TECHNICAL BRIEF

Converged Analytics

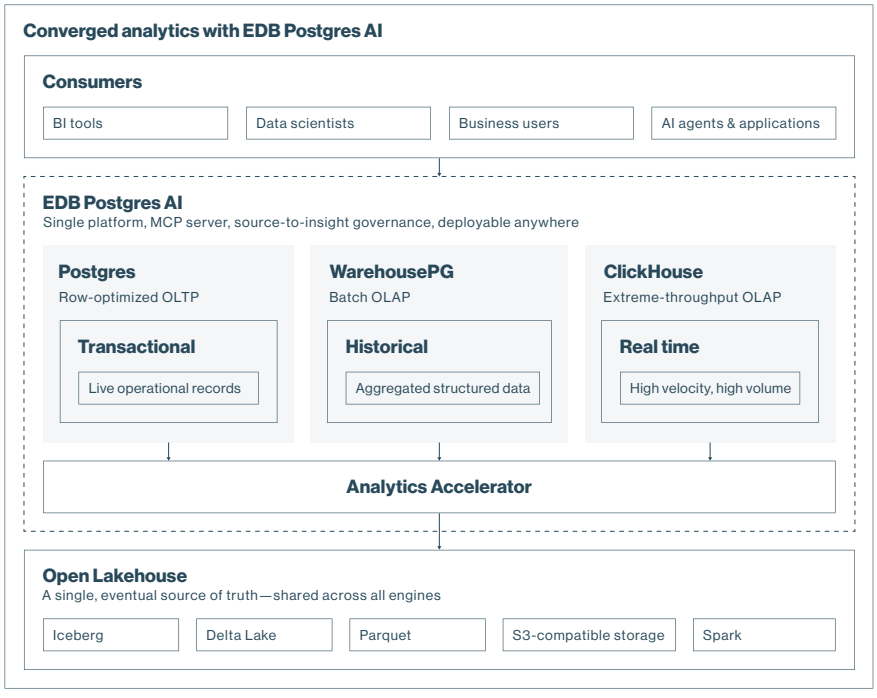
Unify the full data lifecycle on a foundation you control

Solution overview

EDB Postgres AI (EDB PG AI) provides a distributed data architecture designed to unify operational, analytical, and AI workloads. The platform eliminates extract, transform, load (ETL) pipelines by enabling direct, governed access to transactional data in analytical formats using the PGAA (Postgres Analytics Accelerator) extension. Rather than copying data into separate proprietary stores through pipeline-based ETL, it writes once to a shared Apache Iceberg catalog and makes the same current data available to every analytical engine through metadata-driven access paths—eliminating ingestion pipelines and the operational overhead of keeping multiple systems in sync. PGAA works with any standard Iceberg REST Catalog, enabling interoperability with existing data platforms in the organization's ecosystem.

Unlike cloud-native analytics platforms where the processing infrastructure is operated by the vendor, EDB PG AI runs on infrastructure the organization owns and controls. Because every component—Postgres, ClickHouse, WarehousePG—is built on an open source foundation and the platform writes to open Apache Iceberg format, no vendor controls the data format, the query runtime, or the deployment environment. The full stack deploys on-premises, in private or public cloud, or air-gapped, with no dependency on managed cloud services or vendor-controlled infrastructure.

The architecture comprises five integrated subsystems: the Postgres operational data plane, EDB PG AI for ClickHouse, EDB PG AI for WarehousePG, the Apache Iceberg lakehouse layer, and the EDB PG AI Model Context Protocol (MCP). Each subsystem reads from the same catalog without requiring explicit data movement, supporting real-time, historical, and AI workloads from a single Postgres-native foundation.



Core architecture

Postgres operational layer

The Postgres layer is the transactional foundation of the platform—the operational landing zone for live data with full ACID consistency and durability. It serves as the authoritative source of truth for hot operational data, handling write-heavy workloads with complete isolation. Governance policies are enforced here using battle-tested Postgres access and authorization controls: row-level and column-level security, PII masking, and schema-defined access rules enforced at query time.

PGAA and EDB Postgres Distributed

PGAA is a Postgres extension that enables direct, zero-ETL access to large-scale data stored in open table formats—Apache Iceberg, Delta Lake, and Parquet—in object storage, using a vectorized execution engine to offload heavy analytical queries without data movement. Zero-ETL replication from operational Postgres is delivered through PGD, which captures row-level changes via logical decoding and streams them in near real-time to object storage in Iceberg format. ClickHouse and WarehousePG then read from the same Iceberg catalog that PGD and PGAA keep continuously current—eliminating ingestion pipelines entirely.

EDB PG AI for ClickHouse

ClickHouse is an enterprise-supported real-time analytics engine integrated into the EDB PG AI converged analytics stack. Unlike stand-alone ClickHouse deployments that require dedicated ingestion pipelines to populate their data stores, ClickHouse reads directly from the shared Iceberg catalog that PGAA keeps continuously current. This eliminates the ingestion layer entirely and ensures that real-time queries always operate on governed, current data.

ClickHouse is optimized for millisecond-latency responses on large event datasets using columnar storage, vectorized execution, and advanced compression. Within the EDB PG AI architecture, it serves workloads that require sub-second query latency: application log analytics, fraud scoring, live operational dashboards, high-concurrency self-service BI, and observability pipelines.

Key technical characteristics include:

- Reads from shared Apache Iceberg catalog via native Iceberg integration
- Millisecond query latency on event and log datasets
- Columnar storage with vectorized execution for high-throughput aggregations
- Enterprise-support and management within EDB PG AI

EDB PG AI for WarehousePG

WarehousePG is the petabyte-scale batch analytics engine in the EDB PG AI converged analytics stack. Built on Postgres, it distributes query execution across independent compute nodes using a shared-nothing massively parallel processing (MPP) architecture, enabling complex analytical workloads—reporting, historical analysis, joins and aggregations over massive datasets—at consistent throughput under high concurrency, delivering 52% more consistent concurrency performance and 58% lower total cost of ownership than traditional MPP alternatives in benchmark testing. Like ClickHouse, WarehousePG reads from the same shared Iceberg catalog.

WarehousePG handles workloads that require depth and scale rather than millisecond latency: complex BI reports, historical trend analysis, data science, and machine learning (ML) model training. In-database machine learning via MADlib executes analytical algorithms—regression, classification, clustering, text analytics—as distributed SQL functions directly within the MPP cluster, without moving data to external compute. A built-in Foreign Data Wrapper (FDW) enables WarehousePG to query ClickHouse data directly, supporting federated analytics across real-time event data and historical depth through a single SQL interface.

Key technical characteristics include:

- Petabyte-scale shared-nothing MPP architecture built on open source PostgreSQL
- In-database ML via MADlib; Python via PL/Python for data science workflows
- FDW for direct federated queries into ClickHouse
- Deploy on-premises, in sovereign cloud regions, or hybrid

Lakehouse layer (Apache Iceberg)

Where Postgres is the authoritative hot tier for live operational data, the lakehouse layer is the eventually consistent, archival source of truth—continuously synchronized from Postgres via PGD replication and accessible to all analytical engines without data movement.

The lakehouse layer uses Apache Iceberg as the canonical open table format, providing a standardized interface for all analytical workloads. Iceberg supports ACID-compliant table snapshots, enabling consistent reads even as data is continuously updated. Its metadata-driven architecture allows for schema evolution without requiring costly table rewrites, while partition pruning and indexing improve query performance across all engines.

Open formats are central to the architecture's portability guarantees. Apache Iceberg is readable by any compatible engine, and because the engines themselves—ClickHouse, WarehousePG, Postgres—are all open source at their core, no commercial vendor controls the data format. PGAA supports any standard Iceberg REST Catalog, extending that interoperability to any catalog-compatible platform already in use. Because PGAA writes directly into Iceberg tables, the lakehouse layer remains continuously synchronized with operational data, ensuring all analytical engines work from the same current data without duplication.

Analytical execution layer

The analytical execution layer provides right-sized, flexible query execution across the full data lifecycle. Rather than relying on a single engine, the platform offers a spectrum of options—each optimized for a specific workload type and accessible through a consistent SQL interface:

- **Postgres:** Operational query execution with full transactional consistency
- **Apache DataFusion:** In-process low-latency queries for interactive use cases
- **Apache Spark:** For organizations with existing Spark infrastructure, Postgres can offload large-scale distributed processing for bulk transformations and ML pipelines to Apache Spark; optional GPU-accelerated execution with NVIDIA Spark RAPIDS delivers up to 99x faster performance than standard PostgreSQL on analytical workloads
- **ClickHouse:** Millisecond-latency real-time queries for event data, dashboards, fraud scoring, and log analytics
- **WarehousePG:** Petabyte-scale distributed processing for complex reports, historical analysis, and data science workloads

AI access layer

The AI access layer is implemented via MCP server, which provides a standardized interface for AI agents and applications. This layer enables semantic discovery of data, governed query execution, and context-aware retrieval across all engines, including ClickHouse real-time data and WarehousePG historical data, ensuring that AI systems interact with data in a consistent and compliant manner.

By exposing the full analytical estate through a unified interface, the MCP server eliminates the need for custom integrations for each AI application and agent.

Governance architecture

Governance in EDB PG AI is enforced at the Postgres data layer using battle-tested Postgres access and authorization controls. Row-level and column-level security, PII masking rules, and schema-defined access policies are defined and enforced within Postgres, giving organizations strong, proven governance over the operational data at the foundation of the platform.

Every component in the stack is open source, writes to open Apache Iceberg format, and is deployable anywhere—no vendor controls the data format, the query runtime, or the migration path. Audit logging is centralized through the Postgres interface, providing a comprehensive record of all data access and modifications. The customer retains complete control.

Deployment models

EDB PG AI supports deployment across on-premises, hybrid cloud, and air-gapped environments. All platform components—ClickHouse, WarehousePG, and Postgres—deploy together within the EDB PG AI environment, requiring no separate infrastructure footprint. Sovereign cloud deployment satisfies data residency and regulatory jurisdiction requirements, keeping the organization in full control of where data is stored and processed.

Security model

EDB PG AI adds enterprise-grade database security to Postgres. Data redaction policies dynamically mask sensitive column values at query time for specific users or roles, without altering stored data. EDB audit logging provides configurable, fine-grained activity tracking—covering connections, data definition language (DDL) and data manipulation language (DML) operations, and failed authentication attempts—through structured log files suitable for compliance reporting. Profile management enforces enterprise authentication controls: password complexity rules, rotation schedules, and account lockout after repeated failed attempts.

Because ClickHouse, WarehousePG, and all query engines read from the shared Iceberg catalog governed at the Postgres layer, the attack surface is minimized—data never traverses uncontrolled ingestion paths.

Performance characteristics

Metric	Converged analytics
Data availability latency (lakehouse)	<1 sec (near real time) vs. minutes to hours with traditional ETL
Concurrency performance (WarehousePG)	52% more consistent vs. cloud data warehouses
Total cost of ownership (WarehousePG)	58% lower vs. cloud data warehouses
Analytical query performance (Postgres offload to NVIDIA Spark RAPIDS)	Up to 99x faster than standard PostgreSQL for large-scale analytical workloads (TPC-DS benchmark)
Analytical query performance (NVIDIA Spark RAPIDS vs. CPU-only Spark)	Up to 14x faster at 3TB scale (TPC-DS benchmark)
Data freshness	Continuous vs. batch in traditional pipelines
Pipeline overhead	None vs. high overhead with traditional ETL

*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.

Technical comparative analysis

Platform architecture comparison

Capability	EDB PG AI	Databricks	Snowflake	BigQuery
Per-engine ingestion	No—all engines read from shared catalog	No—Lakebase and Lakehouse share a unified storage layer	Yes—load to native tables	Yes—load to native tables
Storage format	Delta Lake + Iceberg (open)	Delta Lake + Iceberg (open)	Proprietary + Iceberg	Proprietary + Iceberg
Governance	Source defined	Catalog based	Platform based	Platform based
Query engine	Multi-engine	Lakebase + Spark/ Photon	Proprietary	Proprietary
AI integration	MCP native (open protocol)	Mosaic AI (proprietary + MCP support)	Cortex AI (proprietary + MCP support)	Vertex AI (proprietary)
Deployment	Hybrid/on-premises/public and private cloud	Cloud-only	Cloud-only	Cloud-only

*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.

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.

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