Core Scientific provides an advanced AI and blockchain infrastructure-as-a-service using AIRI™ (AI-Ready Infrastructure), jointly developed by Pure Storage® and NVIDIA®. The company's customers are realizing quantum increases in actionable insights that were previously unattainable for a wide range of large-scale data analytics projects and initiatives.

BUSINESS TRANSFORMATION

Companies looking to quickly and cost-effectively tap the power of artificial intelligence gain unprecedented levels of tangible insights, identifying new opportunities for growth and changing how they run their businesses.

REGION / COUNTRY

NAM / United States

INDUSTRY

IT Services

"The AIRI infrastructure allows people to turn data into actionable information in ways not previously possible — and that fundamentally changes your business decision-making."

Jim Benedetto, *Chief Data Officer*

CORE SCIENTIFIC ADVANCES AI-ENABLED APPLICATIONS WITH AIRI FROM PURE STORAGE AND NVIDIA

During more than two decades immersed in technology start-ups, Jim Benedetto has learned to see both world-changing opportunities and the challenges that must be overcome to realize them. With experience in starting and growing companies in fields such as social media and content optimization, Benedetto brings over a decade of experience in artificial intelligence and machine learning to Core Scientific as their Chief Data Officer. The company was started in 2018 to provide state-of-the-art infrastructure and software solutions for AI and blockchain applications.

What attracted Benedetto to Core Scientific is what he described as a "fundamental paradigm shift" in how data is now being processed. "For the longest time, the model has been 'collect-store-analyze'. Now, we're starting to 'collect-analyze-store' which really is how the human brain works. We analyze what we see and hear, then decide what we want to retain."

Driving this change, Benedetto added, is the graphics-processing unit (GPU), "which doesn't just enable you to query or analyze data five times faster; it enables you to query data on a scale you never were able to query before." The leading examples of GPUs are produced by NVIDIA, which packages them into its DGX-1® and DGX-2® systems.

As promising as the GPU is, Benedetto sees a major roadblock to achieving the full potential of GPU-accelerated applications for AI and blockchain initiatives. "GPUs are going into data centers that were built for machines based on the CPUs that power PCs and servers. But GPUs are fundamentally different from CPUs and require a very different supporting infrastructure." Consequently, the mission of Core Scientific is to build data centers that are optimized for GPUs and then offer their advanced software platform and infrastructure-as-a-service to enterprises that want to run demanding AI workloads. This approach is specifically designed for forward-looking organizations that want to avoid the significant capital and operating investments necessary to build and operate the required infrastructure on their own.

SOLVING THE STORAGE BOTTLENECK

As Benedetto dug into the details of creating a GPU-optimized infrastructure, he discovered yet another barrier to realizing the full potential of GPUs. "We were finding it difficult to get a DGX-1 above 10% or 15% of its total compute capacity. What we found was that the internal storage built into the DGX-1 was not feeding it data fast enough to maximize its processing power."

COMPANY:

Core Scientific
www.corescientific.com

USE CASE:

- AI Infrastructure Hosting with AIRI
- Blockchain Infrastructure and Hosting with AIRI
- Big Data Analytics with OmniSci on AIRI

CHALLENGES:

- To deliver on its vision of providing an AI-as-a-service infrastructure, Core Scientific needed a linearly scalable storage solution to ensure its GPU investment was fully utilized.

IT TRANSFORMATION:

- Users of Core Scientific's infrastructure see performance improvements of up to 800% on GPU-based applications compared to AWS.
- Data scientists are now able to quickly analyze massive amounts of data on a scale that was not possible with legacy IT systems.

"The almost infinite scalability of FlashBlade is an integral part of our IT architecture and strategy."

Jim Benedetto, ***Chief Data Officer***

Benedetto observed that as users of GPUs saw sub-par performance on their applications, they tended to add more GPUs. "That successfully masks, but doesn't actually solve the real problem," he continued. "The challenge is often perceived as not having enough GPUs; but in reality, most infrastructures do not fully utilize the compute of the GPUs they have. 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."

This realization led Benedetto to conduct an intensive investigation into the storage marketplace, which led him to Pure Storage.

"I have been passionate about storage for more than 20 years and have bought and managed systems from almost every vendor there is. Pure Storage is the first storage vendor I have seen in the last 15 years to come out of nowhere and take on all the major players. Today, Pure is the leader in the all-flash market."

Benedetto said he recalled his first exposure to Pure Storage. "I was being shown slides of their design, and I recognized that it was fundamentally different from any other storage I had ever seen. The key feature of the Pure architecture is horizontal scalability. Everyone claims linear scalability, but FlashBlade™ has true linear scalability."

AIRI: OPTIMIZED FOR ARTIFICIAL INTELLIGENCE APPLICATIONS

Benedetto's evaluation of storage options coincided with the introduction of AIRI (AI-Ready Infrastructure), jointly developed by Pure Storage and NVIDIA. AIRI is a fully integrated software and hardware solution that enables data scientists to jumpstart their AI initiatives in hours, not weeks or months. AIRI is powered by Pure Storage FlashBlade, the industry's first storage platform architected for modern analytics and AI, and four NVIDIA DGX-1 supercomputers, delivering four petaFLOPS of performance with NVIDIA Tesla® V100 GPUs. These systems are interconnected with Arista 100GbE switches, supporting GPUDirect RDMA for maximum distributed training performance.

Benedetto recognized AIRI as the solution to his need for storage optimized for GPUs and AI applications. "With standard DGX-1s, you get around 6 to 8 Gbps throughput, but with AIRI we're talking 75 Gbs for reads. That's what I mean by storage fast enough to optimize the computing power of a GPU."

AIRI systems are the heart of the data centers that Core Scientific is building. "We already have a half-dozen data centers in operation, mostly in Rust Belt locations where there used to be industrial plants that required a lot of electricity. And the availability of power is one of our most important criteria for site selection. Each of our data centers draws anywhere from 30 to 100 MWs, compared to a typical CPU-based data center that pulls maybe 7 to 10 MW."

Core Scientific has contracts in place for more than 650 MW, which Benedetto estimates will support data centers containing as many as 700 DGX-1s. Each data center is cooled by a patented design that uniquely uses computational fluid dynamics to move ambient air — "no raised floors, no HVAC" — which plays to the advantages of the compact form factor of FlashBlade.

CUSTOMERS SEE UNPRECEDENTED PERFORMANCE FROM AIRI-POWERED APPLICATIONS

Customers already are benefitting from the performance delivered by the AIRI-powered infrastructure of Core Scientific. One is Jam City, a leading developer of mobile gaming applications. "They heavily leverage OmniSci to analyze data streamed from millions of

users worldwide to build machine-learning models that will create revenue-increasing insights. The problem they always had was that they could never analyze all the data they wanted to," Benedetto said.

"They were using a dozen GPUs on Amazon Web Services, but each of those GPUs was only utilized at about 5% of capacity. That limited the amount of data they could analyze at around 500 million data points. But they collect billions of data points every day. They migrated their application to Core Scientific infrastructure powered by AIRI — two DGX-1s and FlashBlade — and are now processing 4 billion data points. That's an 8x increase in the volume of data that is processed when compared with AWS."

"Having storage as one of the least complex parts of our infrastructure is something very valuable to me."

Jim Benedetto, ***Chief Data Officer***

The impact of such an improvement goes beyond performance numbers, he observed. "What they found was that it's not just the increase in speed. It's the fact they can do things in terms of data analytics that they weren't able to do before. They can now run queries on *billions* of rows of data. Not only that, line of business personnel throughout the company — not just the data scientists — can learn from the data."

Another Core Scientific customer also has put its demanding OmniSci workloads on the AIRI-powered infrastructure. "Their workload is a real-time SQL-based GPU-accelerated database with very impressive visualization built on top. Their goal was to turn analytics upside-down. Instead of going to executives and asking what they wanted to see in their reports, they wanted to load as much data as possible into OmniSci and give people the ability to ask questions they've never been able to ask before.

"Instead of giving management a handful of metrics, every KPI (key performance indicator) in the company is available. And the data scientists are showing people throughout the company what they can do with the data and how they can benefit from it in their roles. It's very powerful, and changes how the business operates."

The impact of running optimized AI applications may not always be evident, Benedetto observed. "Everybody knows data is important, but companies can actually change how they do business when they have access to more data. The AIRI infrastructure allows people to turn data into actionable information in ways not previously possible — and that fundamentally changes your business decision-making."

A POWERFUL SOLUTION YET EASY TO USE

While positive impact on customers is the most important goal, Benedetto also appreciates the ease of use for Pure Storage products. "Oftentimes storage can become one of the most complex parts of the infrastructure," he noted, "but with Pure we don't need to have anyone on our team with the specialized knowledge that storage systems have traditionally required. Having storage as one of the least complex parts of our infrastructure — the it-just-works part — is something very valuable to me."

Another valuable aspect of AIRI is the fact that it is a reference architecture designed, tested and supported by Pure Storage, NVIDIA and Arista. "Reference architectures eliminate finger-pointing, ensure everyone operate as one team, and give customers the confidence that they're making the right decision."

It's in Benedetto's nature to look ahead, and he sees significant changes down the road from AI applications. "With the explosion of data analytics empowered by technologies like GPUs, 5G, high-performance computing, analytics, and AI, companies will have the potential to make fundamental improvements in the way they do business. But that potential will only be realized if they can tap an infrastructure optimized to take advantage of those technologies. Our goal at Core Scientific is to make that infrastructure and the supporting software platform readily available."

He added, “As our customer base grows, we need the ability to scale our operations easily and cost-effectively. The almost infinite scalability of FlashBlade is an integral part of our IT architecture and strategy.”

Core Scientific’s purpose-built data centers are very complex. “Each component has to be scalable,” Benedetto pointed out. “Horizontally scalable networks have been around for a long time. And GPUs are horizontally scalable as well. But a single namespace, horizontally scalable storage platform is what Pure Storage brings. Having horizontally scalable storage from FlashBlade allows all the other elements of our infrastructure to do their work at maximum efficiency.”

“FlashBlade has true linear scalability.”

Jim Benedetto, ***Chief Data Officer***

With his company making major investments in infrastructure, Benedetto measures the potential return on that investment in a special way. “When you look at an instance where you see an 800% improvement for the same or less cost, that’s a significant ROI. But the real impact goes beyond just numbers. It’s the value a company gets from being able to see every KPI every single day; and the impact on the business when data scientists can ask questions they’ve never been able to ask before, and subsequently apply the answers to make more informed business decisions and increase revenues. And we’re only just beginning to see these benefits.”

PURESTORAGE®

info@purestorage.com

www.purestorage.com/customers