

5.5.2 BIG DATA

L	P
2	2

RATIONALE

The importance of Big Data in various domain disciplines has increased tremendously in recent years. This subject provides an overview of the historical and modern context and operation of Big Data for beginners. It will help the student to study and practice Big Data tools and techniques.

COURSE OUTCOMES

At the end of the open elective, the students will be able to:

CO1: Learn the concept of big data.

CO2: Describe the working of Hadoop.

CO3: Handle Map Reduce.

CO4: Study Hadoop eco systems.

DETAILED CONTENTS

UNIT I

BIG DATA OVERVIEW

Introduction – distributed file system, Big data: definition and taxonomy, Sources of Big Data, characteristics, Benefits of Big Data, Understanding Big Data with Examples. Big data applications, Top 10 industries using Big Data, Big data analytics, Challenges for processing big data.

UNIT II

HADOOP

History of Hadoop, What is Hadoop & Hadoop vendors, Big Data – Apache Hadoop & Hadoop EcoSystem. Hadoop Architecture, How Hadoop clusters work, Hadoop Storage: HDFS Introduction, 5 Dameons of Hadoop and their functionalities: NameNode, Secondary NameNode, DataNode, Job Tracker, and Task Tracker.

MAP REDUCE

MapReduce Introduction, How MapReduce Works, Understanding the Map Reduce architecture
- Writing Hadoop MapReduce Word-Count problem - Loading data into HDFS - Executing the Map phase - Shuffling and sorting - Reducing phase execution.

UNIT IV

HADOOP ECO SYSTEMS

Pig: What is Pig. Introduction to Pig Data Flow Engine. Pig and MapReduce. When Pig should be used, Hive: What is Hive, Architecture of Hive, how Hive Differs from Traditional RDBMS.

PRACTICAL EXERCISES

1. Installation of Hadoop.
2. Setting up a Hadoop cluster.
3. Practice of various Hadoop commands.
4. Map Reduce Word Count problem.
5. Map Reduce Tera Sort problem

RECOMMENDED BOOKS

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”, Wiley, ISBN: 9788126551071, 2015.
2. Chris Eaton, Dirk deroos et al. , “Understanding Big data ”, McGraw Hill, 2012.
3. Tom White, “HADOOP: The definitive Guide”, O Reilly 2012.
4. Big Data and Analytics by Seema Acharya and Subhashini Chellappan; Wiley India.
5. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, “Mining of Massive Datasets”, Cambridge University Press, 2014.
6. The Big Data Revolution : Kindle Edition, by Jason Kolb (Author), Jeremy Kolb.
7. Big Data:Principles and best practices of scalable realtime data systems (Englisch),von Nathan Marz James Warren.
8. Data Mining Methods and Models: wileyindia ,by Daniel T Larose.

9. Pro Apache Hadoop, 2ed by Sameer Wadkar, Madhu Siddalingaiah, Jason Venner; Wiley india,
10. e-books/e-tools/relevant software to be used as recommended by AICTE/HSBTE/NITTTR.

RECOMMENDED WEBSITES

1. <http://nptel.ac.in>
2. <https://ekumbh.aicte-india.org>
3. <http://swayam.gov.in>

INSTRUCTIONAL STRATEGY

This is hands on practice based subject and topics taught in the class should be practiced in the Lab regularly for development of required skills among the students. This subject contains five units of equal weightage.