

1. Introduction to Big Data.
 - What is big data?
 - Limitation of existing systems.
 - Hadoop ecosystem
 - Understanding Hadoop 2.x component.
 - Performing read and write operations.
 - RAC awareness.
 - Installation of Hadoop in virtual machine.

2. Hadoop Architecture and understanding the underlying HDFS.
 - Hadoop Architecture.
 - Horizontal scaling.
 - Movement of only code and not data over network.
 - High availability
 - Scalability : Multiple Name node.
 - HDFS Commands
 - Hadoop configuration files.
 - Password less SSH

3. Hadoop Mapreduce framework.
 - How MapReduce is different from traditional way.
 - Hadoop 2.x Mapreduce architecture and component.
 - Understand processing part i.e. YARN
 - Mapreduce concept
 - Run the basic Mapreducer program.

4. Hadoop MapReduce Framework : go further
 - Understanding Input Splits
 - Mapreduce job submission flow.
 - Performance improvement using combiners.
 - Partitioners.
 - Mapreduce as a whole

5. Mapreduce Advanced.
 - Understanding counters.
 - Map Side Join
 - Reduce side Join
 - MR units.
 - Custom input formats.
 - Sequence file format

6. PIG

- Why and how PIG came into picture.
- Where PIG is a good fit.
- Where PIG should not be used.
- Conceptual data flow.
- Different versions of PIG execution.
- Data models in PIG.
- PIG relational operators
- UDF in PIG : Customized function in Java
- Describe, explain and illustrate.
- Hands on

7. HIVE

- Why and how HIVE came into picture.
- How is this different from PIG.
- Hive architecture and component.
- Where and where not HIVE to be used.
- Data type in HIVE.
- Perform basic HIVE operations.

8. Advance Hive

- HIVE UDF
- Joins in HIVE
- Dynamic partitioning.
- Create UDF for HIVE
- Performance Tuning

9. HBASE, Zookeeper and NoSQL Databases

- Understand NoSQL database
- Understand CAP theorem.
- Comparison of RDBMS and HBASE
- HBASE Architecture.
- How updated is implement on top of HDFS.
- Data model and physical storage in HBASE.
- Execute basic HBASE command
- Data loading techniques in HBASE.
- Understanding Zookeeper

10. Flume, Sqoop, OOZIE

- Implement Flume and Sqoop.
- Understand Oozie.

- Schedule job in Oozie.
- Oozie workflow.
- Demo on Oozie using MR, Pig, Hive, Sqoop

11. Apache Spark

- What is Apache Spark
- Spark Ecosystem
- Spark Components
- History of Spark and Spark Versions/Releases
- Spark a Polyglot
- What is Scala?
- Why Scala?
- SparkContext
- RDD
- Demo on Spark

12. Project