



Data Application Lab

Answer of Quiz for Hive

1. **The partition of hive table has been modified to point to a new directory location. Do I have to move the data to the new location or the data will be moved automatically to the new location?**

Solution: Changing the point of partition will not move the data to the new location. It has to be moved manually to the new location from the old one if you want to move it.

2. **How will you do the data transfer from HDFS to Hive?**

Solution: If data is already present in HDFS then the user doesn't need LOAD DATA that moves the files to the /user/hive/warehouse/. The user just has to define the table using the keyword external that creates the table definition in the hive metastore.

3. **Explain about the different types of join in Hive. (JOIN, FULLY OUTER JOIN, LEFT OUTER JOIN, RIGHT OUTER JOIN)**

Solution: Hive has 4 main types of joins.

JOIN- Similar to Outer Join in SQL

FULL OUTER JOIN – Combines the records of both the left and right outer tables that fulfill the join condition.

LEFT OUTER JOIN- All the rows from the left table are returned even if there are no matches in the right table.

RIGHT OUTER JOIN-All the rows from the right table are returned even if there are no matches in the left table.

Here is the SQL join explanation for you can have a better understanding on their differences:

Inner join

Inner join return rows when there is at least one match of rows between the tables.

Right Join

Right join return rows which are common between the tables and all rows of Right hand side table. Simply, it returns all the rows from the right-hand side table even though there are no matches in the left-hand side table.

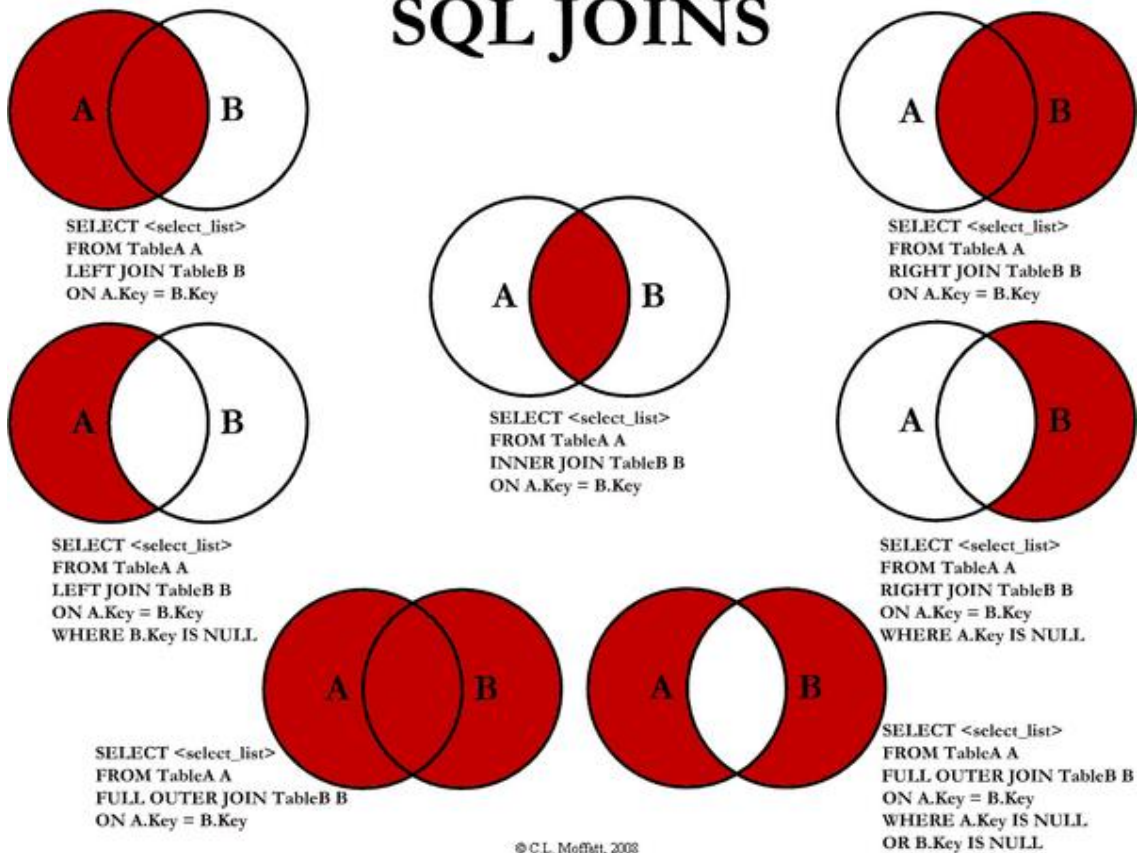
Left Join

Left join return rows which are common between the tables and all rows of Left-hand side table. Simply, it returns all the rows from Left-hand side table even though there are no matches in the Right-hand side table.

Full Join

Full join return rows when there are matching rows in any one of the tables. This means, it returns all the rows from the left-hand side table and all the rows from the right hand side table.

SQL JOINS



4. What is the difference between an external table and a managed table?

Solution: The main difference is that when you drop an external table, the underlying data files stay intact. This is because the user is expected to manage the data files and directories. With a managed table, the underlying directories and data get wiped out when the table is dropped.

5. The Answer of Hive Homework will be published separately.