

Entrega: 13/12/2023

Noms: Mariona Farré Tapias Marc Pérez Guerrero

Activitat 5: Rein

Proves d'execució segons el nombre de nuclis (cores):

Hem intentat fer un parell de proves per veure si el nombre de nuclis afectava a l'eficiència dels scripts del map-reduce, ja que teòricament si hi han més nuclis l'execució dels scripts hauria de ser més ràpida, ja que permet processar les dades en paral·lel.

Hem fet unes quantes proves amb diferents nombres de cores, per veure quina tendència aproximada té cada core:

n cores	Temps primera execució	Temps segona execució	Temps tercera execució	Temps Aproximat
1	2.902473211288452 seconds	2.7202868461608887 seconds	2.5826969146728516 seconds	≈ 2.735 segons
2	2.0907001=49536133 seconds	1.9807045459747314 seconds	1.9018993377685547 seconds	≈ 1.991 segons
4	2.4057908058166504 seconds	2.511936664581299 seconds	2.4956395626068115 seconds	≈ 2.471 segons
8	2.412306785583496 seconds	2.430368185043335 seconds	2.4258131980895996 seconds	≈ 2.423 segons
10	2.32549786567688 seconds	2.639392614364624 seconds	2.4778215885162354 seconds	≈ 2.481 segons
20	3.9236302375793457 seconds	5.0635011196136475 seconds	3.8667328357696533 seconds	≈ 4.285 segons

Es pot observar que en aquests temps d'execució no és més eficient tenir més cores, al contrari, el temps pitjor donat és amb 20 cores.

Ens hem quedat amb el nombre **2 de cores** ja que és el més ràpid i el que ha tret el temps millor de totes les proves.

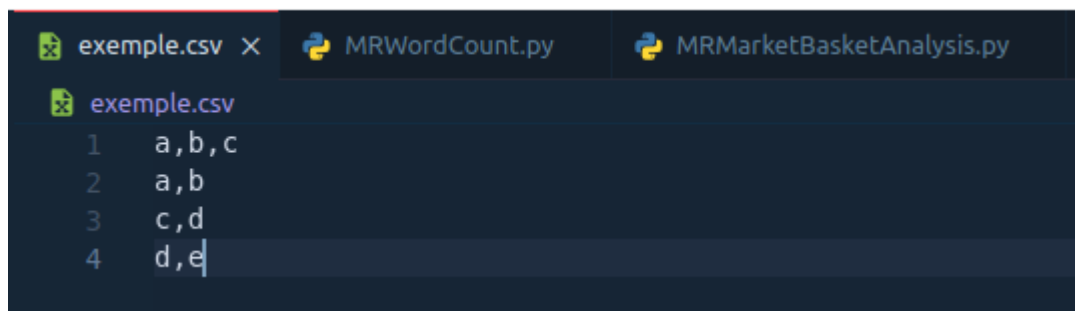
Dificultats i decisions en la implementació:

Les dificultats principals d'aquest treball ha sigut treballar amb mrjob, ja que era una llibreria de python que encara no havíem fet servir, i trobar informació i com poder implementar-ho s'ens ha fet una mica complicat.

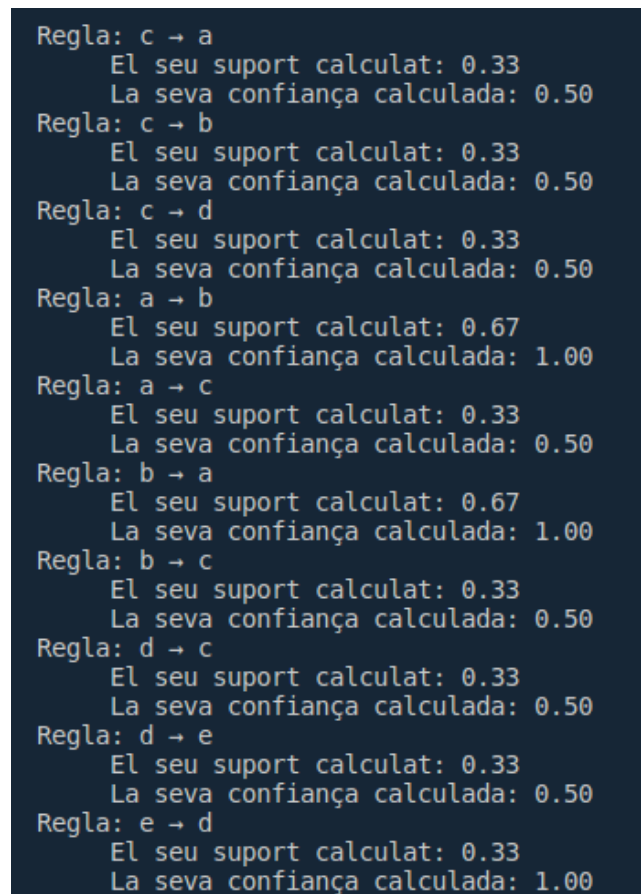
Vem tenir problemes amb la implementació de la lectura de parelles de dades, ja que al principi ens pensàvem que havíem de respectar dins d'una llista l'ordre dels ítems donats, i només fer parelles ítems que estiguessin seguits. Però el resultat ens donava uns números molt baixos, així que canviant el codi vam veure que la lectura de les parelles era per totes les combinacions possibles per cada llista d'ítems.

Per assegurar-nos que es calculava correctament el suport i la confiança, hem fet servir exemples molt més simples per poder veure visualment si els resultats tenien sentit, i així al fer servir el fitxer *groceries.csv* no tenir dubtes dels resultats donats.

Fitxers dels exemples fets:



```
exemple.csv x MRWordCount.py MRMarketBasketAnalysis.py
exemple.csv
1 a,b,c
2 a,b
3 c,d
4 d,e
```



```
Regla: c → a
  El seu suport calculat: 0.33
  La seva confiança calculada: 0.50
Regla: c → b
  El seu suport calculat: 0.33
  La seva confiança calculada: 0.50
Regla: c → d
  El seu suport calculat: 0.33
  La seva confiança calculada: 0.50
Regla: a → b
  El seu suport calculat: 0.67
  La seva confiança calculada: 1.00
Regla: a → c
  El seu suport calculat: 0.33
  La seva confiança calculada: 0.50
Regla: b → a
  El seu suport calculat: 0.67
  La seva confiança calculada: 1.00
Regla: b → c
  El seu suport calculat: 0.33
  La seva confiança calculada: 0.50
Regla: d → c
  El seu suport calculat: 0.33
  La seva confiança calculada: 0.50
Regla: d → e
  El seu suport calculat: 0.33
  La seva confiança calculada: 0.50
Regla: e → d
  El seu suport calculat: 0.33
  La seva confiança calculada: 1.00
```

Pregunta 1.

A partir de la vostra implementació, ompliu la taula següent; els valors de la primera fila ja se us donen calculats.

Fila	Suport	Confiança	Nre. de regles d'associació trobades
1	1%	1%	426
2	1%	25%	96
3	1%	50%	0
4	1%	75%	0
5	5%	25%	4
6	7%	25%	2
7	20%	25%	0
8	50%	25%	0

Els resultats en l'execució del fitxer.py:

```
mariona@mariona-laptop:~/Documentos/REIN/lab5$ python3 MRMarketBasketAnalysis.py --file groceries.csv --ncores 2
No configs specified for local runner
No configs specified for local runner
Time= 1.9370169639587402 seconds

Number of transactions 9835

*****
***** Values and rules to fill the required table *****
*****
Support= 0.01 confidence= 0.01. Rules found 426
Support= 0.01 confidence= 0.25. Rules found 96
Support= 0.01 confidence= 0.5. Rules found 0
Support= 0.01 confidence= 0.75. Rules found 0
Support= 0.05 confidence= 0.25. Rules found 4
Support= 0.07 confidence= 0.25. Rules found 2
Support= 0.2 confidence= 0.25. Rules found 0
Support= 0.5 confidence= 0.25. Rules found 0
```

Pregunta 2.

Doneu la llista de regles d'associació que hàgiu trobat corresponents a les files 4, 5 i 6 de la taula de la pregunta 1.

Hem fet un altre apartat en el codi per veure les regles que compleixen els valors de suport i confiança de les files indicades:

```
print("***** Les normes del suport i la confiança triada ***** ")
print("***** Les normes del suport i la confiança triada ***** ")
print("***** Les normes del suport i la confiança triada ***** ",'\n')
thresholds = [(0.01, 0.75), (0.05, 0.25), (0.07, 0.25)] # Llistes específics a mostrar
n=3
for support, conf in thresholds:
    nr = 0
    n=n+1
    print(f"Fila: {n},f": Regles que compleixen el suport >= {support} i la confiança >= {conf}:")

    for (item1, item2), pair_count in pairs.items():
        item1_count = singles.get(item1, 0)
        sup_parella = pair_count / ntrans
        conf_parella = pair_count / singles[item1]

        if sup_parella >= support and conf_parella >= conf:
            nr += 1
            print(f"    Regla: {item1} → {item2}")
            print(f"    El seu suport calculat: {sup_parella:.2f}")
            print(f"    La seva confiança calculada: {conf_parella:.2f}")

    print(f"Total de regles trobades: {nr}\n")
```

I el resultat de l'execució ens dona les següents regles:

```
***** Les normes del suport i la confiança triada *****
***** Les normes del suport i la confiança triada *****
***** Les normes del suport i la confiança triada *****

Fila: 4 : Regles que compleixen el suport >= 0.01 i la confiança >= 0.75:
Total de regles trobades: 0

Fila: 5 : Regles que compleixen el suport >= 0.05 i la confiança >= 0.25:
Regla: other vegetables → whole milk
El seu suport calculat: 0.07
La seva confiança calculada: 0.39
Regla: rolls/buns → whole milk
El seu suport calculat: 0.06
La seva confiança calculada: 0.31
Regla: whole milk → other vegetables
El seu suport calculat: 0.07
La seva confiança calculada: 0.29
Regla: yogurt → whole milk
El seu suport calculat: 0.06
La seva confiança calculada: 0.40
Total de regles trobades: 4

Fila: 6 : Regles que compleixen el suport >= 0.07 i la confiança >= 0.25:
Regla: other vegetables → whole milk
El seu suport calculat: 0.07
La seva confiança calculada: 0.39
Regla: whole milk → other vegetables
El seu suport calculat: 0.07
La seva confiança calculada: 0.29
Total de regles trobades: 2
```

Els Números total de regles ens quadren amb la pregunta 1, així que ens podem assegurar que són resultats correctes.