

Conceptos y aplicaciones en Big Data

Trabajo Práctico 1 – Modalidad de cursada presencial

MapReduce - Operaciones con conjuntos

Pautas generales

- La entrega consiste en la implementación de Jobs MapReduce, resolviendo todas las consignas presentes en este enunciado.
- 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 31 de octubre de 2022.
- La calificación obtenida en este TP será tomada en cuenta en la nota final de la materia.

Enunciado

Un sitio de e-commerce, almacena para cada mes del año el ID de todos los productos vendidos ese mes. Un mismo ID podría estar repetido en un mismo mes si ese producto se vendió más de una vez en el mismo mes. El dataset solo posee el ID de los productos.

- 1) Implemente un job MapReduce que resuelva la unión de dos datasets (meses).

Ejemplo:

Mayo	Enero	Salida
3	2	3
8	4	8
2	7	4
8		7
		2

- 2) Implemente un job MapReduce que resuelva la intersección de dos datasets (meses).

Ejemplo:

Mayo	Octubre	Salida
3	2	2
8	4	3
2	3	
8		

- 3) Implemente un job MapReduce que permita determinar si dos datasets (meses) son iguales. Dos conjuntos son iguales si ambos tienen los mismos elementos sin importar cuántas veces se repitan dentro de cada conjunto.

Ejemplo:

Mayo	Septiembre
3	2
8	8
2	3
8	

Son iguales

Julio	Agosto
3	8
8	3
2	7
7	

NO son iguales

- 4) Indique para cada uno de los ejercicios anteriores si cada uno de los job implementados se pueden generalizar para N conjuntos. Es decir:
- Un único job que resuelva la unión de N conjuntos.
 - Un único job que resuelva la intersección de N conjuntos.
 - Un único job que permita determinar si N conjuntos son iguales entre si.
- Justifique su respuesta explicando, en caso afirmativo como sería la solución; y en caso negativo, por qué no es posible su resolución con un único Job.