BIG DATA AND ANALYTICS

Course Code: CS8T02 **Credits: 3-0-0-1-4**

UNIT I: Getting an Overview of Big Data **8 Hrs**

What is Big Data? History of Data Management-Evolution of Big Data, Structuring Big Data-Types of Data, Elements of Data, Advantages of Big Data Analytics Introducing Technologies for Handling Big Data Distributed and Parallel Computing for Big Data, Introducing Hadoop, Cloud Computing and Big Data: Cloud Delivery Models, Cloud Services for Big Data, Cloud Providers in Big Data Market, In-Memory Computing TechnologyforBig Data

UNIT II: Big Data Analytics and Technology Landscape **8 Hrs**

Where do we Begin? What is Big Data Analytics? What Big Data Analytics Isn't? Why this Sudden Hype Around Big Data Analytics? Classification of Analytics, Greatest Challenges that Prevent Businesses from Capitalizing on Big Data, Top Challenges Facing Big
Data, Why is Big Data Analytics Important? What Kind of Technologies are we looking Toward to Help Meet the Challenges Posed by Big Data? Data Science, Data Scientist...Your New Best Friend!!! ,Terminologies Used in Big Data Environments. NoSQL (Not Only SQL) ,Hadoop

UNIT III: Introduction to Hadoopand MongoDB **8 Hrs**

:IntroducingHadoop, Why Hadoop? Why not RDBMS? RDBMS versus Hadoop, Distributed Computing Challenges ,History of Hadoop , Hadoop Overview, Use Case of Hadoop ,Hadoop Distributors ,HDFS (Hadoop Distributed File System),Processing Data with Hadoop, Managing Resources and Applications with Hadoop YARN (Yet another Resource Negotiator),Interacting
withHadoop Ecosystem . Introduction to MongoDB: What is and Why MongoDB? Terms used in RDBMS and MongoDB, Data types in MongoDB,MongoDB Query language.

UNIT IV: Introduction to cassandra and MAPREDUCE **7 Hrs**

Apache Cassandra, features, CQL data types, CQLSH, key spaces, CRUD, collections, TTL, using a counter, ALTER commands, import and export, query system tables. **MAPREDUCE Programming :**Mapper, Reducer, Combiner, Partitioner, Searching, Sorting, Compression .

UNIT V: Introduction to Hive and Pig **8 Hrs**

what is Hive? , Hive Architecture, Hive Data Types, Hive File Format, Hive Query Language (HQL), User-defined Function(UDF).
What is Pig? The Anatomy of Pig, Pig on Hadoop , Pig Philosophy, Use Case for Pig: ETL Processing, Pig Latin Overview , Data Types in Pig ,Running Pig , Execution Modes of Pig ,HDFS Commands ,Relational Operators, Eval Function, Complex Data Types ,Piggy Bank, User- Defined Functions (UDF) ,Parameter Substitution , Diagnostic Operator , Word Count

Example using Pig ,When to use Pig? When not to use Pig? Pig at Yahoo! ,Pig versus Hive .

Text Books:

1. Big Data: Black Book :Dt Editorial Services, Dreamtech Press, Edition 2016 (Chapter 1).
2. Big Data and Analytics, SeemaAcharya, SubhashiniChellappan, Infosys Limited, Publication:Wiley India Private Limited,1st Edition 2015.

Reference Books:

1. Hadoop in Practice, Alex Holmes, Manning Publications Co., September 2014, Second Edition.
2. Programming Pig, Alan Gates, O'Reilly, Kindle Publication.
3. Programming Hive, Dean Wampler, O'Reilly, Kindle Publication

Course Outcomes:

1. Identify the characteristics of datasets and compare the trivial data and big data for various applications.
2. Demonstrate an open source software framework called Hadoop and supported tool to empower any meaningful conversation on Big data and analytics.
3. Compare and Contrast different Hadoop supporting tools with traditional tool
4. How Big Data can be analyzed to extract knowledge and apply tools for bigdata analytics