CMR ENGINEERING COLLEGE
UGC Autonomous

(Approved by AICTE-New Delhi, Affiliated to JNTUH and Accredited by NAAC)
Kandlakoya (V), Medchal Road, Hyderabad-501401
Ph:08418 200037, 9247022662,Fax: 08418 200240, www.cmrec.ac.in

Department of CSE- DataScience

LIFT Planner (Laboratory Improvement for Future Trends)

On

BIG DATA ANALYTICS LAB (R22)

2025-2026

BIG DATA ANALYTICS LEARNING

Sl. No	CONTENT
1	OBJECTIVES AND RELEVANCE
2	SCOPE
3	PREREQUISITES
4	SYLLABUS AS PER AUTONOMOUS
5	LEAD EXPERIMENTS
6	LAB SCHEDULE
7	SUGGESTED BOOKS
8	WEBSITES (USEFUL LINKS)
9	EXPERT DETAILS
10	MAPPING OF LAB WITH PROJECT/CONSULTANCY/R & D PROPOSALS
11	INDUSTRIAL VISITS (IIP – INNOVATIVE INDUSTRIAL LEARNING PROGRAM)
12	SHADOW ENGINEERING
13	CALIBIRATION, TESTING AND INSPECTION

14	**MAINTAINANCE AND TROUBLESHOOTING**
15	**ASSESSMENT AND ACCREDITATION PROCEDURE AS PER NABL**
16	**VIRTUAL LABS REFERENCE LINKS**

1.OBJECTIVES AND RELEVANCE &SCOPE

Lab Objectives:

1. Create a Hadoop Cluster

Objective:

- To understand the architecture and components of a Hadoop Cluster.
- To configure and set up a multi-node Hadoop environment.
- To validate Hadoop installation and test its basic functionality.

2. **Implement a Simple Map-Reduce Job for an Inverted Index**

Objective:

- To understand the MapReduce programming model.
- To implement a program that constructs an inverted index from a set of documents.
- To run the MapReduce job on a Hadoop cluster and analyze its output.

3. **Process Big Data in HBase**

Objective:

- To understand the data model and architecture of HBase.
- To create HBase tables and load data into them.
- To perform basic read, write, and delete operations on HBase tables.

4. **Store and Retrieve Data in Pig**

Objective:

- To learn the basic syntax and commands of the Pig Latin language.
- To load, process, and transform datasets stored in HDFS.
- To save processed results to files using Apache Pig.

5. **Perform Data Analysis Using MongoDB**

Objective:

- To understand the document-oriented structure of MongoDB.
- To perform basic database operations such as insert, query, and aggregate.
- To utilize MongoDB for data analysis and extracting actionable insights.

6. Use Power Pivot (Excel) for Big Data Analytics and Charting

a. Big Data Analytics Objective:

- To import large datasets into Excel using Power Pivot.
- To create data models and relationships between tables.
- To perform calculations and aggregations using DAX expressions.

b. Big Data Charting Objective:

- To create interactive and insightful charts and dashboards from large datasets.
- To utilize Power Pivot and Excel's charting tools for visual analytics.
- To gain proficiency in summarizing and presenting big data findings effectively.

Relevance:

1. Create a Hadoop Cluster

Introduction:
Hadoop is an open-source framework for processing and storing big data across a distributed set of servers. This experiment guides you to create a multi-node Hadoop cluster.

Prerequisites:

- Knowledge of Linux shell commands
- Hadoop binaries and Java installed
- At least 2–3 servers configured with SSH access
- Understanding of HDFS and YARN architecture

Relevance:
Setting up a Hadoop Cluster is a foundational skill for working with big data. It prepares you for roles like Big Data Developer, Hadoop Administrator, and Data Engineer.

Scope:
Enables hands-on experience with distributed storage and processing.
Develops ability to configure, manage, and scale Hadoop environments.
Supports deployment of big data applications across nodes.
Lays the foundation for learning advanced big data technologies.

2. Implement a Simple Map-Reduce Job (Inverted Index)

Introduction:
An inverted index is a data structure that maps words to their occurrences in documents. In this lab, you will implement a MapReduce program for building an inverted index from a set of text files.

Prerequisites:

- Understanding of Java programming and basic knowledge of Hadoop
- Hadoop installed and configured
- Sample text files available in HDFS

Relevance:
Inverted indexes form the basis for search and text analytics platforms. This lab is vital for understanding how big data search and retrieval systems work.

Scope:
Enables understanding of the MapReduce programming paradigm.
Develops proficiency in distributed text processing and indexing.
Supports building search and information retrieval applications.
Provides foundation for large-scale data analytics and text mining.

3. Process Big Data in HBase

Introduction:
HBase is a distributed, column-oriented NoSQL database built on HDFS for storing and managing massive amounts of data.

Prerequisites:

- Hadoop and HBase installed and configured
- Understanding of NoSQL databases and HBase architecture
- Knowledge of HBase shell commands or APIs (Java/Python)

Relevance:
HBase is ideal for use cases with billions of rows and low-latency access. It prepares you for big data roles in e-commerce, social media, and IoT.

Scope:
Enables experience with a distributed NoSQL database for big data.
Develops proficiency in data modeling, storage, and retrieval.
Supports low-latency access required for real-time analytics.
Prepares for managing massive, sparse, and schema-less data.

4. Store and Retrieve Data in Pig

Introduction:
Pig is a high-level data flow language for processing large datasets stored in HDFS. This experiment demonstrates data loading, processing, and storage using Pig.

Prerequisites:

- Hadoop environment configured
- Apache Pig installed and configured
- Understanding of Pig Latin syntax and data processing concepts

Relevance:
Pig allows data analysts and engineers to write complex analytics jobs quickly. It's highly relevant for building big data pipelines and batch analytics.

Scope:
Enables proficiency in high-level scripting for data processing.
Develops ability to transform and analyze large HDFS datasets.
Supports building data pipelines for batch analytics.
Bridges the gap between raw big data and actionable results.

5. Perform Data Analysis Using MongoDB

Introduction:
MongoDB is a popular NoSQL database that stores data in a flexible, JSON-like structure. This lab focuses on data insertion, retrieval, and analysis.

Prerequisites:

- Understanding of JSON and NoSQL database concepts
- MongoDB installed and configured
- Knowledge of basic MongoDB commands (`insert`, `find`, `aggregate`)

Relevance:
MongoDB is ideal for working with unstructured and semi-structured data. Its scalability and versatility make it essential for big data engineering and analytics.

Scope:
Enables working with unstructured and semi-structured data.
Develops skills for flexible data storage and quick retrieval.
Supports building analytics applications in a NoSQL environment.
Prepares for roles in big data, IoT, and real-time analytics.

6. Use Power Pivot (Excel) for Big Data Analytics and Charting

Introduction:
Power Pivot allows Excel users to import, model, and analyze big data within a familiar environment. This lab explores creating relationships, calculations, and dashboards.

Prerequisites:

- Microsoft Excel with Power Pivot enabled
- Understanding of Excel and basic data modeling concepts
- Knowledge of creating charts and pivot tables in Excel

Relevance:
Power Pivot bridges big data and business analytics, making it vital for actionable insights and data-driven decisions across industries.

Scope:
Enables integration of big data with Excel for analytics.
Develops proficiency in data modeling, calculations, and dashboards.
Supports actionable insights for business intelligence.
Bridges technical big data platforms with business reporting needs.

3. PREREQUISITES:

1. Basic knowledge of operating systems (Linux/Windows).
2. Java Development Kit (JDK) installed.
3. Hadoop ecosystem components installed and configured.
4. Basic understanding of Big Data concepts.
5. Sufficient system memory and storage for running Big Data tools.
6. Internet connection for downloading required software and datasets.

4. SYLLABUS AS PER AUTONOMOUS

List of Experiments

1. Create a Hadoop cluster
2. Implement a simple map-reduce job that builds an inverted index on the set of input documents (Hadoop)
3. Process big data in HBase
4. Store and retrieve data in Pig
5. Perform data analysis using MongoDB
6. Using Power Pivot (Excel) Perform the following on any dataset
   a. Big Data Analytics
   b. Big Data Charting

5. Lead Experiment

1. Develop a MapReduce program to analyze Uber data set to find the days on which each basement has more trips using the following dataset. The Uber dataset consists of four columns they are dispatching_base_number date active_vehicles trips

2. Write queries to sort and aggregate the data in a table using HiveQL

1. Develop a MapReduce program to analyze Uber data set to find the days on which each basement has more trips using the following dataset. The Uber dataset consists of four columns they are dispatching_base_number date active_vehicles trips.

AIM:

To develop a **MapReduce program** that analyzes the Uber dataset to identify the **days on which each dispatching base** has a higher number of trips, based on the fields **dispatching_base_number, date, active_vehicles, and trips**. This will enable understanding of trip activity patterns across different bases over time.

Uber.java

```
import java.io.IOException; import java.text.ParseException; import java.util.Date;
importorg.apache.hadoop.conf.Configuration;import
org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable; import
org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job; import
org.apache.hadoop.mapreduce.Mapper; import
org.apache.hadoop.mapreduce.Reducer;

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
```

```
importorg.apache.hadoop.mapreduce.lib.output.FileOutputFormat;public class Uber {
public static class TokenizerMapper

extends Mapper<Object, Text, Text, IntWritable>{java.text.SimpleDateFormat format = new
java.text.SimpleDateFormat("MM/dd/yyyy");
String[] days ={"Sun","Mon","Tue","Wed","Thu","Fri","Sat"};

private Text basement = new Text();

Date date = null;

Private int trips;

public void map(Object key, Text value, Context context

) throws IOException, InterruptedException {String line = value.toString();
String[] splits = line.split(",");
basement.set(splits[0]);

                                    try {
date = format.parse(splits[1]);

                                    } catch (ParseException e) {
// TODO Auto-generated catch block
e.printStackTrace();

                                    }
trips = new Integer(splits[3]);

String keys = basement.toString()+ " " +days[date.getDay()];context.write(new Text(keys),
new IntWritable(trips));
}
                                    }
public static class IntSumReducer

extends Reducer<Text,IntWritable,Text,IntWritable> {private IntWritable result
= new IntWritable();
public void reduce(Text key, Iterable<IntWritable> values,Context context

) throws IOException, InterruptedException {


int sum = 0;

for (IntWritable val : values) {sum +=
val.get();
}

result.set(sum); context.write(key,
result);
}

                                    }
public static void main(String[] args) throws Exception {Configuration conf = new
Configuration();

Job job = Job.getInstance(conf, "Uber"); job.setJarByClass(Uber.class);
job.setMapperClass(TokenizerMapper.class); job.setCombinerClass(IntSumReducer.class);
job.setReducerClass(IntSumReducer.class); job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class); FileInputFormat.addInputPath(job, new
```

```
Path(args[0])); FileOutputFormat.setOutputPath(job, new Path(args[1]));
System.exit(job.waitForCompletion(true) ? 0 : 1);
}

}
```

Download dataset:

https://raw.githubusercontent.com/shukladiwakar/Mapreduce-for-Uber-Data-Analysis/master/uberdata

Commands for execution:

```
hadoop com.sun.tools.javac.Main *.java

jar cf Uber.jar *.class

hadoop fs -mkdir -p /user/565/uber/input/

hadoop fs -put -f uberdata /user/565/uber/input/

hadoop jar Uber.jar Uber /user/565/uber/input
 /user/565/uber/output

hadoop fs -cat /user/565/uber/output/part-r-00000
```

output :

```
B02512 Fri     16435
B02512 Mon    11297
B02512 Sat      15026
B02512 Sun    10487
B02512 Thu    15809
B02512 Tue     12041
```

2. Write queries to sort and aggregate the data in a table using HiveQL

AIM:

To write and execute **HiveQL queries** to sort and aggregate data within a Hive table, facilitating the analysis of patterns, trends, and summaries from large datasets.

Make dataset : emo_data.csv

123, Den, 11000, Raphaely
124,Karen,2500,Colmenares
125,Susan,6500,Mavris
126,Jason,3300,Mallin
127,Alexis,4100,Bull
128,Kevin,3000,Feeney
129,Curtis,3100,Davies
130,John,2700,Seo
131,Stephen,3200,Stiles
132,Winston,3200,Taylor
133,James,2500,Marlow
134,Steven,2200,Markle
135,James,2400,Landry
136,Kevin,5800,Mourgos
137,Donald,2600,OConnell
138,Douglas,2600,Grant
139,Girard,2800,Geoni
140,Jean,3100,Fleaur
141,David,4800,Austin

Running HIVE commands:

hive

hive> create table emp (Id int, Name string , Salary float,Department string)

row format delimited fields terminated by
',' ;

hive> LOAD DATA LOCAL INPATH

'/home/hadoop/565/hive_sql/emp_data.csv' INTO TABLE emp;hive> SELECT *

FROM emp;

hive> select * from emp sort by salary desc;

Output:

123	Den	11000.0	Raphaely
125	Susan	6500.0	Mavris
136	Kevin	5800.0	Mourgos
141	David	4800.0	Austin
127	Alexis	4100.0	Bull
126	Jason	3300.0	Mallin
132	Winston	3200.0	Taylor
131	Stephen	3200.0	Stiles

6. LAB SCHEDULE

Week No.	Name of the Program	No. of Lab hours	Text Books	Mode of Assessment
1.	Create a Hadoop cluster	6	T3	Viva and Execution
2.	Implement a simple map-reduce job that builds an inverted index on the set of input documents (Hadoop)	6	T3	Viva and Execution
3.	Process big data in HBase	6	T3	Viva and Execution
4.	Store and retrieve data in Pig	6	T1	Viva and Execution
5.	Perform data analysis using MongoDB	6	T2	Viva and Execution
6.	Using Power Pivot (Excel) Perform the following on any dataset a. Big Data Analytics b. Big Data Charting	3	T4	Viva and Execution
Total no of Lab hours required to complete syllabus		36		

CYCLE 1:

Batches	week-1	week-2	week-3	week-4	week-5	week-6	week-7	week-8
B1	Basic programs & Exp1	Exp.1	Exp.2	Exp.2	Exp.3	Exp.3	Exp.4	Exp.4
B2	Basic programs & Exp1	Exp.1	Exp.2	Exp.2	Exp.3	Exp.3	Exp.4	Exp.4

Batches	week-9	week-10	week-11	week-12	week-13
B1	Exp.5	Exp.5	Exp. 6	LEAD 1	LEAD 2
B2	Exp.5	Exp.5	Exp. 6	LEAD 1	LEAD 2

VIVA SCHEDULE:

Batches	week-1	week-2	week-3	week-4	week-5	week-6	week-7	week-8	week-9	week-10	week-11	week-12	Week 13
B1	viva	viva	viva	viva	viva	viva	Viva	viva	viva	viva	viva	Viva	Viva
B2	viva	viva	viva	viva	viva	viva	Viva	viva	viva	viva	viva	Viva	Viva

Batch 1	**week-1**	**week-2**	**week-3**	**week-4**	**week-5**	**week-6**	**week-7**	**week-8**	**week-9**	**week-10**
SG1	viva									
SG2			viva							
SG3					viva					
SG4							viva			
SG5									viva	

Batch 2	**week-1**	**week-2**	**week-3**	**week-4**	**week-5**	**week-6**	**week-7**	**week-8**	**week-9**	**week-10**
SG1	viva									
SG2			viva							
SG3					viva					
SG4							viva			
SG5									viva	

*SG: Selected Group with a maximum of 6 or 12 students

Scheme of Evaluation:

The scheme of evaluation for internal and external exams is as follows:

LAB INTERNAL:								
Day to Day Evaluation - 10					**Internal Exam - 30**			
Uniform	**Observation & Record**	**Performance Of the Experiment**	**Result**	**Viva**	**Write up**	**Execution & Results**	**Viva**	**Project work**
Marks - 2	Marks - 2	Marks - 2	Marks - 2	Marks - 2	Marks - 05	Marks - 05	Marks - 10	Marks - 10
Total Marks = 40								

S.NO	Write up	Experiment Program	Evaluation of results	Presentation	Viva
1	1. Aim 2. Procedure 3. Program 4. Expected output.	Based on correctness of the program	Based on correctness of the program and Results	Presentation of another program	Based on understanding of Experiment and theoretical questions in the related subjects
Marks	10	15	15	10	10
Total Marks-60					

7. SUGGESTED BOOKS

1. Big Data Analytics, Seema Acharya, Subhashini Chellappan, Wiley 2015.
2. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business, Michael Minelli, Michehe Chambers, 1st Edition, Ambiga Dhiraj, Wiely CIO Series, 2013.
3. Hadoop: The Definitive Guide, Tom White, 3rd Edition, O"Reilly Media, 2012.
4. Big Data Analytics: Disruptive Technologies for Changing the Game, Arvind Sathi, 1st Edition, IBM Corporation, 2012.

REFERENCE BOOK:

1. Big Data and Business Analytics, Jay Liebowitz, Auerbach Publications, CRC press (2013).
2. Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop, Tom Plunkett, Mark Hornick, McGraw-Hill/Osborne Media (2013), Oracle press.
3. Professional Hadoop Solutions, Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Wiley, ISBN: 9788126551071, 2015.
4. Understanding Big data, Chris Eaton, Dirk deroos et al., McGraw Hill, 2012.
5. Intelligent Data Analysis, Michael Berthold, David J. Hand, Springer, 2007.
6. Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, Bill Franks, 1st Edition, Wiley and SAS Business Series, 2012.

8. WEBSITES (USEFUL LINKS)

1. Mastering Big Data Analytics - Great Learning (mygreatlearning.com)
2. **https://nptel.ac.in/courses/106104189**
3. https://medium.com/@ericsson/inverted-index-using-mapreduce-in-hadoop-7d3d21f2ffe0
4. https://www.tutorialspoint.com/hbase/index.htm
5. https://www.tutorialspoint.com/apache_pig/index.htm
6. https://www.w3schools.com/mongodb/
7. https://www.excel-easy.com/data-analysis/power-pivot.html
8. https://www.excelcampus.com/powerpivot/power-pivot-charts/
9. https://www.edureka.co/blog/hadoop-cluster/

9. EXPERT DETAILS

The expert details which have been mentioned below are only a few of the eminent ones known Internationally, Nationally and Locally.

➢ International

1. Carlie Iodine: she is the research director for data science and business analytics for IT leaders.
2. Kirk Borne: He is a principal data scientist at Booz Allen Hamilton. He was an astrophysicist formerly working at NASA.

➢ National:

1. Prof. D.Lakshmi,
Professor, Dept.
of CSE
VIT Bhopal – INDIA

2. Ms.Palak Gupta

Assistant Professor Gmail: palak.gupta@jagannath.org

➢ Local

1. Prof. Shameem Fathima
Professor Dept. Of CSE,
O.U – Hyderabad
E-mail:sameenf@gmail.com
Phone :+91 9848519860
040-27097125

2. Prof. C Krishna Mohan
Professor
Department of Computer Science and Engineering & Dean
(Public and Corporate Relations)
Email : ckm@cse.iith.ac.in Mobile
:94917 12312
Website : www.iith.ac.in/-~ckm/

10. MAPPING OF LAB WITH PROJECT/CONSULTANCY/R & D PROPOSALS

Business Intelligence Dashboard for Taxi Services

Abstract:

In this project, Power Pivot, an advanced Excel add-in, is used to transform and analyze taxi trip data for actionable business insights. The goal is to enable decision-makers to visually understand patterns like daily trips, busiest dispatch bases, average active vehicles, and trends over time. By integrating large datasets into Power Pivot, applying calculated measures, and creating interactive charts, this project simulates a robust, cost-efficient Business Intelligence (BI) solution for taxi service management.

Objectives:

- Import and model large taxi trip datasets in Excel using Power Pivot.
- Perform data cleaning, filtering, and aggregation.
- Create calculated measures (such as average trips per base, trips per day).
- Build interactive dashboards and charts (line charts, bar charts, pivot tables).
- Enable quick trend analysis and data-driven decision making.

Tools and Technologies:

- **Power Pivot (Excel)**: Enables in-memory data processing and modeling.
- **Power Query**: To clean and transform data.
- **Excel Pivot Tables and Charts**: To build interactive dashboards.

Approach:

1. Import the taxi trip dataset (columns: dispatching_base_number, date, active_vehicles, trips).
2. Use Power Query for extracting, cleaning, and shaping the dataset.
3. Build a relational data model within Power Pivot.
4. Create calculated fields (e.g., total trips per base, average trips per day).
5. Visualize the results with Pivot Charts and Slicers.
6. Analyze the results for actionable insights (e.g., busiest dispatch bases, best days for resource allocation).

Expected Outcomes:

- Interactive dashboards providing insights into taxi service efficiency.
- Reports highlighting daily activity trends and busiest dispatch bases.
- Actionable recommendations for optimizing taxi dispatch and pricing.
- An Excel workbook that serves as a cost-efficient BI solution.

11. INDUSTRIAL VISITS (IIP–INNOVATIVE INDUSTRIAL LEARNING PROGRAM)

Students visit schedule different software development MNC companies.

S.No	Type of industry	Nature of industry	Date of visit (Tentative Date)	No. of students participated	Year/branch	remarks
1						
2						

12. Shadow Engineering – Objectives

- To enhance students' understanding and practical knowledge through active participation in laboratory work.
- To strengthen industry-academia collaboration for better exposure and learning opportunities.
- To simulate an industry-like environment within the institution, preparing students for future assignments and professional roles.
- To expose students to real-world engineering tools, technologies, and practices aligned with industry standards.
- To encourage peer learning, mentoring, and collaborative problem-solving through hands-on mini-projects and real-time scenarios.
- To develop professional skills such as teamwork, communication, and project documentation alongside technical proficiency.
- To cultivate an innovation-driven mindset by enabling students to apply theoretical concepts to practical applications.
- To facilitate domain-specific technical training and certification opportunities within the academic setup.
- To prepare students for internships, placements, and higher studies by familiarizing them with the expectations of the engineering profession.
- To continuously assess and improve students' performance through feedback from both academic mentors and industry experts.

13. CALIBIRATION, TESTING AND INSPECTION

Calibration

- A functional Hadoop cluster configured with NameNode, DataNode, and Resource Manager running properly.
- An inverted index listing words and the documents in which they occur, verifying successful data processing.
- An HBase table with dispatching base data that supports quick read and write access for taxi trip records
- Output files containing aggregated statistics such as total trips per dispatch base and average trips per day.
- A structured MongoDB database with collections for taxi trips, supporting quick trend and pattern analyses.
- Interactive dashboards and charts providing insights into taxi trip trends and patterns.

Installation

Aim of this concept is to make and maintain installation procedure for a new equipment or already existing equipment. Download java from https://www.oracle.com/java/technologies/downloads/ Hadoop Tool from https://hadoop.apache.org/releases.html and MongoDB from https://www.mongodb.com/docs/manual/installation/ or internet and install then we are getting this below GUI. According to user requirement select any label or button then proceed.

14. MAINTAINANCE AND TROUBLESHOOTING

Maintenance:

Maintenance and trouble shooting of each equipment in a laboratory must follow the following guidelines:

Maintenance Schedules:

(1) Preventive Maintenance Schedules of lab will be decided by lab in charge along with concerned HOD. The details of schedule should be recorded in the following template of format.

S.No.	Software	Date of Maintenance	Type of Activity	Remarks
1	Operating System	11-3-2025	Software installed	
2	java	9-4-2025	Installation	
3	hadoop	29-4-2025	Installation	

(2) Maintenance Reports should be duly signed by in charges as well as HODs and duly approved by Principal periodically.

15. ASSESSMENT AND ACCREDITATION PROCEDURE AS PER NABL

Accreditation is the formal recognition, authorization and registration of a laboratory that has demonstrated its capability, competence and credibility to carry out the tasks. It provides the feedback to laboratories as to whether they are performing according to the technical competence as per guidelines of NABL (National Accreditation Board for Testing and Calibration Laboratories).

The laboratory should carry out the following important tasks towards getting ready for accreditation from NABL.

1. Preparation of methodology for each experiment
2. Preparation of Standard Operating procedure for each equipment
3. Preparation of Laboratory Manual as per the guidelines specified by Combined Lab Team(CLT) headed by Principal/HOD/Dean/in charge
4. Ensure Effective environmental conditions (temperature, humidity, storage and placement) in the laboratories by implementing proper housekeeping and cleaning of the equipments from dust, dirt etc.
5. Ensure Calibration of instruments/equipment (Only NABL accredited authorized laboratories provide calibration).
6. All the details of Calibration should be included in the format specified exclusively for calibration procedure.
7. Ensure proper implementation of all the documents, formats to be included in the lab manual.
8. Impart training for all the technicians working in labs about the importance of documentation, log sheets, operating procedure of the lab.
9. Incorporate Internal Lab audits for effective functioning of the laboratories. Audits may be once in a month or 3 months or at the end of the semester. The audit schedule will be decided by the Chairman and Principal of the CLT team. Auditors should submit the detailed report of each lab duly signed by them to the Principal.

16. VIRTUAL LABS REFERENCE LINKS

- https://aws.amazon.com/emr/
- https://www.cloudera.com/tutorials.html
- https://www.cloudera.com/tutorials.html
- https://pig.apache.org/docs/r0.17.0/
- https://learn.mongodb.com/
- https://learn.microsoft.com/en-us/power-query/power-query-tutorial