HIVE HEALTHCARE ANALYTICS PROJECT

SQOOP IMPORT COMMANDS FOR ALL QUESTIONS:

```
sqoop import-all-tables \
--connect jdbc:mysql://localhost:3306/healthcare \
--username sqoop \
--password sqoop \
--hive-import \
--m 1
```

Now all problem statements and their solutions:

Problem1: *The healthcare department suspects that some pharmacies prescribe more medicines than others in a single prescription, for them, generate a report that finds for each pharmacy the maximum, minimum and average number of medicines prescribed in their prescriptions.*

```
CREATE EXTERNAL TABLE pharmacy_quantity (
    pharmacyid INT,
    pharmacyname STRING,
    min_quantity INT,
    max_quantity INT,
    avg_quantity DOUBLE
)
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ','
LOCATION '/user/training/hive_1/';
-------------------------------------------------


CREATE EXTERNAL TABLE pharmacy_quantity (
    pharmacyid INT,
    pharmacyname STRING,
    min_quantity INT,
    max_quantity INT,
    avg_quantity DOUBLE
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LOCATION '/user/training/hive_1/';
INSERT OVERWRITE TABLE pharmacy_quantity
SELECT
    a.pharmacyid,
    a.pharmacyname,
    MIN(a.quantity),
    MAX(a.quantity),
    AVG(a.quantity)
FROM (
```

```
    SELECT
        ps.pharmacyid,
        p.pharmacyname,
        c.prescriptionid,
        SUM(c.quantity) AS quantity
    FROM
        pharmacy p
        JOIN prescription ps ON p.pharmacyid = ps.pharmacyid
        JOIN contain c ON c.prescriptionid = ps.prescriptionid
    GROUP BY
        ps.pharmacyid,p.pharmacyname,c.prescriptionid
)as a
GROUP BY
    a.pharmacyid,
    a.pharmacyname;
---------------------------------------
sqoop export \
--connect jdbc:mysql://localhost:3306/healthcare \
--username sqoop \
--password sqoop \
--table pharmacy_quantity \
--export-dir /user/training/hive_1/000000_0 \
--input-fields-terminated-by ',';
```

---

problem2:The State of Alabama (AL) is trying to manage its healthcare resources more efficiently. For each city in their state, they need to identify the disease for which the maximum number of patients have gone for treatment. Assist the state for this purpose.

Note: The state of Alabama is represented as AL in Address Table.

```
CREATE EXTERNAL TABLE treatment_data (
    city string,
    diseasename string,
    patientcount int
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1/';
```

```
CREATE EXTERNAL TABLE treatment_data (
    city string,
    diseasename string,
    patientcount int
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1/';
```

```
INSERT OVERWRITE TABLE pharmacy_quantity
SELECT
    a.pharmacyid,
    a.pharmacyname,
    MIN(a.quantity),
    MAX(a.quantity),
    AVG(a.quantity)
FROM (
    SELECT
        ps.pharmacyid,
        p.pharmacyname,
        c.prescriptionid,
        SUM(c.quantity) AS quantity
    FROM
        pharmacy p
        JOIN prescription ps ON p.pharmacyid = ps.pharmacyid
        JOIN contain c ON c.prescriptionid = ps.prescriptionid
    GROUP BY
        ps.pharmacyid,p.pharmacyname,c.prescriptionid
)as a
GROUP BY
    a.pharmacyid,
    a.pharmacyname;
```

```
------------------------------

sqoop export \

--connect jdbc:mysql://localhost:3306/healthcare \

--username sqoop \

--password sqoop \

--table treatment_data \

--export-dir /user/training/hive_1/000000_0 \

--input-fields-terminated-by ',';

-------------------------------------------
```

problem3:The healthcare department needs a report about insurance plans. The report is required to include the insurance plan, which was claimed the most and least for each disease. Assist to create such a report.

```
CREATE EXTERNAL TABLE plan_disease_counts (

    diseasename string,

    planname1 string,

    planname2 string

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/training/hive_1/';

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

CREATE EXTERNAL TABLE plan_disease_counts (

    diseasename string,
```

```
    planname1 string,
    planname2 string
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1';
INSERT INTO TABLE plan_disease_counts
SELECT A.DISEASENAME,A.PLANNAME,B.PLANNAME FROM (SELECT Y.P
PLANNAME,Y.D DISEASENAME FROM (SELECT X.P,X.D,ROW_NUMBER()
OVER(partition by X.D ORDER BY X.CNT DESC) RN FROM
(SELECT I.PLANNAME P,D.DISEASENAME D,COUNT(C.CLAIMID) CNT FROM
claim C,disease D,treatment T,insuranceplan I WHERE
I.UIN=C.UIN AND C.CLAIMID=T.CLAIMID AND T.DISEASEID=D.DISEASEID
GROUP BY I.PLANNAME,D.DISEASENAME ORDER BY 3)X)Y WHERE
Y.RN=1)A
JOIN
(SELECT Y.P PLANNAME,Y.D DISEASENAME FROM (SELECT
X.P,X.D,ROW_NUMBER() OVER(partition by X.D ORDER BY X.CNT) RN
FROM
(SELECT I.PLANNAME P,D.DISEASENAME D,COUNT(C.CLAIMID) CNT FROM
claim C,disease D,treatment T,insuranceplan I WHERE
I.UIN=C.UIN AND C.CLAIMID=T.CLAIMID AND T.DISEASEID=D.DISEASEID
GROUP BY I.PLANNAME,D.DISEASENAME ORDER BY 3)X)Y WHERE
Y.RN=1)B ON A.DISEASENAME=B.DISEASENAME;
```

```
-------------------------------------------
sqoop export \
--connect jdbc:mysql://localhost:3306/healthcare \
--username sqoop \
--password sqoop \
--table plan_disease_counts \
--export-dir /user/training/hive_1/000000_0 \
--input-fields-terminated-by ',';
-------------------------------------------------
```

problem4: Insurance companies want to assess the performance of their insurance plans. Generate a report that shows each insurance plan, the company that issues the plan, and the number of treatments the plan was claimed for.

```
CREATE EXTERNAL TABLE insurance_data (
   companyname string,
   planname string,
   claimcount int
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1';
-----------------------------------------------
CREATE EXTERNAL TABLE insurance_data (
   companyname string,
   planname string,
   claimcount int
)
```

```
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1';
INSERT OVERWRITE TABLE insurance_data
SELECT I.COMPANYNAME, IP.PLANNAME, COUNT(T.CLAIMID)
FROM INSURANCECOMPANY I
JOIN INSURANCEPLAN IP ON I.COMPANYID=IP.COMPANYID
JOIN CLAIM C ON IP.UIN=C.UIN
JOIN TREATMENT T ON C.CLAIMID=T.CLAIMID
GROUP BY IP.PLANNAME, I.COMPANYNAME
order by desc 3;
----------
sqoop export \
--connect jdbc:mysql://localhost:3306/healthcare \
--username sqoop \
--password sqoop \
--table insurance_data  \
--export-dir /user/training/hive_1/000000_0 \
--input-fields-terminated-by ',';
-----------------------------------------
```

problem5:Jhonny, from the finance department of Arizona(AZ), has requested a report that lists the total quantity of medicine each pharmacy in his state has prescribed that falls under **Tax criteria I** for treatments that took place in 2021. Assist Jhonny in generating the report.

```
CREATE EXTERNAL TABLE pharmacy_data (
   state string,
   pharmacyid int,
   quantity int
```

```
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1';

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

CREATE EXTERNAL TABLE pharmacy_data (
   state string,
   pharmacyid int,
   quantity int
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1';

INSERT OVERWRITE TABLE pharmacy_data
SELECT X.STATE STATE,X.PHARMACYID PHARMACYID ,Y.QNT QUANTITY FROM (SELECT A.STATE
STATE,P.PHARMACYID PHARMACYID FROM PHARMACY P,ADDRESS A WHERE
A.ADDRESSID=P.ADDRESSID)X,
(SELECT P.PHARMACYID PHARMACYID,SUM(C.QUANTITY) QNT
FROM  PRESCRIPTION PR,CONTAIN C,MEDICINE M,TREATMENT T, PHARMACY P
WHERE PR.PRESCRIPTIONID=C.PRESCRIPTIONID AND C.MEDICINEID=M.MEDICINEID
AND T.TREATMENTID=PR.TREATMENTID AND P.PHARMACYID=PR.PHARMACYID
AND YEAR(T.DATE)=2021 AND M.TAXCRITERIA="I" GROUP BY P.PHARMACYID ORDER BY 1)Y
WHERE X.STATE="AZ" AND X.PHARMACYID=Y.PHARMACYID ORDER BY 3;

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

sqoop export \
--connect jdbc:mysql://localhost:3306/healthcare \
--username sqoop \
```

--password sqoop \

--table pharmacy_data \

--export-dir /user/training/hive_1/000000_0 \

--input-fields-terminated-by ',';

---

problem6: Jacob, from insurance management, has noticed that insurance claims are not made for all the treatments. He also wants to figure out if the gender of the patient has any impact on the insurance claim. Assist Jacob in this situation by generating a report that finds for each gender the number of treatments, number of claims, and treatment-to-claim ratio. And notice if there is a significant difference between the treatment-to-claim ratio of male and female patients.

```
CREATE EXTERNAL TABLE treatment_summary (
  gender varchar(10),
  no_of_treatments INT,
  no_of_claims INT,
  treatment_to_claim_ratio FLOAT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LOCATION '/user/training/hive_1';
```

```
CREATE EXTERNAL TABLE treatment_summary (
  gender STRING,
```

```
  no_of_treatments INT,
  no_of_claims INT,
  treatment_to_claim_ratio FLOAT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LOCATION '/user/training/hive_1';
INSERT OVERWRITE TABLE treatment_summary
SELECT P.GENDER,
       COUNT(T.TREATMENTID) AS no_of_treatments,
       COUNT(T.CLAIMID) AS no_of_claims,
       COUNT(T.TREATMENTID) / COUNT(T.CLAIMID) AS
treatment_to_claim_ratio
FROM TREATMENT T
INNER JOIN PERSON P ON T.PATIENTID = P.PERSONID
GROUP BY P.GENDER;
---------------------------------------

sqoop export \
--connect jdbc:mysql://localhost:3306/healthcare \
--username sqoop \
--password sqoop \
--table treatment_summary \
--export-dir /user/training/hive_1/000000_0 \
```

--input-fields-terminated-by ',';

used for all queries to export results in mysql.

---

problem7:The healthcare department wants a pharmacy report on the percentage of hospital-exclusive medicine prescribed in the year 2022.

Assist the healthcare department to view for each pharmacy, the pharmacy id, pharmacy name, total quantity of medicine prescribed in 2022, total quantity of hospital-exclusive medicine prescribed by the pharmacy in 2022, and the percentage of hospital-exclusive medicine to the total medicine prescribed in 2022.

Order the result in descending order of the percentage found.

```
CREATE EXTERNAL TABLE IF NOT EXISTS pharmacy_summary (
  pharmacyid INT,
  pharmacyname varchar(10),
  MedicineQuantity INT,
  HospitalExclusiveMedQuantity INT,
  `Percentage of Hospital exclusive to total medicine` FLOAT
)
PARTITIONED BY (year INT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
LOCATION '/user/training/hive_1';
```

```
CREATE EXTERNAL TABLE IF NOT EXISTS pharmacy_summary (
  pharmacyid INT,
  pharmacyname varchar(10),
  MedicineQuantity INT,
  HospitalExclusiveMedQuantity INT,
  `Percentage of Hospital exclusive to total medicine` FLOAT
)
PARTITIONED BY (year INT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
LOCATION '/user/training/hive_1';
INSERT OVERWRITE TABLE pharmacy_summary PARTITION (year=2022)
SELECT DISTINCT y.pharmacyid, y.pharmacyname, z.cnt2 AS HospitalExclusiveMedQuantity, y.cnt1 AS MedicineQuantity,
  (z.cnt2/y.cnt1)*100 AS `Percentage of Hospital exclusive to total medicine`
FROM
(SELECT p.pharmacyid ph, p.pharmacyname pn, COUNT(c.medicineid) cnt1
 FROM pharmacy p, prescription pr, contain c, medicine m, treatment t
 WHERE p.pharmacyid=pr.pharmacyid AND pr.prescriptionid=c.prescriptionid AND c.medicineid=m.medicineid
 AND t.treatmentid=pr.treatmentid AND YEAR(t.date)=2022
 GROUP BY p.pharmacyid, p.pharmacyname
 ORDER BY 1) y
```

```
JOIN
(SELECT p.pharmacyid ph, p.pharmacyname pn, COUNT(c.medicineid) cnt2
 FROM pharmacy p, prescription pr, contain c, medicine m, treatment t
 WHERE p.pharmacyid=pr.pharmacyid AND
pr.prescriptionid=c.prescriptionid AND c.medicineid=m.medicineid
 AND t.treatmentid=pr.treatmentid AND YEAR(t.date)=2022 AND
m.hospitalexclusive='S'
 GROUP BY p.pharmacyid, p.pharmacyname
 ORDER BY 1) z
ON y.pharmacyid=z.pharmacyid;


ALTER TABLE pharmacy_summary ADD PARTITION (year=2022)
LOCATION 'your_external_table_location/year=2022';
```

```
sqoop export \
--connect jdbc:mysql://localhost:3306/healthcare \
--username sqoop \
--password sqoop \
--table pharmacy_summary \
--export-dir /user/training/hive_1/000000_0 \
--input-fields-terminated-by ',';
```

---

problem8:The healthcare department wants to categorize the patients into the following category.

YoungMale: Born on or after 1st Jan 2005 and gender male.

YoungFemale: Born on or after 1st Jan 2005 and gender female.

AdultMale: Born before 1st Jan 2005 but on or after 1st Jan 1985 and gender male.

AdultFemale: Born before 1st Jan 2005 but on or after 1st Jan 1985 and gender female.

MidAgeMale: Born before 1st Jan 1985 but on or after 1st Jan 1970 and gender male.

MidAgeFemale: Born before 1st Jan 1985 but on or after 1st Jan 1970 and gender female.

ElderMale: Born before 1st Jan 1970, and gender male.

ElderFemale: Born before 1st Jan 1970, and gender female.

Write a SQL query to list all the patient name, gender, dob, and their category.

```
CREATE EXTERNAL TABLE patient_category (
  PatientName varchar(10),
  Gender varchar(10),
  DOB varchar(10),
  Category varchar(10)
)
PARTITIONED BY (year varchar(10), month varchar(10), day varchar(10))
LOCATION '/user/training/hive_1';
```

```
CREATE EXTERNAL TABLE patient_category (
```

```
  PatientName varchar(10),
  Gender varchar(10),
  DOB varchar(10),
  Category varchar(10)
)
PARTITIONED BY (year varchar(10), month varchar(10), day varchar(10))
LOCATION '/user/training/hive_1';
INSERT OVERWRITE TABLE patient_category
PARTITION (year, month, day)
SELECT p.personName AS PatientName,
      p.gender AS Gender,
      pt.dob AS DOB,
      (CASE
         WHEN pt.dob >= "2005-01-01" AND p.gender = "male" THEN
"YoungMale"
         WHEN pt.dob >= "2005-01-01" AND p.gender = "female" THEN
"YoungFemale"
         WHEN pt.dob > "1985-01-01" AND pt.dob < "2005-01-01" AND
p.gender = "male" THEN "AdultMale"
         WHEN pt.dob > "1985-01-01" AND pt.dob < "2005-01-01" AND
p.gender = "female" THEN "AdultFemale"
         WHEN pt.dob > "1970-01-01" AND pt.dob < "1985-01-01" AND
p.gender = "male" THEN "MidAgeMale"
         WHEN pt.dob > "1970-01-01" AND pt.dob < "1985-01-01" AND
p.gender = "female" THEN "MidAgeFemale"
```

WHEN pt.dob < "1970-01-01" AND p.gender = "male" THEN "ElderMale"

ELSE "ElderFemale"

END) AS Category,

YEAR(pt.dob) as year, MONTH(pt.dob) as month, DAY(pt.dob) as day

FROM patient pt

JOIN person p ON pt.patientid = p.personid;

**sqoop export **

**--connect jdbc:mysql://localhost:3306/healthcare **

**--username sqoop **

**--password sqoop **

--table patient_category ****

**--export-dir /user/training/hive_1/000000_0 **

--input-fields-terminated-by ',';

used for all queries to export results in mysql.

---

problem9:Brooke is trying to figure out if patients with a particular disease are preferring some pharmacies over others or not, For this purpose, she has requested a detailed pharmacy report that shows each pharmacy name, and how many prescriptions they have prescribed for each disease in 2021 and 2022, She expects the number of prescriptions prescribed in 2021 and 2022 be displayed in two separate columns.

Write a query for Brooke's requirement.

CREATE EXTERNAL TABLE IF NOT EXISTS disease_counts (

```
  CITY STRING,
  DISEASENAME STRING,
  DISEASECOUNT INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1';
-----------------------------------------
CREATE EXTERNAL TABLE IF NOT EXISTS disease_counts (
  CITY STRING,
  DISEASENAME STRING,
  DISEASECOUNT INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/training/hive_1';
INSERT OVERWRITE TABLE disease_counts
SELECT Y.C CITY,Y.D DISEASENAME,Y.CNT DISEASECOUNT FROM
(SELECT X.CITY C,X.DISEASENAME D,X.COUNT CNT,row_number()
OVER(PARTITION BY X.DISEASENAME ORDER BY COUNT DESC) RN FROM
```

(SELECT A.CITY CITY,D.DISEASENAME
DISEASENAME,COUNT(PR.TREATMENTID) COUNT FROM prescriptions PR
JOIN TREATMENT T ON PR.TREATMENTID=T.TREATMENTID JOIN
DISEASE D ON T.DISEASEID=D.DISEASEID JOIN PERSON P ON
T.PATIENTID=P.PERSONID JOIN ADDRESS A ON
P.ADDRESSID=A.ADDRESSID

group by 1,2)X)Y
WHERE Y.RN<4;

**sqoop export **

**--connect jdbc:mysql://localhost:3306/healthcare **

**--username sqoop **

**--password sqoop **

--table disease_counts****

**--export-dir /user/training/hive_1/000000_0 **

--input-fields-terminated-by ',';

---

problem10:Jhonny, from the finance department of Arizona(AZ), has

requested a report that lists the total quantity of medicine each pharmacy in his

state has prescribed that falls under Tax criteria I for treatments that took

place in 2021. Assist Jhonny in generating the report.

***/**

CREATE TABLE treatment_part_clus(

```
    treatmentID BIGINT,
    patientID BIGINT,
    date STRING,
    claimID BIGINT
)
COMMENT 'Treatment to claim ratio'
PARTITIONED BY (diseaseID STRING) CLUSTERED BY(patientID) INTO 2 BUCKETS
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';

-- insert the data from treatment table
insert overwrite table treatment_part_clus partition(diseaseID)
select treatmentID, patientID, date, claimID, diseaseID from treatment;

--
create external table pharmacy_quantity_2021(
    pharmacyID BIGINT,
    pharmacyName STRING,
    quantity BIGINT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LOCATION '/user/training/hive_9';
```

```
insert overwrite table pharmacy_quantity_2021
select p.`pharmacyID`, p.pharmacyName, sum(k.quantity) `Quantity`
from pharmacy p
inner join (select `addressID` from address_part
            where state='AZ') b
            on b.`addressID` = p.`addressID`
inner join keep k on k.`pharmacyID` = p.`pharmacyID`
inner join medicine m on m.`medicineID`=k.`medicineID`
inner join prescription pr on pr.`pharmacyID` = p.`pharmacyID`
inner join (select `treatmentID` from treatment_part
            where year(date)='2021') a
            on a.`treatmentID`=pr.`treatmentID`
where m.`taxCriteria` = 'I'
group by p.`pharmacyName`, p.`pharmacyID`;


create table pharmacy_quantity_2021(
    pharmacyID BIGINT,
    pharmacyName varchar(40),
    quantity BIGINT
);
```

```
sqoop export \
--connect jdbc:mysql://localhost:3306/healthcare_out \
--username sqoop \
--password sqoop \
--table pharmacy_quantity_2021 \
--export-dir /user/training/hive_9/000000_0 \
--input-fields-terminated-by ',';
```