**ASSIGNMENT-9.1**

1. What is NoSQL data base?

A NoSQL (originally referring to "non SQL" or "non-relational") database provides a mechanism for storage and retrieval of data that is modelled in means other than the tabular relations used in relational databases. 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.

**Key-value stores**

Key-value stores, or key-value databases, implement a simple data model that pairs a unique key with an associated value. Because this model is simple, it can lead to the development of key-value databases, which are extremely performant and highly scalable for session management and caching in web applications.

**Document databases**

Document databases, also called document stores, store semi-structured data and descriptions of that data in document format. They allow developers to create and update programs without needing to reference master schema. Use of document databases has increased along with use of JavaScript and the JavaScript Object Notation (JSON), a data interchange format that has gained wide currency among web application developers, although XML and other data formats can be used as well.

**Wide-column stores**

Wide-column stores organize data tables as columns instead of as rows. Wide-column stores can be found both in SQL and NoSQL databases. Wide-column stores can query large data volumes faster than conventional relational databases.

**Graph stores**

Graph data stores organize data as nodes, which are like records in a relational database, and edges, which represent connections between nodes. Because the graph system stores the relationship between nodes, it can support richer representations of data relationships. Also, unlike relational models reliant on strict schemas, the graph data model can evolve over time and use.

3. What is a column family in HBase?

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.

|  |  |
| --- | --- |
| **Row Key** | **Column Family: {Column Qualifier:Version:Value}** |
| 00001 | CustomerName: {‘FN’: 1383859182496:‘John’, ‘LN’: 1383859182858:‘Smith’, ‘MN’: 1383859183001:’Timothy’, ‘MN’: 1383859182915:’T’} ContactInfo: {‘EA’: 1383859183030:‘John.Smith@xyz.com’, ’SA’: 1383859183073:’1 Hadoop Lane, NY 11111’} |
| 00002 | CustomerName: {‘FN’: 1383859183103:‘Jane’, ‘LN’: 1383859183163:‘Doe’, ContactInfo: { ’SA’: 1383859185577:’7 HBase Ave, CA 22222’} |

The table shows two column families: CustomerName and ContactInfo. When creating a table in HBase, the developer or administrator is required to define one or more column families using printable characters.

1. 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 limit the number based on the design of our model, keeping our use case into mind.

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

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.

1. How does data get managed in HBase?

Hbase is natively supported on Hadoop. The main characteristics that make Hbase an excellent data management platform are fault tolerance, speed and usability. Fault tolerance is provided by automatic fail-over, automatically shared and load balanced tables, strong consistency in row level operations and replication.

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?

INSERT INTO table-name

Specifies the HBase table into which data is to be inserted.

columns

Specifies the order and column names of those columns into which data is to be inserted. If no column names are specified, data is to be inserted into all columns that were listed, and in the order that was specified, when the named table was created.

VALUES (expressions)

Expressions supply values for every column. If a column list is specified, the expression list is evaluated to provide values for those columns. NULL is inserted for those columns that are omitted from the column list.

DISABLE WAL

Disables write-ahead logging for HBase writes and puts. Disabling write-ahead logging increases the performance of write operations, but it can result in data loss if the region servers fail. Your client application is responsible for ensuring data consistency when you use this option.