-----------ROUND 1---------------------------

theory questions:

1.tell about yourself

2.rate your skills

3.explain the projects you have done recently.

4.what steps you follow to analyse a large dataset of more than 10gb

5.what is parquet file format

6.what data we require to analyse the post surger risk of patients

7.what is hdfs,hbase,hive

8.what difference between hadoop and spark

9.what is difference between hive internal and external tables

10.what are recent business usecases you solved

coding questions:

unix:

1.how find whether a particular process is running

2.how to fetch the last 10 lines from a log file

3.how you will run a process continously

python:

1.find the maximum and second maximum element from list

2.find whether given two strings are anograms or not

sql:

1.given a login entries table, asked to find the users with last recent login time.

pyspark dataframes:

1.given dataframe with columns name,age,id...Asked to change name column to upper case and new column age category(teenager,adult,senior citizen) and sort the datframe in the ascending order of id.

-------------ROUND 2---------------------------

1.find employees taking salary more than their manager

|id|name|salary|managerId|

|1|Joe|70000|3|

|2|Henry|80000|4|

|3|Sam|60000|NULL|

|4|Max|90000|0|

|5|Shaen|70000|7|

|6|Mary|80000|8|

|7|Rude|20000|4|

|8|Paul|72000|1|

2.write a program to get this output

input=5

output=[0,1,1,2,3]

Explain the complete approach.

3.write sql query to get this output

FROM,TO,DIST

SEA,SF,300

CHI,SEA,2000

SF,SEA,300

SEA,CHI,2000

SEA,LND,500

LND,SEA,500

LND,CHI,1000

CHI,NDL,180

------------------------------