IT7 C4 Big Data Analytics [3 - 2 4]

Syllabus:

Introduction to Data Analytics Nature of Data, Types of Digital Data, Classification of Digital Data, Structured Data, Semi-Structured Data , Unstructured Data, Characteristics of Data.

Introduction to Big Data: Evolution of Big Data , Definition of Big Data, Challenges of Conventional Systems, Intelligent Data Analysis, Challenges of Big Data Analytic Processes and Tools, Analysis vs Reporting, , Statistical Concepts: Sampling Distributions, Re-Sampling, Statistical Inference - Prediction Error.

Introduction to Big Data Analytics: Big Data Analytics, importance of Big data analytics, Sudden Hype Around Big Data Analytics, Classification of Analytics, Top Challenges Facing Big Data, Kind of Technologies to meet the Challenges Posed by Big Data, Data Science, Role of data scientist, Terminologies Used in Big Data Environment

The Big data technology landscape : NoSQL, Types of No SQL databases, SQL Vs No SQL, why No SQL, Introduction to MongoDB, Data Types in MongoDB, CRUD, Practice examples, Apache Cassandra, Features of Cassandra, CRUD, Practice examples

Hadoop: Introducing Hadoop, comparisons of RDBMS and Hadoop, Distributed Computing Challenges, A Brief History of Hadoop, Hadoop Overview, Business Value of Hadoop, Hadoop Distributors, Hadoop Distributed File System, Processing Data with Hadoop , Introduction to Map reduce, working of Map reduce, Hadoop YARN , Hadoop Ecosystem, HDFS, Hadoop in the Cloud, Case Studies: Real Time Sentiment Analysis and Stock Market Predictions

Frameworks: Applications on Big Data Using Pig and Hive, Data Processing Operators in Pig, Hive Services, HiveQL, Querying Data in Hive, Fundamentals of HBase and ZooKeeper, IBM Info Sphere Big Insights and Streams, Visualizations, Visual Data Analysis Techniques, Interaction Techniques, Systems and Applications, Jasper Report using Jasper Soft

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 8 experiments to be incorporated.

References:

1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer
2. Tom White, Hadoop: The Definitive Guide, Third Edition, O'reilly Media
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Publishing
4. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press
5. Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, John Wiley & sons
6. Glenn J. Myatt, Making Sense of Data, John Wiley & Sons
7. Da Ruan, Guoquing Chen, Etienne E.Kerre, GeertWets, Intelligent Data Mining, Springer
8. Paul Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, Harness the Power of Big Data The IBM Big Data

Platform, Tata McGraw Hill Publications
9. Michael Minelli, Michele Chambers, Ambiga Dhiraj, Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley Publications
10. Zikopoulos, Paul, Chris Eaton, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Tata McGraw Hill Publications
11. Seema Acharya and Subhashini C, Big Data and Analytics, Wiley India