Kadi Sarva Vishwavidyalaya

Faculty of Engineering and Technology

First Year Master of Engineering (Computer Engineering)

In Effect from Academic Year 2017-18

Subject Code: MECE-205-N-B	**Subject Title: BIG DATA ANALYTICS**

Teaching scheme				**Total Credit**	**Evaluation Scheme**					
L	**T**	**P**	**Total**		**Theory**		**Mid Sem Exam**	**CIA**	**Pract.**	**Total**
Hrs	**Hrs**	**Hrs**	**Hrs**		**Hrs**	**Marks**	**Marks**	**Marks**	**Marks**	**Marks**
04	**00**	**02**	**06**	**05**	**03**	**70**	**30**	**20**	**30**	**150**

LEARNING OBJECTIVES:

This course provides practical foundation level training that enables immediate and effective participation in big data projects. The course provides grounding in basic and advanced methods to big data technology and tools, including MapReduce and Hadoop and its ecosystem.

It also teaches fundamental concepts and tools needed to understand the emerging role of business and big data analytics in Organizations.

OUTLINE OF THE COURSE:

Unit No	**Topics**	**Hours**
1	Introduction To Big Data	09
2	Concepts and Techniques	09
3	Advanced Analytics	09
4	Introduction To Hadoop And Its Architecture	12
5	Hadoop Ecosystem And Yarn	07
6	Hive And Hiveql, Hbase	06
7	NoSQL	05
8	Data Base for the Modern Web	07

Total hours (Theory): 64

Total hours (Practical): 32

Total hours: 96

Kadi Sarva Vishwavidyalaya

Faculty of Engineering and Technology

First Year Master of Engineering (Computer Engineering)

In Effect from Academic Year 2017-18

DETAILED SYLLABUS:

Sr. No	Topic	Lecture Hours	Weight age (%)
1	**Introduction To Big Data** • Overview of Big Data: What Is Big Data and its importance? Driving the growth of Big Data • Differentiating between Big Data and traditional enterprise relational data • Four Vs, Drivers for Big data • Stages of analytical evolution • State of the Practice in Analytics • Big Data Analytics in Industry Verticals • Data Analytics Lifecycle • Big data applications • Challenges of Big Data	09	14
2	**Concepts and Techniques** • In-database Analytics • Data Visualization Techniques • Stream Computing Challenges • Main memory data management techniques • Energy-efficient data processing • Benchmarking • Security and Privacy • Failover and reliability	09	14
3	**Advanced Analytics** • R: Operationalizing Basic Data Analytic Methods Using R • Advanced Analytics - Analytics for Unstructured Data - Map Reduce and Hadoop • Why Is MapReduce Necessary? • How Does MapReduce Work? • Understanding inputs and outputs of MapReduce - Data Serialization. • Real-World MapReduce Examples • Algorithms using map reduce • Matrix-Vector Multiplication by Map Reduce.	09	14

4	**Introduction To Hadoop And Its Architecture** • Big Data – Apache Hadoop & Hadoop EcoSystem – Moving Data in and out of Hadoop • Hadoop Architecture • Hadoop Storage: HDFS • Common Hadoop Shell commands • Anatomy of File Write and Read • NameNode, Secondary NameNode, and DataNode, • Hadoop MapReduce paradigm • Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering – Monitoring & Maintenance.	12	19
5	**Hadoop Ecosystem And Yarn** • Hadoop ecosystem components, Schedulers - Fair and Capacity • Hadoop 2.0 New Features- NameNode High Availability, HDFS Federation, MRv2 • YARN, Running MRv1 in YARN	07	11
6	**Hive And Hiveql, Hbase** • Hive Architecture and Installation • Comparison with Traditional Database • HiveQL - Querying Data - Sorting And Aggregating, Map Reduce Scripts, Joins & Subqueries • HBase concepts- Advanced Usage, Schema Design, Advance Indexing - PIG • Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper.	06	09
7	**NoSQL** • What is it? Where It is Used? Why NoSQL? • Types of NoSQL databases • Advantages of NoSQL, Use of NoSQL in Industry • SQL vs NoSQL, NewSQL	05	08

8	**Data Base for the Modern Web** • Introduction to MongoDB key features, Core Server tools, • MongoDB through the JavaScript's Shell, • Creating and Querying through Indexes, Document-Oriented, principles of schema design • Constructing queries on Databases, collections and Documents • MongoDB Query Language	07	11
	Total	**64**	**100%**

INSTRUCTIONAL METHOD AND PEDAGOGY (Continuous Internal Assessment (CIA) Scheme)

- At the start of course, the course delivery pattern, prerequisite of the subject will be discussed.
- Lectures will be conducted with the aid of multi-media projector, black board, OHP etc.
- Attendance is compulsory in lecture and laboratory which carries 10 marks in overall evaluation.
- One internal exam will be conducted as a part of internal theory evaluation.
- Assignments based on the course content will be given to the students for each unit and will be evaluated at regular interval evaluation.
- Surprise tests/Quizzes/Seminar/tutorial will be conducted having a share of five marks in the overall internal evaluation.
- The course includes a laboratory, where students have an opportunity to build an appreciation for the concepts being taught in lectures.
- Experiments shall be performed in the laboratory related to course contents.

STUDENTS LEARNING OUTCOMES:

On successful completion of the course, the student will:

- Understand big data challenges in different domains including social media, transportation, finance and medicine.
- Distinguish computing frameworks for Big Data analytics.
- Analyze scalability and performance of relational model, SQL and emergent systems.
- Understand the capability of No-SQL systems.
- Be aware of Architectures used in big data
- Build secure big data systems.
- Be able to demonstrate the basic operation in big data.
- Learn difference between conventional SQL query language and NoSQL basic concepts
- Learn tips and tricks for Big Data use cases and solutions.
- Study about Big Data processing language Hadoop

- Be able to build and maintain reliable, scalable, distributed systems with Apache Hadoop.
- Apply Hadoop ecosystem components
- Analyze Map-Reduce programming model for better optimization
- Be able to write Map-Reduce based Applications
- Be able to describe about the modern databases.
- Be able to design and build MongoDB based Big data Applications and learn MongoDB query language
- Analyze case studies in big data.

REFERENCE BOOKS:

1. Bill Franks, Taming *The Big Data Tidal Wave*, 1st Edition, Wiley, 2012.
2. Frank J. Ohlhorst, *Big Data Analytics*,1st Edition, Wiley, 2012.
3. Chris Eaton,Dirk derooset al. , "Understanding Big data ", McGraw Hill, 2012.
4. BIG Data and Analytics , Sima Acharya, Subhashini Chhellappan, Willey
5. Tom White, "HADOOP: The definitive Guide", O Reilly 2012.
6. Robert D. Schneider , Hadoop for Dummies, Wiley India.
7. Boris lublinsky, Kevin t. Smith, AlexeyYakubovich, "Professional Hadoop Solutions", Wiley, ISBN: 9788126551071, 2015.
8. MongoDB in Action, Kyle Banker,Piter Bakkum , Shaun Verch, Dream tech Press
9. VigneshPrajapati, "Big Data Analyticswith R and Haoop", Packet Publishing 2013.
10. R in Nutshell , O Reilly.
11. W.N. Venables, D.M Smith, "An introduction to R",

E-RESOURCES:

1. http://bigdatauniversity.com/
2. http://www.coreservlets.com/hadoop-tutorial/
3. http://cran.r-project.org/doc/manuals/R-intro.html

Kadi Sarva Vishwavidyalaya

Faculty of Engineering and Technology

First Year Master of Engineering (Computer Engineering)

In Effect from Academic Year 2017-18

LIST OF PRACTICALS:

Sr. No	Name of Experiment
1	To setup Hadoop and to run sample program using hadoop.
2	To understand the overall programming architecture using Map Reduce API .
3	Store the basic information about students such as roll no, name, date of birth , and address of student using various collection types such as List, Set and Map .
4	Retrieve various types of documents from students collection .
5	To find documents from Students collection.
6	Develop Map Reduce Work Application .
7	Creating the HDFS tables and loading them in Hive and learn joining of tables in Hive.
8	Perform data analytics using Pig, Hive and YARN.
9	Basic CRUD operations in MongoDB.
10	Case Study: To study research papers on the given topic and prepare a report on it.