



# Datastage Training Course Content

## **Introduction to Data warehousing**

- What is Data warehousing and its purpose
- Architecture of Data warehousing
- OLTP Vs Data warehouse Applications
- Data Marts
- Data warehouse Lifecycle with Real-time examples
- Definitions
- ETL Process
- Types of Tables in D/W
- Types of FACTS tables
- Types of DIMENSION tables
- Types of Schemas in D/W
- What is Data Mart
- Warehouse Approaches

## **Data Modelling**

- Introduction to Data Modelling
- Entity Relationship model (E-R model)
- Data Modeling for Data Warehouse
- Dimensions and fact tables
- Star Schema and Snowflake Schemas
- Coverage Tables
- Fact less Tables
- What to look for in modelling tools
- Modelling tools

## **ETL Design process**

- Introduction to Extraction, Transformation & Loading
- Types of ETL Tools
- What to look for in ETL Tools
- Key tools in the market
- ETL Trends & New Solution Options

## **Data stage installation**

- Datastage Installation
- Prerequisites to install Datastage
- Installation process

## **DATASTAGE**

### **Introduction to Datastage version 8.x**

- Datastage Introduction
- IBM Information Server architecture
- Datastage within the IBM Information Server architecture
- Datastage components
- Datastage main functions
- Client components

### **Datastage Administrator:**

- Datastage project Administration
- Editing projects and Adding projects
- Deleting projects
- Cleansing up project files
- Auto purging
- Permissions to users
- Runtime Column Propagation
- Enable Remote Execution of Parallel jobs
- Add checkpoints for sequencer
- Project protect
- .APT Config file

### **Datastage Designer:**

- Introduction to Datastage Designer
- Partitioning Techniques
- Creating the Jobs
- Compiling and Run the Jobs
- Exporting and importing the jobs
- Parameter passing
- System(SMP) & Cluster system(MPP)
- Importing Method(Flat file, Txt, Xls and Database files)
- OSH Importing Method
- Configuration file

- Importing table definitions
- Importing flat file definitions
- Managing the meta data environment
- Dataset management
- Deletion of Dataset
- Importing jobs
- Exporting jobs(Back up)
- Configuration file view
- Explanation of Menu Bar
- Palette
- Passive stages
- Active stages
- Database stages
- Debug stages
- File stages
- Processing stages
- Multiple Instances
- Runtime Column Propagation(RCP)
- Job design overview
- Designer work area
- Annotations
- Creating jobs, deleting jobs
- Compiling jobs
- Batch compiling
- Aggregator stage ,Copy stage
- Change Capture stage, Compress stage
- Filter stage, Funnel stage
- Modify stage
- Join stage, Lookup stages
- Difference between join and Lookup stages
- Merge stage
- Difference between Lookup and Merge stages
- Remove duplicate stage
- Sort stage, Pivot stage
- Surrogate key stage, switch stage
- Types of Lookups
- Types of Transformer stages
- Basic transformer stage
- Transformer stage
- Null handling in Transformer stage

- If Then Else in Transformer
- Stage variables
- Constraints
- Derivations
- Peek stage, Head stage, Tail stage
- Job properties
- Local variables
- Functions in Transformers
- String, Date, Null handling functions
- All properties in all stages
- Slowly changing Dimensions (SCD)
- SCD Type-1
- SCD Type-2
- SCD Type-3
- Implementation of SCD Type-1 in Datastage
- Implementation of SCD Type-2 in Datastage

### **Datastage Director:**

- Introduction to Datastage director
- Datastage Director window
- Jobs status view
- Datastage director options
- Running Datastage jobs
- Validating a job
- Running a job
- Batch Running
- Stopping a job and resetting job
- Monitoring a job
- Job scheduling
- Uncheduling a job
- Rescheduling a job
- Deleting a job
- Unlocking jobs
- View Log file
- Clear log
- Fatal error description
- Warning description
- Info description
- Difference between Compile and Validate
- Difference between Validate and Run

### **JOB SEQUENCER:**

- Arrange job activities in Sequencer
- Triggers in Sequencer
- Reset method
- Recoverability
- Notification Activity
- Terminator Activity
- Wait for file Activity
- Start Look Activity
- Execute Command Activity
- Sequence

### **CONTAINERS:**

- Reusability
- Minimizing complexity
- Local container
- Shared container
- Some jobs in container

### **PARALLEL PROCESSING AND PARTIONING METHODS:**

- Parallel
- Pipeline Parallelism
- Partition Parallelism
- Partitioning and Collecting
- Configuration file
- Fast name, Pools, Resource Disk, Resource Scratch Disk
- Running Job with different nodes
- Symmetric Multi Processing
- Massively Parallel Processing
- Partition techniques
- Round Robin
- Random
- Hash
- Entire
- Same
- Modulus
- Range
- DB2
- Auto

- Datastage components
- Server components
- Clients components
- Datastage Server
- Datastage Repository
- Naming Standards of jobs
- Document preparation
- ETL specs preparation
- Unit test cases preparation

GREENS TECHNOLOGYS



Like and follow with us for more details