Red Hat Storage

Flexible + Scalable Storage

Red Hat’s software-defined storage portfolio: Ceph & Gluster

Patrick Ladd
Technical Account Manager, FSI
pladd@redhat.com
https://people.redhat.com/pladd/

Red Hat Storage Overview

- Software Defined Storage
  - What and Why?
- Red Hat's Portfolio
  - Red Hat Ceph Storage
  - Red Hat Gluster Storage

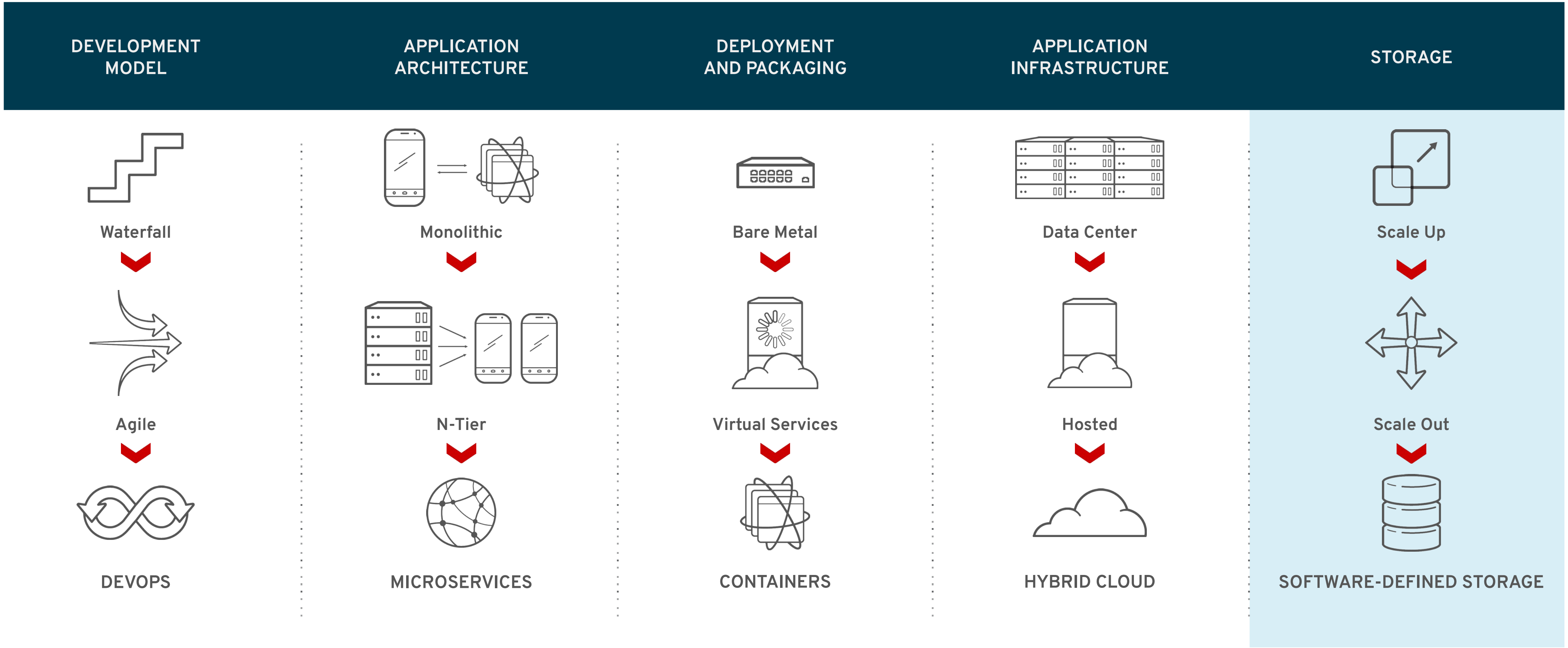
WHAT IS SOFTWARE-DEFINED STORAGE?

SERVER-BASED

CENTRALIZED CONTROL

OPEN ECOSYSTEM

THE ROAD TO SOFTWARE-DEFINED STORAGE

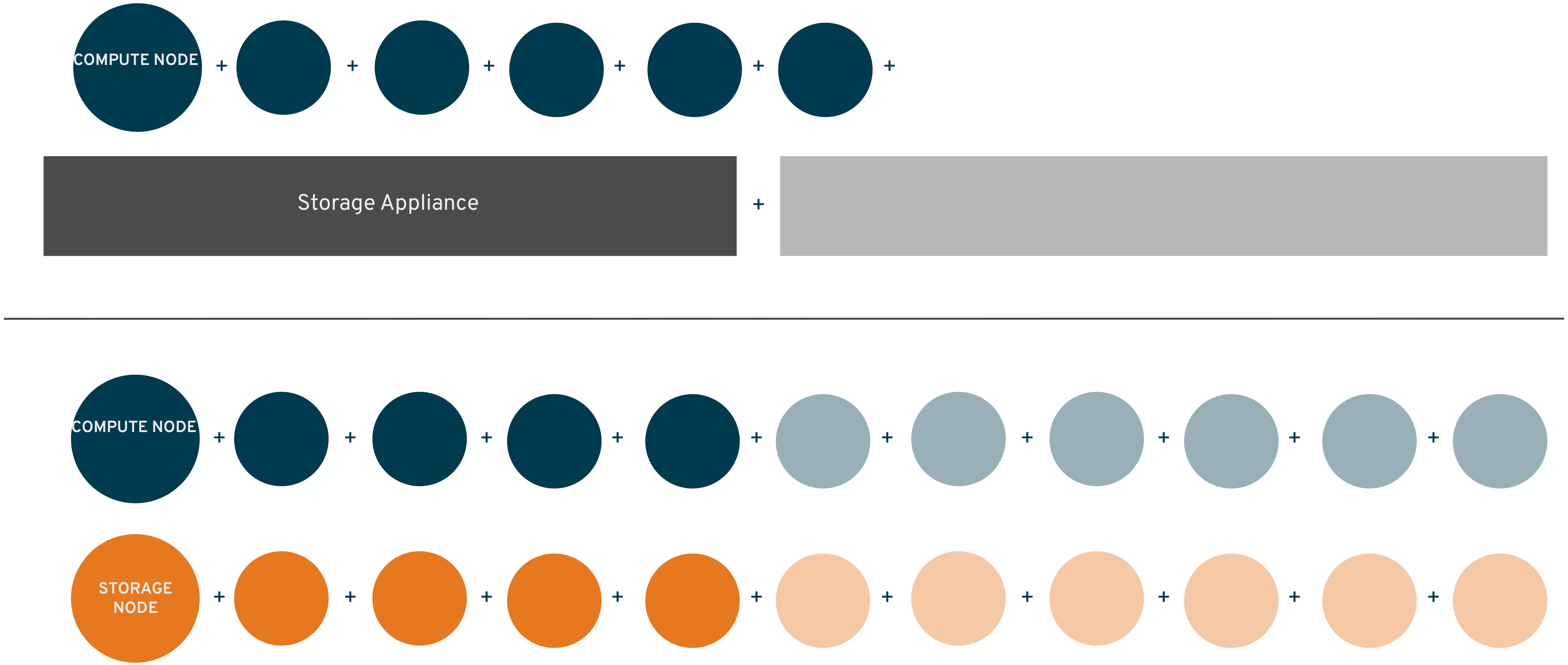


DEVELOPMENT MODEL	APPLICATION ARCHITECTURE	DEPLOYMENT AND PACKAGING	APPLICATION INFRASTRUCTURE	STORAGE
Waterfall	Monolithic	Bare Metal	Data Center	Scale Up
Agile	N-Tier	Virtual Services	Hosted	Scale Out
DEVOPS	MICROSERVICES	CONTAINERS	HYBRID CLOUD	SOFTWARE-DEFINED STORAGE

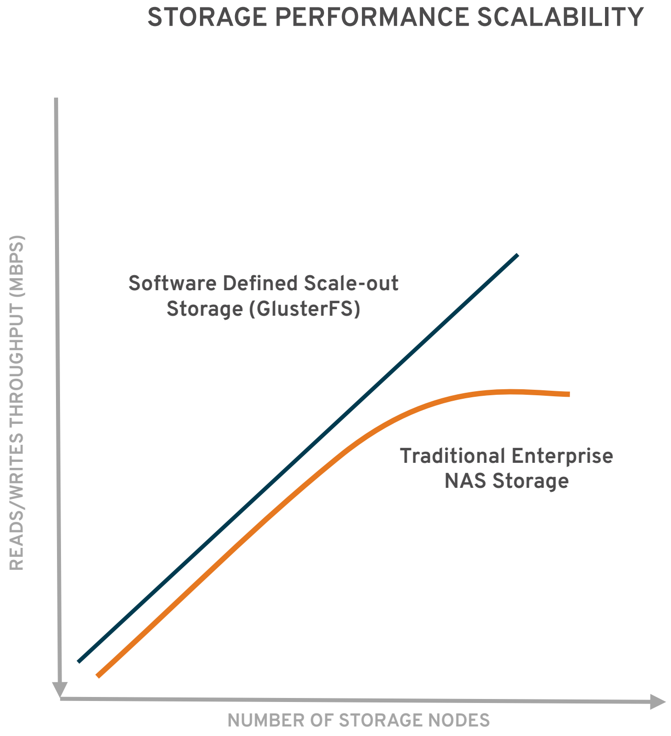
DISRUPTION IN THE STORAGE INDUSTRY

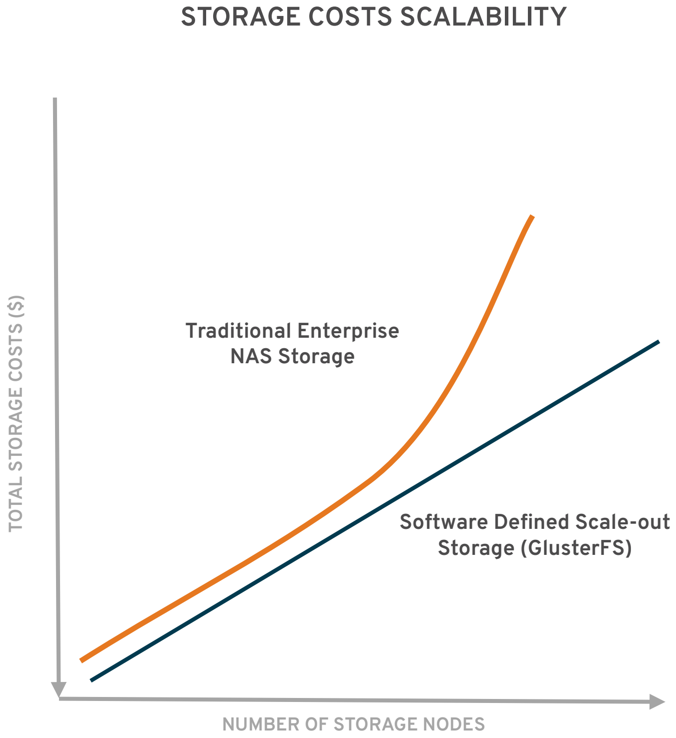
PUBLIC CLOUD STORAGE	← TRADITIONAL APPLIANCES →	SOFTWARE-DEFINED STORAGE
better	COST EFFICIENCY	better
faster	PROVISIONING	faster
more	VENDOR LOCK-IN	less
less	SKILL REQUIRED	more
weaker	GOVERNANCE	stronger
limited	DEPLOYMENT OPTIONS	broad

Virtualized Storage Scales Better



Comparing Throughput and Costs at Scale



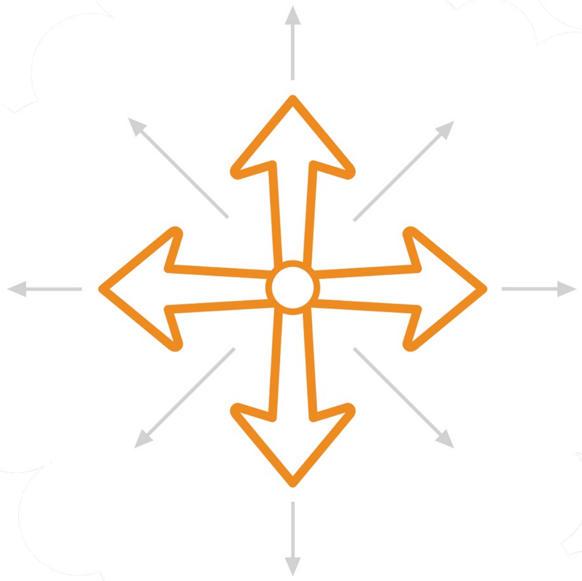


The Robustness of Software

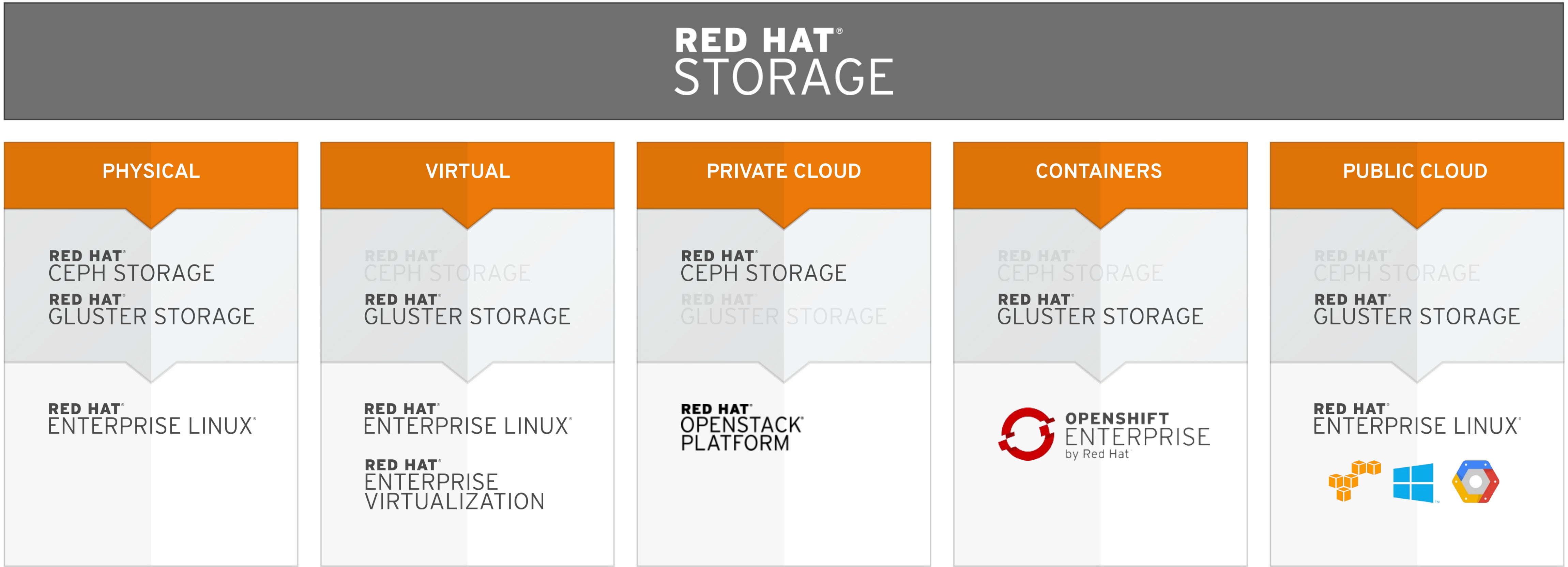
Software can do things hardware can't

Storage services based on software are more flexible than hardware-based implementations

- Can be deployed on bare metal, inside containers, inside VMs, or in the public cloud
- Can deploy on a single server, or thousands, and can be upgraded and reconfigured on the fly
- Grows and shrinks programmatically to meet changing demands



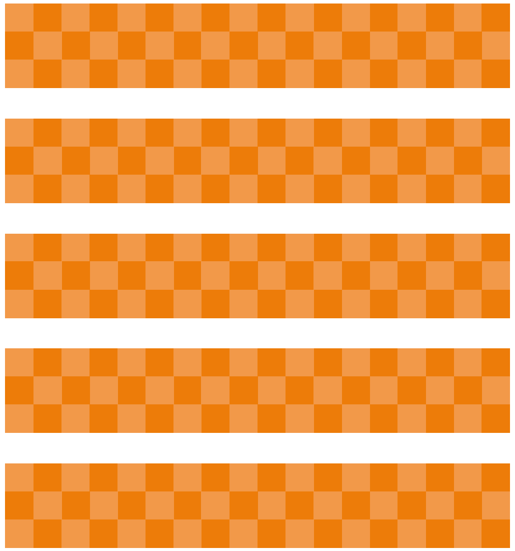
How Storage Fits



RED HAT® STORAGE

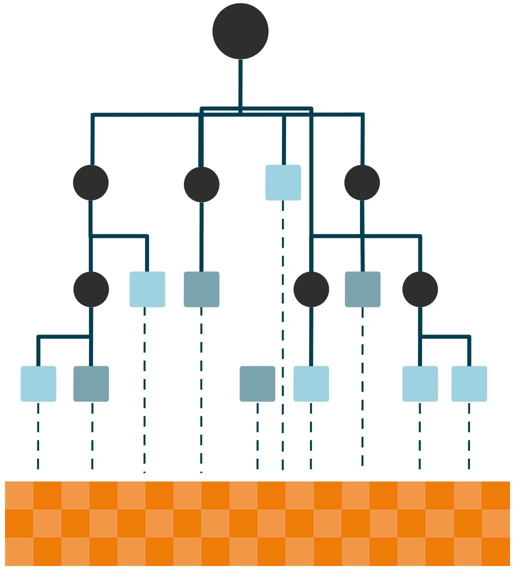
PHYSICAL	VIRTUAL	PRIVATE CLOUD	CONTAINERS	PUBLIC CLOUD
RED HAT® CEPH STORAGE RED HAT® GLUSTER STORAGE	RED HAT® CEPH STORAGE RED HAT® GLUSTER STORAGE	RED HAT® CEPH STORAGE RED HAT® GLUSTER STORAGE	RED HAT® CEPH STORAGE RED HAT® GLUSTER STORAGE	RED HAT® CEPH STORAGE RED HAT® GLUSTER STORAGE
RED HAT® ENTERPRISE LINUX®	RED HAT® ENTERPRISE LINUX® RED HAT® ENTERPRISE VIRTUALIZATION	RED HAT® OPENSTACK® PLATFORM	OPENSHIFT ENTERPRISE by Red Hat	RED HAT® ENTERPRISE LINUX®

DIFFERENT KINDS OF STORAGE



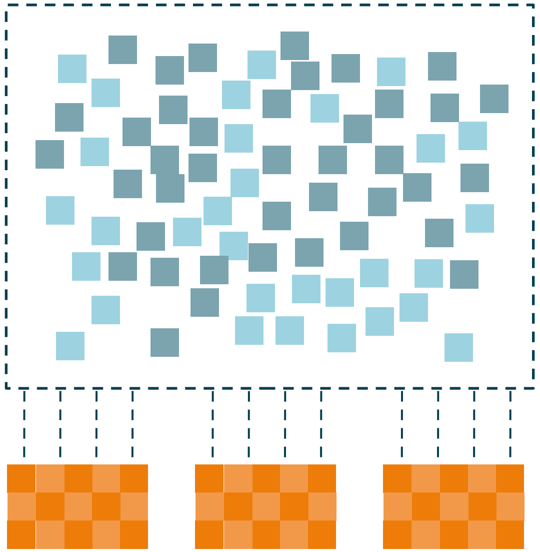
BLOCK STORAGE

Physical storage media appears to computers as a series of sequential blocks of a uniform size.



FILE STORAGE

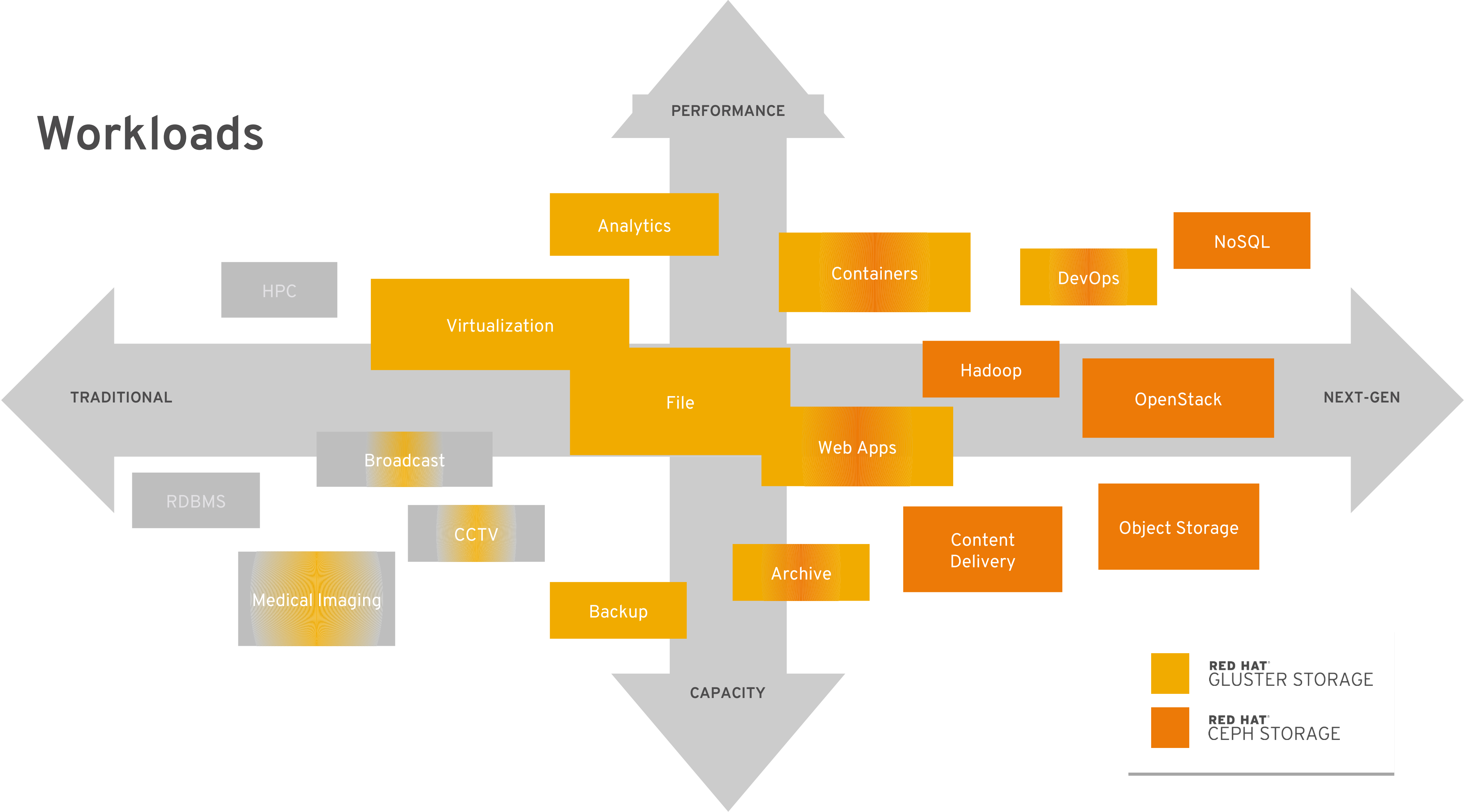
File systems allow users to organize data stored in blocks using hierarchical folders and files.



OBJECT STORAGE

Object stores distribute data algorithmically throughout a cluster of media, without a rigid structure.

Workloads



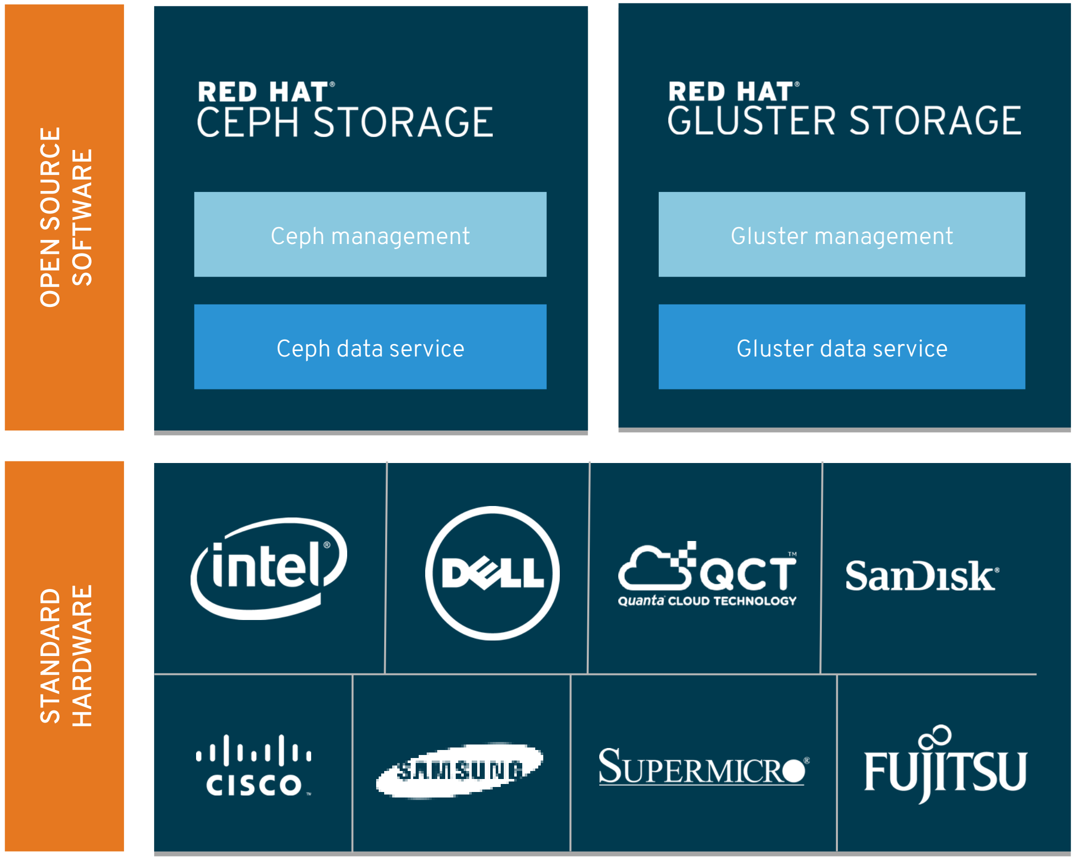
🟨 RED HAT® GLUSTER STORAGE

🟧 RED HAT® CEPH STORAGE

Red Hat Storage Overview

Red Hat Storage

THE RED HAT STORAGE PORTFOLIO



OPEN SOURCE SOFTWARE

RED HAT® CEPH STORAGE

Ceph management

Ceph data service

RED HAT® GLUSTER STORAGE

Gluster management

Gluster data service

STANDARD HARDWARE

- Share-nothing, scale-out architecture provides durability and adapts to changing demands
- Self-managing and self-healing features reduce operational overhead
- Standards-based interfaces and full APIs ease integration with applications and systems
- Supported by the experts at Red Hat

OVERVIEW: RED HAT CEPH STORAGE

RED HAT® CEPH STORAGE

TARGET USE CASES

Cloud Infrastructure

- VM storage with OpenStack® Cinder, Glance Keystone, Manila, and Nova
- Object storage for tenant apps

Rich Media and Archival

S3-compatible object storage

Powerful distributed storage for the cloud and beyond

- Built from the ground up as a next-generation storage system, based on years of research and suitable for powering infrastructure platforms
- Highly tunable, extensible, and configurable, with policy-based control and no single point of failure
- Offers mature interfaces for block and object storage for the enterprise

CUSTOMER HIGHLIGHT: CISCO

Cisco uses Red Hat Ceph Storage to deliver storage for next-generation cloud services

OVERVIEW: RED HAT GLUSTER STORAGE

RED HAT® GLUSTER STORAGE

TARGET USE CASES

Enterprise File Sharing

- Media streaming
- Active Archives

Enterprise Virtualization

Rich Media and Archival

Agile file storage for petabyte-scale workloads

- Purpose-built as a scale-out file store with a straightforward architecture suitable for public, private, and hybrid cloud
- Simple to install and configure, with a minimal hardware footprint
- Offers mature NFS, SMB and HDFS interfaces for enterprise use

CUSTOMER HIGHLIGHT: INTUIT

intuit

Intuit uses Red Hat Gluster Storage to provide flexible, cost-effective storage for its industry-leading financial offerings

Red Hat Gluster Storage

Red Hat
Gluster Storage

Red Hat

GLUSTER FUNDAMENTALS

- Clustered Scale-out General Purpose Storage Platform
- Fundamentally File-Based & POSIX End-to-End
  - Familiar Filesystems Underneath (EXT4, XFS)
  - Familiar Client Access (NFS, Samba, FUSE)
- No Metadata Server
- Standards-Based – Clients, Applications, Networks
- Modular Architecture for Scale and Functionality

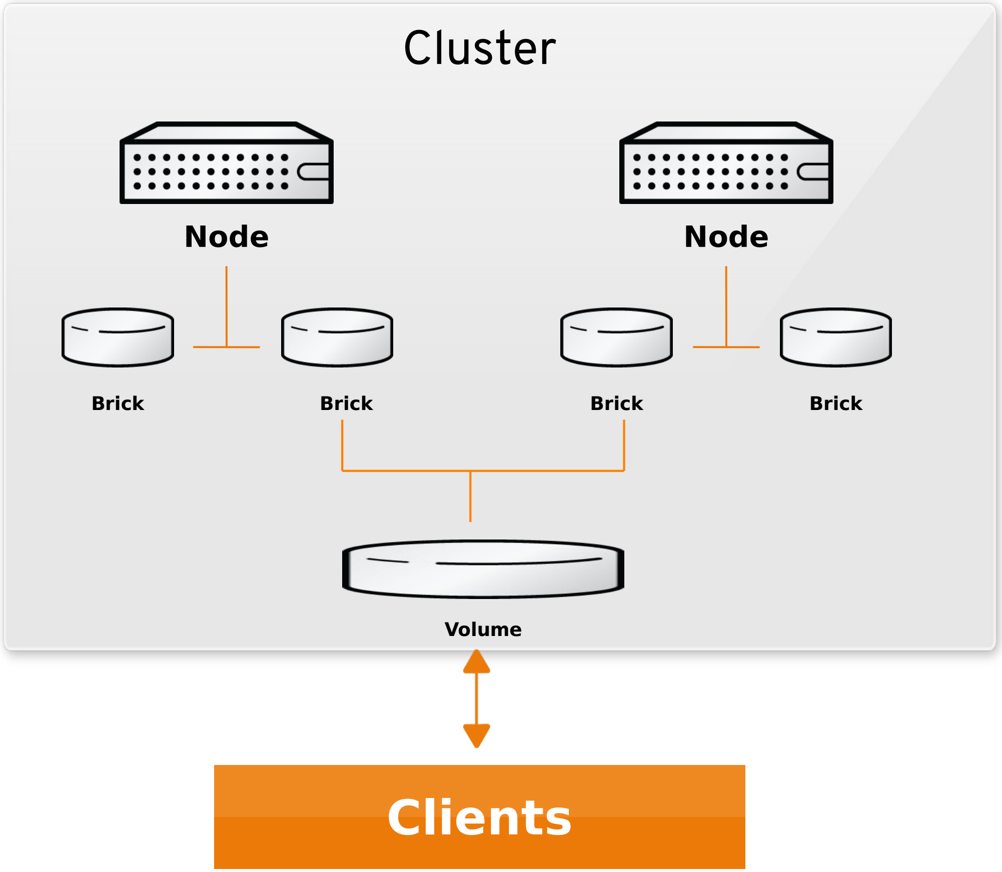
GLUSTER TERMINOLOGY

Cluster: Collection of peer systems

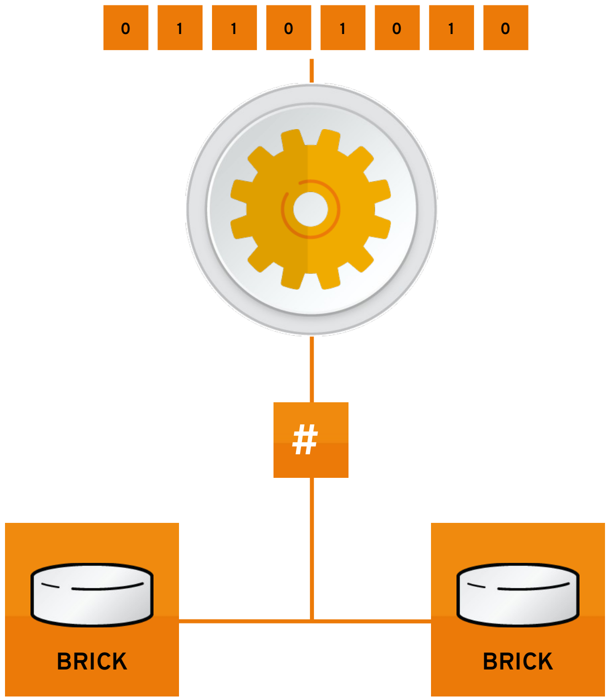
Node: System Participating in Cluster

Brick: Any Linux Block Device

Volume: Bricks taken from one or more hosts presented as a single unit



GLUSTER ELASTIC HASH ALGORITHM



No Central Metadata Server

- Suitable for unstructured data storage
- No single point of failure

Elastic Hashing

- Files assigned to virtual volumes
- Virtual volumes assigned to multiple bricks
- Volumes easily reassigned on-the-fly

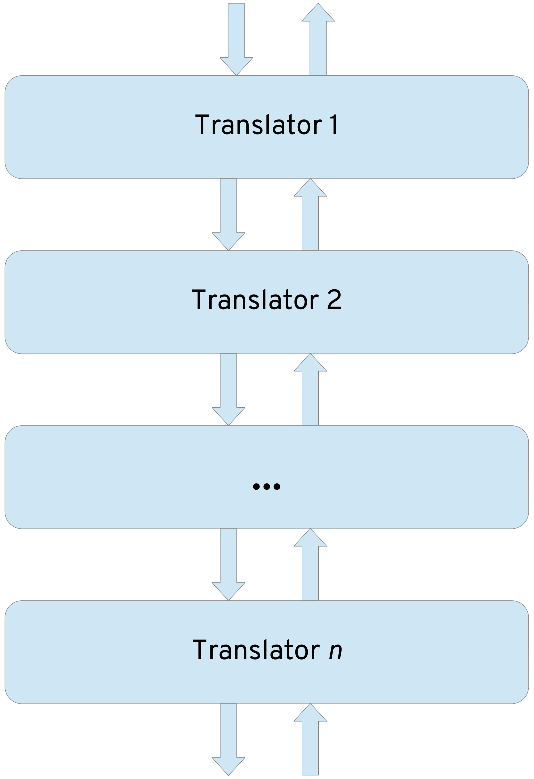
Location Hashed on Filename

- No performance bottleneck
- Eliminates risk scenarios

TRANSLATION LAYERS

Translation layers handle:

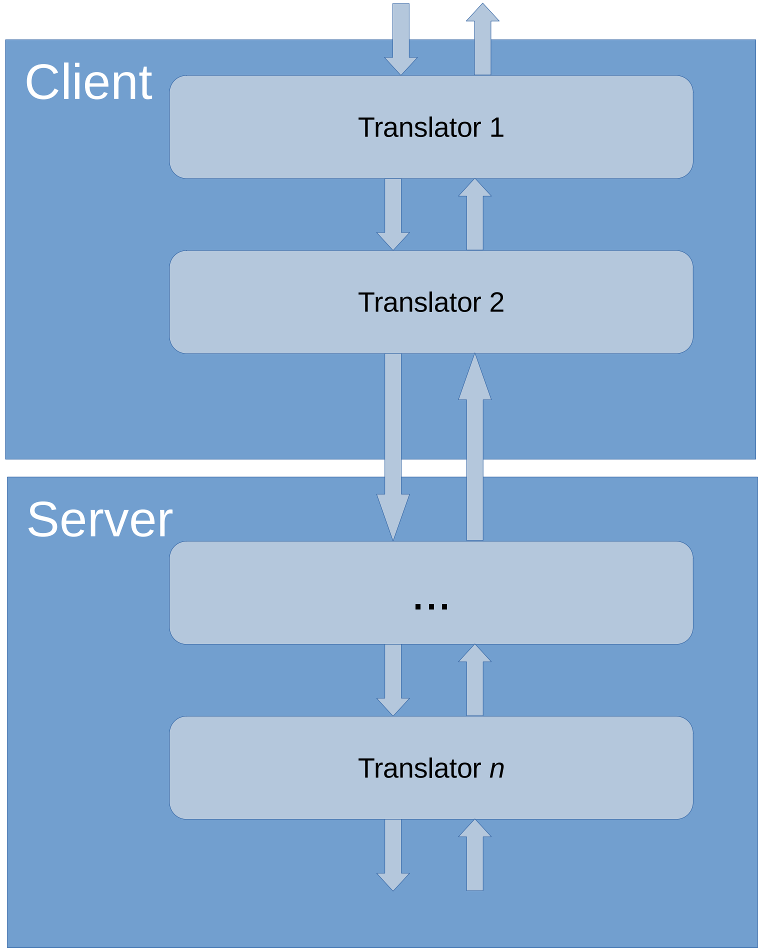
- Data resilience scheme is maintained (replication, erasure coding)
- Metadata is stored and tracked with the object
- Dynamic mapping from virtual volumes to data volumes
- Heal, Rebalance, Bitrot Detection, Geo-Replication, …
- Data translation hierarchy (protocols, encryption, performance, …)
- Health monitoring, alerting, and response



SERVER- AND CLIENT-SIDE TRANSLATORS

Translations layers may be distributed!

- Some layers in the translator stack may be implemented on the client
- Higher performance and efficiency



Client

Translator 1

Translator 2

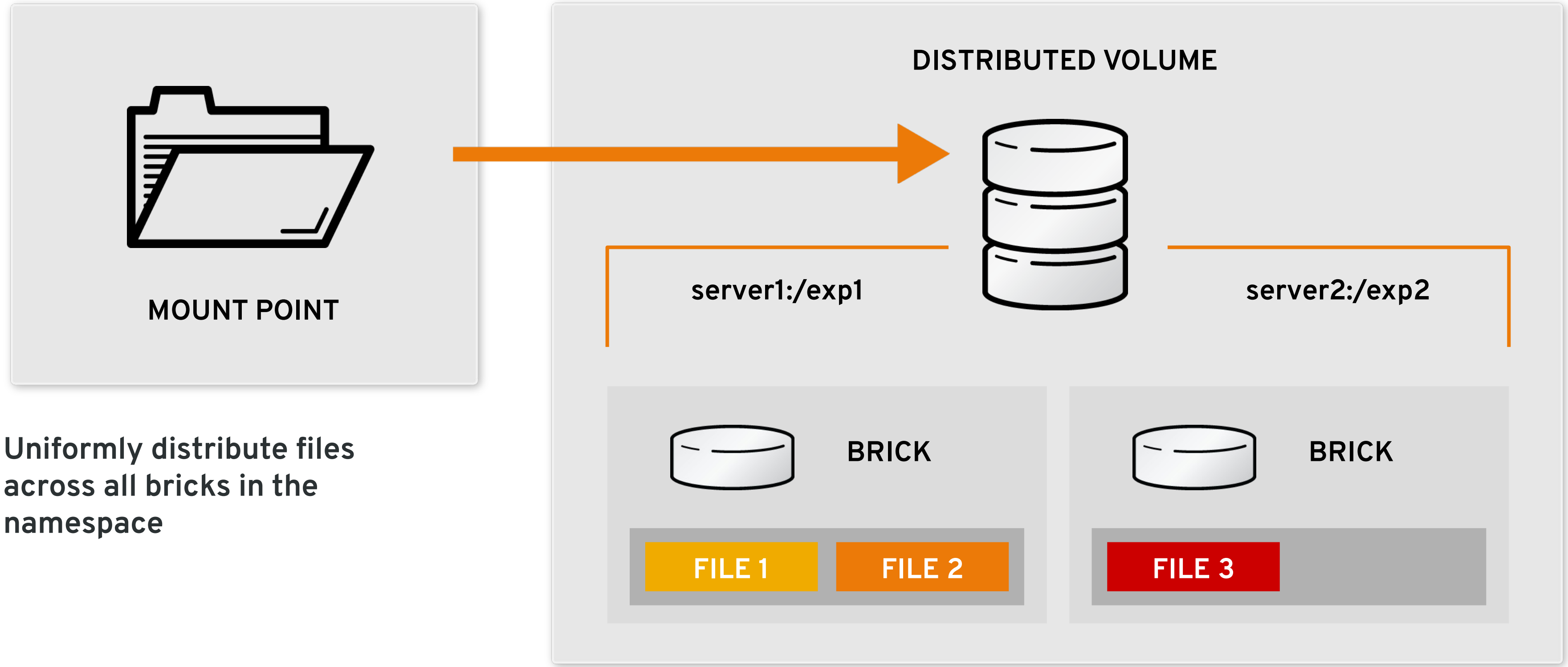
Server

...

Translator *n*

GLUSTER DEFAULT DATA PLACEMENT

Distributed Volume

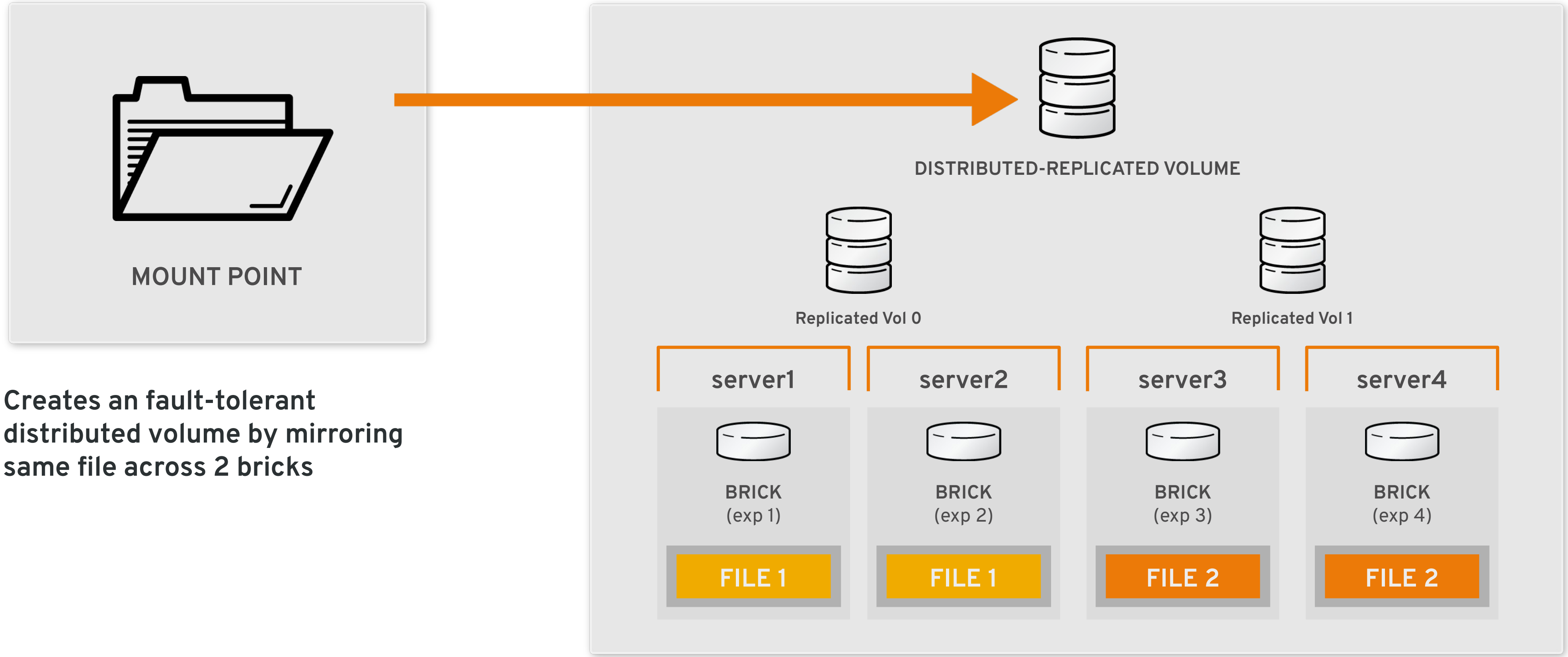


Uniformly distribute files across all bricks in the namespace

Red Hat

GLUSTER FAULT-TOLERANT DATA PLACEMENT

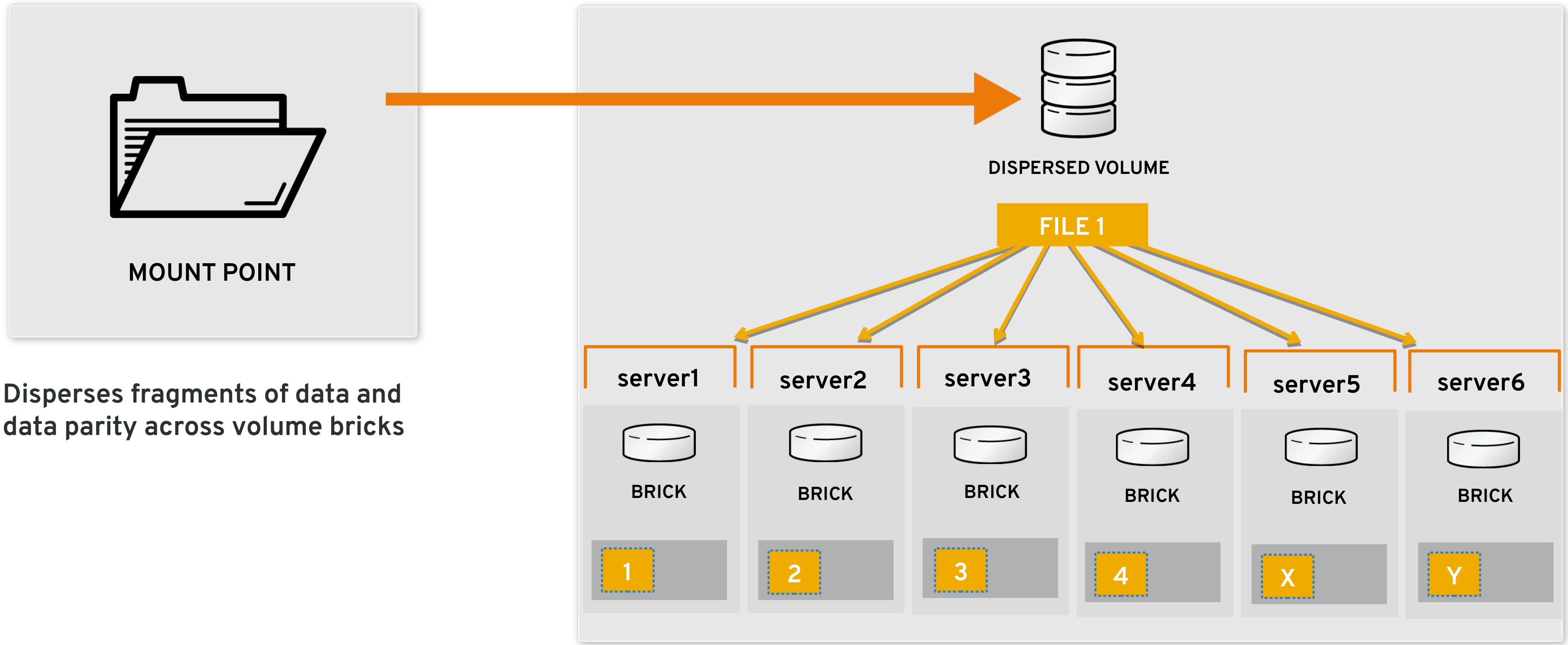
Distributed-Replicated Volume



Creates an fault-tolerant distributed volume by mirroring same file across 2 bricks

GLUSTER ERASURE CODING

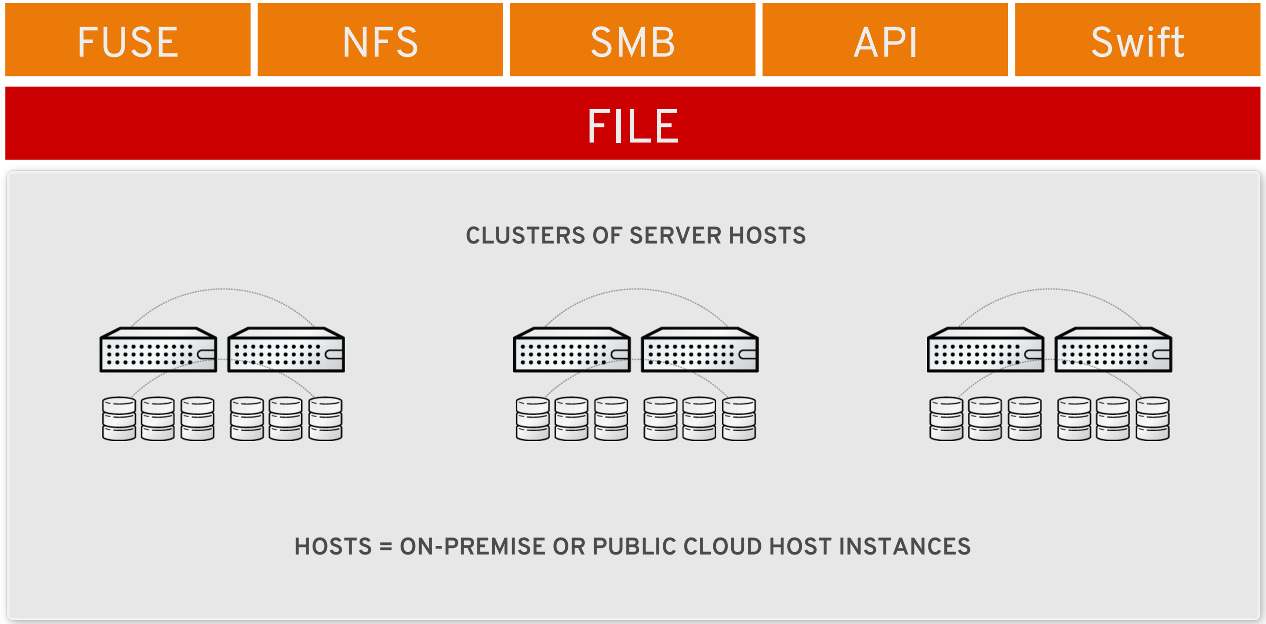
Storing more data with less hardware



Disperses fragments of data and data parity across volume bricks

GLUSTER CLIENT ACCESS

Multi-protocol distributed file system access with optional Swift object translator



FUSE | NFS | SMB | API | Swift

FILE

CLUSTERS OF SERVER HOSTS

HOSTS = ON-PREMISE OR PUBLIC CLOUD HOST INSTANCES

GLUSTER GEO-REPLICATION

Multi-site content distribution

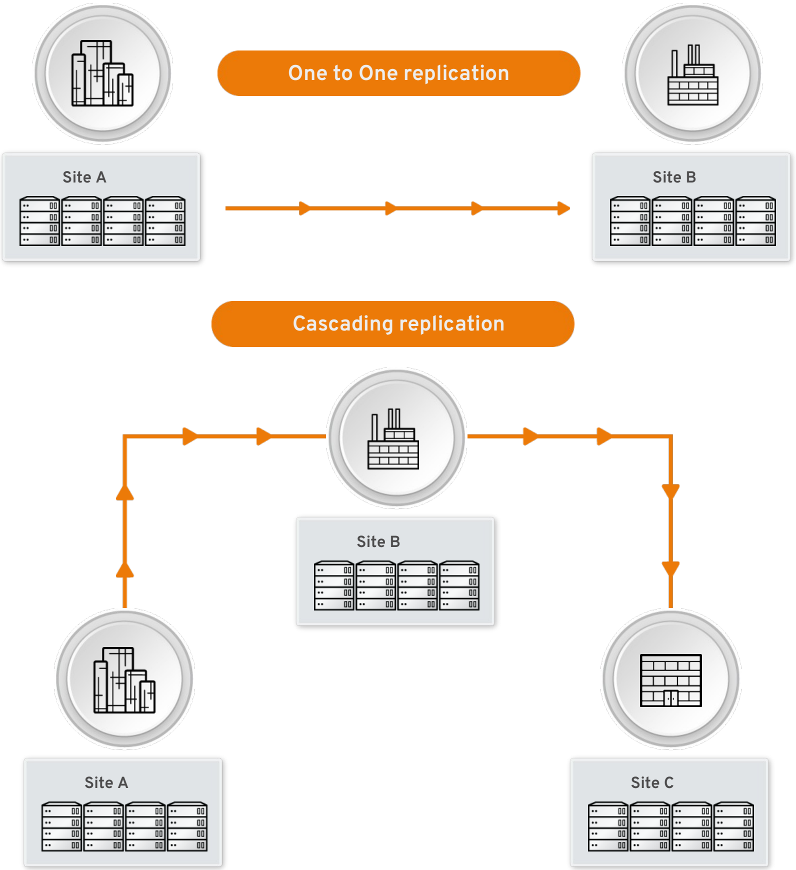
Asynchronous across LAN, WAN, or Internet

Master-slave model, cascading possible

Continuous and incremental

Multiple configurations

- One to one
- One to many
- Cascading

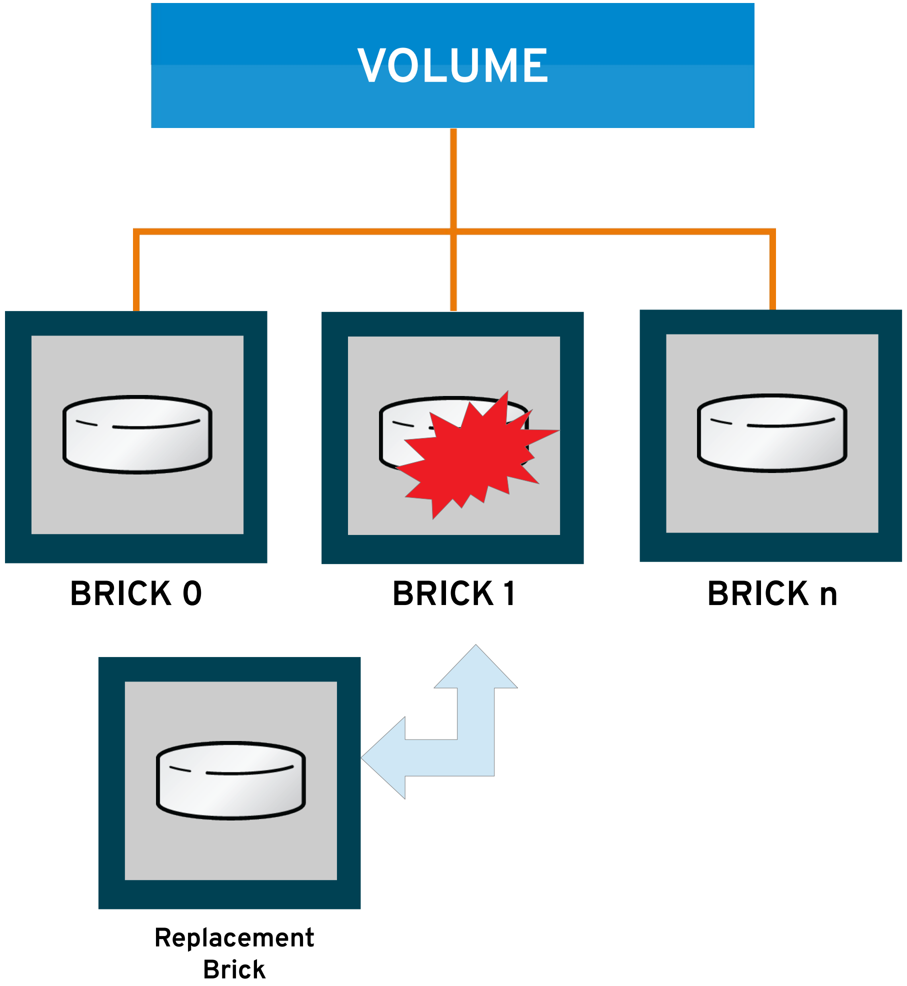


SELF HEALING

- Automatic Repair of Files
  - As they are accessed
  - Periodic via Daemon

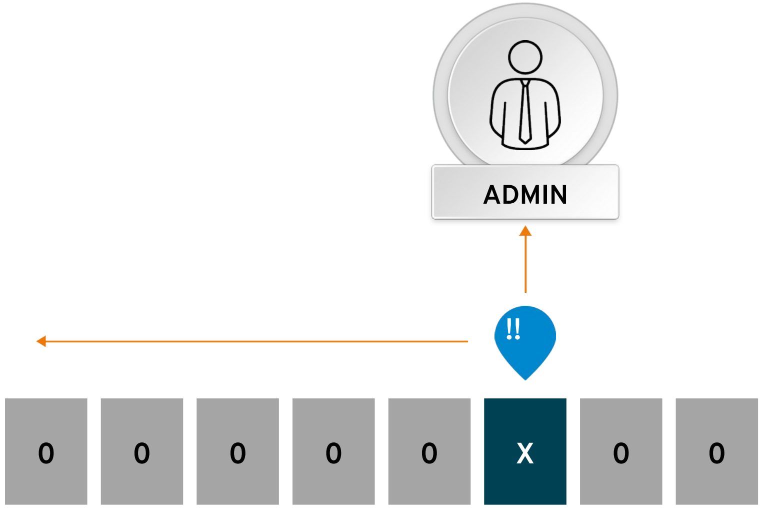
Scenarios:

- Node offline
  - Bricks on node need to be caught up to current
- Node or brick loss
  - New brick needs to be completely rebuilt



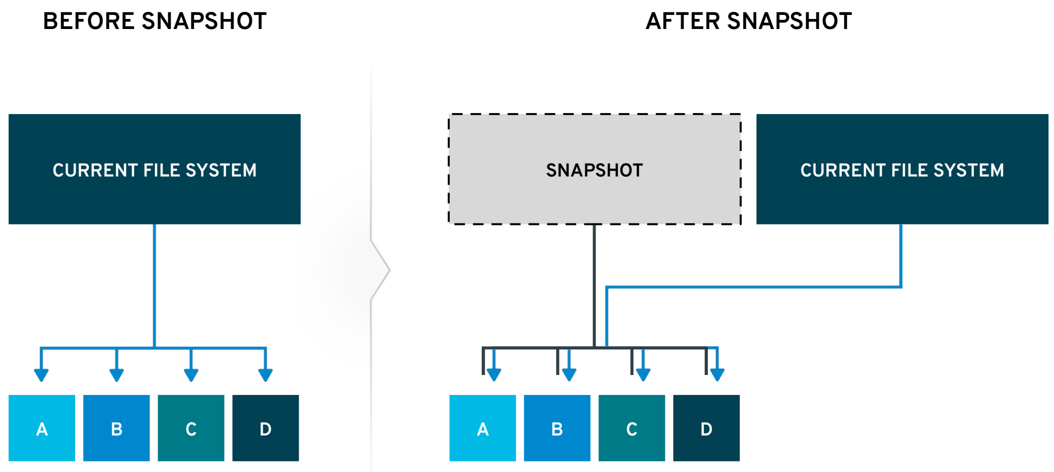
BIT ROT DETECTION

- Scans data periodically for bit rot
- Check sums are computed when files are accessed and compared against previously stored values
- On mismatch, an error is logged for the storage admin



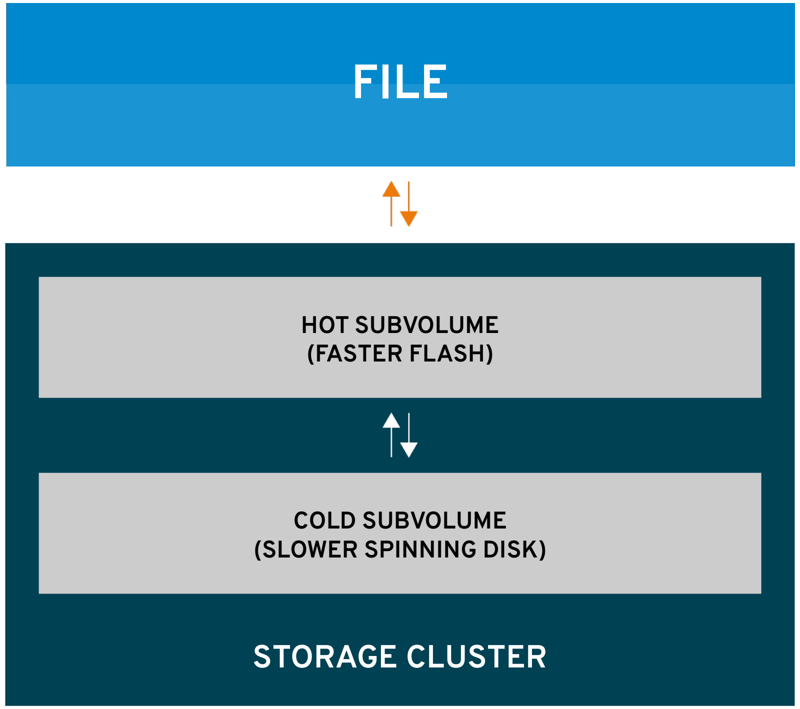
SNAPSHOTS

- Volume level, ability to create, list, restore, and delete
- LVM2 based, operates only on thin-provisioned volumes
- User serviceable snapshots
- Crash consistent image



TIERING

- Automated promotion and demotion of data between “hot” and “cold” sub volumes
- Based on frequency of access
- Cost-effective flash acceleration



Red Hat

QUOTAS

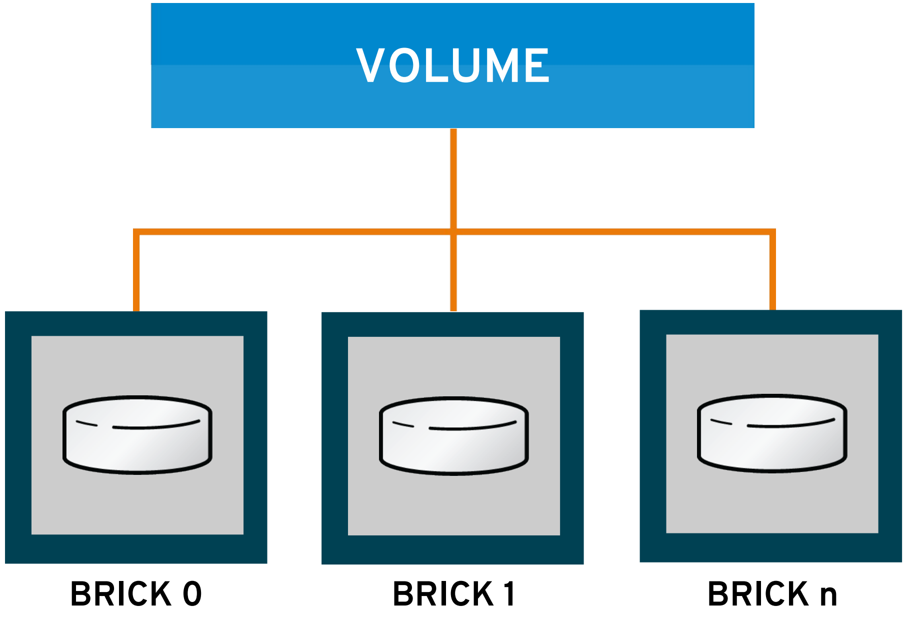
- Control disk utilization at both directory and volume level

Quota Limits

- Two levels of quota limits: Soft (default) and hard
- Warning messages issued on reaching soft quota limit
- Write failures with EDQUAT message after hard limit is reached

Global vs. Local Limits

- Quota is global (per volume)
- Files are psuedo-randomly distributed across bricks

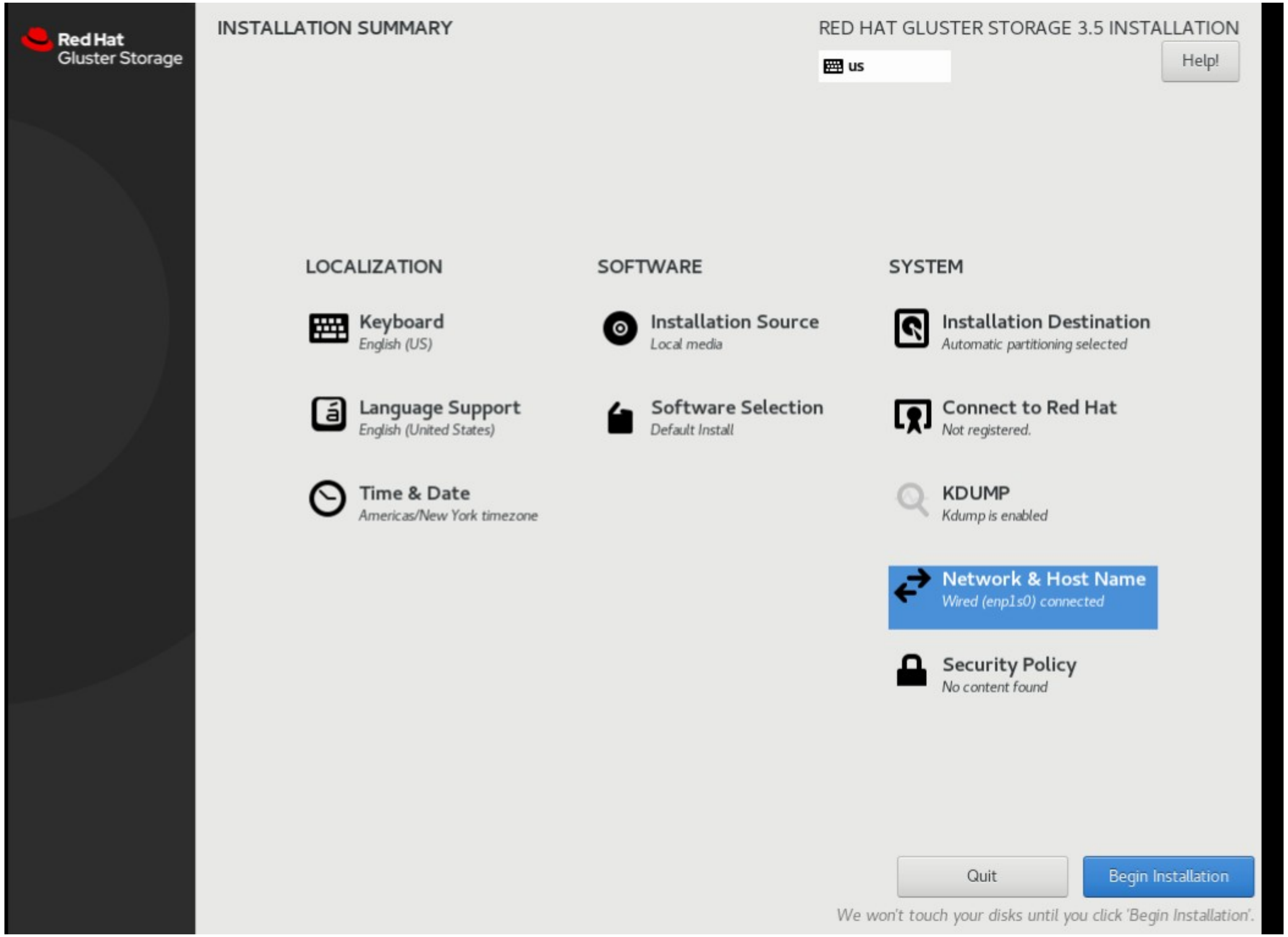


Red Hat Gluster Storage Demo

Red Hat
Ceph Storage

Red Hat

INSTALL



Red Hat Gluster Storage

INSTALLATION SUMMARY

RED HAT GLUSTER STORAGE 3.5 INSTALLATION

us

LOCALIZATION

Keyboard
English (US)

Language Support
English (United States)

Time & Date
Americas/New York timezone

SOFTWARE

Installation Source
Local media

Software Selection
Default Install

SYSTEM

Installation Destination
Automatic partitioning selected

Connect to Red Hat
Not registered.

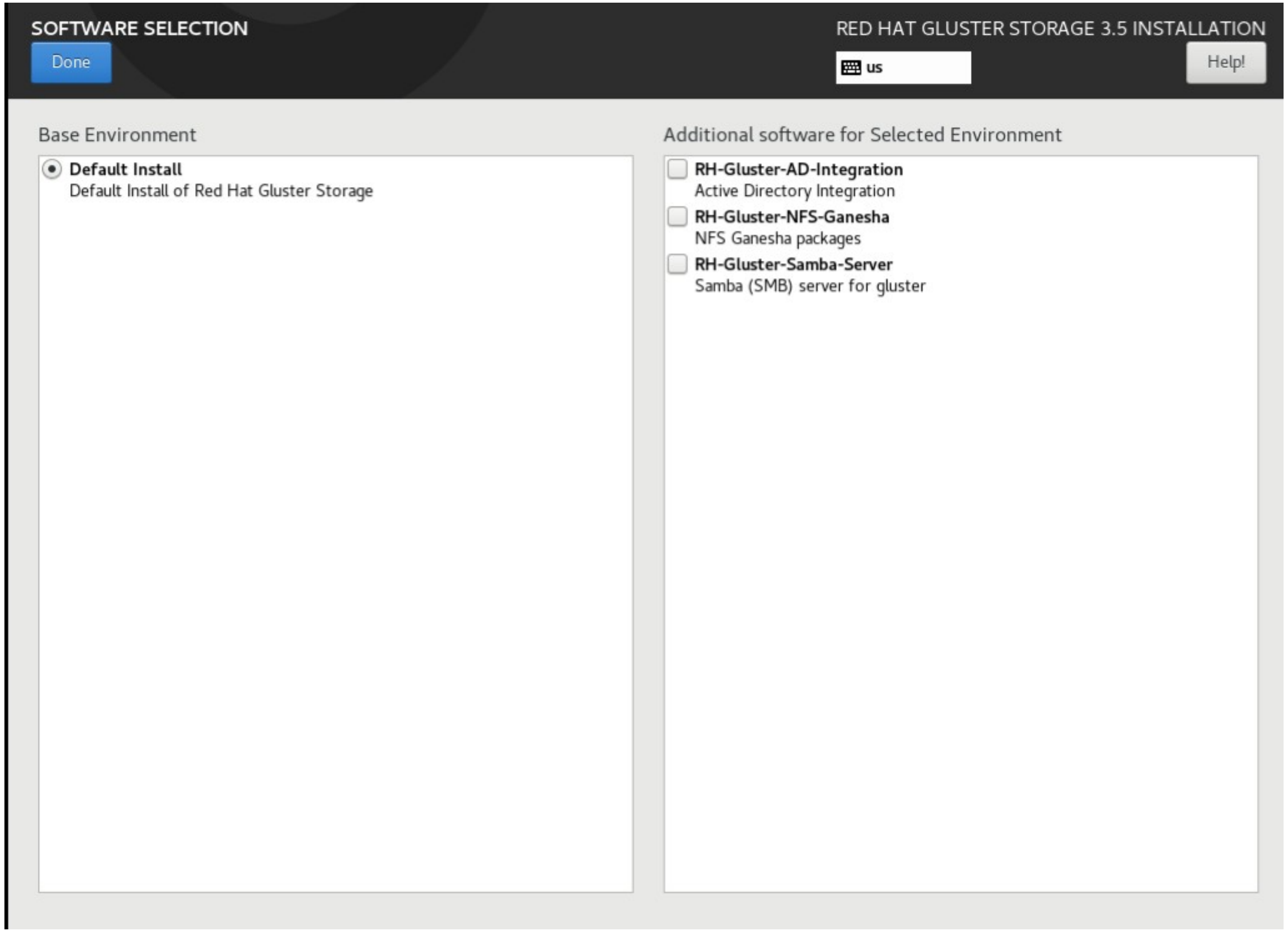
KDUMP
Kdump is enabled

Network & Host Name
Wired (enp1s0) connected

Security Policy
No content found

We won't touch your disks until you click 'Begin Installation'.

INSTALL



SOFTWARE SELECTION

RED HAT GLUSTER STORAGE 3.5 INSTALLATION

Base Environment

● **Default Install**
Default Install of Red Hat Gluster Storage

Additional software for Selected Environment

□ **RH-Gluster-AD-Integration**
Active Directory Integration

□ **RH-Gluster-NFS-Ganesha**
NFS Ganesha packages

□ **RH-Gluster-Samba-Server**
Samba (SMB) server for gluster

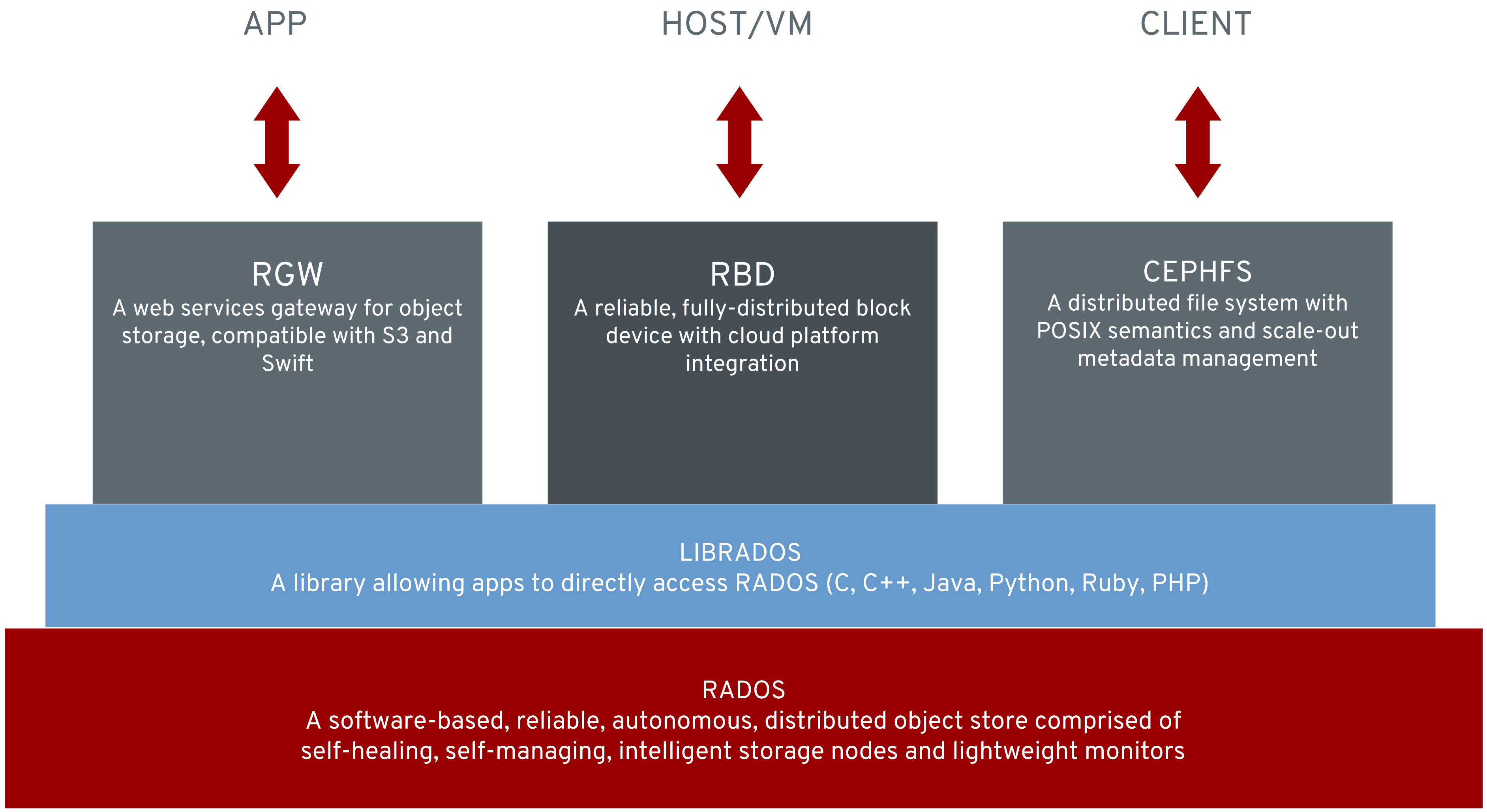
Red Hat
Ceph Storage

Red Hat
Ceph Storage

CEPH FUNDAMENTALS

- Single, efficient, unified storage platform (object, block, file)
- User-driven storage lifecycle management with 100% API coverage
- Integrated, easy-to-use management console
- Designed for cloud infrastructure and emerging workloads

CEPH ARCHITECTURE



APP

HOST/VM

CLIENT

RGW
A web services gateway for object storage, compatible with S3 and Swift

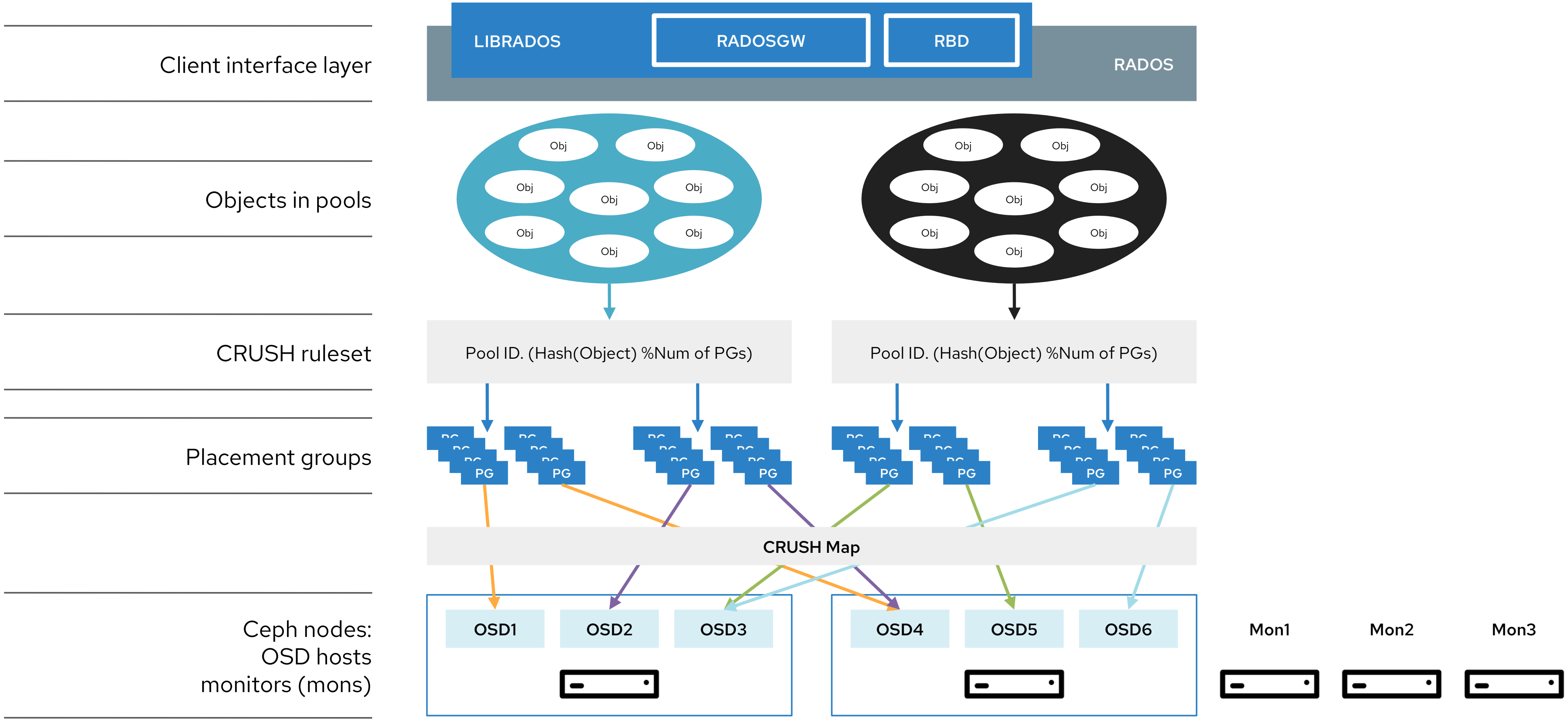
RBD
A reliable, fully-distributed block device with cloud platform integration

CEPHFS
A distributed file system with POSIX semantics and scale-out metadata management

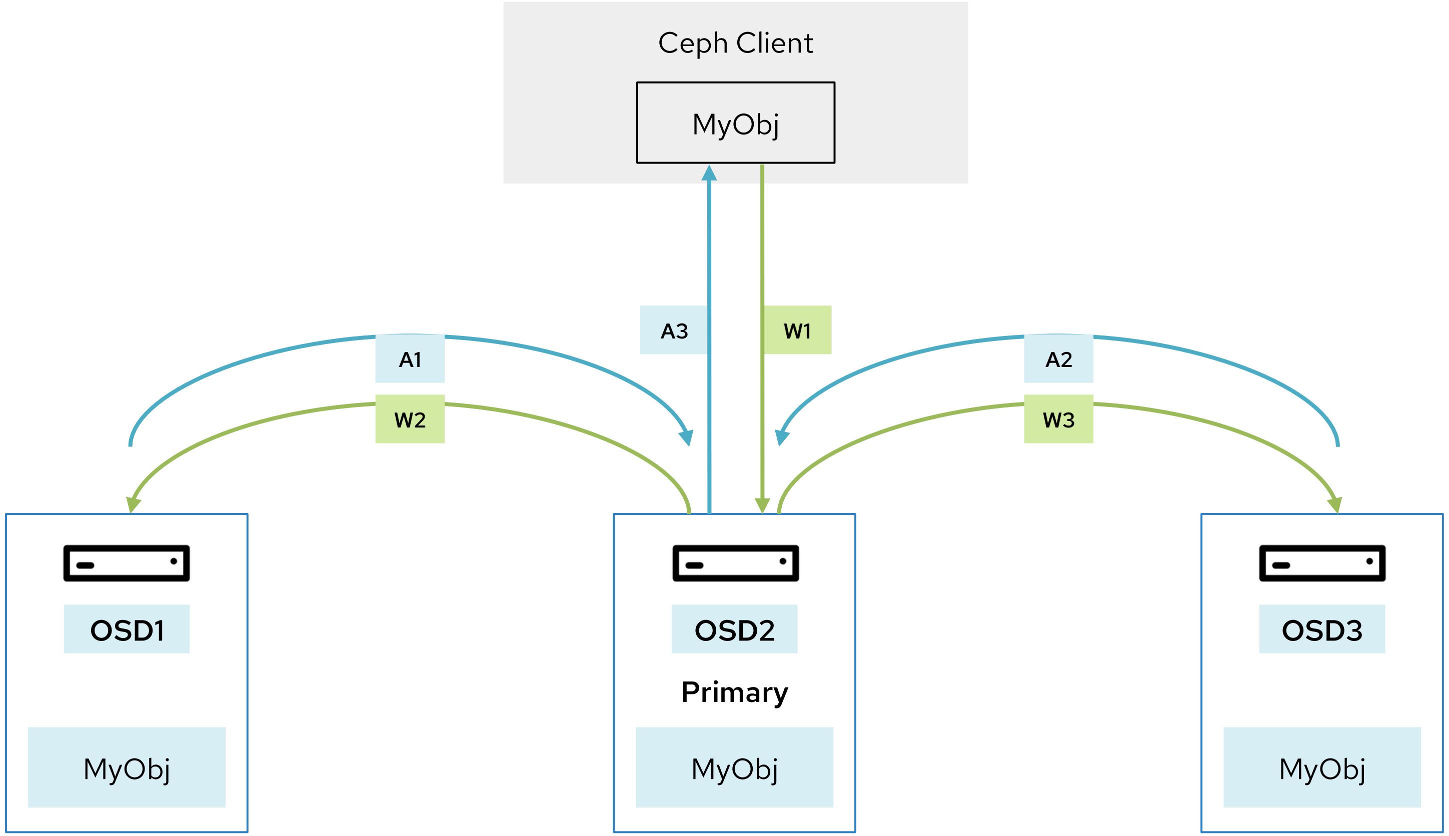
LIBRADOS
A library allowing apps to directly access RADOS (C, C++, Java, Python, Ruby, PHP)

RADOS
A software-based, reliable, autonomous, distributed object store comprised of self-healing, self-managing, intelligent storage nodes and lightweight monitors

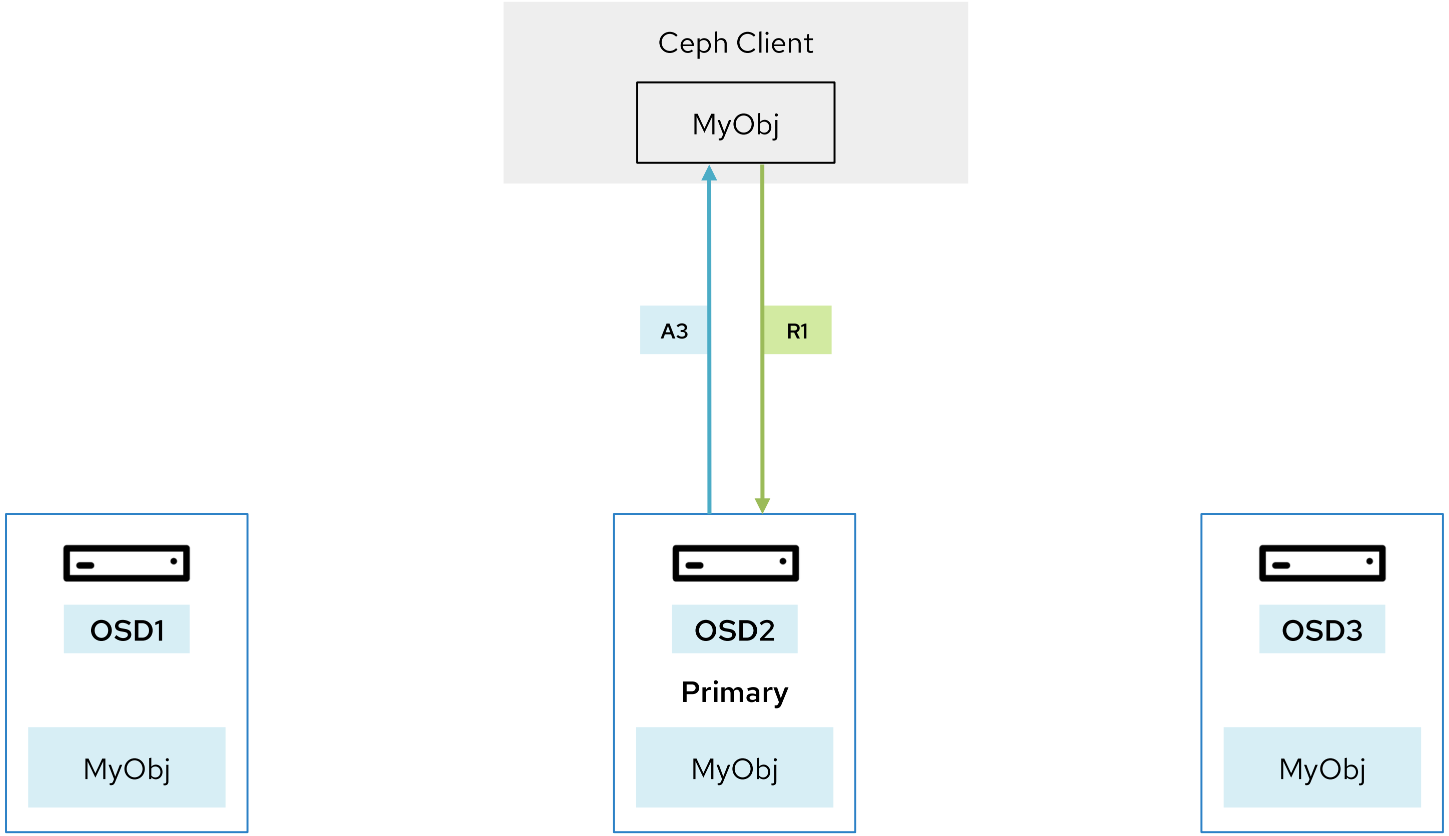
DETAILED ARCHITECTURE



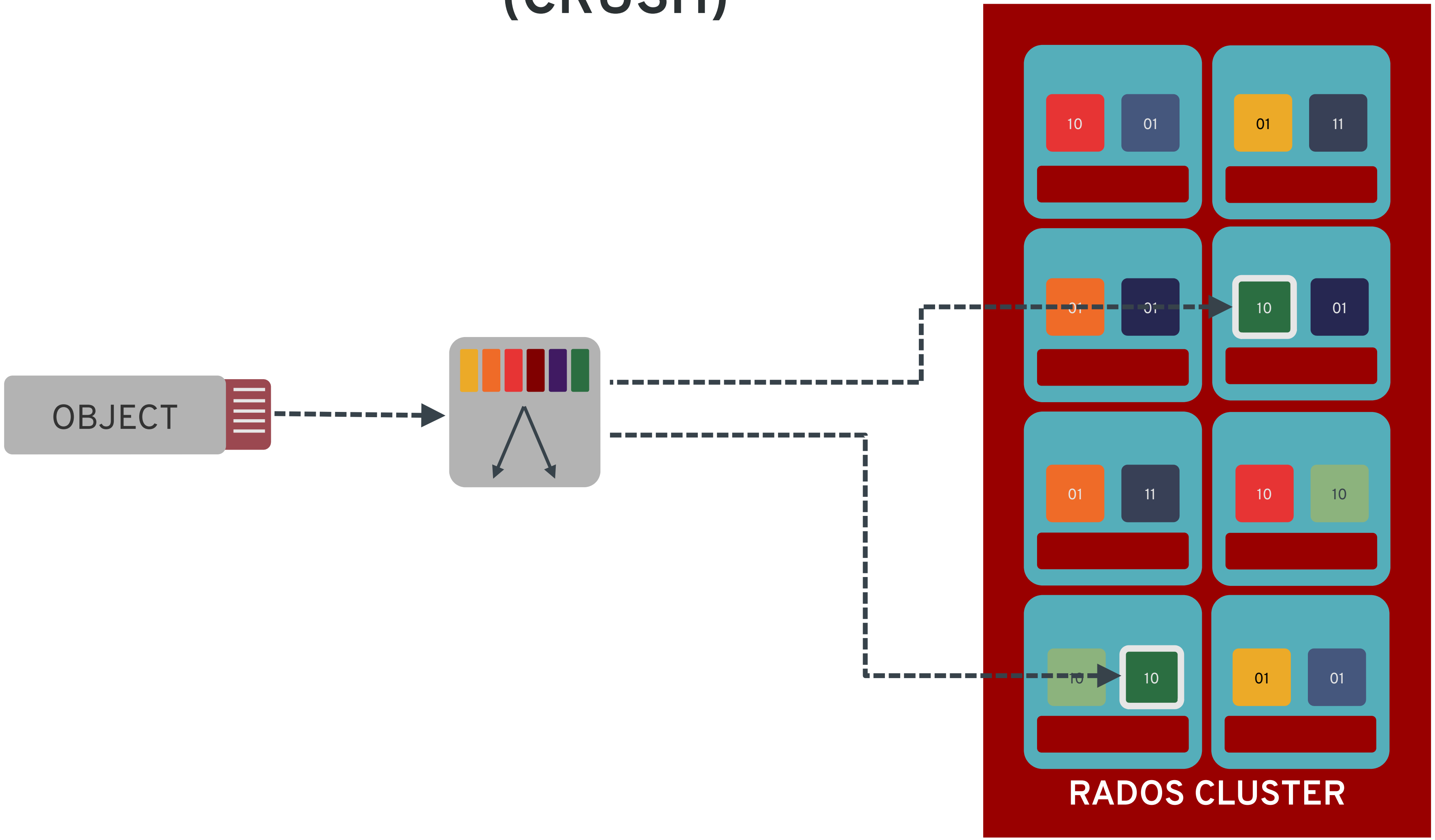
WRITES


Ceph Client
MyObj
A3
W1
A1
W2
A2
W3
OSD1
MyObj
OSD2
Primary
MyObj
OSD3
MyObj

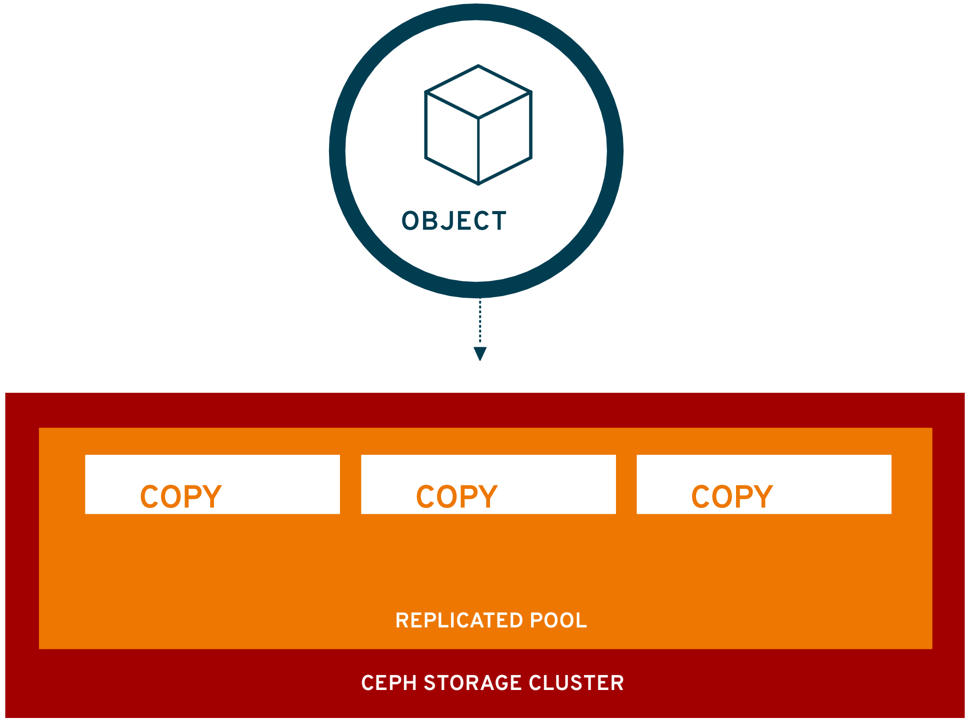
READS

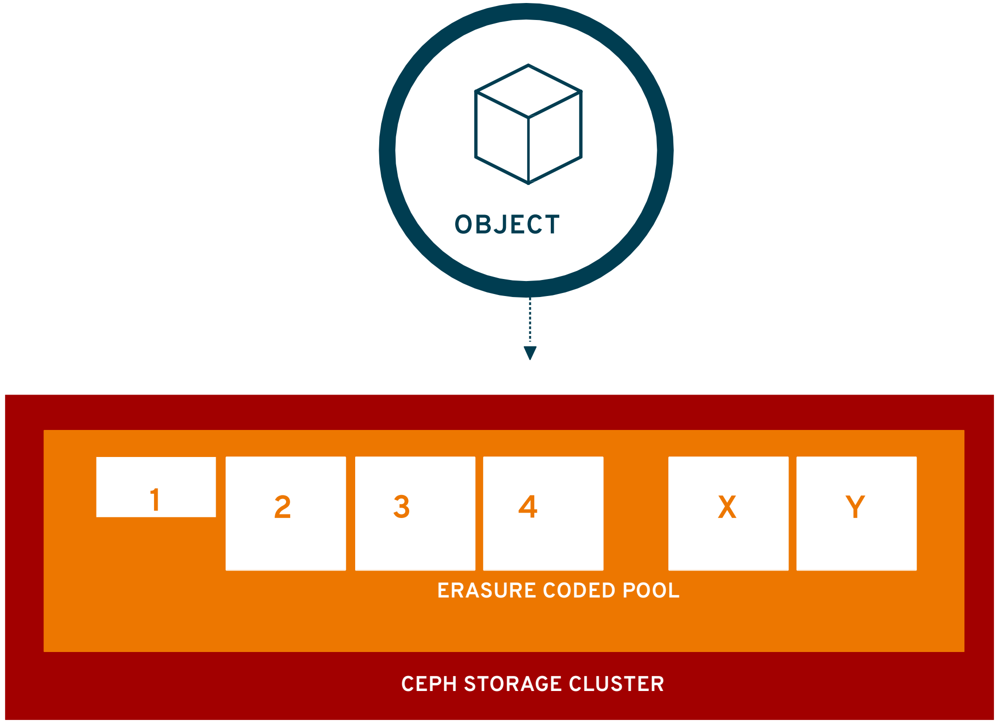

Ceph Client
MyObj
A3
R1
OSD1
MyObj
OSD2
Primary
MyObj
OSD3
MyObj

CEPH DATA PLACEMENT (CRUSH)



CEPH REPLICATION AND ERASURE CODING





FULL COPIES OF STORED OBJECTS

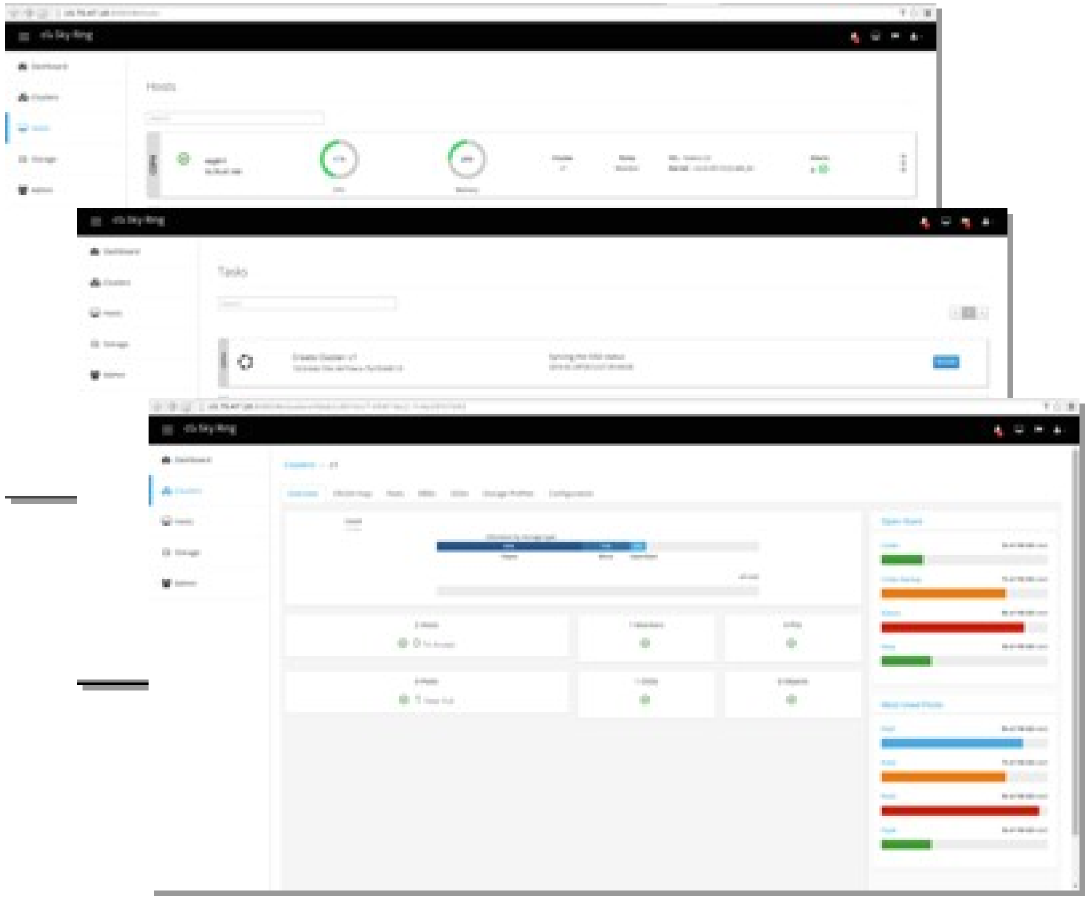
- Very high durability
- Quicker recovery

ONE COPY PLUS PARITY

- Cost-effective durability
- Expensive recovery

STORAGE CONSOLE

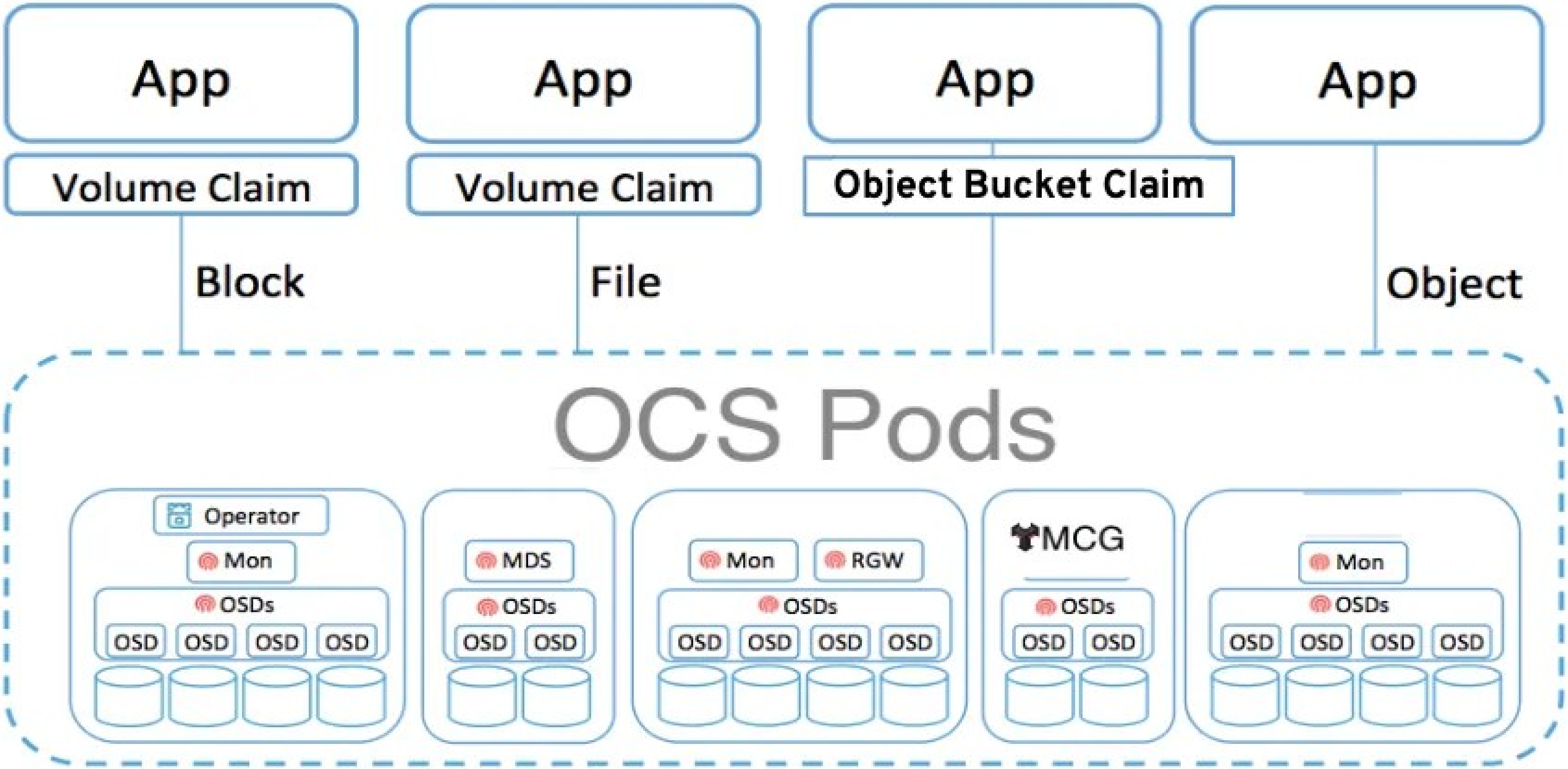
- An easy to use interface for managing cluster lifecycles
- Ansible-based deployment tools for driving granular configuration options from CLI or GUI
- Monitoring and graphs for troubleshooting with statistical information about components



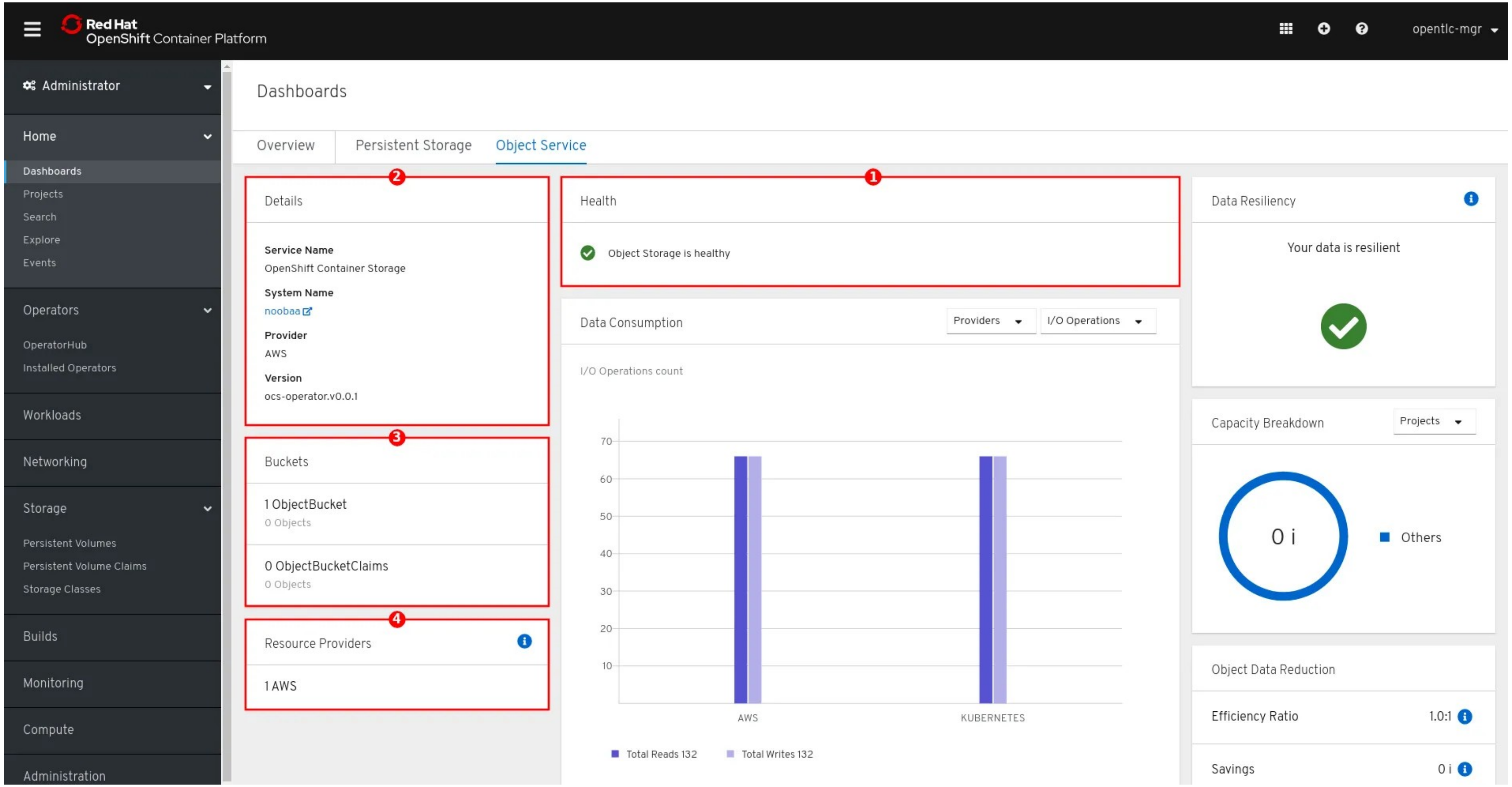
Red Hat OpenShift Container Storage

Red Hat
OpenShift
Container Storage

HYPERCONVERGED STORAGE



OPENSHIFT INTEGRATION



Dashboards

Details

Service Name
OpenShift Container Storage

System Name
noobaa

Provider
AWS

Version
ocs-operator.v0.0.1

Buckets

1 ObjectBucket
0 Objects

0 ObjectBucketClaims
0 Objects

Resource Providers

1 AWS

Health

✓ Object Storage is healthy

Data Consumption

Data Resiliency

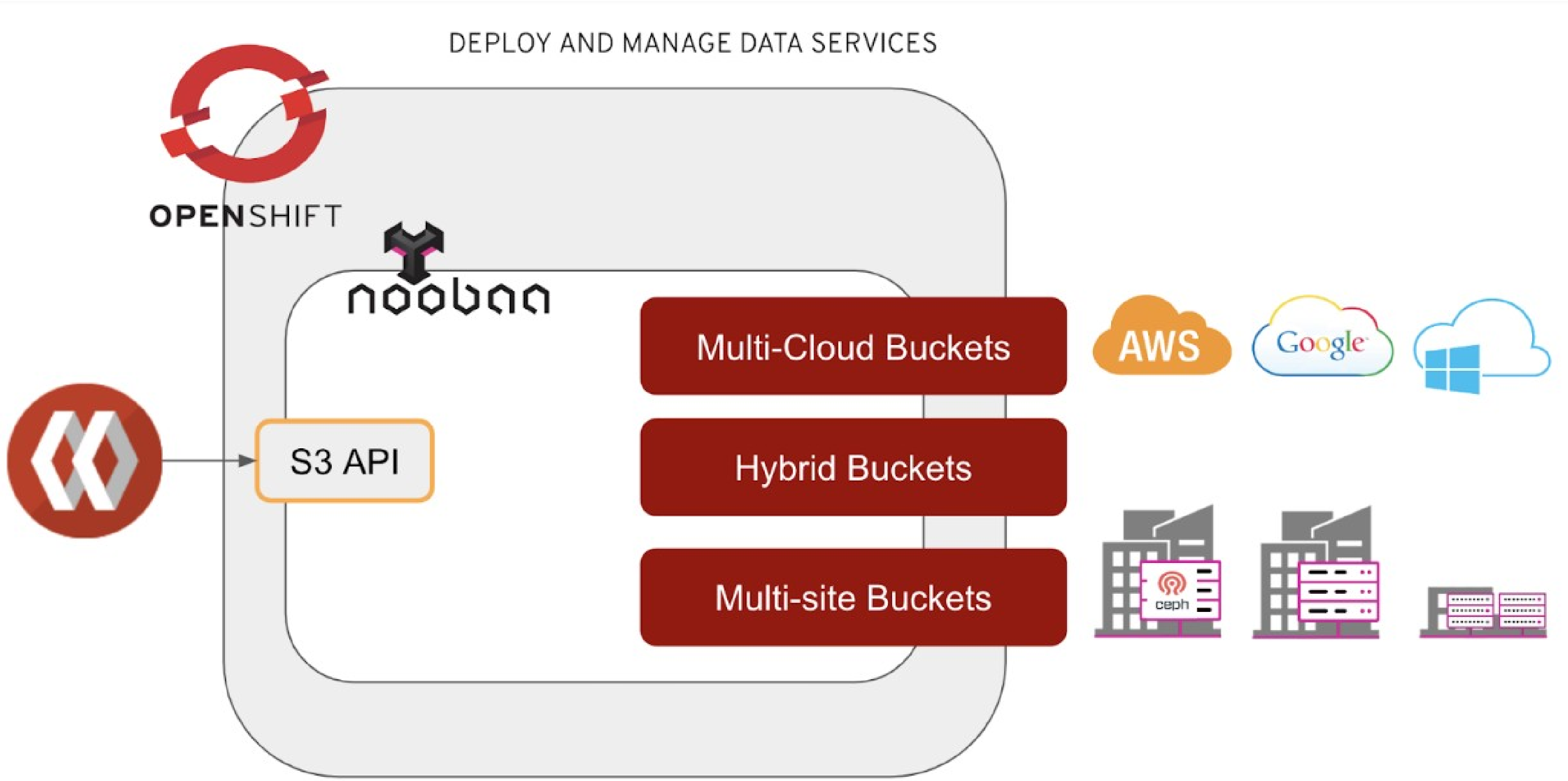
Your data is resilient

Capacity Breakdown

Object Data Reduction

Efficiency Ratio	1.0:1
Savings	0 i

MULTI-CLOUD WITH NOOBAA



Thank you

Red Hat is the world's leading provider of enterprise open source software solutions. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500.

in linkedin.com/company/red-hat

▶ youtube.com/user/RedHatVideos

f facebook.com/redhatinc

twitter.com/RedHat

Red Hat