EOS for Physics Data Storage @CERN 2020 edition

Cristian Contescu, on behalf of the EOS for Physics Operations team @CERN

01.03.2021

Outline

- **2020 in review**
  - Migrations, challenges & improvements
- **New hardware arrived**
  - Exploring new technologies...
  - ...and make them work as expected
- **What lies ahead**
  - LHC Run3 & more...
- **Conclusions**

2020 in review - migration to FUSEx (eosxd)

- **eosxd - actively maintained EOS FUSE implementation**
- **Advantages:**
  - Lower latency
  - More POSIXness
  - Reduced usage of resources (notably RAM)
- **Puppet managed machines are now mounting all EOS instances @CERN via the new FUSE implementation**
- **Some (non-Puppet controlled) clients still use the old FUSE implementation**
  - old FUSE (**eosd**) to be discontinued during EOS version 5 life-cycle

2020 in review - retiring the SRM gateways for EOS

- **SRM gateways at CERN based on BestMan2**
- **BestMan2 not packaged for CentOS 7 (and seemed abandoned)**
- **SLC6 end of life => Nov 2020**
- **Low usage of the SRM gateways & alternatives present**
- **Contacted experiments which were still using SRM and helped them migrate out**
  - with the help of the FTS team as well
- **SRM GWs disabled in September 2020 for:**
  - EOSLHCb
  - EOSCMS
  - EOSPUBLIC

2020 in review - HTTP and XRootD Third-Party-Copy

- **Improving bulk transfers between WLCG sites**
- **HTTP-TPC runs on all EOS for Physics instances @CERN (implemented via the XrdHttp plugin)**
- **XRootD TPC with delegated credentials also deployed and available on most of the Physics instances**
- **Scale up with the size of the EOS cluster**
  - Viable replacement(s) for gateway-ed protocols (e.g.: GridFTP)
  - Higher throughput with no need for extra-hardware
- **Still some issues to be ironed out (notably with HTTP-TPC), but getting there...**

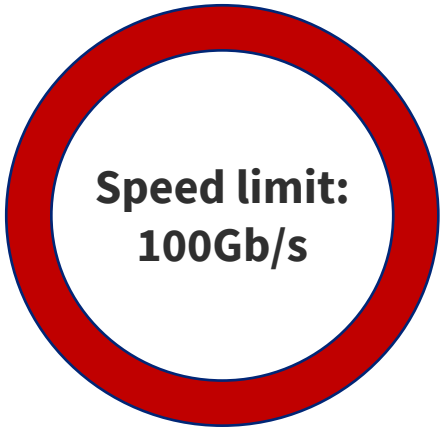
2020 in review - other achievements

- **Successful migration of the production systems to CentOS 7 || CentOS 8**
  - in time for the Scientific Linux 6 end of life
- **Service QoS improvements**
  - New / faster service probe
    - no contention due to (the increasing) number of instances - parallel probe
    - higher running frequency
    - quicker SMS alerting by ruling out false positives in a single run
  - Data durability improvements
    - EOS internal: rewritten EOS FSCK subsystem, now in production (data collection)
    - In-house external tools: finer grained problems classification
  - Reduce the high-load from specific use-cases, e.g.:
    - Introducing off-load redirectors (protecting MGMs from the high load)

New hardware challenges

- **Hello, UEFI!**
  - seamless integration, kudos to the CERN Linux support team
    - installations went smoothly, just needs a special parameter
  - how about re-installations... well, that's a different story
    - Kickstart installation changes system boot order

- **Welcome 100Gbps NICs, as well!**
  - new 100Gbps NICs present in the storage nodes for the AliceO2 pilot
  - squeezing the maximum performance out of the hardware has its ups and downs...



... more on this in my talk tomorrow

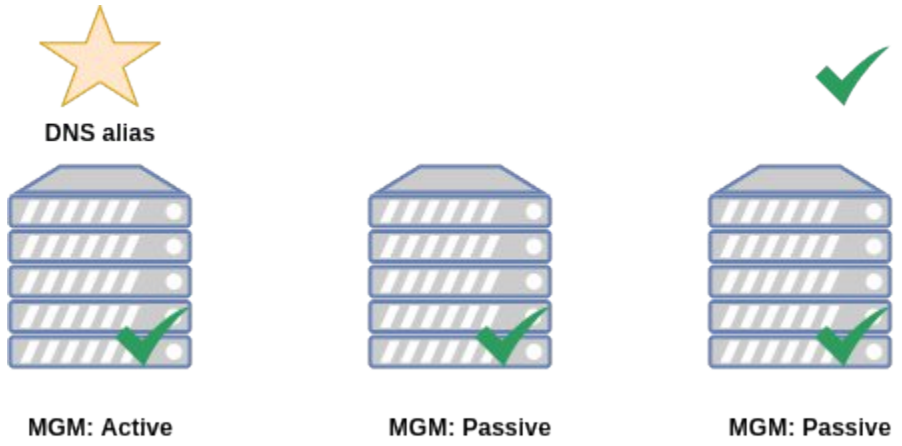
Plans for 2021 - Service Operations (I)

- **Prepare for Run3**
  - close contact with the experiments, understanding their needs
  - commissioning new hardware to sustain the requirements
- **EOSALICEO2 pilot, further developments**
  - test erasure encoded transfers (client-side erasure coding)
  - experiment various failure scenarios and their impact on the instance performance
- **EOSAMS02 take off**
  - new EOS instance for the Alpha Magnetic Spectrometer detector found on board of the ISS
  - splitting the AMS use-case from the EOSPUBLIC instance

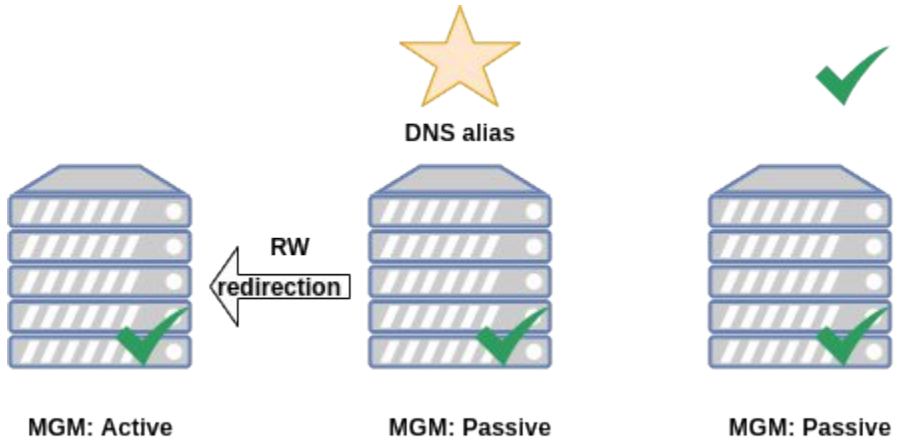
Plans for 2021 - Service Operations (II)

- **EOSPILOT instance created => ready for testing the new EOS and XRootD major release**
- **Ensure high-availability for the data taking period**
  - time to move the highly available setup of EOS to production
  - first step done: EOS configuration is now all migrated to QuarkDB for all Physics instances at CERN
  - evaluating the way forward...

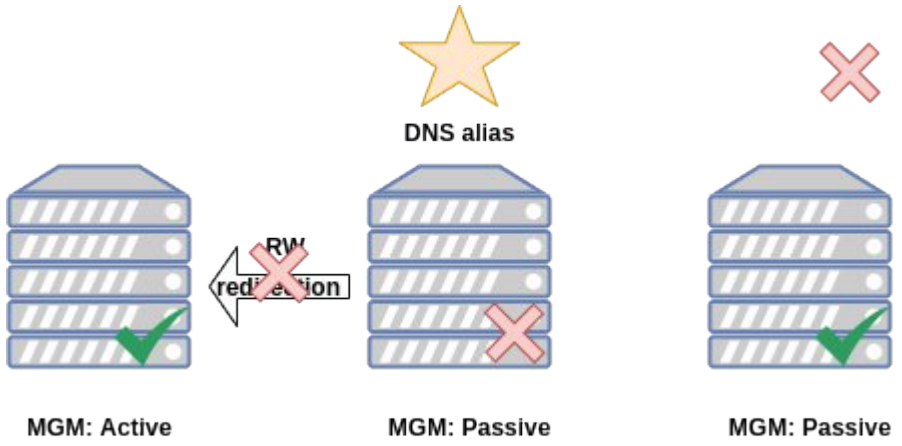
Plans for 2021 - High availability overview



Plans for 2021 - High availability overview



Plans for 2021 - High availability overview



Plans for 2021 - High availability drawbacks & solutions

- **Manual DNS entry change**
  - Someone needs to “wake up” and do it (just semi-improving availability)
- **Automatic DNS change**
  - What if the cluster is unstable?
  - What about DNS caches?
- **Round-robin alias containing all passive+active nodes**
  - Persistent clients would need to be informed about the passive → active change and auto-adjust
  - Clients should retry by using the next value from the gethostbyname()
- **HA router MGM, which will dispatch (redirect) the incoming traffic to a known MGM that is up**
  - The “router” will probe all instances' MGM status and only redirect to available ones

Plans for 2021 - Service Operations (III)

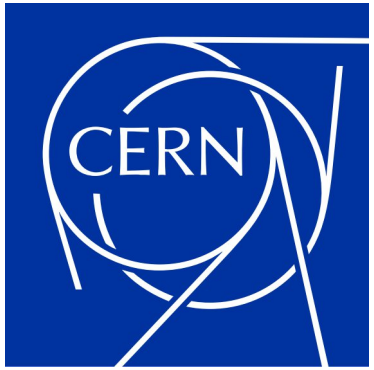
- **Last, but not least: picking up the future OS**
  - RedHat announced change of focus towards CentOS Stream and EOL for CentOS 8 in December 2021
  - CentOS 7 supported until summer 2024
  - AliceO2 storage nodes run on CentOS 8 (performance reasons)
  - Working with the IT Linux support team to see the best way forward

Conclusions

- **Despite the pandemic, 2020 was as active as any other year**
  - eosd => **eosxd** migration
  - SRM gateways decommissioning
  - **HTTP-TPC** and **XRootD TPC** with **credential delegation** in prod
  - Preparation for **LHC Run3**
  - Accommodation of new hardware and configure to **optimal performance**
  - Improvements on the **data durability** ongoing
- **2021 is already having two main challenges**
  - Started walking the path towards **highly available EOS** (especially in the view of the ALICEO2 instance)
  - Follow up the **OS changes** that are ahead of us and plan accordingly

Thank you!

Thank you!
Any questions?



home.cern