



IBM INFOSPHERE DATASTAGE V11.5 - ADVANCED DATA PROCESSING

Course Code: 4282

This 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.

What You'll Learn

- Use Connector stages to read from and write to database tables
- Handle SQL errors in Connector stages
- Use Connector stages with multiple input links
- Use the File Connector stage to access Hadoop HDFS data
- Optimize jobs that write to database tables
- Use the Unstructured Data stage to extract data from Excel spreadsheets
- Use the Data Masking stage to mask sensitive data processed within a DataStage job
- Use the Hierarchical 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

Who Needs to Attend

- Experienced DataStage developers who want learn about more advanced DataStage job techniques



Global Knowledge®

IBM INFOSPHERE DATASTAGE V11.5 - ADVANCED DATA PROCESSING

Course Code: 4282

CLASSROOM LIVE

\$1,700 USD

2 days

Classroom Live Outline

1. Accessing Databases

- Connector Stage Overview
- Connector Stage Functionality
- Error Handling in Connector Stages
- Multiple Input Links
- File Connector Stage
- Handling Database Error
- Parallel jobs with Multiple Connector Input Links
- Using the File Connector Stage to Read and Write HDFS Files

2. Processing Unstructured Data

- Using the Unstructured Data Stage in DataStage Jobs
- Extract Data from an Excel Spreadsheet
- Specify a Data Range for Data Extraction in an Unstructured Data Stage
- Specify Document Properties for Data Extraction
- Processing Unstructured Data

3. Data Masking

- Using the Data Masking stage in DataStage jobs
- Data Masking Techniques
- Data Masking Policies
- Applying Policies for Masquerading Context-Aware Data Types
- Applying Policies for Masquerading Generic Data Types
- Repeatable Replacement
- Using Reference Tables
- Creating Custom Reference Tables
- Data Masking

4. Using Data Rules

- Introduction to Data Rules
- Using the Data Rules Editor
- Selecting Data Rules
- Binding Data Rule Variables
- Output Link Constraints
- Adding Statistics and Attributes to the Output Information
- Use the Data Rules Stage to Valid Foreign Key References in Source Data
- Create Custom Data Rules
- Using Data Rules

5. Processing XML Data

- Introduction to the Hierarchical Stage
- Hierarchical Stage Assembly Editor
- Use the Schema Library Manager to Import and Manage XML Schemas
- Composing XML Data
- Using the HJoin Step to Create Parent-Child Relationships between Input Lists
- Using the Composer Step
- Writing Hierarchical Data to a Relational Table
- Using the Regroup Step
- Consuming XML Data
- Using the XML Parser Step
- Propagating Columns
- Transforming XML Data
- Using the Aggregate Step
- Using the Sort Step
- Using the Switch Step
- Using the H-Pivot Step
- Importing XML Schemas
- Compose Hierarchical Data
- Consume Hierarchical Data
- Transform Hierarchical Data

6. Updating a Star Schema Database

- Surrogate Keys
- Design a Job that Creates and Updates a Surrogate Key Source Key File from a Dimension Table
- Slowly Changing Dimensions (SCD) Stage
- Star Schema Databases
- SCD Stage Fast Path Pages
- Specifying Purpose Codes
- Dimension Update specification
- Design a Job that Processes a Star Schema Database with Type 1 and Type 2 Slowly Changing Dimensions
- Build a Parallel Job that Updates a Star Schema Database with Two Dimensions

Classroom Live Labs

- VMWare-based
- All x86/x64 VMWare training per person



Global Knowledge.

IBM INFOSPHERE DATASTAGE V11.5 - ADVANCED DATA PROCESSING

Course Code: 4282

CLASSROOM LIVE

\$1,590 USD

2 days



IBM INFOSPHERE DATASTAGE V11.5 - ADVANCED DATA PROCESSING

Course Code: 4282

VIRTUAL CLASSROOM LIVE

\$1,700 USD

2 days

Virtual Classroom Live Outline

1. Accessing Databases

- Connector Stage Overview
- Connector Stage Functionality
- Error Handling in Connector Stages
- Multiple Input Links
- File Connector Stage
- Handling Database Error
- Parallel jobs with Multiple Connector Input Links
- Using the File Connector Stage to Read and Write HDFS Files

2. Processing Unstructured Data

- Using the Unstructured Data Stage in DataStage Jobs
- Extract Data from an Excel Spreadsheet
- Specify a Data Range for Data Extraction in an Unstructured Data Stage
- Specify Document Properties for Data Extraction
- Processing Unstructured Data

3. Data Masking

- Using the Data Masking stage in DataStage jobs
- Data Masking Techniques
- Data Masking Policies
- Applying Policies for Masquerading Context-Aware Data Types
- Applying Policies for Masquerading Generic Data Types
- Repeatable Replacement
- Using Reference Tables
- Creating Custom Reference Tables
- Data Masking

4. Using Data Rules

- Introduction to Data Rules
- Using the Data Rules Editor
- Selecting Data Rules
- Binding Data Rule Variables
- Output Link Constraints
- Adding Statistics and Attributes to the Output Information
- Use the Data Rules Stage to Valid Foreign Key References in Source Data
- Create Custom Data Rules
- Using Data Rules

5. Processing XML Data

- Introduction to the Hierarchical Stage
- Hierarchical Stage Assembly Editor
- Use the Schema Library Manager to Import and Manage XML Schemas
- Composing XML Data
- Using the HJoin Step to Create Parent-Child Relationships between Input Lists
- Using the Composer Step
- Writing Hierarchical Data to a Relational Table
- Using the Regroup Step
- Consuming XML Data
- Using the XML Parser Step
- Propagating Columns
- Transforming XML Data
- Using the Aggregate Step
- Using the Sort Step
- Using the Switch Step
- Using the H-Pivot Step
- Importing XML Schemas
- Compose Hierarchical Data
- Consume Hierarchical Data
- Transform Hierarchical Data

6. Updating a Star Schema Database

- Surrogate Keys
- Design a Job that Creates and Updates a Surrogate Key Source Key File from a Dimension Table
- Slowly Changing Dimensions (SCD) Stage
- Star Schema Databases
- SCD Stage Fast Path Pages
- Specifying Purpose Codes
- Dimension Update specification
- Design a Job that Processes a Star Schema Database with Type 1 and Type 2 Slowly Changing Dimensions
- Build a Parallel Job that Updates a Star Schema Database with Two Dimensions

Virtual Classroom Live Labs

- VMWare-based
- All x86/x64 VMWare training per person



IBM INFOSPHERE DATASTAGE V11.5 - ADVANCED DATA PROCESSING

Course Code: 4282

VIRTUAL CLASSROOM LIVE

\$1,590 USD

2 days

Virtual Classroom Live Outline

1. Accessing Databases

- Connector Stage Overview
- Connector Stage Functionality
- Error Handling in Connector Stages
- Multiple Input Links
- File Connector Stage
- Handling Database Error
- Parallel jobs with Multiple Connector Input Links
- Using the File Connector Stage to Read and Write HDFS Files

2. Processing Unstructured Data

- Using the Unstructured Data Stage in DataStage Jobs
- Extract Data from an Excel Spreadsheet
- Specify a Data Range for Data Extraction in an Unstructured Data Stage
- Specify Document Properties for Data Extraction
- Processing Unstructured Data

3. Data Masking

- Using the Data Masking stage in DataStage jobs
- Data Masking Techniques
- Data Masking Policies
- Applying Policies for Masquerading Context-Aware Data Types
- Applying Policies for Masquerading Generic Data Types
- Repeatable Replacement
- Using Reference Tables
- Creating Custom Reference Tables
- Data Masking

4. Using Data Rules

- Introduction to Data Rules
- Using the Data Rules Editor
- Selecting Data Rules
- Binding Data Rule Variables
- Output Link Constraints
- Adding Statistics and Attributes to the Output Information
- Use the Data Rules Stage to Valid Foreign Key References in Source Data
- Create Custom Data Rules
- Using Data Rules

5. Processing XML Data

- Introduction to the Hierarchical Stage
- Hierarchical Stage Assembly Editor
- Use the Schema Library Manager to Import and Manage XML Schemas
- Composing XML Data
- Using the HJoin Step to Create Parent-Child Relationships between Input Lists
- Using the Composer Step
- Writing Hierarchical Data to a Relational Table
- Using the Regroup Step
- Consuming XML Data
- Using the XML Parser Step
- Propagating Columns
- Transforming XML Data
- Using the Aggregate Step
- Using the Sort Step
- Using the Switch Step
- Using the H-Pivot Step
- Importing XML Schemas
- Compose Hierarchical Data
- Consume Hierarchical Data
- Transform Hierarchical Data

6. Updating a Star Schema Database

- Surrogate Keys
- Design a Job that Creates and Updates a Surrogate Key Source Key File from a Dimension Table
- Slowly Changing Dimensions (SCD) Stage
- Star Schema Databases
- SCD Stage Fast Path Pages
- Specifying Purpose Codes
- Dimension Update specification
- Design a Job that Processes a Star Schema Database with Type 1 and Type 2 Slowly Changing Dimensions
- Build a Parallel Job that Updates a Star Schema Database with Two Dimensions

Virtual Classroom Live Labs

- VMWare-based

- All x86/x64 VMWare training per person

Sep 30 - Oct 1, 2019 | 10:00 AM - 6:00 PM EST

Oct 28 - 29, 2019 | 10:00 AM - 6:00 PM EST

Nov 25 - 26, 2019 | 10:00 AM - 6:00 PM EST

Jan 20 - 21, 2020 | 9:00 AM - 5:00 PM EST

Feb 17 - 18, 2020 | 9:00 AM - 5:00 PM EST

Mar 16 - 17, 2020 | 9:00 AM - 5:00 PM EST



IBM INFOSPHERE DATASTAGE V11.5 - ADVANCED DATA PROCESSING

Course Code: 4282

ON-DEMAND

\$720 USD

On-Demand Outline

1. Accessing Databases

- Connector Stage Overview
- Connector Stage Functionality
- Error Handling in Connector Stages
- Multiple Input Links
- File Connector Stage
- Handling Database Error
- Parallel jobs with Multiple Connector Input Links
- Using the File Connector Stage to Read and Write HDFS Files

2. Processing Unstructured Data

- Using the Unstructured Data Stage in DataStage Jobs
- Extract Data from an Excel Spreadsheet
- Specify a Data Range for Data Extraction in an Unstructured Data Stage
- Specify Document Properties for Data Extraction
- Processing Unstructured Data

3. Data Masking

- Using the Data Masking stage in DataStage jobs
- Data Masking Techniques
- Data Masking Policies
- Applying Policies for Masquerading Context-Aware Data Types
- Applying Policies for Masquerading Generic Data Types
- Repeatable Replacement
- Using Reference Tables
- Creating Custom Reference Tables
- Data Masking

4. Using Data Rules

- Introduction to Data Rules
- Using the Data Rules Editor
- Selecting Data Rules
- Binding Data Rule Variables
- Output Link Constraints
- Adding Statistics and Attributes to the Output Information
- Use the Data Rules Stage to Valid Foreign Key References in Source Data
- Create Custom Data Rules
- Using Data Rules

5. Processing XML Data

- Introduction to the Hierarchical Stage
- Hierarchical Stage Assembly Editor
- Use the Schema Library Manager to Import and Manage XML Schemas
- Composing XML Data
- Using the HJoin Step to Create Parent-Child Relationships between Input Lists
- Using the Composer Step
- Writing Hierarchical Data to a Relational Table
- Using the Regroup Step
- Consuming XML Data
- Using the XML Parser Step
- Propagating Columns
- Transforming XML Data
- Using the Aggregate Step
- Using the Sort Step
- Using the Switch Step
- Using the H-Pivot Step
- Importing XML Schemas
- Compose Hierarchical Data
- Consume Hierarchical Data
- Transform Hierarchical Data

6. Updating a Star Schema Database

- Surrogate Keys
- Design a Job that Creates and Updates a Surrogate Key Source Key File from a Dimension Table
- Slowly Changing Dimensions (SCD) Stage
- Star Schema Databases
- SCD Stage Fast Path Pages
- Specifying Purpose Codes
- Dimension Update specification
- Design a Job that Processes a Star Schema Database with Type 1 and Type 2 Slowly Changing Dimensions
- Build a Parallel Job that Updates a Star Schema Database with Two Dimensions

On-Demand Labs

- VMWare-based
- All x86/x64 VMWare training per person



IBM INFOSPHERE DATASTAGE V11.5 - ADVANCED DATA PROCESSING

Course Code: 4282

PRIVATE GROUP TRAINING

2 days

Visit us at www.globalknowledge.com or call us at 1-866-716-6688.

Date created: 9/16/2019 12:31:26 PM

Copyright © 2019 Global Knowledge Training LLC. All Rights Reserved.