Hive Partitioning

1. **Load Data to HDFS somewhere in /user/root directory.**
2. **Create a Non Partitioned table to store the data (Staging Table)**

create external table Unm\_Dup\_Parti (EmployeeID Int,FirstName String,Designation String,Salary Int,Department String) row format delimited fields terminated by "," location '/user/root/HiveTrail';

1. **Create Partitioned Hive Table**

create table Unm\_Parti (EmployeeID Int,FirstName String,Designation String,Salary Int) PARTITIONED BY (Department String) row format delimited fields terminated by ",";

1. **Insert Data into partitioned table**
2. **STATIC PARTITION**

INSERT INTO TABLE Unm\_Parti PARTITION(department='A')

SELECT EmployeeID, FirstName,Designation,Salary FROM Unm\_Dup\_Parti WHERE department='A';

INSERT INTO TABLE Unm\_Parti PARTITION (department='B')

SELECT EmployeeID, FirstName,Designation,Salary FROM Unm\_Dup\_Parti WHERE department='B';

INSERT INTO TABLE Unm\_Parti PARTITION (department='C')

SELECT EmployeeID, FirstName,Designation,Salary FROM Unm\_Dup\_Parti WHERE department='C';

1. **DYNAMIC PARTITION**

**1. set hive.exec.dynamic.partition=true**

This enable dynamic partitions, by default it is false.

**2. set hive.exec.dynamic.partition.mode=nonstrict**

We are using the dynamic partition without a static partition (A table can be partitioned based

on multiple columns in hive) in such case we have to enable the non strict mode. In strict mode we can use dynamic partition only with a Static Partition.

**3. set hive.exec.max.dynamic.partitions.pernode=3**

The default value is 100, we have to modify the same according to the possible no of partitions

**4. hive.exec.max.created.files=150000**

The default values is 100000 but for larger tables it can exceed the default, so we may have to update the same.

INSERT OVERWRITE TABLE Unm\_Parti PARTITION(department) SELECT EmployeeID, FirstName,Designation,Salary,department FROM Unm\_Dup\_Parti;