

Projet Final (NoSql)

Cassandra VS MongoDB

Cassandra et MongoDB sont deux bases de données NoSQL populaires connues pour leur évolutivité, leur haute disponibilité et leur tolérance aux pannes. Alors que Cassandra est une base de données à larges colonnes développée par Apache Software Foundation, MongoDB est une base de données orientée document développée par MongoDB Inc.

Cassandra

Cassandra est un système de gestion de base de données NoSQL , distribué, à larges colonnes et gratuit et open source . Il a été développé par Apache Software Foundation et initialement publié en juillet 2008.

Cassandra est conçu pour gérer de grandes quantités de données sur de nombreux serveurs de base, offrant une haute disponibilité sans point de défaillance unique.

Caractéristiques

Cassandra est une base de données NoSQL distribuée connue pour son évolutivité, sa haute disponibilité et sa tolérance aux pannes. Certaines de ses principales caractéristiques sont les suivantes :

- 1- **Architecture distribuée** : Cassandra est conçue pour fonctionner sur plusieurs nœuds dans plusieurs centres de données, offrant une haute disponibilité et une évolutivité en distribuant les données sur le cluster.
- 2- **Évolutivité linéaire** : à mesure que de nouveaux nœuds sont ajoutés au cluster, Cassandra peut facilement évoluer pour gérer une charge accrue sans temps d'arrêt ni interruption de l'application.
- 3- **Haute disponibilité** : les données sont répliquées sur plusieurs nœuds du cluster, ce qui garantit qu'en cas de défaillance d'un nœud, les données sont toujours accessibles à partir d'autres nœuds, offrant ainsi une disponibilité continue .
- 4- **Tolérance aux pannes** : Cassandra est tolérant aux pannes, ce qui signifie qu'il peut résister aux pannes de nœuds et aux partitions réseau sans perdre de données ni de disponibilité.
- 5- **Sans schéma** : contrairement aux bases de données relationnelles traditionnelles , Cassandra est sans schéma, ce qui vous permet de stocker différents types de données dans la même table sans définir de schéma rigide.

Fonctionnalités de MongoDB

MongoDB est une base de données NoSQL très appréciée qui stocke les données sous forme de documents flexibles et évolutifs. Voici quelques-unes des fonctionnalités clés de MongoDB:

- 1- **Orienté document** : MongoDB stocke les données dans des documents flexibles de type JSON , ce qui facilite le travail avec les données de manière naturelle et intuitive.

- 2- **Sans schéma** : MongoDB ne nécessite pas de schéma prédéfini, ce qui vous permet de modifier facilement la structure de vos documents à mesure que votre application évolue.
- 3- **Hautement évolutif** : MongoDB est conçu pour évoluer horizontalement, vous permettant de faire facilement évoluer votre base de données sur plusieurs serveurs pour gérer des quantités croissantes de données.
- 4- **Hautement disponible** : MongoDB prend en charge les ensembles de réplicas, qui fournissent un basculement automatique et une redondance des données pour garantir une haute disponibilité de vos données.
- 5- **Langage de requête flexible** : MongoDB prend en charge un langage de requête riche qui nous permet d'effectuer des requêtes complexes sur vos données, y compris des requêtes qui s'étendent sur plusieurs documents et collections.
- 6- **Index** : MongoDB prend en charge les index pour améliorer les performances de vos requêtes, notamment les index à champ unique, composés et géospatiaux.

Différence entre Cassandra et MongoDB

- langage de requête

Cassandra :

Langage de requête : Cassandra utilise CQL (Cassandra Query Language) pour interroger les données. CQL est similaire à SQL en termes de syntaxe, mais est conçu spécifiquement pour le modèle de données de Cassandra.

Modèle de données : Cassandra est basé sur un modèle de stockage de données à colonnes larges, ce qui signifie que les données sont stockées dans des lignes et des colonnes. Il est optimisé pour les charges de travail nécessitant beaucoup d'écriture et est conçu pour être hautement évolutif et tolérant aux pannes.

MongoDB :

Langage de requête : MongoDB utilise un langage de requête basé sur des documents de type JSON. Les requêtes sont exprimées à l'aide d'une syntaxe riche et flexible qui permet des requêtes et des agrégations complexes .

Modèle de données : MongoDB utilise une approche orientée document, stockant les données sous forme de documents de type JSON. Il est conçu pour être flexible et évolutif, ce qui le rend adapté à un large éventail de cas d'utilisation.

- Tableau comparatif

Cassandra	MongoDB
Développé par la fondation Apache Software et publié en juillet 2008.	Développé par MongoDB Inc. et initialement publié le 11 février 2009.
Cassandra est écrit uniquement en langage Java.	MongoDB est écrit en langages C++, Go, JavaScript, Python.
L'évolutivité de l'écriture dans Cassandra est très élevée et efficace.	L'évolutivité de l'écriture est limitée dans MongoDB
Les performances de lecture sont très efficaces dans Cassandra car cela prend un temps $O(1)$.	Les performances de lecture ne sont pas aussi rapides dans MongoDB par rapport à Cassandra.
Cassandra n'offre qu'un support superficiel pour les index secondaires, c'est-à-dire que l'indexation secondaire est restreinte.	MongoDB prend en charge le concept d'index secondaires.
Cassandra prend uniquement en charge le format de données JSON.	MongoDB prend en charge les formats de données JSON et BSON.
La méthode de réplication prise en charge par Cassandra est le facteur de réplication sélectionnable.	La méthode de réplication prise en charge par MongoDB est la réplication maître-esclave
Cassandra ne fournit pas de transactions ACID mais peut être configuré pour prendre en charge les propriétés ACID.	MongoDB permet des transactions de documents multiples qui garantissent l'atomicité, la cohérence, l'isolement et la durabilité (ACID) avec l'isolement des instantanés.
Les systèmes d'exploitation serveur pour Cassandra sont BSD, Linux, OS X, Windows.	Les systèmes d'exploitation serveur pour MongoDB sont Solaris, Linux, OS X, Windows.
Des entreprises célèbres comme Hulu, Instagram, Intuit, Netflix, Reddit, etc. utilisent Cassandra.	Des entreprises célèbres comme Adobe, Amadeus, Lyft, ViaVarejo, Craftbase, etc. utilisent MongoDB.

Conclusion

Dans l'ensemble, le choix entre Cassandra et MongoDB dépend des exigences spécifiques de votre application. Si vous avez besoin d'une base de données hautement évolutive optimisée pour les charges de travail à forte intensité d'écriture, Cassandra peut être un meilleur choix. En revanche, si vous avez besoin d'une base de données flexible et sans schéma avec des opérations de lecture et d'écriture rapides, MongoDB pourrait être plus adapté. En fin de compte, il est essentiel d'évaluer les fonctionnalités et les capacités de chaque base de données dans le contexte des besoins de votre application pour prendre une décision éclairée.

Partie pratique :

- Installation mongo sous Mac OS
 - 1- brew tap mongodb/brew
 - 2- brew install [mongodb-community@4.4](#)
 - 3- brew service mongod start
- Installation Cassandra sous Mac OS
 - 1- brew install cassandra
 - 2- brew services start cassandra
- Configuration python
 - 1- pip install pymongo
 - 2- pip install cassandra-driver

Les différentes étapes ci-dessous sont détaillées dans le notebook :

- Création d'un dossier de Projet et un Environnement Virtuel
- Installation de Jupyter, pymongo et cassandra-driver
- Démarrer Jupyter Notebook
- Préparation de la bases de données
- Importer les bibliothèques nécessaires
- Connexion à MongoDB
- Connexion à Cassandra
- Créer un Keyspace et une Table
- Charger des données etudiant dans MongoDB
- Charger des données etudiant dans Cassandra
- Mesurer les performances d'écriture
- Mesurer les performances de lecture