



INDIAN INSTITUTE OF
INFORMATION
TECHNOLOGY

No-SQL and CAPs Theorem for DSAI

Dr. Animesh Chaturvedi

Assistant Professor: IIIT Dharwad

Young Researcher: Heidelberg Laureate Forum

Postdoc: King's College London & The Alan Turing Institute

PhD: IIT Indore MTech: IIITDM Jabalpur



Indian Institute of Technology Indore
भारतीय प्रौद्योगिकी संस्थान इंदौर



PDPM

Indian Institute of Information Technology,
Design and Manufacturing, Jabalpur



Not Only SQL (No-SQL)

No-SQL

- “Not Only SQL” i.e. database systems that work on other than the tabular relations used in relational databases.
- No-SQL is a non-relational database system, which may not necessarily use SQL.
- No-SQL databases are mostly used for real-time and big data applications.
- No-SQL systems may also support SQL-like query languages.
- No-SQL database systems have data structures different from relational databases (key-value, graph, document).
- No-SQL allows a dynamic schema for unstructured data.
- No-SQL databases allow adding new attributes and fields.

No-SQL Properties

- High usage in big data and real-time web applications.
- Most No-SQL database systems follow the CAP theorem.
- Hurdles in adoption of No-SQL stores are low-level query languages, lack of standardized interfaces, and huge investments in existing SQL.
- Less need to pre-plan and pre-organize data, and it's easier to make modifications.
- Some No-SQL systems do not provide all four ACID properties together (atomicity, consistency, isolation and durability).
- “Horizontal” scaling to make clusters of machines, making operations faster.
 - Means add additional servers or nodes as needed to increase load.
- “Vertical” scaling is mostly necessity of SQL based Relational Database
 - to follow ACID properties.

Categories of No-SQL databases

- **Column Oriented:** data is stored as columns instead of rows
 - data is stored in cells grouped in a virtually unlimited number of columns rather than rows.
 - Accumulo, Cassandra, Druid, HBase, Vertica
- **Document-Oriented:**
 - use documents to hold and encode data in standard formats including XML, YAML, JSON (JavaScript Object Notation) and BSON (Binary JSON).
 - documents within a single database can have different data types
 - Clusterpoint, CouchDB, Couchbase, MarkLogic, MongoDB, OrientDB
- **Key-value:** contains many different key value pairs
 - use an associative array (also known as a dictionary or map)
 - Dynamo, FoundationDB, MemcacheDB, Redis, Riak, FairCom ctreeACE, Aerospike, OrientDB
- **Graph:** used to store data related to connections or networks
 - represent data on a graph that shows how different sets of data relate to each other
 - Allegro, Neo4J, InfiniteGraph, OrientDB, Virtuoso, Stardog, RedisGraph
- **Multi-model:** OrientDB, FoundationDB, ArangoDB, Alchemy Database, CortexDB.

CAPS Theorem

CAPS Theorem

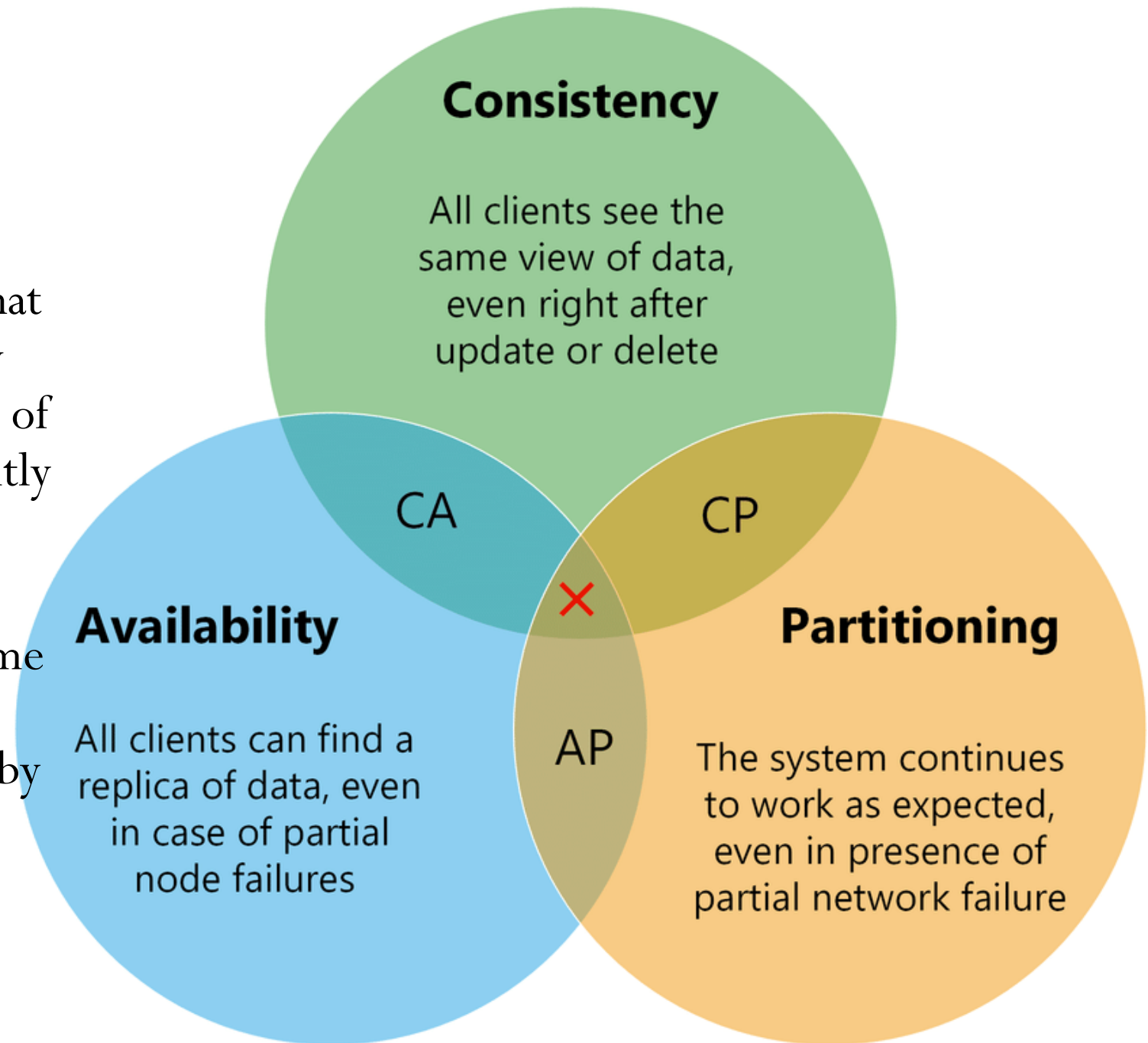
- Also known as Brewer's theorem
- SQL follows ACID properties,
- No-SQL follows the CAP theorem
- The CAP theorem says that
- “It is impossible for a distributed computer system to simultaneously provide three (Consistency, Availability, and Partition) together with guarantees in single instance”
- Although some No-SQL databases — such as IBM’s DB2, MongoDB, AWS’s DynamoDB and Apache’s CouchDB — can also integrate and follow ACID rules

CAPS Theorem

- A distributed data systems allow a trade-off that can guarantee only two of the following three properties (which form the acronym CAP) at any one time:
- **Consistency:** All nodes can view the same data at the same time. Every request receives either the most recent result or an error. MongoDB is an example of a strongly consistent system, whereas others such as Cassandra offer eventual consistency.
- **Availability:** A guarantee that every request will receive a response for success or failure. Every request has a non-error result.
- **Partition:** The system continues to operate irrespective of the loss or failure of a node. Trying to achieve Partition tolerance means any delays or losses between nodes do not interrupt the system operation.

CAPS Theorem

- One such fallacy of distributed computing is that networks are reliable. They aren't. Networks and parts of networks go down frequently and unexpectedly.
- A partition is when the network fails to deliver some messages to one or more nodes by losing them (not by delaying them - eventual delivery is not a partition).



Reference

- Avi Silberschatz, Henry F. Korth, S. Sudarshan, Database System Concepts, Seventh Edition, McGraw-Hill, ISBN 9780078022159
- Wikipedia <https://en.wikipedia.org/wiki/NoSQL>
- Robert Sheldon, 13 April 2021 <https://www.red-gate.com/simple-talk/databases/nosql/how-to-choose-between-sql-and-nosql-databases/>
- Benjamin Anderson “STSM, IBM Cloud Databases”, Brad Nicholson “Senior Database Engineer, IBM Cloud Databases”,
<https://www.ibm.com/cloud/blog/sql-vs-nosql>

תודה רבה

Hebrew

Ευχαριστώ

Greek

Спасибо

Russian

Danke

German

Merci

French

धन्यवादः

Sanskrit

நன்றி

Tamil

شكراً

Arabic

ಧನ್ಯವಾದಗಳು

Kannada

Thank You

English

നന്ദി

Malayalam

Grazie

Italian

ధన్యవాదాలు

Telugu

આભાર

Gujarati

多謝

Traditional Chinese

Gracias

Spanish

ਧੰਨਵਾਦ

Punjabi

धन्यवाद

Hindi & Marathi

多谢

Simplified Chinese

<https://sites.google.com/site/animeshchaturvedi07>

Obrigado

Portuguese

ありがとうございました

Japanese

ขอบคุณ

Thai

감사합니다

Korean