

**ORACLE®**

**ENTERPRISE DATA  
QUALITY**

# **Oracle Enterprise Data Quality 12.2.1.1 New Features Overview**

Integrated Profiling, New Data Services, New Processors

ORACLE WHITE PAPER | JULY 2016



**ORACLE®**



## Table of Contents

Executive Overview	1
Oracle Enterprise Data Quality 12.2.1.1 Release	1
<b>Integrated Profiling and new Configuration API</b>	1
<b>Extended Array Support</b>	1
<b>Additional Customer Data Services and Processors</b>	1
<b>More Extended Attributes in Case Management</b>	2
<b>Processor Security</b>	2
<b>Web Service Monitoring Improvements</b>	3
Conclusion	3

## Executive Overview

Oracle Enterprise Data Quality (EDQ) is Oracle's strategic data quality management platform, used to understand, improve, protect and govern data quality throughout the enterprise.

Oracle Enterprise Data Quality is pre-integrated with a range of Oracle Applications and Technology, including Oracle Siebel Applications, Oracle Customer Hub, Oracle Customer Data Management, Oracle Sales Cloud, Oracle Supplier Cloud, Oracle Data Integrator, Oracle Data as a Service, WebLogic Server and the Oracle Database.

EDQ 12.2.1.1 is the latest generation of the software, enabling integrated profiling, an improved REST API, new out-of-the-box data services, and a range of new processors to improve productivity when working with hierarchical and other multi-value data sets. This whitepaper describes in detail the key new features of the release.

## Oracle Enterprise Data Quality 12.2.1.1 Release

### Integrated Profiling and new Configuration API

EDQ 12.2.1.1 includes a number of new configuration APIs, available as REST web services. The new APIs allow external applications that manage data to generate, run and link to the results of data profiling jobs on their data, so providing integrated profiling capabilities powered by EDQ.

The API can also be used for other integrations with EDQ, for example facilitating external 'gets' of EDQ web service interfaces and reference data sets and contents. The API includes a new interface for querying, running, monitoring and cancelling EDQ jobs over REST web services.

For more details, see [Using the EDQ Configuration API in the Integration Guide](#).

### Extended Array Support

EDQ 12.2.1.1 includes a range of new processors, and extensions to many existing processors, to work with array data, to facilitate improved productivity when working with normalized or hierarchical data sets. Support includes:

- » **Arrays in match.** All match comparisons now support matching using array values, comparing all elements in an array with a string, or with all elements in another array. This makes matching more efficient and avoids 'record explosion' when matching on records that have multiple secondary identifiers, for example when matching people with multiple email addresses, phone numbers, or an ambiguous date of birth, or when matching companies with multiple trading names.
- » **17 new array processors.** 17 new processors are provided to work with Array data, ranging from simple processors to add, concatenate, sort, trim and de-duplicate array attributes, to more complex processors such as Cross Array Element Update, which allows an array of flags to be used to drive transformations to an array of values, for example to scrub all values in an array that fail a check.
- » **Extensions to 20 existing processors.** Many existing processors, such as List Check, Regex Check, Value Check, Regex Replace and many others have been extended to allow them to work with arrays and multiple input attributes. Output flags now indicate check (or transform) success or failure for each value being checked. A new option on each processor drives overall success or failure on the record depending on if any value has been validated or transformed, or if all values have. The Convert processors now work with multiple attributes, for example to convert many strings to numbers in a single operation.

### Additional Customer Data Services and Processors

The EDQ Customer Data Services Pack now includes four additional out-of-the-box data quality web services:

- » **Individual Name.** Provides individual name parsing, standardization, transliteration and validation. Derives gender from name.

- » **Entity Name.** Provides entity name parsing, standardization, transliteration and validation.
- » **Email.** Provides email address parsing, standardization and format validation.
- » **Country.** Provides country and nationality standardization and validation.

The web services have a number of parameters that drive their functionality. Parameters are set by specifying their values in message headers when sending in requests. See the [Customer Data Services documentation](#) for more information.

These new services are powered by reusable tools (published processors) that can be used in any EDQ project, for example to provide standard approaches to gender derivation, individual name transliteration, entity name standardization, country name standardization, and so on.

### More Extended Attributes in Case Management

It is now possible to configure up to 32 extended attributes in Case Management; double the previous limit of 16. Extended attributes are configured in [EDQ local home]/casemanagement/flags.xml. All extended attribute functionality, such as the ability to define valid formats for data entry or constrain values using a picklist or boolean control, is available for the 32 attributes.

### Processor Security

EDQ 12.2.1.1 now allows optional use of security controls which can be used to increase the security level of EDQ processors, for example to block insecure uses of the Script processor. For backward compatibility reasons processor security is off by default, but it can be enabled by adding the following line to [EDQ Local Home]/director.properties:

```
processor.security= off/low/medium/high
```

The processor security option acts in conjunction with the use of a Java Security Manager. The use of a Java Security Manager is controlled by a Java option specified on the server (-Djava.security.manager); this is enabled by default for new installations of EDQ on WebLogic server as it is specified in the setStartupEnv.sh script that sets the Java options for the EDQ server startup group. In other installations it must be manually specified.

The security level of each of the different `processor.security` settings is summarized below:

#### Off

No security restrictions are applied.

#### Low

The following restrictions are applied:

1. The use of the Script processor from the tool palette, for direct use in processes, is disabled if the system is not running with a Java Security Manager.
2. If the system is running with a Java Security Manager, the Script processor is available but is only granted a very small set of default permissions, limited to data processing. Scripts will not be able to make network connections or issue commands outside of the application.

Note that Java processors and scripts that are packaged in jars will run without any restrictions; if a jar contains a **permissions** element the processor will be granted only those permissions.

#### Medium

The same restrictions as Low apply, except that Java processors and scripts that are packaged in jars will be granted a very limited set of permissions. Any additional permissions required by the processor must be listed in

**permissions** elements. To clarify the difference, in 'low' level, all permissions are granted and **permissions** elements **restrict** permissions; in 'medium' level, permissions are limited and **permissions** elements **extend** permissions. In both levels a processor with a **permissions** element will run with exactly the same set of permissions.

## High

The same restrictions as 'medium' apply, except that only processors that are signed by the Oracle EDQ certificate, or by certificates granted to approved partners, will be allowed to include **permissions** elements that grant additional permissions to processors. If a **permissions** element is found in an unsigned jar the processors in the jar will be rejected and will not appear in the processor palette.

## Web Service Monitoring Improvements

EDQ 12.2.1.1 includes new mbeans that can be used for more detailed monitoring of web services. Each web service that exists on the server can be monitored using any JMX client such as Oracle Enterprise Manager Fusion Middleware Control (if installed), a simpler tool such as JConsole, or an external script. Browse to edq/Web Services in the System MBean Browser to find a list of web services. Each web service can be monitored across a range of properties and metrics such as the current number of synchronous requests, the number of messages processed since the web service opened, the last time a message was received, and so on.

Name	Description	Access	Value
1 closeTime	Time when web service was last closed	R	(No value)
2 closeTimeStamp	Time stamp when web service was last closed	R	0
3 concurrent	The current number of synchronous requests	R	0
4 maxConcurrent	The maximum number of concurrent synchronous requests since web service was open...	R	0
5 maxConcurrentMax	The maximum number of concurrent synchronous requests since startup	R	0
6 messages	Number of messages processed since web service was opened	R	0
7 name	Web Service Name	R	AddressClean
8 open	Open/close state of web service	R	0
9 openCount	Number of times web service has been opened since startup	R	0
10 openTime	Time when web service was last opened	R	(No value)
11 openTimeStamp	Time stamp when web service was last opened	R	0
12 processingTime	Processing time for messages since the web services was opene	R	0

Figure 1 - Example of monitoring an EDQ Web Service in Enterprise Manager (Fusion Middleware Control)

## Conclusion

Oracle Enterprise Data Quality (EDQ) 12.2.1.1 is another significant step forward in the evolution of data quality software, providing much improved productivity when working with hierarchical data, even faster time-to-value with customer data, enhanced integration and data stewardship flexibility, and new possibilities for external monitoring and job control.

Oracle continues to set the standard in productivity when understanding, protecting, improving and governing data quality in the enterprise.



**Oracle Corporation, World Headquarters**

500 Oracle Parkway  
Redwood Shores, CA 94065, USA

**Worldwide Inquiries**

Phone: +1.650.506.7000  
Fax: +1.650.506.7200

---

**Integrated Cloud Applications & Platform Services**

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

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. 0615

Oracle Enterprise Data Quality 12.2.1.1 New Features Overview  
July 2016  
Author: Mike Matthews  
Contributing Authors: Richard Evans

