

IBM Storage Technical Strategy and Trends

9.3.2016

Dr. Robert Haas
CTO Storage Europe, IBM
rha@zurich.ibm.com



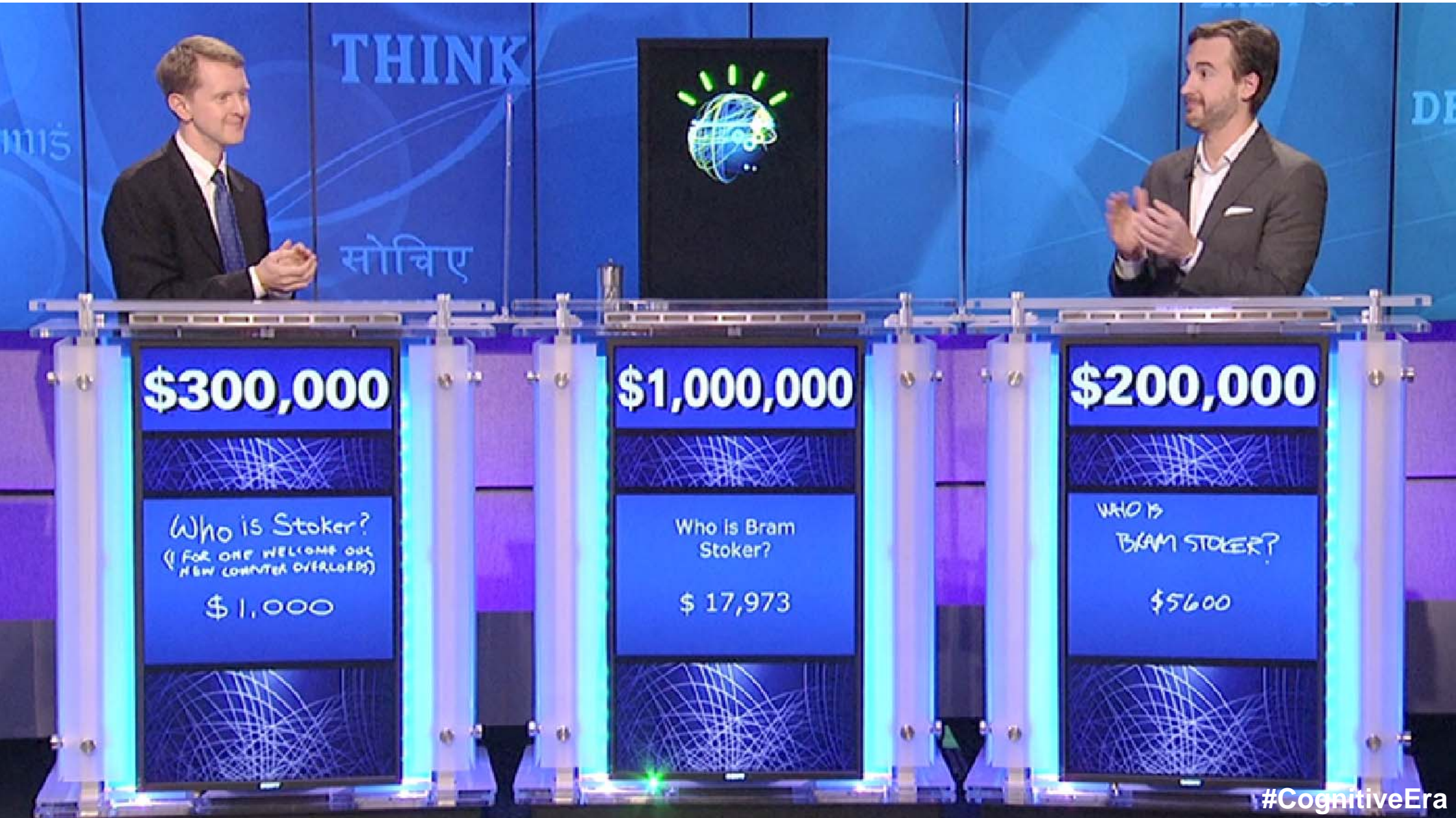


Cognitive Computing: Technologies that will change the business and the needs in terms of data storage

The digital world is generating oceans of data.

Organizations need to fish this ocean for insights.

The ocean of data must be efficiently stored, managed, and protected as well as used by the right applications at the right time.



THINK



सोचिए

\$300,000

Who is Stoker?
(For one welcome out
new computer overlords)
\$1,000

\$1,000,000

Who is Bram
Stoker?
\$ 17,973

\$200,000

WHO IS
BRAM STOKER?
\$5600

#CognitiveEra

The Watson that competed on *Jeopardy!* in **2011** comprised what is now a single API—**Q&A**—built on **five underlying technologies**.

Since then, Watson has grown to a family of **28 APIs**.

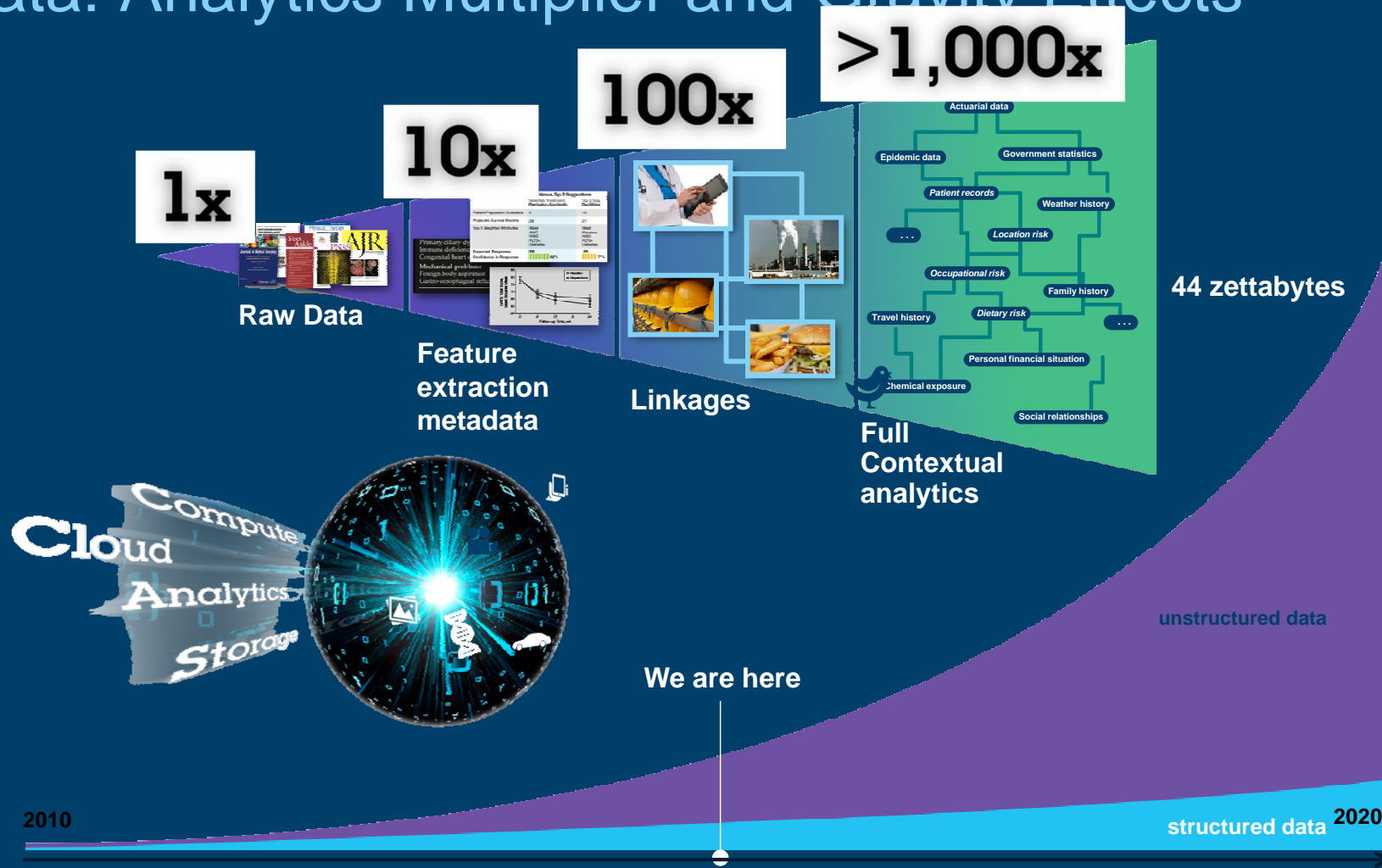
By the end of 2016, there will be nearly **50 Watson APIs**—with more added every year.

Natural Language Processing
Machine Learning
Question Analysis
Feature Engineering
Ontology Analysis

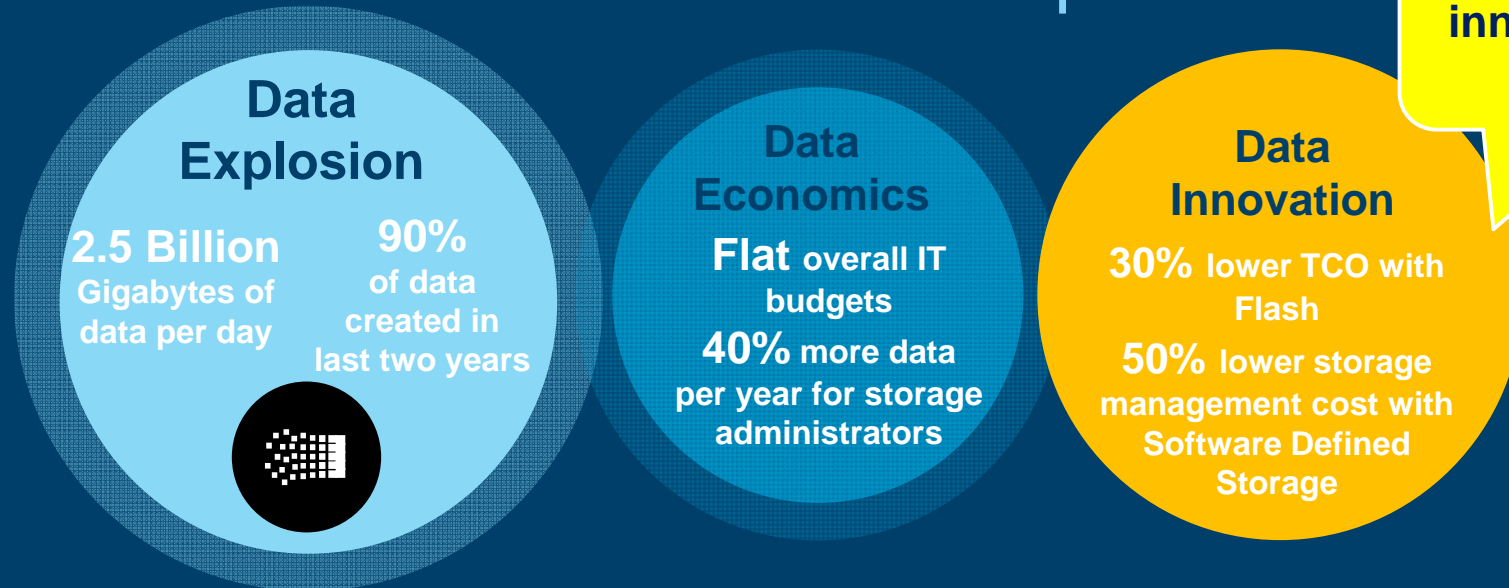


#CognitiveEra

Big Data: Analytics Multiplier and Gravity Effects



The Current Storage Model is being Disrupted by the Explosion of Data and Need for Speed



Unleash the power of innovation to solve this equation

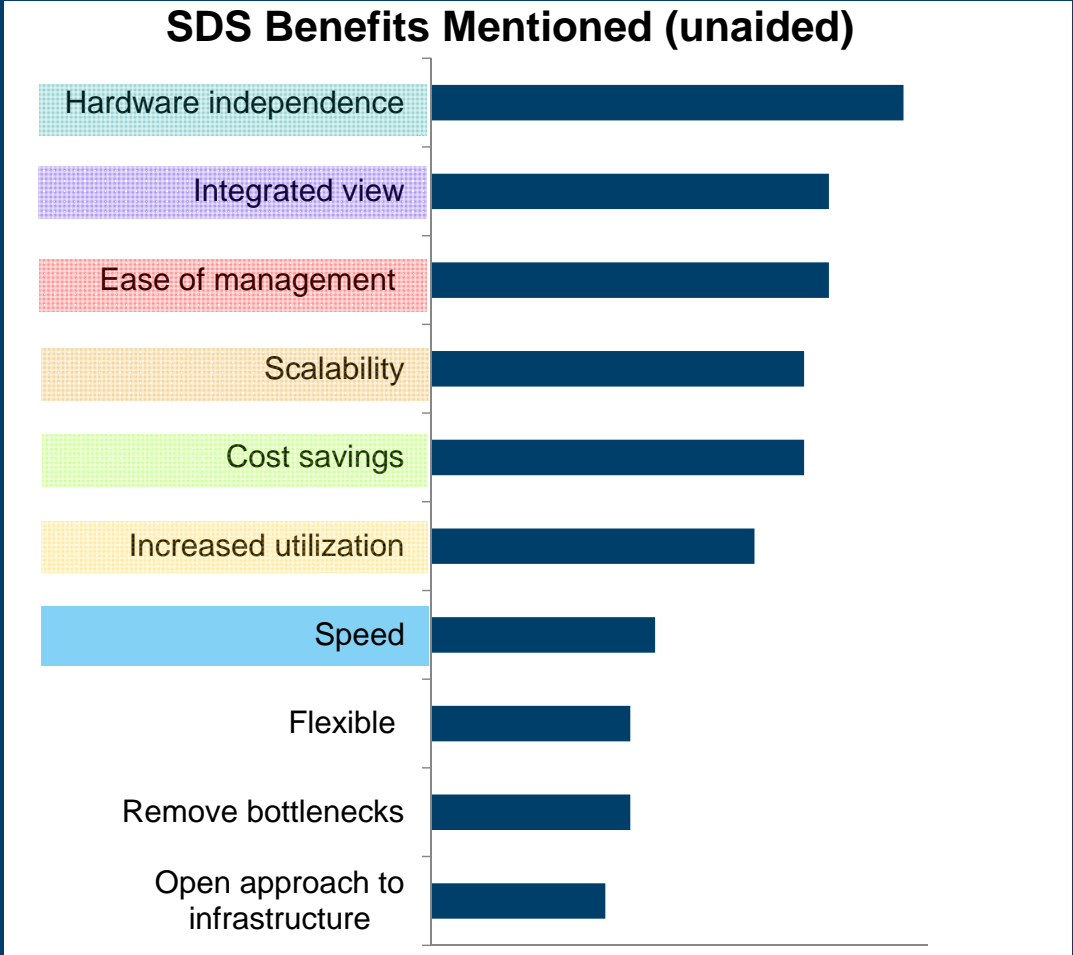
The top two challenges organizations face with IT infrastructure are storage related – Data Management and Cost Efficiency

SOURCE "IT Budgets 2016: A CIO's Guide"

Expected Benefits of Software-defined Storage

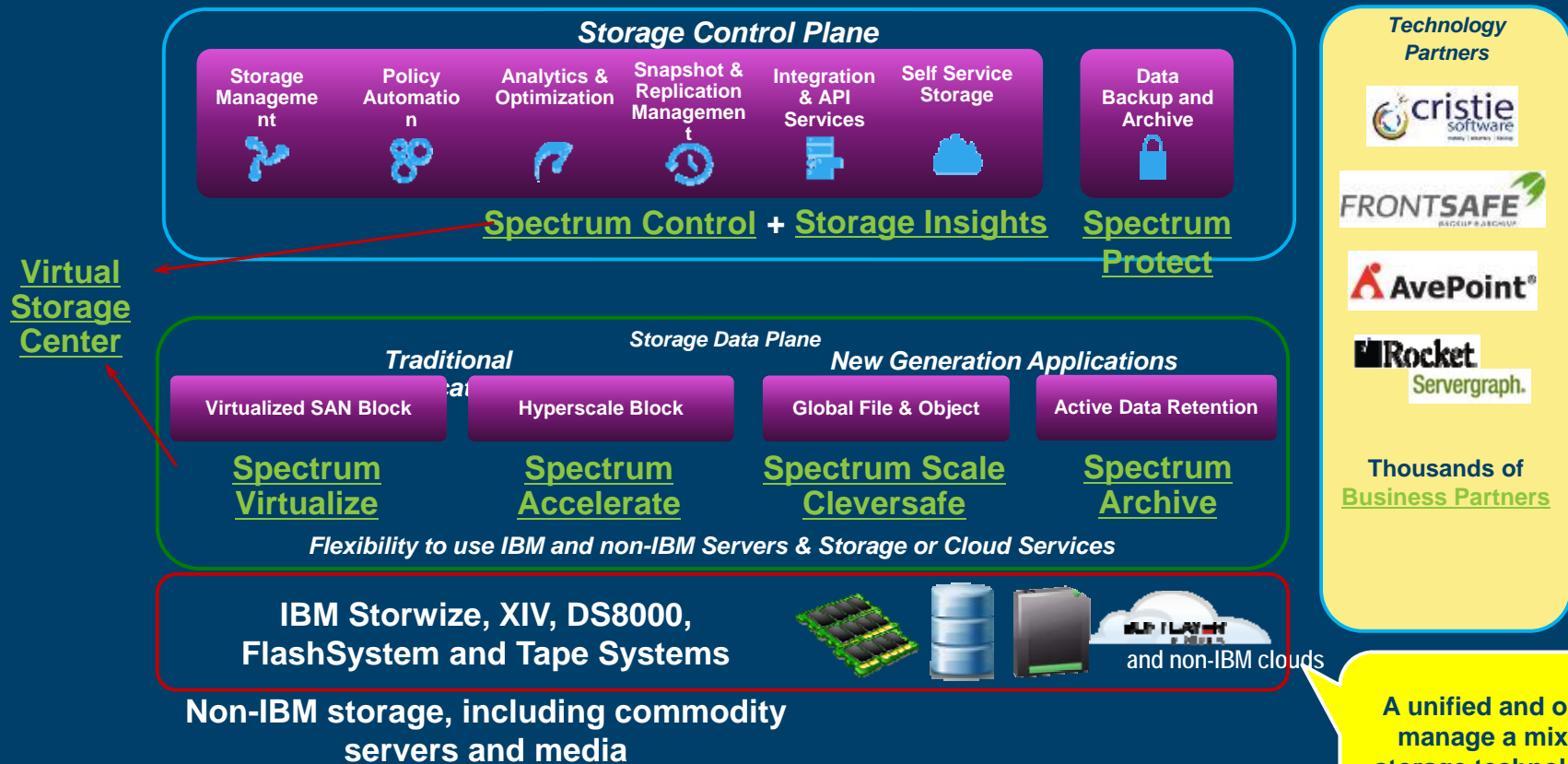
– Client Survey

SDS is seen as the architecture model for storage innovation



Software Defined Storage - Client Survey Findings

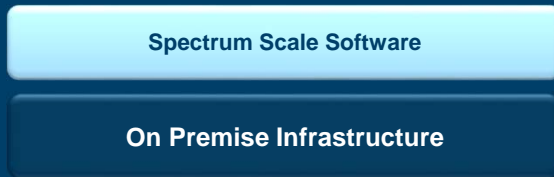
SDS Unified Control Plane Supporting Many Data Planes



Proof Point of Spectrum Scale

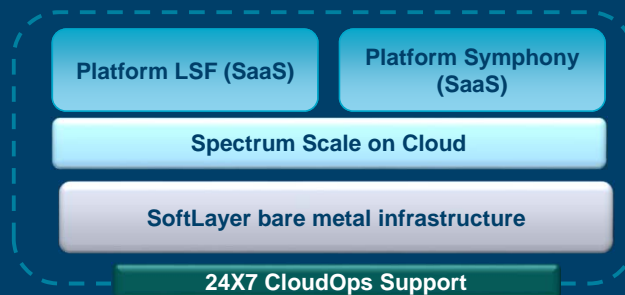


Integrated Offering **IBM's Elastic Storage Server (ESS)**



Software only

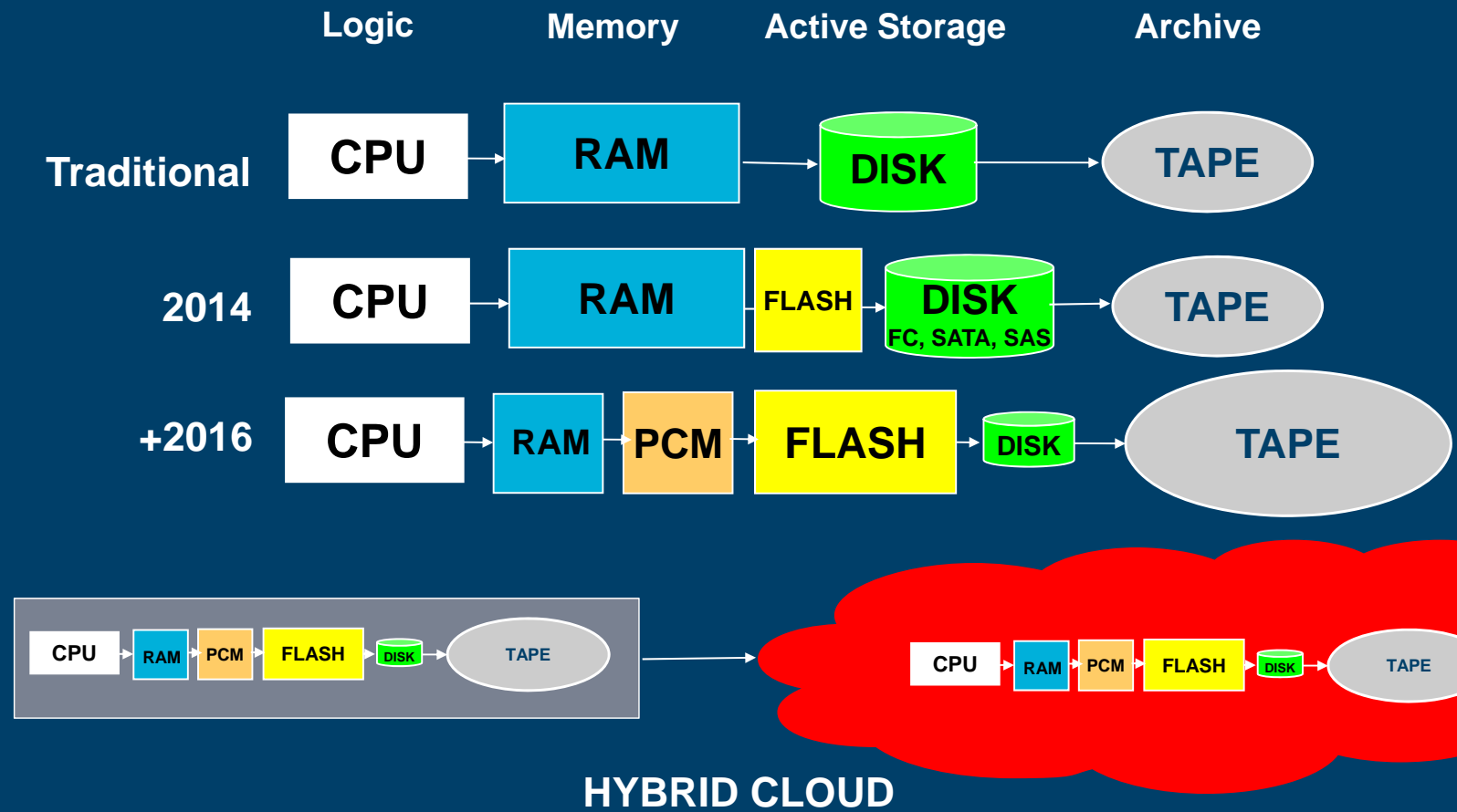
**IBM Spectrum Scale Software, License choices –
Express, Standard, Advanced**



on IBM SoftLayer Cloud
IBM Spectrum Scale on Cloud

Cloud Service
Ready to use, Spectrum Scale on the Cloud

Evolution of the Storage and Memory Hierarchy



Taking Advantage of Hybrid Cloud Storage

1. **Cloud as Remote Storage**
2. **Single Storage View**
3. **Single Workflow View**



Any Storage



Flash Systems

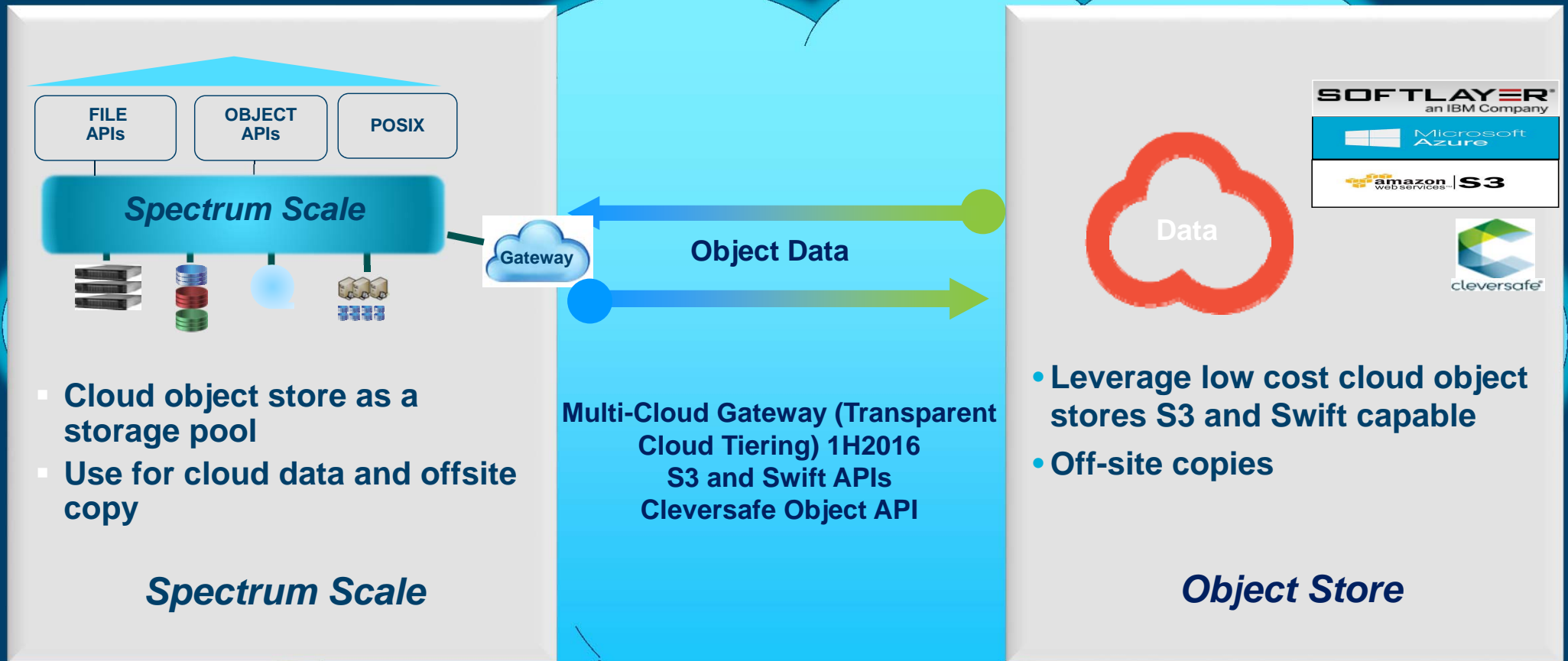


Private, Public
or Hybrid Cloud

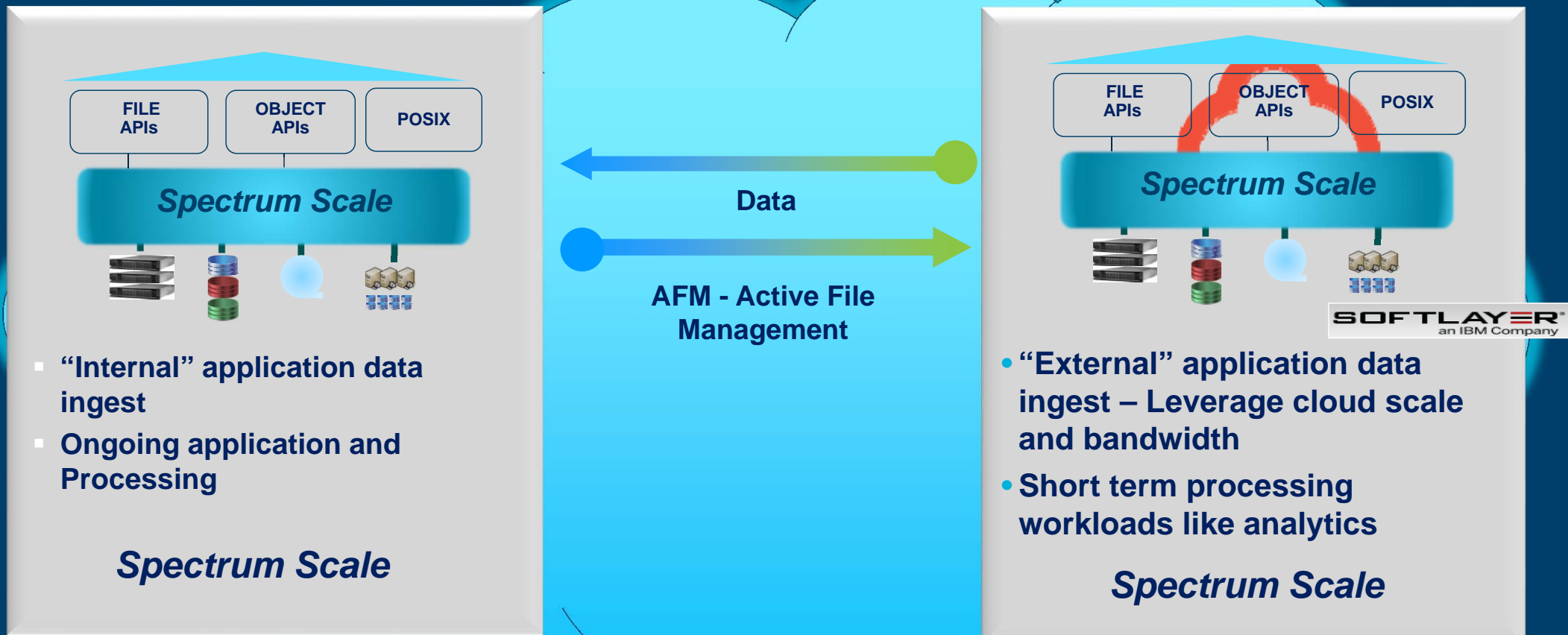
Control	Analytics-driven data management to reduce costs by up to 50 percent
Protect	Optimized data protection to reduce backup costs by up to 38 percent
Archive	Fast data retention that reduces TCO for active archive data by up to 90%
Virtualize	Virtualization of mixed environments stores up to 5x more data
Accelerate	Enterprise storage for cloud deployed in minutes instead of months
Scale	High-performance, highly scalable storage for unstructured data

1. Cloud as Remote Storage

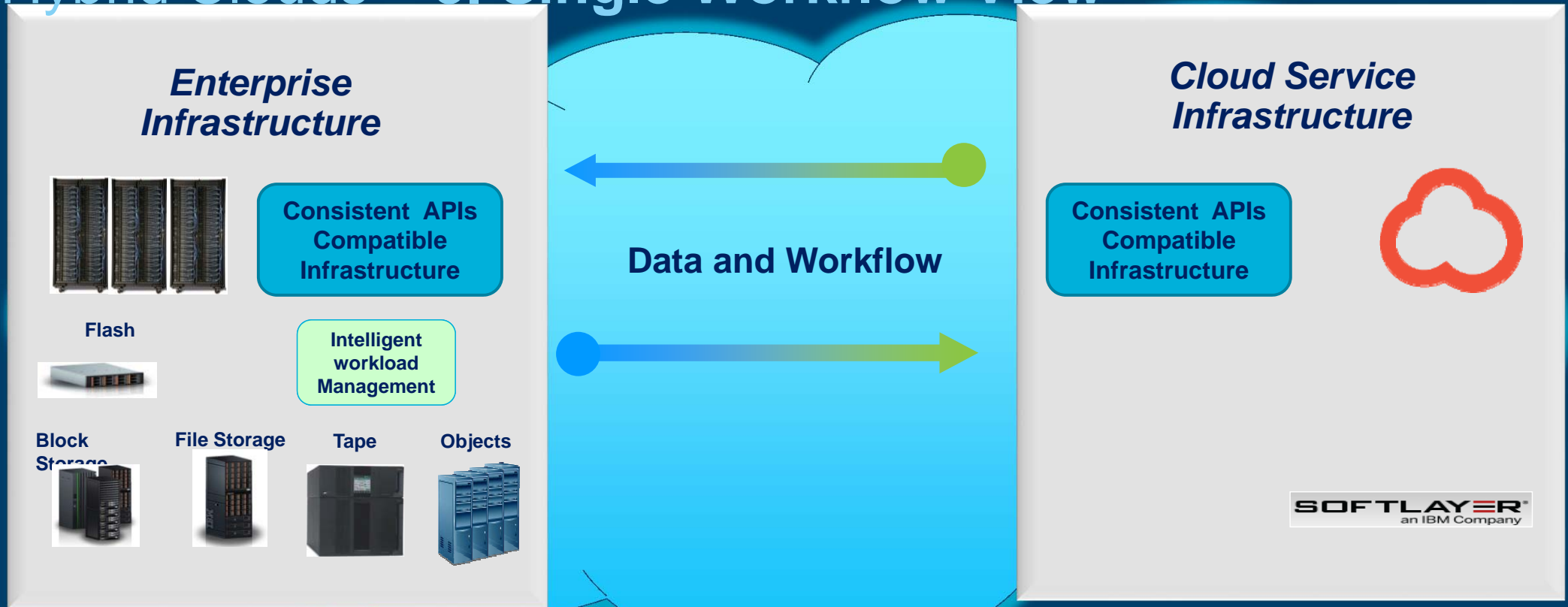
Spectrum Scale Storage Pool in the Cloud



2. Single Storage View - Spectrum Scale as Hybrid Cloud Storage



Hybrid Clouds – 3. Single Workflow View



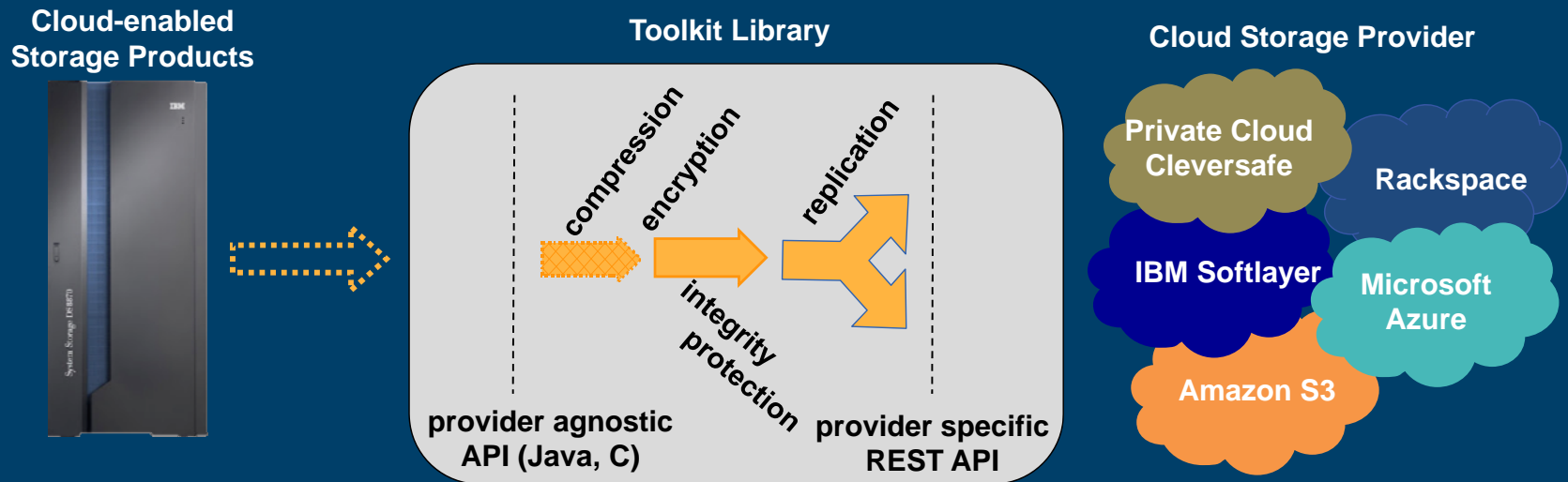
1. Consistent APIs across local IT, private and public cloud.
2. Intelligent data distribution and migration across on-prem, private and public cloud
3. Intelligent workloads management across on-prem, private and public cloud

Multi-Cloud Storage (MCStore) Toolkit

What: A software-defined enterprise cloud storage **gateway**

Why: Address customer concerns regarding cloud **security**, **resilience**, and **vendor lock-in**

Goal: Enable existing storage products to natively support public/private cloud storage



A Software-Defined Enterprise Cloud Storage Gateway for:

transparent data migration

backup and disaster recovery

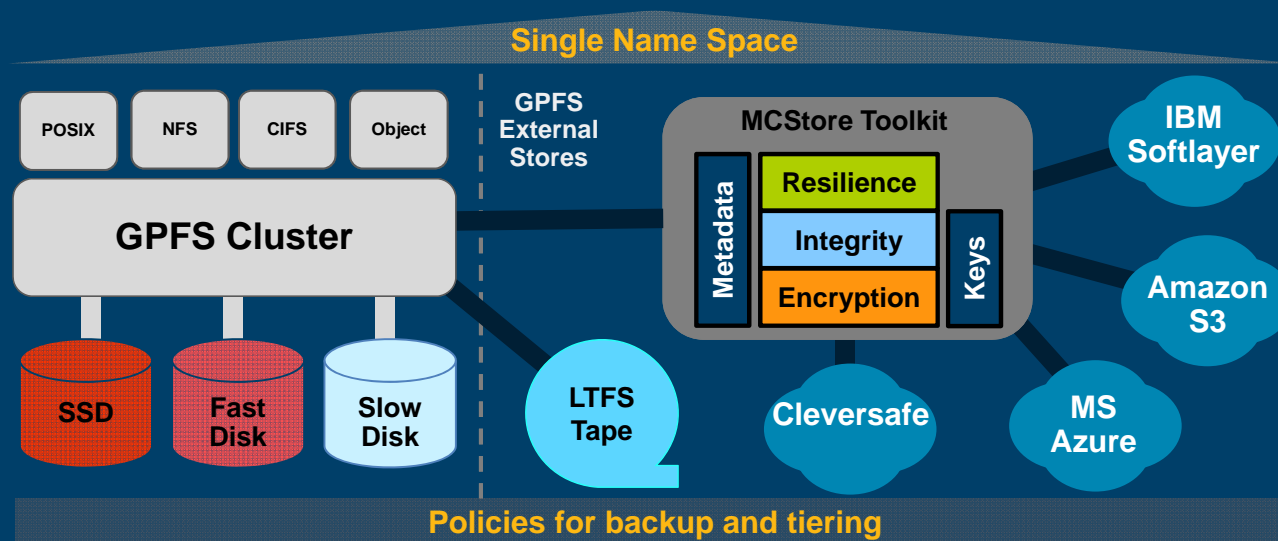
security and high availability

Transparent Cloud Tiering for Spectrum Scale

Goal: Enable a **secure, reliable, transparent** cloud storage tier in Spectrum Scale

Motivation: **Manage data growth** by placing file data

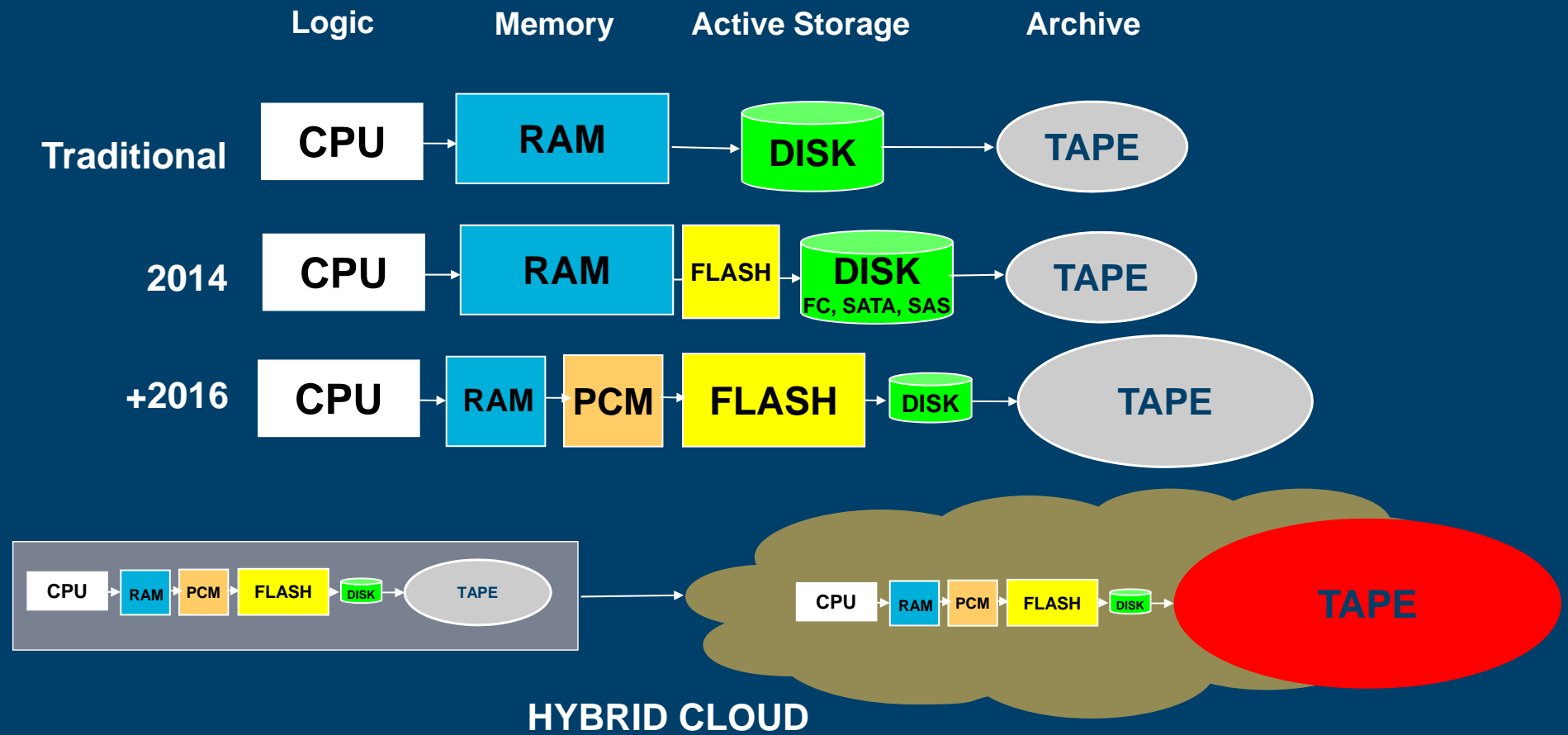
- in the right tier at the right time according to its value
- while being available under one **common name space** at any time
- leveraging the **economy of scale** of the cloud



Value

- Seamless file migration between local disk and cloud
- File system backup for DR
- Efficient data sharing between remote clusters
- Multiple cloud storage tiers (using multiple cloud providers)
- Run workloads locally or in the cloud

Evolution of the Storage and Memory Hierarchy



IceTier: Integrating Tape with OpenStack Swift

Goal

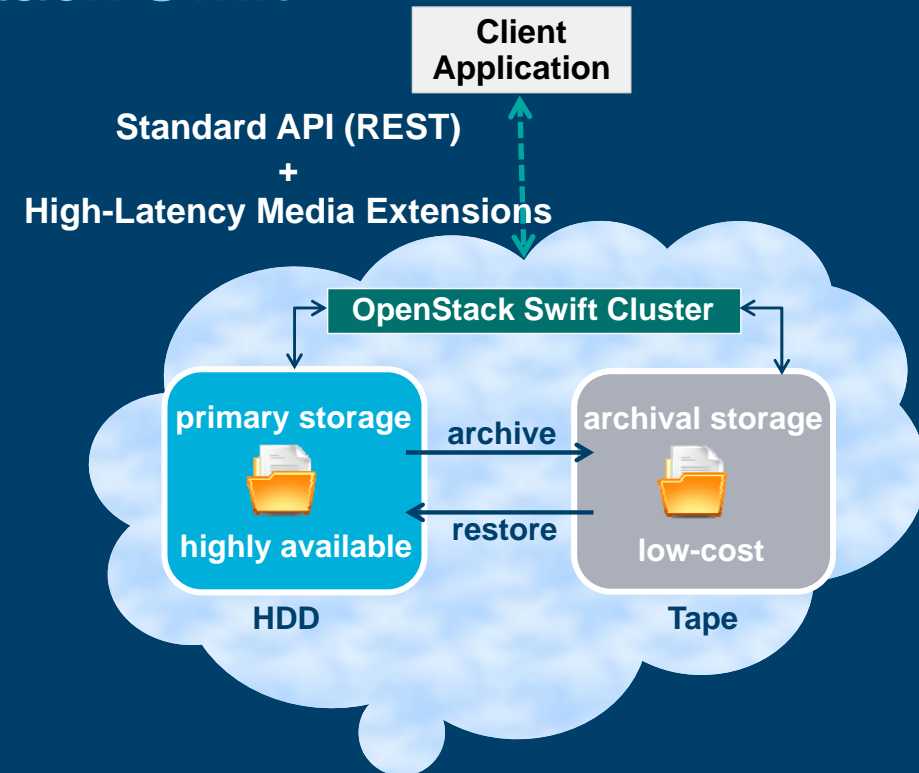
- ❑ Augment cloud object storage with a low-cost, cold storage tier for archive use cases
- ❑ Reduced availability (on the order of minutes)
- ❑ Reduced cost (significantly lower than disk)

Main Idea

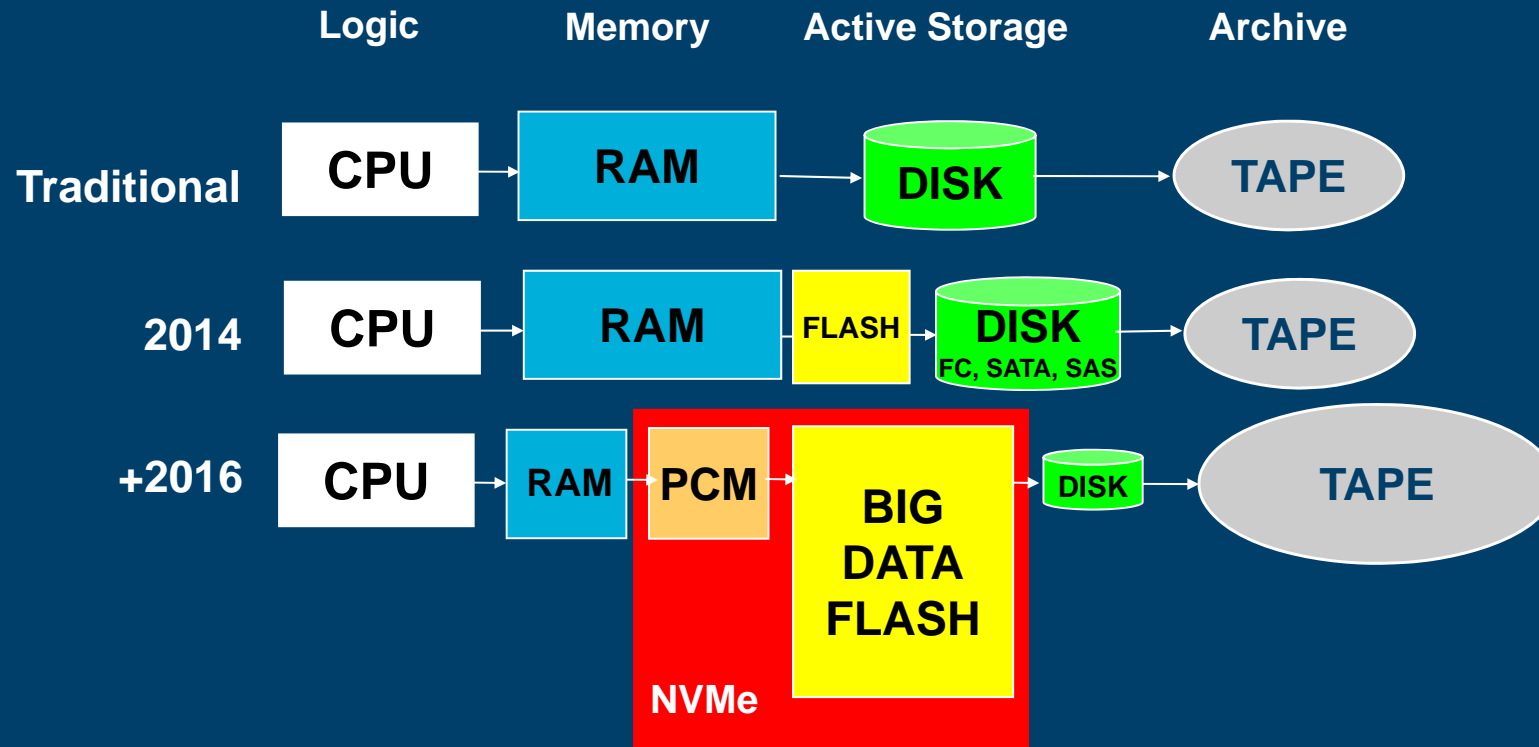
- ❑ Integrate LTFS with a standard disk-based OpenStack Swift installation

Facts about tape

- ❑ Tape is at least 6x cheaper than disk
- ❑ Tape density scaling and cost are projected to be advantageous over disk for the next 10 years



Evolution of the Storage and Memory Hierarchy

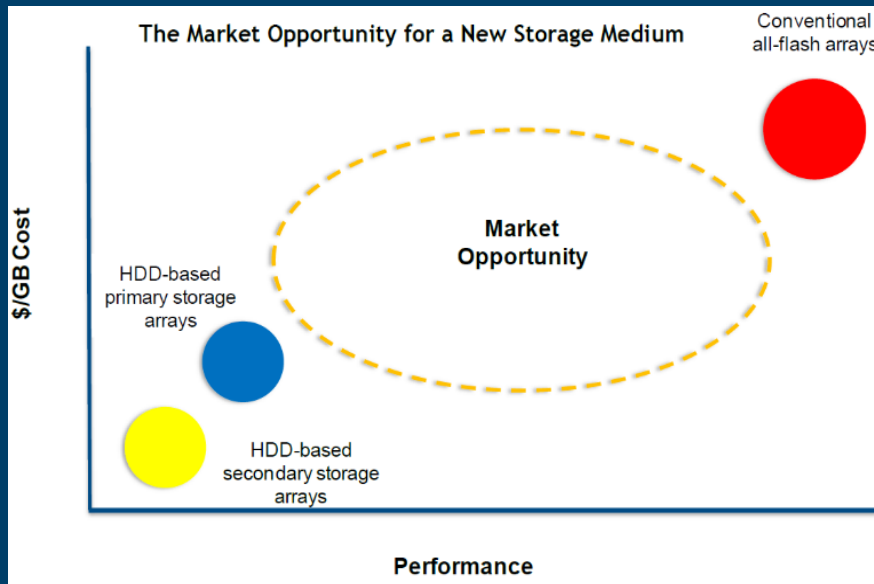


New Category of Big Data Flash

- Many workloads **do not really need the write performance and endurance** of enterprise Flash
 - In certain environments data actually is immutable
- What matters is **high density, low cost, and good read performance**



eBay: "We could live with 1/3rd the number of writes that normal flash supports as long as we could get it for 1/4th the price."

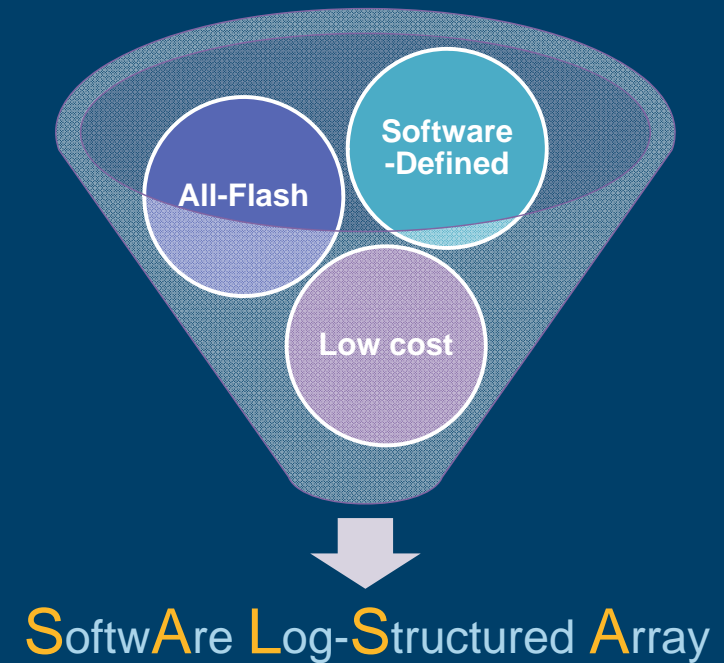


- IDC introduced a new market category of **Big Data Flash** (March 2015)
- Content repositories, media and streaming services, Big Data and analytics, NoSQL, Object storage, Web infrastructure

At < 1 \$/GB for Flash, total acquisition cost becomes the same as an HDD-based solution, with much lower TCO.
- IDC

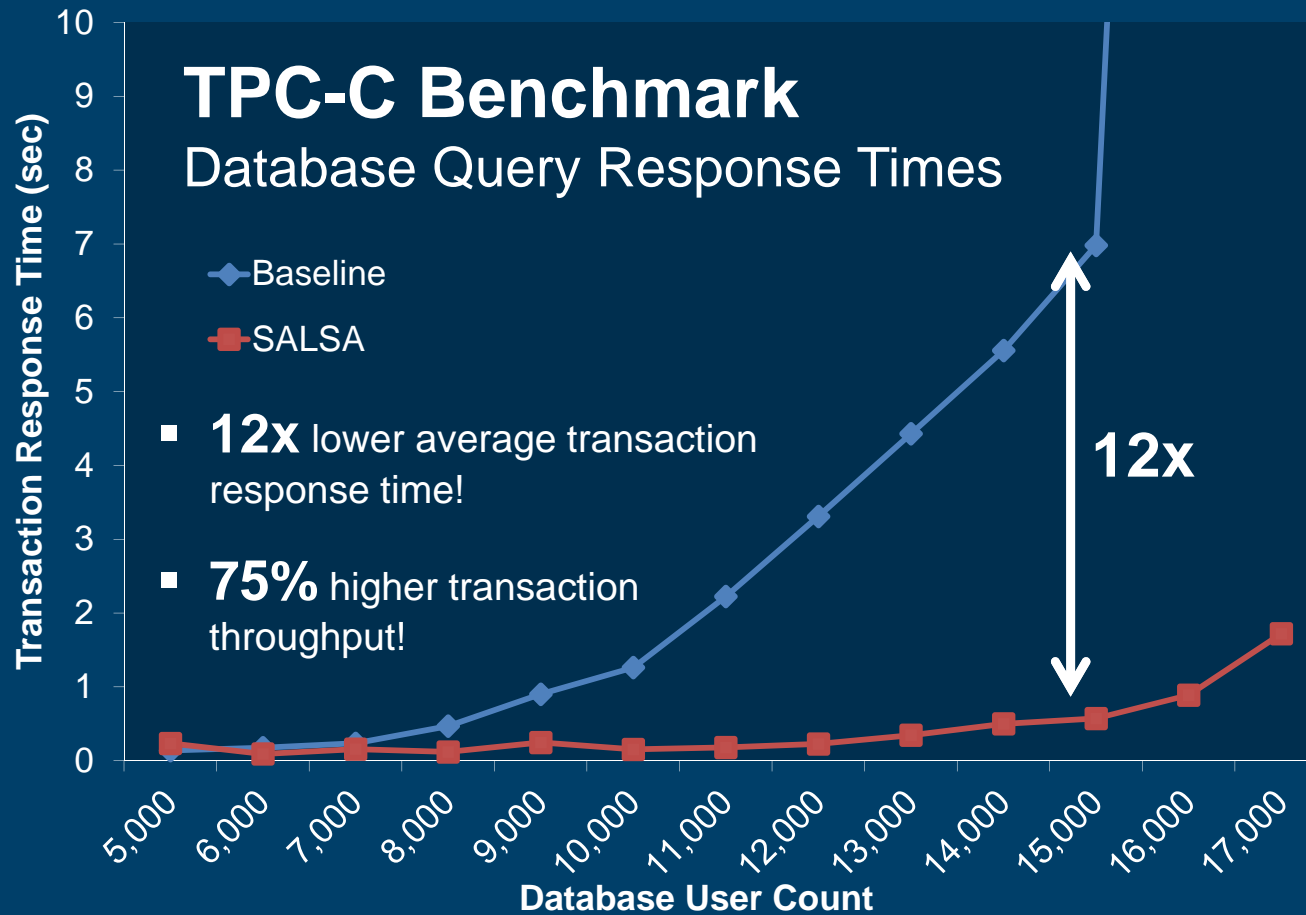
Enabling Use of Low-cost SSDs for Enterprise Workloads

- SALSA (internal code name) is the technology for Spectrum Storage that tunes it for low-cost SSDs with read-dominated workloads
 - SALSA elevates the performance and endurance of SSDs to meet datacenter requirements
- SALSA achieves low Write Amplification by:
 - Transforming access patterns to be as Flash-friendly as possible
 - Implicitly forcing the SSD controller to not do Garbage Collection (GC)
 - Performing GC at a higher level using advanced techniques & more resources – SALSA builds on the IP of FlashSystem controllers



Enables Flash-like performance at HDD-like cost for read-dominated workloads!

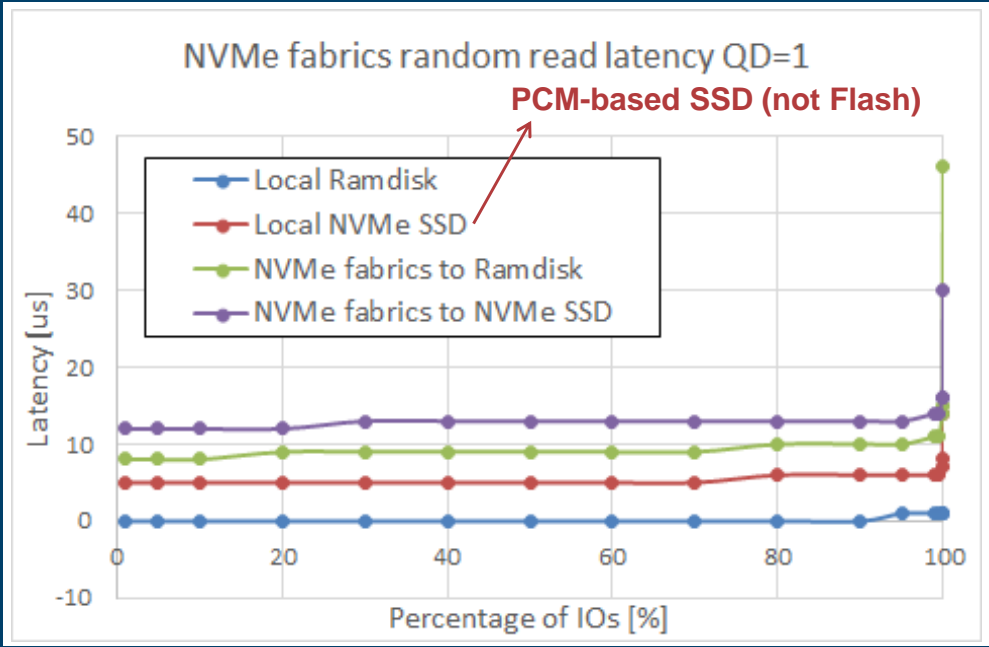
Proof Point of SALSA



NVMe and NVMeF to Unleash Performance of Solid State

- Non-Volatile Memory Express is a standard communications interface and protocol
 - Functionally analogous to SAS & SATA, using PCIe fabric
 - But with significant benefits for high-performance devices
- Low NVMe SSD latency puts pressure on networking: low-latency networking is required!
- NVMe Fabrics is a new driver for NVMe over RDMA

- Lower **latency**: Direct connection to CPU
- Scalable **performance**: 1 GB/s per lane – 4 GB/s, 8 GB/s, ... in one SSD
- Industry **standards**: NVM Express and PCI Express (PCIe) 3.0
- Increased **I/O**: Up to 40 PCIe lanes per CPU socket
- Security** protocols: Trusted Computing Group Opal
- Low **Power** features: Low power link (L1.2), NVMe power states



What's Next?

- Spectrum Scale is at the center of our Software-Defined Strategy for Global File/Object
- Spectrum Scale extends into Cleversafe and hybrid cloud storage solutions using MCStore
- Spectrum Scale and LTFS EE are behind the IceTier “high-latency media” activity in the OpenStack Swift community
- Next steps include enabling Big Data Flash support, and optimizing for the performance of Solid State via NVMe and NVMe over Fabrics
- Stay tuned for HyperScale Converged solutions bringing together Spectrum Scale and Platform Computing components in a highly-integrated package to support containerized workloads and Spark analytics