Idea for a Production Data Challenge

Frank Wuerthwein

UCSD/SDSC

DOMA Access Meeting

May 5th 2020

RAW Data Processing at HL-LHC

- Each of ATLAS and CMS collect roughly 0.5 Exabytes of RAW data per year, archived on tape at the T1s.

- Each want to run a processing campaign of RAW at the end of each data taking year.

- Maybe 2nd one at the end of each running period.

- RAW resides on tape until it is needed for processing.

- Roughly 40% of the RAW is archived in the USA

- 1Tbit/sec for a day ~ 10PB

- An exabyte processing campaign is 1Tbit/sec for 100 days.

Data Challenge

- Process 10PB of data in a day
  - Exercise one day of such a 100 day campaign
- From archive to HPC center and back.

Technical Challenges

- Process 10PB of data in a single day
  - Tape recall
    - How much bandwidth can we achieve from tape?
    - What's reasonable for buffer sizes and tape bandwidth?
  - Manage the limited disk buffer at archival T1
    - Tape recalls will be carousel style, i.e. buffer much smaller than the exabyte dataset.
  - Manage 1Tbit/sec network to an HPC center
    - Network bandwidth needs to be managed with tools like SENSE and AutoGOLE
  - Manage the disk buffer at the HPC center
- Co-schedule processing and all of the above.

FABRIC

FABRIC Core

FABRIC

- FABRIC is an NSF project that will build a network testbed across the USA by 2023 that provides 1Tbit/sec supercore, and a host of features for instrumentation etc.
  - 4 year project started in Fall 2019
  - Testbed will be operated for another N years after it is built.
  - Technical infrastructure strongly aligned with ESNet6 build out
    - ESNet is a collaborator on FABRIC
- It peers with various production networks at each of its endpoints.
  - The sum of US T1 and T2s across ATLAS and CMS will connect to it at >1Tbit/sec
- IT connects up in San Diego in the same data center as Expanse, a 90,000 core AMD x86 cluster.
  - Should also connect to various other HPC centers from DOE & NSF.

Proposal

- We get organized and apply to use this testbed for a variety of tests that build up over time to the 10PB/day data processing challenge. E.g:
    - Learn how to tag traffic.
    - Learn how to use SENSE etc. to schedule networks.
    - Benchmark out entire data transfer chain at Tbit/sec (Rucio, FTS, TPC, SENSE, …)
    - Learn how to co-schedule tape, disk, network and processing
- Do all of the above as a program of work over the next 4 years, with the 10PB/day processing as crowning achievement.
- Do it jointly between ATLAS, CMS, IRISHEP, …

Comments & Questions