

Database replication is the practice of faithfully, comprehensively copying the contents of a relational database into another environment. It is a process of constant updating that serves both operational and analytical purposes.

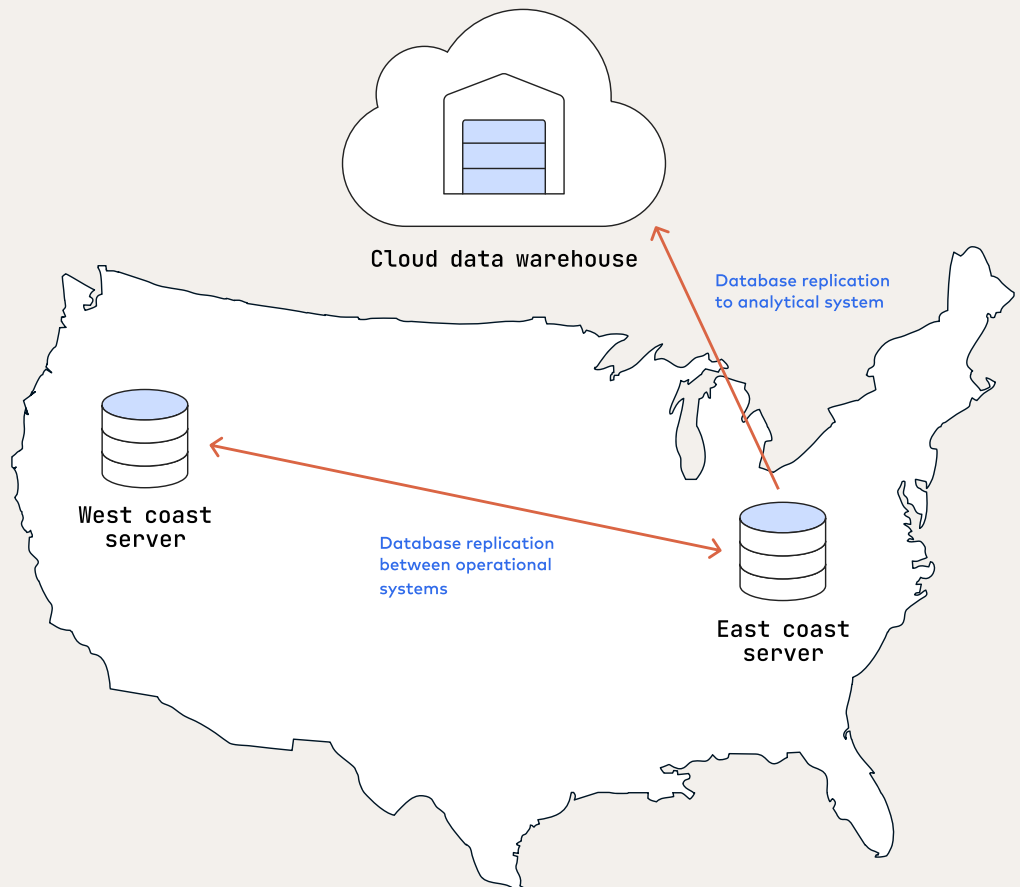
Operations vs. analytics

Operational databases are fundamentally different from analytical databases. The two must not be mistaken for each other.

Operational databases are row-based, designed to efficiently handle small, frequent units of work, i.e. transactions, that concern individual records. The backends of applications are supported by operational databases.

By contrast, analytics databases are column-based, enabling efficient computations over a given column for large numbers of records at a time. Such computations are typical of data analysis.

Database replication for analytics involves moving data from operational databases into an analytical database, such as a cloud data warehouse. Aggregating data into a single source of truth supports reporting, dashboards, predictive models and other analytical uses.

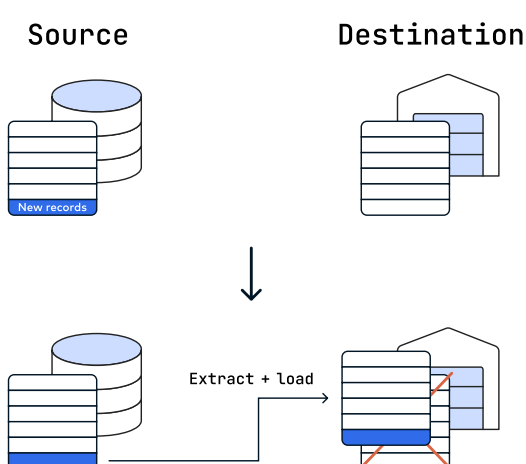


By contrast, database replication for operational databases ensures that the backends of separate environments are synchronized and supply the same data. This data redundancy may be used to spread an application's operational load between multiple servers. For instance, a company may maintain servers for users in specific geographic regions to keep latency lower. It also ensures high availability in case of outages and provides backups for recovery from disasters.

Types of database replication

Snapshot syncs

Copy and replace the entire data set

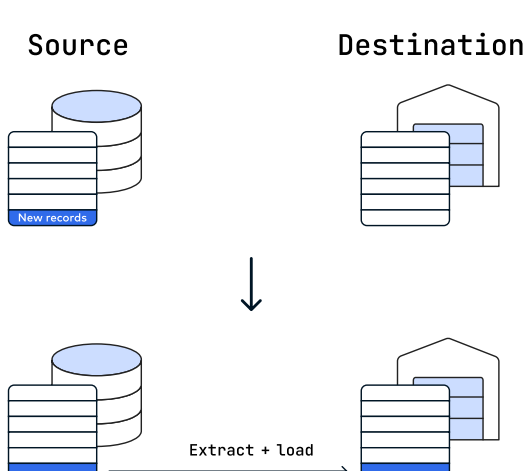


Full or "snapshot" syncs capture the full data set. Full syncs have the following drawbacks:

1. They may take a long time, especially for larger datasets. This often means full syncs are incapable of supporting real-time updates.
2. The large volume of data moved can bog down both the data source and the destination, consuming resources otherwise needed for operations and analytics.
3. They consume excessive network bandwidth and compute expense.

Incremental updates

Only add new records



Incremental updates overcome the disadvantages of full syncs by identifying and replicating changed records only.

This requires the ability to identify changes made to a previous state of the source, a practice often referred to as change data capture (CDC). There are several methods of change data capture; one common method is the use of logs.

At very small increments or batch sizes, CDC enables real-time or streaming updates. CDC is essential to replicating the contents of a database at very high volumes and velocities while minimizing impacts on operational systems.

High-volume database replication is a key Fivetran capability.