



**Scaling Sovereign AI  
with Speed and Efficiency**  
Unlock 2x greater performance  
with EDB PG AI on IBM Power

As enterprises transition into the age of AI, they are demanding more from their data—shifting it from a collection of records into strategic assets that fuel AI agents to take action. Winning businesses have a clear strategy of putting machine labor to work for their customers, generating real-time insights, identifying points of failure, and autonomously scaling their operations. In order to gain this competitive advantage and grow faster, organizations need to choose the right architectural approach, one that goes beyond increasing their infrastructure footprint alone. That's why matching a great database platform with the right 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 performance and full sovereignty. While 95% of enterprises plan to become their own AI and data platform within the next three years, only 13% have successfully moved AI into production at scale. At [EnterpriseDB](#) (EDB), we offer customers a deliberate path forward without compromising on scalability or control. EDB PG AI is the world's first agentic, open source-based sovereign data and AI platform designed to unify transactional, analytical, and AI workloads. This software is supported on [IBM Power11](#), which provides enterprise-grade infrastructure that is purpose-built to operate mission-critical workloads faster and with confidence.

**With EDB PG AI on IBM Power, customers get the same Postgres and features they love, with a reliable, secure, and flexible foundation to operate efficiently and accelerate AI use cases.**

This report quantifies the relational and vector query performance of this particular software-hardware combination of EDB PG AI<sup>1</sup> and IBM Power. The results were measured in total transactions per second (TPS) and queries per second (QPS). The tests conducted compare the resource requirements of using EDB PG AI on IBM Power11 servers (IBM Power) versus Intel Xeon (x86). The following outcomes were assessed:

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

The results of this performance report indicate that **EDB PG AI has up to 2.1x better transactional performance on IBM Power, and up to 2.2x better vector query performance on IBM Power when using pgvector.**<sup>2</sup>

Organizations need business continuity and performance at scale when mobilizing their agentic workforce. 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. This report helps enterprises assess the software-hardware foundation needed to get on the fast path to production. Read further to learn more about the configurations, methodology, and results supporting these outcomes.

1. This test of EDB PG AI includes transactional capabilities only and does not include EDB PG AI Hybrid Manager, Analytics Accelerator, or AI Factory.

2. These are preliminary results. More extensive testing is being conducted and benchmark reports will be published in the near future.

## Configurations

### Transaction performance test

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

<b>Scale factor</b>	1,000
<b>Buffer cache size</b>	3.86
<b>Clients</b>	180
<b>Database engine</b>	EDB Postgres AI (EDB Postgres Advanced Server version 17.6)
<b>IBM Power</b>	IBM Power E1180 (16x16 core)
<b>x86</b>	Intel Xeon 6980P x86 (2x128 core)
<b>VMs</b>	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:

<b>HNSW settings</b>	Euclidean distance, ef_construction=200, m=20, ef_search=200
<b>Database engine</b>	EDB Postgres AI (EDB Postgres Advanced Server version 18.1)
<b>IBM Power</b>	IBM Power E1124 (4x8 core)
<b>x86</b>	Intel Xeon 6960P x86 (2x72 core)
<b>VMs</b>	8 core, Linux 9.6

These tests were conducted under laboratory conditions. 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.

	<b>IBM Power E1180</b> Cores: 60 Storage: 2 TB Virtualization: 4 VMs	<b>2x Intel Xeon 6980P</b> Cores: 160 Storage: 2 TB Virtualization: 4 VMs
<b>Online transaction workload</b>	5,281,120.17 TPS	6,598,363.39 TPS
<b>TPS/core</b>	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.<sup>3</sup> 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. A standard method to measure vector performance is conducting a heavy stress test, referred to as approximate nearest neighbor (ANN), 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 Postgres version 18.1.0
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. Eight 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), queries per core, and recall rate.

	<b>System: Power11 (S1124, 9824-42A), 32 cores, 2 TB</b>	<b>System: Intel Granite Rapids (6960P), 144 cores, 2 TB</b>
<b>Vector search workload (TQS)</b>	2698.92	6,598,363.39 TPS
<b>Queries per core (on an 8-core partition)</b>	88,018.67	41,239.77
<b>Recall rate (Euclidean distance, m=20, ef_search=200, ef_construction=200)</b>	88,018.67	41,239.77

These results show that EDB PG AI with pgvector has 2.2x better 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 organizations, this result represents a functional enterprise-hardened solution optimized for speed and accuracy rather than an experimental test. Supporting a large volume of concurrent AI agents to complete more queries on fewer cores reduces the “AI tax” on your total IT budget, aligning your strategic objectives with cost constraints.

## Conclusion

EDB PG AI on IBM Power provides a reliable solution for enterprises needing to scale their mission-critical and AI workloads across their organizations. Together, EDB PG AI on IBM Power delivers up to 2.2x better vector query search. This means organizations can scale agents to work faster across all teams with rapid and accurate results. You can accelerate AI use cases with confidence knowing that EDB PG AI on IBM Power handles large vector searches for multidimensional data and ensures that AI applications remain responsive and consistent, even under extreme stress.

EDB PG AI on IBM Power delivers up to 2.1x better transactional performance with a greater volume of transactions completed, using fewer cores. This translates into lower maintenance costs and management overhead for businesses looking to accomplish more without expanding their operating budgets. You can modernize your data estate with total sovereignty, ensuring that sensitive data remains within your high-performance, private, controlled environment.

[Contact us](#) to deploy EDB PG AI on IBM Power and get production-ready Postgres in your private data center for fast, reliable transactions and critical AI workloads.