

# Introduction to Apache Hive



# Part 2

## HiveQL Data Manipulation Language (DML)

# What Does Hive DML Do?

- Put data into tables by using one of the “bulk” load operations since Hive has no row-level insert, update, and delete operations.
- Extract data from tables to the filesystem.

# Load Comma-Delimited Data File

```
CREATE TABLE IF NOT EXISTS employees12 (
    name STRING COMMENT 'Employee name',
    salary FLOAT COMMENT 'Employee salary',
    FedTax FLOAT,
    StateTax FLOAT,
    Insurance FLOAT,
    address STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n';
```

```
Mary Smith,80000.0,.2,.05,.1,100 Ontario St. Chicago IL 60601
Todd Jones,70000.0,.15,.03,.1,108 Oak Park Chicago IL 60700
Bill King,60000.0,.15,.03,.1,10 Ontario St. Chicago IL 60700
Sam Kim, 110000.0,.2,.05,.1,101 Uptown Rd. Ithaca NY 14850
Maria Booth, 60000.0,.15,.03,.1,46 South Gate Dr. State College PA 16801
Joe Howe, 75000.0,.2,.5,.1,24 Garden Ave. Ithaca NY 14853
```

# Load Comma-Delimited Data Files

```
LOAD DATA LOCAL INPATH '/empdata-comma.txt'  
OVERWRITE INTO TABLE employees12;
```

The LOCAL keyword:

- When used: path in the local fs; data **copied** into final location.
- When omitted: path in HDFS; data **moved** from path to final location.

```
hive> select * from employees12;  
OK  
John Smith      100000.0      0.2      0.05      0.1      1 Michigan Ave. Chicago IL 60600  
Mary Smith       80000.0 0.2      0.05      0.1      100 Ontario St. Chicago IL 60601  
Todd Jones       70000.0 0.15     0.03      0.1      108 Oak Park Chicago IL 60700  
Bill King        60000.0 0.15     0.03      0.1      10 Ontario St. Chicago IL 60700  
Sam Kim          110000.0      0.2      0.05      0.1      101 Uptown Rd. Ithaca NY 14850  
Maria Booth        60000.0 0.15     0.03      0.1      46 South Gate Dr. State College PA 16801  
Joe Howe          75000.0 0.2      0.5      0.1      24 Garden Ave. Ithaca NY 14853
```

# Load Tab-Delimited Data File

```
CREATE TABLE IF NOT EXISTS employees11 (
    name STRING COMMENT 'Employee name',
    salary FLOAT COMMENT 'Employee salary',
    FedTax FLOAT,
    StateTax FLOAT,
    Insurance FLOAT,
    address STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '\t'
LINES TERMINATED BY '\n';
```

Todd Jones	70000.0	.15	.03	.1	108 Oak Park Chicago IL 60700
Bill King	60000.0	.15	.03	.1	10 Ontario St. Chicago IL 60700
Sam Kim	110000.0	.2	.05	.1	101 Uptown Rd. Ithaca NY 14850
Maria Booth	60000.0		.15	.03	.1 46 South Gate Dr. State College PA 16803
Joe Howe	75000.0		.2	.5	.1 24 Garden Ave. Ithaca NY 14853

# Load Tab-Delimited Data Files

```
LOAD DATA LOCAL INPATH '/empdata-tab.txt'  
OVERWRITE INTO TABLE employees11;
```

```
Time taken: 2.4 seconds  
hive> select * from employees11;  
OK  
John Smith      100000.0      0.2      0.05      0.1      1 Michigan Ave. Chicago IL 60600  
Mary Smith       80000.0 0.2      0.05      0.1      100 Ontario St. Chicago IL 60601  
Todd Jones       70000.0 0.15      0.03      0.1      108 Oak Park Chicago IL 60700  
Bill King        60000.0 0.15      0.03      0.1      10 Ontario St. Chicago IL 60700  
Sam Kim          110000.0      0.2      0.05      0.1      101 Uptown Rd. Ithaca NY 14850  
Maria Booth        60000.0 0.15      0.03      0.1      46 South Gate Dr. State College PA 16801  
Joe Howe          75000.0 0.2      0.5      0.1      24 Garden Ave. Ithaca NY 14853
```

# Load Data of Complex Data Types

```

CREATE TABLE IF NOT EXISTS employees20 (
    name STRING COMMENT 'Employee name',
    salary FLOAT COMMENT 'Employee salary',
    subordinates ARRAY<STRING> COMMENT 'Names of subordinates',
    deductions MAP<STRING, FLOAT>
        COMMENT 'Keys are deductions names, values are percentages',
    address STRUCT<street:STRING, city:STRING, state:STRING, zip:INT>
        COMMENT 'Home address')
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '^'
COLLECTION ITEMS TERMINATED BY '#'
MAP KEYS TERMINATED BY '@'
LINES TERMINATED BY '\n'
STORED AS TEXTFILE;

```

```

John Smith^100000.0^Mary Smith#David Wells^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^1 Michigan Ave.#Chicago#IL#60600
Mary Smith^80000.0^Bill King^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^100 Ontario St.#Chicago#IL#60601
Jason Yang^150000.0^Diana Foster#Kelli Doe#Kevin Hoover^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^101 Uptown Rd.#Ithaca#NY#14850
Ted Wang^120000.0^Diana Johnson#Kevin Li#Dan Jones^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^21 Farm Rd.#Ithaca#NY#14850
Mike McPheron^150000.0^Marvin Xu#Elise Lee^Federal Taxes@0.02#State Taxes@0.05#Insurance@0.1^12 Bush Ave.#Ithaca#NY#14850

```

# Load Data of Complex Data Types

**LOAD DATA LOCAL INPATH '/empdata.txt'**

**OVERWRITE INTO TABLE employees20;**

```
hive> select * from employees20;
OK
John Smith      100000.0      ["Mary Smith","David Wells"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"1 Michigan Ave.", "city":"Chicago", "state": "IL", "zip": 60600}
Mary Smith       80000.0      ["Bill King"] {"Federal Taxes":0.2,"State Taxes":0.05, "Insurance":0.1} {"street": "100 Ontario St.", "city": "Chicago", "state": "IL", "zip": 60601}
Jason Yang       150000.0      ["Diana Foster","Kelli Doe","Kevin Hoover"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street": "101 Uptown Rd.", "city": "Ithaca", "state": "NY", "zip": 14850}
Ted Wang          120000.0      ["Diana Johnson","Kevin Li","Dan Jones"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street": "21 Farm Rd.", "city": "Ithaca", "state": "NY", "zip": 14850}
Mike McPheron     150000.0      ["Marvin Xu","Elise Lee"] {"Federal Taxes":0.02,"State Taxes":0.05,"Insurance":0.1} {"street": "12 Bush Ave.", "city": "Ithaca", "state": "NY", "zip": 14850}
Time taken: 0.58 seconds, Fetched: 5 row(s)
hive>
```

# Load Data of Complex Data Types into a Partitioned Table

```
CREATE TABLE IF NOT EXISTS employees23(
    name STRING COMMENT 'Employee name',
    salary FLOAT COMMENT 'Employee salary',
    subordinates ARRAY<STRING> COMMENT 'Names of subordinates' ,
    deductions MAP<STRING, FLOAT>
        COMMENT 'Keys are deductions names, values are percentages',
    address STRUCT<street:STRING, city:STRING, state:STRING, zip:INT>
        COMMENT 'Home address' )
PARTITIONED BY (country STRING, state STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '^'
COLLECTION ITEMS TERMINATED BY '#'
MAP KEYS TERMINATED BY '@'
LINES TERMINATED BY '\n'
STORED AS TEXTFILE;
```

# Load Data of Complex Data Types into a Partitioned Table (cont'd)

```
LOAD DATA LOCAL INPATH '/empdata-us-il.txt'
OVERWRITE INTO TABLE employees23
PARTITION (country = 'US', state = 'IL');
```

```
John Smith^100000.0^Mary Smith#David Wells^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^1 Michigan Ave.#Chicago#IL#60600
Mary Smith^80000.0^Bill King^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^100 Ontario St.#Chicago#IL#60601
Jason Jones^150000.0^Diana Foster#Kelli Doe#Kevin Hoover^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^101 Uptown Rd.#Ithaca#IL#60666
Bob McPheron^150000.0^Marvin Xu#Elise Lee^Federal Taxes@0.02#State Taxes@0.05#Insurance@0.1^12 Bush Ave.#Chicargo#IL#60606
~
```

```
hive> load data local inpath '/empdata-us-il.txt'
    > overwrite into table employees23
    > partition (country='US', state='IL');
```

```
Loading data to table default.employees23 partition (country=US, state=IL)
Partition default.employees23{country=US, state=IL} stats: [numFiles=1, numRows=0, totalSize=495, rawDataSize=0]
```

```
OK
```

```
Time taken: 1.786 seconds
```

```
hive> _
```

# Query Partitioned Table of Complex Data Types

```
Select * from employees23 ;
```

```
John Smith      100000.0      ["Mary Smith", "David Wells"]      {"Federal Taxes": 0.2, "State Taxes": 0.05, "Insurance": 0.1}      {"street": "1 Michigan Ave.", "city": "Chicago", "state": "IL", "zip": 60600}  US      IL
Mary Smith      80000.0      ["Bill King"]      {"Federal Taxes": 0.2, "State Taxes": 0.05, "Insurance": 0.1}      {"street": "100 Ontario St.", "city": "Chicago", "state": "IL", "zip": 60601}  US      IL
Jason Jones     150000.0      ["Diana Foster", "Kelli Doe", "Kevin Hoover"]      {"Federal Taxes": 0.2, "State Taxes": 0.05, "Insurance": 0.1}      {"street": "101 Uptown Rd.", "city": "Ithaca", "state": "IL", "zip": 60666}  US      IL
Bob McPheron     150000.0      ["Marvin Xu", "Elise Lee"]      {"Federal Taxes": 0.02, "State Taxes": 0.05, "Insurance": 0.1}      {"street": "12 Bush Ave.", "city": "Chicago", "state": "IL", "zip": 60606}  US      IL
Time taken: 0.784 seconds, Fetched: 4 row(s)
```

# Load Data of Complex Data Types into a Partitioned Table (cont'd)

```
LOAD DATA LOCAL INPATH '/empdata-us-ny.txt'  
OVERWRITE INTO TABLE employees23  
PARTITION (country = 'US', state = 'NY');
```

```
nsurance@0.1^1 Square Ave.#New York#NY#10060  
Elise Smith^80000.0^Bill King^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^1  
00 Ontario St.#New York#NY#10001  
Jason Yang^150000.0^Diana Foster#Kelli Doe#Kevin Hoover^Federal Taxes@0.2#State  
Taxes@0.05#Insurance@0.1^101 Uptown Rd.#Ithaca#NY#14850  
Ted Wang^120000.0^Diana Johnson#Kelli Li#Dan Jones^Federal Taxes@0.2#State Taxes  
@0.05#Insurance@0.1^21 Farm Rd.#Ithaca#NY#14850  
Tom McPheron^150000.0^Marvin Xu#Elise Lee^Federal Taxes@0.02#State Taxes@0.05#In  
surance@0.1^12 Bush Ave.#New York#NY#10002
```

```
hive> load data local inpath '/empdata-us-ny.txt'  
> overwrite into table employees23  
> partition (country='US', state='NY');  
Loading data to table default.employees23 partition (country=US, state=NY)  
Partition default.employees23{country=US, state=NY} stats: [numFiles=1, numRows=0,  
totalSize=625, rawDataSize=0]  
OK  
Time taken: 1.717 seconds
```

# Load Data of Complex Data Types into a Partitioned Table (cont'd)

```
LOAD DATA LOCAL INPATH '/empdata-us-ca.txt'
OVERWRITE INTO TABLE employees23
PARTITION (country = 'US', state = 'CA');
```

```
Joann Fraser^100000.0^Mary Smith#David Wells^Federal Taxes@0.2#State Taxes@0.05#
Insurance@0.1^100 Bay Ave.#Long Beach#CA#90814
Elise Smith^80000.0^Bill King^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^1
00 Beach St.#Long Beach#CA#90815
Jason Yang^150000.0^Diana Foster#Kelli Doe#Kevin Hoover^Federal Taxes@0.2#State
Taxes@0.05#Insurance@0.1^101 Uptown Rd.#San Diego#CA#92106
Ted Wang^190000.0^Diana Johnson#Kelli Li#Dan Jones^Federal Taxes@0.2#State Taxes
@0.05#Insurance@0.1^1 Kerrick Rd.#San Diego#CA#92106
Tom McPheron^210000.0^Marvin Xu#Elise Lee^Federal Taxes@0.02#State Taxes@0.05#In
surance@0.1^12 Bush Lane.#San Diego#CA#92106
```

```
hive> load data local inpath '/empdata-us-ca.txt'
    > overwrite into table employees23
    > partition (country='US', state='CA');
```

```
Loading data to table default.employees23 partition (country=US, state=CA)
Partition default.employees23{country=US, state=CA} stats: [numFiles=1, numRows=
0, totalSize=638, rawDataSize=0]
OK
Time taken: 1.734 seconds
```

# Query Partitioned Table of Complex Data Types

See the whole table with three partitions:

Select \* from employees23;

```
hive> select * from employees23;
OK
Joann Fraser    100000.0      {"Mary Smith","David Wells"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"100 Bay Ave.", "city":"Long Beach", "state":"CA", "zip":90814} US CA
Elise Smith     80000.0      {"Bill King"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"100 Beach St.", "city":"Long Beach", "state":"CA", "zip":90815} US CA
Jason Yang      150000.0      {"Diana Foster","Kelli Doe","Kevin Hoover"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"101 Uptown Rd.", "city":"San Diego", "state":"CA", "zip":92106} US CA
Ted Wang         190000.0      {"Diana Johnson","Kelli Li","Dan Jones"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"1 Kerrick Rd.", "city":"San Diego", "state":"CA", "zip":92106} US CA
Tom McPheron    210000.0      {"Marvin Xu","Elise Lee"} {"Federal Taxes":0.02,"State Taxes":0.05,"Insurance":0.1} {"street":"12 Bush Lane.", "city":"San Diego", "state":"CA", "zip":92106} US CA
John Smith      100000.0      {"Mary Smith","David Wells"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"1 Michigan Ave.", "city":"Chicago", "state":"IL", "zip":60600} US IL
Mary Smith       80000.0      {"Bill King"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"100 Ontario St.", "city":"Chicago", "state":"IL", "zip":60601} US IL
Jason Jones     150000.0      {"Diana Foster","Kelli Doe","Kevin Hoover"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"101 Uptown Rd.", "city":"Ithaca", "state":"IL", "zip":60666} US IL
Bob McPheron    150000.0      {"Marvin Xu","Elise Lee"} {"Federal Taxes":0.02,"State Taxes":0.05,"Insurance":0.1} {"street":"12 Bush Ave.", "city":"Chicargo", "state":"IL", "zip":60606} US IL
Sam Johnson     100000.0      {"Mary Smith","David Wells"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"1 Square Ave.", "city":"New York", "state":"NY", "zip":10060} US NY
Elise Smith     80000.0      {"Bill King"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"100 Ontario St.", "city":"New York", "state":"NY", "zip":10001} US NY
Jason Yang      150000.0      {"Diana Foster","Kelli Doe","Kevin Hoover"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"101 Uptown Rd.", "city":"Ithaca", "state":"NY", "zip":14850} US NY
Ted Wang         120000.0      {"Diana Johnson","Kelli Li","Dan Jones"} {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"21 Farm Rd.", "city":"Ithaca", "state":"NY", "zip":14850} US NY
Tom McPheron    150000.0      {"Marvin Xu","Elise Lee"} {"Federal Taxes":0.02,"State Taxes":0.05,"Insurance":0.1} {"street":"12 Bush Ave.", "city":"New York", "state":"NY", "zip":10002} US NY
Time taken: 0.531 seconds, Fetched: 14 row(s)
```

# Query Partitioned Table of Complex Data Types, Using a Condition from a Partition

```
Select * from employees23 where state="NY";
```

```
hive> select * from employees23 where state = 'NY';
OK
Sam Johnson    100000.0      ["Mary Smith","David Wells"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"1 Square Ave.", "city":"New York", "state":"NY", "zip":10060} US NY
Elise Smith    80000.0      ["Bill King"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"100 Ontario St.", "city":"New York", "state":"NY", "zip":10001} US NY
Jason Yang     150000.0      ["Diana Foster","Kelli Doe","Kevin Hoover"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"101 Uptown Rd.", "city":"Ithaca", "state":"NY", "zip":14850} US NY
Ted Wang       120000.0      ["Diana Johnson","Kelli Li","Dan JOnes"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"21 Farm Rd.", "city":"Ithaca", "state":"NY", "zip":14850} US NY
Tom McPheron   150000.0      ["Marvin Xu","Elise Lee"] {"Federal Taxes":0.02,"State Taxes":0.05,"Insurance":0.1} {"street":"12 Bush Ave.", "city":"New York", "state":"NY", "zip":10002} US NY
Time taken: 0.297 seconds, Fetched: 5 row(s)
```

# Query Partitioned Table of Complex Data Types, Using a Condition from a Column (struct) field

```
Select * from employees23 where address.state="NY";
```

```
hive> select * from employees23 where address.state = 'NY';
OK
Sam Johnson    100000.0      ["Mary Smith","David Wells"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"1 Square Ave.", "city":"New York", "state": "NY", "zip": 10060} US NY
Elise Smith    80000.0      ["Bill King"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"100 Ontario St.", "city":"New York", "state": "NY", "zip": 10001} US NY
Jason Yang     150000.0      ["Diana Foster","Kelli Doe","Kevin Hoover"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"101 Uptown Rd.", "city":"Ithaca", "state": "NY", "zip": 14850} US NY
Ted Wang       120000.0      ["Diana Johnson","Kelli Li","Dan Jones"] {"Federal Taxes":0.2,"State Taxes":0.05,"Insurance":0.1} {"street":"21 Farm Rd.", "city":"Ithaca", "state": "NY", "zip": 14850} US NY
Tom McPheron   150000.0      ["Marvin Xu","Elise Lee"] {"Federal Taxes":0.02,"State Taxes":0.05,"Insurance":0.1} {"street":"12 Bush Ave.", "city":"New York", "state": "NY", "zip": 10002} US NY
Time taken: 0.355 seconds, Fetched: 5 row(s)
```

# Query Partitioned Table of Complex Data Types, Using a Condition from a Column

```
Select name, salary, address.city, address.state  
from employees23  
where salary > 120000;
```

```
hive> select name, salary, address.city, address.state  
    > from employees23  
    > where salary > 120000;  
OK  
Jason Yang      150000.0      San Diego      CA  
Ted Wang        190000.0      San Diego      CA  
Tom McPheron    210000.0      San Diego      CA  
Jason Jones     150000.0      Ithaca        IL  
Bob McPheron    150000.0      Chicargo      IL  
Jason Yang      150000.0      Ithaca        NY  
Tom McPheron    150000.0      New York      NY  
Time taken: 0.294 seconds, Fetched: 7 row(s)
```

# Query Partitioned Table of Complex Data Types, Using Conditions from Both a Column and a Partition

```
Select name, salary, address.city, address.state  
from employees23  
where salary > 120000 and (state = 'CA' or state = 'IL');
```

```
hive> select name, salary, address.city, address.state  
    > from employees23  
    > where salary > 120000 and (state = 'CA' or state = 'IL');  
OK  
Jason Yang      150000.0          San Diego        CA  
Ted Wang        190000.0          San Diego        CA  
Tom McPheron    210000.0          San Diego        CA  
Jason Jones     150000.0          Ithaca          IL  
Bob McPheron    150000.0          Chicargo       IL  
Time taken: 0.312 seconds, Fetched: 5 row(s)
```

# The Sub-Directories of a Partitioned Table

Those of the employees23 table:

.../employees23/country=US/state=CA  
.../employees23/country=US/state=IL  
.../employees23/country=US/state=NY

```
[root@sandbox ~]# hadoop fs -ls /apps/hive/warehouse/employees23
Found 1 items
drwxrwxrwx  - root hdfs          0 2017-03-20 04:31 /apps/hive/warehouse/employees23/country=US
[root@sandbox ~]# hadoop fs -ls /apps/hive/warehouse/employees23/country=US
Found 3 items
drwxrwxrwx  - root hdfs          0 2017-03-20 04:31 /apps/hive/warehouse/employees23/country=US/state=CA
drwxrwxrwx  - root hdfs          0 2017-03-20 04:29 /apps/hive/warehouse/employees23/country=US/state=IL
drwxrwxrwx  - root hdfs          0 2017-03-20 04:31 /apps/hive/warehouse/employees23/country=US/state=NY
```

# The Sub-Directories and the Files of a Partitioned Table

## The location of the loaded file and its contents:

```
[root@sandbox ~]# hadoop fs -ls /apps/hive/warehouse/employees23/country=US/state=CA
Found 1 items
-rwxrwxrwx  1 root hdfs      637 2017-03-20 04:31 /apps/hive/warehouse/employees23/country
=US/state=CA/empdata-us-ca.txt
[root@sandbox ~]# hadoop fs -cat /apps/hive/warehouse/employees23/country=US/state=CA/empdata
-us-ca.txt
Joann Fraser^100000.0^Mary Smith#David Wells^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1
^100 Bay Ave.#Long Beach#CA#90814
Elise Smith^80000.0^Bill King^Federal Taxes@0.2#State Taxes@0.05#Insurance@0.1^100 Beach St.#
Long Beach#CA#90815
Jason Yang^150000.0^Diana Foster#Kelli Doe#Kevin Hoover^Federal Taxes@0.2#State Taxes@0.05#In
surance@0.1^101 Uptown Rd.#San Diego#CA#92106
Ted Wang^190000.0^Diana Johnson#Kelli Li#Dan Jones^Federal Taxes@0.2#State Taxes@0.05#Insuran
ce@0.1^1 Kerrick Rd.#San Diego#CA#92106
Tom McPheron^210000.0^Marvin Xu#Elise Lee^Federal Taxes@0.02#State Taxes@0.05#Insurance@0.1^1
2 Bush Lane.#San Diego#CA#92106
```

# Insert Data into Tables from Queries

```
CREATE TABLE inserttbl like employees20;
```

```
INSERT OVERWRITE TABLE inserttbl
```

```
SELECT * FROM employees20  
WHERE salary > 100000.0;
```

```
SELECT * FROM inserttbl;
```

```
hive> select * from inserttbl;  
OK  
Jason Yang      150000.0      ["Diana Foster", "Kelli Doe", "Kevin Hoover"]  {"  
"Federal Taxes":0.2, "State Taxes":0.05, "Insurance":0.1} {"street":"101 Uptown Rd.",  
"city":"Ithaca", "state":"NY", "zip":14850}  
Ted Wang        120000.0      ["Diana Johnson", "Kevin Li", "Dan Jones"]  {"  
"Federal Taxes":0.2, "State Taxes":0.05, "Insurance":0.1} {"street":"21 Farm Rd.",  
"city":"Ithaca", "state":"NY", "zip":14850}  
Mike McPheron    150000.0      ["Marvin Xu", "Elise Lee"]      {"Federal Taxes":  
0.02, "State Taxes":0.05, "Insurance":0.1} {"street":"12 Bush Ave.", "city":  
"Ithaca", "state":"NY", "zip":14850}
```

# Insert Data into Tables from Queries

- With OVERWRITE, any previous contents of the partition (or whole table if not partitioned) are replaced.

```
INSERT OVERWRITE TABLE employees  
PARTITION (country = 'US', state = 'OR')  
SELECT * FROM staged_employees se  
WHERE se.cnty = 'US' AND se.st = 'OR';
```

- Without OVERWRITE or with INTO, Hive appends the data rather than replaces it.

```
INSERT INTO TABLE employees  
PARTITION (country = 'US', state = 'OR')  
SELECT * FROM staged_employees se  
WHERE se.cnty = 'US' AND se.st = 'OR';
```

# An INSERT Syntax: Scan the Data Once and Split it Multiple Ways

```
FROM staged_employees se
INSERT OVERWRITE TABLE employees
    PARTITION (country = 'US', state = 'OR')
        SELECT * WHERE se.cnty = 'US' AND se.st = 'OR'
INSERT OVERWRITE TABLE employees
    PARTITION (country = 'US', state = 'CA')
        SELECT * WHERE se.cnty = 'US' AND se.st = 'CA'
INSERT OVERWRITE TABLE employees
    PARTITION (country = 'US', state = 'IL')
        SELECT * WHERE se.cnty = 'US' AND se.st = 'IL';
```

# Dynamic Partition Inserts

Hive's dynamic partition can infer the partitions to create based on query parameters:

```
INSERT OVERWRITE TABLE employees  
PARTITION (country, state)  
SELECT ..., se.cnty, se.st  
FROM staged_employees se;
```

In this example, Hive determines the values of the partition keys, country and state, from the last two columns in the SELECT clause. Different names in staged\_employees emphasize that the relationship between the source column values and the output partition values is **by position only** and not by matching on names.

# Mix Dynamic and Static Partitions

Static value for the country (**US**) and a dynamic value for the state. The static partition keys must come before the dynamic partition keys.

```
INSERT OVERWRITE TABLE employees
PARTITION (country = 'US', state)
SELECT ..., se.cnty, se.st
FROM staged_employees se
WHERE se.cnty = 'US';
```

# Dynamic Partitions Properties

Name	Default	Description
hive.exec.dynamic.partition	false	Set to true to enable dynamic partitioning.
hive.exec.dynamic.partition.mode	strict	Set to nonstrict to enable all partitions to be determined dynamically.
hive.exec.max.dynamic.partitions.pernode	100	The maximum number of dynamic partitions that can be created by each mapper or reducer. Raises a fatal error if one mapper or reducer attempts to create more than the threshold.
hive.exec.max.dynamic.partitions	+1000	The total number of dynamic partitions that can be created by one statement with dynamic partitioning. Raises a fatal error if the limit is exceeded.
hive.exec.max.created.files	100000	The maximum total number of files that can be created globally. A Hadoop counter is used to track the number of files created. Raises a fatal error if the limit is exceeded.

# Set the Desired Properties Just Before Use

```
hive> set hive.exec.dynamic.partition=true;  
hive> set hive.exec.dynamic.partition.mode=nonstrict;  
hive> set hive.exec.max.dynamic.partitions.pernode=1000;
```

```
hive> INSERT OVERWRITE TABLE employees  
    > PARTITION (country, state)  
    > SELECT ..., se.cty, se.st  
    > FROM staged_employees se;
```

# Create Tables and Load Them in One Query

```
CREATE TABLE ca_employees  
AS SELECT name, salary, address  
FROM employees  
WHERE state = 'CA';
```

- A common use for this feature is to extract a convenient subset of data from a larger table.
- This feature can't be used with external tables.

# Export Data

**INSERT OVERWRITE LOCAL DIRECTORY '/tmp/ca\_employees'**

```
SELECT name, salary, address
FROM employees
WHERE state = 'CA';
```

Independent of how the data is actually stored in the source table, it is written to files with all fields serialized as strings.

```
hive> ! ls /tmp/ca_employees;
```

```
000000_0
```

```
hive> ! cat /tmp/ca_employees/000000_0
```

```
John Doe100000.0201 San Antonio CircleMountain ViewCA94040
```

```
Mary Smith80000.01 Infinity LoopCupertinoCA95014
```

...