

BIG DATA ANALYTICS

(Professional Elective-IV)

Course Code: 20CT1152

L T P C

3 0 0 3

Course Outcomes:

At the end of the Course, the Student will be able to:

CO 1: Relate Big data with Cloud Computing, Hadoop and IoT. (L2)

CO 2: Explain Hadoop Architecture and utilize Hadoop Ecosystem components.(L2)

CO 3: Understand several mechanisms for Big Data generation, acquisition and storage.(L2)

CO 4: Apply various Big Data Analytic Methods.(L3)

CO 5: Apply Big Data Analysis in various Fields.(L3)

UNIT-I

(8 Lectures)

INTRODUCTION:

Dawn of the Big Data Era, Definition and Features of Big Data, Big Data Value, The Development of Big Data, Challenges of Big Data.

RELATED TECHNOLOGIES: Cloud Computing - Cloud Computing Preliminaries, Relationship Between Cloud Computing and Big Data, IoT - IoT Preliminaries, Relationship Between IoT and Big Data, Data Center, Hadoop - Hadoop Preliminaries, Relationship between Hadoop and Big Data. [TEXT BOOK 1]

Learning Outcomes: At the end of the module, the students will be able to:

1. Define the features of Big Data (L1)
2. List the Challenges of Big Data (L1)
3. Explain the preliminaries of various technologies (L2)

UNIT-II

(12 Lectures)

INTRODUCTION TO HADOOP:

Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read. NameNode, Secondary NameNode, and DataNode.

HADOOP ECOSYSTEM:

Hadoop ecosystem components - Schedulers - Fair and Capacity.

Spark:

SPARK BASICS, Using the Spark Shell, RDD Data Types and RDD Creation, RDDs (Resilient Distributed Datasets) in Spark, General RDD Operations: Transformations & Actions, RDD Lineage, RDD Persistence Overview, Distributed Persistence

HIVE :

Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases [TEXT BOOK 2]

Learning Outcomes: At the end of the module, the students will be able to:

1. Understand the Hadoop Architecture (L2)
2. Explain about Hadoop Ecosystem components (L2)
3. Compare the database structures of Spark and Hive (L2)

UNIT-III**(10 Lectures)****BIG DATA GENERATION AND ACQUISITION :**

Big Data Generation-Enterprise Data, IoT Data, Internet Data, Bio medical Data, Data Generation from Other Fields, Big Data Acquisition Data Collection, Data Transportation, Data Pre-processing.

BIG DATA STORAGE:

Storage System for Massive Data, Distributed Storage System, Storage Mechanism for Big Data - Database Technology, Design Factors, Database Programming Model [TEXT BOOK 1]

Learning Outcomes: At the end of the module, the students will be able to:

1. Classify various sources of Big Data generation. (L2)
2. Explain the steps related to Big Data acquisition. (L2)
3. Outline the database Design Factors (L2)

UNIT-IV**(10 Lectures)****BIG DATA ANALYSIS:**

Traditional Data Analysis, Big Data Analytic Methods, Architecture for Big Data Analysis - Real-Time vs. Offline Analysis, Analysis at Different Levels, Analysis with Different Complexity, Tools for Big Data Mining and Analysis. [TEXT BOOK 1]

Learning Outcomes: At the end of the module, the students will be able to:

1. Compare traditional data and Big Data Analysis methods. (L2)
2. Contrast Real-Time vs. Offline Analysis (L2)
3. Apply Big Data with different levels and complexities. (L3).

UNIT-V**(10 Lectures)****BIG DATA APPLICATIONS:**

Application Evolution, Big Data Analysis Fields - Structured Data Analysis, Text Data Analysis, Web Data Analysis, Multimedia Data Analysis, Network Data Analysis, Mobile Traffic Analysis, Key Applications - Application of Big Data in Enterprises, Application of IoT Based Big Data, Application of Online Social Network Oriented Big Data, Applications of Healthcare and Medical Big Data, Collective Intelligence, Smart Grid. [TEXT BOOK 1]

Learning Outcomes: At the end of the module, the students will be able to:

1. Show the Evolution of Big Data Application. (L1)
2. Explain the applications of Big Data in various fields. (L2)
3. Identify various Big Data Analysis Fields. (L3).

TEXT BOOKS:

1. Min Chen, Shiwen Mao, Yin Zhang, Victor C.M. Leung, “*Big Data: Related Technologies, Challenges and Future Prospects*”, Springer; 2014.
2. Tom White, “*Hadoop- The Definitive Guide*”, O’reilly, 4th Edition, April 2015.

REFERENCES:

1. Eric Sammer, "Hadoop Operations", O'Reilley, 2nd Edition., October 2012.
2. VigneshPrajapati, "Big Data Analytics with R and Hadoop", PACKT Publishing, November 2013.

WEB REFERENCES:

1. http://www.jbonneau.com/doc/2012-04-27-big_data_lecture_1.pdf
2. [https://www.ibm.com/developerworks/community/blogs/Susan Visser](https://www.ibm.com/developerworks/community/blogs/Susan%20Visser/entry/flash_book_understanding_big_data_analytics_for_enterprise_class_hadoop_and_streaming_data?lang=en)
Editionntry/flash book understanding big data analytics for enterprise class hadoop and streaming data? langen