

What is H Base? Explain storage mechanism of H Base with an example.

In Previous Years Questions

HBase is an open-source, distributed, non-relational database designed for handling large-scale, real-time data.

It's built on top of the Hadoop Distributed File System (HDFS) and inspired by Google's Bigtable.

H Base key features:

- **Distributed:** Stores data across multiple nodes in a cluster, enabling horizontal scaling and fault tolerance.
- **Non-relational:** Uses a flexible schema with rows, columns, and timestamps, unlike relational databases with fixed tables and relationships.
- **Column-oriented:** Stores data in columns instead of rows, allowing efficient access to specific data points.
- **Versioned:** Each data point has a timestamp, allowing you to access historical versions.

Storage Mechanism

Imagine a library with books representing tables, shelves representing regions, and individual pages representing rows. Each page is further divided into sections (columns) containing specific information (data points). A unique book title identifies each table (row key).

- **Horizontal Scaling:** Adding more shelves (regions) increases the library's capacity.
- **Column-oriented Access:** You can directly access a specific section (column) on a page (row) without flipping through the entire book (table).
- **Versioning:** Each page has revisions (timestamps), allowing you to see past versions of

What is H Base? Explain storage mechanism of H Base with an example.

the information.

Example

Imagine storing website clickstream data in HBase. Each row would represent a user session, with columns for timestamps, visited pages, and actions taken. You could efficiently query for users who visited a specific page within a certain timeframe, regardless of their entire browsing history.