

Hive Partitions & Buckets with Example

Tables, Partitions, and Buckets are the parts of Hive data modeling.

What is Partitions?

Hive Partitions is a way to organizes tables into partitions by dividing tables into different parts based on partition keys.

Partition is helpful when the table has one or more Partition keys. Partition keys are basic

elements for determining how the data is stored in the table.

For Example: -

"Client having Some E –commerce data which belongs to India operations in which each state (38 states) operations mentioned in as a whole. If we take state column as partition key and perform partitions on that India data as a whole, we can able to get Number of partitions (38 partitions) which is equal to number of states (38) present in India. Such that each state data can be viewed separately in partitions tables.

Sample Code Snippet for partitions

1. Creation of Table all states

```
create table all states(state string, District string,Enrolments string)

row format delimited

fields terminated by ',';
```

2. Loading data into created table all states

```
Load data local inpath '/home/hduser/Desktop/AllStates.csv' into table allstates;
```

3. Creation of partition table

```
create table state_part(District string,Enrolments string) PARTITIONED BY(state string);
```

4. For partition we have to set this property

```
set hive.exec.dynamic.partition.mode=nonstrict
```

5. Loading data into partition table

```
INSERT OVERWRITE TABLE state_part PARTITION(state)  
SELECT district,enrolments,state from allstates;
```

6. Actual processing and formation of partition tables based on state as partition key

7. There are going to be 38 partition outputs in HDFS storage with the file name as state name.

We will check this in this step

The following screen shots will show u the execution of above mentioned code



```
hive> create table allstates(state string, District string,Enrolments string)  
1> row format delimited  
  > fields terminated by ',';  
OK
```

creation of table "allstates"

(./images/Hive/120415_1301_Dataoperati7.png).



```
hive> load data local inpath '/home/hduser/AllStates.csv' into table allstates;  
Loading data to table default.allstates  
Table default.allstates stats: [numFiles=1, totalSize=36913]  
2  
OK  
Time taken: 1.459 seconds  
hive> create table state part(District string,Enrolments string) PARTITIONED BY(state string);  
3  
OK  
Time taken: 0.199 seconds
```

Loading Data into "allstates"

creation of partition table "state_part"

(./images/Hive/120415_1301_Dataoperati8.png).

```

hive> set hive.exec.dynamic.partition.mode=nonstrict
>
hive> insert overwrite table state part PARTITION(state) SELECT district,enrolments,state from allstates;
Query ID = hduser_20151104161604_ce10a013-7e6e-4545-94b9-b26dcaad8879
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201511041430_0001, Tracking URL = http://localhost:50030/jobdetails
Kill Command = /usr/local/hadoop-1.2.1/libexec/./bin/hadoop job -kill job_2015110414
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2015-11-04 16:16:13,677 Stage-1 map = 0%, reduce = 0%
2015-11-04 16:16:18,699 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.48 sec
2015-11-04 16:16:19,702 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 1.48 sec
MapReduce Total cumulative CPU time: 1 seconds 480 msec
Ended Job = job_201511041430_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://localhost:54310/user/hive/warehouse/state_part/.hive-staging_hive_2015-11-04_16-16-04_2
Loading data to table default.state part partition (state=null)
Time taken for load dynamic partitions : 5282
Loading partition {state=Haryana}
Loading partition {state=Uttarakhand}
Loading partition {state=Daman_and_Diu}
Loading partition {state=Puducherry}
Loading partition {state=Uttar_Pradesh}
Loading partition {state=Assam}
Loading partition {state=Others}
Loading partition {state=Arunachal_Pradesh}
Loading partition {state=Lakshadweep}
Loading partition {state=West_Bengal}
Loading partition {state=Sikkim}
Loading partition {state=Himachal_Pradesh}
Loading partition {state=Jharkhand}
Loading partition {state=Tripura}
Loading partition {state=Punjab}
Loading partition {state=Tamil_Nadu}
Loading partition {state=Gujarat}

```

Making
partition based
on "state" field

Creation partition tables
using "state" as partition key
during Map reduce process

(./images/Hive/120415_1301_Dataoperati9.png).

```
hduser@datamatics-Ubuntu: /usr/local/hadoop-1.2.1/bin$ ./hadoop dfs -ls /user/hive/warehouse/state_part
Warning: $HADOOP_HOME is deprecated.

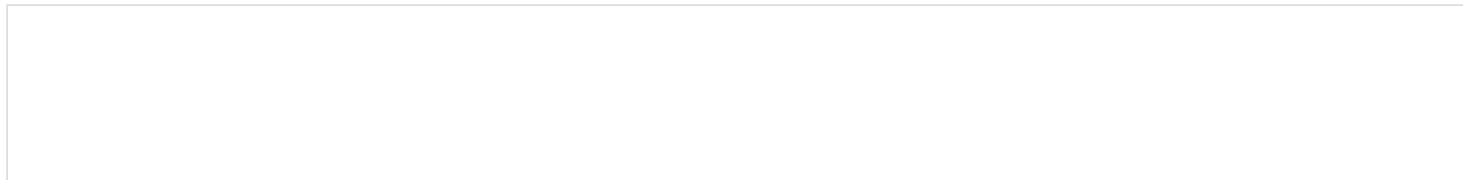
Found 38 items
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Andaman_and_Nicobar_Island
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Andhra Pradesh
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Arunachal Pradesh
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Assam
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Bihar
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Chandigarh
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Chhattisgarh
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Dadra_and_Nagar_Haveli
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Daman_and_Diu
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Delhi
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Goa
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Gujarat
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Haryana
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Himachal Pradesh
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Jammu_and_Kashmir
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Jharkhand
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Karnataka
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Kerala
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Lakshadweep
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Madhya Pradesh
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Maharashtra
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Manipur
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Meghalaya
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Mizoram
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Nagaland
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Odisha
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Others
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Puducherry
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Punjab
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Rajasthan
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Sikkim
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Tamil Nadu
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Tripura
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Uttar Pradesh
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=Uttarakhand
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=West Bengal
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=california
drwxr-xr-x - hduser supergroup 0 2015-11-04 16:16 /user/hive/warehouse/state_part/state=newyork
```

Total 38 states present in table

38 Partition tables stored in HDFS system

(/images/Hive/120415_1301_Dataoperati10.png)

From the above code, we do following things



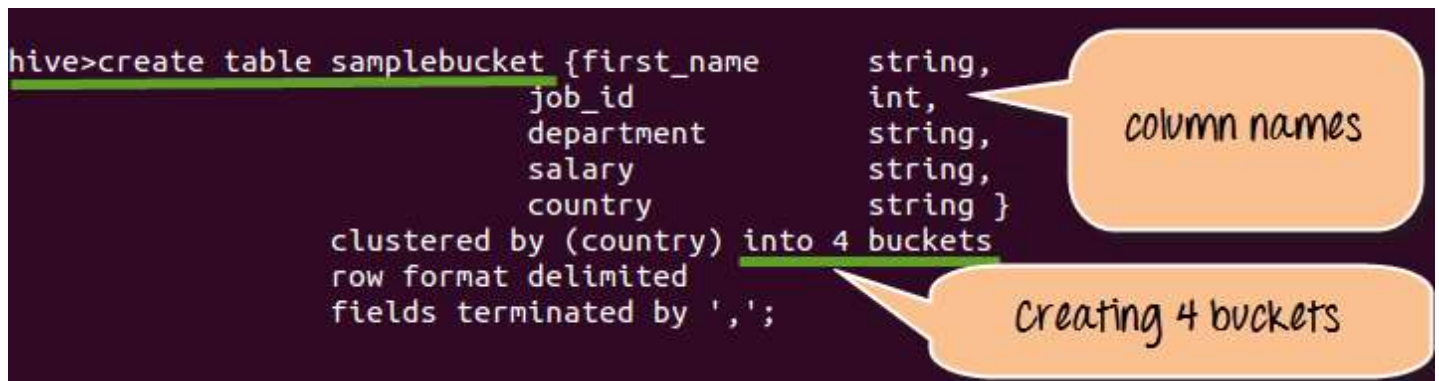
1. Creation of table all states with 3 column names such as state, district, and enrollment
2. Loading data into table all states
3. Creation of partition table with state as partition key
4. In this step Setting partition mode as non-strict(This mode will activate dynamic partition mode)
5. Loading data into partition tablestate_part
6. Actual processing and formation of partition tables based on state as partition key
7. There is going to 38 partition outputs in HDFS storage with the file name as state name. We will check this in this step. In This step, we seeing the 38 partition outputs in HDFS

What is Buckets?

Buckets in hive is used in segregating of hive table-data into multiple files or directories. it is used for efficient querying.

- The data i.e. present in that partitions can be divided further into Buckets
- The division is performed based on Hash of particular columns that we selected in the table.
- Buckets use some form of Hashing algorithm at back end to read each record and place it into buckets
- In Hive, we have to enable buckets by using the **set.hive.enforce.bucketing=true;**

Step 1) Creating Bucket as shown below.



(./images/Hive/120415_1301_Dataoperati11.png)

From the above screen shot

- We are creating sample_bucket with column names such as first_name, job_id, department, salary and country
- We are creating 4 buckets overhere.
- Once the data get loaded it automatically, place the data into 4 buckets

Step 2) Loading Data into table sample bucket

Assuming that "Employees table" already created in Hive system. In this step, we will see the loading of Data from employees table into table sample bucket.

Before we start moving employees data into buckets, make sure that it consist of column names such as first_name, job_id, department, salary and country.

Here we are loading data into sample bucket from employees table.

```
from employees
insert overwrite table samplebucket
select first_name,job_id, department,salary, country;
```

loading Data

(./images/Hive/120415_1301_Dataoperati12.png)

Step 3)Displaying 4 buckets that created in Step 1

```
guru99hive@ubuntu:~/Hadoop_YARN/hadoop-2.2.0/bin$ ./hadoop fs -ls /user/hive/
Found 3 items
--rwx--- 1 guru guru 4602 2015-11-02 09:30 /user/hive/guru99db/samplebucket/
000000 0
--rwx--- 1 guru guru 4602 2015-11-02 09:30 /user/hive/guru99db/samplebucket/
000000_1
--rwx--- 1 guru guru 4602 2015-11-02 09:30 /user/hive/guru99db/samplebucket/
000000 2
--rwx--- 1 guru guru 4602 2015-11-02 09:30 /user/hive/guru99db/samplebucket/
000000 3
```

4 buckets created

(./images/Hive/120415_1301_Dataoperati13.png)

From the above screenshot, we can see that the data from the employees table is transferred into 4 buckets created in step 1.

◀ [Prev \(/hive-create-alter-drop-table.html\)](/hive-create-alter-drop-table.html)

[Report a Bug](#)

[Next > \(/hive-indexes-view-example.html\)](/hive-indexes-view-example.html)

YOU MIGHT LIKE:

HIVE

(/hive-create-alter-drop-table.html) (/hive-create-alter-drop-table.html)

[Hive Create, Alter & Drop Table](#)

(/hive-create-alter-drop-table.html)

HIVE

(/installation-configuration-hive-mysql.html) (/installation-configuration-hive-mysql.html)

[HIVE Installation & Configuration with MYSQL](#)

(/installation-configuration-hive-mysql.html)

HIVE

(/hive-join-subquery.html) (/hive-join-subquery.html)

[Hive Join & SubQuery Tutorial with Examples](#)

(/hive-join-subquery.html)

(/data-operations-



hive.html) (/data-
operations-

hive.html)

**Hive Data Types & Create,
Drop Database**

(/data-operations-hive.html)

(/hive-queries-

implementation.html)



(/hive-queries-

implementation.html)

**Hive Queries: Order By,
Group By, Distribute By,
Cluster By Examples**

(/hive-queries-

implementation.html)

(/introduction-hive.html)



(/introduction-
hive.html)

**What is Hive? Architecture &
Modes**

(/introduction-hive.html)

Hive Tutorials

- [1\) Introduction to Hive \(/introduction-hive.html\)](/introduction-hive.html)
- [2\) Install and Configuration \(/installation-configuration-hive-mysql.html\)](/installation-configuration-hive-mysql.html)
- [3\) Data operations in Hive \(/data-operations-hive.html\)](/data-operations-hive.html)
- [4\) Create, Alter & Drop Table \(/hive-create-alter-drop-table.html\)](/hive-create-alter-drop-table.html)
- [5\) Partitions & Buckets \(/hive-partitions-buckets-example.html\)](/hive-partitions-buckets-example.html)
- [6\) Indexes and View \(/hive-indexes-view-example.html\)](/hive-indexes-view-example.html)
- [7\) Queries and Implementation \(/hive-queries-implementation.html\)](/hive-queries-implementation.html)
- [8\) Join & SubQuery \(/hive-join-subquery.html\)](/hive-join-subquery.html)
- [9\) Query Language & Operators \(/hive-query-language-built-operators-functions.html\)](/hive-query-language-built-operators-functions.html)
- [10\) Function: Built-in & UDF \(/hive-user-defined-functions.html\)](/hive-user-defined-functions.html)
- [11\) Data Extraction Using Hive \(/data-extraction-hive.html\)](/data-extraction-hive.html)



(<https://www.facebook.com/guru99com/>).



(<https://twitter.com/guru99com>).



(<https://www.youtube.com/channel/UC19i1XD6k88KqHlET8atqFQ>).



(<https://forms.aweber.com/form/46/724807646.htm>).

About

[About US \(/about-us.html\)](/about-us.html)

[Advertise with Us \(/advertise-us.html\)](/advertise-us.html)

[Write For Us \(/become-an-instructor.html\)](/become-an-instructor.html)

[Contact US \(/contact-us.html\)](/contact-us.html)

Career Suggestion

[SAP Career Suggestion Tool \(/best-sap-module.html\)](/best-sap-module.html)

[Software Testing as a Career \(/software-testing-career-complete-guide.html\)](/software-testing-career-complete-guide.html)

[Certificates \(/certificate-it-professional.html\)](/certificate-it-professional.html)

Interesting

[Books to Read! \(/books.html\)](/books.html)

[Suggest a Tutorial](#)

[Blog \(/blog/\)](/blog/)

[Quiz \(/tests.html\)](/tests.html)

[eBook \(/ebook-pdf.html\)](/ebook-pdf.html)

Execute online

[Execute Java Online \(/try-java-editor.html\)](/try-java-editor.html)

[Execute Javascript \(/execute-javascript-online.html\)](/execute-javascript-online.html)

[Execute HTML \(/execute-html-online.html\)](/execute-html-online.html)

[Execute Python \(/execute-python-online.html\)](/execute-python-online.html)

© Copyright - Guru99 2019

[Privacy Policy \(/privacy-policy.html\)](/privacy-policy.html)