

# Spectrum Scale Strategy

Wayne Sawdon  
CTO Spectrum Scale Development

Ted Hoover  
Program Director Spectrum Scale Development



# Disclaimer

This information is provided on an "AS IS" basis without warranty of any kind, express or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Some jurisdictions do not allow disclaimers of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

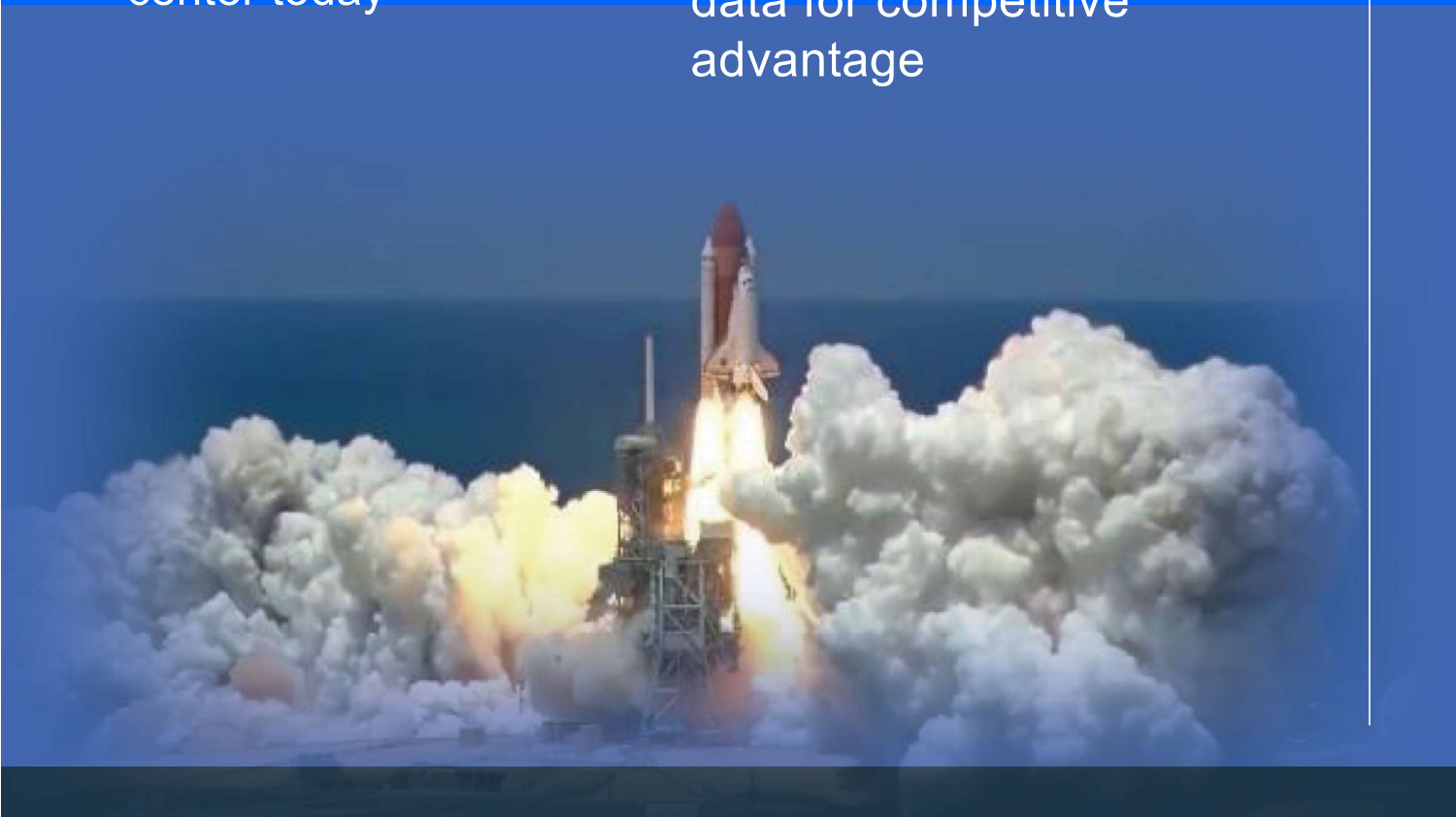
IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

Two simultaneous evolutions are taking shape in the data center today

1. Hybrid multicloud usage
2. Taking advantage of more data for competitive advantage



Companies  
average almost

**5**

private and  
public clouds

**80%**

of companies  
moved their  
applications or  
data from public  
clouds in 2018

### Reasons to migrate from public cloud

- Security
- Performance
- Cost
- Control

*IDC Survey*

# Hybrid multicloud is the platform

**85%**

of companies  
operate in a  
hybrid multicloud  
environment today

**98%**

of companies  
will be hybrid  
multicloud  
in three years

IDC; IBM IBV C-Suite Study; Rightscale

Source: IDC's Cloud and AI Adoption Survey, January 2018



## Thriving in a data driven world

The Economist



The world's most valuable resource is no longer oil, but **DATA**

**20%** of the world's data is searchable and anybody can get to it.

The other **80%** of data is like gold!

"80% of the world's data, whether it is underwriting, pricing, customer experience, risk in loans. That all lives within your data and you don't want to share it, it is **gold**. This is what we do, we help you use that."

Data is unlimited, reproducible, reusable, whose value increases as it is combined and analyzed.

**10X**  
growth in data  
2016 to 2025

Those who can harness the opportunity that lies within their data, gain huge **competitive advantage**.

Two simultaneous evolutions are taking shape in the data center today

1. Hybrid multicloud usage
2. Taking advantage of more data for competitive advantage



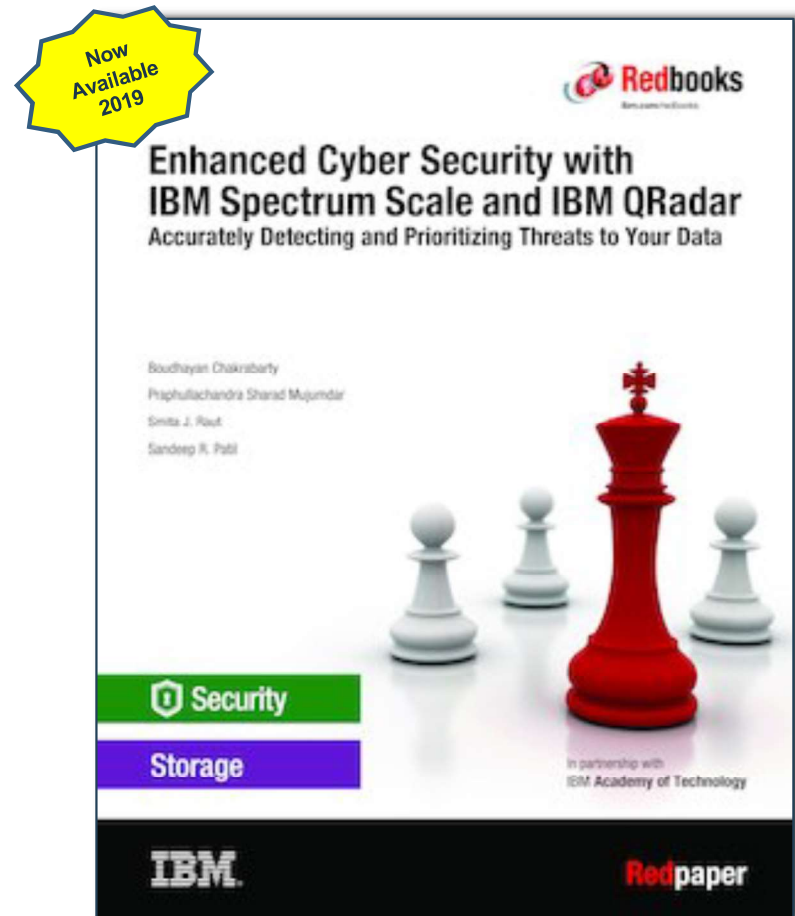
**IBM's ACQUISITION OF RED HAT IN JULY COMPLETELY CHANGED THE CLOUD LANDSCAPE TO BECOME THE WORLD'S #1 HYBRID MULTI-CLOUD PROVIDER.**



## Enhanced Cyber Security with IBM Spectrum Scale and IBM QRadar

- IBM QRadar is a leading Security and Event Information Management (SEIM) system. It analyzes Spectrum Scale's event data in **real time** for early detection of targeted attacks and data breaches.
  - ✓ Proactively safeguard all data residing in Spectrum Scale
  - ✓ Real time alerts of potential attacks and data breaches.
  - ✓ Drastically enhance the cyber threat detection
  - ✓ Immediately stop intruder attacks

**If Data is the 'Crown Jewels' then Storage is the 'Jewel Safe'. Invest to keep your data safe.**



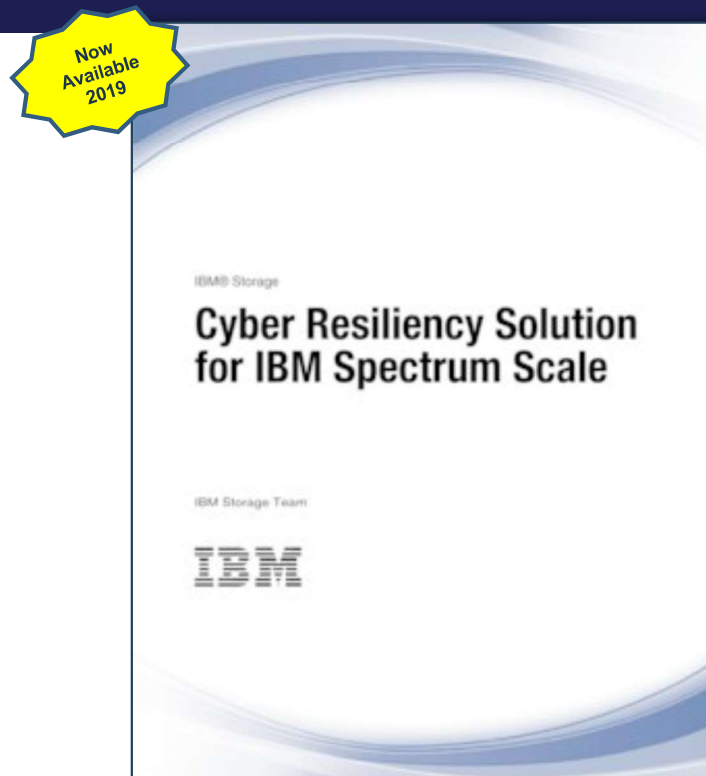
<http://www.redbooks.ibm.com/abstracts/redp5560.html?Open>

# Cyber Resiliency Solution for IBM Spectrum Scale



## Benefits

- Cyberattacks can have a catastrophic impact on an organization, and can result in devastating financial losses.
- Cyber Resiliency refers to an entity's ability to continuously deliver the intended outcome despite cyber events.
- The paper facilitates deployment of the Cyber Resiliency solution for IBM® Spectrum Scale.
- This solution is designed to protect the data on IBM Spectrum™ Scale from external cyberattacks or insider attacks using its integration with IBM Spectrum Protect™ and IBM Tape Storage.



<https://www.redbooks.ibm.com/redpapers/pdfs/redp5559.pdf>



# Next Generation Performance

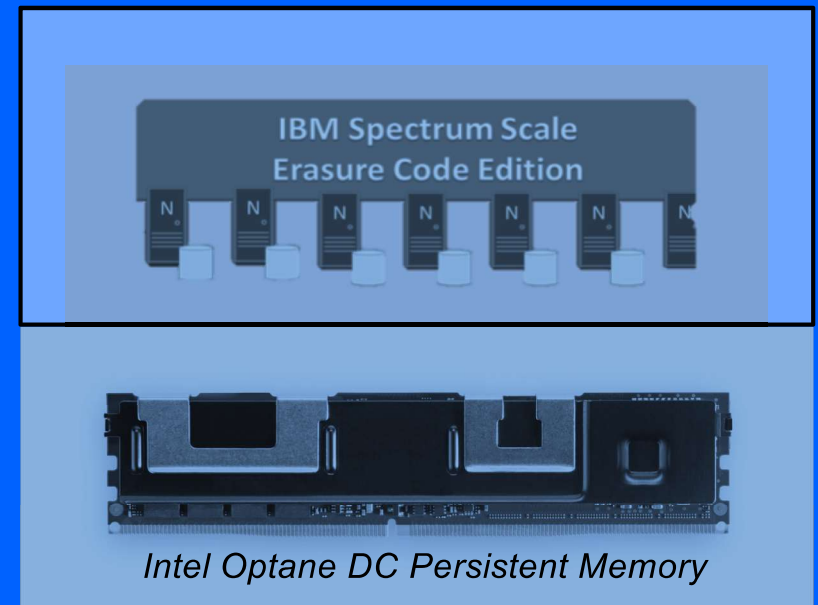
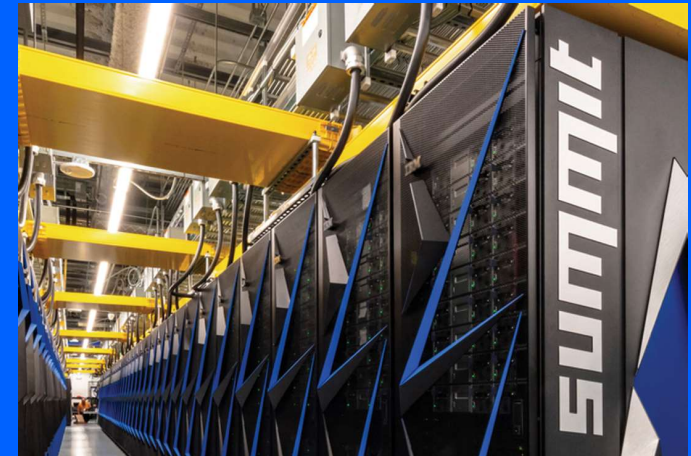
Scale's **Erasure Code Edition (ECE)** was announced in May 2019. Deploy the same ESS erasure encoding across storage rich servers for low cost reliable storage from commodity hardware.

Leverage ECE with **Persistent Memory** to create Extreme Performance to local store.

Project "**Ustore**" to exploit persistent memory:

- ✓ Cooperative consistent client cache
- ✓ Topology Aware Replication with data affinity
- ✓ Read Ahead / Write Behind job scheduling
- ✓ Eventual Durability

Invest to accelerate time to value



# Store everywhere. Run anywhere. Enable Global Collaboration

## Challenge

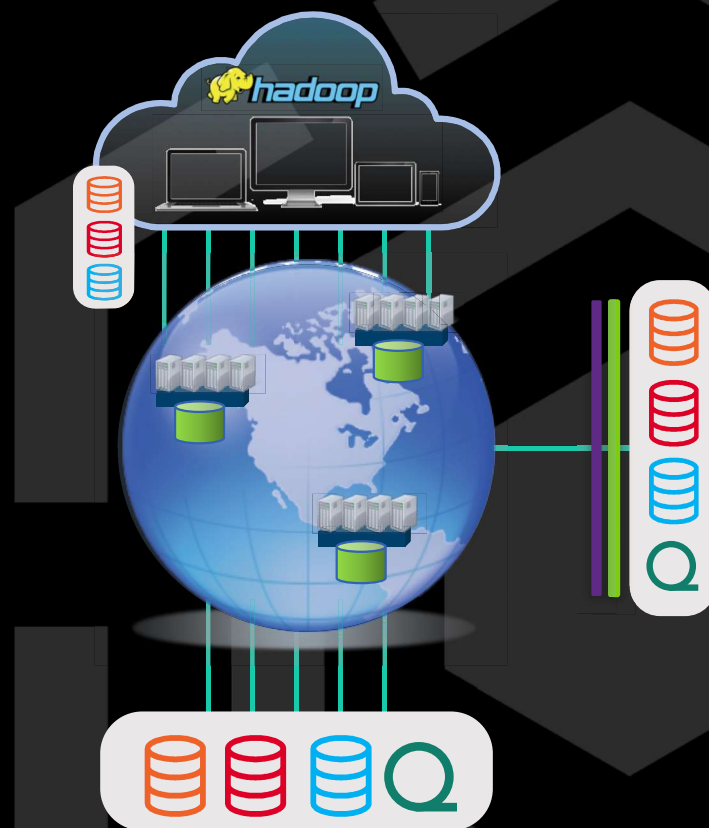
- Multiple sites working on same data
  - High performance access to local cache
  - Consistent metadata and data locking
  - Support for mission-critical transactional replication
  - Manage unreliable, remote sites

## Advanced File Management

- Global namespace with fast, consistent metadata
  - Unifies on-prem deployments and cloud deployments and multi-cloud
  - Active-Active with cache consistency
  - Prefetch data or copy on demand
  - High performance access to local

## Key technology for hybrid multi-cloud deployments

- Invest in connectivity to object store
- Invest in network reliability and performance



# Goal: Deliver High Performance File Services to Containerized Application Workloads

## Support Workloads that Require High Performance File Services

- Analytics & Cognitive
- High Performance Computing
- AI Data Pipeline

## Support the Workload Ecosystem in the Cloud

- Containerized Applications, Storage
- Ephemeral and Persistent Storage Volumes

## Flexible Deployment

- Dynamic Provisioning, Configuration, Upgrade

## Support for Multiple Clouds

- Public, Private, Hybrid

## Support Hybrid Use Cases

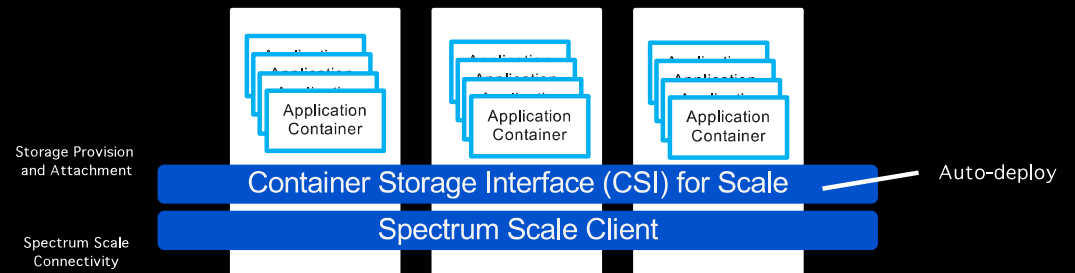
- Cloud Burst – Single Name Space
- Multi Cloud Data Sharing
- Archive
- Data Accelerator (High Performance Tiering)

## Solution Integration (Partners)

# Spectrum Scale Containers Models

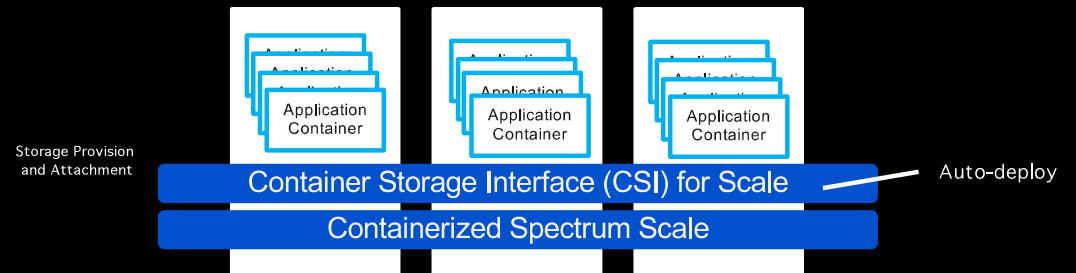
## Storage for Containers

- Container Ready Storage

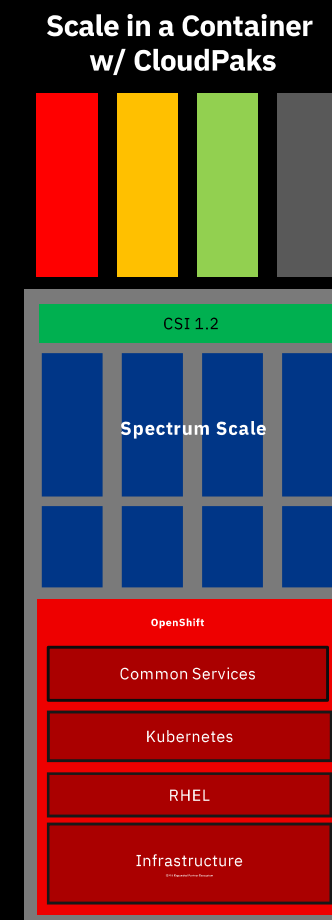
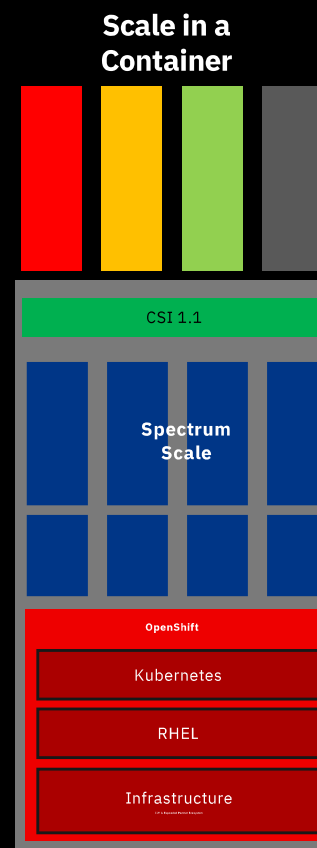
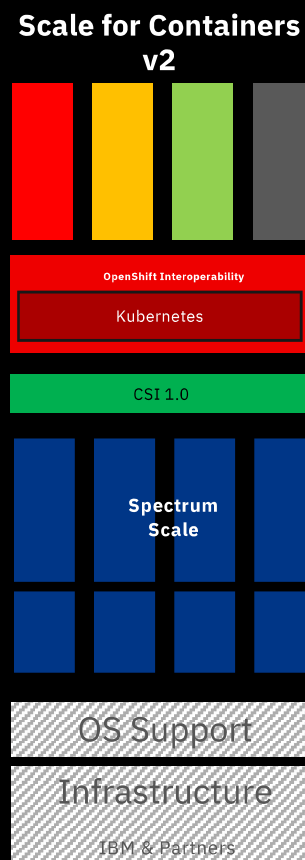
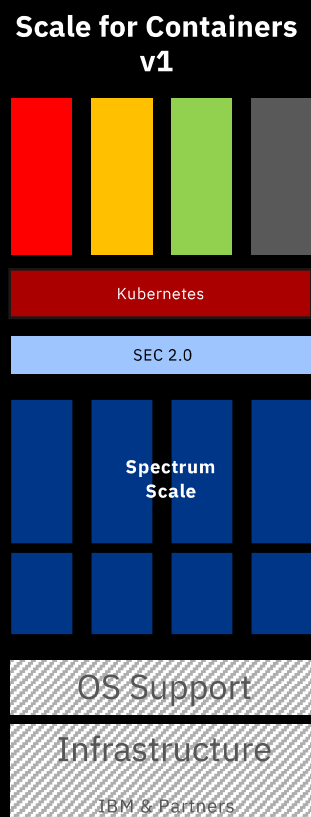
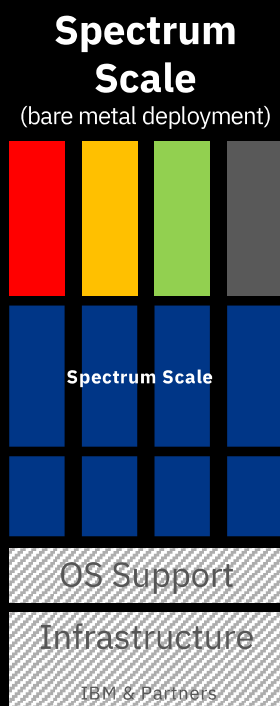


## Storage in Containers

- Containerized Storage

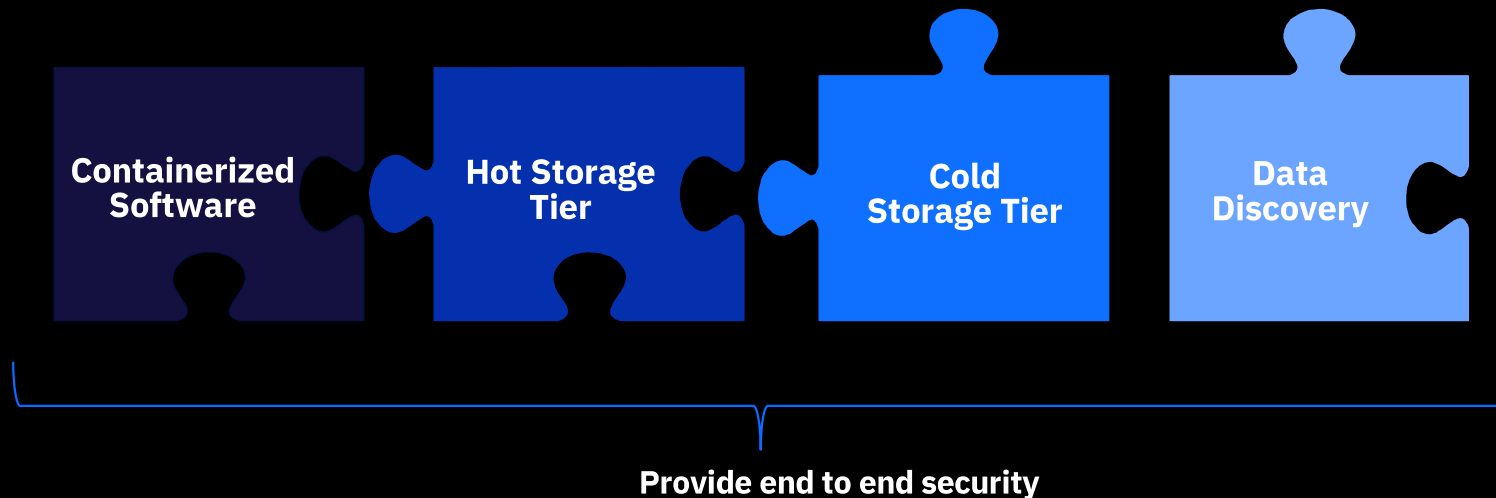


# Evolution of IBM Spectrum Scale Containers



# Data Acceleration for AI and Analytics

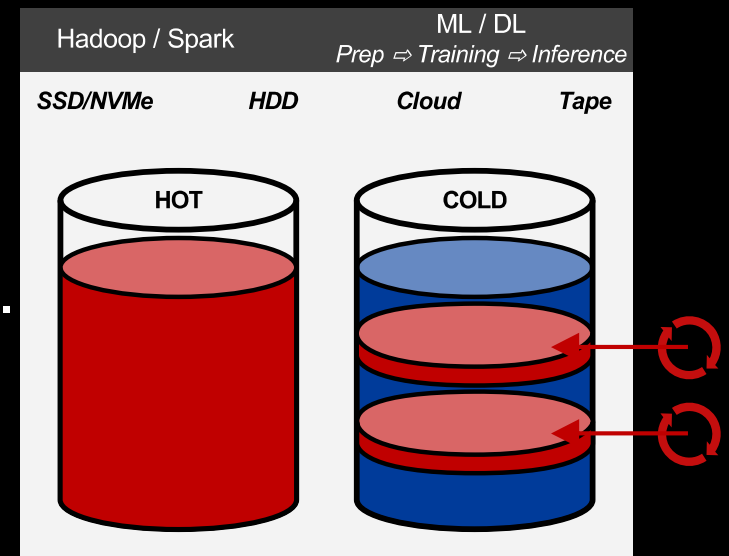
“Daniel, a data scientist, can find and prefetch the data he needs for his ML/DL jobs without any knowledge about Spectrum Scale”



# The Problem

We see:

- Customers across all verticals are creating large PB to EB data stores.
- Vast majority of data is relatively cold, but still required for periodic trend analysis.
- But AI / Analytics require high performance, low latency storage to keep expensive CPU / GPU / TPU / FPGA busy.

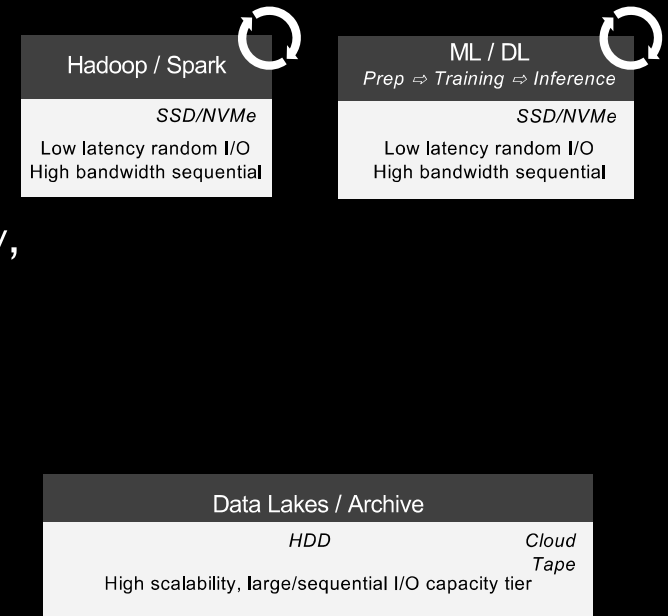




# The Solution **Data Accelerator** for AI and Analytics

## Two-Tiers of Storage:

- **Performance Tier**
  - **Maximize performance of storage: \$/IOP & \$/GB/s are key**
    - Low latency random I/O & High bandwidth sequential
    - Relatively small compared to Capacity tier (say 5-25%)
    - Can be Lower Durability, Lower Availability, Lower Reliability, if Architected properly
    - No Geo-distribution
- **Capacity Tier (aka “Data Lake”)**
  - **Minimize the cost of storage: \$/TB is key**
    - High Durability, Availability, Reliability, Geo-distribution



# IBM Design Thinking

---

READ, PRACTICE, CREATE, ASK

**Tuesday, November 19, 2019 Centennial Ballroom G**

**Abstract:** A collaborative co-creative session to explore how Spectrum Scale can deliver differentiating user experiences to address future needs and technologies. We desire to get first hand information from our customers to get input for our roadmap. Both sessions will have a different focus

**Schedule:**

10:00AM - 11:00AM Cloud and Containers

11:00AM - 12:00PM Data Accelerator for AI



