

## MCA 505D

# Business Analytics and Big Data

### Module 1(10 hours)

Overview business analytics , Foundation and Technologies for decision Making, Descriptive Analytics – Data warehousing, Predictive Analytics – Data Mining, Predictive Analytics – Text Analytics and Text Mining, Predictive Analytics – Web Analytics and Web Mining, Model Based Decision Making, Modeling and Analysis, Knowledge Management and Collaborative Systems, Perspective analytics, Business Analytics: Emerging Trends and Future Impacts

### Module 2(10 hours)

Introduction to Big Data, Big Data characteristics, types of Big Data, Traditional vs. Big Data business approach, Case Study of Big Data Solutions. Hadoop: Core Hadoop Components; Hadoop Ecosystem; Physical **Architecture**; **Hadoop limitations**. **NoSQL: NoSQL business drivers**; **NoSQL** case studies; NoSQL data architecture patterns: Key-value stores, Graph stores, Column family (Bigtable) stores, Document stores, Variations of NoSQL architectural patterns; Using NoSQL to manage big data: What is a big data NoSQL solution? Understanding the types of big data problems; Analyzing big data with a shared-nothing architecture; Choosing distribution models: master-slave versus peer-to-peer; Four ways that NoSQL systems handle big data problems

### Module 3(10 hours)

Distributed File Systems : Physical Organization of Compute Nodes, Large-Scale File-System Organization.

MapReduce: The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of MapReduce Execution, Coping With Node Failures. Algorithms Using MapReduce: Matrix-Vector Multiplication by MapReduce , Relational-Algebra Operations, Computing Selections by MapReduce, Computing Projections by MapReduce, Union, Intersection, and Difference by MapReduce, Computing Natural Join by MapReduce, Grouping and Aggregation by MapReduce, Matrix Multiplication, Matrix Multiplication with One MapReduce Step.

### Module 4(10 hours)

HADOOP RELATED TOOLS: Hbase,data model and implementations, Hbase clients,Hbase examples – praxis.Cassandra ,cassandra data model , cassandra examples, cassandra clients , Hadoop integration. Pig , Grunt , pig data model , Pig Latin, developing and testing Pig Latin scripts. Hive , data types and file formats , HiveQL data definition , HiveQL data manipulation – HiveQL queries

### Module 5

(As per choice of faculty) Portion covered can be tested through Internal evaluation only not to be included in University examination)

#### Text Books:

1. Marc J. Schniederjans Dara G. Schniederjans Christopher M. Starkey “Business Analytics Principles, Concepts, and Applications What, Why, and How” , Pearson Education,2014
2. Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013.
3. Alex Holmes “Hadoop in Practice”, Manning Press, Dreamtech Press.
4. Dan McCreary and Ann Kelly “Making Sense of NoSQL” – A guide for managers and the rest of us, Manning Press.

#### Reference Books

Anand Rajaraman and Jeff Ullman “Mining of Massive Datasets”, Cambridge University Press,  
Tom White, "Hadoop: The Definitive Guide", Third Edition, O'Reilley, 2012.  
Eric Sammer, "Hadoop Operations", O'Reilley, 2012.  
E. Capriolo, D. Wampler, and J. Rutherglen, "Programming Hive", O'Reilley, 2012.