



Devoir surveillé – S1 – 2023/2024

Filière : Ing2_Info	Matière : Bases de données NoSQL et Big Data		Enseignante : Asma KERKENI
Date : 09 / 11 / 2023	Nbr de Crédits : 3	Coefficient : 3	Documents autorisés : Non
Durée de l'épreuve : 1h	Régime d'évaluation : Mixte		Nombre de pages : 5
Nom & Prénom : .....			Matricule : .....
Signature :	Code confidentiel :		Classe : ..... N° Place : .....



NOTE : Répondre directement sur les feuilles de l'examen /

Note

/ 20

Exercice 1 : (8 points)

Dans le Q.C.M. suivant, chaque question admet une ou plusieurs bonne(s) réponse(s). Encadrer la ou les bonne(s) réponse(s).

**N.B** : une réponse juste et deux réponses fausses valent 0.

- Pour un répertoire sous HDFS, le facteur de réplication est :
  - identique au facteur de réplication de ses fichiers
  - 2
  - 4
  - aucune des réponses précédentes
- Dans un cluster Hadoop, si on ajoute **N Go** d'espace disque au NameNode. Combien de Go d'espace supplémentaire deviennent disponibles sur HDFS ?
  - environ **N Go**.
  - environ **N<sup>2</sup> Go**.
  - environ **N/2 Go**.
  - aucune des réponses précédentes
- Pour calculer la moyenne de température par région à partir de données météorologiques ayant la forme : **Région | Date | Température**, que doit faire la fonction Map ?
  - Extraire les températures de chaque enregistrement et les diviser par le nombre total d'enregistrements
  - Extraire les températures de chaque enregistrement et les stocker dans un tableau
  - Extraire les températures de chaque enregistrement et les associer à la région correspondante
  - Extraire les régions de chaque enregistrement et les associer aux températures correspondantes
  - aucune des réponses précédentes



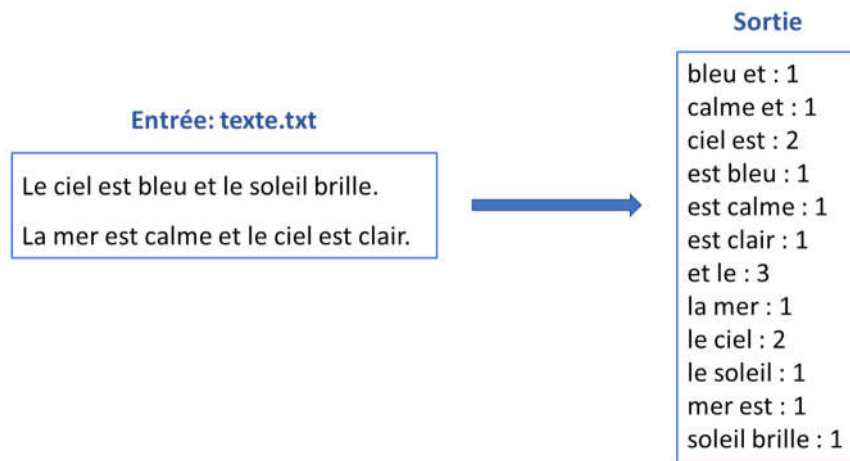
*Ne rien écrire ici*

4. Pour calculer la température maximale à partir de données météorologiques ayant la forme : **Région | Date | Température**, que doit faire la fonction Reduce ?
  - a. Trouver la valeur maximale dans chaque ensemble de températures
  - b. Additionner les températures de chaque enregistrement pour obtenir une somme totale
  - c. Diviser la somme totale des températures par le nombre total d'enregistrements
  - d. Extraire la température maximale de chaque enregistrement et la stocker dans un tableau
  - e. aucune des réponses précédentes
5. La scalabilité :
  - a. peut être assurée par l'augmentation des capacités des machines existantes
  - b. peut être assurée par l'ajout d'autres machines au cluster
  - c. est synonyme de *scale out*
  - d. est la capacité d'un système à gérer une augmentation de charge sans altérer ses performances
  - e. aucune des réponses précédentes
6. Pour répondre à la vélocité des données dans un système Big Data, il est essentiel que le système soit :
  - a. tolérant aux pannes
  - b. scalable
  - c. flexible
  - d. disponible
  - e. aucune des réponses précédentes
7. Quel est le démon responsable de la réplication des données dans Hadoop?
  - a. HDFS
  - b. Task Tracker
  - c. Job Tracker
  - d. NameNode
  - e. aucune des réponses précédentes

8. Quelle est l'action effectuée par la commande **"docker exec"**?
- exécuter une commande à l'intérieur d'un conteneur en cours d'exécution.
  - créer un nouveau conteneur à partir d'une image existante.
  - copier des fichiers entre le système de fichiers local et le système de fichiers du conteneur.
  - afficher des informations détaillées sur un conteneur spécifique.
  - aucune des réponses précédentes

### Exercice 2 : (12 points)

L'objectif de cet exercice est de développer un programme utilisant Hadoop MapReduce pour compter le nombre d'occurrences de bigrams dans un fichier « **texte.txt** » de taille de 128 To. Un bigram est une séquence de deux mots consécutifs dans une phrase. La figure ci-après montre un extrait d'un fichier et le résultat attendu par le programme.



1. Expliquez pourquoi il est nécessaire de diviser le fichier d'entrée en plusieurs blocs de taille égale (ici 64 Mo)?

.....

.....

.....

.....

2. Pour ce cluster, le facteur de réplication est égal à 3. Calculer le nombre de tâches Map nécessaires pour traiter ce fichier ?

.....

.....

.....

.....

.....

3. Illustrer une solution utilisant le modèle de programmation *MapReduce* sur l'exemple.



