ProTech
protechtraining.com

"Charting the Course ...

... to Your Success!"

DataStage Get Started

Course Summary

Description

This course will introduce students to basic functions, capabilities and strengths of this software tool and how DataStage can be used to harness the power of parallel computing hardware for processing massive volumes of data in a minimum amount of time.

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

- Understand what DataStage is and what it can do for you.
- Understand application scalability as it relates to user and data resulting in increased speed.
- Learn how data from various computing architectures can be shared.
- Identify various types of parallelisms and learn about data pipelining and partitioning to generate a parallel dataflow.
- Select partition methods and partition data into smaller segments, process it independently and execute it in parallel with other processes.
- Collect data which will bring back data partitions into a sequential stream.
- Understand Megadata and how to read it from a sequential file, as well as, be able to build metadata for a sequential file.
- Understand Sequence and Copy stages of Enterprise Edition are and be able to identify Lookup, Join and Merge functionality.

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)
- Partitioners and Collectors (how to break up and collect data, how to improve the speed of processing, why 2 CPUs are better than one)
- Overview of Enterprise Edition stages (Sequence stage, copy stage, Row Generator Review examples of Lookup, Join and Merge functions)
- Debugging DataStage (Using the debugging tools within the environment, debugging job design)
- Understanding how Partitioning can impact debugging.

Audience

This course is designed for anyone new to the DataStage Environment.

Duration

Four days