22EINTVAC03	BIG DATA TECHNOLOGIES	L	T	P	C
		3	0	0	0

COURSE OBJECTIVES

- To understand several key big data technologies used for storage, analysis and manipulation of data.
- To recognize the key concepts of Hadoop framework, MapReduce, and Hadoop Ecosystem.
- To analyze various big data based cases studies and prepare a sample big data project.

UNIT–I

Fundamentals of Big Data – Big Data Types – Basics of Distributed Computing – Big Data Technology Components – MapReduce Fundamentals - Defining Big Data Analytics – A Brief History of Hadoop - The Hadoop Foundation and Ecosystem.

UNIT–II

The Hadoop Distributed File system: The Design of HDFS - HDFS Concepts - The Command-Line Interface - Hadoop Filesystems - The Java Interface - Data Flow. Setting Up a Hadoop Cluster: Cluster Specification - Cluster Setup and Installation - SSH Configuration - Hadoop Configuration - YARN Configuration

UNIT–III

MapReduce: Analyzing the Data with Hadoop - Scaling Out - Hadoop Streaming - Hadoop Pipes. Developing a MapReduce Application: The Configuration API - Configuring the Development Environment - Writing a Unit Test - Running Locally on Test Data - Running on a Cluster - Tuning a Job - MapReduce Workflow - Anatomy of a MapReduce Job Run - Failures - Job Scheduling - Shuffle and Sort -Task Execution

UNIT–IV

Pig: Installing and Running - An Example - Pig Latin - User-Defined Functions - Pig in Practice

Hive - Installing Hive - An Example - Running Hive - Comparison with Traditional Databases - HiveQL - Tables - Querying Data - User-Defined Functions

Hbase: Hbasics - Concepts - Installation - Clients – Example

UNIT–V

ZooKeeper: Installing and Running ZooKeeper - An Example - The ZooKeeper Service - Building Applications with ZooKeeper

Sqoop: Getting Sqoop - A Sample Import - Generated Code - Database Imports - Working with Imported Data - Importing Large Objects - Performing an Export.

Case Studies: Hadoop Usage at Last.fm - Hadoop and Hive at Facebook - Nutch Search Engine - Log Processing at Rackspace

TEXT BOOKS

1. Tom White, “Hadoop: The Definitive Guide”, 3rd Edition, O’reilly, 2012.
2. Judith Hurwitz, Alan Nugent, Fern Halper, and Marcia Kaufman, *Big Data for Dummies, First Edition*,For Dummies, 2013.

REFERENCES

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", Wiley, 2015.
2. Borko Furht and Flavio Villanustre, "*Big Data Technologies and Applications", First Edition,* Springer Publishing Company, Incorporated, 2016

COURSE OUTCOMES

At the end of this course, the students will be able to

1. Categorize and summarize big data and its importance.
2. Differentiate various big data technologies like Hadoop, MapReduce, and Hadoop Ecosystem
3. Apply tools and techniques to analyze big data.
4. Earn tips and tricks for big data use cases and solutions.
5. Run and build applications using zookeeper.

Mapping of Course Outcomes (COs) with Programme Outcomes (POs) and Program Specific Outcomes (PSOs)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	-	-	-	-	-	-	-	-	-	-	-	1	-	-
CO2	-	2	2	3	3	-	-	-	-	-	-	-	1	-	-
CO3	3	2	2	-	-	1	-	2	-	-	-	-	1	-	-
CO4	-	-	-	-	-	1	-	2	-	-	-	-	1	-	-
CO5	3	-	-	-	-	1	-	2	-	1	-	-	1	-	-