I Year II Semester   **L   P   C**

Code: 17CS203   **4   0   3**

BIG DATA ANALYTICS

UNIT-I

Data structures in Java: Linked List, Stacks, Queues, Sets, Maps; Generics: Generic classes and Type parameters, Implementing Generic Types, Generic Methods, Wrapper Classes, Concept of Serialization

UNIT-II

Working with Big Data: Google File System, Hadoop Distributed File System (HDFS) – Building block sofe Hadoop (Name node, Data node, Secondary Name node, Job Tracker, Task Tracker), Introducing and Configuring Hadoop cluster (Local, Pseudo-distributed mode, Fully Distributed mode), Configuring XML files.

UNIT-III

Writing Map Reduce Programs: A Weather Dataset, Understanding Hadoop API for Map Reduce Framework (Old and New), Basic programs of Hadoop Map Reduce: Driver code, Mappercode, Reducer code, Record Reader, Combiner, Practitioner

UNIT-IV

Hadoop I/O:The Writable Interface, Writable Comparable and comparators, Writable Classes: Writable wrappers for Java primitives, Text, Bytes Writable, Null Writable, Object Writable and Generic Writable, Writable collections, Implementing a Custom Writable: Implementing a Raw Comparator for speed, Custom comparators

UNIT-V

Pig: Hadoop Programming Made Easier

Admiring the Pig Architecture, Going with the Pig Latin Application Flow, Working through the ABCs of Pig Latin, Evaluating Local and Distributed Modes of Running Pig Scripts, Checking out the Pig Script Interfaces, Scripting with Pig Latin.

Applying Structure to Hadoop data with Hive: Saying Hello to Hive, Seeing How the Hive is Put Together, Getting Started with Apache Hive, Examining the Hive Clients, Working with Hive Data Types, Creating and Managing Databases and Tables, Seeing How the Hive Data Manipulation Language Works, Querying and Analyzing Data

TEXTBOOKS:

1. Big Java 4th Edition, Cay Horstmann, Wiley John Wiley & Sons, INC
2. Hadoop: The Definitive Guide by Tom White, 3rdEdition,O'reilly
3. Hadoop in Action by ChuckLam, MANNING Publ.
4. Hadoop for Dummies by DirkdeRoos, PaulC.Zikopoulos, RomanB.Melnyk, BruceBrown, RafaelCoss

REFERENCEBOOKS:
1. Hadoop in Practice by Alex Holmes,MANNING Publ.
2. Hadoop Map Reduce Cook book, Srinath Perera,Thilina Gunarathne

SOFTWARELINKS:

Hadoop: http://hadoop.apache.org/

Hive: https://cwiki.apache.org/confluence/display/Hive/Home

Piglatin: http://pig.apache.org/docs/r0.7.0/tutorial.html