IBM InfoSphere DataStage v11.5: Advanced Data Processing Training (2M423G)

Course Summary

Description

This two-day course is designed to introduce you to advanced parallel job data processing techniques in DataStage v11.5. In this course you will develop data techniques for processing different types of complex data resources including relational data, unstructured data (Excel spreadsheets), 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.

Topics

- Accessing databases
- Processing unstructured data
- Data masking
- Using data rules
- Processing XML data
- Updating a star schema database

Audience

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

Prerequisites

Before taking this course, students should have a taken a DataStage Essentials course or equivalent.

Duration

Two days

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically

IBM InfoSphere DataStage v11.5: Advanced Data Processing Training (2M423G)

Course Outline

I. Accessing databases
- A. Connector stage overview
- B. Connector stage functionality
- C. Error handling in Connector stages
- D. Multiple input links
- E. File Connector stage
- F. Handling database errors
- G. Parallel jobs with multiple Connector input links
- H. Using the File Connector stage to read and write
- I. HDFS files

II. Processing unstructured data
- A. Using the Unstructured Data stage in DataStage jobs
- B. Processing unstructured data

III. Data masking
- A. Using the Data Masking stage in DataStage jobs
- B. Data masking

IV. Using data rules
- A. Introduction to data rules
- B. Use the Data Rules stage to valid foreign key references in source data
- C. Create custom data rules
- D. Using data rules

V. Processing XML data
- A. Introduction to the Hierarchical stage
- B. Composing XML data
- C. Writing Hierarchical data to a relational table
- D. Using the Regroup step
- E. Consuming XML data
- F. Transforming XML data
- G. Importing XML schemas
- H. Compose hierarchical data
- I. Consume hierarchical data
- J. Transform hierarchical data

VI. Updating a star schema database
- A. Surrogate keys
- B. Slowly Changing Dimensions (SCD) stage
- C. Build a parallel job that updates a star schema database with two dimensions

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically