EJERCICIOS DE HBase

NOMBRE ALUMNO: Tania Batista

Partiendo de los datos de la discografía de álbumes de estudio de Pink Floyd
http://es.wikipedia.org/wiki/Anexo:Discograf%C3%ADa_de_Pink_Floyd
Resolver las siguientes preguntas:

1. Crear un namespace llamado discografía (1 punto)

```
^Cbigdata@bigdata:~$ start-hbase.sh
running master, logging to /home/bigdata/hbase/logs/hbase-bigdata-master-bigdata.out
```

```
bigdata@bigdata:~$ jps
2880 HMaster
2955 Jps
```

```
bigdata@bigdata:~$ hbase shell

HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 2.0.0-alpha4, r5c4b985f89c99cc8b0f8515a4097c811a0848835, Tue Oct 31 16:00:33
Took 0.0090 seconds
hbase(main):001:0>
```

```
hbase(main):012:0> create 'discografia:discos', 'band', 'year', 'disc_name', 'ranking_
Created table discografia:discos
Took 0.7590 seconds
```

2. Crear una tabla llamada discos que contenga sólo las columnas:
grupo, nombre_disco, anio, ranking (a su vez tendrá identificadores para EEUU y UK) (1 punto)
mas facil para mi en ingles -- band, disc_name, year, ranking_us, ranking_uk

create_namespace 'discografia'

```
hbase(main):022:0> create_namespace 'discografia'
Took 0.2364 seconds
```

create 'discografia:discos', 'band', 'disc_name', 'year', 'ranking_us', 'ranking_uk'

```
hbase(main):011:0> create 'discografia:discos', 'band', 'disc_name', 'year', 'ranking_us', 'ranking_uk'

Created table discografia:discos
Took 0.7454 seconds
hbase(main):012:0>
```

scan 'discografia:discos'

```
hbase(main):013:0> scan 'discografia:discos'
ROW                                COLUMN+CELL
0 row(s)
Took 0.0934 seconds
hbase(main):014:0>
```

describe 'discografia:discos'

```
hbase(main):014:0> describe 'discografia:discos'
Table discografia:discos is ENABLED
discografia:discos
COLUMN FAMILIES DESCRIPTION
{NAME => 'band', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', CA
ATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER =
W', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false', COMPR
N => 'NONE', CACHE_DATA_IN_L1 => 'false', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}
{NAME => 'disc_name', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE
CHE_DATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFIL
> 'ROW', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false',
ESSION => 'NONE', CACHE_DATA_IN_L1 => 'false', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}
{NAME => 'ranking_uk', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALS
ACHE_DATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFI
=> 'ROW', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false',
RESSION => 'NONE', CACHE_DATA_IN_L1 => 'false', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}
{NAME => 'ranking_us', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALS
ACHE_DATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFI
=> 'ROW', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false',
RESSION => 'NONE', CACHE_DATA_IN_L1 => 'false', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}
{NAME => 'year', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', CA
ATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER =
W', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false', COMPR
N => 'NONE', CACHE_DATA_IN_L1 => 'false', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}
5 row(s)
Took 0.1011 seconds
hbase(main):015:0>
```

3. Cargar los datos en HBase (se pueden emplear los comandos put o una importación desde PIG).
Nota: algunos años hay más de un disco por lo que la clave será el año y un secuencial,
y no se disponen de todos los datos de rankings por lo que esos datos nulos no será necesario cargarlos (1 punto)

```
hbase(main):120:0> put 'discografia:discos', '1975#1', 'ranking_us', 1
Took 0.0024 seconds
hbase(main):121:0> put 'discografia:discos', '1975#1', 'ranking_uk', 1
Took 0.0073 seconds
hbase(main):122:0> put 'discografia:discos', '1967#1', 'band', 'Pink Floyd'
Took 0.0062 seconds
hbase(main):123:0> put 'discografia:discos', '1967#1', 'disc_name', 'The Piper at the Gates of Dawn'
Took 0.0029 seconds
hbase(main):124:0> put 'discografia:discos', '1967#1', 'year', 1967
Took 0.0017 seconds
```

```
Took 0.0017 seconds
hbase(main):125:0> put 'discografia:discos', '1967#1', 'ranking_us', 131
Took 0.0044 seconds
hbase(main):126:0> put 'discografia:discos', '1967#1', 'ranking_uk', 6
Took 0.0068 seconds
hbase(main):127:0>
hbase(main):128:0* put 'discografia:discos', '1968#1', 'band', 'Pink Floyd'
Took 0.0019 seconds
hbase(main):129:0> put 'discografia:discos', '1968#1', 'disc_name', 'A Saucerful of Secrets'
Took 0.0087 seconds
hbase(main):130:0> put 'discografia:discos', '1968#1', 'year', 1968
Took 0.0092 seconds
hbase(main):131:0> put 'discografia:discos', '1968#1', 'ranking_uk', 9
Took 0.0099 seconds
hbase(main):132:0>
hbase(main):133:0* put 'discografia:discos', '1969#1', 'band', 'Pink Floyd'
Took 0.0057 seconds
hbase(main):134:0> put 'discografia:discos', '1969#1', 'disc_name', 'Music from the Film More'
Took 0.0060 seconds
hbase(main):135:0> put 'discografia:discos', '1969#1', 'year', 1969
Took 0.0166 seconds
hbase(main):136:0> put 'discografia:discos', '1969#1', 'ranking_us', 153
Took 0.0053 seconds
hbase(main):137:0> put 'discografia:discos', '1969#1', 'ranking_uk', 9
Took 0.0201 seconds
hbase(main):138:0>
hbase(main):139:0*
hbase(main):140:0* put 'discografia:discos', '1969#2', 'band', 'Pink Floyd'
Took 0.0204 seconds
hbase(main):141:0> put 'discografia:discos', '1969#2', 'disc_name', 'Ummagumma'
Took 0.0045 seconds
hbase(main):142:0> put 'discografia:discos', '1969#2', 'year', 1969
Took 0.0309 seconds
hbase(main):143:0> put 'discografia:discos', '1969#2', 'ranking_us', 74
Took 0.0137 seconds
hbase(main):144:0> put 'discografia:discos', '1969#2', 'ranking_uk', 5
Took 0.0015 seconds
hbase(main):145:0>
hbase(main):146:0* put 'discografia:discos', '1970#1', 'band', 'Pink Floyd'
Took 0.0021 seconds
hbase(main):147:0> put 'discografia:discos', '1970#1', 'disc_name', 'Atom Heart Mother'
Took 0.0111 seconds
hbase(main):148:0> put 'discografia:discos', '1970#1', 'year', 1970
Took 0.0040 seconds
hbase(main):149:0> put 'discografia:discos', '1970#1', 'ranking_us', 55
```

```
hbase(main):150:0> put 'discografia:discos', '1970#1', 'ranking_uk', 1
Took 0.0103 seconds
hbase(main):151:0>
hbase(main):152:0* put 'discografia:discos', '1971#1', 'band', 'Pink Floyd'
Took 0.0234 seconds
hbase(main):153:0> put 'discografia:discos', '1971#1', 'disc_name', 'Meddle'
Took 0.0088 seconds
hbase(main):154:0> put 'discografia:discos', '1971#1', 'year', 1971
Took 0.0018 seconds
hbase(main):155:0> put 'discografia:discos', '1971#1', 'ranking_us', 70
Took 0.0015 seconds
hbase(main):156:0> put 'discografia:discos', '1971#1', 'ranking_uk', 3
Took 0.0073 seconds
hbase(main):157:0>
hbase(main):158:0* put 'discografia:discos', '1972#1', 'band', 'Pink Floyd'
Took 0.0048 seconds
hbase(main):159:0> put 'discografia:discos', '1972#1', 'disc_name', 'Obscured by Clouds'
Took 0.0024 seconds
hbase(main):160:0> put 'discografia:discos', '1972#1', 'year', 1972
Took 0.0093 seconds
hbase(main):161:0> put 'discografia:discos', '1972#1', 'ranking_us', 46
Took 0.0106 seconds
hbase(main):162:0> put 'discografia:discos', '1972#1', 'ranking_uk', 6
Took 0.0120 seconds
hbase(main):163:0>
hbase(main):164:0* put 'discografia:discos', '1973#1', 'band', 'Pink Floyd'
Took 0.0092 seconds
hbase(main):165:0> put 'discografia:discos', '1973#1', 'disc_name', 'The Dark Side of the Moon'
Took 0.0076 seconds
hbase(main):166:0> put 'discografia:discos', '1973#1', 'year', 1973
Took 0.0170 seconds
hbase(main):167:0> put 'discografia:discos', '1973#1', 'ranking_us', 1
Took 0.0041 seconds
hbase(main):168:0> put 'discografia:discos', '1973#1', 'ranking_uk', 1
Took 0.0118 seconds
hbase(main):169:0>
hbase(main):170:0* put 'discografia:discos', '1975#1', 'band', 'Pink Floyd'
Took 0.0061 seconds
hbase(main):171:0> put 'discografia:discos', '1975#1', 'disc_name', 'Wish You Were Here'
Took 0.0024 seconds
hbase(main):172:0> put 'discografia:discos', '1975#1', 'year', 1975
Took 0.0057 seconds
hbase(main):173:0> put 'discografia:discos', '1975#1', 'ranking_us', 1
Took 0.0094 seconds
hbase(main):174:0> put 'discografia:discos', '1975#1', 'ranking_uk', 1
```

4. Escribir la instrucción y el resultado de consultar el álbum del año 1968 empleando un filtrado (1 punto)

scan 'discografia:discos', {COLUMNS => 'year', FILTER => "ValueFilter(=, 'binary:1968')"}

```
hbase(main):176:0> scan 'discografia:discos', {COLUMNS => 'year', FILTER => "ValueFilter(=, 'binary:1968')"}
ROW                                COLUMN+CELL
 1968#1                            column=year:, timestamp=1521598363418, value=1968
1 row(s)
Took 0.1149 seconds
```

get 'discografia:discos', '1968#1'

```
hbase(main):178:0> get 'discografia:discos', '1968#1'
COLUMN                             CELL
 band:                            timestamp=1521598363336, value=Pink Floyd
 disc_name:                       timestamp=1521598363381, value=A Saucerful of Secrets
 ranking_uk:                      timestamp=1521598363455, value=9
 year:                            timestamp=1521598363418, value=1968
1 row(s)
Took 0.0847 seconds
hbase(main):179:0>
```

5. Añadir la información de los álbumes de estudio de GreenDay (2 puntos)
https://en.wikipedia.org/wiki/Green_Day_discography

```
hbase(main):179:0> put 'discografia:discos', '1990#1', 'band', 'Green Day'
Took 0.0203 seconds
hbase(main):180:0> put 'discografia:discos', '1990#1', 'disc_name', '39/Smooth'
Took 0.0075 seconds
hbase(main):181:0> put 'discografia:discos', '1990#1', 'year', '1990'
Took 0.0157 seconds
hbase(main):182:0> put 'discografia:discos', '1990#1', 'ranking_us', ' '
Took 0.0114 seconds
hbase(main):183:0> put 'discografia:discos', '1990#1', 'ranking_uk', ' '
Took 0.0110 seconds
hbase(main):184:0> put 'discografia:discos', '1991#1', 'band', 'Green Day'
Took 0.0057 seconds
hbase(main):185:0> put 'discografia:discos', '1991#1', 'disc_name', 'Kerplunk'
Took 0.0131 seconds
hbase(main):186:0> put 'discografia:discos', '1991#1', 'year', '1991'
Took 0.0119 seconds
hbase(main):187:0> put 'discografia:discos', '1991#1', 'ranking_us', ' '
Took 0.0143 seconds
hbase(main):188:0> put 'discografia:discos', '1991#1', 'ranking_uk', ' '
Took 0.0133 seconds
hbase(main):189:0> put 'discografia:discos', '1994#2', 'band', 'Green Day'
Took 0.0149 seconds
hbase(main):190:0> put 'discografia:discos', '1994#2', 'disc_name', 'Dookie'
Took 0.0239 seconds
hbase(main):191:0> put 'discografia:discos', '1994#2', 'year', '1994'
Took 0.0101 seconds
hbase(main):192:0> put 'discografia:discos', '1994#2', 'ranking_us', '2'
Took 0.0097 seconds
hbase(main):193:0> put 'discografia:discos', '1994#2', 'ranking_uk', '13'
Took 0.0107 seconds
hbase(main):194:0> put 'discografia:discos', '1995#1', 'band', 'Green Day'
Took 0.0069 seconds
hbase(main):195:0> put 'discografia:discos', '1995#1', 'disc_name', 'Insomniac'
Took 0.0120 seconds
hbase(main):196:0> put 'discografia:discos', '1995#1', 'year', '1995'
Took 0.0076 seconds
hbase(main):197:0> put 'discografia:discos', '1995#1', 'ranking_us', '2'
Took 0.0092 seconds
hbase(main):198:0> put 'discografia:discos', '1995#1', 'ranking_uk', '8'
Took 0.0050 seconds
hbase(main):199:0> put 'discografia:discos', '1997#1', 'band', 'Green Day'
Took 0.0017 seconds
hbase(main):200:0> put 'discografia:discos', '1997#1', 'disc_name', 'Nimrod'
Took 0.0124 seconds
hbase(main):201:0> put 'discografia:discos', '1997#1', 'year', '1997'
Took 0.0057 seconds
hbase(main):202:0> put 'discografia:discos', '1997#1', 'ranking_us', '10'
```

6. Ejecuta varias instrucciones (describiéndolas) para analizar datos anteriores (4 puntos)

- all the CDs that came out AFTER the 2000s, count them

scan 'discografia:discos', {COLUMNS => 'year', FILTER => "ValueFilter(>, 'binary:2000')"}

count 'discografia:discos', {COLUMNS => 'year', FILTER => "ValueFilter(>, 'binary:2000')"}

- all the #1 ranking UK

scan 'discografia:discos', {COLUMNS => 'ranking_uk', FILTER => "ValueFilter(=, 'binary:1')"}

- all the highest ranking US

scan 'discografia:discos', {COLUMNS => 'ranking_us', FILTER => "ValueFilter(=, 'binary:1')"}

- scan and count how many did NOT rank in the US

scan 'discografia:discos', {COLUMNS => 'ranking_us', FILTER => "ValueFilter(=, 'binary: ')"}
count 'discografia:discos', {COLUMNS => 'ranking_us', FILTER => "ValueFilter(=, 'binary: ')"}

- scan and count how many DID rank in the US

scan 'discografia:discos', {COLUMNS => 'ranking_us', FILTER => "ValueFilter(!=, 'binary: ')"}
count 'discografia:discos', {COLUMNS => 'ranking_us', FILTER => "ValueFilter(!=, 'binary: ')"}

- count all Pink Floyd discs

scan 'discografia:discos', {COLUMNS => 'band', FILTER => "ValueFilter(=, 'binaryprefix:Pink Floyd')"}
count 'discografia:discos', {COLUMNS => 'band', FILTER => "ValueFilter(=, 'binaryprefix:Pink Floyd')"}

- count all Green Day discs

scan 'discografia:discos', {COLUMNS => 'band', FILTER => "ValueFilter(=, 'binaryprefix:Green Day')"}
count 'discografia:discos', {COLUMNS => 'band', FILTER => "ValueFilter(=, 'binaryprefix:Green Day')"}