MapReduce

MapReduce for $200

- 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»…). Vous avez 27 travailleurs. Décrivez l'implémentation MapReduce pour ce problème.

MapReduce for $200

- 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»…). Vous avez 27 travailleurs. Décrivez l'implémentation MapReduce pour ce problème.

1) Envoyez un dossier à chaque travailleur.
    1) Faites attention! Tous les dossiers ne contiennent pas la même quantité de fichiers! Vous pouvez attribuer une taille appropriée aux travailleurs ou utiliser moins de travailleurs et envoyer certaines des lettres les moins populaires à un seul travailleur.
2) Les tâches map comptent les mots des fichiers par fichier.
3) Les tâches reduce somment les résultats par dossier et pour tout le système de fichiers.
    1) Vous pouvez avoir plusieurs niveaux de tâches reduce et l'entrée d'une tâche reduce peut être la sortie d'une tâche reduce précédente.

MapReduce for $400

- 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 2019. Décrivez l'implémentation de MapReduce pour ce problème.

MapReduce for $400

- 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 2019. Décrivez l'implémentation de MapReduce pour ce problème.

1. Chaque serveur de données peut exécuter une réplique de l'algorithme maximal.
2. Les tâches map consistent à trouver le maximum dans un seul serveur.
3. Les tâches reduce consistent à trouver le maximum par province et une autre tâche reduce pour trouver le maximum global du pays.

MapReduce for $600

- Nous voulons trouver le joueur avec la plus grande moyenne de buts par match dans toute l'histoire de la NHL. Chaque équipe conserve ses propres statistiques. (Ne considérez pas les équipes qui n'existent pas actuellement). Décrivez l'implémentation MapReduce pour ce problème.

MapReduce for $600

- Nous voulons trouver le joueur avec la plus grande moyenne de buts par match dans toute l'histoire de la NHL. Chaque équipe conserve ses propres statistiques. (Ne considérez pas les équipes qui n'existent pas actuellement). Décrivez l'implémentation MapReduce pour ce problème.

1. Le serveur de données de chaque équipe exécutera une réplique de l'algorithme.
2. Les tâches map consistent à trouver la moyenne par équipe.
3. Les tâches reduce trouveront le maximum sur les moyennes renvoyées.
4. Et la réponse est… Mike Bossy (0,762) des New York Islanders (1977-1987).

MapReduce for $800

- 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 MapReduce pour ce problème.

MapReduce for $800

- 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 MapReduce pour ce problème.

1. Le serveur de données de chaque université exécutera une réplique de l'algorithme grep / search.
2. Les tâches map sont les recherches avec le string «DevOps» sur le titre.
3. Les tâches reduce regrouperont simplement les résultats individuels des tâches map dans une liste.

MapReduce for $1000

- Nous voulons une liste complète de toutes les villes, villages et communautés du monde classés par population par ordre décroissant. Les serveurs de recensement de chaque pays contiennent des données démographiques. Décrivez l'implémentation MapReduce pour ce problème.

MapReduce for $1000

- Nous voulons une liste complète de toutes les villes, villages et communautés du monde classés par population par ordre décroissant. Les serveurs de recensement de chaque pays contiennent des données démographiques. Décrivez l'implémentation MapReduce pour ce problème.

1. Chaque serveur de recensement peut exécuter une réplique de l'algorithme de tri.
2. Les tâches map stockeront les villes par pays.
3. La tâche reduce agrégera la liste totale par ordre décroissant.


DAILY
DOUBLE

MapReduce for $2000

- Nous voulons une liste complète de toutes les villes, villages et communautés du monde classés par population par ordre décroissant. Les serveurs de recensement de chaque pays contiennent des données démographiques. Décrivez l'implémentation MapReduce pour ce problème.
- Question bonus: quel est l'algorithme de tri le plus adapté à ce cas?

MapReduce for $2000

- Nous voulons une liste complète de toutes les villes, villages et communautés du monde classés par population par ordre décroissant. Les serveurs de recensement de chaque pays contiennent des données démographiques. Décrivez l'implémentation MapReduce pour ce problème.
- Question bonus: quel est l'algorithme de tri le plus adapté à ce cas?
- “What is…” mergesort
- En fait, les RDD (ou formats de données similaires) sont des paires clé-valeur, ce qui signifie que le tri peut être très efficace et simple sur les clés.
- Si nous voulons trier par valeur, une solution est avec des «clés composées», où nous pouvons combiner la clé primaire avec la colonne que nous voulons trier, puis trier automatiquement par clé.