Talend for Data Integration and Big Data Certification & Training

About Magna Infotech Career Services (MICS)

MICS is a leading e-learning platform providing live instructor-led, interactive online training. We cater to IT professionals and students across the globe in skill development in all the trending technologies, including *IoT, Blockchain, AI & Machine Learning, Robotics, Big Data, Data Sciences, Business Analytics, Business Intelligence, Database Technologies, Java & Mobile Technologies, System Engineering, Project Management, Programming and many more.*

About the Course

MICS's Talend for DI and Big Data Certification Training is designed to help you master Talend and Big Data Integration Platform using Talend Open Studio. It is a free open source ETL tool using which you can easily integrate all your data with your Data Warehouse and Applications, or synchronize data between systems. You'll also use Talend ETL tool with HDFS, Pig and Hive on real-life case studies.

Module 1: Talend – A Revolution in Big Data

Learning Objectives – In this module, you will get an overview of ETL Technologies and the reason why Talend is referred as the next Generation Leader in Big Data Integration. You will be introduced to various products offered by Talend Corporation till date and its relevance to Data Integration and Big Data. Further you will learn about the TOS (Talend Open Studio), its Architecture, GUI, and how to install TOS.

Topics

- ✓ Working with ETL
- ✓ Rise of Big Data
- ✓ Role of Open Source ETL Technologies in Big Data
- ✓ Comparison with other market leader tools in ETL domain
- ✓ Importance of Talend (Why Talend)
- ✓ Talend and its Products
- ✓ Introduction of Talend Open Studio
- ✓ TOS for Data Integration
- ✓ GUI of TOS with Demo

Hands On

- ✓ Creating a basic job

Module 2: Working with Talend Open Studio for DI

Learning Objectives – Learn to work with various types of Data Source, Target Systems supported by Talend, Metadata and how to read/write from popular CSV/Delimited file and fixed width file. Connect to a Database and read/write/update data and read complex source system like Excel and XML along with some of the basic components like tLog, tMap using TOS.

Topics

- ✓ Launching Talend Studio
- ✓ Working with different workspace directories
- ✓ Working with projects
- ✓ Creating and executing jobs
- ✓ Connection types and triggers
- ✓ Most frequently used Talend components [tJava, tLogRow, tMap]
- ✓ Read & Write Various Types of Source/Target Systems
- ✓ Working with files [CSV, XLS, XML, Positional]
- ✓ Working with databases [MySQL DB]Error handling [tWarn, tDie]
- ✓ Error handling [tWarn, tDie]
- ✓ Metadata management

Hands On

- ✓ Creating a Business Model
- ✓ Adding Components to a Job
- ✓ Connecting the Components
- ✓ Reading and writing Delimited File
- ✓ Reading and writing Positional File
- ✓ Reading and writing XML and Xls/Xlsx Files
- ✓ Connecting Database(MySQL)
- ✓ Retrieving Schema from the Database
- ✓ Reading from Database Metadata
- ✓ Retrieving data from a file and inserting it into the Database
- ✓ Deleting data from Database
- ✓ Working with Logs and Error

Module 3: Basic Transformations in Talend

Learning Objectives – In this module, you will understand Data Mapping and Transformations using TOS. In addition, you will learn how to filter and join various Data Sources and search and sort through them.

Topics

- ✓ Context Variables
- ✓ Data Mapping & tMap
- ✓ Built-in Functions
- ✓ tJava • Routines
- ✓ File Transformation
- ✓ Talend Processing Components
- ✓ tSortRow, tAggregateRow
- ✓ Perform Lookup operation using tJoin
- ✓ tFilter & tReplicate
- ✓ tRowGenerator
- ✓ SubJob (using tRunJob, tPreJob, tPostJob)

Hands On

- ✓ Embedding Context Variables
- ✓ Adding different environments
- ✓ Data Mapping using tMap
- ✓ Using functions in Talend such as:
- ✓ tJava
- ✓ tSortRow
- ✓ tAggregateRow
- ✓ tReplicate
- ✓ tFilter
- ✓ tRowGenerator
- ✓ Perform Lookup operations using tJoin
- ✓ Creating SubJob (using tRunJob, tPreJob, tPostJob)

Module 4: Advance Transformations and Executing Jobs Remotely in Talend

Learning Objectives – In this module, you will understand the Transformation and various steps involved in program looping of Talend, ways to search files in a directory and how to process them in a sequence. You will also learn to export and import Jobs and run them remotely.

Topics

- ✓ Various components of file management (like tFileList, tFileAchive, tFileTouch, tFileDelete)
- ✓ Type Casting (convert datatypes among source-target platforms)
- ✓ Looping components (like tLoop, tForeach)
- ✓ Using FTP components (like tFTPFileList, tFTPFileExists, tFTPGet, tFTPPut)
- ✓ Exporting and Importing Talend jobs
- ✓ How to schedule and run Talend DI jobs externally (using Command line)
- ✓ Parameterizing a Talend job from command line

Hands On

- ✓ Implementing File Management (like tFileList, tFileAchive, tFileTouch, tFileDelete)
- ✓ Type Casting (tConvert and tMap(using Expression Builder))
- ✓ Looping components (like tLoop, tForeach)
- ✓ Using FTP components (like tFTPFileList, tFTPFileExists, tFTPGet, tFTPPut)
- ✓ Exporting and Importing Talend Jobs
- ✓ Parameterizing a Talend Job from command line

Module 5: Big Data and Hadoop with Talend

Learning Objectives - Learn about Big Data and Hadoop concepts, such as HDFS (Hadoop Distributed File System) Architecture, MapReduce, leveraging Big Data through Talend and Talend & Big Data Integration. Learn to setup and use the Talend Open Studio for Big Data. In addition, learn to use Big Data connectors in TOS (Talend offers some 800+ connectors for Big Data environment) and access Hadoop Ecosystem from Talend.

Topics

- ✓ Big Data and Hadoop
- ✓ HDFS and MapReduce
- ✓ Benefits of using Talend with Big Data
- ✓ Integration of Talend with Big Data
- ✓ HDFS commands Vs Talend HDFS utility
- ✓ Big Data setup using Hortonworks Sandbox in your personal computer
- ✓ Explaining the TOS for Big Data Environment

Hands On

- ✓ Understanding Big Data concepts in Hadoop and HDFS
- ✓ Creating a Project and a Job
- ✓ Designing a workflow for a Hadoop job to implement basic Map transformation
- ✓ Connecting to HDFS
- ✓ Transferring files from local filesystem onto HDFS
- ✓ Importing data from HDFS into a Talend job using tHDFSInput
- ✓ Performing basic Map operation using tMap component to join data from two files
- ✓ Aggregate the data and create an expression on top of the results from Map component using tAggregateRow
- ✓ Finally store the results onto HDFS

Talend for Data Integration and Big Data Certification & Training

Module 6: Hive in Talend

Learning Objectives – In this module, you will learn Hive concepts and the setup of Hive environment in Talend. You will learn how to use Hive Big Data connectors in TOS and implement Use Cases using Hive in Talend.

Topics

- ✓ Hive and It's Architecture
- ✓ Connecting to Hive Shell
- ✓ Set connection to Hive database using Talend
- ✓ Create Hive Managed and external tables through Talend
- ✓ Load and Process Hive data using Talend
- ✓ Transform data from Hive using Talend

Hands On

- ✓ Understanding how Hive works in Hadoop
- ✓ We will connect TOS BD to HDFS using tHDFS Connection and to Hive using tHive Connection
- ✓ Create a table in Hive using Talend's tHive,CreateTable component
- ✓ Load data from HDFS into the Hive table using tHiveLoad
- ✓ We will use tMap and tAggregate to perform simple queries to join and transform data in Hive
- ✓ We will store the results onto HDFS and local filesystem1

Module 7: Pig and Kafka in Talend

Learning Objectives – In this module, you will learn the PIG concepts, the setup of Pig Environment in Talend and Pig Big Data connectors in TOS for Big Data and implement Use Cases using Pig in Talend.

Topics

- ✓ Apache Pig
- ✓ Pig Environment in Talend
- ✓ Creating Job using Pig Data Connectors
- ✓ Apache Kafka
- ✓ Creating Job using Kafka Components in TOS for Big data

Hands On

- ✓ Create a program in Pig to Join different data sources and perform aggregation
- ✓ Configuring single node single broker cluster in Kafka

Module 8: In-class Project

Learning Objectives – In this module, you will be developing a Project using Talend DI and Talend BD with MySQL, Hadoop, HDFS, Hive, Pig and Kafka.