



**GOVERNMENT OF TAMILNADU**  
**DIRECTORATE OF TECHNICAL EDUCATION, CHENNAI**  
**STATE PROJECT COORDINATION UNIT**  
**(Established under Canada India Institutional Cooperation Project )**

**CURRICULUM**

Course Name	<b>Big Data analytics using Hadoop for beginners</b>
Course Code	CSE/2020/027
Course Duration	50 Hours
Minimum Eligibility Criteria and Pre-requisites (if any)	ITI/10 <sup>th</sup> /+2/Diploma/Graduates Basic knowledge of computer with database
Course Objectives	The main objectives of the training module is to : <ul style="list-style-type: none"> <li>• Understand fundamentals of Concepts in Bigdata</li> <li>• Understand fundamentals of Hadoop</li> <li>• Use the HDFS file system</li> <li>• Understand basic Map Reduce programming concepts</li> </ul>
Course Outcomes	At the end of training, the participants will be able to <ul style="list-style-type: none"> <li>• Install and work with Hadoop</li> <li>• Debug and run simple Java programs for hdfs.</li> <li>• Develop applications in Java to implement MapReduce program.</li> </ul>
Expected Job Roles	Big Data analytics programmer

**TEACHING AND SCHEME OF EXAMINATION**

Course Code	Course Name	Hours		Assessment Marks		Duration of Examination
				Min	Max	
CSE/2020/027	Big Data analytics using Hadoop for beginners	Theory	15	10	20	3 Hours
		Practical	35	40	80	
		Total	50	50	100	

**CSE/2020/027 - BIG DATA ANALYTICS USING HADOOP FOR BEGINNERS****DETAILED SYLLABUS**

<b>UNIT NO</b>	<b>MODULES</b>	<b>NO.OF.HOURS THEORY</b>
<b>I</b>	<b>INTRODUCTION TO BIG DATA</b>	
1.1	Introduction to Big Data Objectives Need for Big Data Three Characteristics of Big Data Characteristics of Big Data Technology Appeal of Big Data Technology Handling Limitations of Big Data	3
<b>II</b>	<b>INTRODUCTION TO HADOOP</b>	
2.1	Introduction Architecture Ecosystem Components Install and Configure Hadoop	3
<b>III</b>	<b>WORKING WITH HDFS</b>	
3.1	Hadoop distributed file system (HDFS) HDFS Daemons: Name Node Data Node Working with HDFS Commands.	3
<b>IV</b>	<b>DIFFERENT PHASES IN MAP REDUCE</b>	
4.1	Input – Output formats in each phase Modeling Real World applications into Map Reduce Understanding Map Reduce Program Execution	3
<b>V</b>	<b>INTRODUCTION TO APACHE HIVE</b>	
5.1	Introduction to Hive Meta store SQL vs. Hive Hive Query language Managed and External tables Querying data	3
<b>Total Theory Hours</b>		<b>15</b>
<b>Total Practical Hours</b>		<b>35</b>
<b>Total Hours</b>		<b>50</b>

## **PRACTICALS ( 35 HOURS )**

1. Installing Hadoop in Standalone mode.
2. Installing Hadoop in Psuedo Distributed Mode.
3. File management tasks in Hadoop.
4. Implementation of Map Reduce program using Wordcount mapreduce.
5. Implementation of Map Reduce program using Temperature analysis.
6. Implementation of Map Reduce program using Matrix multiplication.
7. Working with Hive.
8. Creating managed tables in Hive.
9. Creating external tables in Hive.
10. Working with table properties, managing tables in Hive.
11. Installing Hadoop in Standalone mode
12. Installing Hadoop in Psuedo Distributed Mode
13. File management tasks in Hadoop
14. Implementation of Map Reduce program using Wordcount mapreduce.
15. Implementation of Map Reduce program using Temperature analysis
16. Implementation of Map Reduce program using Matrix multiplication
17. Working with Hive.
18. Creating managed tables in Hive.
19. Creating external tables in Hive.
20. Working with table properties, managing tables in Hive.

### HARDWARE REQUIREMENT

S.NO	LIST OF TOOLS /EQUIPMENTS
1	Computer
2	Printer

### SOFTWARE REQUIREMENT

S.NO	LIST OF SOFTWARE
1	Java
2	Oracle 11 i

### REFERENCE BOOKS

S.NO	NAME OF THE BOOK	AUTHOR	PUBLISHER
1.	Big Data and Hadoop Paper back – 1 January 2017	V.K.Jain	Kashnna Publishing
2.	Big Data Analytics Paper back – 1 January 2021	<u>Avinash Jha</u> <u>Pinal M. Hansora</u>	Technical Publishing

## ASSESSMENT AND CERTIFICATION

S.No	Criteria for assessment
1.	A trainee will be assessed based on the performance in End Examination for Theory and Practical conducted internally in the Project Polytechnic College for a duration of 3 hours
2.	A trainee must have 75% of attendance to appear for End examination in Theory and Practical.
3.	The assessment for theory part will be based on the marks scored in the end examination on the knowledge bank of questions (1 word/objective type questions)
4.	The assessment for practical part will be based on the marks scored in the end examination conducted by the Project Polytechnic and assessed by the Examiners approved by Strategic Plan Implementation Committee (SPIC) of the project polytechnic.
5.	The criteria for successful completion of training is every trainee should score 50% of marks in theory and practical examination.
6.	On successful completion of training , Certificate will be issued to the participants by the Directorate of Technical Education through the Project Polytechnic.

## END EXAMINATION

### ALLOCATION OF MARKS

S.No	Description	Max.Marks
1.	Theory Examination	20
2.	Practical Examination	
	a)Procedure	10
	b)Execution	30
	c)Output	20
	d)Record	20
<b>Total Marks</b>		<b>100</b>

## THEORY MODEL QUESTION PAPER

### CSE/2020/027 - BIG DATA ANALYTICS USING HADOOP FOR BEGINNERS

(Maximum Marks : 20 )

(N.B: Answer any **twenty** questions)

**20 x 1 = 20 Marks**

1. What is the definition of Big Data technology?
2. What are the Characteristics of Big Data?
3. What are the Steps involved to deploy a Big Data solution?
4. How many forms Big Data could be found?
5. List any two benefits of Big Data Processing.
6. What are the main components of Big Data?
7. What is Hadoop in big data?
8. What are the two main components of Hadoop?
9. Expand HDFS and YARN .
10. What are the two types of Name Nodes in hadoop.
11. What does 'jps' command do?
12. Name the three modes in which Hadoop can run.
13. What is "MapReduce"?
14. What is the syntax to run a "MapReduce" program?
15. What is the command to empty the trash in HDFS?
16. Write the syntax for mkdir command
17. What is the use of ls command in HDFS?
18. What is heartbeat in HDFS?
19. What are the basic parameters of a Mapper?
20. What is a difference between an Input Split and HDFS Block?
21. What does partitioner do?
22. What is the difference between HDFS block and InputSplit?
23. Is it possible to change the default location of a managed table?
24. What are Buckets in Hive?
25. What are the types of tables in Hive?