



DATA STREAMING PERFORMANCE

McKnight Consulting Group © 2024

Self-Hosted Database NoSQL Performance

Testing

Enterprise DB MySQL



RELATED RESEARCH

[Data Stream Processing Ease of Use and TCO](#)
[Benchmark of data pipeline performance & cost](#)

William McKnight
Jake Dolezal

Prepared by



McKnight Consulting Group
www.mcknightcg.com

December 2024

Contents

Executive Summary	3
Platforms	4
EDB	4
MySQL.....	4
Platform Summary.....	4
Setup	5
AWS EC2 Instances.....	5
NoSQL Workload: PG NoSQL Benchmark	5
Results	7
NoSQL Performance.....	7
Key takeaway	9
Conclusion	10
Appendix: Test Configuration	11
About EnterpriseDB	12
About McKnight Consulting Group	13
Disclaimer.....	14

Executive Summary

A self-hosted database provides control, security, cost-effectiveness, flexibility and compliance. It allows organizations to customize and optimize performance, ensuring data resides securely within their infrastructure.

Performance in these databases is critical. By enabling fast transaction processing, low latency, high throughput and scalability, performance profoundly impacts the user experience, business operations and data integrity. Slow performance frustrates users, hindering productivity and satisfaction, while delays in transaction processing affect revenue and competitiveness. Moreover, poor performance increases the risk of data corruption and loss.

Scalability is also compromised by inefficient performance, limiting growth and causing system overload. Furthermore, performance vulnerabilities create security risks, providing entry points for cyber threats. Resource utilization is also affected, wasting computing resources.

High performance ensures seamless transactions, efficient data processing and reliable operations. It maintains organizational credibility, prevents maintenance complexities and safeguards against security threats. Effective performance is essential. Organizations that prioritize performance benefit.

By investing in the most fit database, including consideration of performance, companies can ensure their applications are running smoothly and securely, ultimately leading to better overall business outcomes.

EnterpriseDB (EDB) is a leading provider of enterprise-class data management solutions that help organizations achieve optimal database performance. Their expertise in database optimization and compliance ensures that businesses can operate efficiently and securely, driving success in today's competitive landscape.

We tested EnterpriseDB against popular database MySQL to determine what, if any, would be the performance, cost difference and the potential worth of going beyond enterprise agreements for a self-hosted NoSQL database. EDB offers better NoSQL performance compared to MySQL, particularly as data volume increases.

These tests highlight EDB's compelling value proposition, making it a prime choice for enterprises seeking a single database as a high-performance transactional and NoSQL database without breaking the bank.

Platforms

EDB

EnterpriseDB is an open-source relational database management system based on PostgreSQL. It offers enterprise-class features, scalability and reliability. EnterpriseDB supports various platforms and programming languages. Its security features include encryption, row-level security and multi-factor authentication. EnterpriseDB's flexibility makes it suitable for web, cloud and on-premises applications.

MySQL

MySQL is a relational database management system that stores data in a structured format using tables, rows, and columns. It supports standard SQL and offers features like indexing, caching, and transactions to ensure data consistency and integrity. MySQL is widely used for web applications, supporting a large number of concurrent connections and queries, making it a reliable choice for managing and analyzing large datasets.

Platform Summary

For the NoSQL performance study, we tested the following releases of the platforms:

Vendor	EDB Postgres	MySQL
Tier	Advanced Server	Enterprise Edition
Version	17.2	8.0.38
OS	RHEL 9	RHEL 9

Setup

This section analyzes the methods we used in our NoSQL testing.

AWS EC2 Instances

To perform our testing, we used Amazon Web Services EC2 instances. We chose the following EC2 instance types for their high memory and attached NVMe storage:

Instance Type	i4i.8xlarge
Processors	32 vCPU Intel Xeon Scalable Ice Lake Processors
RAM	256 GB
Disks	Root on EBS gp3 + Data on 2x 3,750GB NVMe SSD Mirrored Disks (RAID1)

For each EC2 instance, we installed the database software as specified in the previous section. See the **Appendix** for all configuration settings used.

NoSQL Workload: PG NoSQL Benchmark

The workload and data used in the NoSQL benchmark were a workload derived from the PG NoSQL Benchmark¹. The original benchmark was only designed for PostgreSQL and MongoDB. We modified the benchmark to add a MySQL suite of functions based on its query syntax. EDB was already fully compatible with the PostgreSQL component of the benchmark, so no modifications for EDB were needed.

Our test performed the following tasks to compare EDB and MySQL:

- Generated large sets of JSON documents with the amount of data ranging from 5,000,000 documents (13GB) to 100,000,000 documents (266GB)
- Loaded the data into each database using EDB's SQL COPY, mongoimport, and MySQL's LOAD DATA LOCAL INFILE commands
- Loaded the same data using each platform's INSERT command
- Executed 4 SELECT queries on each platform

NoSQL Test Measurements

Our benchmark collected three measurements.

¹ More can be learned about the PG NoSQL Benchmark at https://github.com/EnterpriseDB/pg_nosql_benchmark.

- **Bulk load elapsed time** – measures the total time elapsed to load all the JSON via bulk load
- **INSERT elapsed time** – measures the total time elapsed to INSERT all the JSON via bulk load
- **Average SELECT time** – measures the average time to run each SELECT query

These measurements are chosen for their relevance to NoSQL use cases, facilitating comparability across databases and informing optimization, scaling and performance decisions.

Results

This section analyzes the results of our transactional and NoSQL workload scenarios for the platforms we tested.

NoSQL Performance

The following charts show the NoSQL performance of the three databases at various levels of data scale up to 266GB of JSON Data.

Initial Bulk Load of JSON Data

Lower is better

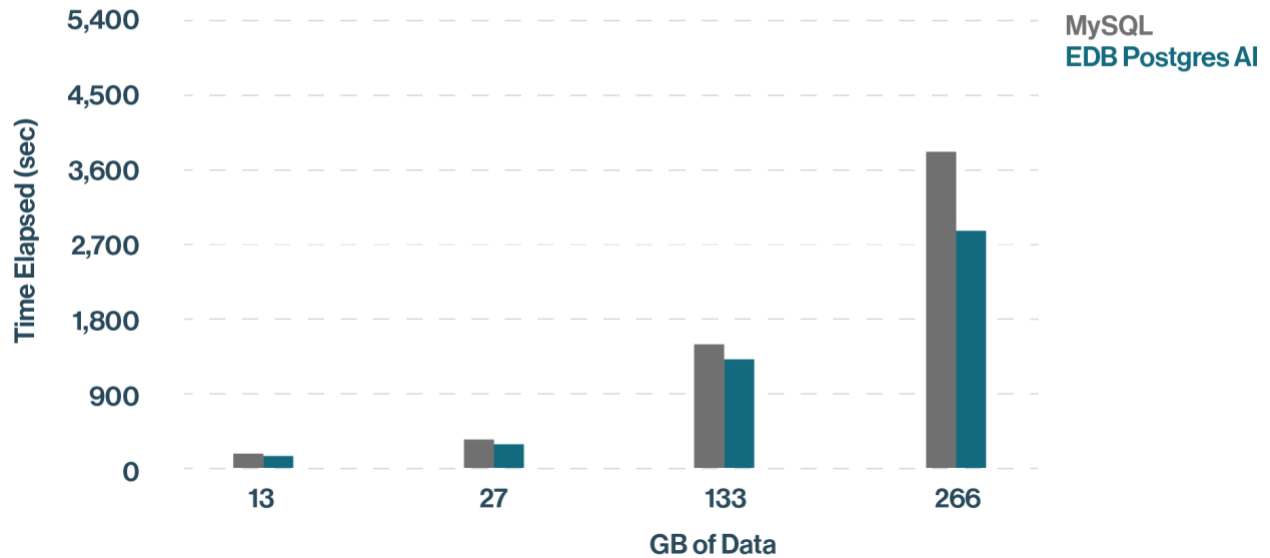


Figure 1: Total Elapsed Time of Initial Bulk Load of JSON Data

Row by Row INSERT of JSON Data

Lower is better

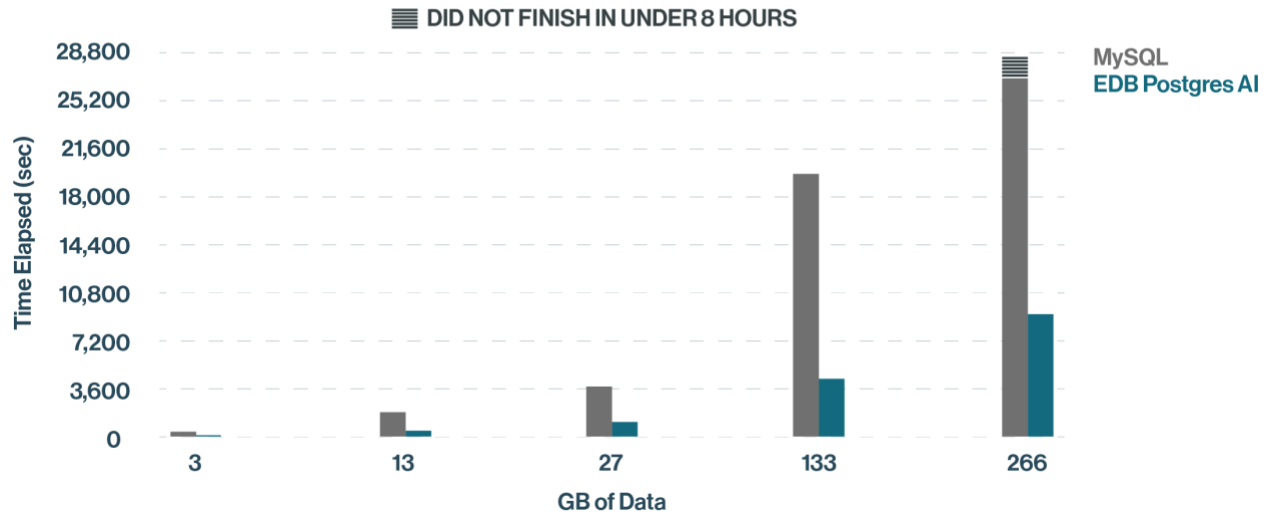


Figure 2: Total Elapsed Time of Row-by-Row INSERT of JSON Data

NOTE: MySQL did not finish loading 266GB row-by-row within 8 hours. That test was ended after 8 hours and marked as incomplete.

Average SELECT Query Runtime

Lower is better

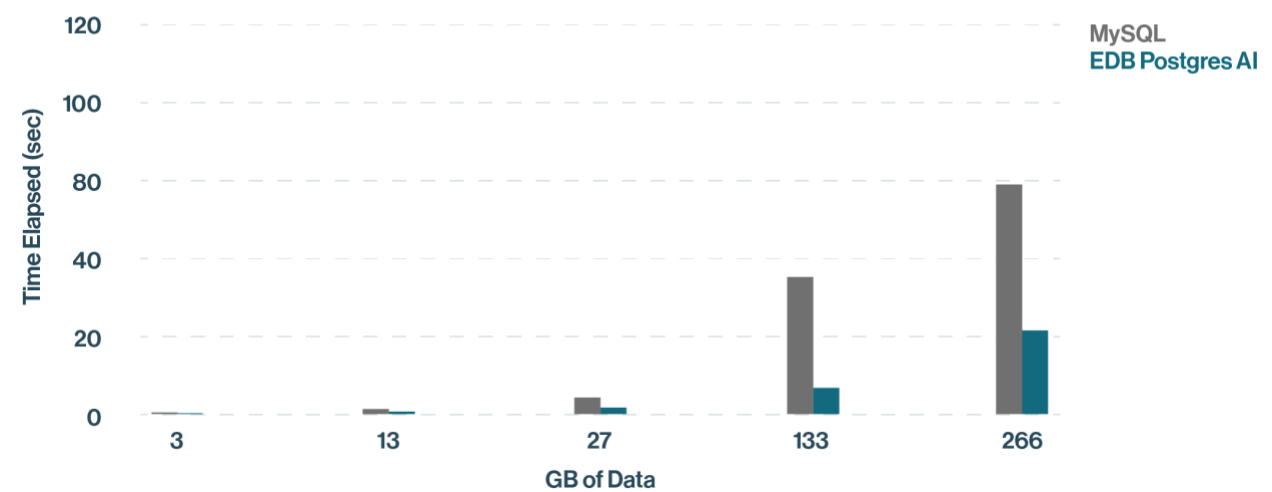


Figure 3: Average Elapsed Time of SELECT queries of JSON Data

The PG NoSQL Benchmark results offered insight into raw EDB and MySQL NoSQL performance with varying amounts of data.

- For bulk loads, EDB was 63% faster than MySQL.
- For row-by-row INSERTs, EDB was over 4 times faster than MySQL.
- For average SELECT query times, EDB was nearly 3 times faster than MySQL.

Key takeaway

EDB outperforms NoSQL performance compared to MySQL even as data volumes increase.

For enterprises demanding high-performance, cost-effective transactional NoSQL databases, EDB stands out as the optimal solution.

Conclusion

EDB also offers superior NoSQL performance compared to NoSQL competitor MySQL, particularly as data volume increases.

By choosing EDB, enterprises boost return on investment, gain budget flexibility and drive strategic growth through innovation and digital transformation, ensuring a forward-thinking technology infrastructure, all in a single database.



Appendix: Test Configuration

We tested EDB under the following configurations and conditions:

For the disk configuration, we moved **pgdata** and **pg_wal** to the mirrored NVMe disks. In the **edb-as-17.service** configuration file, we added:

```
After=tuned.service
```

In **postgresql.conf**, we changed:

```
max_connections = 1000
shared_buffers = 64GB
effective_cache_size = 128GB
work_mem = 8MB
maintenance_work_mem = 64GB
autovacuum_work_mem = 1GB
effective_io_concurrency = 200
maintenance_io_concurrency = 200
max_worker_processes = 128
max_parallel_maintenance_workers = 64
synchronous_commit = on
wal_level = replica
wal_buffers = 512MB
checkpoint_timeout = 1d
checkpoint_completion_target = 0.9
max_wal_size = 1TB
min_wal_size = 64GB
random_page_cost = 1.0
vacuum_cost_limit = 8000
autovacuum = off
autovacuum_freeze_max_age = 800000000
autovacuum_multixact_freeze_max_age = 800000000
max_locks_per_transaction = 512
edb_dynatune = 0
```

About EnterpriseDB

Nearly 1,500 customers worldwide have chosen EDB software, services, and support.

Enterprises and governments around the globe trust EDB to harness the full power of Postgres. With unmatched expertise, EDB ensures high availability, reliability, security, 24x7 global support and advanced professional services, both on premises and in the cloud, to help you control risk, manage costs and scale efficiently. As a leading contributor to the growing Postgres community, EDB is committed to driving technology innovation.

As your organization grows, you need strong strategies for the technologies you bet on. Postgres delivers superior technology, powered by a thriving, vibrant and fast-growing community. To accelerate your Postgres journey, and to run Postgres as your enterprise database standard, you need EDB.

EDB is the heartbeat of Postgres, with hundreds of technologists and developers, and more open source contributions to Postgres than any other company. EDB drives innovation, with enterprise-class software and services that enable businesses and governments globally to harness the full power of Postgres, the world's leading database.



About McKnight Consulting Group

Information Management is all about enabling an organization to have data in the best place to succeed to meet company goals. Mature data practices can integrate an entire organization across all core functions. Proper integration of that data facilitates the flow of information throughout the organization which allows for better decisions – made faster and with fewer errors. In short, well-done data can yield a better run company flush with real-time information... and with less costs.

However, before those benefits can be realized, a company must go through the business transformation of an implementation and systems integration. For many that have been involved in those types of projects in the past – data warehousing, master data, big data, analytics - the path toward a successful implementation and integration can seem never-ending at times and almost unachievable. Not so with McKnight Consulting Group (MCG) as your integration partner, because MCG has successfully implemented data solutions for our clients for over a decade. We understand the critical importance of setting clear, realistic expectations up front and ensuring that time-to-value is achieved quickly.

MCG has helped over 100 clients with analytics, big data, master data management and “all data” strategies and implementations across a variety of industries and worldwide locations. MCG offers flexible implementation methodologies that will fit the deployment model of your choice. The best methodologies, the best talent in the industry and a leadership team committed to client success makes MCG the right choice to help lead your project.

MCG, led by industry leader William McKnight, has deep data experience in a variety of industries that will enable your business to incorporate best practices while implementing leading technology. See www.mcknightcg.com.



Disclaimer

McKnight Consulting Group (MCG) runs all its tests to strict ethical standards. The results of the report are the objective and unbiased results of the application of queries to the simulations described in the report. The report clearly defines the selected criteria and process used to establish the field test. The report also clearly states the data set sizes, the platforms, the methods, etc. that were used. The reader is left to determine for themselves how to qualify the information for their individual needs. The report does not make any claims regarding third-party certification and presents the objective results received from the application of the process to the criteria as described in the report. The report strictly measures performance and does not purport to evaluate other factors that potential customers may find relevant when making a purchase decision. This is a sponsored report. The client chose its configuration, while MCG chose the test, configured the database and testing application, and ran the tests. MCG also chose the most compatible configurations for the other tested platforms. Choosing compatible configurations is subject to judgment. The information necessary to replicate this test is included. Readers are encouraged to compile their own representative configuration and test it for themselves.