

BODS10

SAP Data Services: Platform and Transforms

SAP BusinessObjects - Data Services

Course Outline

Course Version: 96 Revision A

Course Duration: 3 Day(s)

Publication Date: 05-02-2013

Publication Time: 1551



Copyright

Copyright © SAP AG. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. Additionally this publication and its contents are provided solely for your use, this publication and its contents may not be rented, transferred or sold without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors.

Trademarks

- Microsoft®, WINDOWS®, NT®, EXCEL®, Word®, PowerPoint® and SQL Server® are registered trademarks of Microsoft Corporation.
- IBM®, DB2®, OS/2®, DB2/6000®, Parallel Sysplex®, MVS/ESA®, RS/6000®, AIX®, S/390®, AS/400®, OS/390®, and OS/400® are registered trademarks of IBM Corporation.
- ORACLE® is a registered trademark of ORACLE Corporation.
- INFORMIX®-OnLine for SAP and INFORMIX® Dynamic Server™ are registered trademarks of Informix Software Incorporated.
- UNIX®, X/Open®, OSF/1®, and Motif® are registered trademarks of the Open Group.
- Citrix®, the Citrix logo, ICA®, Program Neighborhood®, MetaFrame®, WinFrame®, VideoFrame®, MultiWin® and other Citrix product names referenced herein are trademarks of Citrix Systems, Inc.
- HTML, DHTML, XML, XHTML are trademarks or registered trademarks of W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.
- JAVA® is a registered trademark of Sun Microsystems, Inc.
- JAVASCRIPT® is a registered trademark of Sun Microsystems, Inc., used under license for technology invented and implemented by Netscape.
- SAP, SAP Logo, R/2, RIVA, R/3, SAP ArchiveLink, SAP Business Workflow, WebFlow, SAP EarlyWatch, BAPI, SAPPHIRE, Management Cockpit, mySAP.com Logo and mySAP.com are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world. All other products mentioned are trademarks or registered trademarks of their respective companies.

Disclaimer

THESE MATERIALS ARE PROVIDED BY SAP ON AN "AS IS" BASIS, AND SAP EXPRESSLY DISCLAIMS ANY AND ALL WARRANTIES, EXPRESS OR APPLIED, INCLUDING WITHOUT LIMITATION WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WITH RESPECT TO THESE MATERIALS AND THE SERVICE, INFORMATION, TEXT, GRAPHICS, LINKS, OR ANY OTHER MATERIALS AND PRODUCTS CONTAINED HEREIN. IN NO EVENT SHALL SAP BE LIABLE FOR ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR PUNITIVE DAMAGES OF ANY KIND WHATSOEVER, INCLUDING WITHOUT LIMITATION LOST REVENUES OR LOST PROFITS, WHICH MAY RESULT FROM THE USE OF THESE MATERIALS OR INCLUDED SOFTWARE COMPONENTS.

Contents

Course Overview	v
Course Goals	v
Course Objectives	v
Unit 1: Defining Data Services	1
Defining Data Services	1
Unit 2: Defining Source and Target Metadata	2
Defining Datastores in Data Services	2
Defining Data Services System Configurations	2
Defining a Data Services Flat File Format	2
Defining Datastore Excel File Formats	2
Unit 3: Creating Batch Jobs	3
Creating Batch Jobs	3
Unit 4: Troubleshooting Batch Jobs	4
Setting Traces and Adding Annotations	4
Using the Interactive Debugger	4
Setting up and Using the Auditing Feature	4
Unit 5: Using Functions, Scripts and Variables	5
Using Built-In Functions	5
Using Variables, Parameters and Scripts	5
Unit 6: Using Platform Transforms	6
Using Platform Transforms	6
Using the Map Operation Transform	6
Using the Validation Transform	6
Using the Merge Transform	6
Using the Case Transform	7
Using the SQL Transform	7
Unit 7: Setting Up Error Handling	8
Setting Up Error Handling	8

Unit 8: Capturing Changes in Data	9
Capturing Changes in Data.....	9
Using Source-Based Change Data Capture (CDC).....	9
Using Target-Based Change Data Capture (CDC).....	9
Unit 9: Using Text Data Processing	10
Using the Entity Extraction Transform.....	10
Unit 10: Using Data Services (Integrator) Platform Transforms	11
Using Data Services (Integrator) Platform Transforms	11
Using the Pivot Transform	11
Using the Data Transfer Transform and Performance Optimization	11

Course Overview

SAP BusinessObjects™ Data Integrator 4.0 enables you to integrate disparate data sources to deliver more timely and accurate data that end users in an organization can trust. In this three-day course, you will learn about creating, executing, and troubleshooting batch jobs; using functions, scripts and transforms to change the structure and formatting of data; handling errors; and capturing changes in data.

As a business benefit, by being able to create efficient data integration projects, you can use the transformed data to help improve operational and supply chain efficiencies, enhance customer relationships, create new revenue opportunities, and optimize return on investment from enterprise applications.

Target Audience

This course is intended for the following audiences:

- Solution consultants responsible for implementing data integration projects.
- Power users responsible for implementing, administering, and managing data integration projects.

Course Prerequisites

Required Knowledge

- Basic knowledge of ETL (Extraction, Transformation, and Loading) of data processes

Course Goals

This course will prepare the participant to:

- Stage data in an operational datastore, data warehouse, or data mart.
- Update staged data in batch mode
- Transform data for analysis

Course Objectives

After completing this course, the participant will be able to:

- Integrate disparate data sources
- Create, execute, and troubleshoot batch jobs
- Use functions, scripts, and transforms to modify data structures and format data
- Handle errors in the extraction and transformation process
- Capture changes in data from data sources using different techniques

Unit 1

Defining Data Services

Unit Overview

Data Integrator provides a graphical interface that allows you to easily create jobs that extract data from heterogeneous sources, transform that data to meet the business requirements of your organization, and load the data into a single location. The Data Services platform enables you to perform enterprise-level data integration and data quality functions. Quality functions are discussed in BODS30 Data Quality Services. This unit describes the Data Services platform and its architecture, Data Services objects and its graphical interface, the Data Services Designer.

Lesson: Defining Data Services

Lesson Objectives

After completing this lesson, the participant will be able to:

- Define Data Services objects
- Use the Data Services Designer interface

Unit 2

Defining Source and Target Metadata

Unit Overview

To define data movement requirements in Data Services, you must import source and target metadata. A datastore provides a connection or multiple connections to data sources such as a database. Through the datastore connection, Data Services can import the metadata that describes the data from the source. Data Services uses these datastores to read data from source tables or load data to target tables.

Lesson: Defining Datastores in Data Services

Lesson Objectives

After completing this lesson, the participant will be able to:

- Create various types of Datastores

Lesson: Defining Data Services System Configurations

Lesson Objectives

After completing this lesson, the participant will be able to:

- Define system configurations in Data Services

Lesson: Defining a Data Services Flat File Format

Lesson Objectives

After completing this lesson, the participant will be able to:

- Defining flat file formats as a basis for a Datastore

Lesson: Defining Datastore Excel File Formats

Lesson Objectives

After completing this lesson, the participant will be able to:

- Create a Data Services Excel file format

Unit 3

Creating Batch Jobs

Unit Overview

A data flow defines how information is moved from source to target. These data flows are organized into executable jobs, which are grouped into projects.

Lesson: Creating Batch Jobs

Lesson Objectives

After completing this lesson, the participant will be able to:

- Create a project
- Create and execute a job
- Create a data flow with source and target tables
- Use the Query transform

Unit 4

Troubleshooting Batch Jobs

Unit Overview

To document decisions and troubleshoot any issues that arise when executing your jobs, you can validate your jobs and their components and add annotations to your jobs, work flows and data flows. In addition, you can set various trace options and see the trace results in different logs. You can also use the Interactive Debugger as a method of troubleshooting. Setting up audit points, label, and rules help you to ensure the correct data is loaded to the target.

Lesson: Setting Traces and Adding Annotations

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use descriptions and annotations
- Setting traces on jobs

Lesson: Using the Interactive Debugger

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use the View Data Function
- Use the Interactive Debugger

Lesson: Setting up and Using the Auditing Feature

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use auditing in data flows

Unit 5

Using Functions, Scripts and Variables

Unit Overview

Data Services gives you the ability to perform complex operations using built-in functions. You can extend the flexibility and reusability of objects by writing scripts, custom functions, and expressions using the Data Services scripting language and variables.

Lesson: Using Built-In Functions

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use functions in expressions
- Use the search_replace function
- Use the lookup_ext function
- Use the decode function

Lesson: Using Variables, Parameters and Scripts

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use variables and parameters
- Use the Data Services scripting language
- Create a custom function

Unit 6

Using Platform Transforms

Unit Overview

Platform transforms are optional objects in a data flow that allow you to transform your data as it moves from source to target. In data flows, transforms operate on input data sets by changing them or by generating one or more new data sets. Transforms are added as components to your data flow in the same way as source and target objects. Each transform provides different options that you can specify based on the transform's function. You can choose to edit the input data, output data, and parameters in a transform.

Lesson: Using Platform Transforms

Lesson Objectives

After completing this lesson, the participant will be able to:

- Describe platform transforms

Lesson: Using the Map Operation Transform

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use the Map Operation transform in a data flow

Lesson: Using the Validation Transform

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use the Validation transform

Lesson: Using the Merge Transform

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use the Merge transform

Lesson: Using the Case Transform

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use the Case transform

Lesson: Using the SQL Transform

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use the SQL transform

Unit 7

Setting Up Error Handling

Unit Overview

If a Data Services job does not complete properly, you must resolve the problems that prevented the successful execution of the job. The best solution to data recovery situations is obviously not to get them in the first place. Some of those situations are unavoidable, such as server failures. Others, however, can easily be sidestepped by constructing your jobs so that they take into account the issues that frequently cause them to fail.

Lesson: Setting Up Error Handling

Lesson Objectives

After completing this lesson, the participant will be able to:

- Explain the levels of data recovery strategies
- Use recoverable alternative work flows using a try/catch block with a conditional

Unit 8

Capturing Changes in Data

Unit Overview

The design of your data warehouse must take into account how you are going to handle changes in your target system when the respective data in your source system changes. Data Services transforms provides you with a mechanism to do this. Slow Changing Dimensions (SCD) are dimensions, prevalent in data warehouses, that have data which changes over time. There are three methods of handling these SCDs: no history preservation, unlimited history preservation with new rows and limited history preservation.

Lesson: Capturing Changes in Data

Lesson Objectives

After completing this lesson, the participant will be able to:

- Update data which changes slowly over time

Lesson: Using Source-Based Change Data Capture (CDC)

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use source-based CDC (Change Data Capture)
- Use time stamps in source-based CDC
- Manage issues related to using time stamps for source-based CDC

Lesson: Using Target-Based Change Data Capture (CDC)

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use target-based CDC

Unit 9

Using Text Data Processing

Unit Overview

In this Information Technology age, we are all familiar with the massive explosion of digital data that we have seen in the last decades. In 2003, there were 5 exabytes of data, twice the amount from three years earlier (UC Berkeley). Digital information created, captured and replicated worldwide has grown tenfold in five years (IDC 2008). 95% of digital data is unstructured (IDC 2007). This is the native integration of the text analytics technology acquired in 2007. The Entity Extraction transform is a new feature of Data Services to bring text data onto the platform and preparing it for query, analytics, and reporting.

Lesson: Using the Entity Extraction Transform

Lesson Objectives

After completing this lesson, the participant will be able to:

- Using the Entity Extraction transform

Unit 10

Using Data Services (Integrator) Platform Transforms

Unit Overview

Data Services (Integrator) transforms are used to enhance your data integration projects beyond the core functionality of the platform transforms. These specific transforms perform key operations on data sets to manipulate their structure as they are passed from source to target.

Lesson: Using Data Services (Integrator) Platform Transforms

Lesson Objectives

After completing this lesson, the participant will be able to:

- Using the Data Services (Integrator) Platform transforms

Lesson: Using the Pivot Transform

Lesson Objectives

After completing this lesson, the participant will be able to:

- Use the Pivot transform

Lesson: Using the Data Transfer Transform and Performance Optimization

Lesson Objectives

After completing this lesson, the participant will be able to:

- Describe performance optimization
- Use the Data Transfer transform
- View SQL generated by a data flow