SOLUTION BRIEF · ADI

AI has broken the old storage model

One platform, four tiers, one S3 namespace for the enterprise AI data pipeline

Scality ADI (AI Data Infrastructure) serves enterprise AI, cyber resilience, and sovereign control from one platform. It aligns the right storage media, performance, protection, and economics to each stage of the data lifecycle, at multi-petabyte to exabyte scale.

There is no single AI workload

An AI program is not one job with one storage profile. It is a pipeline, and each stage asks for something different:

- **Training and validation.** Sustained, high-concurrency reads at the lowest achievable latency, so expensive accelerators are never idle.
- **Inference and KV cache.** Small, latency-sensitive access patterns where time to first byte governs the user experience.
- **RAG and vector search.** Large, frequently re-read corpora that must stay queryable, governed, and versioned.
- **Data preparation, retention, and audit.** Petabytes of raw and derived data whose value, and whose cost profile, changes over time.

Forcing all of that onto one architecture produces the same two outcomes every time: a new silo per requirement, or an all-flash estate priced for the hottest workload in the pipeline and paid for by every other one.

2.4%
of AI infrastructure spend goes to storage

18.6x
flash-to-disk price per TB

Source: Storage share: $2.2B of AI infrastructure spend in Q1 2026, IDC Worldwide Quarterly AI Infrastructure Tracker. Flash multiple: 30 TB TLC against 30 TB HDD, August 2026, VDURA Flash Volatility Index via Blocks & Files, 2026-08-11; the same index read 7.0x a year earlier. Basis: third-party published.

What Scality ADI is

Scality ADI became generally available on 2026-05-12.

RING is the engine, ScalityOS the chassis, Scality ADI the platform you drive. The engine is proven from petabyte to exabyte scale; the chassis gives every node, site, and version an appliance-like experience. ADI adds the four-tier architecture, autonomous operations, and a single S3 namespace over all of it.

One platform, four tiers, one S3 namespace

Applications do not change endpoints when data changes temperature.

PIPELINE STAGE	ADI TIER	MEDIA AND ACCESS	WHAT IT DELIVERS
Training, validation, KV cache	GPU-Direct	TLC flash, S3 and S3 over RDMA	Sub-50 microsecond latency, so accelerators stay fed
Data preparation and cleaning	Hot	QLC and NL-SSD, standard S3 plus RDMA	High-concurrency throughput for large-scale prep
RAG, vector search, AI data lakes	Warm	NL-SSD, NL-HDD, HDD	Geo-distributed capacity that stays queryable
Retention, archive, audit evidence	Cold	Tape, cloud	Capacity-optimized long-term custody

Source: Scality ADI solution overview, 2026-05-12; scality.com/adi. Basis: published product architecture.

Lifecycle tiering and hybrid-cloud replication extend the same namespace across on-prem and the major clouds, and to tape, so retention economics are a policy decision rather than a migration project.

The four pillars

- **AIConnect.** High-concurrency S3 and S3 over RDMA data paths for GPU-Direct, training, inference, RAG, and KV cache workloads, at sub-50 microsecond latency. RDMA runs across the full media ladder, on NVMe and on HDD-based servers, and the HDD path is validated and published.
- **MultiScale.** Capacity, performance, and concurrent users grow on independent axes, so no one dimension throttles the rest.
- **CORE5.** Five layers of cyber resilience by design: ransomware-proof writes, geo-replicated recovery, and audit-ready evidence at every tier, not bolted on at one.
- **Guardian.** An AI agent for operations, usable built-in or integrated with your own AI tools and automation workflows over MCP.

© 2026 Scality · scality.com

What autonomous means here

Operational intelligence with bounded, policy-governed execution. Not a black box, and not self-driving. Guardian surfaces insights and recommends actions; humans approve and control every decision. MCP-enabled extensibility lets you integrate your own AI tools and automation workflows directly into ADI operations, so the platform can be driven by your AI stack and not only Scality's, and the same human-in-the-loop principle governs every operation. The system executes inside policies you define. Every action is logged. That is what makes autonomy survivable in a security review.

Source: Scality ADI launch press release, 2026-05-12.

AI-scale performance that survives the budget review

Performance is the entry requirement. What decides the second phase of an AI program is whether the infrastructure still makes sense once the pilot becomes an estate.

- **Power, not just capacity.** Workload-aligned tiering delivers up to 20x better power efficiency than all-flash for typical enterprise data distributions. When the next refresh is capped by megawatts rather than terabytes, that is the constraint that matters.
- **Flash where flash earns it.** Only the stages that need TLC pay for TLC. The rest of the pipeline sits on the media its access pattern justifies, inside the same namespace.
- **Resilience as architecture.** CORE5 puts immutability, erasure coding, secure replication, and audit-ready recovery at every tier. The RING engine underneath carries a publicly stated 14 nines of durability and zero RPO and RTO.
- **No new silo for backup.** ADI is pre-integrated with Veeam, Commvault, and Rubrik over standard S3, so protection lands on the same platform as the AI data it protects.

Proof at scale

>12 EB
in production
Across 70+ countries.

>6T
objects managed
One S3 namespace.

1000+
enterprises in production
In continuous production since 2009.

420
GB/s read
Sustained, with 230-250 GB/s write, on just over 100 x86 nodes.

Source: scality.com/adi; scality.com/products/ring; solved.scality.com HDD validation, 2026-05-26; Scality ADI solution overview, 2026-05-12. Public substantiation only. Basis: measured for the throughput figures; installed-base counts as published.

Where to start

ENTRY POINT	WHO OWNS IT	WHERE IT PAYS OFF
AI data flow	Chief Data and AI Officer, Head of AI Infrastructure or HPC	AI and innovation budget
Cyber resilience	CISO, security architects	Security budget
Sovereign and regulated	Chief Compliance Officer, General Counsel, CISO	Compliance and risk budget
Sustainability	CFO, procurement, ESG	ESG and finance budget

SCALITY

Bring one AI workload and its two neighbours in the pipeline.

We will map each to the tier that serves it, and show what the same estate costs when every stage stops paying flash prices. Talk to your Scality account team or visit **scality.com/adi**