BIG DATA ANALYTICS

Credits: 4 **Semester: VI**

Course code: DS21603 **No. of lecture hours: 60**

Objectives:

• To understand and learn about Big Data.
• To learn the analytics of Big Data.
• To understand MapReduce fundamentals**.**

Outcome: Students will be able to

CO1: Explain the motivation for big data systems and identify the main sources of Big Data in the real world.

CO2: Develop technical skills in predicative and prescriptive modelling to support business decision-making.

CO3: Implement several Data Intensive tasks using the Map Reduce Paradigm.

CO4: Understand Hadoop ecosystem such as YARN and HIVE-QL for structured databases.

CO5: Demonstrate an ability map-reduce programming using PIG and NoSQL databases for storing purpose and process for Big Data Analytics

UNIT-I

Getting an Overview of Big Data: **12 Hrs**

1. What is Big Data? History of Data Management-Evolution of Big Data
1
2. Structuring Big Data, Elements of Big Data
1
3. Big Data Analytics, Careers and Future of Big Data
1

Exploring the Use of Big Data in Business Context:

4. Use of Big Data in Social Networking
1
5. Preventing Fraudulent Activities
1

Introducing Technologies for Handling Big Data:

6. Distributed and Parallel Computing for Big Data
2
7. Introducing Hadoop, Cloud Computing and Big Data
2

Understanding Big Data Technology Foundations:

8. Exploring the Big Data Stack, Virtualization and Big Data
2
9. Virtualization Approaches
1

UNIT-II **12 Hrs**

Understanding Hadoop EcoSystem

1.Hadoop Ecosystem	2
2.Hadoop Distributed File System	4
3.HBase -Architecture, Regions	2
Understanding MapReduce Fundamentals and HBase	
4. The Map Reduce Framework, Uses of MapReduce	2
5. Role of HBase in Big Data Processing	2

UNIT – III	**12hrs**
Processing your Data with MapReduce	
1. Developing Simple MapReduce Application	2
Customizing MapReduce Execution and ImplementingMapReduce program	
2. Controlling MapReduce Execution with InputFormat	1
3. Reading Data with Custom RecordReader	2
4. Organizing Output Data with OutputFormat	2
5. Customizing Data with RecordWriter	2
6. Optimizing MapReduce Execution with Combiner	2
7. Controlling Reducer Execution with Partitioner	1

UNIT – IV	**12hrs**
Understanding Hadoop YARN Architecture:	
1. Background and Advantages of YARN	2
2. YARN Architecture, Working of YARN, YARN Schedulers	2
3. YARN Configurations, YARN Commands, YARN Containers	2
Exploring HIVE:	
4. Introducing Hive, Getting started with Hive	2
5. Data Types and Built-in functions in Hive, Hive DDL	2
6. Data manipulation in Hive, Data Retrieval Queries, Using Joins in Hive	2

UNIT – V	**12hrs**
Analysing Data with Pig	
1. Introducing Pig, Running Pig	1
2. Getting Started with Pig Latin	1
3. Working with Operators in Pig	2
4. Working with Functions in Pig	2
5. Debugging Pig, Error Handling in Pig	1
NoSQL Data Management:	
6. Introduction to NoSQL	1
7. Types of NoSQL Data Models	2
8. Schema-Less Databases, Materialized Views	1
9. Distribution Models, Sharding	1