

Experiment – 10

Demonstrate the concept of Partitioning & Bucketing using HiveQL.

```
hive> show databases;
OK
default
hiveql
rjcstudent
rsk
Time taken: 0.89 seconds, Fetched: 4 row(s)
hive> use rjcstudent;
OK
Time taken: 0.049 seconds
hive> show tables;
OK
student
Time taken: 0.048 seconds, Fetched: 1 row(s)
hive> describe student;
OK
id                int
name              string
course            string
age               int
Time taken: 0.289 seconds, Fetched: 4 row(s)
hive> create table course_part1(ID int,Name string,Age int) PARTITIONED BY(course string)
> ;
OK
Time taken: 0.287 seconds

hive> create table course_part1(ID int,Name string,Age int) PARTITIONED BY(course string)
> ;
OK
Time taken: 0.287 seconds

hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> insert overwrite table course_part1 PARTITION(course) SELECT ID,Name,Age, Course from student;
Query ID = cloudera_20210628205454_a445ef34-3e3e-4252-89c1-62650a533114
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_1624936639194_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1624936639194_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1624936639194_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2021-06-28 20:54:56 Stage-1 map = 0%, reduce = 0%
2021-06-28 20:55:11,092 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.84 sec
MapReduce Total cumulative CPU time: 1 seconds 840 msec
Ended Job = job_1624936639194_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://quickstart.cloudera:8020/user/hive/warehouse/rjcstudent.db/course_part1/.hive-staging_hive_2021-06-28_20-54-32_557_1402664720961742401-1/-ext-10000
Loading data to table rjcstudent.course_part1 partition (course=null)
Time taken for load dynamic partitions : 943
Loading partition {course=Hadoop}
Loading partition {course=Java}
Loading partition {course=Python}
Time taken for adding to write entity : 12
Partition rjcstudent.course_part1{course=Hadoop} stats: [numFiles=1, numRows=2, totalSize=20, rawDataSize=18]
Partition rjcstudent.course_part1{course=Java} stats: [numFiles=1, numRows=2, totalSize=22, rawDataSize=20]
Partition rjcstudent.course_part1{course=Python} stats: [numFiles=1, numRows=2, totalSize=21, rawDataSize=19]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 1.84 sec HDFS Read: 3840 HDFS Write: 266 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 840 msec
OK
Time taken: 42.233 seconds
```

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Hive Editor - Query - Mozilla Firefox

quickstart.cloudera:8888/beanswax/#query

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Hive Editor Query Editor My Queries Saved Queries History

Assist Settings

DATABASE rjstudent

Table name...

course_part1

- id (int)
- name (string)
- age (int)
- course (string)

student

Example: `SELECT * FROM tablename, or press CTRL + space`

Execute Save as... Explain or create a New query

Recent queries Query Log Columns Results Chart

Time	Query	Result
06/22/2021 8:08:40 PM	<code>show databases;</code>	See results...
06/22/2021 8:08:07 PM	<code>show databases;</code>	See results...
06/22/2021 8:02:56 PM	<code>show databases;</code>	See results...
06/22/2021 8:01:57 PM	<code>show databases;</code>	See results...
06/22/2021 8:01:35 PM	<code>show databases;</code>	See results...

[cloudera@quickstart:~] Hue - Hive Editor - Qu...

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Metastore Manager - Table : course_part1 - Mozilla Firefox

quickstart.cloudera:8888/metastore/table/rjstudent/course_part1

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Metastore Manager

ACTIONS

- Import Data
- Browse Data
- Drop Table
- View File Location
- Show Partitions (3)

Databases > rjstudent > course_part1

Columns	Partition Columns	Sample	Properties
Name	Type	Comment	
0 id	int		
1 name	string		
2 age	int		
3 course	string		

[cloudera@quickstart:~] Hue - Metastore Mana...

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Metastore Manager - Table : course_part1 - Mozilla Firefox

quickstart.cloudera:8888/metastore/table/rjstudent/course_part1

Cloudera Hue HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Metastore Manager

ACTIONS

- Import Data
- Browse Data
- Drop Table
- View File Location
- Show Partitions (3)

Databases > rjstudent > course_part1

Columns Partition Columns Sample Properties

	Name	Type	Comment
0	course	string	null

[cloudera@quickstart:~]\$ Hue - Metastore Mana...

Type here to search

09:28 AM 29-06-2021

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Metastore Manager - Table : course_part1 - Mozilla Firefox

quickstart.cloudera:8888/metastore/table/rjstudent/course_part1

Cloudera Hue HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Metastore Manager

ACTIONS

- Import Data
- Browse Data
- Drop Table
- View File Location
- Show Partitions (3)

Databases > rjstudent > course_part1

Columns Partition Columns Sample Properties

	Id	name	age	course
0	3	Harry	23	Python
1	5	Jhon	26	Python

[cloudera@quickstart:~]\$ Hue - Metastore Mana...

Type here to search

09:28 AM 29-06-2021

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Metastore Manager - Table : course_part1 - Mozilla Firefox

quickstart.cloudera:8888/metastore/table/rjstudent/course_part1

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Metastore Manager

ACTIONS

- Import Data
- Browse Data
- Drop Table
- View File Location
- Show Partitions (3)

Databases > rjstudent > course_part1

- Columns
- Partition Columns
- Sample
- Properties

Name	Value
tableName	course_part1
dbName	rjstudent
owner	cloudera
createTime	1624938577
lastAccessTime	0
retention	0
location:hdfs://quickstart.cloudera	8020userhive/warehouse/rjstudent.db/course_part1
inputFormat	org.apache.hadoop.mapred.TextInputFormat
outputFormat	org.apache.hadoop.hive.q.io.HiveIgnoreKeyTextOutputFormat
compressed	false
numBuckets	-1
serdeInfo:SerDeInfo:name	null
serializationLib	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
parameters	{serialization.format=1}
bucketCols	[]
sortCols	[]
parameters	[]
skewedInfo:SkewedInfo:skewedColNames	[]

cloudera@quickstart:~\$ Hue - Metastore Mana...

Type here to search

09:28 AM 29-06-2021

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Hive Editor - Query - Mozilla Firefox

quickstart.cloudera:8888/teeswax/#query

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Hive Editor Query Editor My Queries Saved Queries History

Data sample for course_part1

[View in Metastore Browser](#)

	id	name	age	course
0	3	Harry	23	Python
1	5	Jhon	26	Python

OK

Recent queries

Time	Query	Result
06/22/2021 8:08:40 PM	show databases;	See results...
06/22/2021 8:08:07 PM	show databases;	See results...
06/22/2021 8:02:56 PM	show databases;	See results...
06/22/2021 8:01:57 PM	show databases;	See results...
06/22/2021 8:01:35 PM	show databases;	See results...

cloudera@quickstart:~\$ Hue - Hive Editor - Qu...

Type here to search

09:29 AM 29-06-2021

```
hive> set hive.enforce.bucketing=true;
```

```

hive> create table stu_bucket(ID int,Name string,Course string,Age int)
> clustered by(ID) into 2 buckets
> row format delimited
> fields terminated by ',';

OK
Time taken: 0.145 seconds

hive> insert overwrite table stu_bucket select * from student;
Query ID = cloudera_20210628211818_05459fdc-5036-4139-8e99-932970dc5d20
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 2
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624936639194_0002, Tracking URL = http://quickstart.cloud
era:8088/proxy/application_1624936639194_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1624936639194_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 2
2021-06-28 21:19:19,654 Stage-1 map = 0%, reduce = 0%
2021-06-28 21:19:58,193 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.74 sec
2021-06-28 21:20:24,279 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 3.8
sec
2021-06-28 21:20:28,435 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.
59 sec
MapReduce Total cumulative CPU time: 6 seconds 590 msec
Ended Job = job_1624936639194_0002
Loading data to table rjcstudent.stu_bucket
Table rjcstudent.stu_bucket stats: [numFiles=2, numRows=6, totalSize=101, raw
DataSize=95]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 2   Cumulative CPU: 6.59 sec   HDFS Read: 9775
HDFS Write: 255 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 590 msec
OK
Time taken: 99.403 seconds
hive>

```

The screenshot shows the Cloudera Hue interface running in a virtual machine. The main window displays a data sample for the 'stu_bucket' table. The data is presented in a table with columns: id, name, course, and age. Below the table, there is a list of recent queries with timestamps and links to view results or databases.

id	name	course	age
0 6	Stephen	Java	22
1 4	Jerry	Hadoop	25
2 2	Tom	Java	23
3 5	Jhon	Python	26
4 3	Harry	Python	23
5 1	sam	Hadoop	24

Recent queries:

- 06/22/2021 8:00:07 PM [show databases](#) [View results...](#)
- 06/22/2021 8:02:56 PM [show databases](#) [View results...](#)
- 06/22/2021 8:03:57 PM [show databases](#) [View results...](#)
- 06/22/2021 8:01:35 PM [show databases](#) [View results...](#)

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Metastore Manager - Table : stu_bucket - Mozilla Firefox

Mon Jun 28, 9:21 PM cloudera

Hue - Hive Editor - Query Hue - Metastore Man...

quickstart.cloudera:8888/metastore/table/rjstudent/stu_bucket

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Metastore Manager

ACTIONS

- Import Data
- Browse Data
- Drop Table
- View File Location

Databases > rjstudent > stu_bucket

Columns Sample Properties

	Name	Type	Comment
0	id	int	
1	name	string	
2	course	string	
3	age	int	

cloudera@quickstart:~ Hive2 (~/.Desktop) - g... Hue - Metastore Mana...

Type here to search

09:51 AM 29-06-2021

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - File Browser - Mozilla Firefox

Mon Jun 28, 9:21 PM cloudera

Hue - Hive Editor - Query Hue - File Browser

quickstart.cloudera:8888/filebrowser/view/user/hive/warehouse/rjstudent.db/stu_bucket

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

File Browser

Search for file name Actions Move to trash Upload New

Home / user / hive / warehouse / rjstudent.db / stu_bucket History Trash

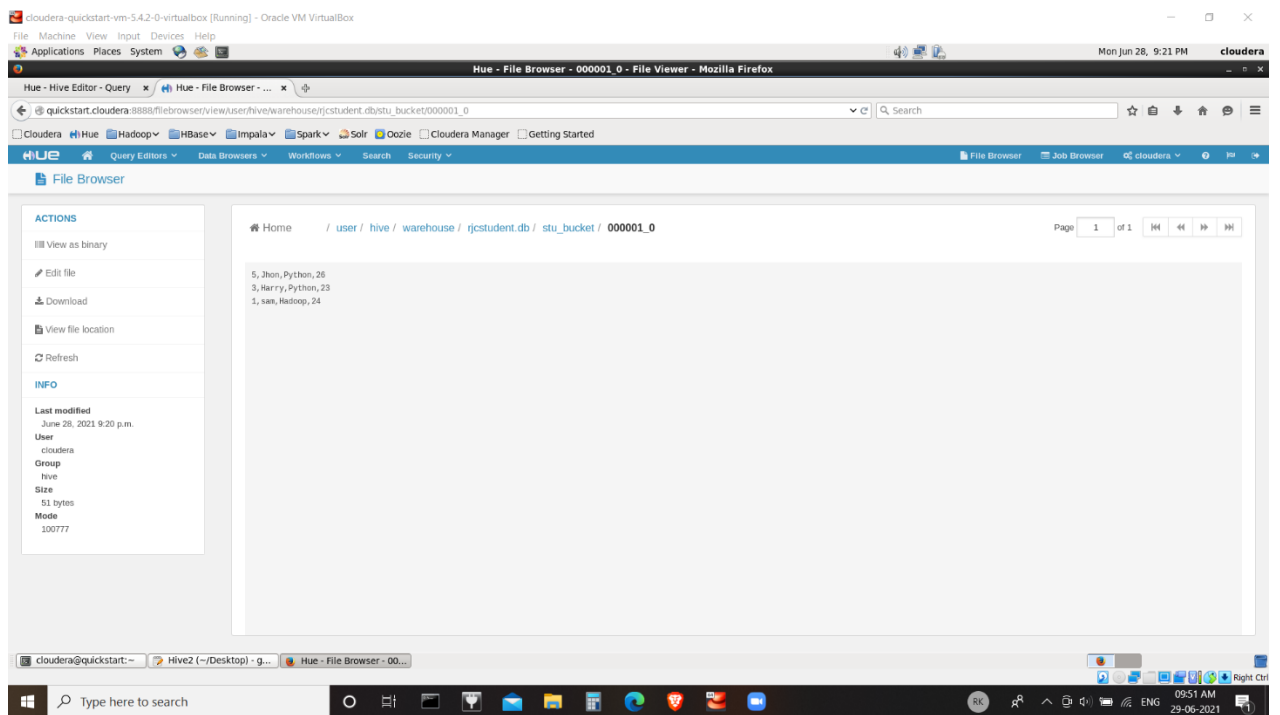
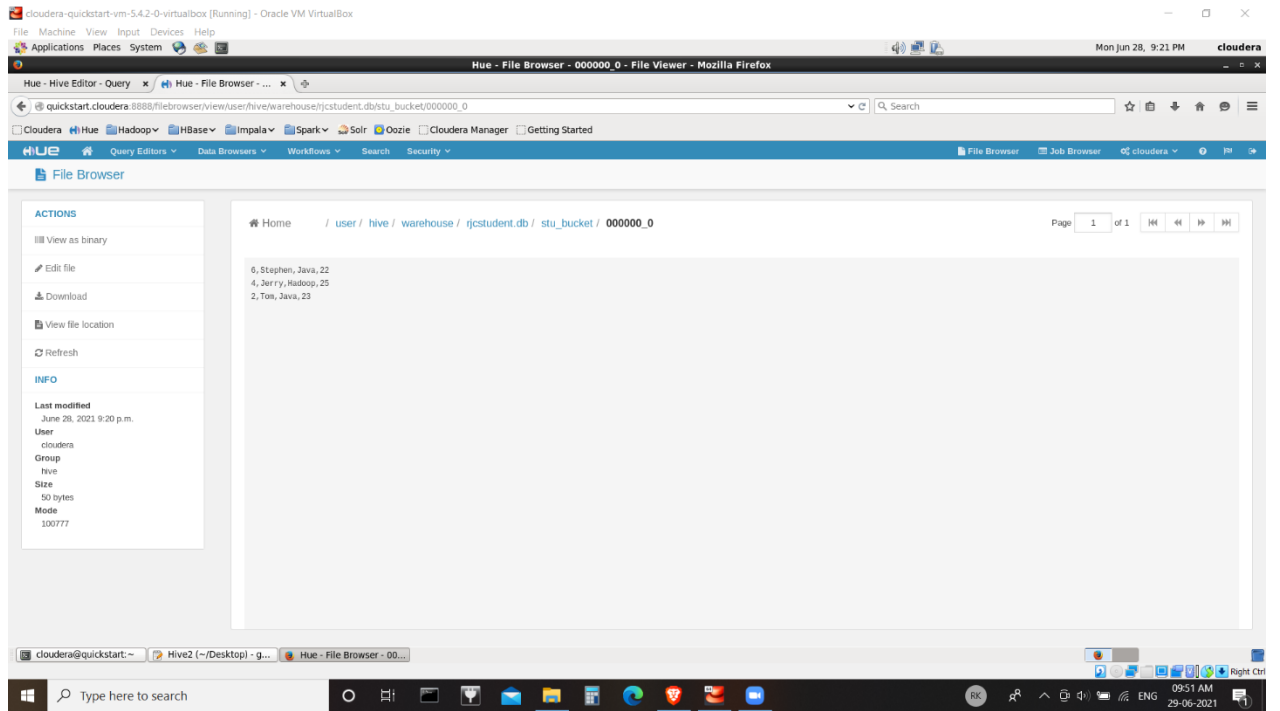
Name	Size	User	Group	Permissions	Date
3		cloudera	hive	drwxrwxrwx	June 28, 2021 09:13 PM
.		cloudera	hive	drwxrwxrwx	June 28, 2021 09:20 PM
0000000_0	50 bytes	cloudera	hive	-rw-rw-rw-	June 28, 2021 09:20 PM
0000001_0	51 bytes	cloudera	hive	-rw-rw-rw-	June 28, 2021 09:20 PM

Show 45 of 2 items Page 1 of 1

cloudera@quickstart:~ Hive2 (~/.Desktop) - g... Hue - File Browser - M...

Type here to search

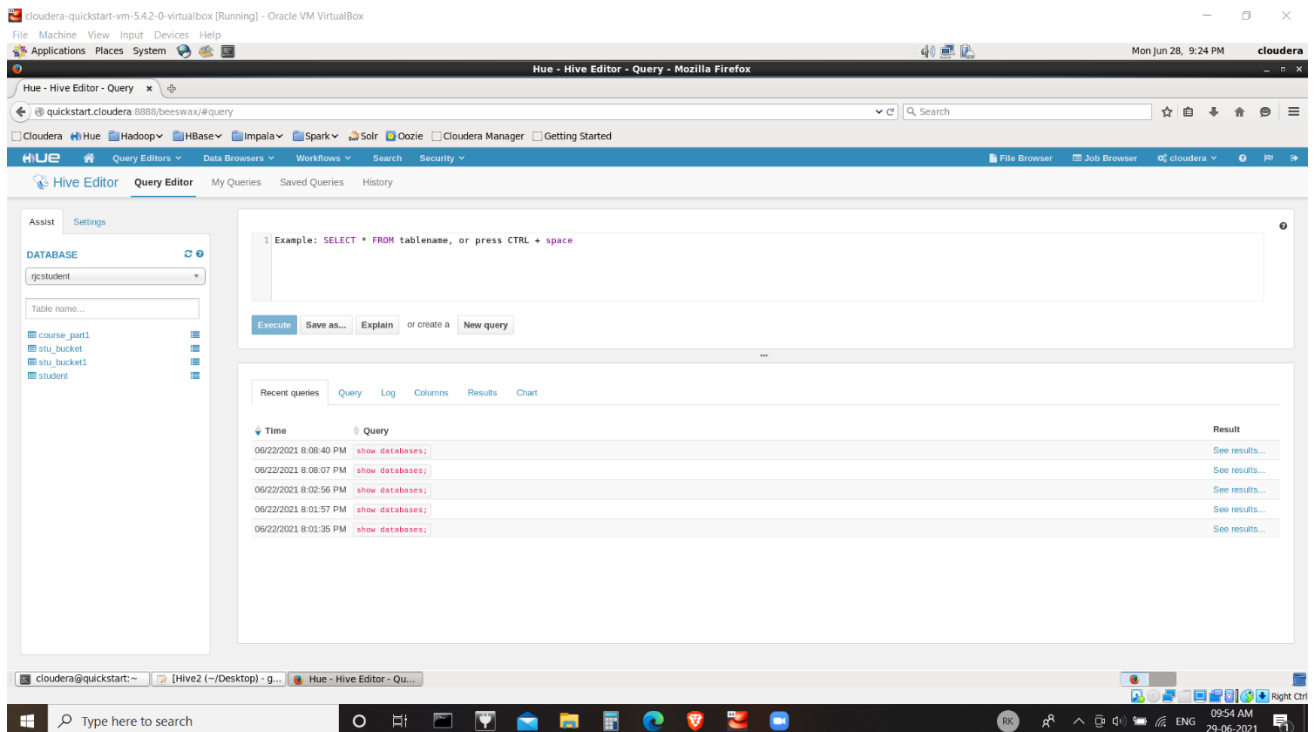
09:51 AM 29-06-2021



```

hive> create table stu_bucket1(ID int,Name string,Course string,Age int)
> clustered by(ID) into 3 buckets
> row format delimited
> fields terminated by ',';
OK
Time taken: 0.119 seconds
hive> insert overwrite table stu_bucket1 select * from student;
Query ID = cloudera_20210628212222_3da99886-73e3-4871-bb6e-fbcce1ef9769
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624936639194_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1624936639194_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1624936639194_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 3
2021-06-28 21:22:44,399 Stage-1 map = 0%, reduce = 0%
2021-06-28 21:22:55,549 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.29 sec
2021-06-28 21:23:27,652 Stage-1 map = 100%, reduce = 44%, Cumulative CPU 3.87 sec
2021-06-28 21:23:30,155 Stage-1 map = 100%, reduce = 72%, Cumulative CPU 5.52 sec
2021-06-28 21:23:31,474 Stage-1 map = 100%, reduce = 78%, Cumulative CPU 6.03 sec
2021-06-28 21:23:32,545 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.75 sec
MapReduce Total cumulative CPU time: 7 seconds 750 msec
Ended Job = job_1624936639194_0003
Loading data to table rjcstudent.stu_bucket1
Table rjcstudent.stu_bucket1 stats: [numFiles=3, numRows=6, totalSize=101, rawDataSize=95]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 3 Cumulative CPU: 7.75 sec HDFS Read: 13014 HDFS Write: 335 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 750 msec
OK
Time taken: 65.187 seconds

```



cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Hive Editor - Query - Mozilla Firefox

Mon Jun 28, 9:24 PM cloudera

Hue - Hive Editor - Query

quickstart.cloudera.8888/beanswax/#query

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Hive Editor Query Editor My Queries Saved Queries History

Data sample for stu_bucket1

[View in Metastore Browser](#)

	id	name	course	age
0	6	Stephen	Java	22
1	3	Harry	Python	23
2	4	Jerry	Hadoop	25
3	1	sam	Hadoop	24
4	5	Jhon	Python	26
5	2	Tom	Java	23

06/22/2021 9:08:07 PM show databases; See results...

06/22/2021 9:02:56 PM show databases; See results...

06/22/2021 9:01:57 PM show databases; See results...

06/22/2021 9:01:35 PM show databases; See results...

cloudera@quickstart: ~ Hive2 (~/Desktop) - g... Hue - Hive Editor - Qu...

Type here to search

09:54 AM 29-06-2021

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - File Browser - Mozilla Firefox

Mon Jun 28, 9:24 PM cloudera

Hue - Hive Editor - Query Hue - File Browser

quickstart.cloudera.8888/filebrowser/view/user/hive/warehouse/jcstudent.db/stu_bucket1

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

File Browser

Search for file name Actions Move to trash Upload New

Home / user / hive / warehouse / jcstudent.db / stu_bucket1 History Trash

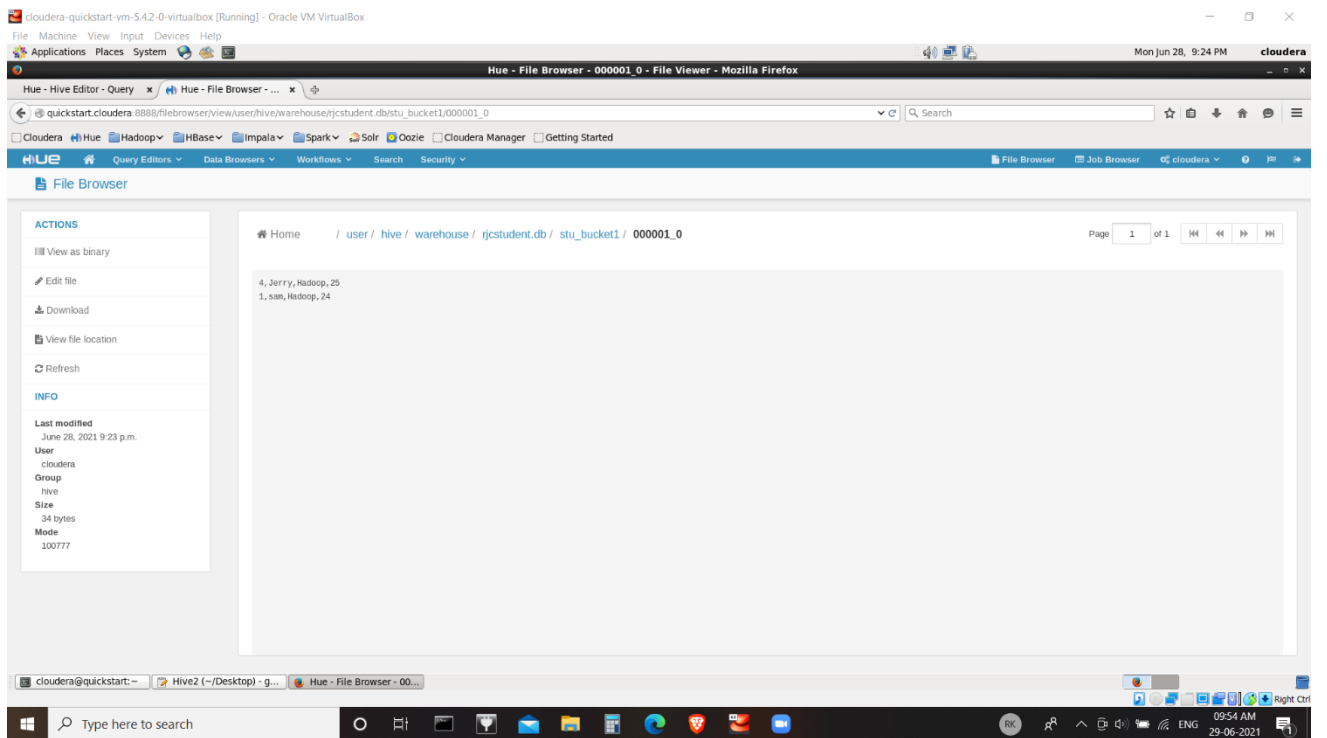
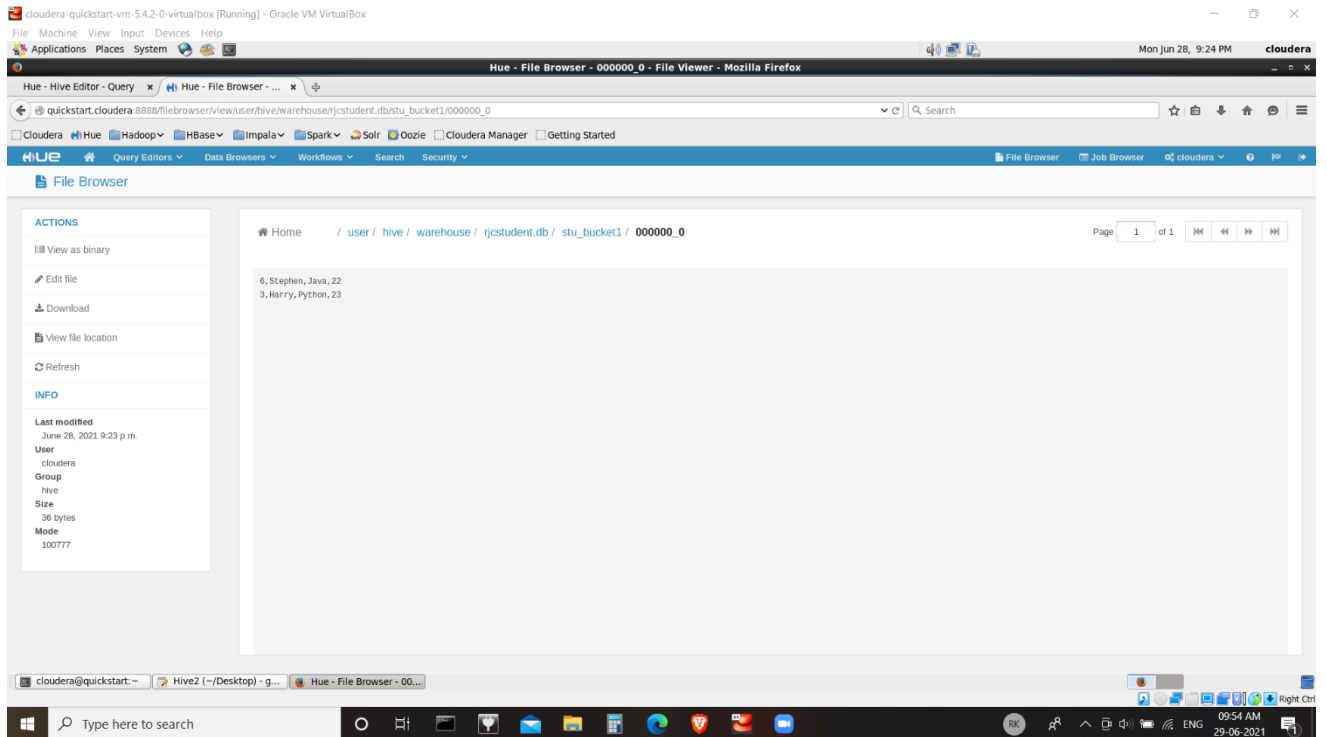
Name	Size	User	Group	Permissions	Date
.		cloudera	hive	drwxrwxrwx	June 28, 2021 09:22 PM
000000_0	36 bytes	cloudera	hive	rw-rw-rw-	June 28, 2021 09:23 PM
000001_0	34 bytes	cloudera	hive	rw-rw-rw-	June 28, 2021 09:23 PM
000002_0	31 bytes	cloudera	hive	rw-rw-rw-	June 28, 2021 09:23 PM

Show 45 of 3 items Page 1 of 1

cloudera@quickstart: ~ Hive2 (~/Desktop) - M... Hue - File Browser - M...

Type here to search

09:54 AM 29-06-2021



cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - File Browser - 000002_0 - File Viewer - Mozilla Firefox

Mon Jun 28, 9:25 PM cloudera

Hue - Hive Editor - Query x Hue - File Browser - ... x

quickstart.cloudera:8888/filebrowser/view/user/hive/warehouse/rjcstudent.db/stu_bucket1/000002_0

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Query Editors Data Browsers Workflows Search Security

File Browser Job Browser cloudera

File Browser

ACTIONS

- View as binary
- Edit file
- Download
- View file location
- Refresh

INFO

Last modified
June 28, 2021 9:23 p.m.

User
cloudera

Group
hive

Size
31 bytes

Mode
100777

Home / user / hive / warehouse / rjcstudent.db / stu_bucket1 / 000002_0

Page 1 of 1

```
5, Zhou, Python, 26
2, Tom, Java, 23
```

cloudera@quickstart:~ Hive2 (~/.Desktop) - g... Hue - File Browser - 00...

Type here to search

09:55 AM 29-06-2021

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Hue - Metastore Manager - Table : stu_bucket1 - Mozilla Firefox

Mon Jun 28, 9:25 PM cloudera

Hue - Hive Editor - Query x Hue - Metastore Man... x

quickstart.cloudera:8888/metastore/table/rjcstudent/stu_bucket1

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Query Editors Data Browsers Workflows Search Security

File Browser Job Browser cloudera

Metastore Manager

ACTIONS

- Import Data
- Browse Data
- Drop Table
- View File Location

Databases > rjcstudent > stu_bucket1

Columns Sample Properties

	Name	Type	Comment
0	id	int	
1	name	string	
2	course	string	
3	age	int	

quickstart.cloudera:8888/filebrowser/view/user/hive/warehouse/rjcstudent.db/stu_bucket1

cloudera@quickstart:~ Hive2 (~/.Desktop) - g... Hue - Metastore Mana...

Type here to search

09:55 AM 29-06-2021