Selective Replication in PostgreSQL

Kranthi Kiran Burada

Swiss PGDay 2024

About me

Database Migration Specialist in AWS

10 years of experience in Database migrations

AWS Certification SME

Linkedin: https://www.linkedin.com/in/burada-kranthi-kiran-03b22b127/

aws

Agenda

Understanding Replication

Why replication

Logical Replication

Selective Replication

Key considerations while configuring Selective Replication

Use Cases

aws

Introduction to Replication

- The Process of copying data from a PostgreSQL database server to another server is called PostgreSQL Replication.

- The intent is one or more standby servers ready to take over operations if the primary server fails.

- The Source database server which sends the data is usually called the Master server.

- The Server receiving the copied data is called the Replica/Standby server.

aws

Why Replication?

High Availability

Refers to the ability of having an up to date copy of your database at all times. This means that in the event of a failure of your main database, the standby copy can be promoted to main and you can start receiving traffic.

aws

Load balancing

Practice of distributing incoming requests to your application in a way that is balanced so that no particular database experiences an uneven workload. With replication, this is possible since multiple copies of the data exist at any point in time.
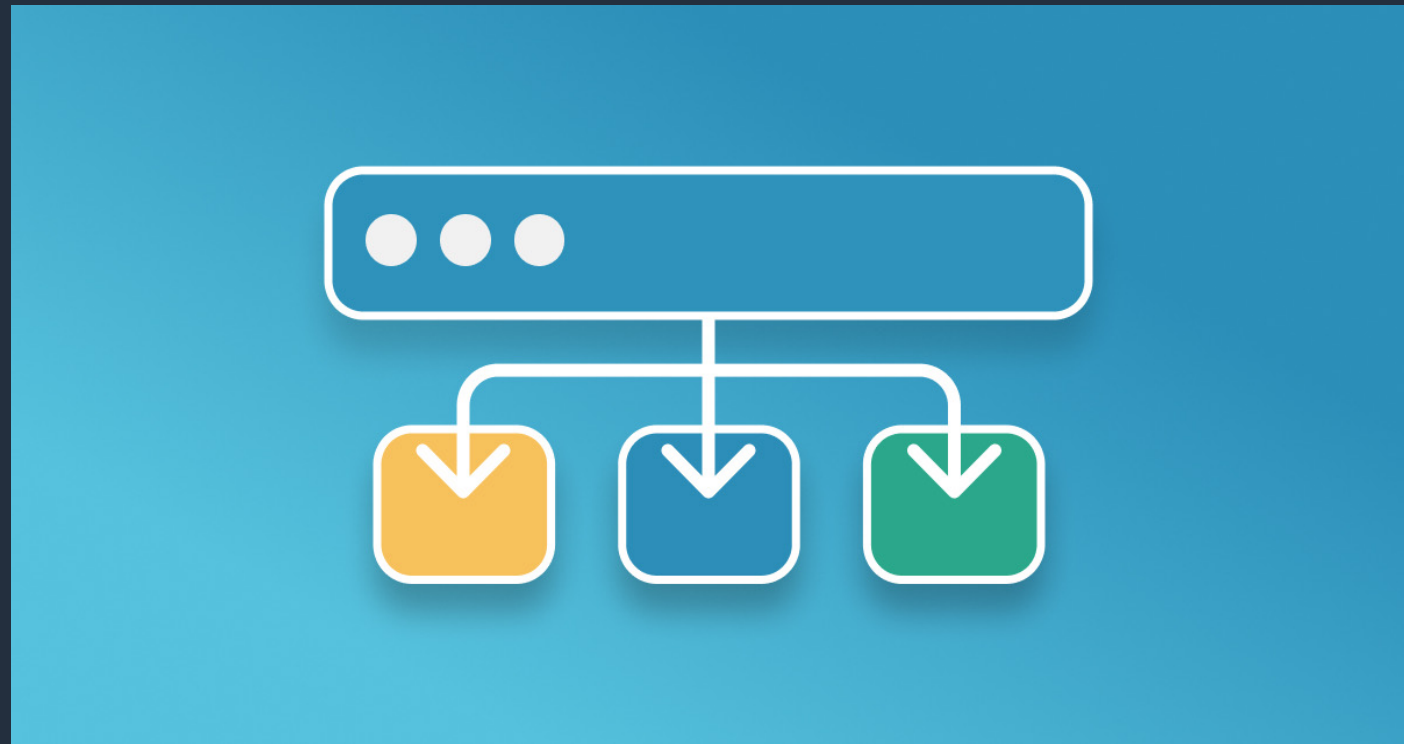
Disaster recovery

Is the need for effective disaster recovery in the event of a systemic failure.

Data Migration

To upgrade database server hardware or patches or even migrating the databases from one database to another

aws

Build in Replication Types

- Physical Replication
    - File Based log shipping
    - Streaming Replication

- Logical Replication

aws

Logical Replication

- Logical replication is a method of replicating data objects and their changes, based upon their replication identity (usually a primary key).

- Logical replication allows fine-grained control over both data replication.

- Logical replication uses a publish and subscribe model with one or more subscribers subscribing to one or more publications on a publisher node.

aws

Logical Replication Architecture
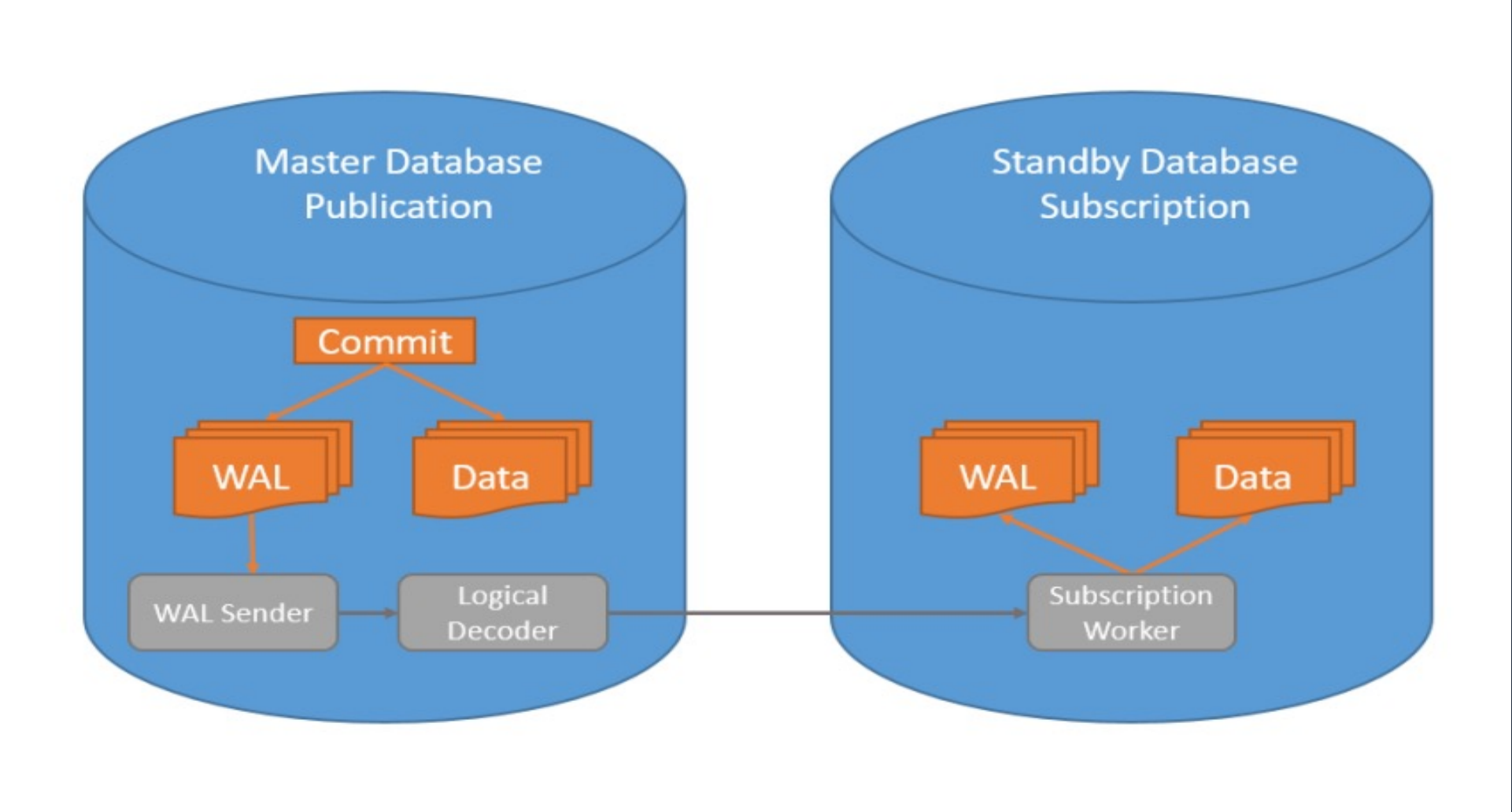
Features of Logical Replication

- Destination server can be used for writes. You can have different indexes and security definition.

- Logical Replication has cross-version/Cross Platform support. Unlike Streaming Replication, Logical Replication can be set between different versions of PostgreSQL.

- Logical Replication can be used for migrations and upgrades

aws

Candidates for Logical Replication

- Consolidate multiple databases into a single database for analytical purposes.

- Replicate data between different major versions of PostgreSQL.

- Send incremental changes in a single database or a subset of a database to other databases.

- Different group of users can access replicated data.

aws

Selective Replication

- From V15 release, PostgreSQL provides more flexibility for managing logical replication

- It introduced **row filtering** and **column lists** for publishers

- Users can choose to replicate a subset of data

aws

Logical Replication: Row filtering

- The publisher can now set a row filter

- The rows gets replicated only if its data satisfies the row filter expression

- Row filter is defined per table

aws

Row filtering: Examples

Example1: Publication of table with one row filter

CREATE PUBLICATION p1 FOR TABLE t1 WHERE (a > 5 AND c = 'NSW');

Example 2: Publication of tables with and without row filters

CREATE PUBLICATION p2 FOR TABLE t1, t2 WHERE (e = 99);

Example 3: Publication of multiple tables with row filters

CREATE PUBLICATION p3 FOR TABLE t2 WHERE (d = 10), t3 WHERE (g = 10);

aws

Key Considerations while using Row filtering

- Row filters use simple expressions

- It cannot contain user-defined functions, operators, types, and collations, system column references or non-immutable built-in functions

- If a publication publishes UPDATE or DELETE operations, the row filter WHERE clause must contain only columns that are covered by the replica identity

aws

Update transformations in Row filtering

The row filter expression is evaluated for both the old and new row

Example:

CREATE PUBLICATION p1 FOR TABLE t1 WHERE (e > 100);

Old Row	New Row	Transformation
No Match (e=30)	No Match (e=50)	Don't Replicate
No Match(e=30)	Match (e=110)	Insert
Match (e=120)	No Match (e=20)	Delete
Match (e=120)	Match (e=130)	Update

aws

Logical Replication: Column Lists

- Each publication can now specify which columns of each table are replicated to subscribers.

- The order of columns in the list is not preserved

- Specifying a column list when the publication also publishes <u>FOR TABLES IN SCHEMA</u> is not supported.

aws

Column List: Example

test_pub=# CREATE TABLE t1(a text, b text, c text, d text, e text);
CREATE TABLE

test_pub=# CREATE PUBLICATION p1 FOR TABLE t1 (b, a, d);
CREATE PUBLICATION

test_sub=# CREATE TABLE t1(a text, b text, d text);
 CREATE TABLE

test_sub=# CREATE SUBSCRIPTION s1

test_sub-# CONNECTION 'host=localhost dbname=test_pub application_name=s1'

test_sub-# PUBLICATION p1;
CREATE SUBSCRIPTION

aws

Use cases for Selective Replication

Reduces network Bandwidth usage and storage requirements

Reduced Network Bandwidth

- Reducing bandwidth usage for data transfer improves replication efficiency

Lower Storage Requirements

- Selective data replication leads to reduced storage which saves cost

Improved Query Performance

- With necessary data, query performance can be optimized as queries only need to process the subset of data relevant to the replica's scope.

aws

Compliance and Privacy Requirements

In environments where data privacy regulations (such as GDPR) or internal compliance policies mandate strict control over the replication of sensitive information

Data Archiving and Purging

In systems where historical data needs to be archived or purged periodically, row filtering can be utilized to replicate only the relevant subset of data to archival or reporting replicas.
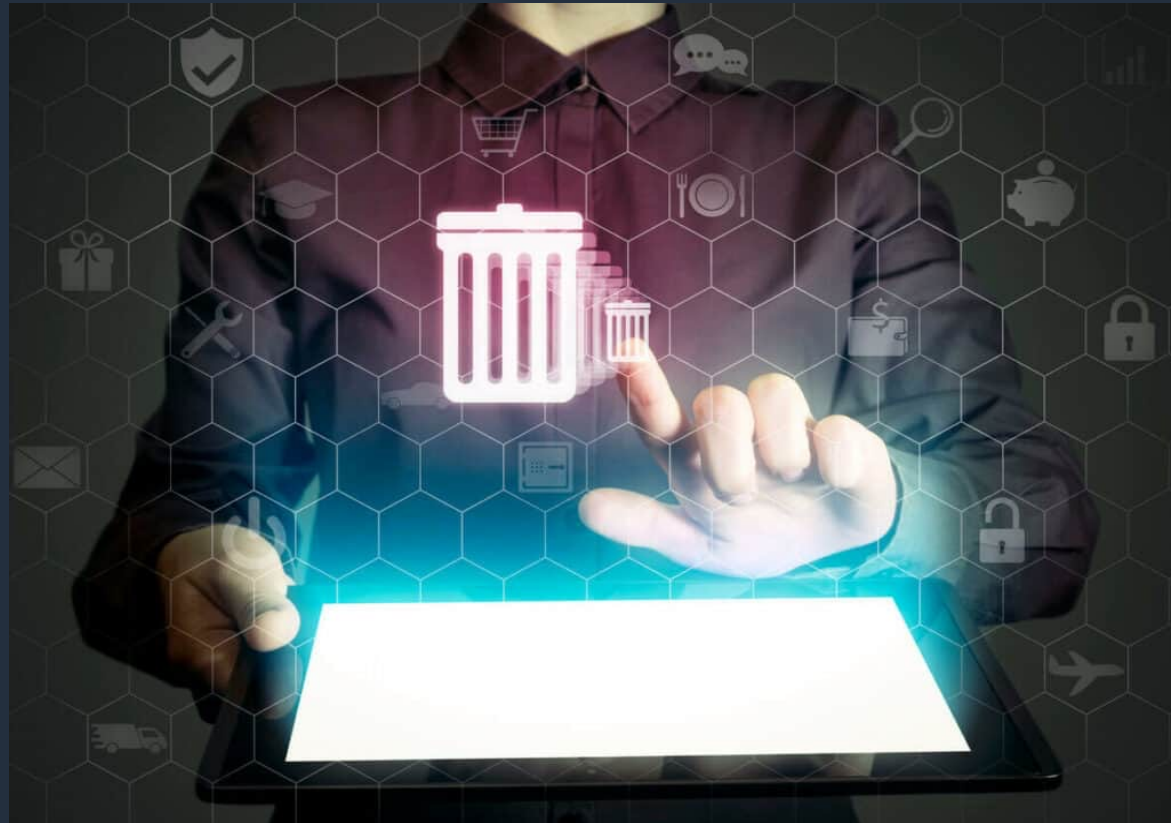
Selective Reporting and Analytics

In analytics or reporting environments where only a subset of data is required for analysis or reporting purposes

Logical Replication New Features & Enhancements

From Version 15

- Users can specify to publish all the tables within a given schema

  - FOR TABLES IN SCHEMA

- pg_stat_subscription_stats view has been added to check the status of errors that occur on the subscription worker

aws

From Version 16

- Subscribers are now able to apply substantial transactions using parallel workers.

- Allow logical replication initial table synchronization to copy rows in binary format.

- New predefined role "pg_create_subscription" which grants users the capability to create logical subscriptions

aws

Thank you!

aws