



Match Blueprints User's Guide

- SAP BusinessObjects Data Services 4.1 (14.1.0)

2012-07-20

Copyright

© 2012 SAP AG. All rights reserved. SAP, R/3, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP BusinessObjects Explorer, StreamWork, SAP HANA and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries. Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Decisions, Web Intelligence, Xcelsius, and other Business Objects products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Business Objects Software Ltd. Business Objects is an SAP company. Sybase and Adaptive Server, iAnywhere, Sybase 365, SQL Anywhere, and other Sybase products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Sybase, Inc. Sybase is an SAP company. Crossgate, m@gic EDDY, B2B 360°, B2B 360° Services are registered trademarks of Crossgate AG in Germany and other countries. Crossgate is an SAP company. All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary. These materials are subject to change without notice. These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

2012-07-20

Contents

Chapter 1	Overview.....	5
Chapter 2	Downloading Blueprint Packages.....	7
2.1	Blueprint versions.....	7
2.2	Available Match blueprints.....	7
2.3	Downloading and setting up blueprints.....	8
Chapter 3	Configuring and Running Jobs.....	11
3.1	Editing the datastore	11
3.1.1	Microsoft SQL Server	11
3.1.2	Other database types.....	12
3.2	Running the jobs.....	13
Index		15

Overview

We've identified a number of common data quality match scenarios that you are likely to perform with SAP BusinessObjects Data Services. For each scenario, we've included a blueprint that is already set up to solve the business problem in that scenario. Each blueprint contains the necessary project, jobs, data flows, file formats, sample data, template tables, and custom functions to run the jobs in your environment with only a few modifications.

You can download the blueprint packages from the SAP Community Network. On the website, we periodically post new and updated blueprints, custom functions, best practices, whitepapers, and other content. You can refer to this site frequently for updated content and use the forums to provide us with any questions or requests you may have. We've also provided the ability for you to upload and share any content that you've developed with the rest of the SAP BusinessObjects Data Services development community (for instructions on uploading content, see *Share Your Knowledge in SCN Topic Spaces* at <https://www.sdn.sap.com/irj/scn/submitcontent>).

Instructions for downloading and installing the content objects are also located on the SAP Community Network website.

Downloading Blueprint Packages

2.1 Blueprint versions

The following table shows the version of the Data Quality Management and Text Data Processing blueprints that can be used for SAP BusinessObjects Data Services. The blueprint version is displayed on the SAP BusinessObjects Data Services Blueprints page of the SAP Community Network website.

Data Services version	Blueprint version	Blueprints available
4.1	4.1	Data Quality Management
XI 4.0	XI 4.0	Data Quality Text Data Processing
XI 3.2	XI 3.2	Data Quality
XI 3.1	XI 3.0	Data Quality
XI 3.0	XI 3.0	Data Quality

2.2 Available Match blueprints

Each blueprints package contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving the matching process.

To see the contents of each blueprint, including jobs and custom functions, see the *Content Objects Summary*.

Blueprint	Description
Data Quality Management Blueprints – Match	Contains miscellaneous jobs configured to illustrate best practice settings for specific Data Quality Management matching use cases.

2.3 Downloading and setting up blueprints

These are the general steps for downloading and setting up Data Quality Management blueprint packages for SAP BusinessObjects Data Services.

Note:

These jobs include Data Cleanse. You should install the SAP-supplied person and firm cleansing package PERSON_FIRM before you import the Match .atl file.

1. To access the SAP Community Network website, go to <https://www.sdn.sap.com/irj/boc/blueprints> in your web browser.
2. Log into your account using your username and password, or create a new account.
3. Review the list of available blueprint packages and other content objects and their descriptions to decide which to download.
4. Select the blueprint package that you want to download. A new page opens.
5. Select **View Document**.
6. Save the .zip file to the Tutorial Files folder in your installed SAP BusinessObjects Data Services path. By default, this folder is installed to \Program Files\SAP BusinessObjects\Data Services\Tutorial Files for 32-bit Windows and \Program Files (x86)\SAP BusinessObjects\Data Services\Tutorial Files for 64-bit Windows. If you are running on UNIX, the Tutorial Files folder exists only on the Windows client workstation, and you should download the .zip file there.
7. In the Tutorial Files folder in Windows Explorer, right-click the .zip file and select to extract the compressed (zipped) folders to the Tutorial Files folder. For example, if you use WinZip for file compression, right-click the .zip file and select **WinZip > Extract to here**.
Extracting creates subfolders and places the files in the appropriate location. The .atl file is saved to the Data Quality Mgmt Samples folder, and the sample data files are saved to the Data Quality Mgmt Samples\Match folder.
8. In the Designer, import the dqm_blueprints_match.atl file. In the Passphrase window, enter dqm_blueprints_match and click **Import**. Click **OK** to close the warning window.
Importing the file adds a project called DataQualityMgmtBlueprintsMatch to your object library. The project contains jobs whose names begin with DqmBlueprintMatch and contain in their name the Data Quality Management use case that they illustrate. Each job contains a data flow with the same name as the job. The import also adds a datastore called DqmBlueprintsMatch to your object library,

and file formats whose names begin with DqmMatchIn and DqmMatchOut for the sample input and output data.

9. If you are running on UNIX, copy the input files to the job server machine and create the same folder structure that is on the Windows client workstation.
10. If you imported the blueprint `.atl` files using a Data ServicesDesigner on 32-bit Windows and use a job server on 64-bit Windows, then you must copy the blueprint files to the Data Services installation of the job server machine.

Related Topics

- [Editing the datastore](#)

Configuring and Running Jobs

3.1 Editing the datastore

After you download the blueprint .zip file to the appropriate folder, unzip it, and import the .atl file in the Designer, you must edit the DqmBlueprintsMatch datastore.

The database that you use for running the blueprints does not need to be the same database that is used for the SAP BusinessObjects Data Services repository. It can be a locally installed database system or any shared database system that you have access to create tables in and read from those tables.

The DqmBlueprintMatch_MatchDNB, DqmBlueprintMatch_SuppressDMA, and DqmBlueprintMatch_SuspectsForReview jobs write to flat files on your file system. However, the DqmBlueprintMatch_MatchDNBPrep, DqmBlueprintMatch_SuppressDMAPrep, and DqmBlueprintMatch_SuspectsForReviewPrep jobs create tables called DQM_BLUEPRINTS_MATCH_DMA, DQM_BLUEPRINTS_MATCH_DNB, DQM_BLUEPRINTS_MATCH_REVIEW_JOB_STATUS, and DQM_BLUEPRINTS_MATCH_REVIEW_SUSPECTS, which must exist before you run the corresponding jobs.

Related Topics

- [Microsoft SQL Server](#)
- [Other database types](#)

3.1.1 Microsoft SQL Server

If you have access to write and read data to tables in Microsoft SQL Server 2000, 2005, or 2008, complete the following steps.

1. Select the **Datastores** tab of the Local Object Library, right-click the DqmBlueprintsMatch datastore and select **Edit**.
2. In the **Edit Datastore DqmBlueprintsMatch** window, enter your repository connection information in place of the four **CHANGE_THIS** values.
3. Click **OK**. If the window closes without an error message, then the database is successfully connected.

4. Before you run the DqmBlueprintMatch_MatchDNB, DqmBlueprintMatch_SuppressDMA, and DqmBlueprintsMatch_SuspectsForReview jobs that access a database, run the DqmBlueprintMatch_MatchDNBPrep, DqmBlueprintMatch_SuppressDMAPrep and DqmBlueprintsMatch_SuspectsForReviewPrep jobs to create the tables.

After you edit the datastore, you should be able to run the blueprint jobs.

3.1.2 Other database types

If you have access to write and read data to tables in another database system (other than Microsoft SQL Server), complete the following steps.

1. Select the **Datastores** tab of the Local Object Library, expand the DqmBlueprintsMatch datastore, and expand the **Template Tables** subfolder.
2. Delete the template tables. Right-click the DQM_BLUEPRINTS_MATCH_DMA, DQM_BLUEPRINTS_MATCH_DNB, DQM_BLUEPRINTS_MATCH_REVIEW_JOB_STATUS, and DQM_BLUEPRINTS_MATCH_REVIEW_SUSPECTS tables, select **Delete**, and select **Yes** to confirm your selection.
3. Delete the datastore. Right-click DqmBlueprintsMatch, select **Delete**, and select **Yes** to confirm your selection.
4. Create a new datastore with the same name as the one you just deleted. In the **Datastores** tab of the Local Object Library, right-click in the white space and select **New**. In the **Datastore** name field, enter the name DqmBlueprintsMatch. In the Database type field, select your database system. Complete the remaining fields with the connection information to the database that you have access to.
5. Click **OK** to close the Create New Datastore window.
6. Open the DqmBlueprintMatch_MatchDNBPrep data flow and delete the target. Then add a new template table with the same name by selecting the Template Table icon from the buttons on the right menu and clicking the data flow canvas. In the Create Template window, enter DQM_BLUEPRINTS_MATCH_DNB for the Template name, and select the DqmBlueprintsMatch datastore in the In datastore field. Click **OK** to close the Create Template window. Connect the last transform to the template table.
7. Open the DqmBlueprintMatch_MatchDMAPrep data flow and delete the target. Then add a new template table with the same name by selecting the Template Table icon from the buttons on the right menu and clicking the data flow canvas. In the Create Template window, enter DQM_BLUEPRINTS_MATCH_DMA for the Template name, and select the DqmBlueprintsMatch datastore in the In datastore field. Click **OK** to close the Create Template window. Connect the last transform to the template table.
8. Open the DqmBlueprintMatch_SuspectsForReviewPrep data flow and delete the two targets. Then add a new template table with the same name by selecting the Template Table icon from the buttons on the right menu and clicking the data flow canvas. In the Create Template window, enter DQM_BLUEPRINTS_MATCH_REVIEW_SUSPECTS for the Template name, and select the DqmBlueprintsMatch datastore in the In datastore field. Click **OK** to close the Create Template window. Connect the top transform called CreateSuspectsTable to the template table. Repeat these

steps to create a template table called `DQM_BLUEPRINTS_MATCH_REVIEW_JOB_STATUS`, and connect the bottom transform called `CreateJobStatusTable` to the template table.

9. Before you run the `DqmBlueprintMatch_MatchDNB`, `DqmBlueprintMatch_SuppressDMA`, and `DqmBlueprintsMatch_SuspectsForReview` jobs that access a database, run the `DqmBlueprintMatch_MatchDNBPrep`, `DqmBlueprintMatch_SuppressDMAPrep` and `DqmBlueprintsMatch_SuspectsForReviewPrep` jobs to create the tables.

After you edit the datastore, you should be able to run the blueprint jobs.

3.2 Running the jobs

Before you run your jobs, you should have already completed the following tasks:

1. Copy the address cleanse reference files.
 - The `DqmBlueprintMatch_SuppressDMA` jobs use only the U.S. address cleanse reference files.
 - While the `DqmBlueprintMatch_MatchDNB` jobs contain some records from countries other than the United States, for the purpose of the blueprint you may copy only the U.S. address reference files.
 - The `DqmBlueprintMatch_SuspectsForReview` jobs contain records from Brazil, France, Germany, India, Mexico, and the United States.
 - The `DqmBlueprintMatch_NPerFirm` and all four of the `DqmBlueprintsMatch_Addr*` jobs include only records from the United States.
2. Install the SAP-supplied person and firm cleansing package `PERSON_FIRM` before running jobs that include data cleansing.
3. Set accurate values in the substitution parameter configuration `Configuration1`:
 - `$$RefFilesAddressCleanse`—Enter the path location where you copied the address cleanse reference files.

Index

A

about blueprints 5
available blueprints 7

B

blueprints
 about 5
 available 7
 downloading 8
 versions 7

D

Data Quality Management Match
 blueprints
 list of 7
datastore
 editing 11
downloading blueprints 8

J

jobs, running 13

M

Match blueprints
 list of 7

R

running jobs 13

S

SAP BusinessObjects Data Services
 Blueprints
 versions 7
SQL Server
 editing datastore 11

V

versions 7

