OPTIMISE DATA CENTER STORAGE WITH PCIE GEN5 SSDS

FOR AI AND BIG DATA

Kingston TECHNOLOGY
BUILT ON COMMITMENT

Foreword and contents

As the world speeds toward a future defined by data, the spotlight turns to the infrastructure supporting it. From generative AI to real-time analytics, the demand for speed, reliability and efficiency is accelerating – and with it, the need for next-generation data center storage. As a result, solid-state drives (SSDs), particularly those built on PCIe Gen5, are fast becoming the backbone of modern IT environments.

But what's driving this shift? How do performance bottlenecks impact GPU utilisation? Why is energy-efficient enterprise storage as critical as IOPS? And how can IT leaders protect data integrity while future-proofing their infrastructure for what's next?

This eBook answers these questions and explores how to improve data center storage performance and key industry trends, from the rising cost of downtime to the promise of sustainable infrastructure. With insights from Kingston flash storage experts, discover how advanced SSDs are transforming data centers for the AI era and beyond.

Table of contents	Pages

Contributors

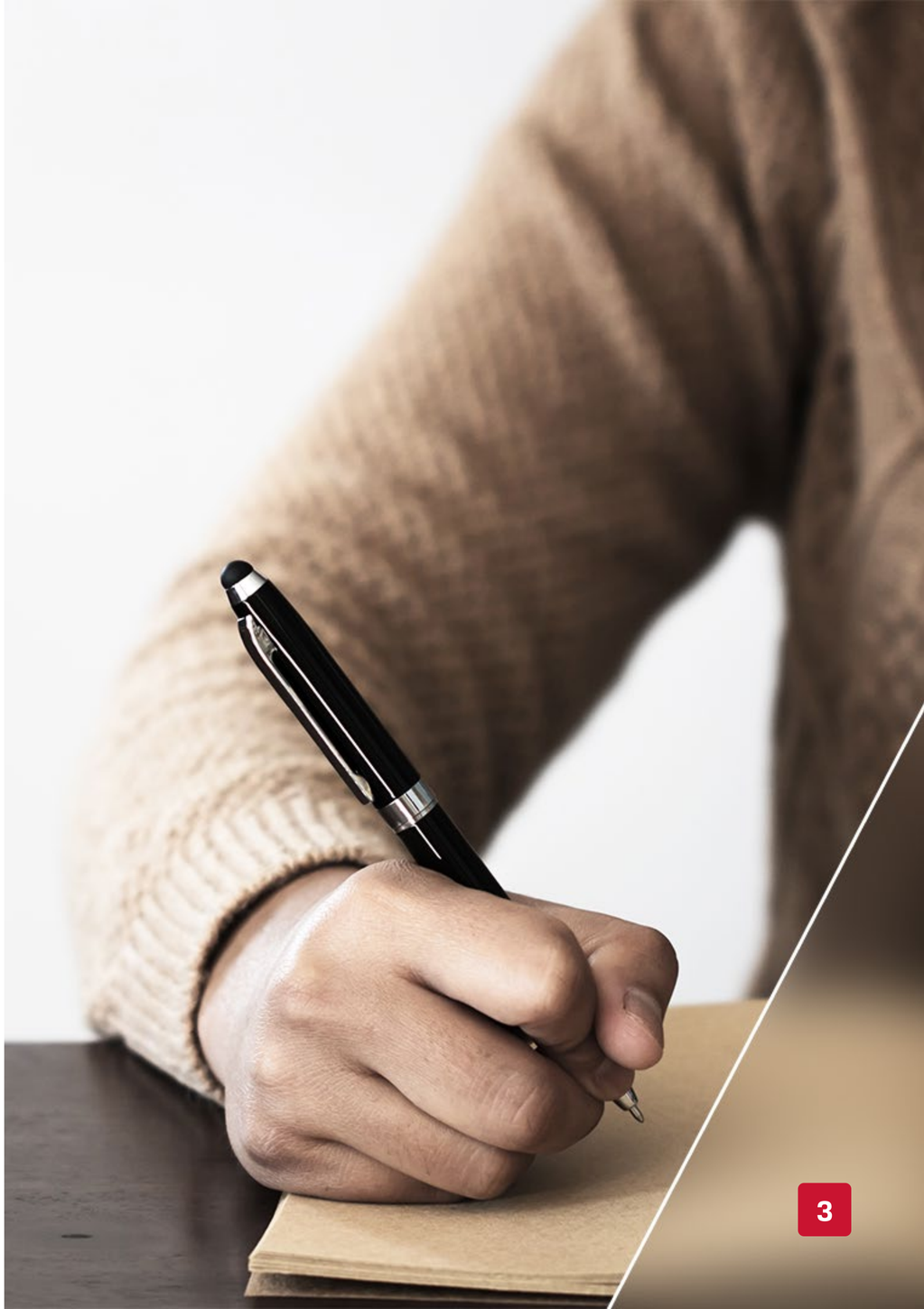
This eBook has been created by two Kingston experts in storage technologies.

Louis Kaneshiro | Kingston Technology

Louis Kaneshiro is the Senior Technology Manager at Kingston Technology. Over his 30 years at Kingston, the last 15 years focused on SSD, he led the Technology Resource Group, now a global team, before launching the SSD Product Engineering Department (SPED). Prior to his current role, Louis served in technical support and as an FAE for Kingston's OEM division.

Tony Hollingsbee | Kingston Technology Europe

Tony has been with Kingston Technology for 23 years, having held various sales roles, including Inside Sales Manager for the company's Flash Memory and SSD product lines. For the past 12 years, he has served as SSD Business Manager for the EMEA region, working closely with both sales and marketing teams to support business development and drive regional growth.



SSDs like the DC3000ME are pivotal in addressing emerging computational trends, particularly in AI, machine learning and edge computing. Our design anticipates the need for higher bandwidth, lower latency and more energy-efficient storage solutions.

Tony Hollingsbee | Kingston Technology Europe

The rising tide of data

The exponential amount of data created and consumed globally is projected to reach over 394 zettabytes by 2028[1]. Much of this data explosion is fuelled by advancements in artificial intelligence (AI), the Internet of Things (IoT) and digital media consumption. AI and machine learning applications generate and consume vast datasets, while emerging technologies like autonomous vehicles and smart cities are driving real-time data generation at an unprecedented scale.

These trends require low-latency, high-throughput storage systems. Enterprise workloads, cloud-native applications and regulatory compliance also contribute to the increasing demand for scalable, secure storage. To accommodate this surge, data centers are evolving their storage infrastructures. The emergence of the Enterprise and Data Center Standard Form Factor (EDSFF) is gaining traction. And innovative storage technologies, such as ceramic-based archival solutions with data retention capabilities exceeding 5,000 years, are being developed to meet the demands of long-term data preservation.

While the current data center storage form factor landscape is wide and varied, traditional form factors like U.2 SSDs, known for their PCIe interface and suitability in performance-centric environments, have been widely adopted. This is especially true amongst the server OEMs and server chassis makers. These advancements collectively aim to support the increasing storage requirements of modern data centers in the face of relentless data proliferation.

While E1.x and E3.x SSDs are gaining in support, the U.2 form factor is far and away the dominant form factor with over 60% of the petabytes in servers residing in a U.2 SSD.

Louis Kaneshiro | Kingston Technology

1. Volume of data/information created, captured, copied and consumed worldwide from 2010 to 2023, with forecasts from 2024 to 2028 (in zettabytes). https://www.statista.com/statistics/871513/worldwide-data-created/

> *With AI at the forefront now, the need for higher performance storage to maximise GPU utilisation is more important than ever.*
>
> **Louis Kaneshiro | Kingston Technology**

Performance bottlenecks and the need for speed

As data volumes soar and applications become more complex, data center managers face mounting challenges. One of the most persistent issues is performance variability. Many enterprise data center storage solutions fail to deliver consistent throughput and latency, leading to inefficiencies in application performance and resource utilisation.

Complex deployments further complicate this landscape, often requiring custom configurations, firmware tuning and intricate orchestration to meet workload-specific needs. Additionally, managing drive longevity is a growing concern. Especially as high-write workloads like logging, caching and AI training place immense stress on SSDs, accelerating wear and risking downtime.

AI workloads rely heavily on GPUs for computation, but unless the underlying storage can keep up, GPU utilisation plummets, leading to wasted investments and missed performance targets.

High-performance SSDs for AI workloads with ultra-low latency and high throughput are critical in sustaining data pipelines that feed GPUs in real time.

Adding to these challenges, the need for speed in data centers has never been greater. Modern workloads, particularly AI, machine learning and real-time analytics, demand instant access to vast amounts of data. Any latency in the storage layer can stall processing, delay insights and reduce overall system efficiency. High-speed storage is no longer a luxury; it's a necessity for high-performance computing resources like GPUs and TPUs.

The cost of downtime and latency

Downtime and latency are further concerns for IT leaders, with the potential for substantial financial loss and operational disruption. Recent studies report that the cost of hourly downtime exceeds $300,000 for 90% of firms, with 41% saying hourly downtime costs $1 million to over $5 million[2].

Latency further compounds these challenges by hindering data access speeds, which is particularly detrimental in data-intensive environments. Delays in data retrieval degrade user experiences, reduce productivity and lead to revenue losses. In high-frequency trading, for example, latency is a critical factor where even a few milliseconds can determine millions in profit or loss.

Advanced SSDs offer a compelling solution to these issues. Leveraging technologies such as PCIe Gen5 interfaces and advanced 3D NAND architectures, modern SSDs deliver significantly faster read and write speeds compared to traditional hard disk drives (HDDs) and SATA SSDs.

This enhanced performance not only accelerates data access; it improves system reliability, reducing the likelihood and impact of downtime. By integrating these advanced SSDs into their infrastructure, businesses can achieve greater operational efficiency and resilience.

> *Next-generation solid-state drives for edge computing provide low-latency, high-speed data processing that supports real-time operations critical for distributed systems.*
>
> **Louis Kaneshiro | Kingston Technology**

2. Cost of Hourly Downtime Exceeds $300,000 for 90% of Firms; 41% of Enterprises Say Hourly Downtime Costs $1 Million to Over $5 Million. https://itic-corp.com/itic-2024-hourly-cost-of-downtime-report/

Reliability and data integrity

In today's digital economy, reliability and data integrity are non-negotiable for data centers managing mission-critical workloads. As businesses increasingly rely on real-time analytics, cloud-native applications and AI-driven systems, even minor data loss or corruption can lead to operational setbacks and reputational damage.

Data reliability depends heavily on the robustness of the storage infrastructure. Factors like drive endurance, error correction mechanisms and failover capabilities are vital in maintaining consistent uptime and data health. Storage media must not only resist physical wear but also detect and correct bit-level errors before they impact applications.

Enterprise-grade SSDs, with technologies like end-to-end data protection, power-loss protection and advanced error correction (such as LDPC), are designed to safeguard data even under extreme workloads. And with the growing prevalence of ransomware and cybersecurity threats, preserving data integrity during attacks is equally critical. Next generation SSDs – built for endurance, consistency and resilience – provide the foundational layer of trust that modern data centers need. By prioritising reliability and data integrity, organisations can protect their most valuable asset: their data.

> *With growing compute demands, Gen5 SSD scalability enables data centers to adapt and meet tomorrow's demands. Features like hardware-assisted power loss protection ensure data integrity; essential in mission-critical applications where data preservation takes priority.*
>
> **Louis Kaneshiro | Kingston Technology**

> *SSD energy efficiency is particularly useful as they maintain high performance while reducing power consumption and overall carbon footprint – a critical consideration for devices likely to run in mid-level or high-end data center deployments.*
>
> **Louis Kaneshiro | Kingston Technology**

Managing energy consumption and sustainability

Along with ensuring reliability and data integrity, reducing data center energy consumption is a top priority for IT leaders. As sustainability goals and operational efficiency converge, and global data volumes soar, so does the energy required to process, store and move that data.

In 2023, global data center demand hovered at 340 TWh, accounting for roughly 1.3% of worldwide electricity use. Multiple sources predict that demand will surge dramatically by 2030, amplifying the pressure on power infrastructure and sustainability efforts[3].

Reducing power consumption isn't just about meeting environmental regulations and corporate ESG (Environmental, Social and Governance) targets. It's critical in lowering utility costs, and ensuring the viability of a data center. As a result, storage optimisation is a key area of focus for today's data center manager.

Traditional spinning hard drives consume more power and generate more heat compared to SATA SSDs. In contrast, while the next-generation PCIe NVMe SSDs are closer to HDDs in terms of power consumption and heat generation, they offer significantly better power efficiency when measured in performance per watt. Thanks to the impressive number of IOPS that PCIe NVMe SSDs can handle, data can be rapidly accessed and processed by host systems, reducing operation times and positively impacting overall energy usage.

High-capacity PCIe NVMe SSDs also enable higher-density storage in smaller footprints. Not only does this support greener operations, but it also enhances the scalability of data infrastructure. In this way, advanced SSDs are pivotal to sustainable, high-performance data center design.

3. Growth in global energy demand surged in 2024 to almost twice its recent average. https://www.iea.org/news/growth-in-global-energy-demand-surged-in-2024-to-almost-twice-its-recent-average

Future-proofing your data center infrastructure

As workloads grow more complex and data volumes surge, adopting scalable, high-performance technologies becomes critical. PCIe 5.0 SSDs are a game-changer in this space, offering double the bandwidth of Gen 4 and significantly reducing latency. It empowers faster data access, enhances GPU utilisation for AI workloads and supports denser storage deployments.

Integrating Gen 5-compatible SSDs and server architectures ensures infrastructure can handle future demands, whether real-time analytics, machine learning or edge computing, without requiring constant, costly hardware overhauls.

When considering Gen 5 data center SSDs, businesses should start by assessing their applications and storage needs to determine whether this level of performance is necessary. This process should begin with the storage architect's evaluation of the performance requirements in the next one, two or five years.

Different data center environments and scale-up/scale-out prediction for storage growth should be considered, along with ROI measurement and evaluation. For this, application telemetry can help gather clinical performance metrics and storage capacity growth expectations.

Current hardware infrastructure should be factored in, as well as current hardware compatibility with existing PCIe 5.0 infrastructure to understand power delivery requirements and validate thermal management solutions.

> *Total cost of ownership should be evaluated not just on initial purchase price, but on performance per watt, reliability metrics and total operational efficiency.*
>
> **Tony Hollingsbee | Kingston Technology Europe**

> *By examining these factors in depth, and considering futureproofing strategies/avoiding pitfalls, organisations can make informed decisions that ensure they are properly equipped to face today's issues and tomorrow's technology advancements.*
>
> **Louis Kaneshiro | Kingston Technology**

Kingston's role in empowering data centers

SSDs such as Kingston's DC3000ME are at the heart of enhancing server and GPU performance. They offer the ability to reduce downtime and handle edge computing, AI infrastructure and intense computational processes with ease.

Representing a huge leap in data center storage performance, the DC3000ME is specifically designed to address the exponential demands of next-generation AI and high-performance computing workloads.

Our key engineering innovations include a revolutionary PCIe 5.0 interface with advanced controller technology. This feature delivers unprecedented read/write speeds, integrated thermal management and a form factor optimised for dense server environments. Such advancements highlight our commitment to shaping a robust and nimble technological ecosystem for future data center generations to benefit from. Here's how:

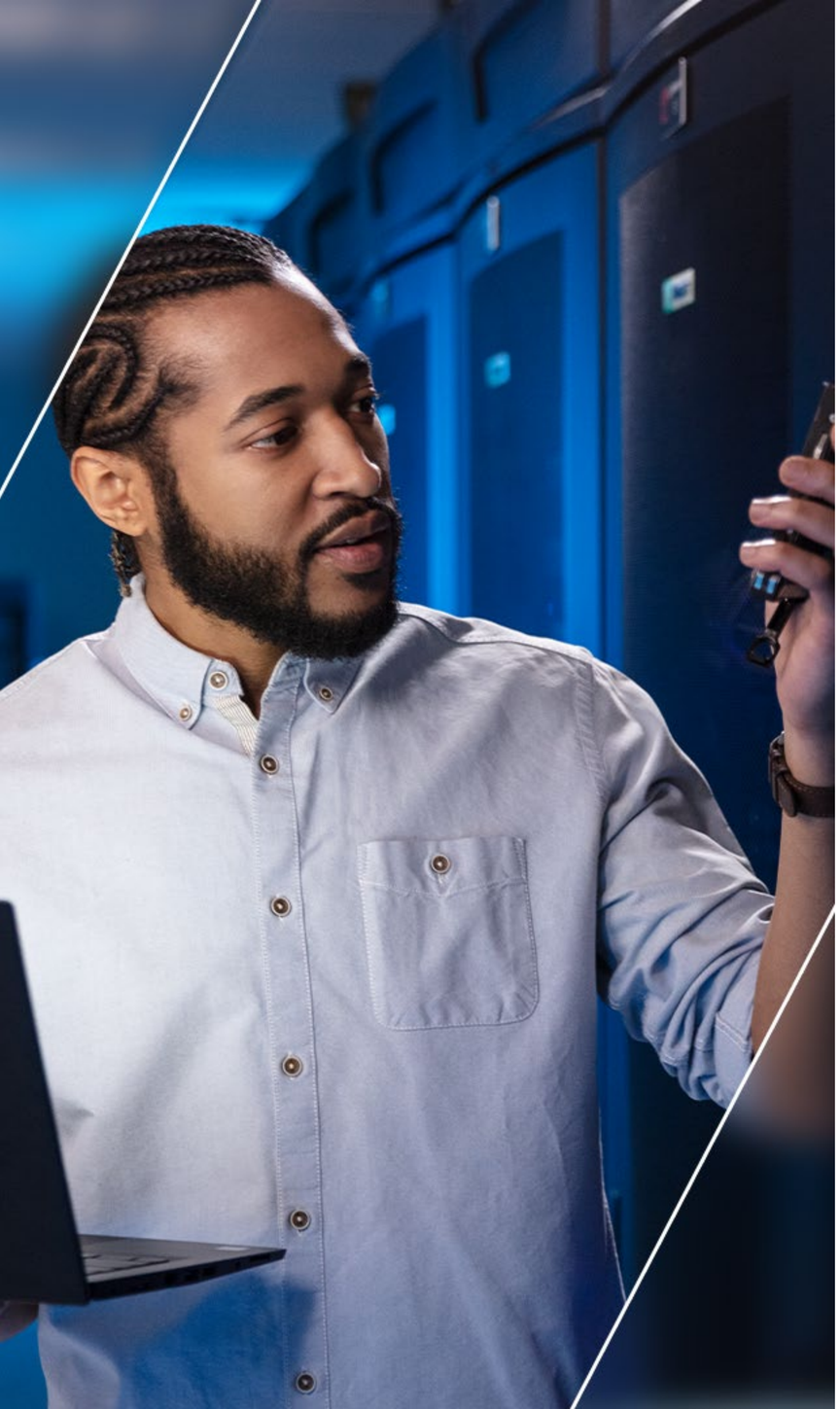
Built for heavy workloads

For AI and ML workloads, the DC3000ME with its Gen5 interface offers superior read throughput and read IOPS. This means that even a single DC3000ME drive can effectively deliver IO to multiple GPUs to effectively saturate performance. The result is faster training times and fewer SSDs needed, especially compared to Gen4, to keep up with multiple GPUs, reducing infrastructure costs while maintaining performance.

DC3000ME complements AI infrastructure with its high bandwidth, easily taking care of data-intensive workloads and enabling seamless GPU utilisation. The ability to serve multiple GPUs simultaneously accelerates AI training and inference activities, making them more efficient and productive. With growing compute demands, Gen5 SSD scalability enables data centers to adapt and meet tomorrow's demands.

> *The DC3000ME performance metrics are exceptional, with sequential read speeds of up to 14,000 MB/s and write speeds of up to 10,000 MB/s. It offers different power stages, allowing for flexibility in power consumption with scalable performance.*
>
> **Tony Hollingsbee | Kingston Technology Europe**

> *With the implementation of the latest PCIe NVMe Gen5 technology and a very well-written firmware Quality of Service implementation, the DC3000ME delivers one of the highest bandwidths and IOPS per bay, while maintaining and sustaining low latency. Both of which are key pillars in enterprise reliability and predictability. This ensures that it can keep pace with the demands of cutting-edge data center workloads.*
>
> **Louis Kaneshiro | Kingston Technology**



Kingston's role in empowering data centers

Optimal performance

Data center managers frequently struggle with performance variability, complex deployment and managing drive longevity.

DC3000ME is manufactured with high reliability specifications to offer consistent performance for optimal uptime. It addresses these challenges with a host-initiated and controller-initiated telemetry functionality to facilitate real-time health monitoring and predictive failure analysis.

This is combined with robust enterprise-grade design with a 5-year warranty and 2 million hours MTBF to avoid downtime and deliver optimal performance.

Enhanced security

The DC3000ME is TCG Opal 2.0 compliant and includes AES 256-bit encryption for robust security. This feature ensures unauthorised users are locked out and data from the drive cannot be stolen if the device is physically ejected. Aside from its enhanced security capabilities, the DC3000ME provides:

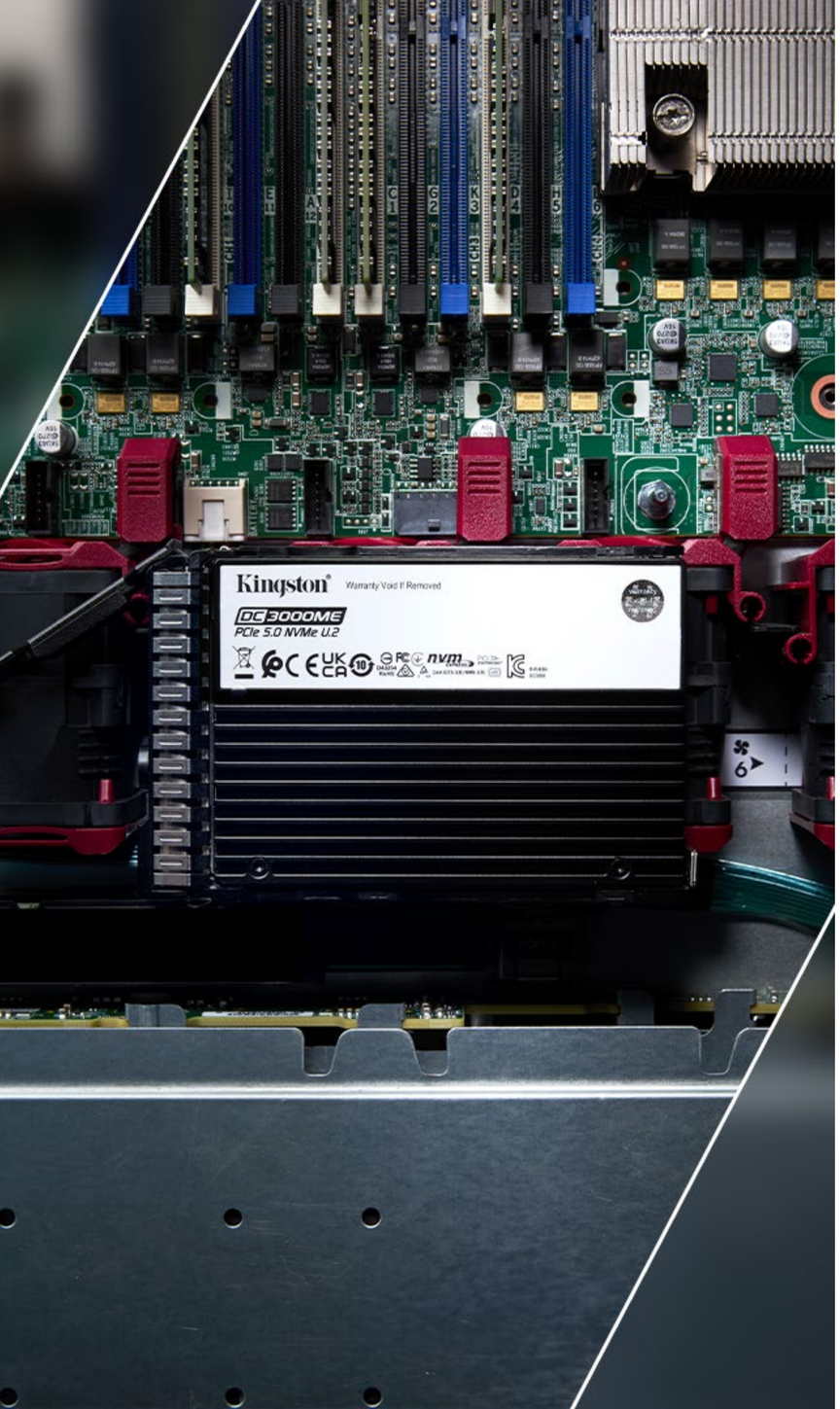
- **Hardware-based power failure protection:** Data integrity persists even with unexpected power failure.
- **NVMe end-to-end data protection:** Providing end-to-end protection of data along the entire storage path.
- **Firmware update without reset:** Facilitating continuous firmware updates without resetting.

Long-term cost savings

As an investment, DC3000ME pays off immediately in high-performance data center edge deployments and AI training use cases with real-time QoS control. Both are essential for application support and scalability, plus high-speed data processing.

Despite the high initial investment, overall long-term cost savings can be made due to its enhanced performance, reduced power requirements and less downtime, making it a strategic economic choice.

> *The DC3000ME delivers significant long-term value through reduced power consumption, minimal performance degradation and extended operational lifespan.*
>
> **Tony Hollingsbee | Kingston Technology Europe**



Kingston's role in empowering data centers

Superior energy efficiency

Power Management is a unique DC3000ME feature that offers flexibility in controlling performance per watt-drive utilisation, keeping operations costs down while reducing carbon footprint and aligning with sustainable goals.

Power failure protection ensures that data is hardened in case of sudden power loss, even at the drive level.

> *Strategic advantages offer support for high bandwidth multitasking without glitching, and decreased power consumption for environmentally friendly endeavours.*
>
> **Louis Kaneshiro | Kingston Technology**

> *The DC3000ME series boasts superior energy efficiency, delivering up to 970MB/s sequential read performance per watt. Through extensive hardware design and firmware optimisation, the DC3000ME series achieves higher hardware utilisation and minimises its impact on server heat dissipation.*
>
> **Louis Kaneshiro | Kingston Technology**

Wide compatibility

The DC3000ME series is compatible with different servers and OEM HBAs, supporting the PCIe Gen5 standard and conforming with the NVMe 2.0 spec.

It supports key features like NVMe-MI 2.0, an industry-standard protocol that enables the discovery, monitoring, configuration and updating of NVMe devices in different operating environments by the server's out-of-band management platform.

For example, Dell iDRAC 9 and the latest Supermicro BMC strictly comply with the NVMe MI 2.0 specification, allowing seamless integration with MI compatible NVMe drives like DC3000ME.

Building for the future

To future-proof their investment, organisations must choose scalable SSD solutions for data centers that meet current and future performance demand – while avoiding underutilisation or overutilisation of drive bays.

Kingston provides the expertise to help guide your decisions. As AI, ML and new technologies advance, our solutions stay focused on supporting the growth and effectiveness of these powerful advancements.

Together, we can accelerate your data center transformation with industry leadership expertise and storage solutions designed to keep pace with rapid change.

Built on Commitment

From big data to IoT devices, including laptops, PCs and wearable technology, Kingston Technology is dedicated to delivering top-tier product solutions, service and support. Trusted by leading PC manufacturers and global cloud providers, we value our long-term partnerships that help us evolve and innovate. We ensure every solution meets the highest standards by prioritising quality and customer care. At every step, we listen, learn and engage with our customers and partners to deliver solutions that make a lasting impact.