# IT Architectures for AI: From POC to Production

Artificial Intelligence is everywhere, from our appliances to our vehicles to all our smart devices, each one of these connected emitters sending out seemingly endless waves of data, every second of every day. We've clearly entered the data era and organizations across the globe are extracting the value of this data to solve many of humanity's greatest challenges by harnessing the power of AI. The increasing value of data is why we talk about adopting a "Data First" strategy. If you plan for it appropriately and build an integrated, optimized IT infrastructure for its growth, your chances of AI success increase dramatically. You'll be able to properly manage, categorize and tag data in place as your AI initiatives mature.

Enterprise adoption of AI, including Machine Learning (ML) and Deep Learning (DL), continues marching forward at a rapid pace thanks to ever increasing amounts of data along with advances in Data Science tools, GPU-accelerated compute, scalable storage and high-speed networking technologies. And organizations are deriving key learnings and best practices along the way. One such key learning is that the successful use of AI technologies is highly dependent on the underlying IT architecture that supports Data Science processes. It's no surprise to Data Architects and allied IT practitioners that these iterative processes of building, training, testing and deploying successful AI models efficiently rely heavily on scalable, high performing compute and storage technologies.

## AI/ML/DL impacts data value

IT infrastructure plays a critical role
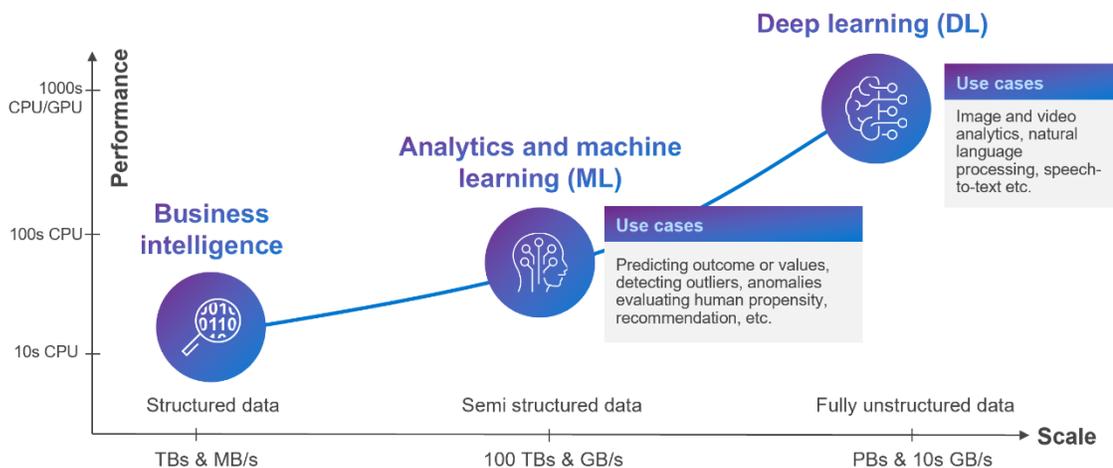


## AI Center of Excellence

An AI Center of Excellence that includes business leaders, Data Scientists, Data Architects and IT practitioners is a great way to get started with AI initiatives. Together the team creates a plan outlining the business problem to be solved, the technology investments required and the bottom-line results expected. And once a use case is decided, the success of the project, from concept to deployment, depends on an integrated and optimized IT infrastructure, one that is reusable, scalable and flexible enough to handle the rapid iteration required to properly train models and drive neural networks. With the business problem mapped to a data science problem, the next step is closing the last mile gap – standing up the IT infrastructure required. Studies have shown that many organization's AI initiatives fail not only due to lack of proper planning but also due to an inadequate IT infrastructure to move beyond a proof of concept (POC) in a cost effective and scalable way.

## Building an AI Proof of Concept

Building an effective and successful AI POC starts with the use of flexible, modular AI Architectures – the compute and storage foundational pieces of an integrated, optimized IT infrastructure. It's best to design for scale from the very beginning, as datasets required for efficient model training grow exponentially over time. Then by paying attention to the storage component at the outset, you'll be well positioned to transition into production once your POC is successful. This ensures you'll be ready with data stored in place as your initiatives mature. Model training is a time-consuming iterative

process. However, if an AI POC is built correctly, model training time can be reduced significantly, thereby increasing the time to value of your data science solution.

## AI Architecture Building Blocks

At Dell Technologies, we use AI in our datacenters, and we understand its complexities and nuances. Our AI Architecture building blocks allow you to holistically plan for and build an end to end, scalable data landscape spanning edge to core to cloud for your AI Initiatives. Use any of these building blocks to build a scalable, high performing and reusable IT infrastructure ready for the iterative model training needs of Data Science from POC through production.

For compute, we have Dell Precision Data Science Workstation, Dell EMC PowerEdge and NVIDIA DGX servers. Dell EMC PowerScale scale-out storage and Dell EMC ECS object storage cover you for on-premises AI storage, while PowerScale for Multi-cloud and PowerScale for Google Cloud help you expand your AI initiatives to the cloud.



# AI Architecture building blocks: POC to Production
From edge to core to cloud
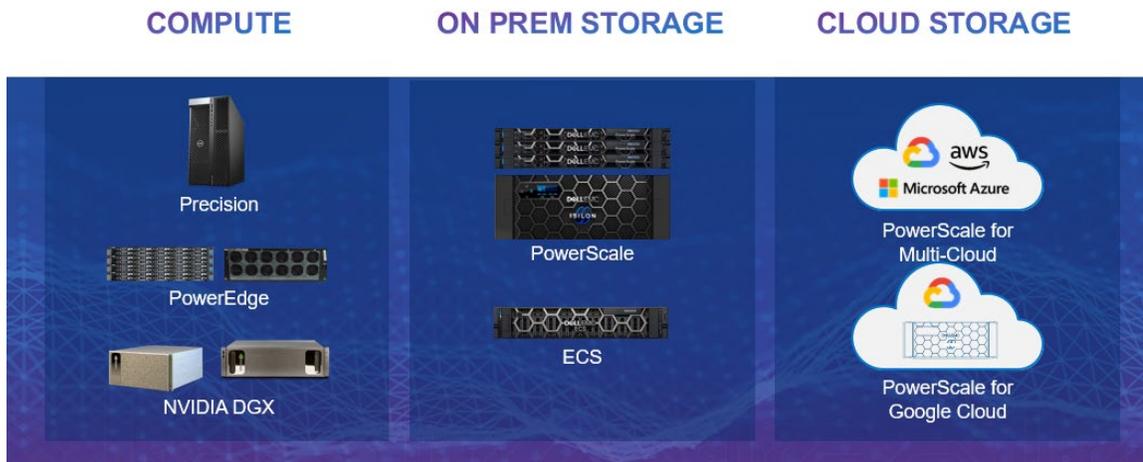
Combine the industry's broadest portfolio of IT Infrastructure products, solutions and services with common AI/ML/DL tools, and you'll have everything you need to get started with your AI initiatives. We support all the common tools and frameworks, and we have reference architectures and validated compute and storage solutions ranging from Splunk analytics to Caffe2 and TensorFlow for DL.



# AI ecosystem aligned with Dell EMC infrastructure
Flexible compute, scalable storage
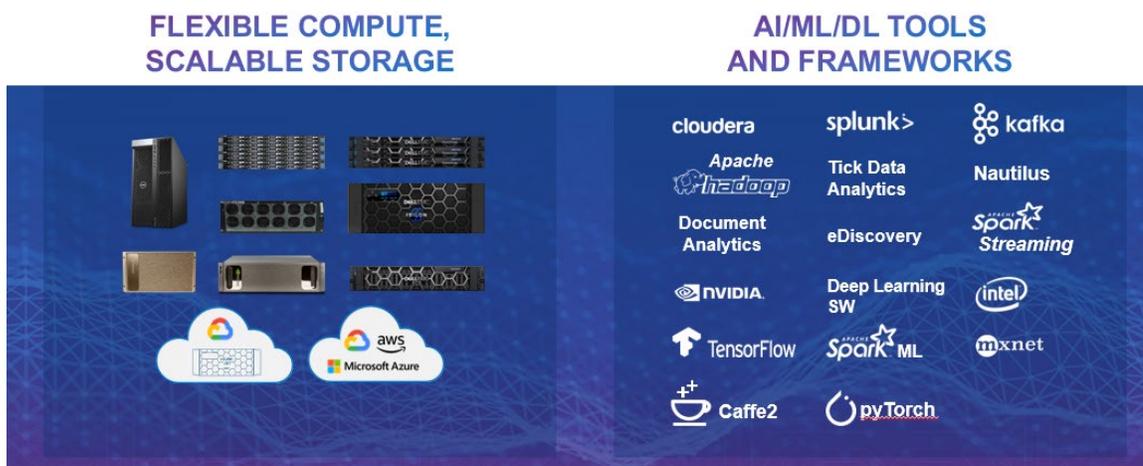
## Scaling to Production

As data science initiatives mature in POC, the true test is now transitioning to production, where you begin to monetize your data. As stated earlier, many organizations struggle with this phase. However, this can be avoided from a storage perspective by keeping your growing data sets in place, on shared storage from the start of POC all the way through to production. With this approach, the time to value of data science can be increased significantly as you won't need costly and time-consuming data migration when moving from POC to production, or when you decide to augment your data science with cloud-based services.

To make it easier for our customers, we have AI Ready Solutions and Reference Architectures built with PowerScale, the industry's leading NAS platform. One example is an AI Starter bundle. Using Precision Data Science Workstations with data stored on PowerScale for a sandbox, Data Scientists can efficiently train Deep Learning models from day one. PowerScale's high performance speeds up the iteration times, allowing for more effective Neural Networks training as the data pipeline grows. And PowerScale's high performance is matched with its scale-out architecture, where you simply drop in additional PowerScale nodes as your storage needs increase while data gets automatically rebalanced when the size of the storage cluster expands.

As you move from POC to production, your data on PowerScale remains in place; you add in PowerEdge servers and PowerSwitch networking to build-out your Dell EMC AI Ready solution for production deployment. Built with PowerScale storage, PowerEdge servers, and PowerSwitch, our AI-ready solutions provide a holistic approach precisely for your production-ready AI initiatives. And with data consolidated on PowerScale, managed by PowerScale OneFS, you have a single file system environment for simple scalability, without costly, time-consuming data migrations. Our customers have been looking for end-to-end solutions, and this is how they come together. Whether in POC or full production, the PowerScale family provides the IT architecture building blocks through to complete solutions for AI initiatives.

## PowerScale Family for AI

PowerScale is made for storing the large datasets needed and created by AI/ML/DL technologies. Organizations of all types and sizes are adopting it for their burgeoning AI initiatives. They're quickly realizing its value and benefits for both POC and production environments. PowerScale eliminates AI I/O bottlenecks with enterprise grade features, high performance, concurrency and scalability delivering faster training and validation of AI models, higher model accuracy, improvements in data science productivity, and maximization of ROI for compute investments. All of this is due to the power, flexibility, scalability and enterprise grade features that are standard on the PowerScale platform.

PowerScale OneFS, the operating system powering the industry's leading scale-out NAS platform PowerScale, offers out of the box enterprise data management and governance capabilities and eliminates data silos. This consolidation allows you to centralize enterprise storage processes including data management, performance management, data protection and data security. This consolidation also increases the economics of data storage, lowers ownership costs, and reduces risk. And with Dell EMC DataIQ, a powerful data management tool that is included with Dell EMC PowerScale, it's easy for anyone to find and understand data across all your file and object data platforms, including non-Dell EMC Storage platforms, spanning on-premises and off-premises deployment.

Simply put, PowerScale delivers the performance and concurrency required for AI/ML/DL workloads. And in the case of GPU accelerated compute, it keeps this key



PowerScale family for AI

| ALL-FLASH | HYBRID | CLOUD |

Isilon F800 | F810

PowerScale F600
PowerEdge based All-NVMe

PowerScale F200
PowerEdge based All-Flash

H400 | H500 | H5600 | H600

**ARCHIVE**

Isilon
A200 | A2000

MULTI-CLOUD
Azure, AWS, Google

NATIVE CLOUD
PowerScale OneFS for Google Cloud

**← PowerScale OneFS →**

PowerScale can join existing Isilon nodes in the same OneFS 9.0 cluster

investment in your compute resources loaded up and busy with the data needed for rapid, effective AI model training and accuracy. The data pipeline stays full and your GPUs stay busy.

## The Future of Data Science is Multi-Cloud

As Data Science matures and evolves, it's becoming clear that lines are blurring with choices for architecting the supporting IT systems. Do you go all in with cloud? Or do you keep data on premises? When it comes to cloud storage, we don't believe it's an either-or proposition. We see cloud not as a destination but rather as an operating model. In many cases, the answers are blended depending on budgets, geography, team size and other contributing factors. At Dell Technologies, we recognize these factors and look to help our customers navigate a world without boundaries to take advantage of the realities in a hybrid, multi-cloud world.

Our Dell Technologies Cloud Storage offerings are designed for flexibility and customer choice. For instance, geographic diversity is important for disaster recovery and business continuity. Or perhaps the business needs to lower costs of on-premise IT investments. One key dependency in AI initiatives, as mentioned previously, is your data. And if your IT storage infrastructure is architected at the outset with a data first mentality, keeping data in place from POC to production, then the value propositions of cloud economics become clearer, especially when considering data egress fees, migration times, and risks to your data.

One key benefit to architecting systems to keep AI data in place is the myriad data science services offered by hyperscalers. By storing data on PowerScale, Data Science practitioners can easily swivel between favorite services. For instance, one popular application is to process data that is generated on premises with industry-specific and/or 3rd party software tools, such as in genomic analysis. Others want to take advantage of specific software services provided by the hyperscalers, building AI/ML/DL algorithms with TensorFlow to help self-driving cars learn. Another example is customers who just need cloud compute for spot capacity or spot pricing, or possibly to take advantage of GPU-powered servers on an on-demand basis for neural network model training.

A key point to emphasize again is that AI initiatives require proper planning and forethought with respect to your data, where it's generated and stored. If your AI data is currently generated and/or stored in one of the public clouds, and you want to use their native compute, AI and other data analytics services, then you may be set. However, if your data isn't generated in the cloud, you're likely better off storing it outside of a public cloud and bursting to cloud for GPU-as-a-service as needed or switch between your favorite compute, AI and Data Analytics services from AWS, Azure and Google Cloud.

At Dell Technologies we provide support for these various AI use cases in the cloud, but let's double-click on our multi-cloud offering. Today AWS, Azure and Google Cloud provide over 500 differentiated data science services, and our customers want to utilize them across the clouds, sometimes for the same workloads and data sets. But moving data between clouds comes with migration risks, huge egress fees, and the long data migration times. So, a big challenge that customers are facing today is how to access the best AI services, platforms and applications they need across various clouds, rather than being limited by a subset of services offered by just a single cloud vendor. Another challenge is the variability of available



## Dell Technologies cloud storage

The performance and scale of Dell EMC unstructured data storage with the economics and application services of the cloud architected with data science in mind

### DIRECTLY CONNECTED TO CLOUD

PowerScale for Multi-Cloud

### NATIVE CLOUD EXPERIENCE

PowerScale for Google Cloud

Data stays in place so that AI Initiatives can scale across disparate cloud-based analytics and AI compute services
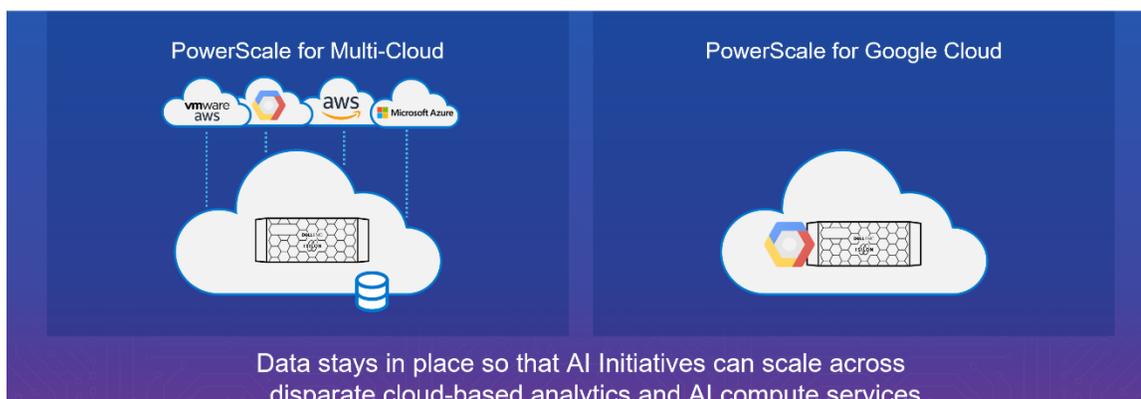
resources, like GPU-powered servers that are critical for AI workloads, not only between cloud providers but also in different regions even within a single provider.

As noted earlier, we offer cloud storage, built on PowerScale, for both the multi-cloud use case and native cloud experience. PowerScale for Multi-cloud is the cloud storage service offering detailed above. It's a managed cloud service offering that's directly connected to the major hyperscale cloud providers through a very fast, direct interconnect. It provides agile, multi-cloud support allowing you to leverage different clouds, easily and quickly switching between them based on applications' needs, to maximize business outcomes. You avoid vendor lock-in by keeping data independent of the cloud, so you don't have to worry about high egress charges, migration risk, or time required to move data.

If, however, you're already invested in Google Cloud then our native cloud offering with them might be for you. PowerScale is offered as a native Google Cloud service based on PowerScale technology and augmented by Dell services and enterprise SLAs. This PowerScale for Google Cloud offering is fully integrated into both Google billing and cloud portal systems. It provides the highest levels of storage to compute performance and tightest cloud integration for all your AI needs.

## About Dell Technologies

Dell Technologies is a trusted leader in infrastructure solutions for data analytics and AI. Dell EMC PowerEdge, PowerScale and PowerSwitch form the foundation of a simple, scalable and future-proof IT infrastructure, giving organizations the agility to transform business operations and the flexibility to adapt to new AI and analytics workflows. Through innovative products, solutions and services, Dell Technologies is helping organizations unlock the value of their data to drive the business. And we stop at nothing to help you achieve success in your use of AI on your Digital Transformation journey.

To learn more about our offerings visit our website
**DellTechnologies.com/StorageforAI**

**D⊗LL**Technologies