From: **German Cancio Melia** German.Cancio.Melia@cern.ch
Subject: our chat today
Date: 10 July 2019 at 16:36
To: Massimo Lamanna Massimo.Lamanna@cern.ch

Hi Massimo,

just some quick notes from our chat today which I'll discuss with our developers. Just let me know if you have comments/corrections!

cheers, Germán

Data input:

- Overall numbers
    - From DAQ to EOSALICEDAQ++ (aka O2 disk farm):
        - O(1500) streams with 2GB files summing up to 100GB/s. (Each stream ~66MB/s). Up to 9PB/day (disregarding inefficiencies)
    - From EOSALICEDAQ++ to EOSCTA:
        - O(100) streams with 2GB files summing up to 30GB/s. (Each stream ~300MB/s). Up to 2.6PB/day to tape (1.3M new files / day, 100 drives, ~20Hz tick rate)
        - EOSCTAALICE instance will be ~800TB of raw(==usable) space.
- Issues:
    - How to do client-side back pressure when writing? ALICE could write faster than we are able to migrate (or our libraries could be stuck), and the buffer could fill up. Unlike CASTOR, EOSCTA has no job slots, so the only mechanism is to hold back if ENOSPC is returned when writing to file - which is painful.
        - a proper back pressure mechanism would need to be provided via XROOT (or not)? Me: Unlikely to have such a thing in the near term
        - as workaround, a "df" or "showquota" command could be used by ALICE for pausing transfers to EOSCTA. How will this actually look like?

Data retrieval - questions:

- ALICE will occasionally reprocess older run data during non-DAQ periods. This may imply retrieving O(10)PB "as fast as possible" onto EOSALICE(DAQ).
- In order to ease collocation efficiency, ALICE should pre-stage (via xrdfs prepare -s) as many files as possible. But can it submit 1-2 days (or more) worth of data (and thereof #files)? Would represent 1.5M-3M files. Would EOSCTA scale? How to do split-up otherwise, what are the limits?
- A back pressure mechanism will exist for files staged from tape to SSD disk, but this resolves only one part of the problem.
- Files need to be picked up (by ALICE) and transmitted to EOSALICE(DAQ) before any GC kicks in. How will ALICE know which files are ready? Polling is unlikely to scale and we have no notification system to inform what files are becoming ready for "pickup".
- Could ALICE use FTS for retrievals? Massimo: "just specify source, destination, and list of O(100K) or O(1M) files to transfer"