WP15 status update

L. Betev

General

WP15 = O_2 data buffer

Objective – design and deliver O2-attached storage; assure its compatibility with distributed Grid storage (disk and tapes) and existing local and distributed data management tools

WP15 operation and structure

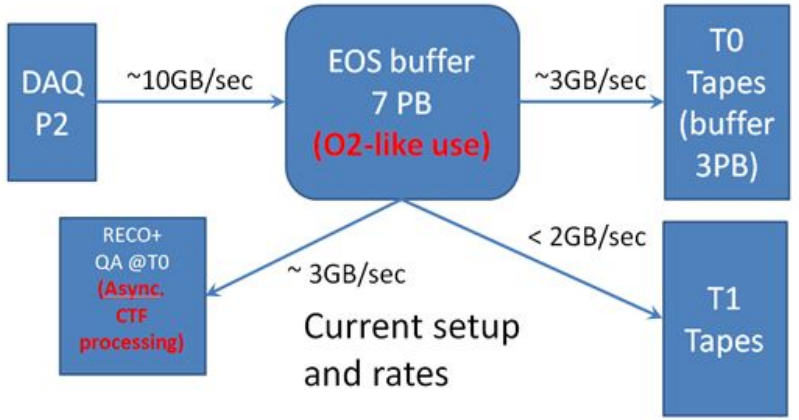
- Minimize duplication of effort and save manpower
- Heavily leveraged with CERN IT know-how and purchase procedures
  - Selection of hardware
  - Tender and procurement
  - Testing and validation
  - EOS installation and configuration
  - Tape interface (CTA)
  - TBD: Operation, maintenance
- WP15 membership
  - LBNL (J. Porter), ORNL (P. Eby, M. Galloway), CERN (L. Betev)
  - CERN IT participation - B. Panzer, E. Bonfillou, A. Peters + EOS developers

Major milestones completion

153: Design of abridged evaluation process - testing in production

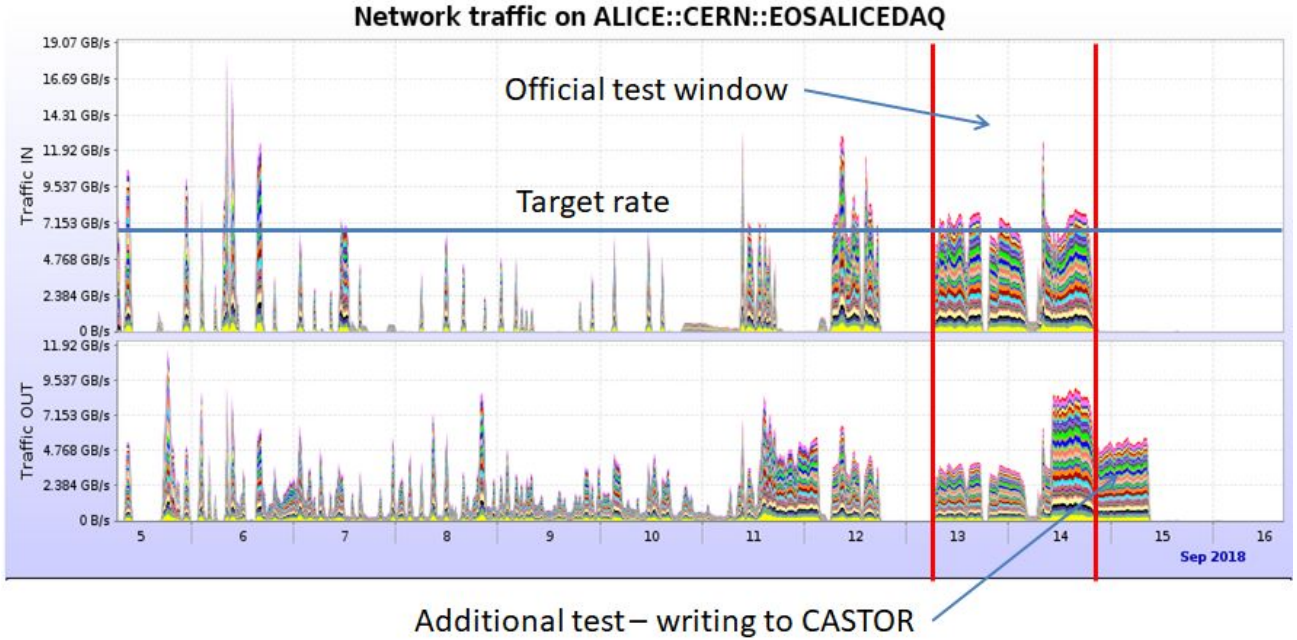
- Executed together with the Pb-Pb data buffer test
- "10% validation" test for the O2 storage
- Includes all data flows foreseen in the TDR
- Scaled data rates

Total I/O rate O2 = 150GB/sec



Key results from the test

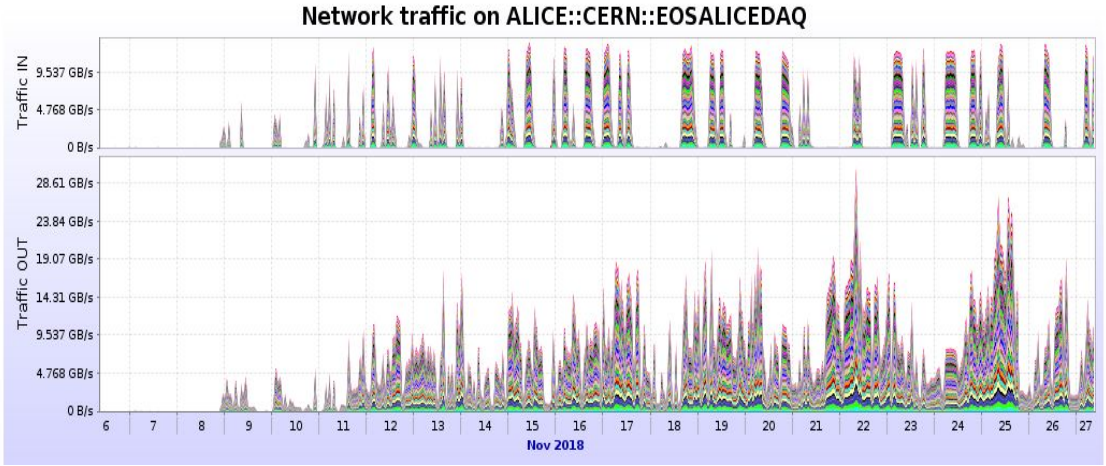
40 hours Pb-Pb data taking simulation



- Even load distribution
- Capacity of pool 6PB (10% O2 buffer)
- 50 concurrent writes (10% of expected EPNs in synchronous mode)
- 600 concurrent reads (10% of expected cores in asynchronous mode)

Key results from the test (cont.)

24 days of Pb-Pb data taking and processing - all parameters well above the initial targets



Data IN (from P2), max 12GB/s

Data OUT (to processing, CASTOR, T1s), max 30GB/s

Completed and ongoing milestones

154: Decision on disk buffer location from CERN IT (following official request)

- Buffer to be located at P2, no space in the CC, ***to be revisited end 2019***

155: Hardware architecture of the O2 buffer - document, with CERN IT

156: Disk pool management with EOS - decision on responsibilities sharing

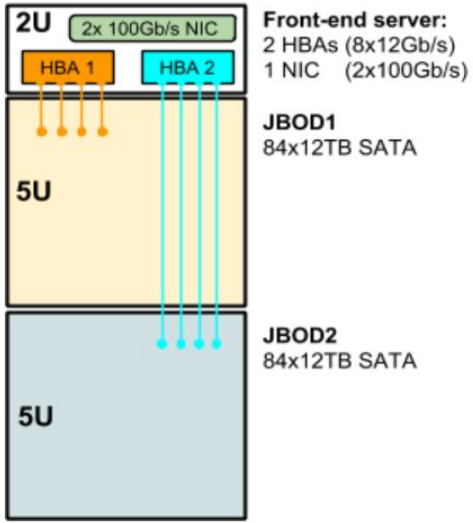
- To be discussed with IT

162: CTA test and evaluation, plan for O2 buffer integration with CTA

- EOS***CTA***ALICE***PPS*** prototype deployed - tests will start soon

O2 buffer hardware architecture

- Size: 60PB, I/O rate: 150GB/sec
- Modular storage - building blocks (SSU) - high density JBODs, managed by a head node with 2x100Gb/s ethernet

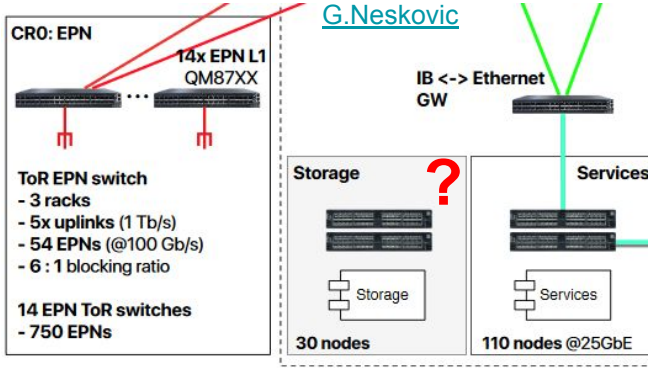


Element	Capacity and size	I/O Rate	Remark
JBOD	Up to 84 3.5” SATA drives, 5U rack space	4x 12Gb/sec channels	Disk I/O rate is not limiting factor
HBA	PCIe 2 channel adapter, hosted in server	4x 12 GB/sec	Max 48Gb/sec/JBOD
Network adapter	PCIe 2 chanel 40Gb/sec NIC, hosted in server	2x 100Gb/sec	Max 200Gb/sec/server
Server	2U rack space server unit	Max 96Gb/sec with 1 NIC channel	

~30 SSUs, total theoretical throughput ~750GB/s

Critical decision on network interface

- The current storage architecture assumes ethernet
  - This is what CERN uses for LAN/WAN
- O2 network architecture (ethernet/infiniband) ongoing discussion
- Consequences for WP15 if IB for storage
  - No common purchase with CERN IT, need to do our own tender
  - No support for testing/configuration/installation and operation - large amount of detailed studies and actions here
  - No sufficient manpower to execute all of the above on our own
- Unknown connectivity aspects with the external world (aka T0/T1s)
  - Likely not a big problem, but requires another dedicated study
- To continue the work on WP15, we need decision on network ~now



Major milestones

- For vertical slice assembly and simulation challenges

157: Prototype disk buffer for O2 - quotes and purchase

- If we want additional capacity, beyond the IT offering
- May 2019

- For O2 buffer delivery

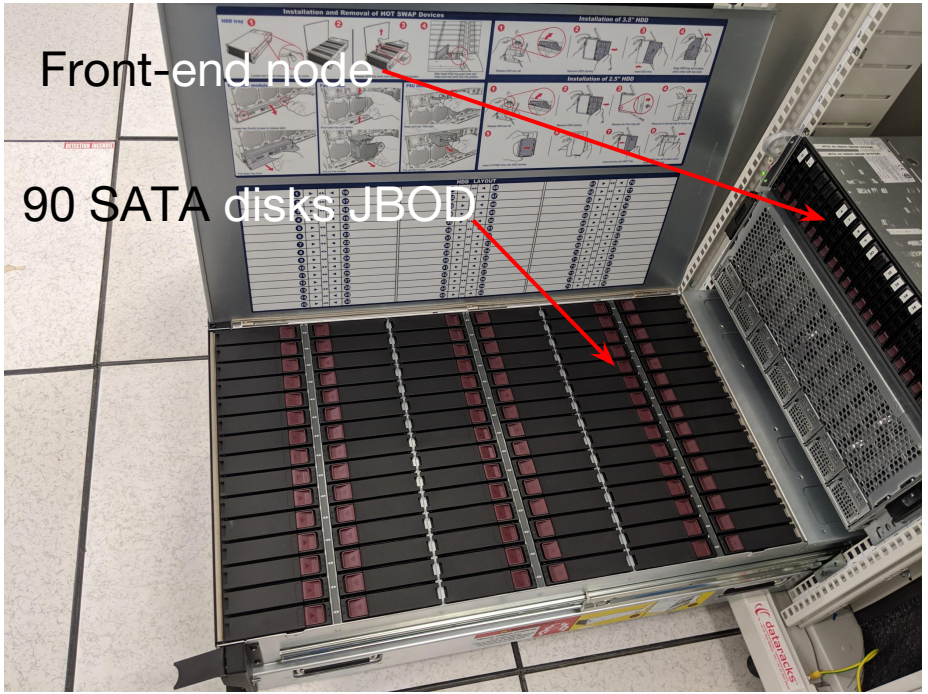
159: Decision to go with IT on buffer purchase

- November 2019

More on the vertical slice

- Offer to use a high density JBOD currently being tested by IT

- If we want to test also redundancy strategy (EOS 2D RAIN)
  - We should plan to buy at least one more node
  - High level of interest/involvement from CERN IT and KISTI
  - ORNL knowledge and experience with ZFS is critical for this step



WP15 - summary

- Milestones defined for the entire project duration
- All milestones completion on track
- Critical decision needs to be taken on network interface
- Next major milestones
  - Purchase ~few PB buffer for O2 prototype (May 2019)
  - Join IT tender for the bulk purchase (November 2019)
  - Actual purchase of 60PB ~mid 2020