

## **Actividad a entregar nº1**

El dataset website tiene información sobre el tiempo de permanencia de sus usuarios en cada una de las páginas del sitio. El formato de los datos del dataset es:

`<id_user, id_page, time>`

Modele una solución MapReduce, sugiriendo la utilización de funciones combiners en los casos que considere necesario, que permita obtener:

- El usuario que más páginas distintas visitó
- La página más visitada (en cuanto a tiempo de visita) por cada usuario siempre que la página haya sido visitada más de  $V$  veces por ese usuario ( $V$  es parámetro de la consulta).
- La página más visitada (en cuanto a cantidad de visitas, sin importar el tiempo de permanencia) por al menos  $U$  usuarios distintos ( $U$  es parámetro de la consulta).

Indique como queda el DAG del proceso completo (las dos consultas)

***Deberá entregar el código en Python de la resolución de las dos consultas y un documento con el DAG del proceso completo.***

***La resolución de este ejercicio debe ser entregado por la mensajería de IDEAS.***