Database Architecture

OceanBase Overview

OceanBase is a distributed SQL database that is designed to be highly scalable, high in performance, and fault-tolerant.

- Database Management System (DBMS)
- supports both hybrid transactional and analytical processing (HTAP) with a single architecture
- atomicity, consistency, isolation, and durability (ACID) capability in both centralized and distributed modes
- Paxos protocol-based multi-replica synchronization algorithm guarantees consistency among multiple data replicas
- compatible with MySQL, easy to migrate from a MySQL database to an OceanBase database
- uses code-based compression to significantly reduce storage space required
- ranks No.1 in the TPC-C benchmark test with a performance result of 707 million tpmC
- ranks No.2 in the TPC-H benchmark test with a performance result of 15 million QphH@30,000 GB

Source: https://en.oceanbase.com/docs/common-oceanbase-database-10000000000829639

The architecture of OceanBase is based on the following **key concepts**:

- **Integrated Architecture** – considers both standalone and distributed features. This means that each layer (SQL layer, transaction layer, storage layer) is designed to consume no extra overhead in both standalone and distributed scenarios.

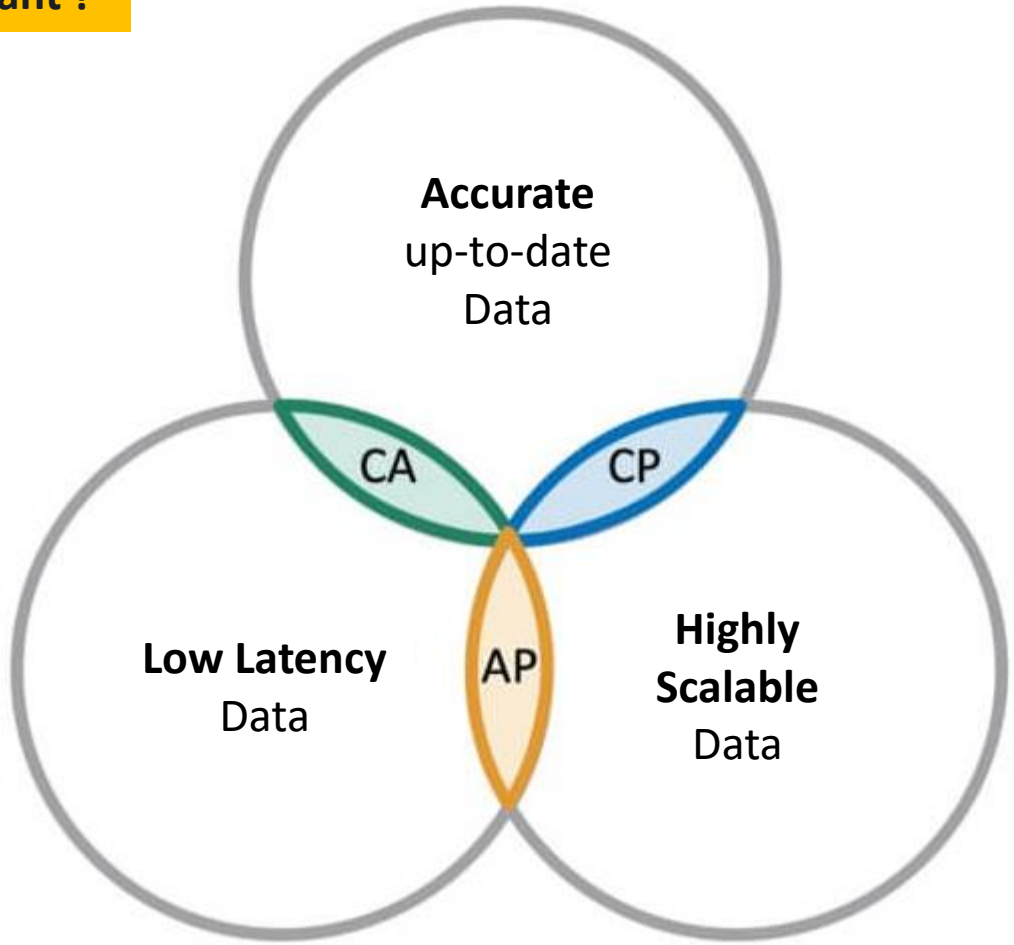
- **Shared-Nothing System Architecture** – the shared-nothing architecture means each server node has its own SQL execution engine and storage engine. The storage engine only accesses the local data on that node, while the SQL engine accesses the global schema and generates the distributed query plan. Query executors visit the storage engine of each node to distribute and gather data among them to execute the query.

- **Replication and Partitioning** - stores replicas of each partition on at least three server nodes in different server clusters. This replication strategy ensures high availability and fault tolerance. The database adopts the Paxos Consensus algorithm to achieve continuous availability with zero recovery point objective (RPO) and less than 8 seconds of recovery time objective (RTO).

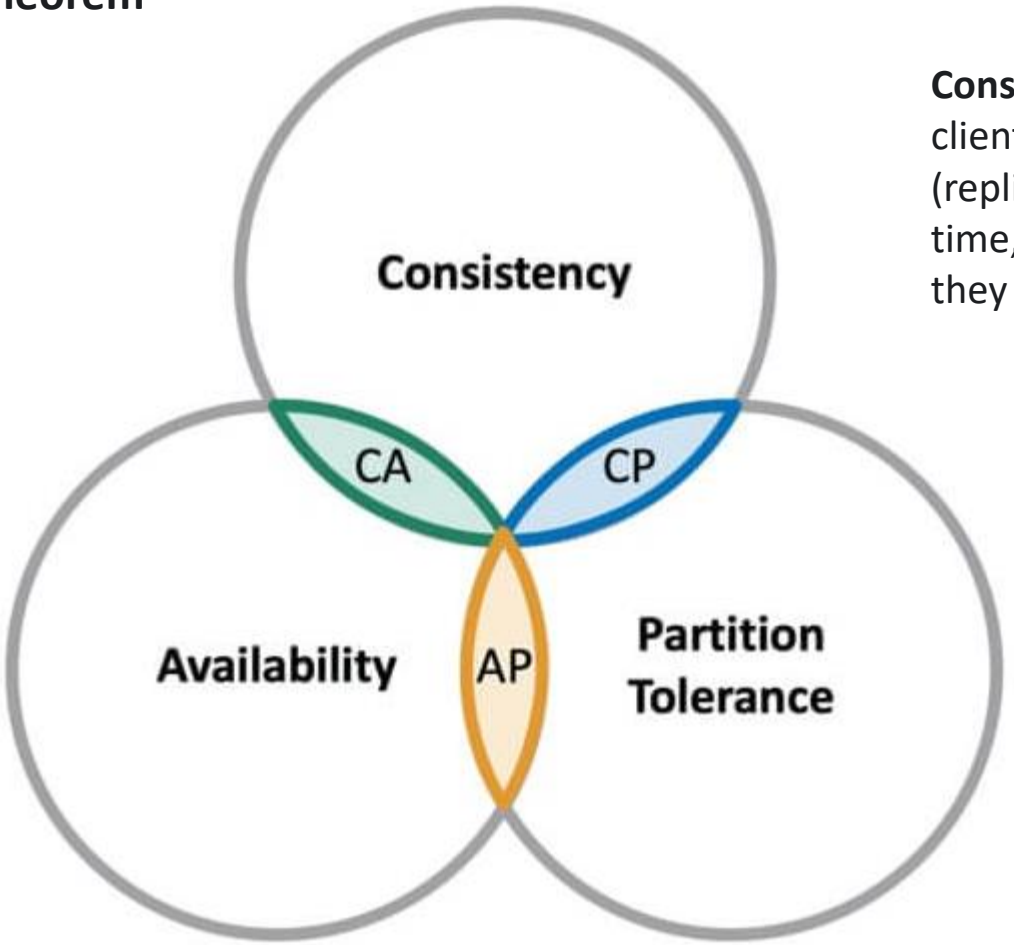
- **MVCC Concurrency Control** - uses Multi-Version Concurrency Control (MVCC) to handle concurrency control. When an operation involves single or multiple partitions on a single server node, it reads the snapshot of that server node. If the operation involves partitions on multiple server nodes, it executes distributed snapshot read.

- **Compatibility and Scalability** - highly compatible with MySQL and supports the relational data model. It is designed to be compatible with common MySQL/Oracle features and protocols, making it easier to migrate businesses from MySQL/Oracle databases to OceanBase. The database also provides high availability, high performance, online scaling, and low costs.

What do customers want ?



Databases in the CAP theorem



Consistency means that all clients see the same (replicated) data at the same time, no matter which node they connect to.

Availability means that that any client making a request for data gets a response, even if one or more nodes are down.

***Partition tolerance** means that the cluster must continue to work despite any number of communication breakdowns between nodes in the system.

*A partition is a communications break within a distributed system, a lost or temporarily delayed connection between two nodes.

Source: https://learning-notes.kovacevic.dev/databases/theory/cap-theorem

1990s

Problem: (physical)
Reading skipped upon hardware impact

Solution: (software)
Read audio data and play from memory

Create a data buffer and feed audio data to RAM within the player, so that audio will not be disrupted while the disk cannot be read due to movement

More details: https://en.wikipedia.org/wiki/Electronic_skip_protection

OceanBase *implements a self-developed storage engine* without resorting to other existing open-source solutions

- **standalone engine** combines some technologies of traditional relational databases and in-memory databases to achieve ultimate performance.
- **distributed engine** implements distributed features (such as linear expansion, Paxos replication, and distributed transactions) to achieve continuous availability.
- not dependent on specific hardware architectures and features cutting-edge compression technology for cost efficiency.
- When merging baseline data with incremental data, OceanBase is optimized with **incremental merging, progressive merging, parallel merging, rotation merging, and I/O isolation**.
- optimization technologies for in-memory databases (including multi-version concurrency and lock-free data structures) to achieve the lowest latency and highest performance.

Source: https://www.alibabacloud.com/blog/an-interpretation-of-the-source-code-of-oceanbase-6-detailed-explanation-of-storage-engine_599324

	OceanBase	MySQL / MariaDB	Oracle SQL
Type of database	**Licensed** **distributed relational database with full SQL support	**Open-source** relational database management system	**Licensed** relational database management system
Developer	Developed by Ant Group (Alibaba Group) 2010	Developed by Michael Widenius 1995 / 2009	Developed by Oracle Corporation 1977
Architecture	Built on a common server cluster with a shared-nothing system architecture	Supports replication and clustering for high availability and fault tolerance	Supports clustering and replication for high availability and fault tolerance
Compatibility	Highly compatible with MySQL and supports the relational data model	Supports the relational data model	Supports the relational data model
Fault tolerance	Stores replicas of each partition on at least three server nodes in different server clusters for fault tolerance	Supports replication and clustering for fault tolerance in a **standalone** database	*Supports clustering and replication for fault tolerance, supports a **distributed** database
Concurrency control	Adopts MVCC concurrency control	Not specified	Not specified
Compression technology	Features cutting-edge compression technology for cost efficiency	Not specified	Not specified
Features	Paxos Consensus algorithm for continuous availability with zero recovery point objective (RPO) and less than 8 seconds of recovery time objective (RTO)	Offers a variety of storage engines to optimize performance and scalability	Offers advanced security features and tools for data protection
Users	For enterprise deployments	For small to medium-sized businesses	Widely used in enterprise environments

*Source: https://www.integrate.io/blog/oracle-vs-mysql

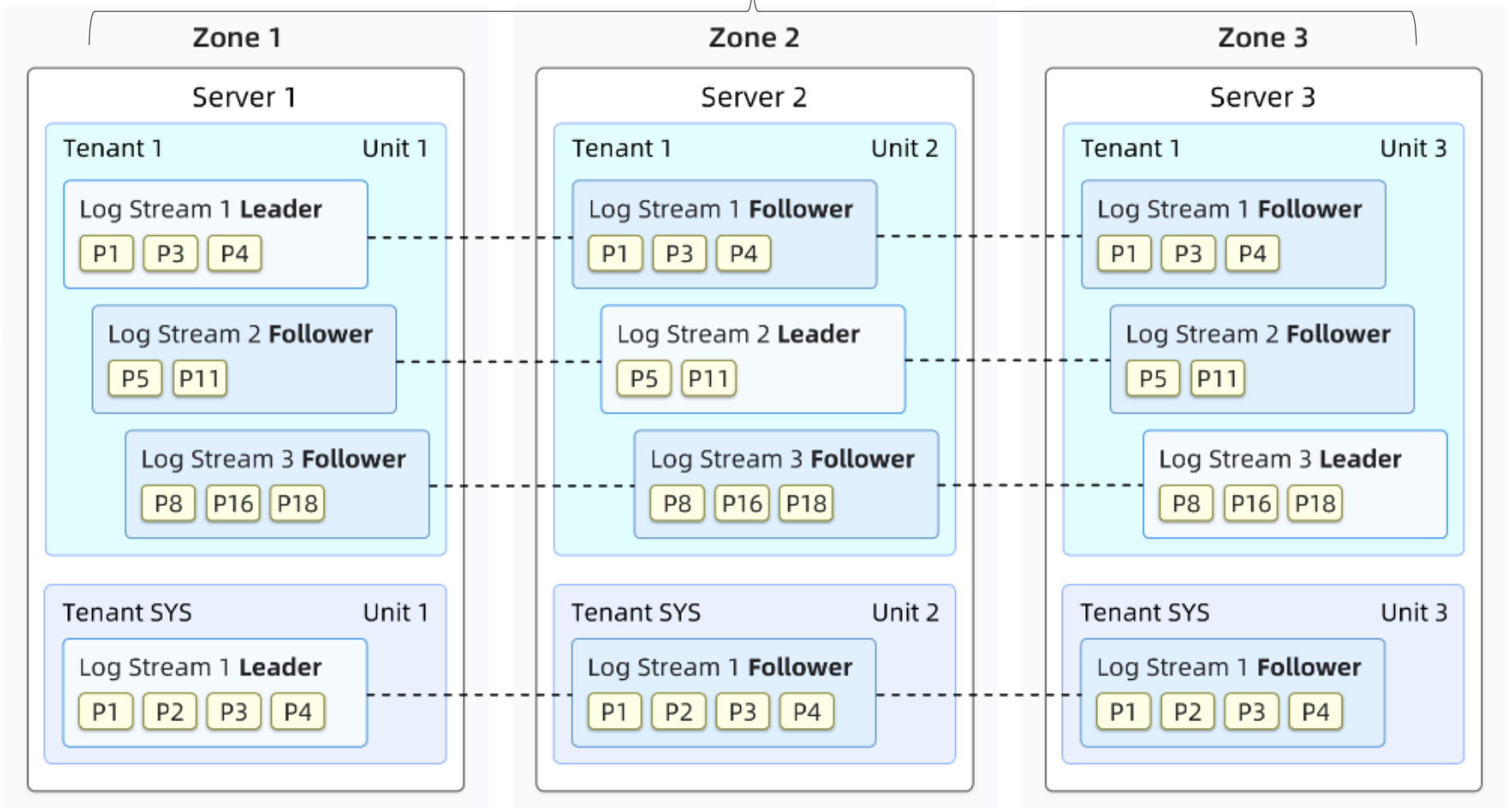
**Source: https://db-engines.com/en/system/OceanBase%3BOracle%3BSAP+Adaptive+Server

OceanBase's architecture is designed to provide a high **performance**, highly **scalable** distributed SQL database that is **easy to use** and compatible with existing **MySQL/Oracle** databases.

Library
(OceanBase **database**)

Librarians
(**datacentres**)

Books
(**data**)




THANK
YOU !