Please find assignment details below:

DB: sandeepv\_test

**1. Using Hadoop command move all those employees data into HDFS directory "/user/your\_user\_name/employees\_data" directory**

vi Consultentdata-1.txt

vi Consultentdata-943.txt

hdfs dfs -mkdir /user/vasistasandeep90\_gmail/employees\_data

hdfs dfs -put Consultentdata-1.txt/user/vasistasandeep90\_gmail/employees\_data

hdfs dfs -put Consultentdata-943.txt/user/vasistasandeep90\_gmail/employees\_data

**2. Create an external Hive table "employees\_Table" representing this "employees\_data". This table will have 5 fields id,age,gender,role and salary.**

CREATE external table employees\_data(id INT, age INT, gender String, role String, salary float) ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde' WITH SERDEPROPERTIES ( "separatorChar" = ",", "quoteChar" = "'","escapeChar" = "//") STORED AS TEXTFILE;

LOAD DATA INPATH '/user/vasistasandeep90\_gmail/employees\_data/Consultentdata-1.txt' INTO table sandeepv\_test.employees\_data;

insert overwrite table employees\_data select distinct \* from employees\_data;

select \* from employees\_data;

*Note: Table created is employees\_data.*

**3. create a new bucketed table "Consultant\_Table\_Bucket" having 4 buckets on the field salary. This table should store the data into columnar format ORC**

create table consultant\_table\_bucket(id int,age int,gender string,role string,salary float) clustered by (salary) into 4 buckets row format delimited fields terminated by ',' stored as ORCfile TBLPROPERTIES('transactional'='true');

show tables;

**4. Insert all those employees whose salary is greater than 5000 into bucketed table "Consultant\_Table\_Bucket". While inserting into "Consultant\_Table\_Bucket" table you need to convert "consultant" role into "BigData Consultant" role.**

insert into table consultant\_table\_bucket select \* from employees\_data where salary > '5000';

update consultant\_table\_bucket SET role = 'BigData Consultant' where role = 'consultant';

**5.Write a Hive query to find out Max, min salary of "BigData Consultant" from the "Consultant\_Table\_Bucket" table.**

select max(salary) from consultant\_table\_bucket where role = 'BigData Consultant';

select min(salary) from consultant\_table\_bucket where role = 'BigData Consultant';