esprit

Se former autrement

ETUDIANT(e)

EXAMEN

Semestre 1

2

Session Principale

Rattrapage

Nom et Prénom :

Code:

Classe :

Module: BIG DATA

Enseignantes :Nesrine BOUAZIZI, Sirine ZAABOUTI, Manel KHAMASSI

Classes: 4SAE

Documents autorisés : OUI

Calculatrice autorisée : OUI

NON

Nombre de pages: 04

NON

Internet autorisée : OUI

NON

Date: 13/01/2022

Heure: 09h00

Durée : 1h30

де

Code

Note

Nom et Signature du

Surveillant

Nom et Signature du

Correcteur

Observations

/20

NB: Les réponses doivent être reportées uniquement sur la feuille réponse

Exercice 1: **OCU (6 points)**

1. **Que signifie la caractéristique vélocité du Big Data ?**

A. La rapidité avec laquelle les données sont générées et traitées **B.** La rapidité avec laquelle les données sont collectées et traitées

C. La rapidité avec laquelle les données sont traitées

D. A et B

2. Hadoop est un :

A. Logiciel open source qui permet le stockage et le traitement de données volumineuses

B. Framework open source qui permet le stockage et le traitement de données volumineuses

C. Système de gestion de fichiers qui permet le stockage et le traitement de données volumineuses

D. Système d'exploitation open source qui permet le stockage et le traitement de données volumineuses **3. La haute disponibilité est assurée dans :**

A. Hadoop vl via le StandBy Name Node **B.** Hadoop v2 via le Secondary Name Node **C.** Hadoop v2 via le StandBy Name Node **D.** Hadoop vl via le Secondary Name Node **4. Le chemin par défaut sous HDFS est :**

A. /home/cloudera

B. /root/cloudera

C. user/cloudera

D. /user/cloudera

NE RIEN ECRIRE

5. **Quelle est la proposition incorrecte à propos d'Apache Hive?**

A. Hive est une solution Data Warehouse intégrée dans Hadoop

B. Hive est une solution adéquate pour les données non structurées

C. Hive utilise comme modèle de base de données le système de gestion de base de données relationnelle **D.** Hive est créé par Facebook

6. Qu'est ce qui se passe suite à la suppression d'une table Hive externe ?

A. Les fichiers de données seront aussi supprimés

B. Les fichiers de données seront gardés sous HDFS

C. Les blocs HDFS seront formatés

D. Aucune des réponses précédentes

7. **Parmi ces composants lequel n'est pas un composant HBASE ?**

A. RegionServer

B. Zookeeper

C. HbaseMaster

D. JDBC driver

8. **Dans HBASE, une table peut être :**

9.

A. Supprimée directement

B. Supprimée après une désactivation

C. Uniquement désactivée

D. Uniquement compressée

Pour exécuter le script 'drivers.pig' en mode Hadoop/MapReduce, on lance la commande :

A. pig -x mapreduce '/home/cloudera/Desktop/drivers.pig'

B. pig -x local '/home/cloudera/drivers.pig'

C. pig -x /user/cloudera/drivers.pig'

D. A et B

10. Quelle est la commande qui permet d'afficher le résultat d'une requête pig sur l'écran ?

A. Filter

B. Store

C. Dump

D. Bet C

2

Exercice 2: (3 points)

Compléter le schéma ci-dessous afin d'expliquer les étapes d'un programme MapReduce permettant de compter le nombre d'occurrences de chaque mot du fichier d'entrée.

Fichier entrée Découpage

Map

Tri

réduction Résultat

French English Italian
Spanish Spanish Italian
French Spanish English

Exercice 3: (5 points)

Soit un fichier d'entrée qui contient une liste des produits, le fichier ayant une taille de 350 Mo.

1. Dans HDFS2, ce fichier sera divisé en combien de blocs ? Quelle est la taille de chaque bloc ?

2. Le fichier '*produits.txt*' est sous l'emplacement '*/home/cloudera*. Quelle est la commande adéquate pour le copier sous le nouvel emplacement '*/User/Cloudera/monDossier*'?

3. Ecrire la requête Hive qui permet de créer la table '*produits*' sachant que le fichier *produits.txt*' a la structure suivante :

Product_id, Category, Price
112, vêtements, 100
113, chaussures, 200
114, accessoires, 300

4. Après avoir chargé la table *'produits* à partir du fichier *'produits.txt',* quelle solution peut-on adopter pour garder ce fichier sous HDFS après la suppression de la table?

5. Soient les deux fichiers 'produitsSuisse.txt' et 'produitsAllemagne.txt':
Quelle solution peut-on adopter pour stocker ces deux fichiers dans une table Hive?

produits Suisse.txt x

price pays

produits Allemagne.txt x

1 Product id category

2

3

4

A

N

M

1 vetements 100

2 accessoires 55

3 chaussures 85 allemagne allemagne allemagne

Exercice 4: (6 points)

produits Allemagne.txt 2 produits Suisse.txt x

1

2

3

4

Product_id category price pays

4 vetements 100

5 55

6 accessoires 85

chaussures

suisse

suisse

suisse

1. En Utilisant les commandes HBASE Shell, créer la table Produits avec le schéma suivant :

Famille de colonnes

description

classification

nom, prix type

2. Ajouter les données suivantes dans la table Produits :

{'prod1', 'gaufrier', '118.9', 'appareil électroménager'}

Colonnes

3. En se basant sur la capture suivante, faites les modifications nécessaires :

```
hbase (main):001:0> get 'Produits', 'prodl', {COLUMN=>'description: prix', VERSIONS =>
 2} COLUMN
  description: prix
  description: prix
 2 row(s) in 0.1000 seconds
```

4. Quel est le nombre de versions sauvegardées par défaut ?

5.

CELL

timestamp=1641338930314, value=99.9

timestamp=1641338326997, value=118.9

A. Ecrire la commande qui permet de supprimer l'ancienne version du prix.

B. La colonne 'prix' va-t-elle être supprimée ? Justifier votre réponse.

Bon courage