Cassandra

1. Подключаемся к Cassandra на worker-2:

/cassandra/bin/cqlsh 10.0.0.18

/cassandra/bin/cqlsh

```
[student3_16@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ /cassandra/bin/cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.8 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh>
```

2. Создаем пространство ключей:

CREATE KEYSPACE student3_16

WITH REPLICATION = {

'class' : 'SimpleStrategy', 'replication_factor' : 1 }

```
cqlsh> CREATE KEYSPACE student3_16
   ... WITH REPLICATION = {
   ... 'class' : 'SimpleStrategy', 'replication_factor' : 1 }
   ...
   ... ;
cqlsh>
```

3. Создаем таблицу и вставляем значения:

CREATE TABLE animals

(id int,

name text,

color text,

size text,

primary key (id, name, color));

insert into animals (id, name, color, size) values (3, 'Deer', 'White', 'Big');

insert into animals (id, name, color, size) values (1, 'Rabbit', 'Gray', 'Small');

insert into animals (id, name, color, size) values (4, 'Cow', 'Black');

insert into animals (id, name, color, size) values (6, 'Cat', 'Orange', 'Small');

```
cqlsh> CREATE TABLE animals
   ... (id int,
   ... name text,
   ... size text,
   ... primary key (id));
InvalidRequest: Error from server: code=2200 [Invalid query] message="No keyspace has been specified. USE a keyspace, or explicitly specify keyspace.tablename"
cqlsh> insert into animals (id, name, size)
   ... values (3, 'Deer', 'Big');
InvalidRequest: Error from server: code=2200 [Invalid query] message="No keyspace has been specified. USE a keyspace, or explicitly specify keyspace.tablename"
cqlsh> use student3_16;
cqlsh:student3_16> CREATE TABLE animals
               ... (id int,
               ... name text,
               ... size text,
               ... primary key (id));
cqlsh:student3_16> insert into animals (id, name, size)
               ... values (3, 'Deer', 'Big');
cqlsh:student3_16>
```

4. Проверяем как работает фильтрация:

select * from animals

where id = 3 and name = '12321';

```
cqlsh:student3_16> select * from animals
               ... where id = 3 and name = '12321';
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this qu
ery despite the performance unpredictability, use ALLOW FILTERING"
cqlsh:student3_16> select * from animals;

 id | name | size
----+------+------
  3 | Deer |  Big

(1 rows)
cqlsh:student3_16> select * from animals where id = 3 and name = 'Big';
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this qu
ery despite the performance unpredictability, use ALLOW FILTERING"
cqlsh:student3_16> select * from animals where id = 3;

 id | name | size
----+------+------
  3 | Deer |  Big

(1 rows)
cqlsh:student3_16> select * from animals where id = 3 and name = 'Deer';
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this qu
ery despite the performance unpredictability, use ALLOW FILTERING"
cqlsh:student3_16> select * from animals where id = 3 and size = 'Big';
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this qu
ery despite the performance unpredictability, use ALLOW FILTERING"
cqlsh:student3_16> select * from animals where id = 3 and size = 'Big' ALLOW FILTERING;

 id | name | size
----+------+------
  3 | Deer |  Big

(1 rows)
cqlsh:student3_16> select * from animals where id = 3 and name = '12321' ALLOW FILTERING;

 id | name | size
----+------+------

(0 rows)
cqlsh:student3_16>
```

```
cqlsh:student3_16> drop table animals;
cqlsh:student3_16> CREATE TABLE animals
               ... (id int,
               ... name text,
               ... color,
               ... size text,
               ... primary key (id, name, color));
SyntaxException: line 4:5 no viable alternative at input ',' (... int,name text,color[,]...)
cqlsh:student3_16> CREATE TABLE animals
               ... (id int,
               ... name text,
               ... color text,
               ... size text,
               ... primary key (id, name, color));
cqlsh:student3_16> insert into animals (id, name, color, size) values (3, 'Deer', 'White', 'Big');
cqlsh:student3_16> insert into animals (id, name, color, size) values (1, 'Rabbit', 'Gray', 'Small');
cqlsh:student3_16> insert into animals (id, name, color, size) values (4, 'Cow', 'Black');
InvalidRequest: Error from server: code=2200 [Invalid query] message="Unmatched column names/values"
cqlsh:student3_16> insert into animals (id, name, color) values (4, 'Cow', 'Black');
cqlsh:student3_16> insert into animals (id, name, color, size) values (6, 'Cat', 'Orange', 'Small');
cqlsh:student3_16> select * from animals;

 id | name   | color  | size
----+--------+--------+-------
  1 | Rabbit |   Gray | Small
  4 |    Cow |  Black |  null
  6 |    Cat | Orange | Small
  3 |   Deer |  White |   Big

(4 rows)
cqlsh:student3_16> select * from animals where id=4 and name = 'Cow' and color = null;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Invalid null value in condition for column color"
cqlsh:student3_16> select * from animals where id=6 and name = 'Cat' and color = null;
[1]+  Stopped                 /cassandra/bin/cqlsh
[student3_16@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ /cassandra/bin/cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.8 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> use student3_16;
cqlsh:student3_16> select * from animals where id=4 and name='Cow' and color='Black';

 id | name | color | size
----+------+-------+------
  4 |  Cow | Black | null

(1 rows)
cqlsh:student3_16> select * from animals where id=4 and name='Cow';

 id | name | color | size
----+------+-------+------
  4 |  Cow | Black | null

(1 rows)
cqlsh:student3_16>
```

```
cqlsh:student3_16> select * from animals where id=4 and name='Cow';

 id | name | color | size
----+------+-------+------
  4 |  Cow | Black | null

(1 rows)
cqlsh:student3_16> select * from animals where id=4;

 id | name | color | size
----+------+-------+------
  4 |  Cow | Black | null

(1 rows)
cqlsh:student3_16> select * from animals where id=4 and color='Black';
InvalidRequest: Error from server: code=2200 [Invalid query] message="PRIMARY KEY column "color" cannot be restricted as preceding column "name" is not restricted"
cqlsh:student3_16> select * from animals where color='Black';
InvalidRequest: Error from server: code=2200 [Invalid query] message="PRIMARY KEY column "color" cannot be restricted as preceding column "name" is not restricted"
cqlsh:student3_16>
```

Как я понял, можно запросить столбцы, которые являются частью первичного ключа, или столбцы, которые имеют вторичный индекс. Т.к. Cassandra является колоночной БД и для выполнения сложной фильтрации Cassandra требует дополнительной фильтрации данных и, возможно, не сможет эффективно выполнить запрос без «ALLOW FILTERING».

По результатам тестирования:

- запрос select * from animals where id = 4 and name = 'Cow' and color = 'Black'; - работает;

- запрос select * from animals where id = 4 and name = 'Cow'; - работает;

- запрос select * from animals where id = 4; - работает;

- запрос select * from animals where id = 4 and color = 'Black'; - не работает;

- запрос select * from animals where color = 'Black'; - не работает.

Можно сделать вывод о том, что важно соблюдать последовательно фильтруемых значений.

5. Сравниваем удаление и вставку пустого значения:

delete id from animals where id = 1;

insert into animals (id, name, size)

values (3, null, null);

```
cqlsh:student3_16> delete id from animals where id = 1;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Invalid identifier id for deletion (should not be a PRIMARY KEY part)"
cqlsh:student3_16> insert into animals (id, name, size)
                ... values (3, null, null);
cqlsh:student3_16> select * from animals;

 id | name | size
----+------+------
  3 | null | null

(1 rows)
cqlsh:student3_16>
```

При попытке удалить id=1 ругается на неправильный идентификатор удаления, так и есть его не существует

При вставке нового значения под id=3, предыдущий который был создан ранее перезаписался

HBase

1. Подключаемся к HBase

hbase shell

create_namespace 'student3_16'

create 'student3_16:animals', 'name', 'size'

```
[student3_16@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/3.1.4.0-315/phoenix/phoenix-5.0.0.3.1.4.0-315-server.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/3.1.4.0-315/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
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 2.0.2.3.1.4.0-315, r, Fri Aug 23 05:15:48 UTC 2019
Took 0.0022 seconds
hbase(main):001:0> create_namespace 'student3_16'
Took 1.2657 seconds
hbase(main):002:0> create 'student3_16:animals', 'name', 'size'
Created table student3_16:animals
Took 88.3212 seconds
=> Hbase::Table - student3_16:animals
```

2. Вставляем значения:

put 'student3_16:animals', '3', 'name', 'Deer'

put 'student3_16:animals', '3', 'size', 'Big'

put 'student3_16:animals', '5', 'name', 'Snake'

put 'student3_16:animals', '3', 'name', 'Doe'

```
hbase(main):006:0> put 'student3_16:animals', '3', 'name', 'Deer'
Took 0.0159 seconds
hbase(main):007:0> put 'student3_16:animals', '3', 'size', 'Big'
Took 0.0071 seconds
hbase(main):008:0> put 'student3_16:animals', '5', 'name', 'Snake'
Took 0.0087 seconds
hbase(main):009:0> put 'student3_16:animals', '3', 'name', 'Doe'
Took 0.0064 seconds
```

3. Удаляем значение:

delete 'student3_16:animals', '5'

```
hbase(main):010:0> delete 'student3_16:animals', '5'
ERROR: wrong number of arguments (2 for 3)

Put a delete cell value at specified table/row/column and optionally
timestamp coordinates.  Deletes must match the deleted cell's
coordinates exactly.  When scanning, a delete cell suppresses older
versions. To delete a cell from  't1' at row 'r1' under column 'c1'
marked with the time 'ts1', do:

  hbase> delete 'ns1:t1', 'r1', 'c1', ts1
  hbase> delete 't1', 'r1', 'c1', ts1
  hbase> delete 't1', 'r1', 'c1', ts1, {VISIBILITY=>'PRIVATE|SECRET'}

The same command can also be run on a table reference. Suppose you had a reference
t to table 't1', the corresponding command would be:

  hbase> t.delete 'r1', 'c1',  ts1
  hbase> t.delete 'r1', 'c1',  ts1, {VISIBILITY=>'PRIVATE|SECRET'}

Took 0.0163 seconds
```

```
hbase(main):018:0> delete 'student3_16:animals', '5', 'name:Snake'
Took 0.0356 seconds
```

4. Делаем запрос к созданной таблице:

get 'student3_16:animals', '5'

```
hbase(main):011:0> get 'student3_16:animals', '5'
COLUMN                                          CELL
 name:                                          timestamp=1619626038406, value=Snake
1 row(s)
Took 0.0557 seconds
hbase(main):012:0> get 'student3_16:animals', '2'
COLUMN                                          CELL
0 row(s)
Took 0.0059 seconds
hbase(main):013:0> get 'student3_16:animals', '0'
COLUMN                                          CELL
0 row(s)
Took 0.0059 seconds
hbase(main):014:0> get 'student3_16:animals', '1'
COLUMN                                          CELL
0 row(s)
Took 0.0053 seconds
hbase(main):015:0> get 'student3_16:animals', '3'
COLUMN                                          CELL
 name:                                          timestamp=1619626045060, value=Doe
 size:                                          timestamp=1619626033420, value=Big
1 row(s)
Took 0.0131 seconds
hbase(main):016:0> get 'student3_16:animals', '5'
COLUMN                                          CELL
 name:                                          timestamp=1619626038406, value=Snake
1 row(s)
Took 0.0060 seconds
```