NetApp Instaclustr

The CAP Theorem With Apache Cassandra® and MongoDB

August 03, 2021 | By **Lewis DiFelice**

MongoDB and Apache Cassandra are both popular NoSQL distributed database systems. In this article, I will review how the CAP and PACELC theorems classify these systems. I will then show how both systems can be configured to deviate from their classifications in production environments.

The CAP Theorem

The CAP theorem states that a distributed sys n ired properties: consistency, availability, and partition t

Consistency (C): Every client sees the same data n the st recent write.

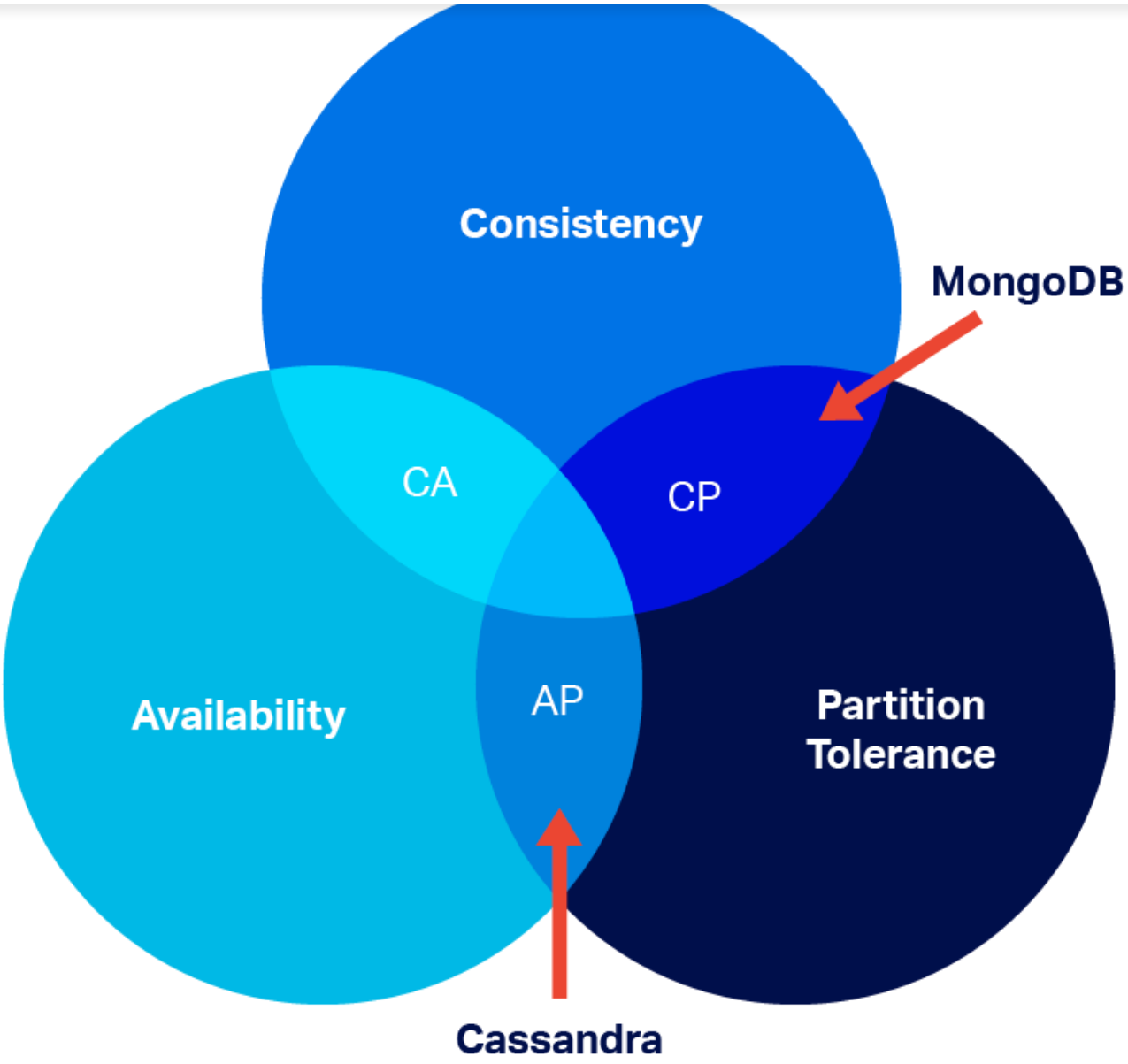
Partition Tolerance (P): The system continues to operate despite one or more breaks in inter-node communication caused by a network or node failure.

Because a distributed system must be partition tolerant, the only choices are deciding between availability and consistency. If part of a cluster becomes unavailable, a system will either:

Safeguard data consistency by canceling the request even if it decreases the availability of the system. Such systems are called CP systems.

Provide availability even though inconsistent data may be returned. These systems are AP distributed systems.

NetApp Instaclustr



According to the CAP theorem, MongoDB is a CP system and Cassandra is an AP system.

CAP theorem provides an overly simplified view of today's distributed systems such as MongoDB and Cassandra. Under normal operations, availability and consistency are adjustable and can be configured to meet specific requirements. However, in keeping with CAP, increasing one state decreases the other. Hence, it would be more correct to describe the default behavior of MongDB or Cassandra as CP or AP. This is discussed in more detail below.

PACELC Theorem

1. It considers the behavior of a distributed system only during a failure condition (the network partition)
2. It fails to consider that in normal operations, there is always a tradeoff between consistency and latency

PACELC is summarized as follows: In the event of a partition failure, a distributed system must choose between Availability (A) and Consistency, else (E) when running normally it must choose between latency (L) or consistency (C).

MongoDB is classified as a PC+EC system. During normal operations and during partition failures, it emphasizes consistency. Cassandra is a PA+EL system. During a partition failure it favors availability. Under normal operations, Cassandra gives up consistency for lower latency. However, like CAP, PACELC describes a default behavior

(As an aside, there are no distributed systems that are AC or PC+EC. These categories describe stand-alone ACID-compliant relational database management systems).

Apache Cassandra vs. MongoDB Architectures

MongoDB is a NoSQL document database. It is a single-master distributed system that uses asynchronous replication to distribute multiple copies of the data for high availability. A MongoDB is a group of instances running mongod and maintaining the same data. The MongoDB documentation refers to this grouping as a **replica set.** But, for simplicity, I will use the MongoDB cluster.

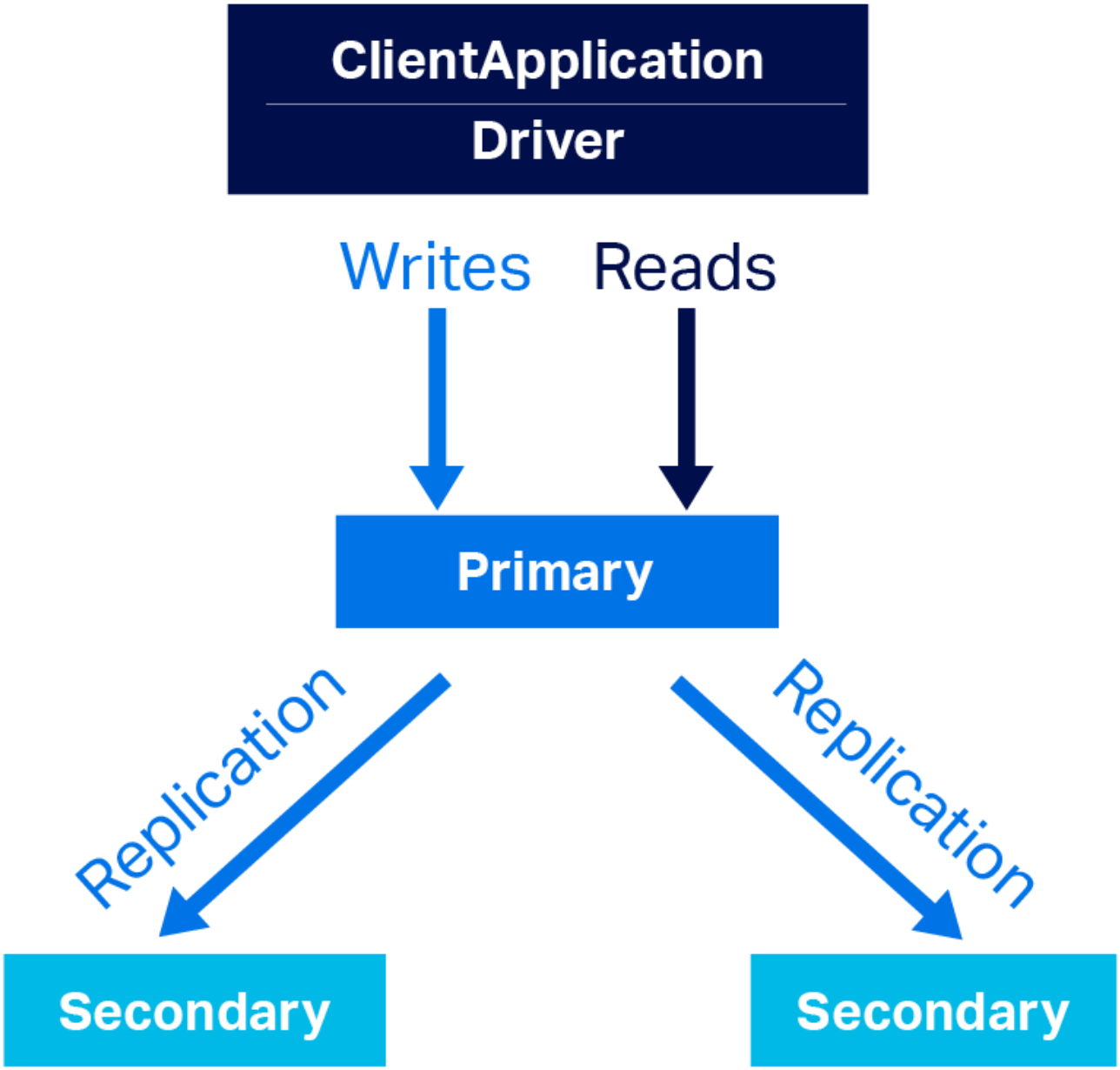
A MongoDB cluster is composed of two types of data-bearing members:

Primary: The primary is the master node and receives all write operations.

Secondaries: The secondaries receive replicated data from the primary to maintain an identical data set.

By default, the primary member handles all reads and writes. Optionally, a MongoDB client can route some or all reads to the secondary members. Writes must be sent to the primary.

If the primary member fails, all writes are suspended until a new primary is selected from one of the secondary members. According to the MongoDB documentation, this process requires up to 12 seconds to complete.

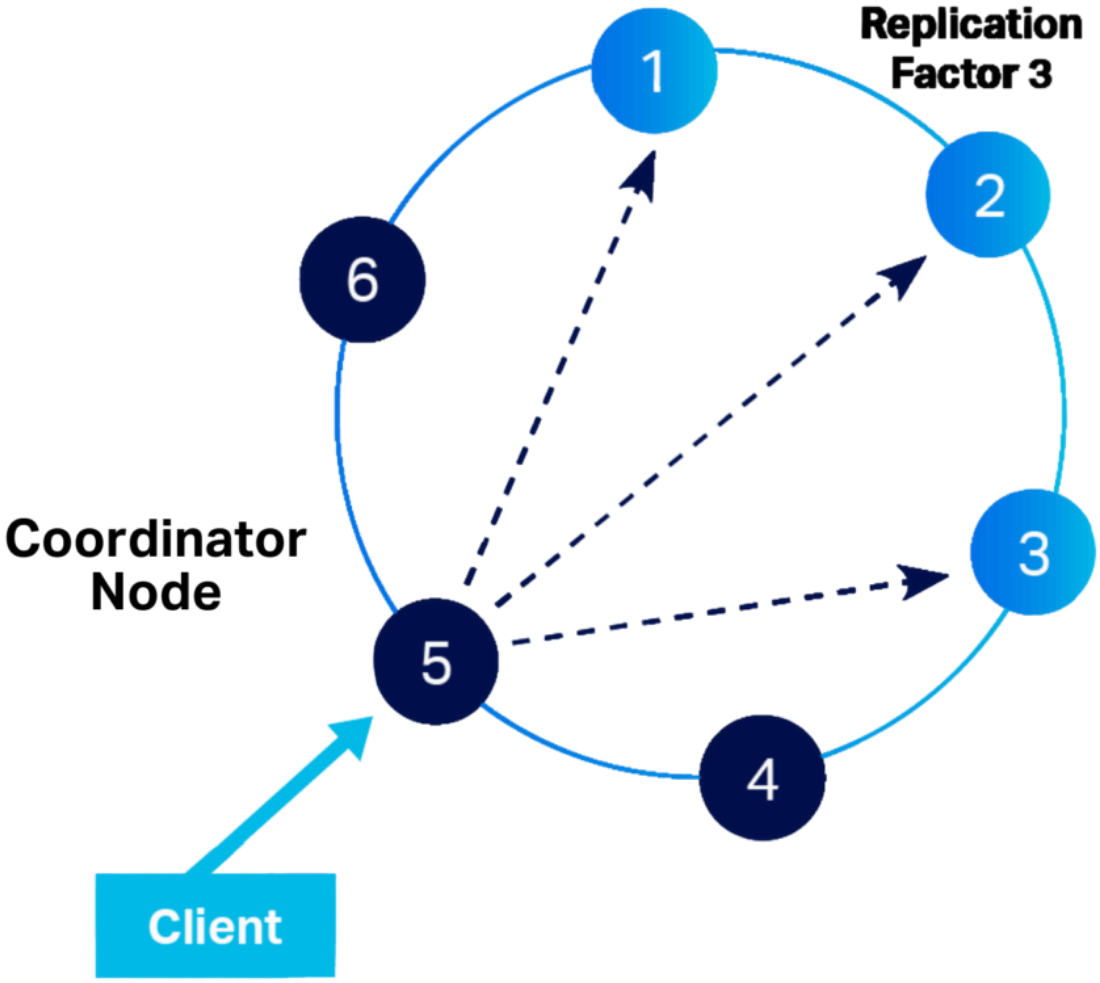


A MongoDB cluster.

A Cassandra cluster is a collection of instances, called nodes, connected in a peer-to-peer "share nothing" distributed architecture. There is no master node and every node can perform all database operations and each can serve client requests. Data is partitioned across nodes based on a consistent hash of its partitioning key. A partition has one or many rows and each node may have one or more partitions. However, a partition can reside only on one node.

NetApp Instaclustr

The node that first receives a request from a client is the coordinator. It is the job of the coordinator to forward the request to the nodes holding the data for that request and to send the results back to the coordinator. Any node in the cluster can act as a coordinator.



Cassandra cluster showing coordinator.