Oracle® Endeca Information Discovery

Glossary

Version 3.2.0 • January 2016

Copyright and disclaimer

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

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

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

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

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

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

A

application (Studio)

A collection of pages and Studio components used for information discovery. A Studio application is bound to a single Endeca Server data domain, which is made up of one or more data sets.

Applications may be public or private. To get access to a private application or private pages in a public application, Studio users are granted application membership.

public application (Studio)

private application (Studio)

application administrator (Studio)

An application member who has complete control over an application. Application administrators can edit the application content, configure the application, and control application membership, including whether a member is an administrator.

The user who creates the application is automatically an application administrator. Application administrators can also assign other users as application administrators.

application member (Studio)

application member (Studio)

A Studio user who has been granted access to view all of the content of a Studio application. Application members cannot edit or configure the application.

Users can be granted application membership by application administrators.

application administrator (Studio)

assignment

When a record in the data domain index has a value for an attribute, this value is referred to as an assignment on a record.

attribute

Basic unit of a record schema. Attributes describe records stored in the index for the data domain. The data domain is hosted by the Oracle Endeca Server.

For data records, attributes provide information about a record. For system records, an attribute is a configuration setting.

The term *attribute* collectively refers to both standard attributes and managed attributes.

- Standard attributes are described by attribute schema records. The attribute schema records that describe standard attributes are known as Property Description Records (PDRs).
- Managed attributes are also described by attribute schema records. The attribute schema records that describe managed attributes are known as Property Description Records (PDRs) and Dimension Description Records (DDRs).

See also attribute schema record, Property Description Record, Dimension Description Record, standard attribute, and managed attribute.

primary key attribute

single-assign attribute

multi-assign attribute

unique attribute

standard attribute

managed attribute

attribute group

A group of attributes defined within a view of application data. In Studio, attributes are often displayed in the context of their groups.

All attributes that are not members of user-defined groups automatically belong to a group with a default name of `Other`.

Additional groups can be created in Studio using the **Attribute Group Manager**.

attribute schema record

Records that describe attributes of attributes. Attribute schema records consist of PDRs and DDRs. Similar to configuration on attributes themselves, attribute schema records can also be configured to be searched, navigated, or filtered out.

Property Description Record (PDR)

Dimension Description Record (DDR)

attribute type

The required format for an attribute value (such as string, integer, boolean, date/time).

attribute value

An assignment from an attribute on a record, used as a tag, or label, to classify a record in your data set. Tagging a record with a value identifies that record as a valid result when a user queries for the value. A record can have more than one assignment from a specific attribute—such an attribute is known as multi-assign or multi-value.

The term *attribute value* applies to values on both standard and managed attributes.

B

base filter (Studio)

base view (Studio)

Bulk Load interface

base filter (Studio)

A filter that is applied automatically when users first display an application.

Each application data set can be configured with a single base filter, which filters the data based on a single attribute value.

base view (Studio)

A base view is the view of an entire data set.

Each data set has a corresponding base view. The base view is created automatically from the physical records and always contains exactly the same records and attributes as the data set.

The base view also uses the name and description of the data set.

See also view.

view (Studio)

dimension (in views)

metric

view query (Studio)

Bulk Load interface

An Oracle Endeca Server interface for loading data into the data domain. It achieves high performance for strictly additive ingest operations of varying amounts of data. An Integrator ETL connector is available for this interface.

Data Ingest Web Service (DIWS)

C

collection

Represent records grouped into logical buckets, by one of their attributes. This attribute must be unique and single-assign, and is known as the collection spec (or a collection primary key). It identifies a collection in the schema, and determines which attributes belong to the collection.

Records within a collection may not have the same set of attributes, but all records in a collection usually have a similar or related purpose in an application powered by the Endeca Server. Examples of collections include products for sale alongside "how to" articles, vehicles alongside warranty claims, structured HR records of employee changes alongside satisfaction surveys.

A single data domain may include many collections. In fact, having collections in the Endeca data domain lets users load data directly into collections within a data domain, and then issue Conversation Web Service queries or EQL queries across (or on top of) all records in all collections. For example, you can select records from each collection, and return results in a join. You can also create filter rules across two different attributes, each belonging to a separate collection.

In Studio, collections from the Endeca Server are known as data sets. For example, a specific Endeca Server connection in Studio may include many data sets.

As opposed to views, which represent virtual sets of records in Studio, collections represent physical records, and reflect how records are grouped when they are loaded into the index for a particular data domain (Endeca Server collection).

See also data sets and filter rules.

cluster

See Oracle Endeca Server cluster or data domain cluster.

data domain cluster

Oracle Endeca Server cluster

component

See Studio component or Integrator ETL component.

Studio component

Integrator ETL component

configuration update

The process of loading configuration changes to the index for the data domain hosted in the Oracle Endeca Server.

Configuration Web Service

Web service that allows you to update the schema and configuration for the records of your data domain.

See also Oracle Endeca Server Web services.

Oracle Endeca Server Web services

Conversation Web Service

Provides the primary means of querying data in the data domain hosted in the Oracle Endeca Server. Used by Studio to query the Oracle Endeca Server.

See also Oracle Endeca Server Web services.

Oracle Endeca Server Web services

D

Data Ingest Web Service (DIWS)

A Web service that provides an interface to ETL tools to load data into the data domains hosted in the Oracle Endeca Server. Besides adding new data records and attribute records (schema) to the data domain, the data ingest operations include modifying (updating) and deleting existing records in the index for the data domain.

See also Oracle Endeca Server Web services.

Bulk Load interface

Oracle Endeca Server Web services

data domain

A logical collection of data and metadata managed by the Endeca Server. Through its interfaces, the Endeca Server allows for the data loading, configuration, and querying of a data domain. A data domain may impose order on subsets of its data through entities (known in Studio as views).

A data domain is the largest unit of data over which the Endeca Server allows queries to be expressed. It represents a discrete set of data and includes indexed data records and system records. (Applications wishing to correlate, join, or display data from multiple data domains must do so themselves.)

Each data domain runs one or more Dgraph processes. Each of the Dgraph processes runs on the Oracle Endeca Server node; depending on the number of Dgraph processes configured for the data domain, it can be hosted by more than one Endeca Server nodes.

To create a data domain hosted in the Oracle Endeca Server, you first create a data domain profile, using either the Endeca Server Web service interfaces or the `endeca-cmd` utility. You can then use this profile to create a new data domain. The Dgraph processes for this data domain will then be allocated and started on the available Endeca Server nodes.

See also Oracle Endeca Server and Endeca Server connection (Studio).

index

data domain cluster

data domain cluster

A set of Dgraph processes that together handle requests for a given data domain.

One of these Dgraph processes is identified as the leader node and is responsible for handling all write operations (updates, configuration changes), while the remaining Dgraph processes serve as read-only follower nodes and serve only non-updating queries to the data domain's index. All Dgraph nodes in a given data domain cluster utilize the same index residing on shared storage. Each data domain cluster node is hosted on a separate Endeca Server node.

The Dgraph process of the Oracle Endeca Server is the main computational module that provides the features of the Endeca Server, such as search, refinement computation, and guided navigation.

See also leader node and follower node.

leader node

follower node

data record

Data records represent the actual data that is being analyzed and manipulated using Studio. These records are the individual items that the user navigates to in an Oracle Endeca Information Discovery application.

Data records generally correspond to traditional records in a source database. Unlike source records, however, data records have been standardized for consistency, and then classified with attribute values.

See also data domain, record, primordial record, system record, and source record.

record

primordial record

source record

data set (Studio)

A set of data from a single source, such as an uploaded file, a Studio data source, or an Endeca Server connection.

Each Studio application can contain one or more data sets. Users can then create views that reference attributes from one or more of the application data sets.

The data sets are contained within the Endeca Server domain for the application.

In Endeca Server, data sets are referred to as collections.

data source (Studio)

A source of data from which users can create a data set for a Studio application.

A data source may be a connection to Oracle BI Server, or a JDBC connection to a relational database.

Dgraph

The name of the process for the data domain hosted in the Oracle Endeca Server. When the Dgraph runs, it creates the index of the data loaded into the data domain. The index contains source records, system records, which constitute the schema, and configuration documents.

See also Oracle Endeca Server, data domain, and data domain cluster.

dimension (in views)

An attribute from a view that can be used for grouping or aggregation. Usually, a dimension is a managed attribute from the physical records.

See also view.

view (Studio)

metric

base view (Studio)

view query (Studio)

Dimension Description Record (DDR)

A system record used to define the behavior of a managed attribute. Each managed attribute has a DDR. The DDR configuration includes rules for displaying the managed attribute and using it in searches.

See also system record, Property Description Record (PDR), and Global Configuration Record (GCR).

E

edge

In an Integrator ETL graph, edges represent data flowing from one component to another.

Each edge is characterized by metadata that you can assign to it. Metadata specifies information about which fields of data are being provided from one component to the next.

Since data produced by one component needs to be received by another component through the edge, the metadata on the edge describe the fields of data that are involved in this operation.

graph

project

Integrator ETL

Integrator ETL component

end user (Studio)

End users are Studio application content consumers. These could include executives seeking a dashboard view as well as others who need to drill through interactive visualizations and reports.

Most end users are application members, able to view application content but not edit it.

See also Studio and application member.

Entity and Collection Configuration Web Service

Web service (also known as `sConfig`) that allows you to create and update collections and views for collections or for data sets.

See also collection, view, and Oracle Endeca Server Web Services.

Oracle Endeca Server Web services

Endeca Query Language (EQL)

EQL is an SQL-like language designed specifically to query and manipulate data from the Oracle Endeca Server. It enables Endeca Server–based applications to examine aggregate information such as trends, statistics, analytical visualizations, comparisons, and more.

Endeca Server connection (Studio)

The source of the data for a Studio application. An Endeca Server connection represents a connection to an Endeca Server data domain.

Endeca Server connections can include role-based security to control access to the data, as well as configuration to control who can connect an application to it.

Each Endeca Server connection represents a data domain in the Endeca Server.

See also data domain.

enrichment (Studio)

An enrichment enhances a data set by creating or editing attributes in an application data set.

For example, an enrichment may extract relevant terms to create a new attribute, or standardize the terminology used in an attribute value. Enrichments rely on Data Enrichment plugins that are installed as part of the Endeca Server software package.

F

filter rule (Endeca Server)

In the Endeca Server, a rule configured to apply a filter across attributes from many collections at once. When you filter a collection by an attribute from a specific filter rule, other collections are refined by other attributes from the same filter rule. For example, when refining a "sales transaction" collection by the sales date, the "Twitter posts" collection is refined by the post date. You can apply filter rules to different types of filters, such as selection filters (used in refinements) and EQL filters (used for range filters).

In Studio, a filter rule is known as a refinement rule. Refinement rules allow users to connect attributes from different data sets in an Endeca Server connection. When end users refine a data set by an attribute that belongs to a refinement rule, other data sets are refined by the other attributes that belong to the refinement rule. For example, when refining a sales transaction data set by the sales date, the Twitter posts data set is refined by the post date.

See also refinement rule (Studio).

follower node

A node in a data domain cluster responsible for processing queries. The follower node does not update the index, although it has access to its latest copy. Each data domain cluster can have more than one follower node. In a single-node data domain cluster, a leader node is also a follower node. All nodes (including follower nodes) must have write access to a shared file system.

See also data domain, data domain cluster, and leader node.

data domain cluster

leader node

full index load

A load of the full index, either as an initial load or a complete reload. This includes all user-defined data for the data domain.

incremental index update

full index refresh

index

full index refresh

The process of reloading the full data set for the data domain into the Oracle Endeca Server. This entails cleaning out the old data and replacing it with the new data. In SQL terminology, full index refresh is equivalent to the "Truncate and load" operation on a table in the database.

Unlike in a full index load, in a full index refresh the Oracle Endeca Server retains some user-defined data for the data domain.

incremental index update

full index load

index

G

Global Configuration Record (GCR)

A single system record used to define global configuration information for the data domain. The GCR includes rules for searches and spelling correction.

See also system record, Property Description Record (PDR), and Dimension Description Record (DDR).

graph

In Integrator ETL, a graphical layout that contains a set of Integrator ETL components.

edge

project

Integrator ETL

Integrator ETL component

guided navigation

The presentation of valid follow-on refinement choices to the user, implemented using the **Available Refinements** and **Selected Refinements** components in Studio.

See also Studio and Studio component.

I

incremental index update

The process of loading changes made to the index of a data domain hosted in the Endeca Server.

full index load

full index refresh

index

index

A general term that refers to the many types of indices created for a data domain by the Dgraph process of the Oracle Endeca Server. For example, in the standard search index, each entry corresponds to a searchable document containing the correct term; in the wildcard search index, each entry corresponds to a document enabled for wildcard search that contains the correct term.

incremental index update

full index load

full index refresh

index configuration documents

A set of XML-based configuration files for the data domain that define how your records, standard attributes, and managed attributes are indexed by the Dgraph process of the Oracle Endeca Server. The index configuration is the mechanism for implementing a number of Oracle Endeca Server features such as record search, value search, snippeting, relevance ranking, precedence rules, and thesaurus entries.

IKM SQL to Endeca Server

The IKM SQL to Endeca Server provides integration and loading modules that enable writing source data to an Endeca Server target within Oracle Data Integrator.

Integrator Acquisition System

The Integrator Acquisition System, or IAS, is a set of components that crawl source data stored in a variety of formats including: file systems, delimited files, JDBC databases, Web servers, and custom data sources. IAS transforms the data, if necessary, and outputs the data to an XML file or a Record Store instance that can be accessed by Integrator ETL for use in the Endeca Server.

Integrator ETL

Integrator ETL is a high-performance data integration platform that lets you extract source records from a variety of source types (from flat files to databases) and send that data to the Data Ingest Web Service, which in turn loads the records into the Oracle Endeca Server.

edge

graph

project

Integrator ETL component

Integrator ETL component

In Integrator ETL, a graphical object that you add to your graph to perform some kind of data manipulation.

Components are characterized by several types, with various components belonging to each type. Some of the component types are readers, writers, transformers, and joiners.

Each component has an input and output port. Even though components are joined sequentially in a graph, all components that are in the same phase of the graph run in parallel.

edge

graph

project

Integrator ETL

K

key-value pair

Assignments on standard attributes for the records in the index of the data domain use key-value pairs, or KVPs. In each KVP, *key* is the name of an attribute, and *value* is an assigned value for this attribute.

See also standard attribute and standard attribute value.

L

leader node

A single Dgraph process in the data domain cluster responsible for receiving and processing updates to the index and configuration. This node can also be configured to process queries, like other nodes. This node is responsible for generating information about the latest versions of the index and propagating this information to the follower nodes.

When you create a new data domain cluster, the Endeca Server software creates and starts the leader node and then the follower nodes. Each data domain cluster (that is not read-only) is created with one leader node and a specified number of follower nodes.

The leader node periodically receives full or incremental index updates. It also receives administration or configuration updates. It is the only node in the data domain cluster that makes updates to the index. After processing updates, the leader publishes a new version of the data and notifies all follower nodes, alerting them to start using the updated version of the index. The follower nodes acquire read-only access to an updated version of the index.

See also follower node, data domain cluster, and Endeca Server cluster.

data domain cluster

follower node

M

managed attribute

An attribute for which a hierarchy of attribute values is attached. Managed attributes are used to support hierarchical navigation. For example, when using a Location attribute to filter records, users may navigate by North America > United States > California.

Managed attributes are described by schema records — Property Dimension Records and Dimension Description Records. An assignment from a managed attribute is known as a managed attribute value (or mval).

See also attribute and standard attribute.

attribute

primary key attribute

single-assign attribute

multi-assign attribute

unique attribute

standard attribute

managed attribute value

Values of a specific attribute organized in an enumerated list or a hierarchy of values. Often, managed attribute values are sources from an external system. Managed attribute values can be added to the Endeca Server as records, and can be manipulated in Studio.

metric

A value calculated from a view attribute and aggregated across a view's dimension value.

A metric is defined using an expression containing an arithmetic formula used for the aggregation. Each metric must contain at least one aggregation function (for example, "SUM(x)" or "AVG(y)"). For example, from the attribute "Sales", you might calculate the metric "Total Sales", which is the total of the values of the "Sales" attribute for each value of a dimension such as region or product line.

Some metrics are defined for a specific instance of a Studio component, to populate a table or generate a graphical visualization. Views may also have predefined metrics that are available to all components.

See also view (Studio) and dimension (in views).

view (Studio)

dimension (in views)

base view (Studio)

view query (Studio)

multi-assign attribute

An attribute for which a record in a data domain may have more than one value. For example, because a book may have more than one author, the Author attribute would be multi-assign.

In Studio, multi-assign attributes are referred to as multi-value.

attribute

primary key attribute

single-assign attribute

unique attribute

standard attribute

managed attribute

N

navigation query

navigation query

Returns a list of matching records based on search or the current navigation choices made by the user. A navigation query uses refinements, to refine the search at each step in the guided navigation process.

O

Oracle Endeca Information Discovery

Oracle Endeca Information Discovery applications guide people to better decisions by combining the ease of search with the analytic power of business intelligence.

Users get self-service access to the data they need without having to specify in advance the queries or views they want. At the same time, the user experience is data driven, continuously revealing the salient relationships in the underlying data for them to explore.

Oracle Endeca Query Language (EQL)

The Oracle Endeca Query Language, or EQL, enables interactive applications that allow users to explore aggregate and statistical views of information using EQL-enabled Studio components.

Oracle Endeca Server

The core search-analytical database. It organizes complex and varied data from disparate source systems into a faceted data model that is extremely flexible and reduces the need for up-front data modeling. This highly-scalable server enables users to explore data in an unconstrained and impromptu manner and to rapidly address new questions that inevitably follow every new insight. Oracle Endeca Server maintains the index of records for the data domain in memory, receives queries, executes them against the stored index, and returns the results.

It is useful to recognize that the term "Endeca Server" may refer to the Endeca Server software package, and to the Endeca Server Java application hosted in the WebLogic Server. Whenever this distinction is needed, the documentation refers to the software package as "the Endeca Server", and to the Java application as the "Endeca Server Java application".

You install the Endeca Server software on Linux or Windows machines (which could be VM images or machines in a private or public cloud) that are running the WebLogic Server and that will host multiple Endeca data domains. Once the Endeca Server package is installed in the WebLogic Server, the WebLogic Server starts the Endeca Server Java application.

The Endeca Server software exposes almost all of its APIs as SOAP web services.

The Endeca Server Java application manages the data domain clusters hosted in the Endeca Server cluster. When you first install the Endeca Server package on multiple machines, it will have no data domains. You then create data domain profiles and use them to create named data domains. The Endeca Server cluster locates ports on which to start the Dgraph processes for the data domain, and identifies which Dgraph configuration flags to use. The Dgraph processes handle requests made to the data domain.

Once a data domain is created, you only need to use the name of the data domain to manage it. You don't need to know which port the Dgraph processes for the data domain are running on, as the Endeca Server keeps track of that information using its Cluster Coordinator services. This name-only reference to the data domains makes it much easier to enable and disable them and perform other data domain management operations.

The Endeca Server has a set of commands, available through `endeca-cmd`, with which you create and control data domains. Optionally, you can use the Web services of the Endeca Server for this purpose.

See also data domain, and Oracle Endeca Server Web services.

Oracle Endeca Server Web services

Oracle Endeca Server cluster

A deployment of multiple Endeca Server instances that host and manage multiple clustered data domains.

The Endeca Server cluster serves as a building block for building other services on top of the Oracle Endeca Server software. It also supports on-premise deployments of Endeca Information Discovery. It runs in a WebLogic domain on a homogeneous cluster of hardware nodes, all of which have access with write permissions to shared storage. It allows system administrators to add additional servers to the cluster over time.

It encapsulates the details of setting up data domain clusters, by keeping details about the location of the Endeca Server nodes and hosted data domain nodes hidden from the end users. (The end users are users of any front-end application, such as Studio, configured on top of a particular data domain cluster.)

The Endeca Server cluster makes decisions about how to allocate resources to data domains, performs load balancing and routing of requests to the appropriate nodes, and maintains availability of data domains. It provides a unified interface for provisioning and managing data domains while automating the process of data domain cluster setup, allocation of processing resources, and configuration.

It lets you create a reusable set of data domain profiles for data domains, such as small, medium, and large. Each of the data domain profiles maps to different allotments of the underlying hardware resources on the Endeca Server nodes. When you create a data domain cluster, a previously defined profile is used initially; a cluster can be changed after it has been created.

See also data domain, data domain cluster, Oracle Endeca Server, and Oracle Endeca Server Web services.

Oracle Endeca Server Web services

Used as interfaces for managing and configuring data domains hosted in the Oracle Endeca Server, as well as for loading data, making configuration changes, and sending queries. Publicly available Web services include the Data Ingest Web Service, the Conversation Web Service, the Configuration Web Service, the Entity and Collection Configuration Web Service (`sConfig`), the Cluster Web Service, and the Manage Web

Service. Additional Web services exist, but their interfaces are used by the internal processes of the Oracle Endeca Server.

Together, they provide an API interface to a data domain hosted in the Oracle Endeca Server.

Data Ingest Web Service (DIWS)

Conversation Web Service

Configuration Web Service

Entity and Collection Configuration Web Service

P

physical records

Records stored in the index for the data domain hosted in the Oracle Endeca Server. Each record is described by a set of attributes, with each attribute having a name, a type, and one or more values.

Physical records are specifically source data records. While schema and application configuration are also stored as index records, for the purposes of views, they are not included in the physical records.

See also schema for records, source record, and system record.

precedence rule

A relationship between two attributes that establishes a navigation or display preference based on a set of predefined criteria, known as triggers for the rule.

primary key attribute

See record specifier attribute.

attribute

single-assign attribute

multi-assign attribute

unique attribute

standard attribute

managed attribute

primordial record

The most basic infrastructure of the schema for the records that are stored in the Dgraph process of the Oracle Endeca Server. Primordial records are created automatically and used as the basis for the system records.

See also record, system record, and data record.

data record

record

source record

private application (Studio)

A Studio application that is only visible to application members. Users who are not members cannot see or select the application.

application (Studio)

public application (Studio)

private page (Studio)

An application page that can only be viewed by application members.

public page (Studio)

project

In Integrator ETL, the location where you create operations to manipulate your data. You can put one or more graphs into a single project.

edge

graph

Integrator ETL

Integrator ETL component

Property Description Record (PDR)

A system record used to define the format and behavior of a single attribute. Each attribute has a PDR. The PDR configuration includes rules for uniqueness, searches, and navigation.

See also system record, Dimension Description Record (DDR), and Global Configuration Record (GCR).

Provisioning Service

The Provisioning Service enables dynamic application creation using data uploaded from the user's desktop, or imported from an Oracle BI Server or using JDBC.

public application (Studio)

A Studio application that is visible to all logged-in Studio users.

All logged-in users can see the application and select it in order to view public pages. Application members can also view private pages.

application (Studio)

private application (Studio)

public page (Studio)

A Studio application page that can be viewed by any logged-in user, including non-members.

Non-logged-in users must use the page URL to get access to the page. They are prompted to log in before the page is displayed.

private page (Studio)

R

record

The fundamental unit of data in the data domain hosted in the Oracle Endeca Server. Records are assigned attribute values. An assignment indicates that a record has a value for an attribute. A record typically has assignments for multiple attributes. For each assigned attribute, the record may have one or more values.

A record may be a primordial record, a data record, or a system record.

See also assignment, attribute, primordial record, data record, record specifier attribute, and system record.

data record

primordial record

source record

record query

Used to return the details for a single record.

record search

A query that returns results based on a user-specified text string by filtering the record set to include only those records that have at least one attribute whose value matches a specified search term (keyword). The result of a record search is a set of records based on the user-defined keyword(s), plus any follow-on query information.

See also value search.

value search

record set specifier

A record set specifier is a set of record attributes that identify a set of records to be updated or deleted.

record specifier attribute

An attribute used to uniquely identify a record in the data domain. An attribute is unique when no two records in this data domain have the same value for it. Each data domain must have at least one record specifier standard attribute. This requirement is enforced when you are adding records to an empty data domain that does not yet have a schema (system records) defined in it.

You can use any attribute as a record specifier as long as an assignment value for the record specifier attribute is guaranteed to be unique. This is achieved when the PDR (Property Description Record) for the record specifier attribute has the `mdex-property_IsUnique` attribute set to `true`. This guarantees that a value for this attribute may be assigned to at most one record. (Note that by default, a standard attribute is not unique. To make a standard attribute unique, you must update the standard attribute configuration in its PDR, before loading any records.)

For example, for a book, the ISBN number could serve as a record specifier attribute, because each book has only one unique value.

See also unique attribute.

refinement

Represents a way to reduce or refine the current query’s record set based on the attribute values.

The key feature of the Endeca Server, known as guided navigation, is that after a user creates a query using record and value search, only valid remaining refinement values are provided to the user to refine that query. This allows the user to reduce the number of matching records without creating an invalid query.

Refinements fall into two categories:

- ***Suggested refinements***, which you can use for guided navigation. When you select a specific refinement, you refine your result set.
- ***Applied refinements***, which show your current location in the guided navigation process. Applied refinements can be explicitly-selected or implicit. Implicit refinements are those that do not narrow your result set.

Working with refinements implies an iterative process of first selecting a top-level refinement in the user interface, expanding and collapsing it, then possibly selecting another refinement and expanding it once, then twice (if it has a hierarchy).

refinement rule (Studio)

A rule configured to link attributes in different application data sets. When end users refine a data set by an attribute that belongs to a refinement rule, the other data sets are refined by the other attributes that belong to the refinement rule.

For example, a data set containing automotive data includes the make and vehicle ID number (VIN) of cars that were sold. Another data set containing warranty claims also includes the make and VIN of cars for which warranty claims were filed.

If there are refinement rules for the make and VIN attributes, then when users refine one data set by a make or VIN, the other data set also is refined by that make or VIN.

See also filter rule (Endeca Server).

relevance ranking

A search interface feature that lets the developer control the order in which record or value search results are displayed to the end user. A relevance ranking module assigns ranking scores to results based on its predetermined criteria, such as the frequency of a user’s query terms in the result text. Modules can be combined to produce a complex ranking strategy for a search interface.

S

schema for records

A set of metadata that describes the data model for your records. During the data modeling process, a data architect for the application powered by Oracle Endeca Information Discovery defines the schema for records. For example, the schema defines which of your attributes are searchable. It also defines display names for the attributes on your records, and other characteristics.

A schema for your data records is itself represented by records. However, unlike data records, which have attributes describing your data, schema records have attributes that describe the schema.

See also Property Description Record (PDR) and Dimension Description Record (DDR).

Property Description Record (PDR)

Dimension Description Record (DDR)

search interface

A named collection of attributes, each of which is enabled for record search. The search interface may include features that control the search behavior, such as relevance ranking modules and partial match. Search interfaces thus allow end users to search multiple attributes simultaneously.

set (EQL)

EQL represents multi-assign attributes (referred to in Studio as multi-value) from the corpus as sets. A set consists of a group of elements, typically derived from the values of a multi-assign attribute. EQL sets are intended to behave like mathematical sets: the order of the elements within a set is not specified (and, in general, not observable). An empty set is a set that contains no elements. All elements in a set must be of the same data type. If the elements in the set come from two multi-assign attributes (for example, by using the INTERSECTION row function), then those two multi-assign attributes must be of the same data type. Sets may not contain duplicate values and sets may not contain other sets.

See also multi-assign attribute.

single-assign attribute

An attribute for which each record can only have one value. For example, a book may have only one ISBN number, so the ISBN attribute would be single-assign.

attribute

primary key attribute

multi-assign attribute

unique attribute

standard attribute

managed attribute

snippeting

The snippeting feature provides the ability to return an excerpt from a record—called a snippet—to an application user who performs a record search query. A snippet contains the search terms that the user provided, along with a portion of the term's surrounding content to provide context. A Web application displays these snippets on the record list page of a query's results. With the added context, users can more quickly choose the individual records they are interested in.

source record

The data that is input into the data domain hosted in the Oracle Endeca Server, and available for viewing and navigation with Oracle Endeca Information Discovery. Source records are supported in a variety of formats.

data record

record

primordial record

standard attribute

Attributes whose values are not organized in an enumerated list or hierarchy. A standard attribute contains information about a record. An assignment from a standard attribute is known as a standard attribute value (or KVP, key-value pair). Each record is described by a set of attribute values.

A standard attribute differs from a managed attribute in that it does not have a hierarchy or an enumeration of attribute values attached to it. Standard attributes (along with managed attributes) are intended for display once the end user has searched or navigated to a record set or an individual record. Standard attributes are described by a type of schema record — Property Description Records (PDRs).

attribute

primary key attribute

single-assign attribute

multi-assign attribute

unique attribute

managed attribute

standard attribute value

An assignment from a non-hierarchical attribute on a record, used as a tag, or label, to classify a record in your data set. For values on standard attributes, use key-value pairs, or KVPs, where *key* is the name of an attribute, and *value* is an assigned value for this attribute.

Studio

A Web-based tool for building and using applications to search and analyze data from Oracle Endeca Server data domains.

Studio applications promote information discovery, allowing Studio users to uncover previously unknown relationships and trends in the data.

application (Studio)

Studio administrator

A Studio user who has the Administrator user role.

Studio administrators can view and edit any Studio application, and also have access to the Studio **Control Panel**, which contains options for configuring and monitoring Studio.

Studio component

A graphical tool displayed on a Studio application page to display and manipulate data from an Endeca Server data domain.

Studio components include functions to:

- Navigate to or search for specific data
- Display detailed information about data
- Display graphical representations of the data
- Manipulate and analyze the data
- Highlight specific values

system record

Used to control the behavior of the schema. Each attribute in a system record represents a configuration setting.

See also Property Description Record (PDR), Dimension Description Record (DDR), and Global Configuration Record (GCR).

Property Description Record (PDR)

Dimension Description Record (DDR)

Global Configuration Record (GCR)

T

thesaurus

The thesaurus feature of the Endeca Server allows it to return matches for related concepts to words or phrases contained in user queries. For example, a thesaurus entry may allow searches for Mark Twain to match text containing the phrase Samuel Clemens.

transactions

Similar to other databases, reading from or updating a data domain in the Oracle Endeca Server is performed via transactions. These transactions have the characteristics of atomicity, consistency, isolation, and durability.

Two types of transactions exist in the Oracle Endeca Server — simple transactions (also referred to as transactions), and an outer transaction.

A ***transaction*** is represented by any Web service request and response (or an administrative command) sent to and received from the Oracle Endeca Server. If the request completes successfully, a transaction is automatically committed. If the request fails, the transaction is rolled back.

In addition to issuing these simple transactions sent as requests to the Oracle Endeca Server, you can also nest them inside a single ***outer transaction***.

Once nested, the transactions are referred to as inner transactions. Nesting inner transactions inside an outer transaction is useful when the application must ensure atomicity, consistency, isolation, and durability of a group of Web service requests. Only one outer transaction per data domain can be running in the Oracle Endeca Server at a time. An outer transaction allows one level of nesting inner transactions inside it.

The Oracle Endeca Server provides a Transaction Web Service as the interface for controlling one or more inner transactions on a particular data domain, inside an outer transaction. You can start an outer transaction, and if all its inner transactions are processed successfully, it is committed to the data domain's index. If any of the operations inside an outer transaction fail, the outer transaction fails to commit and is rolled back. You use Integrator ETL to run graphs that utilize outer transactions.

U

unique attribute

unique attribute

An attribute for which the value must be unique for each record across the data set. This means that the PDR (Property Description Record) for the primary key attribute must have the `mdex-property_IsUnique` attribute set to `true`. For example, for a book, the ISBN number would be a unique attribute.

attribute

primary key attribute

single-assign attribute

multi-assign attribute

standard attribute

managed attribute

V

value query

Used during navigation to return the next level of values in an attribute hierarchy.

value search

A search that finds all of the attribute values that have names containing terms the user provides. The result of a value search is a set of attribute values, organized by attribute. The type-ahead feature in the Studio **Search Box** component returns value search results.

record search

view (Studio)

In Studio, a virtual set of records derived by filtering and grouping the physical records from one or more data sets.

For example, for a list of sales transactions, the transaction records could be used to derive a list of customers and a list of products. Those lists would be views.

In addition to these custom views, each data set has an associated base view, which is generated automatically and which consists of the entire contents of that data set.

Views are made up of attributes and predefined metrics. Some view attributes are also dimensions. Each view has its own metadata, which include its name, type, attribute display names, and the names and definitions of predefined metrics.

dimension (in views)

metric

base view (Studio)

view query (Studio)

view query (Studio)

A query issued against one or more views in Studio.

See also view.

view (Studio)

dimension (in views)

metric

base view (Studio)