Savitribai Phule Pune University

(Formerly University of Pune)

Two year M.Sc. Degree Program in Computer Science

(Faculty of Science & Technology)

M.Sc.- II (Computer Science)

Choice Based Credit System Syllabus

To be implemented from Academic Year

2020-2021

CSDT234A Big Data Analytics Total Credits – 2

Pre-requisites:

- Basic knowledge of Linux working and its commands.
- One must be able to install and uninstall its packages.
- Programming Languages - Programming Languages like Python, Scala, Java is required because it helps to understand Hadoop programming.

Course Objectives:

1. To understand the Big Data challenges & opportunities, its applications
2. Understanding of concepts of map and reduce and functional programming
3. Gain conceptual understanding of Hadoop Distributed File System.
4. To solve the case studies related to real life situations
5. To bridge the gap between academics and industry needs.

Course Outcomes:

- Recognize the characteristics, applications of big data that make it useful to real-world problems.
- Process available data using big data tools hadoop file system and predict outcomes to solve given problem.
- Study & Design various case studies using big data tools/commands and analyse it.

Chapter No.	Topics	# Lectures
1.	**Introduction to Big data**	
	1.1 Big Data :Definition & taxonomy	**5**
	1.2 Sources of Big Data	
	1.3 3V's of Big Data (need for Hadoop)	
	1.4 Varying data structures	
	1.5 Characteristics of Big Data	
	1.6 Applications of Big Data	
	1.7 Challenges in Big Data	
	1.8 **Big Data Implications for Industries** Big Data Analytics for Telecom/Banking/Retail/HealthCare/IT/Operations	

2.	**Emerging Database Landscape**	**3**
	2.1 Scale-Out Architecture, RDBMS Vs Non-Relational Database	
	2.2 Database Workload & its Characteristics	
	2.3 Implication of Big Data Scale on Data Processing	
3.	**Application Architecture & Data Modeling For Big Data And Analytics**	**5**
	3.1 Big Data Warehouse & Analytics	
	3.2 Big data Warehouse System requirements & Hybrid Architectures	
	3.3 Enterprise Data Platform Ecosystem	
	3.4 Big Data and Master Data Management	
	3.5 Understanding data integration Pattern	
	3.6 Big Data Workload Design Approaches	
	3.7 Map-Reduce patterns ,Algorithms and Use Cases	
4.	**The Hadoop Ecosystem**	**8**
	4.1 Introduction to Hadoop	
	4.2 Hadoop Architecture	
	4.3 History of Hadoop-Facebook,Dynamo,Yahoo,Google	
	4.4 Hadoop Components :HDFS, Mapreduce	
	4.5 Introduction to Pig,Hive ,HBase ,Mahout	
	4.6 Installation of single node cluster-installation of java Hadoop configuration	
5.	**Extracting Value From Big Data**	**4**
	5.1 Real Time Analytics	
	5.2 In-Memory Data Grid for real Time Analysis	
	5.3 Map reduce & Real Time Processing ,Use Cases	
6.	**Big Data Analytics Methodology**	**5**
	6.1 Big Data Analytics Methodology-Analyze & evaluate business cases	
	6.2 Develop Business Hypothesis –Analyze outcomes, Build & Prepare Data Sets ,Select & Build Analytical Model ,Design for Big Data scale .Build production ready system ,setting up the Big Data Analytics system ,Gathering data ,measure & monitor	

References:

1. Madhu Jagdeesh,Soumendra Mohanty,Harsha Srivatsa,"Big Data Imperatives: Enterprise Big Data Warehouse,BI Implementations and Analytics",1st Edition, Apress(2013)
2. Frank J.Ohlhorst,"Big Data Analytics:Turning Big Data into Big Money",Wiley Publishers(2012)
3. Cristian Molaro,Surekha Parekh,Terry Purcell,"DB2 11:The Database for Big Data & Analytics",MC Press,(2013)
4. Tom White,"Hadoop-The Definitive Guide,Storage and analysis at internet scale",SPD, O'Really.
5. DT Editorial Services,"Big Data, Black Book-Covers Hadoop2, MapReduce,Hive,YARN, Pig, R and Data Visualization" Dreamtech Press,(2015).
6. Big Data Case Study by Bernard Marr –Willey Publications.

CSDP234A Big Data Analytics Practical Total Credits – 2

NOTE:

It is expected to form teams and ask students to solve these case studies, discuss and work on solutions. (Refer Book No 6 for solving case studies. In-detail explanation for case studies below is given in the said book)

1. Assignment 1: Case study on Facebook
2. Assignment 2: Case Study on IoT Sensors
3. Assignment 3: Case Study on Telecom Industry
4. Assignment 4: Case Study on Banking
5. Assignment 5: Case study on Amazon
6. Assignment 6: Case Study on General Electric –By TCS
7. Assignment 7 : Case Study on Uber
8. Assignment 8: Case Study on Netflix
9. Assignment 9: CDC(Corona Virus and other Pandemics)

Practical's

Note: Slips should be designed on the basis of following topics at college level. The practical's should be taken on the basis of above case studies.

1. Navigating in Hadoop environment [Operational commands in Hadoop environment like moving, copying files. creating directories etc.
2. Understand HDFS
3. Using unix tools
4. Development in Hadoop environment , using various Hadoop tools/utilities
5. Develop mapReduce programs [Assignments] - Develop mapReduce functions either in Java or Python