

# Experience with simple cluster setups

Martin Val'a

February 3, 2017

- Small groups (tens of users) need cluster and maybe expand it later
- Minimum set of components for simple cluster (easy installation)
  - Authentication (FreeIPA - Kerberos, LDAP and CA)
  - Shared Storage for data and home directories (EOS)
  - Software distribution (CVMFS)
  - Job distribution (SLURM)
  - Development, issue tracker, continuous integration and chat (GitLab+Mattermost)
  - Meetings (Indico)
  - Data sync directory (CERNBox)
- 3 independent instances running already and they are happy
- Most of the time clusters are not fully loaded (maybe federation and share resources)

- Hybrid (CPU/GPU/XeonPhi) cluster
  - 1x CPU blade, 2x12 cores x2 hyperthreading = 48 cores, 128Gb RAM
  - 1x MIX blade, 48 cores, 1xNVIDIA Tesla K20X, 1x Intel Xeon Phi Coprocessor 5110, 128Gb RAM
  - 1x PHI blade, 48 cores, 2x Intel Xeon Phi Coprocessor 7120P, 128Gb RAM
  - 2x GPU blades, 48 cores, 3xNVIDIA Tesla K40, 128Gb RAM
  - 4x GPU blades, 56 cores, 2xNVIDIA Tesla K80, 512Gb RAM
  - 2x file servers (48 TB)
- 2 machines for services (VM)
  - Gitlab (<https://gitlab-hlit.jinr.ru/>)
  - Indico (<https://indico-hlit.jinr.ru/>)
  - Monitoring (<https://stat-hlit.jinr.ru/>)
  - CVMFS Strantum 0 and 1 ([hybrilit.jinr.ru](http://hybrilit.jinr.ru))
  - ...
- Thanks to: Oksana Streltsova, Alexander Maiorov, Yuri Butenko, Mikail Matveev and Hybrilit Team

- CPU cluster
  - 50x8 core machines with 1 local disk data storage (local space)
  - 2x file servers (57 TB) (default space)
  - 3 machines for services (VM)
    - Gitlab (<https://gitlab.alice.tlabs.ac.za>)
    - Indico (<https://indico.alice.tlabs.ac.za>)
    - Monitoring (<https://monitor.alice.tlabs.ac.za>)
  - Ixplus machines
- Thanks to: Zinhle Buthelezi, Sehlabaka Qhobosheane, Nathan Boyles and Fabrice Nininahazwe

- CPU educational cluster
- Students in age 15-18 with small programming knowledge
- 2 physical machines
- 12 core machine with all services with outbound connection
- 4 core machine with 4 VM (slurm and EOS storage) local network only
  - n1 - EOS MGM and SLURM manager
  - n2-n4 - EOS FST and SLURM workers
- lxplus machine
- Thanks to: Rudolf Sviantek, Matej Fedor, Dominik Matis, Branislav Beke, Filip Jenik, Jaroslav Murín and other students

- EOS 4.1.XXX (currently 4.1.11/14)
- Storage for user's DATA
- Home directories (We drop one year ago, but now testing with better and faster FUSE)
- Replica mode for user's home directory
- Scratch directory with cache for 1 month (soon)
- Quota system (testing)

- Rename CERN.CH to our kerberos realm and remove gsi in xrd.cf.mgm
- `systemctl eos@*` doesn't work as expected
- `/etc/sysconfig/eos` vs `/etc/sysconfig/eos_env` (`eos_env` is not enough)
- vid gateway add with full hostname (need to also add short name)
- `/eos/instance/mvala` (mvala is owner and he cannot remove directory)
- `eos find --online` is showing files which are not finished with writing
- `eos find --offline` shows file which are offline
- http API request is returning html web page in case of permission denied
- <https://copr.fedorainfracloud.org/coprs/mvala/saske/> for building rpms

# EOS requests still not resolved

- CERNbox rpm are missing (EOS-1700)
- Unable to create home directory //mgm; permission denied (EOS-1666)
- What is d\_mem\_sz\_diff in fsck? (EOS-1610)
- rootjs and EOS (EOS-1409)
- WFE and bash failed (EOS-1604)
- RPM for fedora 24 (EOS-1524)
- Filter file name with regular expression in eos -b find (EOS-1441)



# EOS mgm namespace recovery (EOS-1654)

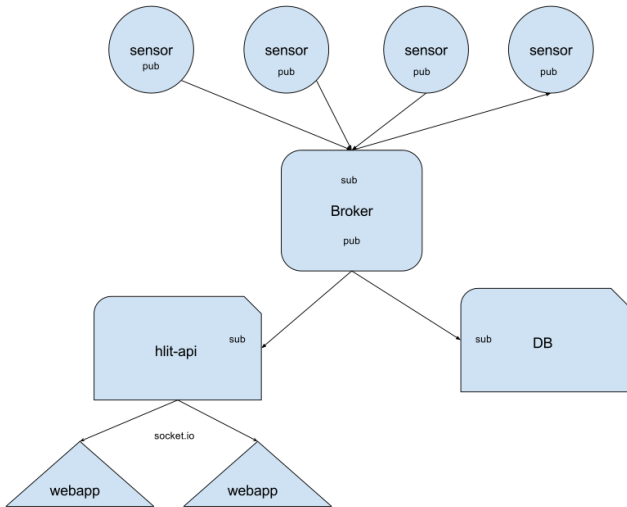
- Didn't save namespace
- I did recover files
- Now i know nice command to find out lfn from fst

```
[root@eosfst1 ~]# getfattr -n user.eos.lfn /var/eos/fs/11/00000000/0000000d
getfattr: Removing leading '/' from absolute path names
# file: var/eos/fs/11/00000000/0000000d
user.eos.lfn="/eos/hybrilit.jinr.ru/scratch/hkdio/test/hkdio.root"
```

- Extend REST api with find and ls commands to JSON (EOS-1697)
- Problem creating file bigger then 10 MB file via ROOT (EOS-1677)
- Problem in registration of disks in eos (EOS-1669) (eosadmin from fst can help)
- and many more.
- Thanks for support

- <https://gitlab.spseke.sk/dmatis/eos-web> (Dominik Matis)
  - <http://lxplus.spseke.sk:10007>
- High school student (18 years old)
- Used in project
  - NodeJS
  - Angular2
  - SocketIO
- Problem with EOS online status (server has to do API requests)
- Maybe EOS can publish via ZeroMQ from MGM like we do in next project

# Projects in development - hlit-api



- Thanks to Alexander Maiorov, Yuri Butenko
- <https://gitlab-hlit.jinr.ru/hybrilit/hlit-api>
- <https://gitlab-hlit.jinr.ru/hybrilit/hlit-webapp>
  - <https://stat-hlit.jinr.ru/>
  - <https://monitor.alice.tlabs.ac.za>
- Project started to monitor HybriLit cluster
- Used in project
  - NodeJS
  - Angular2
  - Socket.io
  - ZeroMQ

- <https://gitlab.spseke.sk/mfedor/SALSA/> (Matej Fedor)
- High school student (18 years old)
- Used in project
  - C++
  - ZeroMQ and czmq
  - Zyre
  - TUI (Thanks Branislav Beke)
- Demo

- 1 Demo with histogram
- 2 Demo with physics case
- 3 What can we do with EOS WFE engine?
  - 1 Can we make 1 milion directories and execute scrip for each?
  - 2 Can we scale?
- 4 Can we map user from differnet kerberos REALM?
  - 1 How can we handle UID of that user?

Thank you