## NoSQL databases

***Explain the basic concepts of NoSQL.***

Next Generation Databases mostly addressing some of the points: being **non-relational, distributed, open source** and **horizontally scalable**. The original intention has been modern web-scale databases. The movement began early 2009 and is growing rapidly. Often more characteristics apply as: **schema-free, easy replication support, simple API, eventually consistent/BASE** (not ACID)**, a huge data amount**, and more.

***Explain the difference of vertical scalability and horizontal scalability.***

Scalability: system can handle growing amounts of data without losing performance.

Vertical Scalability (scale up)1. Add resources (more CPUs, more memory) to a single node

2. Using more threads to handle a local problem

Horizontal Scalability (scale out)

1. Add nodes (more computers, servers) to a distributed system

2. Get more and more popular due to low costs for commodity hardware

3. Often surpasses scalability of vertical approach

***Explain the CAP theorem.***

CAP Theorem: Consistency, Availability, Partition Tolerance

Only 2 of the 3 guarantees can be given in a shared-data system.

Consistency: after an update, all readers in a distributed system see the same data, all nodes are supposed to contain the same data at all times.

Availability: all requests will be answered, regardless of crashes or downtimes.

Partition Tolerance: system continues to operate, even if two sets of servers get isolated

BASE: Basically Available (despite partial failures), Soft State (changes all the time), Eventual Consistency (at some time in future)

***Understand the principles of consistent hashing.***

1. arrange the nodes in a ring and each node is in charge of the hash values in the range between its neighbor node
2. include hash values of all nodes in hash structure
3. calculate hash value of the key to be added/retrieved
4. choose node which occurs next clockwise in the ring
5. if node is dropped or gets lost, missing data is redistributed to adjacent nodes (replication issue)
6. if a new node is added, its hash value is added to the hash table, the hash realm is repartitioned, and hash data will be transferred to new neighbor, no need to update remaining nodes!

***Understand the principles of vector clocks.***

A vector clock for a system of N nodes is an array of N integers

Each process keeps its own vector clock, Vi , which it uses to timestamp local events.

processes piggyback vector timestamps on the messages they send to one another, and there are simple rules for updating the clocks:

***Explain the basic principles and steps of MapReduce.***

1. Sequentially read a lot of data
2. Map: Extract something you care about
3. Group By Key: Sort and Shuffle
4. Reduce: Aggregate, summarize, filter or transform
5. Write the result

***Give an example of applications that benefit from the MapReduce model.***

Matrix-Vector Multiplication