A **CENTER FOR DIGITAL EDUCATION** THOUGHT LEADERSHIP PAPER

Building an Intelligent Data Infrastructure for Higher Education

SPONSORED BY NetApp

Data has become one of the most valuable assets in higher education. When institutions unlock their data, they can improve student outcomes, support groundbreaking research, enhance business operations and make better decisions.

Data has also become more challenging to manage and protect as data volumes increase, needs shift, and security and compliance requirements evolve. Data silos and outdated procurement approaches have exacerbated storage costs and complexity. Data is usually fragmented and spread across on-premises and cloud environments, making it difficult to access for learning, analytics, workflow automation, decision-making and AI-powered applications.

"Higher education organizations need to understand the data in their environment better now than in the past," says Bob Burwell, chief technology officer for state and local government and education at NetApp. "Their environment is not only on premises. It may be in the cloud or edge devices. University researchers may share data with each other in different ways."

To address these issues, higher education institutions need intelligent data infrastructure.

"If you want to be an agile institution, you must ensure your data is not going to slow you down," says Brian Cohen, vice president of the Center for Digital Education. "You don't want it in a format or structure that makes it even harder for you to pivot when you need to."

Data is an incredibly valuable asset for higher education institutions. It's also difficult to manage, access and protect across various environments.

Keeping up with the future

Higher education leaders must address these data challenges:

Massive volumes of structured and unstructured data. Research-focused universities manage petabytes of critical research information, which may include lab notes, geographic information system data, charts, reports and more. Colleges and departments that focus on teaching and learning must be able to use data from multiple systems to enhance the student experience on and off campus. Outdated storage systems, data silos and complex compliance requirements are ongoing hurdles.

"Higher education leaders now have to think about how to move from a data warehouse approach — which isn't ideal for dynamic or unstructured data — to a data lake approach," Cohen says. "Data lakes bring all these different types of structured and unstructured data together from multiple sources, taking higher education to the next step."

Sensitive data. As institutions make their student, research and operational data more accessible for decision-making, research, AI applications and other uses, they must protect sensitive data by implementing policies, governance and controls. Doing that across on-premises, cloud and edge environments is complicated. Institutions must also defend against ransomware and other threats to data availability and integrity.

Shortage of data center space. Many institutions struggle to keep up with storage and data center requirements. As early adopters in individual departments embrace data analytics and AI, they can quickly consume data center space. Institutions need flexible solutions that optimize storage while maintaining high performance.

Tight budgets and high costs. Amid declining enrollments, increasing expenses and state funding cuts, higher education leaders are being asked to do more on a tighter budget. Modern solutions reduce data storage and management costs.

Colleges and departments must be able to use **data from multiple systems** to enhance the student experience on and off campus.

Why you need intelligent data infrastructure

An intelligent data infrastructure uses AI, analytics and software to help you more easily understand, manage, optimize and protect data. It helps you realize the full potential of your data by unifying data storage, simplifying control and management of data services, and building in data protection.

Unifying data storage

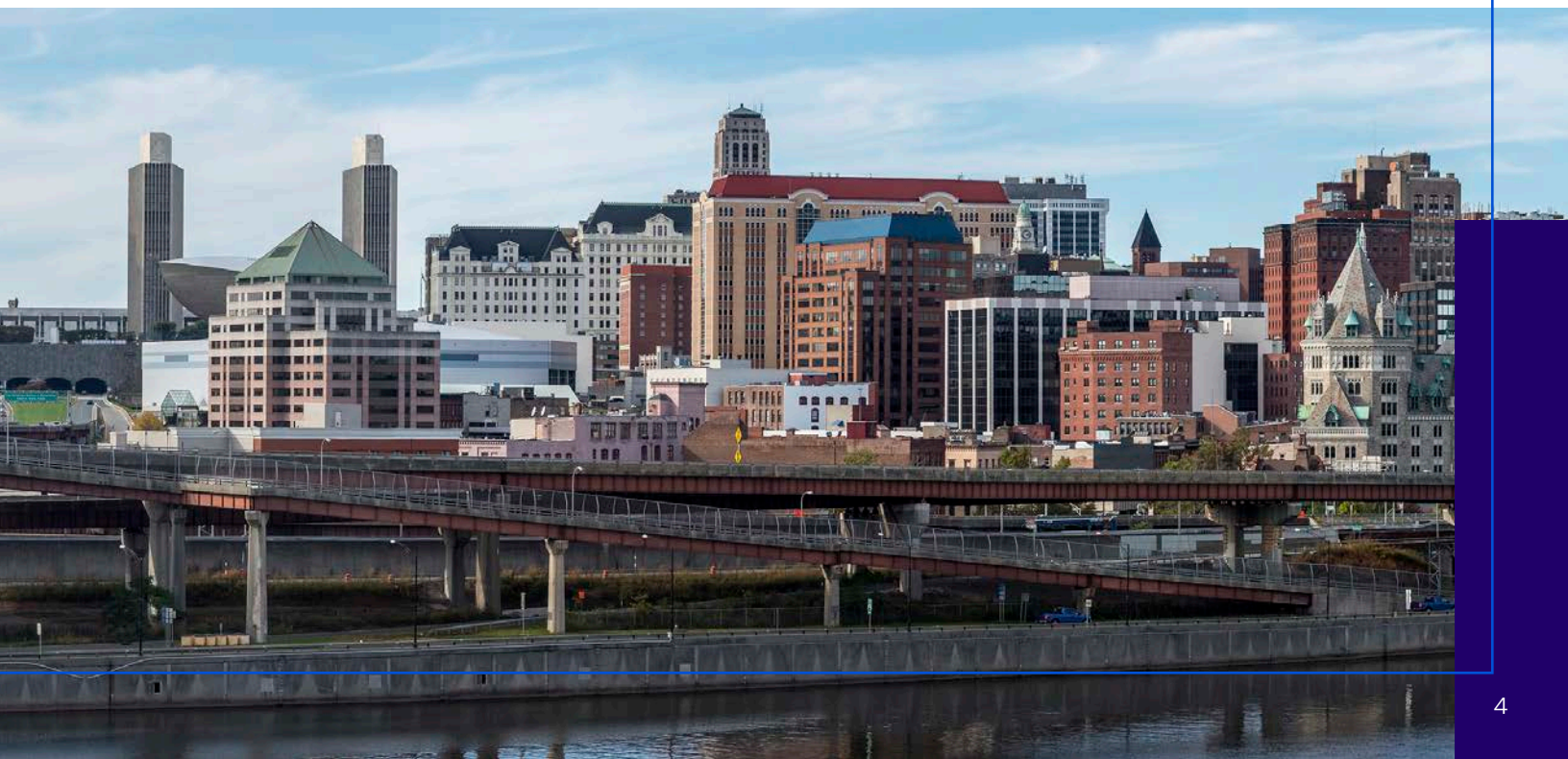
Higher education institutions typically store data in multiple databases, each designed for a specific data type or purpose and each requiring a specific set of protocols to access them.

"Data needs to be agile, so eliminate siloed solutions and wrap classification and governance around your data by deploying tools like NetApp's BlueXP," Burwell says. "This ensures that you understand where personal information may be stored and that you have policies in place to protect it."

As a case in point, Texas A&M University adopted NetApp Keystone storage-as-a-service to accelerate its transition from a siloed environment, where every college or business unit had its own budget and staff, to centralized IT. The university also uses automation to provide self-service access to IT resources with increased efficiency, improved quality, greater flexibility and faster time to deployment.

Similarly, Syracuse University modernized its infrastructure with NetApp AFF. This lowered storage costs for massive, growing data sets while improving school operations and empowering academic research. The university was able to pull more than one petabyte of research data into its data center for greater performance, reliability and availability. Dynamic data tiering minimizes costs for research data that is rarely or never used. Unified storage provides flexibility to expand into the cloud or seamlessly support file-based AI workloads alongside the university's critical block workloads.

An intelligent data infrastructure **lowers data storage costs** and leads to improved school operations and academic research.



"All of a sudden, what was done on a USB drive in the faculty member's office was now in a data center running on high-performance drives, easily accessible for anyone who needs it, at any time," says Eric Sedore, chief technology officer for Syracuse University.[1]

Unified data storage lets you access data anywhere, regardless of protocols. This capability is especially important for taking advantage of AI capabilities and large language models (LLMs) that reside mainly in large-scale cloud environments but must access data that's stored on premises.

Yale New Haven Health in Connecticut built an AI environment — a unified data lake — on NetApp that lets users securely access LLMs to extract data from clinical notes and accelerate research. By integrating NetApp data management tools (the DataOps toolkit) with NVIDIA's NeMo Retriever (a platform for developing custom generative AI), researchers are enhancing clinical and research data management for LLM training and deployment.

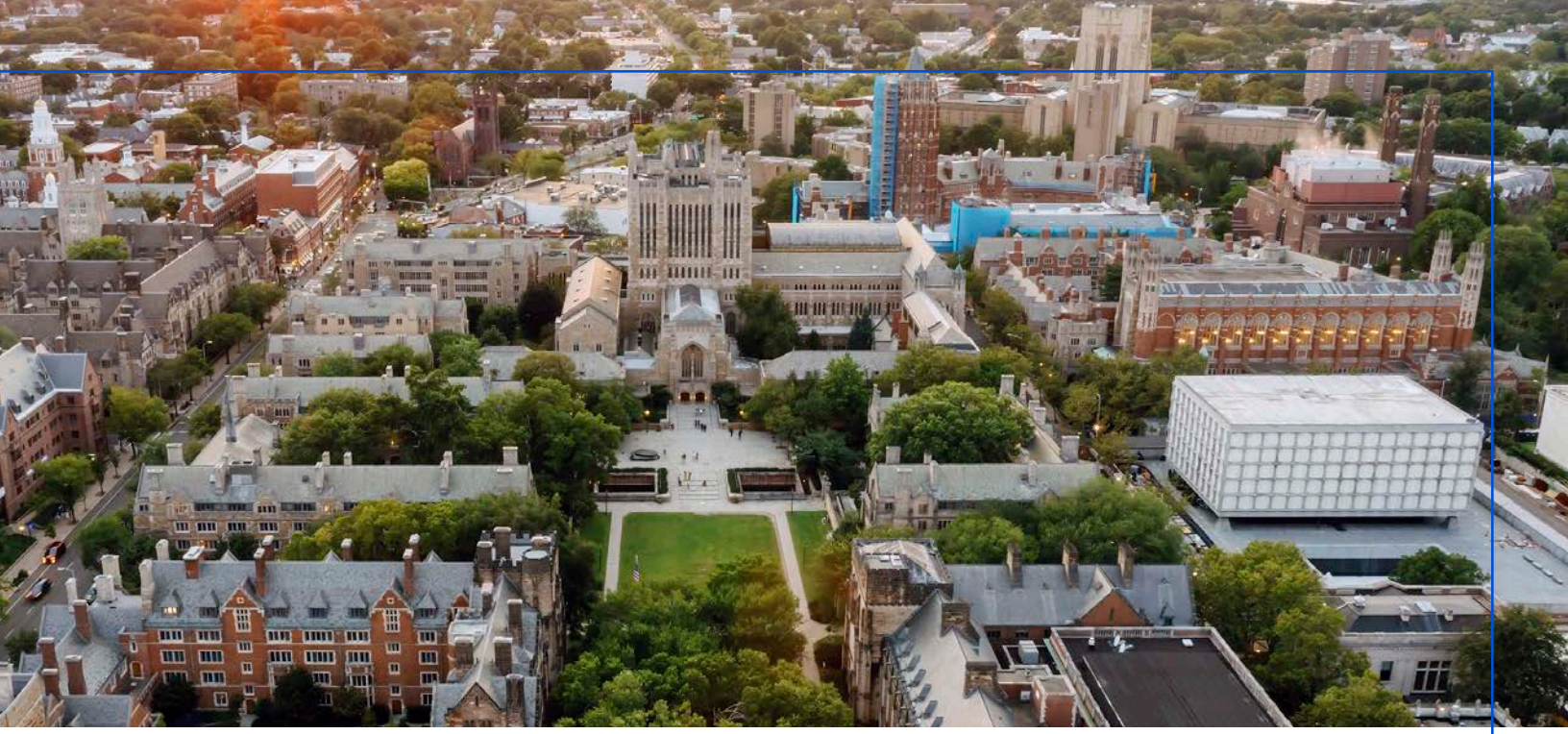
"NetApp was selected because of its price/performance, scalability and the availability of tools that facilitate the transfer of data between platforms," says Wade Schulz, assistant professor of laboratory medicine and computational healthcare researcher at Yale School of Medicine.[2]

In another use case, the University at Albany in New York adopted unified storage for its AI Plus initiative, which will significantly expand the AI supercomputing resources available in the state for teaching and research. The university will use NetApp all-flash storage systems to deliver ultra-high performance to its NVIDIA DGX Clusters, which are purpose-built for AI workloads. This setup offers an enterprise-ready, multi-tenant environment on premises with the potential of having both on-premises and cloud tools available for future growth.

Simplifying control and management

A unified control plane simplifies storage management for your team and makes management and data governance more consistent across disparate environments. It provides a single pane of visibility and a single set of tools for managing storage, data services, licenses and subscriptions across your entire on-premises and cloud environment. You can copy, move or synchronize data between a broad range of file and object storage types and locations with a simple drag-and-drop action.

A unified control plane makes **data governance more consistent** across disparate environments.



The University of Utah Health adopted NetApp BlueXP for simplified control and management of its new hybrid cloud environment. This solution protects personally identifiable information and ensures compliance with Payment Card Industry Data Security Standards (PCI-DSS), HIPAA and other regulations.

Unified control also allows you to continually optimize compute and storage for a specific workload, use case or business requirement. This can reduce costs and improve performance. For example, the Laboratory for Atmospheric and Space Physics (LASP) at the University of Colorado Boulder[3] relies on NetApp AFF and StorageGRID to save 50% on storage costs for massive and growing data sets that support 130 scientists who do ongoing research on everything from climate change to solar physics.

In another example, Western Oregon University recently deployed a NetApp C800 all-flash, capacity-optimized storage system. The university was able to replace three racks of spinning-disk and AFF storage with just half a rack. The smaller footprint will generate significant savings from a power, space and cooling perspective, supporting the university's green initiatives — without sacrificing performance.

Building in data protection

Higher education institutions are under constant attack. Leaders need to make sure ransomware protection, Zero-Trust capabilities and other security measures are built into their data and infrastructure. And if something bad does happen — as when a faulty software patch crashes systems around the world — organizations must be able to recover quickly.

Look for built-in autonomous ransomware protection (ARP), which uses machine learning to detect ransomware automatically and in real time. ARP not only identifies ransomware and prevents its propagation to other locations but also helps institutions speed up their recovery. Within seconds of detecting ransomware, ARP locks data with immutable and indelible snapshots, which means the data cannot be deleted or changed.

Yale University uses layers of protection enabled by NetApp for defense in depth. Data services like Autonomous Ransomware Protection and immutable snapshots have enabled Yale to recover from viruses and ransomware in minutes. The university also implemented a fail-proof disaster recovery strategy for its VMware infrastructure

Data protection services have enabled **Yale University to recover from viruses** and ransomware in minutes.

with a NetApp data replication technology called SnapMirror. The tool helped the university recover seamlessly from the 2024 CrowdStrike outage.

Institutions should also look for storage platforms that use machine learning to automatically classify compliance-related data sets or other sensitive data types stored on the platform. For instance, Western Oregon University implemented a NetApp C800 storage appliance with self-encrypting non-volatile memory express (NVMe) drives. This simplifies compliance with the Gramm-Leach-Bliley Act and other standards and regulations in higher education.

Assessing your readiness

Higher education leaders should ask the following questions to assess their data infrastructure's readiness for the innovation of today and tomorrow.

- **What are we trying to solve?** What are our strategic goals, and why do they matter? Who's driving this initiative, and who will it serve? How do we measure success?
- **Does our data infrastructure design align to our strategic goals?** Is it flexible and scalable? Are we building on a proprietary background, or will we be able to bring in future innovations, including AI and quantum computing?
- **Have we considered our users' goals?** What sources of data will we need to include?
- **Do we understand our data and where it lives?** Are there more cost-effective ways of storing it? Which data do we need to protect? Who or what needs to access it?
- **Are we following interoperability standards?** Are we following our data protocols? Are they open data protocols? Do we have clear policies for data governance, security and privacy?
- **Have we defined the ethical use of data for AI?** How are we using data and for what purposes?
- **How will we bring non-IT users on board?** Are we talking to business users, subject matter experts and end users to understand data principles and issues?

Getting it right

Building intelligent data infrastructure is an iterative process. No institution will get it right the first, second or third time. It's important to have a continuous feedback loop from end users and industry experts to continuously evolve your data strategy.

"Any institution that builds a data strategy today should get industry expertise involved to understand the roadmap and where technology is going," says Cohen. "They should make investments that are sized appropriately, with an understanding that they'll need to pivot and grow that investment, so they'll want something that is interoperable and not proprietary."

1. https://www.netapp.com/pdf.html?item=/media/28168-cs-Syracuse.pdf
2. https://www.netapp.com/blog/how-yale-new-haven-health-transformed-data-infrastructure/
3. https://www.netapp.com/pdf.html?item=/media/28426-cs-CU-Boulder.pdf

This piece was written and produced by the Center for Digital Education Content Studio, with information and input from NetApp.

CENTER FOR DIGITAL EDUCATION

Produced by the Center for Digital Education
The Center for Digital Education is a national research and advisory institute specializing in K-12 and higher education technology trends, policy and funding. The Center provides education and industry leaders with decision support and actionable insight to help effectively incorporate new technologies in the 21st century.

www.centerdigitaled.com

NetApp

Sponsored by the NetApp
NetApp is the intelligent data infrastructure company, combining unified data storage, integrated data services, and CloudOps solutions to turn a world of disruption into opportunity for every customer.

Learn more at **www.netapp.com** or follow us on X, LinkedIn, Facebook, and Instagram.