ceph days

CEPH DAYS Berlin

12./13. November 2025

ceph days

The Need for Speed:

Accelerating OpenStack with NVMe-oF & Ceph

Kritik Sachdeva, Technical Support Professional, IBM

Who am I?

Who am I?

- Technical Support Professional @ IBM
- Explore new technologies and integrating them with each other
- A RHCA Level-XII
- By passion I am an artist - Spend time doing & teaching Calligraphy, and paper quilling

AGENDA

For today

AGENDA

For today

- Storage Gaps in the AI Data Pipeline
- Ceph & NvmeOf
- Request flow architecture in OpenStack
- Integration Demo
- Use Case

ceph days

Storage Gaps in the AI Data Pipeline

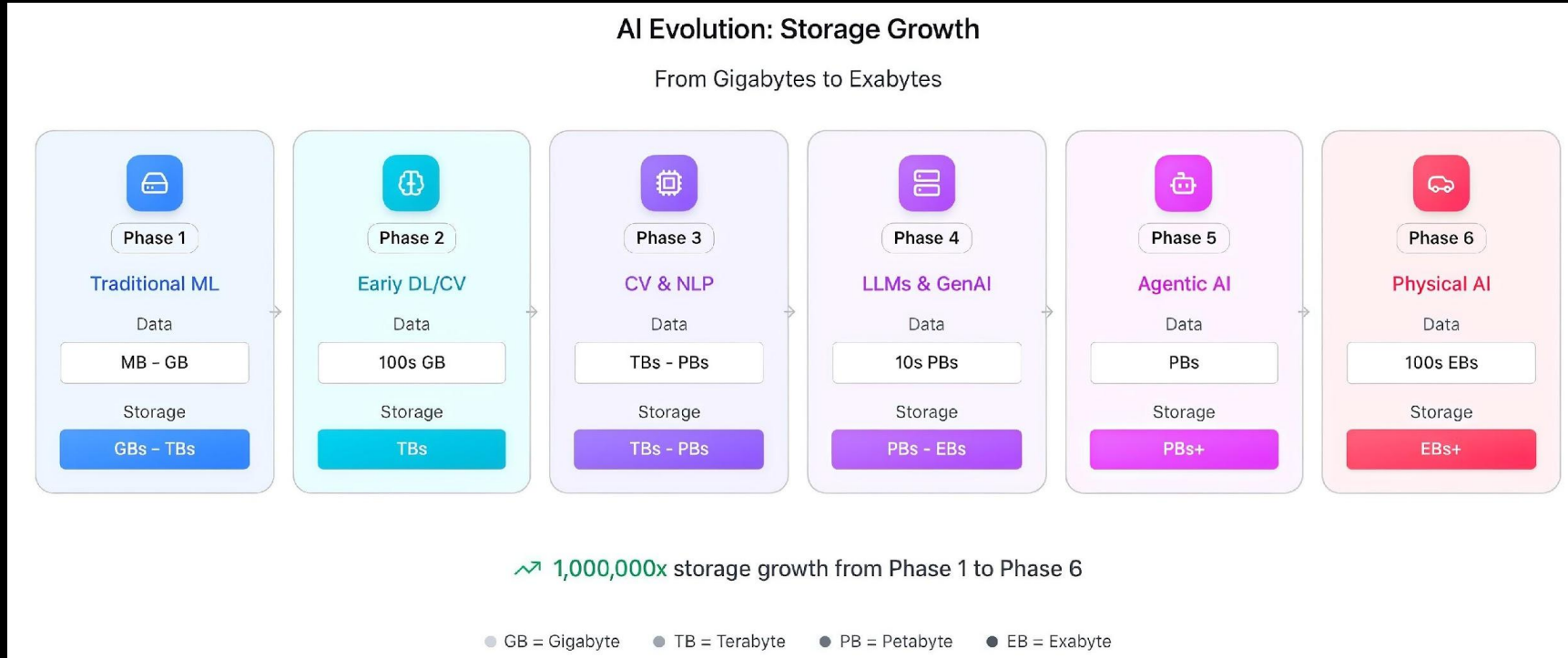
Storage Gaps in the AI Data Pipeline

As AI workloads evolve, storage systems built for throughput are struggling to meet the latency, consistency, and scalability demands of modern ML and analytics pipelines.

Storage Gaps in the AI Data Pipeline

ceph days

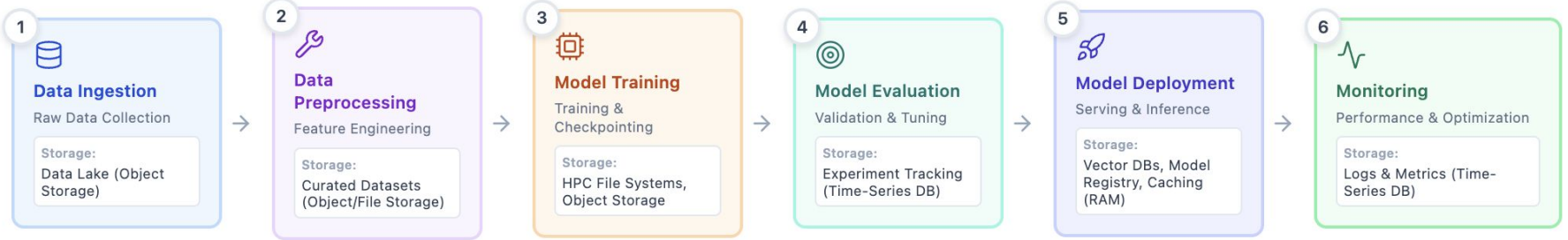
As AI workloads evolve, storage systems built for throughput are struggling to meet the latency, consistency, and scalability demands of modern ML and analytics pipelines.



Storage Gaps in the AI Data Pipeline

ceph days

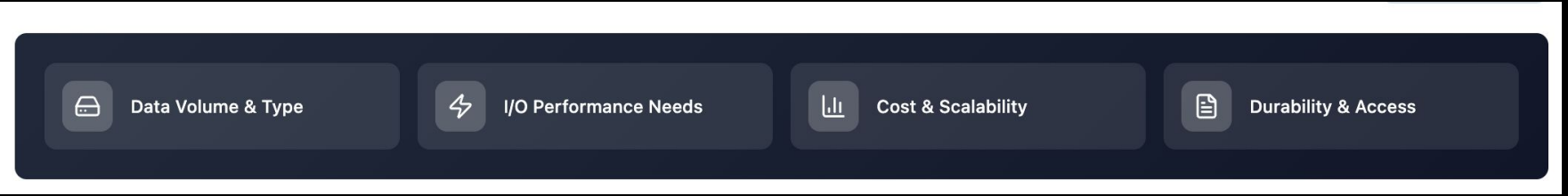
MLOps Pipeline



Storage Gaps in the AI Data Pipeline

ceph days

Key Storage Considerations



Storage Gaps in the AI Data Pipeline

Modern workloads demand more than ever from storage, but traditional protocols create a bottleneck.

- 🧠AI/ML Training: Requires rapid I/O for large datasets and frequent model checkpointing.
- 📊 Vector Databases (RAG): Demand ultra-low latency for real-time semantic search.
- 🧩 High-Transaction Databases: Performance is directly limited by storage IOPS.

Slow storage leaves expensive resources like GPUs waiting, throttling your cloud's potential.

Ceph & NVMe-oF

Ceph

A Software defined storage solution

Ceph

ceph days

A Software defined storage solution

Massively Scalable. Highly Resilient. Unified.

- **Elastic Scale:** Grow from TBs to PBs seamlessly.
- **Self-Healing:** Data integrity and availability built-in.
- **Multi-Protocol:** Block (RBD, NVMe-oF), File (CephFS, NFS), Object (S3).

Your flexible, future-proof data lake.

🚀 NVMe-oF/TCP – High-Speed Storage over Ethernet

NVMe-oF/tcp extends the blazing-fast performance of local NVMe SSDs across the network fabric.

🚀 NVMe-oF/TCP – High-Speed Storage over Ethernet

ceph days

NVMe-oF/tcp extends the blazing-fast performance of local NVMe SSDs across the network fabric.

- Extends NVMe commands over **TCP/IP** networks.
- **Simple deployment –** works on existing Ethernet infrastructure.
- Ideal for **cloud-native, and OpenStack** environments.
- Think of it as: “*iSCSI for NVMe, but faster and more efficient.*”

🚀 NVMe-oF/TCP – High-Speed Storage over Ethernet

ceph days

NVMe-oF/tcp extends the blazing-fast performance of local NVMe SSDs across the network fabric.

- **Low Latency:** Bypasses legacy SCSI command sets for a more direct, efficient data path.
- **Throughput & IOPS:** Designed to saturate modern high-speed networks (TCP & RDMA).
- **Industry Standard:** A widely adopted protocol with broad support.
- It's the modern, high-performance language for network storage.

NVMe-oF Gateway in Ceph

ceph days

A Ceph service (nvmeof-gw) that exposes Ceph RBD images as NVMe/TCP devices.

NVMe-oF Gateway in Ceph

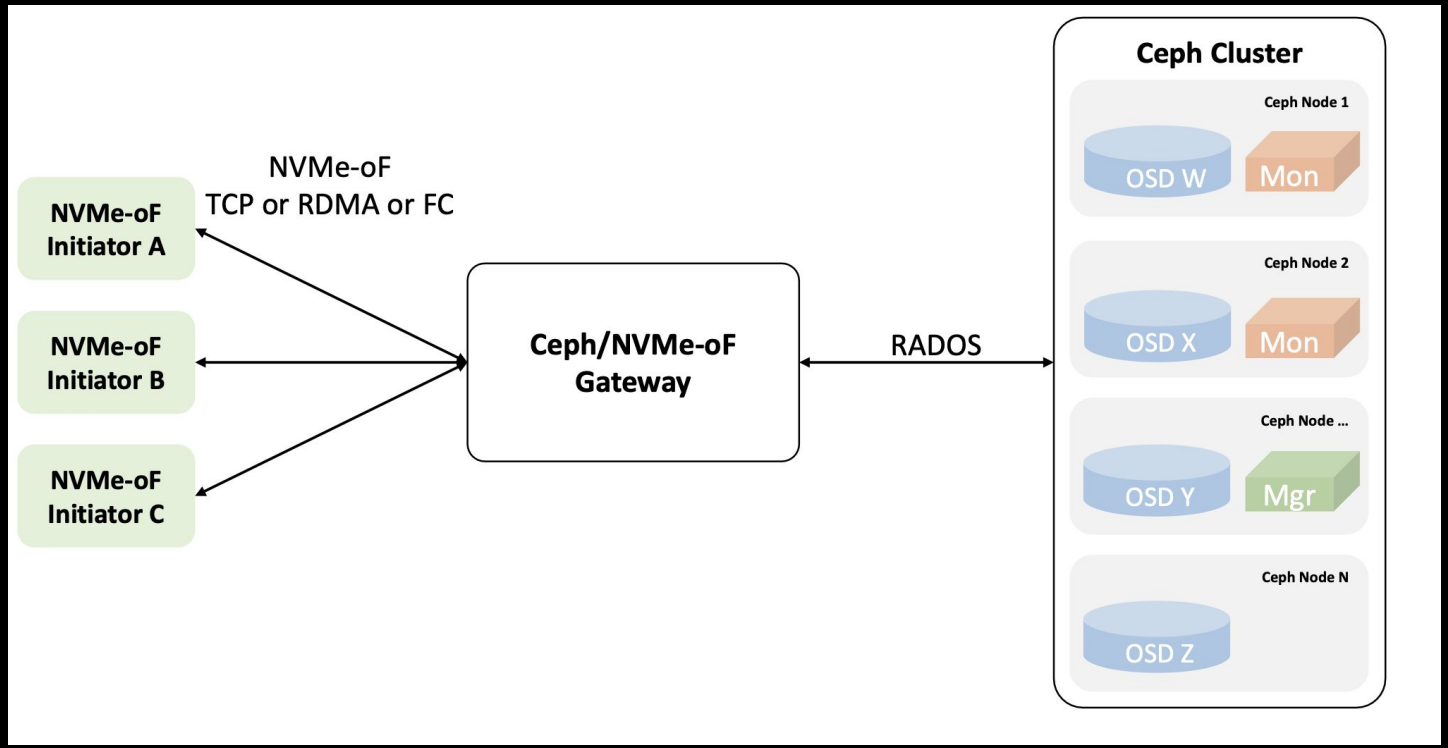
ceph days

A Ceph service (nvmeof-gw) that exposes Ceph RBD images as NVMe/TCP devices.

- Enables **block-level access** to Ceph storage over standard Ethernet networks.
- Integrated via **Ceph Orchestrator (cephadm)** for simple deployment and management.
- Both an NVMe-oF target and a Ceph client. Think of it as a “translator” between Ceph’s RBD interface and the NVME-oF protocol.
- Multi-host access & namespace isolation for scalability.
- **Dynamic configuration** via Ceph CLI or dashboard.

NVMe-oF Gateway in Ceph

A Ceph service (nvmeof-gw) that exposes Ceph RBD images as NVMe/TCP devices.



NVMe-oF Gateway in Ceph

Basic terminologies in NVMe-oF GW in Ceph

NVMe-oF Gateway in Ceph

ceph days

Basic terminologies in NVMe-oF GW in Ceph

- *Namespace*
  - NVMe equivalent to FC and iSCSI LUNs (can be thought as a volume)
  - Defined as a collection of LBAs (Logical Block Addresses)
  - In Ceph Nvme-oF GW a namespace is mapped to an RBD Image

NVMe-oF Gateway in Ceph

ceph days

Basic terminologies in NVMe-oF GW in Ceph

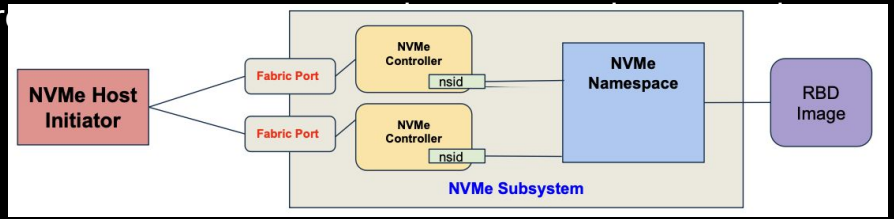
- *Namespace*
  - NVMe equivalent to FC and iSCSI LUNs (can be thought as a volume)
  - Defined as a collection of LBAs (Logical Block Addresses)
  - In Ceph Nvme-oF GW a namespace is mapped to an RBD Image
- *NVMe Subsystem*
  - An entity that contains Namespaces and the other NVMe elements such as NVMe controllers
  - Identify by NQN (NVMe Qualified Name) and unique access controls to which the Initiator connects to target IP/NVMe Subsystem

NVMe-oF Gateway in Ceph

ceph days

Basic terminologies in NVMe-oF GW in Ceph

- *Namespace*
  - NVMe equivalent to FC and iSCSI LUNs (can be thought as a volume)
  - Defined as a collection of LBAs (Logical Block Addresses)
  - In Ceph Nvme-oF GW a namespace is mapped to an RBD Image
- *NVMe Subsystem*
  - An entity that contains Namespaces and the other NVMe elements such as NVMe controllers
  - Identify by NQN (NVMe Qualified Name) and unique access controls to which the Initiator connects to target IP/NVMe Subsystem
- *NVMe IO Controller*
  - Created for every connection between a host and a target NVME-oF fabric address per Subsystem.
  - Receives and pr



NVMe-oF Gateway in Ceph

ceph days

Basic terminologies in NVMe-oF GW in Ceph

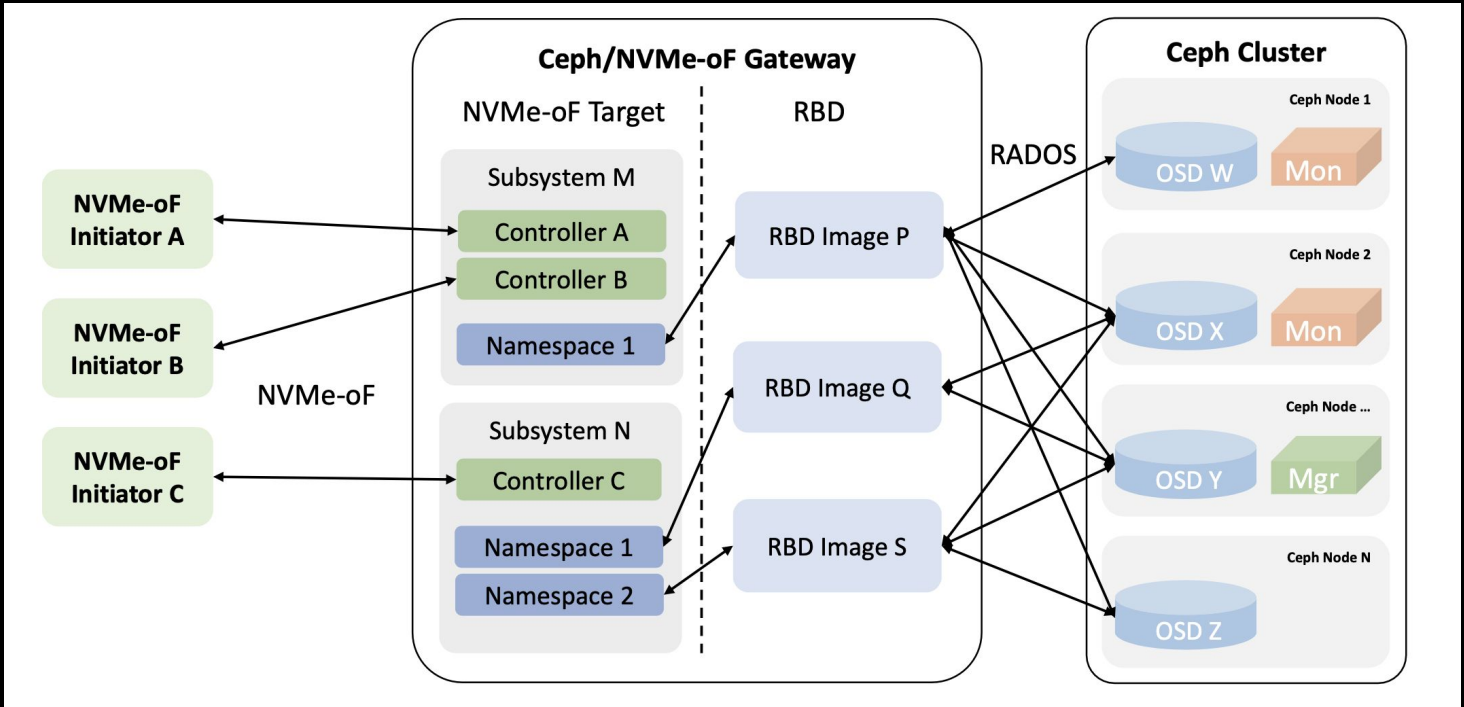
- *Listener/Transport Address*
  - The network interface (IP) and service ID (port) where the gateway listens for NVMe/TCP connections
  - A listener binds a subsystem to a reachable address.
- *Gateway Group / HA domain*
  - A logical group of gateway instances that provide redundancy and failover.
  - If one gateway fails, another in the group can take over I/O for a namespace (active/standby mode)

NVMe-oF Gateway in Ceph

Architecture and components

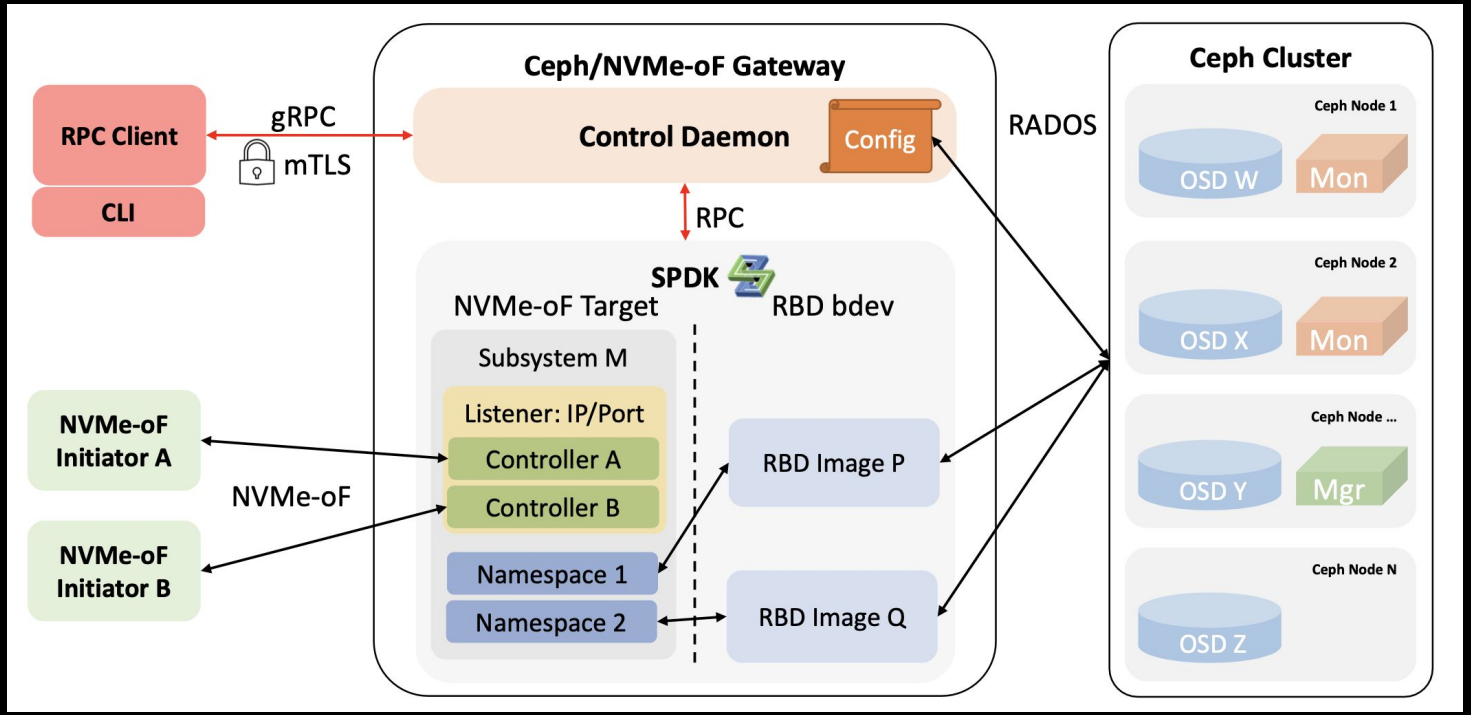
NVMe-oF Gateway in Ceph

Architecture and components



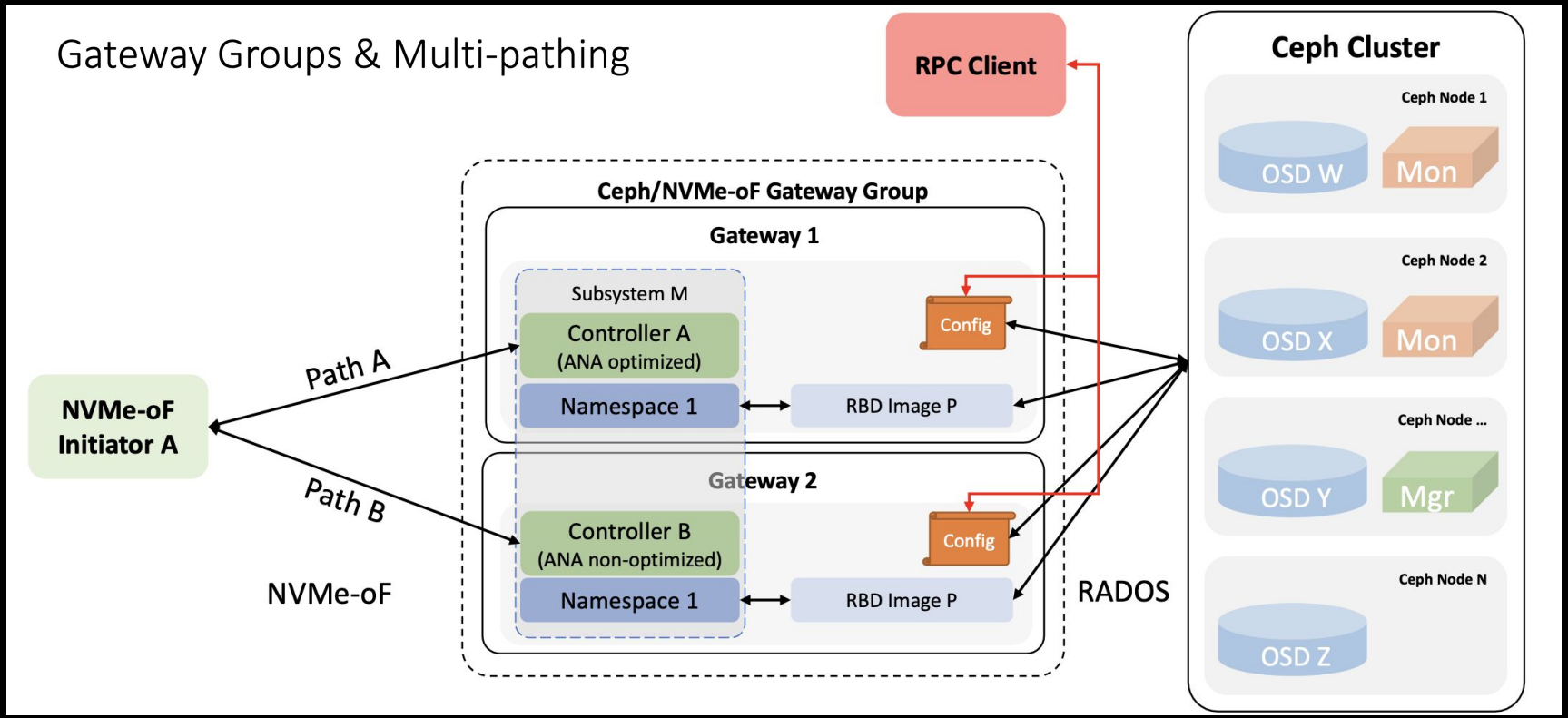
NVMe-oF Gateway in Ceph

Architecture and components



NVMe-oF Gateway in Ceph

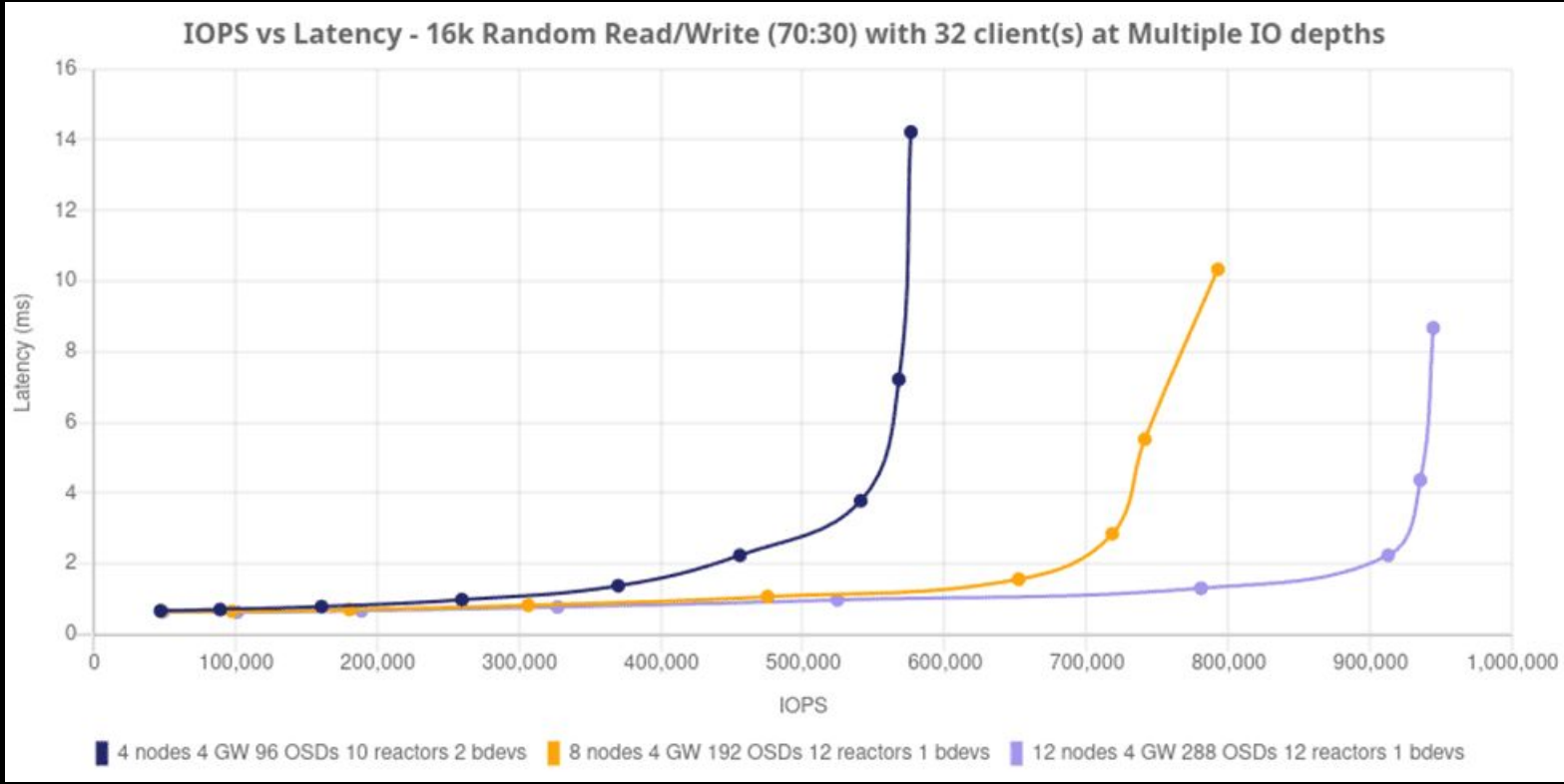
Architecture and components



NVMe-oF Gateway in Ceph

ceph days

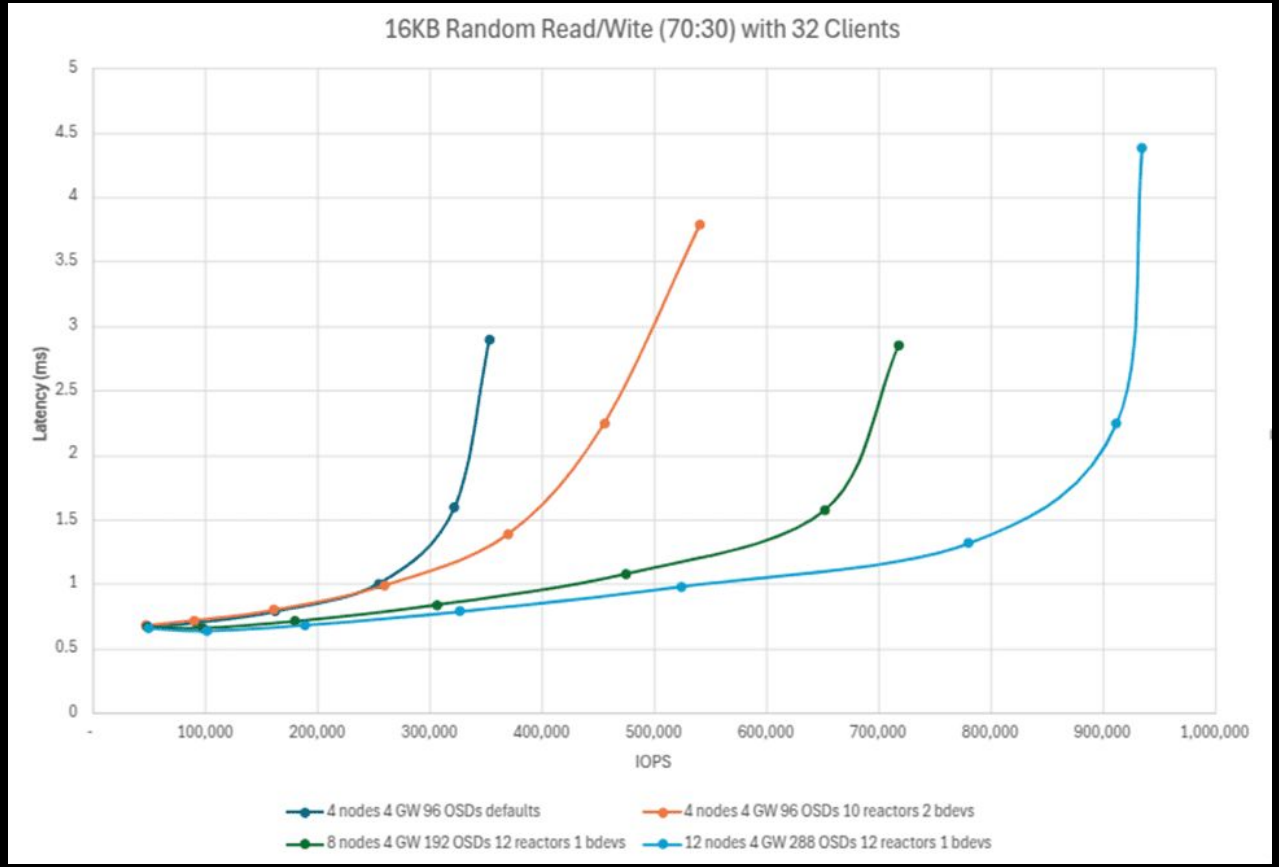
Performance



NVMe-oF Gateway in Ceph

ceph days

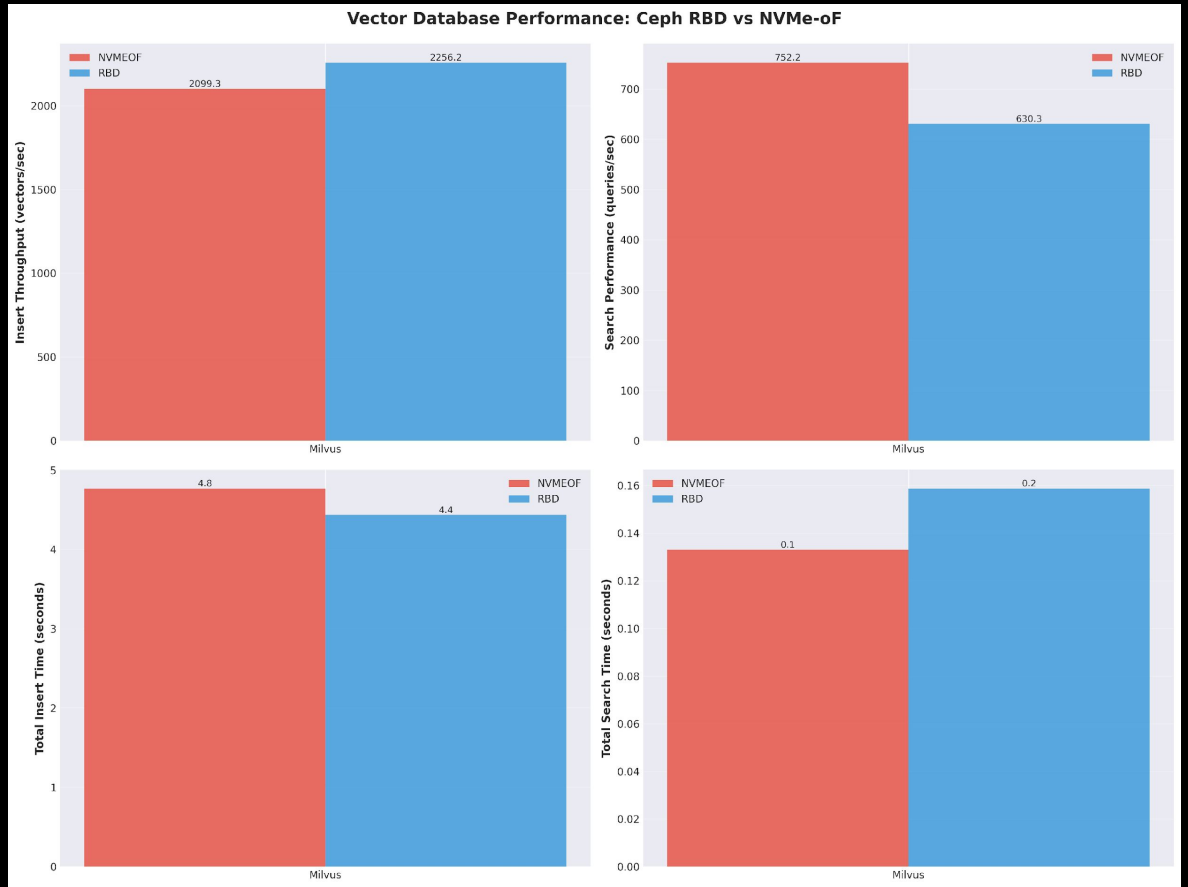
Performance



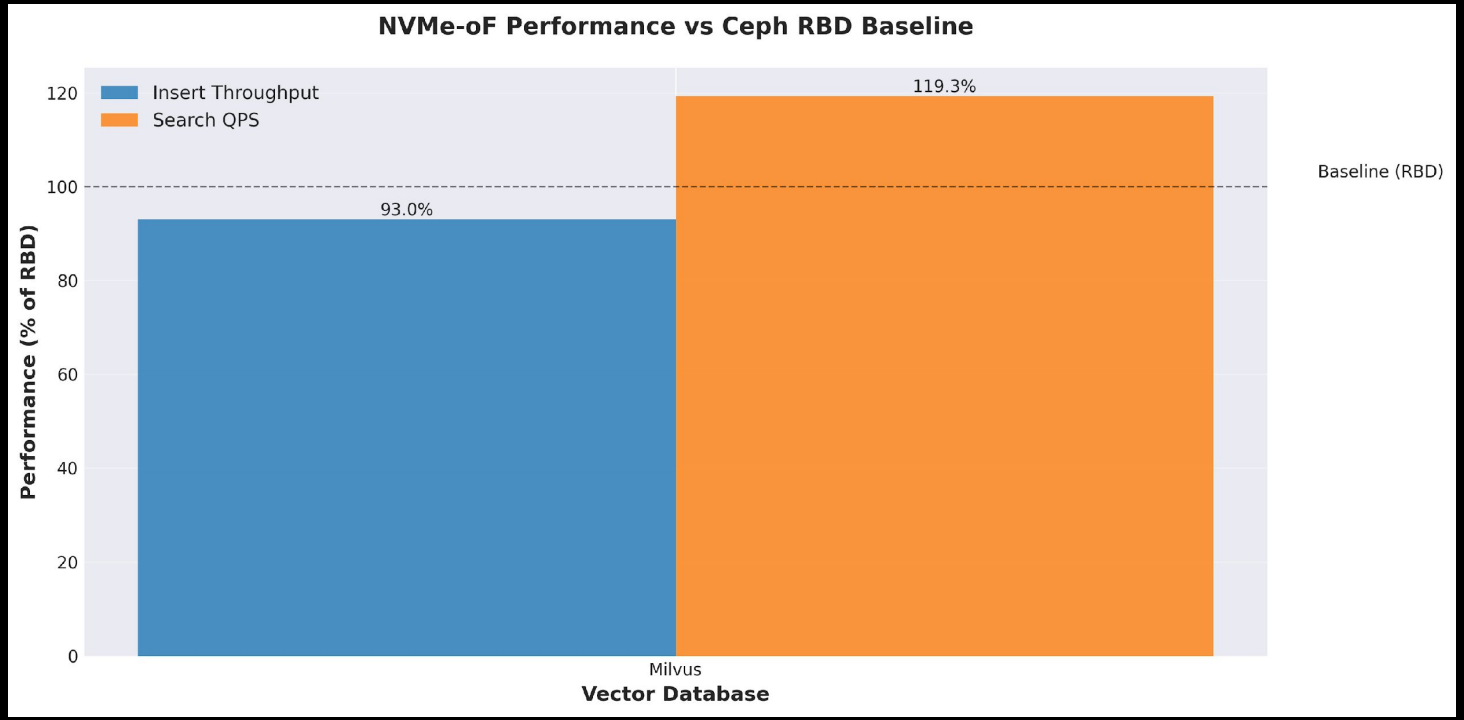
ceph days

Performance Comparison Glimpse

Performance Comparison Glimpse



Performance Comparison Glimpse



Performance Comparison Glimpse

ceph days

Benchmark Results Summary

Database	Storage	Insert Throughput (vec/s)	Search QPS	Insert Time (s)	Search Time (s)
Milvus	NVMEOF	2099.26	752.19	4.76	0.13
Milvus	RBD	2256.17	630.29	4.43	0.16

Connecting OpenStack Volumes through Ceph NVMe-oF

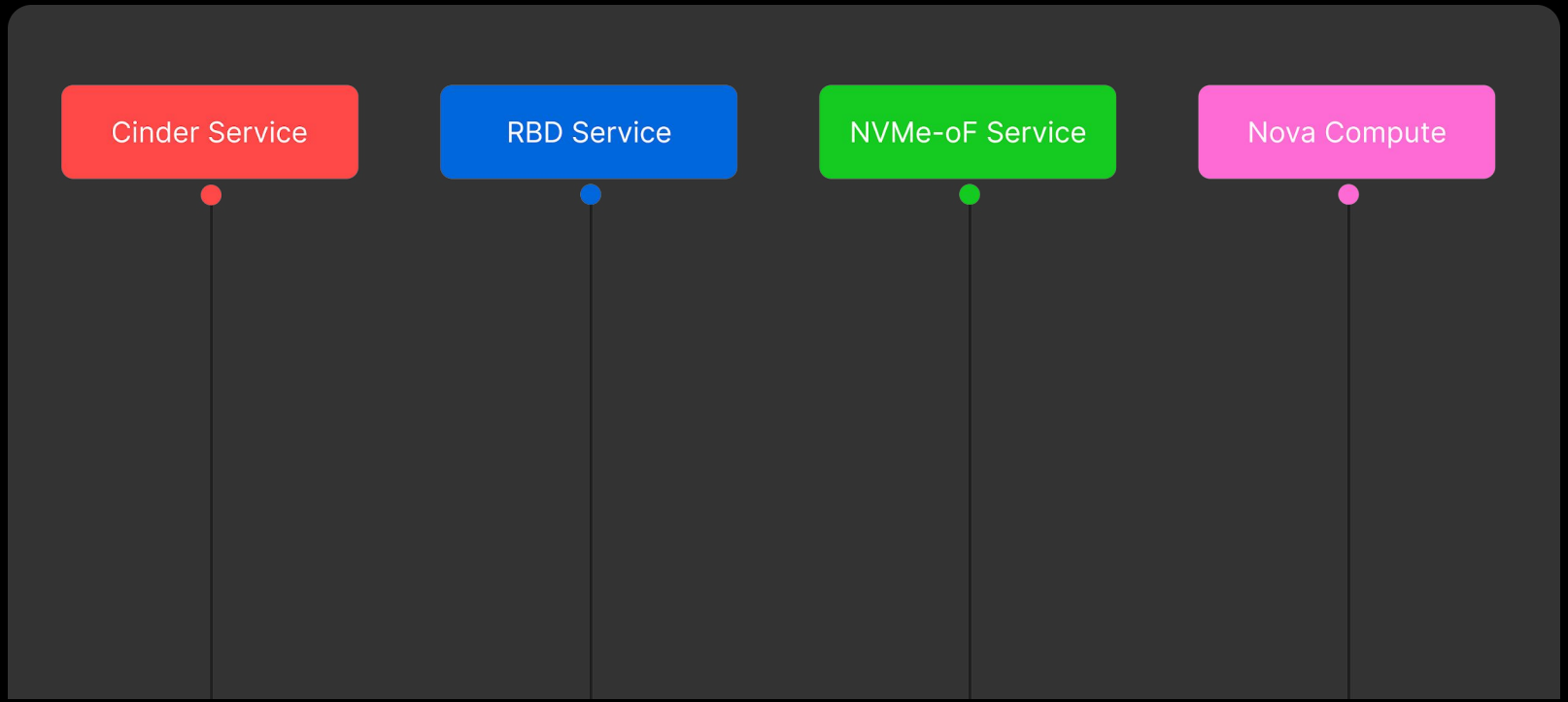
Connecting OpenStack Volumes through Ceph NVMe-oF

This diagram illustrates how OpenStack services (Cinder, Nova) interact with Ceph’s RBD and NVMe-oF Gateway to provision and attach block volumes over TCP.

Connecting OpenStack Volumes through Ceph NVMe-oF

ceph days

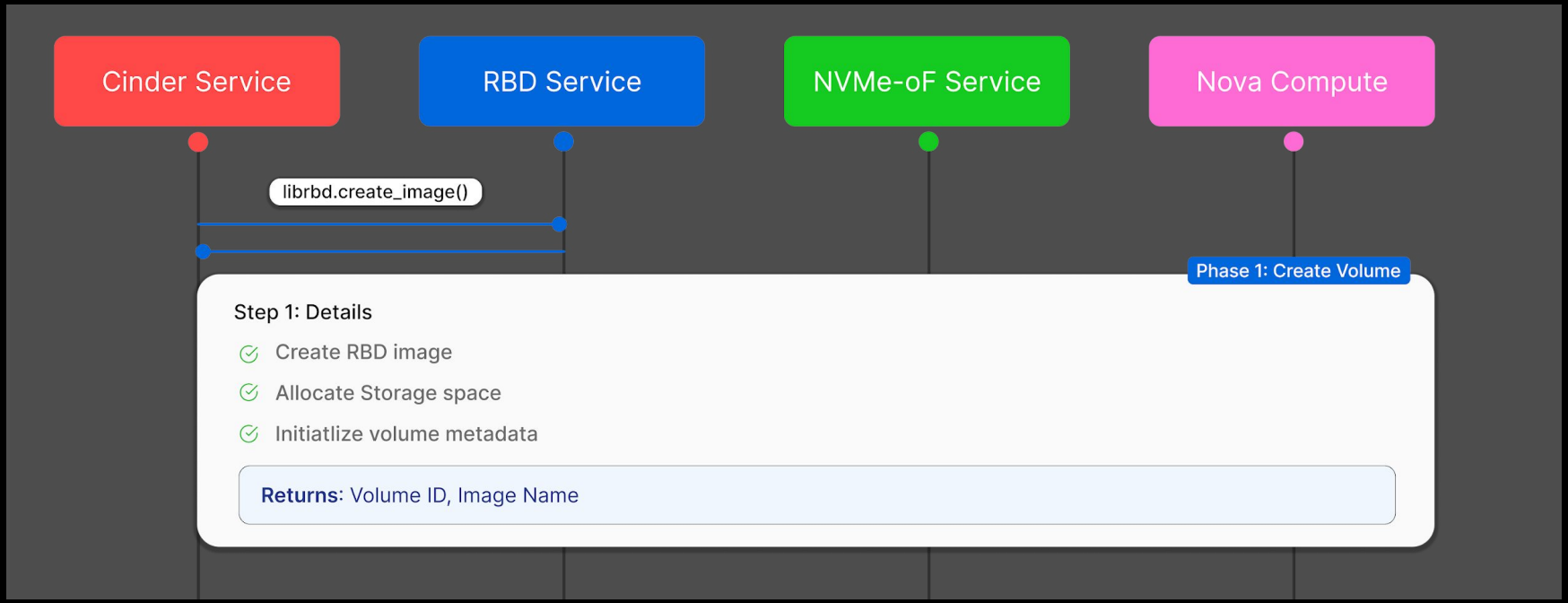
Services interact with each other in this op request flow



Connecting OpenStack Volumes through Ceph NVMe-oF

ceph days

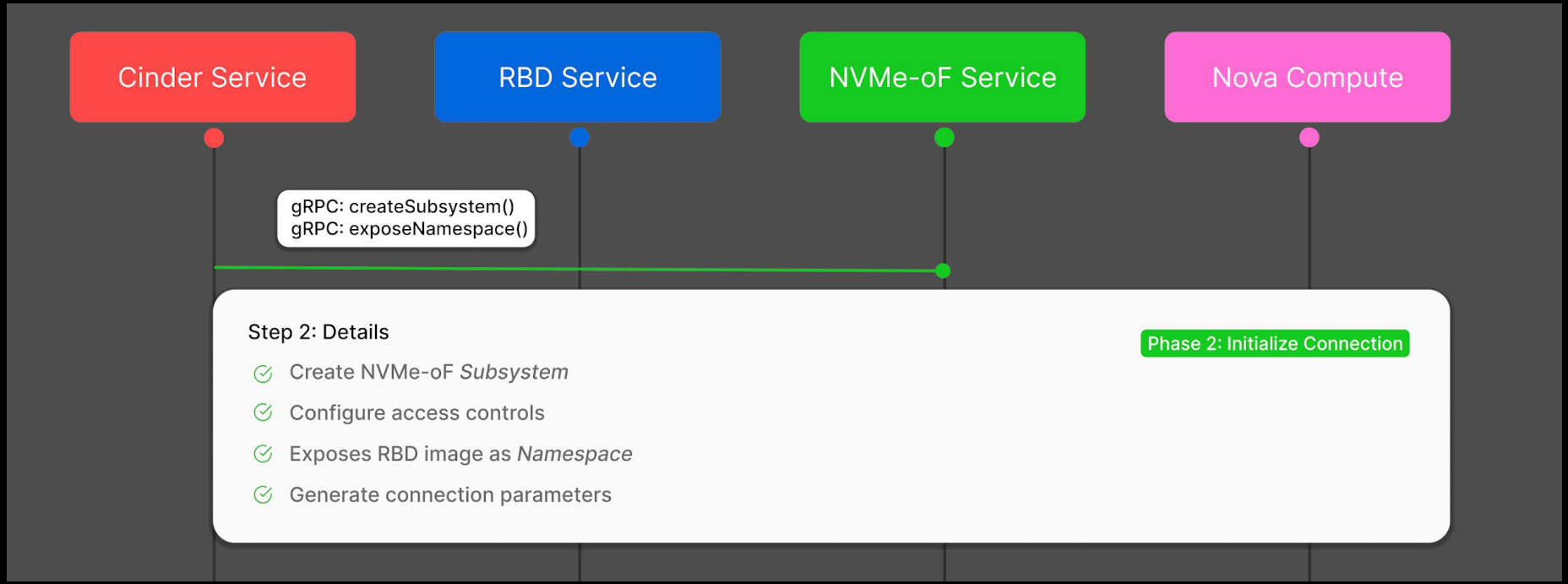
Cinder Volume create request will create an RBD image using librbd module



Connecting OpenStack Volumes through Ceph NVMe-oF

ceph days

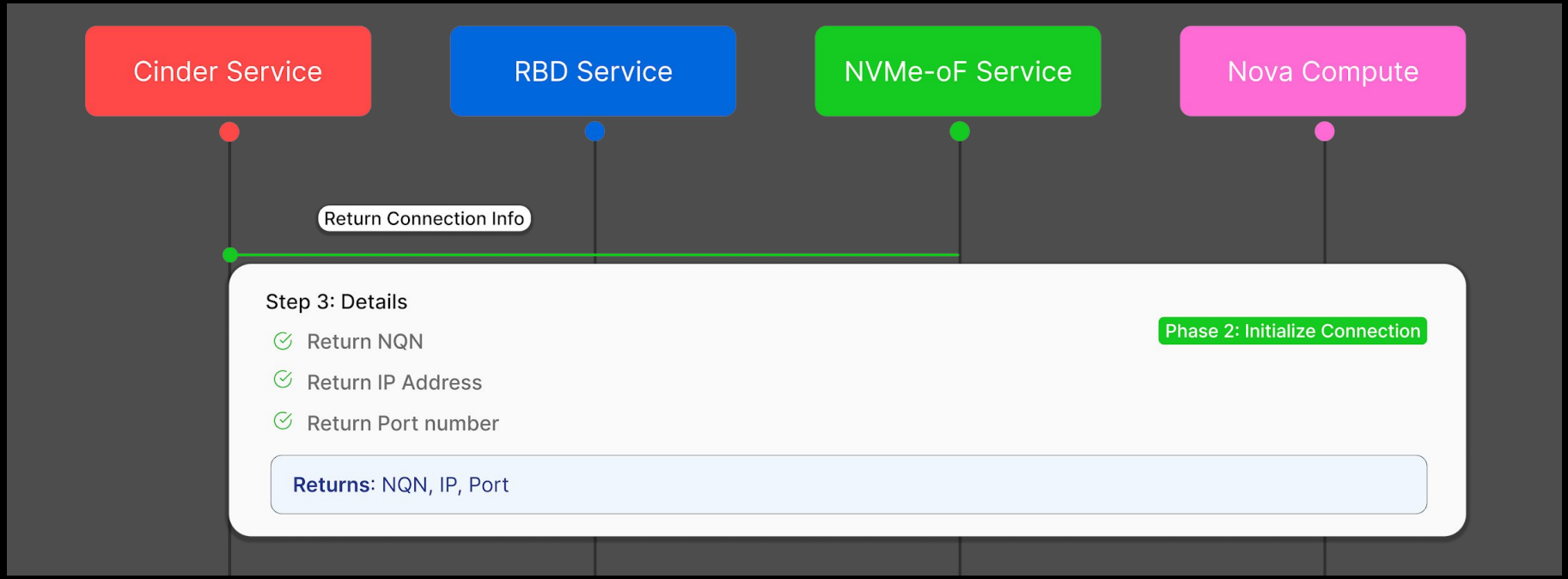
Cinder module - will create create_connection() request which will create relevant resources to attach the volume to a instance



Connecting OpenStack Volumes through Ceph NVMe-oF

ceph days

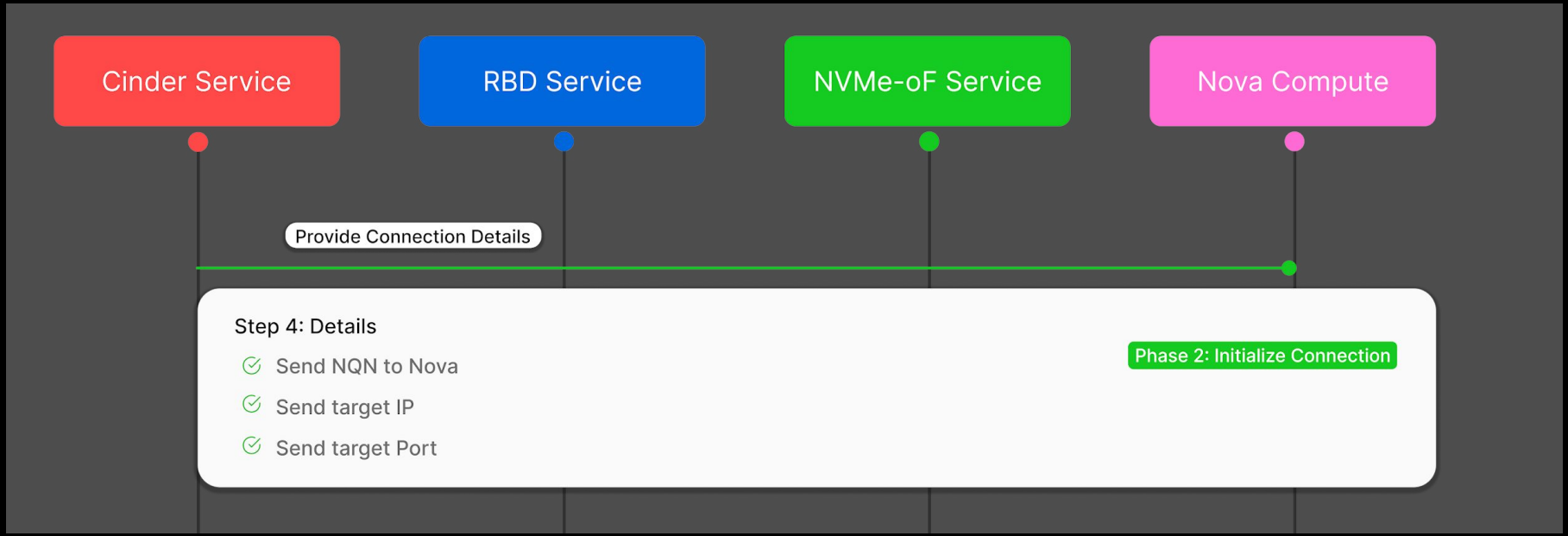
Cinder module receive the connection details from NVMeOF GW in Ceph to connect to



Connecting OpenStack Volumes through Ceph NVMe-oF

ceph days

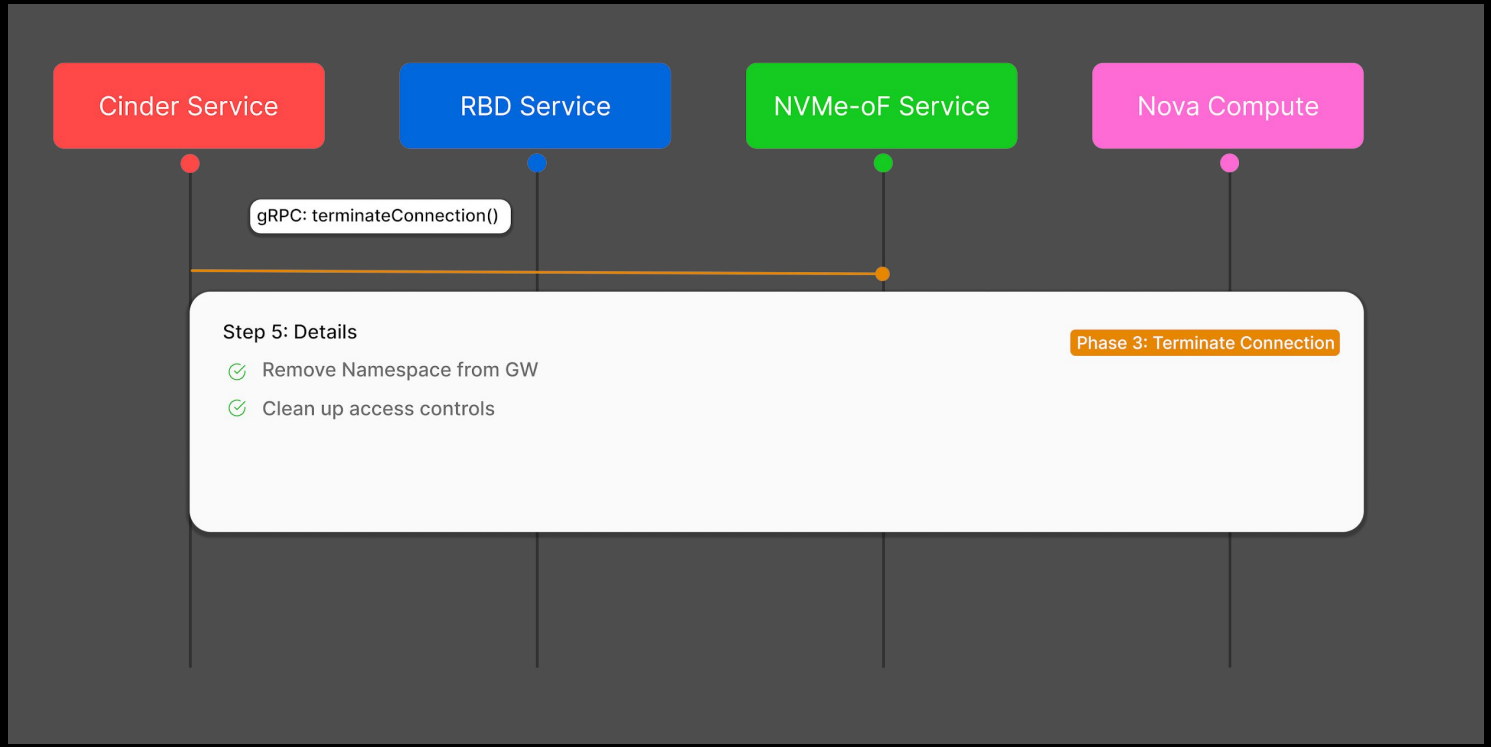
Cinder module provide the connection details received from NVMeOF GW in Ceph to Nova Service



Connecting OpenStack Volumes through Ceph NVMe-oF

ceph days

Cinder module - once volume is removed it will destroy all of the resources of NVMeOf service created earlier



ceph days

Integration Demo

Link:

Use case(s)

Use Case: 1

Powering Private Cloud Vector Databases

Use Case: 1

ceph days

Powering Private Cloud Vector Databases

Problem: Vector databases for RAG and AI search must perform similarity searches across millions of vectors in milliseconds. This requires extremely low-latency storage.

Solution: Run a vector DB on an OpenStack VM, using a Cinder NVMe-oF volume for its indexes.

Benefit:

- **Real-time Performance:** Achieve the low latency needed for interactive AI applications.
- **Scalability & Sovereignty:** Scale your database on your secure, private Ceph cluster.

🏦 Use Case: 2

ceph days

High-Performance Databases

🏦 Use Case: 2

ceph days

High-Performance Databases

Fueling Mission-Critical Applications

- **Challenge:** Traditional storage limits database transaction rates and query response times.
- **Solution:** Host database files & transaction logs on Cinder NVMe-oF volumes.
- **Impact:**
  - Higher Transactions Per Second (TPS).
  - Significantly improved query performance.

ceph days

✨ Conclusion: The Future is Fast

✨ Conclusion: The Future is Fast

- Eliminate storage bottlenecks for demanding workloads.
- Leverage Ceph's robust, distributed architecture.
- Bring cutting-edge performance to your cloud users.

Turn your OpenStack cloud into a serious data-processing engine.

ceph days

Thank you..

Any Questions

Linkedln
@kritiksachdeva

Github
@skritik098

Instagram
@theflourishart

AGENDA - Day 1 (1/2)

ceph days

9am	Welcome, Check-in, Coffee
10am	Welcome Session (Joachim Kraftmayer & Desy Management)
10.15am	Morning Keynote: Cephfs Home @ DESY (Ingo Ebel)
10.50am	Buzzword Bingo: Digital Sovereignty (Markus Wendland & Heiko Krämer)
11.25am	SURFin' the Ceph wave (Jean-Marie de Boer)
11.35am	*Coffee Break & Networking*
12.15pm	It's all about the latency, not the bandwidth! (Wido den Hollander)
12.50pm	Running a small Openstack Cluster with a full NVMe Ceph Cluster (Kevin Honka)
1.20pm	*Lunch Break & Networking*

AGENDA - Day 1 (2/2)

ceph days

2.5pm	Principles for Storage Management (Benedikt Bürk)
3.25pm	Beyond Backup: S3 Data Management with Ceph RGW Tiering, and Chorus (Sirisha Guduru & Artem Torubarov)
5.55pm	*Coffee Break & Networking*
4.35pm	How RGW Stores S3 Objects in Rados (Tobias Brunnwieser)
5.10pm	A generic Ceph sizer for the community optimizing workloads, layouts, and server configs (Matthias Münch)
5.25pm	AI, ML, and the Ceph Advantage: Scalable Storage for Smarter Workflows (Kenneth Tan)
6pm	*Network Reception in Ground Floor with Food Truck & Drinks*

AGENDA - Day 2

ceph days

9.30am	*Welcome & Good Morning Coffee*
10am	**Keynote**
10.35am	Scale Multiple Ceph-clusters Horizontally (Ansgar Jazdzewski)
11.20am	*Coffee Break & Networking*
11.10am	Ceph Rados Gateway as an Interface to Tape Storage (Stuart Hardy & Zaid Bester)
11.45am	The Need for Speed: Accelerating OpenStack with NVMe-oF & Ceph (Kritik Sachdeva)
12.20pm	Speed up your deployments using the Ansible Cephadm collection (Piotr Parczewski)
1.05pm	*Lunch Break & Networking*
2.15pm	Faster CephFS Mirroring with Bounded-Frontier Concurrency (Md Mahamudur Rahaman Sajib)
2.30pm	Ceph-CSI Support for AES-GCM: Challenges and Opportunities (David Mohren)
3pm	*Podium Discussion with Speakers, Wrap Up & Summarize*