# NoSQL Review

Name the categories of NoSQL databases

Name the categories of NoSQL databases

key-value stores

document-based

column NoSQL databases

graphDB

Which of the following is not a reason NoSQL has become a popular solution for some organizations?

A. Better scalability

B. Improved ability to keep data consistent

C. Faster access to data than relational database management systems (RDBMS)

D. More easily allows for data to be held across multiple servers

Which of the following is not a reason NoSQL has become a popular solution for some organizations?

A. Better scalability

B. Improved ability to keep data consistent

C. Faster access to data than relational database management systems (RDBMS)

D. More easily allows for data to be held across multiple servers

When you move from a traditional database to a NoSQL database, which abilities / properties do you lose?

When you move from a traditional database to a NoSQL database, which abilities / properties do you lose?

Referential Integrity Constraint
Ability to join across entire tables (Full table joins)

Strict consistency

At least one of the ACID properties

Ability to use SQL

What are the options in traditional RDBMS for storing Big Data?
What are the problems with each of them?

What are the options in traditional RDBMS for storing Big Data?

Master Slave architecture

- Writes written to master and reads against slaves
- Problems: Reads may be inconsistent, Write propagation, Problems with very large datasets

Sharding

- dividing database horizontally and storing in different nodes
- Problems: Application needs to be partition aware, can no longer join across shards, Loss of referential integrity

What are the options in traditional RDBMS for storing Big Data?
What are the problems with each of them?

What are the options in traditional RDBMS for storing Big Data?

Master Slave architecture

- Writes written to master and reads against slaves
- Problems: Reads may be inconsistent, Write propagation, Problems with very large datasets

Sharding

- dividing database horizontally and storing in different nodes
- Problems: Application needs to be partition aware, can no longer join across shards, Loss of referential integrity

What is the CAP theorem? Where does it apply?

What is the CAP theorem? Where does it apply?

It was proposed by Eric Brewer in 2000. It's applicable to distributed systems and says that out of the following 3:

- Consistency (all nodes see the same data at the same time)
- Availability (a guarantee that every request receives a response about whether it succeeded or failed)
- Partition tolerance (the system continues to operate despite arbitrary partitioning due to network failures)

You can only have 2 at the same time

In case of Big Data, which of the 3 CAP properties is absolutely needed?

In case of Big Data, which of the 3 CAP properties is absolutely needed?

P (Partition tolerance or partitionability)

What is the difference between traditional consistency and the consistency that we can accept for Big Data?

What is the difference between traditional consistency and the consistency that we can accept for Big Data?

Traditional:

strong consistency – ACID(Atomicity Consistency Isolation Durability)

for Big Data:

weak consistency – BASE(Basically Available Soft-state Eventual consistency )

If you were running a very large (~1 bn users) social networking or social data sharing site, which of the 3 CAP properties would you prefer?

If you were running a very large (~1 bn users) social networking or social data sharing site, which of the 3 CAP properties would you prefer?

P + A

with eventual consistency

Name an example of key-value store NoSQL database? What are its advantages and disadvantages?

Name an example of key-value store NoSQL database? What are its advantages and disadvantages?

SimpleDB and Amazon's S3

Pros

- very fast
- very scalable
- simple model
- able to distribute horizontally

*Cons*:

- many data structures (objects) can't be easily modeled as key value pairs

Give an example of a graph NoSQL database

Give an example of a graph NoSQL database

Neo4j