
Examen : Session Principale

Filière : Mastère Info. DS

A.U : 2022-2023

Niveau : Première année

Durée : 01H30

Matière : Introduction au Big Data

Documents : non autorisés

Enseignant: M. SAKKARI

Exercice 1 : (4 pts)

1. Rappelez les principes de map-reduce, quel intérêt a une telle méthode (1 pt)
2. Donnez la définition de map (1 pt)
3. Donnez la définition de reduce (1 pt)
4. Rappelez les étapes d'élaboration d'un traitement dans le paradigme Map-Reduce. (1 pt)

Exercice 2 : (4 pts)

Soit le programme suivant :

```
Opération Map: Entrée: document d
    Pour chaque mot dans d faire:
        Retourner (taille(mot), 1)
Opération Reduce: Entrée: (clé, [val1,...])
    int x = 0;
    Pour chaque val dans [val1,...] faire:
        x++;
    retourner (clé, x);
```

1. Que fait ce programme MapReduce? (1pt)
2. Dérouler sur l'exemple suivant les résultats des différentes étapes du programme ci-dessus. (3 pt)

**Le savoir, c'est le pouvoir.
Savoir pour prévoir, afin de pouvoir.**

Exercice 3 : (7 pts)

Depuis l'apparition de la pandémie du Coronavirus, un Institut garde trace de tous les échantillons reçus des différents gouvernorats, dans le but d'étudier la propagation du virus. Une ligne de données est formatée de la manière suivante:

`id_echantillon,date_prise,gouvernorat,sexe,age,resultat`

Ci-dessous un exemple d'un échantillon :

1652, 25/06/2020, Tunis, H, 35, N

1. Que fait ce programme MapReduce? (1pt)

```
Opération Map: Entrée: document d
  Pour chaque Ligne L dans d faire:
    Retourner (L[3], 1)
Opération Reduce:
  Entrée: (clé, [val1,....])
  int x = 0;
  Pour chaque val dans [val1,....] faire:
    x++;
  retourner (clé, x);
```

2. Ecrire en pseudo-code le programme MapReduce permettant de calculer le nombre de personnes testées positives/négatives par sexe. (3pts)

3. Dérouler sur l'exemple suivant le résultat des différentes étapes de votre programme.(3pts)

```
1652, 25/06/2020, Tunis, H, 35, N
1653, 25/06/2020, Béja, F, 85, P
1654, 26/06/2020, Sfax, F, 25, P
1655, 27/06/2020, Tunis, H, 64, N
1656, 29/06/2020, Sousse, F, 74, P
1657, 29/06/2020, Béja, H, 58, P
```

Exercice 4 : (5 pts)

En quoi consiste la règle des 5V dans le Big Data ?

Bon travail