

Oracle® FMW

Oracle® Enterprise Data Quality Product Data Services User Guide



12.2.1.3.0

E97546-01

July 2018

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

ORACLE®

Copyright © 2017, 2018, Oracle and/or its affiliates. All rights reserved.

Primary Authors: Oracle Corporation,

Contributing Authors:

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	vi
	Documentation Accessibility	vi
	Related Documents	vi
	Conventions	vii
1	About Product Data Services	
2	Installing EDQ-PDS	
	Prerequisites	2-1
	Components	2-1
	Installing PDS	2-1
3	Common Uses of EDQ-PDS	
4	Structure of EDQ-PDS Project	
	Data Stores	4-1
	Staged Data	4-1
	Data Interfaces	4-1
	Processes	4-2
	Reference Data	4-3
	Jobs	4-4
	Exports	4-4
	Web Services	4-5
5	EDQ-PDS Data Interfaces	
	Product Input Interface	5-1
	Match Result Interface	5-3

Batch Key Generation Results Interface	5-9
Realtime Key Generation Results Interface	5-9

6 EDQ-PDS Profiling Process

Profiling Process Structure	6-1
Quickstats	6-1
Price and Custom Number Fields	6-1
MPN and Model Number Pattern Profiling	6-1
All Fields Character Profiling	6-1
Product Description Individual Tokens	6-2
English Words Frequency Count	6-2
Non English Words Frequency Count	6-2
Tokens With No Vowels and No Numbers Frequency Count	6-2
Tokens Containing Numeric Characters Pattern Profiling	6-2
Other Product Description	6-2
Phrase Profiling	6-3
Extract Company	6-3
Extract Color	6-3
Extract Sizes	6-3
Extract Material	6-4
Extract Quantified Units of Measure	6-4

7 Data Preparation

Standardize Product Data	7-1
Standardize The Product Description	7-1
Abbreviate The Product Description	7-2
Standardize The Product Name	7-2
Create String Version of Price	7-2
Standardize The Custom Date Inputs	7-2
Standardize The Custom Strings	7-2
Custom Numbers	7-3
UIDs, EIDs and IEIDs	7-3

8 Key Generation

Structure of EDQ-PDS Key Generation Services	8-1
Published Processor For Creating Key Values	8-2
Key Methods	8-2

9 EDQ-PDS Matching

Match Process	9-1
Compound Comparisons	9-1
Product Description	9-1
Product Name	9-7
Price	9-8
Manufacturer Part Number	9-8
Model number	9-9
Custom Strings Compound Comparisons	9-9
Custom Numbers Compound Comparisons	9-10
Custom Dates Compound Comparisons	9-11
Scores	9-12
Rules	9-14
Inverted EID Eliminations	9-15
Overall Score	9-15
Match Relationship Processing	9-16

10 EDQ-PDS Published Processors

Abbreviate Processor	10-1
Extract Companies Processor	10-3
Extract Materials Processor	10-4
Extract Quantified Units of Measure	10-5

11 Configuration of EDQ-PDS Run Profile and Message Header Options

Configuring Batch Staging Area for External Integrations	11-6
--	------

Preface

Oracle Enterprise Data Quality Product Data Services User's Guide explains using the Production Data Services for Enterprise Data Quality.

Topics

- [Audience](#)
- [Documentation Accessibility](#)
- [Related Documents](#)
- [Conventions](#)

Audience

This document is intended for a user of the EDQ product who wishes to use the Product Data Services toolkit to analyze product data. It is assumed you have a working knowledge of the Enterprise Data Quality product.

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 about EDQ, see the following documents in the Enterprise Data Quality documentation set.

EDQ Documentation Library

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

- *Installing and Configuring Enterprise Data Quality*
- *Administering Enterprise Data Quality*
- *Understanding Enterprise Data Quality*
- *Integrating Enterprise Data Quality With External Systems*

- *Securing Oracle Enterprise Data Quality*
- *Installation and Upgrade Guide*
- *Release Notes*

Find the latest version of these guides and all of the Oracle product documentation at

<http://docs.oracle.com>

Online Help

Online help is provided for all Enterprise Data Quality 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
boldface	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

About Product Data Services

The EDQ-PDS (Product Data Services) pack is a set of tools, processes and services built on top of the EDQ product which can be used for profiling, standardizing and matching product data. It is designed as a starting point for a customer to begin this work, and will usually need modification on the content and structure of the data.

2

Installing EDQ-PDS

This chapter explains how to install the EDQ Product Data Services Pack.

Topics

- [Prerequisites](#)
- [Components](#)
- [Installing PDS](#)

Prerequisites

EDQ-PDS 12c (12.2.1.3.0) requires the following:

The requirements for production systems are as follows:

- 64-bit Operating System.
- 64-bit Java Virtual Machine (JVM).
- Minimum system memory of 8GB, with 4GB allocated to the JVM.
- EDQ 12.2.1.3.0 or later is installed.

Components

PDS is delivered with EDQ. To install the configuration, right-click the server name in EDQ and select **Open Server Package File > pds** folder. The folder contains the following components:

- `EDQ-PDS.dxi` - This contains the starter toolkit processes, jobs and services from which a user may start their PDS implementation.
- `EDQ-PDS - Published Processors.dxi` - This contains the published processors which appear on the Tool Palette under the Product Data tab (as well as some internal processors contained in the Solutions Reference Published Processors tab). They are also used in the EDQ-PDS project above.

Installing PDS

To install EDQ-PDS on the EDQ server:

1. Start the **EDQ Director client**, and log on as a user with the permission to create projects (Administrator or Data Analyst).
2. Right-click on the server name and select **Open Server Package File**. Open the PDS folder and select the `EDQ-PDS.dxi` file.
3. Expand the `EDQ-PDS.dxi` file and drag the whole project onto the Projects node.

4. Repeat steps 2 and 3 for the `EDQ-PDS-Published Processor.dxi` to bring the published processors onto the Published Processor node.
5. Once the projects have been imported, right-click on the `.dxi` files, and select **Close Package File**.

3

Common Uses of EDQ-PDS

EDQ-PDS is designed as a toolkit and starter project for performing profiling, data extraction, and matching on product data. It can be used out of the box, especially for profiling and matching. For extraction and standardization use cases, for example file parsing and standardizing a Product Description into several standard attributes EDQ provides a number of useful processors. Work will be needed to design suitable extraction and standardization processes for the data as product data is by its nature highly variable. The following section describes scenarios in which the EDQ's Product Data Services may be used.

Ways To Use EDQ-PDS Project

Providing Attached Matching Services

EDQ-PDS can be used to provide attached matching services (both batch and real-time) to an application or hub storing Product Data records. The matching services provided are designed to work well with either structured or unstructured product data records, and can be tailored and tuned for added accuracy where needed. The following starting points are recommended:

- Look at the Interfaces, in particular the Product Input interface to which your product data's fields should be mapped. For more information, see [EDQ-PDS Data Interfaces](#)
- Consider how the integration is to be performed. EDQ-PDS is provided with an external data source which can be used to read from an external Oracle database table for the data input. The details of this data store and table can be changed in the run profile., see [Configuring Batch Staging Area for External Integrations](#)
- Check the [Configuration of EDQ-PDS Run Profile and Message Header Options](#) section for seeing the matching and key generation settings that can be changed without modifying the processes.

Optimizing Matching For Your Data

A common use case is that the project will be used as a starter project but with the expectation that the user would like to explore the data, and make significant modifications to the project. In this case the following starting points would be recommended:

- Look at the [Product Input Interface](#) data interface, import and map the product data to this and run it through the [EDQ-PDS Profiling Process](#) to explore your data and look for ways to optimize matching, for example by removing anonymous values or tokens, standardizing abbreviations, or extracting key product attributes that may help with matching. Note that it is useful both to gain an initial understanding of the data through profiling, and to run the data through a batch matching process 'raw' in order to determine how matching may be improved.
- Look at the structure of the project to see what is provided. See, [EDQ-PDS Project Structure](#)

- Look at the [Standardize Product Data](#) for the provided functionality to prepare the product data for matching and key generation. There are default reference data provided with this process which may be modified if required
- Look at the [Key Generation](#) section for provided Key Methods, and also for details of reference data used to prepare input data specifically for Key Generation.
- Look at the [Matching](#) section for the provided matching functionality. This can either be used straight out of the box, tweaked through modifying the weightings of compound comparisons and match threshold, etc, or used as just a starting point for matching functionality, as required.

Providing a Toolkit To Profile, Extract and Standardize Product Data

- PDS provides several published processors designed for use with product data, for extracting information from your data and these come with default data as provided in the project. For that reason it is recommended to start with the project as a whole even if starting from a scratch, so that the relevant data is available for modification as required.
- EDQ includes a number of useful processors for working with Product Data, such as Extract Attributes, Parse, Split Records from Array, and Make Attribute Arrays. For more information, see section `Processor Library` in Oracle Enterprise Data Quality Online Help in Oracle Enterprise Data Quality documentation at <http://docs.oracle.com/en/middleware/>
- It is always recommended to profile any product data sets you are working with, either by using standard EDQ profilers configured from scratch, or by running the data through the provided profiling process, which provides examples of how to analyze key common Product Data attributes such as Product Description. This can provide useful insights into the product data and provide a starting point for extracting or restructuring data. For more information on data preparation, see [Data Preparation](#)

4

Structure of EDQ-PDS Project

The section describes the project elements provided with the EDQ-PDS project.

Topics

- [Data Stores](#)
- [Staged Data](#)
- [Data Interfaces](#)
- [Processes](#)
- [Published Processors](#)
- [Reference Data](#)
- [Jobs](#)
- [Exports](#)
- [Web Services](#)

Data Stores

This section describes the data stores provided with EDQ-PDS project.

Name	Description
Batch Staging Area	Staging area for batch matching services. Input records for matching can be written here by external applications for EDQ to process, and EDQ writes match results back here to a different table.

Staged Data

This section describes the staged data provided with the EDQ-PDS project.

Name	Description
Product Candidates	Snapshot from staging area for batch product data to be input to batch jobs.

Data Interfaces

The EDQ-PDS project consists of data interfaces as shown below.

Name	Description
Batch Key Generation Results	Output interface for batch key generation results.
Match Relationships Internal	Internal interface for processing match relationships.
Match Results	Match results output interface (used for both real-time and batch).
Product Input	Input interface for product data.
Product Internal	Internal interface for product data.
Realtime Key Generation Results	Output interface for real-time key generation results.

Processes

The following processes are provided with the EDQ-PDS project.

Name	Description
Profile Product Data	Profiles and analyzes product data as a starting point for understanding the structure and content of a data set.
Set Key Generation Profile	Obtains and sets the keyprofile setting from the message header or run profile.
Standardize Product Data	Standardizes and abbreviates product data for use in matching and key generation.
Real-time Add Keys	Creates keys for a real-time matching request, to match all drivers together and with all candidates.
Key Generation	Creates keys for use in batch matching and in the real-time key generation service.
Batch Output	Outputs keys to batch one output record per key value.
Real-time Output	Outputs keys for the real-time key generation service, one array of keys per record.
Match Product Data	Performs the matching of product data, in both batch and real-time.
Prepare Match Relationships	Prepares the relationships for match results output, calculating which compound comparisons have been used and constructing rule name outputs.
Generate Reverse Relationships	Prepares match results by ensuring drivers are on the left of relationships and that driver—driver relationships are duplicated.
Match Output	Outputs match results, adding serverid and jobid to results for use in external batch matching integrations.

Example 4-1

Reference Data

The following reference data sets are provided with the EDQ-PDS project.

EDQ-PDS consists of some pre-installed reference data sets as shown below.

Name	Description
Key Generation - Blank Replacements for Abbreviate	Blank reference data for providing the required reference for the Abbreviate processor used in the Key Generation processor.
Key Generation - Product Description Strip Tokens	<p>Tokens that are stripped from a product description attribute before key values are generated, for example single initials such as 'S' which may be too common to form keys but may be significant in matching.</p> <p>Note that these are tokens that you specifically want to remove from key values but not from matching. See, <i>Match Preparation - Product Description Strip Words</i> for the list of tokens that are stripped from both.</p>
Key Generation - Product Description Token Delimiters	Characters to use to delimit tokens in key generation - default is spaces.
Key Generation - Product Description Token Standardization	Token replacements to perform on product description tokens within key generation. These are applied in addition to the standardization and abbreviate performed in the Standardization process.
Match Preparation - Product Description Character strip/standardize	Individual characters to strip or standardize in the product description in preparation for key generation and matching.
Match Preparation - Product Description Strip Words	Tokens to strip from product descriptions for both key generation and matching, for example very common non-identifying words such as 'and' and 'the'.
Match Preparation - Product Description Standardize	Tokens to standardize in product description prior to matching.
Match Preparation - Product Name Character strip/standardize	Individual characters to strip or standardize from product name in match preparation.
Match Preparation - Product Name Standardize	Tokens to standardize in product name prior to matching.
Match Preparation - Strip Vowels	Contains vowels to strip from shortened product description.
Normalize Text- Remove Diacritics	Diacritic characters to remove from product description and name.
Normalize Text- Standardize Accented Characters	Standardization for accented characters.
Date - Formats	Valid formats for custom dates.

Name	Description
Profile - Product Data - Colors	Colors for standardization and extraction.
Profile - Product Data - Companies	Common retail companies for standardization and extraction.
Profile - Product Data - English Dictionary	Dictionary of English words - used for Profiling.
Profile - Product Data - Materials	Common materials for standardization and extraction.
Profile - Product Data - Number Bands	Number bands used for price profiling.
Profile - Product Data - Sizes	Sizes for standardization and extraction of sizes.
Profile - Product Data - Units of Measure Regex	Regular expressions for standardization and extraction of quantified units of measure.
Profile - Strip Letters	Vowels to strip when profiling.

Jobs

This section describes the jobs provided with the EDQ-PDS project.

Name	Description
Batch Product Key Generation	Create product data keys reading data from external batch table.
Batch Product Match	Match product data reading from staged batch data.
Real-time Product Key Generation	Real-time job for performing key generation on product data.
Real-time Product Match	Real-time job for matching product data.
Real-time START ALL	Start all the real-time processes.
Real-time STOP ALL	Stop all the real-time processes.

Exports

This section describes the exports provided with the EDQ-PDS project.

Name	Description
Batch Key Generation Results	Exports key generation results to the batch staging area.
Batch Matches	Exports matching results to the batch staging area.

Web Services

This section describes the web services provided with EDQ-PDS project.

Name	Description
ProductKeyGen	Web service for real-time key generation.
ProductMatch	Web service for real-time matching.

5

EDQ-PDS Data Interfaces

This section describes the EDQ-PDS data interfaces.

Topics

- [Product Input Interface](#)
- [Match Results Interface](#)
- [Batch Key Generation Results Interface](#)
- [Realtime Key Generation Results Interface](#)

Product Input Interface

This section shows the Product Input interface. This interface is used to ingest data into both the default profiling process, and the matching services.

Name	Type	Description	Batch Staging Mapping	Real-time Webservice mapping-match	Real-time Webservice Mappings - key gen
productid	String	Identifier for each product record.	productid	productid	productid
candidate	String	Indicates if the record is a driver (for comparison with all records) or a candidate (for comparison with only drivers). If a candidate, the value should be 1.	candidate	candidate	
productdescription	String	Provides product text description.	productdescription	productdescription	productdescription
productname	String	Product name for the product.	productname	productname	productname
modelnumber	String	Unique identifier for model of product.	modelnumber	modelnumber	modelnumber
manufacturerpartnumber	String	Identifier for product part (unique by manufacturer).	manufacturerpartnumber	manufacturerpartnumber	manufacturerpartnumber
price	Number	Price of this product in a standardized currency for all products.	price	price	price

Name	Type	Description	Batch Staging Mapping	Real-time Webservice mapping-match	Real-time Webservice Mappings - key gen
category	String	Category to group product with others of the same type.	category	category	category
manufacturer	String	Manufacturer of product.	manufacturer	manufacturer	manufacturer
customstring1	String	Custom field for mapping other product data field not mapped elsewhere.	customstring1	customstring1	customstring1
customstring2	String	Custom field for mapping other product data field not mapped elsewhere.	customstring2	customstring2	customstring2
customstring3	String	Custom field for mapping other product data field not mapped elsewhere.	customstring3	customstring3	customstring3
customstring4	String	Custom field for mapping other product data field not mapped elsewhere.	customstring4	customstring4	customstring4
customstring5	String	Custom field for mapping other product data field not mapped elsewhere.	customstring5	customstring5	customstring5
customnumber1	Number	Custom field for mapping other product data field not mapped elsewhere.	customnumber1	customnumber1	customnumber1
customnumber2	Number	Custom field for mapping other product data field not mapped elsewhere.	customnumber2	customnumber2	customnumber2
customnumber3	Number	Custom field for mapping other product data field not mapped elsewhere.	customnumber3	customnumber3	customnumber3
customnumber4	Number	Custom field for mapping other product data field not mapped elsewhere.	customnumber4	customnumber4	customnumber4
customnumber5	Number	Custom field for mapping other product data field not mapped elsewhere.	customnumber5	customnumber5	customnumber5
customdate1	String	Custom field for mapping other product data field not mapped elsewhere.	customdate1	customdate1	customdate1
customdate2	String	Custom field for mapping other product data field not mapped elsewhere.	customdate2	customdate2	customdate2
uid1	String	Unique identifier for product that causes immediate match.	uid1	uid1	uid1

Name	Type	Description	Batch Staging Mapping	Real-time Webservice mapping-match	Real-time Webservice Mappings - key gen
uid2	String	Unique identifier for product that causes immediate match.	uid2	uid2	uid2
uid3	String	Unique identifier for product that causes immediate match.	uid3	uid3	uid3
eid1	String	Identifier for product that causes immediate no match, if different.	eid1	eid1	eid1
eid2	String	Unique identifier for product that causes immediate match, if different.	eid2	eid2	eid2
eid3	String	Unique identifier for product that causes immediate match, if different.	eid3	eid3	eid3
ieid1	String	Identifier for product that causes immediate no match, if the same.	ieid1	ieid1	ieid1
ieid2	String	Identifier for product that causes immediate no match, if the same.	ieid2	ieid2	ieid2
ieid3	String	Identifier for product that causes immediate no match, if the same.	ieid3	ieid3	ieid3

Match Result Interface

This interface is used for the output for matching services in batch and real-time

Match Result Interface

Attribute	Type	Description	Web Service Mappings- Product Match Web Service Output Mapping	Default Batch Staging Mappings- Default Matches Staging Table Mapping
serverid	String	Used to identify the server from which a job was submitted when running a batch job from an external system. Not used in real-time.	-	serverid

Attribute	Type	Description	Web Service Mappings- Product Match Web Service Output Mapping	Default Batch Staging Mappings- Default Matches Staging Table Mapping
jobid	String	Used to identify a distinct job as defined externally when a batch job is run from an external system. Not used in real-time.	-	jobid
masterproductid	String	Master product id.	masterproductid	masterpartyid
matchproductid	String	Matching record id.	matchproductid	matchpartyid
ruleattributes	String	Comma separated list of attributes contributing to the rule.	ruleattributes	ruleattributes
matchscore	Number	Match score.	matchscore	rulescore
rulename	String	Match rule name, this may be taken from a compound score result if the relationship was generated from a compound score based rule.	rulename	rulename
reversedriverflag	String	A flag indicating that an additional, reversed match record has been generated where there is a match between driving records in Batch matching. Valid values are Y and N.	reversedriverflag	reversedriverflag
comparisonresults	String	Comma separated list of attributes contributing to the relationship, and how they matched (e.g. Description Exact, Price Fuzzy).	comparisonresults	comparisonresults
productdescriptionresult	String	Result of rule fragment for product description compound comparison, if it contributed to the output score.	productdescriptionresult	productdescriptionresult
productdescriptionscore	Number	Score of rule fragment for product description compound comparison, if it contributed to the output score.	productdescriptionscore	productdescriptionscore

Attribute	Type	Description	Web Service Mappings- Product Match Web Service Output Mapping	Default Batch Staging Mappings- Default Matches Staging Table Mapping
productdescriptioncategory	String	Category of rule fragment for product description compound comparison, if it contributed to the output score.	productdescriptioncategory	productdescriptioncategory
pricerresult	String	Result of rule fragment for price compound comparison, if it contributed to the output score.	pricerresult	pricerresult
pricescore	Number	Score of rule fragment for price compound comparison, if it contributed to the output score.	pricescore	pricescore
pricecategory	String	Category of rule fragment for price compound comparison, if it contributed to the output score.	pricecategory	pricecategory
mpnresult	String	Result of rule fragment for manufacturer part number compound comparison, if it contributed to the output score.	mpnresult	mpnresult
mpnscore	Number	Score of rule fragment for manufacturer part number compound comparison, if it contributed to the output score.	mpnscore	mpnscore
mpncategory	String	Category of rule fragment for manufacturer part number compound comparison, if it contributed to the output score.	mpncategory	mpncategory
modelnumberresult	String	Result of rule fragment for model number compound comparison, if it contributed to the output score.	modelnumberresult	modelnumberresult
modelnumberscore	Number	Score of rule fragment for model number compound comparison, if it contributed to the output score.	modelnumberscore	modelnumberscore

Attribute	Type	Description	Web Service Mappings- Product Match Web Service Output Mapping	Default Batch Staging Mappings- Default Matches Staging Table Mapping
modelnumbercategory	String	Category of rule fragment for model number compound comparison, if it contributed to the output score.	modelnumbercategory	modelnumbercategory
productnameresult	String	Result of rule fragment for product name compound comparison, if it contributed to the output score.	productnameresult	productnameresult
productnamescore	Number	Score of rule fragment for product name compound comparison, if it contributed to the output score.	productnamescore	productnamescore
productnamecategory	String	Category of rule fragment for product name compound comparison, if it contributed to the output score.	productnamecategory	productnamecategory
customstring1result	String	Result of rule fragment for custom string compound comparison.	customstring1result	customstring1result
customstring1score	Number	Score of rule fragment for custom string compound comparison.	customstring1score	customstring1score
customstring1category	String	Category of rule fragment for custom string compound comparison.	customstring1category	customstring1category
customstring2result	String	Result of rule fragment for custom string compound comparison.	customstring2result	customstring2result
customstring2score	Number	Score of rule fragment for custom string compound comparison.	customstring2score	customstring2score
customstring2category	String	Category of rule fragment for custom string compound comparison.	customstring2category	customstring2category
customstring3result	String	Result of rule fragment for custom string compound comparison.	customstring3result	customstring3result
customstring3score	Number	Score of rule fragment for custom string compound comparison.	customstring3score	customstring3score

Attribute	Type	Description	Web Service Mappings- Product Match Web Service Output Mapping	Default Batch Staging Mappings- Default Matches Staging Table Mapping
customstring3category	String	Category of rule fragment for custom string compound comparison.	customstring3category	customstring3category
customstring4result	String	Result of rule fragment for custom string compound comparison.	customstring4result	customstring4result
customstring4score	Number	Score of rule fragment for custom string compound comparison.	customstring4score	customstring4score
customstring4category	String	Category of rule fragment for custom string compound comparison.	customstring4category	customstring4category
customstring5result	String	Result of rule fragment for custom string compound comparison.	customstring5result	customstring5result
customstring5score	Number	Score of rule fragment for custom string compound comparison.	customstring5score	customstring5score
customstring5category	String	Category of rule fragment for custom string compound comparison.	customstring5category	customstring5category
customnumber1result	String	Result of rule fragment for custom number compound comparison.	customnumber1result	customnumber1result
customnumber1score	Number	Score of rule fragment for custom number compound comparison.	customnumber1score	customnumber1score
customnumber1category	String	Category of rule fragment for custom number compound comparison.	customnumber1category	customnumber1category
customnumber2result	String	Result of rule fragment for custom number compound comparison.	customnumber2result	customnumber2result
customnumber2score	Number	Score of rule fragment for custom number compound comparison.	customnumber2score	customnumber2score
customnumber2category	String	Category of rule fragment for custom number compound comparison.	customnumber2category	customnumber2category
customnumber3result	String	Result of rule fragment for custom number compound comparison.	customnumber3result	customnumber3result

Attribute	Type	Description	Web Service Mappings- Product Match Web Service Output Mapping	Default Batch Staging Mappings- Default Matches Staging Table Mapping
customnumber3score	Number	Score of rule fragment for custom number compound comparison.	customnumber3score	customnumber3score
customnumber3category	String	Category of rule fragment for custom number compound comparison.	customnumber3category	customnumber3category
customnumber4result	String	Result of rule fragment for custom number compound comparison.	customnumber4result	customnumber4result
customnumber4score	Number	Score of rule fragment for custom number compound comparison.	customnumber4score	customnumber4score
customnumber4category	String	Category of rule fragment for custom number compound comparison.	customnumber4category	customnumber4category
customnumber5result	String	Result of rule fragment for custom number compound comparison.	customnumber5result	customnumber5result
customnumber5score	Number	Score of rule fragment for custom number compound comparison.	customnumber5score	customnumber5score
customnumber5category	String	Category of rule fragment for custom number compound comparison.	customnumber5category	customnumber5category
customdate1result	String	Result of rule fragment for custom date compound comparison.	customdate1result	customdate1result
customdate1score	Number	Score of rule fragment for custom date compound comparison.	customdate1score	customdate1score
customdate1category	String	Category of rule fragment for custom date compound comparison.	customdate1category	customdate1category
customdate2result	String	Result of rule fragment for custom date compound comparison.	customdate2result	customdate2result
customdate2score	Number	Score of rule fragment for custom date compound comparison.	customdate2score	customdate2score
customdate2category	String	Category of rule fragment for custom date compound comparison.	customdate2category	customdate2category

Batch Key Generation Results Interface

This interface is used for the output for key generation services in batch only.

Batch Key Generation Results Interface

Attribute	Type	Description
serverid	String	Used to identify the server from which a job was submitted when running a batch job from an external system. Not used in real-time.
jobid	String	Used to identify a distinct job as defined externally when a batch job is run from an external system. Not used in real-time.
productid	String	Unique record id.
keyprofile	String	Key profile used.
keyvalue	String	Key value.
keypriority	String	Priority of key value, used to pick more likely candidates for matching above less likely candidates externally.

Realtime Key Generation Results Interface

This interface is used for the output for key generation services in real-time only.

Attribute	Type	Description
productid	String	Unique product id.
keyprofile	String	The key profile used.
keyvalues	String	Array of key values containing keys for all key methods for this particular record.
keypriorities	String	Array of key priorities for the key values in the key values array.

6

EDQ-PDS Profiling Process

The profiling process is intended to be a starting point for exploring product data as a starting point for a project working on matching, standardization and/or extraction of product data. Out of the box it will give useful information and an insight into any type of product data.

Profiling Process Structure

The profiling process reads in from the standard EDQ-PDS Input interface and Product Input data interface. In order to analyze product data the data being analyzed should be mapped to this interface. A description of the fields in this interface and their intended use can be found in [EDQ-PDS Data Interfaces](#) section in this guide. Note that the profiling process performs the most complicated analysis on the product description field, since product data generally has one, unstructured field which requires the majority of the analysis. Out of the box, the profiling process has the following processors which give the following information.

Quickstats

A quick stats profiler works on all the product data fields. This shows information on how well these fields are populated, whether they are numeric, and how many identical or unique records are in the field.

Price and Custom Number Fields

This is a number profiler for Price and Custom Number fields and is configured with the number bands provided in the Profile –Product Data –Number Bands reference data. This reference data can be modified if required.

MPN and Model Number Pattern Profiling

The MPN and model number inputs are passed through a pattern profiler. This is used to identify common patterns in these fields and potentially highlight any inconsistencies.

All Fields Character Profiling

This is a character profiler to profile the individual characters in all fields with the exception of A-Z characters, a-z characters, and numeric characters from 0–9. This can show unusual characters that are present in the data.

Product Description Individual Tokens

The following processors profile each individual token in the product description.

English Words Frequency Count

It profiles English words to show the frequency of occurrence in the product description. This may show up useful words to strip out (for example – Unclassified), or words that can be used to extract information out of the data – e.g. PACK, SIZE, INCH, etc.

Non English Words Frequency Count

This profiles words that do not contain numeric characters, to look for non English dictionary words. This can show up common company names in the data, or common misspellings.

Tokens With No Vowels and No Numbers Frequency Count

This profiles tokens that contain no numeric characters and no vowels to look for potential abbreviations in the data.

Tokens Containing Numeric Characters Pattern Profiling

This profiles tokens containing numeric characters to look for potential structure of id numbers contained within the description.

Other Product Description

This section contains:

- [Phrase Profiling](#)
- [Extract Company](#)
- [Extract Color](#)

- [Extract Sizes](#)
- [Extract Material](#)
- [Extract Quantified Units of Measure](#)

Phrase Profiling

Phrase profiler profiles the product description to look for common phrases.

Extract Company

Extracts potential company names from the product description, along with the category of the company (e.g. Home and Garden, Fashion and Clothing, etc).

The results show a frequency count of extracted company and a frequency count of the associated category of the companies.

Note, prior to extracting the company data the input product description has punctuation removed, accents standardized and the "&" character converted to "AND ". For this reason data added to the reference data should have any punctuation, accents removed and the "&" character converted as stated, in order for them to match.

The reference data Profile – Product Data – Companies contains the data used for this extraction and contains common retail company names and their categories.

Extract Color

Extracts colors from the product description, standardizing different spellings and potential abbreviations to a common version.

The reference data Profile → Product Data → Colors contains the data used for this extraction and can be modified as required.

The results show a frequency count of extracted color and a frequency count of the standardized version of the extracted color.

Extract Sizes

Extracts potential sizes from the product description, standardizing different spellings and potential abbreviations to a common version.

The reference data Profile – Product Data – Sizes contains the data used for this extraction and can be modified as required.

The results show a frequency count of extracted size and a frequency count of the standardized version of the extracted size.

Extract Material

Extracts materials from the product description, for example Steel, Brass, Cotton, etc.

The reference data Profile – Product Data – Materials contains the data used for the extraction, it also contains a standardized version of the Material although in the majority of cases this is the same as the version being standardized.

The results show a frequency count of extracted material and a frequency count of the standardized version of the extracted material.

Extract Quantified Units of Measure

Extracts quantified units of measure from the data and standardizes them to a common version. For example 1", 1 IN and 1 INCH are all standardized to 1 INCH.

The reference data can be found at Profile → Product Data → Units of Measure Regex. If you wish to add your own values then it can be done by reusing an existing entry and copying the structure of the existing regular expressions.

The results show frequency counts of the number combined with the standardized unit of measure and also a frequency count of the units of measure alone (e.g. count of number of instances of INCH, etc).

7

Data Preparation

Standardize Product Data

This is the process in which the input data is prepared for match. You can modify this process to perform your own data preparation, or modify the reference data which it uses to standardize the data.

Topics

- [Standardize The Product Description](#)
- [Abbreviate The Product Description](#)
- [Standardize The Product Name](#)
- [Create String Version of Price](#)
- [Standardize The Custom Date Inputs](#)
- [Standardize The Custom Strings](#)
- [Custom Numbers](#)
- [UIDs, EIDs and IEIDs](#)

Standardize The Product Description

This consists of:

- Stripping diacritic characters.
- Standardizing accented characters – (é to e for example).
- Upper casing.
- Replacing characters with spaces and/or stripping characters. The reference data used by this is **Match Preparation – Product Description Character Strip/Standardize**, by default it strips most non alpha-numeric characters, and replaces brackets and commas with spaces.
- Standardizes the product description using a replace processor. This uses the **Match Preparation – Product Description Standardize** reference data to replace certain tokens or phrases with others. The provided reference data is empty, and it is likely that a customer will want to add their own entries to it. For example, if the data contains descriptions with the color “WHITE” represented in some places as “WHT”, they could replace “WHT” with “WHITE”.
- Normalize the whitespace in the product description.

The result of this is the **productdescriptionstandardized** field.

Abbreviate The Product Description

Strips words from the product description to create a separate abbreviated version of the product description for use in matching. The list of words to strip is contained in **Match Preparation – Product Description Strip Words**. By default this contains some common words that are unlikely to be useful in matching such as IN, THE, etc. You can add your own records to this reference data.

Standardize The Product Name

This consists of:

- Stripping diacritic characters.
- Standardizing accented characters – (é to e for example).
- Upper casing.
- Replacing characters with spaces and/or stripping characters. The reference data used by this is **Match Preparation – Product Name Character Strip/Standardize**, by default it strips most non alpha-numeric characters, and replaces brackets and commas with spaces.
- Standardizes the product name using a replace processor. This uses the **Match Preparation – Product Name Token Standardize** reference data to replace certain tokens with others. The provided reference data is empty, and it is likely that a customer will want to modify it according to their data. For example, if the data contains descriptions with the colour “WHITE” represented in some places as “WHT”, they could replace “WHT” with “WHITE”.
- Normalize the whitespace.

The result of this is the **productnamestandardized** field.

Create String Version of Price

Takes the price input and converts it to a string. This is mapped to the **pricestring** field.

Standardize The Custom Date Inputs

Takes the custom date inputs (which are both input as strings to the Input interface) and converts them into dates. The valid date formats are contained in the **Date – Formats** reference data. These are mapped to the **customdate1standardized** and **customdate2standardized** fields.

Standardize The Custom Strings

Normalizes the whitespace and uppercases the customstring [1,5] fields. These are mapped to the **customstring[1,5]standardized** fields.

Custom Numbers

Converts the customnumber[1,5] inputs to strings. These are mapped to the **customnumber[1,5] string** fields.

UIDs, EIDs and IEIDs

Normalizes whitespace and uppercases the uid, eid, and ieid fields. These are mapped to the **uid[1,3]standardized**, **eid[1,3]standardized** and **ieid[1,3]standardized** fields.

8

Key Generation

The Key Generation service will take data through the same interface as the matching service and return the Key Values required for matching for the given records.

The returned Key Values can be stored by the calling application for use in selecting their candidates for real-time matching, as in EDQ-CDS. In addition the key generation methods will be used in the batch matching service to create the keys for using prior to matching.

Topics

- [Structure of Key Generation Process](#)
- [Published Processor For Creating Key Values](#)
- [Key Methods](#)

Structure of EDQ-PDS Key Generation Services

The key generation process takes the productdescriptionabbr attribute as produced by the previous standardization process. It applies an additional Character Replace processor to the attribute using the **Key Generation - Product Description Character Strip/Standardize** reference data. By default this strips/replaces with spaces, the remaining punctuation characters that were not removed in the standardization process.

Then it separates this into tokens (based on a reference data **Key Generation - Product Description Token Delimiters**, which by default just contains a space), and standardizes and strips these tokens with the reference data **Key Generation - Product Description Token Standardization** and **Key Generation - Product Description Strip Tokens** respectively.

The tokens are then re-constructed into a single string and have the Abbreviate processor applied (with default settings) to create a string which creates a single, tight product description cluster, which is by default trimmed to 10 characters (and labeled for modification). From the product description tokens, they will be processed to produce the metaphone (by default metaphone 4 but this will be labelled for modification) of the tokens which do not have a numeric character. The tokens containing numeric characters will be unmodified, except they will be trimmed to a maximum of 12 characters length (this will be labeled for modification).

Two key methods are available for use on these processed tokens, one which produces one key per token, and one which produces one key per unique pair of tokens.). By default the maximum number of tokens per record produced by the key per token method will be 10, and the maximum number of tokens used for creating the token pairs will be 6 (producing a maximum of 15 token pairs). These two maximum values will be labeled for modification.

In addition to the key methods on the product description key methods will also be provided on

- productnamestandardized
- modelnumber (whitespace trimmed)
- uid1,2,3standardized (whitespace trimmed)
- manufacturepartnumber (whitespace trimmed)

Each key will have a unique prefix. Each key will be mapped to a single one of the array keyvalue attributes in the internal interface (keyvalue1, keyvalue2, etc).

Published Processor For Creating Key Values

There published processor provided for creating the key values is **PDS – Create Key values and priorities**. This will take keyprofile string, an array of key values, and a single key value as an input, and the key method name as an option. This published processor will check to see if the key method is in the key profile.

Note that the key method must be alpha numerics only and be followed by the ^ character (after which the priority is provided) in the key profile.

It will then prefix the key method name onto each element in the array and/or onto the single key value, and output the key values in an array, along with an array of the same length containing the key priorities (which will be the same for every element).

Key Methods

The following key methods are provided:

Name	Inputs	Key Value Creation	Example
Description Metaphone (DSM)	Standardize dTextMetaph one: Array of metaphone of length 4 tokens from each standardized text token of the description, denoised of tokens that are single character or identified as common tokens i.e. "and", "of", "with", "X", "x".	Create single key on each element of the array. Prefix: DSM	Input: {MXNK, PNSL, PRL} Key Output: {DSM^MXNK, DSM^PNSL, DSM^PRL} Default: ON
Key Method Description Metaphone Pairs (DPM)	Standardize dTextMetaph one: Metaphone of length 4 on each standardized text token, denoised from tokens that are single char or identified common tokens i.e. "and", "of", "with", "X", "x", split into array.	Create key value on each unique pair of tokens (order tokens alphabetically) Prefix: DSM	Input1: {MXNK, PNSL, PRL} Key Output: {DSM^MXNK^PNSL, DSM^PNSL^PRL, DSM^DSM^PRL} Default: OFF

Name	Inputs	Key Value Creation	Example
Key Method Description Simple (DSS)	Description ForDSS.Abbreviated: Single string reconstructed from the processed product description tokens (prior to metaphones being applied), passed through the EDQ-PDS Abbreviate processor (with default settings), trimmed to 10 characters.	Create key value on input Prefix: DPM	Input1: {CHSLTPPNYL} Key Output: {DSS^CHSLTPPNYL} Default: ON
Key Method Manufacturer part number (MPN)	ManufacturerPartNumber Trimmed: Manufacturer part number (whitespace removed).	Create Key value from part number input Prefix: MPM Input1: ManufacturerPartNumberTrimmed	Input1: PN12956 Output: MPN^PN12956 Default: ON
Key Method Model number (MDN)	ModelNumber Trimmed: Model number (whitespace removed)	Create Key value from model number input Prefix: MDM Input1: ModelNumberTrimmed	Input1: ADC-245632 Output: MDN^ADC-245632 Default: ON
Key Method Product name (PRN)	ProductName Trimmed: Product name, with whitespace removed	Create key value from single product name input Prefix: PRN Input1: ProductNameStandardizedTrimmed Length: Full length	Input1: ZINCHEXBOLTSUSSPACK Output: PRN^ZINCHEXBOLTSUSSPACK Default: ON

Name	Inputs	Key Value Creation	Example
Key Method UID full string (UID1..3)	unique id 1..3	Create key value from single UID input Prefix: UID1..3 Input1: Input1:Uid1..3	Input1: ABC123 Prefix: UID1 Output: {UID1^ABC123} Default: ON

9

EDQ-PDS Matching

This section describes the matching services in EDQ-PDS.

Topics

- [Match Process](#)
- [Match Relationship Processing](#)

Match Process

The records with their standardized/derived fields and key values (whether from batch or real-time) will then be passed into the match process, **Match – Product Data**. This takes one input stream and uses “compare against self” to compare records within the input stream.

Clustering

The 20 key value array attributes are mapped for clustering. By default each cluster is set with a maximum of 64000 comparisons permitted. If all data is passed in on the working data input, this is approximately equivalent to a maximum cluster size of 250.

Note that these limits apply to Batch Matching. In real-time matching, the external application or hub selects the candidate records for matching based on the key values that are passed back from the key generation service. Any limits in candidate selection are applied externally.

Compound Comparisons

This sections contains the compound comparisons.

Product Description

Rules are provided that match on the product description using the following principles:

- Matching between both the standardized and abbreviated input.
- Matching between descriptions relating to the same product where the words are out of order.
- Matching between descriptions where one description is a shorter description of the same product (higher priority given to those short descriptions where the “short” description contains more words, to avoid matches based on very sparse descriptions).

- Matching between descriptions where there are character differences, but possible ID values or quantities all match are given a higher weighting (to avoid matches between similar product descriptions that have different sizes).
- Looser matches that allow for typo matches on any token, including those that may be IDs or quantities.

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
P001 Product description exact	100	Exact	Product desc exact = true
P002 Product description stand all words	90	Fuzzy	Product descr stand WMP = 100
P003 Product description stand all words out of order	85	Fuzzy	Product descr stand WMP OOO= 100
P004 Product description stand all words relating to shorter, >= 4 matching words, WMP > 70	84	Fuzzy	Product descr stand WMP shorter = 100, Product descr stand WMC >= 4, Product descr stand WMP OOO >= 70
P005 Product description stand all words relating to shorter out of order, >= 4 matching words, WMP > 70	82	Fuzzy	Product descr stand WMP shorter OOO= 100 Product desc WMC OO >= 4 Product descr stand WMP OOO >= 70
P006 Product description stand all words relating to shorter, >= 4 matching words	80	Fuzzy	Product descr stand WMP shorter = 100 Product WMC OOO >= 4
P007 Product description stand all words relating to shorter out of order, >= 4 matching words	79	Fuzzy	Product descr stand WMP shorter OOO = 100 Product descr WMP OOO >= 4
P008 Product description stand all words relating to shorter, >= 2 matching words	77	Fuzzy	Product descr stand WMP shorter = 100 Product descr WMC OOO >= 2
P009 Product description stand all words relating to shorter out of order, >= 2 matching words	75	Fuzzy	Product descr stand WMP shorter OOO = 100 Product descr WMC OOO >=2

Result	Score	Category	Comparisons
P010 Product description abbr - non number words 1 typo, number words exact	70	Fuzzy	Product descr abbr non number words CED <= 1 Product descr abbr number words exact = true
P011 Product description abbr - all words out of order, number words exact, non number words typos	68	Fuzzy	Product descr abbr has multi tokens = true product descr abbr non number words WMP OO tolerant = 100 Product descr abbr number words WMP OOO = 100
P012 Product description abbr - all words out of order, number words no data, non number words typos	67	Fuzzy	Product descr abbr has multitokens = true product descr abbr non number words WMP OO tolerant = 100 Product descr abbr number words WMP OOO = no data
P013 Product description abbr - all words out of order, number words exact, non number words shortened exact	66	Fuzzy	Product descr abbr has multitokens = true Product descr abbr shortened WMP OOO = 100
P014 Product description abbr - all words out of order relating to shorter, number words exact, non number words typos >= 4 matching words WMP > 70	65	Fuzzy	Product descr abbr WMP shorter OOO tolerant Product descr abbr has multitoken = true Product descr abbr WMP OOO Tolerant >= 70 Product descr abbr WMC abbr tolerant >= 4 Product descr abbr non number words WMP shorter OOO tolerant = 100 Product descr abbr number words WMP shorter OOO = 100

Result	Score	Category	Comparisons
P015 Product description abbr - all words out of order relating to shorter, number words no data, non number words typos >= 4 matching words WMP> 70	64	Fuzzy	Product descr abbr WMP shorter OOO tolerant Product descr abbr has multitoken = true Product descr abbr WMP OOO Tolerant >= 70 Product descr abbr WMC abbr tolerant >= 4 Product descr abbr non number words WMP shorter OOO tolerant = 100 Product descr abbr number words WMP shorter OOO = no data
P016 Product description abbr - all words out of order relating to shorter, number words exact, non number words typos >= 4 matching words	63	Fuzzy	Product descr abbr WMP shorter OOO tolerant Product descr abbr has multitoken = true Product descr abbr WMC abbr tolerant >= 4 Product descr abbr non number words WMP shorter OOO tolerant = 100 Product descr abbr number words WMP shorter OOO = 100
P017 Product description abbr - all words out of order relating to shorter, number words no data, non number words typos >= 4 matching words	62	Fuzzy	Product descr abbr WMP shorter OOO tolerant Product descr abbr has multitoken = true Product descr abbr WMC abbr tolerant >= 4 Product descr abbr non number words WMP shorter OOO tolerant = 100 Product descr abbr number words WMP shorter OOO = no data
P018 Product description abbr - all words out of order relating to shorter, number words exact, non number words typos	61	Fuzzy	Product descr abbr WMP shorter OOO tolerant Product descr abbr has multitoken = true Product descr abbr non number words WMP shorter OOO tolerant = 100 Product descr abbr number words WMP shorter OOO = 100

Result	Score	Category	Comparisons
P019 Product description abbr - all words out of order relating to shorter, number words no data, non number words typos	60	Fuzzy	Product descr abbr WMP shorter OOO tolerant Product descr abbr has multitoken = true Product descr abbr non number words WMP shorter OOO tolerant = 100 Product descr abbr number words WMP shorter OOO = no data
P020 Product description abbr - all words out of order relating to shorter, number words exact, shortened non-number words exact	59	Fuzzy	Product descr abbr has multitokens = true Product descr abbr shortened WMP OOO relating to shorter = 100
P021 Product description stand all words relating to shorter, one record only has one word	58	Fuzzy	Product descr stand WMP shorter OOO
P022 Product description stand one typo	55	Fuzzy	Product descr stand CED <= 1
P023 Product description stand two typos	50	Fuzzy	Product descr stand CED <= 2
P024 Product description stand all words out of order, typos	40	Fuzzy	Product descr stand WMP OOO tolerant = 100
P025 Product description stand all words out of order, relating to shorter, typos	30	Fuzzy	Product descr stand WMP shorter OOO tolerant = 100
P026 Product description abbr CMP > 90, all number words	25	Fuzzy	Product descr abbr CMP >= 90 Product descr abbr number words WMP shorter OOO = 100
P027 Product description abbr CMP > 90, number words no data	25	Fuzzy	Product descr abbr CMP >= 90 Product descr abbr number words WMP shorter OOO = no data

Result	Score	Category	Comparisons
P028 Product description stand LCSP 90 relating to shorter, all number words	23	Fuzzy	Product descr stand Longest Common Substring Percentage \geq 90 Product descr abbr number words WMP shorter OOO = 100
P029 Product description stand LCSP 90 relating to shorter, number words no data	23	Fuzzy	Product descr stand Longest Common Substring Percentage \geq 90 Product descr abbr number words WMP shorter OOO = no data
P030 Product description abbr CMP $>$ 80, all number words	22	Fuzzy	Product descr abbr CMP \geq 80 Product descr abbr number words WMP shorter OOO = 100
P031 Product description abbr CMP $>$ 80, number words no data	22	Fuzzy	Product descr abbr CMP \geq 80 Product descr abbr number words WMP shorter OOO = no data
P032 Product description stand LCSP 80 relating to shorter, all number words	21	Fuzzy	Product descr stand Longest Common Substring Percentage \geq 80 Product descr abbr number words WMP shorter OOO = 100
P033 Product description stand LCSP 80 relating to shorter, number words no data	21	Fuzzy	Product descr stand Longest Common Substring Percentage \geq 80 Product descr abbr number words WMP shorter OOO = no data
P034 Product description abbr CMP $>$ 70, all number words	20	Fuzzy	Product descr abbr CMP \geq 70 Product descr abbr number words WMP shorter OOO = 100
P035 Product description abbr CMP $>$ 70, number words no data	20	Fuzzy	Product descr abbr CMP \geq 70 Product descr abbr number words WMP shorter OOO = no data
P036 Product description stand LCSP 70 relating to shorter, all number words	19	Fuzzy	Product descr stand Longest Common Substring Percentage \geq 70 Product descr abbr number words WMP shorter OOO = 100

Result	Score	Category	Comparisons
P037 Product description stand LCSP 70 relating to shorter, number words no data	19	Fuzzy	Product descr stand Longest Common Substring Percentage >= 70 Product descr abbr number words WMP shorter OOO = no data
P038 Product description abbr CMP > 90	18	Fuzzy	Product descr abbr CMP >= 90
P039 Product description stand LCSP 90 relating to shorter	18	Fuzzy	Product descr stand Longest Common Substring Percentage >= 90
P040 Product description abbr CMP > 80	16	Fuzzy	Product descr abbr CMP >= 80
P041 Product description stand LCSP 80 relating to shorter	16	Fuzzy	
P042 Product description no data	0	No data	Product descr stand exact = no data
P043 Product description conflict	-3		*

Product Name

Rules are provided that match on the product name using the following principles:

- Exact Match.
- Matches containing character differences.
- Matches containing missing words.
- Matches where the words are out of order.

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
PR001 Product name exact	100	Exact	Product name exact = true
PR002 Product name typo	90	Fuzzy	Product name CED <=1
PR003 Product name all words out of order	80	Fuzzy	Product name stand WMP OOO = 100

Result	Score	Category	Comparisons
PR004 Product name all words out of order relating to shorter	60	Fuzzy	Product name stand WMP OOO relating to shorter = 100
PR005 Product name all words out of order relating to shorter typos	40	Fuzzy	Product name stand WMP OOO relating to shorter tolerant= 100
PR006 Product name no data	0	No data	Product name exact= no data
PR007 Product name conflict	-5	Conflict	*

Price

Rules are provided that match on the price using the following principles:

- Exact Matches.
- Character transposition of two digits.
- Small absolute difference between the two numbers.
- Small percentage difference between the two numbers.

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
PR001 Price exact	100	Exact	Price exact = true
PR002 Price string character transposition match	40	Fuzzy	Price string character transposition match = true
PR003 Price absolute difference < 2	30	Fuzzy	Price absolute difference < 2
PR004 Price percentage difference < 10	20	Fuzzy	Price Percentage difference < 10
PR005 Price no data	0	No data	Price string exact = no data
PR0056 Price conflict	-5	Conflict	*

Manufacturer Part Number

Rules are provided that match on the MPN using the following principles:

- Exact match.

- Character transposition of two characters.
- Typographic difference.

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
MPN001 Manufacturer part number exact	100	Exact	Manufacturer part number exact = true
MPN002 Manufacturer part number character transposition	60	Fuzzy	Manufacturer part number character transposition = true
MPN003 Manufacturer part number typos	40	Fuzzy	Manufacturer part number CED <=1
MPN004 Manufacturer part number no data	0	No data	Manufacturer part number exact = no data
MPN005 Manufacturer part number conflict	-10	Conflict	*

Model number

Rules are provided that match on the model number using the following principles:

- Exact Match.
- Character transposition of two characters.
- Typographic difference.

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
MN001 Model number exact	100	Exact	Model number exact = true
MN002 Model number character transposition	60	Fuzzy	Model number character transposition = true
MN003 Model number typos	40	Fuzzy	Model number CED <=1
MN004 Model number no data	0	No data	Model number exact = no data
MN005 Model number conflict	-10	Conflict	*

Custom Strings Compound Comparisons

Two compound comparisons are provided per custom string field, one where exact matches only are of interest, and one where the records being matched may contain differences. It is designed with the expectation that only one of these compound comparisons would be enabled, as per requirement.

The exact compound comparison (designed for when a field is expected to contain ID values or similar) contains only a single rule of an exact match. The other, fuzzy, compound comparison contains typographic match rules and out of order word matches (designed for matching on fields where the contents are more flexible and may have differences for a match to be considered).

Custom String 1..10 Exact

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
CS[N]001 Custom String N exact	100	Exact	Custom String N exact = true
CS[N]005 Custom String N no data	0	No data	Custom String N exact= no data
CS[N]006 Custom String N conflict	-5	Conflict	*

Custom String 1...10 Fuzzy

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
CS[N]001 Custom String N exact	100	Exact	Custom String N exact = true
CS[N]002 Custom String N typos	80	Fuzzy	Custom String N CED <=1
CS[N]003 Custom String N all words	60	Fuzzy	Custom String N stand WMP OOO = 100
CS[N]004 Custom String N all words typos	30	Fuzzy	Custom String N stand WMP OOO relating to shorter = 100
CS[N]005 Custom String N no data	0	No data	Custom String N exact= no data
CS[N]006 Custom String N conflict	-5	Conflict	*

Custom Numbers Compound Comparisons

Two compound comparisons are provided per custom number field, one where exact matches only are of interest, and one where the records being matched may contain differences. It is designed with the expectation that only one of these compound comparisons would be enabled, as per requirement.

The exact compound comparison contains only a single rule of an exact match. The other, fuzzy, compound comparison contains rules that cater for character transposition of digits, small absolute differences between the numbers and small percentage differences between the numbers.

Custom Number Exact

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
CN[N]001 Custom Number N exact	100	Exact	Custom Number N exact = true
CN[N]005 Custom Number N no data	0	No data	Custom Number N exact = no data
CN[N]006 Custom Number N conflict	-5	Conflict	*

Custom Number Fuzzy

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
CN[N]001 Custom Number N exact	100	Exact	Custom Number N exact = true
CN[N]002 Custom Number N character transposition match	80	Fuzzy	Custom number N character transposition match = true
CN[N]003 Custom Number N absolute difference < 2	40	Fuzzy	Custom number N absolute difference < 2
CN[N]004 Price percentage difference < 10	30	Fuzzy	Custom N Percentage difference < 10
CN[N]005 Custom Number N no data	0	No data	Custom Number N exact = no data
CN[N]006 Custom Number N conflict	-5	Conflict	*

Custom Dates Compound Comparisons

Two compound comparisons are provided per custom dates, one where exact matches only are of interest, and one where the records being matched may contain differences. It is designed with the expectation that only one of these compound comparisons would be enabled, as per requirement.

The exact compound comparison contains only a single rule of an exact match. The other, fuzzy, compound comparison contains rules that cater for similar dates as well as an exact match.

Custom Date Exact

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
CD[N]001 Custom Date N exact	100	Exact	Custom Date N exact = true
CD[N]003 Custom Date N no data	0	No data	Custom Date N exact = no data
CD[N]004 Custom Date N conflict	-5	Conflict	*

Custom Date Fuzzy

The below table shows the full details of the rules that are available within this compound comparison.

Result	Score	Category	Comparisons
CD[N]001 Custom Date N exact	100	Exact	Custom Date N exact = true
CD[N]002 Custom Date N similar	60	Fuzzy	Custom Date N similar = true
CD[N]003 Custom Date N no data	0	No data	Custom Date N exact = no data
CD[N]004 Custom Date N conflict	-5	Conflict	*

Scores

There will be one overall score created from the above compound comparisons. It is possible to override the weighting and enablement of each compound comparison in either the message header (real-time) or the run profile, in order to alter the scores received without even needing to modify the match process itself.

Compound Comparison	Weighting	Enabled
Product Description	15	Y

Compound Comparison	Weighting	Enabled
Product Name	9	Y
Model Number	15	Y
MPN	15	Y
Price	1	Y
customstring1exact	1	N
customstring2exact	1	N
customstring3exact	1	N
customstring4exact	1	N
customstring5exact	1	N
customstring6exact	1	N
customstring7exact	1	N
customstring8exact	1	N
customstring9exact	1	N
customstring10exact	1	N
customstring1fuzzy	1	N
customstring2fuzzy	1	N
customstring3fuzzy	1	N
customstring4fuzzy	1	N
customstring5fuzzy	1	N
customstring6fuzzy	1	N
customstring7fuzzy	1	N
customstring8fuzzy	1	N
customstring9fuzzy	1	N
customstring10fuzzy	1	N
customnumber1exact	1	N
customnumber2exact	1	N
customnumber3exact	1	N
customnumber4exact	1	N
customnumber5exact	1	N
customnumber6exact	1	N
customnumber7exact	1	N
customnumber8exact	1	N
customnumber9exact	1	N
customnumber10exact	1	N
customnumber1fuzzy	1	N
customnumber2fuzzy	1	N
customnumber3fuzzy	1	N
customnumber4fuzzy	1	N

Compound Comparison	Weighting	Enabled
customnumber5fuzzy	1	N
customnumber6fuzzy	1	N
customnumber7fuzzy	1	N
customnumber8fuzzy	1	N
customnumber9fuzzy	1	N
customnumber10fuzzy	1	N
customdate1exact	1	N
customdate2exact	1	N
customdate1fuzzy	1	N
customdate2fuzzy	1	N

Rules

This section describes the default set of provided match rules, used to determine whether two records are considered a definite match, a possible match, or not a match.

UID Matches

These rules ensure that any records that have an exact UID in common are considered an automatic match with a score of 100.

Comparison	Value	Score	Decision
UID1 exact	true	100	Match

These rules ensure that any records that have an EID that is different are considered not to be matching.

Comparison	Value	Score	Decision
UID2 exact	true	100	Match

UID3 Exact

Comparison	Value	Score	Decision
UID3 exact	true	100	Match

EID Eliminations

Eliminate on EID1

Comparison	Value	Score	Decision
EID1 exact	true	0	No Match

Eliminate on EID2

Comparison	Value	Score	Decision
EID2 exact	true	0	No Match

Eliminate on EID3

Comparison	Value	Score	Decision
EID3 exact	true	0	No Match

Inverted EID Eliminations

These rules ensure that any records that have an IEID that is the same are considered not be matching.

Eliminate on IEID1

Comparison	Value	Score	Decision
IEID1 exact	false	0	No Match

Eliminate on IEID2

Comparison	Value	Score	Decision
IEID2 exact	false	0	No Match

Eliminate on IEID3

Comparison	Value	Score	Decision
IEID3 exact	false	0	No Match

Overall Score

These rules use the Overall Score compound score to find matches between the product data.

Strong Match

Comparison	Value	Output Score	Decision
Overall score	>=90	Taken from "overall score" compound score.	Match

Intermediate Match

Comparison	Value	Output Score	Decision
Overall score	≥ 70	Taken from “overall score” compound score.	Review

Weak Match

Comparison	Value	Output Score	Decision
Overall score	≥ 0	Taken from “overall score” compound score	Review

Match Relationship Processing

EDQ-PDS processes the relationships (matches) output by the matching process to add in additional information useful for external data stewardship. In particular, it adds a constructed 'rule name' made up of the attributes that have played a part in forming the match, unless the match was formed using a UID, in which case the rule used in the match process is sufficient to describe the match.

10

EDQ-PDS Published Processors

This section contains the details of EDQ-PDS published processors.

Topic

- [Abbreviate Processor](#)
- [Extract Companies Processor](#)
- [Extract Materials Processor](#)
- [Extract Quantified Units of Measure Processor](#)

Abbreviate Processor

The Abbreviate Processor is designed to make an ID-like value from a product description string, and can also be used for general abbreviation of text strings. It takes a single string input and abbreviates it based upon various options, outputting a single string.

Table 10-1 Input Attributes

Attribute Name	Data Type	Description	Mandatory
Input String	String	The input to abbreviate	Y

Table 10-2 Output Attributes

Attribute Name	Data Type	Description
Input.Abbreviated	String	Abbreviated input

Table 10-3 Options

Option Name	Data Type	Default	Description
Characters to replace pre abbreviation	Reference Data	Abbreviate – Characters to Replace	Individual characters to replace pre abbreviation. Default reference data in PDS contains data to standardize accented characters and to replace certain punctuation characters with spaces.
Characters to strip pre abbreviation	Reference Data	Abbreviate – Characters to Strip	Characters to strip pre abbreviation. Default reference data strips diacritics and remaining punctuation characters.

Table 10-3 (Cont.) Options

Option Name	Data Type	Default	Description
Words to replace pre abbreviation	Reference Data	Abbreviate – Word Replacements	Individual words to replace pre abbreviation.
Remove vowels in the middle or end of words? (Y/N)	String	Y	Whether to remove vowels in the middle or end of words. Only applies to those tokens which are of sufficient length (according to Minimum word length setting), and which do not contain any numeric characters.
Remove vowels at the start of words? (Y/N)	String	N	Whether to remove vowels at the start of words. Only applies to those tokens which are of sufficient length (according to Minimum word length setting), and which do not contain any numeric characters.
Replace double consonants with single? (Y/N)	String	Y	Whether to replace double consonants with single. Only applies to those tokens which are of sufficient length (according to Minimum word length) setting, and which do not contain any numeric characters.
Abbreviate tokens containing numeric characters (Y/N)	String	N	Whether to apply the abbreviation options to tokens containing numeric characters.
Don't abbreviate words of this number of	Integer	<Blank>	Minimum length for a token for it to have abbreviation options (first 3 options above) applied.
Truncate words of more than this number of characters (after abbreviation)	Integer	<Blank>	Maximum length for a token after it has been processed according to other options. Any tokens longer than this length will be truncated to this length (truncated characters removed from the end).
Standardize words	Reference Data	<Blank>	Reference data for replacing words prior to abbreviation (but after simple standardization/normalization processing).
Separator for output words	String	<Blank>	Separator to be output between tokens in the output string.

Table 10-4 Examples

Input	Remove vowels at Start	Remove vowels middle/end	Replace double consonants	Abbreviate tokens with numerics	Max token length	Min token length	Output delimiter	Output
Sony Bravia LCD TV, 47", Silver, Abacus	N	Y	Y	N	<Blank>	<Blank>		SNY BRV LCD TV 47 SLVR ABCS
TTB653 SDS Hexarmor Sharpmaster HV 7082 Needlestick-Resistant	N	Y	Y	N	8	<Blank>		TTB653SDS HXRMR SHRPSM ST HV 7082 NDLSTCKR GLVS

Extract Companies Processor

The Extract Companies processor is designed to extract potential company names from an input string. It produces two outputs, the company as extracted and a category for the retail area of this company (e.g. Home and Garden, Clothing and Fashion)

Table 10-5 Input Attributes

Attribute Name	Data Type	Description	Mandatory
Input String	String	The input to extract from	Y

Table 10-6 Output Attributes

Attribute Name	Data Type	Description
ExtractedCategories	String Array	Categories as relating to the extracted Companies.
ExtractedCompanies	String Array	Companies as extracted from the input.
RemainingInput	String	Any of the input string which was not extracted as a Company.

Table 10-6 (Cont.) Output Attributes

Attribute Name	Data Type	Description
Input.Standardized	String	This is a copy of the input string, its inclusion is an artefact of how published processors work.
AttributesExtractedFlag	String (Y/N) (Flag)	Whether any companies were extracted from this record - Yes (Y) or No (N).

Table 10-7 Options

Option Name	Data Type	Default	Description
Companies	Reference Data	Profile – Product Data – Companies	The Companies to extract, contains a lookup column with the original company to look for in the input string and a standardized version of that company.

Extract Materials Processor

The Extract Materials processor is designed to extract potential materials (wood, metal, cotton, etc) from an input string. It produces two outputs, the original material as extracted and a standardized version of this material.

Table 10-8 Input Attributes

Attribute Name	Data Type	Description	Mandatory
Input String	String	The input to extract from	Y

Table 10-9 Output Attributes

Attribute Name	Data Type	Description
StandardizedMaterials	String Array	Standardized Materials as relating to the extracted original Materials.
OriginalMaterials	String Array	Original Materials as extracted from the input.
RemainingInput	String	Any of the input string which was not extracted as a Material.

Table 10-9 (Cont.) Output Attributes

Attribute Name	Data Type	Description
Input.Standardized	String	This is a copy of the input string, its inclusion is an artefact of how published processors work.
AttributesExtractedFlag	String (Y/N) (Flag)	Whether any materials were extracted from this record - Yes (Y) or No (N).

Table 10-10 Options

Option Name	Data Type	Default	Description
Materials	Reference Data	Profile – Product Data – Materials	The Materials to extract, contains a lookup column with the original material to look for in the input string and a standardized version of that material.

Extract Quantified Units of Measure

The Extract Quantified Units of Measure processor is designed to extract potential units of size with a value attached to them (for example 100 ML, 240V, etc) from an input string. It produces three outputs, an array of the standardized type of the extracted measure with the size value attached (for example 100 MILLILITER), an array of the standardized units of measure only (for example MILLILITER), and an array of the quantities only.

Table 10-11 Input Attributes

Attribute Name	Data Type	Description	Mandatory
Input String	String	The input to extract from	Y

Table 10-12 Output Attributes

Attribute Name	Data Type	Description
StandardizedUnitsOfMeasure Array	String Array	The standardized units of measure as extracted from the input (e.g. MILLILITER, INCH).
QuantityOfMeasureArray	String Array	The quantities as extracted from the input (i.e. the numbers only).

Table 10-12 (Cont.) Output Attributes

Attribute Name	Data Type	Description
QuantifiedStandardizedUnitsOfMeasureArray	String Array	The extracted quantities along with the standardized units of measure, e.g. 10 INCH.

Table 10-13 Options

Option Name	Data Type	Default	Description
Regular Expressions to match	Reference Data	Profile – Product Data – Units of Measure Regex	Regular expressions which extract the numeric quantity of the measure along with the measure itself, mapped to the standardized version of this measure.

Table 10-14 Examples

Input	Standardized Units of Measure Array	Quantity of Measure Array	Quantified Standardized Units of Measure Array
100 MM x 200' * 300ft	{100 MILLIMETRE, 200 INCH, 300 FOOT}	{MILLIMETRE, INCH, FOOT}	{100, 200, 300}
WOMEN'S DRESS SIZE 10-12	{10-12 SIZE}	{SIZE}	{10-12}
½"	{1/2 INCH}	{INCH}	{1/2}

11

Configuration of EDQ-PDS Run Profile and Message Header Options

This section gives information on the various configuration parameters that can be overridden when using the real-time matching services by specifying parameters in the message header.

 **Note:**

All parameters affect the matching and scoring of products in the match service, except the 'keyprofile' parameter, which is used in the key generation service.

Name	Type	Default	Description	Run profile param
overallscore.productdescription.weighting	Positive numeric	15	Weighting for the productdescription compound comparison	phase.Product\Match.process.*.overallscore.productdescription.weighting
overallscore.productdescription.enabled	Boolean (Y/N)	Y	Whether to enable the productdescription compound comparison	phase.Product\Match.process.*.overallscore.productdescription.enabled
overallscore.productname.weighting	Positive numeric	9	Weighting for the productname compound comparison	phase.Product\Match.process.*.overallscore.productname.weighting
overallscore.productname.enabled	Boolean (Y/N)	Y	Whether to enable the productname compound comparison	phase.Product\Match.process.*.overallscore.productname.enabled
overallscore.price.weighting	Positive numeric	3	Weighting for the price compound comparison	phase.Product\Match.process.*.overallscore.price.weighting
overallscore.price.enabled	Boolean (Y/N)	Y	Whether to enable the price compound comparison	phase.Product\Match.process.*.overallscore.price.enabled
overallscore.mpn.weighting	Positive numeric	15	Weighting for the mpn compound comparison	phase.Product\Match.process.*.overallscore.mpn.weighting

Name	Type	Default	Description	Run profile param
overallscore.mpn.enabled	Boolean (Y/N)	Y	Whether to enable the mpn compound comparison	phase.Product\Match.process.*.overallscore.mpn.enabled
overallscore.modelnumber.weighting	Positive numeric	15	Weighting for the model number compound comparison	phase.Product\Match.process.*.overallscore.modelnumber.weighting
overallscore.modelnumber.enabled	Boolean (Y/N)	Y	Whether to enable the model number compound comparison	phase.Product\Match.process.*.overallscore.modelnumber.enabled
overallscore.customstring1exact.weighting	Positive numeric	1	Weighting for the custom string 1 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring1exact.weighting
overallscore.customstring1exact.enabled	Boolean (Y/N)	N	Whether to enable the custom string 1 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring1exact.enabled
overallscore.customstring1fuzzy.weighting	Positive numeric	1	Weighting for the custom string 1 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring1fuzzy.weighting
overallscore.customstring1fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom string 1 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring1fuzzy.enabled
overallscore.customstring2exact.weighting	Positive numeric	1	Weighting for the custom string 2 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring2exact.weighting
overallscore.customstring2exact.enabled	Boolean (Y/N)	N	Whether to enable the custom string 2 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring2exact.enabled
overallscore.customstring2fuzzy.weighting	Positive numeric	1	Weighting for the custom string 2 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring2fuzzy.weighting
overallscore.customstring2fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom string 2 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring2fuzzy.enabled
overallscore.customstring3exact.weighting	Positive numeric	1	Weighting for the custom string 3 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring3exact.weighting
overallscore.customstring3exact.enabled	Boolean (Y/N)	N	Whether to enable the custom string 3 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring3exact.enabled

Name	Type	Default	Description	Run profile param
overallscore.customstring3fuzzy.weighting	Positive numeric	1	Weighting for the custom string 3 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring3fuzzy.weighting
overallscore.customstring3fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom string 3 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring3fuzzy.enabled
overallscore.customstring4exact.weighting	Positive Numeric	1	Weighting for the custom string 4 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring4exact.weighting
overallscore.customstring4exact.enabled	Boolean (Y/N)	N	Whether to enable the custom string 4 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring4exact.enabled
overallscore.customstring4fuzzy.weighting	Positive Numeric	1	Weighting for the custom string 4 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring4fuzzy.weighting
overallscore.customstring4fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom string 4 exact compound comparison	phase.Product\Match.process.*.overallscore.customstring4fuzzy.enabled
overallscore.customstring5exact.weighting	Positive numeric	1	Weighting for the custom string 4 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring5exact.weighting
overallscore.customstring5exact.enabled	Boolean (Y/N)	N	Whether to enable the custom string 5 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring5exact.enabled
overallscore.customstring5fuzzy.weighting	Positive numeric	1	Weighting for the custom string 5 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring5fuzzy.weighting
overallscore.customstring5fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom string 5 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customstring5fuzzy.enabled
overallscore.customnumber1exact.weighting	Positive numeric	1	Weighting for the custom string 5 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber1exact.weighting
overallscore.customnumber1exact.enabled	Boolean (Y/N)	N	Whether to enable the custom number 1 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber1exact.enabled
overallscore.customnumber1fuzzy.weighting	Positive numeric	1	Weighting for the custom number 1 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber1fuzzy.weighting

Name	Type	Default	Description	Run profile param
overallscore.customnumber1fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom number 1 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber1fuzzy.enabled
overallscore.customnumber2exact.weighting	Positive numeric	1	Weighting for the custom number 2 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber2exact.weighting
overallscore.customnumber2exact.enabled	Boolean (Y/N)	N	Whether to enable the custom number 2 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber2exact.enabled
overallscore.customnumber2fuzzy.weighting	Positive numeric	1	Weighting for the custom number 2 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber2fuzzy.weighting
overallscore.customnumber2fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom number 2 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber2fuzzy.enabled
overallscore.customnumber3exact.weighting	Positive numeric	1	Weighting for the custom number 3 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber3exact.weighting
overallscore.customnumber3exact.enabled	Boolean (Y/N)	N	Whether to enable the custom number 3 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber3exact.enabled
overallscore.customnumber3fuzzy.weighting	Positive numeric	1	Weighting for the custom number 3 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber3fuzzy.weighting
overallscore.customnumber3fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom number 3 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber3fuzzy.enabled
overallscore.customnumber4exact.weighting	Positive numeric	1	Weighting for the custom number 4 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber4exact.weighting
overallscore.customnumber4exact.enabled	Boolean (Y/N)	N	Whether to enable the custom number 4 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber4exact.enabled
overallscore.customnumber4fuzzy.weighting	Positive numeric	1	Weighting for the custom number 4 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber4fuzzy.weighting
overallscore.customnumber4fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom number 4 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber4fuzzy.enabled

Name	Type	Default	Description	Run profile param
overallscore.customnumber5exact.weighting	Positive numeric	1	Weighting for the custom number 5 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber5exact.weighting
overallscore.customnumber5exact.enabled	Boolean (Y/N)	N	Whether to enable the custom number 5 exact compound comparison	phase.Product\Match.process.*.overallscore.customnumber5exact.enabled
overallscore.customnumber5fuzzy.weighting	Positive numeric	1	Weighting for the custom number 5 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber5fuzzy.weighting
overallscore.customnumber5fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom number 5 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customnumber5fuzzy.enabled
overallscore.customdate1exact.weighting	Positive numeric	1	Weighting for the custom date 1 exact compound comparison	phase.Product\Match.process.*.overallscore.customdate1exact.weighting
overallscore.customdate1exact.enabled	Boolean (Y/N)	N	Whether to enable the custom date 1 exact compound comparison	phase.Product\Match.process.*.overallscore.customdate1exact.enabled
overallscore.customdate1fuzzy.weighting	Positive numeric	1	Weighting for the custom date 1 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customdate1fuzzy.weighting
overallscore.customdate1fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom date 1 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customdate1fuzzy.enabled
overallscore.customdate2exact.weighting	Positive numeric	1	Weighting for the custom date 2 exact compound comparison	phase.Product\Match.process.*.overallscore.customdate2exact.weighting
overallscore.customdate2exact.enabled	Boolean (Y/N)	N	Whether to enable the custom date 2 exact compound comparison	phase.Product\Match.process.*.overallscore.customdate2exact.enabled
overallscore.customdate2fuzzy.weighting	Positive numeric	1	Weighting for the custom date 2 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customdate2fuzzy.weighting
overallscore.customdate2fuzzy.enabled	Boolean (Y/N)	N	Whether to enable the custom date 2 fuzzy compound comparison	phase.Product\Match.process.*.overallscore.customdate2fuzzy.enabled
matchthreshold	Numeric	70	Match threshold	phase.Product\Match.process.*.matchthreshold

Name	Type	Default	Description	Run profile param
keyprofile	String		The key profile to use.	phase.*.process.*.keyprofile

Configuring Batch Staging Area for External Integrations

EDQ-PDS is provided with a configured external data source which can be used to read from an external Oracle database table for the data input, and to write back to the same data source with the results.

If integrating with an external system this data source can be used, and the following configuration options can be changed in the provided run profile.

The `SERVERID` and `JOBID` columns are used to enable processing of multiple batch jobs in parallel so they need to be edited in the run profile accordingly prior to each job submission; if they are not needed then default values can be used.

```
##### Staging Data Configuration Parameters For Batch Jobs #####
# The JNDI data source name and table names may be different dependent on the
installation

# Set the ServerID and JobID you are using here - it will be used to obtain data for
that specific pair of criteria when processing batch data serverid = SERVERID and
jobid = JOBID

# Where clause for candidate snapshots, to obtain data for specific server and job
phase.*.snapshot.*.where = serverid = '${serverid}' AND jobid = '${jobid}'

# Export parameters for specific server and job
phase.*.process.*.serverid = ${serverid}
phase.*.process.*.jobid = ${jobid}

# JNDI data source name for staging schema in database
phase.*.snapshot.*.remotejndi = jdbc/edqstaging
phase.*.export.*.remotejndi = jdbc/edqstaging

# Table names for candidate staging tables (snapshots)
phase.*.snapshot.Product\Candidates.table_name = EDQPDS_CANDIDATES_PROD

# Table names for result staging tables (exports)
phase.*.export.Batch\ Matches.table_name = EDQPDS_MATCHES
phase.*.export.Batch\ Key\ Generation\ Results.table_name = EDQPDS_CLUSTER_KEYS

# Enablement of exports to result staging tables (these are turned off in PDS by
default, uncomment the below to enable)
# job.matchexportenabled = true
# job.keygenerationexportenabled = true
```

Since it is expected that most users of the EDQ-PDS product will not be integrating with external systems, the exports to the external staging tables are turned off by default. When integrating with an external system, these can either be turned on in the jobs themselves, or enabled in the run profile by setting:

```
job.matchexportenabled = true
job.keygenerationexportenabled = true
```