

Question- What is NOSQL Database?

Answer -NoSQL is an approach to database design that can accommodate a wide variety of data models, including key-value, document, columnar and graph formats. NoSQL, which stand for "not only SQL," is an alternative to traditional relational databases in which data is placed in tables and data schema is carefully designed before the database is built. NoSQL databases are especially useful for working with large sets of distributed data.

Question-How does data get stored in NoSQL database?

Answer- Varies based on database type.

For example, key-value stores function similarly to SQL databases, but have only two columns ('key' and 'value'), with more complex information sometimes stored as BLOBs within the 'value' columns. Document databases do away with the table-and-row model altogether, storing all relevant data together in single 'document' in JSON, XML, or another format, which can nest values hierarchically.

Question- What is a column family in Hbase?

Answer - HBase tables are organized by column, rather than by row. The columns are organized in groups called *column families*. When creating a HBase table, we must define the column families before inserting any data.

e.g. create 'employee','saurabh'

here employee is table and saurabh is column family.

Question- How many maximum number of columns can be added to Hbase table?

Answer- There is no hard limit to number of columns in HBase , we can have more than 1 million columns.

Question- Why columns are not defined at the time of table creation in Hbase?

Answer-Columns are not defined at the time of table creation in Hbase **So that you have the flexibility of adding column at the time of entering the data.**

e.g. I have column called as C1 today and the value is V1 today. Suppose tomorrow I get the data feed in which there is column C2 I will add that column at the run time. My data is unstructured I don't know what field will come in future today is C1 and may be future I will be C100. So I need the flexibility the I will add the column value at the time of entering the data whenever I get it.

Question- How does data get managed in Hbase?

Answer- Data in Hbase is organized into tables. Any characters that are legal in file paths are used to name tables. Tables are further organized into rows that store data. Each row is identified by a unique row key which does not belong to any data type but is stored as a bytearray. Column families are further used to group data in rows. Column families define the physical structure of data so they are defined upfront and their modification is difficult. Each row in a table has same column families. Data in a column family is addressed using a column qualifier.

Question- What happens internally when new data gets inserted into Hbase table?

Answer- To write data to HBase, you use methods of the HTableInterface class. You can use the Java API directly, or use HBase Shell, Thrift API, REST API, or another client which uses the Java API indirectly. When you issue a Put, the coordinates of the data are the row, the column, and the timestamp. The timestamp is unique per version of the cell, and can be generated automatically or specified programmatically by your application, and must be a long integer.