



EDB

Postgres® for the AI Generation



NatWest

Unlock the Future of Finance: Exploring the Convergence of Postgres and AI with NatWest

9th October 2024

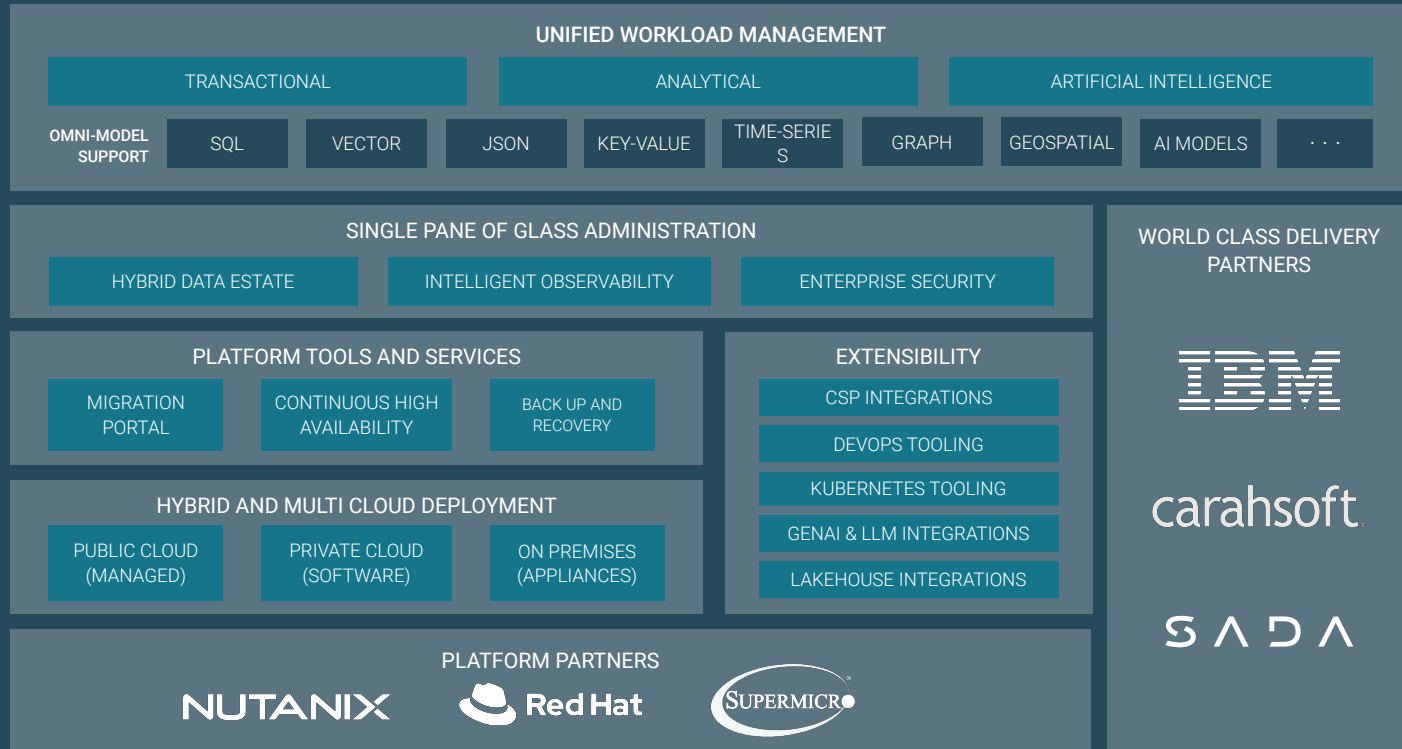


EDB

Postgres for the AI Generation

EDB Postgres AI – One Platform for Transactions, Analytics and AI

The EDB Postgres AI Platform



1. Customer Pain Point – **Sustainable AI:**

“**Many** different **AI projects** and LOBs are implementing their own AI solutions, leading to a **proliferation** of **AI technology** in our company, which **introduces risk** of instabilities, security breaches, human error and lack of skills to our **operational landscape**. But after all, they are all implementing the **same** set of **best practice** AI application **patterns**. How can I establish a **battle-proof platform** that supports these patterns, that can be easily rolled out **for broad adoption** by our LOBs, and that **doesn't bloat** our operational **system landscape**?”



2. Customer Pain Point – **Sovereign AI:**

“We are facing significant hurdles as we are trying to **turn our AI** prototypes and experiments into applications **in production**. We must ensure high standards of data **security and compliance**. As useful as AI providers were in **public cloud** for easy and rapid building of AI applications, this is now the **major blocker** for bringing them into production. We need an **AI platform** that ensures that none of our data leaves a **defined region or data center**.”



3. Customer Pain Point – **Customer 360**

“We have **information** about our customers **spread across** many different **databases** in different clouds and on premise. We need a **hybrid data platform** that **integrates all the data silos** and but **doesn't introduce yet another interface and technology** that I'd need to get acceptance for by all my data producers and consumers.”



4. Customer Pain Point – **Don't Move My Data:**

“We have a **large footprint of business data** in well-oiled data systems that are the backbone of our business. Virtually all our AI and analytics is intrinsically tied to this business data. We need to reduce the **complexity** for preparing this data and **making it available** for analytics and AI. We must also **minimize copying** the data all the time just because we have some analytics or AI to run on it. We need a **Zero-ETL** data platform that can serve our core business **transactions** as well as **analytics** and modern **AI** applications with **real-time data** currency.”



5. Customer Pain Point – **Big Data Scaling Limits & Cost**

“Our business is **growing** and so is our **business data**. Using traditional databases to store **ever growing data volumes** often leads to **technical hurdles** and **cost explosions**. We want to be able to scale our **data** volumes **separately** from the data **compute infrastructure**. But we also want to **avoid yet another database** technology just to get support for **disaggregated data storage**. Ideally, we can use the **same database** platform for **transactional** workloads on **hot data** as well as for **analytics** on the **long tail** of large **historical data** volume.”



6. Customer Pain Point – **HW Acceleration for Analytics & AI**

“**Analytics and AI** data processing is by nature **well-suited** to exploit **GPU** hardware. Prominent examples are **vectorized**, columnar data **processing**, **vector database** processing, or **training and inferencing** of generative **AI models**. But there are **economical hurdles** to assigning **expensive GPUs** to an individual application or workload. We would need a **combined analytics and AI platform** that can host various GPU-accelerated AI and analytic workloads and efficiently leverages a **common pool of GPUs** for that.”

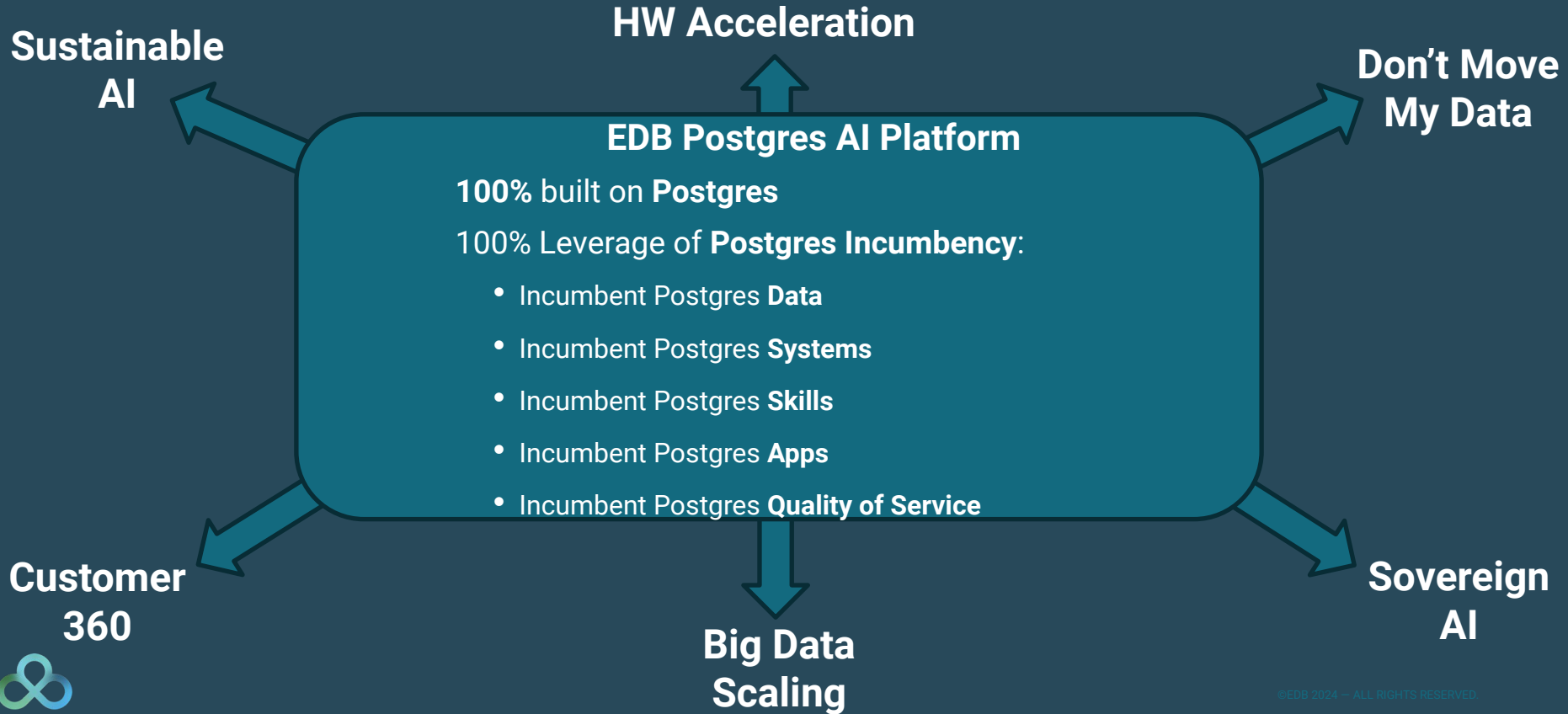


POLL

Which of the pain
points resonate
most with you?

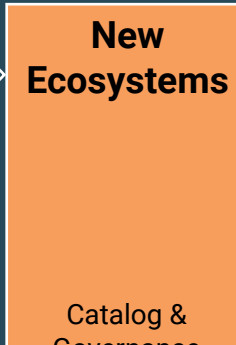
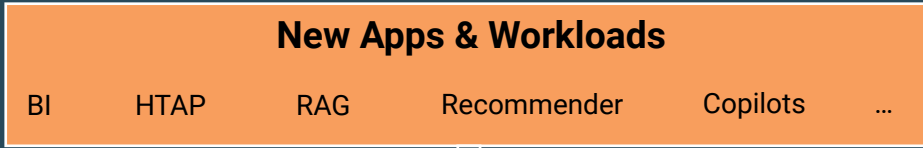


EDB Postgres AI – Maximum Return of Investment in Postgres

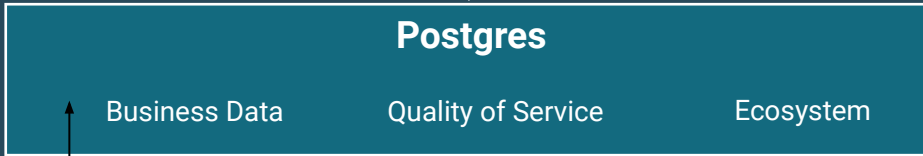


What does EDB Postgres AI add for Analytics & AI?

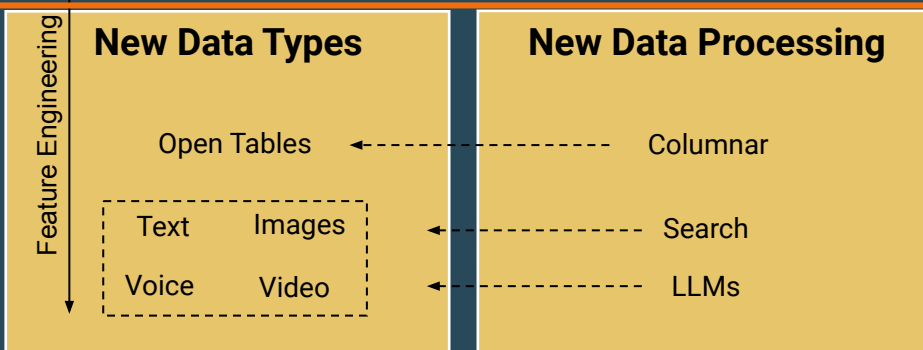
That's what
EDB PG AI
enables



That's what
EDB PG AI
integrates with



That's what EDB is
building in support
of analytics & AI



How: AI & Analytics Capabilities – Strategic Perspective

EDB Postgres AI Platform

1. Data Integration

- External Storage Support
- Tiered Tables
- Lakehouse SQL INSERT
- Migration & Analytic Synch
- Metastore & ETL Vendors

2. Analytic Acceleration

- Columnar Query Engine
- Auto Compaction
- Real-time analytics
- Conversational SQL
- GPU-Accelerated Analytics

3. Search

- PGVECTOR
- Hybrid Search Index
- Text Search
- GPU-Accelerated Search

4. Orchestration


- AI Pipelines
- Embedded model hosting in PG Process
- Text/Image Embeddings
- Auto Embeddings & Retrievers
- In-DB RAG
- AI Feature Engineering
- EDB connectors in AI solution frameworks
- AI Platform Vendors

5. Serving

- Enterprise LLM hosting
- Chat models
- GPU-Accelerated Models
- Model Serving Vendors



Acceleration

 **Wide funnel** for existing data

Scaling analytics w/ columnar engine

Scaling AI search

Broad coverage of application patterns

Scaling hosting of Enterprise AI model

How: AI & Analytics Capabilities – The 2024 Perspective

EDB Postgres AI Platform

1. Data Integration

- External Storage Support
- Tiered Tables
- ~~• Lakehouse SQL INSERT~~
- Migration & Analytic Synchronisation
- ~~• Metastore & ETL Vendors~~

2. Analytic Acceleration

- Columnar Query Engine
- ~~• Auto-Compaction~~
- ~~• Real-time analytics~~
- ~~• Conversational SQL~~
- ~~• GPU Accelerated Analytics~~

3. Search

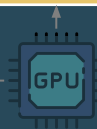
- PGVECTOR
- ~~• Hybrid Search Index~~
- ~~• Text Search~~
- ~~• GPU Accelerated Search~~

4. Orchestration

- AI Pipelines
- Embedded model hosting in PG Process
- Text/Image Embeddings
- Auto Embeddings & Retrievers
- ~~• In-DB RAG~~
- ~~• AI Feature Engineering~~
- ~~• EDB connectors in AI solution frameworks~~
- ~~• AI Platform Vendors~~

5. Serving

- ~~• Enterprise LLM hosting~~
- ~~• Chat models~~
- ~~• GPU Accelerated Models~~
- ~~• Model Serving Vendors~~



Acceleration





EDB

Postgres for the AI Generation

Analytics

PGAA — Postgres Analytics Accelerator

EDB Postgres AI extends the power of Postgres to analytics, with PGAA at the core, integrated tightly with PGD.

PGAA is a Postgres extension for rapid analytics:

- PGAA **separates storage from compute**, and includes support for S3, MinIO or the local filesystem:

```
SELECT pgaa.create_storage_location('sample-data', 's3://pgaa-sample-data-eu-west-1');
```

- PGAA makes **Lakehouse Tables (Delta Lake)** queryable just like any Postgres table:

```
CREATE TABLE public.customer () USING PGAA  
WITH (pgaa.storage_location = 'sample-data', pgaa.path = 'tpch_sf_1/customer');
```

- PGAA has a **Vectorized Query Engine** for **30x faster** queries compared to transactional Postgres:

```
SELECT count(*) AS num_customers from customer;
```

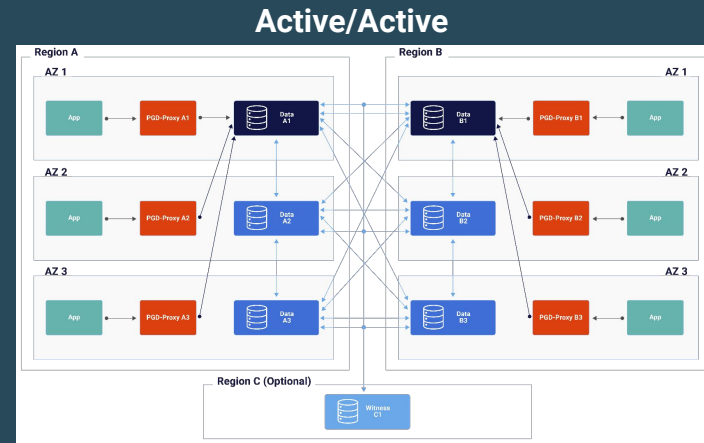
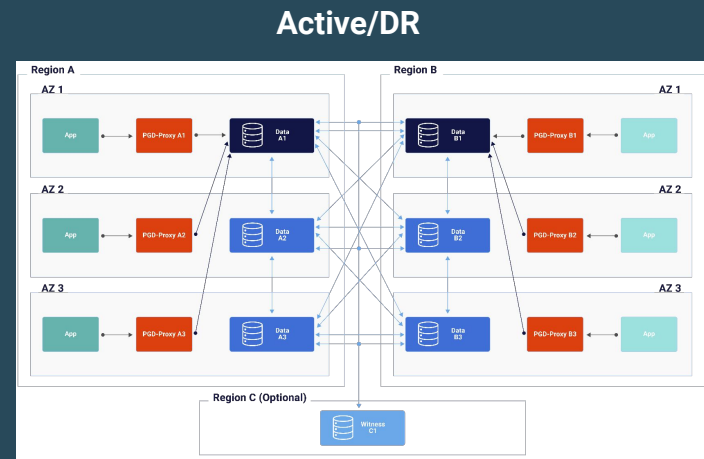
*It's a **Postgres frontend** to the external **Lakehouse** ecosystem.*

*But when combined with **PGD**, it becomes a **real-time analytics** solution...*



EDB's Postgres Distributed (PGD) Maximizes Postgres Availability

- **Fast failover** in seconds
- **5X faster** than native replication
- **Online** maintenance & version upgrades
- **Granular** control of **durability** & **consistency**
- Geo Sharding with **Active-Active** replication
- **Cross cloud** & **Hybrid** cloud deployments



PGD + PGAA — Real-time Analytics for Postgres

EDB Postgres Distributed (PGD) can replicate data directly to object storage and query it with PGAA

PGD now ships with an embedded analytics engine, powered by PGAA:

- Replicate transactional tables **directly to Lakehouse Tables**:

```
CREATE TABLE my_transactional_data(a INT PRIMARY KEY) WITH (pgd.replicate_to_analytics = true);
```

- Optionally, **query tables with the analytical engine**:

```
SET LOCAL bdr.prefer_analytics_engine=true;  
SELECT * FROM x;
```

- Create **Tiered Tables** that offload storage and processing of “cold” data to analytics:

```
CREATE TABLE all_my_data(a timestamp, b int ) PARTITION BY RANGE (a);  
SELECT bdr.autopartition(  
    relation := 'test_part_timestamp',  
    partition_increment := '1 month',  
    partition_initial_lowerbound := CURRENT_TIMESTAMP::text,  
    managed_locally := true,  
    analytics_offload_period := '3 months');
```

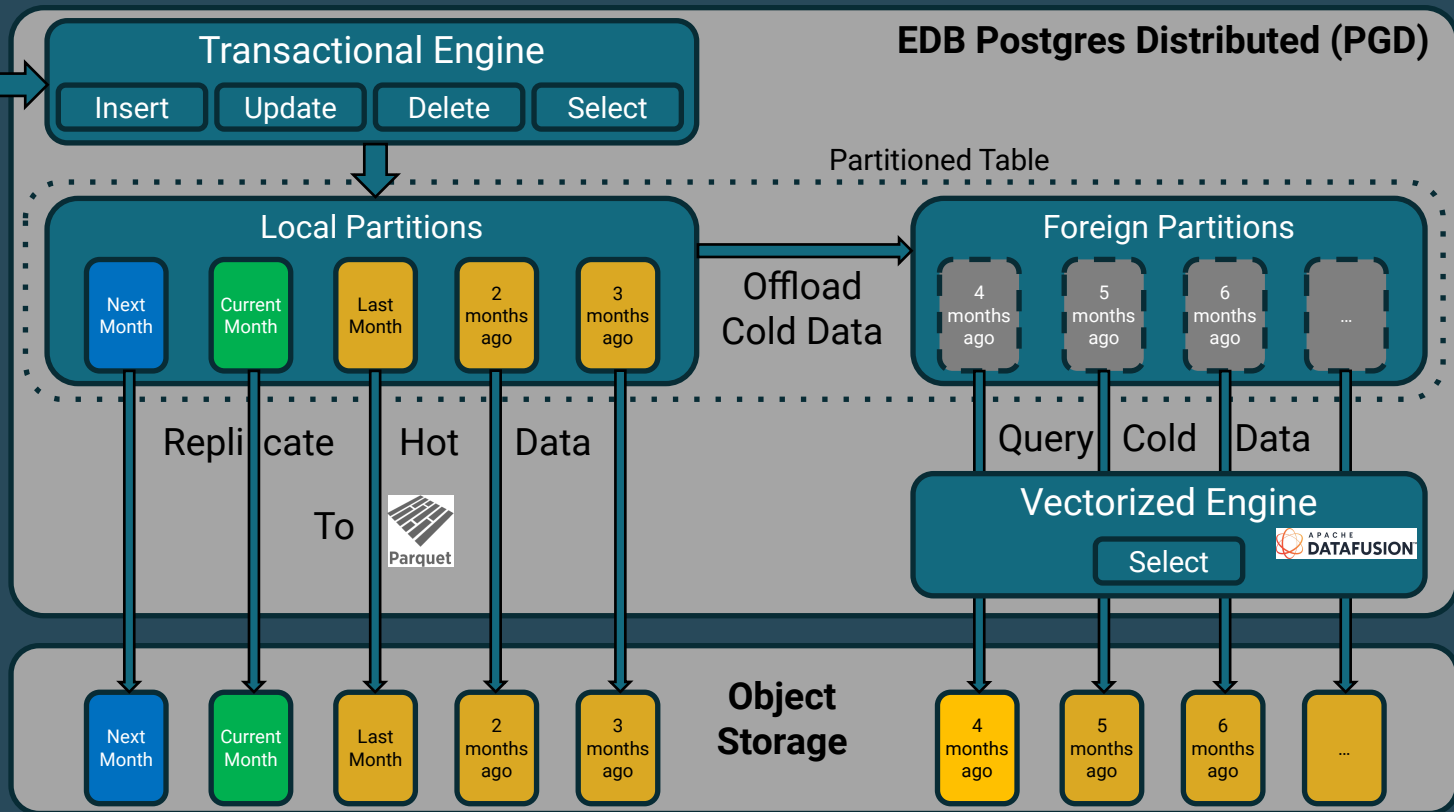


Tiered Tables – Control Storage Costs & Protect Transactions

Transparently offload storage and processing of “cold” data to a separate analytical system & reduce table size on disk



Work with a
Single Version
of all Data



Partition Increment: 1 month
Offload Period: 3 months



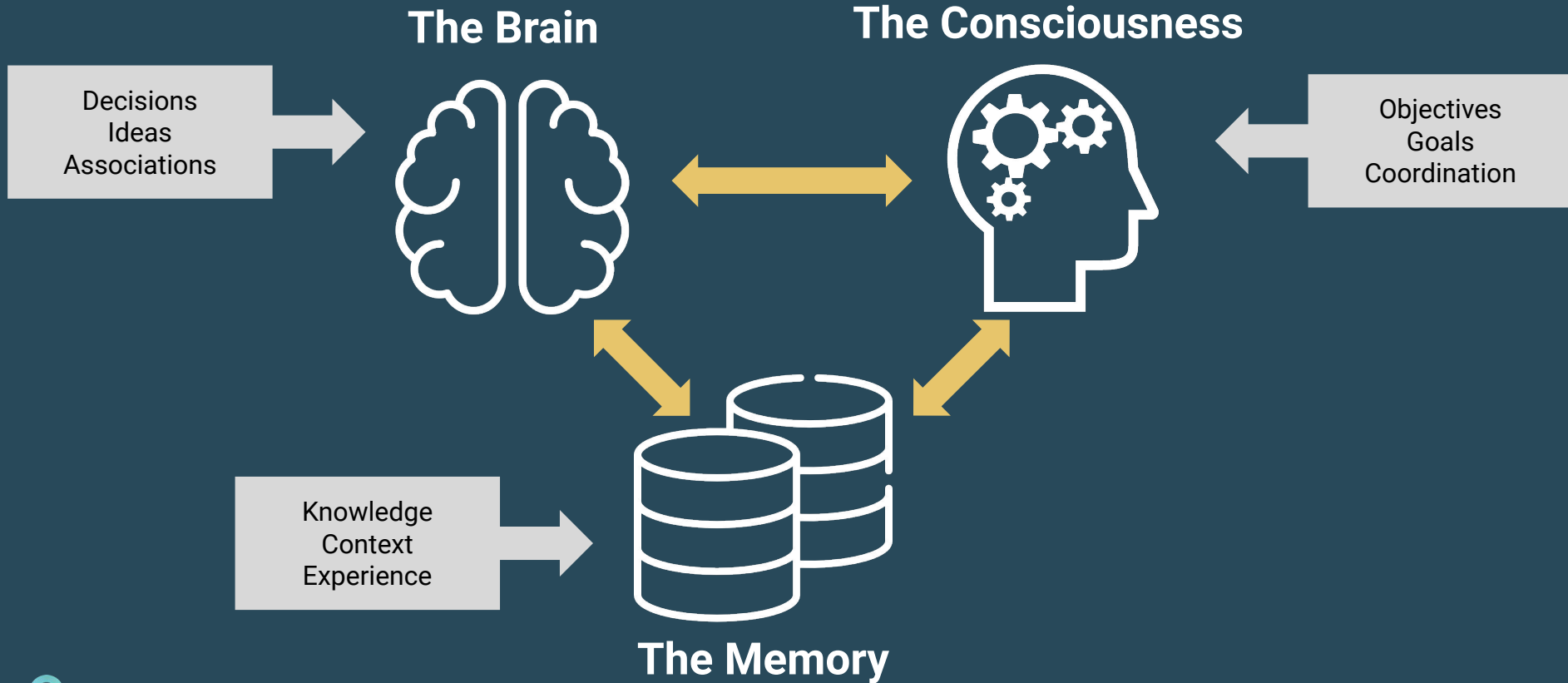


EDB

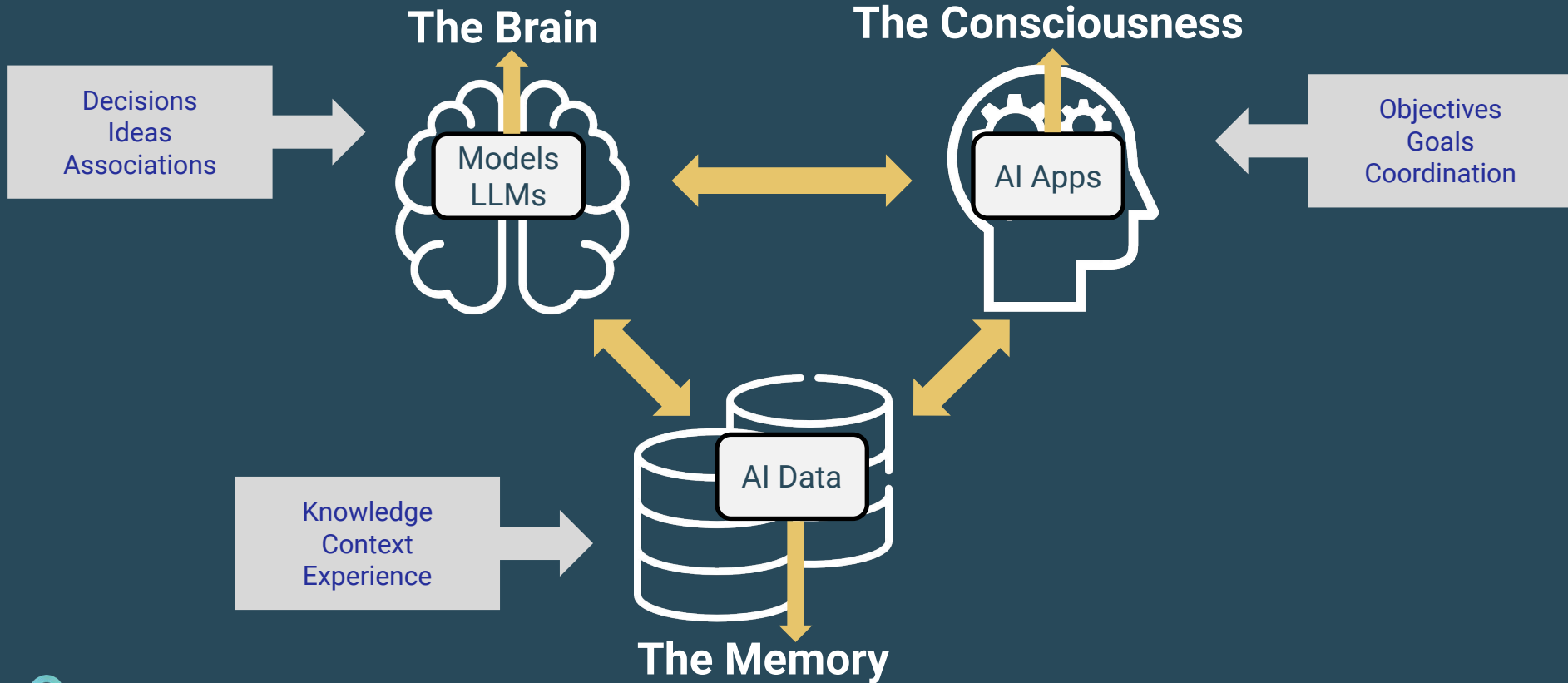
Postgres for the AI Generation

AI

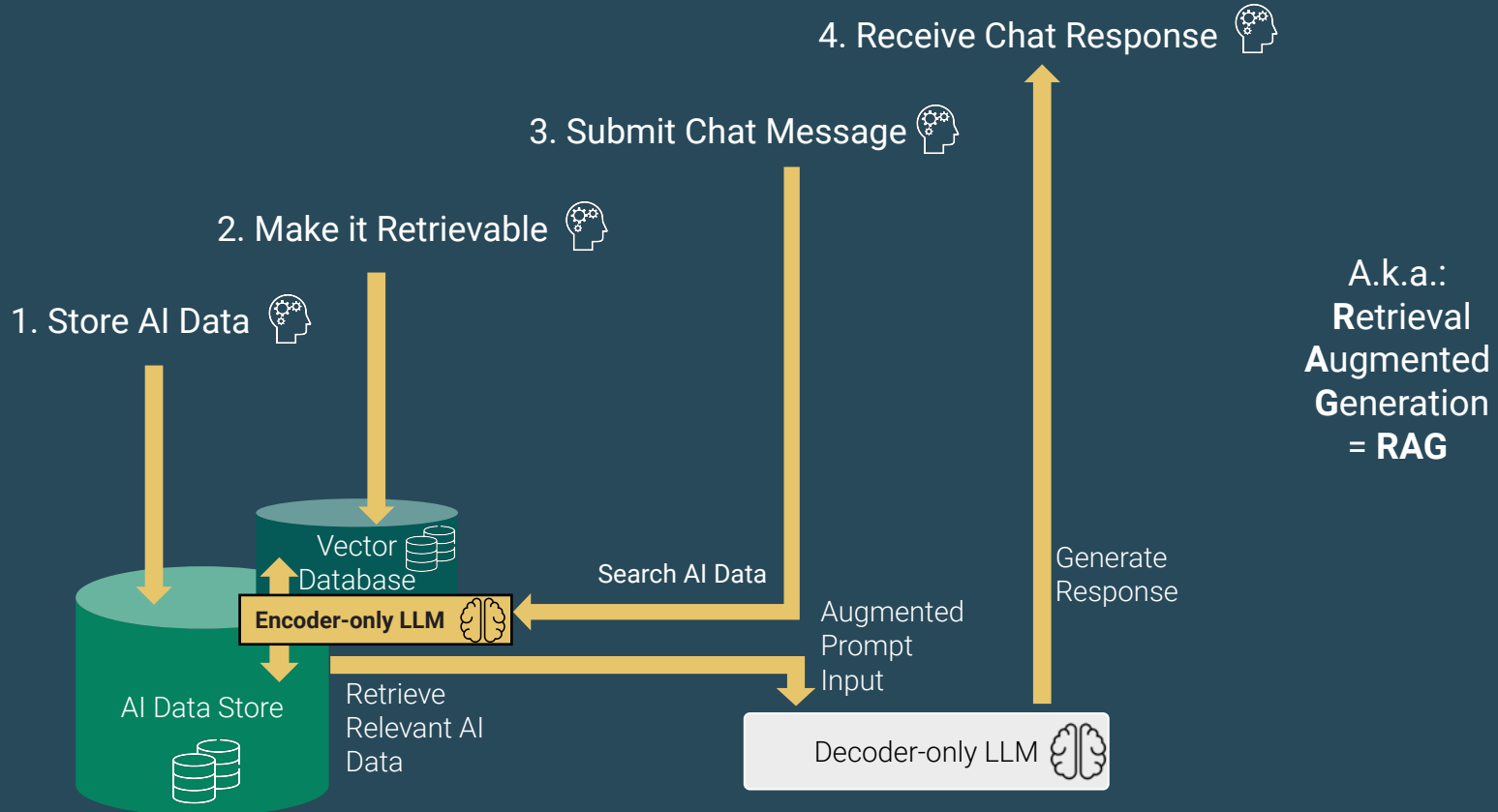
Intelligence



Artificial Intelligence



Chat Bots – The John Doe of Gen AI Applications



POLL

What are your
Generative AI use
cases?



OK, you've built a Chat Bot. Now What?



People love it...



You get more and more users...



People bring more data and more use cases...



People start using it routinely...



Your chat bot becomes **mission critical** for the business!



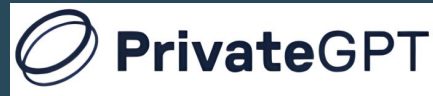
Are you **ready for that??**



Do you have a plan for **sustainable operationalization??**



Growing Number of AI Application Frameworks



- **Popular** approach to **simplify AI solution development**
 - Orchestrating & automating complex AI application flows
 - **Hide** a lot of AI processing **complexity** & **rapid solution development**
- AI Application Framework \leftrightarrow **Data Management Framework**:
 - AI frameworks treat data only as connected data sources
 - You must manage own data storage, repositories and databases



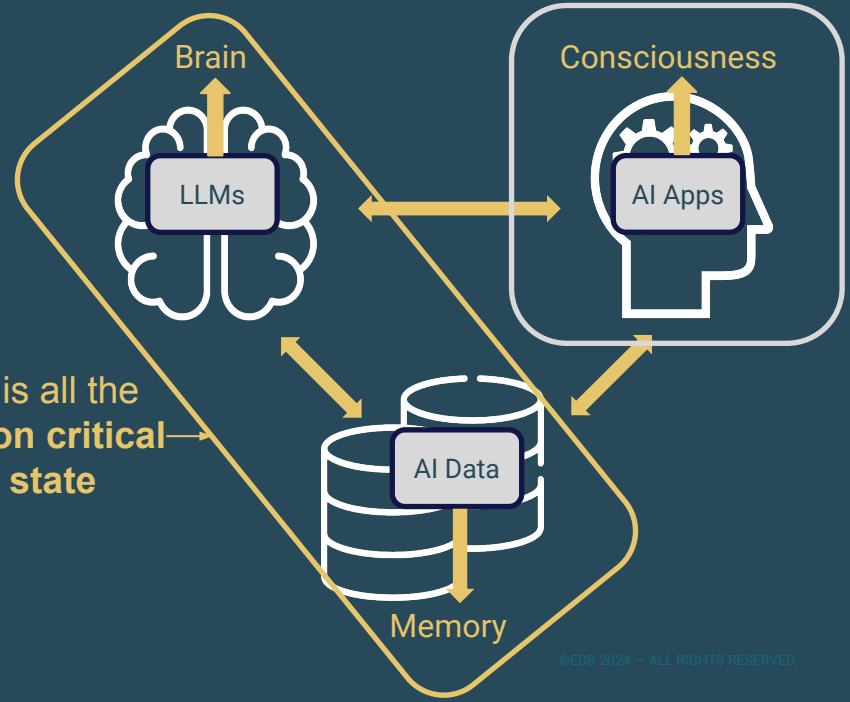
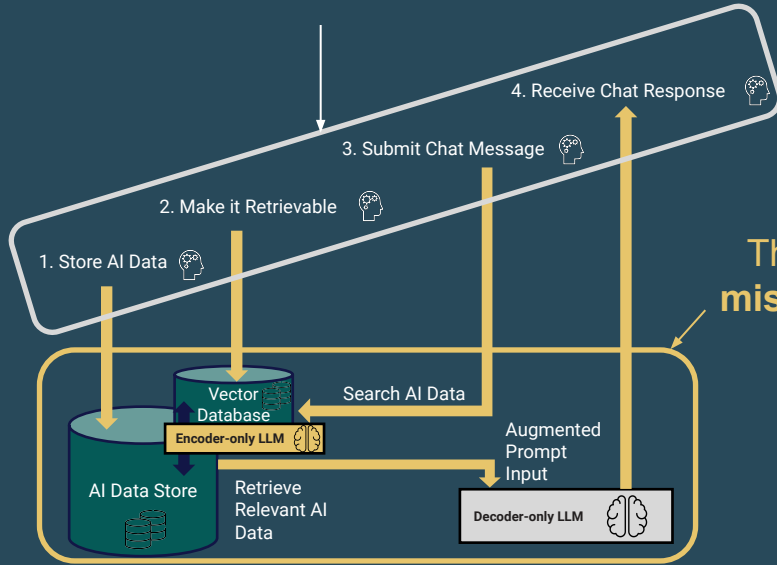
Do LangChain & friends help to

It's The Data,
Stupid!



Not really! They simplify building the AI application only

They don't operationalize the solution,
they only automate your app flow

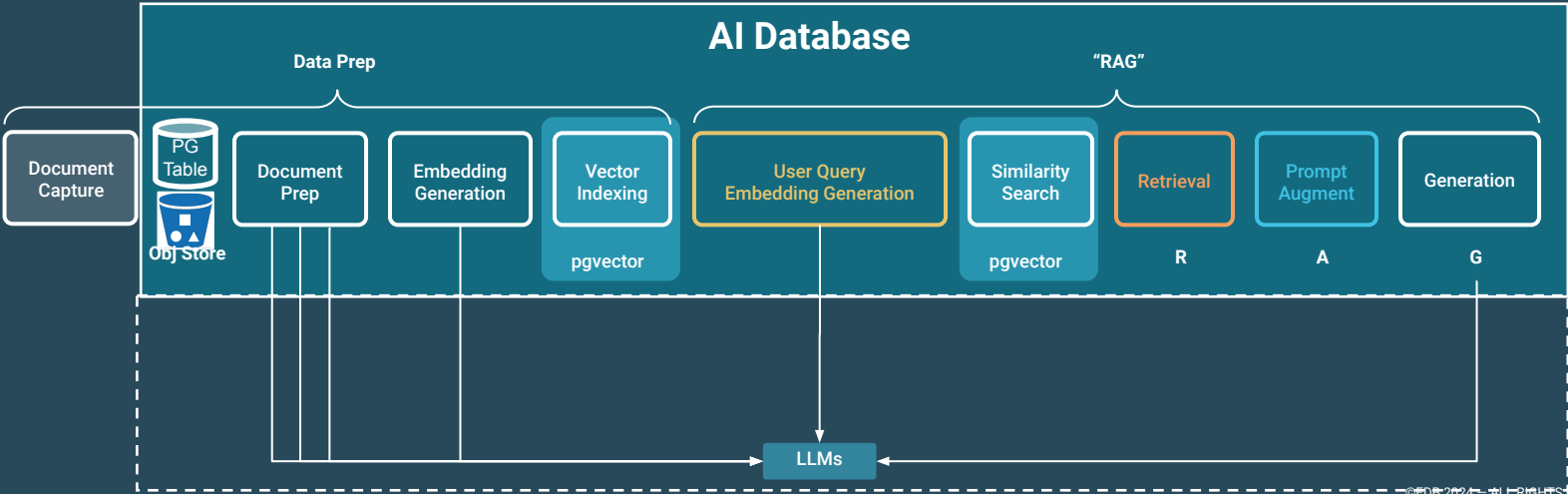


Building GenAI applications with EDB Postgres AI

BEYOND VECTOR SUPPORT

1 Postgres as GenAI Retriever & Generator:
Automating document (and other modalities) prep, embedding generation & vector indexing, providing a simple semantic retriever interface, and even chat completion in database

2 Enabling Sovereign AI for enterprises:
Runs with either, embedded LLMs (in PG memory), external model provider of your choice, or EDB Postgres AI platform hosted models.



POLL

Which vector
databases are you
using or evaluating?



Building AI Solutions with a Vector Database



AI Builder

- **Solution-specific** development
- **Prompt** and **context window** management
- Model **fine-tuning** & serving
- AI data **capture**
- Data **generation** with LLM
- **Automation**
- AI **feature** engineering
- AI data **storage**
- AI data **retrieval**
- LLMs for **embeddings**

Vector Database

- **Vector** Storage & Index
- **Vector** Search

Data & AI Experts

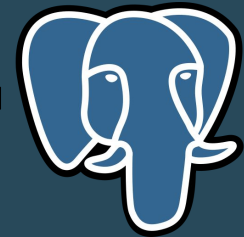


Data Engineer



Data Scientist

Deep
Involvement

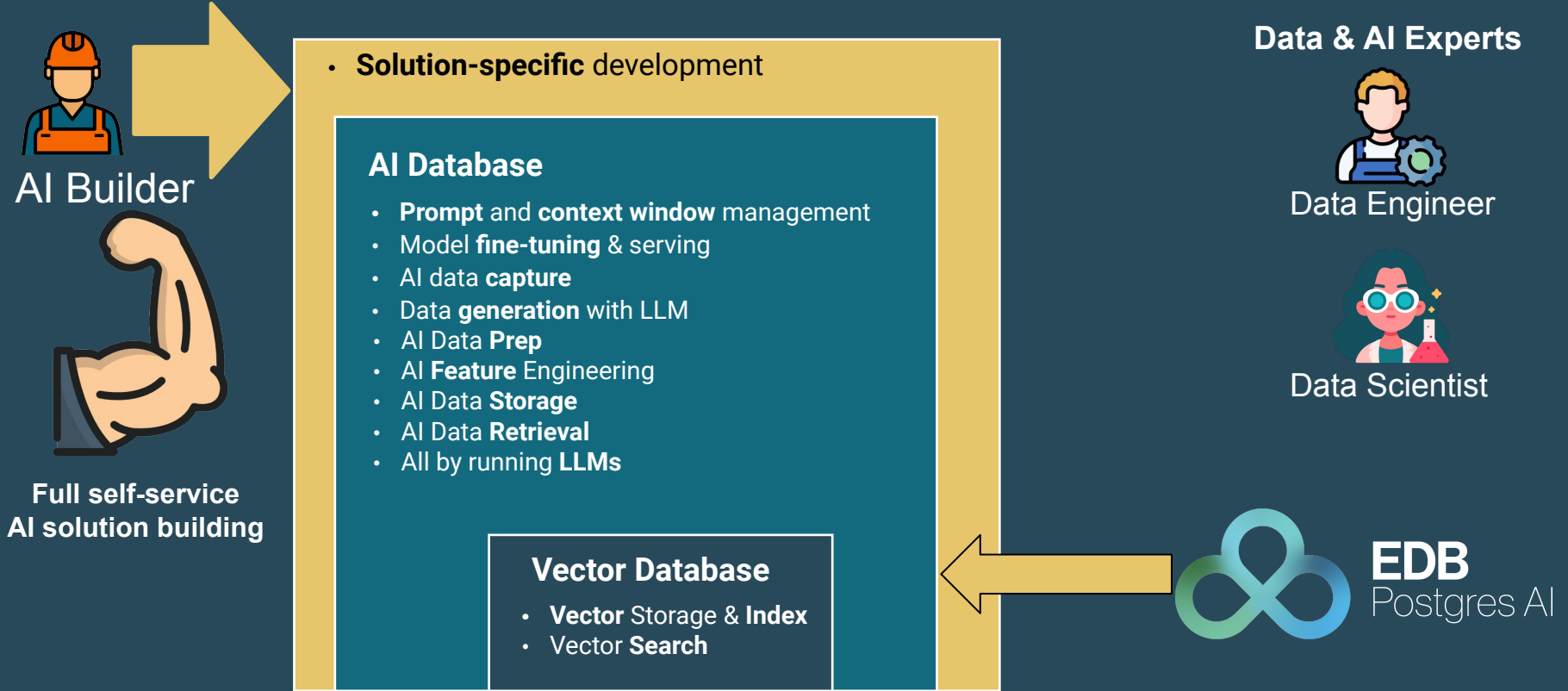


This is
PostgreSQL
today:

PGVECTOR

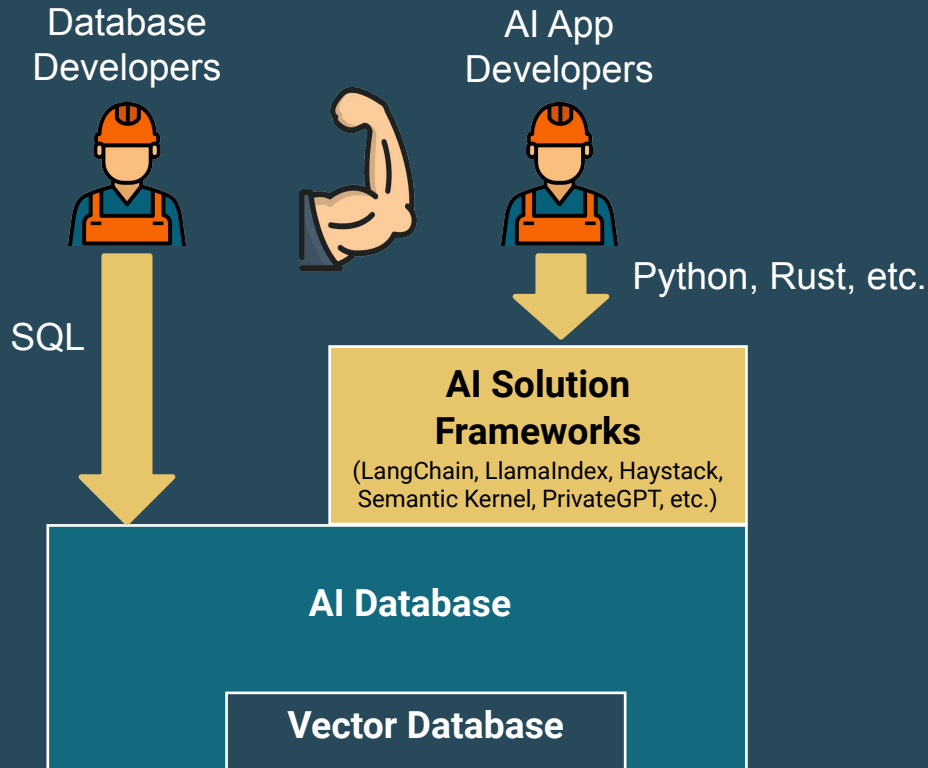


Building AI Solutions with an AI Database



AI Database Consumption Models

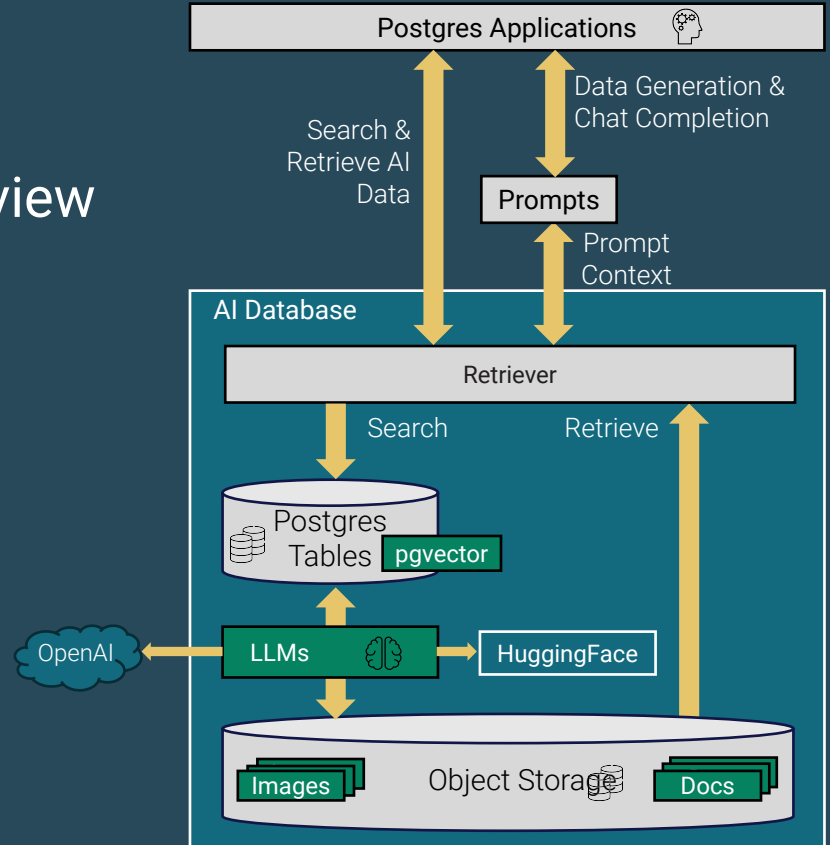
AI Builders:



EDB Postgres AI Database – Tech Preview

Free Access at

<https://info.enterprisedb.com/pgai-preview>



Many Good Reasons to Double Down on Postgres and EDB for Operationalizing AI and Analytic Solutions



Reduce Stack Complexity



Enterprise database with vector capabilities



Simplify AI Data Management



Enhanced Data Utilization



Faster AI Application Deployment





EDB

Postgres® for the AI Generation

This webinar will be available on demand on the **NatWest Content Hub** alongside all of our past dedicated NatWest webinars and content



Thank you

