dCache CTA integration
Status, experience, plans.

Tigran Mkrtchyan
for the dCache collaboration
&
DESY tape team.

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

What is DESY (as storage)

Experiments/Community	Service	Role
EuXFEL, Petra-III, ILC, Accelerator R&D, ...	Source of the data. Primary data site. Provides online, nearline and archival storage.	Tier-0
Belle-II, ...	Provides online and near-line storage.	Tier-1
Atlas, CMS, LHCb	Online only.	Tier-2
H1, Hermes, Hera-B, Zeus , ...	Provides online and archival storage.	Data Preservation

Multiple Faces of Tape

At data source

- High data ingest rate
- Multiple parallel streams
- High durability, multiple copies on different media
- Long-term nearline access
- Small file handling

At analysis facility

- Automatic data accessibility migration
- Bulk recall on periodic basis
- Long-term nearline access
- Recall prioritization

Data Archive

- Manual data accessibility migration
- Long-term preservation
- Automatic technology migration
- Self-healing

Technology in Place at DESY

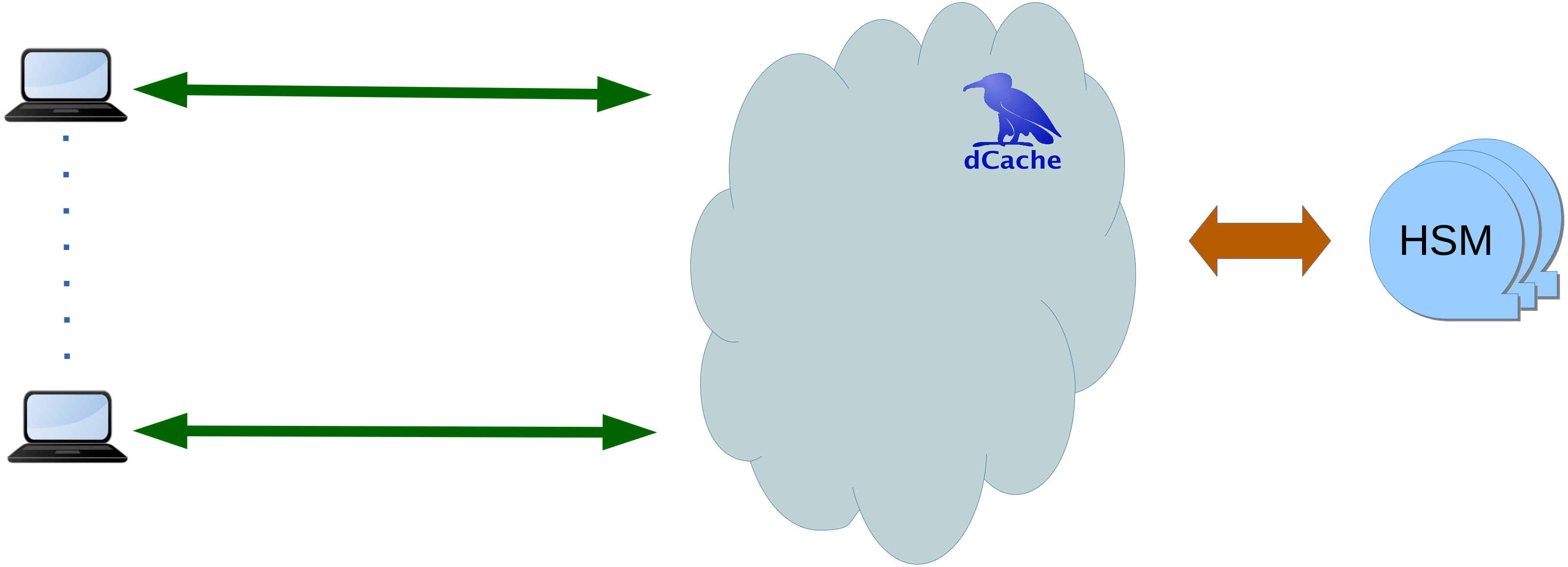
Hardware

- 2x Oracle SL8500 - EOL
  - 26x LTO-8 drives
- 2x IBM TS4500
  - 20x Jaguar
  - 16x LTO-9
  - Different buildings (500m)

Software

- TSM (IBM Spectrum Protect) – classic backup
- dCache – interface to HSM system
  - Scientific Data
  - AFS/Mail backup
- OSM (Open Storage Manager)
  - Since 1994, multiple local modifications (~80%)

dCache+HSM Tandem (DESY)



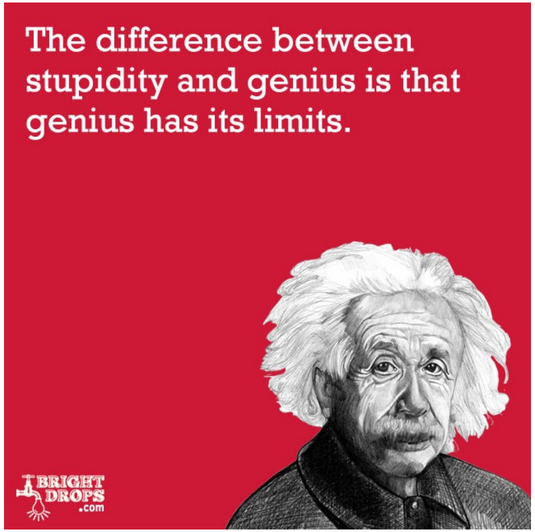
All access to scientific data on tape goes through dCache!

dCache Tape Connectivity

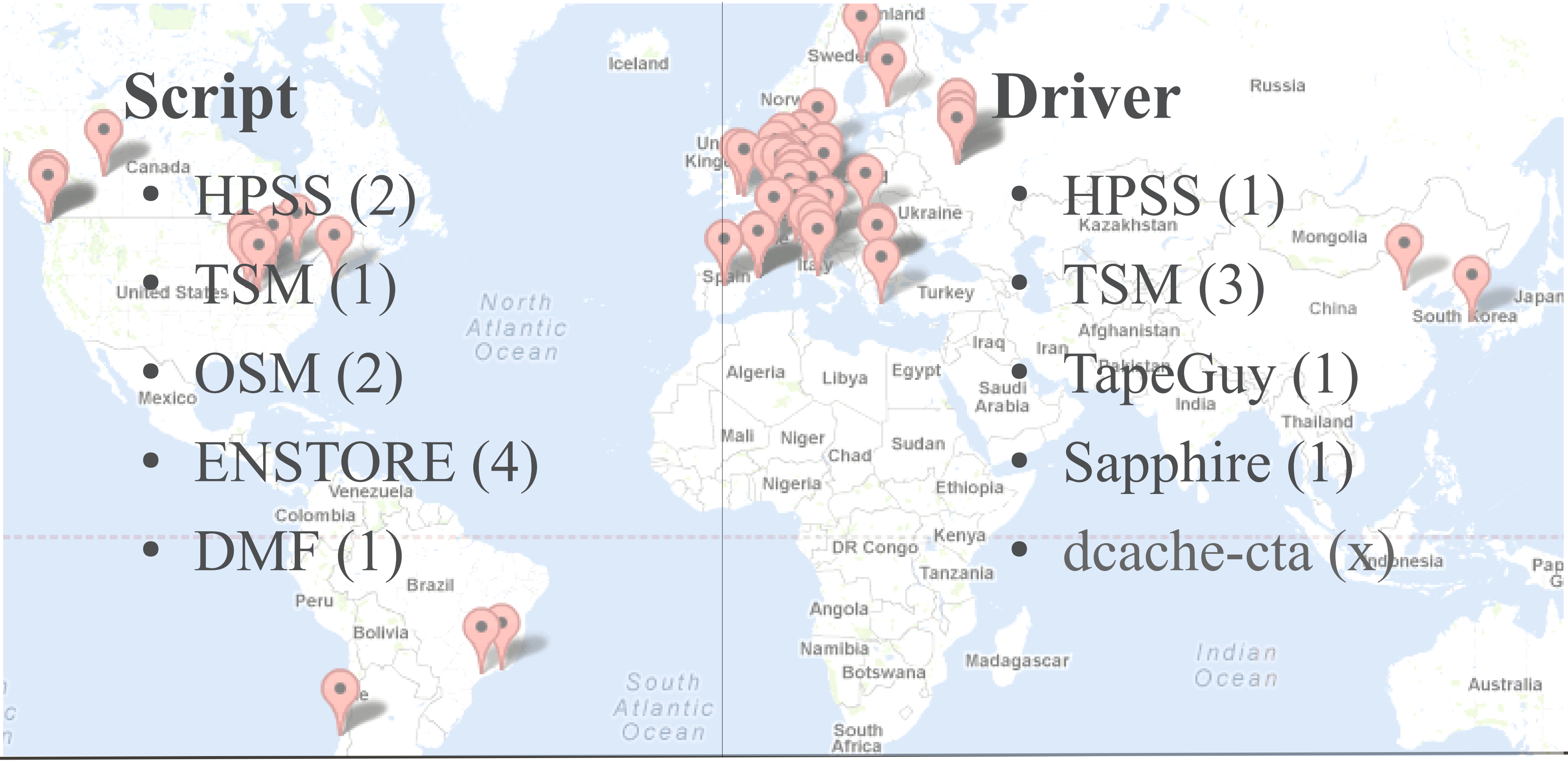
- Write-back / Read-through cache behavior
- Transparent for the users
- Available via all protocols (subject to authorization)
- Multiple HSM on a single instance
- dCache stores tape location as opaque data from HSM

Interfaces to HSM

- External Script
  - Stupid, Simple, Genius …
  - Reference implementation of driver API
- Plugable Driver API:
  - Suitable to create efficient HSM connectivity
  - ENDIT *(Efficient Northern dCache Interface to TSM)*


The difference between stupidity and genius is that genius has its limits.
BRIGHT DROPS .com

dCache+HSM Deployments



Script

- HPSS (2)
- TSM (1)
- OSM (2)
- ENSTORE (4)
- DMF (1)

Driver

- HPSS (1)
- TSM (3)
- TapeGuy (1)
- Sapphire (1)
- dcache-cta (x)

Tape Software Requirements

- Maximize tape HW efficiency
  - Integration into DESY ecosystem
  - Integration with dCache tape interface
- Stable operation for a next decade
- Should be Open-source, adopting open standards
- Wide user and technology community

CTA Test Deployment

- 3x test instances
  - 2x with virtual library (EOS, dCache)
  - 1x TS4500, Jaguar drive (dCache)
- CTA v4.3-3 + dcache patches (65665726)
- PostgreSQL-13
- dCache 7.2.2 (or later)

dCache HSM ⟺ Link

- Files – storage class

```
xfel:SQS-2019@osm
```

- Pools – HSM driver

```
hsm create osm siloA script \
                                   -command=hsmcp.py
```

- Namespace – location URI

```
osm://siloA/xxxxxxxxxxxx
```

dCache HSM Interface

```
// dCache interface to tape system

public interface NearlineStorage {

    void flush(Iterable<FlushRequest> requests);
    void stage(Iterable<StageRequest> requests);
    void remove(Iterable<RemoveRequest> requests);

    void cancel(UUID uuid);

    // driver initialization methods

    ...
}
```

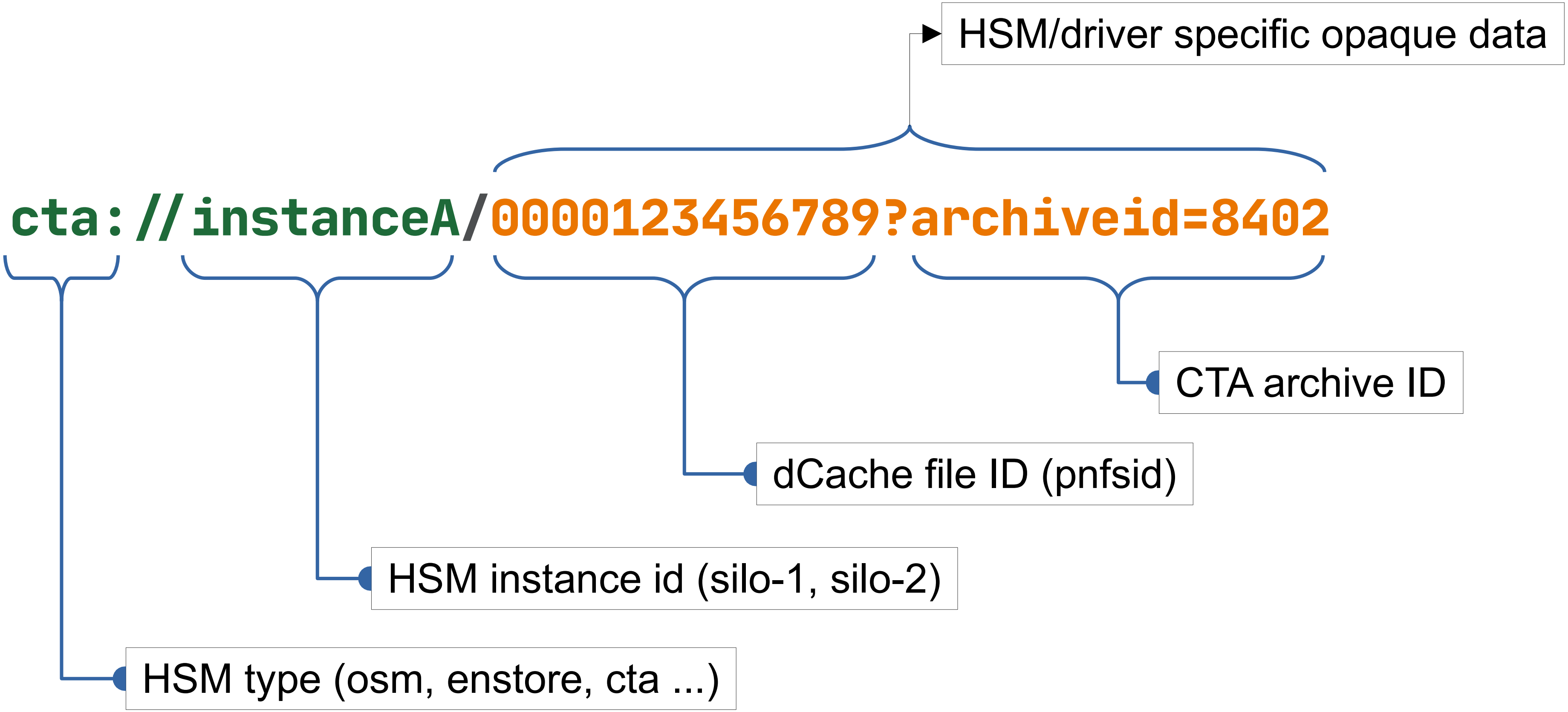
dCache CTA gRPC

```
// gRPC definition of dcache-cta interface

service CtaRpc {
  rpc Version (google.protobuf.Empty) returns (cta.admin.Version) {}

  rpc Archive (ArchiveRequest) returns (ArchiveResponse) {}
  rpc Retrieve (RetrieveRequest) returns (RetrieveResponse) {}
  rpc Delete (DeleteRequest) returns (google.protobuf.Empty) {}
}
```

dCache HSM ⟺ Link



HSM/driver specific opaque data

cta://instanceA/0000123456789?archiveid=8402

CTA archive ID

dCache file ID (pnfsid)

HSM instance id (silo-1, silo-2)

HSM type (osm, enstore, cta ...)

dCache ⟺ CTA Integration

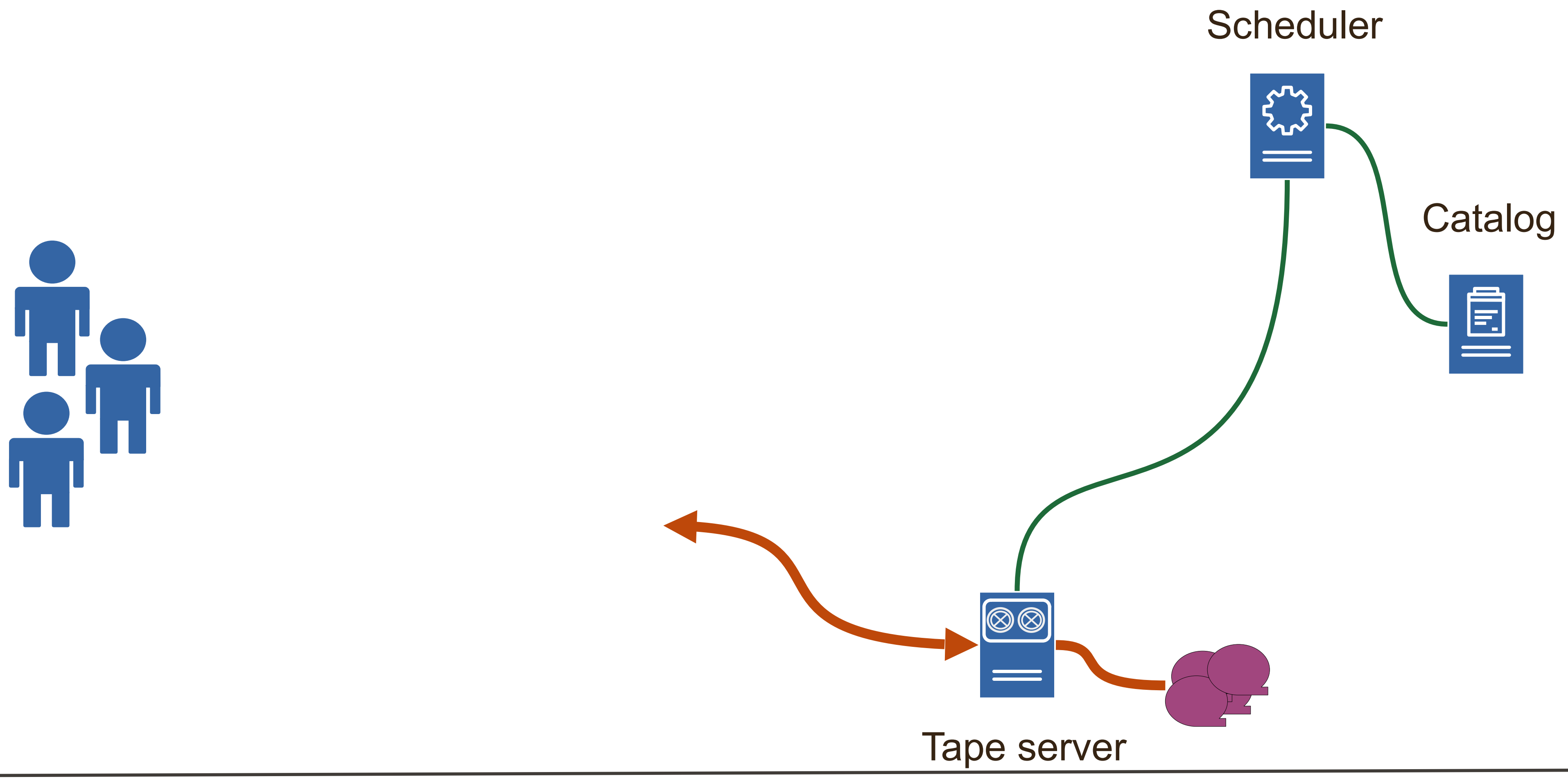
dCache

- Nearline driver to add (dCache ⩾ 7.2.2)
- Can run in parallel with other HSMs
- dCache pre-scheduling must be disabled/reduced
- File path, uid, gid not preserved

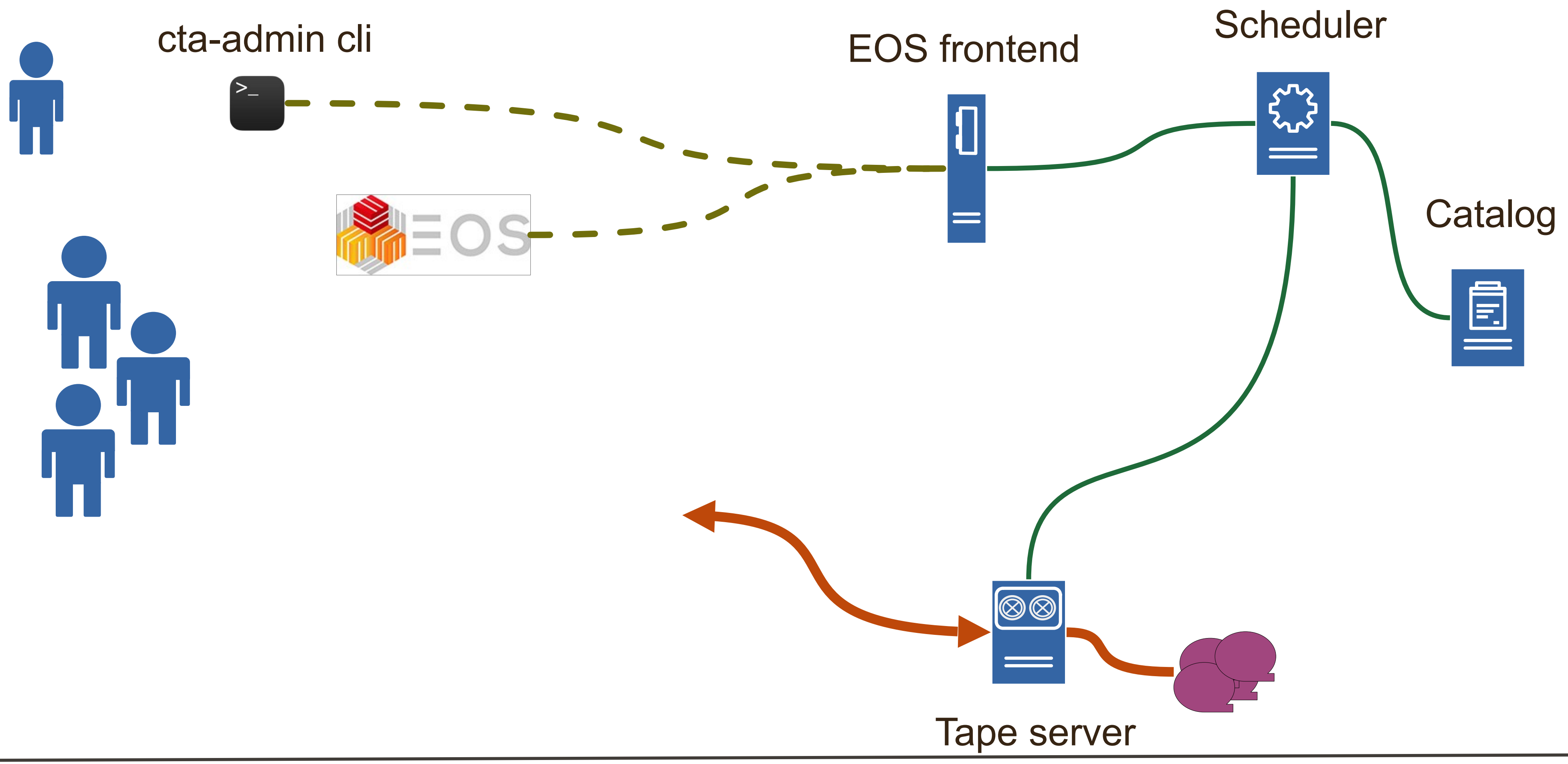
CTA

- Additional *cta-dcache* service and rpm
- Limited to dCache required minimal functionality
  - *cta-frontend still needed*

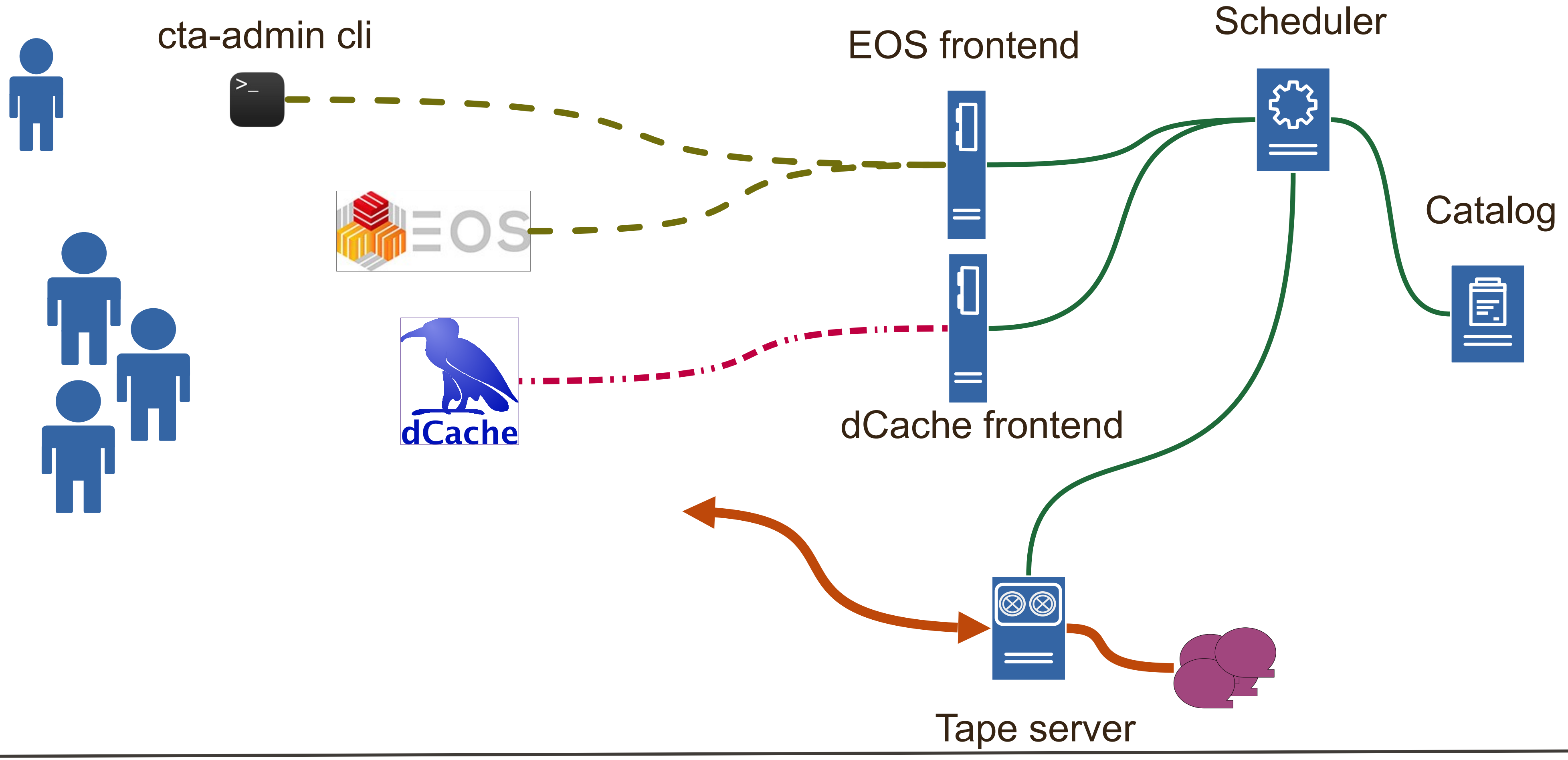
(Extremely) Simplified CTA design



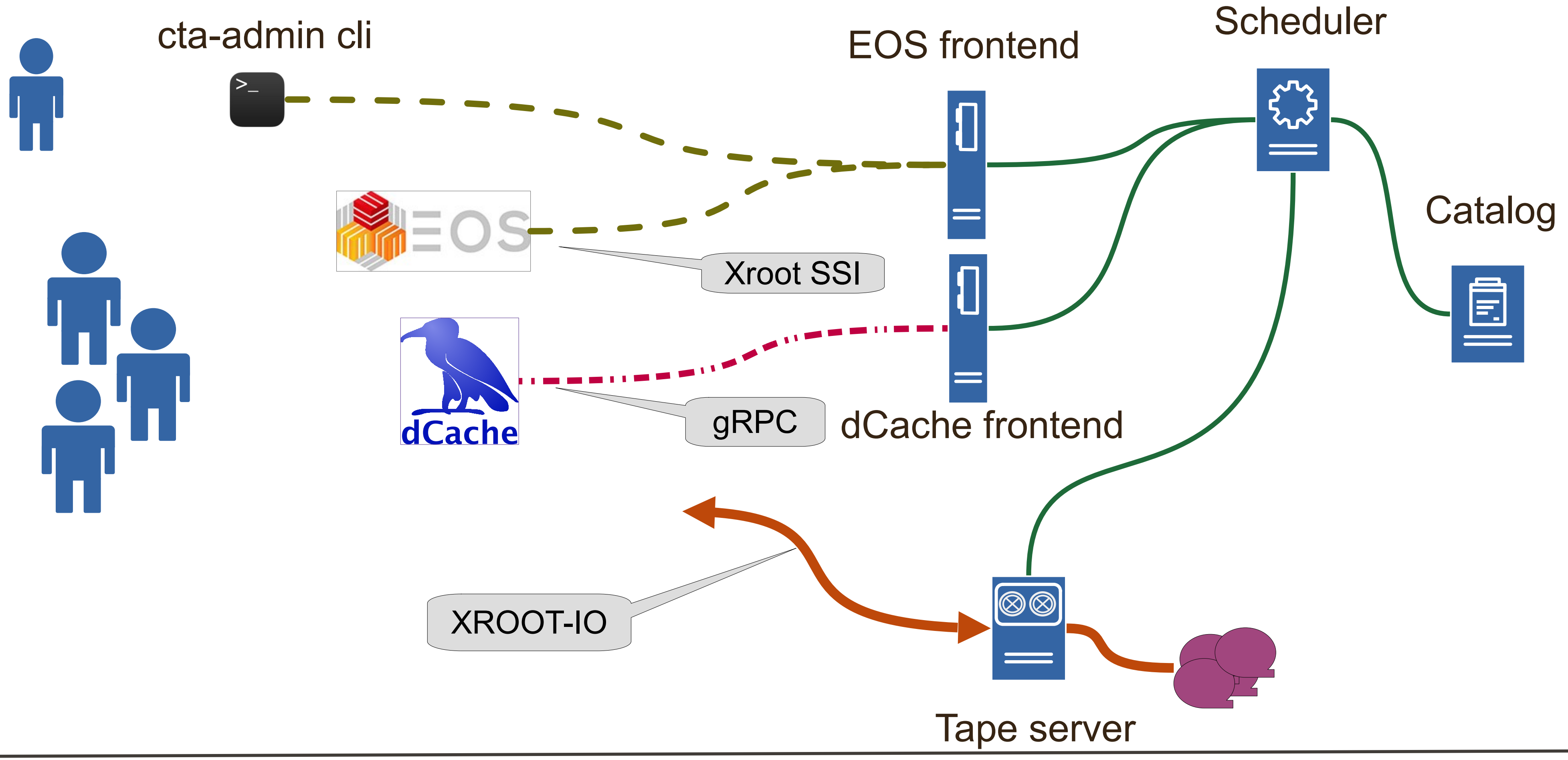
(Extremely) Simplified CTA design



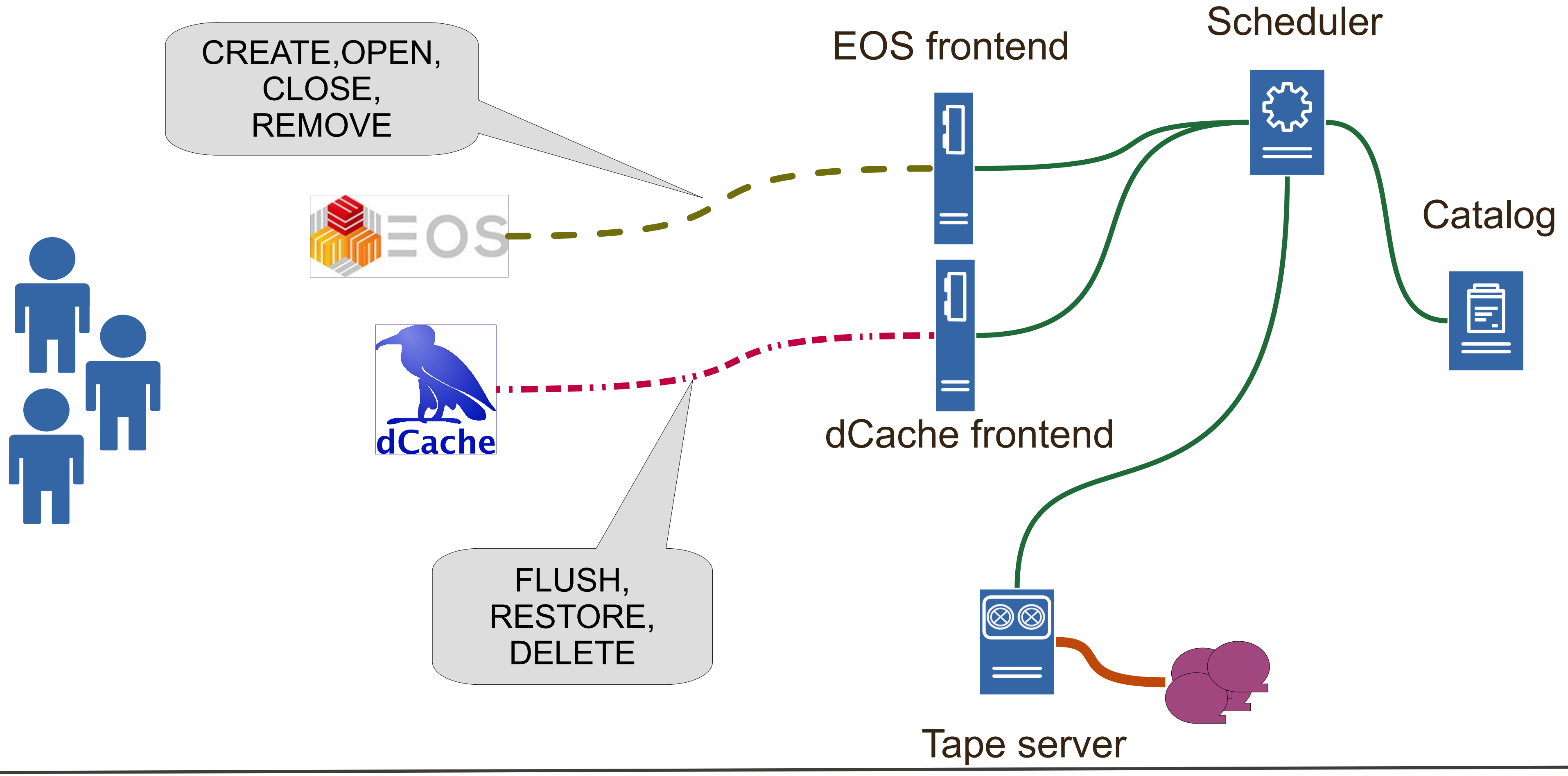
(Extremely) Simplified CTA design



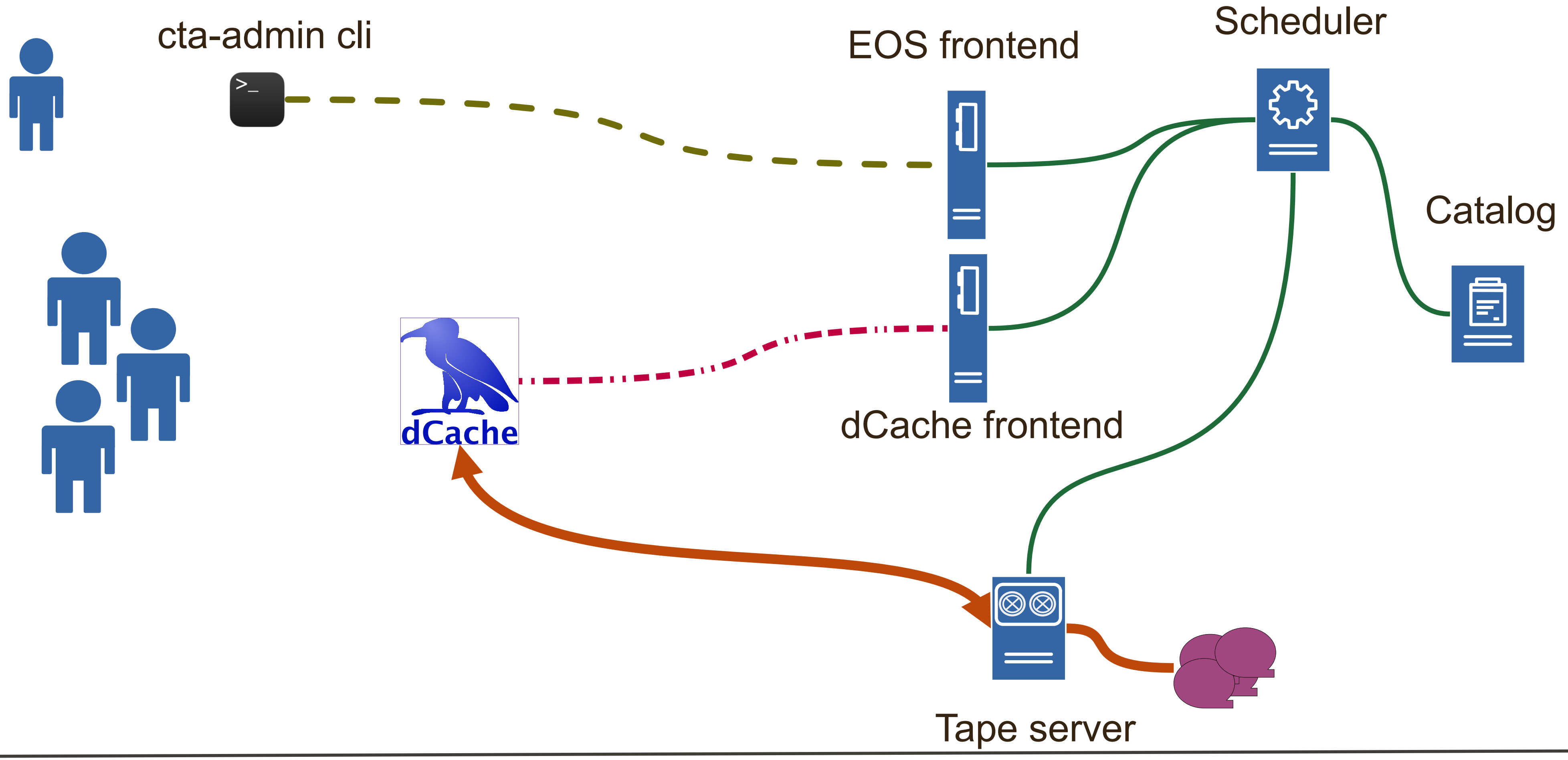
(Extremely) Simplified CTA design



(Extremely) Simplified CTA design



(Extremely) Simplified CTA design



Missing Parts (for DESY)

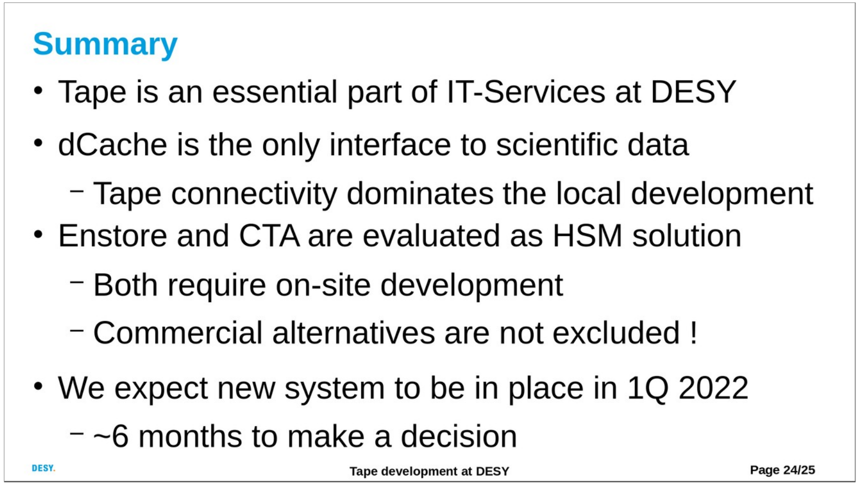
- Multiple tape formats are not supported
  - We still need (at least) to read old files
- OSM ⇒ CTA catalog migration
- dCache support out-of-box
  - No custom builds

Missing Parts (for/at dCache)

- Cancellation of Archive/Restore requests
  - File deletion while flush-to-tape scheduled
- dCache restart handling
  - Requests re-submission
- Switch to HTTP for IO path
- Switch to REST for error/status reporting

Next Steps at DESY

- Better operational experience
  - More FTE on CTA integration, support, development
- Large scale performance tests
  - More tape drives
  - Production HW
- Migration path
  - Copy or DB migration



Summary

- Tape is an essential part of IT-Services at DESY
- dCache is the only interface to scientific data
  - Tape connectivity dominates the local development
- Enstore and CTA are evaluated as HSM solution
  - Both require on-site development
  - Commercial alternatives are not excluded !
- We expect new system to be in place in 1Q 2022
  - ~6 months to make a decision

DESY. Tape development at DESY Page 24/25

Summary (DESY)

- Tape is an essential part of IT-Services
- dCache is the only interface to scientific data on tape
    - Tape related activities dominates the local developments
- We see CTA as the preferred tape software at DESY
    - The architecture matches our demands
    - Seamless integration with dCache
- The final decision should be taken by end of 1Q 2022
    - Massive testing will be done in Dec 2021

More info

- CTA branch with dCache support

  https://gitlab.cern.ch/cta/CTA/-/tree/cta-dcache

- dCache-cta HSM driver

  https://github.com/dCache/dcache-cta

- Documentation

  https://confluence.desy.de/display/~tigran/dCache-CTA+Test+Deployment