EOS Update

D. Duellmann, CERN IT-DSS

Grid Deployment Board
11th May 2011

EOS Goals

- EOS development complements CASTOR at CERN in the disk pool area
  - Focus: Many concurrent analysis jobs
    - T0 and export infrastructure stay unchanged
- Design Principles & Choices
  - Decoupled from Archive
    - EOS has no tape connectivity
    - On disk data is defined by experiment work flow system
  - High performance in-memory name space
    - EOS supports significantly higher meta data access rates than RDBMS approach
  - Tuneable redundancy & performance
    - Via file replication and (later) block encoding
  - Asynchronous h/w replacement
    - EOS should not require immediate admin intervention after typical h/w problems
    - Data should stay accessible

Direct Consequences

- EOS is not an archive
  - CASTOR will stay fully supported
  - CERN will continue to use CASTOR as archive
    - ...but progressively move disk-only pools to EOS in consultation with experiments
    - CMS and ATLAS have identified priority pools and defined a migration plan
    - ALICE and LHCb are interested in disk-archive split but have not yet concrete migration plans to EOS
- EOS impact on CASTOR support?
  - Disk pool support will not be removed from CASTOR
  - CASTOR T1 sites are not be required to follow EOS approach and will be fully supported in their use of CASTOR

From Tests to Production

- Last year to now..
  - several months test period with ATLAS production jobs
  - CMS tested (at lower rate) with heavy ion data
- Performance and reliability has been reported in this meeting (demonstrator review) earlier
  - No significant update yet, since test infrastructure and software have been replaced by production version
- Received request for production service from ATLAS and CMS
  - Interest from ALICE and LHCb
- Defined migration schedule for ATLAS & CMS
  - May: s/w release + h/w available
  - June + July: migration at 1-2 PB scale per experiment from CASTOR to EOS
    - using the experiment workflow systems
  - Total disk volume (CASTOR+EOS) within request
    - no additional storage h/w for experiments

New Functionality in V0.1.0

- Block level checksum support
  - adler, md5, crc32, sha1 and hw accelerated crc32c (blocksizes from 4k -1M).
  - media scrubbing - adaptive file & block level checksum scanning on storage nodes with configurable rescan intervals
- Monitoring & Resource Management
  - storage+IO views by space, disk group, node & filesystem
  - quota nodes with logical & physical space reporting
  - load balancing based on network & disk utilisation
  - average latency measurements & counter for all name space commands by user
- High Availability
  - active-passive failover defined via DNS alias
- Usability
  - simple fuse daemon with statvfs (df) functionality
  - simple ACLs with E-group integration
  - redirection on ENOENT defined on directory level
    - to redirect from EOS to CASTOR during pool migration
  - black-whitelisting of user/groups or hosts (with admin interface)
    - and global system stall to hold client access to storage nodes
- Automation (on-going)
  - automated draining of filesystems
  - automated filesystem rebalancing
  - internal and external transfer queues

Can EOS be used outside CERN?

- Short answer: Not yet…
- Software exists in public repository
  - V0.1 release candidate is now available and will be used for first production phase at CERN
  - Anyone interested in evaluating the code can contact us for pointers to git repository and internal documentation
    - Be warned: at this point this requires development level knowledge and we can not provide much help from the small development / deployment team
- Deployment procedures are emerging together with first service at CERN
  - Again: we can open existing docs for people who have a genuine interest and can abstract from many CERN specific deployment components
  - We can not support any production deployment outside CERN!
- We will regularly report on the progress in the GDB and do think this is currently the most effective way to stay informed.
- After the first deployment phase (several months) at CERN we can review the interest from outside CERN

Can EOS replace XYZ at T1, T2, T3?

- Short Answer: conservative assumption is “No”
- Too early to speculate about that at this point
  - The scalability and performance benefits of the EOS approach are not yet fully confirmed even at CERN
  - The so far good stability of EOS needs to be proven over a longer production phase
  - The operational procedures do not yet fully exist and may still change with s/w releases. One of the main benefits would be reduction in operations coast
  - Also it is not clear if the specific scale targets and goals for CERN will map to similar needs and advantages at other sites
- The current development and deployment teams can not support an external user base during the upcoming consolidation and validation.

Summary

- EOS has moved from proof-of-concept (WLCG demonstrator) to planned production service for ATLAS and CMS
  - planning discussions are well advanced
  - new s/w release being functionality tested now
  - deployment of EOS h/w ongoing in May
  - will stay in contact with other interested experiments
- Project progressing according to plan and show promising results
  - limited manpower and significant change rate in software and deployment procedures is not yet suitable for production deployment outside CERN
- Second half of this year will be crucial for confirming EOS advantages at CERN
  - Results will be publicised here (GDB)
  - Stay tuned...