SESSION 9 ASSIGNMENT 1

1. What is NoSQL data base?

Soln. A NoSQL database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases.

2. How does data get stored in NoSQl database?

Soln. **Key-value data stores**: Key-value NoSQL databases emphasize simplicity and are very useful in accelerating an application to support high-speed read and write processing of non-transactional data. Stored values can be any type of binary object (text, video, JSON document, etc.) and are accessed via a key. The application has complete control over what is stored in the value, making this the most flexible NoSQL model. Data is partitioned and replicated across a cluster to get scalability and availability. For this reason, key value stores often do not support transactions. However, they are highly effective at scaling applications that deal with high-velocity, non-transactional data.

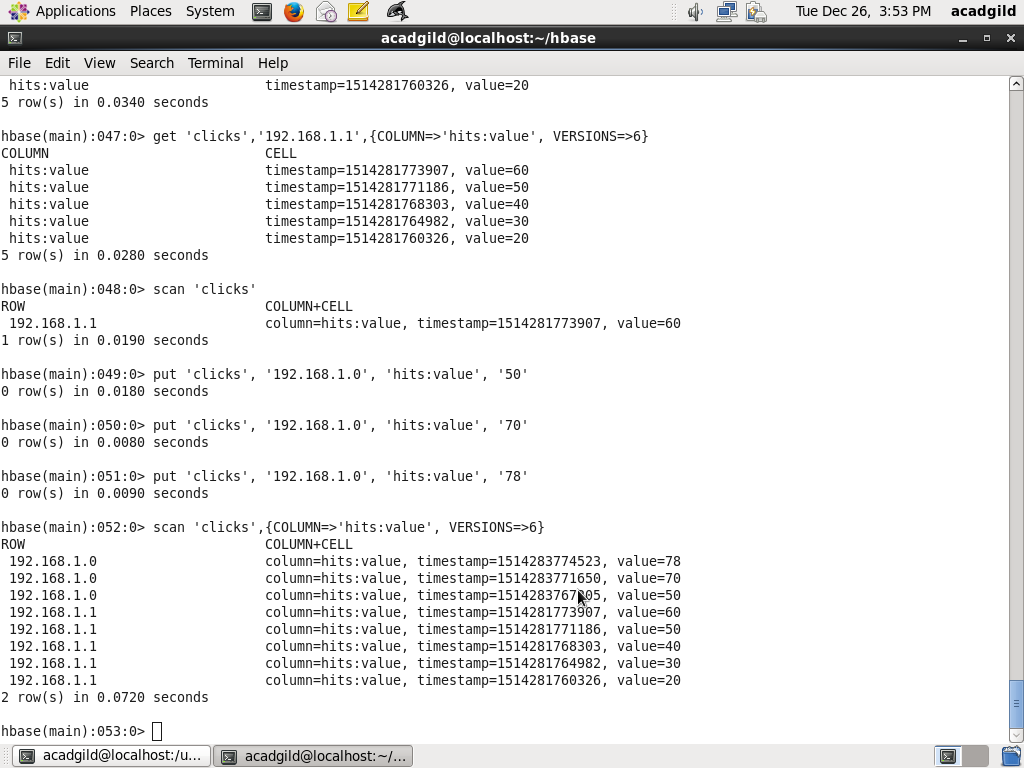
**Document stores**: Document databases typically store self-describing JSON, XML, and BSON documents. They are similar to key-value stores, but in this case, a value is a single document that stores all data related to a specific key. Popular fields in the document can be indexed to provide fast retrieval without knowing the key. Each document can have the same or a different structure.

**Wide-column stores**: Wide-column NoSQL databases store data in tables with rows and columns similar to RDBMS, but names and formats of columns can vary from row to row across the table. Wide-column databases group columns of related data together. A query can retrieve related data in a single operation because only the columns associated with the query are retrieved. In an RDBMS, the data would be in different rows stored in different places on disk, requiring multiple disk operations for retrieval.

**Graph stores**: A graph database uses graph structures to store, map, and query relationships. They provide index-free adjacency, so that adjacent elements are linked together without using an index.

3. What is a column family in HBase?

Soln. In the HBase data model columns are grouped into column families, which must be defined up front during table creation. Column families are stored together on disk, which is why HBase is referred to as a column-oriented data store.



For eg. In above screen shot we can see ‘clicks’ table having hits as column family.

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

Soln. There is actually not any limit of this kind, we must design our model keeping our use case into mind.

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

Soln. Column families need to be defined at the time of table creation whereas columns can be defined dynamically on the go as the requirement comes.

6. How does data get managed in HBase?

Soln. The Hbase data model is different from the model provided by relational databases. Hbase is referred to by many terms like a key-value store, column oriented database and versioned map of maps which are correct. The easiest way of visualizing a Hbase data model is a table that has rows and tables. This is the only similarity shared by Hbase model and the relational model. Data in Hbase is organized into tables.

7. What happens internally when new data gets inserted into HBase table?Soln. When the client issues a Put request, the first step is to write the data to the write-ahead log, the WAL:

- Edits are appended to the end of the WAL file that is stored on disk.

- The WAL is used to recover not-yet-persisted data in case a server crashes.