DataStage Intermediate

Course Summary

Description
The Intermediate course is designed to reinforce the Parallel nature of the datastage environment. A quick review of the parallel nature of the environment will be undertaken. More advanced concepts such as optimization of job design by using Multi-stage take-on of data. Breaking up job execution to take advantage of the parallel nature of the Enterprise engine. Examples such as performing both an aggregation of data and joining of datasets will be explored. Understand how the execution of jobs may be impacted by your APT Configuration file and fin-tuning of the APT file.

Objectives
At the end of this course, students will be able to:

- Clarify understanding of DataStage, all key concepts, parallel processing, partitioning and collecting of data.
- Verify you have everything in place prior to an installation.
- Build an APT configuration file.
- Understand the client-server architecture.
- Know how to be the Administrator and be able to setup projects appropriately.
- Use the Designer to create, edit and check configuration files.
- Identify typical job log messages and identify/debug warnings and errors.
- Understand the additional stages of aggregator join and merge and how they relate to partitions.
- Understand the Transformer, Filter, Switch and Aggregator stages of Enterprise Edition.
- Use an Oracle/DB2 grid with DataStage as it applies to Upsert and Design Update/Insert.

Topics

- Overview of ETL, DataStage and software available, DataStage Enterprise Edition tools (Admin, Designer, Director, and Web Administrator)
- Parallel Concepts (Application scalability, Explicit & Implicit parallelisms, Pipeline and partition parallelisms, Framework of engine)
- Datasets and their types (Virtual –links between stages & Persistent – physical datasets)
- Partitions and Collectors (review of types including hash and random and stages such as aggregator, join and merge which are dependent on partitioners)
- Importing and Exporting of data (how to connect metadata using plugins and ODBC)
- Overview of Enterprise Edition stages which include Transformer, Filter, Switch and Aggregator)
- Complex connection with Oracle, how to use an Oracle grid with DataStage, how Upsert vs. Design update and insert
- More Enterprise Edition stages and how to improve performance such as Row to Column conversion
- Create Subrecord.

Audience
This class is designed for designers in the Datastage Enterprise environment.

Prerequisites
Students should have a minimum of 6 months development experience in the DataStage environment.

Duration
Four days

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically