Main difference between dynamic and static partitioning in Hive

Asked 5 years, 9 months ago   Active 2 years, 7 months ago   Viewed 66k times

▲

16

▼

What is the main difference between static and dynamic partition in Hive? Using individual insert means static and single insert to partition table means dynamic. Is there any other advantage?

hive

⭐

7

🕑

Share Improve this question Follow

edited Apr 6 '17 at 12:07
MC Emperor
16.9k   13   69   106

asked Jun 18 '15 at 17:03
Ronak
261   1   3   13

4 Answers

Active | Oldest | Votes

▲

23

▼

✅

🕑

Partitioning in Hive is very useful to prune data during query to reduce query times.

Partitions are created when data is inserted into table. Depending on how you load data you would need partitions. Usually when loading files (big files) into Hive tables static partitions are preferred. That saves your time in loading data compared to dynamic partition. You "statically" add a partition in table and move the file into the partition of the table. Since the files are big they are usually generated in HDFS. You can get the partition column value form the filename, day of date etc without reading the whole big file.

Incase of dynamic partition whole big file i.e. every row of the data is read and data is partitioned through a MR job into the destination tables depending on certain field in file. So usually dynamic partition are useful when you are doing sort of a ETL flow in your data pipeline. e.g. you load a huge file through a move command into a Table X. then you run a inert query into a Table Y and partition data based on field in table X say day , country. You may want to further run a ETL step to partition the data in country partition in Table Y into a Table Z where data is partitioned based on cities for a particular country only. etc.

Thus depending on your end table or requirements for data and in what form data is produced at source you may choose static or dynamic partition.

Share Improve this answer Follow

edited Aug 30 '18 at 10:37
gravetii
7,608   6   41   66

answered Jun 18 '15 at 18:06
Urvishsinh Mahida
1,170   10   22

@Urvisinh Do you have concrete example for Static and Dynamic partition – user2102237 Jan 18 '17 at 6:00

▲

28

▼

in static partitioning we need to specify the partition column value in each and every LOAD statement.

suppose we are having partition on column country for table t1(userid, name,occupation, country), so each time we need to provide country value

```
hive>LOAD DATA INPATH '/hdfs path of the file' INTO TABLE t1 PARTITION(country="US")
hive>LOAD DATA INPATH '/hdfs path of the file' INTO TABLE t1 PARTITION(country="UK")
```

dynamic partition allow us not to specify partition column value each time. the approach we follows is as below:

1. create a non-partitioned table t2 and insert data into it.

2. now create a table t1 partitioned on intended column(say country).

3. load data in t1 from t2 as below:

```
hive> INSERT INTO TABLE t2 PARTITION(country) SELECT * from T1;
```

4. make sure that partitioned column is always the last one in non partitioned table(as we are having country column in t2)

Share Improve this answer Follow

edited Jan 18 '16 at 10:37

Tunaki
115k   39   277   368

answered Jan 18 '16 at 10:34

Azam Khan
446   5   11

---

2   why partitioned column must be last in non partitioned table? – Khatri Aug 22 '17 at 14:12

1   @Khatri If not it'll *mess up* the data partitioning. I tired a Sample Employee example as `name, salary(partition), designation` . When loading from non-partitioned table to partitioned table, despite specifying the column as `Salary` it was not honored. The data got split by Designation and the worst part was the partition file structure in HDFS was `/user/hive/warehouse/test.db/employee/salary=Technical/000000_0` – Kishore Bandi Mar 6 '18 at 4:04

Static Partition in Hive

0

Insert input data files individually into a partition table is Static Partition Usually when loading files (big files) into Hive tables static partitions are preferred

Static Partition saves your time in loading data compared to dynamic partition You "statically" add a partition in table and move the file into the partition of the table.

We can alter the partition in static partition

You can get the partition column value form the filename, day of date etc without reading the whole big file. If you want to use Static partition in hive you should set property

```
set hive.mapred.mode = strict
```

This property set by default in hive-site.xml Static partition is in Strict Mode You should use where clause to use limit in static partition You can perform Static partition on Hive Manage table or external table.

Dynamic Partition in Hive

single insert to partition table is known as dynamic partition

Usually dynamic partition load the data from non partitioned table

Dynamic Partition takes more time in loading data compared to static partition

When you have large data stored in a table then Dynamic partition is suitable.

If you want to partition number of column but you don't know how many columns then also dynamic partition is suitable

Dynamic partition there is no required where clause to use limit. we can't perform alter on Dynamic partition

You can perform dynamic partition on hive external table and managed table If you want to use Dynamic partition in hive then mode is in nonstrict mode Here is hive dynamic partition properties you should allow

```
SET hive.exec.dynamic.partition = true;

SET hive.exec.dynamic.partition.mode = nonstrict;
```

Share Improve this answer Follow

Dynamic partitioning in HIVE:

-2

```
CREATE TABLE temps_orc_partition_date
(statecode STRING, countrycode STRING, sitenum STRING, paramcode STRING, poc STRING, latitude STRING,
longitude STRING, datum STRING, param STRING, timelocal STRING, dategmt STRING, timegmt STRING, degrees
double, uom STRING, mdl STRING, uncert STRING, qual STRING, method STRING, methodname STRING, state
STRING, county STRING, dateoflastchange STRING)
PARTITIONED BY (datelocal STRING)
STORED AS ORC;
```

move the "datelocal" column to being last in the SELECT. For dynamic partitioning to work in Hive, this is a requirement.

```
INSERT INTO TABLE temps_orc_partition_date
PARTITION (datelocal)
SELECT statecode, countrycode, sitenum, paramcode, poc, latitude, longitude, datum, param, timelocal,
dategmt, timegmt, degrees, uom, mdl, uncert, qual, method, methodname, state, county, dateoflastchange,
datelocal
FROM temps_txt;
```

Share Improve this answer Follow

clarification to 4th column in second answer – Vamshavardhan Reddy Oct 3 '17 at 3:07