

# Oracle® Fusion Middleware

## Integrating Enterprise Data Quality with External Systems



12c (12.2.1.4.0)

E95652-01

September 2019

The Oracle logo, consisting of a solid red square with the word "ORACLE" in white, uppercase, sans-serif font centered within it.

ORACLE®

Oracle Fusion Middleware Integrating Enterprise Data Quality with External Systems, 12c (12.2.1.4.0)

E95652-01

Copyright © 2018, 2019, Oracle and/or its affiliates. All rights reserved.

Primary Author: Oracle Corporation

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, then 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.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information about 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 unless otherwise set forth in an applicable agreement between you and Oracle. 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, except as set forth in an applicable agreement between you and Oracle.

# Contents

## Preface

---

Audience	vii
Documentation Accessibility	vii
Related Documents	vii
Conventions	viii

## 1 Integrating with Subversion

---

Software Requirements	1-1
Understanding the Integration Architecture	1-2
Setting Up a Repository	1-4
Configuring EDQ with Subversion	1-4
Configuring a New EDQ Installation	1-5
Retaining Existing Configuration Information	1-5
Understanding the Integration Elements	1-6
Reviewing a Deployment Example	1-7
Troubleshooting Errors	1-9

## 2 Integrating with IBM Global Name Recognition

---

System Requirements	2-1
Configuring the EDQ Server	2-1
Building the Search Library	2-2
Configuring the GNR Connector	2-3
Creating the EDQ GNR Properties File	2-3
Creating the Search Configuration Files	2-4
Support for GNR 3.2 and GNR 4.2 in Search Configuration Files	2-4

## 3 Integrating with Experian QAS

---

Software Requirements	3-1
Integrating with Experian QAS	3-1

Migrating QAS integrations	3-2
----------------------------	-----

## 4 Integrating with Capscan Matchcode

---

Software Requirements	4-1
Integrating the Capscan Matchcode Libraries into EDQ	4-1
Customizing the Matchcode API	4-2

## 5 Using the Command Line Interface

---

Running the Command Line Interface	5-1
Understanding the Commands and Arguments	5-1
runjob	5-1
runopsjob	5-2
droporphans	5-3
listorphans	5-3
scriptorphans	5-3
list	5-3
showlogs	5-3
shutdown	5-4
version	5-4
Reviewing Examples	5-4
Listing All the Available Commands	5-4
Listing the Available Parameters for a Command	5-5
Running a Named Job	5-5
Running a Named Job in Operations Mode	5-5

## 6 Configuring Additional Database Connections

---

Using JNDI to Connect to Data Stores	6-1
Connecting to an Oracle Database Using tnsnames.ora	6-1
To Configure EDQ to Connect Through TNS	6-1
Connecting to an Oracle Database Using Oracle Internet Directory (LDAP)	6-2

## 7 Configuring EDQ to Process XML Data Files

---

Using Simple XML Data Stores	7-1
Reading Simple XML Files	7-1
Writing Simple XML Files	7-2
Using XML and Stylesheet Data Stores	7-2
Using DN-XML	7-3
Reading Custom XML Files	7-4

Configuring the Data Store	7-6
Writing Custom XML Files	7-6
Configuring the Data Store	7-7

## 8 Using the EDQ Configuration API

---

REST Interface for Projects	8-1
Retrieving a List of EDQ Projects	8-1
Creating a Project	8-2
Deleting a Project	8-2
REST Interface for Data Stores	8-3
Retrieving a List of Data Stores	8-3
Creating a Data Store	8-4
Deleting a data store	8-6
REST Interface for Snapshots	8-6
Retrieving a List of Snapshots	8-7
Creating a Snapshot	8-7
Deleting a Snapshot	8-8
REST Interface for Processes	8-8
Retrieving a List of Processes	8-9
Deleting a Process	8-9
Creating a Simple Process	8-10
REST Interfaces for Jobs	8-11
Retrieving a List of Jobs	8-11
Deleting a Job	8-12
Creating a Simple Job	8-12
Running a Job	8-12
Cancelling a Running Job	8-13
Getting the Status of a Job	8-14
Getting the Details of All Running Jobs	8-14
REST Interface for Reference Data	8-15
Retrieving a List of Reference Data	8-15
Retrieving Contents of Reference Data	8-16
Creating Reference Data	8-17
Deleting Reference Data	8-17
REST Interface for Web Services	8-18
Retrieving a List of Web Services	8-18
Creating or Updating a Web Service	8-19
Deleting a Web Service	8-20
Example: Profiling from an External Application	8-20

## 9 Using the Java Messaging Service (JMS) with EDQ

---

Understanding the Message Queue Architecture	9-1
Uses of JMS with EDQ	9-1
Configuring EDQ to Read and Write JMS Messages	9-2
Defining the Interface Files	9-2
Knowing the <attributes> section	9-2
Knowing the <messengerconfig> Section	9-3
Knowing the <messagebody> section	9-6
Illustrations	9-6

# Preface

Describes how to integrate Enterprise Data Quality with external systems and applications.

## Audience

This document is intended for advanced users of EDQ and administrators responsible for integrating EDQ with third-party applications.

## Documentation Accessibility

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

### Accessible Access to Oracle Support

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

## Related Documents

For more information, see the following documents in the documentation set.

### EDQ Documentation Library

The following publications are provided to help you install and use EDQ:

- 
- 
- 
- 
- 
- 
- 
- 

Find the latest version of these guides and all of the Oracle product documentation at <https://docs.oracle.com>

## Online Help

Online help is provided for all user applications. It is accessed in each application by pressing the **F1** key or by clicking the Help icons. The main nodes in the Director project browser have integrated links to help pages. To access them, either select a node and then press **F1**, or right-click on an object in the Project Browser and then select **Help**. The EDQ processors in the Director Tool Palette have integrated help topics, as well. To access them, right-click on a processor on the canvas and then select **Processor Help**, or left-click on a processor on the canvas or tool palette and then press **F1**.

## Conventions

The following text conventions are used in this document:

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



# 1

## Integrating with Subversion

This chapter describes how to integrate and use EDQ with the Subversion version control system.

The following sections are included:

- [Software Requirements](#)
- [Understanding the Integration Architecture](#)
- [Setting Up a Repository](#)
- [Configuring EDQ with Subversion](#)
- [Understanding the Integration Elements](#)
- [Reviewing a Deployment Example](#)
- [Troubleshooting Errors](#)

### Software Requirements

EDQ supports integration with Subversion 1.6, 1.7, 1.8 and 1.9. For more information about Subversion, see the Apache Subversion website found at <http://subversion.apache.org/>.

 **Note:**

EDQ currently only supports integration with Subversion 1.6, 1.7, 1.8 and 1.9. Attempting to integrate with any other versions will cause an error.

The Subversion server with which EDQ is being integrated must meet these prerequisites:

- Support Hypertext Transfer Protocol (HTTP) and Distributed Authoring and Versioning (DAV) access.
- Require authentication on commit.
- Not require authentication on checkout or update.

When Subversion is integrated with EDQ as a store of configuration information, the following restrictions and limitations apply. Consider the following points before deciding to configure integrated version control using Subversion.

- You cannot update or revert an item that is open in the Director interface or the Subversion server.
- You cannot rename a project once the project is under version control. This is critical in avoiding duplication of reference processor names in a project.
- Deleting a project does not remove it from the Subversion repository.

- Case insensitive name matching is used.

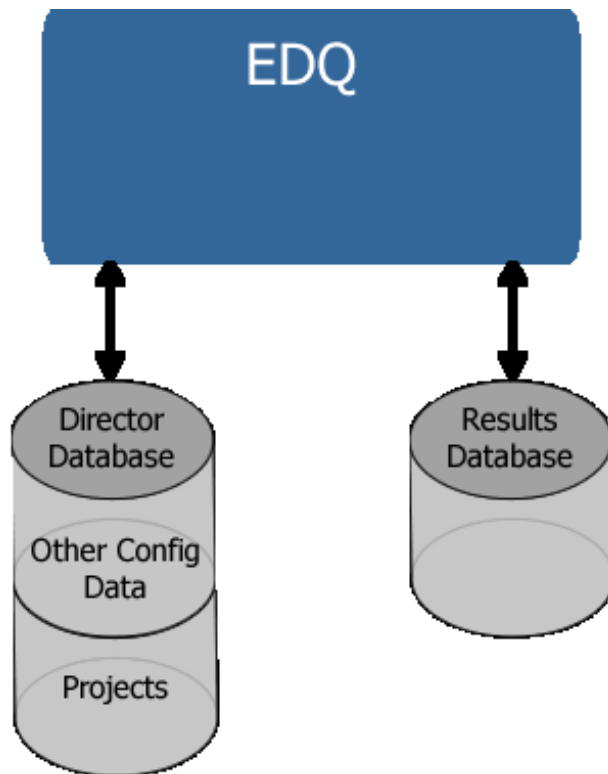
## Understanding the Integration Architecture

The EDQ server can be configured to be aware of a Subversion server as a store of configuration information.

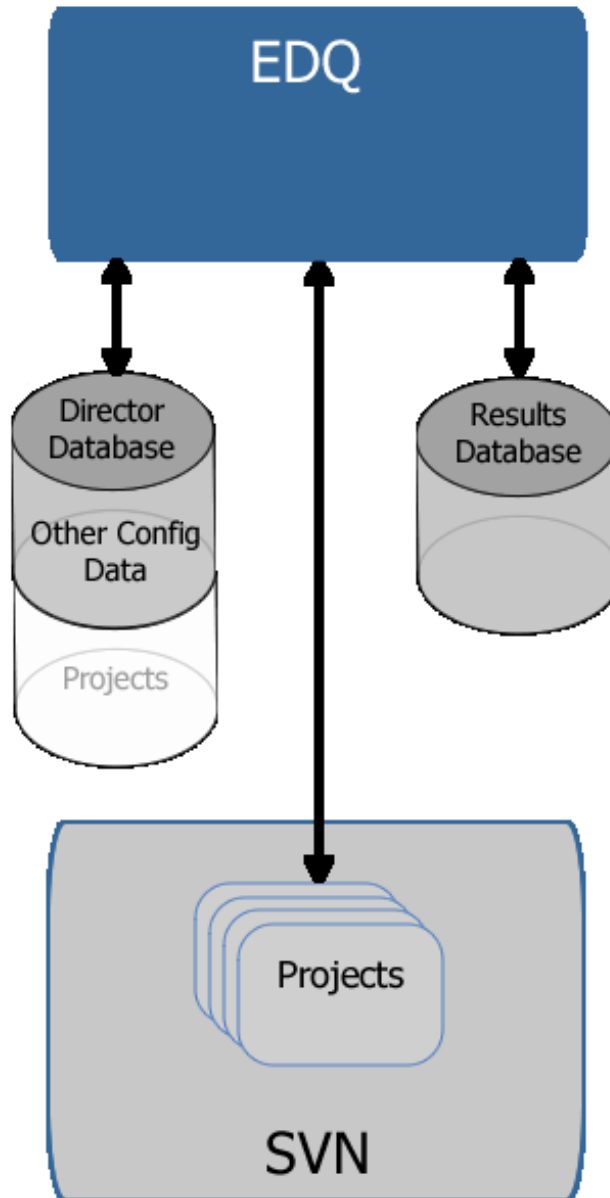
 **Note:**

In this instance, configuration information means information that is managed using the Director UI; for example, projects and system-level data.

In a standard EDQ instance, configuration information, including project information, is stored in the Director database:



The following figure shows an EDQ instance integrated with Subversion:



 **Note:**

The Director database is still required because it contains data derived from the file-mastered configuration that has been normalized to allow querying by the applications.

With EDQ configuration files mastered and stored in a Subversion repository, a Subversion client can be used to commit or otherwise access them. Because EDQ includes an embedded Subversion client, Subversion client operations to control

configuration changes can be performed directly in Director once the EDQ integration with Subversion has been enabled.

## Setting Up a Repository

The first stage of configuration is to create a workspace directory where the checked out data is stored:

1. Create a directory on the disk where desired (for example, `C:\MyRepository`) and then add it and commit it to Subversion.
2. Inside the newly created directory, set the following Subversion property:

```
svn propset edq:systemversion 12.1.3:base .
```

 **Note:**

Set Subversion property to "12.1.3:base" and not to the current version of EDQ.

3. Commit these changes into Subversion. Your workspace now displays these properties:

```
svn proplist -v .  
Properties on '.':  
  edq:systemversion  
  12.1.3:base
```

4. Create the following subdirectories in the newly created directory:
  - Data Stores
  - Hidden Reference Data
  - Images
  - Projects
  - Published Processors
  - Reference Data
5. Add and commit these directories. The repository is now set up correctly for EDQ.

The preceding steps only need to be performed once per repository. All remaining changes can be made using EDQ.

## Configuring EDQ with Subversion

Subversion must be integrated with a fresh installation of EDQ.

 **Caution:**

When an EDQ instance is integrated with Subversion, all pre-existing and other configuration information is lost. To retain this information, you must package and export it first. For further details, see [Retaining Existing Configuration Information](#).

 **Note:**

Oracle recommends that a single workspace be assigned to each instance of EDQ because it is difficult to move between workspaces in a single EDQ instance.

## Configuring a New EDQ Installation

To configure a new EDQ installation:

1. Shut down the application server.
2. Check out the workspace from Subversion. It is not necessary to checkout the whole tree; just the workspace directory itself is required.
3. Edit the `director.properties` file in the `ORACLE_HOME/user_projects\domains\domains\edq_domain\edq\oedq.local.home` directory.
4. Add the following line replacing the directory path with that of the absolute path to your root workspace directory. For example:

```
sccs.workspace = C:\\MyRepository
```

 **Note:**

This example demonstrates the need to escape colon (:) and backslash (\) characters in the path with a backslash. You must also escape space characters in the path with a backslash.

5. Start your EDQ server, and then start Director.
6. Check the top of the `Main0.log` file for an `INFO` message listing the name of the SCCS workspace you added. For example:

```
INFO: 02-Sep-2013 10:05:21: SCCS workspace is C:\MyRepository
```
7. If no errors follow this message, EDQ is configured to use Subversion. If there are errors, see [Troubleshooting Errors](#), for possible solutions.

## Retaining Existing Configuration Information

As previously stated, Subversion must be integrated with a fresh installation of EDQ. Therefore, any pre-existing projects and other configuration items in an EDQ

installation must be packaged before integration begins and then imported to the new installation afterwards:

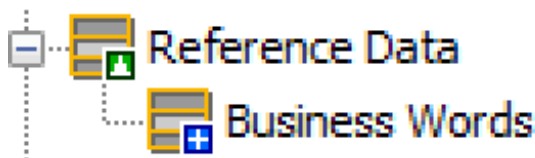
1. Package all configuration items in the current EDQ instance into DXI files.
2. Install a new instance of EDQ with the Subversion integration enabled.
3. Import the DXI files into the new instance, and commit the files to the Subversion workspace.
4. Check that the configuration items are all valid and working correctly.  
Note that all passwords for Data Stores must be re-entered after a configuration import.
5. Decommission the previous instance.

## Understanding the Integration Elements

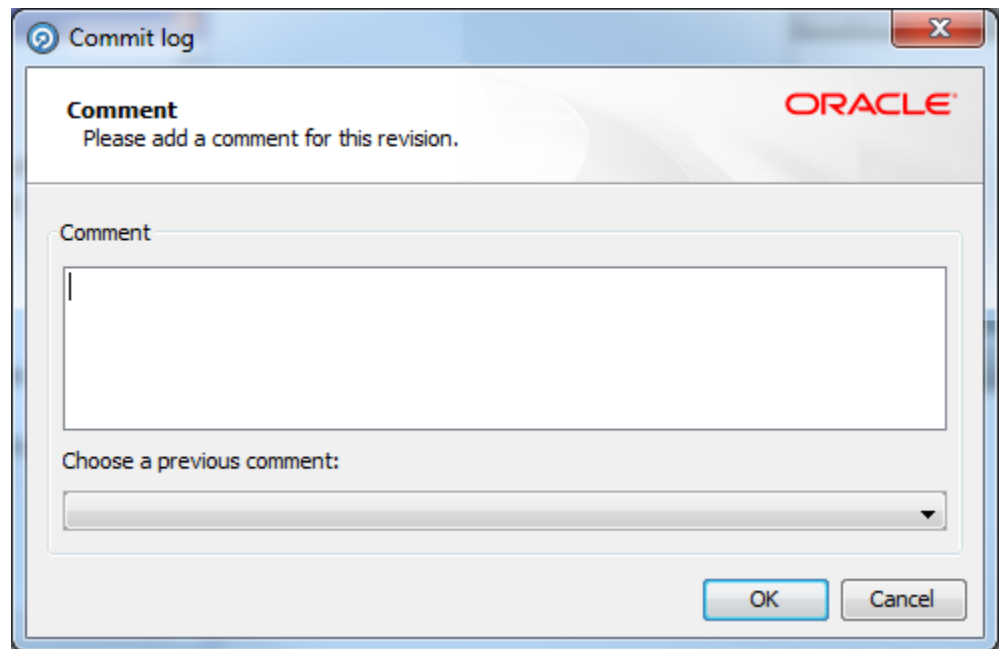
Once EDQ is integrated with Subversion enabled, the following interface elements become visible within the Director application:

- Subversion status icon overlays in Project Browser - There are two icons used to indicate the three possible Subversion statuses of nodes in the Project Browser:
  - No icon - The node (and its sub-nodes) are all up to date.
  - Green icon- This node (and its sub-nodes) have modifications.
  - Blue icon - This node (and its sub-nodes) is new/currently not under Version Control.

For example, the following image shows both icons in use. The Reference Data node is modified (green icon) as one of its sub-nodes has changed. A new piece of Reference Data, Business Words, has been added, and is marked with the blue icon:

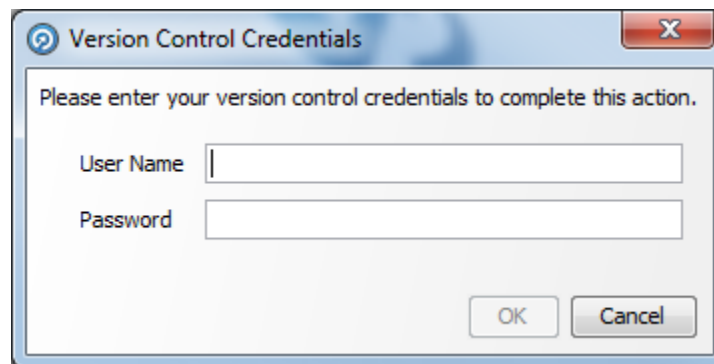


- Version Control tab - The Properties dialog (displayed by right-clicking on an item in the Project Browser and selecting **Properties**) now contains a Version Control tab that describes the state of the item, when it was last updated, its Subversion revision, and whether it is current.
- New context menu for Version Control - The Project Browser right-click menu now contains a Version Control option. When selected, this displays a sub-menu with Subversion options to update, commit, revert, compare or view the log for the item. These options are recursive. For example, if you perform View Log on a single process then you will see the log for this process only, but if you perform View Log on the Processes node you will see changes for all processes.
- Comment and credentials dialogs on commit - When you commit changes to the repository, Director displays the Commit log dialog:



In this dialog you can enter a comment describing the change in the Comment field. Alternatively, you can automatically populate the field by choosing a comment from the list of comments previously entered in the current session.

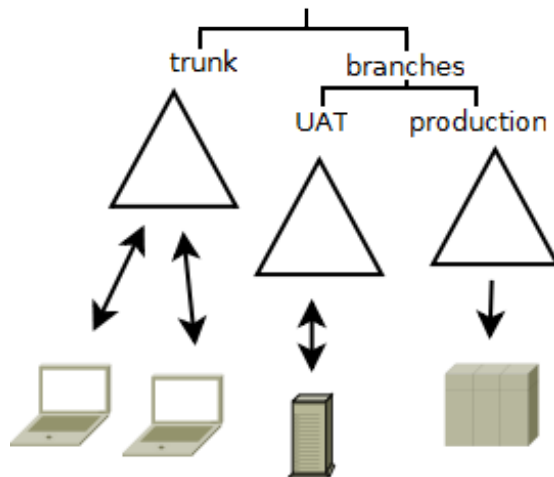
After you click **OK** in the Commit log dialog, Director displays the Version Control Credentials dialog if you have not already provided your credentials in the current session:



In this dialog you enter your user name and password for the Subversion repository and then click **OK**.

## Reviewing a Deployment Example

An example deployment is presented here. In this illustration, there is a single Subversion server that holds three copies of the configuration for four EDQ installations:



The copies of the configuration are:

- **trunk** - the traditional location that all development work is performed on. New features of the configuration are developed and saved here.
- **branches and UAT** - this branch represents the copy of the configuration under UAT testing.
- **branches and production** - this branch represents the production copy of the configuration.

The four EDQ installations using the Subversion server for storing their configuration are:

- Two development laptops where design work and maintenance of existing projects are carried out.
- A UAT server for User Acceptance Testing changes.
- A production server for production runs.

In this example deployment, the laptop users develop configuration for individual projects on their own laptops and then commit changes back to the subversion repository on "trunk". Where the developers are co-operating on developing a project they will periodically update their local installation to pick up changes from the other developer.

At some point development reaches a point where it needs to be released to UAT for testing. A release manager then copies the necessary projects from "trunk" to "UAT" on the subversion server.

For example, the following Subversion command may be used:

```
svn cp -m"Release Project X to UAT" http://svn/repos/config/trunk/ProjectX  
http://svn/repos/config/branches/UAT
```

The test manager then updates the UAT server's projects to load the new configuration into the EDQ server. Over a period of time testing continues. As issues are found they are fixed in the UAT environment and committed back to the subversion repository.



Once UAT environment has achieved an acceptable test level it is promoted to release. This achieved in much the same way as the release from development to UAT. The necessary projects are copied across in the version control repository and then the production server is updated to use this configuration.

## Troubleshooting Errors

You may encounter the following errors for which the cause and solution is provided.

Error	Cause and Solution
Configuration database is not compatible with workspace	The database has been used with a different workspace. This error usually arises occurs when operations have been performed in EDQ before Subversion version control is enabled. There are two solutions: drop and recreate the Director database or reinstall EDQ.
Unable to open an ra_local session to URL	This may occur when trying to commit files to an invalid repository. The EDQ integration is not compatible with file-based repositories (those repositories beginning with <code>file:///</code> or <code>C:\example</code> ). A fully declared <code>http://</code> path to the repository must be made.

# 2

## Integrating with IBM Global Name Recognition

This chapter describes how to integrate with IBM Global Name Recognition (GNR). You can configure EDQ to connect to IBM GNR to facilitate linguistic analysis of names, and linguistically sensitive name searching.

This chapter includes the following sections:

- [System Requirements](#)
- [Configuring the EDQ Server](#)
- [Building the Search Library](#)
- [Configuring the GNR Connector](#)
- [Creating the Search Configuration Files](#)

### System Requirements

To enable EDQ connectivity with IBM GNR, you must have the following:

- EDQ 12c (12.2.1.1.0) installed on 64-bit AIX operating system or Linux operating system running 64-bit Java.
- IBM GNR 4.2.1 (4.2 + 4.2.1 fixpack) or later, including the hotfix based on GNR 4.2.2 (4.2 +4.2.2 fixpack). For more information, see the IBM website at <http://www.ibm.com>.

EDQ does not make use of any of the web services provided by GNR, so you do not need to configure these during GNR installation.

#### Note:

GNR can only be installed on an EDQ instance if you have the required license agreements with both Oracle and IBM.

### Configuring the EDQ Server

The `LD_LIBRARY_PATH` must be set as required for the installation environment.

The EDQ GNR analytic processors use a shared library (`.so`) in the `lib64` directory of the GNR installation. This directory must be specified in an environment variable passed to the EDQ server.

In a Linux 64-bit environment, the environment variable is `LD_LIBRARY_PATH`; for example:

```
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:gnr-installation-dir/lib64
```

*gnr-installation-dir* is the GNR installation path; for example, `/opt/GNR/GNM`.

In an AIX environment, the environment variable name is `LIBPATH` instead of `LD_LIBRARY_PATH`.



#### Note:

The environment variable must be available to the application server process.

## Building the Search Library

The GNR Search Processor uses a native library that must be linked with the GNR libraries.

Oracle supplies these files to create the library:

- Two Makefile templates, one for each platform, that script the building of the search library
- The `namehunter.o` object module file

When building the library on an AIX system, IBM C++ must be available. When building the library on Linux systems, the GCC C++ compiler must be available and it must be of the same version used to create the GNR libraries, as specified in the IBM GNR documentation.

The Makefile template for a 64-bit AIX operating system is as follows:

```
# Build library from object file
# -----

GNR=/opt/GNR/GNM
CFLAGS=-qmkshrojb
LIBS=-lNameHunter -lNameTransliterator -lsicu18n -lsicuuc -lsicudata
SDK=aix61_64-xlc9-release
LIBDIRS=-L$(GNR)/sdk/$(SDK)/lib -L$(GNR)/sdk/icu4c/$(SDK)/lib

all: libnimrod.so

libnimrod.so: namehunter.o
        xlc++_r -q64 $(CFLAGS) -o $@ $? $(LIBDIRS) -lNameHunter -lNameTransliterator
        -lsicu18n -lsicuuc -lsicudata
```

The Makefile template for 64-bit Linux operating system is as follows:

```
# Build 64-bit library from object file
# -----

GNR=/opt/GNR/GNM
CFLAGS=-shared -fPIC

all: libnimrod.so

libnimrod.so: namehunter.o
        g++ -m64 $(CFLAGS) -o $@ $? -L$(GNR)/sdk/rhel4_64-gcc34-release/lib
```

```
-I$(GNR)/sdk/icu4c/rhel4_64-gcc34-release/lib -lNameHunter -lNameTransliterator  
-lsicuil8n -lsicuuc -lsicudata
```

Before running the `Makefile` script for your platform, ensure that the value of `GNR` in the `Makefile` template is set correctly, according to the GNR installation directory location. On the AIX operating system, also ensure that the value of `SDK` is set correctly, according to the system architecture.

After running the edited `Makefile`, the newly created `libnimrod.so` shared library file can be installed anywhere and can be copied to other systems with GNR installs.

## Configuring the GNR Connector

The EDQ GNR connector requires three types of configuration files to integrate it with GNR:

- The `gnr.properties` properties file in the EDQ installation
- The `nameworks.config` configuration file in the GNR installation
- Search configuration files in the EDQ installation

## Creating the EDQ GNR Properties File

The `gnr.properties` file and the `gnr` subdirectory that contains it must be manually created and placed in the `gnr` subdirectory of the EDQ configuration directory. It must contain the following properties:

### **gnr.install**

The GNR installation path. This is the path to the directory containing the following GNR subdirectories:

- `bin`
- `bin64`
- `data` (which contains the GNR data files)
- `lib`
- `lib64`

### **analytics.config**

The absolute location of the `nameworks.config` configuration file in the GNR installation.

### **search.jnilib**

The absolute location of the `libnimrod.so` shared library, which was built using the `Makefile` template.

### **nameworks.config**

During GNR installation, a `nameworks.config` file is created and stored in the GNR data directory.

The critical part of the `nameworks.config` file is the reference files section:

```
[Reference Files]  
NameSifter=/opt/GNR/GNM/data/SifterRules.ibm
```

The `NameSifter` value must refer to the `SifterRules.ibm` file in the GNR installation.

## Creating the Search Configuration Files

Search configuration files are located in the `gnr/search` subdirectory of the EDQ configuration directory. They are read by the connector and used to set parameters for the Search function.

A sample search configuration file named `search.config` is available in the `support/data/search` subdirectory of the EDQ installation. To create a search configuration file, copy this sample file to the `gnr/search` subdirectory of the EDQ configuration directory and edit the copy to suit your needs.

## Support for GNR 3.2 and GNR 4.2 in Search Configuration Files

The search configuration format changed slightly from GNR 3.2 to GNR 4.2, and the EDQ GNR connector supports both versions as far as possible. It also processes data for Organization searches.

The basic differences between the search configuration files in GNR 3.2 and GNR 4.2 are:

- GNR 4.2 specifies the parameter files (for example, tags and variants) in the `[hunter]` section. GNR 3.2 uses the `[search]` section. The EDQ GNR connector looks in the `[hunter]` section first then the `[search]` section.
- The tag and variant files in GNR 4.2 are specified by keys such as `ibmTaqFile` and `custTaqFile`. In GNR 3.2, just `taqFile` is used. The EDQEDQ GNR connector looks for `taqFile`, `ibmTaqFile` and `custTaqFile` and loads each if found. The same rules are used for variant and terms files.
- The generic reg file is set by a specific `genericRegFile` setting; in GNR 3.2 this always defaults to the `anglo` reg file.
- Some settings have been added to the `[parms]` sections, and others have been removed.

# 3

## Integrating with Experian QAS

This chapter describes how to integrate with Experian QAS API and migrate earlier versions of the QAS integration to EDQ versions 8.1.3 or later. This chapter includes the following sections:

- [Integrating with Experian QAS](#)
- [Migrating QAS integrations](#)

### Software Requirements

EDQ includes a connector to the Experian QAS Batch API. You must have an installed version of the Experian QAS Batch API software appropriate for your platform. For more information, see the Experian Data Quality website found at <http://www.qas.com/>.

### Integrating with Experian QAS

Integration of EDQ with Experian QAS Batch is carried out by editing the `gas.properties` file distributed with EDQ. The `gas.properties` file is located in the `ORACLE_HOME/user_projects/domains\domains\edq_domain\edq\oedq.home\gas` directory.

Once both EDQ and the QAS Batch software have been installed, edit the properties in the `gas.properties` as required for your integration. The properties specified in the file are as follows:

Property name	Description	Default Value
<code>gas.install.path</code>	The location of the Experian QAS Batch installation.	C:\Program Files (x86)\QAS\QuickAddress Batch API
<code>gas.qaworld.ini</code>	The path to the <code>QAWorld.ini</code> file to use. This allows you to create and edit copies of the original <code>QAWorld.ini</code> file distributed with Experian QAS Batch API. If no value is specified for this property, it uses the <code>QAWorld.ini</code> file within the Experian QAS Batch API installation.	None

Property name	Description	Default Value
<code>max.number.connections</code>	The maximum number of connections that EDQ will create to the API. This should <b>not</b> be set to a value greater than 32.  Note: There is a known threading and memory issue with Experian QAS Batch API versions 6.85, 6.89 and 6.95, where if maximum usage of 32 instances and 8 threads is reached, the Batch API may crash. This can be avoided by setting this property to a value between 18 and 22 inclusive.	32
<code>connection.pool.timeout</code>	This property specifies the number of milliseconds that a connection can be idle for before it is closed by the pool management functionality. If this is set to -1, idle connections will not be closed.	60,000
<code>connection.pool.timer.interval</code>	This property specifies how often, in milliseconds, the connection pool is scanned for idle connections.	60,000
<code>default.layout</code>	The default layout to use. If no layout with the specified name is available in the specified <code>QAWorld.ini</code> file, this property is ignored.	GBR

## Migrating QAS integrations

Some earlier versions of EDQ (versions prior to 8.1.3) were shipped with a customized version of `QAWorld.ini` that was used instead of the version contained within Experian QAS. It is not possible to migrate automatically from these versions of the Experian QAS integration to the later versions. To migrate an earlier Experian QAS integration, you must:

- Locate the local copy of `QAWorld.ini` and copy any custom layouts specified in this file into the version of `QAWorld.ini` contained within Experian QAS.
- Update the settings in `gas.properties`. See [Integrating with Experian QAS](#).

In addition, any existing processes and results books that make use of QAS processors must be updated as follows:

- Open each configured QAS processor and rename the output attributes to match those in the new `QAWorld.ini` file.
- Open any results books built on results grids from QAS processors and re-map the fields to the new output attribute names.

Refer to the following table for the affected releases for each version of EDQ. All versions of EDQ (previously known as `dn:Director`) prior to 7.2 are affected:

<b>Version</b>	<b>Releases affected</b>
EDQ (dn:Director) 7.2	Release 7.2.9 and all earlier 7.2 releases
EDQ (dn:Director) 8.0	Release 8.0.21 and all earlier 8.0 releases
EDQ (dn:Director) 8.1	Release 8.1.2 and all earlier 8.1 releases



# 4

## Integrating with Capscan Matchcode

This document describes how to integrate address verification and cleansing features from GBGroup Capscan Matchcode with EDQ. This documentation is intended for system administrators responsible for installing and maintaining EDQ applications. This chapter includes the following sections:

- [Software Requirements](#)
- [Integrating the Capscan Matchcode Libraries into EDQ](#)
- [Customizing the Matchcode API](#)

### Software Requirements

You must have the Capscan Matchcode software installed on a system that is accessible to the EDQ Server. For more information, see the GBGroup Matchcode website found at <https://www.gbgroup.com/products/>.

### Integrating the Capscan Matchcode Libraries into EDQ

EDQ includes a connector to the Capscan Matchcode API. EDQ includes a connector to the Capscan Matchcode API from GBGroup. This API provides address verification and cleansing features. This API provides address verification and cleansing features. Integrate the Capscan Matchcode Libraries into EDQ as follows:

1. Copy the `capscan.jar` client API file from the Capscan Matchcode installation to the `ORACLE_HOME/user_projects/domains/edq_domain/servers/edq_server1/tmp/_WL_user/edq/iz3lfy/war/WEB-INF/widgetjars` directory.

The location of the `capscan.jar` file in a Capscan Matchcode installation depends on the installation platform; for example, it is located in the `Capscan\SDK\Matchcode client API\Java` directory on Windows.

#### Note:

You can copy the `capscan.jar` file to a directory other than the default directory given in this step. If you do so, you must edit the `capscan.jar` property of the `capscan.properties` file to specify the location of the file. If you specify a relative path, the path must be relative to one of the directories in the EDQ configuration path.

2. Edit `capscan.properties` file in the `ORACLE_HOME/user_projects/domains/domains/edq_domain/edq/oedq.home/capscan` directory.
3. Edit the `server.host` property to refer to the system where Capscan Matchcode is running.
4. Restart your Server.

## Customizing the Matchcode API

Various aspects of the Capscan Matchcode API behavior can be controlled using the `capscan.properties` file. It allows you to set the following properties:

### **capscan.jar**

The location of the `capscan.jar` client API file in the installation. If you specify a relative path, the path must be relative to one of the directories in the configuration path. The default value is the relative path `capscan/capscan.jar`.

### **server.host**

The IP address of the machine where Capscan Matchcode is running.

### **connection.timeout**

A timeout period, in seconds, after which the Capscan Matchcode API will abort the search and return (a timeout period of zero indicates that there is no time limit on searches).

### **connection.type**

The connection mode to use when communicating with the Capscan Matchcode API. The default connection mode is `CONNECTIONLESS`. The remaining options are:

- `CONNORIENTED`
- `STATELESS`
- `WEBCONNECTION`
- `ONDEMAND`

For information about these connection modes, refer to the Capscan Matchcode API documentation.

### **number.capscan.connections**

The number of connections that EDQ should make to the Capscan Matchcode API.

### **number.threads**

The number of threads that should be used when communicating with the Capscan Matchcode API.

The default contents of the `capscan.properties` file are as follows:

```
# This configuration file is configuring the CapScan processor
# to be able to communicate with the CapScan server
#
# Capscan server name
server.host = 127.0.0.1
#
# Connection timeout in seconds (0 means no time out)
connection.timeout = 30
#
# The connection type to make to the CapScan server.
# Possible values are:
#
# CONNORIENTED
# CONNECTIONLESS
# STATELESS
# WEBCONNECTION
# ONDEMAND
```

```
connection.type = CONNECTIONLESS
#
# The number of connections the director server should
# make to the CapScan server
number.capscan.connections = 1
#
# The number of threads that should be used to
# communicate with the CapScan server
number.threads = 1
```

# 5

## Using the Command Line Interface

This chapter describes how to use the command line interface. This chapter includes the following sections:

- [Running the Command Line Interface](#)
- [Understanding the Commands and Arguments](#)
- [Reviewing Examples](#)

The command line interface, `jmxtools.jar`, provides access to a number of facilities.

### Running the Command Line Interface

The command line interface is distributed as a self contained `.jar` file in the `tools` directory, and is executed by the following command line invocation:

```
java -jar jmxtools.jar commandname arguments
```

The commands and arguments are described in the following section.

### Understanding the Commands and Arguments

The command line interface can run a number of commands and provides functionality including:

- Running jobs
- Listing and dropping orphaned results tables
- Showing user session logs
- Shutting down real-time jobs
- Checking the EDQ version number

The following sections provide a full guide to the commands, arguments and options available.

EDQ also provides support for jobs through the REST-based EDQ Configuration API interfaces. For details on using these interfaces for performing various jobs, see "[REST Interfaces for Jobs](#)" in the "*Integrating Enterprise Data Quality With External Systems*" guide.

`runjob`

The `runjob` command runs a named job in the same way as if running the job using the Director UI. The `runjob` command takes the following arguments:

Argument	Use
<code>-job <i>job_name</i></code>	Specifies the name of the job to run.

Argument	Use
<code>-project project_name</code>	Specifies the name of the project that contains the job .
<code>-u user_name</code>	Specifies the user name to use to connect to the EDQ server. The user must have permission to run jobs and must have permission to the project containing the job.
<code>-p password</code>	Specifies the connecting user's password. If the <code>-p</code> option is not set, EDQ will prompt the user for the password.
<code>-nolockwait</code>	Indicates that if any of the resources used by the job are locked, the job should not wait for them to become available. Instead, it should terminate with a failure code and return control to the command line. The <code>-nolockwait</code> argument takes no extra values.
<code>-nowait</code>	Indicates that the command line should not wait for the job to complete. The <code>-nowait</code> argument takes no extra values.
<code>server:port</code>	Specifies the server and port of the JMX (management) interface.

### runopsjob

The `runopsjob` command runs a named job in the same way as if running the job using the Server Console user interface. This provides additional functionality to the `runjob` command, specifically the use of Run Labels and Run Profiles. Run Labels may be used to store results separately from other runs of the same job. Run Profiles may be used to override externalized configuration settings at runtime.

The `runopsjob` command takes the following arguments:

Argument	Use
<code>-job job_name</code>	Specifies the name of the job to run.
<code>-project project_name</code>	Specifies the name of the project that contains the job
<code>-u user_name</code>	Specifies the user name to use to connect to the EDQ server. The user must have permission to run jobs and must have permission to the project containing the job.
<code>-p password</code>	Specifies the connecting user's password. If the <code>-p</code> option is not set, EDQ will prompt the user for the password.
<code>-nolockwait</code>	Indicates that if any of the resources used by the job are locked, the job should not wait for them to become available. Instead, it should terminate with a failure code and return control to the command line. The <code>-nolockwait</code> argument takes no extra values.
<code>-nowait</code>	Indicates that the command line should not wait for the job to complete. The <code>-nowait</code> argument takes no extra values.
<code>-runlabel run_label_name</code>	Specifies the name of the run label under which you wish to store staged output results. Note that this will override any run label that is specified in a run profile or by <code>-D runlabel = run_label_name</code> .
<code>-props run_profile_name</code>	Specifies the full path to a run profile properties file containing override settings for externalized configuration options in the job.

Argument	Use
-D <i>externalized_option</i> =value	Allows you to override specific externalized options for the job individually. The syntax for the externalized options and values is the same as used in run profile properties files. Note that characters is interpreted by the command line, so some characters will need to be escaped according to the shell conventions of your environment. Also note that any individually specified externalized option settings will override any settings for the same option if these are specified in a run profile used in the same run.
<i>server:port</i>	Specifies the server and port of the JMX (management) interface.

#### droporphans

The `droporphans` command is used to remove any orphaned results tables that may be created when processes are terminated unexpectedly. It should not be run when any jobs or processes are running on the EDQ server.

The `droporphans` command takes the following arguments:

Option	Use
-u <i>user name</i>	Specifies the user name to use to connect to the server. The user must have permission to cancel jobs and must have permission to the project containing the job.
-p <i>password</i>	Specifies the connecting user's password. If the -p option is not set, will prompt the user for the password.
<i>server:port</i>	Specifies the server and port of the JMX (management) interface.

#### listorphans

The `listorphans` command is used to identify any orphaned results tables. The `listorphans` command takes the same arguments as the `droporphans` command.

#### scriptorphans

The `scriptorphans` command creates a list of SQL commands for dropping orphaned results tables. This is useful if you want to review exactly which commands will run on the Results database when you drop tables, or if you want to drop the tables yourself manually.

#### list

The `list` command lists all the available commands.

#### showlogs

The `showlogs` command starts a small graphical user interface application that allows user session logs to be retrieved.

shutdown

The `shutdown` command shuts down all real-time jobs. These are jobs that are running from real-time record providers (web services or Java Message Service).

The `shutdown` command takes the following arguments:

Option	Use
<code>-u user name</code>	Specifies the user name to use to connect to the server. The user must have permission to cancel jobs and must have permission to the project containing the job.
<code>-p password</code>	Specifies the connecting user's password. If the <code>-p</code> option is not set, EDQ will prompt the user for the password.
<code>-nowait</code>	Indicates that the command line should not wait for the job to complete. The <code>-nowait</code> argument takes no extra values.
<code>server:port</code>	Specifies the server and port of the JMX (management) interface.

version

The `version` command is used to identify the version of the currently installed instance of .

Enter the following at the command line:

```
java -jar jmxtools.jar version
```

The version number is returned.

## Reviewing Examples

This section lists several possible invocations of the command line interface:

- [Listing All the Available Commands](#)
- [Listing the Available Parameters for a Command](#)
- [Running a Named Job](#)
- [Running a Named Job in Operations Mode](#)

### Listing All the Available Commands

The following invocation of the command line interface lists all of the available commands:

```
java -jar jmxtools.jar -list
```

The output is as follows:

```
Available launch names:
<Job tools>
runjob Run named job
shutdown Shutdown realtime jobs
runopsjob Run named job in operations mode
```

```
<Logging>
showlogs Show session logs

<Database Tools>
listorphans List orphaned results tables
droporphans Drop orphaned results tables
scriptorphans Create script for dropping orphaned results tables

<System Information>
version Display version number of tools
```

## Listing the Available Parameters for a Command

If the command line interface is invoked by specifying a command without the corresponding parameters, it outputs detailed help for the command. For example, to get detailed help on the `runjob` command, invoke the command line interface as follows:

```
java -jar jmxtools.jar runjob
```

The output is as follows:

```
Usage: runjob -job jobname -project project [-u user] [-p pw] [-nowait] [-
nolockwait] [-sslprops props | -ssltrust store] server:port
```

## Running a Named Job

This example illustrates how to run a named job in a named project on a specific instance (as specified by machine name and port).

To run a job called "rulecheck" in a project called "Audit" on the local machine with a JMX server on port 8090 using a user named "dnadmin", the command is as follows:

```
java -jar jmxtools.jar runjob -job rulecheck -project audit -u dnadmin
localhost:8090
```

The application prompts the user to enter the password for the `dnadmin` user.

## Running a Named Job in Operations Mode

This example illustrates how to run a named job in 'operations mode' in a Windows environment. In operations mode, there is access to the Run Label and Run Profile capabilities so that the configuration of the job can be specified dynamically, and so that the results of the job can be stored by Run Label.

To run a job called "profiling" in a project called "MDM" on a server called "prod01", with a run label of "Nov2011" and a run profile file called `File1.properties`, with a JMX server on port 8090, the command is as follows:

```
java -jar jmxtools.jar runopsjob -job profiling -project MDM -runlabel
Nov2011 -props c:\ProgramData\Oracle\Enterprise Data Quality
\oedq_local_home\File1.properties" -u dnadmin prod01:8090
```



# 6

## Configuring Additional Database Connections

This chapter describes how you can configure additional database connections for use in Director.

- [Using JNDI to Connect to Data Stores](#)
- [Connecting to an Oracle Database Using tnsnames.ora](#)
- [Connecting to an Oracle Database Using Oracle Internet Directory \(LDAP\)](#)

The standard options for Director to connect to data stores are described in the online help. Once implemented, these options appear in the Data Store Configuration step of the New Data Store wizard in Director. For help with using this wizard, see the Director online help.

### Using JNDI to Connect to Data Stores

You can configure EDQ to use a Java Naming and Directory Interface (JNDI) data store connection.

1. Define the JNDI data store. JNDI is provided by the hosting application server. For more information about defining JNDI data sources in Oracle WebLogic Server, see "Using DataSource Resource Definitions" in .
2. In the EDQ data store wizard, specify JNDI as the type of data store, and then specify the JNDI name.

### Connecting to an Oracle Database Using tnsnames.ora

You can configure EDQ to use an Oracle Transparent Network Substrate (TNS) data store connection. To use this connection method, you specify a name from a `tnsnames.ora` file as the data source when using the data sources wizard. Only the `tnsnames.ora` file is needed. No other Oracle client software is needed.

### To Configure EDQ to Connect Through TNS

To connect EDQ through TNS:

1. Set the `oracle.net.tns_admin` Java system property to a local directory that contains the `tnsnames.ora` file.
2. Create a file named `jvm.properties` in your EDQ local configuration directory (`oedq_local_home` by default) and add an entry similar to the following:  
`oracle.net.tns_admin = c:\\temp`. This property may have been set already in the application server when EDQ was installed.

For more information about the `tnsnames.ora` file, see "Configuring the Local Naming Method" in .

## Connecting to an Oracle Database Using Oracle Internet Directory (LDAP)

You can configure EDQ to use an Oracle Lightweight Direct Access Protocol (LDAP) data store connection by setting the required Java system properties. These properties are:

```
dn.oracle.directory.servers = ldap://servername:port
```

```
dn.oracle.default.admin.context = dc=domaincontext1,dc=domaincontext2
```

The first property gives the location of your LDAP servers. The second property sets the context within the LDAP tree. Together, these properties enable EDQ to construct an Oracle and LDAP JDBC connection string, which looks similar to:

```
jdbc:oracle:thin:@ldap://servername:port/  
unicode,cn=Oraclecontext,dc=domaincontext1,dc=domaincontext2
```

# 7

## Configuring EDQ to Process XML Data Files

This chapter describes how can be configured to read and write XML data files. This chapter includes the following sections:

- [Using Simple XML Data Stores](#)
- [Using XML and Stylesheet Data Stores](#)

You can use XML data files in snapshots to read and write the data contained in the file. A snapshot is a staged copy of data in a data store that is used in one or more processes. provides two types of data stores for working with XML data files: Simple XML and XML and Stylesheet. Both are available for server-side and client-side data stores.

### Using Simple XML Data Stores

Simple XML data stores can read and write XML files that have a simple 2-level structure in which the top level tag represents the entity and the lower level tags represent the attributes of the entity. XML files exported from Microsoft Access are an example.

Following is an example of a simple XML file format that could be used with :

```
<dataroot>
  <Person>
    <Id>1</Id>
    <FirstName>Fred</FirstName>
    <LastName>Bloggs</LastName>
    <DateOfBirth>1972-01-31T00:00:00.000+0000</DateOfBirth>
    <Weight>85</Weight>
  </Person>
  <Person>
    <Id>2</Id>
    <FirstName>Jane</FirstName>
    <LastName>Smith</LastName>
    <DateOfBirth>1985-07-16T00:00:00.000+0100</DateOfBirth>
    <Weight>63</Weight>
  </Person>
</dataroot>
```

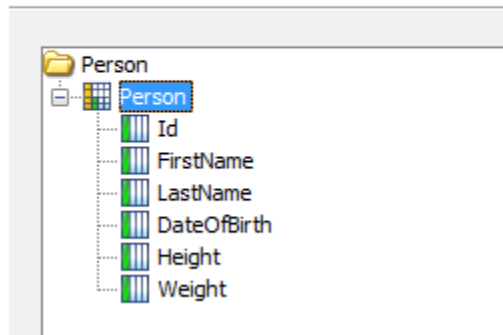
### Reading Simple XML Files

When reads Simple XML files the following occurs:

- The root element name is not used, so it can be anything.
- The record element name appears as the table name in the Table Selection page of the Snapshot Wizard dialog.

**Table Selection**


What data do you want to snapshot?



- The lower level element names appear as the column names in the Column Selection page of the Snapshot Wizard and therefore become EDQ attribute names.

**Column Selection**

Which columns do you want to snapshot?

	Column Name	Data Type
<input checked="" type="checkbox"/>	Id	VARCHAR
<input checked="" type="checkbox"/>	FirstName	VARCHAR
<input checked="" type="checkbox"/>	LastName	VARCHAR
<input checked="" type="checkbox"/>	DateOfBirth	VARCHAR
<input checked="" type="checkbox"/>	Height	VARCHAR
<input checked="" type="checkbox"/>	Weight	VARCHAR

## Writing Simple XML Files

When generating Simple XML files using an export to the data store, the name of the data store defines the record XML element name. The element `Person` in the example in [Using Simple XML Data Stores](#) shows how this appears in the XML.

The XML element names of the lower level tags are taken from the attribute names. EDQ names are encoded to ensure that invalid XML is not generated. For example, space characters in names are replaced by the character sequence `_x0020_`, so an attribute named `Date Of Birth` would generate XML elements in the following format:

```
<Date_x0020_of_x0020_Birth>
```

## Using XML and Stylesheet Data Stores

When there is a requirement to work with XML of a different structure than that of Simple XML, then you use the XML and Stylesheet data stores.

These data stores read and write XML conforming to the DN-XML schema and optionally allow the use of a custom stylesheet to:

- Transform XML from a custom XML format to DN-XML during data snapshot
- Transform XML from DN-XML to a custom XML format during data export

For more information about XML stylesheets, see the W3C website found at <http://www.w3.org/Style/XSL/> and <http://www.w3.org/standards/xml>.

## Using DN-XML

DN-XML is the format by which custom XML can be processed by .

An example of DN-XML is as follows:

```
<dn:data xmlns:dn="http://www.datanomic.com/2008/dnx">
  <dn:record skip="true">
    <dn:value name="Id" type="string"/>
    <dn:value name="FirstName" type="string"/>
    <dn:value name="LastName" type="string"/>
    <dn:value name="DateOfBirth" type="date"/>
    <dn:value name="Height" type="number"/>
    <dn:value name="Weight" type="number"/>
  </dn:record>
  <dn:record>
    <dn:value name="Id">1</dn:value>
    <dn:value name="FirstName">Fred</dn:value>
    <dn:value name="LastName">Bloggs</dn:value>
    <dn:value name="DateOfBirth">1972-01-31</dn:value>
    <dn:value name="Height">1.85</dn:value>
    <dn:value name="Weight">85</dn:value>
  </dn:record>
  <dn:record>
    <dn:value name="Id">2</dn:value>
    <dn:value name="FirstName">Jane</dn:value>
    <dn:value name="LastName">Smith</dn:value>
    <dn:value name="DateOfBirth">1985-07-16</dn:value>
    <dn:value name="Height">1.65</dn:value>
    <dn:value name="Weight">63</dn:value>
  </dn:record>
</dn:data>
```

This is the equivalent DN-XML for the example given in [Using Simple XML Data Stores](#).

Note that the attribute names are defined differently in DN-XML compared with Simple XML. Because DN-XML uses attribute content to specify attribute names, it is possible to create attributes with spaces and other special characters in their names.

In the previous example, the `<dn:record skip="true">` XML element and its contents allows the definition of the structure of the source including the field names and their data types. All other record elements define a row of data in . This is analogous to the header row in a comma-separated values file. The following data types are permitted:

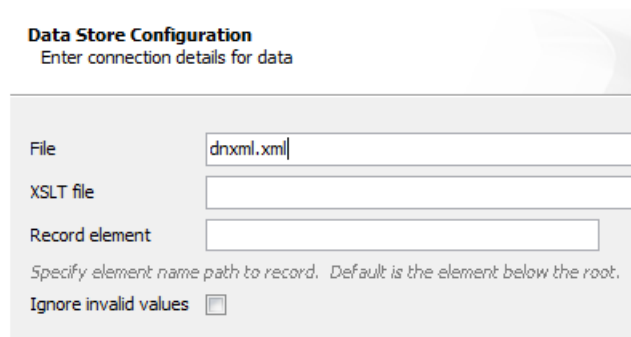
- string
- date
- number

 **Note:**

Date values in DN-XML files should be specified in the XSD date format (ISO 8601). For example, '2008-10-31T15:07:38.6875000-05:00' or without the time component simply as '2008-10-31'.

Within a data record, value elements are used to specify attribute values for the record. The name attribute is used to specify the attribute in question and the text content of the attribute specifies the value for that attribute. For example, the XML fragment, `<dn:value name="FirstName">Fred</dn:value>`, assigns the value 'Fred' to the attribute 'FirstName'.

DN-XML files can be read in to by creating an XML and Stylesheet data store and specifying the location of the XML source file; the XSLT file option should be left blank:



**Data Store Configuration**  
Enter connection details for data

File:

XSLT file:

Record element:

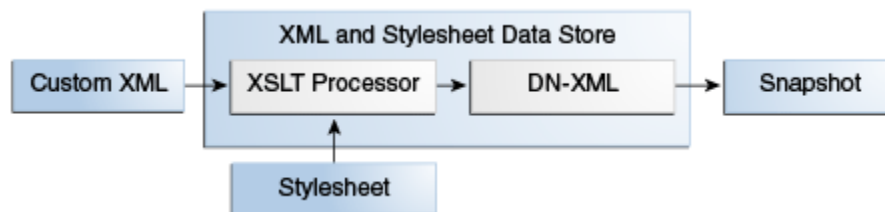
*Specify element name path to record. Default is the element below the root.*

Ignore invalid values

Similarly, can write DN-XML files by exporting data to an XML and Stylesheet data store with the XSLT option left blank.

## Reading Custom XML Files

XML files in custom formats can be read by using the XML and Stylesheet data store configured to use a custom XML stylesheet (XSLT) to transform from the custom schema to the DN-XML schema during data snapshotting.



Following is an example custom XML file that could be read into :

```

<crmdata>
  <contacts>
    <contact id="1">
      <name>

```

```

        <firstname>Fred</firstname>
        <surname>Bloggs</surname>
    </name>
    <dob>1972-01-31</dob>
    <properties>
        <property name="height" value="1.85"/>
        <property name="weight" value="85"/>
    </properties>
</contact>
<contact id="2">
    <name>
        <firstname>Jane</firstname>
        <surname>Smith</surname>
    </name>
    <dob>1985-07-16</dob>
    <properties>
        <property name="height" value="1.68"/>
        <property name="weight" value="63"/>
    </properties>
</contact>
</contacts>
</crmdata>

```

The following XML stylesheet demonstrates one way that the preceding example custom XML can be transformed into a suitable DN-XML format:

```

<xsl:stylesheet version="1.0" xmlns:dn="http://www.datanomic.com/2008/dnx"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:fn="http://www.w3.org/2005/02/xpath-functions">

  <xsl:output method="xml"/>

  <xsl:template match="/">
    <dn:data>

      <!-- Write out the header record -->
      <dn:record skip="true">
        <dn:value name="Id" type="string"/>
        <dn:value name="FirstName" type="string"/>
        <dn:value name="LastName" type="string"/>
        <dn:value name="DateOfBirth" type="date"/>
        <dn:value name="Height" type="number"/>
        <dn:value name="Weight" type="number"/>
      </dn:record>

      <!-- Get each contact record -->
      <xsl:apply-templates select="/crmdata/contacts/contact"/>

    </dn:data>
  </xsl:template>

  <xsl:template match="contact">

    <!-- Write out a data record -->
    <dn:record>
      <dn:value name="Id"><xsl:value-of select="@id"/></dn:value>
      <dn:value name="FirstName"><xsl:value-of select="name/firstname"/></dn:value>
      <dn:value name="LastName"><xsl:value-of select="name/surname"/></dn:value>
      <dn:value name="DateOfBirth"><xsl:value-of select="dob"/></dn:value>
      <dn:value name="Height">

```

```

        <xsl:value-of select="properties/property[@name='height']/@value"/>
    </dn:value>
    <dn:value name="Weight">
        <xsl:value-of select="properties/property[@name='weight']/@value"/>
    </dn:value>
</dn:record>

</xsl:template>

</xsl:stylesheet>

```

## Configuring the Data Store

The data can be read in to by creating an XML and Stylesheet data store and specifying the location of the XML source file and the XSLT file (stylesheet).

XML and Stylesheet Configuration

File: crmdata.xml

XSLT file: crm.xsl

Record element: contacts/contact

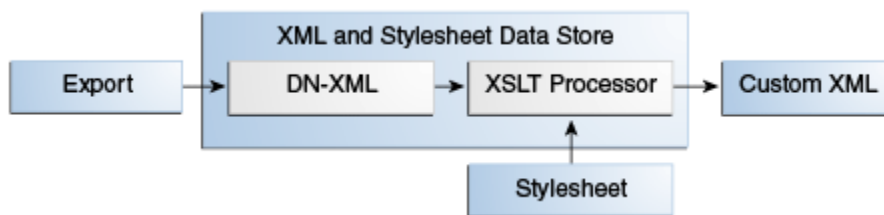
Specify element name path to record. Default is the element below the root.

Ignore invalid values

reads the source XML file in chunks for efficiency breaking up the file on record boundaries. By default uses the element immediately below the root as the record element. If this is not the case in the source XML file then an XPath-style expression to the record element from the root must be specified.

## Writing Custom XML Files

XML files in custom formats can be written by using the XML and Stylesheet data store configured to use a custom XSLT to transform from the DN-XML schema to the custom target schema the during data export.



Following is an example target custom XML format that needs to be generated by :

```

<Report>
  <Person Id="1" FullName="Fred Bloggs"/>
  <Person Id="2" FullName="Jane Smith"/>
</Report>

```



The following XML stylesheet demonstrates one way in which the DN-XML format can be transformed into the target custom XML format:

```
<xsl:stylesheet version="1.0"
  xmlns:dn="http://www.datanomic.com/2008/dnx"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:fn="http://www.w3.org/2005/02/xpath-functions">

  <xsl:output method="xml"/>

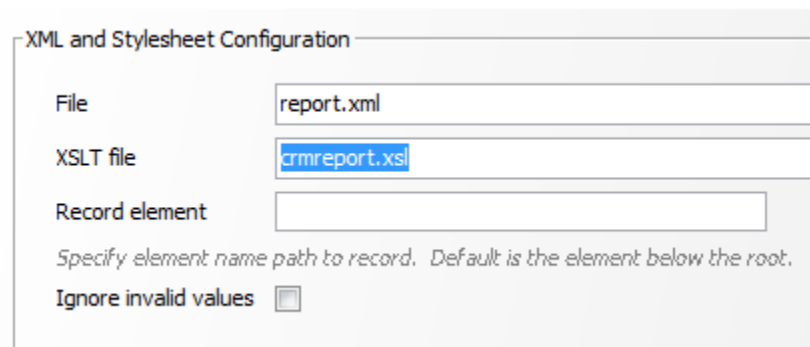
  <xsl:template match="/">
    <Report>
      <xsl:apply-templates select="/dn:data/dn:record"/>
    </Report>
  </xsl:template>

  <xsl:template match="dn:record">
    <Person>
      <xsl:attribute name="Id">
        <xsl:value-of select="dn:value[@name = 'Id']"/>
      </xsl:attribute>
      <xsl:attribute name="FullName">
        <xsl:value-of select="dn:value[@name = 'FirstName']"/>
        <xsl:text> </xsl:text>
        <xsl:value-of select="dn:value[@name = 'LastName']"/>
      </xsl:attribute>
    </Person>
  </xsl:template>

</xsl:stylesheet>
```

## Configuring the Data Store

The data can be written by creating an XML and Stylesheet data store and specifying the destination for the custom XML file and XSLT (stylesheet) file.



The screenshot shows a configuration window titled "XML and Stylesheet Configuration". It contains the following fields and options:

- File:** report.xml
- XSLT file:** crmreport.xsl
- Record element:** (empty text box)
- Specify element name path to record. Default is the element below the root.*
- Ignore invalid values:**

# 8

## Using the EDQ Configuration API

EDQ provides a set of REST-based interfaces that enable you to perform various configuration tasks programmatically, using any preferred programming language. In this chapter, the EDQ service is assumed to be installed at:

```
http://edqserver:8001/edq
```

This chapter provides a detailed description of these interfaces and the operations that can be performed using these interfaces. It includes the following topics:

- [REST Interface for Projects](#)
- [REST Interface for Data Stores](#)
- [REST Interface for Snapshots](#)
- [REST Interface for Processes](#)
- [REST Interfaces for Jobs](#)
- [REST Interface for Reference Data](#)
- [REST Interface for Web Services](#)
- [Example: Profiling from an External Application](#)

### REST Interface for Projects

The REST interface for working with EDQ projects is

```
http://edqserver:8001/edq/config/projects
```

This interface allows you to perform the following tasks:

- [Retrieving a List of EDQ Projects](#)
- [Creating a Project](#)
- [Deleting a Project](#)

### Retrieving a List of EDQ Projects

To get a list of all projects that are available with the current EDQ installation, you need to simply run an HTTP GET operation on the REST interface for EDQ projects, as shown in the following code:

```
GET http://edqserver:8001/edq/config/projects
```

When this code runs successfully, a list of projects is generated in JSON format:

```
[
  {
    "id":10,
    "name":"My Project"
```

```
    },  
    {  
      "id":12,  
      "name": "Scratch Project"  
    }  
  ]
```

## Creating a Project

To create a new project, you need to create a JSON object that describes the project to be created and then send it in the request body of the REST call using an HTTP POST.

For example:

```
POST http://edqserver:8001/edq/config/projects
```

```
{ "name" : "Profile Customer Names" ,  
  "description" : "Profile my customers" }
```

This code returns a response of type "OK" with a response body similar to the following:

```
{"id":14,"name":"Profile Customer Names"}
```

However, if an error occurs while creating a project, then a response of type "500 Internal Server Error" is generated, along with an error message similar to the following:

```
"Profile Customer Names" already exists (Code: 205,130)
```

## Deleting a Project

To delete a project, you need to call HTTP DELETE on the REST interface and specify the project you need to delete, as a query parameter.

There are two ways to specify a project for deletion:

- By ID using `pid=<NN>`
- By name using `pname=<Name>`

To delete a project by ID, do the following:

```
DELETE http://edqserver:8001/edq/config/projects?pid=14
```

To delete a project by name, do the following:

```
DELETE http://edqserver:8001/edq/config/projects?pname=Profile%20Customer%20Names
```

In both cases, the result is a string similar to the following:

```
Project Profile Customer Names deleted
```

If an invalid project is specified then a response of type 406 'Not Acceptable' is returned with an appropriate string message, for example:

```
Bad project ID "14" (Code: 205,454)
```

or

No project named "Profile Customer Names" (Code: 205,453)

## REST Interface for Data Stores

You can query and manipulate data stores in EDQ, using the following interface:

`http://edqserver:8001/edq/config/datasources`

This interface allows you to perform the following tasks:

- [Retrieving a List of Data Stores](#)
- [Creating a Data Store](#)
- [Deleting a data store](#)

### Retrieving a List of Data Stores

To get a list of data stores, call the interface with a valid project name.

For example:

`GET http://edqserver:8001/edq/config/datasources?pid=14`

If successful, an OK response is returned along with a list of data stores in the response body.

For example:

```
[
  {
    "client":false,
    "id":36,
    "name":"Individuals",
    "properties":[
      {
        "name":"quote",
        "value":"\" \"
      },
      {
        "name":"encoding",
        "value":"ISO-8859-1"
      },
      {
        "name":"file",
        "value":"Customer/customerindividuals.csv"
      },
      {
        "name":"cols",
        "value":""
      },
      {
        "name":"project",
        "value":"59"
      },
      {
        "name":"hdr",
        "value":"1"
      },
      {
        "name":"usepr",
```

```

        "value": "0"
      },
      {
        "name": "skip",
        "value": ""
      },
      {
        "name": "sep",
        "value": ", "
      }
    ],
    "species": "servertxt"
  }
]

```

## Creating a Data Store

To create a data store you need to create a JSON object that describes the data store and then POST it to the endpoint specifying the project (by name or id) that will own the data store.

For example:

A server-based .csv file that has been placed in the landing area.

POST <http://edqserver:8001/edq/config/datasources?pid=14>

```

{
  "client": false,
  "name": "Individuals",
  "properties": [
    {
      "name": "quote",
      "value": "\""
    },
    {
      "name": "encoding",
      "value": "ISO-8859-1"
    },
    {
      "name": "file",
      "value": "Customer/customerindividuals.csv"
    },
    {
      "name": "hdr",
      "value": "1"
    },
    {
      "name": "usepr",
      "value": "0"
    },
    {
      "name": "sep",
      "value": ", "
    }
  ],
  "species": "servertxt"
}

```

## A server-based Oracle schema:

POST <http://edqserver:8001/edq/config/datasources?pid=14>

```
{
  "client":false,
  "name":"Staging",
  "properties":[
    {
      "name":"service",
      "value":"sid"
    },
    {
      "name":"sid",
      "value":"orcl"
    },
    {
      "name":"user",
      "value":"staging"
    },
    {
      "name":"port",
      "value":"1521"
    },
    {
      "name":"password",
      "value":"staging"
    },
    {
      "name":"host",
      "value":"localhost"
    }
  ],
  "species":"oracle"
}
```

If successful, an OK response is returned along with the name and ID of the data store, as shown in the following example:

```
{"id":42,"name":"Staging"}
```

The value of the `species` parameter varies depending on the type of data store being used in a project. For example, if you are using the Oracle database, the value of `species` would be "oracle". Each species parameter has its own set of properties.

The following table lists the properties for the species "oracle":

Property	Type	Required	Description
host	String	Yes	The machine hosting the database
port	Number	Yes	Port number of the database
sid	String	Yes	Database Identifier
service	Choice of sid or srv	Yes	Whether the database identifier above is a SID or a SERVICE (srv) name
user	String	Yes	User to log in as
password	String	No	Password for the user
schema	String	No	The schema to use (usually left empty)

The following table lists the properties for the species "servertext":

Property	Type	Required	Description
file	String	Yes	The name and location of the file in the landing area
userpr	Boolean	Yes	Use project specific landing area
hdr	Boolean	Yes	Treat the first line as header
sep	String	No	The field delimiter
quote	Choice	No	Quote character
cols	Integer	No	Number of columns to read
encoding	String	Yes	The character encoding of text
skip	Integer	No	The number of the lines to skip at the start

 **Note:**

For Boolean type use 0 for false and 1 for true.

For quote, the value needs to be a double quote like this "\"", or a single quote like this "'", or an empty value like this "".

For the species, "other", which uses a JDBC connection, the properties are listed in the following table:

Property	Type	Required	Description
driver	String	Yes	The JDBC driver java class
url	String	Yes	The address of the database
user	String	No	The user to log in as
password	String	No	The user's password

## Deleting a data store

To delete a data store call HTTP DELETE on the endpoint by specifying either a data store id or a valid project (by name or id) and a data store name, for example:

```
DELETE http://edqserver:8001/edq/config/datasources?id=42
```

or

```
DELETE http://edqserver:8001/edq/config/datasources?pid=14&name=Staging
```

When the deletion is successful, an OK response is returned without any response body.

## REST Interface for Snapshots

The REST interface for snapshots is:

`http://edqserver:8001/edq/config/snapshots`

It allows you to perform the following tasks:

- [Retrieving a List of Snapshots](#)
- [Creating a Snapshot](#)
- [Deleting a Snapshot](#)

## Retrieving a List of Snapshots

To retrieve a list of snapshots specify a valid project, for example:

```
GET http://edqserver:8001/edq/config/snapshots?pid=14
```

or

```
GET http://edqserver:8001/edq/config/snapshots?pname=Profile%20Customer%20Names
```

If successful, it returns an OK response with a list of snapshots in the response body.

For example:

```
[
  {
    "columns": [
      "TITLE",
      "FULLNAME",
      "GIVENNAMES",
      "FAMILYNAME",
      "NAMETYPE",
      "PRIMARYNAME",
      "ADDRESS1",
      "ADDRESS2",
      "ADDRESS3",
      "ADDRESS4",
      "CITY",
      "STATE",
      "POSTALCODE"
    ],
    "datasource": "Individuals",
    "name": "Individuals",
    "table": "customerindividuals.csv"
  }
]
```

## Creating a Snapshot

To create a snapshot, you need to create a JSON object that describes the snapshot and specify the project where it will be created.

For example:

```
POST http://edqserver:8001/edq/config/snapshots?pid=14
```

```
{
  "name": "Individuals",
  "description": "Customer data",
  "datasource": "Individuals",
  "table": "customerindividuals.csv",
```



```
"columns": [
  "TITLE",
  "FULLNAME",
  "GIVENNAMES",
  "FAMILYNAME",
  "NAMETYPE",
  "PRIMARYNAME",
  "ADDRESS1",
  "ADDRESS2",
  "ADDRESS3",
  "ADDRESS4",
  "CITY",
  "STATE",
  "POSTALCODE"
],
"sampling": {
  "number": 100,
  "offset": 0,
  "ordering": "ascending",
  "count": "true"
}
}
```

If successful, an OK response is returned with the snapshot ID in the response body.

For example:

```
{"id": 68, "name": "Individuals"}
```

## Deleting a Snapshot

To delete a snapshot, you need to specify either a snapshot ID or a valid project and snapshot name.

For example:

```
DELETE http://edqserver:8001/edq/config/snapshots?id=68
```

or

```
DELETE http://edqserver:8001/edq/config/snapshots?pid=14&name=Individuals
```

or

```
DELETE http://edqserver:8001/edq/config/snapshots?pname=Profile%20Customer%20Names&name=Individuals
```

When the specified snapshot is deleted successfully, it returns an OK response with a string message in the response body, which is similar to the following:

```
Snapshot Individuals deleted
```

## REST Interface for Processes

The interface for EDQ processes is:

```
http://edqserver:8001/edq/config/processes
```

Using this interface, you can perform the following tasks:

- [Retrieving a List of Processes](#)
- [Deleting a Process](#)

The following sub-level interface allows you to create a simple profiling process:

```
http://edqserver:8001/edq/config/processes/simpleprocess
```

See [Creating a Simple Process](#) section for details.

## Retrieving a List of Processes

To get a list of processes in a project, you need to call HTTP GET on the processes interface with a valid project name.

Example:

```
GET http://edqserver:8001/edq/config/processes?pid=14
```

or

```
GET http://edqserver:8001/edq/config/processes?pname=Profile%20Customer%20Names
```

An OK response is returned along with a list of processes.

Example:

```
[{"name": "Profile Names", "id": 31}]
```

If the request is not successful, an error response would either be a 404 'Not Found' or 500 'Internal Server Error' along with a string in the response body describing the error.

## Deleting a Process

To delete a process, specify either the process ID, or the project ID or name, and the process name. For example:

```
DELETE http://edqserver:8001/edq/config/processes?id=31
```

or

```
DELETE http://edqserver:8001/edq/config/processes?pid=14&name=Profile%20Names
```

or

```
DELETE http://edqserver:8001/edq/config/processes?pname=Profile%20Customer%20Names&name=Profile%20Names
```

When the deletion is successful, an OK response is returned with a string message and response body, as shown in the following example:

```
Process Profile Names deleted
```

If the deletion is not successful, then either of the following errors along with a response string are returned:

- 404 "Not Found"
- 500 "Internal Server Error"

## Creating a Simple Process

The interface currently only supports creation of simple profiling processes. To create a simple process, you need to create a JSON object that describes the process you want to create and specify the project where it should be created. The HTTP POST operation is used to post this information to the interface.

For example:

```
POST http://edgserver:8001/edg/config/processes/simpleprocess?pid=14
{
  "name": "Profile Names",
  "description": "Profile Individuals Names",
  "reader": {
    "name": "Read from Individuals",
    "stageddata": "Individuals"
  },
  "processors": [
    {
      "name": "Do Quickstats",
      "type": "dn:quickstatsprofiler",
      "columnlist": [
        "GivenNames",
        "FamilyName"
      ]
    },
    {
      "name": "Do Frequency Profiling",
      "type": "dn:attributefrequencycountsprofiler"
    }
  ]
}
```

When the simple process is created successfully, an OK response is returned along with the name and ID of the process in the response body.

For example:

```
{"id":33,"name":"Profile Names"}
```

An error response would be generated in the following cases:

- 404 'Not Found' if the project does not exist
- 400 'Bad Request' if the JSON object is malformed
- 500 'Internal Server Error' if a server error occurs during creation, along with a string in the response body describing the error.

The full list of supported processors is mentioned in the following table:

Processor	Type
Quick Stats Profiler	dn:quickstatsprofiler
Data Types Profiler	dn:datatypesprofiler
Max/Min Profiler	dn:maxandminprofiler
Length Profiler	dn:lengthprofiler

---

Processor	Type
Record Completeness Profiler	dn:recordcompletenessprofiler
Character Profiler	dn:characterprofiler
Frequency Profiler	dn:attributefrequencycountsp profiler
Patterns Profiler	dn:attributepatternsprofiler

---

## REST Interfaces for Jobs

The `/config/jobs` interface performs the following tasks on EDQ jobs:

- [Retrieving a List of Jobs](#)
- [Deleting a Job](#)
- [Creating a Simple Job](#)

The URL for this interface is similar to the following:

```
http://edqserver:8001/edq/config/jobs
```

There is another REST interface, `/jobs` to perform the following tasks:

- [Running a Job](#)
- [Cancelling a Running Job](#)
- [Getting the Status of a Job](#)
- [Getting the Details of All Running Jobs](#)

The URL for this interface is similar to the following:

```
http://edqserver:8001/edq/jobs
```

## Retrieving a List of Jobs

You can get a list of jobs for a project using HTTP GET. You must specify at least one project (there could be more than one project) as the query parameter.

The project can be specified by using the `pid` or `pname` in the query parameter, as shown in the following example:

```
GET http://edqserver:8001/edq/config/jobs?pid=14
```

or

```
GET http://edqserver:8001/edq/config/jobs?pname=Profile%20Customer%20Names
```

The response to this operation would list all the jobs for the specific project or projects, as shown in the following example:

```
[
  {
    "id":99,
    "name":"Profile Names Job"
  },
  {
    "id":98,
```

```

        "name": "Profile Individuals Job"
    }
]

```

## Deleting a Job

To delete a job you need to either specify a valid job ID or a valid project (using one of the query parameters) and a valid job name.

To delete a job by job ID:

```
DELETE http://edqserver:8001/edq/config/jobs?id=99
```

To delete a job by job name:

```
DELETE http://edqserver:8001/edq/config/jobs?pid=14&name=Profile%20Names%20Job
```

or

```
DELETE http://edqserver:8001/edq/config/jobs?pname=Profile%20Customer%20Names&name=Profile%20Names%20Job
```

When the job is deleted successfully, a message appears to confirm the deletion:

```
Job Profile Names Job deleted
```

## Creating a Simple Job

A simple job has a single phase and contains a single process. To create a simple job, you need to specify a valid project that owns the job (using one of the query parameters). Also, you need to create a JSON object describing the job to create.

For example:

```
POST http://edqserver:8001/edq/config/jobs/simplejob?pid=14
{
  "name" : "Profile Names Job" ,
  "process" : "Profile Names" ,
  "description" : "Profile Customer Names"
  "resultsdrilldown" : "none"
}
```

The attribute `resultsdrilldown` can have any of the following values: `none`, `sample`, `limited`, `all`. However, the `sample` and `limited` values have the same implication.

## Running a Job

With the `/jobs/run` interface you can run a named job using HTTP POST. The required parameters are the project name or project ID and the job name. Optionally, you can specify run label and overrides.

The following example shows the URL and associated payload used with running a job named "Real-time Start All" for the project "Profile Customer Name":

```
POST http://edqserver:8001/edq/jobs/run
{
  "project": "Profile Customer Name",
```

```
"job": "Real-time Start All",
"runlabel": "uk",
"runprofile": "profiling",
"overrides": [
  {
    "name": "a",
    "value": "b"
  },
  {
    "name": "c",
    "value": "d"
  }
]
```

The JSON response to this request would be similar to the following:

```
{
  "executionID": 2,
  "runeverywhere": false
}
```

In this example, the job "Real-time Start All" returns the "runeverywhere" value as false, which implies that this job can run only in one place. In such cases, an executionID is returned for the job. This executionID can be used to cancel a job and query a job's status.

However, if the value of "runeverywhere" were true, then only the "jobtype" would be returned in the JSON response. For "runeverywhere" jobs, cancel and query calls are not supported.

## Cancelling a Running Job

To cancel a running job, the HTTP POST operation is used. The interface URL is similar to the following:

```
POST http://edqserver:8001/edq/jobs/cancel
```

You only need the executionID of the job to cancel it. The following example illustrates cancelling a job with the executionID 12.

```
{
  "executionID": 12,
  "type" : "immediate"
}
```

The following options are available with the "type" parameter:

- **immediate:** This option cancels the job as early as possible.
- **keepresults:** This option cancels the job but retains the results that have been generated so far.
- **shutdown:** This options is used to cancel or shutdown a job that runs a web service.

No response is returned when a job is cancelled successfully.

## Getting the Status of a Job

To get the status of an individual job, the `executionID` is passed in the URL. The URL looks similar to the following:

```
GET http://edqserver:8001/edq/jobs/status?xid=executionID
```

For example, if the execution ID of the job "Real-time Start All" is 14, the URL would be:

```
GET http://edqserver:8001/edq/jobs/status?xid=14
```

The JSON response would be as follows:

```
{
  "executionid": 14,
  "project": "Profile Customer Name",
  "job": "Real-time START ALL",
  "server": "edqserver",
  "starttime": "2016-04-20T08:44:48+01:00",
  "endtime": "2016-04-20T08:45:22+01:00",
  "complete": true,
  "status": "finished"}
```

In this example, the "Real-time Start All" job triggers other jobs. Once all the jobs in the project are triggered, the status of the `executionID` 14 shows finished. However, the jobs that are triggered by the "Real-time Start All" job, may still show the status as running.

## Getting the Details of All Running Jobs

The status of all running jobs in a project can be retrieved using the `/jobs/running` interface, which would be represented by a URL similar to the following:

```
GET http://edqserver:8001/edq/jobs/running
```

The following example shows the output in JSON format:

```
{
  "executionid": 4,
  "project": "Profile Customer Name",
  "job": "Real-time Individual Clean",
  "server": "edqserver",
  "starttime": "2016-04-19T10:05:30.74+01:00",
  "endtime": "2016-04-19T16:05:41+01:00",
  "complete": false,
  "status": "running"
}
```

Optionally, you can provide other query parameters such as project name, job name, and run label. For jobs without a run label, omit the run label parameter and set it to empty filter jobs with no run label. The URL in this case would be similar to the following:

```
/jobs/running?[project=project[&job=job][&runlabel=]
```

## REST Interface for Reference Data

Reference data can exist within a project or outside of all projects at system level. To refer to reference data at system level, specify the project ID as 0 or `pid=0`.

The interface for reference data is:

```
http://edqserver:8001/edq/config/referencedata
```

The interface for reference data content is:

```
http://edqserver:8001/edq/config/referencedata/contents
```

You can use this interface to perform the following tasks:

- [Retrieving a List of Reference Data](#)
- [Retrieving Contents of Reference Data](#)
- [Creating Reference Data](#)
- [Deleting Reference Data](#)

### Retrieving a List of Reference Data

To get a list of all reference data defined at system level specify the project using either of the following parameters:

By `pid = 0`:

```
GET http://edqserver:8001/edq/config/referencedata?pid=0
```

By `pname`:

```
GET http://edqserver:8001/edq/config/referencedata?pname=Profile%20Customer%20Names
```

A list of JSON objects, which represent reference data, are returned. The output looks similar to the following:

```
[
  {
    "activerows":2,
    "category":"charactertokeymap",
    "columns":[
      {
        "key":true,
        "name":"Name",
        "type":"STRING",
        "unique":true
      },
      {
        "name":"Value",
        "type":"STRING",
        "value":true
      }
    ],
    "id":39,
    "name":"Tokens",
    "totalrows":2
  }
]
```



```
    }
  ]
}
```

## Retrieving Contents of Reference Data

To list the contents of the reference data, the reference data contents interface is used. You must specify a valid project or use `pid=0` for system level.

For example:

```
GET http://edqserver:8001/edq/config/referencedata/contents?pid=0&id=40
```

or

```
GET http://edqserver:8001/edq/config/referencedata/contents?
pid=0&name=ShortNameMap
```

This returns information about the reference data rows, as shown in the following code snippet:

```
{
  "activerows":4,
  "columns":[
    {
      "key":true,
      "name":"ShortName",
      "type":"STRING",
      "unique":true,
      "value":true
    },
    {
      "key":true,
      "name":"LongName",
      "type":"STRING",
      "value":true
    }
  ],
  "description":"Map short names to long names",
  "id":43,
  "name":"ShortNameMap",
  "rows":[
    {
      "data":[
        "Jeff",
        "Jeffrey"
      ]
    },
    {
      "data":[
        "Jon",
        "Jonathan"
      ]
    }
  ],
  "totalrows":4
}
```

## Creating Reference Data

To create reference data you need to create a JSON object describing the reference data, which you will post to the interface specifying either `pid=0` for system level, or a valid project name.

For example:

```
POST
http://edqserver:8001/edq/config/referencedata?pname=Profile%20Customer%20Names

{
  "name" : "ShortNameMap",
  "description" : "Map short names to long names",
  "columns":
  [
    {
      "key": true,
      "name": "ShortName",
      "type": "STRING",
      "unique": true,
      "value": true
    },
    {
      "key": true,
      "name": "LongName",
      "type": "STRING",
      "value": true
    }
  ],
  "rows":
  [
    {
      "data":
      [
        "Jeff",
        "Jeffrey"
      ]
    },
    {
      "data":
      [
        "Jon",
        "Jonathan"
      ]
    }
  ]
}
```

On successful creation, a response similar to the following is returned:

```
{ "id": 40, "name": "ShortNameMap" }
```

## Deleting Reference Data

To delete reference data, you need to call HTTP DELETE on the reference data interface, with either a valid reference data ID or a valid project (including `pid=0` for system level) and a valid reference data name.

To delete by reference data ID:

```
DELETE http://edqserver:8001/edq/config/referencedata?id=40
```

To delete by reference data name:

```
DELETE http://edqserver:8001/edq/config/referencedata?pid=0&name=ShortNameMap
```

After a successful deletion, the response returns a string message, such as the following:

```
Reference data ShortNameMap deleted
```

## REST Interface for Web Services

The interface for web services is:

```
http://edqserver:8001/edq/config/webservices
```

It allows you to perform the following tasks when you call the respective get, post, or delete operations:

- [Retrieving a List of Web Services](#)
- [Creating or Updating a Web Service](#)
- [Deleting a Web Service](#)

### Retrieving a List of Web Services

You can get a list of web services defined for a valid project by using the HTTP GET operation on the web services interface. To get a list of web services, specify a valid project using either the `pid` or `pname` parameter.

For example:

```
GET http://edqserver:8001/edq/config/webservices?pid=14&pid=20
```

A successful call returns a list of web services, and their input and output interfaces, in the response body:

```
[
  {
    "id":1,
    "inputs":{
      "attributes":[
        {
          "name":"Name",
          "type":"STRING"
        }
      ],
      "multirecord":false
    },
    "name":"Long Names",
    "outputs":{
      "attributes":[
        {
          "name":"LongName",
          "type":"STRING"
        }
      ],

```

```

        "multirecord":false
    }
}
]

```

## Creating or Updating a Web Service

You can create and update web services by creating an appropriate JSON object, which you then POST to the web services interface.

To create a web service you need to specify a valid project (by name or id), as shown in the following example:

POST <http://edqserver:8001/edq/config/webservices?pid=14>

```

{
  "name":"Name Gender",
  "inputs":{
    "attributes":[
      {
        "name":"Name",
        "type":"STRING"
      }
    ],
    "multirecord":false
  },
  "outputs":{
    "attributes":[
      {
        "name":"Gender",
        "type":"STRING"
      }
    ],
    "multirecord":false
  }
}

```

If successful, the name and ID of the web service is returned in the response body.

**Example:**

```
{ "id":4, "name":"Name Gender" }
```

To update a web service, you need a JSON object that is identical in structure, but with an additional ID attribute to identify the existing web service. For an update you do not specify a project.

**Example:**

POST <http://edqserver:8001/edq/config/webservices>

```

{
  "id":4,
  "name":"Name Gender",
  "inputs":{
    "attributes":[
      {
        "name":"First Name",
        "type":"STRING"
      }
    ],
    {

```

```

        "name": "Last Name",
        "type": "STRING"
    }
  ],
  "multirecord": false
},
"outputs": {
  "attributes": [
    {
      "name": "Gender",
      "type": "STRING"
    }
  ],
  "multirecord": false
}
}

```

If successful, the name and ID of the web service is returned in the response body, as shown in the following example:

```
{ "id": 4, "name": "Name Gender" }
```

## Deleting a Web Service

To delete a web service call HTTP DELETE on the web service interface. Specify either the web service ID or a valid project (by name or ID) and the web service name.

For example:

```
DELETE http://edqserver:8001/edq/config/webservices?id=5
```

or

```
DELETE http://edqserver:8001/edq/config/webservices?pid=14&name=Name%20Gender
```

If successful, an OK response is returned but without a response body.

## Example: Profiling from an External Application

Consider a scenario where an external application needs to profile data in a table in an Oracle database, using EDQ. In such a case, you can programmatically profile this table using the REST-based APIs. For this example, a CUSTOMERS table in a CustomerDB database will be used.

To generate and run the profiling job on the CUSTOMERS table, the following tasks are performed:

1. Create a project by using the following URL:

```
POST http://edqserver:8001/edq/config/projects
```

The project name (`pname`) is "Profile Customer". The JSON code is:

```

{
  "name": "Profile Customer",
  "description": "Profiling customers in the CUSTOMERS table"
}

```

2. Create a data store using an Oracle database, CustomerDB, by using the following URL:

POST <http://edqserver:8001/edq/config/datasources?pid=4>

An example JSON code to create the data store is:

```
{
  "client":false,
  "name":"CustomersDB",
  "properties":[
    {
      "name":"service",
      "value":"sid"
    },
    {
      "name":"sid",
      "value":"orcl"
    },
    {
      "name":"user",
      "value":"CRM"
    },
    {
      "name":"port",
      "value":"1521"
    },
    {
      "name":"password",
      "value":"welcome123"
    },
    {
      "name":"host",
      "value":"localhost"
    }
  ],
  "species":"oracle"
}
```

 **Note:**

To determine the `pid` or the project ID for the project "Profile Customer", use the HTTP GET operation with the URL:

GET <http://edqserver:8001/edq/config/projects>

3. Create a snapshot by using the following URL:

POST <http://edqserver:8001/edq/snapshots?pid=4>

The JSON code for creating the snapshot is:

```
{
  "name":"CustomersDB.Customers",
  "description":"Customer details",
  "datasource":"CustomersDB",
  "table":"Customers",
  "columns":[
    "ID",
    "FULLNAME",
  ]
}
```

```

        "GIVENNAME",
        "FAMILYNAME",
        "Street",
        "City",
        "State",
        "PostalCode",
        "State",
        "Phone",
        "Cell",
        "Work",
        "eMail",
        "DoB",
        "Gender",
        "Active",
        "CreditLimit",
        "StartDate",
        "EndDate"
    ],
    "sampling":{
        "number":100,
        "offset":0,
        "ordering":"ascending",
        "count":"true"
    }
}

```

The result for this is displayed as:

```

{
  "id": 85,
  "name": "CustomersDB.Customers"
}

```

The snapshot with the name "CustomersDB.Customers" is created.

#### 4. Create a simple process by using the following URL:

POST <http://edqserver:8001/edq/config/processes/simpleprocess?pid=4>

For this example, a simple process is created with a `Quickstats Profiler` and a `Frequency Profiler`, both profiling only the Name fields. This can be done using the following example JSON.

```

{
  "name":"Profile Names",
  "description":"Profile Customer Names",
  "reader":{
    "name":"Read from Customers",
    "stageddata":"Connection to Customers"
  },
  "processors":[
    {
      "name":"Do Quickstats",
      "type":"dn:quickstatsprofiler",
      "columnlist":[
        "GIVENNAME",
        "FAMILYNAME"
      ]
    },
    {
      "name":"Do Frequency Profiling",
      "type":"dn:attributefrequencycountsprofiler"
    }
  ]
}

```

```
    }  
  ]  
}
```

The response to this request is:

```
{  
  "id": 267,  
  "name": "Profile Names"  
}
```

**5. Create a simple job by using the following URL:**

POST <http://edqserver:8001/edq/jobs/simplejob?pid=4>

```
{  
  "name": "Profile Customer Job",  
  "process": "Profile Names",  
  "description": "Profiling Customer Names",  
  "resultsdrilldown": "none"  
}
```

The response to the request is:

```
{  
  "id": 211,  
  "name": "Profile Customer Job"  
}
```

**6. Run the job using the following URL:**

POST <http://edqserver:8001/edq/jobs/run>

The JSON code for running the job "Profile Customer Job", which in turn would run the profiling process, is:

```
{  
  "project": "Profile Customer",  
  "job": "Profile Customer Job"  
}
```

The response is:

```
{  
  "executionID": 20,  
  "runeverywhere": false  
}
```

Once this job is running, you can check the status of this execution of the job using the following URL:

GET <http://edqserver:8001/edq/jobs/status?xid=20>

Running this URL displays the status of the job, as shown in the following code:

```
{  
  "complete": true,  
  "endtime": "2016-04-29T14:05:41+01:00",  
  "executionid": 1,  
  "job": "Profile Customer Job",  
  "project": "Profile Customer",  
  "server": "edq_server1",  
  "starttime": "2016-04-29T14:05:38+01:00",  
  "status": "finished"  
}
```



If required, you can cancel the job using the following URL:

```
POST http://edqserver:8001/edq/jobs/cancel
```

The JSON code to cancel a job is:

```
{  
  "executionID": 12345,  
  "type" : "immediate"  
}
```

This would cancel the job instantly, without saving the results. For other options that can be used with "type", see [Cancelling a Running Job](#).

To log in to EDQ Director to view the results of a profiling job that was executed successfully, use the following URL:

```
http://edqserver:8001/edq/blueprints/director/jnlp?  
projectid=1&processid=1&processornum=2
```

The `projectid` and `processid` are the same that are generated using the corresponding REST API calls and the `processornum` value is set to 2, which is the first processor after the reader.

This URL opens the Director UI with the focus on the first profiling processor in the job so that its results can be viewed immediately.

An external application may include an option to remove generated jobs, which would execute calls to the relevant deletion calls. The simplest version of this is to delete the whole project. For details on deleting a project, see [Deleting a Project](#).

# 9

## Using the Java Messaging Service (JMS) with EDQ

This document describes how to get started using Java Messaging Service (JMS) technology with Oracle Enterprise Data Quality (EDQ). This documentation is intended for system administrators responsible for installing and maintaining EDQ applications.

This chapter includes the following sections:

- [Understanding the Message Queue Architecture](#)
- [Uses of JMS with EDQ](#)
- [Configuring EDQ to Read and Write JMS Messages](#)
- [Defining the Interface Files](#)
- [Illustrations](#)

### Understanding the Message Queue Architecture

JMS is a flexible technology that allows EDQ to integrate with a variety of different messaging systems and technologies, including WebLogic Server Messaging, WebSphere MQ, Oracle Advanced Queueing (OAQ), and many more.

For simpler architectures, and to allow customers to get started using JMS, EDQ is shipped with ActiveMQ (<http://activemq.apache.org/>) embedded. This can be enabled on an EDQ server simply by adding the following line to `director.properties`:

```
launch.activemq=1
```

Depending on the messaging architecture, an EDQ server may be acting as a 'client' of a message queue, or as a 'server'. Most commonly, EDQ acts as client. Where EDQ acts as a client, it may need additional client jars to be installed; this is the case if the server which EDQ is reading messages from and writing messages to uses a different message queueing technology from the application server EDQ is deployed on. For example, if EDQ is deployed on WebLogic or Tomcat but needs to act as a client to a server running WebSphere MQ, it will need additional client jars to be installed. In this case, the administrator of the MQ system should be consulted to determine which files are needed for a Java application to act as a client.

### Uses of JMS with EDQ

There are two main uses of JMS with EDQ, as listed below: We will focus here on the first use of JMS with EDQ.

- **Consuming and Providing Messages:** You can configure EDQ to read messages from JMS queues, and write messages to them. This can be beneficial where several EDQ servers need to read from a single stream of messages that needs to be persistent to ensure no loss of messages. In this mode, each EDQ

server will consume messages from a queue and only 'commit' when it has finished processing each message. Note that JMS is best used for Asynchronous communication, where EDQ is not expected to return a response to a calling application for a specific message.

- **Using JMS in Triggers:** EDQ may send JMS messages to other systems, or may consume JMS messages to start triggers (for example to use a JMS queue to distribute batch jobs amongst several EDQ servers;). For more details, see the Using JMS in Triggers.

## Configuring EDQ to Read and Write JMS Messages

JMS interfaces to and from EDQ are configured using XML interface files that define:

- The path to the queue of messages
- Properties that define how to work with the specific JMS technology
- How to decode the message payload into a format understood by an EDQ process (for Message Providers – where EDQ reads messages from a queue), or convert messages to a format expected by an external process (for Message Consumers – where EDQ writes messages to a queue).

The XML files are located in the EDQ Local Home directory (formerly known as the config directory), in the following paths:

- buckets/realtime/providers (for interfaces 'in' to EDQ)
- buckets/realtime/consumers (for interfaces 'out' from EDQ)

Once the XML files have been configured, Message Provider interfaces are available in Reader processors in EDQ and to map to Data Interfaces as 'inputs' to a process, and Message Consumer interfaces are available in Writer processors, and to map from Data Interfaces as 'outputs' from a process as shown in Figures 1 and 2 below:

## Defining the Interface Files

An interface file in EDQ consists of three sections, as follows:

- The `<attributes>` section, defining the shape of the interface as understood by EDQ
- The `<messengerconfig>` section, defining how to connect to the JMS queue or topic
- The `<messagebody>` section, defining how to extract contents of a message (e.g. from XML) into attribute data readable by EDQ (for inbound interfaces), or how to convert attribute data from EDQ to message data (e.g. in XML).

## Knowing the `<attributes>` section

The `<attributes>` section defines the shape of the interface. It constitutes the attributes that are available in EDQ when configuring a Reader or Writer. For example the

following excerpt from the beginning of an interface file configures three string attributes that can be used in EDQ, and their names are:

```
<?xml version="1.0" encoding="UTF-8"?>
<realtimedata messenger="jms">
  <attributes>
    <attribute type="string" name="messageID"/>
    <attribute type="string" name="name"/>
    <attribute type="string" name="AccountNumber"/>
  </attributes>
```

[file continues]...

EDQ supports all the standard attribute types and they are:

- string
- number
- date
- stringarray
- numberarray
- datearray

## Knowing the <messengerconfig> Section

The <messengerconfig> section of the interface file configures the settings needed to connect to a given JMS queue or topic on a particular type of JMS technology. The text in <messengerconfig> is parsed as a Java properties file and is merged with the set of properties from the `realtime.properties` file in the EDQ configuration path, with settings in the <messengerconfig> section overriding any matching settings in `realtime.properties`. The `realtime.properties` file therefore allows the configuration of a number of global properties that do not need to be stated in every JMS interface file.

### Note:

- As with all property files stored in the EDQ Local Home directory, properties are themselves merged with a set of 'base' properties in the EDQ Home directory which should not be modified, with any property stated in a file in the Local Home directory used in preference to any property in the EDQ Home directory. For example the `realtime.properties` file in the EDQ Local Home directory will be merged with the file of the same name in the EDQ Home directory to form the final set of properties used.
- The `realtime.properties` values may also be used to set properties for Web Services in EDQ. A prefix of `jms.` is therefore used (in this file only, not in the <messengerconfig> section of a JMS interface file) to set properties for JMS (`ws.` is used for Web Services).

The messenger configuration properties, specified either in the `<messengerconfig>` section of a JMS interface file, or inherited from `realtime.properties`, are used:

- To configure a JNDI name store to look up queues
- To specify the name of the queue/topic and connection factory
- To supply authentication information to the MQ server, if required

### JNDI Setup Properties

- `java.naming.factory.initial` - It denotes the class name of the JNDI initial context factory, used to bootstrap JNDI naming
- `java.naming.provider.url` - the JNDI URL used to connect to the JNDI server or local store

In some cases, authentication is required to connect to the JNDI service. In this case the following properties must be added:

- `java.naming.security.principal` - JNDI user name
- `java.naming.security.credentials` - JNDI password

These JNDI properties are a standard Java feature and for more details, see [JNDI Documentation](#).

### JMS Names

The other properties are:

- `connectionfactory` - the JNDI name of the JMS 'connection factory'. The default for this (if the property is not set) is 'ConnectionFactory' which is correct for several MQ servers.
- `destination` - the JNDI name of the JMS queue or topic. There is no default for this and it is always required.

### Authentication

If the MQ server requires authentication when making connections, add the below properties:

```
username: connection user name
password: connection password
```



#### Note:

These properties are used to authenticate against the MQ server and are used to connect to JNDI. In rare cases, both sets may be required.

### Notes for specific MQ servers

1. ActiveMQ -

For an installation of EDQ that is not on WebLogic, the installed `realtime.properties` file in the EDQ Home directory contains the following settings, suitable for use with the local embedded ActiveMQ server:

```
jms.java.naming.factory.initial=org.apache.activemq.jndi.ActiveMQInitial  
ContextFactory  
jms.java.naming.provider.url=vm://localhost?create=false
```

Therefore, if the local embedded ActiveMQ server is used, the `<messengerconfig>` section of a JMS interface file requires only a destination property setting.

To connect to a remote ActiveMQ server, use a property setting as listed below:

This can be set in the `<messengerconfig>` section to apply to a single interface, or can be set in `realtime.properties` (prefixed with `jms.`) to set the default for all interfaces.

```
java.naming.provider.url=tcp://host:port
```

Remote ActiveMQ servers generally require connection authentication, so the username and password properties will be required. The embedded ActiveMQ server uses EDQ user authentication, so the user account used needs the 'Connect to Messaging System' permission.

To use ActiveMQ on an EDQ server installed on WebLogic, the property settings above need to be entered into the `realtime.properties` in the EDQ Local Home directory. (The settings are present, but commented out, in the file in the Home directory, as it is assumed that WebLogic JMS will normally be used where EDQ is installed on WebLogic.)

## 2. WebLogic JMS -

When running in a full JavaEE application server, such as WebLogic, Glassfish or WebSphere, the JNDI settings for local JMS are configured automatically and the JNDI factory and url information do not need to be specified in any EDQ configuration file.

The recommended approach in these cases is to define the JMS configuration within the application server and then just specify the JNDI connection factory and destination names in `<messengerconfig>`.

WebLogic supports the concept of 'foreign' JMS servers. In this case, you define the connection information in the WebLogic Console and expose the destination(s) and connection factory as names in the native WebLogic JNDI store. This works well with Oracle Advanced Queueing (AQ). See below for a snippet of the configuration of a 'foreign JMS server' pointing at an AQ schema.

When JMS is configured like this, a typical `<messengerconfig>` section might be:

```
<messengerconfig>  
  connectionfactory = jms/cf1  
  destination       = jms/queue1  
</messengerconfig>
```

Here `jms/cf1` and `jms/queue1` are JNDI names defined in WebLogic.

## Knowing the <messagebody> section

This section uses JavaScript to parse message payloads into attributes that EDQ can use for inbound interfaces, and perform the reverse operation (convert EDQ attribute data into message payload data) for outbound interfaces. A function named 'extract' is used to extract data from XML into attribute data for inbound interfaces, and a function named 'build' is used to build XML data from attribute data.

For more details, refer to [Illustrations](#), wherein the scripts are best illustrated using examples.

## Illustrations

### Example 1 – Simple Provider File

The following XML is a simple example of a provider interface file, reading messages from a queue in the path 'dynamicQueues/InputQueue'.

```
<?xml version="1.0" encoding="UTF-8"?>
<realtimedata messenger="jms">
  <attributes>
    <attribute type="string" name="messageID"/>
    <attribute type="string" name="name"/>
    <attribute type="string" name="AccountNumber"/>
    <attribute type="string" name="AccountName"/>
    <attribute type="string" name="Country"/>
  </attributes>
</messengerconfig> destination = dynamicQueues/InputQueue </messengerconfig>
  <incoming>
<messagebody>
<script>
<![CDATA[ function extract(str) { var screeningRequest = new XML(str); var
rec = new Record(); rec.messageID = screeningRequest.individual.messageID;
rec.name = screeningRequest.individual.name; rec.AccountNumber =
screeningRequest.individual.AccountNumber; rec.AccountName =
screeningRequest.individual.AccountName; rec.Country =
screeningRequest.individual.Country; return [rec]; } ]]>
  </script>
  </messagebody>
</eof>
<messageheader name="JMSType" value="tleof"/>
</eof>
  </incoming>
</realtimedata>
```

### Example 2 – Simple Consumer File

The following XML is a simple example of a consumer file representing similar data to the provide file listed above, but with additional attributes, which can be added by an EDQ process:

```
<?xml version="1.0" encoding="UTF-8"?>
<realtimedata messenger="jms">
```

```

    <attributes>
    <attribute type="string" name="messageID"/>
    <attribute type="string" name="AccountNumber"/>
    <attribute type="string" name="AccountName"/>
    <attribute type="string" name="Country"/>
    <attribute type="string" name="AccountType"/>
    </attributes>
    <messengerconfig destination = dynamicQueues/OutputQueue </
messengerconfig>
    <outgoing>
    <messagebody>
    <script>
    <![CDATA[ function build(recs) { var rec = recs[0]; var xml = <Request>
    <individual> <messageID>{checkNull(rec.messageID)}</messageID>
    <AccountNumber>{checkNull(rec.AccountNumber)}</AccountNumber>
    <AccountName>{checkNull(rec.AccountName)}</AccountName>
    <Country>{checkNull(rec.Country)}</Country>
    <AccountType>{checkNull(rec.AccountType)}</AccountType> </individual> </
Request>; return xml.toXMLString(); } function checkNull(value) { return
value != null ? value : "" ; } ]]>
        </script>
    </messagebody>
    </outgoing>
    </realtimedata>

```

### Example 3 – Date Parsing and Formatting

The following example file shows how to define a DATE attribute type, and include date parsing and formatting when decoding the message payload:

```

<?xml version="1.0" encoding="UTF-8"?>
<realtimedata messenger="jms">
<attributes>
<attribute type="date" name="processingDate"/>
</attributes>
<messengerconfig destination = dynamicQueues/HoldQueue </messengerconfig>
<incoming>
<messagebody>
    <script>
    <![CDATA[ var df = Formatter.newDateFormatter("yyyy-MM-
dd'T'HH:mm:ss.SSSZ"); function extract(str) { var screeningRequest = new
XML(str); var rec = new Record();rec.processingDate =
date(screeningRequest.processingDate; return [rec]; } function date(x)
{ var s = x.text().toString(); return s == "" ? null : df.parse(s); } ]]>
        </script>
    </messagebody>
    <eof>
<messageheader name="JMSType" value="tleof"/>
    </eof>
</incoming>
</realtimedata>

```