# ETL y SQL

Basándonos en el archivo “**NBA\_dataset.xlsx**” realizar las siguientes tareas:

Consideraciones

* Crear en todos los casos las estructuras de datos que sean necesarias
* Se deben entregar los archivos .ktr, .kjb y los scripts en caso de haber creado estructuras
* Utilizar identificadores de cargas, registrar de forma manual el log de las cargas en tablas o archivos auxiliares
* No se deben entregar archivos generados por los procesos de carga

## Warm Up:

Repasar funciones de Pentaho data Integration: Joins, Groups, Uniques, Formulas.

## 1)

Crear un proceso de ETL que genere un archivo *.csv* con todos los **nombres** **únicos** de jugadores

## 2)

Crear un proceso de ETL que genere un archivo *.csv* con todos los jugadores y la siguiente información:  
- Identificador  
- Nombre

- Cantidad de partidos jugados

- Minutos **totales** jugados

- Puntos **totales** convertidos

## 3)

Crear un proceso de ETL que alimente 2 tablas:

* En primer lugar la dimensión “Jugador”, donde la información relevante es el identificador y el nombre
* En segundo lugar la tabla de hechos “Promedios”, donde se busca medir, a nivel de granularidad “Jugador” las estadísticas **porcentuales.**

## 4)

Crear un proceso de ETL que genere un único archivo *.xlsx* con todas las estadísticas de los jugadores

## 5)

Extendiendo el proceso del ejercicio 4), guardar únicamente jugadores con nombre único (Si existen dos jugadores de nombre Juan Perez, guardar el primero de ellos y promediar sus estadísticas)