

Strengths and weaknesses of databases cont'd**4. Wide-column store**

Wide-column stores are dynamic column-oriented nonrelational databases. Wide column stores are database management systems that organize related facts into columns. Groups of these columns, called “column families”, have multiple rows and rows may not have same number of columns. These stores are sometimes seen as a type of key-value store but have attributes of traditional relational databases as well. Other names of Wide-Column store databases are “column families,” “columnar databases,” “extensible record stores” or “column-oriented DBMS.”

Wide-column stores use the concept of a keypace instead of schemas. A keypace encompasses column families (similar to tables but more flexible in structure), each of which contains multiple rows with distinct columns. Each row doesn't need to have the same number or type of column. The columns name as well as record keys are not fixed in these databases.

Architecture of Wide-Column Databases uses persistent, sparse matrix, multi-dimensional mapping (row-value, column-value, and timestamp) in a tabular format meant for massive scalability (over and above the petabyte scale). A timestamp determines the most recent version of data. Column Family stores do not follow the relational model, and they aren't optimized for joins. These are mainly used in highly analytical and query-insensitive environments.

Strengths

- This type of database has some benefits of both relational and nonrelational databases. It deals better with both structured and semistructured data than other nonrelational databases, and it's easier to update. Columnar databases compress better than row-based systems. Also, large data sets are simple to explore.

- Columnar databases are highly scalable horizontally and faster at scale because they store data in columns rather than rows. The columns are easy to scale so they can store large volumes of data. This feature also allows users to spread data across many computing nodes and data stores.
- Column stores are very efficient at data compression. Columnar databases are highly compressed compared to conventional relational databases that store data by row. They allow users to optimize storage size.
- Due to their structure, columnar databases perform particularly well with aggregation queries such as SUM, COUNT, AVG, etc.
- Database users can sort and manipulate data directly from a columnar database and there is no need to rely on the application.
- Columnar stores can be loaded extremely fast. A billion row table could be loaded within a few seconds. You can start analysing almost immediately.

Weaknesses

- Wide-column stores are slower than relational databases when handling transactions.
- Good for bulk write and updates but it's a costly affair in lower quantity like individual records.
- Not suitable for Incremental data loading & Queries against only a few rows

Final Verdict

Based on the various parameters, strengths and weaknesses, a wide-column Store database may be used for

- Big data analytics where speed is important
- Reporting systems
- Time Series Data
- Data warehousing on big data
- Extreme write speeds with relatively less velocity reads like Logging
- Sensor Logs [Internet of Things (IOT)]

- User preferences
- Geographic information

Some of the examples of wide-column Store databases are Cassandra, HBase, Microsoft Azure Cosmos, Druid, Hypertable etc.

To be continued...

Assignment

1. What are Wide-Column Store of databases? Where they are used?
2. In what situations, Wide-Column store databases are not suggested?