PURESTORAGE®

AI as a Service, Simplified

How modern infrastructure clears the path for artificial intelligence.

Business success is all about the data. Where once cash was king, business leaders know that their futures hinge on the intelligent use of data. The ability to harness data helps businesses anticipate customer needs and buying trends, customize products and services, launch new business models, and understand market changes as they occur. In short, the intelligent use of data can make the difference in an organization's ability to compete and thrive.

A study by MIT and Pure Storage® shows that 78% of business leaders agree it is a challenge to extract useful insight from data.[1] And for good reason. Datasets are rapidly expanding in volume, velocity, and variety. Data is also flowing in from new, non-traditional sources that increasingly live outside of the data center. This includes social media and the Internet of Things (IoT), which are flooding the enterprise. Thus, much of a company's data remains cold and untapped in warehouses, appliances, and silos. Compounding the challenge, IT resources and talent are often sparse, making it difficult to take full advantage of expanding data and all that it has to offer.

WHY CONSIDER AI?

The challenges of managing and utilizing data are fueling interest in artificial intelligence (AI). In the MIT survey, 83% of respondents agree that AI will change how we think about and process data.[1] Meanwhile, 78% report they've been asked to explore new technologies like AI.

Organizations are investing in AI because they recognize that it will disrupt businesses across every industry. According to Gigabit, as of June 2018, a third of business leaders had committed to investing between $500,000 and $5 million in AI over the next 12 months.[2]

According to McKinsey, 60% of all occupations have at least 30% automatable activities, which signals enormous potential for AI.[3] AI promises increased efficiency and process enhancement. It has the potential to help workers make better decisions, be more creative, and focus strategically on their customers.

According to Gartner, "AI promises to be the most disruptive class of technologies during the next 10 years due to advances in computational power, volume, velocity and variety of data, as well as advances in deep neural networks."[4]

MAKE DISRUPTION WORK FOR YOU

AI is already disrupting multiple sectors by making data more actionable. You can see one example in your living room: Netflix has permanently altered the media landscape with its robust AI capabilities. This includes using AI to accurately recommend what to watch, personalize thumbnails so you're more likely to click, improve streaming quality, scout locations for movie production, and predict when projects might require manual editing.[5] Likewise, food suppliers are using AI to more quickly determine the origin of contamination in supply chains, helping to make recalls small-scale and targeted, instead of large-scale and global.[6]

FIGURE 1: Industries currently benefitting from AI insights.

Analysts also project that AI will cause wide-scale disruption in the development of autonomous and connected vehicles. According to KPMG analysts, there will be more than 150 million connected vehicles on the road by 2020, generating 11 petabytes of data annually.[7] The power of that data, along with the AI capabilities required to operate driverless vehicles, is poised to disrupt the automotive, insurance, wayfinding, rental car, parking, and long-haul trucking industries. Urban infrastructure development and mass transit systems will be affected as well. Over time, experts project that the transportation sector will transition from an individual-car-ownership model to a "mobility-as-a-service" model in which car-ownership rates could plummet.[8]

The central element of these examples is the same: data and how to properly handle it.

THE CHALLENGES IN POWERING INSIGHT

A full 86% of senior executives surveyed in the MIT Technology Review study cite data as the foundation of business decision making; 87% consider it important to business growth.[1] However, given the high volume, quality, and speed of data that organizations collect today, enterprise data warehousing and traditional analytics do not entail sufficient storage or compute power to generate actionable insights at scale or in real time. This reality is borne out by 43% of companies identifying their infrastructure as an obstacle for AI-driven advanced analytics.[1]

Most businesses recognize that AI is a game-changer. Data shows that companies are increasingly interested in employing the strength of AI to advance their positions in the competitive marketplace. It is therefore advantageous to support organizations that want to take advantage of AI as part of a future growth strategy.

However, as a managed service provider (MSP), substantial challenges remain as you try to support those organizations in their journeys to AI readiness:

- **Data:** Unstructured data that comes in from sensors, machines, and user input ranges from small, random objects to large, streaming files. Cleaning and preparing this data to make it usable for AI will tax the storage and computational elements of the IT infrastructure.[9]
- **Infrastructure:** Legacy infrastructure is inadequate to support the modern compute workloads and low-latency transactions required by AI. Most older infrastructure was built to store data rather than query and process it at the speeds needed to deliver modern analytics or AI-derived insights.
- **Future-proof:** Managing the infrastructure complexities and evolution of AI open-source frameworks is a perennial challenge. You need to provide a scalable, yet flexible, infrastructure to support ever-evolving technologies.
- **Expertise:** A sparse talent pool makes it difficult to differentiate between diverse AI services and integrate those services into various analytics platforms, in addition to providing enterprise-grade support.

Established in 2018, Core Scientific builds data centers optimized for demanding, enterprise-level AI workloads. It also offers infrastructure-as-a-service (IaaS) solutions. Using these technologies—and Pure Storage technology —Core Scientific was able to offer customers AI performance improvements of up to 800% compared to their previous Amazon Web Services® (AWS®) cloud computing solutions.

Jam City, a top developer of mobile game applications, depends on AI to obtain revenue-increasing insights from data streamed from millions of users worldwide. Once it engaged Core Scientific to enable enhanced AI capabilities, Jam City experienced an eight-fold increase in the volume of data it could process, compared to AWS. In order to provide enterprise-grade AI IaaS, Core Scientific powers its solution with AIRI™ (AI-ready infrastructure), jointly developed by Pure Storage and NVIDIA.

HOW LEADING MSPS ARE MAKING USE OF AI

Given the imperative to incorporate AI into business practices and decision making, forward-thinking organizations could turn to expert MSPs to fast-track their AI strategies.

MSPs can help customers accelerate digital transformation by partnering with Pure Storage. Pure Storage solutions are part of a data-centric architecture engineered for modern analytics and AI to:

- Accelerate innovation and insights
- Reduce complexity
- Increase agility and value

Companies are looking to capture the value inherent in uncovering patterns in data. While automating business processes and eliminating inefficiencies can provide savings, there could be even greater growth potential in the ability to expose new revenue streams, personalize customer experience, and free workers to focus on strategic priorities.

CRITICAL COMPONENTS OF AI-READY INFRASTRUCTURE

The specialized infrastructure necessary to power AI includes multiple elements. One of the most important components is the graphics processing unit (GPU). GPUs have traditionally delivered the data processing and fast action expected by online gamers. But among the innovations fueling the AI revolution is the discovery that GPUs are also exceptionally good at executing algorithms necessary for AI. Because they use parallel processing, GPUs execute AI's data and algorithm computing many times faster than traditional processors.

It's critical to use storage that harnesses the full compute power of GPUs to create an infrastructure optimized for AI.

"The challenge is often perceived as not having enough GPUs; but in reality, most infrastructures do not fully utilize the compute power of the GPUs they have," notes Jim Benedetto, chief data officer at Core Scientific. "A fundamental aspect of maximizing GPU capacity is getting information streamed into the GPU as fast as you can. So a key building block of a GPU-optimized infrastructure is having the fastest storage possible."[10]

OPPORTUNITY: DELIVER OPTIMIZED STORAGE AND COMPUTE POWER

When you offer your customers storage and compute power optimized to drive actionable AI, you can deliver enterprise-ready solutions that provide data governance and orchestration for many complex data-processing operations. With these solutions, your customers can uncover hidden patterns with increased model intelligence, helping to improve customer satisfaction and retention.

This approach also helps you easily meet requirements for diverse AI use cases, ranging from fraud detection to image recognition. As AI initiatives grow, you can keep up with demand, accelerating data-intensive applications and enabling data engineers to stay productive and innovate over time. Additionally, you can support the needs of developers and data scientists for increased adoption and potentially enable them to uncover opportunities, such as expanding into new markets or developing new lines of revenue growth.

Pure Storage technologies can deliver predictable, ultra-fast performance and response times for any AI workloads—reads or writes, large or small files, and sequential or random data ingestion. The optimized performance is delivered by Pure's AIRI solution, which is designed to maximize GPU utilization. The AIRI solution's integrated FlashBlade™ product delivers flexibility to support disaggregated scaling for future workloads. AI is powered by massively parallel technologies, like deep learning and GPUs, and yet legacy storage systems are full of serial bottlenecks. That's why the FlashBlade product was architected to be massively parallel for AI.

AIRI is the industry's first complete AI-ready infrastructure, architected by Pure Storage and NVIDIA to extend the power of NVIDIA® DGX™ systems. Powered by FlashBlade storage and NVIDIA DGX-1 and DGX-2 servers, AIRI offers a simple, fast, and future-proof infrastructure capable of growing from AIRI "mini" to hyperscale—and meeting AI demands at any scale without downtime.

Companies value a seamless plug-and-play experience that delivers real-time AI insights at scale without having to manage the complexity of retrofitting legacy technology. Being able to rely on easy-to-use infrastructure that is expressly designed for AI simplifies maintenance and operations, enables customers to efficiently engage in AI initiatives, and allows data scientists to focus on strategic priorities for their organizations.

Most legacy IT infrastructures use multiple storage silos for databases, data ingestion, and data analytics, often making them complex to architect and difficult to maintain and upgrade. Additionally, most common storage systems in use today are optimized for long-term archival use, not for fast retrieval and high performance.

In order to ensure maximum throughput, it is critical to locate data near its processors and to spread it across multiple storage devices. "Locating data near the machine learning cluster that will consume it is a necessity," according to Moor Insights & Strategy. Solving for these challenges is critical for creating streamlined, simplified infrastructure solutions optimized to drive AI.

OPPORTUNITY: PROVIDE AI-READY INFRASTRUCTURE

An AI-ready infrastructure gives you the capacity to migrate your customers' data from silos to a data-centric architecture. Consolidation enables real-time access and empowers end-to-end data pipelines for streaming, batch processing, and real-time analytics. Reducing the time it takes to build modern analytics platforms and environments enables customers to jump-start their AI initiatives in hours rather than weeks or months. Thus, the real-time data analytics and accelerated delivery you offer become competitive differentiators that can drive customer adoption and retention.

The Pure Storage data hub helps deliver these operational benefits by providing a data-centric architecture optimized to power analytics and AI. It simplifies and accelerates modern analytics with an elastic, scale-out architecture. Additionally, the Purity operating environment enables consolidation of workloads across a customer's public, private, or on-premises clouds. Pure solutions enable a lower total cost of ownership (TCO) with investment protection and greater flexibility for distributed services. Pure's next-generation data hub, backed by the Evergreen Storage™ subscription model, disaggregates infrastructure and future-proofs AI.

The speed of digital innovation is breathtaking. As new technology comes online, customers want to know that their infrastructure investment is secure and that they can add new capabilities or technologies without downtime or service disruptions. They expect their AI investments to continue yielding results and delivering value.

Creating an optimal AI computing environment relies on critical integration of hardware and software solutions to drive maximum performance and throughput. The use of GPUs is critical in maintaining the processing power needed for AI applications. But GPUs have traditionally been hard to program. Fortunately, NVIDIA has developed a programming language and a wide range of tools to help develop the algorithms that form the foundation of AI and the insights it drives. As a result, the GPU software ecosystem has expanded, allowing a wide variety of new algorithms to power AI. Pairing available AI-ready storage and compute power with programming and algorithms designed to drive AI is the most efficient way to build AI-enabled IT infrastructure.

FIGURE 2: AI-ready infrastructure (AIRI) uses the NVIDIA DGX-1 integrated with the Pure Storage FlashBlade product as a modular building block for scaling AI capacity.

OPPORTUNITY: DELIVER AI AS A SERVICE

When you offer a solution that integrates storage and compute power along with integrated software designed for AI algorithms, you can confidently deliver AI as a service, including pre-validated and scalable frameworks. This level of efficiency enables you to streamline operations, reduce customer-response times, decrease reliance on hard-to-acquire talent for deployment and delivery, and meet service-level agreements (SLAs) with ease.

Pure Storage helps you readily monitor, manage, and plan your AI offerings. The Pure1® platform's proactive support with the NVIDIA GPU Cloud® (NGC®) stack and the AIRI Scaling Toolkit enables you to deliver AI as a service. This allows you to unlock unprecedented productivity for your customers, in addition to streamlining operations. Pure's high availability and predictive capacity planning help eliminate service outages, which simplifies planning and monitoring. You can offer your customers seamless data scaling in addition to upgrades and migrations with zero downtime. As an added benefit, Purity software and Pure Service Orchestrator™ container storage simplify exposing data and AI tasks. This dramatically reduces data bottlenecks and hotspots, while increasing application and microservice integration. As a result, your customers can significantly increase collaboration and reduce costs.

PURE1 PROACTIVE SUPPORT

NVIDIA GPU CLOUD STACK

AIRI SCALING TOOLKIT

WHY PARTNER WITH PURE STORAGE FOR YOUR AI OFFERINGS

AI presents proven potential to unlock market share, reveal new revenue streams, and streamline operations. With early adopters already disrupting markets and entire sectors of the economy, smart businesses will look to their MSPs to accelerate their digital transformations by adding AI services.

There are multiple obstacles to developing the specialized infrastructure to support today's AI-ready compute needs while anticipating future expanded needs. Additionally, it is still a challenge to find qualified AI technology specialists. By partnering with Pure to support your AI offerings, you can overcome these obstacles and offer your customers proven technology that is future-proofed, easy to use, and supported by Pure's qualified experts.

Pure Storage offers build-it-yourself and purpose-built deployment options using the AIRI solution from Pure Storage and NVIDIA. The solution delivers the equivalent of 50 racks of compute and storage capacity at 95% efficiency plus 10 times the space savings over traditional data storage. The Pure Storage data hub simplifies and optimizes AI with a single underlying platform that flattens data, enabling various silos to share and collaborate. The Pure architecture is backed by the modular Evergreen Storage subscription model that scales to a rack and beyond, so you know your investment is future-proofed. Additionally, Pure1 cloud-based management enables your customers' data-science teams to focus on creating algorithms to support the business rather than handling everyday infrastructure operations.

Accelerate your digital transformation with a data-centric architecture engineered for modern analytics and AI. Learn more at **www.purestorage.com/ai**.

Contact your partner representative to discuss a proof-of-concept trial.

1 "Unlock the Intelligence within your Data with AI," MIT Technology Review in partnership with Pure Storage, 2018.

2 "SAP: One third of business leaders to invest over $500,000 in AI in 12 months," Gigabit, June 2018.

3 "What's now and next in analytics, AI, and automation," McKinsey Global Institute, May 2017.

4 "Gartner Says Global Artificial Intelligence Business Value to Reach $1.2 Trillion in 2018," Gartner, April 2018.

5 "How Netflix Uses AI, Data Science, and Machine Learning — From A Product Perspective," Allen Yu, Medium, February 2019.

6 "'Smart' or Not, A.I. Is Playing a Crucial Role in Ensuring Food Safety," SupplyChainBrain, November 2018.

7 "The chaotic middle," KPMG, June 2017.

8 "The rise of mobility as a service," Deloitte Review, 2017.

9 "Storage-Optimized Machine Learning," Moor Insights & Strategy, 2018.

10 "Pure Storage Case Study | Core Scientific," Pure Storage, 2019.