



KM204G: IBM InfoSphere DataStage Essentials 11.5

Course ID #: 0370-426-115-W

Hours: 28

Course Content

Course Description:

This course enables the project administrators and ETL developers to acquire the skills necessary to develop parallel jobs in DataStage. The emphasis is on developers. Only administrative functions that are relevant to DataStage developers are fully discussed. Students will learn to create parallel jobs that access sequential and relational data and combine and transform the data using functions and other job components.

Course Objectives:

After taking this course, you should be able to:

- Describe the uses of DataStage and the DataStage workflow
- Describe the Information Server architecture and how DataStage fits within it
- Describe the Information Server and DataStage deployment options
- Use the Information Server Web Console and the DataStage Administrator client to create DataStage users and to configure the DataStage environment
- Import and export DataStage objects to a file
- Import table definitions for sequential files and relational tables
- Design, compile, run, and monitor DataStage parallel jobs
- Design jobs that read and write to sequential files
- Describe the DataStage parallel processing architecture
- Design jobs that combine data using joins and lookups
- Design jobs that sort and aggregate data
- Implement complex business logic using the DataStage Transformer stage
- Debug DataStage jobs using the DataStage PX Debugger

Target Audience:

Project administrators and ETL developers responsible for data extraction and transformation using DataStage.

Prerequisites:

Basic knowledge of Windows operating system
Familiarity with database access techniques



KM204G: IBM InfoSphere DataStage Essentials 11.5

Course ID #: 0370-426-115-W

Hours: 28

Topics:

- 1. Introduction to DataStage**
- 2. Deployment**
- 3. DataStage Administration**
- 4. Work with Metadata**
- 5. Create Parallel Jobs**
- 6. Access Sequential Data**
- 7. Partitioning and Collecting Algorithms**
- 8. Combine Data**
- 9. Group Processing Stages**
- 10. Transformer Stage**
- 11. Repository Functions**
- 12. Work with Relational Data**
- 13. Control Jobs**