

Ejercicio 1

Dentro del learning path propuesto, como resumen de lo visto hasta el momento, y también con la idea de disponer de una base sobre la que trabajar en las tres sesiones online, vamos a crear un DataWarehouse en nuestro clúster utilizando Hive. Como fuente de datos, utilizaremos unos ficheros con estadísticas del videojuego multijugador League of Legends (https://en.wikipedia.org/wiki/League_of_Legends)

El ejercicio consiste en obtener una serie de información estadística sobre las partidas de League of Legends, siguiendo los siguientes pasos

- Utilizando Hive, crear la estructura de tablas necesaria para almacenar la información de los ficheros champions.txt, participants.txt y stats.txt
 - champions.txt contiene Id y Name de los personajes utilizados por un jugador humano en una partida
 - participants.txt contiene, en cada línea, información de la configuración utilizada por un jugador humano en una partida, relacionando la tabla Champions y la tabla Stats mediante las columnas ChampionID y StatsID
 - stats.txt contiene multitud de indicadores sobre el rendimiento de cada campeón en una partida real. Cada línea representa una partida, e incluye valores como:
 - winner, booleano que nos indica si el campeón ganó o perdió la partida
 - kills, número de campeones que el campeón ha matado en la partida
 - deaths, número de veces que ha muerto el campeón durante la partida
 - totaldamagedealt, daño infligido en la partida
 - totaldamagetaken, daño recibido en la partida
- Cargar en las tablas la información de los ficheros champions.txt, participants.txt y stats.txt
- Crear una consulta en HiveQL que nos muestre la siguiente información de cada campeón:
 - Nombre del campeón
 - Porcentaje de victorias
 - Media de campeones que el campeón ha matado en cada partida
 - Media de veces que ha muerto el campeón en cada partida
 - Total de daño infligido
 - Total de daño recibido
- Almacenar la información de la consulta anterior en una tabla de Hive

Los ficheros champions.txt, participants.txt y stats.txt están disponibles en <https://github.com/martinezmiranda/hadooptraining> dentro de la carpeta Ejercicio 1

Veremos una posible solución a este ejercicio en la primera sesión online, Procesado de Streams sobre Hadoop y Azure Stream Analytics, para la que podéis registraros aquí <https://goo.gl/mjBHRp>. Además, después de la sesión propondremos una aplicación de los nuevos conceptos vistos para ampliar este ejercicio.