EDB Postgres® AI Factory

The fastest way to securely build, test, and launch sovereign AI applications.

Borys Neselovskyi, Senior Sales Engineer
Lucie Zeng, Associate Sales Engineer - AI/Analytics Specialist
December 2025

©EDB 2025 — ALL RIGHTS RESERVED.

EDB POSTGRES AI

Agenda

- Introduction
- Vectors and AI Knowledge Base
- EDB AI Factory
- Q&A

©EDB 2025 — ALL RIGHTS RESERVED.

Introduction

©EDB 2025 — ALL RIGHTS RESERVED.

POLL
How would you rate your experience with Postgres & AI?

- **New to Postgres & AI:** Have not used Postgres or AI
- **Postgres user, AI novice:** Know Postgres but not used AI workflows, dabbled with ChatGPT?
- **AI user, Postgres novice:** Worked in AI / LLMs but not with Postgres
- **Familiar user:** Have built simple AI integrations (e.g. embedding store in Postgres, familiar with pg_vector)

POLL

What best describes your experience level in building and deploying AI applications?

- **Beginner/Just Exploring:** Still learning the fundamentals (e.g., Python, basic models)
- **Intermediate:** Built a few prototypes or hobby projects (e.g., simple ML models, using APIs)
- **Experienced:** Actively developing/deploying AI apps professionally or as advanced side projects (e.g., fine-tuning LLMs, MLOps)
- **Expert/Specialist:** Deep knowledge and leadership in the field (e.g., research, architecting large-scale AI systems)
- **Not Applicable:** I'm interested in AI, but not in development

POLL

What is the single biggest challenge you face when developing AI applications?

- **Data Issues:** Gathering, cleaning, or labeling enough high-quality data
- **Model Performance/Accuracy:** Getting the model to perform reliably in real-world scenarios
- **Deployment/MLOps:** Moving the model from training to production and monitoring it
- **Selecting the Right Tools/Frameworks:** Deciding which language, library, or platform to use
- **Lack of Clear Use Cases/ROI:** Defining a valuable problem to solve with AI

From Postgres the database to **Postgres the platform**

Postgres is the most desired database

35% of enterprises (150+ employees) will consider Postgres for their next project.*

EDB remains the leading contributor to Postgres

150+ database engineers

2 core team members

6 committers

20+ named contributors

Transactional Postgres is not enough in today's market

Moving beyond transactional: AI & analytics are key

New lakehouse & **AI** capabilities for analytics and Gen AI development. Seamless synchronization of data from transactional stores to lakehouse

New Hybrid by design for enterprise customers

New hybrid control plane streamlines provisioning, operations, and maintenance with a consistent user experience across on-premise and public cloud. **New form factors** support on-premises, hybrid, and public cloud environments.

EDB POSTGRES AI

*Source: March 2024, EDB, Brand Study.

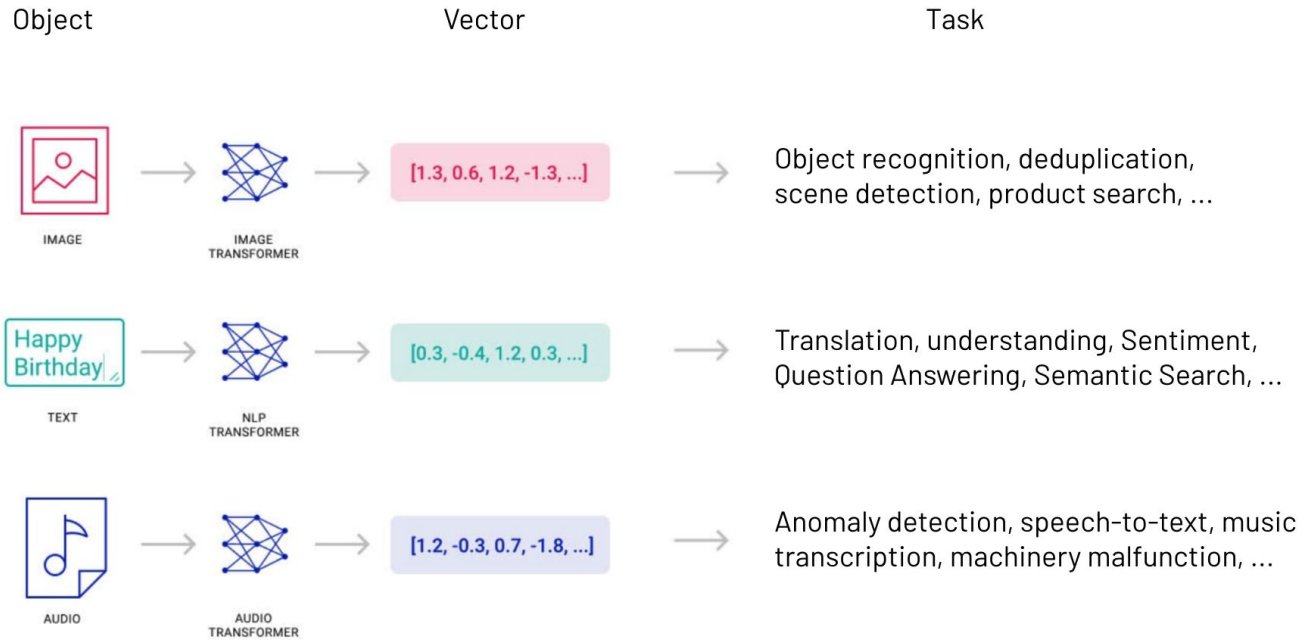
©EDB 2025 — ALL RIGHTS RESERVED.

Intelligent Knowledge Base

EDB POSTGRES AI

©EDB 2025 — ALL RIGHTS RESERVED.

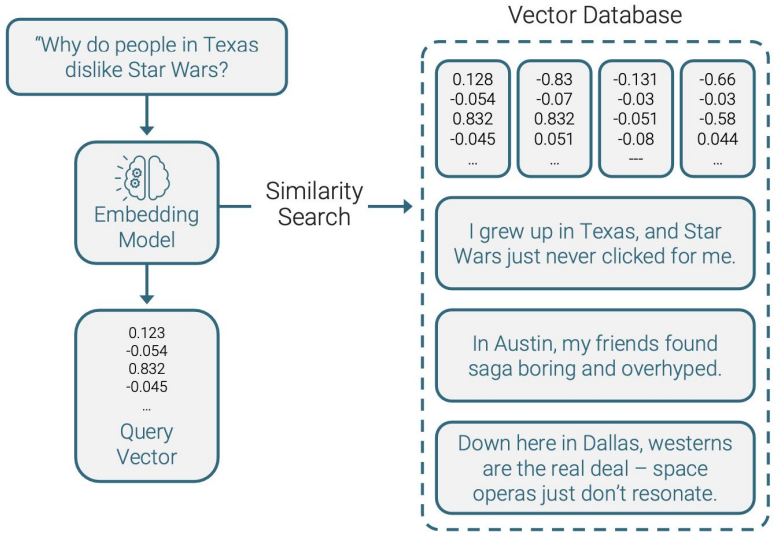
Vector Embeddings: Example

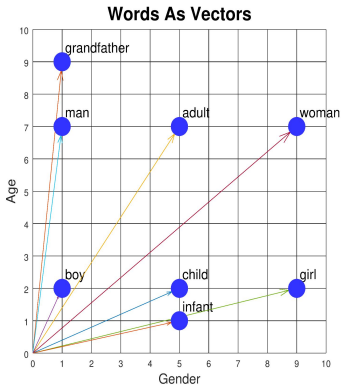


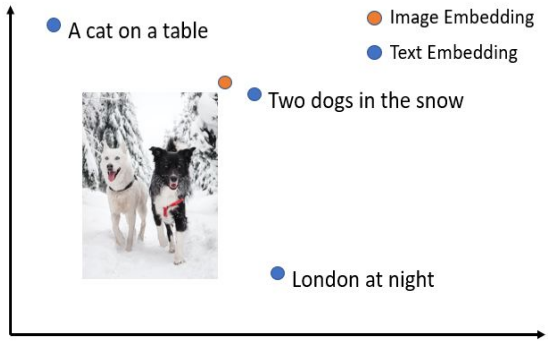
©EDB 2025 — ALL RIGHTS RESERVED.

Vector Embeddings: Example

- Embeddings convert data (like text or images) into a numerical vector.
- This vector captures the item's meaning and relationships.
- Similar items are placed close together in a numerical space.

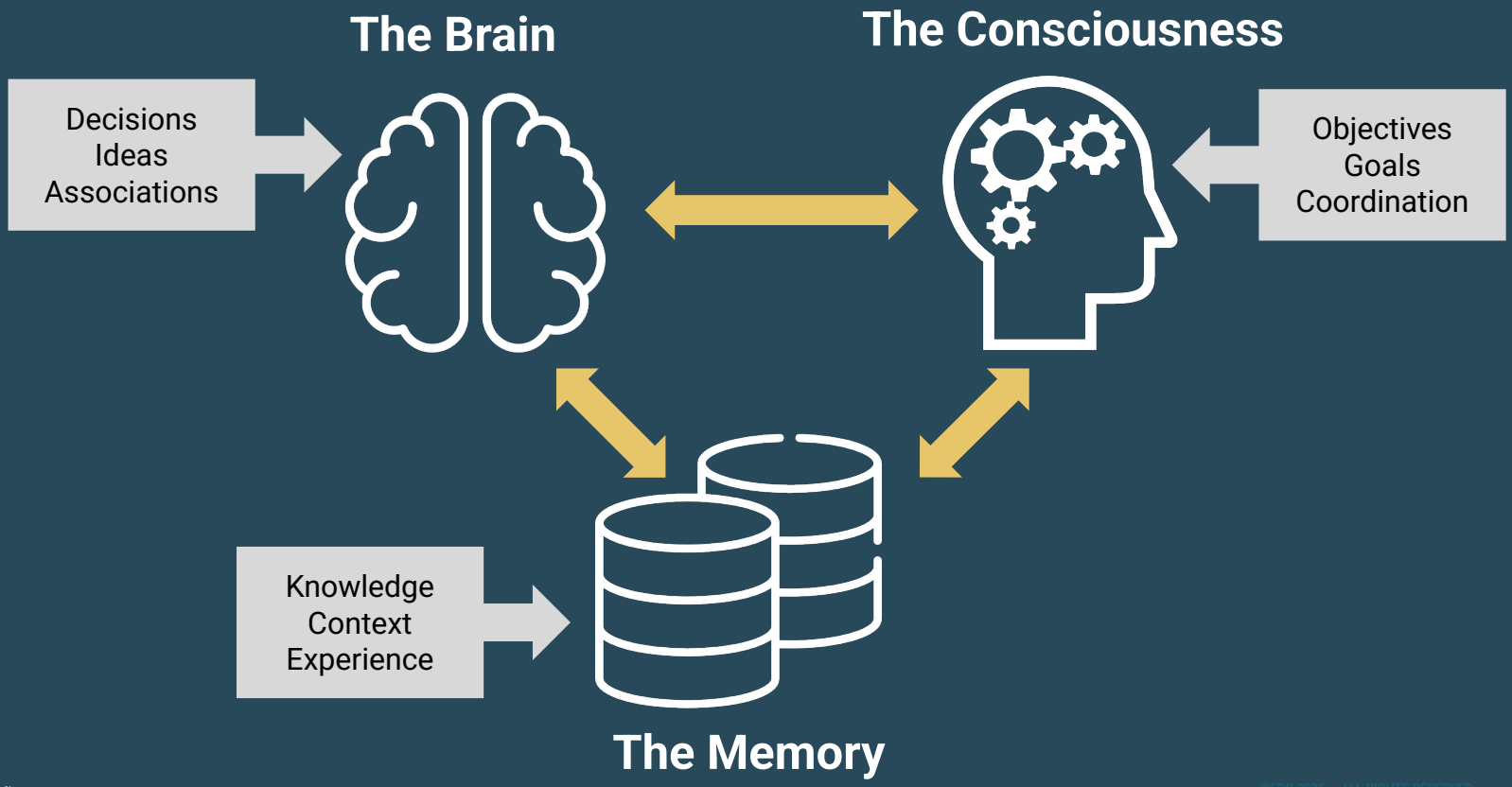






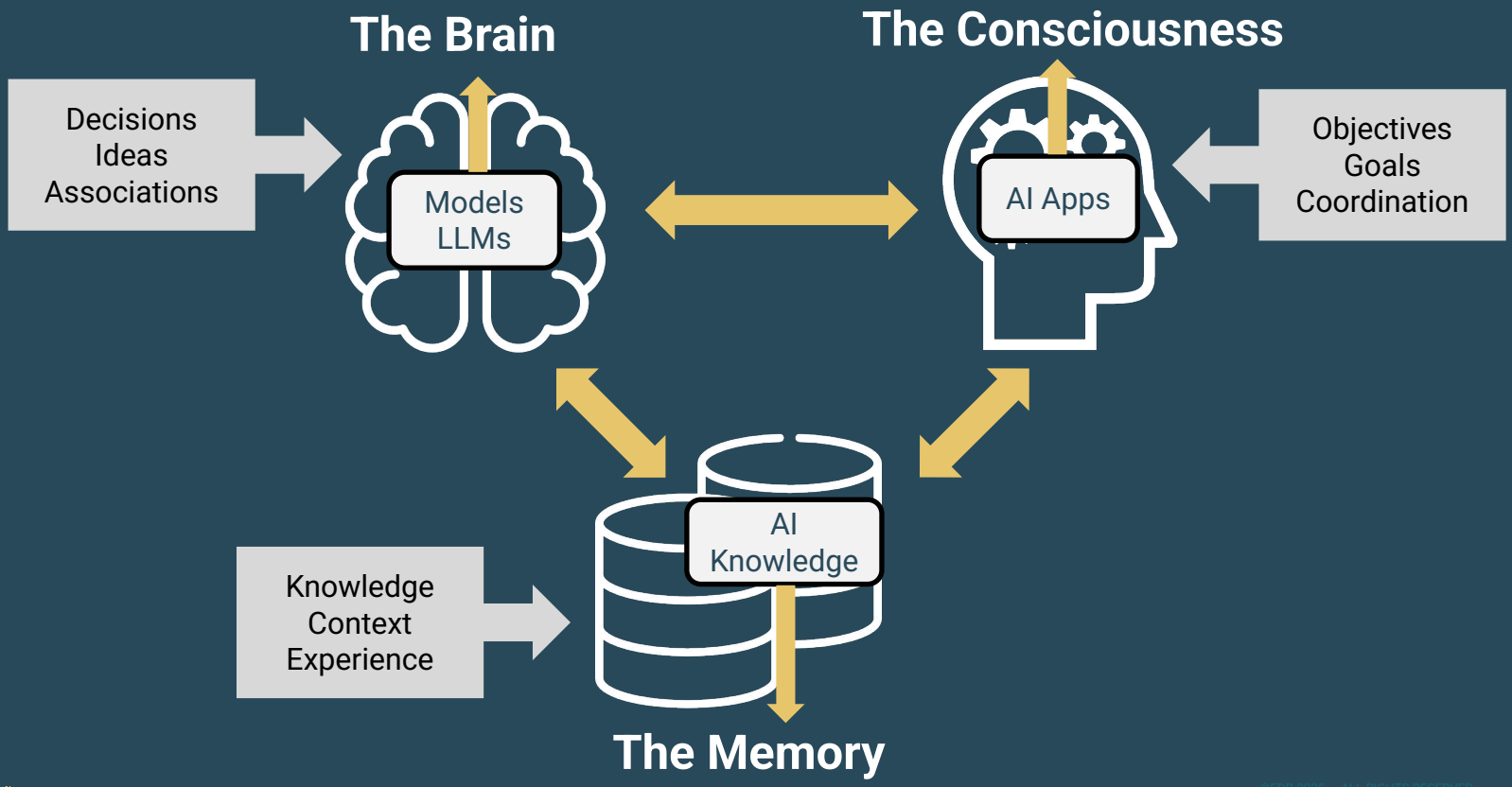
©EDB 2025 — ALL RIGHTS RESERVED.

Intelligence



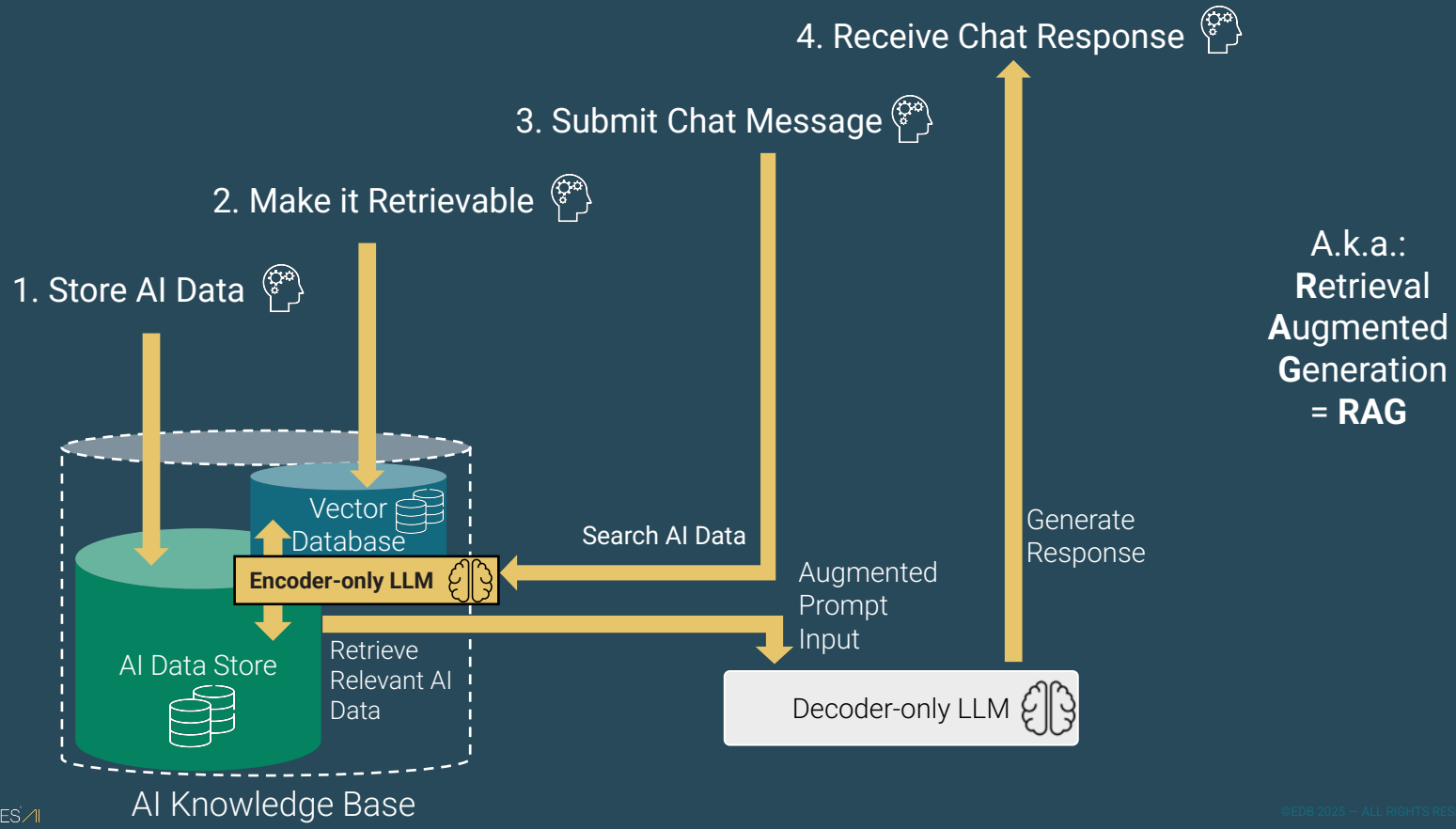
EDB POSTGRES AI
©EDB 2025 — ALL RIGHTS RESERVED.

Intelligence



EDB POSTGRES AI
©EDB 2025 — ALL RIGHTS RESERVED.

Chat Bots – The John Doe of Gen AI Applications



A.k.a.:
Retrieval
Augmented
Generation
= **RAG**

©EDB 2025 — ALL RIGHTS RESERVED.

pgvector Demo

©EDB 2025 — ALL RIGHTS RESERVED.

What is a Knowledge Base?

A Knowledge Base is an indexed store of content optimized for:

Semantic Search	RAG	Hybrid Queries	Explainability
Find relevant content by meaning, not keywords	Enrich LLM outputs with trusted, up-to-date knowledge	Combine metadata filtering with semantic search	Trace responses back to original content sources

It acts as an optimized layer between your raw content and your AI Agents / Assistants.

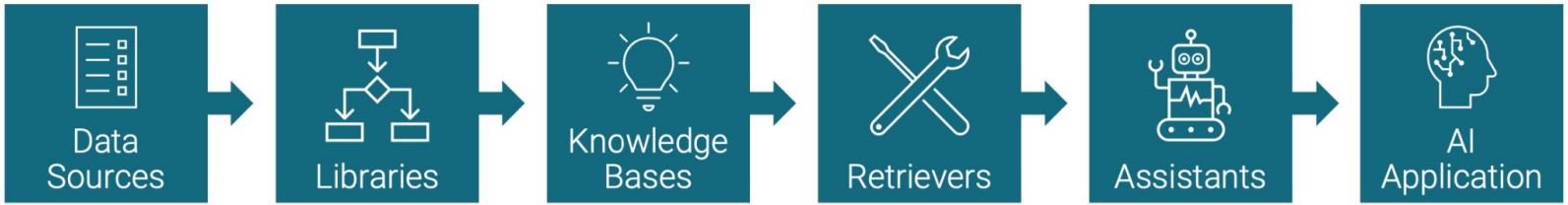
©EDB 2025 — ALL RIGHTS RESERVED.

POLL

When building a GenAI application with a private Knowledge Base (RAG), what is your biggest concern or challenge?

- **Data Synchronization:** Keeping the relational data (metadata) and vector data updated together
- **Tooling Sprawl:** Managing and securing multiple database systems (relational DB + separate vector DB)
- **Retrieval Performance:** Optimizing vector search speed (latency) for a fast user experience
- **Prompt Engineering:** Getting the LLM to use the retrieved context accurately and reliably
- **I haven't built a GenAI application with a knowledge base yet**

Pipeline Flow



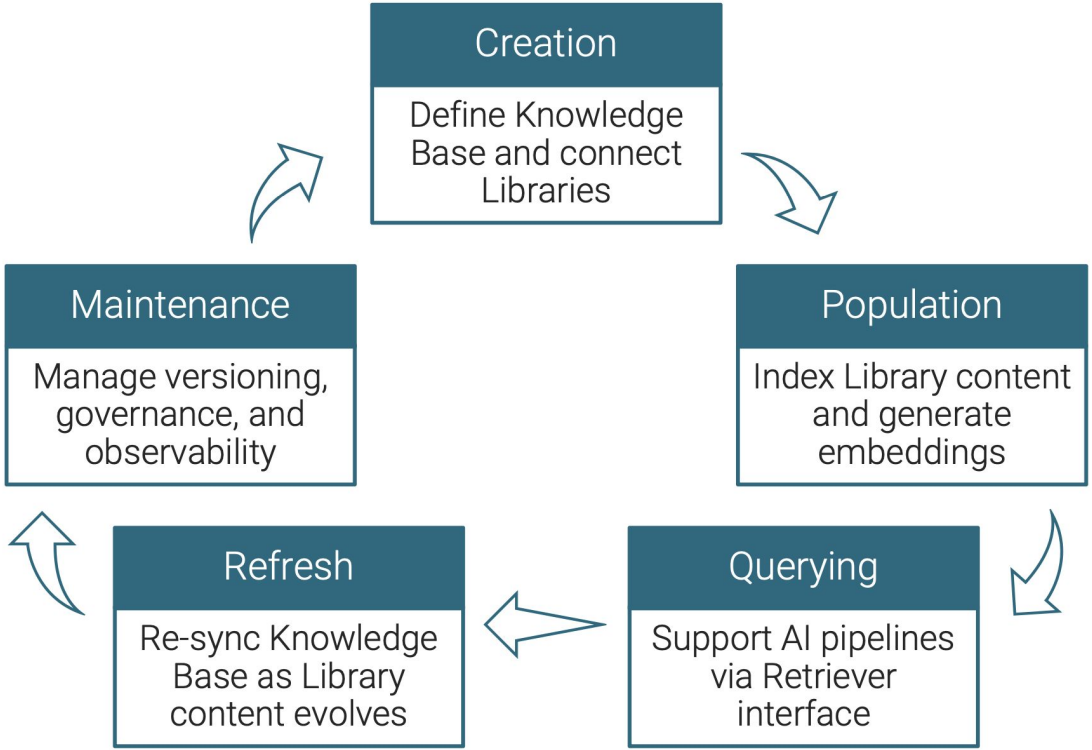
Components:

- Data Sources — Raw content (documents, web pages, APIs)
- Libraries — Processed and structured content collections
- Knowledge Bases — Indexed, query-optimized semantic layer
- AI Applications — Agents, Assistants, APIs, user-facing apps

EDB POSTGRES AI

©EDB 2025 — ALL RIGHTS RESERVED.

Lifecycle of a Knowledge Base



EDB POSTGRES AI

©EDB 2025 — ALL RIGHTS RESERVED.

Need for Automation and acceleration

©EDB 2025 — ALL RIGHTS RESERVED.

Introducing AIDB: Seamless AI Data Management with PostgreSQL

Postgres Extension

AIDB is a Postgres extension designed to integrate AI data seamlessly into your PostgreSQL database.

Manage AI Data

AIDB enables you to process, search, retrieve, and interact with a Large Language Model (LLM) directly within your PostgreSQL database.

Built on pgvector

AIDB is built on top of PostgreSQL's vector data support with pgvector, allowing you to leverage the power of vector-based search and retrieval.

SQL Statements

With AIDB, you can manage and integrate AI data using standard SQL statements, making it easy to work with AI data within your existing database workflows.

EDB POSTGRES AI

©EDB 2025 — ALL RIGHTS RESERVED.

How does AIDB do it? One line of SQL!

Automatically generate and store embeddings for your existing data using a single line of SQL.

Set up a continuous process to automatically create and store new embeddings as your data evolves.

Build custom Retrievers by choosing models tailored to your specific data and use case.

Leverage AIDB's advanced search capabilities to perform both similarity-based searches and hybrid searches that combine multiple criteria.

Seamlessly integrate the retrieved data from AIDB with large language models to generate answers. All inside the Database.

EDB POSTGRES AI

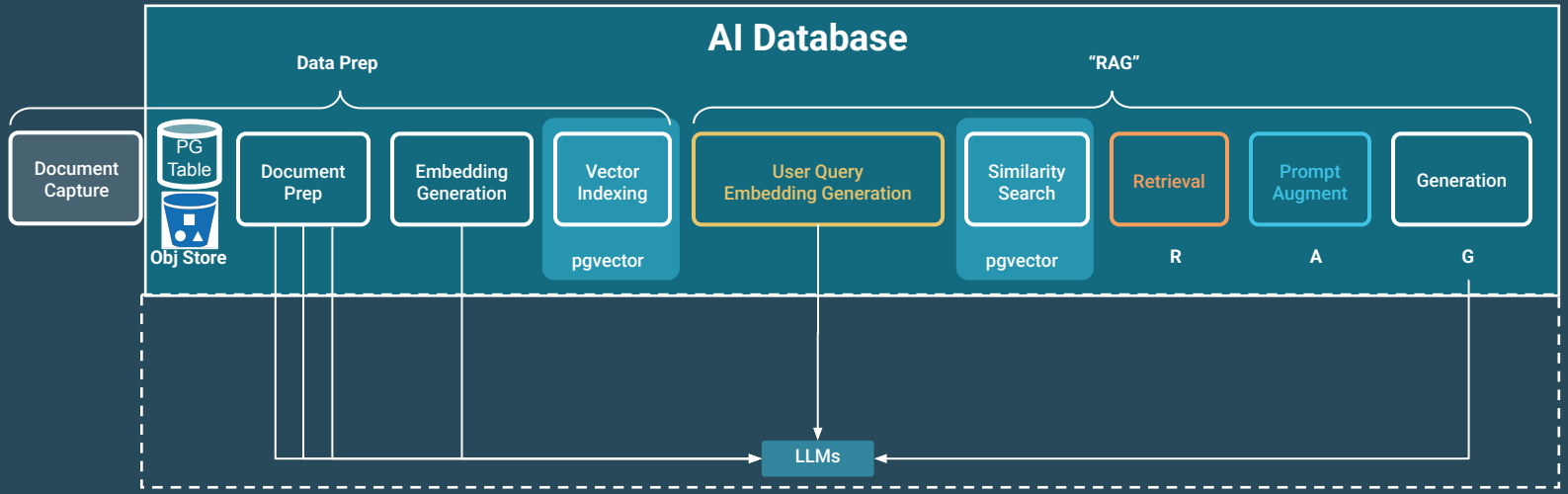
©EDB 2025 — ALL RIGHTS RESERVED.

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.



AIDB Demo

©EDB 2025 — ALL RIGHTS RESERVED.

EDB AI Factory platform

EDB POSTGRES AI

©EDB 2025 — ALL RIGHTS RESERVED.

EDB Postgres AI

Database

- ENTERPRISE POSTGRES SERVER
- ORACLE COMPATIBLE SERVER
- COMMUNITY POSTGRESQL SERVER
- MULTI-MODEL EXTENSIONS
- MANAGEMENT & OBSERVABILITY
- KUBERNETES OPERATORS
- SUPPLY CHAIN SECURITY
- MIGRATION TOOLS
- HIGH AVAILABILITY

Analytics Accelerator

- ANALYTICS ENGINE
- LAKEHOUSE CONNECTOR
- WAREHOUSEPG (OSS, MPP DATAWAREHOUSE)

AI Factory

- VECTOR ENGINE
- AI PIPELINE
- GENAI BUILDER
- AGENT STUDIO
- MODEL SERVING

Hybrid Management

- HYBRID OBSERVABILITY
- HYBRID DBAAS
- DISTRIBUTED HA (99.999%)
- MIGRATIONS

Deploy Anywhere

- SOVEREIGN DATA AND AI FACTORY
- HYBRID SOFTWARE
- MANAGED PLATFORM
- INTEGRATION PARTNERS: AWS, GCP, AZURE, RED HAT OPENSHIFT, IBM, SUPERMICRO, NVIDIA

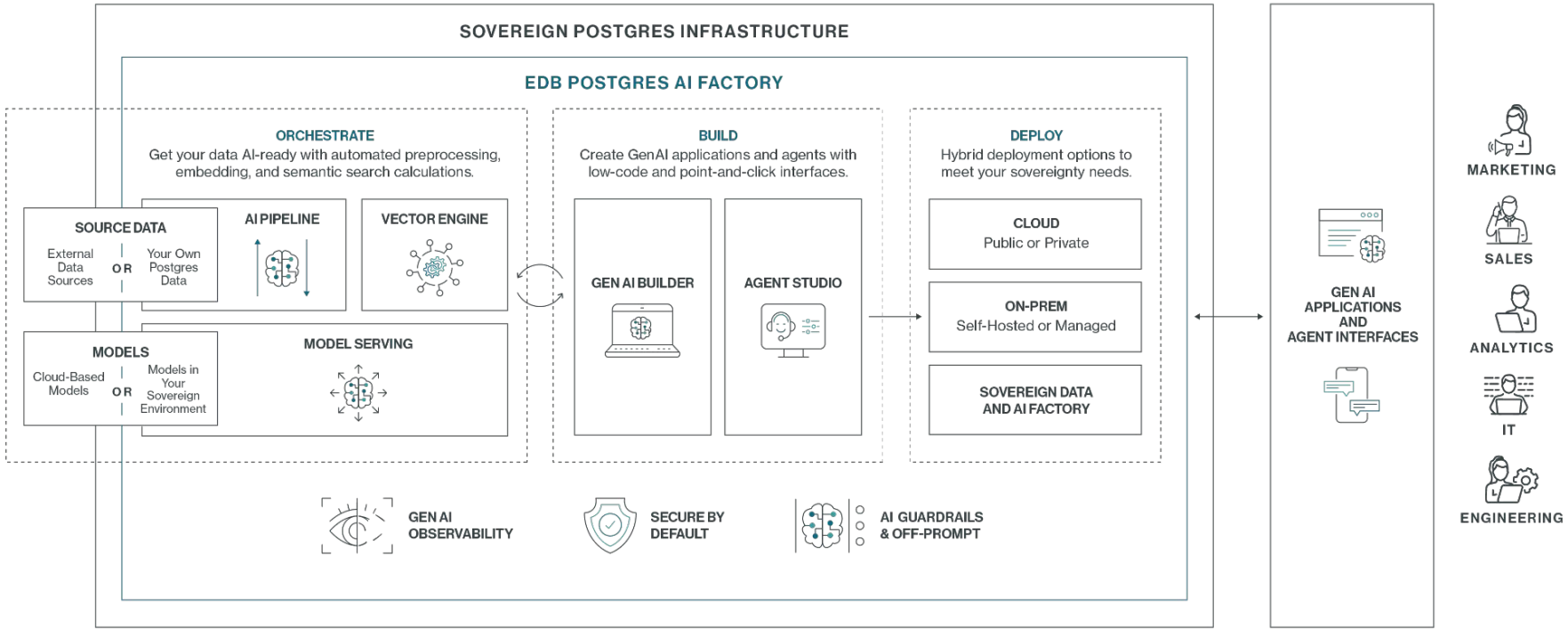
Sovereign Assurance

- MANAGED SERVICES
- PILOT TO PRODUCTION SLAS
- OUTCOME EXECUTION

EDB POSTGRES AI

©EDB 2025 — ALL RIGHTS RESERVED.

AI Factory for Secure Sovereign AI Applications



EDB POSTGRES AI

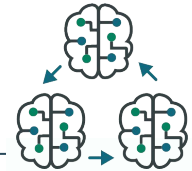
©EDB 2025 — ALL RIGHTS RESERVED.

EDB Postgres AI Factory

WITH FEATURES THAT MAKE CUSTOM, CONTROLLED GENAI ACCESSIBLE FOR EVERY TEAM

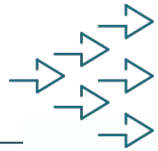
GenAI Builder

- Native MCP Support
- Project workspaces
- Data catalogue integration
- More granular access controls
- Point-and-click interface
- Enterprise security and accuracy
- 3x faster development



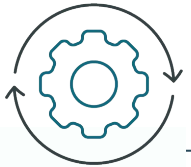
Agent Studio

- Integrated visual genAI workflow designer system
- Deploy AI agents
- Start with open source templates
- Or tailor to your business needs
- Native agent tools for action
- Agent monitoring



AI Pipeline

- Visual AI Pipeline designer
- GPU Acceleration
- Reduce coding effort
- Set-and-forget AI data management
- Auto embedding
- Familiar SQL



Vector Engine

- Two additional new vector index engine options
- Scale out vector database engine
- Single, secure location
- Complete data sovereignty with semantic search
- Open source and Commercial vector types
- Unified data access

Model Serving

- Air gapped model serving
- Bring your own model
- Guardrails models
- Visual auditing
- Eliminate vendor lock-in
- Ability to swap between models
- Maximize hardware ROI
- Flexible deployment

EDB POSTGRES AI

©EDB 2025 — ALL RIGHTS RESERVED.

EDB AI Factory Core Use Cases

Retrieval-Augmented Generation (RAG)

Real-Time Model Inference APIs

Conversational Assistants and Chatbots

Semantic Search Across Enterprise Content

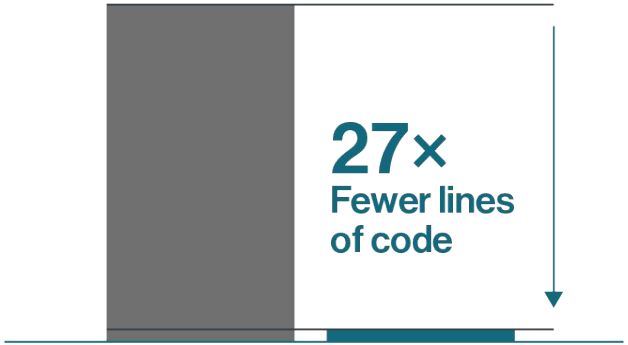
Automated AI-Powered ETL Pipelines

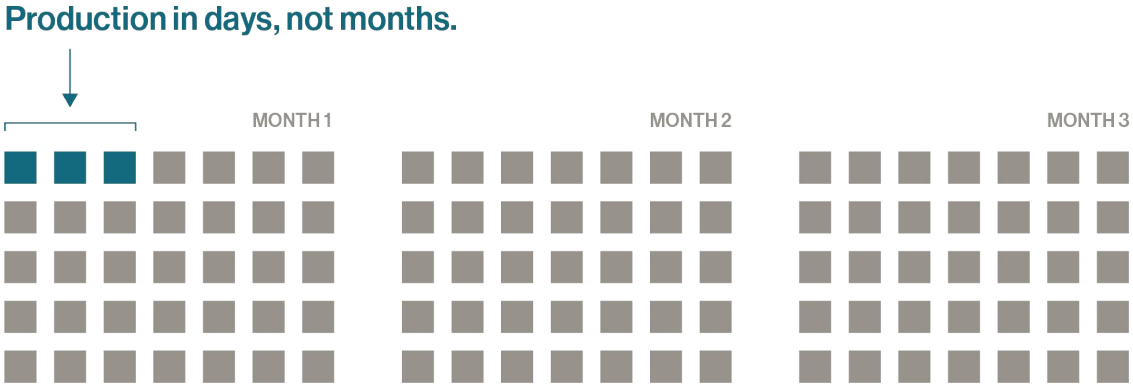
Document Intelligence Pipelines

©EDB 2025 — ALL RIGHTS RESERVED.

Business impacts from EDB Postgres AI Factory

- Turn **every employee** into an GenAI developer
- Enable production GenAI applications in **days instead of months**
- **5 lines of code instead of 135+** to enable automated AI data pipelines
- **3x faster** time to market vs DIY with AWS





©EDB 2025 — ALL RIGHTS RESERVED.

GenAI Builder Demo

©EDB 2025 — ALL RIGHTS RESERVED.

POLL

Which type of Generative AI application are you currently most focused on building or exploring?

- **Internal Knowledge/Doc Chatbot (RAG):** Answering questions using proprietary internal company documents
- **Code Generation/Developer Tools:** Using LLMs for code completion, debugging, or automated testing
- **Advanced Customer Service:** Building conversational AI/chatbots for external customer support
- **Content/Media Generation:** Creating marketing copy, summarizing reports, or generating images/video
- **I haven't started building a GenAI application yet**

Explore resources

Check out the EDB Postgres AI Factory web page for more details

- Dive deeper into EDB Postgres AI use cases:
  - Sovereign AI
  - Cognitive AI
  - Virtual Expert
- Discover more resources about AI Factory:
  - Blog: Take Your First Steps with EDB Postgres AI Factory
  - Webinar: Transforming Contact Centers with AI in the Financial Services Industry
  - Blog: Building a Future-Proof AI Foundation with EDB Postgres AI
  - White Paper: Solving the Vector Database Dilemma: One Platform, 4x Performance, 68% Faster AI Deployment

©EDB 2025 — ALL RIGHTS RESERVED.

Q&A