UNIVERSITY OF CAMBRIDGE

# Recent Developments and Lessons from Lustre

at the University of Cambridge

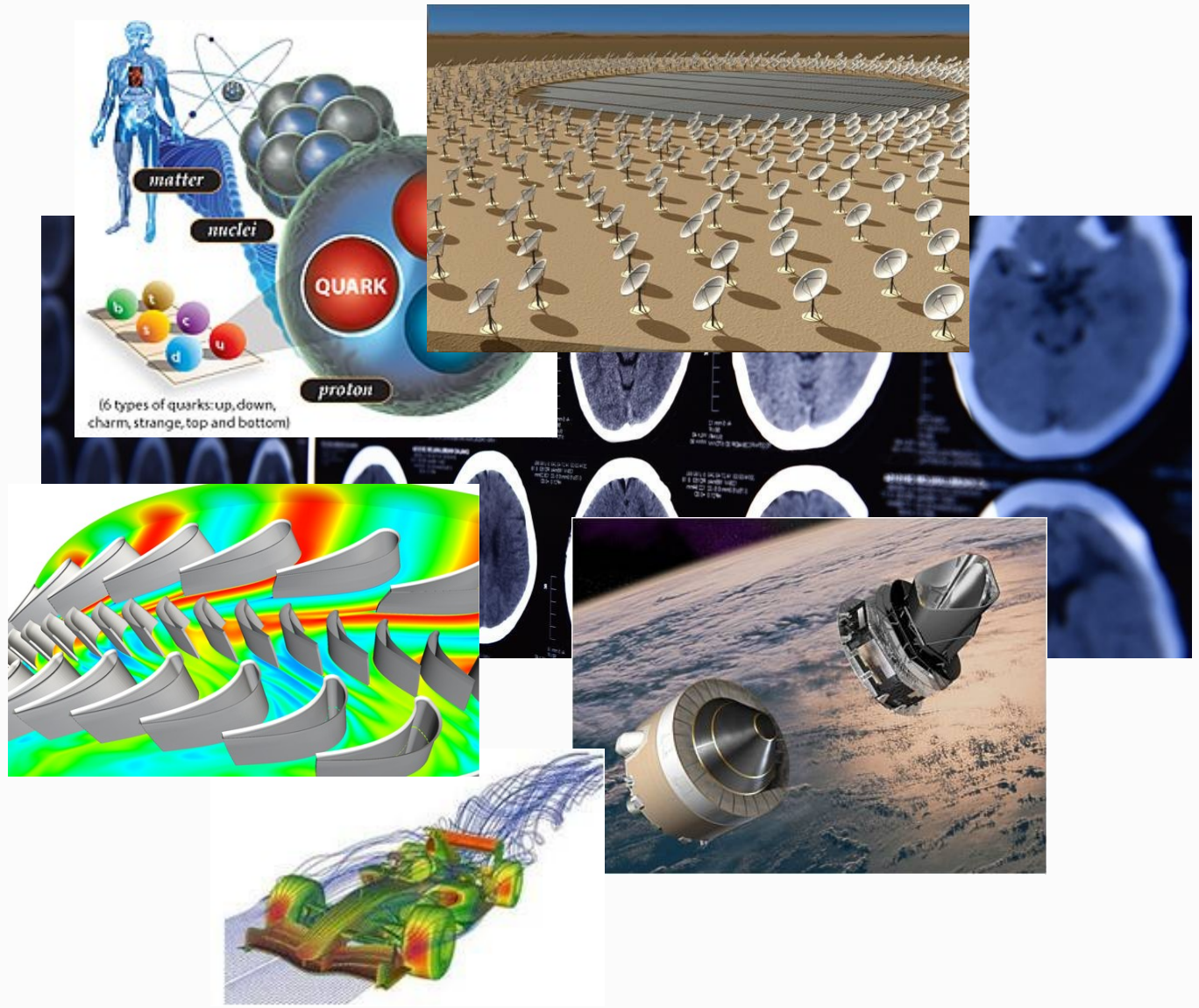
UNIVERSITY OF CAMBRIDGE

# Background

We are Research Computing Services at the University of Cambridge

We serve:

- Cambridge users and students
- Several national research projects (IRIS, DIRAC, UKAEA)
- SKA project users
- UK AI Research Resource



UNIVERSITY OF CAMBRIDGE

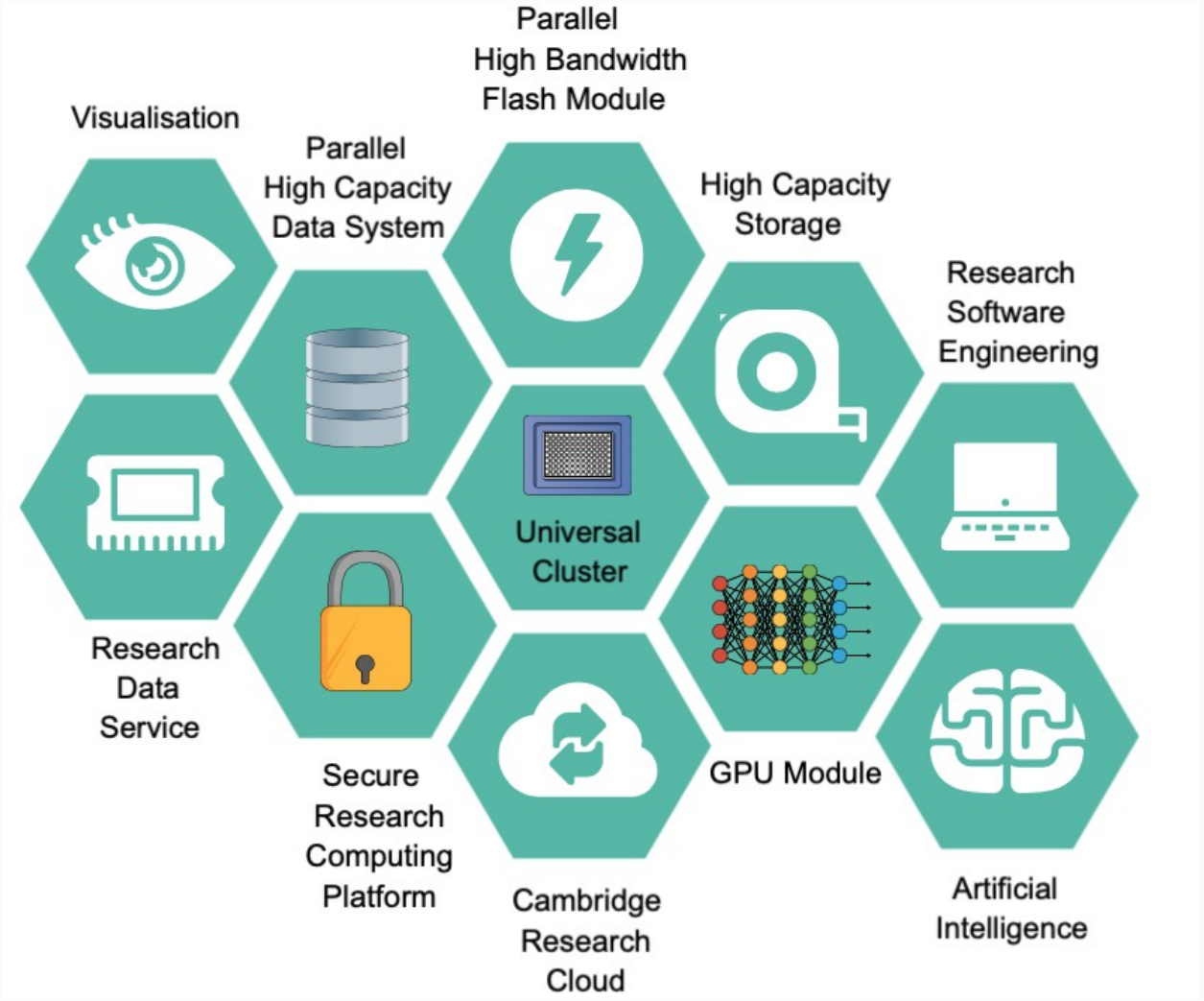
# Background

Services include:

- “Traditional” HPC workloads via SLURM
- IaaS clients via Openstack
- Trusted Research Environments

Served by:

- ~9Pb Isilon
- ~14Pb Tape
- ~8Pb Ceph
- 40Pb Lustre



UNIVERSITY OF CAMBRIDGE

# War Stories

1) Changelogs

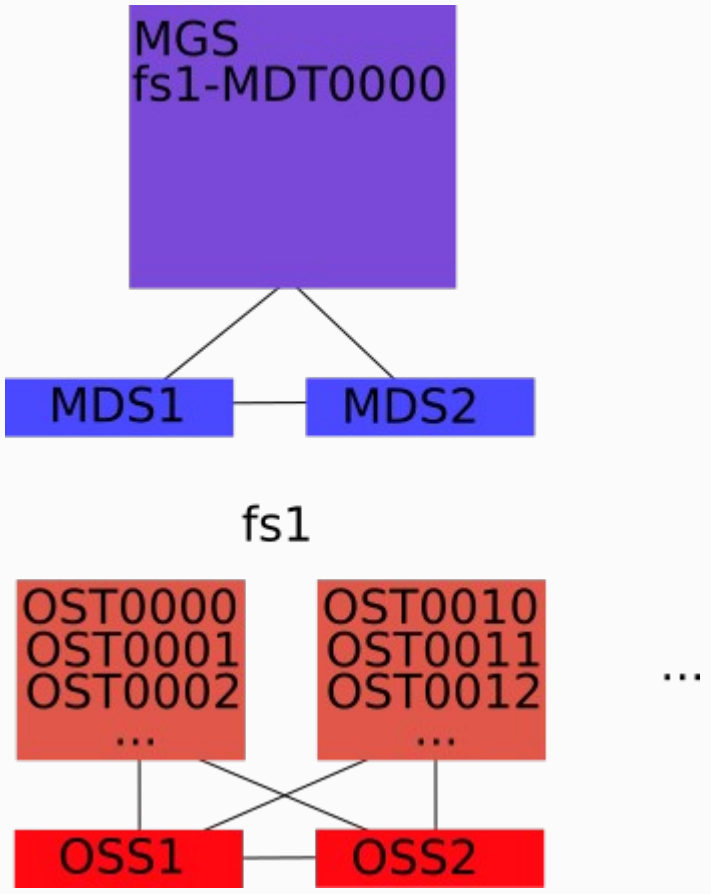
2) Burstbuffer

3) Openstack

UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs

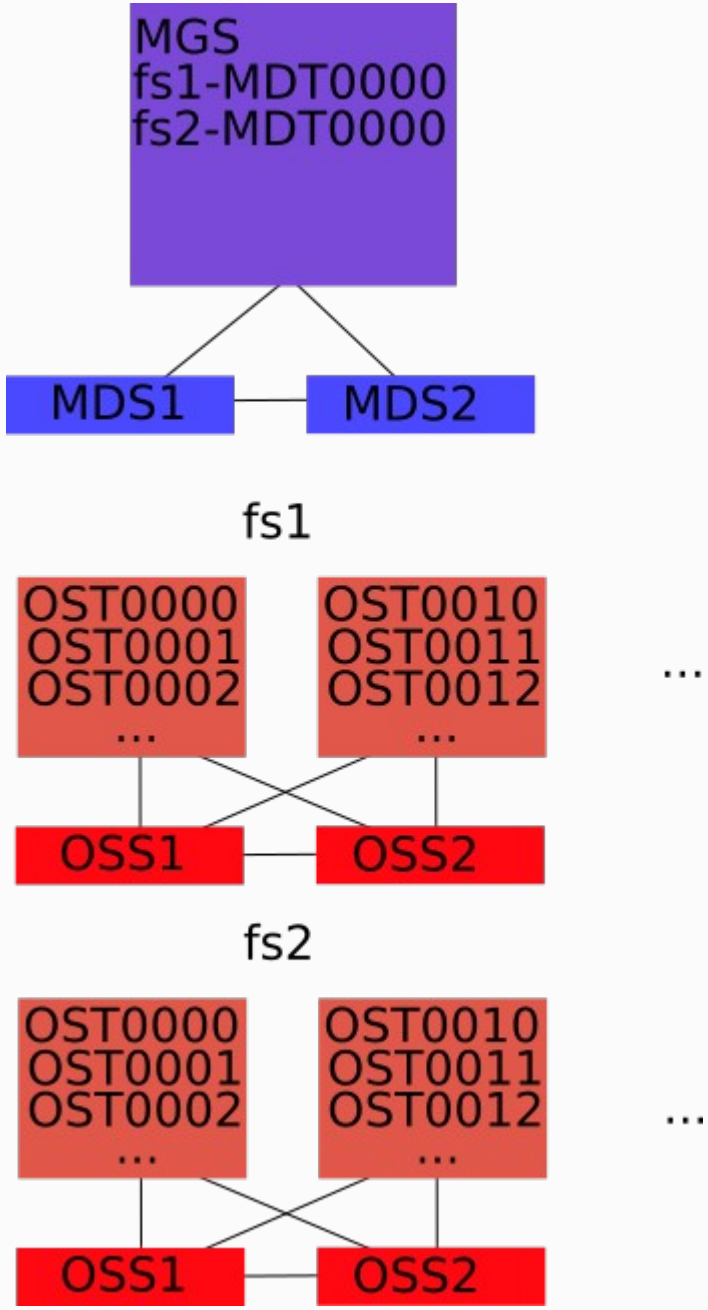
- Initial server deploy @2.7.X
- Size grows as needed for capacity
- No projid, so quota tracked by robinhood



UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs

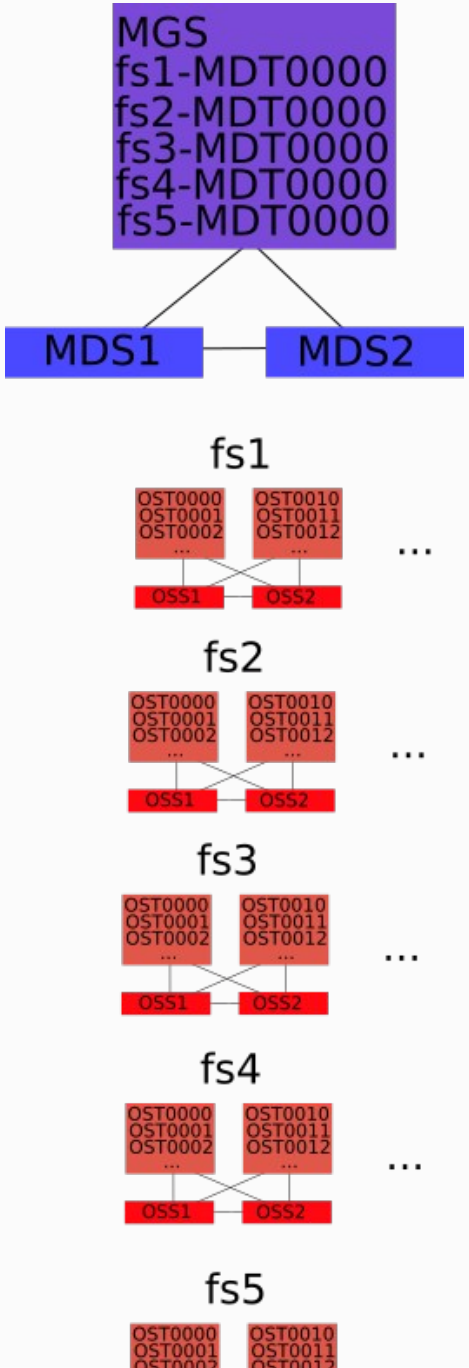
- Problem: Robinhood DB performance insufficient
- Expand by adding new filesystem
- Two DBs, split the workload



UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs

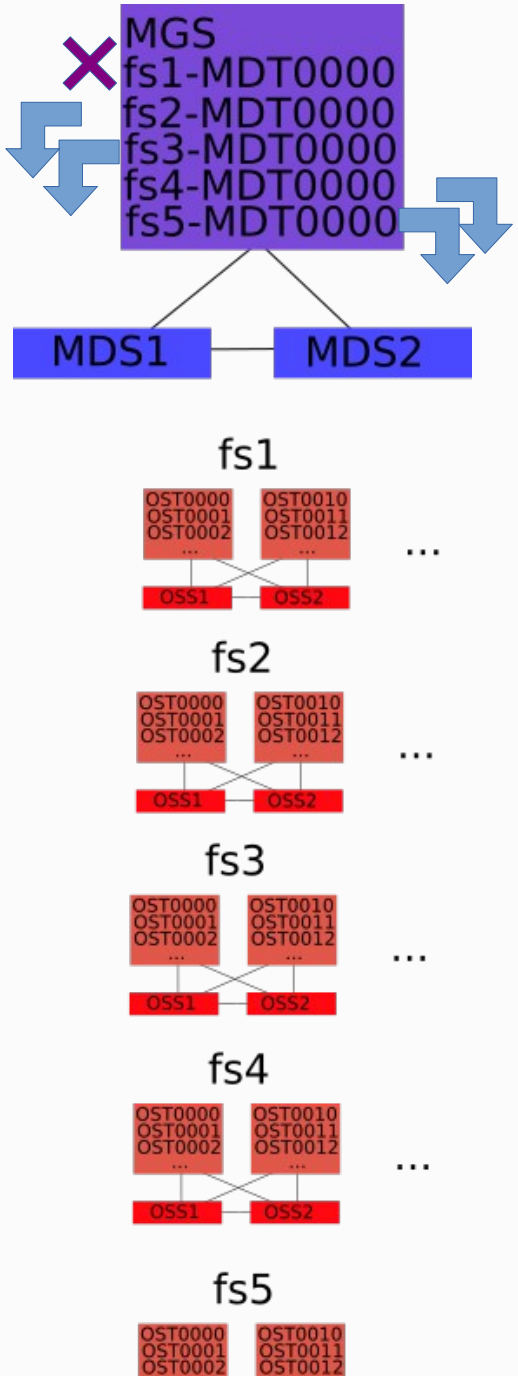
- Natural conclusion:
- Five filesystems, one MDT, one MDS pair
- One got removed due to aging out hardware
- System upgraded to 2.8, 2.10, 2.12



UNIVERSITY OF CAMBRIDGE

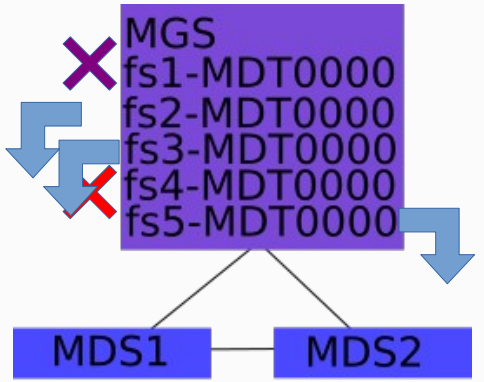
# War Stories 1: Changelogs

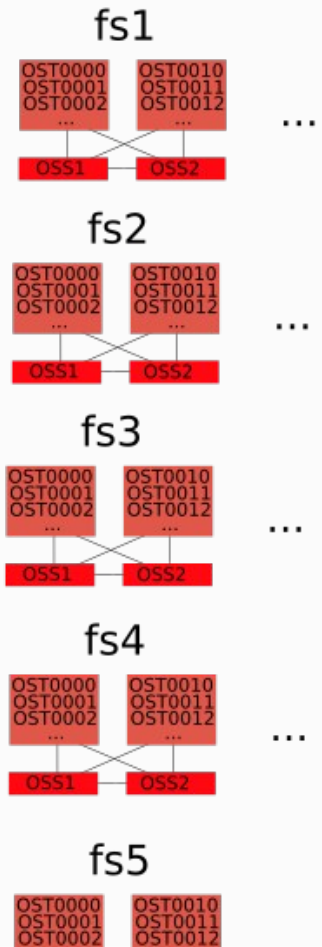
- Today's situation:
- 4 filesystems
  - Approx 2Pb each
- Usually split load on each MDS



UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs



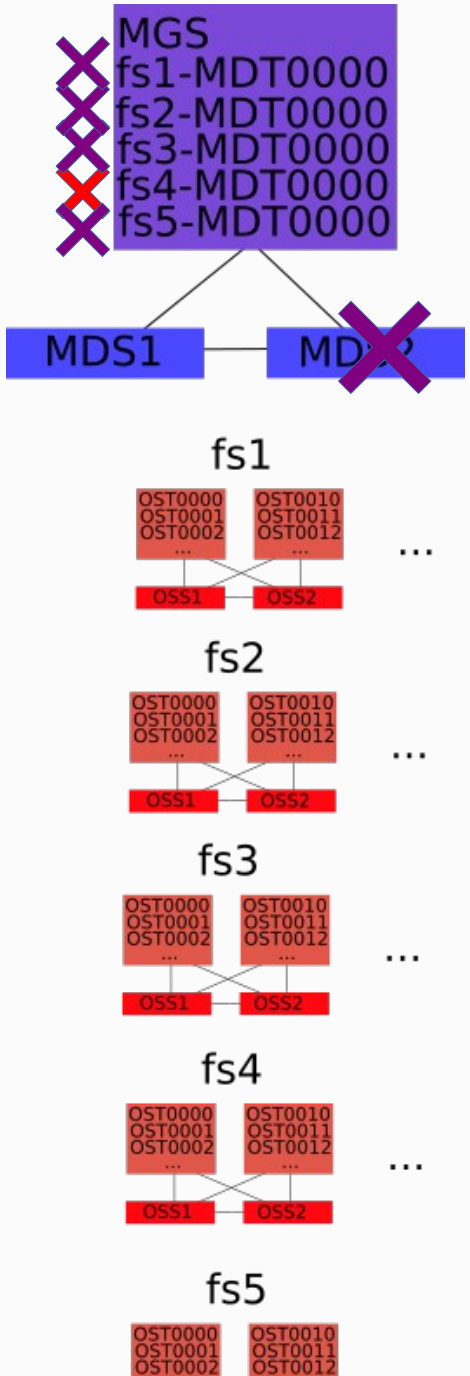

```
• Mar  8 10:39:36 mds2 kernel: LustreError: 14439:0:(mdd_dir.c:1065:mdd_changelog_ns_store())
  fs4-MDD0000: cannot store changelog record: type = 1, name = 'FIN_0001450000', t =
  [0x2000829e7:0x133de:0x0], p = [0x200083cea:0x50eb:0x0]: rc = -28
• Mar  8 10:39:36 mds2 kernel: LustreError: 14439:0 (mdd_dir.c:1065:mdd_changelog_ns_store())
  Skipped 58 previous similar messages
```

- errno.h lookup:
- 28 = ENOSPC “No space left on device”

UNIVERSITY OF CAMBRIDGE

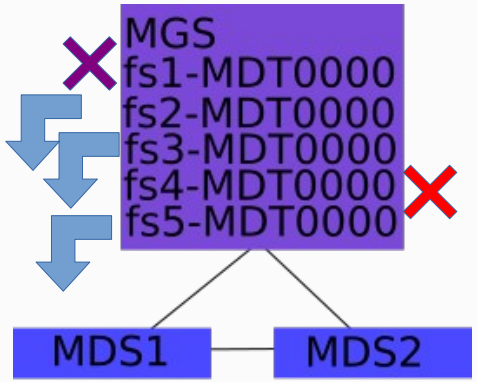
# War Stories 1: Changelogs

- e2fsck mdt
- checked capacity: plenty of space (~1Tb free)
- Retry mount
- Pacemaker kills mds due to timeout
- Force stops other mounts on server
- Attempted diagnostics stop all mounts



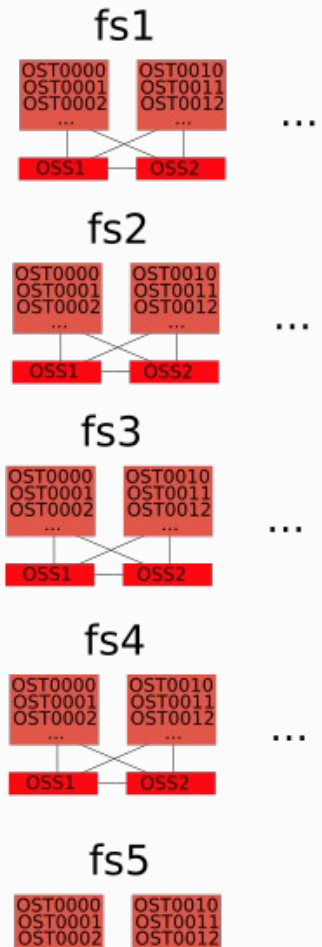
UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs




- Remount all MDTs to MDS1
- Disable pacemaker
- Take backup
- Filesystem seemingly won't mount
- Consider options:
  - Mount as ldiskfs
  - llog_reader on the changelog_catalog file lists
- Wipe the changelog_catalog and all changelog entries?

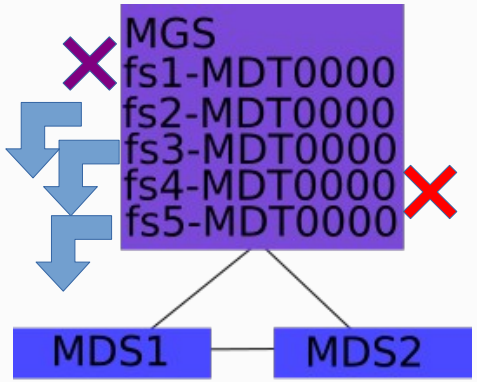
```
• rec #2 type=1064553b len=64 offset 8256
• Header size : 8192       llh_size : 64
• Time : Thu Mar 3 13:35:05 2024
• Number of records: 1    cat_idx: 1
  last_idx: 2
• Target uuid :
• ----------------------
• #02 (064)id=[0x154:0x1:0x0]:0 path=O/1/d20/340
• ...
```

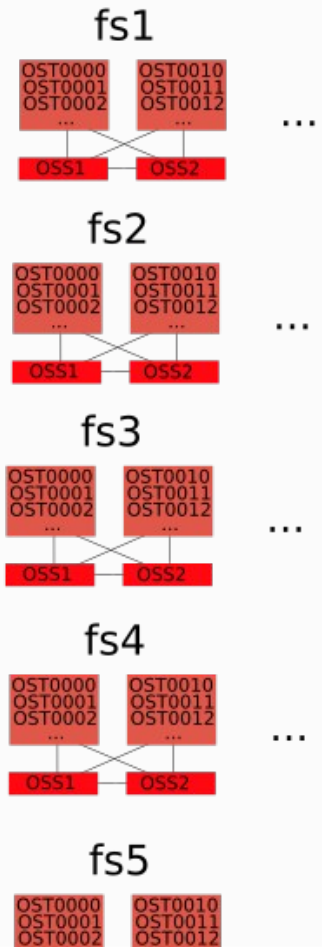


UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs

- One more try
- Mount with
- `-o abort_recov`
- fs4 mounts overnight
- changelog length reports at ~400billion
- Cleared out with
- `lfs changelog_clear`
- No direct intervention needed

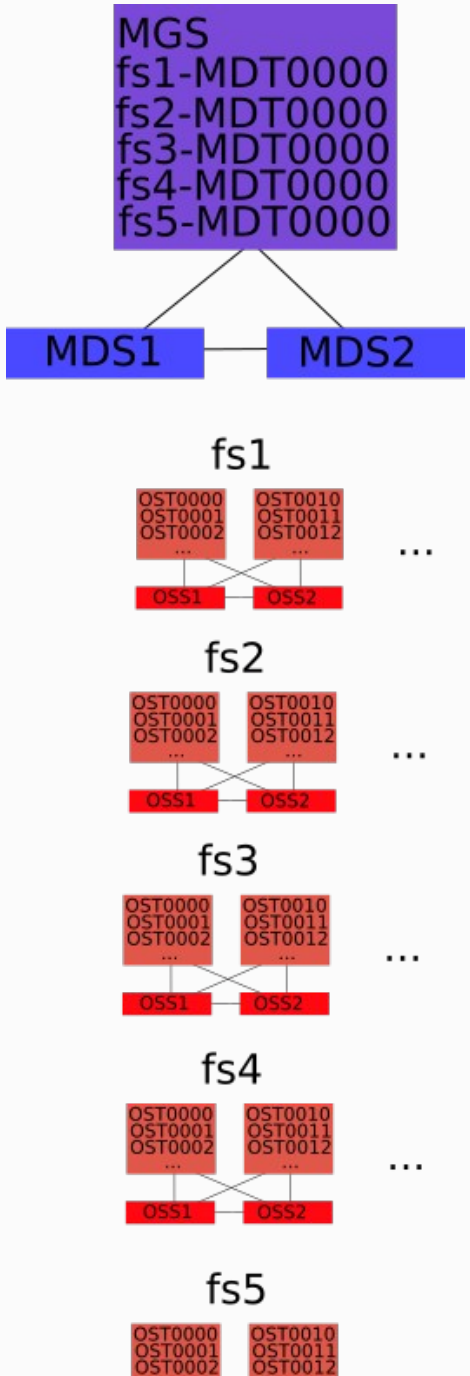


UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs

- Max number of changes in 2.12 is 420 billion
- ~200GB needed

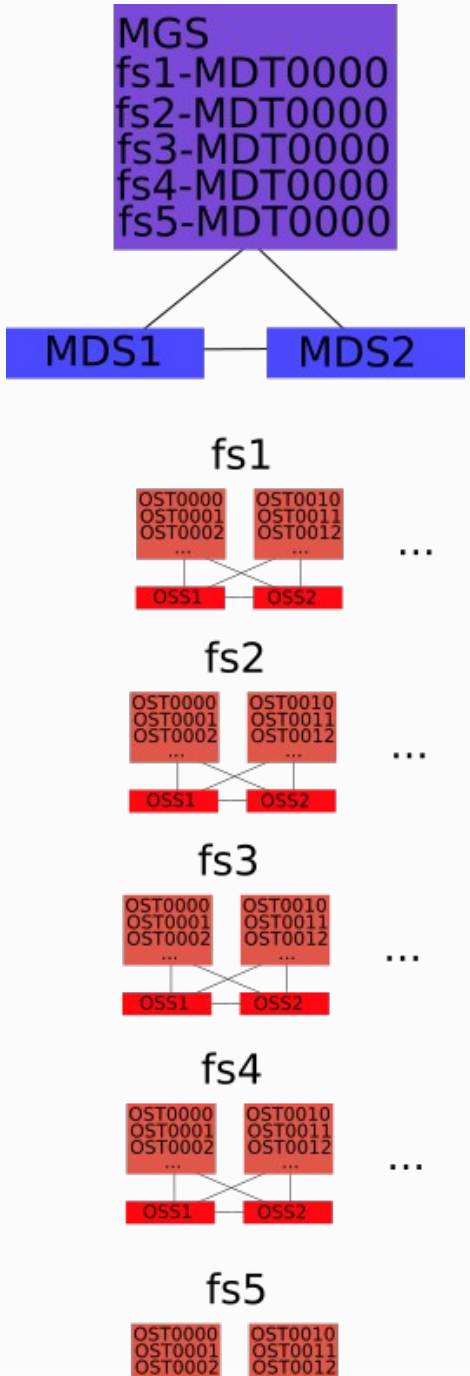


UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs

- Conclusions:
  - Possibly fixed by LU-12871 + LU-14699, but:
- Don’t put many eggs in one basket
- Make sure you have enough space for changelogs
- Monitor changelogs in case of unexpected increases
- Make sure

  `mdd.*.changelog_gc`
- is set to 1!



UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe


- New AI supercomputer

UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

- 4 X 200 Gb/s. 100 GB/s
- 96 CPU cores
- 1TB RAM
- Doing a lot of data processing
- Needs suitable performant filesystem for bursty workloads


- dell.com
- image taken 2025-09-02

UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

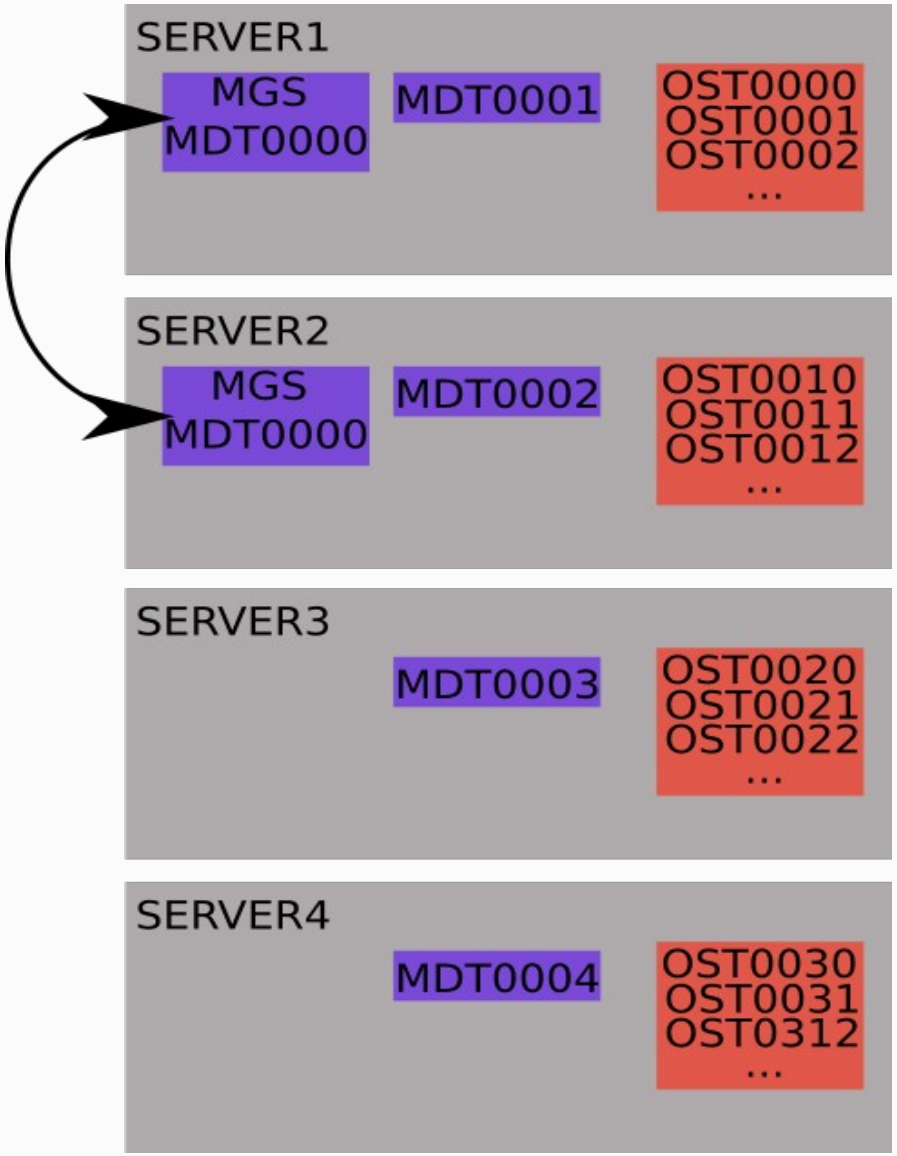
- Lustre built for performance
- Available NVMe's lack resilience
  - JBOF, no switching
- Data not needed to be resilient
- Filesystem uptime important


- serversupply.com
- image taken 2025-09-02

UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

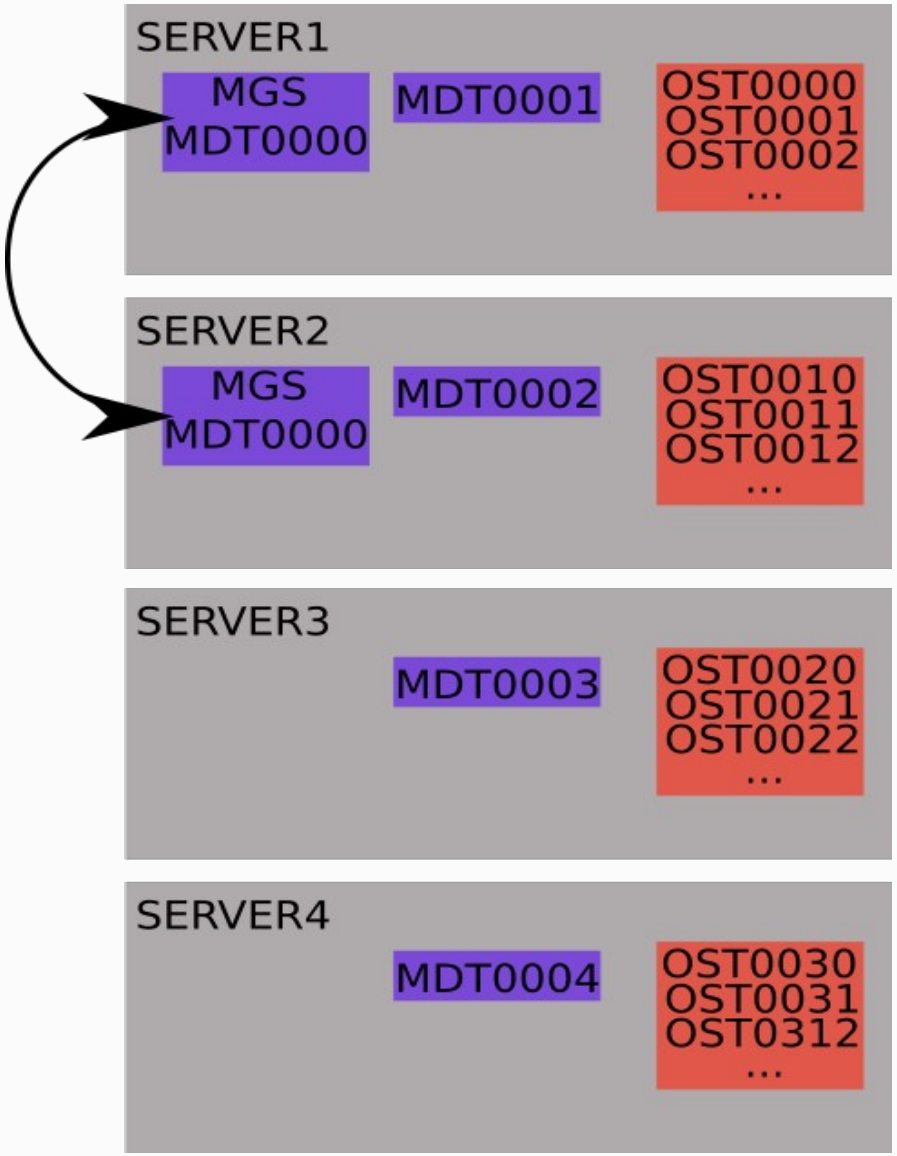
- Only MGS/MDT0000 needed for FS access – replicate these (DRDB, no performance)
- All other MDTs single-pathed
- All OSTs single-pathed
- (Single) failure – drop server out, disable MDT/OSTs in MGS, keep going



UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

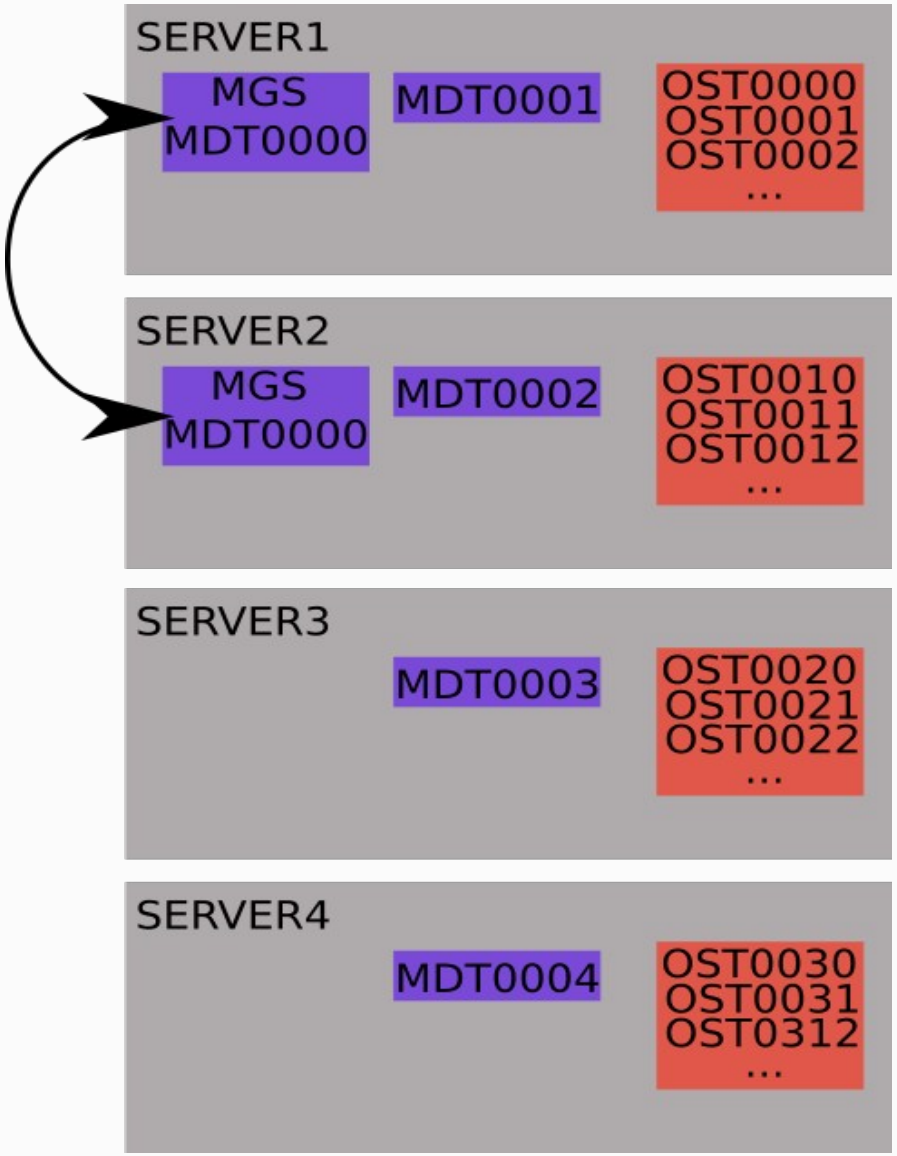
- Each server has 100GB/s (4x200Gb) network to block
- This matches one client's performance per node
  - One client cannot use more than 1 server's worth of bandwidth
- Control placement using pools



UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

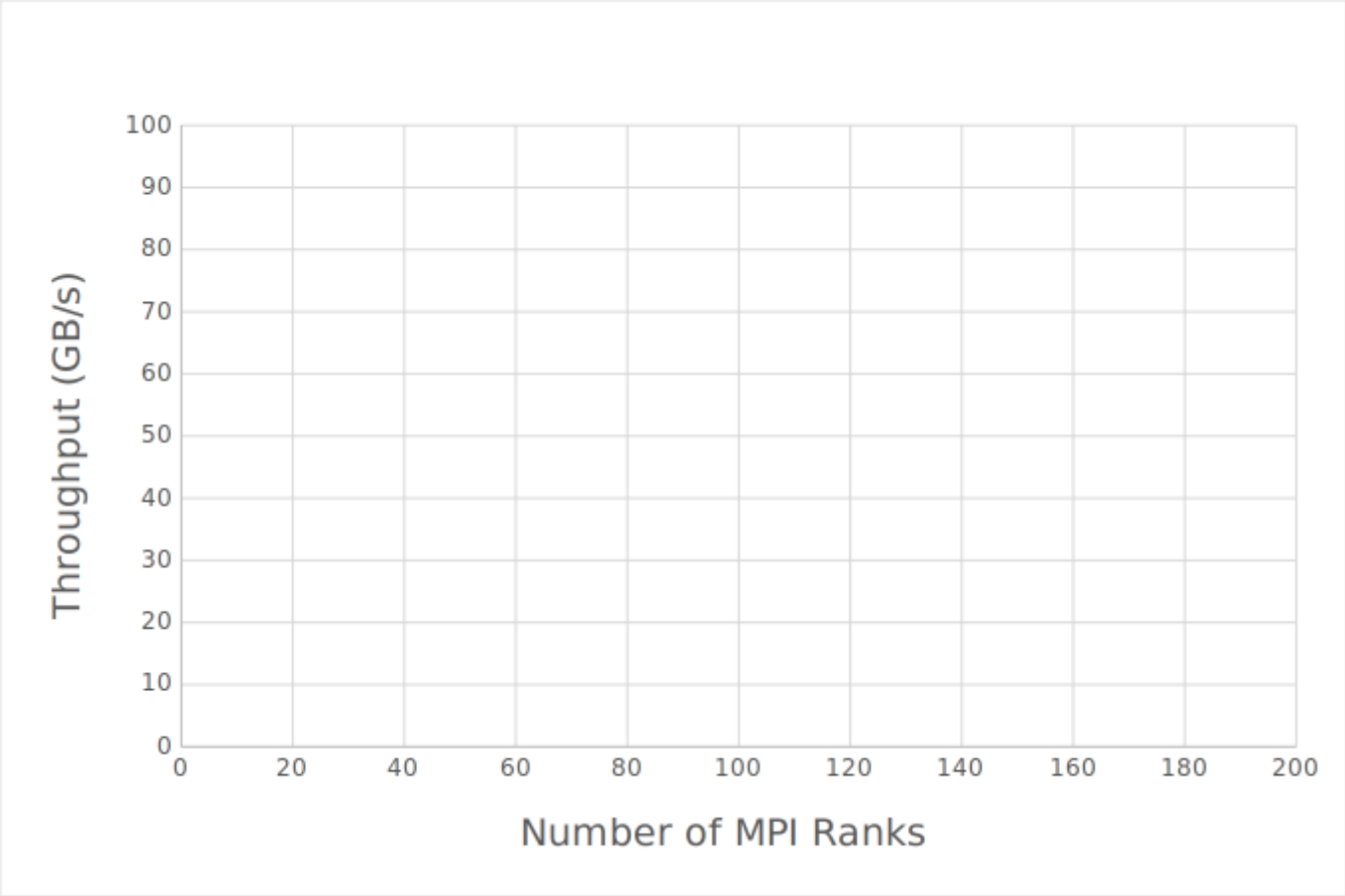
- Control placement using pools
  - Max client llog size bumped into at 64k
  - Cannot have more than 20,000 pool/pool member entries



UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

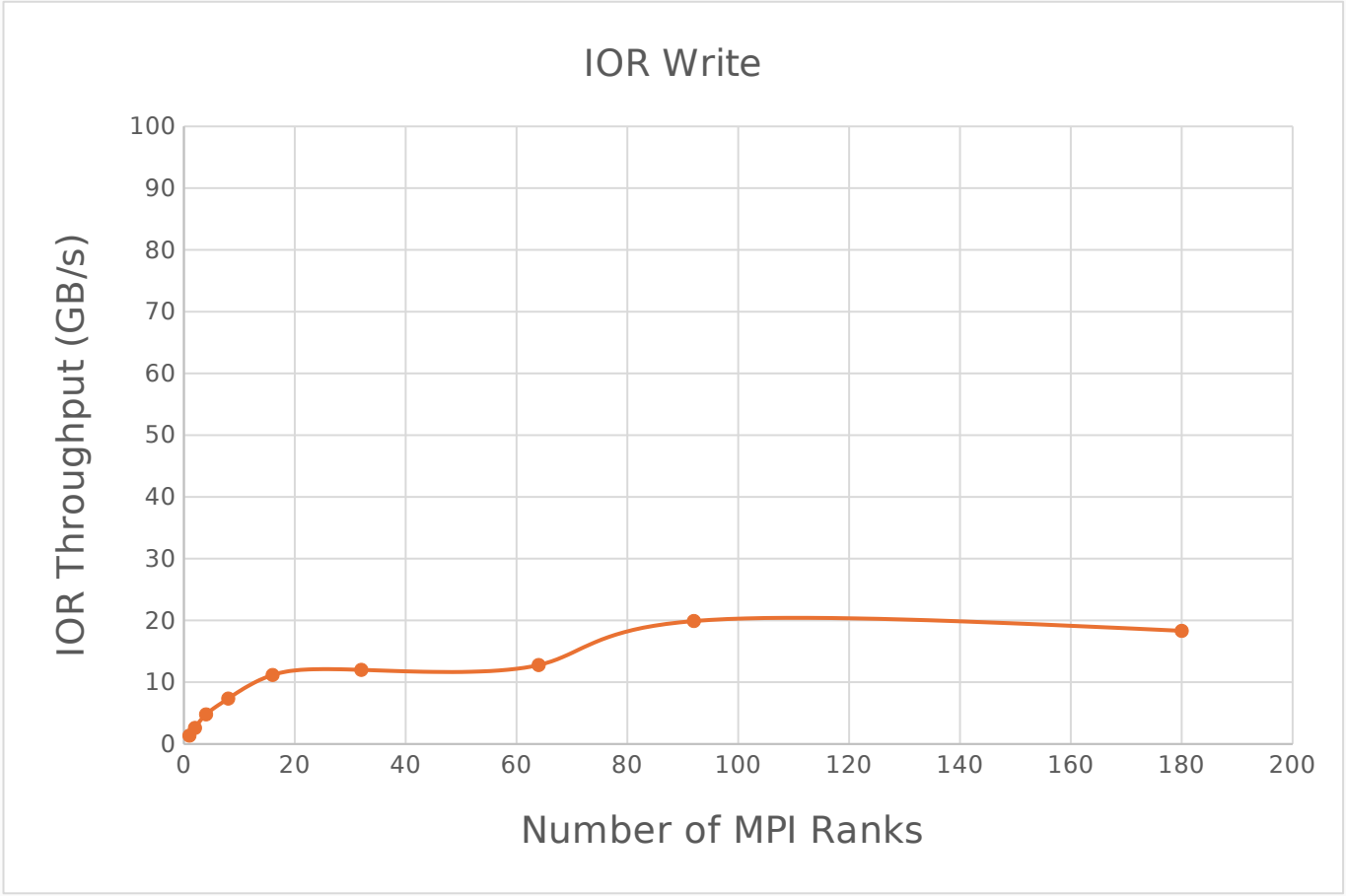
- 1.8 TB/s total theoretical
- Can we prove that?



UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

**Performance 20% of expected**



UNIVERSITY OF CAMBRIDGE

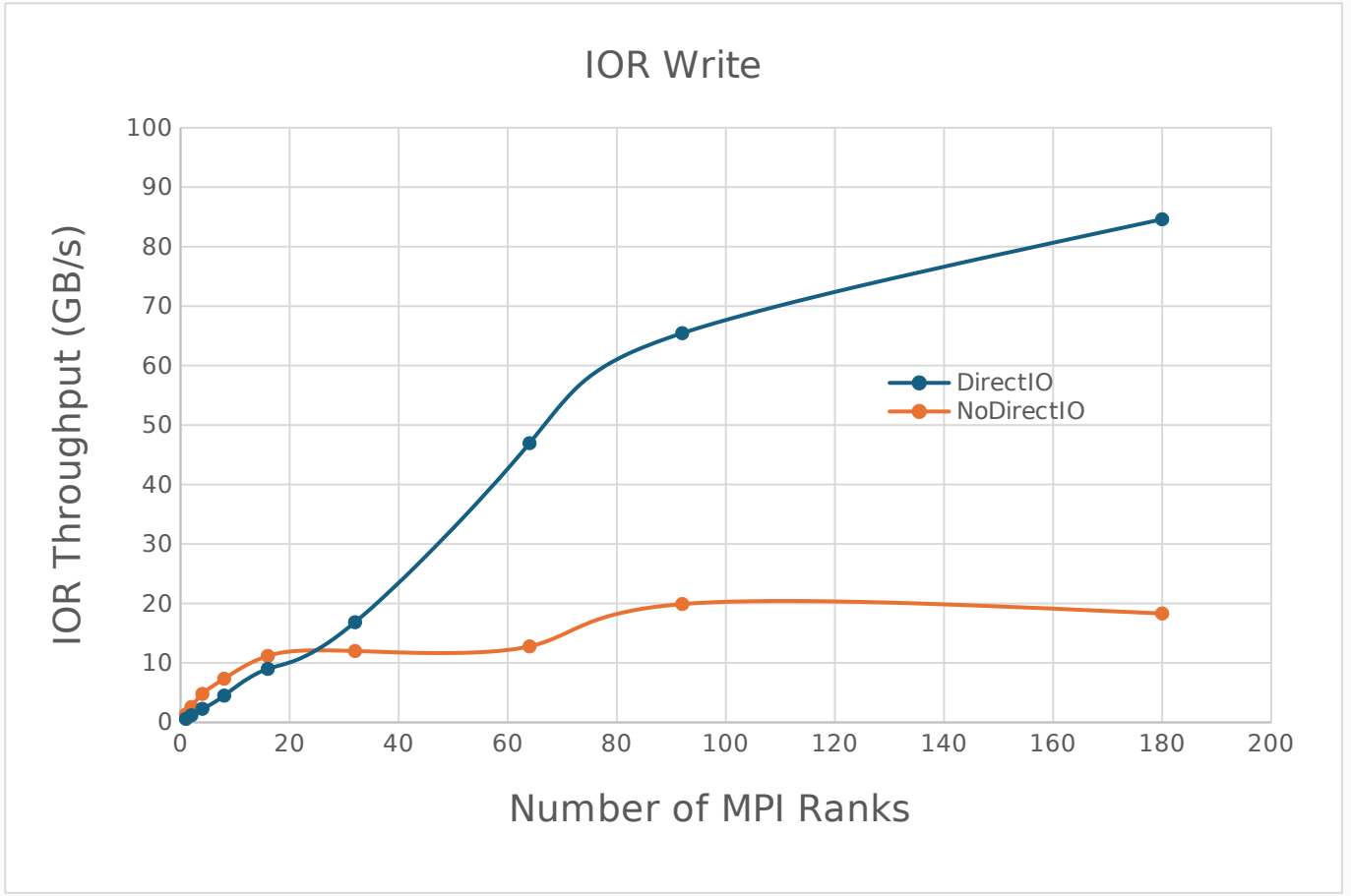
# War Stories 2: NVMe

- There are some clues:
- portal_rotor set to ON, so not limited at 25% per client due to single network connection
- Some issues with page cache a couple of years ago.
- What about O_DIRECT?

UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

IOR Performance with Directio close to linespeed



UNIVERSITY OF CAMBRIDGE

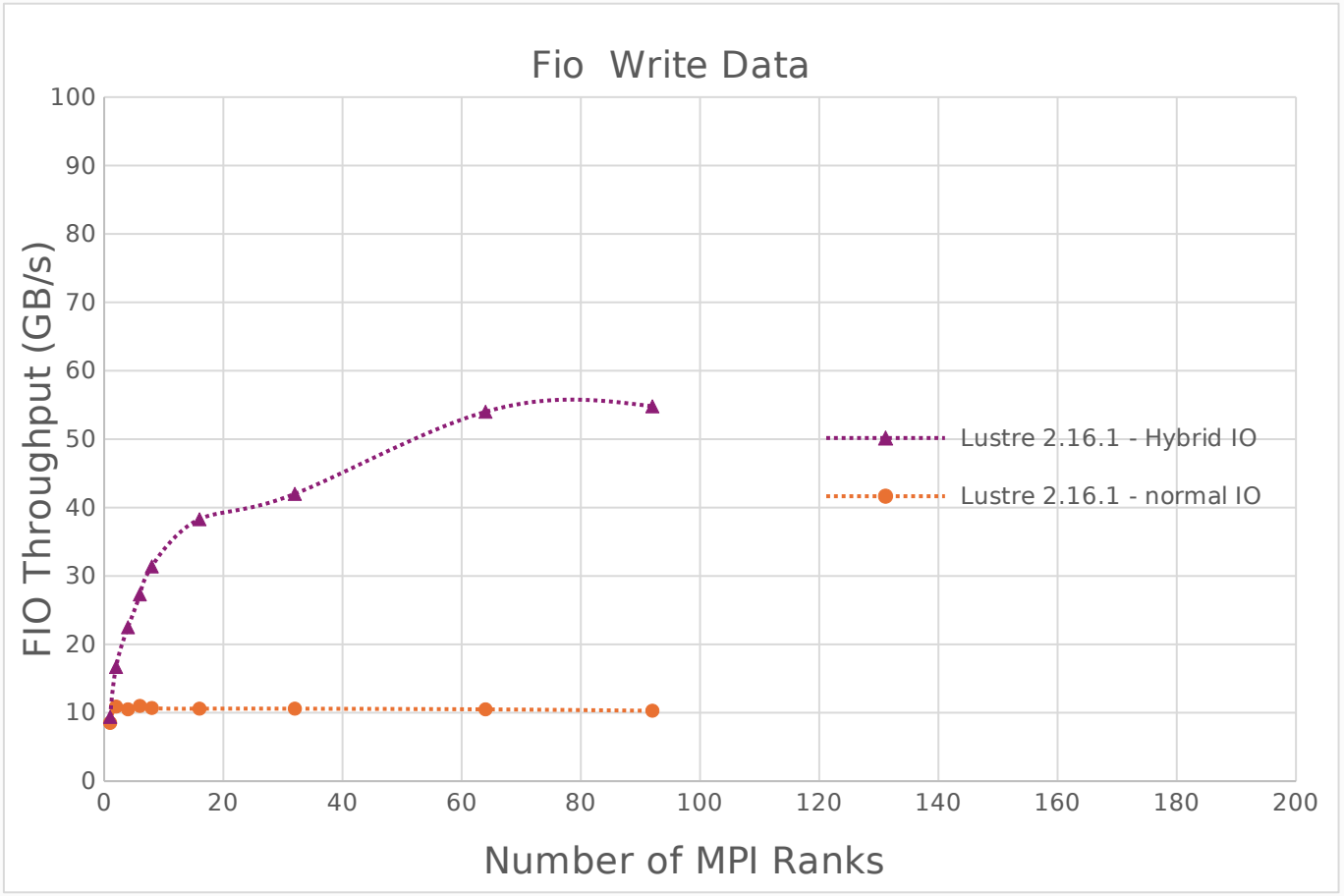
# War Stories 2: NVMe

- Page Cache cannot empty fast enough – varies 5-12 GB/s.
- DirectIO is the way to go, but needs code changes
- Can we do it using HybridIO instead? (new to 2.16)

UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

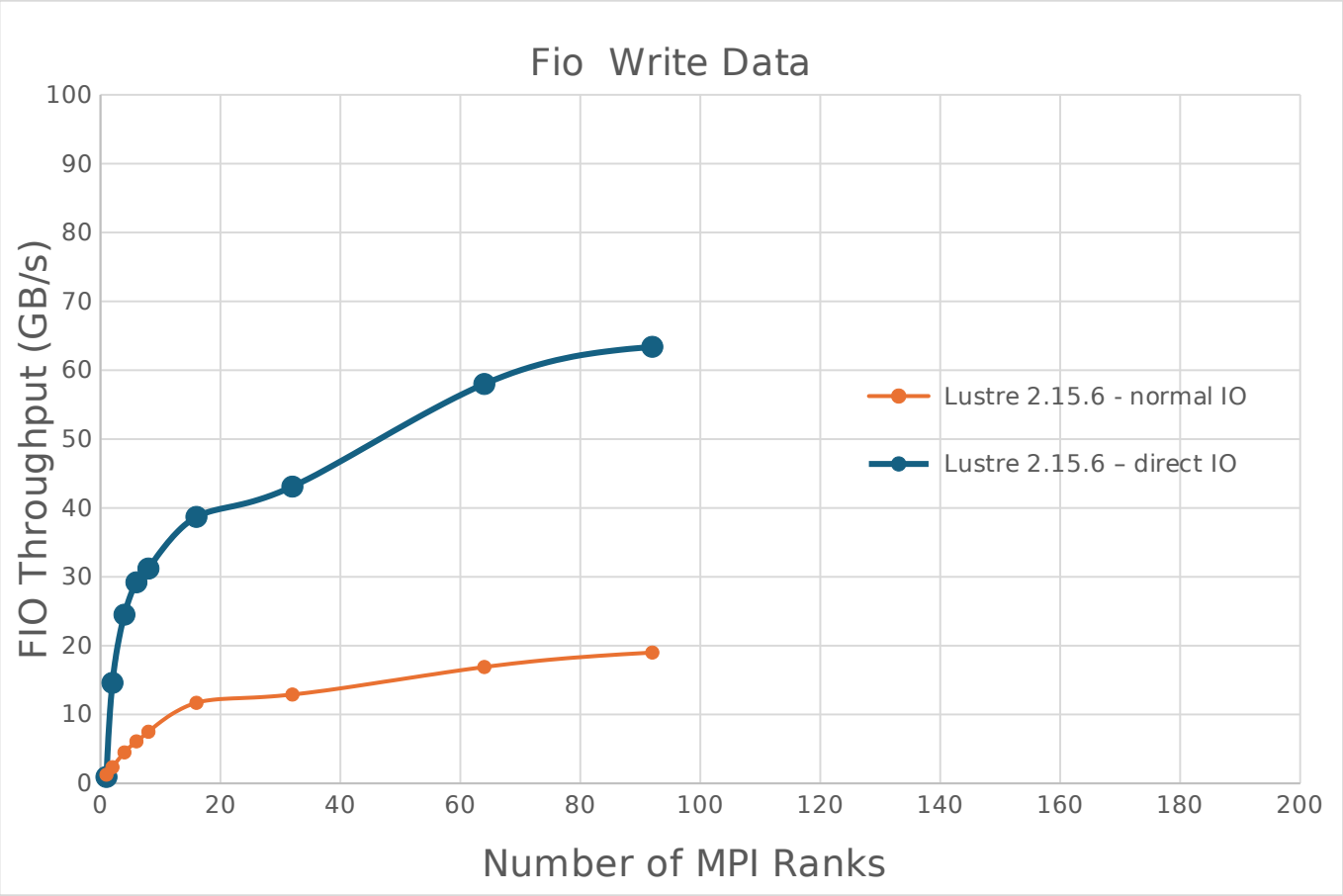
Hybrid IO on - ~55GB/s with FIO benchmark



UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

Compare fio in O_DIRECT



UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe

- Significant Single Node improvement with hybrid IO
- No code changes
- A little less performance than O_DIRECT
- HybridIO untuned – can still be improved

UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack


- Software defined infrastructure
  - “hybrid cloud”
  - satisfies cloud-first strategies
- Powerful central control of multi-tenant infrastructure
- Consistent APIs
  - Good ansible modules
  - Allows for storing physical deployment in code
- Core components:
  - Nova – compute resources
  - Neutron – network resources
  - Ironic – getting Nova to deploy directly onto hardware

UNIVERSITY OF CAMBRIDGE

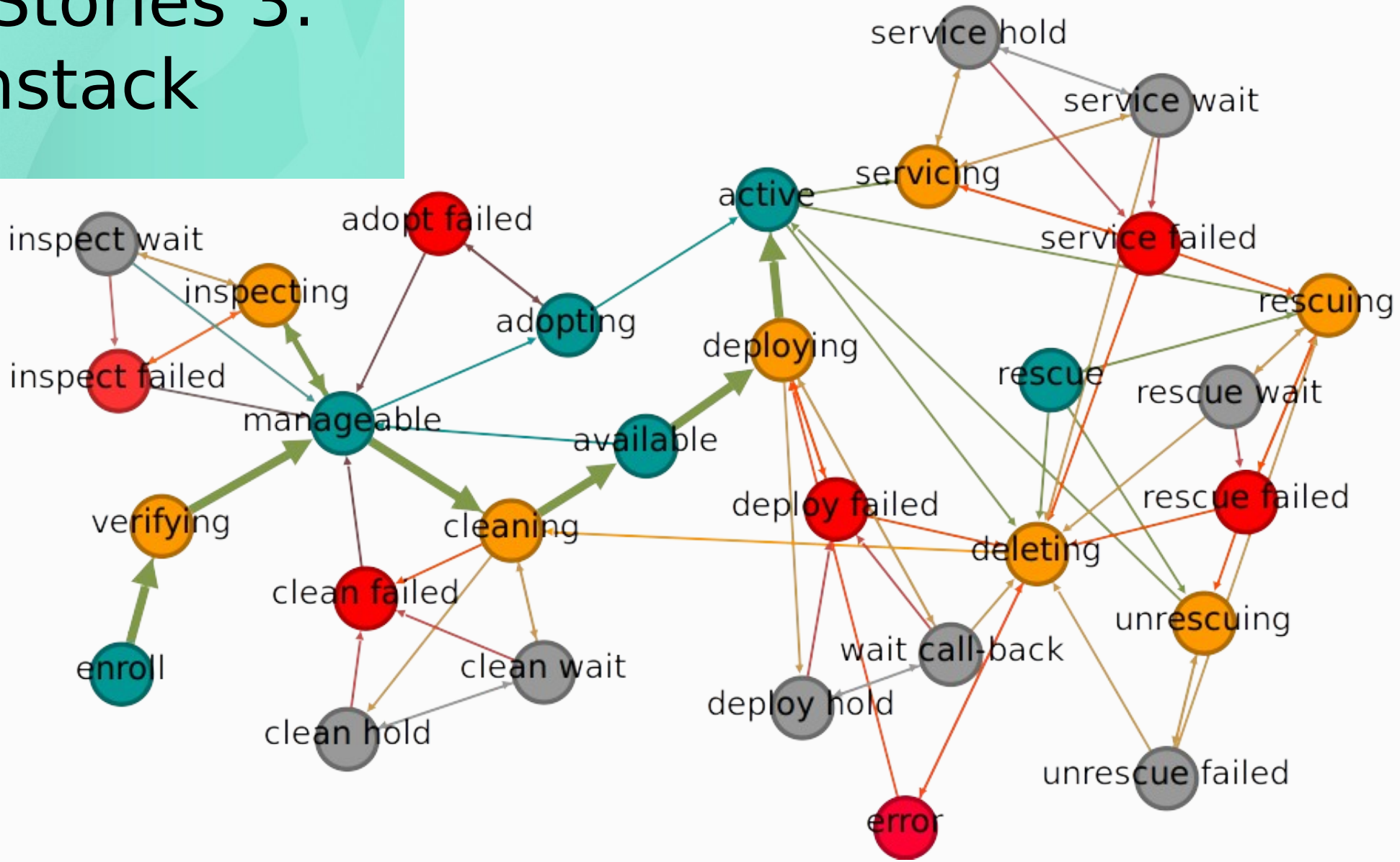
# War Stories 3: Openstack

- Lots of knowledge of Openstack in Cambridge
  - Links to StackHPC for upstream bugfix
- Very powerful for deployment of compute nodes
  - New images once a fortnight
- Can we do this for storage?

- How do we even create openstack servers?

UNIVERSITY OF CAMBRIDGE

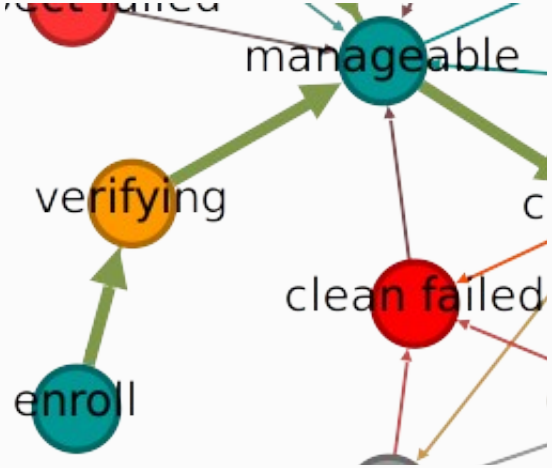
# War Stories 3: Openstack



UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack

- Enroll
  - Simply make openstack aware node exists
  - Find and verify network config



UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack


- Inspect
  - Switch VLAN to inspect VLAN
  - Boot image, run code to look at machine

UNIVERSITY OF CAMBRIDGE

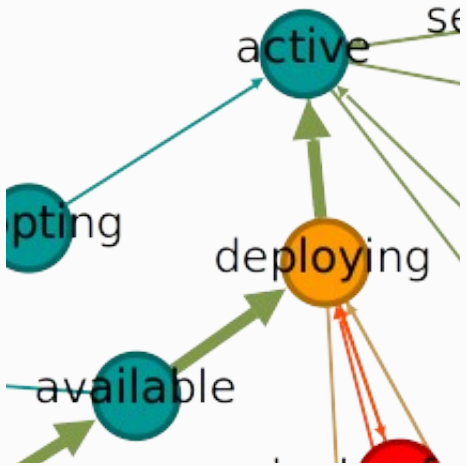
# War Stories 3: Openstack

- Clean
  - Switch VLAN to clean VLAN
  - Delete old OS and make blank
  - Delete all storage

UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack

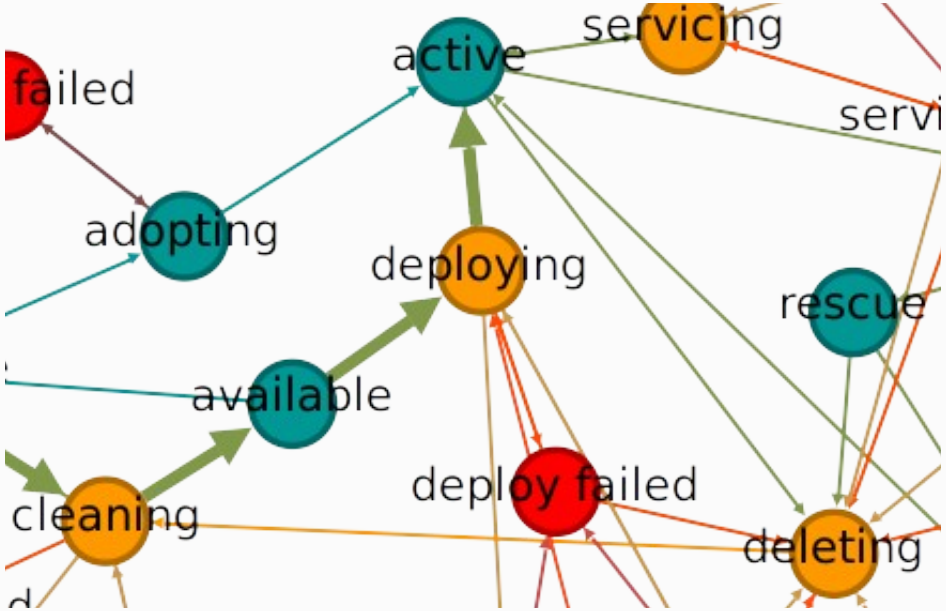
- Deploy
  - Switch VLAN to deploy VLAN
  - Image new OS and boot
  - Switch VLAN to production VLAN



UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack

- Deleted
  - Machine goes back to cleaning step



UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack

- Risks
  - Unmanned cleaning step can wipe attached data
- Mitigations
  - Prevent unacknowledged cleaning step

UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack

- Solutions:
  - Blacklist SAS kernel module in cleaning image
    - Will prevent external storage from being deleted
  - Prevent nodes from going to clean VLAN
    - Stops unacknowledged clean steps through mistakes
  - Lock nodes to prevent deletion

UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack

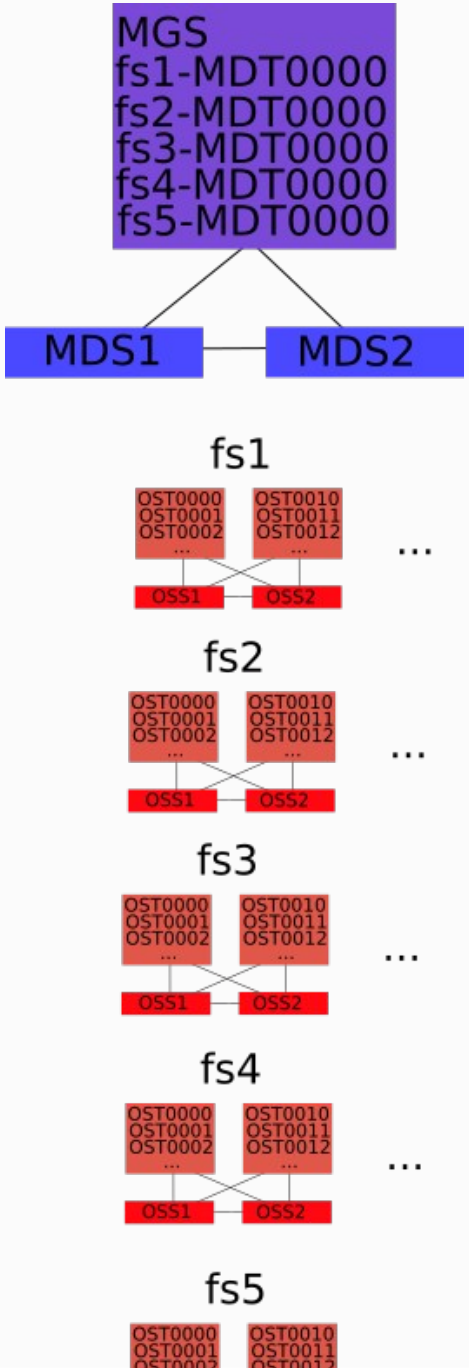
- Conclusions:
  - Works well!
  - Provides test and prod environment that match
  - Allows for more fine-control of power states than IPMI

UNIVERSITY OF CAMBRIDGE

# Thank you!

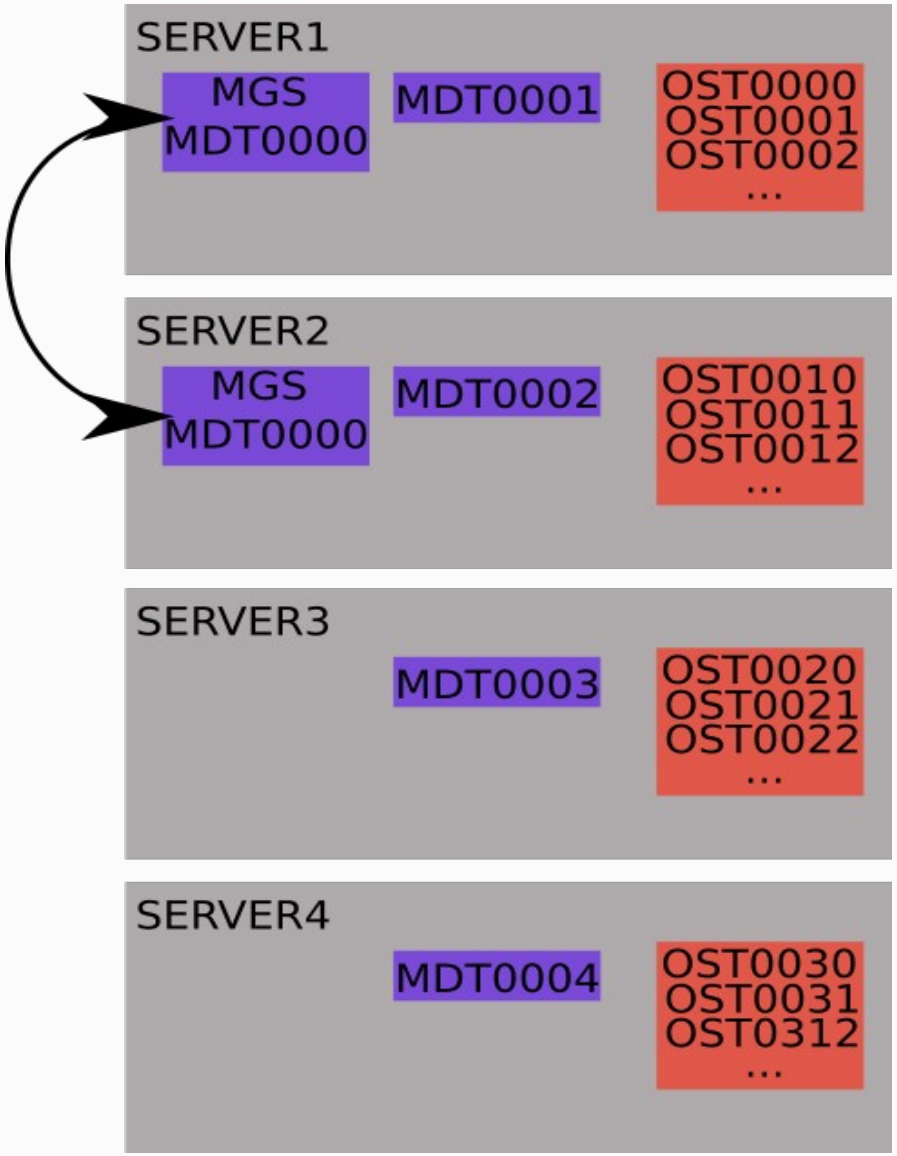
UNIVERSITY OF CAMBRIDGE

# War Stories 1: Changelogs



UNIVERSITY OF CAMBRIDGE

# War Stories 2: NVMe



UNIVERSITY OF CAMBRIDGE

# War Stories 3: Openstack