RESEARCH BRIEF

CAN SOFTWARE REDUCE DATA CENTER ENERGY REQUIREMENTS?

TETRA CA (Collapsive Architecture) HorneSci

Research brief · 2026-09-13
TETRA CA PBC · HorneSci

SECTION 01

EXECUTIVE SUMMARY

Over the past two decades, we have seen software take over every business vertical, every industry. The modern world's economy runs on software. To feed this, data centers have seen 17% compound annual growth rates since ~2018. Even before the current AI and data center boom, software infrastructure cost management garnered considerable attention. By 2030, data centers may consume up to 12% of the US power grid's output.

With AI, the infrastructure shortage has hit critical mass. Computer memory (RAM) has a supply shortage so extreme that Apple raised its prices. Other computer infrastructure components see significant price gains under a data center build-out.

Data center energy usage reflects complex ecosystem factors, but the largest power drain continues to be application servers running software.

Software operates across networks, and to perform even cursory tasks software must process and reprocess data in an endless cycle of store → retrieve → transform → serialize → transmit → deserialize → process → replicate → monitor → secure → store... etc. This constant reprocessing of data introduces unnecessary inefficiencies.

TETRA created the data matrix, a single representation of data that works across the store → process → transmit → read cycles. This means the computer has to do less work to get results. The data matrix uses vector mathematics to optimize efficiency and performance.

HorneSci created DeltaRouting, a customized runtime that optimizes vector calculation efficiency. Based on system level modeling using our working prototype, the benefits of using TETRA and DeltaRouting show potential for a ~40% data center energy reduction.

17%

compound annual growth rate, data centers

since ~2018

12%

of the US power grid's output

data centers, by 2030

~40%

data center energy reduction

TETRA and DeltaRouting, potential

SECTION 02

SOFTWARE AND DATA CENTERS

“Software is eating the world,” business person and investor Marc Andreessen wrote in the Fall of 2011. He cited the *technical ability* of software to drive business objectives, falling operational costs, and the size of the market reachable online. 15 years later, Andreessen’s insights read like prophecy.

> Data centers now number over 12,000 worldwide.

The reason for data centers: software.

Before our modern-day reliance on software systems, how did we manage business? Entertainment? Media? In 2026, entering a discussion about data centers, this may feel distant and irrelevant, but we think the question is highly relevant given software’s role in the economy today.

Commerce has always used ledgers to store, read, and consult information. Based on the ledger updates, letters or missives get sent to other interested parties. Interested parties would update their own ledgers, and in response they would transmit information as-needed.

These interactions would happen across “verticals,” between service providers, business ventures, banking systems… just like we use software to perform today. And this shows us *why* software ate the world: software enables real-time, efficient information storage, processing, retrieval, and transmission.

Software’s value derives from *information’s* value. And information about the real-world drives business and commerce.

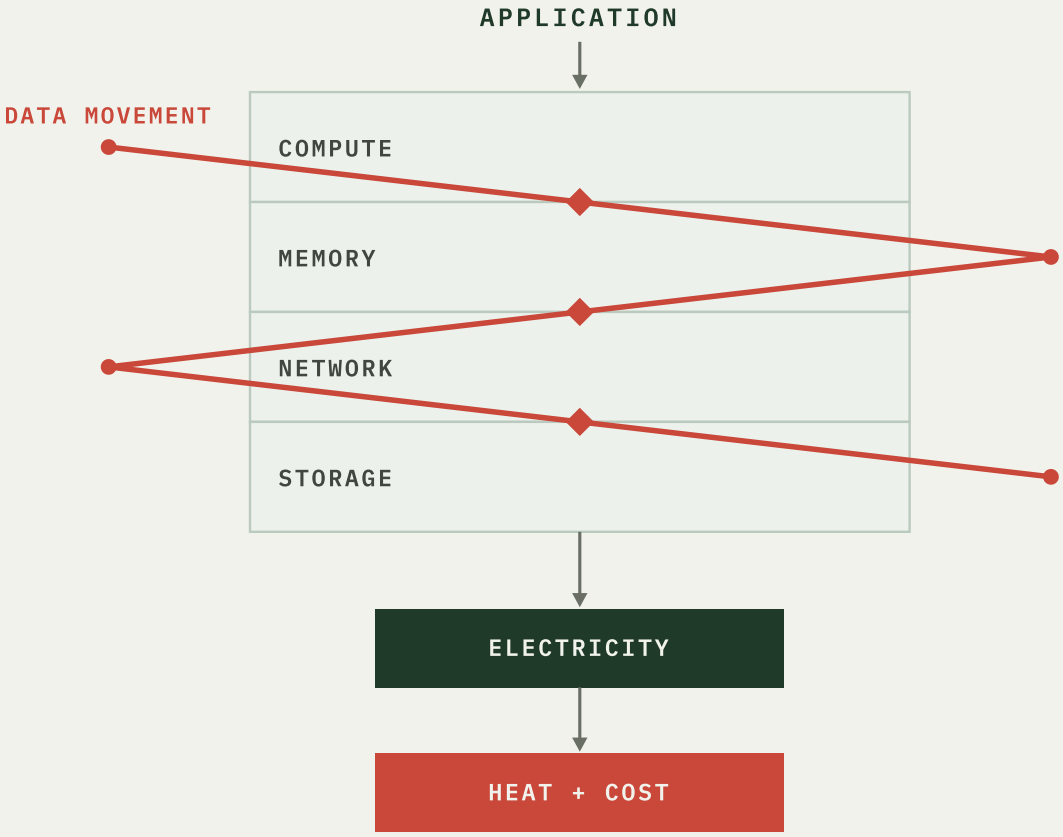
Software, then, *is* the “goods and services” for which data centers exist. The data center, having buildings, space, electrical requirements, and cooling needs, naturally attracts the most attention, similarly to ships and ports serving the physical exchange of goods.

Outside of a few niche industries using robotics, software cannot directly impact the physical world outside of its appearance as the infrastructure supporting it. Data centers.

What does it take to build a data center? Why do they use so much energy? Why do they make so much noise?

At the data center scale, running software means:

- Physical buildings; the foundations building structures
- Computing hardware for memory, processors, and disk storage
- Networking hardware for multiple computers to “talk together” and do tasks in-parallel
- Cooling; operating electronics produces heat, too much heat damages electronics
- Electricity.



As a data center operates, systems work in concert to keep computers running and available for business workloads. Software monitors software, sensors monitor temperature and regulate cooling, electric power meters detect supply and distribution of power to racks of servers. Power disruptions and disk outages get reported in near-real-time.

The data center apparatus itself has grown and expanded, entire systems dedicated to watching other systems.

This interconnected, engineered ecosystem enables us to run software and distribute systems across the internet, worldwide, at real-time speeds. Reliability has skyrocketed, with interdependent systems reporting 99.9% or more percent up-time.

For business-critical systems, the information stored gets repeated over multiple locations (like having a backup). This reliability requires significantly more infrastructure and energy expenditure than operating from a single data center. This cost-for-reliability trade-off exemplifies the complexity of software workload power consumption: often the workload spans *datacenters* now!

Data centers increasingly draw more power, climbing from 415 terawatt hours (TWh) in 2024 to a projected 565 TWh in 2026. On the whole, data centers have been growing at a remarkable 12% compound annualized growth rate (CAGR) *since 2017.* Current projections anticipate, by 2030, the United States will see %6.7 to 12% of its electrical power consumed by data centers.

Total data center electricity use from 2014 through 2028

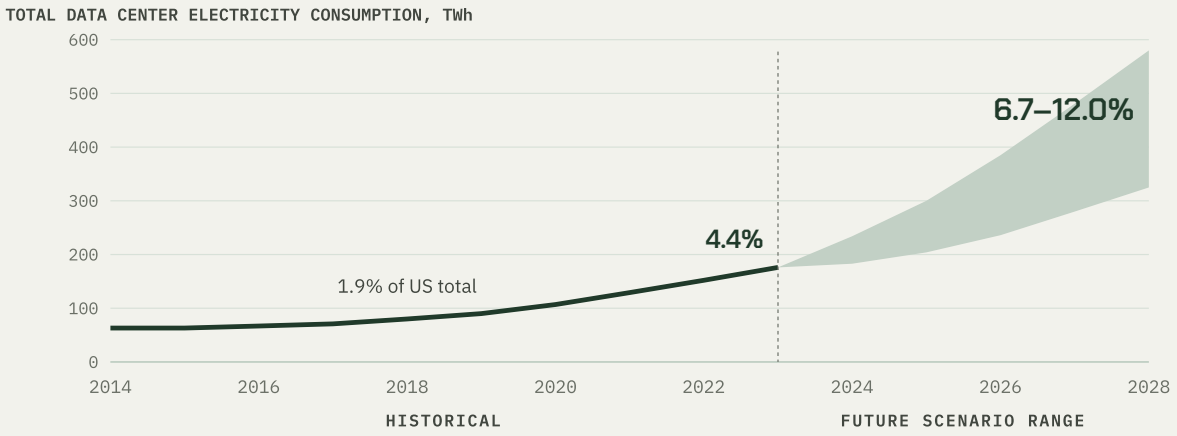


Figure 5.5, redrawn from the 2024 United States Data Center Energy Usage Report (Shehabi et al., 2024). Values read from the published figure.

Remember: Data centers exist to build, operate, and scale software applications. But, most data center discussions about efficiency and environment impact revolve around the *hardware* that we install to run the software.

TETRA CA and HorneSci have partnered to deliver a new solution: we optimize and accelerate existing systems to run efficiently on the hardware available today. Our collaborative efforts invite the question: What if we could do more with less? What impact would efficiency have on the current trajectory of data center build-out and operation?

Server annual electricity usage by type

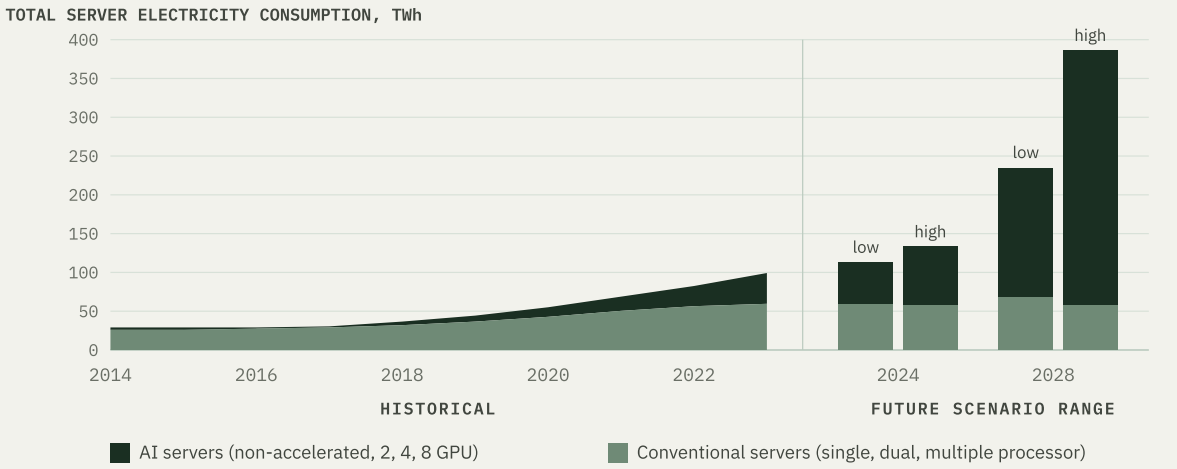


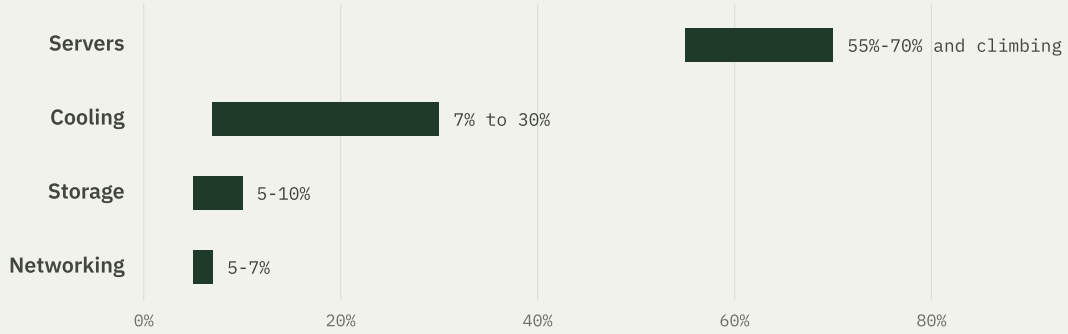
Figure 5.1, redrawn from the 2024 United States Data Center Energy Usage Report (Shehabi et al., 2024). Values read from the published figure; server types grouped.

SECTION 03

WHERE DOES THE ENERGY GO?

A meta-analysis of multiple publicly-available breakdowns indicate these broad categories of data center energy consumption:

- **Cooling:** 7% to 30%. This is where data centers use water. Cooling is also the source of data center noise (“the hum”)
- **Networking:** 5-7%
- **Storage:** 5-10%
- **Servers:** 55%-70% and climbing



Note: data centers have a lot of variability by size and specific hardware configurations, so the estimates here reflect best-effort research.

Note that ~60% of a data center’s electricity consumption feeds “servers.” This is where the “running software” part of a data center gets bucketed. Because the new hardware powering AI applications takes more processing power (2-4x, but falling), AI servers may push “servers” closer to 75% of total consumption.

Things get complex, however, when we ask for more specifics. A “software workload” can mean many things.

Given the tremendous amount of variation in technologies, skills, processes, legacy requirements, and social-technical-political issues, software workload’s energy efficiency has wide variation.

Many variables impact the operational electrical efficiency of a software workload. Yet the *server* bucket of energy spend shows more usage than any other category. This suggests that, for data center energy management, the most influential factors have slipped by unmeasured!

To delve into the details of specific software design decisions and implementation constraints, we must learn a little bit about how information gets products, saved into a computer, searched and retrieved, then restored. We must ask about the efficiency of our current business systems.

Direct water consumption by data center type

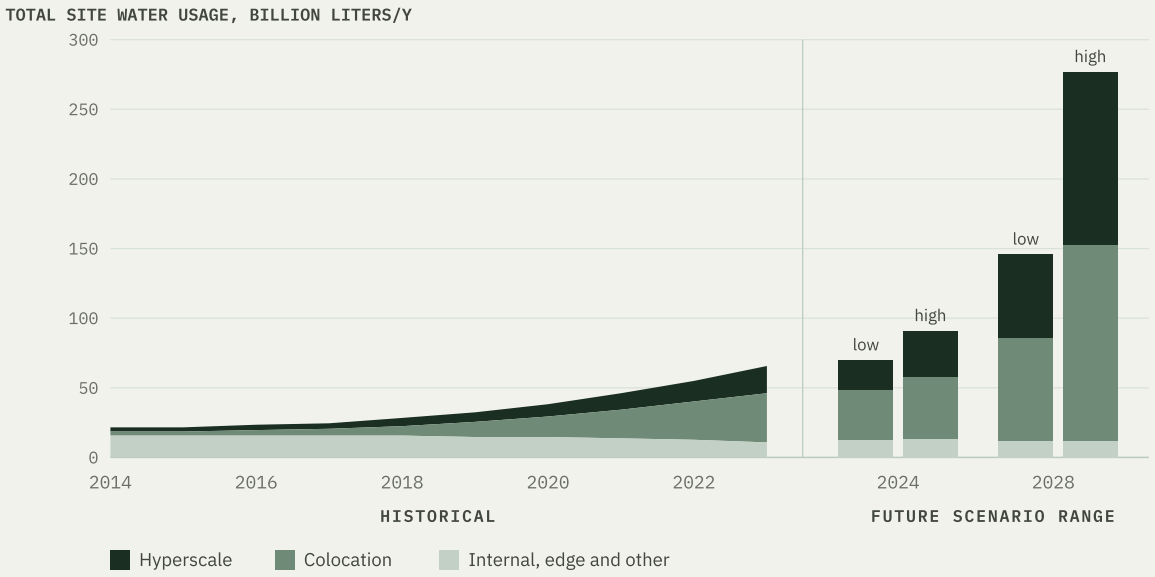


Figure 5.9, redrawn from the 2024 United States Data Center Energy Usage Report (Shehabi et al., 2024). Values read from the published figure; data center types grouped.

Network electricity use by port speed, with InfiniBand switches

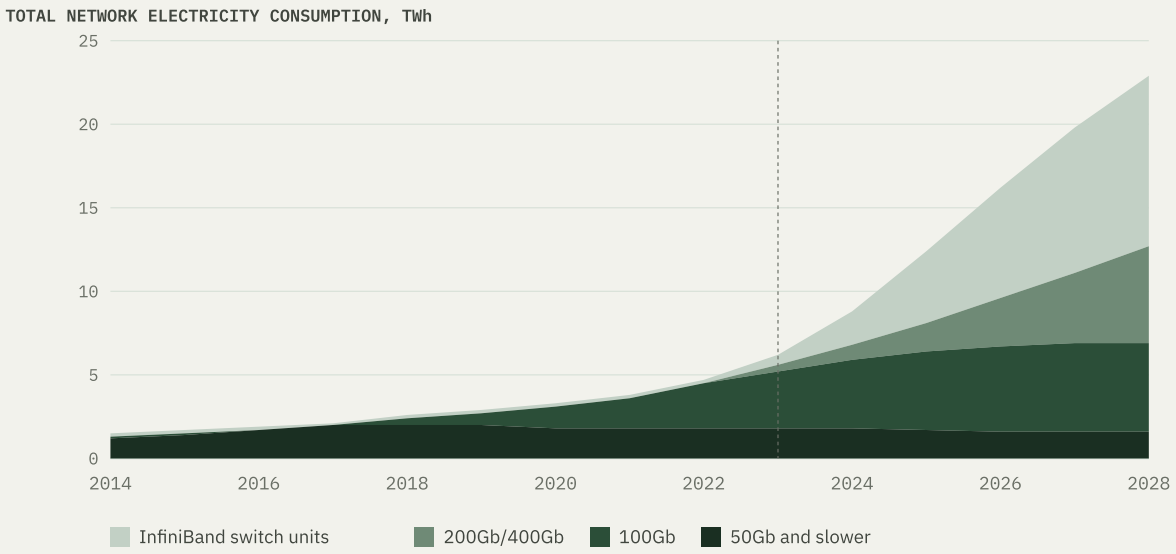


Figure 5.4, redrawn from the 2024 United States Data Center Energy Usage Report (Shehabi et al., 2024). Values read from the published figure; port speeds of 50Gb and slower grouped.

SECTION 04

BUSINESS DATA WORKLOADS

Business data workloads vary greatly. This is because, like writing, the same software “ideas” can be represented with multiple software configurations. Technology trends come and go quickly. New technologies, as well as the technologies that get adopted, may not be the most energy-efficient choices.

Generally speaking, business software workloads rely on:

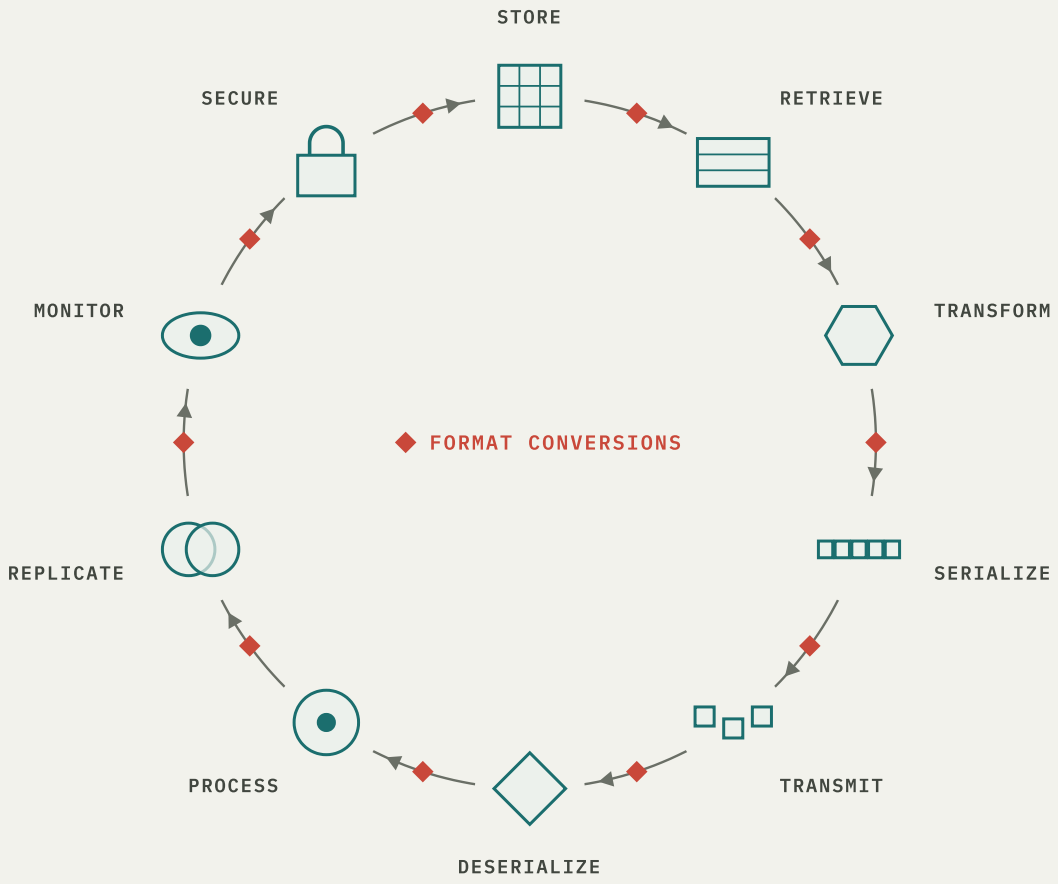
- Operating systems: the program that runs all the other programs on a computer
- Databases: special kind of application that is dedicated to storing and retrieving information
- File stores: computer files, like documents, pictures, whatever is needed
- User-facing apps, web pages, and 3rd party services

Today, almost all business data systems use linear, scalar operations to plan and execute work. “Take this information, hold it in short-term working memory, get that information, add it together, save it, send it, etc...” As hardware has grown faster, our ability to represent these operations has gotten incredibly fast.

Data systems have many requirements to make them safe to use. This includes cybersecurity, identity and access controls, operational monitoring, access and compliance logs. Often these concerns get “outsourced” to other software systems (the “Business to business Software as a Service [B2B SaaS]” ecosystem).

Each integration point or change of format requires computers to do a “translation” between one format and another. These translations are small and inexpensive, but there are many of them. From an app to a business system, from a business system to two or three third-party systems, each system having its own chain reaction of translations, and each database having its own format too! It’s a fractal fan-out.

These practices pervade computer engineering with very few exceptions.



store → retrieve → transform → serialize → transmit → deserialize → process → replicate → monitor → secure → store

Modern engineering designs lean heavily on networking between services. This networking, moving and transmitting the data, frequently costs more than storing the data because data uses more energy to transfer than it uses to stay on a disk. Software data networking costs frequently get ignored when building applications. (Note: the specific energy implications of the movement are outdated (source 2017), so networking has gotten more efficient, but the overall implication that moving data costs more than storing it still holds true.)

Moreover, because business systems grow more complex as they “scale” (serve more users or perform more operations), the current methods for constructing business systems get progressively more complex to operate. These linear, scalar operations need to do more operations to work on big data sets.

As mentioned above, enterprises often need to scale systems across multiple data centers. To keep these systems synchronized, the applications send and receive data continuously to stay up-to-date. This means the storage requirements and systems to manage continues to propagate; more systems, more scaling. Even prior to the technologies discussed below, data workloads could be reduced considerably. Why do businesses pay this ongoing overhead?

Because business data systems tend to minimize up-front investment, resultant software ecosystems often use data poorly. Small inefficiencies get repeated. These unnecessarily inefficiencies tend to get swallowed by the high margins and returns of business software. Generally speaking, traditional software infrastructure may account for 10%-15% of the revenue. On the individual business scale, this seems manageable. At the “software is eating the world” scale, though, the impacts of these design choices aggregate and compound.

Of course, new advancements such as LLM agents and AI services change the math, in which the ongoing percentage of infrastructure costs fluctuate with “tokens,” until an organization has enough technical acumen to host these models internally. Even with self-managed LLM systems, AI generally costs more to operate than traditional software methods.

Software engineering feels abstract and inaccessible to most people. But if we want to have a conversation about data center power usage and technology, we must consider the practice of software engineering.

SECTION 05

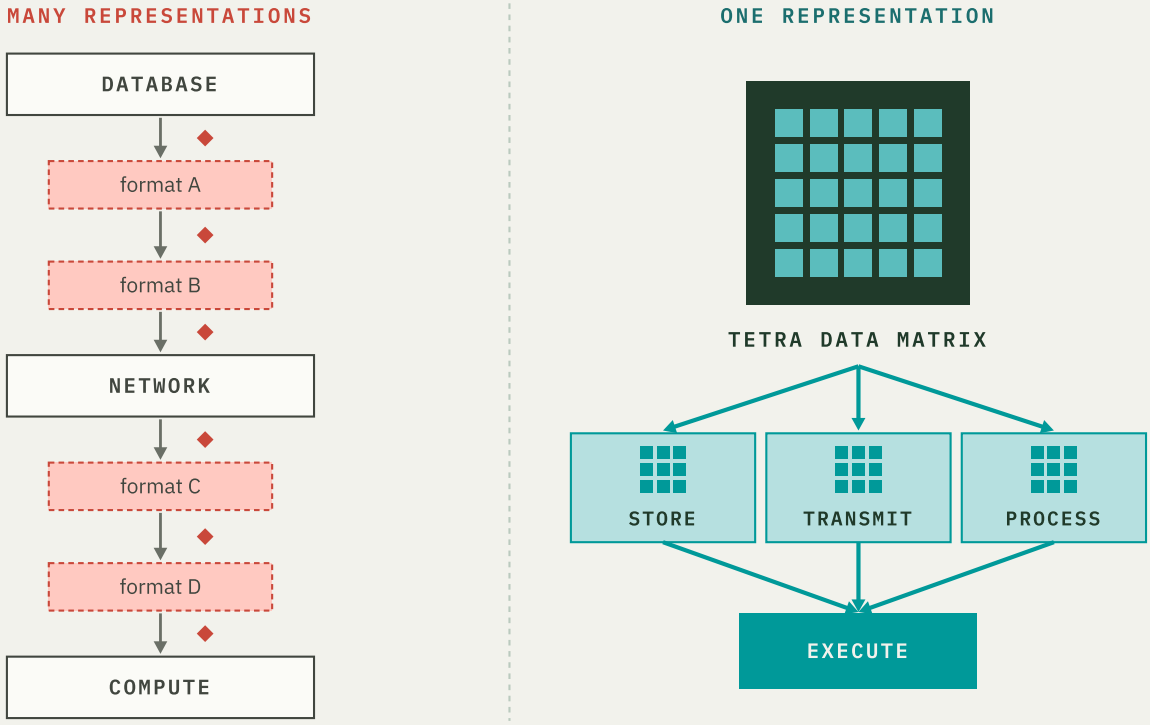
TETRA CA - THE DATA MATRIX

TETRA CA – the CA stands for “collapsive architecture” – operates as a Public Benefit Corporation registered in the State of Delaware.

The company invented, and has a patent application underway, the format to represent all information as a single matrix.

This same matrix format gets saved, sent over the network, or processed. The data matrix can also act as a business logic flow engine, replacing the scalar code used today.

When discussing software workloads above, we mentioned the “translation tax,” current systems must change data formats for every operation. The data matrix circumvents this translation tax entirely. Throughout the data center, the data matrix reduces the data friction.



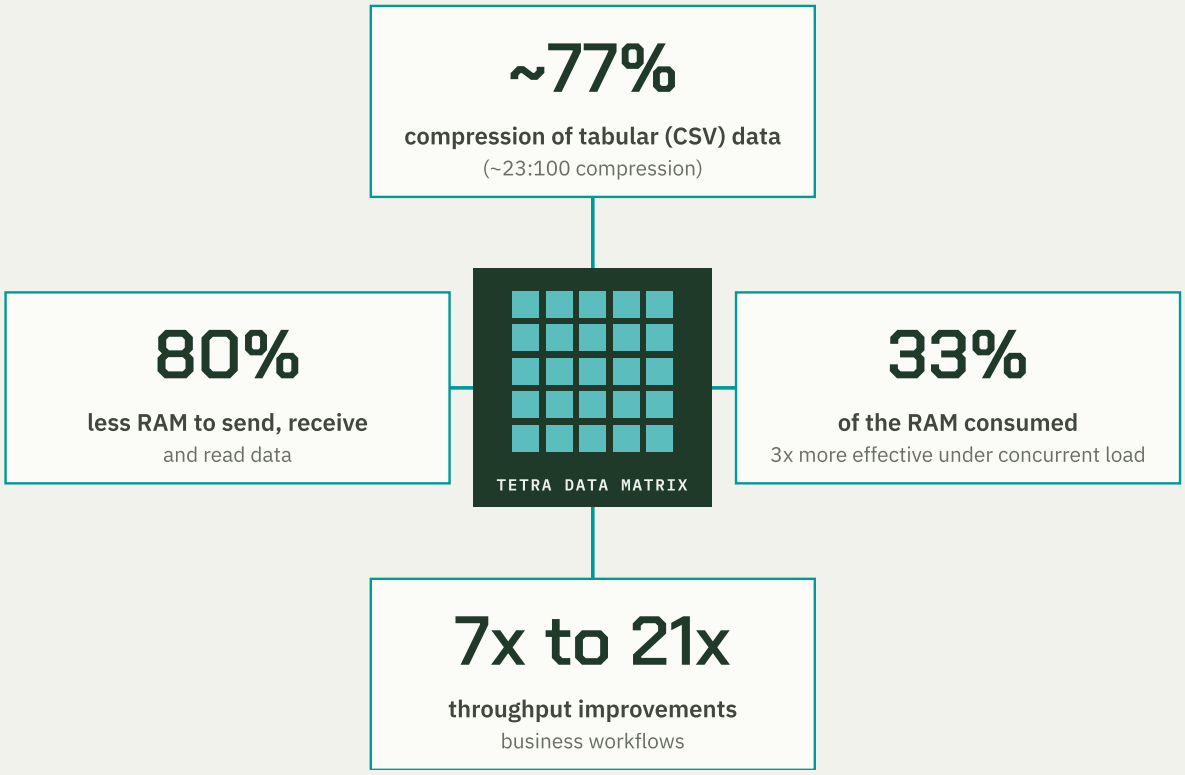
The data matrix leverages math and computer science principles from High Performance Computing (HPC). The same technology that maximizes every watt in a supercomputer drives more efficient data center utilization.

Our data matrix compresses tabular (CSV) data by about 77% (~23:100 compression); this equates to about 1/8th of a normal database's storage footprint (12.5%).

The same format gets stored, transmitted, and processed. As a data transmission format, the data matrix reduces the memory (RAM) requirements of sending, receiving, and reading data by 80%.

As a data system, the matrix database performs 3x more effectively under concurrent load than comparable industry-leading technologies. More data processed and transferred, 33% of the RAM consumed.

Using the data matrix to represent and execute business workflows shows 7x to 21x throughput improvements while decreasing overall memory. The processing utilization goes up, but this reflects greater utilization efficiency. The overall system impedance plummets, resulting in dramatically lower operating costs.



Computers perform matrix math better than they perform "scalar" math. By creating vector-oriented approaches to traditional business logic, TETRA bridges the gap between today's systems and efficient operations.

SECTION 06

HORNESCI – POWERING TOMORROW'S INTELLIGENCE

Hunter Horne founded HorneSci after spending years supporting contracted work within the prestigious United States National Labs ecosystem.

HorneSci optimizes the way that microchips execute a calculation. They developed the Hierarchical Sparse Tensors (HST) framework, productized as "DeltaRouting." They hold two patents and two pending patents for the technology.

Current systems give the operating system instructions, and the operating system delegates a calculation instruction to the microchip. HorneSci observed that this process could be made far more efficient and economical. DeltaRouting leverages telemetry – operational signals emitted by the chip – to plan and execute calculations. These optimized instructions reduce the amount of electrical change (energy movement) required to get a result. Less change means less work, and less work means lower effort for higher throughput.

None of this requires a new chip design. The system works with today's technology.



15x to 21x

tensor-based system acceleration

HorneSci's benchmarks indicate that they accelerate tensor-based systems by 15x to 21x, with reduced memory usage. Moreover, DeltaRouting efficiency gains translate to reduced operating heat.

> The same heat that requires water and noisy fans.

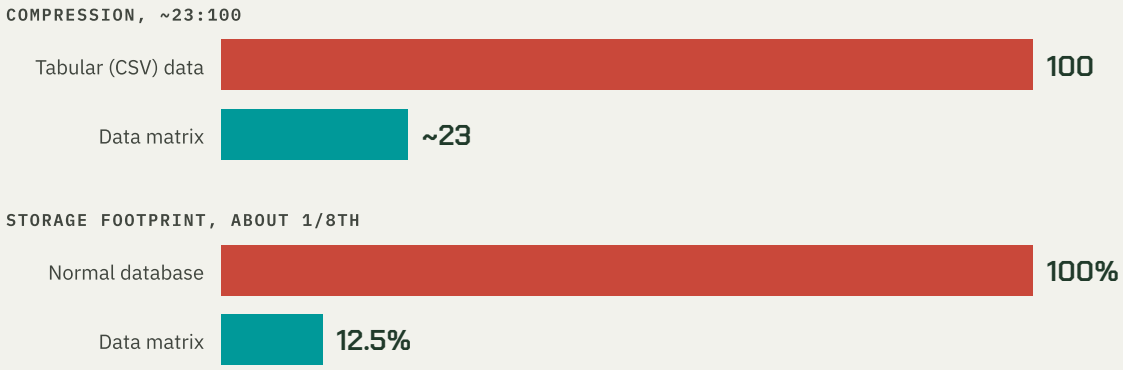
The operations that DeltaRouting optimizes best correspond to how TETRA's data matrix operates. By combining data matrix and DeltaRouting technologies, software optimizations can substantially impact data center energy cost.

SECTION 07

BENCHMARKS AND PROJECTIONS

Current TETRA CA benchmarks do not yet include HorneSci's DeltaRouting acceleration.

Compared to current database systems, we take ~12.5 of the space on disk.



The impedance-free data matrix format across processing, networking, storage, and transmission shows lower memory consumption across business logic workflows. Compared to traditional business workflow engines, we see speed-ups ranging from 7x to over 20x. This shared format that enables storage, processing, and transmitting the data matrix means more interconnected business workloads across fewer machines. This means reduced networking overhead, fewer physical machines, and less energy overhead, still serving identical business workloads.

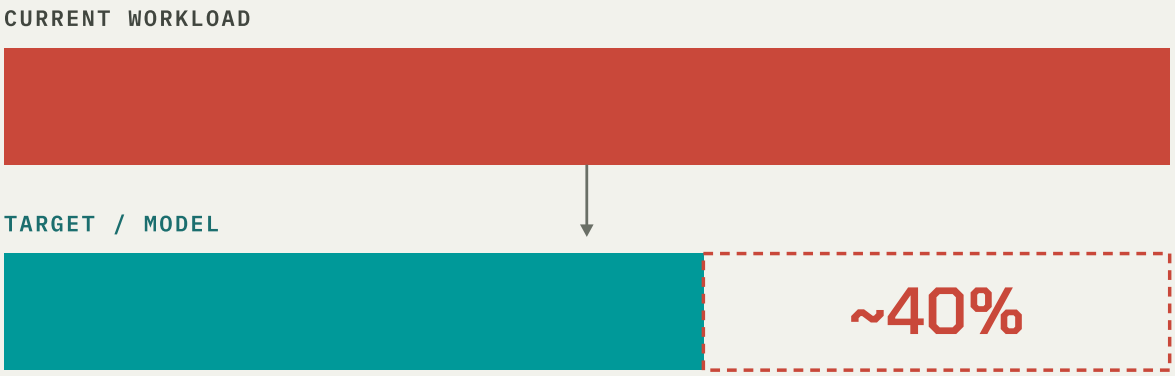
The compression technique applied to the data matrix works across operating systems and processors; the results are deterministic.

Under load, the TETRA CA database (TETRAbase) uses ~1/3rd of the memory compared to industry leading data systems.

These data systems show “sub-linear” scaling. This means that twice the data costs less than twice as much to operate, enabling TETRA to drive greater data center savings than existing systems.

System-level modeling based on benchmarks suggests ~40% total data center energy cuts based on software alone. This is achieved by:

- Combining existing systems onto fewer machines
- Eliminating the “translation tax” between systems and improving networking format efficiency
- Removing compliance, monitoring, security layers (the data matrix can manage these concerns endemically rather than bolting on more translations)



The data matrix reflects a massive step forward for representing information. This, in turn, reduces infrastructure requirements at the hundreds of terawatt hours per year. These savings do not just mean a one-time reduction followed by a new spike.

Based on our benchmarks with the data matrix, we assess the expense to operate the data systems and they grow. We show that the system scales sublinearly; twice the data costs between 60% and 80% more to operate as opposed to >%100 more. This means that the energy savings of the system compound over time, curbing future datacenter growth.



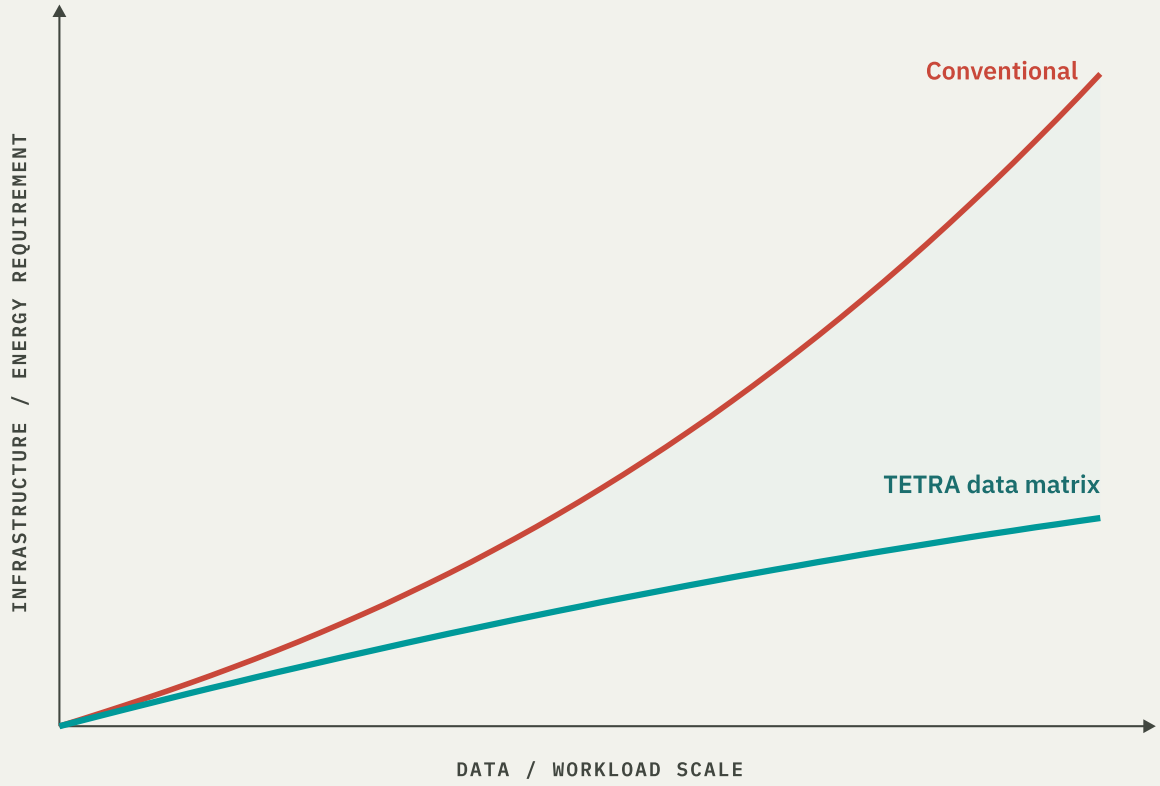
SECTION 08

LESS IS LESS

Compute infrastructure follows non-intuitive scaling laws. This means that 2x the data costs more than 2x to operate. It also means that 2x efficiency gains *reduce* system requirements by more than 2x! In other words, many of the systems that operate data systems exist today because of how software plans and executes work on servers.

As technology like the TETRA Data Matrix and HorneSci's DeltaRouting permeates data centers, we do not just see compute workloads get lighter. Cooling requirements *also drop* precipitously.

HorneSci's microprocessor-based telemetry with operational optimization means the chip does not just perform more work using less energy and heat.



> Less infrastructure scales down, which means massive infrastructure retraction.

SECTION 09

WHAT ABOUT AI?

LLM powered AI agents act upon data systems differently than human users. Bots represent more internet traffic than humans now. Massive parts of our public facing data infrastructure get used entirely by information scraping, non-human entities.

AI systems use more energy than normal systems, AND they use normal systems less efficiently than human users. And LLM driven agent will make multiple requests that a system may not be designed to handle. The unexpected activity can include load and access patterns, requesting data with no index (fast look-up), or requesting tangential, unrelated information.

To address this, the matrix database employs an index-free, geometric lookup pattern. This means depth traversals and index-free lookups become efficient, enabling agents to query data without introducing additional system load.

As far as leveraging the data matrix and DeltaRouting to accelerate LLM/AI operations for more efficiency, this will be investigated by TETRA CA and HorneSci over the coming months.

LLM execution technologies are nascent, and we expect efficiency improvements over the coming years. As AI-independent data storage, processing, and retrieval technology, the data matrix and DeltaRouting enable LLMs to operate upon data more efficiently.

SECTION 10

CLEANER, GREENER TECHNOLOGY

> Software leads the way.

Datacenters exist to host and operate software. The components that operate the software take the largest proportion of energy. While we understand the servers themselves, their motherboards, processors, and energy expenditure at idle or at work, we have an insight gap: the way software gets constructed and configured impacts energy usage at least as much as the business workload itself.

By creating a new technological representation of data, and managing data using efficient vector mathematics, software can do more using existing hardware. End to end, TETRA and HorneSci have demonstrated that the technique reduces software memory and storage while improving throughput.

To fully recognize the benefits of these data management systems, a full-scale changeover would be ideal. However, the economic impacts of this changeover would require massive investment from companies to modernize and rework their data systems. To address this, TETRA CA creates interoperable interfaces that can integrate with existing systems. Via integrations, organizations can leverage the data matrix on their most expensive data operations today.

SECTION 11

BIBLIOGRAPHY

Adamson, J. (2017). *Carbon and the cloud*. Stanford Magazine. https://stanfordmag.org/contents/carbon-and-the-cloud

Andreessen, M. (2011, August 20). *Why software is eating the world*. Andreessen Horowitz. https://a16z.com/why-software-is-eating-the-world/

Autodesk. (n.d.). *A guide to data center construction*. Digital Builder. Retrieved September 12, 2026, from https://www.autodesk.com/blogs/construction/data-center-construction/

Data Center Map. (n.d.). *Data centers*. Retrieved September 12, 2026, from https://www.datacentermap.com/datacenters/

Goldman Sachs. (2024, September 4). *Cloud revenues poised to reach $2 trillion by 2030 amid AI rollout*. https://www.goldmansachs.com/insights/articles/cloud-revenues-poised-to-reach-2-trillion-by-2030-amid-ai-rollout

Hassan, A. (2026, June 17). The cloud has sound: The unrelenting and unseen cost of the A.I. boom. *The New York Times*. https://www.nytimes.com/2026/06/17/us/data-centers-noise-pollution.html

Shehabi, A., Smith, S. J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M. A. B., Holecek, B., Koomey, J., Masanet, E., & Sartor, D. (2024). *2024 United States data center energy usage report* (LBNL-2001637). Lawrence Berkeley National Laboratory. https://eta-publications.lbl.gov/sites/default/files/2024-12/lbnl-2024-united-states-data-center-energy-usage-report.pdf

Tanner, B., Belle, D., Kerry, C. F., Kyosovska, N., Renda, A., Tabassi, E., & Wyckoff, A. W. (2026, April 10). *Global energy demands within the AI regulatory landscape*. Brookings Institution. https://www.brookings.edu/articles/global-energy-demands-within-the-ai-regulatory-landscape/

U.S. Department of Energy. (n.d.). *National laboratories*. Retrieved September 12, 2026, from https://www.energy.gov/national-laboratories

Vertiv. (n.d.). *A beginner's guide to data center cooling systems*. Retrieved September 12, 2026, from https://www.vertiv.com/en-us/insights/articles/educational-articles/a-beginners-guide-to-data-center-cooling-systems/

TETRA CA PBC and HorneSci

tetra-ca.com

inquiries@tetra-ca.com