JUX ONCAMPUS

JGI JAIN DEEMED-TO-BE UNIVERSITY | CENTER FOR SKILLS

CERTIFICATION IN BIG DATA ANALYTICS WITH HADOOP 1

Course Code : OCIT0012

Big data refers to the use of advanced data analytics methods that help extract value from large data sets both structured and unstructured. With the availability of large data sets, there is a need for tools to computationally analyse and help reveal patterns, trends, associations to make meaningful decisions.

Curriculum

Module 1: Introduction to Big Data

Rise of Big Data, Hadoop vs traditional systems, Hadoop Master-Slave architecture, HDFS Architecture, NameNode, DataNode, Secondary Node, JobTracker, TaskTracker.

Module 2: HDFS and MapReduce architecture

Core components of Hadoop, Anatomy of Read and Write data on HDFS, MapReduce architecture Flow, JobTracker and TaskTracker.

Module 3: Hadoop Configuration

Hadoop modes, Hadoop terminal commands, Cluster configuration, Web ports, Hadoop configuration files, Reporting, Recovery, MapReduce in action.

Module 4: Understanding Hadoop MapReduce framework

Overview of the MapReduce framework, Use cases of MapReduce, Anatomy of MapReduce program, Mapper/ Reducer class, Driver code and Combiner and Partitioner.

Module 5: Advance MapReduce - Part 1

Write your own Partitioner, Writing Map and Reduce, Map side/Reduce side Join, Distributed Join, Distributed cache, Counters and joining Multiple datasets in MapReduce.

Module 6: Advance MapReduce - Part 2

MapReduce internals, Input format, Custom input format, Writable and Comparable, Output format, files, Junit and MRUnit testing frameworks.

Module 7: Apache Pig

PIG vs MapReduce, PIG architecture & Data types, PIG Latin relational operators, PIG Latin Join and CoGroup, PIG Latin group and union, Describe, Explain, Illustrate, PIG Latin: File loaders & UDF.

Learning Outcomes

- Understand the concept of BigData
- Understand the concept of Hadoop
- Understand the internals of MapReduce and YARN
- Understand the different modes and distribution of Hadoop
- Write MapReduce job for word count
- Create one node Hadoop cluster

CERTIFICATION IN
BIG DATA ANALYTICS WITH HADOOP 2

Course Code : OCIT0013

Big data refers to the use of advanced data analytics methods that help extract value from large data sets both structured and unstructured. With the availability of large data sets, there is a need for tools to computationally analyse and help reveal patterns, trends, associations to make meaningful decisions.

Curriculum

Module 1: Apache Hive and HiveQL
What is Hive, Hive DDL - Create/Show database, Hive DDL - Create/Show/Drop tables, Hive DML – Load files & Insert data, Hive SQL - Select, Filter, Join, Group By, Hive architecture & components, Difference between Hive and RDBMS

Module 2: Advance HiveQL
Multi-Table Inserts, Joins, Grouping Sets, Cubes, Rollups, Custom map and Reduce scripts, Hive SerDe, Hive UDF, Hive UDAF.

Module 3: Apache Flume, Sqoop, Oozie
Sqoop - How Sqoop works, Sqoop architecture, Flume complex Flow – Multiplexing, Oozie - Simple/Complex flow, Oozie service/ Scheduler, Use cases - Time and data triggers

Module 4: NoSQL Databases
CAP theorem, RDBMS vs NoSQL, Key value stores: Memcached, Riak, Key Value stores: Redis, Dynamo DB, Column Family: Cassandra, HBase, Graph Store: Neo4J, Document Store: MongoDB, CouchDB.

Module 5: Apache Hbase
When/Why to use HBase, Hbase architecture/Storage, Hbase data model, Hbase families/ column families, Hbase master, HBase vs RDBMS, Access Hbase data.

Module 6: Apache Zookeeper
Zookeeper Data model, Znokde Types, What is zookeeper, Sequential Znodes, Installing and configuring, Running zookeeper, Zookeeper use cases.

Module 7: Hadoop 2.0, YARN, Mrv2
MapReduce limitations, HDFS 2: Architecture, HDFS 2: High availability, HDFS 2: Federation, YARN Architecture, Classic vs YARN, YARN multitenancy, YARN capacity scheduler.

Learning Outcomes

- Understand the concept of BigData
- Understand the internals of MapReduce and YARN
- Understand the concept of Hadoop
- Understand the different modes and distribution of Hadoop
- Create one node Hadoop cluster
- Write MapReduce job for word count