JupyterHub for research facilities

Wednesday 15 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 go into depth about Jupyter notebooks themselves, we will share links to talks/lessons on basic Jupyter notebooks in an updated abstract so participants can learn and experiment in advance.

- Jupyter: wikipedia
- Jupyter notebooks: video
- JupyterHub: brief description, video

Workshop outline

1. **Introduction to workshop (5 min)**: *Radovan Bast*
2. **Brief introduction to Jupyter (5 min)**: What is Jupyter and why is it cool? *Thor Wikfeldt*
3. **JupyterHub ≠ x, ∀ x (10 min)**: What is JupyterHub and why? *Richard Darst*
4. **JupyterHub from a sysadmin point of view (20 min)**: What does a sysadmin need to know to quickly set up a JupyterHub deployment? What are the challenges and solutions in operating JupyterHub and integrating Jupyter and HPC resources? *Richard Darst*
5. **Example deployments (20 min)** *Csaba Anderlik, Gergely Sipos, Richard Darst*
6. **Panel Discussion (30 min)**: Q&A - where do we go from here?

Size of poster

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

Presenters: WIKFELDT, Thor (KTH/NeIC);   DARST, Richard (Aalto University);   ANDERLIK, Csaba (University of Bergen);   SIPOS, Gergely (EGI Foundation)

Session Classification: Workshops I

Track Classification: Workshops