



OCP

FUTURE TECHNOLOGIES SYMPOSIUM

OCP Global Summit

November 8, 2021 | San Jose, CA

DSS: High I/O Bandwidth Disaggregated Object Storage System for AI Applications

Mahsa Bayati

Memory Solution Lab Samsung Semiconductor Inc. San Jose, CA

Mahsa.b@samsung.com

Memory Solution Lab Group

Harsh Roogi, Somnath Roy, Ron Lee

h.roogi@samsung.com, som.roy@samsung.com, r2.lee@samsung.com

Outline

- Motivation & Introduction
- Architectural Design
 - Storage server
 - Network server
 - Client server
- Experimental Design
- Conclusions and Future work

Motivation

- Exponential data generation demands *high-bandwidth* and *scalability* storage
 - Data intensive Applications like AI, Deep Learning
- Large amount of generated data is unstructured and object format


Samsung DSS Storage Solution

Serves AI/ Deep learning applications with disaggregated, high-bandwidth, easily scaled, object storage(key-value)

Introduction

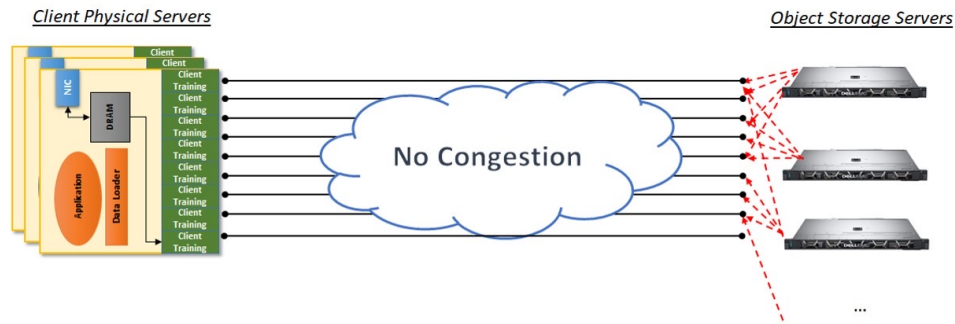
What is DSS Storage?

Disaggregated Storage software

- Implement object Key-Value API on top of NVMe over Fabric (NVMeOF)
- Design explicitly for storing object format
- Support storage remote access protocols (i.e., RDMA), same attributes as NVMeOF
- facilitate disaggregation of storage and computational resources  scalability

Introduction

Architectural Overview

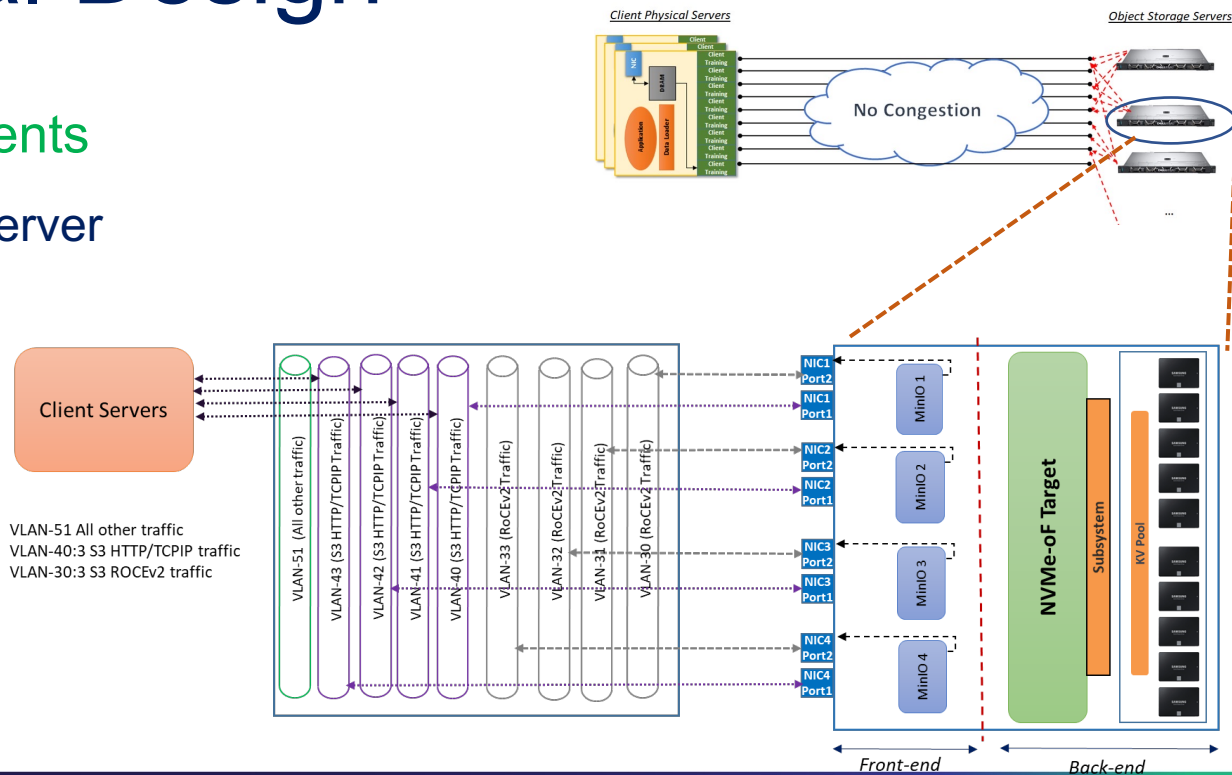


- Besides supporting object storage and scalability, DSS solution can provision the bandwidth demands for each application running on each client server.
- Multiple client session accessing storage → bandwidth inconsistency, congestion

Architectural Design

Architectural components

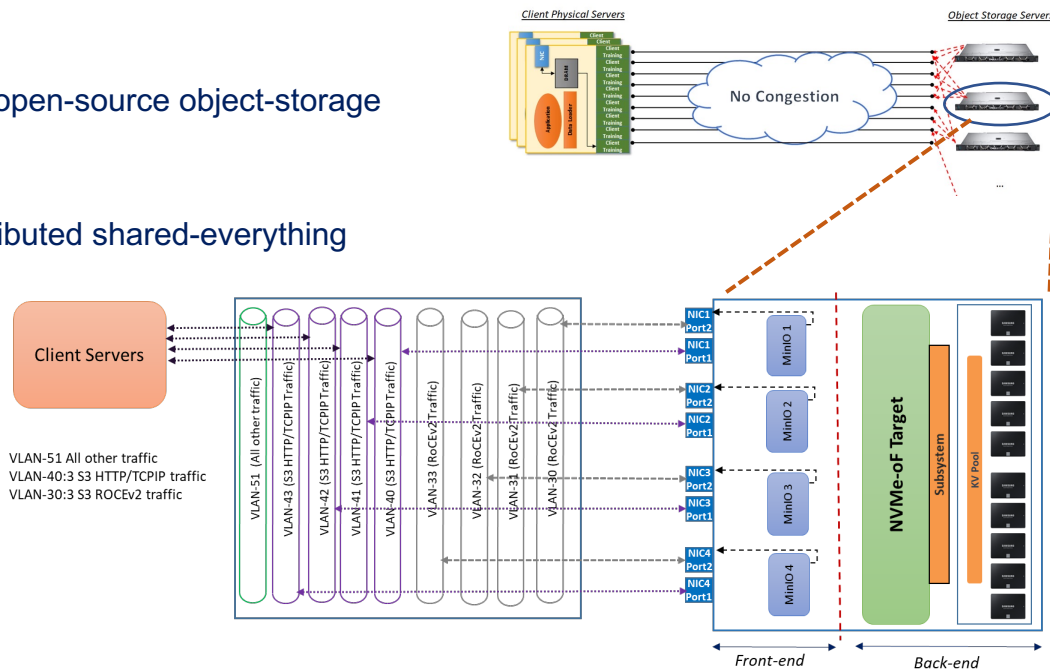
- I. Object storage server
- II. Network setup
- III. Client servers



Architectural Design

I. Storage Server

- Front end
- MinIO well-known Amazon S3 compliant open-source object-storage
Use KV API for data store access
- We modified stock MinIO to run in a distributed shared-everything
Key-Value environment for improved
scaling and performance.
- Erasure coding, for data consistency
faulty drives and random bit flipping



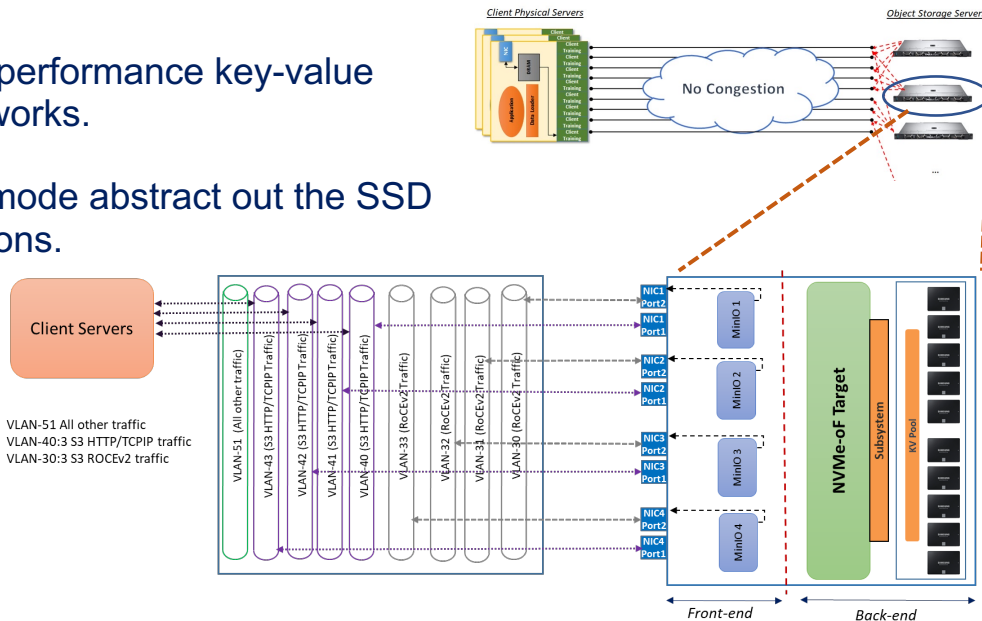
Architectural Design

I. Storage Server

- Back end
- NVMe-oF supported Target stack high-performance key-value services over RDMA and IP-based networks.
- Target application software run in user mode abstract out the SSD devices and perform Key-Value operations.

(I) KV pool, mapped to one/more SSD

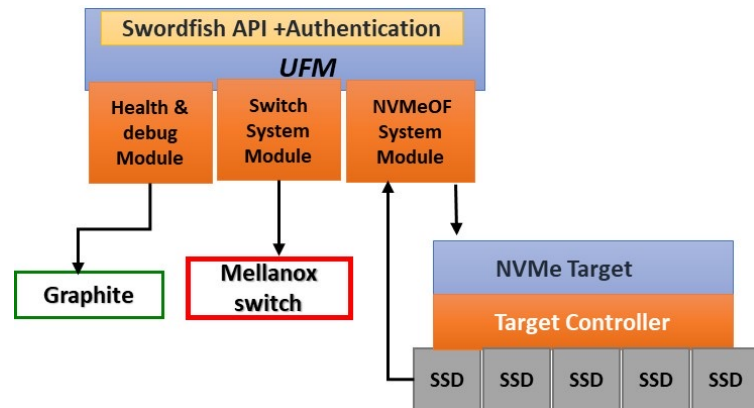
(II) subsystem, pool many SSDs using one or more Namespace/ Container with aggregated performance.



Architectural Design

I. Storage Server

- Back end
- UFM (Unified Fabric Manager) lightweight ecosystem software
- Manages Samsung devices, UFM manages any topology, architecture, and storage (KV-SSD).
- Manages the fabric, discovers, monitors, and configure devices and networks.
- Collects logs and statistics to ensure the cluster is working properly.



Architectural Design

II. Network Setup

DSS supports multiple high-speed Ethernet network ports,

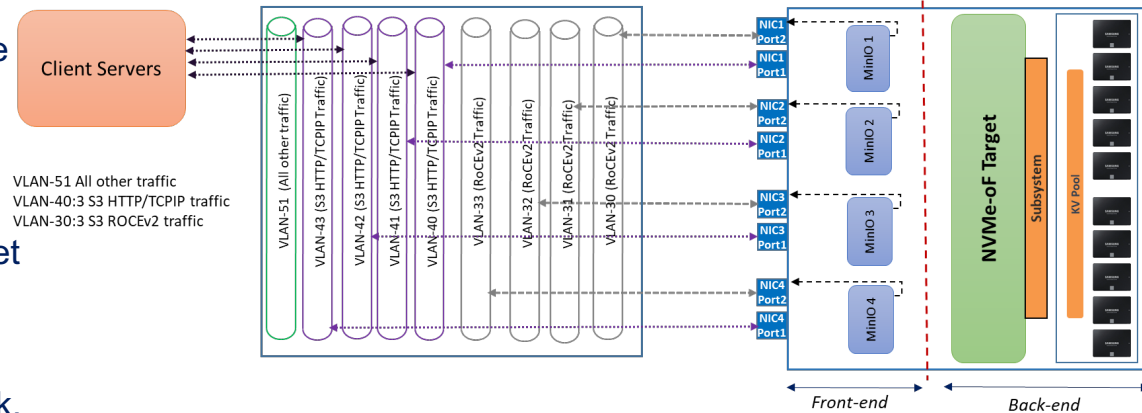
Each storage server has 4x dual port NICs and network software stack supports two different protocols:

TCP/IP (S3 HTTP) Front-end VLANs

- interaction of the client and the storage
- S3 traffic

RoCE v2 traffic Back-end VLANs

- RDMA protocol enables NVMeoF target to access Subsystem then read/write the objects in each drive.
- Clients do not interact with this network.

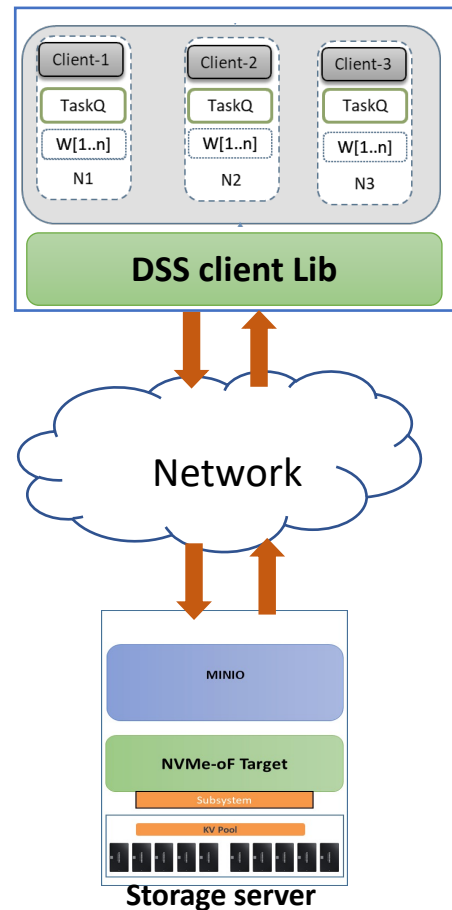


Architectural Design

III. Client Server

Run the application and requesting the data from the storage.

- DSS client library, set of APIs responsible for loading the requested data from storage and distribute data among DSS storage servers.
- DSS client library facilitates the access the storage servers by performing S3 operations PUT/ GET/ DEL/ LIST
- DSS client library takes cluster configurations containing a list of endpoints as input. Maximizes performance by load balancing and distributing the user request to the endpoints.



Experiments and Results

Testbed

- We evaluate DSS architecture using 10 homogeneous storage servers with 16 client servers.

	Storage Server	Client
CPU Type	AMD EPYC 7742 (ROMA)	Dell R740xd
CPU Speed	3.4 GHz	2.6 GHz
Num of Cores	64	24
OS	CentOS	CentOS
NIC	4x Dual 200GbE	2x 100GbE
Storage Node SSD	PM1733 (16x) 4TB	N/A

Experiments and Results

Results

We ran S3 benchmark with 30 TB data of 1MB, 2MB object size.

We measure the throughput with and without Erasure Coding (EC).

DSS storage server achieve around 180- 275 GB/sec for read (GET) 26-38 GB/sec for write (PUT)

1M	Erasure Coding (GB/sec)	No Erasure Coding (GB/sec)
PUT	26.27	38.2
GET	180	275.4

1M	Erasure Coding (GB/sec)
PUT	25.9
GET	267.6

Conclusions & Future work

- Samsung DSS Storage solution is a new object storage system, which deploys Key-Value APIs on NVMeOF SSDs.
- DSS is a disaggregated storage system that features deterministic high I/O bandwidth and scalability over object storage.
- DSS throughput evaluation for read and write on 10 node storage and 16 node client cluster, are around 2 and 1 orders of magnitude accordingly.
- In the future, we want to improve our storage system performance further, by enabling S3 service over RDMA to eliminate http/TCP copy overhead. Also evaluate our system on larger cluster.

Thanks!

Samsung Memory Solution Lab Group

Harsh Roogi h.roogi@samsung.com

Somnath Roy som.roy@samsung.com

Ron Lee r2.lee@samsung.com

Mahsa Bayati mahsa.b@Samsung.com







OCP

FUTURE TECHNOLOGIES SYMPOSIUM

2021 OCP Global Summit | November 8, 2021, San Jose, CA