Database Replication

Data replication is copying data from one database server to another database server.

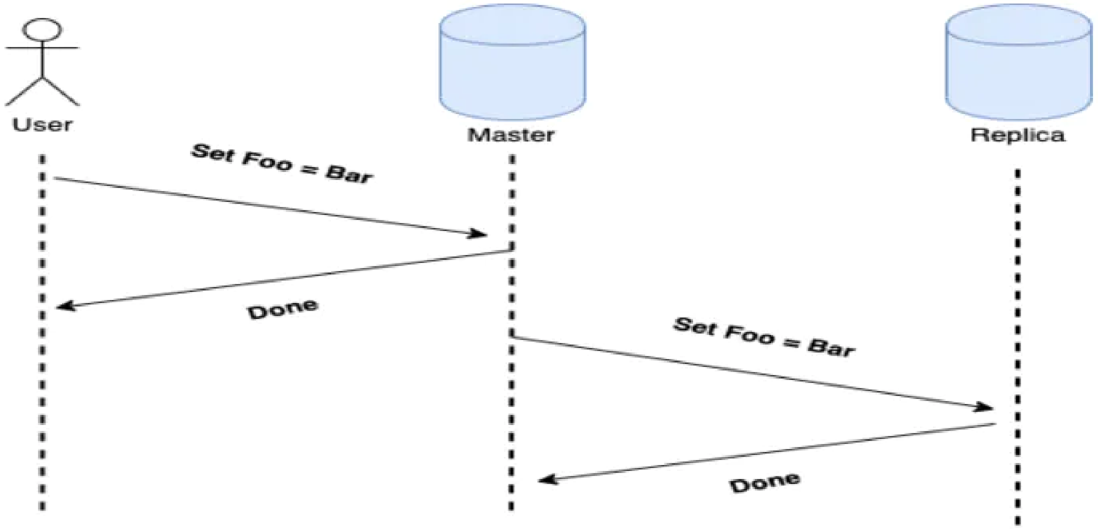
When data is replicated between database servers, so that the information remains consistent throughout the database system and users cannot tell or even know which server in the DBMS they are using, the system is said to exhibit **replication transparency**.

Database replication becomes more complex when it scales up horizontally and vertically. Horizontal scale-up has more data replicas, while vertical scale-up has data replicas located at greater physical distances.

Categorized by the timing of data transfer

Asynchronous(or lazy) replication

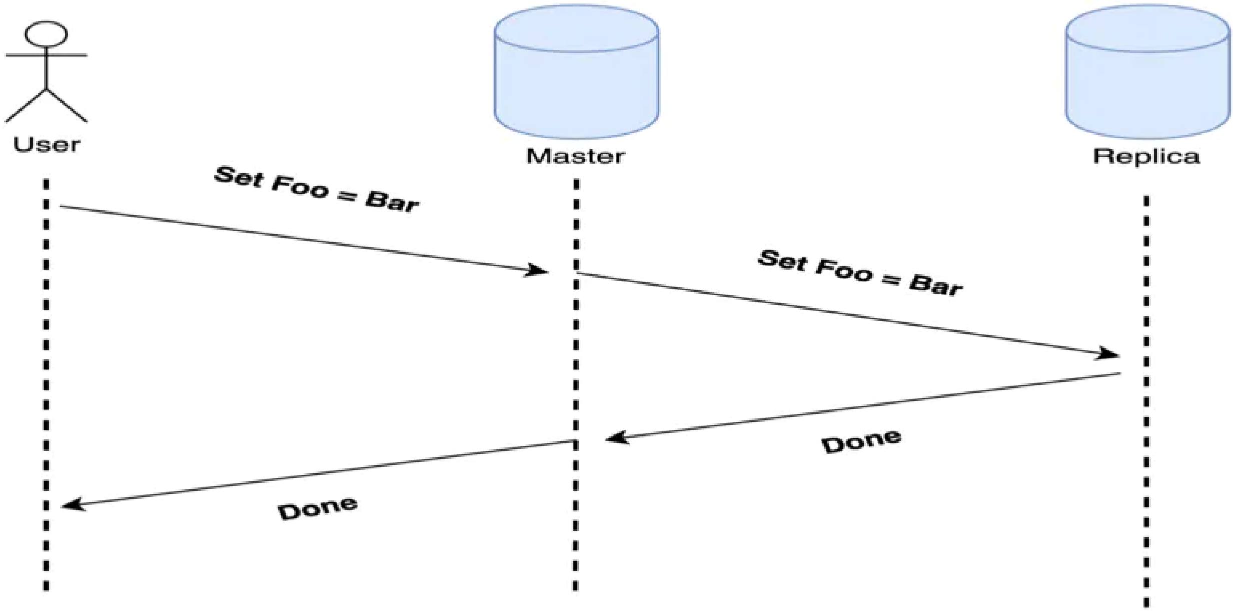
Asynchronous replication is when the data is sent to the master server from the client. Then, the master server pings the client with a confirmation saying the data has been received. From there, it goes about copying data to the replicas at an unspecified or monitored pace.



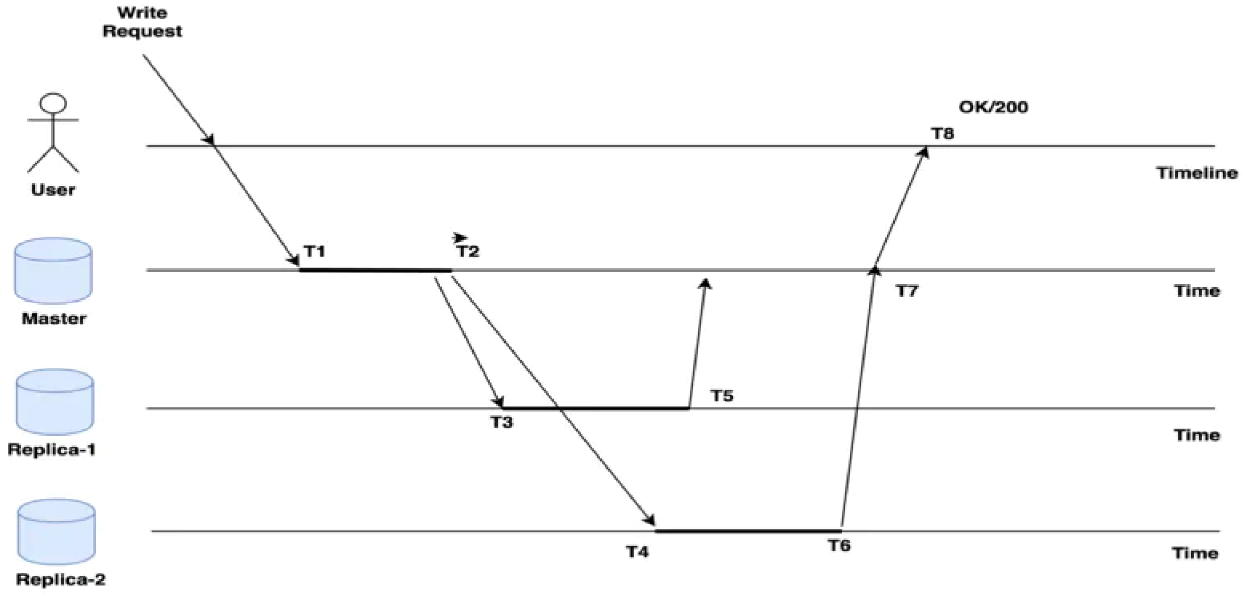
There are two problems with this method one is replication lag and another one is we are weakening our durability guarantees.

Synchronous(or eager) replication

Synchronous replication is when data is copied from the client to the master server and then replicated to all the replica servers **before the client is notified** that data has been replicated.



This takes longer to verify than the asynchronous method, but it presents the advantage of knowing that all data was copied before proceeding. It is performed when reliable and long-term storage is necessary and the business cannot afford to lose any critical data.



Semi-synchronous replication

We can define some replicas to replicate synchronously and others just use asynchronous manner.

Categorized by type of server architecture

Single-Master replication

Master-slave replication enables data from one database server (the **master**) to be replicated to one or more other database servers (the **slaves**). The master **logs** the updates, which then ripple through to the slaves. The slave outputs a message stating that it has received the update successfully, thus allowing the sending of subsequent updates.

Master-slave replication can be **either synchronous or asynchronous**. The difference is simply the timing of the propagation of changes. If the changes are made to the master and slave at the same time, it is synchronous. If changes are queued up and written later, it is asynchronous.

Multi-Master replication

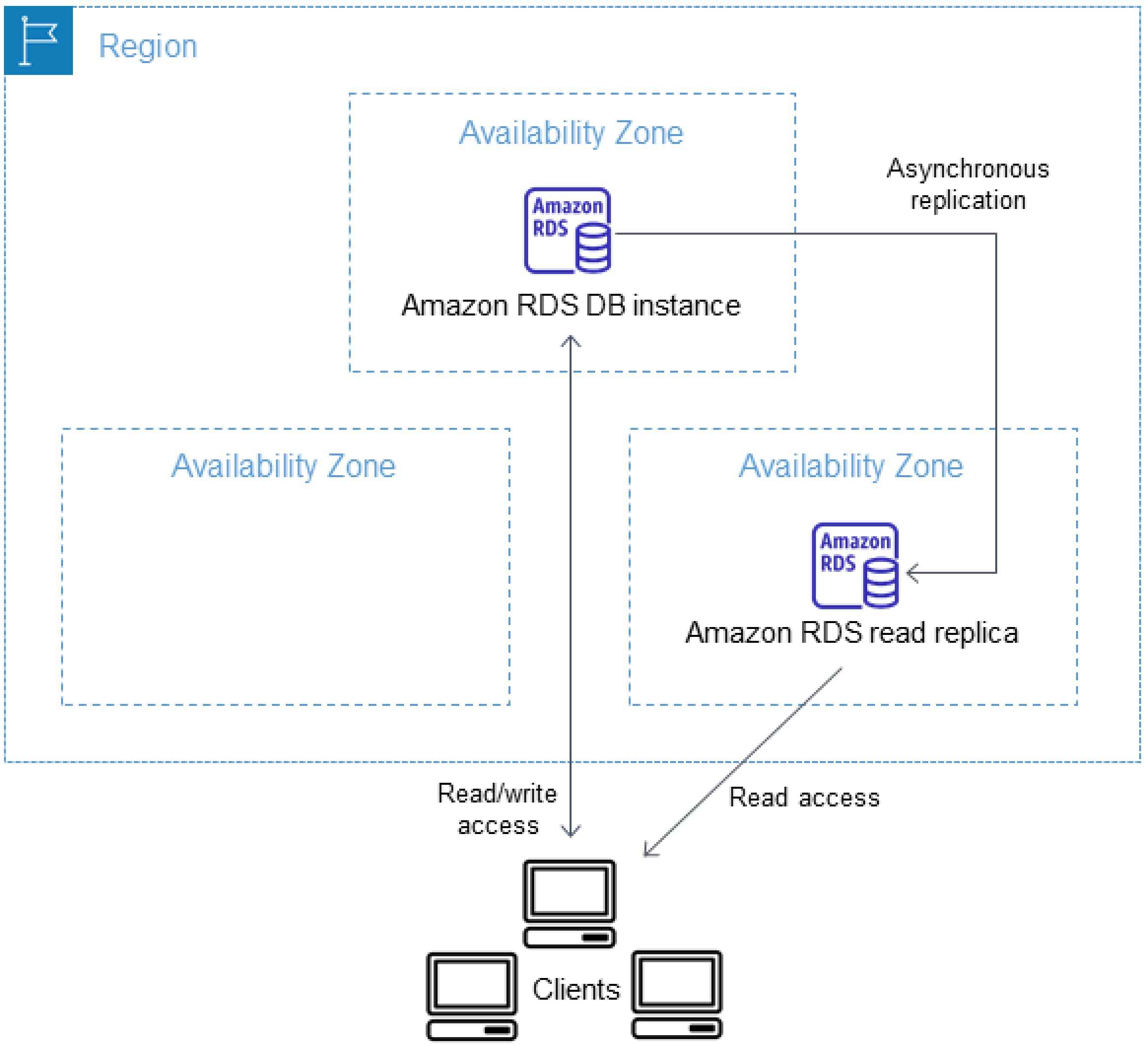
In multi-master replication, updates can be submitted to any database node, and then ripple through to other servers.

The most common challenge that exists in multi-master replication is **transactional conflict prevention or resolution**. Most synchronous (or eager) replication solutions perform conflict prevention, while asynchronous (or lazy) solutions have to perform conflict resolution. For instance, if the same record is changed on two nodes simultaneously, an eager replication system would detect the conflict before confirming the commit and aborting one of the transactions. A lazy replication system would allow both transactions to commit and run a conflict resolution during re-synchronization. The resolution of such a conflict may be based on a **timestamp** of the transaction.

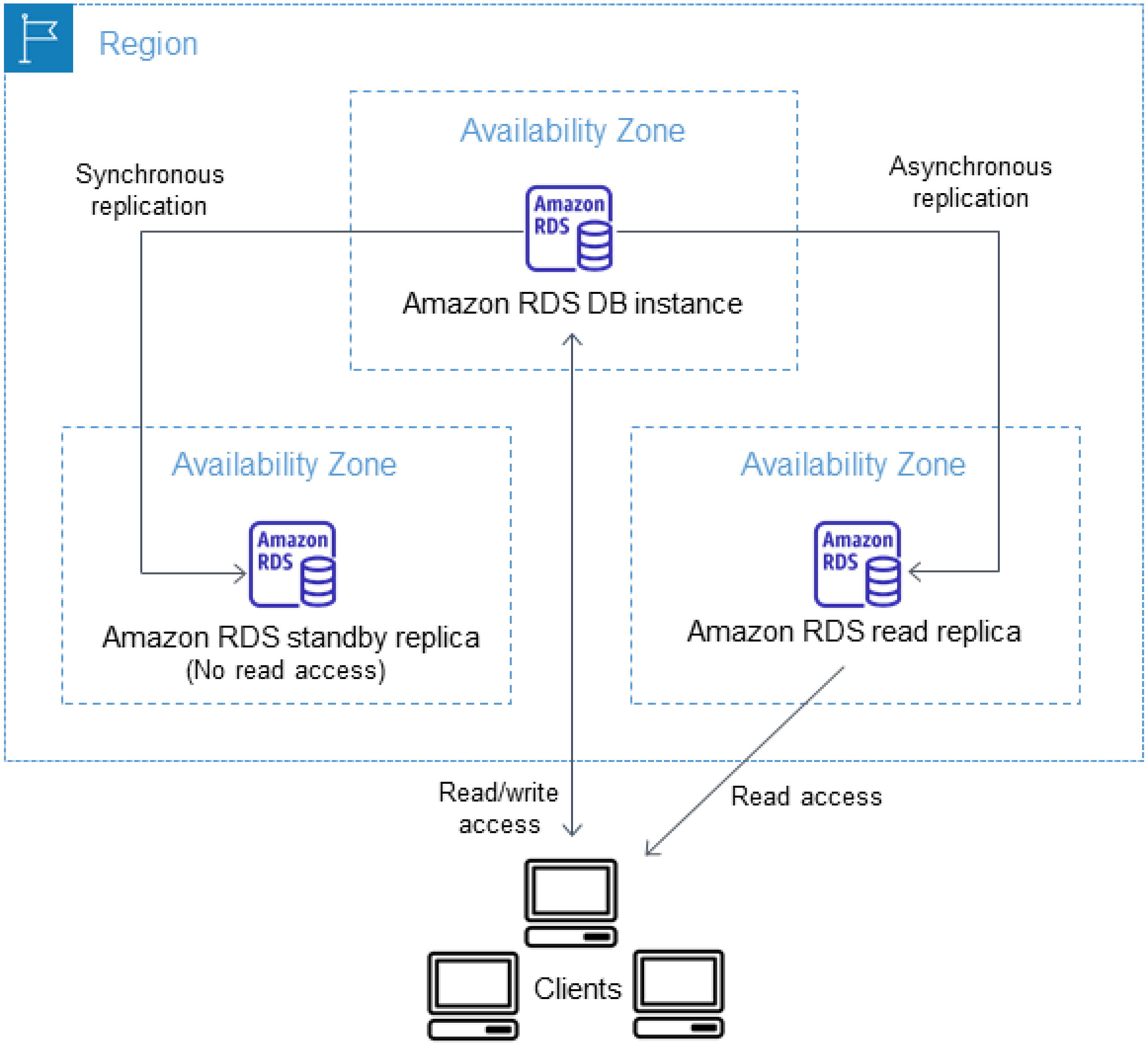
Application

- Amazon Relational Database Service(RDS)

Amazon RDS creates a special type of DB instance called a **read replica** from a **source DB instance**. The source DB instance becomes the primary DB instance. **Updates** made to the **primary** DB instance are **asynchronously** copied to the read replica. You can reduce the load on your primary DB instance by routing read queries from your applications to the read replica.



You can configure a read replica for a DB instance that also has a **standby replica** configured for high availability. Replication with the standby replica is synchronous, and the standby replica can't serve read traffic.


Region
Availability Zone
Synchronous replication
Asynchronous replication
Amazon RDS
Amazon RDS DB instance
Availability Zone
Availability Zone
Amazon RDS
Amazon RDS
Amazon RDS standby replica
(No read access)
Amazon RDS read replica
Read/write access
Read access
Clients

*References

https://searchdatamanagement.techtarget.com/definition/database-replication
https://en.wikipedia.org/wiki/Replication_(computing)
https://www.geeksforgeeks.org/single-master-and-multi-master-replication-in-dbms/
https://en.wikipedia.org/wiki/Multi-master_replication
https://blog.couchbase.com/replication-in-nosql-document-databases-mongo-db-vs-couchbase/
https://thamaraiselvam.dev/synchronous-vs-asynchronous-database-replication
https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_ReadRepl.html