Technical Research Brief

PROWESS

Why Open, Modular AI Data Platforms Win Over Closed, Storage-Embedded AI Data Stacks

AI at scale requires open, orchestrated data pipelines—not storage-bound silos.

How Structural Shifts in Enterprise AI Are Changing Data Platform Requirements

Enterprise AI is undergoing several shifts as organizations move to production-grade deployments. While access to GPU resources continues to expand, the limiting factor for many initiatives is no longer compute, it is data architecture. AI workflows now depend on unstructured data spread across warehouses and data lakes, in addition to object stores and edge locations. Pipelines must sustain continuous data creation and reuse across training, inference, and retrieval-augmented generation (RAG). These demands expose the limits of architectures that focus on storage performance alone.

As a result, enterprises are increasingly prioritizing data platforms that make enterprise-wide data access and processing fundamental platform features. Closed or proprietary systems can force unnecessary data movement, lead to vendor lock-in, and make operations more complex, ultimately hindering the organization's ability to iterate quickly. At the same time, un-optimized data-prep and extract, transform, and load (ETL) software can saddle GPUs with needless reformatting, redundant copies, and avoidable shuffles, which can waste resources and inflate infrastructure costs.

These pressures are driving demand for platforms that enable federated access alongside accelerated processing and native support for AI-driven search across the data layer.

Report Methodology
For this brief, Prowess Consulting reviewed publicly available technical documentation, vendor blogs, and industry press from Dell Technologies and VAST Data. We used these sources to validate key architectural descriptions and confirm the public basis for comparative claims. Our review focused on architectural design, data-access patterns, RAG search capabilities, and operational factors such as density and GPU utilization.

Two Solution Paths: Different Approaches to the AI Data Layer

For this study, Prowess Consulting evaluated Dell™ AI Data Platform and VAST® AI OS because these solutions tackle the challenges facing organizations through very different paths. That split in approach shapes everything from how quickly each platform scales to how much engineering work the user owns and how easily teams can adapt their environments as their AI needs evolve.

Dell AI Data Platform

Dell AI Data Platform is a modular data foundation for AI workloads. It spans storage, compute, and data services on established Dell Technologies infrastructure drawn from the company's portfolio of scale-out storage and server systems.

Rather than centering on a single engine and forcing data consolidation, the platform emphasizes depth by incorporating storage, processing, analytics, and orchestration engines that operate across the entire data estate. This is a design choice that underpins the solution's approach to federated access, accelerated processing, and RAG.

VAST AI OS

VAST's path centers on unifying storage and data services within a single platform and consolidating AI workloads there. Rather than extending across existing infrastructure layers, this approach emphasizes operating data access, query, and AI-related processing within a vertically integrated software environment. Storage, database functions, and execution capabilities are delivered as part of the same platform. This design centers on operating AI workloads within a consolidated platform environment, rather than distributing storage, processing, analytics, and orchestration engines across the broader data estate or emphasizing federated access and engine-level specialization.

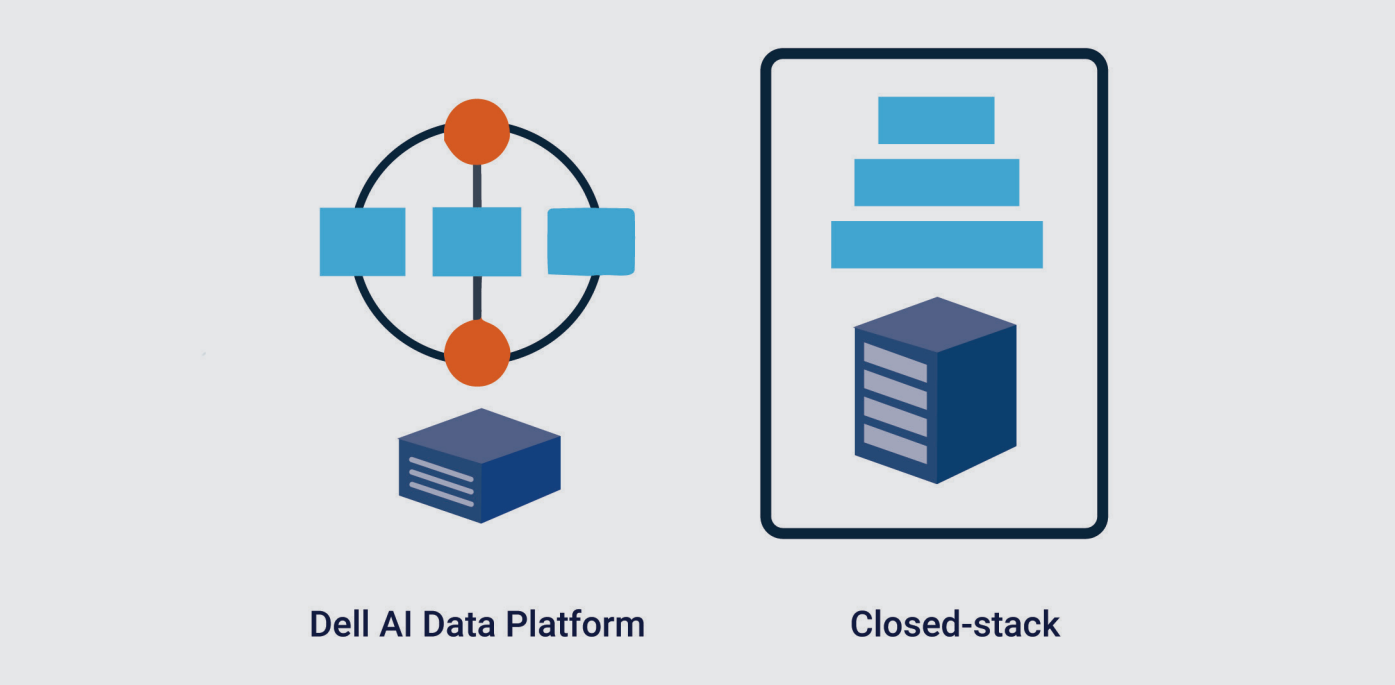

Figure 1. Architectural scope: modular Dell AI Data Platform versus a closed-stack platform like VAST AI OS

Operational Considerations

How a platform handles day-to-day operations often determines whether AI projects launch smoothly or stall in pilot. Integration effort is one of the most overlooked cost drivers in enterprise AI. The more components teams must assemble themselves, the higher the operational burden, and the slower AI initiatives are to reach production. Our comparison found the following impactful differences.

Data Preparation and Enrichment

Dell AI Data Platform streamlines data preparation and enrichment for AI pipelines by blending its data engines and management tools with an architecture built to work across the data sources organizations already use. Instead of requiring teams to gather all their data in a single, vendor-controlled repository, this solution lets teams prepare and reuse data in place. This model preserves existing governance and security practices and lets teams scale storage and processing independently as their needs evolve.

While Dell AI Data Platform makes it easy to enrich data across different sources, VAST AI OS works best when data is already inside its own environment. If organizations want to use external data, teams must often take extra steps to bring that information into VAST AI OS and keep it up to date. Tools like VAST SyncEngine can help with moving and syncing datasets, but these additional steps can make the initial setup take longer, and they can add more complexity as data sources and pipelines evolve.

Practical Impact

When analytics and enrichment require data migration or ongoing synchronization, teams face longer onboarding cycles and higher operational overhead. In contrast, federated analytics and automated enrichment let organizations prepare and reuse data where it lives, shortening time to value and reducing the burden on data engineers.

Data Access and Integration

The Dell Technologies platform lets teams query data across databases, data lakes, or object stores, without copying that data into a new environment. This approach reduces data movement and helps maintain consistent governance as projects grow.

Public descriptions of VAST AI OS deployments show that when organizations need to integrate external data sources with VAST AI OS, they may use tools like SyncEngine to move and sync data into the VAST DataSpace. The data migration step differs from approaches like Dell Technologies' federated access model.

Practical Impact

By supporting federated access to existing data estates, Dell AI Data Platform helps organizations reduce time spent preparing and relocating data, making it easier to extend AI pipelines to new use cases.

Processing and Acceleration

Dell AI Data Platform integrates Apache Spark™ with the NVIDIA® RAPIDS accelerator, allowing data processing and AI workloads to tap GPU acceleration. The platform's analytics engine, powered by Starburst, uses advanced file readers, autonomous indexing, and intelligent reuse of common table expressions to eliminate redundant scans. These architectural advantages are underscored by recent Dell Technologies and NVIDIA benchmark results, which report up to 12× faster vector indexing, 3× faster data processing, and a 19× reduction in time-to-first-token compared with traditional CPU-based workflows—demonstrating the platform's ability to meaningfully accelerate end-to-end AI performance at scale.[1]

Even as VAST Data's hardware foundation evolves, its AI OS data-processing stack still depends on software workflows that lack optimizations designed to eliminate redundant data transformations, excess memory movement, and costly shuffle patterns. When unoptimized, these operations can burden GPU pipelines and reduce overall efficiency.

Practical Impact

AI pipelines depend on fast, predictable, and efficient data processing stages. Only then can teams be free to focus on features instead of infrastructure tuning. In this way, acceleration improves GPU utilization efficiency and can reduce both processing time and the engineering work usually required to tune drivers and libraries. By optimizing data prep workflows and feeding GPUs more efficiently, Dell AI Data Platform can sustain higher GPU utilization efficiency during training and inference, which directly reduces cost per workload.

Search and RAG Workflows

Dell AI Data Platform includes a Data Search Engine, powered by Elastic, designed for AI use cases. It supports vector search, GPU-accelerated index creation, connectors for common enterprise content sources, and hybrid keyword/vector retrieval. These capabilities support both semantic search and traditional enterprise discovery within the same environment.

VAST AI OS supports search and retrieval through native indexing and vector embedding capabilities integrated into its data platform. The system maintains a unified namespace, with search and retrieval operations executed directly within the VAST Data environment. These capabilities are used to support RAG workflows, including both semantic and similarity-based retrieval, only on data stored and managed by the platform.

Practical Impact

RAG workloads can fail when indexing is slow or when content pipelines require extensive custom glue code. The Dell Technologies platform's GPU-accelerated search helps teams deploy production-grade AI assistants and retrieval workflows with fewer moving parts.

Cost and Efficiency Factors

Cost and efficiency often determine whether an AI platform scales smoothly or becomes a budget anchor. Our comparison found meaningful differences in how Dell Technologies and VAST Data solutions approach density, GPU utilization, and system overhead.

Dell PowerScale™ and Dell ObjectScale™ operate as the integrated data foundation of Dell AI Data Platform. These storage platforms unify enterprise-grade file and cloud-native object storage into a single, intelligent data fabric optimized for end-to-end AI workflows. PowerScale delivers massively parallel performance for high-throughput training and inference, while ObjectScale provides elastic, multi-tenant object services for scalable data lakes and long-term model and dataset storage.

Together, they deliver consistent data access, lifecycle management, cyber resilience, and policy-driven governance across the entire AI pipeline—spanning from ingest and preparation to training, fine-tuning, deployment, and protection. This approach ensures organizations can confidently scale AI with predictable performance and enterprise-class reliability.

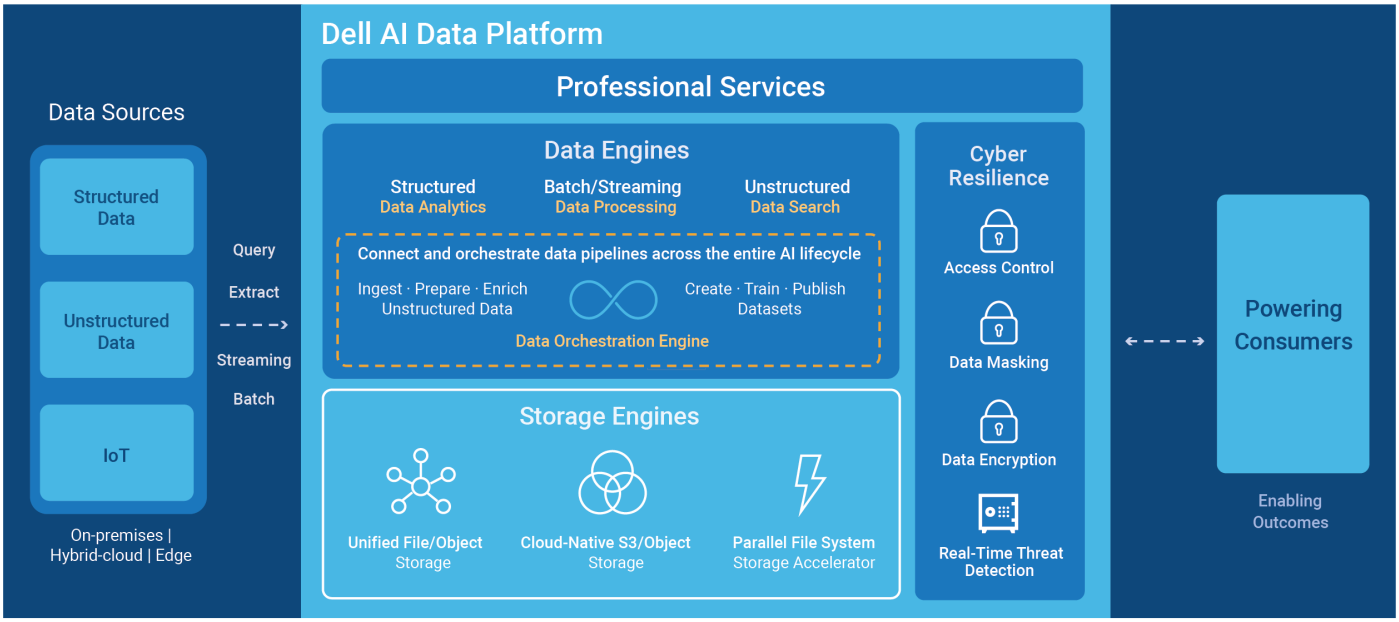

Figure 2. Dell AI Data Platform delivers a modular, end-to-end architecture that connects diverse data sources to AI consumers through integrated data, storage, and cyber-resilience services

Matching Solutions to Business Needs

Based on the cost, efficiency, and operational differences outlined above, we found that the Dell Technologies solution we examined is a strong match for organizations that want a unified AI data foundation with GPU-accelerated processing and support for open table formats. Dell AI Data Platform runs on high-density Dell Technologies systems designed for AI scale. It's also a good choice for teams that prioritize predictable performance for AI pipelines.

On the other hand, we found that the VAST Data solution could suit teams that prefer a storage-centric approach, that do not have an existing AI infrastructure, and that are comfortable owning more of the integration and operational design.

Conclusion

As AI initiatives move from pilot to production, data platform design increasingly shapes operational complexity and costs over time. Platforms with predictable performance that reduce data movement and simplify integration can shorten the path to production. Those that emphasize flexibility at the data layer often shift more responsibility to internal teams to assemble and maintain the surrounding AI stack.

This comparison highlights how architectural choices shape where operational responsibility falls within AI environments. Our comparison found that the Dell Technologies approach consistently shifts more of the integration and operational burden onto the platform itself, while the VAST Data storage-embedded design requires customer teams to assemble and manage more of the surrounding AI stack. We also found differences in how each platform fits into existing enterprise ecosystems, with Dell AI Data Platform designed to integrate alongside established data, analytics, and AI tools, while VAST AI OS favors a more self-contained, storage-centric environment.

A clear-eyed view of these trade-offs helps organizations evaluate AI data platforms based on how much complexity they are willing to carry into production.

To explore Dell AI Data Platform reference architectures and validated designs in more detail, visit **the Dell AI Data Platform page**.

Endnotes

[1] Dell Technologies. "**Dell AI Data Platform with NVIDIA Supercharges Enterprise AI with Breakthrough Data Orchestration and Storage Innovations.**" *PR Newswire.* March 2026.

0426/260004