1. Load batting.csv into a mysql in a database battingdb and table batting

2. Sqoop the details into hdfs.

3. Sqoop the details into hive.

4. Implement a PIG script to

a) Find the total count of participation of G 112

b) Find the player details with "david"

c) Find the average count of "NL"

d) Find the count of teams

5. Implement a Hive script to

a) Find the total count of player details with "david"

b) Create a patition on the TEAMID

c) Create 3 buckets on the partition.

d) Extract the details on player "aaronha01"

e) Find the count of teams

6. Using python,

a) Write a map-reducer program to find the total count of the players

b) Write a map-reducer program to find the total number of the teams.

7. Visualize

Visualize the battings.csv based on the frequency of palyer inclusion yearwise.

8. From halloffame.csv

List the managers.

Find the numbers of votes got year wise by "chancfr01h".

Count the votes got by each person overall.

9. Using hive,partition by year. Then, find the year wise count of participants,

find the total votes got by the players.

10. Using python, map-reducer, find the average votes on the year 1936.