Oracle® Endeca Server

Migration Guide

Version 7.5.1.1 • May 2013

ORACLE

Copyright and disclaimer

Table of Contents

Preface

Oracle® Endeca Server is 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.

About this guide

This guide helps you upgrade your Oracle Endeca Server implementation by describing the major changes between versions 7.4.x and 7.5.x.

Who should use this guide

This guide is intended for system administrators and developers who are upgrading Oracle Endeca Server on Windows or Linux.

Conventions used in this guide

The following conventions are used in this document.

Typographic conventions

This table describes the typographic conventions used when formatting text in this document.

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

Symbol conventions

This table describes the symbol conventions used in this document.

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

Path variable conventions

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

Path variable	Meaning
`$MW_HOME`	Indicates the absolute path to your Oracle Middleware home directory, which is the root directory for your WebLogic installation.
`$DOMAIN_HOME`	Indicates the absolute path to your WebLogic domain home directory. For example, if `endeca_server_domain` is the name of your WebLogic domain, then the `$DOMAIN_HOME` value would be the `$MW_HOME/user_projects/domains/endeca_server_domain` directory.
`$ENDECA_HOME`	Indicates the absolute path to your Oracle Endeca Server home directory, which is the root directory for your Endeca Server installation.

Contacting Oracle Customer Support

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

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

Chapter 1

Upgrading to Oracle Endeca Server Version 7.5.x

This section provides instructions for upgrading to Oracle Endeca Server 7.5.x. Before you start the upgrade processes, check the remaining sections of this guide to learn about the changes that will affect you during or after an upgrade.

Required reading

In addition to reading this document, it is recommended that you read the following documents for important information about the release.

Release Announcement

The Release Announcement outlines the new features that were added in Version 7.5.x.

Release Notes

The Release Notes provide information about known issues and bug fixes for this release.

Installation Guide

The *Oracle Endeca Server Installation Guide* contains installation instructions, and information on how to verify your installation.

Upgrading to Oracle Endeca Server 7.5.x.x

This topic describes how to upgrade to the 7.5.x.x version of Oracle Endeca Server.

To upgrade to Oracle Endeca Server 7.5.x.x from the Oracle Endeca Server 7.4.x:

1. Uninstall Endeca Server version 7.4.x, using instructions from the Installation Guide for the release you have installed.
   If required, first uninstall the Cluster Coordinator.
2. Install Oracle WebLogic Server 10.3.6 and Oracle Application Development Framework Runtime package 11.1.1.6.
   Use the instructions from the *Oracle Endeca Server Installation Guide* (version 7.5.x.).

3. Install the Oracle Endeca Server 7.5.x.
   Use the instructions from the *Oracle Endeca Server Installation Guide* (version 7.5.x.).

After you are finished, you should have a running Oracle Endeca Server. Optionally, you may have created an Endeca data domain (with no source data loaded) to verify that the installation was successful.

Note: If you are upgrading from the Endeca Server 7.5.x to Endeca Server 7.5.x.x, you can uninstall the Endeca Server 7.5.x and the Oracle WebLogic Server, and install the Endeca Server 7.5.x.x. In this case you will need to recreate the data domains after the upgrade. Alternatively, if you want to reuse an already created data domain in the upgraded Endeca Server, then, prior to upgrading from 7.5.x to 7.5.x.x, modify the `EndecaServer.properties` file to specify a new location of the offline directory so that it is outside of the WebLogic Server domain created for Endeca Server 7.5.x, export the data domain to that offline directory, uninstall the Endeca Server while removing its WebLogic domain, install the new version of the Endeca Server (7.5.x.x), create a data domain profile, and import the data domain into the newly installed Endeca Server.

About upgrading client stubs

In this release, each Web service packaged with the Oracle Endeca Server includes a version consisting of major and minor numbers, for example 2.0, where 2 is a major version number and 0 is a minor version number. Therefore, if you are planning to issue requests via stubs that utilize any of the Web services, you must generate new client stubs from the currently supported versions of each Web service.

Note: You only should generate new client stubs if you are planning to use direct requests to the Oracle Endeca Server utilizing any of its packaged Web services. Studio and Integrator use the supported versions of each Web service, therefore no special procedures apply in this case.

For information on changes that took place for each of the Web services, see the chapters in this guide.

For information on versions of Web services supported in this release, see the topic in the *Oracle Endeca Server Installation Guide*.

For information on how Web service versions are assigned and how to avoid version incompatibility, see the section on Web service versions in the *Oracle Endeca Server Developer's Guide*.

Chapter 2

Required Changes

This section describes required changes in the Oracle Endeca Server.

Terminology changes

The following table lists the concepts and terms that have changed in this Endeca Server release compared with the previous release.

New term	Description
Endeca Server Java application	In the previous release, the Endeca Server software included the Endeca Server Java application (as in this release). However, in this release, the Endeca Server Java application is installed and deployed in the Oracle WebLogic Server. This term is also known as the ***Endeca Server***, the ***Endeca Server instance***, or the ***Endeca Server node***.

New term	Description
Endeca Server domain	In the previous release, this term did not exist. In this release, before you deploy the Endeca Server application in the WebLogic Server, a dedicated WebLogic domain must be configured for it. This domain is known as the Endeca Server domain. (This process is described in detail in the *Oracle Endeca Server Installation Guide*.) This term is also known as the ***WebLogic domain for the Endeca Server application***.  **Important:** Do not confuse the term ***Endeca Server domain*** with the term ***Endeca data domain***. The Endeca Server domain refers to the domain created in the WebLogic Server. The term data domain, or Endeca data domain, refers to the logical collection of data and metadata managed by the Endeca Server application running in the WebLogic Server domain.

New term	Description
data domain, or ***Endeca data domain***	In the previous release, this term was known as the ***data store***. In this release, data domain is 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. If the data domain runs multiple Dgraph processes hosted by more than one Endeca Server instances, it is also referred to as the ***data domain cluster*** that runs in the Endeca Server cluster.
Endeca Server cluster	In the previous release, the Endeca Server cluster was represented by a set of Endeca Server instances each running a single Dgraph process. The cluster was managed at the Dgraph level. In this release, the Endeca Server cluster is a deployment of multiple Endeca Server instances that host and manage multiple clustered data domains. The Endeca Server cluster is managed at the Endeca Server instance level. For a complete description of a cluster, including a diagram, see the *Oracle Endeca Server Cluster Guide*.

Integration with the WebLogic Server

The Endeca Server is a J2EE application that runs in a J2EE container in the Oracle WebLogic Server. The Endeca Server uses WebLogic 11g (10.3.6), and Application Development Framework 11g (11.1.1.6).

The Endeca Server software exposes almost all of its APIs as SOAP web services. Starting with the Endeca Server 7.5.0, those Web services are implemented using the Oracle WebLogic Server 11g (WebLogic Server version 10.3.6), inside a hosted web application called the Endeca Server. For information on integrating Endeca Server with the WebLogic Server, see the *Oracle Endeca Server Installation Guide*.

Cannot run Endeca Server as a Windows service

In version 7.4.x, you could create a Windows service for running the Endeca Server in service mode. This functionality is not available in version 7.5.x, as Endeca Server runs as an application of WebLogic Server.

However, note that you can set up a WebLogic Server instance as a Windows service. For details, see the WebLogic document titled *Managing Server Startup and Shutdown for Oracle WebLogic Server*, which is available at: *http://docs.oracle.com/cd/E23943_01/web.1111/e13708/toc.htm*

Security-related changes

In this release, several security-related changes took place.

- **Changes to how SSL mode is used by the Endeca Server**. In this release, you can install the Endeca Server in a secure mode with mutual SSL authentication. In secure mode, Web service requests to the Endeca Server must use HTTPS. Additionally, your browser should be configured for SSL. For SSL-related changes to the `endeca-cmd` utility, see *Changes to endeca-cmd on page 8*.
- **Generating SSL keys**. If you installed Endeca Server in secure mode, you must generate the required SSL keys with the `generate_ssl_keys` utility.
- **Dgraph security**. The security of the Dgraph process has been improved by ensuring that the Endeca Server to the Dgraph communication is secure by default (if you installed the Endeca Server in a secure SSL mode).
- **Cluster security**. In this release, the Endeca Server cluster installation is secured by optionally configured SSL during the installation process.

  If the SSL mode is used, the Endeca Server to Dgraph communication uses SSL also (and cannot be turned off). The Endeca Server to the Cluster Coordinator communication (and the communication between the Cluster Coordinator services themselves) is secure by default (and cannot be turned off).

  For detailed information, see the *Oracle Endeca Server Security Guide* and the *Oracle Endeca Server Installation Guide* (in the section on installing the Endeca Server cluster).

For detailed information about security in the Endeca Server, see the *Oracle Endeca Server Security Guide*.

Changes to creating data domains and the treatment of the Dgraph processes

Due to the changes that occurred for the Endeca Server cluster, the way you create an Endeca data domain (known in the previous release as an "Endeca data store") has changed. Similarly, the administration of the Dgraph process has changed.

These statements summarize the changes:

- In the previous release, system administrators could directly manage a particular Dgraph, as there was a direct correlation between the Dgraph host and port and the Endeca Server host and port.

  In this release, two types of system administrators may be involved — first, those system administrators that administer the Endeca Server cluster, start one or more Endeca Server instances, and create data domain profiles that could be used to create data domains.

  Next, those administrators that manage data domains, create data domains, utilizing the Endeca Server nodes, and the data domain profiles that have been created for them.

  

  Note: Depending on the deployment scenario, the same system administrators can be responsible both for the provisioning of the Endeca Server instances and the creation and management of the data domains.

- In the previous release, the data store was created by running `endeca-cmd` after installing and starting the Endeca Server.

  To create a data domain in this release, after installing and starting the Endeca Server, use the `endeca-cmd` command to first create the data domain profile (or use the default data domain profile), and then use this profile to create a data domain.

- In the previous release, the data store's index (known as "data files"), could be configured to reside on a different directory than the installation of the Endeca Server, by using the `endeca-cmd` command.

  In this release, this setting is controlled by the `EndecaServer.properties` file. It is strongly recommended not to change this setting.

  Note that if the data domain is deployed in the Endeca Server cluster, the configuration of this file must be exactly the same on all Endeca Server instances supporting this data domain in a cluster. Additionally, the index must reside on the same directory on a shared file system that is accessible by all Endeca Server instances. If you have created more than one data domain, all of their index files will reside in the same directory.

  For information about `EndecaServer.properties`, see the *Oracle Endeca Server Administrator's Guide*.

For more information on the management of data domains in this release, see *Cluster changes on page 12* in this guide, and the *Oracle Endeca Server Cluster Guide*.

Changes to endeca-cmd

The `endeca-cmd` command-line script has been significantly enhanced in this release.

A number of command options in `endeca-cmd` have been removed and replaced with other options. The following table highlights the most important changes to the command's options:

Options in release 7.4.0 that are removed	Equivalent options added in release 7.5.1
`start-ds`	`enable-dd`
`stop-ds`	`disable-dd`
`create-ds`	`create-dd`
`attach-ds`	`import-dd`
`detach-ds`	`export-dd`

For detailed information about the commands for Endeca Server version 7.5.1, and other newly added commands in `endeca-cmd`, see the *Oracle Endeca Server Administrator's Guide* and the *Oracle Endeca Server Cluster Guide*.

In addition, in this release, two versions of the `endeca-cmd` script are available — one for SSL use and one for non-SSL use:

- The non-SSL version of the `endeca-cmd` script resides by default in the `endeca-cmd` directory in the root of the Endeca Server installation.

  For example, on Windows the default path is:

  `C:\Oracle\Middleware\EndecaServer7.5.1_1\endeca-cmd`

  The location on Linux will also be in the Endeca Server installation directory. Use this version of the script if your Endeca Server deployment is installed in a non-SSL environment.

- The SSL version of the `endeca-cmd` script resides by default in the `$DOMAIN_HOME/EndecaServer/bin` directory.

  For example, if `endeca_server_domain` is the name of your WebLogic Server domain, then the default path on Windows is:

  `C:\Oracle\Middleware\user_projects\domains\endeca_server_domain\EndecaServer\bin`

  This version lets you issue `endeca-cmd` commands without having to specify the location of the SSL certificates with the `--keystore` and `--truststore` options. It does so by making use of the `EndecaCmd.properties` file, which is located by default in the `$DOMAIN_HOME/config` directory.

  The parameters in the `EndecaCmd.properties` file are automatically set when you run the `generate_ssl_keys` utility to generate the certificates.

Required changes to the Endeca Server interfaces

This section describes required changes that apply to the Web services and other interfaces of the Endeca Server.

Changes to the Data Ingest Web Service

In version 7.5.1, the Data Ingest Web Service has been changed. New operations have been added and some existing operations have been deprecated. This topic discusses the changes.

Deprecated operations

The `ingestRecords` operation of the Data Ingest Web Service has been deprecated and is not guaranteed to be supported in the future releases. Use the `ingestChanges` operation instead, for updating and replacing records and assignments on their attributes.

New operations

The Data Ingest Web Service contains one new operation, `ingestChanges`.

Inside `ingestChanges`, the following elements represent new sub-operations:

- `addRecords`
- `updateRecords`
- `addOrUpdateRecords`
- `deleteRecords`
- `replaceRecords`

Further, inside the `updateRecods` element, the following elements are included, each of which also represents an operation of this Web service:

- `addAssignments`
- `deleteAssignments`
- `wildcardDeleteAssignments`
- `replaceAssignments`

For information and examples of usage of the `ingestChanges` operation, see the *Oracle Endeca Server Data Loading Guide*.

Changed operations

The `IngestDimensionValues` operation has been renamed to `IngestManagedAttributeValues`. The syntax of this operation remains the same, however, some of the nested elements were also renamed to

reflect the fact that operations occur on values for managed attributes (known in previous releases as dimensions).

Additional changes

It is now possible to change a name of the standard attribute in a running data domain.

The operations of the Data Ingest Web Service take EQL (Endeca Query Language) expressions as arguments, which allows you to identify one or more records with the EQL filters. For information and examples, see the *Oracle Endeca Server Data Loading Guide*.

Addition of two new Web services

The Endeca Server has two new Web services — Cluster Web Service for configuring and administering the Endeca Server cluster, and Manage Web Service for managing data domains. The `endeca-cmd` utility relies on both of these interfaces.

For information on the Cluster and Manage Web Services, see the *Oracle Endeca Server Cluster Guide*.

For information on Web service versions, see the *Oracle Endeca Server Developer's Guide*.

Removal of two Web services from public access

The Administration Web Service and the Control Web Service of the Endeca Server have become private interfaces in this release. These interfaces are not intended to be used by Endeca Server clients and are not accessible at the Endeca Server host and port.

The operations from the Control Web Service are available through the `endeca-cmd` utility, which lets you manage data domains, or through the Manage Web Service that, together with the Cluster Web Service, is backing up this utility.

The following table describes how the operations from the Administration Web Service are replaced (or no longer supported) in this release:

Administration Web Service Operation in Endeca Server 7.4.0	**Treatment in Endeca Server 7.5.1**
`createSnapshotOperation`	Use the `endeca-cmd export-dd <data-domain> [--offline-name <exported-domain>]` command, or the `exportDataDomain` operation of the Manage Web Service. These options let you export the index of the specified data domain by taking a snapshot of the index and copying it into an internally stored file. A snapshot represents a copy of the index files only, and does not capture any other characteristics of the data domain, such as the Dgraph configuration or the data domain profile used when it was created.  **Important:** On Linux, you can export a data domain that is either enabled or disabled. On Windows, only a disabled data domain can be exported. When an enabled data domain is exported, it continues to run, but a copy of its index is stored in an offline directory on a shared file system.
`getVersionOperation`	Use `endeca-cmd version`, or the `version` operation of the Manage Web Service, to list the version of the Oracle Endeca Server and the Dgraph processes for each of the enabled data domains hosted in the Endeca Server.
`listJobsOperation` and `cancelJobOperation`	These operations for managing jobs in the Dgraph process are no longer supported, because in this release, the Endeca Server Java application manages the Dgraph process and its underlying jobs. Monitoring the Dgraph jobs directly is no longer a required task for system administrators.

For information on `endeca-cmd`, see the *Oracle Endeca Server Administrator's Guide*.

For information on managing data domains and obtaining a version, see the *Oracle Endeca Server Cluster Guide*.

Cluster changes

In this release, the implementation of the Endeca Server cluster has changed significantly.

The following statements provide a brief listing of the changes. For detailed information, see the *Oracle Endeca Server Cluster Guide* and the *Oracle Endeca Server Installation Guide*.

- **Integration with the WebLogic Server**. In this release, Endeca Server is a J2EE application that runs in the WebLogic Server domain container. WebLogic Server 10.3.6 is used. The Endeca Server utilizes the WebLogic Server native Application Development Framework, specifically JRF, for the Endeca Server interfaces.
- **Cluster use cases**. In the previous release, system administrators could create a single data store on a set of Endeca Server nodes, with a one-to-one mapping between Endeca Server host machines to the Dgraph nodes (one Dgraph process per each Endeca Server running on a single hardware machine). In this release, two installation and deployment scenarios are possible:
  - Installation and deployment of a single Endeca Server instance on a WebLogic Server that is started as a single Admin Server in the WebLogic Server domain created specifically for the Endeca Server Java application.

    In this Endeca Server, one or more data domains can be created. Each of the data domains can run one or more Dgraph processes (where a specific Dgraph can serve only one data domain).
  - Installation and deployment of a multiple-node Endeca Server cluster on a set of physical machines that run in a WebLogic Server domain created for the Endeca Server.

    In the WebLogic domain, a single Admin server controls a set of Managed servers. Managed servers host the Endeca Server instances comprising an Endeca Server cluster. (The Admin Server can host an Endeca Server in a single node deployment, but not in a multiple-node Endeca Server cluster deployment).

    In this multi-node Endeca Server cluster deployment, multiple data domain clusters can be created in the Endeca Server cluster. The system administrators can start by first creating a data domain in a single Endeca Server instance in the development environment, and then redeploying this data domain in an Endeca Server cluster.

    For information on the cluster installation and deployment, see the *Oracle Endeca Server Installation Guide*.
- **Cluster Coordinator changes**. In this release, the Cluster Coordinator is part of the Endeca Server installation and is always installed when you install the Endeca Server. It should no longer be installed separately as in previous release.

  Additionally, instead of starting a new Cluster Coordinator for each data domain cluster (as in the previous release), in this release an ensemble of Cluster Coordinator services runs on a subset of Endeca Servers running in an Endeca Server cluster.
- **Improved clustering capabilities and increased availability**. An ensemble of Cluster Coordinator services tracks availability of Dgraph nodes for multiple data domains hosted in the Endeca Server cluster. Dgraphs are resilient to failure of Cluster Coordinator service (that is, Cluster Coordinator is in itself highly available, if configured in a manner where at least three or a greater than three odd number of Cluster Coordinator services are running on Endeca Server instances.)
- **Leader election in a data domain cluster**. Instead of letting the system administrator determine the data domain's leader node (Dgraph leader node), in this release, the Endeca Server cluster determines which of the Endeca Server nodes will host the Dgraph that will be the leader node for the data domain.

Similarly, instead of manually identifying and starting the follower Dgraph nodes (as was required in the previous release), in this release, the Endeca Server cluster automatically starts the Dgraph nodes, and performs their allocation to Endeca Server nodes that are part of the Endeca Server cluster.

- **Failure recovery.** In the previous release, recovery from a failed Dgraph leader process necessitated additional actions by the system administrator.

  In this release, leader election in the data domain cluster is automatic.

- **Allocation of Dgraphs to Endeca Server instances**. In the previous release, it was difficult to use a hardware cluster to serve multiple data domains because the system administrator needed to manage mapping of leader/follower Dgraphs to the hardware machines manually, by specifying their host names and ports.

  In this release, the allocation of data domain nodes to the Endeca Server nodes is automatic. Moreover, the system administrator managing the Endeca Server cluster can create data domain profiles with different hardware usage to address different deployment scenarios, such as for testing and production data domains.

- **Multiple data domains are supported**. In this release, system administrators can create an Endeca Server cluster to support provisioning of a large number of data domain clusters, while making efficient use of hardware. Some data domains can be hosted on the same multiple Endeca Server nodes to support higher throughput and increased availability of the Endeca Server services for query processing.

  Data domain system administrators can manage a large number of data domains for different tenants, in different sizes and availability constraints, on a hardware cluster that is hosting Endeca Server instances.

- **Cluster security**. In this release, the Endeca Server cluster installation is secured by optionally configured SSL during the installation process. If the SSL mode is used, the Endeca Server to Dgraph communication uses SSL also (and cannot be turned off). The Endeca Server to the Cluster Coordinator communication (and the communication between the Cluster Coordinator services themselves) is secure by default (and cannot be turned off).

- **Cluster interfaces**. To support the Endeca Server cluster implementation, two new interfaces have been added. A Cluster Web Service lets system administrators create Endeca Server clusters on existing hardware. A Manage Web Service lets the system administrators create, clone and delete data domains based on previously predefined data domain profiles, across the entire Endeca Server cluster. It also handles starting data domain nodes (Dgraphs) on appropriate Endeca Server nodes.

  Additionally, the `endeca-cmd` utility has been enhanced in this release to allow accessing a data domain at any Endeca Server node in the Endeca Server cluster. The `endeca-cmd` utility relies on the Cluster and Manage Web Services and lets system administrators create Endeca Server cluster node profiles, list Endeca Server nodes, create data domain profiles, create and manage data domains, and check the health status of the Endeca Server cluster nodes, data domains and data domain nodes. The system administrators in charge of individual data domains can also export their index and re-import it into a new data domain cluster.

- **Routing capabilities in the Endeca Server cluster**. Any node in the Endeca Server cluster can answer queries for any data domain, even if the Dgraph node for that data domain is not running on the Endeca Server node that initially received the query. Endeca Server can route queries to an appropriate node for a given data domain.

  You can also configure the cluster to use session affinity, and use an external load balancer in front of the Endeca Server cluster to provide additional load balancing capabilities.

  For more information, see the *Oracle Endeca Server Cluster Guide*.

Replacement of record and range filters with EQL record filters

In this release, the record and range filters of the Endeca Server have been removed and replaced with the EQL record filters.

EQL record filters let you define arbitrary subsets of the total record set, and dynamically restrict search and navigation results to these subsets. Additionally, EQL filters allow a user, at request time, to specify an arbitrary, dynamic range of values that are then used to limit the records returned for a navigation query.

The Conversation Web Service has two filtering components that allow you to use the Endeca Query Language (EQL) to provide filters for your query using EQL syntax:

- The `DataSourceFilterString` component makes the universe of data that is visible to your query smaller and can be used as a security filter to prevent users from obtaining records that they are not authorized to view.
- The `SelectionFilterString` component is used for additional application-level filtering and specifies the criteria for the final record result set. The returned results are the records that match all of the filters specified in the query. It also determines which data is available for refinement computation.

Either or both of the EQL record filters can be used with a record search or a value search. For detailed information on EQL record filters, see the *Oracle Endeca Server Developer's Guide*.

Endeca Query Language changes

This section describes required and behavioral changes to the Endeca Query Language (EQL).

The following changes in EQL occurred in this release:

- Addition of grouping sets, including `CUBE`, `ROLLUP`, and their helping functions (`GROUPING`, `GROUPING_ID`, `GROUP_ID`).
- `ORDER BY` now can take an arbitrary expression.
- `TO_DATETIME` now supports a numeric interface, in addition to the previous string interface.
- The `CONCAT`, `STRING_JOIN`, `TO_STRING`, and `SUBSTR` functions were added.
- `GROUP BY ManagedAttr:level` is no longer supported. The `ANCESTOR` function can be used as an alternative.
- Optional depth to `GROUP BY` elements is no longer allowed.
- Duplicate `GROUP BY` restriction is removed.
- `GROUP` and `GROUP BY` now respect NULL values and do not discard records with NULL values.
- Aliasing is not longer required for attributes.
- `SELECT *` from intermediate results is added.
- `IS_ANCESTOR` and `IS_DESCENDENT` functions (which previously worked only with managed attributes) now also work with standard attributes.
- Parsing and validation messages from EQL processing have been localized to provide international language support.

- Using single quotes with string values. When using string value comparison operators, make sure that you use single quotes around the text value field. When using numeric value comparison operators, do not use quotes of any kind around the value field.
- Escaping special XML characters. If you are making direct queries against the Conversation Web Service (for example, by using the soapUI tool), you may need to escape some XML characters to prevent parsing errors. For example, you should use the `<` escape character instead of the < (less than) character.

Note: This list only announces the changes but does not describe them in detail. For detailed information on EQL, see the *Oracle Endeca Server EQL Guide*.

Changes to updating spelling dictionaries

The administrative HTTP operation `/admin/datastore?op=updateaspell`, which rebuilds the spelling dictionaries for spelling correction from the data corpus, was removed and replaced with an equivalent `updateSpellingDictionaries` operation in the Manage Web Service.

For more information on `updateSpellingDictionaries`, see the *Oracle Endeca Server Cluster Guide*.

Removed Dgraph flags

This topic lists the flags to the Dgraphs that have been removed in this release.

The following flags have been removed, either because the functionality they provided was removed several releases ago, or because their functionality is implemented in this release through other mechanisms in the Endeca Server. The removed flags fall into three categories, as described below:

- **Flags removed due to changes to the Endeca Server cluster**. The following flags have been removed: `--cmem`, `--threads`, `--coordinator_host`, `--coordinator_port`, `--follower`, and `--read-only`. These statements describe the changes:

Dgraph flags in Endeca Server 7.4.0 that are removed	Treatment in Endeca Server 7.5.1
`--threads`,`--cmem`	Cache size and the number of threads for the Dgraph nodes can now be specified for the data domain profile in the Endeca Server.
`--coordinator_host`,`--coordinator_port`	Cluster Coordinator host and port are now identified during the Endeca Server cluster deployment.
`--follower`	The follower and leader Dgraph nodes (and their host and port) in the data domain cluster are now appointed by the Endeca Server and do not require configuration by the system administrator.
`--read-only`	In this release, you can configure a data domain that will serve as a read-only data domain.

For detailed information, see the *Oracle Endeca Server Cluster Guide*.

- **Flags removed due to changes to thesaurus and international language support**.The following flags have been removed:

Dgraph flags in Endeca Server 7.4.0 that are removed	Treatment in Endeca Server 7.5.1
`--thesaurus_cutoff`, `--thesaurus_multiword_nostem`	The thesaurus cutoff is hard-coded at 3.
`--disable_fast_aspell--latin1`, `--whymatch`, `--whymatch_concise`, `--wordinterp`	These flags are no longer used by the Endeca Server. For detailed information, see the *Oracle Endeca Server Developer's Guide*

- **Other removed flags**. The `xquery-fndoc` and `xquery-path` flags are removed because XQuery is an internal component.

Chapter 3

Behavioral Changes

This section describes behavioral changes to Oracle Endeca Server and its interfaces.

Endeca Server configuration files

In this release, the Endeca Server uses two configuration files.

Server configuration file

The `EndecaServer.properties` file sets global-wide parameters for Endeca Server, such as the default locations of mandatory files and directories. The file is located in the `$DOMAIN_HOME/config` directory. This file is analogous to the Endeca Server version 7.4.x configuration script (`endeca-server.windows.conf.bat` on Windows or `endecaserver. linux.conf` on Linux).

The `EndecaServer.properties` file is automatically created when you create an Endeca Server domain in the WebLogic Server. For the most part, you will not be modifying this file. The file parameters are explained in the *Oracle Endeca Server Administrator's Guide*.

Note: In the Endeca Server cluster, the configuration of this file for a specific data domain must be identical on all Endeca Server nodes that support this data domain.

endeca-cmd configuration file

The `EndecaCmd.properties` file stores locations of the SSL certificates that are needed for use with the SSL version of the `endeca-cmd` utility. The file is located in the `$DOMAIN_HOME/config` directory.

The file is automatically updated with the correct SSL file paths when you run the `generate_ssl_keys` utility to generate the certificates. This means that you should not have to modify the file. In the Endeca Server cluster, the contents of this file must also be identical on all Endeca Server instances, for the specific data domain.

Note that there is no analogous file in the 7.4.x release.

PDR and data type changes

Changes to PDR properties and the `mdex:time` data type may affect the ingest of your attribute schema and source data.

PDR properties

The following two changes affect the creation of a PDR (Property Description Record) for standard attributes:

- The default value for the `mdex-property_IsSingleAssign` property is now `true`. In version 7.4.x, the default was `false`.
- The new `mdex-property_Language` property sets the language for the standard attribute. For details, see *Language property for PDRs on page 19*.

mdex:time time zone requirement

The `mdex:time` data type now requires a time zone. In version 7.4.x, a time zone was not allowed as part of the time representation. In 7.5.x, a time zone is required for both time formats: either a universal (UTC) time or as a local time plus a UTC time zone offset. For details on the `mdex:time` formats, see the *Oracle Endeca Server Data Loading Guide*.

Renaming a standard attribute

The Data Ingest Web Service lets you rename a standard attribute on records in a running data domain.

You use the `replaceAssignments` element in an `updateRecords`, inside the `ingestChanges` request to rename a standard attribute.

For more information, see the *Oracle Endeca Server Data Loading Guide*.

Language support and configuration

You can now configure a specific language for standard attributes and for queries.

Additional supported languages

Besides English, Endeca Server now supports record and value searches in 21 other languages, such as German, French, Chinese, Japanese, and Russian. For a list of supported language codes, see the *Oracle Endeca Server Developer's Guide*.

The `SearchFilter` and `ValueSearchConfig` types have a `Language` attribute to specify the language for the search. This per-query language code enables the Dgraph to select the appropriate dictionary for a given query.

Language property for PDRs

Each PDR (Property Description Record) now has an `mdex-property_Language` field that specifies the language of that standard attribute. The value can either be explicitly set by the user during ingest time or, if not explicitly set, defaults to the global PDR language code for the system. The global PDR language code can be set by the `setPropertyDefaultLanguage` operation described in the next section.

It is recommended that during ingest time, you explicitly set the `mdex-property_Language` field for each of your standard attributes.

Setting the global PDR language code in the Configuration Web Service

The Configuration Web Service has been updated to include two new operations: `setPropertyDefaultLanguage` and `getPropertyDefaultLanguage`.

The `setPropertyDefaultLanguage` operation sets the default language for new standard attributes (PDRs) that are created automatically by the Data Ingest Web Service or the Bulk Load Interface. The default language is also used if the `mdex-property_Language` property is not explicitly set during the creation of a PDR by the Data Ingest Web Service or the Bulk Load Interface. (Note that PDRs created by the Configuration Web Service's `putProperties` and import operations must be fully and explicitly specified.)

The `getPropertyDefaultLanguage` operation returns the default language code that is used for PDRs. The language ID will be either `unknown` (the default) or the language ID that was set by a previous `setPropertyDefaultLanguage` operation.

Diacritic folding behavior

In version 7.4.x, diacritic folding was turned off by default. However, you could use the Dgraph `--latin1` flag to enable diacritic folding.

In version 7.5.x, diacritic folding is the default behavior for all supported languages (including "unknown") during record searches. In addition, the Dgraph `--latin1` flag has been removed. Note that you cannot disable this diacritic folding behavior.

Thesaurus and stop word support

The thesaurus feature is supported in version 7.5.x, similar to version 7.4.x. Note, however, that only one global thesaurus is supported for an Endeca data domain. In other words, language-specific thesauruses are not supported (such as one thesaurus for English, a second for French, and so on).

Stop words are supported only for searches that are marked with the "unknown" language identifier.

See the *Oracle Endeca Server Developer's Guide* for more information on the thesaurus and stop words features.

Addition of the SH data set index

A Sales History data set is supplied as a sample data set for an Endeca data domain.

The SH data set is ready for use out-of-the-box because the files have already been indexed by the Endeca Server. This eliminates the need to load source records into the Endeca Server.

Before you can use the Endeca data domain configured for the SH data set, place the SH indexed files into the offline directory of the Endeca Server, and import them into a newly-created Endeca data domain. For instructions on unpacking the SH index, see the *Oracle Endeca Server Installation Guide*.

Refinement configuration for individual attributes

In the previous release, the Conversation Web Service allowed you to expose refinement values for groups of attributes only. In this release, `RefinementConfig` from the Conversation Web Service can be used on individual attributes, to expose their refinement values.

In other words, in the Conversation Web Service request, you can now use both the `RefinementGroupConfig` and `RefinementConfig` elements in `NavigationMenuConfig` to expose refinement values for attributes. These elements allow you to perform operations on many attributes at once without enumerating all of them.

For more information, see the *Oracle Endeca Server Developer's Guide*.

Performance improvements

This topic provides a summary of Endeca Server 7.5.1 performance improvements.

Performance of the Endeca Server has been improved as follows:

- Faster query evaluation with query-specific optimizations.
- Increased cache usage during query processing with partial result caching.

  Similar to previous releases, all Endeca Server features continue to utilize some form of caching. However, in this release, most of the Endeca Server features benefit from the new type of caching, which utilizes caching of intermediate results during parallel query processing.

  As a result, the Endeca Server utilizes cache more heavily than it did in the previous releases. If the default Dgraph cache size is insufficient for your data domain, consider tuning it. For information on tuning the Dgraph cache, see the *Oracle Endeca Server Administrator's Guide*.
- Reduced peak memory usage during data loading via the Bulk Load interface.

Changes to identifying host and port for the Bulk Load Interface

In the Endeca Server cluster, the way to identify the host and port of the Endeca Server instance for loading data using Bulk Load Interface has changed.

In the previous release, to use the Bulk Load Interface, you had to specify the host and port of the machine on which the Endeca data store was running to the `BulkIngester` constructor of the Bulk Load Interface. If the

data store was running in the Endeca Server cluster, this machine was the one on which the leader Dgraph node was running. Because the system administrator was responsible for configuring the leader node, this was the node whose host and port had to be configured for the Bulk Ingest Interface.

In this release, the initial allocation (and the subsequent appointment) of the leader node occur automatically in the Endeca Server cluster, for each data domain hosted in it. Therefore, before you can specify the host and port of the machine for loading data into the data domain with the Bulk Load Interface, you need to request this information from the Endeca Server cluster. Use the `allocateBulkLoadPort` operation of the Manage Web Service to obtain the host and port of the Endeca Server instance that is allowed to perform data loading operations for the data domain.

Changes in XML representation of records

The format of records that are returned by the Endeca Server in response to the Data Ingest or Conversation Web Service queries has changed.

Records that looked like this:

```
<mdex:record>
  <specKey>specVal</specKey>
  <otherAttribute>value</otherAttribute>
</mdex:record>
```

In this release look like this:

```
<mdex:record>
  <mdex:attribute name="specKey">specVal</mdex:attribute>
  <mdex:attribute name="otherAttribute">value</mdex:attribute>
  <mdex:attribute name="otherAttribute">second value</mdex:attribute>
</mdex:record>
```

The responses from the Data Ingest Web Service and from the Conversation Web Service queries contain records in this format.

In other words, in previous releases, for each Endeca Server attribute on a record, an XML element with that name in the XML representation was created and returned from the Endeca Server in the response. In this release, all the XML elements have the same name (such as `<cs:attribute/>`), but the name of each record's Endeca Server attribute is stored in the XML element's `name` attribute.

Important: Do not confuse Endeca Server attributes with the term "attribute" in XML (which is pertinent to the change described in this topic). The columns on records are known as "attributes" in the Endeca Server.

Changes to how the Dgraph stats page is returned

In the data domain, more than one Dgraph process may be running. You can request the Dgraph stats page either on each Endeca Server node, or for all Dgraph instances at once.

The Dgraph stats page continues to be returned with `http://<host>:<port>/endeca-server/admin/<data_domain>?op=stats` for each machine hosting the Endeca Server for this data domain.

In addition, `endeca-cmd get-dd-status <data-domain>` returns the Dgraph node statistics for each running Dgraph node in the data domain (and not only for one Dgraph node). Note, however, that this information is not formatted in HTML and thus is intended for Oracle Endeca Support only.

Changes to the Endeca Server documentation

The *Oracle Endeca Server Query Language Reference* has been renamed to the *Oracle Endeca Server EQL Guide*. These new guides have been added: the *Oracle Endeca Server Cluster Guide* and the *Oracle Endeca Server Security Guide*.