IDC

ANALYST BRIEF

Sponsored by: NetApp

In the era of 'AI everywhere', flexibility at scale is a critical capability. Ensuring that all required data, regardless of type or location, can be a part of AI workflows will determine the success of new AI initiatives.

Unified Data Architectures Provide Needed Flexibility for AI Workflows

February 2024

Written by: Dave Pearson, Vice President, Worldwide Infrastructure

Introduction

An inability to adapt to changing market conditions and technology requirements has been a major stumbling block for organizations seeking to digitally transform – our reliance on massive amounts of dissimilar data deployed in organizational and technological silos led to difficult integrations, inefficiencies, and failures in governance, security, and data protection. Predictive and interpretative AI, but even more so, generative AI (GenAI) transformation has made it clear that legacy architectures are not sufficient to serve these complex workflows. Greater flexibility is required, in the form of unified data architectures.

Unified data architectures create flexibility at scale by addressing the needs of stakeholders within the enterprise and throughout AI data workflows. A unified architecture satisfies the need for access to high volumes of block, file, and object data, no matter how or where that data is being generated, processed, deployed, or stored.

AI workflows are complex and vary across stages, requiring in turn ultra-high performance, massive scale, and high IOPS or throughput. Parts of these workflows will ultimately reside both in the cloud and on premises for most mature organizations, as well as being split between private and partner-owned resources and IP. This will demand common control planes and management tools, governance, security, and data protection capabilities, and frictionless integrations with cloud and service providers. A unified data architecture provides the flexibility required to accommodate those needs, wherever an organization may be in its journey to AI transformation.

AT A GLANCE

KEY TAKEAWAYS

- As AI practitioners move along the maturity curve of AI transformation, they uncover the unique value of and demands upon their organizations' data – block, file, and object.
- Data must be available for AI initiatives whatever the type or location – not siloed, not requiring significant migration, management, or processing. A unified data architecture can provide that flexibility.
- A unified data architecture can create a flexible approach for AI practitioners to take advantage of the unique capabilities of traditional deployments as well as those of cloud providers.

Definitions

Data architecture

The collective set of models, policies, rules, and standards that govern the type of data collected, stored, and accessed by an organization. Data architecture enables the organization to collect, integrate and analyze data from a variety of sources and to place it into IT systems.

Unified data architecture

A unified data architecture should incorporate the tenets of unified, hybrid and multicloud storage into a coordinated, holistic approach to data storage, utilization, management, protection, governance, and ensures the optimal utilization of data resources for all key workloads no matter the data format, structure, access mechanism, location, or deployment model.

Unified storage

Storage infrastructure, software, and management tools with the ability to store and manage a wide variety of:

- Data formats (i.e., image, video, text, audio) in a single environment.
- Data structures (i.e., structured, unstructured, or semi-structured).
- Data access mechanisms (i.e., block, file, or object).

Hybrid storage

Storage infrastructure, software, and management tools that:

- Enable data to be placed (i.e., stored) across a variety of dissimilar IT systems, deployments, and environments.
- Provide data mobility across a variety of shared (public cloud) and dedicated (private cloud or traditional non-cloud) environments.
- Enable application aware data placement.

Multicloud storage:

A data storage paradigm that

- Enables more than one shared and/or private cloud environment for placement of data.
- Stores and manages dissimilar data in terms of formats, structures, and access mechanisms.
- Benefits from a common control plane, management interface, and data protection and security functionality, where such tools exist.

Benefits

Data is available for AI initiatives

In our December 2023 study on data requirements for AI, IDC found a growing need among end users for increased flexibility within their data environments. Respondents indicated that only structured data was instantly or quickly available for AI projects, while file and object data typically required significant processing, if it was available at all. Most respondents indicated they were still in need of improvement.

"...what we need to improve within our data architecture is the ability to handle multi-modal data..."
– VP of Data Science, Telecommunications, Canada

"...we do not do a good job of putting this data into a modern architecture and this is probably our largest challenge... Otherwise, our AI initiatives will likely fail."
– VP of Data Science, Media & Entertainment, USA

A critical capability of a unified data architecture is to provide flexibility in managing block, file, and object protocols at scale, wherever they reside. Making that data readily available to AI workstreams reduces the time to insights and therefore value, as well as the need for additional skilled resources to manage these processes. Figure 1 shows how respondents see their data types for AI use cases.

Figure 1

Path to Deploying AI Responsibly at Scale

Q. How available are each of the following types of organizational data for use in projects using AI?

	Instantly available for all AI projects	Available with minor processing to prepare	Available with significant processing to prepare	Not available to AI projects
Structured data	7 (29.2%)	11 (45.8%)	6 (25%)	0
Unstructured file data	4 (16.7%)	5 (20.8%)	12 (50%)	3 (12.5%)
Unstructured object data	2 (8.3%)	6 (25%)	11 (45.8%)	5 (20.8%)

Source: IDC, 2024

End users can take advantage of both shared and dedicated data environments

By enabling both hybrid and multicloud deployments, a unified data architecture can allow end users to take advantage of tight integration with cloud providers and native AI services from Amazon, Microsoft, and Google.

"...support for data mobility across shared and dedicated environments is crucial for our operational flexibility...this approach is particularly compatible with our strategy for AI..."
– VP of IT Ops, Manufacturing, USA

"We've realized that we that we need a multi (public) cloud environment to take advantage of some of the unique services and capabilities that each cloud has to offer to maximize the benefits of our AI models and strategy."
– VP of Data Science, Telecommunications, Canada

Increased flexibility in scaling not just capacity, but capabilities

Scaling AI data architectures can happen as needed via validated solutions from industry-leading AI partners, in both dedicated and shared environments. There are unique features valuable to infrastructure deployments on-premises, in hosted environments, and among the various cloud service providers. The greater control over data and ability to contain sensitive data, along with predictable performance and low-latency access to data within on-premises environments is critical for certain types of data and workloads, while the ability to burst and scale capacity, quickly deploy, re-deploy, or end-of-life resources makes shared cloud resources essential for flexible operations. Some respondents in our study pointed to the flexibility and high availability of resources provided in a unified architecture, while others pointed to the maximization of value from AI initiatives.

Considerations

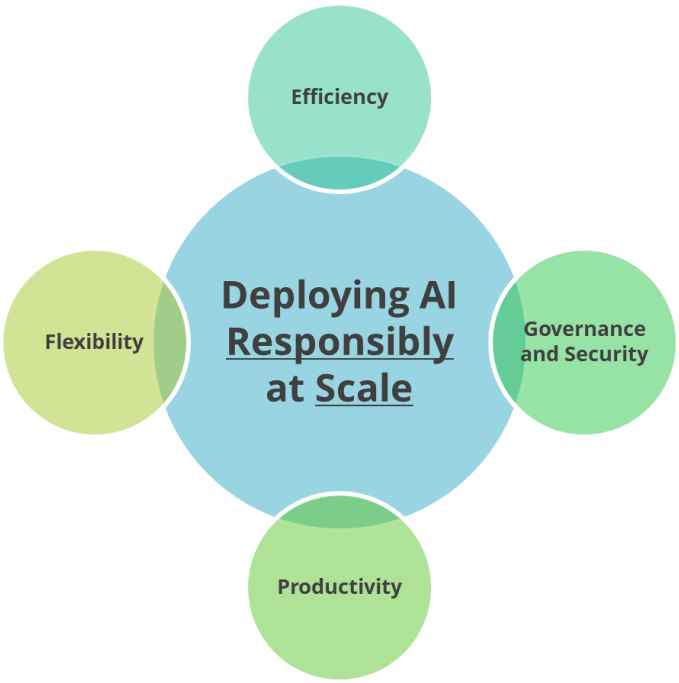
IDC encourages ITDMs to take a holistic approach to data management when deploying AI. Figure 2 shows the four pillars of any responsible and scalable AI initiative. Companies that want to transform themselves with AI must start with a unified data architecture that offers:

- Flexibility. Provides a single-stop access to corporate data for the myriad AI uses cases being implemented across the entire organization.
- Governance. Enables organizationwide enforcement of data security and governance policies, with reliable business outcomes.
- Efficiency. Enables the organization to achieve resource efficiency and sustainability objectives from their AI initiatives relative to their overall datacenter footprint.
- Productivity. Enable all teams involved in AI workflows – IT, Data Science, Data Engineers and Developers - to remain productive, thus ensuring that the desired business outcomes are achieved in a consistent, timely and accurate fashion.

For companies to scale their AI initiatives responsibly and efficiently, they must ensure that one of the design considerations for the AI project is a fit-for-purpose data storage system that enables the implementation of an integrated data pipeline that can scale securely, sustainably, and efficiently.

Figure 2

Path to Deploying AI Responsibly at Scale



Source: IDC, 2024

"...there is the potential for even more rapid change from where things are today, so there needs to be the necessary acknowledgement of that within the organization on how decision making can be altered to cater for that velocity otherwise it risks falling behind the curve."
– Director of IT Operations, Government, UK

Complexity increases when workloads rely on multiple deployment models, especially outside of datacenter. End users report multicloud environments are the most complex and require the greatest assistance from vendors and partners. Offerings need to simplify and de-risk infrastructure choices, and must not create trade-offs for flexibility – performance, scale, management and governance, security, or data protection. By maintaining a unified data architecture across the enterprise, storage infrastructure choices are simplified, and avoid point solutions and silos while enabling the flexibility and adaptability that end users indicate are so important to AI transformation success.

AI practitioners need to understand that providing a unified data architecture for their initiatives is a process, not a point in time. The transition to a modern architecture requires buy-in from a variety of stakeholders within the organization, from the C-suite to IT to finance and even legal departments. While each of these stakeholders have something to gain (operational flexibility, financial flexibility, and improved governance and compliance) those benefits need to be clearly communicated, as is the need for organizations to make decisions that can adapt to the pace of change.

Conclusion

AI transformation has quickly become a business imperative across every industry, segment, and geography. Making infrastructure decisions that enable not only the immediate demands of the organization but also provide a foundation for future initiatives that can provide the flexibility required to adapt to the velocity of change. The requirements of cutting-edge AI workloads (predictive, interpretive, and generative) the various stakeholders involved in AI initiatives are dynamic, and a unified architecture can deliver that required flexibility.

IDC suggests that end users maintain an enterprise-wide view of the data required for AI initiatives and utilize a unified data architecture with a multicloud approach to deployments in support of these new applications and workflows. Beyond the "speeds and feeds" that address the purely technical demands of the workloads, managing data governance, privacy, compliance, security, and data protection are all areas in which inappropriate deployment choices can limit organizational capabilities and reduce the flexibility of the enterprise in the face of this rapidly evolving and developing market.

Making infrastructure decisions that enable not only the immediate demands of the organization but also provide a foundation for future initiatives that can provide the flexibility required to adapt to the velocity of change.

About the Analysts

Dave Pearson, *Research Vice President, Worldwide Infrastructure*

Dave Pearson is Research Vice President for Storage and Converged Systems practice within IDC's worldwide infrastructure research organization. He also oversees IDC Canada's Infrastructure Solutions research practice. He and his team are responsible for IDC's storage, integrated, hyperconverged and composable systems and platforms. This includes storage for performance intensive use cases like high-performance computing, artificial intelligence, and analytics

IDC Custom Solutions

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com

The content in this paper was adapted from existing IDC research published on www.idc.com.