

Devoir de révision

1. Que représente le Big Data :

- A. Une caractéristique de Hadoop capable de traiter une énorme quantité de données en parallèle
- B. Un concept et une plateforme de technologies capables de gérer une énorme quantité de données structurées, semi-structurées et non structurées
- C. Une caractéristique des bases de données capable de convertir des données structurées préexistantes en données brutes non structurées
- D. Une caractéristique des bases de données capable de convertir des données structurées préexistantes en données brutes semi-structurées et non structurées

2. Parmi ces caractéristiques une seule ne correspond pas au Big Data. Laquelle ?

- A. Variété
- B. Volume
- C. Volatilité
- D. Vitesse

3. Hadoop est conçu pour :

- A. Le traitement des transactions
- B. L'accès aux données à faible latence
- C. Le traitement batch sur les fichiers volumineux
- D. Le calcul intensif avec peu de données

4. Le mécanisme qui permet d'éviter la perte de données sous HDFS est :
- A. La réplication
 - B. La partition
 - C. La scalabilité
 - D. YARN
5. Dans HDFS1, les DataNodes envoient périodiquement des :
- A. Fichiers logs
 - B. Heartbeats
 - C. Métadonnées
 - D. Fichiers JSON
6. La configuration du NameNode se fait via des :
- A. Fichiers JSON
 - B. Fichiers logs
 - C. Fichiers XML
 - D. Fichiers CSV
7. Comment changer le facteur de réplication par défaut pour les données stockées dans HDFS ?
- A. Changer la valeur du paramètre dfs.replication de hdfs-site.xml
 - B. Changer la valeur du paramètre dfs.replication.max de hdfs-site.xml
 - C. Changer la valeur du paramètre dfs.replication de core-site.xml
 - D. Changer la valeur du paramètre dfs.replication.max de core-site.xml
8. Laquelle des commandes suivantes va copier le fichier '/user/cloudera/Myfile.txt' dans le répertoire '/user/cloudera/hadoopfile' ?
- A. Hadoop cp Myfile.txt /user/hadoop/hadoopfile
 - B. Hadoop fs -cp Myfile.txt /user/cloudera/hadoopfile
 - C. Hadoop fs -put Myfile.txt /user/hadoop/hadoopfile
 - D. Hadoop fs put Myfile.txt /user/cloudera/hadoopfile
9. L'un des composants de base de Hadoop est :
- A. Mapreduce
 - B. Hbase
 - C. Apache
 - D. Jaql
10. Lequel de ces éléments est responsable de la réplication des données dans Hadoop ?
- A. TaskTracker
 - B. JobTracker
 - C. NameNode
 - D. DataNode
11. Quelles sont les améliorations apportées par YARN par rapport au paradigme MapReduce1 ?

- A. Il est complètement open source
- B. Il sépare la gestion des ressources du traitement des données
- C. Il n'est plus en charge de la gestion des ressources
- D. Il n'est plus en charge du traitement des données

12. Soit la requête HIVE :

```
hive> LOAD DATA LOCAL INPATH '/exam.txt' OVERWRITE INTO TABLE exam;
```

Le terme OVERWRITE signifie que :

- A. Les données existantes dans la table exam seront écrasées lors de l'insertion
- B. Les données insérées seront concaténées aux données existantes dans la table
- C. Une erreur se produira lors de l'insertion
- D. Aucune des réponses

13. Quelle est la solution pour garder un fichier sous HDFS après la suppression d'une table HIVE :

- A. Table externe
- B. Table partitionnée
- C. Table de jointure
- D. Table interne

14. Dans une table partitionnée HIVE, les partitions sont stockées dans :

- A. Un seul répertoire
- B. Des répertoires distincts
- C. Deux répertoires au plus
- D. Deux répertoires au moins

15. Pour créer une table HBase :

- A. Il faut spécifier la liste des colonnes
- B. Il faut spécifier la liste des familles de colonnes
- C. Il faut spécifier les noms des Hfiles
- D. Il faut spécifier les noms des régions

16. Apache HBase est :

- A. Un système de gestion de base de données en colonne qui fonctionne au-dessus de HDFS et qui remplit les propriétés cohérence (consistency) et disponibilité (availability) du théorème CAP
- B. Un système de gestion de base de données en colonne qui fonctionne au-dessus de HDFS et qui remplit les propriétés cohérence (consistency) et partitionnement du théorème CAP
- C. Un système de gestion de base de données en document qui fonctionne au-dessus de HDFS et qui remplit les propriétés cohérence (consistency) et disponibilité (availability) du théorème CAP
- D. Un système de gestion de base de données en document qui fonctionne au-dessus de HDFS et qui remplit les propriétés disponibilité (availability) et partitionnement du théorème CAP

17. Dans HBase, scan permet de

- A.** Afficher une seule ligne
- B.** Afficher une liste de lignes
- C.** Afficher les premières lignes
- D.** Afficher les dernières lignes

18. Pour supprimer une table HBase, on doit :

- A.** Vider la table
- B.** Désactiver la table
- C.** Supprimer les familles de colonnes
- D.** Désactiver les familles de colonnes