

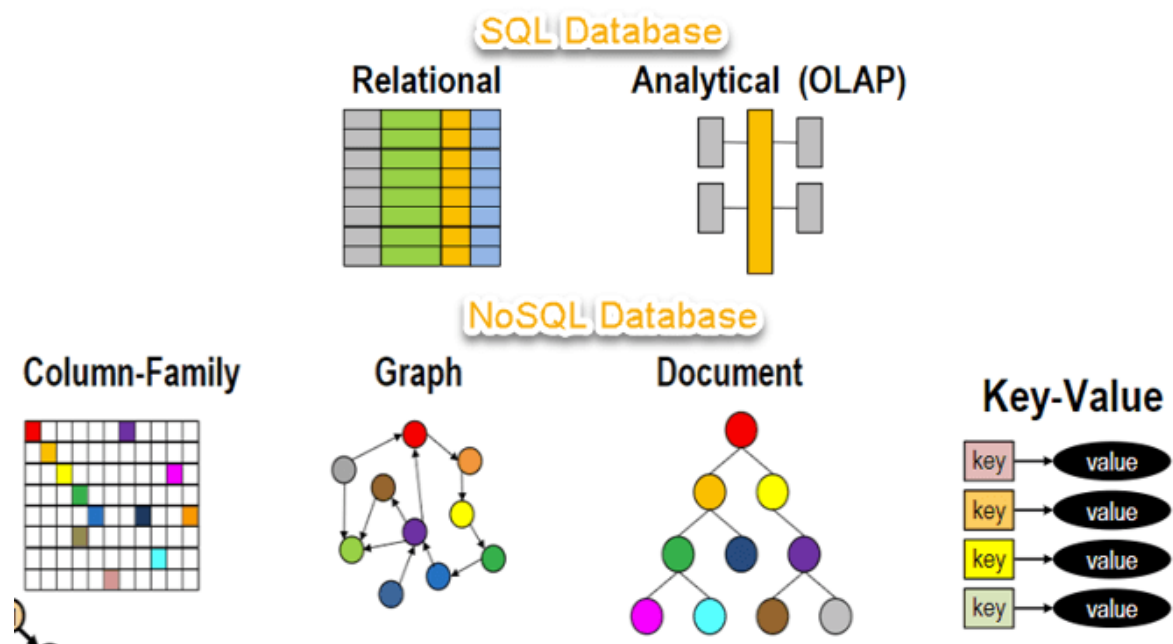
NoSQL Tutorial: Types of NoSQL Databases, What is & Example

What is NoSQL?

NoSQL Database is a non-relational Data Management System, that does not require a fixed schema. It avoids joins, and is easy to scale. The major purpose of using a NoSQL database is for distributed data stores with humongous data storage needs. NoSQL is used for Big data and real-time web apps. For example, companies like Twitter, Facebook and Google collect terabytes of user data every single day.

NoSQL database stands for "Not Only SQL" or "Not SQL." Though a better term would be "NoREL", NoSQL caught on. Carl Strozzi introduced the NoSQL concept in 1998.

Traditional RDBMS uses SQL syntax to store and retrieve data for further insights. Instead, a NoSQL database system encompasses a wide range of database technologies that can store structured, semi-structured, unstructured and polymorphic data. Let's understand about NoSQL with a diagram in this NoSQL database tutorial:



In this NoSQL tutorial for beginners, you will learn NoSQL basics like:

Why NoSQL?

Brief History of NoSQL Databases

Features of NoSQL

Types of NoSQL Databases

Query Mechanism tools for NoSQL

What is the CAP Theorem?

Eventual Consistency

Advantages of NoSQL

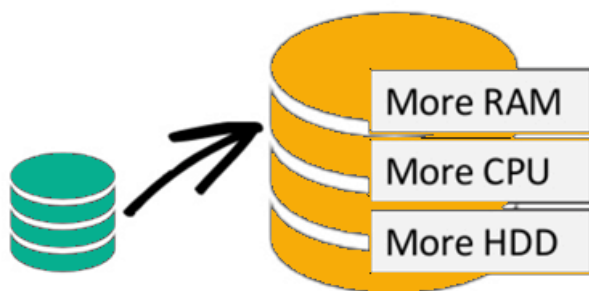
Why NoSQL?

The concept of NoSQL databases became popular with Internet giants like Google, Facebook, Amazon, etc. who deal with huge volumes of data. The system response time becomes slow when you use RDBMS for massive volumes of data.

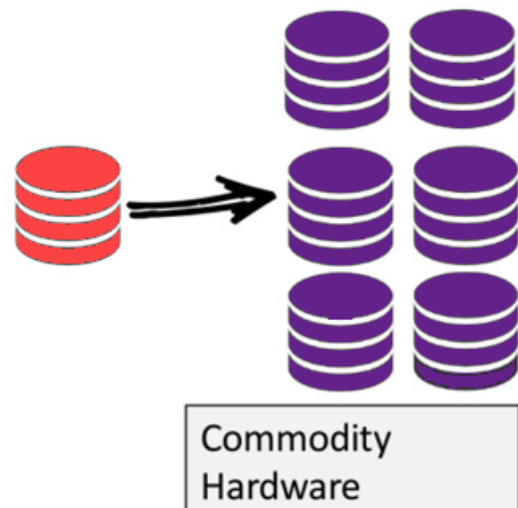
To resolve this problem, we could "scale up" our systems by upgrading our existing hardware. This process is expensive.

The alternative for this issue is to distribute database load on multiple hosts whenever the load increases. This method is known as "scaling out."

Scale-Up (*vertical* scaling):



Scale-Out (*horizontal* scaling):



NoSQL database is non-relational, so it scales out better than relational databases as they are designed with web applications in mind.

Brief History of NoSQL Databases

1998- Carlo Strozzi use the term NoSQL for his lightweight, open-source relational database

2000- Graph database Neo4j is launched

2004- Google BigTable is launched

2005- CouchDB is launched

2007- The research paper on Amazon Dynamo is released

2008- Facebooks open sources the Cassandra project

2009- The term NoSQL was reintroduced

Features of NoSQL

Non-relational

NoSQL databases never follow the relational model

Never provide tables with flat fixed-column records

Work with self-contained aggregates or BLOBs

Doesn't require object-relational mapping and data normalization

No complex features like query languages, query planners,
referential integrity joins, ACID

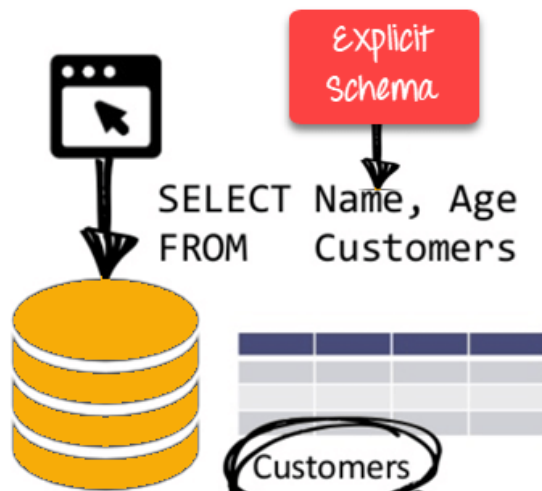
Schema-free

NoSQL databases are either schema-free or have relaxed schemas

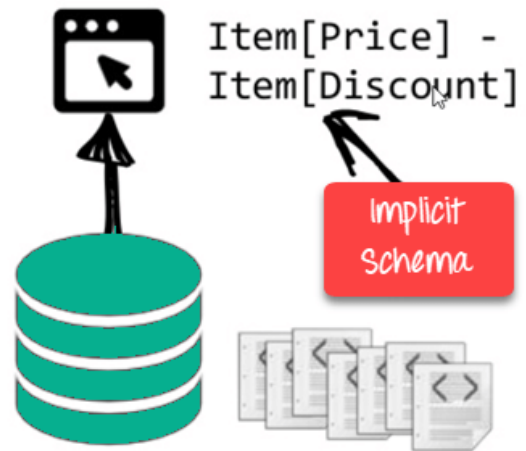
Do not require any sort of definition of the schema of the data

Offers heterogeneous structures of data in the same domain

RDBMS:



NoSQL DB:



Simple API

Offers easy to use interfaces for storage and querying data provided

APIs allow low-level data manipulation & selection methods

Text-based protocols mostly used with HTTP REST with JSON

Mostly used no standard based NoSQL query language

Web-enabled databases running as internet-facing services

Distributed

Multiple NoSQL databases can be executed in a distributed fashion

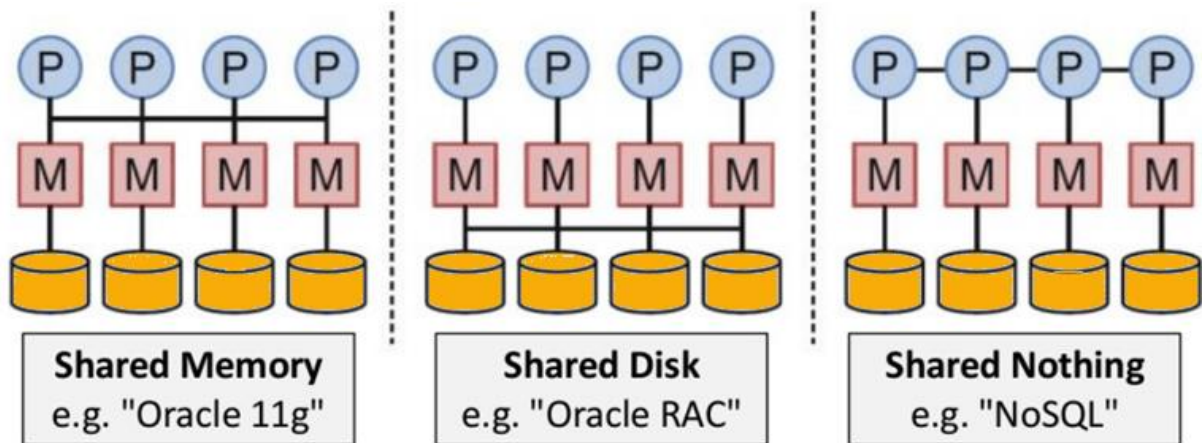
Offers auto-scaling and fail-over capabilities

Often ACID concept can be sacrificed for scalability and throughput

Mostly no synchronous replication between distributed nodes Asynchronous Multi-Master Replication, peer-to-peer, HDFS Replication

Only providing eventual consistency

Shared Nothing Architecture. This enables less coordination and higher distribution.



Types of NoSQL Databases

NoSQL Databases are mainly categorized into four types: Key-value pair, Column-oriented, Graph-based and Document-oriented. Every category has its unique attributes and limitations. None of the above-specified database is better to solve all the problems. Users should select the database based on their product needs.

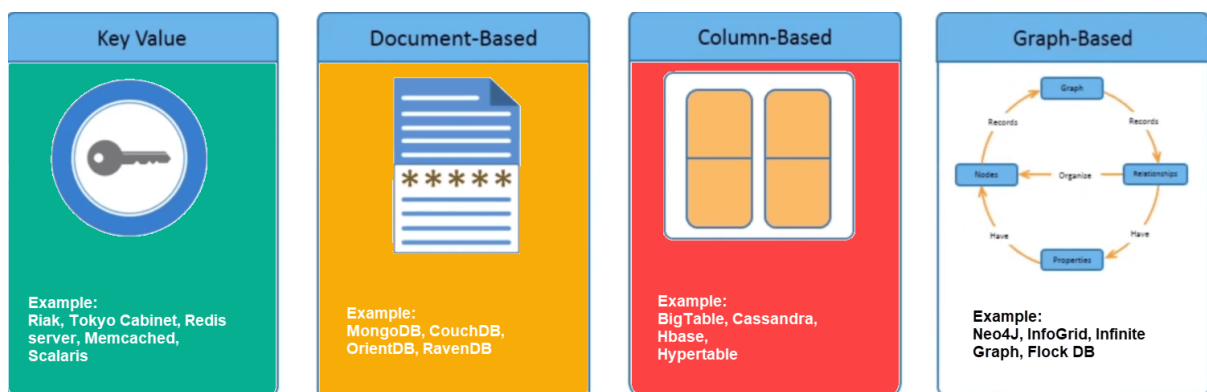
Types of NoSQL Databases:

Key-value Pair Based

Column-oriented Graph

Graphs based

Document-oriented



Key Value Pair Based

Data is stored in key/value pairs. It is designed in such a way to handle lots of data and heavy load.

Key-value pair storage databases store data as a hash table where each key is unique, and the value can be a JSON, BLOB(Binary Large Objects), string, etc.

For example, a key-value pair may contain a key like "Website" associated with a value like "Guru99".

Key	Value
Name	Joe Bloggs
Age	42
Occupation	Stunt Double
Height	175cm
Weight	77kg

It is one of the most basic NoSQL database example. This kind of NoSQL database is used as a collection, dictionaries, associative arrays, etc. Key value stores help the developer to store schema-less data. They work best for shopping cart contents.

Redis, Dynamo, Riak are some NoSQL examples of key-value store DataBases. They are all based on Amazon's Dynamo paper.

Column-based

Column-oriented databases work on columns and are based on BigTable paper by Google. Every column is treated separately. Values of single column databases are stored contiguously.

ColumnFamily			
Row Key	Column Name		
	Key	Key	Key
	Value	Value	Value
	Column Name		
	Key	Key	Key
	Value	Value	Value

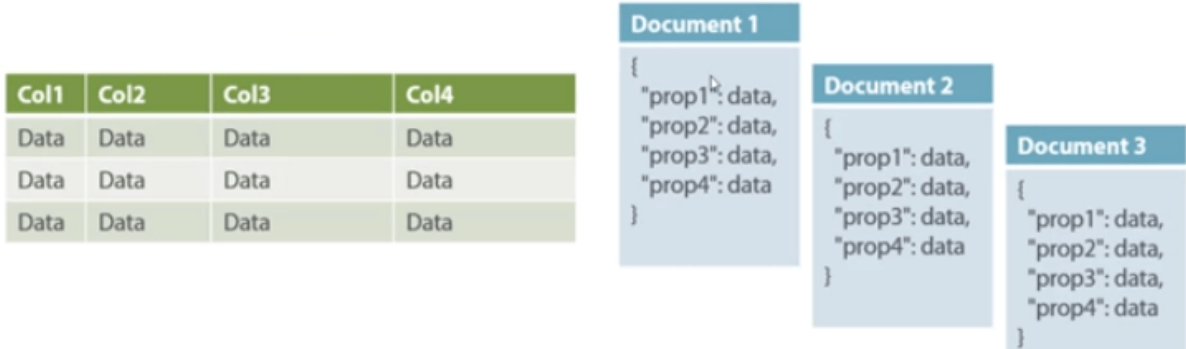
They deliver high performance on aggregation queries like SUM, COUNT, AVG, MIN etc. as the data is readily available in a column.

Column-based NoSQL databases are widely used to manage data warehouses, business intelligence, CRM, Library card catalogs,

HBase, Cassandra, HBase, Hypertable are NoSQL query examples of column based database.

Document-Oriented:

Document-Oriented NoSQL DB stores and retrieves data as a key value pair but the value part is stored as a document. The document is stored in JSON or XML formats. The value is understood by the DB and can be queried.



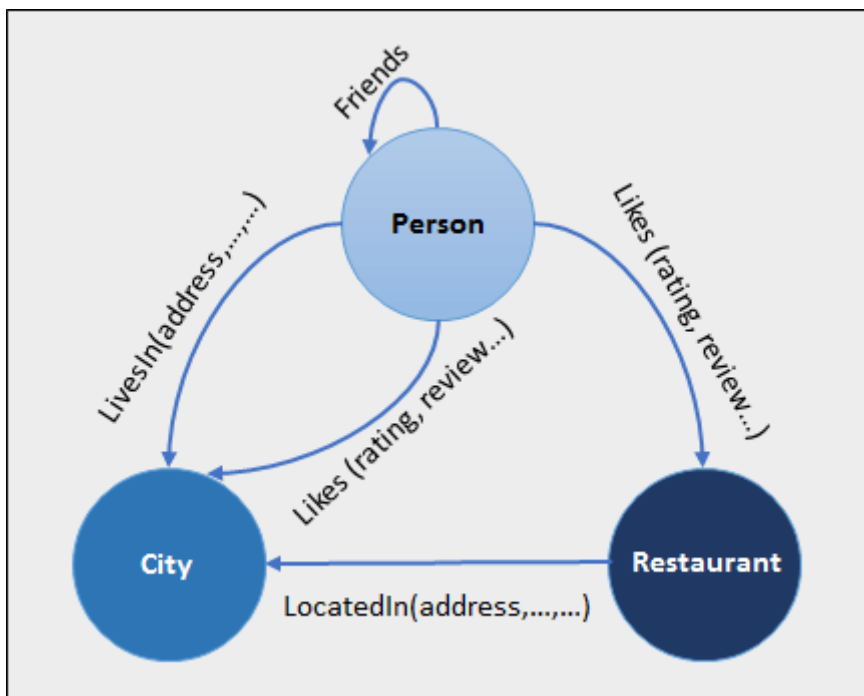
In this diagram on your left you can see we have rows and columns, and in the right, we have a document database which has a similar structure to JSON. Now for the relational database, you have to know what columns you have and so on. However, for a document database, you have data store like JSON object. You do not require to define which make it flexible.

The document type is mostly used for CMS systems, blogging platforms, real-time analytics & e-commerce applications. It should not use for complex transactions which require multiple operations or queries against varying aggregate structures.

Amazon SimpleDB, CouchDB, MongoDB, Riak, Lotus Notes, MongoDB, are popular Document originated DBMS systems.

Graph-Based

A graph type database stores entities as well the relations amongst those entities. The entity is stored as a node with the relationship as edges. An edge gives a relationship between nodes. Every node and edge has a unique identifier.



Compared to a relational database where tables are loosely connected, a Graph database is a multi-relational in nature. Traversing relationship is fast as they are already captured into the DB, and there is no need to calculate them.

Graph base database mostly used for social networks, logistics, spatial data.

Neo4J, Infinite Graph, OrientDB, FlockDB are some popular graph-based databases.

Query Mechanism tools for NoSQL

The most common data retrieval mechanism is the REST-based retrieval of a value based on its key/ID with GET resource

Document store Database offers more difficult queries as they understand the value in a key-value pair. For example, CouchDB allows defining views with MapReduce

What is the CAP Theorem?

CAP theorem is also called brewer's theorem. It states that is impossible for a distributed data store to offer more than two out of three guarantees

Consistency

Availability

Partition Tolerance

Consistency:

The data should remain consistent even after the execution of an operation. This means once data is written, any future read request should contain that data. For example, after updating the order status, all the clients should be able to see the same data.

Availability:

The database should always be available and responsive. It should not have any downtime.

Partition Tolerance:

Partition Tolerance means that the system should continue to function even if the communication among the servers is not stable. For example, the servers can be partitioned into multiple groups which may not communicate with each other. Here, if part of the database is unavailable, other parts are always unaffected.

Eventual Consistency

The term "eventual consistency" means to have copies of data on multiple machines to get high availability and scalability. Thus, changes made to any data item on one machine has to be propagated to other replicas.

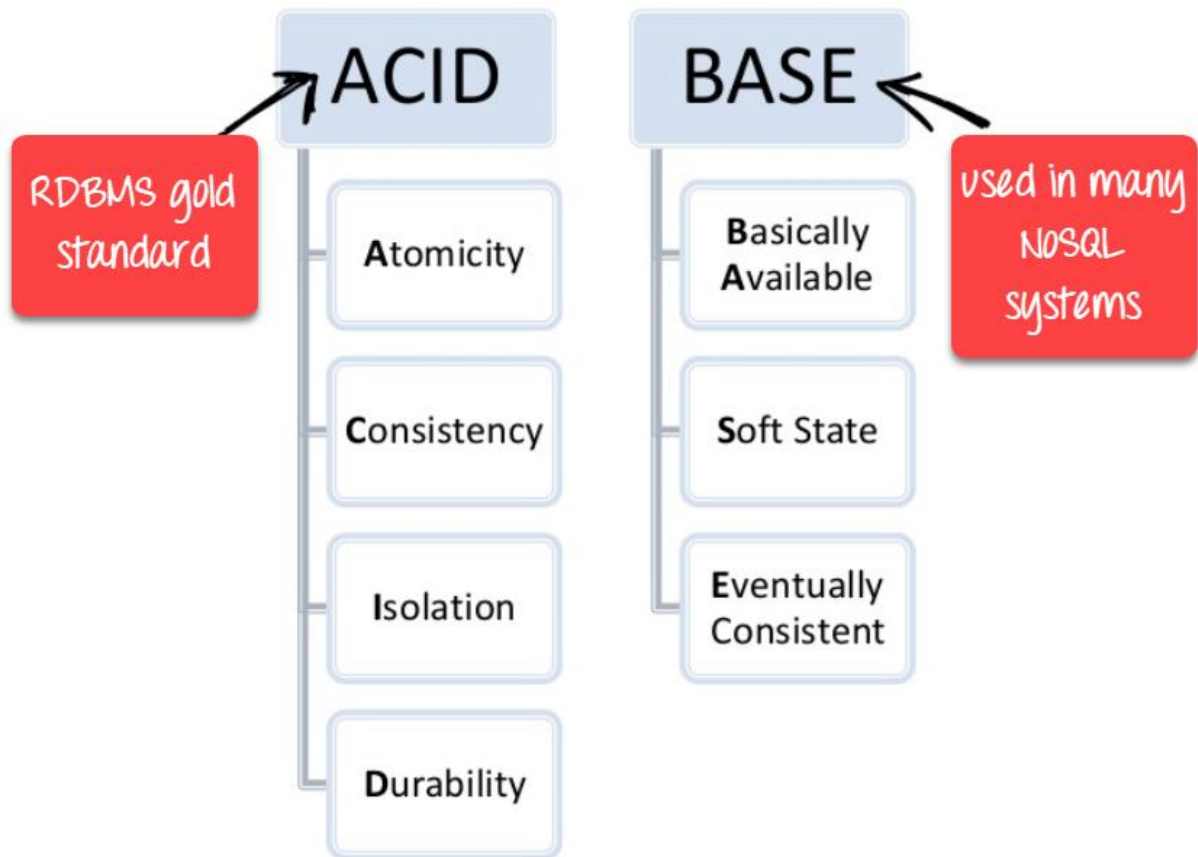
Data replication may not be instantaneous as some copies will be updated immediately while others in due course of time. These copies may be mutually, but in due course of time, they become consistent. Hence, the name eventual consistency.

BASE: Basically Available, Soft state, Eventual consistency

Basically, available means DB is available all the time as per CAP theorem

Soft state means even without an input; the system state may change

Eventual consistency means that the system will become consistent over time



Advantages of NoSQL

Can be used as Primary or Analytic Data Source

Big Data Capability

No Single Point of Failure

Easy Replication

No Need for Separate Caching Layer

It provides fast performance and horizontal scalability.

Can handle structured, semi-structured, and unstructured data with equal effect

Object-oriented programming which is easy to use and flexible

NoSQL databases don't need a dedicated high-performance server

Support Key Developer Languages and Platforms

Simple to implement than using RDBMS

It can serve as the primary data source for online applications.

Handles big data which manages data velocity, variety, volume, and complexity

Excels at distributed database and multi-data center operations

Eliminates the need for a specific caching layer to store data

Offers a flexible schema design which can easily be altered without downtime or service disruption

Disadvantages of NoSQL

No standardization rules

Limited query capabilities

RDBMS databases and tools are comparatively mature

It does not offer any traditional database capabilities, like consistency when multiple transactions are performed simultaneously.

When the volume of data increases it is difficult to maintain unique values as keys become difficult

Doesn't work as well with relational data

The learning curve is stiff for new developers

Open source options so not so popular for enterprises.

Summary

NoSQL is a non-relational DMS, that does not require a fixed schema, avoids joins, and is easy to scale

The concept of NoSQL databases became popular with Internet giants like Google, Facebook, Amazon, etc. who deal with huge volumes of data

In the year 1998- Carlo Strozzi use the term NoSQL for his lightweight, open-source relational database

NoSQL databases never follow the relational model it is either schema-free or has relaxed schemas

Four types of NoSQL Database are 1).Key-value Pair Based 2).Column-oriented Graph 3). Graphs based 4).Document-oriented

NOSQL can handle structured, semi-structured, and unstructured data with equal effect

CAP theorem consists of three words Consistency, Availability, and Partition Tolerance

BASE stands for Basically Available, Soft state, Eventual consistency

The term "eventual consistency" means to have copies of data on multiple machines to get high availability and scalability

NOSQL offer limited query capabilities