

IBM InfoSphere Advanced DataStage Parallel Framework - V11.7

Course Summary

Description

This course covers advanced parallel job development techniques and performance and tuning in DataStage V11.7. In this course you will develop a deeper understanding of the DataStage architecture, including an understanding of the DataStage development and runtime environments. This will enable you to design parallel jobs that are robust, less subject to errors, reusable, and optimized for better performance.

Objectives

At the end of this course, students will be able to understand the parallel framework architecture, performance and tuning, and new features/differences in Version 11.7

Topics

- Describe the parallel processing architecture
 - Environmental Variables, Environment management, Auto purging
 - Runtime Column Propagation (RCP)
 - Add checkpoints for sequencer and NLS configuration
- Partitioning and collecting data
- Sorting data
- Buffering in parallel jobs
- Parallel framework data types
- Reusable components
- Balanced Optimization
- Unstructured stage
- Performance and Tuning Using Operations Console

Audience

Experienced DataStage developers seeking training in more advanced DataStage job techniques and who seek an understanding of the parallel framework architecture.

Prerequisites

Participants should have: IBM InfoSphere DataStage Essentials course or equivalent, and at least one year of experience developing parallel jobs using DataStage.

Duration

Three days

IBM InfoSphere Advanced DataStage Parallel Framework - V11.7

Course Outline

- I. Day 1*
 - A. Introduction to the parallel framework architecture
 - B. Compiling and executing jobs
 - C. Partitioning and collecting data
- II. Day 2*
 - A. Sorting data
 - B. Buffering in parallel jobs
 - C. Parallel framework data types
 - D. Reusable components
- III. Day 3*
 - A. Balanced Optimization
 - B. Unstructured Data stage
 - C. Performance and Tuning Using Operations Console