IBM InfoSphere Advanced DataStage - Parallel Framework 11.3

Information

Length:	3.0 Days
__Ref:	KM403G-X
Delivery method:	Classroom
Price:	AUD

Overview

This course is designed to introduce advanced parallel job development techniques in DataStage V9.1. In this course you will develop a deeper understanding of the DataStage architecture, including a deeper 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.

Public

This advanced course is designed for experienced DataStage developers seeking training in more advanced DataStage job techniques and who are seeking an understanding of the parallel framework architecture.

Prerequisits

You should have:

- DataStage Essentials course **or** equivalent **and** at least one year of experience developing parallel jobs using DataStage.

Objective

- Describe the parallel processing architecture
- Describe pipeline and partition parallelism
- Describe the role and elements of the DataStage configuration file
- Describe the compile process and how it is represented in the OSH
- Describe the runtime job execution process and how it is depicted in the Score
- Describe how data partitioning and collecting works in the parallel framework
- List and select partitioning and collecting algorithms
- Describe sorting in the parallel framework
- Describe optimization techniques for sorting
- Describe sort key and partitioner key logic in the parallel framework
- Describe buffering in the parallel framework

- Describe optimization techniques for buffering
- Describe and work with parallel framework data types and elements, including virtual data sets and schemas
- Describe the function and use of Runtime Column Propagation (RCP) in DataStage parallel jobs
- Create reusable job components using shared containers
- Describe the function and use of Balanced Optimization
- Optimize DataStage parallel jobs using Balanced Optimization

Topics

- Unit 1 - Introduction to the Parallel Framework Architecture
- Unit 2 - Compilation and Execution
- Unit 3 - Partitioning and Collecting Data
- Unit 4 - Sorting Data
- Unit 5 - Buffering in Parallel Jobs
- Unit 6 - Parallel Framework Data Types
- Unit 7 - Reusable components
- Unit 8 - Balanced Optimization