EESSI meeting

2 June 2022

https://github.com/EESSI/meetings/wiki

Agenda

1. Quick introduction by new people
2. EESSI-related meetings in last month
3. Progress update per EESSI layer (incl. bot for software layer)
4. 2021.12 version of pilot repository (+ next pilot version?)
5. AWS/Azure sponsorship update + OCRE funding opportunity
6. Google Summer-of-Code project: Gentoo Prefix on RISC-V
7. Past & upcoming events
8. Q&A

Quick introduction by new people

New people on the call: feel free to introduce yourself!

- Who are you, where do you work, on what?
- Why are you interested in the EESSI project?
- Are you planning to actively contribute,
  and if so, to which aspect(s) of the project?

EESSI-related meetings (1/2)

- Tue 5 May: CernVM-FS coordination meeting
  - Discussion arond impact of (long-running) garbage collection on publishing operations
  - Switching from JIRA to GitHub for issue tracking
  - Discussion around impact of IPS on CernVM-FS encountered by ComputeCanada
  - Corruption issue when updating files in-place to be discussed again at next meeting
  - Extensive notes available at https://hackmd.io/BPMW_kT7RH-V3lITA0DFaw

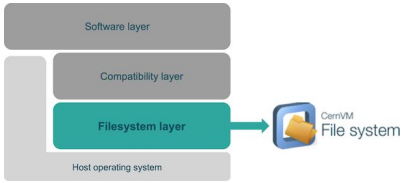
[Kenneth]

EESSI-related meetings (2/2)

- Fri 20 May: monthly sync call with Azure
  - Discussion on possibility of running hackathon-like workshop before summer
  - Working towards CUDA support (see software-layer PR #172, ready for testing!)
  - Status of data repository (cfr. filesystem-layer PR #119)
  - OCRE funding opportunity
- Several additional meetings regarding OCRE call for projects with AWS/Azure
  - More info on separate slide later on…
- Tue 10 May + Wed 1 June: Meetings on bot for building/deploying software layer
  - Fueled by recent efforts by Thomas
  - Notes available at https://hackmd.io/vtrOgU0kTeat2gX5OYKKpQ
  - More info in separate slide later on…

[Kenneth, Alan]

Progress update: filesystem layer

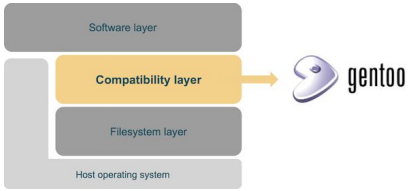


- Separate data repository (`data.eessi-hpc.org`) (PR #112 + PR #119)
  - For hosting large input datasets, etc.
  - Work-in-progress (Bob): set up on Stratum-0, not on Stratum-1 mirrors yet…

- Automatic ingestion of tarballs in S3 bucket is broken…
  - Ingestion of installations for 2021.12 `aarch64/graviton3` failed :(
  - See https://github.com/EESSI/staging/issues/69 (private repo)
  - `Repository pilot.eessi-hpc.org is in a transaction and cannot be repaired`
  - Fixed with a manual `cvmfs_server abort pilot.eessi-hpc.org` by Bob
  - There was no impact on availability of EESSI pilot repository (it seems)

[Kenneth]

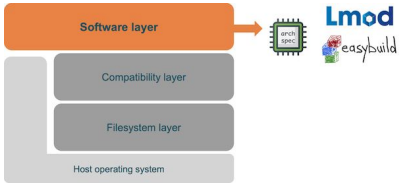
Progress update: compatibility layer



- No security updates required for 2021.06 and 2021.12 compat layers
  (as reported by Gentoo's `glsa-check` tool)
  - Very infrequent security advisories recently (cfr. https://security.gentoo.org/glsa)
  - Time to also start looking into other ways of keeping compat layer secure?
- Google Summer-of-Code (GSoC) project *"RISC-V support for Gentoo Prefix"*
  - Student wrote up a detailed project proposal, showed that he has necessary skills
  - **Project proposal was accepted \o/**
  - See separate slides by Atharva

[Kenneth]

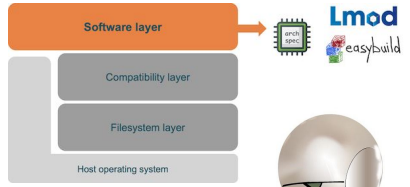
Progress update: software layer



- CUDA support work-in-progress (PR #172)
  - **Testing & feedback welcome!**
- Lmod cache update script fleshed out of software install script (PR #168)
- Fix for CI workflow to test scripts which had Singularity version hardcoded (PR #177)
- Tweak to existing scripts so they can be used by build-and-deploy bot (PR #175 + PR #176)
  - TODO (Kenneth): document purpose of different scripts we have now…
- Initial implementation of build-and-deploy bot by Thomas
  - In `eessi-bot-software-layer` repo, see PR #2 (docs) + PR #4 (bot impl.)
  - Also some changes to PyGHee base library (see PR #2 and PR #3)

[Kenneth, Alan, Thomas]

Bot for building + deploying software layer



- **Work resumed on implementing a bot to build + deploy software installations in software layer \o/**
- Tue 10 May: meeting to discuss overal idea + development plan (see slides + notes)
- Progress since then (by Thomas):
  - Step 1 (done): Set up development environment (see docs in bot PR #2 - to review)
    - Create `smee` channel, register GitHub App, install app in repo(s), run smee client on build system
  - Step 2 (done): Get basic example for PyGHee base library working
    - Fix bug in PyGHee w.r.t. getting unique ID for incoming events (PyGHee PR #2 - merged)
  - Step 3 (WIP, almost done): Rework existing bot on top of PyGHee base library (bot PR #4 - reviewed)
    - **Status: bot starts building software with EasyBuild using easystack file when PR is opened**
  - Already several ideas for improvements to PyGHee + bot (like PyGHee PR #3 - reviewed)
- Wed 1 June: sync meeting (Thomas + Kenneth) to discuss status, PRs, next steps (see notes)

[Thomas, Kenneth]

EESSI pilot repository

NOT FOR PRODUCTION USE!

https://eessi.github.io/docs/pilot

- 2021.06: considered “final” (no further changes, except security updates in compat layer if needed)
- Current status for 2021.12 **[default]**
  - Compatibility layer: in place for `aarch64` / `ppc64le` / `x86_64`
  - Software layer:
    - Software installations included in 2021.06 also in place for 2021.12, incl. GROMACS, OpenFOAM, TensorFlow + Horovod, R + Bioconductor, QuantumESPRESSO
    - Additional software (vs 2021.06): SciPy-bundle with foss/2021a (excl. `ppc64le`), WRF
    - Targets: `aarch64/generic`, `aarch64/graviton2`, `aarch64/graviton3`, `ppc64le/generic` (partial!), `ppc64le/power9le` (partial!), `x86_64/generic`, `x86_64/amd/zen2`, `x86_64/amd/zen3` (Milan), `x86_64/intel/haswell`, `x86_64/intel/skylake_avx512`
  - TODO:
    - Ensure that Lmod cache update is done correctly, includes *all* available modules (first step: PR #168)
    - Bot to automate workflow of adding software to EESSI (to avoid losing time doing it manually)

[Kenneth]

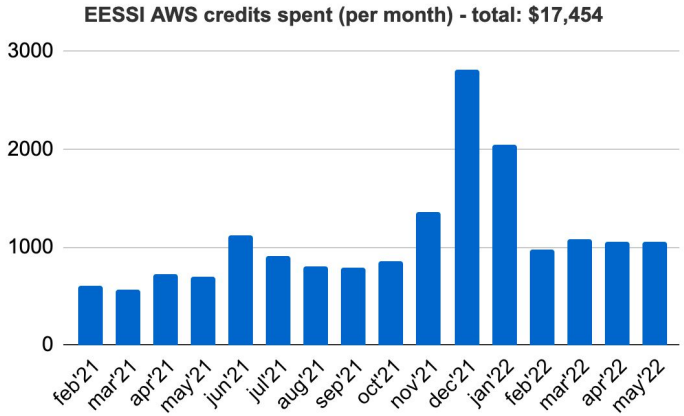
Time for the next pilot version?

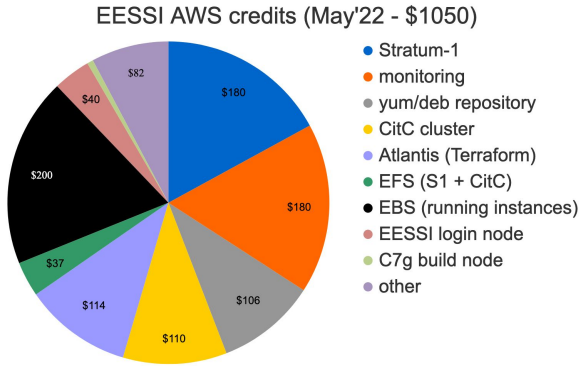
- Small changes to compatibility layer: updated Lmod, …
- Include enhancements/changes that are necessary for CUDA GPU support
- Work towards getting rid of ugly install script, aim for easystack-only
- **Only add software installations via bot, no more manual deployments!**
- Initially include same software installations in software layer, then **gradually expand**
- Also install software with more recent toolchains + more applications
- Stop wasting time with supporting POWER (`ppc64le`), for now?

[Kenneth]

Usage of sponsored AWS credits

- **Ask in #aws-resources Slack channel to get access!**
- Original batch of $25,000 worth of sponsored credits expired on Jan 31, 2022!
- Request for new credits is WIP, extra $10k worth of credits already received to bridge the gap
- **~$5,850 worth of sponsored credits left**
- In May ‘22: ~$1,050 worth of credits spent on Stratum-1 server, monitoring node, CitC cluster, …
- ~$17,454 worth of credits spent in total so far (since Feb’21), all covered by sponsored credits

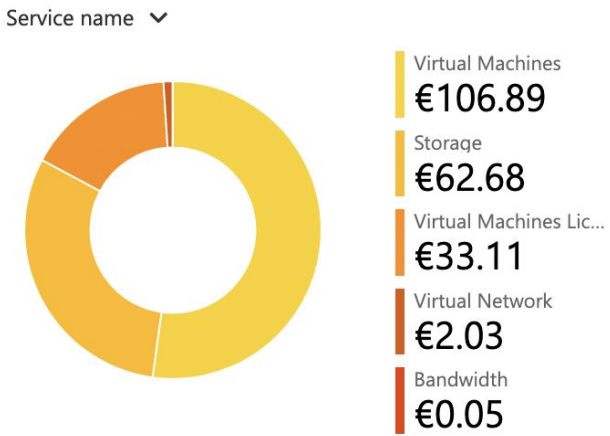


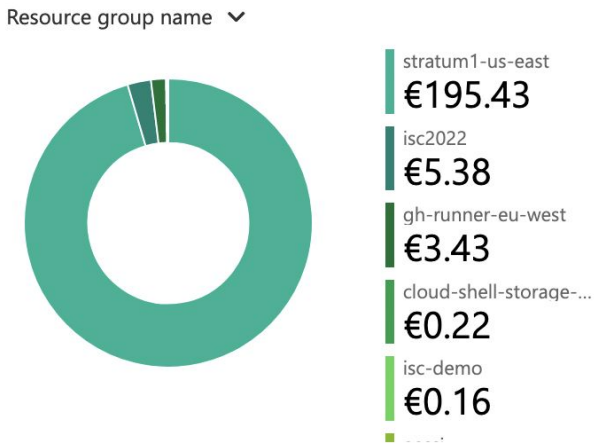


[Kenneth]

Azure sponsorship

- Sponsored credits (€40,000) are being put to good use!
- **Ask in #azure-resources Slack channel to get access!**
- In May'22: ~204 worth of credits spent, pretty much all on Stratum-1
- ~€3,255 worth of credits spent in total (since Sept'21)





[Kenneth]

OCRE Funding call

- https://www.ocre-project.eu
- Considering project with Azure focusing on
  - Resources for CI and the EESSI bot (including GPU software)
  - Training
  - ReFrame testing
- Resource estimate
  - Current spend is ~1000 EUR per month
  - Expect to grow this to 10k/month over the 3 year period, estimating 240K
  - Allowed 25% for “professional services” (which we can use for consultancy work)
  - Total current estimate: 300K

[Alan, Bob, Kenneth]

GSoC project Gentoo Prefix on RISC-V

- Gentoo Prefix is a key component for the EESSI project (compatibility layer)
- RISC-V is one of the target CPU architectures in the EESSI project
- **Goals of GSoC project:**
  - New profile for Gentoo Prefix on RISC-V
  - Make it possible to bootstrap and use a Gentoo Prefix system on RISC-V
  - Test and keyword packages in Gentoo for RISC-V
  - Documentation to port Prefix on new architecture

[Atharva, Kenneth, Guilherme]

GSoC project Gentoo Prefix on RISC-V

Current status:

- Added RISC-V Profile and symlink in `bootstrap-prefix.sh` script
- This allowed stage 1 to proceed
- `scanelf` is expected to be available in host system,

  and hence ncurses failed to compile; adding `pax-utils` in script will solve this.
- Stage 2 of bootstrap goes ahead till compiling GCC
- GCC failed to build as `gcc-multilib` is required to install 32 bit libraries;

  finding a workaround for this, fixing this issue should complete stage 2
- As we proceed further everything is documented in prefix_on_riscv repository
- After this we will can start with stage-3 and fix bugs there…

[Atharva, Kenneth, Guilherme]

GSoC project Gentoo Prefix on RISC-V

References:

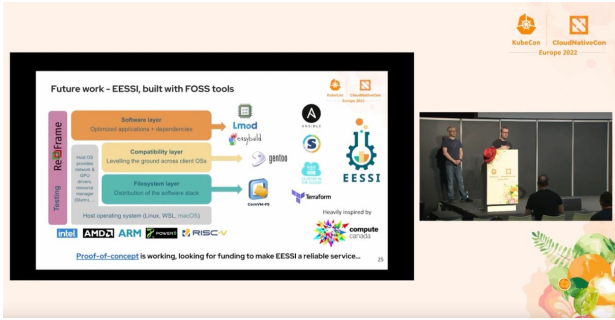
- Description of the GSoC Project on Gentoo wiki
- GitHub repositories I am working on:
  - Documentation of the issues encountered during the process
  - Prefix Profile for RISC-V
  - Fork of Gentoo Prefix
- Gentoo bug on adding ppc64le support to Prefix
- Gentoo Prefix as Physics Software Manager
- RISC-V QEMU image and Chroot environment

[Atharva, Kenneth, Guilherme]

Past events: KubeCon Europe 2022

KubeCon Europe 2022 (Valencia, 16-20 May 2022)

- *“Unlimited Data Science Libraries, One Container Image, No Installation!”*
- Wed 18 May 2022 at 11:55 CEST (“Machine Learning + Data” track)
- Joint talk by Kenneth Hoste (HPC-UGent) & ~~Guillaume Moutier~~ Marcel Hild (Red Hat)
- This work was not using EESSI yet, but planned for “future work”...
- **EESSI was pitched towards the end of the talk, sparked quite a bit of interest!**
- Link to talk abstract + slides
- Recorded talk available on YouTube (EESSI part at 30:40)



[Kenneth]

Past events: ISC'22

ISC 2022 (Hamburg, 29 May - 2 June 2022)

- Half-day EasyBuild tutorial on Sun 29 May (ISC'22 link)
  - Tutorial materials available at https://easybuild.io/tutorial/isc22
- Quick talk on EESSI by Bob at pre-ISC Dell event (cfr. https://dellhpc.org/events/23148)
- Two EESSI presentations by Bob at Microsoft booth at ISC'22 (see booth schedule)

[Kenneth]

Upcoming events: BioHackathon Europe 2022

- Near Paris (7-11 Nov'22) - https://biohackathon-europe.org

- Our project proposal is accepted:
  "Make your own or favourite software available on your cluster with EasyBuild/EESSI"

  https://github.com/elixir-europe/biohackathon-projects-2022/tree/main/16

  Other projects: https://biohackathon-europe.org/projects.html

- Biology-oriented hackathon **but potentially** a good place to:
  - reach new users and developers
  - meet Galaxy users and community

[Sebastien]