

IBM InfoSphere DataStage

Lesson 00:



©2016 Capgemini. All rights reserved.

The information contained in this document is proprietary and confidential. For Capgemini only.

Course Goals and Non Goals



➤ Course Goals

- At the end of this program, participants gain an understanding of basic concepts in ETL.

➤ Course Non Goals

- Implementation of ETL tools.



Pre-requisites



- Fair knowledge of DW concepts

Intended Audience



- Software Engineers and Senior Software Engineers



Table of Contents



➤ Day 1

- Lesson 1: Overview of Datastage
 - 1.1: Introduction
 - 1.2: DataStage History
 - 1.3: DataStage Architecture Overview
 - 1.4: Type of Datasatge produtcts and jobs
 - 1.5: Setting Up Your DataStage Environment

Table of Contents



➤ Day 2

- Lesson 2: Designing Jobs
- 2.1: Introduction to Designer
- 2.2: : Development Workflow and logging into Designer
- 2.3: Import metadata into the Repository
- 2.4: Defining job parameter and parameter Sets
- 2.5: Creating Parallel Jobs
- 2.6: Adding Job Documentation
- 2.7: Compiling a Job

➤ Lab 01

Table of Contents



➤ Day 3

- Lesson 3: Executing and Monitoring Jobs
 - 3.1: Introduction to Datastage Director
 - 3.2: Job execution – interface mode
 - 3.3: Job execution – command mode
 - 3.4: View Job log
 - 3.5: Job Scheduling and Monitoring

➤ Lab 02

Table of Contents



➤ Day 4

- Lesson 4: Various DataStage Stages and Examples
 - 4.1: Type of Stages
 - 4.2: Accessing Sequential Data
 - 4.3: Reading and Writing NULL Values to a Sequential File
 - 4.4: DataSet Stage
 - 4.5: Combining Data (Lookup,Join,Merge,Funnel)

➤ Lab 03, Lab 04

Table of Contents



➤ Day 5

- Lesson 4: Various DataStage Stages and Examples
 - 4.6: Sorting, Aggregating Data And Remove Duplicate
 - 4.7: Transforming Data
 - 4.8: Connectors Stages
 - 4.9: Building SQL to Write to a Table
 - 4.10: Change Capture Stage

➤ Lab 05, Lab 06, Lab 07

Table of Contents



➤ DAY 6

- Lesson 5: DataStage Administration
 - 5.1: Introduction to Administration
 - 5.2: Add/Remove Projects
 - 5.3: User and Group Management
- Lesson 6: Parallel Processing
 - 6.1: Parallelism in Datastage
 - 6.2: Types of Parallelism
 - 6.3: Configuration
 - 6.4: Partitioning and Collecting
 - 6.5: Partitioning Types

➤ Lab 08

Table of Contents



➤ DAY 7

- Lesson 7: Repository Functions, Metadata in the Parallel Framework, Job Control
 - 7.1: Searching the Repository
 - 7.2: Impact Analysis
 - 7.3: Job and Table Difference Reports
 - 7.4: Metadata in the Parallel Framework
 - 7.5: Shared Containers

➤ Lab 09, Lab 10

Table of Contents



➤ DAY 8

- Lesson 7: Repository Functions, Metadata in the Parallel Framework, Job Control
 - 7.6: Job Control
 - 7.7: Flow of Control Stages
 - 7.8: Error Handling
 - 7.9: Restart
- Lesson 8: New Features

➤ Lab 11

References



- Student material:
 - NA
- Book:
 - NA
- Web-site:
 - NA



Next Step Courses (if applicable)



- BI related tool training



Other Parallel Technology Areas



➤ NA