

STORAGE DEVELOPER CONFERENCE



Fremont, CA
September 12-15, 2022

BY Developers FOR Developers

A **SNIA** Event

Behind the scenes for Azure Block Storage unique capabilities

Yuemin Lu

Greg Kramer (presenting on behalf of Malcolm Smith)

Agenda

- Azure Storage Overview
- Unique capabilities & engineering designs
- Q & A

Azure Storage Overview

Azure Storage Overview



Block Storage

Ultra Disks
Premium Disks
Standard Disks

Reliable, persistent, high performing storage for Virtual Machines



Object Storage

Azure Blobs
(w/ Data Lake Gen 2)

Secure, Scalable storage for unstructured data



File Storage

Azure Files
Azure NetApp Files

Lift and shift applications that require file shares to the cloud

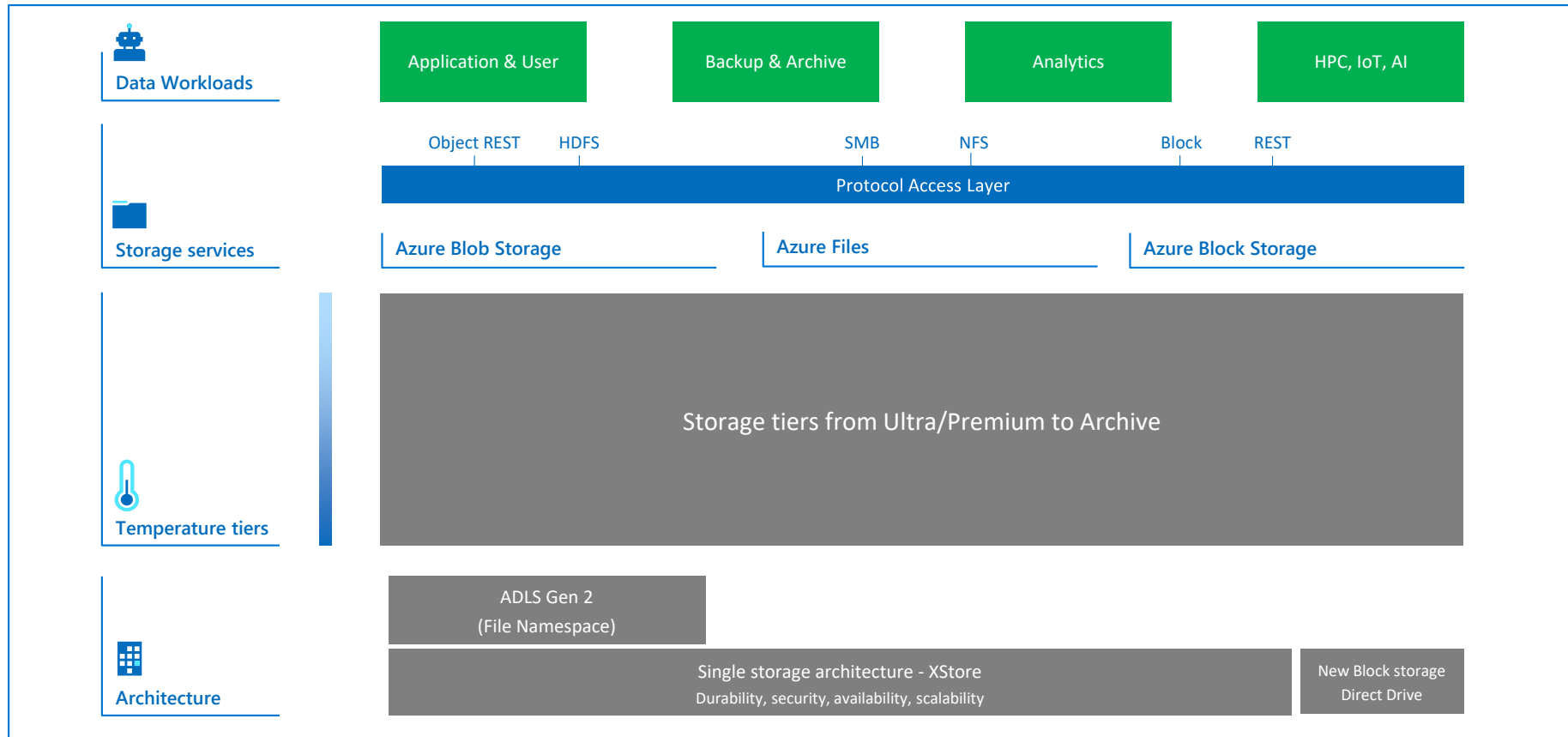


Hybrid Storage






Data Box Gateway/Edge
Azure File Sync
HPC Cache

Secure, intelligent data tiering between on-premises and cloud

Azure Storage Overview



Azure Block Storage for every workload

Azure Disk – Optimized for Virtual Machines					
	Standard HDD	Standard SSD	Premium SSD	Premium SSD v2	Ultra Disk
					
	Low-cost storage	Consistent performance	High performance	Sub-millisecond latency	Sub-millisecond latency
Workloads	Backups, low end file server, test & dev	Big Data, entry-level apps, small DBs, Web Servers	IO-Intensive, Database, production workloads, container volumes	SAP HANA, SAN, Tier-1 workloads	SAP HANA, SAN, Tier-1 workloads
Architecture	XStore			Direct Drive	

Inside Storage - XStore

Front-end layer

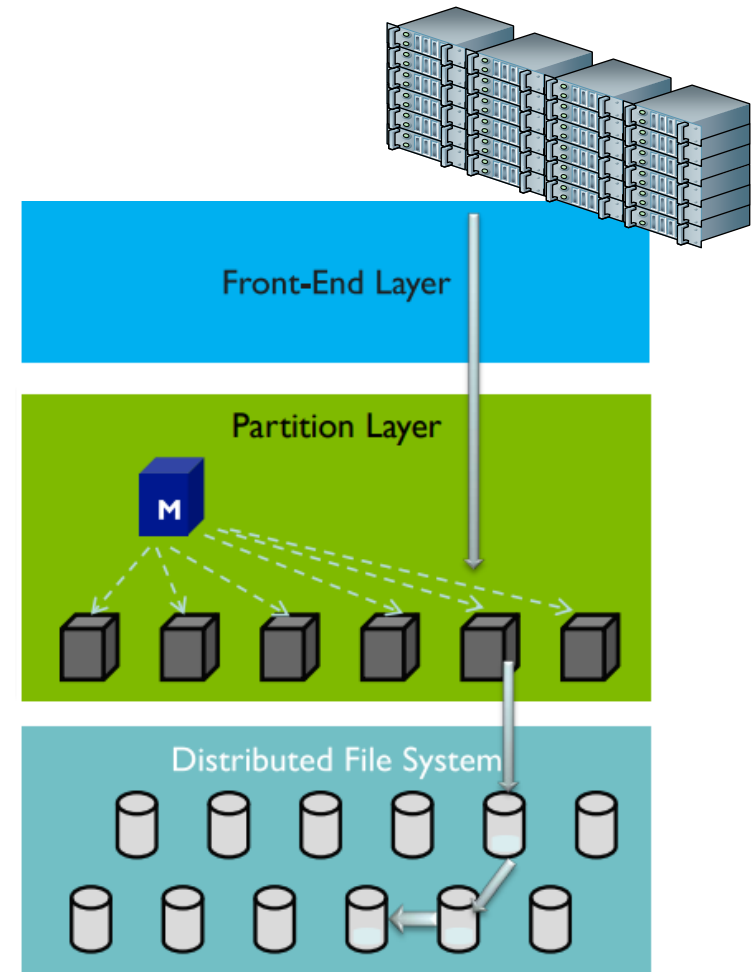
- Protocol endpoint
- Authentication/Authorization
- Metrics/logging

Partition layer

- Understands and manages our data abstractions
- Massively scalable index

Stream layer (Distributed File System)

- Data persistence and replication (JBOD)
- Append-only file system



Unique capabilities & engineering designs

Behind the scenes for Azure unique capabilities

- ✓ Instant snapshot create & restore
- ✓ Support data access over different protocol paths (SCSI, REST)
- ✓ Highly available storage with Zonal Redundancy Storage (ZRS)

This talk is about the engineering design, not specific products built on it

- Just because the architecture allows it, doesn't mean it's available in all products
- Always consult official docs for product capabilities, limitations, guarantees, etc.

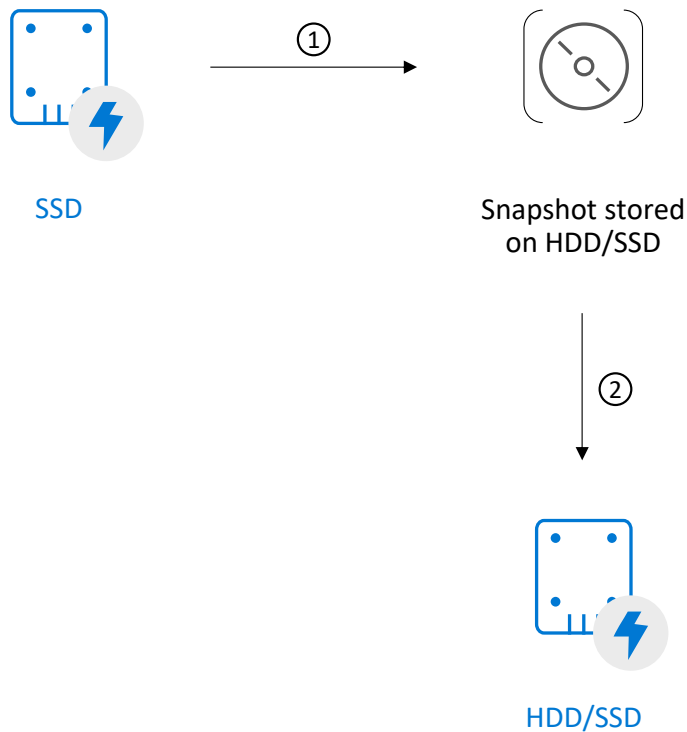
Behind the scenes for Azure unique capabilities

✓ Instant snapshot create & restore

✓ Support data access over different protocol paths (SCSI, REST)

✓ Highly available storage with Zonal Redundancy Storage (ZRS)

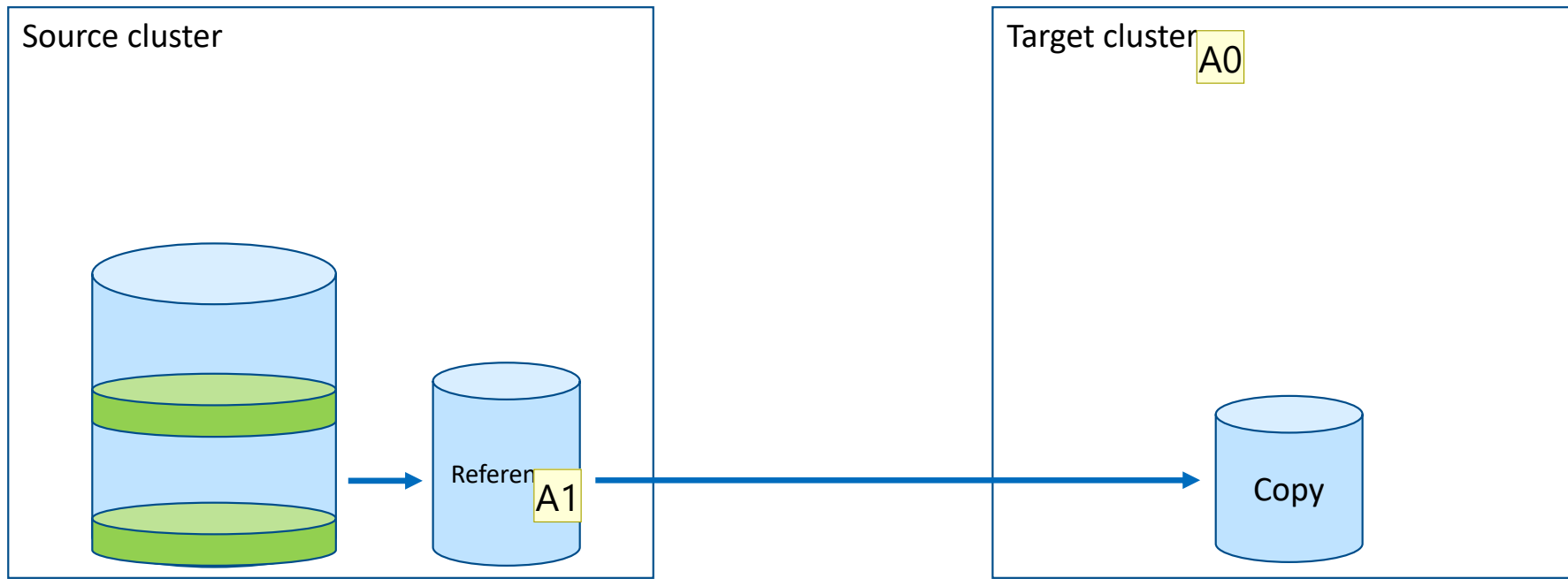
Instant snapshot create & restore



- ✓ Standalone incremental point-in-time backup
- ✓ Stored on storage of choice – HDD/SSD
- ✓ Instantly available for restore and copy

- ✓ Instant restore to storage of choice

Snapshot create



1. Metadata shall be copied to the reference blob
 - Reference to the range of changes

2. Create a new blob on the target cluster
 - Background copy from source to target

Slide 12

A0 What do these arrows mean? If it is the direction of copy then the link to source is not correct.

Author, 2022-09-03T16:39:00.084

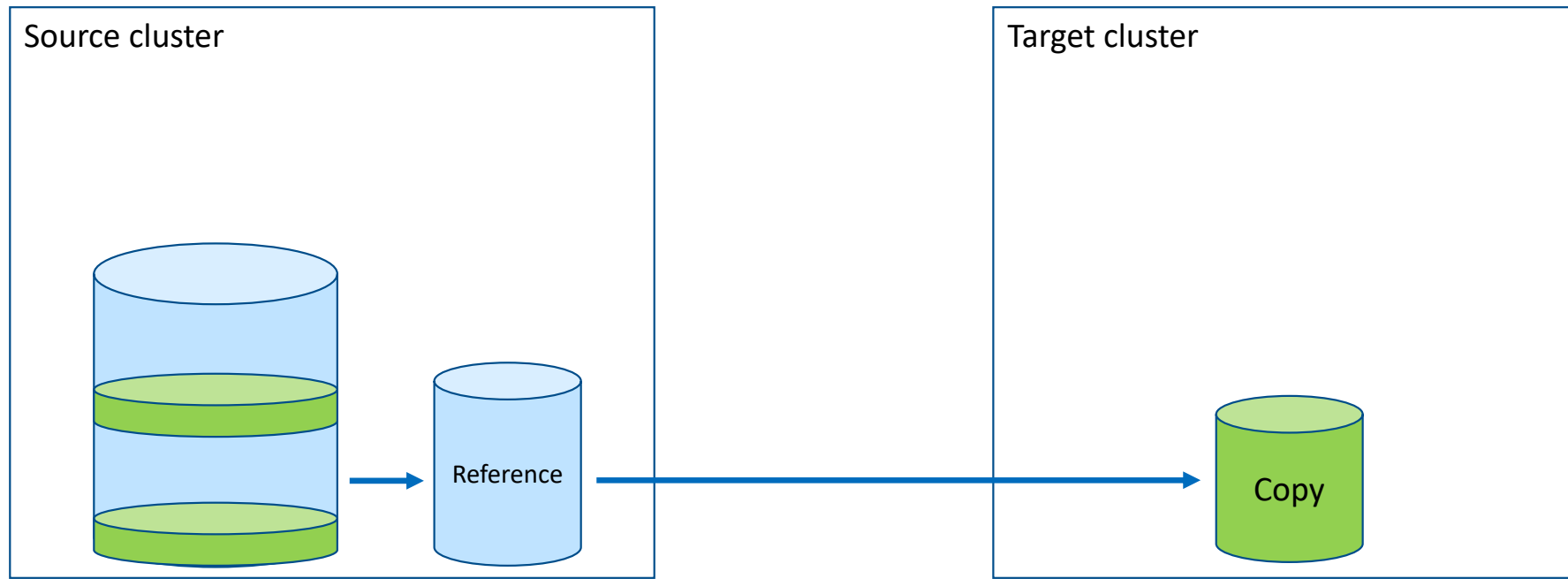
A0 0 Addressed.

Author, 2022-09-06T01:15:11.437

A1 Link is not important in this case. Technically, it is not a link either.

Author, 2022-09-03T16:39:32.977

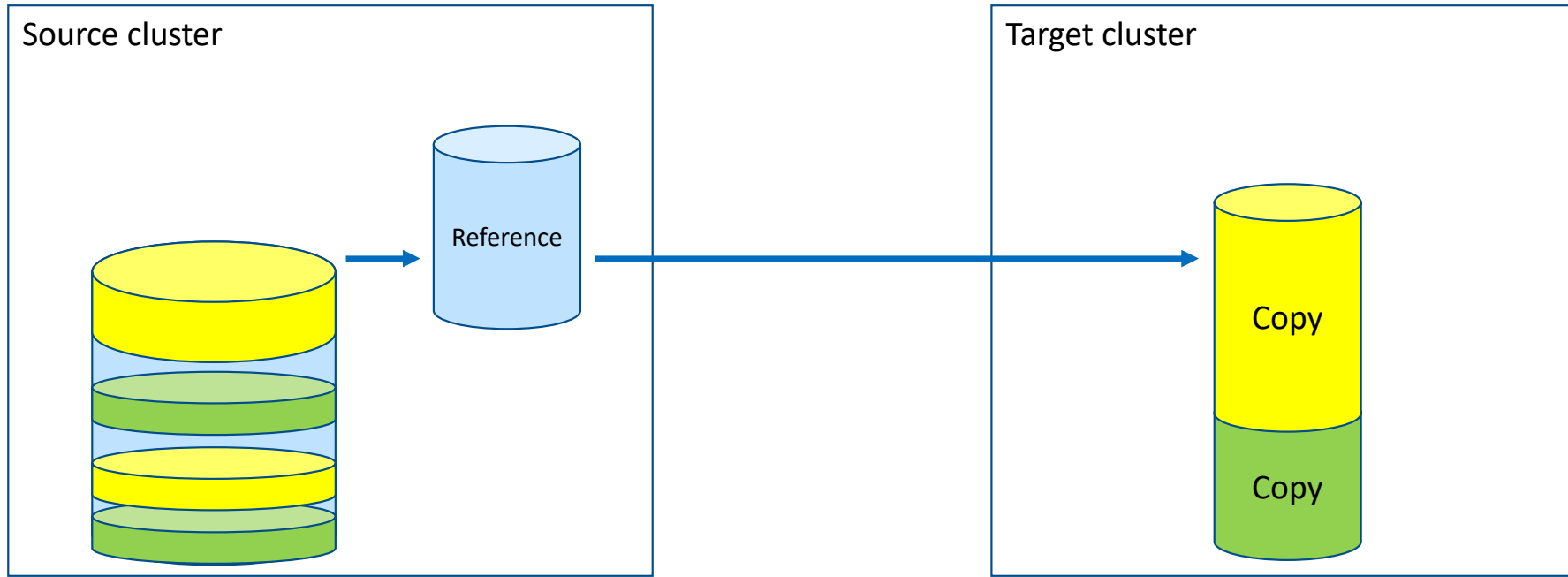
Snapshot create



1. Metadata shall be copied to the reference blob
 - Reference to the range of changes

2. Create a new blob on the target cluster
 - Background copy from source to target

Snapshot create



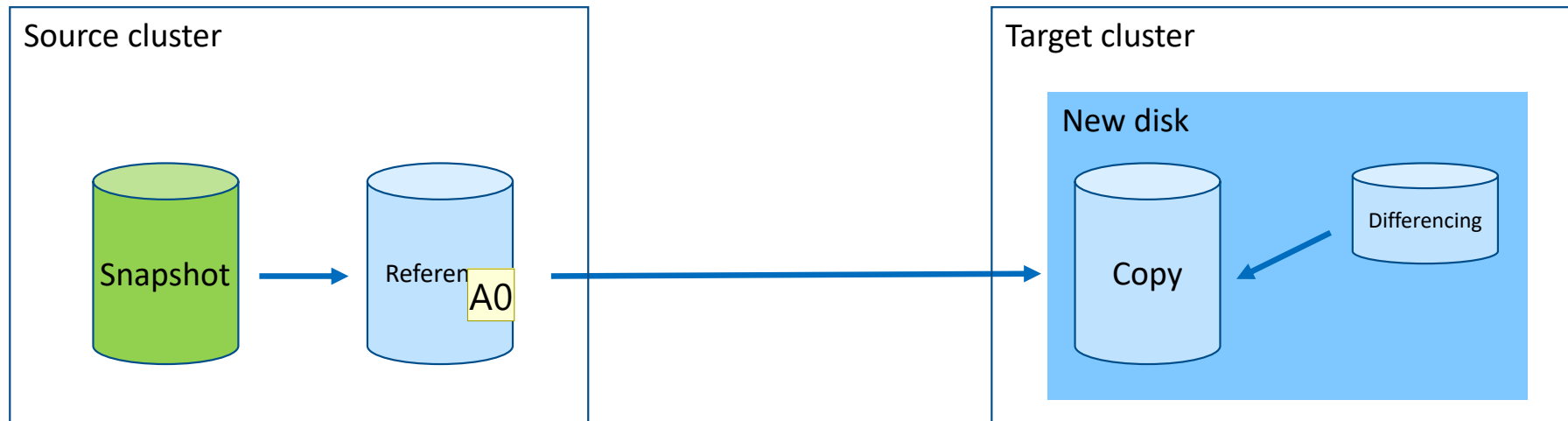
1. Create a new snapshot after new data is written

2. Only incremental changes from the previous snapshot is replicated to the new snapshot

Slide 14

A0 [Mention was removed] to add the export flow
Author, 2022-08-29T20:14:27.886

Snapshot restore

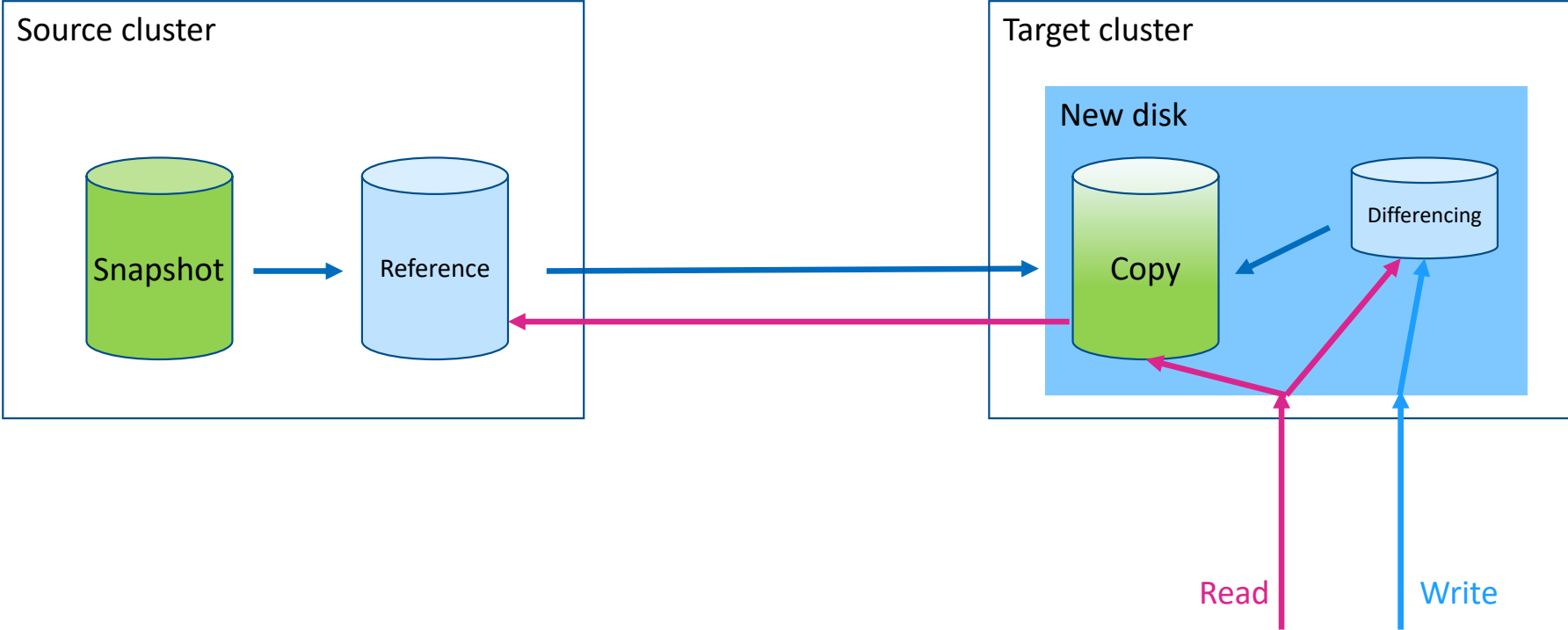


1. Customer data shallow copied to a reference blob
 - Allows snapshot be deleted
2. Create a new blob on the target cluster
 - Background copy from snapshot
3. Create a differencing blob to handle new writes

Slide 15

A0 The snapshot is a CoR from the previous slide. Again, I think we can omit the link on the source side.
Author, 2022-09-03T16:41:00.326

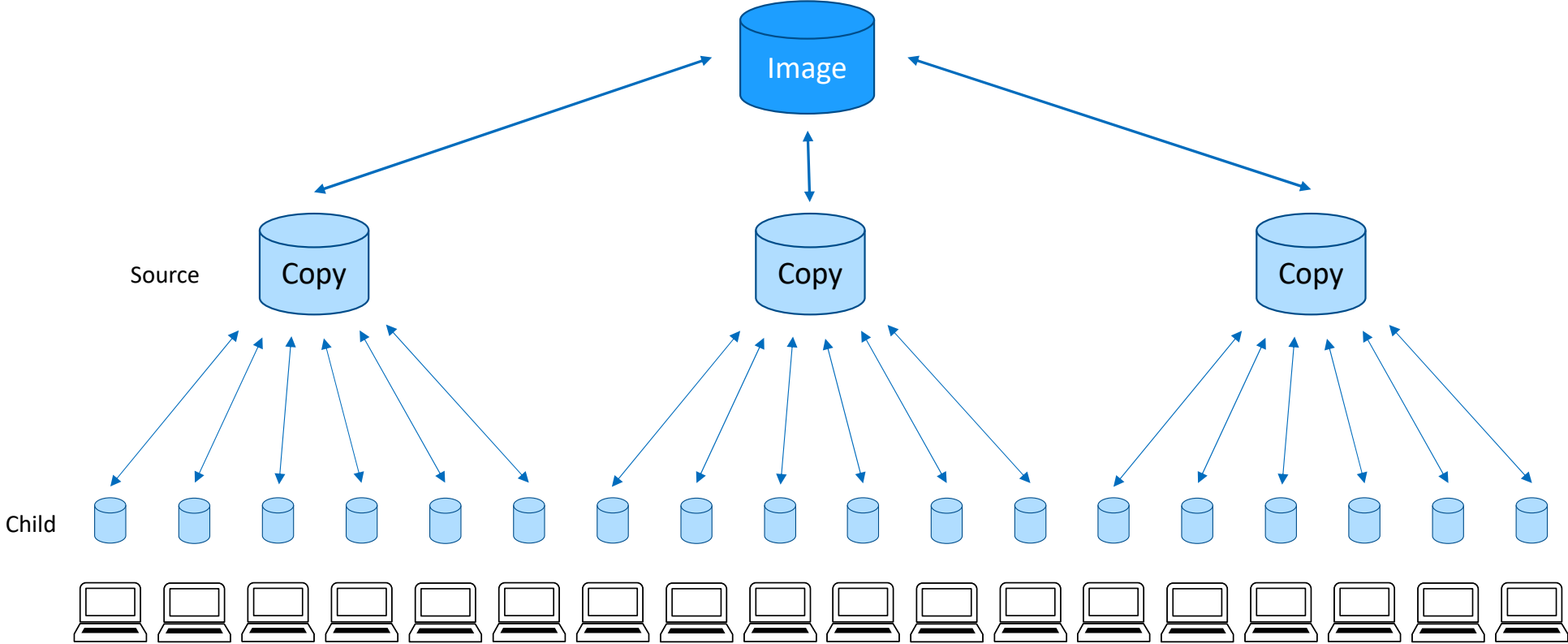
Snapshot restore



Other use case of CoR

- Provision many VMs from a golden image
 - All VMs use the same image for bootstrap
 - All VMs need a unique disk for differences
 - Scale up of new VMs must be fast

Solution: Copy on Read (CoR) Tree

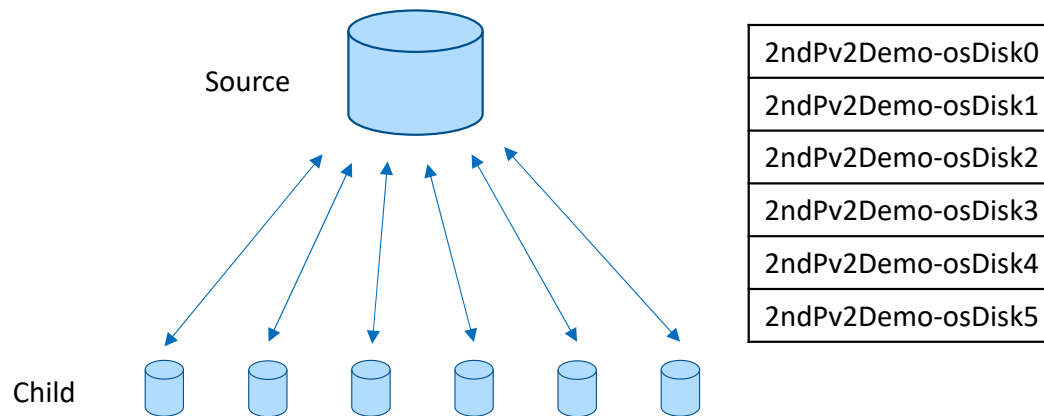


Manage the lifetime of the source CoR

- The lifetime of source CoR must be controlled
- Need a reference on the source to track the dependent children
- Challenge: If we implement a simple reference count with numeric value, how do we ensure that the add reference operation is applied exactly once?
- Consider lost packets, retries, etc in a distributed system

Design choice

- A child should only have a single reference on the source CoR
- Add reference call transactionally record the identity of the child, ensure idempotency

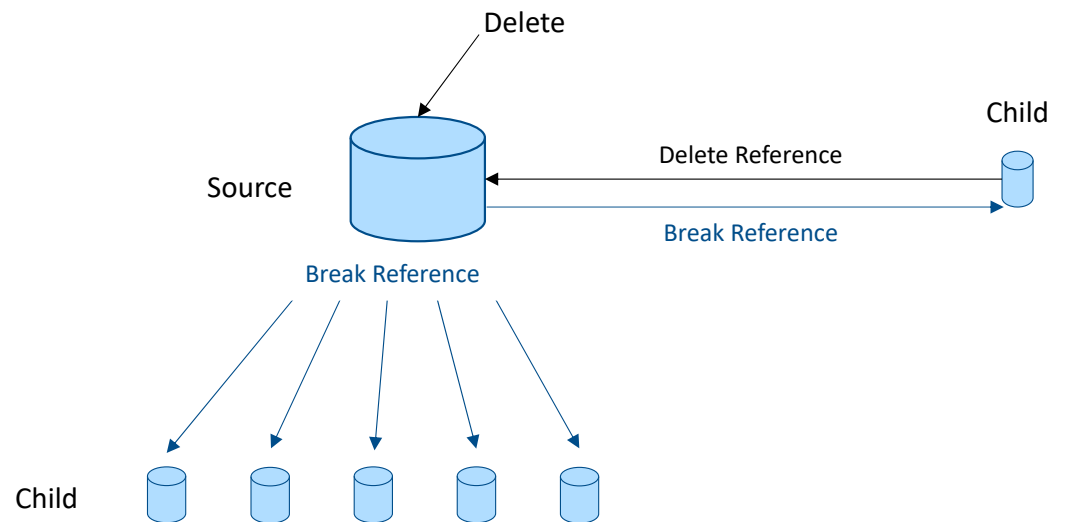


Design choice

- Order of creation of reference or object: Take reference first
- Challenge: How to handle leaked references?

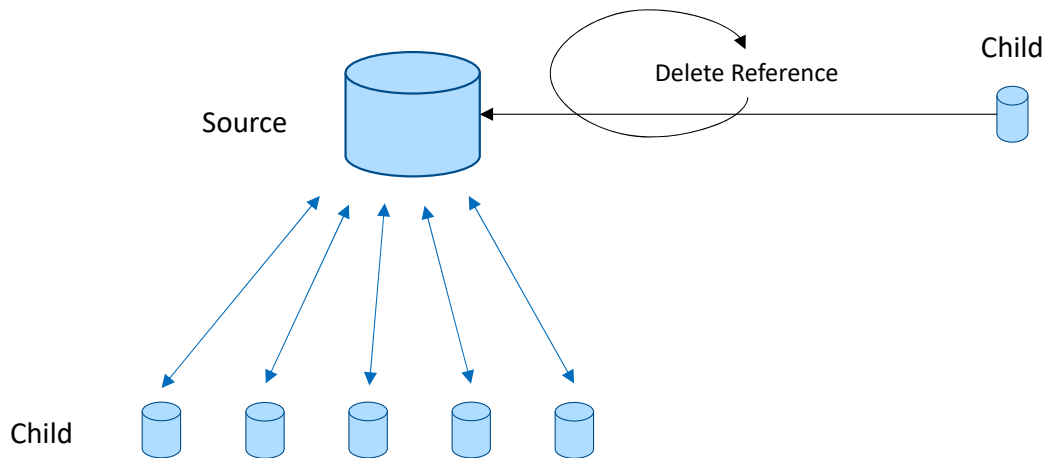
When a source receive a Delete Reference from any child or Delete call:

- Make a Break Reference call to all children
- If the call failed as reference not found, it is treated as a leaked reference



Design choice

- Order of creation of reference or object: Take reference first
- Challenge: How to handle leaked references?



When child is getting deleted:

- Make a Delete Reference call on source
- Implement a workflow to ensure Delete call is successful

Behind the scenes for Azure unique capabilities

- ✓ Instant snapshot create & restore

- ✓ Support data access over different protocol paths (SCSI, REST)

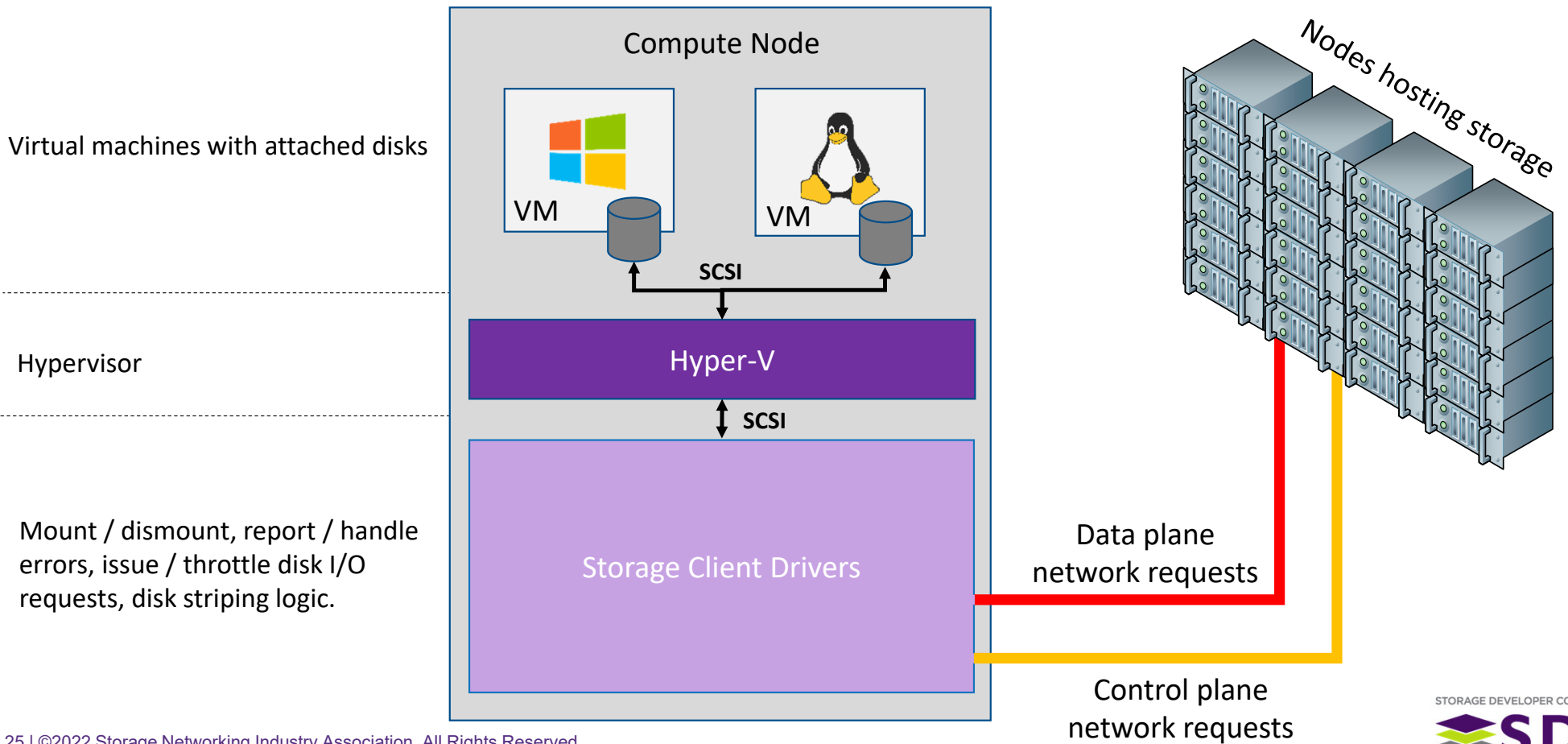
- ✓ Highly available storage with Zonal Redundancy Storage (ZRS)

Support for different protocols

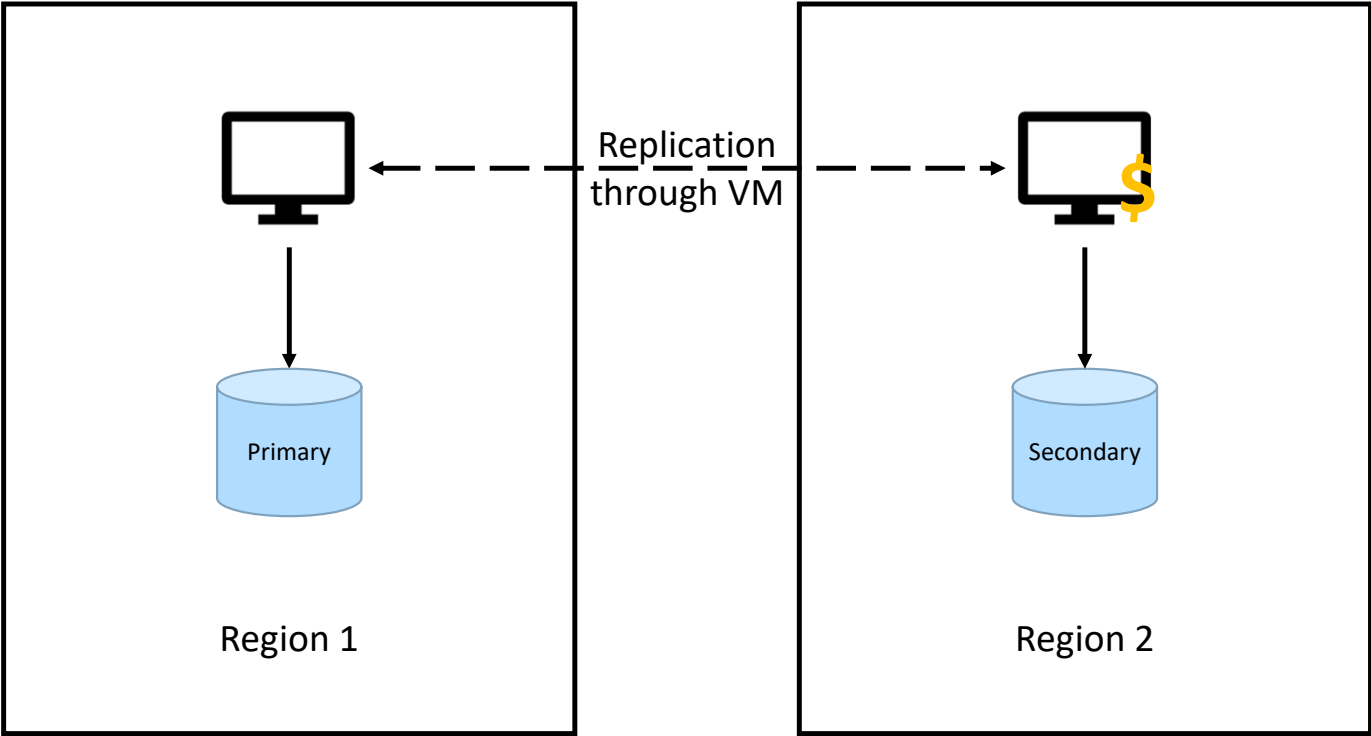
Select the protocol of choice that best suits your use case

- SCSI – Direct access from Azure Virtual Machine
- REST – Flexible data import and export
- Future – Other industry standard block storage protocols

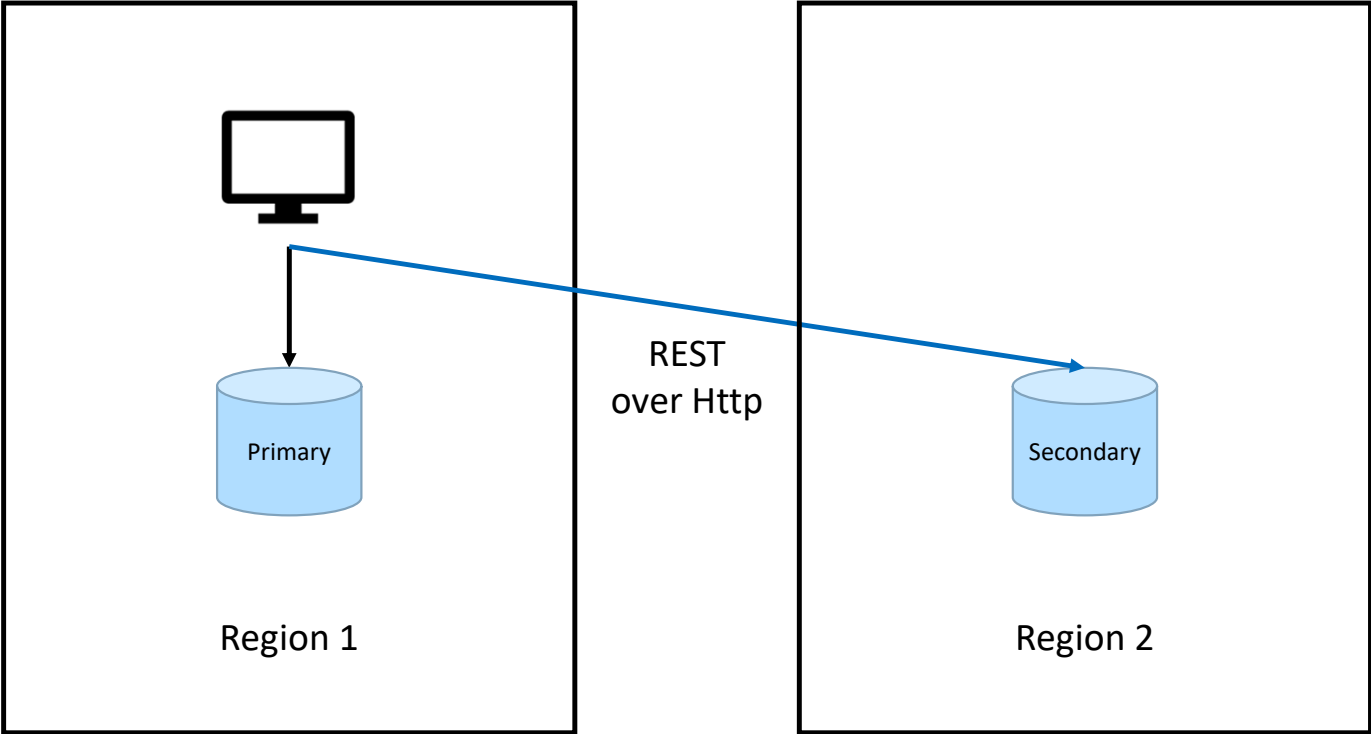
SCSI path



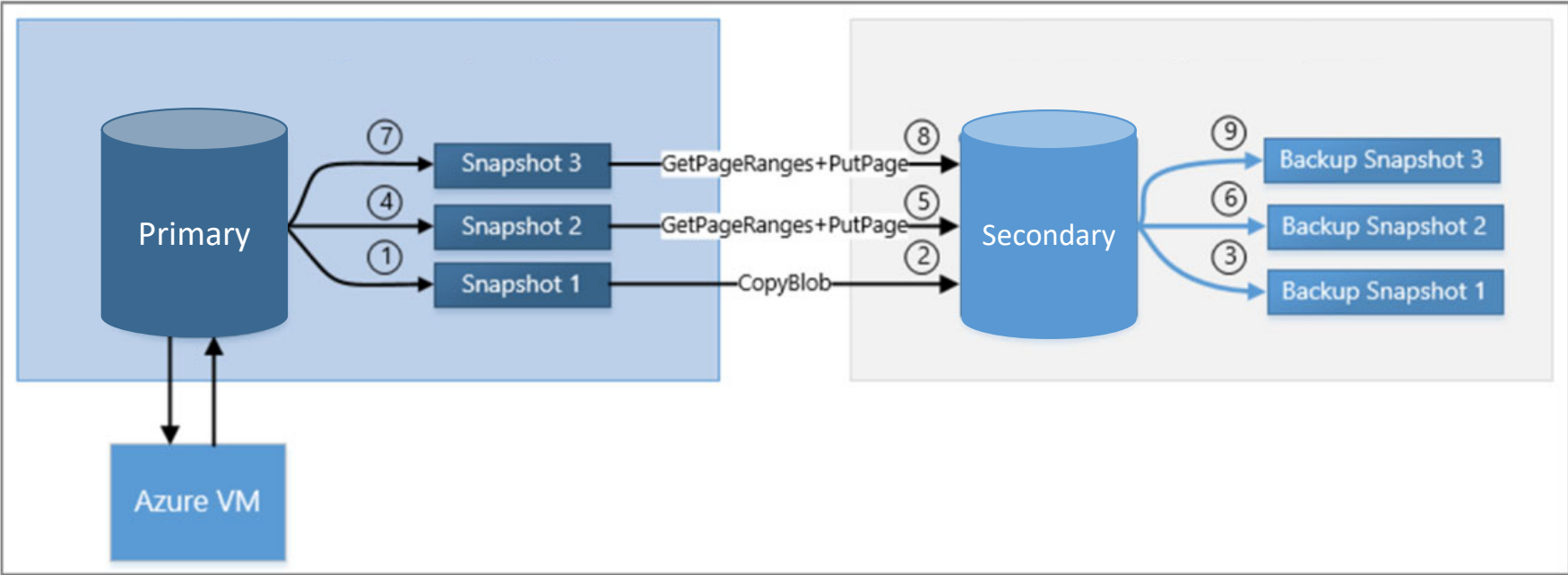
REST use case: DR replication



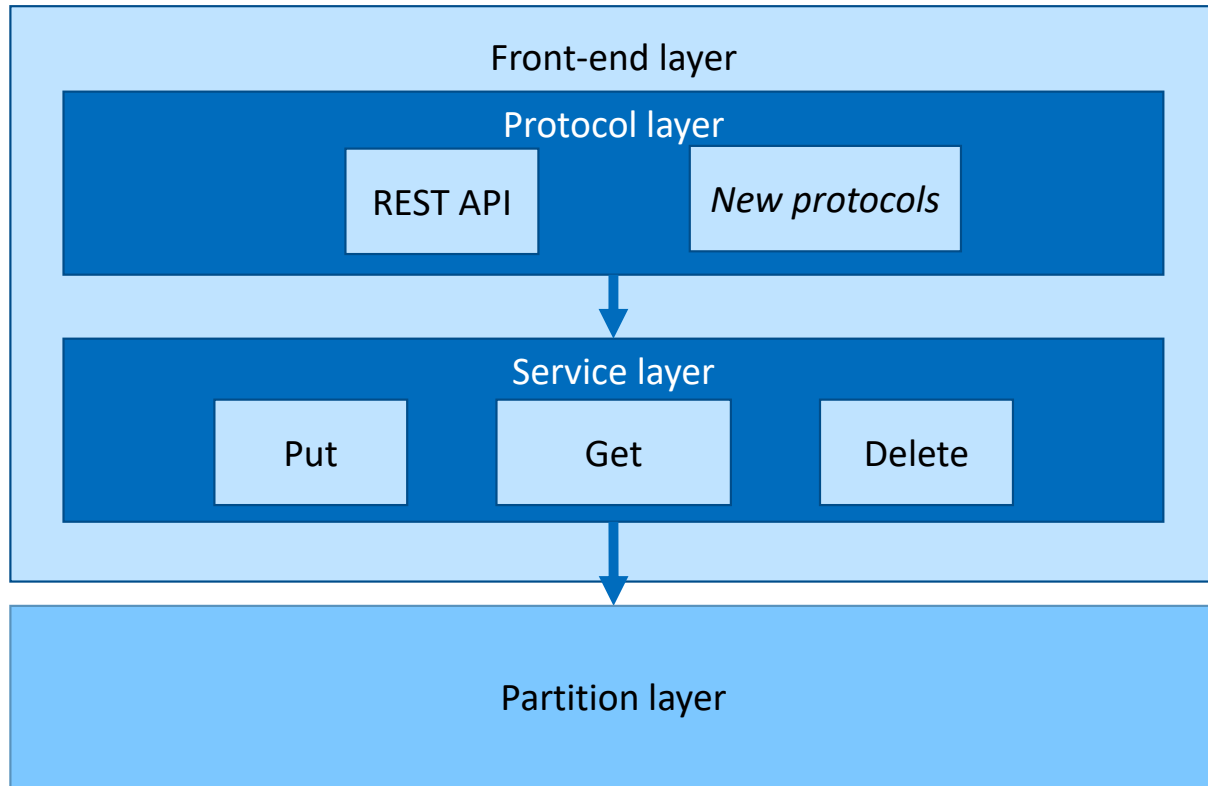
REST use case: DR replication



REST use case: DR replication



REST path



- Protocol frontend, versioning
- Authentication, throttling, logging

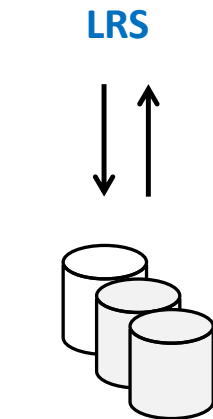
- Business logic of the storage operations

Behind the scenes for Azure unique capabilities

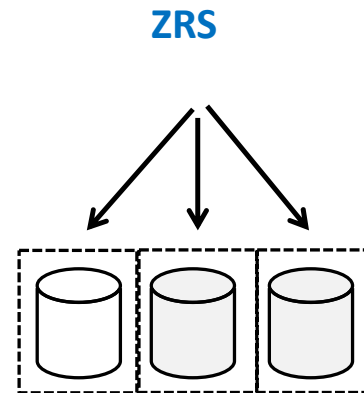
- ✓ Instant snapshot create & restore
- ✓ Support data access over different protocol paths (SCSI, REST)
- ✓ Highly available storage with Zonal Redundancy Storage (ZRS)

Azure Block Storage Redundancy

Single Region

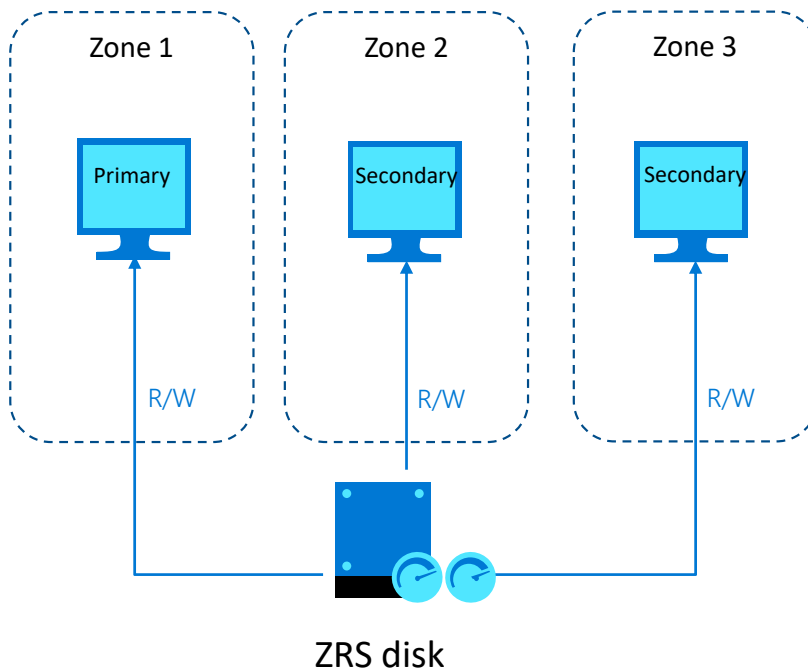


3 replicas
1 region



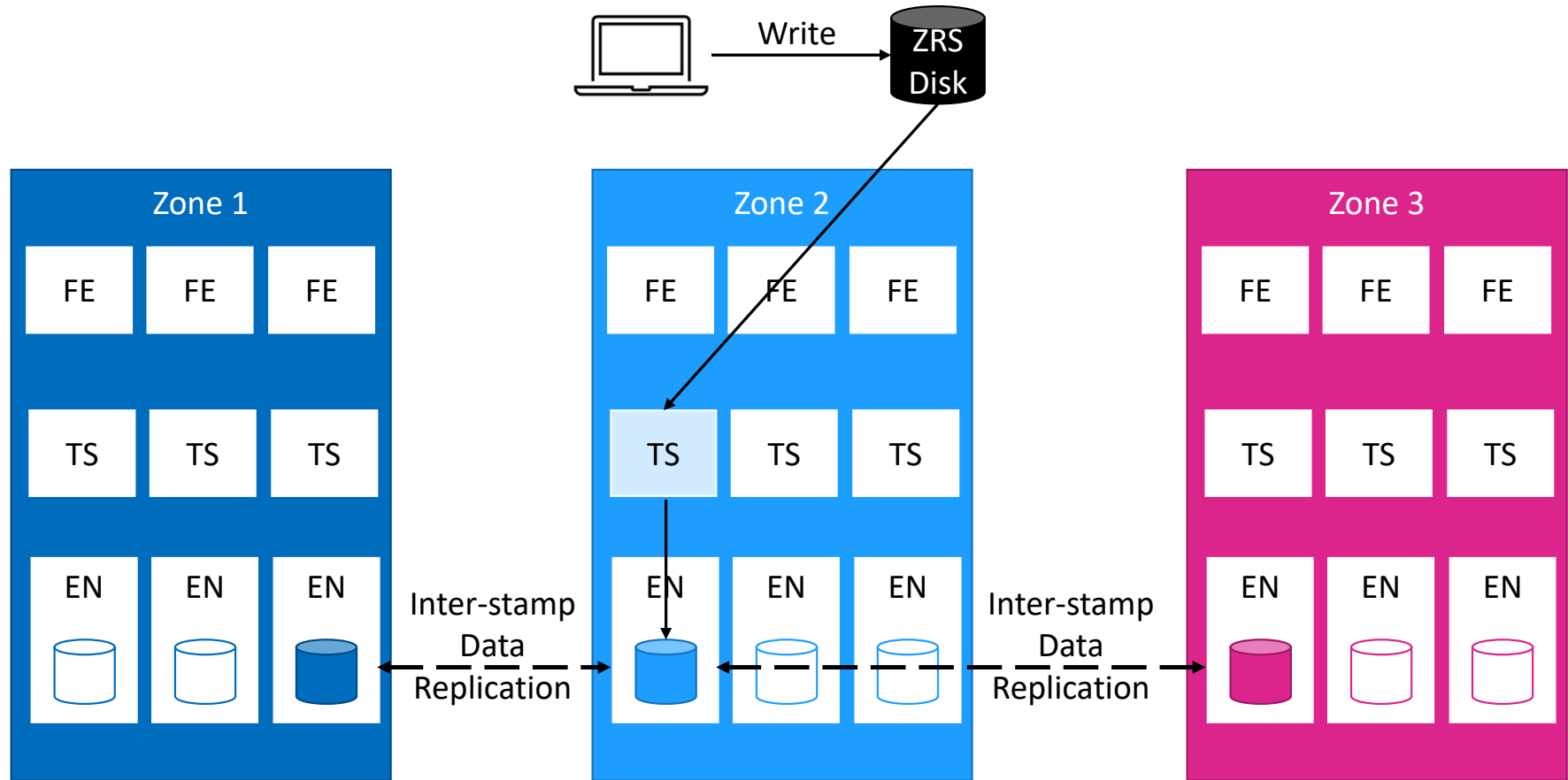
Multiple availability zones
3 replicas
1 region

ZRS use case: HA setup

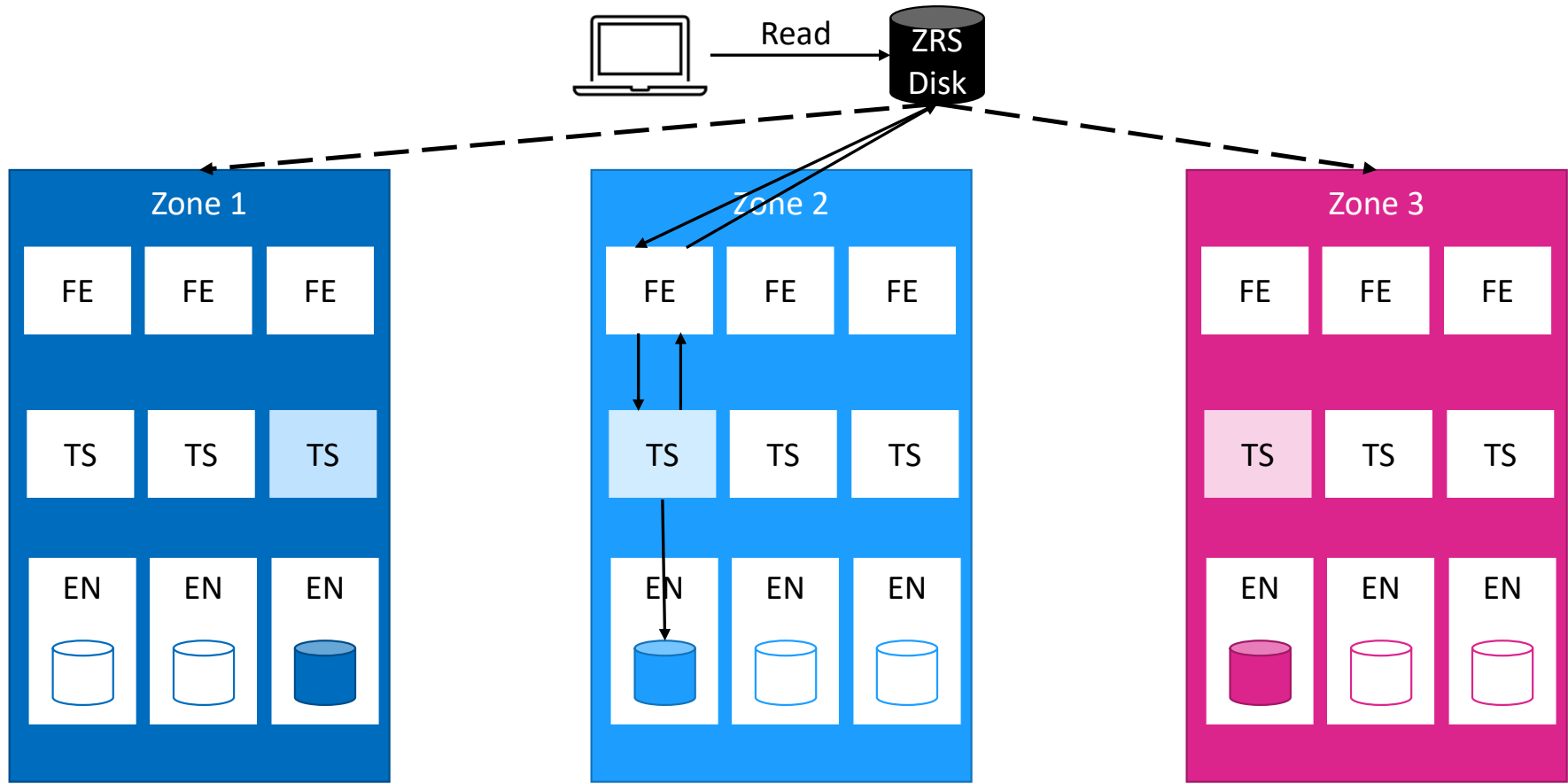


- Data replication taken care by ZRS
- Shared disk attached to primary and secondary VMs
- Fast failover with only 50% additional cost

ZRS implementation



ZRS implementation





Questions?



Please take a moment to rate this session.

Your feedback is important to us.