Lab Record

MCA:MC5191 / MSc:MS5091

Big Data Analytics

Subject Teacher:
Prof. M. K. Rath

Lab Objective & Requirements

Objective:

Upon successful completion of this course, students should be able to:

- Understand and implement the basics of data, statistics and graphs (charts) using Spreadsheet and Python.
- Demonstrate the knowledge of big data analytics and implement different file management task in Hadoop.
- Understand Map Reduce Paradigm and develop data applications using variety of systems.
- Analyze and perform different operations on data using Pig Latin scripts.
- Illustrate and apply different operations on relations and databases using Hive & NoSQL: mongoDB

Software Requirement:

- OS: Linux / Windows 10/11
- Language & Tools: Python 3.x, JAVA, mongoDB (with compass), Spreadsheet (MS-Excel / WPS)

Hardware Requirement:

- Processor: 64-bit multi-core
- RAM: 16GB or Higher

INDEX PAGE (page-1)

Sr. No.	List of Experiments	Page Number
01	Experiment Name1	p - p
02	Experiment Name2	p - p
03	Experiment Name3	p - p
. . .		
. . .	Experiment Name Last	p - p

Lab Record Requirement

[to be submitted at the time of Sessional]

- There are 12 experiments in all.
- Prepare them in physical form, mostly in handwritten type (don’t take printouts except the outputs).
- Out of which a minimum of 2 experiments are expected to be in video form (screen-recording with audio narration).
- The link of these videos are to be mentioned in the Lab record with the respective experiment number.

Experiment: Ogive Graph

Draw an Ogive or cumulative frequency curve for a give set of data e.g. the marks and number of students.

Mentions the steps and necessary codes / formulas to process data and draw charts by using any spreadsheet applications.

Experiment-1: Using Spreadsheet Application.

Experiment-2: Using Python 3.x and its necessary packages.

Show the output in a new page by drawing or by taking printouts.

Experiment: Statistics and Charts / Graphs

Experiment-3: Write a Python script to compute the Mean, Median, Mode, SD, Variance of a given data (list).

E.g. speed = [99,86,87,88,111,86,103,87,94,78,77,85,86]

Three Different Experiments - 4, 5, 6: Create different charts (line, bar, histograms and scatter plots) for visualization of given set of data for basic exploratory data analysis.

Use Pandas Plot & pyplot of Matplotlib

Show the output in a new page by drawing or by taking printouts.

Experiment: NoSQL mongoDB

Experiment-7: Mention the steps with your own sample data to Perform CRUD Operations in MongoDB.

Experiment-8: Using mongoDBCompass import an existing CSV file content to a mongoDB collection.

Show the handwritten step-by-step outputs in each of the above cases.

Experiment: Local File System

Experiment-9: File management Commands to be tested on Linux or Windows:

i) Create/Remove directory.

ii) List files and directories present in a directory.

iii) Create a text file in one folder, store some text in it and then send that to another directory.

iv) Show the content of the above file both in command prompt.

v) Delete, Rename, Move files and directories.

Reference

https://drive.google.com/file/d/1SDm0w7fZHjhWC_wn8OEiTOvd0baIRonu/view?usp=sharing

Experiment: Hadoop File System

Experiment-10:

1. Create the following nested directories on HDFS:
    a. /mca/bda/sessional
    b. /mca/da/mkr/asg2
2. Create the following files:
    a. Empty file ‘sessional.txt’ in /mca/bda/sessional directory
    b. Copy an existing file from the local file system to any of the above hdfs directory.
3. Show the content of the above copied file in the console (terminal) and also in a browser.
4. Give a demo on both get and put hdfs commands.

Experiment: HIVE [CRUD]

Experiment-11:

demo.gethue.com **username: demo** **password: demo**

Database: KSCA, Table: Student

Attributes: (roll_number# numeric, student_name string, cgpa int, email string)

Run the following HIVE commands with your own examples:

Create the above in Hive.

Add 10 records to it. (use either LOAD or INSERT command)

display all attributes and all records

display only student_name with a predicate on cgpa

update cgpa on a particular predicate

Drop all (both table and database) in one command

Experiment: PIG_Latin

Experiment-12:

1. Create a local text file (‘product.txt’) to store a sample dataset with the fields (prod_ID, prod_Name, Unit_Price, Quantity)

2. Load the above dataset to a pig table.

3. Demonstrate the following PIG operations on the above dataset:

- dump
- describe
- explain
- illustrate

Reference

https://drive.google.com/file/d/1AhhOdH8qnhUUG0Ye3nW3D4IjQzduN8Bx/view?usp=sharing

Thank You