CDO BRIEF

Own the Full Data Lifecycle: From Source to Insight

For chief data officers, the challenge is rarely a shortage of governance frameworks. It is that the data estate those frameworks are meant to cover is not fully under the organization's control. Operational data lives in Postgres®. Historical analytics run in a warehouse. Real-time workloads, machine learning pipelines, and archival data each sit in a separate system, managed by a different vendor, often in a proprietary format the organization did not choose. A fragmented estate like this taxes a CDO in four ways at once: spend that climbs unpredictably, insight that lags behind reality, AI agents that cannot reach the full picture, and governance that has to be rebuilt at every boundary. Each is a symptom of the same root cause: The lifecycle is not owned end to end.

EDB Postgres AI (EDB PG AI) closes that gap with a converged analytics approach. Rather than treating unification as an integration project and governance as a downstream correction, it brings the full lifecycle—transactional data, real-time analytics, petabyte-scale history, and AI agent access—onto a single open foundation the organization owns and deploys where it chooses. Ownership becomes a structural property of the architecture: Costs stay predictable, insight stays current, agents see the whole estate, and security stays anchored to a governed Postgres core the organization controls.

> 95% of enterprises want to be their own AI and data platforms. Only 13% are thriving—because they bring their transactional, analytical, and AI workloads together. Those that do see 5x the ROI.

The problem: The cost of a fragmented estate

Traditional analytical architectures split the lifecycle across systems that were never designed to work as one, and the seams are expensive. Data moves between them through extract, transform, load (ETL) pipelines, and data engineering teams spend an average of 44% of their time maintaining those pipelines rather than building analytical value. Every pipeline is a point where insight goes stale, cost accrues, lineage fragments, and governance policies drift. Every new engine added to the estate widens the surface without widening control. For the CDO, this shows up as four compounding pressures: unpredictable consumption-based spend, time-to-insight measured in days, AI initiatives blocked by fragmented or ungoverned data, and a governance estate that grows in complexity faster than it can be enforced.

The solution: Unifying the data lifecycle changes the control equation

At the center of EDB PG AI is the Postgres Analytics Accelerator extension (PGAA), which continuously publishes live Postgres data to a shared open-format Apache Iceberg catalog, with no ETL pipelines, no batch windows, and no impact on transactional performance. Every analytical engine reads from that one catalog. ClickHouse® delivers sub-second queries on live event data. WarehousePG handles petabyte-scale batch analytics, complex reporting, and machine learning. Both are exposed to AI agents through the Model Context Protocol (MCP) server. Because there is one shared source rather than many independent copies, everything a CDO cares about – cost, speed, agent access, and control – is decided once at the foundation instead of relitigated at every system boundary.

Own the estate

EDB PG AI removes the constraints that put most analytical estates on someone else's terms. It starts with two that travel together: the format the data sits in and the engine that reads it. Open formats are only half the equation. Several platforms now store data in open table formats but still run it on a proprietary engine. The organization can read its files anywhere yet remains dependent on one vendor's runtime, pricing, and roadmap to actually query them. EDB PG AI keeps both open – the format and the engine. The engines are open source, using Postgres, WarehousePG, and ClickHouse, the same software enterprises already run independently. So is their storage, from each engine's own open format to the shared Apache Iceberg catalog PGAA publishes to, a vendor-neutral open standard no commercial vendor can alter. Nothing in the estate sits in a proprietary format. Format ownership means the organization can read, query, and migrate its own data without a vendor's permission; engine ownership means it is never locked to a single runtime to do so. And the estate isn't capped at those three. Because the catalog is open Iceberg, any compatible engine can read it, so teams can bring their own, pointing existing Spark and ML workflows at the same data. EDB PG AI's engines are the supported core, not a ceiling.

Deployment control completes the picture. EDB PG AI runs on-premises, in sovereign cloud regions, or in hybrid configurations, so where data lives follows organizational policy, not vendor terms. And because the estate is anchored on Postgres, it rests on a governed core: Access controls and PII masking are defined and enforced right at the Postgres source. That means security lives in one place the organization owns—the same place data-residency requirements such as GDPR and HIPAA are met.

Run on a cost model you control

Most analytical spend is hard to predict because the architecture makes it so: You pay per query, storage and compute scale together even when you need only one, and costs climb whether or not the business is growing. Eighty-four percent of organizations struggle to manage consumption-based billing. EDB PG AI instead offers predictable per-core pricing across the stack, so capacity scales without cost surprises. Organizations consolidating on EDB PG AI have achieved up to 58% lower total cost of ownership than cloud data warehouses. Open table formats add further savings—a 5x smaller storage footprint than proprietary alternatives, with object storage up to 18x more cost-efficient than SSD-based storage. Infrastructure cost grows with the business, not with every query an analyst runs or every model a data scientist trains.

Insight that keeps up with reality

Fragmented pipelines mean insight can never quite catch up. Data has to be rebuilt every time it moves, and time to insight is measured in days. EDB PG AI eliminates the pipeline layer entirely with a zero-ETL architecture. Every engine reads from the same live Iceberg catalog, so time to insight goes from days to now and every data event becomes an opportunity to act. ClickHouse serves real-time queries in milliseconds; WarehousePG handles petabyte-scale analytics with no stale snapshots or pipeline failures at any layer. Because the format is open, the same data is available to Spark-based pipelines with no new integration, and for the heaviest workloads query execution can offload to NVIDIA Spark RAPIDS for GPU acceleration – up to 99x faster. Adding an engine means pointing it at the catalog already in place, and the estate grows without growing in complexity.

Give your agents the full picture

An AI agent is only as good as the context it can reach, and in most estates full context is hard to assemble. Connecting an agent to live data is one integration, to historical depth another, and regulated data adds a compliance review on top. EDB PG AI gives agents one path to the whole estate. Through the MCP server—the same standardized interface agents already use—an agent reaches operational records from Postgres, real-time events from ClickHouse, and petabyte-scale history from WarehousePG in a single reasoning loop. And because agents query through the same interface as human analysts, the access controls and masking enforced at the Postgres source already apply. The governance review done for people covers the agents too, with no duplicate approvals. AI velocity comes from doing the governance work once, at the source, not once per application.

Why this is a strategic advantage

Organizations that can put trusted, current data in front of people and agents—on infrastructure they own, at a cost they can predict—move faster, decide better, and deploy AI with more confidence than competitors still reconciling fragmented estates. Ownership of the full lifecycle is what lets all four advantages hold at the same time and at scale. The value is clearest in financial services, healthcare, and public sector organizations, where the cost of getting data wrong is highest and the payoff from trusted, sovereign data is greatest.

An evaluation framework for CDOs

1. **Format and engine control: Does the organization own both the format its data sits in and the engine that reads it?** EDB PG AI keeps data in open formats—the open storage of Postgres, WarehousePG, and ClickHouse, plus a shared open Apache Iceberg catalog—and runs it on those same open source engines. Open formats without open engines still leave a vendor in the query path; EDB PG AI ensures every future architectural decision stays in the customer's control.
2. **Deployment control: Can the organization choose where data lives?** EDB PG AI deploys on-premises, in sovereign cloud regions, or hybrid—full stack—so where data lives follows organizational policy, not vendor infrastructure terms.
3. **Cost model: Is spend predictable as the estate scales?** Per-core pricing across the stack replaces consumption-based billing, so capacity scales without cost surprises.
4. **Speed and unification: How current is the data, and what does it cost to add capacity?** A shared zero-ETL catalog delivers real-time through historical from a single source, and adding an engine adds no pipelines to build or maintain.
5. **Agent readiness: Can AI agents reach the whole estate through one interface?** Through the MCP server, an agent queries operational records from Postgres, real-time events from ClickHouse, and petabyte-scale history from WarehousePG in a single reasoning loop, with no custom per-system integration.
6. **Governance and control: Does the platform give you one governed core, or a separate policy layer to maintain per system?** EDB PG AI centers the estate on governed Postgres. Access controls and PII masking are enforced at the source, so adding an engine does not mean standing up a new governance implementation.

Data and digital transformation imperatives comparison

Imperative	EDB PG AI	Databricks	Snowflake	BigQuery	Stand-alone ClickHouse
Format control	Open by default—Postgres, WarehousePG, ClickHouse, and Apache Iceberg; no proprietary format	Delta Lake (open; Databricks-optimized)	Proprietary by default; Iceberg tables supported	Proprietary by default; managed Iceberg (REST catalog in preview)	Open source; open formats (Iceberg, Parquet)
Engine ownership	Open source engines—Postgres, WarehousePG, ClickHouse	Open source Spark + proprietary Photon	Proprietary engine	Proprietary engine	Open source engine
Deployment	On-prem, sovereign, or hybrid (full stack, incl. real-time)	Cloud only	Cloud only	Cloud only (GCP)	Self-managed or cloud
Cost model	Predictable per-core pricing	Consumption-based	Consumption/ credits	Consumption-based	Self-managed infra or usage-based (Cloud)
Data movement	Zero-ETL via PGAA; no pipelines	Ingestion (Delta Live Tables)	Ingestion (Snowpipe)	Ingestion (streaming inserts)	Ingestion (Kafka / custom ETL)
Agent access	Whole estate via MCP, same policies as analysts	Integrated within platform	Integrated within platform	Integrated within platform	Limited; external tooling
Governance	Enforced at governed Postgres source	Unity Catalog; platform-bound	Platform-bound	IAM + Dataplex; platform-bound	RBAC separate from source; maintained per system
Vendor lock-in risk	Low—open engines + Iceberg	Medium—platform ecosystem	Medium–high; Iceberg reduces storage lock-in	Medium–high; GCP compute-bound	Low (open source)
Primary strength	Unified, owned lifecycle—real-time, historical, and AI on open, sovereign foundations	ML + lakehouse ecosystem	Elastic warehouse simplicity	Serverless analytics	High-performance real-time OLAP
Primary limitation	Requires open-standards, Postgres-centered strategy	Cloud only	Cloud only; proprietary compute	Requires GCP alignment; cloud only	Governance separate from source; no lifecycle unification

Note: Stand-alone ClickHouse refers to ClickHouse deployed independently without EDB integration. Deployed as part of EDB PG AI, ClickHouse reads from the shared Iceberg catalog PGAA keeps continuously current from the governed Postgres source, and deploys within the same sovereign perimeter as the rest of the estate.

Competitive comparisons are based on publicly available information and are subject to change as vendor offerings evolve and new information is made available. All product names, trademarks, and registered trademarks are the property of their respective owners.

ClickHouse is a trademark of ClickHouse, Inc. https://clickhouse.com

EDB Postgres AI: The sovereign data and AI platform for the agentic enterprise

EDB PG AI brings together a unified data layer, governance, sovereign control and orchestration, and an agent runtime environment, giving enterprises a trusted foundation for AI on infrastructure they own and control. The platform unifies transactional, analytical, and AI workloads in a single Postgres-based architecture—eliminating ETL, data movement, and operational fragmentation. And you choose where and how to deploy: on-premises, cloud, managed, or certified appliance.

The outcome: production-ready sovereign AI in days or weeks, not months.

EDB POSTGRES AI

EDB Postgres® AI (EDB PG AI) is the sovereign data and AI platform for the agentic enterprise. Built on Postgres, the world's leading open source database, EDB PG AI unifies transactional, analytical, and AI workloads in a single governed architecture, on-premises and across clouds. To learn more, visit www.enterprisedb.com.