

1. Parmi les sources de données du big data, on trouve :

- a) Données structurées
- b) Données non structurées
- c) Données semi-structurées
- d) Données locales

Réponse correcte : a), b), c)

2. L'extraction de données quand elle est complète :

- a) Inclut toutes les données disponibles
- b) Sélectionne uniquement les données pertinentes
- c) Est généralement plus rapide
- d) Ne nécessite pas de transformation des données

Réponse correcte : a)

3. Il faut considérer au moment du chargement (le Load dans ETL) :

- a) La vitesse de la lumière
- b) Les coûts de stockage
- c) La phase d'extraction
- d) La phase de transformation

Réponse correcte : b)

4. Le théorème CAP est utilisé pour :

- a) Classer les types de café
- b) Comprendre les bases de données relationnelles
- c) Évaluer la cohérence, la disponibilité et la tolérance aux partitions dans les systèmes distribués
- d) Calculer la capacité d'un serveur

Réponse correcte : c)

5. Dans les systèmes de type CA :

- a) La cohérence est priorisée sur la disponibilité
- b) La disponibilité est priorisée sur la cohérence

- c) La tolérance aux partitions est sacrifiée
- d) Les systèmes ne fonctionnent que localement

Réponse correcte : a)

6. Dans les systèmes de type AP :

- a) La cohérence est priorisée sur la disponibilité
- b) La disponibilité est priorisée sur la cohérence
- c) La tolérance aux partitions est sacrifiée
- d) Les systèmes ne fonctionnent que localement

Réponse correcte : b)

7. Dans les systèmes de type CP :

- a) La cohérence est priorisée sur la disponibilité
- b) La disponibilité est priorisée sur la cohérence
- c) La tolérance aux partitions est sacrifiée
- d) Les systèmes ne fonctionnent que localement

Réponse correcte : a)

8. On utilise l'architecture LAMDA pour :

- a) Stocker des données dans un data lake
- b) Réaliser des opérations de traitement en batch et en temps réel
- c) Concevoir des bases de données relationnelles
- d) Créer des réseaux de neurones artificiels

Réponse correcte : b)

9. On utilise l'architecture KAPA pour :

- a) Écrire des documents en langage Kappa
- b) Construire des data warehouses
- c) Gérer des bases de données relationnelles
- d) Réaliser des calculs complexes en temps réel

Réponse correcte : d)

10. On utilise l'architecture DATA LAKE pour :

- a) Stocker uniquement des données structurées
- b) Stocker des données dans un entrepôt de données
- c) Stocker divers types de données brutes sans structure prédéfinie
- d) Créer un data warehouse

Réponse correcte : c)

11. Le choix du type de base de données dans les systèmes CA dépend de :

- a) La couleur préférée de l'administrateur système
- b) Les besoins en termes de cohérence et de disponibilité du système
- c) Le nombre de partitions du réseau
- d) Le coût de la licence du SGBD

Réponse correcte : b)

12. On se base pour le choix d'une base NOSQL sur :

- a) Le nombre de voyelles dans le nom du système
- b) Les besoins spécifiques en termes de données et de performances
- c) La taille du serveur de base de données
- d) La popularité du système dans les médias sociaux

Réponse correcte : b)

13. Les propriétés ACID et BASE garantissent :

- a) La fraîcheur des données
- b) La disponibilité des données
- c) La cohérence des données
- d) La tolérance aux partitions des données

Réponse correcte : c)

14. Les propriétés ACID garantissent :

- a) L'atomicité, la cohérence, l'isolation et la durabilité des transactions

- b) L'accès à Internet
- c) L'agilité des systèmes
- d) La vitesse de traitement des données

Réponse correcte : a)

15. Les propriétés BASE sont utilisées :

- a) Pour maintenir une base de données relationnelle
- b) Pour assurer la cohérence et l'isolation des données
- c) Pour garantir la disponibilité des données malgré les partitions
- d) Pour accélérer les transactions

Réponse correcte : c)

16. Les types de traitement qu'offre le noyau de Hadoop sont :

- a) Friture et cuisson
- b) Batch et MapReduce
- c) Streaming et réduction
- d) Triage et filtrage

Réponse correcte : b)

17. Le master DataSet est utilisé pour :

- a) Former des ensembles de données intelligents
- b) Gérer les ensembles de données principaux dans un système distribué
- c) Diviser les données en ensembles de test et d'entraînement
- d) Organiser les données en arborescence

Réponse correcte : b)

18. Le master DataSet peut être simulé à :

- a) L'aide d'un logiciel de modélisation 3D
- b) L'aide d'un ordinateur quantique
- c) Des fins de test ou de développement
- d) L'aide de marionnettes

Réponse correcte : c)

19. Dans le big data, les systèmes de fichiers :

- a) Sont obsolètes
- b) Stockent uniquement des données structurées
- c) Gèrent le stockage de grandes quantités de données
- d) Sont utilisés exclusivement pour la programmation en C++

Réponse correcte : c)

20. Dans HDFS, le nombre de blocs écrits :

- a) Dépend du nombre de partitions dans le système
- b) Est fixe et égal à 64 Mo par défaut
- c) Est calculé en fonction de la taille du disque dur
- d) Est illimité

Réponse correcte : b)

21. Le secondary NameNode est installé :

- a) Comme un miroir du NameNode principal
- b) Pour gérer les données de sauvegarde
- c) Pour stocker les fichiers de données
- d) En option, il n'est pas obligatoire

Réponse correcte : b)

22) Le théorème CAP définit la :

- a) Capacité des serveurs
- b) Cohérence, la disponibilité et la tolérance aux partitions dans les systèmes distribués
- c) Portée des protocoles de communication
- d) Configuration des réseaux informatiques

Réponse correcte : b)

23) Le théorème CAP démontre que :

- a) Tout système distribué est parfaitement cohérent
- b) Tout système distribué peut être parfaitement disponible en tout temps
- c) Dans un système distribué, on ne peut garantir simultanément la cohérence, la disponibilité et la tolérance aux partitions
- d) Les systèmes distribués ne sont pas utiles

Réponse correcte : c)

24) Le théorème CAP propose :

- a) De changer le nom du théorème
- b) Un moyen de résoudre tous les problèmes de cohérence et d'indisponibilité dans les systèmes distribués
- c) Une échelle de notation pour les bases de données
- d) De réduire le nombre de serveurs dans un cluster

Réponse correcte : b)

25) Un système AP garantit :

- a) Une disponibilité maximale
- b) Une cohérence stricte
- c) La tolérance aux partitions
- d) Un bon café

Réponse correcte : a)

26) La méthodologie de travail dans un projet big data :

- a) Nécessite des licornes
- b) Suit un processus itératif de collecte, nettoyage, analyse et visualisation des données
- c) Est principalement basée sur la chance
- d) Se résume à l'installation de logiciels de big data

Réponse correcte : b)

27) Le nettoyage de données sert à :

- a) Supprimer toutes les données
- b) Éliminer les données inutiles

- c) Transformer les données en or
- d) Garantir la cohérence des données en éliminant les erreurs et les incohérences

Réponse correcte : d)

28) En Data Science, on utilise le big data pour :

- a) Rendre les données plus petites
- b) Analyser de grandes quantités de données pour en extraire des informations et des modèles
- c) Jouer à des jeux vidéo
- d) Créer des bases de données relationnelles

Réponse correcte : b)

29) Les Data Warehouses sont utilisés pour :

- a) Stocker des données brutes sans aucune transformation
- b) Analyser des données en temps réel
- c) Stocker et gérer des données structurées et historiques pour l'analyse décisionnelle
- d) Fabriquer des entrepôts de données en bois

Réponse correcte : c)

30) Le résultat d'une opération Split est :

- a) La division d'un ensemble de données en deux parties
- b) La multiplication des données
- c) L'addition des données
- d) La suppression de données

Réponse correcte : a)

31) La commande « `hdfs dfs -rm -r` » sert à :

- a) Lister tous les fichiers dans un répertoire HDFS
- b) Supprimer récursivement un répertoire et son contenu dans HDFS
- c) Copier un fichier d'un cluster Hadoop à un autre
- d) Renommer un fichier dans HDFS

Réponse correcte : b)

32) Le SGBD HBASE est :

- a) Un système de gestion de base de données relationnelle
- b) Un système de gestion de base de données NoSQL orienté colonne
- c) Un logiciel de traitement de texte
- d) Un logiciel de navigation web

Réponse correcte : b)

33) La commande scan dans HBASE sert à :

- a) Scanner des documents PDF
- b) Lire des codes-barres
- c) Rechercher des données dans une table HBASE
- d) Supprimer des fichiers dans HDFS

Réponse correcte : c)

34) L'exploration des données sert à :

- a) Voyager à travers l'univers virtuel
- b) Découvrir des informations et des tendances cachées dans les données
- c) Creuser des trous dans la terre
- d) Écrire des rapports d'entreprise

Réponse correcte : b)

35) Les types de données qu'on peut stocker dans BASE :

- a) Des données structurées uniquement
- b) Des données non structurées uniquement
- c) Des données semi-structurées
- d) Des données en alphabet

Réponse correcte : c)

36) Parmi les types de transformation dans ETL :

- a) Extraction, Triage, Langage



- b) Évolution, Traduction, Lecture
- c) Extraction, Transformation, Chargement
- d) Étape, Test, Lecture

Réponse correcte : c)

37) Parmi les types de chargement dans ETL :

- a) Chargement en douceur
- b) Chargement en sautant
- c) Chargement incrémental
- d) Chargement spatial

Réponse correcte : c)

38) Les opérations qu'on peut effectuer sur une DB NoSQL clé-valeur :

- a) Lire, écrire, mettre à jour, supprimer
- b) Ajouter, soustraire, multiplier, diviser
- c) Coudre, tricoter, crocheter, tisser
- d) Danser, chanter, cuisiner, jouer aux échecs

Réponse correcte : a)

39) Parmi les SGBD NoSQL orientés document, on trouve :

- a) MongoDB, Couchbase, Cassandra
- b) MySQL, PostgreSQL, Oracle
- c) Excel, Word, PowerPoint
- d) Hadoop, Spark, Hive

Réponse correcte : a)

40) Parmi les moteurs de traitement micro-batch et temps réel, on trouve :

- a) Apache Kafka, Apache Flink, Apache Spark Streaming
- b) Microsoft Word, Adobe Photoshop, Excel
- c) Twitter, Facebook, Instagram
- d) La météo, la politique, la cuisine

Réponse correcte : a)

J'espère que ces réponses vous sont utiles !