dCache integration with CERN Tape Archive

J. Chodak, M. Karimi, R. Lueken, S. Meyer, T. Mkrtchyan, P. Suchowski, C. Voß

The CERN Tape Archive (CTA) is an open-source storage management system developed by CERN to manage LHC experiment data on tape. Although today CTA's primary target is CERN Tier-0, the data management group at DESY considers the CTA as a main alternative to commercial HSM systems. We are pretty confident that dCache+CTA will address the data archival requirements HEP, Photon Science and EuXFEL at DESY. We hope that other sites that have dCache and tape libraries or plan to have one will benefit from our work and experience.

1. The Motivation

The tape storage plays a key role in data management strategy at DESY. As so called 'Tier-0' storage for PETRA-III and EuXFEL experiments at DESY the tape system must provide: multiple parallel high data ingest streams; support redundant copies for the same data on different media types; bulk recalls, automatic and manual data migration; automatic technology migration.

With dCache as the primary disk system in front of a HSM, the tape system should fulfill the following requirements:

- Maximize tape HW efficiency
  - Integration into DESY ecosystem
  - Integration with dCache tape interface
- Support Enterprise & LTO
- Daily turnover ~1PB
- Stable operation for the next decade
- Should be Open-source, adopting open standards
- Wide user and technology community
- State of the art integration/use development tools (CI/CD)

The CERN Tape Archive (CTA) is an open-source storage system developed by CERN IT storage group to replace the legacy CASTOR system that is used to manage experiment data on tape. Its architecture is designed to meet the requirements of LHC Run 3 as well as HL-LHC.

Out of the box, CTA comes with a *frontend* that communicates with EOS, the disk system deployed at CERN. However, as CTA's queuing system is not EOS aware, other frontend implementations are possible.
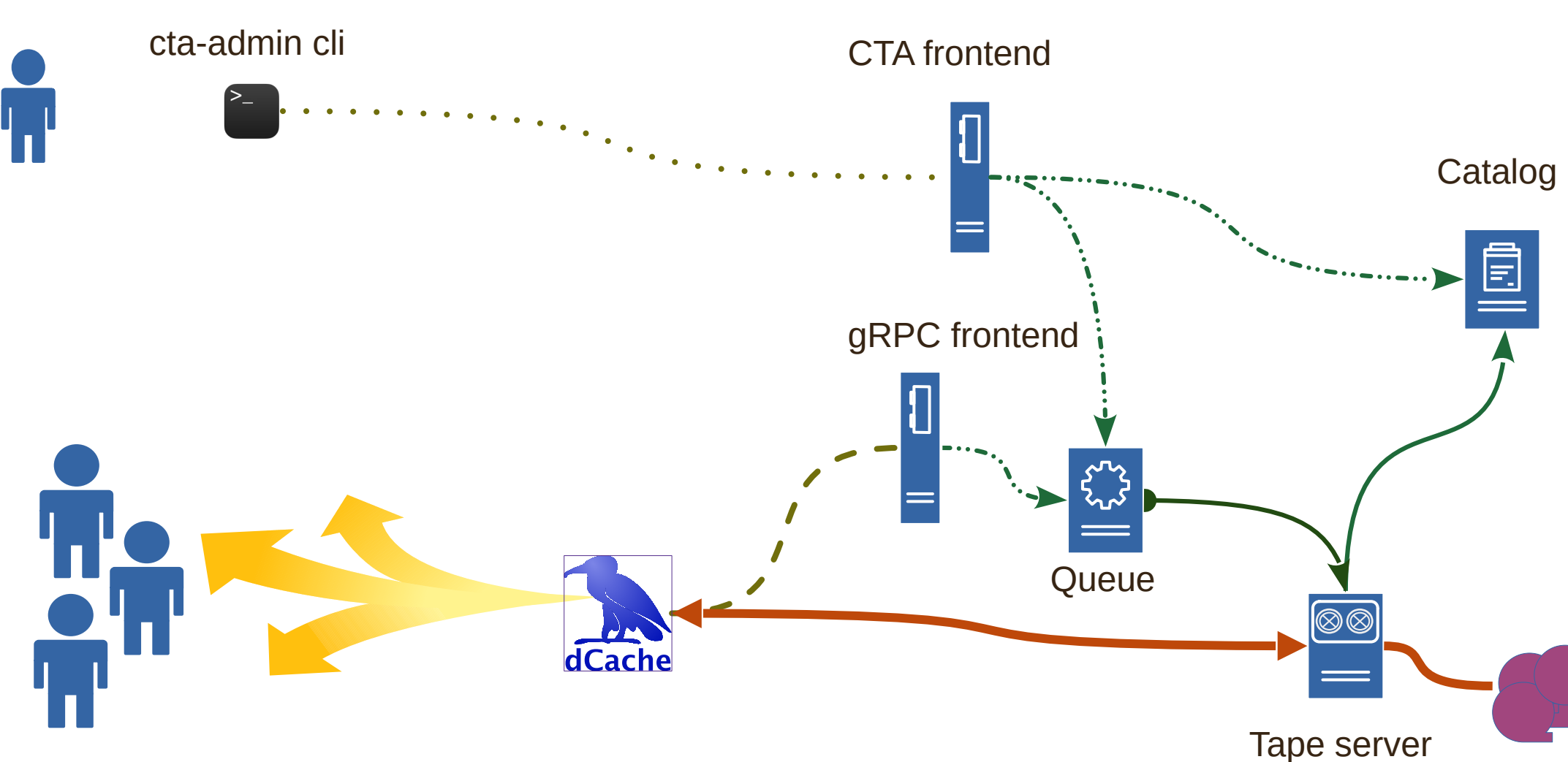
2. The Architecture of CTA

The CTA has two key components: frontend and tape daemon. The frontend accepts the requests, like archive, retrieve, delete or cancel, from the attached disk storage and puts them into request queue. If needed, the file catalog is updated. The scheduling logic is embedded into the tape daemons, which are per-tape drive processes and seek for matching tasks in the request queue. Once the desired number of requests are collected, the data is moved between disk and tape medias.
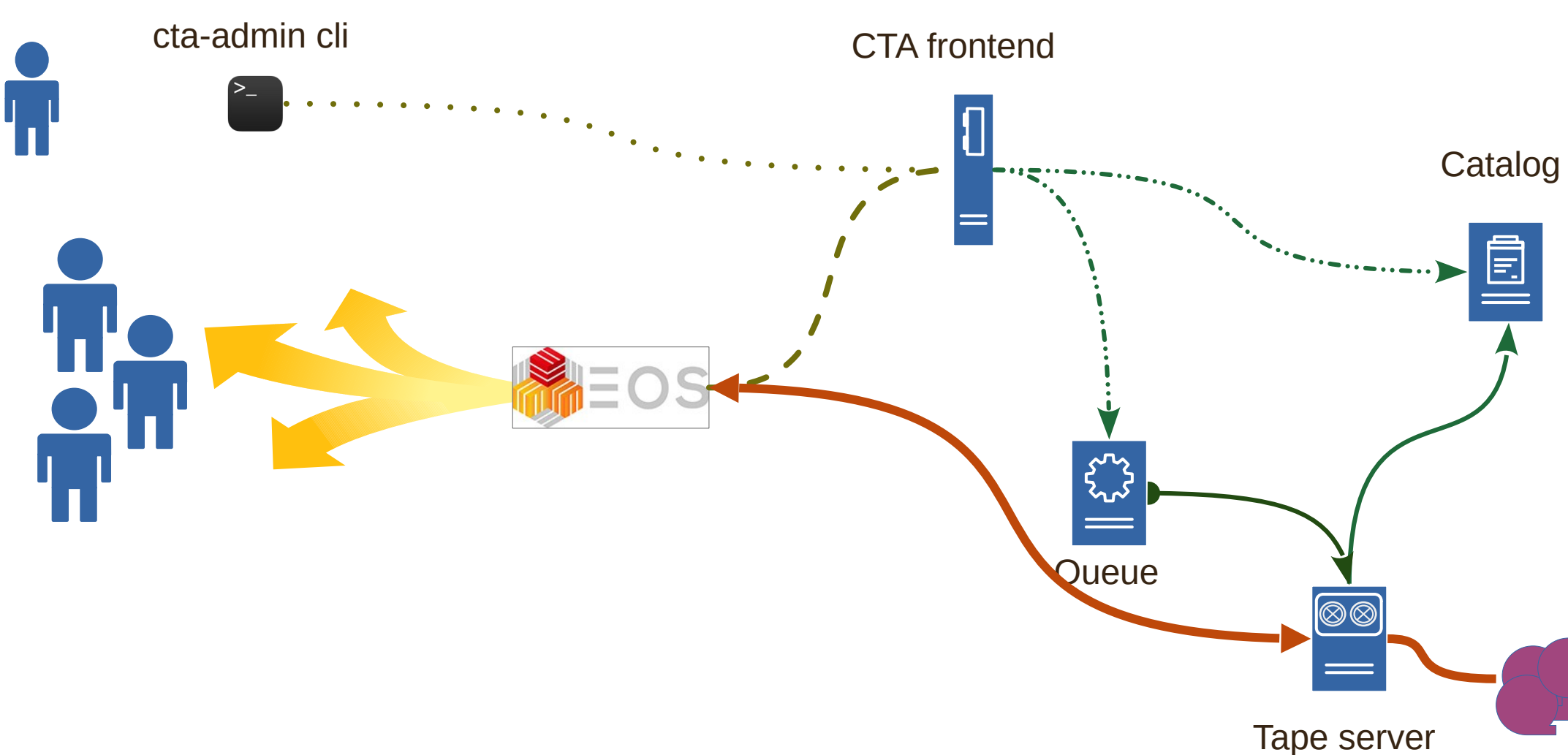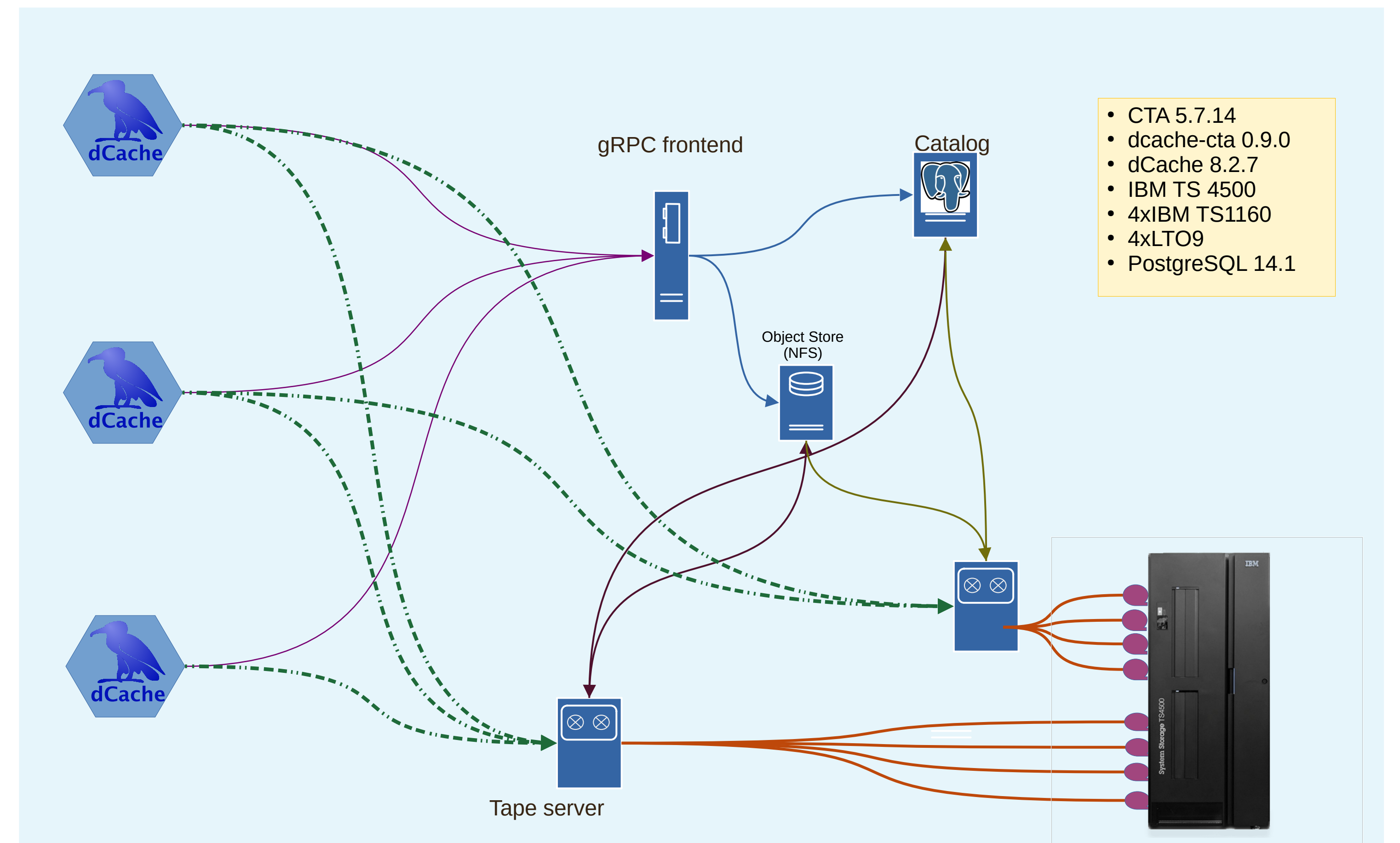

3. Integration with dCache

dCache has a flexible tape interface which allows connectivity to any tape system. There are two ways that a file can be migrated to tape. Ether dCache calls a tape system specific copy command or through interaction via an in-dCache tape system specific driver, called *nearline storage provider*. The latter has been shown (by NDGF, TRIUMF and KIT Tier-1s), to provide better resource utilization and efficiency. Thus, for seamless integration of CTA the dCache developers at DESY have implemented CTA specific nearline storage provider, called dcache-cta, and a corresponding frontend component for CTA. The communication between dCache and the new forntend is based on Google's gRPC library.
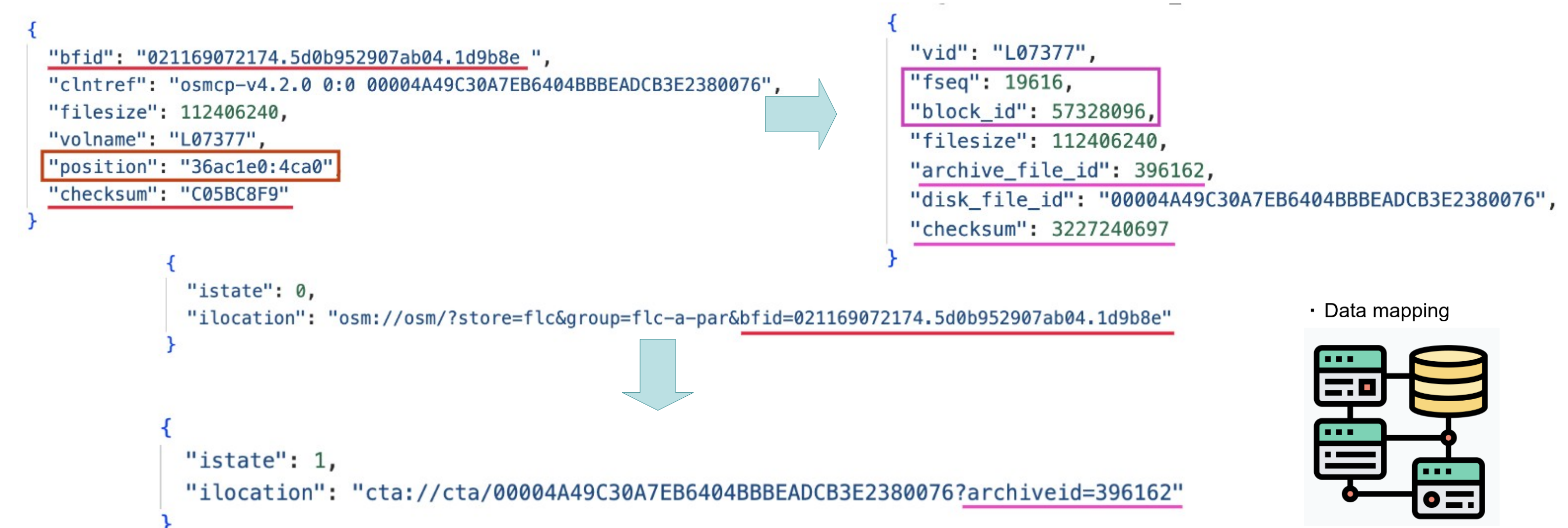

4. CTA Deployment at DESY

- CTA 5.7.14
- dcache-cta 0.9.0
- dCache 8.2.7
- IBM TS 4500
- 4xIBM TS1160
- 4xLTO9
- PostgreSQL 14.1

5. Migration to CTA

Metadata migration only. CTA can read OSM tapes. No physical re-write. All new Data is written in CTA tape format.

6. Current Status

- Seamless integration with dCache is merged into upstream CTA code at CERN
  - Requires dCache 7.2, or later.
  - The latest official CERN release 5.7.14 is deployed at DESY.
  - The proposed dCache interface is under adoption by EOS.
- The existing OSM tape format is supported for READ
  - The code changes are adopted by Fermilab data management team for ENSTORE tape format.
- The OSM tape catalog conversion procedure is ready and exercised multiple times.
- DESY is rolling out dCache+CTA deployment as it's primary solution for scientific data.
- Our deployment is replicated by other HEP sites
  - PIC Barcelona have successfully replicated our setup (currently dCache + ENSTORE).
  - Fermilab have demonstrated running dCache+CTA on their test deployment.
- RAL in UK plans to migrate to PostgreSQL from ORACLE based on our experience

7. Next Steps

- Handling of re-submits on dCache restart
- dCache based tests as a part of standard CTA CI/CD process
- Full monitoring and integration into DESY infrastructure

8. More info

https://www.dcache.org/
https://cta.web.cern.ch/cta/
https://github.com/dCache/dcache-cta