GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND TECHNOLOGY

BIG DATA ANALYTICS
(PROFESSIONAL ELECTIVE II)

Course code: GR15A3111   **L/T/P/C: 3/1/0/4**

III Year II Semester

UNIT - I

INTRODUCTION TO BIG DATA AND HADOOP

Classification of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, RDBMS Vs Hadoop, Analysing Data with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Eco Systems, IBM Big Data Strategy, Introduction to Infosphere BigInsights.

UNIT - II

HDFS (Hadoop Distributed File System)

The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Hadoop I/O, Data Integrity, Compression, Serialization, File-Based Data structures.

UNIT - III

Map Reduce

Map Reduce programming Model, Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats-Map Reduce Types, Input Formats, Output Formats, Map Reduce Features.

UNIT - IV

PIG

Pig: Introduction to PIG, Execution Modes of Pig, Grunt, Pig Latin Editors, Comparison of Pig with Databases, Pig Latin, Pig on Hadoop, User Defined Functions, Data Processing operators.

UNIT - V

Data Stores on Hadoop

Hive: Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions
Hbase: HBasics, Concepts, Schema Design, Loading Data, Online queries, Hbase Versus RDBMS, Introduction to Zookeeper

Text Books
1. Tom White "Hadoop: The Definitive Guide" 4th edition, O'reily Media, 2012.
2. Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015

Reference Books
1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
2. Jay Liebowitz, "Big Data and Business Analytics" Auerbach Publications, CRC press (2013)
3. Tom Plunkett, Mark Hornick, "Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop", McGraw-Hill/Osborne Media (2013), Oracle press.

4. Anand Rajaraman and Jef rey David Ulman, "Mining of Massive Datasets", Cambridge University Press, 2012.
5. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley & sons, 2012.
6. Glen J. Myat, "Making Sense of Data", John Wiley & Sons, 2007
7. Pete Warden, "Big Data Glossary", O'Reily, 2011.
8. Michael Mineli, Michele Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley Publications, 2013.
9. ArvindSathi, "BigDataAnalytics: Disruptive Technologies for Changing the Game", MC Press, 2012
10. Paul Zikopoulos , Dirk DeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corigan , "Harness the Power of Big Data The IBM Big Data Platform ", Tata McGraw Hill Publications, 2012.