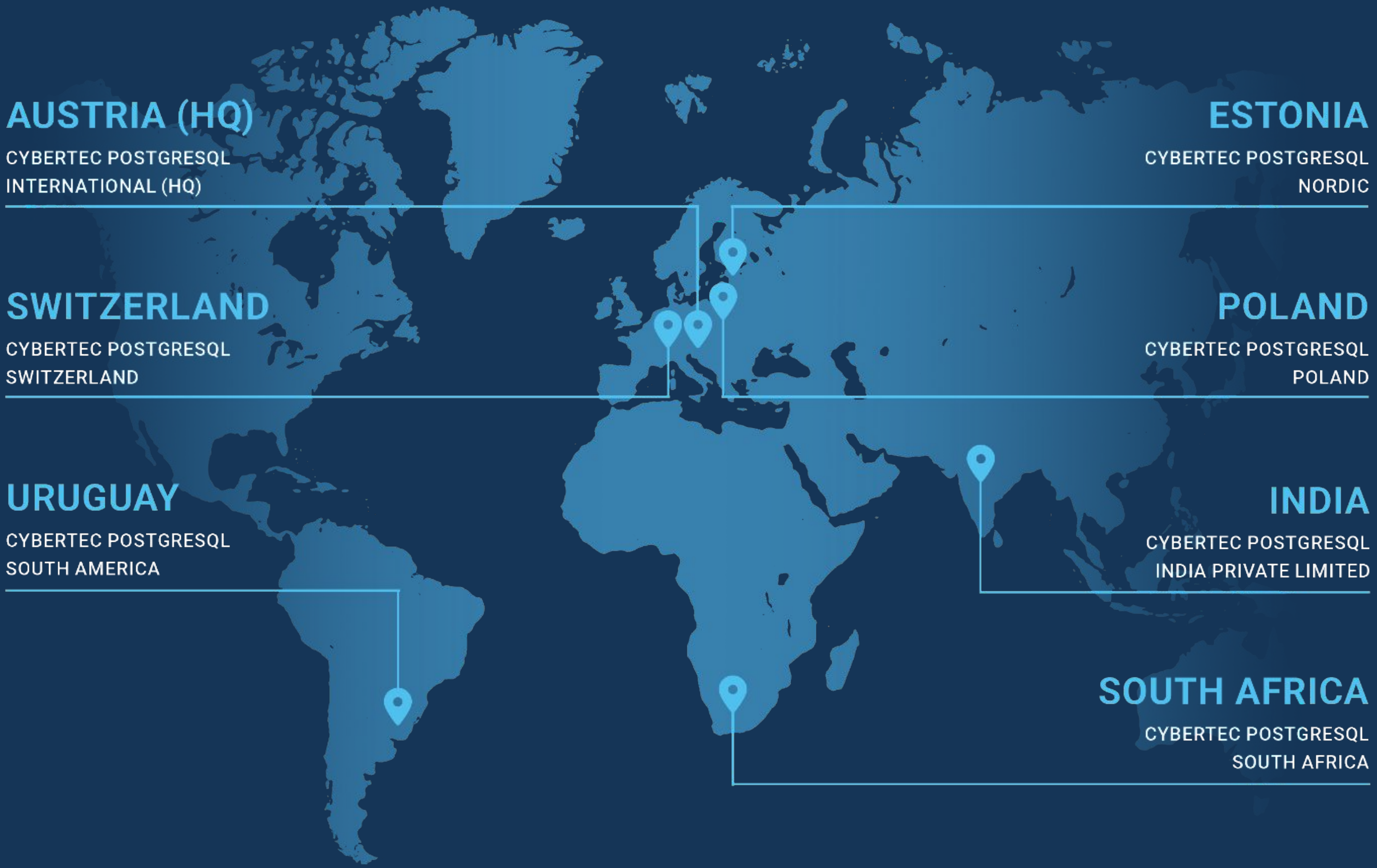
PostgreSQL as Open Source

Vienna 2025

Peter Hofer

CYBERTEC
POSTGRESQL SERVICES & SUPPORT

PROUD CONTRIBUTOR TO
PostgreSQL
the world's most advanced open source database



AUSTRIA (HQ)

CYBERTEC POSTGRESQL
INTERNATIONAL (HQ)

SWITZERLAND

CYBERTEC POSTGRESQL
SWITZERLAND

URUGUAY

CYBERTEC POSTGRESQL
SOUTH AMERICA

ESTONIA

CYBERTEC POSTGRESQL
NORDIC

POLAND

CYBERTEC POSTGRESQL
POLAND

INDIA

CYBERTEC POSTGRESQL
INDIA PRIVATE LIMITED

SOUTH AFRICA

CYBERTEC POSTGRESQL
SOUTH AFRICA

> “*Freedom has many difficulties*
> *& democracy is not perfect,*
> *but we have never had to put*
> *a wall up to keep our people in*”

John F. Kennedy, 1963

> *“Freedom has many difficulties & democracy is not perfect, but we have never had to put a wall up to keep our people in”*

John F. Kennedy, 1963

matches well to software

Closed Source:
What just as happened

“coping audit log data”

Database Compliance and Audit

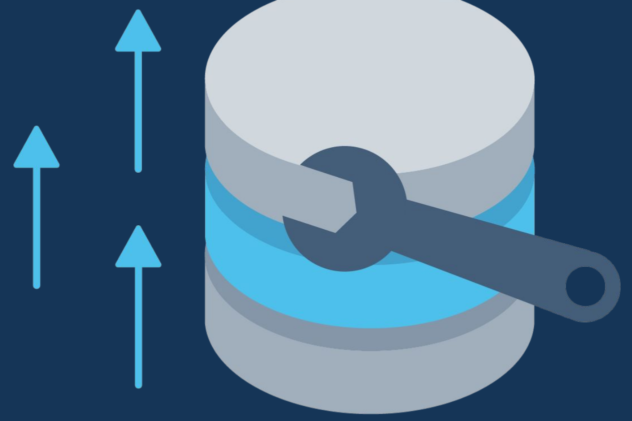
1. Compliance is a significant topic
    - Legal requirements
    - Regulation, certification, specifications

2. Aggregation and analysis of audit data
    - Clear case with PostgreSQL
    - What about closed source?

3. “Some small topics”
    - You begin to understand the differences.

Closed Source

What does that really mean?



Bugs and errors

numerous and disruptive

Knowledge Base

What can one actually know?

The love of detail

Purpose vs. paycheck

Bugs and errors

And how to deal with it

An example

```
DBMS_AUDIT_MGMT.SET_LAST_ARCHIVE_TIMESTAMP(
    audit_trail_type => DBMS_AUDIT_MGMT.AUDIT_TRAIL_UNIFIED,
    last_archive_time => TO_TIMESTAMP(:ts, 'YYYY-MM-DD HH24:MI:SS.FF')
);

DBMS_AUDIT_MGMT.CLEAN_AUDIT_TRAIL(
    audit_trail_type => DBMS_AUDIT_MGMT.AUDIT_TRAIL_UNIFIED,
    use_last_arch_timestamp => true
);
```

- So far, 4 different error messages...
- Identical systems with identical loads
- Occasionally, ancient data remains

Knowledge Base

What can one actually know?

“Let's ask the guru”

PostgreSQL World

- “I'll take a quick look”
- “That makes...”
- “Ah, I know that, wait a moment...”
- … I probably shouldn't have asked …

Commercial world

- “I don't know.”
- “I guess that …”
- “... never seen this before…”

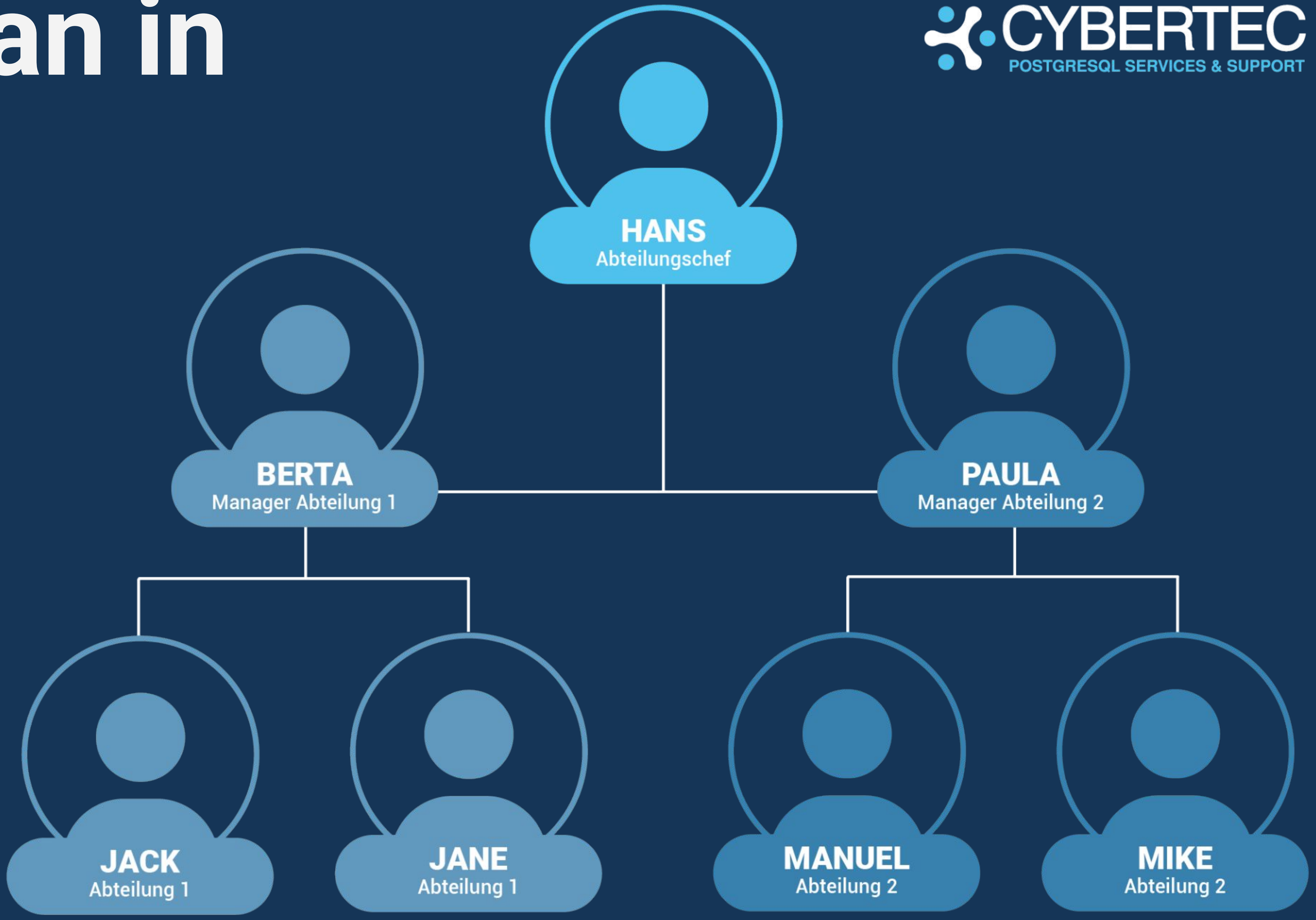
Problem

“ *The realization* ”

How was the colleague supposed to know that, anyway?

What does that mean in real life?

- “Not knowing” = Research
- “Not knowing” = Delay
- “Not knowing” = Risk
- “Not knowing” = Potentially wrong
- “Not knowing” = Frustration
- “Not knowing” = Loss of trust



CYBERTEC
POSTGRESQL SERVICES & SUPPORT

It can be done better

What does open source really mean?

```
test=# SHOW effective_cache_size;
 effective_cache_size
----------------------
 4GB
(1 row)
```

What does this?

```
test=# SHOW effective_cache_size;
 effective_cache_size
----------------------
 4GB
(1 row)
```

Let's take a closer look...

```
hs@system:~/src/postgresql-17.3/src/backend$ grep -r -n -I -l effective_cache_size *
access/gist/gistbuild.c
optimizer/path/costsize.c
utils/misc/postgresql.conf.sample
utils/misc/guc_tables.c
```

GIST? What's happening there?

```
/* subtree must fit in cache (with safety factor of 4) */
   if (subtreesize > effective_cache_size / 4)
       break;
```

Optimizer? Sounds interesting...

```
* costsize.c
 *        Routines to compute (and set) relation sizes and path costs
…
 *      seq_page_cost           Cost of a sequential page fetch
 *      random_page_cost        Cost of a non-sequential page fetch
 *      cpu_tuple_cost          Cost of typical CPU time to process a tuple
 *      cpu_index_tuple_cost  Cost of typical CPU time to process an index tuple
 *      cpu_operator_cost       Cost of CPU time to execute an operator or function
 *      parallel_tuple_cost Cost of CPU time to pass a tuple from worker to leader backend
 *      parallel_setup_cost Cost of setting up shared memory for parallelism
…
 * We also use a rough estimate "effective_cache_size" of the number of
 * disk pages in Postgres + OS-level disk cache.  (We can't simply use
 * NBuffers for this purpose because that would ignore the effects of
 * the kernel's disk cache.)
 *
 * Obviously, taking constants for these values is an oversimplification,
 * but it's tough enough to get any useful estimates even at this level of
 * detail.  Note that all of these parameters are user-settable, in case
 * the default values are drastically off for a particular platform.
```

Optimizer? Sounds interesting...

```
 * index_pages_fetched
 *        Estimate the number of pages actually fetched after accounting for
 *        cache effects.
 *
 * We use an approximation proposed by Mackert and Lohman, "Index Scans
 * Using a Finite LRU Buffer: A Validated I/O Model", ACM Transactions
 * on Database Systems, Vol. 14, No. 3, September 1989, Pages 401-424.
 * The Mackert and Lohman approximation is that the number of pages fetched is
 *      PF =
 *              min(2TNs/(2T+Ns), T)                    when T <= b
 *              2TNs/(2T+Ns)                                   when T > b and Ns <= 2Tb/(2T-b)
 *              b + (Ns - 2Tb/(2T-b))*(T-b)/T   when T > b and Ns > 2Tb/(2T-b)
 * where
 *    T = # pages in table,  N = # tuples in table
 *    s = selectivity = fraction of table to be scanned, b = # buffer pages available
 *
 * We assume that effective_cache_size is the total number of buffer pages
 * available for the whole query, and pro-rate that space across all the
 * tables in the query and the index currently under consideration.  (This
 * ignores space needed for other indexes used by the query, but since we
 * don't know which indexes will get used, we can't estimate that very well;
 * and in any case counting all the tables may well be an overestimate, since
 * depending on the join plan not all the tables may be scanned concurrently.)
```

Optimizer? Sounds interesting...

```
        /*----------
         * Estimate number of main-table pages fetched, and compute I/O cost.
         *
         * When the index ordering is uncorrelated with the table ordering,
         * we use an approximation proposed by Mackert and Lohman (see
         * index_pages_fetched() for details) to compute the number of pages
         * fetched, and then charge spc_random_page_cost per page fetched.
         *
         * When the index ordering is exactly correlated with the table ordering
         * (just after a CLUSTER, for example), the number of pages fetched should
         * be exactly selectivity * table_size.
```

- Yes, the example is simple
- Yes, you still need to know something about the topic
- But: Users and support have a chance!

Love for details

To improve life for everyone

Example: JSON support

PostgreSQL

- early adopter,
- JSON in 2012
- JSONB in 2014.

Oracle

- JSON in 2014,
- native binary JSON in 2021

Oracle entered the JSON world in 2014, but its native binary JSON type only arrived in 2021 – seven years after PostgreSQL's JSONB

Recently: more than 1000 words

```
test=# SELECT 'CREATE TABLE tab' || x || ' (id int) '
       FROM generate_series(1, 3) AS x;
          ?column?
-----------------------------
 CREATE TABLE tab1 (id int)
 CREATE TABLE tab2 (id int)
 CREATE TABLE tab3 (id int)
(3 rows)

test=# \gexec
CREATE TABLE
CREATE TABLE
CREATE TABLE
```

Note: nobody does such a thing

In a nutshell

a few thoughts

> *“Freedom has many difficulties*
> *& democracy is not perfect,*
> *but we have never had to put*
> *a wall up to keep our people in”*

If you have a good solution,
you don't have to lock people up.

Any questions?

Ask anything

CYBERTEC
POSTGRESQL SERVICES & SUPPORT

Hans-Jürgen Schönig

CEO & Founder

Email

hs@cybertec-postgresql.com

Phone

+43 2622 930 22 - 666

www.cybertec-postgresql.com

@cybertec-postgresql

www.youtube.com/@cybertecpostgresql

PROUD CONTRIBUTOR TO

PostgreSQL
the world's most advanced open source database

Our Partners at PGDay Austria

readyset

redgate

Kachelbunt

Cloudway
IT Consulting & Services

PROVENTA

straDATA

OpenSourceDB
Your Trusted Data Partner

FerretDB

TechSurve

paxia
MULTIPLYING POTENTIAL

metalstack
cloud

Nibble
POSTGRES PROFESSIONALS

zebanza

Open Alliance

For PostgreSQL Education

Your Pathway to Verified PostgreSQL Skills

Scan for Updates

oapg-edu.org