Modern Storage for Modern HPC and AI Environments

Ben Trinder
XENON Systems

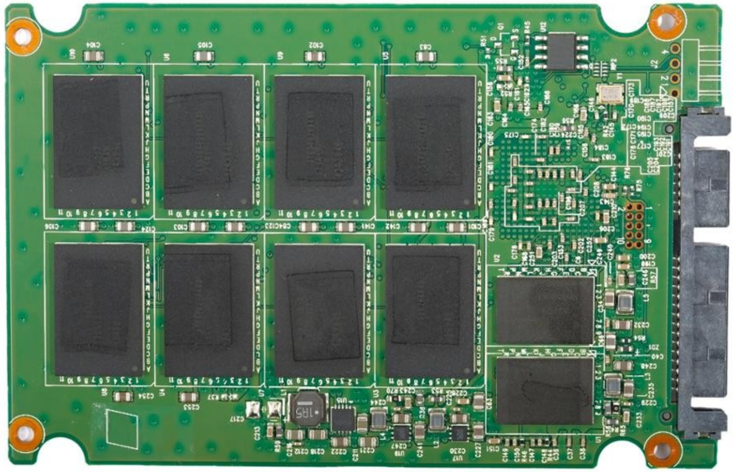
Once upon a time, data storage was simple.

You would buy some hard drives, connect them to your server via a SCSI cable, and life was good.

Then came along the need to share the storage, so we got Fiber Channel switches and arrays.

But soon it wasn't enough to just share the raw disks, you wanted searchable and shareable files, so we added metadata controllers and indexing servers.
And for a time, life was good again.

One day, solid state disks arrived on the market.



NTBscanpix

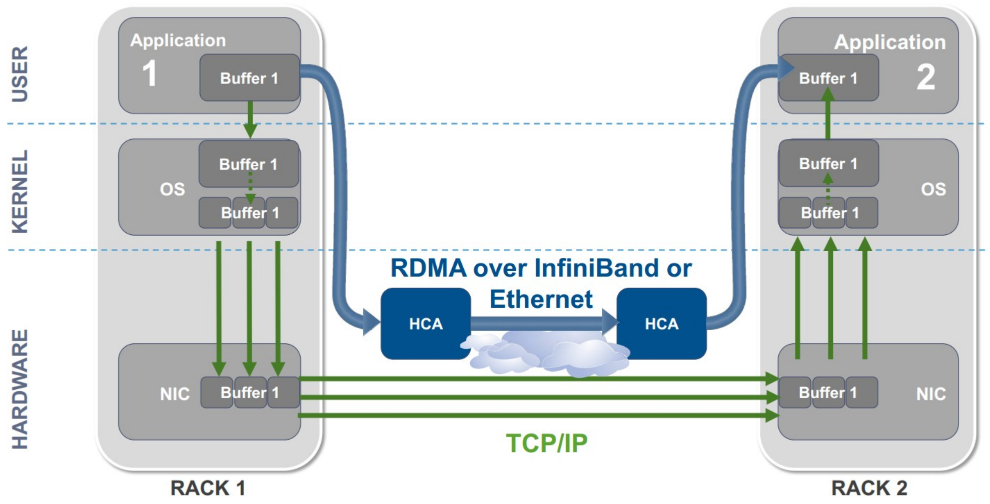
So how do you get from a SAN that uses a dedicated storage fabric in a single rack of compute, to a storage platform that is scalable, available anywhere, performant and reliable, within budget?

A few key concepts for building a modern storage platform:

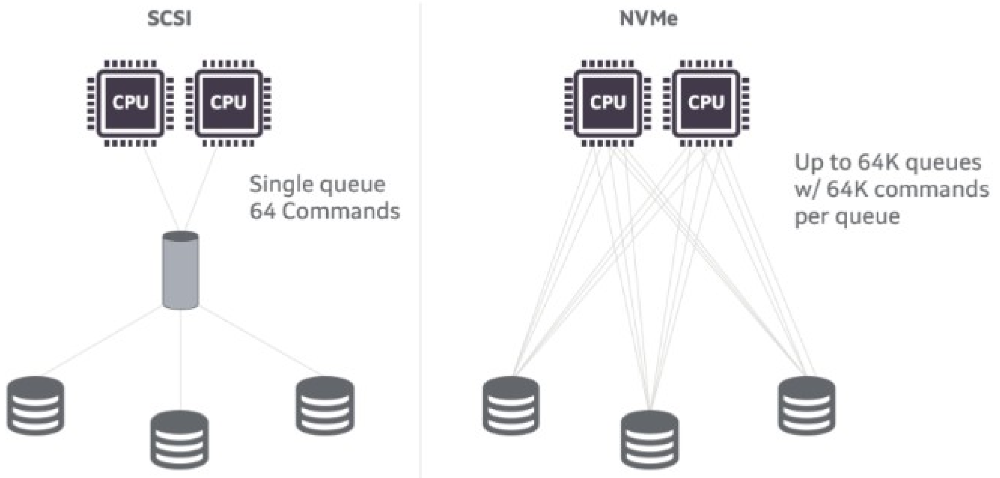
SSD : Solid State Storage – Non-volatile flash memory, NAND chips provide high density storage at relatively low cost.

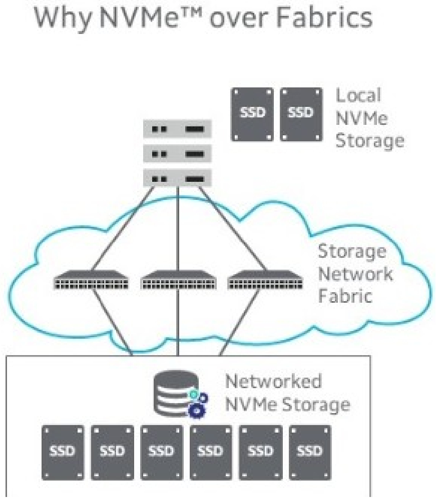
A few key concepts for building a modern storage platform:

RDMA : Remote Direct Memory Access



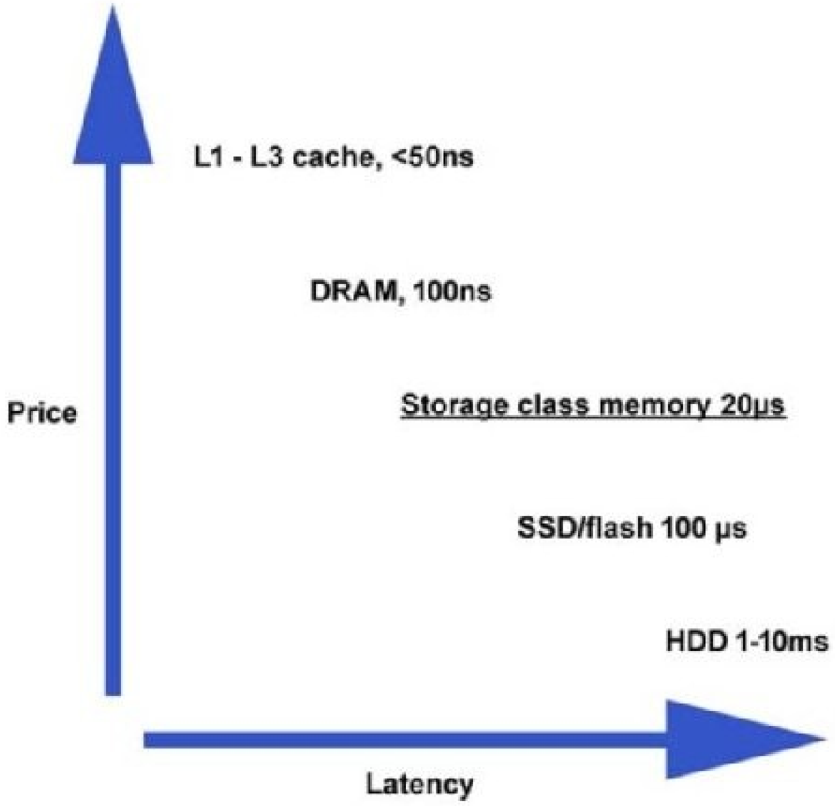
NVMe : Non-volatile Memory Express





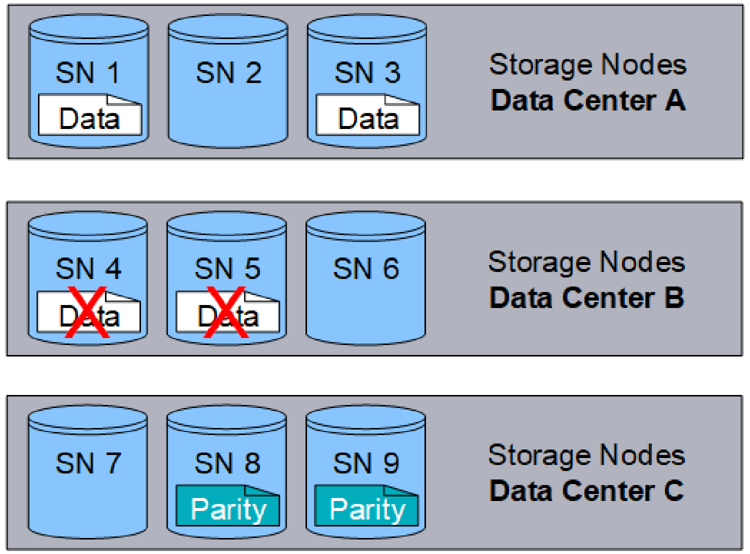
A few key concepts for building a modern storage platform:

DRAM : System Memory – very fast, dynamic storage that does not retain information without power.
SRAM : Cache Memory – super fast, small storage that is ideal for CPU caches.
Storage Class Memory : A type of non-volatile memory that combines the performance of DRAM with the persistence of NAND Flash.

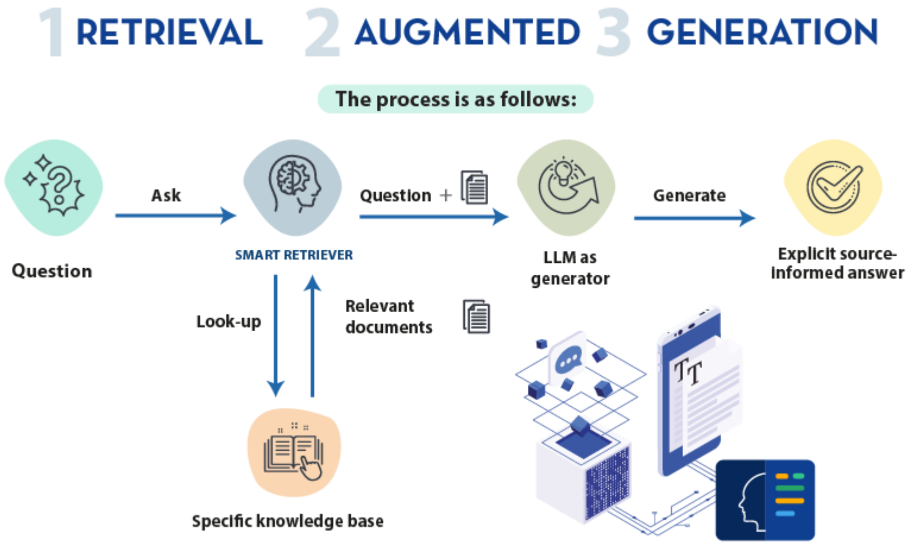


A few key concepts for building a modern storage platform:

Erasure Coding



RAG : Retrieval Augmented Generation



Why do we need to change?

1. Searchability : While a large enough array of spinning disks can provide very impressive data throughput, the simultaneous IO latency severely limits the ability to quickly index the entire storage pool.

2. Redundancy : Traditional RAID controllers require a “rebuild” any time a disk failed. With disk sizes growing and growing, the amount of data that needed rebuilding became impractical.

3. Geo tolerant designs : Business demands require guaranteed availability of data. A single rack. A single data centre. A single city. All can go offline. Storage needs to span multiple geographic regions, and perform even if a city is offline.

4. Users on the Edge of the Internet : Gone are the days where users of the storage would sit at a powerful workstation within 25m of the storage hardware. Researches want to collect data directly in the field and start processing immediately. Teams collaborate on data from all over the world.

5. AI workloads : Up to 57% of traffic on the Internet in 2025 is now AI bots. Traffic on private networks is increasing for AI. Organisations are implementing agentic AI to search through databases, customer records and research data – creating significant load on storage resources on top of the regular human traffic

What does a modern storage platform look like?

- **Fast** : utilising a mix of Storage Class Memory and NVMe SSDs across a number of nodes. IOPS priortised or raw throughput. Sometimes you need both.

- **Highly available** : using Erasure Coding, a large number of smaller nodes, multiple client entry points and geo-spreading

- **Scalable** : architecture that allows additional performance and/or capacity to be added, without downtime or a big performance hit. Sometimes you need to scale both up and down also.

- **Compute ready** : processing power directly in the storage platform to allow databases, AI agents, and other services direct access to the data

- **Cloud native** : storage software that can run both locally on bare metal as well as in public cloud, to allow seamless access to data wherever it is stored – typically built on Kubernetes.

- **Extensible** : Just being fast isn't enough, users need access to all data, in the one platform. Intelligent tiering to spinning disk, and yes, even to tape, ensures the storage platform can grow without costing the earth

How can YOU build a modern storage platform?

There is no single holy grail product, but the technologies we've discussed allow you to build a storage platform that meets your individual budget, performance specs, capacity and growth needs.

Start with:
- How much capacity do you need today?
- What is your annual growth rate?
- Where are your users located?
- How does your data get generated?
- What performance profile do your applications require?
- How often is historical data accessed?
- Do your storage requirements change throughout the year?
- Do you need a native filesystem client for GPUDirect or similar, or is NFS+SMB suitable?
- Do you need any block storage presented to clients?
- Can your data benefit from de-duplication and compression?

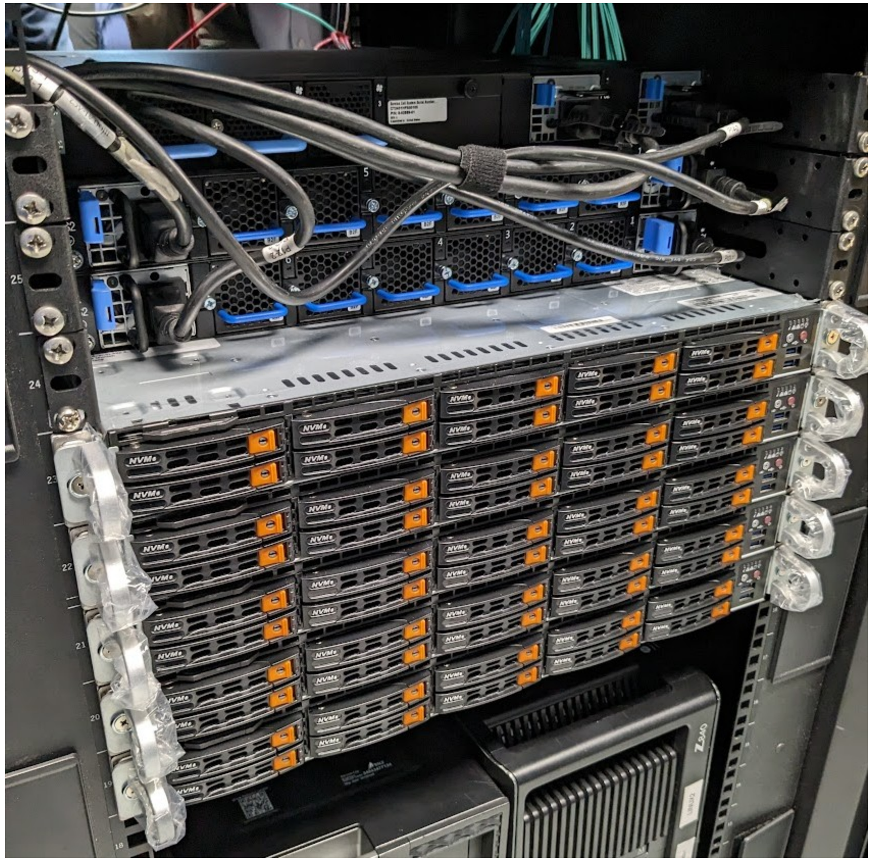
Think about your pain points:
- How much time do researchers spend waiting to access data?
- How often is work repeated due to inability to access data?
- Do your researchers have access to data in all the places they need to use it?

XENON High Performance Computing

Broad stroke types of modern storage

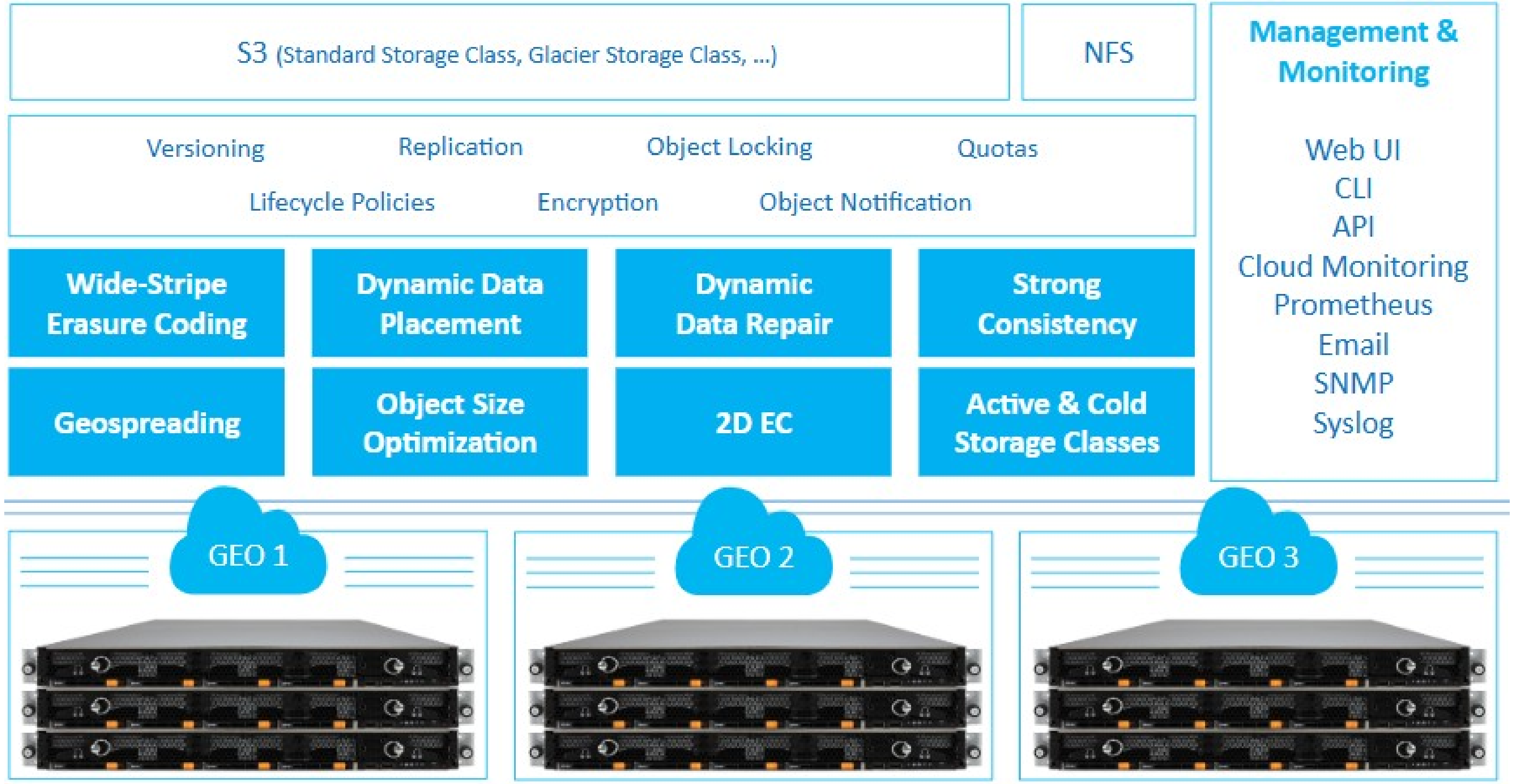
Super fast local flash array - use case : HPC clusters.
Fastest possible NVMe storage with SCM and 100G up to 800G fabric switches, compute nodes in the same rack or DC.

Examples : VAST Data Platform, Quantum Myriad, DDN Infinia



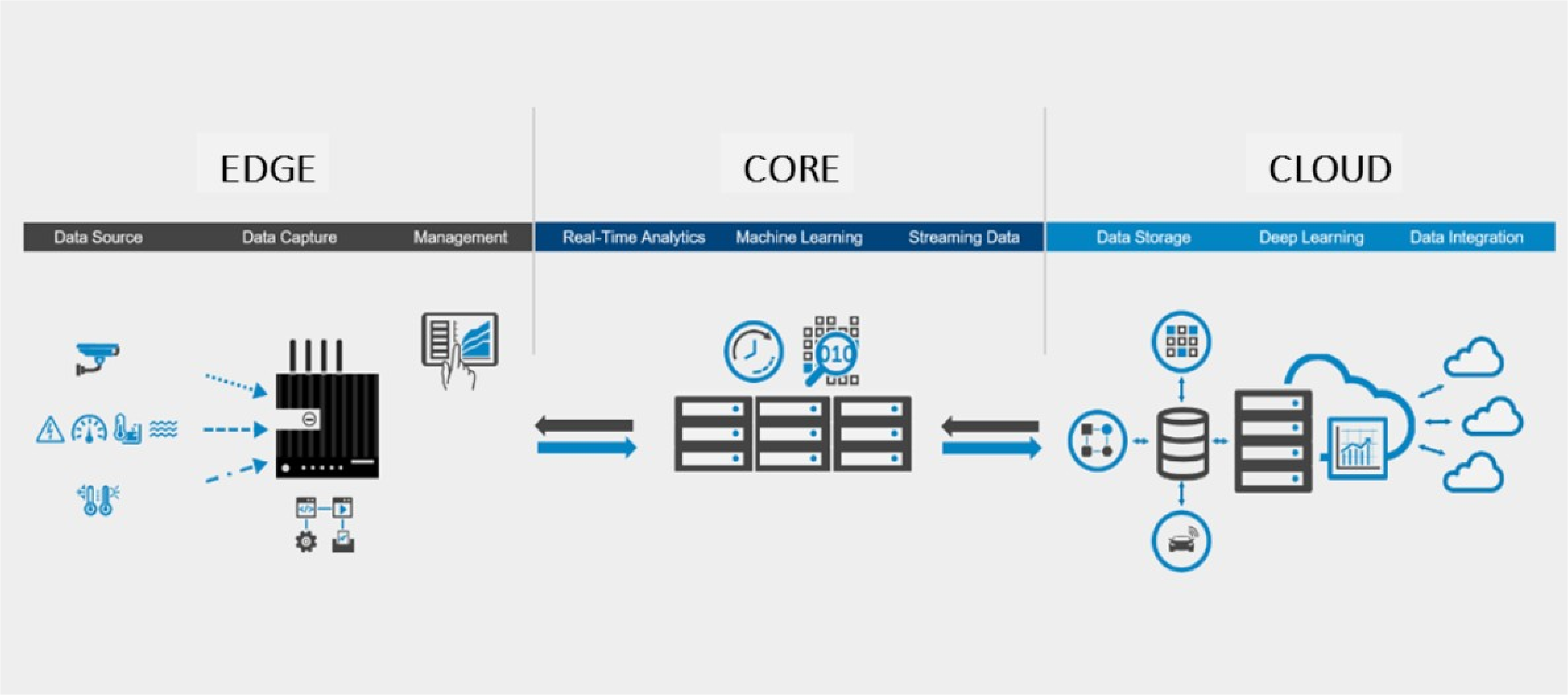
Broad stroke types of modern storage

Geo-redundant object store – use case : Web applications or backup targets.
Very large data capacity, equal performance regardless of user location, 99.9999999% uptime, 10G-25G Ethernet.
Examples : Quantum ActiveScale, Azure GRS, Scality RING



Broad stroke types of modern storage

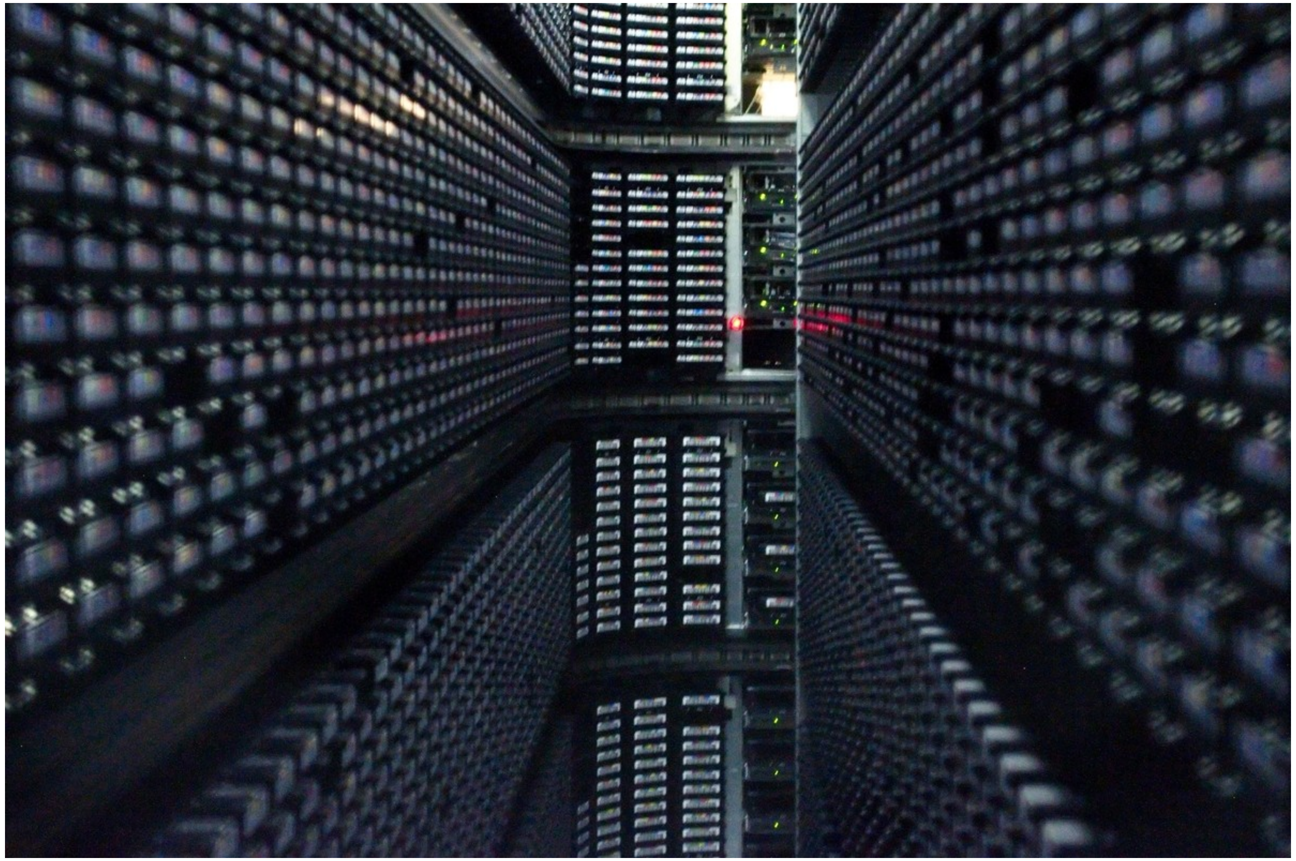
Software-defined storage – use case : flexible data management with unknown future needs.
Start your filesystem in the public cloud, pay by the month. Grow into local storage without downtime.
Examples: WEKA Data Platform, Netapp ONTAP, Hammerspace



Broad stroke types of modern storage

Archive platform - use case: compliance retention of 10 years worth of research.
Larger organisations with multiple user types. Bolt onto a super-fast flash array or other storage type, includes automation to offload data onto more cost effective, higher availability tiers.

Examples: Versity ScoutAM, HPE DMF, IBM HPSS



XENON High Performance Computing

Real world example : HPC Research Cluster with Archive

HPC compute cluster:
- VAST Data Platform : C-Box Compute, D-Box Storage Nodes with SCM and TLC drives
- Versity ScoutAM with Extended Cache and Spectra Logic LTO Library

Explicit Archive vs Implicit Archive policy
- Explicit : users manually move completed jobs from VAST to Versity hot disks
- Implicit : policy rules automatically move aging jobs from hot disks to tape

Disaggregated design:
- Compute, Flash, Hot Archive and Tape Archive can all be scaled separately

Geo-redundant design:
- Versity policy rules create additional copies of data in an offsite archive

Security:
- All internal, isolated network

Limitations:
- No EC across geos : wasted capacity due to 1:1 replication
- No cloud capacity : cannot burst into cloud for short period use

Thank you!

Any questions?

Images in this presentation – all CC BY-NC-SA 2.0
https://commons.wikimedia.org/wiki/File:Buffalo_HD-H250U2.jpg
https://www.flickr.com/photos/pchow98/7114323667 - IBM System Storage DS8800 (Inside)
https://en.wikipedia.org/wiki/File:Telvueprincetonstackedservers.jpg#/media/File:Telvueprincetonstackedservers.jpg
https://ndla.no/nn/r/driftsstotte-im-itk-vg2/lagringsenheter/d8bd4af7c5 - Péter Gudella
Quantum Myriad product photo – Ben Trinder
https://www.rawpixel.com/image/6042705/photo-image-tape-public-domain-technology - Interior of StorageTek tape library at NERSC
HPC Archive product photo – Ben Trinder