

JupyterHub for research facilities

Wednesday, 22 May 2019 11:00 (1h 30m)

Jupyter notebooks combine the accessibility of an interactive web-frontend, the reproducibility of a laboratory notebook, and the collaborative potential of a cloud-based deployment. The accessibility and interactivity lowers the barrier for researchers to prototype, write, and share data analysis pipelines, and the literate programming approach of Jupyter makes it particularly simple to reproduce, reuse, and adjust notebooks by colleagues and peers.

Jupyter has another use: providing access to remote resources via JupyterHub. Many typical JupyterHub deployments have used cloud-based resources for one-off purposes, but there is also good support for JupyterHub as an interface to HPC clusters and other pre-existing research facilities. JupyterHub can provide a stepping stone for light computing on existing clusters - as well as a more user friendly interface for preparation and visualization for existing power users.

In this workshop, we will demonstrate the use of JupyterHub and provide guidance so that attendees can set up their own JupyterHub deployments. There will be a show-and-tell of Jupyter itself and existing JupyterHub deployments. We will go over the basic requirements and practical implementation for a JupyterHub setup. The workshop includes discussion about the difference between traditional batch and interactive workloads, and how the parameters of HPC systems can be tuned to interactive uses. At the conclusion of the workshop, participants will be well prepared to begin deployment of JupyterHub to their own facilities and a Nordic JupyterHub community will begin.

Pre-workshop

Prerequisites: since we do not cover Jupyter itself, we will share links to talks/lessons on basic Jupyter notebooks in an updated abstract so participants can learn and experiment in advance.

Workshop outline

1. **Intro and tour of the Jupyter ecosystem** (talk, 10 min, speaker TBA): What is Jupyter and why is it taking over the world? Most of us have heard about Jupyter, but what is JupyterHub, JupyterLab, repo2docker, mybinder, etc?
2. **Tour of some cloud JupyterHub deployments** (talk + discussion, 10 min, speaker TBA)
3. **Tour of Aalto HPC JupyterHub** (focus on integration to existing infrastructure) (talk + discussion, 10 min, Richard Darst): A brief description of one existing system to frame our final perspective.
4. **JH internals from a sysadmin's point of view** (talk + discussion, 20 min, speaker TBA): JH makes a lot of sense once you know the details. This talk goes into those details.
5. **JH with batchspawner** (talk + discussion, 10 min, Richard Darst): "batchspawner" is a JH spawner for batch systems. It allows harnessing existing batch systems to run the notebooks. It is the core connector of JupyterHub and existing batch systems.
6. **JH with Kubernetes** (talk + discussion, 10 min, speaker TBA): Kubernetes is a container orchestration system, the alternative cloud-like viewpoint for large-scale deployments.
7. **Tuning HPC for interactive uses** (talk + panel, 20 min): Making JH work is relatively easy. But the typical workload of HPC clusters is quite different than that required by the interactive use of Jupyter. How should batch systems be configured so that it can co-exist with Jupyter? This is the hard part.

References

- "I don't like notebooks" presentation by Joel Grus
- SWAN: Jupyter instance at CERN: <https://conferences.oreilly.com/jupyter/jup-ny/public/schedule/detail/68359>

Primary authors: WIKFELDT, Thor (KTH/NeIC); DARST, Richard (Aalto University); BAST, Radovan (UiT/NeIC); RAZICK, Sabry (UiO/NeIC)

Presenter: DARST, Richard (Aalto University)

Session Classification: Workshops I