

# Práctica Impala

Alberto Armijo Ruiz

<b>Base de datos usada</b>	<b>3</b>
<b>Sentencias Impala</b>	<b>3</b>
Proyección	3
Selección.	4
Agrupamiento y cálculo sobre grupos	5

# Base de datos usada

Como base de datos se ha utilizado un conjunto de test de un conjunto de datos sacado de la página web de la Universidad de Irvine, el conjunto de datos se puede encontrar en el siguiente enlace: [DotaGames](#)

Este conjunto de datos contiene 102944 instancias en total; de estos el 10% pertenecen al conjunto de test. Cada una de las instancias de este conjunto de datos contienen los datos de una partida del videojuego *Dota2*. Por cada partida hay 116 columnas, las 4 primeras representan el equipo ganador de la partida, el ID del clúster, el modo de juego de la partida, el tipo de partida ( competitivo, público o privada ); el resto de columnas representan los diferentes personajes del juego, cada una de estas columnas contienen qué equipo lo ha elegido, si es que ha sido elegido.

El conjunto de datos que se ha utilizado es una modificación del conjunto de datos de test; ya que dicho conjunto utiliza valores enteros para representar las diferentes opciones. En el conjunto de datos utilizado estos valores enteros se han cambiado por valores categóricos para darle más sentido a las sentencias dentro de impala. El conjunto de datos se puede encontrar en el siguiente enlace:

<https://drive.google.com/file/d/1c2Pr1-CgL7FTEdc4Z9p56m8vWe9zUxOT/view>

Una vez dentro de la máquina virtual de Cloudera, se debe importar el archivo .csv a la sistema de archivos hdfs. Tras esto, se debe entrar a la shell de Impala y se crea una base de datos, en este caso llamada *practica*. Una vez creada la base de datos se crea una tabla para contener el archivo .csv.

## Sentencias Impala

### Proyección

Como sentencia de proyección se han elegido las columnas de equipo ganador, modo de juego y tipo de juego. En la siguiente imagen se puede ver los resultados de la sentencia.

m	Single Draft	Unranked Game
	Captain's Mode	Unranked Game
	Captain's Mode	Ranked Game
	The Greeviling	Unranked Game
	Captain's Mode	Ranked Game
	Captain's Mode	Unranked Game
	Captain's Mode	Unranked Game
m	Captain's Mode	Unranked Game
m	Captain's Mode	Unranked Game
m	Captain's Mode	Ranked Game
m	Reverse Captain's Mode	Ranked Game
	Captain's Mode	Unranked Game
	Captain's Mode	Ranked Game
m	Captain's Mode	Unranked Game
m	The Greeviling	Unranked Game
m	The Greeviling	Unranked Game
m	Captain's Mode	Ranked Game
	Captain's Mode	Ranked Game

-----+-----+-----+  
295 row(s) in 3.88s  
t.cloudera:21000] > select WinnerTeam,GameMode,GameType from DotaGames

## Selección.

En este apartado se han creado diferentes apartados ejemplos. El primer ejemplo utiliza una la sentencia del apartado anterior y selecciona solamente aquellas filas de partidas *Unranked*. La salida es la siguiente:

Blue Team	The Greeviling	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Red Team	Captain's Mode	Unranked Game
Red Team	Captain's Mode	Unranked Game
Blue Team	Single Draft	Unranked Game
Red Team	Captain's Mode	Unranked Game
Red Team	The Greeviling	Unranked Game
Red Team	Captain's Mode	Unranked Game
Red Team	Captain's Mode	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Red Team	Captain's Mode	Unranked Game
Blue Team	Captain's Mode	Unranked Game
Blue Team	The Greeviling	Unranked Game
Blue Team	The Greeviling	Unranked Game

-----+-----+-----+  
Fetched 6284 row(s) in 2.05s  
[quickstart.cloudera:21000] > select WinnerTeam, GameMode,GameType from DotaGames  
where GameType = 'Unranked Game';

El segundo ejemplo devuelve el número de partidas ganadas cuando se utiliza el personaje Axe en cualquiera de los dos equipos. La salida es la siguiente:

```
[quickstart.cloudera:21000] > select count(WinnerTeam) from DotaGames where axe=
Blue Team' or axe='Red Team';
Query: select count(WinnerTeam) from DotaGames where axe='Blue Team' or axe='Red
Team'
Query submitted at: 2019-03-20 10:47:08 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=ef4edc52720e0add:c65aaa4300000000
+-----+
| count(winnerteam) |
+-----+
| 2232               |
+-----+
Fetched 1 row(s) in 0.19s
```

El tercer ejemplo muestra el número de partidas ganadas cuando no se utiliza en personaje *Ogre Magi* en la partida. La salida es la siguiente:

```
[quickstart.cloudera:21000] > select count(WinnerTeam) from DotaGames where ogre
_magi='None';
Query: select count(WinnerTeam) from DotaGames where ogre_magi='None'
Query submitted at: 2019-03-20 10:56:26 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=aa4d94cfa24cde7d:3d0f348e00000000
+-----+
| count(winnerteam) |
+-----+
| 9999               |
+-----+
Fetched 1 row(s) in 0.17s
```

## Agrupamiento y cálculo sobre grupos

Para las operaciones de agrupamiento y cálculo sobre grupos se muestran dos ejemplos. El primer ejemplo devuelve los tipos de modo de juego que se juegan cuando el tipo de partida es Competitivo (*Ranked*) y el número de partidas jugadas por cada modo.

```
[quickstart.cloudera:21000] > select GameMode, count(GameMode) from DotaGames wh
ere GameType='Ranked Game' group by GameMode;
Query: select GameMode, count(GameMode) from DotaGames where GameType='Ranked Ga
me' group by GameMode
Query submitted at: 2019-03-20 10:41:11 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=d54541ff7bec6d18:b1418e6a000000000
+-----+-----+
| gamemode | count(gamemode) |
+-----+-----+
| Reverse Captain's Mode | 510 |
| Intro | 24 |
| Captain's Mode | 3471 |
+-----+-----+
```

El segundo ejemplo el número de partidas ganadas por cada uno de los equipos. Por lo que se puede ver en la salida, hay una partida que ha quedado en empate, ya que no aparece dentro de ninguno de los dos equipos.

```
part (missing from group by clause): winnerTeam

[quickstart.cloudera:21000] > select WinnerTeam, count(GameType) from DotaGames
group by WinnerTeam;
Query: select WinnerTeam, count(GameType) from DotaGames group by WinnerTeam
Query submitted at: 2019-03-20 10:34:10 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=de46b68e650332b4:35930d76000000000
+-----+-----+
| winnerteam | count(gametype) |
+-----+-----+
| 1 | 1 |
| Blue Team | 5502 |
| Red Team | 4792 |
+-----+-----+
Fetched 3 row(s) in 0.31s
[quickstart.cloudera:21000] > select WinnerTeam, count(GameType) from DotaGames
group by WinnerTeam;
```

El último ejemplo hace una combinación de selección, proyección, agrupamiento y cálculo sobre grupos. En este ejemplo se calcula el número de partidas ganadas por el equipo azul en cada uno de los modos de juego.

```
[quickstart.cloudera:21000] > select GameMode, count(GameMode) from (select WinnerTeam, GameMode from DotaGames where WinnerTeam='Blue Team') bluewinner group by GameMode;
```

```
Query: select GameMode, count(GameMode) from (select WinnerTeam, GameMode from DotaGames where WinnerTeam='Blue Team') bluewinner group by GameMode
```

```
Query submitted at: 2019-03-23 11:28:20 (Coordinator: http://quickstart.cloudera:25000)
```

```
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?query_id=9c4af9efc9606b8f:cec624ee00000000
```

gamemode	count(gamemode)
Reverse Captain's Mode	502
The Greeviling	599
Single Draft	9
Intro	35
Diretide	12
Random Draft	21
Captain's Mode	4242
All Pick	82

```
Fetches 8 row(s) in 0.31s
```