



Conceptos y Aplicaciones de Big Data

Trabajo práctico 1
MapReduce

Prof. Waldo Hasperué
(whasperue@lidi.info.unlp.edu.ar)

Trabajo práctico 1

- Los alumnos pueden conformar un grupo y hacer una única entrega grupal.
- La entrega consiste en la implementación de un proyecto (Java o Python a elección) y un informe respondiendo a todas las consignas presentes en este enunciado.
- La fecha límite de entrega es el 23 de octubre de 2018.
- La calificación obtenida en este TP será tomada en cuenta en la nota final de la materia.

Enunciado

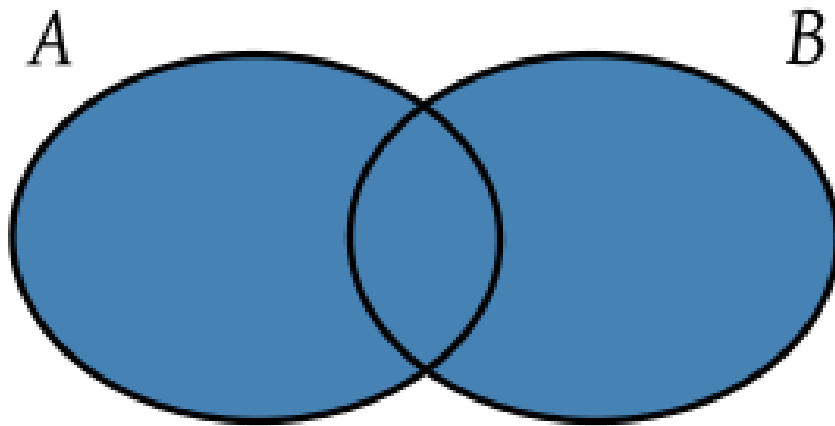
Implementar jobs (procesos map y reduce) que permitan hacer las operaciones básicas entre conjuntos:

- Unión
- Intersección
- Diferencia
- Igualdad
- Complemento
- Cardinalidad
- Pertenencia

para utilizarlos en la resolución de un problema más complejo.

Ejercicio 1 - Unión

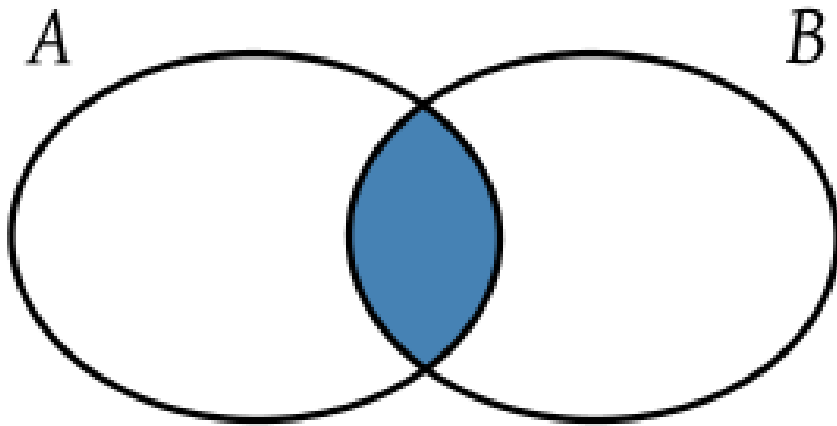
Implemente un job que lea dos conjuntos de datos A y B y genere como salida la unión entre A y B.



A	B	Salida
3	2	3
8	4	8
2	7	4
		7
		2

Ejercicio 2 - Intersección

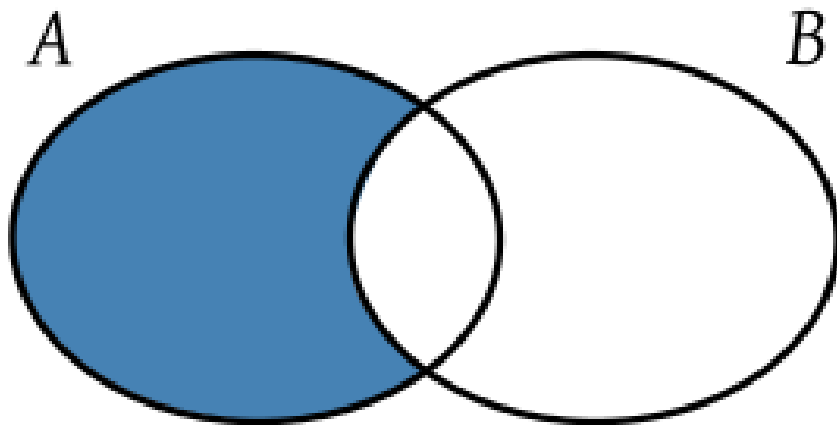
Implemente un job que lea dos conjuntos de datos A y B y genere como salida la diferencia entre A y B.



A	B	Salida
3	2	3
8	4	7
2	7	
5	1	
7		

Ejercicio 3 - Diferencia

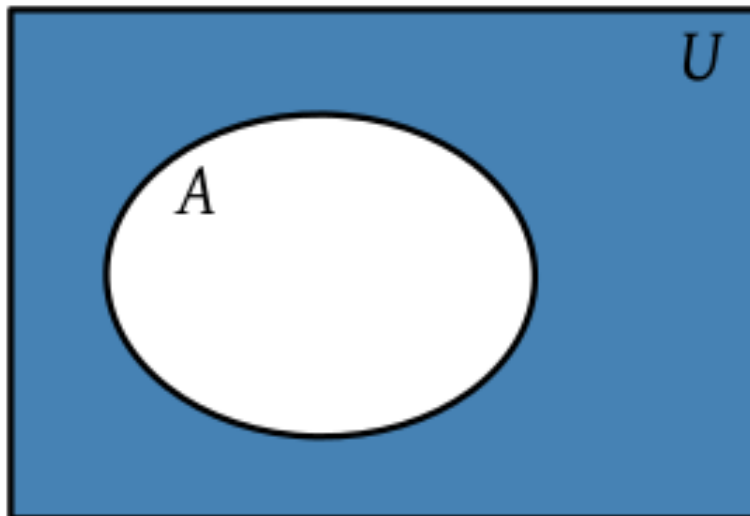
Implemente un job que lea dos conjuntos de datos A y B y genere como salida la diferencia A - B.



A	B	Salida
3	2	3
8	4	5
2	7	8
5	1	
7		

Ejercicio 4 - Complemento

Implemente un job que lea dos conjuntos de datos A y U y genere como salida el complemento de A, donde U es el universo.



A	U	Salida
3	1	1
5	2	4
2	3	
	4	
	5	

Ejercicio 5 - Cardinalidad

Implemente un job que lea un conjunto de datos A y genere como salida la cantidad de elementos que posee el conjunto.

A	Salida
3	5
8	
2	
4	
9	

Ejercicio 6 - Pertenencia

Implemente un job que lea un conjunto de datos A y una lista de elementos E y genere como salida si para cada elemento de E está o no en el conjunto A.

A	B	Salida
3	2	2 SI
8	4	4 NO
2	5	5 NO
6	1	1 SI
1	8	8 SI

Ejercicio 7 - Igualdad

Implemente un job que lea dos conjuntos de datos A y B, tal que permita determinar si A y B son iguales. Dos conjuntos son iguales si tienen los mismos elementos:

- Todos los elementos de A pertenecen a B
- Todos los elementos de B pertenecen a A

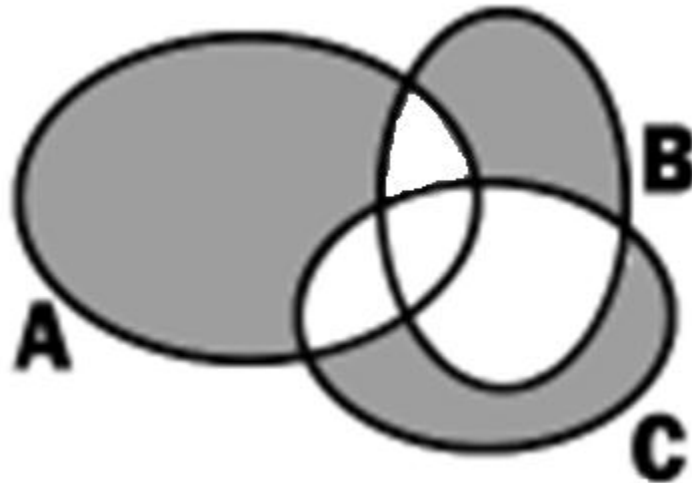
A	B	Salida
3	2	SI
8	3	
2	8	

A	B	Salida
3	2	NO
8	4	
2	7	

Ejercicio 8

Implemente un job que a partir de tres conjuntos A, B y C permita calcular el conjunto X:

$$X = [A - (B \cup C)] \cup [B - (A \cup C)] \cup [C - (A \cup B)]$$



$$U = A \cup B \cup C$$

Ejercicio 9

Implemente un programa que lea tres conjuntos de datos A, B y C y una lista de elementos E y que realice

- El cálculo del conjunto X del ejercicio 8, usando los conjuntos A, B y C.
- Que determine la cardinalidad de X.
- Que determine la pertenencia o no de los elementos de E en X.

Entrega

Deberá entregar:

- La implementación de los ejercicios 1 a 9
- Un documento detallando el funcionamiento de cada job implementado.
 - ¿Qué función cumple el mapper?
 - ¿Qué función cumple el reducer?
 - ¿Trae ventajas el uso de un combiner?