1. What is NoSQL data base?

NoSQL is a class of database management systems (DBMS) that do not follow all of the rules of a relational DBMS and cannot use traditional SQL to query data. The term is somewhat misleading when interpreted as "No SQL," and most translate it as "Not Only SQL," as this type of database is not generally a replacement but, rather, a complementary addition to RDBMSs and SQL.

NoSQL-based systems are typically used in very large databases, which are particularly prone to performance problems caused by the limitations of SQL and the relational model of databases.

2. How does data get stored in NoSQl database?

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.

3. What is a column family in HBase?

Columns in Apache HBase are grouped into column families. All column members of a column family have the same prefix. For example, the columns courses:history andcourses:math are both members of the courses column family. The colon character (:) delimits the column family from the . The column family prefix must be composed ofprintable characters. The qualifying tail, the column family qualifier, can be made of any arbitrary bytes. Column families must be declared up front at schema definition time whereas columns do not need to be defined at schema time but can be conjured on the fly while the table is up an running.

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

HBase currently does not do well with anything above two or three column families so keep the number of column families in your schema low. Currently, flushing and compactions are done on a per Region basis so if one column family is carrying the bulk of the data bringing on flushes, the adjacent families will also be flushed though the amount of data they carry is small. When many column families the flushing and compaction interaction can make for a bunch of needless i/o loading

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

In the HBase data model columns are grouped into column families, which must be defined up front during table creation

6. How does data get managed in HBase?

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. It is not necessary to specify column qualifiers in advance and there is no consistency requirement between row

7. What happens internally when new data gets inserted into HBase table?

Incoming data will be added to the new rows/cells in Hbase table and key values will get updated.