



APACHE HBASE

NOSQL OVER HADOOP LABS

Bin Jiang

04/08/2017

Apache HBase Lab

- **Apache HBase Lab Instructions**

- TrainingOnHDP\docs\Class 7 - NoSQL over Hadoop\lab\Apache HBase Labs.txt

- **Apache HBase Lab Code**

- TrainingOnHDP\HBaseCookBook
- TrainingOnHDP\BehaviorTrackingOnHBase

HBase Essentials Lab

- Start HBase Services

The screenshot displays the Ambari Sandbox interface. The top navigation bar includes the Ambari logo, 'Sandbox' label, '0 ops' and '0 alerts' indicators, and tabs for 'Dashboard', 'Services', 'Hosts', 'Alerts', and 'Admin'. A user dropdown menu shows 'admin'. The left sidebar lists various services: HDFS, YARN, MapReduce2, Tez, Hive, HBase (selected), Pig, Sqoop, Oozie, ZooKeeper, Falcon, Storm, Flume, Ambari Infra, Atlas, Kafka, Knox, Ranger, and Spark2. The main content area is titled 'Summary' and includes tabs for 'Heatmaps' and 'Configs'. A 'Quick Links' dropdown and a 'Service Actions' dropdown are also present. The 'Summary' section shows the 'Active HBase Master' as 'Started' with 'No alerts'. It also indicates '1/1 RegionServers Live' and 'Regions In Transition: 0'. Key metrics include 'Master Started: 1.47 hours', 'Master Activated: 1.47 hours', 'Average Load: 9 regions per RegionServer', and 'Master Heap: 52.5 MB / 1015.7 MB (5.2% used)'. Below the summary is a 'Metrics' section with a 'Last 1 hour' filter. It contains ten sub-sections: 'Reads and Writes', 'Read Latency', 'Write Latency', 'Open Connections', 'Request Handlers', 'Files Local', 'Blocked Updates', 'Cluster CPU', 'Cluster Network', and 'Cluster Disk'. Most of these sections display 'No Data No available data for the time period.', while 'Files Local' shows 'n/a'.

Ambari Sandbox 0 ops 0 alerts Dashboard Services Hosts Alerts Admin admin

Summary Heatmaps Configs Quick Links Service Actions

Summary No alerts

[Active HBase Master](#) Started No alerts Master Started 1.47 hours
[RegionServers](#) 1/1 RegionServers Live Master Activated 1.47 hours
Regions In Transition 0 Average Load 9 regions per RegionServer
Master Heap 52.5 MB / 1015.7 MB (5.2% used)

Metrics Last 1 hour

Reads and Writes	Read Latency	Write Latency	Open Connections	Request Handlers
No Data No available data for the time period.	No Data No available data for the time period.	No Data No available data for the time period.	No Data No available data for the time period.	No Data No available data for the time period.

Files Local	Blocked Updates	Cluster CPU	Cluster Network	Cluster Disk
n/a	No Data No available data for the time period.	No Data No available data for the time period.	No Data No available data for the time period.	No Data No available data for the time period.

HBase Essentials Lab

- Getting to Know the HBase Shell

```
hbase(main):001:0> help
HBase Shell, version 1.1.2.6.3.0-235, r6f982c8e667e7a18451fcbc12c5e2758b178ec78, Mon Oct 30 02:53:33 UTC 2017
Type 'help "COMMAND"', (e.g. 'help "get"' -- the quotes are necessary) for help on a specific command.
Commands are grouped. Type 'help "COMMAND_GROUP"', (e.g. 'help "general"') for help on a command group.

COMMAND GROUPS:

  Group name: general
  Commands: status, table_help, version, whoami

  Group name: ddl
  Commands: alter, alter_async, alter_status, create, describe, disable, disable_all, drop, drop_all, enable, enable_all, exists, get_table, is_disabled, is_enabled, list, locate_region, show_filters

  Group name: namespace
  Commands: alter_namespace, create_namespace, describe_namespace, drop_namespace, list_namespace, list_namespace_tables

  Group name: dml
  Commands: append, count, delete, deleteall, get, get_counter, get_splits, incr, put, scan, truncate, truncate_preserve

  Group name: tools
  Commands: assign, balance_switch, balancer, balancer_enabled, catalogjanitor_enabled, catalogjanitor_run, catalogjanitor_switch, close_region, compact, compact_rs, flush, major_compact, merge_region, move, normalize, normalizer_enabled, normalizer_switch, split, splitormerge_enabled, splitormerge_switch, trace, unassign, wal_roll, zk_dump

  Group name: replication
  Commands: add_peer, append_peer_tableCFs, disable_peer, disable_table_replication, enable_peer, enable_table_replication, list_peers, list_replicated_tables, remove_peer, remove_peer_tableCFs, set_peer_tableCFs, show_peer_tableCFs

  Group name: snapshots
  Commands: clone_snapshot, delete_all_snapshot, delete_snapshot, list_snapshots, restore_snapshot, snapshot, snapshot_all, snapshot_restore

  Group name: configuration
  Commands: update_all_config, update_config
```

HBase Essentials Lab

- Getting to Know the HBase Shell

```
Group name: quotas
Commands: list_quota_snapshots, list_quota_table_sizes, list_quotas, set_quota
```

```
Group name: security
Commands: grant, revoke, user_permission
```

```
Group name: procedures
Commands: abort_procedure, list_procedures
```

```
Group name: visibility labels
Commands: add_labels, clear_auths, get_auths, list_labels, set_auths, set_visibility
```

```
Group name: rsgroup
Commands: add_rsgroup, balance_rsgroup, get_rsgroup, get_server_rsgroup, get_table_rsgroup, list_rsgroups, move_rsgroup_servers, move_rsgroup_tables, remove_rsgroup
```

SHELL USAGE:

Quote all names in HBase Shell such as table and column names. Commas delimit command parameters. Type <RETURN> after entering a command to run it. Dictionaries of configuration used in the creation and alteration of tables are Ruby Hashes. They look like this:

```
{'key1' => 'value1', 'key2' => 'value2', ...}
```

and are opened and closed with curly-braces. Key/values are delimited by the '=' character combination. Usually keys are predefined constants such as NAME, VERSIONS, COMPRESSION, etc. Constants do not need to be quoted. Type 'Object.constants' to see a (messy) list of all constants in the environment.

If you are using binary keys or values and need to enter them in the shell, use double-quoted hexadecimal representation. For example:

```
hbase> get 't1', "key\x03\x3f\xcd"
hbase> get 't1', "key\003\023\011"
```

HBase Essentials Lab

- Getting to Know the HBase Shell

```
hbase(main):004:0> version
1.1.2.2.6.3.0-235, r6f982c8e667e7a18451fcbc12c5e2758b178ec78, Mon Oct 30 02:53:33 UTC 2017

hbase(main):005:0> status
1 active master, 0 backup masters, 1 servers, 0 dead, 8.0000 average load

hbase(main):006:0> create 'mytable', 'columnFamilyA', {NAME => 'columnFamilyB', VERSIONS => 3}
0 row(s) in 2.6230 seconds

=> Hbase::Table - mytable
hbase(main):007:0> list
TABLE
ATLAS_ENTITY_AUDIT_EVENTS
atlas_titan
iemployee
mytable
stocks
5 row(s) in 0.0910 seconds

=> ["ATLAS_ENTITY_AUDIT_EVENTS", "atlas_titan", "iemployee", "mytable", "stocks"]
hbase(main):008:0> describe 'mytable'
Table mytable is ENABLED
mytable
COLUMN FAMILIES DESCRIPTION
{NAME => 'columnFamilyA', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE',
TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
{NAME => 'columnFamilyB', BLOOMFILTER => 'ROW', VERSIONS => '3', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE',
TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
2 row(s) in 1.0240 seconds
```

HBase Essentials Lab

- **Getting to Know the HBase Shell**

```
hbase(main):009:0> put 'mytable', 'rowkey1', 'columnFamilyA:a', 'qwerty'
0 row(s) in 0.5340 seconds

hbase(main):010:0> put 'mytable', 'rowkey2', 'columnFamilyA:a', 'asdfg'
0 row(s) in 0.1030 seconds

hbase(main):011:0> scan 'mytable'
ROW                                COLUMN+CELL
 rowkey1                          column=columnFamilyA:a, timestamp=1514715798125, value=qwerty
 rowkey2                          column=columnFamilyA:a, timestamp=1514715817685, value=asdfg
2 row(s) in 0.4320 seconds

hbase(main):012:0> █
```


- **Working with the Cell Version History**

[illegible]

HBase Essentials Lab

- Incrementing Value

```
hbase(main):022:0> incr 'mytable', 'rowkey1', 'columnFamilyA:c', 0
COUNTER VALUE = 0
0 row(s) in 0.0310 seconds

hbase(main):023:0> incr 'mytable', 'rowkey1', 'columnFamilyA:c', 2
COUNTER VALUE = 2
0 row(s) in 0.0800 seconds

hbase(main):024:0> get 'mytable', 'rowkey1', 'columnFamilyA:c'
COLUMN                                CELL
columnFamilyA:c                       timestamp=1514716460886, value=\x00\x00\x00\x00\x00\x00\x02
1 row(s) in 0.0950 seconds

hbase(main):025:0> incr 'mytable', 'rowkey1', 'columnFamilyA:c'
COUNTER VALUE = 3
0 row(s) in 0.2040 seconds

hbase(main):026:0> |
```

HBase Essentials Lab

- Capturing Shell Output

```
hbase(main):026:0> uit
[root@sandbox-hdp ~]# echo "scan 'mytable'" | hbase shell > scan_results.data
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
[root@sandbox-hdp ~]# cat scan_results.data
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.1.2.2.6.3.0-235, r6f982c8e667e7a18451fcbc12c5e2758b178ec78, Mon Oct 30 02:53:33 UTC 2017

scan 'mytable'
ROW COLUMN+CELL
 rowkey1 column=columnFamilyA:a, timestamp=1514715798125, value=qwerty
 rowkey1 column=columnFamilyA:c, timestamp=1514716492204, value=\x00\x00\x00\x00\x00\x00\x00\x03
 rowkey2 column=columnFamilyA:a, timestamp=1514715817685, value=asdfg
2 row(s) in 2.8110 seconds

[root@sandbox-hdp ~]# hbase shell
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.1.2.2.6.3.0-235, r6f982c8e667e7a18451fcbc12c5e2758b178ec78, Mon Oct 30 02:53:33 UTC 2017

hbase(main):001:0> █
```

HBase Essentials Lab

- **Configuring the Time to Live (TTL) value**

```
hbase(main):003:0> create 'ephimeral', {NAME => 'cf', TTL => '20'}
0 row(s) in 2.7660 seconds

=> Hbase::Table - ephimeral
hbase(main):004:0> put 'ephimeral', 'r1', 'cf:20sec', 'hide-and-seek'
0 row(s) in 0.6490 seconds

hbase(main):005:0> get 'ephimeral', 'r1', 'cf:20sec'
COLUMN                                CELL
cf:20sec                               timestamp=1514717296139, value=hide-and-seek
1 row(s) in 0.1260 seconds

hbase(main):006:0> get 'ephimeral', 'r1', 'cf:20sec'
COLUMN                                CELL
0 row(s) in 0.0770 seconds

hbase(main):007:0> disable 'ephimeral'
0 row(s) in 2.6640 seconds

hbase(main):008:0> drop 'ephimeral'
0 row(s) in 2.6560 seconds

hbase(main):009:0> █
```

HBase Essentials Lab

- Scanning with Limits

```
hbase(main):009:0> truncate 'mytable'
Truncating 'mytable' table (it may take a while):
- Disabling table...
- Truncating table...
0 row(s) in 7.0460 seconds

hbase(main):010:0> put 'mytable', 'r1', 'columnFamilyA:c1', 'r1c1'
0 row(s) in 0.3090 seconds

hbase(main):011:0> put 'mytable', 'r1', 'columnFamilyA:c2', 'r1c2'
0 row(s) in 0.0360 seconds

hbase(main):012:0> put 'mytable', 'r2', 'columnFamilyA:c1', 'r1c1'
0 row(s) in 0.0270 seconds

hbase(main):013:0> put 'mytable', 'r2', 'columnFamilyA:c2', 'r1c2'
0 row(s) in 0.0830 seconds

hbase(main):014:0> scan 'mytable'
ROW                                COLUMN+CELL
r1                                 column=columnFamilyA:c1, timestamp=1514717469486, value=r1c1
r1                                 column=columnFamilyA:c2, timestamp=1514717480530, value=r1c2
r2                                 column=columnFamilyA:c1, timestamp=1514717492878, value=r1c1
r2                                 column=columnFamilyA:c2, timestamp=1514717502904, value=r1c2
2 row(s) in 0.2600 seconds

hbase(main):015:0> █
```

HBase Essentials Lab

- Scanning with Limits

```
hbase(main):015:0> scan 'mytable', {COLUMNS => 'columnFamilyA:c1'}
ROW          COLUMN+CELL
  r1          column=columnFamilyA:c1, timestamp=1514717469486, value=r1c1
  r2          column=columnFamilyA:c1, timestamp=1514717492878, value=r1c1
2 row(s) in 0.0370 seconds

hbase(main):016:0> scan 'mytable', {STARTROW => 'r2'}
ROW          COLUMN+CELL
  r2          column=columnFamilyA:c1, timestamp=1514717492878, value=r1c1
  r2          column=columnFamilyA:c2, timestamp=1514717502904, value=r1c2
1 row(s) in 0.2220 seconds

hbase(main):017:0> scan 'mytable', {FILTER => "(PrefixFilter ('r2'))"}
ROW          COLUMN+CELL
  r2          column=columnFamilyA:c1, timestamp=1514717492878, value=r1c1
  r2          column=columnFamilyA:c2, timestamp=1514717502904, value=r1c2
1 row(s) in 0.1180 seconds

hbase(main):018:0> scan 'mytable', {TIMERANGE => [1514717480530, 1514717492878]}
ROW          COLUMN+CELL
  r1          column=columnFamilyA:c2, timestamp=1514717480530, value=r1c2
1 row(s) in 0.0320 seconds

hbase(main):019:0> scan 'mytable', {TIMERANGE => [1514717480530, 1514717492879]}
ROW          COLUMN+CELL
  r1          column=columnFamilyA:c2, timestamp=1514717480530, value=r1c2
  r2          column=columnFamilyA:c1, timestamp=1514717492878, value=r1c1
2 row(s) in 0.1220 seconds

hbase(main):020:0> □
```

HBase Essentials Lab

- Export Data from HBase Table into Hive Table

```
[root@sandbox-hdp ~]# hive
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender.

Logging initialized using configuration in file:/etc/hive/2.6.3.0-235/0/hive-log4j.properties

    > show databases;
OK
default
foodmart
foodmart_spark
oozie
xademo
Time taken: 14.695 seconds, Fetched: 5 row(s)
hive> create database hbase;
OK
Time taken: 0.759 seconds
hive> use hbase;
OK
Time taken: 0.461 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS mytable_hbase (rowkey STRING, c1 STRING, c2 STRING) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = ':key,columnFamilyA:c1,columnFamilyA:c2') TBLPROPERTIES ('hbase.table.name' = 'mytable');
OK
Time taken: 6.759 seconds
```

HBase Essentials Lab

- Export Data from HBase Table into Hive Table

```
hive> show databases;
OK
default
foodmart
foodmart_spark
hbase
oozie
xademo
Time taken: 0.044 seconds, Fetched: 6 row(s)
hive> SELECT * FROM mytable_hbase;
OK
r1      r1c1    r1c2
r2      r1c1    r1c2
Time taken: 3.827 seconds, Fetched: 2 row(s)
hive> INSERT INTO TABLE mytable_hbase values ("r3", "r3c1", "r3c2");
Query ID = root_20171231113408_4cc1aaf2-5658-482b-9384-9a0d6da0ee1a
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1514669965561_0079)
```

VERTICES	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 33.65 s
OK
Time taken: 89.268 seconds
hive> 
```


HBase Essentials Lab

- Export Data from HBase Table into Hive Table

```
hbase(main):020:0> scan 'mytable'
ROW                                COLUMN+CELL
r1                                column=columnFamilyA:c1, timestamp=1514717469486, value=r1c1
r1                                column=columnFamilyA:c2, timestamp=1514717480530, value=r1c2
r2                                column=columnFamilyA:c1, timestamp=1514717492878, value=r1c1
r2                                column=columnFamilyA:c2, timestamp=1514717502904, value=r1c2
r3                                column=columnFamilyA:c1, timestamp=1514720137243, value=r3c1
r3                                column=columnFamilyA:c2, timestamp=1514720137243, value=r3c2
3 row(s) in 0.1550 seconds

hbase(main):021:0> put 'mytable', 'r4', 'columnFamilyA:c1', 'r4c1'
0 row(s) in 0.0640 seconds

hbase(main):022:0> put 'mytable', 'r5', 'columnFamilyA:c2', 'r5c2'
0 row(s) in 0.0140 seconds

hbase(main):023:0> 
```

HBase Essentials Lab

- Export Data from HBase Table into Hive Table

```
hive> SELECT * FROM mytable_hbase;
OK
r1      r1c1    r1c2
r2      r1c1    r1c2
r3      r3c1    r3c2
r4      r4c1    NULL
r5      NULL    r5c2
Time taken: 1.113 seconds, Fetched: 5 row(s)
hive> CREATE TABLE IF NOT EXISTS mytable_hbase_1 LIKE mytable_hbase ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;
OK
Time taken: 0.725 seconds
hive> INSERT INTO TABLE mytable_hbase_1 SELECT * FROM mytable_hbase;
Query ID = root_20171231114003_f8a34eee-bc92-4534-8d87-0473791a1250
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1514669965561_0079)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01 [=====] 100% ELAPSED TIME: 40.11 s
-----
Loading data to table hbase.mytable_hbase_1
Table hbase.mytable_hbase_1 stats: [numFiles=1, numRows=5, totalSize=61, rawDataSize=56]
OK
Time taken: 49.182 seconds
hive>
```

HBase Essentials Lab

- Import Data into HBase Table from Hive Table

```
hive> CREATE TABLE IF NOT EXISTS mytable_hbase_2 (rowkey STRING, c1 STRING, c2 STRING) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = ':key,columnFamilyA:c1,columnFamilyA:c2') TBLPROPERTIES ('hbase.table.name' = 'mytable_1');
OK
Time taken: 4.371 seconds
hive> INSERT INTO TABLE mytable_hbase_2 SELECT * FROM mytable_hbase_1;
Query ID = root_20171231114454_0189cf10-1290-4fc9-9041-1cfb2d8b4ac4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1514669965561_0079)

-----
      VERTICES      STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
-----
Map 1 ..... SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 23.93 s
-----
OK
Time taken: 29.353 seconds
hive> |
```

HBase Essentials Lab

- **Import Data into HBase Table from Hive Table**

```
hbase(main):023:0> scan 'mytable_1'
ROW                                COLUMN+CELL
r1                                column=columnFamilyA:c1, timestamp=1514720722892, value=r1c1
r1                                column=columnFamilyA:c2, timestamp=1514720722892, value=r1c2
r2                                column=columnFamilyA:c1, timestamp=1514720722892, value=r1c1
r2                                column=columnFamilyA:c2, timestamp=1514720722892, value=r1c2
r3                                column=columnFamilyA:c1, timestamp=1514720722892, value=r3c1
r3                                column=columnFamilyA:c2, timestamp=1514720722892, value=r3c2
r4                                column=columnFamilyA:c1, timestamp=1514720722892, value=r4c1
r5                                column=columnFamilyA:c2, timestamp=1514720722892, value=r5c2
5 row(s) in 0.1660 seconds
```

HBase Essentials Lab

- Import Data into HBase Table from RDBMS

```
sandbox login: root
root@sandbox.hortonworks.com's password:
Last login: Sun Dec 31 10:46:32 2017 from 127.0.0.1
[root@sandbox-hdp ~]# mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 993
Server version: 5.6.38-log MySQL Community Server (GPL)

Copyright (c) 2000, 2017, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show database
-> ;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use
near 'database' at line 1
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| foodmart_spark |
| hive |
| mysql |
| nifi |
| nifi_bak |
| oozie |
| performance_schema |
| ... |
```

HBase Essentials Lab

- **Import Data into HBase Table from RDBMS**

```
[root@sandbox-hdp ~]# sqoop import -Dmapreduce.map.java.opts=-Xmx1024m -Dmapreduce.map.memory.mb=1024 --connect jdbc:mysql://localhost/oozie --table
business --fetch-size 10 --username root -P --hbase-create-table --hbase-table business --column-family f1 --hbase-row-key id -m 1
Warning: /usr/hdp/2.6.3.0-235/accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/12/31 14:35:29 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.3.0-235
Enter password:
17/12/31 14:35:44 INFO manager.MySQLManager: Argument '--fetch-size 10' will probably get ignored by MySQL JDBC driver.
17/12/31 14:35:44 INFO tool.CodeGenTool: Beginning code generation
17/12/31 14:35:46 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `business` AS t LIMIT 1
17/12/31 14:35:47 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `business` AS t LIMIT 1
17/12/31 14:35:47 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/hdp/2.6.3.0-235/hadoop-mapreduce
```

HBase Essentials Lab

- Import Data into HBase Table from RDBMS

```
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=87
HDFS: Number of bytes written=0
HDFS: Number of read operations=1
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=6823745
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=1364749
  Total vcore-milliseconds taken by all map tasks=1364749
  Total megabyte-milliseconds taken by all map tasks=1397502976
Map-Reduce Framework
  Map input records=460942
  Map output records=460942
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=44747
  CPU time spent (ms)=363380
  Physical memory (bytes) snapshot=677965824
  Virtual memory (bytes) snapshot=2982776832
  Total committed heap usage (bytes)=701632512
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
7/12/31 15:00:27 INFO mapreduce.ImportJobBase: Transferred 0 bytes in 1,443.8578 seconds (0 bytes/sec)
7/12/31 15:00:27 INFO mapreduce.ImportJobBase: Retrieved 460942 records.
root@sandbox-hdp ~]#
```


HBase Essentials Lab

- Import Data into HBase Table from RDBMS

```
hbase(main):024:0> describe 'business'
Table business is ENABLED
business
COLUMN FAMILIES DESCRIPTION
{NAME => 'f1', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
1 row(s) in 0.1210 seconds

hbase(main):025:0> scan 'business', {'LIMIT' => 10}
ROW                                COLUMN+CELL
1                                  column=f1:business_account_number, timestamp=1514731102524, value=7594
1                                  column=f1:business_corridor, timestamp=1514731102524, value=
1                                  column=f1:business_end_date, timestamp=1514731102524, value=
1                                  column=f1:business_location, timestamp=1514731102524, value=1598 CARROLL AV\x0ASan Francisco, CA 94124\x0A(37.725018, -122.393442)
1                                  column=f1:business_start_date, timestamp=1514731102524, value=07/01/1987
1                                  column=f1:city, timestamp=1514731102524, value=San Francisco
1                                  column=f1:dba_name, timestamp=1514731102524, value=Devincenzi Trucking Inc
1                                  column=f1:lic_code, timestamp=1514731102524, value=
1                                  column=f1:lic_code_description, timestamp=1514731102524, value=
1                                  column=f1:location_end_date, timestamp=1514731102524, value=
1                                  column=f1:location_id, timestamp=1514731102524, value=0007594-01-001
```

HBase Essentials Lab

- Import Data into HBase Table from RDBMS

```
100003 column=f1:state, timestamp=1514731404991, value=CA
100003 column=f1:street_address, timestamp=1514731404991, value=300 Plymouth Ave
100003 column=f1:supervisor_district, timestamp=1514731404991, value=11
100003 column=f1:transient_occupancy_tax, timestamp=1514731404991, value=false
100004 column=f1:business_account_number, timestamp=1514731404991, value=934555
100004 column=f1:business_corridor, timestamp=1514731404991, value=
100004 column=f1:business_end_date, timestamp=1514731404991, value=
100004 column=f1:business_location, timestamp=1514731404991, value=300 PLYMOUTH AVE\x0ASan Francisco, CA 94112\x0A(
100004 .713198, -122.456063)
100004 column=f1:business_start_date, timestamp=1514731404991, value=01/23/2012
100004 column=f1:city, timestamp=1514731404991, value=San Francisco
100004 column=f1:dba_name, timestamp=1514731404991, value=Lady Ts Family Soul Food
100004 column=f1:lic_code, timestamp=1514731404991, value=
100004 column=f1:lic_code_description, timestamp=1514731404991, value=
100004 column=f1:location_end_date, timestamp=1514731404991, value=
100004 column=f1:location_id, timestamp=1514731404991, value=0934555-07-001
100004 column=f1:location_start_date, timestamp=1514731404991, value=01/23/2012
100004 column=f1:mail_address, timestamp=1514731404991, value=0000 Undeliverable Mail
100004 column=f1:mail_city, timestamp=1514731404991, value=San Francisco
100004 column=f1:mail_state, timestamp=1514731404991, value=CA
100004 column=f1:mail_zipcode, timestamp=1514731404991, value=99999
100004 column=f1:naics_code, timestamp=1514731404991, value=
100004 column=f1:naics_code_description, timestamp=1514731404991, value=
100004 column=f1:neighborhoods_analysis_boundaries, timestamp=1514731404991, value=Oceanview/Merced/Ingleside
100004 column=f1:parking_tax, timestamp=1514731404991, value=false
100004 column=f1:qwnership_name, timestamp=1514731404991, value=Colbert Tanya L
100004 column=f1:source_zipcode, timestamp=1514731404991, value=94112
100004 column=f1:state, timestamp=1514731404991, value=CA
100004 column=f1:street_address, timestamp=1514731404991, value=300 Plymouth Ave
100004 column=f1:supervisor_district, timestamp=1514731404991, value=11
100004 column=f1:transient_occupancy_tax, timestamp=1514731404991, value=false
0 row(s) in 1.7170 seconds
```

HBase Essentials Lab

- **Import Data into HBase Table from RDBMS**

```
hbase(main):026:0> count 'business', INTERVAL => 100000
Current count: 100000, row: 189999
Current count: 200000, row: 3085
Current count: 300000, row: 39850
Current count: 400000, row: 48850
460942 row(s) in 127.1990 seconds

=> 460942
hbase(main):027:0> |
```

HBase Essentials Lab

- Using ImportTsv to load txt to HBase

```
[root@sandbox-hdp ~]# hadoop fs -copyFromLocal /root/TrainingOnHDP/dataset/simple.txt /user/root/simple.txt
[root@sandbox-hdp ~]# cat /root/TrainingOnHDP/dataset/simple.txt
1,tom
2,sam
3,jerry
4,marry
5,john[root@sandbox-hdp ~]# hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator="," -Dimporttsv.columns=HBASE_ROW_KEY,cf simple1
root/simple.txt
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC t
hreads appropriately using -XX:ParallelGCThreads=N
2017-12-31 15:29:23,652 INFO [main] zookeeper.RecoverableZooKeeper: Process identifier=hconnection-0x3d1cfad4 connecting to ZooKeeper ensemble=sandb
ox-hdp.hortonworks.com:2181
2017-12-31 15:29:23,675 INFO [main] zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.6-235--1, built on 10/30/2017 02:26 GMT
2017-12-31 15:29:23,675 INFO [main] zookeeper.ZooKeeper: Client environment:host.name=sandbox-hdp.hortonworks.com
2017-12-31 15:29:23,675 INFO [main] zookeeper.ZooKeeper: Client environment:java.version=1.8.0_151
2017-12-31 15:29:23,675 INFO [main] zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corporation
Total megabyte milliseconds taken by all map tasks=3481730
Map-Reduce Framework
  Map input records=5
  Map output records=5
  Input split bytes=125
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=256
  CPU time spent (ms)=3180
  Physical memory (bytes) snapshot=134279168
  Virtual memory (bytes) snapshot=2103967744
  Total committed heap usage (bytes)=178454528
ImportTsv
  Bad Lines=0
File Input Format Counters
  Bytes Read=38
File Output Format Counters
  Bytes Written=0
```

HBase Essentials Lab

- Using ImportTsv to load txt to HBase

```
hbase(main):027:0> create 'simple1','cf'
0 row(s) in 2.5490 seconds

=> Hbase::Table - simple1
hbase(main):028:0> scan 'simple1'
ROW                                COLUMN+CELL
1                                  column=cf:, timestamp=1514734160884, value=tom
2                                  column=cf:, timestamp=1514734160884, value=sam
3                                  column=cf:, timestamp=1514734160884, value=jerry
4                                  column=cf:, timestamp=1514734160884, value=marry
5                                  column=cf:, timestamp=1514734160884, value=john
5 row(s) in 0.1200 seconds
```

HBase Essentials Lab

- Bulk Load Data From HDFS into HBase

```
[root@sandbox-hdp ~]# hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator="," -Dimporttsv.bulk.output=hfile_tmp5 -Dimporttsv.columns=HBASE_ROW_KEY,cf simple2 /user/root/simple.txt
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
2017-12-31 15:37:45,326 INFO [main] zookeeper.RecoverableZooKeeper: Process identifier=hconnection-0x3d1cfad4 connecting to ZooKeeper ensemble=sandbox-hdp.hortonworks.com:2181
2017-12-31 15:37:45,511 INFO [main] zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.6-235--1, built on 10/30/2017 02:26 GMT
2017-12-31 15:37:45,511 INFO [main] zookeeper.ZooKeeper: Client environment:host.name=sandbox-hdp.hortonworks.com
2017-12-31 15:37:45,511 INFO [main] zookeeper.ZooKeeper: Client environment:java.version=1.8.0_151
2017-12-31 15:37:45,511 INFO [main] zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corporation
2017-12-31 15:37:45,511 INFO [main] zookeeper.ZooKeeper: Client environment:java.home=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.151-1.b12.el6_9.x86_64/j
```

```
[root@sandbox-hdp ~]# hadoop fs -ls hfile*
Found 2 items
-rw-r--r-- 1 root hdfs 0 2017-12-31 15:39 hfile_tmp5/_SUCCESS
drwxr-xr-x - root hdfs 0 2017-12-31 15:39 hfile_tmp5/cf
[root@sandbox-hdp ~]#
```


HBase Essentials Lab

- Bulk Load Data From HDFS into HBase

```
[root@sandbox-hdp ~]# hbase org.apache.hadoop.hbase.mapreduce.LoadIncrementalHFiles hfile_tmp5 simple2
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
2017-12-31 15:44:25,248 INFO [main] zookeeper.RecoverableZooKeeper: Process identifier=hconnection-0x2b91004a connecting to ZooKeeper ensemble=sandbox-hdp.hortonworks.com:2181
2017-12-31 15:44:25,266 INFO [main] zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.6-235--1, built on 10/30/2017 02:26 GMT
2017-12-31 15:44:25,266 INFO [main] zookeeper.ZooKeeper: Client environment:host.name=sandbox-hdp.hortonworks.com
2017-12-31 15:44:25,266 INFO [main] zookeeper.ZooKeeper: Client environment:java.version=1.8.0_151
2017-12-31 15:44:25,266 INFO [main] zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corporation
2017-12-31 15:44:25,266 INFO [main] zookeeper.ZooKeeper: Client environment:java.home=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.151-1.b12.el6_9.x86_64/jre
```

```
hbase(main):029:0> create 'simple2','cf'
0 row(s) in 1.2690 seconds

=> Hbase::Table - simple2
hbase(main):030:0> scan 'simple2'
ROW                                COLUMN+CELL
1                                  column=cf:, timestamp=1514734662844, value=tom
2                                  column=cf:, timestamp=1514734662844, value=sam
3                                  column=cf:, timestamp=1514734662844, value=jerry
4                                  column=cf:, timestamp=1514734662844, value=marry
5                                  column=cf:, timestamp=1514734662844, value=john
5 row(s) in 0.0400 seconds

hbase(main):031:0> █
```


Integration HBase with Hive Lab

- Prepare the dataset

```
[root@sandbox-hdp ~]# hadoop fs -mkdir /user/root/pagecounts
[root@sandbox-hdp ~]# hadoop fs -copyFromLocal /root/TrainingOnHDP/dataset/pagecounts/* /user/root/pagecounts
[root@sandbox-hdp ~]#
```

- Hive CLI

```
hive> CREATE TABLE IF NOT EXISTS PAGECOUNTS (projectcode STRING, pagename STRING, pageviews STRING, bytes STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ' ' LINES TERMINATED BY '\n' STORED AS TEXTFILE LOCATION '/root/TrainingOnHDP/dataset/pagecounts';
OK
Time taken: 2.462 seconds
hive>
```

```
hive> SELECT * FROM PAGECOUNTS LIMIT 10;
OK
aa.b    Special:Statistics      1      837
aa      Main_Page             4     41431
aa      Special:ListUsers      1      5555
aa      Special:ListUsers      1      1052
aa      Special:PrefixIndex/Comparison_of_Guaze%27s_Law_and_Coulomb%27s_Law 1      4332
aa      Special:Search         1      4799
aa      Special:WhatLinksHere/english_Listening_Skills 1      4323
aa      User:Alecs.y           1      12392
aa      User:VolkovBot         1      40994
ab      %25D0%2590%25D1%2586%25D0%25B3%25D1%2583%25D1%258B_%25D0%2598%25D0%25B0%25D2%259B%25D1%2583%25D1%2582%25D3%2599%25D0%25B0%25D0%25B7_%25D0%2590%25D2%25AD%25D0%25B0%25D0%25BA%25D1%2583%25D0%25B0%25D0%25B6%25D3%2599 1      13156
Time taken: 0.519 seconds, Fetched: 10 row(s)
hive>
```

Integration HBase with Hive Lab

- **Hive CLI**

```
hive> CREATE VIEW IF NOT EXISTS PGC(rowkey, pageviews, bytes) AS SELECT concat_ws('/', projectcode, concat_ws('/', pagename, regexp_extract(INPUT_FILE_NAME, 'pagecounts-(\\d{8}-\\d{6})', 1))), pageviews, bytes FROM PAGECOUNTS;
OK
Time taken: 1.018 seconds
hive> CREATE TABLE IF NOT EXISTS PAGECOUNTS_HBASE (rowkey STRING, pageviews STRING, bytes STRING) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = ':key,f1:PAGEVIEWS,f1:BYTES') TBLPROPERTIES ('hbase.table.name' = 'PAGECOUNTS');
OK
Time taken: 3.833 seconds
hive> FROM PGC INSERT INTO TABLE PAGECOUNTS_HBASE SELECT PGC.* WHERE rowkey LIKE 'en/q%' LIMIT 10;
Query ID = root_20171231161900_cb5b3e7b-9c7d-43cc-9afd-e940fefda412
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1514669965561_0084)
```

	VERTICES	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1		SUCCEEDED	19	19	0	0	0	0
Reducer 2		RUNNING	1	0	1	0	0	0

```
VERTICES: 01/02 [=====>--] 94% ELAPSED TIME: 1234.22 s
```

Integration HBase with Hive Lab

- HBase CLI

```
hbase(main):032:0> scan 'PAGECOUNTS'
ROW                                COLUMN+CELL
en/qac/20081001-010000             column=f1:BYTES, timestamp=1514739555161, value=1163
en/qac/20081001-010000             column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
en/qawasameh/20081001-010000        column=f1:BYTES, timestamp=1514739555161, value=981
en/qawasameh/20081001-010000        column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
en/qbama/20081001-010000            column=f1:BYTES, timestamp=1514739555161, value=977
en/qbama/20081001-010000            column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
en/qfn/20081001-010000             column=f1:BYTES, timestamp=1514739555161, value=975
en/qfn/20081001-010000             column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
en/qi/20081001-010000              column=f1:BYTES, timestamp=1514739555161, value=973
en/qi/20081001-010000              column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
en/qiagen/20081001-010000           column=f1:BYTES, timestamp=1514739555161, value=978
en/qiagen/20081001-010000           column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
en/qiki/Hermann_Ebbinghaus/20081001-0 column=f1:BYTES, timestamp=1514739555161, value=995
10000
en/qiki/Hermann_Ebbinghaus/20081001-0 column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
10000
en/qilin/20081001-010000            column=f1:BYTES, timestamp=1514739555161, value=1952
en/qilin/20081001-010000            column=f1:PAGEVIEWS, timestamp=1514739555161, value=2
en/qrcode/20081001-010000           column=f1:BYTES, timestamp=1514739555161, value=960
en/qrcode/20081001-010000           column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
en/qt./20081001-010000              column=f1:BYTES, timestamp=1514739555161, value=975
en/qt./20081001-010000              column=f1:PAGEVIEWS, timestamp=1514739555161, value=1
10 row(s) in 0.0880 seconds
```

Integration HBase with Hive Lab

- Phoenix

The screenshot displays the Ambari web interface for configuring HBase. On the left, a sidebar lists various services, with 'HBase' selected. The main panel shows the 'Configs' tab for the 'Default (1)' group. A notification bar at the top indicates 'ResourceManager Started'. Below this, a table lists three configuration versions: V3 (admin, about a month ago, HDP-2.6), V2 (internal, 2 months ago, HDP-2.6), and V1 (admin, 2 months ago, HDP-2.6). A message bar shows 'admin authored on Sat, Nov 25, 2017 23:44'. The 'Settings' section is currently set to 'Advanced'. The 'Server' section includes sliders for 'HBase Master Maximum Memory' (set to 1GB) and 'HBase RegionServer Maximum Memory' (set to 1GB). The 'Client' section includes sliders for 'Maximum Client Retries' (set to 35) and 'Maximum Record Size' (set to 1MB). A dropdown for 'HBase Region Block Multiplier' is set to 4. The 'Memstore Flush Size' is set to 128MB.

ResourceManager Started

- YARN
- MapReduce2
- Tez
- Hive
- HBase**
- Pig
- Sqoop
- Oozie
- ZooKeeper
- Falcon
- Storm
- Flume
- Ambari Infra
- Atlas
- Kafka
- Knox
- Ranger
- Spark2
- ZooKeeper

Summary Heatmaps Configs Quick Links Service Actions

Group Default (1) Manage Config Groups Filter...

V3 admin about a month ago HDP-2.6 V2 internal 2 months ago HDP-2.6 V1 admin 2 months ago HDP-2.6

admin authored on Sat, Nov 25, 2017 23:44 Discard Save

Settings Advanced

Server

HBase Master Maximum Memory 1GB 0 GB 5.25 GB 10.64 GB

Memstore Flush Size 128MB 32 MB 144 MB 256 MB

HBase RegionServer Maximum Memory 1GB 0 GB 4.25 GB 8.512 GB

HBase Region Block Multiplier 4

Client

Maximum Client Retries 35 5 28 50

Maximum Record Size 1MB 1 MB 15.5 MB 30 MB

Integration HBase with Hive Lab

- **Phoenix**

Timeouts

Zookeeper Session Timeout

Minutes Seconds

HBase RPC Timeout

Minutes Seconds

Security

Enable Authentication

Simple

Enable Authorization

Native

Phoenix SQL

Enable Phoenix

Enabled



Phoenix Query Timeout

Minutes Seconds

Integration HBase with Hive Lab

- Phoenix

```
[root@sandbox-hdp ~]# /usr/hdp/current/phoenix-client/bin/sqlline.py
Setting property: [incremental, false]
Setting property: [isolation, TRANSACTION_READ_COMMITTED]
issuing: !connect jdbc:phoenix: none none org.apache.phoenix.jdbc.PhoenixDriver
Connecting to jdbc:phoenix:
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/2.5.0.0-1245/phoenix/phoenix-4.7.0.2.5.0.0-1245-client.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/2.6.3.0-235/hadoop/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.
17/12/31 17:29:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/12/31 17:29:26 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used because libhadoop cannot be loaded.
Connected to: Phoenix (version 4.7)
Driver: PhoenixEmbeddedDriver (version 4.7)
Autocommit status: true
Transaction isolation: TRANSACTION_READ_COMMITTED
Building list of tables and columns for tab-completion (set fastconnect to true to skip)...
84/84 (100%) Done
Done
sqlline version 1.1.8
0: jdbc:phoenix:>
```

Integration HBase with Hive Lab

- Phoenix

```
0: jdbc:phoenix:> CREATE VIEW "PAGECOUNTS" (pk VARCHAR PRIMARY KEY, "f1".PAGEVIEWS VARCHAR, "f1".BYTES VARCHAR);
No rows affected (6.632 seconds)
0: jdbc:phoenix:> SELECT * FROM "PAGECOUNTS";
```

PK	PAGEVIEWS	BYTES
en/qac/20081001-010000	1	1163
en/qawasameh/20081001-010000	1	981
en/qbama/20081001-010000	1	977
en/qfn/20081001-010000	1	975
en/qi/20081001-010000	1	973
en/qiagen/20081001-010000	1	978
en/qiki/Hermann_Ebbinghaus/20081001-010000	1	995
en/qilin/20081001-010000	2	1952
en/qrcode/20081001-010000	1	960
en/qt./20081001-010000	1	975

```
10 rows selected (0.346 seconds)
0: jdbc:phoenix:>
```


Integration HBase with Hive Lab

- Zeppelin

HBase Page Counts Lab ▶ ⌂ 📄 📌 📁 📄 ⌂ Head 🗑️ ⌂ 🔧 🔒 default ▼

```
%jdbc(hive)
SELECT * FROM hbase.PGC WHERE rowkey = 'aa.b/Main_Page/20081001-010000';
```

FINISHED ▶ ⌂ 📄 📌 📁

📄 📊 📈 📉 📊 📉 📄 ▼

pgc.rowkey	pgc.pageviews	pgc.bytes
aa.b/Main_Page/20081001-010000	1	5509

Integration HBase with Hive Lab

- Zeppelin

%jdbc(phoenix)
SELECT * FROM "PAGECOUNTS";

FINISHED

PK	PAGEVIEWS	BYTES
en/qac/20081001-010000	1	1163
en/qawasameh/20081001-010000	1	981
en/qbama/20081001-010000	1	977
en/qfn/20081001-010000	1	975
en/qi/20081001-010000	1	973
en/qiagen/20081001-010000	1	978
en/qiki/Hermann_Ebbinghaus/20081001-010000	1	995
en/qilin/20081001-010000	2	1952
en/rcode/20081001-010000	1	960

Took 49 sec. Last updated by anonymous at December 31 2017, 1:16:18 PM.

HBASE Use Case - Time Series Database

- Check if the tables exist

← → ↻ ⓘ sandbox-hdp.hortonworks.com:16010/tablesDetailed.jsp ☆ 🌐

APACHE HBASE Home Table Details Procedures Local Logs Log Level Debug Dump Metrics Dump HBase Configuration

User Tables

14 table(s) in set.

Table	Description
ATLAS_ENTITY_AUDIT_EVENTS	'ATLAS_ENTITY_AUDIT_EVENTS', {NAME => 'dt', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELE TTL => 'FOREVER', COMPRESSION => 'GZ', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION
PAGECOUNTS	'PAGECOUNTS', {TABLE_ATTRIBUTES => {coprocessor\$1 => 'org.apache.phoenix.coprocessor.ScanRegionObserver[805306366]', cc 'org.apache.phoenix.coprocessor.UngroupedAggregateRegionObserver[805306366]', coprocessor\$3 => 'org.apache.phoenix.coproc 'org.apache.phoenix.coprocessor.ServerCachingEndpointImpl[805306366]'}, {NAME => 'f1', BLOOMFILTER => 'ROW', VERSIONS => ' DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 't
SYSTEM.CATALOG	'SYSTEM.CATALOG', {TABLE_ATTRIBUTES => {PRIORITY => '2000', coprocessor\$1 => 'org.apache.phoenix.coprocessor.ScanRegion 'org.apache.phoenix.coprocessor.UngroupedAggregateRegionObserver[805306366]', coprocessor\$3 => 'org.apache.phoenix.coproc 'org.apache.phoenix.coprocessor.ServerCachingEndpointImpl[805306366]', coprocessor\$5 => 'org.apache.phoenix.coprocessor.MetaD 'org.apache.phoenix.coprocessor.MetadataRegionObserver[805306367]', coprocessor\$7 => 'org.apache.hadoop.hbase.coprocessor.M => 'org.apache.phoenix.schema.MetadataSplitPolicy'}}, {NAME => '0', BLOOMFILTER => 'ROW', VERSIONS => '1000', IN_MEMORY = => 'FAST_DIFF', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '6
SYSTEM.FUNCTION	'SYSTEM.FUNCTION', {TABLE_ATTRIBUTES => {PRIORITY => '2000', coprocessor\$1 => 'org.apache.phoenix.coprocessor.ScanRegi 'org.apache.phoenix.coprocessor.UngroupedAggregateRegionObserver[805306366]', coprocessor\$3 => 'org.apache.phoenix.coproc 'org.apache.phoenix.coprocessor.ServerCachingEndpointImpl[805306366]', coprocessor\$5 => 'org.apache.phoenix.coproc

HBASE Use Case - Time Series Database

- **Create HBase TABLE**

```
hbase(main):003:0> create 'tsdb-uid', {NAME => 'id', COMPRESSION => 'SNAPPY', BLOOMFILTER => 'ROW'}, {NAME => 'name', COMPRESSION => 'SNAPPY', BLOOMFILTER => 'ROW'}
0 row(s) in 2.7590 seconds

=> Hbase::Table - tsdb-uid
hbase(main):004:0> create 'tsdb', {NAME => 't', VERSIONS => 1, COMPRESSION => 'SNAPPY', BLOOMFILTER => 'ROW'}
0 row(s) in 2.4300 seconds

=> Hbase::Table - tsdb
hbase(main):005:0> create 'tsdb-tree', {NAME => 't', VERSIONS => 1, COMPRESSION => 'SNAPPY', BLOOMFILTER => 'ROW'}
0 row(s) in 1.3030 seconds

=> Hbase::Table - tsdb-tree
hbase(main):006:0> create 'tsdb-meta', {NAME => 'name', COMPRESSION => 'SNAPPY', BLOOMFILTER => 'ROW'}
0 row(s) in 2.2520 seconds

=> Hbase::Table - tsdb-meta
hbase(main):007:0>
```

HBASE Use Case - Time Series Database

- Load Sample Data

```
[root@sandbox-hdp ~]# cd /tmp
[root@sandbox-hdp tmp]# git clone https://github.com/sakserv/opentsdb-examples.git
Initialized empty Git repository in /tmp/opentsdb-examples/.git/
remote: Counting objects: 236, done.
remote: Compressing objects: 100% (71/71), done.
remote: Total 236 (delta 145), reused 236 (delta 145), pack-reused 0
Receiving objects: 100% (236/236), 33.89 KiB, done.
Resolving deltas: 100% (145/145), done.
[root@sandbox-hdp tmp]# vi conf/opentsdb.conf
```

```
# Java Variables
# JAVA_BIN="/usr/jdk64/jdk1.7.0_45/bin/"
JAVA_BIN="/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.151-1.b12.el6_9.x86_64/bin/"
```

```
# HBase Variables
# HBASE_HOME="/usr/hdp/current/hbase-client/"
HBASE_HOME="/usr/hdp/2.6.3.0-235/hbase/"
```

```
CRIMES_DATA="https://s3-us-west-2.amazonaws.com/opentsdb/crime\_data.tar.gz"
WEATHER_DATA="https://s3-us-west-2.amazonaws.com/opentsdb/weather\_data.tar.gz"
SAMPLE_DATA_HDFS_DIR="/tmp/opentsdb_sample_data"
CRIMES_MAPPER=/tmp/opentsdb-examples/bin/mr/crimes_csv_opentsdb_put_mapper.py
CRIMES_REDUCER=/tmp/opentsdb-examples/bin/mr/crimes_csv_opentsdb_put_reducer.py
WEATHER_MAPPER=/tmp/opentsdb-examples/bin/mr/weather_csv_opentsdb_put_mapper.py
# HADOOP_STREAMING_JAR=/usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar
HADOOP_STREAMING_JAR=/usr/hdp/2.6.3.0-235/hadoop-mapreduce/hadoop-streaming-2.7.3.2.6.3.0-235.jar
```

HBASE Use Case - Time Series Database

- Load Sample Data

```
[root@sandbox-hdp ~]# cd /tmp
[root@sandbox-hdp tmp]# git clone https://github.com/sakserv/opentsdb-examples.git
Initialized empty Git repository in /tmp/opentsdb-examples/.git/
remote: Counting objects: 236, done.
remote: Compressing objects: 100% (71/71), done.
remote: Total 236 (delta 145), reused 236 (delta 145), pack-reused 0
Receiving objects: 100% (236/236), 33.89 KiB, done.
Resolving deltas: 100% (145/145), done.
[root@sandbox-hdp tmp]# vi conf/opentsdb.conf
```

```
##
+ echo -e '## Successfully completed OpenTSDB Table Creation on localhost'
## Successfully completed OpenTSDB Table Creation on localhost
+ echo -e '##'
##
+ exit 0
[root@sandbox-hdp opentsdb-examples]# bash -x bin/opentsdb_tsd_install.sh
```

Apache HBase Lab - PowerPoint

```
##
+ echo -e '## Successfully completed OpenTSDB TSD install on localhost'
## Successfully completed OpenTSDB TSD install on localhost
+ echo -e '##'
##
[root@sandbox-hdp opentsdb-examples]# bash -x bin/opentsdb_tcollector_install.sh
```

HBASE Use Case - Time Series Database

- Load Sample Data

```
##  
+ echo -e '## Successfully completed OpenTSDB TSD install on localhost'  
## Successfully completed OpenTSDB TSD install on localhost  
+ echo -e '##'  
##  
[root@sandbox-hdp opentsdb-examples]# bash -x bin/opentsdb_cli_install.sh
```

```
##  
+ echo -e '## Successfully completed OpenTSDB TSD CLI install on localhost'  
## Successfully completed OpenTSDB TSD CLI install on localhost  
+ echo -e '##'  
##  
[root@sandbox-hdp opentsdb-examples]# bash -x bin/opentsdb_load_sample_data.sh
```



HBASE Use Case - Time Series Database

- Start TSDB

```
[root@sandbox-hdp opentsdb-examples]# /usr/lib/opentsdb/build/tsdb tsd --port=15505 --staticroot=/usr/lib/opentsdb/build/staticroot --cachedir=/tmp/tsd_cache --zkquorum=localhost:2181 --zkbasedir=/hbase-unsecure --auto-metric >/var/log/opentsdb/opentsdb.log 2>&1
```

- OpenTSDB Web

← → ↻ ⓘ localhost:15505

 **OPENTSDB**

Graph Stats Logs Version

From To (now) ☐ Autoreload WxH: ☐ Global annotations

metric 1 +

Metric:

Tags

☐ Rate ☐ Rate Ctr ☐ Right Axis

Rate Ctr Max:

Rate Ctr Reset:

Aggregator:

☐ Downsample

Axes **Key** **Style**

	Y	Y2
Label	<input type="text"/>	<input type="text"/>
Format	<input type="text"/>	<input type="text"/>
Range	<input type="text" value="[0:]"/>	<input type="text" value="[0:]"/>
Log scale	<input type="checkbox"/>	<input type="checkbox"/>

Please specify a start time.

HBase Over REST

- Start REST Server

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/hbase/bin/hbase-daemon.sh start rest --port 8086 --infoport 8082
starting rest, logging to /var/log/hbase/hbase-root-rest-sandbox-hdp.hortonworks.com.out
[root@sandbox-hdp ~]#
```

- List of all nonsystem tables

← → ↻ ⓘ localhost:8086

```
ATLAS_ENTITY_AUDIT_EVENTS
PAGECOUNTS
SYSTEM.CATALOG
SYSTEM.FUNCTION
SYSTEM.SEQUENCE
SYSTEM.STATS
atlas_titan
business
iemployee
mytable
mytable_1
simple1
simple2
stocks
```

HBase Over REST

- Version of HBase running on this cluster

← → ↻ ⓘ localhost:8086/version/cluster

1.1.2.2.6.3.0-235

- Cluster status

← → ↻ ⓘ localhost:8086/status/cluster

```
1 live servers, 0 dead servers, 19.0000 average load
1 live servers
  sandbox-hdp.hortonworks.com:16020 1514741195033
    requests=0, regions=19
    heapSizeMB=101
    maxHeapSizeMB=1004
  ATLAS_ENTITY_AUDIT_EVENTS,,1510324862172.ddd6af691cc52c649dc81af34ae462b0.
    stores=1
    storefile=1
    storefileSizeMB=0
    memstoreSizeMB=0
    storefileIndexSizeMB=0
    readRequestsCount=0
    writeRequestsCount=0
    rootIndexSizeKB=2
    totalStaticIndexSizeKB=1
    totalStaticBloomSizeKB=1
    totalCompactingKVs=0
    currentCompactedKVs=0
  PAGECOUNTS,,1514737119693.eebdf03137fbe1458bd90429411da774.
    stores=1
    storefile=1
    storefileSizeMB=0
    memstoreSizeMB=0
    storefileIndexSizeMB=0
    readRequestsCount=20
    writeRequestsCount=0
    rootIndexSizeKB=0
    totalStaticIndexSizeKB=0
    totalStaticBloomSizeKB=0
    totalCompactingKVs=0
    currentCompactedKVs=0
```

HBase Over REST

- Describe the schema of the specified table

```
localhost:8086/tsdb/schema

{ NAME=> 'tsdb', IS_META => 'false', COLUMNS => [ { NAME => 't', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE',
TTL => '2147483647', COMPRESSION => 'SNAPPY', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0' } ] }
```

- List the table regions

```
localhost:8086/PAGECOUNTS/regions

PAGECOUNTS,,1514737119693.eebdf03137fbe1458bd90429411da774. [
  id=1514737119693
  startKey=''
  endKey=''
  location='sandbox-hdp.hortonworks.com:16020'
]
```

Phoenix

- Enter phoenix CLI

```
[root@sandbox-hdp ~]# /usr/hdp/current/phoenix-client/bin/sqlline.py localhost:2181
Setting property: [incremental, false]
Setting property: [isolation, TRANSACTION_READ_COMMITTED]
issuing: !connect jdbc:phoenix:localhost:2181 none none org.apache.phoenix.jdbc.PhoenixDriver
Connecting to jdbc:phoenix:localhost:2181
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/2.5.0.0-1245/phoenix/phoenix-4.7.0.2.5.0.0-1245-client.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/2.6.3.0-235/hadoop/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.
18/01/04 03:27:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/01/04 03:27:16 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used because libhadoop cannot be loaded.
Connected to: Phoenix (version 4.7)
Driver: PhoenixEmbeddedDriver (version 4.7)
Autocommit status: true
Transaction isolation: TRANSACTION_READ_COMMITTED
Building list of tables and columns for tab-completion (set fastconnect to true to skip)...
87/87 (100%) Done
Done
sqlline version 1.1.8
0: jdbc:phoenix:localhost:2181> 
```

Phoenix

- **Create HBase Table using phoenix**

```
0: jdbc:phoenix:localhost:2181> create table if not exists PRICES ( SYMBOL varchar(10), DATE varchar(10), TIME varchar(10), OPEN varchar(10), HIGH va  
rchar(10), LOW varchar(10), CLOSE varchar(10), VOLUME varchar(30), CONSTRAINT pk PRIMARY KEY (SYMBOL, DATE, TIME) );  
No rows affected (5.377 seconds)  
0: jdbc:phoenix:localhost:2181> █
```

Exploring HBase System

- Exploring Zookeeper

```
hbase(main):001:0> zk_dump
HBase is rooted at /hbase-unsecure
Active master address: sandbox-hdp.hortonworks.com,16000,1515035868135
Backup master addresses:
Region server holding hbase:meta: sandbox-hdp.hortonworks.com,16020,1515035985652
Region servers:
  sandbox-hdp.hortonworks.com,16020,1515035985652
/hbase-unsecure/replication:
/hbase-unsecure/replication/peers:
/hbase-unsecure/replication/rs:
/hbase-unsecure/replication/rs/sandbox-hdp.hortonworks.com,16020,1515035985652:
Quorum Server Statistics:
  sandbox-hdp.hortonworks.com:2181
  Zookeeper version: 3.4.6-235--1, built on 10/30/2017 02:26 GMT
  Clients:
    /172.17.0.2:53942[1](queued=0,recved=57,sent=57)
    /172.17.0.2:51518[1](queued=0,recved=60,sent=60)
    /172.17.0.2:43292[1](queued=0,recved=12,sent=12)
    /172.17.0.2:54808[1](queued=0,recved=52,sent=52)
    /172.17.0.2:51852[1](queued=0,recved=61,sent=61)
    /172.17.0.2:35756[1](queued=0,recved=98,sent=98)
    /172.17.0.2:51204[1](queued=0,recved=5399,sent=5482)
    /172.17.0.2:53808[1](queued=0,recved=266,sent=266)
    /172.17.0.2:47284[1](queued=0,recved=12,sent=12)
    /172.17.0.2:40610[1](queued=0,recved=10439,sent=10484)
    /172.17.0.2:51352[1](queued=0,recved=62,sent=62)
    /172.17.0.2:47292[0](queued=0,recved=1,sent=0)

  Latency min/avg/max: 0/0/694
  Received: 17355
  Sent: 17482
  Connections: 12
```



```
hbase(main):002:0> scan 'hbase:meta'
ROW                                COLUMN+CELL
ATLAS_ENTITY_AUDIT_EVENTS,,1510324862 column=info:regioninfo, timestamp=1510324863407, value={ENCODED => ddd6af691cc52c649dc81af34ae462b0, NAME => '172.ddd6af691cc52c649dc81af34ae462b0. ATLAS_ENTITY_AUDIT_EVENTS,,1510324862172.ddd6af691cc52c649dc81af34ae462b0.', STARTKEY => '', ENDKEY => ''}
ATLAS_ENTITY_AUDIT_EVENTS,,1510324862 column=info:seqnumDuringOpen, timestamp=1515036055516, value=\x00\x00\x00\x00\x00\x00\x00\xA3
172.ddd6af691cc52c649dc81af34ae462b0.
ATLAS_ENTITY_AUDIT_EVENTS,,1510324862 column=info:server, timestamp=1515036055516, value=sandbox-hdp.hortonworks.com:16020
172.ddd6af691cc52c649dc81af34ae462b0.
ATLAS_ENTITY_AUDIT_EVENTS,,1510324862 column=info:serverstartcode, timestamp=1515036055516, value=1515035985652
172.ddd6af691cc52c649dc81af34ae462b0.
PAGECOUNTS,,1514737119693.eebdf03137f column=info:regioninfo, timestamp=1514737120818, value={ENCODED => eebdf03137fbe1458bd90429411da774, NAME => 'be1458bd90429411da774. PAGECOUNTS,,1514737119693.eebdf03137fbe1458bd90429411da774.', STARTKEY => '', ENDKEY => ''}
PAGECOUNTS,,1514737119693.eebdf03137f column=info:seqnumDuringOpen, timestamp=1515036062972, value=\x00\x00\x00\x00\x00\x00\x00\x12
be1458bd90429411da774.
PAGECOUNTS,,1514737119693.eebdf03137f column=info:server, timestamp=1515036062972, value=sandbox-hdp.hortonworks.com:16020
be1458bd90429411da774.
PAGECOUNTS,,1514737119693.eebdf03137f column=info:serverstartcode, timestamp=1515036062972, value=1515035985652
be1458bd90429411da774.
PRICES,,1515036533751.bf82afe945974df column=info:regioninfo, timestamp=1515036535383, value={ENCODED => bf82afe945974df989d5e0c29e9c2d77, NAME => '989d5e0c29e9c2d77. PRICES,,1515036533751.bf82afe945974df989d5e0c29e9c2d77.', STARTKEY => '', ENDKEY => ''}
PRICES,,1515036533751.bf82afe945974df column=info:seqnumDuringOpen, timestamp=1515036536521, value=\x00\x00\x00\x00\x00\x00\x00\x02
989d5e0c29e9c2d77.
PRICES,,1515036533751.bf82afe945974df column=info:server, timestamp=1515036536521, value=sandbox-hdp.hortonworks.com:16020
989d5e0c29e9c2d77.
PRICES,,1515036533751.bf82afe945974df column=info:serverstartcode, timestamp=1515036536521, value=1515035985652
989d5e0c29e9c2d77.
SYSTEM.CATALOG,,1514741377001.72e7dc4 column=info:regioninfo, timestamp=1514741379555, value={ENCODED => 72e7dc4436a9d6a649ac01e286d9f908, NAME => '436a9d6a649ac01e286d9f908. SYSTEM.CATALOG,,1514741377001.72e7dc4436a9d6a649ac01e286d9f908.', STARTKEY => '', ENDKEY => ''}
SYSTEM.CATALOG,,1514741377001.72e7dc4 column=info:seqnumDuringOpen, timestamp=1515036065159, value=\x00\x00\x00\x00\x00\x00\x00\x18
436a9d6a649ac01e286d9f908.
SYSTEM.CATALOG,,1514741377001.72e7dc4 column=info:server, timestamp=1515036065159, value=sandbox-hdp.hortonworks.com:16020
436a9d6a649ac01e286d9f908.
```

Exploring HBase System

- Exploring HBase Meta

```
SYSTEM.CATALOG,,1514741377001.72e7dc4 column=info:serverstartcode, timestamp=1515036065159, value=1515035985652
436a9d6a649ac01e286d9f908.
SYSTEM.FUNCTION,,1514741400139.170e4e column=info:regioninfo, timestamp=1514741400668, value={ENCODED => 170e4e4d0d08c4e7e2953239cee20151, NAME => '
4d0d08c4e7e2953239cee20151. SYSTEM.FUNCTION,,1514741400139.170e4e4d0d08c4e7e2953239cee20151.', STARTKEY => '', ENDKEY => ''}
SYSTEM.FUNCTION,,1514741400139.170e4e column=info:seqnumDuringOpen, timestamp=1515036061512, value=\x00\x00\x00\x00\x00\x00\x00\x05
4d0d08c4e7e2953239cee20151.
SYSTEM.FUNCTION,,1514741400139.170e4e column=info:server, timestamp=1515036061512, value=sandbox-hdp.hortonworks.com:16020
4d0d08c4e7e2953239cee20151.
SYSTEM.FUNCTION,,1514741400139.170e4e column=info:serverstartcode, timestamp=1515036061512, value=1515035985652
4d0d08c4e7e2953239cee20151.
SYSTEM.SEQUENCE,,1514741396136.771009 column=info:regioninfo, timestamp=1514741396600, value={ENCODED => 771009a9be83035f138541efa81309ba, NAME => '
a9be83035f138541efa81309ba. SYSTEM.SEQUENCE,,1514741396136.771009a9be83035f138541efa81309ba.', STARTKEY => '', ENDKEY => ''}
SYSTEM.SEQUENCE,,1514741396136.771009 column=info:seqnumDuringOpen, timestamp=1515036062846, value=\x00\x00\x00\x00\x00\x00\x00\x05
a9be83035f138541efa81309ba.
SYSTEM.SEQUENCE,,1514741396136.771009 column=info:server, timestamp=1515036062846, value=sandbox-hdp.hortonworks.com:16020
a9be83035f138541efa81309ba.
SYSTEM.SEQUENCE,,1514741396136.771009 column=info:serverstartcode, timestamp=1515036062846, value=1515035985652
a9be83035f138541efa81309ba.
SYSTEM.STATS,,1514741398598.bb0d2ad0a column=info:regioninfo, timestamp=1514741399183, value={ENCODED => bb0d2ad0ae9338dc156b617059a895e4, NAME => '
e9338dc156b617059a895e4. SYSTEM.STATS,,1514741398598.bb0d2ad0ae9338dc156b617059a895e4.', STARTKEY => '', ENDKEY => ''}
SYSTEM.STATS,,1514741398598.bb0d2ad0a column=info:seqnumDuringOpen, timestamp=1515036056422, value=\x00\x00\x00\x00\x00\x00\x00\x05
e9338dc156b617059a895e4.
SYSTEM.STATS,,1514741398598.bb0d2ad0a column=info:server, timestamp=1515036056422, value=sandbox-hdp.hortonworks.com:16020
e9338dc156b617059a895e4.
SYSTEM.STATS,,1514741398598.bb0d2ad0a column=info:serverstartcode, timestamp=1515036056422, value=1515035985652
e9338dc156b617059a895e4.
atlas_titan,,1510324859831.3e0f2dd293 column=info:regioninfo, timestamp=1510324860739, value={ENCODED => 3e0f2dd2936a55743ba6d5bb8109342b, NAME => '
6a55743ba6d5bb8109342b. atlas_titan,,1510324859831.3e0f2dd2936a55743ba6d5bb8109342b.', STARTKEY => '', ENDKEY => ''}
atlas_titan,,1510324859831.3e0f2dd293 column=info:seqnumDuringOpen, timestamp=1515036063267, value=\x00\x00\x00\x00\x00\x00\x00\x01C
6a55743ba6d5bb8109342b.
atlas_titan,,1510324859831.3e0f2dd293 column=info:server, timestamp=1515036063267, value=sandbox-hdp.hortonworks.com:16020
6a55743ba6d5bb8109342b.
```

Exploring HBase System

- Exploring HBase Meta

```
atlas_titan,,1510324859831.3e0f2dd293 column=info:serverstartcode, timestamp=1515036063267, value=1515035985652
6a55743ba6d5bb8109342b.
business,,1514731689890.c3e44e26a41ab column=info:regioninfo, timestamp=1514731697989, value={ENCODED => c3e44e26a41abc62e32ebffb511aed57, NAME => '
c62e32ebffb511aed57. business,,1514731689890.c3e44e26a41abc62e32ebffb511aed57.', STARTKEY => '', ENDKEY => '2278'}
business,,1514731689890.c3e44e26a41ab column=info:seqnumDuringOpen, timestamp=1515036056558, value=\x00\x00\x00\x00\x00\x00\x03A
c62e32ebffb511aed57.
business,,1514731689890.c3e44e26a41ab column=info:server, timestamp=1515036056558, value=sandbox-hdp.hortonworks.com:16020
c62e32ebffb511aed57.
business,,1514731689890.c3e44e26a41ab column=info:serverstartcode, timestamp=1515036056558, value=1515035985652
c62e32ebffb511aed57.
business,2278,1514731689890.b91fa198e column=info:regioninfo, timestamp=1514731697989, value={ENCODED => b91fa198aaa6e8c6e7df981f70d7dec0, NAME => '
aa6e8c6e7df981f70d7dec0. business,2278,1514731689890.b91fa198aaa6e8c6e7df981f70d7dec0.', STARTKEY => '2278', ENDKEY => ''}
business,2278,1514731689890.b91fa198e column=info:seqnumDuringOpen, timestamp=1515036056448, value=\x00\x00\x00\x00\x00\x00\x05\xBB
aa6e8c6e7df981f70d7dec0.
business,2278,1514731689890.b91fa198e column=info:server, timestamp=1515036056448, value=sandbox-hdp.hortonworks.com:16020
aa6e8c6e7df981f70d7dec0.
business,2278,1514731689890.b91fa198e column=info:serverstartcode, timestamp=1515036056448, value=1515035985652
aa6e8c6e7df981f70d7dec0.
hbase:acl,,1510324824181.33e3e6217353 column=info:regioninfo, timestamp=1510324826017, value={ENCODED => 33e3e6217353514abed742be8ea5790b, NAME => '
514abed742be8ea5790b. hbase:acl,,1510324824181.33e3e6217353514abed742be8ea5790b.', STARTKEY => '', ENDKEY => ''}
hbase:acl,,1510324824181.33e3e6217353 column=info:seqnumDuringOpen, timestamp=1515036050401, value=\x00\x00\x00\x00\x00\x00\x00\x16
514abed742be8ea5790b.
hbase:acl,,1510324824181.33e3e6217353 column=info:server, timestamp=1515036050401, value=sandbox-hdp.hortonworks.com:16020
514abed742be8ea5790b.
hbase:acl,,1510324824181.33e3e6217353 column=info:serverstartcode, timestamp=1515036050401, value=1515035985652
514abed742be8ea5790b.
hbase:backup,,1510324824658.756eb4cb2 column=info:regioninfo, timestamp=1510324826701, value={ENCODED => 756eb4cb2ceefcdd6f946e8dd1e66576, NAME => '
ceefcdd6f946e8dd1e66576. hbase:backup,,1510324824658.756eb4cb2ceefcdd6f946e8dd1e66576.', STARTKEY => '', ENDKEY => ''}
hbase:backup,,1510324824658.756eb4cb2 column=info:seqnumDuringOpen, timestamp=1515036059345, value=\x00\x00\x00\x00\x00\x00\x00\x15
ceefcdd6f946e8dd1e66576.
hbase:backup,,1510324824658.756eb4cb2 column=info:server, timestamp=1515036059345, value=sandbox-hdp.hortonworks.com:16020
ceefcdd6f946e8dd1e66576.
```

Exploring HBase System

- Exploring HBase Meta

```
hbase:backup,,1510324824658.756eb4cb2 column=info:serverstartcode, timestamp=1515036059345, value=1515035985652
ceefcdd6f946e8dd1e66576.
hbase:namespace,,1510324819825.dc3a25 column=info:regioninfo, timestamp=1510324821675, value={ENCODED => dc3a25901f3d97e2dee29f1b56669450, NAME => '
901f3d97e2dee29f1b56669450. hbase:namespace,,1510324819825.dc3a25901f3d97e2dee29f1b56669450.', STARTKEY => '', ENDKEY => ''}
hbase:namespace,,1510324819825.dc3a25 column=info:seqnumDuringOpen, timestamp=1515036062978, value=\x00\x00\x00\x00\x00\x00\x1D
901f3d97e2dee29f1b56669450.
hbase:namespace,,1510324819825.dc3a25 column=info:server, timestamp=1515036062978, value=sandbox-hdp.hortonworks.com:16020
901f3d97e2dee29f1b56669450.
hbase:namespace,,1510324819825.dc3a25 column=info:serverstartcode, timestamp=1515036062978, value=1515035985652
901f3d97e2dee29f1b56669450.
iemployee,,1510325640891.b947073d72b6 column=info:regioninfo, timestamp=1510325641322, value={ENCODED => b947073d72b6714fbbe0e6ebb8ecb03e, NAME => '
714fbbe0e6ebb8ecb03e. iemployee,,1510325640891.b947073d72b6714fbbe0e6ebb8ecb03e.', STARTKEY => '', ENDKEY => ''}
iemployee,,1510325640891.b947073d72b6 column=info:seqnumDuringOpen, timestamp=1515036062739, value=\x00\x00\x00\x00\x00\x00
714fbbe0e6ebb8ecb03e.
iemployee,,1510325640891.b947073d72b6 column=info:server, timestamp=1515036062739, value=sandbox-hdp.hortonworks.com:16020
714fbbe0e6ebb8ecb03e.
iemployee,,1510325640891.b947073d72b6 column=info:serverstartcode, timestamp=1515036062739, value=1515035985652
714fbbe0e6ebb8ecb03e.
mytable,,1514717456054.40d82049387cb4 column=info:regioninfo, timestamp=1514717456619, value={ENCODED => 40d82049387cb4ea53d168ee30637d2e, NAME => '
ea53d168ee30637d2e. mytable,,1514717456054.40d82049387cb4ea53d168ee30637d2e.', STARTKEY => '', ENDKEY => ''}
mytable,,1514717456054.40d82049387cb4 column=info:seqnumDuringOpen, timestamp=1515036056557, value=\x00\x00\x00\x00\x00\x00\x19
ea53d168ee30637d2e.
mytable,,1514717456054.40d82049387cb4 column=info:server, timestamp=1515036056557, value=sandbox-hdp.hortonworks.com:16020
ea53d168ee30637d2e.
mytable,,1514717456054.40d82049387cb4 column=info:serverstartcode, timestamp=1515036056557, value=1515035985652
ea53d168ee30637d2e.
mytable_1,,1514720674117.abef1ce04280 column=info:regioninfo, timestamp=1514720675003, value={ENCODED => abef1ce04280933971ed5993087ba106, NAME => '
933971ed5993087ba106. mytable_1,,1514720674117.abef1ce04280933971ed5993087ba106.', STARTKEY => '', ENDKEY => ''}
mytable_1,,1514720674117.abef1ce04280 column=info:seqnumDuringOpen, timestamp=1515036050340, value=\x00\x00\x00\x00\x00\x00\x0D
933971ed5993087ba106.
mytable_1,,1514720674117.abef1ce04280 column=info:server, timestamp=1515036050340, value=sandbox-hdp.hortonworks.com:16020
933971ed5993087ba106.
```


Exploring HBase System

- Exploring HBase Meta

```
mytable_1,,1514720674117.abef1ce04280 column=info:serverstartcode, timestamp=1515036050340, value=1515035985652
933971ed5993087ba106.
simple1,,1514734026303.e1e0fec5116cb6 column=info:regioninfo, timestamp=1514734027124, value={ENCODED => e1e0fec5116cb6c0634e1182784d5410, NAME => '
c0634e1182784d5410. simple1,,1514734026303.e1e0fec5116cb6c0634e1182784d5410.', STARTKEY => '', ENDKEY => ''}
simple1,,1514734026303.e1e0fec5116cb6 column=info:seqnumDuringOpen, timestamp=1515036059307, value=\x00\x00\x00\x00\x00\x00\x00\x0C
c0634e1182784d5410.
simple1,,1514734026303.e1e0fec5116cb6 column=info:server, timestamp=1515036059307, value=sandbox-hdp.hortonworks.com:16020
c0634e1182784d5410.
simple1,,1514734026303.e1e0fec5116cb6 column=info:serverstartcode, timestamp=1515036059307, value=1515035985652
c0634e1182784d5410.
simple2,,1514734577440.f533b23413d636 column=info:regioninfo, timestamp=1514734577883, value={ENCODED => f533b23413d63629415b4f7548443bb4, NAME => '
29415b4f7548443bb4. simple2,,1514734577440.f533b23413d63629415b4f7548443bb4.', STARTKEY => '', ENDKEY => ''}
simple2,,1514734577440.f533b23413d636 column=info:seqnumDuringOpen, timestamp=1515036054341, value=\x00\x00\x00\x00\x00\x00\x00\x0C
29415b4f7548443bb4.
simple2,,1514734577440.f533b23413d636 column=info:server, timestamp=1515036054341, value=sandbox-hdp.hortonworks.com:16020
29415b4f7548443bb4.
simple2,,1514734577440.f533b23413d636 column=info:serverstartcode, timestamp=1515036054341, value=1515035985652
29415b4f7548443bb4.
stocks,,1514259464002.0b8dfaccaa7b34a column=info:regioninfo, timestamp=1514259465577, value={ENCODED => 0b8dfaccaa7b34a6847d0171e07ef3ae, NAME => '
6847d0171e07ef3ae. stocks,,1514259464002.0b8dfaccaa7b34a6847d0171e07ef3ae.', STARTKEY => '', ENDKEY => ''}
stocks,,1514259464002.0b8dfaccaa7b34a column=info:seqnumDuringOpen, timestamp=1515036057362, value=\x00\x00\x00\x00\x00\x00\x00\x0E
6847d0171e07ef3ae.
stocks,,1514259464002.0b8dfaccaa7b34a column=info:server, timestamp=1515036057362, value=sandbox-hdp.hortonworks.com:16020
6847d0171e07ef3ae.
stocks,,1514259464002.0b8dfaccaa7b34a column=info:serverstartcode, timestamp=1515036057362, value=1515035985652
6847d0171e07ef3ae.
19 row(s) in 2.7030 seconds

hbase(main):003:0> 
```

Credit Card Transaction Monitor

- **Create CustomerAccount Table using HBase Admin and Phoenix**

```
localhost:4200
sandbox login: root
root@sandbox.hortonworks.com's password:
Last login: Thu Jan  4 16:52:30 2018 from 127.0.0.1
[root@sandbox-hdp ~]# java -cp /root/TrainingOnHDP/HBaseCookBook/target/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.hbase.CreditCardCustomerAccountToHBase
log4j:WARN No appenders could be found for logger (org.apache.hadoop.metrics2.lib.MutableMetricsFactory).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
*****is active: true
*****first name: Regina
*****last name: Smith
*****age: 32
*****gender: Female
*****street address: 1234 Tampa Ave.
*****city: Cherry Hill
*****state: NJ
*****zip code: 08003
*****home latitude: 39.919512
*****home longitude: -75.005711
[root@sandbox-hdp ~]#
```

Credit Card Transaction Monitor

- Create CustomerAccount Table using HBase Admin and Phoenix

```
localhost:4200
sandbox login: root
root@sandbox.hortonworks.com's password:
Last login: Thu Jan  4 18:37:33 2018 from 127.0.0.1
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/phoenix/bin/sqlline.py
Setting property: [incremental, false]
Setting property: [isolation, TRANSACTION_READ_COMMITTED]
Issuing: !connect jdbc:phoenix: none none org.apache.phoenix.jdbc.PhoenixDriver
Connecting to jdbc:phoenix:
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/2.6.3.0-235/phoenix/phoenix-4.7.0.2.6.3.0-235-client.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/2.6.3.0-235/hadoop/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.
18/01/04 18:41:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/01/04 18:41:24 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used because libhadoop cannot be loaded.
Connected to: Phoenix (version 4.7)
Driver: PhoenixEmbeddedDriver (version 4.7)
Autocommit status: true
Transaction isolation: TRANSACTION_READ_COMMITTED
Building list of tables and columns for tab-completion (set fastconnect to true to skip)...
132/132 (100%) Done
Done
sqlline version 1.1.8
0: jdbc:phoenix:> select * from "CustomerAccountBI";
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| accountNumber | firstName | lastName | age | gender | streetAddress | city | state | zipcode | latitude | longitude | ipAddr |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 19123         | Regina   | SMITH    | 32  | Female | 1234 Tampa Ave | Cherry Hill | NJ    | 08003   | 39.919512 | -75.005711 |        |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
1 row selected (4.044 seconds)
0: jdbc:phoenix:>
```


Credit Card Transaction Monitor

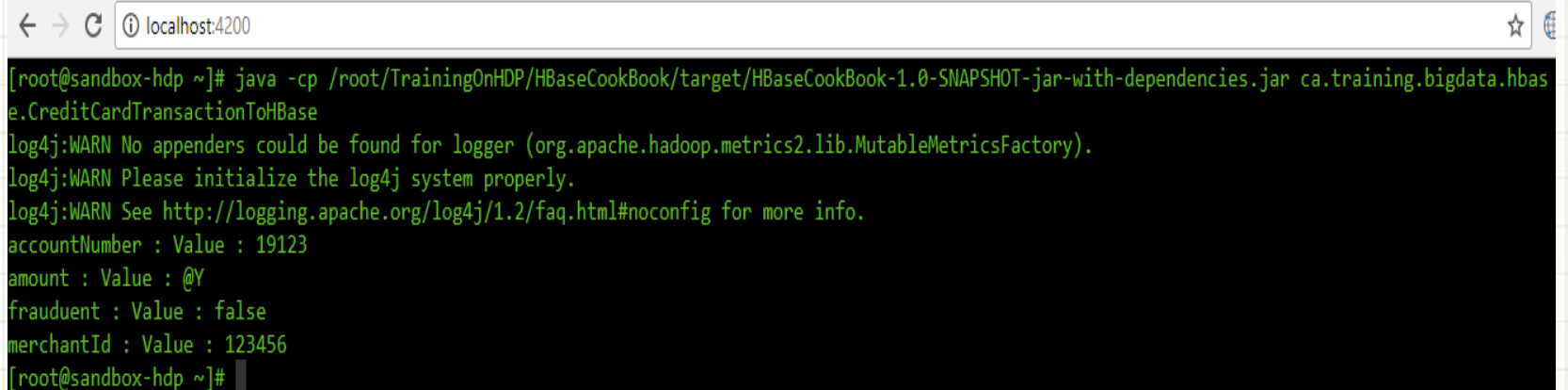
- Create CustomerAccount Table using HBase Admin and Phoenix

```
localhost:4200
sandbox login: root
root@sandbox.hortonworks.com's password:
Last login: Thu Jan  4 18:40:26 2018 from 127.0.0.1
[root@sandbox-hdp ~]# hbase shell
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.1.2.2.6.3.0-235, r6f982c8e667e7a18451fcbc12c5e2758b178ec78, Mon Oct 30 02:53:33 UTC 2017

hbase(main):001:0> scan 'CustomerAccount'
ROW                                COLUMN+CELL
19123                             column=AccountDetails:accountLimit, timestamp=1515091104300, value=20000
19123                             column=AccountDetails:accountNumber, timestamp=1515091104300, value=19123
19123                             column=AccountDetails:accountType, timestamp=1515091104300, value=VISA
19123                             column=AccountDetails:conAmtDev, timestamp=1515091104300, value=4.272822919
19123                             column=AccountDetails:conAmtMean, timestamp=1515091104300, value=16.87915743
19123                             column=AccountDetails:distanceDev, timestamp=1515091104300, value=5.968364997
19123                             column=AccountDetails:distanceMean, timestamp=1515091104300, value=9.173712305
19123                             column=AccountDetails:elecAmtDev, timestamp=1515091104300, value=29.0160567
19123                             column=AccountDetails:elecAmtMean, timestamp=1515091104300, value=98.97291196
19123                             column=AccountDetails:entAmtDev, timestamp=1515091104300, value=5.728492345
19123                             column=AccountDetails:entAmtMean, timestamp=1515091104300, value=39.4743295
19123                             column=AccountDetails:expMonth, timestamp=1515091104300, value=09
19123                             column=AccountDetails:expYear, timestamp=1515091104300, value=19
19123                             column=AccountDetails:gasAmtDev, timestamp=1515091104300, value=7.226414921
19123                             column=AccountDetails:gasAmtMean, timestamp=1515091104300, value=36.9679558
19123                             column=AccountDetails:grocAmtDev, timestamp=1515091104300, value=32.0403594
19123                             column=AccountDetails:grocAmtMean, timestamp=1515091104300, value=83.73396065
19123                             column=AccountDetails:hbAmtDev, timestamp=1515091104300, value=35.58637624
19123                             column=AccountDetails:hbAmtMean, timestamp=1515091104300, value=84.07411631
19123                             column=AccountDetails:isActive, timestamp=1515091104300, value=true
19123                             column=AccountDetails:rAmtDev, timestamp=1515091104300, value=72.17713403
19123                             column=AccountDetails:rAmtMean, timestamp=1515091104300, value=174.1947298
19123                             column=AccountDetails:restAmtDev, timestamp=1515091104300, value=38.0403594
```

Credit Card Transaction Monitor

- **Create TransactionHistory Table using HBase Admin**



A screenshot of a terminal window with a black background and green text. The terminal shows the execution of a Java command to create a table in HBase. The command is: `java -cp /root/TrainingOnHDP/HBaseCookBook/target/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.hbase.CreditCardTransactionToHBase`. The output shows several log4j warnings about the logging system and then displays the values for the transaction: `accountNumber : Value : 19123`, `amount : Value : @Y`, `frauduent : Value : false`, and `merchantId : Value : 123456`. The terminal prompt is `[root@sandbox-hdp ~]#`.

```
[root@sandbox-hdp ~]# java -cp /root/TrainingOnHDP/HBaseCookBook/target/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.hbase.CreditCardTransactionToHBase
log4j:WARN No appenders could be found for logger (org.apache.hadoop.metrics2.lib.MutableMetricsFactory).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
accountNumber : Value : 19123
amount : Value : @Y
frauduent : Value : false
merchantId : Value : 123456
[root@sandbox-hdp ~]#
```

Credit Card Transaction Monitor

- Create TransactionHistory Table using HBase Admin

```
iemployee
mytable
mytable_1
simple1
simple2
stocks
tsdb
tsdb-meta
tsdb-tree
tsdb-uid
22 row(s) in 0.8830 seconds

=> ["ATLAS_ENTITY_AUDIT_EVENTS", "CustomerAccount", "CustomerAccountBI", "PAGECOUNTS", "PRICES", "SYSTEM.CATALOG", "SYSTEM.FUNCTION", "SYSTEM.SEQUENC
E", "SYSTEM.STATS", "TransactionHistory", "atlas_titan", "business", "iemployee", "mytable", "mytable_1", "simple1", "simple2", "stocks", "tsdb", "ts
db-meta", "tsdb-tree", "tsdb-uid"]
hbase(main):043:0> scan 'TransactionHistory'
ROW                                COLUMN+CELL
11111111                            column=Transactions:accountNumber, timestamp=1515092961121, value=19123
11111111                            column=Transactions:accountType, timestamp=1515092961121, value=VISA
11111111                            column=Transactions:amount, timestamp=1515092961121, value=@Y\x00\x00\x00\x00\x00\x00
11111111                            column=Transactions:currency, timestamp=1515092961121, value=CAD
11111111                            column=Transactions:distanceFromHome, timestamp=1515092961121, value=@Y\x00\x00\x00\x00\x00\x00
11111111                            column=Transactions:distanceFromPrev, timestamp=1515092961121, value=@i\x00\x00\x00\x00\x00\x00
11111111                            column=Transactions:fraudent, timestamp=1515092961121, value=false
11111111                            column=Transactions:ipAddress, timestamp=1515092961121, value=
11111111                            column=Transactions:isCardPresent, timestamp=1515092961121, value=yes
11111111                            column=Transactions:latitude, timestamp=1515092961121, value=@C\xF5\xB2\x91\xB8#\xC8
11111111                            column=Transactions:longitude, timestamp=1515092961121, value=\xC0R\xC0]\x91\xAB\x8E\x8F
11111111                            column=Transactions:merchantId, timestamp=1515092961121, value=123456
11111111                            column=Transactions:merchantType, timestamp=1515092961121, value=Retail
11111111                            column=Transactions:transactionId, timestamp=1515092961121, value=11111111
11111111                            column=Transactions:transactionTimeStamp, timestamp=1515092961121, value=\x00\x00\x00\x00\x00\x00\x00\x00
```

HBase Data Transfer

- HBase on Spark - Save Airline Test Data to HBase

```
[root@sandbox-hdp ~]# spark-submit --class ca.training.bigdata.hbase.AirlineTestDataToHBase --driver-memory 1G --executor-memory 1G --master local[1]
/root/TrainingOnHDP/HBaseCookBook/target/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar
SPARK_MAJOR_VERSION is set to 2, using Spark2
18/01/04 19:34:13 INFO SparkContext: Running Spark version 2.2.0.2.6.3.0-235
18/01/04 19:34:15 INFO SparkContext: Submitted application: Airline Test Data to HBase
18/01/04 19:34:15 INFO SecurityManager: Changing view acls to: root
18/01/04 19:34:15 INFO SecurityManager: Changing modify acls to: root
18/01/04 19:34:15 INFO SecurityManager: Changing view acls groups to:
18/01/04 19:34:15 INFO SecurityManager: Changing modify acls groups to:
18/01/04 19:34:15 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); groups with view permissions: Set(); users with modify permissions: Set(root); groups with modify permissions: Set()
18/01/04 19:34:17 INFO Utils: Successfully started service 'sparkDriver' on port 37665.
18/01/04 19:34:17 INFO SparkEnv: Registering MapOutputTracker
18/01/04 19:34:18 INFO SparkEnv: Registering BlockManagerMaster
18/01/04 19:34:18 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
18/01/04 19:34:18 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
18/01/04 19:34:18 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-e81615ce-4a45-4ef5-877f-3910c9176586
18/01/04 19:34:18 INFO MemoryStore: MemoryStore started with capacity 413.9 MB
18/01/04 19:34:18 INFO SparkEnv: Registering OutputCommitCoordinator
18/01/04 19:34:19 INFO log: Logging initialized @9583ms
18/01/04 19:34:19 INFO Server: jetty-9.3.z-SNAPSHOT
18/01/04 19:34:19 INFO Server: Started @10455ms
18/01/04 19:34:20 INFO AbstractConnector: Started ServerConnector@29ad44e3{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
18/01/04 19:34:20 INFO Utils: Successfully started service 'SparkUI' on port 4040.
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@3aacf32a{/jobs,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@38875e7d{/jobs/json,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@d816dde{/jobs/job,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@31c269fd{/jobs/job/json,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@47747fb9{/stages,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@213e3629{/stages/json,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@2a7b6f69{/stages/stage,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@7db0565c{/stages/stage/json,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@52eacb4b{/stages/pool,null,AVAILABLE,@Spark}
18/01/04 19:34:20 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@2a551a63{/stages/pool/json,null,AVAILABLE,@Spark}
```

HBase Data Transfer

- HBase on Spark - Save Airline Test Data to HBase

```
hbase(main):046:0> scan 'airdelaydata'
COLUMN+CELL
row000      column=AirlineID:AirlineID, timestamp=1515094526042, value=row000
row000      column=DayOfWeek:DayOfWeek, timestamp=1515094526042, value=\x00\x00\x00\x00
row000      column=DayOfMonth:DayOfMonth, timestamp=1515094526042, value=\x00\x00\x00\x00
row000      column=FlightDate:FlightDate, timestamp=1515094526042, value=\x00\x00\x00\x00
row000      column=Month:Month, timestamp=1515094526042, value=\x00\x00\x00\x00
row000      column=Quarter:Quarter, timestamp=1515094526042, value=\x00\x00\x00\x00
row000      column=UniqueCarrier:UniqueCarrier, timestamp=1515094526042, value=row000
row000      column=Year:Year, timestamp=1515094526042, value=\x00\x00\x00\x00
row001      column=AirlineID:AirlineID, timestamp=1515094526042, value=row001
row001      column=DayOfWeek:DayOfWeek, timestamp=1515094526042, value=\x00\x00\x00\x01
row001      column=DayOfMonth:DayOfMonth, timestamp=1515094526042, value=\x00\x00\x00\x01
row001      column=FlightDate:FlightDate, timestamp=1515094526042, value=\x00\x00\x00\x01
row001      column=Month:Month, timestamp=1515094526042, value=\x00\x00\x00\x01
row001      column=Quarter:Quarter, timestamp=1515094526042, value=\x00\x00\x00\x01
row001      column=UniqueCarrier:UniqueCarrier, timestamp=1515094526042, value=row001
row001      column=Year:Year, timestamp=1515094526042, value=\x00\x00\x00\x01
row002      column=AirlineID:AirlineID, timestamp=1515094526042, value=row002
row002      column=DayOfWeek:DayOfWeek, timestamp=1515094526042, value=\x00\x00\x00\x02
row002      column=DayOfMonth:DayOfMonth, timestamp=1515094526042, value=\x00\x00\x00\x02
row002      column=FlightDate:FlightDate, timestamp=1515094526042, value=\x00\x00\x00\x02
row002      column=Month:Month, timestamp=1515094526042, value=\x00\x00\x00\x02
row002      column=Quarter:Quarter, timestamp=1515094526042, value=\x00\x00\x00\x02
row002      column=UniqueCarrier:UniqueCarrier, timestamp=1515094526042, value=row002
```


HBase Data Transfer

- HBase on Spark - Load Airline Test Data from HBase

```
[root@sandbox-hdp ~]# spark-submit --class ca.training.bigdata.hbase.AirlineTestDataToHBase --driver-memory 1G --executor-memory 1G --master local[1]
/root/TrainingOnHDP/HBaseCookBook/target/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar
SPARK_MAJOR_VERSION is set to 2, using Spark2
18/01/04 19:34:13 INFO SparkContext: Running Spark version 2.2.0.2.6.3.0-235
18/01/04 19:34:15 INFO SparkContext: Submitted application: Airline Test Data to HBase
18/01/04 19:34:15 INFO SecurityManager: Changing view acls to: root
18/01/04 19:34:15 INFO SecurityManager: Changing modify acls to: root
18/01/04 19:34:15 INFO SecurityManager: Changing view acls groups to:
18/01/04 19:34:15 INFO SecurityManager: Changing modify acls groups to:
```

```
18/01/04 19:46:03 INFO CodeGenerator: Code generated in 133.466096 ms
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| col0|Year|Quarter|Month|DayOfMonth|DayOfWeek|FlightDate|UniqueCarrier|AirlineID|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|row000| 0|    0|    0|        0|        0|        0|    row000|  row000|
|row001| 1|    1|    1|        1|        1|        1|    row001|  row001|
|row002| 2|    2|    2|        2|        2|        2|    row002|  row002|
|row003| 3|    3|    3|        3|        3|        3|    row003|  row003|
|row004| 4|    4|    4|        4|        4|        4|    row004|  row004|
|row005| 5|    5|    5|        5|        5|        5|    row005|  row005|
|row006| 6|    6|    6|        6|        6|        6|    row006|  row006|
|row007| 7|    7|    7|        7|        7|        7|    row007|  row007|
|row008| 8|    8|    8|        8|        8|        8|    row008|  row008|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

```
18/01/04 19:46:03 INFO SparkContext: Invoking stop() from shutdown hook
18/01/04 19:46:03 INFO AbstractConnector: Stopped Spark@384fc774{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
18/01/04 19:46:03 INFO SparkUI: Stopped Spark web UI at http://172.17.0.2:4040
18/01/04 19:46:03 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/01/04 19:46:03 INFO MemoryStore: MemoryStore cleared
18/01/04 19:46:03 INFO BlockManager: BlockManager stopped
18/01/04 19:46:03 INFO BlockManagerMaster: BlockManagerMaster stopped
18/01/04 19:46:04 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
```

HBase Data Transfer

- HBase on Spark - Save SV Episodes to HBase

```
[root@sandbox-hdp ~]# wget https://raw.githubusercontent.com/roberthryniewicz/datasets/master/svepisodes.json svepisodes.json
--2018-01-04 20:00:06-- https://raw.githubusercontent.com/roberthryniewicz/datasets/master/svepisodes.json
Resolving raw.githubusercontent.com... 151.101.124.133
Connecting to raw.githubusercontent.com|151.101.124.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 14880 (15K) [text/plain]
Saving to: "svepisodes.json"

100%[=====] 14,880 --.-K/s in 0.1s

2018-01-04 20:00:07 (102 KB/s) - "svepisodes.json" saved [14880/14880]

--2018-01-04 20:00:07-- http://svepisodes.json/
Resolving svepisodes.json... failed: Name or service not known.
wget: unable to resolve host address "svepisodes.json"
FINISHED --2018-01-04 20:00:07--
Downloaded: 1 files, 15K in 0.1s (102 KB/s)
[root@sandbox-hdp ~]# hadoop fs -put svepisodes.json /tmp
[root@sandbox-hdp ~]#
```


HBase Data Transfer

- HBase on Spark - Save SV Episodes to HBase

```
[root@sandbox-hdp ~]# spark-submit --class ca.training.bigdata.hbase.SVEpisodesToHBase --driver-memory 1G --executor-memory 1G --master local[1] /root/TrainingOnHDP/HBaseCookBook/target/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar
SPARK_MAJOR_VERSION is set to 2, using Spark2
18/01/04 20:15:19 INFO SparkContext: Running Spark version 2.2.0.2.6.3.0-235
18/01/04 20:15:21 INFO SparkContext: Submitted application: SV Episodes to HBase
18/01/04 20:15:21 INFO SecurityManager: Changing view acls to: root
18/01/04 20:15:21 INFO SecurityManager: Changing modify acls to: root
18/01/04 20:15:21 INFO SecurityManager: Changing view acls groups to:
18/01/04 20:15:21 INFO SecurityManager: Changing modify acls groups to:
18/01/04 20:15:21 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); groups with view permissions: Set(); users with modify permissions: Set(root); groups with modify permissions: Set()
18/01/04 20:15:23 INFO Utils: Successfully started service 'sparkDriver' on port 33347.
18/01/04 20:15:23 INFO SparkEnv: Registering MapOutputTracker
18/01/04 20:15:23 INFO SparkEnv: Registering BlockManagerMaster
18/01/04 20:15:23 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
18/01/04 20:15:23 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
18/01/04 20:15:23 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-887833c8-dc0b-4a2f-8819-706aa0108691
18/01/04 20:15:24 INFO MemoryStore: MemoryStore started with capacity 413.9 MB
18/01/04 20:15:24 INFO SparkEnv: Registering OutputCommitCoordinator
18/01/04 20:15:25 INFO log: Logging initialized @9770ms
18/01/04 20:15:25 INFO Server: jetty-9.3.z-SNAPSHOT
18/01/04 20:15:25 INFO Server: Started @10437ms
```

```
root
|-- airdate: string (nullable = true)
|-- airstamp: string (nullable = true)
|-- airtime: string (nullable = true)
|-- id: long (nullable = true)
|-- name: string (nullable = true)
|-- number: long (nullable = true)
|-- runtime: long (nullable = true)
|-- season: long (nullable = true)
|-- summary: string (nullable = true)
|-- url: string (nullable = true)
```

HBase Data Transfer

- HBase on Spark - Save SV Episodes to HBase

airdate	airstamp	airtime	id	name	number	runtime	season	summary	url
2014-04-06	2014-04-06T22:00:...	22:00	10897	Minimum Viable Pr...	1	30	1	Attending an elab...	http://www.tvmaze...
2014-04-13	2014-04-13T22:00:...	22:00	10898	The Cap Table	2	30	1	After a celebrato...	http://www.tvmaze...
2014-04-20	2014-04-20T22:00:...	22:00	10899	Articles of Incor...	3	30	1	While Gavin Belso...	http://www.tvmaze...
2014-04-27	2014-04-27T22:00:...	22:00	10900	Fiduciary Duties	4	30	1	At Peter's toga p...	http://www.tvmaze...
2014-05-04	2014-05-04T22:00:...	22:00	10901	Signaling Risk	5	30	1	Erich convinces...	http://www.tvmaze...
2014-05-11	2014-05-11T22:00:...	22:00	10902	Third Party Insou...	6	30	1	Richard feels thr...	http://www.tvmaze...
2014-05-18	2014-05-18T22:00:...	22:00	10903	Proof of Concept	7	30	1	At TechCrunch Dis...	http://www.tvmaze...
2014-06-01	2014-06-01T22:00:...	22:00	10904	Optimal Tip-to-Ti...	8	30	1	Poised to compete...	http://www.tvmaze...
2015-04-12	2015-04-12T22:00:...	22:00	117409	Sand Hill Shuffle	1	30	2	Season 2 begins w...	http://www.tvmaze...
2015-04-19	2015-04-19T22:00:...	22:00	142992	Runaway Devaluation	2	30	2	Pied Piper could ...	http://www.tvmaze...
2015-04-26	2015-04-26T22:00:...	22:00	142993	Bad Money	3	30	2	Richard mulls a p...	http://www.tvmaze...
2015-05-03	2015-05-03T22:00:...	22:00	142994	The Lady	4	30	2	Richard butts hea...	http://www.tvmaze...
2015-05-10	2015-05-10T22:00:...	22:00	153965	Server Space	5	30	2	Gavin creates int...	http://www.tvmaze...
2015-05-17	2015-05-17T22:00:...	22:00	154580	Homicide	6	30	2	Erich runs into ...	http://www.tvmaze...
2015-05-24	2015-05-24T22:00:...	22:00	155129	Adult Content	7	30	2	The team fields j...	http://www.tvmaze...
2015-05-31	2015-05-31T22:00:...	22:00	155130	White Hat/Black Hat	8	30	2	Richard gets para...	http://www.tvmaze...
2015-06-07	2015-06-07T22:00:...	22:00	155199	Binding Arbitration	9	30	2	Erich wants to t...	http://www.tvmaze...
2015-06-14	2015-06-14T22:00:...	22:00	155200	Two Days of The C...	10	30	2	As the guys await...	http://www.tvmaze...
2016-04-24	2016-04-24T22:00:...	22:00	560883	Founder Friendly	1	30	3	After being uncer...	http://www.tvmaze...
2016-05-01	2016-05-01T22:00:...	22:00	668661	Two in the Box	2	30	3	The new and impro...	http://www.tvmaze...

HBase Data Transfer

- HBase on Spark - Save SV Episodes to HBase

```
hbase(main):048:0> scan 'svepisodes'
ROW
\x00\x00\x00\x00\x00\x00*\x91
\x00\x00\x00\x00\x00\x00*\x91
\x00\x00\x00\x00\x00\x00*\x91
\x00\x00\x00\x00\x00\x00*\x91
\x00\x00\x00\x00\x00\x00*\x91
\x00\x00\x00\x00\x00\x00*\x91
\x00\x00\x00\x00\x00\x00*\x91
\x00\x00\x00\x00\x00\x00*\x91
COLUMN+CELL
column=SVE:Airdate, timestamp=1515096988637, value=2014-04-06
column=SVE:Airstamp, timestamp=1515096988637, value=2014-04-06T22:00:00-04:00
column=SVE:Airtime, timestamp=1515096988637, value=22:00
column=SVE:Name, timestamp=1515096988637, value=Minimum Viable Product
column=SVE:Number, timestamp=1515096988637, value=\x00\x00\x00\x00\x00\x00\x00\x01
column=SVE:Runtime, timestamp=1515096988637, value=\x00\x00\x00\x00\x00\x00\x00\x1E
column=SVE:Season, timestamp=1515096988637, value=\x00\x00\x00\x00\x00\x00\x00\x01
column=SVE:Summary, timestamp=1515096988637, value=Attending an elaborate launch party, Richard and his computer programmer friends - Big Head, Dinesh and Gilfoyle - dream of making it big. Instead, they're living in the communal Hacker Hostel owned by former programmer Erlich, who gets to claim ten percent of anything they invent there. When it becomes clear that Richard has developed a powerful compression algorithm for his website, Pied Piper, he finds himself courted by Gavin Belson, his egomaniacal corporate boss, who offers a $10 million buyout by his firm, Hooli. But Richard holds back when well-known investor Peter Gregory makes a counteroffer.
column=SVE:Url, timestamp=1515096988637, value=http://www.tvmaze.com/episodes/10897/silicon-valley-1x01-minimum-viable-product
\x00\x00\x00\x00\x00\x00*\x91
\x00\x00\x00\x00\x00\x00*\x92
\x00\x00\x00\x00\x00\x00*\x92
\x00\x00\x00\x00\x00\x00*\x92
\x00\x00\x00\x00\x00\x00*\x92
\x00\x00\x00\x00\x00\x00*\x92
\x00\x00\x00\x00\x00\x00*\x92
\x00\x00\x00\x00\x00\x00*\x92
\x00\x00\x00\x00\x00\x00*\x92
COLUMN+CELL
column=SVE:Airdate, timestamp=1515096988637, value=2014-04-13
column=SVE:Airstamp, timestamp=1515096988637, value=2014-04-13T22:00:00-04:00
column=SVE:Airtime, timestamp=1515096988637, value=22:00
column=SVE:Name, timestamp=1515096988637, value=The Cap Table
column=SVE:Number, timestamp=1515096988637, value=\x00\x00\x00\x00\x00\x00\x00\x02
column=SVE:Runtime, timestamp=1515096988637, value=\x00\x00\x00\x00\x00\x00\x00\x1E
column=SVE:Season, timestamp=1515096988637, value=\x00\x00\x00\x00\x00\x00\x00\x01
column=SVE:Summary, timestamp=1515096988637, value=After a celebratory party at the Hacker Hostel, Richard and Erlich learn that Peter Gregory won't pay up until they deliver a viable business plan that includes a slimmed-down staff. A desperate Richard hires former Belson underling Jared, and they set about trying to trim the fat. While Gilfoyle and Dinesh prove essential, Big Head's place in the company is less certain.
\x00\x00\x00\x00\x00\x00*\x92
column=SVE:Url, timestamp=1515096988637, value=http://www.tvmaze.com/episodes/10898/silicon-valley-1x02-the-cap-table
```

HBase Data Transfer

- HBase on Spark - Load SV Episodes from HBase

```
[root@sandbox-hdp ~]# spark-submit --class ca.training.bigdata.hbase.SVEpisodesFromHBase --driver-memory 1G --executor-memory 1G --master local[1] /r
oot/TrainingOnHDP/HBaseCookBook/target/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar
SPARK_MAJOR_VERSION is set to 2, using Spark2
18/01/04 20:27:48 INFO SparkContext: Running Spark version 2.2.0.2.6.3.0-235
18/01/04 20:27:52 INFO SparkContext: Submitted application: SV Episodes from HBase
18/01/04 20:27:52 INFO SecurityManager: Changing view acls to: root
18/01/04 20:27:52 INFO SecurityManager: Changing modify acls to: root
18/01/04 20:27:52 INFO SecurityManager: Changing view acls groups to:
18/01/04 20:27:52 INFO SecurityManager: Changing modify acls groups to:
18/01/04 20:27:52 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); groups w
ith view permissions: Set(); users with modify permissions: Set(root); groups with modify permissions: Set()
```

id	airdate	airstamp	airtime	name	number	runtime	season	summary	url
10897	2014-04-06	2014-04-06T22:00:...	22:00	Minimum Viable Pr...	1	30	1	Attending an elab...	http://www.tvmaze...
10898	2014-04-13	2014-04-13T22:00:...	22:00	The Cap Table	2	30	1	After a celebrato...	http://www.tvmaze...
10899	2014-04-20	2014-04-20T22:00:...	22:00	Articles of Incor...	3	30	1	While Gavin Belso...	http://www.tvmaze...
10900	2014-04-27	2014-04-27T22:00:...	22:00	Fiduciary Duties	4	30	1	At Peter's toga p...	http://www.tvmaze...
10901	2014-05-04	2014-05-04T22:00:...	22:00	Signaling Risk	5	30	1	Erlich convinces...	http://www.tvmaze...
10902	2014-05-11	2014-05-11T22:00:...	22:00	Third Party Insou...	6	30	1	Richard feels thr...	http://www.tvmaze...
10903	2014-05-18	2014-05-18T22:00:...	22:00	Proof of Concept	7	30	1	At TechCrunch Dis...	http://www.tvmaze...
10904	2014-06-01	2014-06-01T22:00:...	22:00	Optimal Tip-to-Ti...	8	30	1	Poised to compete...	http://www.tvmaze...
117409	2015-04-12	2015-04-12T22:00:...	22:00	Sand Hill Shuffle	1	30	2	Season 2 begins w...	http://www.tvmaze...
142992	2015-04-19	2015-04-19T22:00:...	22:00	Runaway Devaluation	2	30	2	Pied Piper could ...	http://www.tvmaze...
142993	2015-04-26	2015-04-26T22:00:...	22:00	Bad Money	3	30	2	Richard mulls a p...	http://www.tvmaze...
142994	2015-05-03	2015-05-03T22:00:...	22:00	The Lady	4	30	2	Richard butts hea...	http://www.tvmaze...
153965	2015-05-10	2015-05-10T22:00:...	22:00	Server Space	5	30	2	Gavin creates int...	http://www.tvmaze...
154580	2015-05-17	2015-05-17T22:00:...	22:00	Homicide	6	30	2	Erlich runs into ...	http://www.tvmaze...
155129	2015-05-24	2015-05-24T22:00:...	22:00	Adult Content	7	30	2	The team fields j...	http://www.tvmaze...
155130	2015-05-31	2015-05-31T22:00:...	22:00	White Hat/Black Hat	8	30	2	Richard gets para...	http://www.tvmaze...
155199	2015-06-07	2015-06-07T22:00:...	22:00	Binding Arbitration	9	30	2	Erlich wants to t...	http://www.tvmaze...
155200	2015-06-14	2015-06-14T22:00:...	22:00	Two Days of The C...	10	30	2	As the guys await...	http://www.tvmaze...
560883	2016-04-24	2016-04-24T22:00:...	22:00	Founder Friendly	1	30	3	After being uncer...	http://www.tvmaze...
668661	2016-05-01	2016-05-01T22:00:...	22:00	Two in the Box	2	30	3	The new and impro...	http://www.tvmaze...

Extend HBase

- Coprocessor

```
hbase(main):015:0> disable 'airdelaydata'
0 row(s) in 2.2690 seconds

hbase(main):016:0> alter 'airdelaydata', METHOD => 'table_att', 'Coprocessor'=>'/user/root/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar|ca.train
aining.bigdata.hbase.coprocessor.AirDelayDataObserver|1073741823'
Updating all regions with the new schema...
5/5 regions updated.
Done.
0 row(s) in 2.0210 seconds

hbase(main):017:0> describe 'airdelaydata'
Table airdelaydata is DISABLED
airdelaydata, {TABLE_ATTRIBUTES => {METADATA => {'Coprocessor$1' => '/user/root/HBaseCookBook-1.0-SNAPSHOT-jar-with-dependencies.jar|ca.train
ing.bigdata.hbase.coprocessor.AirDelayDataObserver|1073741823'}}}
COLUMN FAMILIES DESCRIPTION
{NAME => 'AirlineID', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL
=> 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
{NAME => 'DayOfWeek', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL
=> 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
```

```
hbase(main):018:0> enable 'airdelaydata'
0 row(s) in 4.4040 seconds

hbase(main):019:0> get 'airdelaydata', 'row008'
COLUMN                                CELL
CF:CP                                timestamp=1515726849607, value=You can not see the detail of row008
1 row(s) in 0.0280 seconds
```