



Big Data – Detailed Syllabus

UNIT 1: Introduction to Big Data

Topics:

- Definition and Overview of Big Data
- Characteristics of Big Data – Volume, Velocity, Variety, Veracity, Value (5Vs)
- Traditional vs Big Data Systems
- Big Data Applications and Use Cases
- Challenges in Big Data Management
- Data Growth Trends and Sources of Big Data (Social Media, IoT, Sensors, etc.)

Learning Outcome:

Understand what Big Data is, why it matters, and its impact on industries.

UNIT 2: Big Data Technologies and Ecosystem

Topics:

- Big Data Architecture Overview
- Hadoop Ecosystem Overview
- HDFS (Hadoop Distributed File System) – Features and Architecture
- MapReduce Framework – Working Mechanism, Mapper & Reducer Functions
- YARN (Yet Another Resource Negotiator)
- Data Ingestion Tools – Sqoop, Flume
- Data Processing Tools – Pig, Hive

Practical:

- Install Hadoop in pseudo-distributed mode
- Perform basic HDFS operations
- Write and execute a MapReduce job

Learning Outcome:

Understand Hadoop components and perform basic Big Data operations.

UNIT 3: NoSQL Databases

Topics:

- Need for NoSQL Databases
- Comparison: RDBMS vs NoSQL

- Types of NoSQL Databases:
 - Key-Value (Redis)
 - Document-Oriented (MongoDB)
 - Column-Oriented (HBase)
 - Graph-Based (Neo4j)
- CAP Theorem (Consistency, Availability, Partition Tolerance)
- Data Modeling and CRUD Operations in NoSQL

Practical:

- Create and query collections in MongoDB
- Implement CRUD operations

Learning Outcome:

Learn how NoSQL databases handle unstructured data efficiently.

UNIT 4: Apache Spark Framework**Topics:**

- Introduction to Apache Spark
- Spark vs Hadoop MapReduce
- Spark Architecture (Driver, Executors, Cluster Manager)
- RDD (Resilient Distributed Datasets) Concepts
- Transformations and Actions in Spark
- Spark SQL and DataFrames
- Spark Streaming Overview
- Introduction to MLlib and GraphX

Practical:

- Run Spark applications using PySpark
- Perform basic transformations and actions

Learning Outcome:

Gain understanding of in-memory Big Data processing using Spark.

UNIT 5: Data Analytics and Visualization**Topics:**

- Introduction to Data Analytics
- Types of Analytics – Descriptive, Predictive, Prescriptive
- Data Preprocessing: Cleaning, Integration, Transformation

- Data Analysis Tools: R, Python (NumPy, Pandas)
- Data Visualization Tools: Tableau, Power BI, Matplotlib, Seaborn
- Big Data Analytics Life Cycle

Practical:

- Analyze datasets using Python or PySpark
- Create visualizations using Tableau/Power BI

Learning Outcome:

Learn to analyze and visualize Big Data for decision-making.

UNIT 6: Machine Learning with Big Data**Topics:**

- Introduction to Machine Learning Concepts
- Supervised vs Unsupervised Learning
- Integration of ML with Big Data Tools
- Using Spark MLlib for Machine Learning Tasks
- Case Studies:
 - Recommendation Systems
 - Sentiment Analysis
 - Fraud Detection

Practical:

- Build and train ML models in Spark MLlib
- Evaluate model performance

Learning Outcome:

Learn how to apply ML techniques on large datasets using Big Data tools.

UNIT 7: Big Data in Cloud and Real-Time Processing**Topics:**

- Introduction to Cloud Computing for Big Data
- Cloud Platforms for Big Data: AWS, Google Cloud, Azure
- Big Data Services: Amazon EMR, Google BigQuery, Azure HDInsight
- Real-time Data Processing Frameworks: Apache Storm, Kafka, Flink
- Data Pipeline Tools: Apache NiFi, Talend, Airflow

Learning Outcome:

Understand how Big Data is managed and analyzed on cloud environments.

UNIT 8: Big Data Security, Privacy, and Ethics

Topics:

- Security Challenges in Big Data
- Authentication and Access Control in Hadoop
- Data Encryption and Privacy Preservation
- Data Governance and Compliance (GDPR, HIPAA)
- Ethical Use of Big Data

Learning Outcome:

Learn about security, privacy, and governance in Big Data systems.

UNIT 9: Case Studies and Applications

Case Studies:

- Big Data in Healthcare
- Big Data in E-commerce and Retail
- Big Data in Banking and Finance
- Big Data in Social Media Analytics
- Big Data in IoT

Mini Project Ideas:

- Analyzing social media data using Spark
- Customer segmentation using Big Data tools
- Real-time traffic data analysis using Kafka + Spark Streaming

□ Suggested Tools & Technologies

Category	Tools/Frameworks
Data Storage	HDFS, HBase, Cassandra, MongoDB
Processing	Hadoop, Spark, Flink, Storm
Analytics	Python, R, Hive, Pig
Visualization	Tableau, Power BI, Matplotlib
Cloud Platforms	AWS, Google Cloud, Azure
Workflow	Apache NiFi, Airflow

Learning Outcomes

After completing this syllabus, you will be able to:

- ✓ Understand Big Data fundamentals and ecosystem
- ✓ Use Hadoop and Spark for large-scale data processing
- ✓ Store and manage data using NoSQL databases
- ✓ Analyze and visualize Big Data using analytics tools
- ✓ Implement Big Data solutions on cloud platforms
- ✓ Apply ML techniques on Big Data