

Week 6

Felipe Rodriguez

Bellevue University

DSC 650 Big Data

Professor Nasheb Ismaily

January 21, 2024

Exercise 1: Create a table named 'students' with a column family 'details'.

```
[hbase(main):005:0> create 'students', 'details'
Created table students
Took 11.4392 seconds
=> Hbase::Table - students
hbase(main):006:0> █
```

Exercise 2: Verify that the table has been created.

```
[hbase(main):006:0> list
TABLE
students
1 row(s)
Took 0.0519 seconds
=> ["students"]
hbase(main):007:0> █
```

Exercise 3: Add data to the 'students' table. Let's assume each student has a unique ID, a first name, and a last name.

```
[hbase(main):007:0> put 'students', '1', 'details:firstName', 'John'
Took 0.3108 seconds
[hbase(main):008:0> put 'students', '1', 'details:lastName', 'Doe'
Took 0.0086 seconds
[hbase(main):009:0> get 'students'. '1'
SyntaxError: (hbase):9: syntax error, unexpected tSTRING_BEG
get 'students'. '1'

[hbase(main):010:0> get 'students', '1'
COLUMN                                CELL
details:firstName                      timestamp=2024-01-21T18:39:42.325, value=John
details:lastName                      timestamp=2024-01-21T18:39:58.941, value=Doe
1 row(s)
Took 0.0659 seconds
hbase(main):011:0> █
```

Exercise 6: Add sample data to the ‘orders’ table using the composite key:

Exercise 7: Query the ‘orders’ table to retrieve details of all orders placed by the customer with ID ‘101’.

```
hbase(main):012:0> put 'orders', '101:20230806', 'orderDetails:item', 'Laptop'
Took 0.0245 seconds
hbase(main):013:0> put 'orders', '102:20230806', 'orderDetails:item', 'Smartphone'
Took 0.0092 seconds
hbase(main):014:0> scan 'orders', {STARTROW => '101:', ENDROW => '101:~'}
ROW                                COLUMN+CELL
101:20230806                        column=orderDetails:item, timestamp=2024-01-21T18:42:06.173, value=Laptop
1 row(s)
Took 0.0234 seconds
hbase(main):015:0>
```

Exercise 8: Generate random data for the ‘students’ table.

```
hbase(main):020:0> (2..100).each do |i|
hbase(main):021:1*   first_name = "Student#{i}"
hbase(main):022:1>   last_name = "LastName#{i}"
hbase(main):023:1>   put 'students', "#{i}", 'details:firstName', first_name
hbase(main):024:1>   put 'students', "#{i}", 'details:lastName', last_name
hbase(main):025:1> end
Took 0.0133 seconds
Took 0.0073 seconds
Took 0.0063 seconds
Took 0.0073 seconds
Took 0.0063 seconds
Took 0.0067 seconds
Took 0.0050 seconds
Took 0.0122 seconds
Took 0.0064 seconds
Took 0.0068 seconds
Took 0.0062 seconds
Took 0.0070 seconds
Took 0.0059 seconds
Took 0.0065 seconds
Took 0.0062 seconds
Took 0.0061 seconds
```

Exercise 9: Scan the ‘students’ table to verify data insertion.

```
Took 0.0044 seconds
hbase(main):027:0> scan 'students'
ROW                                COLUMN+CELL
1                                column=details:firstName, timestamp=2024-01-21T18:39:42.325, value=John
1                                column=details:lastName, timestamp=2024-01-21T18:39:58.941, value=Doe
10                                column=details:firstName, timestamp=2024-01-21T18:46:28.153, value=Student10
10                                column=details:lastName, timestamp=2024-01-21T18:46:28.162, value=LastName10
100                                column=details:firstName, timestamp=2024-01-21T18:46:29.278, value=Student100
100                                column=details:lastName, timestamp=2024-01-21T18:46:29.281, value=LastName100
11                                column=details:firstName, timestamp=2024-01-21T18:46:28.169, value=Student11
11                                column=details:lastName, timestamp=2024-01-21T18:46:28.181, value=LastName11
12                                column=details:firstName, timestamp=2024-01-21T18:46:28.190, value=Student12
12                                column=details:lastName, timestamp=2024-01-21T18:46:28.199, value=LastName12
13                                column=details:firstName, timestamp=2024-01-21T18:46:28.207, value=Student13
13                                column=details:lastName, timestamp=2024-01-21T18:46:28.214, value=LastName13
14                                column=details:firstName, timestamp=2024-01-21T18:46:28.222, value=Student14
14                                column=details:lastName, timestamp=2024-01-21T18:46:28.230, value=LastName14
15                                column=details:firstName, timestamp=2024-01-21T18:46:28.238, value=Student15
15                                column=details:lastName, timestamp=2024-01-21T18:46:28.266, value=LastName15
16                                column=details:firstName, timestamp=2024-01-21T18:46:28.279, value=Student16
16                                column=details:lastName, timestamp=2024-01-21T18:46:28.288, value=LastName16
17                                column=details:firstName, timestamp=2024-01-21T18:46:28.297, value=Student17
17                                column=details:lastName, timestamp=2024-01-21T18:46:28.305, value=LastName17
18                                column=details:firstName, timestamp=2024-01-21T18:46:28.313, value=Student18
18                                column=details:lastName, timestamp=2024-01-21T18:46:28.320, value=LastName18
19                                column=details:firstName, timestamp=2024-01-21T18:46:28.326, value=Student19
19                                column=details:lastName, timestamp=2024-01-21T18:46:28.333, value=LastName19
2                                column=details:firstName, timestamp=2024-01-21T18:46:28.018, value=Student2
2                                column=details:lastName, timestamp=2024-01-21T18:46:28.027, value=LastName2
20                                column=details:firstName, timestamp=2024-01-21T18:46:28.341, value=Student20
```

HBASE Data Manipulation

1. Update First Names

```
hbase(main):029:0> (2..50).each do |i|
hbase(main):030:1*   first_name = "Scholar#{i}"
hbase(main):031:1>   put 'students', "#{i}", 'details:firstName', first_name
hbase(main):032:1> end
Took 0.0133 seconds
Took 0.0043 seconds
Took 0.0073 seconds
Took 0.0039 seconds
Took 0.0034 seconds
Took 0.0034 seconds
```

```
hbase(main):033:0> scan 'students'

COLUMN=CELL
ROW
1 column=details:firstName, timestamp=2024-01-21T18:39:42.325, value=John
1 column=details:lastName, timestamp=2024-01-21T18:39:58.941, value=Doe
10 column=details:firstName, timestamp=2024-01-21T19:08:59.156, value=Scholar10
10 column=details:lastName, timestamp=2024-01-21T18:46:28.162, value=LastName10
100 column=details:firstName, timestamp=2024-01-21T18:46:29.278, value=Student100
100 column=details:lastName, timestamp=2024-01-21T18:46:29.281, value=LastName100
11 column=details:firstName, timestamp=2024-01-21T19:08:59.160, value=Scholar11
11 column=details:lastName, timestamp=2024-01-21T18:46:28.181, value=LastName11
12 column=details:firstName, timestamp=2024-01-21T19:08:59.164, value=Scholar12
12 column=details:lastName, timestamp=2024-01-21T18:46:28.199, value=LastName12
13 column=details:firstName, timestamp=2024-01-21T19:08:59.169, value=Scholar13
13 column=details:lastName, timestamp=2024-01-21T18:46:28.214, value=LastName13
14 column=details:firstName, timestamp=2024-01-21T19:08:59.174, value=Scholar14
14 column=details:lastName, timestamp=2024-01-21T18:46:28.230, value=LastName14
15 column=details:firstName, timestamp=2024-01-21T19:08:59.178, value=Scholar15
15 column=details:lastName, timestamp=2024-01-21T18:46:28.266, value=LastName15
16 column=details:firstName, timestamp=2024-01-21T19:08:59.182, value=Scholar16
16 column=details:lastName, timestamp=2024-01-21T18:46:28.288, value=LastName16
17 column=details:firstName, timestamp=2024-01-21T19:08:59.186, value=Scholar17
17 column=details:lastName, timestamp=2024-01-21T18:46:28.305, value=LastName17
18 column=details:firstName, timestamp=2024-01-21T19:08:59.190, value=Scholar18
18 column=details:lastName, timestamp=2024-01-21T18:46:28.320, value=LastName18
19 column=details:firstName, timestamp=2024-01-21T19:08:59.193, value=Scholar19
19 column=details:lastName, timestamp=2024-01-21T18:46:28.333, value=LastName19
2 column=details:firstName, timestamp=2024-01-21T19:08:59.112, value=Scholar2
2 column=details:lastName, timestamp=2024-01-21T18:46:28.027, value=LastName2
20 column=details:firstName, timestamp=2024-01-21T19:08:59.197, value=Scholar20
20 column=details:lastName, timestamp=2024-01-21T18:46:28.348, value=LastName20
21 column=details:firstName, timestamp=2024-01-21T19:08:59.201, value=Scholar21
21 column=details:lastName, timestamp=2024-01-21T18:46:28.369, value=LastName21
22 column=details:firstName, timestamp=2024-01-21T19:08:59.204, value=Scholar22
22 column=details:lastName, timestamp=2024-01-21T18:46:28.389, value=LastName22
23 column=details:firstName, timestamp=2024-01-21T19:08:59.207, value=Scholar23
23 column=details:lastName, timestamp=2024-01-21T18:46:28.409, value=LastName23
24 column=details:firstName, timestamp=2024-01-21T19:08:59.210, value=Scholar24
24 column=details:lastName, timestamp=2024-01-21T18:46:28.429, value=LastName24
25 column=details:firstName, timestamp=2024-01-21T19:08:59.213, value=Scholar25
25 column=details:lastName, timestamp=2024-01-21T18:46:28.442, value=LastName25
26 column=details:firstName, timestamp=2024-01-21T19:08:59.217, value=Scholar26
```

2. Add a Middle Name

```
hbase(main):034:0> (51..75).each do |i|
hbase(main):035:1*   middle_name = "Midname#{i}"
hbase(main):036:1>   put 'students', "#{i}", 'details:midName', middle_name
hbase(main):037:1> end
Took 0.0109 seconds
Took 0.0044 seconds
Took 0.0041 seconds
Took 0.0045 seconds
Took 0.0057 seconds
Took 0.0037 seconds
Took 0.0037 seconds
Took 0.0029 seconds
Took 0.0038 seconds
Took 0.0026 seconds
Took 0.0022 seconds
```

3. Modify Last Names

```
hbase(main):039:0> (76..100).each do |i|
hbase(main):040:1*   last_name = "LastName#{i}_Modified"
hbase(main):041:1>   put 'students', "#{i}", 'details:lastName', last_name
hbase(main):042:1> end
Took 0.0137 seconds
Took 0.0098 seconds
Took 0.0051 seconds
Took 0.0030 seconds
Took 0.0075 seconds
Took 0.0029 seconds
```

```
76      column=details:lastName, timestamp=2024-01-21T19:21:48.513, value=LastName76_Modified
77      column=details:firstName, timestamp=2024-01-21T18:46:29.066, value=Student77
77      column=details:lastName, timestamp=2024-01-21T19:21:48.521, value=LastName77_Modified
78      column=details:firstName, timestamp=2024-01-21T18:46:29.075, value=Student78
78      column=details:lastName, timestamp=2024-01-21T19:21:48.532, value=LastName78_Modified
79      column=details:firstName, timestamp=2024-01-21T18:46:29.085, value=Student79
79      column=details:lastName, timestamp=2024-01-21T19:21:48.537, value=LastName79_Modified
8      column=details:firstName, timestamp=2024-01-21T19:08:59.148, value=Scholar8
8      column=details:lastName, timestamp=2024-01-21T18:46:28.129, value=LastName8
80      column=details:firstName, timestamp=2024-01-21T18:46:29.095, value=Student80
80      column=details:lastName, timestamp=2024-01-21T19:21:48.541, value=LastName80_Modified
81      column=details:firstName, timestamp=2024-01-21T18:46:29.106, value=Student81
81      column=details:lastName, timestamp=2024-01-21T19:21:48.549, value=LastName81_Modified
82      column=details:firstName, timestamp=2024-01-21T18:46:29.116, value=Student82
82      column=details:lastName, timestamp=2024-01-21T19:21:48.552, value=LastName82_Modified
83      column=details:firstName, timestamp=2024-01-21T18:46:29.125, value=Student83
83      column=details:lastName, timestamp=2024-01-21T19:21:48.556, value=LastName83_Modified
84      column=details:firstName, timestamp=2024-01-21T18:46:29.133, value=Student84
84      column=details:lastName, timestamp=2024-01-21T19:21:48.561, value=LastName84_Modified
85      column=details:firstName, timestamp=2024-01-21T18:46:29.143, value=Student85
85      column=details:lastName, timestamp=2024-01-21T19:21:48.565, value=LastName85_Modified
86      column=details:firstName, timestamp=2024-01-21T18:46:29.151, value=Student86
86      column=details:lastName, timestamp=2024-01-21T19:21:48.569, value=LastName86_Modified
```


4. Bulk delete

```
hbase(main):048:0> (90..100).each do |i|
hbase(main):049:1*   delete 'students', "#{i}", 'details:firstName'
hbase(main):050:1>   delete 'students', "#{i}", 'details:lastName'
hbase(main):051:1> end
Took 0.0177 seconds
Took 0.0056 seconds
Took 0.0054 seconds
Took 0.0070 seconds
```

```
hbase(main):053:0> (90..100).each do |i|
hbase(main):054:1*   delete 'students', "#{i}", 'details:lastName'
hbase(main):055:1> end
Took 0.0057 seconds
Took 0.0050 seconds
Took 0.0038 seconds
Took 0.0035 seconds
```

```
85 column=details:firstName, timestamp=2024-01-21T18:46:29.143, value=Student85
86 column=details:lastName, timestamp=2024-01-21T18:46:29.151, value=LastName85_Modified
86 column=details:firstName, timestamp=2024-01-21T18:46:29.151, value=Student86
87 column=details:lastName, timestamp=2024-01-21T18:46:29.163, value=LastName86_Modified
87 column=details:firstName, timestamp=2024-01-21T18:46:29.163, value=Student87
88 column=details:lastName, timestamp=2024-01-21T18:46:29.172, value=LastName87_Modified
88 column=details:firstName, timestamp=2024-01-21T18:46:29.172, value=Student88
89 column=details:lastName, timestamp=2024-01-21T18:46:29.181, value=LastName88_Modified
89 column=details:firstName, timestamp=2024-01-21T18:46:29.181, value=Student89
90 column=details:lastName, timestamp=2024-01-21T18:46:29.181, value=LastName89_Modified
90 column=details:firstName, timestamp=2024-01-21T19:08:59.151, value=Scholar9
91 column=details:lastName, timestamp=2024-01-21T18:46:28.145, value=LastName9
89 row(s)
Took 0.0988 seconds
hbase(main):057:0> █
```

5. Data Retrieval

```
[hbase(main):068:0> get 'students', '40'
COLUMN
  details:firstName      CELL
                        timestamp=2024-01-21T19:08:59.270, value=Scholar40
  details:lastName      CELL
                        timestamp=2024-01-21T18:46:28.654, value=LastName40
1 row(s)
Took 0.0113 seconds
[hbase(main):069:0> get 'students', '60'
COLUMN
  details:firstName      CELL
                        timestamp=2024-01-21T18:46:28.884, value=Student60
  details:lastName      CELL
                        timestamp=2024-01-21T18:46:28.896, value=LastName60
  details:midName        CELL
                        timestamp=2024-01-21T19:13:23.589, value=Midname60
1 row(s)
Took 0.0082 seconds
[hbase(main):070:0> get 'students', '80'
COLUMN
  details:firstName      CELL
                        timestamp=2024-01-21T18:46:29.095, value=Student80
  details:lastName      CELL
                        timestamp=2024-01-21T19:21:48.541, value=LastName80_Modified
1 row(s)
Took 0.0093 seconds
[hbase(main):071:0> get 'students', '90'
COLUMN
  0 row(s)
Took 0.0043 seconds
hbase(main):072:0> █
```