

QCM NOSQL ET BIG DATA

1. Qu'est-ce que NoSQL ?

- A. Une base de données relationnelle
- B. Une base de données orientée objet
- C. Une base de données non structurée
- D. Une base de données basée sur XML

Réponse : C

2. Quelle est la principale caractéristique de MongoDB ?

- A. Modèle de données en colonnes
- B. Modèle de données en graphes
- C. Modèle de données clé-valeur
- D. Modèle de données document

Réponse : D

3. Quel langage est couramment utilisé pour interagir avec MongoDB ?

- A. SQL
- B. JavaScript
- C. Python
- D. Java

Réponse : B

4. Qu'est-ce que HBase ?

- A. Un système de fichiers distribué
- B. Un moteur de recherche
- C. Un système de gestion de base de données relationnelles
- D. Un système de gestion de base de données orientées colonnes

Réponse : D

5. Quelle est la structure de stockage des données dans HBase ?

- A. Orientée document
- B. Orientée colonnes
- C. Orientée objets
- D. Orientée graphes

Réponse : B

6. Quel est le langage de requête utilisé pour interagir avec HBase ?

- A. SQL
- B. HQL (HBase Query Language)
- C. NoQL (No Query Language)
- D. JSON

Réponse : A

7. Qu'est-ce que Spark ?

- A. Un système de fichiers distribué
- B. Un moteur de traitement distribué
- C. Un système de gestion de base de données relationnelles
- D. Un framework de programmation front-end

Réponse : B

8. Dans Spark, quelle abstraction est utilisée pour représenter les ensembles de données distribuées ?

- A. RDD (Resilient Distributed Dataset)
- B. DataFrame
- C. Dataset
- D. SparkSQL

Réponse : A

9. Quel langage de programmation est principalement utilisé avec Spark ?

- A. Java
- B. Python
- C. C++
- D. Ruby

Réponse : B

10. Quelle est la principale différence entre une base de données relationnelle et une base de données NoSQL ?

- A. La structure des données
- B. Les performances
- C. La sécurité
- D. La capacité de montée en charge

Réponse : A

11. Quel est le principal avantage de MongoDB par rapport aux bases de données relationnelles ?

- A. La flexibilité du schéma
- B. Les performances de lecture
- C. La sécurité des données
- D. La prise en charge des transactions ACID

Réponse : A

12. Quel est le principal avantage de HBase par rapport aux bases de données relationnelles ?

- A. La prise en charge des requêtes complexes
- B. La capacité de stocker de grands volumes de données
- C. La sécurité des données
- D. La facilité de mise en place et de maintenance

Réponse : B

13. Quelle est la structure de données principale utilisée par HBase pour stocker les données ?

- A. Tableaux
- B. Documents
- C. Graphes
- D. Colonnes

Réponse : D

14. Quelle est la principale fonctionnalité de Spark Streaming ?

- A. L'analyse en temps réel des flux de données
- B. L'optimisation des requêtes
- C. L'intégration avec les bases de données NoSQL
- D. Le traitement par lots de données

Réponse : A

15. Qu'est-ce que l'optimisation de requêtes dans Spark ?

- A. L'amélioration des performances des requêtes SQL
- B. L'utilisation de l'indexation pour accélérer les recherches
- C. L'optimisation des opérations de transformation et d'action sur les ensembles de données
- D. L'amélioration de la tolérance aux pannes dans un cluster Spark

Réponse : C

16. Quel est le principal avantage de l'utilisation de NoSQL dans un environnement Big Data ?

- A. La flexibilité du schéma
- B. La facilité de requêtage
- C. La conformité ACID
- D. La compatibilité avec les outils de BI traditionnels

Réponse : A

17. Qu'est-ce que la réplication dans MongoDB ?

- A. La copie des données sur plusieurs serveurs
- B. L'indexation des données pour une recherche rapide
- C. La division des données en plusieurs fragments
- D. L'optimisation des requêtes pour une meilleure performance

Réponse : A

18. Quelle est la méthode recommandée pour effectuer des opérations de lecture/écriture dans HBase ?

- A. L'utilisation de transactions ACID
- B. L'utilisation de langage de requête spécifique (HQL)
- C. L'utilisation de l'API Java HBase
- D. L'utilisation de l'API REST de HBase

Réponse : C

19. Qu'est-ce que Spark MLlib ?

- A. Un moteur de recherche basé sur Spark
- B. Une bibliothèque de machine learning intégrée à Spark
- C. Un langage de requête pour interagir avec Spark
- D. Un outil de visualisation de données

Réponse : B

20. Qu'est-ce que la partitionnement des données dans MongoDB ?

- A. La division des données en fragments pour une meilleure performance
- B. La copie des données sur plusieurs serveurs pour la redondance
- C. La création d'index pour une recherche rapide
- D. L'optimisation des requêtes pour une meilleure performance

Réponse : A

21. Quel est le langage de requête utilisé dans SparkSQL ?

- A. SQL
- B. HQL
- C. Python
- D. R

Réponse : A

22. Qu'est-ce que la cohérence des données dans HBase ?

- A. La garantie que les données sont toujours disponibles
- B. La garantie que les données sont répliquées sur plusieurs serveurs
- C. La garantie que les données sont cohérentes à un instant précis
- D. La garantie que les données sont stockées en utilisant le même schéma

Réponse : C

23. Qu'est-ce que la transformation de données dans Spark ?
- A. La modification des données stockées dans une base de données
 - B. L'extraction, la transformation et le chargement de données
 - C. La création de requêtes SQL complexes
 - D. L'optimisation des requêtes pour une meilleure performance

Réponse : B

24. Quel est le principal inconvénient des bases de données NoSQL ?
- A. La complexité de la modélisation des données
 - B. Les performances inférieures par rapport aux bases de données relationnelles
 - C. La faible sécurité des données
 - D. Le manque de support pour les transactions ACID

Réponse : D

25. Qu'est-ce que le sharding dans MongoDB ?
- A. La division des données en fragments pour une meilleure performance
 - B. La copie des données sur plusieurs serveurs pour la redondance
 - C. La création d'index pour une recherche rapide
 - D. L'optimisation des requêtes pour une meilleure performance

Réponse : A

26. Quelle est la différence entre les données stockées en colonnes et les données stockées en documents ?
- A. La structure de stockage des données
 - B. Les performances de lecture
 - C. La flexibilité du schéma
 - D. La capacité de requêtage

Réponse : A

27. Quel est le principal avantage de l'utilisation de HBase dans un environnement Big Data ?

- A. La flexibilité du schéma
- B. La facilité de requêtage
- C. La conformité ACID
- D. La capacité de stocker de grands volumes de données

Réponse : D

28. Qu'est-ce que la distribution des données dans Spark ?

- A. La copie des données sur plusieurs serveurs
- B. La division des données en fragments
- C. La création d'index pour une recherche rapide
- D. L'optimisation des requêtes pour une meilleure performance

Réponse : B

29. Qu'est-ce que la tolérance aux pannes dans HBase ?

- A. La garantie que les données sont toujours disponibles
- B. La garantie que les données sont répliquées sur plusieurs serveurs
- C. La garantie que les données sont cohérentes à un instant précis
- D. La garantie que les données sont stockées en utilisant le même schéma

Réponse : A

30. Quelle est la fonction principale de Spark Streaming ?

- A. L'analyse en temps réel des flux de données
- B. L'optimisation des requêtes
- C. La création de visualisations de données
- D. Le traitement par lots de données

Réponse : A