

Course Brochure

DWH – DATA STAGE

Overview

•A Data Warehouse is a storage area, which holds the complete data specific to the organization so that efficient queries can be built on it for Analysis and Reporting purpose. Data Warehouse is a subject-oriented database where data is associated to a single organizational process, often called entity. A Data Warehouse is integrated with various source databases, which provides enormous strategic implication for Business Intelligence. The Data Warehouse on the other hand does not cater to real time operational requirements of the enterprise. It is a storehouse of current and historical data extracted from either internal or external data source(s).

Pre-requisites

•Knowledge of Oracle is essential.

Applications

COURSE CONTENTS

Datawarehouse Fundamentals:

- ❖ Data warehouse
- ❖ Data mart
- ❖ Purpose of Data warehouse
- ❖ Datawarehouse tools
- ❖ ETL,OLAP and Modeling tools
- ❖ ETL Developer roles and responsibilities
- ❖ ETL specifications
- ❖ Staging areas
- ❖ ODS
- ❖ Validation
- ❖ Transformation Rules(Business Rules)
- ❖ Full Load
- ❖ Delta Load
- ❖ Dimensional Modeling and Relational Modeling
- ❖ Start Schema
- ❖ Snow flake schema
- ❖ Multi Star Schema
- ❖ Confirmed Dimension
- ❖ Factless Fact table
- ❖ Additive facts
- ❖ Semi additive facts
- ❖ Non-additive facts

Datastage 8.1

- ❖ Server Jobs and Parallel Jobs difference
- ❖ Parallel
- ❖ Pipeline Parallelism
- ❖ Partition Parallelism
- ❖ Partitioning and Collecting
- ❖ Configuration file
- ❖ Fastname,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 testcases preparation

Datastage Administrator

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

Datastage Director

- ❖ Introduction to Datastage director
- ❖ Datastage Director window
- ❖ Jobs status view
- ❖ Datastage director options

COURSE CONTENTS

- ❖ Running Datastage jobs
- ❖ Validating a job
- ❖ Running a job
- ❖ Batch Running
- ❖ Stopping a job and resetting job
- ❖ Monitoring a job
- ❖ Job scheduling
- ❖ Unscheduling a job
- ❖ Rescheduling a job
- ❖ Deleting a job
- ❖ Unlocking jobs
- ❖ View Logfile
- ❖ Clear log
- ❖ Fatal error description
- ❖ Warning description
- ❖ Info description
- ❖ Difference between Compile and Validate
- ❖ Difference between Validate and Run

Datastage Designer

- ❖ Introduction to Datastage Designer
- ❖ Importing table definitions
- ❖ Importing flat file definitions
- ❖ Managing the meta data environment
- ❖ Dataset management
- ❖ Deletion of Dataset
- ❖ Routines
- ❖ 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
- ❖ Validating jobs
- ❖ Parameter passing
- ❖ ODBC, SQL Server stages
- ❖ Dataset, sequential file, fileset and Lookup fileset
- ❖ Difference between dataset and Fileset
- ❖ Formats of sequential file
- ❖ Delimetter and final delimiter
- ❖ Multiple reader from sequential file
- ❖ Aggregator stage ,Copy stage
- ❖ Change Capture stage, Compress stage
- ❖ Decode stage, Encode stage, Difference 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

COURSE CONTENTS

- ❖ 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 T ype-1 in Datastage
- ❖ Implementation of SCD T ype-2 in Datastage

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
- ❖ Sequencer

Containers

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