PRODUCED BY

MIT Technology Review

IN PARTNERSHIP WITH

PURESTORAGE®

Modern Storage Accelerates Data Insights, Speeding Innovation

Deep learning opens up new worlds of possibility in artificial intelligence, enabled by advances in computational capacity, the explosion in data, and the advent of deep neural networks. But data is evolving quickly and legacy storage systems are not keeping up. Advanced AI applications require a modern all-flash storage infrastructure that is built specifically to work with high-powered analytics.

Everyone remembers the experience of learning to drive a car: nerve-wracking first forays behind the wheel with a parent or instructor. Before long, however, the new driver gets used to operating the vehicle safely in a variety of conditions. Now imagine the prospect of training a self-driving car to understand what it means to drive safely. Unimaginable amounts of data are needed for deep learning (DL) applications in driverless vehicles: Call it driver's ed without the driver.

Audi and Tesla are among the first auto manufacturers to offer self-driving features in their cars so the driver can just sit back and relax. Audi is now working to offer vehicles with the next level of autonomous capabilities, building cars that can recognize and handle every possible scenario for highway driving. Tesla will increasingly rely on radar as well as ultrasonic technology for sensing and data-gathering to form the basis for its autonomous autopilot algorithms. The amount of data needed to train these self-driving applications is almost beyond comprehension.

Indeed, some estimates place the amount of autonomous-vehicle training data, in general, in the realm of petabytes. The problem, for many companies in many industries: Legacy disk-based storage systems cannot deliver data to the servers fast enough to keep them all busy. As DL—a subset of machine learning (ML), and an advanced form of artificial intelligence (AI)—begin to take hold across industries, companies often find their storage systems are not equal to these new data- and compute-intensive applications that will propel their organizations forward. The answer: high-performance, all-flash storage systems designed to support data-intensive applications involving parallel processing.

The new reality is that big data analytics drives game-changing insights and helps companies leapfrog over their competitors, transforming whole industries. DL is now enabling applications such as Google Brain, which (among other things) learned to identify whether an image contains the address of a building, based on its being "trained" with massive amounts of related data. We are just at the beginning of applying DL technology broadly; the possibilities are endless.

"This will become a competitive differentiator…The benefits of flash are so transformative that if you're not using it, you are behind the curve."

-SCOTT SINCLAIR
SENIOR ANALYST,
ENTERPRISE STRATEGY
GROUP

Data is the Fuel for AI and It's Changing

"Deep learning was first used by the hyperscale big data players," says Robert Ober, chief architect at NVIDIA. "The early traction was in revenue-generating things: ad placement, recommendations, Wall Street decision-making. The second area has been customer experience and stickiness," he adds. This includes things like image tagging and image translation, voice and speech recognition, video filtering, news feeds, and autonomous driving. The third wave includes Internet of Things, or IoT; failure prediction; fraud and threat detection; and surveillance.

Preeminent AI expert Andrew Ng believes DL will transform virtually every industry in the near term. Ng cites, as an example, the rise of voice-interface devices such as Amazon Alexa, which require high-quality speech recognition.

"Data is the fuel that keeps these advanced applications' engines going."

-ROY KIM, DIRECTOR OF PRODUCT MARKETING AT PURE STORAGE

The use of speech recognition, a DL application, has tripled since 2015, according to Ng's former firm, Baidu Research. Systems will need high-powered flash storage to be able to carry out the data-intensive "learning" processes required by speech recognition. Without that piece of the equation, there won't be enough bandwidth and IOPS (input/output operations per second) to train the system, and storage will become a bottleneck. According to Facebook, their DL workloads require a well-balanced system able to drive both ExaFLOPS of compute and billions of IOPS for storage performance.

This is because modern ML and DL workloads are completely different from the previous generation of file server applications. File sizes vary from small up to exceptionally large, and the access patterns may be random. Data access may be real-time or batched. With apps and data types evolving quickly, the most important thing is to have an elastic IT infrastructure that can deliver uncompromised performance for any type of unstructured data, for any modern workload.

"Data is the fuel that keeps these advanced applications' engines going," says Roy Kim, director of product marketing at Pure Storage, a maker of all-flash storage systems. "The nature of data is changing, and you need a platform purpose-built to deliver massive performance for modern data."

NVIDIA's Ober advises those getting started with DL to heed Ng's words: "The data is the differentiator." That means the amount and quality of the data used for training the neural net is what will determine the value you get out of it.

AI Needs a Modern Data Platform

If you look more closely at hard disk drives (HDDs) or solid-state drives (SSDs), it is not surprising that they are not up to the task of handling AI workloads. HDDs were state of the art when they made their debut in 1956, and they have been a mainstay of enterprise computing for decades. But as a modern storage system, HDDs fall short. Their performance is not sufficient for AI-related computing, and the spinning disks are vulnerable to failure and resulting data loss. Even newer SSDs employing flash memory chips (as opposed to hard disks) are designed in software and hardware enclosures of HDDs, unable to keep up with the massively parallel processing in AI, ML, and DL workloads.

These cutting-edge applications demand an all-flash storage system that is built from the ground up with a parallel architecture. DL training workloads require storage to deliver high performance for random access to any data, of any size. Without this critical piece of the equation, the big data insights will be paltry—like buying a Ferrari with the fuel-injection system of a Hyundai, turning the elite racing machine into a plodding suburban vehicle. DL needs maximum read performance to keep training computers busy. Without modern storage infrastructure, there is no point to investing in the latest big data analytics tools.

To be specific: What is needed now is a scale-out storage solution based on flexible and powerful flash technology purpose-built for advanced analytics. Flash memory has already taken hold within consumer mobile devices. Smartphones, iPods, iPads, and the like are all based on flash technology. Enterprise storage is just beginning to catch up, moving away from HDDs in favor of updated storage media such as flash—and companies looking to cash in on the big data analytics revolution will need to make the transition.

FlashBlade: The Modern Platform for Data Analytics

Storage has to be designed to be massively parallel and efficient to handle demanding new workloads. The FlashBlade™ system from Pure Storage was built to accelerate modern analytics applications. Powered by Purity software, it is a parallel platform capable of delivering high-performance access to an almost limitless number of objects and files for tens of thousands of clients concurrently.

Flash alleviates the problems associated with spinning disk storage. With disk-based storage, IT tends to overprovision the system to hedge against the inherent weaknesses of mechanical disks, says Scott Sinclair, senior analyst at Enterprise Strategy Group. "[Overprovisioning] causes problems—running out of space, consuming lots of power and then needing more power to cool all the heat. And there are many more components that can fail," he says. All of these challenges are greatly alleviated by flash. "With flash-based systems such as FlashBlade, you can dramatically reduce the cost of managing and maintaining your infrastructure" while maintaining the flexibility that is essential with ever-changing workloads.

With flash-based storage systems, you can dramatically reduce the cost of managing and maintaining your infrastructure.

When it's no longer ignored, enterprise storage can itself be a tool for transformation in the age of AI. Data is the foundation of insight—and how the storage system handles the data is of critical importance for analytics. "This will become a competitive differentiator, how efficient your IT infrastructure is, your ability to develop and deliver digital products and better understand your customers," Sinclair says. "The benefits of flash are so transformative that if you're not using it, you are behind the curve."

There are other solutions that may work with AI and ML, but they are built on legacy storage technologies that will slow down AI development time significantly. Only FlashBlade was designed from the ground up to meet the demands of advanced data analytics.

Powering Data Intelligence

Researchers at the RISELab at the University of California-Berkeley, successor to the AMPLab, which invented the popular Apache® Spark™ technology, employ the latest tools to accelerate DNA sequencing in pursuit of precision-medicine treatments to help patients. They soon discovered that exponential increases in the volume of genomic data demanded new approaches to mass storage. Adding

FlashBlade has reduced by almost two-thirds the time needed to sequence and analyze data-intensive DNA samples. “FlashBlade allows us to maintain high throughput while horizontally scaling bandwidth,” says Frank Austin Nothaft, a graduate student at RISELab.

Another FlashBlade customer also experienced significant performance improvements with its upgraded infrastructure. At this web-based company, the data science team is experiencing data processing times reduced from days to hours, yielding a significant boost in productivity.

Data is critically important, helping enterprises deliver their next major innovation with faster time to market or to gain a competitive edge. Enterprises now need to consider the essential storage piece of the equation when thinking about AI in general and deep learning in particular. The Pure Storage elastic scale-out storage solution is the only technology that is specifically designed for AI, ML, and DL, helping to accelerate business outcomes.

Visit **purestorage.com/analytics** for more information.

About Pure Storage

Pure Storage (NYSE:PSTG) helps companies push the boundaries of what’s possible. Pure’s end-to-end data platform - including FlashArray, FlashBlade and our converged offering with Cisco, FlashStack – is powered by innovative software that’s cloud-connected for management from anywhere on a mobile device and supported by the Evergreen business model. The company’s all-flash based technology, combined with its customer-friendly business model, drives business and IT transformation with solutions that are effortless, efficient and evergreen. With Pure’s industry-leading Satmetrix-certified NPS score of 83.7, Pure customers are some of the happiest in the world, and include organizations of all sizes, across an ever expanding range of industries.

About MIT Technology Review Custom

Built on more than 115 years of excellence in technology journalism, MIT Technology Review Custom is the arm of global media company MIT Technology Review that creates and distributes custom content. Our turnkey solutions include everything from writing, editing, and design expertise to multiple options for promotional support. Working closely with clients, our expert custom-editorial staff develops a range of high-quality, relevant content, delivering it to users when and where they want it—in digital, print, online, and in-person experiences. Everything is customized to fit clients’ content marketing goals and position them as thought leaders aligned with the authority on technology that matters.

www.technologyreview.com/media