

Ejercicio Hbase – Anna Lawrenc

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

Año Album	Ranking EEUU	Ranking UK
1967 The Piper at the Gates of Dawn	131	6
1968 A Saucerful of Secrets	—	9
1969 Music from the Film More	153	9
1969 Ummagumma (Disco 2)	74	5
1970 Atom Heart Mother	55	1
1971 Meddle	70	3
1972 Obscured by Clouds	46	6
1973 The Dark Side of the Moon	1	1
1975 Wish You Were Here	1	1
1977 Animals	3	2
1979 The Wall	1	3
1983 The Final Cut	6	1
1987 A Momentary Lapse of Reason	3	3
1994 The Division Bell	1	1
2014 The Endless River	3	1

Resolver las siguientes preguntas:

1. Crear un namespace llamado discografía

hbase shell

```
create_namespace 'Discografía'
```

```
hbase(main):002:0> create_namespace 'Discografía'
0 row(s) in 0.0550 seconds

hbase(main):003:0> describe_namespace 'Discografía'
DESCRIPTION
{NAME => 'Discografía'}
1 row(s) in 0.0070 seconds

hbase(main):004:0> █
```

```
hbase(main):004:0> list_namespace
NAMESPACE
Discografía
default
ejercicios
hbase
masterHadoop
5 row(s) in 0.0740 seconds

hbase(main):005:0>
```

2. Crear una tabla llamada discos que contenga sólo las columnas: grupo, nombre_disco, año, ranking (a su vez tendrá identificadores para EEUU y UK)

create 'Discografia:Discos' ,'Disco','Grupo','Anio','Ranking'

```
hbase(main):005:0> create 'Discografia:Discos' ,'Disco','Grupo','Anio','Ranking'
0 row(s) in 1.2900 seconds

=> Hbase::Table - Discografia:Discos
hbase(main):006:0> describe 'Discografia:Discos'
Table Discografia:Discos is ENABLED
Discografia:Discos
COLUMN FAMILIES DESCRIPTION
{NAME => 'Anio', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONING => 'NONE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
{NAME => 'Disco', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONING => 'NONE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
{NAME => 'Grupo', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONING => 'NONE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
{NAME => 'Ranking', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONING => 'NONE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
4 row(s) in 0.1750 seconds

hbase(main):007:0> █
```

3. Cargar los datos en HBase (se pueden emplear los comandos put o u 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

```
put 'Discografia:Discos','r1','Grupo','Pink Floyd'
put 'Discografia:Discos','r2','Grupo','Pink Floyd'
put 'Discografia:Discos','r3','Grupo','Pink Floyd'
put 'Discografia:Discos','r4','Grupo','Pink Floyd'
put 'Discografia:Discos','r5','Grupo','Pink Floyd'
put 'Discografia:Discos','r6','Grupo','Pink Floyd'
put 'Discografia:Discos','r7','Grupo','Pink Floyd'
put 'Discografia:Discos','r8','Grupo','Pink Floyd'
put 'Discografia:Discos','r9','Grupo','Pink Floyd'
put 'Discografia:Discos','r10','Grupo','Pink Floyd'
put 'Discografia:Discos','r11','Grupo','Pink Floyd'
put 'Discografia:Discos','r12','Grupo','Pink Floyd'
put 'Discografia:Discos','r13','Grupo','Pink Floyd'
put 'Discografia:Discos','r14','Grupo','Pink Floyd'
put 'Discografia:Discos','r15','Grupo','Pink Floyd'
```

```
put 'Discografia:Discos','r1','Disco','The Piper at the Gates of Dawn'
put 'Discografia:Discos','r2','Disco','A Saucerful of Secrets'
put 'Discografia:Discos','r3','Disco','Music from the Film More'
put 'Discografia:Discos','r4','Disco','Ummagumma (Disco 2)'
put 'Discografia:Discos','r5','Disco','Atom Heart Mother'
put 'Discografia:Discos','r6','Disco','Meddle'
put 'Discografia:Discos','r7','Disco','Obscured by Clouds'
put 'Discografia:Discos','r8','Disco','The Dark Side of the Moon'
put 'Discografia:Discos','r9','Disco','Wish You Were Here'
put 'Discografia:Discos','r10','Disco','Animals'
put 'Discografia:Discos','r11','Disco','The Wall'
put 'Discografia:Discos','r12','Disco','The Final Cut'
put 'Discografia:Discos','r13','Disco','A Momentary Lapse of Reason'
put 'Discografia:Discos','r14','Disco','The Division Bell'
put 'Discografia:Discos','r15','Disco','The Endless River'
```

```
put 'Discografia:Discos','r1','Ranking:EEUU',31
put 'Discografia:Discos','r2','Ranking:EEUU',-
put 'Discografia:Discos','r3','Ranking:EEUU',153
put 'Discografia:Discos','r4','Ranking:EEUU',74
put 'Discografia:Discos','r5','Ranking:EEUU',55
put 'Discografia:Discos','r6','Ranking:EEUU',70
put 'Discografia:Discos','r7','Ranking:EEUU',46
put 'Discografia:Discos','r8','Ranking:EEUU',1
put 'Discografia:Discos','r9','Ranking:EEUU',1
put 'Discografia:Discos','r10','Ranking:EEUU',3
put 'Discografia:Discos','r11','Ranking:EEUU',1
put 'Discografia:Discos','r12','Ranking:EEUU',6
put 'Discografia:Discos','r13','Ranking:EEUU',3
put 'Discografia:Discos','r14','Ranking:EEUU',1
put 'Discografia:Discos','r15','Ranking:EEUU',3
```

```
put 'Discografia:Discos','r1','Ranking:UK',6
put 'Discografia:Discos','r2','Ranking:UK',9
put 'Discografia:Discos','r3','Ranking:UK',9
put 'Discografia:Discos','r4','Ranking:UK',5
put 'Discografia:Discos','r5','Ranking:UK',1
put 'Discografia:Discos','r6','Ranking:UK',3
put 'Discografia:Discos','r7','Ranking:UK',6
put 'Discografia:Discos','r8','Ranking:UK',1
put 'Discografia:Discos','r9','Ranking:UK',1
put 'Discografia:Discos','r10','Ranking:UK',2
put 'Discografia:Discos','r11','Ranking:UK',3
put 'Discografia:Discos','r12','Ranking:UK',1
put 'Discografia:Discos','r13','Ranking:UK',3
put 'Discografia:Discos','r14','Ranking:UK',1
put 'Discografia:Discos','r15','Ranking:UK',1
```

s,can 'Discografia:Discos'

```
r1 column=Anio:, timestamp=1475527676041, value=1967
r1 column=Disco:, timestamp=1475527211279, value=The Piper at the Gates of Dawn
r1 column=Grupo:, timestamp=1475527415932, value=Pink Floyd
r1 column=Ranking:EEUU, timestamp=1475527963929, value=31
r1 column=Ranking:UK, timestamp=1475528125677, value=6
r10 column=Anio:, timestamp=1475527787969, value=1977
r10 column=Disco:, timestamp=1475527211643, value=Animals
r10 column=Grupo:, timestamp=1475527416418, value=Pink Floyd
r10 column=Ranking:EEUU, timestamp=1475527983768, value=3
r10 column=Ranking:UK, timestamp=1475528125943, value=2
r11 column=Anio:, timestamp=1475527788006, value=1979
r11 column=Disco:, timestamp=1475527211722, value=The Wall
r11 column=Grupo:, timestamp=1475527416497, value=Pink Floyd
r11 column=Ranking:EEUU, timestamp=1475527983793, value=1
r11 column=Ranking:UK, timestamp=1475528125963, value=3
r12 column=Anio:, timestamp=1475527788044, value=1983
r12 column=Disco:, timestamp=1475527211771, value=The Final Cut
r12 column=Grupo:, timestamp=1475527416643, value=Pink Floyd
r12 column=Ranking:EEUU, timestamp=1475527983825, value=6
r12 column=Ranking:UK, timestamp=1475528125988, value=1
r13 column=Anio:, timestamp=1475527788083, value=1987
r13 column=Disco:, timestamp=1475527211805, value=A Momentary Lapse of Reason
r13 column=Grupo:, timestamp=1475527416716, value=Pink Floyd
r13 column=Ranking:EEUU, timestamp=1475527983851, value=3
r13 column=Ranking:UK, timestamp=1475528126027, value=3
r14 column=Anio:, timestamp=1475527788115, value=1994
r14 column=Disco:, timestamp=1475527211832, value=The Division Bell
r14 column=Grupo:, timestamp=1475527416779, value=Pink Floyd
r14 column=Ranking:EEUU, timestamp=1475527983887, value=1
r14 column=Ranking:UK, timestamp=1475528126071, value=1
r15 column=Anio:, timestamp=1475527788145, value=2014
r15 column=Disco:, timestamp=1475527211868, value=The Endless River
r15 column=Grupo:, timestamp=1475527416809, value=Pink Floyd
r15 column=Ranking:EEUU, timestamp=1475527983922, value=3
r15 column=Ranking:UK, timestamp=1475528126096, value=1
```

4. Escribir la instrucción y el resultado de consultar el álbum del año 1968

Busco el album del año 1968:

```
scan 'Discografia:Discos', { COLUMNS => 'Anio', FILTER => "ValueFilter( =, 'binary:1968' )" }
```

```
hbase(main):003:0> scan 'Discografia:Discos', { COLUMNS => 'Anio', FILTER => "ValueFilter( =, 'binary:1968' )" }
ROW                COLUMN+CELL
 r2                  column=Anio:, timestamp=1475527787643, value=1968
1 row(s) in 0.1030 seconds
hbase(main):004:0> █
```

Es el album con clave r2

Busco las características del album con clava r2:

```
get 'Discografia:Discos','r2'
```

```
hbase(main):002:0> get 'Discografia:Discos','r2'
COLUMN              CELL
 Anio:                timestamp=1475527787643, value=1968
 Disco:               timestamp=1475527211330, value=A Saucerful of Secrets
 Grupo:               timestamp=1475527415972, value=Pink Floyd
 Ranking:UK           timestamp=1475528125715, value=9
4 row(s) in 0.0310 seconds
hbase(main):003:0> █
```

5. Añadir la información de los álbumes de estudio de GreenDay wget

https://en.wikipedia.org/wiki/Green_Day_discography

```
put 'Discografia:Discos','r16','Grupo','Green Day'
put 'Discografia:Discos','r17','Grupo','Green Day'
put 'Discografia:Discos','r18','Grupo','Green Day'
put 'Discografia:Discos','r19','Grupo','Green Day'
put 'Discografia:Discos','r20','Grupo','Green Day'
put 'Discografia:Discos','r21','Grupo','Green Day'
put 'Discografia:Discos','r22','Grupo','Green Day'
put 'Discografia:Discos','r23','Grupo','Green Day'
put 'Discografia:Discos','r24','Grupo','Green Day'
put 'Discografia:Discos','r25','Grupo','Green Day'
put 'Discografia:Discos','r26','Grupo','Green Day'
put 'Discografia:Discos','r27','Grupo','Green Day'
put 'Discografia:Discos','r28','Grupo','Green Day'
put 'Discografia:Discos','r29','Grupo','Green Day'
put 'Discografia:Discos','r30','Grupo','Green Day'

put 'Discografia:Discos','r16','Disco','39/Smooth'
put 'Discografia:Discos','r17','Disco','Kerplunk'
put 'Discografia:Discos','r18','Disco','Dookie'
put 'Discografia:Discos','r19','Disco','Insomniac'
put 'Discografia:Discos','r20','Disco','Nimrod'
put 'Discografia:Discos','r21','Disco','Warning'
put 'Discografia:Discos','r22','Disco','American Idiot'
put 'Discografia:Discos','r23','Disco','21st Century Breakdown'
put 'Discografia:Discos','r24','Disco','¡Uno!'
put 'Discografia:Discos','r25','Disco','¡Dos!'
put 'Discografia:Discos','r26','Disco','¡Tré!'
put 'Discografia:Discos','r27','Disco','Revolution Radio'
put 'Discografia:Discos','r28','Disco','Bullet in a Bible'
put 'Discografia:Discos','r29','Disco','Awesome as Fuck'
put 'Discografia:Discos','r30','Disco','On the Radio'

put 'Discografia:Discos','r16','Ranking:EEUU',-
put 'Discografia:Discos','r17','Ranking:EEUU',-
put 'Discografia:Discos','r18','Ranking:EEUU',2
put 'Discografia:Discos','r19','Ranking:EEUU',2
put 'Discografia:Discos','r20','Ranking:EEUU',10
put 'Discografia:Discos','r21','Ranking:EEUU',4
put 'Discografia:Discos','r22','Ranking:EEUU',1
put 'Discografia:Discos','r23','Ranking:EEUU',1
put 'Discografia:Discos','r24','Ranking:EEUU',2
put 'Discografia:Discos','r25','Ranking:EEUU',9
put 'Discografia:Discos','r26','Ranking:EEUU',13
put 'Discografia:Discos','r27','Ranking:EEUU',-
put 'Discografia:Discos','r28','Ranking:EEUU',8
put 'Discografia:Discos','r29','Ranking:EEUU',14
put 'Discografia:Discos','r30','Ranking:EEUU',-

put 'Discografia:Discos','r16','Ranking:UK',-
put 'Discografia:Discos','r17','Ranking:UK',-
put 'Discografia:Discos','r18','Ranking:UK',13
put 'Discografia:Discos','r19','Ranking:UK',8
put 'Discografia:Discos','r20','Ranking:UK',11
put 'Discografia:Discos','r21','Ranking:UK',4
put 'Discografia:Discos','r22','Ranking:UK',1
put 'Discografia:Discos','r23','Ranking:UK',1
put 'Discografia:Discos','r24','Ranking:UK',2
put 'Discografia:Discos','r25','Ranking:UK',10
put 'Discografia:Discos','r26','Ranking:UK',31
put 'Discografia:Discos','r27','Ranking:UK',-
put 'Discografia:Discos','r28','Ranking:UK',6
put 'Discografia:Discos','r29','Ranking:UK',14
put 'Discografia:Discos','r30','Ranking:UK',-
```

scan 'Discografia:Discos'

```
r16                  column=Anio:, timestamp=1475527788199, value=1990
r16                  column=Disco:, timestamp=1475527865228, value=39/Smooth
r16                  column=Grupo:, timestamp=1475527846722, value=Green Day
r17                  column=Anio:, timestamp=1475527788247, value=1992
r17                  column=Disco:, timestamp=1475527865271, value=Kerplunk
r17                  column=Grupo:, timestamp=1475527846745, value=Green Day
r18                  column=Anio:, timestamp=1475527788303, value=1994
r18                  column=Disco:, timestamp=1475527865291, value=Dookie
r18                  column=Grupo:, timestamp=1475527846775, value=Green Day
r18                  column=Ranking:EEUU, timestamp=1475528002983, value=2
r18                  column=Ranking:UK, timestamp=1475528140036, value=13
r19                  column=Anio:, timestamp=1475527788335, value=1995
r19                  column=Disco:, timestamp=1475527865317, value=Insomniac
r19                  column=Grupo:, timestamp=1475527846812, value=Green Day
r19                  column=Ranking:EEUU, timestamp=1475528003011, value=2
r19                  column=Ranking:UK, timestamp=1475528140058, value=8
```

...

```

r20 column=Anio:, timestamp=1475527788375, value=1997
r20 column=Disco:, timestamp=1475527865353, value=Nimrod
r20 column=Grupo:, timestamp=1475527846834, value=Green Day
r20 column=Ranking:EEUU, timestamp=1475528003040, value=10
r20 column=Ranking:UK, timestamp=1475528140095, value=11
r21 column=Anio:, timestamp=1475527788398, value=2000
r21 column=Disco:, timestamp=1475527865376, value=Warning
r21 column=Grupo:, timestamp=1475527846863, value=Green Day
r21 column=Ranking:EEUU, timestamp=1475528003062, value=4
r21 column=Ranking:UK, timestamp=1475528140118, value=4
r22 column=Anio:, timestamp=1475527788440, value=2004
r22 column=Disco:, timestamp=1475527865415, value=American Idiot
r22 column=Grupo:, timestamp=1475527846885, value=Green Day
r22 column=Ranking:EEUU, timestamp=1475528003096, value=1
r22 column=Ranking:UK, timestamp=1475528140142, value=1
r23 column=Anio:, timestamp=1475527788465, value=2009
r23 column=Disco:, timestamp=1475527865439, value=21st Century Breakdown
r23 column=Grupo:, timestamp=1475527846910, value=Green Day
r23 column=Ranking:EEUU, timestamp=1475528003123, value=1
r23 column=Ranking:UK, timestamp=1475528140164, value=1
r24 column=Anio:, timestamp=1475527788494, value=2012
r24 column=Disco:, timestamp=1475527865459, value=\xC2\xA1Uno!
r24 column=Grupo:, timestamp=1475527846938, value=Green Day
r24 column=Ranking:EEUU, timestamp=1475528003151, value=2
r24 column=Ranking:UK, timestamp=1475528140195, value=2
r25 column=Anio:, timestamp=1475527788529, value=2012
r25 column=Disco:, timestamp=1475527865496, value=\xC2\xA1Dos!
r25 column=Grupo:, timestamp=1475527846969, value=Green Day
r25 column=Ranking:EEUU, timestamp=1475528003183, value=9

```

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

Busco los álbumes con primera plaza en el ranking:

```
scan 'Discografia:Discos', { COLUMNS => 'Ranking', FILTER => "ValueFilter( =, 'binary:1' )" }
```

```

hbase(main):010:0> scan 'Discografia:Discos', { COLUMNS => 'Ranking', FILTER => "ValueFilter( =, 'binary:1' )" }
ROW COLUMN+CELL
r11 column=Ranking:EEUU, timestamp=1475527983793, value=1
r12 column=Ranking:UK, timestamp=1475528125988, value=1
r14 column=Ranking:EEUU, timestamp=1475527983887, value=1
r14 column=Ranking:UK, timestamp=1475528126071, value=1
r15 column=Ranking:UK, timestamp=1475528126096, value=1
r22 column=Ranking:EEUU, timestamp=1475528003096, value=1
r22 column=Ranking:UK, timestamp=1475528140142, value=1
r23 column=Ranking:EEUU, timestamp=1475528003123, value=1
r23 column=Ranking:UK, timestamp=1475528140164, value=1
r5 column=Ranking:UK, timestamp=1475528125800, value=1
r8 column=Ranking:EEUU, timestamp=1475527983726, value=1
r8 column=Ranking:UK, timestamp=1475528125887, value=1
r9 column=Ranking:EEUU, timestamp=1475527983750, value=1
r9 column=Ranking:UK, timestamp=1475528125909, value=1
9 row(s) in 0.0380 seconds

hbase(main):011:0> █

```

Busco los álbumes con primera plaza en el ranking en UK:

```
hbase(main):011:0> scan 'Discografia:Discos', { COLUMNS => 'Ranking:UK', FILTER => "ValueFilter( =, 'binary:1' )" }
```

```

hbase(main):011:0> scan 'Discografia:Discos', { COLUMNS => 'Ranking:UK', FILTER => "ValueFilter( =, 'binary:1' )" }
ROW COLUMN+CELL
r12 column=Ranking:UK, timestamp=1475528125988, value=1
r14 column=Ranking:UK, timestamp=1475528126071, value=1
r15 column=Ranking:UK, timestamp=1475528126096, value=1
r22 column=Ranking:UK, timestamp=1475528140142, value=1
r23 column=Ranking:UK, timestamp=1475528140164, value=1
r5 column=Ranking:UK, timestamp=1475528125800, value=1
r8 column=Ranking:UK, timestamp=1475528125887, value=1
r9 column=Ranking:UK, timestamp=1475528125909, value=1
8 row(s) in 0.0420 seconds

hbase(main):012:0> █

```

Busco los álbumes de Pink Floyd:

```
scan 'Discografia:Discos', { COLUMNS => 'Grupo', FILTER => "ValueFilter( =, 'binary:Pink Floyd' )" }
```

```
hbase(main):012:0> scan 'Discografia:Discos', { COLUMNS => 'Grupo', FILTER => "ValueFilter( =, 'binary:Pink Floyd' )" }
ROW
COLUMN+CELL
The Piper at the Gates of Dawn      column=Grupo:, timestamp=1475524489846, value=Pink Floyd
r1                                  column=Grupo:, timestamp=1475527415932, value=Pink Floyd
r10                                  column=Grupo:, timestamp=1475527416418, value=Pink Floyd
r11                                  column=Grupo:, timestamp=1475527416497, value=Pink Floyd
r12                                  column=Grupo:, timestamp=1475527416643, value=Pink Floyd
r13                                  column=Grupo:, timestamp=1475527416716, value=Pink Floyd
r14                                  column=Grupo:, timestamp=1475527416779, value=Pink Floyd
r15                                  column=Grupo:, timestamp=1475527416809, value=Pink Floyd
r2                                   column=Grupo:, timestamp=1475527415972, value=Pink Floyd
r3                                   column=Grupo:, timestamp=1475527416003, value=Pink Floyd
r4                                   column=Grupo:, timestamp=1475527416042, value=Pink Floyd
r5                                   column=Grupo:, timestamp=1475527416082, value=Pink Floyd
r6                                   column=Grupo:, timestamp=1475527416154, value=Pink Floyd
r7                                   column=Grupo:, timestamp=1475527416238, value=Pink Floyd
r8                                   column=Grupo:, timestamp=1475527416292, value=Pink Floyd
r9                                   column=Grupo:, timestamp=1475527416359, value=Pink Floyd
16 row(s) in 0.0490 seconds
hbase(main):013:0> █
```

Busco los álbumes anteriores al año 1980:

```
scan 'Discografia:Discos', { COLUMNS => 'Anio', FILTER => "ValueFilter( <, 'binary:1980' )" }
```

```
hbase(main):013:0> scan 'Discografia:Discos', { COLUMNS => 'Anio', FILTER => "ValueFilter( <, 'binary:1980' )" }
ROW
COLUMN+CELL
r1                                  column=Anio:, timestamp=1475527676041, value=1967
r10                                 column=Anio:, timestamp=1475527787969, value=1977
r11                                 column=Anio:, timestamp=1475527788006, value=1979
r2                                  column=Anio:, timestamp=1475527787643, value=1968
r3                                  column=Anio:, timestamp=1475527787705, value=1969
r4                                  column=Anio:, timestamp=1475527787745, value=1969
r5                                  column=Anio:, timestamp=1475527787773, value=1970
r6                                  column=Anio:, timestamp=1475527787808, value=1971
r7                                  column=Anio:, timestamp=1475527787866, value=1972
r8                                  column=Anio:, timestamp=1475527787903, value=1973
r9                                  column=Anio:, timestamp=1475527787934, value=1975
11 row(s) in 0.0400 seconds
hbase(main):014:0> █
```