

Module	Key Concepts	Assignments
Introduction to Big Data	• Introduction to Big Data • Concepts of Big Data • Velocity • Volume • Variety • Facts about Big Data • Big Data Ecosystem Overview	
Cloud computing, and big data interaction	• Introduction: Know what economic advantages are driving the adoption of Cloud Computing • Virtualization, cloud computing, and big data interact to create exciting business opportunities • Principal architectural components of a cloud and their organization • Services and orchestration on cloud: IaaS, PaaS, and SaaS • Compare IaaS, PaaS, and SaaS • Services that major Cloud companies provide and how they provide them	
Apache Hadoop	• HDFS ◦ HDFS Architecture ◦ NameNode ◦ SecondaryNameNode ◦ DataNode ◦ Configuring Hadoop Environment on Ubuntu ◦ Different modes of installation ◦ Pseudo Distributed Mode ◦ Standalone mode ◦ Fully Distributed Mode ◦ HDFS Vs Regular File System	Preparing Big Data Environment
MapReduce	• MapReduce ◦ Introduction to MapReduce ◦ MapReduce Engine ◦ JobTracker ◦ TaskTracker ◦ MapReduce Programming Model ◦ Mapper Class	Hadoop MapReduce

Module	Key Concepts	Assignments
	○ Reducer Class	
Streaming Systems / Apache Storm	• Streaming • Advanced storm • Storm Internals • Introduction to Storm • Spark Streaming • YARN + Storm from scratch	Storm
NoSQL Databases	• CAP theorem & eventual consistency • Cassandra • Hive – The Data Warehouse in Hadoop ○ Concepts of Hive ○ Hive Architecture ○ Metastore ○ Driver ○ Thrift Server ○ Web Interface ○ JDBC / ODBC ○ CLI ○ Introduction to HQL (Hive Query Language) ○ The Hive Data Model ○ Partitions ○ Data types ○ Hive Configuration ○ Sample Hive Queries and commands	Working with Hive
NoSQL Databases, Spark	• HBase – Distributed Columnar Database ○ NoSQL Movement ○ HBase Architecture ○ Region Servers ○ HBase Storage ○ Introduction to Zookeeper ○ Entities of Zookeeper ○ Leader ○ Follower ○ Observer ○ Zookeeper Data Model	HBase

Module	Key Concepts	Assignments
	○ Configuring HBase and Zookeeper ○ HBase Examples ○ HBase Use Cases • Pig ○ Pig Execution Mode ○ Local Mode ○ MapReduce Mode ○ Pig Engine ○ Pig Latin Scripts ○ Interactive Mode ○ Batch Mode ○ Configuring PIG ○ Sample PIG Scripts • Tez • Spark	
Moving Data to and from HDFS	Sqoop and Flume • Introduction to Sqoop • Sqoop Connectors to RDBMS • Importing Data from Sqoop to Hive • Sqoop Commands • Introduction to Flume • Flume Data Model • Flume Examples • Use Cases of Sqoop and Flume	Using sqoop and flume to move data to and from HDFS
Hadoop Ecosystem and Components	• Mahout • Oozie • Phoenix • Squirrel • Cassandra • MongoDB	Assignement : MongoDB
Graph Processing and Machine Learning	• Graph processing ○ GraphX • Machine learning • Mahout • SparkML	Applications (Graph, Machine Learning)

