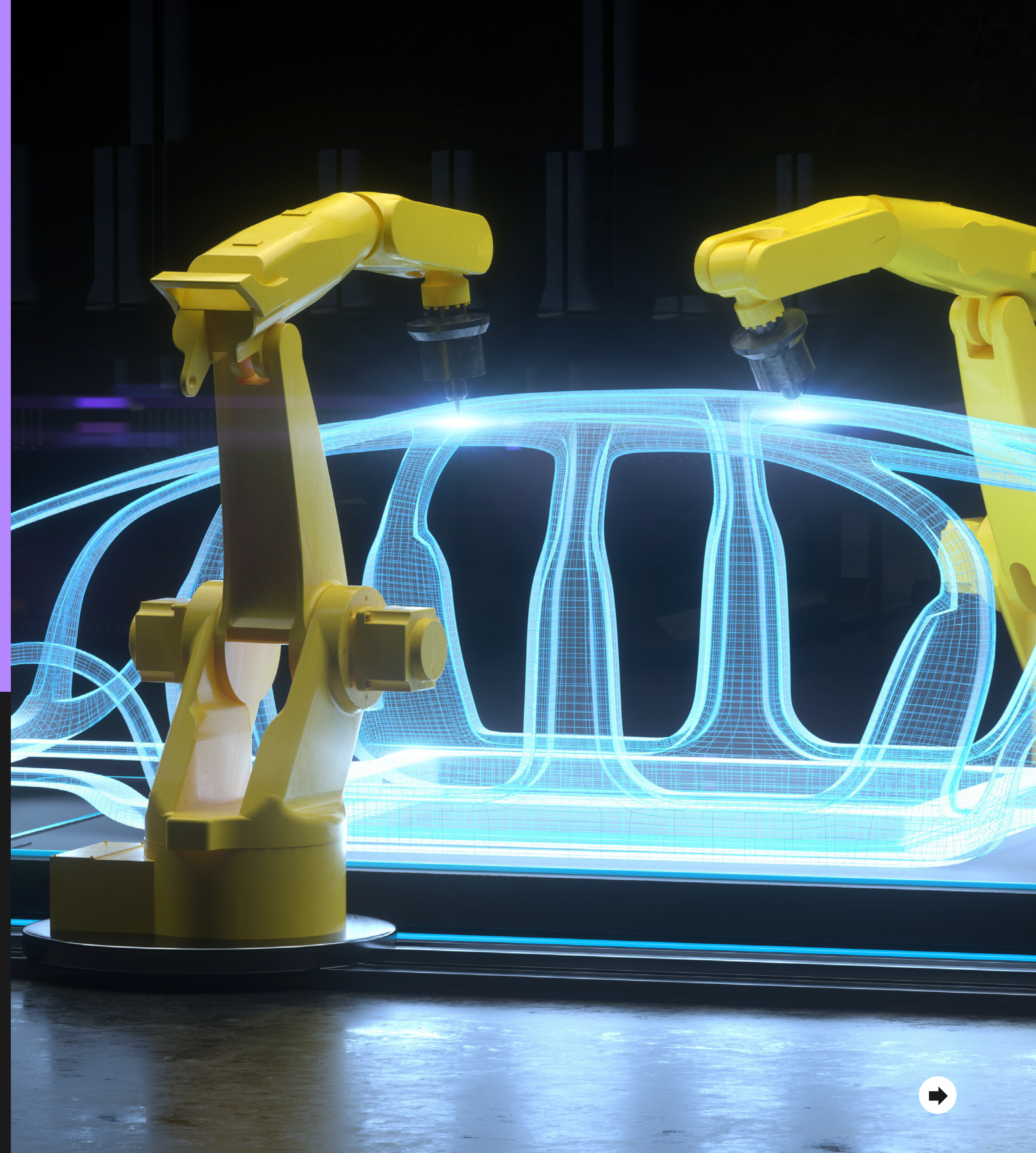


E-BOOK

Transforming industry with data-driven AI

➔ netapp.com/ai

 **NetApp**



Leading data-driven AI

Executive summary

Artificial intelligence (AI) is transforming thousands of companies in industries worldwide. The rapid advancements require massive amounts of data. But AI is only as good as the data that it uses; data quality and movement are also critical. That means you need a smart, powerful, and trusted data architecture to unleash the full power of AI.

Wherever your organization is on the journey to AI, this e-book can help you chart a more direct path to success.

Read on to learn about:

- The importance of maximizing your data pipeline
- How breaking down data silos improves the flow of data
- The powerful business benefits of performance and scalability
- Recommendations for meeting AI architecture challenges
- Transforming your business with trusted AI solutions

➔ netapp.com/ai



Contents

- 1. Data-driven AI 4 →
- 2. Master your data pipeline 6 →
- 3. Integrate your data fabric across edge, core, and cloud 9 →
- 4. Deliver the power of performance with scalability 10 →
- 5. Overcome AI architecture challenges 14 →
- 6. Smart, powerful, and trusted AI solutions from NetApp 16 →
- 7. Accelerate your journey to AI 18 →

1. Data-driven AI

From increasing customer engagement, to streamlining processes, to creating a sustainable competitive advantage, AI has opened a vast well of business opportunity across industries.

And thousands of companies worldwide are working hard to tap into its value. AI is expected to top IT investment agendas for at least 5 years, affecting digital transformation in every industry.

With the help of AI, insights that were previously too complex to envision are now well within striking distance. And it's no longer all about quantitative data; today's powerful deep learning and machine learning technologies can also use images and voice to further advancements in data analysis.

This e-book can help you speed your journey to AI regardless of where your organization is on its digital transformation path. Learn how an AI architecture that's smart, powerful, and trusted can provide a platform for unlocking the potential of your data.

“There was nearly universal acknowledgment—96%—that big data and AI efforts were yielding results”¹

— NewVantage Partners, Big Data and AI Executive Survey 2021

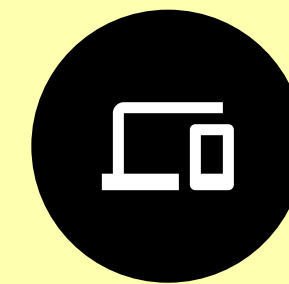
Data-driven AI

Data—the driving force of AI

Large amounts of data are fundamental to AI success. So all your important data needs to be usable, accessible, and protected. Of course, that's easier said than done. In today's enterprises, the data sources that feed AI systems are significantly different from those of the past. They provide better quality data, but there are more of them in more places. We're talking about distributed, diverse, and dynamic tangles of complexity.

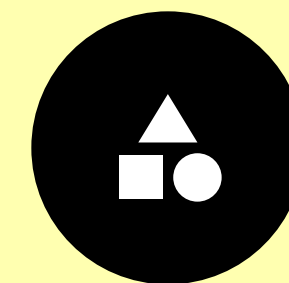
PPDAI Doing the Imaginable

PPDAI is transforming financial services in China by bringing online services to the underserved—residents who lack credit bureau data. As the country's first internet financing firm, PPDAI relies on vast amounts of data from diverse sources to serve its 60 million customers. In the absence of credit scores, PPDAI uses machine learning and AI to analyze mobile, social network, and behavior data and to assess risk. And the process is fast. People who apply and qualify for loans using PPDAI's mobile app receive their money within minutes. The reach and success of the PPDAI platform is possible only through rapid, secure data movement and analysis.



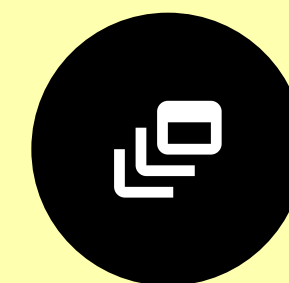
Distributed.

To solve a big challenge using AI, the data you need no longer lives only in your data center. It could come from Internet of Things (IoT) devices at the edge of your network, from the cloud, or even from third-party sources.



Diverse.

Video. Audio. Images. Machine-generated data. Transactional. And much more... AI projects often require a range of structured and unstructured data sources. Just when you thought you had a handle on complexity, new data sources and types can further complicate data management and governance.



Dynamic.

AI datasets are always growing and changing. So, it isn't easy to keep track of where data is, where it came from, and where archives live.

Consider an organization that's using data aggregated from worldwide sources, including mobile applications, social media, and point-of-sale devices. To gain actionable insights, the organization must keep all those data sources up to date in close to real time.

2. Maximize your data pipeline

Just how important is data management to machine learning and deep learning processes? It's like high-quality practice for a pro sports team. Ignoring fundamentals can quickly lead to mistakes and shoddy play, despite the caliber of players. And IT teams that are new to AI often underestimate the relationship. An MIT Sloan survey identified the following common data challenges when adopting AI.

Lack of data.

AI algorithms depend on data to predict real-world events. To create an AI algorithm to anticipate failures, you need more than just data from normal operations. You also need plenty of log data from failures that have occurred.

Unexpected legwork.

Identifying the right data sources, collecting the data, and preparing it can be time-consuming.

Unclear data ownership.

You might have difficulty gaining access to some valuable data because it's proprietary or because ownership is uncertain or contested.

Connecting the data dots.

Particularly in large organizations, data can be fragmented across multiple corporate systems, complicating the training process.

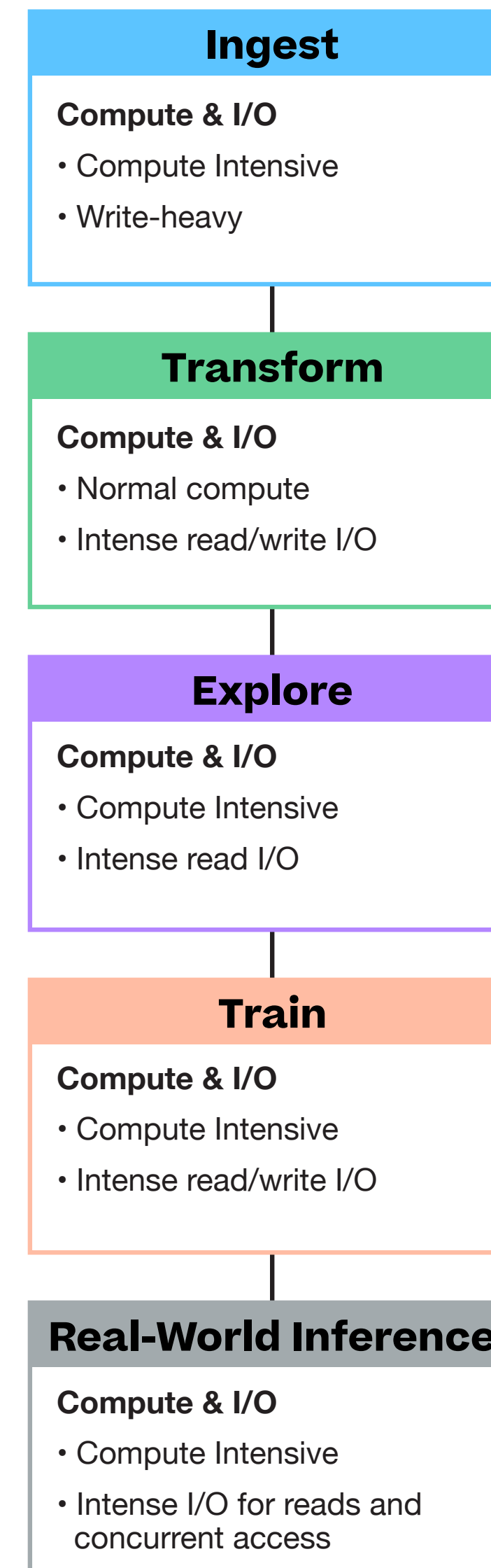
These challenges all tie back to data flow. And overlooking any one of them could put your AI efforts on the skids. That's why your data pipeline is so critical.

Maximize your data pipeline

What is a data pipeline?

A data pipeline enables data to flow smoothly from the point of creation through the entire AI process. And the creation points are virtually infinite... An IoT device in a factory? Yes. Customer equipment? You bet. A social media platform or corporate system of record? Yep, yep.

Steps along the pipeline include those shown in Figure 1.



Data creation often occurs at the edge of your network on things like intelligent point-of-sale devices. And IoT-connected devices and sensors are only becoming more prevalent across all industries.

Before you can do any training, you need to normalize data. Preprocessing options include a data lake, an Amazon Simple Storage Service (Amazon S3) object store in the cloud, or a file store on premises.

During the explore stage, you figure out which deep learning models are most promising based on your datasets and the outcome that you want.

Training inference models involves periodically moving data into the training cluster. It's an iterative process that typically takes place at regular intervals throughout the life of a model.

Trained and tested inference models are deployed in the real world. Models are often stored in a DevOps-style repository where they benefit from ultralow-latency read access.

Figure 1) A well-designed data pipeline enables data to flow freely through multiple AI stages, meeting the unique I/O requirements of each stage and preventing bottlenecks.

Maximize your data pipeline

Smart inhalers use an edge-to-core-to-cloud approach to improve patient health

Whether you're performing AI processes in the cloud or on premises, a data pipeline must be able to accommodate unique compute and I/O needs at each stage. And a data storage approach that works well across the entire pipeline is essential. Consider how compute and I/O requirements change as data progresses through the pipeline with smart inhalers.

In the United States, 25 million people suffer from asthma—that's 1 in every 13 people. Imagine the suffering you might eliminate by correlating inhaler usage with information about location, weather, air quality, and pollen counts in real time to help patients avoid triggers. It's possible by adding a few sensors to asthma inhalers and relying on AI.

Cambridge Consultants, a NetApp customer and AI innovation partner, demonstrated the potential of a smart inhaler solution that relied on NVIDIA GPUs and NetApp data management. And it's a great example of determining the requirements of AI at scale:

- Data flows from thousands of devices at the edge.
- That data is combined with outside datasets during training in a GPU-accelerated data center.
- The resulting inference model is deployed in the cloud to analyze new data points and to identify and act on trigger events.

3. Integrate your data fabric across edge, core, and cloud

Are you eager to take advantage of AI technologies to introduce new services and to enhance insights from company data? Good for you!

Just don't forget that AI workflows are complex. Some move from the edge of your network to your core data centers and then to the cloud. Others might start at the edge of your network and go directly to the cloud, and then from there go to the core or to the edge again.

Yarn balls aside, data can also require extensive preprocessing before training—work that might be done at the edge, in your core data centers, in the cloud, or in all three. AI models can be trained in the core or in the cloud, and you might need to archive the data in the cloud for future use or for compliance reasons.

NetApp unites data across the AI pipeline so that you can make it available wherever it's needed.

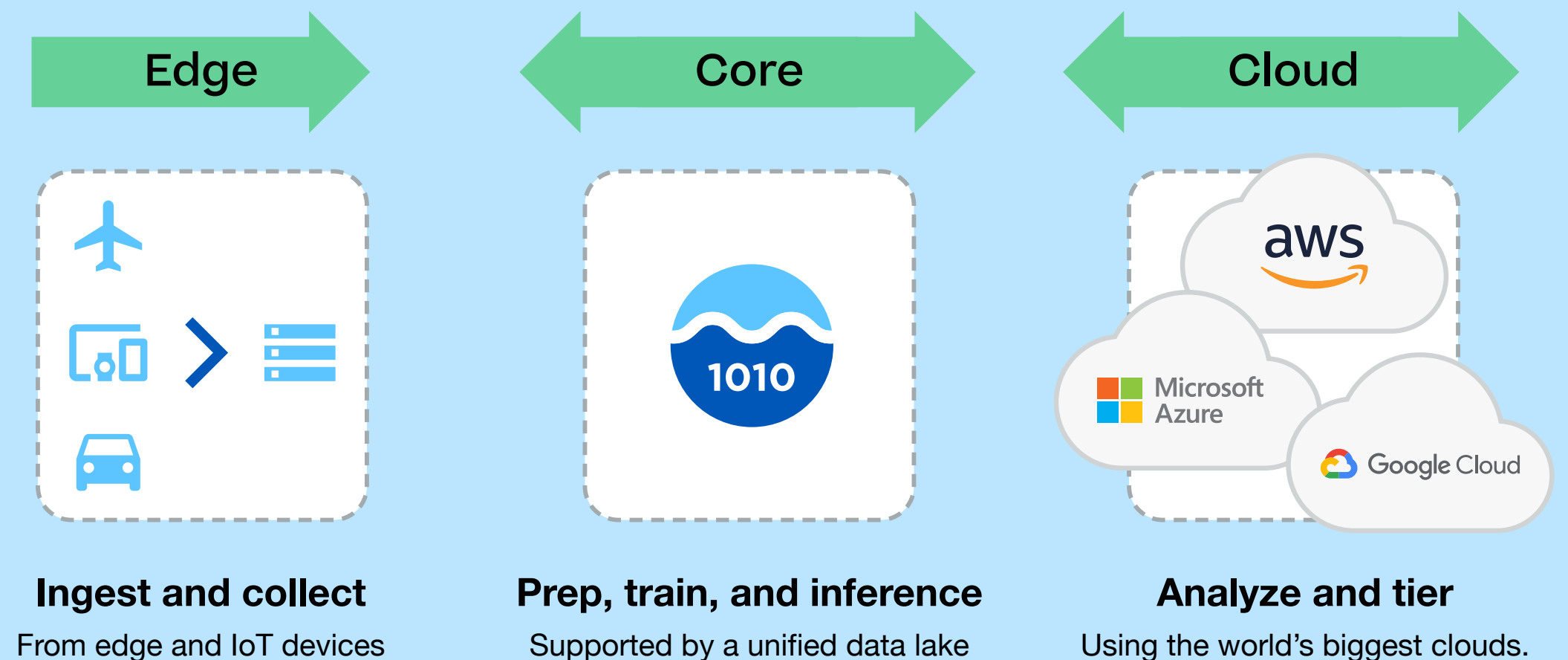


Figure 2) Only NetApp enables you to integrate their data fabric across edge, core, and cloud.



4. Deliver the power of performance with scalability

Here's a sentence worth reading twice: For successful AI, the data pipeline requires extreme performance and scalability at every stage—at the edge, in the core, and in the cloud. The fact is that too many organizations underestimate the challenges that come with marshaling and managing vast quantities of data. And it's usually a painful lesson.

Bottlenecks at any point in your pipeline idle expensive infrastructure, increase costs, and waste the time of data scientists. Those consequences alone are bad enough, but for many use cases, bottlenecks put outcomes at risk. For example, with a smart inhaler, getting a late alert about a trigger event as you're struggling to regain your breath after an asthma attack doesn't exactly add value.

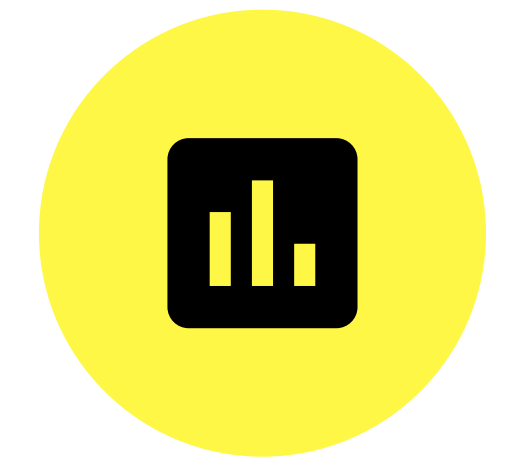
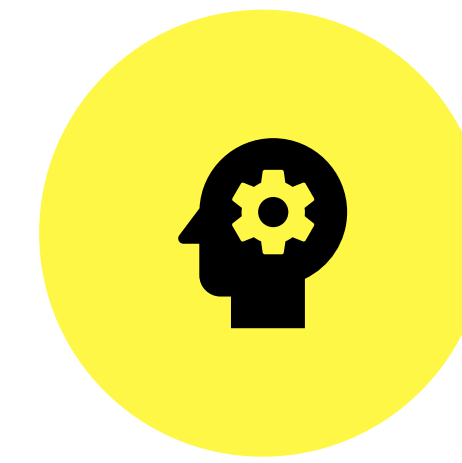
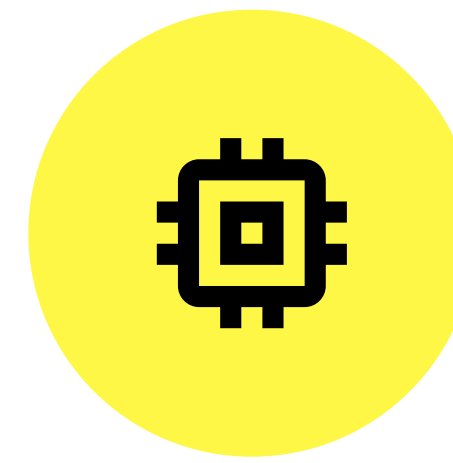
Deliver the power of performance with scalability

GPU power/performance: A delicate balancing act

To generate actionable insights that are based on the most current data, some models must be trained in close to real time. But if GPUs aren't matched with equally powerful data-access performance, achieving immediacy can be nearly impossible. So what does it take to reap the performance benefits of expensive GPUs?

High-performance data storage.

Whether you're performing AI training on premises or in the cloud, you need GPUs to run as efficiently as possible. Leaving them sitting idle while waiting for data prolongs each training cycle, reduces your training productivity, and puts big dents in the bottom line.



Why Are GPUs so important to Deep Learning?

In the early 2000s, computer scientists realized that the GPU video manipulation capabilities could be applied to a wider variety of parallel computing problems.

The deep learning algorithms at the heart of today's data-driven AI require a huge volume of matrix multiplication that's executed in parallel. That need makes GPUs, or similar parallel processing hardware, essential for data exploration and inference model training. Today's GPUs—from companies such as NVIDIA—provide hundreds or thousands of simple cores.

GPU performance is growing rapidly, so your data pipeline needs to keep up. It's like the difference between taking a high-end sports car onto a clear racetrack versus taking it onto a freeway in rush hour traffic... Despite looking cool and fast, you're at the mercy of fellow drivers.

Deliver the power of performance with scalability

Deliver the power of performance with scalability

The type of data in your AI pipeline affects the I/O pattern, which might affect the speed with which training runs are completed. Unstructured data can be coalesced into a set of streams, creating sequential I/O. Sources such as databases, sensor logs, file logs, and emails must be accessed by using random reads.

These two types of I/O—sequential and random—are quite different. It's common to use both unstructured and structured data sources to train an inference model, so your architecture must support both types of I/O to avoid bottlenecks that stall the pipeline.

Seamlessly transition from prototype to production

As you move from prototype to production, it's critical to consider the needs of your various data users: data engineers and data scientists. Data scientists, for example, can't wait months for the data pipeline environment to go into production. They're looking for rapid deployment, whether on premises or in the cloud. And they expect rapid scaling of the environment after you're in production.

It can be difficult to predict up front how successful an application is going to be. Success can mean much more data and more frequent training. And that creates greater demands for both GPU and I/O, as well as for the resources that you need to support the additional load. That's why it's well worth considering powerful architectures that are easier to bring online.

Deliver the power of performance with scalability

Take AI to the next level

After your team has a few AI projects in production, you can build on what you've created to expand to new AI use cases. Look for opportunities to use a similar AI approach to streamline business processes, to create a competitive advantage, or to expand into new markets.

How can you achieve success? Focus on scaling out the infrastructure and processes that you've already created and incorporating the latest, best-in-class capabilities with trusted solutions. And remember, up-front investments in AI can be significant whether they're in the cloud or on premises.

You might start by provisioning GPUs in the cloud, and then find that your data is growing so fast that deploying an on-site training cluster is more cost effective. Or you might start on site and burst to the cloud as needed if your data grows in spurts. Flexibility is the key to success.

“AI will be viewed as an architecture and will help unearth radical transformation in business processes and worker performance”²

— IDC Market Analysis Perspective, Worldwide Artificial Intelligence Software, 2020

5. Overcome AI architecture challenges

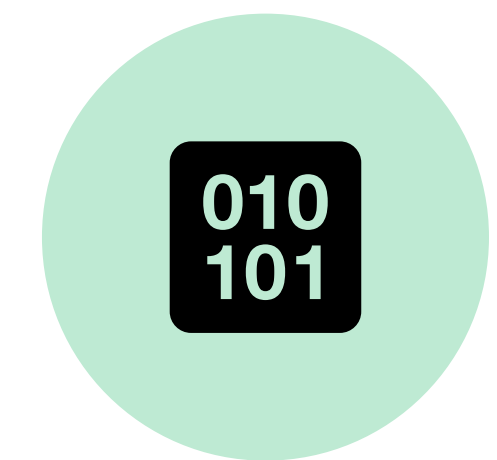
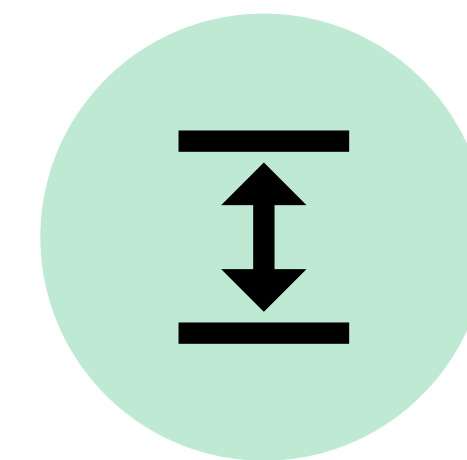
What do you think? Does building effective AI infrastructure from the edge to the core to the cloud sound complex? It can be, but there are also ways to build your infrastructure so that it can support significant growth.

Avoid point solutions

An AI data pipeline can consist of five or more distinct stages. If you rely on point solutions in each stage, it's a recipe for complexity. An optimal AI architecture enables you to use the best compute and data storage solution for the requirements in each stage, while providing uniform data management and data efficiency across all stages.

Choose technologies that offer superior scale

An AI project can start small but then scale to a size that's hard for even IT savants to predict. That means your data pipelines might have to scale to accommodate petabytes or even exabytes of data. After all, a single autonomous vehicle generates a terabyte or more of data for every hour of operation, so you can see how some training sets grow beyond 500PB. Processing petabytes of data requires a balance of tremendous I/O bandwidth and compute performance. And if your solution can't scale to meet your capacity and performance requirements as they increase... well, prepare to put your initiative on hold while you rearchitect operations.



Overcome AI architecture challenges

Think hybrid multicloud to provide for new technologies

The AI field is evolving quickly, so it's essential to be ready for change. Gaining the ability to deploy projects on premises or to choose from best-in-class AI services in the cloud will help you maximize flexibility. An end-to-end AI pipeline is your foundation for success. It's how you keep workloads humming no matter where your data lives—or where it moves to—from the edge to the core to the cloud.

Choose integrated, validated solutions for rapid deployment

Are you looking for ways to reduce the time that's required for AI infrastructure deployment? The key is to choose validated and integrated solutions whenever possible. A proven, integrated solution can help you avoid costly errors and reduce your time to insight.

It can also help you:

- Eliminate design guesswork by delivering predictable performance that scales.
- Reduce deployment complexity by simplifying procurement, installation, and troubleshooting.
- Streamline support by eliminating multiple points of contact.

Remember: Make sure that any solution you choose scales across a wide range of performance and capacity. And be sure to look for one that doesn't lock you into different point solutions at every stage of the data pipeline.

6. Smart, powerful, and trusted AI solutions from NetApp

At the end of the day, it can be very difficult to satisfy all your AI objectives unless you get help with both AI infrastructure and software. That's why it pays to work with specialists. We can't fix your city's traffic flow to make your supercar commute more fun. But we are data flow experts, and we can help you unleash the full power of your data by designing and implementing a strategy for AI infrastructure that supports innovation at scale across a hybrid multicloud environment.

Only NetApp enables you to integrate your data fabric across edge, core, and cloud, and offers smart, powerful, and trusted solutions to help you unlock the potential of data science by using machine learning and deep learning. Choose NetApp® solutions to enable a future-proof platform that will support your data-driven AI journey from predictive analytics to autonomous decisions.



Smart, powerful, and trusted AI solutions from NetApp

Businesses are adopting AI faster than ever to stay competitive, drive growth, and decrease expenses. Whether you're powering chatbots, predictive maintenance, or genomic medicine, successful AI deployments depend completely on data. You need a smooth data pipeline on premises, in the cloud, and for hybrid cloud deployments. NetApp AI solutions can help.

NetApp helps you build a tailored data fabric to speed your journey to AI. Only NetApp enables you to integrate your data fabric no matter where it resides.

With NetApp solutions, you can:

- Simplify data management with industry-leading technology that streamlines AI deployments so your data scientists can focus on science instead of IT.
- Seamlessly integrate data on premises, in the cloud, and in hybrid cloud environments.
- Configure your AI infrastructure in ~20 minutes with Ansible integration.
- Run **5 times more data through your data pipeline**, decreasing your time from data to insight.
- Reduce the time to copy a dataset from **days to seconds**.
- Keep data secure no matter where it resides by using the AI industry's most complete native set of data protection and data security features.

[Discover what's possible with NetApp AI solutions.](#)

Table 2) The NetApp Data Fabric meets data management challenges across the edge, the core, and the cloud.

7. Accelerate your journey to AI

To learn more about the entire NetApp AI solutions portfolio, including NetApp ONTAP® AI, visit netapp.com/ai.

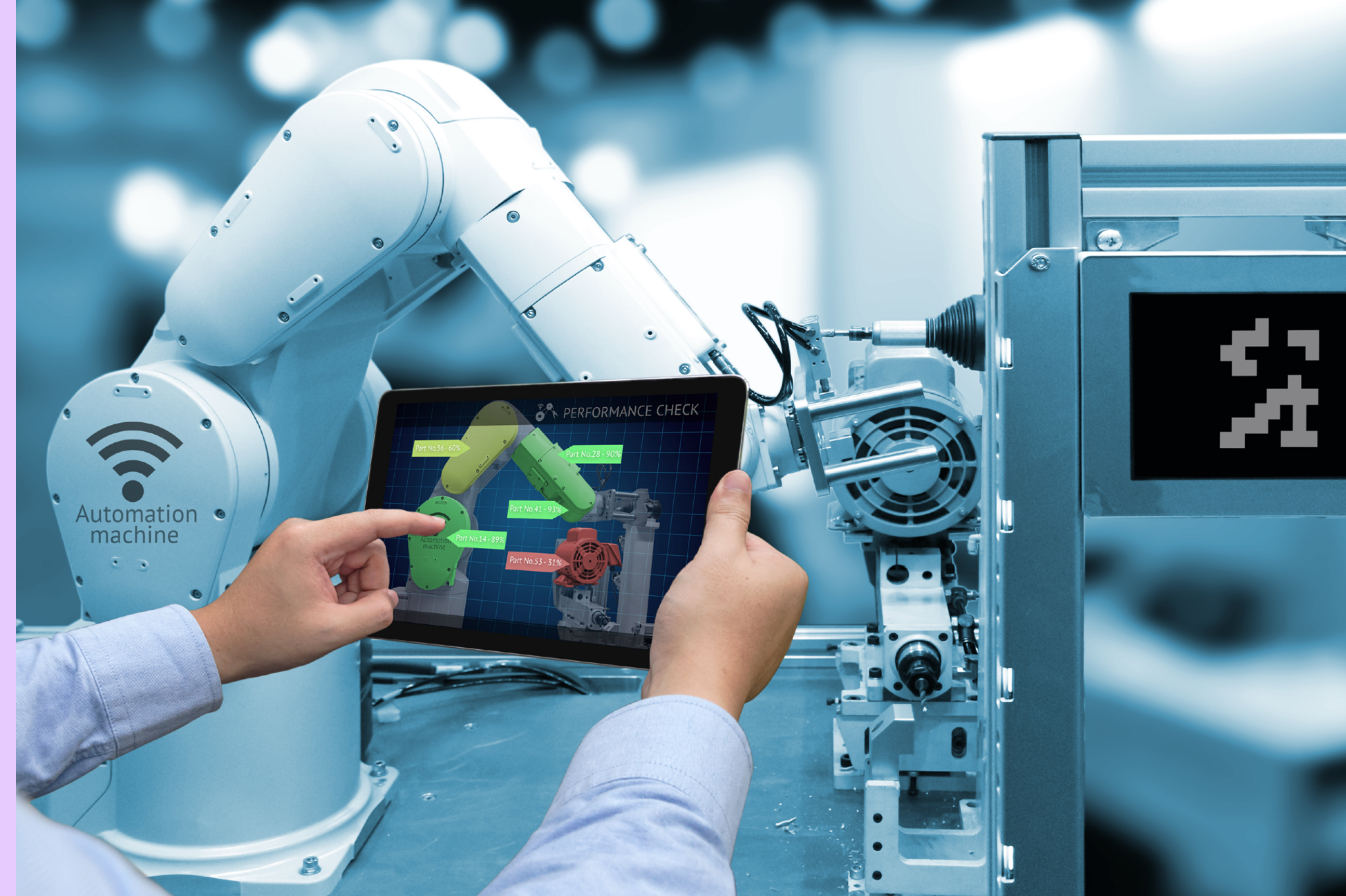
[Discover 10 good reasons for choosing NetApp AI solutions.](#)

Questions? Connect directly with a NetApp AI solution specialist today.

 **Meet with our specialists.**

¹ NewVantage Partners, [Big Data and AI Executive Survey 2021](#).

² IDC Market Analysis Perspective, [Worldwide Artificial Intelligence Software, 2020](#).



About NetApp

In a world full of generalists, NetApp is a specialist. We're focused on one thing, helping your business get the most out of your data. NetApp brings the enterprise-grade data services you rely on into the cloud, and the simple flexibility of cloud into the data center. Our industry-leading solutions work across diverse customer environments and the world's biggest public clouds.

As a cloud-led, data-centric software company, only NetApp can help build your unique data fabric, simplify and connect your cloud, and securely deliver the right data, services, and applications to the right people—anytime, anywhere.