**UNIT 1**

1. **Evolution of Big Data.**
2. **Introduction of Big Data.**
3. **Best Practices for Big Data Analytics.**
   1. **Best Practices.**
      1. Define Your goal.
      2. Choose the right tools
      3. Clean and preprocess your data.
      4. User machine learning Algorithm.
      5. Visualize yours results.
      6. Monitor and optimize.
      7. Ensure data security and privacy.
      8. Stay up-to-date with latest technologies.
4. **Big Data Characteristics.**
   1. **Characterizes:**
      1. Volume
      2. Velocity
      3. Varity
      4. Veracity
5. **Validating The Promotion of the Value of Big Data.**
6. **Big Data use cases.**
   1. **Uses cases:** 
      1. Retail and E-commerce
      2. Healthcare.
      3. Finance
      4. Manufacturing and Supply Chain.
7. **Perception and Quantification of Value.**
8. **Understanding Big Data Storage**
   1. **Challenges in Big Data Storage**
      1. Volume.
      2. Velocity.
      3. Variety.
      4. Veracity.
   2. **Technologies and Approaches for Big Data Storages.**
      1. Distributed File System.
      2. Cloud Storage.

***UNIT 2***

1. **General Overview of High-Performance Architecture**
2. **HDFS (Hadoop Distributed File System):**
   1. **Components:**
      1. Name Node.
      2. Data Node.
      3. Block
      4. Secondary Name Node.
      5. Client.
3. **MapReduce:**
   1. **MapReduce Frameworks (type) :**
      1. Vanilla MapReduce.
      2. Apache Sparks.
4. **YARN (Yet Another Resource Negotiator**
   1. **Types:** 
      1. Vanilla YARN.
      2. Kubernetes.
5. **Big Data Overview Analysis of data at Rest-Hadoop Analytics:**
6. **Limitation of existing distributed systems:**
   1. **Limitations:**
      1. Complexity.
      2. Latency.
      3. Programing Complexity.
      4. Limited Support for SQL.
      5. Storage Overheads.
      6. Community Support’s
7. **Hadoop Approach.**
8. **Hadoop Architecture.**
   1. **Architecture:**
      1. Hadoop Distributed File System (HDFS).
         1. NameNode
         2. DataNode.
      2. Resources Management.
         1. YARN (Yet Another Resources Negotiator).
      3. Data Storage.
         1. HDFS
      4. Data Processing.
         1. Map Reduce.
         2. Apache Spark.
         3. Apache Hive.
         4. Apache Pig.
         5. Apache Tez
         6. Data Visualization and Reporting.
         7. Security.
         8. Monitoring and Management.
         9. Metadata Management.
9. **Distributed File System: HDFS and GPFS.**
   1. **HDFS**
      1. **Type:**
         1. Vanilla HDFS.
         2. HDFS Federation**.**
   2. **GPFS**
      1. **Type:**
         1. Standard Spectrum Scale.
         2. Spectrum Scale for AI.
10. **Internal of Hadoop MR Engine:**
    1. **Type:**
       1. Apache Spark
       2. Apache Tez