

Is Your Storage Ready for AI?

Accelerating highly variable AI storage workloads while improving storage efficiency is crucial to realizing the full value of AI. Learn how to get started building an accelerated, efficient and scalable AI data storage pipeline.

Key Benefits

- Manage complex AI data with an accelerated, efficient and scalable data storage pipeline
- Modernize your storage to be AI-ready: Store more, store more efficiently, and plan for scalability
- Tune your storage infrastructure for AI with a common storage pipeline and then optimize for AI workloads
- Intel® Optane™ media optimizes, stores, and moves larger, more complicated datasets through the AI pipeline

Today, Artificial Intelligence is transforming everything from medicine to manufacturing, from smart cities to self-driving cars.

And it's growing by leaps and bounds. A recent Gartner study reported that "The number of enterprises implementing Artificial Intelligence grew 270 percent in the past four years and tripled in the past year".¹ Chris Howard, a Gartner VP, summed it up: "If you are a CIO and your organization doesn't use AI, chances are high that your competitors do and this should be a concern".¹

"Artificial intelligence, whether machine learning or deep learning, requires IT organizations to think about data and storage architecture differently from those supporting more traditional enterprise workloads."

- Patrick Moorhead

The catch is that AI requires data. Lots of it. And lots of different kinds at lots of different times.

The global consultancy IDC projects that "the Global DataSphere will balloon to 175 zettabytes by 2025".² (A zettabyte is roughly 1000 exabytes or a billion terabytes; if each terabyte in a zettabyte were a kilometer, it would be equivalent to 1,300 round trips to the moon and back.)³

Most companies are not ready for that. A recent IBM report suggested that the average factory analyzes less than 1% of its data in real time.⁴

Today, every business struggles with efficient storing of increasing data. Yet among the dizzying accolades of big data and AI, storage is rarely discussed. The first step is to understand how complicated AI data can be. It is often talked about as the three Vs: Volume, Velocity, and Variety.

- Volume**
 In AI, as training data grows, the algorithms get smarter. Tesla is building their own AI infrastructure, which they say will “process the thinking algorithms” for their Autopilot autonomous driving software. To do this, they are amassing 1.3 billion miles of driving data. And Microsoft required five years of continuous speech data to teach computers to talk.⁵ Neil Stobart of Cloudbian points out that “managing these datasets requires storage systems that can scale without limits”.⁶ When it comes to data and AI, “more is more.”
- Velocity**
 IDG says that data is becoming more and more critical, which means it must be able to be accessed instantly: “The trend away from consumers and entertainment as the primary creators of data will see enterprises as the (source) of 60% of the world's data by 2025. We are transitioning from a period in which information has been transformed from analog to digital to one in which digital information will increasingly

be a critical part of systems required for everyday life-critical systems that use analytics, machine learning, and IoT. Nearly 20% of the world's data will be critical to our daily lives by 2025, and nearly 10% of that will be “hypercritical”.⁷

- Variety**
 Variety refers to the format of the data. And as businesses look to improve customer experiences, run more efficiently and stay competitive they are analyzing data across a wider range of formats. To illustrate, imagine a retail company with the objectives of optimizing in-store experiences, increasing customer loyalty, driving higher per visit purchases, and improving supply chain efficiency. Accomplishing these goals could involve ingesting data from online purchases, social media engagement, SCM data, customer service and returns, in-store cameras, customer in-store location, and on-shelf monitors. That's a wide variety of files, clicks, texts, videos, machine data, and bluetooth signals.

The solution to the three V's is to create a common data pipeline underlying the various AI functions with a tier optimized for space-efficient capacity/scaling, and another tier optimized for performance storage/scaling. If you get this pipeline right, you can begin to resolve your current data issues—and be more equipped to address the complicated AI data storage architecture.

How can you start building an efficient, scalable, future-proof AI infrastructure?

Modernize to AI-Ready Storage

Store more

The first thought might be: More data? Get more drives. But adding HDDs doesn't work.

When they scale, performance per Gigabyte decreases, risk of failure from moving parts increases, and you are left with a large, inefficient, expensive footprint of dead-end technology. Even

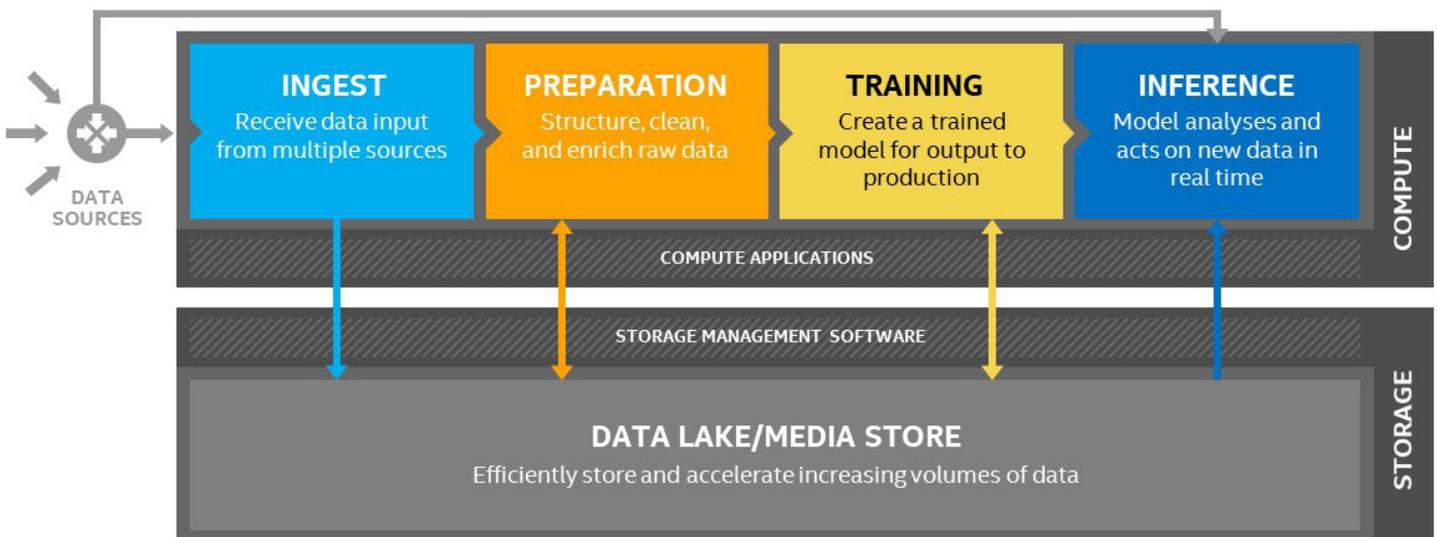


Figure 1
 Implement a common data pipeline to manage the highly complex data of AI workloads.

SATA-based solid state drives (SSDs) may not have sufficient performance for some AI implementations and, like HDDs, they are built on an interface that is not innovating for performance, manageability, and capacity scaling.

Store more efficiently

Today, more and more businesses are opting for PCI Express (PCIe), which supports the Non-Volatile Memory Express (NVMe) protocol, providing the required low latency and high throughput demanded by AI workloads.

Mary Brascomb on DataCenterKnowledge.com notes that “The higher performance and lower system requirements of NVMe as the interconnect for flash drives and arrays are driving acceleration in the enterprise”.⁸ Many leading storage solutions utilized for AI storage systems (Ceph, VMware vSAN, Microsoft Azure Stack HCI) are taking advantage of these hardware advances to optimize efficiency.

Plan for scalability

Data volumes continue to grow and AI model complexity is increasing. Storage infrastructure needs scalability to keep pace.

A roadmap of innovations on the PCIe/NVMe interface offer better IOPS/TB and deliver continued advances in serviceability, performance, manageability and form factor, optimizing storage footprints and operational efficiency. This has led to PCIe shipments outpacing both SAS and SATA shipments combined. With PCIe, AI’s unpredictable mixes of random and sequential reads and writes across variable workloads and fluctuating sizes becomes manageable.

The combination of Intel’s storage technologies, including Intel® Optane™ SSDs and Intel® 3D NAND SSDs, are ideally suited to address AI storage constraints. Intel Optane SSDs provide the unique combination of breakthrough performance and low latency needed for the performance storage tier, while Intel® QLC technology offers space-efficient, large capacity.

Tune Your Storage Infrastructure for AI

Build a common storage pipeline

A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale. AWS points out that this means “you can store your data as-is, without having to first structure the data, and run different types of analytics—from dashboards and visualizations to big data processing, real-time analytics, and machine learning to guide better decisions”.⁹

An Aberdeen survey shows companies with data lakes increased organic growth by 9% more than like companies with traditional siloes. A data lake gives

you one source of truth, with unified and democratized access to all your information.¹⁰

Optimize for AI workloads

Datasets can arrive in the pipeline as petabytes, move into training as gigabytes of structured and semi-structured data, and complete their journey as trained models in the kilobyte size.

Workloads are also variable, starting with ingest of 100% writes, progressing to preparation where they can reach 50/50 read/write mix, and finally shifting to training and inference of 100% read. To complicate matters, access patterns are wide ranging, from ingest as sequential to training as highly randomized (to help improve model accuracy.) And with the variability comes the demanding requirements of high throughput and extremely low latency, no matter the workload.

Traditional NAND SSDs may be strained to meet these requirements across the full data pipeline. Intel Optane SSDs, however, offer the breakthrough performance required for the IO demands of AI workloads. For example, Intel® Optane™ technology’s consistently low latency can enable faster time to trained models, and high throughput in presence of buffer destage can improve ingest performance.

A thorough analysis of current and anticipated needs for each stage will ensure your performance storage is AI-ready, with scalable, future-proof infrastructure.

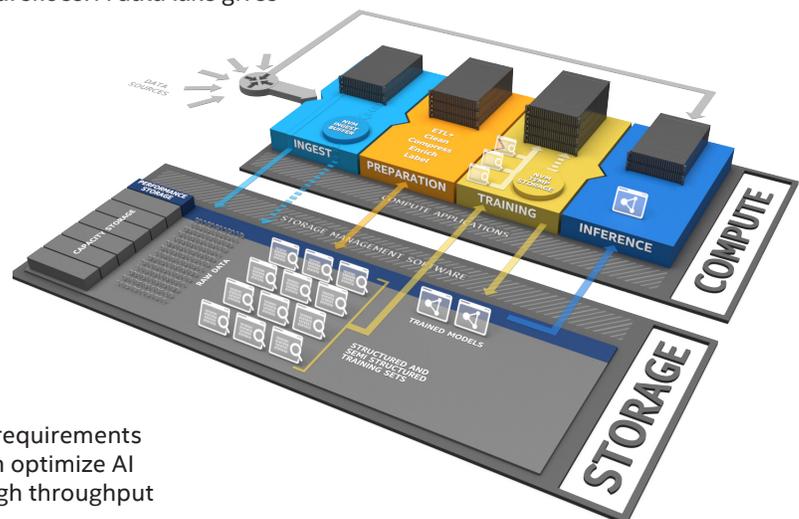


Figure 2

Across the data pipeline, from ingest to inference, IO requirements are highly variable and extremely demanding. You can optimize AI workloads with Intel storage solutions that provide high throughput and low latency, as well as high capacity.

Who's Doing It Right?

High performance Intel Optane SSDs and space-efficient capacity 3D NAND SSDs are quickly gaining traction across both OEMs and end customers as the industry recognizes the important role storage performance, capacity, and scalability play in improving AI efficiency and outcomes.

Some companies employing a modern AI infrastructure are Dell EMC, Baidu, and iFLYTEK.

Dell EMC partnered with Intel to create a solution to deliver storage capabilities for the full AI life cycle. It includes Dell PowerEdge servers with Dell EMC network switches, Isilon storage, and an optimized software stack. And Intel® Optane™ SSDs provide lower latency and higher throughput than standard NAND PCIe SSDs.¹¹

Patrick Moorhead, a leading technology analyst, notes the importance of the underlying software, EMC Isilon, pointing out that it simplifies and optimizes data storage at every stage of the AI workflow, "specifically low-latency block storage for real-time response during ingest, data prep and production inference".¹²

Moorhead summarizes the benefit: "Artificial intelligence, whether machine learning or deep learning, requires IT organizations to think about data and storage architecture differently from those supporting more traditional enterprise workloads. The attributes of the data are different. The complexity of the analytics is different. The needs of the consumers of that data are different. The ability to keep accelerated compute nodes fed with data is paramount. The Dell EMC Isilon-based AI solutions are designed for precisely these needs".¹²

VAST Data has been selected by a variety of industry-leading customers who are adopting AI to modernize their application environment. VAST has partnered with Intel to deliver an AI optimized storage solution which eliminates the compromises of storage scale, performance and efficiency to help organizations harness the power of AI as they evolve their data agenda in the age of machine intelligence. VAST as part of its Universal Storage Platform uniquely utilizes Intel Optane SSDs and Intel QLC 3D NAND technologies to break down the barriers to performance, capacity and scale - where flash economics make it possible for customers to afford flash for all of their AI training data.

According to IDC's Eric Burgener, Research Vice President, Infrastructure Systems, Platforms, and Technologies, "This design gives a single Universal Storage Platform system the ability to handle the low latencies required by transactional workloads as well as the high degrees of data access concurrency required by artificial intelligence/machine learning/ deep learning and other big data analytics workloads."¹⁴

Baidu has gained widespread recognition for its work in the area of search technologies. With over 100 billion pages, 2,000 Petabytes (PB) of data stored and 100PB of data processed per day¹, Baidu is well-versed in the technological challenges brought about by the storage of massive unstructured small files.

Baidu AI Cloud is following in the footsteps of their successful public cloud deployments with private cloud storage, including a new high-performance all-flash object storage solution with Intel® Optane™ DC SSDs and Intel® QLC 3D NAND SSDs that has been deployed for AI training and Media Asset Management.

The Intel® Optane™ SSD is used as cache to optimize the read efficiency and synchronization latency, boosting the metadata processing speed. Four Intel® SSD D5-P4320 drives are included with each storage server providing the large capacity storage.

The high price-to-performance ratio of the Intel® QLC 3D NAND SSDs ensure the high performance of this solution while effectively lowering the Total Cost of Ownership (TCO) for the system. [Read more](#) about the Baidu all-flash AI case study.

iFLYTEK focuses on voice recognition software and voice-based internet/mobile products. They are exploring AI applications in cognitive fields, specifically their "Super Brain Project" which "seeks to emulate human brain neurons in order to give the company's intelligent speech devices rudimentary human thinking capabilities."

For this deep simulation to work, massive data training is required, which means enormous computing workloads. The deep learning infrastructure of speech recognition links computing resources to a parallel file system over high speed networks, where the compute engine is used in various types of training and computing.

To create the necessary low latency and high throughput performance required, iFLYTEK paired 2nd Gen Intel® Xeon® Scalable processors with Intel® Optane™ SSDs. The processor was well-suited to high-load parallel computing situations, and was reliable and scalable under high-performance workloads—perfect for iFLYTEK's complex neural deep learning networks. [Read more.](#)

"The success of ML and AI initiatives relies on orchestrating effective data pipelines that provision the high quality of data in the right formats in a timely manner during the different stages of the AI pipeline."

- Gartner, Three Ways That AI Will Impact Your Data Management and Storage Strategy¹³

Intel® Optane™ media, considered the first major memory and storage breakthrough in over 25 years, optimizes, stores, and moves larger, more complicated datasets through the AI pipeline. As Dell EMC, Baidu and iFLYTEK can attest, Intel® Optane™ technology is a great place to start. Read more about Intel® Optane™ technology.

Learn more

How Intel Storage Accelerates AI Deployments

Storage and AI: Financial Services Invests in Big Data and Automation

Watch: Building a Modern AI Data Storage Pipeline

Watch: Intel® Optane™ Technology Fuels Financial Services



To learn more visit www.intel.com/AIStorage

- 1 Source: <https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have>
- 2 Source: <https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html>
- 3 Source: <https://blogs.cisco.com/sp/the-zettabyte-era-officially-begins-how-much-is-that>
- 4 Source: <https://www.ibm.com/industries/manufacturing/smart-manufacturing-technology>
- 5 Source: <https://www.bloomberg.com/news/articles/2016-12-20/the-tesla-advantage-1-3-billion-miles-of-data>
- 6 Source: <https://www.cbronline.com/emerging-technology/ai-machine-learning-important-data-storage-requirements/>
- 7 Source: <https://www.computerweekly.com/news/450416206/Analytics-internet-of-things-to-drive-data-volumes-to-163ZB-by-2025>
- 8 Source: <https://www.datacenterknowledge.com/storage/how-latest-nvme-specs-make-flash-storage-data-center-better>
- 9 Source: <https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/>
- 10 Source: <https://s3-ap-southeast-1.amazonaws.com/mktg-apac/Big+Data+Refresh+Q4+Campaign/Aberdeen+Research+-+Angling+for+Insights+in+Today's+Data+Lake.pdf>
- 11 Source: Intel® Optane™ SSD DC P4800X/P4801X product brief <https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/data-center-ssds/optane-ssd-dc-p4800x-p4801x-brief.html>
- 12 Source: https://www.dellemc.com/resources/en-us/asset/analyst-reports/products/storage/h17841_ar_enterprise_machine_and_deep_learning_with_intelligent_storage.pdf
- 13 Source: <https://www.gartner.com/en/documents/3888265/three-ways-that-ai-will-impact-your-data-management-and->
- 14 Source: <https://vastdata.com/resources/idc-technology-spotlight-vast-data/>