



StorageNewsletter - Predictions 2026 - Recap Opinions and Ideas - Jan. 21, 2026

Company	Prediction 1	Prediction 1 Detail	Prediction 2	Prediction 2 Detail	Prediction 3	Prediction 3 Detail
9livesdata	Re-emergence of Cost-Reducing Technologies and Techniques	The cost of hardware such as DRAM, flash, and even HDDs is expected to continue rising sharply. This trend may prompt a resurgence of established technologies like tape, offering a viable alternative for organisations seeking cost-effective solutions. Storage systems that help manage capacity consumption—through deduplication, compression, high utilization support, dynamic capacity sharing, thin provisioning, and similar techniques—are likely to experience broader market acceptance.	Anticipated Stronger AI Backlash	Beyond traditional concerns of AI replacing humans, a more immediate impact will be felt: consumers facing smartphone price increases and unaffordable gaming hardware; IT projects with significant infrastructure demands exceeding allocated budgets. These challenges are likely to provoke public outcry, strengthening anti-AI sentiment and even causing regulatory intervention. While market forces are expected to eventually address these issues, such corrections are unlikely to be effective in 2026.	no more predictions	no more predictions
Akave cloud	Agentic AI will Reshape Storage Architecture	By 2026, 40% of job roles in the G2000 (the world's 2,000 largest enterprises) will involve working with AI agents. Not humans with AI tools. Autonomous systems processing datasets at machine speed. That shift forces storage systems to behave very differently than the human-driven patterns they're built for. A human opens 50 files a day. An agent opens 50,000 per hour. Agents pull entire datasets and run continuously. Governance has to be built in the	Where Trust becomes Verification: the x402 Protocol will Unlock the First \$1B in Machine-Driven Tran	As AI agents become autonomous, they require autonomous transaction capabilities to pay for data, APIs, compute, and services without human intervention.The x402 protocol, a new internet-native payment standard from Coinbase and supported by Akave Cloud and others, addresses this by repurposing HTTP 402 ("Payment Required") into a programmatic payment handshake.	Geopatration will become Strategic, Sovereign Clouds are Key Geopatration is accelerating faster	European businesses are moving out of the US owned clouds. By 2026, 30% of organisations with sovereignty requirements will move to new localised cloud providers. Enterprises will discover that regional storage does not equal sovereignty. As the US CLOUD Act creates extraterritorial exposure that Frankfurt data centres don't comply with - many organisations' mistakenly
Arctecta	Data-intensive science and collaboration will drive the need for scalable research data platforms.	As research data grows exponentially in volume, variety, and velocity, traditional management practices that are heavily dependent on ad hoc, dispersed individual and departmental efforts are failing catastrophically. Institutions will need to fundamentally rethink long-term data management strategies to keep pace with this surge and ensure data remains accessible. Organizations that are proactive in their approach will accelerate discovery and innovation.	An AI-ready data infrastructure will be essential to optimizing AI model training and inference.	The real AI leaders won't be those with the biggest models, but those with the most unified, AI-ready data fabrics. Integrated platforms will replace fragmented data stacks, offering built-in vector database support, unified metadata, and pipeline orchestration that can move quickly, adapt to new models, and scale insights across the enterprise. This approach will provide a faster route to turning AI into value and will future-proof data infrastructures. Data readiness is AI readiness.	Data security and governance will become an ethical imperative.	An organization's credibility now depends as much on the integrity of its data infrastructure as on the integrity of its findings. In this high-stakes environment, immutability, traceability, and governance aren't just operational necessities; they're ethical imperatives. Metadata-driven systems are becoming a crucial operating backbone, automating access, retention, and policy enforcement while enabling secure collaboration across global locations. Organizations that thrive will be those that d
Astera Labs	The End of One-Size-Fits-All: Why 2026 Brings the Accelerator Explosion	We're about to see an explosion of accelerators, each laser-focused on specific workloads. Efficiency will soar as AI chip architectures drill down to the silicon level, optimizing distinct mathematical engines for peak performance, and integrate novel I/O interfaces to get data in and out at a faster pace.	Multi-Protocol Mandate: Flexibility Becomes the New AI Infrastructure Standard	Say goodbye to rigid, single connectivity solutions for rack-scale designs. The future belongs to hybrid AI fabrics solutions that can blend multiple open and proprietary protocols for scale-up deployments. Flexibility isn't optional—it's the new standard.	Inference at Scale Demands Memory Tiering: From Niche to Necessity	Large-scale inference workloads will demand tiered memory architectures. Whether accelerating KV Cache or powering Retrieval-Augmented Generation, tiering memory—via disaggregation or directly attached to each accelerator in the cluster—will move from niche to mainstream.
Backblaze	Neoclouds stop being "GPU clouds" — and become full AI-workflow clouds	In 2026, the fastest-growing cloud category won't be hyperscalers or GPU clouds — it will be "neoclouds" that handle datasets, movement, feature stores, fine-tuning, inference, end-to-end. AI traffic is 10x more intense than traditional workloads — and increasingly dominated by fewer, heavy endpoints. Providers that solve data ingest, caching, egress freedom, and training data pipelines (not just raw GPUs) will capture the next wave of AI spending.	Data movement will explode AI budgets	The biggest AI cost surprise won't be GPUs going forward but it dataset movement, prompting CFOs to require "egress clauses" in every AI contract. 95% of orgs already report surprise cloud fees; AI is only accelerating that. Flow-heavy AI pipelines (training refreshes, retrieval-augmented inference) create massive, expensive churn. Teams shift to egress-friendly, transparent storage providers as a direct response.	Multi-cloud AI pipelines become the default — with storage as the neutral control plane.	By late 2026, most serious AI teams run models across 2–4 GPU providers — but keep their data anchored in a neutral storage cloud they control. Crucial to maintaining this multi-vendor, best-of-breed approach will be terabit ingest + free egress architectures make it rational to burst to whichever GPU fleet is cheapest or fastest. Storage becomes the home base, and compute becomes ephemeral and interchangeable.
Cerabyte	Active archives will play a central role in ensuring high-value datasets remain instantly accessible	Organizations will increasingly adopt a combination of active archives, intelligent tiering, and hybrid cloud architectures to optimize storage utilization at scale. Tiering is necessary to group large datasets and assign them levels of importance and priority. An active archive serves this purpose well, as it allows data to be relegated to a lower tier while still being available rapidly should it be needed by the AI engine. - Eric Polet, Director of Product Marketing, Arctecta	Tape for long-term storage	Driven by exponential data growth and the need for lower-cost, energy-efficient, long-term storage, tape is poised to become a cornerstone of active archival tiers within hybrid storage architectures. - Marc Steinhilber, CEO, BDT Media Automation GmbH	Active archives will require sophisticated data analytics	The industry hit the storage wall as we predicted last year - rising lead times, media and stock prices tell the story. Active archives require sophisticated data analytics, as archives evolve from data dumps to data sources. Data storage must be accessible, sustainable, and affordable to unleash the full potential of AI.
Cohesity	Prediction 1: "No trend" is the defining trend of 2026	2026 cybersecurity will be defined by extreme unpredictability; attackers exploit volatility, often for chaos. Success demands adaptive architecture, dynamic defenses, and a resilient mindset. Organizations must focus on adapting and recovering from any attack, not merely predicting specific threats.	Prediction 2: Cyber maturity is rewriting the insurance playbook	By 2026, cyber insurance will link directly to a company's measurable cyber maturity. Insurers will reward robust resilience, making premiums risk-based and fostering shared accountability. With 92% of organizations globally facing legal or regulatory consequences, including fines or lawsuits post-attack, investment in cyber resilience becomes a quantifiable asset.	Prediction 3: Data control and Smart governance in Europe	European enterprises are prioritizing enhanced control over their data, especially in cloud environments. This drives demand for intelligent, automated data governance. The emerging concept of "self-aware data" promises to empower organizations with granular oversight, ensuring compliance and robust data integrity throughout the region. This is a key concept for AI adoption, as 81% of organizations feel that generative AI adoption outpaces their ability to securely mitigate associated risks.



StorageNewsletter - Predictions 2026 - Recap Opinions and Ideas - Jan. 21, 2026

Company	Prediction 1	Prediction 1 Detail	Prediction 2	Prediction 2 Detail	Prediction 3	Prediction 3 Detail
CTERA	The End of the "Single Cloud of Failure"	After a string of major cloud outages exposes the brittleness of centralization, 2026 will mark a definitive industry shift toward hybrid resilience. Organizations that take availability seriously will no longer bet everything on one provider, driving a significant move to architectures where storage is replicated across on-prem environments, regional facilities, and multiple public clouds. This move is not about abandoning the cloud, but about building for survival when any single component fails	The Reckoning with "Shadow AI"	The quiet leakage of sensitive corporate data into public AI tools will become a full-blown governance crisis. Enterprises will awaken to the reality that they have a "shadow" copy of their internal world, which is discoverable and uncontrolled, living on third-party servers. In response, 2026 will see a push to bring AI workloads and their resulting data back inside the corporate security perimeter, making private, governed AI storage solutions a top priority to regain control over sensitive intellectual property	Navigating the AI Walled Gardens	The biggest data management headache of 2026 will be the fragmentation of data access, as tech giants begin walling off their ecosystems. Platforms will increasingly block or "authenticate" which AI agents can interact with their services, creating a new layer of competitive gatekeeping. This will force a strategic rethink of data mobility and integration, as organizations must now navigate a landscape where open access is replaced by a permissioned system, complicating everything from analytics to automation
Databahn	AI Neutrality Will Become a Non-Negotiable Requirement for Enterprises	In 2026, companies will demand the freedom to run the right model for the right job without being locked into one vendor's vision. A neutral pipeline will let teams choose models based on value and governance, not on who owns the storage layer. This shift will mirror what happened with cloud neutrality and will become a board-level conversation.	Telemetry Will Be Routed by Purpose, Not by Habit or Vendor Defaults	Storage and SIEM bills will force a reset. Instead of sending everything everywhere, enterprises will start routing data based on security value. High-value detections will go to analytics platforms. Broader telemetry will move to cheaper storage. Agentic AI will guide these decisions by scoring relevance in real time. This will reshape how organizations think about data retention.	AI Observability Will Become Essential to Pipeline Governance	In 2026, AI systems will need the same level of observability we expect from traditional infrastructure. Teams will track how models enrich data, detect drift, and shape routing decisions. They will require full transparency into versioning and confidence thresholds. This visibility will be critical for trust and for proving that AI decisions align with policy and compliance obligations.
Datadobi	Data governance will shift from a compliance checkbox to a strategic enabler of AI and analytics.	GenAI and advanced analytics demand high-quality inputs, but most organizations struggle with unstructured data management. As businesses increasingly recognize the challenges this creates, governance frameworks will evolve beyond compliance into a foundation for enterprise intelligence. Those who can also break down silos and provide enterprise-wide visibility will unlock faster collaboration	Legal and risk teams will play a critical role in shaping enterprise data strategies.	Regulatory pressure around data protection, sovereignty, and AI ethics is intensifying. Legal teams that once reacted to risks hidden in file shares or outdated records will increasingly partner with IT and data leaders to proactively shape governance policies. By ensuring visibility and defensible processes across petabyte-scale environments, they will become strategic enablers, reducing litigation risk, su	AI Is Only as Smart as Your Data: Why CDOs Must Take Control	Gen AI and advanced analytics promise important insights, but only if the underlying data is clean, compliant, and accessible. CDOs need to take control of their data and bring order to unstructured chaos, ensuring that everyone can work with trusted inputs, allowing AI initiatives to be built on solid ground.
DDN	The Era of Intelligent Infrastructure	AI has moved beyond experimentation to scale, efficiency, and sovereignty. Mastering data—how it's moved, managed, and operationalized—drives innovation, control, and speed, while smart infrastructure reduces wasted energy and cost.	The Rise of AI Sovereignty	Nations and enterprises see AI sovereignty as digital sovereignty. Secure, governed, and portable data across clouds and sovereign platforms ensures leadership, compliance, and energy-conscious operations.	Intelligence per Watt	AI-driven energy demand is soaring toward 1,000 TWh by 2027. Organizations deploying optimized, autonomous infrastructure can maximize utilization, reduce waste, and lead in performance, sustainability, and economic resilience.
ExaGrid	2026 is the Year that Ransomware Recovery Drives Data Protection Decisions	Ransomware attacks continue to rise, disrupting business, impacting reputations, and costing organizations millions. We've seen a record number of customers realize that they are more exposed than they thought. They now know that they can't only depend on having ransomware prevention in place and that they need another line of defense ready if it fails. They need an air gap copy, immutable data objects, and delayed deletes or locking, and are seeking more AI automation to catch attacks early.	Artificial Intelligence Rollouts Will Continue to Use a Large Share of IT and Service Budget	There is no doubt that the AI cat is out of the bag, and all companies are running to take advantage in order to position themselves against their competition by making employees more productive and by analyzing data to make fast and strategic business, product, and other decisions. IT data centers and service providers will be challenged to find the balance between their core app, storage, networking, etc. infrastructure spend and their AI spend.	SSD Market Share Will Pick Up Steam in 2026	In 2026, the conversion of HDD to SSD will pick up momentum as customers look for faster storage to accelerate user productivity, faster backups, faster restores, and lower power and cooling costs. The price of SSD is still much higher than HDD, but the gap continues to close and there is a point where customers will be willing to pay the difference for performance and power savings. There is no doubt that the future is SSD, and the only question is when does the big move from HDD to SSD occur.
Grau Data	Data Preparation Overtakes Storage as the Real Cost	By 2026, enterprises will recognize that storage capacity and performance are no longer the main cost drivers in AI pipelines. The dominant expense will come from repeatedly parsing, tokenizing, embedding, and enriching the same unstructured data. GPU consumption and redundant preprocessing will exceed raw storage spend, driving demand for intelligence captured once and reused across workflows.	Storage Systems Become Workflow-Aware	In 2026, storage platforms will increasingly support AI and analytics workflows by exposing insight, context, and lineage alongside data. Most discovery and reasoning will occur without opening files, while direct access becomes the exception. Storage will evolve from a passive repository into an intelligence-aware layer that supports governance, analytics, and automation.	AI-Driven Workflows Redefine Storage Buying Decisions	By 2026, organizations will evaluate storage solutions based on how effectively they support AI-driven workflows, not just capacity or performance. Buyers will favor systems that preserve context, reduce repeated data handling, and keep data usable across models, tools, and time—blurring the line between storage, data management, and AI enablement.
Hammerspace	The End of Data Fragmentation as AI's Silent Killer	In 2026, enterprises will need to confront fragmented data estates. The industry will recognize the major limitation to AI adoption isn't GPU supply; it's data access speed, consistency, and reach. Organizations will shift investment from more compute to unified data platforms, making existing infrastructure AI-ready. Those who treat data fragmentation as an architectural flaw to be eliminated will win the AI race. Performance, cost efficiency, and scalability will flow from this unification.	Energy and Efficiency Drive Infrastructure Innovation	The sheer scale of inference and GenAI workloads will force a reckoning with power and efficiency. By 2026, new infrastructure technologies — from smarter data orchestration layers to energy-aware storage and compute systems — will emerge as enterprises seek to manage costs and sustainability pressures. We expect infrastructure vendors to compete not only on speed and scale, but also on their ability to tame energy consumption while maintaining enterprise-class performance.	The Year of the AI Predictionary — Where Efficiency Defines Intelligence	The world's compute capacity is now bound by energy and data movement, not transistors. As a result, efficiency will become the new scoreboard of AI progress, measured in tokens-per-watt, throughput-per-rack, and time-to-insight. AI Predictionaries will rise as the modern equivalent of industrial power plants, unifying data, compute, and automation into tightly orchestrated systems that transform raw information into actionable intelligence at unprecedented speed.
Hitachi Vantara	Data Defines Industry Leaders in 2026 (Sheila Rohra, CEO)	"Data will become the ultimate differentiator. Organizations that can connect, govern and trust their data will be the ones that move faster and make smarter decisions in an AI-driven world. But achieving this will mean disrupting traditional data practices and building new capabilities for speed and trust." - Sheila Rohra, CEO, Hitachi Vantara	Power and Density Will Drive Data Center Modernization (Octavian Tanase, Chief Product Officer)	"AI and data-intensive workloads demand more compute and storage, straining energy use. Flash storage offers better performance with lower power consumption so organizations can reduce costs and prioritize AI innovation. Companies will adopt systems that enable them to consolidate applications using high-density or flash-based technology rather than traditional hard-disk environments. Enterprises will modernize around these compact architectures that yield business and sustainability returns."	Solutions-Focused Vendors Are Poised To Win In 2026 (Todd Hemenway, VP, Americas Partners)	"The right combination of compute will become paramount, which may be a combination of high-end compute with storage, in file, block and/or object formats. The challenge for vendors will be unlocking the right combination by providing clear partner guidance and reference designs while understanding what's best for the customer from both an access and cost standpoint. Those who simplify that complexity and design the best mix for performance, cost and scale will be poised to win in 2026."



StorageNewsletter - Predictions 2026 - Recap Opinions and Ideas - Jan. 21, 2026

Company	Prediction 1	Prediction 1 Detail	Prediction 2	Prediction 2 Detail	Prediction 3	Prediction 3 Detail
HYCU	Agentic AI will become the largest new source of unintentional data loss.	AI workflows and autonomous agents will start to operate across SaaS platforms, APIs, and data stores with limited human oversight. These systems will copy, transform, and redistribute sensitive data at a pace that existing data protection controls were never designed to handle. The most damaging incidents will not come from external attackers but from poorly governed agents acting exactly as designed, exposing data through over-permissioned access, persistent memory, and uncontrolled automation	SaaS security will become the primary strategic priority for IT.	As systems of record fully move into SaaS platforms, enterprise risk will concentrate around identity, access, and data hygiene rather than endpoints or networks. AI initiatives will force organizations to confront how fragmented, overexposed, and poorly understood their SaaS data has become. Data protection has lagged behind SaaS adoption for years, and in 2026 closing that gap will be required to safely deploy AI, not simply to meet compliance requirements.	IT will be forced to radically consolidate to fund AI initiatives.	Rising costs for infrastructure, model access, data pipelines, and licensing will collide with years of unchecked SaaS sprawl. IT leaders will be forced to reduce vendors, eliminate redundant platforms, and simplify architectures in order to sustain AI investments. As a result, IT and security teams will regain control over which technologies are adopted, once again shaping innovation through architectural discipline rather than experimentation alone.
HyperBunker	Offline recovery copies will become an audit expectation.	Within the next 24 months, regulators, auditors, and insurers will increasingly expect organizations in regulated sectors and supply chains to demonstrate access to physically isolated recovery copy. Logical immutability alone will no longer be considered sufficient to prove recoverability.	Ransomware response will shift focus from prevention to recovery certainty	As ransomware groups continue to bypass preventive controls through credential theft and trusted access, organizations will place greater emphasis on guaranteed recovery outcomes. Investment decisions will increasingly prioritize technologies that ensure a last-known-clean restore even when production systems and backups are compromised.	Backup platforms and recovery vaults will separate into distinct categories	The market will increasingly differentiate between connected backup platforms optimized for availability and dedicated disaster recovery vaults designed for isolation and worst-case scenarios. Organizations will deploy these as complementary layers.
Infinidat	The Continued Rise of AI Applications and Workloads to Accelerate Enterprise Transformation	In 2026, AI deployments will continue to increase exponentially, with massive implications for enterprise storage. For many enterprises, AI will move from a pilot phase into core enterprise operations. This signals a pivot to AI-centric storage architectures and higher performance solutions for retrieval-based AI workloads. The AI blueprint that is poised to become the 'AI architecture of the year' for 2026 is the Retrieval-Augmented Generation (RAG) workflow deployment architecture.	Cyber Storage Resilience to Make Safeguards More Proactive and Preemptive	In 2026, a shift will accelerate in the enterprise market from reliance on traditional detection and response in cybersecurity to prioritizing proactive cyber defense and preemptive measures to make storage infrastructure more cyber secure. Enterprise cyber storage resilience and recovery will become mandatory for enterprises to have a comprehensive cybersecurity strategy. Next-generation data protection capabilities, such as automated cyber protection and cyber detection, will be in demand.	Increasing Demand for Power Efficiency to Reshape Data Infrastructure	The need for more power will continue to increase in data centers in 2026, leading to new mandates to reallocate and free up power sources for power-hungry AI workloads and applications. Not only is AI a major Predictionor in increased power consumption, but non-AI tech is also demanding more power. Enterprises will be looking to generate power savings by reducing the use of energy in their enterprise storage infrastructure, requiring the adoption of more power-efficient storage systems.
Komprise	Unstructured Data Preparation will emerge as key to delivering AI ROI	As AI pilot projects are starting to go mainstream, enterprises are seeing anywhere from 60% (source: Gartner) to 90% (source: Deloitte) of AI projects fail. IT leaders are realizing that poor AI ROI is due to poor data quality. With 90%+ of enterprise data now being unstructured, curating and preparing it is key to realizing AI ROI. The shift from a heavy focus on structured data to unstructured data preparation will accelerate in 2026.	CIOs Prioritize Cost over Everything Else	Storage costs will rise sharply in 2026 due to both HDD and Flash component cost increases. More data, more costs and more security risks is what CIOs are dealing with in 2026 according to the Komprise 2026 State of Unstructured Data Management. Deferring storage purchases, shrinking ransomware defense costs, and cutting AI expenses will be important priorities for unstructured data management. Cold data tiering, unstructured data classification and curated AI ingest will rise to the top.	Security becomes embedded feature of unstructured data management	Unstructured data is a significant security risk: it is unmanaged and growing quickly across silos. It is highly vulnerable to ransomware attacks which can lurk unnoticed in the chaos. Unstructured data is also being fed to AI without controls for sensitive information, causing data governance risks. Securing unstructured data, classifying it and enabling proper handling of sensitive data during AI ingestion will be expected features of unstructured data management.
Leil Storage	HAMR time arrival	Now the manuPredictionuring output and number of units are limited due to early stage and consequently higher costs associated with this. Once adoption grows, and potentially the second vendor joins it, we can expect the technology spreads more actively, also for a reason of the necessity of pushing shipped capacity higher which is possible with HAMR.	SMRs becoming a default HDD	All random workloads are either on flash or migrating there. HDDs are getting bigger while IOPS do not. Therefore the omni prevalent pattern of accessing HDDs for workloads they are good with, will be sequential. And sequential + capacity gains + price efficiency + energy efficiency are what SMR drive is.  We believe that this will help keep HDDs relevant for the foreseeable future.	Transition of materials used in AI from tape to "actionable" storage	A lot of data which is already used by AI or will begin to be used by AI, various training material is still on tape. We believe that the intensity of usage of this material will trigger the transition to "actionable" storage, e.g. HDDs as flash is too expensive for bulk data which does not have to be hot. SMR based storage can be a great landing spot for this.
Lightbits Labs	NVMe-oF Becomes the Default Data Fabric	NVMe-oF will evolve from a performance niche into the ubiquitous backbone of Tier-1 and Tier-2 data fabrics as organizations accelerate the replacement of legacy protocols such as iSCSI and Fibre Channel. Fueled by the demands of generative AI and hyperscale microservices, latency tolerance will shrink to single-digit microseconds, driving widespread adoption of NVMe-TCP and accelerating the move to faster networks (e.g. 400GbE).	Intelligent Block Storage Becomes the Brain of AI Data Pipelines	The stopgap of block storage as a cache for vector databases will give way to intelligent systems in which storage converges with compute-- executing operations close to the data, eliminating the need for costly in-memory caches. Next-gen software-defined systems powered by NVMe/TCP will slash latency, lower storage costs, and make PB-scale RAG and generative AI workloads faster, simpler, and economically viable.	Block Storage Becomes the Backbone of Cyber Resilience	Cyber extortion and stricter regulations will redefine data security. A shift from recovery alone to recovery and prevention will see block storage emerge as a gold standard for backups--offering tamper-proof immutability, instant recovery, and built-in zero-trust defense. As organizations overhaul their DR strategies, block storage will emerge as the foundation of a secure, compliant, and resilient data protection plan.
Lucidity	Unpredictable AI Workloads Will Push Block Storage Costs Higher in 2026	AI and agentic workflows will surge in 2026, sharply increasing cloud block storage usage. As these workloads scale horizontally, demand high IOPS and throughput, and create unpredictable bursts and task branches, teams will continue to over-allocate storage to stay safe. The result: significant waste and an outsized share of already tight cloud budgets consumed by block storage.	Storage debt becomes a board issue as waste, ESG demands and rising cloud costs force strict KPIs	Explosive data growth from AI, analytics, and SaaS is driving massive storage expansion--and hidden waste. As pressure from shareholders, regulators, and ESG frameworks increases, unmanaged storage becomes a financial and environmental liability. By 2026, organizations will set storage efficiency benchmarks and require quarterly audits, turning "storage debt" into a governance priority with executive visibility.	By 2026, storage right-sizing becomes the top cloud cost action, finally overtaking compute	As AI adoption accelerates and data-heavy workloads grow, storage is becoming the fastest-rising part of cloud spend--surpassing compute for many organizations. With block storage waste often exceeding 60%, the industry will shift from reactive cost-cutting to proactive, data-driven optimization. Tools will pivot toward automated storage right-sizing and tiering as the highest-impact savings lever, marking a major turning point for FinOps and the broader cloud optimization market.
N2W	Outages drive the industry towards a shadow cloud future	As hyperscaler outages continue to make headlines, organizations are losing patience with "all-in-one" cloud dependency. 2026 will accelerate a shift toward smaller, regional clouds and multi-cloud strategies. Organizations will also start relying on secondary, lesser-known cloud providers as hidden backups, creating a "shadow cloud" layer. Hyperscalers will feel the pressure, scrambling to reassure customers with aggressive marketing campaigns, new reliability guarantees, and promises of ironcl	The rise of invisible clouds and predictive failover	As organizations lose trust in even the largest hyperscalers, the next wave will include clouds that don't rely on human intervention. Adopting "invisible cloud" architectures where workloads automatically fail over across different providers, regions, or even private clouds will emerge. AI tools will help predict outages before they happen and reroute traffic and data in real time. As a result, businesses will stop caring about which cloud they're using at any moment.	The rise of invisible clouds and predictive failover	As organizations lose trust in even the largest hyperscalers, the next wave will include clouds that don't rely on human intervention. Adopting "invisible cloud" architectures where workloads automatically fail over across different providers, regions, or even private clouds will emerge. AI tools will help predict outages before they happen and reroute traffic and data in real time. As a result, businesses will stop caring about which cloud they're using at any moment.



StorageNewsletter - Predictions 2026 - Recap Opinions and Ideas - Jan. 21, 2026

Company	Prediction 1	Prediction 1 Detail	Prediction 2	Prediction 2 Detail	Prediction 3	Prediction 3 Detail
Nakivo	The rise of autonomous data storage	We expect the storage industry to move past simple scalability to adopt intelligent and inherently secure architectures. When it comes to AI-native storage, instead of adding AI features, new systems will be built from the ground up with ML models at their core. These platforms will be able to self-manage, automatically tier data, detect anomalies, and improve performance. That's autonomous storage we're talking about.	Unified management with edge-cloud-hybrid ecosystems	Another area we see skyrocketing is the edge-cloud-hybrid ecosystems or smart multi-tiered storage. Organisations will operate across edge, on-premises, and public/hybrid cloud with seamless data mobility and unified management. In Prediction, some of our enterprise customers have already adopted this model. On that front, we expect growth in storage analytics and tools for real-time monitoring and automated management to enable visibility across different tiers.	New technologies enabling data reduction, automated tiering and carbon-aware storage	Technologies that enable powerful data reduction (like deduplication, compression, and swap data) and automate tiering to carbon-aware storage will definitely be at the forefront. Meanwhile, legacy disk-based systems and traditional HDDs will steadily lose ground to more efficient and scalable alternatives, such as next-generation multi-hundred-terabyte QLC SSDs.
NetApp	AI graduates from pilots to production—powered by Intelligent Data Infrastructure	Businesses will shift from trials to production by prioritizing unified, well-governed, and accessible data—and by cutting the cost and complexity of data preparation. Intelligent Data Infrastructure is the key: it automates ingestion, transformation, lineage, quality, access, and policy, reducing total cost of ownership (TCO) while accelerating AI readiness. Weak data foundations cause most AI failures; strengthen them to turn ambition into ROI.	Cyber resilience goes AI-first—embedded, self-healing defense	AI-driven detection, predictive analytics, and autonomous response will become standard practice, shrinking time to contain threats and reducing false positives. AI-driven resilience is the only way to stop ever growing threats and AI-driven threats. The breakthrough is embedding these capabilities within Intelligent Data Infrastructure—continuous, contextual security that's self-healing rather than bolted on. The result is stronger resilience at lower operational overhead.	Unified hybrid multi-cloud unlocks AI anywhere—run models where the data lives	The future isn't choosing one cloud; it's unifying and governing data across on-prem and hybrid multi-cloud to enable 'Alatthedata.' Intelligent Data Infrastructure reduces latency, enforces compliance, and leverages native model integrations across all hyper-scalers so teams can run AI where it makes the most sense. The payoff is faster outcomes without lock-in.
Nexsan	AI Is Coming to the Storage Layer — and That's Good News for Cyber Resilience	As we move into 2026, cyberattacks are becoming faster, smarter, and more destructive. IT leaders are finally accepting what many of us have said for years — data storage and data security can no longer be separate disciplines. Storage has always been part of the security conversation, but now it's becoming something more: an intelligent participant in cyber defence. The arrival of AI in the storage layer marks a turning point in how organizations prevent, contain, and recover from ransomware.	Smarter Storage, Stronger Prevention	AI isn't replacing storage's traditional mission of availability, integrity, and recoverability — it's enhancing it. Modern storage systems are learning to recognize abnormal data access, detect ransomware-like behaviour and even trigger automatic protective actions such as isolating affected volumes or preserving immutable snapshots. This doesn't make storage a security product — it makes it a smarter, more resilient foundation that supports security operations with real-time insight and cont	Aligning Storage and Security Goals	As cyberthreats evolve, the best organizations will align storage and security strategies. CISOs already view storage not as capacity, but as part of the resilience infrastructure — expected to meet not only performance SLAs but also recovery and compliance SLAs. AI will deepen that connection, giving both teams shared visibility into data behavior, anomalies, and recovery readiness.
Object First	IT will favor backup storage solutions that don't rely on human oversight	Secure data storage has become a frontline component of cybersecurity, and customers will see it as a part of their active defense and resilience strategy. 2026 will demand solutions that deliver strong security controls while remaining easy to deploy and manage across environments. Therefore, Immutability and zero-trust architectures will be wildly acknowledged as core requirements for companies and the entities that insure them.	Cyber resilience will finally become a continuous investment	There has been a widening gap between the cyber resilience an organization thinks it has and how capable its cyber defenses actually are. Many times, resilience planning stops at detection, instead of focusing on data integrity/recoverability through immutable and segmented data backups – the most underutilized and effective disaster recovery tool. In 2026, we will see the gap between perceived cyber resilience and actual capabilities decrease, with more prioritization of immutable storage.	Cyber resilience will finally become a continuous investment	There has been a widening gap between the cyber resilience an organization thinks it has and how capable its cyber defenses actually are. Many times, resilience planning stops at detection, instead of focusing on data integrity/recoverability through immutable and segmented data backups – the most underutilized and effective disaster recovery tool. In 2026, we will see the gap between perceived cyber resilience and actual capabilities decrease, with more prioritization of immutable storage.
Oxibox	Cyber vaulting mainstream	Cyber vaults—isolated clean rooms storing immutable backups completely separated from production networks—will become standard in 2026. Rising attack frequency and insurer demands for proven recovery capabilities are driving vendors to democratize pricing. Mid-market companies will access what was once reserved for banks and critical infrastructure. When ransomware strikes, organizations restore from verified-clean data in hours, not weeks. The shift from premium to essential is inevitable.	SaaS backup urgency	Many organizations assume Microsoft 365 and Google Workspace automatically protect their data—they don't. Native retention is limited, and deletion or ransomware can be permanent. As attackers increasingly target cloud productivity suites via phishing and token theft, data loss incidents are surging. Third-party SaaS backup solutions provide independent, immutable copies outside the blast radius. In 2026, expect rapid adoption as businesses realize shared responsibility means their cloud data is	AI-powered resilience	AI will transform backup security in 2026 by detecting threats before damage spreads. Machine learning analyzes backup patterns—file change rates, entropy levels, access behaviors—to spot ransomware encryption in real time. Anomalies trigger automatic isolation of compromised data and alerts. AI also predicts storage failures, optimizes backup schedules, and validates recovery integrity without manual intervention. This shift from reactive to predictive protection dramatically reduces recovery t
Parallel Works	Neo Cloud Providers Challenge Hyperscaler Dominance	Specialized GPU-focused cloud providers — often delivering services 4x less than Amazon, Google, or Microsoft's cost — will carve out a meaningful share of AI workloads. Their pricing models, flexibility, and regional presence will give mid-sized enterprises and research institutions a more viable entry point into advanced AI and HPC workloads. This new tier of "neo clouds" will increasingly become a bridge between expensive public cloud offerings and private infrastructure ownership.	Hybrid Multi-Cloud as the Default Model	By 2026, hybrid and multi-cloud architectures will be the standard for HPC and AI, replacing the one-size-fits-all approach of monolithic on-prem systems. Workloads will dynamically move across on-prem, cloud, and specialized resources (GPUs, quantum, etc.) to balance performance, cost, and compliance. Cloud bursting and heterogeneous workload placement will no longer be differentiators — they will be table stakes for competitiveness in AI-driven industries.	The Rise of Private AI	Enterprises will increasingly move away from fully relying on public hyperscalers and toward private or semi-private AI infrastructure. Neo cloud providers will take center stage as hedge funds, defense contractors, and other data-sensitive organizations will begin leveraging them for GPU access and to manage Kubernetes environments. This will then allow them to transition to owning their own AI systems. This shift reflects a growing desire for control, cost predictability and sovereignty in how
Peer Software	Active-passive architectures die	2026 will mark the end of active-passive data architectures as real-time AI and globally distributed pipelines demand constant availability. Organisations will no longer accept delayed/corrupted sharing of project files across locations, idle backup systems or slow, disruptive failovers. Instead, active-active architectures, where every site participates, synchronises instantly and actively serves traffic, will become the expected baseline for resilience, efficiency and performance.	AI consolidation accelerates	In 2026, AI consolidation will intensify as major vendors acquire smaller AI, data and edge-platform companies to expand their ecosystems and speed up innovation. But the real challenge won't be the deals themselves: it will be integrating the many incompatible systems they bring. The companies that can rapidly unify data, metadata and file services across newly merged environments will deliver value faster and win market trust.	Metadata becomes a critical AI advantage	As AI fuels explosive data growth, metadata will become a decisive competitive advantage in 2026. Organisations will move away from replicating entire datasets and instead rely on metadata-driven orchestration to determine exactly which slices of data each AI, ML or analytics workflow needs. Rich metadata that captures context, relevance, access patterns and intent will guide how data moves, cutting costs and accelerating pipelines.



StorageNewsletter - Predictions 2026 - Recap Opinions and Ideas - Jan. 21, 2026

Company	Prediction 1	Prediction 1 Detail	Prediction 2	Prediction 2 Detail	Prediction 3	Prediction 3 Detail
Phison	AI Focus Will Move Beyond Access to Experience	-This year, AI access expanded through edge use cases, similar to the early Internet where access came first, and user experience improvements quickly followed. Just as smartphones once hit a bandwidth bottleneck, today's edge AI is facing a memory-bandwidth limitation that hinders user experience and revenue-generating applications driven by inferencing.	AI Focus Will Move Beyond Access to Experience	In 2026, development teams will focus on accelerating inferencing by using solutions that expand the KV cache space to unlock full GPU power. The industry will also see significant increases in memory bandwidth within platforms such as DGX Spark. These enhancements in KV cache and memory bandwidth bring cost efficiency by reducing reliance solely on GPU VRAM, while providing the capacity and performance needed for high-quality AI user experiences.	Talent Gap and aiDAPTIV+	-The AI talent gap is not going away, especially for organizations outside of Big Tech so we can't rely on hiring our way out of it. With aiDAPTIV+ and similar infrastructure, we are packaging years of storage and GPU expertise into platforms that are easier to deploy and operate. In 2026, that kind of turnkey foundation enables customers to focus their scarce experts on models and applications instead of wrestling with the underlying infrastructure.
Plakar	Production AI agents will necessitate continuous data protection	As autonomous AI agents execute complex operations in production, the industry lacks mature sandboxing standards to contain them. Without this safety layer, the risk of rapid data corruption becomes unmanageable. In 2026, the only viable defense is a shift to high-frequency, continuous backups. Organizations must be able to rewind datasets by minutes to survive inevitable agent errors and safely sustain AI adoption.	Ransomware hunting backups will enforce zero-knowledge storage	Attacks targeting backup software and infrastructure will intensify, aiming to eliminate the safety net first. To counter this, 2026 will make key separation the new standard. Backup infrastructure must become blind, holding encrypted data without ever having access to the decryption keys. This architectural guarantee will be the only way to ensure that a compromised backup system cannot be held for ransom.	The AI threat gap will give birth to Resilience-as-a-Service	With attackers using AI to professionally execute high-speed breaches, SMBs will find themselves outgunned. The operational burden of defense will become too high to manage internally. We predict 2026 will mark the true emergence of Resilience-as-a-Service (RaaS), a new model where the market begins shifting toward automated, outsourced providers to close the dangerous gap left by the cloud safety illusion.
PoINT Software & Systems	Software Made in Europe	Europe's digital future depends on trust, transparency, and technological independence. To strengthen Europe's digital sovereignty, software products "Made in Europe" will become highly important in 2026.	Homogeneous integration of tape storage in data centers	Advances in tape system integration, including compatibility with S3 object storage, metadata-driven access, and seamless connection to cloud workflows, will make S3-to-Tape systems increasingly popular in data centers as an additional storage class. Consequently, S3-to-Tape solutions will play a central role in data centers in 2026.	Geo-distributed archives based on tape technology	By 2026, the deployment of geo-distributed archives leveraging modern tape libraries is expected to accelerate across enterprises and data center environments. This development is driven by ongoing data growth, increasing energy and storage costs, and growing demands for data resilience and regulatory compliance.
Pure Storage	Organisations will stop using one solution for cyber resilience	The historical approach to cyber resilience (considering it as a stand alone issue, where one vendor can protect an entire company) will be put to bed. Organisations will move away from using point solutions and embrace the wider ecosystem as they understand they can't go it alone. An interconnected framework can help prevent a ripple effect when an attack happens - users should be able to identify and halt an attack. A properly integrated framework is vital to mitigate risk and speed recovery	AI and data sovereignty will drive massive cloud repatriation	The dual issues of AI and data sovereignty are driving concerns about where data is stored, and how organisations can maintain trust and guarantee access. In order to extract value from AI, it's critical for organisations to know where their most important data is and that it's ready for use. Data sovereignty concerns are driving organisations to reconsider their cloud strategy. Rising geopolitical tensions and regulatory pressure will shape nations' data centre strategies in 2026.	Terabytes per Watts should become an industry key metric	Efforts to reduce energy use have fallen down the political and business agenda in recent months, it remains a priority for some. Industry standards on the way data storage efficiency is measured should be updated and will almost always include energy per watt. Terabytes per Watts (TBe/W), which measures the amount of data stored per unit of energy, should be introduced. This is a relevant and clear measurement capturing real-world energy use, is simple, vendor neutral, and accurate benchmark
QStar Technologies	NVMe Shortages in 2026 will lead to optimization of already owned primary storage	Demand for NVMe drives and systems will significantly outstrip supply in 2026, leading to increased prices and longer lead times. Organizations will need to maximize their existing NVMe capacity by automatically migrating data to lower cost storage solutions, such as disk or tape. Either can be represented as file (NAS) or object (Cloud) and the use of tools and products to create an active archive environment will complement overstretched NVMe based high performance storage systems.	Ransomware attacks are increasingly sophisticated, targeting backup and requiring immutable storage	Ransomware attacks are on the increase. Backups are increasingly being targeted. To counter this, many cloud providers have added immutability as a feature to their products to prevent third-party encryption of critical data. However, the simplest and cheapest immutable technology is tape. Because LTO tape is a sequential technology, backup data can never be overwritten, only appended. In addition, it allows for a simple air-gap approach by physically removing media from the environment.	AI-Ready data storage infrastructures need cost-effective, secure and accessible archive solutions	AI systems massive datasets are heavily accessed for short periods and then can remain inactive for longer periods. Inactive data should be placed on secure, cost-effective storage, and be accessible when needed. Tape libraries, when integrated into a tiered system, will be more prevalent in data centers, as they can stream data at very high rates by using many tape drives in parallel. Tape also uses significantly less power, allowing maxed out data centers to use under-utilized space.
Savartus	The evolution of cold storage	Cold storage solutions will evolve to provide near-instant access (within seconds) to archived data, making it truly "active" rather than dormant. - Pete Paisley, Owner, MagStor	The rise of geo-distributed active archives based on S3-to-tape technology	By 2026, the deployment of geo-distributed active archives leveraging modern tape libraries is expected to accelerate across enterprises and data center environments. Advancements in tape system integration, such as S3 object storage compatibility, metadata-driven access, and seamless connection to cloud workflows, are transforming S3-to-Tape systems into geo-aware active archives. - Thomas Thalmann, CEO, PoINT Software & Systems GmbH	Companies that adopt active data archive solutions will be driven data challenges	Exponential growth in attack surfaces, vectors and points of entry Required recovery of minimum business operations without ransomware payment Regulatory enforcement that punishes non-compliance heavily Rising cost of infrastructure and energy Corporate sustainability mandates Increasing volumes of AI-derived data with long-term retention requirements
ScaleFlux	AI spending budgets exceed supplier capacity	The appetite for AI infra, as exemplified by announced spending plans from the US hyperscalers, Neoclouds, and even the US government (Stargate), is astronomical and growing. In 2026, the semiconductor supply chain will replace grid power as the gating Predictor in just how many billions of dollars can be spent on AI infra. AI infra demands for memory and storage components exceed the worldwide fabrication capacity for both DRAM and NAND Flash, impacting GPU cluster deployments.	Memory and Storage Innovation take center stage	Demand for HBM, DDR, and NAND Flash will far exceed supply, spurring revolutionary innovation in Memory and Storage components with CXL memory and StorageNext SSDs. CXL will help with effective DRAM supply by (1) enabling better utilization of DRAM through dynamic allocation and (2) enabling re-deployment of RDIMMs from systems being retired. Revolutionary StorageNext controllers will enable the fine-grained, highly parallelized data access that GPUs demand. Demos begin in 2026.	New year, new SSD density and performance records – 512TB and PCIe 6 eSSDs.	The first PCIe 6 SSDs were announced in 2025. 2026 will see several additional vendors announce and sample PCIe 6 eSSDs, setting new throughput and IOPs records for the AI and data center marketplace. Initial PCIe 6 drives will emphasize performance for deployment closer to GPUs and CPUs. In the High-Density swim lane, 512TB eSSDs designed with QLC will provide the greatest storage density in terms of both TB/rack unit and TB/watt.
Scality	Storage becomes a strategic differentiator, not a utility	With "AI token economics," organisations will start measuring cost per token - making storage efficiency, data-tiering and access patterns critical.	Hybrid & sovereign-cloud demand surges	As compliance, data-sovereignty and cyber-resilience become mandatory, businesses will seek vendors who can deliver scalable object storage, sovereign data control and audit-ready architectures.	New service opportunities for the channel	MSPs, resellers and distributors can shift from selling hardware alone to delivering higher-value services including storage strategy, data-governance, compliance-ready architectures and cyber-resilience offerings.



StorageNewsletter - Predictions 2026 - Recap Opinions and Ideas - Jan. 21, 2026

Company	Prediction 1	Prediction 1 Detail	Prediction 2	Prediction 2 Detail	Prediction 3	Prediction 3 Detail
SIOS Technology	Cybersecurity Will Redefine the Role of High Availability	The rising wave of cybersecurity threats is transforming how enterprises view HA clustering. In 2026, HA will not only be about achieving 99.99% uptime—it will also serve as a vital tool for maintaining security resilience. More organizations will use HA clusters to enable rapid, low-risk patching and updates, ensuring systems remain both highly available and protected against emerging threats.	Continuous Availability: The New Foundation for Trusted AI	AI and ML workloads will run more frequently on distributed clusters and GPU-intensive systems, where downtime creates costly disruptions. In 2026, IT admins will demand high availability solutions that simplify complex AI stacks and expose full visibility into data, storage, and node health. Continuous availability becomes a prerequisite for AI reliability and trust.	Observability Becomes Essential for Complex IT Environments	As IT infrastructures expand across on-premises, cloud, hybrid, and multi-cloud environments, visibility into application performance and health and interdependencies of the elements of the IT stack will become mission-critical. In 2026, observability will emerge as a key differentiator for HA solutions, allowing IT teams to identify and resolve issues before they impact uptime. The most successful HA platforms will provide deep insights across the full stack—from hardware to application layer.
Solidigm	Increased Focus on Efficiency	With the massive increase of data center growth based on the AI explosion, organizations will be looking for ways to further boost sustainability efforts without losing the value of necessary storage and compute power. Optimized locations for data center footprints with the smallest environmental impact, innovative solutions like full liquid cooling, high-capacity storage and other innovations will be top of mind in 2026 as folks look to make every watt and square inch count.	Reduced Reliance on HDDs	Storage-related focus and needs are shifting; there's no question about that. As AI continues to drive up storage capacity, data center architects must reevaluate what success their storage provides for the best performance across the increasing and various workloads. As AI continues to increase the need for warmer data in workloads being managed, more HDDs will be swapped for SSD storage related solutions.	Rack Level Innovation	We are seeing a complete networking rework and innovation in and between racks in addition to the evolution of how storage solutions are being structured, delivered, and integrated with compute for AI deployments. As we continue seeing this in 2026, we expect demands for storage to be further optimized and designed at the rack level, working together with compute and networking to deliver data to higher performance for AI workloads. As we move into the future, there's a dramatic change coming—be
Starfish Storage	Powerful and holistic data management tools will become strategic assets	The advent of AI means data silos will need to be eliminated on a grand scale. Consolidating data, sifting through what is relevant and what is not, will require powerful data insight, data movement, and collaboration capabilities. With the size and value of unstructured data increasing, questions about how to store it, who owns it, what data types are being created, what its value is, will require combining traditional storage capacity management and detailed data management capabilities.	The Great Data Clean-up is coming	There will be a hunt for poorly cataloged, orphaned, and expired data that is taking up significant storage space in an effort to free up space for AI initiative data. There will be a push to empower data creators to manage and clean up their data, to better manage valuable data assets, and match data value to storage media costs.	Data management tools will need to be storage agno...	Owners of multi-PB, multi-vendor file systems will need deep insights into their data, regardless of the diversity of their data ecosystem. Data management tools will need to track both data and associated metadata as it is ushered through AI and scientific workflows, from system to system, into tape libraries and cold storage areas, and back again as data is reused and re-explored. Single platform catalogs will give way to storage-agnostic solutions.
StorMagic	Cloud reliability risk will lead to more on-prem deployments	The AWS outage reminded everyone how risky it is to depend too heavily on centralised cloud infrastructure. When a single technical issue can disrupt entire operations at a massive scale, CIOs are realising that stability requires balance. In 2026, more organisations will move toward proven on-premises hyperconverged infrastructure for mission-critical applications at the edge.	AI shifts to the edge	Most AI conversations in 2025 have been about the cloud and all of the investment happening to build bigger and bigger datacentres. However, a shift is starting to happen for AI at the edge. Training large models in the cloud will remain essential, but inferencing and real-time decision-making are better handled closer to where data is created at remote sites. Enterprises will move to keep their most critical applications and information closer to home, on their own servers at local edge sites.	2026 will be the year of 'back to basics' for IT infrastructure	The year ahead will favour a shift toward simplicity, uptime and management. There are so many ways an IT team can architect and deploy infrastructure, and there has been a trend to build everything like you would a datacentre. This comes with a cost - overly complex systems that are difficult to manage and typically very expensive. The organisations that succeed will be those that figure out how to avoid downtime with simple and reliable on-prem infrastructure to run local applications.
StorONE	2026 Outlook: The All-Flash Fantasy Collapses	By 2026, the storage industry will face the full consequences of a decade-long illusion: the belief that flash could replace every tier of storage. Flash price inflation isn't a short-term hiccup, it's a structural reality. Cost ratios of 1:10 and higher against HDD are making All-Flash design a financial liability that most enterprises can no longer ignore.	2026 Outlook: The All-Flash Fantasy Collapses	Budgets are breaking. Major data initiatives are being paused, scaled back, or canceled entirely. Business leaders are discovering that "performance everywhere" comes with a tax they simply cannot pay. Meanwhile, data growth isn't slowing for anyone. AI workloads, digital transformation, compliance retention, cybersecurity, these forces are accelerating the demand for affordable capacity. And when storage becomes the bottleneck, innovation stalls.	2026 Outlook: The All-Flash Fantasy Collapses	2026 will be the year organizations admit a critical truth: performance must be focusedThe shift ahead isn't optional. The industry will rapidly abandon fragile, multi-system data-shuttling tiering schemes and demand unified intelligent tiering built into the storage fabric itself with one data path, one set of controls, and no operational ambiguity. The verdict is in: the All-Flash era was a luxury. 2026 begins the era of Hybrid Intelligence; sustainable, resilient, and economically grounded.
StorPool Storage	AI Adoption Will Continue to Rise Becoming Even Farther-Reaching	Some experts are predicting that the rise of AI will change the business landscape at least as much as internet adoption has. AI's use by businesses will increase, whether drawing insights from data warehouses, powering real-time changes in IT environments, replacing some jobs, and acting as a digital assistant for other jobs. AI watchdog groups will continue to raise red flags about the over-dependence on AI, while AI itself becomes ubiquitous.	Malware – Particularly Ransomware – Will Accelerate, Continuing to Be Businesses' Greatest Fear	The "bad actors" and criminals will step up their efforts to access personal data and to use ransomware for extortion. Data thieves will continue to set their sights on big targets, retailers, credit card companies, banks, and other databases where they can maximize their return by capturing as many users' data as possible Ransomware extortionists will continue to attack critical infrastructure, hospitals and other healthcare facilities, emergency services, and power grids.	The VMware Exodus Will Become a Stampede	Not all organizations are in a position to migrate from VMware to an alternative right now. Others still are covered by their previous support contracts and have been able to postpone their migration. Broadcom is not interested in supporting small and medium businesses, having opted to focus on large enterprises. Customers will see deadlines looming and accelerate their migration plans.
Swissbit	Start of the Memory Supercycle	The biggest allocation crunch in decades is hitting NAND and DRAM alike. NAND is essential for AI storage and edge systems; DRAM underpins bandwidth and speed for training and inference. Vendors are prioritizing HBM and high-end NAND while fast-tracking the phase-out of legacy nodes. Capacity is largely sold out into 2026, prices surge, and allocation remains difficult even at premium rates. Analysts call it a "Memory Supercycle" that reshapes and consolidates the market.	Memory becomes a strategic architectural pillar	Memory is no longer a commodity but a core design element for modern systems. AI-, edge- and hybrid-cloud workloads need modular, flexible, and energy-efficient NAND and DRAM that integrate with data-centric designs. Storage becomes a lever for performance, scalability, and power optimization. Security and lifecycle management – data-at-rest, firmware integrity – become key differentiators.	Edge computing rises (including on-device AI)	Edge computing remains a major growth engine as processing moves closer to machines, vehicles, and industrial equipment. This reduces latency, saves bandwidth, and drives demand for compact, efficient storage with high performance and strong write endurance. On-device AI reinforces the trend – inference shifts local, raising requirements for memory performance, security, and reliable update capability.



StorageNewsletter - Predictions 2026 - Recap Opinions and Ideas - Jan. 21, 2026

Company	Prediction 1	Prediction 1 Detail	Prediction 2	Prediction 2 Detail	Prediction 3	Prediction 3 Detail
Toshiba Electronics	AI and Cloud Drive Storage Growth in 2026	AI workloads, alongside the continued growth of hyperscale cloud computing and IoT, are set to be the primary drivers of storage demand. Global data requirements are forecast to reach nearly 20,000 exabytes by 2029, almost double the 2025 level, underscoring the need for scalable, efficient, and reliable storage solutions. The surveillance sector will also drive new requirements, as AI-powered video analytics and always-on, high-resolution monitoring demand both higher capacity and performance.	HDDs Remain Core to Storage Infrastructure in 2026	HDDs will remain the backbone of large-scale, cost-efficient storage. Standardisation around 3.5-inch SAS/SATA HDDs supports long-term infrastructure investments and simplifies integration and upgrades, limiting the adoption of alternative form Predictionors and interfaces such as NVMe, multi-actuator drives, and SMR. So much so that in 2026, high-capacity, 3.5-inch nearline drives are projected to account for 68% of total shipped capacity.	High-Capacity HDDs Set to Transform Data Storage in 2026	A major development is the verification of 12-disk stacking technology, which adds two additional platters to the standard 10-disk 3.5-inch stack and combines it with Microwave-Assisted Magnetic Recording (MAMR). This advancement enables higher-density drives while maintaining mechanical stability, in-plane accuracy, and reliability, paving the way for 40TB-class HDDs targeted for launch in 2027.
TrueNAS	Sub-30PB storage becomes commoditized	Storage systems below 30PB will see less differentiation and lower pricing power. Enterprises will expect strong reliability and support by default. Vendors that cannot differentiate above the commodity line will struggle.	Rapid growth of workload-specific solutions	Enterprises will shift toward turnkey, workload-focused designs. Buyers will prioritize complete solutions built for specific workflows. They will avoid assembling components and will demand prescriptive architectures.	Data sovereignty and sovereign AI rise in priority	Enterprises will elevate data sovereignty and sovereign AI requirements to board-level concerns. They will expect greater control of data, metadata, and AI pipelines. This will influence how storage platforms are evaluated and deployed.
Tuxera	Flash grows, and QLC is building its presence	Looking into 2026, the shift to flash drive technology will continue. Quad-level cell (QLC) offering higher density and power efficiency than spinning drives, which are becoming popular for general-purpose workloads, will push into the market further. NVMe is being used as a caching layers and to support higher numbers of RDMA workloads. Alongside these moves, Ultra-high-IOPS, and ultra-dense SSDs, will enhance GPU environments supporting the endless rise of AI and ML.	Balanced protocol structures for modern workflows	Storage platform acquisition decisions are driven by workflows, important Predictionors such as file vs block, and native client vs network protocol. This traditionally results in a narrow solution choice. In 2026, we will start to see cost-effective Unified Storage Systems that will allow data to move seamlessly, offering block alongside file with SMB and NFS all working in sync, meaning workflows are not restricted by protocol choice.	Cloud costs potentially push data back to on-premises, could hybrid be the default state?	With NVMe costs rising due to AI/ML demand, this in turn could push cloud storage prices higher. These hikes could force a data flow return to on-premises. For some users, this hybrid, or repatriation, approach could offer the savings needed to continue to be profitable. On the other hand, we could see CSP's stockpiling of media to combat the potential rise. Storage efficiencies will also drive a change in strategy. TCO is more important now than ever before.
Vdura	AI workloads demand intelligent balance, not just speed.	"By 2026, infrastructure will be judged not only on raw performance but on its ability to intelligently balance workloads across flash and hybrid tiers. Enterprises will expect platforms that can dynamically optimize cost while sustaining high speed throughput, ensuring that AI models scale without bottlenecks. The winners will be those who deliver parallel performance that adapts in real time to shifting demands." Ken Claffey, CEO, VDURA	Unified data fabrics replace storage silos	"The era of fragmented storage is ending. By 2026, enterprises will demand a single namespace spanning flash and capacity tiers, eliminating the inefficiencies of siloed systems. Intelligent orchestration will automatically move data to where it's needed most, whether for analytics, compliance, or AI training, creating a seamless data fabric that accelerates innovation while reducing operational overhead." Erik Salo, SVP of Product and Operations, VDURA	Operational simplicity becomes non negotiable	"Complexity will no longer be tolerated at scale. Infrastructure must deploy in hours, expand in minutes, and self optimize without manual tuning. By 2026, operational simplicity will be the baseline expectation, not a differentiator. Organizations will gravitate toward platforms that abstract away complexity, enabling IT teams to focus on outcomes rather than configuration, and proving that simplicity is the ultimate measure of resilience." Erik Salo, SVP of Product and Operations, VDURA
Versity Software	Data Growth Is Exploding, Pushing More Archives to Exascale	In 2026, data growth will continue to accelerate as AI training sets, scientific instruments, and rich media expand rapidly. As primary storage becomes increasingly expensive, organizations will shift more of their growing datasets into archives and find those archives reaching exabyte scale, driving adoption of software-defined architectures that scale linearly across disk and tape. Multi-exabyte environments will become increasingly visible.	Multi-Site Replication Becomes Mission-Critical Infrastructure	By 2026, multi-site replication will transition from a premium feature to a fundamental requirement for enterprise storage deployments. Driven by ransomware resilience, disaster recovery mandates, and increasingly stringent data sovereignty regulations, organizations will implement active-active and active-passive replication architectures as standard practice. At Versity, most new deployments now ship with native replication, so every object written to tape or disk is automatically protected.	The Rise of S3-to-Tape and Unified Archive Access	S3-to-tape workflows are becoming the foundation of modern archive architectures. As cloud egress and AI compute costs rise, organizations are moving cold and warm data into S3-compatible, tape-backed archives while retaining cloud for burst compute. POSIX access remains essential, driving demand for unified namespaces that present the same data through both S3 and POSIX without migration—delivering cloud agility with the economics and durability required for exabyte-scale preservation.
Wasabi Technologies	2026 Increased IT investment will be at the foundations of AI	AI is already embedded in daily business, and few technologies drive IT forward as strongly. Yet today's IT and data architectures still limit AI's ability to scale, preventing measurable returns. By 2026, the priority won't be buying more AI tools, but building an efficient, affordable data pipeline that turns experiments into value. The winners of the AI era will be those who invest smartly in storage, compute and networking.	Storage will be the front line of ransomware recovery	2026 will make it clear that cybersecurity is no longer just about keeping attackers out – it's about ensuring an organisation can recover quickly, stay compliant, and continue operating under pressure. In an increasingly sophisticated threat landscape, storage will become the front line of recovery. Businesses will need to rebuild resilience from the core, treating storage as the epicenter of continuity and ensuring primary and backup data cannot be compromised.	Businesses will need to look beyond data sovereignty slogans and get specific	Across Europe, hyperscalers are adjusting to growing data-sovereignty pressures by working with local partners, but it remains unclear whether these efforts truly deliver sovereignty. In 2026, organisations will need to focus on the specifics of their own operations to define what sovereignty means for them—where data is stored, how it moves, and who has access. Success will come when vendors and customers examine their sovereignty requirements and prioritise what matters most.
Western Digital	AI meets its infrastructure reckoning	The race to scale AI will collide head-on with the physical limits of power, space, and sustainability. Expect a paradigm shift that incorporates the concept of active archives energy-aware, cyber-resilient tiers where cold data moves from cloud and disk to modern tape systems that consume virtually no power at rest yet remain immediately accessible. Ted Oade, Director of Product Marketing, Spectra Logic	Cloud-based, active archives will no longer be thought of as secondary storage	We expect demands for instant access to archived content to only increase - perhaps double or triple over the next few years - as adoption of AI, analytics, threat hunting, media workflows, and compliance accelerate. Data has gravity and cloud-based archives are a way to balance storage costs with demand for accessibility and we expect demand for more active "always available" storage to grow unabated. - George Hamilton, Director of Product Marketing, Wasabi	AI infrastructure will demand smarter access to all data	Traditional architectures are struggling to keep up with the demands of real-time analytics, model training, and inference. What's emerging is a need for infrastructure that's not only high-performance, but also flexible enough to span edge, core, and cloud environments. To support the full AI lifecycle, systems must deliver consistent performance while also enabling intelligent access to archival data.
XenData	Hybrid deployments for large active archives	The use of hybrid configurations that combine on-premises storage and cloud will continue to grow. This is especially the case for large active media archives where on-premises storage provides high performance and cost-effectiveness and, when combined with cloud object storage, the solution provides a high level of data protection.	n/a	n/a	n/a	n/a



StorageNewsletter - Predictions 2026 - Recap Opinions and Ideas - Jan. 21, 2026

Company	Prediction 1	Prediction 1 Detail	Prediction 2	Prediction 2 Detail	Prediction 3	Prediction 3 Detail
Xinnor	Storage shortage will push companies to look for more efficient ways to get capacity and performance	Storage market faces unprecedented shortage from hyperscaler and Neocloud demand. The situation intensifies in 2026 with GPU farm deployments requiring more capacity. Organizations will adopt innovative approaches: deploying massive capacity drives including >100TB QLC, and improving usable capacity by replacing inefficient mirroring and replication with parity RAID protection to maximize every terabyte.	Flash TCO will become more and more compelling	With 250TB NVMe drives arriving, flash capacity advantage over HDD jumps from 3-6X, delivering major benefits in real estate (TB/server) and power consumption (Watt/TB). Organizations will consider flash beyond high-performance tiers for cold storage. Flash gains ground over DRAM as memory costs in AI projects soar, with organizations exploring flash alternatives to replace HBM memory wherever feasible.	Storage will become more relevant in AI project	Today the priority of any organization working on AI has been to secure GPU and most of the time storage has been an afterthought. With the explosion of data needed and generated by AI models, the critical role played by storage in maximizing GPU load and the on-going shortage that is increasing its cost, storage will become a priory element in an AI strategy. With that, there will be more attention on how storage is deployed in order to maximize the performance and minimize inefficiencies.