# SM.3 Separation of archives and disk pools/caches

## Description of topic(s)

This topic deals with the usage, configuration and management of tape archives and disk pools at Tier-1 sites.

## State of play

All of the LHC experiments seem to be working fine with (or towards) splitting disk caches from tape archives. ALICE, ATLAS and LHCb are split, while CMS has a work plan in progress.

Use of disk and tape by LHC VOs at RAL, an example multi-VO Tier-1:

- ATLAS have separate service classes for T0D1 and T1D0. Production jobs can read from either, while user analysis jobs can read from T0D1 only. The vast majority of outbound transfers are from T0D1. Data on T0D1 can be archived to tape using FTS to copy from T0D1 to T1D0.
- LHCb also have separate T0D1 and T1D0 service classes. Production jobs can read from T1D0, while analysis jobs read only from T0D1. Data on the disk caches can be archived to tape when necessary using FTS.
- The CMS setup is almost entirely T1D0, so all files, both from inbound transfers and output from jobs (apart from small temporary files), go to tape.

## Requirements and principles

The LHC experiment workflows set the requirements for this topic. Experiments need the ability to:

- be able to keep defined samples on disk, both for reprocessing and redistribution to other sites;
- allow (limited or not) user analysis on sites having custodial responsibilities, without destructively impacting their archival system. It is important to ensure that users don't read files that need to be recalled from tape, as this random access won't efficiently make use of the tape system and may impact production activities;
- process samples without immediately writing to the archive. For example, this enables the ability to run reprocessing at one Tier-1 and archive at another, or allow validation to be carried out before archiving.

## Technologies

Storage systems at the Tier-1 sites:

- dCache
- CASTOR
- StoRM/GPFS/TSM

Managing data transfers between caches and archives:

- FTS

## Recommendations and Observations

1. The experiments require split disk and tape, and are happy using FTS to transfer data from the disk cache to the disk buffer in front of the tape system.
2. None of the experiments want to "drop" HSM, as it brings useful functionality.
3. The experiments prefer the idea of using a single system (e.g. FTS) to manage both transfers internal to Tier-1s as well as external transfers, as dealing with two different systems would be significantly more complex. In other words, the interaction between disk and tape within a Tier-1 should not be considered any differently than any other data transfer. So, whether data is resident at a Tier-2 or on a disk cache at a Tier-1, it can be archived in exactly the same way.
4. Follow the development of FTS-3, and let the experiment operations teams work with the developers to ensure that we converge on something that will meet our needs.

## Unanswered Questions / Points of Contention

1. How should data movement between caches and archives be managed? Currently FTS seems to be the only tool available for scheduling and managing data placement, however, is there any other concept or architecture that would fit the problem better?
2. There are two views on how data on the tape archive should be accessed: either read directly from the tape buffer, or copied to an external disk cache first.
    - ATLAS and LHCb prefer to read directly from T1D0, and copying data from T1D0 to T0D1 is a rare occurrence (e.g. a recovery operation). In the alternative view both pre-staging and pinning can be considered as copying data from T1D0 to T0D1.

## Impact and Timescales

## Risks