LUSTRE USER GROUP 2023

Bringing Lustre to the masses through a fully-managed cloud service

Darryl Osborne

Principal Solutions Architect
Amazon Web Services

Agenda

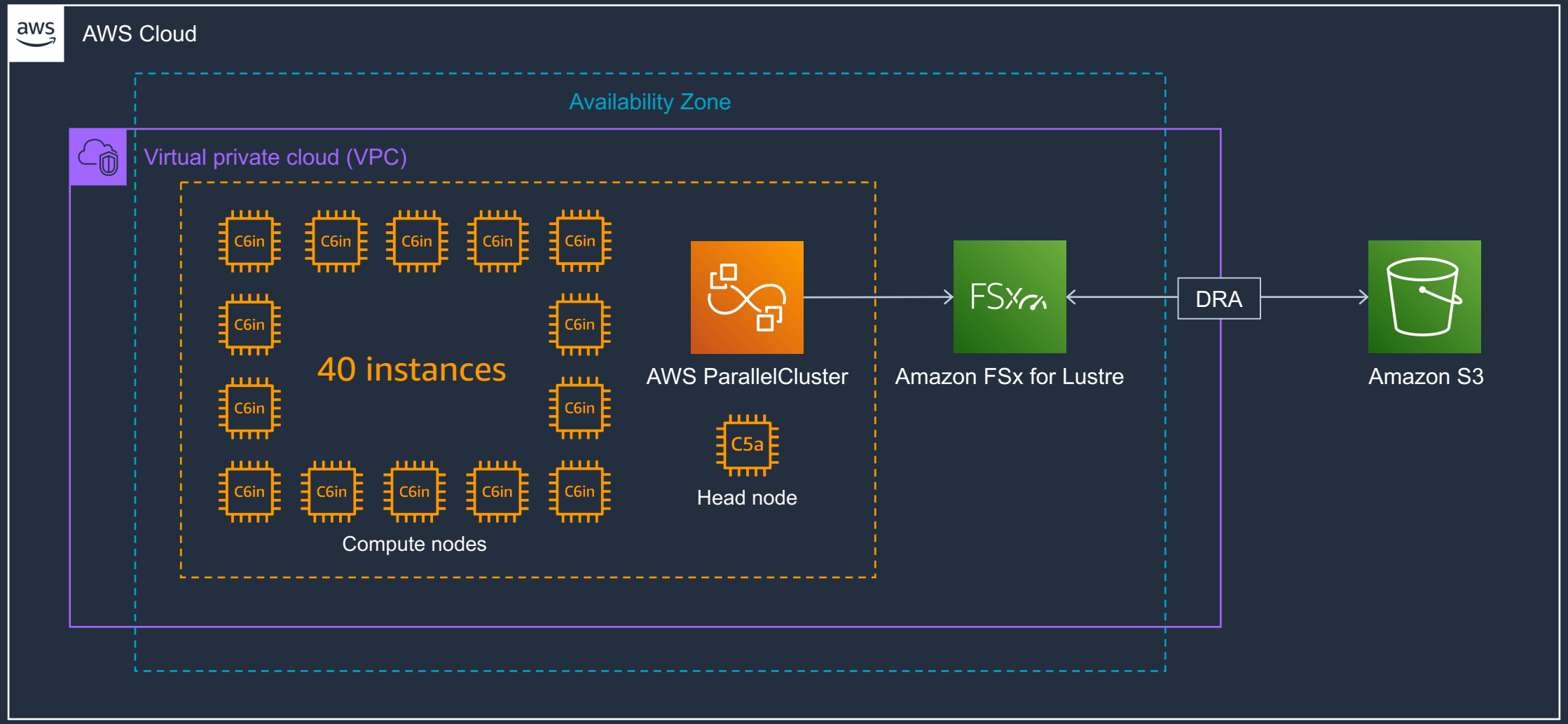
Demo

Presentation

Juggle

Demo

From nothing to 200+ GB/s in 30 minutes or less



Agenda

Demo – From nothing to 200+ GB/s in 30 minutes or less

Use cases

Architecture

HSM solution using Amazon S3

Performance

Q&A

Use cases

RIVIAN

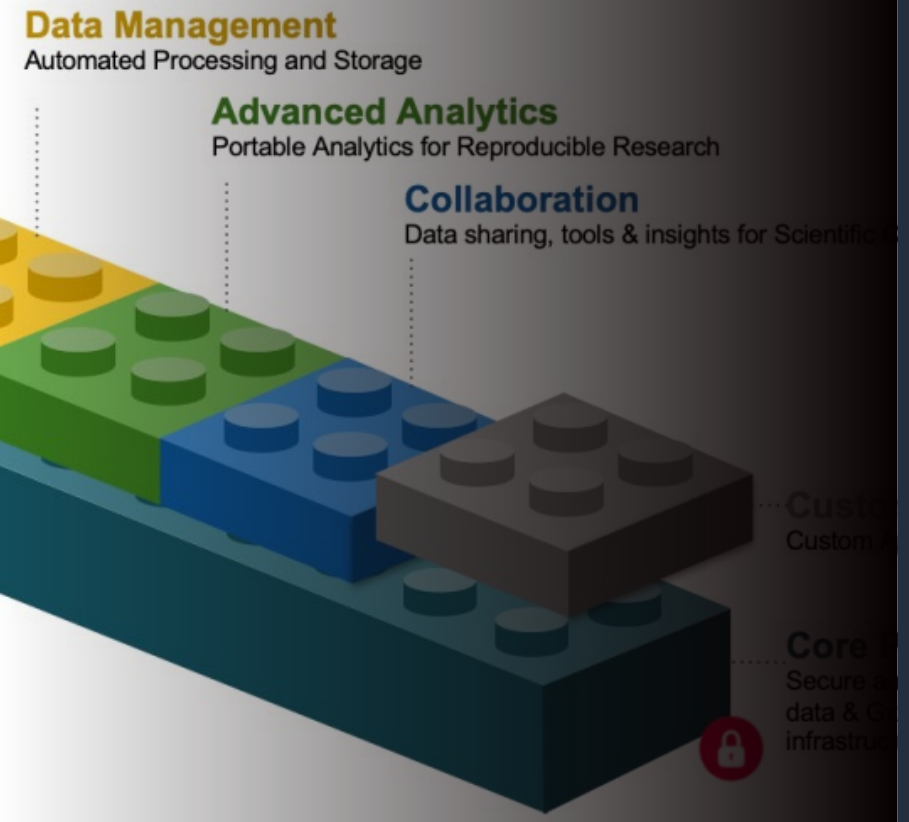
Rivian used **Amazon FSx for Lustre and Amazon EC2 to** support new concepts, crash and vibration testing, and simulations and achieved a up to a 56% workload acceleration.

"This is accelerating adoption across the board."

Madhavi Isanaka
Chief Information Officer, Rivian

FSx

Learn More



Roche

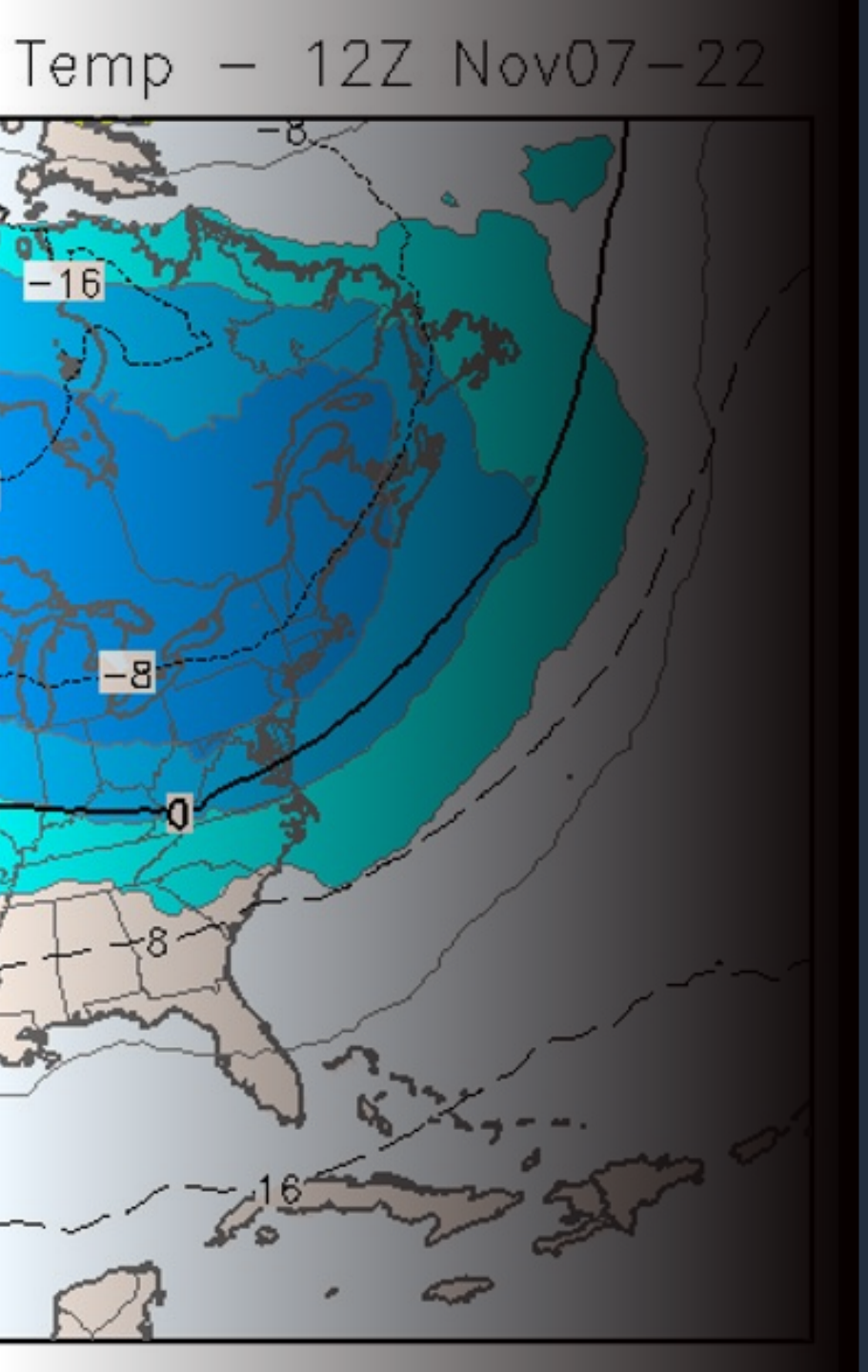
Image files that previously took 2–3 days for processing are now ready in hours, and modular electronic health record datasets get processed within minutes.

“Roche is taking steps closer towards its mission to **provide every patient with the best treatment possible in the fastest time**.”

Mustaqhusain Kazi, Head of Personalized Healthcare, Pharma Informatics at Roche

FSx

Learn More



MAXAR

Maxar uses AWS to deliver forecasts 58% faster than weather supercomputer

"Maxar used **Amazon FSx for Lustre** in our AWS HPC solution for running NOAA's numerical weather forecasting model. This allowed us to reduce compute time by 58%, generating the forecast in about 45 minutes for a much more cost-effective price point. Maximizing our AWS compute resources was an incredible performance boost for us."

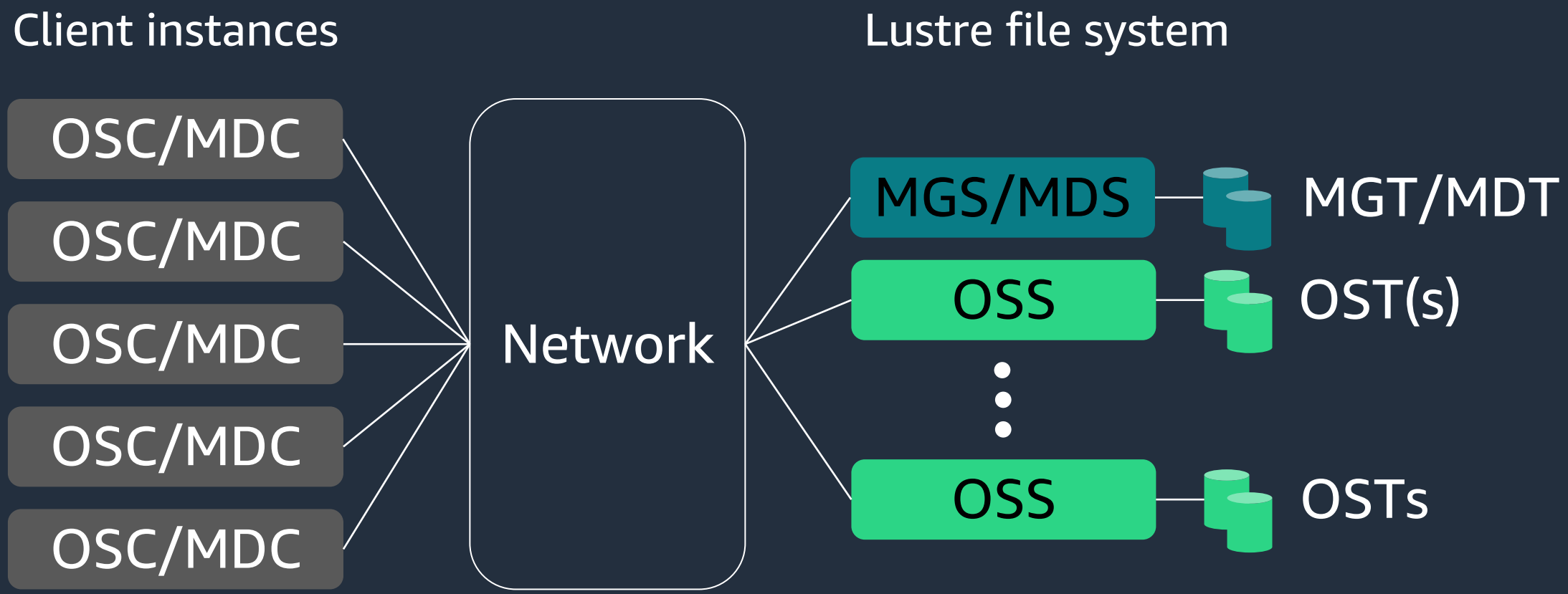
Stefan Cecelski,
PhD Senior Data Scientist & Engineer, Maxar Technologies

FSx

Learn More

Architecture

Lustre architecture

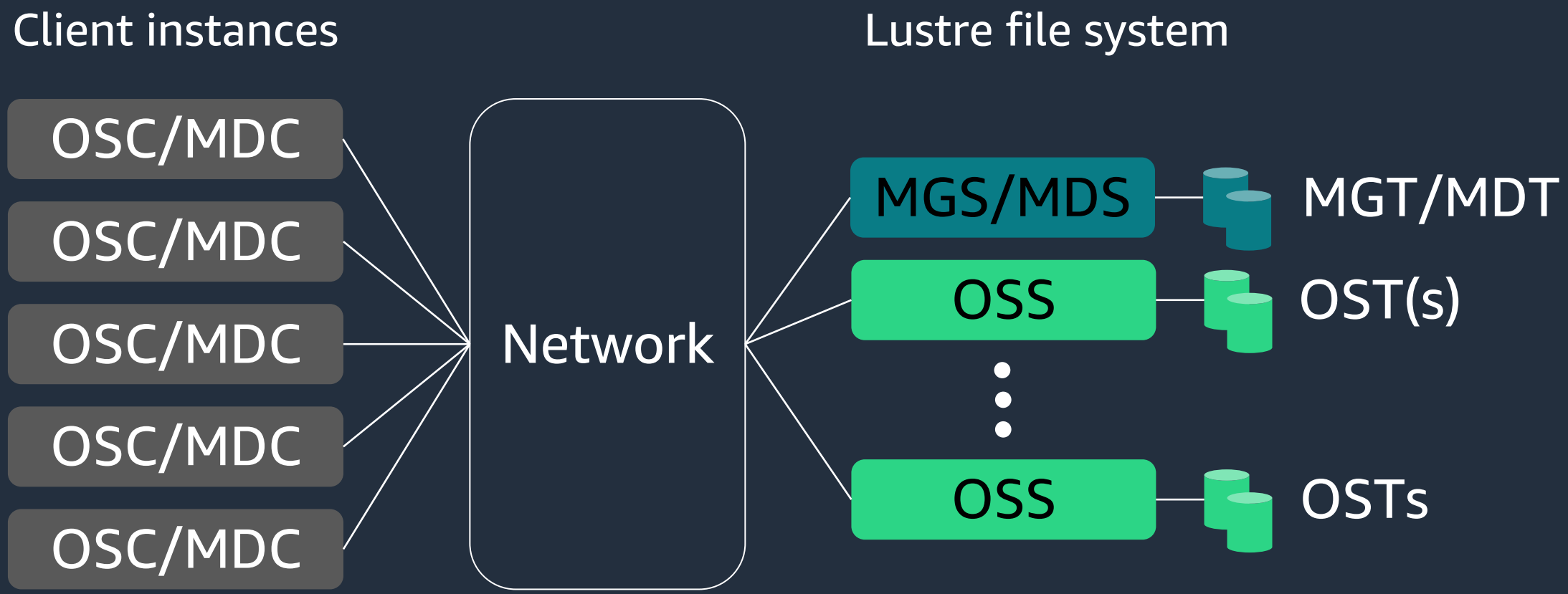


Object Storage Client (OSC) | Metadata Client (MDC) | Management Server (MGS) | Management Target | Metadata Server (MDS) | Metadata Target (MDT) | Object Storage Server (OSS) | Object Storage Target (OST)

Storage and deployment types

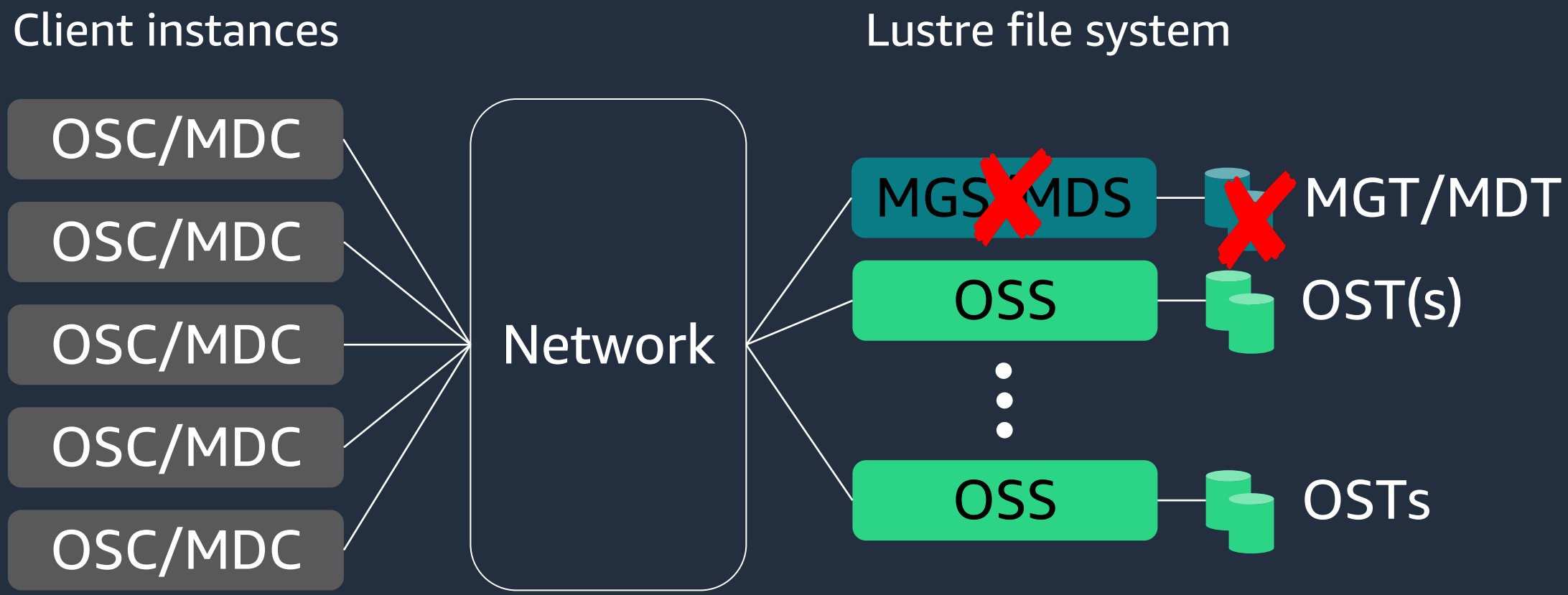
Storage type	Deployment type
HDD	Persistent
SSD	Scratch
	Persistent

Lustre architecture



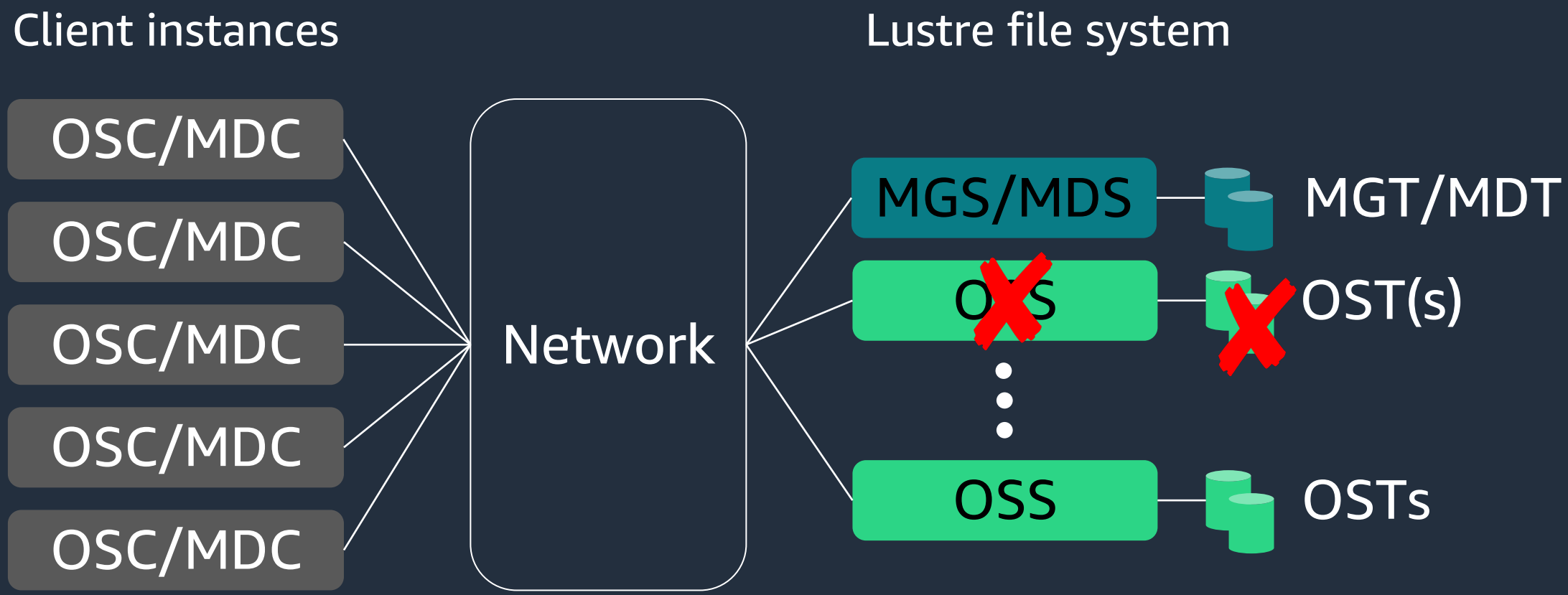
Object Storage Client (OSC) | Metadata Client (MDC) | Management Server (MGS) | Management Target | Metadata Server (MDS) | Metadata Target (MDT) | Object Storage Server (OSS) | Object Storage Target (OST)

Lustre architecture and persistent file systems



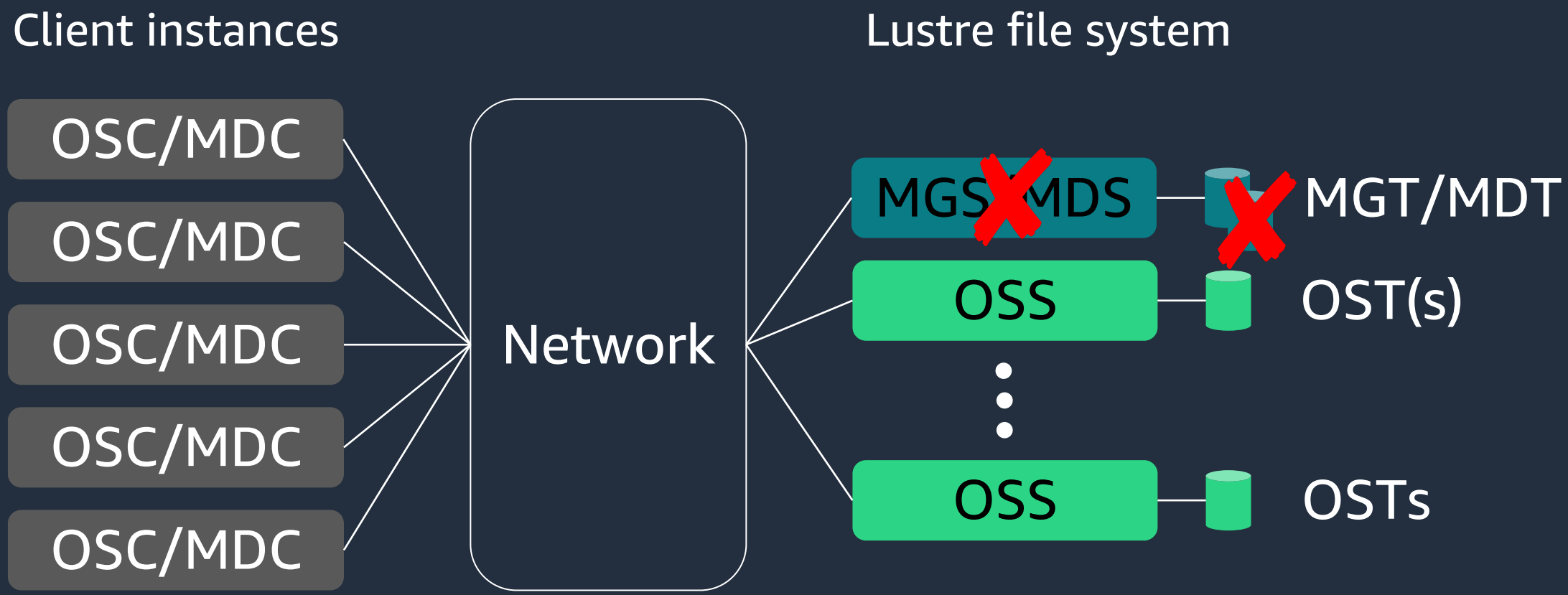
Object Storage Client (OSC) | Metadata Client (MDC) | Management Server (MGS) | Management Target | Metadata Server (MDS) | Metadata Target (MDT) | Object Storage Server (OSS) | Object Storage Target (OST)

Lustre architecture and persistent file systems



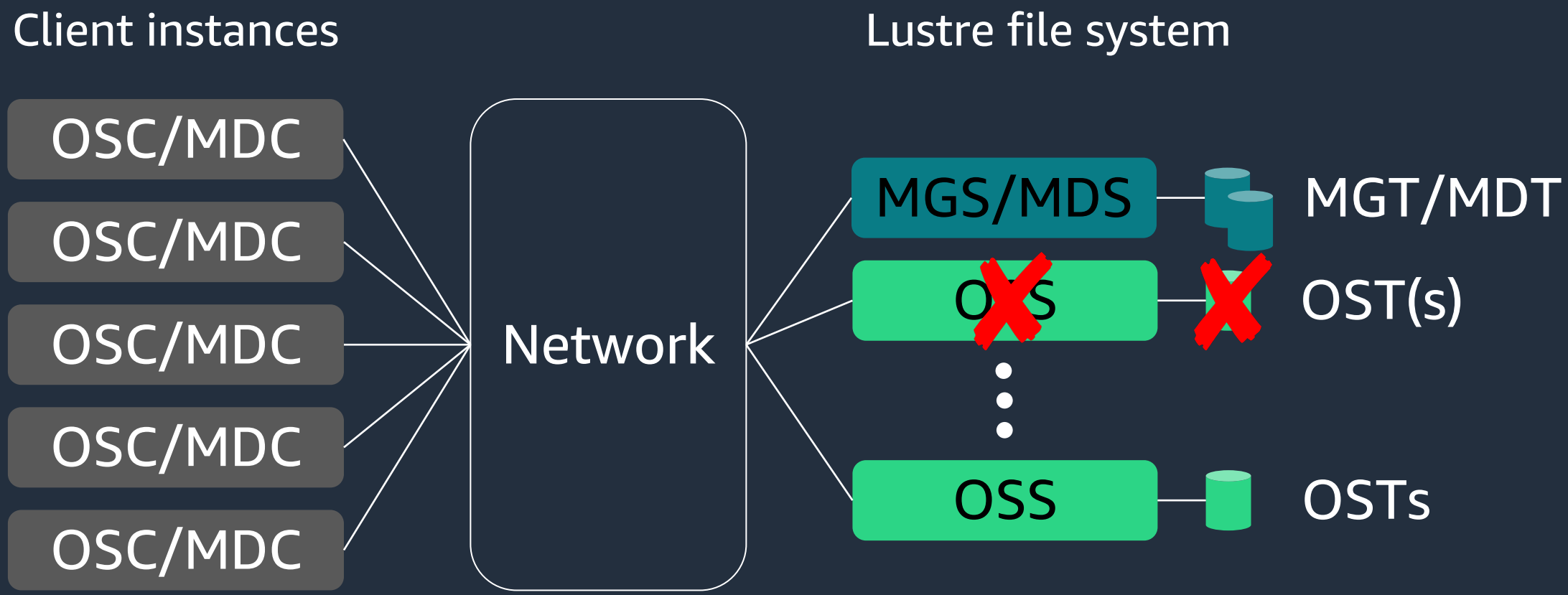
Object Storage Client (OSC) | Metadata Client (MDC) | Management Server (MGS) | Management Target | Metadata Server (MDS) | Metadata Target (MDT) | Object Storage Server (OSS) | Object Storage Target (OST)

Lustre architecture and scratch file systems



Object Storage Client (OSC) | Metadata Client (MDC) | Management Server (MGS) | Management Target | Metadata Server (MDS) | Metadata Target (MDT) | Object Storage Server (OSS) | Object Storage Target (OST)

Lustre architecture and scratch file systems



Object Storage Client (OSC) | Metadata Client (MDC) | Management Server (MGS) | Management Target | Metadata Server (MDS) | Metadata Target (MDT) | Object Storage Server (OSS) | Object Storage Target (OST)

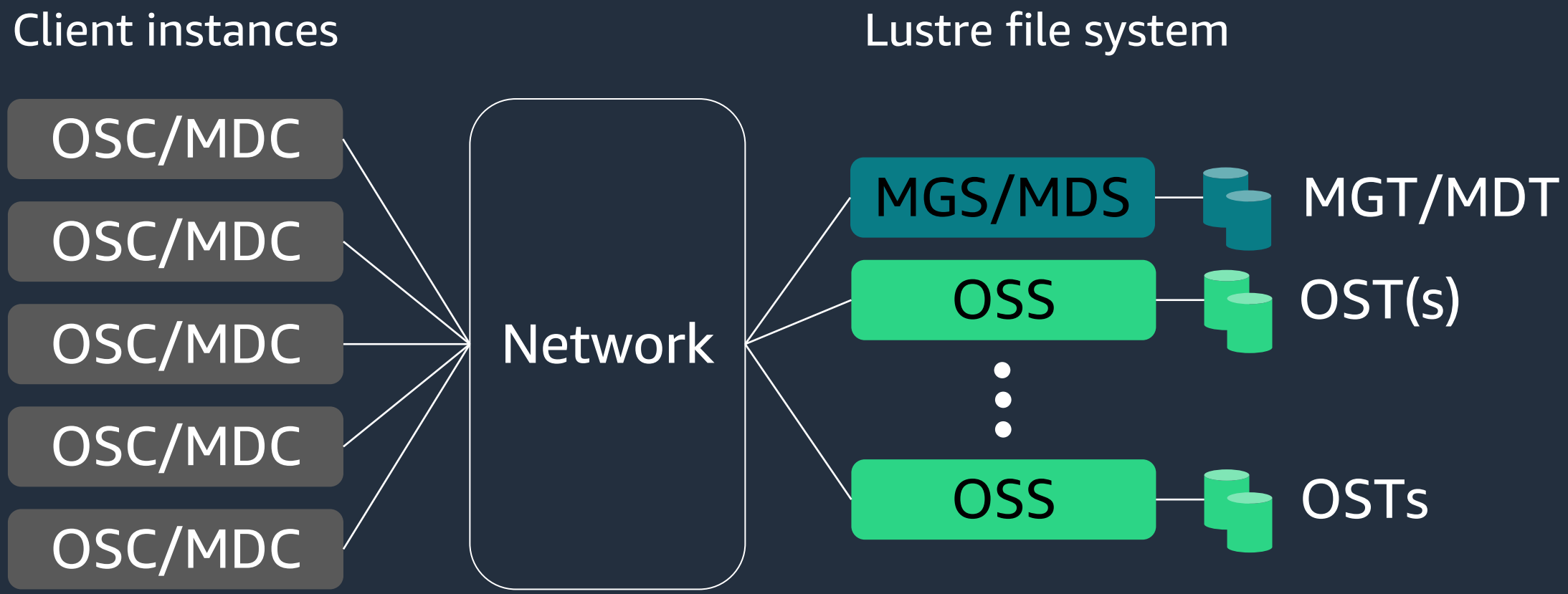
Storage and deployment types

Storage type	Deployment type	Disk storage throughput (MB/s per TiB of storage)	SSD read cache throughput (MB/s per TiB of cache*)	Price per GB-month**
HDD	Persistent	12	-	$0.025
			200	$0.041
		40	-	$0.083
			200	$0.099
SSD	Scratch	200	-	$0.140
	Persistent	125	-	$0.145
		250	-	$0.210
		500	-	$0.340
		1000	-	$0.600

* Read cache sized at 20% of HDD storage capacity

** US East (N. Virginia) pricing

Lustre architecture



Object Storage Client (OSC) | Metadata Client (MDC) | Management Server (MGS) | Management Target | Metadata Server (MDS) | Metadata Target (MDT) | Object Storage Server (OSS) | Object Storage Target (OST)

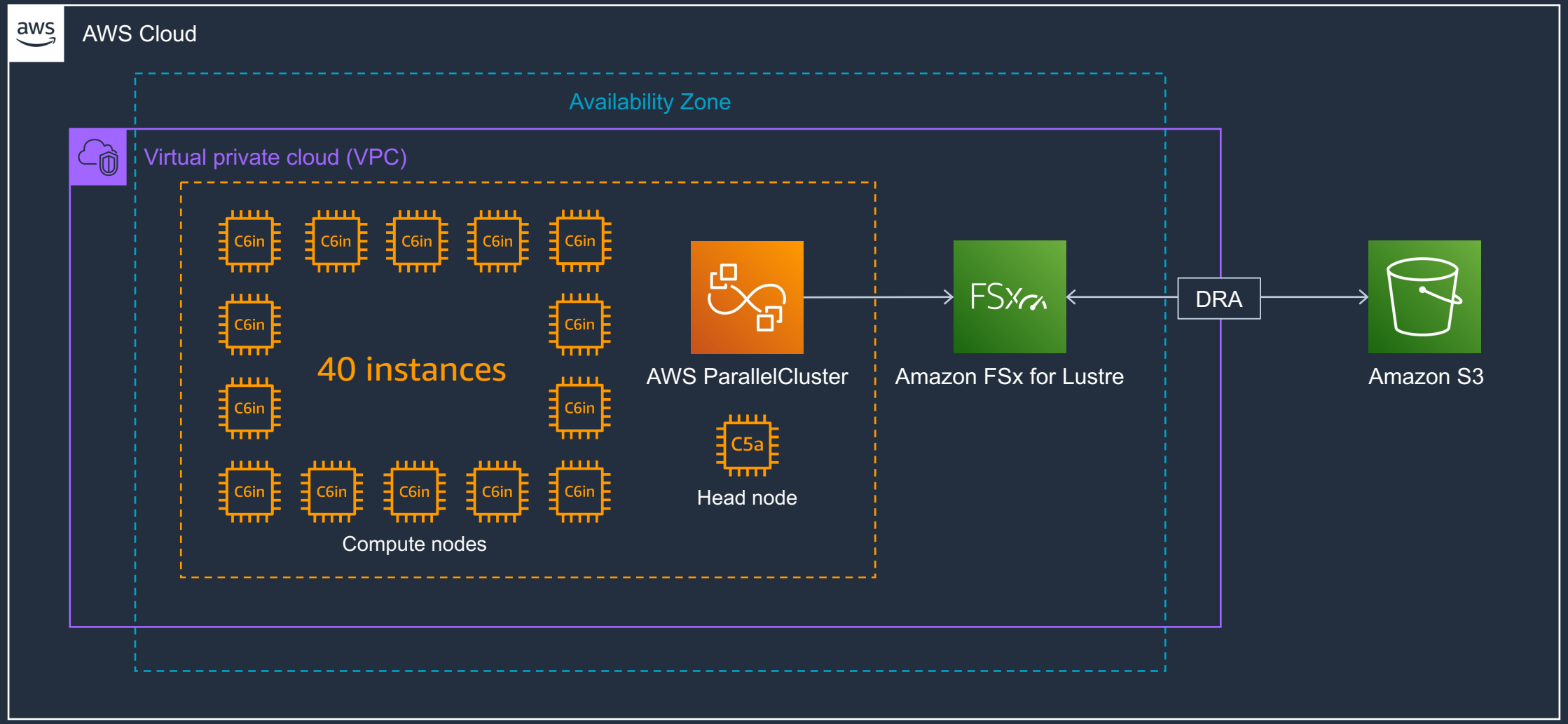
Storage and deployment types

Storage type	Deployment type		Minimum size (TiB)	Incremental size (TiB)	OSS count (1 per x TiB)	OST size (TiB)	OST count per OSS	MDS and MDT count
HDD	Persistent	12 MB/s	6.0	6.0	6.0	1.5	4	1 and 1
		40 MB/s	1.8	1.8	1.8	1.8	1	
SSD	Scratch		1.2	2.4	2.4	1.2	2	
	Persistent							

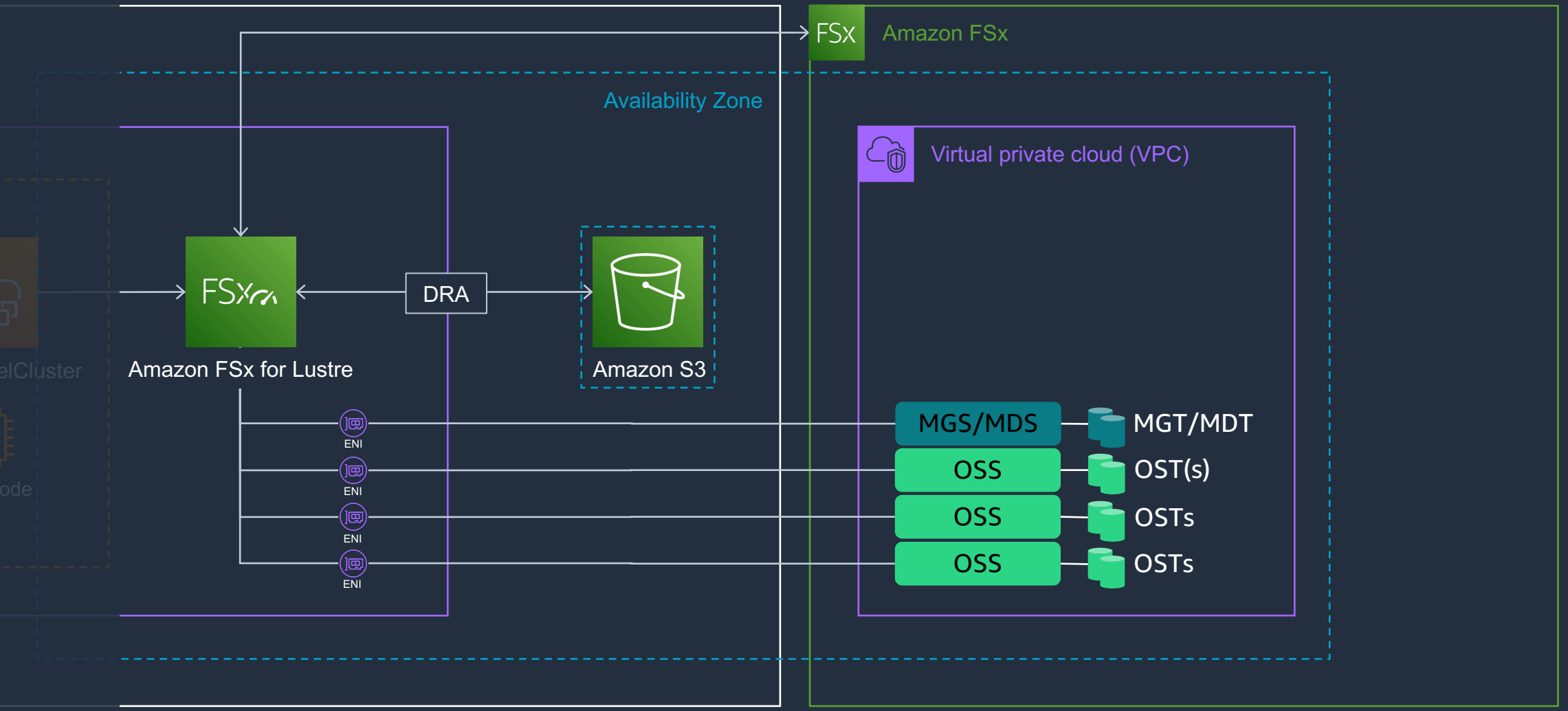
Example:

SSD Persistent 100.8 TiB: 1 MDS/MDT = 1 Elastic Network Interface (ENI)
42 OSSs (100.8 ÷ 2.4) = 42 Elastic Network Interfaces (ENIs)
84 OSTs (100.8 ÷ 1.2)

Demo environment

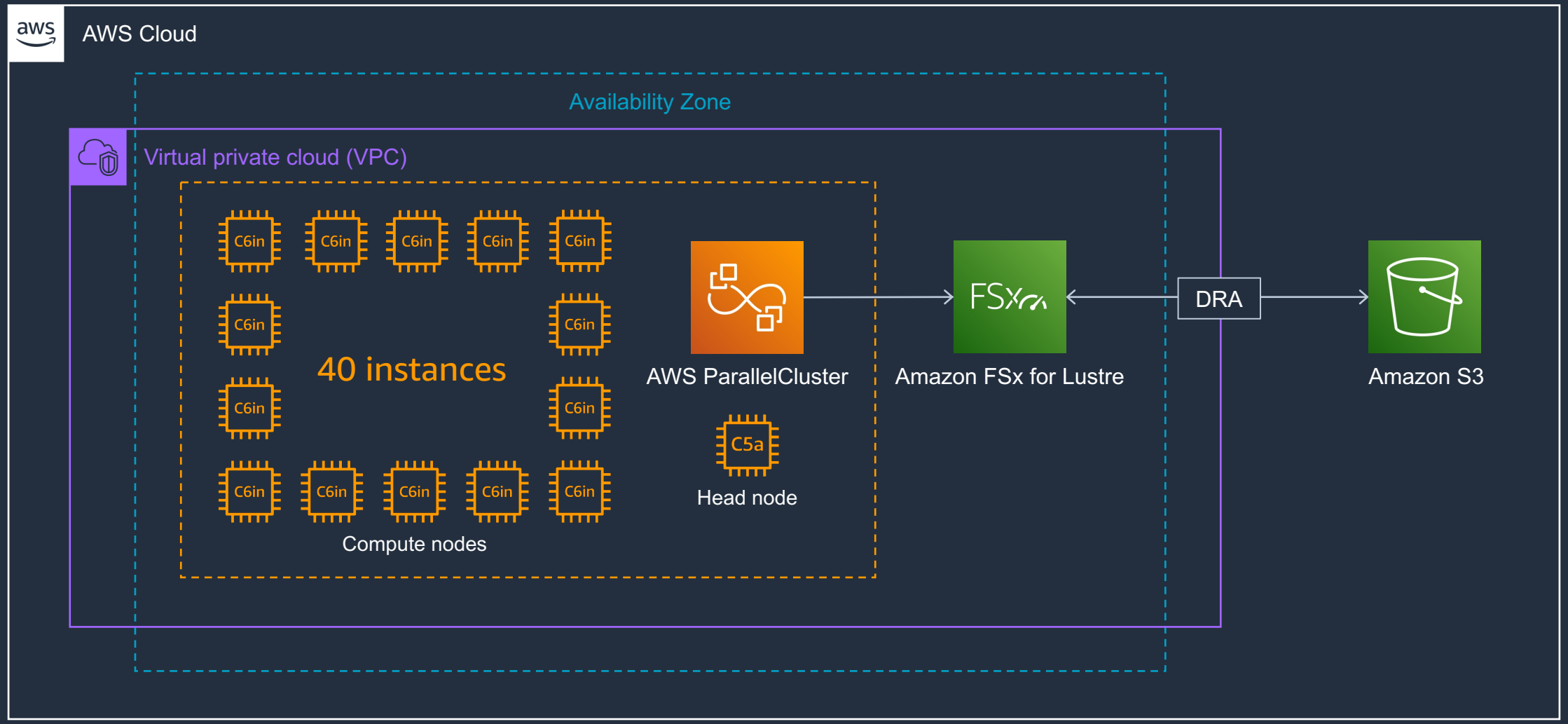


Demo environment

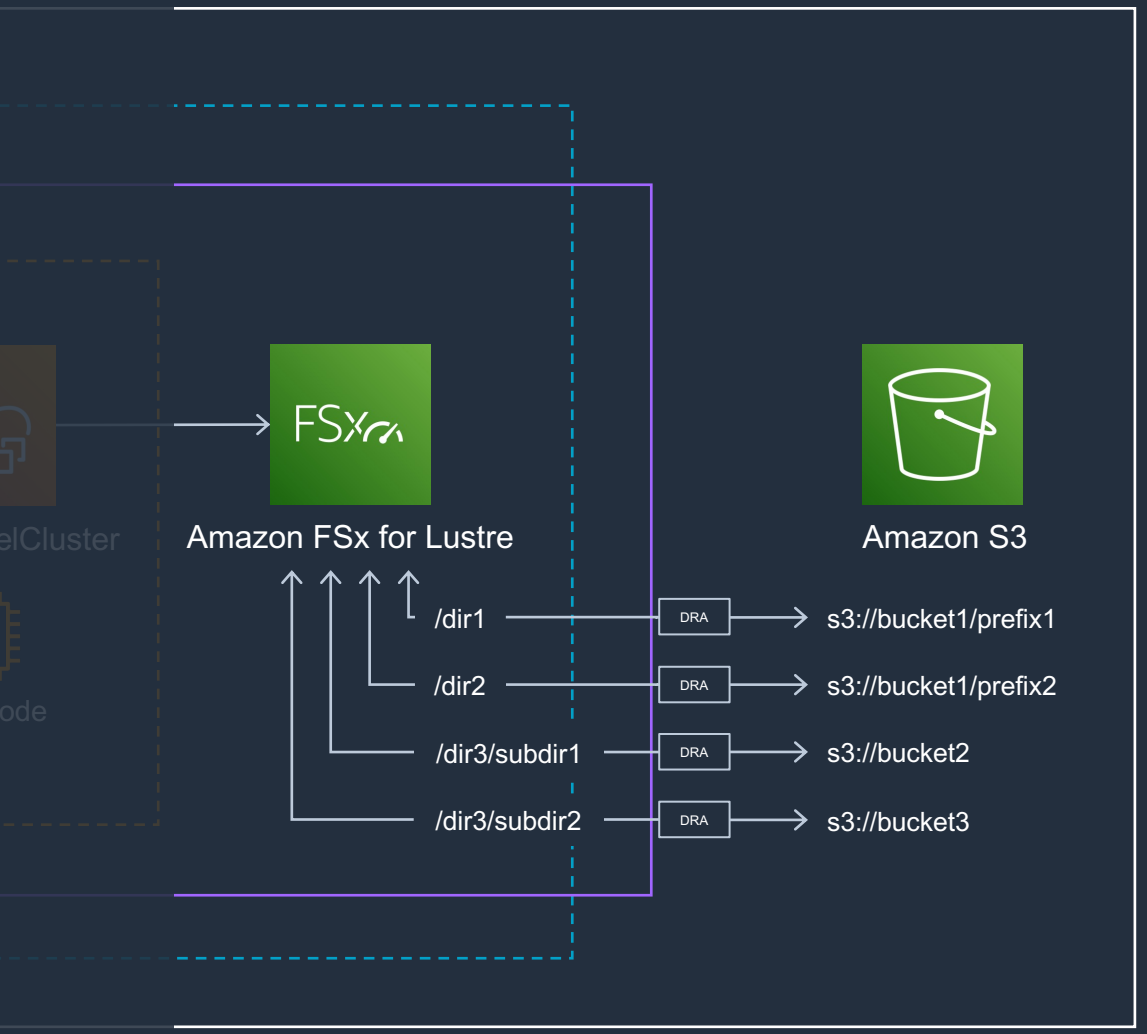


HSM using Amazon S3

Demo environment



Hierarchical Storage Management (HSM) using Amazon S3



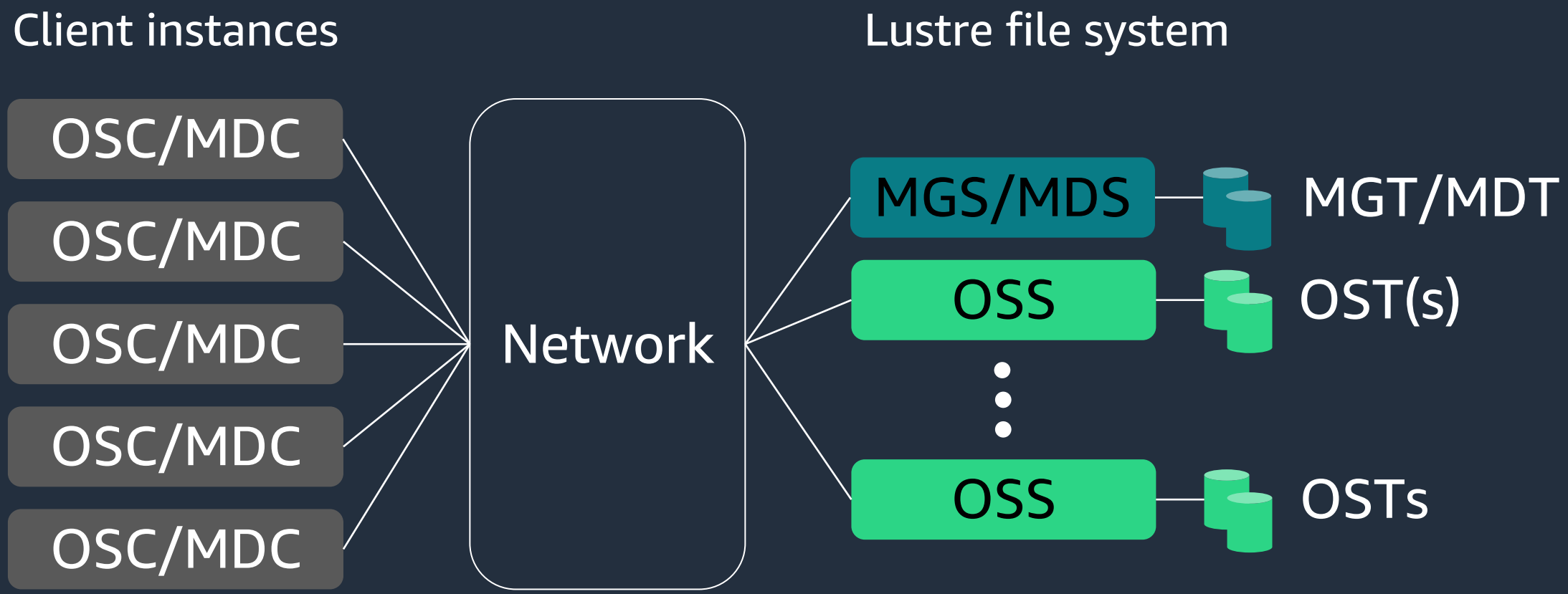
Data Repository Association (DRA)

- Up to eight (8) per file system
- DRA path is an S3 bucket or prefix
- Links file system path to a DRA path
- Cannot overlap file system paths
- Cannot overlap DRA paths
- Import policy – DRA path updates propagated to file system path
- Export policy – File system path updates propagated to DRA path
- 1:1 mapping between file system path and object keys

Demo – HSM solution on Amazon S3

Demo – hsm_restore

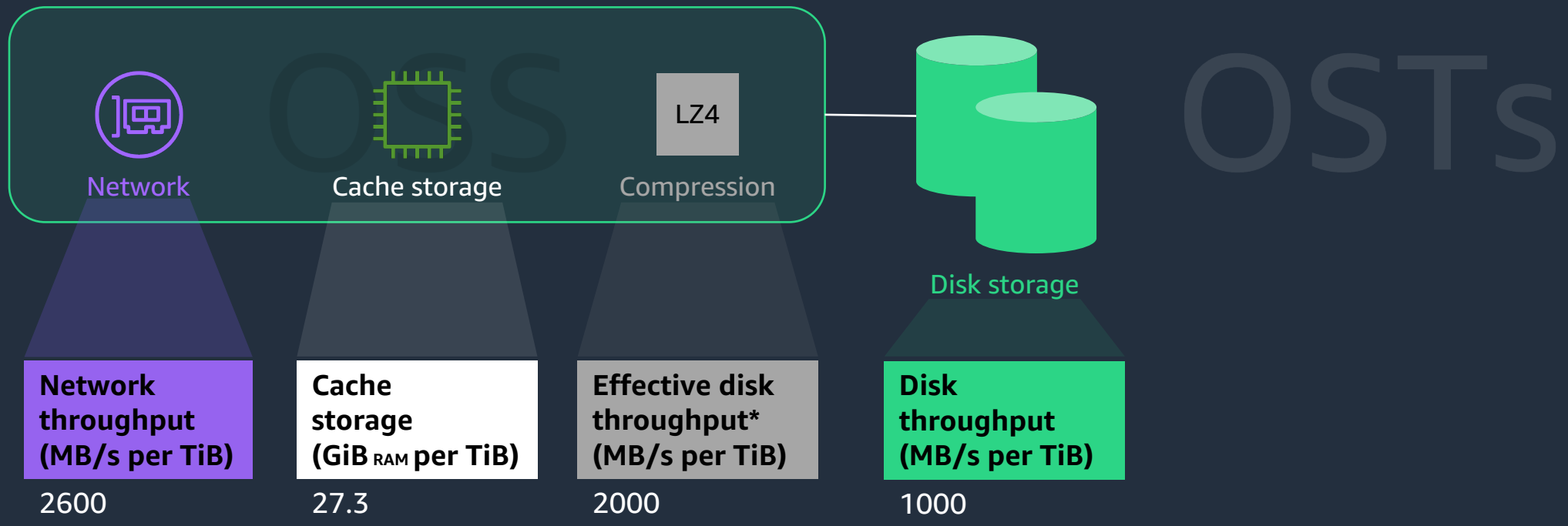
Lustre architecture



Object Storage Client (OSC) | Metadata Client (MDC) | Management Server (MGS) | Management Target | Metadata Server (MDS) | Metadata Target (MDT) | Object Storage Server (OSS) | Object Storage Target (OST)

Lustre architecture

Lustre architecture and performance



SSD Persistent 2 1000 MB/s per TiB
* 2:1 compression ratio

Demo – Parallel cluster read & auto export

Feature summary

HDD and SSD storage types

Persistent and scratch deployment types

LZ4 compression

Configurable file striping (PFLs)

Online storage capacity increases

AWS service integrations

Storage quotas

Root squash

Encryption at rest and in transit

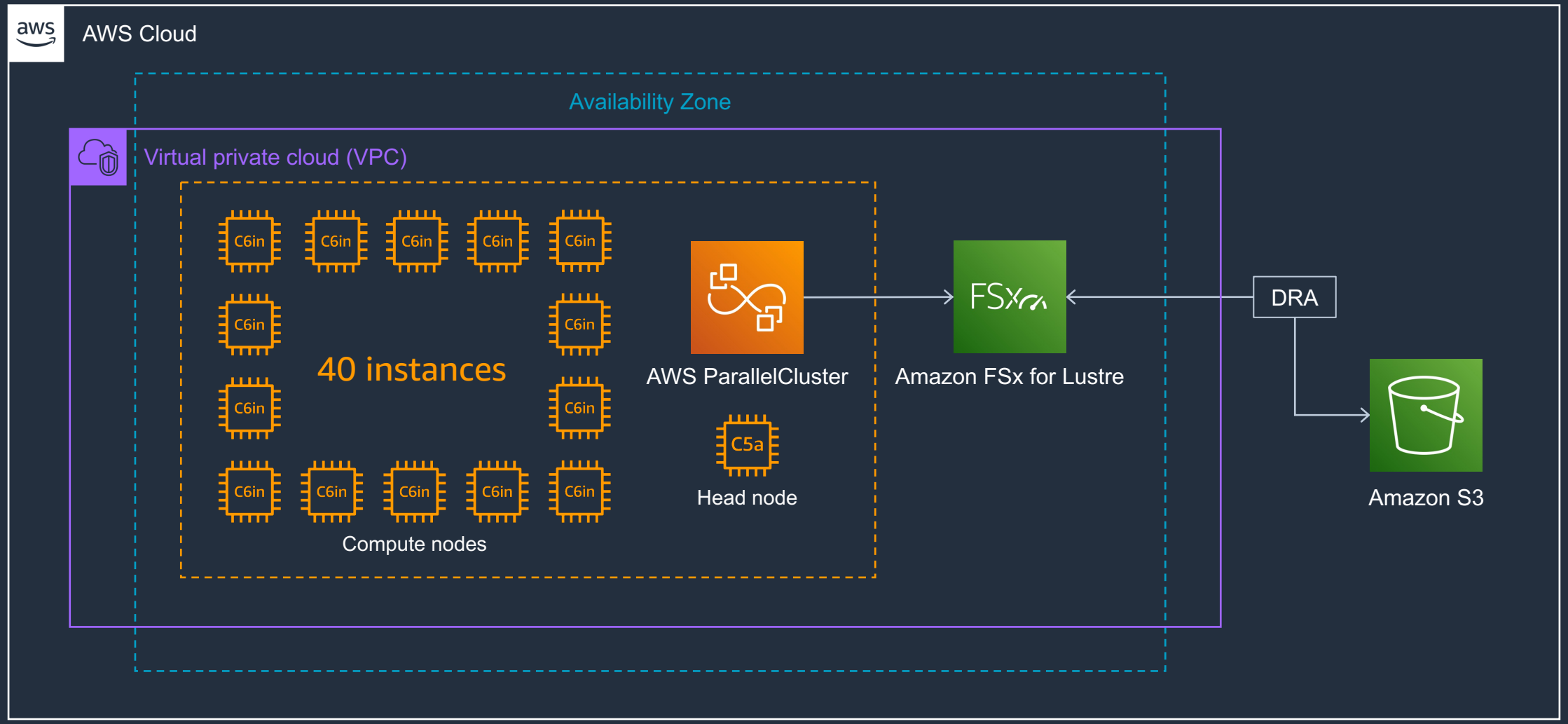
HSM solution using Amazon S3

Automatic backups

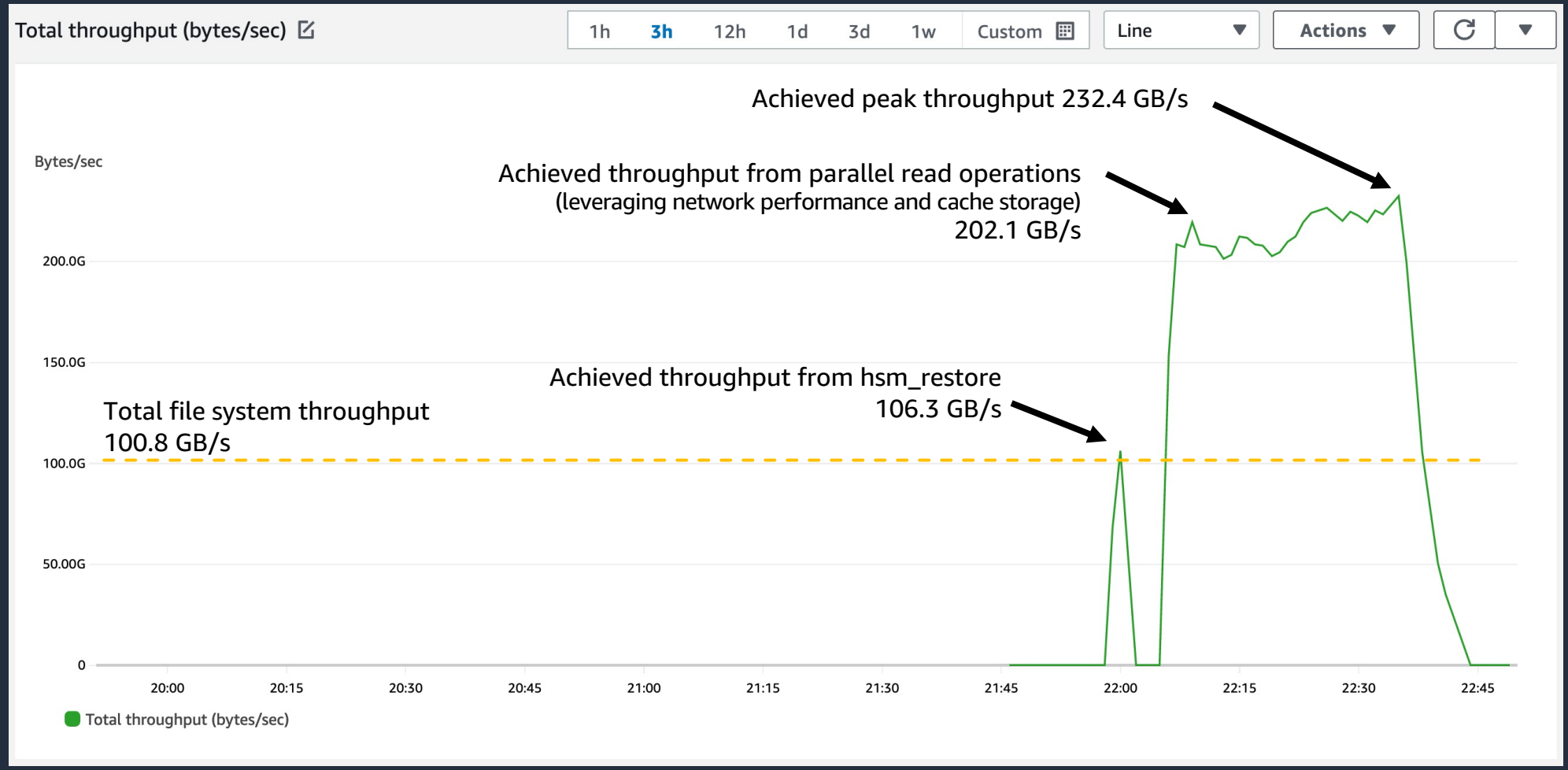
Weekly maintenance window

Q&A

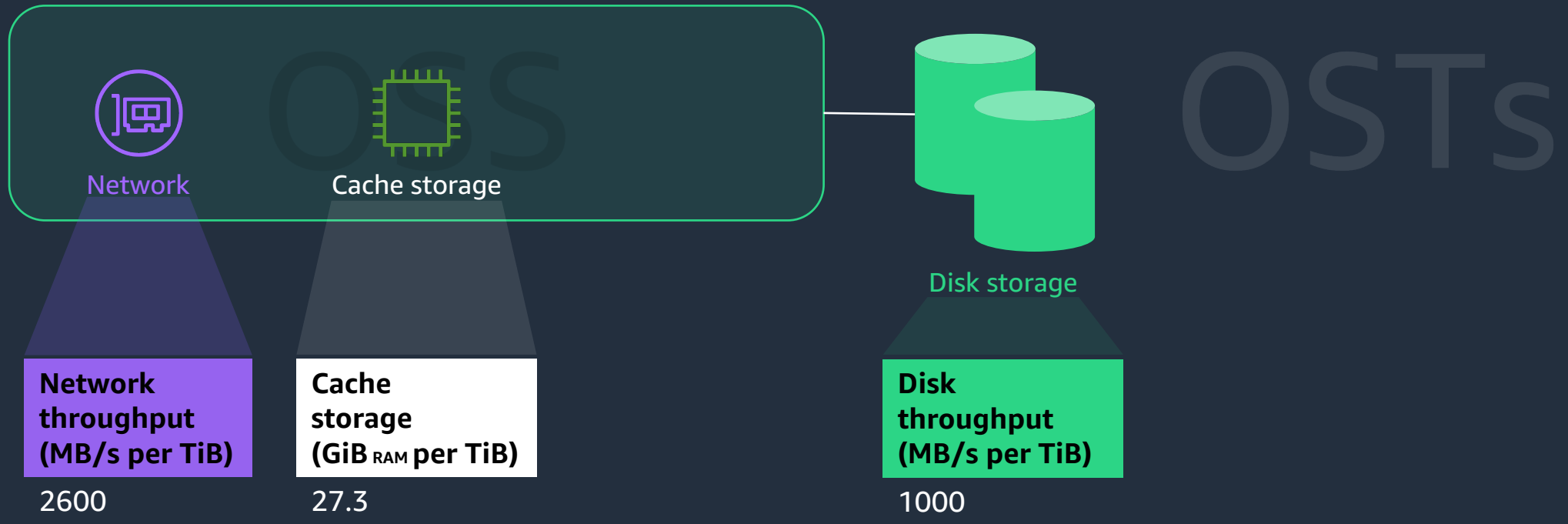
From nothing to 200+ GB/s in 30 minutes or less



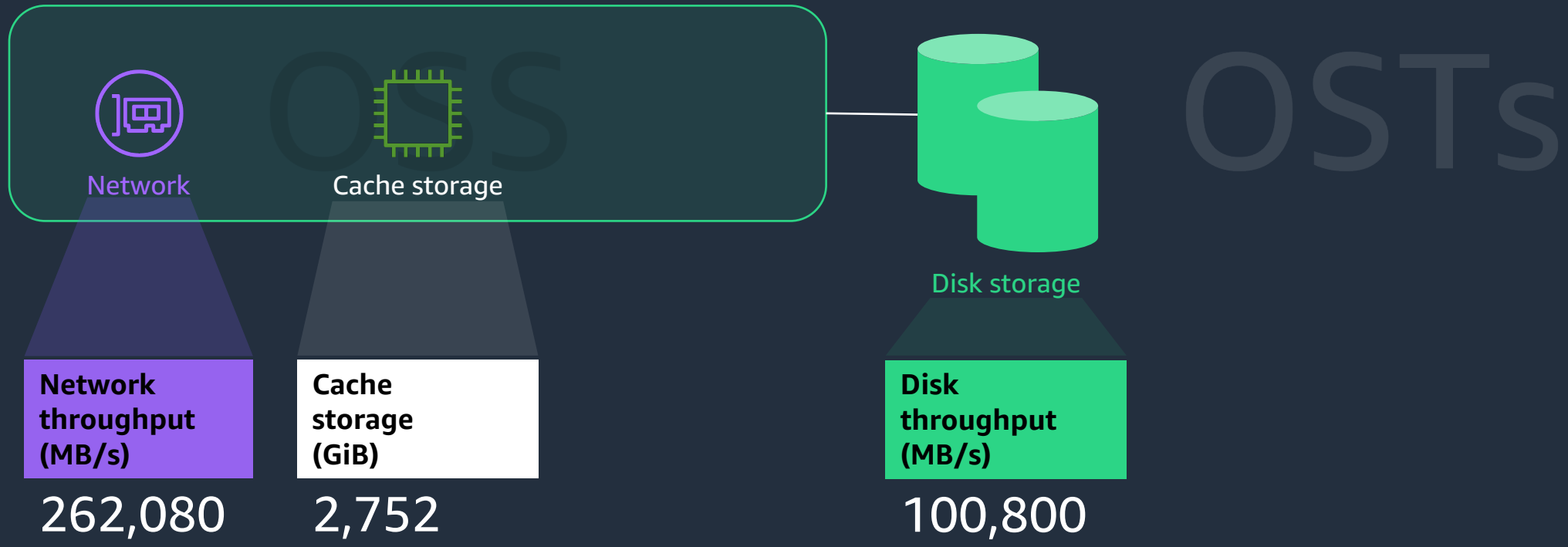
Demo results – Persistent 2 SSD 100.8 TiB at 1000 MB/s per TiB



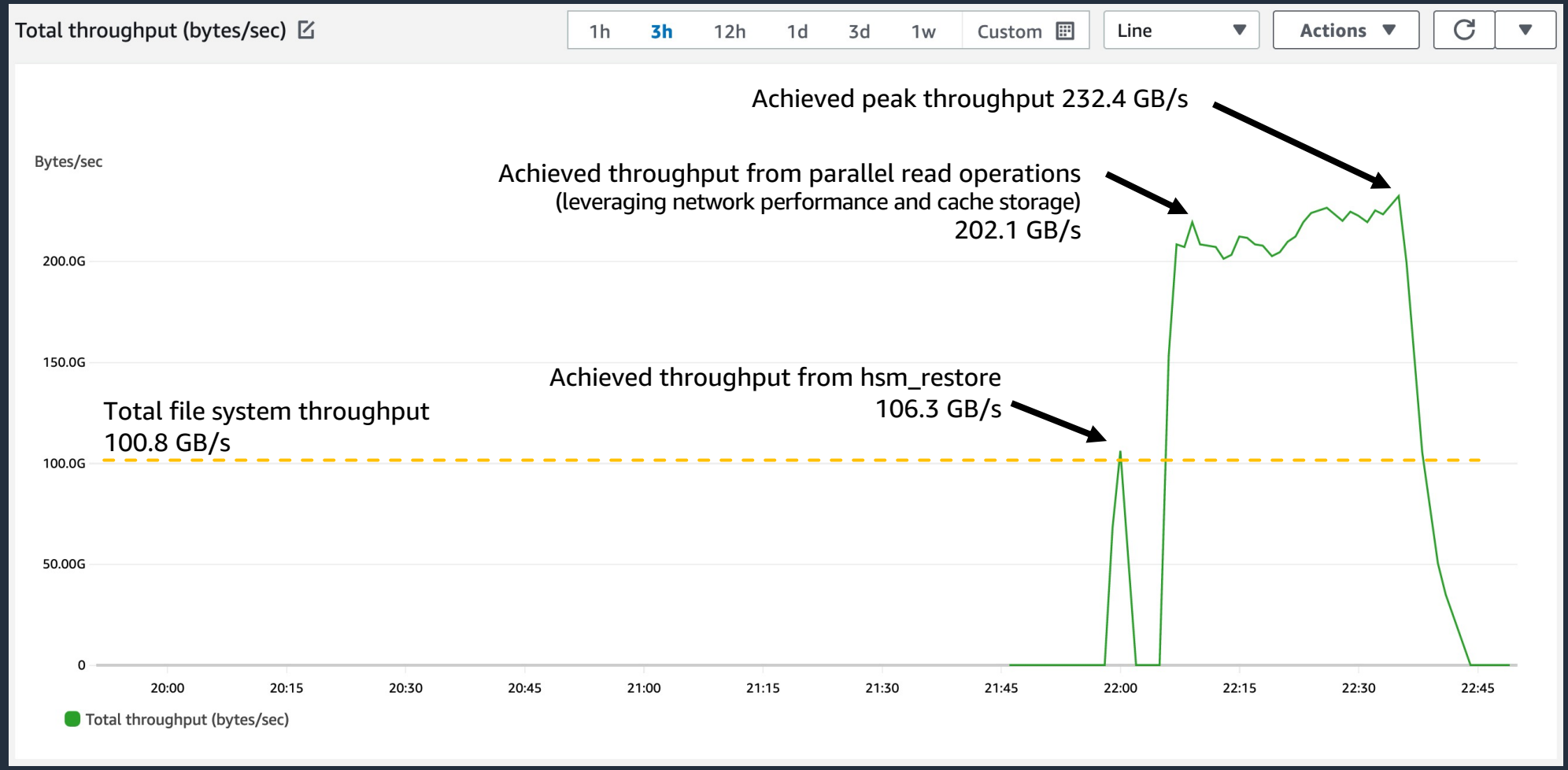
Demo results – Persistent 2 SSD 100.8 TiB at 1000 MB/s per TiB



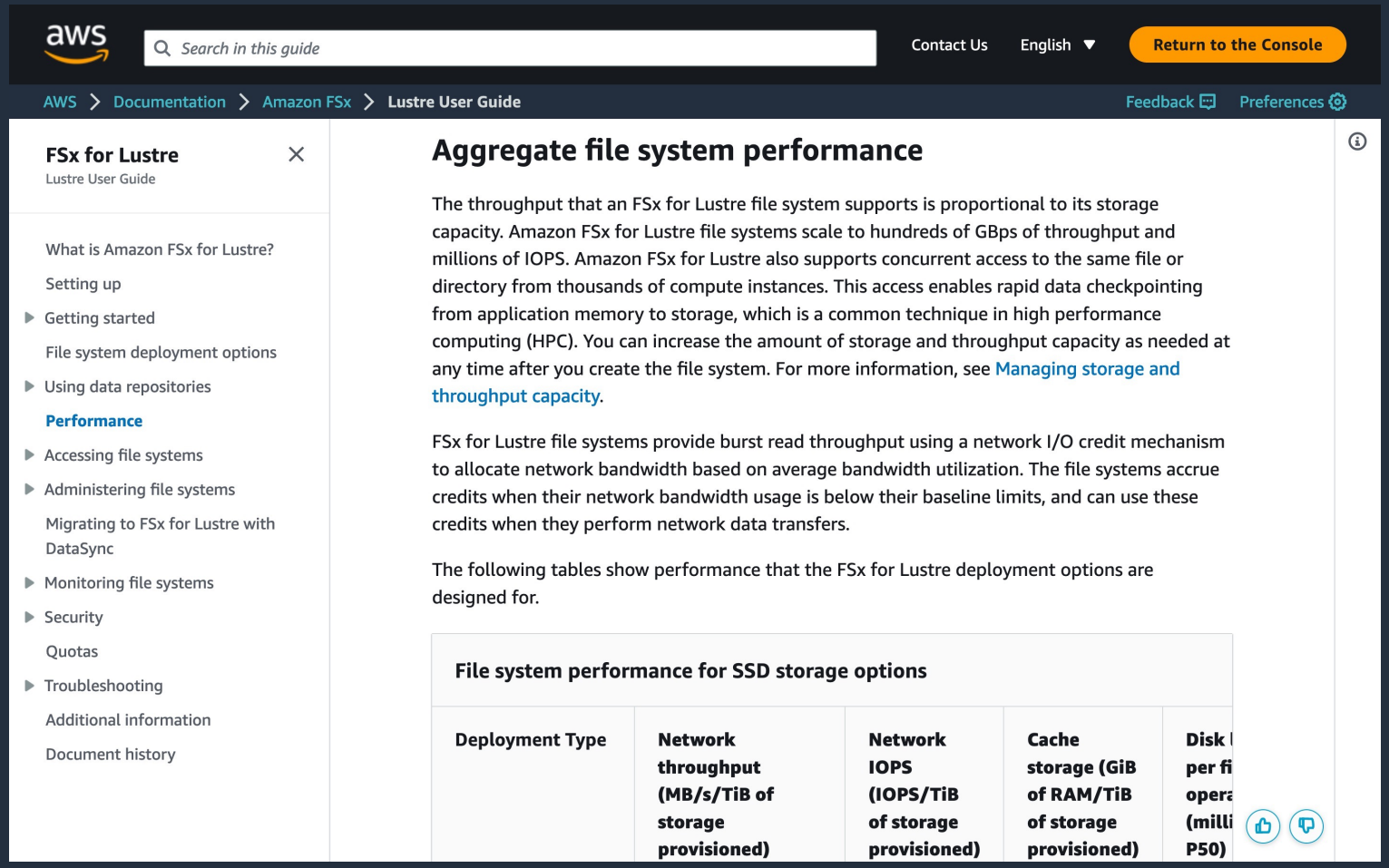
Demo results – Persistent 2 SSD 100.8 TiB at 1000 MB/s per TiB



Demo results – Persistent 2 SSD 100.8 TiB at 1000 MB/s per TiB



Lustre architecture and performance



https://docs.aws.amazon.com/fsx/latest/LustreGuide/performance.html

Thank you!

Darryl Osborne

darrylsosborne

darrylo@amazon.com