ICHEC
Irish Centre for High-End Computing

FROM RESEARCH TO INDUSTRY
cea

Integrating Phobos - an open-source tape-capable object store – as Lustre HSM backend

LAD'20

Ciarán O' Rourke & Sebastien Gougeaud

15th October 2020

Who are we?

- Performance Engineering Group @ ICHEC.
- Irish Centre for High-End Computing @ National University of Ireland Galway.
- Hubs in Dublin and Galway Ireland.
- Collaborative project with CEA and DDN.

Who are we?

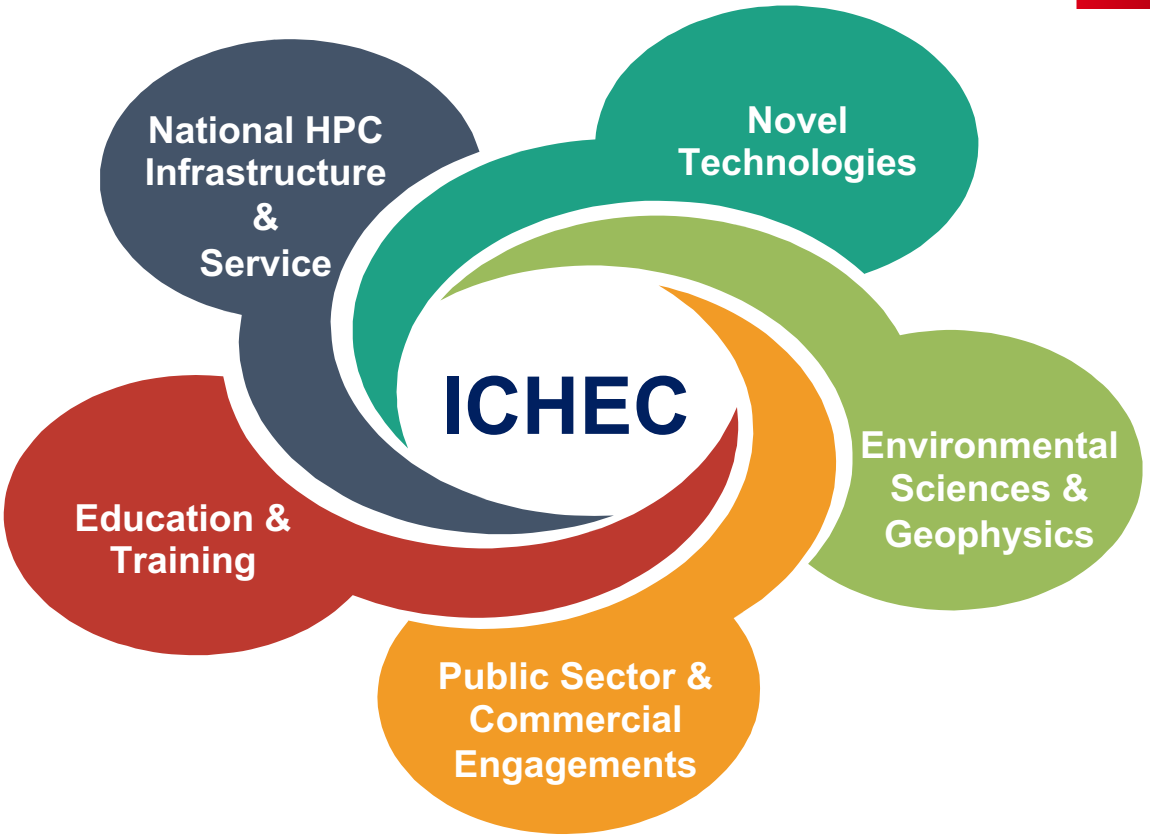

National HPC Infrastructure & Service
Novel Technologies
ICHEC
Environmental Sciences & Geophysics
Education & Training
Public Sector & Commercial Engagements

Table of Contents

Motivation

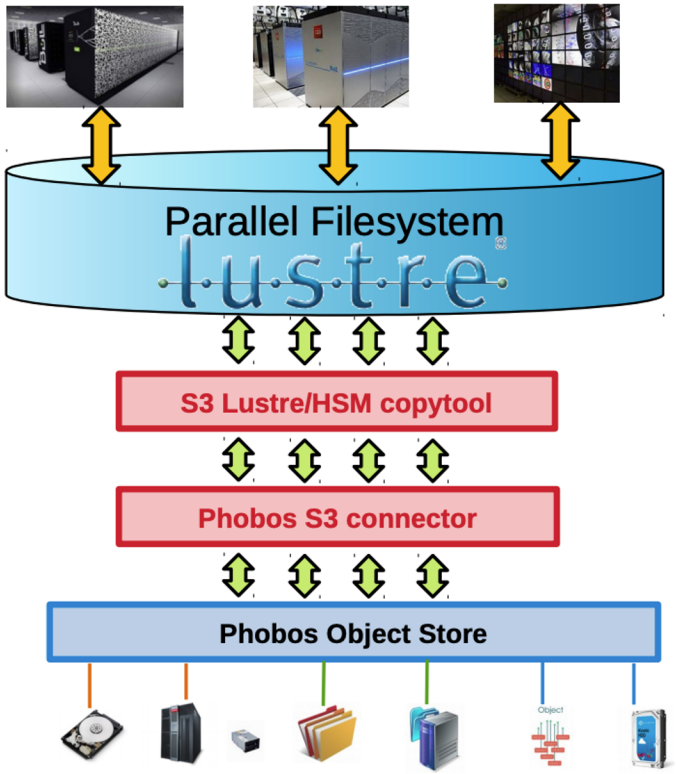
- Exascale computing will significantly increase workloads on storage systems.
- Huge amounts of data storage and ingestion will be required.
- This will require extremely scalable storage systems at reasonable prices.
- Tape libraries provide safe long term storage at low costs and zero energy usage.
- Object stores have proved their scalability.

Motivation

- Provide extension to Lustre parallel storage onto cost effective central object storage.
- Phobos provides tape object storage at scale (and more).
- Integrate Phobos with Lustre HSM, using a commonly used data access paradigm.
- S3 Interface is a generic/commonly used data access paradigm used in Cloud Computing and supported by many object stores.

Motivation

- Create/use open source tools to enable Lustre HSM with Phobos backend.
- S3 HSM CopyTool.
- S3 Web server for interfacing with managed object stores.



Overview of Phobos

Parallel Heterogeneous Object Store

- Developed by the CEA since late 2014
- ~44'000 code lines in C (core) & Python (CLI)
- LGPL 2.1 licence

Goal: handle a heterogeneous distributed set of storage resources

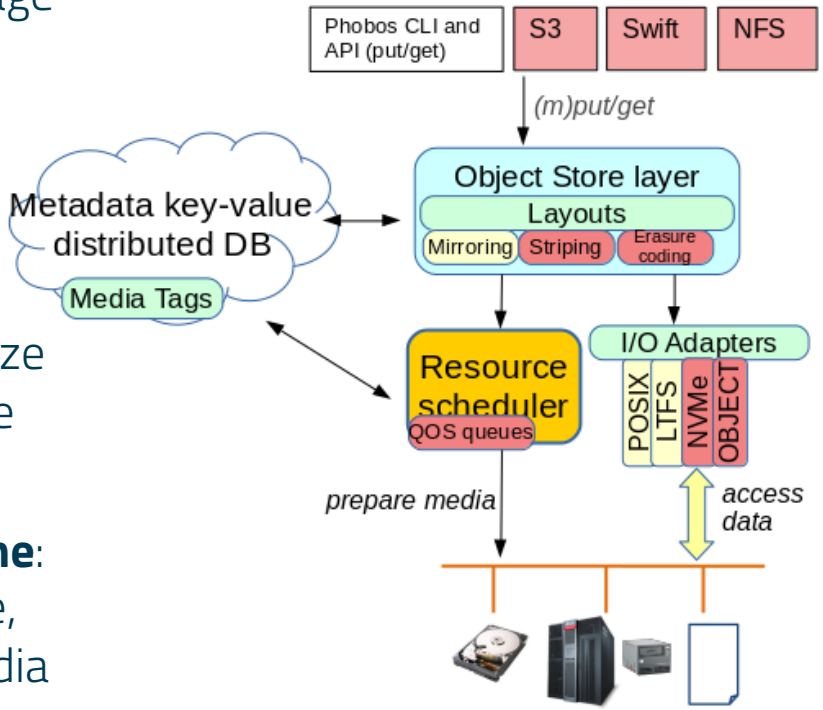
- Tapes, hard disks, file systems, etc.
- Optimized I/O for each technology

Used in production for France Genomique since 2016

- Multi-petabyte genomics datasets

How Phobos works?

- **I/O adapters**: multiple storage technologies
- **Layouts**: performance and fault-tolerance
- **Tags**: storage partitionning
- **Resource scheduler**: optimize tape fill rate, number of tape mounts
- **Key-value metadata scheme**: distributed NoSQL database, saved within objects on media (recovery, tape import)



Phobos roadmap

Features

- Deletion, versioning
- Media lifecycle, migration

Performance

- Multi-server parallelism
- Allocation optimisation

Administration

- GUI, monitoring
- Production requests

Deimos

Delivery Endpoint Interface for Managing Object Storage

- Webserver with S3 Interface for interchangeable storage backends.
- Currently supports PUT, GET, HEAD, and object listing S3 features.
- Supports S3 Authentication mechanism.
- Open source project that can be found at, https://git.ichec.ie/performance/storage/deimos.
- Maintained currently by ICHEC.

Deimos

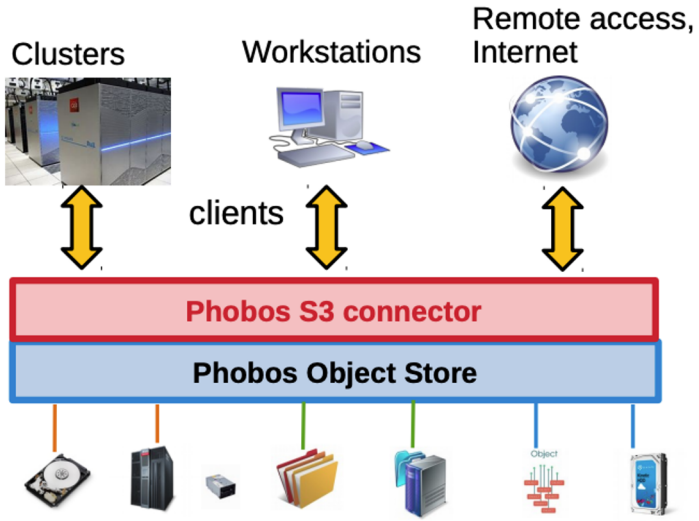
Implementation

- Modularised approach, with storage, stream and server protocol.
- Webserver implementation based on Proxygen, Facebook's open source library.
- Highly parallelisable.
- FIFO-no-copy approach used to handle data streams.

Deimos

Deimos and Phobos

- Deimos can be used as a Phobos S3 connector.
- Enabling S3 objects to be sent to and retrieved from a Phobos managed object store on the available storage media.



Deimos

Future Work

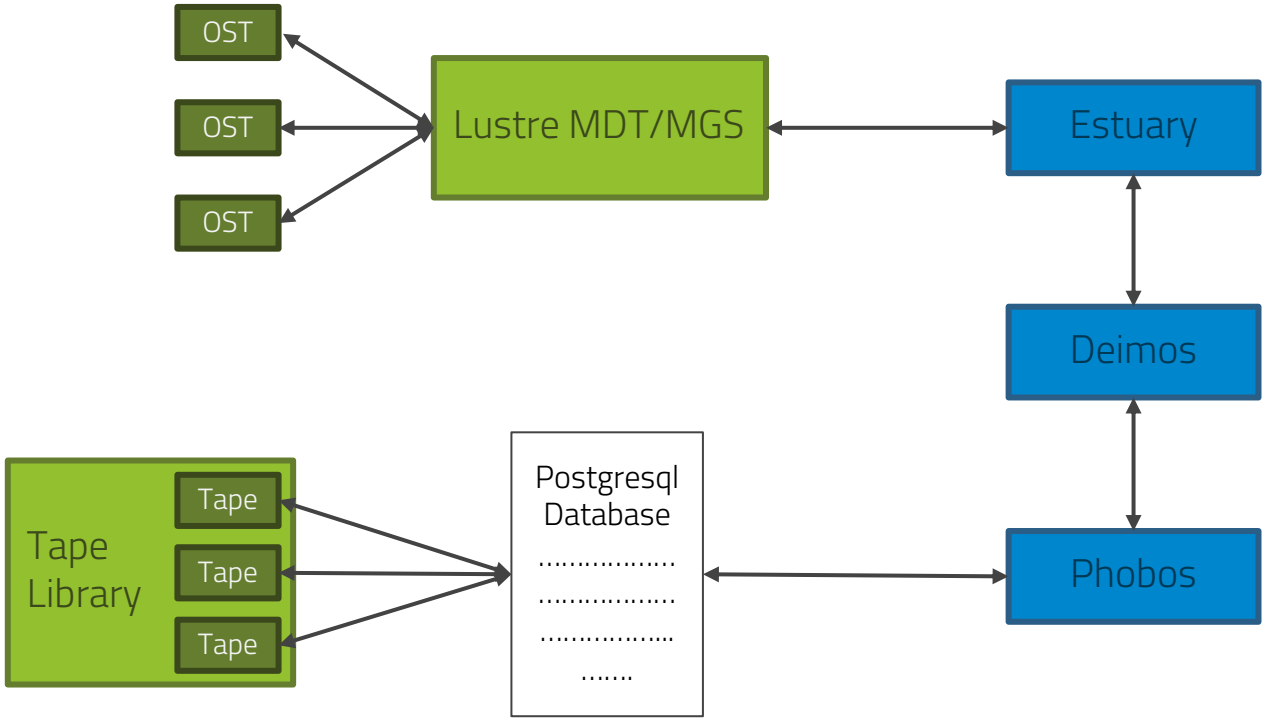
- Support additional S3 functionality.
  - List Buckets
  - Delete Objects
  - Versioning
  - Multipart upload
  - User management and permissions
- Distributed Deimos system.
- Extend to include OpenStack Swift interface.

Estuary

S3 HSM copyTool

- HSM copytool enabling HSM with Lustre and Object Storage via a S3 interface.
- Forked from ComputeCanada lustre-obj-copytool.
- Main change was to update to new lustre version and update to new version of libs3.
- Allows pathway from Lustre to Deimos via S3.
- Open source project found at, https://git.ichec.ie/performance/storage/estuary.
- Maintained by ICHEC.

Full Pipeline



Example Setup

Two machines:

Martin

Running Lustre and the copytool Estuary

Dieter

Running the web server Deimos and Phobos

Example

Martin

Archive a file from Lustre using Estuary

Release the file

Restore the file

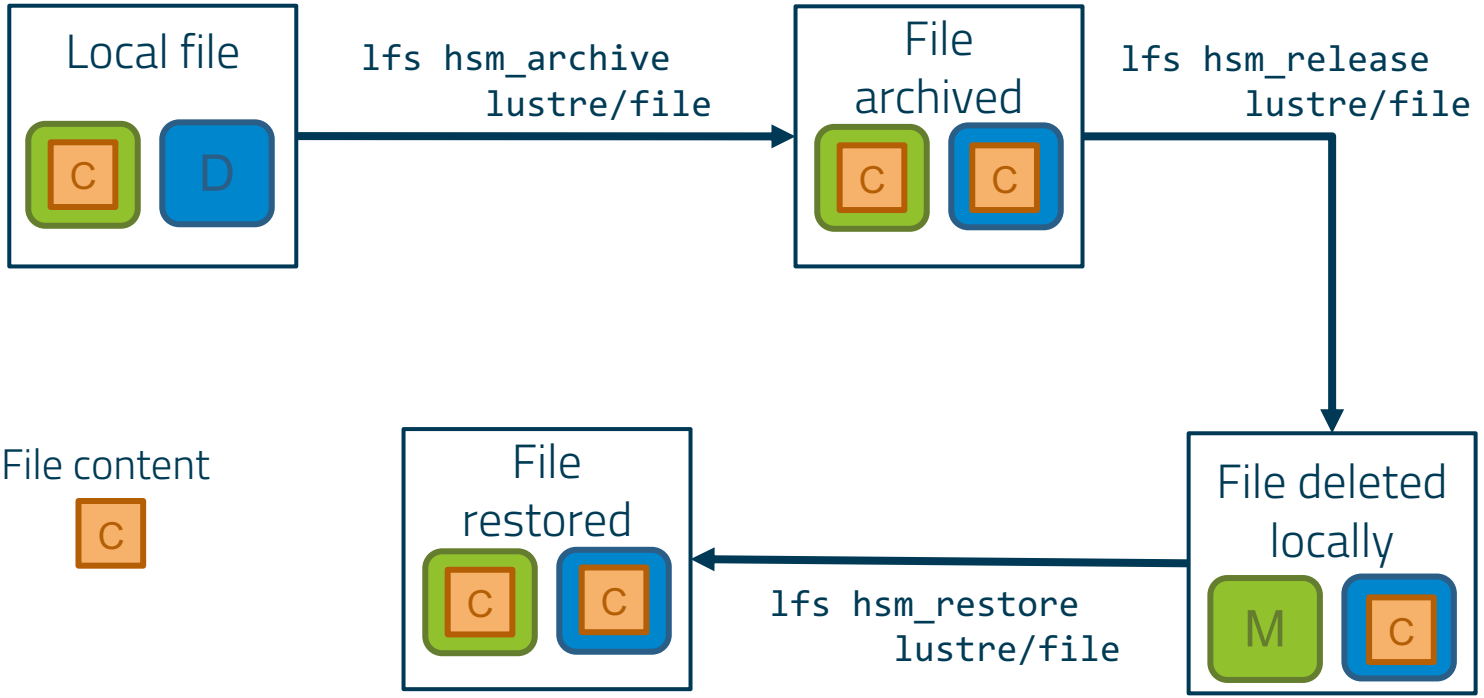
Dieter

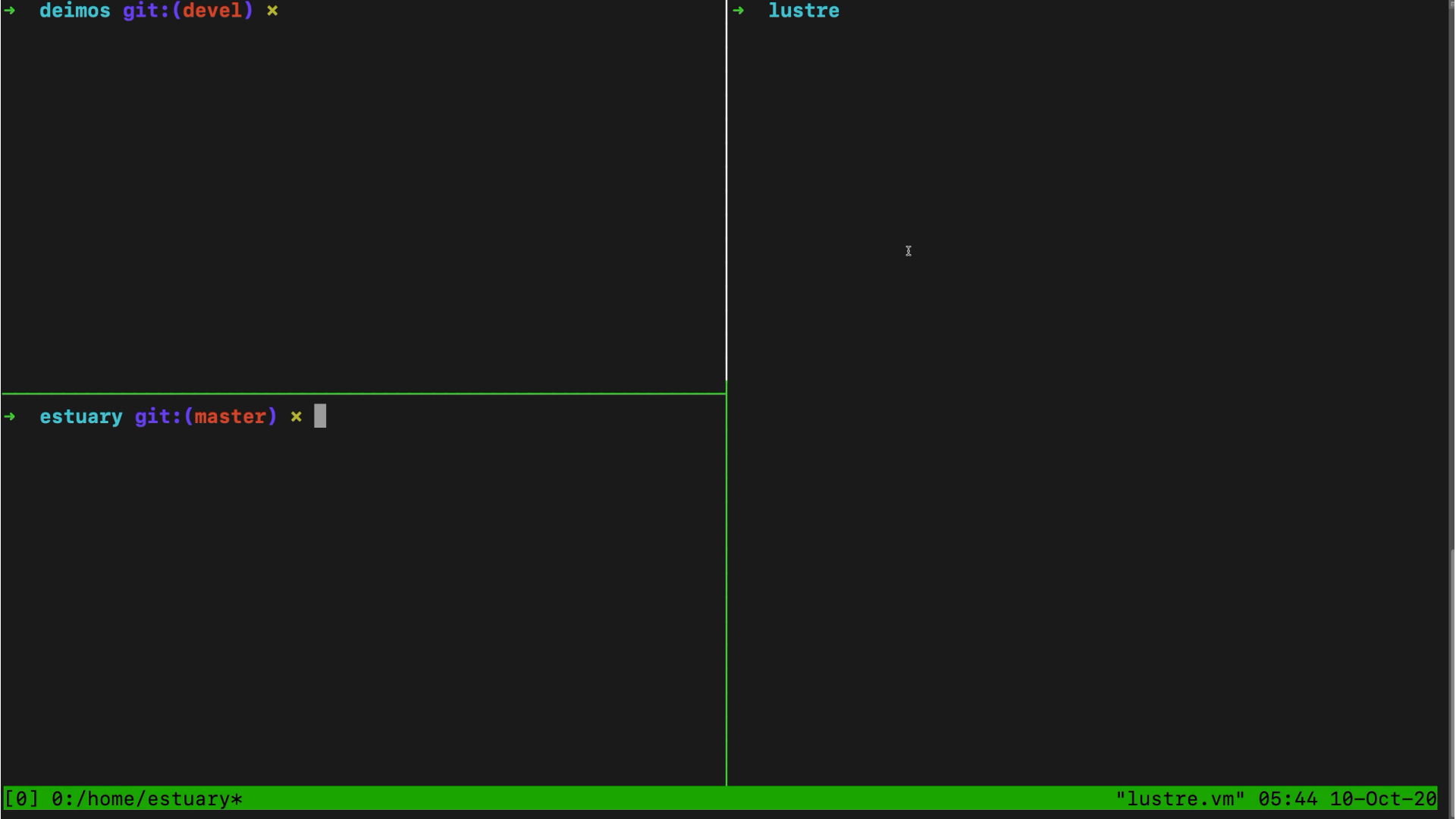
Accept the file contents in Deimos

Save it to Phobos

Send it back

Example


Local file
C
D
lfs hsm_archive
lustre/file
File
archived
C
C
lfs hsm_release
lustre/file
File content
C
File
restored
C
C
lfs hsm_restore
lustre/file
File deleted
locally
M
C


→ deimos git:(devel) ×
→ lustre
→ estuary git:(master) ×
[0] 0:/home/estuary*
"lustre.vm" 05:44 10-Oct-20

ICHEC
Irish Centre for High-End Computing

Thank You