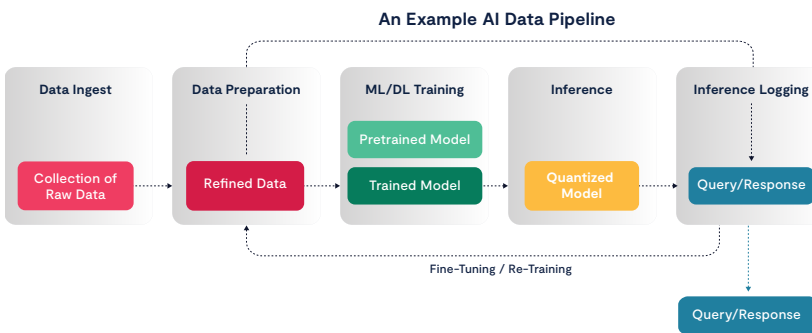


# Simplifying and Scaling AI Workloads



## Introduction (The Challenge)

Enterprises are rapidly adopting Artificial Intelligence (AI) to gain new insights and competitive advantage, but they are encountering significant challenges in designing and constructing optimal infrastructure. AI models require large datasets for training, and businesses are increasingly reliant on AI for critical tasks, highlighting the need for extremely reliable, secure, and high-performance infrastructure. Enterprises require a straightforward solution that can be deployed rapidly, deliver exceptional performance for AI workloads, and seamlessly scale as AI initiatives expand. The infrastructure must be resilient, multi-tenant, and comply with strict enterprise security standards, while also being user-friendly for enterprise IT teams to support.



## The Solution

VAST Data and Arista Networks have joined forces to offer a simplified AI infrastructure solution suitable for organizations at any stage of their AI journey. This solution can seamlessly scale to meet the needs of large GPU cloud service providers and Enterprises alike. The VAST Data Platform’s innovative Disaggregated Shared-Everything (DASE) architecture allows customers to independently scale compute and capacity to enable true “scale as you grow” economics and performance. It achieves this by deploying stateless controllers that access high-performance NVMe devices, providing all-flash file, object, and tabular data services optimized for GPU-accelerated computing. The foundation of this architecture relies on Arista Networks’ high-speed, low-latency networking to ensure resilient data access.

## Why VAST & Arista?

AI workloads can demand complexity from any infrastructure environment but Arista and Vast deliver simplicity, observability, and scalability for any situation, enabling organizations to flourish in a data-driven, AI-focused world.



### Simplicity

Network and storage compositibility enables a best-of-breed approach to a scalable, high-performance, and zero-trust based AI storage infrastructure utilizing Arista and Vast Data.



### Observability

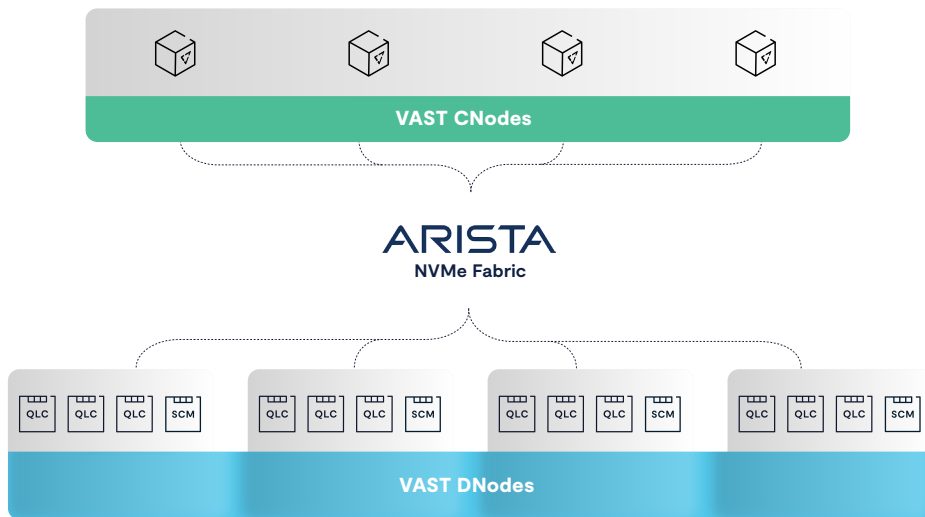
Ensuring that data is flowing from edge to core to cloud in order to meet the needs of data-hungry algorithms and services is a critical job that can be easily managed using Arista and Vast Data’s complementary observability tools.



### Scalability

Arista and Vast Data offer a scalable, performance-driven solution that can grow with your business, ensuring support for AI’s ever-increasing demands.

With VAST and Arista, organizations can consolidate enterprise and AI workloads onto a unified, multi-tenant infrastructure designed to streamline their AI data pipelines, simplify management, and reduce complexity. Together, VAST Data and Arista provide a secure, scalable, and high-performance infrastructure that empowers enterprises to unlock the full potential of AI, from initial exploration to production-scale deployments, while meeting the strictest security and reliability requirements.



Since the bandwidth and scale requirements for AI networks will vary from customer to customer and application to application, Arista Networks, an industry leader in AI-driven networking for large data centers, campus, and routing environments, offers a wide breadth of hardware platforms to cater to different verticals. In conjunction with Vast’s disaggregated, shared everything (DASE) infrastructure, they’ve come together to provide a sample set of configurations that scale with changing business requirements.

## Conclusion

The rise of Large Language Models (LLMs) drives an ever-increasing demand for GPUs and storage capacity. Modern AI applications necessitate expansive clusters, often comprising thousands of GPUs and storage devices, with the ability to scale to tens of thousands as demands escalate. With GPU speeds doubling every other year, it is crucial to prevent compute and network bottlenecks through scalable network design. While application teams enhance compute capacity, network teams must meticulously evaluate the interconnect to ensure optimal performance. Networks that balance performance with simplicity, observability, and scalability are poised to be at the forefront of this AI revolution.

Similarly, AI’s demand on storage to provide multi-protocol, scalable, and performance-oriented access to both structured and unstructured data requires reimagining storage solutions. Creating an independently scalable, secure, and multi-tenant solution utilizing the performance and ubiquity of Ethernet demonstrates that current and future challenges can easily be accounted for using proven technologies.

Arista and Vast Data together represent the ideal combination of simplicity, observability, and scalability through proven technologies like Ethernet and NVMe over Fabrics. Utilizing their combined resources, customers can be assured that no matter the challenge presented by AI, they are ready.

## About Arista

Arista Networks is an industry leader in data-driven, client to cloud networking for large data center/AI, campus and routing environments. Arista’s award-winning platforms deliver availability, agility, automation, analytics and security through an advanced network operating stack.

For more information, visit [arista.com](https://arista.com)

## About VAST Data

VAST Data is the data platform company built for the AI era. As the new standard for enterprise AI infrastructure, organizations trust the VAST Data Platform to serve their most data-intensive computing needs. VAST Data empowers enterprises to unlock their data’s full potential by providing simple, scalable, and architected AI infrastructure to power deep learning and GPU-accelerated data centers and clouds. Launched in 2019, VAST Data is the fastest-growing data infrastructure company in history.

For more information, visit [vastdata.com](https://vastdata.com)



[Scan Here](#) learn more about how VAST and Arista.