Debian GNU/Linux for Scientific Research

Andreas Tille

Debian

Online, 20. June 2024

What is Debian Science

- *Debian Pure Blend*
- ... rather an umbrella to grow specific Blends
- Virtual place where scientists in Debian can meet
- Not ¨competing¨ with other scientific Blends
- Maintenance of common scientific tools
- Entry point for scientists who want to contribute
- Other interesting teams:
  - Debian HPC: *Wiki*, *Mailing list*
  - Debian Math: *tasks page*, *Mailing list*
  - *Debian Med*
  - *Debian PAN*

Organisation

- *Common mailing list*
- *Common Salsa Repository*
- *Blends Web sentinel listing tasks*
- *IRC #debian-science on irc.oftc.net*
- *Debian Wiki*

Mentoring

- *Mentoring of the Month (MoM)*
- *Sponsoring of Blends*
- Packaging workshop at Max Planck Institute Kognitions- und Neurowissenschaften Leipzig (October 2006)
- *Debian for Scientific Facilities Days* at ESRF Grenoble (June 2012)
- Packaging workshop at Max Planck Digital Library Munich (July 2014)
- Several live packaging workshops
- ➜ I was payed to hold packaging workshops (+ doing packaging work)

Who is using Debian

- *List of Debian users contains lots of scientific institutions*
- Just picking a view examples (that are not (yet) on this list)

Max Planck Institute for Gravitational Physics, Hannover

- Running large HTC cluster
- World's largest research institute specializing in general relativity
- 1,000 compute nodes (41,000 cores); 300 GPUs; >95% busy 24/7
- HTCondor for day job scheduling
- Bare metal to minimal OS: FAI
- Why using Debian
  - Many packages available without third party repos
  - Existing in-house knowledge for package building, tooling
  - Problem: Commercial software usually does not (officially) support Debian
    (support file creation, firmware updates)

IGWN Debian Repositories

- *IGWN Debian Repositories*
- Checked their *bookworm repository* and found
  `apptainer` where some packaging effort exists *inside Debian*
  `bzip3` which is in Debian in the same version
  `ca-certificates-java` which is in Debian in a later version
  `igwn-cmake-macros`, `igwn-htcondor-config` local purpose
  `lal` example for software developed locally
- ➜ Better talking to Debian first

Linux at CERN [1]

- Accelerator front-ends
  Linux Team provides Limited Debian Support
- Discussing risk mitigation by adding Debian
  RHEL contract ends in 2029 + AlmaLinux and RHEL are entangled

[1] *Preparing a Multi-Ecosystem Linux strategy at CERN*

Wellcome Sanger Institute

- World-leading genomics research institue
- Has (had?) employed up to three Debian Developers
- Started with Debian, moved to CentOS because many other Academic HPC centres were using it
- Now running Ubuntu + OpenStack and Ceph [2]

[2] *Posting on Debian Med mailing list*

EDF

- French electricity producer and provider
- Most supercomputers in use at EDF were running Debian until 2020
- Not possible anymore to buy a Debian based top 500 supercomputer
  (none of the top sellers support Debian as a prerequisite)[3]
- Workstations of people in R&D and nuclear engineering departments are equipped with (very close) derivative of Debian
- Calculation codes are packaged in Debian: `openturns`, `stopt`, `code-saturne`, `syrthes` and others
- Packages above maintained in Debian Science or Debian Math
- In case RHEL has to be used those packages are partly installed as Debian based singularity containers

[3] *Posting on Debian Science mailing list*

ELIXIR

- European life sciences infrastructure
- Strong cooperation with Debian Med
- Makes not only use of packages but also metadata for classification

Institute of Neurosciences and Medicine (INM)

- Cluster at *Institute of Neurosciences and Medicine* is running Debian
- *Publication*: `FAIRly big: A framework for computationally reproducible processing of large-scale data`
  Co-author Michael Hanke is Debian Developer

- ➜ General hint for publications: IMHO each publication should be accompanied by some container that can do the data processing reproducible even years later (hopefully)

Do it yourself

- Do not take Debian as a finished product but something you can influence
- Turn Debian into something that fits your needs
- Debian developers are happy to support you

Joining forces with upstream

- FIS GT.M (MUMPS database)
- HTCondor: sponsoring of packages by Tim Theisen (upstream)
  see *Web archive of Debian HPC mailing list*

Reproducibility

- *Reproducible builds*
- *Snapshots of every released Debian package*
- ➜ Establish reproducible containers
  some guarantee your container still builds in future without divergence

This talk is available at
https://people.debian.org/~tille/talks/
Andreas Tille <tille@debian.org>