Study of NoSQL Database Along With Security Comparison

Ankita A. Mall [1], Jwalant B. Baria [2]

[1] Student, Computer Engineering Department, Government Engineering College, Modasa, Gujarat, India
ank.fetr@gmail.com[1]

[2] Assistant professor, Computer Engineering Department, Government Engineering College, Modasa, Gujarat, India

ABSTRACT

This paper considers the top 5 open source NoSQL Database Redis, MongoDB, CouchDB, Hyper table, and Cassandra. NoSQL Database is leading database in the aspect of the industry and the current requirement of technology. Each database is designed in such a way, in which their performance can be evaluated based on the flexibility, security, durability etc feature provided by them. The database should be selected, with respect to the need of the application. Some database gives worst performance if they are not flexible with the application.

Introduction

In today technology, the way requirement of data changes rapidly, the mechanism to store such data should also be changed. NoSQL is used to satisfy this requirement as it do not require pre knowledge of format of data, it can accept the rapid changes in data in less amount of time. NoSQL is a term to individualize a system from a traditional method to store data or RDBMS.

In a more precise way it can be define as, collection of concepts that allow fast and efficient processing on data with focus on performance, reliability and agility.

NoSQL have following advantage over SQL such as:

- More number of data format over SQL: - It support format key-value stores, graph databases, column-family stores, document stores and even rows in tables.
- Free of joins:-It do not constraint the database to have join, it can easily perform this mechanism by using interface to extract data.
- Schema free:-E-R diagram is not essential in NoSQL
- Support multiple processors:-It allows storing database on several processors and speed performance is very high.
- Linear scalability supported:-It allows adding database in already setup database, to improve performance.

There are four types of NoSQL data bases:-

1. Key-Value databases,
2. Column oriented database
3. Document oriented databases
4. Graph databases.

These all database follow similar structure of RDBMS database. The database server can contain more than one database. Each database might contain more than one table. The terminologies vary from one database to other.

1. KEY-VALUE

This database based on paper published by Dynamo Whitepaper under Amazon. It allow data to be stored in simple <key>=<value> format, where retrieval of data will be based on key.

- Data Model:-The database will contain many key spaces and each key space can have many variables to store key value pairs. It can be represented as key in column wise and value in row. It is suitable for building simple, non-complex, high available applications. Since most of Key Value Databases support in memory storage, can be used for building cache mechanism.

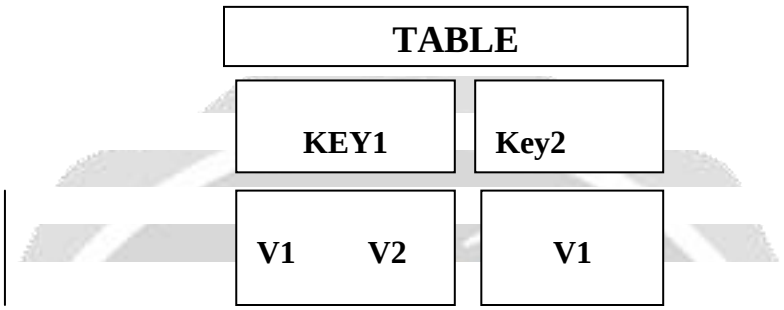

Figure 1.Key-Value database table

The schema-less format of key-value database look like Riak. The can be synthetic or auto-generated. Value of key can be BLOB, JSON, String etc.[1]

It uses hash table for storing key and value, so the time complexity is O (1).It make the access fast and storing the record. It consists of unique key and pointer to point a particular item of data. Hash table do not actually bind them physically, there can be identical keys in different hash table bucket[2].performance is enhance by great degree because of cache mechanism. In order to read value of data, one should know the hash bucket and key value in order to get value.

CAP theorem in key-value database is great in availability and partition concept but failed in consistency.

Key	Value
India	B-25, Sector-58, Noida, India – 201301
Bardoli	394601
Surat	{city of diamond}

Figure 2.Key-Value database stored example

This key/value type database allow clients to read and write values using a key as follows[3]:

- Get (key), returns the value associated with the provided key.
- Put (key, value), associates the value with the key.
- Multi-Get (key1, key2, .., key N), returns the list of values associated with the list of keys.

- Delete (key), removes the entry for the key from the data store.

2. **Document Store NoSQL Database**

Document store NoSQL databases storage is analogous to key-value databases in that there's also similar relationship alike key and a value. Data is stored as a value. Its associated key is the unique identifier for that value[4].The only difference in document store and key-value is that, in a document database, the value contains structured or semi-structured data. This structured/semi-structured value is referred to as a document and can be in XML, JSON or BSON format. It is preferred in Blogging platforms, Content management system, analytics platforms, e-commerce.

Examples of document store NoSQL databases are MongoDB, Apache CouchDB, and Elastic search. It's used to run complex search queries, which contain a proper structure or format.

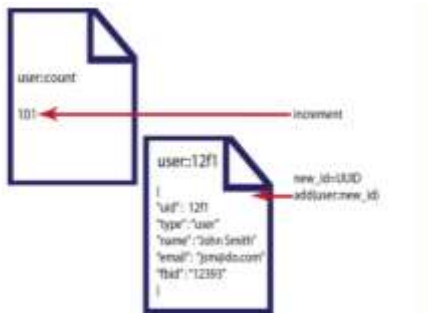

Figure 3. Document Store NoSQL structure

One key difference between a key-value store and a document store is that the document store type embeds attribute metadata associated with stored content, which essentially provides a way to query the data based on the contents[3]

{office Name:"3Pillar Noida", {Street: "B-25, City:"Noida", State:"UP", Pin code:"201301"}}
{office Name:"3Pillar",{Boulevard:"Coriolanus Brediceanu No. 10", Block:"B, 1st Floor", City: "Timisoara", Pin code: 300011"}}
{office Name:"3Pillar Latitude:"40.748328", Longitude:"-73.985560"}}

Figure 4. Document Store NoSQL example

For example, in the above example, one could search for all documents in which "City" is "Noida" that would deliver a result set containing all documents associated with any "3Pillar Office" that is in that particular city.

CouchDB uses JSON, to store data, JavaScript as its query language using Map Reduce and HTTP for an API. Data and relationships are not stored in tables as is a norm with conventional relational databases but in fact are a collection of independent documents. The fact that document style databases are schema-less makes adding fields to JSON documents a simple task without having to define changes first.

3. Column store NoSQL database

In column-oriented NoSQL database, data are arranged with respect to column, in cells. Columns are logically grouped in family of column. Column family can be unlimited no of column that can be created at defining of schema or run time. Read and write are performing on column, instead of traditional way of inserting data in row. Column family is group of similar data that are accessed together usually. As example, Student information and parent's detail, but information is not in same order.[3] The main advantages of storing data in columns over relational DBMS are fast search/access and data aggregation. Different rows are stored in different places on the disk while columnar databases store all the cells corresponding to a column as a continuous disk entry, thus making the search/access faster.

Column-oriented NoSQL are used in Content management systems, Service that have aspiring usage, System that require heavy writes request etc. Examples are Cassandra and Apache, Hadoop, Hbase.

Figure **5**. Column-Store Nosql database

In comparison, most relational DBMS store data in rows, the benefit of storing data in columns, is fast search/ access and data aggregation. Relational databases store a single row as a continuous disk entry. Different rows are stored in different places on disk while Columnar databases store all the cells corresponding to a column as a continuous disk entry thus makes the search/access faster.

For example: To query the titles from a bunch of a million articles will be a painstaking task while using relational databases as it will go over each location to get item titles. On the other hand, with just one disk access, title of all the items can be obtained.

The best known examples are Google's Big Table and HBase & Cassandra that were inspired from Big Table.

Big Table, for instance is a high performance, compressed and proprietary data storage system owned by Google. It has the following attributes:

- **Sparse** – some cells can be empty
- **Distributed** – data is partitioned across many hosts
- **Persistent** – stored to disk
- **Multidimensional** – more than 1 dimension
- **Map** – key and value
- **Sorted** – maps are generally not sorted but this one is

4. Graph database

A graph database, is called a graph-oriented database, it uses graph theory to store, map and query relationships. A graph database is collection of nodes and edge. Each node represents relationship between 2 nodes. Every node represent a unique identifier, set of edges expressed by key-value pair. Examples are Neao4j, ArangoDB etc

Figure **6**. Graph database Nosql database

- Edges and node are used to represent database and store data.
- Relationship among node is represented, with the help of edge and arrange in some organized way.
- Relationship among node and database are defined some properties

Graph databases are an prominent choice for working with connected data . Data that contains lots of interconnected relationships. Graph databases are much more suited to displaying relational data than relational databases.

It is mostly suited for social network, real-time, product recommendations, access management and much more. It do not contain any fixed schema which bind its property .Any schema is just a mirror or simply a reflection of the data that has been entered. Schema grows as more varied data entered into it.

5. Comparison of Nosql database and their performance parameter[2]

Data Model	Perfor-mance	Scalable	Compl-exity	Function
RD	Variable	Variable	moderate	algebra
KD	high	high	none	Variable
GD	Variable	Variable	high	Graph theory
DS	high	Variable	low	Variable
CS	high	high	low	Minimal

Figure 7. **Comparison of Nosql database and their performance parameter**

RD-Relational Database KD-key-store database GD-Graph Database Ds-Document Store CS-Column Store

6. The Security Comparison of the Top 5 Open Source NoSQL Databases[1]

SECURITY ISSUES	**MD**	CD	COD	RD
DFE	No encrypt	No encrypt	No encrypt	No encrypt
CSA	Weak	Weak	SSL	No authen no encrypt

ICA	Weak	Weak	SSL	No authen no encrypt
SI	Vulnerable	Not vulnerable	Vulnera-ble	Not vulnerable
DSA	Not vulnerable	Vulnerable	Vulnera-ble	Not vulnerable

Figure 8. **The Security Comparison of the Top 5 Open Source NoSQL Databases**

MD- MONGODB COD- COUCHDB R- Redis CD- CASSANDRA DFE- Data files encryption CSA- Client/server authentication ICA- Inter-cluster Authentication SI- Script Injection DSA- Denial of service attack .

7. **References**

[1] V. J. Dindoliwala, R. D. Morena, and A. K. Store, "Survey on Security Mechanisms In NoSQL Databases," vol. 8, no. 5, pp. 333–338, 2017.

[2] A. Nayak, A. Poriya, and D. Poojary, "Type of NOSQL Databases and its Comparison with Relational Databases," Int. J. Appl. Inf. Syst., vol. 5, no. 4, pp. 16–19, 2013.

[3] Yishan Li;Sathiamoorthy Manoharan, "A performance comparison of SQL and NoSQL\ndatabases," 978-1-4799-1501-9/13- 2013 Ieee, pp. 15–19, 2013.

[4] Dadapeer, N. . Indravasan, and A. G, "A Survey on Security of NoSQL Databases," Int. J. Innov. Res. Comput. Commun. Eng., vol. 4, no. 4, pp. 5249–5254, 2016.