SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Artificial Intelligence and Data Science

Course Name – 19AD501 Big Data Analytics

III Year / V Semester

Unit 5 – Big Data Database

Topic – CAP Theorem

CAP Theorem

- The CAP theorem, or Brewer's theorem, is a fundamental theorem within the field of system design.

- The CAP theorem states that a distributed system can only provide two of three properties simultaneously: **consistency, availability, and partition tolerance**.

- The theorem formalizes the tradeoff between consistency and availability when there's a partition
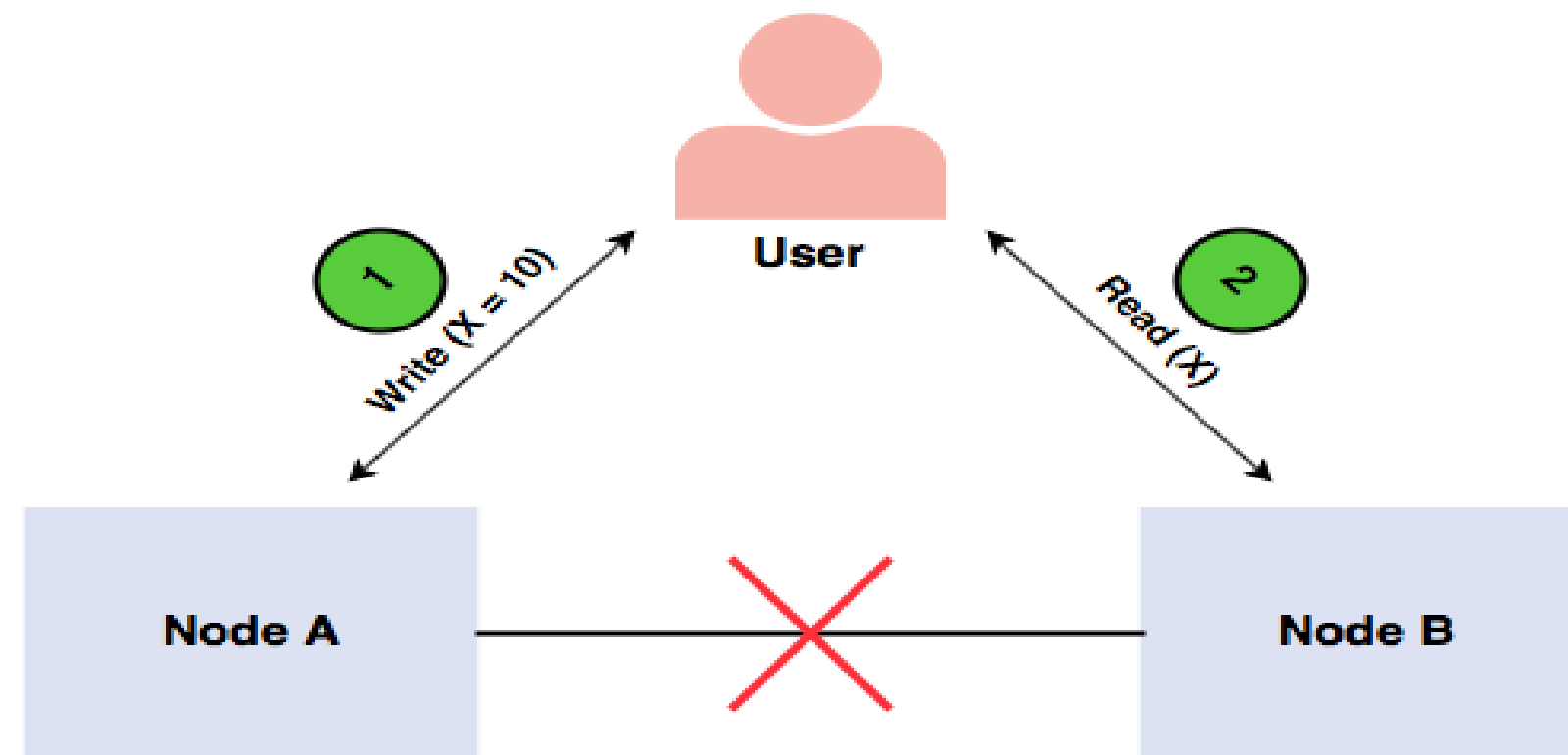
CAP Theorem

Consistency

- In a consistent system, all nodes see the same data simultaneously.

- If we perform a read operation on a consistent system, it should return the value of the most recent write operation.

- The read should cause all nodes to return the same data.

- All users see the same data at the same time, regardless of the node they connect to.

- When data is written to a single node, it is then replicated across the other nodes in the system.
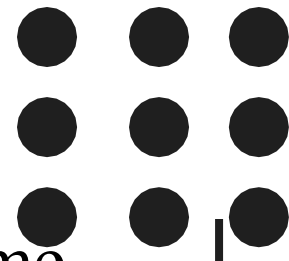
Availability

When availability is present in a distributed system, it means that the system remains operational all of the time.
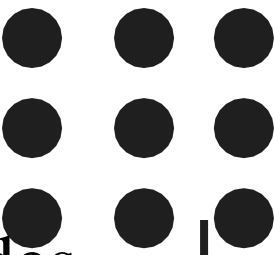
Every request will get a response regardless of the individual state of the nodes.

This means that the system will operate even if there are multiple nodes down.

Unlike a consistent system, there's no guarantee that the response will be the most recent write operation.

CAP Theorem

Partition tolerance

When a distributed system encounters a partition, it means that there's a break in communication between nodes.

If a system is partition-tolerant, the system does not fail, regardless of whether messages are dropped or delayed between nodes within the system.

To have partition tolerance, the system must replicate records across combinations of nodes and networks.
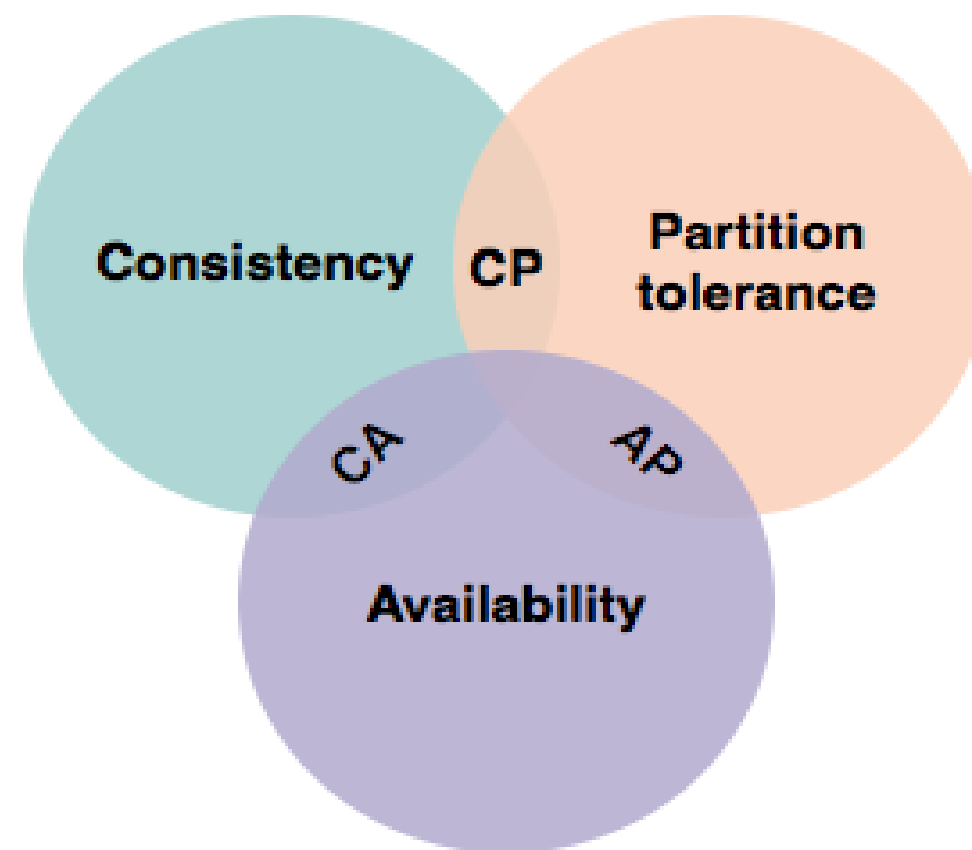
CAP Theorem

CAP theorem NoSQL databases

NoSQL databases are great for distributed networks.

They allow for horizontal scaling, and they can quickly scale across multiple nodes.

When deciding which NoSQL database to use, it's important to keep the CAP theorem in mind.

NoSQL databases can be classified based on the two CAP features they support

CAP Theorem

Consistency and Availability databases

CA databases enable consistency and availability across all nodes.

Unfortunately, CA databases can't deliver fault tolerance.

In any distributed system, partitions are bound to happen, which means this type of database isn't a very practical choice.

Some relational databases, such as **PostgreSQL**, **MySQL** etc allow for consistency and availability.

CAP Theorem

Consistency and Partition tolerance databases

CP databases enable consistency and partition tolerance, but not availability.

When a partition occurs, the system has to turn off inconsistent nodes until the partition can be fixed.

Example - MongoDB, HBase,Redis, BigTable

CAP Theorem

Availability Partition tolerance databases

AP databases enable availability and partition tolerance, but not consistency.

In the event of a partition, all nodes are available, but they're not all updated.

For example, if a user tries to access data from a bad node, they won't receive the most up-to-date version of the data.

Example-Riak, Cassandra, CouchDB, DynamoDB

THANK YOU