

1. What is NoSQL data base?

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.

2. How does data get stored in NoSQL database?

Graph stores are used to store information about networks of data, such as social connections. Graph stores include Neo4J and Giraph. Key-value stores are the simplest NoSQL databases. Every single item in the database is stored as an attribute name (or 'key'), together with its value.

3. What is a column family in HBase?

A HBase table is comprised of one or more column families, each of which is stored in a separate set of regionfiles sharing a common key. To express it in terms of an RDBMS, a column family is roughly analogous to a RDBMS table with the rowkey as a clustered primary key index.

4. How many maximum number of columns can be added to HBase table?

There is no hard limit to number of columns in HBase , we can have more than 1 million columns but usually three column families are recommended (not more than three)

5. Why columns are not defined at the time of table creation in HBase?

It is so because:

- if you are retrieving data that a row key maps to, you'd get data from all column families related to the row that the row key identifies
- if you are retrieving data which a particular column family maps to, you'd get all column qualifiers and associated data (maps with timestamps as keys and corresponding values)
- if you are retrieving data that a particular column qualifier maps to, you'd get all timestamps (versions) for that column qualifier and all associated values.