CSL7090
Software & Data Engineering

Lecture #16 Document database, column store, Distributed databases, Consistency

Course Instructors:

Sumit Kalra, Assistant Professor, IIT Jodhpur
Ram Subramaniam, Co-founder - My Money Karma

Google Classroom Code: cezptvv

Document Database

JSON

    Json schema, reusable schemas

REST

    Get, post, put, delete

    Stateless Communication, Cacheable information, uniform interface,

Column Store

A column oriented relational database

Buffer management, Homogeneity, Data locality, Column Insertion

Tuple reconstruction, tuple Insertion

Column Compression

     Run-length encoding, bit-vector encoding, dictionary encoding, frame of reference encoding, differential encoding, null suppression,

Distributed Databases

Scaling Horizontally

Load balancing, flexible scalability, heterogeneous nodes, symmetric configuration, decentralized control

Distribution Transparency

Access transparency, location transparency, replication transparency, migration transparency

Failures

Server failure, message failure, link failure, network partition,

Consistency

Strong Consistency

Weak Consistency
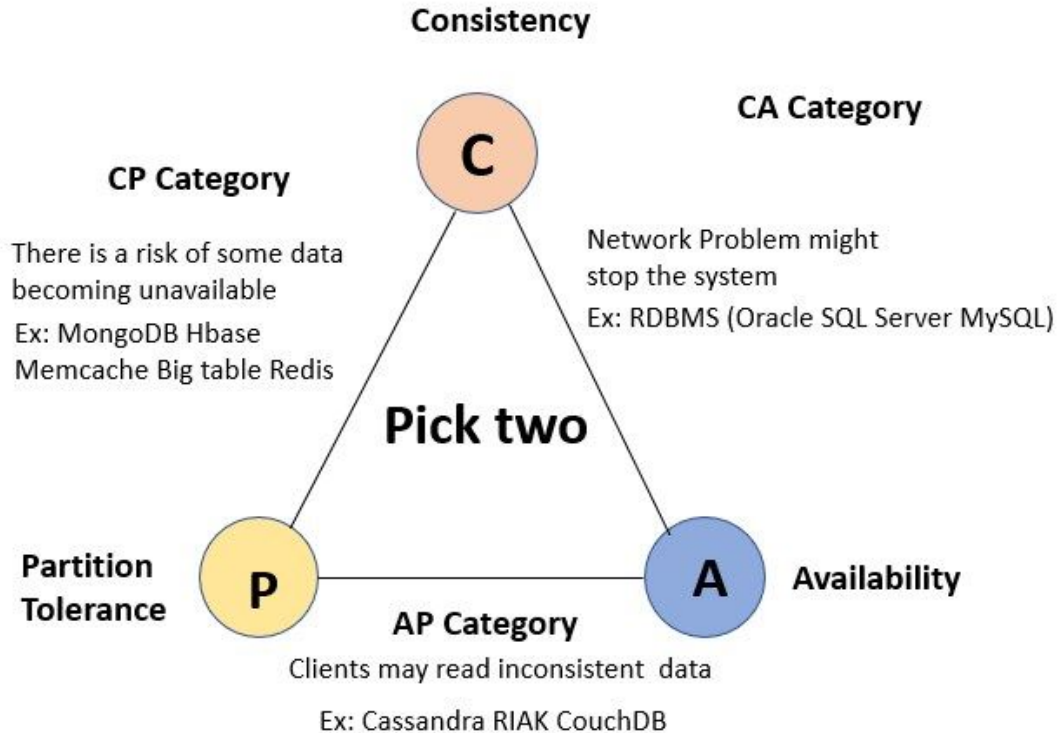
Data Centric Consistency

    Eventual, Causal

Client Centric Consistency

    Read your writes, Monotonic Reads, Writes follows reads, Monotonic writes, Consistent prefix

Consistency Trade-offs (CAP)

CAP Theorem Explained

Consistency

CA Category

CP Category

C

Network Problem might stop the system

Ex: RDBMS (Oracle SQL Server MySQL)

There is a risk of some data becoming unavailable

Ex: MongoDB Hbase Memcache Big table Redis

Pick two

Partition Tolerance

P

A

Availability

AP Category

Clients may read inconsistent data

Ex: Cassandra RIAK CouchDB

www.educba.com

?

Any Questions?

Homework:

Create Google Cloud account

Next Class:
Data Intensive Processing Systems

Contact: sumitk@iitj.ac.in