Course Curriculum:

1. > Understanding Big Data and Hadoop

**Learning Objectives -** In this module, you will understand Big Data, the limitations of the existing solutions for Big Data problem, how Hadoop solves the Big Data problem, the common Hadoop ecosystem components, Hadoop Architecture, HDFS, Anatomy of File Write and Read, Rack Awareness.

**Topics -** Big Data, Limitations and Solutions of existing Data Analytics Architecture, Hadoop, Hadoop Features, Hadoop Ecosystem, Hadoop 2.x core components, Hadoop Storage: HDFS, Hadoop Processing: MapReduce Framework, Anatomy of File Write and Read, Rack Awareness.

2. > Hadoop Architecture and HDFS

**Learning Objectives** - In this module, you will learn the Hadoop Cluster Architecture, Important Configuration files in a Hadoop Cluster, Data Loading Techniques.

**Topics** - Hadoop 2.x Cluster Architecture - Federation and High Availability, A Typical Production Hadoop Cluster, Hadoop Cluster Modes, Common Hadoop Shell Commands, Hadoop 2.x Configuration Files, Password-Less SSH, MapReduce Job Execution, Data Loading Techniques: Hadoop Copy Commands, FLUME, SQOOP.

3. > Hadoop MapReduce Framework - I

**Learning Objectives -**In this module, you will understand Hadoop MapReduce framework and the working of MapReduce on data stored in HDFS. You will learn about YARN concepts in MapReduce.

**Topics -**MapReduce Use Cases, Traditional way Vs MapReduce way, Why MapReduce, Hadoop 2.x MapReduce Architecture, Hadoop 2.x MapReduce Components, YARN MR Application Execution Flow, YARN Workflow, Anatomy of MapReduce Program, Demo on MapReduce.

4. > Hadoop MapReduce Framework - II

**Learning Objectives** - In this module, you will understand concepts like Input Splits in MapReduce, Combiner & Partitioner and Demos on MapReduce using different data sets.

**Topics** - Input Splits, Relation between Input Splits and HDFS Blocks, MapReduce Job Submission Flow, Demo of Input Splits, MapReduce: Combiner & Partitioner, Demo on de-identifying Health Care Data set, Demo on Weather Data set.

5. > Advance MapReduce

**Learning Objectives -**In this module, you will learn Advance MapReduce concepts such as Counters, Distributed Cache, MRunit, Reduce Join, Custom Input Format, Sequence Input Format and how to deal with complex MapReduce programs.

**Topics -**Counters, Distributed Cache, MRunit, Reduce Join, Custom Input Format, Sequence Input Format.

6. > Pig

**Learning Objectives -**In this module, you will learn Pig, types of use case we can use Pig, tight coupling between Pig and MapReduce, and Pig Latin scripting.

**Topics -** About Pig, MapReduce Vs Pig, Pig Use Cases, Programming Structure in Pig, Pig Running Modes, Pig components, Pig Execution, Pig Latin Program, Data Models in Pig, Pig Data Types.

Pig Latin : Relational Operators, File Loaders, Group Operator, COGROUP Operator, Joins and COGROUP, Union, Diagnostic Operators, Pig UDF, Pig Demo on Healthcare Data set.

7. > Hive

**Learning Objectives -** This module will help you in understanding Hive concepts, Loading and Querying Data in Hive and Hive UDF.

**Topics -**Hive Background, Hive Use Case, About Hive, Hive Vs Pig, Hive Architecture and Components, Metastore in Hive, Limitations of Hive, Comparison with Traditional Database, Hive Data Types and Data Models, Partitions and Buckets, Hive Tables(Managed Tables and External Tables), Importing Data, Querying Data, Managing Outputs, Hive Script, Hive UDF, Hive Demo on Healthcare Data set.

8. > Advance Hive and HBase

**Learning Objectives -**In this module, you will understand Advance Hive concepts such as UDF, dynamic Partitioning. You will also acquire in-depth knowledge of HBase, Hbase Architecture and its components.

**Topics -**Hive QL: Joining Tables, Dynamic Partitioning, Custom Map/Reduce Scripts, Hive : Thrift Server, User Defined Functions.

HBase: Introduction to NoSQL Databases and HBase, HBase v/s RDBMS, HBase Components, HBase Architecture, HBase Cluster Deployment.

9. > Advance HBase

**Learning Objectives -**This module will cover Advance HBase concepts. We will see demos on Bulk Loading , Filters. You will also learn what Zookeeper is all about, how it helps in monitoring a cluster, why HBase uses Zookeeper.

**Topics -**HBase Data Model, HBase Shell, HBase Client API, Data Loading Techniques, ZooKeeper Data Model, Zookeeper Service, Zookeeper, Demos on Bulk Loading, Getting and Inserting Data, Filters in HBase.

10. > Oozie and Hadoop Project

**Learning Objectives -**In this module, you will understand working of multiple Hadoop ecosystem components together in a Hadoop implementation to solve Big Data problems. We will discuss multiple data sets and specifications of the project. This module will also cover Flume & Sqoop demo and Apache Oozie Workflow Scheduler for Hadoop Jobs.

**Topics -**Flume and Sqoop Demo, Oozie, Oozie Components, Oozie Workflow, Scheduling with Oozie, Demo on Oozie Workflow, Oozie Co-ordinator, Oozie Commands, Oozie Web Console,