Oracle® Retail Advanced Science Engine

User Guide

Release 14.1

E59125-01

December 2014

ORACLE®

Oracle Retail Advanced Science Engine User Guide, Release 14.1

E59125-01

Primary Author: Judith Meskill

reconfigurations, reassembly or reverse assembly, re-engineering or reverse engineering and recompilations or reverse compilations of the VAR Applications or any derivatives of the VAR Applications. You acknowledge that it shall be a breach of the agreement to utilize the relationship, and/or confidential information of the VAR Applications for purposes of competitive discovery.

The VAR Applications contain trade secrets of Oracle and Oracle's licensors and Customer shall not attempt, cause, or permit the alteration, decompilation, reverse engineering, disassembly or other reduction of the VAR Applications to a human perceivable form. Oracle reserves the right to replace, with functional equivalent software, any of the VAR Applications in future releases of the applicable program.

Contents

3 Demand Transference

4 Advanced Clustering

Send Us Your Comments

Oracle Retail Advanced Science Engine User Guide, Release 14.1

Oracle welcomes customers' comments and suggestions on the quality and usefulness of this document.

Your feedback is important, and helps us to best meet your needs as a user of our products. For example:

- Are the implementation steps correct and complete?
- Did you understand the context of the procedures?
- Did you find any errors in the information?
- Does the structure of the information help you with your tasks?
- Do you need different information or graphics? If so, where, and in what format?
- Are the examples correct? Do you need more examples?

If you find any errors or have any other suggestions for improvement, then please tell us your name, the name of the company who has licensed our products, the title and part number of the documentation and the chapter, section, and page number (if available).

> **Note:** Before sending us your comments, you might like to check that you have the latest version of the document and if any concerns are already addressed. To do this, access the Online Documentation available on the Oracle Technology Network Web site. It contains the most current Documentation Library plus all documents revised or released recently.

Send your comments to us using the electronic mail address: retail-doc_us@oracle.com

Please give your name, address, electronic mail address, and telephone number (optional).

If you need assistance with Oracle software, then please contact your support representative or Oracle Support Services.

If you require training or instruction in using Oracle software, then please contact your Oracle local office and inquire about our Oracle University offerings. A list of Oracle offices is available on our Web site at http://www.oracle.com.

Preface

This guide describes the Oracle Retail Advanced Science Engine user interface. It provides step-by-step instructions to complete most tasks that can be performed through the user interface.

Audience

This User Guide is intended for retailers and analysts.

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.

Related Documents

For more information, see the following documents in the Oracle Retail Modeling Engine documentation set:

- *Oracle Retail Advanced Science Engine Installation Guide*
- *Oracle Retail Advanced Science Engine Release Notes*
- *Oracle Retail Advanced Science Engine Implementation Guide*
- *Oracle Retail Advanced Science Engine Security Guide*

Customer Support

To contact Oracle Customer Support, access My Oracle Support at the following URL:

https://support.oracle.com

When contacting Customer Support, please provide the following:

- Product version and program/module name

- Functional and technical description of the problem (include business impact)
- Detailed step-by-step instructions to re-create
- Exact error message received
- Screen shots of each step you take

Review Patch Documentation

When you install the application for the first time, you install either a base release (for example, 14.1) or a later patch release (for example, 14.1.1). If you are installing the base release or additional patches, read the documentation for all releases that have occurred since the base release before you begin installation. Documentation for patch releases can contain critical information related to the base release, as well as information about code changes since the base release.

Improved Process for Oracle Retail Documentation Corrections

To more quickly address critical corrections to Oracle Retail documentation content, Oracle Retail documentation may be republished whenever a critical correction is needed. For critical corrections, the republication of an Oracle Retail document may at times not be attached to a numbered software release; instead, the Oracle Retail document will simply be replaced on the Oracle Technology Network Web site, or, in the case of Data Models, to the applicable My Oracle Support Documentation container where they reside.

This process will prevent delays in making critical corrections available to customers. For the customer, it means that before you begin installation, you must verify that you have the most recent version of the Oracle Retail documentation set. Oracle Retail documentation is available on the Oracle Technology Network at the following URL:

http://www.oracle.com/technetwork/documentation/oracle-retail-100266.html

An updated version of the applicable Oracle Retail document is indicated by Oracle part number, as well as print date (month and year). An updated version uses the same part number, with a higher-numbered suffix. For example, part number E123456-02 is an updated version of a document with part number E123456-01.

If a more recent version of a document is available, that version supersedes all previous versions.

Oracle Retail Documentation on the Oracle Technology Network

Documentation is packaged with each Oracle Retail product release. Oracle Retail product documentation is also available on the following Web site:

http://www.oracle.com/technetwork/documentation/oracle-retail-100266.html

(Data Model documents are not available through Oracle Technology Network. These documents are packaged with released code, or you can obtain them through My Oracle Support.)

Documentation should be available on this Web site within a month after a product release.

Conventions

The following text conventions are used in this document:

Convention	Meaning
boldface	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
italic	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
`monospace`	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.

1

Getting Started

This chapter provides an overview of Oracle Retail Advanced Science Engine (ORASE). It includes the following sections:

About Oracle Retail Advanced Science Engine

The Oracle Retail Advanced Science Engine is an analytical product that consists of three modules: Customer Decision Trees, Demand Transference, and Advanced Clustering. You may have access to all three of these modules or to a subset of them.

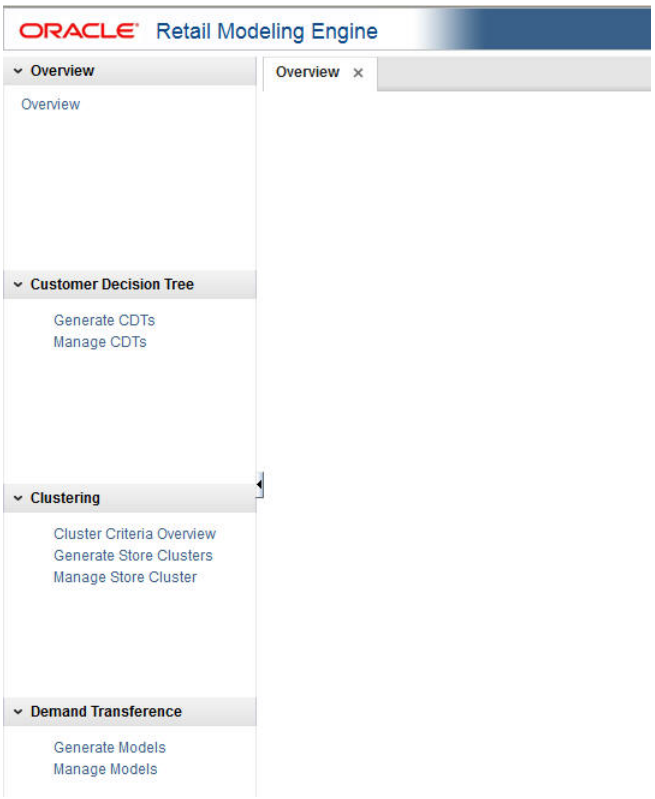
Installing and Configuring ORASE

For information about the installation of the ORASE, see *Oracle Retail Advanced Science Engine Installation Guide*.

Overview

Once you log into ORASE, you see an interface that is divided into three panels. The following figure shows the task list panel.

Figure 1–1 ORASE Task Panel



The panel on the left lists the modules you have access to, and for each module, a list of the main tabs for the product. You can access the tabs from this panel.

The center panel is where you do all the work within the module. The work is organized into a series of five stages. A process trail at the top of the panel provides access to each stage.

The panel on the right is used for supplementary information that can help you as you work through the stages. For example, certain stages include histograms that can assist you in analyzing the calculation results of a specific stage. Other stages include additional business details that can assist you in understanding the information in a specific stage.

Process

This guide contains a chapter for each of the three modules with detailed instructions. In general, each module contains three tabs, an overview tab, a generation tab, and a management tab. The overview tab allows you to monitor the status of your work. The generation tab is the area where you progress through a series of stages in which you configure parameters and run calculations that inform your analysis. The management tab allows you to manage certain tasks that are outside the main application workflow.

Customer Decision Trees

When you use the CDT module, you following this general iterative process to create and manage CDTs:

- Setup. You define the domain of calculation by selecting the categories and time frame for the calculation of the CDTs.

- Data Filtering. You configure filters that remove input data that might cause errors in the calculation or that might lead to inaccurate or unreliable answers.
- Calculation. You calculate a specific version of a CDT. Different versions of the CDT based on different configurations can be calculated and compared.
- Evaluation. You examine the results of the calculation and determine the reliability and accuracy of the answers. You can prune inaccurate or unreliable results.
- Escalation. When you are satisfied with the results, you can set the escalation path to fill in the holes for partitions whose CDTs were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy. Then, you can set a version of a CDT as complete.

Within each stage, you set the required parameters. In some stages, once you do that you click **Run** to initiate the process. To determine when a run is complete, go to the Overview tab and click the **Refresh** icon in order to see the current status of each stage. If you make changes to a stage, you must re-run that stage and all the relevant stages that follow it as the results of those stages are made invalid by the changes you made.

Demand Transference

When you use the DT module, you follow this general iterative process to create and manage DT models:

- Data Setup. You define the categories to be used in the DT calculation.
- Data Filtering. You configure filters that remove input data that might cause errors in the calculation or that might lead to inaccurate or unreliable results.
- Similarity Calculation. You calculate similarities and assess the results of the calculation.
- Elasticity Calculation. You calculate the assortment elasticities and assess the results in terms of substitutable demand, which is the percentage of demand of a SKU that is retained when the SKU is deleted from the stores where it is selling.
- Escalation. When you are satisfied with your results, you can set the escalation path you can set the escalation path to fill in the holes for partitions whose DT models were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy. Then you can set a version of the DT model as complete.
- Manage Models. Use this tab to set time intervals for evaluating your results and to override the value for the maximum substitutable demand percentage.

Advanced Clustering

When you use the Advanced Clustering module, you follow this general iterative process to create and manage clusters:

- Cluster Criteria. You define the criteria for the cluster. This information includes the type of data used to define the cluster, the nodes, and the time period.
- Explore Data. You can view the data for the cluster you defined in this stage.
- Cluster Setup. You create various scenarios based on a specified number of clusters and execute those scenarios.
- Cluster Results. You can view the results of the scenarios you executed and approve or reject clusters based on those results.
- Cluster Insights. You can view a summary of the information for each scenario.

Overview of the User Interface

The user interface includes functionality described in this section.

Process Train

You can use the process train to navigate through the stages of each module. You can also use the **Back** button and the **Next** button to move through the train. The color changes for each stage once you visit that stage. Certain stages require you to run that stage before you can go to the next stage.

View Menu

The View Menu provides access to a variety of functionality that you can use to customize the display of the tables in the user interface.

Embedded Help

Embedded help, which you access by clicking the Question Mark icon, provides additional information about the type of details required by certain fields.

Process Indicator

At the top of the user interface, in the right-hand corner, is a process indicator that you can use to monitor the status of a user action such as clicking Next to go to the next stage.

Search

In certain cases, you can customize your search, using advanced search capabilities to specify the search criteria.

Icons

The following icons are used in ORASE.

Table 1–1 Icons Used in ORASE

Icon	Icon Name	Icon Description
	Add	Add a category for CDT or DT. Add a cluster in Clustering. Create a new scenarios in Clustering.
	Approve Version	Approve a version of a CDT.
	Calculate	Initiate calculation of substitutable demand percentages in DT.
	Compare Two CDTs	Look at two CDTs side-by-side in the CDT Editor.
	Complete	Indicates the CDT is ready to be activated.
	Delete an entry in a table. Delete a scenario in Clustering.	Delete

Table 1–1 (Cont.) Icons Used in ORASE

Icon	Icon Name	Icon Description
	Detach	Detach the table from the user interface for better viewing.
	Duplicate	Make a copy of a scenario.
	Edit	Edit the category attributes (CDT and DT)
	Embedded Help	Indicates that embedded help is available for the adjacent field.
	Execute	Execute the scenario.
	Export to Excel	Export the selected data to Excel.
	Go To Top	Adjusts table display.
	Go Up	Adjusts table display.
	No	Indicates No in Clustering.
	Query By Example	Provides access to a text entry field at the top of each column that you can use to search for data by an initial set of characters.
	Refresh	Update the table display or the stage status.
	Revert	Reverts the DT calculation.
	Save	Save the scenario.
	See Similarities	See similarities in DT Similarities Display.
	Set Version As Complete	Set a CDT version as complete.
	Select Date	Access a calendar in order to select a specific date.
	Show As Top	Adjusts table display.
	View One CDT	Access a CDT in the CDT Editor.
	Withdraw From Approved Version	Un-approve a version of a CDT.
	Yes	Acts as an indication that something exists in Clustering.

Buttons

Buttons are used for navigation and to perform certain actions.

Table 1–2 Buttons

Name	Description
Action	Provides access to Save, Approve, and Reject.
Advanced	Provides access to advanced search functionality.
Approve	Used to approve a CDT or a cluster.
Back	Used to navigate the process train.
Cancel	Cancels the action.
Complete	Used to make a CDT or DT model active.
Next	Moves to the next stage.
Reject	Used to navigate the process train.
Reset	Resets the values to the original ones.
Run	Initiates a run.
Search	Provides access to search functionality.
Stop	Stops a process.

Browser Settings

The supported browsers include Internet Explorer 11 and Google Chrome 28 or newer.

Concurrent Browser Sessions

Users should not log into more than one browser session at the same time using the same user name.

Localization

The default language for the application is English. If you are using a different language on your computer, you should adjust the language settings on your browser as appropriate.

Setting Up Internet Explorer

To configure Internet Explorer:

1. Open Internet Explorer.
2. From the **Tools** menu, select **Internet Options**.
3. From the **Internet Options** dialog box, click the **Security** tab.
4. From the **Security** tab, click **Local intranet**, or, if you have been instructed to do so by your Systems Administrator, **Trusted sites**, and then click the **Sites** button.

   If you selected Local intranet, go to step 5. If you selected Trusted sites, go to Step 6.
5. On the **Local Intranet** dialog box, click the **Advanced** button.
6. On the resulting **Local intranet** or **Trusted sites** dialog box, add the application URL if it is not already listed.

To do so, type the application URL in the **Add this Web site to the zone** text box. Click **Add**. When the URL appears in the Web sites list, click **OK**.

7. If the **Local Intranet** dialog box from step 5 is still open, click **OK** to close it.
8. Based on the selection your made in step 4, from the **Security Tab** of the **Internet Options** dialog box, select either **Local intranet** or **Trusted sites**. Click the **Custom Level** button.
9. In case you have Pop-up Blocker enabled, add the host name from the application URL as an exception using the following steps:
   a. On the **Internet Options** dialog box, click the **Privacy** tab.
   b. On the **Privacy** tab, in the **Pop-up Blocker** section, click **Settings**.
   c. On the **Pop-up Blocker Settings** dialog box, enter the host name in the **Address of website to allow** field, and click **Add**.
   d. Click **Close**.
10. On the Internet Options dialog box, click **OK** to return to the browser.

Supported Characters for Text Entry

The following characters are valid for text input: all letters, all numbers, and the following characters: '_', '#', '%', '*', '$', ' ', ',' & '-'

Login

Once the application is installed, you can access the application using the following URL, which redirects you to the login page:

http://<address>:<port>/cdm/faces/oracle/retail/rse/cdm/fe/view/page/CentralizedDemandModelling.jspx

To log into the application, enter the user name and password assigned to you during the installation procedure. See the installation guide for details.

After logging in to the application, you will be redirected to the main application page, located at this URL. This URL may be bookmarked for later use and navigated to directly.

Histograms

Certain stages have associated histograms that can help you analyze the data presented in that stage. You can adjust the way the histogram presents the data in two ways. You can select the number of bins that are used to display the data. In addition, you can select how the bins are defined: Equiwidth or Custom. Each of these options uses a specific algorithm to determine how the bins are defined.

The Equiwidth approach takes the minimum and maximum values in a set of numbers and divides that range into equally sized bins. For example, using the numbers from 1 to 100 with 10 bins, the histogram shows bins for 1-10, 11-20, and so on. If specific bins have no value represented (for example, if all the values are in the range of 1-10 and 91-100), then the histogram will not show that bin in the UI. Additionally, the histogram data series ranges are shown using the actually minimum and maximum values for each of the bins. So rather than showing a range of 1-10, if the only value available was a 5, then the range for the first bin would appear as 5-5.

In the Custom approach, each of the bins has an equal number of values represented, while the minimum and maximum number associated with the bin is adjusted.However, the bins are defined using distinct values instead of all the available values. The bins may or may not be of equal height, depending on how diverse the numbers are.

The two approaches differ in that the Custom approach only shows fewer bins than requested if there are fewer distinct values than what was requested for the number of bins.

The two approaches are similar in that both handle the Min/Max value display in a similar manner, using actual data values that are associated with the bin.

To determine which approach to use, you should consider what type of data you are trying to see and the amount of detail you want. For example, if you are trying to set a data filter value, and you want to do so using a common value, you may be able to see where the majority of the data falls using one of the algorithms, while the other algorithm may help you pinpoint a specific value within the range. The Equiwidth approach is negatively affected by values that are at the extreme ends of a value being binned. This can cause the majority of the data values to appear in a single bin. The Custom approach puts a greater emphasis on a value that is repeatedly found in a dataset. Depending on the values being charted, you may find that one of the approaches presents better data than the other approach.

2

Customer Decision Trees

This chapter describes the use of the Customer Decision Tree (CDT) module.

Introduction

A customer decision tree is a decision support tool that uses a tree-like graph to model a customer hierarchical decision-making process for a specific product. The branches of the tree provide a visual representation of the choices a customer makes and the order of importance of various product characteristics. Transaction data is used in the analysis.

A customer decision tree identifies the decisions a customer makes when choosing a particular product. The decision tree is produced by algorithms that analyze historical customer sales data. It illustrates how customers shop and how they evaluate the importance of different product attributes when making buying decisions. Such information can be useful to a retailer in terms of product selection and display.

The CDT module consists of three tabs: Overview, Generate CDTs, and Manage CDTs. You use the Overview tab to keep track of the status of each stage within the Generate CDTs tab. You use the Manage CDTs tab to assess the set of CDTs you and other users have created using the Generate CDTs tab, setting versions as complete, approving versions, and deleting versions, as necessary.

Overview of CDT Process

When you use the CDT module, you follow this general iterative process to create and manage CDTs:

- Setup. You define the domain of calculation by selecting the categories and time frame for the calculation of the CDTs.
- Data Filtering. You configure filters that remove input data that might cause errors in the calculation or that might lead to inaccurate or unreliable answers.
- Calculation. You calculate a specific version of a CDT. Different versions of the CDT based on different configurations can be calculated and compared.
- Evaluation. You examine the results of the calculation and determine the reliability and accuracy of the answers. You can prune inaccurate or unreliable results.
- Escalation. When you are satisfied with the results, you can set the escalation path to fill in the holes for partitions whose CDTs were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy. Then, you can set a version of a CDT as complete.

Within each stage, you set the required parameters. In some stages, once you do that you click **Run** to initiate the process. To determine when a run is complete, go to the Overview tab and click the **Refresh** icon in order to see the current status of each stage. If you make changes to a stage, you must re-run that stage and all the relevant stages that follow it as the results of those stages are made invalid by the changes you made.

Overview Tab

The Overview tab displays information that you can view and use to monitor the progress of the CDT stages as well as to view some aggregate statistics and the CDT results from the last successful run.

This tab contains the following sections:

- CDT Generation Stage Status
- Aggregate Statistics
- CDT Calculation Report
- CDT Score Distribution

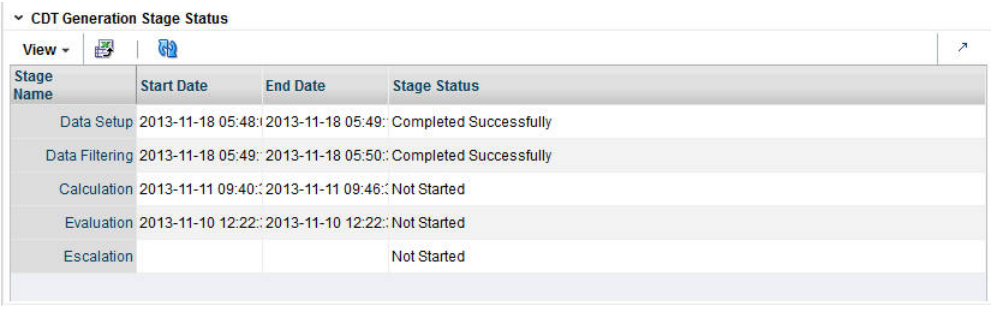
CDT Generation Stage Status

The CDT Generation Stage Status table displays the status of each of the CDT stages.

Click the **Refresh** icon to update the fields and see the latest status for each stage.

In cases where the status is either Failed or Stopped/Paused, or the run never completes, you should consult the database logs in RSE_LOG_MSG as well as the WebLogic console logs in your WebLogic domain in order to troubleshoot the problem.

Figure 2–1 CDT Generate Stage Status



You can use the status information to monitor the progress of each stage. It contains the following fields, which can be arranged and viewed, but not modified.

Table 2–1 CDT Generation Stage Status Fields

Field Name	Description
Stage Name	A row exists in the table for each CDT stage that provides detailed status information. The five stages are Setup, Data Filtering, Calculation, Evaluation, and Escalation.
Start Date	The date and time when a run for the stage most recently started.
End Date	The date and time when a run for the stage most recently ended.

Table 2–1 (Cont.) CDT Generation Stage Status Fields

Field Name	Description
Stage Status	The current status of the stage: Not Started, Not Started (Scheduled for Later), Processing, Completed Successfully, Completed with Errors, Stopped/Paused, Cancelled, or Failed.

The following provides an explanation of the different values for the status of a stage.

Table 2–2 Status Explanations

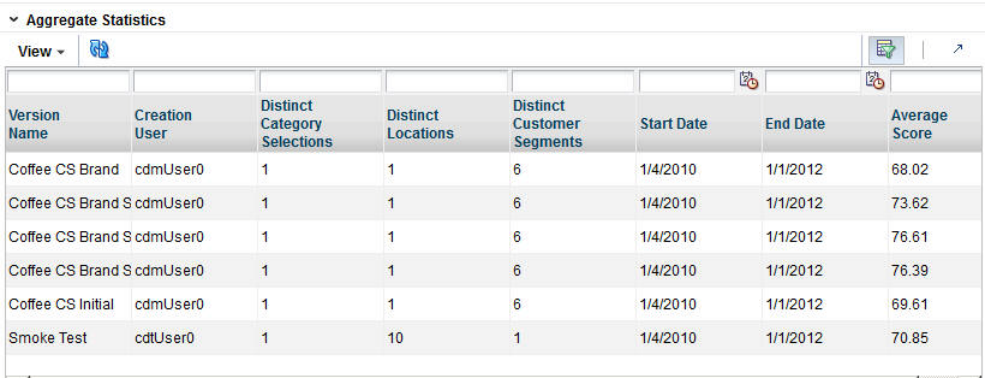
Stage Status	Description
Not Started	This indicates that the stage has not yet been initiated. This can occur when you first begin to create a CDT or when you update an existing version.
Not Started (Scheduled for Later)	The stage is scheduled to start after the previous stage is complete.
Processing	This indicates that the stage is currently being processed.
Stopped/Paused	This indicates either that the stage was stopped by the application or that the user has chosen to stop the process (for example, to make a change to an option). In the later case, once any background processing has stopped, the user can re-run the stopped stage.
Completed Successfully	This indicates that the stage has been successfully processed.
Completed with Errors	This indicates that the stage was able to complete the processing of all requested CDTs, but one or more CDTs encountered errors during the run and were not able to complete successfully. This most commonly occurs when the data used for a CDT is not available or is too sparse to produce a result.
Failed	This indicates that a problem occurred during the processing. If the failure occurs during the Calculation stage, and only a few CDT results did not produce a CDT score, then the user can continue without those results since pruning will remove them. If the failure occurs during any of the other stages, the user must re-run the stage.
Cancelled	This indicates that the database has cancelled the execution of the stage either because of missing data or an exception. The user should review the RSE_LOG_MSG log to determine the problem.

Aggregate Statistics

The Aggregate Statistics table displays statistical details about the existing CDT versions.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 2–2 CDT Aggregate Statistics



Version Name	Creation User	Distinct Category Selections	Distinct Locations	Distinct Customer Segments	Start Date	End Date	Average Score
Coffee CS Brand	cdmUser0	1	1	6	1/4/2010	1/1/2012	68.02
Coffee CS Brand S	cdmUser0	1	1	6	1/4/2010	1/1/2012	73.62
Coffee CS Brand S	cdmUser0	1	1	6	1/4/2010	1/1/2012	76.61
Coffee CS Brand S	cdmUser0	1	1	6	1/4/2010	1/1/2012	76.39
Coffee CS Initial	cdmUser0	1	1	6	1/4/2010	1/1/2012	69.61
Smoke Test	cdtUser0	1	10	1	1/4/2010	1/1/2012	70.85

Each existing version has a row in the table. The table contains the following fields, which can be arranged and viewed but not modified.

Table 2–3 Aggregate Statistics Fields

Field Name	Description
Version Name	User-created name that identifies this CDT version. The name must be unique to the user.
Creation User	The login name of the person who created this version.
Distinct Category Selections	The number of categories associated with this version.
Distinct Locations	The number of locations associated with this version.
Distinct Customer Segments	The number of customer segments associated with this version.
Start Date	The beginning date associated with the CDTs in this version.
End Date	The end date associated with the CDTs in this version.
Average Score	The average of the CDT scores for a version. A CDT score ranges from 0 to 100. A higher value indicates a better score.

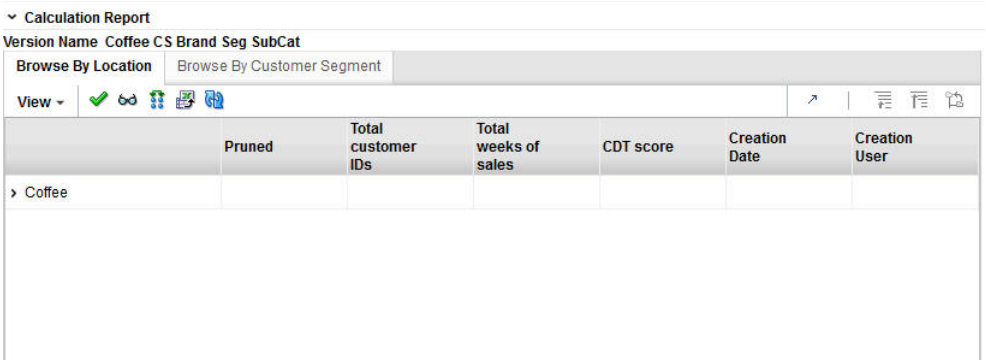
Calculation Report

The Calculation Report displays the CDT results for the indicated version from the last successful run, if one has occurred.

Here you can review information by Location or by Customer Segment. You can also set a version as complete, view and edit a CDT, or compare two CDTs.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 2–3 CDT Calculate Report



The Calculation Report contains the following fields:

Table 2–4 Calculation Report Fields

Field Name	Description
Partition	A partition is the combination of category, segment, and location. The column identifies the names of all nodes in the tree structure in the table. The node type may be either category, location, or customer segment.
Pruned	The number of CDTs removed from the list of usable CDTs. CDTs are removed that do not meet the filtering thresholds. During the Escalation stage, the escalation process makes adjustments for the CDTs that are removed.
Total Customer IDs	The number of customers used in the calculation of the CDT.
Total Weeks of Sales	The total number of weeks of sales used in the calculation of the CDT. This provides an indicates of the amount of sales data used in the calculation. All the CDTs in a given version should have a similar value; if this is not the case, the results should be evaluated.
CDT Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Creation Date	The date when the version whose data is displayed was created.
Creation User	The login name of the user who created the version.

The following icons are available for the Calculation Report:

Set Version as Complete

Each partition must have an active CDT. The active CDT is used by other applications that require a CDT.

All CDTs within a version are activated when you click **Set Version as Complete**.

If a CDT exists in more than one version then the most recently activated version takes precedence.

If a version contains multiple categories and is Complete and then a different version with a partial overlap of categories is later marked as Complete, then only the overlapping categories are replaced in the new version.

If a version is overwritten then the CDTs in that version are no longer active.

When a CDT is at risk of going from active to inactive, you will see a warning message.

To determine whether or not a CDT is active, go to the Manage CDTs tab. The Browse by Categories or by Versions table displays a flag that indicates whether or not a CDT is active.

View One CDT

This icon provides access to the CDT Editor. For details about this functionality, see Using the CDT Editor.

Compare Two CDTs

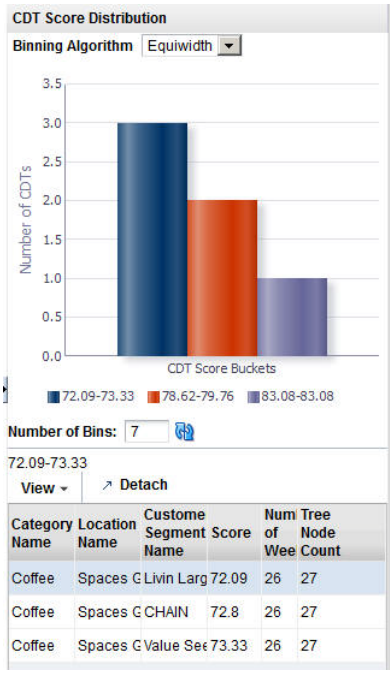
This icon provides access to CDT Compare. For details about this functionality, see Comparing Two CDTs

CDT Score Distribution Histogram

The CDT Score Distribution histogram displays, for the current calculation and the current user, the range of CDT scores and their frequencies. CDT scores range from 0 to 100; a higher number indicates a better score. You can use the information in this histogram to assess the quality of your CDTs.

For detailed information about using the histograms, see Chapter 1, "Getting Started.".

Figure 2–4 CDT Score Distribution



Below the histogram you see a table that provides details about the histogram. This information can help you assess the quality of a CDT. Click on a specific bin in the histogram to populate the table with information about that bin. You see a list of all the specific partitions for the bin, along with the score for each, the number of weeks of data used to calculate the score, and the number of tree nodes for each partition for the selected bin.

Table 2–5 CDT Score Distribution Fields

Field Name	Description
Category Name	The category name of the partition for the specific CDT score.
Location Name	The location name of the partition for the specific CDT score.
Customer Segment Name	The customer segment name of the partition for the specific CDT score.
Score	The CDT score for the defined partition. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Number of Weeks	The number of weeks of data used to calculate the CDT score.
Tree Node Count	The number of tree nodes present in the CDT tree structure used to calculate the CDT score.

Generate CDT Tab

The Generate CDTs tab is used to configure, run, evaluate, modify, and deploy a CDT. The process is divided into five stages that must be run in order. You can return to a stage you have already completed and make changes, but if you do, you must rerun that stage and all the stages that follow that stage, as the calculations are invalidated by the modifications you just made to the settings in the stage you changed.

To monitor the progress of any stage, go to the Overview tab and click the **Refresh** icon.

The five CDT stages are:

Table 2–6 Generate CDTs Tab: Stages

Stage Name	Description
Data Setup	Select CM groups and define the time intervals for the data to be used in the calculation.
Data Filtering	Filter out input data that may result in inaccurate or unreliable answers.
Calculation	Calculate the CDT.
Evaluation	Assess the reliability and accuracy of the results of the calculation. Prune unreliable or inaccurate answers.
Escalation	Set the escalation path for the CDT. Use the Escalation Report to evaluate the results of the escalation.

Data Setup

The Data Setup stage provides two sections to configure: Category Selections and Time Interval Setup. In this stage, you select the category or categories you want to calculate CDTs for and specify the time interval for the calculation.

Process

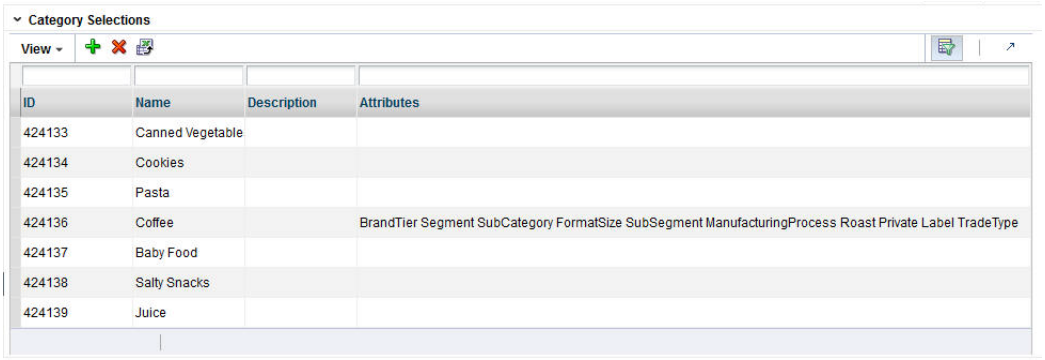
Here is the high-level process for setting up the data for CDT.

1. Select the category or categories you want to calculate CDTs for.
2. Set up the time intervals for the CDT calculation.
3. Click **Next** to go to the Data Filtering stage.

Category Selections

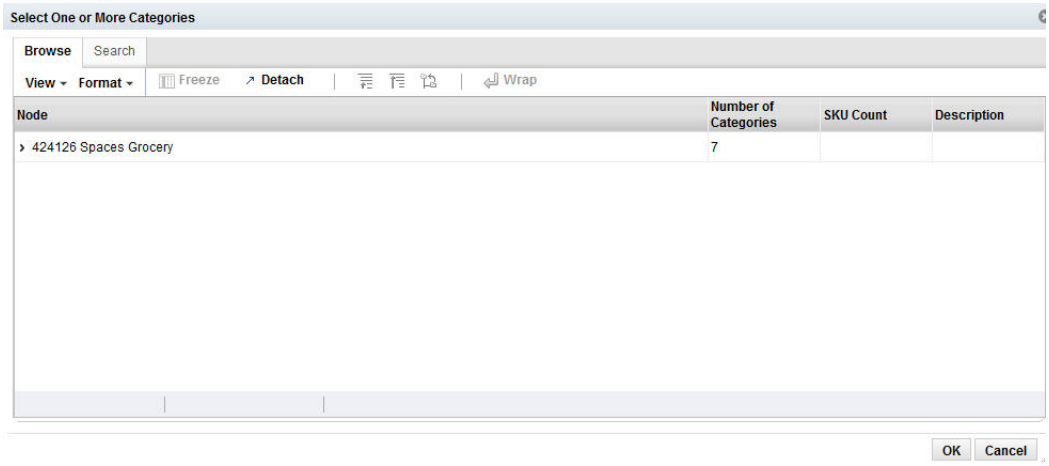
Use this table to add categories you want to include in the CDT calculation or delete categories that you want to remove from the CDT calculation.

Figure 2–5 CDT Category Selections



To display a list of available nodes and the categories included in those nodes, click the **Add** icon. You see the Select One or More Categories dialog box, which contains two tabs, Browse and Search. You can use either tab to find the categories you are looking for.

Figure 2–6 CDT Categories Browse



The Browse tab displays a table with the following fields:

Table 2–7 Category Selections: Browse

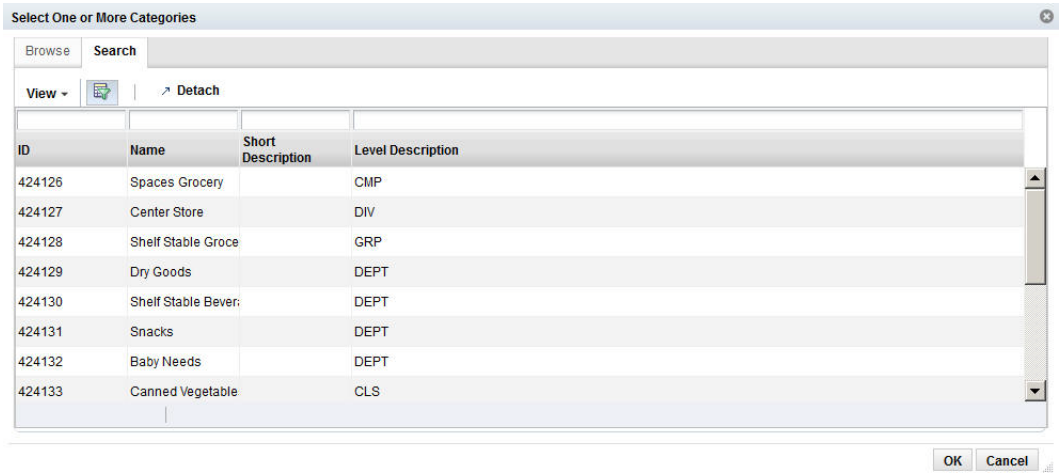
Field Name	Description
Partition	The node tree structure can be expanded in order to view its categories.
Number of Categories	The number of categories within the node that has been selected. The number can help you understand the amount of processing required for the calculation.
SKU Count	The number of SKUs in a category. A category with too few SKUs may not produce good CDTs.

Table 2–7 (Cont.) Category Selections: Browse

Field Name	Description
Description	A description that provides additional information about the category.

Select the category or categories within the node that you want to be part of CDT and click **OK**.

Figure 2–7 CDT Categories Search



The Search tab displays a table with the following fields:

Table 2–8 Category Selections: Search

Field Name	Field Description
ID	An external code used to identify the category in other systems such as Category Management.
Name	The category name.
Short Description	A description that provides additional information about the category.
Level Description	A description of the level of the merchandise hierarchy that the node belongs to.

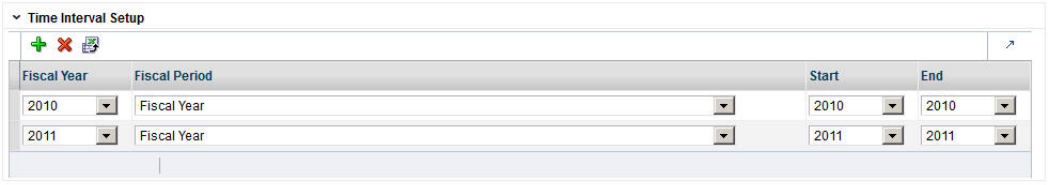
Select the category or categories within the node that you want to be part of CDT and click **OK**.

Your selections are displayed in the Category Selections table.

Time Interval Setup

The time interval defines the time span for the historical data that is used to calculate the CDT. Only the specified weeks of sales data are used in the calculation. A group of intervals can be defined. Gaps between intervals are permitted; however, intervals cannot overlap. A six-month period is recommended.

Figure 2–8 CDT Time Interval Setup



To define a time interval for the CDT to be generated, click the **Add** icon. A row displays in the dialog box. Select from the following drop-down menus in order to define the time interval:

Table 2–9 Time Interval: Menus

Menu Name	Description
Fiscal Year	The fiscal year for the time interval.
Fiscal Period	The fiscal period within the fiscal year (Fiscal Quarter, Fiscal Period, or Fiscal Week).
Start	This defines when the time interval specified in Fiscal Period begins.
End	This defines when the time interval specified in Fiscal Period ends.

After selecting the category and specifying the time interval, click **Next** to go to the Data Filtering stage.

Data Filtering

The Data Filtering stage applies to all the categories and time intervals that you select in the Data Setup stage.

Process

Here is the high-level process for setting up and running data filtering.

1. Enter the appropriate values into the Filter Setup text entry boxes.
2. Click **Run** in order to filter the data.
3. Review the filtering results in the Data Filtering Summary table and the Data Filtering histograms.
4. After reviewing the results, if necessary, make changes to the values for the filters in Filter Setup and re-run the stage.
5. When you are satisfied with the results, click **Next** to go to the Calculation stage.

Filter Setup

You configure the following filters in order to filter out data you consider unacceptable for the calculation of the CDT. Note that the two attribute filters listed in Table 2–10 are stored and used during the Calculation stage. You also set additional data filtering parameters in the Calculation stage.

Figure 2–9 CDT Filter Setup

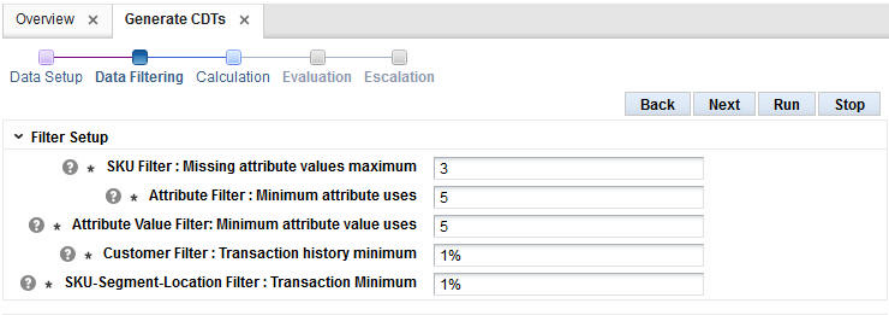


Table 2–10 Data Filters

Data Filter Name	Data Filter Description
SKU Filter: Missing attribute values maximum	Each SKU is defined by its attribute values. If a certain percentage of the attribute values are not defined, then the product definition is not accurate. A SKU with too many missing attribute values should be filtered out. The default value is 25% of the total attribute values (that is, a SKU with greater than 25% missing attribute values is not included in the calculation of the CDT).
Attribute Filter: Minimum attribute uses	An attribute that is used by only a few SKUs should be filtered out. The default value is 5% of the total SKUs in the category (that is, The data for an attribute that is used by fewer than 5% of the SKUs is not included in the calculation of the CDT).
Attribute Value Filter: Minimum attribute value uses	An attribute value that is used by only a few SKUs should be filtered out. The default value is 5% of the SKUs in a category (that is, the data for an attribute value that is used by fewer than 5% of the SKUs is not included in the calculation of the CDT).
Customer Filter: Transaction history minimum	Customers with short transaction histories are considered outliers. You assign a percentage value that is applied to the median number of transactions for all customers. Such customers are filtered out. The default value is 10% (that is, a customer who has fewer than 10% of the median number of transactions for all customers is not included in the calculation of the CDT).
SKU-Segment-Location Filter: Transaction minimum	SKUs that have few transactions for a given location-segment partition are considered outliers. You assign a percentage value that is applied to the median number of transactions for the SKUs in a specific partition. Such transactions are filtered out. The default value is 10% (that is, a SKU that is involved in fewer that 10% of the median number of transactions for a specific partition is not included in the calculation of the CDT).

Once you have configured the filters, click **Run** to start the filtering process.

Data Filtering Summary

The following information is provided after the filtering is complete and quantifies the amount of data filtered out for the three indicated filters for sales units, sales amounts, transaction counts, SKU counts, and customer counts. Use this information to assess the effects of the pruning.

Figure 2–10 CDT Data Filtering Summary

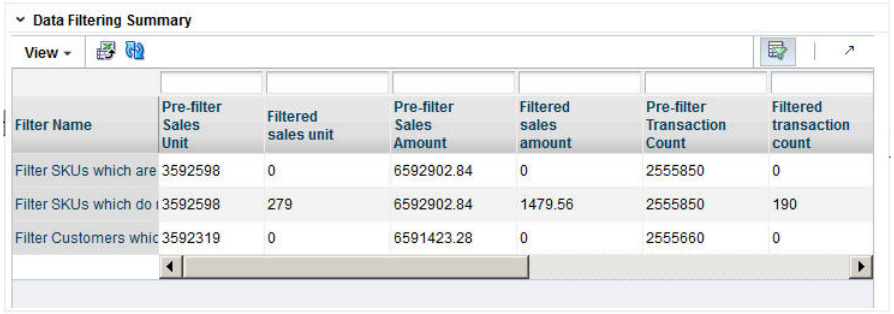


Table 2–11 Filter Data Summary Fields

Field Name	Field Description
Filter Name	The following filter names are listed: Filter SKUs that are missing too many attribute values. This maps to the SKU Filter: Missing attribute values maximum in the Filter Setup. Filter SKUs that do not have enough sale transaction history. This maps to the Customer Filter: Transaction history minimum in the Filter Setup. Filter Customers that do not have typical transaction history. this maps to the SKU-Segment-Location Filter: Transaction minimum in the Filter Setup.
Pre-filter Sales Unit	Amount prior to the application of the filter.
Filtered Sales Unit	Amount remaining after the application of the filter.
Pre-filter Sales Amount	Amount prior to the application of the filter.
Filtered Sales Amount	Amount remaining after the application of the filter.
Pre-filter Transaction Count	Amount prior to the application of the filter.
Filtered Transaction Count	Amount remaining after the application of the filter.
Pre-filter SKU Count	Amount prior to the application of the filter.
Filtered SKU Count	Amount remaining after the application of the filter.
Pre-filter Customer Count	Amount prior to the application of the filter.
Filtered Customer Count	Amount remaining after the application of the filter.

Data Filtering Histograms

The following histograms illustrate the effects of filtering. You can use the information displayed in the histograms to adjust the configuration of the filters in order to eliminate outlier data. If you modify the filters, you must re-run the stage.

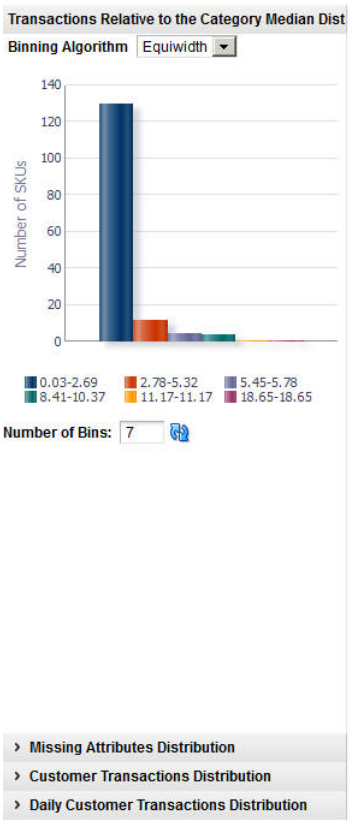
Table 2–12 Filter Histograms

Histogram Name	Description
Transactions Relative to the Category Median Distribution	Displays the number of transactions relative to the category median for a SKU's category.
Missing Attribute Distribution	Displays the number of attributes that have missing values per number of SKUs.

Table 2–12 (Cont.) Filter Histograms

Histogram Name	Description
Customer Transaction Distribution	Displays the number of transactions in historical data per number of customers.
Daily Customer Transaction Distribution	Displays the relative number of transactions per day per number of customers.

Figure 2–11 CDT Transactions Relative Histogram



Once you are satisfied with the pruning results, click **Next** to go to the Calculation stage.

Calculation

The Calculation stage creates the CDTs for all of the partitions that you selected in the Data Setup stage. A separate CDT calculation occurs for each partition. For example, if you selected two CM nodes, and the system also has two customer segments and three locations, then that is a total of 2 x 2 x 3 = 12 CDTs at the lowest level. In addition, there are CDTs at the higher levels (for example, above location).

Process

Here is the high-level process for calculating CDTs.

1. Enter a unique name for the version of the CDTs to be calculated.
2. Enter values for lowest tree level and minimum percentage of SKUs.
3. Use the check boxes to indicate whether or not only top level processing calculation should occur for location or customer segment.

4. Customize the ranking of the category attributes if necessary.
5. Click **Run** to start the calculation.
6. Review the calculation results in the Calculation Report and the CDT Score Distribution histogram.
7. After reviewing the results, if necessary, make changes to the values for the calculation in Version Setup and re-run the stage.
8. When you are satisfied with the results, click **Next** to go to the Evaluation stage.

Version Setup

At the top of this stage you see three text boxes and two check boxes that you use to configure the levels at which the calculation occurs.

Figure 2–12 CDT Version Setup

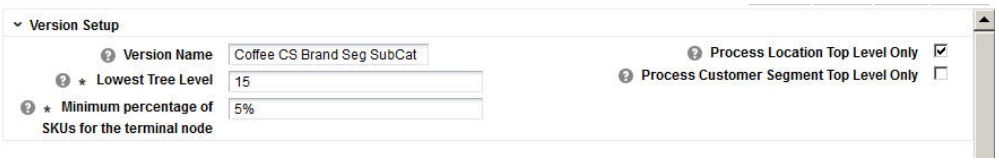


Table 2–13 Calculation Stage: Fields

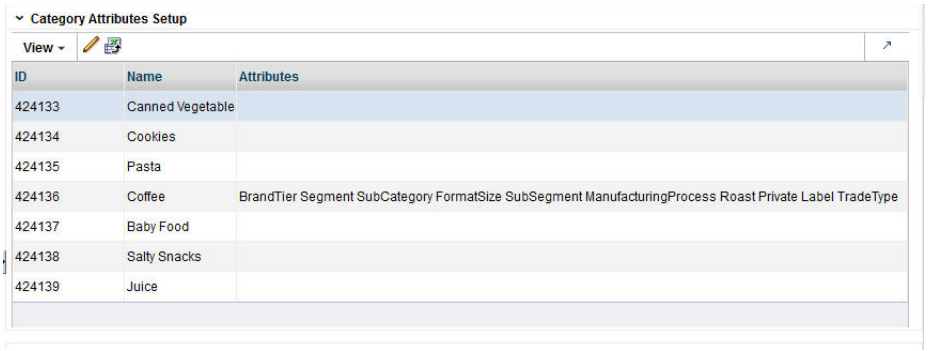
Field Name	Field Description
Version Name	Assign each version of a CDT calculation with a name. This allows you to create and save more than one version of a CDT. The version name you assign here is used in the Calculation Report, Aggregate Statistics table, and in the Manage CDTs tab. Version names can be re-used; however, if the version name in question has active CDTs, then you will see a warning that the active CDTS will be removed from the version if you do re-use the version name.
Lowest Tree Level	Use this option to define the maximum number of levels of CDTs to be calculated. The default value is 15. The number of levels in a CDT can at most be equal to the number of attributes used by the SKUs in the category plus 1. (The number 1 is used for the top level of the tree.) When determining the maximum number of levels, you should consider how many attributes should be represented in the tree. (For example, if you only want to see the top six attributes in the CDT, then set the lowest level to 7.)
Minimum Percentage of SKUs for the Terminal Node	Use this option to make sure the terminal nodes have a sufficient number of SKUs. For example, if the current CDT partition contains 100 SKUs, then any one node in the tree containing fewer than 5 SKUs will be considered a terminal node and the tree will not expand further along that branch of the tree.
Process Location Top Level Only	Check this option if you want CDTs to be calculated for the Location Chain *only*. You can select this option in order to decrease the amount of time it takes the system to perform the calculation.
Process Customer Segment Top Level Only	Check this option if you want CDTs to be calculated for the Customer Segment Chain *only*. You can select this option in order to decrease the amount of time it takes the system to perform the calculation.

Category Attributes Setup

The application determines a specific ranking order for the category attributes. You can optionally change this order and create your own ranking of the attributes for a category by ordering the attributes from most important to least important. The Category Attributes Setup table lists the categories you are calculating CDTs for. Use this list to select a category to edit.

Click the **Edit** icon to access this functionality. You can adjust the ranking for all the attributes or a subset of the attributes.

Figure 2–13 CDT Category Attributes Setup



The Category Attributes Setup dialog box contains the following fields. For each attribute you want to rank, select the Top Attribute check box and enter a value for the Attribute Rank. Click **OK**.

Figure 2–14 CDT Category Attributes Setup

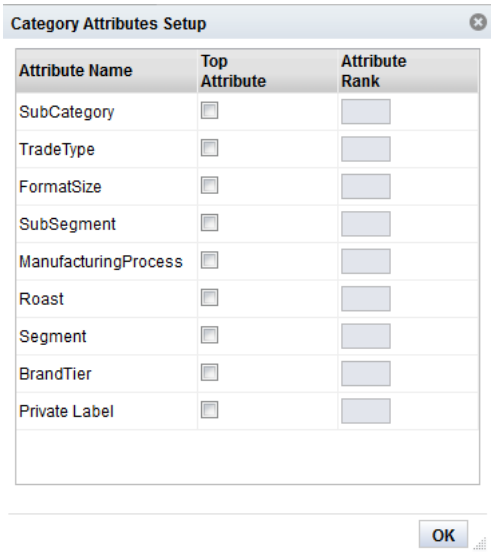


Table 2–14 Category Attributes Setup Dialog Box

Field	Description
Attribute Name	Identifies the attribute.

Table 2–14 (Cont.) Category Attributes Setup Dialog Box

Field	Description
Top Attribute	Check to enable ranking. This indicates that this attribute must be one of the first attributes in the resulting tree. This box is checked automatically for functional fit attributes and cannot be deselected. A functional fit attribute is one that a customer cannot choose a different product with a different attribute value. For example, windshield wiper blades must be selected according to the car manufacturer's requirements. The numerical rank is optional. If you do not enter rank values, then the system will determine the order of all the top attributes first and then put any remaining attributes below the top attributes in the tree.
Attribute Rank	Enter a value for each attribute, starting with the value 1, to indicate the relative rank of the attribute. A value can be entered only if the attribute has been identified as a top attribute. You do not have to specify the rank if Top Attribute is checked. If multiple attributes are checked, then the system rank orders them.

Once you have configured the parameters for the calculation, click **Run** to start the process.

Calculation Report

The Calculation Report displays the CDT results for the indicated version from the last successful run (if one has occurred). Here you can review information by Location or by Customer Segment. You can also set a version as active, view and edit a CDT, or compare two CDTs.

Figure 2–15 CDT Calculation Report

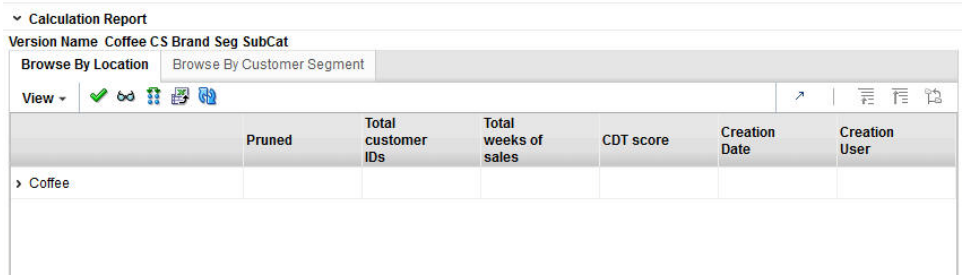


Table 2–15 Calculation Report Fields

Field Name	Description
Partition	A partition is the combination of category, segment, and location. The column identifies the names of all nodes in the tree structure in the table. The node type may be either category, location, or customer segment.
Pruned	The number of CDTs removed from the list of usable CDTs. CDTs are removed that do not meet the filtering thresholds. During the Escalation stage, the escalation process makes adjustments for the CDTs that are removed.
Total Customer IDs	The number of customers used in the calculation of the CDT.

Table 2–15 (Cont.) Calculation Report Fields

Field Name	Description
Total Weeks of Sales	The total number of weeks of sales used in the calculation of the CDT. This provides an indicates of the amount of sales data used in the calculation. All the CDTs in a given version should have a similar value; if this is not the case, the results should be evaluated.
CDT Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Creation Date	The date when the version whose data is displayed was created.
Creation User	The login name of the user who created the version.

The following icons are available for the Calculation Report:

Set Version as Complete

Each partition must have an active CDT. The active CDT is used by other applications that require a CDT.

All CDTs within a version are activated when you click **Set Version as Complete**.

If a CDT exists in more than one version then the most recently activated version takes precedence.

If a version contains multiple categories and is Complete and then a different version with a partial overlap of categories is later marked as Complete, then only the overlapping categories are replaced in the new version.

If a version is overwritten then the CDTs in that version are no longer active.

When a CDT is at risk of going from active to inactive, you will see a warning message.

To determine whether or not a CDT is active, go to the Manage CDTs tab. The Browse by Categories or by Versions table displays a flag that indicates whether or not a CDT is active.

View One CDT

This icon provides access to the CDT Editor. For details about this functionality, see Using the CDT Editor.

Compare Two CDTs

This icon provides access to CDT Compare. For details about this functionality, see Comparing Two CDTs

CDT Score Distribution Histogram

The CDT Score Distribution histogram displays, for the current calculation, the range of CDT scores and their frequencies. CDT scores range from 0 to 100; a higher number indicates a better score. You can use the information in this histogram to assess the quality of your CDTs.

Below the histogram you see a table that provides details about the histogram. This information can help you assess the quality of a CDT. Click on a specific bin in the histogram to populate the table with information about that bin. You see a list of all the specific category name/location name/customer segment name partitions for the bin,

along with the score for each, the number of weeks of data used to calculate the score, and the number of tree nodes for each partition.

Table 2–16 CDT Score Distribution Fields

Field Name	Description
Category Name	The category name of the partition for the specific CDT score.
Location Name	The location name of the partition for the specific CDT score.
Customer Segment Name	The customer segment name of the partition for the specific CDT score.
Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Number of Weeks	The number of weeks of data used to calculate the CDT score.
Tree Node Count	The number of tree nodes used to calculate the CDT score.

When you have finished with the calculation process, click **Next** to go to the Evaluation stage.

Evaluation

Once the CDT has been generated, you can evaluate it and make adjustments to parameters used in the CDT calculation during the Evaluation stage. Only the current version of the CDT, the one just generated in the Calculation stage, can be assessed in the Evaluation stage.

Process

Here is the high-level process for using the Evaluation stage to prune the results of the CDT calculation.

1. If you have determined that you need to prune the results of the calculation, enter appropriate values into the text boxes in the Pruning Setup area.
2. Click **Run** to start the pruning process.
3. Review the pruning results in the Pruning Results section and in the Pruning histograms.
4. After reviewing the results, if necessary, make changes to the values for the pruning and re-run the stage.
5. When you are satisfied with the results, click **Next** to go to the Escalation stage.

Pruning Setup

You can make changes by pruning the CDT based on the customer count, the SKU count, the Tree-level count, and the minimum CDT score. To do this, enter a minimum value for each pruning filter.

Figure 2–16 CDT Pruning Setup

Once you have completed setting up the filters, click **Run** to begin the processing. After you review the pruning results, you can change the values for the filters if you find it necessary. Once you make changes, you must run the stage again in order to see the results of your changes.

Table 2–17 Pruning Setup: Filters

Filter Name	Filter Description
Minimum Customer Count	The minimum number of customers to be used in the CDT calculation.The default is 1000.
Minimum SKU Count	The minimum number of SKUs to be used in the CDT calculation. The default is 10.
Minimum Tree Level Count	The minimum number of levels to be used in the CDT calculation. The default is 2.
Minimum CDT Score	The minimum measure of the quality of a CDT to be used in the CDT calculation. The default is 25.

Pruning Results

The Pruning Results are located below Pruning Setup and display information that can help you assess the effects of the values you provided for the pruning filters.

Figure 2–17 CDT Pruning Results

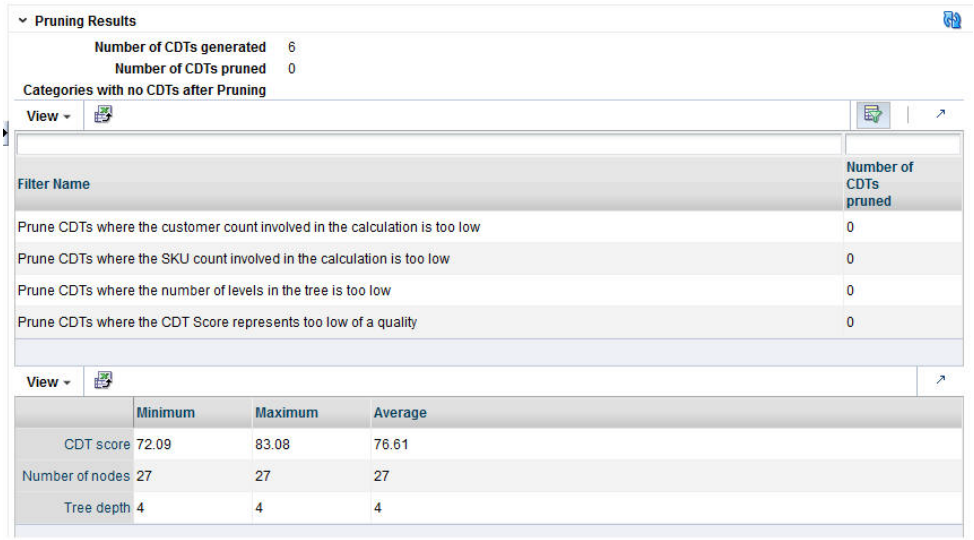


Table 2–18 Pruning Results

Field Name	Field Description
Number of CDTs Generated	The number of CDTs that were generated by the Calculation stage.
Number of CDTs Pruned	The number of CDTs that were pruned after the filters were applied.
Categories with no CDTs After Pruning	The names of the categories from which all CDTs were pruned.

Below this list is a table that identifies the number of CDTs that have been pruned by the following filters.

- Prune CDTs where the customer count involved in the calculation is too low.
- Prune CDTs where the SKU count involved in the calculation is too low.
- Prune CDTs where the number of levels in the tree is too low.
- Prune CDTs where the CDT score represents too low of a quality.

A second table provides an overview of the pruning results, including the minimum, maximum, and average values for the CDT score, the number of nodes, and tree depth.

The number of nodes indicates the size of the tree. If a CDT contains few nodes, this can indicate a problem with the data or that too many nodes were excluded during the Calculation stage (because of a parameter setting).

Tree depth also indicates the size of the tree. This value can be used in conjunction with the Lowest Tree Level setting in the Calculation stage to analyze the results in terms of the number of levels in the tree.

Evaluation Report

The Evaluation Report displays the CDT results for the indicated version from the last successful run, if one has occurred. Here you can review information by Location or by Customer Segment. You can also set a version as active, view and edit a CDT, or compare two CDTs.

Table 2–19 Evaluation Report Fields

Field Name	Description
Partition	A partition is the combination of category, segment, and location. The column identifies the names of all nodes in the tree structure in the table. The node type may be either category, location, or customer segment.
Pruned	The number of CDTs removed from the list of usable CDTs. CDTs are removed that do not meet the filtering thresholds. During the Escalation stage, the escalation process makes adjustments for the CDTs that are removed.
Total Customer IDs	The number of customers used in the calculation of the CDT.
Total Weeks of Sales	The total number of weeks of sales used in the calculation of the CDT. This provides an indicates of the amount of sales data used in the calculation. All the CDTs in a given version should have a similar value; if this is not the case, the results should be evaluated.
CDT Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Creation Date	The date when the version whose data is displayed was created.
Creation User	The login name of the user who created the version.

The following icons are available for the Evaluation Report:

Set Version as Complete

Each partition must have an active CDT. The active CDT is used by other applications that require a CDT.

All CDTs within a version are activated when you click **Set Version as Complete**.

If a CDT exists in more than one version then the most recently activated version takes precedence.

If a version contains multiple categories and is Complete and then a different version with a partial overlap of categories is later marked as Complete, then only the overlapping categories are replaced in the new version.

If a version is overwritten then the CDTs in that version are no longer active.

When a CDT is at risk of going from active to inactive, you will see a warning message.

To determine whether or not a CDT is active, go to the Manage CDTs tab. The Browse by Categories or by Versions table displays a flag that indicates whether or not a CDT is active.

View One CDT

This icon provides access to the CDT Editor. For details about this functionality, see Using the CDT Editor.

Compare Two CDTs

This icon provides access to CDT Compare. For details about this functionality, see Comparing Two CDTs

Three histograms are provided in the Evaluation stage (need details):

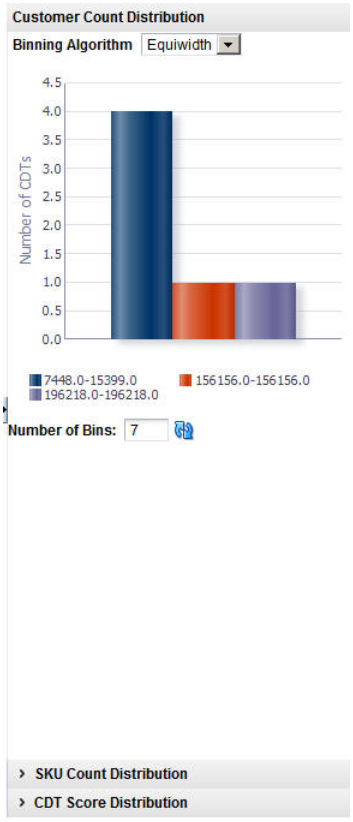
Evaluation Histograms

The following histograms are displayed in the Evaluation stage. You can use the information from the histograms to understand the data that was pruned by the filters.

Table 2–20 Evaluation Histograms

Name	Description
Customer Count Distribution	Displays the number of CDTs for a specific customer count.
SKU Count Distribution	Displays the number of CDTs for a specific SKU count.
CDT Score Distribution	Displays the CDT score distribution.

Figure 2–18 CDT Customer Count Distribution



When you have completed the evaluation of the results, click **Next** to go to the Escalation stage.

Escalation

The Escalation stage is used to fill in the holes for partitions whose CDTs were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy.

Process

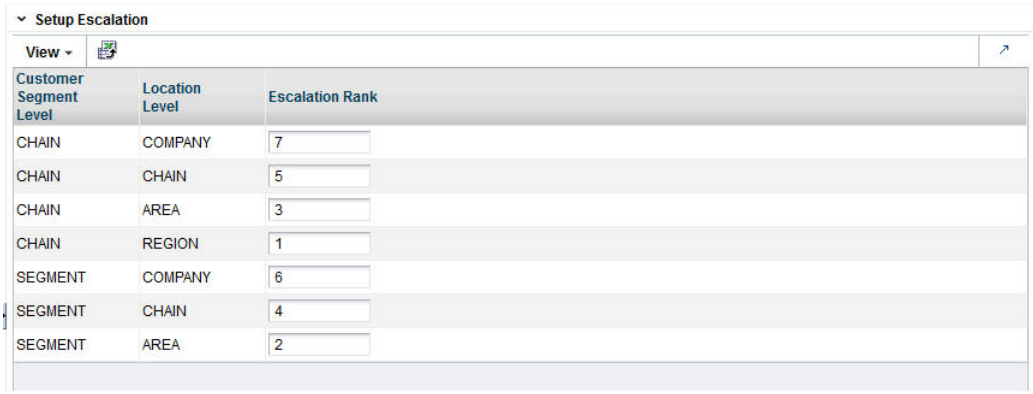
Here is the high-level process for setting up an escalation.

1. Enter a series of numbers to indicate the escalation rank, which determines the order in which the escalation occurs.
2. Click **Run** to start the escalation process.
3. Review the escalation results in the Escalation Report.
4. After reviewing the results, if necessary, make changes to the escalation ranks and re-run the stage.
5. When you are satisfied with the results, you can complete the version and make the version active so that it is available for other applications to use.

Setup Escalation

Escalation occurs along the segment hierarchy and the location hierarchy. Here is an example of an escalation path:

Figure 2–19 CDT Setup Escalation



The following fields are required to set up the escalation.

Table 2–21 Setup Escalation

Field	Description
Customer Segment Level	Identifies the customer segment level in the escalation.
Location Level	Identifies the location level in the escalation.
Escalation Rank	Used to assign the ranks for the escalation, which determines the order in which the escalation occurs.

Here is an example of an escalation path.

Table 2–22 Example of Escalation Path

Segment Level	Location Level	Escalation Rank
Segment chain	Location chain	8
Segment chain	Region	7
Segment chain	Location area	6
Segment chain	Store cluster	5
Segments	Location chain	4
Segments	Region	3
Segments	Location area	2
Segments	Store cluster	1

You fill in the order of numbers. Every row must have an ordering number, and no ordering number can be reused.

The escalation path is specific to the user and the current version that the user is working on.

The default ordering is to go up the location hierarchy first, and then up the segment hierarchy, as shown in the example above. The reason is that the segment hierarchy has only two levels, and so its top level is very general.

Escalation Report

The Escalation Report breaks down the numbers to provide counts for the number of partitions filled with higher-level CDTs and the number of partitions that have not

been changed by escalation. In addition, the fraction of CDTs for each partition is displayed.

Figure 2–20 CDT Escalation Report

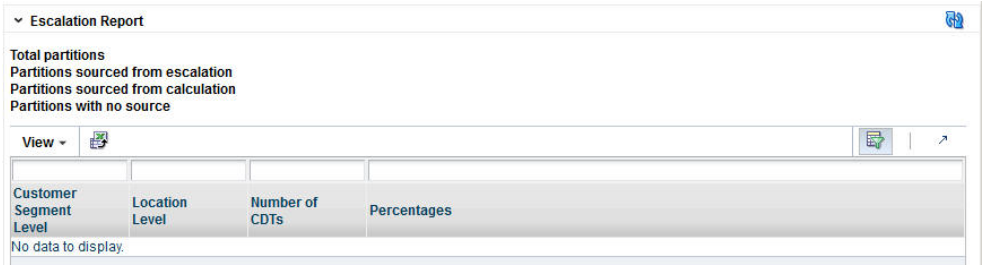


Table 2–23 Escalation Report

Field Name	Description
Total Partitions	The number of partitions in the version.
Partitions Sourced from Escalation	The number of partitions removed during escalation.
Partitions sourced from Calculation	The number of partitions removed during calculation.
Partitions with No Source	A partition that does not have CDTs assigned to it because all CDTs related to the partition have been pruned.
Customer Segment Level	Identifies the customer segment level.
Location Level	Identifies the location level.
Number of CDTs	The number of partitions that are trying to have a CDT assigned. This is generally the number of customer segments by the number of locations.
Percentages	The percentage of partitions that have been assigned a CDT from a given escalation level.

Completion of Process

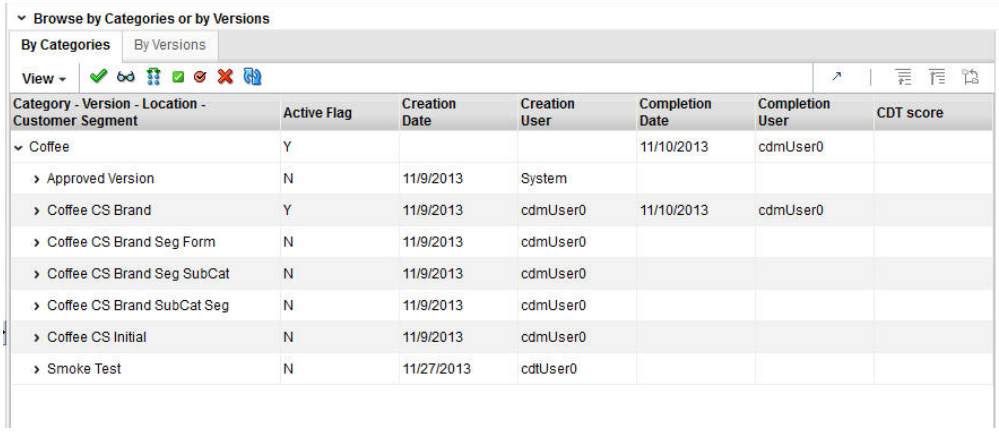
When a version is complete, the results for the version are activated so that other applications can use the information. The similarity data that has been calculated during the generation process is also activated for use.

After the completion of this step, the intermediate results from each stage is removed from the database and can no longer be used.

Manage CDTs Tab

The Manage CDTs tab is used to control which version of a CDT is active or approved. You can also access the functionality to edit a CDT or compare two CDTs.

Figure 2–21 Manage CDTs



You can find the CDT you are interested in by:

- Browsing by categories or versions
- Searching by version name or user name.

Set Version as Complete

Each partition must have an active CDT. The active CDT is used by other applications that require a CDT.

All CDTs within a version are activated when you click **Set Version as Complete**.

If a CDT exists in more than one version then the most recently activated version takes precedence.

If a version contains multiple categories and is Complete and then a different version with a partial overlap of categories is later marked as Complete, then only the overlapping categories are replaced in the new version.

If a version is overwritten then the CDTs in that version are no longer active.

When a CDT is at risk of going from active to inactive, you will see a warning message.

To determine whether or not a CDT is active, go to the Manage CDTs tab. The Browse by Categories or by Versions table displays a flag that indicates whether or not a CDT is active.

View One CDT

This button provides access to the CDT Editor. For details about this functionality, see Using the CDT Editor.

Compare Two CDTs

This button provides access to CDT Compare. For details about this functionality, see Comparing Two CDTs.

Approve a Version

Only one approved CDT is permitted for a given CM node/segment/location combination. If you approve a second CDT with the same combination, it overwrites the first one.

If an Approved CDT exists, then the Approved CDT is returned to the requesting application.

If no Approved CDT exists, then the Active CDT is returned to the requesting application.

If no Approved CDT exists and no Active CDT exists, an error is returned to the requesting application.

Approval can be done for individual CDTs or for an entire version; activation is for an entire version.

An approved version is used by the Category Management application. An Active version is used by Demand Transference.

Unapprove

Use this to change the status of a CDT that was previously approved. Select a single CDT record in the data table and click the Unapprove button to remove it from the approved version.

Delete a Version

If you delete a version, it cannot be activated. You cannot delete an approved version, but you can remove a CDT from an approved version.

Browse by Category or Version

You can browse by category or version. The following information is displayed:

Table 2–24 Browsing by Category or Version

Field Name	Field Description
Version...Customer Segment	Partition
Active Flag	Y = Version is active. N = version in not active.
Creation Date	Date on which version was created.
Creation User	User who created the version.
Activation Date	Date on which the version was made active.
Activation User	User who made the version active.
CDT Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.

Search

You can search by version name or creation user name. In addition, you can design a search with specific criteria and save that search for future use.

Using the CDT Editor

You can view or edit a CDT using the CDT Editor, which you access by clicking the **Edit** button. The CDT Editor displays the Customer Segment Name and Location Name for the CDT. It provides tools to navigate through the display using Zoom, Pan, Center, and Layout functionality.

The following functionality is available:

- Add a Child
- Edit
- Delete
- Delete Branch
- Copy Branch
- Paste Branch
- Save and Approve

Add a Child

To add a child node to the CDT, do the following:

1. Click **Add Child**. You see the Add CDT Node window.
2. Select an attribute from the **Attribute** drop-down menu.
3. Select a value from the **Attribute Value** drop-down menu.
4. Select the Create separate branch per value, if applicable. When more than one of the attribute values are selected, this field create a separate node for every attribute value selected.

> **Note:** A node can be created by selecting one or more attribute values. If multiple attribute values are chosen, and the check box to create a separate branch is not selected, one node that represents multiple attribute values is created.
>
> If one branch is created for all values of an attribute, select all attribute values and only one node will appear with the attribute value of ALL.
>
> If all the attribute values for an attribute are used, either by creating one node or separate nodes, a child cannot be added to the node, and the **Add Child** is disabled. Creating a node with ALL does not associate the specific attribute values with the node. Instead, it encompasses any attribute value.

5. Click **OK**.

Editing Nodes

Editing is allowed on all nodes as long as there are additional attribute values available. The selections from the node being edited are preselected in the dialog. The attribute cannot be changed, so the attribute drop-down menu is not disabled. Any attribute values that are in use by other siblings are not be displayed in the list.

Deleting Nodes

Nodes can be deleted from the tree. There are two options for deleting: Delete and Delete Branch.

Selecting Delete deletes the selected node. Delete is available for any leaf node (a node without any children). It is also available for a non-leaf node that represents all the attributes for an attribute value: the children under this node move under the parent.

Selecting Delete Branch deletes the entire branch of nodes under the selected node, but not the selected node itself.

Copying/Pasting in Customer Decision Tree Editor

Copy and Paste functionality is provided to copy nodes from one branch to another. Copy Branch is enabled when a node is selected that has children. The Copy Branch function copies the full branch of children of the selected node, but not the selected node itself. **Paste Branch** is enabled when a branch has been copied and the selected node is a leaf node.

Viewing a Customer Decision Tree

The Customer Decision Trees have the potential to get large and occupy more space than the screen real estate allows for. The hierarchy viewer component used to display the Customer Decision Tree provides several features to assist in viewing the Customer Decision Tree effectively.

Expanding and Collapsing Branches

One way to limit the amount of space taken up by the Customer Decision Tree is to collapse branches of the tree. For any node that has children, a small triangle appears at the bottom of the box for that node. Hovering over the triangle enlarges it and display an option to collapse that node if it is expanded and expand the node if it is collapsed.

Moving the Tree

If the entire Customer Decision Tree is not visible on one screen, the Customer Decision Tree can be moved to make other parts of the tree visible. The view can be moved by either clicking and dragging or by using the panning controls in the control bar for the hierarchy viewer.

Zooming

The hierarchy viewer provides some controls for zooming in and out to allow more or less of the tree to be in view at a time. Zooming out shrinks the size of the nodes which may make them difficult to read.

Comparing Two CDTs

You can also compare two CDTs using the same CDT Editor functionality. You select the two CDTs you want to compare from the list. Both CDTs are displayed side by side.

3

Demand Transference

This chapter described the use of the Demand Transference (DT) module.

Introduction

Demand Transference (DT) helps you to compare products based on their similarities in order to determine what, if any, products customers might buy if the product they want to buy is for some reason unavailable. In this way, planning and ordering can be optimized. DT calculates similarities by comparing the attributes of the two products. If you are using CDT in conjunction with DT, you also have available the similarities calculated by CDT, which are based on customer-supplied transaction data.

The DT module consists of three tabs: Overview, Generate Models, and Manage Models. You use the Overview tab to keep track of the status of each stage during the main work you do with the application within the Generate Models tab. You use the Manage Models tab to evaluate the demand elasticity results and override the Maximum Substitutable Demand Percentage value if that is needed.

Overview of DT Process

When you use the DT module, you follow this general iterative process to create and manage DT models:

- Data Setup. You define the categories to be used in the DT calculation.
- Data Filtering. You configure filters that remove input data that might cause errors in the calculation or that might lead to inaccurate or unreliable results.
- Similarity Calculation. You calculate similarities and assess the results of the calculation.
- Elasticity Calculation. You calculate the assortment elasticities and assess the results in terms of substitutable demand, which is the percentage of demand of a SKU that is retained when the SKU is deleted from the stores where it is selling.
- Escalation. When you are satisfied with your results, you can set the escalation path you can set the escalation path to fill in the holes for partitions whose DT models were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy. Then you can set a version of the DT model as complete.
- Manage Models. Use this tab to set time intervals for evaluating your results and to override the value for the maximum substitutable demand percentage.

Overview Tab

The Overview tab displays information that you can view and use to monitor the progress of the DT stages as well as to view some aggregate statistics and the DT results from the last successful run.

This tab contains the following sections:

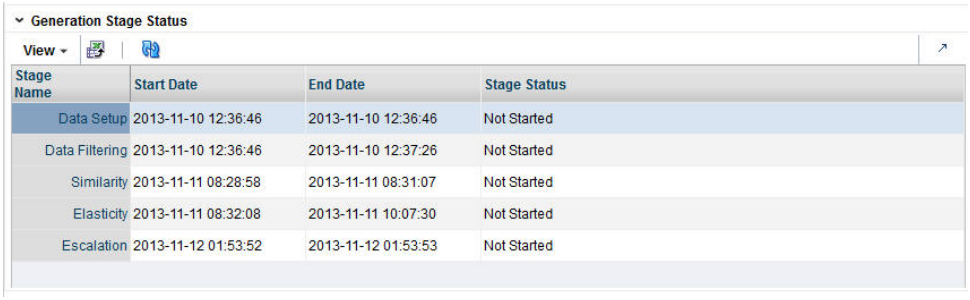
- Generation Stage Status
- Aggregate Statistics
- Calculation Report

Generation Stage Status

The Generation Stage Status table displays the current status of each of the DT stages.

Click the **Refresh** icon to update the fields and see the latest status for each stage.

Figure 3–1 DT Generate Stage Status



You can use the status information to monitor the progress of each stage. It contains the following fields, which can be arranged and viewed, but not modified.

Table 3–1 Generation Stage Status Fields

Field Name	Description
Stage Name	A row exists in the table for each DT stage that provides detailed status information. The five stages are Data Setup, Data Filtering, Similarity, Elasticity, and Escalation.
Start Date	The date and time when a run for the stage most recently started.
End Date	The date and time when a run for the stage most recently ended.
Stage Status	The current status of the stage: Not Started, Not Started (Scheduled for Later), Processing, Completed Successfully, Stopped/Paused, Cancelled, or Failed.

The following table provides an explanation of the different values for the status of a stage.

Table 3–2 Stage Status Values

Stage Status	Description
Not Started	This indicates that the stage has not yet been initiated. This can occur when you first begin to create a DT model or when you update an existing version.

Table 3–2 (Cont.) Stage Status Values

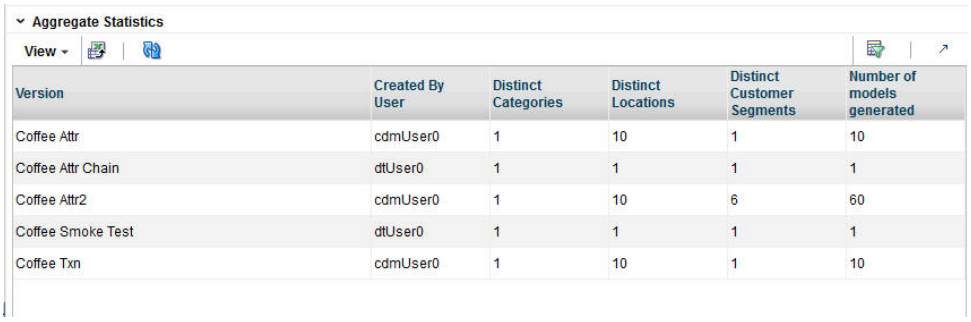
Stage Status	Description
Not Started (Scheduled for Later)	This indicates that the stage is scheduled to start after the previous stage is complete.
Processing	This indicates that the stage is currently being processed.
Stopped/Paused	This indicates either that the stage was stopped by the application or that the user has chosen to stop the process (for example, to make a change to an option). In the later case, once any background processing has stopped, the user can re-run the stopped stage.
Completed Successfully	This indicates that the stage has been successfully processed.
Completed with Errors	This indicates that the stage was able to complete the processing of all requested CDTs, but one or more CDTs encountered errors during the run and were not able to complete successfully. This most commonly occurs when the data used for a CDT is not available or is too sparse to produce a result.
Cancelled	This indicates that the database has cancelled the execution of the stage either because of missing data or an exception. The user should review the RSE_LOG_MSG log to determine the problem.
Failed	This indicates that a problem occurred during the processing.

Aggregate Statistics

The Aggregate Statistics table displays statistical details about the existing DT model versions.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 3–2 DT Aggregate Statistics



Each existing version has a row in the table. The table contains the following fields, which can be arranged and viewed but not modified.

Table 3–3 Aggregate Statistics Fields

Field Name	Description
Version	User-created name that uniquely identifies this DT model version
Created By User	The user name of the person who created this version.
Distinct Categories	The number of categories associated with this version.
Distinct Locations	The number of locations associated with this version.

Table 3–3 (Cont.) Aggregate Statistics Fields

Field Name	Description
Distinct Customer Segments	The number of customer segments associated with this version.
Number of Models Generated	The number of DT models that have been calculated for this version.

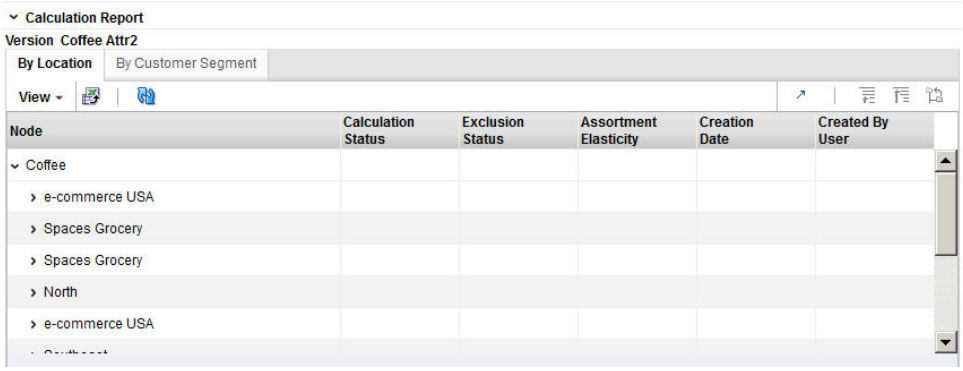
Calculation Report

The Calculation Report displays the DT model results from the last successful run, if one has occurred.

Here you can review information by Location or by Customer Segment.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 3–3 DT Calculation Report



The Calculation Report contains the following fields:

Table 3–4 DT Calculation Report Fields

Field Name	Description
Node	Identifies the node name.
Calculation Status	**Yes** indicates that the calculation is complete. **No** indicates that the calculation is not complete.
Exclusion Status	Yes indicates that data has been pruned. No indicates that data has not been pruned.
Assortment Elasticity	A number calculated by the application that is a parameter in the DT model. A larger magnitude indicates larger overall transference.
Creation Date	The date when the version whose data is displayed was created.
Created By User	The login name of the person who created the version.

Generate Models Tab

The Generate Models tab is used to configure, run, evaluate, modify, and deploy a DT model. The process is divided into five stages that must be run in order. You can return to a stage you have already completed and make changes, but if you do, you must re-run that stage and all the stages that follow that stage, as the calculations are invalidated by the modifications you just made to the settings in that stage.

The five stages are:

Table 3–5 Generate Models Tab: Stages

Stage Name	Description
Data Setup	Select the nodes for the DT model calculations.
Data Filtering	Filter out input data that may result in inaccurate or unreliable answers.
Similarity Calculation	Calculate the similarities in customer demand.
Elasticity Calculation	Calculate the assortment elasticities for customer demand.
Escalation	Set the escalation path for the DT model. Use the Escalation Report to evaluate the results of the escalation.

Data Setup

The Data Setup stage is used to add and delete the categories to be used in the DT model generation process.

Process

Here is a high-level process for setting up the data for DT.

1. Select the category or categories you want to calculate DT models for.
2. Click **Next** to go to the Data Filtering stage.

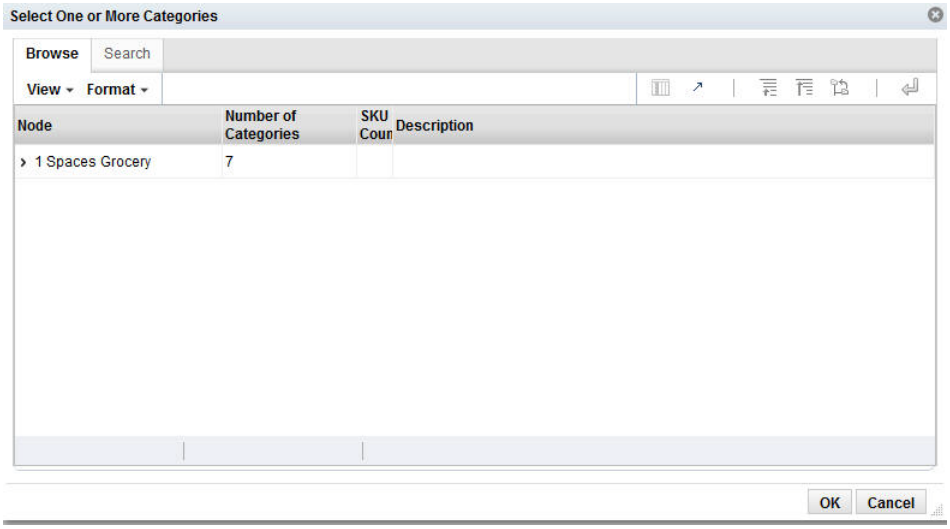
Category Selections

Use this table to add categories you want to include in the DT model calculation or delete categories that you want to remove from the DT model calculation.

Figure 3–4 DT Category Selections

To display a list of available nodes and the categories included in those nodes, click the **Add** icon. You see the Select One or More Categories dialog box, which contains two tabs, Browse and Search. You can use either tab to find the categories you are looking for.

Figure 3–5 DT Browse Categories



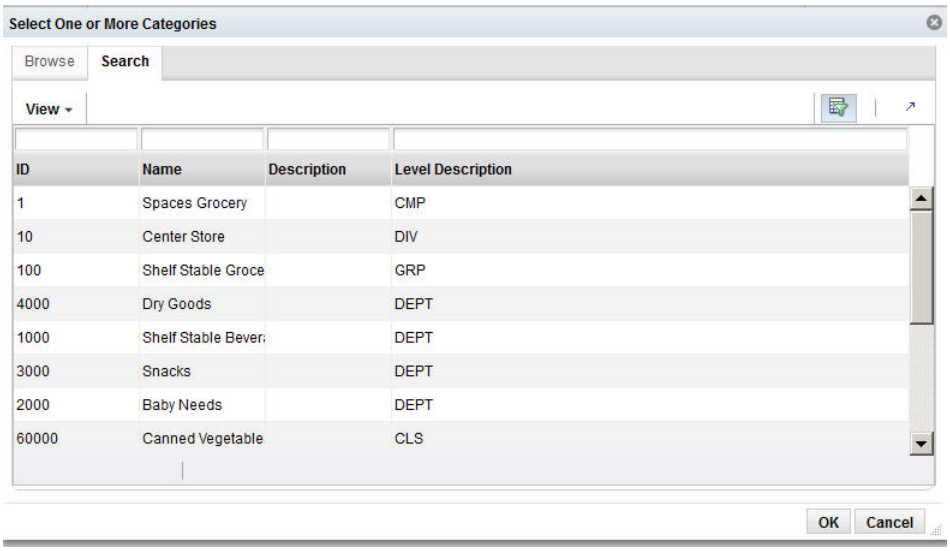
The Browse tab displays a table with the following fields:

Table 3–6 Category Selections: Browse

Field Name	Description
Node	The node tree structure can be expanded in order to view its categories.
Number of Categories	The number of categories within the node that has been selected. The number can help you understand the amount of processing required for the calculation.
SKU Count	The number of SKUs in a category. A category with too few SKUs may not produce good DT models.
Description	A description that provides additional information about the category.

Select the category or categories within the node that you want to be part of the DT model and click **OK**.

Figure 3–6 DT Search Categories



The Search tab displays a table with the following fields:

Table 3–7 Category Selections: Search

Field Name	Field Description
ID	An external code used to identify the category in other systems such as Category Management.
Name	The category name.
Short Description	A description that provides additional information about the category.
Level Description	A description of the level of the merchandise hierarchy that the node belongs to.

Select the category or categories within the node that you want to be part of the DT model and click **OK**.

Your selections are displayed in the Category Selections table.

After selecting the categories, click **Next** to go to the Data Filtering stage.

Data Filtering

The Data Filtering stage applies to all the categories that you select in the Data Setup stage. You should set the filters based on the histograms for each filter. The histograms help identify what data is actually outlier data, as compared to the rest of the data. In most cases, the default settings should be sufficient. However, if a histogram shows a flatter distribution, then you should consider modifying the default settings.

Process

Here is the high-level process for setting up and running data filtering.

1. Enter the appropriate values into the Filter Setup text entry boxes.
2. Click **Run** in order to filter the data.
3. Review the filtering results in the Data Filtering Summary table and the Data Filtering histograms.

4. After reviewing the results, if necessary, make changes to the values for the filters in Filter Setup and re-run the stage.
5. When you are satisfied with the results, click **Next** to go to the Similarity Calculation stage.

Filter Setup

You configure the following filters in order to filter out data you consider unacceptable from the calculation of the DT model.

Figure 3–7 DT Filter Setup

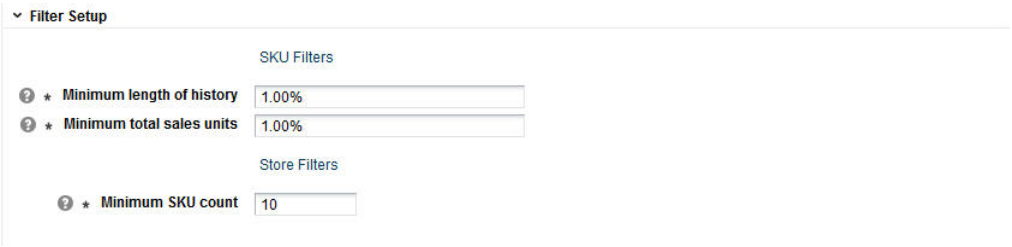
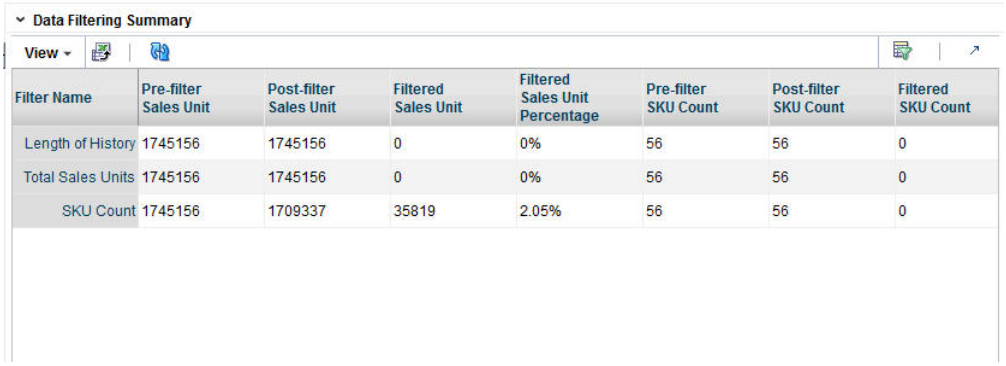


Table 3–8 Data Filters

Filter Name	Description
Minimum Length of History	This filter prunes SKU-segment-store combinations that have a short transaction history. The threshold is defined as a percentage of the median value for the category. The default value 1%.
Minimum Total Sales Units	This filter prunes SKU-segment-store combinations that have a small number of total sales units during a given sales history for a specified customer segment and store. The threshold is defined as a percentage of the median value for the category. The default value 1%.
Minimum SKU Count	This filter is applied after the above two filters and looks at the remaining data to determine if a store does not have enough SKUs. The threshold is defined as a set number of SKUs per store. The default value is 10 SKUs.

Data Filter Summary

Figure 3–8 DT Data Filter Summary



Filter Name	Pre-filter Sales Unit	Post-filter Sales Unit	Filtered Sales Unit	Filtered Sales Unit Percentage	Pre-filter SKU Count	Post-filter SKU Count	Filtered SKU Count
Length of History	1745156	1745156	0	0%	56	56	0
Total Sales Units	1745156	1745156	0	0%	56	56	0
SKU Count	1745156	1709337	35819	2.05%	56	56	0

The following information is provided after the filtering is complete and quantifies the amount of data filtered out for baseline history, total sales units, sales amounts, and SKU counts. Use this information to assess the effects of filtering.

Click the **Refresh** icon to update the fields and see the latest information for this table.

Table 3–9 Data Filter Summary Fields

Field Name	Field Description
Filter Name	The relevant filter of the three listed above.
Pre-filter Sales Unit	Amount prior to the application of the filter.
Post-filter Sales Unit	Amount remaining after the application of the filter.
Filtered Sales Unit	Amount filtered.
Filtered Sales Unit Percentage	Amount filtered, expressed as a percentage.
Pre-filter SKU Count	Amount prior to application of filter.
Post-filter SKU Count	Amount remaining after application of filter.
Filtered SKU Count	Amount filtered.
Filtered SKU Count Percentage	Amount filtered, expressed as a percentage.

Data Filtering Histograms

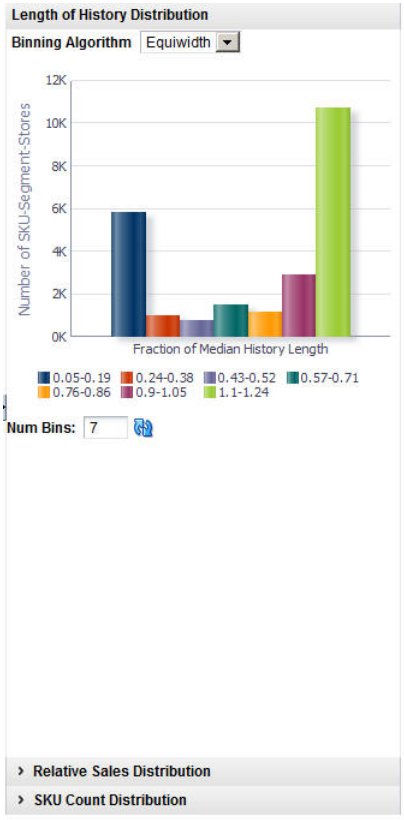
The following histograms illustrate the effects of filtering. You can use the information displayed in the histograms to adjust the configuration of the filters in order to eliminate outlier data. If you modify the filters, you must re-run the stage.

For information about adjusting the display of the histograms, see the Chapter 1, "Getting Started."

Table 3–10 Data Filtering Histograms

Histogram Name	Description
Length of History Distribution	Displays the percentage of median history length relative to the number of SKU-segment-stores.
Relative Sales Distribution	Displays the percentage of median category sales relative to the number of SKU-segment-stores.
SKU Count Distribution	Displays the percentage of the median category SKU count relative to the number of segment-stores.

Figure 3–9 DT Length of History Distribution Histogram



Similarity Calculation

Similarity in demand can be determined using either transaction-based data or attribute data. You can calculate the similarity using each type of data, if both types are available. You can only view the most recent run in the UI, so in order to compare runs, you must query the database to obtain the results from earlier runs.

Each category has its own set of similarities, relevant to the SKUs that are in that category. A similarity is calculated for each pair of historical SKUs in a category.

If transaction-based similarities are available, it is recommended that you use them instead of attribute-based similarities. Note that transaction-based similarities are only available through the Customer Decision Tree application.

Process

Here is the high-level process for calculating similarities:

1. Enter a unique name for the version of the DT model to be calculated.
2. Select the source of the data to be used: transaction-based or attributed-based. Transaction-based data is only available from the Customer Decision Tree generation.
3. Use the check boxes to indicate whether or not only top level processing should occur for location or customer segment.
4. Customize the ranking of the category attributes if necessary.
5. Click **Run** to start the calculation.

6. Review the calculation results in the Similarity Display.
7. After reviewing the results, if necessary make changes to the values for the calculation in the Version Setup and Category Attribute Setup and then re-run the stage.
8. When you are satisfied with the results, click **Next** to go to the Elasticity Calculation stage.

Version Setup

At the top of this stage you see two text boxes and two check boxes that you use to configure the parameters for the calculation.

Figure 3–10 DT Version Setup

Table 3–11 Version Setup: Fields

Field Name	Field Description
Version Name	Assign each version of a DT model calculation with a name. This allows you to create and save more than one version of a DT model. The version name you assign here is used in the Calculation Report, Aggregate Statistics table, and in the Manage DTs tab. Version names can be re-used; however, if the version name in question has active DT models, then you will see a warning that the active DT models will be removed from the version if you do re-use the version name.
Select the Source of Similarities	Use this option to define the type of data used in the calculation: transaction data or attribute data. Transaction-based data uses similarities calculated by Customer Decision Tree using transaction-based data. Attribute-based data calculates similarities within DT based on the attribute values associated with every SKU in the category.
Process Location Top Level Only	Check this option if you want DT models to be calculated for the Location Chain *only*. You can select this option in order to decrease the amount of time it takes the system to perform the calculation.
Process Customer Segment Top Level Only	Check this option if you want DT models to be calculated for the Customer Segment Chain *only*. You can select this option in order to decrease the amount of time it takes the system to perform the calculation.

Once you select the source for the similarities, you will see either Category Attribute Setup, if you have selected to use attribute-based similarities, or Transaction-based Similarity Availability Per Category, if you have selected to use transaction-based similarities.

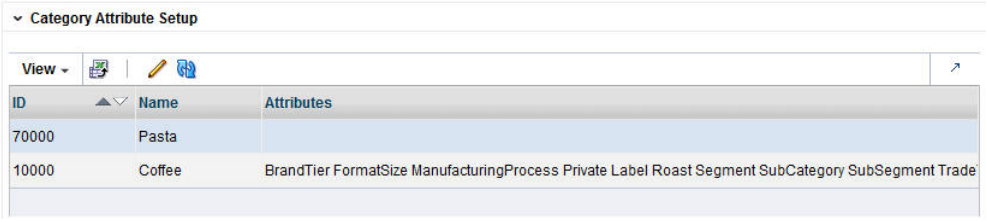
Category Attribute Setup

The application, using historical data, determines a specific weight for the category attributes. You can optionally change this weight and assign your own weight to the attributes for a category.

The weights indicate the importance of the attribute to the customers when they are making purchasing decisions. The attribute with the highest weight is the one the customer considers first when making a purchase. The system-generated weights are determined by the application from historical sales data. However, if a user disagrees with those weights, the user can override them. For example, in the case of coffee, the system may assign a weight of 0.7 to brand and 0.2 to size. This indicates that brand is historically more important to the customer than size when purchasing coffee. If the user disagrees with this analysis and thinks that brand and size are actually much closer together, the user can assign a weight of 0.5 to brand and 0.4 to size.

The Category Attribute Setup table displays the following:

Figure 3–11 DT Category Attribute Setup



Highlight the category you want to adjust the weights for and click the **Edit** icon. You see the Category Attributes Setup dialog box.

Figure 3–12 DT Category Attribute Setup

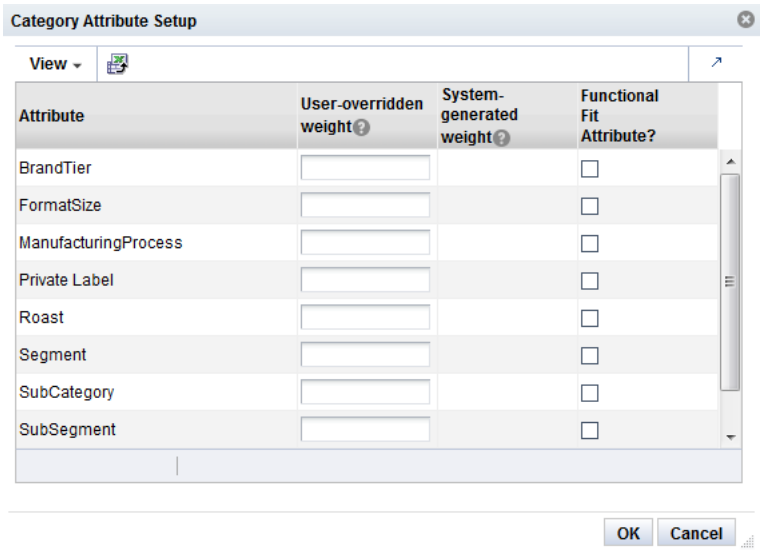


Table 3–12 Category Attributes Setup

Field	Description
ID	An external code used to identify the category in other systems such as Category Management.
Name	The category name.
Attribute	The specific attribute you are configuring.

The Category Attributes Setup pop-up lists the categories you are calculating DT models for. The system-assigned weights are also displayed.You can adjust the weight for all the attributes or a subset of the attributes.

The Category Attributes Setup dialog box contains the following fields. For each attribute you want to assign a custom weight to, enter a number between 0.000 and 1.000. For attributes that have no substitutes (such as windshield wipers of a specific length), the Functional Fit check box is checked by the system, so that similarities are not calculated for these attributes. When you are finished configuring the category attributes, click **OK**.

Table 3–13 Category Attribute Setup Fields

Field	Description
Attribute	The category attribute to assign a weight to.
User-Overridden Weight	The user-defined weight for the attribute.
System-Generated Weight	The system-generated weight for the attribute.
Functional Fit Attribute?	This is checked by the system if the attribute has no substitutes.

After you have finished configuring the similarity parameters, click **Run** to calculate the similarities. You see the results via the Similarity display table.

Transaction-Based Similarity Availability Per Category

The Transaction-Based Similarity table displays the following:

Table 3–14 Transaction-Based Similarity Availability Per Category

Field	Description
ID	An external code used to identify the category in other systems such as Category Management.
Name	The category name.
Description	A description that provides additional information about the category.
Available	A flag that indicates that a CDT version that contains data for this category has been made active.
Available As Of	Indicates the date that the CDT version was activated. This information can help you identify whether the CDT results are recent, or if they are potentially too old to use. For example, if the CDT data became available 2 years ago, you may consider that data to be out of date.

Similarity Display

The Similarity Display table shows the list of SKUs for which similarities have been calculated so that you can sort and analyze the results. You can search through the list of results by Category Name, Location, or customer Segment.

Click the Refresh icon to update the fields and see the latest information in this table.

Figure 3–13 DT Similarity Display

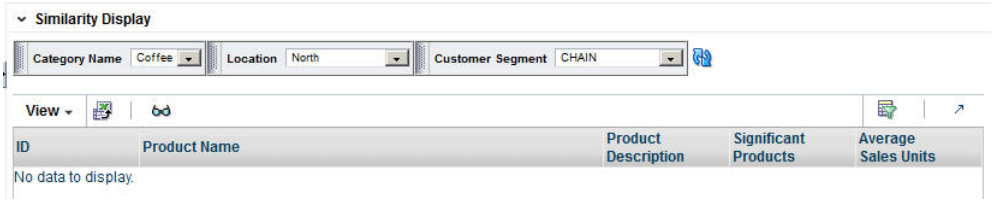


Table 3–15 Similarity Display

Field	Description
ID	An external code used to identify the category in other systems such as Category Management.
Product Name	The name identifying the product.
Product Description	A detailed description of the product.
Significant Products	A list of products that have High or Very High similarity. The threshold for how many products are considered significant can be configured in the database.
Average Sales Units	The average number of units used in the calculation.

Click the **See Similarities** icon to see detailed results for a specific set of SKUs.

The detailed results include the following fields:

Table 3–16 Similarity Display Results

Field Name	Field Description
ID	The product SKU.
Product Name	The name identifying the product.
Product Description	A detailed description of the product.
Similarity Strength	An indication of the similarity for the product: Very High, High, Medium, Low, Very Low.
Similarity Value	The calculated value for the similarity, from 0 to 1. A higher value indicates a higher degree of similarity.
Similarity Code	The numeric value associated with the similarity: 4 = Very High, 3 = High, 2 = Medium, 1 = Low, 0 = Very Low.

When you are satisfied with the Similarity results, click **Next** to go to the Elasticity Calculation stage.

Elasticity Calculation

During the Elasticity Calculation stage, the assortment elasticity is calculated. You do not configure any parameters. Click **Run** to initiate the calculation.

The assortment elasticity should not be a positive value because the transference model does not work properly if the value is positive. In addition, it should not be a null value because a null value indicates that the calculation of assortment elasticity failed and did not produce an assortment elasticity value. If an assortment elasticity value is positive, it must be replaced with a negative value. The replacement occurs during the escalation process.

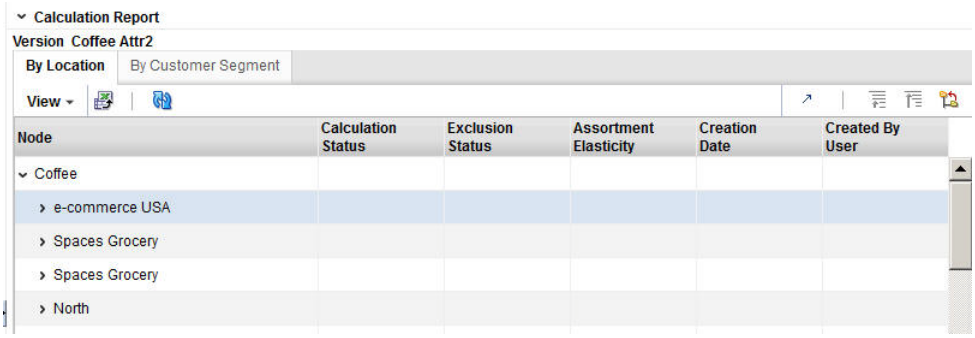
Process

The elasticity calculation is a background process. You use this stage to view the results. Note that the substitutable demand information is displayed and percents for the DT models are calculated after you set the time intervals within the Manage Models tab.

Calculation Report

The Calculation Report lists the status of the elasticity calculation and the exclusion, either by Location or by Customer Segment. An assortment elasticity is calculated for each category/location/segment combination selected during the Data Setup stage and the Calculation stage. The numerical result of the calculation, an output of DT generation, is used by Manage Models to calculate substitutable demand percentages.

Figure 3–14 DT Calculation Report



The Calculation Report has two tabs: By Location and By Customer Segment. Each tab has the following fields:

Table 3–17 Calculation Report Fields

Field	Description
Node	The node tree structure can be expanded in order to view its categories
Calculation Status	Indicates whether or not (Yes/No) the elasticity calculation has occurred.
Exclusion Status	Indicates whether or not (Yes/No) pruning has occurred.
Creation Date	The date when the version whose data is displayed was created.
Created By User	The login name of the person who created the version.
Assortment Elasticity	A number calculated by the application that is a parameter in the DT model. A larger magnitude indicates larger overall transference.

Pruning Report

The Pruning Report displays statistics about the results of data pruning.

Figure 3–15 DT Pruning Report

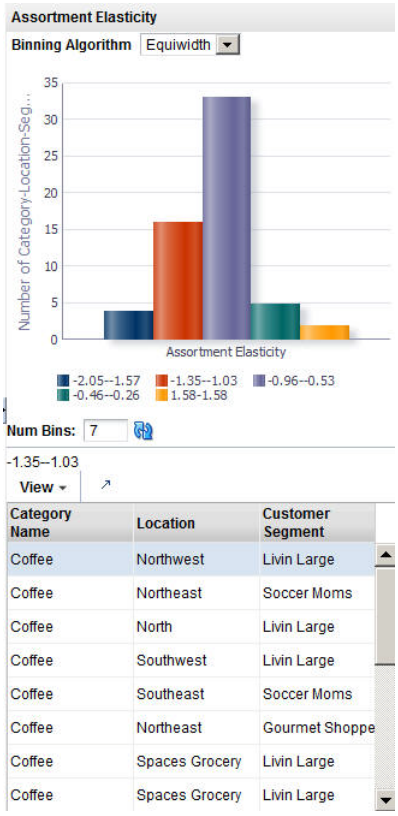
The Pruning Report contains the following fields:

Table 3–18 Pruning Report Fields

Field	Description
Number of Models Generated	The number of DT models produced by the calculation.
Number of Models Pruned	The number of DT models pruned by the calculation.
Categories with All Models Pruned	The names of the categories for which all DT models have been pruned.
Positive or Null Assortment Elasticity	The Assortment Elasticity (AE) value is a negative number used in calculating substitutable demand percentages and demand transference effects. A positive AE value may be produced as a result of missing or unreliable input data for some partitions. Such meaningless transference effects must be pruned. In addition, if a null AE value is generated, it must also be pruned.
Substitutable Demand	Substitutable demand is a measure of how much demand is retained by the rest of the assortment when an item that is removed. When the item is removed, a portion of its demand is transferred to the remainder of items in the assortment. These values are populated when you set up the time intervals in the Manage Models tab and run the calculation there.

Assortment Elasticity Histogram

Figure 3–16 DT Assortment Elasticity Histogram



Escalation

The Escalation stage is used to fill in the holes for partitions whose DT models were removed during pruning by setting up a search path through the customer segment hierarchy and the location hierarchy. The DT models used to fill in the holes are not used by Manage Models calculations.

Process

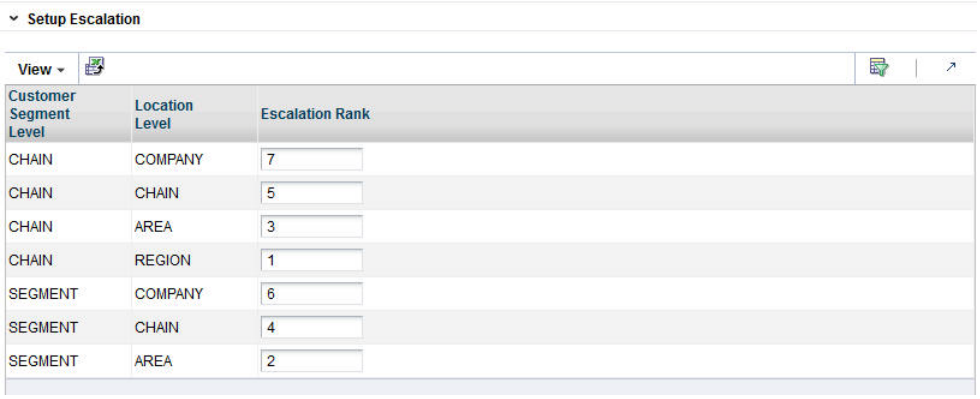
Here is the high-level process for setting up an escalation.

1. Enter a series of numbers to indicate the escalation rank, which determines the order in which the escalation occurs.
2. Click **Run** to start the escalation process.
3. Review the escalation results in the Escalation Report.
4. After reviewing the results, if necessary, make changes to the escalation ranks and re-run the stage.
5. When you are satisfied with the results, you can complete the version and make the version active so that it is available for other applications to use.

Setup Escalation

Escalation occurs along the customer segment hierarchy and the location hierarchy. Here is an example of an escalation path:

Figure 3–17 DT Setup Escalation



The following fields are required to set up the escalation.

Table 3–19 Setup Escalation

Field	Description
Customer Segment Level	Identifies the customer segment level in the escalation.
Location Level	Identifies the location level in the escalation.
Escalation Rank	Used to assign the ranks for the escalation, which determines the order in which the escalation occurs.

Here is an example of an escalation path.

Table 3–20 Example of Escalation Path

Segment Level	Location Level	Escalation Rank
Segment chain	Location chain	8
Segment chain	Region	7
Segment chain	Location area	6
Segment chain	Store cluster	5
Segments	Location chain	4
Segments	Region	3
Segments	Location area	2
Segments	Store cluster	1

You fill in the order of numbers. Every row must have an ordering number, and no ordering number can be re-used.

The escalation path is specific to the user and the current version that the user is working on.

The default ordering is to go up the location hierarchy first, and then up the segment hierarchy, as shown in the example above. The reason is that the segment hierarchy has only two levels, and so its top level is very general.

Escalation Report

The Escalation Report breaks down the numbers to provide counts for the number of positions filled with higher-level DT models and the number of partitions that have

not been changed by escalation. In addition, the fraction of DT models for each partition is displayed.

Figure 3–18 DT Escalation Report

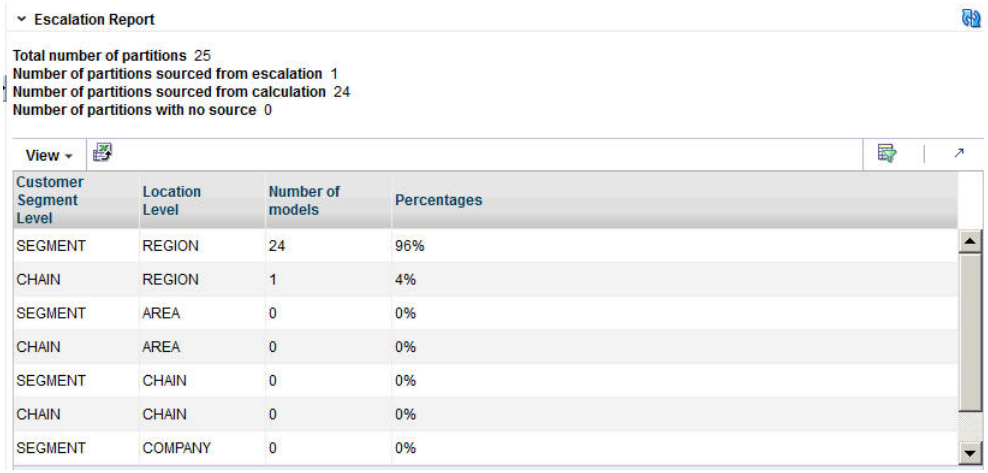


Table 3–21 Escalation Report

Field Name	Description
Total Number of Partitions	The number of partitions in the version.
Number of Partitions Sourced from Escalation	The number of partitions removed during escalation.
Number of Partitions Sourced from Calculation	The number of partitions removed during calculation.
Number of Partitions with No Source	A partition that does not have a model assigned to it because all models related to the partition have been pruned.
Customer Segment Level	Identifies the customer segment level.
Location Level	Identifies the location level.
Number of Models	The number of partitions that are trying to have a model assigned. This is generally the number of customer segments by the number of locations.
Percentages	The percentage of partitions that have been assigned a model from a given escalation level.

Completion of Process

When a version is complete, the results for the version are activated so that other applications can use the information. The similarity data that has been calculated during the generation process is also activated for use.

After the completion of this step, the intermediate results from each stage is removed from the database and can no longer be used.

Be aware that once a version is completed, it cannot be completed again unless a different version is completed first. Changes made to the version's data after completing it will not be copied to the relevant output tables.

Manage Models Tab

You can set up various time intervals to use in the evaluation of a version of a DT and configure the value for maximum substitutable demand and see how different maximum values affect the substitutable demand. This allows you to change the maximum value to one you find more suitable. You can see the percentage of demand of a SKU that is retained when the SKU is deleted from the stores where it is selling. This is the substitutable demand percentage. In this way you can evaluate the accuracy and usability of the elasticity calculation.

Substitutable demand is a measure of how much demand is retained by the rest of the assortment when an item that is removed. When the item is removed, a portion of its demand is transferred to the remainder of items in the assortment. This portion is considered the retained demand. If the magnitude of the assortment elasticity is larger, then the amount retained will be higher. By examining the retained demand, you can evaluate the assortment elasticity value to see if its magnitude is too large. The key value to examine is the maximum substitutable demand percentage. For a given category, you may decide that this value is too large.

Process

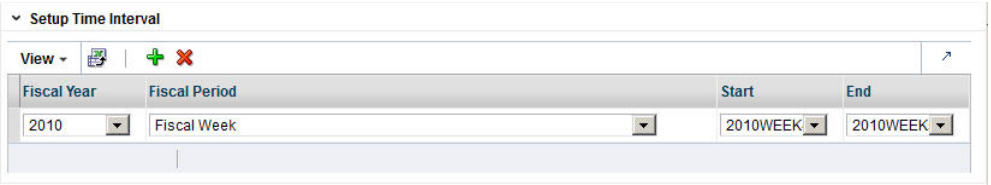
Here is the high-level process for determining suitable substitutable demand values and thus suitable DT models.

- Set up the time intervals you are interested in.
- Select the versions you want to evaluate.
- Click the **Calculate** icon to obtain an initial set of percentages.
- Enter various values for maximum substitutable demand and use the **Edit** icon to enter override values for Maximum Substitutable Demand Percent and the **Calculate** icon to determine the impact.
- You can use the **Revert** icon to restore the original percentage values.
- When you are satisfied with the results, you can choose which version to make active using the **Set Version as Complete** icon.

Time Interval Setup

The time interval defines the span of time for the sales history to be used to determine the amount of history that is retained when SKUs are dropped. A group of intervals can be defined. Gaps between intervals are permitted; however, intervals cannot overlap.

Figure 3–19 DT Setup Time Interval



You should select a time interval where the historical assortments are reasonably representative of the assortments that will be used in the Category Management application. Because the time interval is used to calculate the substitutable demand information, selecting a representative interval provides substitutable demand information that is highly relevant to the actual application of demand transference in

Category Management. Typically, the most representative time period is a recent time interval, since that is generally when assortments are most similar to the current assortments. If you use a time period that is not recent, you run the risk of using assortments that are not as similar to the current ones. You should also make sure not to select an interval that is too large, because a large interval necessarily includes several assortment changes within that interval. An interval size of approximately four weeks is recommended.

The fields that define a time interval are:

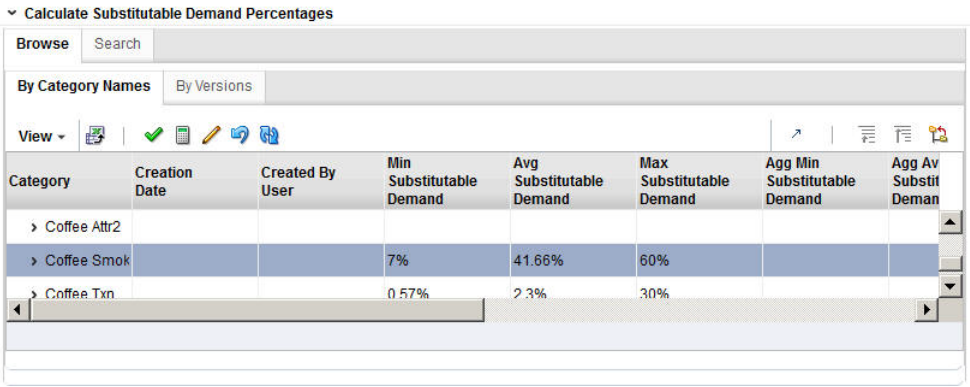
Table 3–22 Setup Time Interval: Fields

Field Name	Field Description
Fiscal Year	The fiscal year for the time interval.
Fiscal Period	The fiscal period within the fiscal year (Fiscal Quarter, Fiscal Period, or fiscal Week).
Start	The time unit when the time interval specified in Fiscal Period begins.
End	The time unit when the time interval specified in Fiscal Period ends.

Calculate Substitutable Demand Percentages

You can vary the value for the maximum substitutable demand percentage and see the impact on selected categories or versions.

Figure 3–20 DT Calculate Substitutable Demand Percentages



Browsing and Searching

You can browse by Category Names or by Versions. You can also search by name or by creation user. You see the following DT model data displayed:

Table 3–23 Browsing

Field Name	Description
Category/Version	The name of the category.
Creation Date	The date when the version was created.
Created By User	The user name of the person who created this version.
Min Substitutable Demand	The percentage value for the minimum substitutable demand.
Avg Substitutable Demand	The percentage value for the average substitutable demand.

Table 3–23 (Cont.) Browsing

Field Name	Description
Max Substitutable Demand	The percentage value for the maximum substitutable demand.
Agg Min Substitutable Demand	The aggregated value for the minimum substitutable demand.
Agg Avg Substitutable Demand	The aggregated value for the average substitutable demand.
Agg Max Substitutable Demand	The aggregated value for the maximum substitutable demand.
Calculation Status	Indication of whether or not (Yes/No) the calculation is complete.
Completion Status	Indication of whether or not (Yes/No) the status is active.
Completion Date	The date when the status became active.
Completion User	The user name of the person who activated the version.

Once you have made a selection, click the **Edit** icon and enter an override value between 0% and 100% for the Maximum Substitutable Demand Percent. Click the **Calculate** icon to initiate the calculation. To revert the calculation, click the **Revert** icon.

Once you have determined the substitutable demand value you want, you can click the **Complete** icon to make the version active.

Escalation Report

The Escalation Report breaks down the numbers to provide counts for the number of positions filled with higher-level DT models and the number of partitions that have not been changed by escalation. In addition, the fraction of DT models for each partition is displayed.

Figure 3–21 DT Escalation Report

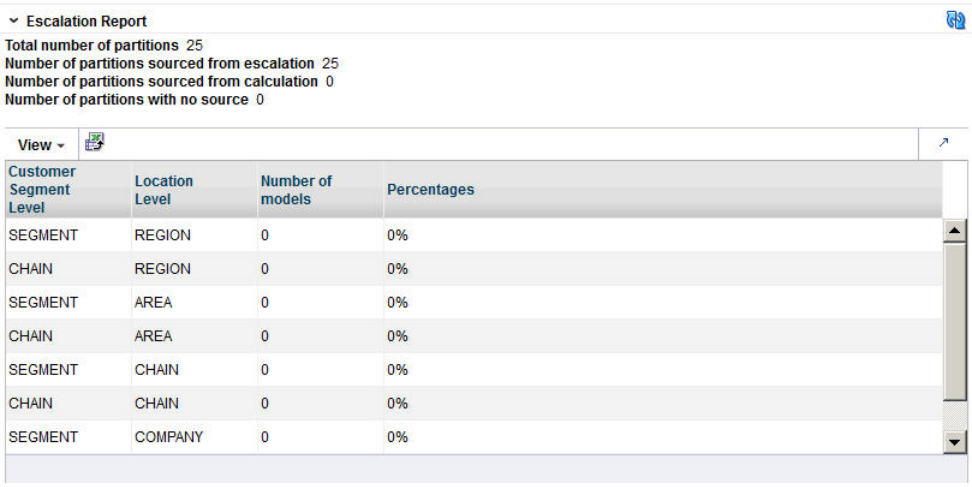


Table 3–24 Escalation Report

Field Name	Description
Total Number of Partitions	The number of partitions in the version.

Table 3–24 (Cont.) Escalation Report

Field Name	Description
Number of Partitions Sourced from Escalation	The number of partitions removed during escalation.
Number of Partitions Sourced from Calculation	The number of partitions removed during calculation.
Number of Partitions with No Source	A partition that does not have a model assigned to it because all models related to the partition have been pruned.
Customer Segment Level	Identifies the customer segment level.
Location Level	Identifies the location level.
Number of Models	The number of partitions that are trying to have a model assigned. This is generally the number of customer segments by the number of locations.
Percentages	The percentage of partitions that have been assigned a model from a given escalation level.

Histogram

The Substitutable Demand Distribution histogram displays the distribution of the substitutable demand values for SKU/stores.

4

Advanced Clustering

This chapter describes the Advanced Clustering module.

Introduction

Advanced Clustering lets you create store clusters based on common features such as customer demographics in order to manage merchandise assortments and pricing strategies in a targeted way. Clusters can help retailers to understand who shops in their stores and what their preferences are.

You can create clusters based on consumer profiles, store attributes, product attributes, a mixture of attributes, or performance.

The Advanced Clustering application optimizes clusters in order to determine the minimum number of clusters that best describes the historical data used in the analysis and that best meets your business objectives, which you define during the design of your clusters.

You can use Advanced Clustering to execute localized or consumer-centric assortments and for pricing. In addition, the application can help you when forecasting, for example, if you want to cluster stores based on similar seasonal patterns. You can also use the application for allocation, by clustering stores based on similar selling patterns.

Features

The key features of Advanced Clustering include:

Dynamic clustering on the fly. You can define a hierarchy cluster of stores based on a store attribute such as format and then cluster further using performance in order to determine which stores have high, medium, and low sales.

Continuous and discrete (non-continuous or text) attributes. You can cluster using a store attribute such as format and performance metrics such as retail sales.

Recommended optimal number of clusters. You can compare the improvement in the significance of increasing the number of cluster centers.

What-if scenarios. You can create what-if scenarios and compare them in order to determine the best one.

Cluster scores. The application uses scoring to indicate which clusters fall below defined thresholds and as a result requires manual intervention.

Overview of Advanced Clustering Process

When you use the Advanced Clustering module, you follow this general process to create and manage clusters, working in the Generate Store Clusters tab and the Manage Store Clusters tab:

- Cluster Criteria. View summary data about existing clusters and define the characteristics of new clusters.
- Explore Data. Examine the supporting data for the cluster you defined.
- Cluster Setup. Create various scenarios based on a specified number of clusters.
- Cluster Results. View the scenario results and compare scenarios.
- Cluster Insights. Gain understanding of implications of results through contextual information.
- Manage Store Clusters. Manage existing clusters.

Cluster Criteria Overview Tab

The Cluster Criteria Overview tab displays a list of completed runs and details, including the status, of each. You can click on the run name in order to access it within the Generate Store Clusters tab.

Table 4–1 Cluster Criteria Overview Tab

Field	Description
Name	The run ID and user-assigned name of the cluster.
Cluster By	The Cluster By option used for the cluster.
Created By	The name of the user who created the cluster.
Last Updated By	The name of the user who most recently updated the cluster.
Last Updated On	The date when the cluster was most recently updated.
Status	The current status of the cluster. Values include Created, Executed, Approved, and Rejected.
Period Count	A count for the calendar.
Merchandise Count	A count for the merchandise nodes.
Location Count	A count for the location nodes.

The following cluster criteria are the defaults:

Consumer Profile. Stores that have a similar consumer segment profile such as stores where "soccer moms" shop or stores where "empty nesters" shop can be grouped together.

Store Attribute. Stores that have similar store properties and demographics can be grouped together. For example, you can use formats such as main anchor, downtown, mall, or standalone as the cluster criteria.

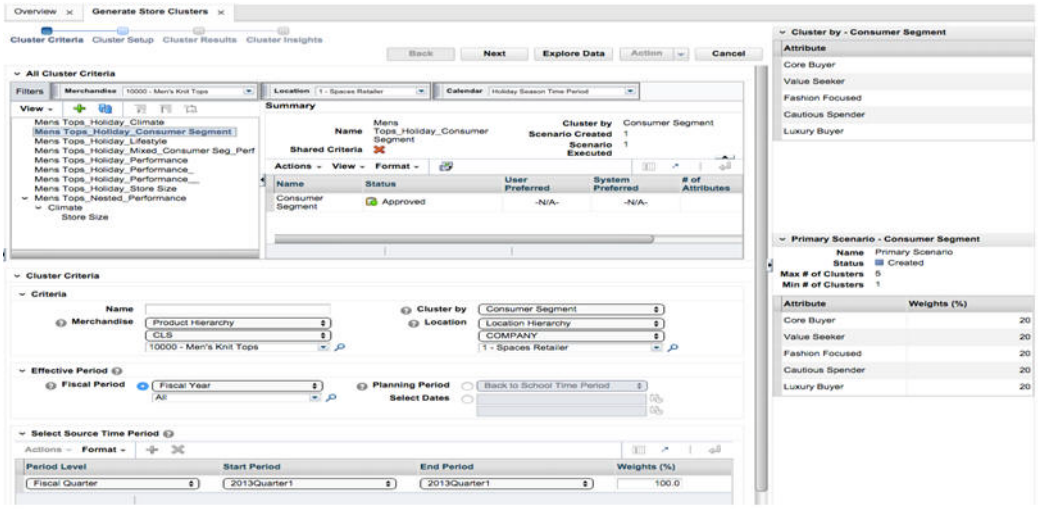
Product Attributes. Stores that have similar product attribute store shares can be grouped together. For example, you can differentiate between stores selling premium brands, stores selling standard brands, and stores selling niche brands.

Performance Criteria. Stores that have similar sales can be grouped together in order to identify high, medium, and low volume stores.

Generate Store Clusters Tab

The Generate Store Clusters tab is used to create clusters and then model the clusters with various scenarios in order to determine the best clusters. It consists of four stages: Cluster Criteria, Cluster Setup, Cluster Results, and Cluster Insights.

Figure 4–1 Generate Store Clusters



Cluster Criteria Stage

In this stage, you can view summary data about existing clusters and define the characteristics of new clusters.

Process

Here is a high-level process for defining a cluster.

1. Provide a unique name for the cluster.
2. Define the type of data used to characterize the cluster.
3. Select merchandise and location nodes.
4. Define the time period for the cluster.
5. Define the historical time period for the data.

All Cluster Criteria

In this area of the page you can view information about existing clusters.

Use the View list to select an existing cluster. You can tailor your search for existing clusters by Merchandise, Location, and Calendar. Once you select a cluster, the defining details for that cluster are displayed in the Worksheet area.

In addition, a context menu allows you to:

- Copy a cluster definition. Once you have copied it, you can modify it.
- Delete a cluster definition.
- View a cluster definition. The characteristics of the cluster are displayed in the Cluster Criteria area.

- Created a nested cluster. In this way you can sub-divide an existing cluster in order to analyze it further. Once you create a nested cluster, the name "Nested of <name of original cluster>" appears in the Cluster Criteria area. You can then define its characteristics in the same way you define any cluster.

Once you have highlighted a cluster to examine, details about that cluster are displayed in the Summary area. The details include information about the cluster and the scenarios created for that cluster.

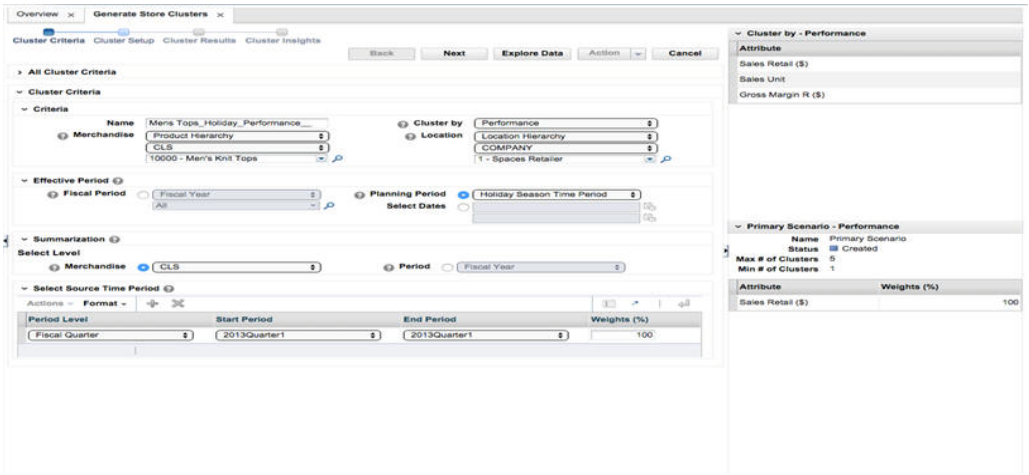
Table 4–2 Summary Details

Field Name	Description
Cluster By	Set of attributes to be used in the creation of the cluster: Consumer Profiles, Store Attributes, Product Attributes, Performance Attributes, or Mixed Attributes.
Shared Criteria	A check mark indicates more than one merchandise or location node is used in the cluster.
Merchandise Type	The merchandise type.
Scenario Created	The number of scenarios created for the cluster.
Scenario Executed	The number of scenarios executed for the cluster.
Location Type	The location type.
Name	The name assigned to each scenario that has been created for the cluster.
Status	Created, Executed, Execution Failed, Ready for Approval, Approved, Rejected.
User Preferred	Indicates whether or not the user prefers the cluster.
System Preferred	Indicates whether or not the system prefers the cluster.
# of Attributes	The number of attributes that were used in the cluster.
Max. # of Clusters	You provide a value for how many maximum clusters centers the clustering process should consider.

Cluster Criteria

In this area of the page, you define the characteristics of a new cluster. Note that multiple hierarchies are supported in order to facilitate comparisons between clusters. For example, you can compare clusters for market and retail location hierarchy.

Figure 4–2 Cluster Criteria



The following information defines a cluster:

Table 4–3 New Cluster Definition

Field Name	Description
Name	A unique name to identify the cluster.
Cluster By	The key feature that defines the cluster. Choose from Consumer Profile, Store Attributes, Product Attribute Profile, Product Performance, or Mixed Attributes.
Merchandise	Once you choose the merchandise level for the cluster, you must also choose the node. The nodes are specific to the merchandise level you select.
Location	Once you choose the location level for the cluster, you must also choose the node. The nodes are specific to the location level you select.
Hierarchy	The type of merchandise/location hierarchy that is used to create clusters.

Effective Period

You can define a time interval for the cluster by either choosing a period from the list provided or by selecting a start date and an end date. This option is available for Consumer Profile, Store Attributes, and Product Attribute Profile.

To define the Effective Period, you select either Planning Period, Fiscal Period, or Select Date:

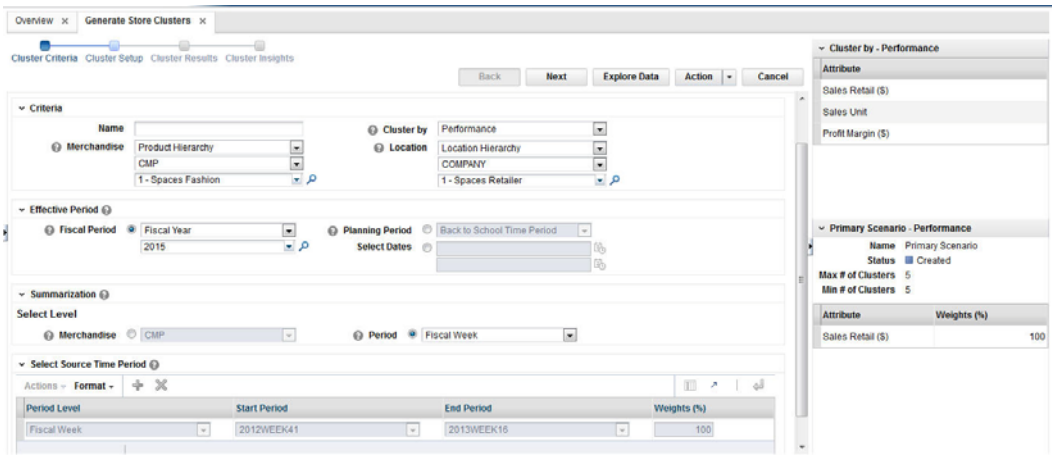
Table 4–4 Effective Period

Option	Description
Fiscal Period	If you select this option, choose the period and the sub-divisions of that period from the drop-down lists.
Planning Period	Select from the range of values provided for period. Planning periods are user-defined buying periods for a season or a season subset.
Select Dates	If you select this option, choose the start and end dates using the calendar pop-up.

Summarization

If you are clustering by Product Performance, this option is available. You can select either Select Merchandise Level or Period. Summarization allows you to aggregate data. Summarization allows you to cluster over by calendar or merchandise.

Figure 4–3 Summarization



Advanced Clustering uses the selected dimension of a hierarchy (calendar or merchandise) to summarize data and then consider all positions of the dimensions in the clustering process. For example, when you use category/week sales data to generate store clusters, you can select the week dimension in the calendar hierarchy as summarization. It internally considers all weeks (week1...week52) as attributes and clusters stores based on their weekly sales patterns.

Select Source Time Period

If you are clustering by Product Performance or Product Attribute Profiles, you can define the time period for the historical data to be used for the calculation. You can specify more than one time period and assign different weights to different periods in order to place more or less emphasis on different periods.

Source time periods are available for all cluster criteria. With product/performance or mixed criteria, when performance metrics are used for clustering, these define the historical data used for the calculation. This time period also defines the historical data used to display BI when sales metrics are shown.

Table 4–5 Select Source Time Period

Field	Description
Period Level	Select from Fiscal Year, Fiscal Quarter, Fiscal Period, or Fiscal Week.
Start Period	Once you select the Period Level, you select the starting sub-division within that period.
End Period	Once you select the Period Level, you select the ending sub-division within that period.
Weight (%)	Used to define the weight given to the historical data from the defined time period.

Contextual Area

When you are creating a new cluster, you can see details about the following two parameters that can help you understand the clustering you are creating.

Cluster By <variable> This list displays according to the choice you selected for the Cluster By option. It shows the name of each attribute and the relative weight of each attribute that characterize the Cluster By option you selected.

Effective Period This list displays the time period you selected for the cluster definition. This information is available only for planning periods, where it provides the start and end dates of the planning period.

Explore Data Stage

Use this stage to examine data for the cluster you defined. You can view the store that provides input into the clustering process. This screen can be accessed by clicking **Explore Data** from any other stage.

Process

Here is the high-level process for exploring cluster data.

1. In this stage you can only view the data, so the only actions you can perform are drilling down through the data in the table and altering the arrangement of the table.

Summary

This area lists the criteria you initially selected to define the cluster.

Figure 4–4 Clustering Summary

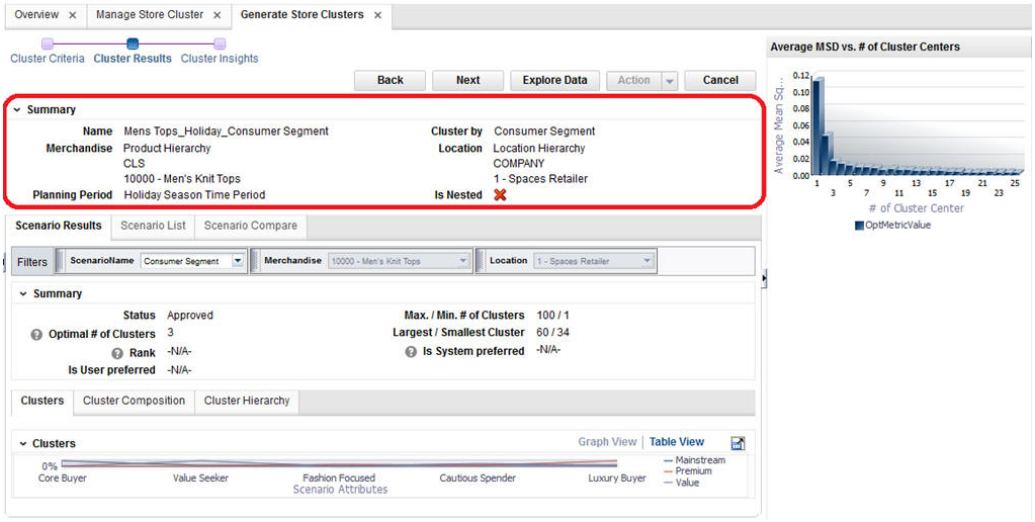


Table 4–6 Explore Data: Summary

Field	Description
Name	The name you provided for the cluster in the Cluster Criteria stage.
Cluster By	The key feature that defines the cluster. Choose from Consumer Profile, Store Attributes, Product Attribute Profile, or Product Performance.
Merchandise	The merchandise level and nodes for the cluster.
Location	The location level and nodes for the cluster.

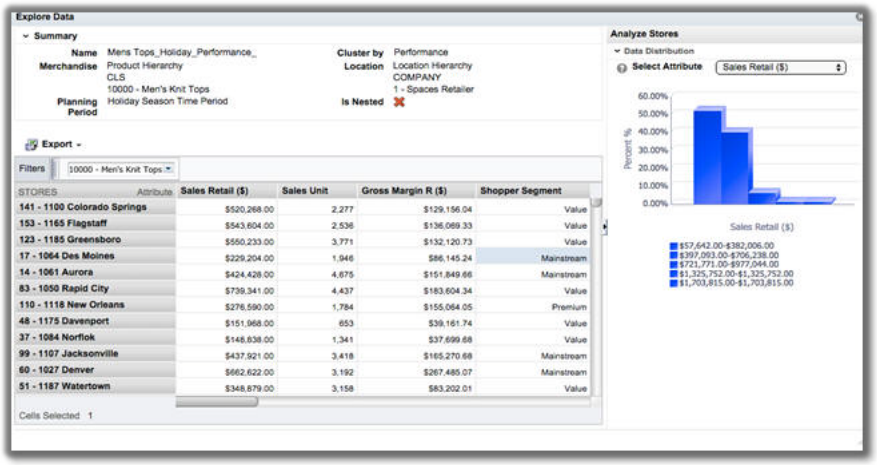
Table 4–6 (Cont.) Explore Data: Summary

Field	Description
Fiscal Period	The time period for the cluster.
Is Nested	Indicates whether or not the cluster is nested within another cluster.
Merchandise Hierarchy Type	Provides details about which type of hierarchy the cluster criteria have been created for.

View Stores

This area displays a nested list of the stores in the cluster you have defined and data for each store for each of the relevant attributes for the Cluster By option you selected to define the cluster. You can see data at the aggregated level as well as at the individual level. Filters are provided so that you can filter the display, for example, by category. You can see aggregated data at a higher level as well as at the store level. Export to Excel is possible for all rows or selected rows.

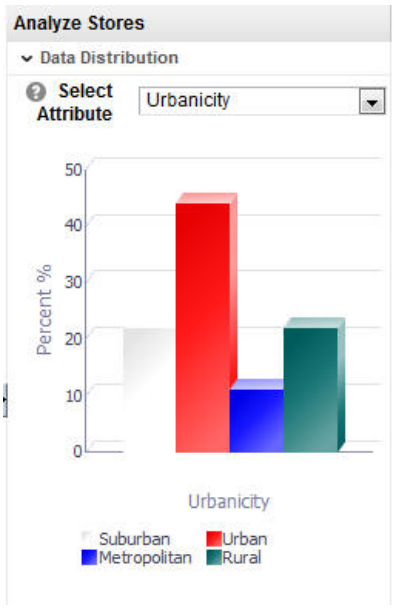
Figure 4–5 Explore Data



Contextual Area

This area provides graphical illustration of the detailed data distribution about the cluster. In Explore Data, the BI displays the data distribution of the location by each participating attribute as well as other configured informational attributes. Advanced Clustering identifies the bins based on the underlying data and displays the histograms. It provides the percentage of stores that are present in a location. For example, a company may have 45 percent of stores in urban areas.

Figure 4–6 Clustering Analyze Stores



Cluster Setup Stage

Use this stage to create various scenarios based on a specified number of clusters. Such scenarios allow you to experiment with different numbers of clusters. You can either define the maximum and minimum number of clusters or define a specific number of clusters to be generated. Once the scenarios are generated, different scenarios can be compared.

You can use this stage to perform what-if analysis by defining one or more scenarios that are based on a specified number of clusters and attributes. Using these scenarios, you can experiment with different numbers of clusters, participating attributes, and their weights. You can either define the maximum number or the minimum number of clusters or alternatively define a specific number of clusters that you want to be generated. You can also use other features in this stage to copy or delete scenarios.

Process

Here is the high-level process for setting up scenarios.

1. Either select the name of a cluster you want to modify or enter a name for the new cluster you want to create.
2. If you want the application to optimize the total number of clusters, enter minimum and maximum values for the number of clusters.
3. If you want the application to generate a specific number of clusters, enter that value. In this case, the application does not optimize the number of clusters.
4. Optionally, configure the weights assigned to the attributes. The total must add up to 100%. Use a value of 0% if you do not want a specific attribute to be part of the cluster.
5. Click the **Execute** icon to execute the scenario. Once the processing is complete, you see the results in the Cluster Results stage.
6. To see a list of all scenarios and the status for each, go to the Scenario List tab.

7. To compare the defining characteristics of two different scenarios, go to the Scenario Compare tab.

Summary

This section summarizes the characteristics of the cluster you are creating scenarios for. It lists the following information:

Table 4–7 Cluster Setup: Summary

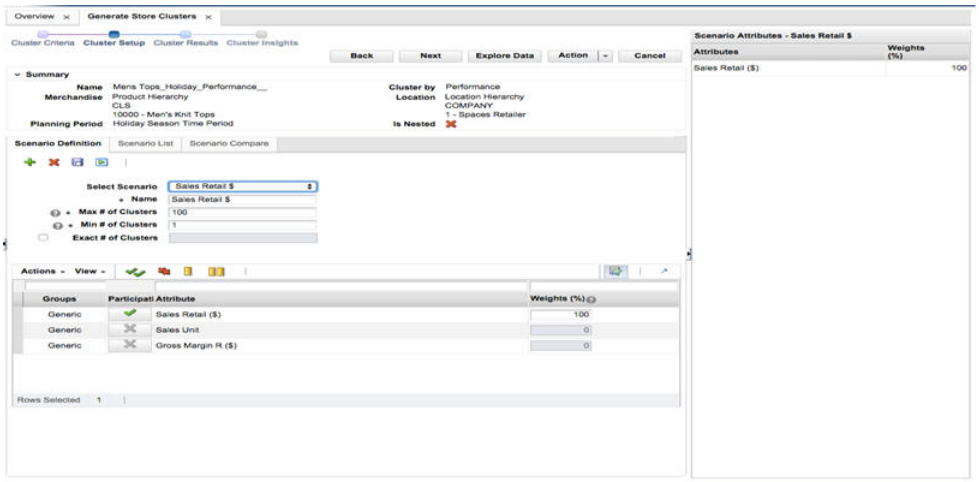
Field	Description
Name	The name you provided for the cluster in the Cluster Criteria stage.
Cluster By	The key feature that defines the cluster. Choose from Consumer Profile, Store Attributes, Product Attribute Profile, Product Performance, or Mixed Attribute.
Merchandise	The merchandise level and the nodes for the cluster.
Location	The location level and the nodes for the cluster.
Fiscal Period	This display varies, depending on the Cluster By option you selected the time period you selected. In general, it tells you the time period you defined for the cluster calculation.
Is Nested	Indicates whether or not the cluster is nested within another cluster.

Scenario Definition Section

This area has three tabs: Scenario Definition, Scenario List, and Scenario Compare.

Scenario Definition Tab

Figure 4–7 Clustering Scenario Definition



The following information is needed to define a scenario.

Table 4–8 Scenario Definition

Field Name	Description
Select Scenario	You can select an existing scenario if you want to modify it.

Table 4–8 (Cont.) Scenario Definition

Field Name	Description
Name	A unique name that identifies the scenario being defined.
Max. # of Clusters	Set the maximum number for the total number of clusters that can be generated. The application determines the optimal number of clusters during the generation process.
Min. # of Clusters	Set the minimum number for the total number of clusters that can be generated. The application determines the optimal number of clusters during the generation process.
Exact # of Clusters	Check box to indicate that the exact number of clusters should be generated. The application does not determine the optimal number of clusters.

Attributes

The Attributes table is used to define which attributes are included in the cluster criteria and the weights that should be assigned to each participating attribute. The following information defines the attributes that are participating or non-participating.

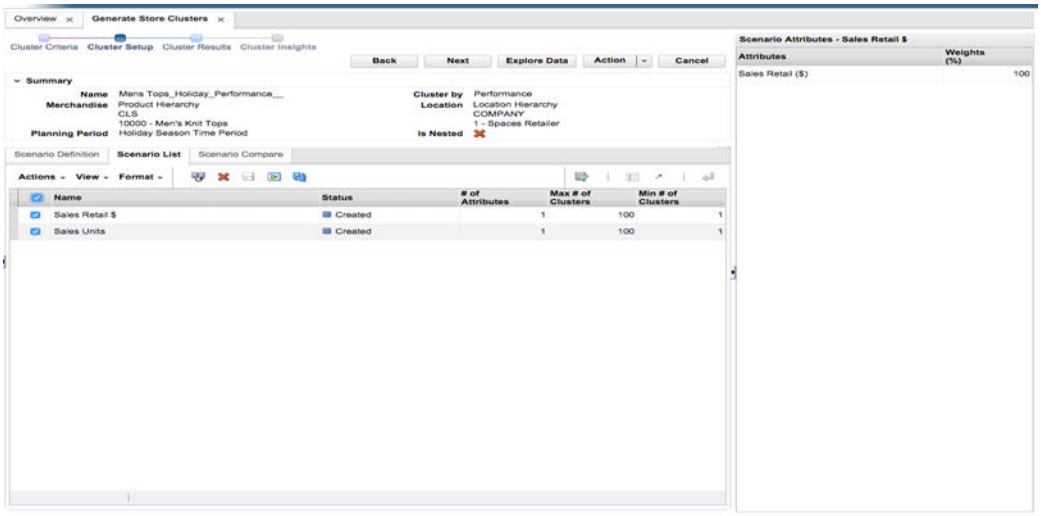
Table 4–9 Attributes

Field Name	Description
Participating	A check in this column indicates that the attributes participates in the cluster criteria.
Groups	Identifies the group.
Attributes	A description of the attribute.
Weights	The weight assigned to the attribute. All participating attributes can have the same weight or each participating attribute can have a unique weight. The total of all the weights must add up to 100%.

Scenario List Tab

The Scenario List summarizes the characteristics for each scenario.

Figure 4–8 Scenario List



You can make a copy a a specific scenario in order to modify it in some way, delete a specific scenario, execute a specific scenario, or save a specific scenario.

Table 4–10 Scenario List

Field Name	Description
Name	The unique name that identifies the scenario.
Status	Created, Executed, Execution Failed, Ready for Approval, Approved, Rejected.
# of Attributes	The number of attributes is defined by the Cluster By option you select and the weights you optionally assign.
Max. # of Clusters	If you provided a value for this in the scenario definition, that number is displayed here.
Min. # of Clusters	If you provided a value for this in the scenario definition, that number is displayed here.

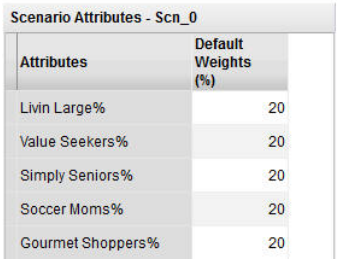
Scenario Compare Tab

You can select two scenarios from the list to compare. The scenarios you select from the Scenario list are shown side-by-side to allow you to compare them.

Contextual Area

Scenario Attributes shows a list of attributes and weights.

Figure 4–9 Clustering Scenario Attributes



Cluster Results Stage

After you select a scenario and execute it, you can see the results in this stage. The application uses the data and the parameters you defined in order to group stores together that are most similar according to the characteristics you selected and to separate stores that are most dissimilar.

Process

You use this stage to view the results of executing a scenario. Then you can approve or reject a cluster.

Summary

This section summarizes the characteristics of the cluster whose results you are examining. It lists the following information:

Table 4–11 Cluster Setup: Summary

Field	Description
Name	The name you provided for the cluster in the Cluster Criteria stage.
Cluster By	The Cluster By option you selected for the cluster in the Cluster Criteria stage.
Merchandise	The merchandise level and nodes for the cluster.
Location	The location level and nodes for the cluster.
Fiscal Period	Calendar Period/Calendar Node is displayed for the Performance Cluster by option. Planning Period is displayed for all the other Cluster By options.
Is Nested	Indicates whether or not the cluster is nested within another cluster.

Scenario Results Section

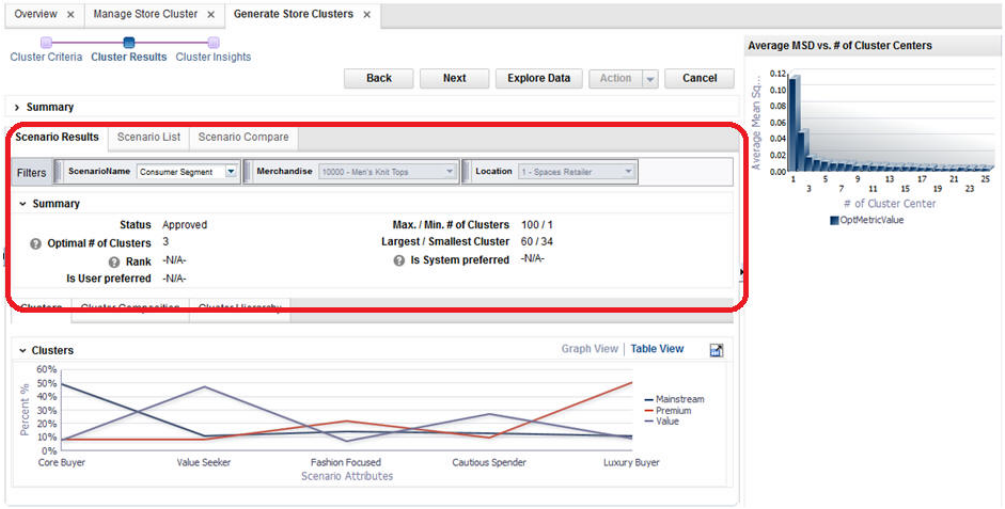
The Scenario Results section consists of three tabs: Scenario Results (with two sub-tabs for Clusters and Cluster Composition), Scenario List, and Scenario Compare.

You can filter the results by Scenario Name (for Clusters), Merchandise, and Location, and Calendar (for performance).

Scenario Results

The Scenario Results tab has three sub-tabs: Clusters, Cluster Composition, and Cluster Hierarchy.

Figure 4–10 Clustering Scenario Results



The Summary section provides an overview of the characteristics of the clusters.

Table 4–12 Scenario Results Summary

Field	Description
Status	Ready for Approval, Approved, or Rejected.
Optimal # of Clusters	The optimal number of clusters determined by the optimization.

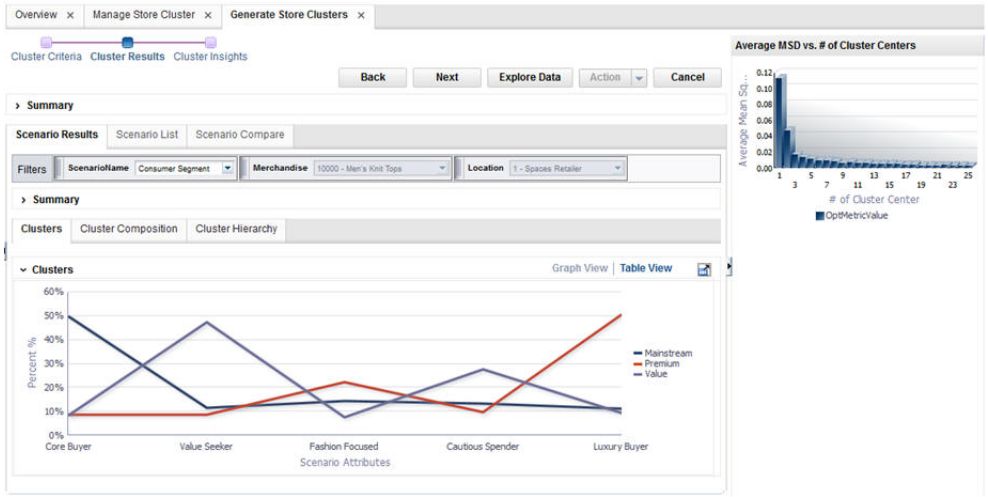
Table 4–12 (Cont.) Scenario Results Summary

Field	Description
Rank	The application compares executed scenarios and ranks them. A value of 1 indicates the best scenario.
Max./Min # of Clusters	The number you provided for the maximum and minimum number of clusters to calculated.
Largest/Smallest Cluster	Provides the sizes of the largest cluster and the smallest cluster in order to show the range of values.
Is System Preferred	Indicates whether or not the system prefers the cluster.
Is User Preferred	Indicates whether or not the user prefers the cluster.

The Clusters section provides the cluster results for each individual cluster in the scenario in either a Graph View or a Table View. The attributes displayed depend on the Cluster By option chosen in the Cluster Criteria stage.

The Graph View shows the percentage for each attribute in the cluster.

Figure 4–11 Cluster Results Graph View



The Table View, shown in Figure 4–12, provides details that can help you analyze the cluster.

Figure 4–12 Cluster Results Table View

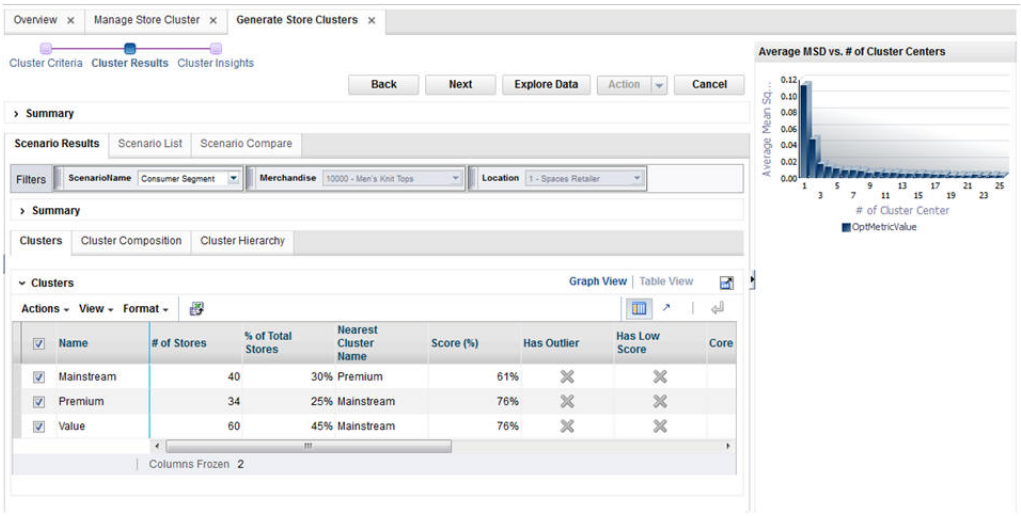


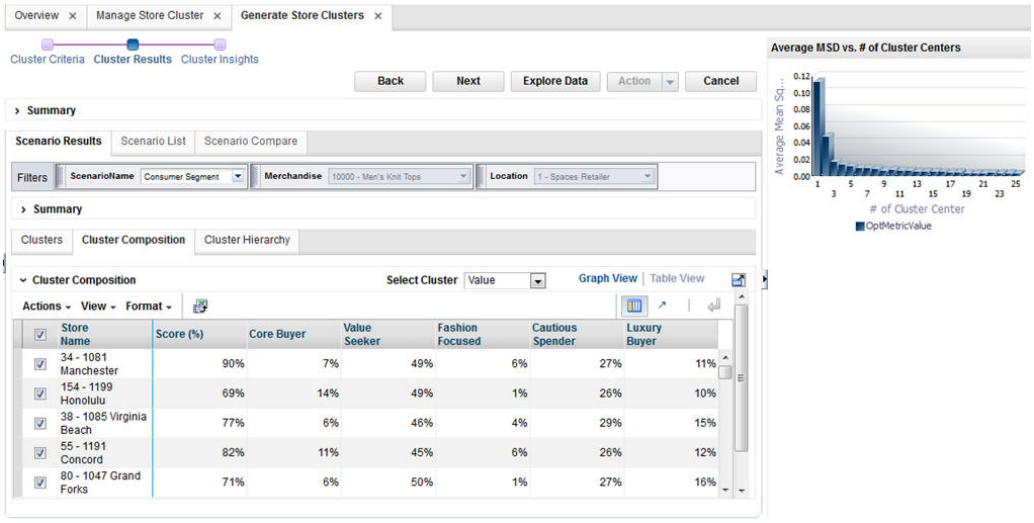
Table 4–13 Scenario Results - Clusters: Table View

Field	Description
Name	The name you assigned to the cluster.
# of Stores	The number of stores in the cluster.
% of Total Stores	The percentage of the total stores that the number of stores represents.
Nearest Cluster Name	The name of the cluster that is most similar to this cluster.
Score %	This value is calculated at the level of store and then averaged to the cluster. The probability, expressed as a percent, of a store being present in this cluster rather than any of the other clusters.
Has Outlier	Indicates a cluster with the number of stores below a threshold. For example, the number of stores are below certain percentage of the number of stores in a cluster.
Has Low Score	The score threshold can be defined in two ways: The default threshold is calculated as a probability of a store that exist in any one of the clusters. The user can further override this threshold at the time of deployment by each Cluster By.
Attributes and their %	For each attribute specific to the Cluster By option, the value indicates the percentage that attribute represents within the total cluster.

The Clusters Composition sub-tab breaks down the cluster into its component parts and shows the percentages for each attribute.

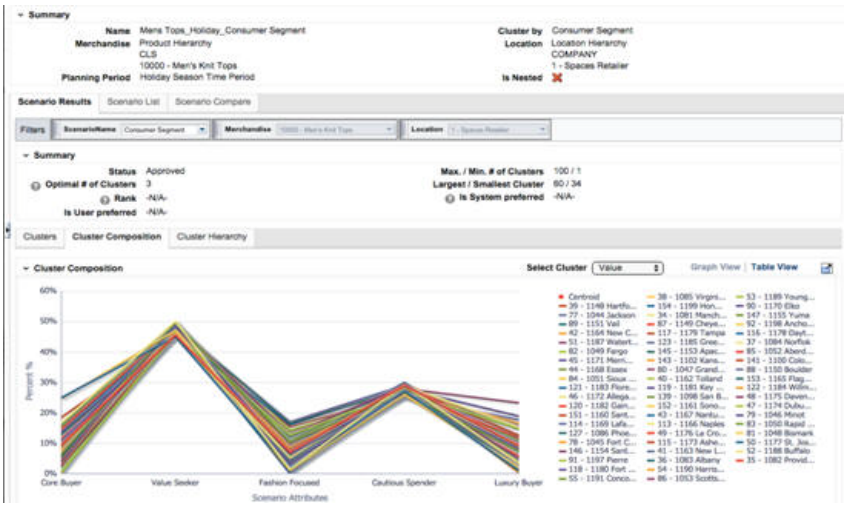
The Table View, shown in Figure 4–13, shows attributes and score percent.

Figure 4–13 Cluster Composition Table View



The Graph View, shown in Figure 4–14, shows the centroid of a cluster and allows the user to compare each store with the cluster centroid.

Figure 4–14 Cluster Composition Graph View



The Cluster Hierarchy sub-tab breaks down the cluster into its component parts and shows store metrics.

Scenario List Section

The Scenario List section contains one table with details about each cluster. In this section, you can approve or reject a cluster.

After a cluster is approved, it is available for other applications.

Figure 4–15 Clustering Scenario List

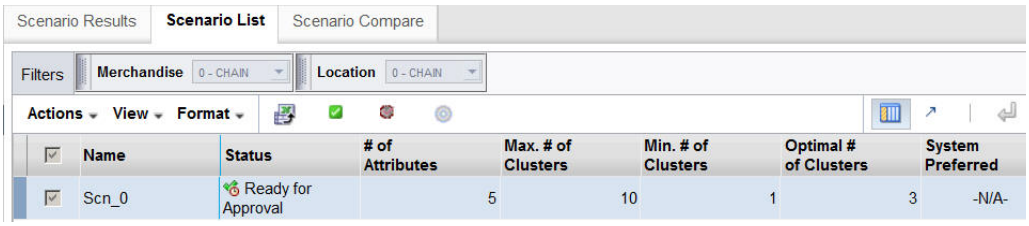


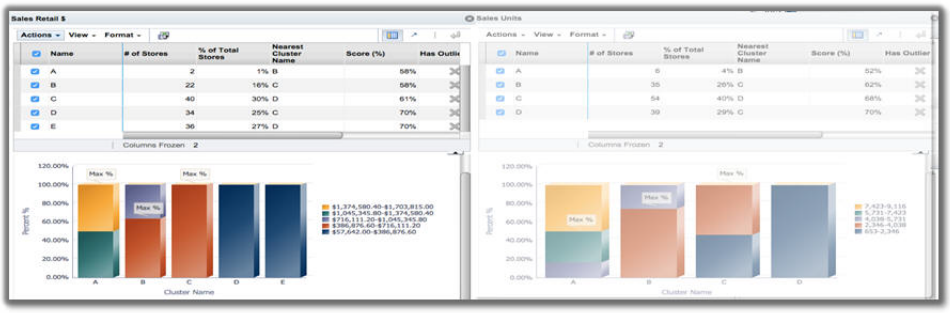
Table 4–14 Scenario List

Field	Description
Name	Name assigned to the scenario.
Status	Ready for Approval, Approved, or Rejected.
# of Attributes	The total number of attributes used, as determined by the weight assigned to each attribute.
Max. # of Clusters	The value used for the maximum in the scenario execution, if this option used.
Min. # of Clusters	The value used for the minimum in the scenario execution, if this option used.
Optimal # of Clusters	The value used for the optimal number of clusters in the scenario execution, if this option used.
System Preferred	Indicates whether the scenario is the one the application prefers.
User Preferred	Indicates whether the scenario is one the user prefers.
Rank Sequence	Indicates the ranking the scenario is given by the application.

Scenario Compare

The Scenario Compare section shows two clusters of your choosing side by side so that you can compare the main results of each, using the same characteristics used in Scenario Results and Scenario List.

Figure 4–16 Compare Scenarios



The information displayed includes:

Table 4–15 Scenario Compare

Field Name	Description
Max. # of Clusters	The value used for the maximum in the scenario execution, if this option used.

Table 4–15 (Cont.) Scenario Compare

Field Name	Description
Min. # of Clusters	The value used for the minimum in the scenario execution, if this option used.
Optimal # of Clusters	The value used for the optimal number of clusters in the scenario execution, if this option used.
Rank	The value for the rank.
Is System Preferred	Indicates whether the scenario is the one the application prefers.
Is User Preferred	Indicates whether the scenario is the one the user prefers.
Smallest Cluster Size	The size of the smallest cluster.
Largest Cluster Size	The size of the largest cluster.
Is Outlier	Indicates a cluster with the number of stores below a threshold. For example, the number of stores are below certain percentage of the number of stores in a cluster.
Attributes	A list of relevant attributes.

Contextual - Average MSD vs # Cluster Centers

This graph indicates how the system identifies the best number of clusters for a given data set. It starts with a small number of cluster centers and searches for the number beyond which there is little improvement in the mean squared distance (MSD). At this point, increasing the number of cluster centers any more only decreases the MSD by a small amount, and the marginal improvement is small.

Scenario Optimality

You can review optimality in the Cluster Results step. Optimality is used to indicate whether or not increasing the number of clusters will result in an improvement in the results. The application calculates optimality using the mean square distance (MSD). The MSD is calculated for each cluster until it reaches the user-defined maximum number of clusters. Optimality is achieved once the value for the MSD ratio plateaus.

Cluster Rank

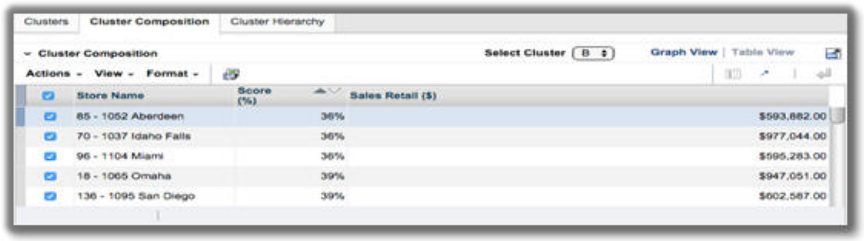
You can see the ranking of all scenarios in the Cluster Results step. The scenario with the highest rank is designated as System Preferred. The ranking is based on the following:

- How many similar stores are contained in the cluster.
- How well separated the clusters are from each other.

Cluster Scores

The application provides scores for clusters, based on the calculated threshold score. The score is based on the assumption that each store has an equal chance of being a member of the cluster. A high score indicates that the store is close to the centroid. A low score indicates that the store is an outlier.

Figure 4–17 Cluster Scores



Outliers An outlier store is defined as one that is beyond a specified distance from the centroid. An outlier cluster is one in which the number of stores is below a defined threshold.

New Stores and Stores with a Poor History

Advanced Clustering supports three post-processing rules in order to allocate stores that are new or that have a poor history. These rules can be configured for each criterion and can be changed during deployment.

- Like Stores. This rule allocates new stores or stores with a poor history to the same clusters that the like location belongs to. It requires a feed to Advanced Clustering that defines the mapping between the location and like locations. This mapping can be configured by merchandise, and one location can be mapped to multiple locations with different weights. For example, a like location can be used to correct a store with poor history or to allocate a new store to a valid performance cluster.
- Largest Clusters. This rule allocates new stores or stores with a poor history to the largest cluster identified by Advanced Clustering. Stores can be allocated to a bigger group of stores. For example, a store that has not yet formed a consumer or customer base can be allocated to the largest cluster.
- Cohesive Clusters. This rule allocates new stores or stores with a poor history to the most compact cluster identified by Advanced Clustering. Stores can be allocated to a compact group of stores. For example, stores can be assigned to a cluster that has not been affected because of outliers.

Insights Stage

The final stage displays a list that shows, for each cluster, a second cluster that is closest to it, as well as a score. This also allows user to create manual clusters in a scenario, rank scenario, mark scenario as user preferred, rename clusters and override any cluster composition.

Each cluster is identified by the following information:

Table 4–16 Insights Stage

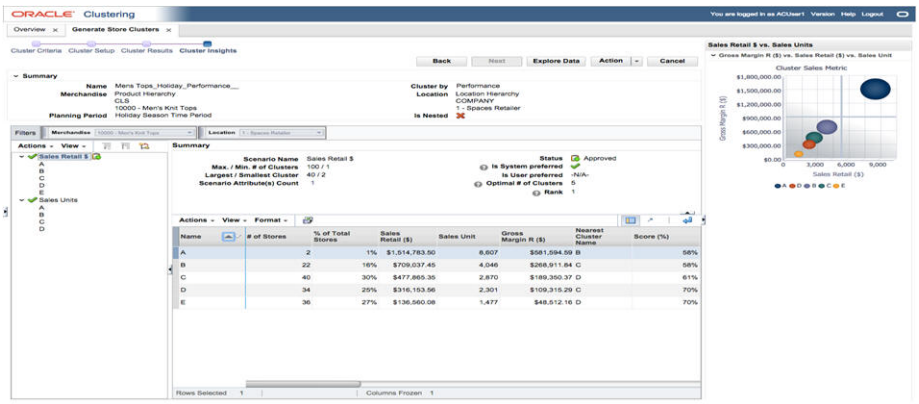
Field Name	Description
Name	The name assigned to the cluster.
Nearest Cluster Name	The name of the cluster that is most similar to the named cluster.
# of Stores	The number of stores in the cluster.
Is Outlier	Whether or not the cluster is considered an outlier. If it is an outlier, you may want to review that store.
Gross Margin ($)	Financial information about the store.

Table 4–16 (Cont.) Insights Stage

Field Name	Description
Score %	This value is calculated at the level of store and then averaged to the cluster. The probability, expressed as a percent, of a store being present in this cluster rather than any of the other clusters.
Has Low Store	Indicates a cluster that falls below a defined threshold.
% Total Stores	The percentage of the total stores that the number of stores represents.
Sales Retail	Financial information about the store.
Sales Unit	Financial information about the store.

You use this stage to view all scenarios, examine and compare sales metrics for the various clusters, and manage clusters by approving or rejecting, merging or deleting clusters.

Figure 4–18 Clustering Insights



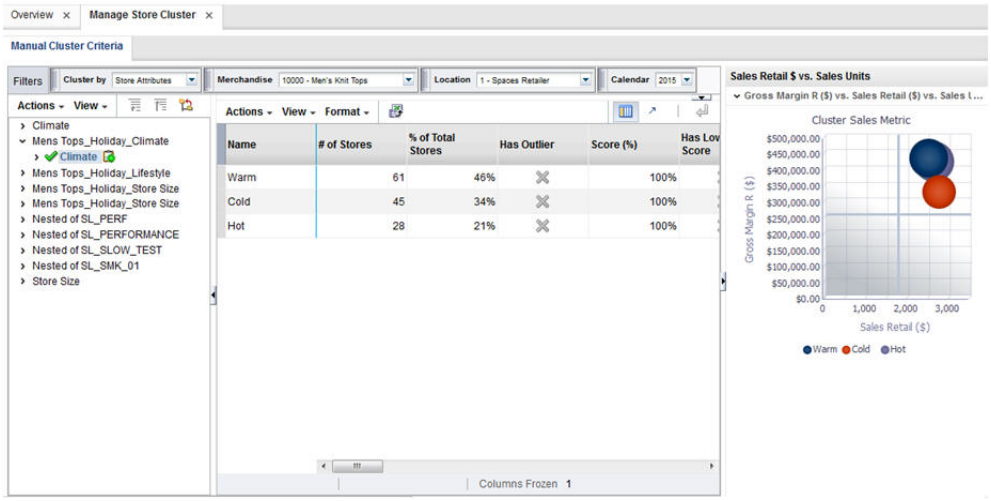
Contextual Information

The four graphs displayed in the Clustering Insights stage are also displayed in the Manage Store Clusters tab. For a discussion, see Contextual Information.

Manage Store Clusters Tab

You can use the Manage Store Clusters tab to view a list of cluster criteria and associated summary details. You can create manual clusters in a scenario, rank/approve/reject scenario, mark scenario as user preferred, rename clusters and override any cluster composition within a scenario.

Figure 4–19 Clustering Manage Cluster Criteria



The following information is displayed.

Table 4–17 Manage Store Clusters

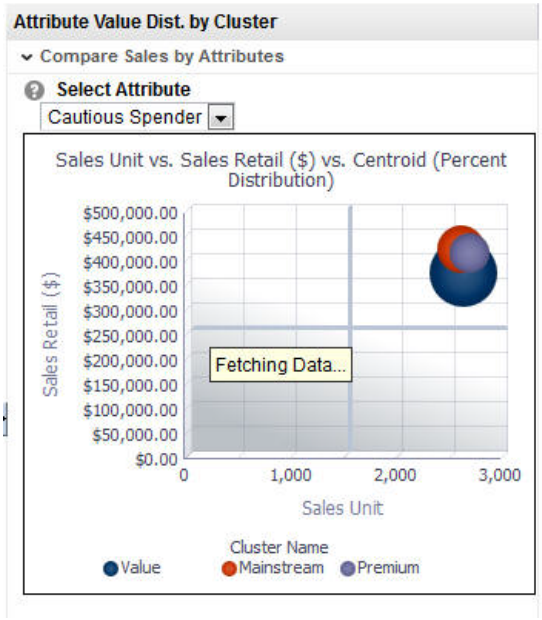
Field	Description
Is Nested	Indicates whether or not the cluster is nested within another cluster.
Is Deployed	Indicates that the cluster has been deployed.
Is Shared	A check mark indicates more than one merchandise or location node is used in the cluster.
Scenario Created	The number of scenarios that were created.
Scenario Executed	The number of scenarios that were executed.
Name	The name of the scenario.
Status	Ready for Approval, Created, Approved, and Rejected.
# Attributes	The total number of attributes used, as determined by the weight assigned to each attribute.
Max # Clusters	The value used for the maximum number of cluster centers in the scenario execution, if this option used.
Min # Clusters	The value used for the minimum number of cluster centers in the scenario execution, if this option used.
Optimal # Clusters	The value used for the optimal number of clusters in the scenario execution, if this option used.
System Preferred	Indicates whether the scenario is the one the application prefers.
Rank Sequence	Indicates the ranking the scenario is given by the application.

Contextual Information

The following four charts are available.

Attribute Value Dist By Cluster. The chart shows a comparison of sales with the average value distribution by clusters. The Y axis is sales retail and the X axis is sales units. The Z bubble size shows the average value for the selected attribute for the cluster.

Figure 4–20 Attribute Value Dist. by Cluster

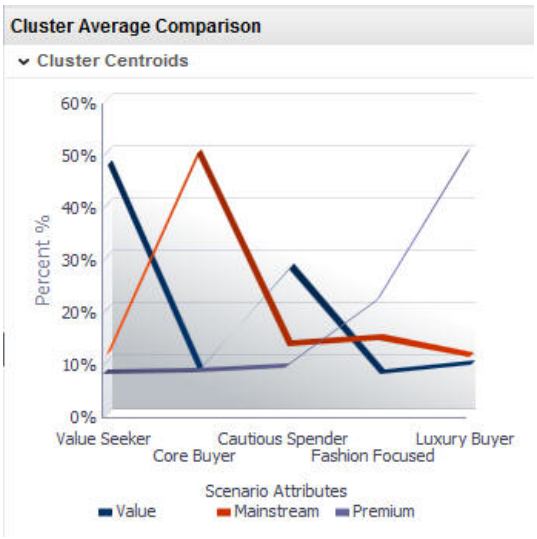


Sales Retail vs. Sales Units. The chart shows a comparison of sales retail and sales unit. The Y axis is sales retail and the X axis is sales unit. The Z bubble size shows the gross margin retail for a cluster.

Figure 4–21 Sales Retail vs. Sales Units

Cluster Average Comparison. The chart overlays the centroid of the cluster and provides a comparison of cluster averages so that you can rename the clusters based on the information in the graph.

Figure 4–22 Cluster Average Comparison



Cluster Comparison. The chart shows the stacked contribution of each attribute by percentage for each cluster. You can use this chart to determine which cluster contribute the most for an attribute by viewing the clusters side by side.

Figure 4–23 Cluster Comparison