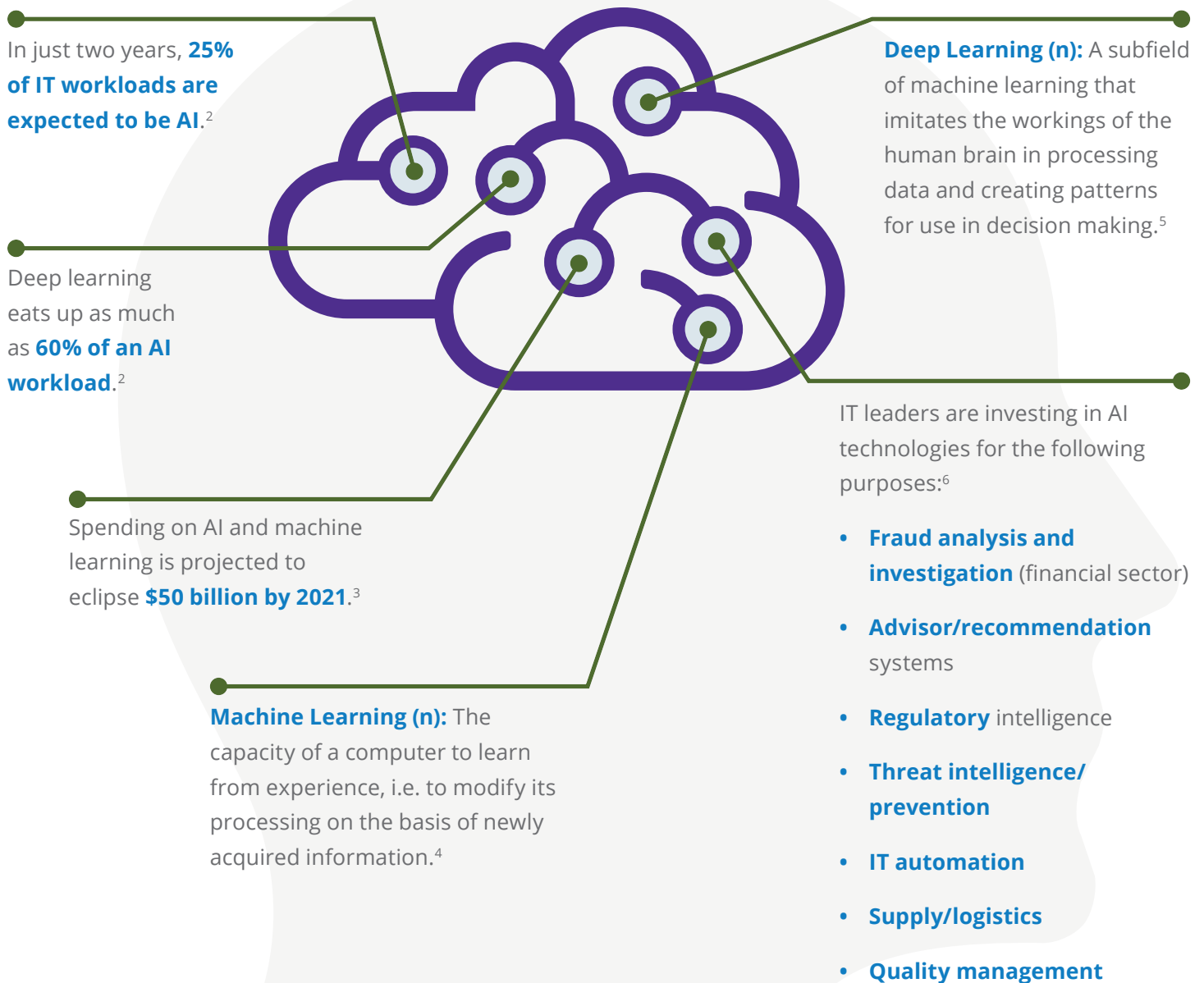


Why Artificial Intelligence Necessitates **Going Beyond Traditional Infrastructure**

Creating a Data Pipeline that Enables AI, Machine
Learning and Deep Learning

Artificial intelligence is here. **Is your infrastructure ready?**

Adopting artificial intelligence (AI) technologies is the top initiative for 61% of IT leaders.¹ Many of them are focused on investing in solutions like machine learning (ML) and deep learning (DL), two AI techniques that require massive stores of data and are extremely compute intensive. This rapid, impending adoption presents server and storage infrastructure with never-before-seen performance demands and leads many decision makers to ask, "Is today's infrastructure enough?"



Bottom line— IT leaders are ready to hit the go button on AI! Only 23% don't know what AI software or algorithms they need to meet their needs.²

Hitting the Infrastructure Wall with AI

For many organizations, today's infrastructure is not enough to handle the demands of AI, ML and DL applications. Many lack the high-performing, scalable storage solutions to house large amounts of data, the processing power to review, label and transform this data or both. What's more, infrastructure at these companies is often focused too much on core processing performance, which neglects data aggregation (especially at the edge) and cloud environments — two components that are essential to creating a data pipeline optimized to run AI applications. The results of this narrow focus: disruptive bottlenecks, wasted costs and frequent data losses.

- **Over 70% of companies have run into limitations** with their on-premise AI infrastructure.⁶
- **Over 90% of companies have run into limitations** with cloud-based AI infrastructure.⁶



Over 90% of companies have run into limitations with cloud-based AI infrastructure.⁶

Creating Your **Data Pipeline**

Going from edge to core to cloud

The key to optimizing your data pipeline for AI is taking a holistic approach, one that addresses the critical areas across your infrastructure — your edge, core and cloud environments.

Traditional approaches

Building this type of holistic pipeline is nothing new. Many organizations use basic commodity hardware or cloud services to piece together a pipeline with enough scalability and firepower to run AI applications. Not only is this approach expensive, but it often leads to bottlenecks throughout the pipeline that prevent AI applications from ever achieving maximum productivity.

A holistic approach

The alternative is to go beyond the capabilities of traditional infrastructure. The right technologies at all five phases of your data pipeline help ensure a smooth flow of data and optimal performance for AI, ML and DL solutions.

1 Data Ingest

Without the right infrastructure, the edge of your network can quickly become a massive bottleneck in your data pipeline, especially if you have a large number of ingestion points. Enterprises with multiple geographically dispersed branches, large mobile workforces or those streaming data from manufacturing or industrial equipment can encounter particular challenges in this regard.

Preventing bottlenecks at the edge requires you to have the tools to selectively pass data using tiered management. Edge analytics enables you to create tiers of data service where you can prioritize data to be passed down the pipeline to feed your AI applications. These tiers of service ensure data is transformed to meet the efficiency and security requirements of your AI applications. They also ensure no critical data gets left behind in the transition from the edge to the core. Because of the important role edge analytics plays in your AI data pipeline and the large volumes of data being processed, the infrastructure supporting it at the edge must be extremely reliable, have vast amounts of computing power and include super low-latency storage.

Another critical AI-enabling technology at the data ingest phase is the smart data mover, a component that delivers data from edge infrastructure to the core. Traditionally, this process involves moving data wholesale without any transformation. Inevitably, this bloats bandwidth requirements and slows down data movement. The “next-generation” approach involves using data movers that can transform your data passing through the pipeline and reduce your data footprint.

2 Data Prep

After being delivered from the edge, your raw data gets collected in your data lake (either in the cloud or on premise) for the next phase in the pipeline — data prep, where data is preprocessed and normalized before the training phase.

In the past, data lakes took the form of object or file stores. These solutions have their limitations and are not optimal for supporting AI workloads. Object storage is great for cloud archiving because it is scalable and flexible, but it isn’t built for performance and is often too slow to support the rapid data flow that drives your AI technologies. In contrast, file stores are built for batch processing and don’t deal well with the small file workloads delivered from the edge.

The solution is a data lake that combines the features of object and file storage. One solution that checks the boxes on all necessary capabilities saves you from spending more on additional, unnecessary storage. As an example, many companies use NetApp® AFF, a 2U all-flash storage array, to meet their batch processing and small-file workload needs so their training clusters can be fed at the regular intervals they require.



3 Training

The training phase of your pipeline is critical to the performance of AI technologies like ML and DL. ML/DL solutions are “trained” by inputting data from the data lake, running it through their specific algorithms and outputting a result that is compared to the target. The model gets adjusted based on that comparison, and the final result is the inference model that will eventually be deployed in the field.

This phase often requires the use of graphics processing units (GPUs) that deliver high input/output (I/O) bandwidth and a constant flow of data so they never go underutilized and your mission-critical data never gets lost. For this reason, a data lake with high-bandwidth, low-latency nodes that enable constant streaming of data into your training cluster is a must. The right “next-generation” infrastructure may even equip you to support many I/O streams simultaneously, ensuring your GPUs remain 100% busy and you can produce quality inference models faster.



4 Deployment

After training, resulting inference models get deployed in a repository where they can be tested and validated by your teams. These repositories might be on premise or in the cloud. Either way, your teams need seamless control and management of the data and any security or compliance requirements must be met. For these reasons, the cloud is an attractive deployment environment.

Seamless data control, management and sovereignty in the cloud is crucial during the deployment phase. If your data lake lives in the cloud, make sure your team has the ability to move, store and manage your AI data in public cloud environments like Amazon Web Services (AWS) or Microsoft Azure. If your data lake is on premise, it benefits to store your data as close to the cloud as possible. This can be accomplished by using cloud-connected colocation facilities that keep your data private but give you a dedicated on-ramp to public cloud environments when you need them. This deployment model enables you to take advantage of the performance and data sovereignty benefits of a public cloud service like AWS without sacrificing control or visibility over your data.

5 Archiving

Most organizations archive cold data indefinitely and, more often than not, this archiving takes place in the cloud. While this data won't be immediately used for your AI deployments, it still needs to be secure and accessible. Many decision makers who are leading AI initiatives turn to technologies that allow them to migrate cold data to secure, low-cost object stores that automatically tier the data. This ensures high volumes of AI data get stored in an organized, easily accessible way that meets data sovereignty requirements and doesn't lock them into a specific cloud environment or vendor.

Start Creating Your AI Data Pipeline

Artificial intelligence's impact on IT is ubiquitous — and it's inevitable. No matter whether you've already invested in AI or you're still considering its application at your organization, it pays to consider whether your infrastructure can meet the demands of AI today and in the future.

As with most things in business, there is no one-size-fits-all solution. Using the ConRes Software-Defined Data Center (SDDC) Framework, we work closely with you to create a holistic solution that ensures every element of your infrastructure and services contributes to meeting the demands of your business. This includes:

- **Completing a virtual network assessment** to understand all traffic across your network.
- **Defining a plan to roll out a next-generation infrastructure solution** to meet your performance requirements.
- **Offering best-in-class services and support** that makes your staff self sufficient.

As experts in IT and networking for more than 50 years, we offer leading next-generation data center solutions from providers like **NetApp** so you can optimize performance and reduce costs and complexities across your IT infrastructure.

¹MESQL, 2017. *2018 Outlook: Machine Learning and Artificial Intelligence*.

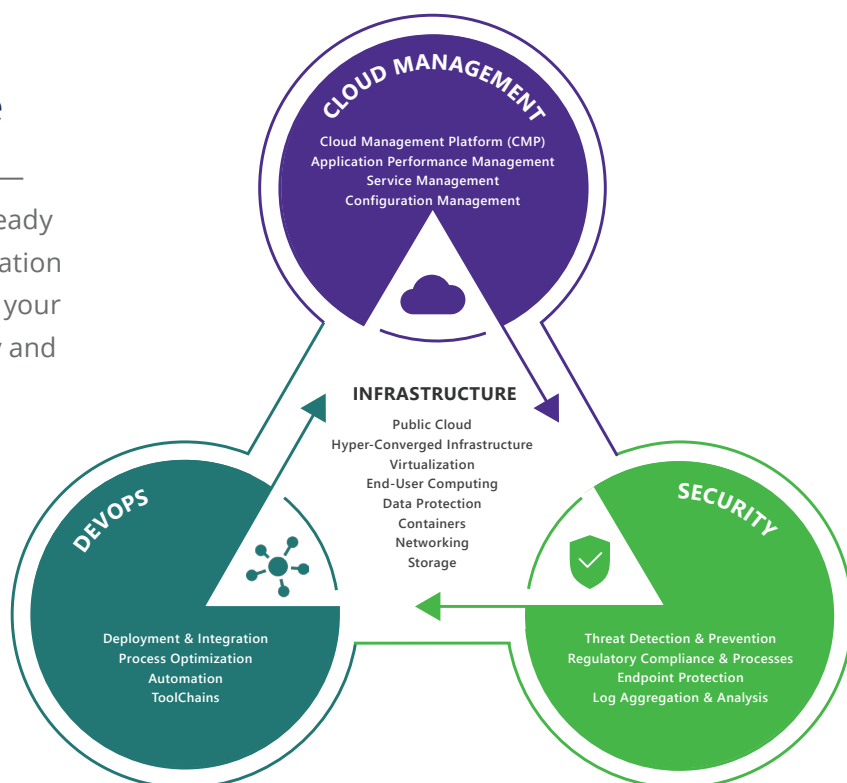
²IDC, 2017. *Considering IT Infrastructure in the AI Era*.

³IDC, 2017. *IDC Spending Guide Forecasts Worldwide Spending on Cognitive and Artificial Intelligence Systems to Reach \$57.6 Billion in 2021*.

⁴Oxford Living Dictionaries. Machine Learning.

⁵Investopedia. Deep Learning.

⁶Rutten, P. & Schubmehl, D., 2017. *Hitting the Wall with Server Infrastructure for Artificial Intelligence*. IDC.



Our team has earned more than 500 technology certifications and our configuration, testing and validation facility enables unparalleled validation of customer solutions in a live environment. Simply put, we offer the most advanced IT solutions delivered by the most experienced team while providing you with industry-leading personal service and support.

Why NetApp for Your AI Data Pipeline?

NetApp offers you the technology and services to future-proof your data pipeline for AI, machine and deep learning workflows. Through products like its AFF A-Series storage technology and ONTAP® data management software, NetApp delivers:

- **Balanced** performance
- **A seamless flow of data** from edge to core to cloud
- **Powerful**, in-place analytics

