

TUTORIAL A

HBASE SHELL COMMANDS:

CONNECTING TO EMR CLUSTER:

```
Last login: Thu Apr 11 13:48:11 on console
s-MacBook-Air ~ % ssh -i ~/Downloads/MyNewMacKey.pem hadoop@ec2-3-238-235-161.compute-1.amazonaws.com
The authenticity of host 'ec2-3-238-235-161.compute-1.amazonaws.com (3.238.235.161)' can't be established.
ED25519 key fingerprint is SHA256:7yjFQCYCp4gAPkmdwDXk78HvXJlqKRSrLOfKWPR8a4.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-3-238-235-161.compute-1.amazonaws.com' (ED25519) to the list of known hosts.

A newer release of "Amazon Linux" is available.
Version 2023.4.20240319:
Version 2023.4.20240401:
Run "/usr/bin/dnf check-release-update" for full release and version update info

_#_
~\#### Amazon Linux 2023
~~\#####
~~\#####
~~\###|
~~\#/
~~V--> https://aws.amazon.com/linux/amazon-linux-2023
~~
~~--
~~--
~~-/m/'

Last login: Thu Apr 11 19:05:19 2024

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRR
E::::::::::::::::::E M:::~::~M M:::~::~M R:::~::~~::~R
EE::::::::::::::::::E M:::~::~M M:::~::~M R:::~::~~::~~::~R
E:::~::~ EEEEE M:::~::~M M:::~::~M RR:::~::~ R:::~::~
E:::~::~E M:::~::~M:::~::~M M:::~::~M R:::~::~ R:::~::~
E:::~::~EEEEEEEEEE M:::~::~M M:::~::~M M:::~::~M R:::~::~~::~~::~R
E:::~::~~::~~::~E M:::~::~M M:::~::~M M:::~::~M R:::~::~~::~~::~R
E:::~::~EEEEEEEEEE M:::~::~M M:::~::~M M:::~::~M R:::~::~~::~~::~R
E:::~::~E M:::~::~M M:::~::~M M:::~::~M R:::~::~ R:::~::~
E:::~::~E EEEEE M:::~::~M MMM M:::~::~M R:::~::~ R:::~::~
EE:::~::~EEEEEEEEEE M:::~::~M M:::~::~M R:::~::~ R:::~::~
E:::~::~~::~~::~E M:::~::~M M:::~::~M RR:::~::~ R:::~::~
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRR RRRRRR

[hadoop@ip-172-31-67-64 ~]$ hbase shell
```

STARTED USING HBASE AND CREATING TABLES:

```
hbase:001:0> status
1 active master, 0 backup masters, 1 servers, 0 dead, 2.0000 average load
Took 1.7447 seconds
hbase:002:0> status 'simple'
active master: ip-172-31-67-64.ec2.internal:16000 1712862361500
0 backup masters
1 live servers
ip-172-31-66-138.ec2.internal:16020 1712861857600
requestsPerSecond=0.0, numberOfOnlineRegions=2, usedHeapMB=50, maxHeapMB=3055, numberOfStores=4, numberOfStorefiles=0, storefileUncompressedSizeMB=0, storefileSizeMB=0, memstoreSizeMB=2, rootIndexSizeKB=0, readRequestsCount=12, filteredReadRequestsCount=0, writeRequestsCount=7, rootIndexSizeKB=0, totalStaticIndexSizeKB=0, totalStaticBloomSizeKB=0, totalCompactingKVs=0, currentCompactedKVs=0, compactionProgressPct=NaN, coprocessors=[MultiRowMutationEndpoint, RefreshFilesEndpoint]
0 dead servers
Aggregate load: 0, regions: 2
Took 0.0289 seconds
hbase:003:0> status 'summary'
1 active master, 0 backup masters, 1 servers, 0 dead, 2.0000 average load
Took 0.0253 seconds
hbase:004:0> status 'detailed'
version 2.4.17-amzn-4
0 regionsInTransition
active master: ip-172-31-67-64.ec2.internal:16000 1712862361500
0 backup masters
master coprocessors: []
1 live servers
ip-172-31-66-138.ec2.internal:16020 1712861857600
requestsPerSecond=0.0, numberOfOnlineRegions=2, usedHeapMB=44, maxHeapMB=3055, numberOfStores=4, numberOfStorefiles=1, storefileUncompressedSizeMB=1, storefileSizeMB=1, compressionRatio=1.0000, memstoreSizeMB=1, rootIndexSizeKB=1, readRequestsCount=12, filteredReadRequestsCount=0, writeRequestsCount=7, rootIndexSizeKB=1, totalStaticIndexSizeKB=1, totalStaticBloomSizeKB=1, totalCompactingKVs=0, currentCompactedKVs=0, compactionProgressPct=NaN, coprocessors=[MultiRowMutationEndpoint, RefreshFilesEndpoint]
"hbase:meta,1"
    numberOFStores=3, numberOFStorefiles=0, storeRefCount=0, storefileUncompressedSizeMB=0, lastMajorCompactionTimestamp=0, storefileSizeMB=0, memstoreSizeMB=1, readRequestsCount=6, writeRequestsCount=5, rootIndexSizeKB=0, totalStaticIndexSizeKB=0, totalStaticBloomSizeKB=0, totalCompactingKVs=0, currentCompactedKVs=0, compactionProgressPct=NaN, completeSequenceId=1, dataLocality=0
    "hbase:namespace,1712862470799.0ba2819a472bdfaab2812a6a90c2dd909"
    numberOFStores=1, numberOFStorefiles=1, storeRefCount=0, storefileUncompressedSizeMB=1024, lastMajorCompactionTimestamp=0, storefileSizeMB=1, compressionRatio=0.0010, memstoreSizeMB=0, readRequestsCount=6, writeRequestsCount=2, rootIndexSizeKB=1, totalStaticIndexSizeKB=1, totalStaticBloomSizeKB=1, totalCompactingKVs=0, currentCompactedKVs=0, compactionProgressPct=NaN, completeSequenceId=6, dataLocality=1
0 dead servers
Took 0.0926 seconds
=> #<java:LjavaUtil:Collections:UnmodifiableRandomAccessList:0x4fa0ee7e>
hbase:005:0> version
2.4.17-amzn-4, rUnknown, Fri Dec 8 01:45:07 AM UTC 2023
Took 0.0050 seconds
hbase:006:0> create 'myFirstTable', {NAME => 'familyOne', VERSIONS => 5}
Created table myFirstTable
Took 0.8407 seconds
=> Hbase::Table -> myFirstTable
```

```

=> Hbase::Table - myFirstTable
hbase:007:0> create 'mySecondTable', {NAME => 'familyOne'}, {NAME => 'familyTwo'}, {NAME => 'familyThree'}
Created table mySecondTable
Took 1.2360 seconds
=> Hbase::Table - mySecondTable
hbase:008:0> create 'myThirdTable', 'familyOne', 'familyTwo', 'familyThree'
Created table myThirdTable
Took 1.2081 seconds
=> Hbase::Table - myThirdTable
hbase:009:0> create 'myFourthTable', {NAME => 'familyOne', VERSIONS => 1, TTL => 259200, BLOCKCACHE => true}
Created table myFourthTable
Took 1.1515 seconds
=> Hbase::Table - myFourthTable
hbase:010:0> create 'myFifthTable', {NAME => 'familyOne', CONFIGURATION => {'hbase.hstore.blockingStoreFiles' => '10'}}
Created table myFifthTable
Took 1.1645 seconds
=> Hbase::Table - myFifthTable
hbase:011:0> alter 'myFirstTable', {NAME => 'familyOne', VERSIONS => 3}
Updating all regions with the new schema...
1/1 regions updated.
Done.
Took 2.3580 seconds
hbase:012:0> alter 'mySecondTable', 'familyOne', {NAME => 'familyTwo', IN_MEMORY => true}, {NAME => 'familyThree', VERSIONS => 5}
Updating all regions with the new schema...
1/1 regions updated.
Done.
Took 1.8123 seconds
hbase:013:0> alter 'myThirdTable', {NAME => 'familyOne', METHOD => 'delete'}
Updating all regions with the new schema...
1/1 regions updated.
Done.
Took 1.7440 seconds
hbase:014:0> alter 'myThirdTable', 'delete' => 'familyOne'
Updating all regions with the new schema...
1/1 regions updated.
Done.
Took 1.7512 seconds
hbase:015:0> alter 'myFirstTable', MAX_FILESIZE => '134217728'
Updating all regions with the new schema...
1/1 regions updated.
Done.
Took 1.7209 seconds
hbase:016:0> describe 'myFirstTable'
Table myFirstTable is ENABLED
myFirstTable, {TABLE_ATTRIBUTES => {MAX_FILESIZE => '134217728'}}
COLUMN FAMILIES DESCRIPTION
{NAME => 'familyOne', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '3', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

1 row(s)
Quota is disabled
Took 0.0640 seconds
hbase:017:0> disable 'mySecondTable'
Took 0.4101 seconds

```

PUT AND COUNT COMMANDS:

```

[hbase:020:0> is_disabled 'mySecondTable'
true
Took 0.0180 seconds
=> true
[hbase:021:0> is_enabled 'mySecondTable'
false
Took 0.0101 seconds
=> false
[hbase:022:0> drop 'mySecondTable'
Took 0.3692 seconds
[hbase:023:0> exists 'myFirstTable'
Table myFirstTable does exist
Took 0.0076 seconds
=> true
[hbase:024:0> list
TABLE
myFifthTable
myFirstTable
myFourthTable
myThirdTable
4 row(s)
Took 0.0463 seconds
=> ["myFifthTable", "myFirstTable", "myFourthTable", "myThirdTable"]
[hbase:025:0> list 'myFirstTable'
TABLE
myFirstTable
1 row(s)
Took 0.0127 seconds
=> ["myFirstTable"]
[hbase:026:0> put 'myFirstTable', '100', 'familyOne:price', '$45'
Took 0.1507 seconds
[hbase:027:0> put 'myFirstTable', '100', 'familyOne:quantity', '250'
Took 0.0064 seconds
[hbase:028:0> put 'myFirstTable', '101', 'familyOne:price', '$42'
Took 0.0059 seconds
[hbase:029:0> put 'myFirstTable', '101', 'familyOne:price', '$35'
Took 0.0074 seconds
[hbase:030:0> put 'myFirstTable', '101', 'familyOne:quantity', '300'
Took 0.0066 seconds
[hbase:031:0> put 'myFirstTable', '100', 'familyOne:price', '$55'
Took 0.0137 seconds
[hbase:032:0> count 'myFirstTable'
2 row(s)
Took 0.0943 seconds
=> 2
[hbase:033:0> count 'myFirstTable', INTERVAL => 100000
2 row(s)
Took 0.0188 seconds
=> 2
[hbase:034:0> count 'myFirstTable', CACHE => 1000
2 row(s)
Took 0.0074 seconds
=> 2
[hbase:035:0> count 'myFirstTable', INTERVAL => 10, CACHE => 1000
2 row(s)
Took 0.0248 seconds
=> 2
[hbase:036:0> delete 'myFirstTable', '100', 'familyOne:price', 1450245
Took 0.0104 seconds
[hbase:037:0> deleteall 'myFirstTable', '100'
Took 0.0067 seconds
[hbase:038:0> incr 'myFirstTable', '100', 'familyOne:price'
COUNTER VALUE = 1
Took 0.0409 seconds

```

DELETE, INCREMENT, GET, AND SCAN COMMANDS:

```
hbase:038:0> incr 'myFirstTable', '100', 'familyOne:price'
COUNTER VALUE = 1
Took 0.0409 seconds
hbase:039:0> incr 'myFirstTable', '100', 'familyOne:quantity', 10
COUNTER VALUE = 10
Took 0.0133 seconds
hbase:040:0> get 'myFirstTable', '100'
COLUMN                                CELL
familyOne:price                       timestamp=2024-04-11T19:25:12.283, value=\x00\x00\x00\x00\x00\x00\x01
familyOne:quantity                   timestamp=2024-04-11T19:25:20.492, value=\x00\x00\x00\x00\x00\x00\x0A
1 row(s)
Took 0.0510 seconds
hbase:041:0> get 'myFirstTable', '100', {TIMERANGE => [1400000, 1500000]}
COLUMN                                CELL
0 row(s)
Took 0.0115 seconds
hbase:042:0> get 'myFirstTable', '100', {COLUMN => 'familyOne:price'}
COLUMN                                CELL
familyOne:price                       timestamp=2024-04-11T19:25:12.283, value=\x00\x00\x00\x00\x00\x00\x01
1 row(s)
Took 0.0188 seconds
hbase:043:0> get 'myFirstTable', '101', {COLUMN => ['familyOne:price', 'familyOne:quantity']}
COLUMN                                CELL
familyOne:price                       timestamp=2024-04-11T19:23:52.222, value=$35
familyOne:quantity                   timestamp=2024-04-11T19:23:59.450, value=300
1 row(s)
Took 0.0139 seconds
hbase:044:0> get 'myFirstTable', '101', {COLUMN => 'familyOne:price', TIMESTAMP => 1450445}
COLUMN                                CELL
0 row(s)
Took 0.0177 seconds
hbase:045:0> get 'myFirstTable', '100', {COLUMN => 'familyOne:price', VERSIONS => 4}
COLUMN                                CELL
familyOne:price                       timestamp=2024-04-11T19:25:12.283, value=\x00\x00\x00\x00\x00\x00\x01
1 row(s)
Took 0.0141 seconds
hbase:046:0> scan 'myFirstTable', {COLUMNS => ['familyOne:price', 'familyOne:quantity'], LIMIT => 5, STARTROW => '100'}
ROW                                    COLUMN+CELL
100                                    column=familyOne:price, timestamp=2024-04-11T19:25:12.283, value=\x00\x00\x00\x00\x00\x00\x01
100                                    column=familyOne:quantity, timestamp=2024-04-11T19:25:20.492, value=\x00\x00\x00\x00\x00\x00\x0A
101                                    column=familyOne:price, timestamp=2024-04-11T19:23:52.222, value=$35
101                                    column=familyOne:quantity, timestamp=2024-04-11T19:23:59.450, value=300
2 row(s)
Took 0.0150 seconds
hbase:047:0> scan 'myFirstTable', {COLUMNS => ['familyOne:price', 'familyOne:quantity'], LIMIT => 10, STARTROW => '100', STOPROW => 105}
ROW                                    COLUMN+CELL
ERROR: undefined method `to_java_bytes' for 105:Integer
Did you mean?  to_java

For usage try 'help "scan"'

Took 0.3460 seconds
hbase:048:0> scan 'myFirstTable', {COLUMNS => 'familyOne:price', TIMERANGE => [1303668804, 1303668904]}
ROW                                    COLUMN+CELL
0 row(s)
Took 0.0069 seconds
```

TUTORIAL B

CREATING AND POPULATING TABLES WITH HBASE:

```

Took 0.0246 seconds
hbase:057:0> create 'customer','ID',{NAME=>'Address',VERSIONS=>3},'Demographics',{NAME=>'Payments',VERSIONS=>3}
Created table customer
Took 1.1511 seconds
=> Hbase::Table - customer
hbase:058:0> put 'customer', '171', 'Cust_Name:FirstName', 'Madelyn'
Took 0.0146 seconds
hbase:059:0> put 'customer', '171', 'Cust_Name:LastName', 'Hensley'
Took 0.0150 seconds
hbase:060:0> put 'customer', '171', 'Address:StreetAddress', '10075 Thierer Plaza'
Took 0.0152 seconds
hbase:061:0> put 'customer', '171', 'Address:StreetAddress', '615 McCallie Ave'
Took 0.0058 seconds
hbase:062:0> put 'customer', '171', 'Address:City', 'New York'
Took 0.0075 seconds
hbase:063:0> put 'customer', '171', 'Address:State', 'New York'
Took 0.0097 seconds
hbase:064:0> put 'customer', '171', 'Address:State', 'Tennessee'
Took 0.0057 seconds
hbase:065:0> put 'customer', '171', 'Demographics:Gender', 'M'
Took 0.0061 seconds
hbase:066:0> put 'customer', '171', 'Payments:Internet', 'True'
Took 0.0195 seconds
hbase:067:0> put 'customer', '171', 'Payments:Internet', 'True'
Took 0.0077 seconds
hbase:068:0> put 'customer', '171', 'Payments:Mobile', 'False'
Took 0.0096 seconds
hbase:069:0> put 'customer', '171', 'Payments:Mobile', 'True'
Took 0.0072 seconds
hbase:070:0> put 'customer', '171', 'Payments:Mobile', 'False'
Took 0.0071 seconds
hbase:071:0> put 'customer', '172', 'Cust_Name:FirstName', 'Lony'
Took 0.0077 seconds
hbase:072:0> put 'customer', '172', 'Cust_Name:LastName', 'Foster'
Took 0.0055 seconds
hbase:073:0> put 'customer', '172', 'Address:State', 'Texas'
Took 0.0047 seconds
hbase:074:0> put 'customer', '172', 'Address:Zip', '13498'
Took 0.0074 seconds
hbase:075:0> put 'customer', '172', 'Demographics:BirthYear', '1981'
Took 0.0068 seconds
hbase:076:0> put 'customer', '172', 'Demographics:Gender', 'F'
Took 0.0065 seconds
hbase:077:0> put 'customer', '172', 'Payments:CreditCard', 'True'
Took 0.0075 seconds
hbase:078:0> put 'customer', '172', 'Payments:Mobile', 'False'
Took 0.0137 seconds
hbase:079:0> put 'customer', '172', 'Payments:Mobile', 'False'
Took 0.0059 seconds
hbase:080:0> put 'customer', '172', 'Payments:Mobile', 'False'
Took 0.0060 seconds
hbase:081:0> put 'customer', '171', 'Address:Zip', '81377'
Took 0.0064 seconds
hbase:082:0> put 'customer', '171', 'Address:Zip', '37415'
Took 0.0102 seconds
hbase:083:0>

```

```

hbase:083:0> scan 'customer'
ROW COLUMN+CELL
171 column=Address:City, timestamp=2024-04-10T23:59:52.578, value=New York
171 column=Address:State, timestamp=2024-04-11T00:00:07.809, value=Tennessee
171 column=Address:StreetAddress, timestamp=2024-04-10T23:59:42.230, value=615 McCallie Ave
171 column=Address:Zip, timestamp=2024-04-11T00:02:24.297, value=37415
171 column=Cust_Name:FirstName, timestamp=2024-04-10T23:59:13.070, value=Madelyn
171 column=Cust_Name:LastName, timestamp=2024-04-10T23:59:25.876, value=Hensley
171 column=Demographics:Gender, timestamp=2024-04-11T00:00:13.877, value=M
171 column=Payments:Internet, timestamp=2024-04-11T00:00:31.972, value=True
171 column=Payments:Mobile, timestamp=2024-04-11T00:00:52.684, value=False
172 column=Address:State, timestamp=2024-04-11T00:01:16.128, value=Texas
172 column=Address:Zip, timestamp=2024-04-11T00:01:23.951, value=13498
172 column=Cust_Name:FirstName, timestamp=2024-04-11T00:01:02.530, value=Lony
172 column=Cust_Name:LastName, timestamp=2024-04-11T00:01:09.155, value=Foster
172 column=Demographics:BirthYear, timestamp=2024-04-11T00:01:29.987, value=1981
172 column=Demographics:Gender, timestamp=2024-04-11T00:01:40.116, value=F
172 column=Payments:CreditCard, timestamp=2024-04-11T00:01:47.325, value=True
172 column=Payments:Mobile, timestamp=2024-04-11T00:02:11.313, value=False
2 row(s)
Took 0.1227 seconds

```