

Apache Hive Joins

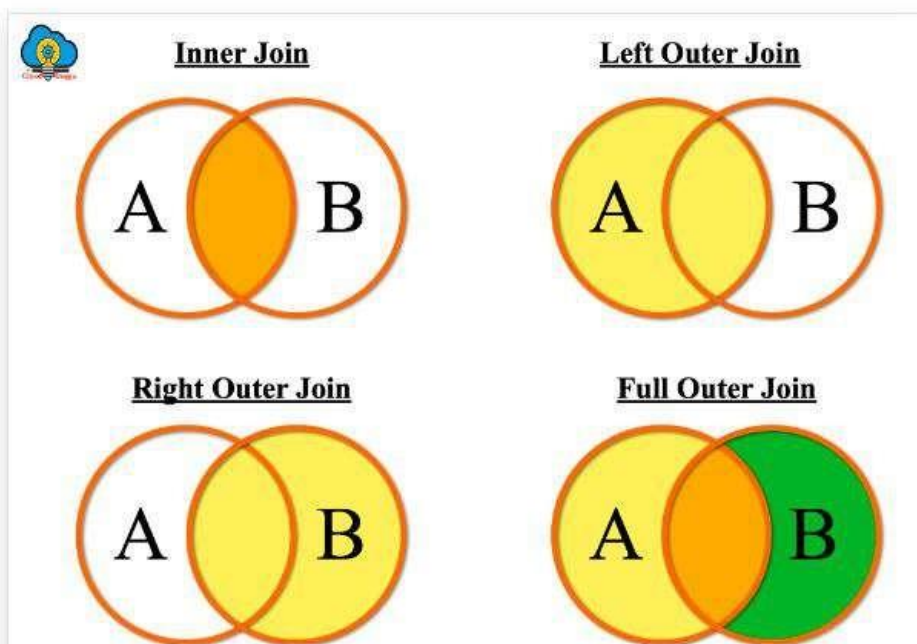


Apache Hive JOINS are used to combine columns from one (self-join) or more tables by using values common to each. Using join we can fetch corresponding records from two or more tables. It is almost similar to SQL joins.

Apache Hive provides four types of joins which are mentioned below.

1. Inner Join
2. Left Outer Join
3. Right Outer Join
4. Full Outer Join

The following graph is the representation of Apache Hive Joins using table A and table B.



Hive Join Syntax

join_table:

```
table_reference [INNER] JOIN table_factor [join_condition]
| table_reference {LEFT|RIGHT|FULL} [OUTER] JOIN table_reference join_condition
| table_reference LEFT SEMI JOIN table_reference join_condition
| table_reference CROSS JOIN table_reference [join_condition] (as of Hive 0.10)
```

table_reference:

```
table_factor
| join_table
```

table_factor:

```
tbl_name [alias]
| table_subquery alias
| ( table_references )
```

join_condition:

```
ON expression
```

Hive Join Example

For this example, we have created two tables, the first table name is “cloudduggudb.employees_data” and the second table name is “cloudduggudb.department_data” and their create statements and insert statements are mentioned below.

Create Table Statement:

Create a statement for table “cloudduggudb.employees_data”.

```
create table cloudduggudb.employees_data(emp_id INT,emp_name STRING,job_name STRING, manager_id INT, salary INT, dep_id INT );
```

Insert Table Statement:

Insert statement for table “cloudduggudb.employees_data”.

```
INSERT INTO cloudduggudb.employees_data VALUES(68319, 'KAYLING', 'PRESIDENT', 68319, 6000, 1001);
INSERT INTO cloudduggudb.employees_data VALUES(66928, 'BLAZE', 'MANAGER', 68319, 2750, 3001);
INSERT INTO cloudduggudb.employees_data VALUES(67832, 'CLARE', 'MANAGER', 68319, 2957, 1001);
INSERT INTO cloudduggudb.employees_data VALUES(65646, 'JONAS', 'MANAGER', 65646, 1600, 2001);
INSERT INTO cloudduggudb.employees_data VALUES(67858, 'SCARLET', 'ANALYST', 65646, 1600, 2001);
INSERT INTO cloudduggudb.employees_data VALUES(69062, 'FRANK', 'ANALYST', 66928, 6000, 2001);
INSERT INTO cloudduggudb.employees_data VALUES(63679, 'SANDRINE', 'CLERK', 66928, 1400, 2001);
```

Create Table Statement:

Create a statement for table “cloudduggudb.department_data”.

```
create table cloudduggudb.department_data(dep_id INT, dep_name STRING, dep_location STRING);
```

Insert Table Statement:

Insert statement for table “cloudduggudb.department_data”.

```
INSERT INTO cloudduggudb.department_data VALUES(1001, 'ACCOUNTING', 'NEW YORK');
```

```
INSERT INTO cloudduggudb.department_data VALUES(3001, 'RESEARCH', 'DALLAS');
```

```
INSERT INTO cloudduggudb.department_data VALUES(2001, 'SALES', 'CHICAGO');
```

