**Hadoop Course Contents**

1. Introduction to BigData and What the use in real time (30 minutes)
2. Hadoop Basic Architecture, Yarn and How it is different from Traditional RDBMS systems. (1 hour)
3. Hdfs File system and basic commands (1 hour)
4. Sqoop

* a. what is Sqoop (30 minutes)
* b. Its Optimization Techniques (30 minutes)

1. Hive

* What is hive and query (30 minutes)
* Loading data, managed and external Table, Analytical Functions (45 minutes minutes)
* Hive optimizations (30 minutes)
* Hive vs Tez Vs Impala (15 minutes)
* Impala and its advantages. (30 minutes)

1. Oozie

* what is oozie and what is the need. (15 minutes)
* how to schedule our jobs in OOzie (45 minutes)

1. Kafka

* what is apache kafka and its architecture. (45 minutes)
* how to use Kafka in real time (30 minutes)

1. Different file system and its advantages (30 minutes)
2. Hadoop Interview Questions (1 hour)

Section 3 to 8 will have 1 or 2 hours of hands-on session for each, depending upon requirement.

**Spark Course Contents (using python)**

1. Introduction to spark, Hadoop Vs Spark and its use. (1 hour)
2. Spark Architecture and its components. (1 hour)
3. Different modules in Spark and its storage system. (30 minutes)
4. RDD (1 hour)
5. Data Frame and Spark SQL (2 hour)
6. Spark Streaming. (2 hour)
7. Spark Optimization Techniques (2 hours)
8. Spark Interview Questions (1 hour)

Section 4 to 8 will have 1 or 2 hours of hands-on session for each, depending upon requirement.