KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing

Category: AI, Data & Analytics | **Vendor:** IBM

Duration: 16.00 hours (2 days) **13.0 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Online

Course Overview

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.

About This Course

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.

Who Should Attend

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

Learning Outcomes

Upon successful completion of this course, participants will be able to:

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

Detailed Course Outline

Accessing databases

- Connector stage overview
- - Use Connector stages to read from and write to relational tables
- - Working with the Connector stage properties
- Connector stage functionality
- - Before / After SQL
- - Sparse lookups
- - Optimize insert/update performance
- Error handling in Connector stages
- - Reject links
- - Reject conditions
- Multiple input links
- - Designing jobs using Connector stages with multiple input links
- - Ordering records across multiple input links
- File Connector stage
- - Read and write data to Hadoop file systems
- Demonstration 1: Handling database errors
- Demonstration 2: Parallel jobs with multiple Connector input links
- Demonstration 3: Using the File Connector stage to read and write HDFS files

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.
- Demonstration 1: Processing unstructured data

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
- Demonstration 1: Data masking

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
- Demonstration 1: Using data rules

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
- Topic 6: Transforming XML data
- - Using the Aggregate step
- - Using the Sort step
- - Using the Switch step
- - Using the H-Pivot step
- Demonstration 1: Importing XML schemas
- Demonstration 2: Compose hierarchical data
- Demonstration 3: Consume hierarchical data
- Demonstration 4: Transform hierarchical data

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
- Demonstration 1: Build a parallel job that updates a star schema database with two dimensions

Additional Course Details

Nexus Humans KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing training program is a workshop that presents an invigorating mix of sessions, lessons, and masterclasses meticulously crafted to propel your learning expedition forward.

This immersive bootcamp-style experience boasts interactive lectures, hands-on labs, and collaborative hackathons, all strategically designed to fortify fundamental concepts.

Guided by seasoned coaches, each session offers priceless insights and practical skills crucial for honing your expertise. Whether you're stepping into the realm of professional skills or a seasoned professional, this comprehensive course ensures you're equipped with the knowledge and prowess necessary for success.

While we feel this is the best course for the KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing course and one of our Top 10 we encourage you to read the course outline to make sure it is the right content for you.

Additionally, private sessions, closed classes or dedicated events are available both live online and at our training centres in Dublin and London, as well as at your offices anywhere in the UK, Ireland or across EMEA.

Frequently Asked Questions

Q: What delivery options are available for KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing?

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
- Traditional Instructor-Led Classroom Training (ILT)
- On-site delivery at your offices anywhere in United Kingdom
- Private dedicated courses customized for your team

Q: How many CPD hours does this course provide?

The 2-day KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing course provides up to 13.0 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing training?

The training takes place over 2 day(s), with each day lasting approximately 16.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing?

Yes, we provide corporate training, dedicated training, and closed classes for KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing. Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

Q: Why choose Nexus Human for KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your KM423 IBM InfoSphere DataStage v11.5 - Advanced Data Processing training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

Nexus Human

Professional Training & Development

Email: info@nexushuman.com

Website: www.nexushuman.com

Phone: +353 1 889 8700 (Ireland) | +44 20 8142 8961 (UK)

© 2026 Nexus Human. All rights reserved. This brochure was generated on 14/11/2026.