IDC

ANALYST BRIEF

Sponsored by: Pure Storage

The life science industry is on a journey to bring new drugs to the market faster. Flash arrays offer an effective solution to help the industry cope with the dual challenges of exploding data volumes and the need to accelerate innovation.

The New AI-Enabled Life Science Stack: Building Scalable, Smart, and Secure Storage Solutions with Flash Arrays

June 2025

Written by: Dr. Nimita Limaye, Research Vice President, Life Science R&D Strategy and Technology

Dealing with the Data Deluge

Data has been the fundamental unit of innovation for the life science industry. While patient centricity is top of mind, the life science industry has become a very data-centric world. As the focus on precision medicine and personalized care continues to scale and as the focus on cell and gene therapies continues to grow, there is a need to bring together complex and heavy workloads. These include multiomic data, including genomic, proteomic, epigenomic, and transcriptomic; imaging data, including high-resolution microscopy, digital pathology, and other clinical images; and vast clinical trials data sets. Real-world data, including structured and nonstructured data, claims data, and electronic health record data, is also being leveraged to accelerate drug approvals and extend labeling claims.

Today's life science industry faces the challenge of managing real-time access to massive and highly sensitive data in a compliant and cost-effective manner. Overseeing and storing vast amounts of data can be expensive, requiring significant investment in storage infrastructure and technologies.

Flash arrays offer an effective solution to help the industry cope with these and other challenges.

AT A GLANCE

KEY STATS

- 50% of healthcare and life science (HLS) companies identified optimization of AI infrastructure — including compute, storage, network, and cloud resources — as their top 2025 priority to secure AI workloads cost effectively. This will allow organizations to scale AI's adoption across the entire business and IT stack (source: IDC's *Future Enterprise Resiliency and Spending Survey,* February 2025).
- 44% of the HLS industry considered increasing the level of automated IT infrastructure operations as the most important infrastructure modernization goals for the past two years (source: IDC's *Future Enterprise Resiliency and Spending Survey,* April 2025).

KEY TAKEAWAY

The healthcare and life science industries are seeking AI-optimized, low-latency, cost-effective storage infrastructure solutions that will provide flexibility and scalability to accelerate innovation.

Definitions

- **Flash memory:** This is a type of nonvolatile computer storage that can be electrically erased and reprogrammed.
- **Nonvolatile memory:** Flash memory retains data even when the power is turned off, unlike RAM (random-access memory).
- **All-flash array (AFA):** An AFA is a storage infrastructure that contains only flash memory drives instead of spinning-disk drives. It is an enterprise-grade data storage system that utilizes only flash memory (solid state drives) for data storage, unlike traditional systems that use hard disk drives (HDDs) or a combination (hybrid arrays).

Laying Out an Array of Benefits for Life Sciences

Flash arrays benefit the life science industry in the following ways:

- **Compounded benefits of high-performance computing (HPC) and fast data access:** As generative AI (GenAI) gains importance, the combination of fast data access offered by flash arrays and HPC is needed to rapidly train large AI models used in genomics and drug discovery, such as those for protein structure prediction, molecular docking, and virtual screening.
- **Modular architecture and scalable storage:** The life science industry faces critical data storage challenges in the areas of drug discovery and genomics research. Flexible and scalable solutions that accommodate increasing data volumes with a pay-as-you-go model are essential to address R&D's evolving demands. The modular design architecture can help organizations rightsize their infrastructure to meet specific needs and manage costs more effectively.
- **Business continuity:** Evergreen architecture with nondisruptive upgrades is critical to avoid disruptions. Even a single day's delay in launching a drug can cost millions of dollars.
- **Speed to insights:** The healthcare and life science organizations need to accelerate modeling for genomics, drug discovery pipelines, point-of-care decisions for patients, and precision care. Flash arrays offer significantly faster read/write speeds and lower latency compared with traditional hard disk drive storage.
- **Lower power consumption:** Unlike HDDs, AFAs have no moving parts, so no energy needs to be spent on running them. Reduced cooling requirements significantly lower power consumption needs.
- **Smaller datacenter footprint:** AFAs can achieve higher storage density within a smaller physical footprint compared with HDDs, thus saving space.
- **Deep storage to scale cost efficiencies:** The life science industry requires cost-effective solutions that strike a balance between the need for rapid access to critical data and the need for the long-term archival of less frequently accessed data. Deep storage solutions can offer more cost-effective solutions for storing less frequently accessed data, such as raw sequencing data generated by next-generation sequencing (NGS) platforms, vast libraries of chemical compounds, experimental data, preclinical data from animal models, and clinical trial results data to ensure regulatory compliance.

- **Automation of workload management:** The life science industry needs to invest its time and resources on scaling innovation. Solutions that can automate workload placement, provisioning, and scaling can empower developers to provision and manage storage resources programmatically, without requiring manual intervention from IT administrators.
- **A single pane of glass approach:** Having a unified platform that integrates data from many sources can simplify operations and reduce complexity by eliminating the need to manage each array individually.
- **Optimization using an AI lens:** In today's day and age, it's all about AI. Leveraging AI to optimize storage performance, and capacity planning, and to enable anomaly detection to proactively preempt cybersecurity concerns seems like a no-brainer. And the life science industry is looking for these solutions.

Considerations

The need to accelerate innovation despite major shake-ups caused by geopolitical turmoil, tariffs, and drug price controls has put significant cost pressure on the life science industry to do more with less. Data powers innovation, and the industry is dealing with a data tsunami. Data is flowing in all shapes, sizes, and forms. There is a need to store the data cost effectively and access it when needed. However, data storage infrastructure can be very expensive. Therefore, modular solutions that can provide tiered access to data — enabling faster access to critical data while lowering costs for accessing heavy workloads that are not needed on a daily basis — can address some of these problems. Automating the governance and planning of data storage solutions by leveraging AI also helps significantly reduce storage overheads.

While AFAs may require a higher initial investment, in the long term, the total cost of ownership (TCO) could potentially be lower, so organizations need to analyze this option carefully. Integrating new flash arrays with legacy systems and existing IT infrastructure can be complex and require careful planning. Migrating large volumes of data from existing storage systems to flash arrays can also be complex and time-consuming. Due consideration needs to be given to all of these factors when transitioning to using AFAs.

Conclusion

Leveraging flash arrays and AI-enabled, optimized storage infrastructure to deal with data-heavy workloads is imperative for the life science industry. Flexibility, scalability, deep storage capabilities, modular and evergreen architecture, and automated workload management are the ask of an industry that is struggling to carry the burden of escalating data storage costs. While flash arrays could enable the life science industry to address these challenges, organizations must carefully examine the up-front investments needed and the data migration and system integration challenges.

Flash arrays can silently transform massive petabytes of biological noise into life-saving innovations, powering drug discovery and accelerating time to market.

About the Analyst

Dr. Nimita Limaye, *Research Vice President, Life Science R&D Strategy and Technology*

Dr. Nimita Limaye provides research-based advisory and consulting services as well as market analysis on key topics related to R&D strategy and technology in the life science industry. Her research focuses extensively on AI/GenAI. She is the recipient of the 2024 DIA Global Inspire award and is the past chair of the board of SCDM.

MESSAGE FROM THE SPONSOR

Pure Storage provides an efficient, reliable, and agile data storage platform for the life sciences industry to optimize genomics workflows, expedite drug discovery, and AI-initiative readiness and adoption at scale. Our platform delivers cost savings for all of your data needs to improve care outcomes, maximize operational efficiency, and drive clinical innovation.

Learn how Pure Storage can help scale your storage needs.

IDC Custom Solutions

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

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
blogs.idc.com
www.idc.com