IBM Spectrum Fusion HCI

OpenShift that is simple to own and manage

Spectrum Scale German User Meeting 2022
Cologne, Germany – October 19-20, 2022

mdietz@de.ibm.com (IBM Spectrum Scale RAS Architect)

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.

What clients are asking for

Self-managed OpenShift
for Cloud Paks and containers
on-premises

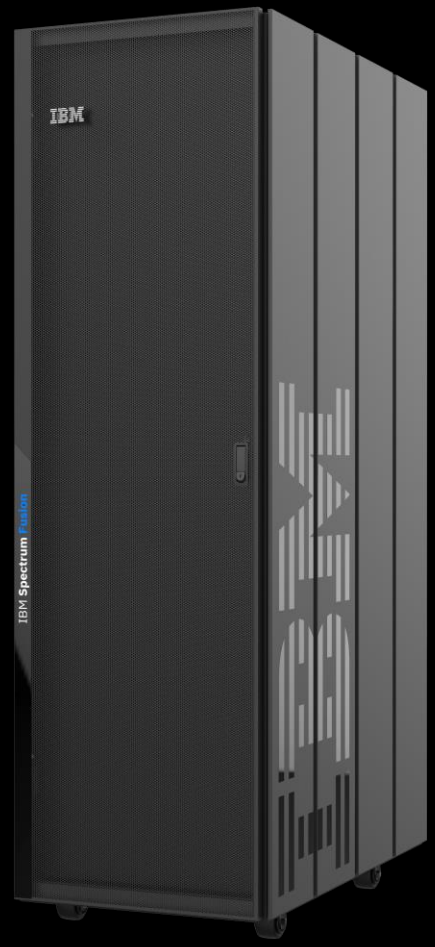
But there are challenges:

- Setting up OpenShift clusters require significant IT resources
- Lifecycle management of OpenShift clusters requires new skills
- Getting data into the OpenShift cluster, efficiently and securely
- Transitioning projects into production

IBM **Spectrum Fusion HCI**

Bare-metal OpenShift cluster-in-a-box

Enables clients to get Cloud Pak and container workloads **out of pilot and into production *faster***



IBM Spectrum Fusion **HCI**

Pre-engineered self-managed Kubernetes that is easy to own

Integrated system

Combines :

- x86 servers, NVMe storage, high-speed switches
- Red Hat OpenShift
- Spectrum Fusion data services

Pre-engineered, factory integrated OpenShift application platform

Hardware Infrastructure + RED HAT® OPENSHIFT + IBM Spectrum Fusion

Integrated system
Architected for modern OpenShift workloads

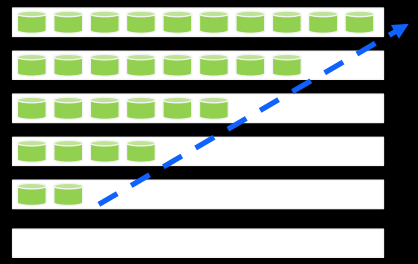
Enterprise Data Services for OpenShift
To harden and protect mission-critical applications

Ideal self-managed OpenShift platform
for Cloud Paks and certified operators on-premises

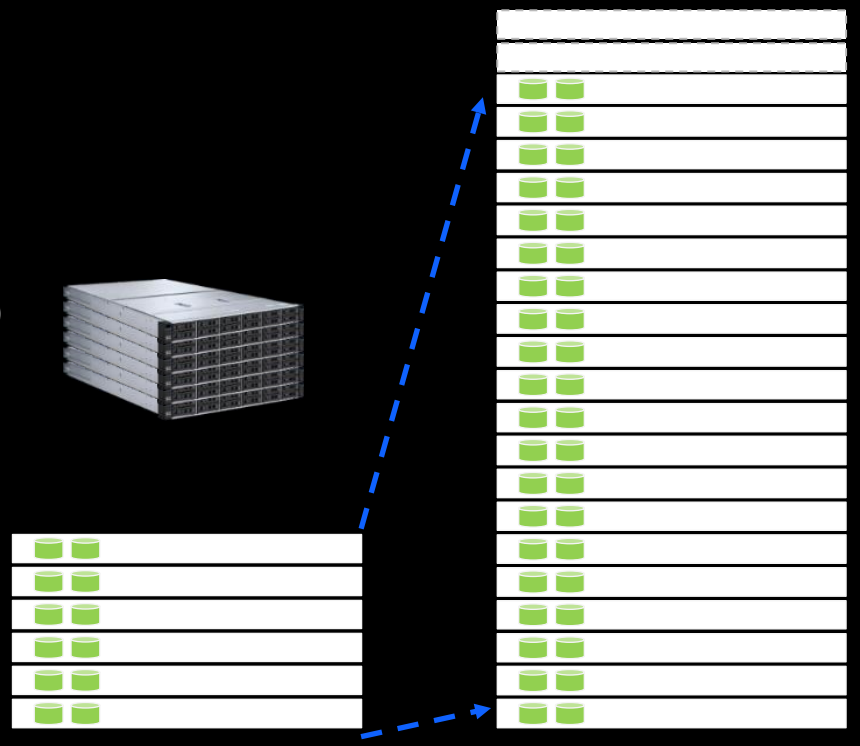
IBM Spectrum Fusion HCI

6-20 hyperconverged 1U compute/storage nodes for independent scaling of capacity and compute performance

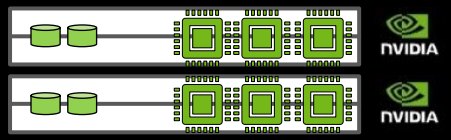
Current minimal configuration:
42U rack
2x Ethernet ToR switches (100GbE)
2x Ethernet management switches
6x Storage-rich x86 compute node



Storage scalability
0-10 × 7.68TB NVMe SSDs (in pairs) per node, equally distributed except for zero.



Compute scalability
6-20 × x86 nodes
(increment of 2, min 2 SSDs each)



Pairs of 2U GPU enhanced nodes with AMD CPUs and 3x NVIDIA A100 GPUs

Optional GPU-accelerated 2U nodes

Optional 1-2 AFM Gateway Nodes

Why Spectrum Fusion HCI?

AGILITY and COST

- Simple to deploy, scale, and operate
- Flexible configuration options
- Reduce OpenShift license costs by 5x
- Simplify data management

ENTERPRISE GRADE

- Outage resistant hardware architecture
- Integrated Spectrum Fusion data services
- Expert Care with SPOC(2) support (24x7x365)

PERFORMANCE

- Dedicated high-speed storage network
- High performance RWX/RWO file system
- Optional GPU acceleration

1. Requires data center network configuration, which is not included in this timing
2. Single Point of Contact

Coming Soon

Integrated OpenShift data services platform

IBM Spectrum Fusion HCI

OpenShift data services platform software

Red Hat OpenShift

vmware

IBM Spectrum Fusion

OpenShift data services platform software on Public Cloud

Red Hat OpenShift

Google Cloud

Microsoft Azure

aws

IBM Cloud

IBM Spectrum Fusion

Spectrum Fusion HCI: **Architecture**

Spectrum Fusion HCI

Software Architecture

IBM Cloud Paks
Data, Security, Integration,
Network Automation,
Business Automation,
Watson AI Ops

Client/Partner
Any OpenShift application
NVIDIA CUDA
Edge applications
…

Open Source
AI/ML: TensorFlow, PyTorch,
NumPy, run.ai, etc.
PostgreSQL, MongoDB
…

GUI and REST API

Spectrum Fusion Data Services

DR Services

HA Services

Data Access Services

Configuration Management

Storage

Compute

Networking

Monitoring and Health

Health

Capacity

Performance

Serviceability and Upgrade

Logs / Audit

Events

Call Home

Upgrade

IBM Cloud Satellite

OpenShift Virtualization

OpenShift Container Platform (Bare Metal)
Build and run cloud-native applications

Red Hat OpenShift

100% automation: Setup & configuration

Highly available infrastructure
Compute, Storage and Network

Spectrum Fusion

Bare Metal is Better

High Performance (1)

- Eliminate hypervisor to reduce overhead
- More resources available for workloads

Reduce Cost, Simplify Operations

- Reduce OpenShift license cost (2)
- Eliminate VM operational complexity

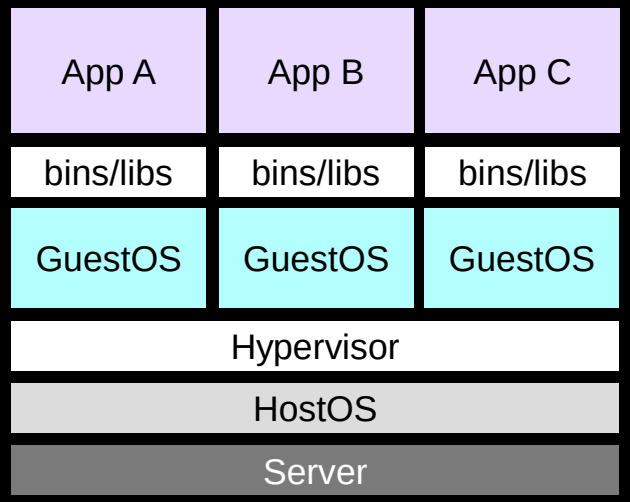
Improve Security

- Immutable CoreOS reduces attack surface
- Sandboxed containers (3)

Support Windows and Linux VMs

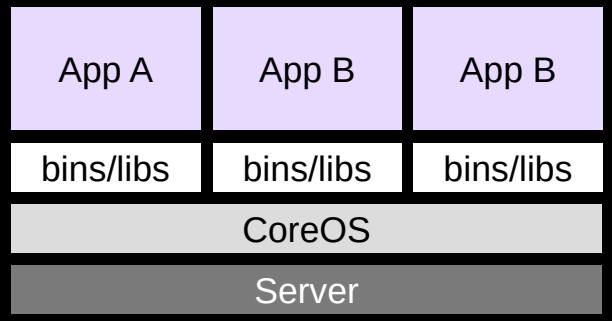
- Manage with OpenShift Virtualization (4)

Good



Virtualized Infrastructure with many layers

Better



Fusion on bare metal OpenShift

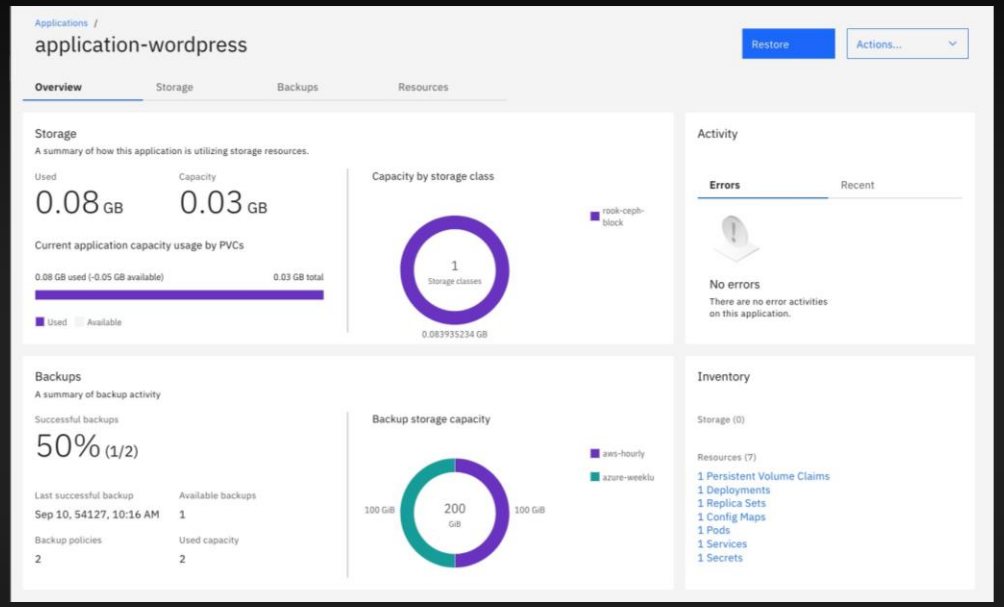
(1) VMs can add up to a 25% performance penalty
(2) OpenShift core-pair licensing can cost 5x or more of bare metal licensing for some configurations
(3) https://docs.openshift.com/container-platform/4.8/sandboxed_containers/understanding-sandboxed-containers.html
(4) https://cloud.redhat.com/learn/topics/virtualization/

Spectrum Fusion **HCI**

Two ways to manage

Management UI

- Guided experience
- Simplified workflows



“As code” via custom resources

- Kubernetes native APIs
- Check configuration into git for easy deployment into multiple clusters (GitOps)

```yaml
apiVersion: data-protection.isf.ibm.com/v1alpha1
kind: PolicyAssignment
metadata:
  name: wordpress-aws-daily
  namespace: ibm-spectrum-fusion-ns
spec:
  application: wordpress
  backupPolicy: aws-daily
  runNow: true
```

Spectrum Fusion **HCI**

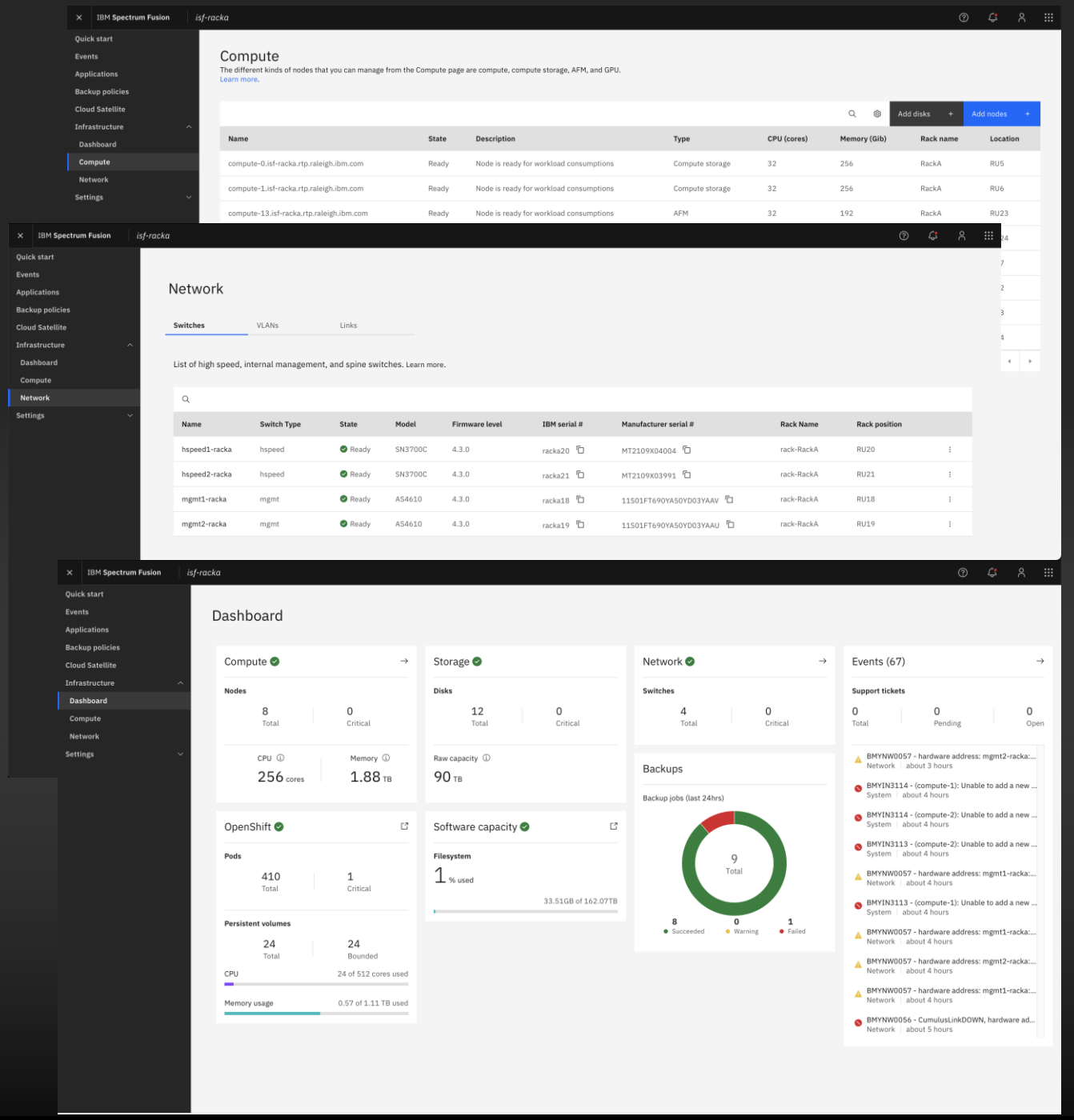
Integrated hardware management

Red Hat OpenShift Operators are used for managing the hardware of Spectrum Fusion HCI

The inventory of all compute nodes and Ethernet switches is presented using an intuitive GUI

The status of the hardware is shown along with menus of available actions

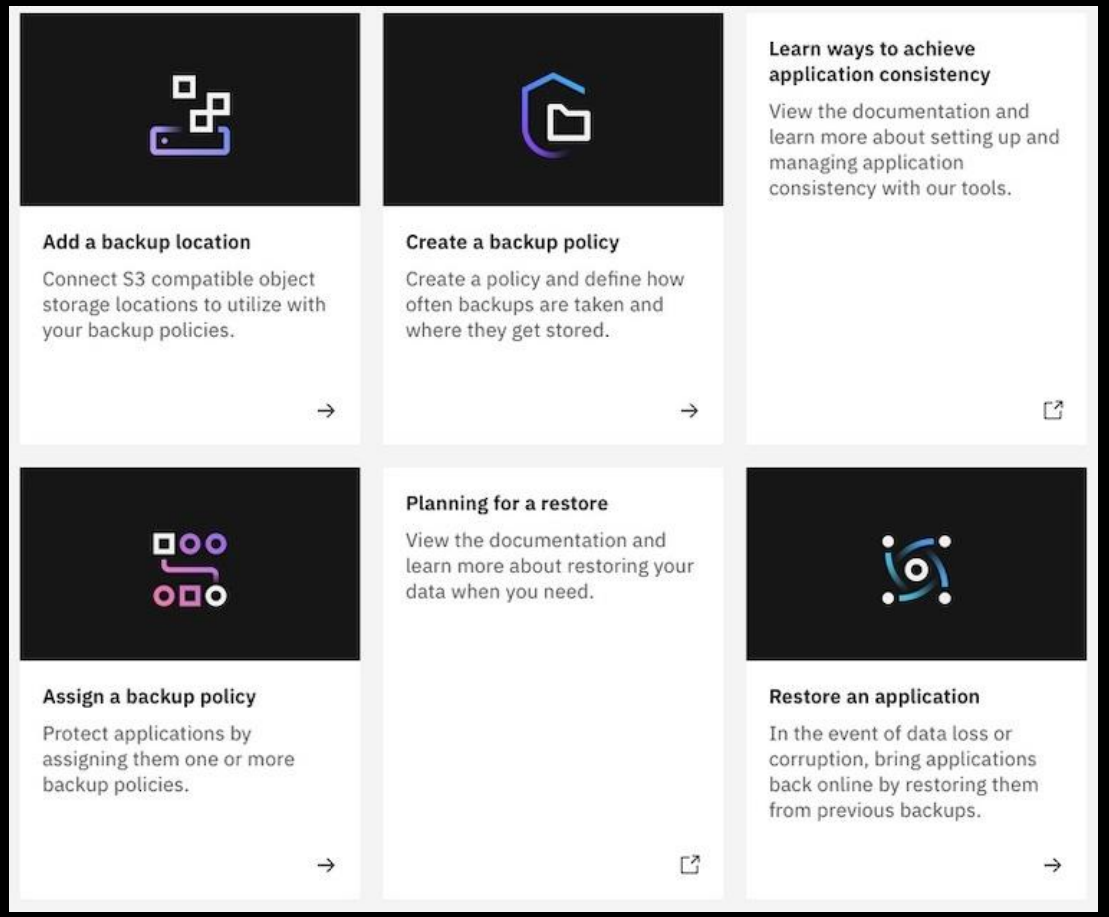
Regular updates of the firmware of all components are provided



Spectrum Fusion

Protect Application Data

Backup and restore



Basic B/R features

1. **Connect to object storage locations** to safely store backups
2. **Define policies**
   - How frequently to backup
   - Where to store the backups
   - How long to keep the backups
3. **Assign policies to applications**
4. **Restore applications from backups**

Spectrum Fusion

Protect Application Data

Backup and restore

```
groups:
- name: <group-name – must be different from group-names in Application CR>
- parent: <group-name from Application CR> (Application CR itself represent implicit default parent group)
- <Include/exclude/select clauses – subset of those in Application CR>
resources:
- <resourcesetname>
-   name: <resourcename> (optional)
-   yaml: <inline yaml>
hooks:
- name: <hookname>
- type: <exec|scale|check>
- <hook-specific parameters>
    - e.g. namespace: <$<group-name>|<namespace>>
    - e.g. nameSelector: <workload/pod-name with wildcards>
    - e.g. labelSelector: <label selector for workload or Pod resource>
- Ops/chks:
    - <hook operations or checks respectively>
workflow:
- resource: <resourcesetname>
- hook: <hookname>/<opname>(<args>)
- group: <ref to group defined herein or in Application CR>
- hook: <hookname>/<opname>(<args>)
- group: <ref to group defined herein or in Application CR>
```

Advanced B/R features

1. **Backup workflows, aka recipes**
   Define complex backup workflows in YAML.
   - What to backup
   - How to backup
2. **Consistency groups**
   Instant crash consistent snapshot of multiple PVs without application pause
3. **Application consistent backups**

Spectrum Fusion **HCI**

Integrated lifecycle services

Feature	Why it Matters
Rolling firmware upgrades for servers and switches	**Avoid application outages** Updating firmware to fix critical security issues requires that the affected hardware be taken offline. Fusion HCI enables workload / traffic to be rerouted around the affected device so the device can be taken offline and updated without causing an application outage. You're given flexibility on how many nodes to upgrade so that you control the length of your change control window and Fusion automates the upgrades in a non-disruptive manner.
Firmware tested by IBM	**Avoid application outages** Firmware incompatibilities between system components can result in outages. IBM tests firmware compatibility in Fusion HCI components before releasing updates.
Scale-out storage	**Add storage without application outage** Add storage by either adding drives to existing servers or by adding more servers with drives.
Scale-out servers	**Add cores / vCPU without application outage** Plug-in additional servers, bring them online through the Fusion UI by adding them to your OpenShift cluster on your own schedule.
Integrated call home	**Automate problem ticketing for critical system errors to accelerate resolution**
Event management	**Monitor and manage events through OpenShift**

Spectrum Fusion **HCI**

Value of storage architecture

Scale-out parallel file system

- Enormous scalability with software-based declustered RAID protection
- Very high performance no additional RAID hardware
- Performance scales as you add nodes

Durable and robust storage

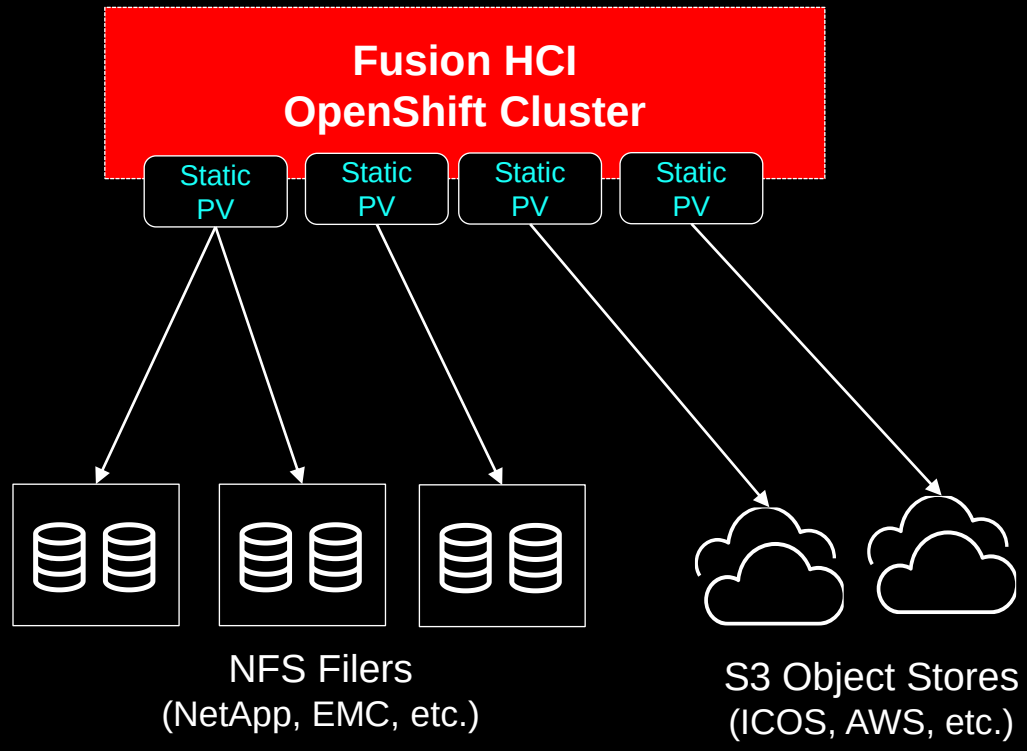
- ❖ Distributes data across nodes and drives for higher durability *without* the cost of replication
- End to end checksum identifies and corrects errors introduced by network or media
- Withstands multiple failures; Rapid recovery and rebuild

High storage efficiency

- 66% efficient with default 4+2p erasure coding
- 75% efficient with 8+3p optional erasure coding
- Lower cost than 33% efficient 3-way replication

Spectrum Fusion **HCI**

Access data through Spectrum Scale AFM



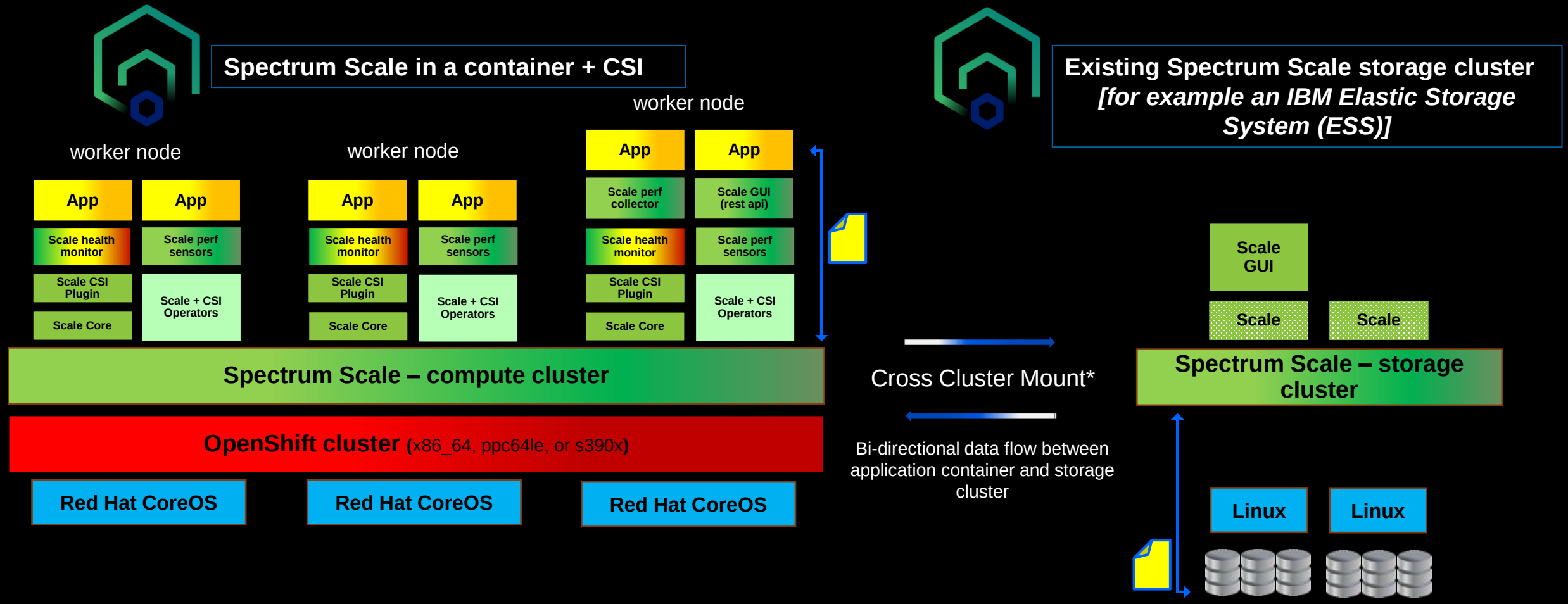
Storage administrators can **eliminate data copies** and **reduce data administration overhead**

- Bring NAS Filers and S3 Storage under a single directory and make it transparent to the application
- Intelligent data mirroring enables accessing remote data at local file system speeds
- Transparent and online migration of data from remote storages to Fusion HCI file system without downtime

Spectrum Fusion HCI: **Spectrum Scale Integration Details**

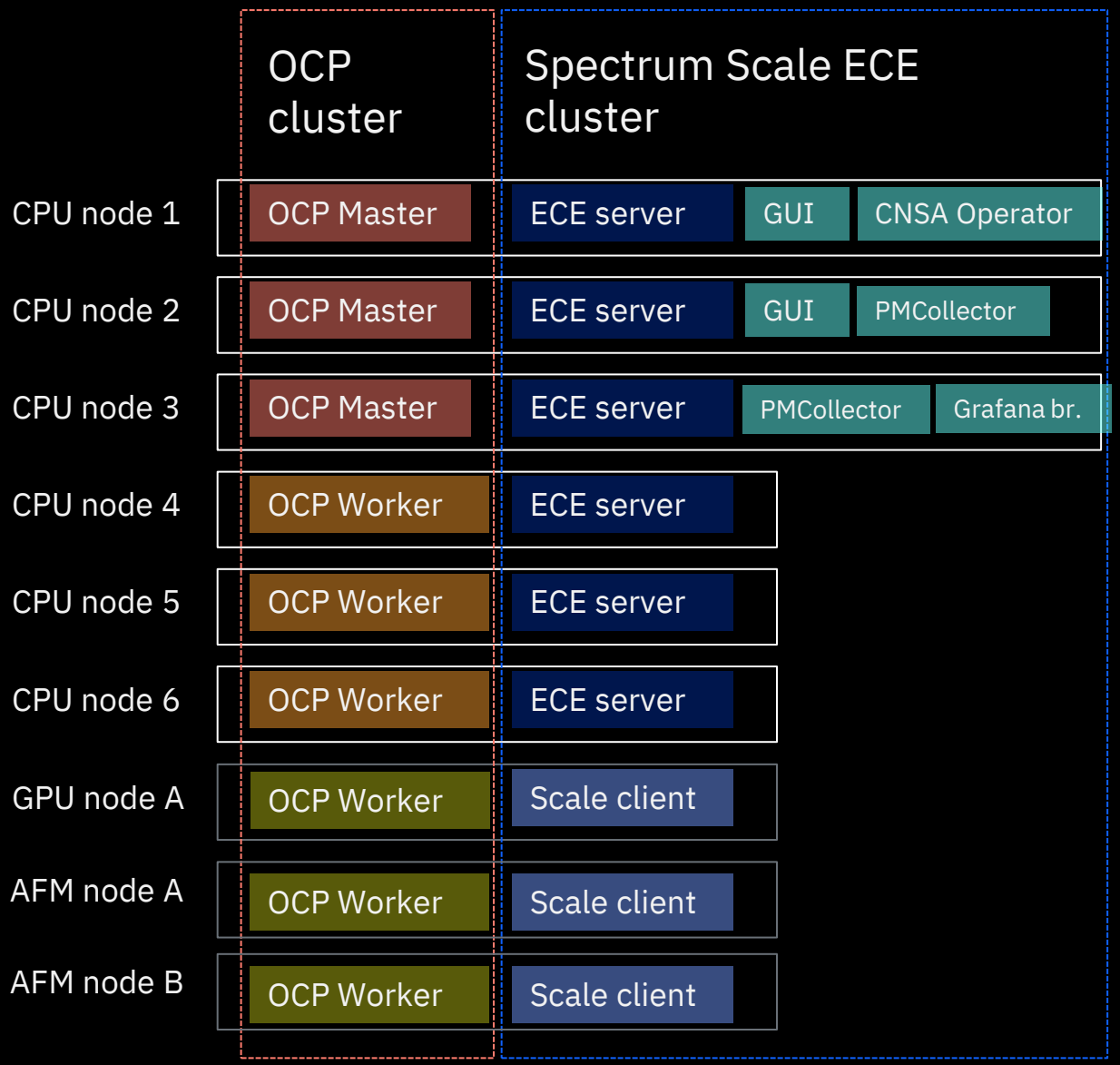
Spectrum Fusion builds upon:
IBM Spectrum Scale Container Native Storage Access (CNSA)

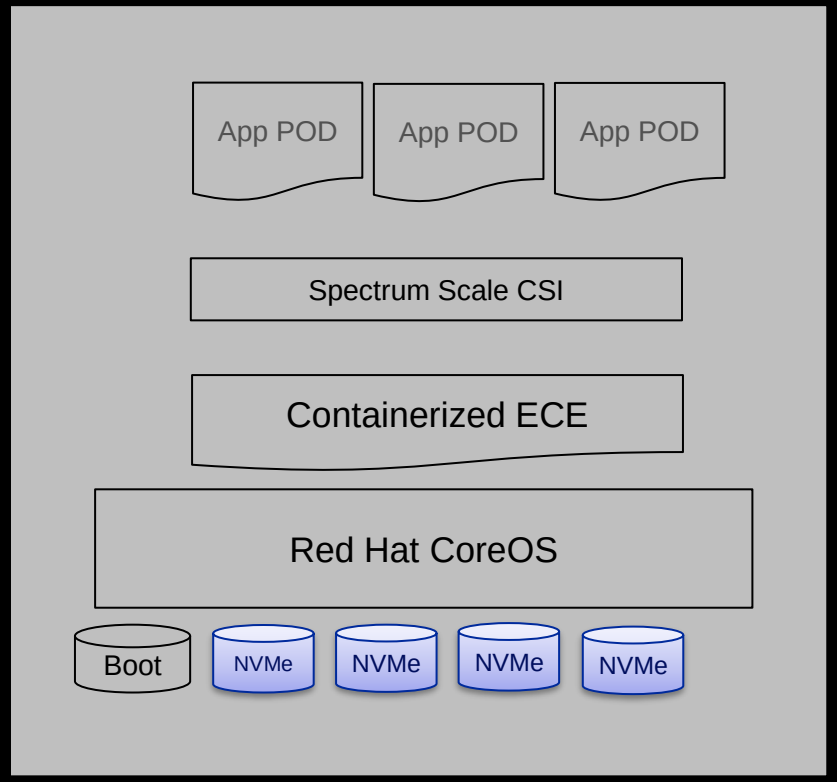
Cluster Overview



*Spectrum Fusion HCI leverages ECE with builtin storage instead of cross cluster mount

Fusion HCI – Spectrum Scale ECE Deployment Model



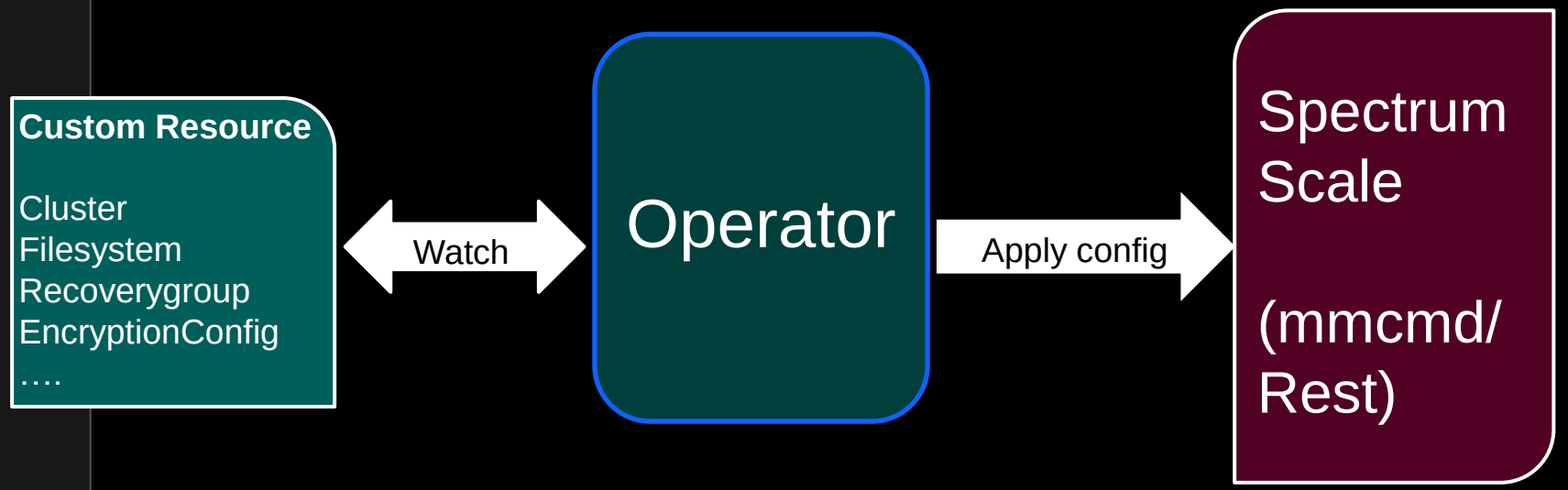


Spectrum Scale ECE SW:

- Best of the breed high performance cluster file system
- Declustered erasure coding
- Scale out solution
- Containerized for easy deployment
- AI optimization
- Global data access

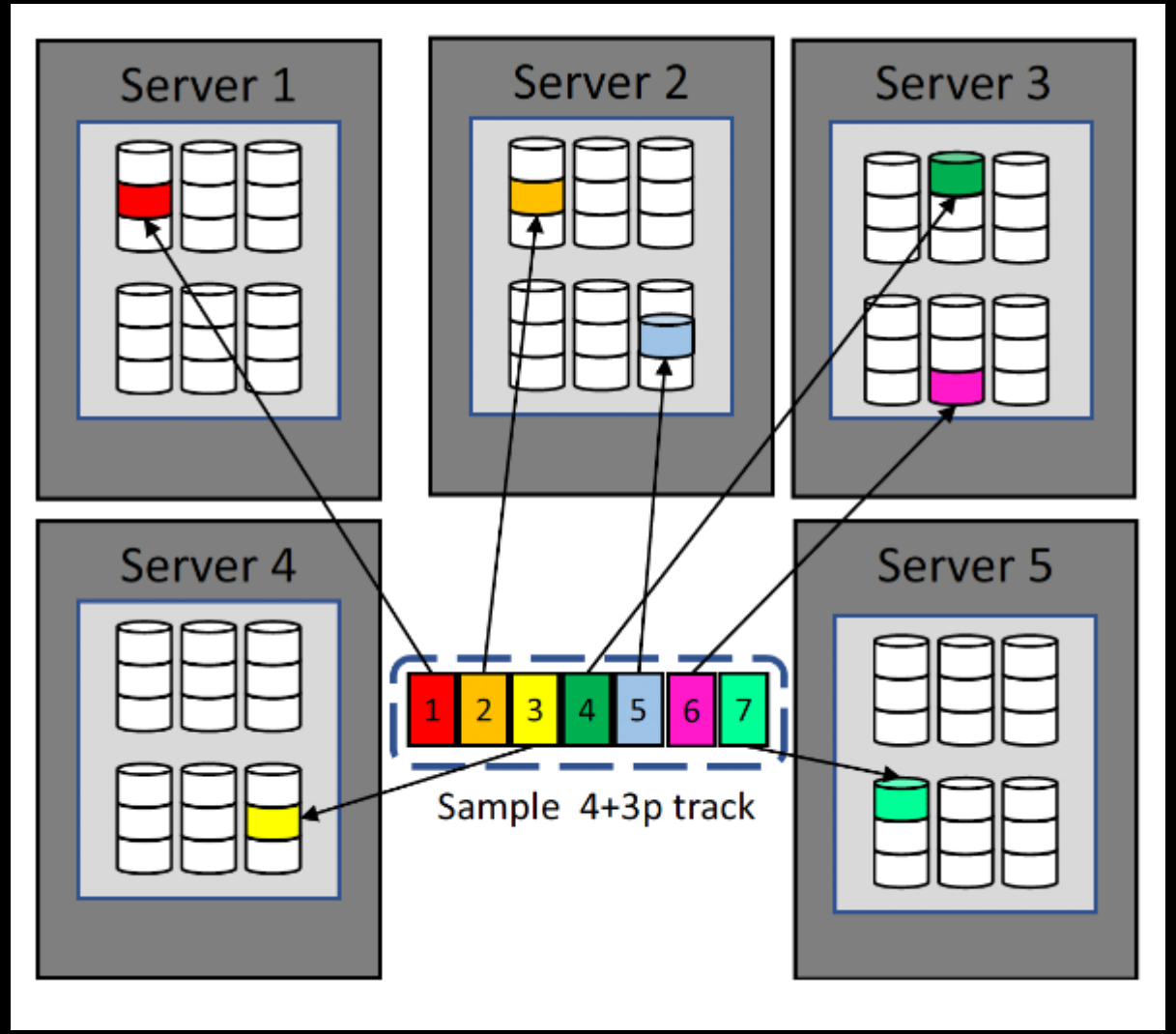
IaC - Configuration through Custom Resources (CRs)

```
apiVersion:
scale.spectrum.ibm.com/v1beta1
kind: Filesystem
metadata:
  labels:
    app.kubernetes.io/instance: ibm-
spectrum-scale
    app.kubernetes.io/name: cluster
  name: scale_fs1
  namespace: ibm-spectrum-scale
spec:
  vdiskNSD:
    vdiskSets:
    - blockSize: 1M
      declusteredArray: DA1
      nsdUsage: dataAndMetadata
      raidCode: 4+2P
      recoveryGroups:
      - rg1
      setSize: 100%
      storagePool: system
```



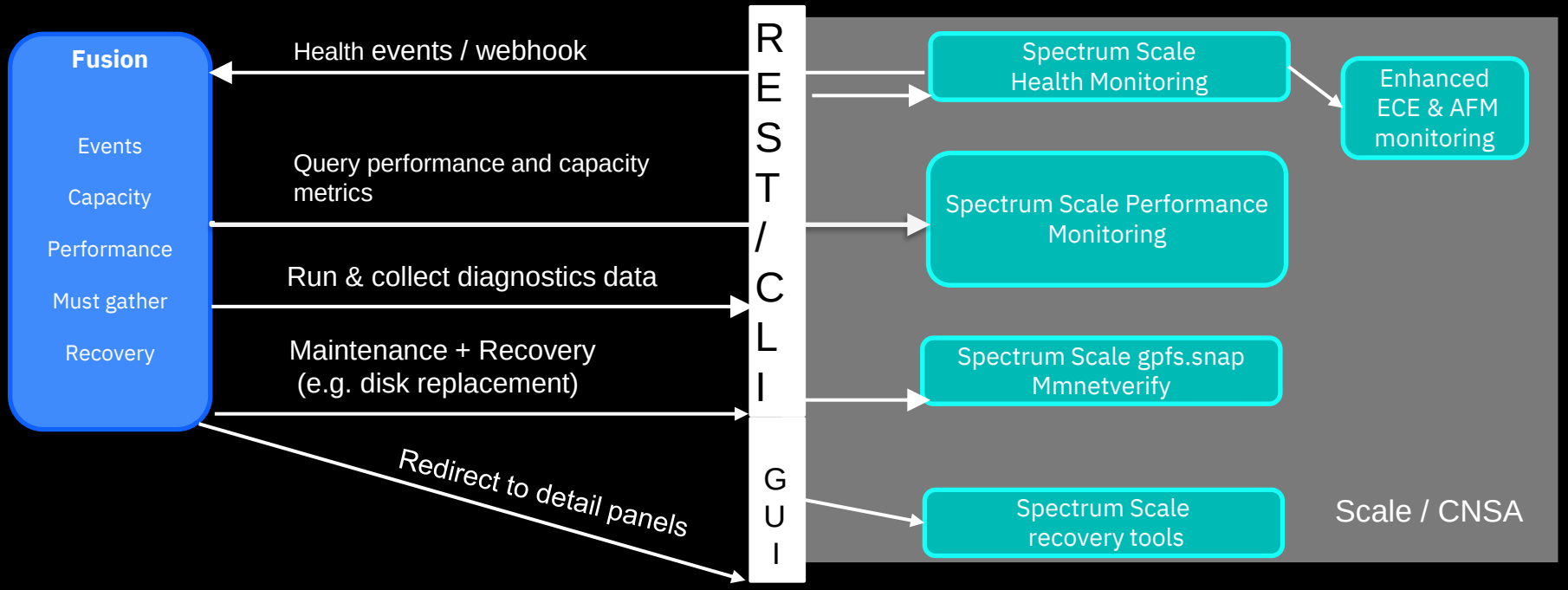
Spectrum Scale ECE

- Spectrum Scale running on storage rich servers connected to each other with a high speed network infrastructure
- ECE supports several erasure coding options and brings much better storage efficiency, e.g. 4+2P, 4+3P, 8+3p and 8+2p Reed Solomon Code
- End to end checksums
- Declustered RAID across nodes
  - Tracks are distributed across the nodes in a recovery group
  - Ability to tolerate multiple node and device failures while preserving access to the data



Fusion/Scale Single pane of glass for system monitoring

- Fusion UI has a central event view and shows selected performance metrics (e.g. capacity)
  - Fusion UI uses existing REST API to query Scale performance data
  - Fusion Eventmanager registers to mmhealth webhook and forwards events to OCP
- Scale UI is used for:
  - Scale Health overview
  - Collect diagnostics data (gpfs.snap),
  - Recovery actions (e.g. disk replacement)



Fusion/Scale : Call Home Integration

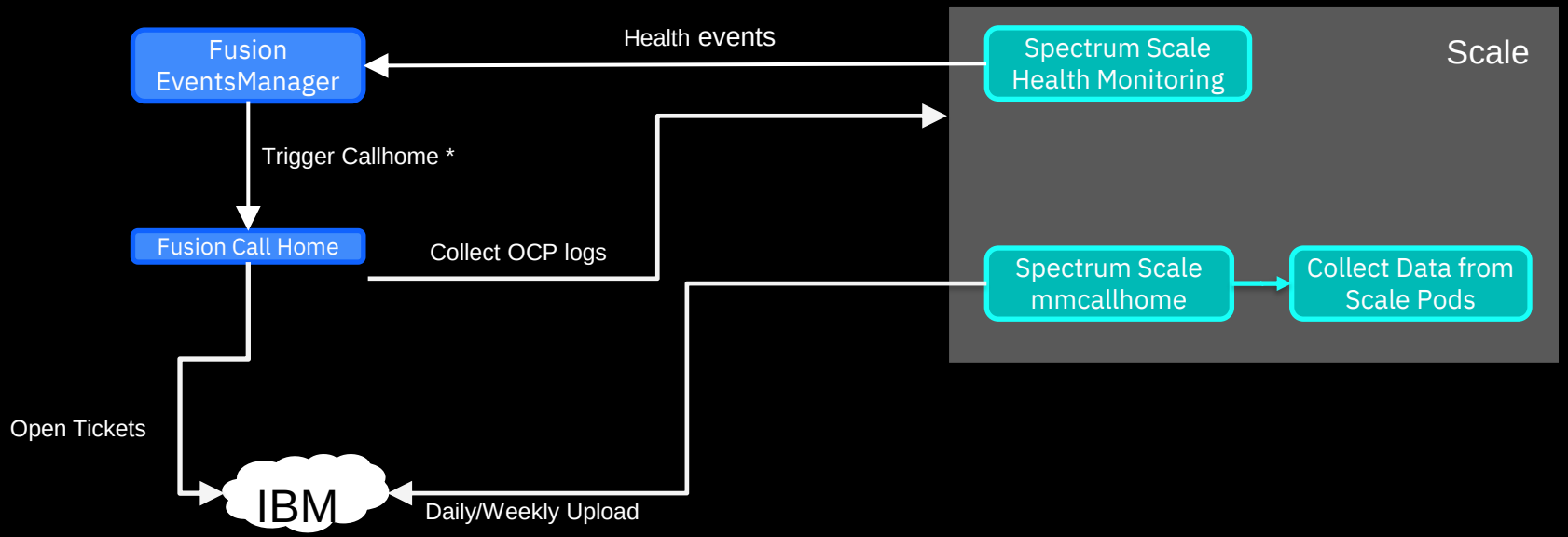
- Spectrum Scale Callhome is configured by Fusion setup:
  - Provide daily/weekly callhome collection
  - Scale callhome is independent from Fusion Callhome

- Fusion will open tickets for selected Scale health events
  - Scale sends events to Fusion through webhook
  - Event name controls if event opens a ticket (allow-list)
    - E.g. gnr_pdisk_replaceable, fsstruct errors, ..

Features

Automatic Ticket opening

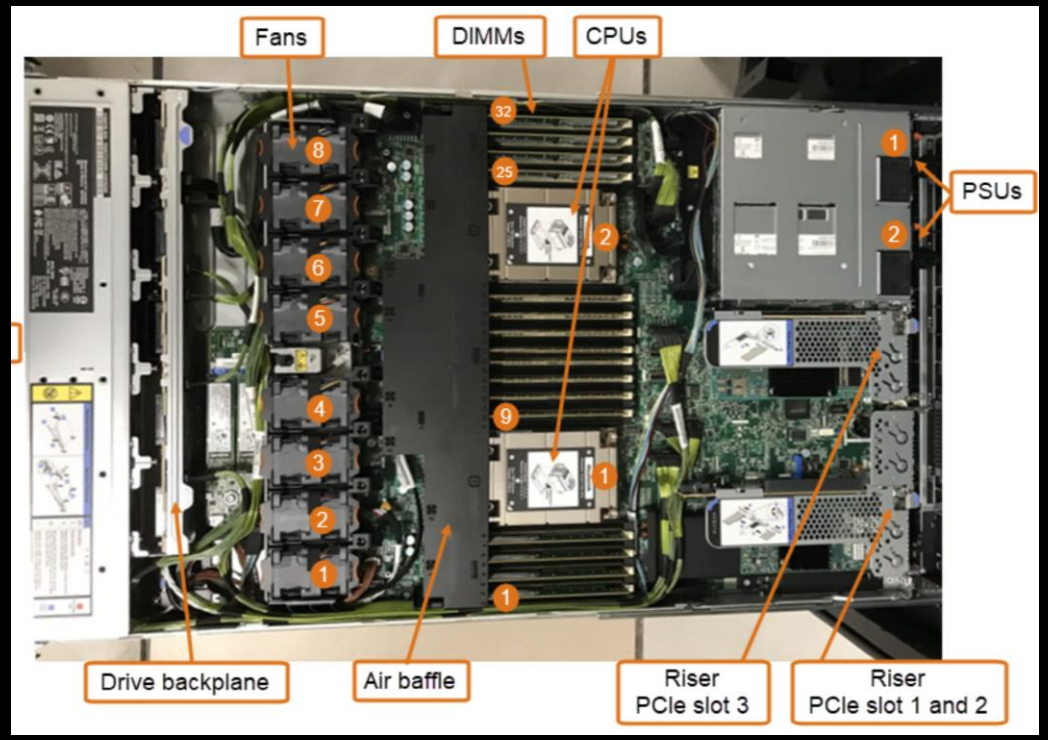
Daily/Weekly data collector for support



Spectrum Fusion HCI 2.3 Update

New in 3Q2022 - V2.3

Fusion HCI now can be ordered with 64-core compute nodes.



Today:

Fusion HCI only supports 32-core nodes

What's changing:

Clients can use StorM to configure Spectrum Fusion HCI with 32 and 64 core compute nodes

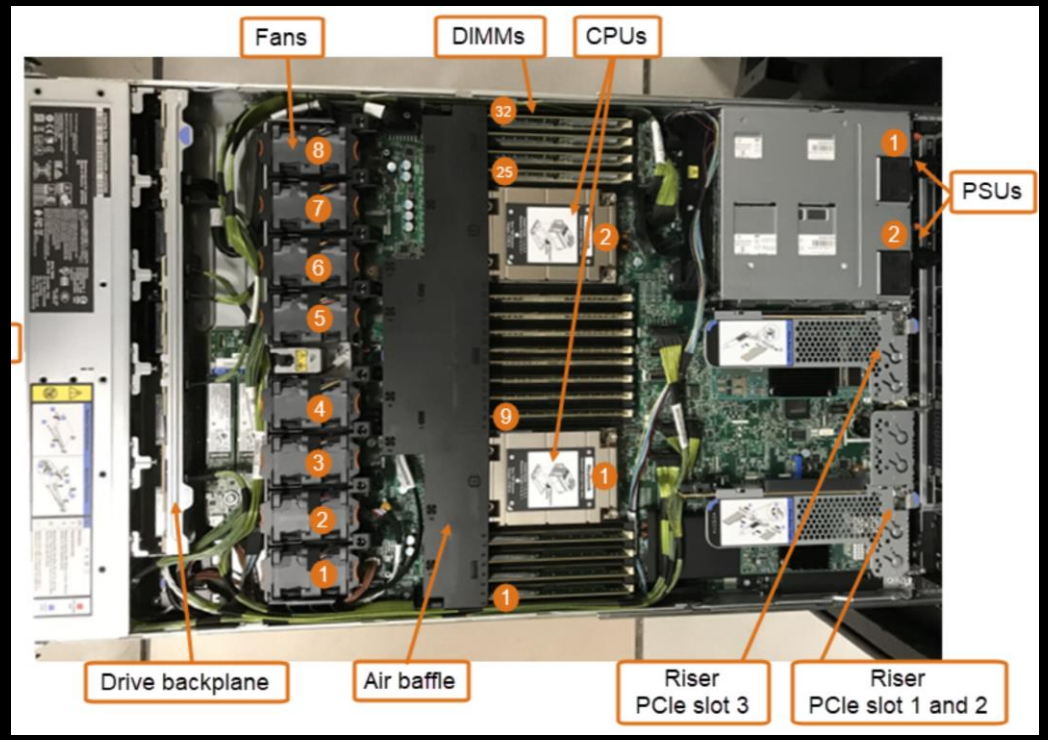
- Flexible: Mix 32 & 64 core nodes within a cluster; control nodes and worker nodes
- 1TB RAM and 2 TB RAM

Benefit:

- Density - more cores per unit of space
- Economics - reduced price per core
- Run workloads on control nodes

New in 3Q2022 - V2.3

Example Configurations



Example 1: High density entry

- 3 x 32 core control nodes (storage)
- 3 x 64 core worker nodes (storage)
- *You still need a minimum of 6 nodes for ECE*

Example 2: Compute rich

- 3 x 32 core control nodes (storage)
- 3 x 32 core worker nodes (storage and infra)
- 1 to 14 64 core worker nodes (either)

Example 3: Max Cores for work

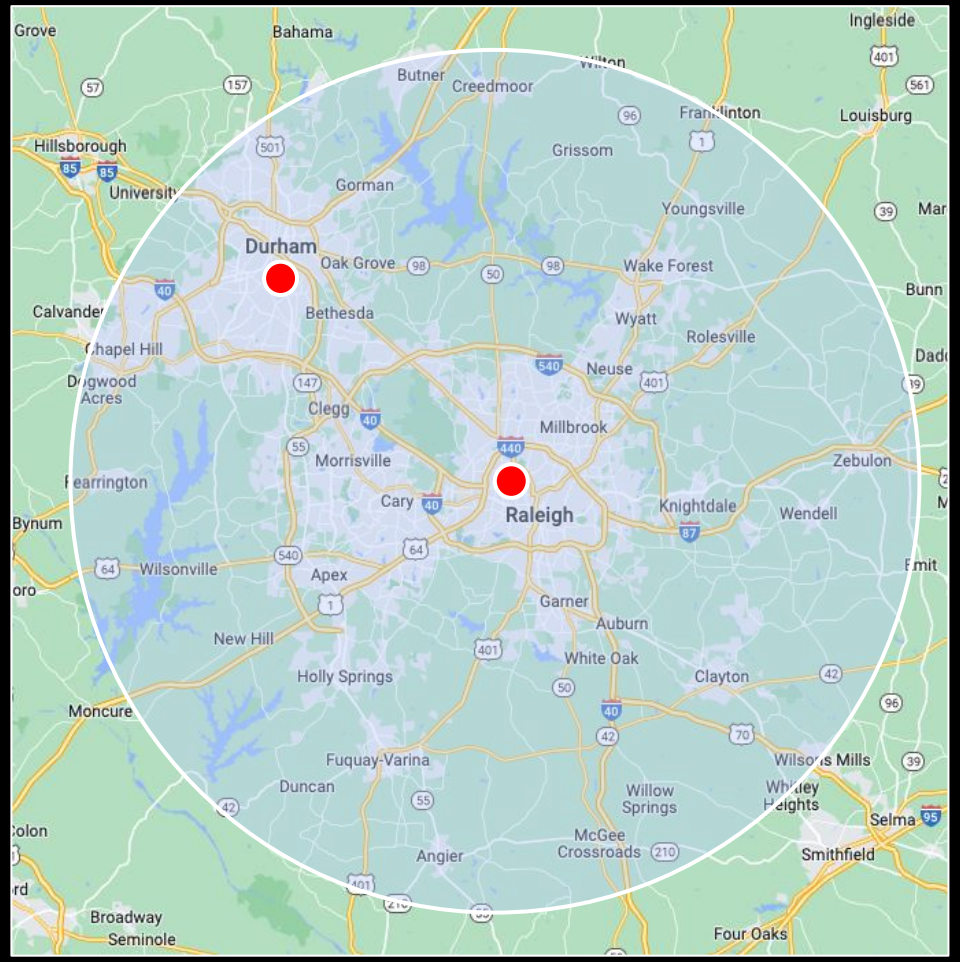
- 3 x 64 core control and worker nodes (storage)
- 3 x 64 core worker nodes (storage)
- 1 to 14 64 core worker nodes (either)
- Must license control nodes

Example 4: Some infrastructure nodes

- 3 x 32 core control nodes
- 3 x 64 core worker nodes
- 1 x 32 core infrastructure node

New in 3Q2022 - V2.3

Fusion HCI – Metro sync DR



Picture source: Google Maps

Today:

It is not possible to configure a pair of Fusion HCI racks in a data sharing relationship for the purpose of improving application availability.

What's changing:

Allow two HCI racks within a metropolitan area network to be paired in a data sharing relationship. Data will be replicated synchronously between the storage on each rack to achieve RPO = 0. RTO can be in minutes.

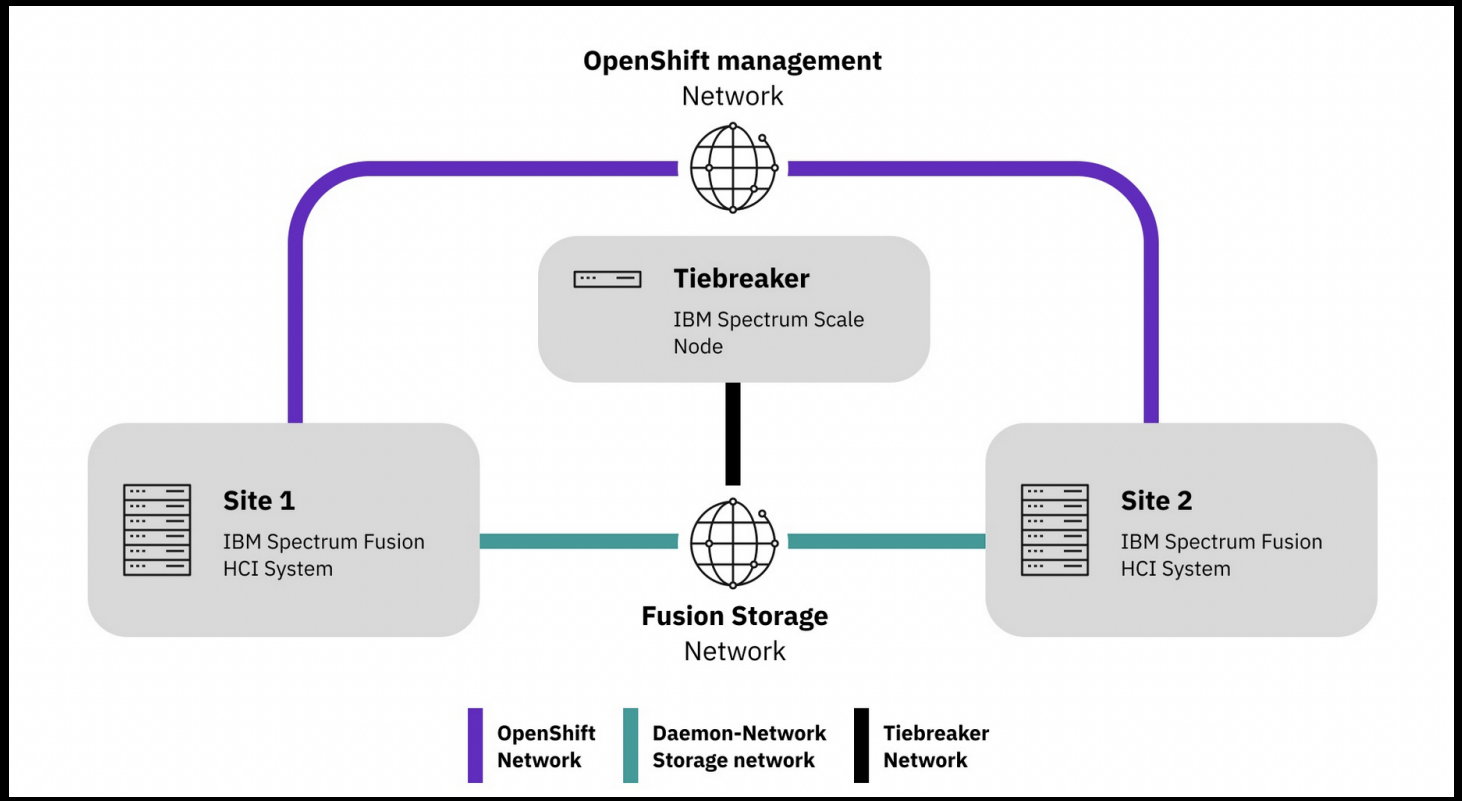
Benefit:

Clients can maintain application availability in the presence of:

- Unplanned outages – loss of facilities
- Planned outages – facilities maintenance
- Application rollout strategies

New in 3Q2022 - V2.3

Fusion HCI – Metro sync DR



Setup

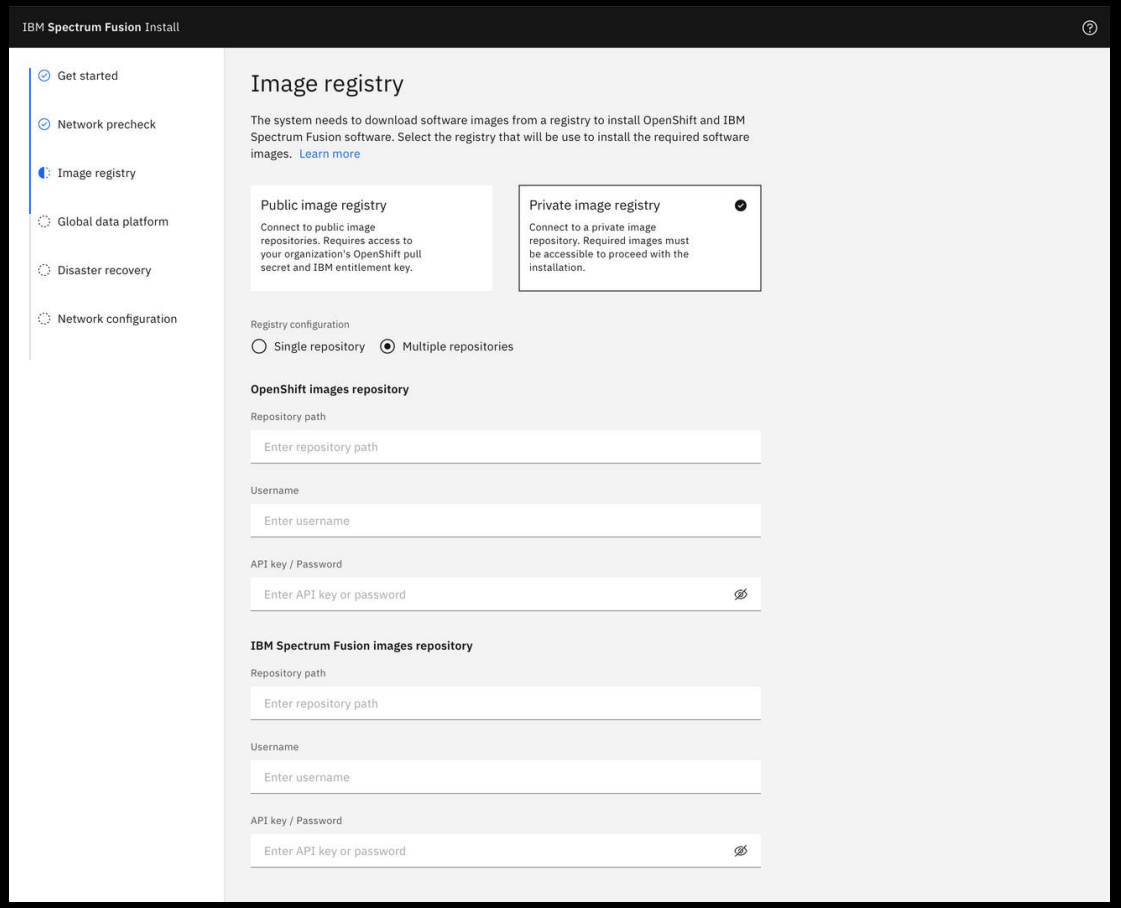
- Set up OpenShift network
- Set up Spectrum Fusion storage network
- Set up tie breaker VM
- Configure Spectrum Fusion as primary / secondary system
- Enable Metro DR on application level

Failover

- HCI A and HCI B see the exact same data - *RPO is zero*
- Trigger relocations manually or automated
  GUI, CLI, K8s Custom Resource, RHACM (roadmap)
- Relocation is a ‘pull’ action - *RTO in seconds*

New in 3Q2022 - V2.3

Fusion HCI - Install from private Artifactory image registry



Example: Select public or private image registry

Today:

Fusion HCI must have line of sight to the IBM and Red Hat registries, direct or via proxy. It is not possible to install a Fusion HCI without access to the internet.

What's changing:

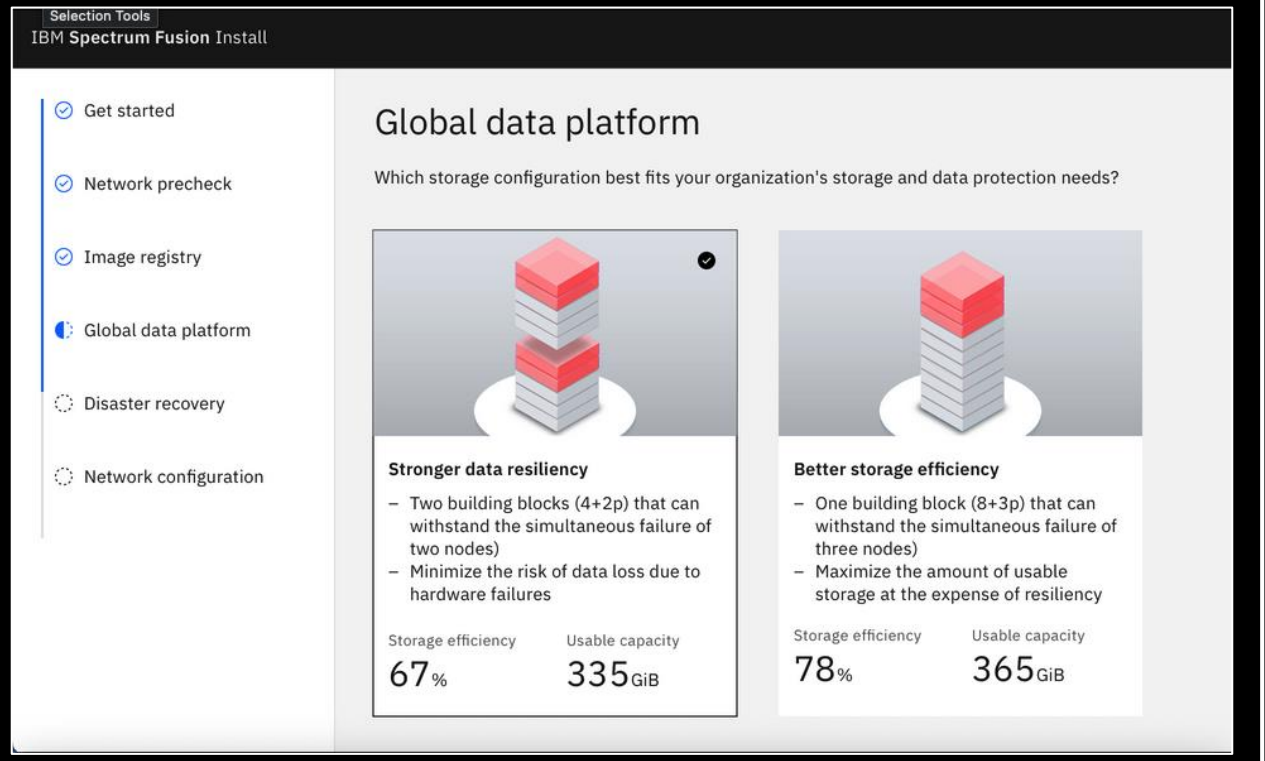
Spectrum Fusion HCI can now take advantage of a client's private enterprise registry they use for installing OpenShift on machines that do not have access to the internet. Support is limited to Artifactory (no Quay, no Docker registry)

Benefit:

Security – clients can stage install and update images in a protected internal registry to keep Fusion HCI isolated from the internet

New in 3Q2022 - V2.3

Fusion HCI – Configurable erasure coding and recovery groups.



Today:

A Fusion HCI rack is limited to 4+2p erasure coding across a single recovery group that can tolerate 2 failures.

What's changing:

At install time, options will be provided to enable a Fusion HCI rack to be configured with 8+3p erasure coding to improve storage efficiency or 2 recovery groups with 4+2p erasure coding. A rack must have minimum of 12 nodes to support configurable erasure coding and recovery groups.

Benefit:

Clients can have more control over rack storage efficiency and resiliency, optimizing to meet their specific requirements. By breaking a rack into multiple smaller recovery groups, recovery time is reduced.

Spectrum Fusion HCI: **Cloud Paks** link

Spectrum Fusion HCI
Cloud Pak Compatibility Sept 19, 2022

IBM Documentation | Search in IBM Spectrum Fusion

IBM Spectrum Fusion

Change version

2.3

Show full table of contents

Filter on titles

- IBM Spectrum Fusion documentation
- What's new
- Product overview
- **IBM Cloud Paks support for IBM Spectrum Fusion**
- Deploying IBM Spectrum Fusion
- Managing services
- Managing workloads
- Configuring encryption for storage
- Accessing remote IBM Spectrum Scale storage cluster in IBM Spectrum Fusion
- Sharing data
- Classifying data
- Protecting data
- Disaster recovery

IBM Spectrum Fusion / 2.3 /

Feedback Product list

IBM Cloud Paks support for IBM Spectrum Fusion

Last Updated: 2022-09-13

Mapping of IBM Cloud Paks with IBM Spectrum Fusion.

IBM Cloud Paks	Version
IBM Cloud Pak for Data	4.5 or later
IBM Cloud Pak for Integration	2021.3 or later
IBM Cloud Pak for Security	1.9 or later
IBM Cloud Pak for Network Automation	2.2 or later
IBM Cloud Pak for Watson AIOPs	3.3 or later

For an example of storage class for use with IBM Cloud Paks, IBM Spectrum Scale version (All), see Dynamic provisioning of Persistent Volume in filesystem.
To use IBM Cloud Paks in IBM Spectrum Fusion, you must create a storage class with specific parameters that includes `shared: true`. For more information about how to create storage classes, see Storage provisioning using Container Storage Interface driver.

- **Validation tools for IBM Cloud Paks**
  It is recommended to run the storage readiness and performance suites on your IBM Spectrum Fusion environments. The suites run on OpenShift® Container Platform to evaluate the setup and performance of your storage setup.

Spectrum Fusion **HCI for AI**

Spectrum Fusion HCI

Accelerate enterprise AI and incorporate AI/ML insights into business operations

Scale-out parallel file system

Cost effective performance for AI/ML workloads with up to 3.6 M IOPS and 220 GiB/s throughput

NVIDIA A100 Tensor Core GPUs

Choice of 3 or 6 NVIDIA A100 GPUs

Discover, catalog, and classify data

Scan unstructured data on NFS filers, S3 object stores, and SMB shares to build meta-data catalogs

Access unstructured data anywhere

provides transparent high-performance access to unstructured data on NFS filers and S3 object stores

RED HAT® OPENSHIFT

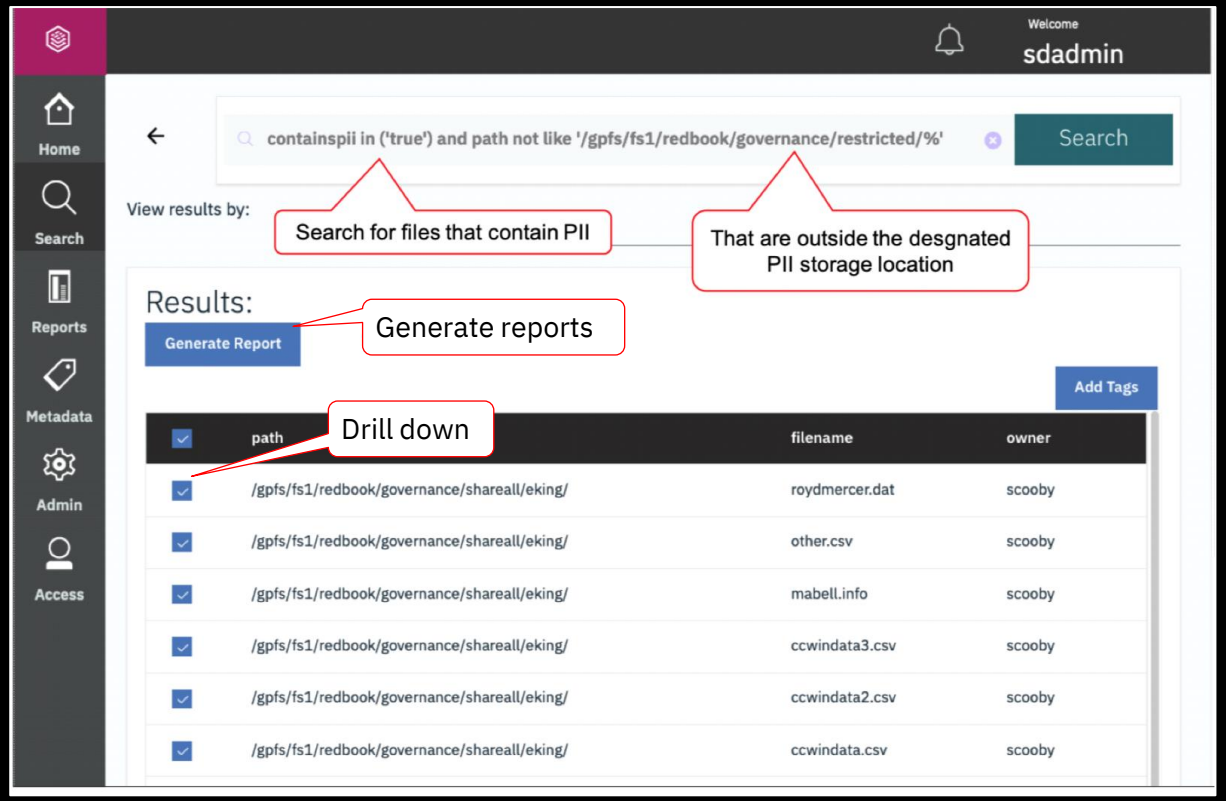
IBM Spectrum Fusion

NVIDIA

Spectrum Fusion

Catalog and classify data

Spectrum Discover



Description:

- Catalog unstructured data on NAS, SMB shares, and S3 object stores
- Tag and classify data
- Data Scientists use the catalog to find data to feed AI/ML workflows

Benefit:

- Enable data scientists to be more productive
- Storage administrators maintain data oversight and governance
- Ensure data scientists always have access to fresh data

Spectrum Fusion HCI: **vs VMware**

3 ways Spectrum Fusion HCI **saves you money** vs VMware

Reduce storage cost by 50%

Fusion HCI's 66% efficient erasure coded storage vs 33% efficient 3-way replicated storage that is common with OpenShift deployments

Improve compute resource utilization by up to 20% (1)

By eliminating hypervisor resource overhead

Reduce OpenShift subscription costs by up to 90%

On dual-socket 64 core x86 servers with bare-metal OpenShift subscriptions vs core-based OpenShift subscriptions required on VM based deployments

(1) Kubernetes on Bare Metal vs. VMs: It's Not Just Performance

Differentiation

Spectrum Fusion HCI
mission critical application hosting platform

	Why it matters
1	**Turnkey solution for mission critical workloads** • Balanced; redundant; dedicated high-speed storage network • Includes Spectrum Fusion mission critical data services • Designed to meet needs of mission critical workloads
2	**Cost** • Competitive hardware costs (with usual IBM discounts) • Reduce OCP subscription costs by 80% vs OCP on VMware • Eliminate cost of VMware • Reduce storage cost by 50% vs 3-way replication • Reduce administrative costs with lifecycle management
3	**Performance** • High-performance parallel, scalable file system • Enterprise grade NVMe drives • Dedicated high speed storage network for predicable scaling under load
4	**Ready for Enterprise AI and MLOps** • High performance file system for unstructured data • Available NVIDIA A100 GPUs • Data Virtualization and Data Discovery

How to say Thank You in 31 different languages on YouTube: https://www.youtube.com/watch?v=w3RnrfVaYAs