

Diseño de una Almacén de Datos NoSQL

- title
- author
- description
- rating
- ratingCount
- reviewCount
- genre
- book_link
- image

Diseñar un almacén de datos noSQL para la gestión de la información recogida sobre los libros que mejor se adapte a nuestras necesidades. Para este diseño debes de tener en cuenta las diferentes gestiones que se desean realizar con los datos, que quedan representadas en las siguientes consultas.

Consultas:

- Todos los libros de un género
- Datos de un libro
- Libros más populares o con mejor rating
- Libros de un autor concreto

Además, no únicamente hay que diseñar una base de datos que funcione, sino que además se busca un diseño que sea eficiente según las necesidades descritas anteriormente.

Debido a lo explicado, se pide realizar un informe en el que se detalle tanto la tecnología a utilizar y el por qué, y el esquema del propio almacén de datos.

Posible solución (resumido)

HBASE:

Orientado a columnas

Lecturas y escrituras consistentes

Distribución automática de la información

Recuperación automática ante fallos

Procesamiento en el servidor

Gestión integrada

HDFS

Primero, las imágenes se guardarán en otro tipo de base de datos y se consultarán utilizando URIs.

Claves secuenciales para leer más rápido. Perjudica en la escritura, pero es permisible dado que éstas se realizarán cada mucho tiempo

Tall-Thin con clave compuesta, dado que favorecen las lecturas del tipo 'Scan'.

Clave: Title-author. Favorece las dos búsquedas más importantes

Tabla: 3 familias con 3 accesos independientes

CF data (description, genre)

CF values (rating, ratingCount, reviewCount)

CF links (book_link, image)