ANALYST CONNECTION

Sponsored by: Pure Storage

Flash arrays represent the data storage solution of choice for the increasingly data-intensive workloads, real-time insights, and AI-powered, GxP-compliant, and tiered data storage needs that will power precision medicine in the life science industry.

Addressing Critical Data Storage Challenges for the Life Science Industry: Intelligent Flash Arrays

July 2025

Questions posed by: Pure Storage

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

Q. What are the data storage challenges facing the life science industry today, especially in drug discovery and genomics research?

A. With the increasing focus on precision medicine and the development of cell and gene therapies, there is an ongoing effort to bring together massive data sets of multiomic (i.e., genomic, proteomic, transcriptomic, metabolomic, and epigenomic), imaging, and unstructured data (e.g., lab notes, imaging data, 3D models) to accelerate drug discovery and drive the innovation of therapies for key disease areas, such as oncology and rare diseases. A critical need exists for high-performance storage platforms with a modular and flexible architecture that can enable organizations to rightsize their infrastructure to meet growing and fluctuating R&D needs for data storage and to manage costs effectively. The storage platform should ensure data integrity, as this is crucial for the reproducibility and credibility of scientific research and to accelerate time to market. Not all data needs to be accessed at the same time; however, storing data for long durations and being able to pull it up as needed is essential for the life science industry to ensure compliance. On the other hand, easy, real-time access across the globe for teams that are collaborating to develop new drugs is essential to maintain a rapid pace of innovation. Hence, a tiered, flexible, and scalable platform is a must.

Q. What are the alternatives to flash arrays for data storage in the life science industry and why are flash arrays better?

A. The modular evergreen architecture, deep storage capability, smaller datacenter footprint, lower power consumption, low latency, and high performance that flash arrays offer serve to power the fast analysis of data-intensive, compute-heavy, and increasingly AI-driven workloads in areas such as drug discovery and genomics, making them the data storage solution of choice for the life science industry. Flash arrays offer high speeds and very low latency, which are critical for high-throughput applications such as genome sequencing, AI-driven drug discovery, and real-time diagnostics. High data processing speeds power big data analytics, which are key when working with petabyte-scale genomic data sets or heavy imaging workloads for digital pathology. High performance and fast access are crucial for model training and inference.

While hard disk drives (HDDs) represent a more widely adopted technology, cost less, and are more suitable for cold or infrequently accessed data, they have slower performance, higher latency, and higher power and cooling needs. Flash arrays offer higher-density storage in smaller footprints, with lower cooling and power requirements. Fewer moving parts decrease the likelihood of mechanical failure rates, driving data integrity and good industry practices (GxP)–compliant storage in regulated environments, making the performance of flash arrays more predictable than HDDs.

Network-attached storage (scalable file-level storage accessible over a network) is not optimized for high-throughput analytics and low-latency workloads, which are increasingly important for the life science industry. Storage area networks with HDDs or hybrid drives (flash + HDD) provide enterprise-level, reliable, and scalable centralized storage across multiple servers. However, these can be costly and complex and would still be slower than all-flash for high-performance workloads. Cloud-based storage solutions are highly scalable and well suited for data archival and collaboration and offer a pay-as-you-go model; however, organizations may still need to address latency issues, ongoing operational costs, and compliance and data residency concerns.

Q. Why is deep storage so important for the life science industry?

A. Modular solutions that provide tiered access to data are the need of the hour. They enable faster access to critical data and lower the costs of accessing heavy workloads that are not needed on a daily basis. Examples include the raw sequencing data generated by next-generation sequencing (NGS) platforms, vast libraries of chemical compounds, or massive data sets of preclinical data, including data from animal models, in vitro experiments, and high-throughput screening. Data archiving must ensure compliance, provide easy access to data during audits and inspections, and allow access to critical data as and when needed to rapidly fuel innovation while costing less for storing data that does not need to be accessed regularly.

Intelligent deep data storage solutions are the answer for storing data that organizations need to retain for compliance or long-term analysis ("cold" or "deep" data) while not increasing the latency of access to important data that is used in day-to-day operations ("hot" or "warm" data).

These deep data storage platforms increase the durability and accessibility of deep data.

Q. Are the data explosion and AI revolution fueling the demand for scaling data storage in the life science industry?

A. A massive data explosion is occurring in the life science industry, with the volume of data generated increasing exponentially from 743EB in 2024 to 3,242EB in 2028, reflecting a CAGR of 44.5% (see *Worldwide Enterprise Global DataSphere by Vertical Industry Forecast, 2024–2028*, IDC #US52712424, November 2024). While data has been available for a while, the power that GenAI brings to the table to rapidly generate insights from these massive data sets and the ability of AI agents to transform genomic workflows have exponentially scaled the importance of data for the life science industry. This is evident from the fact that 56% of the life science industry has reported that its top GenAI use case in production in R&D is drug discovery (source: IDC's *Life Sciences Generative AI Survey,* August 2024), and a fourth of the

life science industry is already investing significantly in agentic AI (see *Critical Advice Sought from IT Leadership Regarding the Impact of Agentic AI on Healthcare and Life Sciences*, IDC #US53145625, January 2025).

The data explosion and AI revolution are working in tandem to fuel the demand for intelligent, compliant, and high-performance data solutions in the life science industry. Thus, it is not surprising that a third of the healthcare and life science industries believe they will need to significantly increase their storage capacity to achieve their goals for using agentic AI in business processes. AI can help forecast future capacity needs, optimize resource allocation, detect anomalies, predict potential failures, provide recommendations for enhancing storage security, and identify potential ransomware threats. It can also efficiently categorize data as cold versus warm, helping quickly move cold data into the deeper recesses of the storage infrastructure while pulling it out swiftly when needed.

Q. How do you handle data backups and disaster recovery to minimize the risk of data loss and ensure business continuity?

A. Although the life science industry has always been known to be a "patient centric" industry, with the force of AI powering transformation, "data centricity" has taken center stage. The life science industry powers innovation, and ensuring the retention of sensitive data, which constitutes its core intellectual property, is absolutely critical. Hence, it is essential to have robust disaster recovery systems in place that ensure the continuity and resilience of operations and address the potential impact of disruptions. Flash arrays play a key role in moving the needle from "reactivity" to "resiliency."

Flash arrays play a significant role in enhancing disaster recovery capabilities, enabling the efficient storage and quick retrieval of large data sets, which is vital for disaster recovery scenarios. They offer high-performance infrastructure, providing high data transfer rates, fast processing speeds, and the rapid retrieval of all data, including "tier 0" data (critical data that needs immediate recovery) in a multistage backup process. They also offer low recovery time objectives, the time by which a system, application, or business process must be restored after a disruptive event. In addition, cloud-based disaster recovery–as-a-service (DRaaS) solutions eliminate the need for organizations to maintain a dedicated secondary disaster recovery site, thus preventing high costs and paying only for the resources that they need.

Cost-efficient disaster recovery systems will serve as the backbone of the innovative life science industry, powering breakthroughs in diagnostics and personalized medicine.

About the Analyst

Dr. Nimita Limaye, *Research Vice President, Life Sciences 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 at https://www.purestorage.com/solutions/industries/life-sciences.html

IDC Custom Solutions

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

IDC Custom Solutions produced this publication. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. This IDC material is licensed for external use, and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.