

Training: IBM

IBM InfoSphere DataStage engine Administration for Information Server



TRAINING GOALS:

This course teaches Information Server and/or DataStage administrators to configure, manage, and monitor the DataStage Engine which plays a crucial role in Information Server. It not only runs high performance parallel ETL jobs designed and built in DataStage. It also supports other Information Server products including Information Analyzer, QualityStage, and Data Click. After introducing DataStage parallel jobs and the Engine that runs them, the course describes DataStage project configuration, the Engine's development and runtime environments, and the Engine's data source connectivity. In addition the course explains how to import and export DataStage objects, how to run and monitor DataStage jobs through the command line and GUI, and how to use some important Engine utilities.

After completing this course, you should be able to:

- Describe how the DataStage Engine is used within Information Server
- Describe how the engine achieves high performance through partition parallelism
- Describe the DataStage job compilation process
- Describe the engine runtime process and environment
- Describe the purpose and format of the DataStage configuration file
- Describe how to configure the engine using environment variables
- Describe how to configure data source connections
- Describe the ways of running DataStage jobs
- Describe how to monitor DataStage jobs and the runtime environment
- Describe engine workload management
- Import and export DataStage jobs and other engine objects

This course is recommended for those who will be administering Information Server and DataStage.

CONSPECT:

Unit 1: Introduction to the Information Server (DataStage) engine

- Information Server architecture
- How the DataStage Engine is used in Information Server
- DataStage Engine features

Unit 2: Elements of a DataStage job

- Anatomy of a DataStage parallel job
- Explore the elements of an example DataStage job, including its stages and job parameters
- Understand the OSH that is generated by the job during compile

Unit 3: Engine architecture

- Partition parallelism

- Runtime architecture
- Configuration files
- The “Score”

Unit 4: Engine project configuration

- DataStage project authorizations
- Runtime Column Propagation (RCP)
- Environment variables

Unit 5: Configuring database connectivity

- Configuring ODBC data sources
- Configuring native database connections

Unit 6: Running DataStage jobs

- Running jobs from the command line
- Monitoring jobs in Director and the Operations Console
- Workload management
- Starting and stopping the Engine

Unit 7: Engine utilities

- Data set utilities
- Multiple job compile
- Resource estimation tool

Unit 8: Importing and exporting DataStage objects

- DataStage Designer exports and imports
- Command line exports and imports

REQUIREMENTS:

It is recommended that students take KM510 “IBM InfoSphere Administrative Tasks for Information Server v11.5” course.

Difficulty level

