

# What Does it Take to Build a Postgres Specialized Data Movement Tool?

- Sai Srirampur, CEO and Co-founder of PeerDB

# Why Postgres?

Open Source

constraints

Extensions

PostGIS/Geospatial

HLL, TOPN, citus

Foreign Data Wrapper

Rich SQL

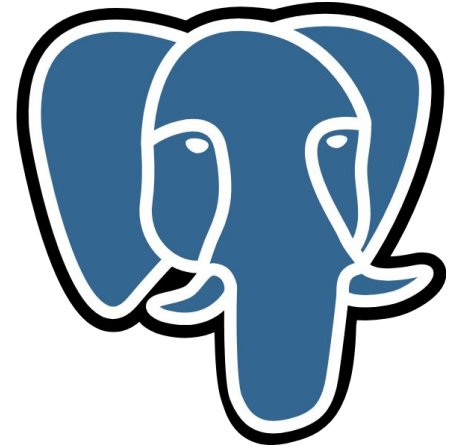
CTES

Window Functions

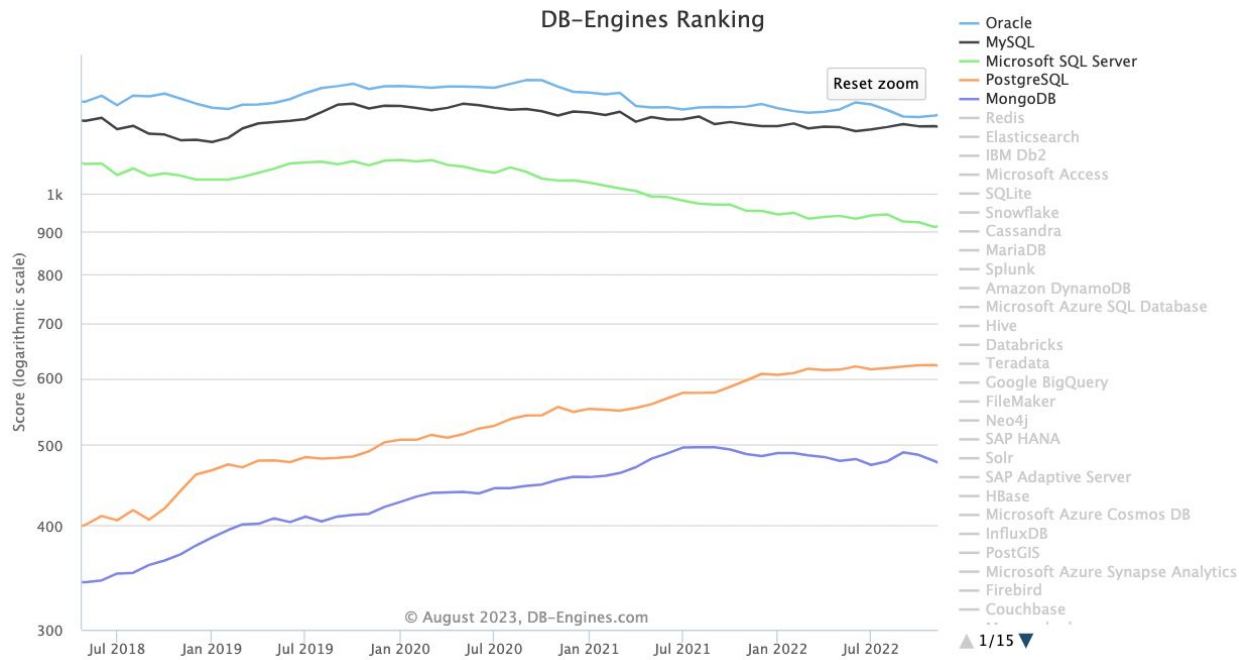
Full Text Search

Datatypes

JSONB



# Postgres is becoming the database of the world



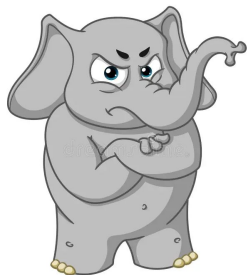
**PostgreSQL lacks a  
first-class data-movement  
tool**

**Most ETL Tools are  
generalized and not built  
for Postgres**



## Slow

Syncing TBs can take days to weeks  
Cannot support real-time syncs



## Not Postgres Native

Data Types - ARRAYS, GEOSPATIAL.  
TOAST Columns



## Unreliable

Not built to handle scale  
Pose risk to Postgres - replication slot growth



## Cost Prohibitive

Unreasonable pricing models  
that can cause costs to spiral out of control

**Solution: Data movement  
tool tailored to Postgres**

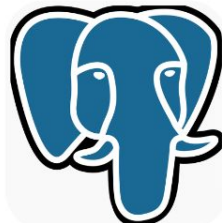
# PeerDB is Built For Postgres

Binary Compressed Data Transfer  
w/ Avro

Native Monitoring and metrics

No Kafka or other  
external deps

Parallel reading of CDC  
replication slots



CTID based initial load



Resilient Orchestration

Binary Copy For  
Homogeneous Transfer

Streaming Server Side Cursors

Query Based Streaming

Parallel Snapshotting



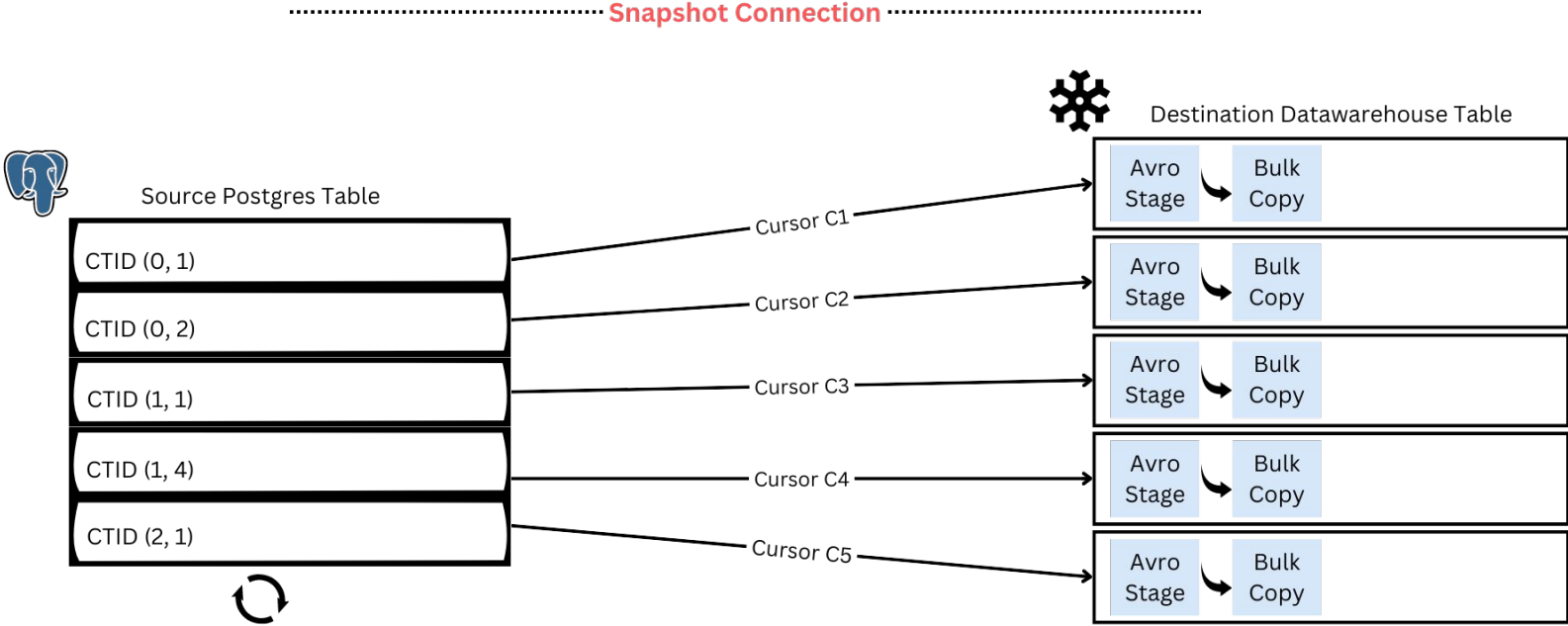
DBT  
Integration

Toast Column Support

Lossless Unified  
Datatypes

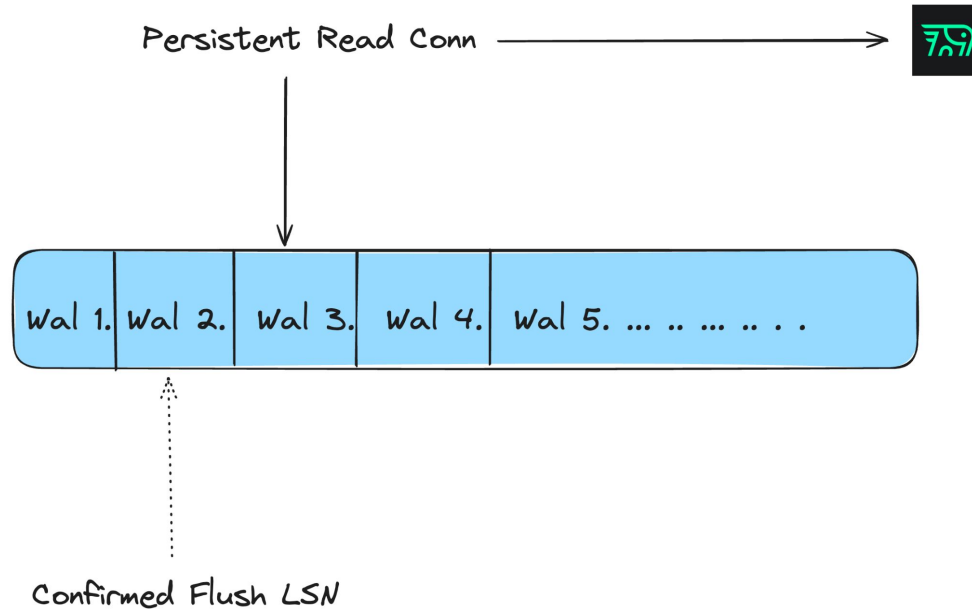


# Parallel Snapshotting



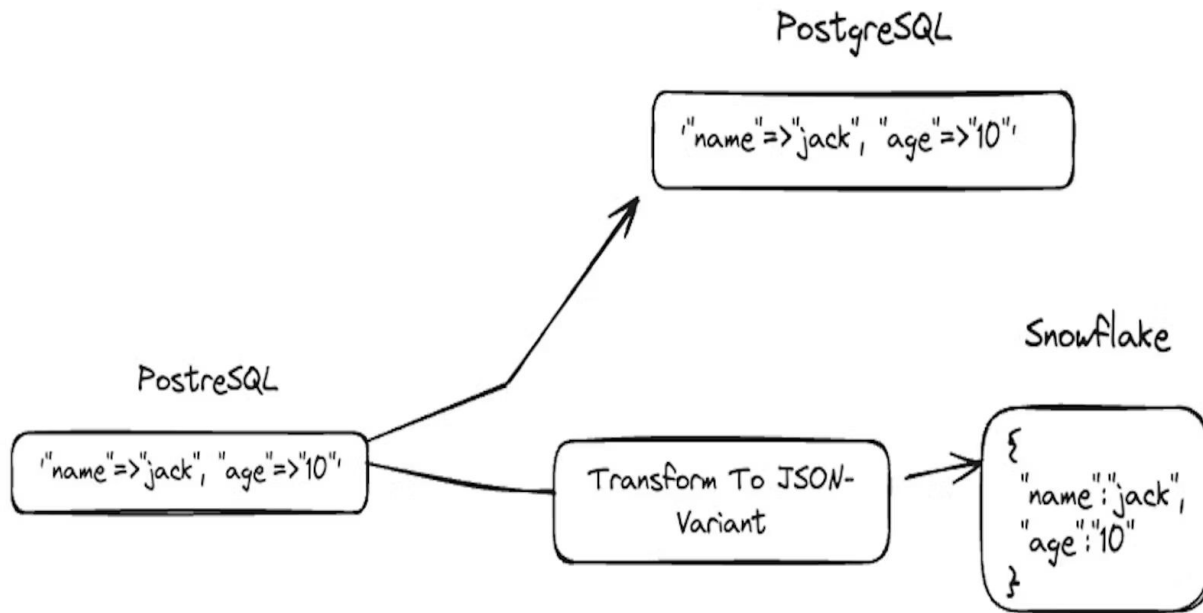
**Move Terabytes within Hours instead of Days**

# Always Consume the replication slot



**Your Postgres Database is Crash Safe**

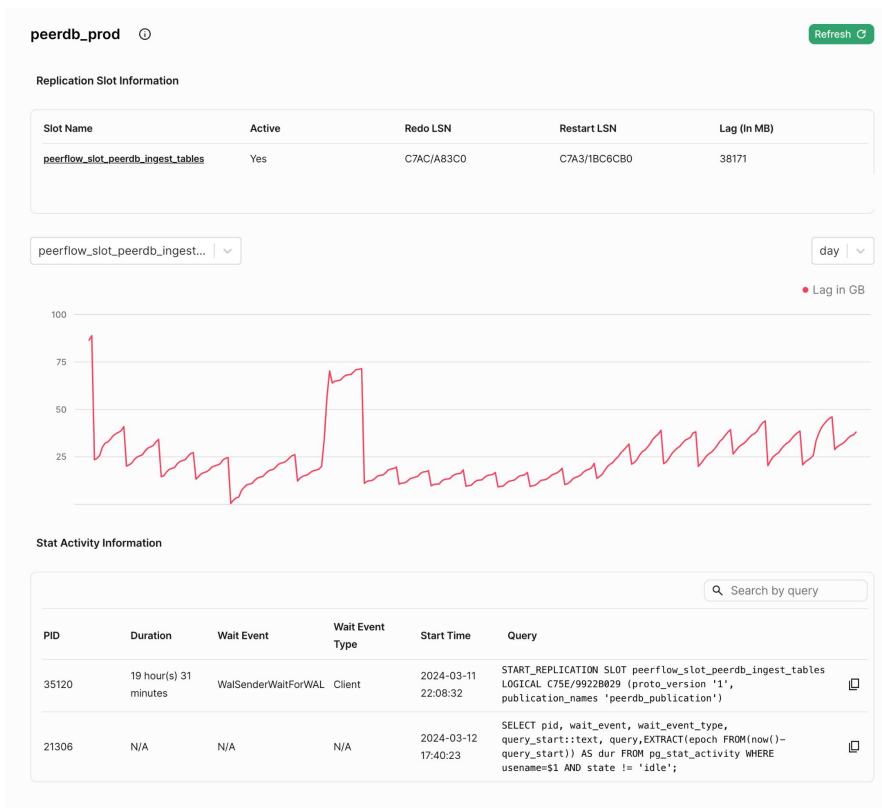
# Replicating Data Types Natively



**Avoids transformations to reduce tech debt and costs**

# Postgres Native Monitoring

1. Replication Slot Growth / Lag
2. Number of connections
3. Activity on Postgres - Wait Events



# A few ideas to implement in upstream PostgreSQL

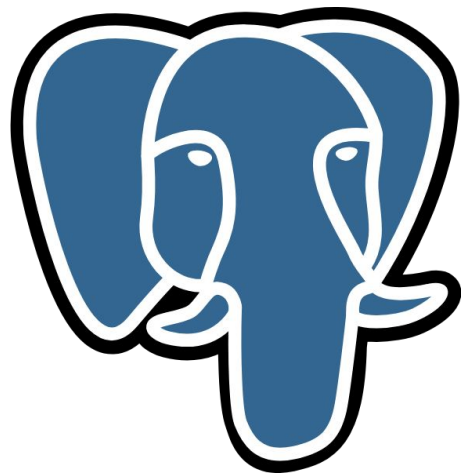
Support replicating schema changes

Replication Slots on Remote storage

Improve CDC (logical decoding) performance

compressed data formats - avro, parquet

Direct ingestion from cloud storage - S3, Blob



# Thank You!



**Sai Srirampur,**

CEO and Co-founder, PeerDB

<https://github.com/PeerDB-io/peerdb>

@saisrirampur

