Talend : Data Integration Developer

Duration : 04 Days

Course Contents

Day 1

Getting started with Talend Data Integration

- Starting Talend Studio
- Creating your first Job
- Running a Job
- Using the component help
- Designing a Job using best practices
- Documenting a Job

Working with files

- Working with delimited files
- Working with hierarchical files

Working with databases

- Creating tables in MySQL databases
- Reading data from MySQL database tables
- Applying best practices

Using repository metadata

- Using delimited file metadata
- Using XML file metadata
- Using database metadata
- Using generic schemas
- Updating metadata

Processing data

- Mapping data using tMap
- Joining data using tMap
- Capturing join rejects
- Filtering data and capturing filtering rejects
- Using other data processing components

Day 2 :

Using contexts and context variables

- Creating a built-in context variable
- Connecting to databases using context variables
- Creating a context group in the repository
- Loading context variables from a flow

Building executables and Docker images from data integration Jobs

- Building a stand-alone Job
- Building a new version of the Job
- Building a Docker image

Controlling execution

- Managing files
- Processing files
- Managing Job execution using a master Job

Handling errors

- Detecting and handling basic errors
- Raising a warning

Day 3:

Working with web services

- Accessing a SOAP web service

Use case: Creating a master sales table from different data sources

- Setting up a customer table
- First challenge
- Setting up a sales table
- Joining data
- Performing calculations
- Second challenge
- Creating a master Job

Connecting to a remote repository

- Creating a remote connection

SVN in Studio

- Copying a Job to a branch
- Comparing Jobs
- Resetting a branch

Reference project

- Setting up and using a reference project

Day 4:

Remote Job execution

- Creating and running a Job remotely

Resource usage and basic debugging

- Using Memory Run to view real-time resource usage
- Debugging Jobs using Debug Run

Activity Monitoring Console (AMC)

- Configuring statistics and logging
- Using Activity Monitoring Console (AMC)

Parallel execution

- Writing large files
- Writing to databases
- Automatic parallelization
- Partitioning

Joblets

- Creating a Joblet from an existing Job
- Creating a Joblet from scratch
- Triggering Joblets

Unit test

- Creating a unit test

Change data capture

- Examining databases
- Configuring the CDC database
- Monitoring changes
- Updating a warehouse
- Resetting the database