Phone: 1-732-325-1126 | **E-mail:** nanduv@gmail.com

IBM WebSphere DataStage Course Contents:

Data Warehousing – An Overview

- Data Warehouse Definition
- OLTP vs OLAP
- Data Warehouse Process Overview
- Staging Area
- Datamart

DWH – Design Concept

- Star schema
- Snowflake schema
- Fact constellations

Data modeling techniques

- E/R modeling
- Dimensional modeling

Database design methodology for data warehouses

Case Study: Real time Retail industry scenario DWH design

Datastage tool Overview

- Datastage Server
- Datastage Parallel Extender
- Comparison between Datastage Server Vs Parallel

Introduction to Datastage PX

- DataStage Architecture
- DataStage Clients
    - Designer
    - Director
    - Administrator

DataStage Workflow (Runtime architecture)

Types of DataStage Job

- Parallel Jobs
- Server Jobs
- Job Sequence

Setting up DataStage Environment

- DataStage Administrator Properties
- Defining Environment Variables
- Importing Table Definitions

Creating Parallel Jobs

- Design a simple Parallel job in Designer
- Compile your job
- Run your job in Director
- View the job log
- Command Line Interface (dsjob)

Accessing Sequential Data

- Sequential File stage
- Data Set stage
- Create jobs that read from and write to sequential files
- Read from multiple files using file patterns
- Use multiple readers
- Null handling in Sequential File Stage

Platform Architecture

- Describe parallel processing architecture Describe pipeline &
- Partition parallelism
- List and describe partitioning and collecting algorithms
- Describe configuration files
- Explain OSH & Score

Sorting and Aggregating Data

- Sort data using in-stage sorts and Sort stage
- Combine data using Aggregator stage
- Remove Duplicates stage

Combining Data

- Combine data using the Lookup stage
- Combine data using merge stage
- Combine data using the Join stage
- Combine data using the Funnel stage

Transforming Data

- Understand ways DataStage allows you to transform data
- Create column derivations using user defined code and system functions
- Filter records based on business criteria
- Control data flow based on data conditions

Working with Relational Data

- Import Table Definitions for relational tables.
- Create Data Connections.
- Use SQL Builder to define SQL Select statements.
- Use SQL Builder to define SQL Insert and Update statements.
- Use the DB2 Enterprise stage.

Metadata in Parallel Framework

- Explain schemas.
- Create schemas.
- Explain Runtime Column Propagation (RCP).
- Build a job that reads data from a sequential file using a schema

Job Control

- Checkpoint/Restart
- Looping
- Expression Support
- User Variables Stage
- Automatically Handling Aborts
- Terminator Stage ~ aborting a sequence
- Enhanced Variable Support

- Enhancements to other activity stages:

Performance tuning techniques

DataStage Tips and Tricks

Best practices overview

- Standards
- Development guidelines
- Component usage
- DataStage Data Types
- Partitioning data
- Collecting data
- Sorting
- Stage specific guidelines