Join multiple tables in Hive

Asked 3 years, 7 months ago   Active 3 years, 6 months ago   Viewed 6k times

Below is the data set

```
Table1
col1,col2
key1,k1
key2,k2
key3,k3


Table2
col1,col3
key1,k11
key2,k22
key4,k44


Table3
col1,col4
key1,k111
key2,k222
key5,k555
```

I need to join the 3 tables based on col1. Below is my query

select a.col1,a.col2,b.col3,c.col4 from table1 a full outer join table2 b full outer join table3 c;

The expected output is shown below

Expected output:

```
col1,col2,col3,col4
key1,k1,k11,k111
key2,k2,k22,k222
key3,k3,   ,
key4,   ,k44,
key5,   ,   ,k555
```
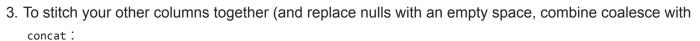
hadoop   hive

Share Follow

asked Aug 17 '17 at 17:58
user4119502

Do you want it to be a cross join? Maybe you mean `table1.col1 = table2.col1 and table1.col1 = table3.col1` ? —
Andrew Aug 17 '17 at 21:28

2 Answers

Active Oldest Votes

▲

3

▼

✓

🕑

1. Based on your desired result, you need to join all your tables on col1.

2. To ensure that you always have a value in col1, you need to `coalesce` them together (coalesce gives you the first non-null value).

3. To stitch your other columns together (and replace nulls with an empty space, combine coalesce with `concat` :

Putting that all together:

```
select
coalesce(t1.col1,t2.col1,t3.col1) as col1,
concat(coalesce(t1.col2,' '),',',coalesce(t2.col2,' '),',',coalesce(t3.col2,' '))
from
table1 t1
full join table2 t2
  on t1.col1 = t2.col1
full join table3 t3
  on t1.col1 = t3.col1
```

Share Follow

answered Aug 17 '17 at 21:51

Andrew
7,296 2 22 41

---

key1 k1,k11,k111 key2 k2,k22,k222 key3 k3, , key4 ,k44, key5 , ,k555 Not getting comma value after first column – user4119502 Aug 17 '17 at 22:36

▲

2

▼

🕑

Since you are doing full outer join, it will be a good idea to do `COALESCE` for column `col1`

Something like:

```
select
    COALESCE(a.col1, b.col1, c.col1) as col1,
    a.col2,
    b.col3,
    c.col4
from
    table1 a
    full outer join table2 b
      on t1.col1 = t2.col1
    full outer join table3 c
      on t1.col1 = t3.col1;
```

We need to have join condition as well to avoid cartesian product in the result set.

Share Follow

edited Aug 19 '17 at 1:27

answered Aug 17 '17 at 18:07

Ambrish
3,417 2 21 37

Above query returned 27 rows.. where as the output expected is 5 rows. – user4119502 Aug 17 '17 at 20:39

that was good. how do i do it, when i have table1.col1 = table2.col1 and table2.col7=table3.col1 – Naveen Kumar Apr 25 '20 at 2:58