IBM InfoSphere Advanced DataStage - Advanced Data Processing V9.1

Information

Length:	2.0 Days
__Ref:	KM422G-X
Delivery method:	Classroom
Price:	EUR

Overview

This course is designed to introduce you to advanced parallel job data processing techniques in DataStage V9.1. In this course you will develop data techniques for processing different types of complex data resources including relational data, unstructured data (Excel spreadsheets), Hadoop HDFS ("big data") files, and XML data. In addition, you will learn advanced techniques for processing data, including techniques for masking data and techniques for validating data using data rules. Finally, you will learn techniques for updating data in a star schema data warehouse using the DataStage SCD (Slowly Changing Dimensions) stage. Even if you are not working with all of these specific types of data, you will benefit from this course by learning advanced DataStage job design techniques, techniques that go beyond those utilized in the DataStage Essentials course.

Public

This advanced course is for Experienced DataStage developers seeking training in more advanced DataStage job techniques and who seek techniques for working with complex types of data resources.

Prerequisits

You should:

- complete DataStage Essentials course **or** equivalent
- **and** have at least one year of experience developing parallel jobs using DataStage

Objective

- Use Connector stages to read from and write to database tables
- Handle SQL errors in Connector stages
- Use the Unstructured Data stage to extract data from Excel spreadsheets
- Use the Big Data stage to read from and write to Hadoop HDFS files
- Use the Data Masking stage to mask sensitive data processed within a DataStage job
- Use the XML stage to parse, compose, and transform XML data
- Use the Schema Library Manager to import and manage XML schemas

- Use the Data Rules stage to validate fields of data within a DataStage job
- Create custom data rules for validating data
- Design a job that processes a star schema data warehouse with Type 1 and Type 2 slowly changing dimensions
- Use the Surrogate Key Generator stage to generate surrogate keys

Topics

- Unit 1: Accessing Databases
- Unit 2: Processing Unstructured Data
- Unit 3: Processing Big Data
- Unit 4: Data Masking
- Unit 5: Processing XML Data
- Unit 6: Using Data Rules
- Unit 7: Updating a Star Schema Database

All units are accompanied by hands-on lab exercises.