

Governement Polytechnic For Women, Faridabad

Academic Session: July'2019-Dec'2019

Lesson Plan

Course Name: **BIG DATA**

Semester: **V**

Date of commencement of classes: **22nd July 2019**

No of hours allotted per week as per timetable: **3 T + 3 P**

Week No	Module/Topic Details with sub and sub-sub topic	No of Hours Planned	Text books/ Reference
		Theory	Books
JULY 2019			
1 (22-26)	Unit-I INTRODUCTION TO BIG DATA • Introduction – distributed file system, • Big data: definition and taxonomy, • Sources of Big Data, characteristics, Benefits of Big Data, Understanding Big Data with Examples. • Big data applications, Top 10 industries using Big Data, Big data analytics, Challenges for processing big data.	1	[1], [2]
2 (29-31)	• Benefits of Big Data, Understanding Big Data with Examples. • Big data applications, • Top 10 industries using Big Data, • Big data analytics, • Challenges for processing big data.	4	[1], [2]
AUGUST 2019			
3 (1-2)*	Unit II INTRODUCTION TO HADOOP • History of Hadoop, • What is Hadoop & Hadoop vendors • Big Data – Apache Hadoop & Hadoop EcoSystem.	1	[1], [2], [9]
4 (5-15)	• Hadoop Architecture, • How Hadoop clusters work, • Hadoop Storage: HDFS Introduction	2	[1], [2], [9]

5 (18-29)*	• 5 Dameons of Hadoop and their functionalities: • NameNode, • Secondary NameNode, • DataNode, • Job Tracker, • Task Tracker.	3	[1], [2], [4]
SEPTEMBER 2019			
6 (2-6)	UNIT –III INTRODUCTION TO MAP REDUCE • MapReduce Introduction, • How MapReduce Works, • Understanding the Map Reduce architecture	2	[1], [2], [4]
7 (12-20)	• Writing Hadoop MapReduce Word-Count problem • Loading data into HDFS -	3	1], [2], [9]
8(23-30)	• Executing the Map phase – • Shuffling and sorting – • Reducing phase execution.	4	
OCTOBER 2019			
9 (1-10)	UNIT –IV HADOOP ECO SYSTEMS • Pig: What is Pig. • Introduction to Pig Data Flow Engine . • Pig and MapReduce. • When Pig should be used, Hive:	3	[1], [2], [9]
10 (15-24)	• What is Hive, • Architecture of Hive, • how Hive Differs from Traditional RDBMS.	2	[1], [2], [9]
NOVEMBER 2019			
11 (4-8)	UNIT –V INTRODUCTION TO R • Reading and getting data into R – • ordered and unordered factors – • arrays and matrices	2	[1], [2], [4]

12 (11-15)	• lists and data frames – • reading data from files – • probability distributions – • statistical models in R.	2	[1], [2]
------------	---	---	----------

TEXT BOOKS:

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”, Wiley, ISBN: 9788126551071, 2015.
2. Chris Eaton, Dirk deroos et al. , “Understanding Big data ”, McGraw Hill, 2012.
3. Tom White, “HADOOP: The definitive Guide” , O Reilly 2012.
4. Big Data and Analytics by Seema Acharya and Subhashini Chellappan; Wiley India

RECOMMENDED BOOKS:

5. Vignesh Prajapati, “Big Data Analytics with R and Haoop”, Packet Publishing 2013.
6. Nina Zumel, John Mount, “Practical Data Science with R”, Manning Publications, 2014.
7. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, “Mining of Massive Datasets”, Cambridge University Press, 2014.
8. Mark Gardener, “Beginning R - The Statistical Pr ogramming Language”, John Wiley & Sons, Inc., 2012.
9. A Simple Introduction to DATA SCIENCE: BOOK ONE (New Street Data Science Basics 1) Kindle Edition ,by Lars Nielsen Noreen Burlingame .
10. A Simple Introduction to Data Science: BOOK TWO (New Street Data Science Basics 2) Kindle Edition by Lars Nielsen
11. The Big Data Revolution : Kindle Edition, by Jason Kolb (Author), Jeremy Kolb
12. Big Data:Principles and best practices of scalable realtime data systems (Englisch),von Nathan Marz James Warren

Subject Faculty
Ms. Karishma Shukla