



BONAM VENKATA CHALAMAYYA INSTITUTE OF TECHNOLOGY & SCIENCE

(AUTONOMOUS)

DEPARTMENT OF CSE - ARTIFICIAL INTELLIGENCE & DATA SCIENCE

III Year II Semester

L	T	P	C
3	0	0	3

BIG DATA ANALYTICS

COURSE OBJECTIVES:

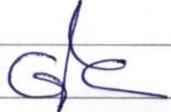
- Optimize business decisions and create competitive advantage with Big Data analytics
- Introducing Java concepts required for developing map reduce programs
- Derive business benefit from unstructured data
- Imparting the architectural concepts of Hadoop and introducing map reduce paradigm
- To introduce programming tools PIG & HIVE in Hadoop ecosystem.

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 blocks of Hadoop (Namenode, Datanode, Secondary Namenode, 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, Mapper code, Reducer code, Record Reader, Combiner, Practitioner

UNIT-IV: Stream Memory and Spark: Introduction to Streams Concepts– Stream Data Model and Architecture , Stream computing, Sampling Data in a Stream , Filtering Streams ,Counting Distinct Elements in a Stream , Introduction to Spark Concept , Spark Architecture and components , Spark installation , Spark RDD(Resilient Distributed Dataset) – Spark RDD operations.

				
Dr. A S N CHAKRAVARTHY, Professor, Department of CSE, UCEK JNTUK, Kakinada	Dr. JIMSON MATHEW Professor Dept of Computer Science and Engg. Indian Institute of Technology Patna	Prof. CHAPRAM SUDHAKAR Professor, Department of CSE, National Institute of Technology, Warangal - 506 004 Telangana, INDIA	Mr. RAJESH BOBBURI COO & Director, HighQ Labs Pvt Ltd, Rajahmahendravaram	Mrs. G VIJAYA KUMARI Assoc Professor & HoD Department of CSE-AI & DS, AIML B.V.C.I.T.S. Batlapalem

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 Analysing data

TEXT BOOKS:

1. Wiley & Big Java 4th Edition, Cay Horstmann, Wiley John Sons, INC
2. Hadoop: The Definitive Guide by Tom White, 3rd Edition, O'reilly

REFERENCE BOOKS:

1. Hadoop in Action by Chuck Lam, MANNING Publ.
2. Hadoop for Dummies by Dirk deRoos, Paul C.Zikopoulos, Roman B.Melnyk, Bruce Brown, Rafael Coss
3. Hadoop in Practice by Alex Holmes, MANNING Publ.
4. Big Data Analytics by Dr. A.Krishna Mohan and Dr.E.Laxmi Lydia
5. Hadoop Map Reduce Cookbook, SrinathPerera, ThilinaGunarathne

Software Links:

1. Hadoop: <http://hadoop.apache.org/>
2. Hive: <https://cwiki.apache.org/confluence/display/Hive/Home>
3. Piglatin: <http://pig.apache.org/docs/r0.7.0/tutorial.html>

				
Dr. A S N CHAKRAVARTHY, Professor, Department of CSE, UCEK JNTUK, Kakinada	Dr. JIMSON MATHEW Professor Dept of Computer Science and Engg. Indian Institute of Technology Patna	Prof. CHAPRAM SUDHAKAR Professor, Department of CSE. National Institute of Technology, Warangal - 506 004 Telangana, INDIA	Mr. RAJESH BOBBURI COO & Director, HighQ Labs Pvt Ltd. Rajamahendravaram	Mrs. G VIJAYA KUMARI Assoc Professor & HoD Department of CSE-AI & DS, AIML B.V.C.I.T.S. Batlapalem