

Introduction to JupyterHub @ bwUniCluster

Samuel Braun, SCC, KIT



Funding:



Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST



Bundesministerium
für Bildung
und Forschung

www.bwhpc.de / www.nhr.kit.edu

Outline

- Motivation
- Project Jupyter
- JupyterHub @ bwUniCluster
 - Access
 - Resource selection – default/advanced
 - Jupyter software stacks
 - JupyterHub – Step-by-Step
- Miscellaneous Topics
 - Add software
 - Jupyter w/o JupyterHub
- Outlook
- Questions

Reference: Jupyter @ KIT

Most information can be found at

bwHPC Wiki:

https://wiki.bwhpc.de/e/Jupyter_at_SCC

NHR@KIT Wiki:

<https://www.nhr.kit.edu/userdocs/jupyter>

The screenshot shows a wiki page titled "Jupyter at SCC". The page content discusses using Jupyter as an alternative to SSH for accessing HPC resources. It includes a table of contents with sections like "Short description of Jupyter", "Access requirements", "Login process", "Selection of the compute resources", "JupyterLab", "Log out", "Selection of software", and "Installation of further software". Below the table of contents, there is a section titled "1 Short description of Jupyter" which provides a brief overview of what Jupyter is and how it works.

The screenshot shows the "Overview" page of the NHR@KIT User Documentation. The page title is "Overview" and it includes a table of contents for "Jupyter". The main content area describes Jupyter as an interactive supercomputing tool accessible via SSH. It mentions that Jupyter runs in a browser and allows users to execute commands on HPC systems, including GPUs. The page also lists URLs for different hardware resources: HoreKa, bwUniCluster 2.0, and HAICORE.

Motivation

Why Jupyter?

HPC – Classical

- SSH
- **High Entry Hurdles**
 - Choice of resources
 - Linux
 - Tools for connection and data transfer
- Remote-Visualization
 - VNC? X11?
- State of the art for **advanced requirements!**

Why Jupyter?

HPC – Jupyter

- Web browser
 - No data transfer for analysis
 - No additional software

■ Low Entry Hurdles

- Intermediate performance requirements
- Interactive visualization of data
- Prototyping

Project Jupyter

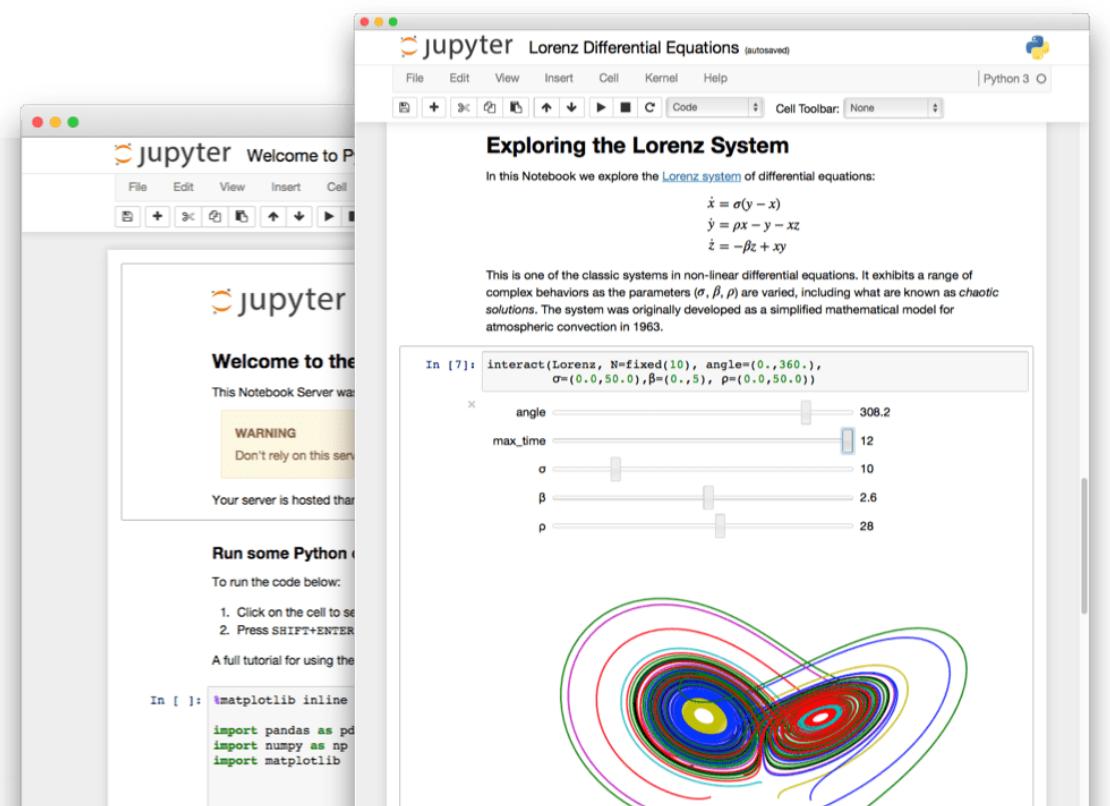
Project Jupyter

- Spin-off from project IPython
- Jupyter: Core languages
 - Julia
 - Python notebook
 - R
- Language agnostic
- Jupyter kernels
 - IPython
 - IJulia
 - IRKernel
 - >100 other kernels



Jupyter Notebook

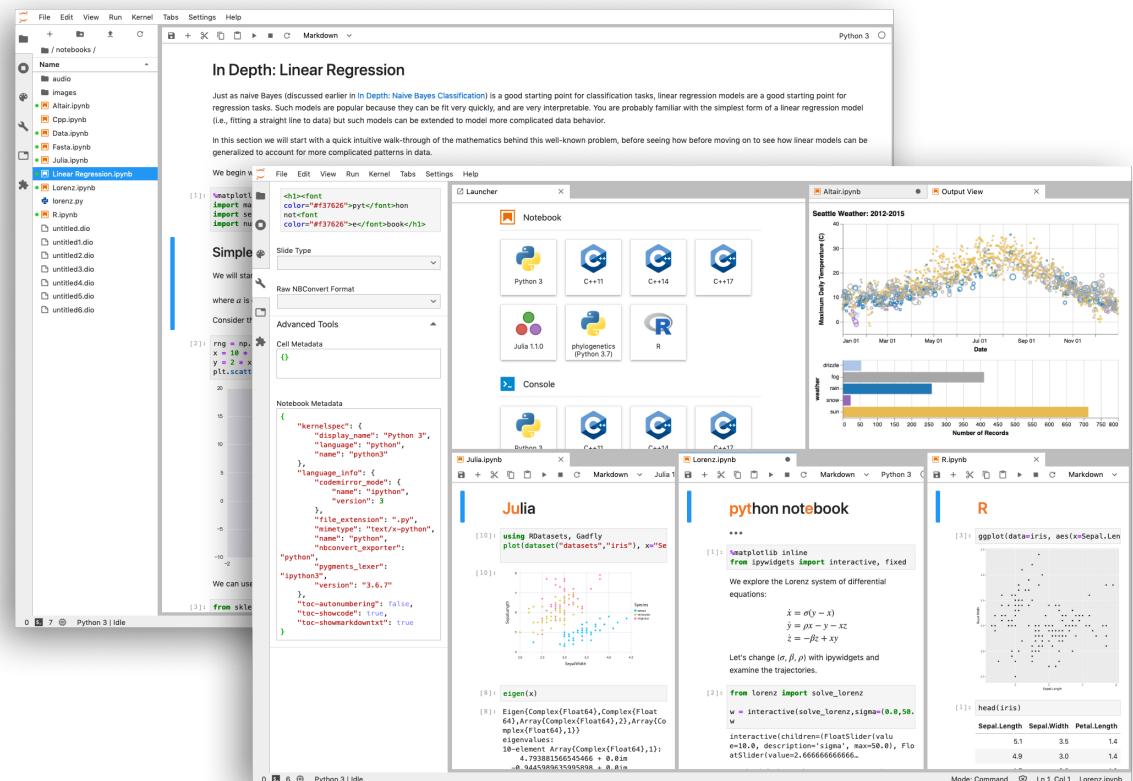
- Open-source web application
- Create and share documents
 - Live code
 - Equations
 - Visualizations
 - Narrative text
 - HTML5 is the limit...
- Execute code in browser
 - ... on bwUniCluster
- .ipynb file
 - JSON document



<https://jupyter.org/>

JupyterLab

- User interface for Project Jupyter
- Arrange documents/activities in tabs/blocks
 - Notebook
 - Terminal
 - Text editor
 - File browser
 - Rich outputs
 - ...



<https://jupyter.org/>

JupyterHub

- Multi-user server for Jupyter Notebooks
- User management and authentication
- Spawning and proxying
- HPC context
 - Choice of resources
 - Slurm integration



JupyterHub @ bwUniCluster

Registration

- **Registration @ bwUniCluster 2.0**
 - bwUniCluster entitlement
 - Web registration + questionnaire
- **Set service password**
 - bwIDM → bwUniCluster 2.0 → Set service password
- Register a software or hardware token (alias **2FA**)
 - KIT users: <https://my.scc.kit.edu/token>
 - “The Länd” users: bwIDM → My Tokens

Accessing bwUniCluster

- Only within **network**
 - ... of **KIT**
 - ... of your **home institution**
- ... otherwise establish **VPN** connection
- **SSH**
 - ssh <userid>@uc2.scc.kit.edu
 - TOTP prompt (first)
 - Service password (second)
- **Jupyter**
 - (modern) Web browser: <https://uc2-jupyter.scc.kit.edu>
 - **ONE successful login via SSH required**
 - ... otherwise there is no \$HOME
 - ... spawning will fail (timeout)

Selection of Resources – Normal

Select your resources

The grayed out fields contain a reasonable preselection of resources.
Other values can be selected in advanced mode.

Number of CPU-cores:

Number of GPUs:

Runtime:

Partition:

Amount of memory:

JupyterLab-Basemodule:

Advanced Mode:

Spawn

“Normal” mode

- Number of CPU cores **OR** GPUs
- Runtime
- Jupyter Basemodule
- Grayed out fields: **Sane pre-selection** of resources

Auto-Reservation for GPUs

- Mo – Fr, 8 AM – 8 PM
- Availability indicator

Number of GPUs:
Good availability
Number of GPUs:
Limited availability
Number of GPUs:
Currently not available

Spawn

- Starts JupyterLab in interactive Slurm session
- Connects/proxies to that session

Selection of Resources – Advanced

Select your resources

The grayed out fields contain a reasonable preselection of resources.
Other values can be selected in advanced mode.

Number of CPU-cores:

Number of GPUs:

Runtime:

Partition:

Amount of memory:

JupyterLab-Basemodule:

Advanced Mode:

Reservation:

Account:

Mount LSDF:

Use BEEOND:

Spawn

„Advanced“ mode

- Free choice of resources
- No grayed out fields
- No auto reservation

Reservation

Account

LSDF

BEEOND

Jupyter Software Stacks

- Lmod modules
 - jupyter/base
 - jupyter/tensorflow
- JupyterLab lives inside venv
 - `--system-site-packages` enabled/visible
 - Possible **interference** with `pip --user` installs (!)
- Access via
 - Drop-down menu in JupyterHub: “**JupyterLab-Basemodule**”
 - `module load jupyter/base` or `jupyter/tensorflow`

Step-by-Step Jupyter @ bwUniCluster (1/4)

- Go to <https://uc2-jupyter.scc.kit.edu> and click on „Login“
 - ...or go directly to <https://uc2-jupyter.scc.kit.edu/hub/login>
- Choose your **home organization** and continue

Jupyter@UC2

Jupyter@UC2

Overview

Jupyter Software Stacks

Login

The login to JupyterHub consists of the following steps, you will be redirected accordingly:

- Choose your home organization
- Enter username and password
- Enter the one-time password (second factor)

Click on "Enter JupyterHub"

Login ➔

Documentation

- [bwHPC Wiki](#): Information about the use of bwUniCluster 2.0+GFB-HPC.
- [Jupyter Wiki](#): General information about Jupyter and its use on the bwUniCluster.
- [Service Description](#): Service description on the KIT:SCC website.

Software

Software stacks

The software for data analysis and machine learning is developing rapidly. We therefore provide an up-to-date [overview](#) of the currently preinstalled software packages.

Installing own software

KIT - Föderierte Dienste

[KIT](#)

Impressum | Datenschutz | KIT | Deutsch

Föderierte Dienste am KIT

Willkommen

Sie wurden von einem Dienst hierher weitergeleitet, um sich zu authentifizieren:
Unicloud Jupyter

Heimatorganisation merken:

Föderation:

Suchfilter:

Heimatorganisation:

- HTWG Konstanz
- Hochschule Mannheim
- Hochschule Offenburg
- Hochschule Pforzheim IdP
- Hochschule Ravensburg-Weingarten
- Reutlingen University / Hochschule Reutlingen
- Hochschule der Medien Stuttgart
- Hochschule für Forstwirtschaft Rottenburg (HFR)
- Hochschule für Musik Freiburg
- Hochschule für Technik Stuttgart
- HWU Nürtingen-Geislingen
- Hochschule für öffentliche Verwaltung und Finanzen Ludwigsburg
- Karlsruher Institut für Technologie (KIT)

FORTFAHREN

Step-by-Step Jupyter @ bwUniCluster (2/4)

- Login to your home organization
 - Username + password
 - 2FA

Web Anmeldedienst <https://idp.scc.kit.edu/idp/profile/SAML2/Redirect/SSO?execution=e1s1>

KIT
Karlsruhe Institute of Technology

Shibboleth Identity Provider

Anmelden

Sie wurden von den **Föderierte Dienste am KIT** hier weitergeleitet und befinden sich nun auf einem Server des KIT. Bitte melden Sie sich mit Ihrem KIT-Account (z.B. ab1234 als Mitarbeiter oder xxxx als Student) und Ihrem Passwort an.

Benutzername:
ab1234

Passwort:

Wenn Sie Probleme in den KIT-Diensten angezeigt haben, können Sie sich mit Ihren lokalen Kontakt personen unterreden. Klicken Sie auch mit Ihrem lokalen Kontakt personen.

Windows Login VERWENDEN

Die oben bezeichnete Webseite des Serviceanbieters bittet Sie, sich bei Ihrer Heimteinrichtung anzumelden.

Sie bekommen auf der Folgeseite die Daten angezeigt, um deren Übermittlung der Serviceprovider bitten. Sie können dies bestätigen und damit den Vorgang fortsetzen oder durch Schließen des Fensters abbrechen. Haben Sie denselben Service bereits einmal genutzt, werden Sie nur dann erneut nach einer Bestätigung gefragt, wenn sich der Datenumfang oder der Name des Serviceanbieters geändert hat. Wenn Sie auf jeden Fall nochmal sehen möchten, welche Daten zur Übermittlung vorgesehen sind, aktivieren Sie bitte nachstehende Option.

Bitte zeige mir für diesen Serviceprovider erneut an, welche Daten gesendet werden sollen.

ANMELDEN

Wenden Sie sich bei technischen Problemen auf dieser Seite an den [ServiceDesk](#) des SCCs. Bitte teilen Sie dem Servicedesk den Serviceprovider und die Nutzerkennung mit, die Sie zur Anmeldung verwenden wollten. Weitere Informationen über Shibboleth finden Sie auf der [Shibboleth Dienstseite](#) des SCCs.

KIT - Föderierte Dienste <https://bwidm.scc.kit.edu/user/twofa-login.xhtml>

KIT
Karlsruhe Institute of Technology

Föderierte Dienste am KIT

Login mit zweitem Faktor

Ihre Token werden an dieser Stelle nur angezeigt. Um Änderungen vorzunehmen besuchen Sie bitte die folgende Webseite:
<https://my.scc.kit.edu/token>

Um die angeforderte Aktion durchzuführen, muss ein zweiter Faktor eingegeben werden. Bitte geben Sie einen beliebigen zweiten Faktor aus der unten stehenden Liste ein um fortzufahren.

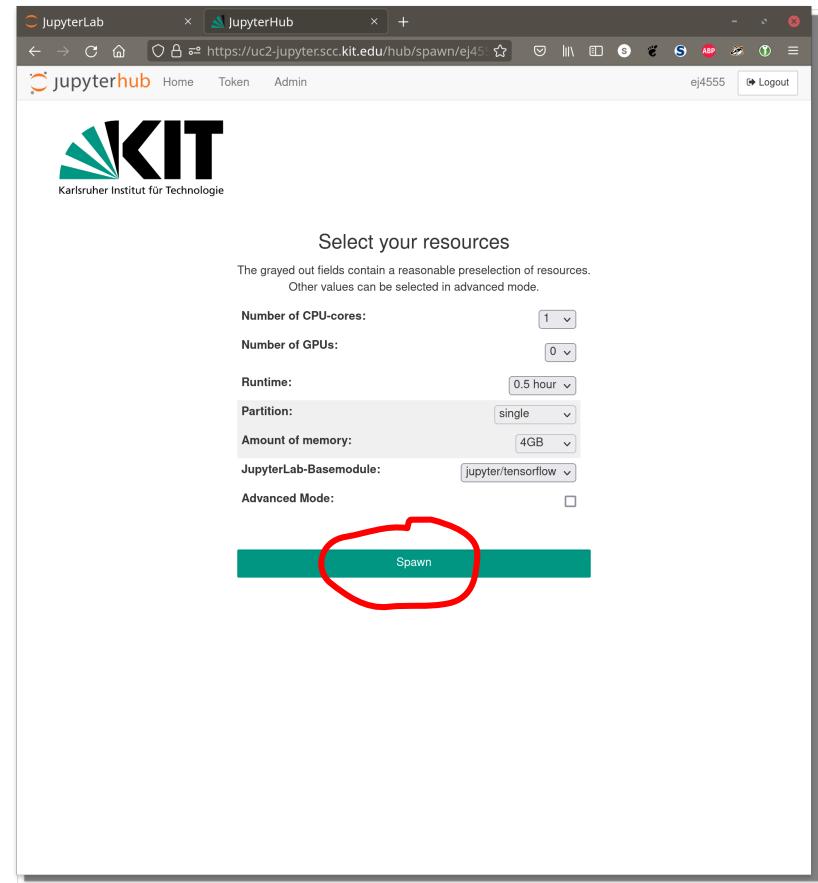
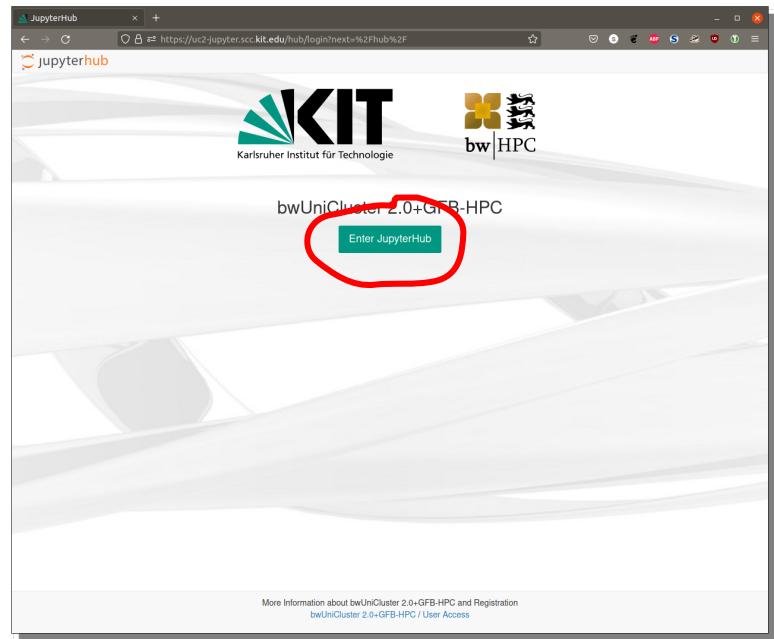
Aktueller code: **dhnhhhshsggggtnk**

FINDEN

OATH0001B4ED Backup TAN Liste	TOTP0001E310 Smartphone App	UBCM016880C4 Yubikey
UBCM0169FE9D Yubikey	2608719411986 Hardware TOTP	

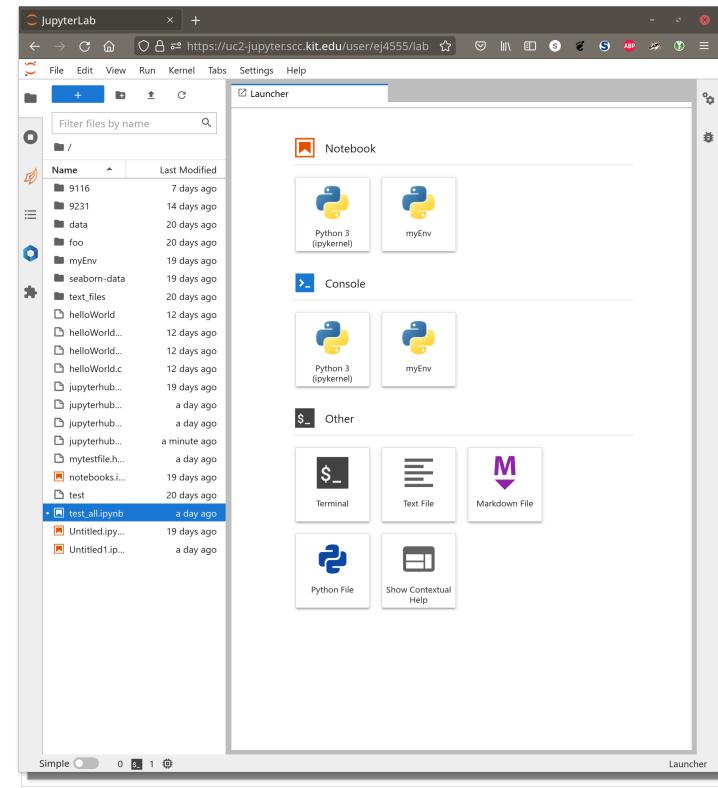
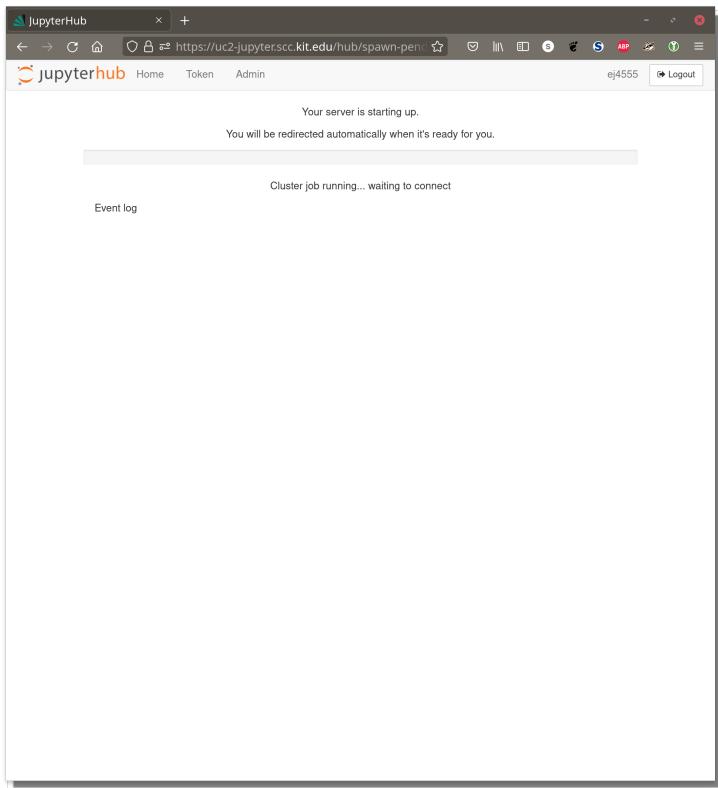
Step-by-Step Jupyter @ bwUniCluster (3/4)

- Click “Enter JupyterHub”
- Select resources and click “Spawn”



Step-by-Step Jupyter @ bwUniCluster (4/4)

- Spawning may take a while
 - ... timeout after 10 minutes
- JupyterLab runs on compute node on bwUniCluster



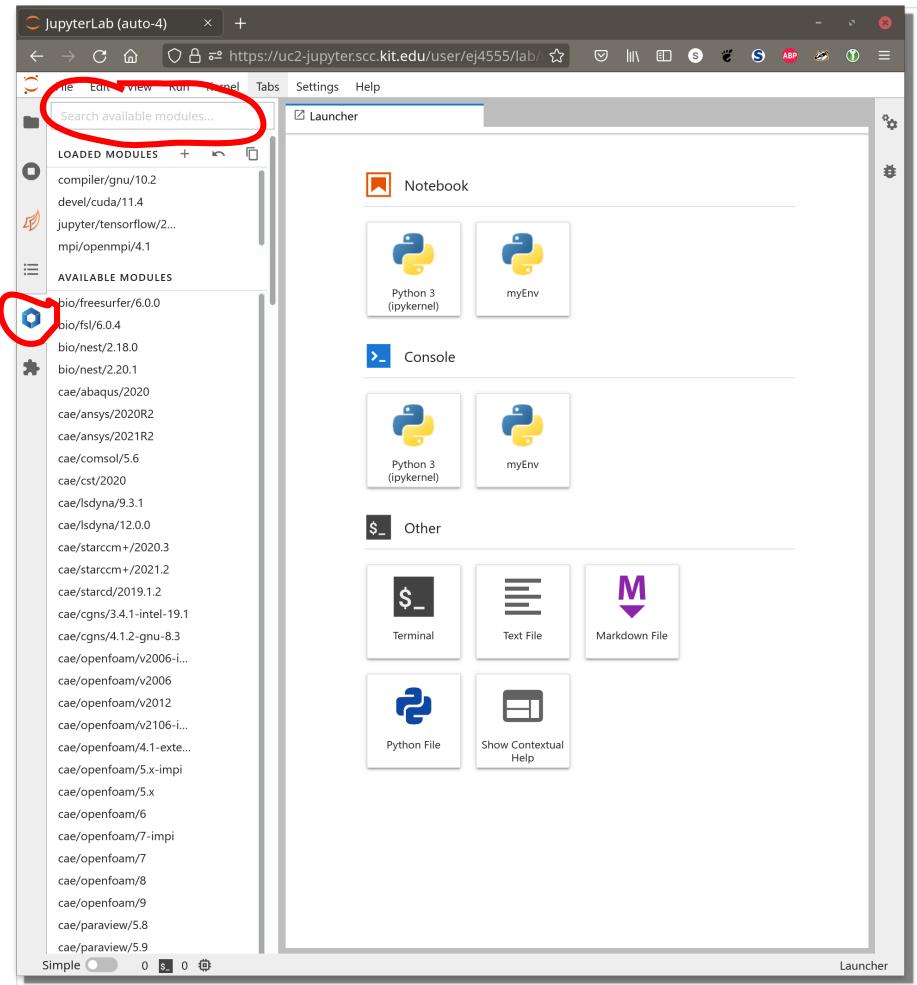
Miscellaneous Topics

Add Software

■ Python: Use virtual environments

```
> python3.8 -m venv myEnv  
> source myEnv/bin/activate  
> pip install <myPackage>
```

- Load Lmod modules
 - blue button
 - search field
- Kernel restart required



Jupyter w/o JupyterHub

- You **don't need** JupyterHub ...
- Login to UC2, start interactive job

```
> ssh <userID>@uc2.scc.kit.edu  
> salloc -p gpu_8 --gres=gpu:1 --time=30:00
```

- Wait till job is running, remember compute node ID (nodeID) and install and/or start Jupyter Notebook or JupyterLab ...

```
> module load jupyter/base  
> jupyter notebook --no-browser --port=8888 --ip 0.0.0.0
```

- From your local terminal: Establish SSH tunnel to compute node

```
> ssh -L 8888:<nodeID>:8888 uc2.scc.kit.edu
```

- Open <http://localhost:8888> in your web browser

Outlook

WIP: BYO Jupyter Container

- Connect **containerized** Jupyter with JupyterHub @ bwUniCluster
- **Docker images** from any registry
 - For complicated/intrusive software stacks
 - Optimized software stacks
 - Intel, e.g.
intel/intel-optimized-tensorflow
 - Nvidia, e.g.
nvcr.io#nvidia/tensorflow:21.10-tf2-py3
 - AMD, e.g.
rocm/tensorflow:rocm4.3.1-tf2.6-dev
- Possible **root access** (sic!)
 - **Yes, you can**

```
sudo apt-get install <myPackage>
```

Select your resources

The grayed out fields contain a reasonable preselection of resources.
Other values can be selected in advanced mode.

Number of CPU-cores: 1

Number of GPUs: 0

Runtime: 0.5 hour

Partition: single

Amount of memory: 4GB

JupyterLab-Basemode: Container Mode

Advanced Mode:

Container Mode:

--container-image **jupyter/base-notebook** (highlighted with red circles)

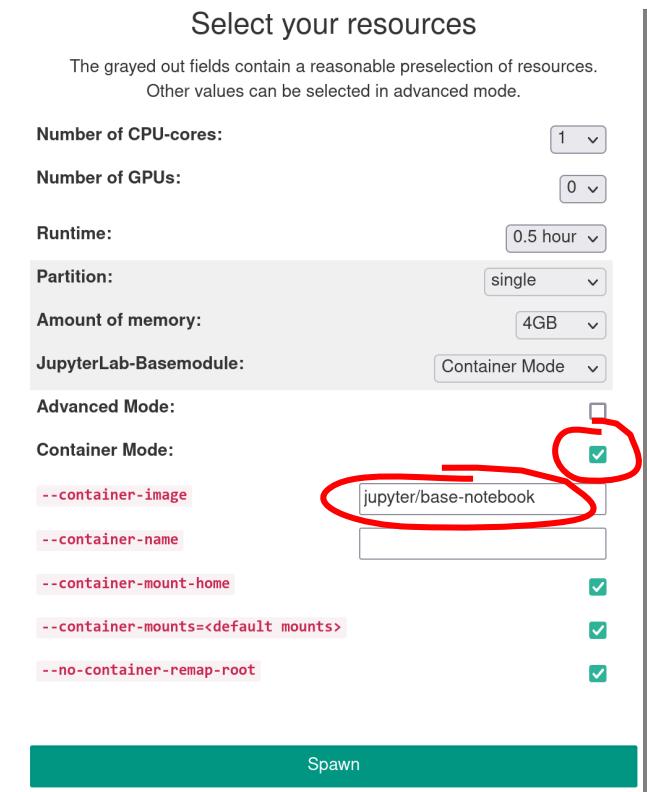
--container-name

--container-mount-home

--container-mounts=<default mounts>

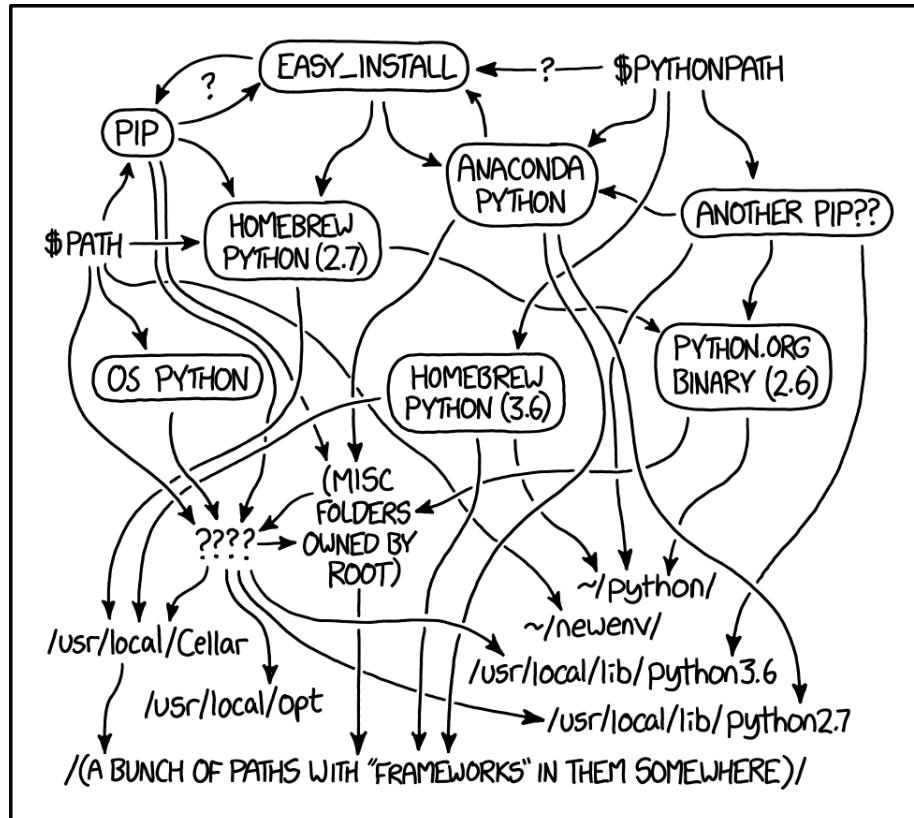
--no-container-remap-root

Spawn



Thank you for your attention!

Questions?



<https://xkcd.com/1987>

MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.