

# Almacenamiento en Apache HBase



**Apache HBase** es una base de datos de código abierto y escalable, siendo un subproyecto de **Apache Hadoop**. Está modelada a partir de Google BigTable y está escrita en Java. Es una parte del ecosistema de Hadoop que proporciona una lectura/escritura real en tiempo real de los datos en el sistema de archivos Hadoop. Antes de todo esto, empezó como un proyecto por Powerset en 2007 para procesar masivas cantidades de datos para la búsqueda en lenguaje natural.

**Apache HBase** es un sistema de almacenamiento orientado a columnas, por tanto no sigue el esquema relacional y no admite SQL. Otra característica que tiene es que usa HDFS como su sistema de almacenamiento. Proporciona rápidas búsquedas para grandes tablas y también proporciona acceso de baja latencia a filas individuales de miles de millones de registros. Almacena los datos de cualquier tipo: estructurado, semiestructurado y no estructurado, sin ningún modelado inicial.

La estructura de la organización de los datos por parte de **Apache HBase** se hace por regiones, en donde una región es una unidad de distribución y disponibilidad. Estas regiones se dividen cuando crecen demasiado, en donde el tamaño de región máxima es un parámetro de ajuste.

Está diseñado para soportar consultas de conjuntos de datos masivos. **Apache HBase** está optimizado para rendimiento de lectura. Para las escrituras busca mantener la consistencia. En contraste con **Cassandra** que es eventualmente consistente, **Apache HBase** no ofrece varios ajustes de nivel de consistencia. Por tanto, las escrituras pueden ser lentas.

Un gran problema que tiene esta base de datos es el de no tener un lenguaje intuitivo para realizar las consultas, ya que utiliza un lenguaje pseudo-javascript para realizar tablas y crear consultas, cosa que en la mayoría de bases de datos es en Java.

Las principales compañías que utilizan **Apache Hadoop** son: Facebook, TrendMicro, Explorys, Tumblr, LinkedIn, Spotify y Netflix.

**Título:** Recuperación y acceso a la información - **Almacenamiento en Apache HBase**

**Autor:** Carlos Flores Martín.

Publicado en Abril, 2017