EOS AT THE VIENNA T2

EOS Workshop 2024

Erich Birngruber, erich.birngruber@gmi.oeaw.ac.at

Ümit Seren, uemit.seren@gmi.oeaw.ac.at

Vienna Biocenter, Austrian Academy of Sciences

https://www.clip.science/

VIENNA T2: COMPUTE

- Project CLIP (CLoud Infrastructure Project), OpenStack based

- VBC Compute center for HTC workloads (Biology, physics, various others)

- Job scheduler is SLURM (+ HT-Condor CE)

- 200 compute nodes, 250TB scratch filesystem, 100GBit/s Ethernet interconnect
  - CPU / compute nodes (total ca 8000c)
  - High-memory nodes (2TB)
  - GPU nodes 112 GPUs total

- Shared "public" partitions for generic workload
  - Separate partition for grid workload: 40 nodes, ca 1500 cores
  - Multiple experiments workload: Alice, Belle2, CMS
  - QoS + Fairshare for balancing between experiments

VIENNA T2: EOS STORAGE

- Deployed early 2020, production mid-2020 (EOS v4.7.x)

- Today: EOS 5.2.17

- RAIN 5+2 layout

- Converged instance for 3 experiments
  - Alice
  - Belle2
  - CMS

- Extra services
  - FUSEx mount on SLURM computes
  - For other physics data, and easy access

VIENNA T2: EOS HARDWARE

- On baremetal commodity servers, 6.3 PB total raw capacity

- 3x JBOD with 84 HDDs @ 12TB as building blocks
  - Added 2x JBODs @ 20TB in 6FSTs in 2022

- EOS installation
  - 3 mgm with quarkdb (on NVMe SSD)
  - 15 FSTs
    - 28 x 12TB (9 FSTs) or 20TB (6 FSTs) HDDs attached (dual SAS-3 12 Gbit/s)
    - 3 FSTs per JBOD
    - 2x 40Gbit/s Ethernet

- Campus 40 Gbit/s uplink

- Current usage: ca 60+%, quarkdb ca 152GB (39 mio. files, 4.59 mio. dirs)

VIENNA T2: OPERATIONAL TASKS 2023

- Major update to EOS 5 ('22)

- EOS system extension, doubling capacity ('23)
  - Deploying additional hardware: 6 FSTs (28 disks @ 20TB), 2 JBOds (@ 84 slots)
  - Fully automated deployment: Host + EOS + Firewall

- EOS rolling software upgrade automation ('23)
  - Manage repo mirror, Quiesque IO, stop service, install packages, (reboot), check service status

- update to EOS 5.2.0 ('23)
  - Multiple improvements for CMS
  - Scitoken, ZTN support

- Certificate renewal automation ('23)

OPERATIONAL TASK: UPDATE AUTOMATION

- Ansible and Satellite (Pulp) are used for system management

- Workflow
  - Satellite mirrors upstream repos (continously)

- Playbook 1
  - Snapshot of repo gives consistent state across machines

- For each host
  - Update 1 FST at a time (set node to off, wait, udpate, reboot, check)
    - Continue only if success
    - Don't interrupt requests, RAIN N+2 layout keeps all data available
    - MGM updated last (minimal service interruption)

- Update the whole instance (15+3) unattended

OPERATIONAL TASK: CERTIFICATE RENEWAL

- Sectigo is certificate authority for GEANT organisations

- They provide an ACME endpoint
  - But not for Grid certificates 💔

- They provide Ansible automation for the API
  - But you have to pay extra for it 🤡

- However, the API is quite simple for renewing purposes
  - Create key, CSR, request auto-renew cert (if it does not exist)
  - For each host: find existing / matching certificate in API (matching CN)
  - Download "newest" certificate
  - No service restart required, xrootd will detect new certs (checks every 8h)
  - Update ca-policy packages, while we're at it

VIENNA T2: SUMMARY / OUTLOOK

- Next steps
  - Use upstream monitoring metrics for prometheus, https://github.com/cern-eos/eos_exporter
  - Update OS on EOS nodes from Centos7 to Alma9? – experiences with in place update?
  - Deploy CERNBox

- Whishlist
  - EOS client auth vs Alice token auth for (fuse) compute nodes (quota, find, etc)

- Thanks to the active community: Elvin, Andreas, EOS forums!

  Thanks to the organisers of the TechWeekStorage24, CS3 and EOS workshop!

VIENNA T2: CODE

- https://www.clip.science

- https://github.com/CLIP-HPC/clip-grid-eos (private repo)
  - Ansible playbooks
  - Prometheus exporter
  - Other config details
  - (ask us for collab invitation Github @ebirn, @timeu)