

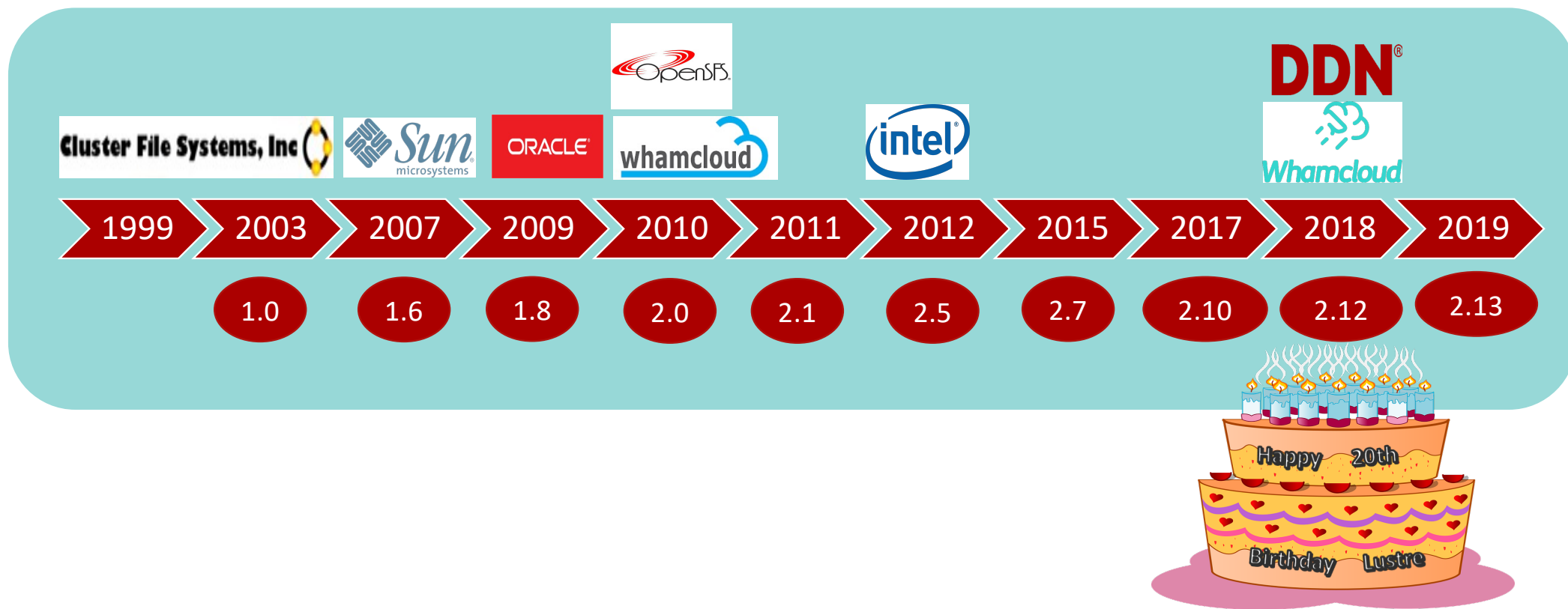


Next Generation DDN Exascaler

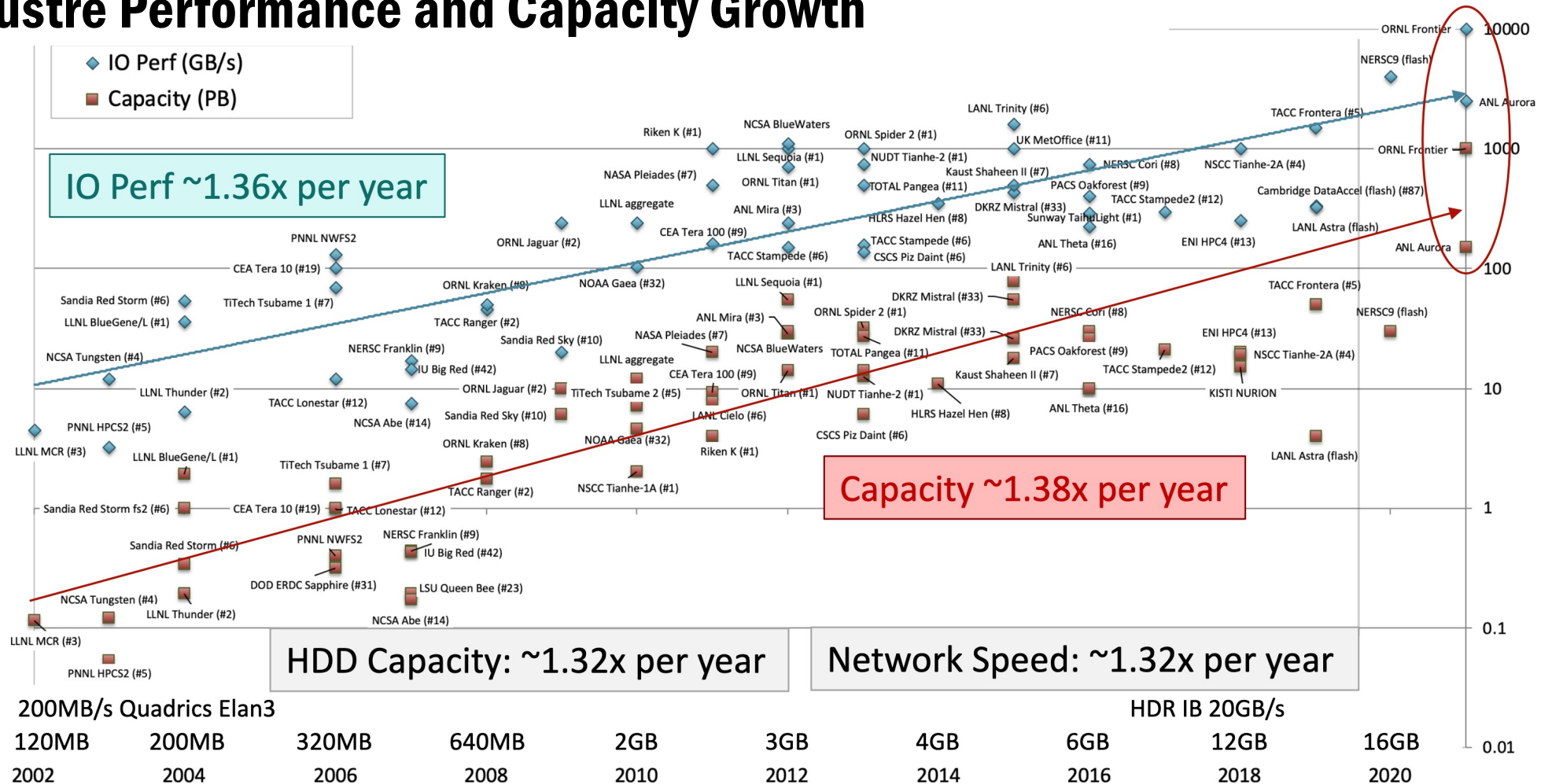
Shuichi Ihara

History of Lustre

Shows the robustness of open source technology in the face of organizational changes

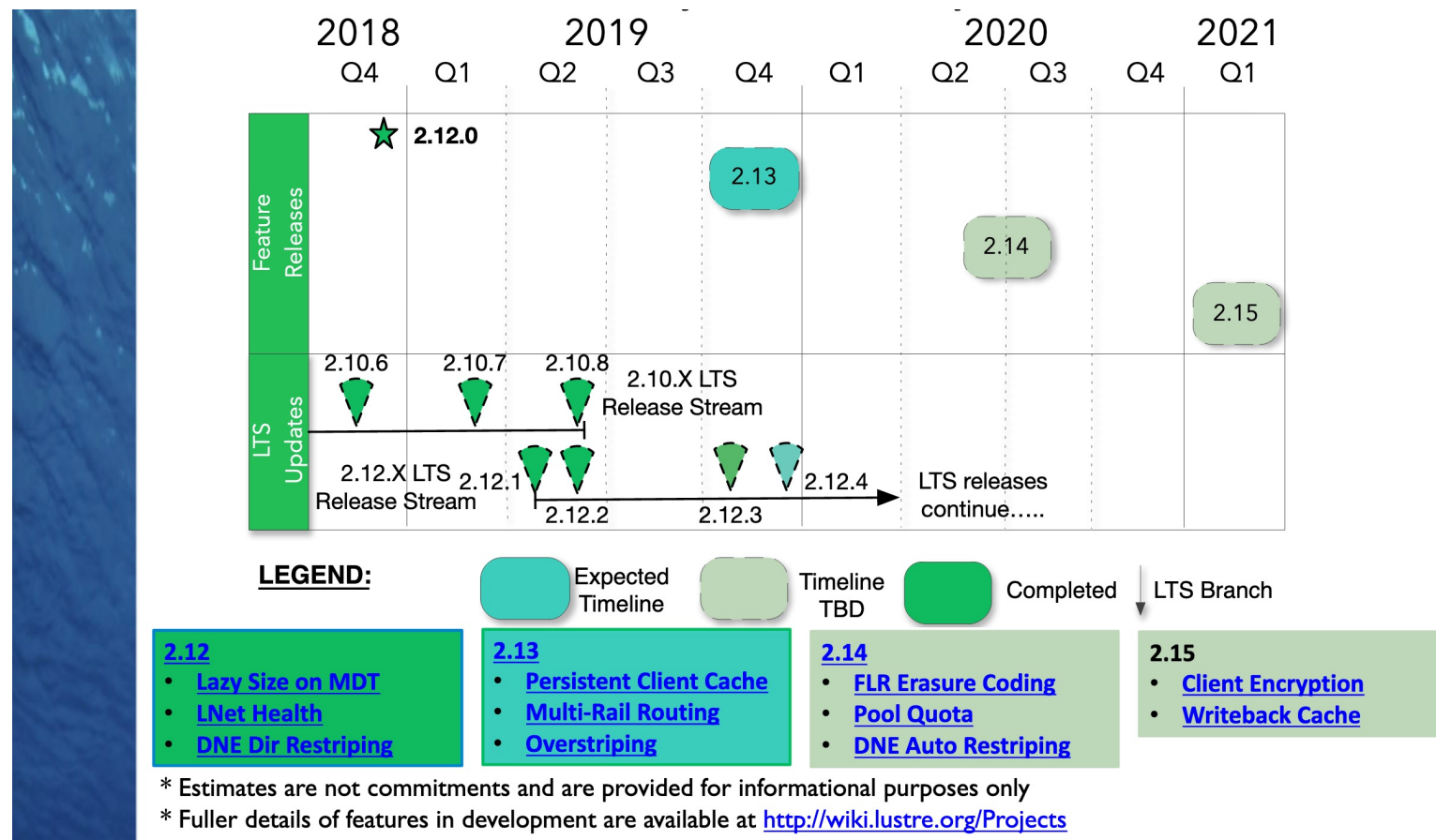


Lustre Performance and Capacity Growth



Source: [Rock Hard Lustre](#), Nathan Rutman, Cray (with updates for recent years); [Disk Drive Prices \(1955-2019\)](#), John C. McCallum

Lustre Community Loadmap

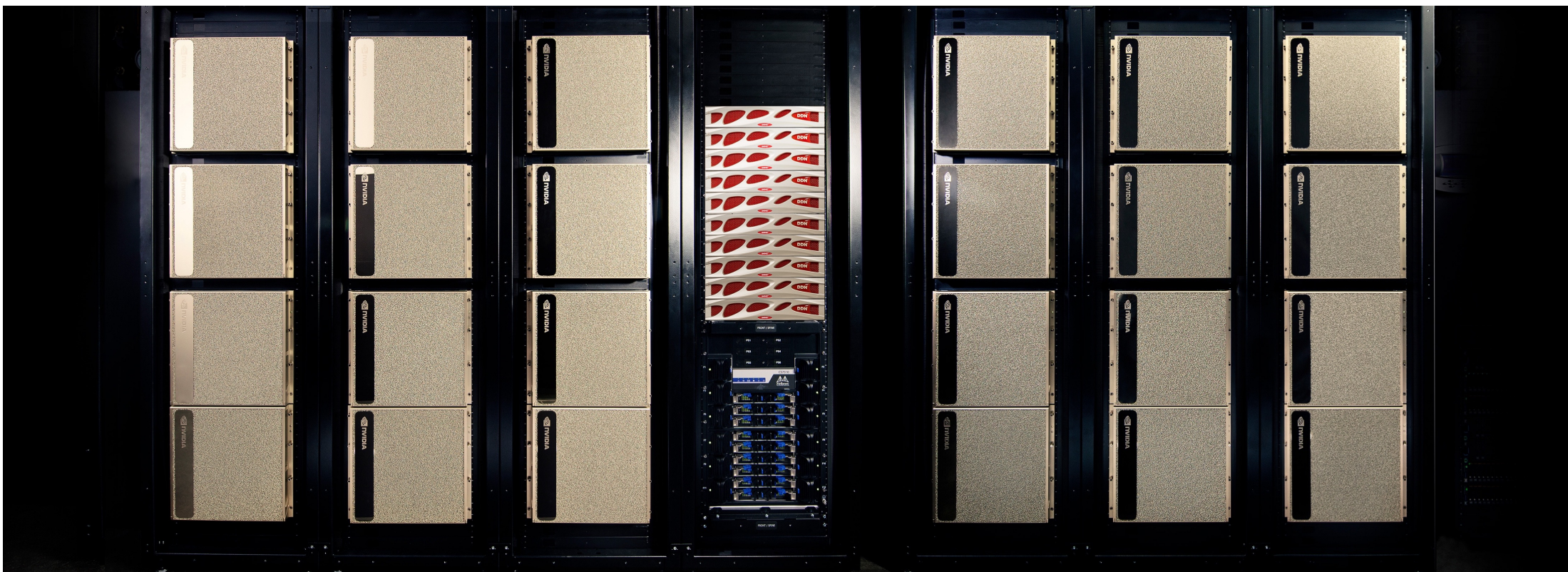




DDN Exascaler

DDN AI400~~X~~

THE SUPERPOD ACCELERATOR



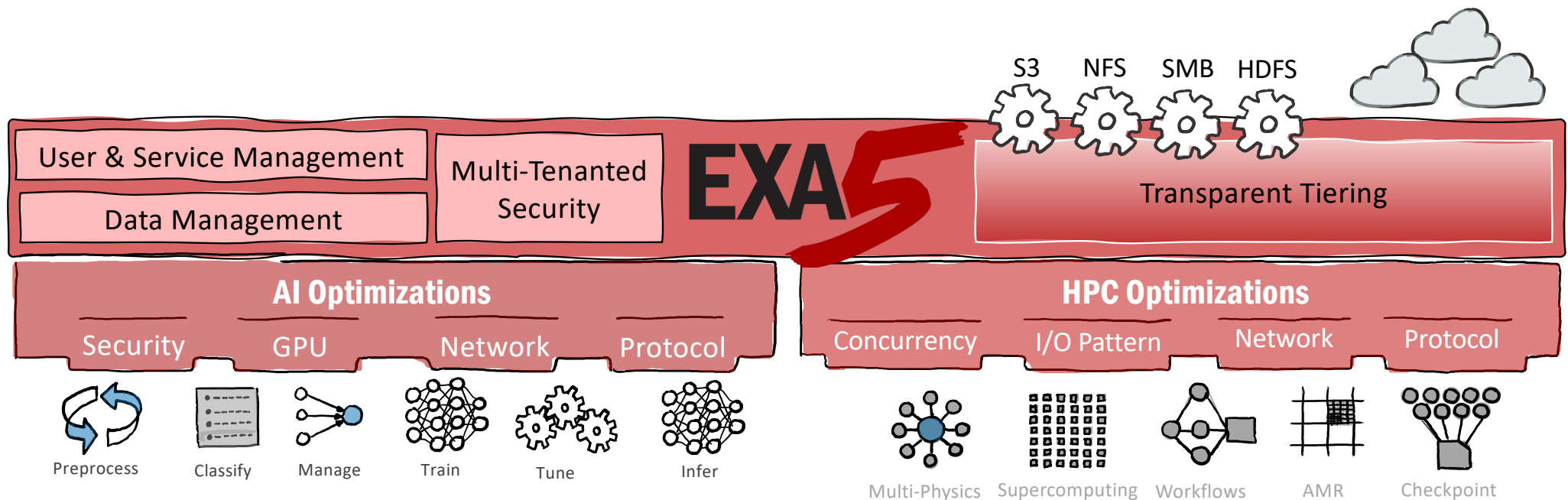


**OVER 25 NEW
FEATURES!**

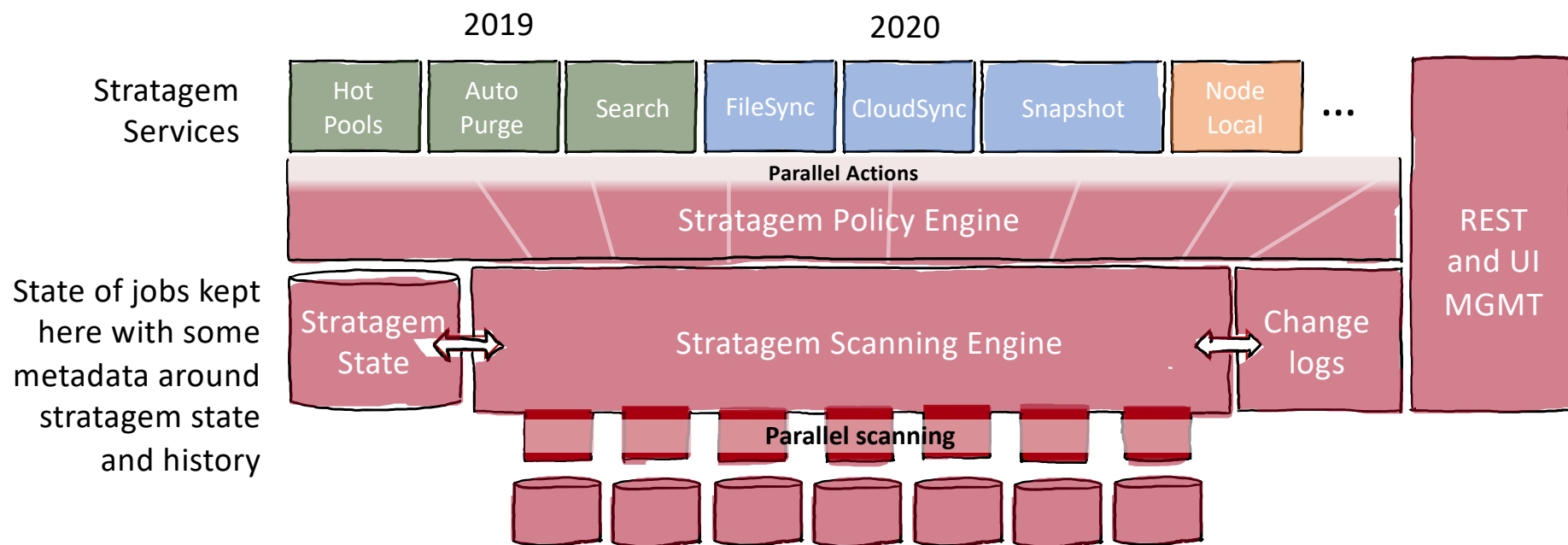
STRATAGEM	HOT POOLS	TRIM/UNMAP	NFS/CIFS DATA SERVICES	A ³ I SOLUTIONS
DATAFLOW INTEGRATION	WORKLOAD ANALYTICS	EXA5 IN THE CLOUD	CONTAINER NATIVE CLIENT	DDN INSIGHT
SINGLE CLIENT PEAK PERF	ENHANCED MULTI RAIL	DATA ON METADATA	DIRECTORY RESTRIPING	PROGRESSIVE FILE LAYOUT
LAZY SIZE ON MDT	T10DIF END-TO-END	IML	LOCK AHEAD	AUDIT FACILITIES
NVIDIA GPUDIRECT TO STORAGE	PERSISTENT CLIENT CACHE	MULTIRAIL ROUTING	OVERSTRIPING	SELF EXTENDING LAYOUTS

The Intelligent, Optimized Environment for AI and HPC

- ▶ Deep Optimizations for both AI and HPC delivers for the highest efficiencies and the right capabilities
- ▶ *Your data in the right place at the right time*

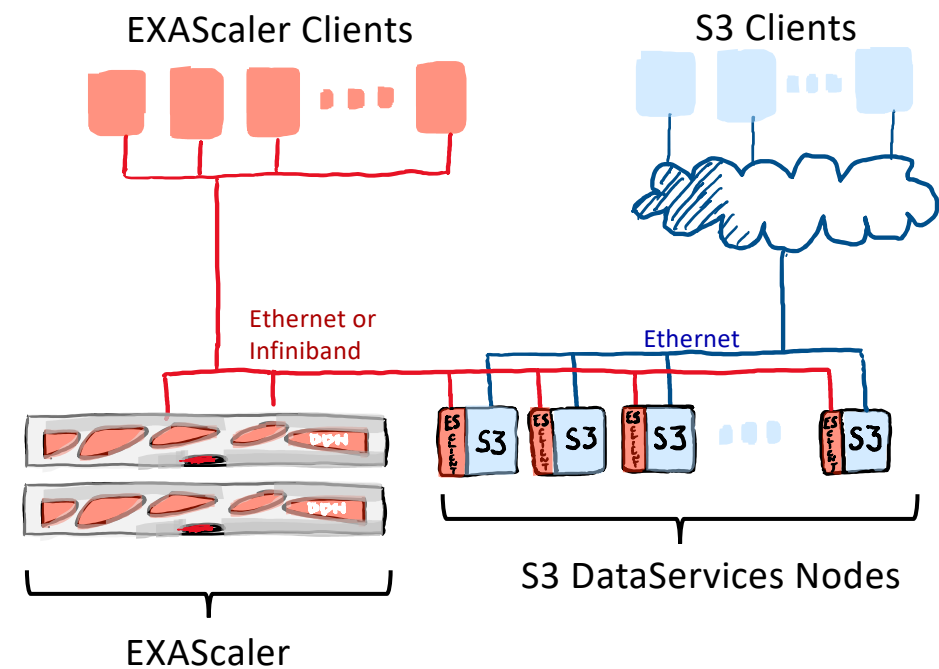


EXA5 Stratagem Data Management Architecture



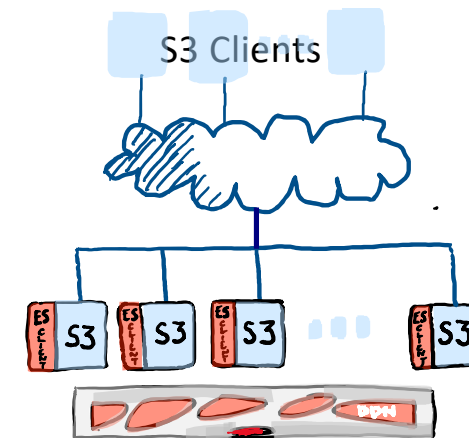
EXAScaler S3 Data Services

- ▶ Highly Available Scale-Out S3 Data Services With Single S3/File Namespace
- ▶ Initial v1 Release Requires at least 2 External Data Services Nodes (min 32GB RAM, 8 Core (TBD))
- ▶ Data stored via the S3 API shall be stored on EXAScaler as files in the file system.
 - **V1:** PUT/GET through S3 and read-only through EXAScaler File interface.
 - **V2:** adds write capability through EXAScaler as Files so that files are accessible as S3 objects
- ▶ Files located in EXA at
~~"/lustre/ddncloudStorage/<bucket>/<filename>..."~~
"/lustre/ddncloudStorage/<ACCESSKEY>/<bucket>/<filename>..."
- ▶ S3 object metadata shall be written into EXAScaler as extended attributes so that they are searchable through filesystem tools.
- ▶ Management REST API for system admin



S3 Data Services Performance

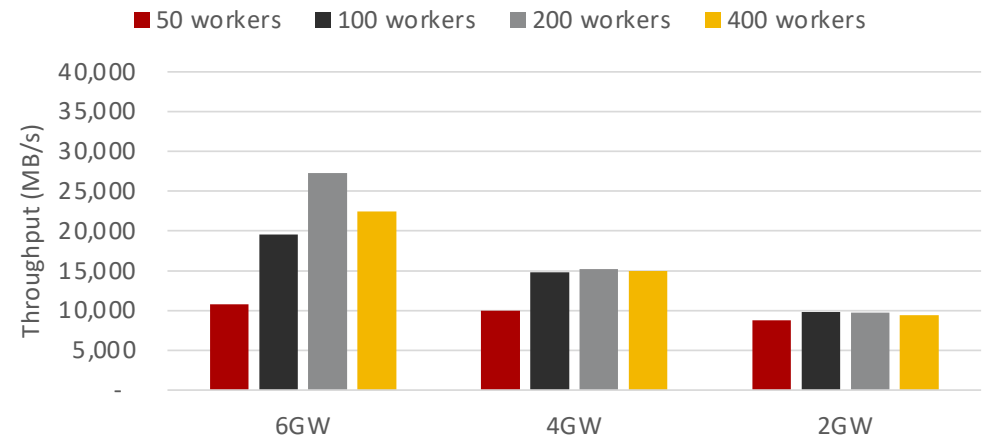
- ▶ Single Server Delivers around 5 GB/s for Read and for Write
- ▶ Scaling up to 6 servers gets over 30GB/s for Read and 25 GB/s for Write



S3 Data Services Node Scaling - READ



S3 Data Services Node Scaling - WRITE



Examples – S3 DataService

Data access from S3 and Lustre Native Client

▶ Assumed two users "user1" and "user2" on Lustre namespaces

- Lustre mounted /share0 on clients.
- Each user have different access key and secret key on S3 Dataservice.
- Configured those keys with S3 application (e.g. s3cmd or AWS CLI)

▶ Create new S3 bucket and PUT data into bucket

```
user1$ s3cmd mb s3://mybucket
```

```
user1$ s3cmd put testfile s3://mybucket/testfile
```

▶ Access file from Lustre Namespace

```
user1$ ls -l /share0/ddncloudStorage/mybucket/testfile
```

```
-rw-r--r-- 1 user1 users 104857600 Feb  6 17:04 /share0/ddncloudStorage/mybucket/testfile
```

▶ Set ACLs to share data to other users

```
user1$ s3cmd setacl --acl-public s3://mybucket/testfile (ACL set to Public)
```

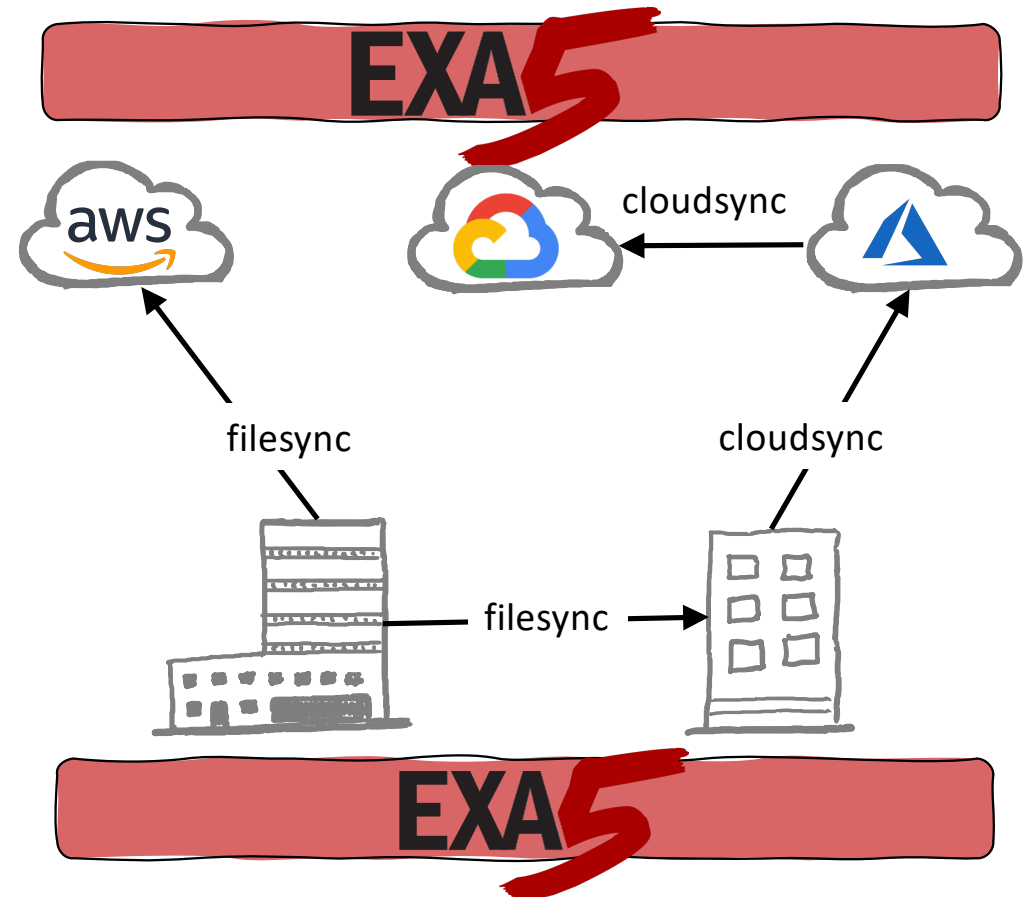
```
user1$ s3cmd setacl --acl-grant=read:fb9dc133a3cf2775a6a159c316a92cfb6b1162f3 s3://mybucket/testfile (ACL to user2)
```

▶ GET data by other user

```
user2$ s3cmd get testfile s3://mybucket/testfile
```

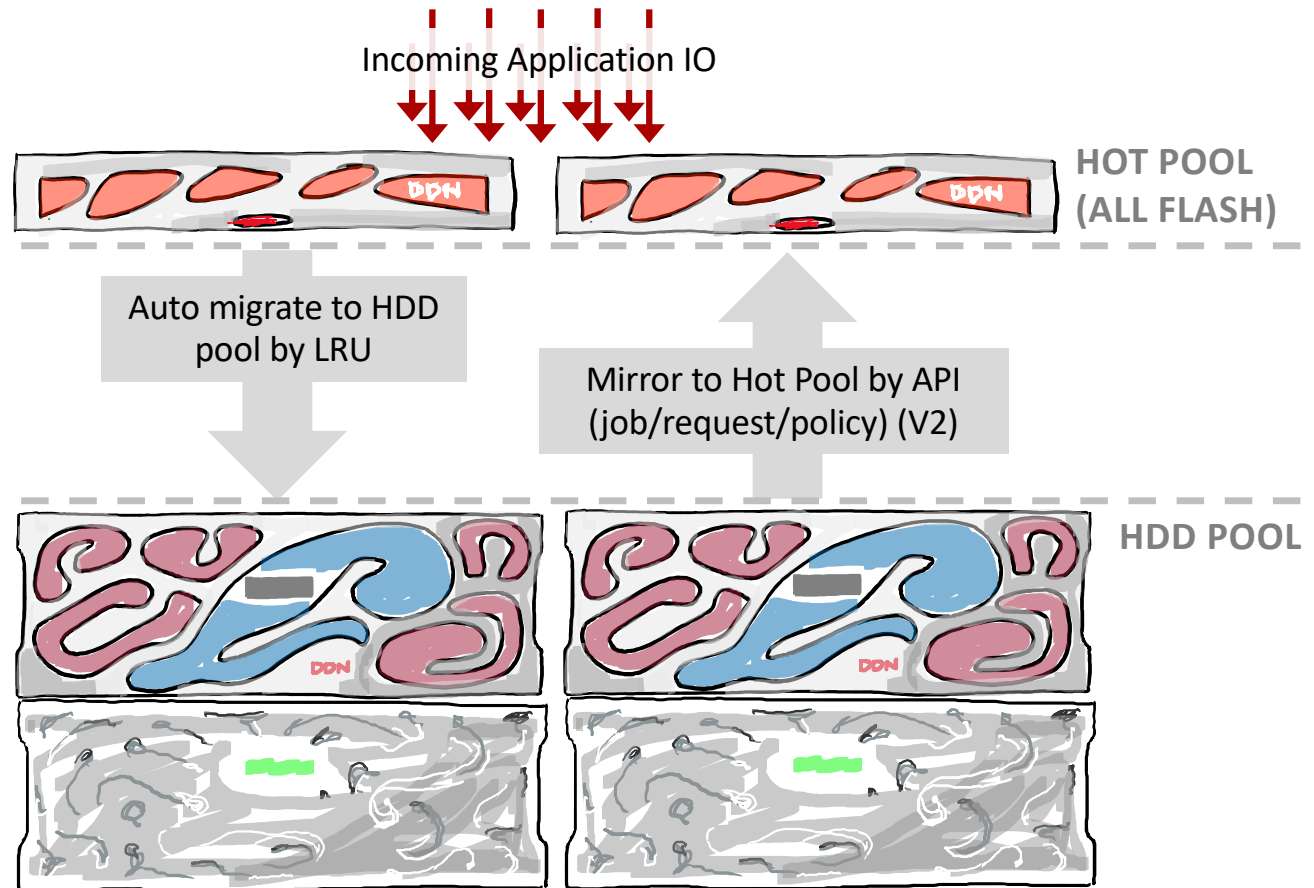
Enabling Hybrid cloud workflow implementations

- ▶ *Stratagem Data Management brings the ability to track and move your data between filesystems and object stores*
- ▶ *Dump your filesystem to S3 for Archive or Data Transfer*
- ▶ *Spin up new Filesystems from S3 at speed reducing your long term cloud costs*

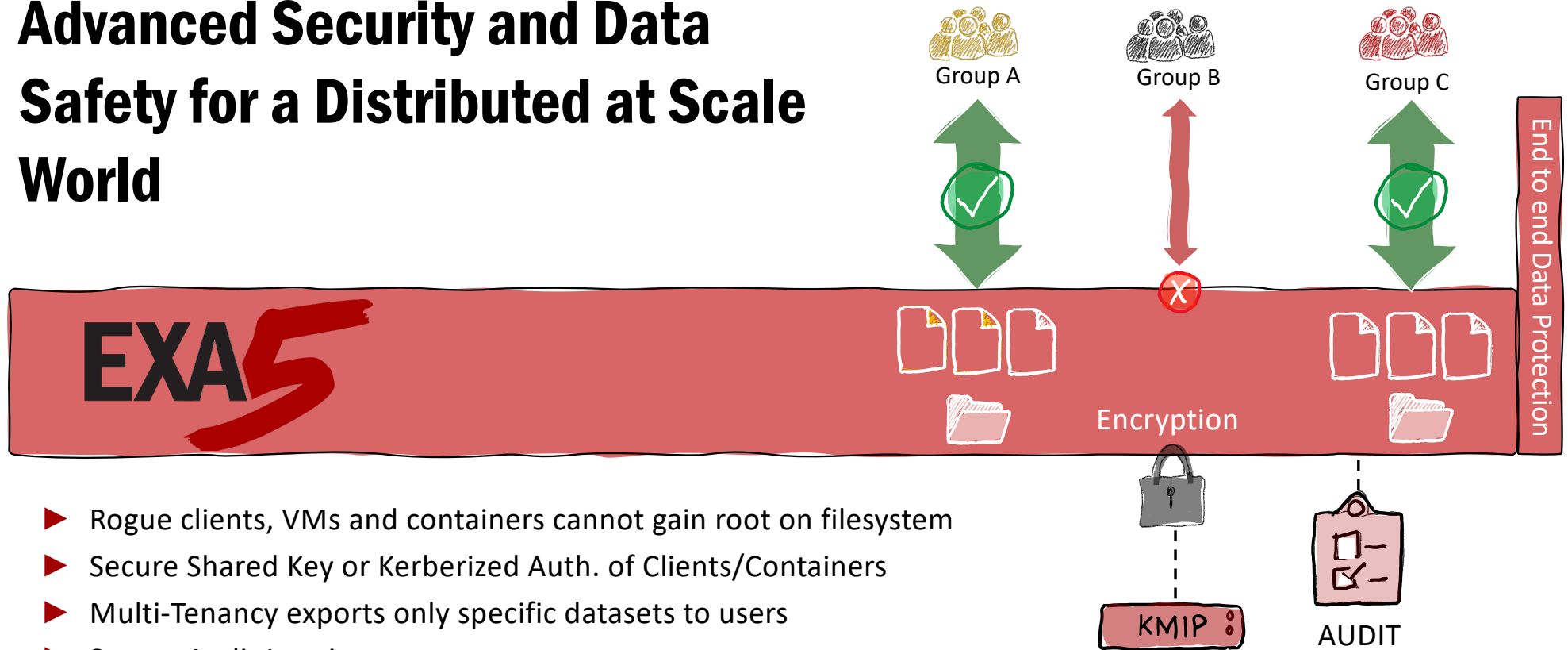


Hot Pools: At Scale Transparent Tiering

- ▶ Hot Pools V1 actively migrates old, least accessed files from flash tier to HDD tier
- ▶ API provided for use cases with externally driven data movement (e.g. scheduler directives)
- ▶ Hot Pools V2 adds capability for also promoting hot objects in the HDD pool into SSD pool
- ▶ Hot Pools uses Stratagem File Scanning & File Level Replication



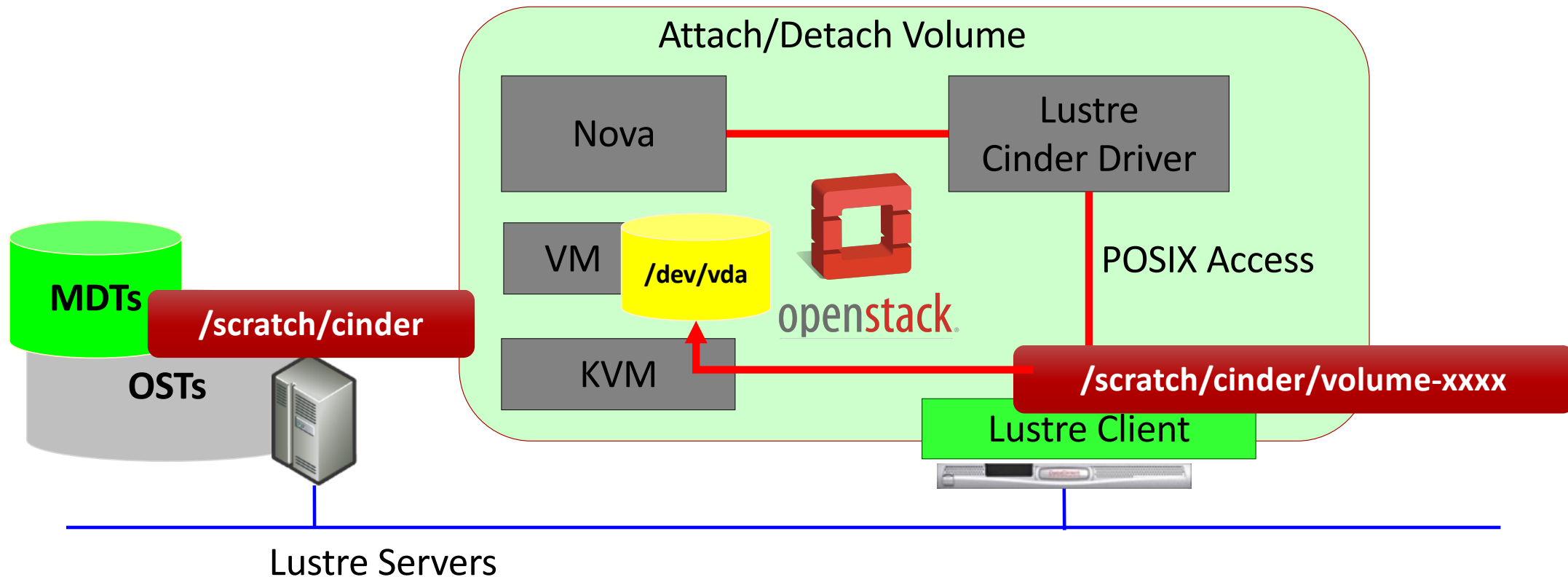
Advanced Security and Data Safety for a Distributed at Scale World



- ▶ Rogue clients, VMs and containers cannot gain root on filesystem
- ▶ Secure Shared Key or Kerberized Auth. of Clients/Containers
- ▶ Multi-Tenancy exports only specific datasets to users
- ▶ Secure Audit Logging
- ▶ Data at Rest Encryption for Drives and NVMe*

Storage Platform for Virtual Machine Environment

Lustre Cinder driver



How Lustre Cinder driver works(1)

► Cinder Configuration (/etc/cinder/cinder.conf)

```
[lustre]
```

```
volume_driver = cinder.volume.drivers.lustre.LustreDriver
```

```
lustre_share_host = 10.0.10.193@o2ib30:10.0.10.192@o2ib30
```

```
lustre_share_path = /scratch/cinder
```

```
volume_backend_name = lustre
```

► Lustre automatically mounted for OpenStack

```
[root@devstack~]# mount -t lustre
```

```
10.0.10.193@o2ib30:10.0.10.192@o2ib30:/scratch/cinder on
```

```
/opt/stack/data/cinder/mnt/71ee0200412a18cf142a396734dbb1a4 type lustre  
(rw,lazystatfs)
```

How Lustre Cinder driver works(2)

► Enabled Lustre Cinder Driver

```
[root@devstack~]# openstack volume service list
```

Binary	Host	Zone	Status	State	Updated At
cinder-backup	devstack	nova	enabled	up	2017-05-21T22:39:31.000000
cinder-scheduler	devstack	nova	enabled	up	2017-05-21T22:39:36.000000
cinder-volume	devstack@lustre	nova	enabled	up	2017-05-21T22:39:30.000000

► Volume Creation

```
[root@devstack~]# openstack volume create --size 1024 --image CentOS7.3 ¥ devstack-vm01-vda
```

IO-500 on Containerized Lustre Client

10 nodes challenges (SC19)

Test Case	Baseline	After Optimization
ior_easy_write	25.881 GB/s	28.618 GB/s
mdtest_easy_write	143.882 kiops	152.840 kiops
ior_hard_write	2.784 GB/s	2.960 GB/s
mdtest_hard_write	32.326 kiops	79.646 kiops
find	1735.410 kiops	810.420 kiops
ior_easy_read	39.941 GB/s	41.723 GB/s
mdtest_easy_stat	455.032 kiops	451.974 kiops
ior_hard_read	8.985 GB/s	42.190 GB/s
mdtest_hard_stat	44.918 kiops	449.926 kiops
mdtest_easy_delete	88.524 kiops	132.756 kiops
mdtest_hard_read	20.410 kiops	172.593 kiops
mdtest_hard_delete	16.354 kiops	75.146 kiops
SCORE	B/W 12.6807 GB/s IOPS 91.4099 kiops TOTAL 34.0462	B/W 19.65 GB/s IOPS 207.632 kiops TOTAL 63.8746

