



NoSQL Base de Datos Escalable y
Distribuida

Presenta: Fernando Arevalo

Fecha: 26 de agosto del 2024

Introducción a Apache HBase

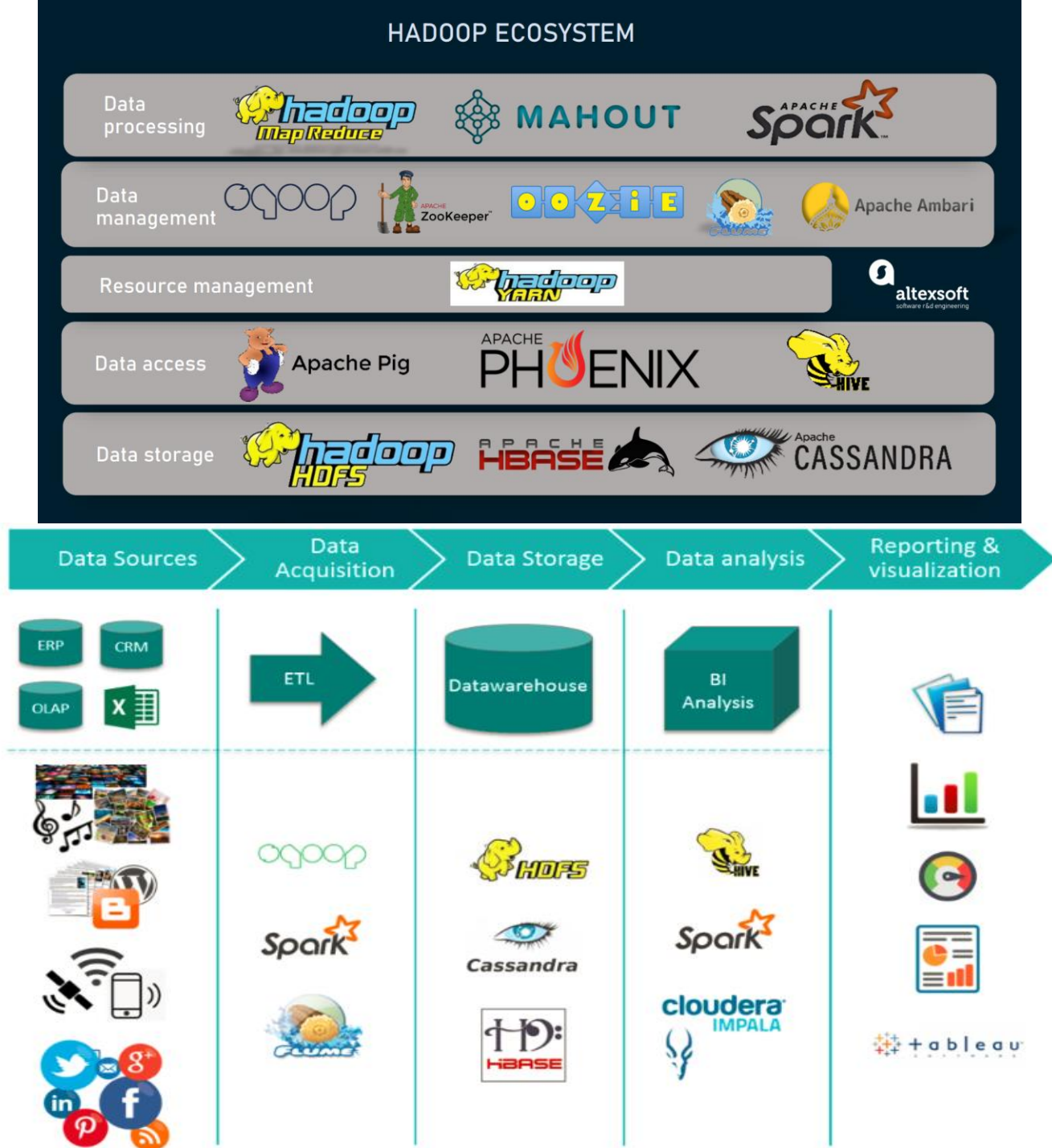
¿Qué es Apache HBase?

- - Base de datos NoSQL distribuida, diseñada para manejar grandes cantidades de datos.
- - Inspirada en Google Bigtable, se ejecuta sobre el sistema de archivos distribuido de Hadoop (HDFS).

¿Qué es HDFS?

- HDFS es un sistema de archivos distribuido que maneja grandes conjuntos de datos que se ejecutan en hardware básico. Se utiliza para escalar un solo clúster de Apache Hadoop a cientos (e incluso miles) de nodos. HDFS es uno de los componentes principales de Apache Hadoop, siendo los otros MapReduce y YARN. HDFS no debe confundirse ni reemplazarse por Apache HBase, que es un sistema de administración de bases de datos no relacionales orientado a columnas que se encuentra en la parte superior de HDFS y puede soportar mejor las necesidades de datos en tiempo real con su motor de procesamiento en memoria.

Almacenamiento de datos



Arquitectura de HBase

Características principales

- - Alta escalabilidad horizontal.
- - Modelado basado en tablas con filas y columnas.
- - Soporte para operaciones de lectura/escritura en tiempo real.

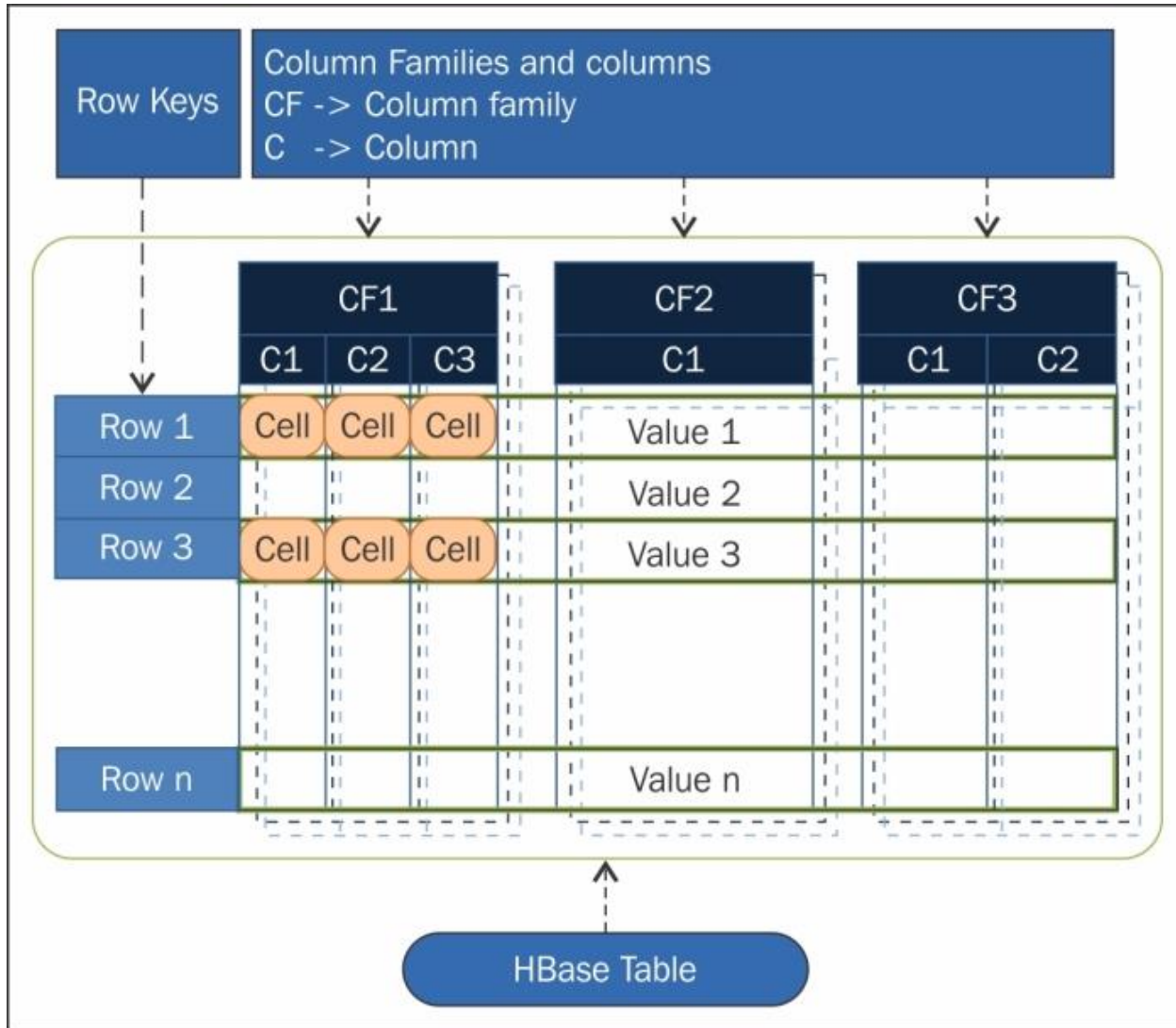
Componentes principales

- - HMaster: Coordina las operaciones, maneja la asignación de regiones.
- - RegionServer: Gestiona la lectura y escritura de datos para las regiones asignadas.
- - Zookeeper: Coordina y mantiene el estado de la red, esencial para la alta disponibilidad.

Regiones y tablas

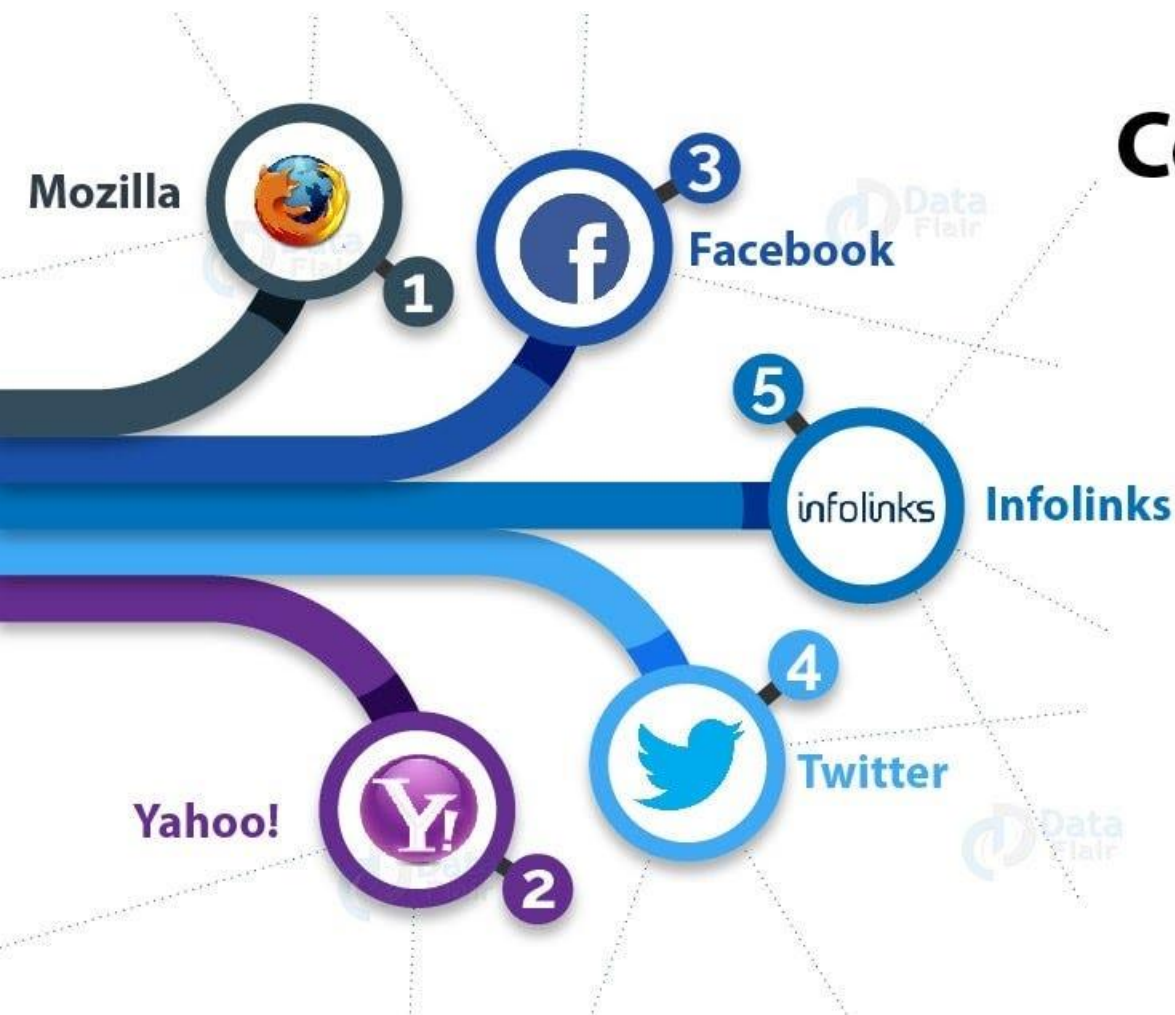
- - Las tablas se dividen en regiones, que son el conjunto de filas.
- - Las regiones se asignan a diferentes RegionServers para equilibrar la carga.

Ejemplo de estructura de tablas Hbase:



Casos de Uso

- Manejo de grandes volúmenes de datos
- - Ideal para almacenar y gestionar billones de filas y columnas.
- - Usado en aplicaciones como análisis de datos, recomendaciones personalizadas, y registros históricos.
- Consultas en tiempo real
- - Soporte para consultas rápidas en datasets masivos.
- - Ejemplos: redes sociales, análisis en tiempo real de datos de sensores.



Companies Using

APACHE
HBASE

Ventajas y Desventajas

Ventajas

- - Alta escalabilidad y capacidad de manejo de grandes volúmenes de datos.
- - Integración nativa con Hadoop y otros componentes del ecosistema.
- - Soporte para operaciones en tiempo real.

Desventajas

- - Curva de aprendizaje más pronunciada en comparación con otras bases de datos NoSQL.
- - Menos adecuado para operaciones complejas de agregación o uniones (joins).

Comparativa con otras tecnologías

HBase vs. RDBMS

- - HBase es preferible para datos no estructurados o semiestructurados, mientras que RDBMS es mejor para datos estructurados con relaciones complejas.

HBase vs. Cassandra

- - HBase está optimizado para consistencia fuerte, mientras que Cassandra se enfoca más en la disponibilidad.

edureka!

- 1 Description
- 2 **Data Model**
- 3 Implementation Language
- 4 Query Language
- 5 Performance
- 6 Security
- 7 Replication Methods
- 8 Competitive Advantages
- 9 Application Areas
- 10 Market Metrics

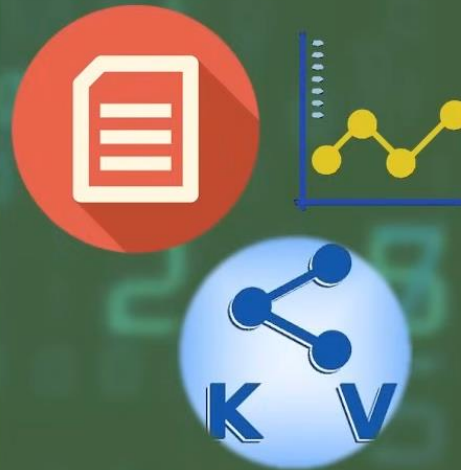


cassandra

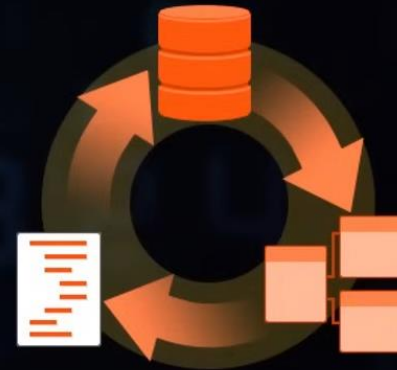
Key Spaces



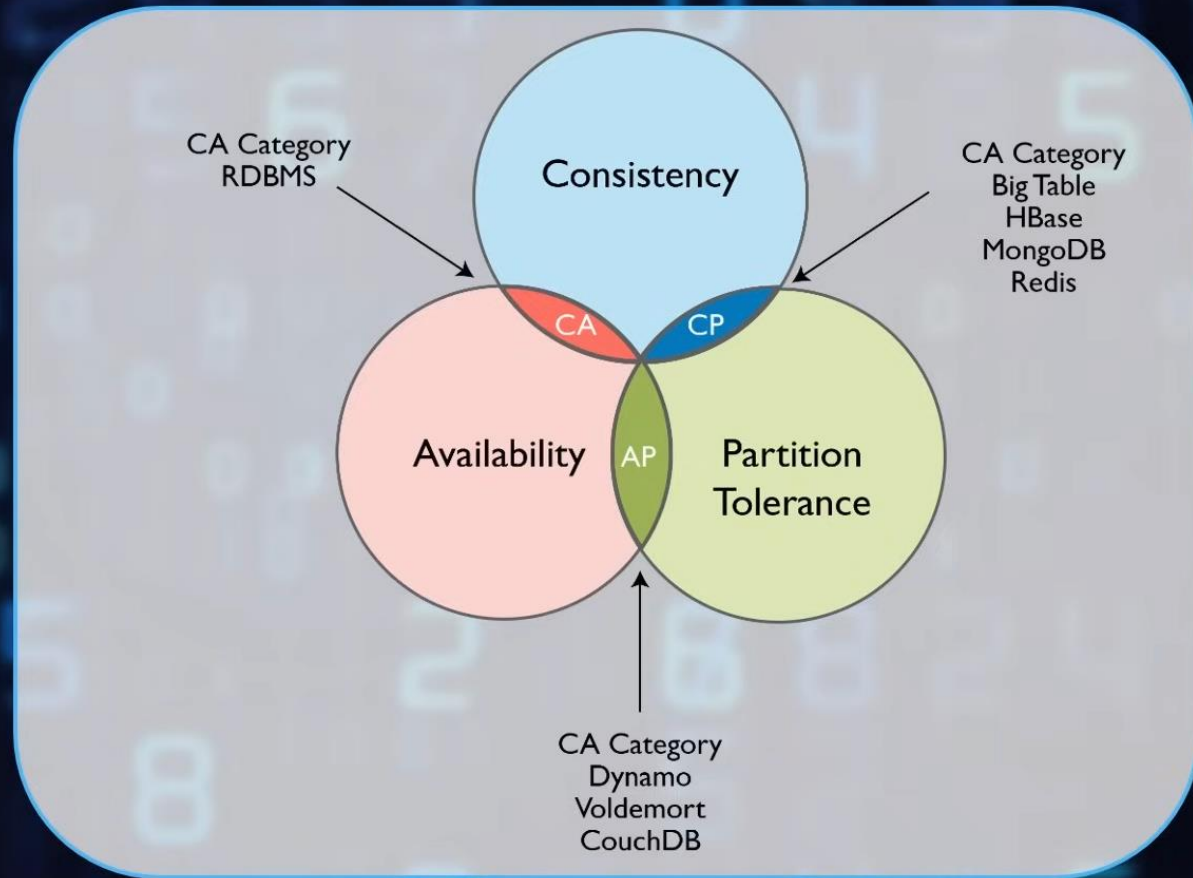
Flexible Schema



Column Oriented Database



$C \rightarrow$ Consistency
 $A \rightarrow$ Availability
 $P \rightarrow$ Partition Tolerance



edureka!

- 1 Description
- 2 Data Model
- 3 Implementation Language
- 4 Query Language
- 5 **Performance**
- 6 Security
- 7 Replication Methods
- 8 Competitive Advantages
- 9 Application Areas
- 10 Market Metrics



cassandra



Comparatively better



edureka!

- 1 Description
- 2 Data Model
- 3 Implementation Language
- 4 Query Language
- 5 Performance
- 6 Security
- 7 Replication Methods
- 8 Competitive Advantages
- 9 **Application Areas**
- 10 Market Metrics



edureka!

- 1 Description
- 2 Data Model
- 3 Implementation Language
- 4 Query Language
- 5 Performance
- 6 Security
- 7 Replication Methods
- 8 Competitive Advantages
- 9 Application Areas
- 10 Market Metrics



cassandra



40% of Fortune 100 companies



40 million downloads



7% of the companies in the world

Conclusiones

- HBase es una solución poderosa para gestionar grandes volúmenes de datos que requieren acceso rápido y escalabilidad.
- Es esencial evaluar el caso de uso específico para determinar si HBase es la mejor opción en comparación con otras tecnologías de bases de datos.

A tall, slender glass filled with a golden beer and a thick, white head of foam. The text "BIG DATA & BREWS" is printed in a bold, white, sans-serif font with a slight shadow effect, positioned in the center of the glass. The background is a blurred outdoor scene with trees and a building, suggesting a brewery or a pub setting.

**BIG DATA
& BREWS**

Gracias



Referencias bibliograficas

- edureka! (2018, October 17). Cassandra vs MongoDB vs HBase | Difference Between Popular NoSQL Databases | Edureka [Video]. YouTube. <https://www.youtube.com/watch?v=QlqylUequeis>
- Big Data & Brews. (2013, November 19). How HBase works [Video]. YouTube. https://www.youtube.com/watch?v=ISrNUyMR_Ek
- OpenAI. (2024). ChatGPT (August 2024 Version) [Large language model]. OpenAI. <https://www.openai.com/chatgpt>