

Resumen MongoDB

Dillan Almendares Barrantes - 2020033336

Debido a que las bases de datos relacionales no estaban siendo capaces de servir para el análisis de sus datos por el ritmo con que aumentaban y que la estructura de estas bases de datos complicaba el uso de sistemas distribuidos. Para solventar esto se solía recurrir a bases de datos NoSQL pero a cambio de mayor flexibilidad se sacrificaban la integridad de las bases de datos SQL al igual que reducían el valor de los datos para ciertas aplicaciones valiosas.

En el 2009 se lanzó MongoDB creando una nueva clase de bases de datos de uso general. Esto lo logra manteniendo los mejores aspectos de las bases de datos relacionales y las NoSQL, en MongoDB en vez de tablas, se utilizan documentos más flexibles porque se mapean de la manera que los desarrolladores lo programen. En lugar de guardar los datos en columnas y filas, se usan JSON. Este tipo de base de datos de documentos permite almacenar cualquier tipo de datos y a los documentos se les puede cambiar la estructura fácilmente, permitiendo a desarrolladores programar más rápido sus aplicaciones. Es posible agregar campos sin alterar otros documentos y cada documento puede estructurarse de la manera que sea más conveniente para la aplicación, ya sea como pares Key-Value, como aristas y nodos de un grafo; o como estructuras anidadas que representen relaciones.

Las principales ventajas de este modelo basado en documentos son:

1. Es más intuitivo

Es más sencillo y rápido para programar con este modelo al poder mapear los documentos con los objetos del código, de esta manera no hay necesidad de descomponer los datos de una tabla o de ejecutar JOINS para obtener la información porque en los documentos se suele guardar cerca los datos que se acceden simultáneamente.

2. Esquema flexible

El esquema de los documentos es dinámico, no hace falta predefinirlo en la base de datos, así las nuevas funcionalidades de la aplicación se pueden añadir sin tener que hacer grandes migraciones al esquema. Para tener control de la estructura de los documentos, MongoDB ofrece **schema validation** para agregar reglas a los campos como los tipos de datos que pueden tener o rangos de valores, esto ahorra programar las validaciones en la capa de aplicación.

3. Universal

La flexibilidad en la estructura de los documentos junto al extendido uso de JSON amplía los tipos de aplicación donde puede servir este modelo de base de datos. El formato en el que se guardan los datos en MongoDB es **BSON** (Binary JSON) que es una representación binaria de JSON. Dicha codificación extiende la representación de JSON para incluir más tipos de datos como int, long, date, floating point y decimal128. Para las aplicaciones esto significa mayor facilidad para procesar, ordenar y comparar datos.

Las principales diferencias entre bases de datos son la expresión de su language de consultas, su capacidad de indexar y sus controles para la integridad de los datos. En el caso de MongoDB estas propiedades le corresponden a las siguientes secciones:

MongoDB Query API

Aporta maneras sencillas de acceder, agrupar, transformar y analizar datos. Puede unificar consultas entre bases de datos, soporta cargas de trabajo transaccionales y archivar datos en un data lake con la misma API.

MongoDB Aggregation Pipeline

Permite transformar y analizar datos mediante un pipeline multietapas donde pasan los documentos. En las etapas básicas del pipeline se aplican filtros que sirven como consulta que modifican la forma del documento. Otras operaciones de pipeline pueden agrupar y ordenar documentos por campos seleccionados, además se pueden hacer agregaciones de arrays.

ACID transactions

Se puede mantener el mismo nivel de aislamiento de las bases de datos relacionales por lo que no se pierde la capacidad de consistencia.

MongoDB usa replica sets y fragmentación para tener escalabilidad y buena disponibilidad.

Replica sets

Cada uno de los hasta 50 replica sets tiene una copia de los datos y pueden estar distribuidos entre nodos, datacenters o regiones. Si un nodo principal se llega a caer, se consigue un reemplazo del cluster rápidamente, elección que toma en cuenta parámetros como: time stamps, estado de conectividad o prioridad a replicas definidas por usuarios.

Privacidad y seguridad

Para proteger el acceso a los datos MongoDB tiene las siguientes funcionalidades:

- **Autenticación**
- **Autorización**
- **Auditorías:** Existe un log de auditoría donde se registra toda actividad y cambios.
- **Aislamiento de red:** La información de los usuarios de MongoDB Atlas está aislada de los otros usuarios.
- **Encriptación:** Los datos se pueden encriptar mientras pasan por la red, durante su uso en la base de datos o cuando no se están usando y permanecen en disco o en backups.