

**B. TECH. SEMESTER – VII (CE)**  
**SUBJECT: BIG DATA AND ANALYTICS**  
**SYLLABUS & SCHEME (W.E.F. 2019)**

Teaching Scheme (Hours/Week)			Examination Scheme (Marks)					
Lectures	Tutorial	Practical	Theory (3 hrs.)	Sessional (1.25 hr.)	Practical	T.W.	Total	Credits
4	0	2	60	40	25	25	150	5

**A. OBJECTIVES OF THE COURSE:**

The tombs of data are no more futuristic because of the Volume, Velocity add Variety of demand driven data sources. To handle the Big Data, technological advancement is the key to stay up on the edge. Today's happening is the footprint of tomorrows and hence the Analytics of Big Data can help us generate the results we need to build a better tomorrow. The horizontal scalability infrastructure and up to the mark usage is more than necessity of the time.

**Designed in collaboration  
with  
Infosys Limited  
under part of Infosys Campus Connect Program.**

**B. DETAILED SYLLABUS:**

- 1) Types of Digital Data (Structured, Semi-Structured, Unstructured)
- 2) Introduction to Big Data
- 3) The Big Data Technology Landscape
  - a. NoSQL – NewSQL
  - b. Hadoop - Introduction to Eco system.
- 4) Hadoop – Distributed File System and Processing using MapReduce.
- 5) Introduction to MapReduce Programming
- 6) Introduction to Big Data Analytics
- 7) Data Storage and Handling (Apache Cassandra/mongoDB)
- 8) Querying Data using Hive/Pig like components
- 9) Data Reporting Tools (i.e. Community Edition : Jasper Soft )

## 10) The realm of Data Science

### C. LEARNING OUTCOMES:

The end to end steps, followed from data handling to reporting visionaries or any individual for day to day life, is the goal of efforts laid in the subject. Students shall become familiar of notions of Big Data. The importantly storage options and processing for results. Not only programmatically but also querying is possible and interfacing to different components for Analytics.

### D. RECOMMENDED TEXT BOOK:

1. Big Data and Analytics – Seema Acharya and Subhashini C – Wiley India

### E. REFERENCE BOOKS:

1. Hadoop: The Definitive Guide by Tom White
2. Big Data Analytics: Methods and Applications by B. L. S. Prakasa Rao (Editor), S. B. Rao (Editor)

Online References:

<https://bigdataanalytics.mit.edu/>

<https://www.edx.org/learn/big-data>

### F. LIST OF EXPERIMENTS TO BE PERFORMED UNDER THIS SUBJECT HEAD

Sr. No.	Aim
1	Recording types of data and various file formats. Identifying data sources. Handling traditionally to start with at small scale.
2	Configuring and experiencing Hadoop Stand Alone Mode. Start/stop script and monitoring.
3	Interfacing to Distributed File System – HDFS commands and gui file system browser usage
4	Write a map-reduce program to count the frequencies of words from distributed storage source and understand the phases involved in map-reduce programming.
5	Map-reduce solution to a custom problem at hand. Molding the problem into map-reduce architecture.
6	Connecting to NoSQL database/s and querying to provide analysis using api like aggregation, etc. To be able to successfully import/export from CSV.
7	Crud operations and more using hive query language.
8	Establishing user equivalence and configuring Hadoop Cluster. Experiencing Big Data handling / scaling horizontally. Monitoring the health of cluster.
9	Leveraging machine learning using Mahout like tools.

10	Delivering Reports including visualizations and customized properties.
----	--