



Data archive solution in PostgreSQL



PAWAN SHARMA · Follow

3 min read · Oct 21, 2023



20

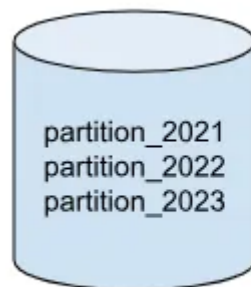


1



Waiting from long time to write something for community. finally today I am here with this article not sure if this will help others or not.

*There are multiple options to archive the data , the data which is rarely or sometime used by the application team in different cases and the most common cases are, which I have figured out that the most of the application team will keep the last year or the older data for auditing purpose. where we perform the very less **INSERT** or **UPSERT** and mostly **SELECT**'s queries to fetch the data.*

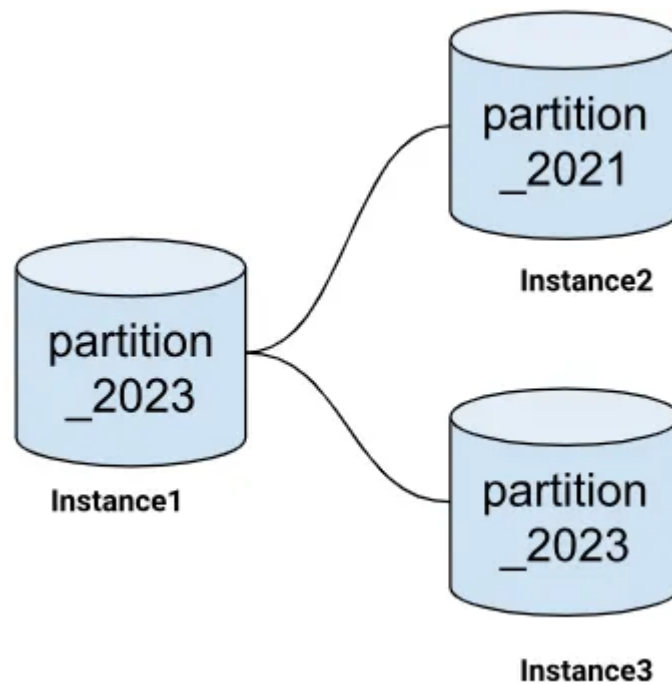


Instance1

Total DB Size: 5TB
Partition_2021 Size: 2TB
Partition_2022 Size: 2TB
Partition_2023 Size: 1TB

How to archive data for such use cases.

1. *Partition's table approach, where will detach the partition which is not in used.*
2. *Take a backup and keep it safe for longer duration and restore it if there is a need in future.*
3. *Create a separate DB cluster, take a dump and restore the data.*
4. *There is another approach to archive the date using partition tables using FDW.*



Advantages of partition based FDW data archive solution.

1. **Single Endpoint(hostname) for application:** *Using this approach the hostname or application DB endpoint will always remain same.*
2. **Backup size will reduce:** *We can take backup of each instance separately. which will help to reduce the backup size*
3. **Restore Time will reduce:** *In case of any issue we can quickly bring up the DB cluster (Instance 1) with partition table_2023.*

4. *Operational efforts will reduce. Keeping the distribute data on each shards, the operational day to day efforts will reduce.*
5. *Using this approach we can easily perform any kind of transaction like select/insert /update and delete directly through app.*
6. *Easy to Delete/Drop the unused data.*

How ?

1. *Create FDW extension on instance1.*

```
// Instance1

CREATE DATABASE appdb;

CREATE TABLE EMP(id INT, name TEXT, time TIMESTAMP NOT NULL ) PARTITION BY RANGE

CREATE TABLE EMP_2023(id INT, name TEXT, time TIMESTAMP NOT NULL );

CREATE EXTENSION postgres_fdw;
```

2. *Create Foreign server and user mapping on instance1 for accessing the instance2 and instance3.*

```
CREATE SERVER instance2 FOREIGN DATA WRAPPER postgres_fdw OPTIONS (host 'instanc
CREATE SERVER instance3 FOREIGN DATA WRAPPER postgres_fdw OPTIONS (host 'instanc
```

```
CREATE USER MAPPING FOR postgres SERVER instance2 OPTIONS (user 'bob', password
CREATE USER MAPPING FOR postgres SERVER instance2 OPTIONS (user 'bob', password
```

3. *Create child partition table on instance2 and instance3 respectively.*

```
CREATE TABLE EMP_2021 (id INT, name TEXT, time TIMESTAMP NOT NULL ); // Instance2
CREATE TABLE EMP_2022 (id int, name text, time TIMESTAMP NOT NULL); // Instance3
```

4. *Create a foreign table on instance1.*

```
ATTACH PARTITION OF EMP for values from ('2021-01-01') to ('2021-12-31') SERVER instance2;
ATTACH PARTITION OF EMP for values from ('2022-01-01') to ('2022-12-31') SERVER instance3;
```

5. *Dump the data of partition emp_2021 and partition emp_2022 from instance1 and restore it on instance2 and instance3 respectively.*

Using the above approach you can access the data of each partition from instance1 and whenever purge of data is required for unused partitions. Decommission that particular instance without affecting the other partition tables.

Conclusion:

There are multiple options and ways to archive data in market, but according to the future requirement of your data and business cases, we can decide which approach we have used to archive our data. so that we can fulfill the business requirements quick and with less operational efforts.

Thanks you for reading the article...

Data Archiving

Postgres



Written by PAWAN SHARMA

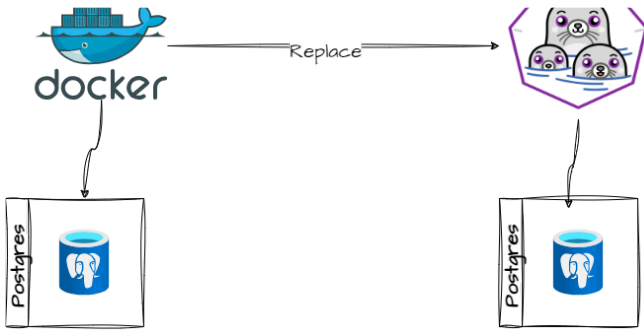
40 Followers

Postgres DBA

Follow



More from PAWAN SHARMA



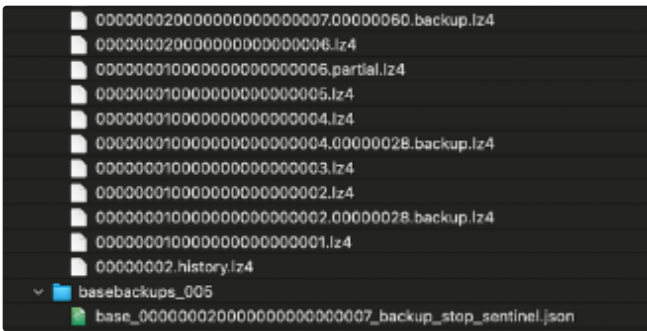
PAWAN SHARMA

Run PostgreSQL with Podman as Docker Container

Podman: Podman is an alternative and similar to Docker, Podman is open source tool whic...

2 min read · Jan 2, 2024

10

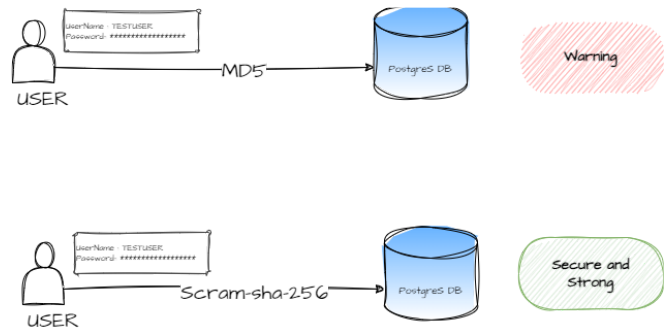


PAWAN SHARMA

Wal-g Backup Solution in PostgreSQL

3 min read · Dec 6, 2023

11



PAWAN SHARMA

Password Encryption In PostgreSQL

Hello and Welcome back. In my last article, I covered how passwords play an important...

5 min read · Mar 8, 2024

21 1



PAWAN SHARMA

Postgres replication lag using debezium connector.

2 min read · Apr 29, 2024

3

See all from PAWAN SHARMA

Recommended from Medium

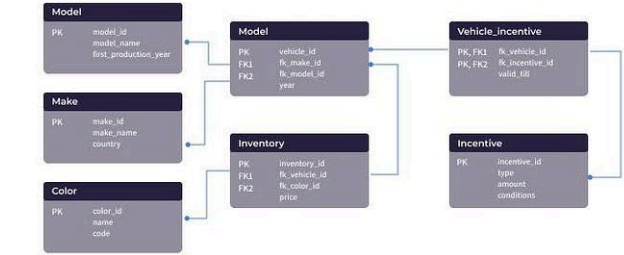


 Adyen in Adyen Tech

Efficient Data Cleanup in Partitioned PostgreSQL Tables...

By Dwarka Rao, Database Engineer & Aakash Agarwal, Java Software Engineer.

9 min read · Mar 22, 2024



 Kamruzzaman Kamrul

Auto-Increment Integer IDs vs. UUIDs in Database Tables: Which ...

2 min read · Dec 22, 2023



Lists



Staff Picks

638 stories · 964 saves



Stories to Help You Level-Up at Work

19 stories · 605 saves



Self-Improvement 101


20 stories · 1798 saves



Productivity 101

20 stories · 1651 saves



 Shadab Mohammad in Oracle Developers

Enhancing OCI PostgreSQL Performance with PgBouncer

Leveraging PgBouncer for Efficient PostgreSQL Connection Management

10 min read · Apr 19, 2024

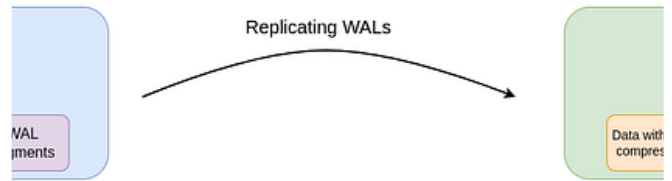



 Oliver Rau

PostgreSQL Multi-Active Replication - Part 2: Native...

This article, the second installment in our series on replication techniques with...

9 min read · Mar 20, 2024

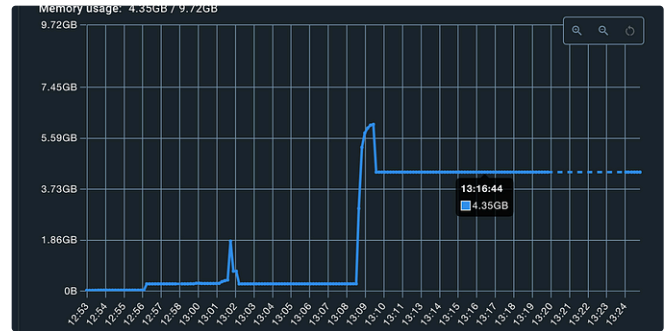


 Anas Anjaria

Improving Postgres Replication: Our Dedicated Disk Move

Diving into Dedicated WAL Disks: A Postgres Perspective

3 min read · Dec 19, 2023



 Hüseyin Demir

PostgreSQL Configuration Chronicles: Memory Related...

We explore the fundamental role memory plays within the architecture of PostgreSQL...

5 min read · Apr 24, 2024



[See more recommendations](#)