# BBDD NOSQL
# Uso de Cassandra

**José Manuel Bustos Muñoz**

# Ejercicios

1. (**0,5 puntos**) Crear el **keyspace wc2014**.

Con la instrucción "*Describe keyspaces;*" se listan los keyspaces definidos.

```
cqlsh> Describe keyspaces;

system_traces  system_schema  system_auth  system  system_distributed
```

Creamos el keyspace wc2014 con la sentencia: "***Create keyspace wc2014 with replication = {'class':'SimpleStrategy', 'replication_factor':1};***".

Al volver a poner "*Describe keyspaces;*" ya saldrá en la lista el nuevo keyspace creado.

```
cqlsh> Create Keyspace wc2014 with replication = {'class':'SimpleStrategy', 'rep
lication_factor':1};
cqlsh> Describe keyspaces;

system_schema  system_auth  system  wc2014  system_distributed  system_traces
```

2. (**2,5 puntos**) Crear la tabla **players**, justificando el tipo de dato de cada campo que se cree.

Primero miramos el fichero con los datos que se quieren exportar a la tabla players, para ver las columnas que habrá que crear en la tabla: grupo, equipo, numero, posicion, nombre, dia, mes, ano, club, liga y capitan.

| Grupo | Equipo | Numero | Posicion | Nombre | Dia | Mes | Ano | Club | Liga | Capitan |
|---|---|---|---|---|---|---|---|---|---|---|
| A | Brasil | 1 | PORTERO | Jefferson | 2 | 1 | 1983 | Botafogo | Brasil | NO |
| A | Brasil | 2 | DEFENSA | Dani Alves | 6 | 5 | 1983 | FC Barcelona | Espana | NO |
| A | Brasil | 3 | DEFENSA | Thiago Silva (c) | 22 | 9 | 1984 | Paris Saint-Germain | Francia | SI |
| A | Brasil | 4 | DEFENSA | David Luiz | 22 | 4 | 1987 | Chelsea | Inglaterra | NO |
| A | Brasil | 5 | MEDIOCENTRO | Fernandinho | 4 | 5 | 1985 | Manchester City | Inglaterra | NO |
| A | Brasil | 6 | DEFENSA | Marcelo | 12 | 5 | 1988 | Real Madrid | Espana | NO |
| A | Brasil | 7 | DELANTERO | Hulk | 25 | 7 | 1986 | Zenit Saint Petersburg | Rusia | NO |
| A | Brasil | 8 | MEDIOCENTRO | Paulinho | 25 | 7 | 1988 | Tottenham Hotspur | Inglaterra | NO |
| A | Brasil | 9 | DELANTERO | Fred | 3 | 10 | 1983 | Fluminense | Brasil | NO |
| A | Brasil | 10 | DELANTERO | Neymar | 5 | 2 | 1992 | FC Barcelona | Espana | NO |
| A | Brasil | 11 | MEDIOCENTRO | Oscar | 9 | 9 | 1991 | Chelsea | Inglaterra | NO |
| A | Brasil | 12 | PORTERO | Julio Cesar | 3 | 9 | 1979 | Toronto | Canada | NO |
| A | Brasil | 13 | DEFENSA | Dante | 18 | 10 | 1983 | Bayern Munich | Alemania | NO |

Viendo el significado de cada columna y los datos almacenados en cada una en el fichero, se deciden asignar los siguientes tipos a cada columna:

- **grupo**: tipo varchar. Almacena un carácter alfanumérico que identifica al grupo en el que está encuadrado el equipo.
- **equipo**: tipo varchar. Almacena el equipo o selección con la que participa el jugador en el mundial.
- **numero**: tipo int. Almacena el dorsal de la camiseta del futbolista.
- **posicion**: tipo varchar. Almacena la posición en la que juega el jugador.
- **nombre**: tipo varchar. Almacena el nombre del jugador.
- **dia**: tipo int. Guarda el día de nacimiento del futbolista.
- **mes**: tipo int. Guarda el mes de nacimiento del futbolista.
- **year**: tipo int. Guarda el año de nacimiento del futbolista.
- **club**: tipo varchar. Almacena el equipo de fútbol en el que milita el jugador.

- **liga**: tipo varchar. Almacena la liga en la que participa el club al que pertenece el jugador, el país de dicha liga.
- **capitan**: tipo varchar. Almacena un valor entre dos disponibles “SI/NO” que hace referencia a sí el jugador en cuestión era capitán de su equipo o si no lo era respectivamente.

Se opta por asignar “*varchar*” como tipo a las columnas que almacenen texto o caracteres alfabéticos. No se ha encontrado diferencia alguna entre asignar a estas columnas con el tipo “*varchar*” o con el tipo “*text*”, parece según lo visto y leído que ambos tipos sirven para el cometido que aquí se desea.

El resto de columnas serán de tipo int, ya que almacenan números enteros, que en este caso no pasaran de un número de 4 cifras que tendrá máximo un valor aproximado de ‘2000’, o a lo sumo al ser de 4 cifras el año un valor de ‘9999’.

Antes de crear la tabla se utiliza la sentencia “**use wc2014;**” para trabajar sobre el keyspace correcto donde queremos crear la tabla.

Para crear la tabla se utiliza la sentencia: “**Create table players (grupo varchar, equipo varchar, numero int, posicion varchar, nombre varchar, dia int, mes int, year int, club varchar, liga varchar, capitan varchar, PRIMARY KEY (equipo, numero, nombre) );**”.

```
cqlsh:wc2014> Use wc2014;
cqlsh:wc2014> Create Table players (grupo varchar, equipo varchar, numero int, p
osicion varchar, nombre varchar, dia int, mes int, year int, club varchar, liga
varchar, capitan varchar, PRIMARY KEY (equipo, numero, nombre) );
cqlsh:wc2014>
```

Con la primera opción se crean como clave primaria las columnas de equipo, número y nombre, ya que se piensa que en base a estas tres columnas se podría identificar únicamente a cada jugador participante del mundial.

Otra opción para crear la tabla es asignar de clave primaria a las columnas por las que posteriormente se van a filtrar las consultas a realizar: si el jugador era capitán, la posición en la que juega el jugador, y el día, mes y año de nacimiento.
“***Create table players (grupo varchar, equipo varchar, numero int, posicion varchar, nombre varchar, dia int, mes int, year int, club varchar, liga varchar, capitan varchar, PRIMARY KEY (capitan, year, mes, dia, posicion) );***”

```
cqlsh:wc2014> create table players (grupo varchar, equipo varchar, numero int, p
osicion varchar, nombre varchar, dia int, mes int, year int, club varchar, liga
varchar, capitan varchar, PRIMARY KEY (capitan, year, mes, dia, posicion) );
cqlsh:wc2014> _
```

3. (**1,5 puntos**) Importar los datos del fichero **csv** a la tabla **players**. ¿Cuántos elementos has importado? ¿Has tenido algún problema? Si has tenido algún problema, ¿puedes indicar qué pasaba y cómo lo has solucionado?.

Primero se realiza un recuento de lineas del fichero a importar, y tiene 737 líneas contando la cabecera con el nombre de las columnas, por lo que tendríamos 736 filas que se corresponderían con futbolistas que participaron en el mundial.

```
vagrant@vagrant-ubuntu-trusty-64:~/apache-cassandra-3.11.0$ wc -l WorldCup2014.c
sv
737 WorldCup2014.csv
```

Para importar los datos del fichero en la tabla se utiliza la sentencia: "***COPY players (grupo, equipo, numero, posicion, nombre, dia, mes, year, club, liga, capitan) from '/home/vagrant/apache-cassandra-3.11.0/WorldCup2014.csv' with delimiter = ';' and reader = 'true';***"

```
cqlsh:wc2014> COPY players (grupo, equipo, numero, posicion, nombre, dia, mes, y
ear,club, liga, capitan) from '/home/vagrant/apache-cassandra-3.11.0/WorldCup201
4.csv' WITH delimiter = ',' and header = 'true';_
```

Se procesan las líneas del fichero.

```
cqlsh:wc2014> Copy players (grupo, equipo, numero, posicion, nombre, dia, mes, y
ear, club, liga, capitan) from '/home/vagrant/apache-cassandra-3.11.0/WorldCup20
14.csv' with delimiter = ';' and header = 'true';
Using 1 child processes

Starting copy of wc2014.players with columns [grupo, equipo, numero, posicion, n
ombre, dia, mes, year, club, liga, capitan].
Processed: 736 rows; Rate:     718 rows/s; Avg. rate:    1190 rows/s
736 rows imported from 1 files in 0.619 seconds (0 skipped).
```

Se puede hacer un recuento sobre la tabla al terminar para comprobar que se han importado todos los registros del fichero.

```
cqlsh:wc2014> select count(*) from players;

 count
-------
   736

(1 rows)
```

A la hora de realizar la importación se han encontrado un par de problemas:

1. Primero dio error la importación, y se tuvo que ir a ver el fichero de error generado en la importación, y se vio que ocurría porque se estaba poniendo como delimitador al copiar el carácter ‘,’ y el correcto es ‘;’.

```
vagrant@vagrant-ubuntu-trusty-64:~/apache-cassandra-3.11.0$ ls
bin                                          javadoc
CHANGES.txt                                  lib
conf                                         LICENSE.txt
data                                         logs
doc                                          NEWS.txt
import_wc2014_players.err.20171022_180836    NOTICE.txt
import_wc2014_players.err.20171023_182117    pylib
import_wc2014_players.err.20171023_182512    tools
interface                                    WorldCup2014.csv
```

2. Si al crear la tabla se optaba por poner como claves primarias las columnas por las que se dijo que iban a realizar las búsquedas: capitan, year, mes, dia, posicion; al importar los datos del fichero y recontar no se obtenían todos los registros, sino 720, menos de los que contiene el fichero.

```
cqlsh:wc2014> copy players (grupo, equipo, numero, posicion, nombre, dia, mes,ye
ar, club, liga, capitan) from '/home/vagrant/apache-cassandra-3.11.0/WorldCup201
4.csv' with delimiter = ';' and header = 'true';
Using 1 child processes

Starting copy of wc2014.players with columns [grupo, equipo, numero, posicion, n
ombre, dia, mes, year, club, liga, capitan].
Processed: 736 rows; Rate:     991 rows/s; Avg. rate:    1559 rows/s
736 rows imported from 1 files in 0.473 seconds (0 skipped).
cqlsh:wc2014> select count(*) from players;

 count
-------
   720

(1 rows)
```

Al optar por crear la tabla con las columnas de nombre, equipo y numero, si se importan todos los registros del fichero.
Lo que ocurre es que entonces luego al realizar consultas y querer filtrar por columnas que no son clave primaria se debe optar o por crear primero un índice con la columna requerida, o utilizar la sentencia “*allow filtering*” que permite filtrar por una columna que no esté como primary key.

4. (**1,5 puntos**) Listar todos los jugadores que fueron capitanes, identificando en qué equipo jugaba el capitán más longevo.

Con la sentencia "***select nombre, equipo, capitán from players where capitan = 'SI';***" se obtiene el listado de los futbolistas que eran capitanes en el mundial.
Si la columna capitan no era clave primaria, primero se crea un índice para filtrar por esta columna.

```
cqlsh:wc2014> create index on players (capitan);
cqlsh:wc2014> select nombre, equipo, capitan from players where capitan = 'SI';
```

Se obtiene la lista con los 32 capitanes. Las columnas a mostrar pueden ser las del ejemplo o añadir/eliminar según la información del registro que se quiera mostrar.

Por comodidad visual se listan sólo algunas columnas.

```
       Thiago Silva (c) |                Brasil |      SI
      Asamoah Gyan (c) |                 Ghana |      SI
       Mario Yepes (c) |              Colombia |      SI
      G?khan Inler (c) |                 Suiza |      SI
     Claudio Bravo (c) |                 Chile |      SI
    Steven Gerrard (c) |             Inglaterra |      SI
        Bryan Ruiz (c) |            Costa Rica |      SI
      Samuel Eto'o (c) |               Camerun |      SI
     Didier Drogba (c) |       Costa de Marfil |      SI
   Noel Valladares (c) |              Honduras |      SI
     Clint Dempsey (c) |                 USA |      SI
     Makoto Hasebe (c) |                 Japon |      SI
       Hugo Lloris (c) |              Francia |      SI
      Philipp Lahm (c) |              Alemania |      SI
    Rafael Marquez (c) |                Mexico |      SI
Giorgos Karagounis (c) |                Grecia |      SI
      Lionel Messi (c) |             Argentina |      SI
   Vincent Enyeama (c) |               Nigeria |      SI
      Mile Jedinak (c) |             Australia |      SI
    Roman Shirokov (c) |                 Rusia |      SI
  Gianluigi Buffon (c) |                Italia |      SI

(32 rows)
```

Si se ha creado la tabla con las primary key para las búsquedas se puede obtener el capitán más longevo de la forma:

"***Select nombre, equipo, year from players where capitan = 'SI' order by year asc limit 1;***"

Se hace un limit 2 con la misma sentencia para observar que sólo hay uno nacido en el primer año que aparece y por tanto es el mayor.

```
cqlsh:wc2014> select nombre, equipo, year from players where capitan = 'SI' orde
r by year asc limit 2;

 nombre                 | equipo   | year
------------------------+----------+------
        Mario Yepes (c) | Colombia | 1976
 Giorgos Karagounis (c) |   Grecia | 1977

(2 rows)
cqlsh:wc2014> select nombre, equipo, year from players where capitan = 'SI' orde
r by year asc limit 1;

 nombre          | equipo   | year
-----------------+----------+------
 Mario Yepes (c) | Colombia | 1976

(1 rows)
```

5. (**1 punto**) Contabilizar el número total de jugadores que participaron en dicho Mundial.

Con la sentencia “select * from players;” se saca la lista de todos los registros y sus columnas, y con count se hace el recuento obteniendo el número de registros que se corresponderá con el número de jugadores participantes del mundial:
“***select count(*) from players;***”

```
cqlsh:wc2014> select count(*) from players;

 count
-------
   736

(1 rows)
```

6. (**1 punto**) Listar los jugadores que tenían menos de 30 años (de 1 a 29).

Con la sentencia "***select nombre, equipo, year from players where year > 1984***" se obtendría una lista con los jugadores que al menos han nacido en 1985 y por tanto durante el mundial era seguro que tenían un máximo de 29 años.
Si year no es una columna clave primaria se puede crear un índice antes de la consulta o en la misma utilizar allow filtering.
Por comodidad visual se listan sólo algunas columnas.

7. (**1 punto**) Listar el jugador más mayor y más pequeño que jugó el mundial.

Si se ha creado la tabla con las primary key para las búsquedas se puede obtener el jugador más longevo de la forma:

"***Select nombre, equipo, year from players where capitan in ('SI','NO') order by year asc limit 1;***"
Como capitan puede tener dos valores, Si o No, y es clave primaria la utilizamos en la cláusula where con IN y así se cubren todos los jugadores y se pueden ordenar en base al año de nacimiento.
Se hace un limit 2 con la misma sentencia para observar que sólo hay uno nacido en el primer año que aparece y por tanto es el mayor.

Por comodidad visual se listan sólo algunas columnas.

```
cqlsh:wc2014> select nombre, equipo, year from players where capitan in ('SI','N
O') order by year asc limit 2;

 nombre          | equipo   | year
-----------------+----------+------
 Faryd Mondragon | Colombia | 1971
 Mario Yepes (c) | Colombia | 1976

(2 rows)
cqlsh:wc2014> select nombre, equipo, year from players where capitan in ('SI','N
O') order by year asc limit 1;

 nombre          | equipo   | year
-----------------+----------+------
 Faryd Mondragon | Colombia | 1971

(1 rows)
```

Y el jugador más joven:

"***Select nombre, equipo, year from players where capitan in ('SI','NO') order by year desc limit 1;***"

```
cqlsh:wc2014> select nombre, equipo, year from players where capitan in ('SI','N
O') order by year desc limit 2;

 nombre         | equipo     | year
----------------+------------+------
 Fabrice Olinga |    Camerun | 1996
      Luke Shaw | Inglaterra | 1995

(2 rows)
cqlsh:wc2014> select nombre, equipo, year from players where capitan in ('SI','N
O') order by year desc limit 1;

 nombre         | equipo  | year
----------------+---------+------
 Fabrice Olinga | Camerun | 1996

(1 rows)
```

8. (**1 punto**) Listar los jugadores que fueron delanteros, mostrando únicamente la siguiente información: nombre y el equipo al que pertenecían.

Con la sentencia "***select nombre, equipo from players where position = 'DELANTERO';***" se obtiene la lista de jugadores que su demarcación es la de delantero, mostrando únicamente las columnas de nombre y equipo como explícitamente se solicita. Se obtienen 139 jugadores que fueron delanteros en el mundial.
Si la posición no fuera parte de la clave primaria se crea antes un índice con dicha columna.

```
cqlsh:wc2014> create index on players (posicion);
cqlsh:wc2014> select nombre, equipo from players where posicion = 'DELANTERO';
```

```
          Shola Ameobi |    Nigeria
           Uche Nwofor |    Nigeria
            Tim Cahill |  Australia
         Mathew Leckie |  Australia
          Adam Taggart |  Australia
          Yuri Zhirkov |      Rusia
     Aleksandr Kokorin |      Rusia
   Aleksandr Kerzhakov |      Rusia
         Aleksei Ionov |      Rusia
     Aleksandr Samedov |      Rusia
     Maksim Kanunnikov |      Rusia
       Mario Balotelli |     Italia
       Antonio Cassano |     Italia
         Alessio Cerci |     Italia
         Ciro Immobile |     Italia
       Lorenzo Insigne |     Italia

(139 rows)
```