

Creating Lookup Table

Command to create the Lookup Table

<Command>

```
CREATE TABLE lookup_data_hbase
(
    `card_id`          STRING,
    `ucl`              DOUBLE,
    `score`            INT,
    `postcode`         STRING,
    `transaction_dt`   TIMESTAMP
) stored BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH
serdeproperties ("hbase.columns.mapping"=
":key, lookup_card_family:ucl,
lookup_card_family:score, lookup_transaction_family:postcode, lookup_transact
ion_family:transaction_dt"
) tblproperties ("hbase.table.name" = "lookup_data_hive");
```

<Command to see the table created>

<Command>

```
describe 'lookup_data_hive'
```

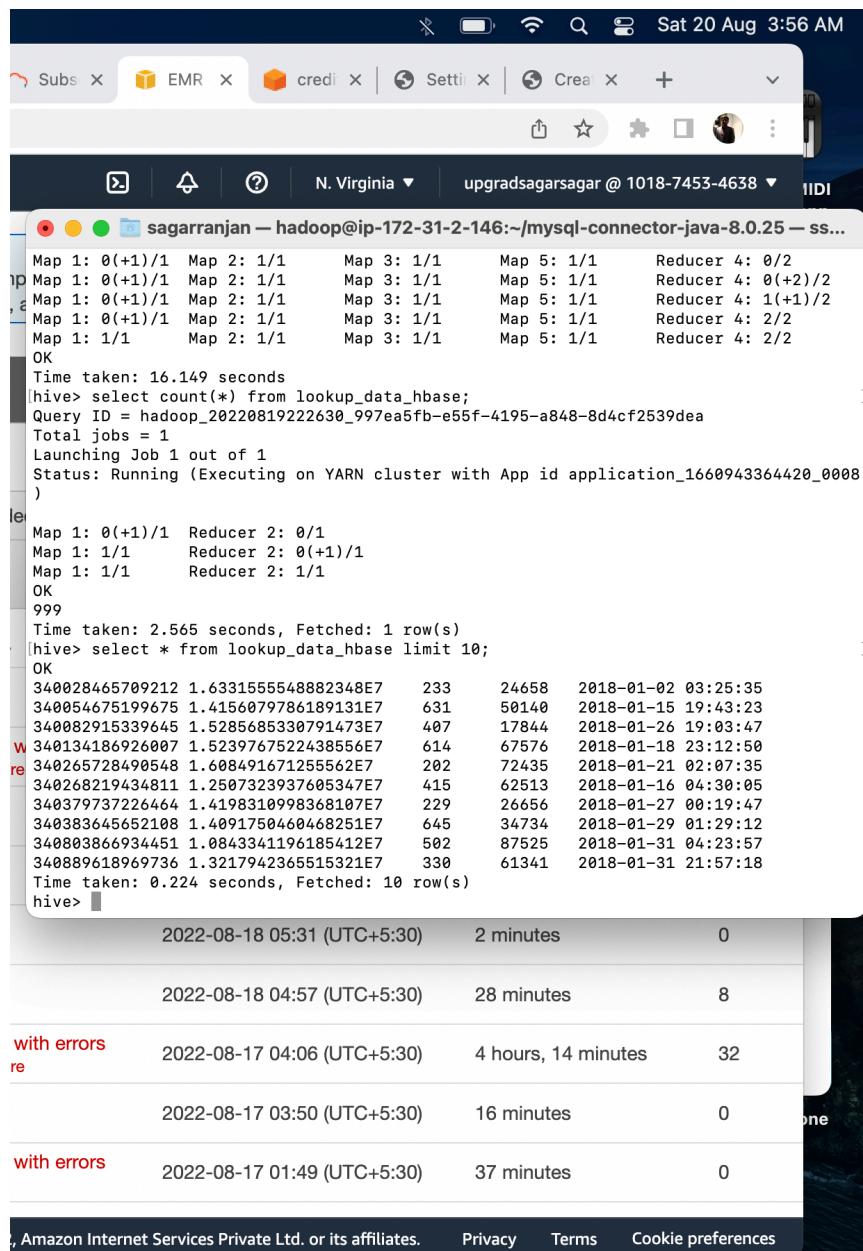
```
COUNT 'lookup_data_hive'
```

```
scan 'lookup_data_hive'
```

```
SELECT Count(*) FROM lookup_data_hbase;
```

Screenshot of the created table

<Screenshot of the table created>

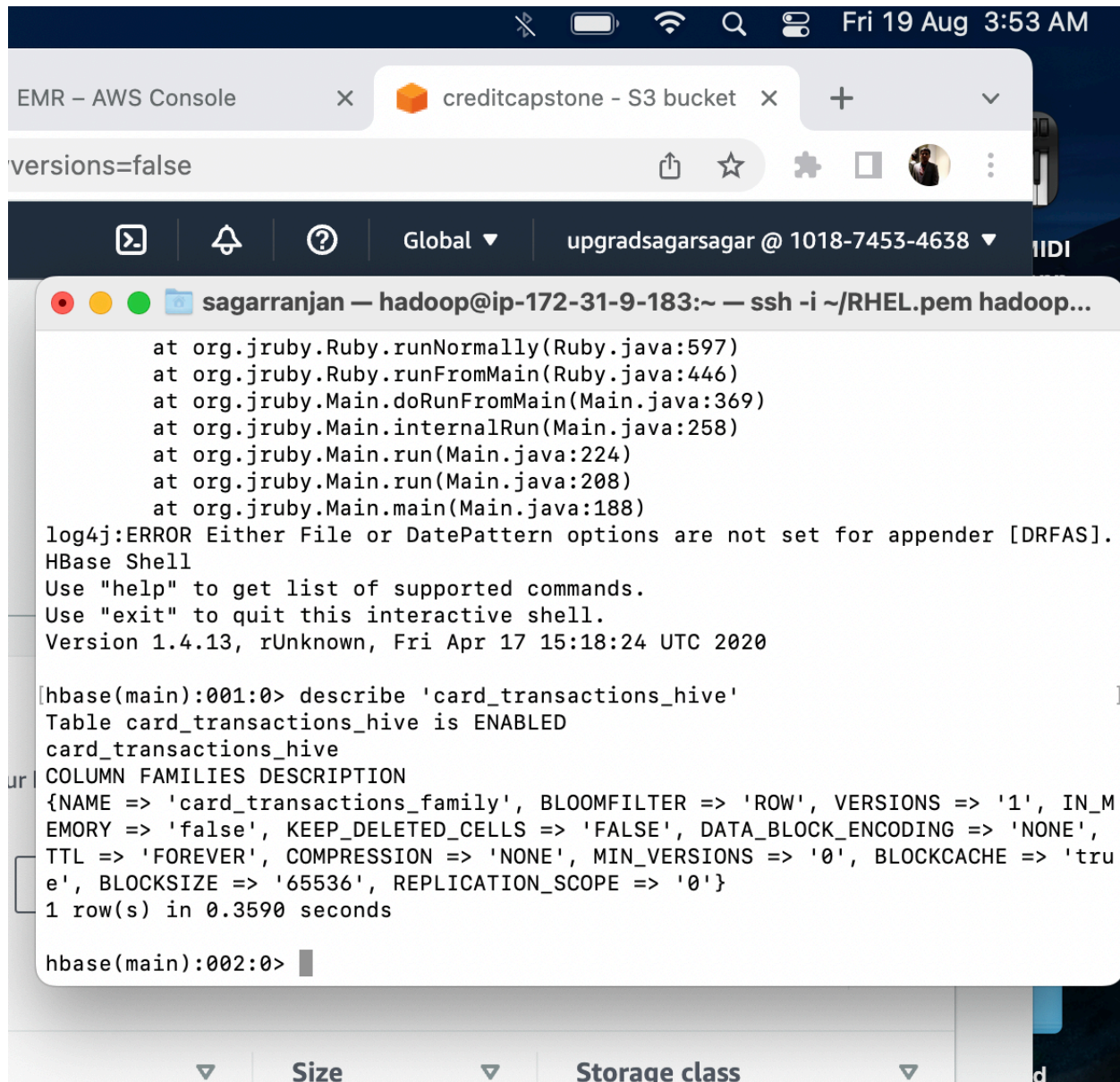


The screenshot shows a terminal window with the following content:

```
sagarranjan — hadoop@ip-172-31-2-146:~/mysql-connector-java-8.0.25 — ss...
Map 1: 0(+1)/1 Map 2: 1/1 Map 3: 1/1 Map 5: 1/1 Reducer 4: 0/2
Map 1: 0(+1)/1 Map 2: 1/1 Map 3: 1/1 Map 5: 1/1 Reducer 4: 0(+2)/2
Map 1: 0(+1)/1 Map 2: 1/1 Map 3: 1/1 Map 5: 1/1 Reducer 4: 1(+1)/2
Map 1: 0(+1)/1 Map 2: 1/1 Map 3: 1/1 Map 5: 1/1 Reducer 4: 2/2
Map 1: 1/1 Map 2: 1/1 Map 3: 1/1 Map 5: 1/1 Reducer 4: 2/2
OK
Time taken: 16.149 seconds
[hive> select count(*) from lookup_data_hbase;
Query ID = hadoop_20220819222630_997ea5fb-e55f-4195-a848-8d4cf2539dea
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1660943364420_0008)
Map 1: 0(+1)/1 Reducer 2: 0/1
Map 1: 1/1 Reducer 2: 0(+1)/1
Map 1: 1/1 Reducer 2: 1/1
OK
999
Time taken: 2.565 seconds, Fetched: 1 row(s)
[hive> select * from lookup_data_hbase limit 10;
OK
340028465709212 1.6331555548882348E7 233 24658 2018-01-02 03:25:35
340054675199675 1.4156079786189131E7 631 50140 2018-01-15 19:43:23
340082915339645 1.5285685330791473E7 407 17844 2018-01-26 19:03:47
340134186926007 1.5239767522438556E7 614 67576 2018-01-18 23:12:50
340265728490548 1.608491671255562E7 202 72435 2018-01-21 02:07:35
340268219434811 1.2507323937605347E7 415 62513 2018-01-16 04:30:05
340379737226464 1.4198310998368107E7 229 26656 2018-01-27 00:19:47
340383645652108 1.4091750460468251E7 645 34734 2018-01-29 01:29:12
340803866934451 1.0843341196185412E7 502 87525 2018-01-31 04:23:57
340889618969736 1.3217942365515321E7 330 61341 2018-01-31 21:57:18
Time taken: 0.224 seconds, Fetched: 10 row(s)
hive>
```

	2022-08-18 05:31 (UTC+5:30)	2 minutes	0
	2022-08-18 04:57 (UTC+5:30)	28 minutes	8
with errors	2022-08-17 04:06 (UTC+5:30)	4 hours, 14 minutes	32
	2022-08-17 03:50 (UTC+5:30)	16 minutes	0
with errors	2022-08-17 01:49 (UTC+5:30)	37 minutes	0

Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences



The screenshot shows a mobile browser interface. At the top, the status bar displays 'Fri 19 Aug 3:53 AM'. The browser has two tabs: 'EMR - AWS Console' and 'creditcapstone - S3 bucket'. The address bar shows 'versions=false'. Below the browser, a terminal window is open with the title 'sagarranjan — hadoop@ip-172-31-9-183:~ — ssh -i ~/RHEL.pem hadoop...'. The terminal output shows Ruby stack traces, an HBase error message, and the execution of the 'describe' command for the 'card_transactions_hive' table.

```

at org.jruby.Ruby.runNormally(Ruby.java:597)
at org.jruby.Ruby.runFromMain(Ruby.java:446)
at org.jruby.Main.doRunFromMain(Main.java:369)
at org.jruby.Main.internalRun(Main.java:258)
at org.jruby.Main.run(Main.java:224)
at org.jruby.Main.run(Main.java:208)
at org.jruby.Main.main(Main.java:188)
log4j:ERROR Either File or DatePattern options are not set for appender [DRFAS].
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.13, rUnknown, Fri Apr 17 15:18:24 UTC 2020

[hbase(main):001:0> describe 'card_transactions_hive'
Table card_transactions_hive is ENABLED
card_transactions_hive
COLUMN FAMILIES DESCRIPTION
{NAME => 'card_transactions_family', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_M
EMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE',
TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'tru
e', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
1 row(s) in 0.3590 seconds

[hbase(main):002:0> ]

```

Below the terminal window, a table header is visible with columns: Size, Storage class.

```

Sat 20 Aug 3:18 AM
Subs x EMR x credi x | Setti x | Crea x + v
sagarranjan — hadoop@ip-172-31-2-146:~/mysql-connector-java-8.0.25 — ss...

FILE: Number of write operations=0
HDFS: Number of bytes read=87
HDFS: Number of bytes written=19980
HDFS: Number of read operations=4
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=174384
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=3633
  Total vcore-milliseconds taken by all map tasks=3633
  Total megabyte-milliseconds taken by all map tasks=5580288
Map-Reduce Framework
  Map input records=999
  Map output records=999
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=86
  CPU time spent (ms)=2230
  Physical memory (bytes) snapshot=318828544
  Virtual memory (bytes) snapshot=3279949824
  Total committed heap usage (bytes)=330301440
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=19980
22/08/19 21:48:16 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 18.2798 seconds (1.0674 KB/sec)
22/08/19 21:48:16 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-2-146 mysql-connector-java-8.0.25]$

```