Lok Jagruti Kendra University
University with a Difference

Diploma
in
Cloud Computing &
Big Data

Course Code:025100603

Big Data & Big Data Tools

Programme / Branch Name			Diploma in Cloud Computing & Big Data			
Course Name	Big Data & Big Data Tools			**Course Code**		025100603
Course Type	HSSC	BSC	ESC	PCC	OEC	PEC

Legends: HSSC: Humanities and Social Sciences Courses   BSC: Basic Science Courses
ESC: Engineering Science Courses   PCC: Program Core Courses
OEC: Open Elective Courses   PEC: Program Elective Courses

1. Teaching and Evaluation Scheme

Teaching Hours / Week / Credits				Evaluation Scheme			
L	**T**	**P**	**Total Credit**	**CCE**	**SEE (Th)**	**SEE (Pr)**	**TOTAL**
3	0	4	5	50	50	50	150

Legends:

L: Lectures   T: Tutorial   P: Practical
CCE: Continuous & Comprehensive Evaluation
SEE (Th):   Semester End Evaluation (Theory)
SEE (Pr):   Semester End Evaluation (Practical)

2. Prerequisites

✓ Basic knowledge of SQL (queries and sub queries)

3. Rationale

In today's data-driven world efficient operations of organizations in all sectors increasingly depend on effective use of large amounts of data. Big data and big data tools can help you examine this data to uncover hidden patterns, correlations, and other insights. It is a rapidly growing field and skills in this field are among the most sought after today.

4. Objectives

✓ The theory should be taught and practical should be carried out in such a manner that students are able to acquire different learning out comes in cognitive, psychomotor and affective domain to demonstrate following course outcomes.
- To understand the Big Data and its applications.
- To learn use of Apache Hadoop in Big Data.
- To learn NoSQL databases used to store Big Data
- To understand various tools of Hadoop Ecosystem.

5. Contents

Unit No.	Topics	Sub-Topic	Learning Outcomes	% Weightage	Hours
1	**Introduction to Big Data**	1.1 Introduction 1.2 Four Vs of Big Data 1.3 Types of Big Data 1.4 Traditional vs Big Data Approach 1.5 Big Data Analytics 1.6 Big Data Applications 1.7 Case Studies	● To understand the Big Data ● To understand Four V's ● To learn How Big Data is used in Real World	15	6
2	**Introduction to Hadoop**	2.1 History of Hadoop 2.2 Hadoop Distributed File System 2.3 Components of Hadoop 2.4 Analyzing the Data with Hadoop 2.5 Hadoop Cluster 2.6 Cluster Specification 2.7 Cluster Setup and Installation 2.8 Security in Hadoop 2.9 Administering Hadoop	● To learn Hadoop Distributed File System ● To understand Components and setting up a Hadoop Cluster ● To understand Hadoop Security	30	12
3	**MapReduce**	3.1 MapReduce Introduction 3.2 Architecture 3.3 MapReduce Algorithm 3.4 The Map Tasks 3.5 The Reduce Tasks 3.6 Coping with Node Failures 3.7 MapReduce jobs	● To understand the MapReduce ● To understand MapReduce Algorithm ● To understand MapReduce jobs	20	8
4	**NoSQL**	4.1 Introduction of NoSQL 4.2 Uses of NoSQL 4.3 Types of NoSQL Databases 4.4 NoSQL business drivers 4.5 NoSQL data architecture 4.6 Advantages of NoSQL 4.7 NoSQL in Big Data 4.8 SQL Vs NoSQL	● To understand NoSQL database ● To understand its advantages, features, etc. ● To use NoSQL in Big Data	15	8
5	**Hadoop Ecosystem**	5.1 Introduction to Pig Data Flow Engine 5.2 Data processing using Pig 5.3 Introduction of Hive 5.4 Architecture of Hive 5.5 Hive Differs from traditional RDBMS 5.6 HiveQL	● To understand use of Pig. ● To understand the Hive and its services. ● To learn fundamentals of	20	8

		5.7 Querying Data in Hive 5.8 Fundamentals of HBase and Zookeeper	HBase and Zookeeper.		

	Total Hours	**42**

6. List of Practicals / Exercises

The practical/exercises should be properly designed and implemented in an attempt to develop different types of skills so that students can acquire the competencies/programme outcomes. Following is the list of practical exercises for guidance.

Sr. No.	Practicals /Exercises	Key Competency	Hours
1	Install and configure big data tools like Hadoop, Pig, Hive for Windows.	To install various big data tools.	4
2	Perform various Hadoop HDFS Operations: Copying File to Hadoop. Copy from Hadoop File system and deleting file. Moving and displaying files in HDFS.	To learn basic operations of Hadoop HDFS.	2
3	Implement Distributed Cache in Hadoop and Spark.	To apply concept of Distributed Cache.	2
4	Implement matrix manipulations like multiplication, sorting, indexing, etc. using MapReduce.	To learn matrix manipulations using MapReduce.	6
5	Implement Map side join in Hadoop.	To apply Map side join.	4
6	Implement Reduce side join in Hadoop.	To apply Reduce side join.	4
7	Write a program to implement word count program using MapReduce.	To understand Word count problem	4
8	To be able to perform basic operation of NoSQL databases.	To understand NoSQL operation.	2
9	Install HBase on windows with the help of Zookeeper.	To be able to install ZooKeeper.	2
10	Create a table in Hbase and insert five student detail in table and retrieve from it.	To implement basic Hbase operations.	2
11	Retrieve multiple rows of student detail than delete the single row in Hbase.	To be able to perform various operations in Hbase	2
12	Implement basic program to show date and time with the help of Pig.	To show date and time in Pig	2
13	Write a program to show data of 10 employees in Pig with the help of tuple and bag.	To show data using Bag of Tuple.	2
14	Write a program to load data from file into relation in Pig with the help of filter.	To load data from file into relation.	2

14	Implementing any one Clustering algorithm (K-Means/CURE) using MapReduce.	To implement any clustering Algorithm.	4
15	Implementing any one Association Rule algorithm using MapReduce.	To implement any Association Rule Algorithm.	4
16	To learn about the Hive commands using HiveQL/HQL (DDL and DML).	To perform HiveQL operations.	4
17	Create HIVE Database and Descriptive analytics-basic statistics, visualization using Hive/PIG	Use HiveQL to analyse and visualize data.	4
		Total Hours	**56**

7. Suggested specification Table with hours

Unit No.	Chapter Name	Teaching Hours	Distribution of Topics According to Bloom's Taxonomy					
			R %	U %	App %	C %	E %	An %
1	Introduction to Big Data	6	30	30	20	-	10	10
2	Introduction to Hadoop	12	20	30	30	5	5	10
3	MapReduce	8	20	30	20	10	10	10
4	NoSQL	8	30	30	20	5	5	10
5	Hadoop Ecosystem	8	20	20	30	10	10	10

8. Text Books

1) Hadoop: The Definitive Guide, Tom White, O'Reily Media
2) Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, Bill Franks, Wiley
3) Hadoop in Action, Chuck P. Lam, Mark W. Davis, Ajit Gaddam, Manning Publications

9. Reference Books

1) Big Data and Analytics, Seema Acharya and Subhashini Chellappan, Wiley India
2) Mining of Massive Datasets, Anand Rajaraman and Jeff Ullman, Cambridge University Press,
3) Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Michael Minelli, Michele Chambers, Ambiga Dhiraj, Wiley India

10. Open Sources (Website, Video, Movie)

1) https://nptel.ac.in/courses/106104189
2) https://www.coursera.org/learn/introduction-to-big-data-with-spark-hadoop
3) https://www.simplilearn.com/big-data-and-analytics/big-data-and-hadoop-training
4) https://hadoop.apache.org/docs/current/
5) https://www.udemy.com/course/the-ultimate-hands-on-hadoop-tame-your-big-data