IBM DATASTAGE 11.5 (5 DAYS)
By Dr. Vishwanath Rao

- ➢ **Introduction to DataStage**
  - List and describe the uses of DataStage
  - List and describe the DataStage clients
  - Describe the DataStage workflow
  - Describe the two types of parallelism exhibited by DataStage parallel jobs
  - Explain the purpose of the DataStage configuration file
- ➢ **Deployment**
  - Describe what a deployment domain consists of
  - Describe different domain deployment options
  - Describe the installation process
  - Start the Information Server
- ➢ **DataStage Administration**
  - Open the Information Server Web console
  - Create new users and groups
  - Assign Suite roles and Component roles to users and groups
  - Give users DataStage credentials
  - Log onto DataStage Administrator
  - Add a DataStage user and specify the user's role
  - Specify DataStage global and project defaults
  - List and describe important environment variables
- ➢ **Working With Metadata**

- Log onto DataStage Designer
- Navigate around DataStage Designer
- Import and export DataStage objects to a file
- Import a table definition for a sequential file

- ➢ **Creating Parallel Jobs**
  - Design a parallel job in DataStage Designer
  - Define a job parameter
  - Use the Row Generator, Peek, and Annotation stages in the job
  - Compile the job
  - Run the job
  - Monitor the job log
  - Create a parameter set and use it in a job

- ➢ **Accessing Sequential Data**
  - Use the Sequential File stage in a DataStage job
  - Read from a sequential file using the Sequential File stage
  - Write to a sequential file using the Sequential File stage
  - Create reject links from the Sequential File stage
  - Work with nulls in sequential files
  - Use the Data Set stage to read and write to a data set file

- ➢ **Partitioning and Collecting**
  - Describe parallel processing architecture
  - List and describe partitioning and collecting algorithms
  - View the configuration file used by a job
  - Describe the parallel job compilation process

- View the OSH
- View the Score

- ➢ **Combining Data**
  - Combine data using the Lookup stage
  - Define range lookups
  - Combine data using Merge stage
  - Combine data using the Join stage
  - Combine data using the Funnel stage

- ➢ **Sorting and Aggregating Data**
  - Sort data using in-stage sorts and Sort stage
  - Provide summary calculations using Aggregator stage
  - Remove duplicate rows using the Remove Duplicates stage

- ➢ **Transforming Data**
  - Use the Transformer stage in parallel jobs
  - Define constraints using the DataStage expression editor
  - Define derivations using the DataStage expression editor
  - Define stage variables and use them in constraints and derivations
  - Handle nulls within the Transformer
  - Use loop processing in the Transformer

- ➢ **Working with Relational Data**
  - Import table definitions for relational tables
  - Create data connections
  - Use ODBC and DB2 Connector stages in a job
  - Use SQL Builder to define SQL statements in Connector stages

- Use multiple input links into Connector stages to update multiple tables within a single
  transaction

- Create reject links from Connector stages to capture rows with SQL errors

- ➢ **Job Control**
  - Use the DataStage job sequencer to build a job that controls a sequence of jobs

  - Use Sequencer links and stages to control the sequence a set of jobs run in

  - Use Sequencer triggers and stages to control the conditions under which jobs run

  - Pass information in job parameters from the master controlling job to the controlled jobs

  - Define user variables

  - Enable restart

  - Handle errors and exceptions

- ➢ **Processing unstructured data**
  - List examples of unstructured data

  - Extract data from a Microsoft Excel spreadsheet

  - Specify a data range for data extraction in an Unstructured Data stage

• Specify document properties for data extraction

➢ **Data masking**

- Describe the DataStage Data Masking Pack

- Understand how to apply policies for masquerading context aware data types

- Understand how to apply policies for masquerading generic data types

- Understand how hash lookup policy works

- Create a data masking job
  ➢ **Using data rules**

- Use the Data Rules stage in a DataStage job to validate fields in source data

- Use the Data Rules stage to valid foreign key references in source data

- Create custom data rules
  ➢ **Processing XML data**

• Introduction to the Hierarchical stage
o Hierarchical stage Assembly editor
o Use the Schema Library Manager to import and manage XML schemas • Composing XML data
o Using the HJoin step to create parent-child relationships between input lists
o Using the Composer step

- Writing Hierarchical data to a relational table

- Using the Regroup step

- Consuming XML data

o Using the XML Parser step
o Propagating columns • Transforming XML data
o Using the Aggregate step o Using the Sort step
o Using the Switch step
o Using the H-Pivot step

➢ **Reusable components**

- Create a schema file

- Read a sequential file using a schema

- Describe Runtime Column Propagation (RCP)

- Enable and disable RCP

- Create and use shared containers