EOS Performance Evaluation Framework

WP15 O^2 Disk Buffer Project

EOS Workshop, CERN
5 Feb. 2018

Pete Eby - ebypi@ornl.gov

Slides: http://cern.ch/go/CNF6
Gitlab: https://gitlab.cern.ch/alice-o2-wp15

OAK RIDGE National Laboratory

CADES

UT-BATTELLE

Disk Buffer in a nutshell

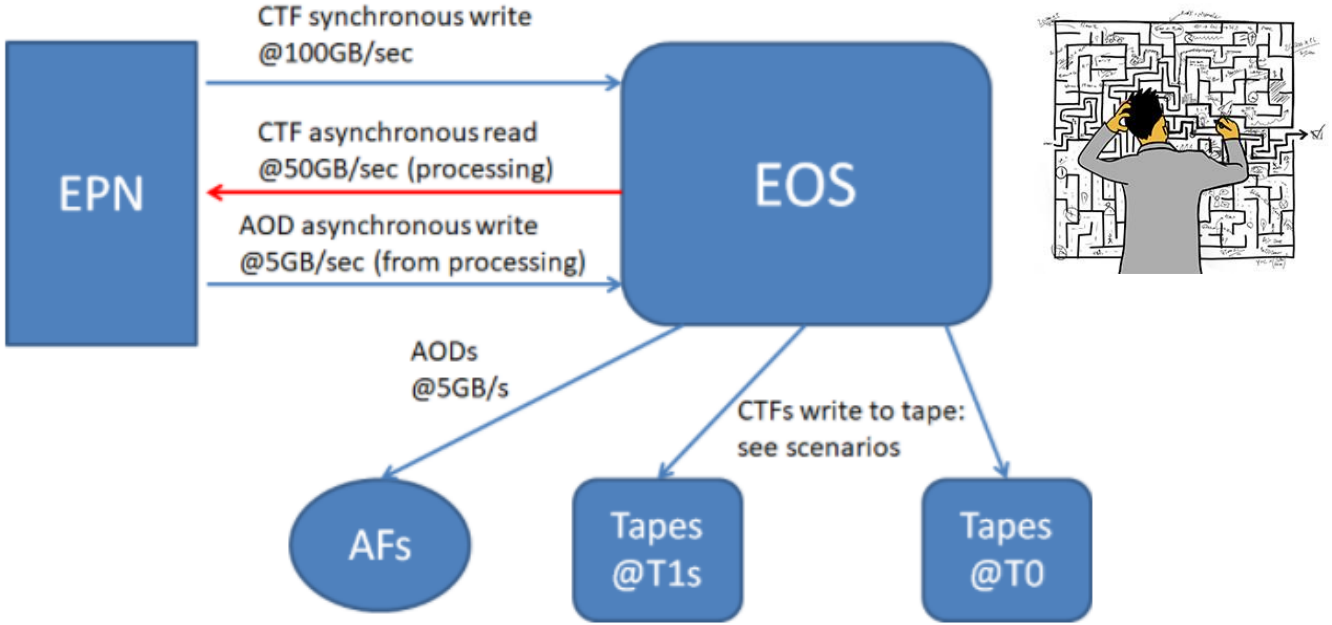

Figure 1 - Interaction between O2 EPNs, the O2 disk buffer and data export to analysis facilities and tapes - peak performance. Average rates are lower.

O2 Glossary

AF - Analysis Facility: a dedicated computing cluster running exclusively organised physics analysis,

AOD - Analysis Object Data: resulting from CTF processing, contains all information necessary for physics analysis

CTF - Compressed Time Frame: 20ms snapshot, containing 1000 events. One time frame duration is 11245 orbits/s = 22.8ms, CTF is immutable

EPN - Event Processing Node: a node, which collects the partial Time Frames from the FLPs, assembles them in a complete TF, performs second level compression and produces the final CTF.

ESD - Event Summary Data: resulting from CTF processing, contains all information necessary for detector calibration.

FLP - First Level Processing node: a node which collects part of a detector uncompressed RAW data and does a first-level compression, producing a partial Time Frame.

Test Framework Stakeholders - Who profits?

1. O2 Disk Buffer WP 15 Project
   - During peak synchronous Pb-Pb acquisition
     - EOS Disk buffer will receive ~2PB a day
   - Need to ensure the 60PB system meets performance requirements

2. EOS and wlcg community - def.self()
   - Evaluate SE performance for new and existing deployments
   - Using agreed upon synthetic (fio, ior, etc.) and simulated data tests
     - Same yardstick
   - Under productions levels of concurrency
   - Which are reproducible
   - For initial benchmarking, or troubleshooting
   - Easily share results

Evaluation and Objectives

O^2 Disk Buffer IO requirements defined in WP15 capacity estimate paper

Throughput Summary*

- During peak synchronous Pb-Pb acquisition - with simultaneous processing
  - 105 GB/sec writes 320 GB/sec
  - 55 GB/sec read 120 GB/sec
- Mix workloads
  - Synchronous and asynchronous processing

Test Framework Goals

Validate O^2 disk buffer storage environment through the development of an EOS testing framework which uses synthetic (fio, etc.) and simulated O^2 workloads under expected levels of concurrency for standardized, reproducible results and SE performance analysis.

Test Parameters

O2 disk buffer testing outline derived during LBNL Nov 2017 meetings identified high level parameter categories, including

- Synthetic (fio, etc.) vs simulated workflow tests
- Concurrency vs raw throughput
  - Optimizing number and size of file systems
  - Possibly mapped to number of jobs from FLPs and EPNs
- Operation under limiting conditions
  - Max iops, max occupancy, full filesystems, rebuilds, etc
- Access control requirements
- EOS/xrootd Internals
- Network simulation
- Failure scenarios and impact

Hardware Configurations

- 9PB test system available Q3
  - Testing framework under development to be used to iterate through various storage configurations and tunings.
- Parity method to be determined
  - hw, sw, zfs, rain, etc.
- Determine optimum eos fsids, group size, blocksize and alignment, etc. for sustained write performance and subsequent mixed workload
- Remote reconfiguration of raid, filesystem tunables, etc.
- Drives types, capacities
  - Capacity vs greater spindle count of smaller drives, more eos filesystems, etc.
- Filesystem tunings adjusted / re-optimized for lifecycle workloads

Test Framework Next Steps...

- Initial code implementations
  - Define modules for environment data collection, testing, reporting, monitoring, etc.
- Obtain access to CERN-IT evaluation system
  - Begin iterative testing of various storage layouts, parity, filesystem tunings, etc.
- Reporting and sharing
  - Results committed to community git repo
  - Summarize environment and performance results
  - Comparison and visualization

Call for collaborators!

What this too? This is the perfect time to collaborate

- Input from EOS users is desired
  - Design and implementation suggestions
  - Code contributions
  - Testing and feedback

How to contribute?

- Email: ebypi@ornl.gov and alice-o2-wp15@cern.ch
  - Subscribe to alice-o2-wp15 via CERN egroups

- GitLab: https://gitlab.cern.ch/alice-o2-wp15/eos-testing-framework
  - Use that issue tracker!