

Hive Lab

(base) [bigdata@localhost ~]\$ **start-all.sh**

This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh

Starting namenodes on [localhost]

localhost: starting namenode, logging to /home/bigdata/hadoop-2.7.3/logs/hadoop-bigdata-namenode-localhost.localdomain.out

localhost: starting datanode, logging to /home/bigdata/hadoop-2.7.3/logs/hadoop-bigdata-datanode-localhost.localdomain.out

Starting secondary namenodes [0.0.0.0]

0.0.0.0: starting secondarynamenode, logging to /home/bigdata/hadoop-2.7.3/logs/hadoop-bigdata-secondarynamenode-localhost.localdomain.out

starting yarn daemons

starting resourcemanager, logging to /home/bigdata/hadoop-2.7.3/logs/yarn-bigdata-resourcemanager-localhost.localdomain.out

localhost: starting nodemanager, logging to /home/bigdata/hadoop-2.7.3/logs/yarn-bigdata-nodemanager-localhost.localdomain.out

(base) [bigdata@localhost ~]\$ **hive**

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/home/bigdata/hive/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/home/bigdata/hadoop-2.7.3/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/bigdata/hive/lib/hive-common-2.1.0.jar!/hive-log4j2.properties Async: true

Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. tez, spark) or using Hive 1.X releases.

hive> **show databases;**

OK

db1

default

Time taken: 1.248 seconds, Fetched: 2 row(s)

hive> **create database employee;**

OK

Time taken: 0.639 seconds

hive> **use employee;**

OK

Time taken: 0.01 seconds

hive> **create database nti;**

OK

Time taken: 0.037 seconds

hive> **show databases;**

OK

```

db1
default
employee
nti
Time taken: 0.012 seconds, Fetched: 4 row(s)
hive> use nti;
OK
Time taken: 0.01 seconds
hive> describe database nti;
OK
nti          hdfs://localhost:9000/user/hive/warehouse/nti.db      bigdataUSER
Time taken: 0.011 seconds, Fetched: 1 row(s)
hive> create table employee(ID int, name string, salary float, Age int);
OK
Time taken: 1.361 seconds
hive> SHOW TABLES IN nti;
OK
employee
Time taken: 0.038 seconds, Fetched: 1 row(s)
hive> show tables;
OK
employee
Time taken: 0.013 seconds, Fetched: 1 row(s)
hive> describe employee;
OK
id           int
name         string
salary       float
age          int
Time taken: 0.754 seconds, Fetched: 4 row(s)
hive> insert into employee values (1, 'Marwa', 1000, 30);
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider
using a different execution engine (i.e. tez, spark) or using Hive 1.X releases.
Query ID = bigdata_20210922124927_c71f9b87-8899-4875-aaf2-fd333d450d51
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_1632306522146_0001, Tracking URL =
http://localhost:8088/proxy/application_1632306522146_0001/
Kill Command = /home/bigdata/hadoop-2.7.3/bin/hadoop job -kill job_1632306522146_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2021-09-22 12:49:49,063 Stage-1 map = 0%, reduce = 0%
2021-09-22 12:49:56,823 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.65 sec
MapReduce Total cumulative CPU time: 1 seconds 650 msec

```

Ended Job = job_1632306522146_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 directory hdfs://localhost:9000/user/hive/warehouse/nti.db/employee/.hive-staging_hive_2021-09-22_12-49-27_649_668752906316954909-1/-ext-10000

Loading data to table nti.employee

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 1.65 sec HDFS Read: 4659 HDFS Write: 86 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 650 msec

OK

Time taken: 30.684 seconds

hive> **insert into employee values (2, 'Jamila', 1000, 24);**

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. tez, spark) or using Hive 1.X releases.

Query ID = bigdata_20210922125026_7263bf27-266d-4f2d-9a03-4eaf9b8d475d

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_1632306522146_0002, Tracking URL =

http://localhost:8088/proxy/application_1632306522146_0002/

Kill Command = /home/bigdata/hadoop-2.7.3/bin/hadoop job -kill job_1632306522146_0002

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2021-09-22 12:50:33,766 Stage-1 map = 0%, reduce = 0%

2021-09-22 12:50:40,122 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.55 sec

MapReduce Total cumulative CPU time: 1 seconds 550 msec

Ended Job = job_1632306522146_0002

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 directory hdfs://localhost:9000/user/hive/warehouse/nti.db/employee/.hive-staging_hive_2021-09-22_12-50-26_865_8979185472376487462-1/-ext-10000

Loading data to table nti.employee

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 1.55 sec HDFS Read: 4663 HDFS Write: 87 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 550 msec

OK

Time taken: 15.652 seconds

hive> **insert into employee values (3, 'Mazen', 3000, 40);**

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. tez, spark) or using Hive 1.X releases.

Query ID = bigdata_20210922125102_19bb20b1-456b-46b6-acd5-f6f2917c2c49

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_1632306522146_0003, Tracking URL =
http://localhost:8088/proxy/application_1632306522146_0003/
Kill Command = /home/bigdata/hadoop-2.7.3/bin/hadoop job -kill job_1632306522146_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2021-09-22 12:51:09,740 Stage-1 map = 0%, reduce = 0%
2021-09-22 12:51:17,227 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.45 sec
MapReduce Total cumulative CPU time: 1 seconds 450 msec
Ended Job = job_1632306522146_0003
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 directory hdfs://localhost:9000/user/hive/warehouse/nti.db/employee/.hive-
staging_hive_2021-09-22_12-51-02_339_2893455336244887926-1/-ext-10000
Loading data to table nti.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 1.45 sec HDFS Read: 4662 HDFS Write: 86 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 450 msec
OK
Time taken: 17.275 seconds
hive> insert into employee values (4, 'Ahmed', 4000, 40);
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider
using a different execution engine (i.e. tez, spark) or using Hive 1.X releases.
Query ID = bigdata_20210922125135_1c100e59-464d-4329-85c8-998e83653838
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_1632306522146_0004, Tracking URL =
http://localhost:8088/proxy/application_1632306522146_0004/
Kill Command = /home/bigdata/hadoop-2.7.3/bin/hadoop job -kill job_1632306522146_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2021-09-22 12:51:42,210 Stage-1 map = 0%, reduce = 0%
2021-09-22 12:51:48,507 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.37 sec
MapReduce Total cumulative CPU time: 1 seconds 370 msec
Ended Job = job_1632306522146_0004
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 directory hdfs://localhost:9000/user/hive/warehouse/nti.db/employee/.hive-
staging_hive_2021-09-22_12-51-35_397_8211404414197429749-1/-ext-10000
Loading data to table nti.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 1.37 sec HDFS Read: 4662 HDFS Write: 86 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 370 msec

OK

Time taken: 14.434 seconds

hive> **insert into employee values (5, 'Samir', 1000, 30);**

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. tez, spark) or using Hive 1.X releases.

Query ID = bigdata_20210922125207_5e556736-9205-4a99-925b-f424519c718b

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_1632306522146_0005, Tracking URL =

http://localhost:8088/proxy/application_1632306522146_0005/

Kill Command = /home/bigdata/hadoop-2.7.3/bin/hadoop job -kill job_1632306522146_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2021-09-22 12:52:13,888 Stage-1 map = 0%, reduce = 0%

2021-09-22 12:52:20,244 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.29 sec

MapReduce Total cumulative CPU time: 1 seconds 290 msec

Ended Job = job_1632306522146_0005

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 directory hdfs://localhost:9000/user/hive/warehouse/nti.db/employee/.hive-staging_hive_2021-09-22_12-52-07_108_2248375321562286092-1/-ext-10000

Loading data to table nti.employee

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 1.29 sec HDFS Read: 4662 HDFS Write: 86 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 290 msec

OK

Time taken: 14.391 seconds

hive> **select * from employee;**

OK

1 Marwa 1000.0 30

2 Jamila 1000.0 24

3 Mazen 3000.0 40

4 Ahmed 4000.0 40

5 Samir 1000.0 30

Time taken: 0.105 seconds, Fetched: 5 row(s)

hive> **select age from employee;**

OK

30

24

40

40

30

Time taken: 0.093 seconds, Fetched: 5 row(s)

```
hive> select name , age from employee;
```

```
OK
```

```
Marwa 30
```

```
Jamila 24
```

```
Mazen 40
```

```
Ahmed 40
```

```
Samir 30
```

```
Time taken: 0.085 seconds, Fetched: 5 row(s)
```

```
hive> create table trail(ID int, name string, salary float, Age int);
```

```
OK
```

```
Time taken: 0.058 seconds
```

```
hive> alter table employee change salary salary double;
```

```
OK
```

```
Time taken: 0.159 seconds
```

```
hive> describe employee;
```

```
OK
```

```
id          int
```

```
name        string
```

```
salary      double
```

```
age         int
```

```
Time taken: 0.036 seconds, Fetched: 4 row(s)
```

```
hive> describe formatted employee;
```

```
OK
```

```
# col_name          data_type          comment
```

```
id          int
```

```
name        string
```

```
salary      double
```

```
age         int
```

Detailed Table Information

```
Database:          nti
```

```
Owner:            bigdata
```

```
CreateTime:       Wed Sep 22 12:35:28 EET 2021
```

```
LastAccessTime:   UNKNOWN
```

```
Retention:        0
```

```
Location:         hdfs://localhost:9000/user/hive/warehouse/nti.db/employee
```

```
Table Type:       MANAGED_TABLE
```

```
Table Parameters:
```

```
    last_modified_by    bigdata
```

```
    last_modified_time  1632308207
```

```
    numFiles            5
```

```
    numRows             5
```

```
    rawDataSize         86
```

totalSize 91
transient_lastDdlTime 1632308207

Storage Information

SerDe Library: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat: org.apache.hadoop.mapred.TextInputFormat
OutputFormat: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed: No
Num Buckets: -1
Bucket Columns: []
Sort Columns: []
Storage Desc Params:
 serialization.format 1

Time taken: 1.337 seconds, Fetched: 34 row(s)

```
hive> CREATE TABLE airports(  
  > name STRING,  
  > country STRING,  
  > area_code INT,  
  > code STRING)  
  > ROW FORMAT DELIMITED  
  > FIELDS TERMINATED BY ',';
```

OK

Time taken: 0.067 seconds

```
hive> create external table emp_ext (ID int, name string,salary float,age int)  
  > ROW FORMAT DELIMITED  
  > FIELDS TERMINATED BY ','  
  >;
```

OK

Time taken: 0.061 seconds

```
hive> describe formatted emp_ext;
```

OK

# col_name	data_type	comment
id	int	
name	string	
salary	float	
age	int	

Detailed Table Information

Database: nti
Owner: bigdata
CreateTime: Wed Sep 22 13:07:38 EET 2021
LastAccessTime: UNKNOWN
Retention: 0

Location: hdfs://localhost:9000/user/hive/warehouse/nti.db/emp_ext

Table Type: EXTERNAL_TABLE

Table Parameters:

COLUMN_STATS_ACCURATE	{\"BASIC_STATS\": \"true\"}
EXTERNAL	TRUE
numFiles	0
numRows	0
rawDataSize	0
totalSize	0
transient_lastDdlTime	1632308858

Storage Information

SerDe Library: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

InputFormat: org.apache.hadoop.mapred.TextInputFormat

OutputFormat: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat

Compressed: No

Num Buckets: -1

Bucket Columns: []

Sort Columns: []

Storage Desc Params:

field.delim	,
serialization.format	,

Time taken: 0.039 seconds, Fetched: 35 row(s)

HDFS

(base) [bigdata@localhost ~]\$ **hdfs dfs -put file:///home/bigdata/Desktop/airports.csv /user/hive**

(base) [bigdata@localhost ~]\$ **hdfs dfs -mkdir /user/hive/external**

(base) [bigdata@localhost ~]\$ **hdfs dfs -put file:///home/bigdata/Desktop/data1.txt /user/hive/external**