

Commencé le mardi 9 mars 2021, 14:12

État Terminé

Terminé le mardi 9 mars 2021, 14:42

Temps mis 29 min 58 s

Note 19,00 sur 20,00 (95%)

Question **1**

Terminer

Note de 5,00 sur 5,00

Nommez la fonctionnalité qui rend Spark plus rapide que Hadoop. Expliquez pourquoi.

La fonctionnalité qui rend Spark plus rapide que Hadoop sont les RDDs. Ici, on parle du Resilient Distributed Dataset. En effet, Spark fonctionne en mémoire. En d'autres termes, Spark est efficace lorsqu'on parle de traitement rapide des données ainsi qu'un processus itératif. D'ailleurs, en mémoire, Spark réalise des tâches en mémoire 100 fois plus rapidement que Hadoop et 10 fois plus rapidement lorsque c'est sur disque. Effectivement, la cause est qu'on utilise moins d'écritures et de lectures par cycle.

Commentaire :

Question 2

Terminer

Note de 5,00 sur 5,00

Dans vos propres mots, définissez des bases de données de wide-column. Nommez deux de leurs avantages.

Les bases de données wide-column sont des bases de données de type NoSQL. Au niveau conceptuel, les wide-columns se rapprochent des bases de données relationnelles. Cependant, la notion de colonnes est assez flexible. Effectivement, il est possible de regrouper des attributs dans des familles de colonnes. En quelques sortes, on pourrait considérer que les wide-columns sont des bases de données de type key-value en 2 dimensions. Il faut savoir qu'ils possèdent des colonnes et des rangées.

Au niveau des avantages, on pourrait penser au fait qu'ils peuvent supporter de grandes quantités de données. Aussi, la flexibilité des familles colonnes améliore la maintenabilité des données et la gestion des données variantes.

Commentaire :

Question 3

Terminer

Note de 9,00 sur 10,00

Vous développez un système de prévision météorologique. Le système reçoit les données météorologiques toutes les 30 minutes, il reformate les données et les stocke dans une base de données. Après une journée, le système utilise les données du jour pour mettre à jour un modèle de prédiction entraîné et stocke la prédiction mise à jour dans une base de données. Quel type d'architecture de traitement (Lambda ou Kappa) utiliseriez-vous et pourquoi? Décrivez l'architecture du système (vous pouvez fournir un diagramme).

L'architecture de traitement que je choiserais serait Lambda. En effet, on mentionne que les données sont reçues à chaque 30 minutes et elles sont traitées. Alors, l'architecture Lambda permet de traiter ces informations de façon efficace. Il faut savoir que ce n'est pas du vrai temps réel même si on reçoit des données à chaque 30 minutes. On doit avoir une architecture qui supporte bien le temps réel, mais pas que. Il faut des réponses rapides et il faut un système qui permet de supporter plusieurs mises à jour de façon efficace. Voici une architecture :

Commentaire :

On manque l'architecture mais la réponse est bonne.

Aller à...