

Conceptos y aplicaciones en Big Data

Trabajo Práctico 1 – Modalidad de cursada presencial

MapReduce

Pautas generales

- La entrega consiste en la implementación de Jobs MapReduce, resolviendo todas las consignas presentes en este enunciado. Se deberá entregar el código fuente implementado y un documento con el DAG de cada uno de los problemas planteados.
- Los alumnos pueden conformar grupo de no más de dos integrantes y hacer una única entrega grupal.
- La entrega se realiza por la mensajería del curso en IDEAS.
- La fecha límite de entrega es el 13 de octubre de 2025.

Enunciado

Ha sido contratado para el análisis de uso de un videojuego online donde millones de jugadores se conectan a diario para combatir con otros jugadores. En un combate, un jugador “retador” elige un oponente (jugador “retado”) y como resultado del combate el retador obtiene un puntaje y el tiempo en segundos que duró el combate. Por cada combate, los servidores del videojuego almacena la siguiente información:

- ID_Jugador_Retador
- ID_Jugador_Retado
- Puntaje obtenido (por el retador)
- Tiempo del combate en segundos

Todos los jugadores pueden participar cuántas veces deseen y por cada combate obtienen un puntaje y también se registra el tiempo que duró el combate. Un jugador puede combatir todas las veces que quiera.

Un fragmento del dataset de este jugador podría ser:

ID_Jug_Retador	ID_Jugador_Retado	Puntos	Tiempo
231	492	1054	621
231	492	2068	504
231	98	789	302
492	501	5462	955

Semanalmente los administradores del juego quieren obtener algunas estadísticas para regalarle premios a los mejores jugadores y armar el ranking semanal. Al finalizar la semana se desea saber.

- 1) El jugador más “retador” y el jugador más “retado”.
- 2) El jugador que más puntos obtuvo en promedio :

$$PP_i = \frac{\text{puntaje_total_de_todos_los_combates}_i + 1}{\text{cantidad_de_combates_como_retador}_i + 1}$$
 (el +1 en el numerador y en el denominador es para evitar divisiones por cero en el caso que un jugador no haya “retado” a nadie, en cuyo caso tendrá el puntaje mínimo de 1).
- 3) Todos los jugadores que “retaron” a más de H oponentes **distintos** (H es un parámetro de la consulta).
- 4) El top 10 de los jugadores con mejor puntaje heroico. El puntaje PH de un jugador i se calcula como:

$$PH_i = \alpha \left(\sum_{j=1}^n PH_j \cdot PP_i / PP_j \right) + (1 - \alpha)$$

donde n es la cantidad de jugadores “retados” por el jugador i , PH_j es el puntaje P del j -ésimo jugador retado por i , PP_i el puntaje promedio del jugador i , PP_j el puntaje promedio del j -ésimo jugador retado por i y α es un parámetro del algoritmo.

Cómo para calcular el puntaje PH de un jugador i previamente hay que calcular el puntaje de los jugadores a quienes retó (y a su vez éstos pueden haber retado al jugador i), el cálculo se realiza de manera iterativa.

Se comienza asignando el mismo puntaje heroico con un valor arbitrario a todos los jugadores (tiempo 0), con esos puntajes arbitrarios se calcula el puntaje usando la fórmula detallada para obtener los nuevos puntajes (tiempo 1), con éstos se calculan los nuevos puntajes (tiempo 2) y así se continúa hasta alcanzar una convergencia en los valores. Al final de este documento hay un ejemplo que explica este cálculo iterativo.

Implemente la cantidad de jobs necesarios para resolver las cuatro consultas. Utilice el archivo jugadores.txt para realizar las pruebas (Se ofrecen los resultados de cada consulta).

Aclaración

Suponga que la cantidad de jugadores y de combates es “Big Data”.

En cada job planteado en la solución piense si una función combiner contribuye o no para la optimización del job. En caso de contribuir implemente dicha función.

Ejemplo del cálculo del puntaje de los jugadores

Supongamos que cuatro jugadores hicieron los siguientes retos:

Jugador 1 → retó a los jugadores 2 y 3

Jugador 2 → retó a los jugadores 1 y 3

Jugador 3 → retó al jugador 4

Jugador 4 → no retó a nadie

y que al calcular el puntaje promedio de cada jugador da los siguientes valores:

Jugador 1 → 12 puntos en promedio

Jugador 2 → 11 puntos en promedio

Jugador 3 → 18 puntos en promedio

Jugador 4 → 1 punto (no retó a nadie, tiene el puntaje mínimo)

y usamos un valor de α de 0.9.

1. Tiempo 0: Inicialmente todos tienen el mismo puntaje: 1 (valor arbitrario)

Jugador 1 → 1 punto heroico

Jugador 2 → 1 punto heroico

Jugador 3 → 1 punto heroico

Jugador 4 → 1 punto heroico

2. Tiempo 1: Se calculan los nuevos puntajes heroicos usando los puntajes heroicos del tiempo 0:

$$\begin{aligned}\text{Jugador 1} &\rightarrow \alpha * (PH_2 * PP_1 / PP_2 + PH_3 * PP_1 / PP_3) + (1 - \alpha) = \\ &= 0.9 * (1 * 12 / 11 + 1 * 12 / 18) + (1 - 0.9) = 1.6818\end{aligned}$$

$$\begin{aligned}\text{Jugador 2} &\rightarrow \alpha * (PH_1 * PP_2 / PP_1 + PH_3 * PP_2 / PP_3) + (1 - \alpha) = \\ &= 0.9 * (1 * 11 / 12 + 1 * 11 / 18) + (1 - 0.9) = 1.475\end{aligned}$$

$$\text{Jugador 3} \rightarrow \alpha * (PH_4 * PP_3 / PP_4) + (1 - \alpha) = 0.9 * (1 * 18 / 1) + (1 - 0.9) = 16.3$$

$$\text{Jugador 4} \rightarrow \alpha + (1 - \alpha) = 0.9 + (1 - 0.9) = 1$$

3. Tiempo 2: Se calculan los nuevos puntajes heroicos usando los puntajes heroicos del tiempo 1:

$$\begin{aligned}\text{Jugador 1} &\rightarrow \alpha * (PH_2 * PP_1 / PP_2 + PH_3 * PP_1 / PP_3) + (1 - \alpha) = \\ &= 0.9 * (1.475 * 12 / 11 + 16.3 * 12 / 18) + (1 - 0.9) = 11.3281\end{aligned}$$

$$\begin{aligned}\text{Jugador 2} &\rightarrow \alpha * (PH_1 * PP_2 / PP_1 + PH_3 * PP_2 / PP_3) + (1 - \alpha) = \\ &= 0.9 * (1.6818 * 11 / 12 + 16.3 * 11 / 18) + (1 - 0.9) = 10.4525\end{aligned}$$

$$\text{Jugador 3} \rightarrow \alpha * (PH_4 * PP_3 / PP_4) + (1 - \alpha) = 0.9 * (1 * 18 / 1) + (1 - 0.9) = 16.3$$

$$\text{Jugador 4} \rightarrow \alpha + (1 - \alpha) = 0.9 + (1 - 0.9) = 1$$

4. Tiempo 3: Se calculan los nuevos puntajes heroicos usando los puntajes heroicos del tiempo 2:

$$\begin{aligned}\text{Jugador 1} &\rightarrow \alpha * (PH_2 * PP_1 / PP_2 + PH_3 * PP_1 / PP_3) + (1 - \alpha) = \\ &= 0.9 * (10.4525 * 12 / 11 + 16.3 * 12 / 18) + (1 - 0.9) = 20.1424\end{aligned}$$

$$\begin{aligned}\text{Jugador 2} &\rightarrow \alpha * (PH_1 * PP_2 / PP_1 + PH_3 * PP_2 / PP_3) + (1 - \alpha) = \\ &= 0.9 * (11.3281 * 11 / 12 + 16.3 * 11 / 18) + (1 - 0.9) = 18.4107\end{aligned}$$

$$\text{Jugador 3} \rightarrow \alpha * (PH_4 * PP_3 / PP_4) + (1 - \alpha) = 0.9 * (1 * 18 / 1) + (1 - 0.9) = 16.3$$

$$\text{Jugador 4} \rightarrow \alpha + (1 - \alpha) = 0.9 + (1 - 0.9) = 1$$

Y así se continúa iterando hasta alcanzar una convergencia con la precisión deseada. En este ejemplo luego de 100 iteraciones los puntajes heroicos de los jugadores, cambiando en el segundo o tercer decimal, son:

$$\text{Jugador 1} \rightarrow 98.8402$$

$$\text{Jugador 2} \rightarrow 90.6079$$

$$\text{Jugador 3} \rightarrow 16.3$$

$$\text{Jugador 4} \rightarrow 1$$