Course: BTech **Semester:** 7

Prerequisite: Database Management system, SQL

Course Objective: Big data analytics is the often complex process of examining big data to uncover information such as hidden patterns, correlations, market trends and customer preferences that can help organizations make informed business decisions

Teaching and Examination Scheme

Teaching Scheme					Examination Scheme					Total
					Internal Marks			External Marks		
Lecture Hrs/Week	Tutorial Hrs/Week	Lab Hrs/Week	Hrs/Week	Credit	T	CE	P	T	P	
3	0	0	0	3	20	20	-	60	-	100

SEE - Semester End Examination, **T** - Theory, **P** - Practical

Course Content

W - Weightage (%) , **T** - Teaching hours

Sr.	Topics	W	T
1	**Introduction:** What is in Store? , Classification of Digital Data: Structured, Semi Structured & Un Structured , Evolution of Big Data , Definition of Big Data - Volume - Velocity ±Variety, Challenges of Big Data , Why Big Data?, Traditional Business Intelligence (BI) versus Big Data , industry examples of big data , What is Big Data Analytics? , Data Science	20	9
2	**Nosql Data Management:** Introduction to NoSQL, Types of NoSQL, Why NoSQL? , Advantages of NoSQL, Comparison of SQL, NoSQL and NewSQL , aggregates , key-value and document data models, graph databases, map-reduce, partitioning and combining	20	9
3	**Basics Of Hadoop:** What is Hadoop?, Brief History of Hadoop , Why Hadoop? , RDBMS versus Hadoop , Hadoop Components , High Level Architecture of Hadoop , Key Advantages & Features of Hadoop , Data format ,Hadoop distributed file system (HDFS) , Processing Data with Hadoop. Map Reduce Interface:Overview of Map Reduce, Map-Reduce workflows, anatomy of Map-Reduce job run, shuffle and sort ,task execution ,input formats , output formats.	40	18
4	**Hadoop Related Tools:** Overview of HBase, Pig introduction, Pig data model, Hive, data types and file formats, HiveQL data definition, HiveQL data manipulation, HiveQL queries, Pig Latin Overview , Pig versus Hive, Using JSON , Overview of Cassandra, Jasper Reports.	20	9

Reference Books

1.	**Hadoop: The Definitive Guide by Tom White, Third Edition, O'Reilley. (TextBook)** By Tom White
2.	**Understanding Big data** By Chris Eaton,Dirk derooset al. | McGraw Hill, Pub. Year 2012
3.	**Hadoop Operations** By Eric Sammer | O'Reilley, Pub. Year 2012
4.	**Big data analytics with R and Hadoop, VigneshPrajapati,SPD.** By VigneshPrajapati
5.	**Big Data and Analytics** By Seema Acharya and Subhashini C | Wiley India
6.	**Programming Hive, E. Capriolo, D. Wampler, and J. Rutherglen, O'Reilley** By E. Capriolo, D. Wampler, and J. Rutherglen
7.	**MongoDB in Action** By Kyle Banker, Piter Bakkum ,Shaun Verch | Dream tech Press
8.	**HBase: The Definitive Guide, Lars George, O'Reilley**

Course Outcome
After Learning the Course the students shall be able to:
1. Understand the Big Data flow 2. Solve problems using MapReduce technique 3. Implement single-node/multimode Hadoop cluster 4. Differentiate between conventional SQL query language and NoSQL 5. Apply the various technologies and tools associated with Big Data such as HDFS, Map Reduce, Pig, Hive, MongoDB

Course: BTech **Semester:** 7

Prerequisite: Database Management system, SQL

Course Objective: Big data analytics is the often complex process of examining big data to uncover information --such as hidden patterns, correlations, market trends and customer preferences -- that can help organizations make informed business decisions

Teaching and Examination Scheme

Teaching Scheme					Examination Scheme					Total
Lecture Hrs/Week	Tutorial Hrs/Week	Lab Hrs/Week	Hrs/Week	Credit	Internal Marks			External Marks		
					T	CE	P	T	P	
0	0	2	0	1	-	-	20	-	30	50

SEE - Semester End Examination, **T** - Theory, **P** - Practical

Course Outcome

After Learning the Course the students shall be able to:
1. Understand the Big Data flow 2. Solve problems using the MapReduce technique 3. Implement single-node/multimode Hadoop cluster 4. Differentiate between conventional SQL query language and NoSQL 5. Apply the various technologies and tools associated with Big Data such as HDFS, Map Reduce, Pig, Hive, MongoDB

List of Practical

1.	**To understand the overall programming architecture using Map Reduce API.**
2.	**Write a program of Word Count in Map Reduce over HDFS.**
3.	**Basic CRUD operations in MongoDB.**
4.	**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.**
5.	**Basic commands available for the Hadoop Distributed File System.**
6.	**Basic commands available for HIVE Query Language.**
7.	**Basic commands of HBASE Shell.**
8.	**Creating the HDFS tables and loading them in Hive and learn joining of tables in Hive.**