Oracle® Endeca Information Discovery

Glossary

Version 2.3.0 • April 2012

ORACLE

Copyright and disclaimer

A

Administration Web Service

Used by IT engineers and administrators to integrate the Oracle Endeca Server and its reporting with third-party IT tools.

Oracle Endeca Server Web services

assignment

When a record has a value for an attribute, it is referred to as an assignment.

attribute

Basic unit of a record schema. Attributes describe records in 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 by the administrator or power user. Attributes are displayed in the context of their groups in Studio. 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 using Studio.

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 record is known as multi-assign.

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

B

Base view

A base view is a view of the entire data store.

It is created automatically from the physical records and always contains exactly the same records and attributes as the physical records.

See also view.

view

dimension (in views)

metric

view query

Bulk Load interface

An Oracle Endeca Server interface that is intended to achieve high performance for strictly additive ingests of varying amounts of data. An Integrator connector is available for this interface.

Data Ingest Web Service (DIWS)

C

cluster

A cluster is composed of a set of Oracle Endeca Server nodes each of which can serve query requests. Only one node is identified as the leader node; All other nodes are follower nodes. There is one copy of the Oracle Endeca Server index that is shared and used by all Oracle Endeca Server nodes. Nodes can be added or removed dynamically, without having to stop the cluster.

The Cluster Coordinator provides communication between the nodes in the cluster and notifies the nodes about index updates and updates to the configuration.

See also leader node and follower node.

leader node

follower node

component

See Studio standard component or Integrator component.

Studio standard component

Integrator component

configuration update

The process of loading changes made to the configuration of the Oracle Endeca Server index.

Configuration Web Service

Web service that allows you to update the schema and configuration.

See also Oracle Endeca Server Web services.

Oracle Endeca Server Web services

Conversation Web Service

Provides the primary means of querying data 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

Studio

D

Data Ingest Web Service (DIWS)

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

See also Oracle Endeca Server Web services.

Bulk Load interface

Oracle Endeca Server Web services

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 store, record, primordial record, system record, and source record.

record

primordial record

source record

data store

A discrete set of data, representing particular data records. Includes system records and indexed data records for the particular data store.

To create an Endeca data store, run the `create-ds` command of the Oracle Endeca Server. This command creates the data store, attaches it to the server, and starts the Dgraph process for that data store. You can later attach and detach the same data store, as well as start and stop it.

See also the Oracle Endeca Server.

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

metric

Base view

view query

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 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

Integrator component

end user (Studio)

End users are Studio content consumers. This includes executives seeking a dashboard view as well as others who need to drill through interactive visualizations and reports. Typically, Studio is configured so that end users cannot access the edit controls found on the edit view of each component.

See also Studio and power user.

Studio

power user (Studio)

Endeca Query Language (EQL)

EQL is a 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.

Entity Configuration Web Service

Web service that allows you to create and update views for a data source.

See also Endeca Web Services and view.

Oracle Endeca Server Web services

F

follower node

full index load

full index refresh

follower node

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

See also leader node and cluster.

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.

incremental index update

full index refresh

index

full index refresh

The process of reloading the full data set 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.

incremental index update

full index load

index

G

Global Configuration Record (GCR)

A single system record used to define global configuration information. 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, a graphical layout that contains a set of Integrator components.

Guided Navigation

The presentation of valid follow-on refinement choices to the user, implemented using the Guided Navigation and Breadcrumbs components in Studio.

See also Studio and component.

I

incremental index update

index

index configuration documents

Integrator

Integrator component

incremental index update

The process of loading changes made to the data records, which updates the data files (index) in the data store.

full index load

full index refresh

index

index

A collective term that refers to the many types of indices in 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. Other indices are also used.

incremental index update

full index load

full index refresh

index configuration documents

A set of XML-based configuration files that define how your records, standard attributes, and managed attributes are indexed by 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.

Integrator

Integrator 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 component

Integrator component

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

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

K

key-value pair

key-value pair

Assignments 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.

See also standard attribute and standard attribute value.

L

leader node

Liferay Portal

leader node

A single node in a cluster of Oracle Endeca Server nodes responsible for processing queries and for receiving updates to the index and to the configuration. This node is responsible for obtaining information about the latest index and propagating this information to the follower nodes through the Cluster Coordinator.

Each cluster must have one and only one leader node. All nodes must have write access to a shared file system on which the Oracle Endeca Server index is stored. The modules outside the cluster of Oracle Endeca Server nodes (such as connectors in Integrator and components in Studio) must have access to the leader node.

See also follower node and cluster.

cluster

follower node

Liferay Portal

Portal technology upon which Studio is built. Studio extends basic Liferay functionality to provide enhanced user management, security, and cross-component interaction, as well as performance-optimized communication with Oracle Endeca Servers.

For more detailed information, consult the Liferay documentation and forums, available at *http://www.liferay.com*. For information on the version of Liferay currently employed, see the *Oracle Endeca Information Discovery Installation Guide*.

Studio

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 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 managed in Studio.

metric

An expression defined for a view. The expression contains an arithmetic formula used for aggregation when querying against the view. Each metric must contain at least one aggregation function (for example, "SUM(x)" or "AVG(y)").

For example, a view that includes the attribute "Sales" may also contain a metric "Total Sales", which is the total of the values of the "Sales" attribute.

Views can have any number of metrics.

See also view.

view

dimension (in views)

Base view

view query

multi-assign attribute

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

attribute

primary key attribute

single-assign attribute

unique attribute

standard attribute

managed attribute

N

navigation query

navigation query

Used to return a list of matching records based on the current navigation or search.

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 data files of your records in memory, receives queries, executes them against the stored data files, and returns the results.

When used in the context of a process running inside installed software, Oracle Endeca Server is a Java-based utility for creating and administering data stores. Only one Oracle Endeca Server can be installed on a machine. Each Endeca data store has a Dgraph process that handles requests made to that store. For example, if you have used the Oracle Endeca Server to create three Endeca data stores and they have all been started, then three Dgraph processes are running on that machine.

See also Oracle Endeca Web services.

Oracle Endeca Server Web services

index

Oracle Endeca Server Web services

Include the Data Ingest Web service, the Conversation Web service, the Configuration Web service, the Entity Configuration Web service, the Administration Web service, the EQL Parser Web service, and the Control Web service.

Together, they provide an API interface to an Oracle Endeca Server implementation.

Data Ingest Web Service (DIWS)

Conversation Web Service

Configuration Web Service

Administration Web Service

Entity Configuration Web Service

P

physical records

Records stored in the Oracle Endeca Server index. 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 Oracle Endeca Server records, for the purposes of views, they are not included in the physical records.

power user (Studio)

Power users configure Studio content. One example would be a business analyst who configures Studio for end users and determines what components and data they can access.

See also Studio and end users (Studio).

Studio

end user (Studio)

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

An attribute used to uniquely identify a record. A primary key attribute must be both unique and single-assign. For example, for a book, the ISBN number could be a primary key attribute, because each book has only one unique value.

See also unique attribute and single-assign attribute.

attribute

single-assign attribute

multi-assign attribute

unique attribute

standard attribute

managed attribute

primordial record

The most basic infrastructure of an 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

project

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

edge

graph

Integrator

Integrator 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).

R

record

The fundamental unit of data in an 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, 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

refinement

An attribute whose values may be used to reduce or refine the current query's record set.

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.

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 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 customizable, component-based portal application, built on the Liferay Portal, that offers an interactive Guided Navigation user experience across a range of structured and unstructured enterprise data. Granular layout and configuration control enable users to manage and personalize their own experiences within the portal.

end user (Studio)

power user (Studio)

Studio standard component

A Studio portlet created and provided by Oracle, containing logic needed to retrieve and manipulate data from the Oracle Endeca Server and render results. Each Studio standard component provides specific Oracle Endeca Server features or other application support.

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 allows the system 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 store 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 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 store 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 store on the server. If any of the operations inside an outer transaction fail, the outer transaction fails to commit and you can roll it back manually. You use the Integrator 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. 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 a search box in Studio returns value search results.

record search

view

A virtual set of records derived from the physical records in a data source by filtering and grouping.

For example, for a data source consisting of 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.

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

dimension (in views)

metric

Base view

view query

view query

A query issued against one or more views.

See also view.

view

dimension (in views)

metric

Base view