

## LOG8430 - H2021 - Rappel- Mégadonnées

1. Nous développons une application pour une société commerciale. La société gère les contrats et les factures des transactions commerciales. Les données doivent être stockées de manière sécurisée et persistante pendant une longue période. L'entreprise fonctionne au niveau mondial avec un grand nombre de transactions. Quel modèle de données choisiriez-vous et pourquoi?

- ☐ Key-value, car les factures doivent être récupérées par leur clé
- ☐ Document, car les factures sont des fichiers et peuvent être recherchées par contenu
- ☐ Column-wide, car les factures peuvent avoir une structure variable.

2. Wikipedia est une encyclopédie en ligne où chaque entrée est une page et peut contenir des liens vers d'autres entrées. Wikipédia organise également ses entrées en catégories et sous-catégories pour faciliter la recherche. Quel modèle de données choisiriez-vous pour implémenter ce système et pourquoi?

- ☐ Key-value, car la structure de chaque page peut être différente.

- ☐ Document, car chaque page est une entrée et nous pouvons rechercher par son contenu.
- ☐ Graph, car les pages sont connectées les unes aux autres par des relations complexes.
- ☐ Wide-column, car c'est le modèle le plus proche d'une base de données relationnelle.

3. Nous avons un système de surveillance des ressources cloud. À des intervalles assez fréquents et rapides, le système envoie des mesures (CPU, mémoire, disque, réseau) pour chaque ressource (machine virtuelle). Quel modèle de données choisiriez-vous et pourquoi?

- ☐ Document, les mesures pour chaque ressource sont stockées dans un fichier.
- ☐ Key-value, les données sont petites et indexées par ressource
- ☐ Wide-column, pour tenir compte des valeurs manquantes

4. Les bases de données pour les données biologiques existent depuis un certain temps déjà. Ils contiennent des données sur les gènes, les protéines, les organismes. Les entités ont des attributs, mais il est possible de découvrir de nouveaux attributs dans le futur. Diverses analyses et outils existent déjà pour nous aider à étudier le monde naturel. Quel modèle de données utiliseriez-vous et pourquoi?

- ☐ Base de données relationnelle, chaque type est une table et les relations sont modélisées avec des clés étrangères
- ☐ Graph, pour capturer les relations complexes entre les entités
- ☐ Wide-column, pour tenir compte des outils existants et des nouveaux attributs.

5. Nous voulons développer un programme pour compter les mots de milliers de documents organisés dans des dossiers avec la première lettre du nom du fichier («A», «B», «C»...). Décrire l'implémentation de MapReduce pour ce problème (nombre de travailleurs, cartographie des tâches, réduction des tâches)

- ☐ Un travailleur par lettre, mappez le nombre de tâches par document, réduisez les agrégats de tâches pour tous les documents
- ☐ Un travailleur par lettre, mappez le nombre de tâches par lettre, réduisez le nombre de tâches pour toutes les lettres

- ☐ Affectez les travailleurs en fonction de la distribution des mots pour les lettres anglaises, mappez le nombre de tâches par travailleur, réduisez les agrégats de tâches pour tous les travailleurs.

6. Nous avons la base de données de l'ARC pour l'impôt sur le revenu. L'ARC a plusieurs serveurs de données à travers le pays. Nous voulons savoir qui a payé le plus d'impôts en 2018. Décrivez l'implémentation de MapReduce pour ce problème (nombre de travailleurs, tâches map et reduce).

- ☐ Un travailleur par province, la map trouve le maximum par province, la reduce le maximum de trouvailles de toutes les provinces
- ☐ Un travailleur par province, la map trie les déclarations de revenus par province, la reduce le tri des déclarations de revenus pour toutes les provinces
- ☐ Répartissez les travailleurs en fonction de la population, la map trouve le maximum de trouvailles par travailleur, la reduce trouve le maximum de trouvailles parmi tous les travailleurs.

7. Nous voulons trouver le joueur avec le plus de buts moyens par match dans toute l'histoire de la LNH. Chaque équipe conserve ses propres statistiques. (Ne considérez pas les équipes qui n'existent pas actuellement). Décrivez l'implémentation de MapReduce pour ce problème (nombre de travailleurs, tâches map et reduce).

- ☐ Un travailleur par équipe, le map trouve les moyennes par équipe, le reduce le nombre maximum de trouvailles pour tous les joueurs
- ☐ Un travailleur par équipe, le map trouve la moyenne maximale par équipe, le reduce le nombre maximal de trouvailles pour tous les joueurs
- ☐ Attribuez des ouvriers en fonction de l'âge d'une équipe, le map trouve la moyenne maximale par équipe, le reduce le nombre maximal de trouvailles pour tous les joueurs.

8. Chaque université détient toutes les thèses soutenues dans leur bibliothèque respective. Nous souhaitons rechercher dans le monde entier des thèses contenant le terme «DevOps» dans leur titre. Décrivez l'implémentation de MapReduce pour ce problème (nombre de travailleurs, tâches map et reduce).

- ☐ Un travailleur par bibliothèque universitaire, la map trouve tous les titres DevOps, la reduce les résultats agrégés dans une liste
- ☐ Un travailleur par université, la map trouve tous les titres DevOps, la reduce les résultats agrégés dans une liste
- ☐ Un travailleur par université, la map trouve tous les titres DevOps, la reduce trouve la thèse avec la plupart des citations.

9. Nous voulons une liste complète de toutes les villes, villages et communautés de l'ordre mondial

par population par ordre décroissant. Les serveurs de recensement de chaque pays contiennent des données démographiques. Décrivez l'implémentation de MapReduce pour ce problème (nombre de travailleurs, tâches map et reduce).

- ☐ Un travailleur par pays, la map trie les villes par pays, la reduce les agrégats de toutes les villes dans une liste triée
- ☐ Un travailleur par serveur de recensement, la map trie les villes par serveur de recensement, la reduce les agrégats de toutes les villes dans une liste triée

Terminé

Optimisé par



SurveyMonkey®

[Créez un sondage](#) en quelques clics !

[Avis de confidentialité](#) et [Avis relatif aux cookies](#)