**INTRODUCTION TO HADOOP AND BIG DATA**

* What is Big Data?
* What are the challenges for Processing Big Data?
* What technologies Suports big data?
* What is Hadoop?
* Why Hadoop?
* Histry of Hadoop
* Use case of Hadoop
* RDBMS vs HADOOP
* When to use and when not to use Hadoop
* Ecosystem tour

**HDFS**: Hadoop File System

* Significance of HDFS in Hadoop
* Features oh HDFS
* 5 Daemons of Hadoop
* Name Node and its Functionality
* Data Node and its Functionality
* Secondary Name Node and its Functionality
* Job Tracker and its Functionality
* Task Tracker and its Functionality
* Data Storage in HDFS
* Introduction about blocks
* Data replication
* Accessing HDFS
* Download HADOOP
* Install and set-up of Hadoop

**MApe Reduce**:

* Map Reduce Story
* Map Reduce Architecture
* How Map Reduce Works
* developing Map Reduce
* Map Reduce Programing Model
* Diffrent phases of Map Rerduce Algorithm
* Diffrent Data types in Map Reduce
* How Write a Basic Map Reduce Program
* Driver code
* Maper code
* Reducer code
* creating I/O Formats in Map Reduce jobs
  + Input Formate
  + Value Input Formate
  + File input Formate

**PIG:**

* Introduction to Apache Pig
* MApReduce vs Apache Pig
* SQL vs Apache pig
* Diffrent datatypes in pig
* Modes of Execution in Pig
* Grunt Shell
* Loading Data
* Exploring Pig
* Latin Commands

**HIVE:**

* Hive Introduction
* Hive architecture
* Hive vs RDBMS
* HiveQl and the shell
* Managing Tables(External vs managed)
* Data types and schemas
* Partitioning and Buckets

**HBASE:**

* Architecture and schema design
* HBASE vs RDBMS
* HMAster and Region Servers
* column Families and Regions
* Write Pipeline
* Read pipeline
* HBAse Commands

**SQOOP:**

* Sqoop Introduction
* Why Sqoop - Sqoop vs Flume
* Key Features of Sqoop
* Sqoop Architecture & Working
* Sqoop Commands

**SCALA COURSE SYLLABUS**

* **Module 1 -  Introduction**
  1. Introduction to Scala
  2. Creating a Scala Doc
  3. Creating a Scala Project
  4. The Scala REPL
  5. Scala Documentation
* **Module 2 - Basic Object Oriented Programming**
  1. Classes
  2. Immutable and Mutable Fields
  3. Methods
  4. Default and Named Arguments
  5. Objects
* **Module 3 - Case Objects and Classes**
  1. Companion Objects
  2. Case Classes and Case Objects
  3. Apply and Unapply
  4. Synthetic Methods
  5. Immutability and Thread Safety
* **Module 4 - Collections**
  1. Collections overview
  2. Sequences and Sets
  3. Options
  4. Tuples and Maps
  5. Higher Order Functions
* **Module 5 - Idiomatic Scala**
  1. For expressions
  2. Pattern Matching
  3. Handling Options
  4. Handling Failures
  5. Handling Futures

**spark - COURSE SYLLABUS**

Module 1 - Introduction to Spark - Getting started

1. What is Spark and what is its purpose?
2. Components of the Spark unified stack
3. Resilient Distributed Dataset (RDD)
4. Downloading and installing Spark standalone
5. Scala and Python overview
6. Launching and using Spark’s Scala and Python shell

Module 2 - Resilient Distributed Dataset and DataFrames

1. Understand how to create parallelized collections and external datasets
2. Work with Resilient Distributed Dataset (RDD) operations
3. Utilize shared variables and key-value pairs

Module 3 - Spark application programming

1. Understand the purpose and usage of the SparkContext
2. Initialize Spark with the various programming languages
3. Describe and run some Spark examples
4. Pass functions to Spark
5. Create and run a Spark standalone application
6. Submit applications to the cluster

Module 4 - Introduction to Spark libraries

1. Understand and use the various Spark libraries

Module 5 - Spark configuration, monitoring and tuning

1. Understand components of the Spark cluster
2. Configure Spark to modify the Spark properties, environmental variables, or logging properties
3. Monitor Spark using the web UIs, metrics, and external instrumentation
4. Understand performance tuning considerations