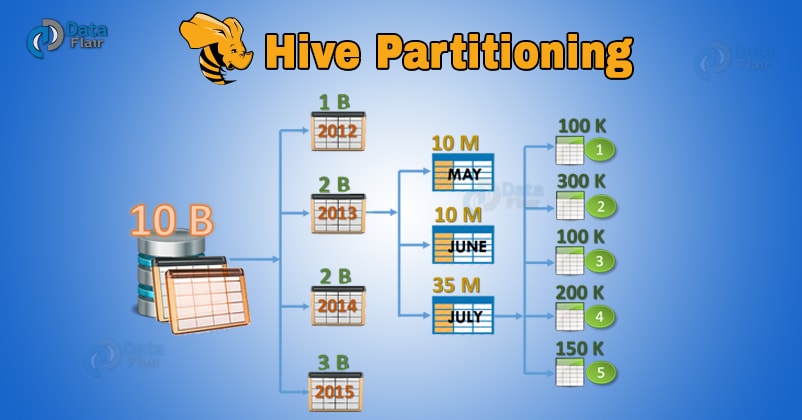
**Hive Partitions, Types of Hive Partitioning with Examples**

[*24 Aug, 2016*](https://data-flair.training/blogs/2016/08/24/) *in* [*Hive Tutorials*](https://data-flair.training/blogs/category/hive/) *by* [*Support*](https://data-flair.training/blogs/author/df-hd-team/)

<https://data-flair.training/blogs/apache-hive-partitions/>

## **1. Objective**

The [**Hive tutorial**](http://data-flair.training/blogs/apache-hive-tutorial/) explains about the Hive partitions. This blog will help you to answer what is Hive partitioning, what is the need of partitioning, how it improves the performance? Partitioning is the optimization technique in Hive which improves the performance significantly. Apache Hive is the data warehouse on the top of [**Hadoop**](http://data-flair.training/blogs/hadoop-introduction-tutorial-quick-guide/), which enables ad-hoc analysis over structured and semi-structured data. Let’s discuss Apache Hive partitioning in detail.

[](https://d2h0cx97tjks2p.cloudfront.net/blogs/wp-content/uploads/apache-hive-partitioning.jpg)

## **2. What are Hive Partitions?**

**Apache Hive** organizes tables into **partitions**. Partitioning is a way of dividing a table into related parts based on the values of particular columns like date, city, and department. Each table in the hive can have one or more partition keys to identify a particular partition. Using partition it is easy to do queries on slices of the data.

## **3. Why is Partitioning Important?**

In the current century, we know that the huge amount of data which is in the range of petabytes is getting stored in[**HDFS**.](http://data-flair.training/blogs/comprehensive-hdfs-guide-introduction-architecture-data-read-write-tutorial/) So due to this, it becomes very difficult for Hadoop users to query this huge amount of data.

The Hive was introduced to lower down this burden of data querying. Apache Hive converts the SQL queries into [**MapReduce**](http://data-flair.training/blogs/hadoop-mapreduce-introduction-tutorial-comprehensive-guide/) jobs and then submits it to the [**Hadoop cluster**](http://data-flair.training/blogs/install-configure-apache-hadoop-2-7-x-on-ubuntu/). When we submit a SQL query, Hive read the entire data-set. So, it becomes inefficient to run [**MapReduce jobs**](http://data-flair.training/blogs/hadoop-mapreduce-job-execution-flow/) over a large table. Thus this is resolved by creating partitions in tables. Apache Hive makes this job of implementing partitions very easy by creating partitions by its automatic partition scheme at the time of table creation.

In Partitioning method, all the table data is divided into multiple partitions. Each partition corresponds to a specific value(s) of partition column(s). It is kept as a sub-record inside the table’s record present in the HDFS. Therefore on querying a particular table, appropriate partition of the table is queried which contains the query value. Thus this decreases the I/O time required by the query. Hence increases the performance speed.

## **4. How to create Partitions in Hive?**

[](https://data-flair.training/blogs/category/quiz/hadoop-quiz/)

To create data partitioning in Hive following command is used-

CREATE TABLE table\_name (column1 data\_type, column2 data\_type) PARTITIONED BY (partition1 data\_type, partition2 data\_type,….);

## **5. Hive Data Partitioning Example**

**[Learn Hadoop from Industry Experts](https://data-flair.training/big-data-hadoop/" \t "_blank)**

Now let’s understand data partitioning in Hive with an example. Consider a table named **Tab1**. The table contains client detail like id, name, dept, and yoj( year of joining). Suppose we need to retrieve the details of all the clients who joined in 2012. Then, the query searches the whole table for the required information. But if we partition the client data with the year and store it in a separate file, this will reduce the query processing time. The below example will help us to learn how to partition a file and its data-

The file name says file1 contains client data table:

1. tab1/clientdata/file1
2. id, name, dept, yoj
3. 1, sunny, SC, 2009
4. 2, animesh, HR, 2009
5. 3, sumeer, SC, 2010
6. 4, sarthak, TP, 2010

Now, let us partition above data into two files using years

1. tab1/clientdata/2009/file2
2. 1, sunny, SC, 2009
3. 2, animesh, HR, 2009
4. tab1/clientdata/2010/file3
5. 3, sumeer, SC, 2010
6. 4, sarthak, TP, 2010

[](https://data-flair.training/blogs/category/interview-questions/hadoop-interview-questions/)

Now when we are retrieving the data from the table, only the data of the specified partition will be queried. Creating a partitioned table is as follows:

1. CREATE TABLE table\_tab1 (id INT, name STRING, dept STRING, yoj INT) PARTITIONED BY (year STRING);
2. LOAD DATA LOCAL INPATH tab1’/clientdata/2009/file2’OVERWRITE INTO TABLE studentTab PARTITION (year='2009');
3. LOAD DATA LOCAL INPATH tab1’/clientdata/2010/file3’OVERWRITE INTO TABLE studentTab PARTITION (year='2010');

**[Fascinated by Big Data? Enroll Now!](https://data-flair.training/big-data-hadoop/" \t "_blank)**

## **6. Types of Hive Partitioning**

Till now we have discussed Introduction to Hive Partitions and How to create Hive partitions. Now we are going to introduce the types of data partitioning in Hive. There are two types of Partitioning in Apache Hive-

* Static Partitioning
* Dynamic Partitioning

Let’s discuss these types of Hive Partitioning one by one-

### **6.1. Hive Static Partitioning**

* 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 the table and move the file into the partition of the table.
* We can alter the partition in the static partition.
* You can get the partition column value from the filename, day of date etc without reading the whole big file.
* If you want to use the Static partition in the 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 the static partition.
* You can perform Static partition on Hive Manage table or external table.

### **6.2. Hive Dynamic Partitioning**

* Single insert to partition table is known as a dynamic partition.
* Usually, dynamic partition loads the data from the 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 the Dynamic partition is suitable.
* If you want to partition a number of columns 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 the Dynamic partition.
* You can perform dynamic partition on hive external table and managed table.
* If you want to use the Dynamic partition in the hive then the mode is in non-strict mode.
* Here are Hive dynamic partition properties you should allow

## **7. Hive Partitioning – Advantages and Disadvantages**

Let’s discuss some benefits and limitations of Apache Hive Partitioning-

**a) Hive Partitioning Advantages**

* Partitioning in Hive distributes execution load horizontally.
* In partition faster execution of queries with the low volume of data takes place. For example, search population from Vatican City returns very fast instead of searching entire world population.

**b) Hive Partitioning Disadvantages**

* There is the possibility of too many small partition creations- too many directories.
* Partition is effective for low volume data. But there some queries like group by on high volume of data take a long time to execute. For example, grouping population of China will take a long time as compared to a grouping of the population in Vatican City.
* There is no need for searching entire table column for a single record.

## **8. Conclusion**

Hope this blog will help you a lot to understand what exactly is partition in Hive, what is Static partitioning in Hive, What is Dynamic partitioning in Hive. We have also covered various advantages and disadvantages of Hive partitioning. If you have any query related to Hive Partitions, so please leave a comment. We will be glad to solve them.

**See Also-**

* [Apache Hive Architecture](http://data-flair.training/blogs/apache-hive-architecture/)
* [Hive Partitioning vs Bucketing](http://data-flair.training/blogs/hive-partitioning-vs-bucketing/)