

CS204 - DATABASE

MANAGEMENT SYSTEMS

ASSIGNMENT

Submitted by,

GOKULNATH M PRABHU

Class : CS4B

Roll No : 21

INTRODUCTION TO NoSQL DATABASES

A NoSQL originally referring to non-SQL or non-relational is a database that provides a mechanism for storage and retrieval of data. This data is modeled in means other than the tabular relations used in relational databases. Such databases came into existence in the late 1960's but did not obtain the NoSQL moniker until a surge in popularity in early twentyfirst century. NoSQL databases are used in real-time web applications and big data and their use is increasing over time. NoSQL systems are also sometimes called Not only SQL to emphasize the fact that they may support SQL-like query languages.

A NoSQL database includes simplicity of design, simpler horizontal scaling to clusters of machines and finer control over availability. Many NoSQL stores compromise consistency in favour of availability, speed and partition tolerance. Barriers to the greater adoption of NoSQL stores include the use of low-level query languages, lack of standardized interface and huge previous investments in existing relational databases.

Advantages:

1. High scalability
2. High availability

Disadvantages:

1. Narrow focus
2. Open-source. So no reliable standard
3. Management challenge
4. GUI is not available
5. No approach to backup of data
6. Large document size.

Types of NoSQL databases:

NoSQL document based database

1. MongoDB falls in the category of NoSQL document based database
2. Key value store : Memcached, Redis, Coherence
3. Tabular : Hbase, Big Table, Accumulo
4. Document based : MongoDB, CouchDB, Cloudant.

KEY - VALUE DB

A key-value database or key-value store, is a data storage paradigm designed for storing, retrieving and managing associative arrays and a data structure more commonly known as a dictionary or hash table. Key-value systems treat the data as a single opaque collection which may have different fields for every record.

Redis or Remote Dictionary Server is an in-memory data structure store, used as a distributed, in-memory key-value database, cache and message broker with optional durability.

Main Characteristics of Key-Value DB (Redis)

1. It is exceptionally fast and can perform about 110,000 sets per second and about 81,000 gets per second.

2. Supports such data types

Redis data types include

* Strings - text or binary data upto 512 MB in size

* Lists

* Bit maps

* Sets

* Hyperlog Logs

* Sorted sets

* Streams

* Hashes

* Geospatial.

3. Simplicity and ease-of-use

Redis enables you to write traditionally complex code with fewer, simpler lines.

4. Replication and Persistence

Redis employs a primary-replica architecture and supports asynchronous replication, where data can be replicated to multiple replica servers. For persistence, Redis supports point-in-time backups.

5. Operations are atomic which ensures that if two clients concurrently access redis server, they will get the updated value.

6. Multifaceted tool and can be used in a number of usecases like caching, messaging queues, any short lived data, etc.

DOCUMENT DB

A document-oriented database or document store, is a computer ~~store~~ program and data storage system designed for storing, retrieving and managing document-oriented information, also known as semi-structured data.

Document oriented db are inherently a subclass of key-value store. The difference lies in the way the data is processed. A document DB relies on internal structure in the document in order to extract metadata that the database engine uses for further optimization.

Main Characteristics of Document DB (MongoDB)

1. Support ad hoc queries: In MongoDB, you can search by field, range query and it also supports regular expression searches.

2. It supports indexing
3. MongoDB supports Master Slave replication
4. Duplication of data. MongoDB can run over multiple servers. The data is duplicated to keep the system up and also keep its running condition in case of hardware failure.
5. Load balancing. MongoDB has an automatic load balancing configuration because of data placed in shards.
6. Supports map reduce and aggregation tools.
7. It is a schema-less database written in C++.
8. Provides high performance and stores files of any size easily without complicating the stack.
9. Easy to administer in the case of failures
10. It also supports:
 - * JSON data model with dynamic schemas
 - * Auto-sharding for horizontal scalability
 - * Built in replication for high availability.

COLUMN - FAMILY DB

A column-family is a database object that contains columns of related data. It is a tuple that consists of a key-value pair, where the key is mapped to a value that is a set of columns.

Cassandra is an open source, distributed and decentralized/distributed storage system (database) for managing very large amounts of structured data spread out across the world.

Main characteristics of Column-Family DB (Cassandra)

1. Elastic scalability: Cassandra is highly scalable if allows to add more hardware to accommodate more customers & more data as per requirement.
2. Always on architecture: Cassandra has no single point of failure and it is continuously available for business critical applications that cannot afford a failure.
3. Fast linear-scale performance: It increases throughput as you increase the no. of nodes in the cluster. Therefore it maintains a quick response time.
4. Flexible data storage: Cassandra accommodates all possible data formats including structured, semi-structured and unstructured.
5. Easy data distribution: Cassandra provides the flexibility to distribute data where you need by replicating across multiple data centers.
6. Transaction support: Cassandra supports Atomicity, Consistency, Isolation and Durability ~~(ACID)~~ (ACID).
7. Fast writes: Cassandra is designed to run on cheap commodity hardware.

GRAPH DB

A graph DB is a database that uses graph structures for semantic queries with nodes, edges and properties to represent and store data. The graph relates the data items in the store to a collection of nodes and edges, the edges representing the relationships between the nodes. The relationships allow data in the store to be linked together directly and in many cases, retrieved with one operation.

AragoDB is a free and open-source native multi-model database system. The database system supports three data models (key/value, documents, graphs) with one database core and a unified query language.

Main characteristics of graph DB (AragoDB)

1. It is a native multi-model paradigm.
2. A single query language is enough to retrieve data out of the database even though it supports multi-model.
3. AragoDB allows clients, such as browsers, to interact with the database with HTTP API, the API being resource oriented.
4. It is extendable with Javascript
5. It supports transactions and is compliant with Atomicity, Consistency, Isolation and Durability (ACID) properties.