MCA- E306 Big Data Analytics				
	L	T	P	C
	4	0	0	4

Course objective:

1. Understand the Big Data Platform and its Use cases
2. Provide an overview of Apache Hadoop
3. Provide HDFS Concepts and Interfacing with HDFS
4. Understand Map Reduce Jobs
5. Provide hands on Hadoop Eco System
6. Apply analytics on Structured, Unstructured Data.

Course outcomes:

1. Identify Big Data and its Business Implications.
2. List the components of Hadoop and Hadoop Eco-System
3. Access and Process Data on Distributed File System
4. Manage Job Execution in Hadoop Environment
5. Develop Big Data Solutions using Hadoop Eco System

Introduction: Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analyzing Data with Unix tools, Analyzing Data with Hadoop, Hadoop Streaming, Hadoop Echo System

HDFS (Hadoop Distributed File System): The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.

Map Reduce: Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.

Pig: Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators.

Hive: Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions.

HBase: HBasics, Concepts, Clients, Example, HBase Versus RDBMS.

Spark: RDD, Shared variables, Anatomy of a Spark Job Run, Executers and Cluster managers

Recommended Books:

1. Tom White, Hadoop: A Definitive Guide, O'Reilly Media
2. Seema Acharya, Big Data and Analytics, Wiley

HEAD
Department of Computer Science
Gurukul Kangri Vishwavidyalaya
Haridwar (UK) - 249404