RK UNIVERSITY

SYLLABUS

Course Title	Big Data Analytics	
Course Code	CP908 (Elective - II)	
Course Credit	Lecture	: 03
	Practical	: 01
	Tutorial	: 00
	Total	: 04

Course Learning outcomes

At the end of the course students will be able to:

- **Understand** and **describe** the Big Data flow.
- **Explore** the use of SQL-like queries and No SQL in a big data environment.
- **Understand** and **Use** cutting edgetechnologies and tools such as HDFS, MapReduce, Hive, Pig, Python, R and RStudio for Data Analytics.
- **Utilize** and **Apply** the Data Analytics lifecycle to Big Data analytics projects.

Detailed Syllabus

Sr. No.	Name of chapter & details	Hours Allotted
	Section – I	
1	**Introduction to Digital Data:** Structured Data, Semi-Structured Data, Unstructured Data **Introduction to Big Data:** What is Big Data?, Characteristics of Big Data, Challenges, Why is Big Data Important?, Data Science is multi-disciplinary, Big data stack	06
2	No SQL, No SQL vs. RDBMS, Mongo DB- CRUD Operations, Cassandra- CRUD Operations	07

3	**Introduction to Hadoop:** Introduction to Hadoop, Business Value of Hadoop, High Level Architecture of Hadoop, Hadoop Distributed File System, HDFS Architecture, Working with HDFS Commands, Features of HDFS, Processing Data with Hadoop, Managing Resources and Applications with Hadoop YARN, Hadoop Eco System	**08**
Section – II		
4	**Introduction to Map Reduce:** Introduction to MapReduce, Developing MapReduce applications, How Map Reduce works, Types & Formats, MapReduce features, Map Reduce Example	**06**
5	**Hive:** Introduction to Hive, Introduction to Hive Command Line Interface, Hive Data Types and File Formats, Hive Query Language (DDL,DML, Joins, Group by, Sort by, Order by, Views, Index), Hive - UDF (User Defined Function), Hive - Analytical Functions, Data Extraction using Hive. **Pig:** Introduction to Pig, Use Case for Pig, Pig Data Types and File Formats, Interacting with Grunt Shell, Pig Latin, Piggybank Functions, Pig - UDF (User Defined Function), Data Extraction using Pig.	**08**
6	Big Data in Cloud **Introduction to Python:** Introduction to Python, Data Preparation using Python, Data Analysis using Python. **Introduction to R:** Introduction to R, Data Preparation using R, Data Analysis using R.	**07**

Instructional Method and Pedagogy

- Lecture will be conducted with the aid of multi-media projector, blackboard, OHP etc.
- Students are provided with Lecture notes and hand outs for pre reading.
- Edmodo is used as learning management systems for engaging students in off time.
- Students are engaged through various active learning activities.
- Assignments based on Course contents will be given to the students at the end of each unit/topic and will be evaluated at regular interval.

Reference Books

- Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill
- Tom White (Goodreads Author), Doug Cutting (Foreword by), Hadoop: The Definitive Guide, O'Reilly, 1st Edition, 2009
- Viktor Mayer-Schonberger, Kenneth Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and Think, Mariner Books, 2014
- Alan Gates, Programming Pig, O'Reilly, 1st Edition, 2011
- Judith Hurwitz , Alan Nugent, Fern Halper, Marcia Kaufman, Big Data For Dummies, John Wiley & Sons, 2013
- Big Data Now, O'Reilly Media, 2nd Edition, 2012