Bigdata Lab SEE Guidelines

- One program out of the five lab programs for execution.
- Evaluation Criteria (Write-up – 15, Program Execution – 25, VIVA-10).
- If the student opts for change of program in the exam, 10 Marks will be reduced for Write-up.
- In the write-up, the student has to write the Queries/Commands/Core logic(JAVA) based on the given program.

1. Mongo DB – No SQL

a. Create database **CRICKET** and Make Collection With name **TeamIndia** and execute following queries

b. Insert the following Records Into **TeamIndia** Collection

no	name	salary	role
1	ST	2000	BAT
2	MSD	1500	WK
3	YS	1000	ALR
4	RD	1000	BAT
5	RS	500	BAT
6	BK	500	BAT
7	VK	300	BWL
8	JB	400	BWL
9	HP	400	ALR
10	VS	300	BAT

c. Display Data in Proper Format

d. Update Salary Of Player where Name is "VK" by +6000

e. Update Salary Of All Players by giving an increment of +4000 each

f. Update role of "MSD" as "C and WK"

g. Add a New Field remark to document with name "RS" set Remark as WC

h. Find Document From the TeamIndia collection where name begins with S

i. Find Document From the TeamIndia collection where name ends with K

j. Display Documents where in TeamIndia collection Field have BAT, ALR

k. Display Documents where in TeamIndia collection Field not have BAT, ALR

2. HDFS Commands

Do the following activities using suitable HDFS command

a. Listing files and directories.

b. Creating directory

c. Create an empty file

d. Copy files/directories from local file system to hdfs

e. Print the content of the file

f. Copy files/directories from hdfs store to local file system

g. Move file from local to hdfs

h. Copy files within hdfs

i. Move files within hdfs

j. Delete a file from hdfs recursively

k. Size of each file in directory

l. Total size of directory/file

m. Last modified time of directory or path

n. Append the content to file

3. Map Reduce (Programs)
Use the Hadoop framework to write a custom MapReduce program to perform word count operation on a custom data set of your choice.

4. Map Reduce (Programs)
Use the Hadoop framework to write a MapReduce program to read a .csv file into a single node Hadoop cluster containing following fields:

sl_no	card _name	user_name	amount_withdrawn

Implement the following:

1. Count the Number of transactions done by each user
2. Find the total amount of money transacted by each user

5. Hive SQL

a. Create the following tables

Bank(Bank_id:integer, Bname: string, blocation:string)

Customer(Cust_id: integer, Cname: string, income:float, accid:int dob:date)

Account(accid int, cust_id: int,bnk_id: int);

Write the HiveQL queries for the above database schema

b. Insert 5 records using INSERT command.

c. Demonstrate the Alter command for the above tables

d. Rename the table Account to Accounts.

e. Rename the column name “blocation” to “location”.

f. Retrieve all the customers who have account in AXIS bank

g. Retrieve the customer with maximum income and minimum income

h. Retrieve all the customers who have account

i. Retrieve all the customers whose income is not more than 20000

j. Retrieve all the customers whose dob not between 10-10-2010 and 1-2-2012 group based on cust_id and sort the results in descending order

k. Retrieve view which contains all the fileds from account and cname in it. Demonstrate update operation