



## Sub Committee Engineering – Computer Science/IT for Curriculum Development

**Course Title:** Big Data Analytics

**Course Code:**

**Number of Credits:** 4

**Level :** 4

**Learning Outcome (s):** The students are able to:

Differentiate between Name Node, Secondary Name Node, Data Node.

Compare MapReduce-1 and MapReduce-2 frameworks for solving Big data problems

Differentiate Hive and RDBMS.

Experiment pig queries and examine the performance.

Test Sorting, Aggregate functions in HiveQL.

Analyze Query execution performance with in-memory databases like Apache Spark.

### Pre-learning:

Knowledge of Big data, No-SQL is desirable

### Pedagogy:

Classroom teaching

Demonstration in labs

Case Study

### Course Outline:

Sr.No.	Topic	Hours
1	Hadoop Architecture: Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., Name Node, Secondary Name Node, and Data Node, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering – Monitoring & Maintenance, Hadoop ecosystem components - Schedulers - Fair and Capacity, Hadoop 2.0 New Features- NameNode, High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.	15

2	Hive: What is Hive?, Architecture of Hive, Hive services, Hive Clients, How it differs from RDBMS, Introduction to HiveQL, Data Types and file formats in Hive, File encoding, Common problems while working with Hive. HiveQL: Managed and external tables, Understand storage formats, Sorting And Aggregating, Map Reduce Scripts, Joins & Subqueries, Writing user defined functions, Data types and schemas, Querying data.	15
3	PIG Concepts of Pig , Installation of a Pig engine, Queries, Case Study	15
4	Apache Spark RDD, RDD Transformations & Meaning, Advanced Spark programming	15
	<b>Total</b>	<b>60</b>

### Books Recommended:

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. Vignesh Prajapati, "Big Data Analytics with R and Haoop", Packet Publishing 2013.
5. Tom Plunkett, Brian Macdonald et al, "Oracle Big Data Handbook", Oracle Press, 2014.
6. <http://www.bigdatauniversity.com/>
7. Jy Liebowitz, "Big Data and Business analytics",CRC press, 2013

### Suggested Evaluation Methods:

#### A. Continuous Assessment

- a) Unit test 1
- b) Unit Test 2
- c) Assignments
- d) Presentation
- e) Quizzes

#### B. End Semester Examination

Written Exam

### Benchmarked against similar courses in other national/ international universities /organizations

S. No.	Name of the Course	Name of University where it is offered
1	Big data Analytics	SRM university, Karantaka
2.	Hadoop Platform and Application Framework	University of California

Name of Member	Dr. Shraddha Phansalkar	Dr. Swati Ahirrao
Designation	Associate Professor (CS & IT)	Associate Professor (CS & IT)
Org. / Inst.	SIT	SIT
Signature		

Name of Experts	Dr. Sandeep deshmuKh
Designation	Data Analyst
Org. / Inst.	freelancer
Signature	

Signature of Dean:

Date: