AMRITSAR COLLEGE OF ENGINEERING AND TECHNOLOGY

Syllabus

for

Big Data Analytics

ACCS-16503

Theory

B.Tech (CSE) – V Semester

**Section- A(Y Type)**

An Overview of Big Data and Big Data Analytics, Big Data sources, Application areas of Big Data. Understanding Hadoop and its Ecosystem. Brief intro to Hadoop Ecosystem components: Hadoop Distributed File System, MapReduce, YARN, HBase, Hive, Pig, Sqoop, ZooKeeper, Flume, Oozie, Ambari. Understanding a Hadoop cluster.

**Section- B(X Type)**

Overview of HDFS. Architecture of HDFS, Advantages and disadvantages of HDFS, HDFS Daemons, HDFS Blocks, HDFS file write and read, NameNode as SPOF, Hadoop HA, heartbeats, block reports and rereplication, Safemode of Namenode, Hadoop fs commands: cat, ls, put, get, rm, df, count, fsck, balancer, mkdir, du, copyfromlocal, copytolocal.

**Section- C(X Type)**

Hadoop fs commands: expunge, chmod, chown, chgrp, setrep, stat. Hadoop dfsadmin commands. Introduction to Apache Pig, Need of Pig, Installation of Pig, Execution modes of Pig, Pig – Architecture, Grunt shell and basic utility commands, Data types and Operators in Pig, Analysing data stored in HDFS using Pig, Pig operators for Data analysis: Dump, Describe, Explanation, Illustration, Store.

**Section- D(X Type)**

Group, cogroup, join, split, filter, distinct, foreach, order by, limit operators. Functions in Pig: Eval functions, Load and store functions, Bag and tuple functions, String functions, Date time functions, Math functions, Case Studies: Analyzing various datasets with Pig.

**References:**

1. Big Data, Black Book by DT Editorial Services, Dreamtech Press.

2. Hadoop – The Definitive Guide 3rd Edition, Tom White/ OReilly-Yahoo press

3. Hadoop in Action, Chuck Lam/Manning

4. Hadoop – Beginner’s Guide, Garry Turkington/Packt Publishing