Big Data Analytics - I

Credit Distribution, Eligibility and Pre-Requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical / Practice		
Big Data Analytics-I	**2**	**1**	**0**	**1**	**Class XII**	**Basic knowledge programming language (preferably Java), Shell scripting, SQL Query**

Learning Objectives

The Learning Objectives of this course are as follows:

- To understand the Big Data platform and its uses.
- Provide an overview of Apache Hadoop.
- Provide HDFS concepts and Interfacing with HDFS.
- Provide an overview of Map Reduce Programming.

Learning Outcomes

The Learning Outcomes of this course are as follows:

- After studying this course, students will be able to identify Big Data and its Business Implications.
- After studying this course, students will be able to list the components of Hadoop and Hadoop Eco-Systems.
- After studying this course, students will be able to access and process data on distributed file system.
- After studying this course, students will be able to manage job execution in Hadoop environment.

Syllabus:

Unit-1: Understanding Big Data **(2 hours)**

Data Storage and Analysis-The process of data analysis, Characteristics of Big Data, Big Data Analytics, Typical Analytical Architecture, Requirement for new analytical architecture, Challenges in Big Data Analytics – Need of big data frameworks

Unit-2: Foundations of Big Data Systems (4 hours)

Getting started with Hadoop, Requirement of Hadoop Framework, Design principle of Hadoop –Comparison with other system, Understanding Hadoop Ecosystem: Hadoop Components – Hadoop 1 vs Hadoop 2

Unit-3: HDFS (Hadoop Distributed File System) (4 hours)

The Design of HDFS, Hadoop Daemon's – HDFS Commands, HDFS Concepts, Command Line Interface, Hadoop file system interfaces-Loading data into HDFS, read/write process to HDFS

Unit-4: Introduction to Parallel Programming with Map Reduce (5 hours)

Map Reduce Programming: I/O formats, Map side join, Reduce Side Join, Secondary sorting, Pipelining Map Reduce jobs (Map Reduce Execution Pipeline)- Map, Shuffle and Sort, Reduce

Practical Exercises (30 hours)

- Downloading and installing Hadoop.
- Understanding different Hadoop modes. Startup scripts, Configuration files.
- Hadoop Implementation of file management tasks, such as Adding files and directories, retrieving files and Deleting files.
- Run a basic word count Map reduce program to understand map reduce paradigm: To count words in a given file, to view the output file, and to calculate the execution time.
- Map Reduce Program to analyse time-temperature statistics and generate report with max/min temperature.

Essential/recommended readings

- Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.
- Tom White, "Hadoop: The Defective Guide", O'Reilly, 4th Edition, 2015.
- Donald Miner, Adam Shook, "Map Reduce Design Pattern", O'Reilly, 2012.

Examination scheme and mode:

Evaluation scheme and mode will be as per the guidelines notified by the University of Delhi