

PgObject - a self-describing pg-typed structured data type (like JSON but Much Better)

PostgreSQL Conference Europe 2024,
Athens 2024.10.25

hannuk@google.com



Speaker introduction



Hannu Krosing
Cloud SQL / PostgreSQL
hannuk@google.com

Working with PostgreSQL since it was called Postgres95 (and also played around with Postgres 4.2 - without the "SQL" - a little before that).

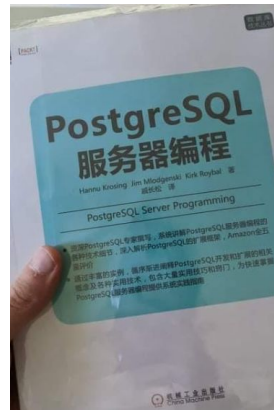
My oldest *surviving* post on postgresql-hackers@ mailing list archives is from January 1998, proposing using index for fast ORDER BY queries with LIMIT.

The first DBA at Skype, where I wrote patches for making **VACUUM** able to **work on more than one table in parallel** and invented the sharding and remote call language **p1/proxy** to make it easy to use PostgreSQL in an infinitely scalable way.

Have written books, **PostgreSQL 9 Admin Cookbook** and **PostgreSQL Server Programming**

After Skype I did 10+ years of PostgreSQL consulting all over the world as part of 2ndQuadrant.

For last few years he has been a PostgreSQL Database Engineer at Google working mostly with Cloud SQL.



SQL is a strongly typed language ...

... So

- When you write a query, it always knows what are the data types
- But we also have type **any** which can mean any type
- And we have even tables which have type **anyarray** in some system tables - `pg_attribute`, `pg_statistic`, `pg_stats_ext_exprs`, `pg_stats`
- But I can not make such table yourself:

```
create table test as select * from pg_stats;
```

```
ERROR: 42P16: column "most_common_vals" has pseudo-type anyarray
```

```
LOCATION: CheckAttributeType, heap.c:581
```



BUT sometimes I want to store "anything"

- Audit trigger saving changes
- Conflicting data row in logical replication (and soon active/active)
- Dynamic, but still structured, user data
 - From some NoSQL databas
 - From AI model output
 - Result of analysis
 - "Anything" really



SQL already has `text`, `JSON`, `JSONB` ?

- But JSON is "lossy" - it loses lot of info about the data, including its type
- It is mostly "just text"
- It can be wasteful
- It does not automatically learn about new types, and in PostgreSQL you can easily add your own if needed



Solution: let's make a PostgreSQL Object type

It has to store anything PostgreSQL can

- Scalar types
- Arrays
- Row Types (Table Types)

But in a way that keeps the type info, meaning together with type info.

And it should have also reasonable indexability and preferably same ordering rules people expect from other structured self-describing types (Python, JavaScript Object, BSON, ...)

The type itself is straightforward (JSON, but binary and with PG type info)
Ordering and other corner-case behaviours need to be agreed upon



PgObject - a
self-describing pg-typed
structured data type
- like JSON but so Much Better

PostgreSQL Conference Europe 2024,
Athens 2024.10.25



hannuk@google.com

