

- ❖ **Yes-M Systems** offers the unique opportunity to aspiring fresher's and experienced professionals to get real time experience in ETL Data warehouse tool – IBM DataStage.

- ❖ **Course Description**

With this training the professional will get critical knowledge and skills, to design and develop an efficient and scalable DataStage solution using minimum number of components. Also the professional can control data partitioning and provide minimum data skew to a complex enterprise level business problem.

The expert should be able to tune a parallel application to determine where bottlenecks exist and how to eliminate them. The trained professional can work with complex data import and export scenarios and understand the designing of parallel solutions without relying on automated insertion of partitioning and sort operations. The candidate clearly understands what an inserted buffer operation is, can enable environment variables to collect performance Statistics and can determine the appropriate response to the collected data.

- ❖ **Support**

- IBM Certification assistance.
- Resume and Interview preparation assistance.

-
- ❖ **Course Duration: 40 Hours**

 - ❖ **Customizable: Yes**

 - ❖ **Pre-requisites: Students should have strong knowledge in the following topics**
 - RDBMS

 - Data Warehousing concepts

❖ **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 Sequences

❖ **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

- ❖ ETL Architecture overview
- ❖ Loading the EDW
- ❖ Types of EDW tables
- ❖ ETL Transformation
- ❖ ETL Naming convention
- ❖ History and incremental data approach
- ❖ Record Versioning – SCD Types
- ❖ ETL Default values
- ❖ Error and reject record handling
 - Reject handling with the Sequential File stage
 - Reject handling with the Lookup stage
 - Reject handling with the Transformer stage
 - Reject handling with Target Database stages
- ❖ Error Reprocessing
- ❖ Rollback and restart ability

Other classes that we offer



-
- QA
 - Business Analyst
 - QA and BA as a single course (Integrated QA+BA) at a reduced fee.
 - Healthcare as a Domain Class
 - Healthcare BA
 - Certification QA
 - **QTP**
 - **ISTQB**
 - **Quality Center**
 - Java
 - .Net
 - Oracle Financials (11i and R12)
 - Java
 - Project Management
 - SAP FICO, SAP BASIS, SAP HR
 - SAP Testing, SAS and Advance LoadRunner classes

Contact us for details [Yes-M Systems, <http://myyesm.com>, 678-643-7777]