ESG WHITE PAPER

Accelerate Artificial Intelligence Initiatives with DDN and NVIDIA at Any Scale

Enterprise and HPC End-users Describe How DDN's A³I Solutions with NVIDIA DGX™ Systems Accelerate Artificial Intelligence Initiatives

By Scott Sinclair, ESG Senior Analyst and Monya Keane, ESG Senior Research Analyst

June 2021

This ESG White Paper was commissioned by DDN and NVIDIA and is distributed under license from ESG.

CONTENTS

INTRODUCTION

Data now plays such a prominent role in determining business success that 98% of organizations surveyed by ESG are in some phase of data-driven digital transformation. The goals of these near-ubiquitous initiatives include attaining greater operational efficiency, transforming to fuel revenue growth, and unlocking new market opportunities. Specifically, ESG found that:

> **56%** of respondents sought to become more operationally efficient.

> **40%** sought to provide a superior customer experience.

> **38%** sought to develop new innovative products and services.

> **25%** sought to develop entirely new business models.[1]

Investing in artificial intelligence (AI) and machine learning (ML) can be a key to success in digital transformation. However, given that the adoption of AI/ML-based initiatives is relatively recent and rapid, it can be challenging to determine how to architect the environment properly. And the size of that challenge only increases as the organization scales.

A need clearly exists for IT infrastructure itself to enable and empower the considerable investments organizations are making in their data scientist teams. Relying on hardware with traditional functionality is not enough. IT teams need both expert people and expert tools to achieve business success through AI.

DDN and NVIDIA are leaders in infrastructure. But more importantly, both vendors are experts in AI/ML and serve as valuable partners to many organizations that are ramping up AI/ML environments. With the introduction of NVIDIA's DGX™ A100 systems with eight NVIDIA A100 GPUs and two second generation AMD EPYC™ processors, these allied vendors have raised the bar once more. DDN and NVIDIA's technical collaboration has yielded turnkey infrastructure that delivers a number of benefits for enterprise IT shops and their end-users at any scale, with the parallelism of DDN's A3I solutions ensuring the full utilization of the NVIDIA A100 GPUs.

DDN and NVIDIA invited ESG to interview several of their joint customers that have been running production AI environments to better understand their needs and to confirm the value that DDN and NVIDIA are providing.

THE ARTIFICIAL INTELLIGENCE RACE AND WHAT IT MEANS TO MODERN BUSINESS

ESG research sheds light on the rapid rise of AI initiatives among businesses, including how businesses are leveraging AI to create value. Business have been rapidly investing in AI and ML projects because they work. Consider that:

> **82%** of organizations saw value from AI/ML in six months or less.

> Organizations with AI/ML in production are twice as likely to identify as being among the leaders in their competitive markets **(29% versus 15%)** than those still in the early stages of AI/ML implementation.

> Organizations with AI/ML in production are more than twice as likely to report being in a very strong business position **(55% versus 26%)** than those still in early implementation stages.[2]

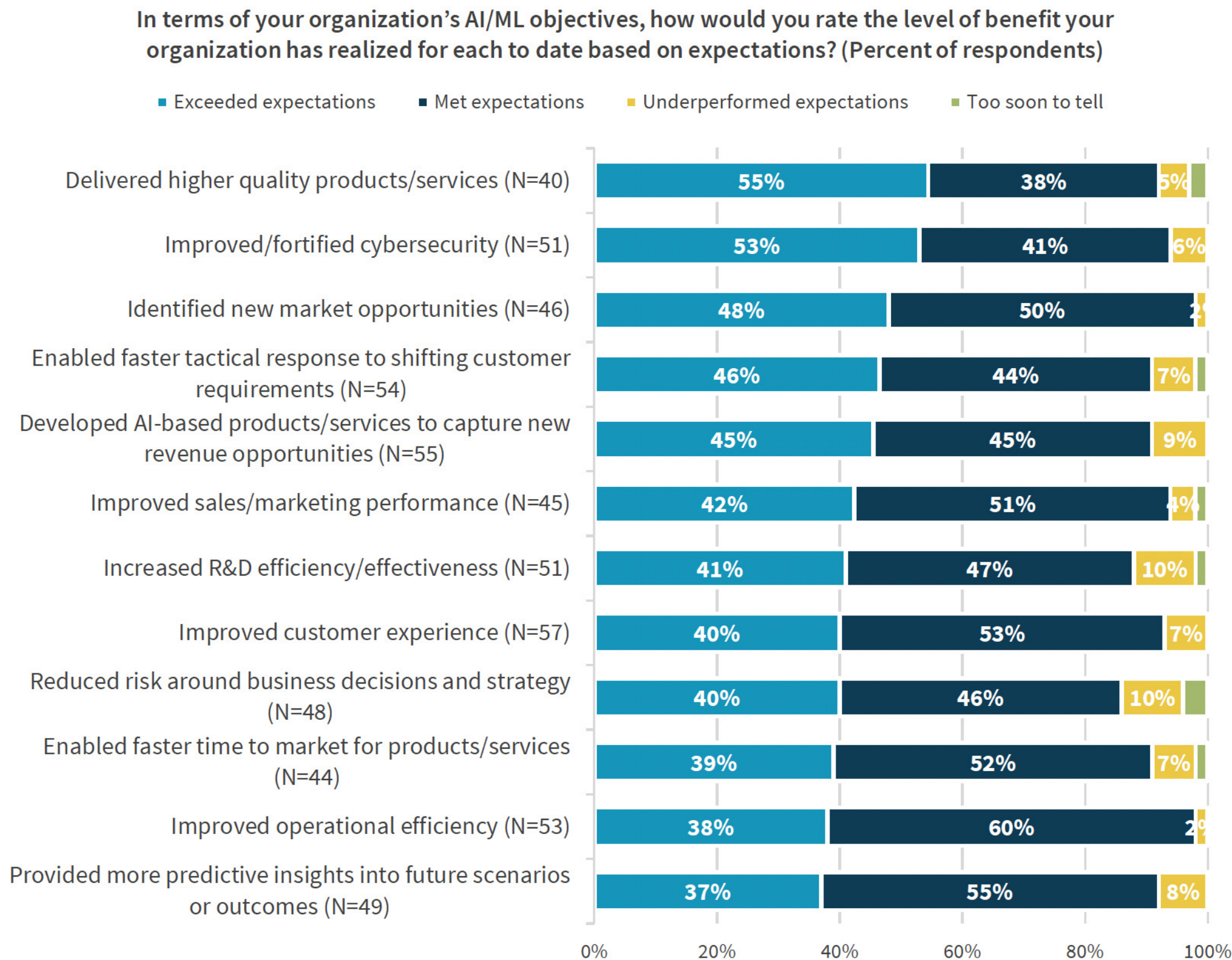[1] Source: ESG Research Report, 2021 Technology Spending intentions Survey, January 2021.

[2] Source: ESG Master Survey Results, Supporting AI/ML Initiatives with a Modern Infrastructure Stack, May 2021.

[3] Ibid.

Figure 1 shows that regardless of an organization's specific objectives, their AI/ML projects are meeting or exceeding their expectations at a staggering rate. 3 That success is fueling increased interest and investment in AI initiatives—63% of organizations leveraging AI expect to accelerate those investments in 2021.4 But as organizations accelerate AI adoption and usage, IT is presented with the burden of delivering an infrastructure environment that can keep pace.

Figure 1. Success Rates in AI/ML Projects by Objectives

In terms of your organization's AI/ML objectives, how would you rate the level of benefit your organization has realized for each to date based on expectations? (Percent of respondents)

- Exceeded expectations
- Met expectations
- Underperformed expectations
- Too soon to tell

Objective	Exceeded	Met	Underperformed
Delivered higher quality products/services (N=40)	55%	38%	5%
Improved/fortified cybersecurity (N=51)	53%	41%	6%
Identified new market opportunities (N=46)	48%	50%	2%
Enabled faster tactical response to shifting customer requirements (N=54)	46%	44%	7%
Developed AI-based products/services to capture new revenue opportunities (N=55)	45%	45%	9%
Improved sales/marketing performance (N=45)	42%	51%	4%
Increased R&D efficiency/effectiveness (N=51)	41%	47%	10%
Improved customer experience (N=57)	40%	53%	7%
Reduced risk around business decisions and strategy (N=48)	40%	46%	10%
Enabled faster time to market for products/services (N=44)	39%	52%	7%
Improved operational efficiency (N=53)	38%	60%	2%
Provided more predictive insights into future scenarios or outcomes (N=49)	37%	55%	8%

Source: Enterprise Strategy Group

CHALLENGES WITH DEPLOYING INFRASTRUCTURE FOR AI

For IT teams, supporting an AI infrastructure adds challenges and complexity. Overall, 75% of IT organizations surveyed by ESG report that IT has become more complex in just the past two years. 28% of those organizations identified the need to incorporate emerging technologies such as AI/ML as a driver of that IT complexity. Compounding the problem is the fact that 36% of IT organizations report problematic skill shortages in the area of AI/ML.[5]

[4] Source: ESG Research Report, 2021 Technology Spending intentions Survey, January 2021
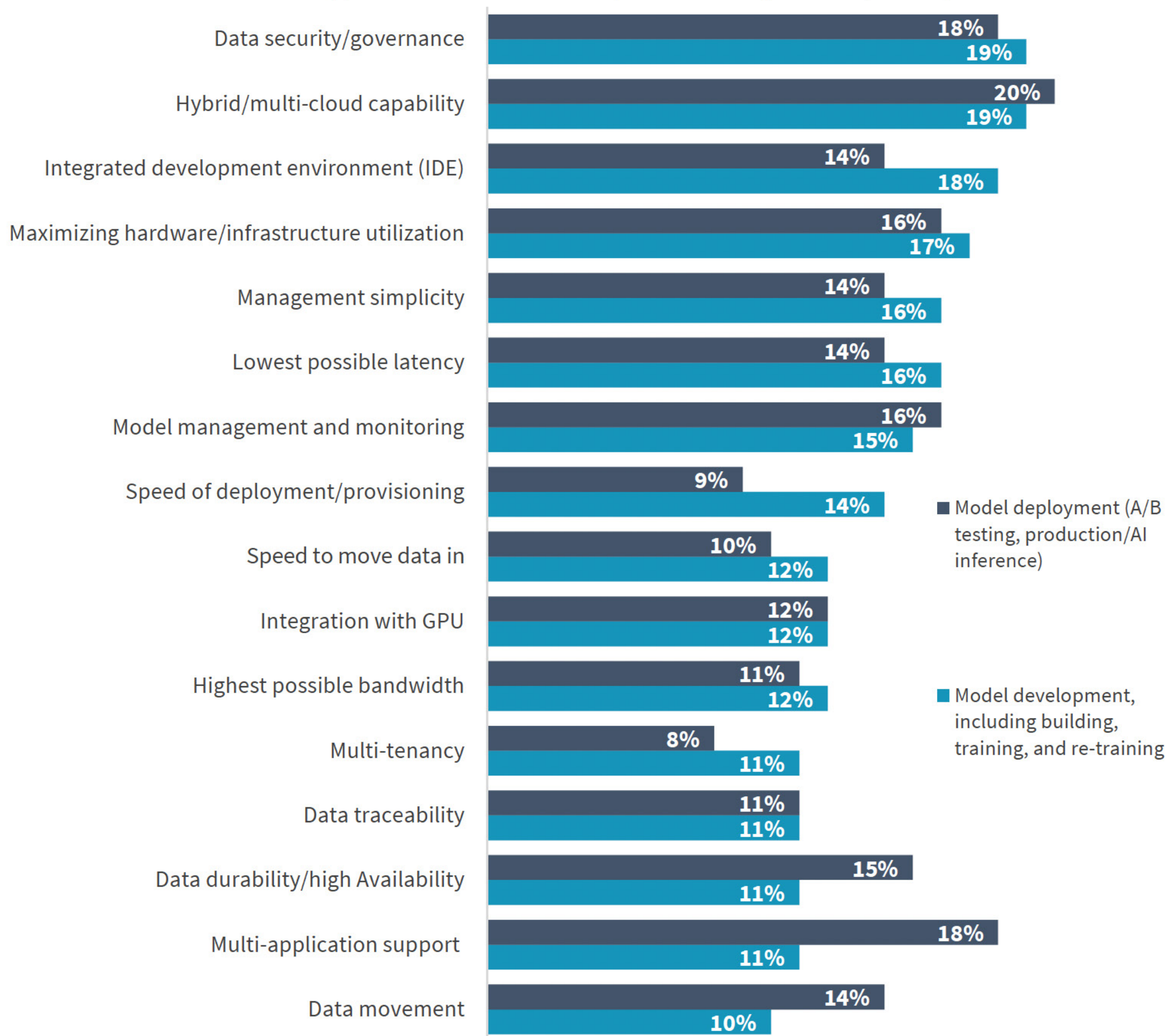[5] Ibid.

The right storage infrastructure makes a tremendous difference to the success of AI projects. When ESG asked IT decision makers who support AI initiatives to identity the weakest links of their AI infrastructure stack, 22% identified data storage. Selecting the right data storage architecture is critical to AI/ML success.[6]

Figure 2 describes the top infrastructure-related considerations for supporting AI/ML initiatives, with a focus on model development and deployment.7 It highlights key considerations, including data governance and data security—AI/ML projects are very likely to be associated with sensitive data. Additionally, organizations are looking for good hardware utilization with the lowest possible latency and integration with the AI-accelerating GPU.

Figure 2. Infrastructure Considerations for AI/ML Model Development and Deployment

For each part of the data pipeline, which of the following are—or likely will be—the most important in your organization's consideration of the infrastructure solution(s) used to support its AI/ML initiatives? (Percent of respondents, N=325)

Consideration	Model deployment (A/B testing, production/AI inference)	Model development, including building, training, and re-training
Data security/governance	18%	19%
Hybrid/multi-cloud capability	20%	19%
Integrated development environment (IDE)	14%	18%
Maximizing hardware/infrastructure utilization	16%	17%
Management simplicity	14%	16%
Lowest possible latency	14%	16%
Model management and monitoring	16%	15%
Speed of deployment/provisioning	9%	14%
Speed to move data in	10%	12%
Integration with GPU	12%	12%
Highest possible bandwidth	11%	12%
Multi-tenancy	8%	11%
Data traceability	11%	11%
Data durability/high Availability	15%	11%
Multi-application support	18%	11%
Data movement	14%	10%

Source: Enterprise Strategy Group

[6] Source: ESG Master Survey Results, Supporting AI/ML Initiatives with a Modern Infrastructure Stack, May 2021.
[7] Ibid.

THE VALUE OF DDN'S A3I SOLUTIONS WITH NVIDIA DGX™ SYSTEMS—INSIGHTS FROM USERS

DDN and NVIDIA are both leaders in addressing the infrastructure needs of production AI environments. NVIDIA is a leader in GPU technology and solutions for AI environments, and DDN is a leader in intelligent storage infrastructure for high-performance computing AI environments. Together, they have released reference architectures that simplify the selection, configuration, purchase, and deployment of AI infrastructure utilizing anywhere from one DGX system to hundreds. These architectures are designed for power and simplicity, delivering an immediate boost to AI application performance while removing the complexity associated with managing data- and performance- hungry AI workloads.

ESG had the opportunity to speak to several IT leaders who support AI/ML environments and validate the value of the combined solutions DDN and NVIDIA provide.

Use Case One—An Interview with the Storage Engineering Product Lead, Financial Industry

This organization chose DDN and NVIDIA because it wanted to build a shared data science environment for all of its data scientists, "getting away from everyone having their own workstation under their desk," as the storage engineering product lead put it. "We needed to have actual knowledge of what was being run, instead of having our people stealthily doing who knows what," they said.

> "
> With this high-performing storage, we churn through more data. No GPUs sit idle.
>
> _____
>
> **Storage Engineering Product Lead,** Financial Industry

This is a financial institution, and it must be very careful with how data is protected. Data needs to be in a secure, protected environment, and the IT group needs to know where all of it lives. The product lead said, "With machine learning as a service, our teams can do what they need to, and we receive meaningful telemetry about what is happening on the system, who is using it, how busy it is, and what they were running on it. And we can then back it up properly. We can't back up someone's workstation under a desk so easily."

Every bank has a machine learning initiative to support fraud detection, analytics, and more. AI is mission- critical, and if the bank is using it for fraud prevention and the system goes down, any fraud it might have prevented earlier can now be committed.
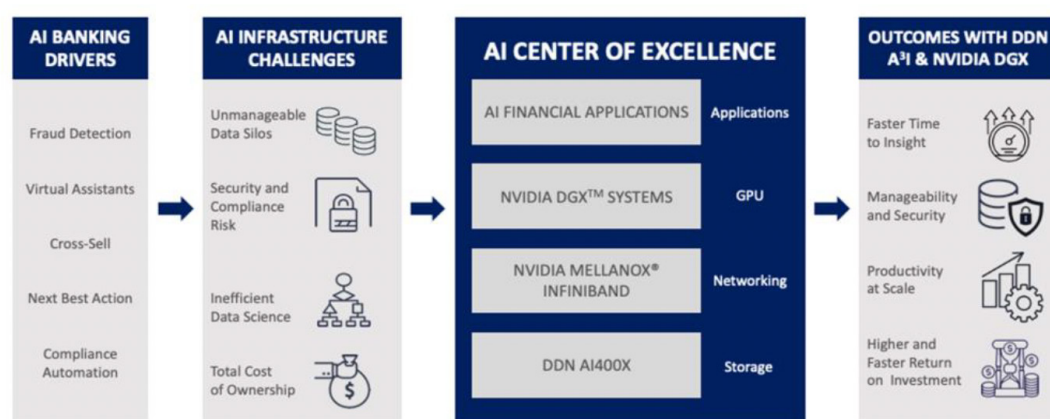
The challenge was manageability. "Data science teams know everything about data science but nothing about infrastructure, nothing about networking, and nothing about security," they said. By leveraging DDN storage, they were able to simplify their organization environment, saying "I am managing a much bigger infrastructure environment but one that has far fewer devices [which is easier to manage]."

Another positive factor is that AI software is included with the solution. Instead of having to build their own, they use AI/ML libraries from NVIDIA that are pre-packaged. "The AI/ML libraries are pre-packaged and pre-engineered with all the kinks worked out of them. It's a major time saver for us," they said.

Perhaps the biggest improvement of all is driven by DDN storage, enabling the

data scientists to look at more data. "This is big, fast storage—we can look at more workloads, which means we have more days, months, and years of data we can sift through. We also get much more granularity, which means the results are much more precise," they said.

This is where speed comes into play and where DDN delivers the required performance. Being able to read and write data at very high speeds is important: GPUs drive huge amounts of data. "I need to be able to get that data quickly to and from the many GPUs in our system to keep them busy," they said. "With this high-performing storage, we churn through more data. No GPUs sit idle."

Use Case Two—An Interview with the Director for Research Computing, Higher Education

This university wants all of its students to be ready for the new world of AI, regardless of their field of study. The school chose DDN and NVIDIA to enable that vision. In the past few years, leveraging of AI has exploded at the university. Researchers are working with bigger data sets, which often include massive numbers of small files, and developing more tools that leverage GPUs.

> "
> ```
> It saves everyone time, and it
> allows me to take on projects
> I wouldn't have been able to
> before.
> ```
> ———————
> Director for Research Computing, Higher Education

The director for research computing said, "Gene sequencing programs, for example, create lots of little files. And any students working with statistics are creating lots of little files, too. The DDN/NVIDIA architecture accelerates all of it. If we can make those applications work ten times faster, students get their results in a couple of hours instead of putting in a month of work. We deployed this new system, and we had no more complaints."

The platform's management capabilities have definitely met the university's goals. "I have a much larger group in the engineering department who feel comfortable handling this system. It saves everyone time, and it allows me to take on projects I wouldn't have been able to before. It makes us and our users happier, and that is good."

Use Case Three—An Interview with the Head of Research Computing, Higher Education

The head of research computing at this educational institution calls its two DDN systems "high-performance workhorses for the life sciences." Life science data [sets are] large by nature and requires a lot of memory to perform analyses. Individual studies can consume tens or even hundreds of terabytes of capacity. The university needed greater simplicity, reliability, and resilience.

Leveraging the DDN/NVIDIA solution was a crucial step in moving from being a science department, in which each group had its own machines and small file stores, toward creating a single environment. IT delivers "shares" of the cluster to the various groups, allowing researchers to pool their efforts to tackle bigger problems. "We are linking together scientists in different disciplines. They all collaborate in one space now," they said. "That would not have been possible before."

The institution also can handle volumes of data that it never could before—and at great speed. "It's a centralized system that enables us to redirect compute upon request if we are in a publishing race and have to hit a deadline."

The research head reports that performance for all workload types is superlative. They have nearly 1,000 users who all focus on separate fields of science. "These are all varied workloads," they said. "Somebody is always doing something to this platform that is [data-intensive]. But DDN always stays up; it always answers the requests regardless of how many overlapping processes it must perform. It never goes quiet."

> "
> ```
> We are linking together
> scientists in different
> disciplines...That would not have
> been possible before.
> ```
> ———————
> Head of Research Computing, Higher Education

This university once thought of a parallel file system as being a very hard-to-build, fragile-to-maintain environment. But the DDN/NVIDIA platform now in place is working so well that a dedicated administrator to support it is no longer necessary.

"We still haven't felt a need to have a full-time storage person," the research computing head reported.

THE BIGGER TRUTH

Organizations using the combination of DDN's A³I solutions and NVIDIA DGX™ A100 systems are seeing AI/ML initiatives develop much faster and more cost effectively—with fewer issues.

These are proven, tested reference architectures that are easy to roll out, especially versus a traditional supercomputer that would require months of planning before even starting the process of acquiring hardware, not to mention putting it all together and tuning everything correctly.

DDN, working in tight collaboration with NVIDIA, makes it possible to implement an operational platform in weeks, not months or years. And after consolidating their AI/ML storage environments with NVIDIA and DDN, organizations begin experiencing a transformational level of value. By breaking down data silos, eliminating data movement and data preparation time performed by data scientists, and getting AI data under the control of enterprise IT, companies can scale transformative value even faster, while ensuring both data security and robust availability.

With AI/ML, leveraging the right infrastructure that can deliver the right level of scale and performance simplifies operations for the IT organization. Today's organizations need an AI infrastructure solution that will de-risk their AI growth plans. The secret to success is choosing a solution that not only begins adding value immediately, but also will scale to whatever growth levels the organization will need going forward. DDN's A³I solutions are deployed with more than a thousand NVIDIA DGX™ A100 systems, supporting production AI use cases around the world. DDN has demonstrated its ability to drive AI applications in the most complex NVIDIA DGX SuperPOD™ environments and now has POD reference architectures designed to address enterprise AI challenges at any scale.

Most importantly, DDN and NVIDIA together are alleviating huge headaches for data scientists with these robust, certified reference architecture solutions—solving the biggest business problems using AI infrastructure and data-driven techniques.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

www.esg-global.com   contact@esg-global.com   508.482.0188