



Swarm 64 DA Technical Overview

High-performance PostgreSQL extensions increase query performance to enable larger-scale reporting and dashboarding, easier Oracle and SQL Server migration, and open source data warehousing.

Swarm64 DA Overview

The Swarm64 Data Accelerator (Swarm64 DA) software extends free, open source PostgreSQL with high-performance features that help it run even faster and more cost-effectively.

Swarm64 DA is especially useful for databases that support reporting, dashboarding, or analytic workloads where query performance, scalability, and concurrency are important.

At a high level...four things to know about Swarm64 DA

Swarm64 DA accelerates PostgreSQL by enhancing it with

Swarm64 DA accelerates PostgreSQL by enhancing it with

- Smarter query planning,
- More parallel processing,
- Columnar indexing, and
- Faster I/O.

Swarm64 DA comfortably handles databases that are many terabytes in size, and with many concurrent users.

Swarm64 DA installs as an extension to

Swarm64 DA installs as an extension to free, open source PostgreSQL rather than requiring you to migrate to a custom fork or switching to a different database entirely.

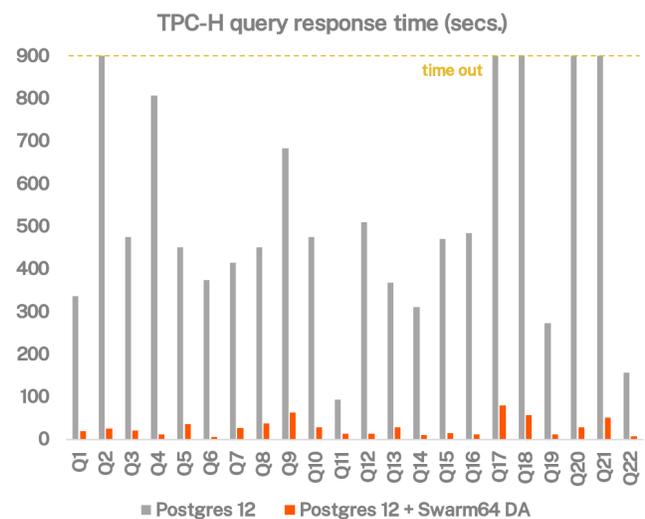
Faster query performance enables you to expand usage of PostgreSQL beyond transaction processing (OLTP).

Use Swarm64 DA-accelerated PostgreSQL for query-intensive projects like large-scale reporting, data warehousing, and analytics instead of expensive commercial alternatives like Oracle, SQL Server, and Netezza.

Swarm64 DA is developed by a team of high-performance database experts who are deeply experienced with PostgreSQL query planning and execution, parallel computing, low-level I/O optimizations, and hardware acceleration.

The team develops extensions to PostgreSQL, and also contributes open source patches, bug fixes and reference designs back to PostgreSQL.

This white paper describes the Swarm64 Data Accelerator (Swarm64 DA) software in detail—how it works, who it helps (and does not help), and how to get started using it.



Swarm64 DA-accelerated PostgreSQL use cases

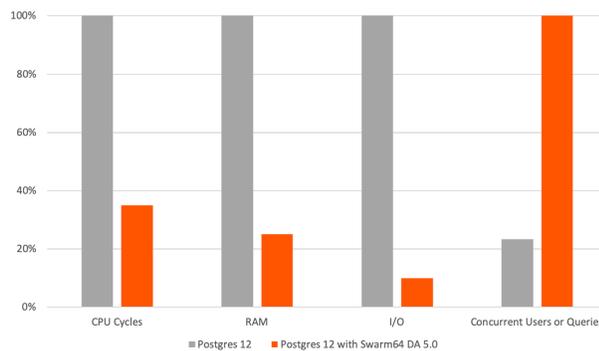
Swarm64 DA accelerates PostgreSQL query performance, which enables people to i gYZYŹ cdYb'gci fW'Dcgh fYGE @Zf'Uk jXyf'j UrjYmicZi gY'WUgYg--especially systems of insight or applications with large-scale reporting or dashboarding requirements.

Open source alternative to query-intensive Oracle & SQL Server projects

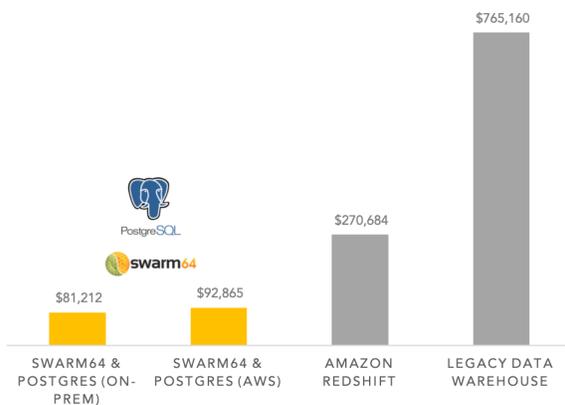
For mixed workloads or query-intensive projects that might require columnstore indexing in SQL Server or Oracle's in-memory columnstore, (or even a columnar database like Vertica, Greenplum, or MemSQL), Swarm64 DA-accelerated PostgreSQL is a money-saving alternative that offers comparable performance.

Increase PostgreSQL "server density" and profitability

Swarm64 DA increases PostgreSQL's CPU, memory, and I/O efficiency, which results in each PostgreSQL being even more productive. "Server density," as a large manufacturer called it, allowed their PostgreSQL servers to handle 4x more users without any hardware changes or migrations.



9.5X LOWER 3-YEAR COST (30TB DATA WAREHOUSE)



Open source data warehousing

Using Swarm64 DA-accelerated PostgreSQL as an alternative to IBM, Oracle, Amazon Redshift, or other costly data warehouses can save a company millions in IT costs.

One healthcare company compared the price-performance of Swarm64 DA-accelerated PostgreSQL against their legacy data warehouse. With Swarm64 DA, query performance was comparable to the legacy appliance, and the annual cost of Swarm64 DA and Postgres was 9.5x lower...a savings of \$670K USD over 3 years.

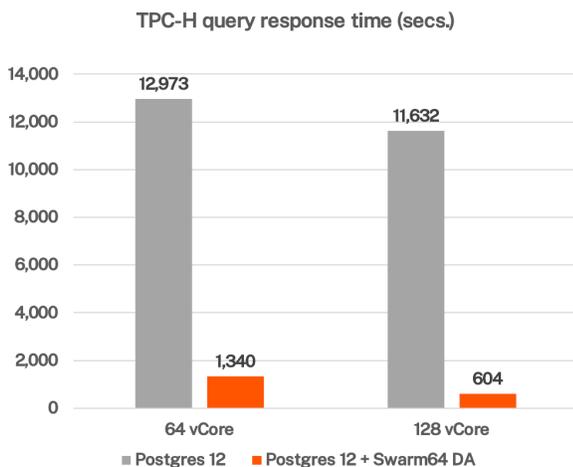
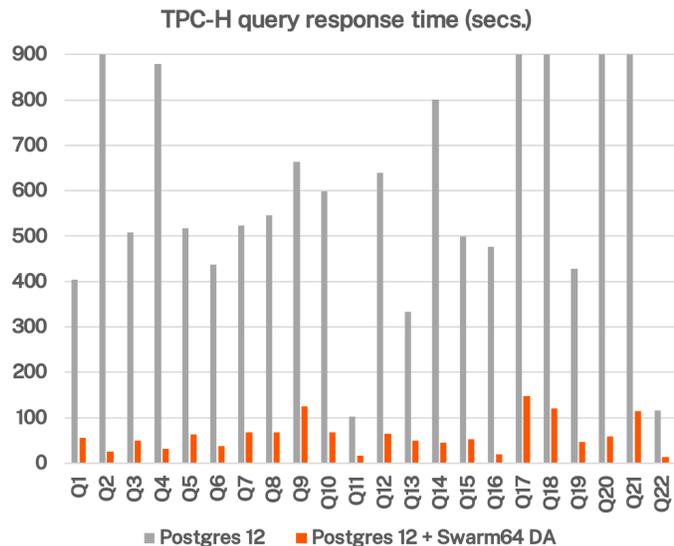
Swarm64 DA-accelerated PostgreSQL performance

Swarm64 DA software extends PostgreSQL with transaction-safe query engine and indexing improvements that accelerate query performance by an order of magnitude or more. The following benchmarks illustrate the accelerative effects of Swarm64 DA on PostgreSQL.

6x to 28x faster TPC-H

TPC-H is a popular SQL data warehousing benchmark. With 1TB of data, running on a 64-vcore Amazon EC2 r5d.16xlarge instance, Swarm64 DA improves PostgreSQL v12 query performance as follows:

- 10x faster overall (all 22 queries)
- Individual queries improving from 6x (Q13) to 28x (Q4) faster
- 5 queries timed out in standard PostgreSQL vs. all queries completed with Swarm64 DA



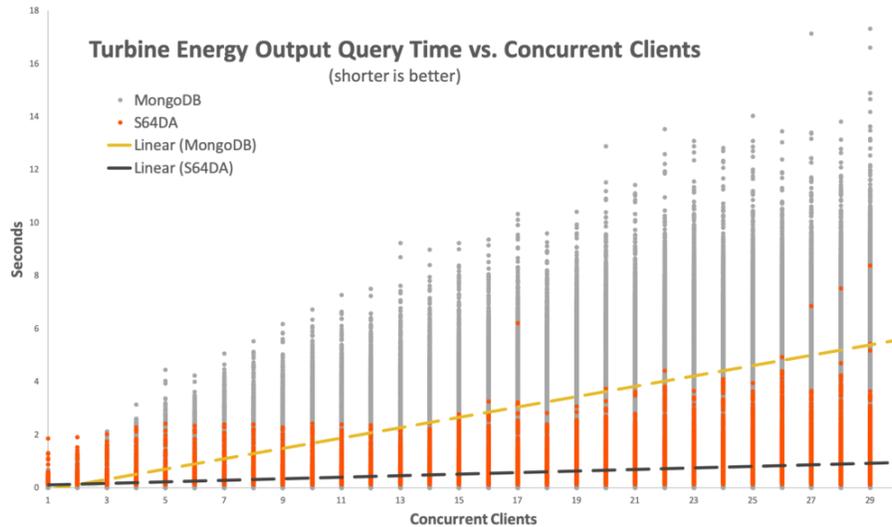
Performance scales linearly

Swarm64 DA extensions enable PostgreSQL query performance to scale linearly as CPU cores are added to the server.

As shown in the chart to the left, doubling the number of cores on the server mentioned above (from 64 virtual cores to 128), doubles the Swarm64 DA-accelerated PostgreSQL query performance. PostgreSQL now runs 20x faster with Swarm64 DA than without (with individual queries running 7x to 60x faster).

Customer benchmark: 5x better scaling, 80% lower cost than MongoDB

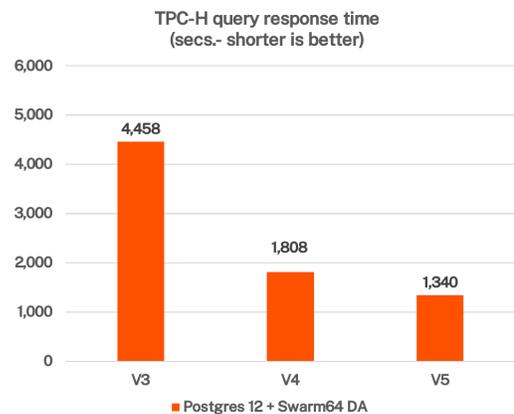
Wind turbine analytics software company, Turbit Systems, chose Swarm64 DA-accelerated PostgreSQL over MongoDB for their one TB time series database. Swarm64 DA enabled query response times to remain sub-second for 5x more concurrent users than MongoDB. This enabled Turbit to grow their user base 5x without increasing their cloud database infrastructure costs.



Partner benchmark - Continuous Swarm64 improvement

A long-time Swarm64 DA user benchmarked the performance of Swarm64 DA over several releases of Swarm64 DA. As the chart shows, the Swarm64 DA acceleration effect has tripled from version 3.0 in 2019 to version 5.0 in 2020.

At Swarm64, we are continuously developing new ways for Swarm64 DA to accelerate PostgreSQL. User feedback drives our product roadmap, providing us with new query patterns to accelerate.



Technology: How Swarm64 DA accelerates PostgreSQL

Swarm64 DA is a PostgreSQL extension

Swarm64 DA extends PostgreSQL by providing a collection of tools to increase the performance of your workloads. The default PostgreSQL functionality remains available while Swarm64 DA uses various official PostgreSQL mechanisms to add functionality: modules, hooks and data access methods. Swarm64 DA accelerates PostgreSQL queries at every stage:

1. During the query planning step,
2. When accessing the data, and
3. During query execution.

Improved query planning

During query planning, Swarm64 DA adds the following functionality:

Query rewriting

Common query patterns are transparently optimized with a behind-the-scenes SQL-to-SQL transformation that enables PostgreSQL to find much better query plans--for faster queries--while the semantics of the query remain the same.

Optimized Cost Functions

A good plan can save a lot of work; this statement is also true for PostgreSQL queries. When PostgreSQL's internal statistics are up to date and the predictions work well, queries complete faster. Reversely, many queries that take hours to finish are in fact suffering from planning errors. Swarm64 DA improves PostgreSQL query planning by providing better estimation methods for bigger datasets and mechanisms to automatically keep the statistics up to date.

Workload management

The way PostgreSQL allocates resources for parallel query execution can lead to errors or long-running queries due to CPU, memory, and disk overloads. With Swarm64 DA, PostgreSQL demonstrates greater stability and throughput when executing multiple queries in parallel thanks to the Swarm64 DA workload management system.

I/O reduction with columnstore indexing

Swarm64 DA uses a columnstore index to accelerate data access. The Swarm64 DA PostgreSQL columnstore index enables the user to keep all their native PostgreSQL tables with existing constraints and indexes, while adding a columnar view on top to enable faster queries.

The columnstore index caches the data of the related table in a compressed, column-oriented format, thus providing data locality for similar values. A query that uses the columnstore index benefits from more efficient I/O and improved CPU utilization. Both these properties lead to faster query execution.

After the columnstore index is created, the PostgreSQL planner uses it automatically in your queries. Like other PostgreSQL indexes, the columnar index has full transaction safety, crash-safety and replication support. This allows you to have the indexes maintained on the primary node and have updates propagated to all (read-only) replicas that then use these indexes immediately without any overhead.

Accessing the heavily compressed data selectively--and only the columns required for the query--provides the read speed required for fast query execution.

Improved query execution

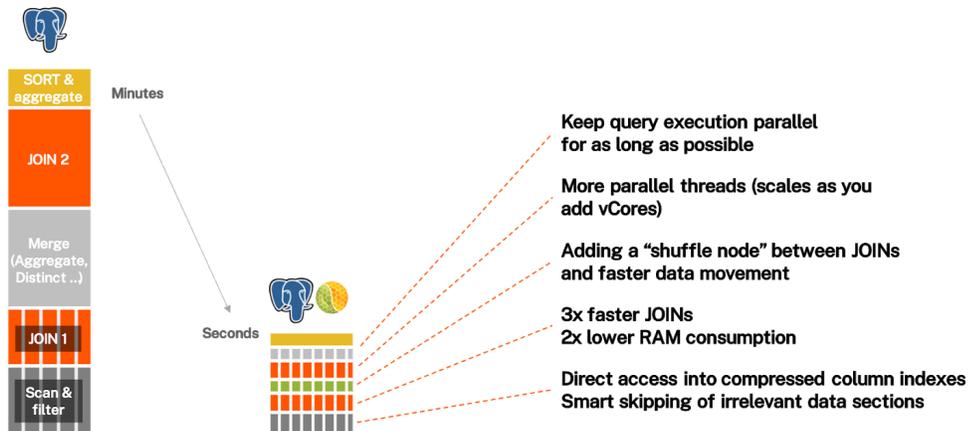
During query execution the plan is processed step by step. Within PostgreSQL, the individual steps taken are referred to as “execution nodes.”

Optimized query execution nodes accelerate JOINS and other query patterns

Swarm64 DA implements custom variants of standard PostgreSQL’s most expensive functions, such as the parallel hash join. This function is often used when querying large portions of multiple tables, e.g. for grouping data, running a report, or updating a dashboard. The optimized execution node for the hash join will run approx. 3x faster with half the memory requirements. This enables joining billions of rows in seconds and running more concurrent queries on the system.

Increased parallelism and better data locality

PostgreSQL query parallelism is improved by the Swarm64 DA shuffle node operator, which significantly partitions the data. So, in each parallel PostgreSQL worker on the server, data locality is improved and the query plan can stay parallel much longer. This dynamic partitioning, along with a custom high-performance hash join implementation, significantly reduces the query runtime.



Faster data movement

As millions of database tuples are passed along the nodes of the query plan, their transfer speed becomes essential. Swarm64 DA optimizes these operations to take as few CPU cycles as possible to free up system resources for concurrent queries and faster processing.

Result: faster queries

The sum of the improvements and added features during query planning, data access and query execution powers the performance gains shown in the previous section. When queries use one or two of these features, a measurable speed-up is visible. **K \ Yb`h Ymii gY'a cghcZ h YgY`ZUhi fYg`j`hik j``HU_YDcgh fYGE @j`ei YmfYgdcbgY`hja Y`Zca ` \ ci fg`hc`a]bi hYg`cf` Zca `a]bi hYg`hc`gYW`bXg`"**

Implementing Swarm64 DA

Swarm64 DA runs anywhere that PostgreSQL does

Cloud or on-premises

You can deploy a containerized version of Swarm64 DA on your own data center, or on any cloud platform. Swarm64 DA can also be launched as an AWS service via the Amazon Marketplace.

Standard PostgreSQL or EDB Postgres - Version 11 and higher

Swarm64 DA extends free, open source PostgreSQL. Swarm64 DA also works with EDB Postgres Advanced Server.

Recommended hardware (Amazon EC2 r5d.8xlarge)

Swarm64 DA parallelism scales with the number of virtual cores available to the server, we recommend that you be prepared to run Swarm64 DA-accelerated PostgreSQL on a server with at least 24 virtual cores. As a starting point, we typically recommend using a server similar to the Amazon EC2 r5d.8xlarge instance, which is configured with:

- 32 virtual cores
- 256 GB RAM
- 1.2TB NVMe SSD (or a SW RAID of SSDs)

Container support

Swarm64 DA can be deployed along PostgreSQL server and tools in a Docker container for immediate deployment and use. It can also be run in a VM or on bare metal.

If it works with PostgreSQL, it works with Swarm64 DA...

Swarm64 DA is a good PostgreSQL citizen. It extends PostgreSQL without requiring any SQL or application code changes. So application software, PostgreSQL tooling, ETL, reporting, and other database software will continue to work with Swarm64 DA-accelerated PostgreSQL databases.

