

Zhastay Yeltay

HBASE PRACTICE

Note: The requirements for the practical task are just texts or scripts/codes are in bold, the other simple italic and yellowed text and screenshots are my answers.

CONTENTS

1.	Notes regarding the practice	2
2.	Connect to the the HBase environment	2
3.	General Details (10).....	3
4.	Table and Data Creation (30)	4
5.	Qurey Data (60)	5
6.	Delete Table (10)	6

1. NOTES REGARDING THE PRACTICE

- When you see the dollar sign in a command you need to execute, note this is just a command prompt indication. You do not need to actually write the “\$”, as it is not part of the command.
- For all practices below – write all the commands you have used in a “Practice answers document”.

When you are complete, you can submit this document for review.

2. CONNECT TO THE THE HBASE ENVIRONMENT

- Verify the “hbase” environment is up and running: `docker ps -a`
- Open a BASH session to the practice environment `docker exec -it hbase /bin/bash`

3. GENERAL DETAILS (10)

- View the HBase related scripts in the HBase “Bin” directory
 - Hint: The location of the base directory should be in “/opt”
 - You should see ~25 scripts

```
root@6860663283f5:/opt# ls
hbase hbase-server replace-hostname
root@6860663283f5:/opt# cd hbase
root@6860663283f5:/opt/hbase# ls
CHANGES.md RELEASENOTES.md bin conf hbase-webapps lib logs
root@6860663283f5:/opt/hbase# cd bin
root@6860663283f5:/opt/hbase/bin# ls
considerAsDead.sh graceful_stop.sh hbase-common.sh hbase-daemons.sh local-master-backup.sh region_mover.rb
replication start-hbase.sh zookeepers.sh
draining_servers.rb hbase hbase-config.sh hbase-jruby local-regionserver.sh region_status.rb
rolling-restart.sh stop-hbase.sh
get-active-master.rb hbase-cleanup.sh hbase-daemon.sh hibr.rb master-backup.sh regionserver.sh
shutdown_regionserver.rb test
root@6860663283f5:/opt/hbase/bin#
```

- Enter HBase Shell

```
root@6860663283f5:/opt/hbase/bin# hbase shell
2024-02-04 12:01:12,585 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.1.2, r1dfc418f77801fbfb59a125756891b9100c1fc6d, Sun Dec 30 21:45:09 PST 2018
Took 0.0059 seconds
hbase(main):001:0>
```

- Get the HBase status in different levels
 - Retrieve base status, simple status, and detailed status

\$status

\$status 'simple'

\$status 'detailed'

```
hbase(main):001:0> status 1
1 active master, 0 backup masters, 1 servers, 0 dead, 2.0000 average load
Took 0.7750 seconds
hbase(main):002:0> status 'simple' 2
active master: 6860663283f5:16000 1707047448763
0 backup masters
1 live servers
6860663283f5:16020 1707047452222
requestsPerSecond=0.0, numberOfOnlineRegions=2, usedHeapMB=64, maxHeapMB=2907, numberOfStores=4, numberOfStorefiles=4, storefileUncompressedSizeMB=0, storefileSizeMB=0, memstoreSizeMB=0, storefileIndexSizeKB=0, readRequestsCount=11, filteredReadRequestsCount=0, writeRequestsCount=7, rootIndexSizeKB=0, totalStaticIndexSizeKB=0, totalStaticBloomSizeKB=0, totalCompactingKVs=0, currentCompactedKVs=0, compactionProgressPct=NaN, coprocessors=[MultiRowMutationEndPoint]
0 dead servers
Aggregate load: 0, regions: 2
Took 0.0482 seconds
hbase(main):003:0> status 'detailed' 3
version 2.1.2
0 regionsInTransition
active master: 6860663283f5:16000 1707047448763
0 backup masters
master coprocessors: []
1 live servers
6860663283f5:16020 1707047452222
requestsPerSecond=0.0, numberOfOnlineRegions=2, usedHeapMB=64, maxHeapMB=2907, numberOfStores=4, numberOfStorefiles=4, storefileUncompressedSizeMB=0, storefileSizeMB=0, memstoreSizeMB=0, storefileIndexSizeKB=0, readRequestsCount=11, filteredReadRequestsCount=0, writeRequestsCount=7, rootIndexSizeKB=0, totalStaticIndexSizeKB=0, totalStaticBloomSizeKB=0, totalCompactingKVs=0, currentCompactedKVs=0, compactionProgressPct=NaN, coprocessors=[MultiRowMutationEndPoint]
"hbase:meta,1"
numberOfStores=3, numberOfStorefiles=3, storefileUncompressedSizeMB=0, lastMajorCompactionTimestamp=0, storefileSizeMB=0, memstoreSizeMB=0, readRequestsCount=9, writeRequestsCount=5, rootIndexSizeKB=0, totalStaticIndexSizeKB=0, totalStaticBloomSizeKB=0, totalCompactingKVs=0, currentCompactedKVs=0, compactionProgressPct=NaN, completeSequenceId=9, dataLocality=0.0
"hbase:namespace,1707047465011.adf77489040c33826fd2014ec00b511e."
numberOfStores=1, numberOfStorefiles=1, storefileUncompressedSizeMB=0, lastMajorCompactionTimestamp=0, storefileSizeMB=0, memstoreSizeMB=0, readRequestsCount=2, writeRequestsCount=2, rootIndexSizeKB=0, totalStaticIndexSizeKB=0, totalStaticBloomSizeKB=0, totalCompactingKVs=0, currentCompactedKVs=0, compactionProgressPct=NaN, completeSequenceId=6, dataLocality=0.0
0 dead servers
Took 0.0444 seconds
=> #Java:Javutil::collections::UnmodifiableRandomAccessList:0x503556cb>
hbase(main):004:0>
```

- o List all filters

```
hbase(main):007:0> show_filters
DependentColumnFilter
KeyOnlyFilter
ColumnCountGetFilter
SingleColumnValueFilter
PrefixFilter
SingleColumnValueExcludeFilter
FirstKeyOnlyFilter
ColumnRangeFilter
ColumnValueFilter
TimestampsFilter
FamilyFilter
QualifierFilter
ColumnPrefixFilter
RowFilter
MultipleColumnPrefixFilter
InclusiveStopFilter
PageFilter
ValueFilter
ColumnPaginationFilter
Took 0.0255 seconds
=> #<Java::JavaUtil::HashMap::KeySet:0x10f20d38>
hbase(main):008:0> _
```

- o List all tables

```
hbase(main):009:0> list
TABLE
0 row(s)
Took 0.1492 seconds
=> []
hbase(main):010:0>
```

4. TABLE AND DATA CREATION (30)

- Create a table with the name: “employees” with the following column families
 - personal_data
 - Store 2 versions for this column
 - professional_data
 - Store 4 versions for this column

```
create 'employees', {NAME => 'personal_data', VERSIONS => 2}, {NAME => 'professional_data', VERSIONS => 4}
```

```
hbase(main):013:0> create 'employees', {NAME => 'personal_data', VERSIONS => 2}, {NAME => 'professional_data', VERSIONS => 4}
Created table employees
Took 1.0660 seconds
=> Hbase::Table - employees
hbase(main):014:0>
```

- List all tables

```
hbase(main):014:0> list
TABLE
employees
1 row(s)
Took 0.0309 seconds
=> ["employees"]
hbase(main):015:0>
```

- Insert data for ten employees
 - The id of each employee must be a unique value
 - Insert employee's id to be 1 - 10
 - Fill the following data
 - personal_data
 - first_name
 - surname
 - age
 - professional_data
 - role
 - expertise

```
put 'employees', '1', 'personal_data:first_name', 'Albert'
put 'employees', '1', 'personal_data:surname', 'Einstein'
put 'employees', '1', 'personal_data:age', '23'
put 'employees', '1', 'professional_data:role', 'Junior Developer'
put 'employees', '1', 'professional_data:expertise', 'Data Engineer'
```

```
hbase(main):015:0> put 'employees', '1', 'personal_data:first_name', 'Albert'
Took 0.5560 seconds
hbase(main):016:0> put 'employees', '1', 'personal_data:surname', 'Einstein'
Took 0.0046 seconds
hbase(main):017:0> put 'employees', '1', 'personal_data:age', '23'
Took 0.0053 seconds
hbase(main):018:0> put 'employees', '1', 'professional_data:role', 'Junior Developer'
Took 0.0086 seconds
hbase(main):019:0> put 'employees', '1', 'professional_data:expertise', 'Data Engineer'
Took 0.0069 seconds
hbase(main):020:0>
```

- Scan employee table to print all rows

```
hbase(main):085:0> scan 'employees'
ROW                                COLUMN+CELL
1                                  column=personal_data:age, timestamp=1707065226097, value=23
1                                  column=personal_data:first_name, timestamp=1707065225979, value=Albert
1                                  column=personal_data:surname, timestamp=1707065226050, value=Einstein
1                                  column=professional_data:expertise, timestamp=1707065237422, value=Data Engineer
1                                  column=professional_data:role, timestamp=1707065226153, value=Junior Developer
10                                 column=personal_data:age, timestamp=1707065413285, value=42
10                                 column=personal_data:first_name, timestamp=1707065413246, value=Gilbert
10                                 column=personal_data:surname, timestamp=1707065413266, value=George
10                                 column=professional_data:expertise, timestamp=1707065413336, value=Database Manager
10                                 column=professional_data:role, timestamp=1707065413314, value=Senior Developer
2                                  column=personal_data:age, timestamp=1707065411918, value=33
2                                  column=personal_data:first_name, timestamp=1707065411837, value=Alan
2                                  column=personal_data:surname, timestamp=1707065411880, value=Walker
2                                  column=professional_data:expertise, timestamp=1707065411997, value=Data Software Engineer
2                                  column=professional_data:role, timestamp=1707065411954, value=Middle Developer
3                                  column=personal_data:age, timestamp=1707065412137, value=26
3                                  column=personal_data:first_name, timestamp=1707065412056, value=Robert
3                                  column=personal_data:surname, timestamp=1707065412097, value=Wayne
3                                  column=professional_data:expertise, timestamp=1707065412213, value=Data Analytics Engineer
3                                  column=professional_data:role, timestamp=1707065412175, value=Strong Middle Developer
4                                  column=personal_data:age, timestamp=1707065412326, value=28
4                                  column=personal_data:first_name, timestamp=1707065412268, value=Hello
4                                  column=personal_data:surname, timestamp=1707065412294, value=Helson
4                                  column=professional_data:expertise, timestamp=1707065412399, value=Product Manager
4                                  column=professional_data:role, timestamp=1707065412361, value=Senior Manager
5                                  column=personal_data:age, timestamp=1707065412497, value=35
5                                  column=personal_data:first_name, timestamp=1707065412440, value=Henry
5                                  column=personal_data:surname, timestamp=1707065412468, value=Johnson
5                                  column=professional_data:expertise, timestamp=1707065412562, value=Project Manager
5                                  column=professional_data:role, timestamp=1707065412525, value=Middle Manager
6                                  column=personal_data:age, timestamp=1707065412663, value=40
6                                  column=personal_data:first_name, timestamp=1707065412603, value=Steven
6                                  column=personal_data:surname, timestamp=1707065412629, value=Kosinski
6                                  column=professional_data:expertise, timestamp=1707065412731, value=Data Engineer
6                                  column=professional_data:role, timestamp=1707065412699, value=Lead Developer
7                                  column=personal_data:age, timestamp=1707065412846, value=31
7                                  column=personal_data:first_name, timestamp=1707065412770, value=Carlson
7                                  column=personal_data:surname, timestamp=1707065412808, value=FromRoof
7                                  column=professional_data:expertise, timestamp=1707065412909, value=Big Data
7                                  column=professional_data:role, timestamp=1707065412881, value=Junior Developer
8                                  column=personal_data:age, timestamp=1707065413017, value=40
8                                  column=personal_data:first_name, timestamp=1707065412949, value=Octavian
8                                  column=personal_data:surname, timestamp=1707065412989, value=August
8                                  column=professional_data:expertise, timestamp=1707065413076, value=Big Data
8                                  column=professional_data:role, timestamp=1707065413046, value=Technical Director
9                                  column=personal_data:age, timestamp=1707065413166, value=46
9                                  column=personal_data:first_name, timestamp=1707065413121, value=Julian
9                                  column=personal_data:surname, timestamp=1707065413147, value=Caesar
9                                  column=professional_data:expertise, timestamp=1707065413217, value=Product Manager
9                                  column=professional_data:role, timestamp=1707065413193, value=Junior Manager
10 row(s)
Took 0.0902 seconds
hbase(main):086:0>
```

- Get all data of employee with id 7

```
hbase(main):086:0> get 'employees', '7'
COLUMN                                CELL
personal_data:age                     timestamp=1707065412846, value=31
personal_data:first_name               timestamp=1707065412770, value=Carlson
personal_data:surname                 timestamp=1707065412808, value=FromRoof
professional_data:expertise            timestamp=1707065412909, value=Big Data
professional_data:role                 timestamp=1707065412881, value=Junior Developer
1 row(s)
Took 0.1172 seconds
hbase(main):087:0>
```

- Update age and role of employee number 3

```
put 'employees', '3', 'personal_data:age', '55'
put 'employees', '3', 'professional_data:role', 'Senior HR'
hbase(main):087:0> put 'employees', '3', 'personal_data:age', '55'
Took 0.0202 seconds
hbase(main):088:0> put 'employees', '3', 'professional_data:role', 'Senior HR'
Took 0.0051 seconds
hbase(main):089:0>
```

- Get all data of employee with id 3 and make sure updates applied

```
hbase(main):089:0> get 'employees', '3'
COLUMN                                CELL
personal_data:age                     timestamp=1707065812422, value=55
personal_data:first_name               timestamp=1707065412056, value=Robert
personal_data:surname                 timestamp=1707065412097, value=Wayne
professional_data:expertise            timestamp=1707065412213, value=Data Analytics Engineer
professional_data:role                 timestamp=1707065814245, value=Senior HR
1 row(s)
Took 0.0134 seconds
hbase(main):090:0>
```

5. QUREY DATA (60)

- Query all record in employees table

```
hbase(main):090:0> scan 'employees'
ROW COLUMN+CELL
1 column=personal_data:age, timestamp=1707065226097, value=23
1 column=personal_data:first_name, timestamp=1707065225979, value=Albert
1 column=personal_data:surname, timestamp=1707065226050, value=Einstein
1 column=professional_data:expertise, timestamp=1707065237422, value=Data Engineer
1 column=professional_data:role, timestamp=1707065226153, value=Junior Developer
10 column=personal_data:age, timestamp=1707065413285, value=42
10 column=personal_data:first_name, timestamp=1707065413246, value=Gilbert
10 column=personal_data:surname, timestamp=1707065413266, value=George
10 column=professional_data:expertise, timestamp=1707065413336, value=Database Manager
10 column=professional_data:role, timestamp=1707065413314, value=Senior Developer
2 column=personal_data:age, timestamp=1707065411918, value=33
2 column=personal_data:first_name, timestamp=1707065411837, value=Alan
2 column=personal_data:surname, timestamp=1707065411880, value=Walker
2 column=professional_data:expertise, timestamp=1707065411997, value=Data Software Engineer
2 column=professional_data:role, timestamp=1707065411954, value=Middle Developer
3 column=personal_data:age, timestamp=1707065812422, value=55
3 column=personal_data:first_name, timestamp=1707065412056, value=Robert
3 column=personal_data:surname, timestamp=1707065412097, value=Wayne
3 column=professional_data:expertise, timestamp=1707065412213, value=Data Analytics Engineer
3 column=professional_data:role, timestamp=1707065814245, value=Senior HR
4 column=personal_data:age, timestamp=1707065412326, value=28
4 column=personal_data:first_name, timestamp=1707065412268, value=Hello
4 column=personal_data:surname, timestamp=1707065412294, value=Helloson
4 column=professional_data:expertise, timestamp=1707065412399, value=Product Manager
4 column=professional_data:role, timestamp=1707065412361, value=Senior Manager
5 column=personal_data:age, timestamp=1707065412497, value=35
5 column=personal_data:first_name, timestamp=1707065412440, value=Henry
5 column=personal_data:surname, timestamp=1707065412468, value=Johnson
5 column=professional_data:expertise, timestamp=1707065412562, value=Project Manager
5 column=professional_data:role, timestamp=1707065412525, value=Middle Manager
6 column=personal_data:age, timestamp=1707065412663, value=40
6 column=personal_data:first_name, timestamp=1707065412603, value=Steven
6 column=personal_data:surname, timestamp=1707065412629, value=Kosinski
6 column=professional_data:expertise, timestamp=1707065412731, value=Data Engineer
6 column=professional_data:role, timestamp=1707065412699, value=Lead Developer
7 column=personal_data:age, timestamp=1707065412846, value=31
7 column=personal_data:first_name, timestamp=1707065412770, value=Carlson
7 column=personal_data:surname, timestamp=1707065412808, value=fromRoof
7 column=professional_data:expertise, timestamp=1707065412909, value=Big Data
7 column=professional_data:role, timestamp=1707065412881, value=Junior Developer
8 column=personal_data:age, timestamp=1707065413017, value=40
8 column=personal_data:first_name, timestamp=1707065412949, value=Octavian
8 column=personal_data:surname, timestamp=1707065412989, value=August
8 column=professional_data:expertise, timestamp=1707065413076, value=Big Data
8 column=professional_data:role, timestamp=1707065413046, value=Technical Director
9 column=personal_data:age, timestamp=1707065413166, value=46
9 column=personal_data:first_name, timestamp=1707065413121, value=Julian
9 column=personal_data:surname, timestamp=1707065413147, value=Caesar
9 column=professional_data:expertise, timestamp=1707065413217, value=Product Manager
9 column=professional_data:role, timestamp=1707065413193, value=Junior Manager
10 row(s)
Took 0.1148 seconds
hbase(main):091:0>
```

- Get all data of employee with id 3 and the 3 last versions of his column families: personal_data, professional_data

```
get 'employees', '3', {COLUMNS => ['personal_data', 'professional_data'], VERSIONS => 3}
hbase(main):098:0> get 'employees', '3', {COLUMNS => ['personal_data', 'professional_data'], VERSIONS => 3}
COLUMN CELL
personal_data:age timestamp=1707065812422, value=55
personal_data:age timestamp=1707065412137, value=26
personal_data:first_name timestamp=1707065412056, value=Robert
personal_data:surname timestamp=1707065412097, value=Wayne
professional_data:expertise timestamp=1707065412213, value=Data Analytics Engineer
professional_data:role timestamp=1707065814245, value=Senior HR
professional_data:role timestamp=1707065412175, value=Strong Middle Developer
1 row(s)
Took 0.0158 seconds
hbase(main):099:0>
```

OR

```
scan 'employees', {FILTER => "RowFilter(=,'binary:3')", VERSIONS => 3}
```

```
hbase(main):095:0> scan 'employees', {FILTER => "RowFilter(=,'binary:3')", VERSIONS => 3}
ROW COLUMN+CELL
3 column=personal_data:age, timestamp=1707065812422, value=55
3 column=personal_data:age, timestamp=1707065412137, value=26
3 column=personal_data:first_name, timestamp=1707065412056, value=Robert
3 column=personal_data:surname, timestamp=1707065412097, value=Wayne
3 column=professional_data:expertise, timestamp=1707065412213, value=Data Analytics Engineer
3 column=professional_data:role, timestamp=1707065814245, value=Senior HR
3 column=professional_data:role, timestamp=1707065412175, value=Strong Middle Developer
1 row(s)
Took 0.0292 seconds
hbase(main):096:0>
```


- Get all data of employees with age bigger or equals to 40

```
scan 'employees', { FILTER => "SingleColumnValueFilter('personal_data', 'age', >=, 'binary:40')"
```

```
hbase(main):099:0> scan 'employees', { FILTER => "SingleColumnValueFilter('personal_data', 'age', >=, 'binary:40')"
```

ROW	COLUMN+CELL
10	column=personal_data:age, timestamp=1707065413285, value=42
10	column=personal_data:first_name, timestamp=1707065413246, value=Gilbert
10	column=personal_data:surname, timestamp=1707065413266, value=George
10	column=professional_data:expertise, timestamp=1707065413336, value=Database Manager
10	column=professional_data:role, timestamp=1707065413314, value=Senior Developer
3	column=personal_data:age, timestamp=1707065812422, value=55
3	column=personal_data:first_name, timestamp=1707065412056, value=Robert
3	column=personal_data:surname, timestamp=1707065412097, value=Wayne
3	column=professional_data:expertise, timestamp=1707065412213, value=Data Analytics Engineer
3	column=professional_data:role, timestamp=1707065814245, value=Senior HR
6	column=personal_data:age, timestamp=1707065412663, value=40
6	column=personal_data:first_name, timestamp=1707065412603, value=Steven
6	column=personal_data:surname, timestamp=1707065412629, value=Kosinski
6	column=professional_data:expertise, timestamp=1707065412731, value=Data Engineer
8	column=professional_data:role, timestamp=1707065412699, value=Lead Developer
8	column=personal_data:age, timestamp=1707065413017, value=40
8	column=personal_data:first_name, timestamp=1707065412949, value=Octavian
8	column=personal_data:surname, timestamp=1707065412989, value=August
8	column=professional_data:expertise, timestamp=1707065413076, value=Big Data
9	column=professional_data:role, timestamp=1707065413046, value=Technical Director
9	column=personal_data:age, timestamp=1707065413166, value=46
9	column=personal_data:first_name, timestamp=1707065413121, value=Julian
9	column=personal_data:surname, timestamp=1707065413147, value=Caesar
9	column=professional_data:expertise, timestamp=1707065413217, value=Product Manager
9	column=professional_data:role, timestamp=1707065413193, value=Junior Manager

```
5 row(s)
Took 0.0485 seconds
hbase(main):100:0>
```

- Get only role value of all employees with age bigger than 35

```
scan 'employees', { FILTER => "SingleColumnValueFilter('personal_data', 'age', >, 'binary:35') AND MultipleColumnPrefixFilter('professional_data', 'role')"
```

```
hbase(main):104:0> scan 'employees', { FILTER => "SingleColumnValueFilter('personal_data', 'age', >, 'binary:35') AND MultipleColumnPrefixFilter('professional_data', 'role')"
```

ROW	COLUMN+CELL
10	column=professional_data:role, timestamp=1707065413314, value=Senior Developer
3	column=professional_data:role, timestamp=1707065814245, value=Senior HR
8	column=professional_data:role, timestamp=1707065412699, value=Lead Developer
9	column=professional_data:role, timestamp=1707065413046, value=Technical Director
9	column=professional_data:role, timestamp=1707065413193, value=Junior Manager

```
5 row(s)
Took 0.0160 seconds
hbase(main):105:0>
```

- Count the number of all employees

```
count 'employees'
```

```
hbase(main):103:0> count 'employees'
10 row(s)
Took 0.0171 seconds
=> 10
hbase(main):104:0>
```

- Count the number of employees with age less than 40

```
count 'employees', { FILTER => "SingleColumnValueFilter('personal_data', 'age', <, 'binary:40')"
```

```
hbase(main):105:0> count 'employees', { FILTER => "SingleColumnValueFilter('personal_data', 'age', <, 'binary:40')"
```

```
5 row(s)
Took 0.0074 seconds
=> 5
hbase(main):106:0>
```

- Delete the newer age (that updated in topic 4) for employee with id 3

```
delete 'employees', '3', 'personal_data:age'
```

```
hbase(main):106:0> delete 'employees', '3', 'personal_data:age'
Took 0.0242 seconds
hbase(main):107:0> _
```

- Get the data of employee with id 3 and validate his age reverted to first value

```
get 'employees', '3', {COLUMN => 'personal_data:age', VERSIONS => 3}
```

```
hbase(main):107:0> get 'employees', '3', {COLUMN => 'personal_data:age', VERSIONS => 3}
```

COLUMN	CELL
personal_data:age	timestamp=1707065412137, value=26

```
1 row(s)
Took 0.0105 seconds
hbase(main):108:0>
```

6. DELETE TABLE (10)

- o Delete table employees

disable 'employees'

drop 'employees'

exists 'employees'

```
hbase(main):108:0> exists 'employees'
Table employees does exist
Took 0.0248 seconds
=> true
hbase(main):109:0> disable 'employees'
Took 0.8247 seconds
hbase(main):110:0> drop 'employees'
Took 0.2776 seconds
hbase(main):111:0> exists 'employees'
Table employees does not exist
Took 0.0040 seconds
=> false
hbase(main):112:0>
```

1
2
3
✓