



MongoDB - Introduction et principes généraux

Lucas Filippi
Sébastien Ferrandez

Une nouvelle façon de penser le modèle de données

« MongoDB is just one big, dumb data store »

Schema-less → Simplicité

L'objectif premier est aussi la performance ! On stocke et on extrait des données, le plus rapidement/efficacement possible !

Exige une remise en cause des acquis techniques

Les outils de Mongo

Un cœur écrit en C++

Des données au format BSON (*binary* JSON)

Un *shell* JavaScript pour la ligne de commande

Des *drivers* pour les différents langages

Rappel des principes ACID

- Atomicité : c'est tout...ou rien !
- Cohérence : on va d'un état cohérent à un autre
→ intégrité des données
- Isolation : la transaction est la seule à voir ce qu'elle modifie, les autres ne verront ça qu'une fois la transaction terminée
- Durabilité (toutes les opérations sont en log et donc facilitent la reprise sur incident)

NoSQL vs SQL...round 1, fight !

- Table, row et column deviennent *collection*, *document* et *field* dans la terminologie Mongo
- Les index : OUI
- Les jointures : NON
- Les transactions : NON

(dé)Normalisation

- Données embarquées = dénormalisation
- Données référencées = normalisation
- Aucune des 2 méthodes meilleure que l'autre
- Ex: une commande référence des produits avec des id = NORMALISATION
une commande embarque en son sein les produits = DENORMALISATION, fini les références induites par la présence d'ids

(dé)Normalisation

- Données embarquées : 1 seek sur le disque, 1 a/r vers la base.
- On évite les jointures en remontant un document avec ses sous-documents *embedded*
- Attention, limite max d'un document = 16MB (à ce jour)

(dé)Normalisation

- Comment choisir ? Tout dépend du type de l'appli...Plutôt de la lecture ou de l'écriture ?
- Dénormaliser **accélère** les lectures mais ralentit les écritures
- Normaliser garantit l'**intégrité**
- Le but ultime : une seule requête pour tout ! :)

Avantages

- La performance comme but ultime...
- ...tout comme la simplicité !
- Flexibilité (exemple des documents)
- Mise à l'échelle horizontale et verticale facilitée
- Des références sérieuses dans tous les domaines (e-commerce, news, ad servers, analytiques, réseaux sociaux etc.)

Inconvénients

- Technologie nouvelle...donc manque de maturité sur certains points
- Exige de revoir ses pratiques (jointures...)
- On considère la rapidité au détriment parfois de l'intégrité des données
- Pas de prise en charge des transactions