Scaling Sovereign AI with Speed and Efficiency

Unlock 3.5x greater vector query performance with EDB Postgres AI and IBM Power

enterprisedb.com

As enterprises transition into the agentic era, they are demanding more from their data, shifting it from a collection of static records into strategic assets that fuel AI agents to take action. To do this, 95% of enterprises plan to become their own AI and data platform over the next two years. However, only 13% have successfully moved AI into production at scale. Winning businesses know that having a viable strategy of putting machine labor to work for their customers (generating real-time insights, identifying points of failure, and autonomously scaling their operations) requires the right architectural foundation. That's why matching a great data and AI platform with a consistent, reliable hardware foundation can help you extract the most out of Postgres®.

EDB Postgres AI (EDB PG AI) on IBM Power high-performance servers gives users the ability to operationalize AI workloads with the world's most popular database, all with superior production-grade performance and full sovereignty. Together, EnterpriseDB (EDB) and IBM offer customers a deliberate path forward without compromising on scalability or control. EDB PG AI gives enterprises an agentic data and AI platform to unify transactional, analytical, and AI workloads, while IBM Power11 provides enterprise-grade infrastructure to accelerate mission-critical workloads.

With EDB PG AI on IBM Power, you get the Postgres, extensions, and features you love, with a reliable, secure, and flexible foundation to accelerate core operational and new agentic use cases.

This benchmark quantifies the relational and vector query performance of this particular software-hardware combination of EDB PG AI[1] and IBM Power. The results measure total transactions per second (TPS) and queries per second (QPS), comparing the performance and resource requirements of using EDB PG AI on IBM Power11 servers (IBM Power) versus Intel Xeon (x86). The following outcomes were assessed:

1. The transactional performance of EDB PG AI on IBM Power versus x86
2. The vector query performance of EDB PG AI with pgvector on IBM Power versus x86

The results of this performance benchmark indicate that **EDB PG AI has up to 2.1x better transactional performance on IBM Power, and 3.5x vector query performance on IBM Power when using pgvector.[2]**

EDB PG AI on IBM Power simultaneously meets the demands of high-volume transactions and complex vector searches, without moving data across fragmented systems or risking sovereignty. Organizations need this business continuity and performance at scale when mobilizing their agentic workforce. This benchmark helps enterprises assess the software-hardware foundation needed to get on the fast path to agent production. Read further to learn more about the configurations, methodology, and results supporting these outcomes.

1. *This test of EDB PG AI includes core database capabilities only and does not include EDB PG AI Hybrid Manager, Analytics Accelerator, or Agent Factory.*
2. *These results are in line with the Composite Performance Metrics (CPM) used by PrecisionIT to evaluate the efficiency of running databases, such as Postgres, on IBM Power hardware.*

Configurations

Transaction performance test

The performance test of EDB PG AI on IBM Power versus x86 used the following configurations:

Scale factor	1,000
Buffer cache size	20 GB
Clients	180
Database engine	EDB Postgres AI (EDB Postgres Advanced Server version 17.6)
IBM Power	IBM Power E1180 (16x16 core)
x86	Intel Xeon 6980P x86 (2x128 core)
VMs	4 total, Linux 9.6

Vector performance test

The performance test of EDB PG AI on IBM Power versus x86, based on internal testing running ANN vector search, used the following configurations:

HNSW settings	ef_construction=200, m=20, ef_search=85
Database engine	EDB Postgres AI (EDB Postgres Advanced Server Version 18.4)
IBM Power	IBM Power E1180 (4x64 core)
x86	Intel Xeon 6960P x86 (2x72 core)
Memory	1TB
VMs	16 core, Linux 9.6
Operating system	RHEL 9.7

These tests were conducted under laboratory conditions, with pgvector optimized on Power11 using detailed build instructions. Individual results can vary based on workload size, use of storage subsystems, and other conditions.

Transaction performance test

Methodology

This test evaluated the resource requirements of using EDB PG AI on IBM Power versus x86 for any given workload. The test measured read-only performance using pgbench, which is the standard benchmarking tool for Postgres. A heavy workload simulation (in which all cores of the system are used) with concurrent users was conducted. The following requirements were implemented for this test:

1. The same pgbench workload was used across both IBM Power and x86 servers with a simulation of 180 clients.
2. The same EDB PG AI software stack was used on IBM Power and x86 servers.
3. Virtual machines (VMs) operated at 100% CPU utilization to maximize cores, usage, and subscription.
4. Four VMs operated in parallel to stress the physical server and to represent the client production environment.

Results

The results were measured using total transactions per second (TPS), and efficiency was quantified by calculating TPS per core.

	IBM Power E1180 **Cores:** 60 **Storage:** 2 TB **Virtualization:** 4 VMs	**2x Intel Xeon 6980P** **Cores:** 160 **Storage:** 2 TB **Virtualization:** 4 VMs
Online transaction workload	5,281,120.17 TPS	6,598,363.39 TPS
TPS/core	88,018.67	41,239.77

The results of this test show that you can complete up to 2.1x more transactions per workload per core with EDB PG AI on IBM Power versus x86.[3] By running your mission-critical workloads with fewer cores on highly dependable servers, you can transition from high-maintenance to streamlined operations with an AI-ready architecture built for modernization. For enterprises, this means consolidated administration efforts, reduced infrastructure footprint, and lower total cost of ownership (TCO).

3. These results are in line with the Composite Performance Metrics (CPM) used by PrecisionIT to evaluate the efficiency of running databases, such as Postgres, on IBM Power hardware.

pgvector performance test

Methodology

This test compared resource requirements of EDB PG AI with pgvector on IBM Power versus x86. pgvector is an open source extension for Postgres that allows store, query, and index vectors to be implemented directly within a database. The test uses a standard vsbt method to measure vector search performance: approximate nearest neighbor (ANN), a heavy stress test using hierarchical navigable small world (HNSW) index. This is a modified workload test specifically designed to simulate real-world generative AI and vector search traffic. Instead of searching data populated in simple rows and columns, an ANN-modified workload replaces those simple tasks with complex vector mathematics. The database is asked to perform a concurrent set of high-dimensional similarity search queries and the test measures the number of such queries that the system can serve per second.

In order to run this test with the closest comparison possible, the following methodology was used:

1. The same vector data, HNSW index configuration, and search queries across x86 and IBM Power servers
2. The same EDB PG AI Database version (Postgres Version 18.4.0)
3. VMs operating at 100% CPU utilization to maximize cores, usage, and subscription
4. Virtualization: IBM Power using PowerVM on Power and x86 using KVM
5. 16 cores for both IBM Power and x86
6. Total memory encryption (TME) turned on for x86, since Power11 has memory encryption turned on by default

Results

The results of this test were measured in total queries per second (TQS) and TQS per core using a recall rate of 97.04%.

Number of Cores	P11 TQS	P11 TQS per core	X86 TQS	X86 TQS per core	Ratio
16	18355.95	1147.25	5307.23	331.7	3.46

These results show that EDB PG AI with pgvector has 3.5x better vector query performance per core on IBM Power versus x86. AI agents rely on vector searches to retrieve accurate, relevant information in milliseconds and make real-time decisions. For enterprises, this result represents a functional, hardened solution optimized for speed and accuracy rather than an experimental test. When supporting a large volume of concurrent queries, your autonomous agents can run with fewer resources. This lowers the tax AI has on your total IT budget, preventing cost barriers from hindering strategic AI adoption.

Conclusion

EDB PG AI on IBM Power offers a robust, enterprise-grade foundation for scaling mission-critical AI. By delivering up to 3.5x faster vector queries and 2.1x better transactional performance on fewer cores, this solution ensures your agents remain responsive and accurate—even under extreme stress. This translates into lower maintenance costs and management overhead for businesses looking to modernize their infrastructure for the agentic era, without expanding IT budgets. With EDB PG AI on IBM Power, you gain the performance required for high-volume AI workloads while maintaining total data sovereignty within your private, controlled environment.

Learn more about EDB PG AI on IBM or contact us to deploy production-ready Postgres and accelerate your strategic AI initiatives with confidence.

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 is the first open, enterprise-grade sovereign data and AI platform, with a secure, compliant, and fully scalable environment, on premises and across clouds. Supported by a global partner network, EDB Postgres AI unifies transactional, analytical, and AI workloads, enabling organizations to operationalize their data and LLMs where, when, and how they need them. For more information, visit enterprisedb.com.

© EnterpriseDB Corporation 2026. All rights reserved.

EDB POSTGRES AI