

Branch Name:	Information Technology
Semester/Year:	Semester VII / Final Year
Subject Title:	Analytics in Big Data
Subject Code:	1ET1050701
Pre-requisite:	-

Course Objective:

- To learn fundamentals of big data analytics.
- To learn and understand management of big data in big data applications.
- To study Hadoop and its architecture.
- To learn No SQL over conventional SQL.

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Theory (Marks)		Practical (Marks)		Total (Marks)
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
04	-	02	05	70	30	30	20	150

Subject Contents			
Sr. No	Topic	Total Hours	Weightage (%)
1	Big Data: An Introduction Why Big data? Importance of Big data. Characteristics of Big Data, Four V's of Big data and their meaning. What is Big Data Analytics? Big Data Examples and Sources. Big Data Technologies: An overview.	06	12
2	Big Data Processing: Big Data ecosystem, The Hadoop Framework, Hadoop Building blocks, Data processing in Hadoop architecture. Map Reduce operations.	08	13
3	Hadoop Distributed File Systems (HDFS): HDFS: Introduction, Hadoop Architecture, Hadoop Map Reduce tasks, job, task trackers. Map Reduce example.	10	25
4	Hive and HBase: Architecture of Apache Hive, HiveQL, Database management in HiveQL, Apache HBase Introduction, Its functions, Pig, Zookeeper, Cluster management using apache pig and zookeeper.	08	15
5	Big Data Manipulation: What is NoSQL? Data manipulation in NoSQL database, Difference between SQL and NoSQL. Industry applications of NoSQL. NoSQL advantages. What is Mongo DB? Features of mongo DB, CRUD operations, Using the mongo DB shell, Querying using indexes in mongo DB, Collections and Documents, Schema designing, Replication and Sharding.	10	15

6	SPARK: SPARK: Introduction, installation of Spark, Using SPARK shell, programming of RDDs.	06	20
---	--	----	----

Course Outcome:

After learning the course, Students will be able to do the following:

- Students can identify key areas of Big data and can learn the applications of the same.
- The students will be able to differentiate SQL and NoSQL concepts and their applicability.
- Usage of Hadoop can help students to manage distributed data systems by incorporating Map-Reduce operations.
- By learning MongoDB query language students can have exposure to building and maintaining big data applications.

List of Reference Books:

1. BIG Data and Analytics , Sima Acharya, Subhashini Chhellappan, Willey
2. Understanding Big Data- Chris Eaton,Dirk derooset, McGraw-Hill.
3. Big Data Analytics with R and Hadoop- Vignesh prajapati, Packt Pub Ltd.
4. HADOOP: The definitive Guide, Tom White, O Reilly.
5. Karau, “Learning Spark”, O'Reilly.

List of Suggested titles of Experiments:

1. To study Hadoop installation & configuration
2. To understand architecture of Map Reduce API.
3. To study MongoDB installation & configuration.
4. Perform basic CRUD operations in MongoDB
5. Perform text search operations in MongoDB.
6. To study Installation of Hive. Create the HDFS tables and load them in Hive.
7. To analyze and study various special purpose diodes such as varactor diode and photo diode.
8. To learn joining of tables in Hive.

e-Resources:

1. <https://nptel.ac.in/courses/106104189/>
2. <https://cognitiveclass.ai/courses/what-is-big-data/>
3. <https://datafloq.com/>