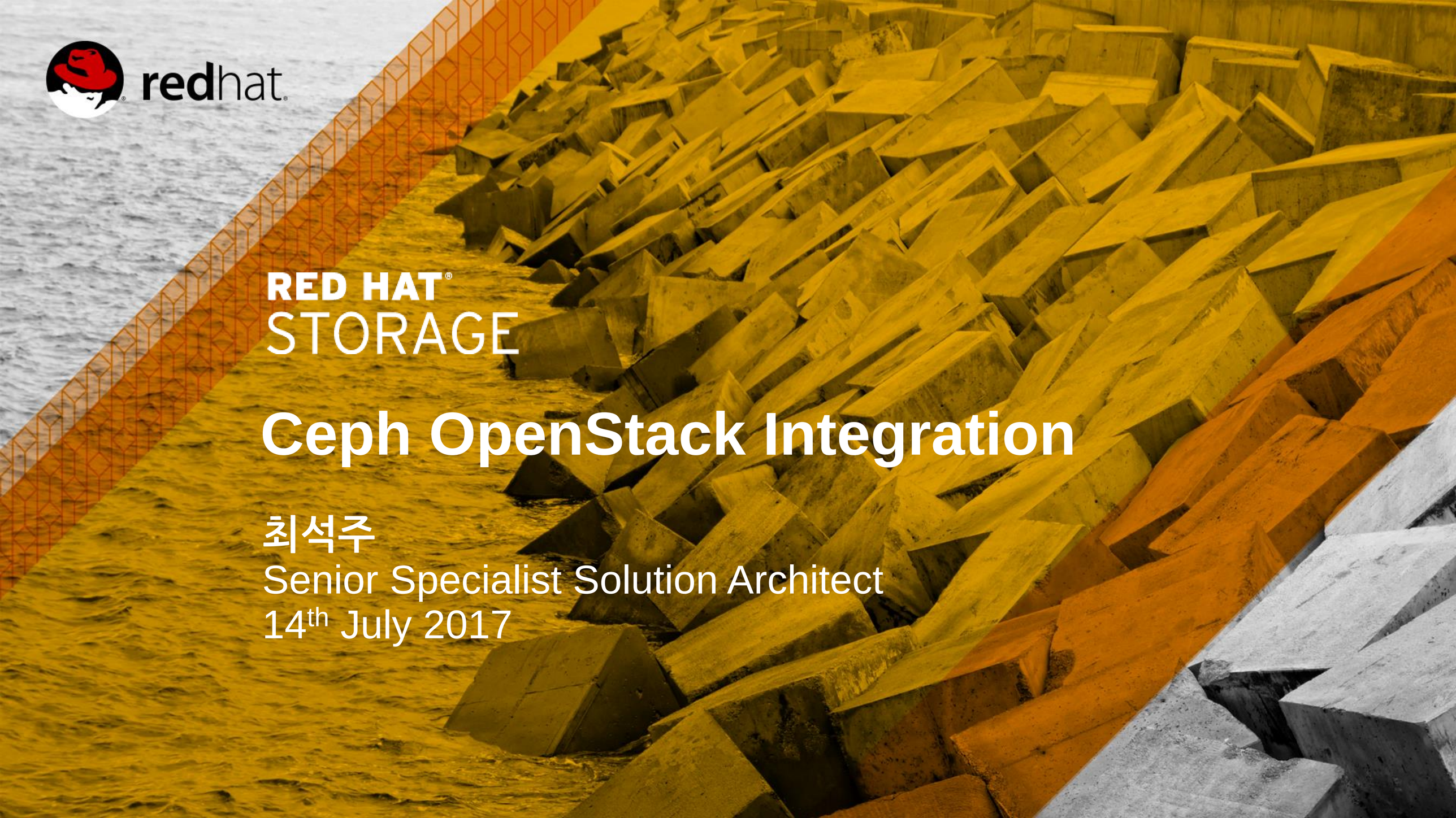


RED HAT®
STORAGE

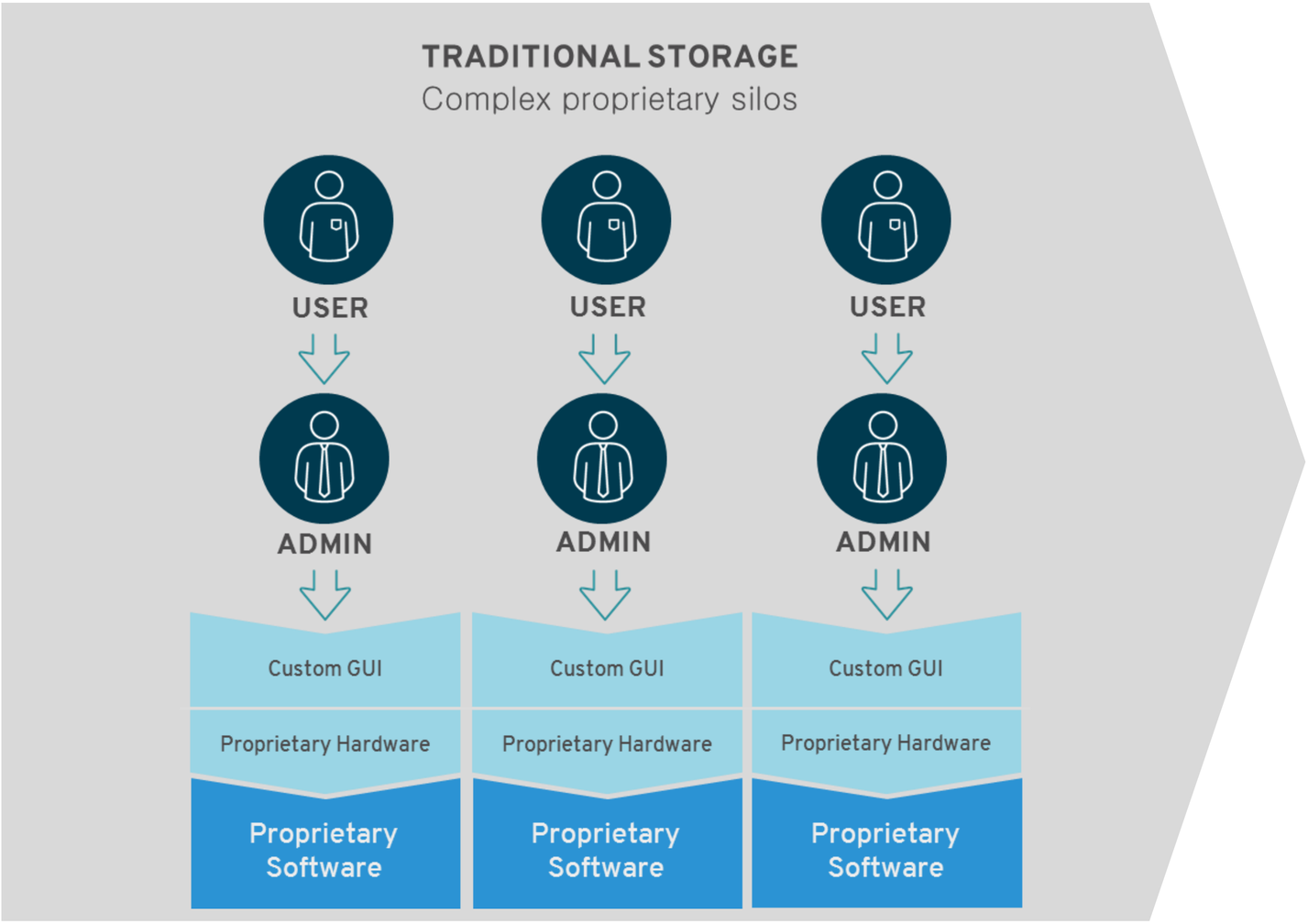
Ceph OpenStack Integration

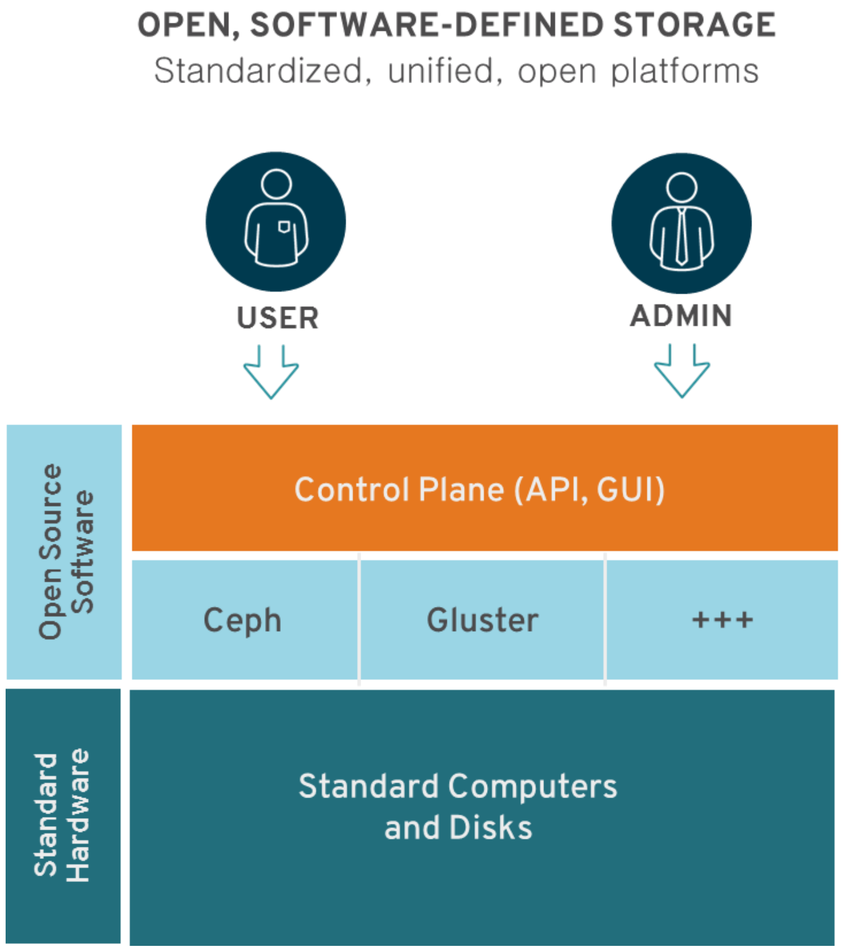
최석주
Senior Specialist Solution Architect
14th July 2017

목차

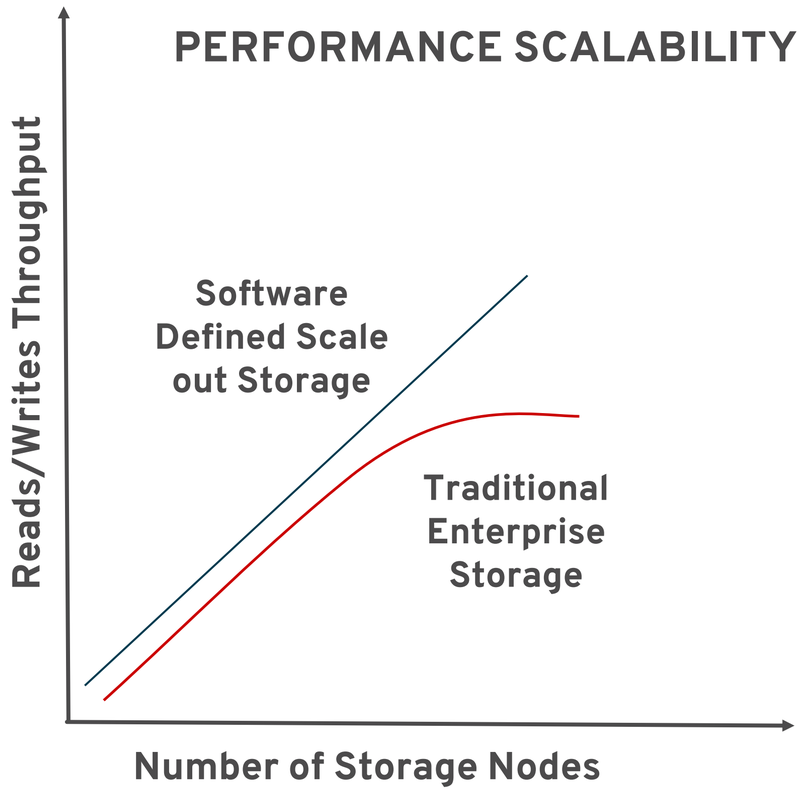
redhat.

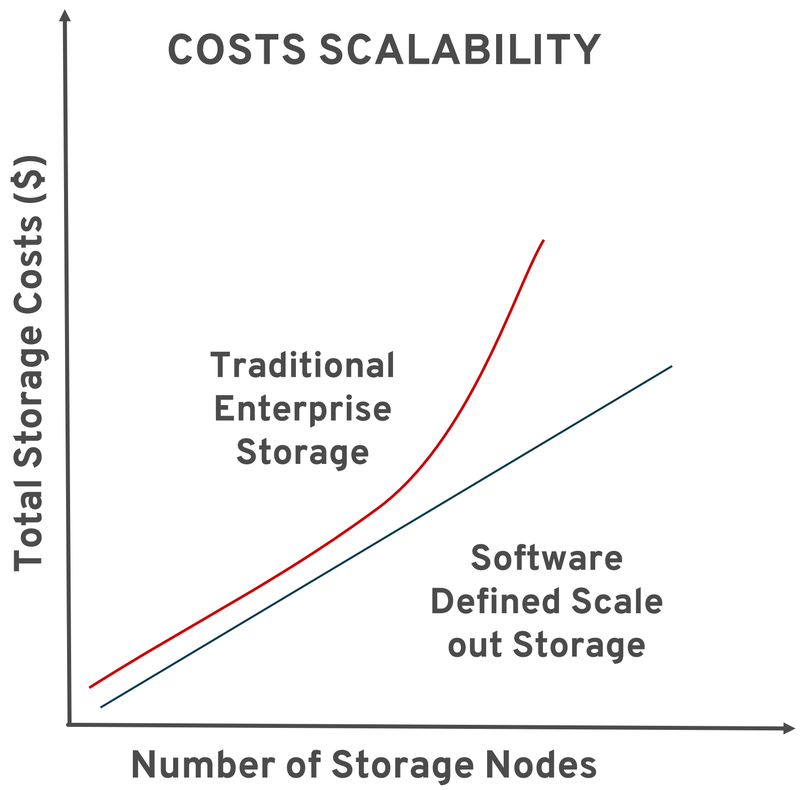
STORAGE IS EVOLVING





Significant Advantage Over Traditional Storage





WHY THIS MATTERS

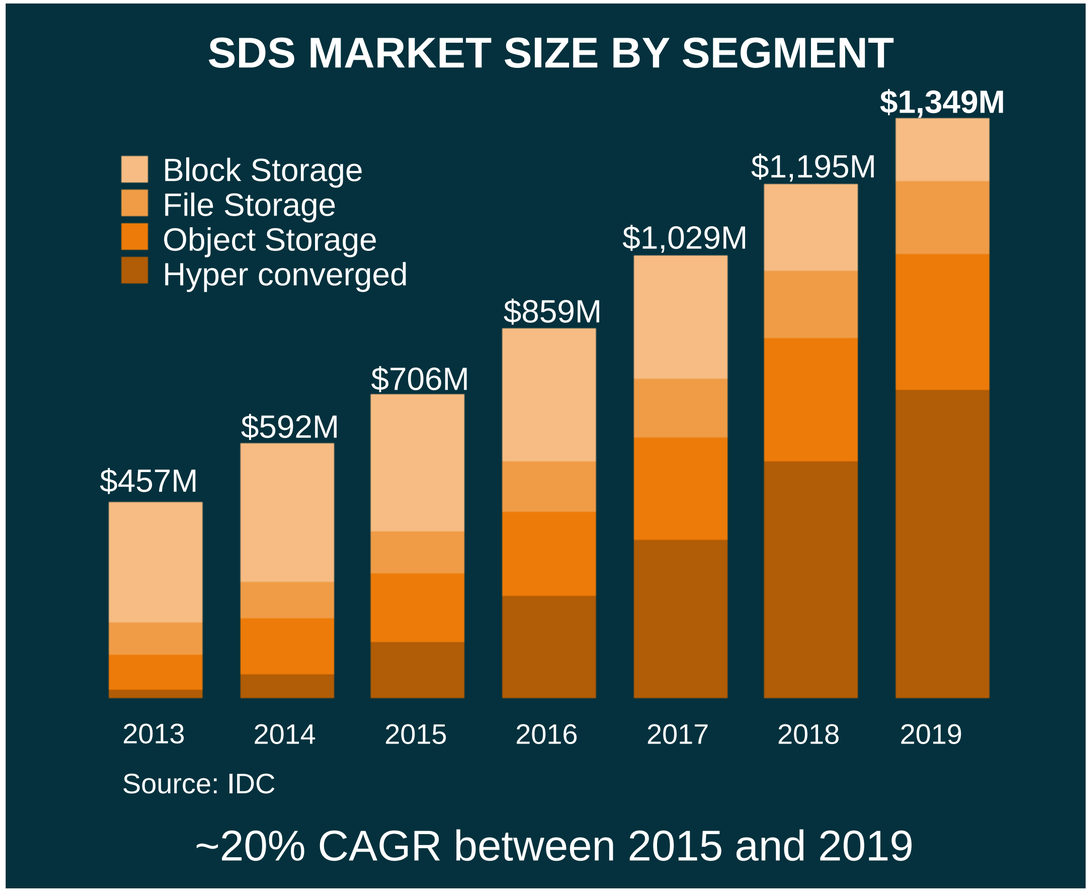
PROPRIETARY HARDWARE	Common, off-the-shelf hardware	Lower cost, standardized supply chain
SCALE-UP ARCHITECTURE	Scale-out architecture	Increased operational flexibility
HARDWARE-BASED INTELLIGENCE	Software-based intelligence	More programmability, agility, and control
CLOSED DEVELOPMENT PROCESS	Open development process	More flexible, well-integrated technology

Rising tide of software-defined storage

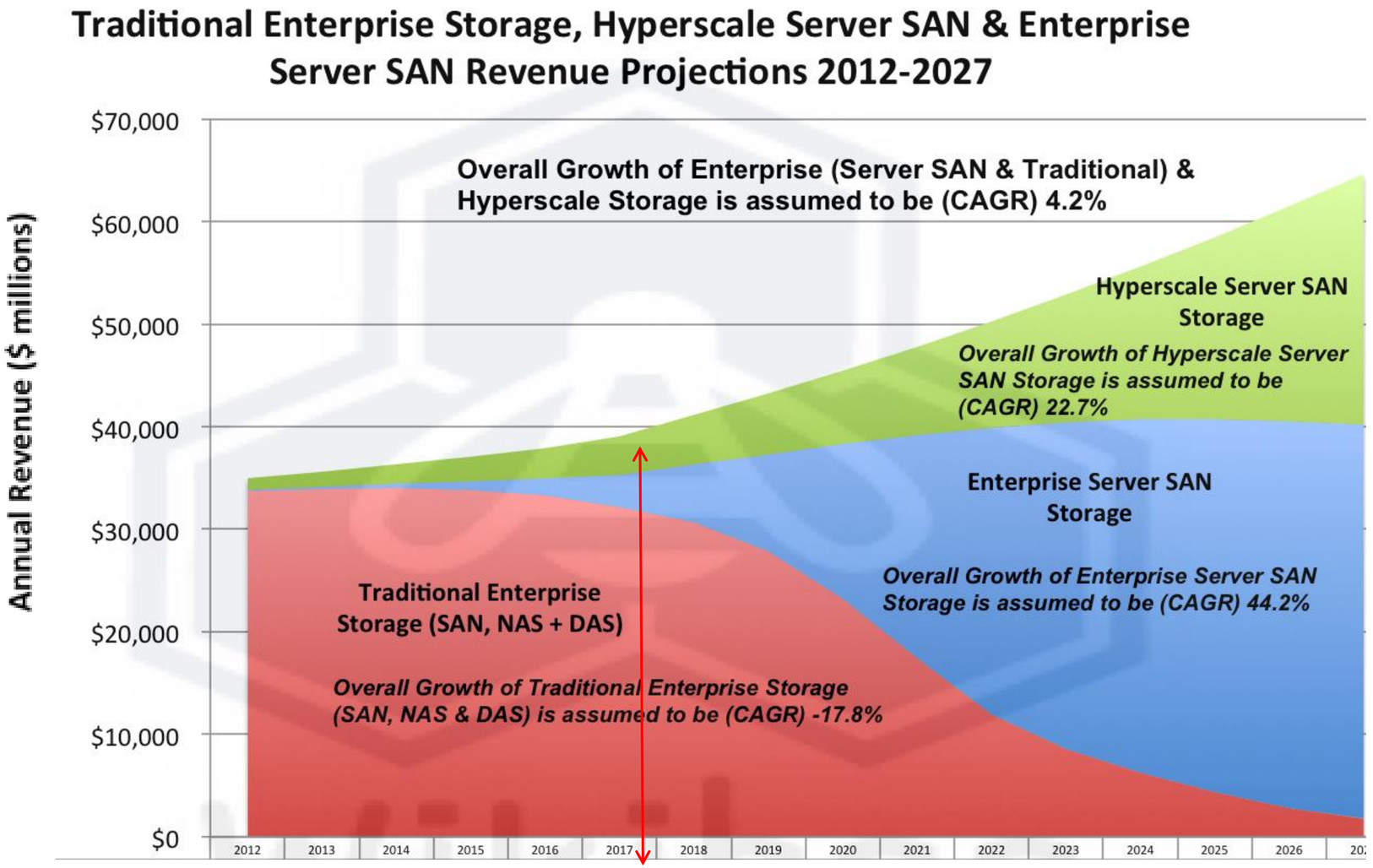
"By 2016, server-based storage solutions will lower storage hardware costs by 50% or more."
Gartner: "IT Leaders Can Benefit From Disruptive Innovation in the Storage Industry"

"By 2020, between 70-80% of unstructured data will be held on lower-cost storage managed by SDS environments."
Innovation Insight: Separating Hype From Hope for Software-Defined Storage

"By 2019, 70% of existing storage array products will also be available as software only versions"
Innovation Insight: Separating Hype From Hope for Software-Defined Storage



Storage revenue projection



wikibon.org The Rise of Server SAN, http://wikibon.org/wiki/v/The_Rise_of_Server_SAN

2012 대용량 스토리지로 SDS를 채용하기 시작함

2014 전통적인 엔터프라이즈 스토리지 매출이 정점에 다다름

2015 오픈 소스 기반의 클라우드 플랫폼과 SDS가 일반화되기 시작함

2021 SDS 기반 스토리지 매출이 전통적인 엔터프라이즈 스토리지 매출을 앞서기 시작함

2023 초대용량 SDS 매출이 전통전인 엔터프라이즈 스토리지 매출을 앞서기 시작함

2025 대부분의 전통적인 엔터프라이즈 스토리지는 레거시 시스템을 지원하기 위해 존재함

목차

redhat.

CEPH 스토리지

ceph

TARGET USE CASES

Cloud Storage
• OpenStack® VM lifecycle storage with Glance, Cinder, Keystone, Nova

• Persistent object and block storage for tenant apps

Media & Big Data Storage
• S3 and S3A compatible

Powerful distributed storage for the cloud and beyond

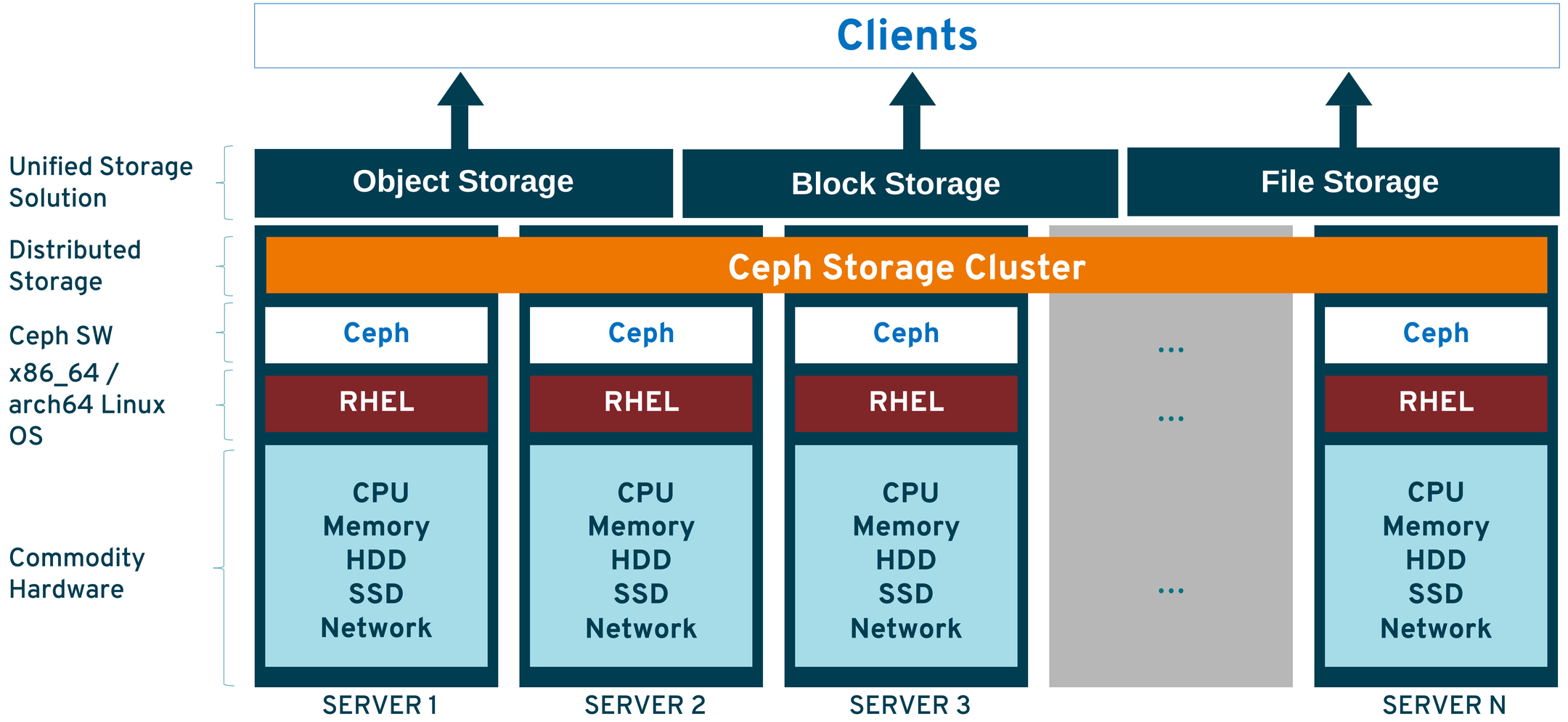
- Low-cost software-defined storage on commodity servers
- Open source
- Flexible block, object, and file storage
- Production-ready data protection and self-healing
- Massively scalable
- Leading storage for OpenStack

CUSTOMER HIGHLIGHT: CISCO

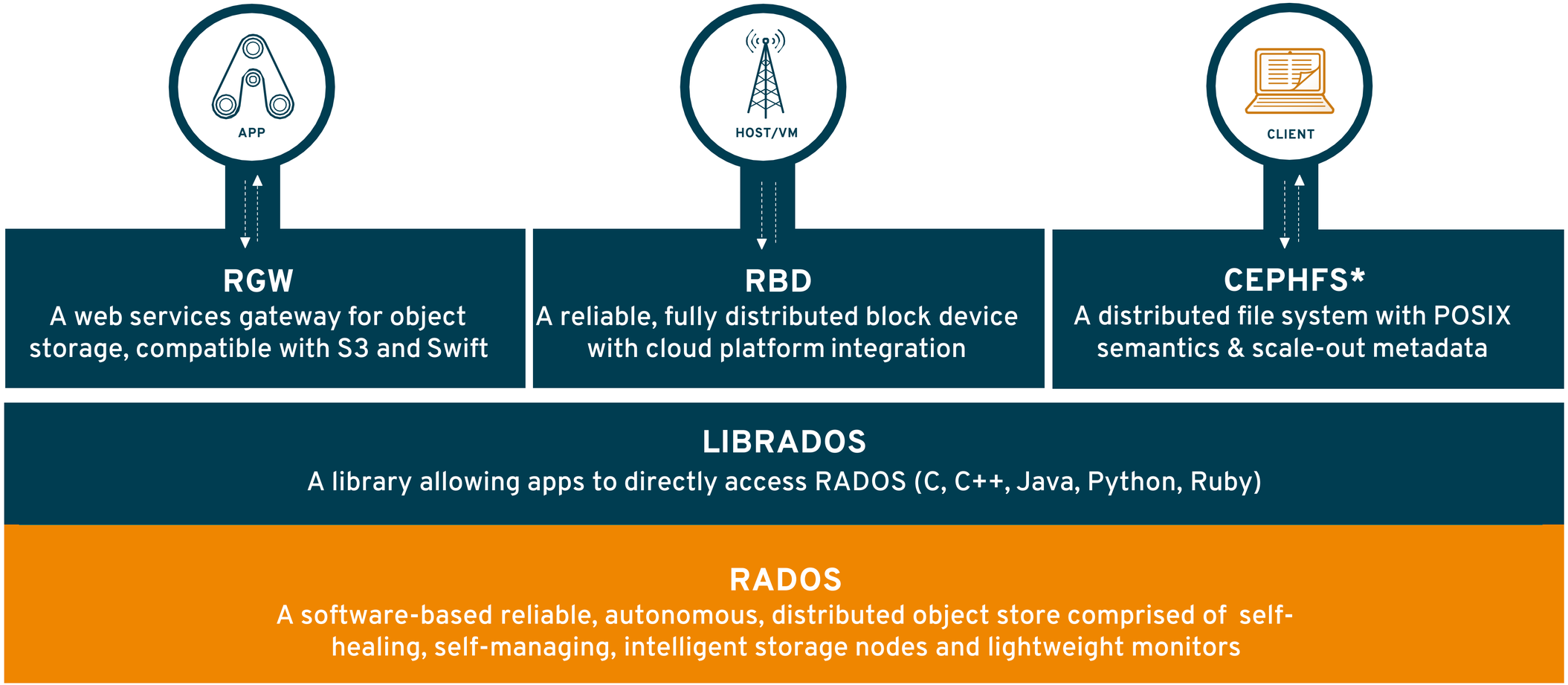
CISCO

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

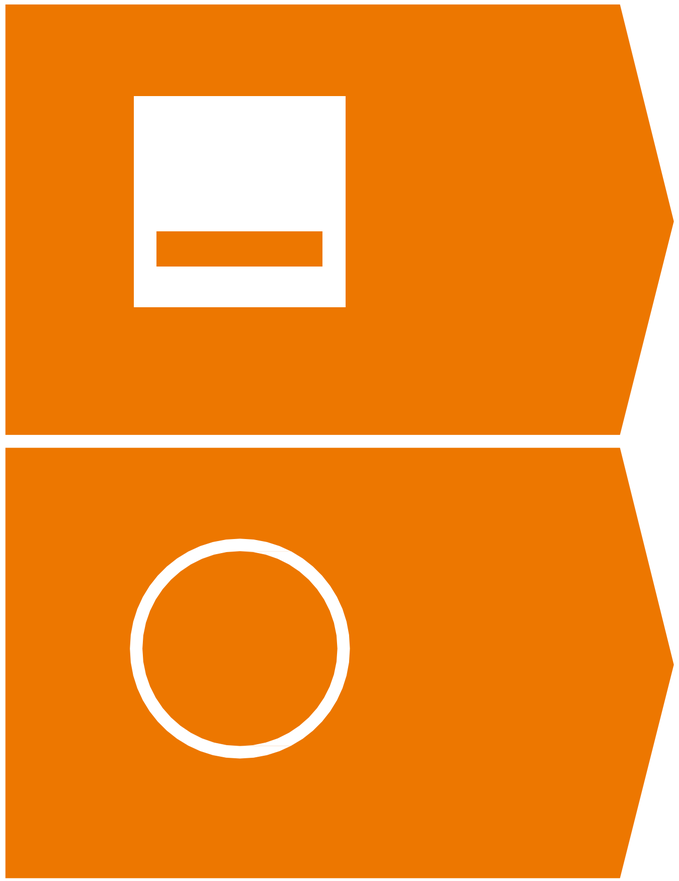
CEPH 아키텍처



CEPH 구성요소 : RADOS



CEPH 구성요소 : RADOS
RELIABLE AUTONOMOUS DISTRIBUTED OBJECT STORE



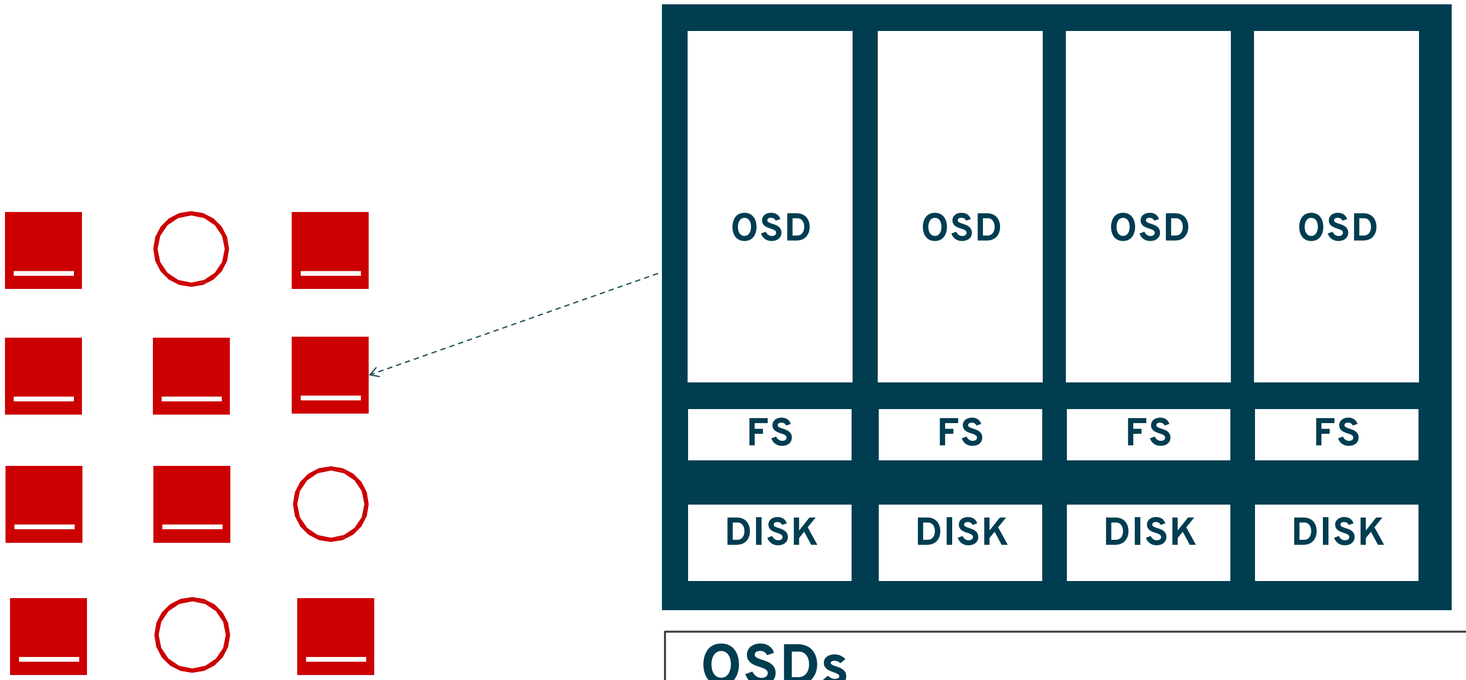
OSDs (Object Storage Daemons)

- 10s to 10000s in a cluster
- One per disk (or one per SSD, RAID group…)
- Serve stored objects to clients
- Intelligently peer for replication & recovery
- Minimum 3 per cluster

MONs (Monitors)

- Maintain cluster membership and state
- Track health of the cluster
- Provide consensus for distributed decision-making
- Small, odd number
- These do not serve stored objects to clients (not in the data path)
- Minimum 3 per cluster

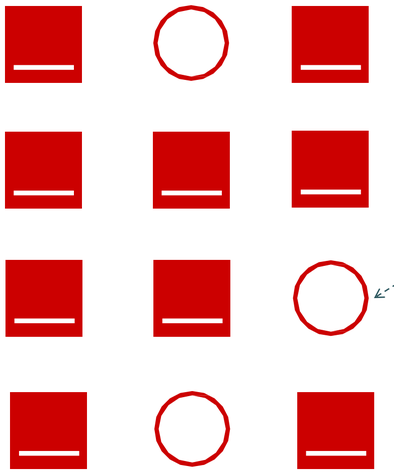
CEPH 구성요소 : RADOS
OSD - OBJECT STORAGE DAEMON



OSDs

- 10s to 10000s in a cluster
- One per disk (or one per SSD, RAID group…)
- Serve stored objects to clients
- Minimum 3 per cluster

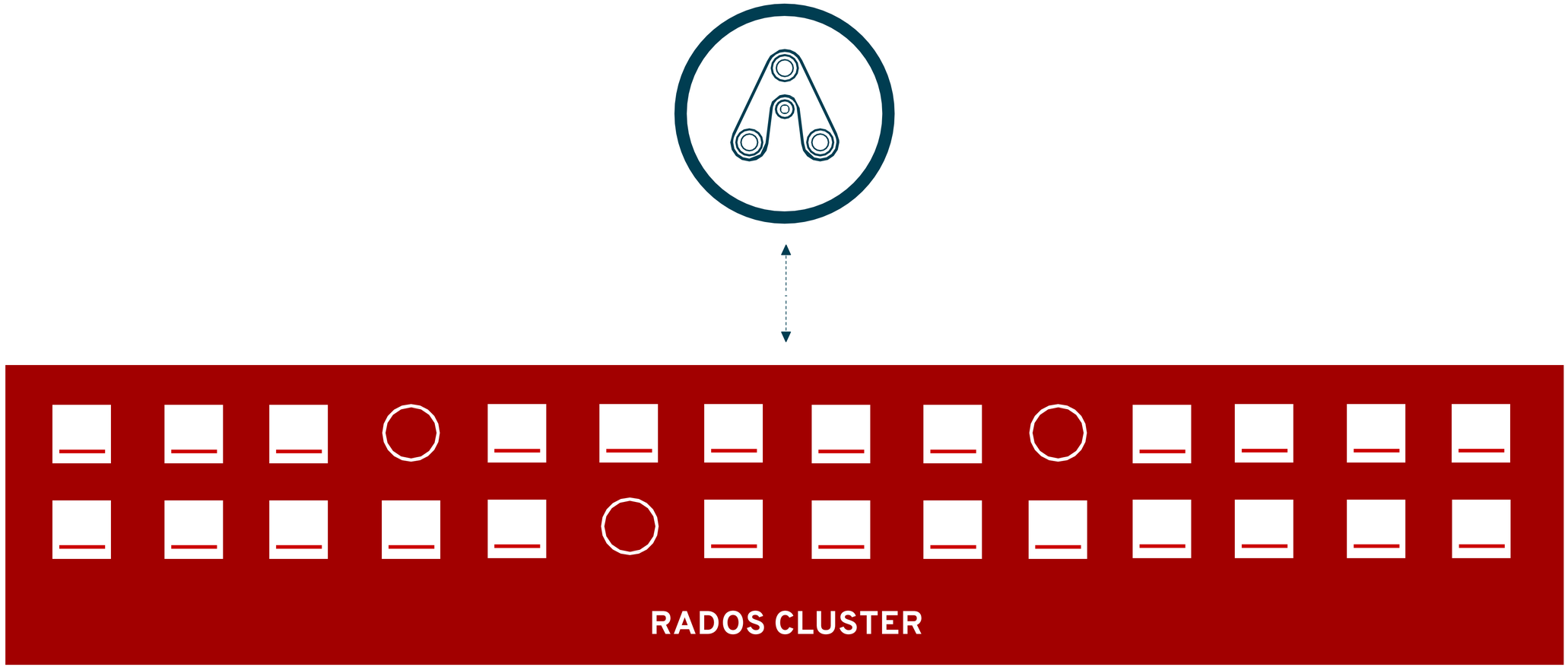
CEPH 구성요소 : RADOS
MON - MONITORS



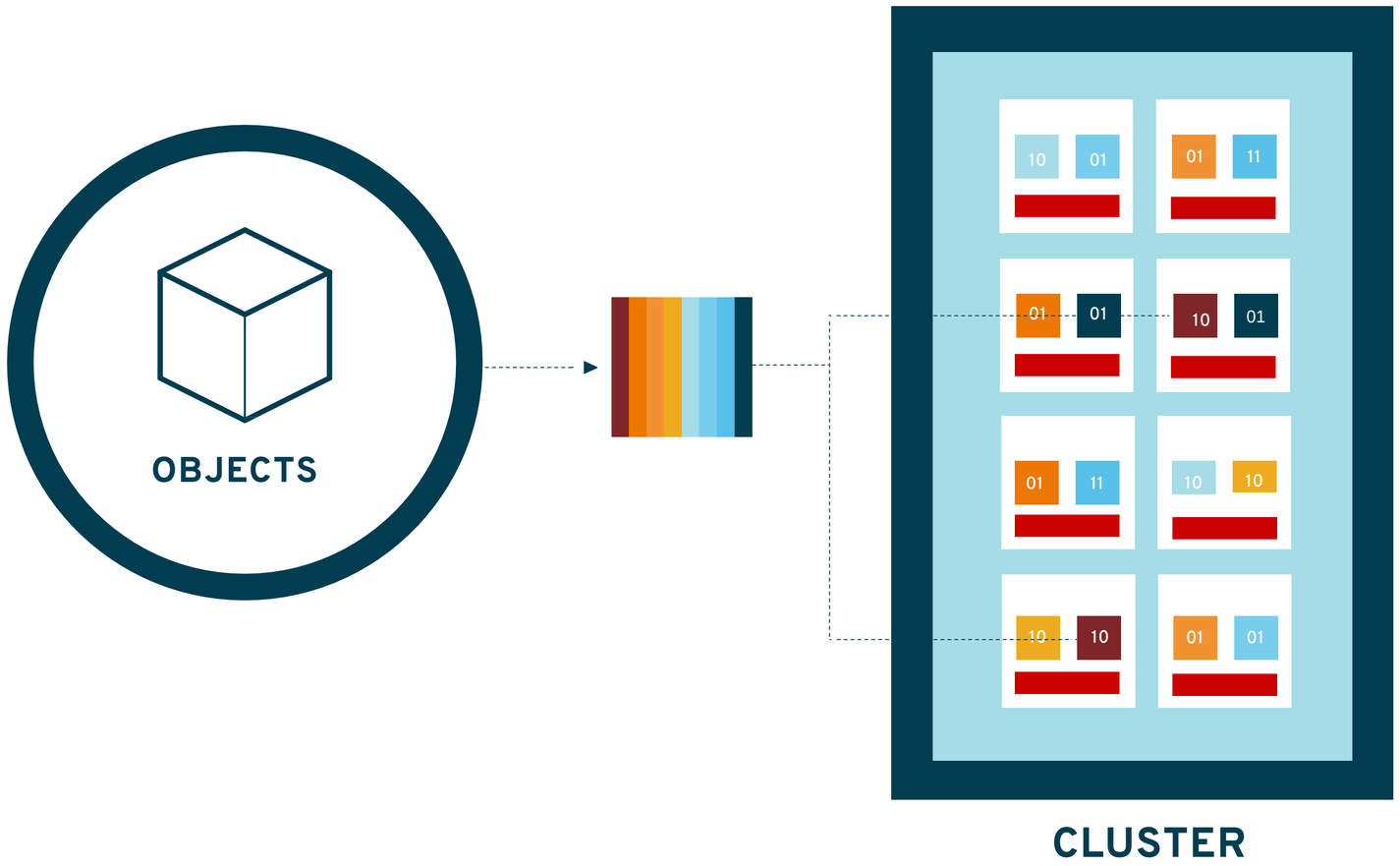
Monitors

- Maintain cluster membership and state
- Track health of the cluster
- Provide consensus for distributed decision-making
- Small, odd number
- These do not serve stored objects to clients (not in the data path)
- Minimum 3 per cluster

CEPH 구성요소 : RADOS
RADOS CLUSTER



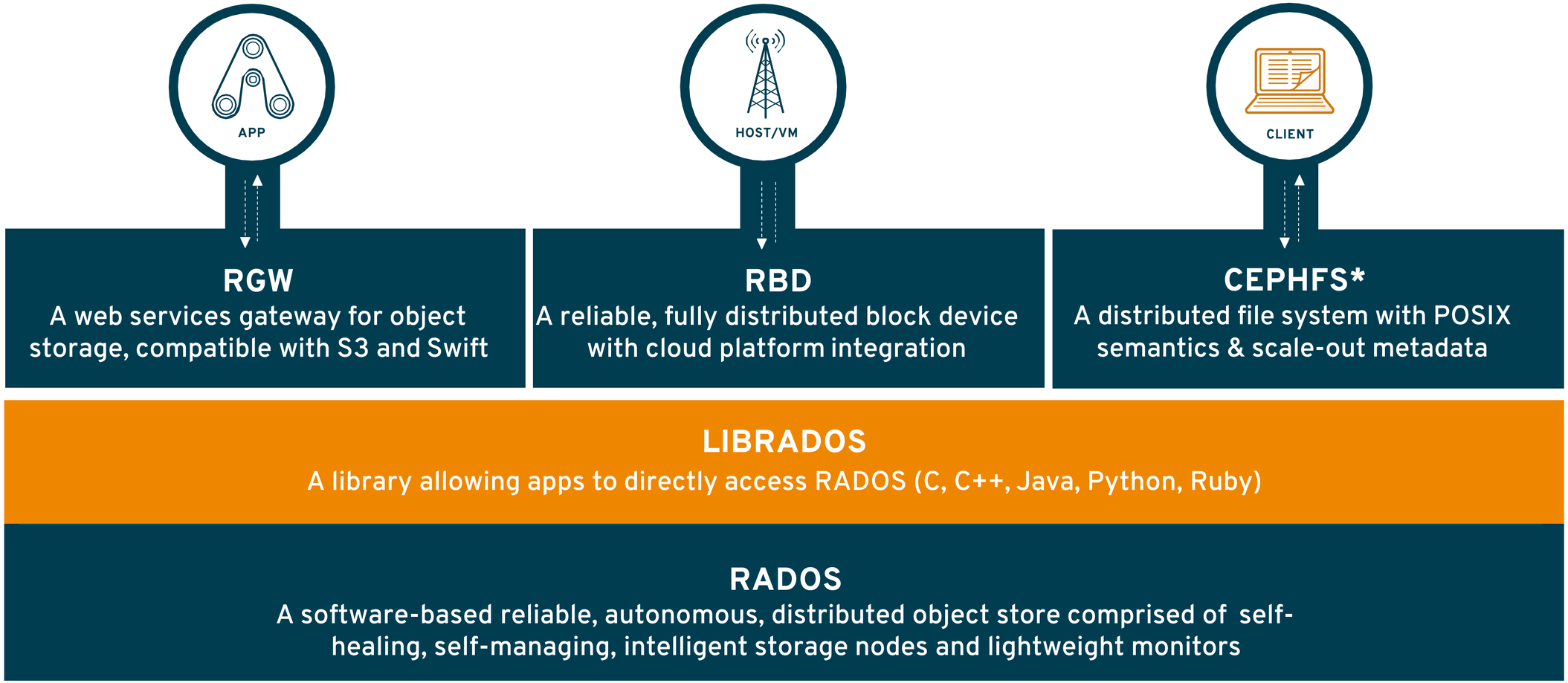
CEPH 구성요소 : RADOS
CRUSH - DYNAMIC DATA PLACEMENT / QUICK CALCULATION



CRUSH

- Pseudo-random placement algorithm
  - Fast calculation, no lookup
  - Repeatable, deterministic
- Statistically uniform distribution
- Stable mapping
  - Limited data migration on change
- Rule-based configuration
  - Infrastructure topology aware
  - Adjustable replication
  - Weighting

CEPH 구성요소 : LIBRADOS

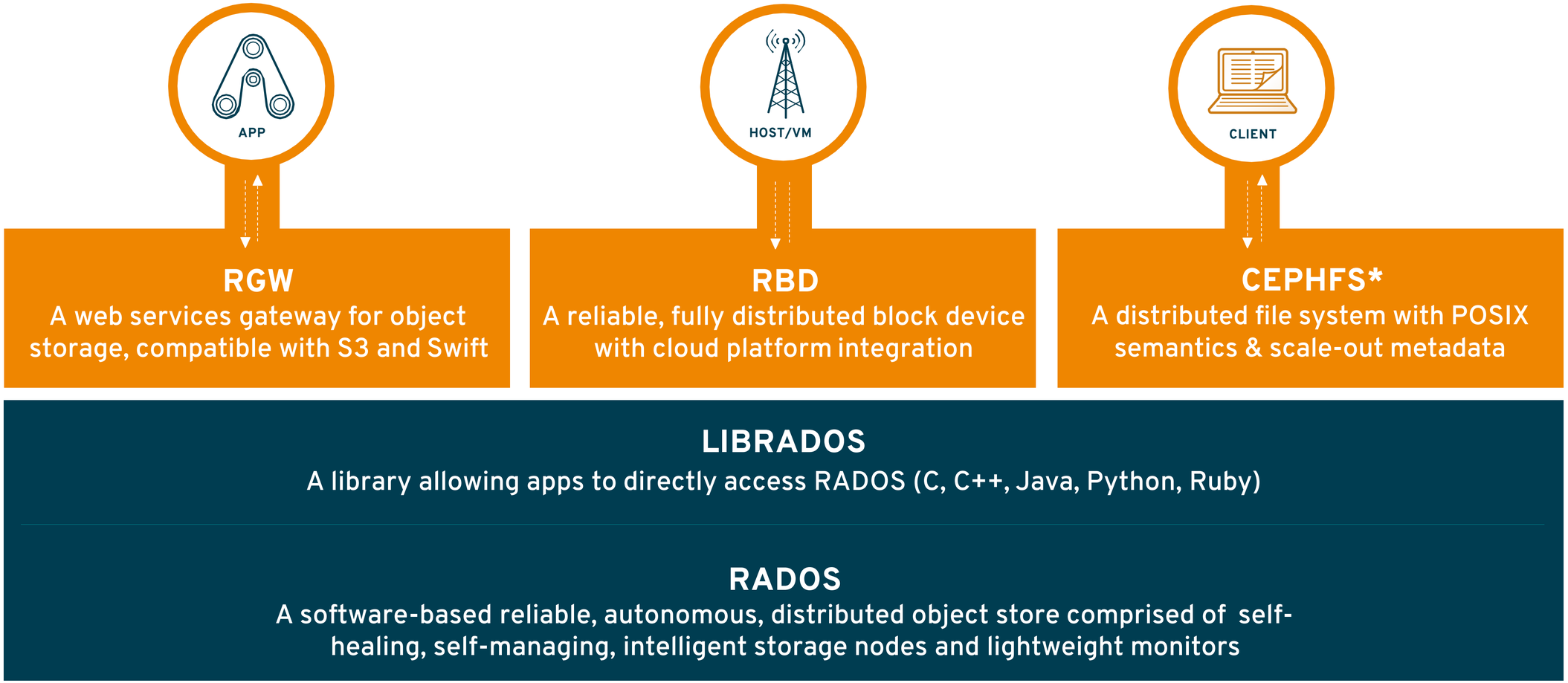


CEPH 구성요소 : LIBRADOS

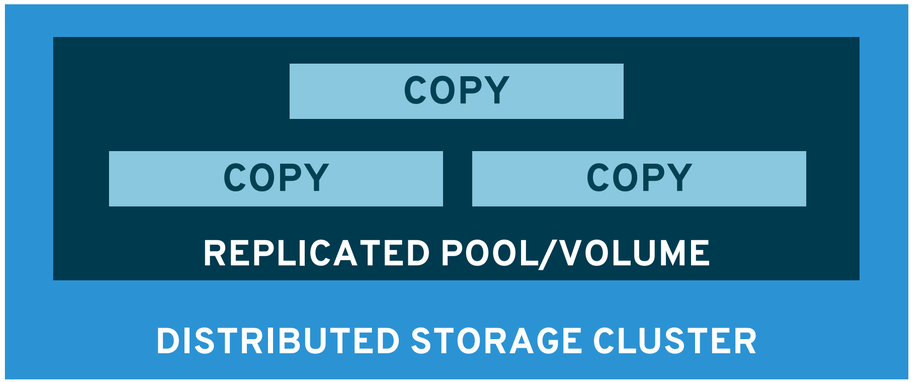
LIBRADOS

- Direct access to RADOS for applications(C, C++, Python, PHP, Java, Erlang)
- Direct access to storage nodes
- No HTTP overhead – fast, socket-based connection

CEPH 구성요소 : INTERFACE



CEPH 데이터 보호

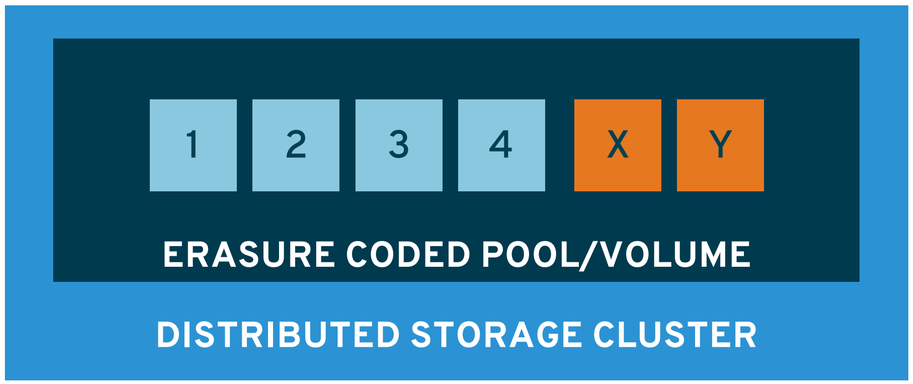


Performance-optimized

Create copies of each object or file stored within them.

Deliver optimal performance, with multiple copies available for reading. However, these back-ends can be expensive.

Fast performance and recovery.



Capacity-optimized

Corrupted or lost data is mathematically reconstructed using fragments stored elsewhere in the system.

This consumes far less space than replication, although hardware failures can affect performance.

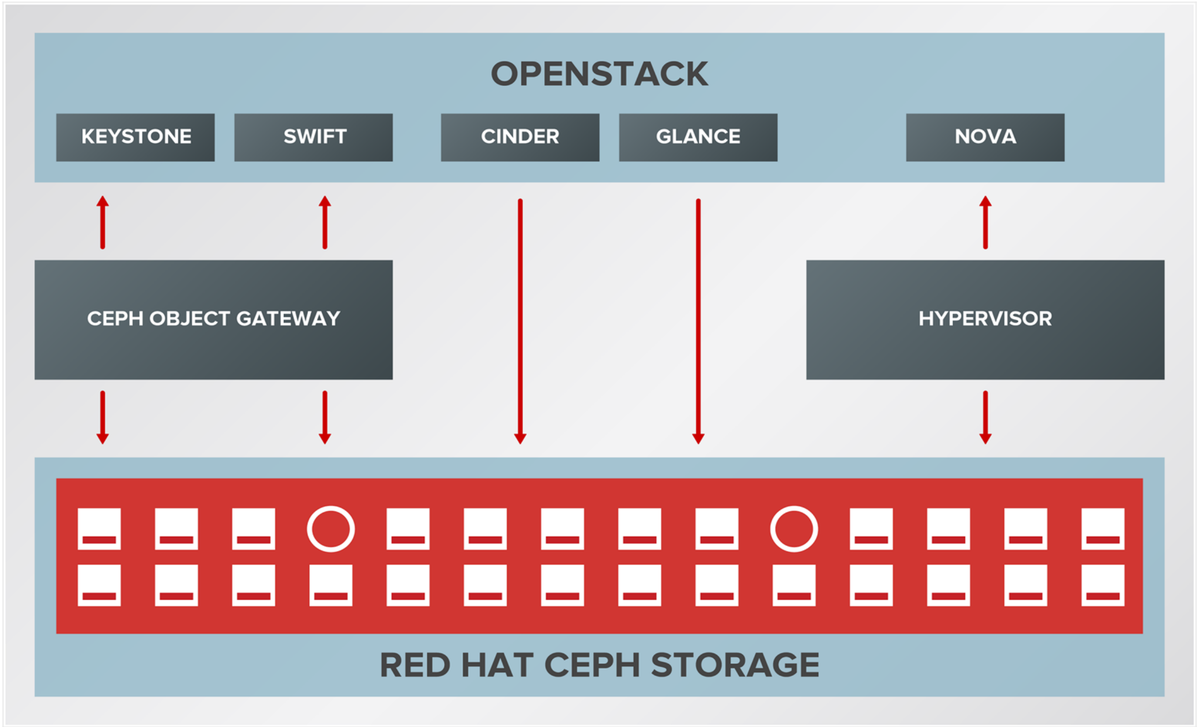
목차

redhat.

CEPH storage Use Cases

- OpenStack
- Web App Storage (STaaS/Cloud Storage)
- Media Repository
- Big Data : Data Lake for Analytics
- Backup
- Enterprise File Sync and Share

OPENSTACK



FEATURES

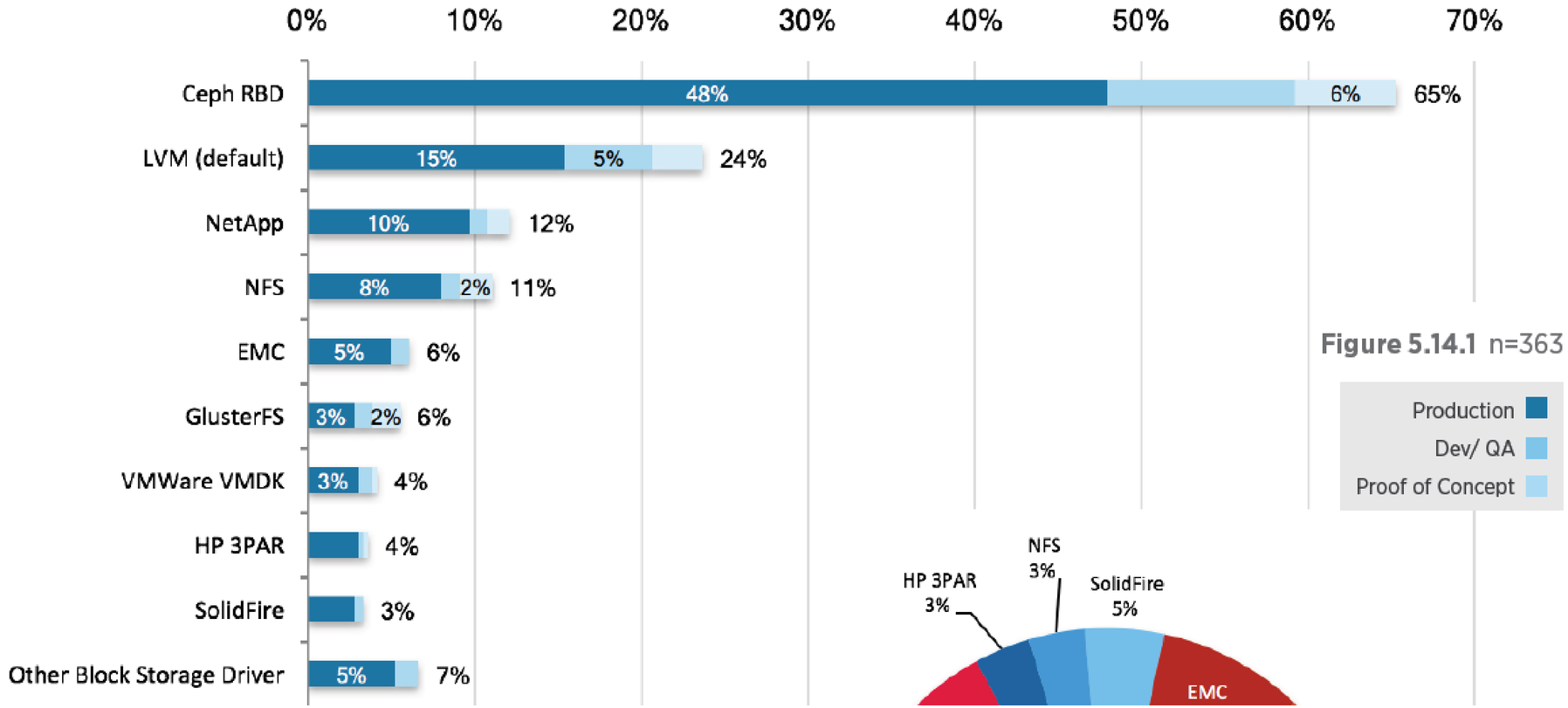
- Full integration with Nova, Cinder and Glance
- Single storage for images and ephemeral and persistent volumes
- Copy-on-write provisioning
- Swift-compatible object storage gateway
- Full integration with Red Hat Enterprise Linux OpenStack Platform

BENEFITS

- Provides both volume storage and object storage for tenant applications
- Reduces provisioning time for new virtual machines
- Requires no data transfer of images between storage and
  compute nodes
- Offers unified installation experience with Red Hat Enterprise Linux OpenStack Platform

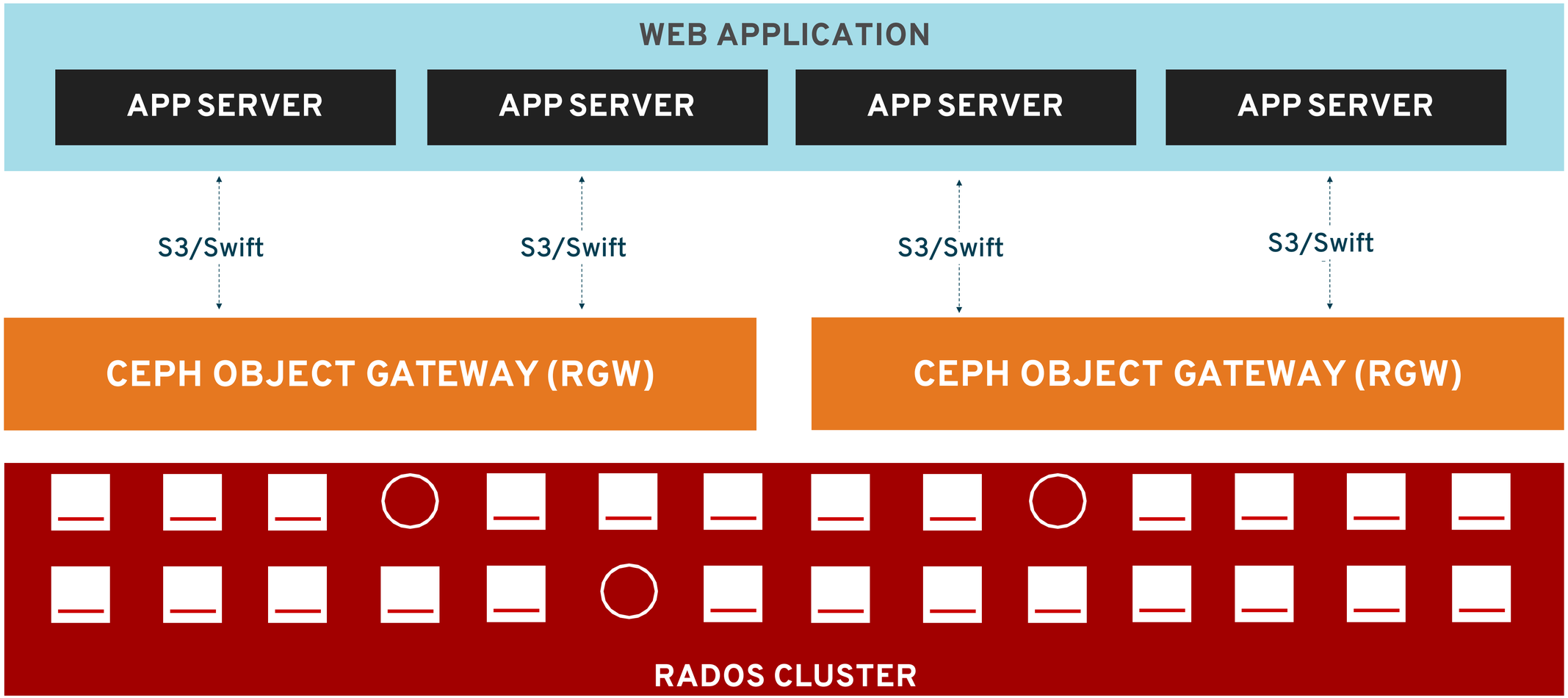
OPENSTACK Survey (2017 April)

Which OpenStack block storage (Cinder) drivers are in use?

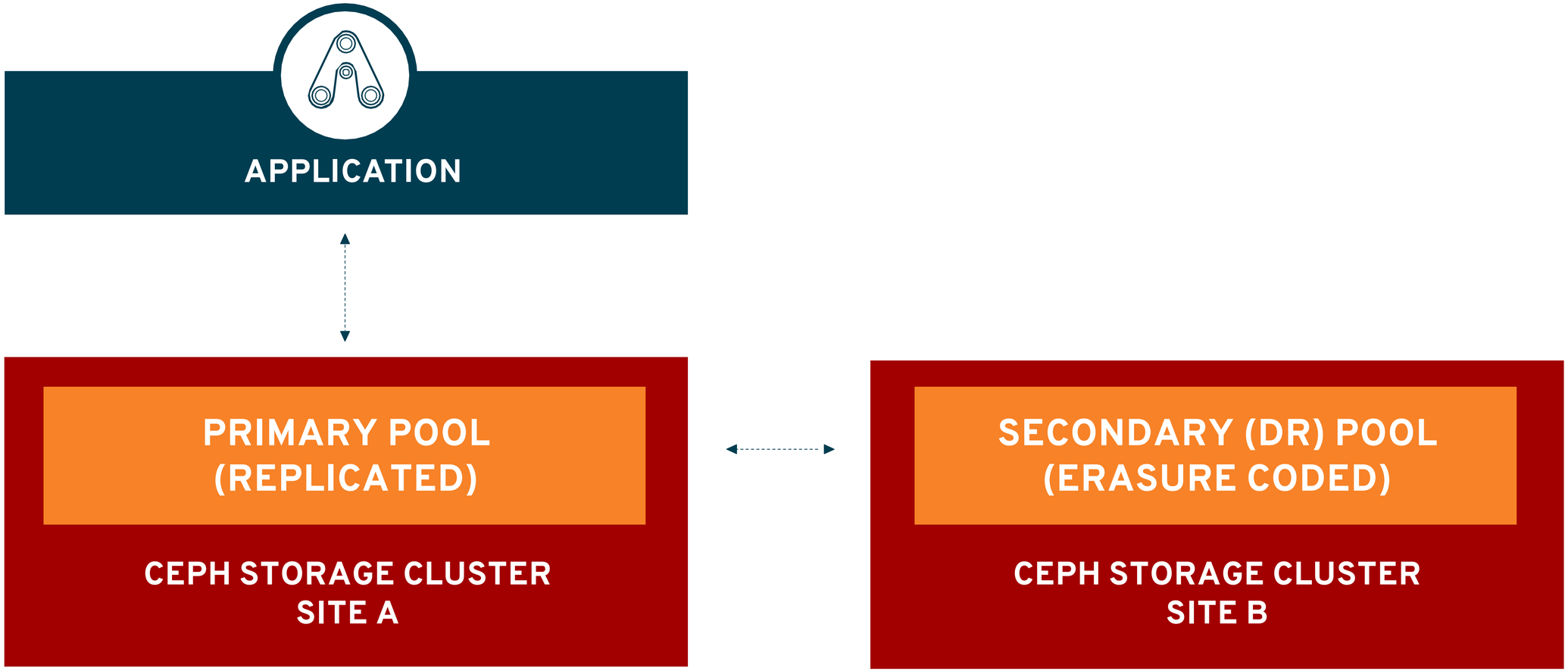


https://www.openstack.org/assets/survey/April2017SurveyReport.pdf

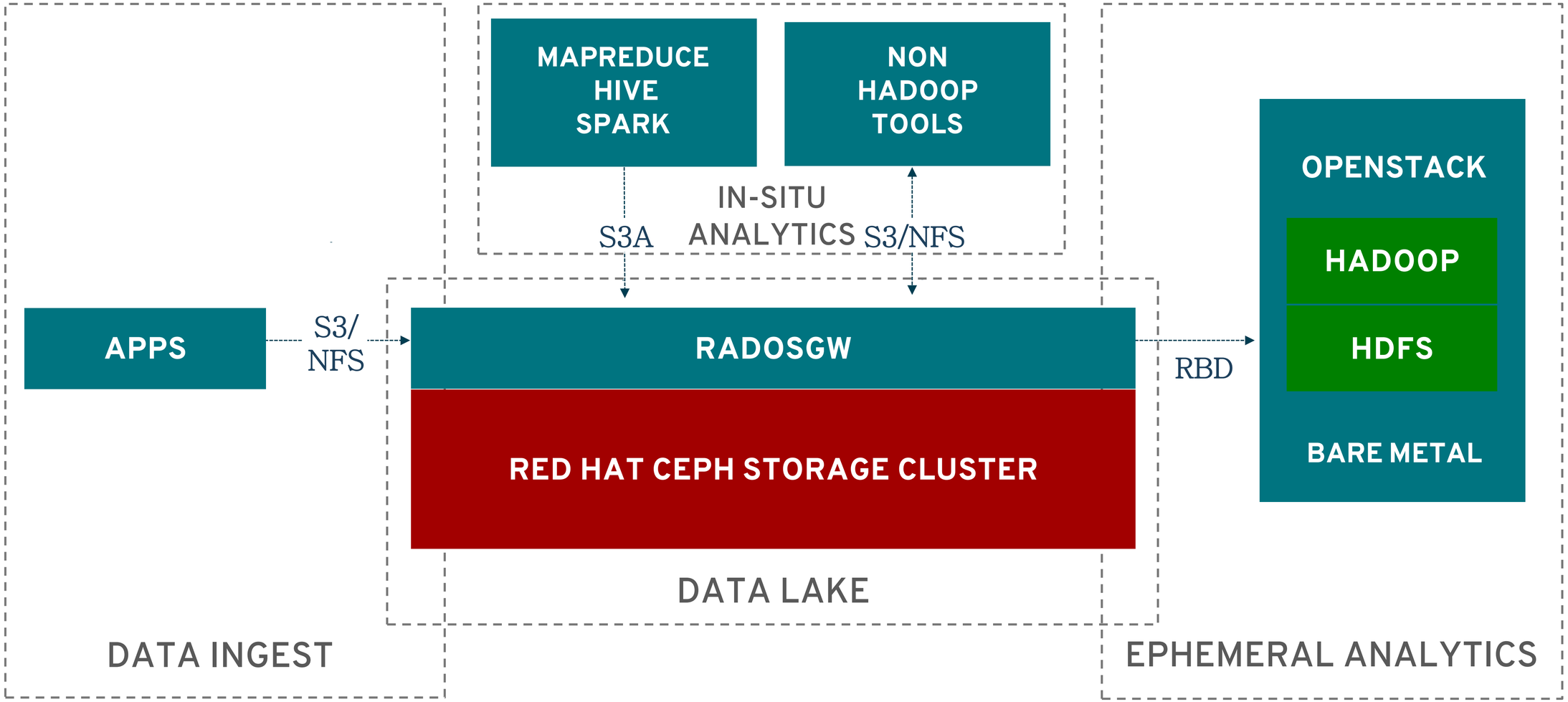
WEB APPLICATION STORAGE



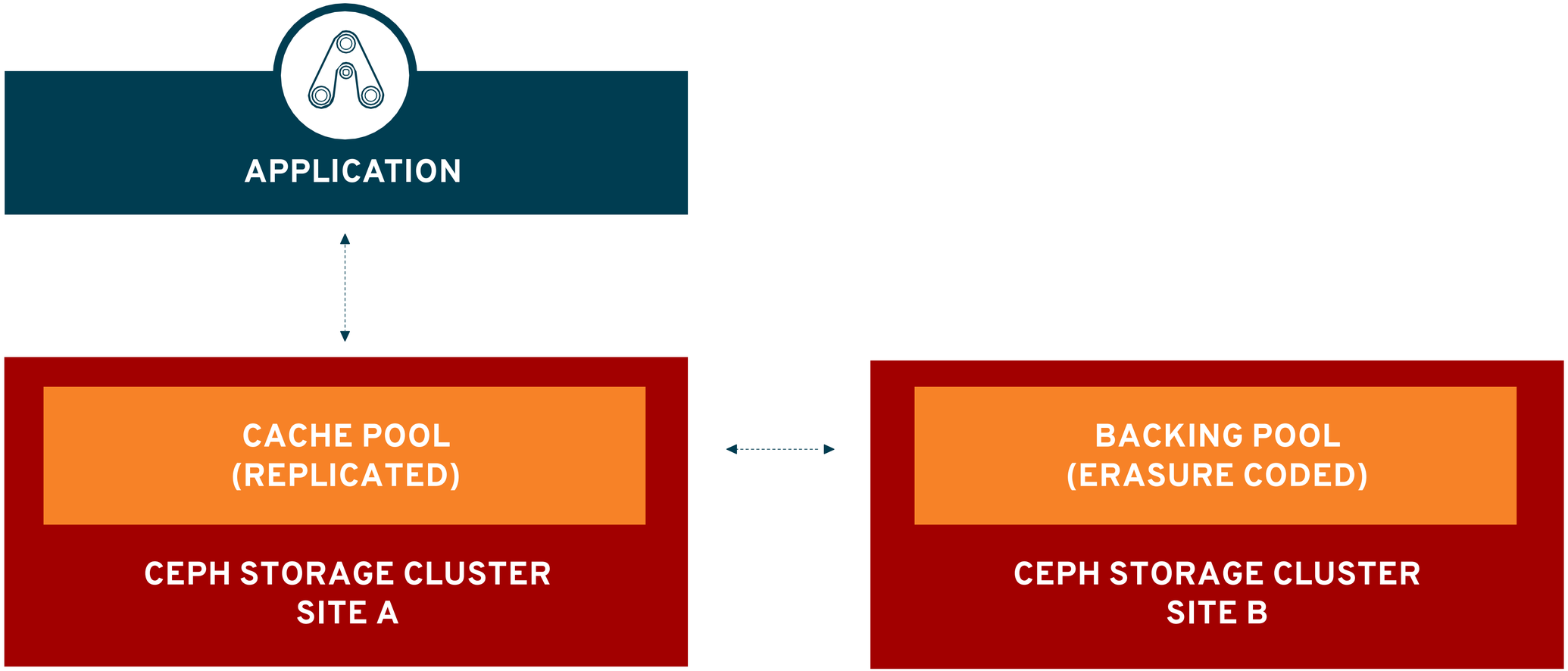
MEDIA REPOSITORY



DATA LAKE FOR ANALYTICS



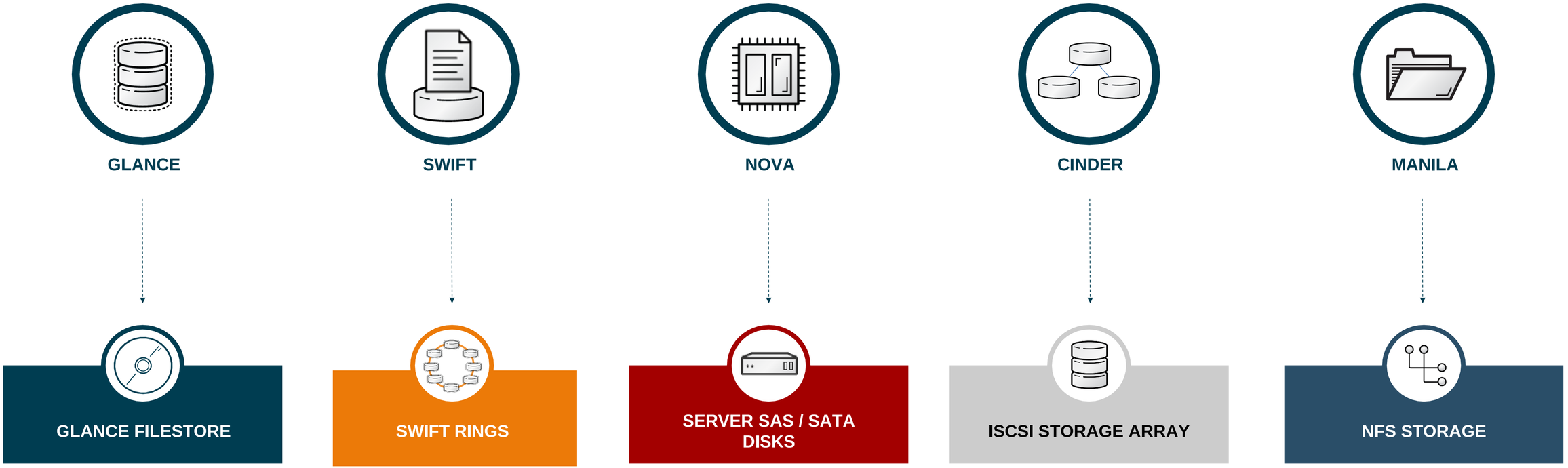
BACKUP / ARCHIVE



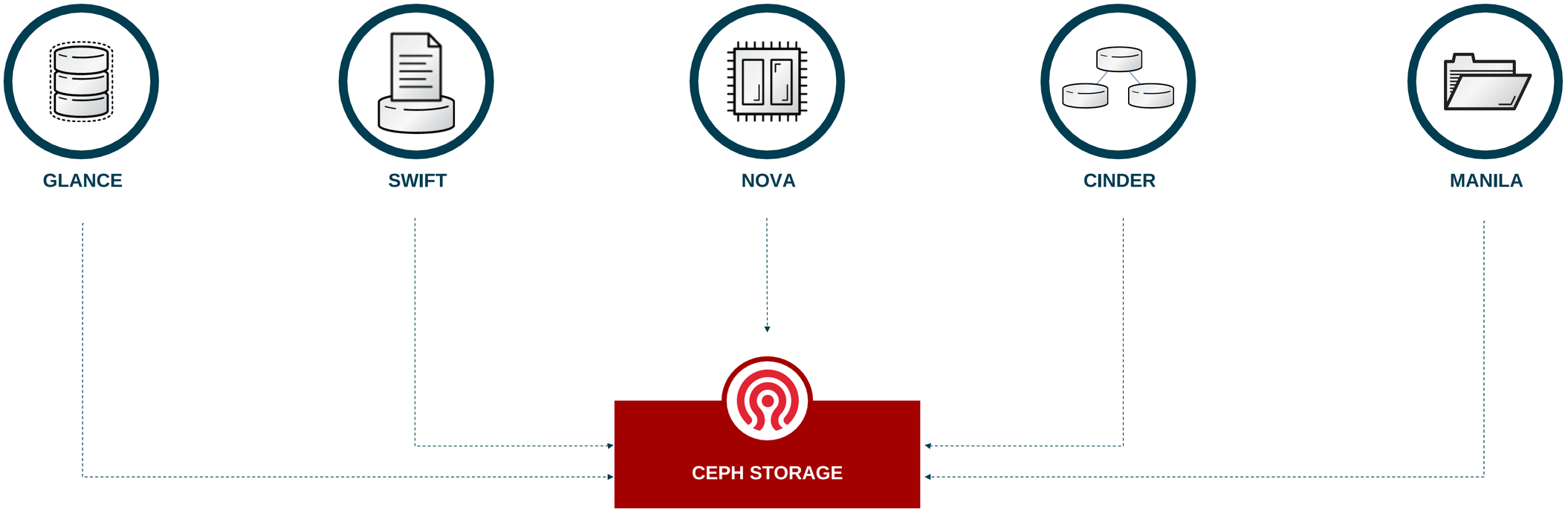
목차

redhat.

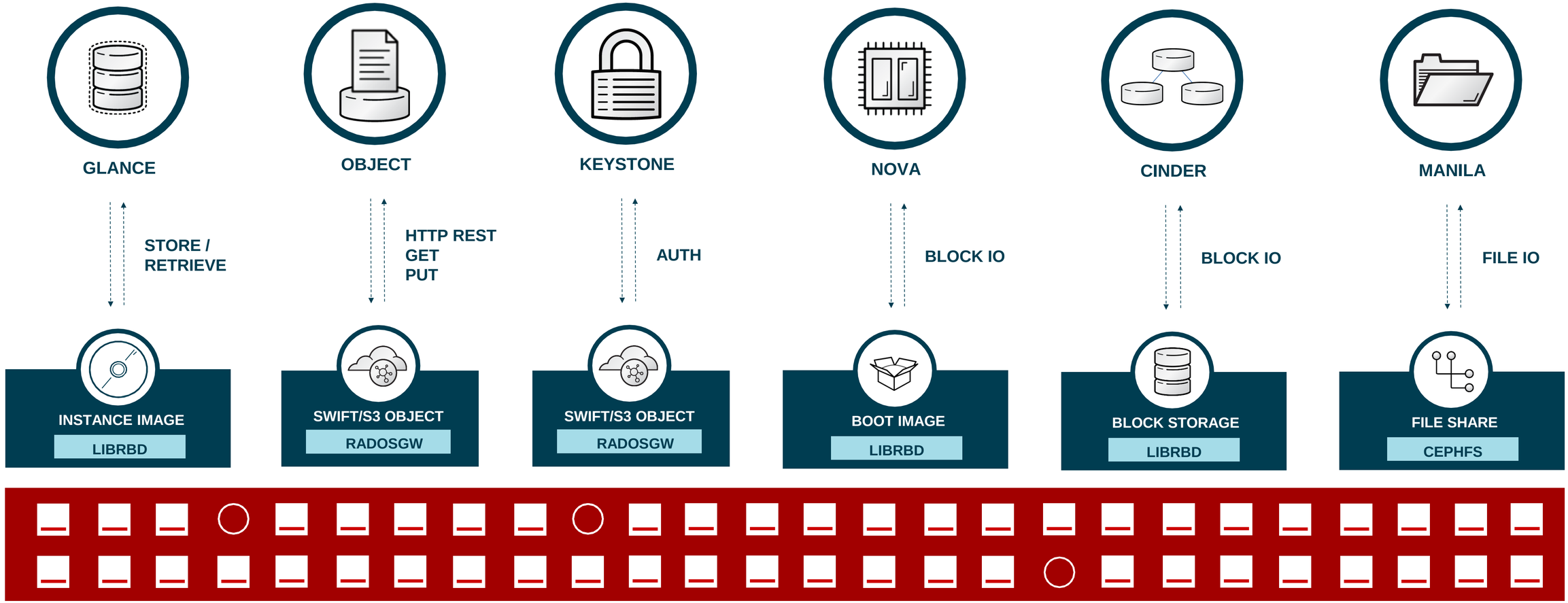
OPENSTACK WITHOUT CEPH



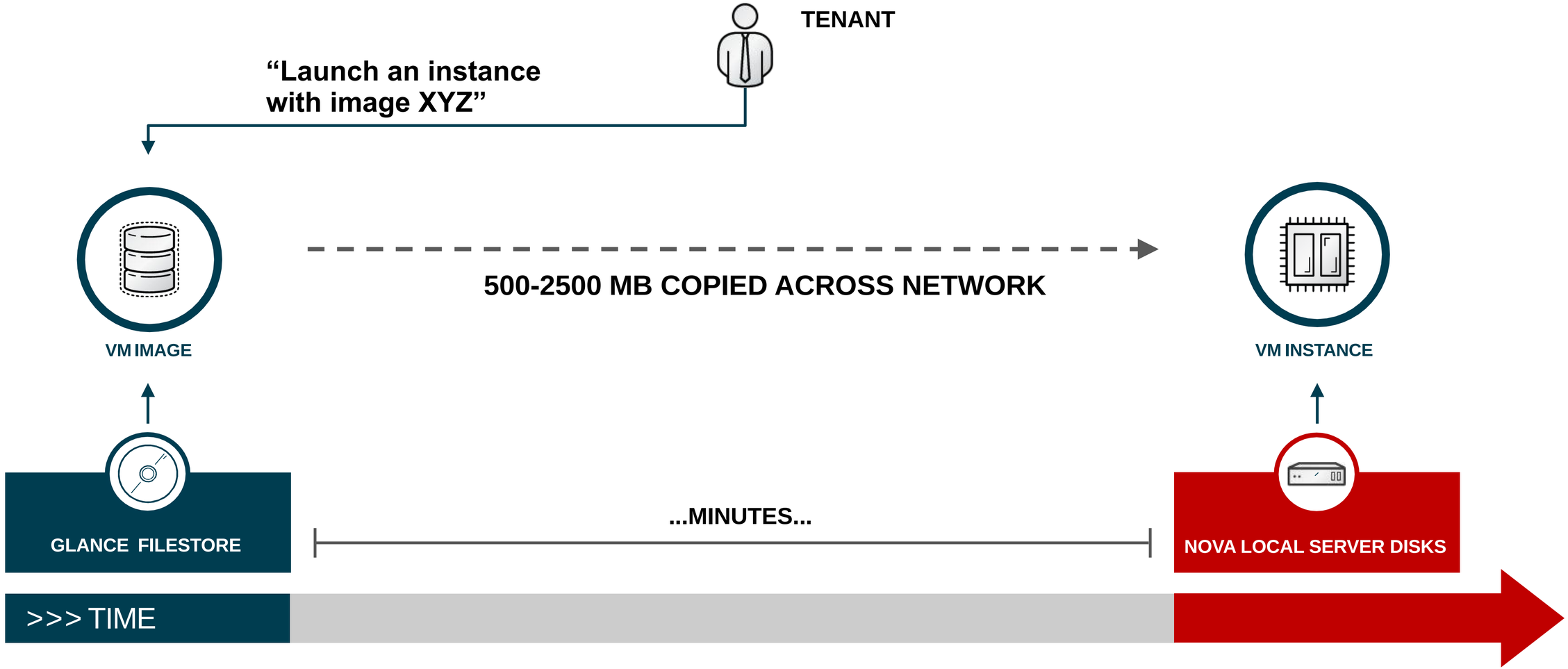
OPENSTACK WITH CEPH



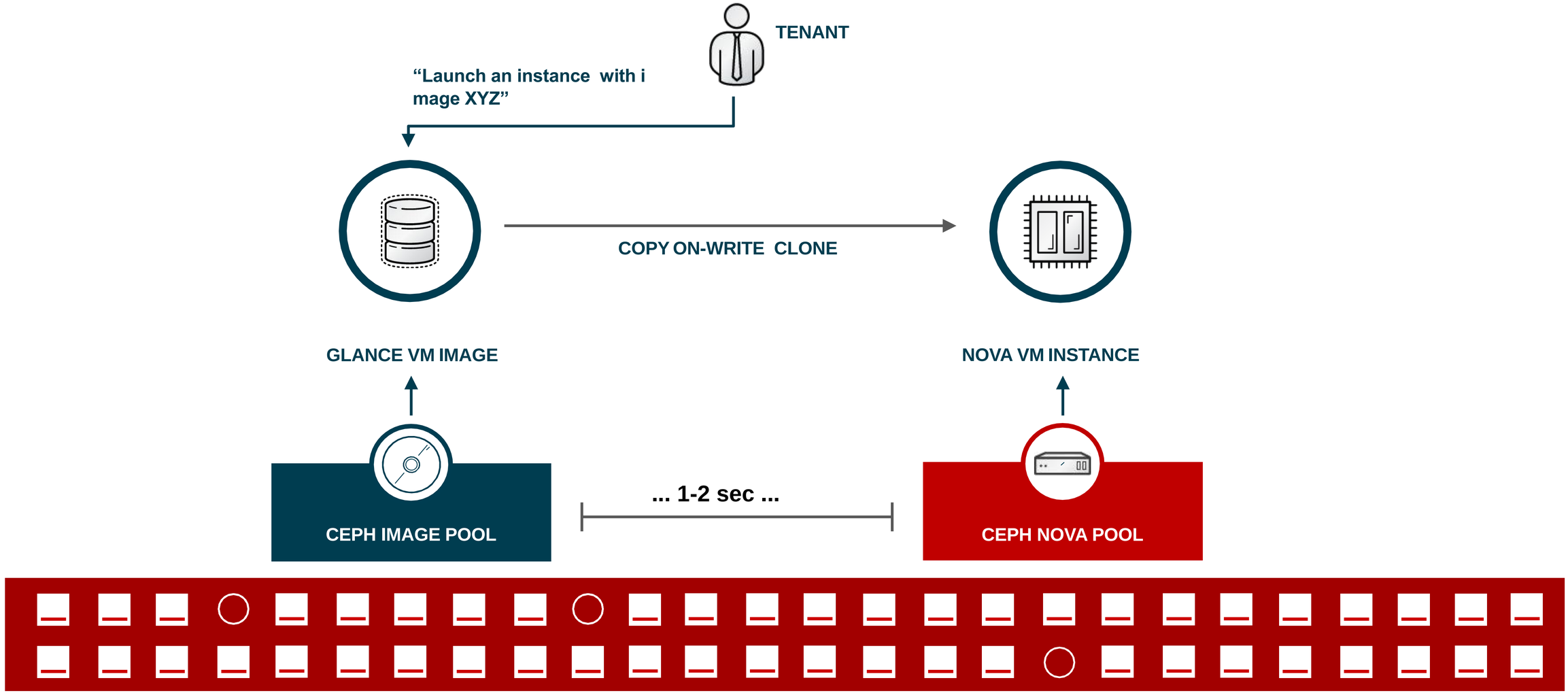
OPENSTACK INTEGRATION WITH CEPH



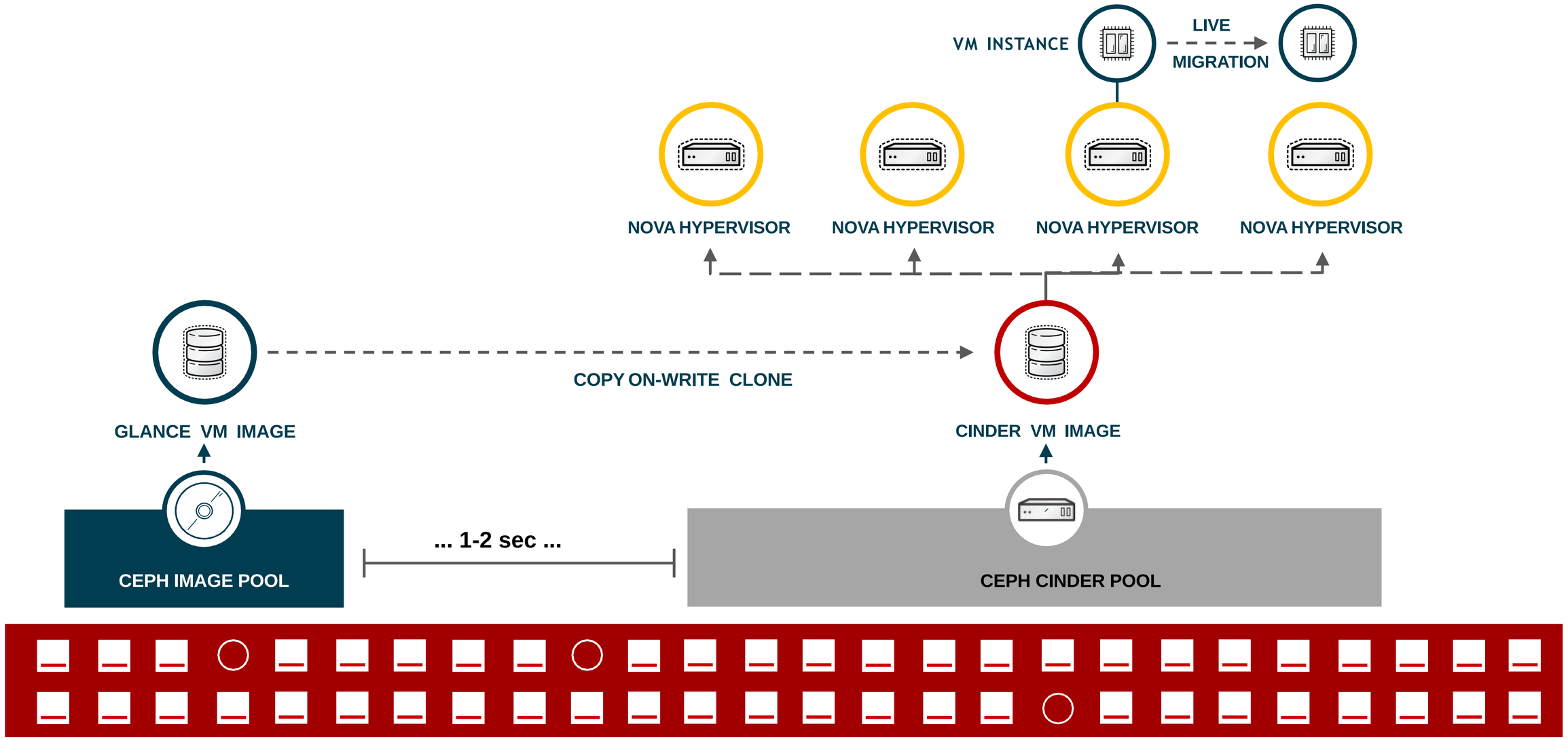
SLOW INSTANCE BOOT WITHOUT CEPH



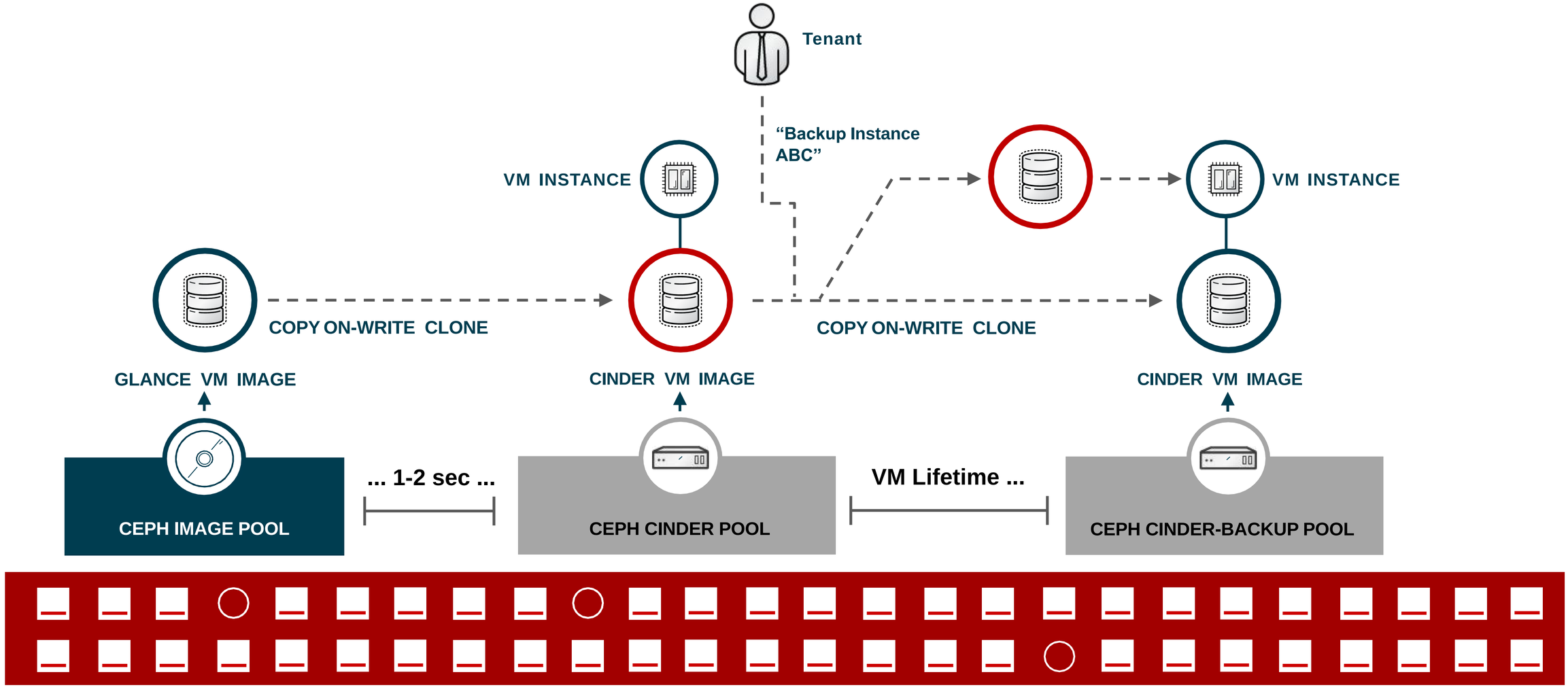
FAST INSTANCE BOOT WITH CEPH



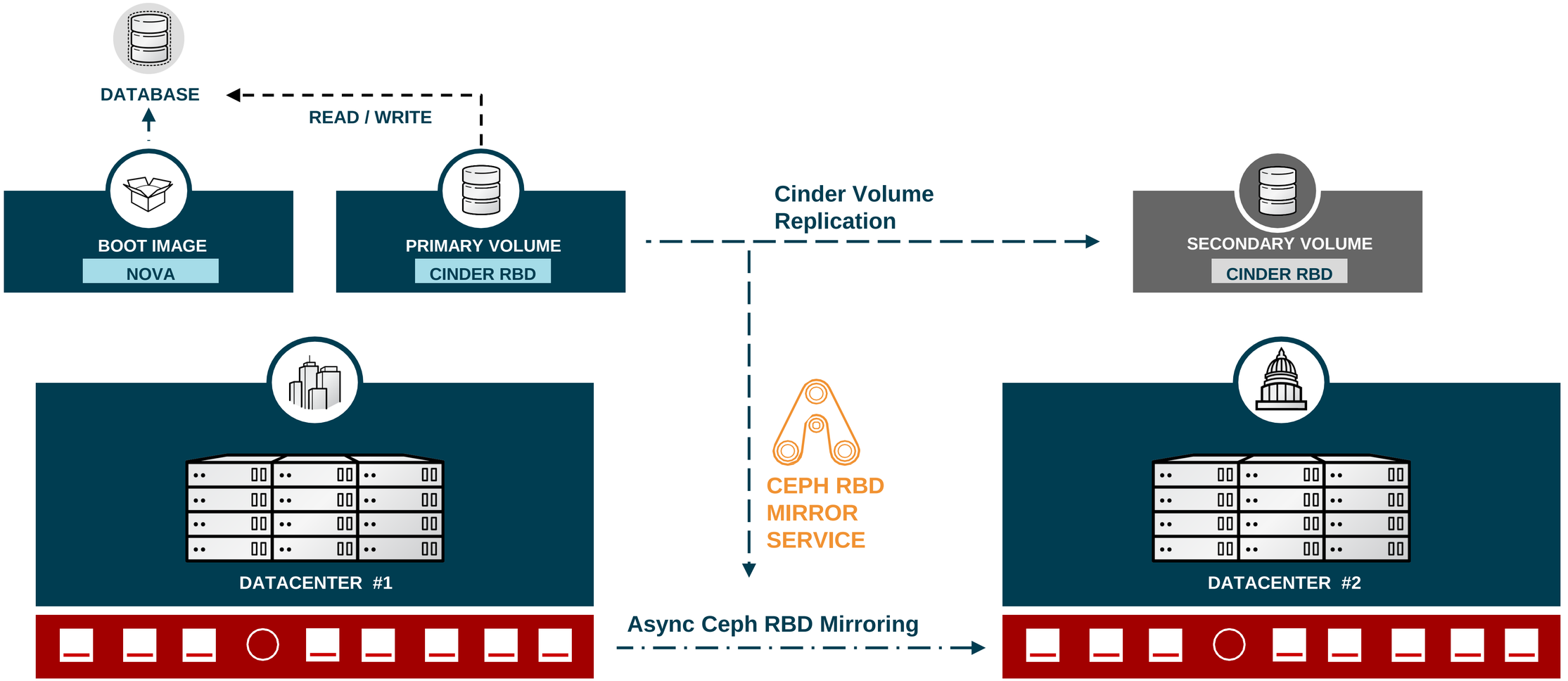
VM HIGH-AVAILABILITY WITH CEPH AND CINDER



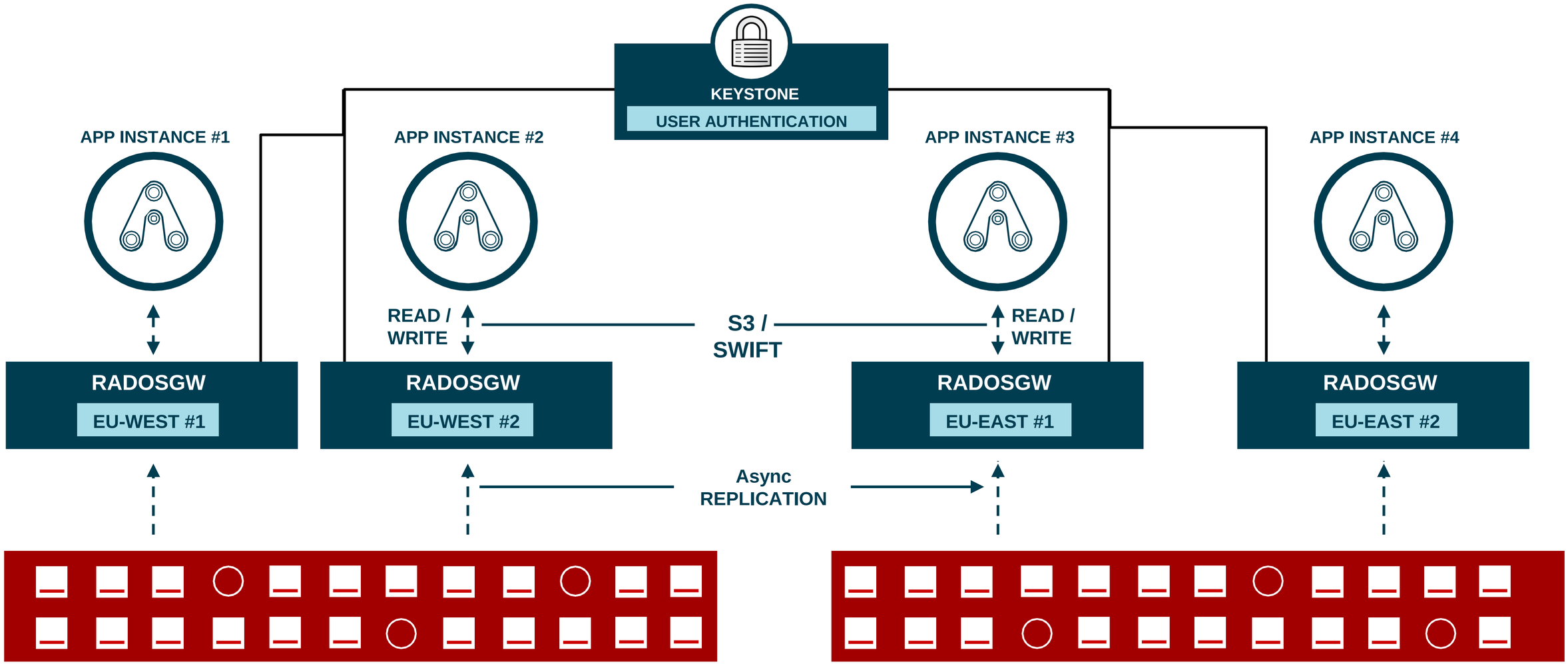
VM BACKUP WITH CEPH AND CINDER



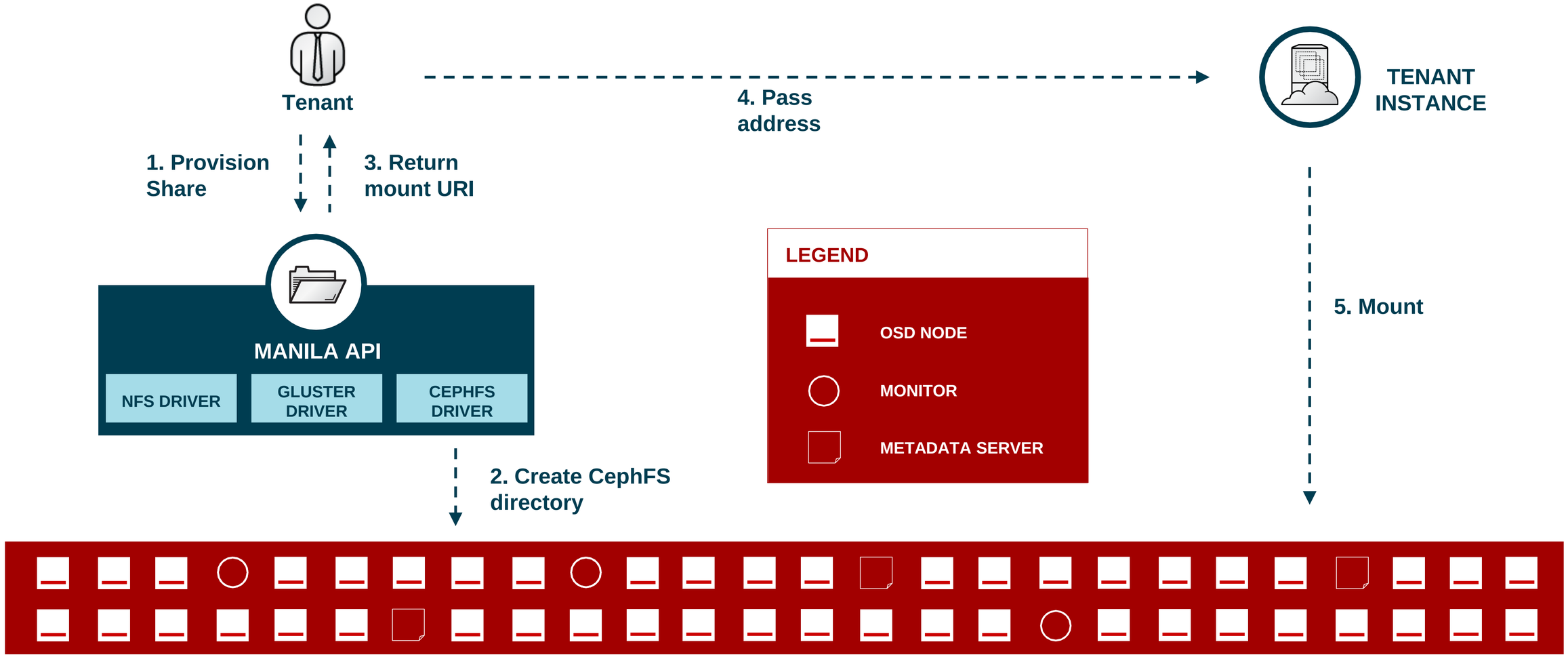
GEO-REPLICATED VM STORAGE WITH CEPH



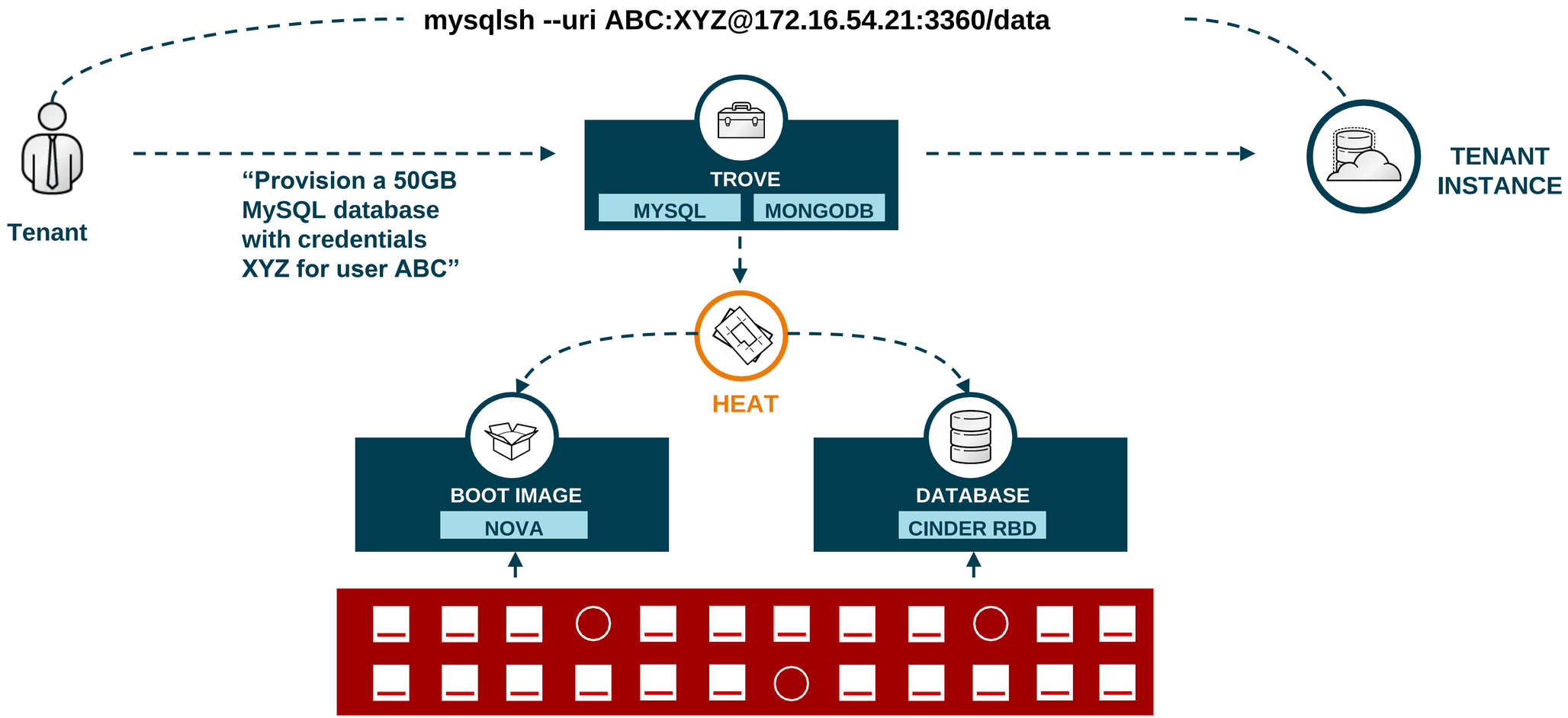
MULTI-SITE S3 STORAGE WITH CEPH RGW



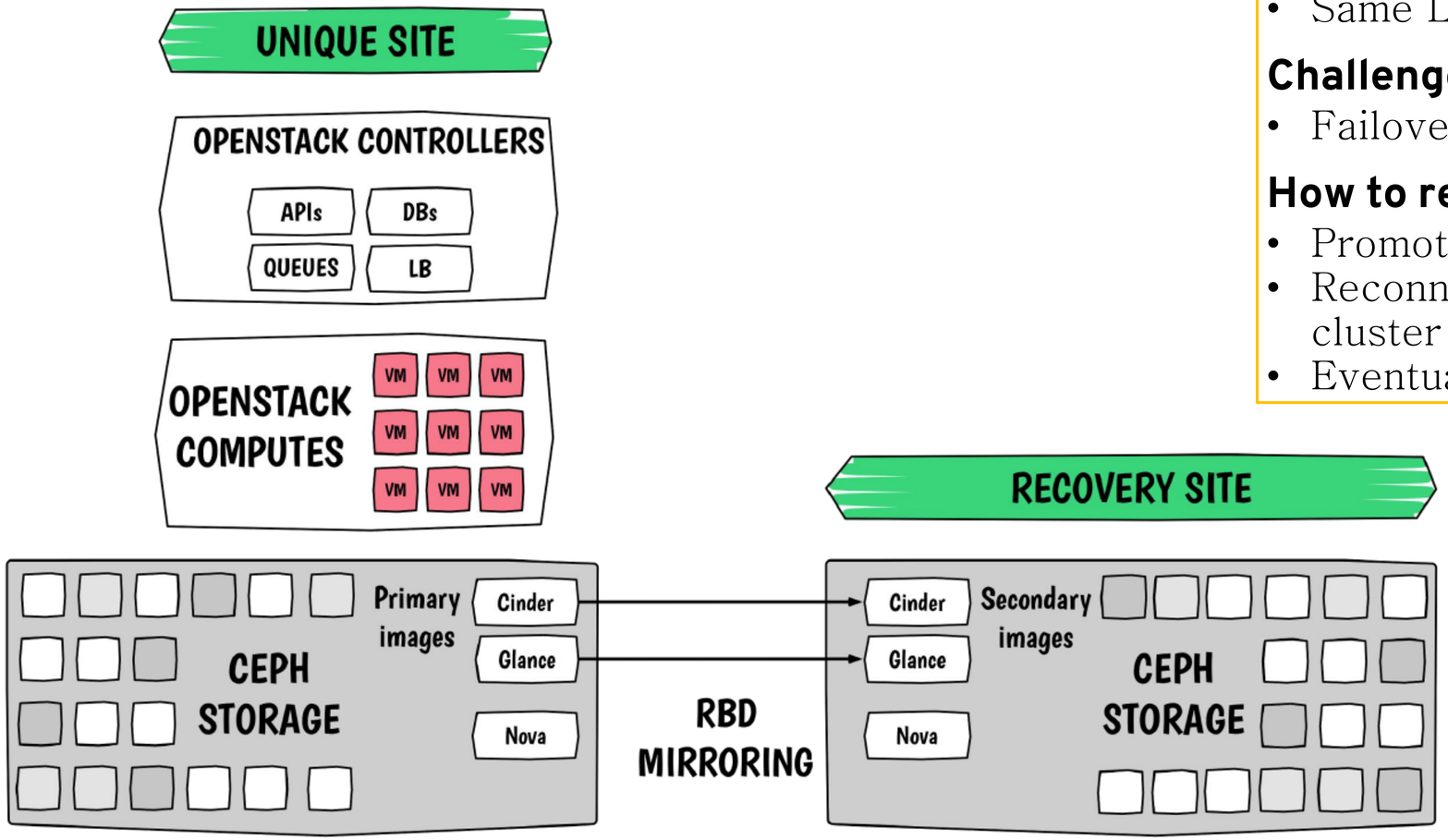
FILESHARE-AS-A-SERVICE WITH CephFS (MANILA)



DATABASE-AS-A-SERVICE WITH CEPH (TROVE)



재해복구 시나리오 #1



Properties:

- Single OpenStack site
- A data recovery site
- Both sites have with the same cluster FSID
- Same L3 segment

Challenge:

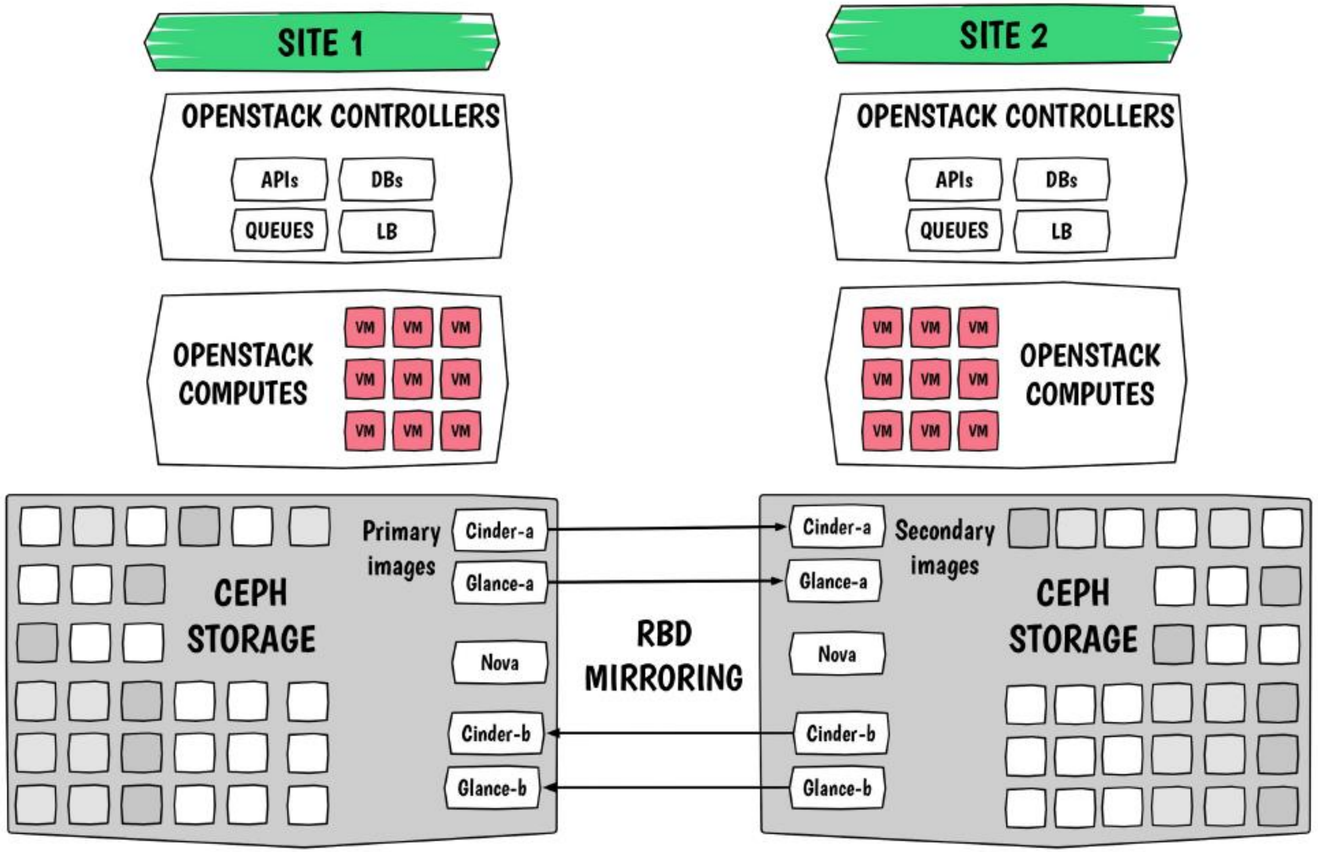
- Failover procedure

How to recover?

- Promote secondary site
- Reconnect all the services to the recovery cluster
- Eventually move back to the primary site

재해복구 시나리오 #2

No Shared Keystone



Properties:

- Keystone on the controllers (as usual)
- Individual login on each region/site
- Both sites have each other's data
- Both sites have the same cluster FSID

Challenge:

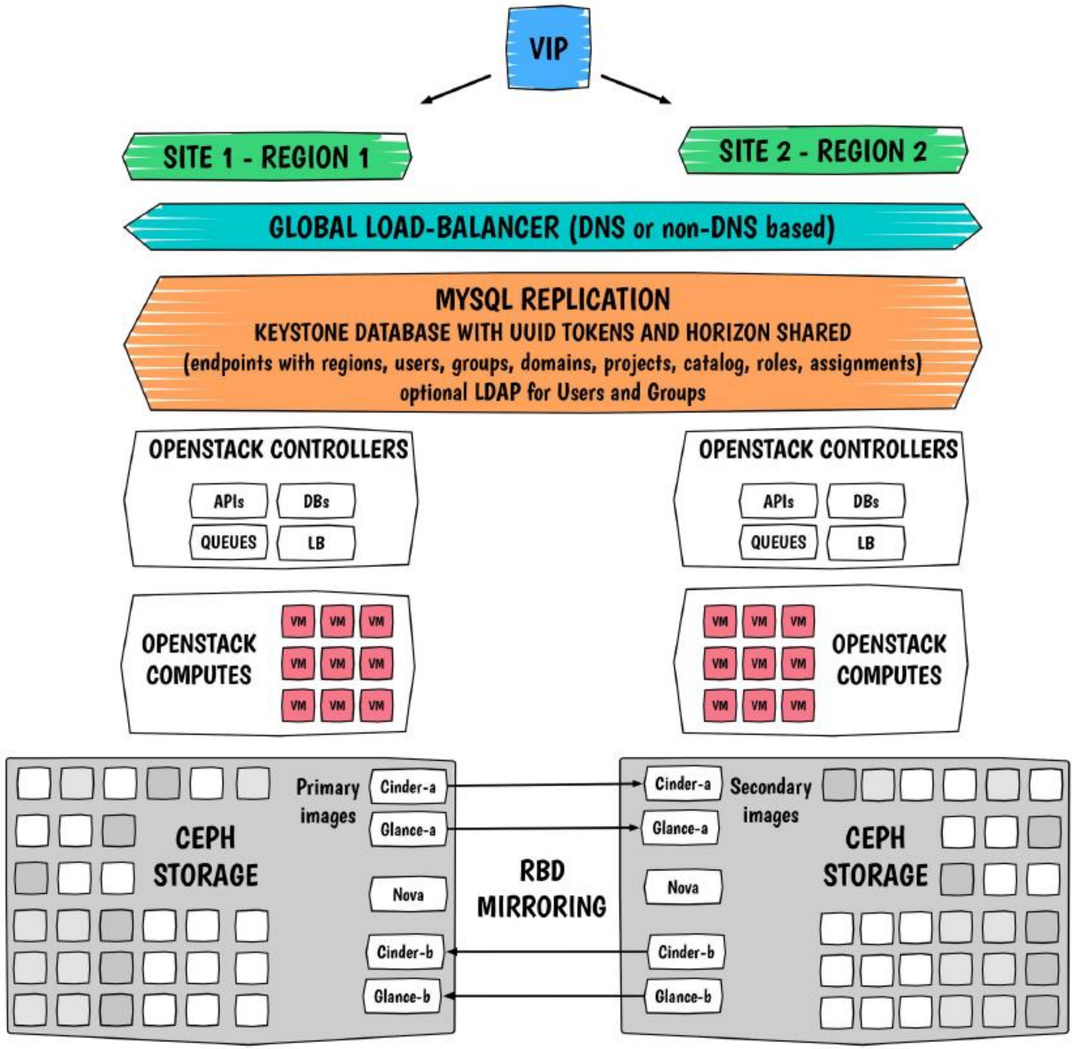
- Replicate metadata for images and volumes

How to recover?

- Promote the secondary site
- Import DB records in the survival site

재해복구 시나리오 #3

Shared Keystone with Region



Properties:

- Shared Keystone
- Keystone centralized and replicated DB
- Both sites have each other's data
- Works with N sites
- Both sites have with the same cluster FSID

Challenge:

- Replicate UUID tokens
- MySQL cross-replication over WAN
- Requires low latency and high bandwidth
- Fernet tokens are not ready yet

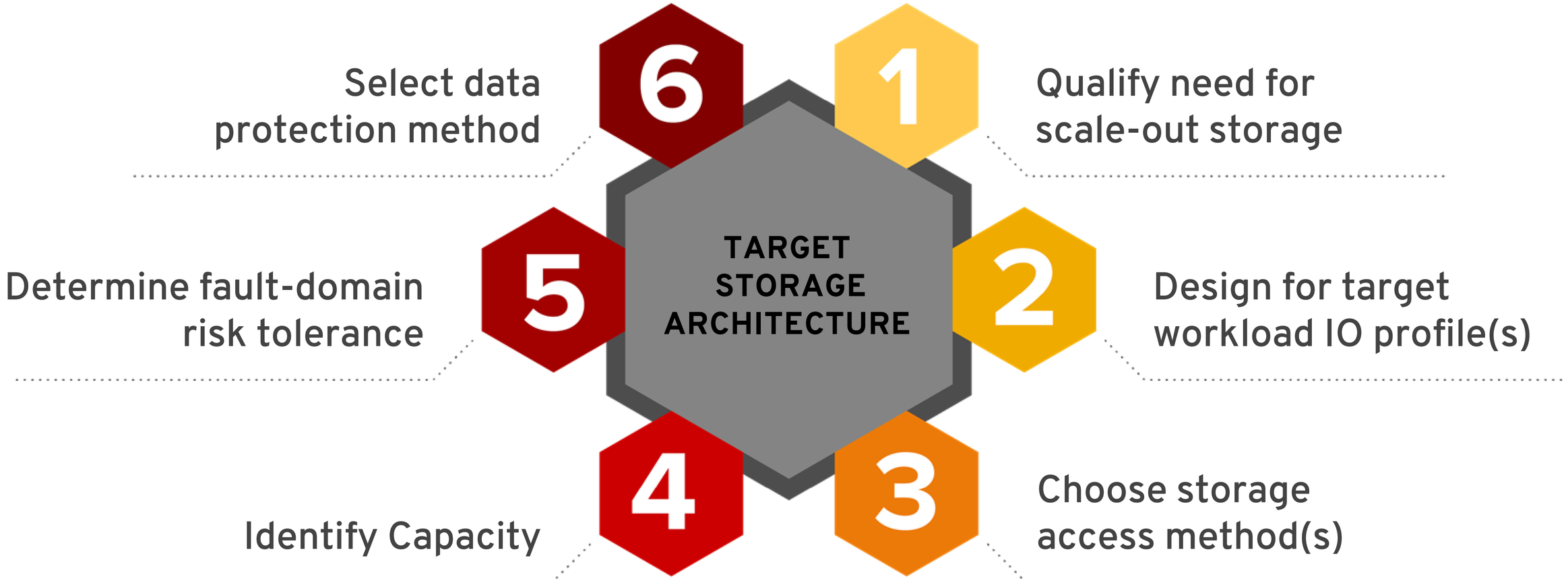
How to recover?

- Promote the secondary site

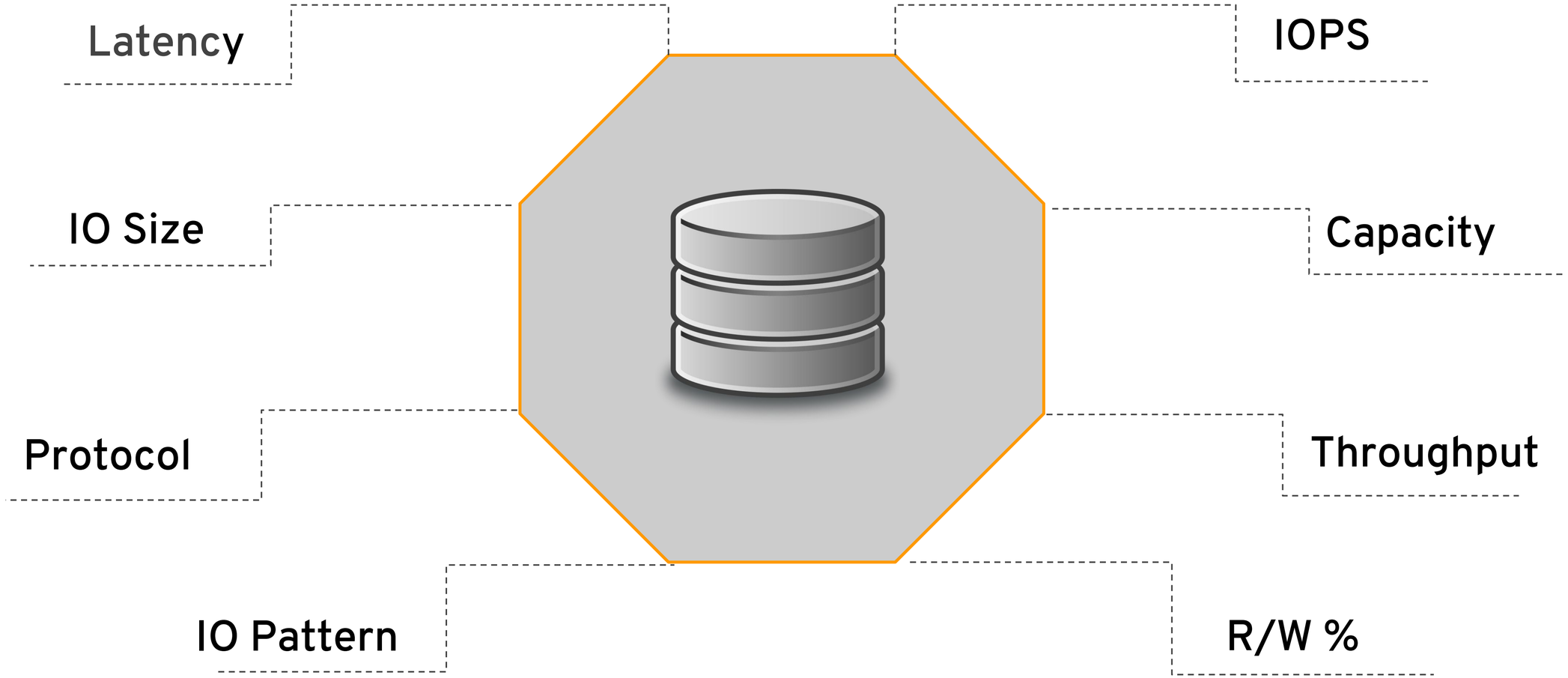
목차

redhat.

STORAGE DESIGN



UNDERSTANDING YOUR WORKLOADS



OTHER FACTORS

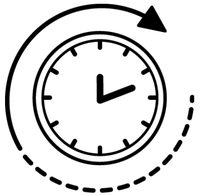
Hyper-convergence

x-86 Architecture

Reduced Costs

Disaster Recovery

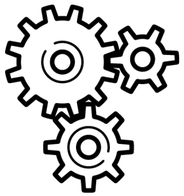
Backup / Replication



BCP

Management

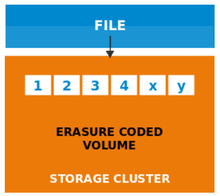
Monitoring



Operations

Thin Provisioning

Dedup/ compression

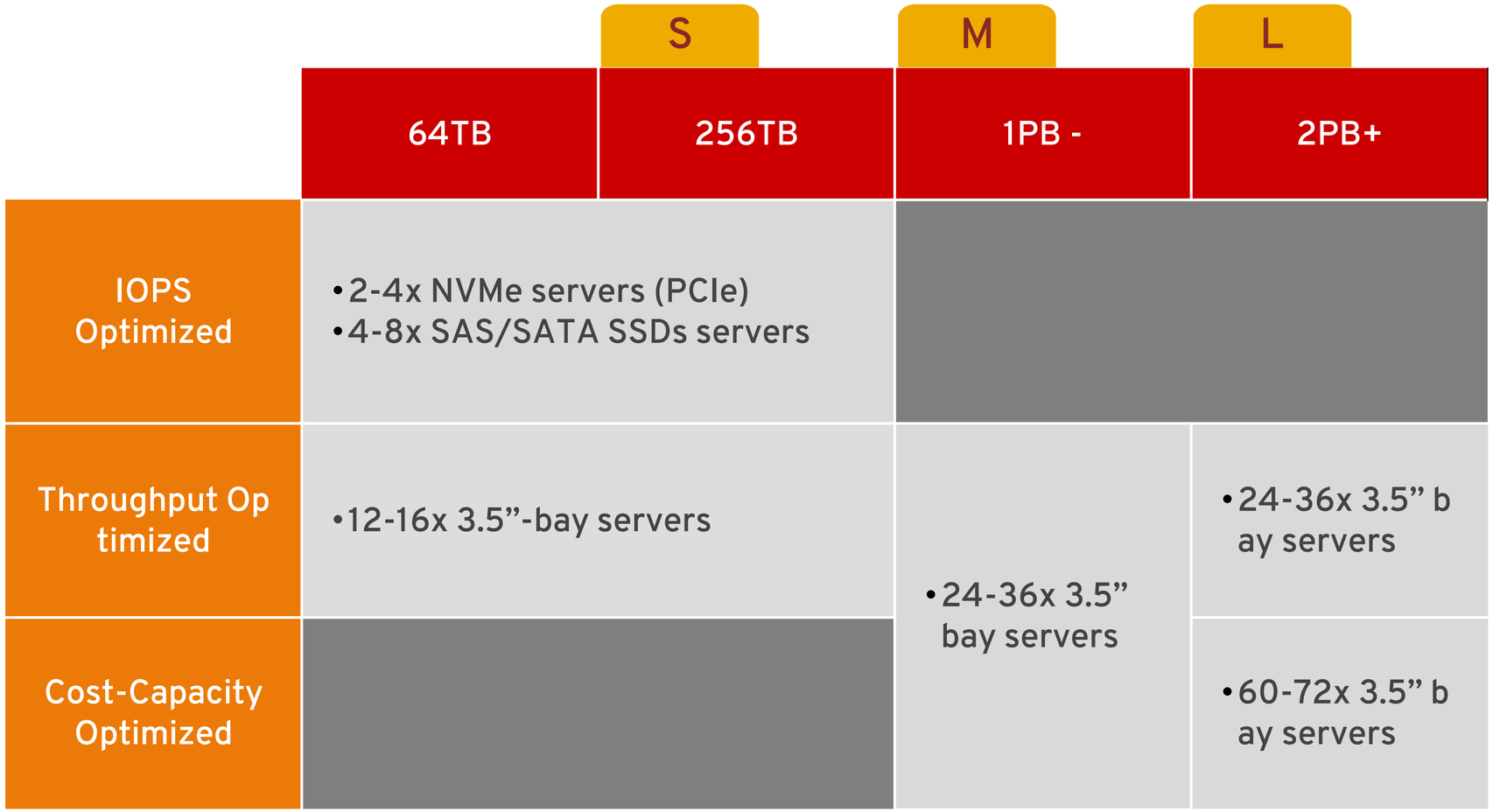


Advanced features

Ceph is not just scale out capacity

IOPS Optimized	Throughput Optimized	Cost / Capacity Optimized
NVMe SSD in SLED chassis	SSD, HDD in standard / dense chassis	HDD in dense / ultra-dense chassis
High IOPS / GB Smaller, random IO Read / write mix	High MB/s throughput Large, sequential IO Read / write mix	Low cost / GB Sequential IO Write mostly
Use Case: MySQL	Use Case: Rich Media	Use Case: Active Archives

BROAD SERVER SIZE TRENDS



SERVER CONFIGURATION RECOMMENDATIONS

	64TB	256TB	1PB -	2PB+
IOPS Optimized	• Ceph block (RBD) • NVMe SSD w/ co-located write journals, OR SATA SSD w/ NVMe write journals • Multiple OSDs per flash drive • 10 Xeon® cores per NVMe SSD, OR 4 cores per SATA SSD • 2x or 3x replication (with backup)			
Throughput O ptimized	• Ceph block or object (RBD or RGW) • HDDs w/ NVMe or SATA SSD write journals • 1 Xeon® core per 2 HDDs • Single OSD per HDD • 10GbE or 40GbE with 12+ HDD per chassis • 3x replication			
Cost-Capacity Optimized			• Ceph object (RGW) • HDD drives with no SSD journal • 1 Xeon® core per 2 HDDs • Single OSD per HDD • Erasure-coded	

목차

redhat.

Some Technical References

RHCS Test Drive : Hands-on Lab for Ceph
- ★ http://bit.ly/ceph-test-drive

RHCS Hardware Selection Guide
- ★ http://bit.ly/RHCS-hardware-selection-guide

RHCS Hardware Configuration Guide
- ★ http://bit.ly/RHCS-hw-configuration-guide

MySQL on RHCS Reference Architecture
- ★ http://bit.ly/MySQL_DB-on-RHCS

RHCS on Intel CPUs and SSDs Config Guide
- ★ http://bit.ly/RHCS-on-Intel

RHCS Ready Supermicro Server SKUs
- ★ http://bit.ly/RHCS-SuperMicro-SKU

RHCS on CISCO UCS Servers
- ★ http://bit.ly/RHCS-on-Cisco-UCS

RHCS on QCT Servers Perf & Sizing Guide
- ★ http://bit.ly/RHCS-on-QCT

RHCS on Supermicro Servers Perf & Sizing Guide
- ★ http://bit.ly/RHCS-on-SuperMicro

RHCS on DELL EMC PE 730xd Servers Perf & Sizing Guide
- ★ http://bit.ly/RHCS-on-DellEMC-PE730xd

RHCS on DELL EMC DSS 7000 Servers Perf & Sizing Guide
- ★ http://bit.ly/RHCS-on-DellEMC-DSS7000

RHCS on Samsung Sierra Flash Array Perf & Sizing Guide
- ★ http://bit.ly/RHCS-on-Samsung-flash-array

RHCS Ready QCT Server SKUs
- ★ http://bit.ly/RHCS-QCT-SKU

RHCS on SanDisk Infiniflash
- ★ http://bit.ly/RHCS-on-Sandisk-Infiniflash

RHCS and RHOSP HCI Ref. Arch
- ★ http://bit.ly/RHCS-RHOSP-HCI

감사합니다

Web: www.redhat.com/storage

Blog: http://redhatstorage.redhat.com/

Twitter: www.twitter.com/redhatstorage

Facebook: www.facebook.com/RedHatStorage/

YouTube: www.youtube.com/user/redhatstorage

Slideshare: www.slideshare.net/Red_Hat_Storage

redhat.