

Estructuras de Datos

Universidad Nacional de Tres de Febrero

Guía de ejercicios 6

Map Reduce

Hay que recordar que MapReduce tal como lo vimos en clase utiliza una arquitectura distribuida, que puede ser un cluster de computadoras, en las cuales hay una que cumple la función de master, que subdivide el problema y distribuye los subproblemas entre las otras computadoras del cluster, que se encargan de mapear es decir aplicar la función map a los datos para obtener resultados intermedios que se almacena y luego la computadora master distribuye tareas entre las otras computadoras para combinar los resultados intermedios para obtener el resultado final. Esta es la etapa del reduce. En Python existen las funciones map y reduce que nos van a permitir introducir el tema de MapReduce con algunas limitaciones, ya que las mismas trabajan sobre iterables. Map toma como argumentos una función y una colección iterable y le aplica la función a cada elemento de la colección, por ejemplo: Por ejemplo dada la lista [1,2,3,4,5] elevar al cuadrado sus elementos:

In []:

```
l=[1,2,3,4,5]
l2=map(lambda x: x*x, l)
for x in l2:
    print(x)
```

Para usar la función reduce es necesario importar functools. Dada la lista [1,2,3,4,5], elevar al cuadrado todos los elementos y sumarlos

In []:

```
import functools

l2=map(lambda x: x*x, [1,2,3,4,5])
res=functools.reduce(lambda x, y: x+y, l2, 0)
print(res)
```

Ejercicio 1

Dada una lista de enteros, calcular el producto de los elementos elevados al cubo

In []:

Ejercicio 2

Dada una lista de enteros encontrar el máximo en valor absoluto. Por ejemplo, dada la lista:

```
[5, -200, -7, 6]
```

Debe devolver 200.

Sugerencia: la etapa de map debe calcular el valor absoluto de cada elemento y en la etapa de reduce quedarse con el máximo

In []:

Ejercicio 3

Dada una lista de tuplas (artículo, cant, precio_unitario). Calcular el monto total de la venta.

Por ejemplo:

```
[("cuadernos",10,100),("lapiceras",20,2),("libros",5,150)]
```

Map debe mapear cada tupla a su valor total de venta reduce debe sumar todas las ventas

```
map ->[1000,40,750]
```

```
reduce->1790
```

In []:

Ejercicio 4

Se tienen datos de temperaturas en un archivo.txt con el siguiente formato y se quiere calcular la temperatura máxima de cada año. Formato:

Columnas 1 a 4: año

Columnas 5-6: mes

Columnas 7-8: día

Columna 9: signo de la temperatura (+ / -)

Columnas 10 a 12: temperatura * 10 (por ejemplo 346 se lee como 34.6 y 999 significa dato faltante)

Columna 13: Calidad del dato (si la calidad no es 1 o 2, el dato no sirve).

Escribir en Python dos funciones Map y Reduce con las siguientes especificaciones:

Map:

Lee cada línea de texto, separa año, temperatura y calidad y si la calidad es la adecuada y el dato no es faltante. Genera una línea y la escribe en un archivo de texto intermedio de la forma aaaa,ttt.

Reduce:

Lee el archivo de texto intermedio y produce el resultado deseado: año, máxima temperatura de cada año.

In []:

In []: