Deleting a page .....	86
<b>Chapter 17: Adding and Configuring Components</b> .....	<b>87</b>
Adding and deleting components on a page .....	87
Renaming components .....	88
Configuring components .....	88
Selecting and configuring the data displayed on a component .....	89
Selecting the view to use for a component .....	89
Selecting the attributes to use on a component .....	90
Selecting the aggregation method to use for a metric .....	93
Configuring the format of values displayed on a component .....	93
Configuring component actions .....	96
Configuring the Actions menu for a component .....	97
About the component Actions menu .....	97
Selecting the actions to include in the Actions menu .....	98
Configuring a Pass Parameters Actions menu action .....	99
Configuring a Refinement Actions menu action .....	100
Configuring actions for displayed values .....	100
Selecting the target page for a refinement or hyperlink .....	102
Configuring cascading for dimension refinement .....	104
Configuring hyperlinks to external URLs .....	105
About component hyperlinks .....	105
Indicating whether to encode inserted attribute values .....	106
About using non-HTTP protocol links .....	107
Recommendations for attribute values that are complete URLs .....	107
Configuring pagination options for components .....	108
Previewing a component .....	108
Summary of available components .....	109
Containers .....	109
Data record displays .....	110
Charts .....	113
Other data visualizations .....	116
Web-based content .....	122
<b>Chapter 18: Managing Access to Projects</b> .....	<b>123</b>
Configuring the project type .....	123
Adding and removing project members .....	124
<b>Chapter 19: Configuring and Deleting Projects</b> .....	<b>129</b>
Localizing the project name, description, page names, and component titles .....	129
Deleting a project using the Catalog .....	130

## Part VII: Viewing and Using Projects

<b>Chapter 20: Navigating Within a Project</b> . . . . .	<b>132</b>
Navigating to a project . . . . .	132
Navigating within a project . . . . .	132
Displaying information about the project data . . . . .	133
<b>Chapter 21: Using Common Component Functions</b> . . . . .	<b>134</b>
Paging through component data . . . . .	134
Using a component to refine data . . . . .	135
Exporting data from a component . . . . .	136
Printing from a component . . . . .	137
Displaying details for a component item . . . . .	138
Comparing items from a component . . . . .	138
<b>Chapter 22: Creating and Using Bookmarks</b> . . . . .	<b>141</b>
About bookmarks . . . . .	141
Displaying and navigating to bookmarks . . . . .	141
Creating, editing, and deleting bookmarks . . . . .	142
Generating and emailing bookmark links . . . . .	144
Bookmark data saved for each component and panel . . . . .	146
<b>Chapter 23: Creating and Managing Snapshots</b> . . . . .	<b>148</b>
About snapshots and snapshot galleries . . . . .	148
Displaying the list of snapshots and snapshot galleries . . . . .	148
Creating and managing individual snapshots . . . . .	149
Creating a snapshot . . . . .	149
Managing saved snapshots . . . . .	150
Emailing or exporting the entire snapshot list . . . . .	153
Creating and managing snapshot galleries . . . . .	153
Creating a snapshot gallery . . . . .	154
Editing and deleting snapshot galleries . . . . .	155
Emailing a snapshot gallery . . . . .	157
Exporting a snapshot gallery to a PDF file . . . . .	157

## Part VIII: Managing Data for a Project

<b>Chapter 24: Managing Data Sets for a Project</b> . . . . .	<b>159</b>
Adding a data set to an existing project . . . . .	159
Removing a data set from a project . . . . .	159
<b>Chapter 25: Transforming a Project Data Set</b> . . . . .	<b>161</b>
About transforming data . . . . .	161
Selecting the data to transform . . . . .	162
Toggling the Transform data grid view . . . . .	163
Filtering and sorting the displayed attributes and values . . . . .	163
Deleting records from the project data set . . . . .	164

Adding transformations to a transformation script . . . . .	165
Overview of transforming an attribute . . . . .	166
Adding an attribute . . . . .	166
Changing the data type for an attribute . . . . .	168
Editing attribute values . . . . .	168
Displaying whitespace characters in attribute values . . . . .	168
Changing the capitalization of string values . . . . .	169
Trimming spaces from string values . . . . .	169
Performing mathematical operations on numeric values . . . . .	169
Selecting subsets of date/time values . . . . .	170
Splitting and grouping values . . . . .	170
Splitting string attribute values . . . . .	170
Grouping attribute values . . . . .	171
Binning numeric attribute values . . . . .	171
Creating a custom transformation script . . . . .	172
Managing and executing the transformation script . . . . .	173
Reordering, editing, and deleting transformations . . . . .	173
Running the transformation script against the project data set . . . . .	174
Creating a new data set from the transformed data . . . . .	175
<b>Chapter 26: Linking Project Data Sets . . . . .</b>	<b>176</b>
About data set links . . . . .	176
Creating a data set link . . . . .	176
Editing a link key . . . . .	179
Managing aggregation methods for a linked view . . . . .	180
Controlling automatic filtering between linked data sets . . . . .	183
<b>Chapter 27: Defining Views of Project Data . . . . .</b>	<b>185</b>
About views . . . . .	185
About base views . . . . .	186
About linked views . . . . .	186
Displaying the list of views for a project . . . . .	188
Displaying details for a view . . . . .	189
Creating views . . . . .	189
Creating a new view . . . . .	190
Copying an existing data view . . . . .	190
Configuring a view . . . . .	192
Editing and localizing the view display name and description . . . . .	192
Editing the view definition . . . . .	193
Guidelines for using EQL to define a view . . . . .	194
Configuring the view attributes . . . . .	198
About the view attributes list . . . . .	198
Editing and localizing an attribute display name and description . . . . .	200
Determining the available subsets of date/time attributes . . . . .	201
Configuring the default display format for an attribute . . . . .	203
Selecting the available and default aggregation methods for an attribute . . . . .	205
Aggregation methods and the data types that can use them . . . . .	207

Configuring the Available Refinements behavior for base view attributes . . . . .	210
Defining predefined metrics for a view . . . . .	211
About predefined metrics . . . . .	211
Adding a predefined metric to a view . . . . .	212
Configuring a predefined metric for a view . . . . .	213
Deleting a predefined metric . . . . .	214
Previewing the content of a view . . . . .	214
Determining whether a custom view is published . . . . .	214
Deleting a view . . . . .	215
<b>Chapter 28: Managing Attribute Groups for a Project . . . . .</b>	<b>216</b>
About attribute groups and the Attribute Groups page . . . . .	216
About groups for linked views . . . . .	217
Displaying the groups for a selected view . . . . .	218
Displaying attribute information for a group . . . . .	219
Displaying the list of attributes for a group . . . . .	220
Displaying the full list of metadata values for an attribute . . . . .	220
Selecting the metadata columns to display for each attribute . . . . .	220
Creating an attribute group . . . . .	221
Deleting an attribute group . . . . .	222
Configuring an attribute group . . . . .	222
Changing the group name and use . . . . .	222
Configuring the group membership . . . . .	224
Configuring the sort order for a group . . . . .	225
Saving the attribute group configuration and reloading the attribute group cache . . . . .	225
<b>Chapter 29: Exporting Project Data . . . . .</b>	<b>226</b>
About exporting project data . . . . .	226
Exporting project data . . . . .	227

## Part IX: Working with Components

<b>Chapter 30: Containers . . . . .</b>	<b>230</b>
Component Container . . . . .	230
About the Component Container . . . . .	230
Configuring a Component Container . . . . .	230
Tabbed Component Container . . . . .	231
About the Tabbed Component Container . . . . .	231
Configuring a Tabbed Component Container . . . . .	232
<b>Chapter 31: Data Record Displays . . . . .</b>	<b>234</b>
Results List . . . . .	234
About the Results List . . . . .	234
Using the Results List . . . . .	235
Configuring a Results List . . . . .	236
Default configuration for a Results List . . . . .	236
Selecting the attributes to display for each record . . . . .	236

Configuring the attributes in a record . . . . .	237
Configuring the sorting options for a Results List . . . . .	238
Configuring the images to display next to each record . . . . .	239
Configuring the Record Details link and Actions menu for a Results List . . . . .	242
Configuring display and pagination options for a Results List . . . . .	243
Results Table . . . . .	243
About the Results Table . . . . .	243
Using a Results Table . . . . .	244
About the Results Table display . . . . .	244
Selecting the column set to display on a Results Table . . . . .	245
Changing the layout of a Results Table . . . . .	246
Showing and hiding columns . . . . .	246
Changing the column display order . . . . .	247
Changing the aggregation method used for a metric . . . . .	247
Showing and hiding summary rows on an aggregated table . . . . .	248
Showing and hiding conditional formatting . . . . .	248
Changing the Results Table sort order . . . . .	248
Sorting a record list table or an aggregated table without summary rows . . . . .	248
Sorting an aggregated table with summary rows . . . . .	249
Using Results Table values to refine the data . . . . .	250
Navigating to other URLs from a Results Table . . . . .	250
Exporting data from a Results Table . . . . .	250
Configuring a Results Table . . . . .	250
Selecting the type of Results Table . . . . .	251
Setting up the data columns for a record list Results Table . . . . .	251
About the default record list column configuration . . . . .	251
Determining whether to use the default record list column configuration . . . . .	251
Available data for a record list table . . . . .	252
Selecting the persistent columns for a record list table . . . . .	252
Creating and configuring column sets for a record list table . . . . .	253
Setting up the data columns for an aggregated Results Table . . . . .	254
Adding dimension columns to an aggregated table . . . . .	254
Adding column sets to an aggregated table . . . . .	255
Configuring columns for a Results Table . . . . .	256
Highlighting Results Table values that fall within a specified range . . . . .	258
Adding action columns for row-level Results Table actions . . . . .	259
Available row-level actions . . . . .	259
Selecting the action columns to include . . . . .	260
Configuring a Hyperlink action column . . . . .	260
Configuring a Refinement action column . . . . .	262
Configuring how each Hyperlink or Refinement action column is displayed . . . . .	263
Configuring conditions for enabling Results Table actions . . . . .	265
Configuring the default Results Table sort order and summary row display . . . . .	266
Configuring the Results Table display and navigation . . . . .	267

<b>Chapter 32: Charts</b> .....	<b>268</b>
Chart .....	268
About the Chart component and its sub-types .....	268
Displaying details for chart values .....	274
Using dimension values to refine the data .....	274
Configuring a Chart component .....	275
Configuring the color palette for the chart display .....	280
Sorting the values on the chart .....	280
Paging controls on a chart .....	281
Saving a snapshot of the chart .....	281
Box Plot .....	282
About Box Plots .....	282
Selecting the Box Plot display options .....	283
Using a Box Plot to refine the data .....	284
Histogram Plot .....	284
About the Histogram Plot .....	285
Selecting the Histogram Plot options .....	285
Using a Histogram Plot to refine data .....	287
Parallel Coordinates Plot .....	287
About the Parallel Coordinates Plot .....	288
Selecting data to display on a Parallel Coordinates Plot .....	289
Using a Parallel Coordinates Plot to refine data .....	289
<b>Chapter 33: Other Data Visualizations</b> .....	<b>291</b>
Map .....	291
About the Map .....	292
Configuring the Map connection to Oracle MapViewer .....	292
Using the Map .....	294
Types of map layers .....	294
Showing and hiding map layers .....	295
Using the numbered points list .....	296
Changing the display of point and heat map layers .....	297
Searching the map .....	298
Displaying details for a map point .....	300
Configuring the Map .....	301
About the default Map configuration .....	301
Adding, editing, and deleting map layers .....	302
Configuring a map layer .....	303
Selecting the view to use for the map layer .....	303
Selecting the map layer type .....	303
Defining the map points for the map layer .....	304
Configuring the layer name and display options .....	306
Selecting and configuring the details to display for each map point .....	307
Configuring pagination and sorting for the map layer .....	309
Determining which map layers are visible .....	310
Configuring general display settings for the Map .....	310

Pivot Table . . . . .	311
About the Pivot Table . . . . .	311
Using the Pivot Table . . . . .	312
Navigating through Pivot Table data . . . . .	312
Displaying summaries and highlighting in the Pivot Table . . . . .	312
Sorting the Pivot Table dimension values . . . . .	313
Changing the layout of the Pivot Table . . . . .	313
Using the Pivot Table to refine the project data . . . . .	314
Configuring the Pivot Table . . . . .	315
Configuring the layout of the Pivot Table . . . . .	315
About configuring the Pivot Table layout . . . . .	315
Adding dimensions and metrics to the Pivot Table . . . . .	316
Changing the orientation of the Pivot Table . . . . .	317
Adding dimensions and metrics . . . . .	317
Configuring a dimension . . . . .	318
Configuring a metric . . . . .	318
Allowing end users to configure the table layout . . . . .	319
Highlighting Pivot Table metric values that fall within a specified range . . . . .	320
Configuring display options for the Pivot Table . . . . .	320
Summarization Bar . . . . .	321
About the Summarization Bar . . . . .	321
Using the Summarization Bar . . . . .	321
Configuring the Summarization Bar . . . . .	323
Default configuration of a Summarization Bar . . . . .	323
Adding, editing, and deleting summary items . . . . .	323
Selecting the type of summary item . . . . .	324
Configuring metric and dimension value spotlight summary items . . . . .	325
Configuring the metric for a metric value summary item . . . . .	325
Configuring the values for a dimension value spotlight summary item . . . . .	326
Configuring the display and actions . . . . .	328
Configuring conditional formatting . . . . .	329
Configuring flag summary items . . . . .	330
Configuring display options for the Summarization Bar display . . . . .	333
Tag Cloud . . . . .	333
About the Tag Cloud . . . . .	334
Using the Tag Cloud . . . . .	335
Selecting the type of Tag Cloud display . . . . .	335
Selecting the values to display on the Tag Cloud . . . . .	336
Configuring the Tag Cloud . . . . .	337
Selecting the available dimensions for the Tag Cloud . . . . .	337
Configuring the available Tag Cloud dimensions . . . . .	337
Configuring the relative size or order of the Tag Cloud terms . . . . .	338
Configuring the available Tag Cloud metrics . . . . .	339
Configuring the general Tag Cloud display options . . . . .	340
Thematic Map . . . . .	340
About the Thematic Map . . . . .	341

Configuring the Thematic Map connection to Oracle MapViewer .....	342
Using the Thematic Map .....	343
Timeline .....	345
About the Timeline .....	345
Adding and editing metric timelines .....	346
Setting the displayed date unit and date range for the metric timelines .....	348
Saving the current timeline configuration as the default .....	349
<b>Chapter 34: Web-Based Content .....</b>	<b>350</b>
IFrame .....	350
About the IFrame .....	350
Using the IFrame .....	351
Configuring an IFrame .....	351
Web Content Display .....	352
About the Web Content Display .....	353
Configuring a Web Content Display .....	353

# Preface

Oracle Big Data Discovery is a set of end-to-end visual analytic capabilities that leverage the power of Hadoop to transform raw data into business insight in minutes, without the need to learn complex products or rely only on highly skilled resources.

## About this guide

This guide contains information about using Oracle Big Data Discovery to explore, analyze, and transform data.

## Who should use this guide

This guide is written for business analysts and data scientists who want to explore and analyze data within the Hadoop ecosystem.

Oracle Big Data Discovery is designed for those who look for insights in data:

- Data Analyst
- Data Scientist
- Analytics Manager
- Problem Solver
- Casual Analyst

## Conventions used in this document

The following conventions are used in this document.

### Typographic conventions

The following table describes the typographic conventions used in this document.

Typeface	Meaning
<b>User Interface Elements</b>	This formatting is used for graphical user interface elements such as pages, dialog boxes, buttons, and fields.
Code Sample	This formatting is used for sample code phrases within a paragraph.
Variable	This formatting is used for variable values. For variables within a code sample, the formatting is <i>Variable</i> .
File Path	This formatting is used for file names and paths.

## Symbol conventions

The following table describes symbol conventions used in this document.

Symbol	Description	Example	Meaning
>	The right angle bracket, or greater-than sign, indicates menu item selections in a graphic user interface.	File > New > Project	From the File menu, choose New, then from the New submenu, choose Project.

## Path variable conventions

This table describes the path variable conventions used in this document.

Path variable	Meaning
\$MW_HOME	Indicates the absolute path to your Oracle Middleware home directory, which is the root directory for your WebLogic installation.
\$DOMAIN_HOME	Indicates the absolute path to your WebLogic domain home directory. For example, if <code>bdd_domain</code> is the domain name, then the <code>\$DOMAIN_HOME</code> value is the <code>\$MW_HOME/user_projects/domains/bdd_domain</code> directory.
\$BDD_HOME	Indicates the absolute path to your Oracle Big Data Discovery home directory. For example, if <code>BDD1.0</code> is the name you specified for the Oracle Big Data Discovery installation, then the <code>\$BDD_HOME</code> value is the <code>\$MW_HOME/BDD1.0</code> directory.
\$DGRAPH_HOME	Indicates the absolute path to your Dgraph home directory. For example, the <code>\$DGRAPH_HOME</code> value might be the <code>\$BDD_HOME/dgraph</code> directory.

## Contacting Oracle Customer Support

Oracle Customer Support provides registered users with important information regarding Oracle software, implementation questions, product and solution help, as well as overall news and updates from Oracle.

You can contact Oracle Customer Support through Oracle's Support portal, My Oracle Support at <https://support.oracle.com>.

# **Part I**

## **Introduction to Big Data Discovery**



## Chapter 1

---

# Overview

This section describes the user workflow in Big Data Discovery and the data workflow. It helps you make your way through finding, exploring, and transforming data using Big Data Discovery.

[\*Value of Using Big Data Discovery\*](#)

[\*Your tasks within Big Data Discovery\*](#)

[\*About the Workflow in Big Data Discovery\*](#)

## Value of Using Big Data Discovery

Big Data Discovery lowers the bar for working with big data, and also expands the set of users who can directly engage with it.

Traditional analytics projects place most effort upfront, where 80% of effort is focused on data preparation, and only 20% on data analysis. They try to predict all possible business questions before letting you engage with the data. This includes modeling, locating and acquiring data, as well as building ETL pipelines. In such tools, business analysts need to first design a model to answer predicted questions, as well as hunt down data sources and manipulate feeds to fit into the designed model. This way only 20% of effort is actually focused on data analysis.

With big data, its complexity compounds the upfront cost of data upload. Increased volume, variety and velocity of data only add complexity, effort and cost required for upfront data loading. Significantly more time is needed to understand and manipulate new source data before it is ready to be published for analysis. This requires highly qualified engineers with expert skills.

Oracle Big Data Discovery Big Data Discovery addresses these problems:

- It expands the set of users who are able to engage directly with the data and therefore contribute to big data projects.
- It makes data preparation 20% of the effort so you can focus 80% of your time on executing and improving your business. This lets you find and analyze data, discover insights and then decide and act without having to prepare data upfront.

# Your tasks within Big Data Discovery

Big Data Discovery is a product for you as a data scientist, analyst, or a data explorer. This topic discusses the goals and needs you typically have and how Big Data Discovery can help you address them.

## Data scientist's tasks

As a data scientist or analyst, you may have these goals and needs:

- **Must solve complex questions using a set of disjointed tools.** You solve complex problems and address unique issues. You start with numerous and dynamic ill-formed questions with unpredictable needs for data, visualization, and discovery capabilities. To address these ad-hoc questions, you rely on open-source and custom-created discovery tools. You often use tools in combination with many other tools, and go in and out of them to get the results you need.

In Oracle Big Data Discovery, this fragmented workflow between tools is replaced with a single workflow that is natively part of the Hadoop ecosystem.

- **Need to collaborate.** Together with other members of your team, you work with big data that is arriving from multiple external and internal sources. Other team members may consume the results of your findings. You also improve and publish insights and prototypes of new offerings.
- **Must make sense of data.** To accomplish your goals, you often need to create and employ insights. You make sense of data by data collection, data cleansing and data analysis.

Big Data Discovery does just that — it lets you make sense of raw data you have and arrive at insights, by letting you collect, join, change and analyze your data.

- **Must generate ideas and insights.** You want to generate actionable insights that can be applied to change the business; enhance existing products, services and operations; and identify, define and prototype new data-driven products, services, and offerings.
- **Must validate, trace back, reiterate and share your hypotheses.** You often need to provide new perspectives to address old problems. The path to the solution is usually exploratory, and involves:
  - Hypotheses revision. You must explore multiple hypotheses in parallel, supported by multiple data sets and interim data products.
  - Hypotheses validation. You need to frame and test hypotheses formally and informally. This requires validation and learning of experimental methods and results done by others.
  - Data recollection and transparency. You want to ensure that all stages of the analytical effort are transparent and can be repeated and shared with others, so that you can reconstruct or repeat existing analytical workflows that you have created previously. This requires linear history of all steps and activities.

Oracle Big Data Discovery allows you to do just that, by saving your projects, data sets and transformation scripts.

## Summary

To conclude, in Big Data Discovery, you can achieve any of these goals:

- Continue to work within the Hadoop ecosystem, and use Big Data Discovery as the visual representation of data found in Hive tables.
- Create custom discovery applications to suit your own business needs.

- Run any type of discovery on data, — for example, understanding customers — comprehensively, on an ongoing basis. You can run discovery on data repeatedly, transparently to others, and on different large-scale data sets arriving from various sources.
- Serve as the original author of all discovery solution elements: data sets, information models, discovery applications, and transformation scripts.
- Run discovery repeatedly, transparently to others in your group, and on different large-scale data sets arriving from various sources.
- Publish insights to decision-making groups and social forums inside your organization.
- Roll your own applications from scratch, iterate and then publish them to wider groups of users.

## Results of Working with Big Data Discovery

As a result of using Oracle Big Data Discovery, you can create either or all of the following:

- A model, such as another data set that illustrates the business hypothesis, or answers a specific business question.
- A product, such as a set of transformations that could be applied to all other similar data sets.
- An insight, such as a conclusion that can be easily illustrated and presented to others in your organization, by using several saved visualizations.
- An infrastructure (a set of tools that includes Big Data Discovery as a component).
- A tool for further analysis of future data sets.

As an outcome, you reach insights that guide you to immediate action, and the intuition and understanding you need to perform feature selection and build more sophisticated mathematical models. You also can make business decisions based on predictions of future outcomes and create data products for new features and services.

## About the Workflow in Big Data Discovery

Oracle Big Data Discovery enables exploratory data analysis. It is structured around the following workflow.

- After you log in, you see the **Catalog**, in which you can select either projects (if they have already been created), or data sets:

 Big Data Discovery



This is where you can **Find** source data placed into the Hive database, or the project which you or your teammates have already created.



**Note:** To move from finding the data set to **Explore**, **Transform**, or **Discover**, you are first asked to save the data set into a project. Alternatively, you can find and select an existing project.

- Next, you can **Explore** source data by looking at various visualizations of aspects of the data (outliers, scatter plots, etc). Also, in **Explore**, you observe and use the results of automatic data profiling, sampling and enrichments to grasp basic data characteristics.

### Explore ▾ Transform ▾ Discover ▾

This is where you pick source data sets that are of interest to you for further exploration, mashup with other data sets, and analysis.

- You can move to **Transform**, where you transform the data to remove inconsistencies, bring further clarity by changing types or other characteristics, and continue with analysis.

### Explore ▾ Transform ▾ Discover ▾

From **Transform**, you can export the data sets from BDD and import it into Hive, or into other systems (in .avro or .csv formats) for further exploration and analysis in tools you have at your disposal.

- In **Discover** you arrive at insights about the data and save your projects for sharing with other teams, to enable further utilization of your findings, be they exploratory analysis, hypothesis, or other insights.

### Explore ▾ Transform ▾ Discover ▾

*Find*

*Profile and Enrich Data*

*Explore*

*Transform*

*Discover: Analyze and Obtain Results*

*Filter, Use Guided Navigation, and Search*

*Understand the Data Flow*

## Find

The number of data sets in any Hadoop reservoir can be huge. Big Data Discovery provides the visual face to Hadoop. In Studio's **Catalog**, you can use search and Guided Navigation to find data sets and existing projects relevant to your analytical purpose.

After you log in, you see the **Catalog**:

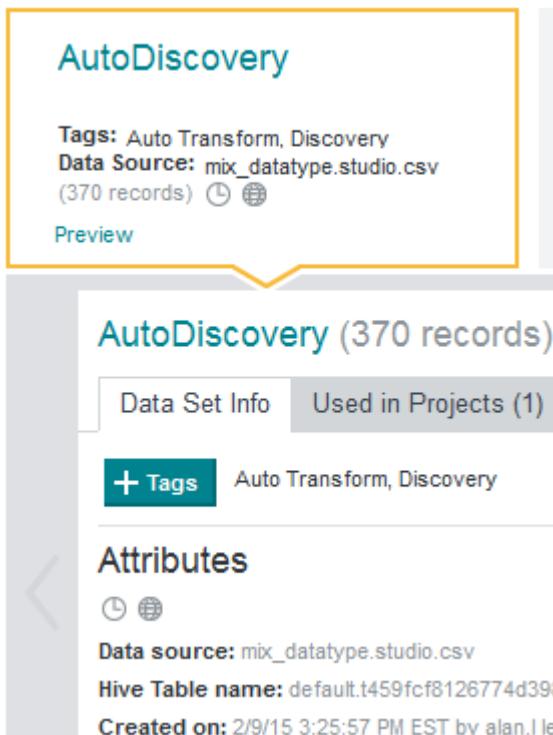
**ORACLE® Big Data Discovery**



Here, you can:

- Find source data sets
- Find existing projects
- Add a data set by using the interface for personal data upload.

This area lets you take a quick look at a data set, using **Preview**. **Preview** relies on the data set's enriched metadata to provide a summary — enough to help you decide if you'd like to move to the exploration phase:



The screenshot shows the AutoDiscovery interface. At the top, there is a yellow-bordered box containing the title "AutoDiscovery" and metadata: "Tags: Auto Transform, Discovery", "Data Source: mix\_datatype.studio.csv (370 records)", and a "Preview" button. Below this, the main content area has a title "AutoDiscovery (370 records)" and a navigation bar with "Data Set Info" and "Used in Projects (1)". Under "Data Set Info", there is a "Tags" button with "Auto Transform, Discovery" listed. The "Attributes" section shows "Data source: mix\_datatype.studio.csv", "Hive Table name: default.t459fcf8126774d398", and "Created on: 2/9/15 3:25:57 PM EST by alan.lie".

Through **Catalog**, you can:

- Browse through your Hadoop cluster to discover relevant data sets based on name, metadata, attribute names, topics, dates, locations, or key terms.
- Search, filter and use Guided Navigation, for finding the data sets.
- Find out which other data sets are recommended to you by Big Data Discovery, based on your interests, overlap in values, and activity of others.
- Use the same experience that is familiar to you from browsing online commerce sites, for searching and Guided Navigation through discovered data sets.

As an outcome of this workflow, you find the data sets and projects you need for your analysis.

#### [About the Workflow in Big Data Discovery](#)

## Profile and Enrich Data

This topic summarizes how Big Data Discovery helps to profile and enrich new data.

As an IT user, you can add new data as files in HDFS using your favorite data integration technologies. As an end user, you can use Studio to upload files, such as Excel and CSV, or pull data from a database using your credentials, landing data in a personal sandbox in HDFS. In either case, Big Data Discovery performs these activities for you:

- Profiles the data, inferring types, classifying content, and understanding value distributions.
- Lists the most interesting data sets first and indicates what they contain.
- Decorates the data with metadata, by adding profile information.
- Enriches the data, by extracting terms, locations, sentiment, and topics, and storing them as new data in HDFS.
- Takes a random sample of the data of the specified size.
- Indexes the data and prepares it for fast search and analysis in BDD.

As an outcome, the data you need resides in HDFS, is indexed by the Dgraph and enriched by BDD, and is ready for your inspection and analysis.

## Explore

You can think of **Explore** area as a guided tour of new data sets, freeing you from the need to manually enter a series of common R commands and trying to extract initial meaning from the results, to better understand what's inside the data set.

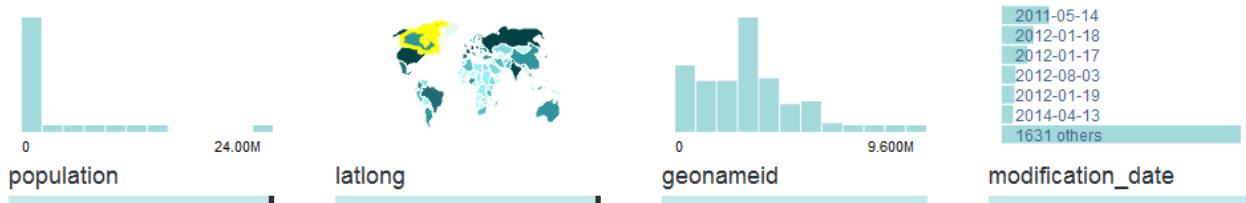
The **Explore** space in Studio is found here:

**Explore** ▾ **Transform** ▾ **Discover** ▾

**Explore** provides an attribute-focused visual summary of the data, summarizing value distributions, data quality gaps, and relationships.

**Explore** presents visualizations that give you the most insight into the data set — visualizations are most suitable for each specific attribute's data type and value distribution.

Visualizations are automatically composed, to save your time and effort at this early stage in the process. When you have a better understanding of the data set, you can compose your own visualizations on data that you have identified as worthy of further analysis.



Here are some of the questions for which Big Data Discovery provides answers, when you use **Explore**:

- What fields are in my data? Do I understand what they are?
- How are values distributed? BDD uses descriptive statistics (mean, median, mode, quintiles) and visuals (histograms, box plots)
- How dirty/messy is the data? Are there missing values or outliers?
- What relationships exist between fields? BDD uses numeric correlations and visuals.
- BDD provides you with out-of-the-box guided analytics experience. It configures the canvas with the right visualizations to explain the data based on your goals, to let you get an overview, explore data quality, and explain available outcomes.



**Note:** Even if you could explore a large data set at full scale, you always want to start by exploring a representative sample and later check against full data.

As an outcome, you understand the data sets you've been exploring.

#### [About the Workflow in Big Data Discovery](#)

## Transform

New data is usually dirty, and Big Data only makes it dirtier. Through exploration, data quality issues often come to the surface. You must triage them and, in many cases, correct, clean, or change data in some way.

The interactive **Transform** area in Studio helps you turn dirty data into data that is fit for advanced analysis. **Transform** helps you isolate data quality problems and lets you quickly apply a number of data transformations to clean or restructure the data.

Here is how you can find **Transform** in Studio:

Explore ▾ Transform ▾ Discover ▾

**Transform** presents a spreadsheet-style view of data, organized either by attribute's value list view, or row view. Here are examples of how you can transform data:

- You can focus on just a few basic transformations needed to support your visualization or analysis, without compromising quality and credibility of your results. You do this by running common data transformations, such as filling missing values, converting data types, collapsing values, reshaping, pivoting, grouping, merging, and extracting.
- You can use live **Preview** to see the set of transformations in action before applying them to the entire data set in Studio.

Because data transformation and data exploration are two integral aspects of any discovery project, you can seamlessly transition between **Transform** and **Explore** within a unified interface, which promotes a smooth workflow.

As an outcome, the data you need is stored in HDFS, indexed by Big Data Discovery, transformed and ready for your analysis.

#### *[About the Workflow in Big Data Discovery](#)*

## Discover: Analyze and Obtain Results

Using **Discover** in Studio, you analyze the data and discover insights hidden in data. **Discover** lets you create meaningful outcomes and results of data exploration and analysis.

The **Discover** area of Studio is found here:

 **Explore** ▾ **Transform** ▾ **Discover** ▾

Some of the examples of the insights you can create with **Discover** are:

- Use search and Guided Navigation to identify subsets of data for deeper analysis.
- Use interactive visualizations to let the data speak for itself, in raw or aggregated forms, in text or in numbers. Statistical techniques (regression, correlation) and machine learning tools (classification, clustering) are used in Big Data Discovery to drive these visualizations.
- Join (or combine) multiple data sets for new insights.
- Predict outcomes. You do this by creating projects, which serve as models for different outcomes. These models automatically understand the weight of different factors in contributing to given outcomes without requiring experts to hand-code rules or heuristics.
- Apply the same projects to new data or to existing data altered to examine new scenarios. (This is known as "what if" analysis.)
- Export your data sets from Big Data Discovery back into Hadoop, so that you can run complex models built in other tools (such as R, or Python) on your custom-created data sub-sets, such as on different customer segments.

#### *[About the Workflow in Big Data Discovery](#)*

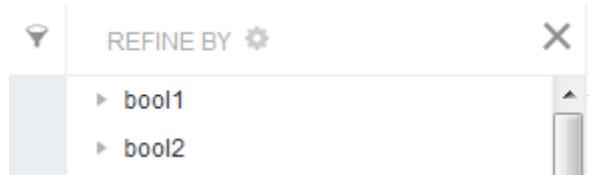
## Filter, Use Guided Navigation, and Search

The following examples explain how search filtering and Guided Navigation work in Big Data Discovery.

The **Find Data Sets and Projects**, **Explore**, **Transform** and **Discover** areas in Studio make use of powerful search and data-driven Guided Navigation capabilities of Oracle Big Data Discovery.

In **Find Data Sets and Projects**, refinements are always available on the side.

In **Explore**, **Transform** and **Discover**, the refinements area uses this icon:



Using refinements, either from the **Refine by** menu, or from the refinement icon, you can filter data sets by attributes. The attribute filter applies to data sets across your entire user experience in BDD. Using the attribute filters you can, for example, select all numeric attributes in a data set in **Explore**, and then switch to **Transform**. **Transform** is then filtered by all numeric attributes.

If you highlight or dismiss any attributes, your project reflects these selections throughout your workflow, when you move from **Find Data Sets and Projects** to **Explore** to **Transform** and back.

You can also search data sets at any stage in your workflow:



If you use search in **Explore**, and then switch to **Transform**, results shown in **Transform** for this data set are filtered by search.

## Understand the Data Flow

This topic provides an overview of the life cycle of a data set as it flows through Big Data Discovery.

Before we proceed with describing the data flow, here are some considerations:

- BDD does not update or delete source Hive tables. BDD only creates new tables to represent new data sets. This way, your source Hive tables remain intact if you want to work with them outside of Big Data Discovery.
- Most of the actions in the data flow take place because you select them and control which actions you want to run. Indexing is a step that runs automatically.

Here are high-level steps of the data flow through BDD:

1. **Create** the data set. You create a data set in BDD using one of the two methods:

- Uploading source data using Studio. In this release, you can upload source data in MS Excel or CSV files. When you upload source data, BDD creates a corresponding Hive source table based on the source data file.
- Running the Data Processing CLI to discover Hive tables and create data sets in Studio based on source Hive tables. Each source Hive table has a corresponding data set in Studio. For details about using the Data Processing CLI, see the *Data Processing Guide*.

The **Catalog** of data sets in Studio displays all data sets whether they are based on files or based on source Hive tables.

2. Optionally, you can choose to **process** the data set. The data enrichment process samples the data set and runs the appropriate data enrichment modules against it. This includes Language Detection, TF/IDF, Geocoding Address, Geocoding IP, and Reverse Geotagger modules. The results of the data enrichment process are stored in the data set and not in the Hive tables.

Although the Data Processing components of BDD perform this phase in the life cycle as part of creating the data set, it's useful to call out this step as a separate logical process that is optionally enabled. For more information about Data Processing, see the *Data Processing Guide*.

3. **Index** the data set. Next, the Dgraph process creates binary index files that represent the data set (and other configuration). The Dgraph accesses the index files to respond to Studio queries.
4. **Transform** the data set. When you use Studio, there are two operations that affect the life cycle of a data set: You can create a new data set, and you can commit a transform script to modify a data set. Both operations are available from **Transform**. As the name suggests, creating a new data set creates a new Hive table.

If you commit a transform script's changes, Studio writes the changes to the Dgraph and stores the changes in the Dgraph's index, but Studio does not create a new Hive table for the data set. You are modifying the data set in the Dgraph but not the source table itself.

5. **Delete** the data set. When you delete a data set from Studio, the data set is removed from the **Catalog** and therefore it is no longer accessible in Studio. However, deleting the data set does not remove the corresponding source Hive table on which the data set is based.

If the Hive database administrator deletes a Hive table from the database, BDD detects that the table was deleted and removes the corresponding data set from the **Catalog** in Studio.

It's important to call out that BDD does not update or delete source Hive tables. BDD only creates new tables to represent new data sets. (You may need to ask your Hive data base administrator to remove old tables as necessary to keep the database clean.)

## **Part II**

### **Logging in and Managing Your Account**



## Chapter 2

# Getting Access to Big Data Discovery

Users can have accounts that are specific to Big Data Discovery, or may use single sign-on to log in using an existing account.

[Options for access to Big Data Discovery](#)

[Logging in to Big Data Discovery](#)

## Options for access to Big Data Discovery

An administrator can create and manage users within Big Data Discovery, or Studio may be integrated with an external user management system.

For details about configuring how user accounts are managed, including how to integrate with LDAP and single sign-on (SSO), see the *Oracle Big Data Discovery Administrator's Guide*.

## Logging in to Big Data Discovery

If you are not logged in automatically using single sign-on, then when you navigate to Big Data Discovery, the **Welcome** page is displayed to allow you to log in with either your Big Data Discovery or LDAP credentials.

To log in to Big Data Discovery from the **Welcome** page:

1. Assuming that your user name is your email address, in the **Email Address** field, type your email address.  
If Big Data Discovery is configured to use a screen name instead of an email address, then the field label is updated to reflect the type of user name.
2. In the **Password** field, type your Big Data Discovery password.

To have Big Data Discovery remember your login information, check the **Remember me** checkbox.

3. Click **Sign In**.

The **Catalog** is displayed.

4. If you forget your Big Data Discovery password, then you can have Big Data Discovery reset it and email it to you.

Note that Big Data Discovery cannot reset LDAP passwords.

To create a request to reset a Big Data Discovery password:

- (a) Click the **Forgot your password?** link.

The **Forgot Your Password?** panel is displayed.

**Forgot Your Password?**

[Sign In](#)

Email Address:



Text Verification:

**Send New Password**

To return to the login panel, click **Sign In**.

- (b) In the **Email Address** field, type your email address.

This must be the address associated with your Big Data Discovery user account.

- (c) In the **Text Verification** field, type the text displayed above the field.

- (d) Click **Send New Password**.

Big Data Discovery resets your password and sends it to your email address.



## Chapter 3

# Managing Your Big Data Discovery Account

The **My Account** page, available from the user menu, provides options to update your user account information, including your password and your preferred locale.

*[Changing your Big Data Discovery password](#)*

*[Selecting the locale to use for Big Data Discovery](#)*

## Changing your Big Data Discovery password

If you sign in using a Big Data Discovery account (as opposed to single sign-on or LDAP), then you can use the **My Account** page to change your Big Data Discovery password.

To change your Big Data Discovery password:

1. In the right corner of Studio, click **User Options>My Account**.

The **My Account** page displays.

**My Account**

\* Required

**User Details**

Screen Name: \*

Password:

Email Address: \*

Retype Password:

First Name: \*

**Display Settings**

Locale:

Middle Name:

Last Name: \*

Time Zone:

Role: Restricted User

► INHERITED ROLES

**Cancel** **Save**

2. In the **Password** field, type the new password.
3. In the **Retype Password** field, type the new password again.
4. To save the new password, click **Save**.

Big Data Discovery verifies that the values you typed matched and that the new password is valid.

## Selecting the locale to use for Big Data Discovery

Big Data Discovery is available in multiple locales. You can select the locale you want to use.

[\*About locales\*](#)

[\*Configuring the locale for your user account\*](#)

[\*Selecting a locale from the user menu\*](#)

### About locales

In Big Data Discovery, the locale determines the language in which the user interface is displayed, as well as the default formatting for data values such as numbers and dates.

Big Data Discovery is configured with a default locale, as is each user account.

Big Data Discovery supports the following locales:

- Chinese - Simplified
- English - US
- Japanese
- Korean
- Portuguese - Brazilian
- Spanish

Note that this is a subset of the languages supported by the Dgraph.

For more details on how Big Data Discovery determines the locale to use, and how to configure the available and default locales, see the *Administrator's Guide*.

## Configuring the locale for your user account

When your user account is first created, the preferred locale is set to **Use Browser Locale**, which indicates to use the browser locale for Big Data Discovery. If Big Data Discovery doesn't support the browser locale, then Big Data Discovery uses its default locale.

To select a different preferred locale for your user account:

1. From the user menu, select **My Account**.

The **My Account** page is displayed.

The screenshot shows the 'My Account' page with the following fields filled in:

Screen Name: rwiggum	Password:
Email Address: ralph.wiggum@ssotest.com	Retype Password:
First Name: Ralph	Display Settings
Middle Name:	Locale: United States - English
Last Name: Wiggum	Time Zone: (UTC ) Coordinated Universal Time
Role: Restricted User	
▶ INHERITED ROLES	
<b>Cancel</b>	<b>Save</b>

2. Under **Display Settings**, from the **Locale** drop-down list, select your preferred locale.
3. To save the change, click **Save**.

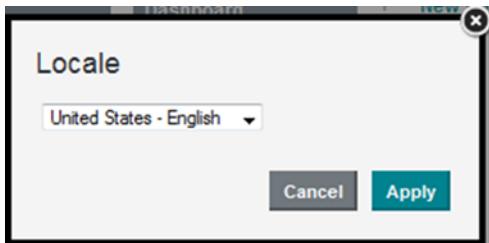
## Selecting a locale from the user menu

The user menu includes an option to select a locale.

If you change the locale from the user menu while logged in to Big Data Discovery, your account is updated accordingly.

To select the locale to use:

1. From the user menu, select **Change locale**.
2. On the change locale dialog, from the **Locale** drop-down list, select the locale.



3. Click **Apply**.

The Big Data Discovery user interface is updated to use the selected locale. This includes:

- User interface labels
- Names of attributes displayed on components
- Values of attributes
- Formatting of data values

Values listed on the **Selected Refinements** panel are not affected.

If you are logged in, then your user account also is updated to use the selected locale as your preferred locale.



## Chapter 4

# Using Big Data Discovery on an iPad

You can use the Safari browser on an iPad running iOS 7 or later to log in to Big Data Discovery and view projects. You cannot use an iPad to create, configure, or export projects.

While the iPad can support most component functions, the component export option is disabled.

## **Part III**

### **Searching, Filtering, and Navigating in Big Data Discovery**



## Chapter 5

# Using Refinements for Filtering

Throughout Big Data Discovery, you can use the refinements panel to refine projects, data sets, or data.

[\*About using refinements for filtering\*](#)

[\*Using a value list to select refinements\*](#)

[\*Using a range filter to select refinements\*](#)

[\*Using date/time values for refinement\*](#)

[\*About implicit refinements\*](#)

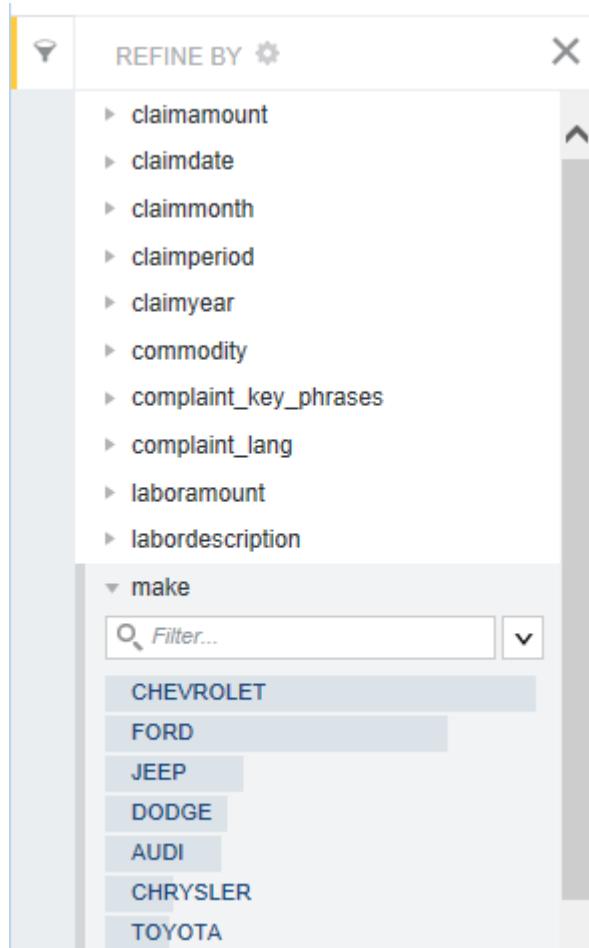
## About using refinements for filtering

Throughout Big Data Discovery, you can use the **Refine By** panel to filter the currently displayed items and data.

On the **Catalog**, the **Refine By** panel allows you to filter the lists of available data sets and projects based on characteristics such as who created them and the types of data they contain.

On the **Explore** and **Transform** pages, the **Refine By** panel allows you to filter data based on attribute values. The **Refine By** panel is collapsed by default.

On the refinements panel, the refinements are displayed in collapsible categories or groups. For each refinement, you can also show and hide the available values.



The screenshot shows the 'REFINE BY' panel with a list of refinements on the left. The refinements are listed as follows:

- claimamount
- claimdate
- claimmonth
- claimperiod
- claimyear
- commodity
- complaint\_key\_phrases
- complaint\_lang
- laboramount
- labordescription

Below this list, there is a collapsed category labeled 'make'. A 'Filter...' input field is positioned above the list of car brands. The brands listed are:

- CHEVROLET
- FORD
- JEEP
- DODGE
- AUDI
- CHRYSLER
- TOYOTA

Depending on the refinement, you can either select a specific value or values, or specify a range of values.

## Using a value list to select refinements

For a value list on the refinements panel, you may be able to select either one or many values to refine by. You also may be able to select a refinement to show items that do NOT have the selected values.

### Displaying and sorting the available values

For a value list, Big Data Discovery initially shows the first 20 values. To display the first 100 values, click **More**. After that, each time you click **More**, the next 100 values are displayed. You can display up to a total of 2000 values.

When all of the values are displayed, to return to displaying the first 20 values only, click **Fewer**.

If you can search for a value, then the filter field is enabled.

By default, the values are displayed in descending order based on the number of matches for each value. The shaded bar indicates the relative number of matches.



The screenshot shows a list of countries under the 'productioncountry' refinement. The values are listed in descending order of matches: United States (darkest grey bar), Germany, Canada, and Korea (lightest grey bar). A 'Select All' button is at the bottom.

You can also display the values in alphabetical order.



The screenshot shows the same list of countries, but they are now sorted alphabetically: Canada, Germany, Korea, and United States. A 'Select All' button is at the bottom.

To change the sort order, click the drop-down arrow for the item, then select the type of sort to use.



The screenshot shows the 'Sort' menu open for the 'productioncountry' refinement. The 'Number of records' option is selected and highlighted with a yellow background. Other options include 'Number of records' (with arrows for ascending/descending), 'Alphabetically' (with arrows for ascending/descending), and 'Alphabetically' (with arrows for ascending/descending).

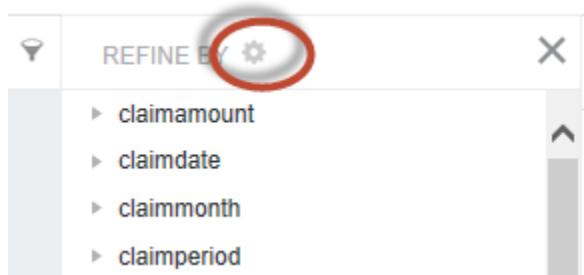
## Changing the metric used to sort attribute values

By default, if attribute values are not sorted alphabetically, they are sorted based on the number of items or records that contain that value.

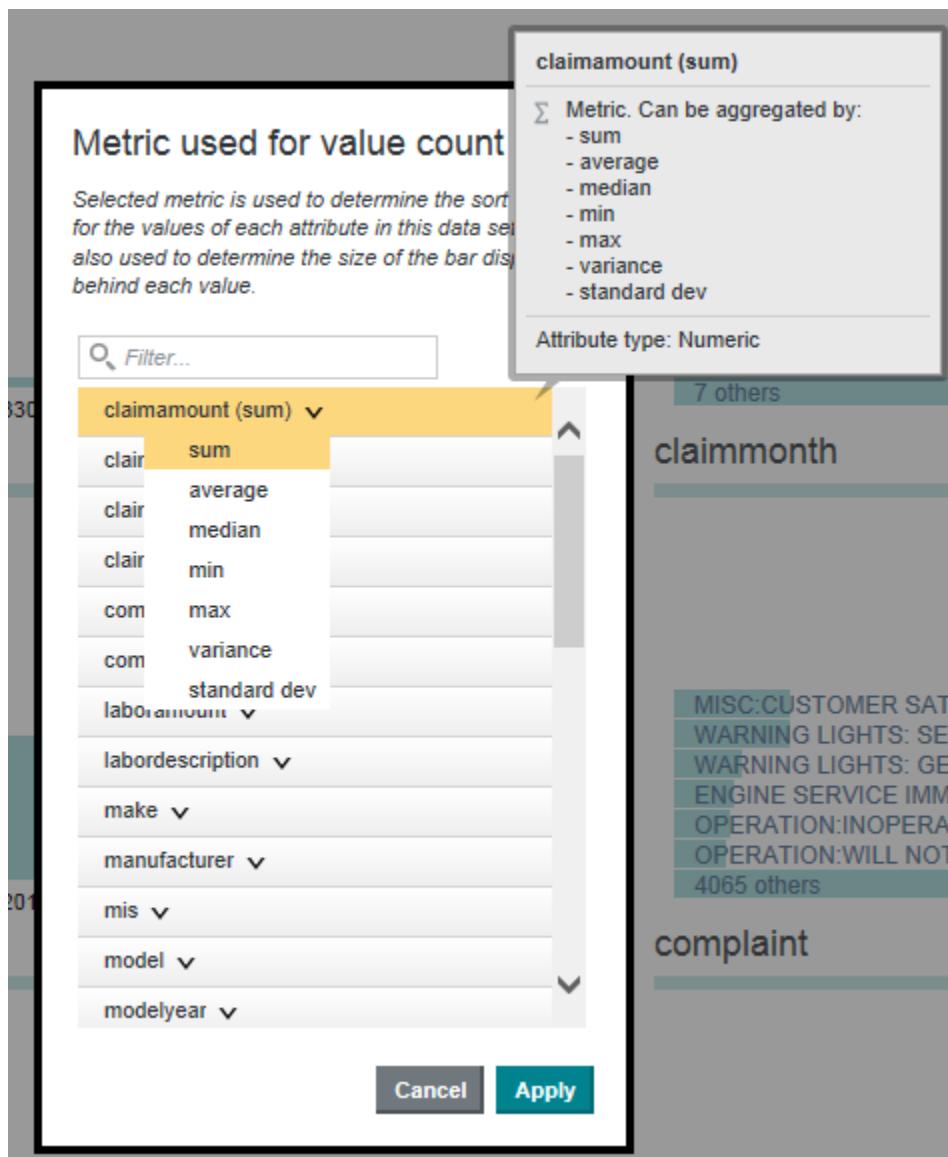
You can also select a different metric to use for the sorting within a data set. For example, you can sort the values of a Region attribute based on the total sales for that region.

To select the metric to use for sorting:

1. Click the configuration icon in the **Refine By** header.



2. On the dialog, click the attribute you want to use. Use the drop-down list next to the attribute to select the aggregation method.



### 3. Click **Apply**.

The sort drop-down list is updated to reflect the currently selected metric value.

## Selecting a single refinement value

To select a single refinement value, click the value in the list.

When you click a value:

- The data is filtered to only display records with that value.
- The selected value is added to the **Selected Refinements** panel.
- If the attribute does not allow you to select multiple values, or if there are no matching records for any of the other values, then the attribute is disabled on the **Refine By** panel.

When you Ctrl-click a value, it is added to the multi-select queue.

## Selecting a negative refinement for a refinement value

You can also apply a negative refinement for a refinement value. Negative refinement means that the data is refined to only include records that do NOT have the selected attribute value.

When you move the mouse over a refinement value, the negative refinement icon displays to the right of the value.

To only include items that have that value, click the value.

To only include records that do NOT have that value, click the negative refinement icon. When you do a negative refinement, then the negative refinement icon is displayed in front of the refinement.

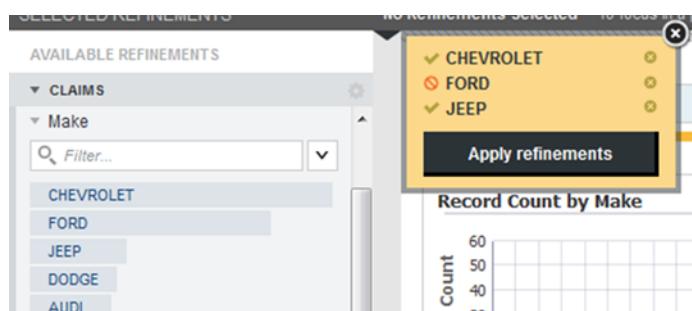
## Selecting multiple values for refinement

If you can select multiple values for a refinement, then the list of values includes a **Select All** button to refine by all of the displayed values.



To refine by a single value, click the value.

To refine by multiple values, Ctrl-click each value. The selected values are added to the multi-select queue. You can also Ctrl-click the negative refinement icon to add a negative refinement for a value to the list.



After selecting all of the values you want to add, to use those values for filtering, click **Apply refinements**. When you filter by multiple values:

- Depending on the refinement, the data is filtered to include records that either:
  - Have any one or more of the selected values
  - Have all of the selected values
- The selected values are added to the **Selected Refinements** panel.
- The selected values are removed from the refinements panel.

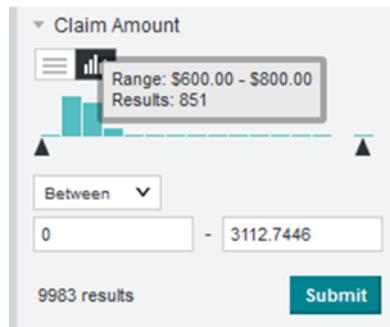
You can still select the other remaining values. If you select all of the values, then the attribute is disabled on the refinements panel.

## Using a range filter to select refinements

On the refinements panel, range filters allow users to refine by a range of values instead of by specific values. Range selections are used by default for numeric values, date values, and time values.

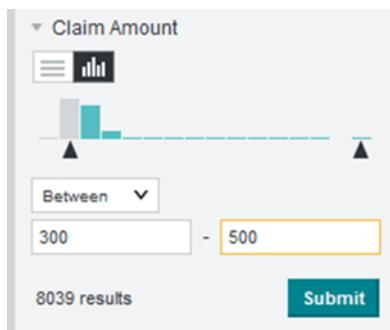
Each filter displays the current minimum and maximum values for the selected attribute.

Range filters also display a histogram, which shows how the values are distributed within the range. Hover over a histogram bar to see the range of values for that bar and the number of matching records in that range.



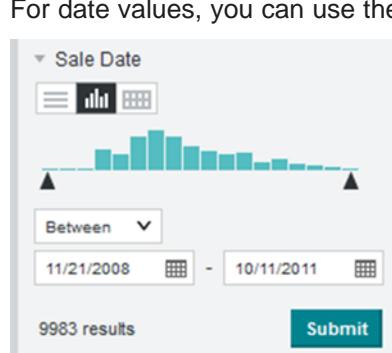
To use a range filter:

1. To select a range of values:
  - (a) From the drop-down list, select whether to define the range:
    - Between two values or two dates or times (the default)
    - Less than a value, or before a date or time
    - Greater than a value, or after a date or time



- (b) To specify the value or values to use, either:
  - Enter the values in the fields.

For date values, you can use the date picker to select the dates.



For time values, you can use either the displayed format (for example, 9:00 am or 1:00 pm) or 24-hour format (for example, 09:00 or 13:00).

- Use the histogram slider bar to set the values.

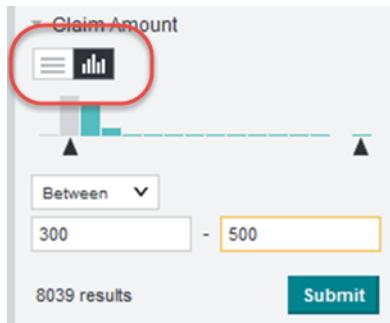
(c) Click **Submit**.

The selected range is added as a refinement on the **Selected Refinements** panel.



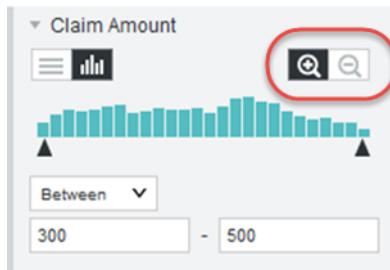
However, the attribute remains available for you to select a different range. If you submit a different range value, the existing refinement is replaced by a refinement for the new range.

2. A toggle icon allows you to toggle between using a range filter and using a value list.



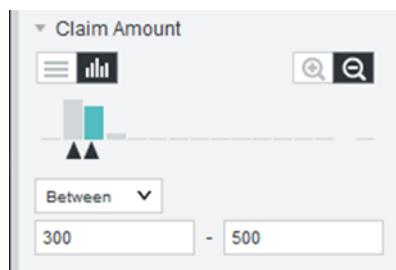
The value list behaves the same way as any other value list. See [Using a value list to select refinements on page 37](#).

3. After you submit a range filter refinement, you can use the range type toggle to change the displayed range on the histogram. This does not affect the selected refinement.



You can either display:

- The current range of available values based on all of the current refinements, including refinements to the current attribute.
- The full range of available values based on all of the current refinements, except for refinements to the current attribute.



## Using date/time values for refinement

For date/time values, the refinements panel provides multiple options for selecting refinements.

Selection Type	Description
Date/time value list	<p>Allows you to select from a list of values for either:</p> <ul style="list-style-type: none"> <li>• Year</li> <li>• Year-Month</li> <li>• Year-Month-Day</li> <li>• Year-Month-Day-Hour-Minute-Second</li> </ul> <p>For example, you could refine the data to only include dates in 2004, or dates in May of 2009.</p> <p>For date/time attributes, the available options for an attribute come from the attribute configuration.</p>
Range filter	<p>Allows you to provide a standard value range for the date/time value.</p> <p>For example, you can refine a date/time attribute to only include values from January 1, 2001 through June 30, 2003, or refine the projects list to only include projects created from December 1 through December 31, 2014.</p>
Recurring date/time unit	<p>Allows you to select a specific unit for:</p> <ul style="list-style-type: none"> <li>• Month</li> <li>• Day of the month</li> <li>• Hour of the day</li> <li>• Minute of the hour</li> <li>• Second of the minute</li> </ul> <p>For example, you could only include:</p> <ul style="list-style-type: none"> <li>• Dates in May of any year</li> <li>• Dates on the third of any month</li> <li>• Date values with time values during the noon hour</li> </ul> <p>You can also select values from different units. For example, you could refine the data to only include dates on May 6 of any year.</p> <p>For date/time attributes, the available values for an attribute come from the attribute configuration.</p>

As you refine by each option, the available values for the other options are also updated. For example, if you refine by May 2004 in a value list, then the range filter is limited to May 2004, and for the recurring date, May is the only available month.

## Using the value lists for a date/time attribute



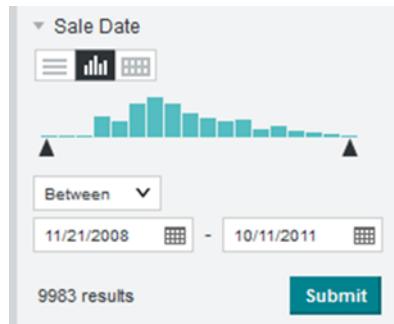
The screenshot shows a refinement interface for the 'ClaimDate' attribute. A dropdown menu is open, showing the 'Year' option selected. Below the menu, a list of years from 2010 to 2012 is displayed, with 2010 highlighted. At the bottom of the list is a 'Select All' button.

To use a date value list to select a refinement:

1. Click the value list button for the date refinement.
2. From the drop-down list, select the type of value (Year, Year-Month, etc.) to display.
3. In the list, click the value to refine by.

The value list works the same way as any other value list. See [Using a value list to select refinements on page 37](#).

## Using the range filter for a date/time refinement



The screenshot shows a refinement interface for the 'Sale Date' attribute. A dropdown menu is open, showing the 'Between' option selected. Below the menu, two date fields are displayed: '11/21/2008' and '10/11/2011'. At the bottom of the interface are buttons for '9983 results' and 'Submit'.

To use a range filter to select a date/time refinement value:

1. Click the range filter button for the date/time refinement.
2. From the drop-down list, select the comparison type.
3. Use the date field or fields to provide the date values.

4. If the date/time field includes the time, then to show or hide fields for specifying the time, click the time icon.

The screenshot shows a date/time refinement interface for 'Order Date'. It includes a 'Between' dropdown, two date pickers (3/21/2013 and 1/5/2014), and two time pickers (13:30:00 and 9:44:00). The 'Submit' button is visible at the bottom right. The interface is part of a larger refinement panel with other fields like 'Claim Date' and 'Month'.

5. Click **Submit**.

## Using recurring date or time values for a date/time refinement

The screenshot shows a date/time refinement interface for 'Claim Date'. It includes a 'Month' dropdown, which is expanded to show a 'Day' dropdown. The 'Day' dropdown contains a grid of dates from 01 to 31. The interface is part of a larger refinement panel with other fields like 'Order Date' and 'Month'.

To use the recurring value option to select a date/time refinement:

1. Click the recurring value button for the date/time refinement.
2. Expand the unit for which you want to make a selection.
3. Click the value you want to select for that unit.

## About implicit refinements

A refinement to one attribute can affect the available refinements for other attributes. This is referred to as an implicit refinement.

For example, if vehicle sales data includes both the make and the model, then when users refine by the vehicle model Mustang, which is produced by Ford, then the make attribute is implicitly refined by Ford.

If the data also includes the sales month, but no Mustangs were sold in January, then January is removed as an available value for the sales month.

Enabling filtering between linked data sets also leads to implicit refinements.

Implicit refinements are not displayed on the **Selected Refinements** panel. However, if there are no available refinements for an attribute because of implicit refinements, the attribute is disabled, and Big Data Discovery displays a message explaining that the attribute is not available because of implicit refinements.



## Chapter 6

# Using the Big Data Discovery Search Feature

The Big Data Discovery search feature provides quick access to data sets, projects, pages, and data from throughout Big Data Discovery.

*About the Big Data Discovery search feature*

*Displaying and using the results of a search*

*Notes about keyword searches*

## About the Big Data Discovery search feature

The search box is available throughout the Big Data Discovery user interface.

You can start a search from:

- **Catalog**
- Pages within a project
- **Project Settings** pages for a project
- **Control Panel** pages

Depending on where you search from, the search results can include:

- Attributes and attribute values
- Bookmarks and snapshots
- Projects and pages
- Data sets

For each type of search result, you have different options for using the result.

You can also use the search text to complete a basic keyword search. Within projects, keyword searches are used to look for search terms in attribute values with searchable string attributes. In the **Catalog**, keyword searches are used to look for the search terms in the information about the available projects and data sets.

## Displaying and using the results of a search

To start a search, begin typing your search text into the search box. As you type, Big Data Discovery displays the matching search results.

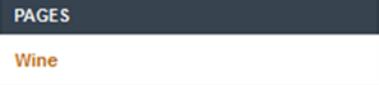
The search results are divided into panels to reflect the different types of results.

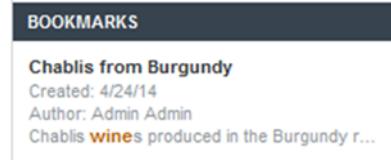
For searches from the **Catalog**, the results are also divided into results from data sets, projects, keyword search, and project refinements.

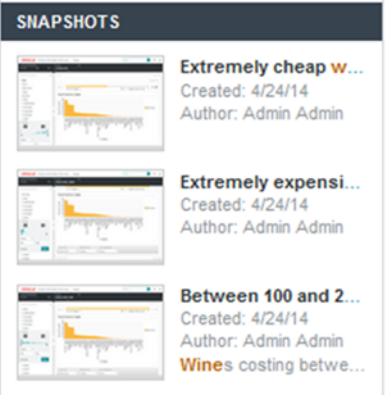
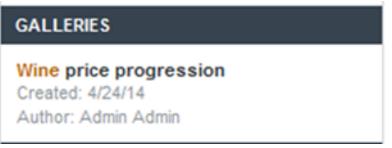
For other searches, there is a single set of panels.

Each section contains a different type of search result. Each type of result has an associated action you can perform from that result. The possible search results include:

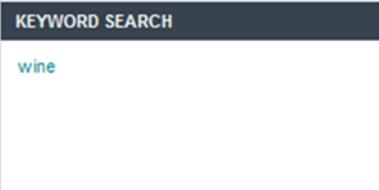
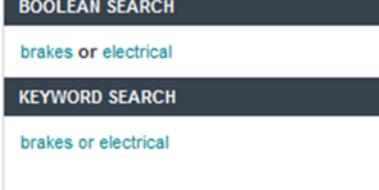
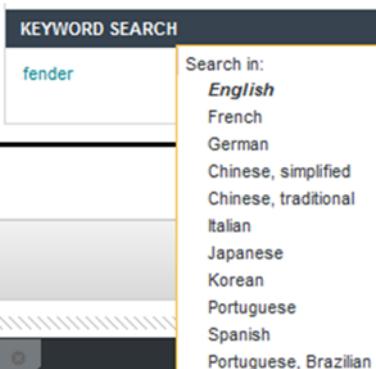
Result Type	Description	Available Action
<b>Attributes</b>	<p>Only included in searches from within a project, including:</p> <ul style="list-style-type: none"> <li>From a project page</li> <li>From <b>Project Settings</b></li> <li>From the <b>Control Panel</b>, if displayed from within a project</li> </ul> <p>A list of attributes whose names match the entered search terms. Next to each attribute is the name of the attribute group that it belongs to.</p> <p>The attributes are grouped by data set.</p> <div data-bbox="437 882 829 1072" style="border: 1px solid #ccc; padding: 5px;"> <p><b>WINE ATTRIBUTES</b></p> <ul style="list-style-type: none"> <li>▶ <a href="#">Wine Type - Other</a></li> <li>▶ <a href="#">P_WineID - Other</a></li> <li>▶ <a href="#">P_Winery - Other</a></li> </ul> </div>	<p>When searching from a project page, to display the available values for refining by the attribute, click the attribute name.</p> <p>When searching from <b>Project Settings</b> or the <b>Control Panel</b>, to display the <b>Views</b> page, click the attribute name. The base view for the data set containing the attribute is selected.</p>
<b>Refinements</b>	<p>Only included in searches from a project page.</p> <p>A list of attribute values, grouped by data set and attribute, that match the entered search terms.</p> <div data-bbox="437 1294 829 1607" style="border: 1px solid #ccc; padding: 5px;"> <p><b>WINE REFINEMENTS</b></p> <ul style="list-style-type: none"> <li><a href="#">P_Winery - Other</a></li> <li><a href="#">Australian Domaine Wines</a></li> <li><a href="#">Bovlei Winery</a></li> <li><a href="#">Brother &amp; Sister Winery</a></li> <li><a href="#">Evans Family Winery</a></li> <li><a href="#">Evans Wine Company</a></li> <li><a href="#">Good Hope Wines</a></li> </ul> </div>	<p>To refine the data using a displayed value, click the value.</p> <p>You can also use Ctrl-click to add a value to the multi-select queue.</p>

Result Type	Description	Available Action
<b>Projects</b>	<p>Included in all searches.</p> <p>A list of projects where the search term matches any of the following:</p> <ul style="list-style-type: none"> <li>• Project name or description</li> <li>• Page names</li> <li>• Data sets</li> <li>• Bookmarks</li> <li>• Snapshots</li> </ul> <div data-bbox="437 699 816 973">  <p>PROJECTS</p> <p>Asimo-ie-20142404-08 Created: 4/24/14 Author: Admin Admin</p> <p>Data sets: <a href="#">Wine</a></p> <p>Pages: <a href="#">Wine</a></p> </div> <p>The results only include projects that you have access to.</p>	<p>To navigate to the project, click the project name.</p> <p>You can also click individual matching items such as pages, bookmarks, or snapshots.</p>
<b>Pages</b>	<p>Included in all searches.</p> <p>A list of project pages that you have access to, and whose name matches the entered search terms.</p> <div data-bbox="437 1262 816 1347">  <p>PAGES</p> <p><a href="#">Wine</a></p> </div> <p>If searching from within a project, then the <b>Pages</b> section only includes matching pages for that project. Matching pages from other projects are listed under <b>Projects</b>.</p> <p>Otherwise, the results include matching pages for any project that you have access to, and are grouped by project.</p>	<p>To navigate to the page, click the page name.</p>

Result Type	Description	Available Action
<b>Bookmarks</b>	<p>Included in all searches.</p> <p>A list of bookmarks whose name or description matches the entered search terms, or that include refinements matching the entered search terms.</p> <div data-bbox="425 523 816 686">  <p>BOOKMARKS</p> <p>Chablis from Burgundy Created: 4/24/14 Author: Admin Admin Chablis wines produced in the Burgundy r...</p> </div> <p>If searching from within a project, then the <b>Bookmarks</b> section only lists matching bookmarks for that project. Matching bookmarks from other projects are listed under <b>Projects</b>.</p> <p>Otherwise, the results include matching bookmarks for any project that you have access to, and are grouped by project.</p>	<p>To display the associated page based on the bookmark, click the bookmark name.</p>

Result Type	Description	Available Action
<b>Snapshots</b>	<p>Included in all searches.</p> <p>A list of snapshots whose name or description matches the entered search terms, or that include refinements matching the search terms.</p>  <p>If searching from within a project, then the <b>Snapshots</b> section only lists matching snapshots for that project. Matching snapshots for other projects are listed under <b>Projects</b>.</p> <p>Otherwise, the results include matching snapshots for any project that you have access to, and are grouped by project.</p>	<p>To display the snapshot details, click the snapshot name.</p> <p>If you are searching from outside of a project, then Big Data Discovery first navigates to the project, then displays the snapshot details dialog.</p> <p>See <a href="#">Managing saved snapshots on page 150</a>.</p>
<b>Galleries</b>	<p>Included in all searches.</p> <p>A list of snapshot galleries whose names match the entered search terms.</p>  <p>If searching from within a project, then the <b>Galleries</b> section only lists matching galleries for that project. Matching snapshots for other projects are listed under <b>Projects</b>.</p> <p>Otherwise, the results include matching galleries for any project that you have access to, and are grouped by project.</p>	<p>To display the gallery details, click the gallery name.</p> <p>If you are searching from outside of a project, then Big Data Discovery first navigates to the gallery, then displays the gallery details dialog.</p>

Result Type	Description	Available Action
<b>Project Settings</b>	<p>Only included in searches from a project page or from <b>Project Settings</b>, for project or Big Data Discovery administrators.</p> <p>A list of pages in <b>Project Settings</b> whose names match the entered search terms.</p> <div data-bbox="437 523 816 677" style="border: 1px solid #ccc; padding: 5px; background-color: #f9f9f9;"> <p><b>PROJECT SETTINGS</b></p> <p><a href="#">Manage Views</a></p> </div>	<p>To navigate to the page, click the page name.</p>
<b>Control Panel</b>	<p>Included in all searches, for Big Data Discovery administrators only.</p> <p>A list of pages in the <b>Control Panel</b> whose names match the entered search terms.</p> <div data-bbox="437 903 816 1058" style="border: 1px solid #ccc; padding: 5px; background-color: #f9f9f9;"> <p><b>CONTROL PANEL</b></p> <p><a href="#">Manage Users</a></p> <p><a href="#">Manage User Groups</a></p> </div>	<p>To navigate to the page, click the page name.</p>
<b>Enterprise Data</b>	<p>Included in all searches.</p> <p>A list of available data sets with matches in their name or description.</p> <div data-bbox="437 1227 816 1381" style="border: 1px solid #ccc; padding: 5px; background-color: #f9f9f9;"> <p><b>ENTERPRISE DATA</b></p> <p><a href="#">Wine</a> Last modified: 3/13/14</p> </div>	<p>To display the data set in <b>Explore</b>, click the data set name.</p>

Result Type	Description	Available Action
<b>Keyword Search</b>	<p>Only included for searches from a project page or from the <b>Catalog</b>.</p> <p>Contains a link with the search text.</p> <p>For projects, the search is within the attribute values from the project data.</p> <p>On the <b>Catalog</b>, the search is within the information about the available data sets and projects.</p>  <p>If you include Boolean syntax in the search text, then Big Data Discovery also displays a <b>Boolean Search</b> section to allow you to complete a Boolean search.</p> 	<p>To complete the text search, click the search term, or press <b>Enter</b>.</p> <p>You can also click the search icon next to the search field.</p> <p>For project data, if the data is available in more than one locale, then when you hover the mouse over the keyword in the results, a drop-down icon is displayed.</p>  <p>To select the locale for the search, click the icon, then click the locale you want to use for the search.</p>  <p>Once you select a locale, Big Data Discovery uses it until you change it. If the current search locale is different from your locale, then Big Data Discovery displays the current search locale next to the keyword.</p>

## Notes about keyword searches

When you complete a keyword search in a project, the data is refined to only include records with the matching search terms. In the **Catalog**, the list of projects and data sets are filtered to only include items with the matching search terms.

### Search mode

For non-Boolean searches, Big Data Discovery looks for items that have all of the terms. If it finds any, it stops looking and returns those records.

If none of the items have all of the search terms, then records are returned based on the partial search rules:

- The item must contain at least one of the search terms.
- The item cannot be missing more than two of the search terms.

For example, if you type California red berry sweet, then:

- An item containing only one term would not be a match.
- An item containing two or more terms would be a match.

### Search refinements

When you complete a text search, the **Selected Refinements** panel is updated to add a search refinement.



For more information on how refinements are displayed, see [About the Selected Refinements panel on page 58](#).

### Displaying snippets for search terms

On the **Results List** and **Results Table** components, the search terms may be highlighted. For attributes that support snippetting, the search snippet is displayed. The snippet displays the portion of the attribute value that contains the search terms.

## Availability of Keyword Search and Boolean Search results

Keyword Search and Boolean Search results, as shown here, may not display in all data sets:

For Keyword Search and Boolean Search to be available in a data sample, there has to be at least one column in the data sample with string values that are larger than 200 characters on average.

## Effect of relevance ranking

For non-Boolean searches, relevance ranking can affect the display order of results on the following components:

Component	Effect of Relevance Ranking
<b>Results List</b>	If there is no default sort order for the <b>Results List</b> , and you haven't applied a sort to the list, then when a search is submitted, the results are sorted based on search relevance.  If a specific sort order has been applied to the table, then the relevance ranking settings are not applied.
<b>Results Table</b>	If there is no default sort order for the <b>Results Table</b> , and you haven't applied a sort to the table, then when a search is submitted, the results are sorted based on search relevance.  If a specific sort order has been applied to the table, then the relevance ranking settings are not applied.
<b>Map</b>	When the search is submitted, the results are automatically sorted based on search relevance.  A <b>Search Relevance</b> option is added to the <b>Sorted by</b> drop-down list.  You can then switch between the <b>Search Relevance</b> option and the other available sorting options.



## Chapter 7

---

# Managing Selected Refinements

The **Selected Refinements** panel displays all refinements currently applied to the **Catalog** or to data sets within a project.

[About the Selected Refinements panel](#)

[Removing refinements from the Selected Refinements panel](#)

## About the Selected Refinements panel

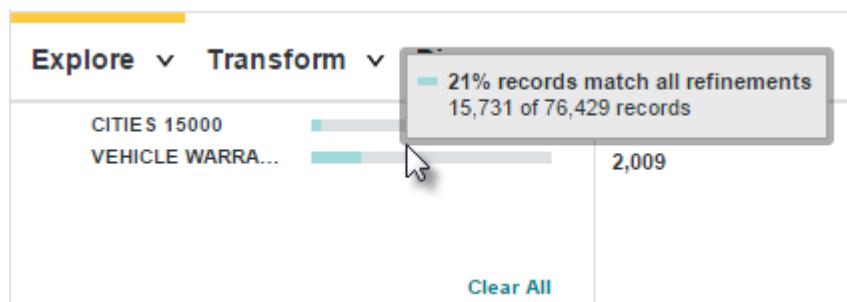
The **Selected Refinements** panel displays all of your currently selected refinements and enables you to modify and remove them. It appears at the top of the page whenever you apply a refinement to the **Catalog** or to a data set within a project.

### The data set list

Within a project, the left side of the **Selected Refinements** panel lists the data sets that are currently being filtered.

CITIES 15000	<div style="width: 20px; height: 10px; background-color: #ccc; border: 1px solid #ccc;"></div>	vehicle_model_year	<div style="width: 20px; height: 10px; background-color: #0070C0; border: 1px solid #ccc;"></div>	vehicle_make	<input checked="" type="radio"/> CHEVROLET	<input type="radio"/> FORD	<input type="radio"/> DODGE	<input type="radio"/> JEEP	asciiname (300)	<input type="radio"/> Victoria	<input type="radio"/> San Luis	<input type="radio"/> Richmond	<input type="radio"/> Greenville
<a href="#">Clear All</a>													

Next to the name of each data set is a bar that indicates the percentage of its records that match the refinements. You can hover your mouse over this bar to view the data set's details, which include the exact percentage and number records that match.

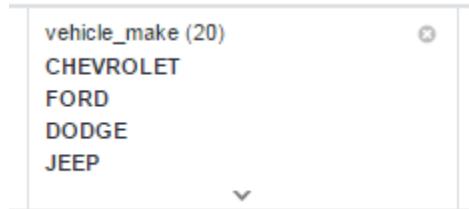


### Refinement tiles

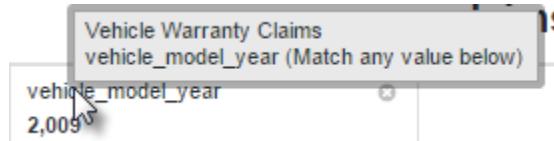
The **Selected Refinements** panel displays each refinement as a tile that contains the name of the refinement and its value.



If the refinement has multiple values, the tile displays the full list in the order you selected them. If there are more selected values than the tile can display at once, you can click the arrow at the bottom of the tile to view the entire list.



You can hover over the refinement's name to view what it applies to. In the **Catalog**, this shows whether it applies to a project or a data set; within a project, it shows the name of the data set it applies to.



If the **Selected Refinements** panel contains too many tiles to display at once, you can use the arrows to the right and left of the panel to scroll through them.

## Hiding the Selected Refinements panel

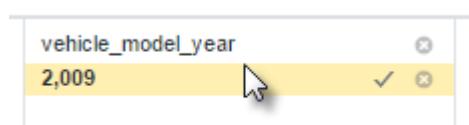
Within a project, you can collapse the **Selected Refinements** panel so that only the list of data sets is displayed.



## Negative refinements

By default, all refinement values are positive, meaning the data is filtered to include records that match them. You can reverse this and filter for records that *don't* match a refinement by making it negative.

To make a positive refinement value negative, hover your mouse over its name and click the check mark.



The checkmark turns into a slashed circle and the value is crossed out. If there are multiple values selected, the negative value will be moved to the bottom of the list.



## Refinements in linked data sets

When data sets in a project are linked, the data set list on the **Discover** page displays a line connecting them and each data set's details specify the number of its records that are unavailable because of the linking.

The screenshot shows the Oracle Big Data Discovery interface. At the top, the ORACLE logo is visible. Below it, the 'Explore' and 'Transform' tabs are selected. In the center, there are two data set tiles: 'CITIES 15000' and 'VEHICLE WARRA...'. A tooltip is displayed over the 'VEHICLE WARRA...' tile, stating: '8% records match all refinements' (284 of 3,630 remaining\* records) and '\* 43,324 records are unavailable as a result of linking to Vehicle Warranty Claims'. At the bottom right, a 'Clear All' button is visible.

Refinements applied to one of the data sets are automatically applied to the other.

## Removing refinements from the Selected Refinements panel

You can use the **Selected Refinements** panel to remove refinements and individual refinement values. You can also use it to remove all selected refinements at once.

To remove a single refinement, click the delete icon in its tile.

The screenshot shows the 'Selected Refinements' panel. It displays a single refinement tile for 'vehicle\_model\_year' with the value '2,009'. A delete icon (a small red square with a white circle) is highlighted with a red box.

To remove a single refinement value, hover your mouse over the value's name and click the delete icon. To remove all values, click the delete icon for the refinement.

The screenshot shows the 'Selected Refinements' panel. It displays a refinement tile for 'vehicle\_model\_year' with the value '2,009'. A delete icon for the value '2,009' is highlighted with a red box.

To remove all currently selected refinements, click **Clear All** in the leftmost tile of the **Refinements Panel**.

The screenshot shows the 'Refinements Panel'. It displays two refinement tiles: 'CITIES 15000' and 'VEHICLE WARRA...'. Below these is a 'Selected Refinements' panel showing a refinement for 'vehicle\_model\_year' with the value '2,009'. At the bottom, a 'Clear All' button is highlighted with a red box.

## **Part IV**

### **Finding and Browsing Available Data Sets and Projects**



## Chapter 8

# About the Catalog

The **Catalog** is the main entry point for Big Data Discovery. It provides access to everything that you have access to. Clicking the logo in the header brings you back to the **Catalog** from any location in Studio.

On the **Catalog**, each project and data set is represented by a tile.

## About the Projects list

To display a list of projects you have access to, click **Projects**.

Projects are sets of pages containing visualizations such as charts and tables. They are used to visualize data from one or more data sets.

If you haven't searched or filtered the list, then the **Projects** tab displays groups of:

- Projects you have viewed recently
- Projects that have been viewed the most by all users

## About the Data Sets list

To display the list of data sets you have access to, click **Data Sets**. Restricted users do not have permission to view **Data Sets**.

Data sets are collections of data that you can explore and add to projects.

If you haven't searched or filtered the list, then the **Data Sets** tab displays groups of:

- Data sets you've used recently
- Data sets that have been used the most by all users
- Data sets that were added most recently



## Chapter 9

# Filtering and Searching Data Sets and Projects

From the **Catalog**, you can display the full lists of projects and data sets. You can also use the **Refine By** panel to filter the **Catalog** lists, or use the search field to search for a project or data set.

### Displaying the full list of projects or data sets

From the **Catalog**, to display the full results list of projects or data sets, click the **View all** link on the **Projects** or **Data Sets** tab, or the **View more** link for the project or data set groups.

### Filtering the Catalog lists

The **Refine By** panel provides options for filtering the **Catalog** lists.

Many of the refinement options are associated with the data in a data set. For example, when you refine the **Catalog** to only show items with geocode attributes, then data sets are included in the results if a project data set includes a geocode attribute.

The available refinements for the **Catalog** are grouped into the following categories:

Refine By Category	Description
<b>Usage</b>	Allow you to options to filter data sets and projects to only display items you created.
<b>Content</b>	Provides options to filter the <b>Catalog</b> items based on values in the associated data.  For example, you can filter items based on whether the data contains date/time values or geocodes, or contains values in a particular language.
<b>Metadata</b>	Provides options to filter the <b>Catalog</b> items based on other data about the associated item.  For example, you can filter items based on: <ul style="list-style-type: none"><li>• The author of the data set or the project</li><li>• Tags</li><li>• When the item was last modified</li><li>• The number of attributes or records in the data set</li><li>• Where the data came from</li></ul>

For details on using refinements for filtering, see [\*Using Refinements for Filtering on page 35\*](#).

## Searching the Catalog

You can also use the search field to search for catalog items. See [\*Using the Big Data Discovery Search Feature on page 47\*](#).



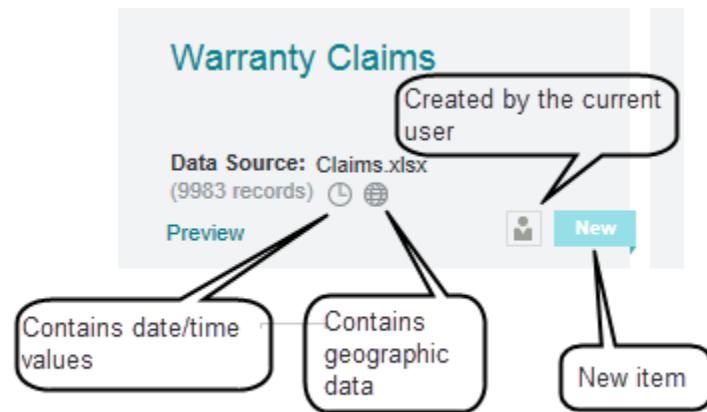
## Chapter 10

# Browsing Information About Data Sets or Projects

From the **Catalog** you can sort the displayed items, and get additional information about each project and data set.

### Icons used on the Catalog tiles

On the **Catalog**, each tile uses icons to provide additional information about the item and its associated data.



These icons include:

Icon	Definition
	Indicates an item that was created within the last week.
	Indicates an item that you created (the currently logged in user).
	Indicates data that includes date/time values.
	Indicates data that includes geocode values.

This same information about a data set is also accessible by clicking the **Show/Hide Data Set Properties** icon.

## Sorting the lists of data sets or projects

At the top of the **Catalog** results page is a sorting drop-down list to allow you to sort the list of data sets or projects.

From the drop-down list, select the value to use for the sort. To switch between ascending and descending order, click the sort order icon.

## Displaying a preview of a data set

The **Preview** link for each data set displays a small sample (the first 15 records and 50 attributes) of the data set.

## Displaying details for a data set or project

Clicking the tile displays a panel with additional details about the item, including:

- Who created it
- When it was created and last updated
- Tags
- Available actions
- The number of times it has been viewed
- For projects:
  - Data sets used in the project
  - Pages in the project
  - Numbers of snapshots and bookmarks
- For data sets:
  - Types of data in the data set
  - Projects where the data set is used
  - Other data sets used in the same projects as this data set

You can use the paging icons at the left and right to page through the details panels for the previous or next data set or project.

In addition to accessing the details from the tile, you can also display a subset of the details by clicking the **Show Data Set Properties** icon in **Explore** and **Transform**.



## Chapter 11

# About Tags

From the **Catalog**, you can add one or more tags to a data set or a project. Tags are words or phrases that provide meaningful information about a data set or project. Once you add a tag, you can then find data sets and projects that contain those tags using the **Catalog** search feature and the **Catalog** feature to **Refine By** either **Project Tags** or **Data Set Tags**.

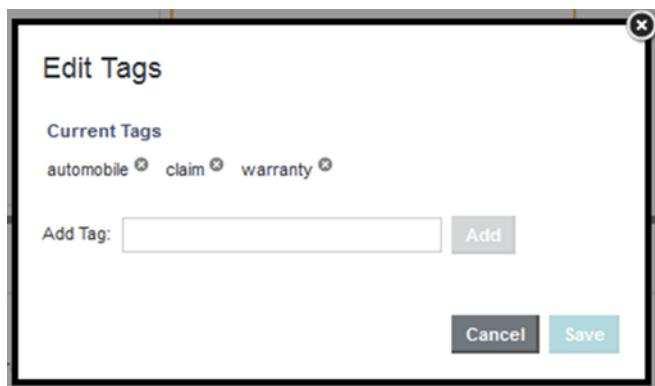
### *Adding and removing tags for catalog items*

## Adding and removing tags for catalog items

Each data set and project can have associated tags. From the **Catalog**, you can update the tags.

To add and remove tags for data sets and projects:

1. In the **Catalog**, locate the data set or project that you want to add a tag.
2. Click the **Tags+** button on the details panel, or select the **Edit Tags** option.
3. To add a tag, type the tag text in the **Add Tag** field, then click **Add**.



4. To remove a tag, click its delete icon.
5. Click **Save**.

# **Part V**

## **Creating, Exploring, and Deleting Data Sets**



## Chapter 12

# Creating a New Data Set

From the **Catalog**, you can create a new data sets.

### *Loading data from a file*

If you are a User, Power User or Administrator, you can create a new data set by uploading data from an Excel or CSV file. A user with the role Viewer cannot create a data set.

By default, any file you upload is processed to result in a maximum data set sample size of up to 1,000,000 records. Files that contain more than 1,000,000 records are down sampled to approximately 1,000,000 records. Files with less than 1,000,000 records contain the full record set in the data set sample size. If necessary, you can increase the default setting from 1,000,000 records to another value. For details, see the `bdd.maxRecordsToProcess` documentation in the *Oracle Big Data Discovery Administrator's Guide*.

Note that if your file has columns with identical headings, then in the new data set, the columns are converted into a single multi-value attribute. For example, if your data includes:

Item	Color	Color	Color
T-Shirt	Red	Blue	Green
Sweatshirt	Red	White	

Then in the final data set, the result is:

Item	Color
T-Shirt	Red, Blue, Green
Sweatshirt	Red, White

To create a data set from an uploaded file:

1. On the **Catalog**, click **+Data Set**.  
From within a project, you can also add a new data set.
2. On the **New Data Set**, provide a name and description for the new data set.
3. Click **Upload a File**.
4. Click **Browse** to locate and select the file.

5. Configure the details for the file.

For an Excel file, you can select the worksheet to use and indicate whether the data includes a header row.

For a CSV file, you can:

- Select the delimiter used for the file
- Select the locale for the data
- Configure the number of rows to skip from the top of the file
- Indicate whether the data includes a header row

6. Click **Done**.

A new data set, based on the file, is available in the **Catalog**.

If the file upload fails due to a connection timeout, the file may be too large to upload from Studio. You can work around this issue by asking your Hive database administrator to import the source file into a Hive table, and then you run the Data Processing CLI utility to process the table. After data processing, a new data set, based on the file, is available in the **Catalog**.



## Chapter 13

---

# Exploring a Data Set

**Explore** allows you to explore a data set, either from the **Catalog** or from within a project.

[About exploring a data set](#)

[Selecting the data to explore](#)

[Switching Explore between tile view and table view](#)

[Filtering and sorting the attributes in Explore](#)

[Working with an individual attribute in Explore](#)

[Using the scratchpad](#)

## About exploring a data set

**Explore** allows you to explore the structure and data values in a data set.

**Explore** ▾ **Transform** ▾ **Discover** ▾

Using **Explore**, you can find out:

- What attributes are available
- The types of values for those attributes
- The distribution of those values across the data set
- The quality of the values – in particular, whether values are missing
- How attributes correlate to each other

After exploring a data set, you should have a better idea of the visualizations you may want to use to show the data values and relationships, and the transformations that you may want to apply.

## Selecting the data to explore

You can explore a data set from the **Catalog** or from within a project.

### Selecting a data set to explore

To explore a data set, click the data set name in the **Catalog**.

From the data sets view of a project, to explore a project data set, click the data set tab.

## Explore header information and links

At the top of **Explore** is a summary of the data you are looking at. For example:



This summary includes:

- The number of records in the sample being used
- The number of records in the entire data set
- The number of records in the sample that match the current refinements
- The number of attributes in the data set
- Attributes that have been marked as favorites
- A **Data Type** drop down to locate attributes by data type (string or numeric)
- A **Name** drop down to locate attributes by name
- A **Hidden** control to display any attributes that were hidden.

The page header can also contain links to:

- Show/Hide the Scratchpad
- Show/Hide data set properties

## Switching Explore between tile view and table view

By default, **Explore** displays each attribute as a tile containing a visualization showing the distribution of values for that attribute.

The default visualization is based on the attribute's data type.

A data quality bar shows the percentage of valid values (blue) and null values (black). You can click a section of the data quality bar to refine the data to only include records with valid, invalid, or null values.

You use the toggle switch at the top right of the list to switch between the tile view and a table view. The table view contains the data set records.

## Filtering and sorting the attributes in Explore

Within **Explore**, you can decide which attributes you want to see, and the order in which to display them. If you are within a project, these preferences carry over to **Transform**, and also apply to all users of the project.

### Setting attributes as hidden or favorite

As you look over the attributes, you may discover attributes that you are very interested in, or that don't have as much value and you don't want to see.

For attributes that you are most interested in, you can click **Add to favorites** in the attribute panel. These attributes move to the front of the sorted list.

For attributes that you are not at all interested in, you can hide the attribute by clicking **Hide from view** in the attribute panel. This removes the attribute from the display.

The favorite and hide options display when you hover over an attribute tile, and when you display the attribute details.

If you mark an attribute as favorite or hidden, when exploring the data set from the **Catalog**, the attribute is displayed a favorite or hidden for all users who explore the data set from the **Catalog**.

## Filtering the displayed attributes

The attribute filter bar allows you to determine which attributes to display.

You can show or hide attributes based on:

- The data type for the attribute.

Check the checkbox next to each data type to include.

- The first letter of the attribute name.

Type an attribute name into the search box to locate attributes whose names start with that letter.

- Whether the attribute is starred or hidden.

To only show starred attributes, click **Favorites**. To only show hidden attributes, click **Hidden**.

You can also use the filter field to filter the attributes based on text in the attribute name.

## Sorting the displayed attributes

When you sort based on an attribute, the selected attribute is moved to the front of the list, and the other attributes are displayed in order based on that attribute. In **Explore**, you can sort the attributes:

- Based on an algorithm that places attributes with more information potential at the top of the list. This is the default.
- Alphabetically by name
- Based on the relationship to a selected attribute.

This sort suggests how closely related the attribute is to others in the data set and sorts the attributes accordingly. This sort suggests what one attribute can tell you about the distribution of other attributes. For example, you might sort a `complaint` attribute in a manufacturing data set and find that a `supplier name` attribute is closely related.

This sort does not work with attributes that have hundreds of distinct values, so those attributes are marked as ineligible for sort by relationship. Also, this sort does not evaluate null values; attributes that have a large number of null values are moved toward the end of the sorted list.

To sort alphabetically by name, select **By name** from the **Sort** drop-down list.

To sort based on the relationship to a selected attribute, select **Relationship to an attribute** from the **Sort** drop-down list and choose an attribute to sort by.

# Working with an individual attribute in Explore

From **Explore**, you can filter the values, select a different visualization, and display a details panel with additional options.

## Filtering the attribute values

The **Available Refinements** panel allows you to filter attribute values by selecting specific values or ranges of values. The refinements are added to the **Selected Refinements** panel at the top of the page.

You can also use the individual visualizations to refine the data based on the attribute values.

## Displaying a different visualization for an attribute

In tile view, each attribute may have multiple visualization options. For example, for a numeric attribute, you might be able to switch between a histogram or a box plot.

If there are multiple visualization options, then to switch to a different visualization, use the paging arrows to the left and right of the attribute visualization.

## Displaying the details panel for an attribute

When you click the attribute name on the attribute tile, the details panel for the attribute is displayed.

You can use the paging icons to the left and right of the details panel to page through the details panels for each attribute.

The panel includes options to:

- Indicate the attribute is a favorite or hidden
- Sort the attribute values
- Add the attribute to the scratchpad

The details panel also contains the following tabs:

Tab	Description
<b>Overview</b> tab	Displays the attribute visualization, with a <b>Summary Statistics</b> section listing information about the attribute and its values. Also displays thumbnails representing the available visualizations for the attribute. To change to a different visualization, either use the paging icons to the left and right of the visualization, or click the visualization thumbnail.

Tab	Description
<b>Details tab</b>	<p>Displays some details about the attribute, including:</p> <ul style="list-style-type: none"> <li>Display name</li> <li>Enrichments that were applied to the original values</li> <li>Data type</li> </ul> <p>It also displays the list of values for the attribute, sorted by relative frequency within the data sample.</p>

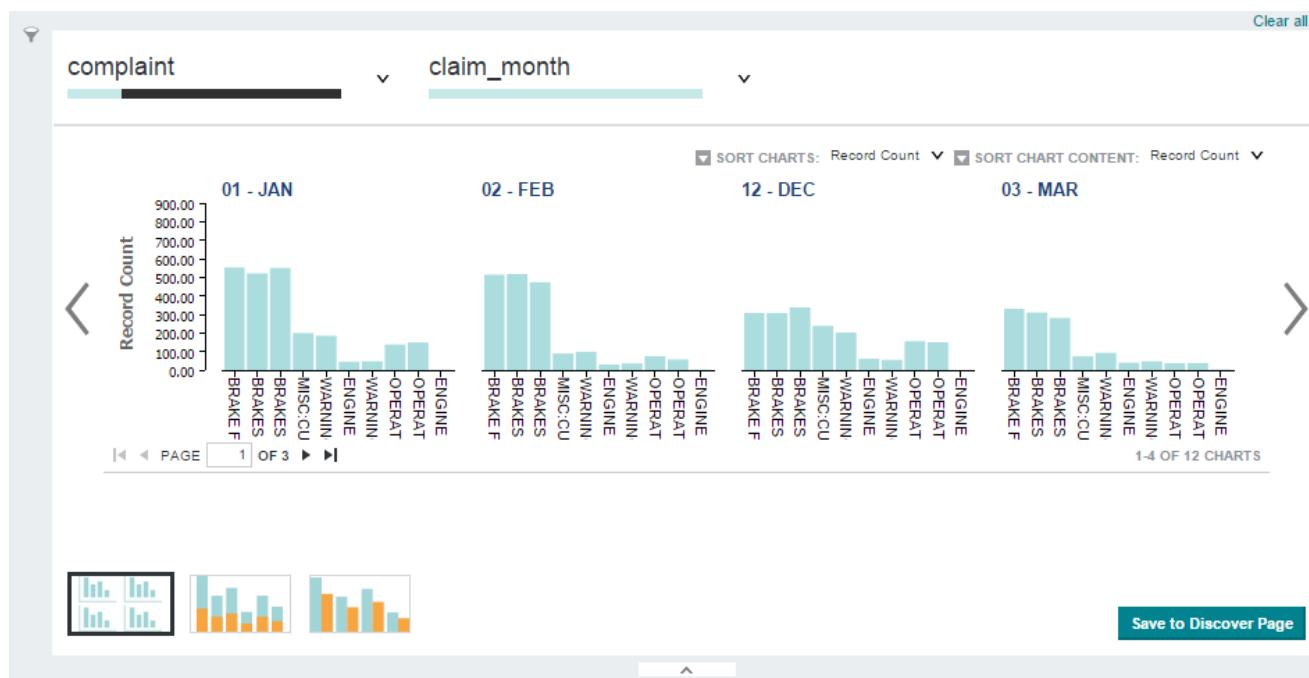
## Using the scratchpad

You can use the scratchpad component on the **Explore** page to try out different visualizations with one or more attributes before adding them to a **Discover** page.

### About the scratchpad

The scratchpad consists of:

- The scratchpad drop zone, which lists the attributes currently displayed on the scratchpad
- The scratchpad visualizations



The visualizations available in the scratchpad can include charts, box plots, histograms, and maps. These function similarly to the chart components; for information, see [Charts](#).



**Note:** The scratchpad cannot display more than 1000 records at once. You must refine or remove any attributes that exceed this limit.

The type and number of visualizations available for an attribute or combination of attributes varies. If the scratchpad contains only one attribute, it displays the same visualization as the attribute's tile and allows you to switch to the other visualizations available in the attribute's details panel. If the scratchpad contains multiple attributes, it displays the most appropriate visualization for the current set.

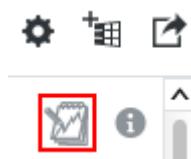


**Note:** Some combinations of attributes are incompatible and cannot be displayed by the scratchpad at all. If this occurs, you must try a different combination.

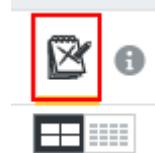
You can page through the available visualizations by clicking the arrows on either side of the scratchpad. You can also display a specific visualization by clicking its thumbnail at the bottom of the scratchpad.

## Displaying and hiding the scratchpad

The scratchpad displays when you add an attribute to it, or when you click the **Show Scratchpad** icon at the top right of the **Explore** page.



To hide the scratchpad, click either the **Hide Scratchpad** icon or the white arrow at the bottom of the scratchpad.

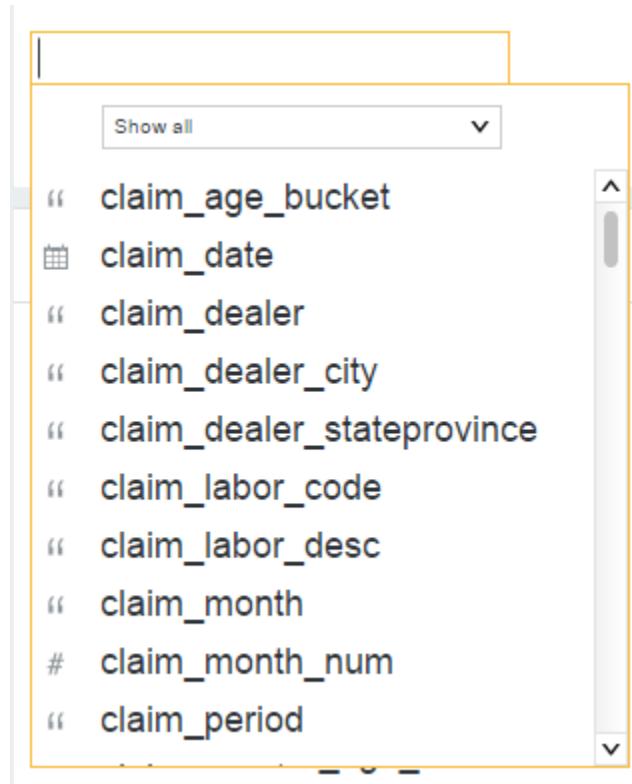


## Adding attributes

You can add attributes to the scratchpad in multiple ways:

- Hover your mouse over an attribute tile and click **Add to scratchpad**.
- Click the down arrow on an attribute tile to open its details panel and click **Add to scratchpad**.

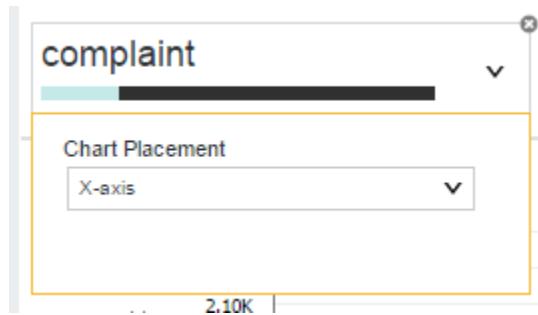
- Click anywhere within the scratchpad drop zone and either enter the name of the attribute you want to add or select it from the list of available attributes.



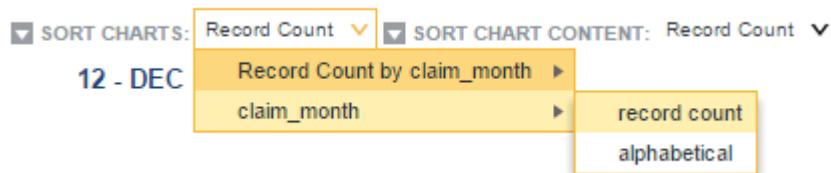
As you add attributes to the scratchpad, the other attributes may be sorted to put the most relevant first.

## Configuring visualizations

Each type of visualization is configurable; for example, you can change the chart placement and sort order of a histogram plot. To view options for configuring the visualization, click the down arrow next to the attribute's name in the scratchpad drop zone.



You can also modify the sort order using the **Sort** drop down menus, when available.



## Refining visualizations

Any refinements applied to your project are automatically applied to visualizations within the scratchpad. You can also click on specific values within the visualization to set them as refinements.

You can modify and remove your selected refinements using the **Selected Refinements** panel. For more information, see [Managing Selected Refinements](#).

## Removing attributes

To remove an individual attribute from the scratchpad, hover your mouse over its name in the drop zone and click the delete icon.

To remove all currently selected attributes, click **Clear all** in the top right of the scratchpad.

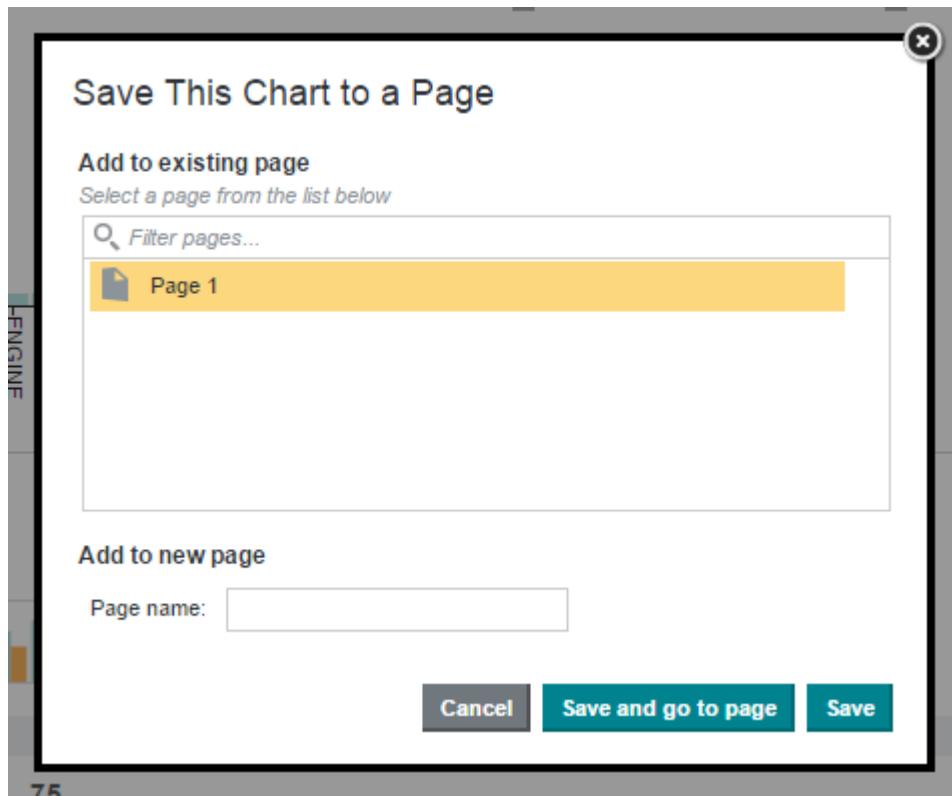
## Saving a visualization to a Discover page

If you are working within a project, you can save your visualization to a **Discover** page when you are satisfied with it.



**Note:** If you are not working within a project, you must create one before saving your visualization.

To save a visualization to a **Discover** page, click **Save to Discover** page, then either select an existing page to save it to, or provide a name for a new page.





## Chapter 14

# Deleting Data Sets

From the **Catalog**, you can delete data sets.

### *Deleting a data set*

If you are a User, Power User or Administrator, you can delete any data set from the **Catalog**. A user with the role Viewer cannot delete a data set.

You cannot delete a data set that is currently being used in a project.

To delete a data set, click **Delete** on the data set details panel.



**Note:** Deleting a data set from the **Catalog** does not delete its corresponding Hive table in the Hive database. For details about the relationship between a data set and its Hive table, see the *Data Processing Guide*.

# **Part VI**

## **Creating and Configuring Projects**



## Chapter 15

---

# Creating a project

### *About projects*

[Creating a project using the Catalog](#)

[Creating a project from within a data set](#)

## About projects

Any edit operation, such as transforming data, must take place within a project rather than in a data set. Creating a new project also provides several other features that are not available when exploring a data set. In a project you can:

- Create and save project pages in **Discover**.
- Add multiple data sets, and link them in the **Data Set Manager** editor. In other words, a project provides the structure to join, aggregate, and transform multiple data sets. Note however, that a project does not make a copy of a data set.
- Modify attribute metadata in the **Attribute Groups** editor. For example, you can preserve changes for display name settings or dimension settings as part of project.
- Create and modify data views in the **Project Settings>Data Views**.

There are several ways to create a project:

- From the **Catalog**, you can immediately create a new project before you explore or transform the data. This is a good option if you know you are going to make changes to the data and prefer to create a new project right away. In other words, you typically do this *early in the discovery process*.
- From **Explore**, you can create a new project after examining the source data. You typically do this *later in the discovery process*.
- Or, you can create a new project only after being prompted by Studio during an edit operation that you want to perform. You typically do this *only if required*.

Users with a role of Administrator, Power User, or User can create new projects. Users who only have the Restricted User role cannot create projects.

## Creating a project using the Catalog

To create a new project using the Catalog:

1. From the **Catalog**, select a data set.
2. Expand the data set preview menu and click **Add to Project**.
3. Provide a name and description for the new project.

4. Click **Add**.

The new project opens to the **Explore** page.

## Creating a project from within a data set

You can also create a new project while you explore a data set.

To create a new project from within a data set:

1. From **Explore** page, click **Add to Project**.
2. Provide a name and description for the new project.
3. Click **Add**.

The new project opens to the **Explore** page.



## Chapter 16

# Managing Project Pages

A project can contain any number of pages where you create components to visualize data.

[Adding a page to a project](#)

[Renaming a page](#)

[Changing the layout of a page](#)

[Changing the data set binding for a page](#)

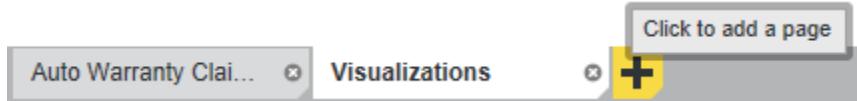
[Deleting a page](#)

## Adding a page to a project

When you first create a project, it contains a single empty page. From the project pages view, you can then add additional pages.

To add a page to a project:

1. Select **Discover**.
2. In the page footer, click the add page icon next to the page tabs.



3. On the **Add Page** dialog, type a name for your new page, and if your project contains multiple data sets, also select a primary data set to be the main focus for this page.

(This is page-level data binding. Any components you add to this page are bound to the primary data set by default.)

4. Click **Add**.

The new page is added to the project and then automatically displayed, so that you can add components to it.

See [Adding and Configuring Components on page 86](#).

## Renaming a page

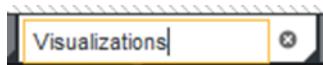
After you create a page, you can change the display name that appears on its tab.

When you rename a page, you are only changing the display name, so you do not have to change any existing page transitions.

To rename a page:

1. Click the page tab.
2. Click the page tab again.

The page name is displayed in an editable field.



3. In the field, type the new name.
4. To cancel the name change, and restore the previous name, press **Esc**.
5. To save the new name, click outside of the tab, or press **Enter**.

You can also use the **Localization** page to edit and localize the page name. See [Localizing the project name, description, page names, and component titles on page 129](#).

## Changing the layout of a page

You can select the layout to use for each page in your project.

The page layout determines how components are organized on each page. A page may have one, two, or three columns. One column width may be fixed, or all of the column widths may be proportional.

To change the layout of a page:

1. In a project, select **Discover**.
2. In the right bar, click the pencil icon to edit page settings.
3. Expand **Page Templates** and select the layout you want to apply to the page.

The new layout is applied to the page, with any existing components organized accordingly.

## Changing the data set binding for a page

Each page within a project is data bound to a specific data set. This approach provides a more granular page-level data binding rather than project-level data binding. If necessary, you can change the data set associated with a particular page.

To change the data set binding for a page:

1. Select **Discover**.
2. In the footer, select the page you want to modify.
3. In the right bar, click the pencil icon to edit page settings.
4. Expand **Page Data Settings** and select a different primary data set from the drop-down list.

## Deleting a page

You can delete pages from a project. When you delete a page, any components on the page also are deleted.

To delete a page:

1. Click the delete icon in the top corner of the page that you want to delete.



2. To delete the page, click **OK**.

When you delete a page, any bookmarks and snapshots created for that page are also deleted.



## Chapter 17

# Adding and Configuring Components

On each page in a project, you add and configure components. Components provide tools to explore and analyze the data.

[\*Adding and deleting components on a page\*](#)

[\*Renaming components\*](#)

[\*Configuring components\*](#)

[\*Selecting and configuring the data displayed on a component\*](#)

[\*Configuring component actions\*](#)

[\*Configuring pagination options for components\*](#)

[\*Previewing a component\*](#)

[\*Summary of available components\*](#)

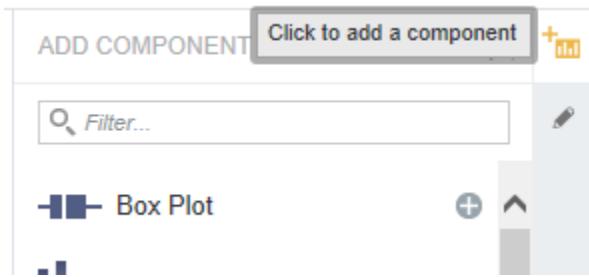
## Adding and deleting components on a page

If you are able to edit a project, then you can add and delete components. The components provide functions to display and create different visualizations of the project data.

To add and delete components on a page:

1. To add a component:
  - (a) Click the add component icon in the upper right of the panel.

For example:



- (b) To display a description of a component, hover the mouse over the component icon.
- (c) To find a specific component, use the search field.

- (d) To add a component to the page, either:
- Click the add icon for the component. If the page contains more than one column, then the component is added to the leftmost column.
  - Drag the component from the add page/component panel to the appropriate location on the page.
2. To delete a component from a page, click its delete icon.



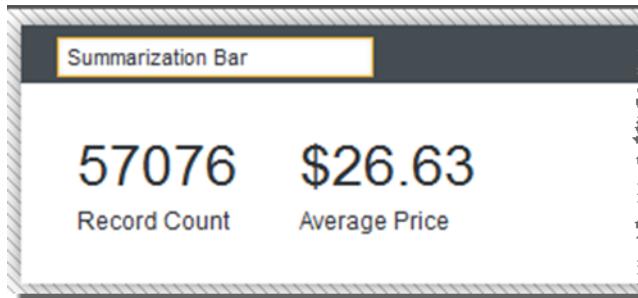
## Renaming components

After you add a component to your project, you can rename it to reflect your own terminology or the content being displayed.

To rename a component:

1. Click the title of the component.

The title becomes editable.



2. In the field, type the new title for the component.
3. Press **Enter**.

The component now uses the new title.

You can also use the **Localization** page to edit and localize the component title. See [Localizing the project name, description, page names, and component titles on page 129](#).

## Configuring components

You can customize components to determine details for the displayed data and formatting. Many components have a workable default configuration, while others require you to provide a specific configuration.

The general process for editing a component is the same for each component. For detailed information about configuring specific components, see the component documentation.

To configure a component:

1. In the component title bar, click the configuration icon.



The edit view for the component is displayed.

The edit view covers the entire page. You cannot see or edit any other components.

2. Select the data to use for the component.
3. Edit the other settings for the component.
4. To save your changes, click **Save**.
5. To revert any changes since the last save, click **Revert to Last Save**.
6. To exit the edit view, click the **Exit** link at the top right corner of the edit view.

If you have unsaved changes, then you are prompted to choose whether to save the changes before exiting, discard the changes before exiting, or cancel the exit and return to the edit view.

## Selecting and configuring the data displayed on a component

Most components display data from a specific view. You select the view to use, and then select and configure values from that view.

*[Selecting the view to use for a component](#)*

*[Selecting the attributes to use on a component](#)*

*[Selecting the aggregation method to use for a metric](#)*

*[Configuring the format of values displayed on a component](#)*

### Selecting the view to use for a component

When configuring a component, you select the view to use for the component data.

Most components are tied to a specific view from the project data sets. Other components can be associated with more than one view. For example, for the **Map** component, each map layer is tied to a view. For the **Summarization Bar** component, each summary item is tied to a view.

On the component edit view, to select the view to use:

1. Click the **Data Selection** tab for the component or sub-component.

The tab contains the list of views for the project. If a view is not valid for the current component or sub-component, then it is disabled.

For example, for the **Map** component, views that do not include a geocode attribute cannot be used for a map layer.

#### Bizwine Denormalized: Data Selection

Select the data view for this layer to use. Views that do not include a valid geocode attribute cannot be used for the map.

View Name	Type	Data Set(s)	?	Record Identifier	Description
<input checked="" type="radio"/> Bizwine Denorma...	Base	Bizwine Denorm...	?	Transaction Id	The classic bizwine data set, in which dimensional data about wine s...
<input type="radio"/> Customer	Custom	Bizwine Denorm...	?	Customer Id	List of customers with a summary of their credit and transaction infor...
<input type="radio"/> Customers - No g...	Custom	Bizwine Denorm...	?	Customer Id	List of customers with a summary of their credit and transaction infor...

In addition to the view name and description, the view list includes whether the view is a base or custom view, the data sets the view is created from, and the identifying attributes for the view.

2. To display the list of attributes in a view, click its information icon.
3. To select a view to use, click its radio button.

## Selecting the attributes to use on a component

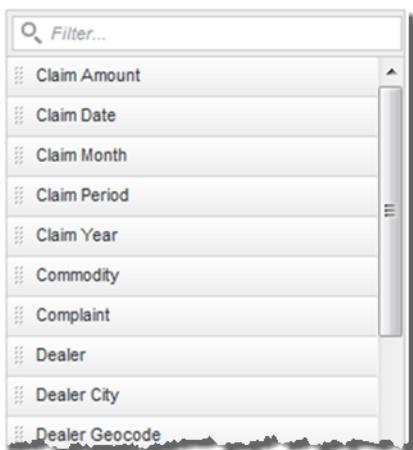
For many components, you need to select the attributes to use for the component.

If you only need to select a single value, then the component edit view includes a button or link to display the list of available attributes.

Metric: Select Metric...

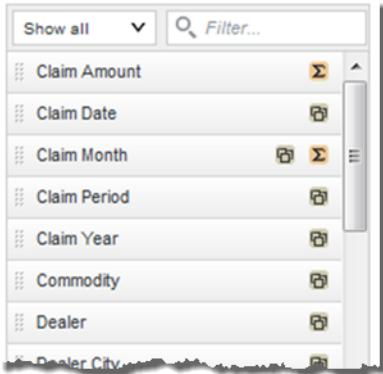
If you can set up lists of values, then the list of available attributes is displayed directly on the edit view, and you can drag and drop attributes from the list.

In some cases, such as **Results List** or a record list **Results Table**, you are simply selecting attributes to display, without any aggregation. The available attributes list then only displays attributes.



For other components, such as **Chart** and **Pivot Table**, you are selecting aggregated metric values, and the dimensions to use for the aggregation. For components that display aggregated data, the attributes list:

- Uses icons to indicate whether each attribute can be used as a metric or a dimension



Note that multi-value date/time attributes cannot be used as dimensions, and are not included in the list.

- Includes the predefined metrics from the selected view
- Includes the following system metrics:

Metric Name	Description
<b>Number of records</b>	<p>Displays the number of records in the underlying data set that have the combination of dimension values for the current row.</p> <p>For example, a component is aggregated using the dimensions Country and Product Line. The current item has United States for the country and Apparel for the product line.</p> <p>The number of matching records is the number of records that have the value United States for Country and Apparel for Product Line.</p> <p>This metric is displayed on the component as "Record Count".</p>
<b>Number of records with values</b>	<p>Displays the number of records that have a value (any value other than NULL) for a selected attribute for the current dimension values.</p> <p>When you add this metric, you are prompted to select the attribute to use. You can only select an attribute that has the count aggregation enabled.</p> <p>For example, a component is aggregated using Country and Product Line. The current item has United States for the country and Apparel for the Product line.</p> <p>Of the 15 records with United States as the Country and Apparel as the product line, 4 records have Color set to Blue, 3 have Color set to Green, and 8 have no value for Color.</p> <p>The number of records with values for the Color attribute is then 7.</p>

Metric Name	Description
<b>Number of unique values</b>	<p>Displays the number of unique values (other than NULL) for a selected attribute for the current dimension values.</p> <p>When you add this metric, you are prompted to select the attribute to use. You can only select an attribute that has the count distinct aggregation method enabled.</p> <p>For example, a component is aggregated using Country and Product Line. The current item has United States for the country and Apparel for the product line.</p> <p>Of the 15 records with United States as the Country and Apparel for the product line, 4 records have Color set to Blue, 3 have Color set to Green, and 8 have no value for Color.</p> <p>The number of unique values for Color is then 2.</p>

To select values for a component:

1. To select a single value:

- (a) Click the link or button.
- (b) On the attribute list dialog, click the attribute to select.

You can use the filter field to search for a particular attribute.

- (c) Click **Apply**.

When you select a metric attribute, Big Data Discovery assigns the default aggregation method for that attribute.

When you select the **Number of records with values** or **Number of unique values** metric, Big Data Discovery prompts you to select the attribute to use.

2. To populate lists of values for a component:

- (a) To add an attribute to a list, drag it from the available attributes list to the appropriate destination location on the component edit view.

You can use the filter field to search for a particular attribute.

When you begin dragging the attribute, Big Data Discovery highlights the locations where you are allowed to drop it.

For example, if you are dragging a metric, you cannot drop it into a dimensions list.

- (b) Drop the attribute into the destination location.

When you drop an attribute into a metrics list, Big Data Discovery assigns the default aggregation method for that attribute.

When you drop a **Number of records with values** or **Number of unique values** attribute, Big Data Discovery prompts you to select the attribute to use.

3. To configure a selected attribute, click its edit icon.
4. To remove a selected attribute, click its delete icon.

## Selecting the aggregation method to use for a metric

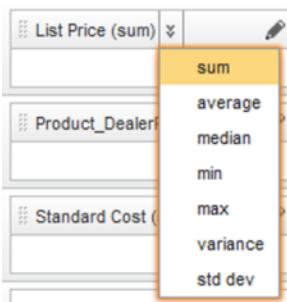
For metrics other than system metrics or predefined metrics, when you first select a metric to add to a component, Big Data Discovery uses the default aggregation method for that attribute. You can then select a different aggregation method.

For information on aggregation methods and how they work, see [Aggregation methods and the data types that can use them on page 207](#).

You configure the default and available aggregation methods for each attribute on the **Data Views** page. For information on configuring the available and default aggregation methods, see [Selecting the available and default aggregation methods for an attribute on page 205](#).

To change the aggregation method used for a metric value:

1. On the edit view of the component, the selected metric includes a drop-down icon to allow you to select a different aggregation method.



2. The configuration dialog for the metric also contains an aggregation method drop-down list you can use to change the aggregation method.



## Configuring the format of values displayed on a component

For values displayed on a component, you can configure the display format.

Displayed values on a component could be:

- Actual attribute values from a view
- Predefined metrics from a view
- Component-specific metrics calculated from view attribute values

For attributes and predefined metrics, the default display format is configured in the **Data Views** page. See [Configuring the default display format for an attribute on page 203](#). You can then either use this format, or use a customized format.

For component-specific metrics, the default display depends on the type of aggregation:

Aggregation Method	Default Metric Format
Sum	Default display format for the original attribute.
Average	
Median	
Min	
Max	
Set	
Arb	
Variance	Default format (adjusted for locale) for a decimal number.
Standard Deviation	
Count (Number of records with values)	Default format (adjusted for locale) for an integer number.
Count Distinct (Number of unique values)	

Configuration dialogs for displayed values can include a **Value Formatting** section to control the display format. In the drop-down lists and radio button labels, **Use default** indicates to use the default format configured on the **Data Views** page.

At the top of the display settings is a sample value showing the current configuration adjusted for your locale.

From the configuration dialog, to set the display format for a value displayed on a component:

1. To customize the display of a numeric value:

- If the number is formatted by default as a regular number, then you can change the display to be a currency value or a percentage. Note that if you select the percentage format, the value is multiplied by 100.

To change the type of formatting for a number, click the radio button next to the formatting type to use.

If the number is formatted by default as a currency or a percentage, then you cannot select a different type of number formatting.

- If you have changed the formatting of a number to be a currency value, then from the **Currency** drop-down list, select the type of currency.

If the number was originally formatted as a currency, then the currency selection is displayed as a read-only value.

(c) If the number is a percentage, then from the **Percentage** drop-down list, select whether to display the percentage (%) sign with the value.

(d) Under **Decimal Places**, to use the decimal place configuration from the **Data Views** page, leave the **Default** radio button selected.

To display the full raw number, up to 6 decimal places, click the **Up to 6 digits** radio button.

To specify the number of decimal places, click the **Fixed length** radio button, then in the field, type the number of decimal places to display.

(e) From the **Grouping Separator** drop-down list, select whether to display the grouping separator (used to separate thousands).

2. For numeric values, the **Advanced Formatting** section provides additional format options.

For each of the advanced items, you can choose to have the display use the default configured for the view, or select a specific option for this particular component. Under **Advanced Formatting**:

Advanced Formatting Changes below will override localized formatting based on each user's locale.

Number Format: **varies based on locale**

Decimal Separator: **varies based on locale**

Grouping Separator: **varies based on locale**

(a) From the **Number Format** drop-down list, select how to display negative numbers.

(b) From the **Decimal Separator** drop-down list, select the character to use for the decimal point.

(c) From the **Grouping Separator** drop-down list, select the character to use for the grouping (thousands) separator.

3. For a Boolean attribute, to set the values to display for each Boolean value:

See a sample true / false (United States | English) i

Basic Formatting Boolean Display:  Default  Custom Values

True:

False:

(a) Click the **Custom Values** radio button.

(b) In the **True** field, type the value to display if the attribute value is 1 (True).

(c) In the **False** field, type the value to display if the attribute value is 0 (False).

4. For a geocode, to set a specific number of decimal places to display for the latitude and longitude values:

▼ VALUE FORMATTING

See a sample 44.756530 -91.473097 (United States | English) i

Basic Formatting Decimal Places:  Default  Fixed length 6 digits

(a) Click the **Fixed length** radio button.

(b) In the field, type the number of decimal places.

5. For a date/time attribute:

- (a) From the **Date Display** drop-down list, select the format to use for the date.

For each format, the display is controlled by the locale. For example, for some locales, the month is displayed first, and for others the day is displayed first.

- (b) From the **Time Display** drop-down list, select the format to use for the time.

6. For a time attribute, from the **Time Display** drop-down list, select the format to use for the time.

7. If multiple values are displayed, either because the selected attribute is multi-value, or the set aggregation is being applied, then the configuration includes the **Multi-Value Formatting** section.

You can configure the number of values to display, and the separator to use between the values.

- (a) Under **Number of values to display**, to use the default, leave the **Default** radio button selected. To specify a number, click the **Maximum at** radio button, then type the number in the field.
- (b) From the **Multi-value Separator** drop-down list, select the character to use to separate the values.

## Configuring component actions

Part of configuring a component is configuring any available fort that component, such as refining data and displaying details. Some actions are specific to a displayed value, while others may apply to the entire component.

[Configuring the Actions menu for a component](#)

[Configuring actions for displayed values](#)

[Selecting the target page for a refinement or hyperlink](#)

[Configuring cascading for dimension refinement](#)

[Configuring hyperlinks to external URLs](#)

## Configuring the Actions menu for a component

A component **Actions** menu contains options that apply either to the entire component or to selected items from the component.

### About the component Actions menu

A component can include an **Actions** menu to allow end users to perform component-level actions such as printing the component and exporting data.

Some components also support options that require users to first select one or more items, such as comparing records, refining by attribute values in selected records, or using POST to pass parameters to a URL.

The available actions for a component **Actions** menu are:

Action	Description
<b>Print</b>	Allows users to print the component.
<b>Export</b>	Allows users to export data from a component.
<b>Advanced Export</b>	Allows users to export data from a component with more delimiter options than <b>Export</b> .
<b>Compare</b>	<p>Displays the <b>Compare</b> dialog to allow users to do a more detailed comparison of selected items.</p> <p>The <b>Actions</b> menu can only contain one <b>Compare</b> option, which is enabled by default.</p> <p>If the view used for the component does not have any identifying attributes, then the <b>Compare</b> checkbox is disabled and locked.</p>
<b>Pass Parameters</b>	<p>Allows users to create a hyperlink that generates an HTTP POST request containing attribute values from the selected records as POST parameters.</p> <p>The <b>Pass Parameters</b> option is only available for the <b>Results List</b> and <b>Results Table</b> components.</p> <p>The <b>Actions</b> menu can contain multiple <b>Pass Parameters</b> options, each with a different label, URL, and parameters.</p> <p>By default, there are no <b>Pass Parameters</b> options in the menu.</p>

Action	Description
<b>Refinement</b>	<p>Allows users to refine the data using selected attribute values from the selected records. The attributes must allow users to refine using multiple values (have multi-or or multi-and refinement behavior).</p> <p>For example, users could select three records and then refine the data to only include records that have values for available colors and available sizes that are used in those selected records.</p> <p>The <b>Refinement</b> option is only available for the <b>Results List</b> and <b>Results Table</b> components.</p> <p>The actions menu can contain multiple <b>Refinement</b> options, each with a different label and attributes.</p> <p>By default, there are no <b>Refinement</b> options in the menu.</p>

## Selecting the actions to include in the Actions menu

On the **Actions** tab of the component edit view, you use the **Actions menu** list to select the available options in the **Actions** menu. For some actions, you can only enable or disable them. For other actions, you can add one or more instances to the menu.

To select the options to include in the **Actions** menu:

1. To exclude the **Print** option, uncheck the **Print** checkbox. To restore the **Print** option in the **Actions** menu, check the checkbox.

There are no configuration options for the **Print** option.

2. To exclude the **Export** option, uncheck the **Export** checkbox. To restore the **Export** option in the **Actions** menu, check the checkbox.

There are no configuration options for the **Export** option.

3. To exclude the **Compare** option, uncheck the checkbox. To restore the **Compare** option in the **Actions** menu, check the **Compare** checkbox.

For the **Compare** option, you can configure the label to display in the column heading for each record. To configure the column heading:

- (a) Click the edit icon for the **Compare** option.

- (b) On the **Compare Configuration** dialog, in the **Header** field, type the header to display.

The default is `Record {counter}`, which indicates to display the word "Record" plus the column number based on the order in which the records were selected. For example, Record 1, Record 2, etc.

In addition to the `{counter}` token, you can also use the `{attributeKey}` token, where `attributeKey` is the key name (not the display name) of an attribute for which to display the associated value. For example, entering `{Name}` would indicate to use the value of the `Name` attribute for the record as the column heading.

The attribute must be in a group that is configured to display for record details. If the attribute is not included in the attributes on the **Compare** dialog, then the value is not displayed in the column heading.

4. The **Actions menu** list initially does not contain any **Pass Parameters** options. You can add one or more of these options, each with unique settings:
  - (a) To add a **Pass Parameters** option, click the **+Action** button. From the drop-down menu, select **Pass parameters**.
  - (b) By default, the option is enabled when you add it. To exclude the option, uncheck its checkbox. To reenable the option, check the checkbox.
  - (c) To remove a **Pass Parameters** option, click its delete icon.
5. The **Actions** menu list initially does not contain any **Refinement** options. You can add one or more of these options, each with unique settings:
  - (a) To add a **Refinement** option, click the **+Action** button. From the drop-down list, select **Refinement**.
  - (b) By default, the option is enabled when you add it. To exclude the option, uncheck its checkbox. To reenable the option, check the checkbox.
  - (c) To remove a **Refinement** option, click its delete icon.
6. To change the display order of the **Actions** menu options, drag each option to the appropriate location in the list.

## Configuring a Pass Parameters Actions menu action

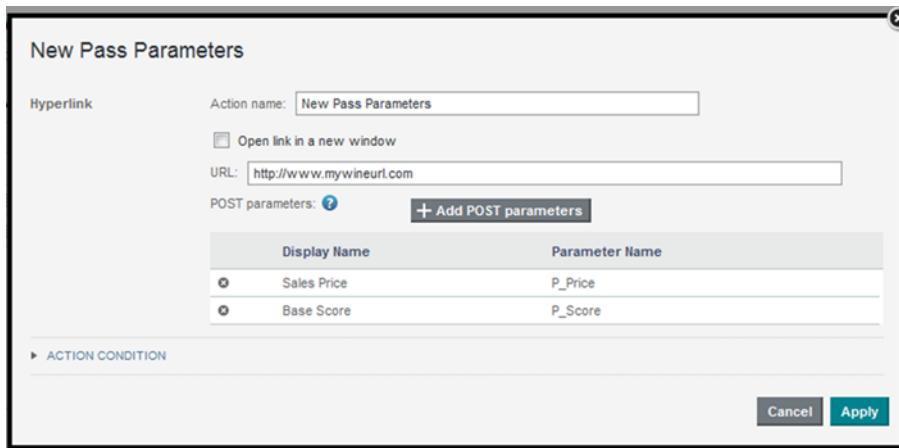
For a **Pass Parameters** option in a component **Actions** menu, you can configure the URL and whether to display it in a new browser window. You can also include attribute values as POST parameters.

To configure a **Pass Parameters** option:

1. Click its edit icon.
2. On the configuration dialog, in the **Action name** field, type the label to use for this option.
3. To display the URL in a new browser window, check the **Open link in a new window** checkbox.
4. In the **URL** field, type the URL to post to.  
Make sure that the URL is correctly formed, and that special characters are properly encoded. You must provide the full URL, starting with the protocol (HTTP, HTTPS, etc.).
5. To add attribute values as POST parameters:
  - (a) Click the **+ Add POST parameters** button.  
The list of available attributes is displayed.
  - (b) For each attribute you want to add, check its checkbox.

- (c) When you are finished selecting the attributes to add, click **Apply**.

The selected attributes are displayed in a list.



- (d) To remove a parameter, click its delete icon.

6. To save the action configuration, click **Apply**.

## Configuring a Refinement Actions menu action

For a **Refinement** option in a component **Actions** menu, you can configure the attributes to use for the refinement and the page on which to apply the refinement.

To configure a **Refinement** option:

1. Click its edit icon.
2. On the configuration dialog, in the **Action name** field, type the label to use for this option.
3. In the attribute list, check the checkbox next to each attribute to include in the refinement.

Note that you can only select attributes that allow you to select multiple values for refinement. On the **Data Views** page, these attributes have **Refinement Behavior** set to **Multi-Or** or **Multi-And**.

4. Use the **Target Page** setting to specify the page on which to execute the refinement.

To stay on the current page, click the **Current page** radio button.

To navigate to a different page, click the **Other page** radio button, then in the field, specify the page.

For information on specifying a different page in the project, see [Selecting the target page for a refinement or hyperlink on page 102](#).

5. To save the configuration, click **Apply**.

## Configuring actions for displayed values

For many components, you can allow users to click displayed values in order to refine data, display record details, or navigate to a specified page or external URL.

When configuring a displayed value, there is an **Actions** section to configure the action.

To select and configure an action for a displayed value:

1. If more than one action is available, then from the action drop-down list, select the action.

The available actions are:

Action Name	Description
<b>No Action</b>	Indicates that the value is not hyperlinked. For attributes that do not support refinement, this is the default option.
<b>Show Details</b>	When users click the value, the <b>Record Details</b> dialog is displayed and is populated with the details for that record or row.
<b>Refinement</b>	When users click the value, the data is refined to only include records with that value. If the displayed attribute supports refinement, then this action is selected by default.
<b>Hyperlink</b>	When users click the value, they navigate to the specified URL. A hyperlink can be to another page in the same project, or to an external URL. Hyperlinks to external URLs can include attribute values as parameters.

2. For a **Show Details** action, use the **Window Name** setting to configure the title of the **Record Details** dialog.
3. For a **Refinement** action, use the **Target Page** setting to specify the page to navigate to when users click the value.

To stay on the current page, click the **Current page** radio button.

To navigate to a different page, click the **Other page** radio button, then in the field, specify the page.

For details on providing a target page for refinement, see [Selecting the target page for a refinement or hyperlink on page 102](#).

4. For a **Hyperlink** action:
  - (a) To display a tooltip for the value containing a description of the action, check the **Display action description in tooltip** checkbox. In the **Action Description** field, type a description of the hyperlink action.
  - (b) To display the destination in a new browser window, check the **Open link in a new window** checkbox. The box is checked by default.
  - (c) In the **URL** field, type the hyperlink destination. The link can be to:
    - A different page in the same project. For details on providing a target page for a hyperlink, see [Selecting the target page for a refinement or hyperlink on page 102](#).

- An external URL. For an external URL, make sure the URL is correctly formed, and that you have properly encoded special characters.

You must provide the full URL, starting with the protocol (HTTP, HTTPS, FTP, etc.)

A hyperlink to an external URL can include attribute values.

The values may be query parameters:

```
http://www.acme.com/index.htm?p1="Red"&p2="1995"
```

Or may be part of the URL path:

```
http://www.acme.com/wines/1995/
```

To add attribute values to an external URL:

- Click **+ Add URL Parameters**.
- On the add parameters dialog, in the attribute list, check the checkbox next to each attribute to add.
- When you are finished selecting attributes, click **Apply**.

The selected attributes are displayed in a table, with each attribute assigned an ID to use when inserting the attribute into the URL.

The attributes are also inserted as query parameters, where the parameter name is the attribute key, and the parameter value is `{IDNumber}`, where `IDNumber` is the ID for that attribute. For example: `http://www.acme.com/index.htm?Region={0}&WineType={1}`

By default, the value is encoded. To not encode the value, change the format to `{ {IDNumber} }`. For example: `{ {0} }`

You can also use the ID numbers to insert the attribute values manually.

For details on component hyperlinks and encoding inserted attribute values, see [Configuring hyperlinks to external URLs on page 105](#).

- To remove a URL parameter from the table, click its delete icon.

If you did not edit the inserted query parameter, then Big Data Discovery also removes it from the URL.

If you did edit the inserted query parameter, then you must remove the parameter from the URL manually.

If you inserted the attribute value manually, then you also must remove it manually.

## Selecting the target page for a refinement or hyperlink

You can configure refinement actions to display a different page. You can also create hyperlink actions that navigate to a different page.

For example, when users refine by a value on a **Results Table** component on one page, users could be redirected to a different page. Note, however, that the refinement affects all of the components where it applies, no matter what page the components are displayed on.

On the component edit view, the target page setting for a refinement action uses radio buttons to indicate whether to stay on the current page or navigate to a different page.



## Using the internal page name

When specifying the target page for a refinement or hyperlink, you must use the internal page name used in the URL, not the display name used on the page tab.

Big Data Discovery creates the internal name automatically when you create the page. The internal name removes spaces and special characters.

For example, if the page name displayed on the page tab might be **Data Results**, then the page name in the URL might be **data-results**.

While you can change the display name for a page, the internal page name does not change.

## Selecting a tab on a Tabbed Component Container

If the target page includes a **Tabbed Component Container** component, then to specify the tab that is selected, you append to the page name:

```
#tabComponentName[tabNumber]
```

Where:

- *tabComponentName* is the name of the **Tabbed Component Container**.
- *tabNumber* is the number (1, 2, 3, etc.) of the tab to select.

So for example, for the following target value:

```
analyze#Sales Numbers[1]
```

- The end user is redirected to the **analyze** page.
- On the page, the first tab of the **Sales Numbers** tabbed component is selected.

To select the tab to display for multiple tabbed components, use a double colon (:) to delimit the components.

For example, for the following target value:

```
analyze#Sales Numbers[1]::Quarterly Forecast[2]
```

- The user is redirected to the **analyze** page.
- On the **Sales Numbers** tabbed component, tab 1 is selected.
- On the **Quarterly Forecast** tabbed component, tab 2 is selected.

## Using component IDs to specify a Tabbed Component Container

Because the double colon (:) is part of the target page syntax, you should avoid using it in your tab titles. You also should avoid multiple tabbed component containers with duplicate titles.

If you cannot avoid these naming features, then when defining a target page, you must use a component's ID rather than its name.

To find a component's ID:

1. Hover your mouse over the tab until the URL appears in the browser status bar.
2. Extract the `p_p_id` parameter from the URL.

`/dashboard?p_p_id=nested_tabs_INSTANCE_7r3D&p_p_lifecycle=1&p_p_state=`

For example, for the following target value:

```
analyze#nested_tabs_INSTANCE_0CbE[2]::nested_tabs_INSTANCE_Ja6E[1]
```

- The end user is redirected to the `analyze` page.
- On the tabbed component with ID `nested_tabs_INSTANCE_0CbE`, tab 2 is selected.
- On the tabbed component with ID `nested_tabs_INSTANCE_Ja6E`, tab 1 is selected.

## Configuring cascading for dimension refinement

When users refine by dimension values on a component, you can configure the dimension to use cascading.

Cascading means that when the data is refined to a single value for the dimension value, the component is updated to use a different dimension.

When there are no more dimensions to cascade to, the component remains on the last dimension in the cascade.

For example, a component includes a Country dimension. The Country dimension is configured to cascade to State and then to Supplier. With this configuration:

1. When users refine the data to only show records for the United States, the component uses the State dimension (for states within the United States).
2. If users then refine the data to only show records for California, the component uses the Supplier dimension (for suppliers within California).
3. If users then refine the data by a specific supplier, the component displays the data for the selected supplier, and the cascade stops.

Note that for multi-or or multi-and dimensions, where users can refine by more than one value, then if the data is refined by any one of the dimension values, the cascade continues to the next dimension, even if there are still available values to refine by.

When configuring a dimension, to configure cascading for refinement:

1. In the **Dimension Cascade** section, check the **Enable dimension cascade** checkbox.

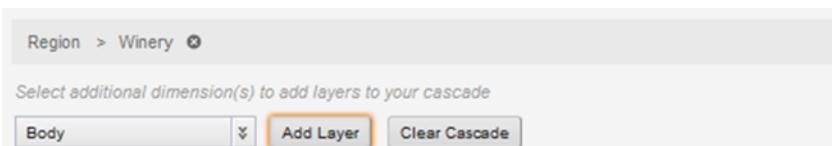
The cascade configuration is displayed. The current dimension is automatically at the top of the cascade.



A drop-down list is displayed for you to select the next dimension in the cascade.

2. To add a new layer to the cascade, select the dimension for that layer from the drop-down list, then click **Add Layer**.

The new layer is added to the end of the cascade.



3. To remove a layer from the cascade, click the delete icon next to that layer.
4. To clear the entire cascade, click **Clear Cascade**.

## Configuring hyperlinks to external URLs

Components can include hyperlinks to external URLs, which can present security risks.

### About component hyperlinks

Some components can include hyperlinks from the displayed records to external URLs.

In some cases, the URLs are static, with the exact same target displayed for every record.

More commonly, a URL contains one or more variable placeholders, each representing an attribute from the underlying data. At runtime, Big Data Discovery replaces the placeholders with the actual attribute values for the current record.

For example, for a list of products, a **Results List** component might include a hyperlink to the product web page. To find the correct page, the hyperlink includes the product identifier as a dynamic URL parameter.

When configuring these hyperlinks and inserting dynamic URL parameters, you should be aware of the security risks and recommendations associated with them, including:

- Encoding versus not encoding the inserted values
- Using non-HTTP (file or FTP) links

- Storing full URLs as attribute values

## Indicating whether to encode inserted attribute values

When you insert an attribute value as a URL parameter, you can specify whether Big Data Discovery encodes the value before inserting it into the URL.

URLs must use ASCII characters, and must avoid using certain reserved characters used as part of the URL syntax. URL encoding replaces any non-conforming characters with a percent sign followed by a hexadecimal numeric representation of the character.

For example, the forward slash (/) is a reserved character. If a URL includes the phrase "CD/DVD", then URL encoding would change that to "CD%2FDVD".

To insert parameters into a URL, you first select a list of attributes, each of which is assigned a number. You then use the number to insert the parameter.

ID	Display Name
0	Price Range
1	Score Range
2	Winery

The format uses the single and double braces to indicate encoding versus not encoding. By default, the values are encoded, and the parameters are enclosed by a single set of braces:

`{parameterNumber}`

For example:

`http://www.acme.com/index.htm?p1={0}&p2={1}`

To have Big Data Discovery not encode a hyperlink parameter, use a double set of braces:

`{ {parameterNumber} }`

For example:

`http://www.acme.com/index.htm?p1={{0}}&p2={{1}}`

It is important to understand that avoiding encoding can pose a security risk, because data is rendered in the user's browser without intervention, potentially allowing scripts or HTML to be injected directly into a user's browser.

Similarly, if the attribute contains reserved characters that are disallowed in a URL, the resulting hyperlink may be invalid or incorrect and therefore not work for the end user.

As a result, using the non-encoding syntax is not recommended. Unless you know that the underlying data is already encoded, you should always use the encoding syntax.

## About using non-HTTP protocol links

While Big Data Discovery does not prevent you from inserting links that use a non-HTTP protocol, such as file:// or ftp://, using these types of links is not recommended.

Most modern browsers are configured by default to disallow file URLs on web pages retrieved over HTTP. This is a security precaution to ensure web pages cannot execute files on a user's system.

When users click these types of links, the browser ignores them and does not display the link target.

Similarly, not all browsers are guaranteed to support FTP links, and users may not have FTP clients available to get access to content hosted over this protocol.

If users need to reach content on a file system or FTP server, it is recommended that you set up a configuration such as a web servlet to stream files through an HTTP interface. This allows all of your links to be secure HTTP links, and ensures consistent access for users with all browser configurations.

## Recommendations for attribute values that are complete URLs

It is possible that your source data contains complete URLs (including the protocol, host, port, and path) that are ingested when the data is loaded. For example, company data could include links to pages on the company web site.

If you store the full value, and then insert the value as the hyperlink URL, you would not be able to encode it. Because Big Data Discovery does not parse the attribute value as part of an absolute URL, encoding it would corrupt the URL, and the link would not work.

For example, `http://www.mycompany.com/page1` would become  
`http%3A%2F%2Fwww.mycompany.com%2Fpage1`.

Because of this, storing the full URL in your data is not recommended.

For these types of attributes, during the data ingest process, it is recommended that you use one of the following approaches:

### 1. Use one or more attributes to store only the parameter values for each record's URL.

When you configure the URL in a component, you would then manually type the standard part of the URL into the component configuration, and use encoded attributes for the query string parameters. For example:

`http://server.mycompany.com/path/to?file={0}`

0 is a number from the list of selected parameters.

You can only use this approach if all of the URLs have the same structure. If this is not the case, then you may need to use one of the other approaches.

### 2. Store the structural portions of the URL in a separate attribute from the parameter values.

In this approach, the structural portions of the URL such as the protocol, hostname, port, and context path/delimiters are stored in one or more attributes. The parameter portions of the URL that represent identifiers are stored in separate attributes as in approach 1 above.

When you enter the URL in the component, you would not encode the structural attributes, but would encode the parameters. For example:

`{{0}}/path/to/{{1}}?file={{2}}`

0, 1, and 2 are numbers from the list of selected parameters.

3. Use a single attribute for the full URL, but have the data ingest process encode any non-structural portions of the URL, such as query string parameters.

This prevents script injection and addresses any disallowed characters. When you enter the URL, you can then use the attribute value without further encoding. For example:

{ { 0 } }

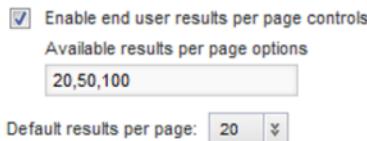
0 is a number from the list of selected parameters.

## Configuring pagination options for components

For components that allow end users to page through the displayed data, you can configure the number of results to display per page, and whether to allow end users to change that number.

On the **Display Options** tab of the component edit view, to configure the pagination options:

1. To allow end users to select the number of results to display on each page, check the **Enable end user results per page controls** checkbox.



2. If you are allowing end users to select the number of results per page, then:
  - (a) In the **Available results per page options** field, type a comma-separated list of the options.
  - (b) From the **Default results per page** drop-down list, select the default number of results to display per page.
3. If you are not allowing end users to select the number of results per page, then in the **Results per page** field, type the number of results to display per page.

## Previewing a component

When configuring a component, once the minimum configuration is established, a preview of the end user view is displayed.

The preview is designed to show the default display for the component.

Not all of the component functions are necessarily available. For example, you can't print or export from the component preview.

To preview the component:

1. To display the component preview, click the **Show Preview** button.

The button is displayed at the bottom of the component edit view, and is only enabled when the component is configured enough for the preview to display.

The **Show Preview** button changes to **Hide Preview**.

The component preview is displayed below the preview buttons.

When the preview is displayed, it is updated automatically whenever you make changes to the component.

- When the preview is being updated, a **Cancel** link is displayed. To stop updating the preview, click the **Cancel** link.

The update stops, and an **Update** button is displayed.

To complete the update, click the **Update** button.

- To hide the preview, click **Hide Preview**.

The **Hide Preview** button changes back to **Show Preview**. If the **Update** button is currently displayed, it is removed.

## Summary of available components

Here is a brief overview of each of the components. The components are grouped by the type of function they provide.

*Containers*

*Data record displays*

*Charts*

*Other data visualizations*

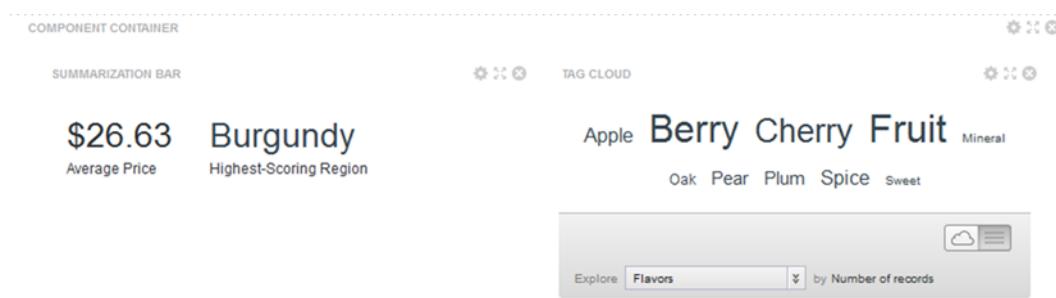
*Web-based content*

## Containers

Containers allow users to group components on the page.

### Component Container

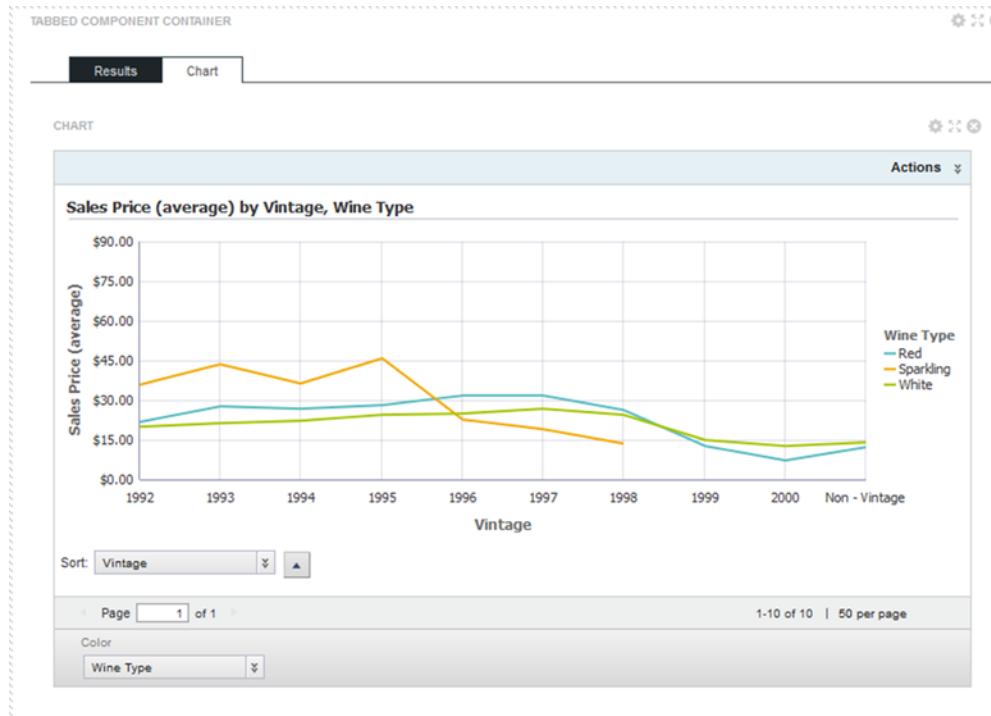
Used to group components on the page.



For the **Component Container**, users can select the layout to use.

## Tabbed Component Container

Used to create a tabbed interface. Each tab can contain a different set of components.



The **Tabbed Component Container** configuration includes:

- The number of tabs to display
- The tab display order
- The label on each tab
- The layout of each tab

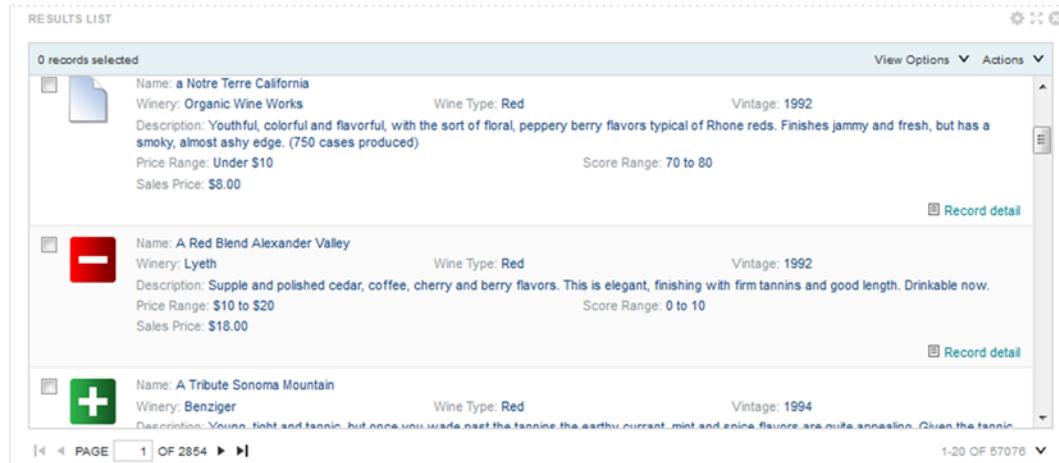
## Data record displays

Data record displays allow users to view lists of records.

### Results List

Displays the list of records for the current refinement. The list format is similar to regular Web search results. Each record contains a selected set of attributes.

Users can export and print the list. The component can also include options to compare records, display details for a record, and refine the data.



The screenshot shows a 'RESULTS LIST' component with the following details:

- 0 records selected**
- View Options** and **Actions** buttons
- Record detail** links for each record
- Record 1:** Name: a Notre Terre California, Winery: Organic Wine Works, Wine Type: Red, Vintage: 1992, Description: Youthful, colorful and flavorful, with the sort of floral, peppery berry flavors typical of Rhone reds. Finishes jammy and fresh, but has a smoky, almost ashy edge. (750 cases produced), Price Range: Under \$10, Score Range: 70 to 80, Sales Price: \$8.00
- Record 2:** Name: A Red Blend Alexander Valley, Winery: Lyeth, Wine Type: Red, Vintage: 1992, Description: Supple and polished cedar, coffee, cherry and berry flavors. This is elegant, finishing with firm tannins and good length. Drinkable now., Price Range: \$10 to \$20, Score Range: 0 to 10, Sales Price: \$18.00
- Record 3:** Name: A Tribute Sonoma Mountain, Winery: Benziger, Wine Type: Red, Vintage: 1994, Description: Young, light and tannic, but once you Wade past the tannins, the earthy, currant, mint and anise flavors are quite appealing. Given the tannin

Pagination: 1 OF 2854 | 1-20 OF 57076

The **Results List** component configuration includes:

- The attributes to include for each record, including whether to use the attributes for refinement or to display related content
- The sorting options for the list
- The images to display next to each record
- The display and pagination options for the list

## Results Table

Displays a table containing either:

- A list of records for the current refinement



The screenshot shows a 'RESULTS TABLE' component with the following details:

- Identification** dropdown and **0 RECORDS SELECTED** button
- VIEW OPTIONS** and **ACTIONS** buttons
- Wine ID** column is sorted **Wine ID ▲**
- Records:**
  - 34707: Alentejo Conventual, Red, 1992, Adega Cooperativa de Po...
  - 34708: Alentejo Monte Velho, Red, 1992, Herdade do Esporao...
  - 34709: Alentejo Tinto da Anfora, Alentejo, Red, 1992, J.P. Vinhos...
  - 34710: Alentejo Tinto da Talha, Alentejo, Red, 1992, Roquevale...
  - 34711: Alentejo Vinha do Monte, Red, 1992, Sogrape...
  - 34712: Alentejo, Red, 1992, Adega Cooperativa de B...
  - 34713: Alliance California, Red, Rhone Blend, 1992, R.H. Phillips...
  - 34714: Aloxe-Corton, Burgundy Cote de Be..., 1992, Bouchard Pere & Fils...
  - 34715: Aloxe-Corton, Burgundy Cote de Be..., 1992, Jaffelin...
  - 34716: Aloxe-Corton, Burgundy Cote de Be..., 1992, Louis Latour...
- Pagination: 1 OF 2854 | 1-20 OF 57076

- A set of metrics aggregated by one or more dimensions



The screenshot shows a 'RESULTS TABLE' component with the following configuration and data:

- RESULTS TABLE** (Header)
- 0 RECORDS SELECTED** (Header)
- VIEW OPTIONS ▾** and **ACTIONS ▾** (Header)
- Columns:** Region, Vintage, Record Count, Price (average) ▾, Review Score (aver... ▾)
- Data Rows:**
  - Alsace (1992, 116, \$18.95, 84.04)
  - Alsace (1993, 105, \$35.15, 86.13)
  - Alsace (1994, 145, \$37.09, 85.88)
  - Alsace (1995, 188, \$32.99, 86.23)
  - Alsace (1996, 260, \$28.80, 86.29)
  - Alsace (1997, 270, \$30.36, 85.46)
  - Alsace (1998, 196, \$23.83, 85.30)
  - Alsace (1999, 11, \$12.73, 83.55)
  - Alsace Summary** (1,291, \$29.46, 85.67)
  - Argentina** (4009, 10, \$10.17, 81.48)
  - Grand Summary (474 results)** (55,099, \$26.74, 84.84)
- Pagination:** 1 OF 28 (Page 1-20 of 555)

Users can export and print the list. They also may be able to use the **Compare** dialog to compare selected rows.

The table can include links to display the **Record Details** dialog, refine the data set, or display related content. Selected values also may be highlighted.

The **Results Table** component configuration includes:

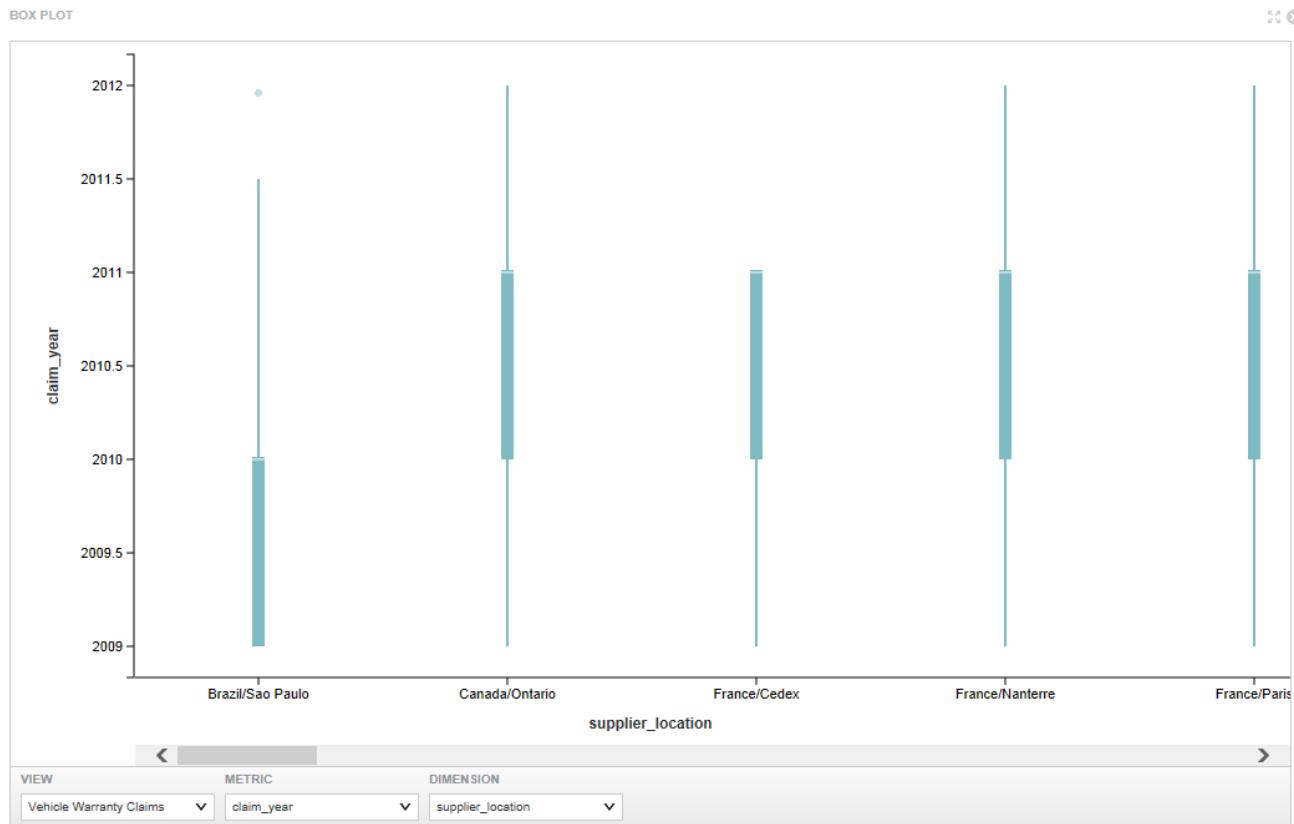
- The type of table
- The columns to display in the table
- Whether to highlight values that fall within selected ranges
- The available actions for the table
- The pagination and sorting options for the table

## Charts

Big Data Discovery provides several types of charts to use to visualize data and the distribution of data values.

### Box Plot

A **Box Plot** provides a capsule summary of values for a metric for each value of another attribute. For example, you can show a summary of the values for Number Ordered against each value of Product Category.



For a **Box Plot**, users can select:

- The view to use
- The metric to plot on the Y axis
- The dimension to plot on the X axis

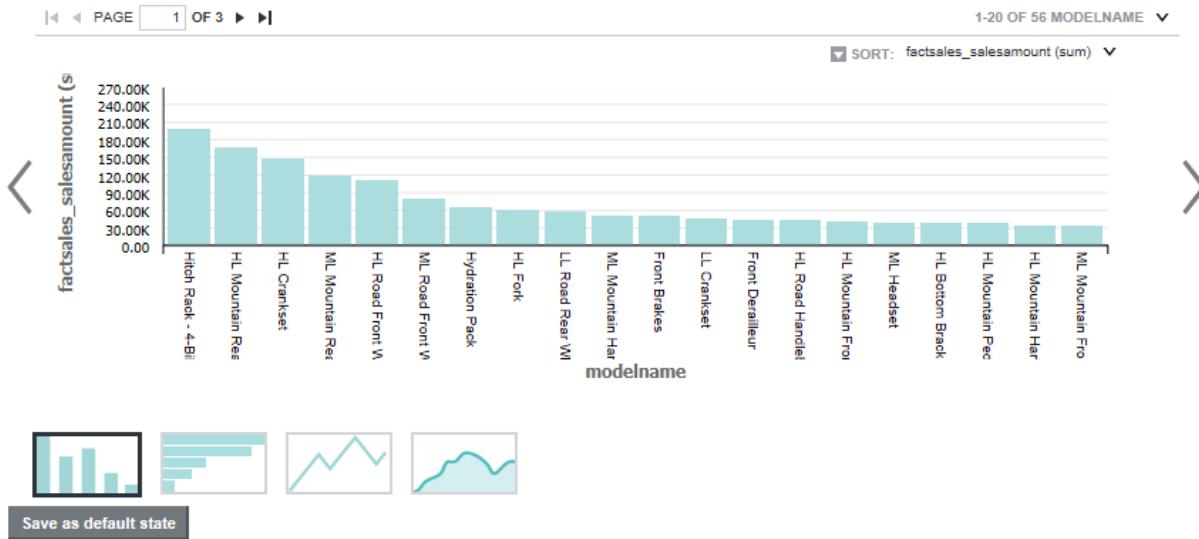
### Chart component

The **Chart** component displays data using standard chart formats, including:

- Bar charts
- Line charts
- Area charts

- Pie charts
- Scatter charts (including binning as an option to scatter charts)
- Bubble charts

Users may be able to select different metric or dimensions in order to change the chart display. They can also save the currently displayed chart as an image.

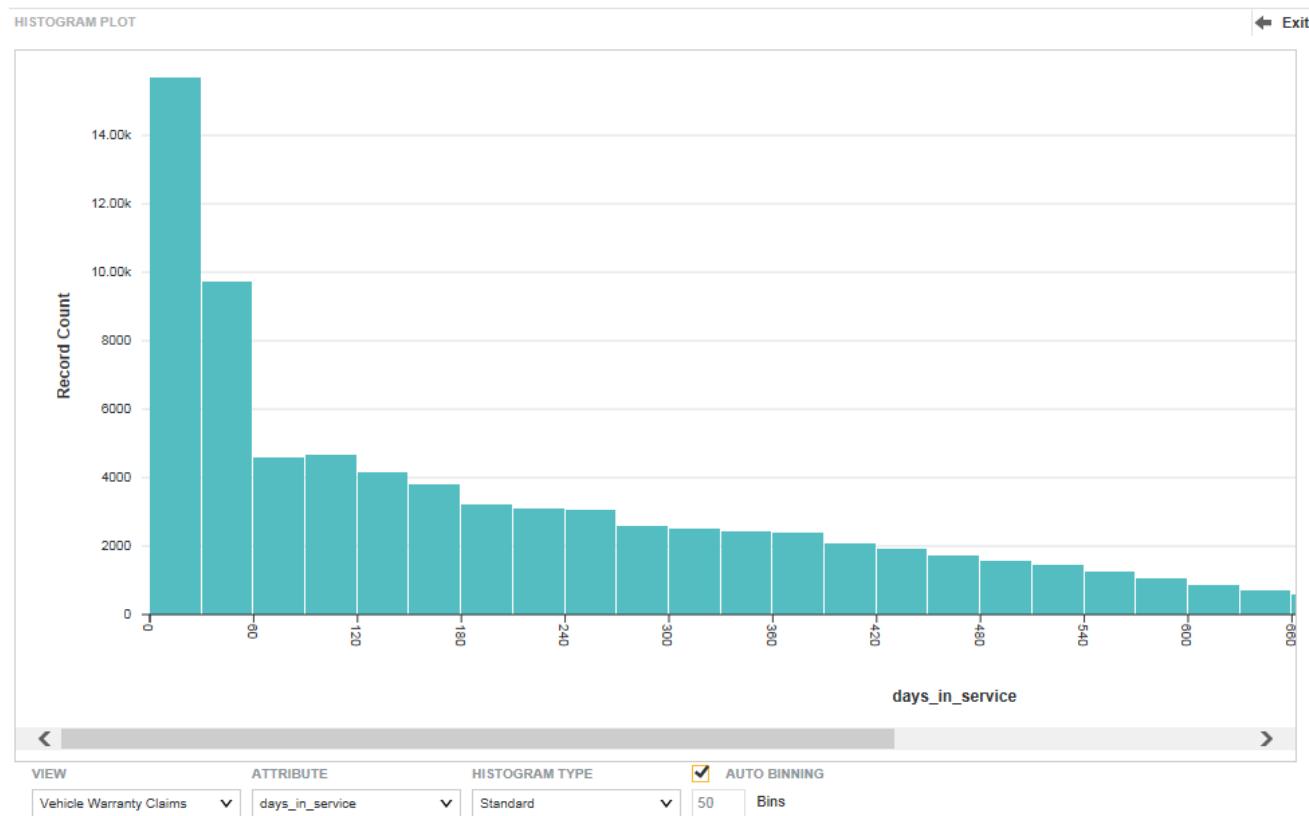


The **Chart** component configuration includes:

- The chart type
- The dimensions and metrics to display on the chart
- The chart style, including options for displaying the chart legend and axes
- Whether end users can change the chart metrics and dimensions

## Histogram Plot

A **Histogram Plot** is a basic visualization for showing the distribution of values for a single metric. The values are grouped into bins, each representing a range of values. Users can then see the number of values within each range.

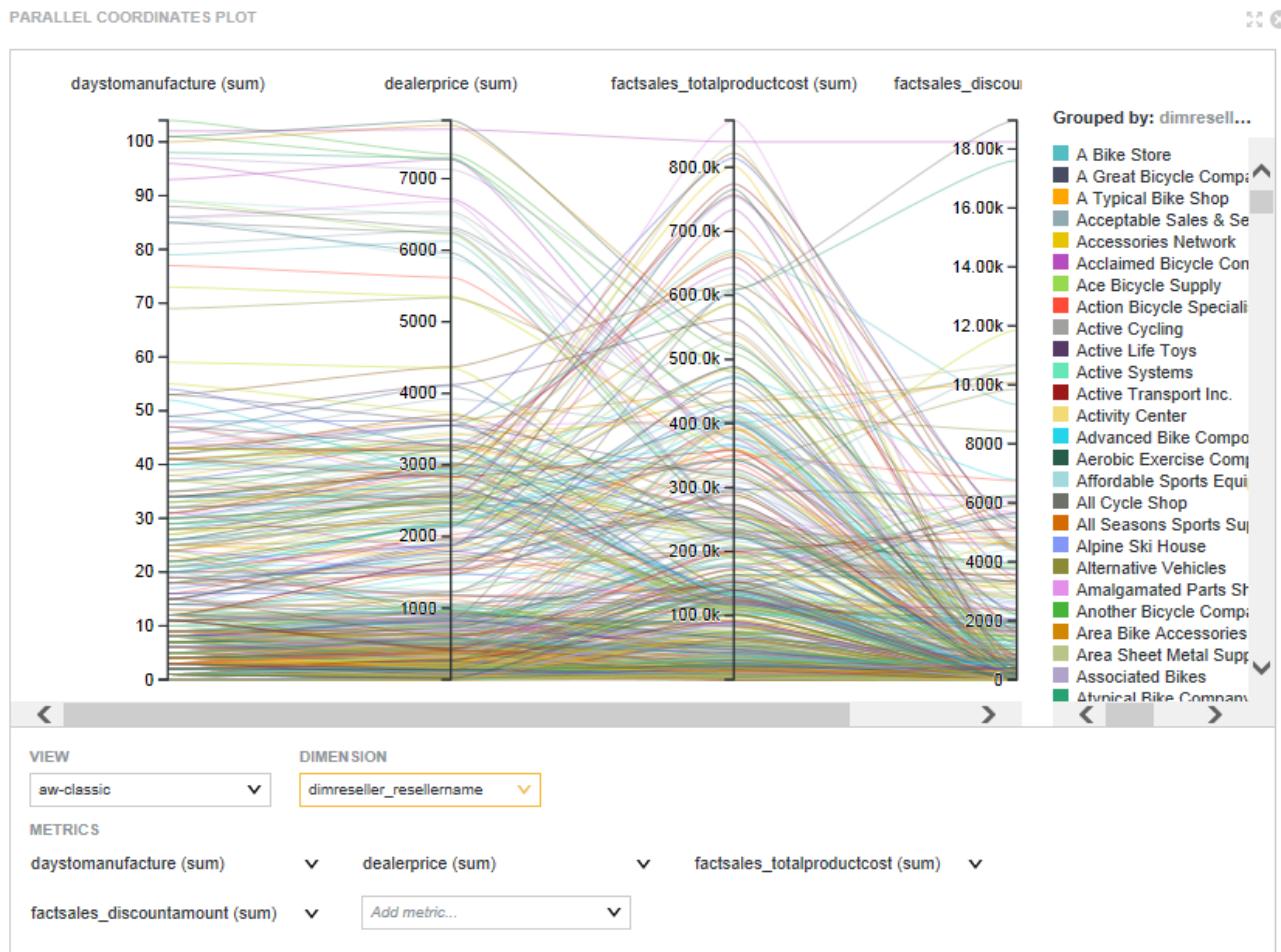


For a **Histogram Plot**, users can select:

- The view to use for the chart data
- The metric to plot on the X axis
- The type of histogram to display
- The number of bins to display on the X axis

## Parallel Coordinates Plot

The **Parallel Coordinates Plot** plots the values of multiple metrics for each value of a single dimension.



For a **Parallel Coordinates Plot**, users can select:

- The view to use for the chart data
- The metrics to display on the chart. For each metric, you also select the aggregation method to use.
- The dimension to use

## Other data visualizations

In addition to charts, Big Data Discovery provides data visualizations for generating more detailed or alternate views of the data.

### Map

Displays one or more map layers. Each layer represents a set of geographic locations.

You can search for locations, and display details for a specific location.



The **Map** component configuration includes:

- The lists of locations, including:
  - The geocode attribute representing the location
  - Display options for the map layer points
  - Sorting and pagination options
  - The attributes to include in the location details
- The available search options

## Pivot Table

Generates a table that allows users to perform comparisons and identify trends across several cross sections of data.

Users can export the data to a spreadsheet, and may be permitted to change the table layout.

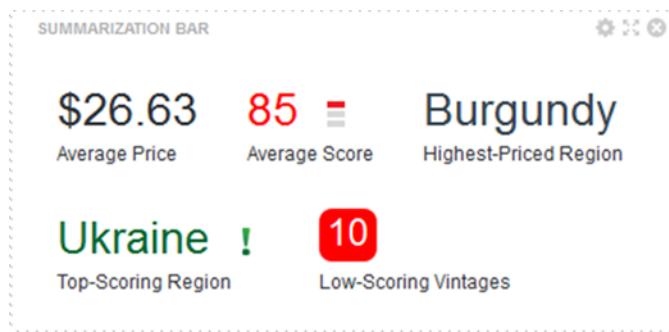
PIVOT TABLE								
Configure								
Vintage ▲		1992		1993		1994		
Region	Designation	Price(average)	Review Sco...	Price(average)	Review Sco...	Price(average)	Review Sco...	Price(average)
Alsace	Auslese			\$38.00	90.00			
	Best Buy	\$9.00	88.00	\$9.00	88.00	\$10.00	85.00	
	Botrytis	\$19.00	87.00			\$185.00	90.00	
	Brut	\$20.00	85.00	\$15.00	85.00			
	Cellar Select	\$43.50	92.50	\$40.00	92.00			
	Cellar Selecti...	\$43.50	92.50	\$40.00	92.00			
	Cremant	\$20.00	85.00	\$15.00	85.00			
	Cuvee	\$22.45	84.85	\$27.60	86.07	\$31.29	86.00	
	Grand Cru	\$27.60	85.88	\$42.88	86.47	\$45.76	86.51	
	Highly Reco...			\$34.00	90.67			
	Late Harvest	\$30.00	87.00					
	Reserve	\$13.61	81.89	\$16.00	84.40	\$29.00	84.25	
	Rose	\$20.50	88.50	\$56.29	87.14	\$40.77	88.23	
	Sec							

The **Pivot Table** component configuration includes:

- The metrics and dimensions to display
- Highlighting for specific metric values
- Available actions
- Other display options such as the table height and the column width

## Summarization Bar

Display a set of summary items that contain summary values from the data. A summary value can be a metric (such as average sales), a dimension value associated with the lowest or highest value of a metric (such as the region with the highest sales), or a number of flags. Flags indicate dimension values for which associated metric values have passed a specified threshold.

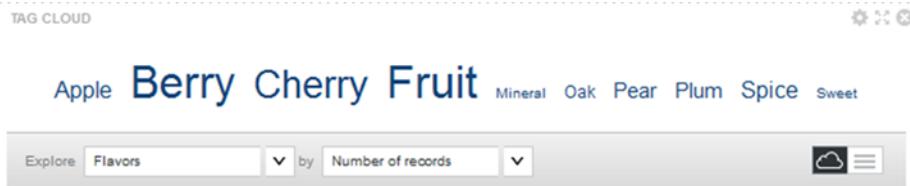


The **Summarization Bar** configuration includes:

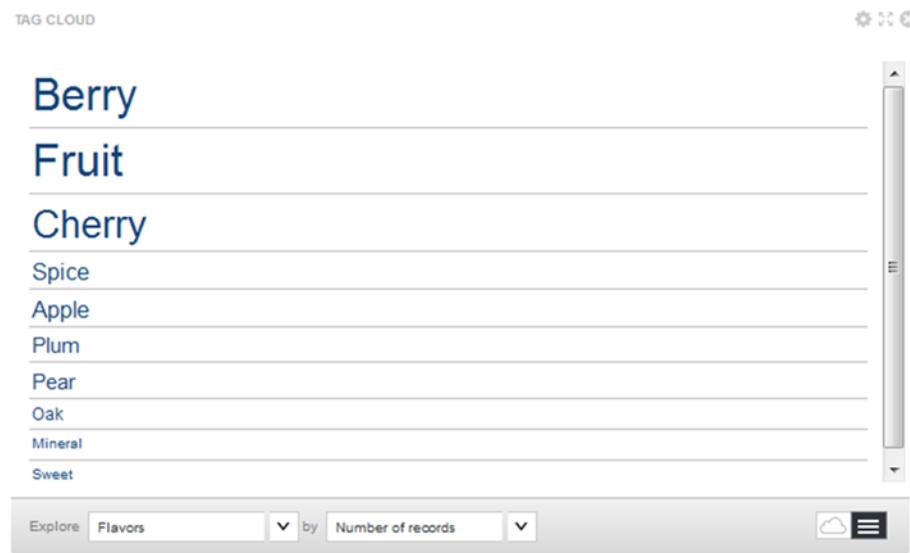
- The metric and dimension values to display
- Conditional formatting to highlight the summary item for specific ranges of metric values
- The text size for the summary item title and value

## Tag Cloud

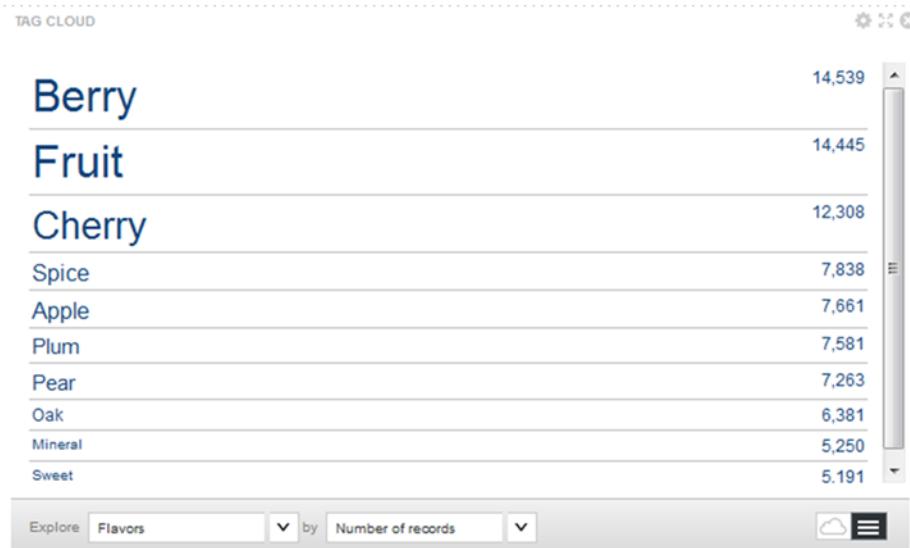
Displays a distribution of terms based on the value of an associated metric. The font size reflects the metric value. The higher the metric value, the larger the font.



The component can display in cloud view, with the terms in alphabetical order, or in list view, with the terms ordered by the metric value.



For both formats, the **Tag Cloud** can display the associated metric value:



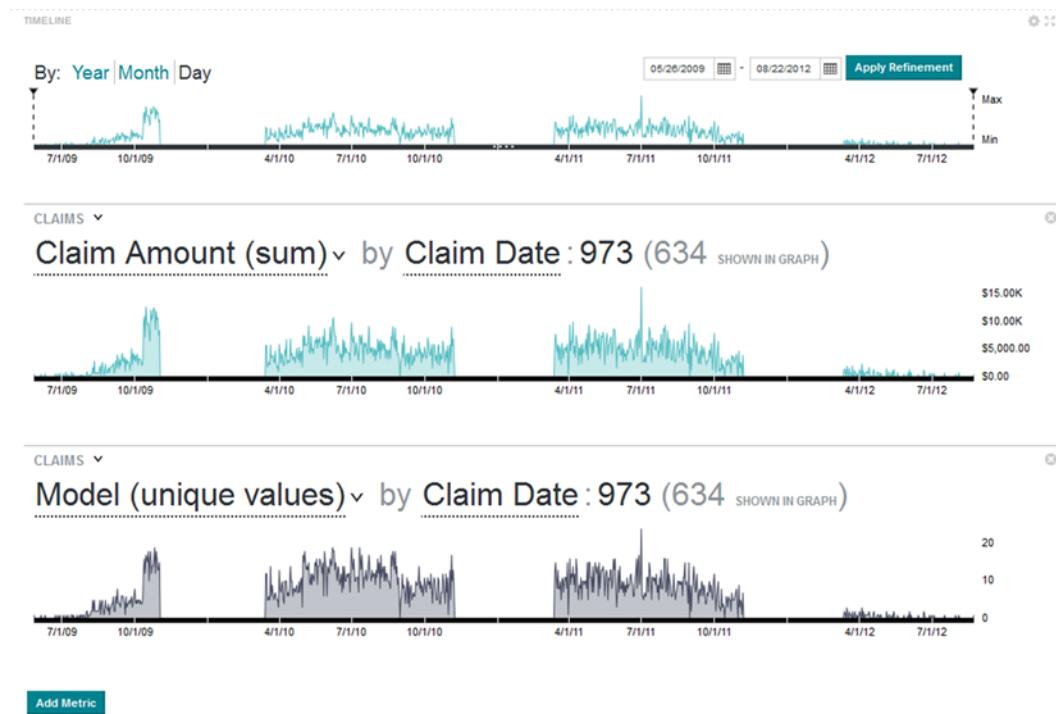
Users can refine by the displayed terms.

The **Tag Cloud** component configuration includes:

- The available dimensions (**Tag Cloud** terms) and metrics (values used to organize the terms)
- Whether to display the metric value next to each **Tag Cloud** term
- The number of terms to display
- The default display format (cloud of list)
- The range of text sizes for the displayed terms

## Timeline

Displays one or more metric values over the period of time. It consists of a master timeline to set the date/time range, and individual metric timelines showing the value of a metric for a selected date/time attribute.



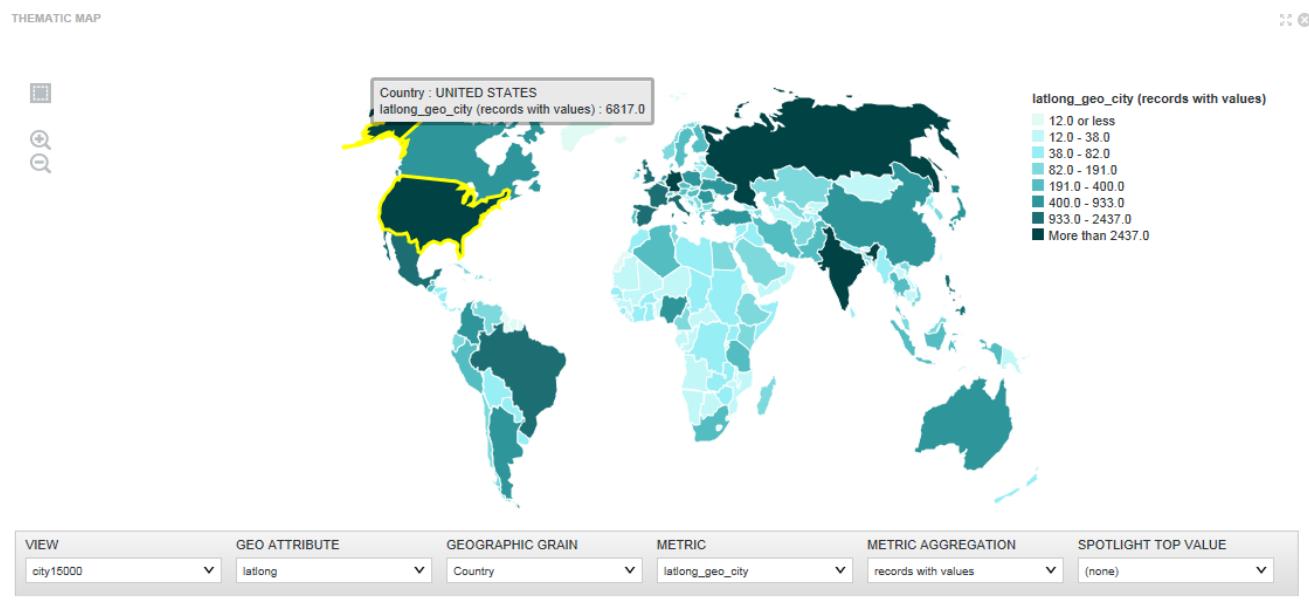
Users can:

- Add and remove metric timelines
- Select the metric and date/time attribute to use for each timeline
- Select a specific subset of the available date/time range to display on the metric timelines

There is no edit view for the component, but administrators can save a configuration to be the default displayed when users first display the page.

## Thematic Map

The **Thematic Map** component plots regions on a map and distinguishes the regions by color and record count per metric.



For a **Thematic Map**, you can select:

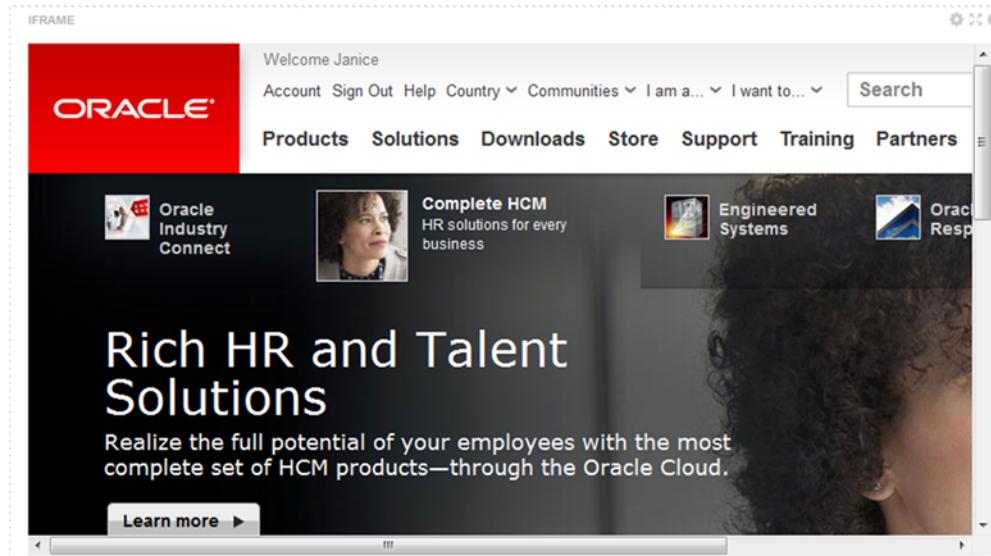
- The data view to use for the chart data
- The geocode attribute to display on the map.
- The geographic grain used to display the map. For example, you could display the map by country, region, or county, etc.

## Web-based content

These components are used to display web-based content on a page. The content may be custom HTML content created and stored in Big Data Discovery, or may be content from external URLs.

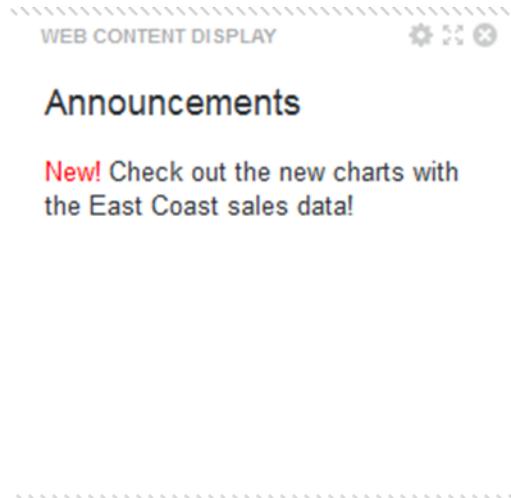
### IFrame

The **IFrame** component displays the content of an external URL on the page.



### Web Content Display

The **Web Content Display** component displays HTML content. The content is created and edited within the component.





## Chapter 18

# Managing Access to Projects

Projects can have different levels of access for different users.

*Configuring the project type*

*Adding and removing project members*

## Configuring the project type

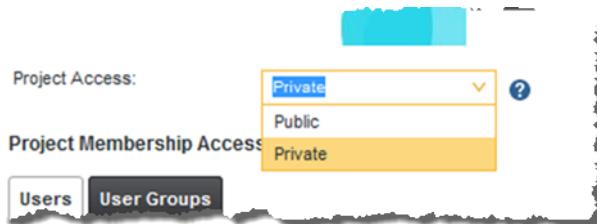
The project type determines whether the project is visible to users.

The project types are:

Project Type	Description
Public	<p>The project is visible to all logged-in users, and all logged-in users can select the project in order to view public pages.</p> <p>Members can also see private pages.</p> <p>Membership must be granted by a project administrator.</p>
Private	<p>The project is visible only to members.</p> <p>Membership must be granted by a project administrator.</p> <p>Projects are by default private projects.</p>

From the **Project Settings** page, to change the project type for a project:

1. On the **Project Configuration** page, from the **Project Access** drop-down list, select the project type.



2. To save the change, click **Save**.

## Adding and removing project members

Within a project, a user may either be a simple member or an administrator. Project administrators can add and remove project members, and determine whether a member is a project administrator.

The different project roles have different access privileges to the project content and membership. The roles are:

Role	Description
<b>Administrator</b>	<p>Administrators have full access to all features in Studio.</p> <p>Administrators can:</p> <ul style="list-style-type: none"> <li>• View all projects</li> <li>• Create and delete data sets and projects</li> <li>• Configure and manage all projects</li> <li>• Use all of the <b>Control Panel</b> features</li> </ul>
<b>Power User</b>	<p>Power users can:</p> <ul style="list-style-type: none"> <li>• View projects, based on the project and page type and their projects membership</li> <li>• Create and delete data sets and projects</li> <li>• Configure and manage projects for which they are an administrator</li> <li>• Transform data sets, if they are assigned the Project Author role</li> <li>• Export to HDFS and create new data sets</li> <li>• Edit their account information</li> </ul> <p>Power users cannot:</p> <ul style="list-style-type: none"> <li>• Access <b>Control Panel</b> features</li> </ul>
<b>User</b>	<p>Users can:</p> <ul style="list-style-type: none"> <li>• View projects, based on the project and page type and their project membership.</li> <li>• Create and delete data sets and projects</li> <li>• Transform data sets, if they are assigned the Project Author role</li> <li>• Edit their account information</li> </ul> <p>Users cannot:</p> <ul style="list-style-type: none"> <li>• Create new data sets or projects</li> <li>• Access <b>Control Panel</b> features</li> <li>• Export to HDFS</li> </ul>

Role	Description
<b>Restricted User</b>	<p>This is the default user role for new users. It has the most restricted privileges and is essentially a read-only role. This is the default user role for new users.</p> <p>Restricted users can:</p> <ul style="list-style-type: none"> <li>• View projects, based on the project and page type and their project membership.</li> </ul> <p>Restricted users cannot:</p> <ul style="list-style-type: none"> <li>• Create new data sets or projects</li> <li>• Access <b>Control Panel</b> features</li> <li>• Export to HDFS or create new data sets</li> <li>• Configure and manage projects</li> </ul>
<b>Project Author</b>	<p>Project authors are similar to administrators but have privileges for a particular project. Also, an administrator can assign the Project Author role to a Power User, Restricted User, and User. (This allows them to configure and manage projects in a way that their primary role would otherwise not allow.)</p> <p>Project authors can:</p> <ul style="list-style-type: none"> <li>• Configure and manage projects for which they are an author</li> <li>• Transform data</li> </ul> <p>Project authors cannot:</p> <ul style="list-style-type: none"> <li>• Create new data sets</li> <li>• Access <b>Control Panel</b> features</li> </ul>
<b>Project Restricted User</b>	<p>Project Restricted Users are similar to Users but are restricted to a particular project.</p> <p>Project Restricted Users can:</p> <ul style="list-style-type: none"> <li>• View projects, based on the project and page type and their project membership.</li> </ul> <p>Project restricted users cannot:</p> <ul style="list-style-type: none"> <li>• Create new data sets or projects</li> <li>• Access <b>Control Panel</b> features</li> <li>• Export to HDFS</li> <li>• Access <b>Project Settings</b> features</li> </ul>

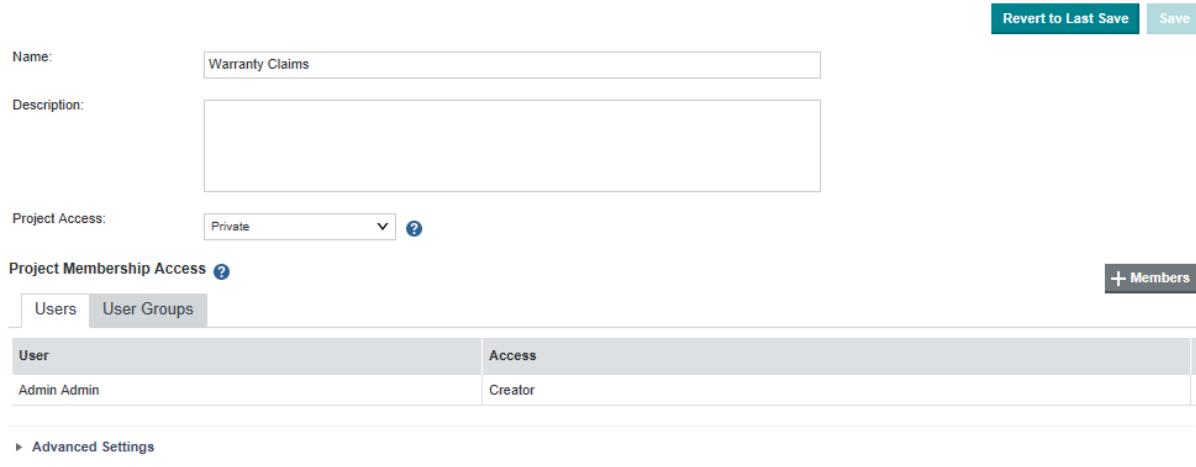
To add and remove project members:

1. On the project menu bar, click the project settings icon.

The project settings page is displayed, with the **Project Configuration** option selected automatically.

At the bottom of the page is the current list of project members. The **Users** tab lists the individual users that belong to the project, and the **User Groups** tab lists the user groups that belong to the project.

#### Project Configuration



The screenshot shows the 'Project Configuration' page. At the top, there are 'Name' and 'Description' fields, both empty. Below these are 'Project Access' and 'Project Membership Access' sections. The 'Project Membership Access' section includes tabs for 'Users' (selected) and 'User Groups', and a '+ Members' button. A table lists a single member: 'Admin Admin' with 'Creator' access. At the bottom, there are 'Advanced Settings' and 'Save' buttons.

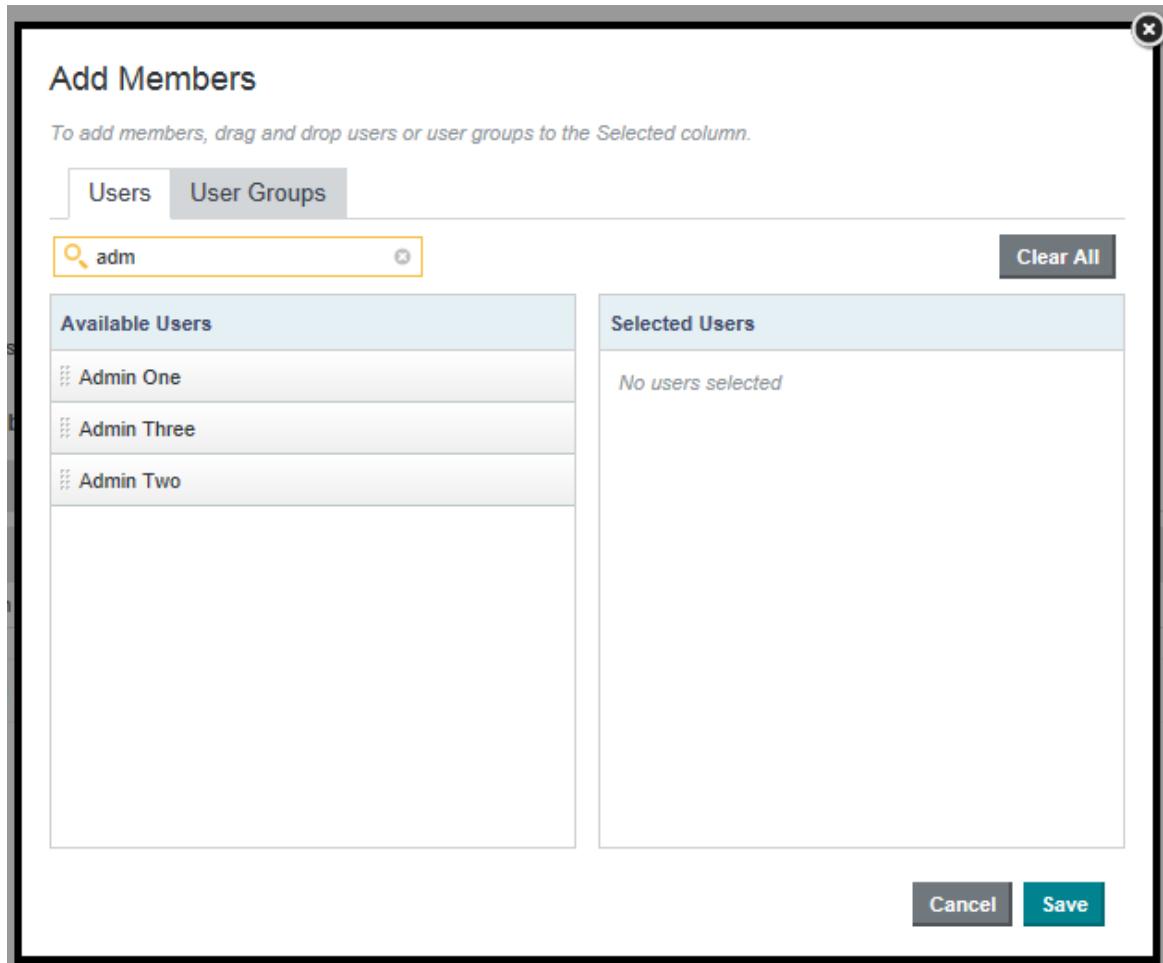
User	Access
Admin Admin	Creator

2. To add new members to the project:

- (a) Click the **+ Members** button above the member list.

On the **Users** and **User Groups** tab, the **Available Users** list shows the available users or groups.

- (b) To select a user or group to add as a new member, drag the user or group from the **Available Users** list to the **Selected Users** list.



- (c) To clear the list of selected users or groups, click **Clear All**.  
 (d) When you finish selecting the new members, click **Save**.

The new members are added to the list on the **Project Configuration** page. They are initially added as project members.

Project Membership Access <small>?</small>		<small>+ Members</small>
		<small>Users</small>
		<small>User Groups</small>
User	Access	
Admin Admin	Creator	
Ralph Wiggum	Project Restricted User	<small>x</small>

3. On the **Users** tab, you can use **Access** drop-down list next to each member to change them into an administrator.

User groups cannot be made into administrators.

4. To remove a user or group from the project membership, click the delete icon next to that user or group.

You cannot remove the original creator of the project.

5. To save the changes to the project membership, click **Save**.



## Chapter 19

# Configuring and Deleting Projects

Project administrators and Big Data Discovery administrators can configure and delete projects.

[Localizing the project name, description, page names, and component titles](#)

[Deleting a project using the Catalog](#)

## Localizing the project name, description, page names, and component titles

On the **Project Settings** page, you use the **Localization** page to localize the project name, description, page names, and component titles.

If you localize these items, then when end users select a different locale, these items can be displayed in the correct languages.

To localize the project name, description, page names, and component titles:

1. On the project menu bar, click the settings icon.
2. In the project settings menu, click **Localization**.

The **Localization** page contains a table listing:

- Project name
- Project description
- Pages
- Under each page, the list of components

Localization		
Select a locale to customize text for that language and location. To change default values for all locales, edit "United States   English" locale values.		
Element	Use Current Default	Override Default
Project Name	<input type="radio"/>	<input checked="" type="radio"/> Sales Discovery Dashboard
Project Description	<input type="radio"/>	<input checked="" type="radio"/> Dashboard for exploring recent sales data.
Page Title	<input type="radio"/>	<input checked="" type="radio"/> Wine
Component: Chart	<input checked="" type="radio"/> Chart	<input type="radio"/>
Component: Results Table	<input checked="" type="radio"/> Results Table	<input type="radio"/>

For the project name and description, you can only localize these values in other languages. You use the **Project Configuration** page to change the default values. For page names and component titles, you can change the value for both the default language and for other languages.

3. To provide localized values for a specific locale:
  - (a) From the locale drop-down list, select the locale.

By default, each item in the list has the **Use Current Default** radio button selected, indicating to use the value for the default locale.

Localization

Select a locale to customize text for that language and location.  
To change default values for all locales, edit "United States | English" locale values.

Revert to Last Save Save

Element	Use Current Default	Replace Default
Project Name	<input checked="" type="radio"/> Sales Discovery Dashboard	<input type="radio"/>
Project Description	<input checked="" type="radio"/> Dashboard for exploring recent sales data.	<input type="radio"/>
Page Title	<input checked="" type="radio"/> Wine	<input type="radio"/>
Component: Chart	<input checked="" type="radio"/> Gráfico	<input type="radio"/>
Component: Results Table	<input checked="" type="radio"/> Tabla de resultados	<input type="radio"/>

- (b) For each item you want to localize, click the **Replace Default** radio button.  
A field for the localized value is displayed.
  - (c) In the field, type the localized value, then press **Enter**.
4. To save the changes to the localization, click **Save**.

## Deleting a project using the Catalog

Both project administrators and Big Data Discovery administrators can delete individual projects from the **Catalog** or they can use the **Projects** page on the **Control Panel** to delete multiple projects at a time.

For details about using the **Control Panel** to delete projects, see the *Administrator's Guide*.

To delete a project using the Catalog:

1. On the **Catalog** landing page or results page, click the project tile.
2. On the project details panel, click the delete link.

When you delete a project, Big Data Discovery automatically deletes any bookmarks and snapshots for that project.

# **Part VII**

## **Viewing and Using Projects**



## Chapter 20

# Navigating Within a Project

From the **Catalog**, you can select a project to view. You can also display information about the project and its associated data sets.

*[Navigating to a project](#)*

*[Navigating within a project](#)*

*[Displaying information about the project data](#)*

## Navigating to a project

From the **Catalog**, to navigate to a project, click the project name. The project details panel also contains a link to display the project.

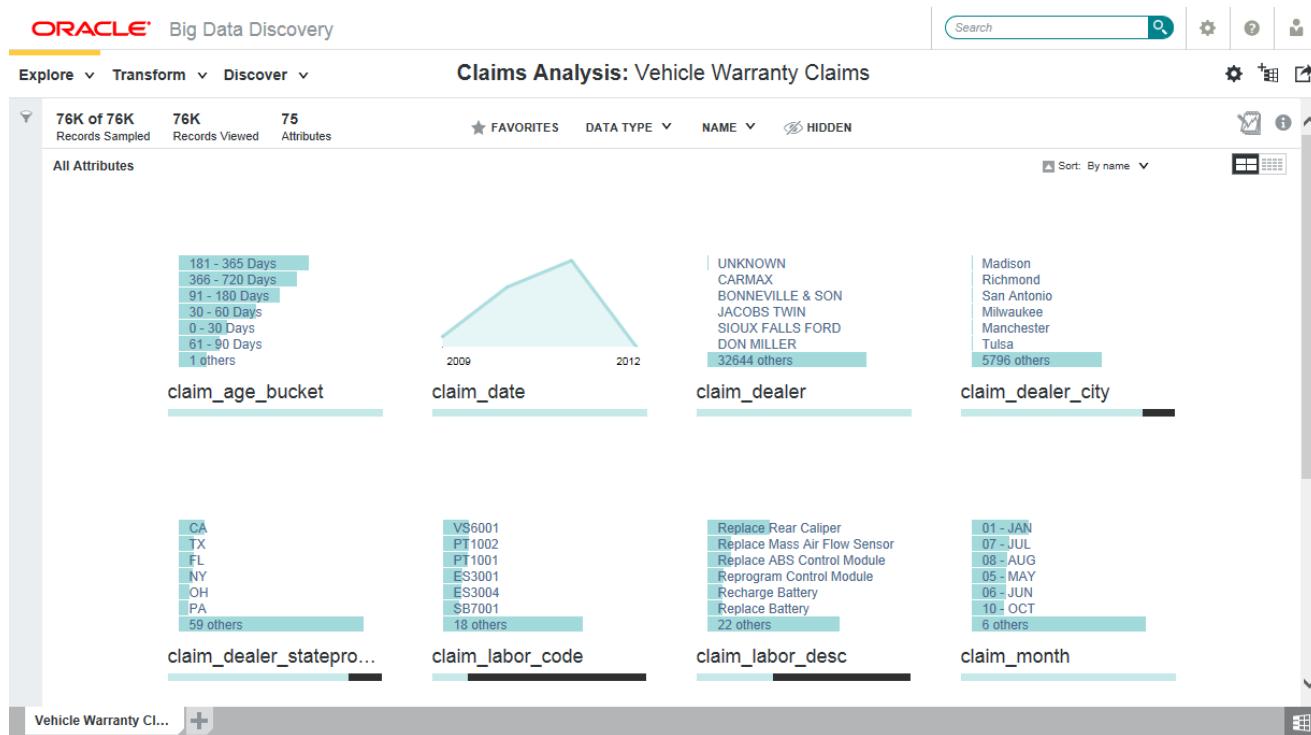
## Navigating within a project

Projects initially open on the **Explore** page, to allow you to explore one or more data sets in a project.

In **Explore**, each data set is represented on a tabbed page. For details on exploring a data set, see *[Exploring a Data Set on page 70](#)*.

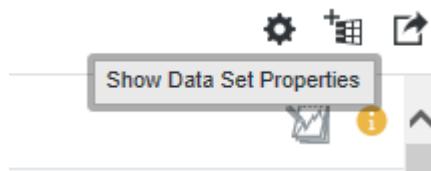
You can also switch to the **Transform** or **Discover** pages, and, if you have role privileges, you can also display project settings. In **Discover**, each tabbed page provides an area to add components to analyze the

data and discover insights hidden in data. For details about using specific components, see [Working with Components on page 229](#).



## Displaying information about the project data

In the header of both the **Explore** and **Transform** pages, there is an information icon that provides information about the data sets used in the project:



To display information about the project, click the **Show Data Set Properties** icon.

For each data set, the information includes:

- The data source name
- When the data was last updated
- Information about the status of the data set, to indicate when a data set is loading or has a problem
- Information about other related data sets in the project

The **Data Set Relationships** tab shows how the data sets are linked.



## Chapter 21

# Using Common Component Functions

These features are common to many components.

*[Paging through component data](#)*

*[Using a component to refine data](#)*

*[Exporting data from a component](#)*

*[Printing from a component](#)*

*[Displaying details for a component item](#)*

*[Comparing items from a component](#)*

## Paging through component data

Components that display a set of data may include a pagination toolbar. From the pagination toolbar, you can navigate through the data, and also determine the amount of data to display per page.

To page through component data:

1. At the left of the pagination toolbar are the navigation controls. To navigate using the paging toolbar:



- To navigate to the next or previous page, click the next or previous page button.
- To navigate to the first or last page, click the first or last page icon.
- To jump to a specific page, type the page number in the field, then press **Enter**.

2. The right of the toolbar shows the current subset of records being displayed (for example, "1-50 of 500"). To change the amount of data displayed on each page:
  - (a) Click the drop-down icon next to the record subset information.
  - (b) From the pop-up list, select the number of items to display per page.



## Using a component to refine data

The **Available Refinements** panel and **Search Box** functions are specifically designed for refinement, but other components can allow you to click attribute values in order to refine the data.

When you click a value that is enabled for refinement, other components that are tied to data from the same data set are updated to reflect the change.

Refinements are always added to the **Selected Refinements** panel, to allow you to:

- See what you have refined by
- Remove refinements

See [About the Selected Refinements panel on page 58](#).

### Refining from a list of values

For a list of values:

- To refine by an individual value, click that value
- To refine by all of the values, click the icon that is displayed after the list, then click **Refine by all values**.

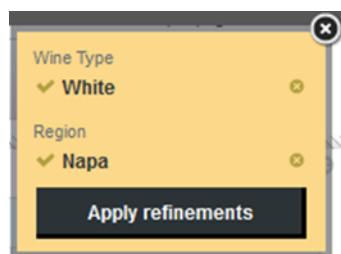


If you can only refine by one value at a time, then the **Refine by all values** option is not displayed.

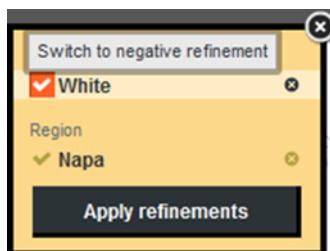
### Selecting multiple values to refine by

You can also create a list of values to refine by. You can select the values from different attributes and components, including components on different pages. For example, you could refine by a Region value displayed on a **Results Table**, and a Wine Type value displayed on a **Summarization Bar**.

To do this, Ctrl-click each value. The values must all be enabled for refinement. When you Ctrl-click a value, it is added to the list of values to refine by. You can also Ctrl-click a **Refine by all values** link to add all of the values to the list. As you navigate between pages, the list remains displayed until you either confirm or cancel the refinement.



To toggle a selection between positive and negative refinement, click the icon in front of the value.



If the list of refinements does not match any records, Big Data Discovery displays a warning message.

To remove a value from the list, click its delete icon. If you remove all of the values, then the list is closed.

By default, the list is either at the right of the page, if you selected the first value from a component, or immediately to the right of the **Available Refinements** panel, if you selected the first value from there. You can, however, use the top bar to drag the list to another location on the page.

The list only applies to the current page. If you navigate to a different page, the list closes automatically.

To refine the data using the selected values, click **Apply refinements**.

To close the list and cancel the refinement, click the close icon.

## Date hierarchies and dimension cascading

Date/time values have an implicit hierarchy that uses the available subsets of date/time units. For example, if a component is currently displaying the year for a date value, then when you refine by a specific year, Big Data Discovery could display data for the months of that year. If you then refine by a specific month, Big Data Discovery could display data for the days of that month.

Components may also be configured to allow cascading of values. Cascading means that when you refine by a specific attribute value, the component is updated to display values from a different attribute. For example:

1. When you refine by the country United States, the component displays data for all of the states in the United States.
2. When you refine by California, the component displays data for cities in California.

## Exporting data from a component

The **Actions** menu for a component can include an option to export data from the component to a CSV file that uses UTF-16LE encoding.

The export option is not available if you are viewing the project on an iPad.

The exported file includes:

Component	Exported Data
<b>Pivot Table</b>	Contains the data used to create the <b>Pivot Table</b> display.
<b>Record Details</b>	Contains a single row for the displayed data.

Component	Exported Data
<b>Results List</b>	You can select specific records to export. The exported data contains the displayed attributes for the exported records.
<b>Results Table</b>	You can select the specific rows to export. The data includes all of the columns from all of the available column sets. For aggregated tables, only the displayed dimensions are included in the export.

To export data from a component:

1. From the **Actions** menu, select **Export**.

On the export dialog, if the amount of data currently selected to export is greater than the maximum allowed, a warning is displayed indicating that the data will be truncated to the maximum allowed.

2. In the **Name** field, set the name of the exported file.

By default the file name is *ComponentNameYYYY-MM-DD*.

For example, if you export a **Results Table** component on October 5, 2013, then the default file name for the exported data is *ResultsTable2013-10-05.csv*.

3. By default, the exported file uses commas to separate the fields. Under **CSV Format**, to use a custom separator:

(a) Click the **Custom separator** radio button.

(b) From the **Fields separated by** drop-down list, select the character to use to separate the fields.

(c) From the **Fields enclosed by** drop-down list, select whether to use quotes to enclose the values.

You can use single quotes, double quotes, or no quotes.

For example, if you use a pipe as the separator, and enclose the value in single quotes, a row from the exported file might look something like:

```
| '10234' | 'Bourdeaux' | 'Valmaison' | 'Bourdeaux Philibert Bourgeouis' | '$5' |
```

4. Under **Value Formatting**, click a radio button to indicate whether to export the raw data before formatting, or the formatted values.
5. If the component has individual records selected, then under **Export**, click a radio button to indicate whether to export the selected records only, or all of the records.
6. Click **Export**.
7. When the export is completed, you are prompted to save or open the file.

The prompt varies based on the current browser.

## Printing from a component

The component **Actions** menu can include an option to print the component.

When you select the **Print** option, Big Data Discovery displays a preview of the printout.

At the top of the printout are any refinements that have been applied to the data.

For **Results Table**, only the current page is printed.

For **Pivot Table**, only the currently displayed data is printed.

## Displaying details for a component item

Some component hyperlinks display the **Record Details** dialog, which displays information from the current component item. For example, you could display details for a row in a **Results Table**, a record in a **Results List**, or a location on a **Map** component.

The **Record Details** dialog displays attributes within the context of their attribute groups.

If the original item reflects a record from a data set, the **Record Details** dialog displays all of the attribute groups that an administrator configured to include in record details. For example, a **Results List** item may only display a few attributes, but the **Record Details** dialog contains all of the applicable attribute groups for the associated record.

For aggregated data, the **Record Details** dialog only includes the displayed data:

- The dimensions are displayed in their attribute groups
- The metrics are displayed in the Other group

From the **Record Details** dialog, you can print the displayed data, or export the data to a spreadsheet. See [Printing from a component on page 137](#) and [Exporting data from a component on page 136](#).

## Comparing items from a component

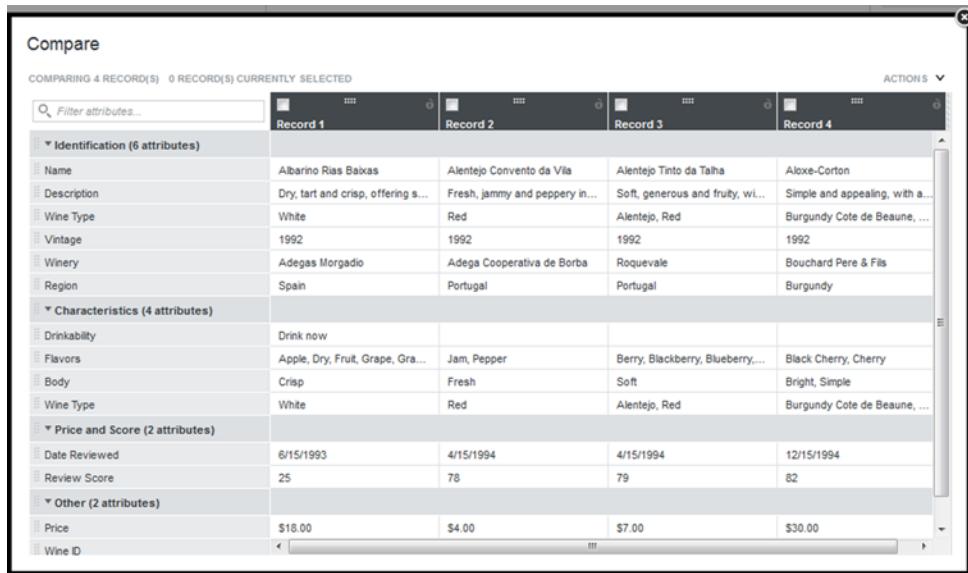
Some components allow you to compare selected items. If the component supports the compare function, then the component **Actions** menu includes a **Compare** option.

### Selecting the items to compare

On the component, to select the records to compare using the **Compare** component:

1. Check the checkbox for each item you want to compare.
2. From the **Actions** menu, select **Compare**.

The selected items are displayed on the **Compare** dialog.



The screenshot shows the 'Compare' dialog with four records listed: Record 1, Record 2, Record 3, and Record 4. The table displays various wine attributes across these records. The 'Identification' group shows Name, Description, Wine Type, Vintage, Winery, Region, and other details. The 'Characteristics' group shows Drinkability, Flavors, Body, and Wine Type. The 'Price and Score' group shows Date Reviewed, Review Score, and Price. The 'Other' group shows Wine ID. Record 3 is highlighted as the baseline record.

	Record 1	Record 2	Record 3	Record 4
Name	Albarino Rias Baixas	Alentejo Convento da Vila	Alentejo Tinto da Talha	Aloxe-Corton
Description	Dry, tart and crisp, offering s...	Fresh, jammy and peppery in...	Soft, generous and fruity, wi...	Simple and appealing, with a...
Wine Type	White	Red	Alentejo, Red	Burgundy Cote de Beaune, ...
Vintage	1992	1992	1992	1992
Winery	Adegas Morgadio	Adega Cooperativa de Borba	Roquevale	Bouchard Pere & Fils
Region	Spain	Portugal	Portugal	Burgundy
Drinkability	Drink now			
Flavors	Apple, Dry, Fruit, Grape, Gra...	Jam, Pepper	Berry, Blackberry, Blueberry,...	Black Cherry, Cherry
Body	Crisp	Fresh	Soft	Bright, Simple
Wine Type	White	Red	Alentejo, Red	Burgundy Cote de Beaune, ...
Date Reviewed	6/15/1993	4/15/1994	4/15/1994	12/15/1994
Review Score	25	78	79	82
Price	\$18.00	\$4.00	\$7.00	\$30.00
Wine ID	4	11	12	13

If the component items reflect records from the data set, then the **Compare** dialog displays all of the attribute groups configured to be used for record details. For example, a **Results Table** row may only display a few attributes, but the **Record Details** dialog contains all of the applicable attribute groups for the associated record.

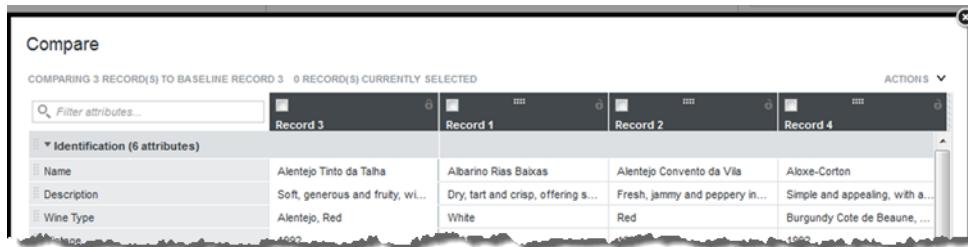
If the original items were aggregated, then the **Compare** dialog only includes the displayed data. The dimensions used for aggregation are displayed in their attribute groups, and the metrics are displayed in the Other group.

To restore the initial display for the **Compare** dialog, from the **Actions** menu, select **Reset view**.

## Selecting a baseline item

To select an item to be the baseline, in the heading, click the lock icon.

The selected item becomes the first column in the list, and cannot be moved. The lock icon is changed to indicate that the item is the baseline. If another item was previously selected as the baseline, it becomes a non-baseline item.



The screenshot shows the 'Compare' dialog with four records listed: Record 3, Record 1, Record 2, and Record 4. The table displays various wine attributes across these records. The 'Identification' group shows Name, Description, Wine Type, and other details. Record 3 is highlighted as the baseline record.

	Record 3	Record 1	Record 2	Record 4
Name	Alentejo Tinto da Talha	Albarino Rias Baixas	Alentejo Convento da Vila	Aloxe-Corton
Description	Soft, generous and fruity, wi...	Dry, tart and crisp, offering s...	Fresh, jammy and peppery in...	Simple and appealing, with a...
Wine Type	Alentejo, Red	White	Red	Burgundy Cote de Beaune, ...
Vintage	1992	1992	1992	1992

To remove the designation as a baseline record, click the lock icon again. The record column remains at the left of the table, but can now be moved.

You can drag non-baseline records left or right in the display, to allow you to do a side-by-side comparison of selected records.

## Managing the displayed attributes and attribute groups

You also can drag and drop attribute groups to move them up or down in the display. You can reorder attributes within a group, but cannot drag attributes to other groups.

To expand or collapse an attribute group, click the group name.

To expand all of the attribute groups, from the **Actions** menu, select **Expand all groups**.

To collapse all of the attribute groups, from the **Actions** menu, select **Collapse all groups**.

You can also use the filter field to search for a specific attribute.

## Highlighting differences between the items

To highlight the differences in attribute values, from the **Actions** menu, select **Highlight Differences**.

If you haven't selected a baseline, then the **Compare** dialog highlights attribute values that are not the same across all of the selected items.

▼ Identification (6 attributes)				
Name	Albarino Rias Baixas	Alentejo Convento da Vila	Alentejo Tinto da Talha	Aloxe-Corton
Description	Dry, tart and crisp, offering s...	Fresh, jammy and peppery in...	Soft, generous and fruity, wi...	Simple and appealing, with a...
Wine Type	White	Red	Alentejo, Red	Burgundy Cote de Beaune, ...
Vintage	1992	1992	1992	1992
Winery	Adegas Morgadio	Adega Cooperativa de Borba	Roquevare	Bouchard Pere & Fils
Region	Spain	Portugal	Portugal	Burgundy

If there is a baseline, then in the non-baseline items, the **Compare** dialog highlights attribute values that are different from the baseline record:

▼ Identification (6 attributes)				
Name	Alentejo Tinto da Talha	Albarino Rias Baixas	Alentejo Convento da Vila	Aloxe-Corton
Description	Soft, generous and fruity, wi...	Dry, tart and crisp, offering s...	Fresh, jammy and peppery in...	Simple and appealing, with a...
Wine Type	Alentejo, Red	White	Red	Burgundy Cote de Beaune, ...
Vintage	1992	1992	1992	1992
Winery	Roquevare	Adegas Morgadio	Adega Cooperativa de Borba	Bouchard Pere & Fils
Region	Portugal	Spain	Portugal	Burgundy

When the highlighting is displayed, to remove the highlighting, from the **Actions** menu, select **Hide Highlights**.

## Removing items from the display

To remove records from the **Compare** dialog:

1. Check the checkbox for each record to remove.
2. From the **Actions** menu, select **Remove selected**.



## Chapter 22

# Creating and Using Bookmarks

---

You can create, navigate to, and send links to bookmarks.

*[About bookmarks](#)*

*[Displaying and navigating to bookmarks](#)*

*[Creating, editing, and deleting bookmarks](#)*

*[Generating and emailing bookmark links](#)*

*[Bookmark data saved for each component and panel](#)*

## About bookmarks

Bookmarks allow you to save a given navigation and component state so you can return to it at a later time or email it to other users.

Note that a bookmark is only guaranteed to work if the page that it points to has not been changed. If components have been added, removed, or modified, then the bookmark may no longer be valid.

When a page or project is removed, then any associated bookmarks for that page or project are also removed.

When a user is removed, then any bookmarks created by that user are also removed.

## Displaying and navigating to bookmarks

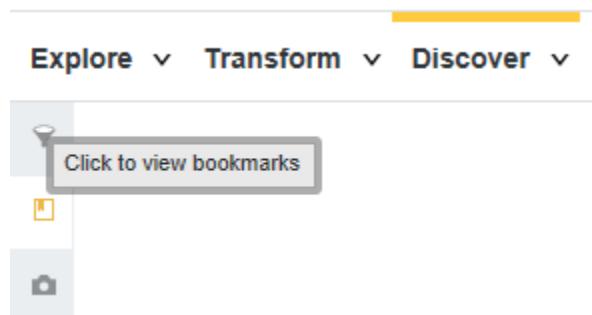
The bookmarks panel displays bookmarks for the current project created by the current user, and can include shared bookmarks created by other users.

To display and navigate to bookmarks:

1. Select **Discover** if you haven't already.

2. Click the bookmarks icon in the left panel.

For example:



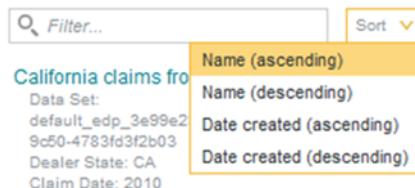
On the bookmarks panel, the **My Bookmarks** list contains bookmarks created by the current user. If a bookmark is configured to be available to other users, then it is marked with a shared icon. The **Shared Bookmarks** list contains bookmarks created by other users and configured to be available to all project users.

3. To search for a specific bookmark within a list, use the filter field.

As you type, Big Data Discovery filters the list to display bookmarks for which the entered text is in the bookmark name, bookmark description, or the refinements associated with the bookmark.

4. To sort a list of bookmarks, use the **Sort** drop-down list at the top of the bookmark list.

You can either sort the list alphabetically by the bookmark name, or based on the order in which the bookmarks were created.



5. To navigate to a specific bookmark, click the bookmark name.

## Creating, editing, and deleting bookmarks

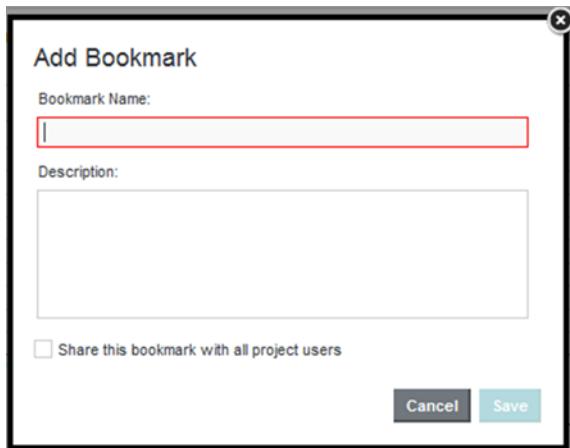
You can create new bookmarks. You can also edit and delete bookmarks that you have created.

To create, edit, and delete bookmarks:

1. To create a new bookmark:
  - (a) Select **Discover** if you haven't already.
  - (b) Click the bookmarks icon.

- (c) On the bookmarks panel, click **Add New Bookmark**.

The **Add Bookmark** dialog is displayed.



- (d) In the **Bookmark Name** field, enter a name for the bookmark.

All bookmarks you create must have a unique name.

- (e) In the **Description** field, provide an optional description for the bookmark.

- (f) To make the bookmark a shared bookmark available to all users with access to the current project, check the **Share this bookmark with all project users** checkbox.

- (g) Click **Save**.

2. To display the details for the bookmark, click the information icon next to the bookmark name.

The bookmark details include the bookmark name, bookmark description, when it was created, and the refinement state represented by the bookmark. For a shared bookmark, the information includes the user who created it.

The bookmark **Actions** menu includes option to edit, email, retrieve the URL for, and delete the bookmark. You can only edit and delete bookmarks that you created.

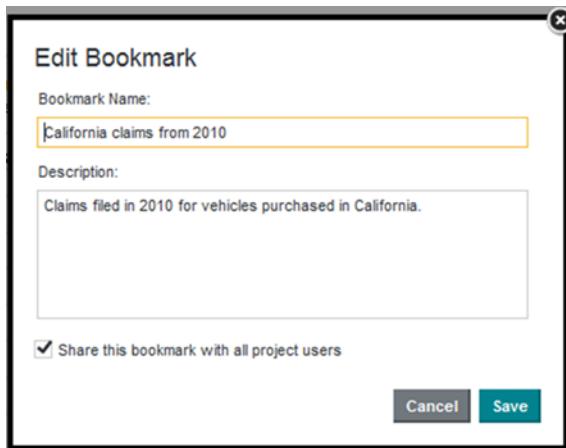


3. To edit an existing bookmark that you created:

- (a) From the bookmark **Actions** menu, select **Edit**.

- (b) On the **Edit Bookmark** dialog, make the necessary edits to the bookmark.

You can change the bookmark name and description, and whether the bookmark is shared.



- (c) Click **Save**.

4. To delete a bookmark that you created, from the bookmark **Actions** menu, select **Delete**.

## Generating and emailing bookmark links

Users can get access to bookmarks using a hyperlink. From the **Actions** menu on the bookmark details panel, you can generate the link to the bookmark, or email the link.

Users who try to navigate to a bookmark using a link are prompted to log in to Big Data Discovery. Even if they are sent the link, users cannot navigate to a bookmark for a project or page they do not have access to.

Note that in order to be able to send snapshots by email, the outbound email server for Big Data Discovery must be configured. For information on how to configure the outbound email server, see the Administrator's Guide.

To generate or email a link to a bookmark:

1. To generate the link to a bookmark, from the bookmark **Actions** menu, select **Create Link**.

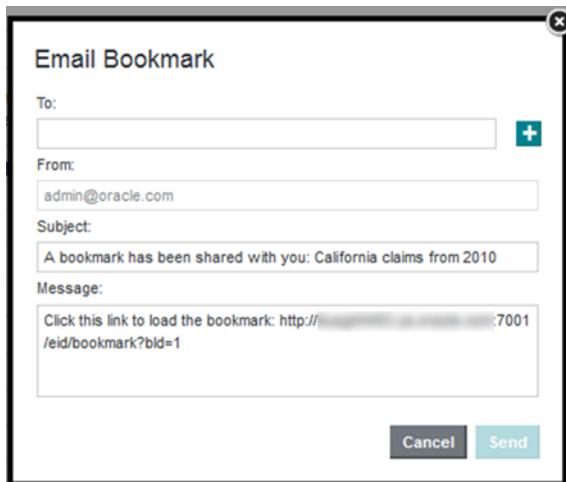
The **Bookmark Link** dialog is displayed, with the link displayed in a field. The link is selected automatically, so that you can immediately copy it.



2. To email a link to a bookmark:

- (a) From the bookmark **Actions** menu, select **Email**.

The **Email Bookmark** dialog is displayed, with the **Subject** and **Message** fields pre-populated.



- (b) In the **To** field, type the email address for the recipient.

If you are sending the snapshot to multiple recipients, then for each additional recipient, click the **+** icon to add a field for the email address.

- (c) The **From** field contains your email address, which is used as the reply-to address on the email message.

Note that this is not the **From** address. The **From** address on the email is the address associated with the outbound email server.

For information on configuring the outbound email server, see the *Administrator's Guide*.

- (d) In the **Subject** field, you can customize the subject line for the email.

- (e) In the **Message** field, you can add any additional text to the email message.

Make sure you do not change the bookmark link.

- (f) To send the email, click **Send**.

## Bookmark data saved for each component and panel

For the different components and panels on a page, bookmarks can include different levels of detail.

Bookmarks save the following information:

Component or Panel	Persisted States
<b>Available Refinements</b>	<ul style="list-style-type: none"> <li>• Whether the panel is expanded or collapsed</li> <li>• Which data set is expanded</li> <li>• Expanded and collapsed groups and attributes</li> <li>• Toggle selection for the refinement type (value list vs. range filter vs. recurring date)</li> </ul>
<b>Box Plot</b>	<ul style="list-style-type: none"> <li>• Selected Y-axis metric</li> <li>• Selected X-axis attribute</li> <li>• Boxes per page</li> </ul>
<b>Chart</b>	<ul style="list-style-type: none"> <li>• Selected metric</li> <li>• Selected dimensions</li> </ul>
<b>Histogram Plot</b>	<ul style="list-style-type: none"> <li>• Selected metric</li> <li>• Number of bins</li> </ul>
<b>Map</b>	<ul style="list-style-type: none"> <li>• Records per page</li> <li>• Sort order</li> <li>• Map searches</li> </ul>
<b>Parallel Coordinates Plot</b>	<ul style="list-style-type: none"> <li>• Selected view</li> <li>• Displayed metrics</li> <li>• Aggregation method</li> <li>• Selected dimension</li> </ul>
<b>Pivot Table</b>	None
<b>Results List</b>	<ul style="list-style-type: none"> <li>• Sort order</li> <li>• Records per page</li> </ul>
<b>Results Table</b>	<ul style="list-style-type: none"> <li>• Sort order</li> <li>• Number of records per page</li> </ul>

Component or Panel	Persisted States
<b>Selected Refinements</b>	<ul style="list-style-type: none"> <li>• Whether the panel is expanded or collapsed</li> <li>• Whether refinements for each data set are hidden</li> <li>• Display order of the refinement tiles</li> <li>• Whether each refinement tile is expanded</li> <li>• Scroll position for navigating through the refinement tiles</li> </ul>
<b>Summarization Bar</b>	None
<b>Tabbed Component Container</b>	Selected tab
<b>Tag Cloud</b>	<ul style="list-style-type: none"> <li>• Display format (cloud or list)</li> <li>• Selected dimension</li> <li>• Selected metric</li> </ul>
<b>Timeline</b>	<ul style="list-style-type: none"> <li>• Selected time frame</li> <li>• Selected date/time attributes</li> <li>• Selected metrics</li> </ul>



## Chapter 23

# Creating and Managing Snapshots

Big Data Discovery allows you to create and manage snapshots of project pages.

[\*About snapshots and snapshot galleries\*](#)

[\*Displaying the list of snapshots and snapshot galleries\*](#)

[\*Creating and managing individual snapshots\*](#)

[\*Emailing or exporting the entire snapshot list\*](#)

[\*Creating and managing snapshot galleries\*](#)

## About snapshots and snapshot galleries

A snapshot captures an image of a component or page at a specific moment in time. You can save these snapshots in Big Data Discovery.

You can email individual snapshots, save them as image files, and return to the project state reflected by the snapshot.

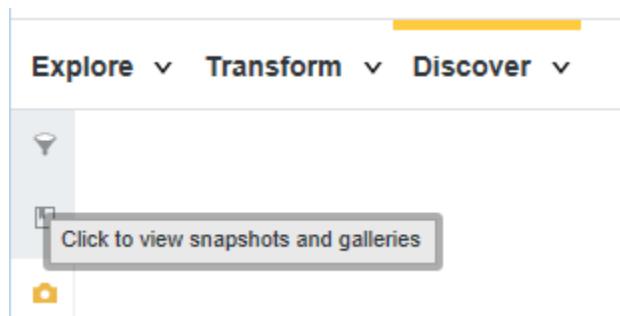
You can also create galleries of related snapshots. For example, a snapshot gallery could show the progression of chart data as different refinements are applied.

You can email a snapshot gallery, and save a gallery to a PDF file.

## Displaying the list of snapshots and snapshot galleries

If you click the snapshot (camera) icon, available on the **Discover** page, Big Data Discovery displays the snapshots panel.

For example:



The snapshots panel includes:

- The **Snapshots** list, showing the snapshots you have created for this project.

- The **Galleries** list, showing the snapshot galleries you have created for this project.

To find a specific snapshot, use the search field at the top of the **Snapshots** list. As you type, Big Data Discovery filters the list to display snapshots with the entered text in the snapshot name, snapshot description, or the refinements associated with the snapshot.

To find a specific gallery, use the search field at the top of the **Galleries** list. As you type, Big Data Discovery filters the list to display galleries with the entered text in the gallery name or gallery description.

## Creating and managing individual snapshots

You can create, view details, and perform actions on individual snapshots.

[Creating a snapshot](#)

[Managing saved snapshots](#)

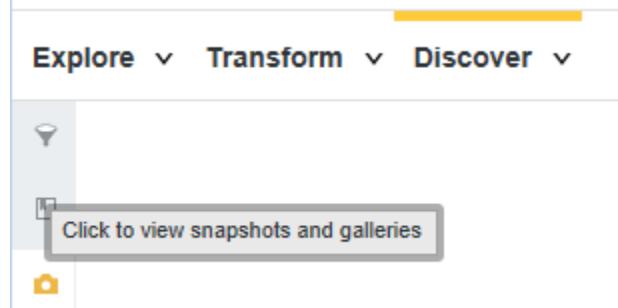
### Creating a snapshot

From either the project footer or the snapshots panel, you can create snapshots of a project page. The snapshots are added to the **Snapshots** list.

To create a new snapshot:

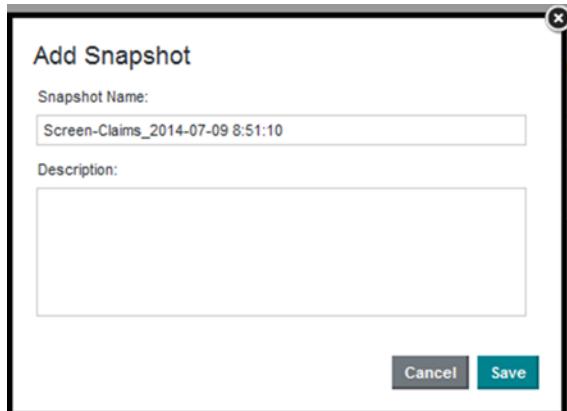
1. Select **Discover** if you haven't already.
2. Click the snapshot (camera) icon to display the Snapshots panel

For example:



3. In the project footer click **New snapshot** or on the snapshots panel, click **New snapshot**.

Both bring up the same dialog. The **Add Snapshot** dialog displays. The dialog includes a generated default name for the snapshot.



4. In the **Snapshot Name** field, update the name for the snapshot.
5. In the **Description** field, type a description of the snapshot.
6. Click **Save**.

## Managing saved snapshots

From the **Snapshots** list for a project, you can edit, email, save, and delete snapshots. You can also use the snapshot as a bookmark to display the project page in the state reflected by the snapshot.

Note that in order to be able to send snapshots by email, the outbound email server for Big Data Discovery must be configured. For information on how to configure the outbound email server, see the *Administrator's Guide*.

To manage the snapshots in the **Snapshots** list:

1. To display details for a snapshot, click the snapshot.

The snapshot details include the name, description, creation date, source of the data, and the refinements that were in place when the snapshot was created.

**Ford claims**

Claims for Ford models

ORACLE® Big Data Discovery

REFINEMENTS

CREATED: 7/14/14

REFINEMENTS

Data Set: default\_edp\_7925f989-6f85-4fb0-bf46-775e57fo4de1  
make: FORD

Apply this refinement state

The **Actions** menu contains options to edit, email, save, and delete the snapshot. You can also navigate to the refinement state represented by the snapshot.

2. To edit a snapshot:
  - (a) From the snapshot **Actions** menu, select **Edit**.
  - (b) On the edit dialog, you can edit the name and description of the snapshot.

**Ford claims**

Snapshot Name: Ford claims

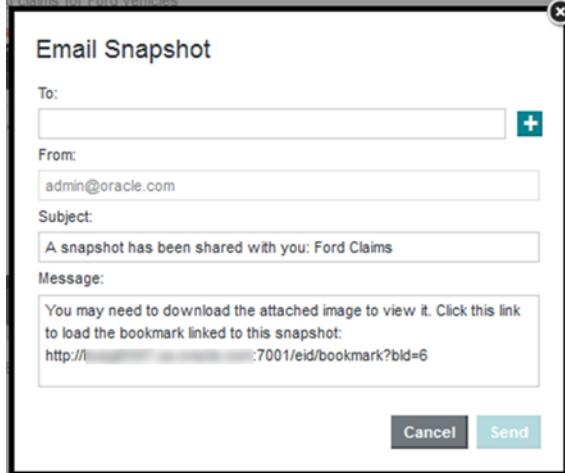
Description: Claims for Ford models.

Cancel Save

- (c) To save the changes, click **Save**.

3. To email a snapshot:
  - (a) From the snapshot **Actions** menu, select **Email**.

The **Email Snapshot** dialog is displayed.



- (b) In the **To** field, type the email address of the recipient.

If you are sending the snapshot to multiple recipients, then for each additional recipient, click the **+** icon to add a field for the email address.

- (c) The **From** field shows your email address as the "Reply to:" address on the email message.

Note that this is not the **From** address. The **From** address on the email is the address associated with the outbound email server.

For information on configuring the outbound email server, see the *Administrator's Guide*.

- (d) In the **Subject** field, edit the subject line of the email.

- (e) In the **Message** field, edit the text for the email message.

- (f) To send the snapshot, click **Send**.

4. To save a snapshot to a file, from the **Actions** menu, click the save to disk icon.

Depending on your browser, you are prompted to open or save the image.

5. To display the project page using the refinement state associated with the snapshot, on the snapshot details, click **Apply this refinement state**.

6. To delete a snapshot, from the snapshot **Actions** menu, select **Delete**.

You are prompted to confirm that you want to delete the snapshot from the list. If the snapshot is part of a gallery, it is still stored there.

## Emailing or exporting the entire snapshot list

The **Actions** menu on the snapshots panel includes options to email or export the entire list of snapshots. Both functions save the images to a single PDF file.

Note that in order to be able to send snapshots by email, the outbound email server for Big Data Discovery must be configured. For information on how to configure the outbound email server, see the *Administrator's Guide*.

To email or save the entire set of snapshots:

1. To email all of the snapshots:
  - (a) In the **Actions** menu, click **Email All**.
  - (b) On the email dialog, in the **To** field, type the email address of the recipient.  
To add a field for an additional recipient, click the **+** button.
  - (c) In the **Subject** field, type the subject line for the email.
  - (d) In the **Message** field, type the text of the email.
  - (e) Click **Send**.

The snapshots are sent in a PDF file attached to the email.
2. To save all of the snapshots to a PDF file:
  - (a) In the **Actions** menu, click **Save All to Disk**.
  - (b) Use the file save dialog to name and save the file.

## Creating and managing snapshot galleries

You can group snapshots into snapshot galleries, which you can then email to other users.

[Creating a snapshot gallery](#)

[Editing and deleting snapshot galleries](#)

[Emailing a snapshot gallery](#)

[Exporting a snapshot gallery to a PDF file](#)

## Creating a snapshot gallery

You can create snapshot galleries to group a related set of snapshots. For example, you can put together a "slide show" of screen shots showing a series of refinement states.

From the snapshots panel, to create a new snapshot gallery:

1. On the snapshots panel, click **New Gallery**.

The new gallery dialog is displayed.



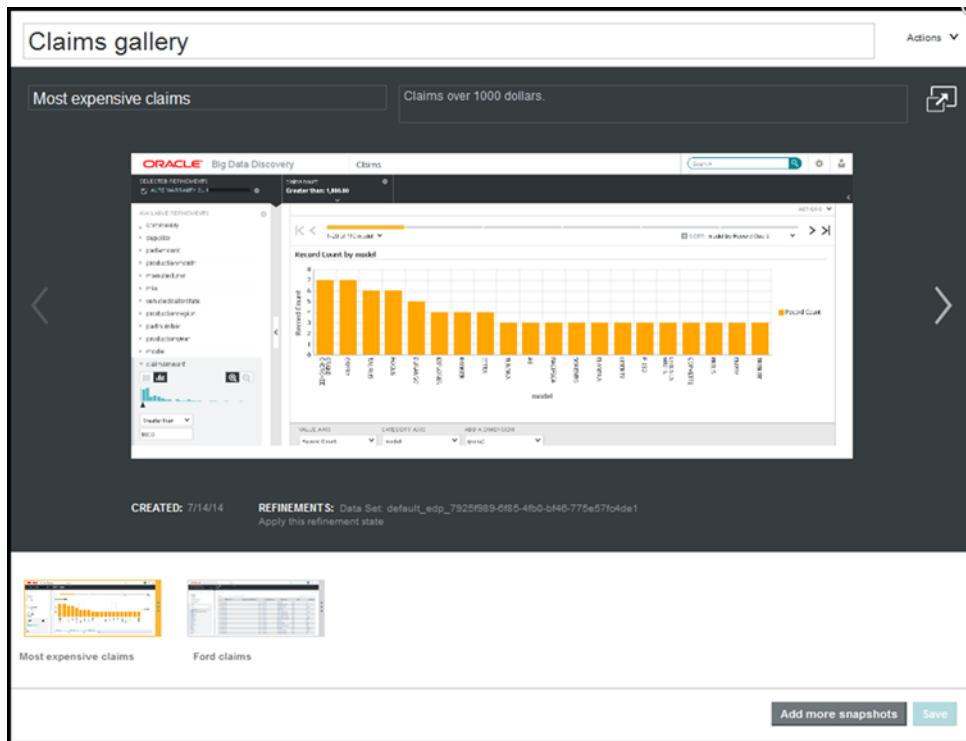
2. In the name field, type the name of the new gallery.
3. In the list of snapshots, click each snapshot you want to add to the new gallery.  
To deselect a snapshot, click it again.

4. To save the new gallery, click **Save**.

The gallery panel for the new gallery is displayed.

The first snapshot is selected, and is displayed at the top of the panel.

At the bottom of the panel is the full list of snapshots in the gallery.



## Editing and deleting snapshot galleries

Within a snapshot gallery, you can add snapshots to the gallery, delete snapshots from the gallery, and change the order of the snapshots in the gallery. For each snapshot, you can provide a name and description that is specific to the gallery.

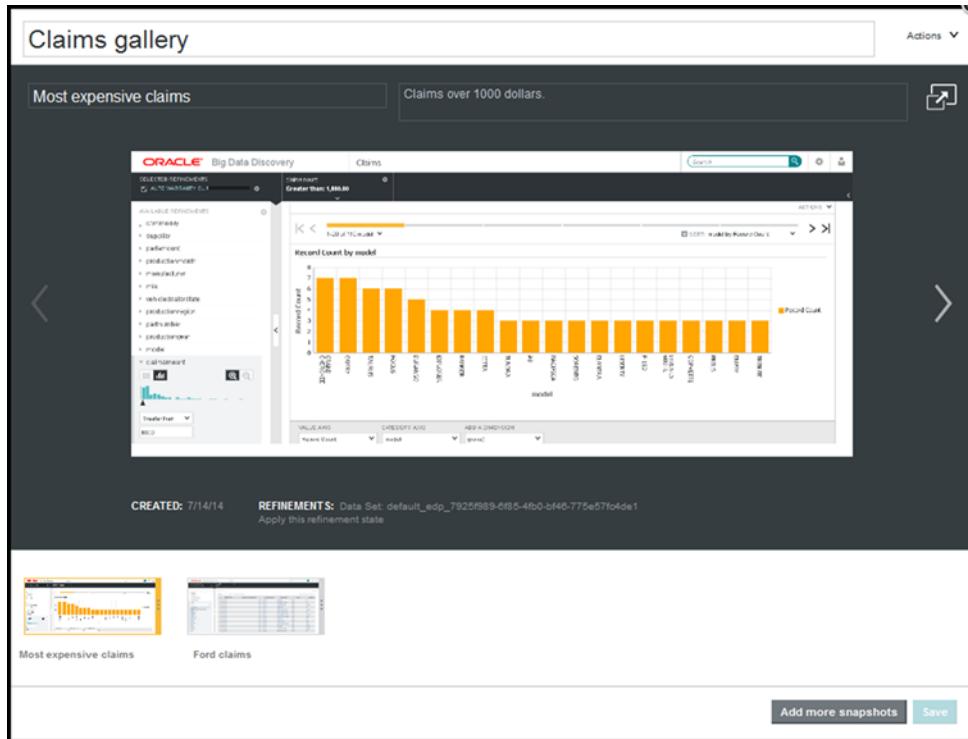
To edit and delete a snapshot gallery:

1. To display the details for a gallery, in the **Galleries** list, click the information icon for the gallery.

The gallery panel for the selected gallery is displayed.

The first snapshot is selected by default, with the full list of snapshots in the gallery displayed at the bottom of the panel.

You can use the next and previous icons to page through the snapshot list.



2. To determine the order of the snapshots in the gallery, drag and drop the snapshots in the gallery snapshots list.
  3. To display and edit the details for a snapshot in the gallery, in the snapshots list, click the snapshot. You can use the name and description fields at the top of the panel to customize the name and description of the snapshot in the context of the gallery.

This name and description is specific to this instance of the snapshot within the gallery.
  4. To page through the snapshots in full screen mode:
    - (a) Click the full screen icon at the top right.
    - (b) Use the next and previous icons to page through the snapshots.
    - (c) To exit the full screen mode, either click the return icon, or press the **Esc** key.
  5. To add snapshots to the gallery:
    - (a) Click **Add more snapshots**.
    - (b) On the add snapshots dialog, click each snapshot that you want to add.

To deselect the snapshot, click it again.
    - (c) When you have selected and deselected the snapshots as needed, click **Save**.
  6. To delete a snapshot from the gallery:
    - (a) Hover the mouse over the snapshot.
    - (b) Click the delete icon.

7. To save the changes to the snapshot gallery, click **Save**.
8. To delete a snapshot gallery, in the **Galleries** list, click the delete icon for the snapshot gallery.

## Emailing a snapshot gallery

You can email an entire snapshot gallery to one or more email addresses. The gallery is attached to the message as a PDF file.

Note that in order to be able to send a snapshot gallery by email, the outbound email server for Big Data Discovery must be configured. For information on how to configure the outbound email server, see the *Administrator's Guide*.

To email a snapshot gallery:

1. In the **Galleries** list, click the information icon for the snapshot gallery.
2. In the gallery **Actions** menu, click **Email**.
3. On the **Email Gallery** dialog, in the **To** field, type the email address of the recipient.  
To add another recipient, click the + button.
4. In the **Subject** field, edit the subject line of the email.
5. In the **Message** field, edit the text for the email message.
6. Click **Send**.

## Exporting a snapshot gallery to a PDF file

You can export an entire snapshot gallery to a PDF file.

To export a snapshot gallery:

1. In the **Galleries** list, click the information icon for the snapshot gallery.
2. In the gallery **Actions** menu, click **Save to Disk**.
3. Use the file save dialog to name and save the PDF file.

# **Part VIII**

## **Managing Data for a Project**



## Chapter 24

# Managing Data Sets for a Project

You can add or remove data sets within a project. A project must contain at least one data set. Also, the **Data Set Manager** provides a means to enable or disable links among multiple data sets in a project.

*[Adding a data set to an existing project](#)*

*[Removing a data set from a project](#)*

## Adding a data set to an existing project

Big Data Discovery provides several options for adding a data set to a project.

### From the Catalog

On the **Catalog**, the data set details panel includes an option to add the data set to a project.

Click **Add to Project**, then select the project.

### From Explore

When exploring a data set from the **Catalog**, the page header includes an option to add the data set to a project.

Click **Add to Project**, then select the project.

### From within a project

From the **Explore** or **Transform** pages of a project, to add a data set to the project, click the **+** icon to the right of the data set tabs.

You can select an existing data set, or create a new data set.

## Removing a data set from a project

You can have any number of data sets associated with a project. If a data set is no longer useful in a particular project, you can remove the data set from that project. This removal affects only the active project. It does not affect any other projects that a data set may be added to.

To remove a data set from a project:

1. Select either **Explore** or **Transform**.
2. In the footer, locate the data set you want to remove from the project

3. Click the remove icon for the data set.

For example:





## Chapter 25

# Transforming a Project Data Set

The **Transform** page provides options for correcting and enriching a project data set.

[About transforming data](#)

[Toggling the Transform data grid view](#)

[Filtering and sorting the displayed attributes and values](#)

[Deleting records from the project data set](#)

[Adding transformations to a transformation script](#)

[Managing and executing the transformation script](#)

[Creating a new data set from the transformed data](#)

## About transforming data

The **Transform** page allows you to make corrections and enhancements to a project data set. The transformations are within the scope of the project and do not affect the data set in the **Catalog** for other Studio users.

Explore ▾ **Transform** ▾ Discover ▾

For example, you can use **Transform** to modify:

- Incorrect data types
- Incorrect formatting, such as extra spaces or inconsistent capitalization
- Missing values
- Invalid or inconsistent values. For example, an attribute may have multiple versions of the same value (Wal-Mart, Walmart, Wal\*Mart).

You can also use **Transform** to provide additional information about the data or provide new ways to use the values. For example, you can:

- Use grouping or binning to create groups or ranges of attribute values
- Split attributes to pull out specific pieces of information from a value
- Identify commonly used terms
- Analyze sentiment

**Transform** also provides options to remove attributes or records from the project data set.

Any number of transformations can be combined into a transformation script. You run the script against the project data set. And if desired, you can create a new data set in the **Catalog** based on the transformed version.

## Transform page header information and links

At the top of **Transform** is a summary of the project data set. For example:



This summary includes:

- The number of records in the sample being used
- The number of records in the entire data set
- The number of records in the sample that match the current refinements
- The number of attributes in the data set
- Attributes that have been marked as favorites
- A **Data Type** drop down to locate attributes by data type
- A **Name** drop down to locate attributes by name
- A **Hidden** control to display any attributes that were hidden

The page header can also contains:

- A link to show/hide the Custom Transform Editor
- A link to show/hide data set properties
- A value indicating the number of transforms that have been applied to this project data set.

## Size of the data set to transform

If the source data is larger than one million records, the source data is automatically sampled during the ingest process to give you a data set limited to a maximum of one million records.

If the source data is smaller than one million records, the project contains the full data set rather than a sampled data set.

## Selecting the data to transform

The **Transform** displays the project data set. You can select a single data set on a tabbed page, or a linked data set if you have enabled linking.

To select the data to transform:

1. In the **Catalog**, select a project whose data set you want to transform.
2. In the header, click **Transform**.

## Toggling the Transform data grid view

Most of the **Transform** page consists of the data grid. In the grid, each column represents an attribute from the project data set.

The data grid can display either:

- An initial list of records from the data sample. In this view, each row represents a record.
- Lists of attribute values. For each attribute, this view lists the unique values found in the data sample, listed in descending order by frequency.

To switch to the record list, click the rows icon. To switch to the value list view, click the values icon.

The header for each attribute column contains:

- An icon to indicate the data type for that attribute
- The attribute menu, used to get access to transformation options for that attribute
- A data quality bar to indicate the number of records with valid or missing values for the attribute. You can click either group of values in the quality bar to refine the data.

## Filtering and sorting the displayed attributes and values

**Transform** provides sorting and filtering options to help you focus on the attributes you want to work with.

### Setting attributes as hidden or favorite

The **Favorite** and **Hide** options allow you to mark attributes that you are more interested in or that you do not want to see. These options only apply within a project.

To set an attribute as a favorite, in the attribute menu, under **Other Actions**, click **Star**. You can then use the **Favorites** filter option to only display the attributes you mark as a favorite.

To hide an attribute, in the attribute menu, under **Other Actions**, click **Hide**. The attribute is removed from the grid. You can use the **Hidden** filter option to only display the hidden attributes.

When you select **Favorite** or **Hide**, the menu option is checked. To remove the favorite or hidden flag for the attribute, click the option again.

When you set an attribute as hidden or favorite in the **Transform** page, the attribute is also treated the same way in **Explore** and all users of the project.

### Filtering the displayed attributes

In the **Transform** header are options for filtering the displayed attributes based on:

- Whether they are favorites or hidden
- The attribute data type
- The attribute name

## Filtering the attribute values

You can use the **Available Refinements** panel to filter the displayed data based on attribute values. See [Using Refinements for Filtering on page 35](#).

You can also use the data quality bar to refine the data to only include records with valid, invalid, or missing values.

The refinement state can be included in custom transformations, and also used to select records to delete from the project data set.

## Sorting the attributes and values

The sort menu on the **Transform** page allows you to sort the attributes based on either the original sort order, or by the attribute name.

To sort the attribute values, click the attribute column heading.

- In rows view, the attribute values are sorted alphabetically. Clicking the column toggles between ascending and descending order.
- In values view, the attribute values are sorted based on their relative frequency in the sample data. Clicking the column toggles between displaying the most used or least used values at the top of the list.

## Deleting records from the project data set

You can use the refinement state as a basis for deleting records from the project data set. You can either delete records that match the refinement state, or delete records that do NOT match the refinement state.

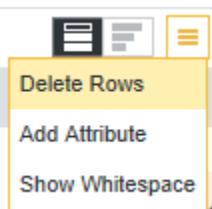
For example, you might want to delete all records with null values for certain attributes. Or you may want to delete all records except for those with Canada as the country value.

If you delete the records, then they are permanently removed from the project data set. They are still in the original data set.

To delete records from the project data set:

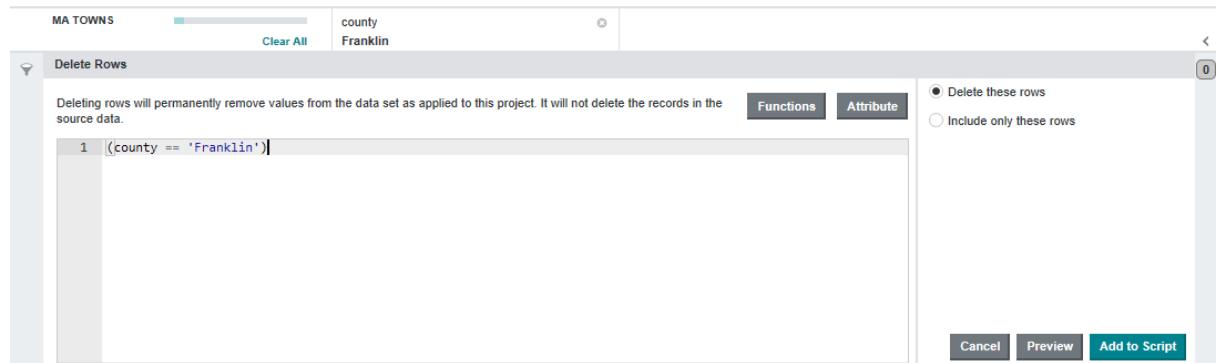
1. On **Explore**, refine the data set to either represent the records you want to keep or exclude.
2. Select **Transform**.
3. From the transform grid menu, select **Delete Rows**.

For example:



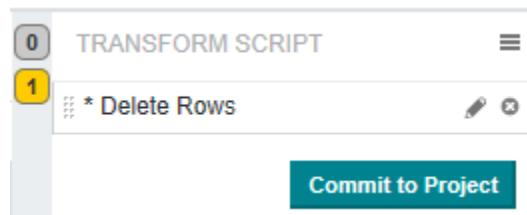
4. On the **Delete Rows** editor, verify that the transform script reflects the refinement state you selected.

For example, this transform script deletes records that have a county value of Franklin:



5. Click the radio button to indicate whether to delete records that match the script, or delete records that do NOT match the script.
6. If desired, click **Preview** to see the how the script affects the data set, or click **Add to Script** to save the transform.
7. Click the numbered bubble to expand the **Transform Script** window.
8. Click **Commit to Project** to run the script. If necessary, click the pencil icon to edit the transform before committing it.

For example:



## Adding transformations to a transformation script

You can add any number of transformations to a transformation script. The script can include quick transformations, available from the attribute menu, and also custom transformations that you create in the **Transformation Editor**.

[Overview of transforming an attribute](#)

[Adding an attribute](#)

[Changing the data type for an attribute](#)

[Editing attribute values](#)

[Splitting and grouping values](#)

[Creating a custom transformation script](#)

## Overview of transforming an attribute

Both quick transforms and custom transforms follow the same basic flow.

The high-level process:

1. From the attribute menu, select the transformation that you want to apply. If the transformation does not require any configuration, then it is added immediately to the transformation script. When you commit the script to the project, the transformed values overwrite the original attribute values. If the transformation does require configuration, then the **Transform Editor** displays.
2. On the **Transform Editor** panel, provide the transformation-specific configuration.
3. Determine the target attribute for the transformed values. Some transformations require that you use the transformation to create a new attribute. In this case, in the **New Attribute Name** field, type the name for the new attribute.  
Other transformations allow you to choose whether to create a new attribute, or use the transformed values to overwrite the current attribute.
  - The default option is to create a new attribute. In the **New Attribute Name** field, type the name for the new attribute.
  - To instead overwrite the current attribute, click the **Apply to existing attribute**.
4. To display the effect of the transformation on the **Transform** data grid, click **Preview**.
5. To add the transformation to the transformation script, click **Add to Script**.

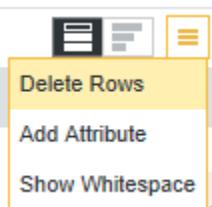
## Adding an attribute

You can add a new attribute to a project data set by specifying Groovy code to produce the attribute and by specifying a new attribute name and data type for the result. You then preview the change. If the result is what you expect, you can add it to the transformation script.

To add an attribute:

1. Select **Transform**.
2. From the transform grid menu, select **Add Attribute**.

For example:



- In the **Transformation Editor**, specify the Groovy code necessary add the attribute.

This code can be as simple or as complicated as necessary for your use case.

In this simple example, the `toUpperCase` function examines the `asciiname` attribute, converts its values to upper case, and creates a new attribute called `Name`:

The Transformation Editor interface shows a Groovy script in the top-left pane:

```
1 toUpperCase(asciiname)
```

On the right, a configuration panel for a new attribute is shown:

- New Attribute Name:** Name
- Data Type:** String
- Single Assign

Below these panes is a preview of the data in a tabular format. The preview shows the following columns: asciiname, dem, elevation, geonameid, latlong, latlong\_geo\_city, latlong\_geo\_country, and latlong\_g. The data rows are:

asciiname	dem	elevation	geonameid	latlong	latlong_geo_city	latlong_geo_country	latlong_g
Kenilworth	31	34	5,100,007	40.676490 -74.290700	Kenilworth	US	07033
Achim	14		2,959,681	53.014160 9.026300	Achim	DE	26832
Sarbogard	107		3,045,564	46.886920 18.620410	Sárboágard	HU	7002
Arras	73		3,036,784	50.293010 2.781860	Arras	FR	62001 ced
Boora	26		1,337,233	24.850980 89.371080	Boora	BD	5621

Buttons at the bottom right of the preview pane include Cancel, Preview, and Add to Script.

- In **New Attribute Name**, specify the name of the attribute you want to add.
- From the **Data Type** list, choose the data type of the new attribute.
- If necessary, uncheck **Single Assign** to make the new attribute multi-assign.
- Click **Preview** to see the results of running the function displayed in the tabular view, or click **Add to Script** to both preview the new attribute in tabular view and also save the transform to the transform script.

Preview simply highlights the new column for the attribute in yellow but does not save the change to the transform script.

In this simple example, **Add to Script** was selected and you can see the new attribute `Name` to the left of `asciiname`.

The Transformation Editor interface shows the same Groovy script and configuration panel as the previous screenshot. The preview of the data now includes an additional column to the left of `asciiname`, labeled `Name`, which contains the uppercase values of the original `asciiname` column.

Name	asciiname	dem	elevation	geonameid	latlong	latlong_geo_city	latlong_geo_country	latlong_g
KENILWORTH	Kenilworth	31	34	5,100,007	40.676490 -74.290700	Kenilworth	US	07033
ACHIM	Achim	14		2,959,681	53.014160 9.026300	Achim	DE	26832
SARBOGARD	Sarbogard	107		3,045,564	46.886920 18.620410	Sárboágard	HU	7002
ARRAS	Arras	73		3,036,784	50.293010 2.781860	Arras	FR	62001 ced
BOGRA	Bogra	26		1,337,233	24.850980 89.371080	Bogra	BD	5621
FALCONARA MARITTIMA	Falconara Marittima	15	5	3,177,250	43.615480 13.374630	Falconara Marittima	IT	
BASODA	Basoda	410		1,276,752	23.851530 77.936520	Básoda	IN	
CAMPBELLTOWN	Campbelltown	84		2,172,586	-34.066670 150.816670	Campbelltown	AU	

If you are done making changes to the project data set, you can commit the changes. See [Running the transformation script against the project data set on page 174](#).

## Changing the data type for an attribute

You can change the data type assigned to an attribute.

For example, an attribute might contain mostly date values, but be classified as a string because of missing values or an unrecognized format. To take advantage of the features associated with date/time attributes, you can change the data type to date/time, then use other transformations to clean up the values.

Any data type can be transformed to a string. For other data types:

- Strings can be converted to date/time, numeric, or geocode data types
- Date/time values can be converted to numbers, or to times
- Long numbers can be converted to dates or geocodes

When converting values to or from date/time or time values, you must specify the date/time format to use. The acceptable formats use the Java SimpleDateFormat syntax

(<http://docs.oracle.com/javase/7/docs/api/java/text/SimpleDateFormat.html>).

When converting to geocode values, you must specify the latitude and longitude.

To change the data type of an attribute:

1. Select **Transform**.
2. Locate the attribute you want to change.
3. From the attribute menu, select **Quick Transformation>Convert to....**

## Editing attribute values

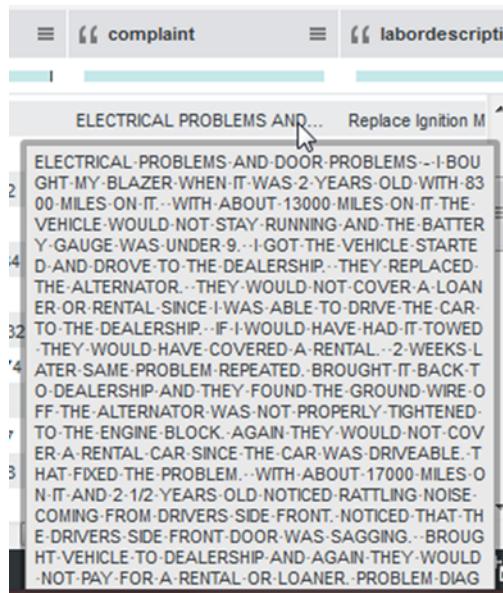
Some of the available transformations are used to make basic changes to attribute values.

### Displaying whitespace characters in attribute values

To help gauge the amount of cleanup that might be needed, you can display spaces, tabs, and paragraph breaks as whitespace characters.

To toggle whether to display the whitespace characters, in the data grid menu, click the **Show Whitespace** option.

When the whitespace characters are enabled, then when you hover the mouse over a value, the tooltip displays the full value with the whitespace characters.



## Changing the capitalization of string values

When you change the capitalization, the transformation is immediately added to the transformation script. There is no configuration, and the change always overwrites the selected attribute.

From the attribute menu, under **Quick Transformation**:

- To change the value to be all uppercase, select **String Transformation>Uppercase**.
- To change the value to be all lowercase, select **String Transformation>Lowercase**.
- To change the value to use title case, select **String Transformation>Titlecase**.

## Trimming spaces from string values

The **Trim** transformation removes white spaces from the beginning or end of string attribute values.

From the attribute menu, under **Quick Transformation**, select **String Transformation>Trim**.

When you trim white spaces, the transformation is immediately added to the transformation script. There is no configuration, and the change always overwrites the selected attribute.

## Performing mathematical operations on numeric values

For numeric attributes, you can perform mathematical operations. The available operations change based on whether the attribute is specifically an integer, long, or double value.

The options can include:

- Getting the absolute value
- Rounding decimal values

- Change a real number to the largest previous integer (floor) or the smallest following integer (ceiling).

To display the available numeric transformations, from the attribute menu, under **Quick Transformation**, select **Numeric Transformation**.

When you select a numeric transformation, it is immediately added to the transformation script. There is no configuration, and the change always overwrites the selected attribute.

## Selecting subsets of date/time values

For date/time attributes, you can truncate the date/time values to only include a specific subset of the date/time units.

For example, if the attribute values include both the date and a time, you can truncate the values to only include the date.

To truncate the values for a date/time attribute, from the attribute menu, under **Quick Transformation**, select **Date Transformation>Truncate Date**.

When you select the truncate transformation, it is immediately added to the transformation script. There is no configuration, and the change always overwrites the selected attribute.

## Splitting and grouping values

**Transform** allows you to split attribute values into two separate attributes, or create groups or bins of attribute values.

### Splitting string attribute values

The split transformation creates two new attributes from the original value. You can only split string attributes.

For example, the values of a Vehicle Type attribute for vehicles consists of the vehicle make and model separated by a colon (Honda:Civic, Toyota:Camry). You could then use the split transformation to create a Make attribute and a Model attribute.

To split string attribute values:

1. Select **Transform**.
2. Locate the attribute you want to change.
3. From the attribute menu, select **Quick Transformation>String Transformation**.
4. On the configuration panel, choose a **Split method** option and its corresponding qualifier.

You can split the value:

- Based on a delimiter.

For example, you might split the attribute after a comma or colon. If the delimiter occurs multiple times, the value is split after the first occurrence.

- At a specific position in the text.

For example, you might split the attribute after the fourth character.

So if a Vehicle Type included the year (1995 Civic, 2003 Camry), then by splitting the value after the fourth character, you could create a Year attribute and a Model attribute.

- After specific text.  
For example, split the attribute after the word "type".

The split transformation creates two new attributes. After configuring how to split the value, provide the names for the new attributes.

## Grouping attribute values

Grouping attribute values takes the attribute values and puts them into groups, with each group given a unique name.

You can use the groups to replace the current attribute values, or create a new attribute.

You might replace the current attribute values if you want to standardize values that are actually the same value, but with slightly different formatting or spelling across the data set.

For example, Walmart might show up as Wal-Mart, walmart, and Wal\*Mart. To fix these inconsistencies, you could create a Walmart group that includes all of these variants.

You might create a new attribute if you want to provide an attribute with a higher level of abstraction.

For example, you might want to group the values of a State attribute into a new Region attribute. You could then create visualizations based on either State or Region.

To group the values for an attribute:

1. Select **Transform**.
2. Locate one of the attributes that you want to group.
3. Go to the attribute menu and under **Advanced Transformation**, select **Group Values**.
4. Click **+ Value Group**.

A new empty group is added to the list, with a default name of "New Value Group N".

Each group name becomes an attribute value.

5. To rename the group, double-click the group name. Type the new name, then press **Enter**.
6. To populate a group, drag values from the value list to the group.
7. Either specify a new attribute name for the group, or apply the new attribute group to the attribute you selected in Step 2.
8. Click **Add to Script**.

## Binning numeric attribute values

The binning transformation takes values for a numeric attribute and groups them into bins. The bins are then used as the values of a new attribute.

For example, from a Price attribute, you might create a new Price Range attribute such that:

- If Price is less than 20 dollars, then Price Range is "Inexpensive".
- If Price is greater than 20 dollars but less than 40 dollars, then Price Range is "Moderately Priced".
- If Price is greater than 40 dollars but less than 60 dollars, then Price Range is "Expensive".
- If Price is greater than 60 dollars, then Price Range is "Very Expensive".

To bin the values of a numeric attribute:

1. Select **Transform**.
2. Locate one of the attributes that you want to bin.
3. From the attribute menu, under **Advanced Transformation**, select **Data Cleansing>Bin Values**.
4. For each bin that you add, specify the following:
  - The bin name.  
The bin names are used as the display name for the new attribute value.
  - The low value for the bin range.  
For the low value, you indicate whether values in the bin must be greater than or greater than or equal to the specified value.
  - The high value for the bin range.  
For the high value, you indicate whether values in the bin must be less than or less than or equal to the specified value.
5. For values that don't fit in any of the bins, you can either:
  - Place them in another bin that you name
  - Leave them as is
6. Provide the name of the new attribute to contain the bins. When you bin values, you must always create a new attribute.
7. Click **Add to Script**.

## Creating a custom transformation script

You can create a custom transformation script by writing custom Groovy code in the **Transformation Editor**.

The editor also provides access to lists of functions and attributes to insert into a transformation script. The individual Groovy functions that you can run as part of a script are documented in the "Transform Function Reference" section of the *Oracle Big Data Discovery Extensions Guide*.

Also note that like other scripting languages, Groovy has a set of reserved keywords that have special meanings and therefore can't be used as variable or function names in transformation scripts. For details, see "Unsupported Groovy language features and Reserved Keywords" section of the *Oracle Big Data Discovery Extensions Guide*.

To create a custom transformation script:

1. Select **Transform**.
2. To display the **Transformation Editor**, either:
  - From the Transform header, select **Show Transformation Editor**.  
If you select this option, then you can either create a new attribute, or replace the values in a selected attribute.
  - From the data grid menu, select **Add Attribute**.  
If you select this option, then you always create a new attribute.

3. To insert the current refinement state into the script, click the **Use refinement state as a conditional statement** link.
4. To insert a function into the script:
  - (a) Click the **Functions** button.
  - (b) To find a specific function, use the filter field.
  - (c) To add a function, either double-click it, or drag and drop it into the script.
5. To insert an attribute name into the script:
  - (a) Click the **Attributes** button.
  - (b) To find a specific attribute, use the filter field. You can filter the list by name or data type.
  - (c) To add an attribute, either double-click it, or drag and drop it into the script.
6. If you are creating a new attribute, then provide the attribute name in **New Attribute Name**.
7. From the **Data Type** list, select the data type to assign to the resulting attribute.
8. Click **Add to Script**.

## Managing and executing the transformation script

The transformation script tracks and executes the transformations.

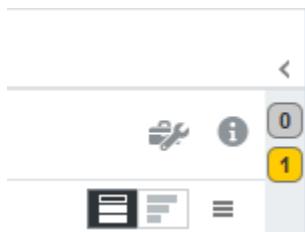
*[Reordering, editing, and deleting transformations](#)*

*[Running the transformation script against the project data set](#)*

### Reordering, editing, and deleting transformations

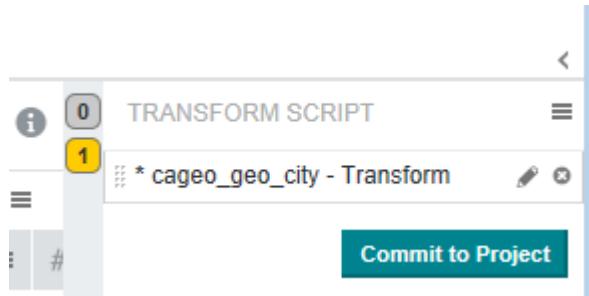
The transformation script is contained in a collapsible panel at the right of the **Transform** page.

As you create each additional transformation, it is added to the end of the transformation script. The number of uncommitted transforms in a script display in the yellow bubble. For example, this transformation script has one uncommitted transform:



To display the list of transforms in the transformation script, click the bubble.

Transformations that haven't been committed to a project data set are marked by \*. For example:



From the **Transform Script** editor, you can:

- Change the order in which to commit the transformations to the project.  
The transformations are committed in the order they appear. To change the order, drag and drop the transformations to the new location in the list.
- Edit individual transformations.  
To edit a transformation, click its pencil icon. You can modify the Groovy code in the **Transformation Editor** as necessary.
- Delete individual transformations.  
To delete a transformation, click its delete icon.

## Running the transformation script against the project data set

The project data set is not affected by the transformations until you run the transformation script against the project's data set. Studio commits the transformations to the data sample that is currently associated with the project. Also, if you later change the sample size, the transformations that you have already applied are automatically applied to the sample.

To run the transformation script against the project data set:

1. Select **Transform** if you haven't already.
2. Expand the **Transformation Script** panel by clicking the grey bubble that indicates the project contains transforms that have not yet been committed.
3. Click **Commit to Project**.

**Commit to Project** runs the transformations against the project data set but does not modify the underlying data set as it appears in the **Catalog** to other Studio users. This means that other Studio users can explore the untransformed data set in the **Catalog**, and project users can explore the transformed data set.

If you want to run transformations and make them available to other Studio users, commit the transformations to the project, and then create a new data set. The resulting new data set contains the transform changes and is available in the **Catalog**. See [Creating a new data set from the transformed data on page 175](#).

## Creating a new data set from the transformed data

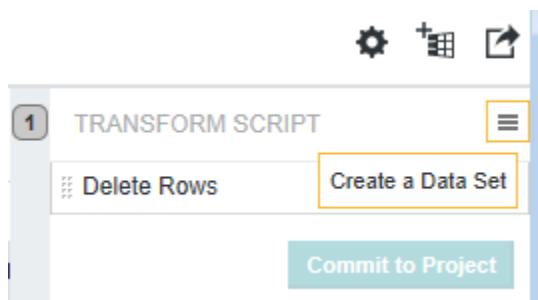
After running a transformation script, you can create a new data set in the **Catalog** based on the transformed project data. This is sometimes called forking a data set.

If you create a new data set, Studio runs the transformation script against the entire data set, not just the data sample in the project, and Studio applies the transformations, filters, and enrichments to the new data set. A new sample is generated for the new data and the data set is available in the **Catalog**.

To create a new data set from the transformed data:

1. From the transformation script menu, select **Create a Data Set**.

For example:



The grid at the top of the dialog shows a preview of the data, based on the current transformation script.

2. In the **New Data Set Name** field, type the name of the new data set.
3. If desired, provide any notes about the new data set in the **New Data Set Description** field.
4. In the **New Hive Table Name** field, provide a table name for this data set in the Hive database. By default, the table name matches the data set name.
5. If desired, modify the **New Hive Table Data Directory** field. By default, new tables are stored under /user/bbd. For example, if you create a new data set named `customers`, it is stored as /user/bbd/customers.
6. Click **Save**.
7. Click **Ok** to dismiss the information message.

(You can check the progress of this process by looking at the Oozie Web UI tool.)

After BDD finishes creating a new table in the Hive database and performing some additional data processing, the data set becomes available in the **Catalog**.



## Chapter 26

# Linking Project Data Sets

You can link together project data sets that have corresponding attributes.

[About data set links](#)

[Creating a data set link](#)

[Editing a link key](#)

[Managing aggregation methods for a linked view](#)

[Controlling automatic filtering between linked data sets](#)

## About data set links

You create a link between two data sets by selecting one or more attributes from one data set and associating them with corresponding attributes from another data set. These corresponding pairs of attributes are called a link key.

For each data set that is linked to others, Big Data Discovery creates a linked view, which contains attributes from that data set, as well as those it is linked to. For more information, see [About linked views on page 186](#).

While you cannot edit the linked views themselves, you can modify the data that appears in them by adding or removing link key attributes, changing the methods used to aggregate attributes within the view, and enabling or disabling automatic filtering.

## Creating a data set link

You create a link between two data sets by setting one or more corresponding attributes from each as the link key. You can do this from either the **Data Set Manager** or the **Data Set Relationships** tab of the data set details footer panel.

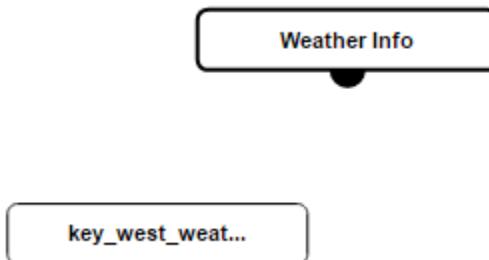
To create a data set link:

1. Open either the **Data Set Manager** or the **Data Set Relationships** panel:
  - To open the **Data Set Manager**, click **Project Settings**, then select **Data Set Manager**.
  - To open the **Data Set Relationships** panel, click the data set details icon in the footer, then select the **Data Set Relationships** tab.

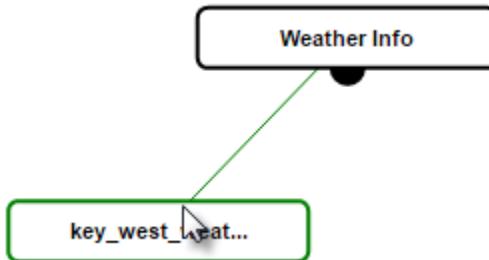
A diagram of the data sets in your project displays.

2. In the diagram, click one of the data sets to select it.

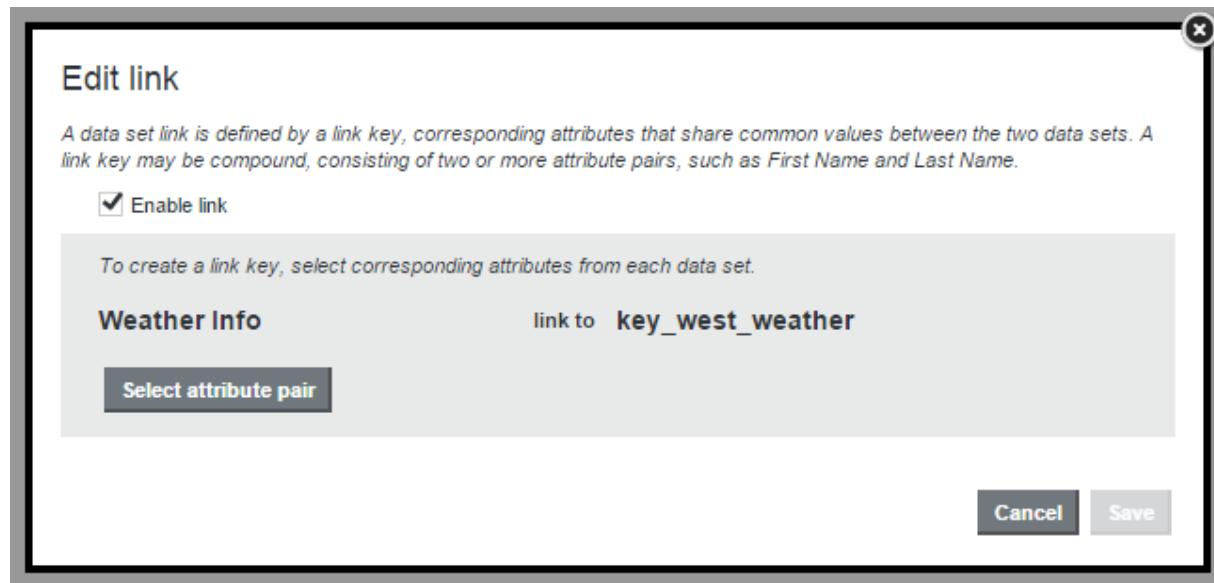
A handle appears below the selected data set.



3. Click and drag the handle to the data set you want to link to and release the mouse when the second data set becomes highlighted in green.



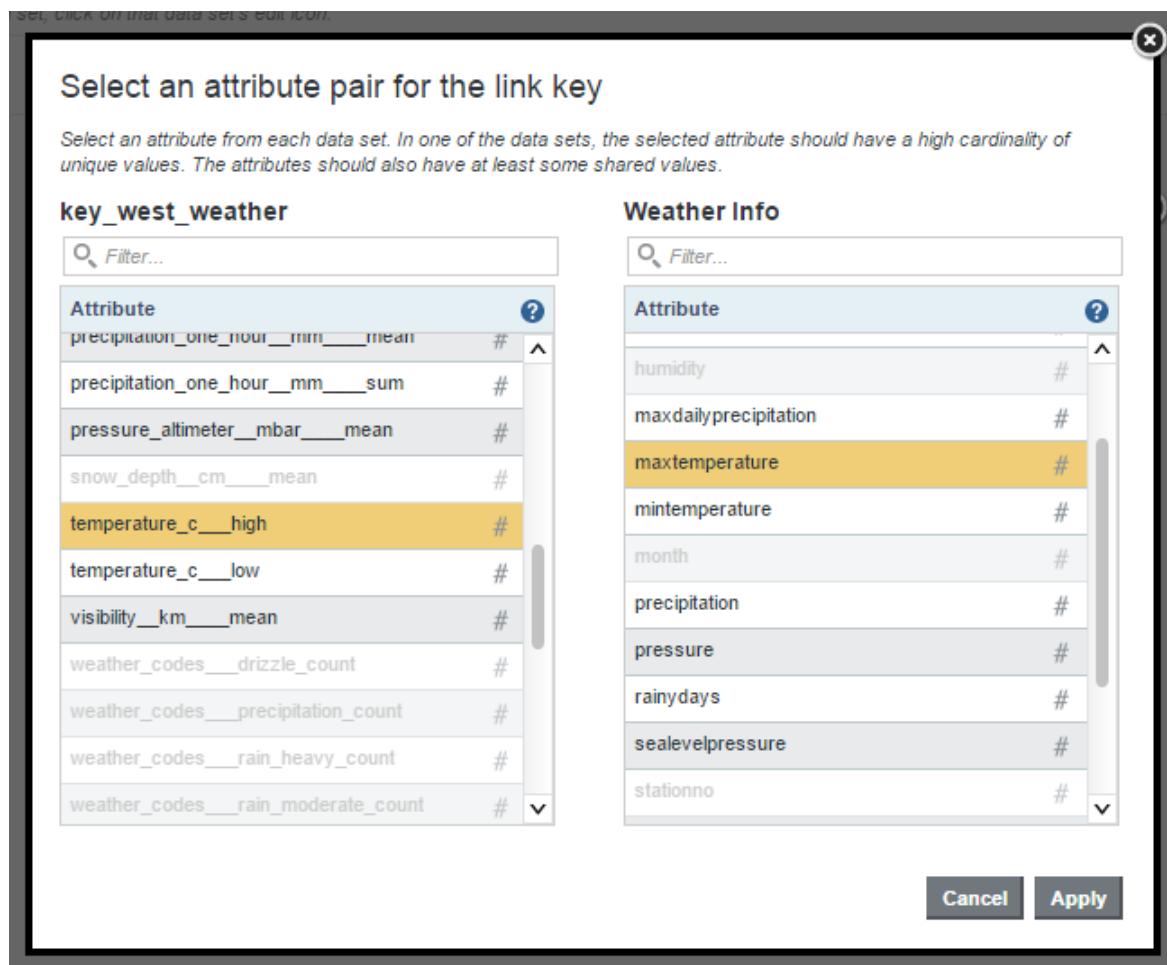
The **Edit Link** dialog opens.



4. Verify that the **Enable link** check box is checked.  
Unchecking this disables the link.
5. Add at least one attribute pair:
  - Click **Select attribute pair**.

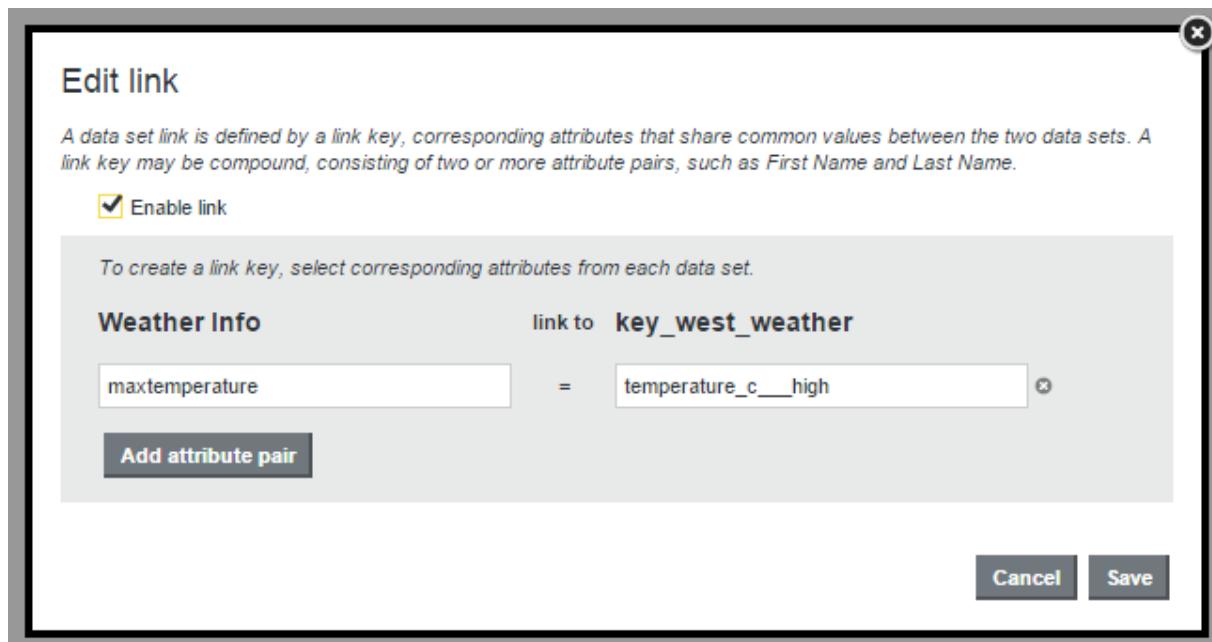
- (b) In the attribute pair selection dialog, select an attribute from each data set.

When you select an attribute from one list, the other is automatically filtered for attributes of the same data type.



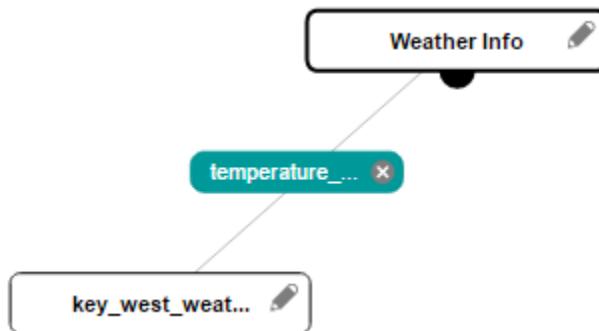
(c) Click **Apply**.

The **Edit link** dialog is updated with the selected attribute pair.



## 6. Click **Save**.

The diagram displays the two linked data sets with the link key between them.



## Editing a link key

You can edit a link key by adding or removing attribute pairs.

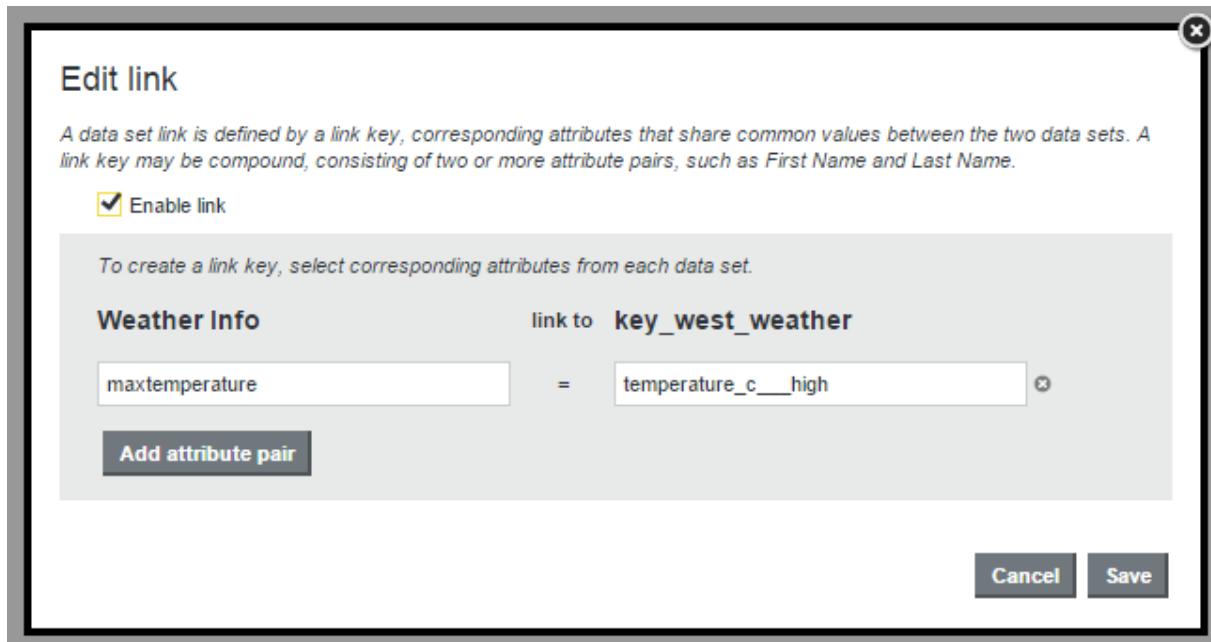
To edit the link key for a data set link:

1. Open either the **Data Set Manager** or the **Data Set Relationships** panel:
    - To open the **Data Set Manager**, click **Project Settings**, then select **Data Set Manager**.
    - To open the **Data Set Relationships** panel, click the data set details icon in the footer, then select the **Data Set Relationships** tab.

A diagram of the data sets and linked views in your project displays.

2. Click the link you want to edit.

The **Edit link** dialog displays.



3. Make the desired changes:

- Click **Add attribute pair** to add a new attribute pair.
- Click the delete icon next to an attribute pair to remove it.

4. When you are satisfied with your changes, click **Save**.

The combined views in your project are updated to reflect your changes.

## Managing aggregation methods for a linked view

When you first create a data set link, the attributes in the linked view are aggregated by the default aggregation methods set for their data types. However, you can change the aggregation methods used for individual attributes from either the **Data Set Manager** or the **Data Set Relationships** tab of the data set details footer panel.

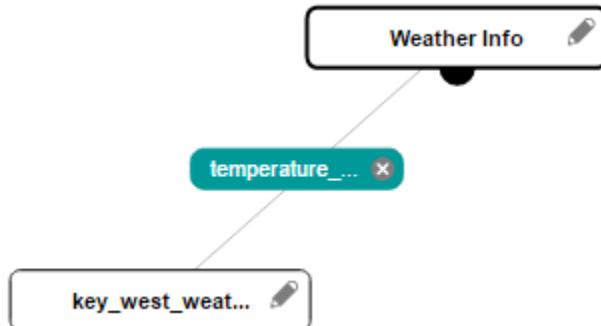
For more information on aggregation methods, see [Selecting the available and default aggregation methods for an attribute](#).

To manage the aggregation methods for a linked view:

1. Open either the **Data Set Manager** or the **Data Set Relationships** panel:
  - To open the **Data Set Manager**, click **Project Settings**, then select **Data Set Manager**.

- To open the **Data Set Relationships** panel, click the data set details icon in the footer, then select the **Data Set Relationships** tab.

A diagram of the data sets and linked views in your project displays.



- Click the pencil icon for the view you want to modify.

The manage aggregation dialog displays.

If the selected view contains more than one data set, the dialog contains a drop down menu for selecting the data set whose aggregation methods you want to modify.

The dialog also contains a table that displays sample values from the selected data set. The two leftmost columns contain values from the Record ID and Link Key attributes of the selected view.

The remaining columns contain attributes and values from the selected data set. The aggregation method used for each attribute is listed below the attribute's name.

Manage aggregation for the key\_west\_weather view

When a data set is linked to other data sets, a custom view is automatically created. For the data sets that are linked to this data set, use the table below to configure the aggregation for the data set attributes.

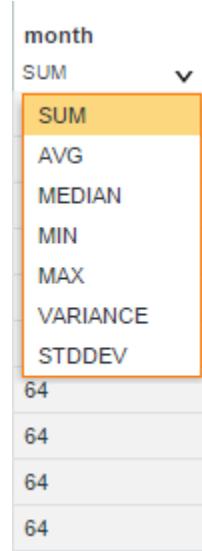
key_west_weather data (730 records)		Weather Info data						
PRIMARY_KEY	temperature_...	month	rainydays	vaporpressure	precipitation	humidity	windspeed	
Record ID	Link Key	SUM	SUM	SUM	SUM	SUM	SUM	
0-0-210	33.9	43	32.2		373.1			
0-1-177	33.9	43	32.2		373.1			
0-0-209	33.9	43	32.2		373.1			
0-0-203	33.9	43	32.2		373.1			
0-1-176	33.3	64	51.8	11.4	638.4		2.2	
0-0-233	33.3	64	51.8	11.4	638.4		2.2	
0-0-208	33.3	64	51.8	11.4	638.4		2.2	
0-0-232	33.3	64	51.8	11.4	638.4		2.2	
0-1-163	33.3	64	51.8	11.4	638.4		2.2	
0-0-226	33.3	64	51.8	11.4	638.4		2.2	

**Close** **Save**

3. If the selected view is linked to multiple data sets, select the name of the data set you want to modify from the drop down menu.
4. Click the down arrow next to the aggregation method you want to modify.

Note that this will not be available for String attributes.

A drop down menu displays, listing the available aggregation methods for that attribute.



5. Select one of the available aggregation methods, then click **Save**.

The dialog is updated to display the new values for the modified attribute.

Manage aggregation for the key\_west\_weather view

Your aggregations were successfully modified

When a data set is linked to other data sets, a custom view is automatically created. For the data sets that are linked to this data set, use the table below to configure the aggregation for the data set attributes.

key_west_weather data (730 records)		Weather Info data						
PRIMARY_KEY	temperature_...	month	rainydays	vaporpressure	precipitation	humidity	windspeed	
Record ID	Link Key	MEDIAN	SUM	SUM	SUM	SUM	SUM	
0-0-350	19.4	4	147.4	133.6	1,763	717	29.7	
0-0-345	19.4	4	147.4	133.6	1,763	717	29.7	
0-1-17	19.4	4	147.4	133.6	1,763	717	29.7	
0-1-364	22.8	5	133.9	136.8	1,789.1	613	26	
0-1-359	22.8	5	133.9	136.8	1,789.1	613	26	
0-1-6	22.8	5	133.9	136.8	1,789.1	613	26	
0-1-355	22.8	5	133.9	136.8	1,789.1	613	26	
0-1-35	18.3	5	98.7	103.5	1,370.7	579	23.1	
0-0-76	22.8	5	133.9	136.8	1,789.1	613	26	
0-1-1	22.8	5	133.9	136.8	1,789.1	613	26	

**Close** **Save**

6. When you are finished, click **Close**.

## Controlling automatic filtering between linked data sets

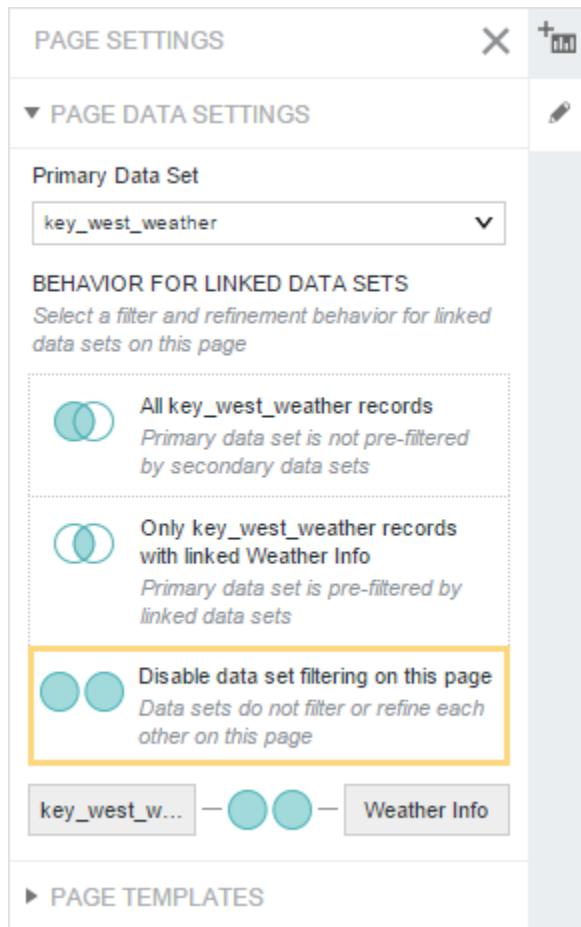
Automatic filtering of linked data sets is enabled by default, but you can disable it for individual project pages from the **Page Data Settings** panel.

When automatic filtering is enabled for a pair of linked data sets, you can filter both data sets simultaneously using attributes from their link key. For example, two data sets called Sales and Products are linked based on product SKU and have automatic filtering enabled. If you filter Sales for a specific range of SKU values, Products is also filtered for the same range of values automatically, and vice versa.

To control automatic filtering for a pair of linked data sets:

1. On the project page for the linked view, click the pencil icon in the right bar.
- The **Page Settings** panel opens.
2. Click **Page Data Settings**.
- The panel displays the data settings for the current page. By default, the **All <data set> records** option is selected

3. Select the **Disable data set filtering on this page**.



The page is updated to reflect the change. Refinements applied to both data sets are now only applied to their original data sets.



## Chapter 27

# Defining Views of Project Data

You can use the **Views** page to configure a set of views for project data. Views are used by components to determine the data to display.

[About views](#)

[About base views](#)

[About linked views](#)

[Displaying the list of views for a project](#)

[Displaying details for a view](#)

[Creating views](#)

[Configuring a view](#)

[Previewing the content of a view](#)

[Determining whether a custom view is published](#)

[Deleting a view](#)

## About views

A view is a logical collection of information that is derived from the records in project data sets.

For example, the original records in a data set may represent individual sales transactions. However, from those transaction records, you could derive:

- A list of products
- A list of customers
- A summary of sales data grouped by year

Each of those could be a view.

All views are made up of attributes. Some attributes may be the physical attributes from a data set, while others may be calculated from the original data.

For example, for a list of products derived from a list of sales transactions, the product name comes directly from the data, while the total sales for that product is calculated from the individual sales records.

Views can optionally have predefined metrics, which are values calculated from view attributes using EQL expressions.

For example, for sales transaction data, a predefined metric might be the profit margin.

## About base views

Each data set is automatically configured with a base view, which contains all of the physical attributes from the data set. The base view uses the name and description from the data set.

While you cannot change the definition of a base view, or add predefined metrics to a base view, you can configure base view attributes.

## About linked views

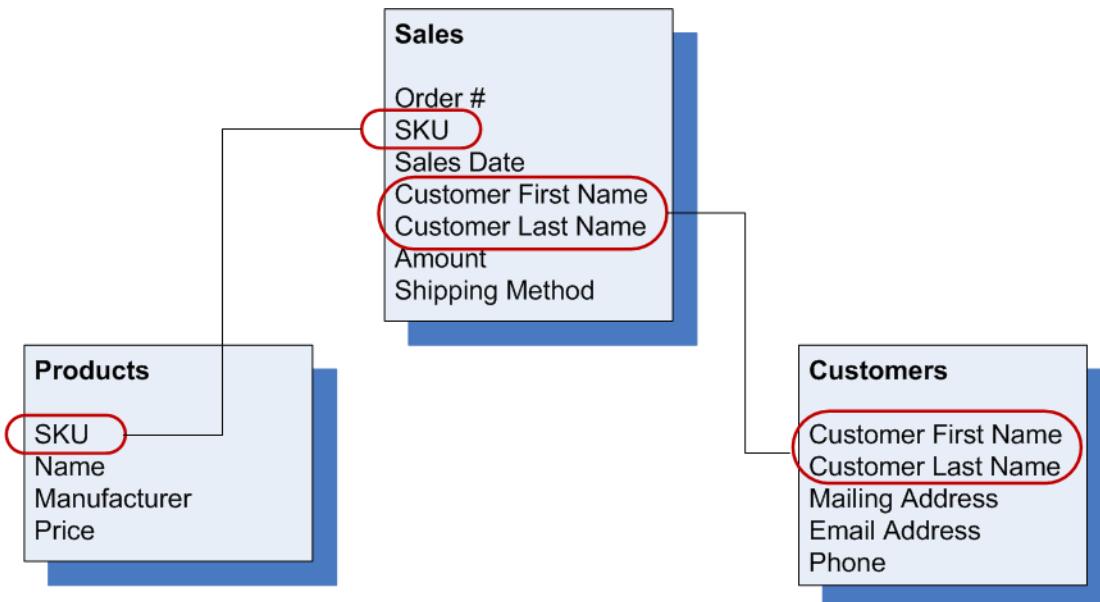
Linked views are created automatically when you link two data sets.

For each data set that has links to other data sets, Big Data Discovery creates a combined linked view that contains:

- All of the attribute for that data set
- The attributes from the other data sets, aggregated based on the link key

Linked views are read-only, and do not allow for any configuration or customization. You also cannot copy a linked view.

So for the following links among the Products, Customers, and Sales data sets:



Big Data Discovery creates the following linked views:



Linked View	Description
Products - linked	<p>Contains the Products records with Sales attributes added in. The Sales attributes are aggregated by product SKU.</p> <p>For example:</p> <ul style="list-style-type: none"> <li>The sales amount (from Sales) for each product is the total for all of the sales for the product.</li> <li>The customer name (also from Sales) for each product is a list of customers that purchased the product.</li> </ul>
Customers - linked	<p>Contains the Customers records with Sales attributes added in. The Sales attributes are aggregated by the Customer First Name and Last Name.</p> <p>For example:</p> <ul style="list-style-type: none"> <li>The sales amount (from Sales) for each customer is total for all sales to that customer.</li> <li>The product SKU (from Sales) for each customer is a list of SKUs for products purchased by that customer.</li> </ul>

Linked View	Description
Sales - linked	<p>Contains the Sales records with the Products and Customers attributes added in.</p> <p>The Products information is aggregated by product SKU, and the Customers information is aggregated by the Customer First Name and Customer Last Name.</p> <p>In this case, because each sales record has a single product and customer, there is still only a single value for each product and customer field.</p>

## Displaying the list of views for a project

On the project settings page, the **Data Views** page allows you to view and edit the views for the project data.

To display the list of views for a project:

1. On the project menu bar, click the settings icon.
2. In the project settings menu, click **Data Views**.

On the **Data Views** page, the **Data Views** list is populated with the views for the project. For each view, the list shows the display name, key, type, source data sets, and identifying attributes for the view records. The first base view is selected by default.

### Data Views

Configure the data views for your project. Select a view from the list at the top to see and configure its details.

Views							<a href="#">+ View</a>
Publish	View Name	View Key	Type	Data Set(s) <a href="#">?</a>	Record Identifier(s)	Actions	
✓	Product	edp_cli_edp_39361033...	Base	Product	PRIMARY_KEY	<a href="#">Edit</a>	
✓	Sales	edp_cli_edp_7a6bae5d...	Base	Sales	PRIMARY_KEY	<a href="#">Edit</a>	
✓	Product - linked	edp_cli_edp_39361033...	Linked	Sales, Product	PRIMARY_KEY	<a href="#">Edit</a>	
✓	Sales - linked	edp_cli_edp_7a6bae5d...	Linked	Product, Sales	PRIMARY_KEY	<a href="#">Edit</a>	

Selected View: Product (edp_cli_edp_39361033-51e8-4e08-810e-914054a1046c_bu\$g 0307_wls_ENTITIES_10850)																																															
Display name:		Default view description:		Record Identifier(s):		Actions																																									
Product <a href="#">?</a>		<a href="#">?</a>		PRIMARY_KEY		<a href="#">Edit</a>																																									
<b>Attributes</b>																																															
<table border="1"> <thead> <tr> <th>Display Name</th> <th>Attribute Key</th> <th>Refinable</th> <th>Multi-Value</th> <th>Data Type</th> <th>Standard Format</th> <th>Dimension</th> <th>Available Aggregations</th> </tr> </thead> <tbody> <tr> <td>product_category</td> <td>product_category</td> <td>✓</td> <td>---</td> <td>“ String</td> <td>(string)</td> <td>✓</td> <td>set, Number of records</td> </tr> <tr> <td>product_class</td> <td>product_class</td> <td>✓</td> <td>---</td> <td>“ String</td> <td>(string)</td> <td>✓</td> <td>set, Number of records</td> </tr> <tr> <td>product_color</td> <td>product_color</td> <td>✓</td> <td>---</td> <td>“ String</td> <td>(string)</td> <td>✓</td> <td>set, Number of records</td> </tr> <tr> <td>product_daysto...</td> <td>product_daysto...</td> <td>✓</td> <td>---</td> <td># Number (long)</td> <td>Number</td> <td>✓</td> <td>sum, average, media</td> </tr> </tbody> </table>								Display Name	Attribute Key	Refinable	Multi-Value	Data Type	Standard Format	Dimension	Available Aggregations	product_category	product_category	✓	---	“ String	(string)	✓	set, Number of records	product_class	product_class	✓	---	“ String	(string)	✓	set, Number of records	product_color	product_color	✓	---	“ String	(string)	✓	set, Number of records	product_daysto...	product_daysto...	✓	---	# Number (long)	Number	✓	sum, average, media
Display Name	Attribute Key	Refinable	Multi-Value	Data Type	Standard Format	Dimension	Available Aggregations																																								
product_category	product_category	✓	---	“ String	(string)	✓	set, Number of records																																								
product_class	product_class	✓	---	“ String	(string)	✓	set, Number of records																																								
product_color	product_color	✓	---	“ String	(string)	✓	set, Number of records																																								
product_daysto...	product_daysto...	✓	---	# Number (long)	Number	✓	sum, average, media																																								

## Displaying details for a view

To display the details for a view, in the views list, click the view name.

The view details section is populated with

- Whether the view is published.

If the view is not published, then it is not available to use for components. You cannot publish views that do not have a valid view definition.

- View name
- View description
- Identifying attributes for the view
- Attributes list.

The attributes list does not display for views that do not have a valid view definition.

- Predefined metrics list.

The predefined metrics list does not display for views that do not have a valid view definition.

✓	Sales	edp_cli_edp_7a6bae5d...	Base	Sales	PRIMARY_KEY	
✓	Product - linked	edp_cli_edp_39361033...	Linked	Sales, Product	PRIMARY_KEY	
✓	Sales - linked	edp_cli_edp_7a6bae5d...	Linked	Product, Sales	PRIMARY_KEY	

Selected View: Sales (edp\_cli\_edp\_7a6bae5d-4109-41bd-962d-4fcf2a1d6061\_busgl0307\_wls\_ENTITIES\_10850)

[View Definition](#) [Preview](#) [Revert to Last Save](#) [Save View](#)

Display name: Sales [?](#)

Default view description: [?](#)

Record Identifier(s): PRIMARY\_KEY

**Attributes**

Display Name	Attribute Key	Refinable	Multi-Value	Data Type	Standard Format	Dimension	Available Aggregat
sales_activityd...	sales_activitydates	✓	✓	Date/Time 	DateTime 	<input type="checkbox"/> 	set 
sales_calendar...	sales_calendarqu...	✓	---	# Number (long) 	Number 	<input checked="" type="checkbox"/> 	sum, average, m...
sales_calendar...	sales_calendarse...	✓	---	# Number (long) 	Number 	<input checked="" type="checkbox"/> 	sum, average, m...
sales_calendar...	sales_calendaryear	✓	---	# Number (long) 	Number 	<input checked="" type="checkbox"/> 	sum, average, m...
sales_carriertr...	sales_carriertrack...	✓	---	!! String 	(string) 	<input checked="" type="checkbox"/> 	set, Number of re...
sales_customers...	sales_customerpo...	✓	---	!! String 	(string) 	<input checked="" type="checkbox"/> 	set, Number of re...
sales_daynam...	sales_daynameof...	✓	---	!! String 	(string) 	<input checked="" type="checkbox"/> 	set, Number of re...

## Creating views

From the **Views** page, you can either create a new empty view, or make a copy of an existing view.

[Creating a new view](#)

[Copying an existing data view](#)

## Creating a new view

One option for creating a new view is to create a completely new view.

To create a new view:

1. Select **Project Settings > Data Views**.
2. On the views list, click **+ View**.

The **Create View Definition** dialog displays.

3. In the **Display Name** field, type a name for the view.
4. In the **Description** field, type a description of the view.
5. In the EQL text area, enter the view definition.

For details on defining a view definition, see [Guidelines for using EQL to define a view on page 194](#).

To the right of the text area is the **Available Attributes** list, which contains the list of base attributes. You can show or hide this list.

You can use the attributes list as a reference as you compose the view definition. To add an attribute to the definition, you can also double-click or drag and drop the attribute from the list. Big Data Discovery automatically adds quotes around the inserted attribute, in case the attribute name contains reserved characters in EQL.

Note that if you include a base attribute that has locale-specific versions, then for those locale-specific versions to be available in the new view, you must manually include them in the view definition.

6. By default, when you save a new view, it is automatically published. To not publish the new view automatically, uncheck the **Publish when saving** checkbox.
7. To validate the definition, click **Validate**.  
If the definition is not valid, then Big Data Discovery displays error messages to help you troubleshoot the definition.
8. To save the new view, click **Save**.  
If the view definition is valid, then the view is saved and can be published.  
If the view definition is not valid, then you are prompted to confirm that you want to save an invalid view. If you save the view, it is flagged as invalid and cannot be published.

## Copying an existing data view

You can create a data view by copying another data view. You cannot copy a linked data view. The new view initially has the same definition, attributes, and metrics as the original view. Big Data Discovery also copies the attribute groups.

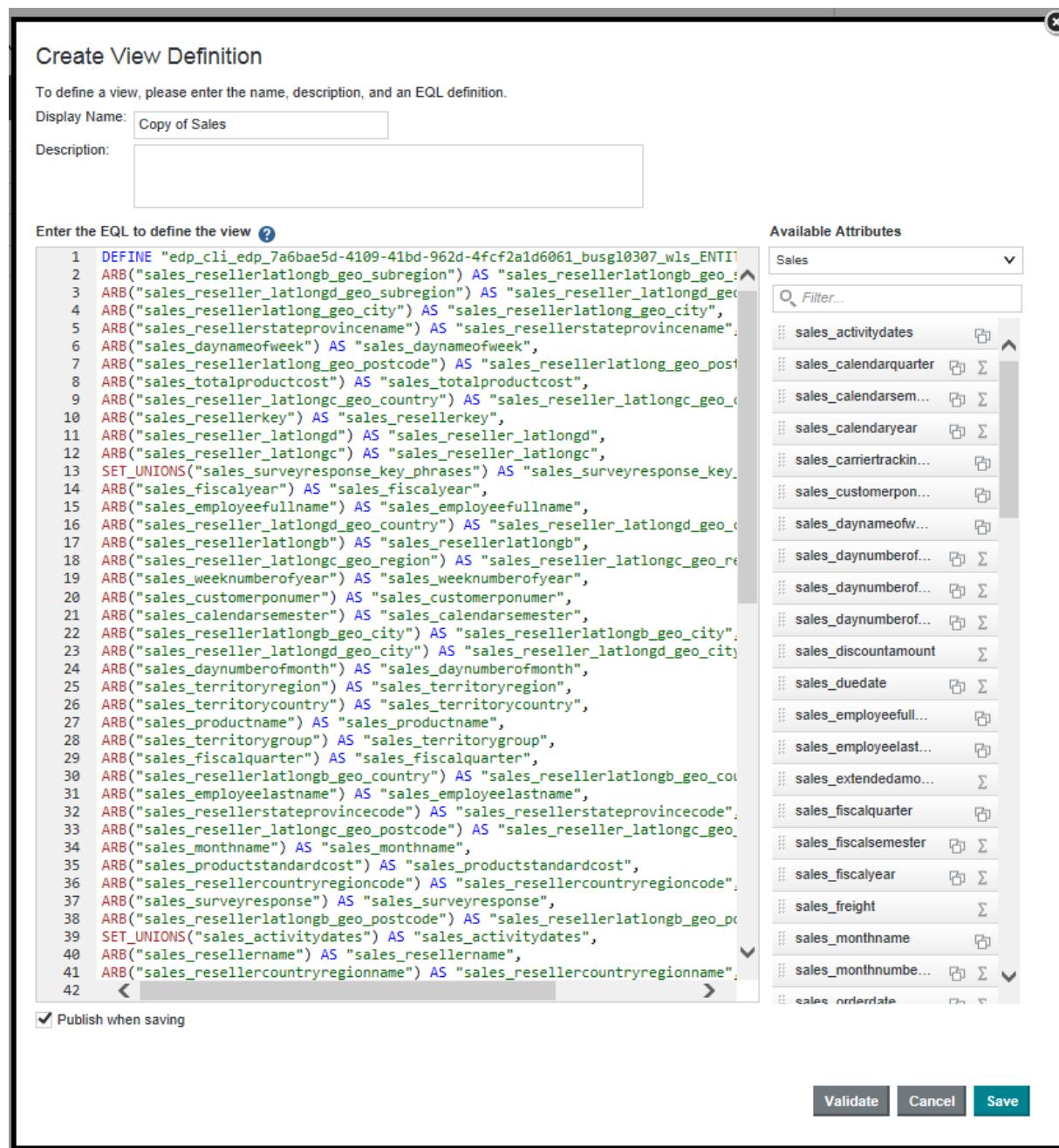
To copy an existing:

1. Select **Project Settings**.
2. Select **Data Views**.
3. Identify the view you want to copy.

4. In the **Actions** column, click the copy icon for the data view you want to copy.

The **Create View Definition** dialog is displayed. By default:

- The view display name is "Copy of <copiedViewName>".
- The view key in the view definition is set to "<copiedViewKeyName>Copy".
- The view description and the rest of the view definition are identical to the copied view.



5. Update the view name, description, view key, and definition as needed.

Note that you cannot change the view key after you save the new view, so make sure you update the value before saving.

Remember also that the view key must be NC-Name compliant.

6. By default, when you save a new view, it is automatically published. To not publish the new view automatically, uncheck the **Publish when saving** checkbox.
7. To validate the new view, click **Validate**.
8. To save the new view, click **Save**.

## Configuring a view

From the **Views** page, you can edit each view's display name, description, definition, attributes, and predefined metrics.

[Editing and localizing the view display name and description](#)

[Editing the view definition](#)

[Guidelines for using EQL to define a view](#)

[Configuring the view attributes](#)

[Defining predefined metrics for a view](#)

## Editing and localizing the view display name and description

On the **Data Views** page, when you select a view from the view list, the view details include the **Display name** and **Default view description** fields.

Selected View: Sales (edp\_cli\_edp\_7a0bae5d-4109-41bd-962d-4fcf2a1d0061\_busgl0307\_wls\_ENTITIES\_10850)

Selected View: Sales (edp_cli_edp_7a0bae5d-4109-41bd-962d-4fcf2a1d0061_busgl0307_wls_ENTITIES_10850)	
Display name:	Sales
Default view description:	
Record Identifier(s):	PRIMARY_KEY

View Definition   Preview   Revert to Last Save   Save View

The display name and description are used when showing the list of available views to assign to a component.

For custom views, you can create localized versions of the name and description.

Base views use the name and description of the data set. To edit or localize the name and description of a base view, you edit the data set.

Linked views use automatically generated names and descriptions.

To edit and localize the view display name and description for a custom data view:

1. Copy a base view if you haven't already. This is your custom data view.
2. To edit the default display name and description:
  - (a) In the **Display name** field, type the display name for the view.

- (b) In the **Default view description** field, type a description of the view.
3. To localize the view display name and description, click the localize icon next to either field.
4. To localize the view display name:
  - (a) On the **Configure Locale Options** dialog, click the **Display Name** tab.

The **Default display name** field contains the default name. If you edit the name here, it will be reflected on the main **Views** page.
  - (b) From the locale drop-down list, select a locale for which to provide a version of the display name.
  - (c) To use the default display name for the selected locale, leave the **Use default value** checkbox checked.
  - (d) To provide a localized display name, uncheck the checkbox. In the field, type the localized name.
5. To localize the view description:
  - (a) On the **Configure Locale Options** dialog, click the **Description** tab.

The **Default description** text area contains the default description. If you edit the description here, it is reflected on the main **Views** page.
  - (b) From the locale drop-down list, select a locale for which to provide a version of the description.
  - (c) To use the default description for the selected locale, leave the **Use default value** checkbox checked.
  - (d) To provide a localized description, uncheck the checkbox. In the text area, type the localized description.
6. After providing all of the localized versions of the name and description, click **Apply**.
7. To save the changes to the view, click **Save View**.

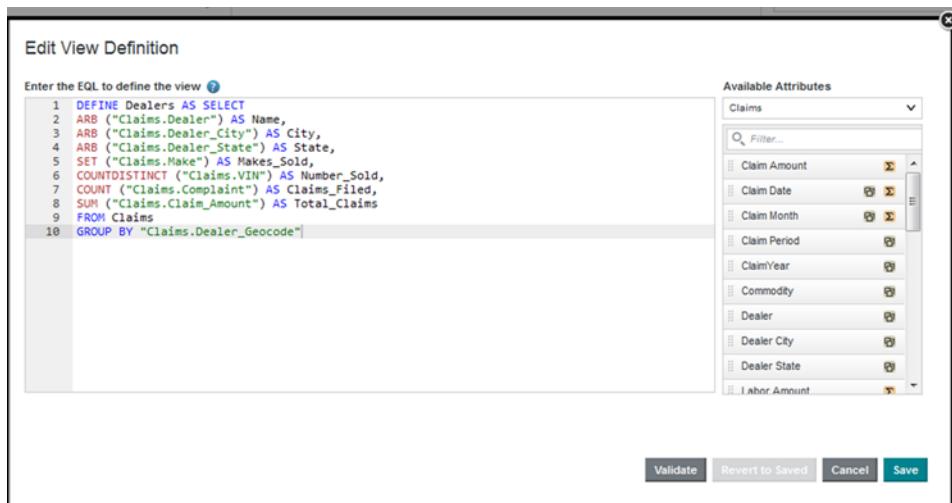
## Editing the view definition

From the **Views** page, after you create a custom view, you can update the view EQL definition. Note that if you change the definition, you could affect components that are using the view, as well as other views that may use attributes from this view.

To edit a view definition:

1. Select **Project Settings>Data Views**.
2. In the view list, click the view.
3. Click **Edit Definition**.

4. On the **Edit View Definition** dialog, update the view definition.



For details on the rules for setting a view definition, see [Guidelines for using EQL to define a view on page 194](#).

5. To validate the new definition, click **Validate**.

If the definition is invalid, then Big Data Discovery displays error messages to help you troubleshoot the definition.

6. To save the new definition, click **Save**.

If the view definition is valid, then the attribute and metrics lists are updated to reflect the new definition.

If the view definition is not valid, then you are prompted to confirm that you want to save an invalid view. If you save the view, it is flagged as invalid and cannot be published.

## Guidelines for using EQL to define a view

The view definition EQL defines the content of the view. When defining a view, you need to keep the following guidelines in mind. For complete details on EQL syntax, see the *EQL Reference*.

### View **DEFINE** and **SELECT** statements

A view definition may contain multiple **DEFINE** and **SELECT** statements.

The last statement must be a **DEFINE** statement that contains all of the view attributes. The name of that **DEFINE** statement is used as the view key, and cannot be changed. If you try to change the view key for an existing view, the corresponding components will no longer work.

### Providing the source view for the view attributes

For all custom view definitions, you must use the **FROM** parameter to provide the original source of the view attributes.

For example, if you create a copy of a base view, you will see that the view definition includes the view key for that base view.

```

DEFINE Customers as SELECT
ARB(Customer_Name) as Customer_Name,
ARB (Customer_Address) as Customer_Address,
ARB(Cities) as Cities,
ARB(States) as States,
ARB(CustomerZip) as CustomerZip,
ARB(LatLong) as LatLong,
ARB(Business_Types) as Business_Types,
ARB(CustomerAgreedDaysCredit) as CustomerAgreedDaysCredit,
ARB(Credit_Rating) as Credit_Rating,
ARB(CustomerDiscount) as CustomerDiscount,
SUM(Number_of_Cases_Sold) as TotalCustomerCases,
CountDistinct(Transaction_Id) as TotalCustomerTransactions,
SUM(GrossDollars) as TotalCustomerGross,
SUM(MarginDollars) as TotalCustomerMargin,
SET(Shipping_Companies) as ShippingCompaniesUsed
FROM "wine-sales"
GROUP BY Customer_Id

```

When identifying the source view, make sure to use the view key, and not the view display name.

Also, when referring to another view, you can only refer to the final **DEFINE** statement that identifies the attributes for other view. You cannot refer to any intermediate **DEFINE** or **SELECT** statements in that view definition.

## Grouping a view

All custom views should have at least one grouping (GROUP BY) attribute.

Big Data Discovery uses the GROUP BY attributes as the identifying attributes for the view records.

If there are no GROUP BY attributes, then there are no identifying attributes, and the view cannot be used for **Record Details** or **Compare**. On component edit views, the **Record Details** and **Compare** options are then automatically disabled.

## Using multi-value attributes for grouping

When using a multi-value attribute for grouping, you must indicate whether to group by the individual values, or by the sets of values assigned to the records.

- To group by the sets of values, use GROUP BY.
- To group by the individual values, use GROUP BY MEMBERS. The syntax for GROUP BY MEMBERS is slightly different from GROUP BY:

```
GROUP BY MEMBERS(attributeName) as attributeName
```

For example, the following data shows available product colors, and the number of outlets that carry each product:

ItemColors	AvailableOutlets
Red, Blue	3
Red	5
Red, Blue, White	6

ItemColors	AvailableOutlets
Red, Blue, White	4
Red, Blue	1

If you are calculating the total number of available outlets grouped by the item color, then:

- If you use `DEFINE NewView as SELECT TotalOutlets as sum(AvailableOutlets) FROM Sales GROUP BY ItemColors`, then the result is:

ItemColors	TotalOutlets
Red, Blue	4
Red	5
Red, Blue, White	10

- If you use `DEFINE NewView as SELECT TotalOutlets as SUM(AvailableOutlets) FROM Sales GROUP BY MEMBERS(ItemColors) as ItemColors`, then the result is:

ItemColors	TotalOutlets
Red	19
Blue	14
White	10

## Providing aggregation methods when grouping view attributes

When defining a view, if you are using grouping, then you must provide an aggregation method for all of the attributes in the view, except for the grouping attributes.

If you are using aggregation to derive attributes from the physical attributes, then you can use the applicable aggregation methods for the attribute data type. See [Aggregation methods and the data types that can use them on page 207](#).

In some cases, you may want to display the actual values of physical attributes in a grouped view. For example, when you are grouping by product identifier, then you may want to include the product name, product description, and available sizes. When using grouping for a view, to include the actual values of physical attributes:

- For a single-value attribute, use the `ARB` aggregation method. For example:

```
SELECT ARB(Region) as Region
```

Note that if there are different values for an attribute within the grouping, Big Data Discovery chooses a single, arbitrary value.

- For a multi-value attribute:

- To get a single, arbitrary value for the attribute from each record, use the ARB aggregation method. For example:

```
SELECT ARB(ItemColors) as ItemColors
```

- To get the complete set of values for the records, use the SET\_UNIONS aggregation method. For example:

```
SELECT SET_UNIONS(ItemColors) as ItemColors
```

If you are not using grouping, then you do not need an aggregation method. For example:

```
SELECT ProductName as ProductName
```

In this case, you do not need to specify the name, so you could also use:

```
SELECT ProductName
```

## Naming view attributes

When naming attributes in a view:

- Do not change the name of (also referred to as aliasing) attributes from the physical data.

For example, for a "Region" attribute, when you add the attribute to a view, define it as "Region".

```
SELECT ARB(Region) AS Region
```

Do not define the attribute under a different name, such as:

```
SELECT ARB(Region) AS RegionList
```

End users can only refine by attributes that are present in the physical data. If you alias an attribute, then end users cannot use it for refinement.

- Do not give a derived attribute the same name as an attribute from the physical data.

For example, for an attribute that averages the values of the "Sales" attribute, do not define the new attribute as:

```
SELECT AVG(Sales) as Sales
```

Instead, use something like:

```
SELECT AVG(Sales) as AvgSales
```

This is also to prevent end users from trying to refine by an attribute that is not in the physical data.

## Sample view definition

Here is an example of an EQL query for generating a list of products, including price, available colors, and sales numbers, from a single data set consisting of individual transactions.

```
DEFINE Products AS SELECT
ARB(ProductSubcategoryName) AS ProductSubcategoryName,
ARB(ProductCategoryName) AS ProductCategoryName,
ARB(Description) AS Description,
SET_UNIONS(AvailableColors) AS AvailableColors,
AVG("FactSales_SalesAmount") AS AvgSales,
```

```

SUM("FactSales_SalesAmount") AS SalesSum,
AVG(StandardCost) AS AvgStandardCost,
AVG(ListPrice) AS AvgListPrice,
(AvgListPrice - AvgStandardCost) AS Profit
FROM Sales
GROUP BY ProductName

```

## Configuring the view attributes

For each attribute in the view, you can configure the display name and aggregation methods.

### About the view attributes list

On the **Views** page, the attributes list for a view consists of the attributes from the view definition. This includes both the defined attributes and the grouping attributes.

The list includes the following information about each attribute:

Attribute List Column	Description
<b>Display Name</b>	<p>The display name for the attribute.</p> <p>Used to identify the attribute throughout the UI.</p> <p>You can edit and localize the display name at any time.</p> <p>See <a href="#">Editing and localizing an attribute display name and description on page 200</a>.</p>
<b>Attribute Key</b>	<p>The attribute identifier.</p> <p>Read-only. You cannot change the attribute key unless you update the view definition.</p>
<b>Refinable</b>	<p>Whether the attribute can be used for refinement.</p> <p>Only attributes from the physical data can be used for refinement.</p> <p>Read-only.</p>
<b>Multi-Value</b>	<p>Whether the attribute can be assigned multiple values.</p> <p>For a base view, this indicates whether the physical attributes are configured to be multi-value.</p> <p>For views you create, an attribute can be multi-value if you used the set aggregation for that attribute in the view definition.</p> <p>Read-only.</p>

Attribute List Column	Description
<b>Data Type</b>	<p>The data type for the attribute.</p> <p>You cannot change the data type.</p> <p>For base views, for date/time attributes, you can configure the available subsets of date/time units that can be used for filtering and display. See <a href="#">Determining the available subsets of date/time attributes on page 201</a>.</p>
<b>Standard Format</b>	<p>The default display format for the attribute values.</p> <p>For most data types, you can customize the default display format.</p> <p>String attributes can only be customized if they are multi-value.</p> <p>See <a href="#">Configuring the default display format for an attribute on page 203</a>.</p>
<b>Dimension</b>	<p>Whether the attribute can be used as a dimension for aggregating metric values.</p> <p>To allow the attribute to be used as a dimension, check the checkbox.</p> <p>Note that multi-value date/time attributes cannot be used as dimensions.</p>
<b>Available Aggregations</b>	<p>The available aggregation methods for aggregating the values for this attribute into a metric.</p> <p>You can enable or disable available aggregation methods.</p> <p>See <a href="#">Selecting the available and default aggregation methods for an attribute on page 205</a>.</p>
<b>Default Aggregation</b>	<p>The default aggregation method.</p> <p>You can change the default aggregation method.</p> <p>See <a href="#">Selecting the available and default aggregation methods for an attribute on page 205</a>.</p>
<b>Refinement Behavior</b>	<p>Displayed and editable for base views only.</p> <p>The type of refinement allowed for the attribute.</p> <p>Indicates whether users can select more than one value to refine by.</p> <p>See <a href="#">Configuring the Available Refinements behavior for base view attributes on page 210</a>.</p>

Attribute List Column	Description
<b>Include in Available Refinements</b>	<p>Displayed and editable for base views only.</p> <p>Whether to include the attribute on the <b>Available Refinements</b> panel.</p> <p>If this box is not checked, then even if the attribute belongs to a group that is included on the <b>Available Refinements</b> panel, the attribute is not displayed.</p> <p>See <a href="#">Configuring the Available Refinements behavior for base view attributes on page 210</a>.</p>
<b>Description</b>	<p>A short description of the attribute.</p> <p>You can edit and localize the attribute description at any time.</p> <p>See <a href="#">Editing and localizing an attribute display name and description on page 200</a>.</p>

## Editing and localizing an attribute display name and description

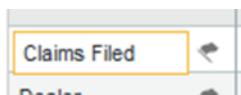
The attribute display name is used on the Big Data Discovery user interface as the default label for the attribute. By default, the attribute display name is the same as the attribute key. The description provides additional information. You can edit both values, and also create localized versions for each of the supported locales.

In the attributes list on the **Views** page, to edit and localize an attribute's display name and description:

1. To edit the display name:

- (a) Click the display name.

The name displays in an editable field.



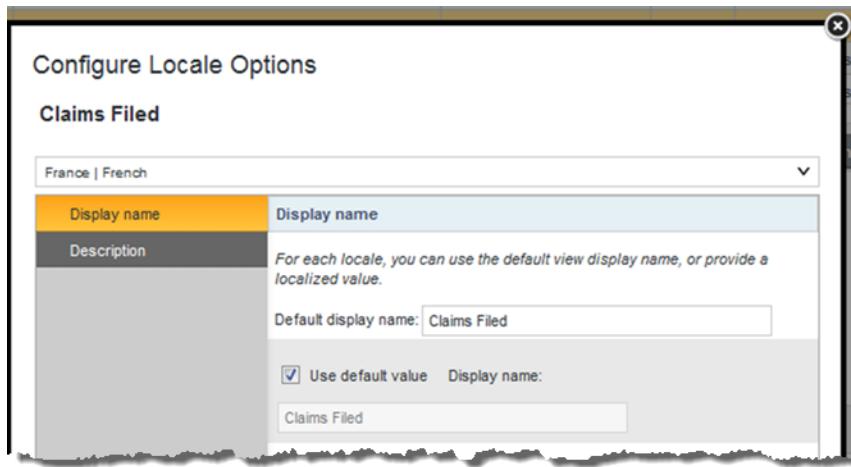
- (b) In the field, type the new name.

2. To localize the display name, in the **Display Name** column, click the localize icon.

On the **Display Name** tab of the **Configure Locale Options** dialog, the **Default display name** field contains the value you entered in the **Display Name** column. If you edit the value here, it also is updated in the attributes list.

For each locale:

- From the locale drop-down list, select the locale for which to configure the display name value.



- To use the default value, leave the **Use default value** checkbox checked.
  - To provide a localized value for the selected locale, uncheck the checkbox. In the **Display name** field, enter the localized value.
  - After entering the localized values, click **Apply**.
- From the **Description** column, to edit the description:
    - Click the description.
    - The description displays in an editable field.
    - In the field, type the new description.
  - To localize the description, in the **Description** column, click the localize icon.

On the **Description** tab of the **Configure Locale Options** dialog, the **Default description** field contains the value you entered in the column. If you edit the value here, it also is updated in the attributes list.

For each locale:

- From the locale drop-down list, select the locale for which to configure the description.
- To use the default description, leave the **Use default value** checkbox checked.
- To provide a localized value for the selected locale, uncheck the checkbox. In the **Description** text area, enter the localized value.
- After entering the localized values, click **Apply**.

- To save the changes to the attribute display name and description, click **Save View**.

## Determining the available subsets of date/time attributes

For a date/time attribute, you can select the available combinations of date "parts" that can be used to display and filter by the attribute.

For example, you may want to allow users to filter by the specific month or combination of month and year.

On a component, users may choose to display an individual date/time element (Year, Month, Day, Hour, Minute, Second), or may be able to select one of the following date/time combinations:

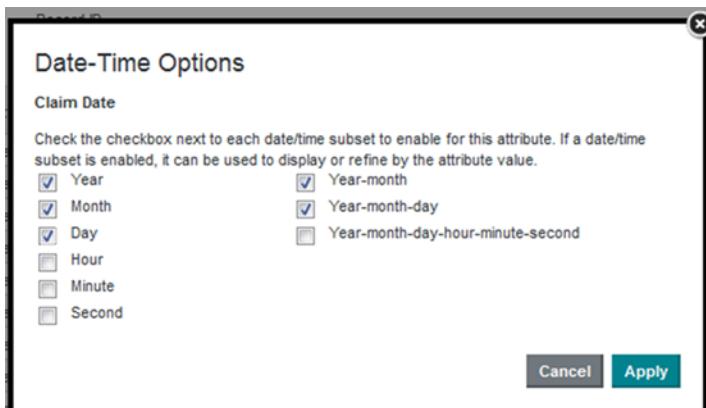
Date/Time Display Option	Sample Value
Year-Month	January 2012
Year-Month-Day	January 1, 2012
Year-Month-Day-Hour-Minute-Second	January 1, 2012 3:15:05 am

To configure the available subsets of date/time units:

1. In the **Data Type** column for a date/time attribute, click the edit icon.

	Data Type	Standard
	String	(String)
	String	(string)
	Date/Time	DateTim
	String	(string)
	Date/Time	DateTim
	String	(string)
	String	(string)

2. On the **Date-Time Options** dialog, check the checkbox next to each option to enable.



3. Click **Apply**.

## Configuring the default display format for an attribute

By default, the display format for attribute values is based on the attribute's data type plus the user's current locale. You can change the default display format used for each attribute. Within specific components, you can then make some updates to the display format.

Note that you only configure the display format for the default locale version of the attribute. You do not configure formatting for the locale-specific attributes. Because the formatting configuration already takes the current locale into account, you do not need to configure the formatting for the locale-specific attributes separately.

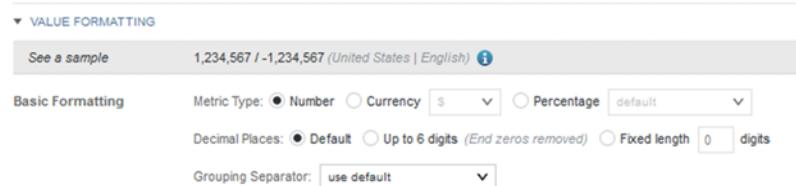
From the attributes list on the **Views** page, to change the default display format for an attribute:

1. Click the edit icon in the **Standard Format** column.

The **Format Option** dialog is displayed.

At the top of the dialog is a sample value that displays the current formatting selection based on your current locale.

2. For numeric attributes:



- (a) Under **Metric Type**, click the radio button to indicate the type of number.

A numeric attribute may be displayed as a regular number, a currency value, or a percentage.

Note that if you set a numeric attribute to be a currency or percentage, that selection cannot be overridden at the component level.

- (b) If the number is a currency value, then from the **Currency** drop-down list, select the type of currency. This selection determines the symbol to display next to each value.
- (c) If the number is a percentage, then from the **Percentage** drop-down list, select whether to display the percentage (%) sign with the value.

Note that if you use percentage as the format, Big Data Discovery automatically multiplies the value by 100. So if the actual value is 0.05, Big Data Discovery displays the value as 5%.

- (d) Under **Decimal Places**, to display all numbers using the same default number of decimal places, leave the **Default** radio button selected.

To display all of the decimal places from the original value, up to 6 places, click the **Up to 6 digits** radio button. Big Data Discovery truncates any decimal places after 6, and removes any trailing zeroes.

To specify the number of decimal places to use for all numbers, click the **Fixed length** radio button, then in the field, type the number of decimal places to display.

- (e) From the **Grouping Separator** drop-down list, select whether to display the grouping separator (used to separate thousands).

- (f) The **Advanced Formatting** section provides additional format options.

Advanced Formatting *Changes below will override localized formatting based on each user's locale.*

Number Format:

Decimal Separator:

Grouping Separator:

For each of the advanced items, you can choose to have the display determined automatically based on the user's locale, or select a specific option.

- (g) From the **Number Format** drop-down list, select how to display positive and negative numbers.
- (h) From the **Decimal Separator** drop-down list, select the character to use for the decimal point.
- (i) From the **Grouping Separator** drop-down list, select the character to use for the grouping (thousands) separator.
3. For a Boolean attribute, you can determine the values to display for each Boolean value.

By default, 1 is displayed as the localized version of "True", and 0 is displayed as the localized version of "False". To set specific (but non-localized) values to display:

See a sample true / false (United States | English) *?*

Basic Formatting

Boolean Display:  Default  Custom Values

True:

False:

- (a) Click the **Custom Values** radio button.
- (b) In the **True** field, type the value to display if the attribute value is 1 (True).
- (c) In the **False** field, type the value to display if the attribute value is 0 (False).
4. For a geocode, to set a specific number of decimal places to display for the latitude and longitude values:

▼ VALUE FORMATTING

See a sample 44.756530 -91.473097 (United States | English) *?*

Basic Formatting

Decimal Places:  Default  Fixed length  digits

- (a) Click the **Fixed length** radio button.
- (b) In the field, type the number of decimal places.

5. For a date/time attribute:

See a sample      1/18/99 (United States | English) ?

Basic Formatting      Date Display:  ▼

Time Display:  ▼

- (a) From the **Date Display** drop-down list, select the format to use for the date.

The selected format is adjusted for locale. For example, for some locales the month displays first, and for others the day displays first.

- (b) From the **Time Display** drop-down list, select the format to use for the time.

6. For a time attribute, from the **Time Display** drop-down list, select the format to use for the time.

See a sample      3:30 PM (United States | English) ?

Basic Formatting      Time Display:  ▼

7. For all value types, the **Multi-Value Formatting** section indicates how to display multiple values.

▼ MULTI-VALUE FORMATTING

Multi-Value Formatting      Only apply when a record has a multiple values or as the result of a set aggregation.  
Changes to any of the settings below will override localized formatting based on each user locale.

Multi-value Separator:  ▼

The setting is applied whenever multiple values are displayed for an attribute, including when an attribute is by definition multi-value, and when the set aggregation is applied to an attribute.

From the **Multi-value Separator** drop-down list, select the character to use to separate the values.

8. To confirm the format configuration, click **Apply**.
9. To save the changes to the view settings, click **Save View**.

## Selecting the available and default aggregation methods for an attribute

Aggregation methods are used to aggregate attribute values into metrics displayed on components such as **Chart**, **Pivot Table**, and **Results Table**.

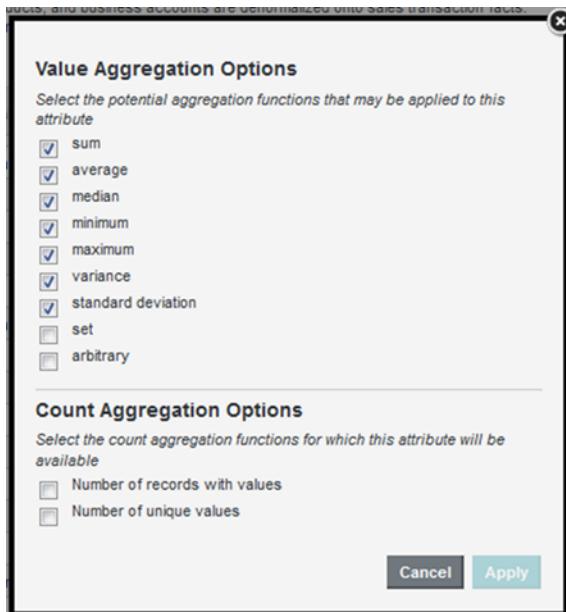
Based on its data type, each attribute can support specific aggregation methods. You can then configure attributes to only support a subset of the applicable aggregation methods. For details on the available aggregation methods, see [Aggregation methods and the data types that can use them on page 207](#).

Each attribute also is assigned a default aggregation method. You can change the default used for each attribute.

From the attributes list on the **Views** page, to configure the available and default aggregation methods for an attribute:

- To configure the available aggregation methods for an attribute:
  - In the **Available Aggregations** column, click the edit icon.

- (b) On the aggregation methods dialog, to enable an available aggregation method, check the checkbox.



- (c) To disable an aggregation method, uncheck the checkbox.

- (d) To confirm the settings, click **Apply**.

2. From the **Default Aggregation** drop-down list, select the default aggregation to use.

	Default Aggregation	
	set	
	sum	
	sum	
	average	
	median	
	minimum	
	maximum	
	variance	
	standard deviation	
	set	

3. To save the changes to the view configuration, click **Save View**.

## Aggregation methods and the data types that can use them

Aggregation methods are types of calculations used to group attribute values into a metric for each dimension value. For example, for each country (each value of the Country dimension), you might want to retrieve the total value of transactions (the sum of the Sales Amount attribute).

On the component edit view, you select the aggregation method from a drop-down list. When using these aggregation methods in definitions of views and predefined metrics, you use the EQL syntax.

The available aggregation methods are:

Aggregation Method	Description	EQL Syntax
sum	<p>Calculates the total value for the metric.</p> <p>You can use this aggregation method for numbers.</p> <p>You cannot use this aggregation method for multi-value attributes.</p>	<code>sum(attributeName)</code>
average	<p>Calculates the average value for the metric.</p> <p>You can use this aggregation method for numbers, dates, and times.</p> <p>You cannot use this aggregation method for multi-value attributes.</p>	<code>avg(attributeName)</code>
median	<p>Calculates the median value for the metric.</p> <p>You can use this aggregation method for numbers, dates, and times.</p> <p>You cannot use this aggregation method for multi-value attributes.</p>	<code>median(attributeName)</code>
minimum	<p>Selects the minimum value for the metric.</p> <p>You can use this aggregation method for numbers, dates, and times.</p> <p>You cannot use this aggregation method for multi-value attributes.</p>	<code>min(attributeName)</code>
maximum	<p>Selects the maximum value for the metric.</p> <p>You can use this aggregation method for numbers, dates, and times.</p> <p>You cannot use this aggregation method for multi-value attributes.</p>	<code>max(attributeName)</code>

Aggregation Method	Description	EQL Syntax
variance	<p>Calculates the variance (square of the standard deviation) for the metric values.</p> <p>You can use this aggregation method for numbers, dates and times.</p> <p>You cannot use this aggregation method for multi-value attributes.</p>	variance( <i>attributeName</i> )
standard deviation	<p>Calculates the standard deviation for the metric values.</p> <p>You can use this aggregation method for numbers, dates and times.</p> <p>You cannot use this aggregation method for multi-value attributes.</p>	stddev( <i>attributeName</i> )
set	<p>Instead of performing a calculation, creates a list of all of the unique values.</p> <p>You can use this aggregation method for any attribute.</p>	set( <i>attributeName</i> ) For multi-value attributes, use set_unions( <i>attributeName</i> ).
arb	<p>Instead of performing a calculation, arbitrarily selects one of the values.</p> <p>You can use this aggregation method for any attribute.</p> <p>The arb aggregation method is typically not provided as an option for aggregating metric values on components.</p>	arb( <i>attributeName</i> )

When configuring components, the following count aggregation options are applied using separate system metrics.

For the other aggregation methods, such as sum and average, you select an attribute, and then apply the aggregation method. For the count aggregation methods, you first select the system metric, then select the attribute to use.

When using these aggregation methods in definitions of views and predefined metrics, you use the EQL syntax.

Aggregation Method	Description	EQL Syntax
Number of records with values	<p>Calculates the number of records with a value for the selected attribute.</p> <p>You can use this aggregation method for all data types.</p> <p>You cannot use this aggregation method for multi-value attributes.</p> <p>You apply this aggregation by adding the <b>Number of records with values</b> system metric to the component.</p>	count( <i>attributeName</i> )
Number of unique values	<p>Calculates the number of unique values for the metric.</p> <p>You can use this aggregation method for all data types.</p> <p>You cannot use this aggregation method for multi-value attributes.</p> <p>You apply this aggregation by adding the <b>Number of unique values</b> system metric to the component.</p>	countdistinct( <i>attributeName</i> )

To show how each of the aggregation methods work, we'll use the following data:

Country	Amount of Sale	Shipping Company
US	125.00	UPS
US	50.00	UPS
US	150.00	FedEx

Based on these values, the aggregation for Amount of Sale and Shipping Company for the US would be:

Aggregation Method	Amount of Sale (US)	Shipping Company (US)
sum	325.00	Cannot be aggregated
avg	108.33	Cannot be aggregated
median	125.00	Cannot be aggregated
min	50.00	Cannot be aggregated

Aggregation Method	Amount of Sale (US)	Shipping Company (US)
max	150.00	Cannot be aggregated
variance	2708.33	Cannot be aggregated
standard deviation	52.04	Cannot be aggregated
set	125.00, 50.00, 150.00	UPS, FedEx
arb	Could be either 125.00, 50.00, or 150.00	Could be either UPS or FedEx
Number of records with values	3	3
Number of unique values	3	2

## Configuring the Available Refinements behavior for base view attributes

For each attribute in a base view, the **Views** page attributes list allows you to configure settings for how to handle selection and sorting of attribute values on the **Available Refinements** panel.

In the attributes list for a base view, to configure the value selection and sorting for each attribute:

1. From the **Refinement Behavior** drop-down list, select whether users can select multiple values, and if so, whether the records must contain all of the selected values.

Refinement Behavior	Include in Available Refinements
Single	<input checked="" type="checkbox"/>
Single	<input checked="" type="checkbox"/>
Multi-AND	<input checked="" type="checkbox"/>
Multi-OR	<input checked="" type="checkbox"/>
Single	<input checked="" type="checkbox"/>

The available options are:

Option	Description
Single	Indicates that end users can only refine by one value at a time.

Option	Description
<b>Multi-OR</b>	<p>Indicates that end users can refine by more than one value at a time.</p> <p>For multi-or, a record matches if it has at least one of the selected values.</p> <p>So if an end user selects the values Red, Green, and Blue, then matching records only need to have one of those values (Red or Green or Blue).</p> <p>Note that date/time attributes are always multi-or. You cannot change the refinement behavior for a date/time attribute.</p>
<b>Multi-AND</b>	<p>Indicates that end users can refine by more than one value at a time.</p> <p>For multi-and, a record matches only if it has all of the selected attribute values. Multi-and should only be used with multi-value attributes.</p> <p>So if an end user selects the values Red, Green, and Blue, then matching records must have all of those values (Red and Green and Blue).</p>

2. In the **Include in Available Refinements** column, to include the attribute on the **Available Refinements** panel if it belongs to a group that is displayed, check the checkbox.

## Defining predefined metrics for a view

Custom views can also have predefined metrics, which are values calculated from the view attributes.

### About predefined metrics

Predefined metrics are values calculated from view attributes. For example, if the attributes for a view include product price and cost information, you could calculate predefined metrics such as the ratio of retail price to cost.

Predefined metrics are optional. By default, views do not have predefined metrics. They are always added manually.

Predefined metrics should only be used for calculated values that are only meaningful within the context of the view, and that stand on their own as values. When configuring a component, you can assign simple aggregations such as sum or average to any attribute, so you don't need to create those as predefined metrics. Predefined metrics are more appropriate for more complex values such as ratios.

On the **Views** page, the **Metrics** section, which lists the predefined metrics, is displayed below the view attribute list. By default, the section is collapsed. To expand the section, click the section heading.

Metrics						
Display Name	Attribute Key	Multi-Value	Data Type	Standard Format	Definition	Description
Sales Share	SalesShare	—	Number (double)	Number	sum(Sales_SalesAmount)/G...	Ratio of sales to...

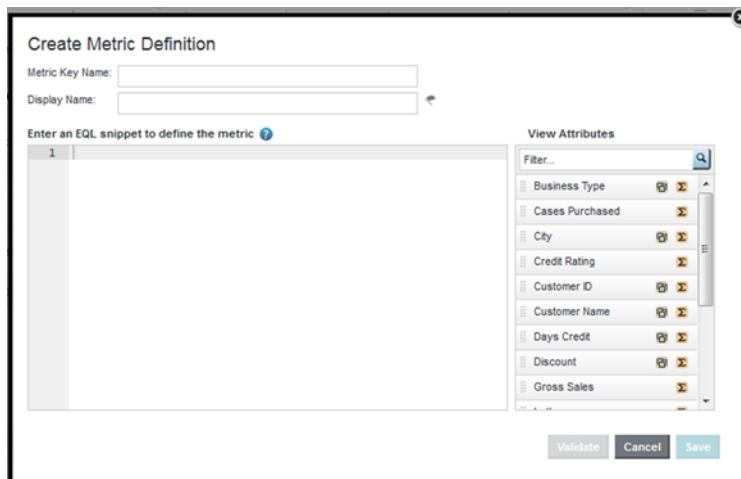
## Adding a predefined metric to a view

Adding a predefined metric to a view consists of specifying the metric key and EQL definition.

On the **Views** page, to add a predefined metric to a view:

1. From the **Metrics** list, click **+ Metric**.

The **Create Metric Definition** dialog is displayed.



2. On the **Create Metric Definition** dialog, in the **Metric Key Name** field, type the key name for the metric.

The key name must be NCName-compliant.

Once you save the new metric, you cannot change the key name.

3. In the **Display name** field, type the display name for the metric.

This is the name used to select the metric to add to a component. You can edit and localize the display name at any time.

4. In the text area, type the EQL expression for calculating the predefined metric value.

The expression is an EQL snippet that is inserted into a **SELECT** statement when a query using the predefined metric is issued. The snippet can only refer to attributes from the current view. For example:

```
count(1) WHERE (SalesAmountSum > 3000000)
```

To the right of the EQL text area is the list of attributes in the current view for you to use as a reference. You can show or hide the attribute list.

For information on aggregation methods, including the EQL syntax for including them in predefined metrics, see [Aggregation methods and the data types that can use them on page 207](#). For detailed information on EQL syntax, see the *EQL Reference*.

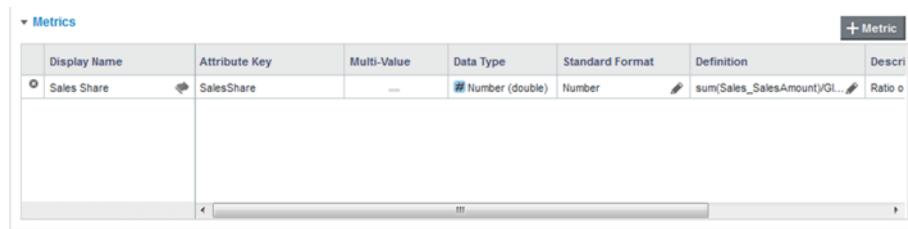
5. To validate the expression, click **Validate**.

If the expression is invalid, Big Data Discovery displays error messages to help you troubleshoot.

6. To confirm adding the metric, click **Save**.

## Configuring a predefined metric for a view

From the predefined metrics list on the **Views** page, you can configure the metric name, description, definition, and display. The display name and definition can be localized for each of the available locales.



Metrics						
	Display Name	Attribute Key	Multi-Value	Data Type	Standard Format	Definition
1	Sales Share	SalesShare	...	Number (double)	Number	sum(Sales_SalesAmount)/G1...

To configure a predefined metric:

1. In the **Display name** field, type a display name for the predefined metric.

Display names do not need to be NCName-compliant.

2. To localize the display name, in the **Display name** column, click the localize icon.

On the **Configure Locale Options** dialog, the **Default display name** field contains the value you entered in the column. If you edit the value here, it also is updated in the metrics list. For each locale:

(a) From the locale drop-down list, select the locale for which to configure the display name value.

(b) To use the default value, check the **Use default value** checkbox.

(c) To provide a localized value for the selected locale, uncheck the checkbox. In the **Display name** field, enter the localized value.

(d) After entering the localized values, click **Apply**.

3. To configure the display format for a predefined metric, click the edit icon in the **Standard Format** column.

The display format configuration is the same as that for attributes. See [Configuring the default display format for an attribute on page 203](#).

4. To change the definition of a predefined metric, click the edit icon in the **Definition** column.

5. In the **Description** field, type a brief description of the predefined metric.

6. To localize the description, in the **Description** column, click the localize icon.

On the **Configure Locale Options** dialog, the **Default description** field contains the value you entered in the column. If you edit the value here, it also is updated in the metrics list. For each locale:

- (a) From the locale drop-down list, select the locale for which to configure the description.
- (b) To use the default description, check the **Use default value** checkbox.
- (c) To provide a localized value for the selected locale, uncheck the checkbox. In the **Description** field, enter the localized value.
- (d) After entering the localized values, click **Apply**.

7. To save any changes to the predefined metrics, click **Save View**.

## Deleting a predefined metric

You can delete predefined metrics from a view. Before deleting a predefined metric, try to make sure that it is not being used by a component.

From the **Views** page, to delete a predefined metric from the current view, click the delete icon for that metric.

To save the change to the view configuration, click **Save View**.

## Previewing the content of a view

From the **Views** page, you can display a preview list of the records in a view.

To preview the contents of a view:

1. Select **Project Settings>Data Views**.
2. In the views list, click the view.
3. Click **Preview**.

The **Preview** dialog is displayed, containing a list of records for the view.

The list only displays the view attributes. Because the preview list is not grouped by a specific dimension, it cannot display the predefined metrics.

4. To close the **Preview** dialog, click **Close**.

## Determining whether a custom view is published

For custom views, the **Publish** checkbox controls whether the view is published. Only published views are available to use in a project.

Base views are always published.

Custom views can only be published if the view definition is valid.

## Deleting a view

From the **Views** page, you can delete custom views. You cannot delete base views or linked views. If you delete a view, any component that uses that view can no longer display correctly. Also, any other view whose definition includes references to this view becomes invalid.

Note that Big Data Discovery cannot automatically check whether a view is being used in the definition of another view.

To delete a view, in the **Actions** column of the view list, click the delete icon for that view.



## Chapter 28

# Managing Attribute Groups for a Project

From the **Attribute Groups** page, you can create, edit, and delete attribute groups for each view in the project data.

[About attribute groups and the Attribute Groups page](#)

[About groups for linked views](#)

[Displaying the groups for a selected view](#)

[Displaying attribute information for a group](#)

[Creating an attribute group](#)

[Deleting an attribute group](#)

[Configuring an attribute group](#)

[Saving the attribute group configuration and reloading the attribute group cache](#)

## About attribute groups and the Attribute Groups page

Attribute groups are used to group attributes within a view. For example, for a view containing sales information, one group might contain attribute related to the payment (credit card type, cost, billing address) and another group might contain attributes related to the shipping (shipping address, shipping company, shipping cost).

For each group, you select the specific attributes in the group.

### ATTRIBUTE GROUPS

Display Name	Attribute Key	Info	Data Type	Dimension
Sales Price	P_Price	1	#	—
Price Range	PriceRange	1	“	✓
Date Reviewed	P_DateReviewed	1	“	✓
Base Score	P_Score	1	#	—
Score Range	ReviewScore	1	“	✓
Designation	Designation	1	“	✓

From the **Attribute Groups** page, you can create and edit the attribute groups for the current project data. To display the **Attribute Groups** page:

1. Select **Project Settings**.
2. Click **Attribute Groups**.

## About groups for linked views

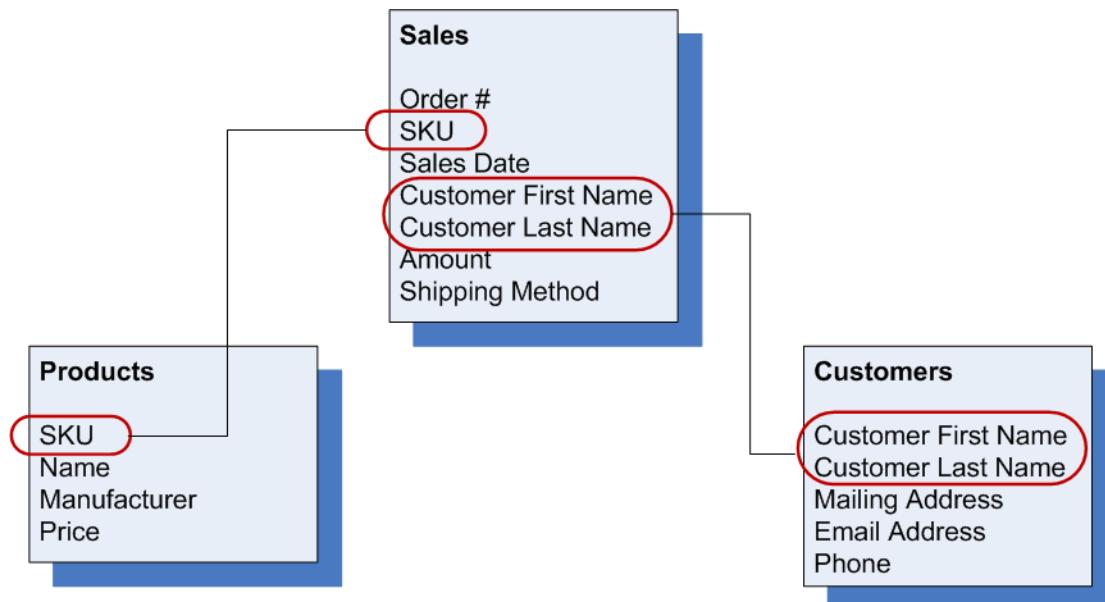
Big Data Discovery automatically creates a linked view for each linked data set. For each linked view, Big Data Discovery also creates a default set of attribute groups.

The default attribute groups for a linked view consist of:

- The attribute groups from the source data set
- For each data set that the source data set is linked to, an attribute group containing all of the attributes in that data set, except for the link key attributes. The group is named for the data set.

For example, as shown in the following diagram:

- The Sales and Products data sets are linked by the SKU attribute
- The Sales and Customers data sets are linked using the Customer First Name and Customer Last Name attributes.



For these links, Big Data Discovery creates the following views:

- Sales - linked
- Products - linked
- Customers - linked

If the base views for the data sets contain the following groups:

Data Set	Base View Groups
Sales	Order Creation Customer Information Shipping Details
Products	Product Summary Product Pricing
Customers	Customer Name Customer Contact Information

Then for the linked views, Big Data Discovery creates the following default attribute groups:

Linked View	Default Attribute Groups
Sales - linked	Order Creation Customer Information Shipping Details Products (all Product attributes, except for SKU) Customers (all Customer attributes, except for Customer First Name and Customer Last Name)
Products - linked	Product Summary Product Pricing Sales (all Sales attributes, except for SKU)
Customers - linked	Customer Name Customer Contact Information Sales (all Sales attributes, except for Customer First Name and Customer Last Name)

These are default groups only. You can create, edit, and remove groups for linked views the same way you would for base or custom views.

## Displaying the groups for a selected view

When the **Attribute Groups** page is first displayed, the groups for the first base view are displayed by default. You can then change the selected view.

To display the groups for a selected view:

to select the view for which to display the attribute groups:

1. Select **Project Settings**.
2. Select the **Attribute Groups** page.
3. From the **View** drop-down list, select the view.
4. If there are unsaved changes in the current view, you are prompted to either save the changes, discard the changes, or cancel and return to the current view.

When you select a view, the **Attribute Groups** section is updated to display the groups for the current view.

The screenshot shows the 'Attribute Groups' page with the following structure:

- Permanent Groups:** Contains 'All' and 'Other' groups, each with a pencil icon for editing.
- Created Groups:** Contains 'Identification', 'Source', 'Characteristics', and 'Price and Score' groups. The 'Price and Score' group is highlighted with a yellow background, indicating it is the selected view.
- A 'Group' button is located at the top right of the list.

The **Permanent Groups** list contains the **All** group and the **Other** group. These groups are automatically part of every view.

- The **All** group always contains all of the attributes in the view.
- The **Other** group always contains the attributes that are not members of any of the created groups.

The **Created Groups** list contains the groups you create for this view, as well as groups created automatically for linked data set views.

For the **Created Groups** list, you can set the default display order.

## Displaying attribute information for a group

For each group on the **Attribute Groups** page, you can display a list of attributes and attribute metadata.

*Displaying the list of attributes for a group*

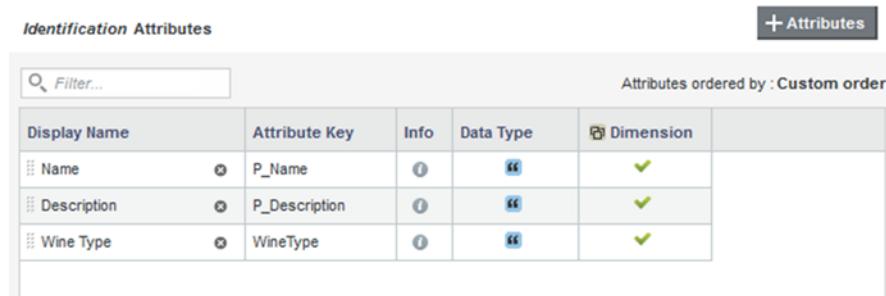
*Displaying the full list of metadata values for an attribute*

*Selecting the metadata columns to display for each attribute*

## Displaying the list of attributes for a group

From the **Attribute Groups** list on the **Attribute Groups** page, to display the current contents of a group, click the group name. The attribute list is updated to display the current membership of the group.

The attribute list consists of attribute metadata for each attribute. By default, the list includes the attribute name, attribute key, data type, and whether the attribute is a dimension.



Identification Attributes					
Attributes ordered by : Custom order					
Display Name	Attribute Key	Info	Data Type	Dimension	
Name	P_Name	ⓘ	“	✓	
Description	P_Description	ⓘ	“	✓	
Wine Type	WineType	ⓘ	“	✓	

## Displaying the full list of metadata values for an attribute

From the **Attribute Groups** page, to display the full list of metadata values for an attribute, hover the mouse over the information icon in the **Info** column.

The metadata includes the list of groups that the attribute belongs to.

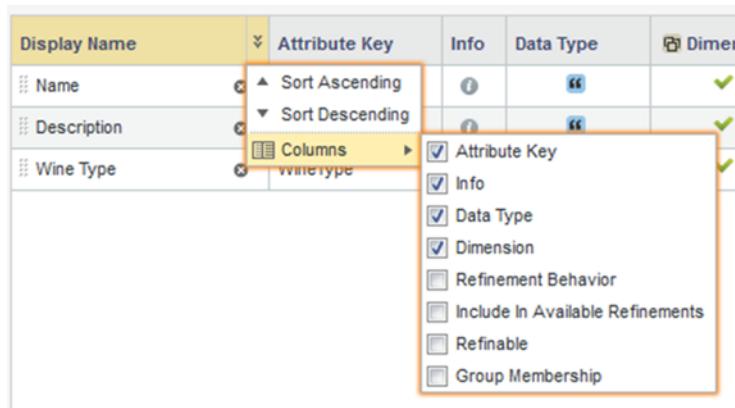
## Selecting the metadata columns to display for each attribute

By default, the attribute list on the **Attribute Groups** page displays basic information for each attribute. You can add columns to and remove columns from the attribute list.

To change the displayed columns:

1. Hover the mouse over the right edge of any column.  
A drop-down arrow is displayed.
2. Click the drop-down arrow.
3. In the drop-down menu, click **Columns**.  
The full list of available columns is displayed.

Columns currently in the list are checked.



Display Name	Attribute Key	Info	Data Type	Dimension
Name	<input type="checkbox"/> Sort Ascending <input type="checkbox"/> Sort Descending	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Description	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Wine Type	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

4. To add a column, check the checkbox.

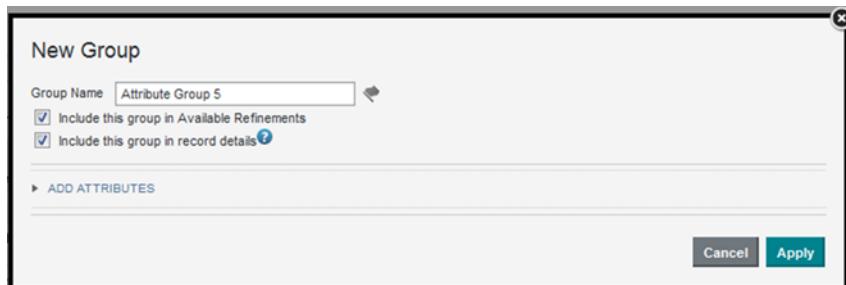
To remove a displayed column, uncheck the checkbox.

## Creating an attribute group

When you create a group, you give the group a name, select default options for how the group is used, and then select the attributes that belong to the group.

From the **Attribute Groups** page, to create a new attribute group:

1. Click **+Group**.
2. On the **New Group** dialog, in the **Group name** field, enter the name of the new group.



3. For a group in a base view, to have this group enabled on the **Available Refinements** panel, check the **Include this group in Available Refinements** checkbox.

Note that because the **Available Refinements** panel only uses base views, this checkbox is not displayed for groups created in other views.

Also, even if you enable the group, the attributes in the group are only displayed if they are configured on the **Views** page to display on the **Available Refinements** panel. See [Configuring the Available Refinements behavior for base view attributes on page 210](#).

4. To have this group enabled when displaying record details information for this view, including on the **Record Details** and **Compare** dialogs, and in the default configuration for a **Results Table** component, check the **Include this group in record details** checkbox.

5. Configure the group membership.

See [Configuring the group membership on page 224](#).

6. After completing the configuration, click **Apply**.

The new group is added to the group list. You can then drag the group to set its default display order.

## Deleting an attribute group

From the **Attribute Groups** page, you can delete attribute groups.

You cannot delete the default **Other** or **All** attribute groups.

Deleting an attribute group does not delete its attributes from the view or from other attribute groups.

To delete an attribute group, in the **Created Groups** list, click the delete icon for the attribute group. The attribute group is removed from the **Created Groups** list.

When you save the attribute group changes for the current view, then the attribute group is no longer displayed on components.

If an attribute does not belong to any of your other groups, it is restored to the default **Other** group.

## Configuring an attribute group

From the **Attribute Groups** page, for each group, you can configure the display name, the group membership, and the attribute display order.

[Changing the group name and use](#)

[Configuring the group membership](#)

[Configuring the sort order for a group](#)

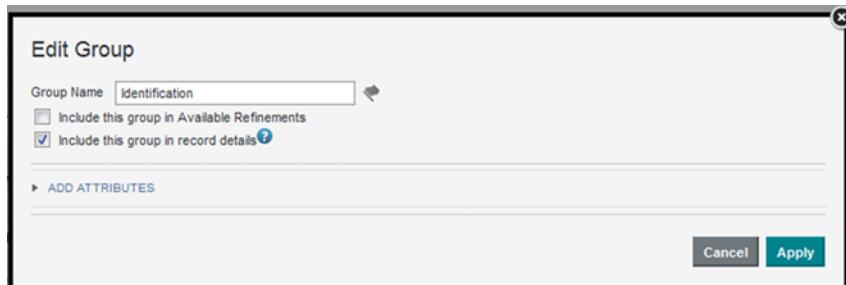
## Changing the group name and use

The group name is used as the display name for the group on the Big Data Discovery user interface. You can edit and localize the name. You can also change whether a group is used in **Available Refinements** and for record detail display.

To change the name and default use of an existing group:

1. In the groups list, click the edit icon for the group.

2. On the **Edit Group** dialog, in the **Group Name** field, type the new display name for the group.

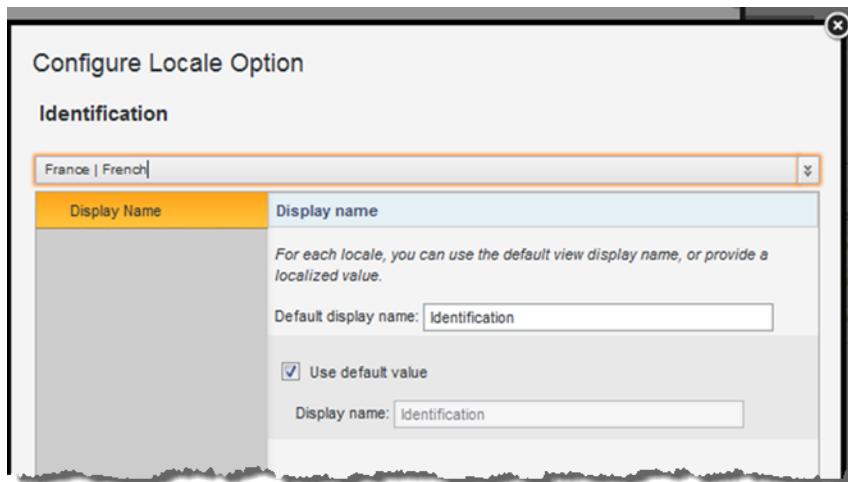


3. To localize the display name, click the localize icon next to the **Group Name** field.

On the localization dialog, the **Default display name** field contains the value you entered in the **Group Name** field. If you edit the value here, it also is updated on the **Edit Group** dialog and in the group list.

For each locale:

- (a) From the locale drop-down list, select the locale for which to configure the display name value.



- (b) To use the default value, leave the **Use default value** checkbox checked.  
 (c) To provide a localized value for the selected locale, uncheck the checkbox. In the **Display name** field, enter the localized value.  
 (d) After entering the localized values, click **Apply**.

4. For a group in a base view, to have this group enabled on the **Available Refinements** panel, check the **Include this group in Available Refinements** checkbox.

Note that because the **Available Refinements** component only uses base views, this checkbox is not displayed for groups in other views.

Also, even if you enable the group, the attributes in the group are only displayed if they are configured on the **Views** page to display on the **Available Refinements** panel. See [Configuring the Available Refinements behavior for base view attributes on page 210](#).

5. To have this group enabled when displaying record details information for this view, including on the **Record Details** and **Compare** dialogs, and in the default configuration for a **Results Table** component, check the **Include this group in record details** checkbox.
6. To save the changes, click **Apply**.

## Configuring the group membership

For each group, you select the specific attributes that belong to the group.

If an attribute has locale-specific versions, you must add them manually to the group membership.

From the **Attribute Groups** page, to add and remove attributes for a group:

1. To add attributes:
  - (a) Display the **New Group** or **Edit Group** dialog.

The **New Group** dialog is displayed when you create a new group.

To display the **Edit Group** dialog for an existing group, either:

- From the groups list, click the edit icon.
- From the attributes list for the group, click the **+Attributes** button.

The **Add attributes** section contains the list of attributes in the view that do not belong to the group.

ADD ATTRIBUTES				
Filter...		Show all		
	Display Name	Attribute Key	Info	Group Membe...
<b>+ Add</b>	Base Score	P_Score	<small>Info</small>	Price and Score
<b>+ Add</b>	Body	Body	<small>Info</small>	Characteristics
<b>+ Add</b>	Date Reviewed	P_DateReviewed	<small>Info</small>	Price and Score
<b>+ Add</b>	Designation	Designation	<small>Info</small>	Price and Score
<b>+ Add</b>	Drinkability	Drinkability	<small>Info</small>	Characteristics
<b>+ Add</b>	Flavors	Flavors	<small>Info</small>	Characteristics
<b>+ Add</b>	Price Range	PriceRange	<small>Info</small>	Price and Score

Selected Attributes (0)  
No attributes selected.

- (b) To filter the list, either:
  - Use the search field above the attribute list to search based on the attribute display name.
  - Use the data type drop-down list to only display attributes for a specific data type.
- (c) In the attribute list, click the **+Add** button next to each attribute you want to add to the group.

The attributes are added to the **Selected Attributes** list. The **+Add** button for the selected attributes is disabled.

To remove a selected attribute from the list of attributes to add, click the delete icon for that attribute. The **+Add** button for that attribute is then re-enabled.

- (d) To add the selected attributes, click **Apply**.
2. To remove an attribute from a group, in the attribute list, click the delete icon for that attribute.

## Configuring the sort order for a group

For each group, you can either sort the list using a metadata value, or set the sort order manually. The sort order on the **Attribute Groups** page determines the display order of attributes on components.

By default, the group is sorted by attribute display name in alphabetical order. For attributes that do not have a display name, the attribute key is used.

The **Attributes ordered by** field at the top of the attribute list shows the current sort order, including the metadata used and whether the sort is in ascending or descending order.

When you click a column heading, the sort order is updated to use that value for the sort. Clicking the same column heading twice switches the sort between ascending and descending order.

You also can move an attribute to a specific location in the list. When you do that, the sort order is updated to be **Custom order**.

Identification Attributes					
Attributes ordered by : Custom order					
Display Name	Attribute Key	Info	Data Type	Dimension	
Name	AttributeKey	Info	Data Type	Dimension	

If you then click a column heading, the sort order changes to use that metadata value.

## Saving the attribute group configuration and reloading the attribute group cache

Changes to the group configuration do not become permanent until you save them. To immediately reflect the saved group configuration changes on the project components, reload the attribute group cache.

When you have unsaved changes, you can use the **Revert to Last Save** button to undo those changes.

To save the current group configuration, click **Save**.

When Big Data Discovery first loads the project data, it stores the attributes and groups in the cache. To reload the cache, so that the project components immediately reflect any updates to the group configuration, click **Reload Attribute Cache**.



## Chapter 29

---

# Exporting Project Data

You can export data from project views and **Results Tables** to either HDFS or your computer.

[About exporting project data](#)

[Exporting project data](#)

## About exporting project data

You can export data from project views and **Results Tables**. This creates a file based on your data that you can either store in HDFS or download to your computer.

The exported data is based on the current refinement state of your project, so it will not include attributes that have been filtered out or hidden. You will be able to preview a small sample of the exported data, as well as a list of attributes that will not be included, before you export.

When you export project data, you must specify whether you want the file exported to HDFS or your computer. If you export to HDFS, you must specify a directory to save the file to. If you export to your computer, the file is automatically downloaded once it has been generated.

You must also select the type of file your data is exported to. This can be one of the following:

- Delimited
- Avro
- A Hive table. This option is only available when you export to HDFS.



**Note:** The exported data uses attribute display names by default. However, if you are exporting to an Avro file, any attribute display names that contain invalid Avro characters will be replaced with their attribute keys.

## Transposing attributes

If the data set you're exporting contains multi-value attributes, you can have some or all of them transposed. Transposed attributes are replaced with sets of Boolean attributes representing their possible values.



**Note:** You cannot transpose attributes when exporting to an Avro file.

For example, consider a multi-value attribute named Color that has the unique values "Red", "Blue", and "Green". If the Color attribute were transposed, it would be replaced in the exported data by the Boolean attributes Red, Blue, and Green. A record that originally had its Color attribute set to "Red, Blue" would now contain the following:

- A Red attribute set to "True"
- A Green attribute set to "False"

- A Blue attribute set to "True"

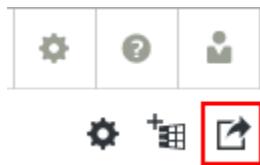
Because multi-value attributes can have numerous possible values, you can specify a limit for the number of new columns added for each transposed attribute. Columns will only be added for the top values until the limit is reached.

## Exporting project data

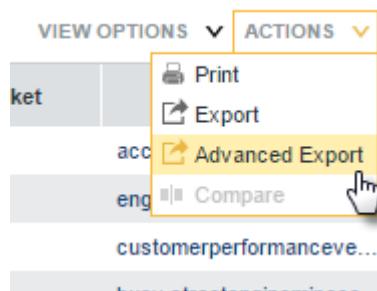
You can export data from a project view or a **Results Table** to either HDFS or your computer.

To export project data:

1. If you are exporting from a project view, click the **Export** icon in the top right corner of the page:



If you are exporting from a **Results Table**, select **Actions > Advanced Export**.



The **Export Data** dialog opens.

2. If you are exporting from a project view, select the data set to export from the **Data Selection** drop down list.

The **Preview** displays the first three records of the selected data set.

3. In the **Filename** field, enter a name for the file the data will be exported to.

By default, this is the name of the selected data set followed by the current date; for example, `Vehicle_Warranty_Claims_03022015`.

4. From the **Export to** drop down list, select the location you want to export the file to:

- **HDFS**
- **My Computer**

5. If you are exporting to HDFS, do the following:

- (a) In the **File Path** field, enter the path to the HDFS directory you want to export to.

By default, this path begins with `/user/bdd/`.

- (b) If you want to create a new Hive table based on the exported data, select the **Create Hive table** check box and enter a name for the table in the **Table name** field.

By default, this is the name of the data set followed by the current date; for example, `Vehicle_Wattanty_Claims_030215`.

Optionally, enter a description for the table in the **Add description** field.

6. From the **Type** drop down list, select the type of file to export to:

- **Delimited**
- **Avro**

Note that if you select **Avro**, you will not be able to transpose any of your data's multi-value attributes.

7. If you are exporting to a delimited file, select delimiters for fields and multi-value attributes.

If you select **Other...**, you must also provide a new delimiter. This must be a single unicode character with a value of 127 or less.

8. If you want to transpose your data's multi-value attributes:

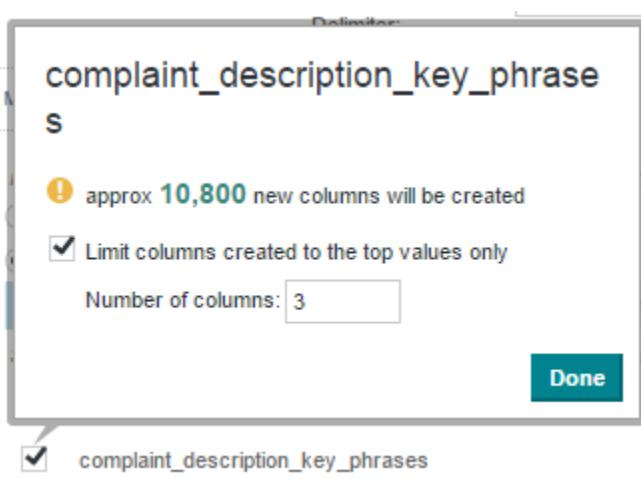


**Note:** This option is *not* available for Avro files.

- (a) Click **Multi-value attribute handling** and select the **Transpose unique values** radio button.
- (b) Select the attribute(s) you want to transpose.

You can select all attributes by checking the **Apply to all multi-value attributes** check box. However, if you want to limit the number of new columns created for each transposed attribute, you must select them one at a time.

- (c) In each selected attribute's pop up, optionally check the **Limit columns created to the top values only** check box and enter a column limit.



This prevents your data from becoming cluttered with columns from attributes that have large numbers of possible values.

9. Click **Export**.

# **Part IX**

## **Working with Components**



## Chapter 30

---

# Containers

Container components allow you to group components on a page.

[Component Container](#)

[Tabbed Component Container](#)

## Component Container

The **Component Container** groups components into a single component.

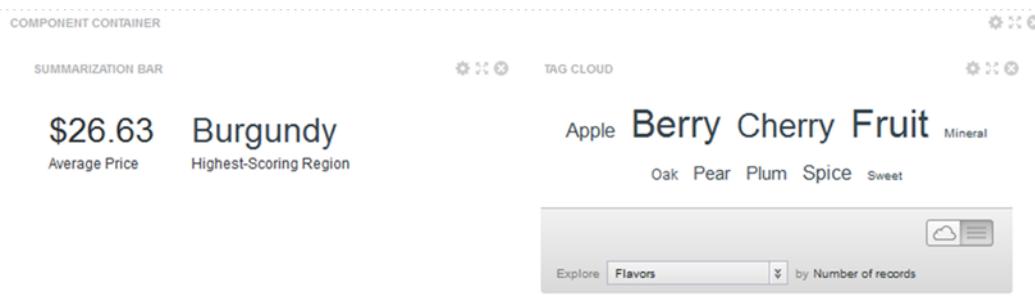
[About the Component Container](#)

[Configuring a Component Container](#)

### About the Component Container

The **Component Container** allows you to organize components by grouping them in a single container.

The components can then be viewed and moved as a set.



Do not add a **Component Container** to a **Tabbed Component Container** or to another **Component Container**.

### Configuring a Component Container

For a **Component Container**, you can configure the layout to use.

On the component edit view, under **Page Layout**, click the radio button next to the layout you want to use for the container.

After configuring the **Component Container**, you can drag other components onto the component and configure them as usual.

## Tabbed Component Container

The **Tabbed Component Container** groups components into multiple tabs.

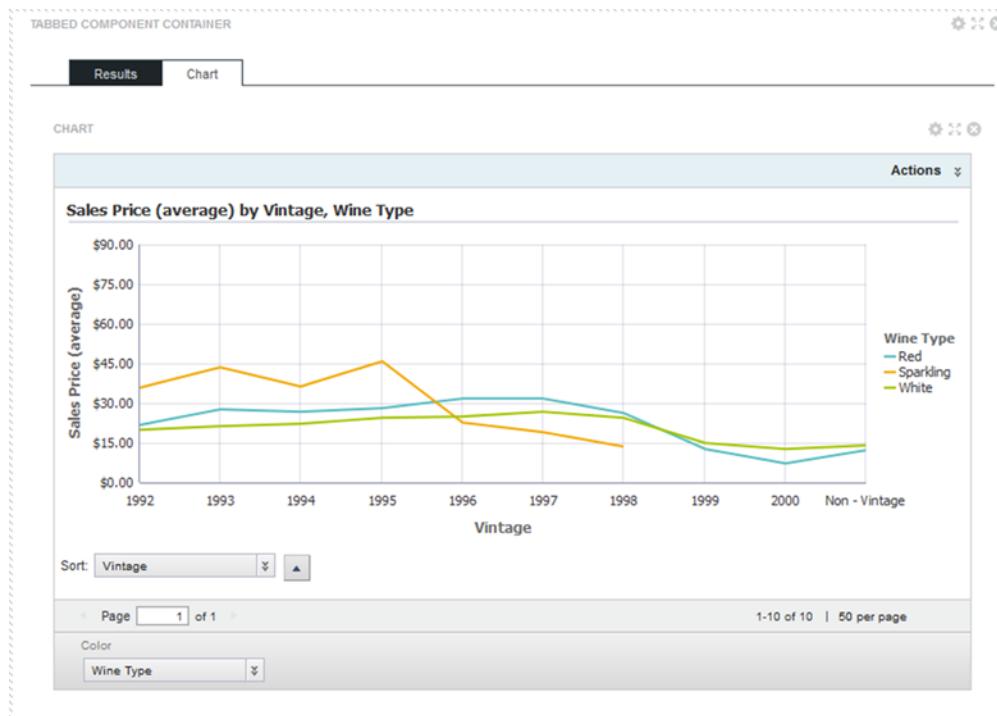
[About the Tabbed Component Container](#)

[Configuring a Tabbed Component Container](#)

### About the Tabbed Component Container

The **Tabbed Component Container** creates a tabbed interface.

Each tab can contain a different set of components.



Do not add a **Component Container** to a **Tabbed Component Container**.

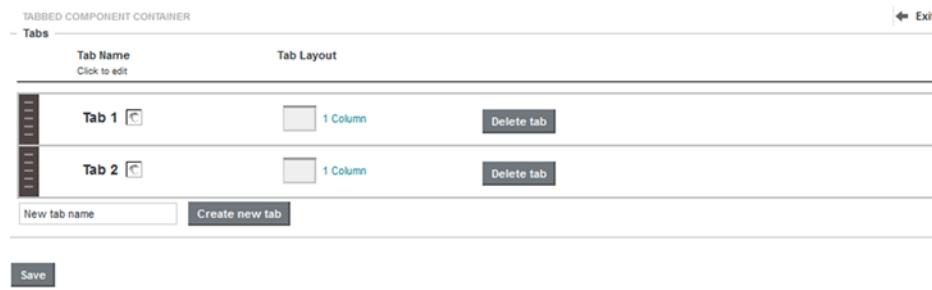
You also cannot add a **Tabbed Component Container** to another **Tabbed Component Container**.

## Configuring a Tabbed Component Container

For a **Tabbed Component Container** component, you can add, rename, and delete tabs, and localize the tab labels.

To configure a **Tabbed Component Container**:

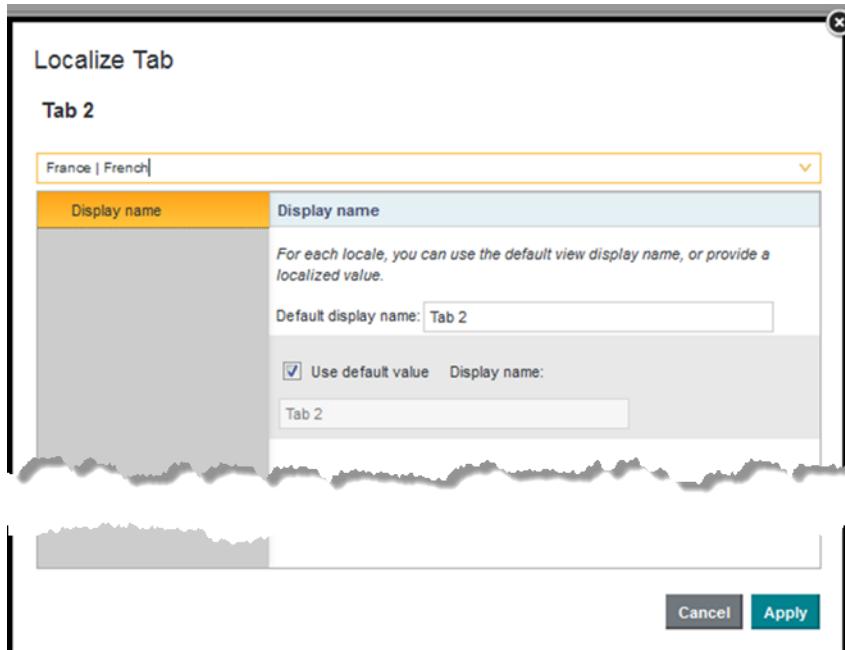
1. On the component edit view, to add a new tab to the component:



- (a) In the **New tab name** field, type the name of the new tab.  
(b) Click **Create new tab**.
2. To change the order of the tabs, use the tab drag handle at the left to move the tab to the correct location.
3. To delete a tab, click the **Delete tab** button for the tab.
4. To rename a tab:
  - (a) Click the tab name.  
The tab name becomes an editable field.
  - (b) In the field, type the new label for the tab.
  - (c) Click **OK**.
5. To localize the name of the tab:
  - (a) Click the localization icon next to the tab name.

- (b) On the **Localize Tab** dialog, from the locale drop-down list, select the locale for which to provide a value.

By default, each locale uses the display name you provided on the edit view.



- (c) To provide a custom value, uncheck the **Use default value** checkbox.
- (d) In the **Display name** field, type the localized version of the tab name.
- (e) When you have finished providing the localized values for the tab name, click **Apply**.
6. To change the layout of a tab:
- Click the layout name.  
The list of available layouts is displayed, with the current layout selected.
  - Click the radio button for the layout you want to use.
  - Click **OK**.

After configuring the **Tabbed Component Container**, you can drag other components onto each tab and configure them as usual.



## Chapter 31

---

# Data Record Displays

These components provide a detailed view of records for the current refinement.

[Results List](#)

[Results Table](#)

## Results List

The **Results List** component displays a list of records in a list format similar to regular web search results.

[About the Results List](#)

[Using the Results List](#)

[Configuring a Results List](#)

## About the Results List

The **Results List** component is designed to provide a meaningful summary of results, and is especially useful for records that include unstructured data such as long text fields.

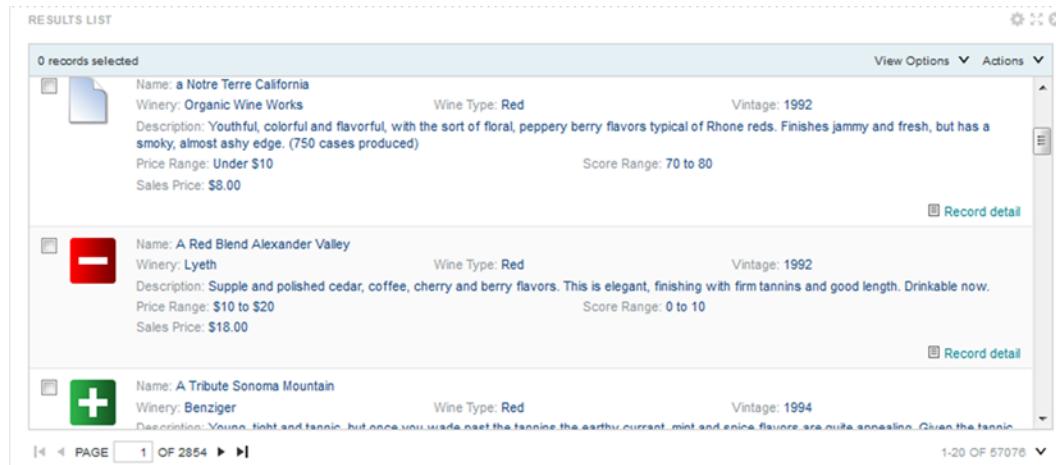
In the list, each record includes a selected set of attributes.

The screenshot shows the Oracle Big Data Discovery interface with the 'RESULTS LIST' component. The component displays three wine records in a list format. Each record includes a checkbox, an icon, and a summary of attributes: Name, Winery, Wine Type, Vintage, and Description. The 'View Options' and 'Actions' dropdown menus are visible at the top right. The bottom of the screen shows a navigation bar with 'PAGE 1 OF 254' and '1-20 OF 57078'.

Record	Name	Winery	Wine Type	Vintage	Description
1	a Notre Terre California	Organic Wine Works	Red	1992	Youthful, colorful and flavorful, with the sort of floral, peppery berry flavors typical of Rhone reds. Finishes jammy and fresh, but has a smoky, almost ashy edge. (750 cases produced)
2	A Red Blend Alexander Valley	Lyeth	Red	1992	Supple and polished cedar, coffee, cherry and berry flavors. This is elegant, finishing with firm tannins and good length. Drinkable now.
3	A Tribute Sonoma Mountain	Benziger	Red	1994	Young, tight and tannic, but once you taste the tannins, the earthy, currant, mint and anise flavors are quite appealing. Given the tannin

## Using the Results List

The **Results List** component is used to view record information. The list looks very similar to the search results from a standard web search.

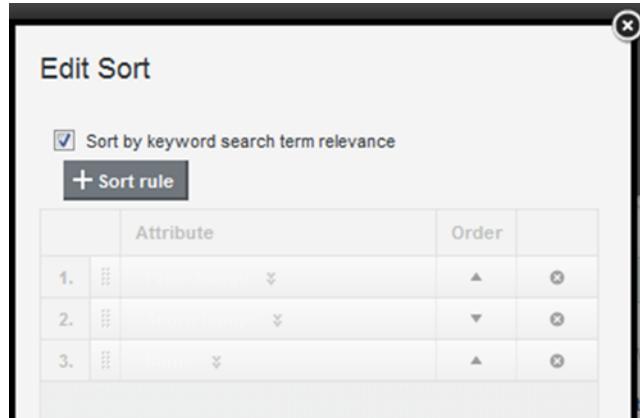


The screenshot shows a 'RESULTS LIST' interface with the following details:

- 0 records selected**
- View Options** and **Actions** menus.
- Three wine records listed:
  - Name: a Notre Terre California**  
Winery: Organic Wine Works  
Wine Type: Red  
Vintage: 1992  
Description: Youthful, colorful and flavorful, with the sort of floral, peppery berry flavors typical of Rhone reds. Finishes jammy and fresh, but has a smoky, almost ashy edge. (750 cases produced)  
Price Range: Under \$10  
Sales Price: \$8.00
  - Name: A Red Blend Alexander Valley**  
Winery: Lyeth  
Wine Type: Red  
Vintage: 1992  
Description: Supple and polished cedar, coffee, cherry and berry flavors. This is elegant, finishing with firm tannins and good length. Drinkable now.  
Price Range: \$10 to \$20  
Sales Price: \$18.00
  - Name: A Tribute Sonoma Mountain**  
Winery: Benziger  
Wine Type: Red  
Vintage: 1994  
Description: You can taste the tannins, the earthy, currant, mint and anise flavors are quite appealing. Given the tannin
- Record detail** links for each record.
- Page navigation: **1 OF 2854**
- Total records: **1-20 OF 57076**

The **View Options** menu includes a **Sort** option to allow you to sort the records based on the displayed attributes.

If there is currently a search refinement in place, then when you select the **Sort** option, on the **Edit Sort** dialog, the **Sort by keyword search term relevance** checkbox is checked and enabled.



The 'Edit Sort' dialog is open, showing the following configuration:

- Sort by keyword search term relevance** checkbox is checked.
- + Sort rule** button.
- A table for defining sort rules:
 

	Attribute	Order
1.	▼	▲
2.	▼	▼
3.	▼	▲

You can then uncheck the checkbox in order to sort by one of the displayed attributes.

To restore the sort order back to the default display, from the **View Options** menu, select **Reset to default**.

The displayed attributes can be links to refine the data, display record details, or navigate to a URL. Each record can also have a separate link to display the **Record Details** dialog with the details for that record.

From the **Actions** menu, you may be able to print and export the results. You may also be able to select records in order to compare them using the **Compare** dialog.

See [Using Common Component Functions on page 133](#).

## Configuring a Results List

When the **Results List** component is first added, it displays the first five refinable attributes for each record. You determine how the component is displayed and how end users can interact with each displayed attribute.

For information on selecting the data to use for the component, see [Selecting the view to use for a component on page 89](#).

### Default configuration for a Results List

When you first add a **Results List** component to a page it includes five attributes - the identifying attributes from the first base view, plus up to four additional attributes.

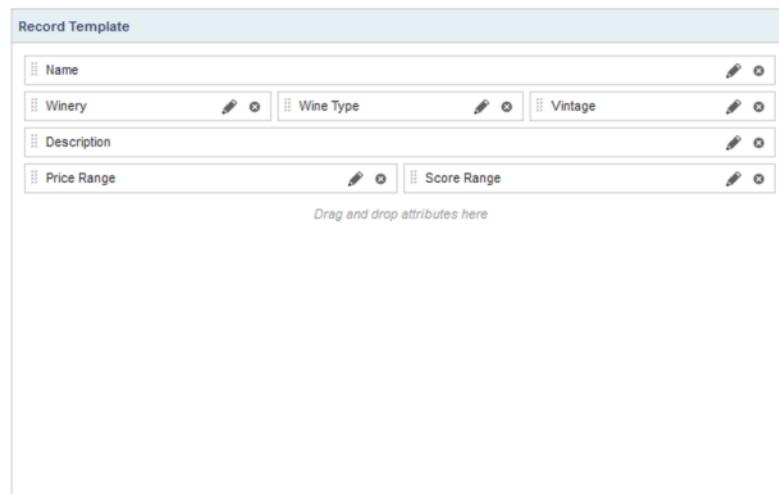
The attributes are each displayed in a separate row. The identifying attributes are displayed at the top, and in a slightly larger, bold font.



### Selecting the attributes to display for each record

On the **List Template** tab of the **Results List** component edit view, you select the attributes to display for each record.

The record display consists of multiple rows. Each row can contain values for up to 3 attributes. You can add additional rows and additional blocks to each row. For information on selecting attributes for a component, see [Selecting the attributes to use on a component on page 90](#).



To select the attributes to display for each **Results List** record:

1. To add an attribute to the template, drag the attribute from the attributes list to the appropriate location in the list template.  
You can use the attribute to create a new row or add a new block to an existing row.
2. To change the location of an attribute, you can drag it to another location in the template.
3. To remove an attribute from the template, click its delete icon.

## Configuring the attributes in a record

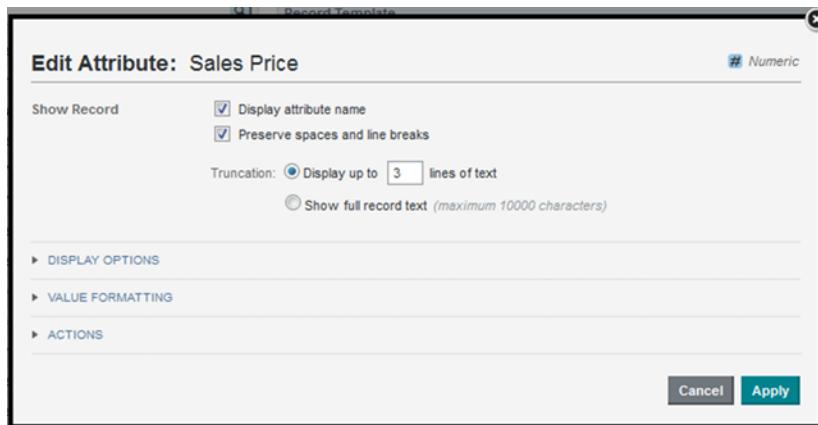
For each attribute selected to display for a **Results List** component, you configure the display and behavior of the attribute value.

Note that for date/time values, the **Results List** component automatically uses the most specific subset enabled for the attribute on the **Views** page. For example, if Year-Month and Year-Month-Day are enabled, then Year-Month-Day is always used. If Year-Month-Day-Hour-Second is enabled, then that value is always used. See [Determining the available subsets of date/time attributes on page 201](#).

From the **List Template** tab of the **Results List** edit view, to configure an attribute:

1. Click the edit icon for that attribute.

The **Edit Attribute** dialog is displayed.



2. By default, the **Display attribute name** checkbox is checked, indicating to display the attribute name as well as the value. To only display the value, uncheck the checkbox.
3. To preserve any line breaks and white space in the attribute value, check the **Preserve spaces and line breaks** checkbox. The box is checked by default.

This option is useful for longer text values that may include some HTML formatting. Except for the white space and line breaks, all HTML code is stripped from the attribute value.

4. Under **Truncation**, to determine how longer values will be truncated:
  - To display the value up to a specific number of lines of text, click the **Display up to** radio button. In the field, type the number of lines to display.
  - To display the full text of the value, up to the maximum allowed, click the **Show full record text** radio button.

5. Under **Display Options**, use the **Text Format** settings to configure the text formatting for the value. This includes:
  - The size of the text
  - Whether to display the value in bold type
  - The horizontal alignment within the template block
6. For a multi-value attribute, use the **Multi-value Display** setting to indicate whether to display the values as a comma-separated string, or in rows, with one value per row.
7. If the attribute has formatting options, then use the **Value Formatting** section to configure the format. For multi-value attributes, the **Multi-Value Formatting** section allows you to configure the number of values and the delimiter. For details on formatting displayed values, including values for multi-value attributes, see [Configuring the format of values displayed on a component on page 93](#).
8. Under **Actions** drop-down list, configure whether the attribute value is clickable, and if so, what happens when end users click the value. For details on configuring actions for displayed values, see [Configuring actions for displayed values on page 100](#).
9. To save the attribute configuration, click **Apply**.

## Configuring the sorting options for a Results List

You can configure whether end users can change the sort order for the list.

### Sort Options

*Options for setting the Sort Options.*

**Enable end user sorting controls**

**Default Sort Order:** + Sort rule

	Attribute	Order	
1.	Price Range ▼	▲	✖
2.	Score Range ▼	▼	✖
3.	Name ▼	▲	✖

On the **Sort Options** tab of the **Results List** edit view, to configure the sorting options for the component:

1. By default, end users can change the sort order, and the **Enable end user sorting controls** checkbox is checked. To not allow end users to control the sort order, uncheck the checkbox.
2. To configure the default sort order:
  - (a) To add a new sort rule, click the **+Sort rule** button.

- (b) To change the attribute assigned to a sort rule, in the **Attribute** column, click the drop-down arrow next to the attribute name, then select the new column.
- You can only sort using attributes that are displayed on the component.
- (c) To change the sort direction for a sort rule, in the **Order** column, click the sort direction toggle.
- (d) To change the order in which to apply the sort rules, drag each sort rule to the appropriate location in the list.
- (e) To remove a sort rule, click its delete icon.

## Configuring the images to display next to each record

For a **Results List** component, you can configure an image to display next to each record.

RESULTS LIST

0 records selected

	Name: A Red Blend Alexander Valley	Winery: Lyeth	Wine Type: Red	Vintage: 1992
	Description: Supple and polished cedar, coffee, cherry and berry flavors. This is elegant, finishing with firm tannins and good length. Drinkable now.			
	Price Range: \$10 to \$20		Score Range: 0 to 10	
	Sales Price: \$18.00			
	<a href="#">Record detail</a>			
	Name: A Tribute Sonoma Mountain	Winery: Benziger	Wine Type: Red	Vintage: 1994
	Description: Young, tight and tannic, but once you wade past the tannins the earthy currant, mint and spice flavors are quite appealing. Given the tannic strength, cellar into 2002. A blend of Cabernet Sauvignon, Merlot, Cabernet Franc and Petit Verdot. (215 cases produced)			
	Price Range: \$20 to \$30			

Big Data Discovery supports the following image sources:

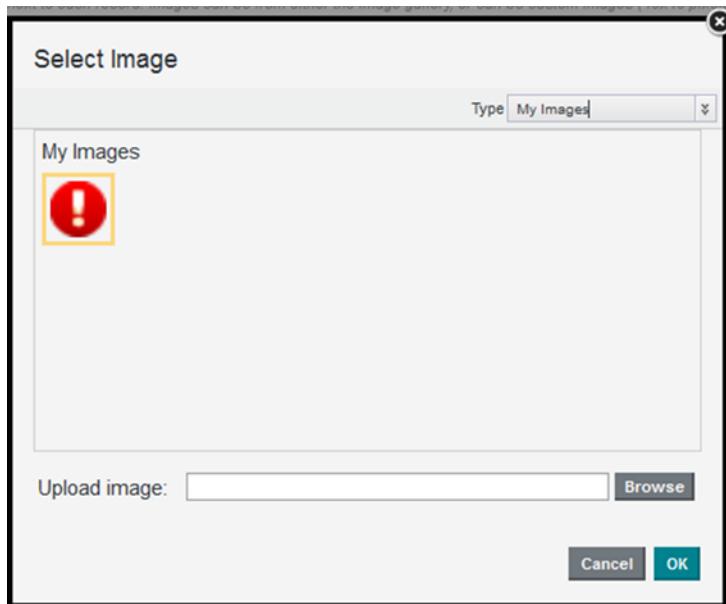
Image Type	Description
Icons selected from the image gallery	<p>Big Data Discovery provides a basic set of icon images. If you use this option, you can assign one of these images to each value of a selected attribute.</p> <p>The icon images provided by Big Data Discovery include:</p> <ul style="list-style-type: none"> <li>• "Document" images with different colors</li> <li>• File type images</li> <li>• Sentiment images</li> <li>• Social media images</li> </ul> <p>This option is typically used when the icons represent a category or other high-level classification of the records.</p> <p>You can also upload your own images to use as icons on the <b>Results List</b>.</p> <p>Big Data Discovery supports the .gif, .jpeg/.jpg, .bmp, and .png file types.</p> <p>When you upload images, they become available to use for any <b>Results List</b> component in the same project.</p>

Image Type	Description
Images from a URL that you provide	<p>If you select this option, then you provide the URL for the image. The URL can include tokens to represent the values of selected attributes.</p> <p>This is typically used when most of the records in the system have associated images published to a Web or application server (for example, a unique preview image for products represented by each record).</p>

From the **Images** tab of the **Results List** edit view, to configure the images:

1. To not display an image next to each record, under **Type of image**, click the **None** radio button.  
This is the default configuration.
2. To display an icon from a set provided by Big Data Discovery for each value of a selected attribute:
  - (a) Under **Type of image**, click the **Icon from gallery** radio button.
  - (b) To select the attribute to use, click the **Select Attribute** button. Click the attribute you want to use, then click **Apply**.  
The attribute selection dialog displays the attributes that can be used for the image selection. The attribute must have fewer than 15 values, and must be displayed on the component.
  - (c) For each attribute value, to select the image to use, click the **Select** button.
  - (d) On the **Select Image** dialog, click the image you want to use.  
To never display an image for a value, click the **No Image** option.
  - (e) To upload your own image, click the **Browse** button to search for and select the image. After selecting the file, click **Upload**.

The uploaded image is added to a new **My Images** category on the **Select Image** dialog.



If the image is too large, Big Data Discovery crops it.

- (f) Click **OK**.

The list is updated with the selected image.

Type of image:  None  Icon from gallery  Thumbnail from URL (40X40 pixels)

Select an attribute to determine the icon to be displayed:

Score Range

Attribute Value	Icon to Display
60 to 70	<input type="button" value="Select"/>
70 to 80	<input type="button" value="Select"/>
80 to 90	<input type="button" value="Select"/>
90 to 100	

Default image to display if selected image is not available:

None  Change Icon

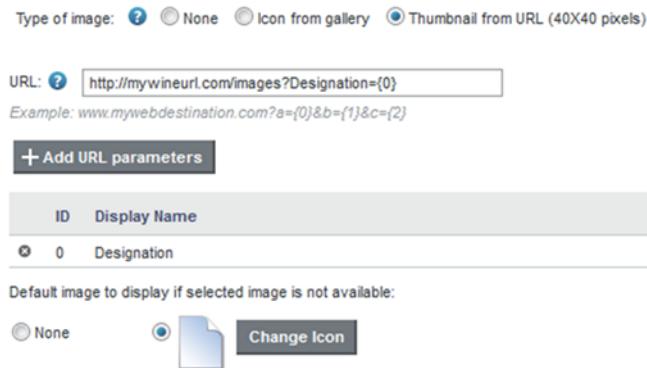
- (g) To clear a selected image, and display the default image for that attribute value, click the delete icon.
- (h) To select a different image for an attribute value, click the edit icon.

3. To display images from a URL:

- (a) Under **Type of image**, click the **Thumbnail from URL** radio button.
- (b) In the field, type the URL for the image file.
- (c) To add attribute values as parameters in the URL, click **Add URL Parameters**.
- (d) On the add parameters dialog, in the attribute list, check the checkbox next to each attribute to add.
- (e) When you are finished selecting attributes, click **Apply**.

The selected attributes are displayed in a table, with each attribute assigned an ID to use when inserting the attribute into the URL.

The attributes are also inserted as query parameters, where the parameter name is the attribute key, and the parameter value is `{IDNumber}`, where `IDNumber` is the ID for that attribute. For example: `http://www.mywines.com/images?Designation={0}`



By default, the value is encoded. To not encode the value, change the format to `{ {IDNumber} }`. For example: `{ {0} }`

You can also use the ID numbers to insert the attribute values manually.

For details on component hyperlinks and encoding inserted attribute values, see [Configuring hyperlinks to external URLs on page 105](#).

- (f) To remove a URL parameter from the table, click its delete icon.

If you did not edit the inserted query parameter, then Big Data Discovery also removes it from the URL.

If you did edit the inserted query parameter, then you must remove the parameter from the URL manually.

If you inserted the attribute value manually, then you also must remove it manually.

4. The default image setting determines the default image to display if Big Data Discovery can't find the configured image, or, for icons, if you selected the **No Image** option for a value. You can also use this setting if you want to display the same image for all of the records.

To not display a default image, click the **None** radio button.

To select a default image, click the other radio button, then click the **Change Icon** button.

On the **Select Image** dialog, click the image you want to use, then click **OK**.

## Configuring the Record Details link and Actions menu for a Results List

For the **Results List** component, you can disable or enable the **Record Details** link that displays for each record. You can also configure the **Actions** menu for the component.

On the **Actions** tab of the **Results List** edit view, if the **Enable Record Details action** checkbox is checked, then on the end user view, the **Record detail** link is displayed next to each record. For the **Record Details** action, you can configure the title to display on the **Record Details** dialog.

From the **Actions** tab, you can also control the available options in the **Actions** menu. In addition to the **Print**, **Export**, and **Compare** options, you can add **Refinement** options, to refine by attribute values from selected records.

For details on configuring the **Actions** menu for a component, see [Configuring the Actions menu for a component on page 97](#).

## Configuring display and pagination options for a Results List

For a **Results List** component, you can configure the component height, whether to enable pagination, and, if pagination is enabled, whether to allow end users to select the number of results per page.

On the **Display Options** tab of the **Results List** edit view, to configure the pagination and navigation options for a **Results List** component:

1. In the **Component height** field, type the height in pixels for the component.
2. Use the remaining fields to configure the pagination options. See [Configuring pagination options for components on page 108](#).

## Results Table

The **Results Table** component displays a set of data in a table format.

[About the Results Table](#)

[Using a Results Table](#)

[Configuring a Results Table](#)

## About the Results Table

The **Results Table** component displays a set of data in a tabular format.

The data displayed in the **Results Table** component is either:

- A flat list of records from a selected view. Each row represents a single record. The columns contain attribute values for that record.

RESULTS TABLE							
Identification		Data				Actions	
		Name	Description	Wine Type	Vintage	Winery	
<input type="checkbox"/>	34707	Alentejo Conventual	Some plum and cherry fla...	Red	1992	Adega Cooperativa de Po...	Portugal
<input type="checkbox"/>	34708	Alentejo Monte Velho	This shows supple, flesh...	Red	1992	Herdade do Esporao	Portugal
<input type="checkbox"/>	34709	Alentejo Tinto da Anfora	Smells like a melange of s...	Alentejo, Red	1992	J.P. Vinhos	Portugal
<input type="checkbox"/>	34710	Alentejo Tinto da Talha	Soft, generous and fruity...	Alentejo, Red	1992	Roquevale	Portugal
<input type="checkbox"/>	34711	Alentejo Vinha do Monte	Here's a bruiser. Deeply ...	Red	1992	Sogrape	Portugal
<input type="checkbox"/>	34712	Alentejo	Fruity, concentrated and ...	Red	1992	Adega Cooperativa de B...	Portugal
<input type="checkbox"/>	34713	Alliance California	Rustic, rough-and-tumble...	Red, Rhone Blend	1992	R.H. Phillips	Other
<input type="checkbox"/>	34714	Aloxe-Corton	Simple and appealing, wit...	Burgundy Cote de Be...	1992	Bouchard Pere & Fils	Burgundy
<input type="checkbox"/>	34715	Aloxe-Corton	Interesting floral and rhub...	Burgundy Cote de Be...	1992	Jaffelin	Burgundy
<input type="checkbox"/>	34716	Aloxe-Corton	Light and sweet-tasting, ...	Burgundy Cote de Be...	1992	Louis Latour	Burgundy

- An aggregated list of metric values calculated from view attribute values. The metric values are aggregated using dimension values displayed at the left of the table. Each row contains the values for a unique combination of dimension values.

RESULTS TABLE

0 RECORDS SELECTED

VIEW OPTIONS ▾ ACTIONS ▾

Region	Vintage	Record Count	Price (average) ▾	Review Score (aver... ▾
Alsace	1992	116	\$18.95	84.04
Alsace	1993	105	\$35.15	86.13
Alsace	1994	145	\$37.09	85.88
Alsace	1995	188	\$32.99	86.23
Alsace	1996	260	\$28.80	86.29
Alsace	1997	270	\$30.36	85.46
Alsace	1998	196	\$23.83	85.30
Alsace	1999	11	\$12.73	83.55
Alsace Summary		1,291	\$29.46	85.67
Argentina	1992	10	\$10.42	81.50
Grand Summary (474 results)		55,099	\$26.74	84.84

14 < PAGE 1 OF 28 > 1-20 OF 555 ▾

## Using a Results Table

You can page through and sort the results. You may be able to select the columns to display, and use displayed values to refine the data or display related content.

For information on using common component functions such as paging, printing, exporting, comparing, and displaying details, see [Using Common Component Functions on page 133](#).

## About the Results Table display

A **Results Table** always contains a set of persistent columns, and can also contain one or more column sets.

Persistent columns are always displayed to the left of the table.

- For a record list, any column can be a persistent column. One use of persistent columns in a record list might be to display identifying information for each record, such as an ID or name.
- For aggregated tables, the persistent columns always contain the dimension values used to aggregate the metric values.

For example, the table could have persistent columns for the Country and Product Line dimensions. The values in the other columns are then aggregated by the country and product line values.

For an aggregated table, the persistent columns are by default locked to the left of the table, and do not scroll horizontally with the other columns. For a record list, the persistent columns are unlocked by default.

The columns for the currently selected column set display after the persistent columns. For aggregated tables, the columns contain the aggregated metric values.

Aggregated tables can also contain summary rows. There can be summary rows to aggregate values across a dimension value, as well as a summary row that aggregates the values across the entire table. The same aggregation method is used for a metric throughout the table.



RESULTS TABLE

0 RECORDS SELECTED

VIEW OPTIONS ▾ ACTIONS ▾

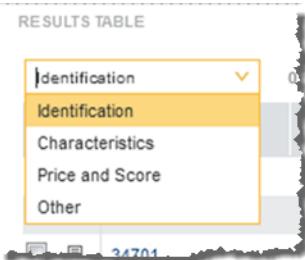
Region	Vintage	Record Count	Price (average) ▾	Review Score (aver... ▾)
Alsace	1992	116	\$18.95	84.04
Alsace	1993	105	\$35.15	86.13
Alsace	1994	145	\$37.09	85.88
Alsace	1995	188	\$32.99	86.23
Alsace	1996	260	\$28.80	86.29
Alsace	1997	270	\$30.36	85.46
Alsace	1998	196	\$23.83	85.30
Alsace	1999	11	\$12.73	83.55
Alsace Summary		1,291	\$29.46	85.67
Argentina	1999	10	\$10.42	81.58
Grand Summary (474 results)		55,099	\$26.74	84.84

PAGE 1 OF 28 1-20 OF 555

## Selecting the column set to display on a Results Table

If a **Results Table** has multiple available column sets, then a drop-down list of the column sets is displayed above the table.

From the column set drop-down list, select the column set you want to display.



The column set columns are displayed to the right of the persistent columns.

## Changing the layout of a Results Table

You may be able to change the **Results Table** layout, including the columns that are displayed, the display order, and whether to display summary rows for an aggregated table.

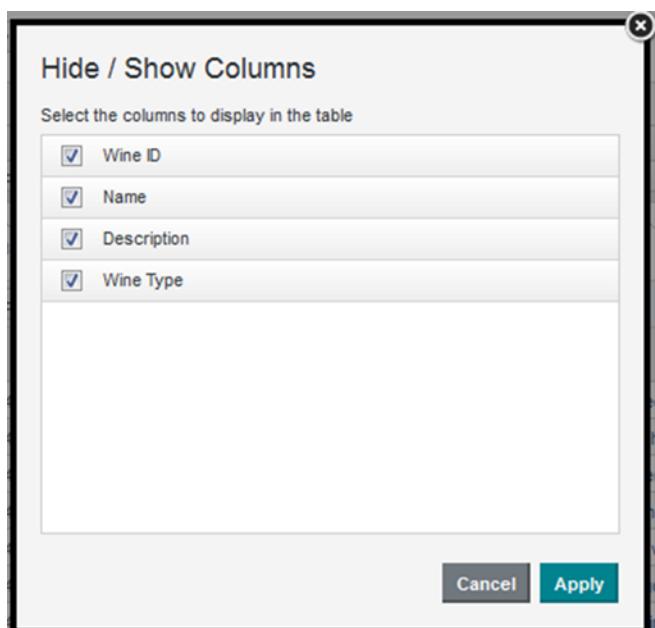
### Showing and hiding columns

You may be able to show and hide columns in the **Results Table**. For aggregated tables, when you show a hidden dimension, it is always added to the dimension columns area.

To show and hide columns:

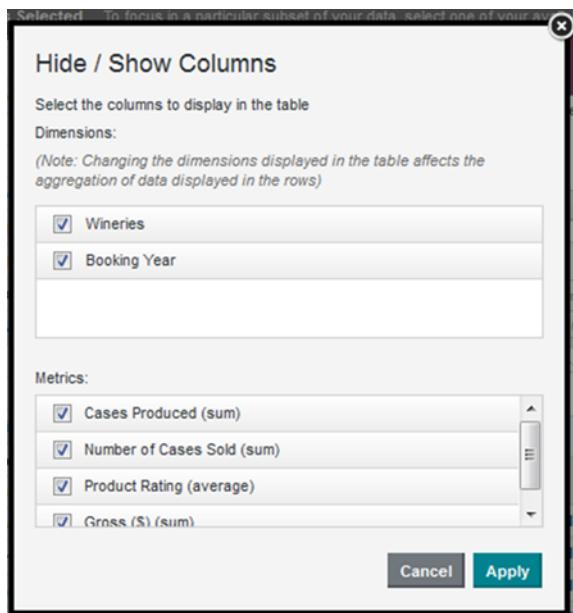
1. From the **View Options** menu, select **Hide / Show Columns**.

The complete list of available columns is displayed.



Columns that are currently displayed have their checkbox checked.

For aggregated tables, the dimension and metric columns are in separate lists.



2. To hide a currently displayed column, uncheck its checkbox.
3. To display a currently hidden column, check its checkbox.
4. When you have finished selecting the columns, click **Apply**.
5. To restore the default column configuration, from the **View Options** menu, select **Reset table to default**.

### Changing the column display order

You can change the display order of the currently displayed **Results Table** columns.

To move a column, click the column heading, then drag the column to the new location.

You can only move persistent columns or dimensions within the persistent columns area of the table. You also cannot move columns from a column set into the persistent columns.

To restore the default column configuration, from the **View Options** menu, select **Reset table to default**.

### Changing the aggregation method used for a metric

In an aggregated **Results Table**, each metric has an associated aggregation method (for example, sum or average). If the aggregation is not built into the metric, then you can assign a different aggregation method.

For example, you could change a column to display the average of the values across the selected dimension values instead of the sum.

To change the aggregation method for a metric:

1. Click the drop-down arrow immediately to the right of the column heading.
2. From the drop-down list, select the aggregation method.

## Showing and hiding summary rows on an aggregated table

If an aggregated **Results Table** is configured to allow summary rows, then you can show or hide those rows.

To show or hide the dimension summary rows:

1. Click **View Options**.
2. In the **View Options** menu, to show or hide the summary rows for each dimension, check or uncheck the **Summary rows** checkbox.
3. To show or hide the summary row for the entire table, check or uncheck the **Grand summary row** checkbox.

## Showing and hiding conditional formatting

A **Results Table** can include conditional formatting, where individual cells that contain numbers are highlighted based on the value in the cell. For example, the table may be configured to highlight in red all Sales values that are lower than 1000.

If the formatting is available, then to show or hide the formatting:

1. Click **View Options**.
2. In the **View Options** menu, to show the formatting, check the **Conditional formatting** checkbox.
3. To hide the formatting, uncheck the **Conditional formatting** checkbox.

## Changing the Results Table sort order

You can sort the **Results Table**. The sorting works slightly differently when an aggregated table has summary rows.

### Sorting a record list table or an aggregated table without summary rows

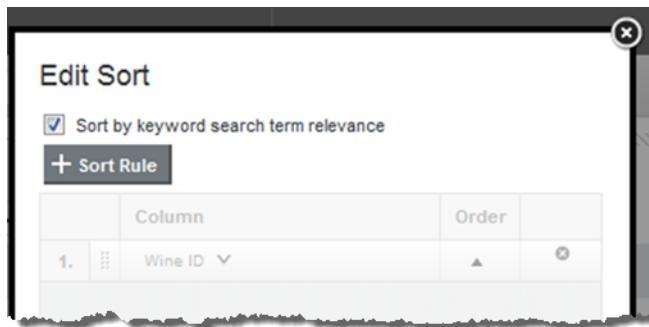
For a record list **Results Table**, or an aggregated table without summary rows, you can sort the table using any of the displayed columns. You can also create a compound sort, where the table is sorted by multiple columns in order.

So for example, you could set up a compound sort to sort by Country, then by Year. So all of the records for the United States would be grouped together, with those records then sorted by Year.

To set the sorting for a table:

1. To sort by a single column, click the column heading.  
The rows are sorted by the column values in ascending (A-Z or low-high) order.  
To reverse the sort order, click the column heading again.
2. To set up a compound sort:
  - (a) In the **View Options** menu, click **Sort**.  
The **Edit Sort** dialog is displayed. It shows the current sort order for the table.  
On the dialog, you create the list of sort rules. Each rule includes the column to sort by and the sort direction. The rule at the top of the list is used first.

Note that if there is currently a search refinement in place, then on the **Edit Sort** dialog, the Sort by keyword search term relevance checkbox is checked and enabled.



You can then uncheck the checkbox in order to sort based on the table columns.

- (b) To add a sort rule to the list, click **+Sort Rule**.

An empty sort rule is added to the bottom of the list.

	Column	Order	
1.	Wine ID ▼	▲	✖
2.	Select a column ▼	▲	✖

- (c) To select a column to use for a sort rule, click the drop-down arrow for the column name, then from the drop-down list, select the column.

You cannot select a column that is already being used in a sort rule.

- (d) To switch the sort direction for a sort rule, click the sort direction arrow in the **Order** column.
- (e) To change the order in which to apply the sort rules, drag and drop each rule to the appropriate location in the list.
- (f) To remove a sort rule, click its delete icon.
- (g) To save and implement the compound sort, click **Apply**.

### Sorting an aggregated table with summary rows

For an aggregated **Results Table** that includes summary rows, when you sort by a specific column other than the first dimension, the sort is applied within the parent dimension.

For example, an aggregated table is grouped by Country, then by Year, then by Product Line. The table includes Total Sales and Average Sales for each combination of Country and Year, with summary rows to show the total and average across the entire year and the entire country.

So if you sort by Year, the year values are only sorted in the context of each Country. If you sort by Total Sales, the sales values are sorted within each year.

## Using Results Table values to refine the data

You can use the values in a **Results Table** to refine the data.

The table can include options to:

Option	Description
Refine by a specific attribute value	For a record list table, the displayed values may be hyperlinked to allow you to refine by that value. When you click the hyperlink, the data is refined to only include records with that attribute value.
Refine by multiple attributes in the same row	A record list table can also include a column with a link to refine by selected attribute values in the row.
Refine by attribute values from multiple rows	<p>The <b>Actions</b> menu can also include options to refine by selected attribute values from the selected rows.</p> <p>To use this option, you must first check at least one row. You can select multiple rows across different pages.</p> <p>To uncheck all of the selected rows, click delete icon next to the number of selections.</p>

## Navigating to other URLs from a Results Table

A **Results Table** can include hyperlinks to other URLs. The hyperlink can pass parameters in the form of attribute values from a row or rows.

For individual rows, a column can contain a hyperlink to another URL. The URL can include attribute values for that record as part of the URL or as query parameters for the URL.

The **Actions** menu can also contain options to send attribute values from selected rows as POST parameters.

## Exporting data from a Results Table

**Results Table** can include two export options - the standard export common to all components, and an advanced export that allows to export data to HDFS.

In the **Actions** menu, the **Export** option performs the standard export of data from a component. See [Exporting data from a component on page 136](#).

The **Advanced Export** option allows you to export all of the table columns as an Avro or CSV file to either HDFS or the desktop. This option includes the same configuration options as the export option in the project footer. For details on these options, see [Exporting project data on page 227](#).

## Configuring a Results Table

When the **Results Table** is first added, it uses the default configuration. You can then configure the table type, layout, and available end user options.

For information on selecting the view to use for the component, see [Selecting the view to use for a component on page 89](#).

## Selecting the type of Results Table

A **Results Table** may be either a flat list of records or an aggregated list of calculated values grouped by dimension values.

On the **Table Type** tab of the **Results Table** edit view, click the type of table you want to create.

- To create a flat list of records, click **Details table**.
- To create an aggregated list of metric values, click **Analytics table**.

If you change the table type after you have configured the table layout, then you lose the entire table configuration.

## Setting up the data columns for a record list Results Table

A record list table contains attributes from the selected view.

### About the default record list column configuration

For a record list **Results Table**, the default column configuration contains the identifying attributes (for a base view) and the view group-by attributes (for other views), and the default record detail groups. The identifying or group-by attributes are used for the persistent columns, and the record detail groups are used for the column sets.

Each view is defined using group-by attributes. For example, from a set of sales data, you could create a view listing the available products. The view might be grouped by product ID, and so the grouping attribute would be the product ID.

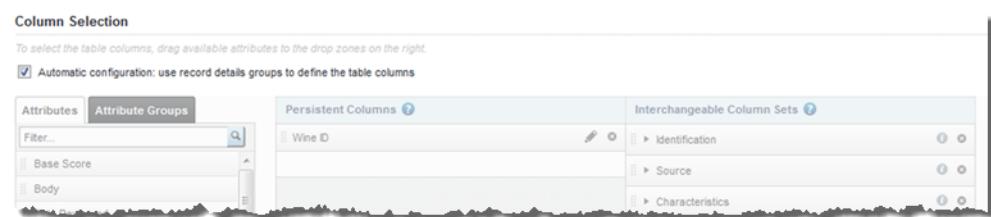
Base views do not have group-by attributes, but have attributes that are flagged as identifying attributes for the record. For a data set created from a file upload, the display name of the identifying attribute is **PRIMARY\_KEY**.

Each view can be configured with a set of attribute groups. Each group is configured to indicate whether to use it by default when viewing record details.

### Determining whether to use the default record list column configuration

For a record list **Results Table**, you can either use the default configuration, or manually select the columns.

On the **Column Selection** tab of the **Results Table** edit view, the **Automatic configuration: use record details groups to define the table columns** checkbox determines whether the table uses the default column configuration.



By default, the box is checked, and the rest of the tab is disabled.

To use a custom column configuration, uncheck the checkbox. The default configuration remains, but the lists are enabled to allow you to change the configuration.

If you check the checkbox again, the default configuration is restored, and any customization is lost.

### Available data for a record list table

On the left of the **Column Selection** tab of the **Results Table** edit view for a record list table are the available items to add to a record list table, if you are not using the default column configuration.

In the list, the **Attributes** tab contains the complete list of attributes in the selected view. It does not include predefined metrics. You can add these attributes either to the **Persistent Columns** list or to manually created column sets.

The **Attribute Groups** tab contains the list of attribute groups in the selected view. You can use an attribute group to create a column set. To display the list of attributes in a group, click the information icon for that group.

### Selecting the persistent columns for a record list table

On the **Column Selection** tab of the **Results Table** edit view for a record list table, the **Persistent Columns** list contains the columns that are always displayed. In the default column configuration, the list contains the identifying or group-by attributes for the view.

Note that you can create a record list table containing only persistent columns.

By default, the persistent columns are not locked, meaning that on the end user view they can scroll horizontally in the end user view.

If you use the default column configuration, you cannot lock the persistent columns.

If you use a custom column configuration, then you can lock the persistent columns. If you lock a column, it is moved to the top of the **Persistent Columns** list, and is locked to the left of the table in the end user view. Locked columns do not scroll horizontally.

Persistent Columns		
Wine ID		
Name		
Wine Type		

For general information on selecting attributes to use for a component, see [Selecting the attributes to use on a component on page 90](#).

On the **Column Selection** tab, to manually select the columns for the **Persistent Columns** list:

1. To add a column to the **Persistent Columns** list, drag an attribute from the **Attributes** tab to the empty slot in the **Persistent Columns** list.
2. To set the display order of the columns, drag each attribute to the appropriate location in the list. The column at the top of the list is displayed at the far left of the table.
3. To remove a column from the **Persistent Columns** list, click its delete icon.

For information on configuring columns, including how to lock persistent columns, see [Configuring columns for a Results Table on page 256](#).

## Creating and configuring column sets for a record list table

On the **Results Table**, a column set is a group of columns. While the persistent columns are always displayed, only one column set is displayed at a time. If there are multiple column sets, then end users use a drop-down list to select the column set to display.

Column sets are optional. You can configure a table with only persistent columns.

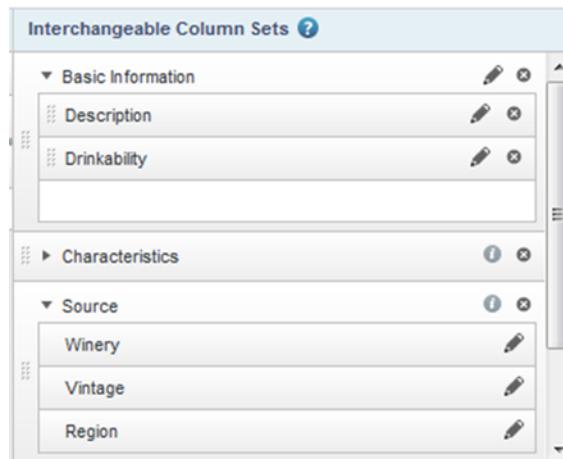
For a record list table, a column set may be either:

- An attribute group from the original view. For this type of column set, you cannot change the column set name, and you cannot add or remove columns from the set.
- A manually created column set. For this type of column set, you name the set and select the attributes that are in the set.

For the default column configuration, the columns sets consist of the attribute groups configured to be used by default for record details.

On the **Column Selection** tab of the edit view, the column sets are displayed in the **Interchangeable Column Sets** list. If you are not using the default column configuration, then the list always contains an empty set, used to add a new set to the list.

You can expand and collapse each column set to show or hide the list of columns.



The order of the column sets in the list determines the display order in the end user drop-down list. The column set at the top of the list is selected by default on the end user view.

To create and configure column sets:

1. To create a new column set from an attribute group, drag the group from the **Attribute Groups** tab to the **Interchangeable Column Sets** list.

The column set uses the attribute group name. You cannot change the name of a column set created from an attribute group.

The column set uses the attributes and attribute display order configured for the attribute group. For column sets created from an attribute group, you cannot add or remove columns, and you cannot change the display order of the columns.

2. To create and configure manual column sets:
  - (a) To create a new manual column set, drag an attribute from the **Attributes** tab into the empty column set slot.

A new empty set slot is added to the bottom of the list. The default name for a new manual column set is "New set" plus a number if needed to differentiate the name from other column sets.

For information on selecting attributes for a component, see [Selecting the attributes to use on a component on page 90](#).
  - (b) To edit the name of a manual column set, click its edit icon.
  - (c) To add an attribute to a manual column set, drag the attribute from the **Attributes** tab to the empty slot in the column set.

When you add an attribute to a set, a new empty attribute slot is added.
  - (d) To control the display order of the columns in a manual column set, drag the attribute to the appropriate location in the list. You can also drag attributes to other manually created sets.
  - (e) To remove a column from a manual column set, click its delete icon.
3. To determine the display order of the sets in the end user drop-down list, drag each set to the appropriate location in the list.
4. To remove a column set, click its delete icon.

For information on configuring individual columns, see [Configuring columns for a Results Table on page 256](#).

## Setting up the data columns for an aggregated Results Table

For an aggregated table, you select the dimensions to use for the aggregation, and the metrics to display.

### Adding dimension columns to an aggregated table

On the **Column Selection** tab of an aggregated **Results Table** edit view, the **Dimension Columns** list contains the dimensions used to aggregate the metrics in the column sets. Dimension columns are always displayed to the left of the table.

By default, the dimension columns are locked, meaning that they do not scroll horizontally in the end user view.



If you do not select any dimension columns, then the table consists of a single row, with the metric values aggregated across the entire data set.

To manage the **Dimension Columns** list:

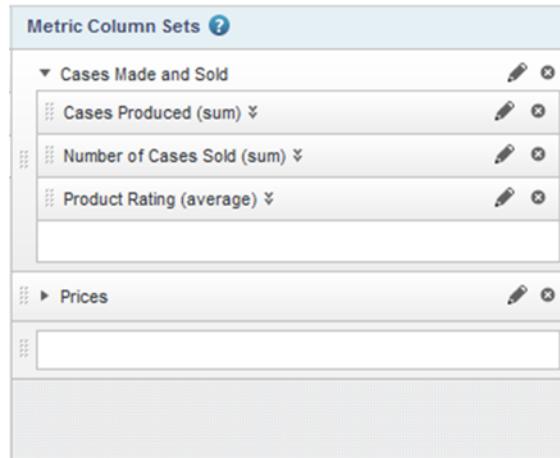
1. To add a dimension to the list, drag the dimension from the attributes list to the **Dimension Columns** list.  
For additional information on selecting dimensions for a component, see [Selecting the attributes to use on a component on page 90](#).
2. To determine the display order of the dimension columns, drag each dimension to the appropriate location in the list.
3. To remove a dimension column, click its delete icon.

For information on configuring columns, see [Configuring columns for a Results Table on page 256](#).

### Adding column sets to an aggregated table

For an aggregated **Results Table**, a column set is a group of metric values. All metric values display in column sets. While the dimension columns are always displayed, only one column set is displayed at a time. If there are multiple column sets, then end users use a drop-down list to select the column set to display.

On the **Column Selection** tab of the edit view, the **Metric Column Sets** list contains the column sets. Each column set can be collapsed to allow you to see more of the list.



The list always contains an empty column set slot, to allow you to add new sets.

The display order of the **Metric Column Sets** list determines the order in which they display in the end user drop-down list. The column set at the top of the list is selected by default when the component is first displayed.

To manage the column sets:

1. To add a new column set, drag a metric from the attributes list into the empty column set slot.

For information on selecting metrics for a component, see [Selecting the attributes to use on a component on page 90](#).

A new column set is created containing the selected metric, and a new empty column set slot is added.

The default name for a new column set is "New set", plus a number if needed to differentiate the name from other column sets.

2. To add a column to an existing column set, drag a metric into the empty metric column slot for the column set.
3. To determine the order of the column sets in the end user drop-down list, drag each set to the appropriate location in the list.
4. Within each column set, to determine the display order of the columns, drag each column to the appropriate location in the column set.
5. To configure the column set, click its edit icon.

From the **Edit column set** dialog, you can configure the column set name.

6. To remove a column set, click its delete icon.

For information on configuring columns, see [Configuring columns for a Results Table on page 256](#).

## Configuring columns for a Results Table

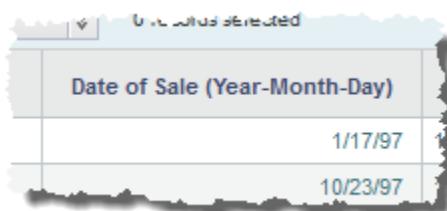
From the **Column Selection** tab of the **Results Table** edit view, you can configure the width, formatting, and other options for all of the table columns.

Note that for a record list table, if you are using the default column configuration, then you cannot configure the columns. Only default settings are used.

To configure the display of a column:

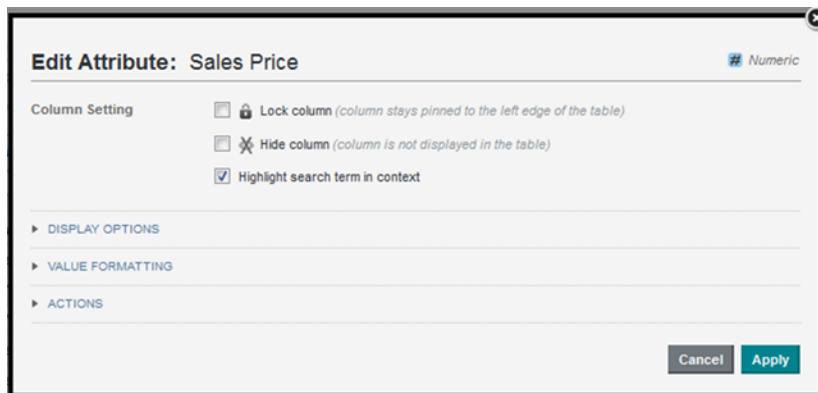
1. From the **Column Selection** tab, to display the **Edit column** dialog, click the edit icon for the column.
2. For aggregated tables, you can select the aggregation method for each metric.  
See [Selecting the aggregation method to use for a metric on page 93](#).
3. For a date/time attribute, from the **Date/time Subset** drop-down list, select the date/time subset to display.

The selected date/time subset is included in the column heading on the end user view.



4. For persistent columns and dimensions, you can lock the column, so that it is always visible at the left of the table.

To lock the column, check the **Lock column** checkbox.



To unlock the column, uncheck the checkbox. If the column is not locked, it is still at the left of the table, but can scroll horizontally.

5. By default, all of the columns are visible. However, you can configure a column to be hidden when the table is first displayed.

To hide the column by default, check the **Hide column** checkbox.

On the **Column Selection** tab, hidden columns are flagged with an icon.

6. For record list tables, you can indicate whether to highlight search terms that appear in the attribute value. To highlight search terms, check the **Highlight search term in context** checkbox.

7. For dimension columns in an aggregated table, you can use the **Attribute Cascade** section to configure a dimension cascade when users refine by the dimension value.

For information on configuring a dimension cascade, see [Configuring cascading for dimension refinement on page 104](#).

8. In the **Display Options** section:

- (a) Under **Alignment**, to select the horizontal alignment to use for the column value, click its radio button.
- (b) Under **Width**, to override the default width, click the **Custom** radio button, then in the field, type the width in pixels of the column.

9. The **Value Formatting** section is used to set the format for displaying the column values.

For details on formatting displayed values, see [Configuring the format of values displayed on a component on page 93](#).

10. The **Actions** section allows you to configure the action that occurs when users click the value.

When the dialog is first displayed, the section is collapsed. To expand or collapse the section, click the section heading.

For general information on configuring column actions, see [Configuring actions for displayed values on page 100](#).

In addition to the basic action configuration, for **Results Table** columns, you can use the **Action Condition** setting to also set a condition to determine whether to enable the action based on the

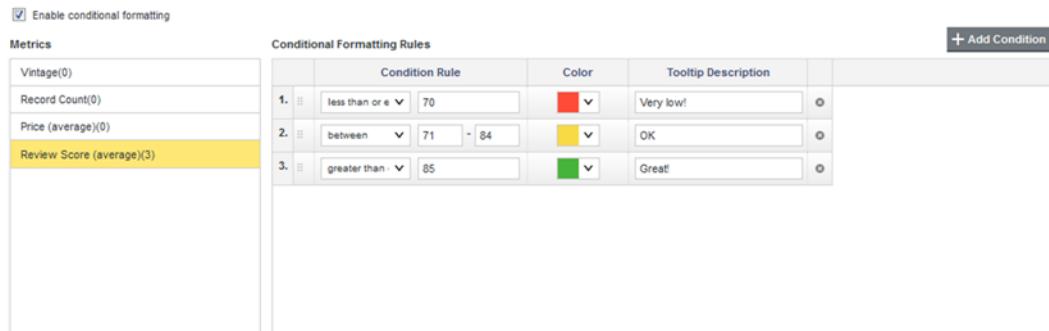
value of an attribute in the table. For example, you can configure the action to only be enabled if the **Date Reviewed** column has a non-null value.

Action Condition      If **Date Reviewed**   is **not null** then **enable this action**

For details on configuring action conditions, see [Configuring conditions for enabling Results Table actions on page 265](#).

## Highlighting Results Table values that fall within a specified range

The **Conditional Formatting** tab of the **Results Table** edit view allows you to highlight specific numeric values in a **Results Table**.



Conditional Formatting Rules			
	Condition Rule	Color	Tooltip Description
1.	less than or equal to 70	Red	Very low!
2.	between 71 and 84	Yellow	OK
3.	greater than 85	Green	Great!

To configure whether to use conditional formatting, and to set the values to highlight:

1. To allow conditional formatting, make sure that the **Enable conditional formatting** checkbox is checked.  
If the checkbox is not checked, then there is no conditional formatting.
2. In the list, click the attribute or metric for which to configure conditional formatting.  
Only numeric attributes and metrics are included in the list.  
Each item in the list includes the number of conditional formatting rules configured for it.
3. To add a new range of values to highlight for that attribute, click **Add condition**.
4. To configure a condition:
  - (a) From the **Condition Rule** drop-down list, select the type of comparison.
  - (b) In the field (or fields, for the **between** or **a range from** options), enter the value(s) against which to do the comparison.  
Note that for the **between** option, the values are inclusive. So if you specify a range between 20 and 30, values of 20 and 30 also are highlighted.
  - (c) From the **Color** drop-down list, select the color to use for the highlighting.
  - (d) In the **Tooltip Description** field, type a brief description of the highlighting. The description is included in the tooltip for each highlighted cell.

5. The conditions are applied based in the order they are listed. So a condition at the top of the list has a higher priority than a condition lower in the list.  
To change the priority of the conditions, drag and drop them to the appropriate location in the list.
6. To remove a condition, click its delete icon.

## Adding action columns for row-level Results Table actions

You can add actions columns to the **Results Table** to allow end users to perform an action on the specific row.

For information on configuring the **Actions** menu, see [Configuring the Actions menu for a component on page 97](#).

### Available row-level actions

For row-level action columns in a **Results Table**, the link to the action can be either a text string or an icon.

The available row-level actions are:

Action	Description
<b>Detail</b>	<p>Displays the <b>Record Details</b> dialog, which contains data associated with the selected row.</p> <p>For record list tables, the component displays the attribute groups configured on the <b>Views</b> page to be used to display record details.</p> <p>For aggregated tables, the component displays the columns from the current row.</p> <p>There can be only one <b>Detail</b> action column, which is disabled by default.</p> <p>If the view used for the table does not have any identifying attributes, then the <b>Detail</b> checkbox is disabled and locked.</p> <p>For a <b>Detail</b> action column, you can configure the title of the <b>Record Details</b> dialog.</p>
<b>Hyperlink</b>	<p>Allows users to navigate to a specified URL. The link can include attribute values from the current row.</p> <p>Users can create multiple <b>Hyperlink</b> action columns, each with unique settings.</p> <p>By default, there are no <b>Hyperlink</b> action columns.</p>
<b>Refinement</b>	<p>Refines the data based on one or more attribute values from the current row.</p> <p>Users can create multiple <b>Refinement</b> action columns, each with unique settings.</p> <p>By default, there are no <b>Refinement</b> action columns.</p>

## Selecting the action columns to include

On the **Actions** tab of the **Results Table** edit view, you use the **Record actions** list to manage the action columns. For the **Detail** action, you can only enable or disable the action. For other actions, you can add one or more instances to the table.

To select the row-level action columns:

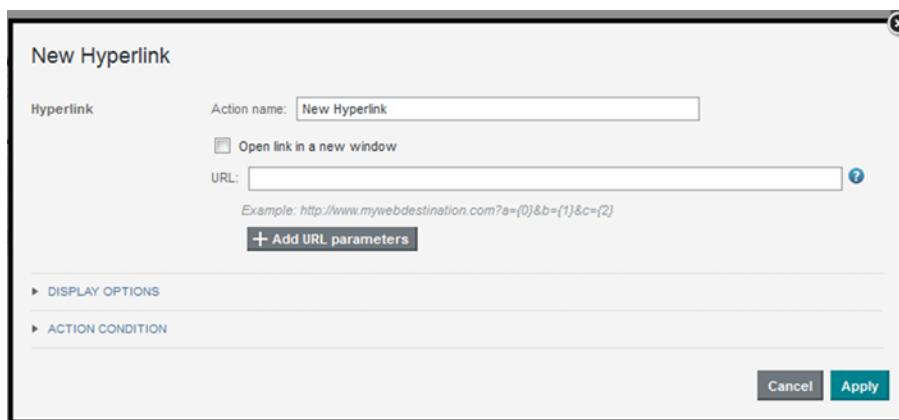
1. To include the **Detail** action, check its checkbox. To exclude the **Detail** action, uncheck the checkbox.
2. The table initially does not contain any **Hyperlink** options. You can add one or more of these columns, each with unique settings:
  - (a) To add a **Hyperlink** action column, click the **+ Action** button. From the drop-down menu, select **Hyperlink**.
  - (b) By default, the column is enabled when you add it. To exclude the action column, but not remove it, uncheck its checkbox.
  - (c) To remove a **Hyperlink** action column from the list, click its delete icon.
3. The table initially does not contain any **Refinement** action columns. You can add one or more of these columns, each with unique settings:
  - (a) To add a **Refinement** action column, click the **Action** button. From the drop-down menu, select **Refinement**.
  - (b) By default, the column is enabled when you add it. To exclude the action column, but not remove it, uncheck its checkbox.
  - (c) To remove a **Refinement** action column, click its delete icon.

## Configuring a Hyperlink action column

For a **Hyperlink** row-level action column in a **Results Table**, you can configure the name, URL, and whether to display the URL in a new browser window. You can also include attribute values as URL parameters.

To configure a **Hyperlink** row-level action column:

1. On the **Actions** tab of the **Results Table** edit view, click the edit icon for the **Hyperlink** action.
2. In the **Action name** field, type a name for the action.



3. To display the hyperlink in a separate browser window, check the **Open link in a new window** checkbox.

4. In the **URL** field, type the URL to link to.

Make sure that the URL is correctly formed, and that any special characters are properly encoded.

5. The URL can include attribute values. The values could be query parameter names or values:

`http://www.acme.com/index.htm?p1="Red"&p2="1995"`

Or could be part of the URL path:

`http://www.acme.com/wines/1995/`

To add attribute values to the URL:

- Click **Add URL Parameters**.

- On the add parameters dialog, in the attribute list, check the checkbox next to each attribute to add.

- When you are finished selecting attributes, click **Apply**.

The selected attributes are added to a table, with each attribute assigned an ID number.

The attributes are also inserted as query parameter values, where the parameter name is the attribute key, and the parameter value is `{IDNumber}`, where `IDNumber` is the ID for that attribute.

For example: `http://www.acme.com/index.htm?Region={0}&WineType={1}`

ID	Display Name
0	Price Range
1	Vintage
2	Wine Type

By default, the value is encoded. To not encode the value, change the format to `{ {IDNumber} }`.

For example: `{ {0} }`

You can also use the ID numbers to insert the attribute values manually.

For details on component hyperlinks and encoding inserted values, see [Configuring hyperlinks to external URLs on page 105](#).

- To remove a URL parameter from the table, click its delete icon.

If you did not edit the inserted query parameter, then Big Data Discovery also removes it from the URL.

If you did edit the inserted query parameter, then you must remove the parameter from the URL manually.

You must also remove manually any attribute values that you added manually.

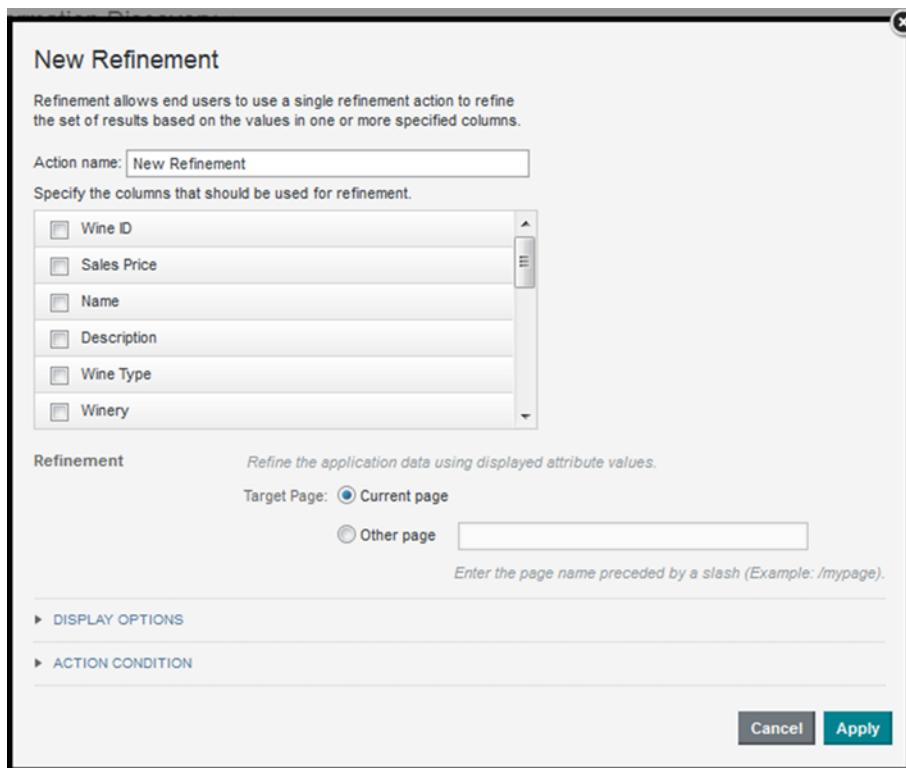
6. To save the configuration, click **Apply**.

## Configuring a Refinement action column

For a **Refinement** action column on a **Results Table**, you can select the attributes to use for the refinement, and the page on which to apply the refinement.

To configure a **Refinement** action column:

1. On the **Actions** tab of the **Results Table** edit view, click the edit icon for the action.
2. On the configuration dialog, in the attributes list, check the checkbox next to each attribute to use for the refinement.



Note that for aggregated tables, you can only use the dimension columns for refinement.

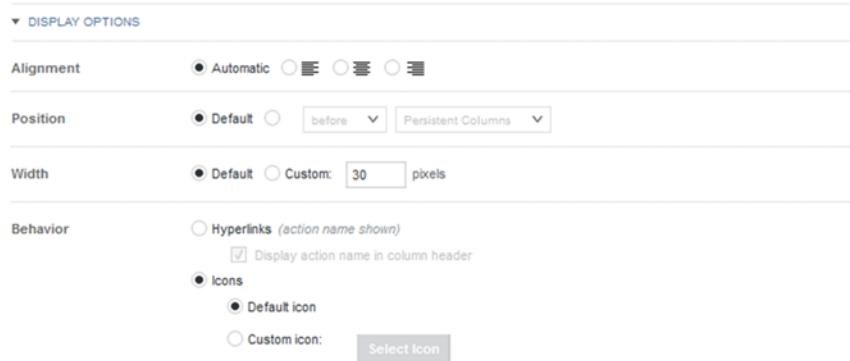
3. Use the **Target Page** setting to specify the page to display when users do the refinement.
- For information on specifying a different page in the project, see [Selecting the target page for a refinement or hyperlink on page 102](#).

## Configuring how each Hyperlink or Refinement action column is displayed

The configuration dialog for each **Hyperlink** or **Refinement** action column in a **Results Table** contains a **Display Options** section. This section is collapsed by default, and is the same for each action. It determines how the column is displayed in the table.

In the **Display Options** section, to configure the display for an action column:

1. To expand or collapse the section, click the section heading.



2. Under **Alignment**, click a radio button to indicate how to align the action icon or hyperlink within the column.
3. By default, action columns display at the far left of the table, in front of the persistent columns. Under **Position**, to change the column location:
  - (a) Click the radio button for the custom location.
  - (b) From the location drop-down list, select whether to display the column before or after the selected column set.
  - (c) From the column sets drop-down list, select the column set in which to display the action column. The drop-down list includes all of the column sets in the table, plus an entry to display the column in the persistent columns.
4. Under **Width**, to have the width set automatically, click the **Default** radio button. To set a specific width, click the **Custom** radio button. In the field, type the default width in pixels.
5. Under **Behavior**, configure how to display the action column to end users.
 

To display the action name as a hyperlink:

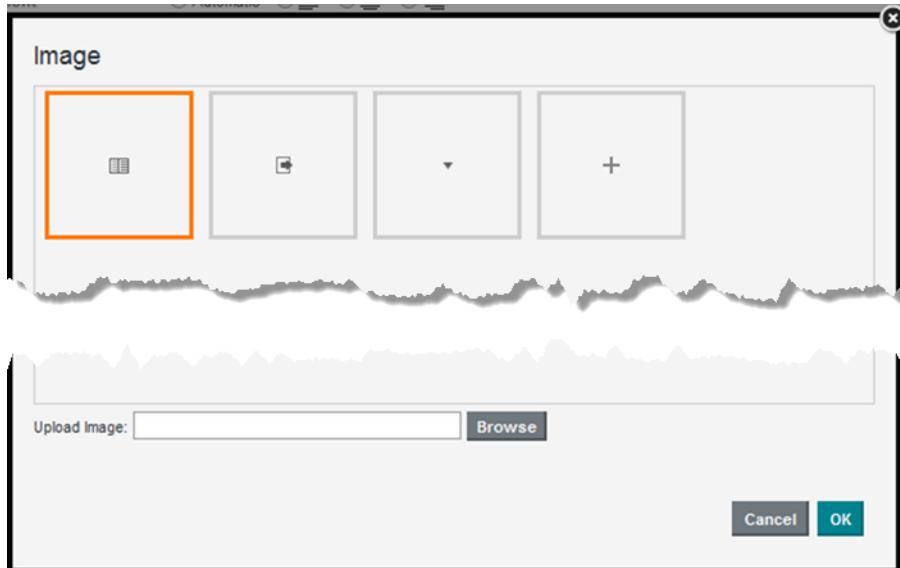
  - (a) Click the **Hyperlinks** radio button.
  - (b) To also display the action name in the column heading, check the **Display action name in column header** checkbox.

To display a clickable icon:

  - (a) Click the **Icons** radio button. This is the default. When you display an icon in the action column, the column heading is empty.
  - (b) To display an icon other than the default, click the **Custom icon** radio button.

- (c) To select the custom icon, click the **Select Icon** button.

The image selection dialog is displayed. The dialog lists any images that have been uploaded for any **Results Table** component in the current project.

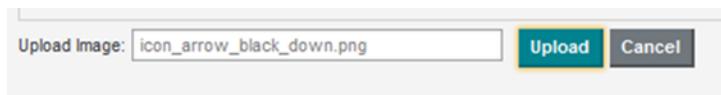


The list does not include images from other component types or projects.

- (d) To search for and select a new image file to upload, click the **Browse** button.

Big Data Discovery supports the .gif, .jpeg/.jpg, .bmp, and .png file types.

- (e) After you select the file, to clear the current selection, click **Cancel**.

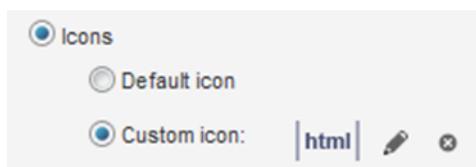


- (f) To add the image to the list, click **Upload**.

- (g) To select an image to use for the action, click the image in the list, then click **OK**.

If the image is too large, Big Data Discovery crops it.

- (h) To change to a different image, click the edit icon next to the selected image.



- (i) To remove the selected image, click the delete icon.

## Configuring conditions for enabling Results Table actions

For most **Results Table** actions, including column actions, action columns, and the **Actions** menu, you can set up a condition to determine whether the action is enabled.

Note that for the **Actions** menu, you can only configure conditions for the **Refinement** and **Pass Parameters** actions. For action columns, you can only configure conditions for the **Refinement** and **Hyperlink** actions. **Print, Export, Compare**, and **Details** actions never have conditions.

For column values, the **Action Condition** fields display in the **Column Actions** section of the column configuration dialog.

Action Condition      If **Select Attribute** is  then

For **Actions** menu and action column actions, the configuration dialog contains a separate **Action Condition** section.

Action Condition      If **Select Attribute** is  then

▼ ACTION CONDITION

To configure the condition to control whether the action is available:

1. To select the attribute for the condition, click the **Select Attribute** button. On the attribute selection dialog, click the attribute to use, then click **Apply**.  
You can select any of the columns in the table.
2. From the **is** drop-down list, select the comparison operator to use for the condition.  
The options are:

Option	Description
equal to	Indicates that the condition applies if the column value is equal to the comparison value.
not equal to	Indicates that the condition applies if the column value is not equal to the comparison value.
null	Indicates that the condition applies if the column value is null.
not null	Indicates that the condition applies if the column value is a value other than null.

3. For the equal to and not equal to comparisons, in the value field, type the value to use for the comparison.
4. From the **then** drop-down list, select whether to enable or disable the action if the condition is met.

When an action has an associated condition, then on the end user view:

Action Type	Effect of Action Condition
Column values	For rows that should not have the action, the column value is not hyperlinked.
Action columns	For rows that should not have the action, the action column is empty. It does not display the icon or hyperlink.
<b>Actions</b> menu	<p>For <b>Actions</b> menu options, users must first select at least one row to perform the action on.</p> <ul style="list-style-type: none"> <li>If all of the selected rows should not have the action, the action is disabled (grayed out) in the <b>Actions</b> menu.</li> <li>If at least one of the selected rows can have the action, the action is enabled.</li> </ul>

## Configuring the default Results Table sort order and summary row display

The **Sorting & Summaries** tab on the **Results Table** edit view allows you to configure the default sort order to use for the table. For aggregated tables, the tab also determines whether to enable the dimension and table summary rows.

For record list tables and aggregated tables without dimension summary rows, you create a list of sort rules, which are applied in order. For each sort rule, you can select a different column and determine whether to sort in ascending or descending order.

Default Sort Order:		+ Sort Rule	
	Column	Order	
1.	Winery ▼	▲	✖
2.	Vintage ▼	▲	✖
3.	Name ▼	▲	✖

For aggregated tables with dimension summary rows, the sort rules use the dimension columns from left to right, with each dimension sorted in ascending order. You cannot change the order of the sort rules, and you cannot add or remove sort rules. You can only:

- Change the sort direction of each dimension (ascending or descending)
- Change the attribute used in the last sort rule. By default, it is the last dimension in the dimension list, but you can instead select a metric column. The values in the metric column then determine the order of the values for the last dimension column.

Note that on the end user view, only the currently visible columns can affect the sort order. Dimensions that are hidden by default are not available to include in the default sort order.

On the **Summaries & Sorting** tab, to configure the sorting:

- In the **Maximum number of sort rules** field, type the maximum number of rules that can be used in a compound sort.

2. For record list tables, or aggregated tables that do not display dimension summary rows, to configure the list of sort rules:
  - (a) To add a new sort rule, click the **+Sort Rule** button.  
If the maximum number of sort rules already has been added, then the button is disabled.
  - (b) To change the column assigned to a sort rule, click the drop-down arrow next to the column name, then select the new column.
  - (c) To switch the sort direction for a sort rule, click sort arrow in the **Order** column.
  - (d) To change the order in which to apply the sort rules, drag each sort rule to the appropriate location in the list.
  - (e) To remove a sort rule, click its delete icon.
3. For an aggregated table that does display dimension summary rows, to configure the sort rules:
  - (a) To switch the sort direction for a sort rule, click the sort direction icon in the **Order** column.
  - (b) To change the column used in the last sort rule, click the drop-down arrow next to the column name, then select the new column.
4. For an aggregated table, the **Summaries** section of the **Sorting & Summaries** tab also allows you to configure whether to display summary rows by default:
  - (a) By default, the dimension summary rows are not displayed. To display by default the dimension summary rows, check the **Show summary rows by default** checkbox.
  - (b) By default, the grand summary row is displayed. To not show the grand summary row, uncheck the **Show a grand summary row by default** checkbox.

## Configuring the Results Table display and navigation

The **Display Options** tab of the **Results Table** edit view contains options for displaying and navigating through the table.

To configure these settings:

1. Under **Table height**, in the text field, type the number of rows to display at one time.
2. Use the pagination fields to configure whether users can select the number of results per page, and the available options. See [Configuring pagination options for components on page 108](#).
3. In the **Default column width** field, type the default width in pixels of a table column.
4. To allow end users to show and hide columns, check the **Allow end user to show and hide columns** checkbox.

If the checkbox is checked, then the **View Options** drop-down menu includes the **Hide / Show Columns** option.



## Chapter 32

---

# Charts

Big Data Discovery provides chart components that you can use to visualize data and the distribution of data values. The chart components include the Chart, Box Plot, Histogram Plot, and the Parallel Coordinates Plot.

[Chart](#)

[Box Plot](#)

[Histogram Plot](#)

[Parallel Coordinates Plot](#)

## Chart

The **Chart** component displays a graphical chart based on the project data.

[About the Chart component and its sub-types](#)

[Displaying details for chart values](#)

[Using dimension values to refine the data](#)

[Configuring a Chart component](#)

[Configuring the color palette for the chart display](#)

[Sorting the values on the chart](#)

[Paging controls on a chart](#)

[Saving a snapshot of the chart](#)

## About the Chart component and its sub-types

The **Chart** component supports several sub-types such as bar, line, scatter, pie, etc. and includes options for you to select the specific data to display.

For information on using common component functions such as exporting and printing, see [Using Common Component Functions on page 133](#).

The chart sub-types are:

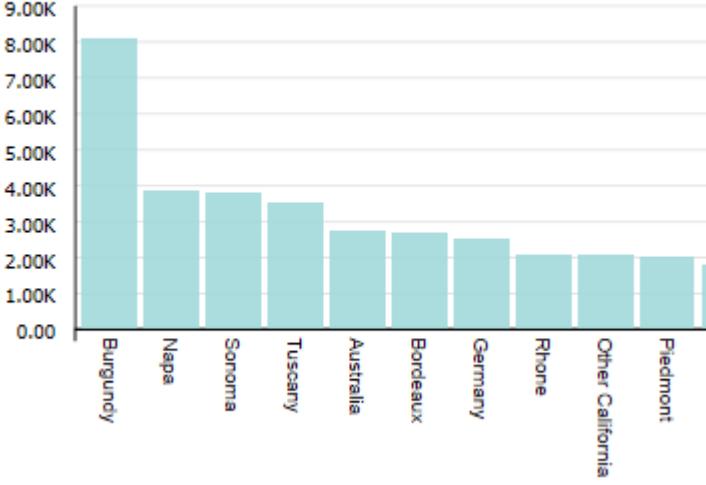
Chart Type	Description																						
<b>Bar</b>	<p>Bar charts show one or more metric values aggregated across a group dimension.</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Sales (K)</th> </tr> </thead> <tbody> <tr><td>Burgundy</td><td>8.2</td></tr> <tr><td>Napa</td><td>4.0</td></tr> <tr><td>Sonoma</td><td>4.0</td></tr> <tr><td>Tuscany</td><td>3.5</td></tr> <tr><td>Australia</td><td>2.8</td></tr> <tr><td>Bordeaux</td><td>2.8</td></tr> <tr><td>Germany</td><td>2.5</td></tr> <tr><td>Rhone</td><td>2.2</td></tr> <tr><td>Other California</td><td>2.2</td></tr> <tr><td>Piedmont</td><td>2.2</td></tr> </tbody> </table> <p>For example, a bar chart could show both the total sales and projected sales for each region.</p> <p>Bar charts are good for more precise comparisons of one or more values.</p> <p>On a basic bar chart, for each group dimension value, the metric values are displayed as a cluster of bars.</p> <p>Each metric value may be subdivided using a series dimension value. For example, the chart could show the total sales and projected sales per region for each product line.</p> <p>If there is a series dimension, then for each group dimension value, the chart displays a separate bar for each combination of series metric and series dimension.</p> <p>You can also use the trellis option to display a separate smaller chart for each series dimension value.</p>	Region	Sales (K)	Burgundy	8.2	Napa	4.0	Sonoma	4.0	Tuscany	3.5	Australia	2.8	Bordeaux	2.8	Germany	2.5	Rhone	2.2	Other California	2.2	Piedmont	2.2
Region	Sales (K)																						
Burgundy	8.2																						
Napa	4.0																						
Sonoma	4.0																						
Tuscany	3.5																						
Australia	2.8																						
Bordeaux	2.8																						
Germany	2.5																						
Rhone	2.2																						
Other California	2.2																						
Piedmont	2.2																						

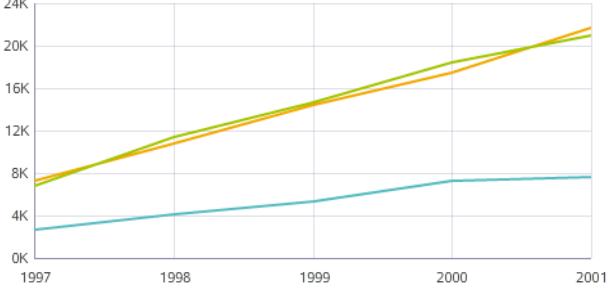
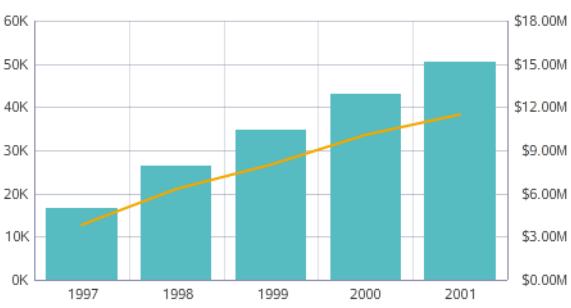
Chart Type	Description
<b>Line/Area</b>	<p>Line charts show one or more metric values aggregated across a group dimension.</p>  <p>For example, a line chart could show the change in total sales and the change in sales margins over several years.</p> <p>Line charts are usually designed to show changes or trends.</p> <p>For each group dimension value, the metric value is a data point on the line. Each metric is plotted on a separate line.</p> <p>Each metric value may be subdivided using series dimensions. For example, the chart could show the change in total sales per year for each region.</p> <p>Each combination of metric and series dimension displays as a separate line on the chart.</p> <p>You can also use the trellis option to display a separate smaller chart for each series dimension value.</p>
<b>Bar-Line chart</b>	<p>Bar-Line charts show two or more metric values aggregated across a group dimension.</p>  <p>The values for some metrics are displayed as bars, and the values for others as lines.</p> <p>In the single-axis version of the chart, both the bars and lines are plotted against the left vertical axis.</p> <p>If the lines and bars have different scales, then you can use the dual-axis version, where the bars are plotted against the left vertical axis and the lines against the right vertical axis.</p>

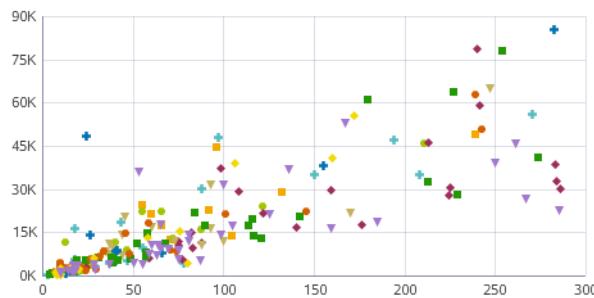
Chart Type	Description
------------	-------------

**Scatter**

Scatter charts display data points, with each point representing a dimension value.

The color dimension determines the number of colors, and the detail dimension controls the number of points in each color. Each color also uses a different shape.

The location of each data point corresponds to the values of the X-axis and Y-axis metrics.



Scatter charts allow you to show correlations between metrics. Hovering the mouse over a data point displays the complete set of dimension and metric values for that point.

You can also use the trellis option to display a separate, smaller chart for each color dimension value.

Scatter charts may also have points gathered into bins. This is called a binned scatter plot. For example, a bin may contain all points where both the X and Y values are between 0 and 50. A bin to the right contains all points where the X value is between 50 and 100, and the Y value is between 0 and 50.

The chart automatically switches between bins of points and individual points based on the number of points in the chart. The bins are shaded to indicate the relative number of points in that bin. So the more points contained in the bin, the darker the shading. For example:

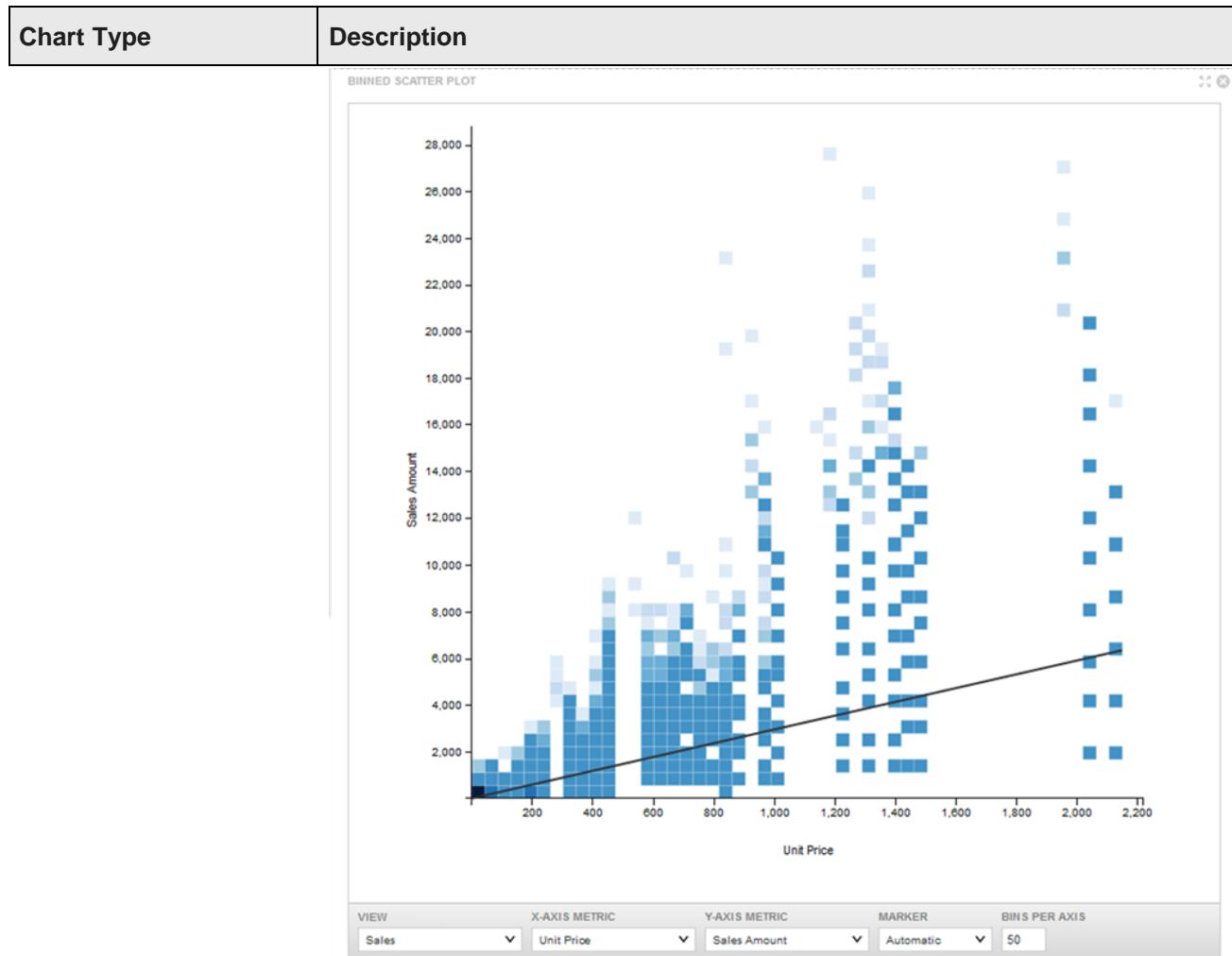
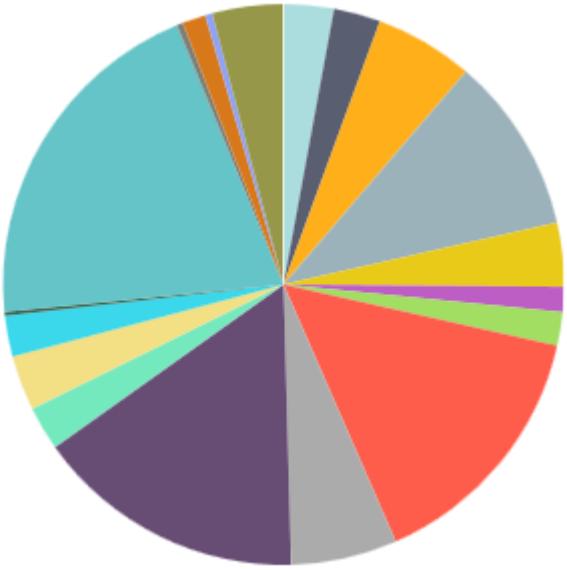


Chart Type	Description
<b>Bubble</b>	<p>Bubble charts are similar to scatter charts, but display data points as bubbles, with each bubble representing a dimension value. The color dimension controls the number of colors, and the detail dimension controls the number of bubbles in each color.</p> <p>The location of each bubble corresponds to the values of the X-axis and Y-axis metrics. The size of each bubble indicates the relative value of a third metric.</p> <p>Bubble charts allow you to show correlations among multiple metrics. Hovering the mouse over a bubble displays the complete set of dimension and metric values for that bubble.</p> <p>You can also use the trellis option to display a separate, smaller chart for each color dimension value.</p>

Chart Type	Description
<b>Pie chart</b>	<p>A pie chart shows a single metric aggregated across a group dimension.</p>  <p>For example, a pie chart could show the number of transactions for each region. Pie charts are good for doing a quick assessment of how each value contributes to the whole. The values are displayed as a pie wedge for each dimension value. The size of the wedge reflects the relative size of the metric value.</p>

## Displaying details for chart values

When you hover over a chart object (for example, line, bar, wedge, or bar section), it is highlighted and a tool tip displays with the exact dimension and metric values for that section.

## Using dimension values to refine the data

You may be able to use the dimension values to refine the data in a chart as you click a particular section.

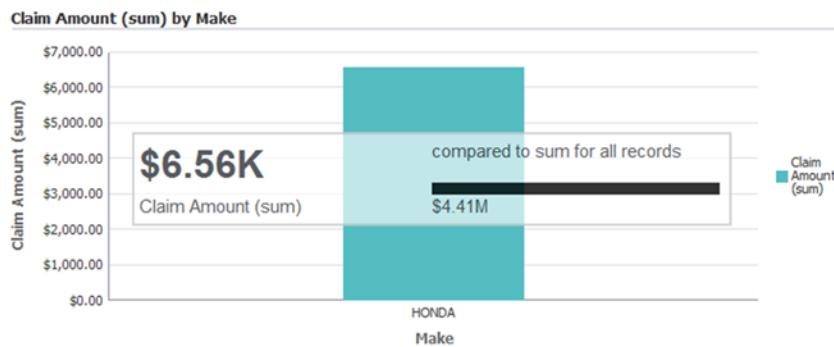
If the dimensions can be used for refinement, then:

- When you click a data point on the chart, the data is refined by all of the applicable dimension values.
- When you click a label on the category axis, the data is refined by that value only.
- When you click a legend entry, the data is refined by that value only.

For example, if the data is grouped by sales year, then when you click the axis label for 2010, the data is refined to only include records for 2010, and the selected value is added to the **Available Refinements** panel.

Chart dimension values may also allow for cascading. See [Using a component to refine data on page 135](#).

If a chart is left with a single dimension value, with no way to cascade, then the chart displays a panel listing the metric values for the current dimension value compared to the same value for all of the records.



## Configuring a Chart component

Studio creates a new **Chart** component based on a set of default values. You can re-configure the defaults with different chart data and chart display options. This is often necessary to change at least the X and Y-axis settings.

The default **Chart** component configuration starts with:

- A basic vertical Bar chart
  - The base view of the default project data set
  - For the chart metrics, the first 20 available metrics in the view.
- Each metric uses its default aggregation method. The available metrics also include the Record Count system metric, which is selected by default.
- For the chart group (category axis) and series (color) dimensions, allows end users to select from the first 20 available dimensions in the view.

By default, the category axis displays the first dimension. The chart is initially displayed with no color dimension selected or trellis dimension selected. You can then select a color dimension and trellis dimension if desired.

To configure a Chart component:

1. Select the **Discover** page, locate the chart you want to configure, and click the pencil icon to display the **Chart Settings** panel.
2. From the **Data View** list, select a data set to associate with this particular chart instance.  
This allows data set binding per chart instance. For general information on selecting a view for a component, see [Selecting the view to use for a component on page 89](#).
3. From the **Chart Type** list, select one of the following chart types:
  - **Bar** - Bar and Line/Area charts can display multiple metrics at the same time. Bar and Line/Area charts support both group dimensions (category axis) and series (color) dimensions.
  - **Line/Area**
  - **Bar-Line** - Bar-Line charts can display multiple bar and line metrics at the same time. Bar-Line charts only support group dimensions (category axis). They do not support series dimensions.

- **Scatter** - Scatter and bubble charts both support X-axis and Y-axis metrics. Only one X-axis and Y-axis metric can be used at a time. Scatter and bubble charts support color and detail dimensions.
- **Bubble** - Bubble charts also support a third metric, the bubble size metric. Only one size metric can be used at a time.
- **Pie** - Pie charts can display one metric at a time. Pie charts only support group dimensions (wedge colors). They do not support series dimensions.

The chart type you select affects the configuration options that display next.

4. Modify any of the following Chart configuration options as necessary:

*Table 32.1: Chart configuration options*

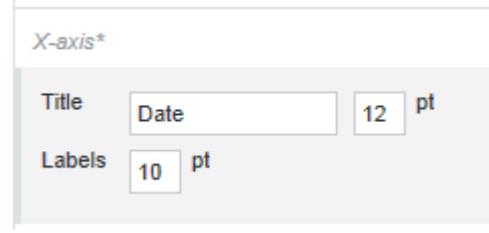
Chart Setting	Description
<b>X-axis</b>	Specifies which attribute values to display on the X-axis. To change the attribute, delete the current attribute, click <b>Add</b> , and select a new attribute.
<b>X-axis&gt;Format</b>	<p>Clicking <b>Format</b> allows you to specify a <b>Title</b> value for the X-axis. The <b>Label</b> specifies the point size of each value on the X-axis. For example:</p> 
<b>Y-axis</b>	Specifies which attribute values to display on the Y-axis. To change the attribute, delete the current attribute, click <b>Add</b> , and select a new attribute.

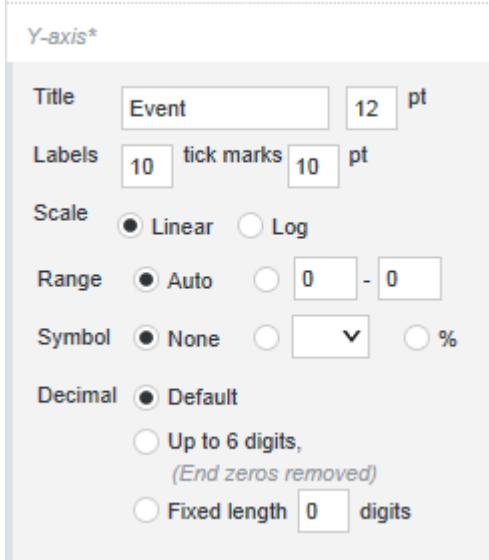
Chart Setting	Description
Y-axis>Format	<p>Clicking <b>Format</b> allows you to specify a <b>Title</b> value for the Y-axis and the <b>Label</b> specifies the point size of each value on the Y-axis. For example:</p> 

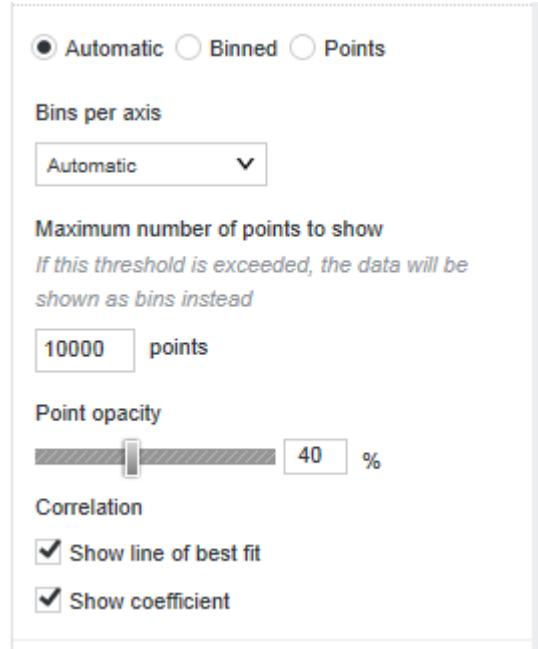
Chart Setting	Description
(Scatter options)	<p>Applies only to charts of type <b>Scatter</b>.</p> <p>With the <b>Automatic</b> option, the chart automatically switches between bins and points based on the number of points in the chart. If the number of points is lower than 10,000 points, then points are displayed. Otherwise, the chart displays bins.</p> <p>You can adjust this value in <b>Maximum number of points to show</b>.</p> <p>For example:</p> 

Chart Setting	Description
<b>Trellis</b>	<p>Specifies an attribute to display the metric values for each series dimension value as a set of separate smaller charts, called a trellis.</p> <p>For example, here is a standard stacked bar chart with both a group dimension and a series dimension:</p> <p>And here is the same chart displayed using a trellis for the series dimension:</p>
<b>Color</b>	<p>Specifies which attribute values to distinguish by color. If unspecified, the attribute values display in light blue.</p> <p>To change the attribute, delete the current attribute, click <b>Add</b>, and select a new attribute.</p>
<b>Orientation</b>	<p>Indicates whether to create vertical bars or horizontal bars. (Does not apply to Pie charts.) For example:</p> <div data-bbox="736 1543 1274 1727" style="border: 1px solid #ccc; padding: 10px; width: 330px; height: 87px;"> <p>Orientation</p> <p><input checked="" type="radio"/> Vertical bars</p> <p><input type="radio"/> Horizontal bars</p> </div>
<b>Display Settings&gt;Chart height</b>	<p>Specifies the height, in pixels, of the chart.</p>

Chart Setting	Description
<b>Display Settings&gt;Chart width</b>	Specifies the width, in pixels, of the chart. <b>Auto</b> adjusts the width of the chart to the width of the browser window.
<b>Display Settings&gt;Show recommended chart types</b>	Enabling this option displays thumb-nail sized chart alternatives underneath the main chart.
<b>Display Settings&gt;Show chart title</b>	Specifies a title in the upper left corner of the chart. By default, charts do not have a title and this option is disabled. Also, you can specify the point size of the text for the title.

- When you are finished, close the **Chart Settings** panel by clicking the X icon.

## Configuring the color palette for the chart display

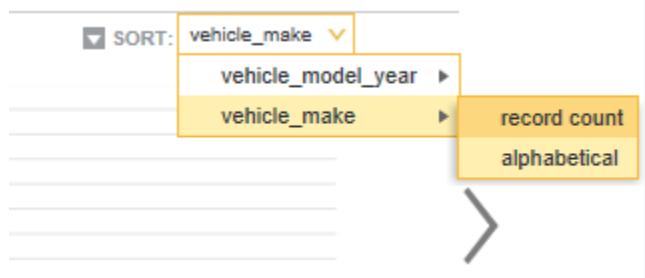
The colors used to display the charts are based on the `df.defaultChartColorPalette` setting of the **Control Panel>Studio Settings** page.

For details on how to configure Studio Settings, see the *Administrator's Guide*.

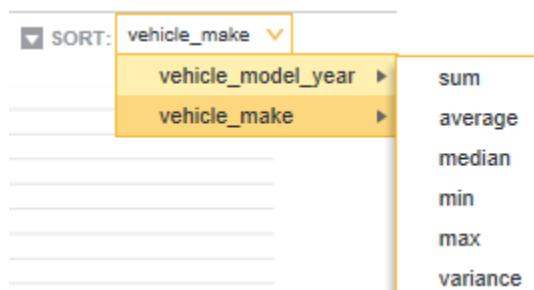
## Sorting the values on the chart

You can sort the values on a chart by either the values on the X-axis or the Y-axis. The sort options available in the sort drop down list depend on the data type of the attribute. You can sort string attributes alphabetically and numeric attributes by a variety of functions such as sum, average, median, etc.

For example, string attributes have the following:



And number attributes have the following:



To sort the values on the chart:

1. Make sure you have already configured a Chart component with values that are appropriate for the axes.
2. From the **Sort** drop-down list at the right of the chart, select the attribute to use for the sort and a sort option.
3. To change the sort order between ascending and descending, click the sort direction toggle.

## Paging controls on a chart

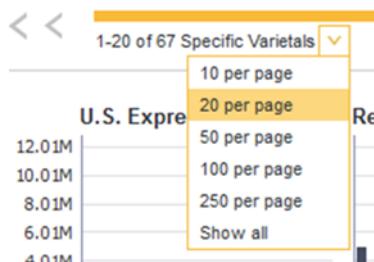
If the chart data does not fit on a single page, then Big Data Discovery provides a paging toolbar to use to page through the data.

The toolbar is divided into sections, with each section representing a page of data. The currently selected page is highlighted. To navigate to a specific page, click that section of the toolbar.



You can also use the paging icons at either end of the toolbar. At the left are icons to navigate to the first page and previous page, and at the right are icons to navigate to the next page or last page.

The left of the paging toolbar lists the current subset of values being displayed (for example, 1-50 of 500). To change the number of values displayed on each page, use the drop-down list next to the displayed subset.



## Saving a snapshot of the chart

You can create a snapshot of the currently displayed chart on the **Chart** component. The snapshot includes the current refinements to the chart data.

To save a snapshot of the chart:

1. Click the camera icon in the left bar to expose the **Snapshots** pane.
2. Select **New Snapshot**.
3. On the **Add Snapshot** dialog:
  - (a) Provide the name of the snapshot.
  - (b) Provide an optional description of the chart snapshot.

(c) Click **Save**.

The snapshot displays in the **Snapshots** pane. You can click to view the snapshot and also perform a number of addition actions from the **Actions** menu (edit, email, save to disk, or delete).

## Box Plot

A **Box Plot** shows the range of metric values associated with the values of another attribute.

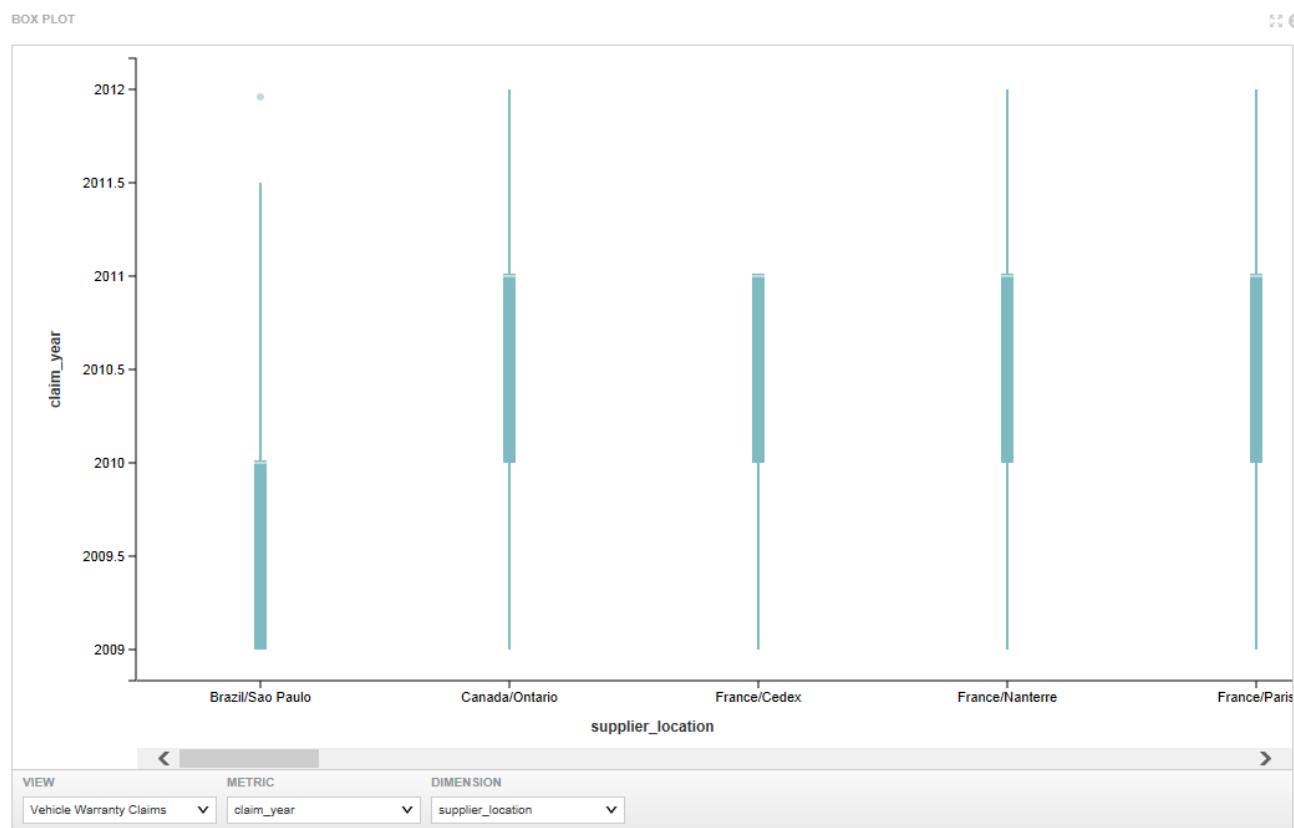
[About Box Plots](#)

[Selecting the Box Plot display options](#)

[Using a Box Plot to refine the data](#)

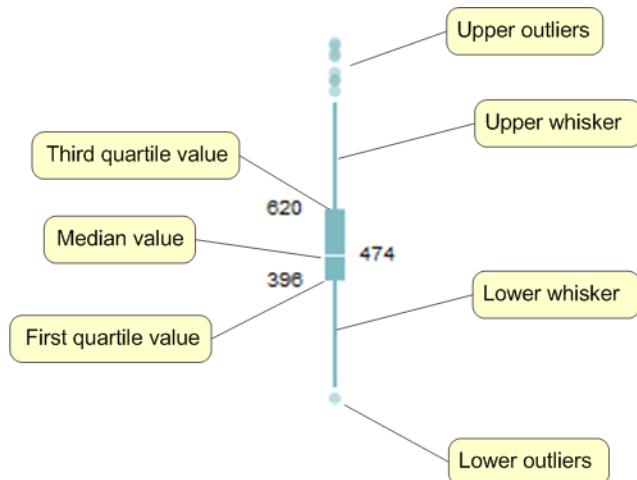
## About Box Plots

A **Box Plot** provides a capsule summary of metric values for each value of a selected string attribute.



For example, you can show a summary of the values for Number Ordered against each value of Product Category.

For each dimension value, the **Box Plot** displays a box showing:



- First and third quartile metric values for that dimension value
- Median metric value for that dimension value
- Upper and lower whisker lines.

To calculate the end points of the whisker lines, Big Data Discovery first gets the difference of the third and first quartile values, then multiplies it by 1.5.

For the lower whisker line, it then subtracts the resulting value from the first quartile value. The lower whisker line also cannot extend past the absolute minimum value.

For the upper whisker line, it adds the value to the third quartile value. The upper whisker line also cannot extend past the absolute maximum value.

- Any outlier values that fall above or below the whisker line end points

## Selecting the Box Plot display options

On a **Box Plot** component, you can select the data to display on the chart. There is no edit pane for a **Box Plot** component. The controls to configure the component are at the bottom of the chart itself.

To select the **Box Plot** display options:

1. At the bottom of the chart, locate the **View** drop-down list, and select the view to use for the chart data.

For example:

VIEW	METRIC	DIMENSION
Vehicle Warranty Claims	claim_year	supplier_location

2. From the **Metric** drop-down list, select the metric to display on the Y-axis.
3. From the **Dimension** drop-down list, select the string attribute to display on the X-axis.

## Using a Box Plot to refine the data

You can use the plot for a single attribute value to refine by one or more of the quartiles for the metric value.

To use a **Box Plot** to refine the data:

1. To refine by a single quartile of metric values, double-click that quartile.

Big Data Discovery adds two refinements to the **Selected Refinements** panel:

- One refinement is for the X axis attribute value.
- The other refinement is for the range of metric values represented by the quartile you selected.

2. To refine by values from more than one quartile:

- (a) To select the quartiles, click the mouse, then drag until the quartiles you want to include are highlighted.

- (b) Double-click the selected area.

Big Data Discovery adds two refinements to the **Selected Refinements** panel:

- One refinement is for the X axis attribute value.
- The other refinement is for the range of metric values represented by the quartiles you selected.

## Histogram Plot

A **Histogram Plot** shows the distribution of values for an attribute.

[\*About the Histogram Plot\*](#)

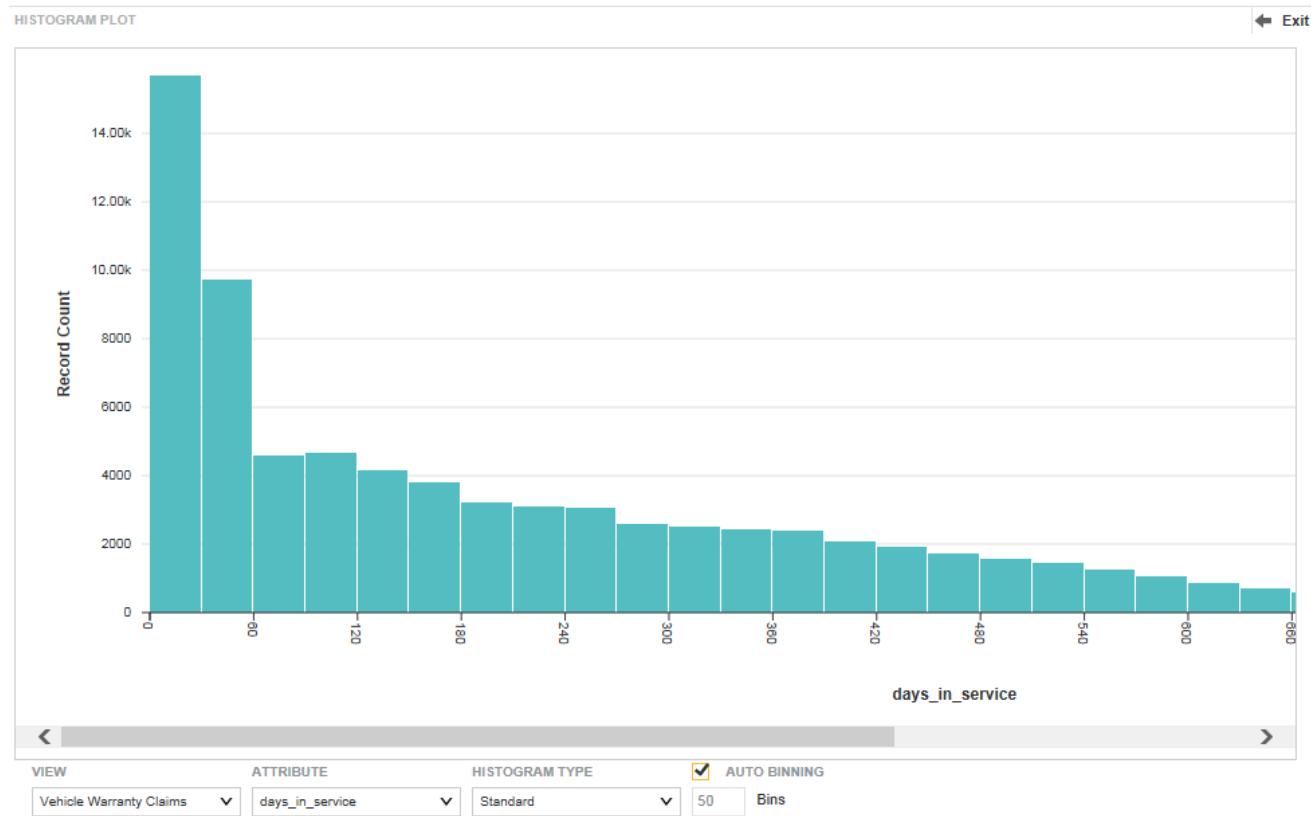
[\*Selecting the Histogram Plot options\*](#)

[\*Using a Histogram Plot to refine data\*](#)

## About the Histogram Plot

A **Histogram Plot** is a basic visualization for showing the distribution of values for a single metric.

On the **Histogram Plot**, the X axis contains bars with binned data reflecting ranges of attribute values.



For example, the first bin might reflect values between 0 and 50, the second bin values between 50 and 100, and so on.

The Y axis shows the number of records with values that fall within each range of values.

To display the range of values for a selected bin, hover the mouse over the bar for that binned range.

## Selecting the Histogram Plot options

When viewing a **Histogram Plot**, you can select the view, the metric, the type of histogram, and the number of bins.

To select the **Histogram Plot** options:

1. From the **View** drop-down list, select the view to use as the source of the histogram data.
2. From the **Attribute** drop-down list, select the metric for which to plot the value ranges on the X axis.

3. From the **Histogram Type** drop-down list, select the type of histogram to display.

The options are:

Histogram Type	Description
<b>Standard</b>	The standard histogram shows the actual number of records for each range of attribute values. Standard histograms also include a kernel density line.
<b>Relative</b>	The relative histogram shows the relative frequency for each range of attribute values. Relative histograms also include a kernel density line.
<b>Cumulative</b>	The cumulative histogram adds up the number of records through each range of values. So each point in the histogram represents the total number of records with values up to and including that range. For example: <ul style="list-style-type: none"><li>• The bar for range 1-10 shows the number of records with a value between 1 and 10.</li><li>• The bar for range 10-20 shows the number of records with a value between 1 and 20.</li><li>• The bar for range 20-30 shows the number of records with a value between 1 and 30.</li></ul>

4. By default, the **Auto Binning** checkbox is checked, indicating that Big Data Discovery automatically determines the number of bins to display. To set the number manually:

- Uncheck the **Auto Binning** checkbox.
- In the **Number of Bins** field, type the number of bins to display on the X axis. After typing the new value, to reflect the change, click outside the field.

The larger the number of bins, the smaller the range of values in each bin.

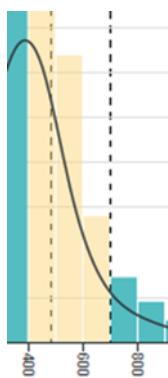
For example, if you have 50 bins, then each bin might represent a range of 50 (0-50, 50-100, etc.). If you have 100 bins, then each bin might represent a range of 20 (0-20, 20-40, etc.).

## Using a Histogram Plot to refine data

On a **Histogram Plot**, you can drag the mouse to select multiple bars, then refine the data based on the metric values for those bars.

To use a **Histogram Plot** to refine the data:

1. Click the **Histogram Plot**, then drag until all of the bars you want to include are highlighted.



2. Release the mouse, then double-click the selected area.

The metric attribute is refined to that selected range of values.

## Parallel Coordinates Plot

The **Parallel Coordinates** plot shows the value of multiple metrics across a single dimension.

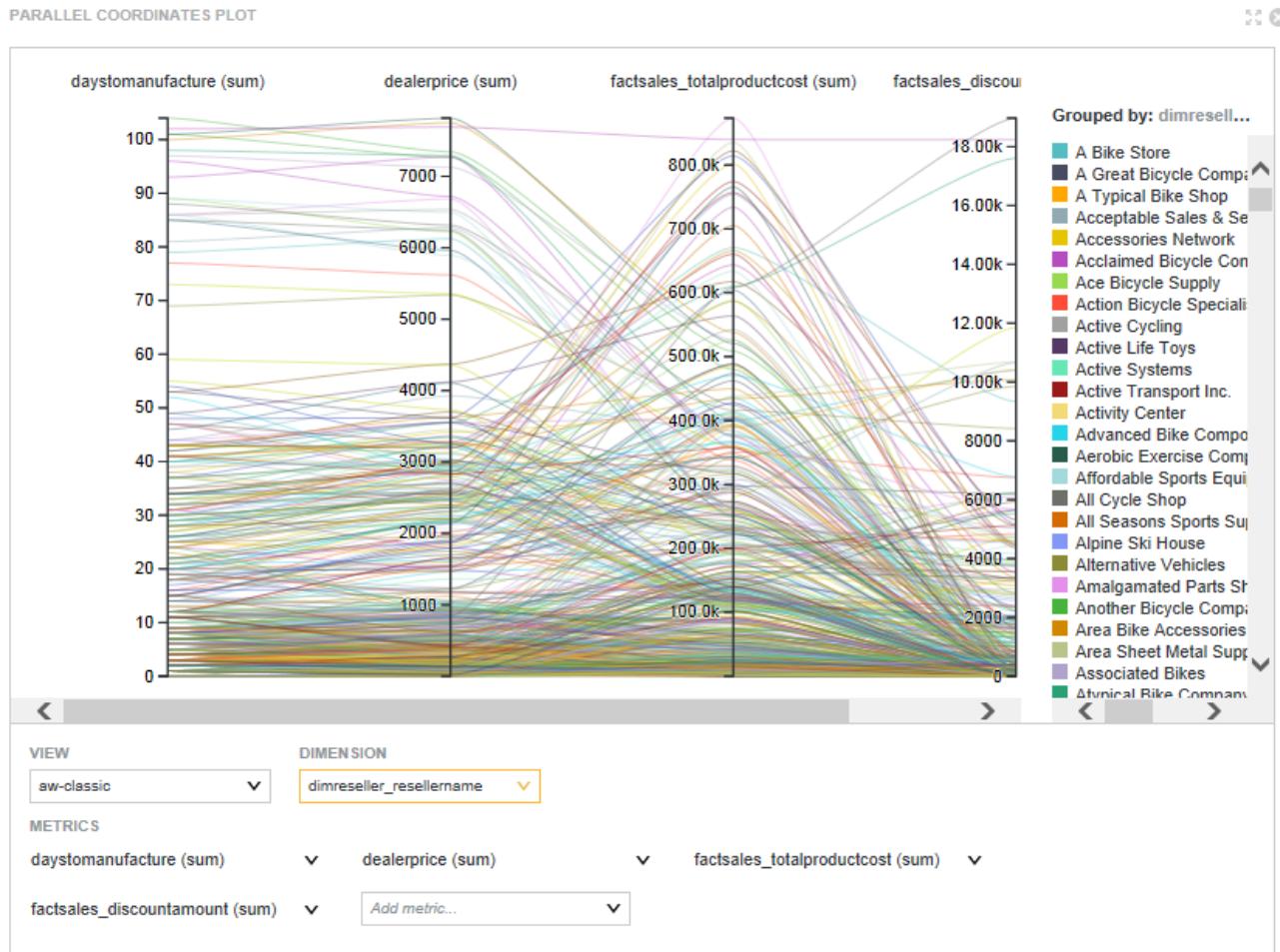
[\*About the Parallel Coordinates Plot\*](#)

[\*Selecting data to display on a Parallel Coordinates Plot\*](#)

[\*Using a Parallel Coordinates Plot to refine data\*](#)

## About the Parallel Coordinates Plot

A **Parallel Coordinates Plot** displays the values of several metrics for each value of a single grouping dimension.

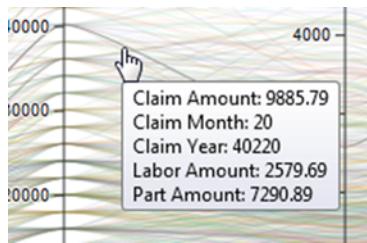


On the chart:

- Each metric is represented by a vertical axis.
- Each dimension value is represented by a colored line, which crosses each metric axis based on the metric value for the dimension value.

To focus on the metric values for a specific dimension value, hover the mouse over either a specific line or a specific entry in the chart legend.

The line for that dimension value is highlighted, and the rest of the lines grayed out. A tooltip is displayed listing the metric values.



For example, a **Parallel Coordinates Plot** could display the average price, average rating, number of cylinders, and average mileage for each value of car model.

## Selecting data to display on a Parallel Coordinates Plot

For a **Parallel Coordinates Plot**, you can select the view to use, and the metrics and dimension to display.

To select the data to display on a **Parallel Coordinates Plot**:

1. From the **View** drop-down list, select the view to use as the source of the plot data.
2. To add a new metric to the chart, select the metric to add from the **Add Metric...** drop-down list. You can select any numeric attribute.
3. To select the aggregation method for a metric, click its drop-down icon and select the operation. Note that when the data is refined to a single record, then on the **Parallel Coordinates Plot**, the aggregation is removed, and the chart shows the actual metric values for the record.
4. To remove a metric from the chart, click its delete icon.
5. From the **Dimension** drop-down list, select the grouping dimension to display on the chart.
6. To change the order of the metric axes:
  - (a) Hover the mouse over the axis label at the top of the axis you want to move.
  - (b) Click and drag the axis to the new location.

## Using a Parallel Coordinates Plot to refine data

From a **Parallel Coordinates Plot**, you can use the dimension values to refine the data.

You can either:

- Use the legend or plotted lines to select individual dimension values.
- Refine the dimension to only include values associated with a specific range of metric values.

For example, you can refine the Model attribute to only include models for which the average price is between \$15,000 and \$20,000.

To use a **Parallel Coordinates Plot** to refine the data:

1. To refine by a specific dimension value, either:
  - In the chart legend, click the value

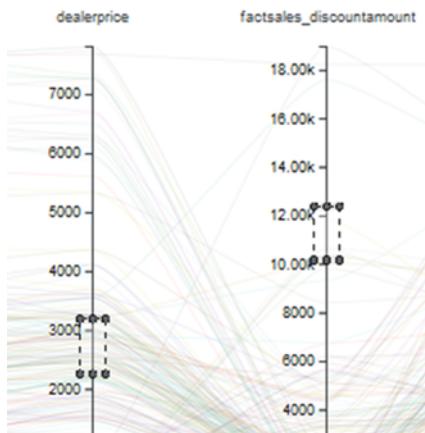
- On the chart, click the line plotted for the value

You can also use Ctrl-click to select multiple individual values.

2. To refine by dimension values that are within a range of metric values:

- (a) Use the mouse to select a range of values on a metric axis.

You can create multiple selected areas on different axes.



To clear the selected areas, click the **Clear selections** link above the chart legend.

- (b) Double-click any of the selected areas.

For all lines that run through the selected ranges, the associated group dimension values are added to the **Selected Refinements** panel.



## Chapter 33

# Other Data Visualizations

---

These components provide other options for visualizing values in your data.

[\*Map\*](#)

[\*Pivot Table\*](#)

[\*Summarization Bar\*](#)

[\*Tag Cloud\*](#)

[\*Thematic Map\*](#)

[\*Timeline\*](#)

## Map

The **Map** component displays one or more sets of geographic locations.

[\*About the Map\*](#)

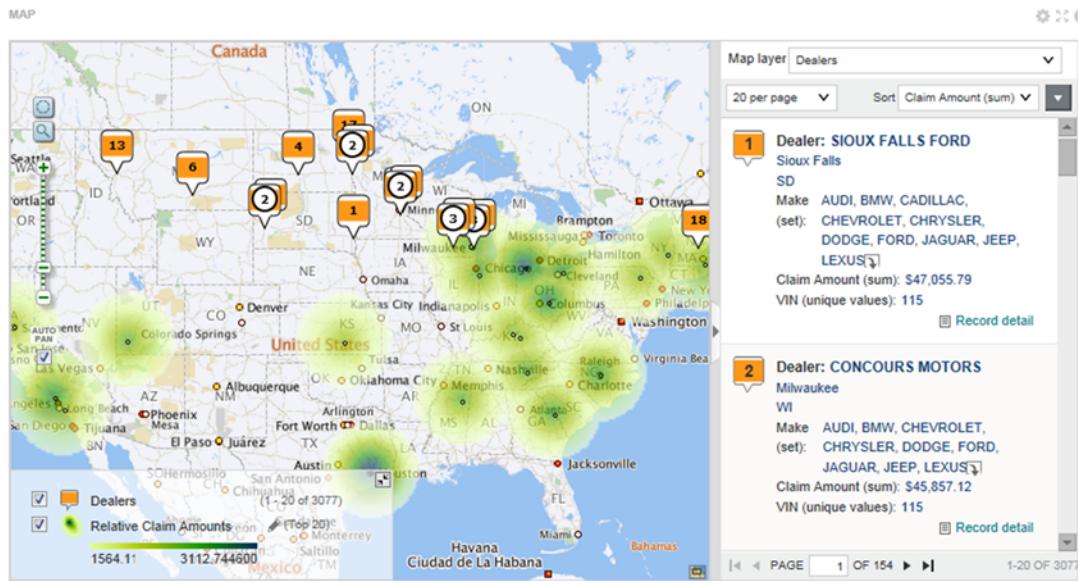
[\*Configuring the Map connection to Oracle MapViewer\*](#)

[\*Using the Map\*](#)

[\*Configuring the Map\*](#)

## About the Map

The **Map** component allows you to analyze data based on geographic location. It can only be used if the data contains at least one geocode attribute.



## Configuring the Map connection to Oracle MapViewer

The **Map** component uses Oracle's MapViewer (11.1.1.7.2 or greater) to display the map. Big Data Discovery includes framework settings to configure the connection to MapViewer.

By default, Big Data Discovery is configured to use the public instance of MapViewer. If you are using that instance, and your browser has access to the Internet, then you should not need to make any configuration changes.

Big Data Discovery includes the following framework settings related to the MapViewer connection. If you are using your own instance of MapViewer, or if your browser does not have Internet access, then you will need to

change these settings. For details on configuring framework settings in Big Data Discovery, see the *Administrator's Guide*.

Framework Setting	Description
df.mapLocation	<p>The URL for the MapViewer eLocation service. The eLocation service is used for the text location search, to convert the location name entered by the user to latitude and longitude. By default, this is the URL of the global eLocation service. If you are using your own internal instance, and do not have Internet access, then set this setting to "None", to indicate that the eLocation service is not available. If the setting is "None", Big Data Discovery disables the text location search. If this setting is not "None", and Big Data Discovery is unable to connect to the specified URL, then Big Data Discovery disables the text location search. Big Data Discovery then continues to check the connection each time the page is refreshed. When the service becomes available, Big Data Discovery enables the text location search.</p>
df.mapViewer	<p>The URL of the MapViewer instance. By default, this is the URL of the public instance of MapViewer. If you are using your own internal instance of MapViewer, then you must update this setting to connect to your MapViewer instance.</p>
df.mapTileLayer	<p>The name of the MapViewer Tile Layer. By default, this is the name of the public instance. If you are using your own internal instance, then you must update this setting to use the name you assigned to the Tile Layer.</p>

## Using the Map

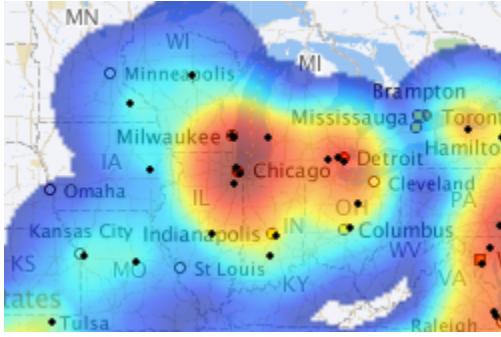
You use the **Map** component to view geographic locations.

### Types of map layers

Each **Map** component is configured with one or more map layers. Each map layer uses a specific layer type.

The available map layer types are:

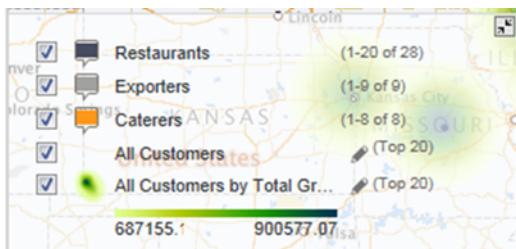
Layer Type	Description
Numbered point layer	<p>A numbered point layer displays a numbered point for each location on the map.</p>  <p>The layer includes a separate list containing the details for each point. The <b>Map</b> component can display multiple numbered point layers at the same time, with color used to differentiate the layers.</p>
Point layer	<p>A point layer displays a point for each location on the map.</p>  <p>The <b>Map</b> component can display multiple point layers at the same time, with color used to differentiate the layers.</p>

Layer Type	Description
Heat map layer	<p>A heat map layer can display:</p> <ul style="list-style-type: none"> <li>• A point for each location on the map.</li> <li>• A shaded cloud with color gradients to show either the relative density of the points on the map, or the change in an associated metric value between locations. The cloud colors are also reflected on the map points.</li> </ul>  <p>You can choose whether to display just the points, just the cloud, or both.</p> <p>The <b>Map</b> component can only display one heat layer at a time.</p>

## Showing and hiding map layers

At the bottom left of the **Map** component is the layer list, which shows the available layers configured for the component.

You can use the icon at the top right to expand or collapse the layer list.



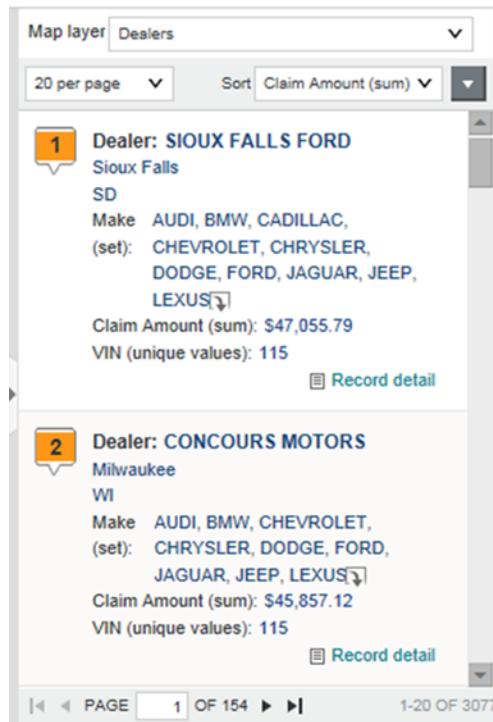
The layer list contains a checkbox for each numbered point and point layer. To hide a numbered point or point layer, uncheck its checkbox.

If any heat layers are available for the **Map** component, then there is a single heat layer checkbox, with a drop-down list containing all of the available heat layers. From the heat layer drop-down list, select the heat map to display. To not display a heat layer on the map, uncheck the heat layer checkbox.



## Using the numbered points list

When a numbered point layer is displayed on the **Map** component, the component by default displays the numbered point list to the right of the map.



You can use the expand/collapse icon to the left of the list to show or hide the list.

The list also uses the standard pagination to allow you to page through the list and control the number of points to display per page. See [Paging through component data on page 134](#).

From the numbered point list:

1. If there is more than one numbered point layer, then from the **Map Layer** drop-down list, select the layer for which to display the numbered point list.

2. To change the sort order for the numbered point list:

- (a) From the **Sort** drop-down list, select the attribute to use for the sort.

Note that if the data is currently refined using a keyword search, then the list also includes a **Search Relevance** option.

- (b) To switch the sort direction, click the sort direction toggle. When you change the sort for a numbered point layer, it affects the points displayed on the map.

## Changing the display of point and heat map layers

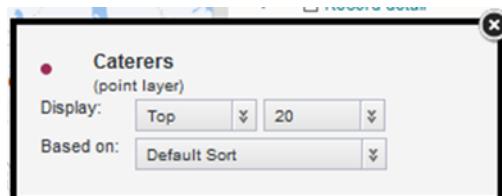
From the layer list, for point and heat map layers, you can change the amount of information displayed.

To adjust the display of point and heat map layers:

1. In the layer list, the point layer name is followed by the number of points displayed. To change the point layer display:



- (a) Click the edit icon for the point layer.
  - (b) On the layer dialog, from the **Display** drop-down list, select whether to display the top or bottom set of points.

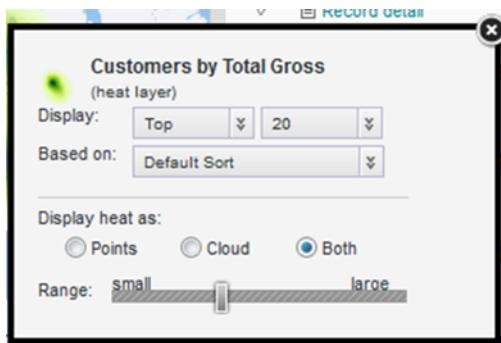


- (c) From the number drop-down list, select the number of points to display.
  - (d) From the **Based on** drop-down list, select the value to use to determine the top or bottom points.

2. In the layer list, the heat layer information includes the number of points. To change the heat layer display:



- Click the edit icon for the heat map.
- On the layer dialog, from the **Display** drop-down list, select whether to display the top or bottom set of points.



- From the number drop-down list, select the number of points to display.
- Under **Display heat as**, click the radio button to indicate whether to display only the points, only the cloud, or both the points and the cloud.
- Use the **Range** slider to determine the range for calculating the cloud.

The range determines how much a point affects the cloud calculation. The larger the range, the more impact each point has on the cloud calculation.

## Searching the map

The **Map** component can include both text and range filter search tools to allow you to find specific locations.

The text search allows you to display locations based on their proximity to a specified place name. For example, you can display locations within 10 miles of Boston, Massachusetts.



**Note:** The version of MapViewer used by the component may be limited to selected geographic areas. If you try to do a text search based on a location outside those geographic areas, then an error is displayed.

The range search allows you to only display locations in a selected area of the map.

To do text and range searches on the **Map** component:

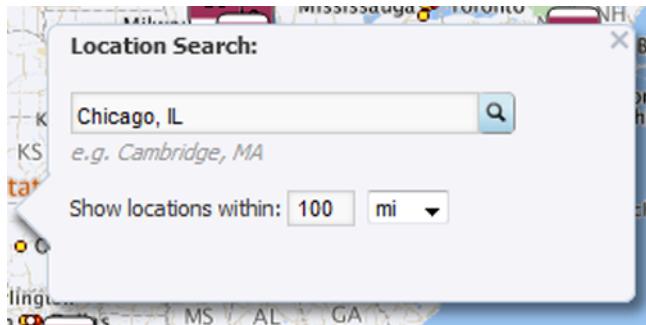
1. To do a text search:

- (a) Click the text search icon.

The **Location Search** pop-up is displayed.

- (b) In the field, type the search text.

- (c) In the **Show locations within** field, specify the number of miles or kilometers, then select the unit of measurement.



- (d) Click the search icon.

The map is updated to only show pins for locations within the specified area.

If the **Auto Pan** checkbox is checked, the map shifts automatically to display all of the pins for the first page of results.

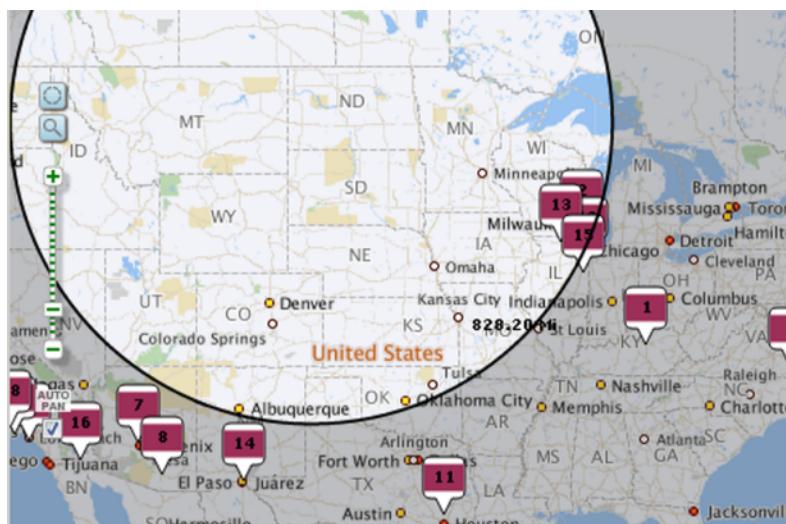
A **Range Filter** entry also is added to the **Selected Refinements** panel.

2. To do a range search:

- (a) Click the range search icon.

- (b) On the map, click and drag the mouse from the middle to the edge of the area to select.

As you drag the mouse, the distance from the point you started at is displayed.



- (c) When you have selected the area you want, release the mouse.

The map is updated to only display pins for locations within that selected area.

A **Range Filter** entry also is added to the **Selected Refinements** panel.

If the **Auto Pan** checkbox is checked, the map shifts automatically to display all of the pins for the first page of results.

## Displaying details for a map point

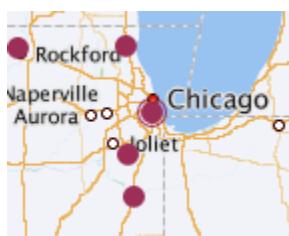
When you click a map point, Big Data Discovery displays additional information about that location.



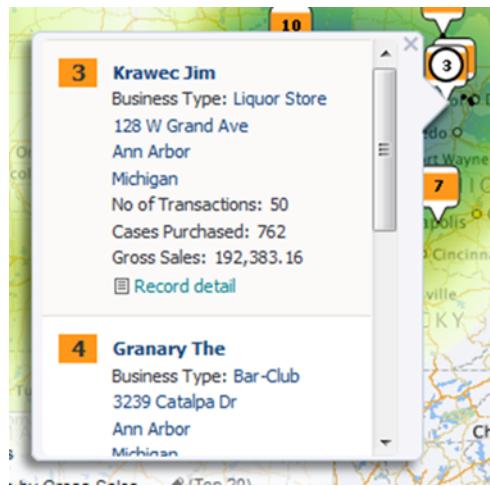
If several numbered points are clustered together, then the **Map** component displays the number of points in the cluster.



For point layers, when there is a cluster of points, Big Data Discovery displays a circle around the cluster.



When you click a cluster of points, the **Map** component displays the details for all of the points in the cluster.



The values in the point details may be configured to allow you to refine the data. See [Using a component to refine data on page 135](#).

The point details may also include a link to display record details. See [Displaying details for a component item on page 138](#).

## Configuring the Map

The **Map** component configuration allows you to configure the map layers.

### About the default Map configuration

When the **Map** component is first added, it uses the default configuration. For each base view that has a geocode attribute, the component includes a numbered point layer. The layer name is set to the view name.

The details for the default layers contain the identifying attributes for the view.

The automatically created layers are visible by default.

**Map Layers**

Create, order, and manage the information layers that appear in your map.

**Point Layers**

 Bizwine Denormalized

**Heat Layers**

No heat layers have been configured.

**+ Layer** **▼**

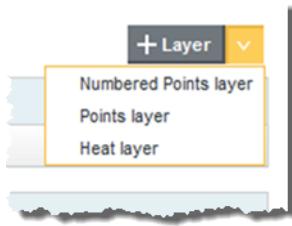
## Adding, editing, and deleting map layers

From the **Map Layers** tab of the **Map** component edit view, you manage the list of layers.

To add, edit, and delete map layers:

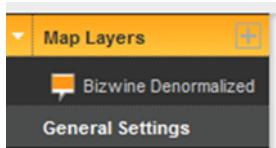
1. To add a new layer to the map, either:

- Hover the mouse over the **+Layer** button, then in the drop-down list, click the type of layer to add.



For details on the map layer types, see [Selecting the map layer type on page 303](#).

- In the edit view menu, click the **+** button next to the **Map Layers** entry in the edit view menu.



A new layer is added to the list.

If you selected the layer type when you added the layer, then the **Data Selection** tab is displayed. See [Selecting the view to use for the map layer on page 303](#).

If you did not select a layer type, then by default the layer is a numbered point layer, and the **Layer Type** tab is displayed. See [Selecting the map layer type on page 303](#).

2. To edit a layer, click the layer name.
3. To change the name of a layer:
  - (a) Double-click the layer name.  
The name is displayed in an editable field.
  - (b) Type the new name.
  - (c) Press **Enter**.
4. To delete a layer, click its delete icon.

## Configuring a map layer

For each map layer, you select the data to use and configure the display.

### Selecting the view to use for the map layer

The **Data Selection** tab for the map layer contains the list of views from the project data. Map layers can only use views that contain a geocode attribute. The geocode attribute cannot be multi-value.

Views that do not contain a geocode attribute are disabled. For details on selecting views for a component, see [Selecting the view to use for a component on page 89](#).

### Selecting the map layer type

By default, a new map layer is a numbered point layer. After you create the layer, you can select the type of layer.

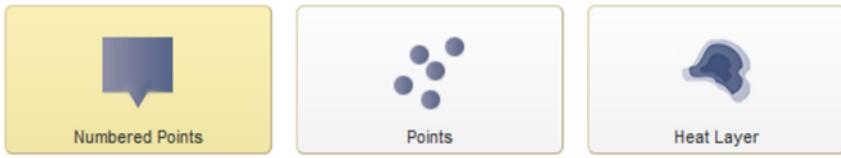
The available layer types are:

Layer Type	Description
<b>Numbered Points</b>	Displays each location on the map as a numbered point. The numbered points are also displayed in a corresponding list. End users can page through the list of numbered points. This is the default layer type for a new map layer.
<b>Points</b>	Displays each location on the map as a point.
<b>Heat Layer</b>	Can display: <ul style="list-style-type: none"><li>A point for each location on the map. The points are color coded based on the value of a selected metric.</li><li>Shaded clouds to show the change in metric values across a geographic area.</li></ul> End users can display one or the other or both.

On the **Layer Type** tab for a layer, to select the layer type, click the icon for the type.

**Customers: Layer Type**

Select the style to use to render the map layer



**Numbered Points**

- Records appear on the map as numbered pins
- Each point can show more details when clicked
- A corresponding list appears with details for each numbered pin on the page
- Numbered points layers support pagination

### Defining the map points for the map layer

On the **Points Definition** tab for the map layer, you select the geocode attribute to use for the map points, and whether the map layer uses aggregated points.

For an aggregated layer, you also select the group dimensions to use for the aggregation.

For a heat map layer, you select the metric to use to determine the point colors.

To define the map points:

- From the **Geospatial attribute** drop-down list, select the geocode attribute to use for the map locations.

**Customers: Points Definition**

Select attributes to define the layer's points

Geospatial attribute:

LatLong 

Detail points  Aggregated points

- Detail point map layers render the data points based on the grain of the records in the data view

- You also select whether the map locations need to be aggregated.

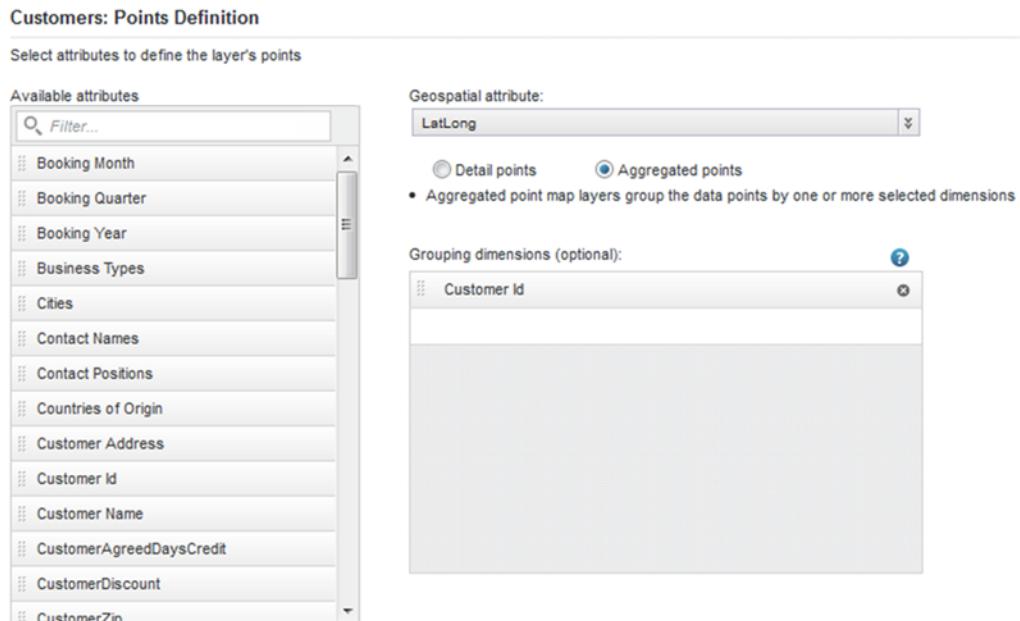
- To not aggregate the map locations, click the **Detail points** radio button.
- To aggregate the map locations, click the **Aggregated points** radio button.

For example, if the view contains a simple list of stores, with the geocode being the store location, you would not need to aggregate the locations, and could use the detail points option.

But if the view contains a list of sales transactions, with the geocode being the store location for the sale, then to generate a single list of store locations, you might aggregate by a dimension such as the store name.

3. If you are creating an aggregated map layer, then you configure the list of grouping dimensions to use for the aggregation.

- (a) To add a dimension to the list, drag a dimension from the **Available attributes** list to the **Grouping dimensions** list.



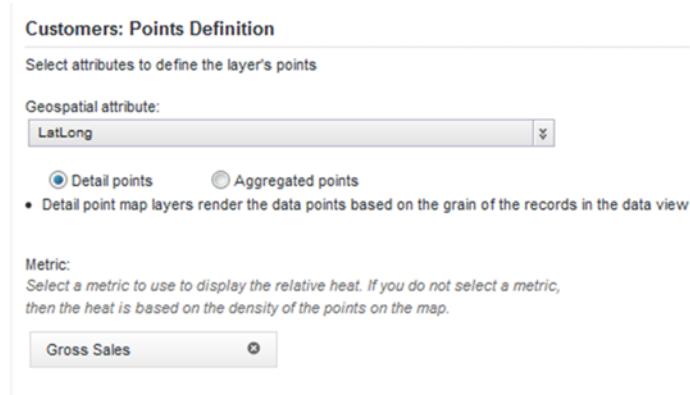
- (b) To determine the order in which to apply the aggregation, drag each dimension to the appropriate location in the list. The dimension at the top of the list is applied first.
- (c) To remove a dimension from the list, click its delete icon.

4. For a heat map layer, you can select a metric to use for setting the layer point and heat cloud colors.

For example, if the locations are stores, then the heat map layer colors could reflect the total sales at each store.

If you do not select a metric, then the heat map colors are based on the relative density of the points on the map. So if the locations are stores, then the heat map colors would change based on the number of stores in a given area.

To select the metric, click the **Select Metric** button. On the metric dialog, click the metric, then click **Apply**.



For predefined metrics or system metrics, the aggregation method is built in.

For other attributes, the default aggregation method is assigned. You can then use the drop-down list to select a different aggregation method.

To select a different metric, click the delete icon, then click **Select Metric**.

### Configuring the layer name and display options

On the **Layer Properties** tab for the map layer, you configure the layer name and options for displaying the map locations, and, for heat map layers, the cloud overlay.

On the **Layer Properties** tab:

1. In the **Layer name** field, type the name of the layer.

The layer name is used to identify the layer on the map legend.  
For numbered point layers, the name is also used in the layer drop-down list on the numbered point list.
2. If the view contains multiple geocode attributes, then from the **Geo filtering attribute** drop-down list, select the attribute to use when searching or filtering the locations.

By default, the map layer uses the same geocode attribute for both the map points and the filtering.
3. For point layers and heat map layers, from the **Size of points** drop-down list, select the size of the map location points.
4. For numbered point and point map layers, from the **Layer color** drop-down list, select the color to use to display the points.

5. For a heat map layer, to configure the color and metric value range:

<b>Color Options</b> <p>Layer color: </p> <p><input checked="" type="radio"/> Dynamically determine min / max values on color ramp</p> <p><input type="radio"/> Fixed values on color ramp</p> <p>Minimum: <input type="text"/></p> <p>Maximum: <input type="text"/></p>	<b>Heat Options</b> <p><input type="radio"/> Show heat points by default</p> <p><input type="radio"/> Show heat cloud by default</p> <p><input checked="" type="radio"/> Show both by default</p> <p><input checked="" type="checkbox"/> Enable end user to adjust heat range</p>
---	---

- (a) From the **Layer color** drop-down list, select the color range to use for the map points and cloud.
- (b) To automatically calculate the minimum and maximum metric values for the color range, click the **Dynamically determine min/max values on color ramp** radio button.
- (c) To specify the minimum and maximum values, click the **Fixed values on color ramp** radio button, then in the **Minimum** and **Maximum** fields, type the minimum and maximum values.

6. For a heat map layer, under **Heat Options**:

- (a) Click the radio button to indicate whether to display by default the location points, heat cloud, or both.
- (b) To allow end users to adjust the heat range, check the **Enable end user to adjust heat range** checkbox.

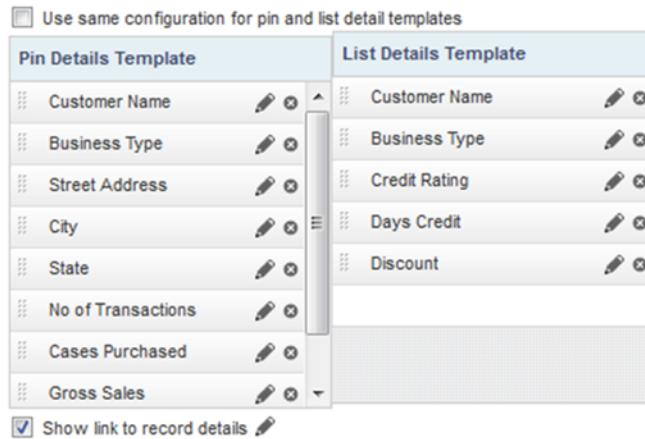
### Selecting and configuring the details to display for each map point

On the **Details Template** tab for the map layer, you configure the information to display when users click a point on the map. For detail points, you can also determine whether to include a separate link to display record details.

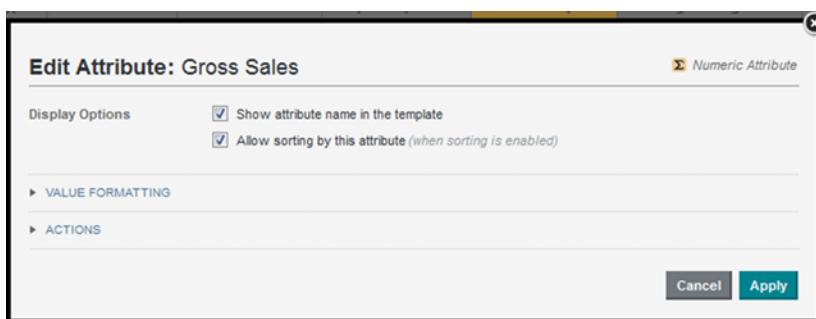
<b>Available attributes</b> <p><input type="button" value="Filter..."/></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Business Type</li> <li><input type="checkbox"/> Cases Purchased</li> <li><input type="checkbox"/> City</li> <li><input type="checkbox"/> Credit Rating</li> <li><input type="checkbox"/> Customer ID</li> <li><input type="checkbox"/> Customer Name</li> <li><input type="checkbox"/> Days Credit</li> <li><input type="checkbox"/> Discount</li> <li><input type="checkbox"/> Gross Sales</li> </ul>	<p><input checked="" type="checkbox"/> Use same configuration for pin and list detail templates</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th colspan="2" style="text-align: left; padding: 2px;">Pin Details Template</th> </tr> <tr> <td style="width: 15%; padding: 2px;"><input type="checkbox"/> Customer Name</td> <td style="width: 15%; text-align: right; padding: 2px;"><input type="button"/> <input type="button"/></td> </tr> <tr> <td><input type="checkbox"/> Business Type</td> <td style="text-align: right; padding: 2px;"><input type="button"/> <input type="button"/></td> </tr> <tr> <td><input type="checkbox"/> Street Address</td> <td style="text-align: right; padding: 2px;"><input type="button"/> <input type="button"/></td> </tr> <tr> <td><input type="checkbox"/> City</td> <td style="text-align: right; padding: 2px;"><input type="button"/> <input type="button"/></td> </tr> <tr> <td><input type="checkbox"/> State</td> <td style="text-align: right; padding: 2px;"><input type="button"/> <input type="button"/></td> </tr> <tr> <td><input type="checkbox"/> No of Transactions</td> <td style="text-align: right; padding: 2px;"><input type="button"/> <input type="button"/></td> </tr> <tr> <td><input type="checkbox"/> Cases Purchased</td> <td style="text-align: right; padding: 2px;"><input type="button"/> <input type="button"/></td> </tr> <tr> <td><input type="checkbox"/> Gross Sales</td> <td style="text-align: right; padding: 2px;"><input type="button"/> <input type="button"/></td> </tr> </table> <p><input checked="" type="checkbox"/> Show link to record details <input type="button"/></p>	Pin Details Template		<input type="checkbox"/> Customer Name	<input type="button"/> <input type="button"/>	<input type="checkbox"/> Business Type	<input type="button"/> <input type="button"/>	<input type="checkbox"/> Street Address	<input type="button"/> <input type="button"/>	<input type="checkbox"/> City	<input type="button"/> <input type="button"/>	<input type="checkbox"/> State	<input type="button"/> <input type="button"/>	<input type="checkbox"/> No of Transactions	<input type="button"/> <input type="button"/>	<input type="checkbox"/> Cases Purchased	<input type="button"/> <input type="button"/>	<input type="checkbox"/> Gross Sales	<input type="button"/> <input type="button"/>
Pin Details Template																			
<input type="checkbox"/> Customer Name	<input type="button"/> <input type="button"/>																		
<input type="checkbox"/> Business Type	<input type="button"/> <input type="button"/>																		
<input type="checkbox"/> Street Address	<input type="button"/> <input type="button"/>																		
<input type="checkbox"/> City	<input type="button"/> <input type="button"/>																		
<input type="checkbox"/> State	<input type="button"/> <input type="button"/>																		
<input type="checkbox"/> No of Transactions	<input type="button"/> <input type="button"/>																		
<input type="checkbox"/> Cases Purchased	<input type="button"/> <input type="button"/>																		
<input type="checkbox"/> Gross Sales	<input type="button"/> <input type="button"/>																		

To configure the map location details:

1. For a numbered point layer, you can configure a different set of details to display in the numbered points list and when users click the point on the map. To enable the different sets of details, uncheck the **Use same configuration for pin and list detail templates** checkbox.



2. To add an item to a details template, drag it from the **Available attributes** list to the template. See [Selecting the attributes to use on a component on page 90](#).
3. To determine the display order of the template items, drag each item to the appropriate location in the list.
4. To remove an item from a template, click its delete icon.
5. To configure the item display, click its edit icon. On the configuration dialog:
  - (a) For a metric other than a system or predefined metric, from the **Aggregation Function** drop-down list, select the aggregation method to use.
  - (b) For a date/time attribute, from the **Date/time Subset** drop-down list, select the subset of date/time units to display (for example, Year, Year-Month, Year-Month-Day).
  - (c) To include the attribute display name as well as the value, check the **Show attribute name in the template** checkbox.

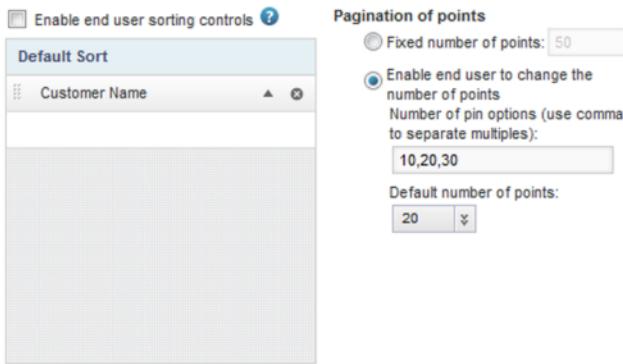


- (d) To allow users to sort by the attribute, check the **Allow sorting by this attribute** checkbox.
- (e) Under **Value Formatting**, configure the display format for the value. See [Configuring the format of values displayed on a component on page 93](#).

- (f) Under **Actions**, configure the action for when the user clicks the value. For information on configuring actions for displayed values, see [Configuring actions for displayed values on page 100](#).
- (g) To save the configuration, click **Apply**.
6. If the points are not aggregated, then to display a link to record details, check the **Show link to record details** checkbox. For aggregated points, you cannot display a link to record details, and the checkbox is not displayed.

## Configuring pagination and sorting for the map layer

On the **Sorting and Pagination** tab for the map layer, you configure the available sort options and pagination for the layer.



To configure the sorting and pagination:

1. To allow end users to change the sort order, check the **Enable end user sorting controls** checkbox. If this checkbox is checked, then end users can sort the map using any of the values in the details template.
2. To configure the default sort order:
  - (a) To select the attributes to include in the default sort, drag the attributes from the **Available attributes** list to the **Default Sort** drop zone.
  - (b) To determine the order in which to apply the sort, drag each attribute to the appropriate location in the list.
  - (c) For each sorting attribute, use the sort order icon to indicate whether to sort in ascending or descending order.
  - (d) To remove an item from the default sort, click its delete icon.
3. To always display the same fixed number of locations on the map:
  - (a) Click the **Fixed number of points** radio button.
  - (b) In the field, enter the number. For a numbered points layer, this is the number of points to display per page. Users can then page through the list.
4. To allow end users to change the number of locations to display:
  - (a) Click the **Enable end user to change number of points** radio button.

- (b) In the **Number of pin options** field, type a comma-separated list of options for the number of locations to display.
- (c) From the **Default number of points** drop-down list, select the default number of locations to display.

## Determining which map layers are visible

If you are using the default map layer configuration, then all of the automatically generated map layers are visible on the end user view of the **Map** component. If you are not using the default configuration, then you can control whether the layer is visible.

On the **Map Layers** tab, to not display the map layer on the end user view of the **Map** component, uncheck the layer's checkbox.

## Configuring general display settings for the Map

On the **General Settings** tab of the **Map** component edit view, you configure general settings for the component as a whole. These settings are not specific to a map layer.

On the **General Settings** tab:

1. Under **Filtering**:

**Filtering**

- Enable geospatial filtering via marquee**  
 **Enable geospatial range filtering via text search**

- (a) To allow users to search the map using the range search, check the **Enable geospatial range filtering via marquee** checkbox.
- (b) To allow users to search the map using the location text search, check the **Enable geospatial filtering via text search** checkbox.

2. Under **Display Settings**:

**Display Settings**

- Show the numbered list overlay panel by default when numbered point layers are visible on the map**

Map height:  pixels

Units of measure:  

- (a) If the map includes a numbered pin layer, then to display the numbered pin list by default when the component is first displayed, check the **Show numbered list overlay panel by default** checkbox.
- (b) In the **Map height** field, type the height in pixels for the **Map** component.

- (c) From the **Units of measure** drop-down list, select the unit of measurement to use for the map.

To use the preferred unit of measurement based on the current locale, select **use locale default**.

Otherwise, you can have the map always use either miles or kilometers.

## Pivot Table

The **Pivot Table** component is a more complex table containing calculated and grouped values.

[About the Pivot Table](#)

[Using the Pivot Table](#)

[Configuring the Pivot Table](#)

### About the Pivot Table

The **Pivot Table** component allows users to perform comparisons and identify trends across several cross sections of data.

The values in the header rows and columns represent every possible grouping of the selected data fields.

Each body cell contains a metric value that corresponds to the intersection of the values in the heading rows and columns. In the following example, each cell contains either the average price or average score for a specific combination of region, designation, and vintage.

		Vintage ▲		1992		1993		1994			
Region	Designation	Price(averag...	Review Sco...	Price(averag...							
Alsace	Auslese			\$38.00	90.00						
	Best Buy	\$9.00	88.00	\$9.00	88.00	\$10.00	85.00				
	Botrytis	\$19.00	87.00					\$185.00	90.00		
	Brut	\$20.00	85.00	\$15.00	85.00						
	Cellar Select	\$43.50	92.50	\$40.00	92.00						
	Cellar Selecti...	\$43.50	92.50	\$40.00	92.00						
	Cremant	\$20.00	85.00	\$15.00	85.00						
	Cuvee	\$22.45	84.85	\$27.60	86.07	\$31.29	86.00				
	Grand Cru	\$27.60	85.88	\$42.88	86.47	\$45.76	86.51				
	Highly Reco...			\$34.00	90.67						
	Late Harvest	\$30.00	87.00								
	Reserve	\$13.61	81.89	\$16.00	84.40	\$29.00	84.25				
	Rose	\$20.50	88.50	\$56.29	87.14	\$40.77	88.23				
	Sec										

Cells may also be highlighted based on the displayed value. In the example, values are highlighted in red if they are below a certain number, and in green if they are above a certain number.

## Using the Pivot Table

In addition to viewing the **Pivot Table** data, you may be able to change the table layout.

For information on using common component functions such as paging, exporting, and printing, see [Using Common Component Functions on page 133](#).

### Navigating through Pivot Table data

You use the scroll bars to scroll through the **Pivot Table** rows and columns. You can also use the scroll bars to jump to a location.

When you hover the mouse over a location on a scroll bar, Big Data Discovery calculates and displays the dimension values for that relative location in the **Pivot Table**.



For the vertical scroll bar, the row dimension values are shown. For the horizontal scroll bar, the column dimension values are shown.

If you then click the scroll bar, the **Pivot Table** jumps to that location.

### Displaying summaries and highlighting in the Pivot Table

The **Pivot Table** component can include summaries of individual row or column dimensions as well as summary rows and columns for the entire table. Individual values may also be highlighted. The **Pivot Table** may be configured to allow you to show or hide the summaries, and to show or hide the highlighting.

If you can control these features, then a **View Options** button is displayed at the top of the table.

To enable and disable these options:

1. Click the **View Options** button.  
A drop-down list of checkboxes is displayed. There are options for enabling and disabling the highlighting, the dimension summaries, and the table summaries.
2. To show the value highlighting, check the **Conditional Formatting** checkbox.  
To hide the highlighting, uncheck the **Conditional Formatting** checkbox.
3. To show the summaries for each dimension, check the **Summaries** checkbox.  
To hide the dimension summaries, uncheck the **Summaries** checkbox.
4. To show the table summaries, which aggregate all of the values in a row or column, check the **Grand Summary** checkbox.  
To hide the table summaries, uncheck the **Grand Summary** checkbox.

## Sorting the Pivot Table dimension values

For each **Pivot Table** row and column dimension, you can determine the order in which to display the dimension values.

At the top left of the **Pivot Table** are the headings with the dimension names.



To switch the display order of the values for a dimension, click the dimension heading.

## Changing the layout of the Pivot Table

You may be able to change the layout of the **Pivot Table**, including the dimensions and metrics that are displayed and the row and column display order.

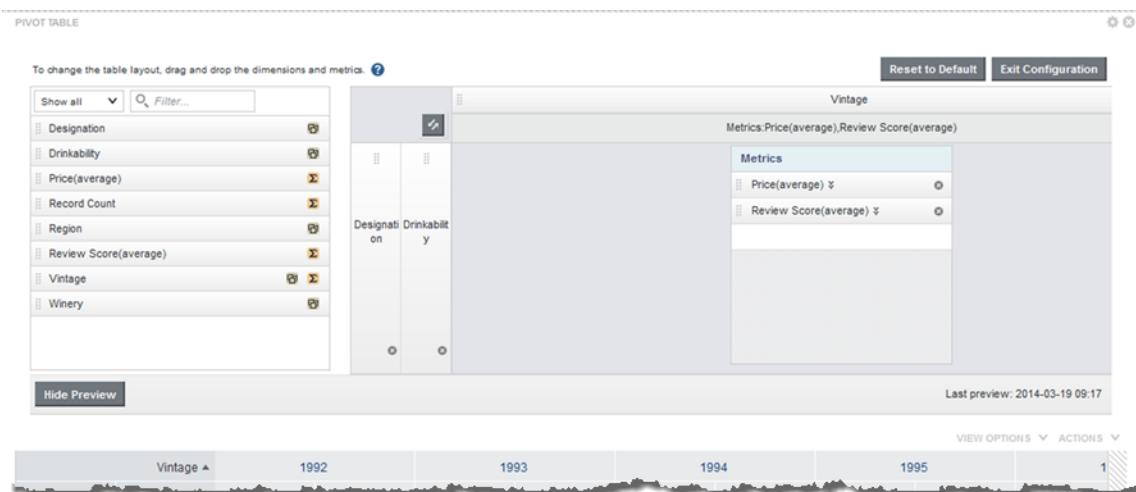
If the **Pivot Table** is configured to allow you to change the table layout, then a **Configure** button is displayed at the top of the component.

To change the layout of the **Pivot Table** table:

1. Click the **Configure** button.

On the configuration view, the list at the left shows the available metrics and dimensions that are not currently displayed on the table.

The rest of the configuration view shows a mockup of the displayed dimension columns and rows, with the list of displayed metrics in the middle. A metrics placeholder indicates whether the metrics are displayed as rows or columns.



2. To change the order of the displayed dimension rows or columns, drag the dimension to the new location.
  3. To remove a dimension or metric, click its delete icon.

The dimension or metric is removed from the table and added to the available list.
  4. To add a new dimension, drag the dimension from the available list to the appropriate location in the row or column groups.
  5. To add a new metric, drag the metric from the available list to the appropriate location in the metrics list.

If the metrics are displayed as columns, then the metric at the top of the list is displayed in the leftmost column.

If the metrics are displayed as rows, then the metric at the top of the list is displayed in the top row.
  6. To swap all of the rows and columns, including the metrics, click the swap icon in the top left corner of the layout.
  7. To only switch whether the metrics are displayed as rows or columns, click the swap icon in the metrics placeholder.
  8. To change the display order of the metric rows or columns, drag each metric to the appropriate location in the list.
  9. As you make changes, a preview of the resulting table is displayed below the configuration. It is updated automatically as you make changes.
    - (a) To hide the preview, click the **Hide Preview** button.
    - (b) To show the preview again, click **Show Preview**.
  10. To revert the layout back to the original default, click **Reset to Default**.
  11. To exit the configuration view, click **Exit Configuration**.
- The currently displayed layout is used.

## Using the Pivot Table to refine the project data

You can use the **Pivot Table** dimension values to refine the project data.

If a dimension allows refinement, then when you click the dimension values in the row and column headings, the data is refined by that value.

When you click a metric value, then the data is refined by all of the dimension values that apply to that cell and that allow refinement.

So for example, if a cell displays an average price of \$25.55 for Red wines in the Bordeaux region for the year 1992, then when you click the value \$25.55, the data is refined by the values:

- Red for the Wine Type attribute
- Bordeaux for the Region attribute
- 1992 for the Vintage attribute

For more details on refining by data on a component, including date hierarchies and cascading, see [Using a component to refine data on page 135](#).

## Configuring the Pivot Table

For a **Pivot Table** component, you select the view to use and configure how the table is displayed.

### Configuring the layout of the Pivot Table

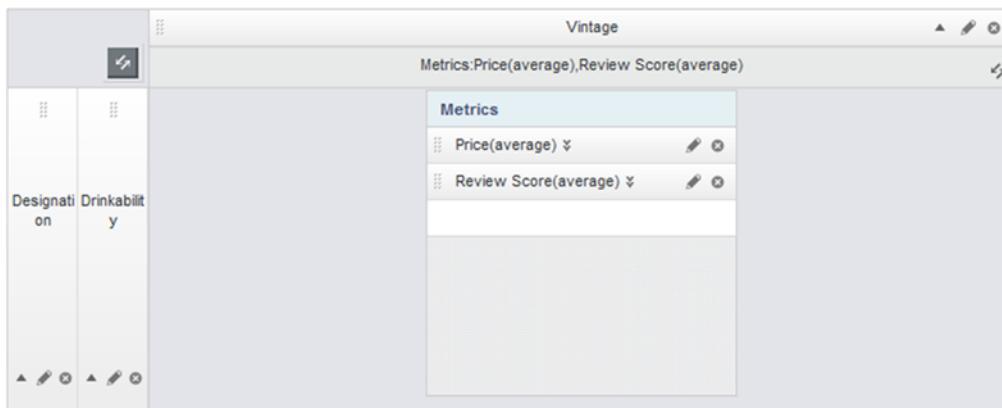
Once you select the view to use, you can configure the layout of the dimensions and metrics on the **Pivot Table** table.

#### About configuring the Pivot Table layout

You use the **Table layout** tab of the **Pivot Table** edit view to configure the layout of the **Pivot Table**.

The layout includes:

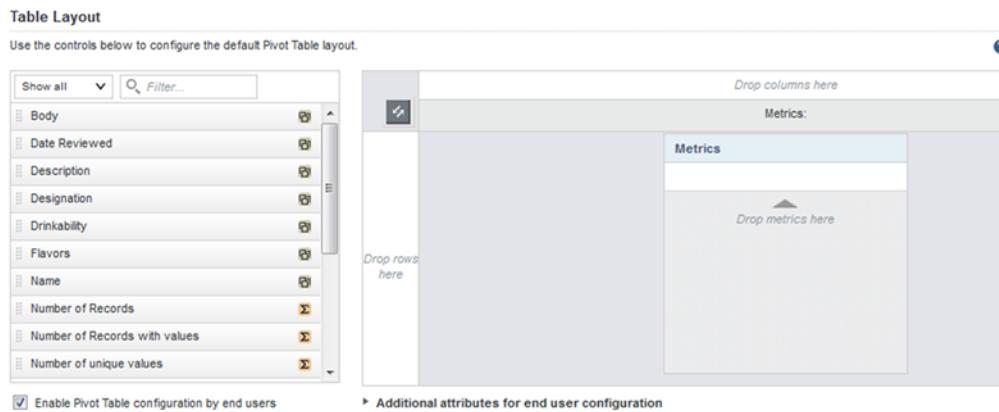
- The dimensions to use to aggregate the **Pivot Table**
- How to group the dimension values into nested rows and columns. The table must have at least one row-based dimension.
- The metrics to display on the **Pivot Table**
- Whether to display those metrics as rows or columns. By default, the metrics display as columns, with the headings in the last row of column headings.



## Adding dimensions and metrics to the Pivot Table

Each **Pivot Table** consists of dimension rows and columns to determine the aggregation, and metrics to determine the displayed values.

At the right of the **Table Layout** tab is a mockup of the **Pivot Table** format, with drop zones to add dimensions to the row and column groups, and to add metrics. A placeholder block indicates whether the metrics are displayed as columns or rows.



To populate the default view of the **Pivot Table**:

1. To add a dimension to a row or column group, drag the dimension from the list to either the row or column group drop zone.

You must add at least one dimension to the row group zone.

If there are existing row or column dimensions, you can drop the new dimension before or after those dimensions.

You can also drag and drop the dimensions within the mockup to change the order of the row and column groups.

For additional information on selecting values to display on a component, see [Selecting the attributes to use on a component on page 90](#).

2. To add a metric to the **Pivot Table**, drag the metric from the list to the metric drop zone.

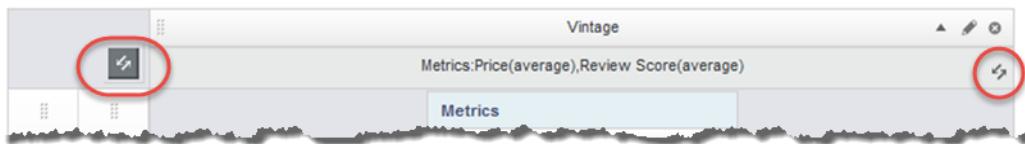
The order of the metrics in the list determines the order (from left to right for columns or top to bottom for rows) of the metrics in the **Pivot Table**. The placeholder also lists the selected metrics.

The screenshot shows the 'Vintage' Pivot Table edit view. The 'Metrics' placeholder on the right side of the table contains two items: 'Price(average)' and 'Review Score(average)'. Each item has a small edit icon to its right. The table itself has two columns labeled 'Designation' and 'Drinkability'.

► Additional attributes for end user configuration

### Changing the orientation of the Pivot Table

The **Table Layout** tab on the **Pivot Table** edit view includes options to switch whether dimensions and metrics are displayed as rows or columns.



To switch the **Pivot Table** row and column groups, click the swap icon at the top left of the mockup. The swap includes the metrics, so for example if your metrics were displayed as columns, when you click **Swap**, they are changed to display as rows.

To only change the location of the metrics, click the swap icon on the metrics placeholder.

### Adding dimensions and metrics

If you are allowing end users to change the configuration of the **Pivot Table**, you can provide additional available dimensions and metrics that are not displayed on the default view of the table.

On the **Table Layout** tab of the **Pivot Table** edit view, the **Available attributes for end user configuration** section below the **Pivot Table** mockup contains the lists of these available dimensions and metrics.

▼ Additional attributes for end user configuration

Dimensions	Metrics
Region	Number of Records
Winery	

To add the additional dimensions and metrics:

1. To expand or collapse the section, click the section heading.
2. To add additional available dimensions, drag the dimensions from the attributes list to the **Dimensions** list.
3. To add additional available metrics, drag the metrics from the attributes list to the **Metrics** list.

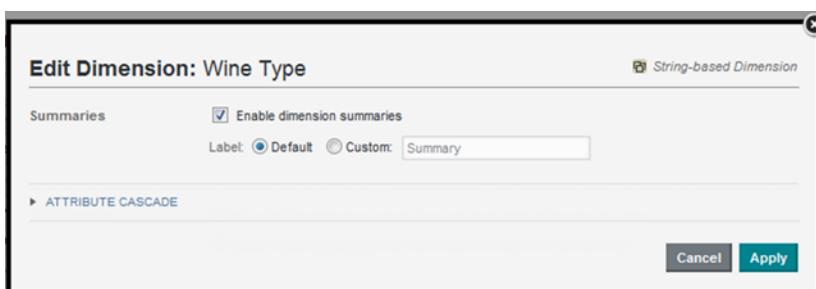
### Configuring a dimension

For each **Pivot Table** dimension, you can configure the label for the dimension, whether to display a summary row or column for the dimension, and rules for refining by the dimension value.

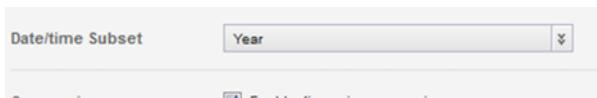
From the **Table Layout** tab of the **Pivot Table** edit view, to configure a dimension:

1. Click the edit icon for the dimension you want to configure.

The **Edit Dimension** dialog is displayed.



2. If the dimension is a date/time attribute, then from the **Date/time subset** drop-down list, select the default date/time subset to display.



By default, Big Data Discovery selects the largest date/time unit that is available for that attribute.

3. On the edit dimension dialog, to display a summary row or column for the dimension, check the **Enable dimension summaries** checkbox.
4. Under **Label**, by default, the summary label is "Summary", and the **Default** radio button is selected. To provide a custom label, click the **Custom** radio button, then type the new label into the field.
5. Use the **Attribute Cascade** section to configure cascading for the dimension. For details on dimension cascading and how to configure it, see [Configuring cascading for dimension refinement on page 104](#).
6. To save the dimension configuration, click **Apply**.

### Configuring a metric

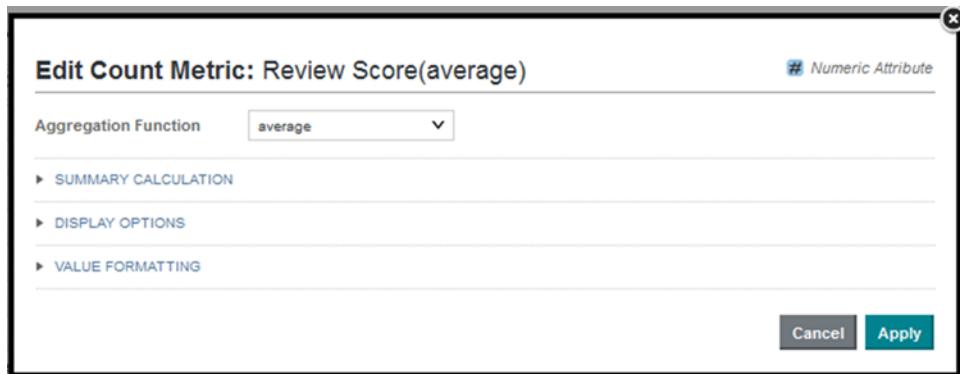
For a **Pivot Table** metric, you can configure the aggregation method for the metric, the tooltip for summary values, and the format of the metric value.

For information on selecting the aggregation method for a metric, see [Selecting the aggregation method to use for a metric on page 93](#).

To configure all of the options for a metric:

1. On the **Table Layout** tab, click the edit icon for the metric you want to edit.

The **Edit Count Metric** dialog is displayed.



2. The **Summary Calculation** section contains the **Summary Tooltip** field, used to add a description to the tooltip displayed when users hover over the mouse over a metric value in a summary row or column.



3. The **Display Options** section contains a setting to determine the horizontal alignment of the value in the metric column. When the dialog is first displayed, the section is collapsed.

To expand or collapse the section, click the section heading.



To select the alignment, click the radio button.

4. The **Value Formatting** section contains settings to customize the format of the metric values.

For details on formatting displayed values, see [Configuring the format of values displayed on a component on page 93](#).

5. To save the metric configuration, click **Apply**.

### Allowing end users to configure the table layout

You can configure the **Pivot Table** to allow end users to change the layout.

On the **Table Layout** tab of the **Pivot Table** edit view, to allow end users to change the layout, check the **Enable Pivot Table configuration by end users** checkbox.



## Highlighting Pivot Table metric values that fall within a specified range

The **Conditional Formatting** tab of the **Pivot Table** edit view allows you to highlight specific metric values.



To configure whether to use conditional formatting, and select the values to highlight:

1. To allow conditional formatting, make sure that the **Enable conditional formatting** checkbox is checked. It is checked by default.  
If the checkbox is not checked, then there is no conditional formatting.
2. In the **Metrics** list, click the metric for which to configure conditional formatting.  
For each metric, the list displays the number of conditional formatting rules created for it.
3. To add a new range of values to highlight for that metric, click **+Add Condition**.
4. To configure a condition:
  - (a) From the **Condition Rule** type drop-down list, select the type of comparison.
  - (b) In the field (or fields, for the **between** and **a range from** options), enter the value(s) against which to do the comparison.  
Note that for the **between** option, the values are inclusive. So if you specify a range between 20 and 30, values of 20 and 30 also are highlighted.
  - (c) From the **Color** drop-down list, select the color to use for the highlighting.
  - (d) In the **Tooltip Description** field, type the text to display in the tooltip for a highlighted cell.
5. The conditions are applied based in the order they are listed. So a condition at the top of the list has a higher priority than a condition lower in the list.  
To change the priority of the conditions, drag and drop them to the appropriate location in the list.
6. To remove a condition, click its delete icon.

## Configuring display options for the Pivot Table

The **Display Options** tab of the **Pivot Table** edit view contains options for setting size and other display options for the **Pivot Table** table.

To configure the **Pivot Table** display options:

1. Under **Table height**, in the field, type the number of rows that are visible at a time.
2. In the **Column width** field, type the width in pixels to use for each column.
3. Under **Grid display**, to hide any empty rows or columns, check the **Hide empty rows/columns** checkbox.
4. Under **Summaries**, use the checkboxes to indicate whether to display the summary and grand summary rows by default.

## Summarization Bar

The **Summarization Bar** displays one or more summary values from the project data.

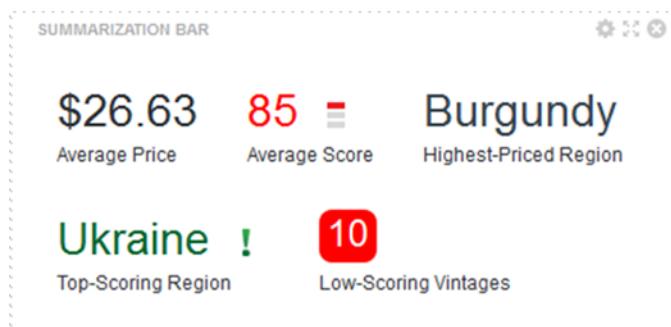
[About the Summarization Bar](#)

[Using the Summarization Bar](#)

[Configuring the Summarization Bar](#)

### About the Summarization Bar

The **Summarization Bar** allows users to quickly view metric or dimension values that summarize aspects of the underlying data.



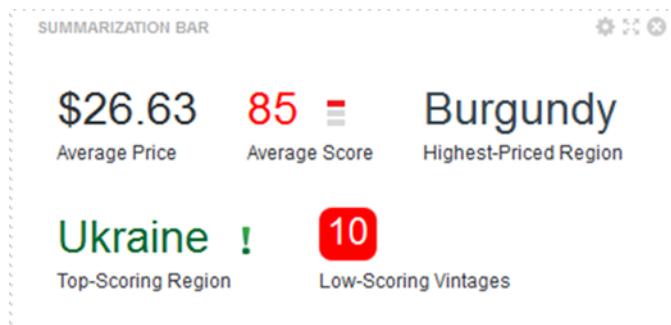
### Using the Summarization Bar

The **Summarization Bar** displays one or more summary items. You may be able to display additional information about an item, or use an item to refine the data or navigate to another page or site.

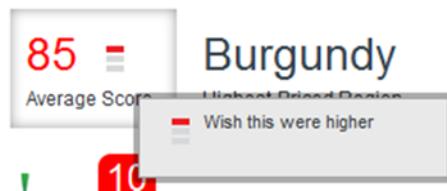
Each summary item contains one of the following types of values:

Summary Item	Value Description
Metric value	The actual value of a specific metric. For example, an item could display the total sales or average profit.
Dimension value	The dimension value associated with either the top or bottom value of an associated metric value. For example, an item could display the product category that has the highest total sales.
Number of flags	The number of flags. Flags are based on the value of a metric for a selected dimension or dimensions. For example, a flag summary item could reflect the number of product categories for which the total number of sales is greater than 500.

**Summarization Bar** values can use icons and color highlighting to provide additional context for that value.



You can also display a tooltip containing a description and value details.



For flags, you can click the number to display the list of values.



You also may be able to use summary item values to navigate to a different page or URL. For dimension values, you may also be able to use the value to refine the data. See [Using a component to refine data on page 135](#).

## Configuring the Summarization Bar

For a **Summarization Bar** component, you create and configure the summary items to display.

### Default configuration of a Summarization Bar

When you first add a **Summarization Bar** component to a project page, it automatically displays a single metric value summary item that shows the number of records in the base view of the first data set.

The screenshot shows the 'Summary Items' tab in the Summarization Bar edit view. At the top, there is a note: 'Create and manage the summary items to display on the component. To set the display order, drag and drop the items.' Below this is a 'plus item' button and a dropdown menu. A single summary item is listed: '# Record Count'.

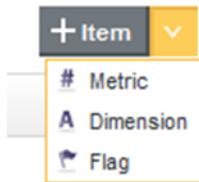
### Adding, editing, and deleting summary items

On the **Summarization Bar** edit view, the **Summary Items** tab displays the current summary items.

The screenshot shows the 'Summary Items' tab in the Summarization Bar edit view. It displays five summary items: '# Average Price', '# Average Score', 'A Highest-Priced Region', 'A Top-Scoring Region', and 'P Low-Scoring Vintages'. Each item has a delete icon to its right. The 'plus item' button is at the top right.

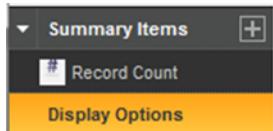
To add, edit, and delete summary items:

1. To add a new summary item to the **Summarization Bar**, either:
  - Click the **+Item** drop-down icon, then in the drop-down list, click the type of summary item to add.



For details on the summary item types, see [Selecting the type of summary item on page 324](#).

- In the edit view menu, click the **+** button next to the **Summary Items** entry in the edit view menu.



A new item is added to the list.

If you selected the item type when you added the item, then the **Select Data** tab is displayed. See [Selecting the view to use for a component on page 89](#).

If you did not select an item type, then by default the item is a metric item, and the **Select Item Type** tab is displayed. See [Selecting the type of summary item on page 324](#).

2. To edit the configuration of a summary item, click the item name.
3. To edit the name of a summary item:
  - (a) On the **Summary Items** tab, double-click the item name.  
The name is displayed as an editable field.
  - (b) Type the new name, then press **Enter**.
4. To return to the **Summary Items** list after editing a summary item, click the **Summary Items** option in the left menu.
5. To determine the order in which to display the items on the end user view, drag each item to the appropriate location.
6. To delete a summary item, click its delete icon.

## Selecting the type of summary item

On the **Select Item Type** tab for a summary item configuration, you select the type of summary item you want to create.

To select the item type to use, click it.

The options are:

Summary Item Type	Description
<b>Metric Value</b>	Displays a single aggregated metric value, such as total sales or average score.
<b>Dimension Value Spotlight</b>	Displays a single dimension value. The displayed dimension value is associated with the top or bottom value for a single metric.  For example, you could display the name of the product category that has the highest or lowest total sales.

Summary Item Type	Description
<b>Flag</b>	<p>Displays the number of values for a selected dimension or combination of dimensions that meet a set of metric conditions.</p> <p>For example, you could display the number of combinations of product category and sales region that had more than 500 sales and for which the percentage of total sales was at least 90% of expected sales.</p>

## Configuring metric and dimension value spotlight summary items

For metric and dimension spotlight summary items, you configure the values, how to display them, and whether to apply conditional formatting.

### Configuring the metric for a metric value summary item

For a metric summary value, on the **Define Item** tab of the **Summarization Bar** edit view, you select the metric value to display.

**Average Price**

Define the metric value to display on this summary item.

Metric:  ▼ ✎ ✖

Display Name:

Use metric name:  
Sales Price (avg)

Custom name:  
Average Price

▶ DISPLAY OPTIONS

▶ ACTIONS

To select and configure the summary item value for a metric summary item:

1. To select the metric:
  - (a) Click the **Select Metric** button.

**Average Price**

Define the metric value to display on this summary item.

Metric: Select Metric...

- (b) On the **Select a Metric** dialog, click the metric you want to use. You can use the filter field to search for a specific metric.
- (c) Click **Apply**.

2. To select a different aggregation method for a metric, you can use the drop-down list next to the metric name.

You cannot select a different aggregation method for a predefined metric.

3. To select a different metric:

- (a) Click the delete icon for the current metric.
- (b) Click the **Select Metric** button.
- (c) On the **Select a Metric** dialog, click the metric you want to use.
- (d) Click **Apply**.

4. To configure the metric:

- (a) Click its edit icon.
- (b) For metrics that are not predefined, the configuration dialog includes a drop-down list to select the aggregation method.
- (c) The **Value Formatting** section allows you to customize the format of the displayed value.

For details on customizing a displayed value on a component, see [Configuring the format of values displayed on a component on page 93](#).

### Configuring the values for a dimension value spotlight summary item

For a dimension value spotlight summary item, on the **Define Item** tab of the **Summarization Bar** edit view, you select the dimension and metric to use.

**Highest-Priced Region**

Select the dimension to display, and the metrics to use to determine the dimension value.

Dimension:   

Top value  Bottom value

Metric:   

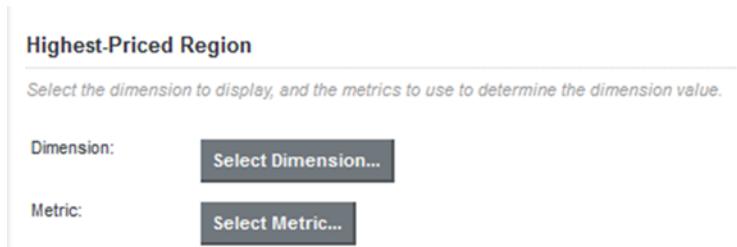
Display Name:  Use dimension & metric names:  
Top Region by Sales Price (avg)  
 Custom name:





To select the summary item values:

1. To select the dimension for which to display a value:
  - (a) Click the **Select Dimension** button.



- (b) On the **Select a Dimension** dialog, click the dimension you want to use.
  - (c) Click **Apply**.
2. To display the dimension value with the highest metric value, click the **Top Value** radio button.  
To display the dimension value with the lowest metric value, click the **Bottom Value** radio button.
3. To select a different dimension:
  - (a) Click the delete icon for the current dimension.
  - (b) Click the **Select Dimension** button.
  - (c) On the **Select a Dimension** dialog, click the dimension you want to use.
  - (d) Click **Apply**.
4. To select the metric value:
  - (a) Click the **Select Metric** button.
  - (b) On the **Select a Metric** dialog, click the metric you want to use. You can use the filter field to find a specific metric.
  - (c) Click **Apply**.
5. To select a different aggregation method for the metric, you can use the drop-down list next to the metric name.  
You cannot select a different aggregation method for a predefined metric.
6. To select a different metric:
  - (a) Click the delete icon for the current metric.
  - (b) Click the **Select Metric** button.
  - (c) On the **Select a Metric** dialog, click the metric you want to use.
  - (d) Click **Apply**.
7. To configure the metric or dimension:
  - (a) Click its edit icon.
  - (b) For date/time dimensions, from the **Date/time Subset** drop-down list, select the date/time subset to display on the component.  
By default, the largest available subset is used.

- (c) For metrics that are not predefined, the configuration dialog includes a drop-down list to select the aggregation method.
- (d) For both the dimension and the metrics, the **Value Formatting** section allows you to customize the format of the displayed value.

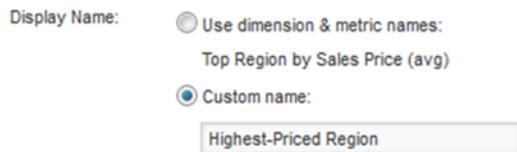
For details on customizing a displayed value on a component, see [Configuring the format of values displayed on a component on page 93](#).

### Configuring the display and actions

For metric and dimension value spotlight summary items, the summary item includes a label above or below the displayed metric or dimension value. You can also configure the display options and action for the summary item.

On the **Summarization Bar** edit view, on the **Define Item** tab for the summary item:

1. The **Display Name** setting determines the label displayed below the metric or dimension value.



You can either display the dimension and metric names, or use a custom name. The custom name is also used as the name of the summary item.

2. Under **Display Options**, use the **Tooltip** setting to provide a description of the summary item.



In the tooltip text, use the `{MetricValue}` token to represent the metric value.

3. The **Width** setting determines the maximum width of the summary item block.

You can use the default maximum width, or set a custom maximum width.

If the value doesn't fit within the maximum width of the summary item, then Big Data Discovery truncates it.

4. For a metric summary item, you can allow users to click the metric value in order to navigate to a different page in the project or to an external URL.

To enable the hyperlink, you must at least provide the page name or URL. See [Configuring actions for displayed values on page 100](#).

5. For a dimension value spotlight item, you can allow users to click the dimension value in order to either refine by the value, or to navigate to a different page in the project or to an external URL.

Under **Actions**, from the **Type** drop-down list, select the action. For details on configuring an action from a component value, see [Configuring actions for displayed values on page 100](#).

## Configuring conditional formatting

For a metric or dimension spotlight value summary item, you can configure different formatting based on the metric value.

Display	When value is...	Description
	less than <input type="text" value="85"/>	Wish this were higher
	greater than <input type="text" value="90"/>	Great score!

**DISPLAY OPTIONS**

Set summary item text to black for all conditions

Only display this summary item if it meets a condition

On the **Conditional Formatting** tab for a summary item, to configure the conditional formatting:

1. Select the type of value to use to determine whether the condition is met:

- (a) Click the radio button to indicate the value to use.

The options are:

Comparison Type	Description
<b>Metric value</b>	Indicates to apply the conditional formatting based on the actual summary item value.
<b>Percentage of another metric</b>	Indicates to apply the conditional formatting based on the ratio of the summary item metric value to another metric.  For example, if a summary item displays total sales, then you could configure formatting based on the ratio of total sales to expected sales.  So you might configure the total sales to display in green if they are greater than or equal to 100% of the expected sales.

- (b) For a ratio to another metric, to select the other metric, click the **Select Metric** button.

- (c) On the **Select a Metric** dialog, click the metric you want to use, then click **Apply**.

2. From the **Style** drop-down list, select the style to use for the conditional formatting.

The style determines the available set of icons to display on the summary item.

3. To add a condition to the list, click the **+Condition** button.

4. To configure a condition:

- (a) From the **Display** drop-down list, select the icon to display when the condition is met.

Note that you can have multiple rows that use the same display. For example, you may want to use the same formatting to highlight both very low and very high outlying values.

- (b) Under **When value is**, select the comparison to use, then enter the appropriate values.
  - (c) In the **Description** column, type a description to display in the summary item tooltip when the condition is met.
5. To remove a condition, click its delete icon.
  6. In the **Display Options** section:
    - (a) By default, the text color on the summary item matches the icon color, if color is used. To always display the summary item value in black, check the **Set summary item text to black for all conditions** checkbox.
    - (b) To only display the summary item if the value meets one of the conditions, check the **Only display this summary item if it meets a condition** checkbox.
- For example, if you create a condition for total sales for when the value is less than 1,000,000, then if this checkbox is checked, the summary item does not display at all when total sales are greater than or equal to 1,000,000.

## Configuring flag summary items

For a flag summary item, on the **Define Flags** tab of the **Summarization Bar** edit view, you select the dimensions and set up the metric conditions.

If you select multiple dimensions, then the conditions are applied to combinations of those dimension values.

For example, if you select the product category and region dimensions, and set the flag condition to total sales greater than 500, then the flag summary item is displayed if any combination of product category and region has total sales greater than 500.

To configure a flag summary item:

1. In the **Display Name** field, type the text to display on the summary item.

The screenshot shows a configuration dialog for a flag summary item named 'Low-Scoring Vintages'. The 'Display Name' field contains 'Low-Scoring Vintages'. The 'Flag Color' dropdown is set to red. Under 'Select Dimension(s)', the dimension 'Vintage' is listed. There are buttons for 'Edit' and 'Delete' next to the dimension list.

The display name is also used as the name of the summary item.

2. From the **Flag Color** drop-down list, select the background color to use for the number of values that match the flag criteria.
3. To add a dimension to the flag summary item:
  - (a) Click the **+ Dimension** button.
  - (b) On the **Select a Dimension** dialog, click the dimension you want to use.

(c) Click **Apply**.

If you select a date/time attribute as a flag dimension, then by default, Big Data Discovery uses the largest available date/time subset. For example, if the attribute has Year, Year-Month, and Year-Month-Day enabled, then the Year subset is used.

4. The dimension values are applied in the order they are listed. To change the order, drag each dimension to the appropriate location in the list.
5. To remove a dimension from the list, click its delete icon.
6. Use the **Define Conditions** list to select the list of conditions that must be met in order for the flag summary item to display.

A flag summary item is only displayed on the end user view if at least one value matches all of the conditions in the list.

Define Conditions

Define the conditions (metrics and thresholds) that determine when a dimension value is flagged. A dimension value is only flagged if ALL of the conditions are met. To change the order of the conditions, drag them up or down in the list. [?](#) [+ Condition](#)

Condition based on	Metric	Threshold
Metric value	Base Score (avg)	less than 90

- (a) To add a condition to the list, click the **+Condition** button.
- (b) To remove a condition from the list, click its delete icon.
- (c) To change the order in which to apply the conditions, drag and drop the conditions to the appropriate location in the list.
7. For each condition in the **Define Conditions** list:
  - (a) From the **Condition Based On** drop-down list, select whether the condition is based on:
    - A single metric value. For example, display the flag summary item if any of the dimension values have more than 500 sales.
    - A ratio of two metrics. For example, display the flag summary item if the total sales for any of the dimension values is greater than 90% of the projected sales.
  - (b) To select a metric, click the **Select Metric** button.
  - (c) To select a different aggregation method for a metric, you can use the drop-down list next to the metric name. You cannot select a different aggregation method for a predefined metric.
  - (d) In the **Threshold** column, configure the condition comparison. From the drop-down list, select the comparison operator, then in the field or fields, type the comparison value or values.
8. To configure a dimension or metric:
  - (a) Click its edit icon.
  - (b) For date/time dimensions, from the **Date/time Subset** drop-down list, select the date/time subset to display on the component.  
By default, the largest available subset is used.
  - (c) For metrics that are not predefined, the configuration dialog includes a drop-down list to select the aggregation method.

- (d) For both metrics and dimensions, the **Value Formatting** section allows you to customize the format of the displayed value.

For details on customizing a displayed value on a component, see [Configuring the format of values displayed on a component on page 93](#).

9. Under **List of flags**:

List of Flags

Max number of flags to display:

Overview content:

- (a) In the **Max number of flags to display** field, type the maximum number of flags to display when users click the flag summary item.

For example, even if 20 dimension values match the flag conditions, you can restrict the list of flags to only show 10.

- (b) In the **Overview content** field, type a brief description to display above the list.

10. Under **Display Options**:

DISPLAY OPTIONS

Width: Summary items are truncated if they don't fit within the defined maximum width.

Default max width (350 pixels)  Custom max width  pixels

Display name for column of dimensions:

Use default name  Custom name

Select metrics to display as columns:

Use default name  Custom name

- (a) Under **Width**, to customize the maximum width of the item, click the **Custom max width** radio button, then in the field, type the width in pixels.

- (b) Under **Display name for column of dimensions**, specify the column heading for the dimension column on the list of flags.

By default, the dimension names are used. To specify a custom name, click the **Custom name** radio button, then in the field, type the custom heading.

- (c) Under **Select metrics to display as columns**, to include the metric values in the list of flags, check the checkbox.

- (d) If you are displaying the metrics, then to customize the column heading for the metric values, click the **Custom name** radio button, then in the field, type the custom heading.

11. In the **Actions** section, configure the page to display when end users refine by a dimension value in the list of flags.

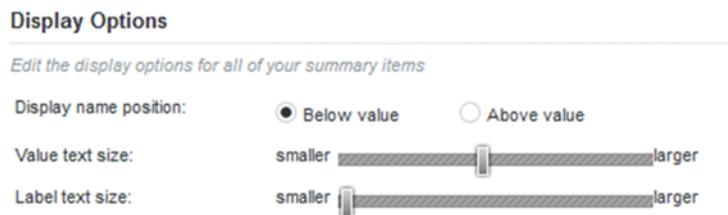
For information on configuring a target page for refinement, see [Selecting the target page for a refinement or hyperlink on page 102](#).

## Configuring display options for the Summarization Bar display

The **Display Options** tab of the **Summarization Bar** edit view contains settings to control the label position and text size for all of the summary items.

To configure the display options:

1. On the **Summarization Bar** edit view, click the **Display Options** tab.



2. Under **Display name position**, click a radio button to indicate whether to display the summary item display name above or below the summary item value.
3. Under **Value text size**, use the slider to adjust the size in pixels of the summary item value. You can set the size anywhere between 11 and 48 pixels. The default size for the summary item value is 28 pixels.
4. Under **Label text size**, use the slider to adjust the size in pixels of the summary item display name. You can set the size anywhere between 11 and 48 pixels. The default size for the summary item display name is 12 pixels.

## Tag Cloud

The **Tag Cloud** component displays a collection of terms from a selected attribute. The terms are highlighted or ordered based on the value of an associated metric.

[About the Tag Cloud](#)

[Using the Tag Cloud](#)

[Configuring the Tag Cloud](#)

## About the Tag Cloud

The **Tag Cloud** component allows users to quickly compare a set of displayed terms based on the value of an associated metric, such as the number of times the term occurs.

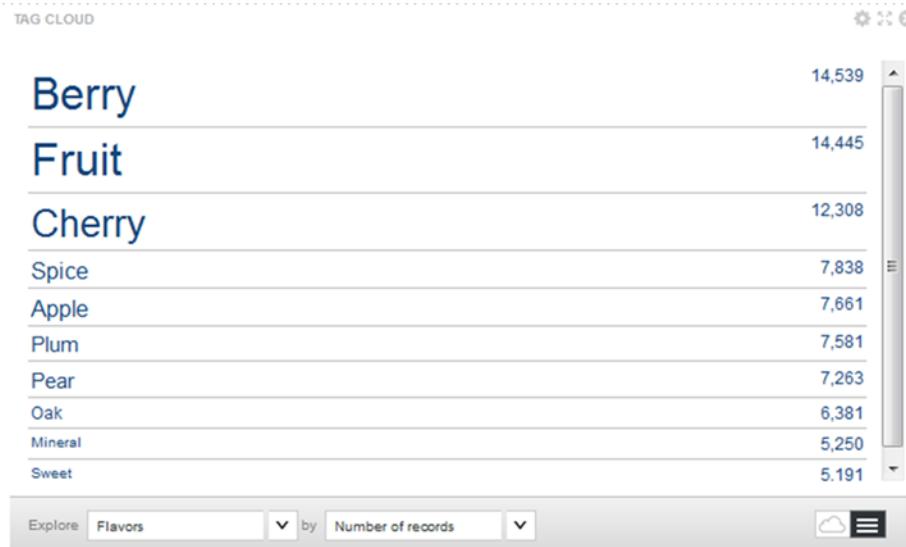
For example, a **Tag Cloud** component might display a list of wine flavors, organized based on the number of wines that include each flavor.



The component can also display the metric value associated with each term.



The **Tag Cloud** terms can display in a cloud, or as a simple list.



## Using the Tag Cloud

On the **Tag Cloud**, you can select the display type (cloud or list) and the terms to display. You may also be able to use the tag cloud terms to refine the data.

For information on using common component functions such as refining data, see [Using Common Component Functions on page 133](#).

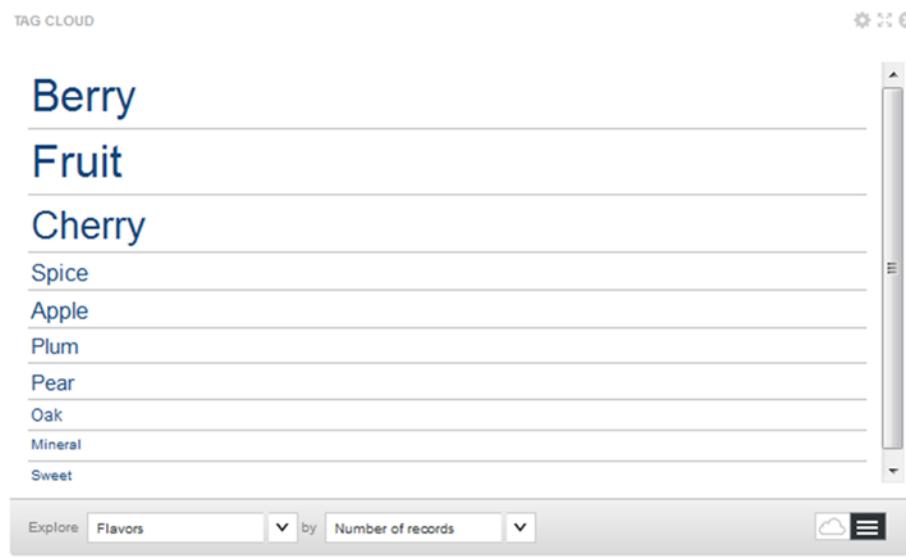
### Selecting the type of Tag Cloud display

You can display the **Tag Cloud** in either cloud layout or list layout.

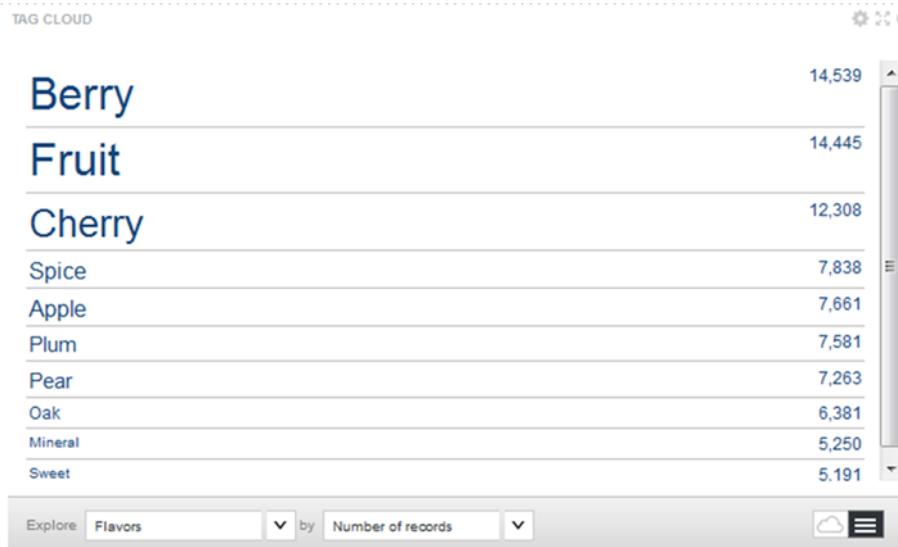
In cloud layout, the terms display in alphabetical order, and are not separated by a line break.



In list layout, each term displays on a separate line. The terms are displayed in descending order based on the value of the associated metric.



For either layout, each term may be followed by the value of the associated metric.



The **Tag Cloud** component provides a toggle icon to allow you to toggle between the two layouts. On the toggle icon, the currently displayed format is highlighted.

To toggle to the other layout, click the toggle icon.

## Selecting the values to display on the Tag Cloud

A **Tag Cloud** component can be configured with multiple options for the set of terms to display and the metric used to determine the relative size or display order of those terms.

If there are multiple sets of terms available, then the component includes an **Explore** drop-down list. From the **Explore** drop-down list, select the set of terms to display.

If there are multiple metrics available, then a **by** drop-down list is displayed. From the **by** drop-down list, select the metric to use to compare the displayed terms.



## Configuring the Tag Cloud

For a **Tag Cloud** component, you select the available values to display and the format in which to display them

### Selecting the available dimensions for the Tag Cloud

The terms displayed on the **Tag Cloud** are values from a selected dimension. From the edit view of the **Tag Cloud** component, you can configure a list of available dimensions for end users to select from.

On the **Configuration** tab of the **Tag Cloud** edit view, the **Dimensions** list contains the available dimensions.

Dimensions (terms in cloud) 	Metrics (size of terms) 
Flavors  	Number of records  
Designation  	Sales Price (average)  
	Base Score (average)  

By default, the **Tag Cloud** component contains the first available multi-value dimension that allows refinement.

For additional information on selecting attributes for a component, see [Selecting the attributes to use on a component on page 90](#).

To select the available **Tag Cloud** dimensions:

1. To add a dimension to the **Dimensions** list, drag it from the attributes list.  
Note that you cannot use date/time attributes in a **Tag Cloud**.
2. To determine the order in which to display the dimensions in the **Explore** drop-down list, drag each dimension to the appropriate location in the list.  
The dimension at the top of the list is selected by default when the component is first displayed.
3. To remove a dimension from the **Dimensions** list, click its delete icon.

### Configuring the available Tag Cloud dimensions

For each **Tag Cloud** dimension, you can configure the capitalization to use for the displayed terms. You can also set up a cascading hierarchy for when users refine by a **Tag Cloud** term.

To configure the **Tag Cloud** dimensions:

1. In the **Dimensions** list, click the edit icon for the dimension you want to configure.  
The **Edit Dimension** dialog is displayed.



2. The **Attribute Cascade** section allows you to enable dimension cascading when users refine by a **Tag Cloud** term.

For details on how to configure a dimension cascade, see [Configuring cascading for dimension refinement on page 104](#).

3. The **Value Formatting** section allows you to control the capitalization of the displayed terms.



#### Under Value Formatting:

- To display the **Tag Cloud** terms exactly as they are stored in the data, click the **Default (no change)** radio button.
- To display all of the terms in lowercase, click the **Lower case** radio button.

4. To save the configuration, click **Apply**.

## Configuring the relative size or order of the Tag Cloud terms

On the **Tag Cloud**, the value of the selected metric determines the relative size of the **Tag Cloud** terms. For list view, it also determines the display order.

On the **Configuration** tab of the **Tag Cloud** edit view, the **Metrics** list contains the available metrics for end users to select from.

Dimensions (terms in cloud)	Metrics (size of terms)
Flavors	Number of records
Designation	Sales Price (average)
	Base Score (average)

For additional information on selecting attributes to use for a component, see [Selecting the attributes to use on a component on page 90](#).

In addition to regular attributes and system metrics, the **Tag Cloud** also provides a Relevancy metric. The Relevancy metric applies when the data has been refined, and determines how relevant each term is to the current refinement. To determine the relevancy of each term, Big Data Discovery compares the frequency of the term in the current refinement to the frequency in the data as a whole.

- If a term occurs more frequently in the current refinement than in the data as a whole, then that term is more relevant, and is larger.
- If a term does not occur any more frequently in the current refinement, then the term is less relevant, and is smaller.

For example, a **Tag Cloud** contains wine types. The Sparkling wine type occurs in 30% of the records. When users refine by a specific region, if the Sparkling wine type now occurs in more than 30% of the records, then the Sparkling wine type is more relevant. If the Sparkling wine type only occurs in 30% or fewer of the records, then it is less relevant.

To select the available **Tag Cloud** metrics:

1. To add a metric to the **Metrics** list, drag the metric from the attributes list to the **Metrics** list.  
Note that you cannot use date/time attributes in a **Tag Cloud**.
2. To determine the order in which the metrics display in the **by** drop-down list, drag each metric to the appropriate location in the list.  
The metric at the top of the list is selected by default when the **Tag Cloud** component is first displayed.
3. To remove a metric from the list, click its delete icon.

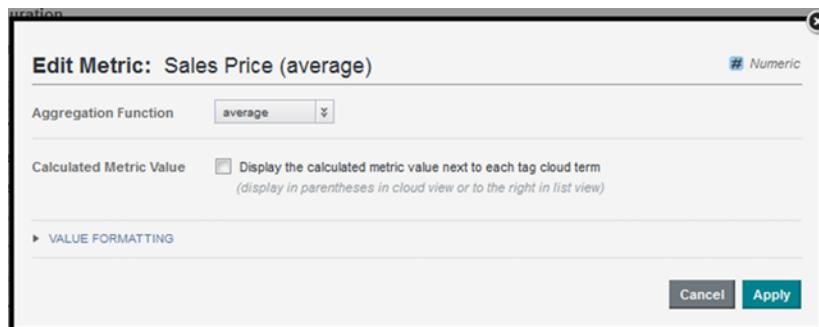
## Configuring the available Tag Cloud metrics

For each **Tag Cloud** metric, you can configure the aggregation function to use. You also determine whether to display the metric values on the **Tag Cloud**, and the format to use for those values.

On the **Configuration** tab of the **Tag Cloud** edit view, to configure a metric:

1. In the **Metrics** list, click the edit icon for the metric.
2. On the **Edit Metric** dialog, from the **Aggregation Function** method drop-down list, select the aggregation method to use to calculate the metric value.

For information on selecting a metric aggregation method, see [Selecting the aggregation method to use for a metric on page 93](#).



3. To display the calculated metric value for each **Tag Cloud** term, check the **Display calculated metric value next to each tag cloud term** checkbox.
4. If you are displaying the metric values, then use the **Value Formatting** section to customize the display format.  
For details on customizing the format of values displayed on a component, see [Configuring the format of values displayed on a component on page 93](#).
5. To save the configuration, click **Apply**.

## Configuring the general Tag Cloud display options

The **Display Options** tab of the **Tag Cloud** edit view provides additional display configuration options for the **Tag Cloud**.

The configuration includes the number of terms to display, the default format, and the text size for the terms.

**Display Options**

Configure the size, format, and optional features for the tag cloud.

Maximum number of keywords:  Enter a value between 1 and 100

Maximum component height for list view:  pixels

Tag Cloud format:  Cloud  List  
 Allow end user to toggle format

Text size:   32px

To configure the display:

1. In the **Maximum number of keywords** field, type the maximum number of terms to display on the **Tag Cloud**.
2. In the **Maximum component height for list view** field, type the maximum height in pixels of the **Tag Cloud** component when displayed in list view.  
If the list does not fit within the specified height, then end users can scroll through the list.
3. Under **Tag Cloud** format:
  - (a) Click a radio button to select the default display format (**Cloud** or **List**) for the tag cloud.
  - (b) By default, end users can toggle between cloud and list view. To prevent them from changing the format, uncheck the **Allow end user to toggle format** checkbox.
4. Under **Text size**, use the sliders to set the minimum and maximum size in pixels of the **Tag Cloud** terms.

By default, the text size ranges from 10 pixels to 32 pixels.

You can set the text size down to a minimum of 8 pixels, and up to a maximum of 48 pixels.

You can also put the sliders on top of each other to set a single text size for all of the terms. For example, for a list view that includes the number of values next to each term, you may just want to display all of the terms at a size of 11 pixels.

## Thematic Map

The **Thematic Map** component plots regions on a map and distinguishes the regions by a metric and aggregation method that you choose.

### *About the Thematic Map*

### [Configuring the Thematic Map connection to Oracle MapViewer](#)

### [Using the Thematic Map](#)

## About the Thematic Map

The **Thematic Map** component shows the data graphically and allows you to refine the data by regions including country, region, county, etc. You can refine data by drilling down into more granular regions in a thematic map and see the results associated with a sub-region.

The component can only be used if the data set contains at least one geocode attribute along with five related attributes that qualify the geocode attribute with additional region information.

The naming syntax for the geocode attribute varies slightly depending on whether it is a latitude/longitude geocode or an IP address/address geocode. This naming difference only affects the main geocode attribute but not the five related attributes.

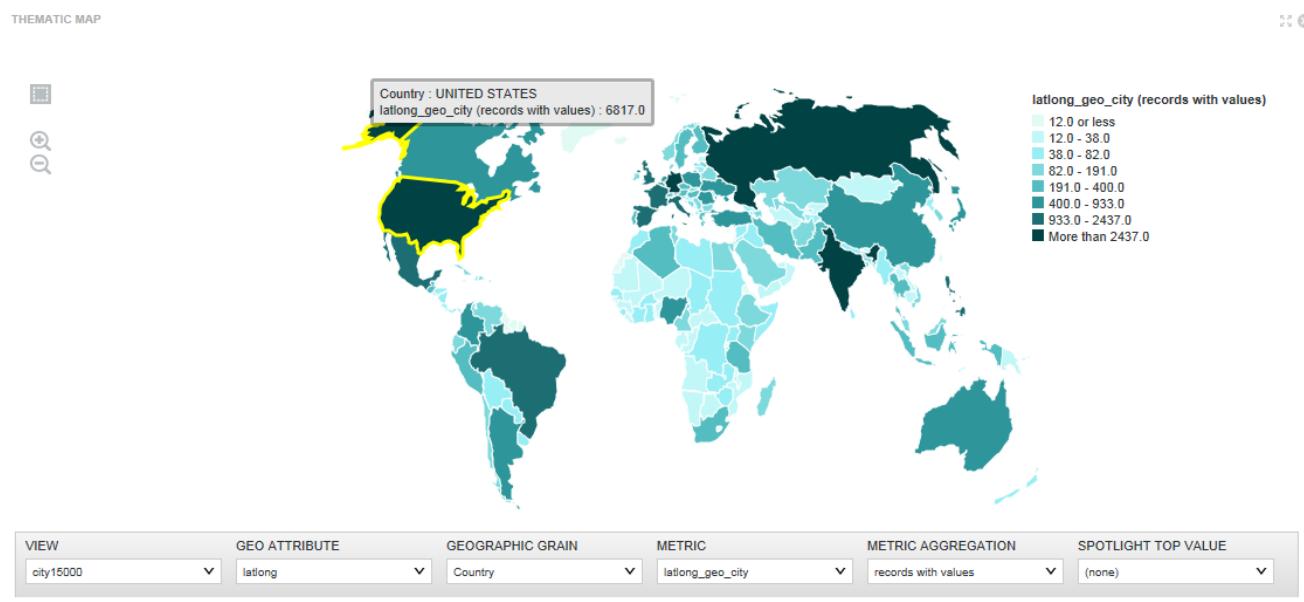
Here is the minimum set of attributes required to render a **Thematic Map** component:

- <attributeName> for geocode attributes or <attributeName>\_geo\_geocode for IP address or address attributes
- <attributeName>\_geo\_country
- <attributeName>\_geo\_region
- <attributeName>\_geo\_subregion
- <attributeName>\_geo\_regionid This Id attribute must be present in the data set but it is not exposed in the tabular view of **Explore**.
- <attributeName>\_geo\_subregionid This Id attribute must be present in the data set but it is not exposed in the tabular view of **Explore**.

The five related attributes are produced in one of two ways:

- The data enrichment process creates the region attributes automatically during data processing. Remember that auto-enrichment is enabled in Studio using the `bdd.enableEnrichments` setting on the **Control Panel>Data Processing Settings** page.
- You can create the region attributes manually by writing a custom transformation script on the **Transform** page of Studio.

For example, if you create the region attributes manually and they are based on an IP attribute, you call the `geotagIPAddressGetGeocode` transform function to create a geocode attribute based on an IP attribute. (See the *Extensions Guide* for details about functions.)



## Configuring the Thematic Map connection to Oracle MapViewer

The **Thematic Map** component uses Oracle's MapViewer (11.1.1.7.2 or greater) to display the map. Big Data Discovery includes framework settings to configure the connection to MapViewer.

By default, Big Data Discovery is configured to use the public instance of MapViewer. If you are using that instance, and your browser has access to the Internet, then you should not need to make any configuration changes.

Big Data Discovery includes the following framework settings related to the MapViewer connection. If you are using your own instance of MapViewer, or if your Web browser does not have Internet access, then you will need to change these settings. For details on configuring framework settings in Big Data Discovery, see the *Oracle Big Data Discovery Administrator's Guide*.

Framework Setting	Description
df.mapViewer	<p>The URL of the MapViewer instance.</p> <p>By default, this is the URL of the public instance of MapViewer.</p> <p>If you are using your own internal instance of MapViewer, then you must update this setting to connect to your MapViewer instance.</p>

## Using the Thematic Map

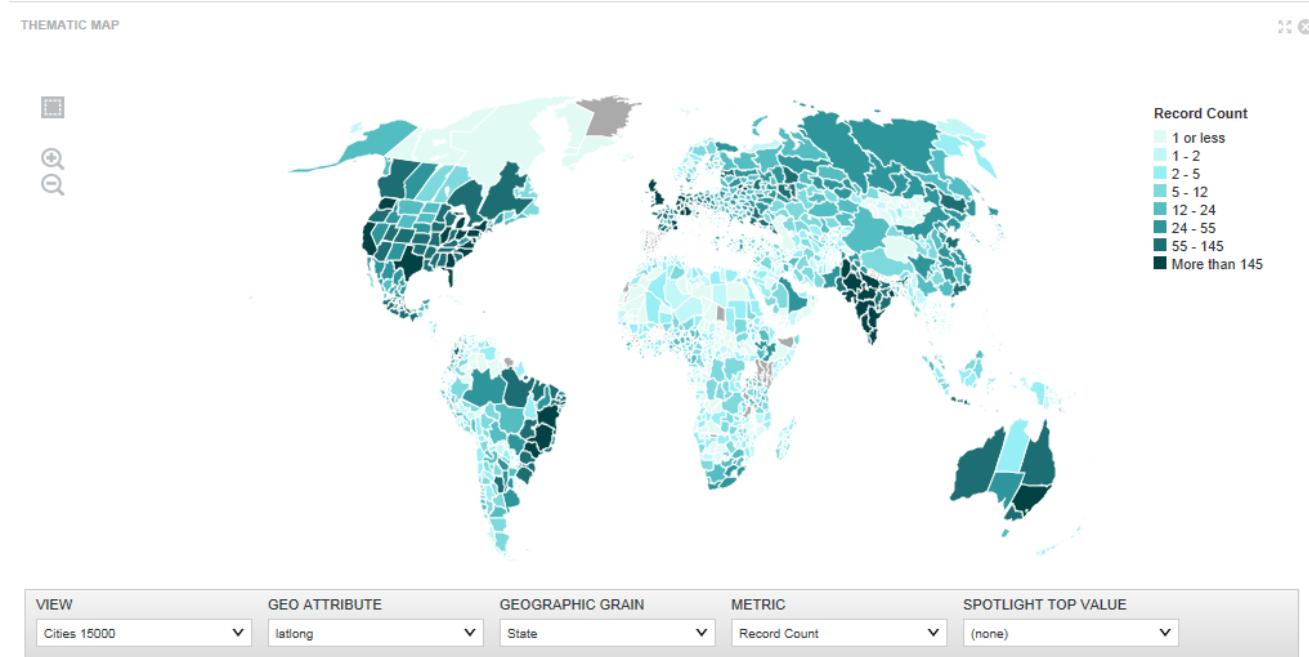
You can create a new **Thematic Map** component in either the scratchpad or the **Discover** page.

### Choosing the geocode attribute for the thematic map

From the **Geo Attribute** drop-down, you can select an attribute name on which to aggregate records in the thematic map.

### Choosing a geographic grain region for the thematic map

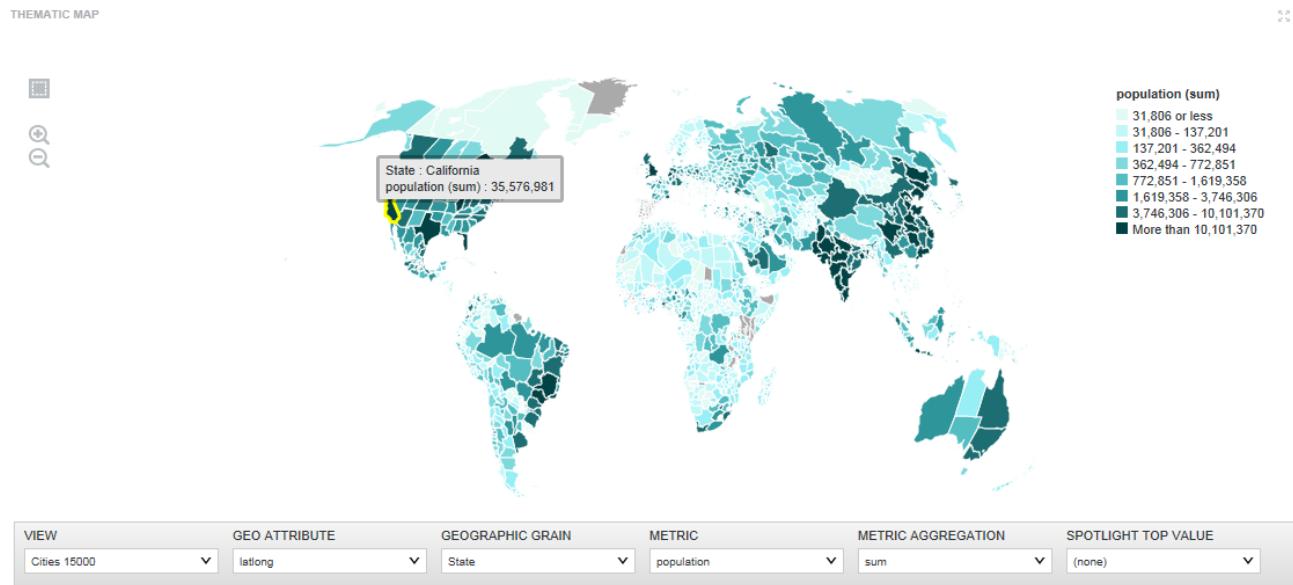
From the **Geographic Grain** drop-down, you can select a region name on which to aggregate records in the thematic map. For example, this thematic map aggregates records based on State rather than Country.



### Choosing a metric and metric aggregation

From the **Metric** drop-down, you can select a metric name on which to aggregate records in the thematic map. Optionally, from the **Metric Aggregation** drop-down, you can select a method to aggregate the value of the

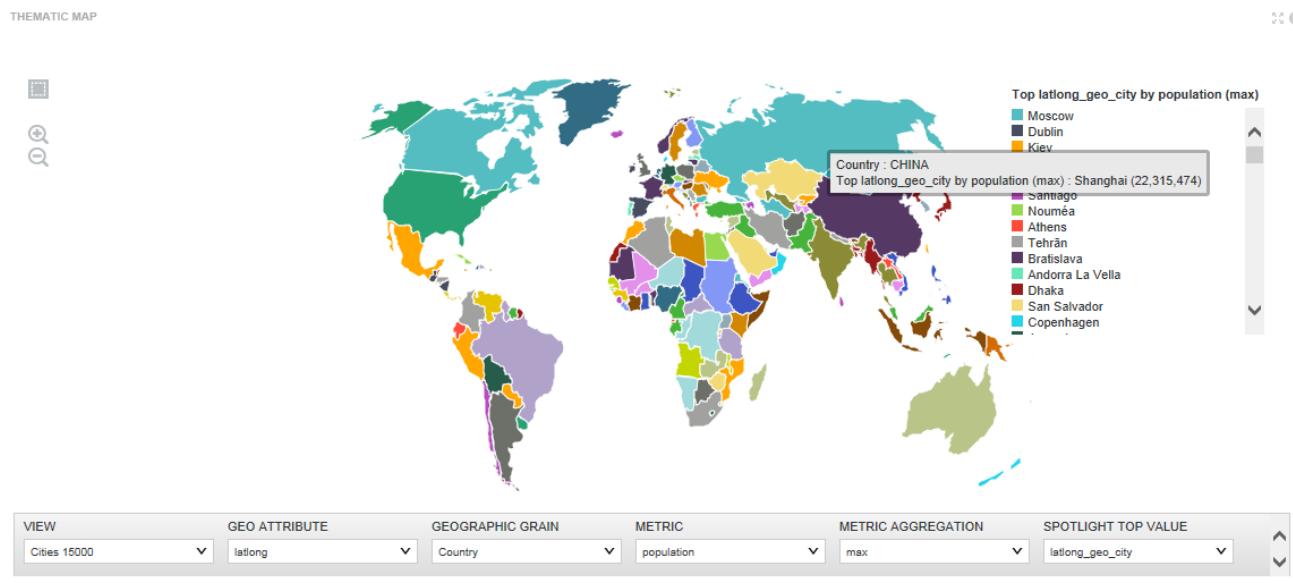
metric you selected in the **Metric** option. For example, this thematic map first aggregates records based on State and then sums the records for the state based on population.



## Spotlighting a value in the map

You can choose from the **Spotlight Top Value** drop-down to display the region (State) according to the largest value of the metric you selected in the **Metric** option (population).

For example, this map shows the city in each country that has the maximum population. The grain is set to country, and the metric and metric aggregation instruct the component to calculate records based on maximum population, and lastly, the **Spotlight Top Value** groups the records by city. So you can see that in China, the city of Shanghai has the largest population:



## Zooming in/out and redrawing the map

You can use Zoom In and Zoom Out features on the left edge of the component to manipulate the map.

You can multi-select regions using both Marque drawing tool and the usual ctrl click method. You can also drag the thematic map to different locations on the page and then use the tools to drill down into finer views of the data.

## Timeline

The **Timeline** component displays one or more metrics over a selected time period.

[About the Timeline](#)

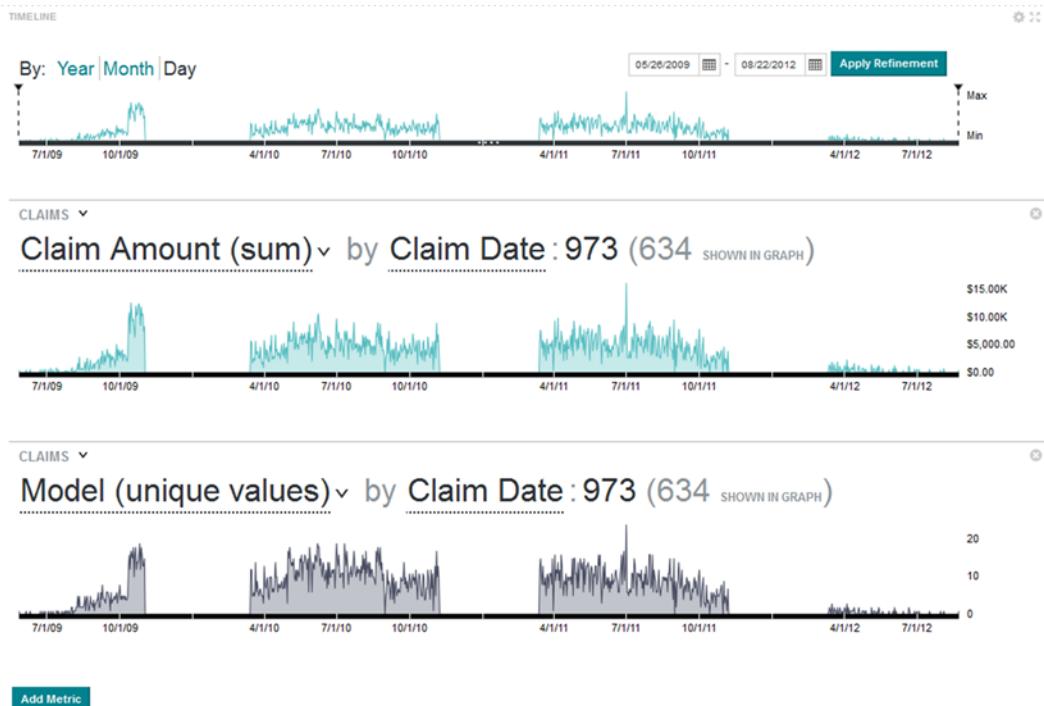
[Adding and editing metric timelines](#)

[Setting the displayed date unit and date range for the metric timelines](#)

[Saving the current timeline configuration as the default](#)

## About the Timeline

The **Timeline** component allows you to see changes in the value of one or more metrics over the same time period. For example, one timeline could track total sales against the sale date, and another could track total claims against the claim date.



The component consists of:

Timeline Type	Description
Master timeline	<p>Used to set the range of dates to use on all of the metric timelines. The full range of available dates is based on the earliest and latest dates that are available for the date attributes displayed on the metric timelines.</p> <p>From the master timeline, you can also select a date subset to use on the charts. You can only select date parts that are available for all of the selected date attributes.</p> <p>Note also that on the master timeline, the metric values are plotted based on the relative minimum and maximum values for each metric. The metrics do not use a single scale.</p>
Metric timelines	<p>Each metric timeline is a bar or line chart that is associated with a specific data set. To plot the chart, the metric timeline uses:</p> <ul style="list-style-type: none"> <li>• A date/time attribute, for the horizontal axis of the chart. If a data set does not include at least one date/time attribute, then you cannot use it on the <b>Timeline</b> component.</li> <li>• A metric value to plot against the date/time values</li> </ul> <p>Each metric timeline includes the number of matching records for the current refinement.</p>

You can add and remove metric timelines, and use a date range to refine the data.

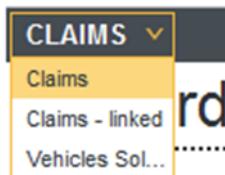
Administrators can also save a configuration as the default to display for all users.

## Adding and editing metric timelines

On the **Timeline** component, you can add multiple metric timelines. For each metric timeline, you select the view, date/time attribute, and metric to use for the display. You can also change the display order of the metric timelines.

To add and edit metric timelines:

1. To add a new metric timeline, click the **Add Metric** button.
2. On each metric timeline, from the view drop-down list, select the view to use for the timeline.

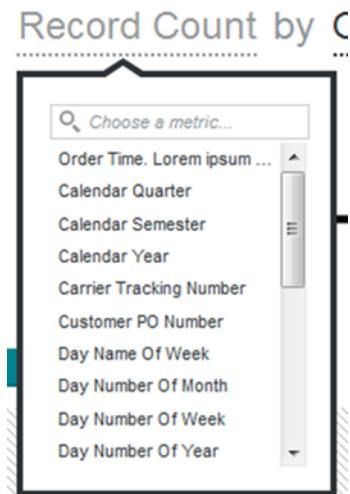


If the view does not include at least one date/time attribute that can be used as a dimension, then it cannot be used for a timeline.

3. To select the metric to display on the metric timeline:

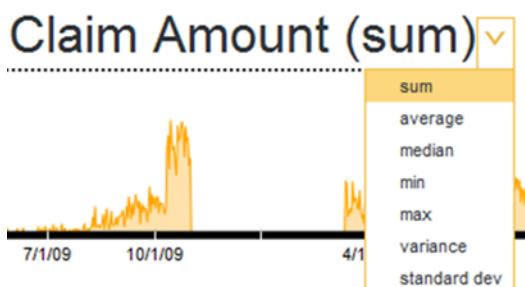
- (a) Click the metric name.

Big Data Discovery displays a list of available attributes to use for the metric.



You can also use the search field to find a specific attribute.

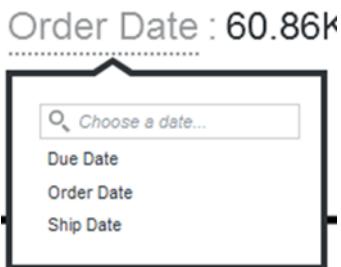
- (b) Click the attribute you want to use for the metric.
  - (c) If the metric is not a predefined metric, then use the aggregation method drop-down list to select the aggregation method to use to calculate the metric value.



- To select the date/time attribute to use on the metric timeline:

- Click the attribute name.

Big Data Discovery displays the list of available date/time attributes.



You can also use the search field to find a specific attribute.

- Click the date/time attribute you want to use.

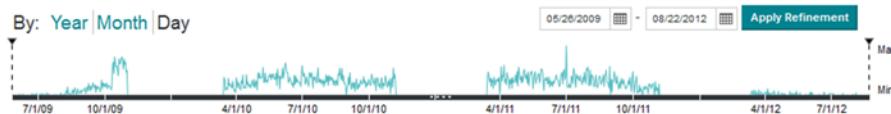
- To change the display order of the metric timelines, drag the timeline to the new location.

- To delete a metric timeline, click its delete icon.

If there is only one metric timeline, then you cannot delete it. You can only change its configuration.

## Setting the displayed date unit and date range for the metric timelines

On the **Timeline** component, you can use the master timeline to refine the project data based on a specific date/time range, or refine by a date value on a metric timeline. You can also change the date unit to use on the horizontal axis for the charts.



Note that if date/time refinements have already been applied, then those values are displayed above the master timeline.

To change the displayed date subset and date unit for the metric timelines:

- Above the master timeline are the available date units that you can use (for example, Year, Month, Day). To display a different date part on the metric timelines, click the date part.

You can only use date units that are available for all of the selected date/time attributes. So if one date/time attribute supports Year, Month, and Day, but the other only supports Year and Day, then you can only select from Year and Day.

- To use the master timeline to refine the data:

- Select the date/time range.

To do this, you can either use the sliders on the master timeline, or use the date picker fields.

When change the date range, then for each metric timeline, in parenthesis next to the number of records for the current refinement, the **Timeline** component displays the number of records that will match if you applied the new date range as a refinement.



- (b) Click **Apply Refinement**.

The date range refinements are added to the **Selected Refinements** panel.

The metric timelines are updated to show the selected date range, and the number of records for each metric timeline is also updated.

3. To refine by a date value on a metric timeline, double-click the point on the metric timeline.

The value for that point is added as a refinement for the date attribute displayed on that timeline.

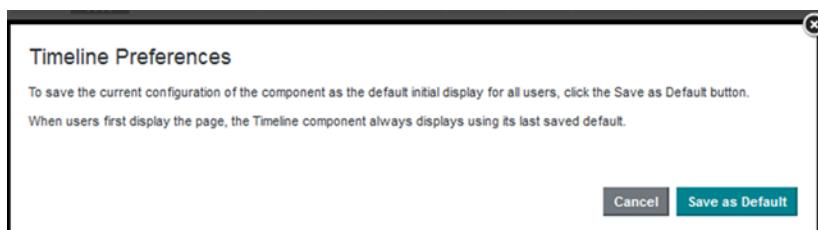
## Saving the current timeline configuration as the default

If you are able to edit the project, then you can also save the current configuration of the **Timeline** component as the default for all users.

When users first display the page, the **Timeline** component initially displays using the last saved default.

To save the current timeline configuration as the default:

1. Click the component configuration icon.
2. On the **Timeline Preferences** dialog, click **Save as Default**.





## Chapter 34

---

# Web-Based Content

These components are used to display web-based content on a page. The content may be custom content created and stored in Big Data Discovery, or may be content from external URLs displayed within a component.

### *IFrame*

#### *Web Content Display*

## IFrame

The **IFrame** component can display the content of any external URL.

#### *About the IFrame*

#### *Using the IFrame*

#### *Configuring an IFrame*

## About the IFrame

The **IFrame** component allows users to view from within a project content from an external URL.

For example, you might want to display your company's home page, or provide access to tools to allow users to share their findings.



## Using the IFrame

The **IFrame** component displays the content of the URL associated with the component.

Link targets from the URL also display within the **IFrame** component, unless they are specifically configured to display in another browser window.

## Configuring an IFrame

For an **IFrame** component, you configure the URL to display, and determine whether to require additional authentication in order to view the content.

To configure an **IFrame** component:

1. On the **IFrame** edit view, in the **Source URL** field, type the URL.

The screenshot shows the 'IFRAME' edit view. The 'General' section contains a 'Source URL' field with the value 'http://www.oracle.com' and a 'Relative to Context Path' checkbox. The 'Authentication' section contains an 'Authenticate' checkbox. The 'Advanced' section contains an 'HTML Attributes' text area with the following code:  
`border=0  
bordercolor=#000000  
frameborder=0  
height-maximized=600  
height-normal=300  
hspace=0  
scrolling=auto`  
A 'Save' button is at the bottom.

2. If the URL is relative to the context path for Big Data Discovery, check the **Relative to Context Path** checkbox.
3. If the page you are embedding in the **IFrame** component requires authentication, then you can provide the authentication information as part of the component configuration. To do this:

- (a) Check the **Authenticate** checkbox.

The authentication fields are displayed.

The screenshot shows the 'Authentication' configuration section. It includes an 'Authenticate' checkbox (checked), an 'Authentication Type' dropdown set to 'Basic', a 'User Name' text field, and a 'Password' text field.

- (b) From the **Authentication Type** drop-down list, select the type of authentication.

**Basic** authentication simply provides the user name and password required by the embedded page.

**Form** authentication uses POST or GET to validate the user.

- (c) For **Basic** authentication, to specify the user name and password, enter the values in the fields.
- (d) For **Form** authentication, from the **Form Method** drop-down list, indicate whether to use GET or POST to validate the user.

The screenshot shows a configuration dialog for authentication. The 'Authentication' section is expanded. Under 'Authenticate', the checkbox is checked. 'Authentication Type' is set to 'Form'. 'Form Method' is set to 'Post'. For 'User Name', there are two fields: 'Field Name' and 'Value'. For 'Password', there are also 'Field Name' and 'Value' fields. At the bottom, the 'Hidden Variables' field contains the value 'var1=hello;var2=world'.

For the user name and password, provide the field names and values to send.

In the **Hidden Variables** field, provide any hidden variables to include in the authentication request.

4. Use the **HTML Attributes** field to provide any additional display parameters for the content.
5. To save the configuration, click **Save**.
6. To return to the end user view, click **Exit**.

## Web Content Display

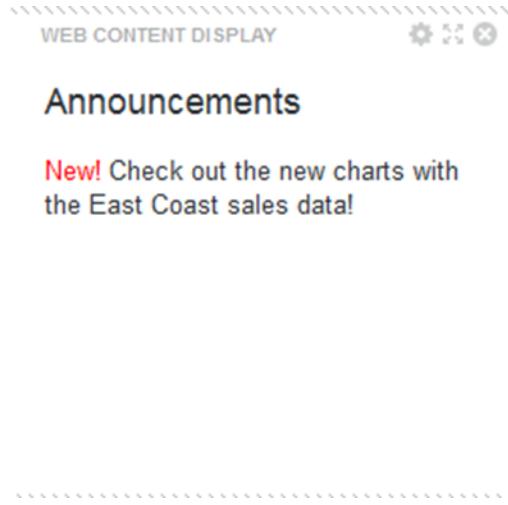
The **Web Content Display** component allows you to create and display HTML content in your project.

[About the Web Content Display](#)

[Configuring a Web Content Display](#)

## About the Web Content Display

The **Web Content Display** component allows you to display HTML-based content on a project page. For example, you could use the component for a brief welcome note, announcement, or instructions.

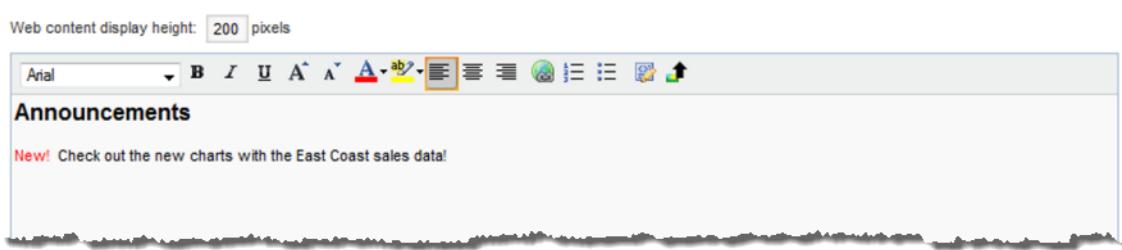


## Configuring a Web Content Display

For the **Web Content Display** component, you provide the content to display. You can also customize the component height.

To configure the **Web Content Display** component:

1. By default, the component is 200 pixels high. Use the **Web content display height** field to change the height.
2. Use the text editor to create or edit the component content.



The text editor provides options to:

- Select the font
- Highlight text
- Change the text alignment
- Create numbered and bulleted lists

Note that you cannot include script tags or a separate style tag. If you do include these items, they are removed when you save the component.

3. To add a hyperlink:
  - (a) Select the link text.
  - (b) In the text editor toolbar, click the hyperlink icon.
  - (c) On the hyperlink dialog, in the **URL** field, type the URL.



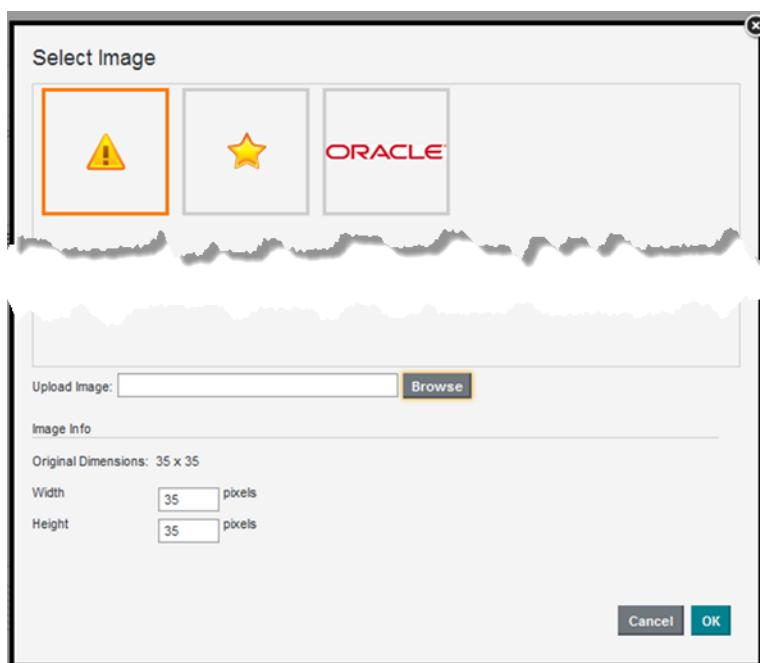
- (d) You can also edit the link text.
  - (e) To save the link, click **OK**.

When end users click links in a **Web Content Display** component, the target URL automatically displays in a new browser window.

To edit an existing link, click anywhere in the link text, then click the hyperlink icon.

4. To insert an image:
  - (a) In the text editor toolbar, click the image upload icon.

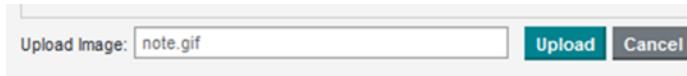
The image upload dialog is displayed. The dialog displays any images that have been uploaded for this or any other **Web Content Display** component in the current project.



The list does not include images uploaded in other components or other projects.

The images are stored in the database.

- (b) To use an existing image in the list, click the image, then click **OK**.
- (c) If you need to upload a new image, click the **Browse** button to search for and select the image.

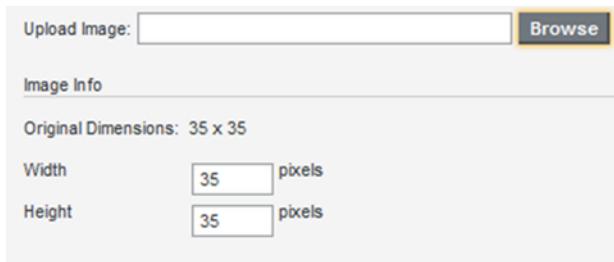


Big Data Discovery supports the .gif, .jpeg/.jpg, .bmp, and .png file types.

To clear the image file selection, click **Cancel**.

- (d) To upload the selected file, click **Upload**.

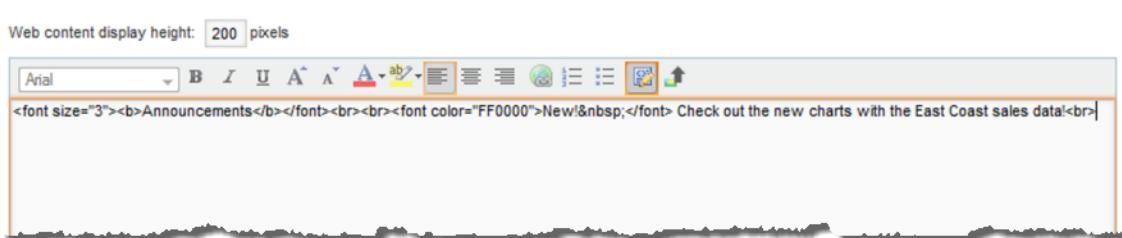
The image is added to the list, and the dialog is updated to display the image width and height.



You can use the **Width** and **Height** fields to customize the displayed size of the image. The image always maintains its original aspect ratio. So for example when you edit the width, the height is automatically updated.

- (e) To insert the currently selected image, click **OK**.

5. To display the source HTML, click the source view icon.



# Glossary

## attribute

**Attributes** are named pieces of information about a set of records. They consist of names and values of fields on a record, and by this, contain the core information on record.

Just like columns describe rows in a table, attributes describe records in the Dgraph. Each set of attributes is specific to a particular data set of records. For example, a data set consisting of a store's products might contain attributes like "item name", "size", "color", and "SKU", for which records can have values.

If you think of a table representation of records, a record is a row and an attribute name is a column header, where attributes values are values in each column.

Most attributes that appear in Big Data Discovery appear in the underlying source data, but in Big Data Discovery you can also create new attributes, change, or delete attributes within the project. (These changes are not persisted to the source data.) Also, some attributes are the result of transformations that Big Data Discovery runs on the data it discovers.

An attribute's configuration in the collection schema controls three characteristics of each attribute — required (or not), unique (or not), and having a single or multiple assignments. In other words, an attribute's configuration in the schema determines whether an attribute is:

- **Required.** For a required attribute, each record must have at least one value assigned to it.
- **Unique.** For unique attributes, two records can never have the same assigned value.
- **Single-value or multi-value** (also known as **single-assign** and **multi-assign**). Indicates whether a record may have at most one value, or can have more than one value assignments for the same attribute.

**Single-value attributes** can at most have one assigned value on a record. For example, each item can have only one SKU.

**Multi-value attributes** allow for multiple assigned values on a single record. For example, a Color attribute may allow multiple values for a specific record.

These characteristics of attributes, along with attribute type in the index, originate in the schema maintained in the Dgraph. In addition to these characteristics of attributes, Studio has additional characteristics, such as refinement mode or a metric flag. Studio also lets you localize attribute descriptions and display names.

See also "data set", "index", "schema", "record", "type (for attributes)", and "value".

## base view

**The base view** for a data set represents the fundamental attributes in a project data set. Base views represent the data "as is".

You can create custom views, which are useful for data aggregation, computation, and visualization.

**Custom views** include only data from specific selected attributes (or columns) in the underlying data. They provide different ways of looking at data. Each custom view has a definition expressed as an EQL statement. Custom views do not eliminate the base view, which is always present in the system.

You can create multiple custom views in your project in parallel to each other.

**Linked views** are created automatically when you join data sets. They are broadened views of data. Linked views widen a base view by joining the original data set with another data set.

## Big Data Discovery Cluster

A **Big Data Discovery Cluster** is a deployment of Big Data Discovery components on any number of nodes.

Nodes in the deployment have different roles:

- **CDH nodes** represent machines in the Cloudera (CDH) cluster. A pre-existing CDH cluster is assumed when you deploy Big Data Discovery. Some of the machines from the pre-existing CDH cluster may also become the nodes on which components of Big Data Discovery (those that require running in Hadoop) are deployed.
- **Weblogic Server nodes** are machines on which Java-based components of Big Data Discovery — Studio and Dgraph Gateway — run inside Weblogic Server. At deployment time, you can add more than one of these WebLogic Server machines (or nodes) to the BDD cluster.
- **Dgraph-only nodes** represent machines in the Big Data Discovery cluster on which the Dgraph instance is running. They themselves form a Dgraph cluster within the Big Data Discovery cluster deployment.

You have multiple options for deploying Big Data Discovery in order to use hardware efficiently. For example, you can co-locate various parts of the Big Data Discovery on the same nodes. Several recommended deployment scenarios are provided to help you efficiently deploy Big Data Discovery:

- Learning/demo deployment on one or two machines.
- Production deployment on a set of 6 machines:
  - 3 run a CDH cluster only
  - 2 run Weblogic Server with Studio and the Dgraph Gateway
  - 1 runs the Dgraph

The number of nodes in the production deployment can be fewer than 6 (with some software co-located), or can be more.

See the deployment diagrams and co-location considerations in the *Installation and Deployment Guide*.

## Catalog

A **Catalog** is an area of the Studio application that lists:

- Data sets available to you
- Projects that you have access to

The **Catalog** contains options to create a new data set, explore a data set, or navigate to an existing project.

When the Data Processing component of Big Data Discovery runs, the available data sets are discovered by Big Data Discovery in the Hive database, profiled, and then presented as a list in the **Catalog**.

You can then use the **Catalog** to identify data sets of interest to you, by navigating and filtering data sets and projects based on data set metadata and various characteristics of projects. You can also display additional details about each data set or project, for further exploration.

The very first time you log in to Big Data Discovery, the **Catalog** may display discovered data sets only, but not projects. After you or members of your group create and share projects, the **Catalog** displays them when you log in, in addition to the available data sets.

## custom views

**Custom views** are useful for data aggregation, computation, and visualization. Comparing with base views that contain fundamental data for records, custom views include only data from specific selected attributes (or columns) in the underlying data. This way, custom views provide different ways of looking at data. Each custom view has a definition expressed as an EQL statement.

Custom views do not eliminate the base view, which is always present in the system. You can create multiple custom views in your project in parallel to each other.

See also "base view" and "linked views".

## data processing workflow

Data processing workflow (or, simply, **data processing**) is a stage in Big Data Discovery processing that includes:

- Discovery of data in Hive
- Creating a sample of a data set
- Running a select set of enrichments on data
- Profiling the data
- Indexing the sample

Data processing is performed automatically when Big Data Discovery starts. Administrators can control some parts of the data processing workflow. For information, see the *Big Data Discovery Data Processing Guide*.

See also "index", "enrichments", "sampling", and "profiling".

## data set

In the context of Big Data Discovery, a **data set** is a logical unit of data, which corresponds to source data such as a CSV file, an Excel file, or a Hive table. A data set is a result of data processing workflow which includes data processing, enriching, and indexing the source data.

The data set becomes available in Studio as an entry in the **Catalog**. A data set may include enriched data and transformations applied to it from **Transform**. Each data set has a corresponding set of Dgraph index files.

A data set in Big Data Discovery can be created in different ways:

- When Big Data Discovery starts and you run its data processing workflow
- When you upload personal data files in Studio (CSV or Excel files)
- When you use Transform features to create a new data set based on the transform script

See also "attribute", "index", "schema", "record", "type (for attributes)", and "value".

## **data set import (personal data upload)**

**Data set import** (or **personal data upload**) is the process of manually creating a data set from Studio, by uploading data from an Excel or delimited (CSV) file.

## **Dgraph**

The **Dgraph** empowers search analytical processing of the data sets. It handles requests users make to data sets. The Dgraph uses proprietary data structures and algorithms to provide real-time responses to client requests for analytic processing and data summarization.

The Dgraph stores the index created after source data is discovered during the data processing stage in Big Data Discovery. After the index is stored, the Dgraph receives client requests through the Studio application tier, queries the index, and returns the results.

The Dgraph is designed to be stateless. This design requires that a complete query is sent to it for each request. The stateless design facilitates the addition of Dgraph processes (at deployment time) for load balancing and redundancy — any replica of a Dgraph can reply to queries independently of other replicas.

## **Dgraph Gateway**

The **Dgraph Gateway** is a Java-based interface in Big Data Discovery for the Dgraph, providing:

- Routing of requests to the Dgraph instances
- Caching
- Handling of cluster services for the Dgraph instances, using the ZooKeeper package from CDH

In BDD, the Dgraph Gateway and Studio are two Java-based applications co-hosted on the same WebLogic Server. You can have several WebLogic Servers in your deployment, each hosting an additional instance of Studio and Dgraph Gateway. In this case, your deployment typically includes an external load balancer in front of WebLogic Servers.

## **Discover**

**Discover** is one of the three main modes, or major areas of Studio, along with **Explore** and **Transform**. As a user, you work within one of these three modes at a time.

**Discover** provides an intuitive visual discovery environment where you can compose and share discovery dashboards using a wide array of interactive data visualization components. It lets you

mash up disparate data sources to find new insights and publish them across the enterprise using bookmarks, snapshots, and discovery narratives.

Unlike **Explore** and **Transform**, **Discover** is where you create persistent visualizations of your data and share them with other users of your project.

## enrichments

**Enrichments** are modules in Big Data Discovery that allow you to change the data presented in Big Data Discovery for exploration and analysis.

Enrichments extract semantic information from raw data for exploration and analysis. For example, Big Data Discovery has enrichments for finding administrative boundaries, such as states or counties, from Geocodes and IP addresses. It also has text enrichments that use advanced statistical methods to pull out entities, places, key phrases, sentiment, and other items from long text fields.

Some enrichments allow you to add additional derived meaning to your data sets, for example, you can derive positive or negative sentiment from the records in your data set. Other enrichments let you address invalid or inconsistent values.

Some enrichments run automatically during the data processing workflow if profiling shows that an attribute is suitable for a given enrichment. During data processing, Big Data Discovery discovers data in Hive tables, and performs data set sampling and initial data profiling.

Enrichments are derived from a data set's additional information such as terms, locations, the language used, sentiment, and key phrases. Big Data Discovery determines which enrichments are useful for each discovered data set, and automatically runs them on samples of the data. As a result of automatically applied enrichments, additional derived attributes (columns) are added to the data set, such as geographic data, or a suggestion of the detected language.

The data sets with this additional information appear in **Catalog**. This provides initial insight into each discovered data set, and lets you decide whether the data set is a useful candidate for further exploration and analysis.

In addition to automatically-applied enrichments, you can apply enrichments to a project data set from the **Transformation Editor** in **Transform**. From **Transform**, you can configure parameters for each type of enrichment. In this case, an enrichment is simply another type of available transformation.

See also "transformations".

## Enterprise Manager Plug-in for Big Data Discovery

**The Enterprise Manager Plug-in for Big Data Discovery** extends Oracle Enterprise Manager Cloud Control to add support for monitoring, diagnosing, and managing Big Data Discovery components.

There are three target types included with the Enterprise Manager plug-in:

- The Cluster target
- The Dgraph target
- The Studio target

## Explore

**Explore** is an area of Studio's user interface where users explore a single data set, either from the **Catalog**, or from within a project.

**Explore** provides a machine-guided tour of the data. Each attribute is represented by a tile with an interactive data visualization, based on profiling of the attribute. These visualizations let you instinctively engage with raw Hadoop data to find interesting patterns and triage messy data.

Use **Explore** to analyze the attributes and values in a data set. You explore a single data set at a time.

**Explore** displays the available attributes in the data set. The attributes are initially sorted by name. You can filter the displayed attributes, and change the sort order.

For each attribute, **Explore** provides a set of visualizations that are most suitable for that attribute's data type and value distribution.

Exploring a data set does not make any changes to that data set, but allows you to assemble a visualization using one or more data set attributes, and save it to a project page.

## exporting to HDFS/Hive

**Exporting to HDFS/Hive** is the process of exporting the results of your analysis from Big Data Discovery into HDFS/Hive.

From the perspective of Big Data Discovery, the process is about exporting the files from Big Data Discovery into HDFS/Hive. From the perspective of HDFS, you are importing the results of your work from Big Data Discovery into HDFS. In Big Data Discovery, the **Dgraph HDFS Agent** is responsible for exporting to HDFS and importing from it.

The exporting to HDFS process is not to be confused with data set import, also known as personal data upload, in which you explicitly add a data set to HDFS by uploading a file.

## index

The **index** represents the contents of the data sets that can be queried by Dgraph, in Big Data Discovery.

The index is what empowers analytical processing. It exists both in persistent files in memory and on disk.

Index refers to the entire set of files and to the logical structures into which the information they contain is organized internally. Logical structures describe both the contents and structure of the data set (schema).

The index stores data in way that allows the query engine (the Dgraph) to effectively run interactive query workloads, and is designed to allow efficient processing of queries and updates.

When you explore data records and their attributes, Big Data Discovery uses the schema and its index to allow you to filter records, identify their provenance (profiling), and explore the data using available refinements.

See also "attribute", "data set", "schema", "record", "refinement", "type (for attributes)", and "value".

## linked views

**Linked views** are created automatically when you join data sets. They are broadened views of data. Linked views widen a base view by joining the original data set with another data set.

See also "base view" and "custom views".

## metadata

Each data set includes various types of **metadata** — higher-level information about the data set attributes and values.

Basic metadata is derived from the characteristics of data sets as they are registered in Hive during data processing. This is called **data profiling**. Big Data Discovery performs initial data profiling and adds metadata, such as Geocode values, derived from running various data enrichments.

As you explore and analyze the data in Big Data Discovery, additional metadata is added, such as:

- Which projects use this data set
- Whether the source data has been updated

Some metadata, such as attribute type, or multi-value and single-value for attributes, can be changed in **Transform**. Other metadata uses the values assigned during data processing.

In addition, various types of attribute metadata are available to you in Studio. They include:

- Attribute display names and descriptions
- Formatting preferences for an attribute
- Available and default aggregation functions for an attribute

## Oracle Big Data Discovery

**Oracle Big Data Discovery** is a set of end-to-end visual analytic capabilities that leverage the power of Hadoop to transform raw data into business insight in minutes, without the need to learn complex products or rely only on highly skilled resources.

It lets you find, explore and analyze data, discover insights, decide and act.

The Big Data Discovery software package consists of these main components:

- **Studio**, which is the front-end Web application for the product, with a set of unified interfaces for various stages of your data exploration:

You use the **Catalog** to find data sets, and **Explore** to explore them.

You can then add data sets to projects, where you can analyze them, or use **Transform** to apply changes to them.

You can also export data to Hive, for further analysis by other tools such as R. Both **Explore** and **Transform** are part of the area in the user interface known as **project**. Note that you can explore data sets that are part of a project, as well as source data sets that are not included into any project.

- **Dgraph Gateway**, which performs routing of requests to the Dgraph instances that perform data indexing and query processing.

- The **Dgraph** is the query engine of Big Data Discovery.
- **Data Processing**, which runs in CDH and performs discovery, sampling, profiling, enrichment and transformations of source data.

## profiling

**Profiling** is a step in the data processing workflow that discovers the characteristics of source data, such as a Hive table or a CSV file, and the attributes it contains.

In Big Data Discovery, profiling results in the creation of metadata used to enhance the user experience by intelligently automating configuration. Metadata discovered during profiling includes attribute names, attribute data types, attribute's cardinality (a number of distinct values a record has from the attribute), and time when a data set was created and updated. For example, a specific data set can be recognized as a collection of structured data, social data, or geographic data.

Using **Explore**, you can look deeper into the distribution of attribute values or types.

Using **Transform**, you can adjust or change some of these metadata. For example, you can replace null attribute values with actual values, or fix other inconsistencies, such as change an attribute that profiling judged to be a String value into a number.

## project

A **project** is a container for data sets and user-customized pages in Studio.

In a project, you can create pages with visualizations, such as charts and tables. These visualizations allow you to perform additional analysis and discovery across one or more data sets.

From within a project, you can:

- Explore a data set
- Transform a data set
- Link data sets
- Create custom views of data set data

You can save and share projects with other users.

## record

A **record** is a collection of assignments, known as values, on attributes. Records belong to data sets.

For example, for a data set containing products sold in a store, a record can have an item named "t-shirt", be of size "S", have the color "red", and have an SKU "1234". These are **values** on attributes.

If you think of a table representation of records, a record is a row and an attribute name is a column header, where attribute values are values in each column.

## refinement state

A **refinement state** is a set of filter specifications (attribute value selections, range selections, searches) to narrow a data set to a subset of the records.

## sample

A **sample** is an indexed representative subset of a data set that you interact with in Studio. Data processing draws a simple random sample from the underlying Hive table, then creates an index for the Dgraph to allow search, interactive analysis, and exploration of data of unbounded size.

The default size of the sample is one million records, but you can optionally specify the size of the sample, or take a sample of different size.

## sampling

**Sampling** is a step in the data processing stage in Big Data Discovery, when a sample is taken. Working with data at very large scales causes latency and reduces the interactivity of data analysis. To avoid these issues in Big Data Discovery, you work with a sampled subset of the records from large tables discovered in HDFS. Using sample data as a proxy for the full tables, you can analyze the data as if using the full set.

During data processing, a random sample of the data is taken. The default sample size is 1 million records. Administrators can adjust the sample size. If a source Hive table contains fewer records than the currently specified sample size, then all records are fetched.

## schema

**Schema** defines all of the attributes in the data set, including all of the characteristics of each attribute.

See also "attribute", "data set", "index", "record", "type (for attributes)", and "value".

## scratchpad

The **scratchpad** is a part of **Explore** that lets you quickly compose visualizations using multiple attributes. When you add an attribute to the scratchpad, either by clicking on a tile, or by using typeahead within the scratchpad itself, the scratchpad automatically renders a best-practice data visualization based on all of the attributes in the scratchpad. This lets you concentrate on the data, rather than on chart configuration. In addition to automatically rendering a visualization, the scratchpad provides several alternative visualizations for the attributes, allowing you to quickly switch to an alternative view, without having to change any configuration. From within a project, you can save a scratchpad visualization to a page in **Discover** where you can use it within a discovery dashboard and apply a more fine-grained configuration.

## source data

**Source data** can be a CSV file, an Excel file, or a Hive table. All source data is visible in Hadoop, is stored in HDFS, and is registered as a Hive table.

Any source Hive tables can be discovered by the data processing workflow in Big Data Discovery. Data processing takes a random sample of a specific size, and creates a data set in Dgraph, visible in the **Catalog**, for possible exploration and selection.

Once a sampled source data appears in the **Catalog**, it becomes a Big Data Discovery data set, and already represents a sample of the source Hive table.

## Studio

**Studio** is a front-end Web application for Big Data Discovery. Studio is a component of Big Data Discovery that contains user interfaces and provides tools for users to create and manage data sets and projects, as well as administrator tools for managing user access and other settings.

Projects and settings are stored in a relational database.

## Transform

**Transform** is one of the three main areas of Studio, along with **Explore** and **Discover**. **Transform** is where you make changes to your project data set. It allows you to edit the data set values and schema, either to clean up the data, or to add additional values.

**Transform** unlocks data cleansing, manipulation and enrichment activities that are commonly restricted to rigid ETL processes. **Transform** lets you easily construct a transformation script through the use of quick, user-guided transformations as well as a robust, Groovy- based list of custom transform functions.

In **Transform**, you interactively specify a list of default enrichments and transformations, or you can write your own custom transformations. You can preview the results of applying the transformations, then add them to a transformation script that you can edit, apply to the project data set, and save.

You can either apply the transform script to your current project, or create a new data set using the transformation script. Note that when a transformation script is applied to the project data set (which is a sample), it creates a new version of the project data set and publishes it to the **Catalog**. You can also create a new data set using the transformation script. This creates a new full data set in Hadoop, thus unlocking the transformed data for exploration in Big Data Discovery as well as in other applications and tools within Hadoop.

## Transformation Editor

The **Transformation Editor** is a part of **Transform** in Studio where you transform your data and create derived attributes using the Groovy language. Along with Groovy support, the **Transformation Editor** gives you access to a list of additional, easy-to-use custom transform functions (based on Groovy) that let you speed up the process of data conversion, manipulation and enrichment.

## transformations

**Transformations** are changes to a project data set. For example, you can do any of the following transformations:

- Change data types

- Change capitalization of values
- Remove attributes or records
- Split columns into new ones (by creating new attributes)
- Group or bin values
- Extract information from values

Transformations can be thought of as a substitute for an ETL process of cleaning your data. Transformations can be used to overwrite an existing attribute, modify attributes, or create new attributes.

Most transformations are available directly as specific options in **Transform**. Some transformations are enrichments.

You can use the Groovy scripting language and a list of custom, predefined Groovy-based transform functions available in Big Data Discovery, to create a transformation script.

When a transformation script is applied, it creates a new version of the project data set and publishes it to the **Catalog**. You can either apply the transform script to your current project, or create a new data set using the transformation script:

- When you apply the transformation script to the project, no new entry is created in the Catalog, but the current project does see the effects of the transform script.
- When you create a new data set using the transformation script, a new data set entry is added to the **Catalog** for use by other projects. That new data set is a new sample of the original source Hive table after the transformation script is applied. Creating a new data set in this way does not apply the transformation script to the current project.

See also "enrichments", "Transform", and "Transformation Editor".

## **type (for an attribute)**

An attribute's **type** determines the possible values that can be assigned to that attribute. Examples of attribute types include Boolean, Integer, String, Date, Double, and Geocode.

String attributes have additional characteristics related to text search.

See also "attribute", "data set", "index", "schema", "record", and "value".

## **value (for an attribute)**

An attribute's **value** is an assignment to an attribute for a specific record.

For example, for a data set containing products sold in a store, a record may have:

- For an attribute with the name "Item Name", an assignment to the attribute's value "t-shirt"
- For an attribute named "Color", an assignment to the attribute's value "red"
- For an attribute named "SKU", an assignment to the attribute's value "1234"

See also "attribute", "data set", "index", "schema", "record", and "type (for attributes)".

# Index

## A

actions  
Actions menu, configuring for a component 97  
configuring for displayed values 100  
Results Table, action columns 259  
Results Table, conditions for actions 265  
target page, specifying 102  
Actions menu, configuring for components 97  
aggregation methods  
  EQL syntax 207  
  list of 207  
  selecting for a metric 93  
attribute groups  
  attribute metadata, displaying full 220  
  attribute metadata, selecting columns to display 220  
  attributes, displaying the list of 220  
  creating 221  
  deleting 222  
  displaying on the Attribute Groups page 218  
  linked views 217  
  renaming 222  
  selecting attributes 224  
  sort order, configuring 225  
Attribute Groups page  
  attribute group cache, reloading 225  
  attribute groups, creating 221  
  attribute groups, deleting 222  
  attribute groups, renaming 222  
  attribute groups, saving 225  
  attribute groups, selecting attributes 224  
  attribute metadata, displaying full 220  
  attribute metadata, selecting columns to display 220  
  attributes, displaying for a group 220  
  displaying groups for a view 218  
  sort order, configuring for an attribute group 225  
attributes  
  aggregation methods 207  
  available aggregation methods, selecting 205  
  Available Refinements behavior, configuring 210  
  default aggregation method, selecting 205  
  descriptions, editing and localizing 200  
  display names, editing and localizing 200  
  selecting for components 90  
  storing URLs as values 107  
Available Refinements  
  about 36  
  bookmarks, saved data for 146  
  dates, refining by 44  
  implicit refinements, about 46

linked data sets, controlling automatic filtering 183  
range filters, using for refinement 41  
value list, using for refinement 37

## B

Big Data Discovery, about  
  data scientist's tasks 17  
Big Data Discovery: workflow 18  
Binned Scatter Plot  
  about 271  
Bin Values transformation 171  
bookmarks  
  about 141  
  bookmarks list, displaying 141  
  creating 142  
  deleting 144  
  editing 143  
  emailing 145  
  links, generating 144  
  navigating to 141  
  saved data for components and panels 146  
Box Plot  
  about 282  
  bookmarks, saved data for 146  
  display options, selecting 283  
  refining from 284

## C

Catalog  
  about 62  
  data sets, displaying details 66  
  data sets, previewing 66  
  data sets, sorting 66  
  filtering 63  
  icons on the tiles 65  
  projects, displaying details 66  
  projects, navigating to 132  
  projects, searching 64  
  projects, sorting 66  
  tags, about 67  
  tags, adding 67  
Chart component  
  about 268  
  bookmarks, saved data for 146  
  changing the chart type 275  
  chart values, displaying details 274  
  color palette, configuring 280  
  data selection 276  
  default configuration 275  
  paging 281  
  snapshot, creating 281  
  sorting 280

- trellis display 279  
using to refine data 274
- Component Container**  
about 230  
configuring 230
- components**  
actions, configuring for component values 100  
Actions menu, about 97  
Actions menu, selecting actions 98  
adding 87  
attributes, selecting 90  
Box Plot 282  
Chart 268  
comparing items from 138  
Component Container 230  
deleting 87  
dimension cascades, configuring 104  
displaying Record Details 138  
editing 88  
exporting data 136  
formatting displayed values 93  
Histogram Plot 284  
hyperlinks, configuring external 105  
IFrame 350  
localizing titles 129  
Map 291, 340  
pagination, configuring 108  
paging through 134  
Parallel Coordinates Plot 287  
Pass Parameters action, configuring 99  
Pivot Table 311  
previewing 108  
printing 137  
Refinement Actions menu option, configuring 100  
renaming 88  
Results List 234  
Results Table 243  
Summarization Bar 321  
Tabbed Component Container 231  
Tag Cloud 333  
Timeline 345  
using for refinement 135  
view, selecting 89  
Web Content Display 352
- D**
- data, exporting from projects 226  
data sets  
deleting 80  
deleting from projects 159  
linking, about 176  
links, aggregation methods for linked view 180  
links, creating 176  
links, editing the link key 179  
projects, adding to 159  
tags, about 67  
tags, adding 67
- Discover, portion of Studio 23
- E**
- enrichment, about 21
- EQL**  
aggregation method syntax 207  
predefined metric definitions 212  
view definition guidelines 194
- Explore**  
attributes, filtering 72  
attributes, sorting 72  
attribute values, filtering 74  
attribute visualization, changing 74  
data set, selecting 71  
details panel, displaying 74  
display type, toggling 72  
scratchpad, saving 78  
scratchpad, using 75
- Explore, portion of Studio 21
- F**
- filtering  
how it works 24
- Find, portion of Studio 20
- G**
- Group Values transformation 171  
Guided Navigation  
how it works 24
- H**
- Histogram Plot  
about 285  
bookmarks, saved data for 146  
display options, selecting 285  
refining data from 287
- hyperlinks from components  
encoding 106  
non-HTTP protocol links 107  
URLs as attribute values 107
- I**
- IFrame component  
about 350  
configuring 351  
using 351
- iPad, using to view projects 34
- L**
- layout, changing for a page 85  
locales  
configuring user preferred 32  
selecting from the user menu 32  
supported 31
- logging in 27

**M**

Map component  
 about 292, 341  
 bookmarks, saved data for 146  
 default configuration 301  
 details template, configuring for a layer 307  
 display settings, configuring 310  
 filtering, configuring 310  
 heat map layer, changing the display 297  
 map layer display options, configuring 306  
 map layer name, configuring 306  
 map layer points, defining 304  
 map layers, adding 302  
 map layers, configuring visibility 310  
 map layers, deleting 302  
 map layers, editing 302  
 map layers, renaming 302  
 map layers, selecting the view 303  
 map layers, showing and hiding 295  
 map layers, types of 294  
 map layer type, selecting 303  
 map points, displaying details for 300  
 numbered points list, using 296  
 Oracle MapViewer, configuring the connection to 292, 342  
 pagination, configuring 309  
 point layer, changing the display 297  
 range search, executing 298  
 sorting, configuring 309  
 text search, executing 298  
 view, selecting for a map layer 303

**P**

pages  
 adding 84  
 deleting 86  
 layout, changing 85  
 localizing names 129  
 renaming 84  
 target page, specifying for actions 102

Parallel Coordinates Plot  
 about 288  
 bookmarks, saved data for 146  
 data, selecting 289  
 refining data from 289

Pass Parameters action, configuring 99

password  
 changing 30  
 resetting forgotten 27

Pivot Table component  
 about 311  
 bookmarks, saved data for 146  
 column width, configuring 320  
 conditional formatting, configuring 320  
 dimensions, adding 316  
 dimensions, adding additional available 317  
 dimensions, configuring 318  
 end user configuration, enabling 319  
 end user layout configuration 313

hiding empty rows and columns 320  
 highlighting, showing and hiding 312  
 metrics, adding 316  
 metrics, adding additional available 317  
 navigating in 312  
 sorting dimension values 313  
 summaries, default display 320  
 summaries, showing and hiding 312  
 swapping rows and columns 317  
 table height, configuring 320  
 using for refinement 314

predefined metrics  
 adding to a view 212  
 configuring 213  
 deleting from a view 214

profiling, about 21

project roles  
 assigning to users 126  
 types of 124

projects  
 about 82  
 attribute groups, configuring 216  
 creating from Catalog 82  
 creating from data set 83  
 data sets, adding 159  
 data sets, deleting 159  
 data sets, displaying information about 133  
 deleting 130  
 exporting data 227  
 exporting data, about 226  
 localizing the name and description 129  
 navigating within 132  
 pages, adding 84  
 pages, changing the layout 85  
 pages, data binding 85  
 pages, deleting 86  
 pages, renaming 84  
 project roles, assigning to users 126  
 project type, configuring 123  
 tags, adding 67  
 users, adding 126  
 users, removing 126  
 viewing on an iPad 34  
 views, configuring for project data 185

Project Settings page  
 Attribute Groups page 216  
 Localization page 129  
 Views page 188

project type, configuring for a project 123

**R**

range filters  
 attribute value list, toggling to 43  
 histogram range, changing 43  
 range, selecting 41

Refinement Actions menu option, configuring 100

refinements  
 about 36

- dates, refining by 44  
implicit refinements, about 46  
range filters, using for refinement 41  
value list, using for refinement 37
- Results List component  
about 234  
Actions menu, configuring 242  
attributes, configuring 237  
attributes, selecting to display 236  
bookmarks, saved data for 146  
default configuration 236  
images, configuring 239  
pagination, configuring 243  
Record Details link, enabling 242  
sorting, configuring 238
- results of working with BDD 18
- Results Table component  
about 243  
action columns, available 259  
action columns, configuring the display 263  
action columns, selecting 260  
actions, configuring conditions for 265  
aggregated table, adding dimension columns 254  
bookmarks, saved data for 146  
columns, changing the display order 247  
column sets, adding to an aggregated table 255  
column sets, configuring for a record list table 253  
column sets, selecting on the end user view 245  
columns, formatting 256  
columns, showing and hiding 246  
conditional formatting, configuring 258  
conditional formatting, showing and hiding 248  
data, exporting 250  
default configuration, about 251  
default configuration, enabling and disabling 251  
display options, configuring 267  
Hyperlink action column, configuring 260  
metrics, changing the aggregation method 247  
navigating to other URLs 250  
persistent columns, selecting 252  
record list table, available data 252  
Refinement action column, configuring 262  
sorting, configuring 266  
sorting without summary rows 248  
sorting with summary rows 249  
summary rows, configuring 266  
summary rows, showing and hiding 248  
table type, selecting 251  
types of rows and columns 244  
using to refine data 250
- search  
about 48  
keyword search, match mode 56  
relevance ranking, effect of 57  
results, displaying 48  
results, using 48  
snippets, displaying 56, 57
- Selected Refinements  
bookmarks, saved data for 147  
data set list 58  
linked data sets 60  
negative refinements 59  
refinements, removing 60  
refinement tiles 58  
Selected Refinements panel, hiding 59
- snapshots  
creating 149  
deleting 152  
details, displaying 151  
editing 151  
emailing all 153  
emailing individual 152  
galleries, creating 154  
galleries, deleting 155  
galleries, editing 155  
galleries, emailing 157  
galleries, exporting to PDF 157  
galleries, full screen mode 156  
refinement state, navigating to 152  
saving all to a PDF 153  
saving to a file 152  
snapshots panel, displaying 148
- Split transformation 170
- Studio  
user interface, Discover 23  
user interface, Explore 21  
user interface, Find 20  
user interface, Transform 22
- Summarization Bar component  
about 321  
action, configuring for summary items 328  
bookmarks, saved data for 147  
conditional formatting, configuring 329  
default configuration 323  
dimension value, selecting for a dimension spotlight item 326  
display name, configuring for summary items 328  
display name location, configuring 333  
flag items, configuring 330  
item type, selecting for a summary item 324  
metric value, selecting for a dimension spotlight item 326  
metric value, selecting for a metric summary item 325  
summary items, adding 323  
summary items, deleting 323  
summary items, editing 323  
text size, configuring 333  
tooltip, configuring for summary items 328

## S

- Scatter  
bin values, displaying 278

using 321  
width, configuring for summary items 328

## T

Tabbed Component Container  
about 231  
bookmarks, saved data for 147  
configuring 232

Tag Cloud component  
about 334  
bookmarks, saved data for 147  
dimensions, configuring 337  
dimensions, selecting available 337  
display, configuring 340  
display type, selecting 335  
metrics, configuring 339  
metrics, selecting available 338  
values, selecting to display 336

Thematic Map component  
attribute, selecting 343

Timeline component  
about 345  
bookmarks, saved data for 147  
date range, setting 348  
date unit, setting 348  
default configuration, saving 349  
metric timelines, adding 346  
metric timelines, deleting 346  
metric timelines, editing 346  
metric timelines, reordering 346  
refinement, using for 348

Transform  
about 161  
attributes, adding custom 172  
attributes, favorite 163  
attributes, filtering 163  
attributes, hiding 163  
binning values 171  
capitalization, changing for strings 169  
custom transformation, creating 172  
data grid, toggling the display 163  
data set, creating new 175  
data set, forking 175  
data set, selecting 162  
data set size determining 162  
data type, changing for an attribute 168  
date/time values, truncating 170  
deleting records 164  
grouping values 171  
numeric transformations, performing 169  
page header, about 162  
splitting values 170  
transformation, adding attribute 166  
transformation, configuring 166  
Transformation Editor, using 172  
transformation script, applying 174  
transformations, deleting 173  
transformation, selecting 166  
transform script, reordering 173

trimming spaces 169  
whitespace characters, displaying 168

Transform, portion of Studio 22

Trim transformation 169

## V

views  
attribute aggregation methods, selecting  
available 205  
attribute aggregation methods, selecting  
default 205  
attribute descriptions, editing and  
localizing 200  
attribute display names, editing and  
localizing 200  
attributes, default display format 203  
Available Refinements behavior for attributes,  
configuring 210  
base views, about 186  
copying 190  
creating a new view 190  
date/time subsets, configuring available 201  
deleting 215

description, editing and localizing 192  
displaying for a project 188  
display name, editing and localizing 192  
editing the definition 193  
EQL guidelines 194  
linked views, about 186  
predefined metrics, adding 212  
predefined metrics, configuring 213  
predefined metrics, deleting 214  
previewing 214  
publishing custom 214  
selecting for a component 89

Views page  
attribute aggregation methods, selecting  
available 205  
attribute aggregation methods, selecting  
default 205  
attribute descriptions, editing and  
localizing 200  
attribute display names, editing and  
localizing 200  
attributes, default display format 203  
attributes list columns 198  
Available Refinements behavior,  
configuring 210  
date/time subsets, configuring available 201  
displaying 188  
predefined metrics, adding to a view 212  
predefined metrics, configuring 213  
predefined metrics, deleting 214  
view definition, editing 193  
view description, editing and localizing 192  
view display name, editing and localizing 192  
views, copying 190  
views, creating new 190  
views, deleting 215  
views, displaying details 189

views, previewing 214  
views, publishing custom 214

about 353  
configuring 353  
workflow  
in Big Data Discovery 18

## W

Web Content Display component