

Finding PostgreSQL place in the AI space

Floor Drees

@DevOpsBarbie

@floord@hachyderm.io



PostgreSQL ****and**** AI

- PG as a data source
- PG as a key component of an AI solution
- PG as a target

whoami

- Open Source Community @ Aiven
- Ex-Microsoft, Grafana Labs
- Devopsdays organizer
- Community over Code organizer
- Meetup (you guessed it) organizer



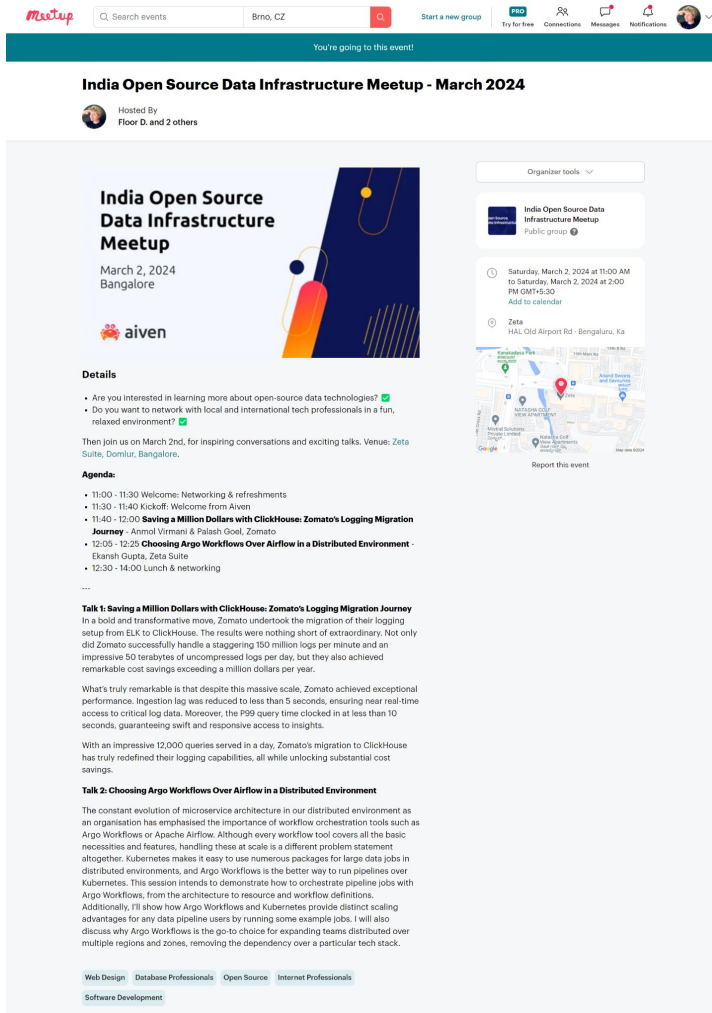
Open Source Data Infrastructure Meetup

March 2, 11:00AM

Zeta, HAL Old Airport Rd, Bengaluru

- Having a Million Dollars with ClickHouse: Zomato's Logging Migration Journey - Anmol Virmani & Palash Goel, Zomato
- Choosing Argo Workflows Over Airflow in a Distributed Environment - Ekansh Gupta, Zeta Suite

<https://www.meetup.com/india-open-source-data-infrastructure-meetup/>



The screenshot shows the Meetup event page for "India Open Source Data Infrastructure Meetup - March 2024". The page is hosted by "Floor D, and 2 others". The event is scheduled for March 2, 2024, in Bangalore at Zeta, HAL Old Airport Rd. The event details include a schedule of talks and networking, and a description of the topics to be discussed, such as Zomato's logging migration and Argo Workflows. The page also features a map of the venue and a list of tags including Web Design, Database Professionals, Open Source, and Internet Professionals.

India Open Source Data Infrastructure Meetup - March 2024

Hosted By
Floor D, and 2 others

India Open Source Data Infrastructure Meetup
March 2, 2024
Bangalore

aiven

Details

- Are you interested in learning more about open-source data technologies?
- Do you want to network with local and international tech professionals in a fun, relaxed environment?

Then join us on March 2nd, for inspiring conversations and exciting talks. Venue: Zeta Suite, Domtur, Bangalore.

Agenda:

- 11:00 - 11:30 Welcome: Networking & refreshments
- 11:30 - 11:40 Kickoff: Welcome from Aiven
- 11:40 - 12:00 **Saving a Million Dollars with ClickHouse: Zomato's Logging Migration Journey** - Anmol Virmani & Palash Goel, Zomato
- 12:05 - 12:25 **Choosing Argo Workflows Over Airflow in a Distributed Environment** - Ekansh Gupta, Zeta Suite
- 12:30 - 14:00 Lunch & networking

...

Talk 1: Saving a Million Dollars with ClickHouse: Zomato's Logging Migration Journey

In a bold and transformative move, Zomato undertook the migration of their logging setup from ELK to ClickHouse. The results were nothing short of extraordinary. Not only did Zomato successfully handle a staggering 150 million logs per minute and an impressive 50 terabytes of uncompressed logs per day, but they also achieved remarkable cost savings exceeding a million dollars per year.

What's truly remarkable is that despite this massive scale, Zomato achieved exceptional performance. Ingestion lag was reduced to less than 5 seconds, ensuring near real-time access to critical log data. Moreover, the P99 query time clocked in at less than 10 seconds, guaranteeing swift and responsive access to insights.

With an impressive 12,000 queries served in a day, Zomato's migration to ClickHouse has truly redefined their logging capabilities, all while unlocking substantial cost savings.

Talk 2: Choosing Argo Workflows Over Airflow in a Distributed Environment

The constant evolution of microservice architecture in our distributed environment as an organisation has emphasised the importance of workflow orchestration tools such as Argo Workflows or Apache Airflow. Although every workflow tool covers all the basic necessities and features, handling these at scale is a different problem statement altogether. Kubernetes makes it easy to use numerous packages for large data jobs in distributed environments, and Argo Workflows is the better way to run pipelines over Kubernetes. This session intends to demonstrate how to orchestrate pipeline jobs with Argo Workflows, from the architecture to resource and workflow definitions.

Additionally, I'll show how Argo Workflows and Kubernetes provide distinct scaling advantages for any data pipeline users by running some example jobs. I will also discuss why Argo Workflows is the go-to choice for expanding teams distributed over multiple regions and zones, removing the dependency over a particular tech stack.

Web Design Database Professionals Open Source Internet Professionals
Software Development

PostgreSQL as a data source





- Your Postgres already contains data;
- Postgres is often the first data touchpoint;
- Postgres is easy to query and integrate.

PostgreSQL as a data source

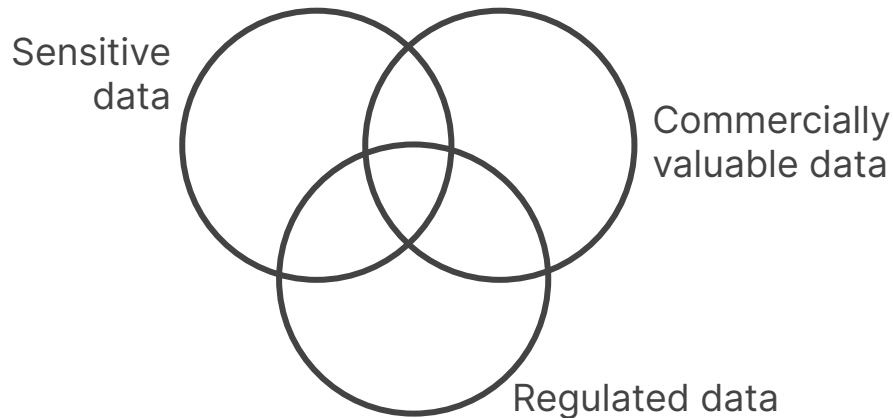
- Fine-tuning

-  >  > 

- Prompt assembly

-  How does product A work?
-  Hi Customer AAA, I see you had errors using Feature X, do you want to know more?
-  Yes, thanks.
-  You can use with Feature X in Product B by ...

AI and risk



- Anonymisation
- Masking
- Aggregation

TECH

Google researchers say they got OpenAI's ChatGPT to reveal some of its training data with just one word

Beatrice Nolan Dec 4, 2023, 12:26 PM CET

Share Save



Google researchers managed to get ChatGPT to reveal its training data, a study says.
OLIVIER DOULIERY/Getty Images

A team of Google researchers say they've found a way to extract some of [ChatGPT's](#) training data.

In a paper published last week, the researchers said certain keywords forced the bot to divulge sections of the dataset it was trained on.

In one example [published in a blogpost](#), the model gave out what appeared to be a real email address and phone number after being prompted to repeat the word "poem" forever. Worryingly, the researchers said the release of personal information often happened when they ran the attack.

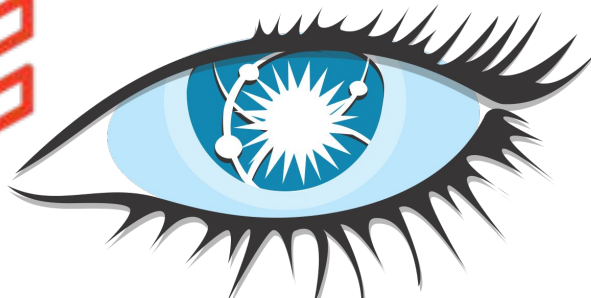
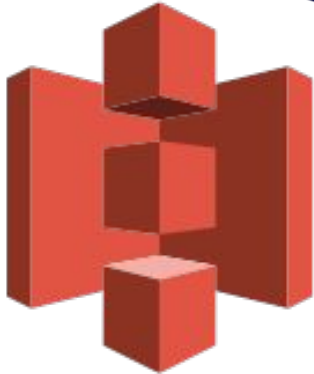
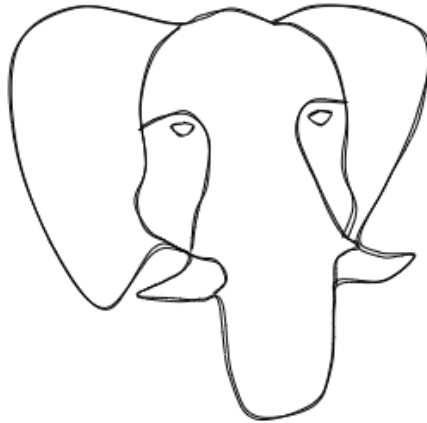
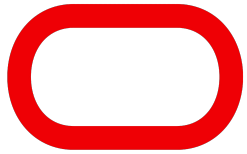
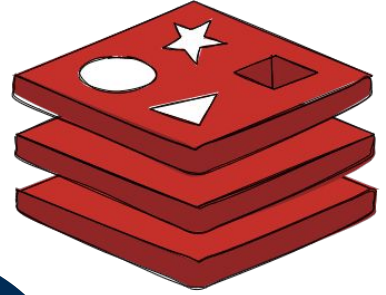
A similar leak of training data was also achieved when the model was asked to repeat the word "company" forever in [another example](#).

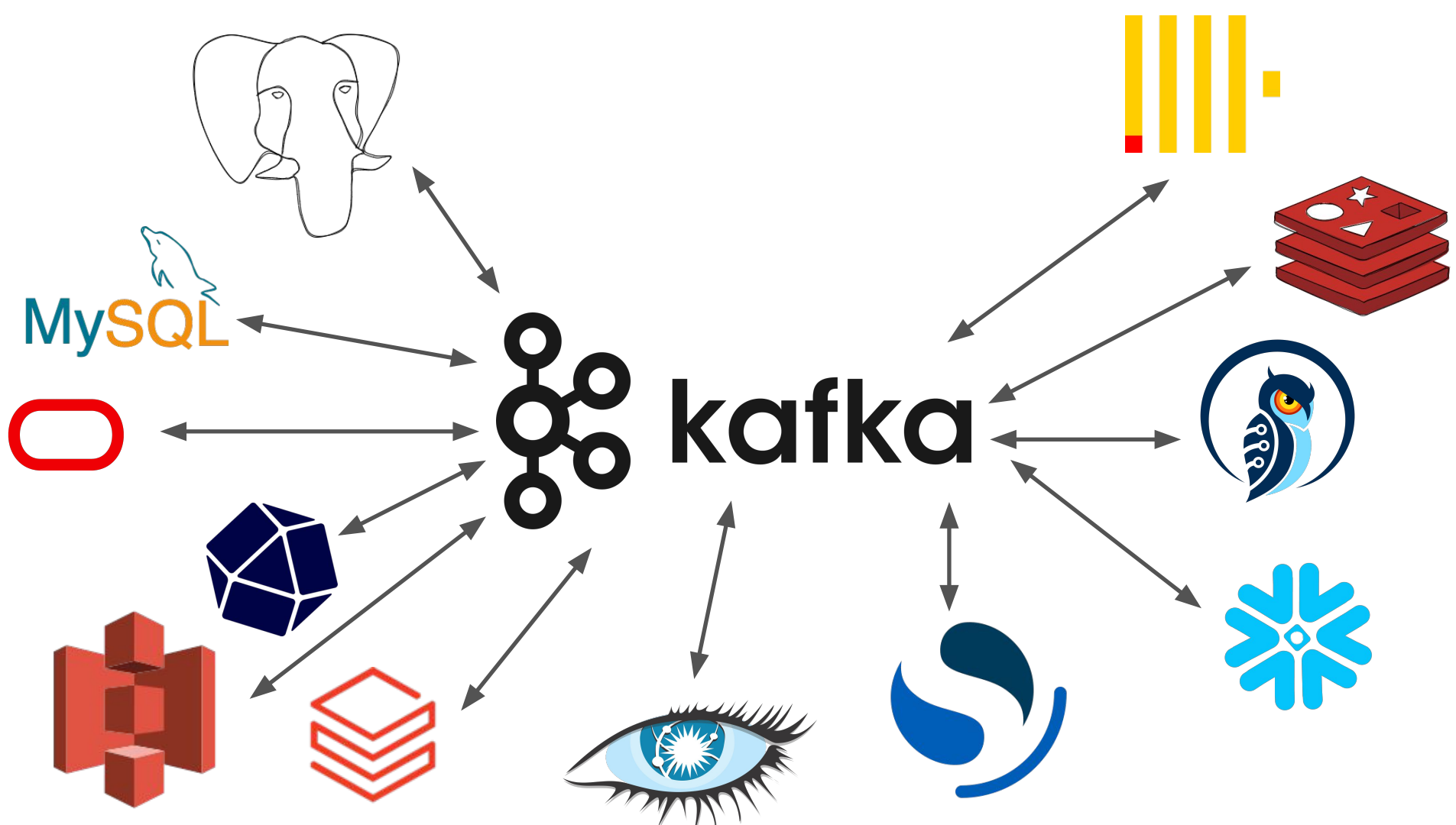
The researchers, who called the simple attack "kind of silly," said in the blogpost: "It's wild to us that our attack works and should've, would've, could've been found earlier."

They said in the paper with only \$200 worth of queries they were able to "extract over 10,000 unique verbatim memorized training examples."

"Our extrapolation to larger budgets (see below) suggests that dedicated adversaries could extract far more data," they added.

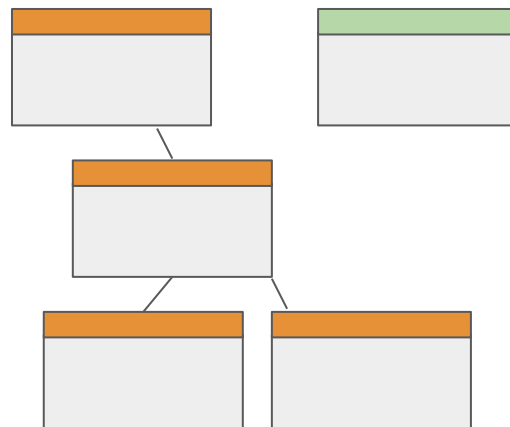
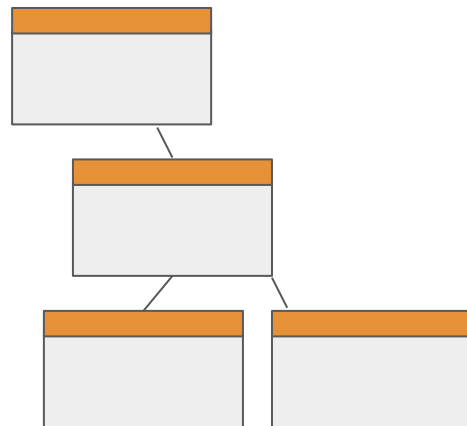
MySQL





PostgreSQL **and** Kafka

- Query-based
- Log-based
 - The Outbox pattern
 - Dedicated table
 - Logical decoding messages



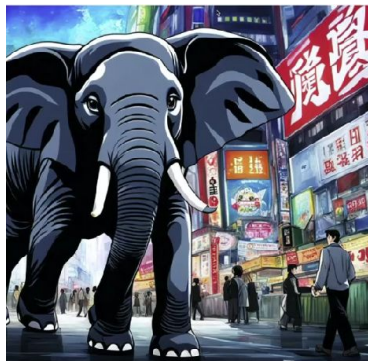
PostgreSQL as a key component

AI in PostgreSQL:

- Labels
- Numbers
- Embeddings / vectors

Must-watch

- Matt Cornillon: How I found my Pokémon cards thanks to Postgres: an AI journey (PGConf.EU 2023)
- <https://www.youtube.com/watch?v=QofVrNDSby4>



How I found my
Pokémon cards
thanks to
Postgres

 aiven
Matt Cornillon
PGConf.EU 2023
Senior Solution Architect



PostgreSQL as a key component

<https://github.com/pgvector/pgvector>

SELECT picture_name

FROM pictures

ORDER BY

embedding ↔ '[1,0.6656,0.9888, ...]'

The screenshot shows the GitHub repository for pgvector. The repository is titled "pgvector" and is described as "Open-source vector similarity search for Postgres". The repository has 64 branches and 28 tags. The repository is owned by "ankane" and has 1000 commits. The repository is licensed under the MIT license. The repository is also available as a Docker image, a Helm chart, and a Terraform provider. The repository is also available as a Docker image, a Helm chart, and a Terraform provider. The repository is also available as a Docker image, a Helm chart, and a Terraform provider.

Open-source vector similarity search for Postgres

- exact and approximate nearest neighbor search
- L2 distance, inner product, and cosine distance
- any language with a Postgres client

Plus ACID compliance, point-in-time recovery, RDNs, and all of the other great features of Postgres

Installation

Linux and Mac

Compile and install the extension (supports Postgres 12+)

```
cd /tmp
git clone --branch v0.6.0 https://github.com/pgvector/pgvector.git
cd pgvector
make
make install # may need sudo
```

See the [installation index](#) if you run into issues

You can also install it with [Docker](#), [Protonline](#), [2023n API](#), [Vary](#), [pds](#), or [conda-forge](#) and it comes preinstalled with [Postgres.app](#) and many [hosted providers](#). There are also instructions for [GitHub Actions](#).

Windows

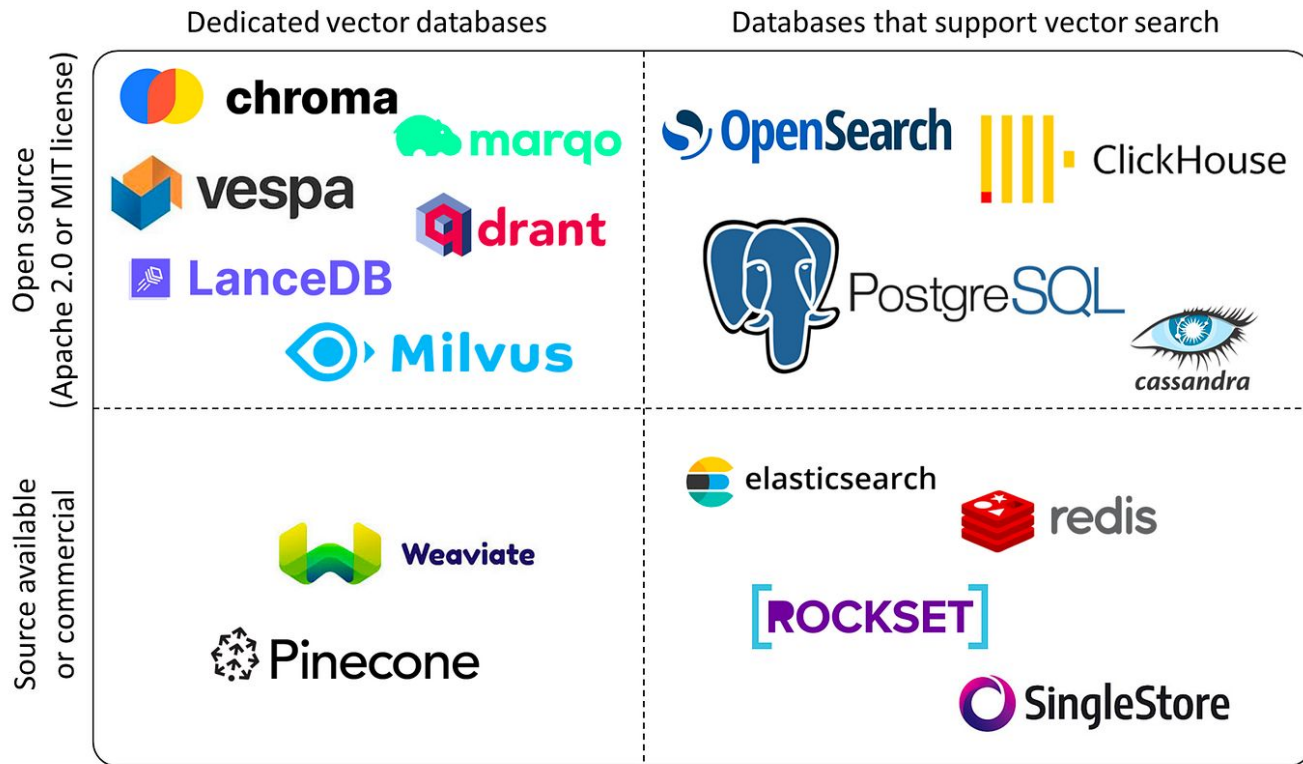
Ensure [C++ support in Visual Studio](#) is installed, and run:

```
cmd /c "C:\Program Files\Microsoft Visual Studio\2022\Community\VC\Auxiliary\Build\vcvars64.bat"
```

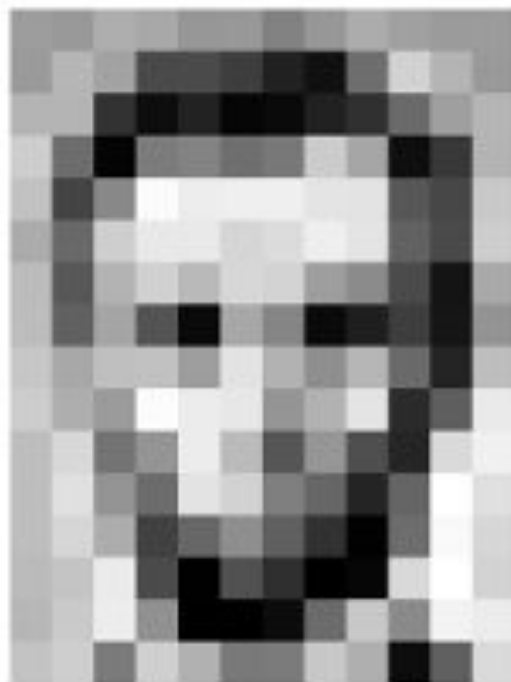
Note: the exact path will vary depending on your Visual Studio version and edition

Then use `make` to build:

```
set PGROOT=C:\Program Files\PostgreSQL\15
git clone --branch v0.6.0 https://github.com/pgvector/pgvector.git
```

Source: <https://datacamp.com/blog/the-top-5-vector-databases>



187	183	174	168	160	152	129	101	172	161	155	156
185	182	163	74	75	52	33	17	719	270	180	154
180	180	50	14	34	6	10	30	48	168	159	181
204	108	6	124	131	111	120	204	166	76	66	180
194	68	137	261	257	239	236	238	207	67	71	181
172	156	267	283	283	214	270	238	238	66	74	204
188	66	170	309	180	218	211	156	159	76	90	169
189	97	166	84	70	148	134	71	31	62	22	148
194	168	181	163	156	227	178	143	183	168	36	190
205	174	165	252	236	231	168	176	228	43	90	134
190	216	156	140	234	187	66	160	78	38	218	247
193	224	147	166	227	210	127	159	36	161	255	134
190	214	173	66	102	143	66	66	2	169	249	218
187	186	235	76	1	61	47	0	6	217	256	211
183	202	237	140	0	0	13	128	200	128	243	236
195	206	123	207	177	121	121	200	176	13	66	218



187	183	174	168	160	152	129	101	172	161	155	156
185	182	163	74	75	52	33	17	719	270	180	154
180	180	50	14	34	6	10	30	48	168	159	181
204	108	6	124	131	111	120	204	166	76	66	180
194	68	137	261	257	239	236	238	207	67	71	181
172	156	267	283	283	214	270	238	238	66	74	204
188	66	170	309	180	218	211	156	159	76	90	169
189	97	166	84	70	148	134	71	31	62	22	148
194	168	181	163	156	227	178	143	183	168	36	190
205	174	165	252	236	231	168	176	228	43	90	134
190	216	156	140	234	187	66	160	78	38	218	247
193	224	147	166	227	210	127	159	36	161	255	134
190	214	173	66	102	143	66	66	2	169	249	218
187	186	235	76	1	61	47	0	6	217	256	211
183	202	237	140	0	0	13	128	200	128	243	236
195	206	123	207	177	121	121	200	176	13	66	218

Must-watch

- Francesco Tisiot: Chihuahua vs Muffin with PostgreSQL and pgvector
- <https://www.youtube.com/watch?v=QofVrNDSby4>
- Jupyter notebook: <https://go.aiven.io/muffin-vs-chihuahua>



AI in PostgreSQL

- plpython3u 
- PostgresML 
- ...

AI in PostgreSQL

- plpython3u 🐍
- PostgresML 🦉
- ...

```
SELECT pgml.transform(  
  task ⇒ 'text-classification',  
  inputs ⇒ ARRAY[  
    'I love how amazingly simple ML has become!',  
    'I hate doing mundane and thankless tasks. 😞'  
  ]  
) AS positivity;
```

AI in PostgreSQL

- plpython3u 🐍
- PostgresML 🦉
- ...

```
SELECT pgml.transform(  
  task ⇒ 'text-classification',  
  inputs ⇒ ARRAY[  
    'I love how amazingly simple ML has become!',  
    'I hate doing mundane and thankless tasks. 😞'  
  ]  
) AS positivity;
```

positivity

```
[  
  {"label": "POSITIVE", "score":  
0.9995759129524232},  
  {"label": "NEGATIVE", "score":  
0.9903519749641418}  
]
```

Move Compute where Data is

- No Data Movement
- No Stale Data
- Security
- Reuse of existing skills
- Increased variety of workloads

Move Data where Compute is

- More choice of tooling
- Different compute needs
- Data scientist are usually NOT in DB
- Separation of concerns

PostgreSQL as the target

"Work smarter, not more"

Developer



- From test to query (or: SQL)
- Optimize queries
- Optimize code (ORM)
- Suggestions and learning

DBA



- Observability
- Performance insights
- Improvement suggestions

PostgreSQL as the target

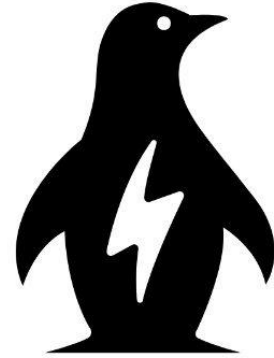
Keep sight of the balance between automatic optimisation vs performance stability.



EverSQL

EverSQL is the fastest way to optimize your PostgreSQL & MySQL databases automatically.












It's used by over 100,000 engineers as a PostgreSQL & MySQL optimizer, to optimize SQL queries.



<https://www.eversql.com/>

EverSQL-Aiven-Gift

sample-458
pg-6963205[← Back to project](#)

-  Overview
-  Integrations
-  Network
-  Metrics
-  Logs
-  Query statistics
-  Current queries
-  Users
-  Databases
-  Pools
-  Backups

My Organization / sample-458 / pg-6963205 / Overview

**pg-6963205**PostgreSQL 15.4  EOL: 2027-11-11: OK  Running  Nodes 1





Get started with Aiven for PostgreSQL®

Set up, secure, and integrate your Aiven service with a step-by-step guide.

[Get started](#)

Connection information

[Quick connect](#)

Service URI	postgres://CLICK_TO_REVEAL_PASSWORD@pg-6963205-sample-458.aivencloud.com:21472/defaultdb?sslmode=require	
Database Name	defaultdb	
Host	pg-6963205-sample-458.aivencloud.com	
Port	21472	

How does EverSQL do all that?

- Query history
- Database metadata
- 🕶️
- balance between accuracy vs confidentiality

```
select *  
from ORDERS  
where  
    CUSTOMER='Floor'
```

PostgreSQL ****and**** AI

- PG as a data source
- PG as a key component of an AI solution
- PG as a target

What's next?

- <https://opensource.org/deepdive/drafts/the-open-source-ai-definition-draft-v-0-0-5>
- <https://opensource.org/blog>

Out of scope issues

The Open Source AI Definition doesn't say how to develop and deploy an AI system that is ethical, trustworthy or responsible, although it doesn't prevent it. We support the efforts to discuss the responsible development, deployment and use of AI systems, including through appropriate government regulation, as a separate conversation.



The Open Source AI Definition – draft v. 0.0.5

version 0.0.5

[Leave comments for this text](#)

Note: This document is made of three parts: A preamble, stating the intentions of this document; the Definition of Open Source AI itself; and a checklist to evaluate licenses.

This document follows the definition of AI system adopted by the Organization for Economic and Co-operation Development (OECD)

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

More information about definitions of AI systems on OSI's blog.

Preamble

Why we need Open Source Artificial Intelligence (AI)

Open Source has demonstrated that massive benefits accrue to everyone when you remove the barriers to learning, using, sharing and improving software systems. These benefits are the result of using licenses that adhere to the Open Source Definition. The benefits can be summarized as: autonomy, transparency, and collaborative improvement.

Everyone needs these benefits in AI. We need essential freedoms to enable users to build and deploy AI systems that are reliable and transparent.

Out of scope issues

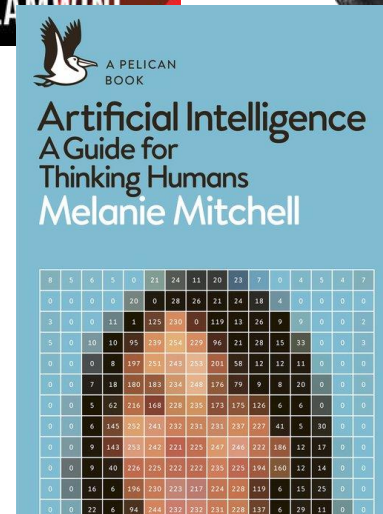
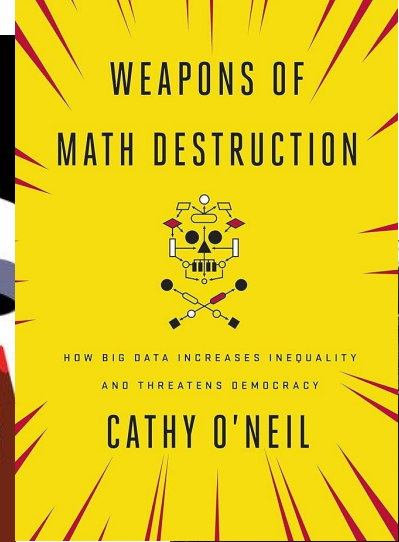
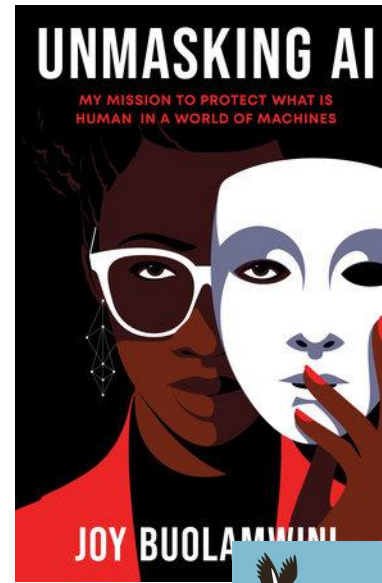
The Open Source AI Definition doesn't say how to develop and deploy an AI system that is ethical, trustworthy or responsible, although it doesn't prevent it. We support the efforts to discuss the responsible development, deployment and use of AI systems, including through appropriate government regulation, as a separate conversation.

What is Open Source AI

To be Open Source, an AI system needs to be available under legal terms that grant the freedoms in:

Must-reads

- Unmasking AI, by Dr. Joy Buolamwini
- Race after Technology, by Ruha Benjamin
- Weapons of Math Destruction, by Cathy O'Neil
- Artificial Intelligence: A Guide for Thinking Humans, by Melanie Mitchell



Thank you

Floor Drees

floor@aiven.io

@DevOpsBarbie

@floord@hachyderm.io

