Apache Hive: Alter, Views, Joins and Subqueries

Alter:

In Hive, the **ALTER** statement is used to make changes to existing database objects like tables and databases.

Alter Table: You can use ALTER TABLE to modify the structure of an existing table. This includes adding, dropping, or renaming columns, changing table properties, or renaming the table itself.

Task 1: Rename all the column names of u.data table in Hive using Alter command. Also rename the table name itself.

-- Rename a column

ALTER TABLE table_name CHANGE COLUMN old_column_name new_column_name data_type;

-- Rename the table

ALTER TABLE old_table_name RENAME TO new_table_name;

Views:

In Hive, a view is a virtual table that represents the result of a pre-defined query. Views allow you to encapsulate complex SQL queries into a named object, which can then be queried like a regular table. Views provide several benefits, including

1. Simplifying complex queries,
2. Abstracting the underlying data model, and
3. Controlling access to sensitive data by restricting the columns visible to different users.

Views do not store data themselves; they only represent the result of a query on the underlying tables. Therefore, any changes made to the underlying tables are reflected in the view.

Hive views do not support the INSERT, UPDATE, or DELETE operations. They are read-only.

Task 2: Create a View on u.data table. The view should contain grouping on each user, the number of films rated by the user and the average rating given by the user.

```
CREATE VIEW user_view AS

SELECT user_id, count(film_id) AS no_of_films, avg(rating) as avg_rating

FROM user_ratings

GROUP BY user_id;
```

```
select * from user_view;
```

Joins:

In Hive, joins are used to combine rows from two or more tables based on a related column between them. Hive supports various types of joins, including inner joins, outer joins (left, right, and full), and cross joins.

Inner Join:

An inner join returns rows when there is a match in both tables based on the join condition.

Left Outer Join:

A left outer join returns all rows from the left table and matching rows from the right table. If there is no match, NULL values are returned for the columns from the right table.

Right Outer Join:

A right outer join returns all rows from the right table and matching rows from the left table. If there is no match, NULL values are returned for the columns from the left table.

Task 3: Join the user_view with user_info table and get the combined results.

```
select uv.user_id, ui.age, ui.occupation, uv.no_of_films, uv.avg_rating from user_view as uv

inner join user_info ui

on uv.user_id = ui.user_id;
```

Subquery

In Hive, subqueries allow you to nest one query (the inner query) inside another query (the outer query). Subqueries can be used in various parts of a SQL statement, such as the SELECT, FROM, WHERE, and HAVING clauses.

Task 4: Get the total number of films rated by the occupation of the user using subquery.

ud.occupation	no_of_films
administrator	7479
artist	2308
doctor	540
educator	9442
engineer	8175
entertainment	2095
executive	3403
healthcare	2804
homemaker	299
lawyer	1345
librarian	5273
marketing	1950
none	901
other	10663
programmer	7801
retired	1609
salesman	856
scientist	2058