





THE
FRANCIS
CRICK
INSTITUTE



Imperial College
London



CANCER
RESEARCH
UK

Crick Site Update



Steve Hindmarsh
Head of Scientific Computing, The Francis Crick Institute

Spectrum Scale User Group
10th December 2021

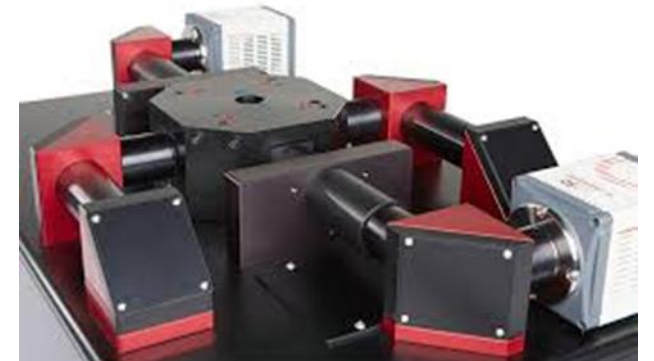
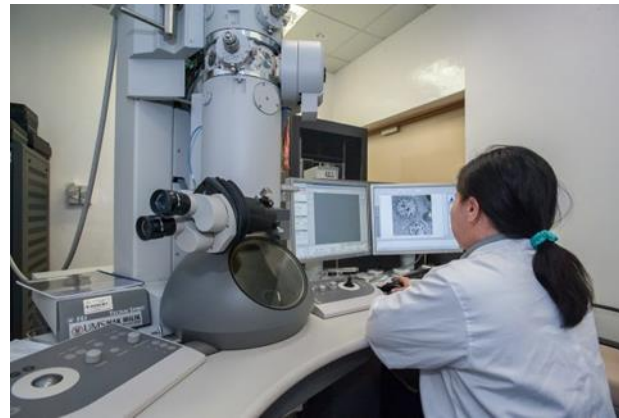
Current research data storage environment

- 11 PiB DDN GRIDScaler (2 x SFA12KXE + 2 x GS7K)
- OEM Spectrum Scale v4 licences tied to hardware
- 10 x MEDIAScaler NAS/protocol nodes
- FDR InfiniBand
- TSM/Spectrum Protect backup & archive to off-site IBM TS4500 with type 3592 enterprise ('Jaguar') drives
- ~9 PiB used, ~2 Billion files



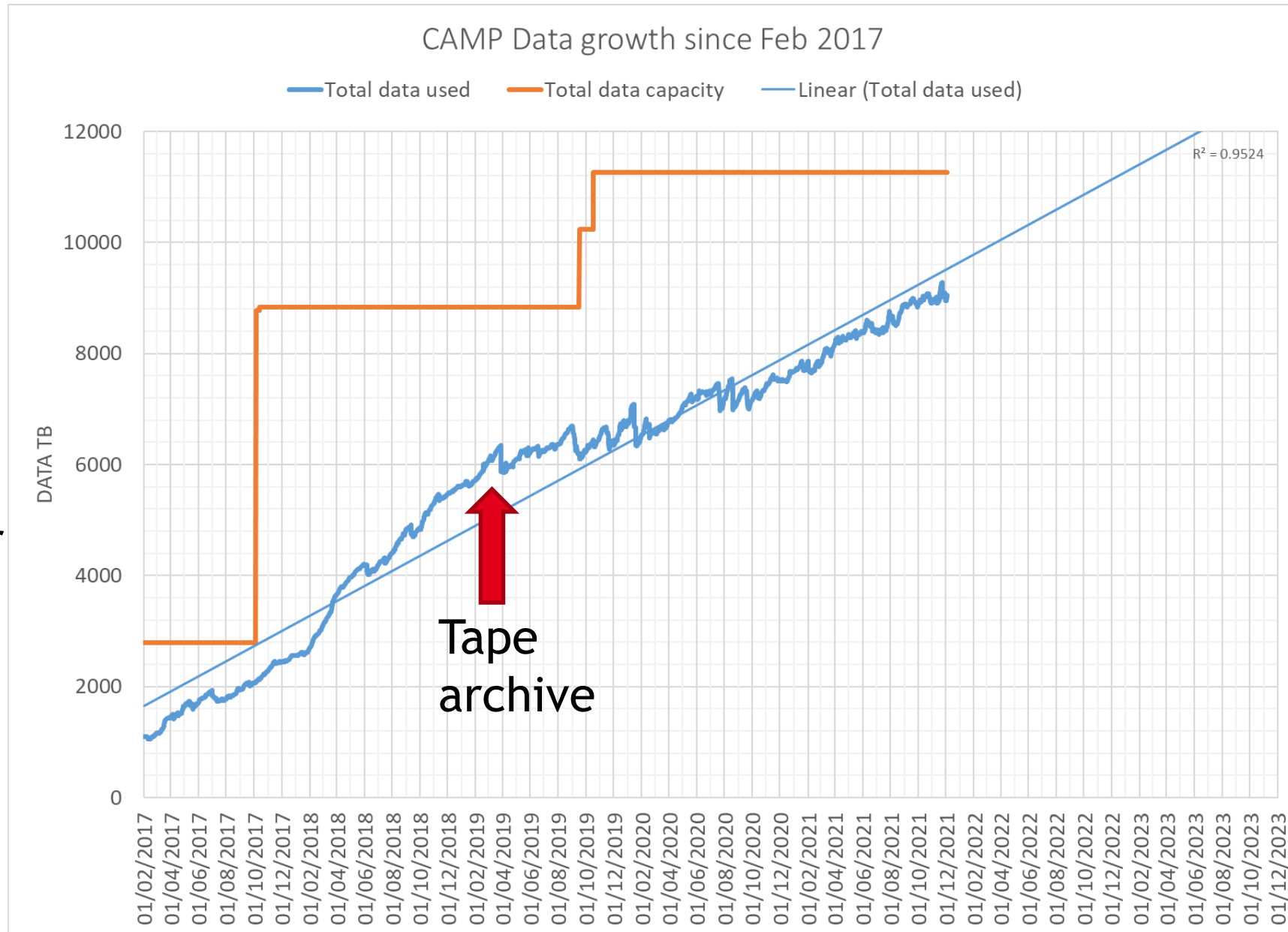
Storage management

- All 100+ labs and 15 core facilities have quotas (between 5 TB and 1.5 PB!)
- Peer review process for quota increase
- Data sources: DNA/RNA sequencing, cryo-EM, Electron Microscopy, Light Microscopy, 'omics etc.



Breathe in - data

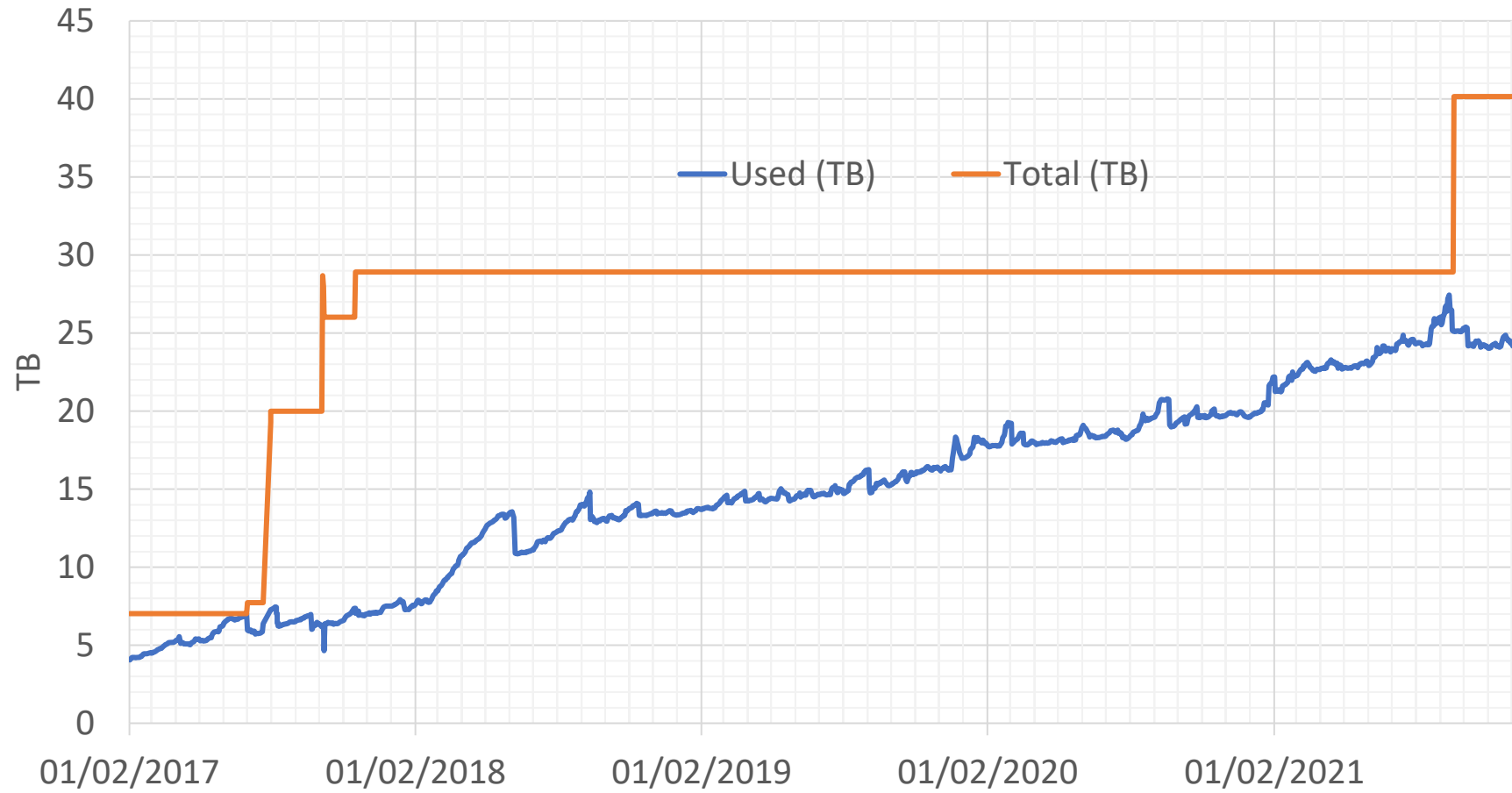
- Manage rapidly growing data with fixed data capacity since Oct 2019
- No budget for extra storage until 2022
- Tape archive from March 2019



Breathe in - metadata

- SSDs added with 2017 disk expansion
- SSD Fast pool re-purposed for metadata this year

CAMP Metadata usage



Spectrum Scale v4 to v5 upgrades...

- Purchased DDN 18-month maintenance extension to end of 2022
- v4 EOS...
- IBM/DDN breakup...
- Marriage guidance!
- Purchased new v5 licences from IBM to run on DDN controllers
- CentOS > RHEL
- GRIDScaler v4 > v5 by DDN

New storage system needed...

Procurement during 2021 for replacement system:

- At least 15 PiB usable
- Easily scalable to >30 PiB over 5 years
- HDD + NVME for capacity and performance
- Tape archive still supported
- Need predictable linear expansion cost for grant-funded projects
- Cloud comparison...

OCF/Lenovo solution

Finalised November 2021:

- DSS-G 15.5 PiB HDD + 1.2 PiB NVMe
- OEM Spectrum Scale
- HDR InfiniBand (connected to existing FDR via new EDR core)
- New Ice Lake DSS-G appliances expandable to 10 disk shelves (3 appliances x 5 PiB each expandable to 10 PiB)
- New large grant funded projects can buy extra disk shelves
- Shipping Q1 2022, in service by April 2022



OCF

Lenovo



IBM
Spectrum
Scale

Migration etc.

- Lenovo storage will be ready for new research data from April
- OCF managed ~9 PiB data migration from DDN > Lenovo using Atempo Miria software, estimate April-October 2022
- DDN hardware to be retired December 2022

I'm on the move...

NBI Partnership

Healthy Plants, Healthy People, Healthy Planet



Acknowledgements



- Crick Research Computing Platforms/HPC Team
- DDN
- OCF
- Lenovo
- IBM



crick.ac.uk