

## **Big Data Analytics**

### **Course Objectives**

- To introduce big data tools & Information Standard formats.
- To understand the basic concepts of big data.
- To learn Hadoop, HDFS and MapReduce concepts.
- To teach the importance of NoSQL.
- To explore the big data tools such as Hive, HBase and Pig.

### **UNIT I**

**Big Data and Analytics:** Classification of Digital Data: Structured Data- Semi Structured Data and Unstructured Data.

Introduction to Big Data: Characteristics – Evolution – Definition - Challenges with Big Data - Other Characteristics of Data - Big Data - Traditional Business Intelligence versus Big Data - Data Warehouse and Hadoop.

Environment Big Data Analytics: Classification of Analytics – Challenges - Big Data Analytics important - Data Science - Data Scientist - Terminologies used in Big Data Environments – Basically Available Soft State Eventual Consistency - Top Analytics Tools

### **UNIT II**

**Technology Landscape:** NoSQL, Comparison of SQL and NoSQL, Hadoop - RDBMS Versus Hadoop - Distributed Computing Challenges – Hadoop Overview - Hadoop Distributed File System - Processing Data with Hadoop - Managing Resources and Applications with Hadoop YARN - Interacting with Hadoop Ecosystem

### **UNIT III**

**Mongodb and Mapreduce Programming:** MongoDB: Mongo DB - Terms used in RDBMS and Mongo DB - Data Types - MongoDB Query Language.

MapReduce: Mapper – Reducer – Combiner – Partitioner – Searching – Sorting – Compression

### **UNIT IV**

**Hive:** Introduction – Architecture - Data Types - File Formats - Hive Query Language Statements – Partitions – Bucketing – Views - Sub- Query – Joins – Aggregations - Group by and Having – RCFile - Implementation - Hive User Defined Function - Serialization and Deserialization.

## UNIT V

**Pig:** Introduction - Anatomy – Features – Philosophy - Use Case for Pig - Pig Latin Overview - Pig Primitive Data Types - Running Pig - Execution Modes of Pig - HDFS Commands - Relational Operators - Eval Function - Complex Data Types - Piggy Bank - User-Defined Functions - Parameter Substitution – Diagnostic Operator - Word Count Example using Pig - Pig at Yahoo! - Pig Versus Hive

### Text Book:

1. Seema Acharya, Subhashini Chellappan, "Big Data and Analytics", Wiley Publications, First Edition, 2015

### Reference Book:

1. Judith Huruwitz, Alan Nugent, Fern Halper, Marcia Kaufman, "Big data for dummies", John Wiley & Sons, Inc. (2013)
2. Tom White, "Hadoop The Definitive Guide", O'Reilly Publications, Fourth Edition, 2015
3. Dirk Deroos, Paul C. Zikopoulos, Roman B. Melnky, Bruce Brown, Rafael Coss, "Hadoop For Dummies", Wiley Publications, 2014
4. Robert D. Schneider, "Hadoop For Dummies", John Wiley & Sons, Inc. (2012)
5. Paul Zikopoulos, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill, 2012 Chuck Lam, "Hadoop In Action", Dreamtech Publications, 2010

### Course Outcomes

On the successful completion of the course, students will be able to

CO1:	To understand, illustrate and evaluate the concepts and techniques of Data Science, Big Data Analytics and its tools	K1-K6
CO2:	To collaborate, apply and review the computing for big data in Hadoop, and NoSQL environment.	K1-K6
CO3:	To comprehend, implement and review the concepts of data science and big data analytics projects using MapReduce, and MongoDB	K1-K6
CO4:	To understand, use and analyze the concepts of big data analytics projects using HIVE database.	K1-K6
CO5:	To illustrate, develop and review the concepts of PIG database in Hadoop environment.	K1-K6

**K1- Remember, K2- Understand, K3- Apply, K4- Analyze, K5 Evaluate, K6- Create**

### Mapping with Programme Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	-	-	-	-	L	-	-	-	-	-	-
CO2	S	-	M	-	M	L	-	-	-	-	-	-
CO3	S	-	S	-	S	L	-	-	-	S	S	S
CO4	S	-	S	-	S	L	-	-	-	S	S	S
CO5	S	-	S	-	S	L	-	-	-	S	S	S

**S- Strong; M-Medium; L-Low**

**Big Data Analytics Lab**

**Course Objectives**

- To teach the fundamental techniques for handling the big data tools.
- To familiarize the tools required to manage big data.
- To analyse big data using Hadoop, MapReduce, Hive, and Pig
- To teach the fundamental principles in achieving big data analytics with scalability and streaming capability
- To enable students to have skills that will help them to solve complex.

**List of Programs**

1. Implement File System Shell Commands for HDFS in Hadoop Environment
2. Write a Mapreduce program using single reduce function for finding Maximum and Minimum Number
3. Write a Mapreduce program using multiple reduce function for Word Count in an given Text document
4. Implement the following using Pig Latin Input and Output Operations Relational Operations
5. Implement the following using Pig Latin User Defined Functions Advanced Relational Operations
6. Write a Word Count program using Pig Latin Script
7. Write a program to find a maximum temperature using Pig Latin Script
8. Implement the following using Hive commands Handling the Database Creating and Manipulating table
9. Implement Simple Queries for database using Mongo
10. Implement Simple Queries for collections using Mongo

## Course Outcomes

On the successful completion of the course, students will be able to

CO1:	Understand and develop conceptually how Big Data is stored and implement it using different tools	K1-K6
CO2:	Comprehend and implement programs for data storage in HDFS and table manipulation using Big Data tools in Hadoop environment	K1-K6
CO3:	Understand and Critically analyse existing Big Data datasets and implementations the solutions for it using MongoDB	K1- K6
CO4:	Understand and examine existing Big Data datasets and implementations the solutions using HIVE database	K1- K6
CO5:	Comprehend and review existing datasets and implementations the solutions to handle it using PIG	K1- K6

**K1- Remember, K2- Understand, K3- Apply , K4- Analyze, K5- Evaluate, K6- Create**

## Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	S	M		M	S	-	-	-	-	-	-
CO2	S	M	S	S	S	M	-	-	-	-	-	-
CO3	S	S	S	S	S	S	-	-	-	-	-	-
CO4	S	M	S	S	S	M	-	-	-	-	-	-
CO5	S	S	S	S	S	S	-	-	-	-	-	-

**S- Strong; M-Medium; L-Low**