DSC 650 - Big Data

Week 6 Assignment

Eyram Kueviakoe

04/17/2024

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase

Use "exit" to quit this interactive shell.

For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell

Version 2.3.6, r7414579f2620fca6b75146c29ab2726fc4643ac9, Wed Jul 28 22:24:42 UTC 2021

Took 0.0021 seconds

hbase(main):001:0> create 'students', 'details'

Created table students

Took 1.9336 seconds

=> Hbase::Table - students

hbase(main):002:0>
```

Screenshot of the table creation command and its output

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
hbase(main):056:0* list
TABLE
students
l row(s)
Took 0.0411 seconds
=> ["students"]
hbase(main):057:0> []
```

Screenshot of the tables listed in HBase

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase

hbase(main):056:0* list

TABLE

students
1 row(s)

Took 0.0411 seconds
=> ["students"]
hbase(main):057:0> put 'students', 'l', 'details:firstName', 'John'

Took 0.3007 seconds
hbase(main):058:0> put 'students', 'l', 'details:lastName', 'Doe'

Took 0.0108 seconds
hbase(main):059:0> []
```

Screenshot of the commands used to add data and their outputs.

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase

hbase (main):059:0> get 'students', '1'

COLUMN

details:firstName
timestamp=2024-04-17T18:41:10.770, value=John
details:lastName
timestamp=2024-04-17T18:41:21.486, value=Doe

1 row(s)

Took 0.0691 seconds
hbase (main):060:0>
```

Screenshot of the query (get 'students', '1') and its output

```
hbase (main):060:0>
hbase (main):060:0> create 'orders', 'orderDetails'
Created table orders
Took 0.6625 seconds

> Hbase: (Table - orders', '101:20230806', 'orderDetails: item', 'Laptop'
Took 0.0378 seconds
hbase (main):062:0> put 'orders', '102:20230806', 'orderDetails:item', 'Smartphone'
Took 0.0077 seconds
hbase (main):063:0> but 'orders', '102:20230806', 'orderDetails:item', 'Smartphone'
Took 0.0077 seconds
hbase (main):063:0> but 'orders', 'STARTROW => '101:', ENDROW => '101:~')
ROW
COLUMN+CELL
101:20230806
Column=orderDetails:item, timestamp=2024-04-17T18:45:30.420, value=Laptop
1 row(s)
Took 0.0187 seconds
hbase (main):064:0>
```

Screenshot of the commands used to query the data with composite key and their outputs

```
dsc650-kueviakoe: ~/dsc650-kueviakoe: √dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase & c650-kueviakoe: √dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
 hbase(main):066:0*
 hbase(main):067:0*
 hbase(main):068:0* (2..100).each do |i|
 hbase(main):069:1*
                                                                                      first name = "Student#{i}"
                                                                                             last name = "LastName#{i}"
 hbase (main):070:1>
                                                                                                put 'students', "#{i}", 'details:firstName', first_name
put 'students', "#{i}", 'details:lastName', last_name
 hbase(main):071:1>
 hbase(main):072:1>
 hbase(main):073:1> end
 Took 0.0163 seconds
 Took 0.0071 seconds
 Took 0.0068 seconds
 Took 0.0062 seconds
 Took 0.0056 seconds
 Took 0.0056 seconds
 Took 0.0052 seconds
 Took 0.0057 seconds
 Took 0.0053 seconds
 Took 0.0053 seconds
 Took 0.0058 seconds
 Took 0.0064 seconds
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
                                                                                                                                          COLUMN+CELL
                                                                                                                                          column=details:firstName, timestamp=2024-04-17T18:41:10.770, value=John
                                                                                                                                          column=details:lastName, timestamp=2024-04-17T18:41:21.486, value=Doe
                                                                                                                                         column=details:lastName, timestamp=2024-04-17718:49:17.870, value=Student10 column=details:lastName, timestamp=2024-04-17718:49:17.878, value=LastName10 column=details:firstName, timestamp=2024-04-17718:49:18.987, value=Student100 column=details:lastName, timestamp=2024-04-17718:49:18.995, value=LastName10
                                                                                                                                          column=details:firstName, timestamp=2024-04-17718:49:17.885, value=Student11 column=details:lastName, timestamp=2024-04-17718:49:17.892, value=LastName11
                                                                                                                                         column-details:lastName, timestamp=2024-04-1/T18:49:17.892, Value=LastName11 column-details:firstName, timestamp=2024-04-17T18:49:17.905, value=Student12 column-details:lastName, timestamp=2024-04-17T18:49:17.905, value=LastName12 column-details:firstName, timestamp=2024-04-17T18:49:17.911, value=Student13 column-details:firstName, timestamp=2024-04-17T18:49:17.918, value=LastName13 column-details:firstName, timestamp=2024-04-17T18:49:17.929, value=Student14 column-details:lastName, timestamp=2024-04-17T18:49:17.936, value=LastName14 column-details:firstName, timestamp=2024-04-17T18:49:17.934, value=Student15 column-details:firstName, timestamp=2024-04-17T18:49:17.935, value=LastName18 column-details:firstName, timestamp=2024-04-17T18:49:17.935, value=LastName18 column-details:firstName, timestamp=2024-04-17T18:49:17.935, value=LastName18 column-details:firstName, timestamp=2024-04-17T18:49:17.935, value=LastName18 column-details:firstName timestamp=2024-04-17T18:49:17.935, value=2024-04-17T18 column-details:firstName timestamp=2024-04-17T18 column-details:firstName timestamp=2024-04-17T18 column-details:firstName timestamp=2024-04-17T18 column-details:firstName timestamp=2024-04-17T18 column-details:firstName timestamp=2024-04-17T18 column-details:firstNa
    12
12
    14
14
15
                                                                                                                                         column=details:firstName, timestamp=2024-04-17T18:49:17.943, value=Student15 column=details:lastName, timestamp=2024-04-17T18:49:17.949, value=LastName15 column=details:firstName, timestamp=2024-04-17T18:49:17.955, value=Student16 column=details:lastName, timestamp=2024-04-17T18:49:17.961, value=LastName16 column=details:firstName, timestamp=2024-04-17T18:49:17.967, value=Student17 column=details:firstName, timestamp=2024-04-17T18:49:17.977, value=Student18 column=details:firstName, timestamp=2024-04-17T18:49:17.987, value=LastName18 column=details:firstName, timestamp=2024-04-17T18:49:18.003, value=Student18 column=details:firstName, timestamp=2024-04-17T18:49:18.003, value=Student18 column=details:firstName, timestamp=2024-04-17T18:49:18.003, value=Student19 column=details:firstName, timestamp=2024-04-17T18:49:18.003, value=Student19
                                                                                                                                          Column=details:Iffs.Name, timestamp=2024-04-17T18:49:18.003, Value=Student19 column=details:firstName, timestamp=2024-04-17T18:49:17.755, value=Student2 column=details:firstName, timestamp=2024-04-17T18:49:17.767, value=Student2 column=details:firstName, timestamp=2024-04-17T18:49:18.018, value=Student20
                                                                                                                                          column=details:lastName, timestamp=2024-04-17T18:49:18.939, value=LastName96
                                                                                                                                          column=details:IastName, timestamp=2024-04-17T18:49:18.9949, value=LastName97 column=details:IastName, timestamp=2024-04-17T18:49:18.955, value=EastName97 column=details:IastName, timestamp=2024-04-17T18:49:18.961, value=Student98 column=details:IastName, timestamp=2024-04-17T18:49:18.965, value=LastName98
                                                                                                                                           column=details:firstName, timestamp=2024-04-17T18:49:18.971, value=Student99
                                                                                                                                          column=details:lastName, timestamp=2024-04-17T18:49:18.980, value=LastName99
  00 row(s)
```

Screenshot of the commands used for data generation and their outputs

HBase data manipulation

1. Update First Names: For students with IDs from 2 to 50, change the first name prefix from Student to Scholar. For instance, Student3 should become Scholar3.

Code:

```
(2..50).each do |i|

first_name = "Scholar#{i}"

put 'students', "#{i}", 'details:firstName', first_name

end
```

ena

rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase Took 0.1600 seconds hbase(main):084:0> hbase(main):084:0> (2..50).each do |i| first name = "Scholar#{i}" hbase(main):085:1* hbase(main):086:1> put 'students', "#{i}", 'details:firstName', first name hbase(main):087:1> end Took 0.0092 seconds Took 0.0039 seconds Took 0.0075 seconds Took 0.0040 seconds Took 0.0044 seconds Took 0.0034 seconds Took 0.0036 seconds

```
ook 0.0044 seconds
                                                                                                                                 COLUMN+CELL
                                                                                                                                column=details:lastName, timestamp=2024-04-17T18:41:21.486, value=Doe column=details:firstName, timestamp=2024-04-17T19:14:58.635, value=Scholar10
                                                                                                                                 column=details:lastName, timestamp=2024-04-17T18:49:17.878, value=LastName10 column=details:firstName, timestamp=2024-04-17T18:49:18.987, value=Student100 column=details:lastName, timestamp=2024-04-17T18:49:18.995, value=LastName100
                                                                                                                                 column=details:firstName, timestamp=2024-04-17T19:14:58.640, value=Scholar11
                                                                                                                                column-details:firstName, timestamp=2024-04-17T19:14:58.644, value=Scholar12 column-details:firstName, timestamp=2024-04-17T19:14:58.644, value=Scholar12 column-details:firstName, timestamp=2024-04-17T19:14:58.648, value=Scholar13 column-details:firstName, timestamp=2024-04-17T19:14:58.648, value=Scholar13 column-details:lastName, timestamp=2024-04-17T18:49:17.918, value=LastName13
                                                                                                                                column=details:firstName, timestamp=2024-04-17T19:14:58.653, value=Scholar14 column=details:lastName, timestamp=2024-04-17T18:49:17.936, value=LastName14 column=details:firstName, timestamp=2024-04-17T19:14:58.658, value=Scholar15
                                                                                                                                  column=details:lastName, timestamp=2024-04-17T18:49:17.949, value=LastName15
                                                                                                                                column-details:firstName, timestamp=2024-04-17T19:14:58.662, value=Scholar16 column-details:lastName, timestamp=2024-04-17T19:14:58.667, value=Scholar17 column-details:firstName, timestamp=2024-04-17T19:14:58.667, value=Scholar17 column-details:lastName, timestamp=2024-04-17T19:14:78.667, value=Scholar17 column-details:lastName, timestamp=2024-04-17T18:49:17.977, value=LastName17
                                                                                                                                 column=details:firstName, timestamp=2024-04-17T19:14:58.671, value=Scholar18 column=details:lastName, timestamp=2024-04-17T18:49:17.995, value=LastName18 column=details:firstName, timestamp=2024-04-17T19:14:58.675, value=Scholar19
                                                                                                                                 column=details:lastName, timestamp=2024-04-17T18:49:18.010, value=LastName
                                                                                                                                column-details:firstName, timestamp=2024-04-17T19:14:58.590, value=Scholar2 column-details:firstName, timestamp=2024-04-17T19:14:58.590, value=Scholar2 column-details:firstName, timestamp=2024-04-17T19:14:58.679, value=Scholar20 column-details:lastName, timestamp=2024-04-17T19:14:58.679, value=Scholar20 column-details:lastName, timestamp=2024-04-17T19:14:89:18.024, value=LastName20
                                                                                                                                 column=details:firstName, timestamp=2024-04-17T19:14:58.683, value=Scholar21 column=details:lastName, timestamp=2024-04-17T18:49:18.040, value=LastName21 column=details:firstName, timestamp=2024-04-17T19:14:58.689, value=Scholar22
                                                                                                                                 column=details:lastName, timestamp=2024-04-17T18:49:18.071, value=LastName2
                                                                                                                                column=details:IastName, timestamp=2024-04-17T19:14:58.694, value=Scholar23 column=details:IastName, timestamp=2024-04-17T19:14:58.696, value=EastName23 column=details:IastName, timestamp=2024-04-17T19:14:58.697, value=Scholar24 column=details:IastName, timestamp=2024-04-17T19:14:58.697, value=Scholar24 column=details:IastName, timestamp=2024-04-17T18:49:18.107, value=LastName24
                                                                                                                                 column=details:firstName, timestamp=2024-04-17T19:14:58.701, value=Scholar25
column=details:lastName, timestamp=2024-04-17T18:49:18.120, value=LastName25
                                                                                                                                  column=details:firstName, timestamp=2024-04-17T19:14:58.706, value=Scholar2
```

2. Add a Middle Name: For students with IDs from 51 to 75, add a middle name column under the details column family. The middle name should follow the pattern MidName#{i}.

Code:

```
(51..75).each do |i|

middle_name = "MidName#{i}"

put 'students', "#{i}", 'details:middleName', middle_name

end
```

```
🧬 rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
Took 0.1767 seconds
hbase(main):089:0> (51..75).each do |i|
                      middle name = "MidName#{i}"
hbase(main):090:1*
                      put 'students', "#{i}", 'details:middleName', middle_name
hbase(main):091:1>
hbase(main):092:1> end
Took 0.0258 seconds
Took 0.0102 seconds
Took 0.0070 seconds
Took 0.0087 seconds
Took 0.0099 seconds
Took 0.0041 seconds
Took 0.0046 seconds
Took 0.0043 seconds
Took 0.0036 seconds
Took 0.0051 seconds
Took 0.0038 seconds
Took 0.0036 seconds
Took 0.0035 seconds
Took 0.0076 seconds
Took 0.0039 seconds
Took 0.0031 seconds
Took 0.0036 seconds
Took 0.0041 seconds
Took 0.0107 seconds
Took 0.0034 seconds
Took 0.0032 seconds
Took 0.0038 seconds
Took 0.0047 seconds
Took 0.0068 seconds
=> 51..75
```

3. Modify Last Names: For students with IDs from 76 to 100, append _Modified to the last name. So, LastName76 should be updated to LastName76_Modified.

Code:

```
(76..100).each do |i|

put 'students', "#{i}", 'details:lastName', "LastName#{i}_Modified"

end
```

```
hbase(main):094:0> (76..100).each do |i|
hbase(main):095:1* put 'students', "#{i}", 'details:lastName', "LastName#{i} Modified"
hbase (main):096:1> end
Took 0.0218 seconds
Took 0.0037 seconds
Took 0.0033 seconds
Took 0.0030 seconds
Took 0.0040 seconds
Took 0.0027 seconds
Took 0.0028 seconds
Took 0.0027 seconds
Took 0.0025 seconds
Took 0.0025 seconds
Took 0.0032 seconds
Took 0.0034 seconds
Took 0.0029 seconds
Took 0.0034 seconds
Took 0.0034 seconds
Took 0.0034 seconds
Took 0.0030 seconds
Took 0.0030 seconds
Took 0.0032 seconds
Took 0.0048 seconds
Took 0.0036 seconds
Took 0.0055 seconds
Took 0.0038 seconds
hbase(main):097:0>
```

```
column-details:firstName, timestamp=2024-04-17718:49:18.722, value=Student79
column-details:lastName, timestamp=2024-04-17720:214.622, value=StatName79 Modified
column-details:lastName, timestamp=2024-04-17720:214.622, value=StatName79 Modified
column-details:lastName, timestamp=2024-04-17718:49:17.851, value=StatName89
column-details:lastName, timestamp=2024-04-17718:49:17.851, value=StatName89
column-details:lastName, timestamp=2024-04-17718:49:18.752, value=StatName80
column-details:lastName, timestamp=2024-04-17718:49:18.747, value=StatName81
column-details:firstName, timestamp=2024-04-17718:49:18.747, value=StatName81
column-details:firstName, timestamp=2024-04-17718:49:18.747, value=Student81
column-details:firstName, timestamp=2024-04-17718:49:18.747, value=Student81
column-details:firstName, timestamp=2024-04-17718:49:18.749, value=Student83
column-details:firstName, timestamp=2024-04-17718:49:18.749, value=Student83
column-details:firstName, timestamp=2024-04-17718:49:18.749, value=Student84
column-details:firstName, timestamp=2024-04-17718:49:18.749, value=Student84
column-details:firstName, timestamp=2024-04-17718:49:18.789, value=Student84
column-details:firstName, timestamp=2024-04-17718:49:18.800, value=Student84
column-details:firstName, timestamp=2024-04-17718:49:18.800, value=Student86
column-details:firstName, timestamp=2024-04-17720:02:14.664, value=LastName86 Modified
column-details:firstName, timestamp=2024-04-17720:02:14.669, value=Student87
column-details:firstName, timestamp=2024-04-17720:02:14.669, value=Student87
column-details:firstName, timestamp=2024-04-17720:02:14.669, value=Student87
column-details:firstName, timestamp=2024-04-17720:02:14.669, value=Student89
column-details:firs
```

4. Bulk Delete: Delete all the details for students with IDs from 90 to 100.

Code:

end

(90..100).each do |i|

deleteall 'students', "#{i}"

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase

hbase(main):045:1* deleteall 'students', "#{i}"

hbase(main):046:1> end

Took 0.0053 seconds

Took 0.0040 seconds

Took 0.0045 seconds

Took 0.0069 seconds

Took 0.0038 seconds

Took 0.0061 seconds

Took 0.0060 seconds

Took 0.0060 seconds

Took 0.0060 seconds
```

5. Data Retrieval: After all modifications, retrieve and display the details for students with IDs 40, 60, 80, and 90 to verify changes.

Code:

=> 90..100

```
[40, 60, 80, 90].each do |id|
get 'students', "#{id}"
end
```

```
hbase(main):049:0> [40, 60, 80, 90].each do |id| hbase(main):050:1* get 'students', "#{id}" hbase(main):051:1> end
 COLUMN
 details:firstName
details:lastName
                                                                                   timestamp=2024-04-17T19:14:58.783, value=Scholar40 timestamp=2024-04-17T18:49:18.318, value=LastName40
1 row(s)
Took 0.0777 seconds
  details:firstName
details:lastName
                                                                                   timestamp=2024-04-17T18:49:18.547, value=Student60 timestamp=2024-04-17T18:49:18.552, value=LastName60 timestamp=2024-04-17T19:28:45.788, value=MidName60
  details:middleName
1 row(s)
Took 0.0161 seconds
 COLUMN
                                                                                   timestamp=2024-04-17T18:49:18.735, value=Student80
  details:firstName
  details:lastName
                                                                                   timestamp=2024-04-17T20:02:14.625, value=LastName80_Modified
1 row(s)
Took 0.0109 seconds
 COLUMN
Took 0.0064 seconds
=> [40, 60, 80, 90]
hbase(main):052:0>
```