Elective IV: 1. Big Data Analytics
Subject Code - MCA305

Max. Marks: 100 **Min Marks : 40**

Course Outcomes

After the completion of course, Student must be able

- To understand the building blocks of Big Data.
- To articulate the programming aspects of cloud computing (map Reduce etc). Also get the knowledge about the big data programming languages apache, pig, hive and spark.
- To understand the specialized aspects of big data with the help of different big data applications.
- To represent the analytical aspects of Big Data along with the knowledge of bigdata database such as mongodb and nosql.
- To know the recent research trends related to Hadoop File System, MapReduce and Google File System etc

Syllabus

UNIT-I Introduction Concept of Big Data

Big Data- Define Data, Web Data, Classification of Data- Structured, Semi-Structured, and Unstructured. Big Data Definitions, Challenges of Conventional system, Why We Need Big Data, Difference between Big Data and Small Data, Importance of Big Data.Big Data Characteristics(4V's Volume, Velocity, Variety,and Veracity), Big Data Types, Big Data Handling Techniques.Complexity of Big Data, Big Data Processing Architectures, Big Data Technologies, Big Data Business Value. Big Data AnalyticsApplication. Big Data Challenges and Future Scope.

UNIT-II INTRODUCTION TO HADOOP AND HADOOP ARCHITECTURE

Big Data - Apache Hadoop&HadoopEcoSystem:Hadoop Core Component, Features of Hadoop, The Hadoop Distributed File System: HDFS data Storage, Hadoop Physical Organization, HDFS Commands, MapReduce Framework, MapReduce Programming Model,MapReduce Map task,Reduce Task and MapReduceExecution,Hadoop YARN, Hadoop2 Execution Model, Hadoop Ecosystem Tools, Hadoop Ecosystem.

UNIT-III NoSQL Big Data Management, Mongo DB

NoSQL: What is it?, Where It is Used Types of NoSQL databases, Why NoSQL?, Advantages of NoSQL, Use of NoSQL in Industry, SQL vsNoSQL, NoSQLDataStore, NoSQL Data Architecture pattern, NOSQL to Manage Big Data.**Data Base for the Modern Web:** Introduction to MongoDB, features of MongoDB, Data Types,Mongo DB Query Language and Database Command.

UNIT-IV Hive and Pig:

Pig: Apache Pig, Application of Apache Pig, Feature, Pig Architecture, Pig- Grunt Shell, Installing Pig, Pig Latin Data Model, Pig Latin and Developing Pig Latin Scripts: Apache Pig Execution, Commands.
HIVE AND HIVEQL.**Hive:** Introduction, Characteristics, limitation, Hive Architecture and Installation, Comparison with Traditional Database (RDBMS), Hive Datatype and File Formats, Hive Data Model,Hive Integration and Workflow Steps, Hive Built-in Functions, HiveQL.

UNIT 5:Hadoop Environment, Analytics and Spark

Installing HBase, Fundamentals of HBase, Running MapReduce jobs on HBase (table input/output), Zookeeper.**Visualizations**: Visual Data Analysis techniques, and interaction techniques.**SPARK:** Introduction to Data Analysis with Spark, Downloading Spark and Getting Started, Programming with RDDs, Machine Learning with MLlib.

RECOMMENDED BOOKS:

1. **Big Data Analytics**, *Raj Kamal and PreetiSaxena*, McGrawHill Education
2. **Big Data: Black Book**, *DT Educational Services*, Dreamtech Press
3. **Big Data Analytics**,*SeemaAcharya&ShubhashiniChellappan*, Wiley India
4. **Big Data Analytics**,*M. Vijayalakshmi&RadhaShankarmani*, Wiley India