CASTOR to CTA Migration

Michael Davis

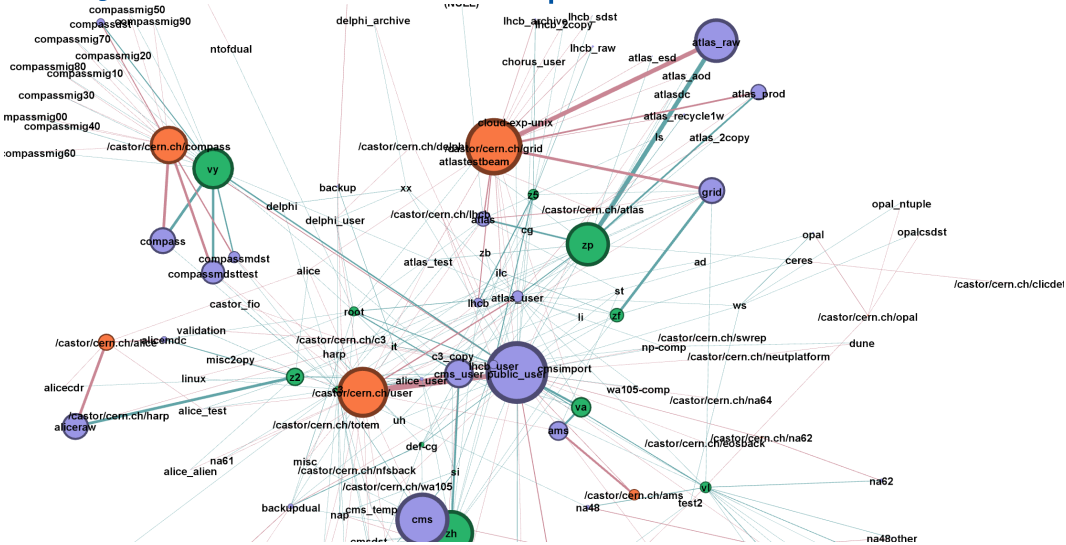
What will be migrated?

- Migration is **metadata only**
  - Number of files matters
  - Size of files is irrelevant
- CASTOR Catalogue → CTA Catalogue
- CASTOR Namespace → EOS Namespace

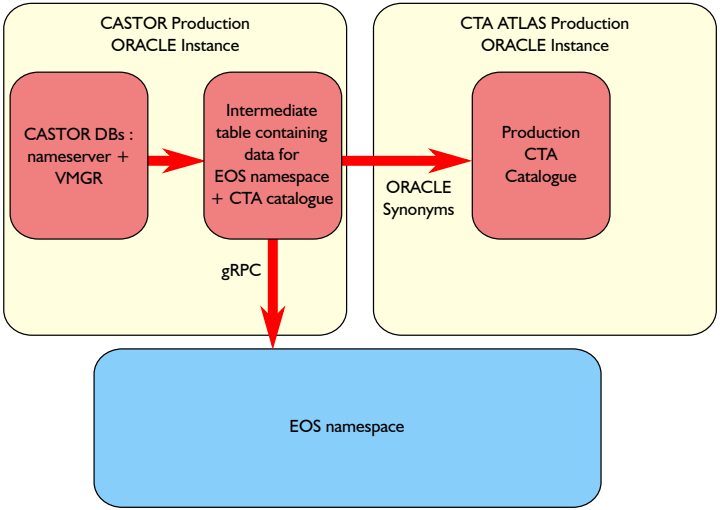
What will not be migrated?

- Files with no tape copy
  - zero-length files
  - deleted files
- (some) Permissions and Access Control
  - CASTOR Access Control Lists
  - POSIX `S_UID`, `S_GID`, `S_VTX` (sticky) bits
- Symbolic links

Analysis of the namespace



Migration Process



Migration Process

- Select tape pool(s) to be migrated
- Disable the tapes in CASTOR
  - Subsequent metadata operations on these files (delete, rename) will not be migrated!
- Copy metadata to intermediate table
- Populate CTA catalogue
- Populate EOS namespace
- Enable tapes in CTA
- Disaster Recovery
  - CTA is prohibited from writing to tapes imported from CASTOR
  - To return a tape to CASTOR, disable the tape in the CTA catalogue and re-enable the tape in CASTOR

CASTOR Catalogue to CTA Catalogue Migration

- Source and target DBs are both Oracle
- Possibility to migrate to a different DB post-migration
  - This option implies further development work
- Migrate tapepool by tapepool
  - Atomic unit for migration is the file/storage class
    Largest CASTOR fileclass is ≈60 million files
  - Rollback in case of failed migration

CASTOR Namespace to EOS Namespace Migration

- Migrate target directories to EOS namespace
- Migrate files to EOS namespace
- Current injection rate into EOS around 1700 Hz (6 million files per hour)
- Failures are recorded and can be replayed
- No automatic rollback in case of a partial migration

```
For each storage class:
  Migrate Catalogue
  Migrate Namespace
```

To Do List

- RAL has two CASTOR production instances. Will this map to two CTA instances?
- Profile CASTOR files by number of files per tape pool/file class
- Analyse CASTOR DB for corner cases and legacy issues
- Set up a test EOS+CTA instance and do a dry-run of the migration


CERN