HPC.NRW

NATIONALES HOCHLEISTUNGS RECHNEN

EDIH Rheinland

WEST AI KI-Servicezentrum

PPCES: Machine and Deep Learning

Hands-on Exercises

DAS KOMPETENZNETZWERK FÜR HOCHLEISTUNGSRECHNEN.

General Information about Course Material

HPC.NRW

- **Hands-on material and slide decks can be found on the PPCES website**
  - https://blog.rwth-aachen.de/itc-events/en/event/ppces-2025/
- **Download and extract material on desired machine**
  - Separate folders for scikit-learn and PyTorch
  - scikit-learn available as regular Python (`.py`) and Jupyter Notebooks (`.ipynb`)

```
# change to the directory where you want to save your material
cd <working directory>

# download material and examples
wget https://hpc.rwth-aachen.de/ppces/ppces2025-ML-DL-labs.tar.gz

# unpack the tar file
tar -xzvf ppces2025-ML-DL-labs.tar.gz
```

3 Ways to Access the HPC Cluster

HPC.NRW

Classic / Batch Mode

Login via SSH
SLURM Workload Manager

Module System
Apptainer (Singularity)

Remote Desktop

Login via FastX
Graphical Interface

Rest is the same as in Batch Mode

Interactive Mode

JupyterHub/Lab
Web Interface

Predefined Profiles

Accessing CLAIX – Hardware

HPC.NRW

- **Dedicated hardware (and accounting) for the workshop**
  - 4 CLAIX-2023-HPC nodes
  - 10 CLAIX-2023-ML nodes

	MPI & OpenMP	ML/DL
Advanced reservation	`PPCES-m`	`PPCES-g`
Compute time project	`lect0138`	

Example in batch script file:

```
#SBATCH --reservation=PPCES-m
#SBATCH --account=lect0138
```

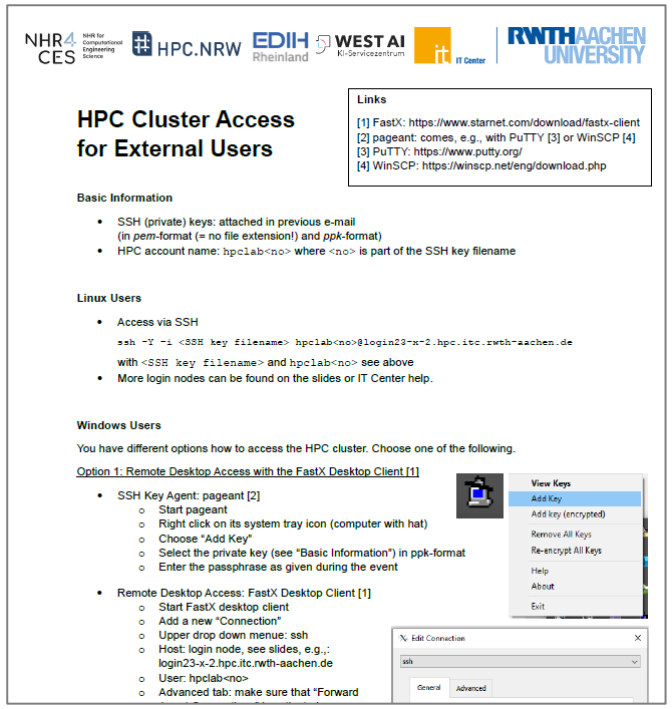
- Login nodes (full list here)
  - login23-2.hpc.itc.rwth-aachen.de (SSH)
  - login23-3.hpc.itc.rwth-aachen.de (SSH)
  - login23-x-1.hpc.itc.rwth-aachen.de (X-Server, remote desktop sessions, web access available)
  - login23-x-2.hpc.itc.rwth-aachen.de (X-Server, remote desktop sessions, web access available)

Accessing CLAIX – Accounts

- **Members of RWTH (or affiliated persons)**
  - HPC account + Two-factor-authentication (2FA) required:
    - https://regapp.itc.rwth-aachen.de/
    - Instructions (Step 3 – Step 5)
  - If provided during survey, all permissions for dedicated hardware have already been granted
  - If not, please contact the IT center staff during the lab session!

- **External participants**
  - Temporary account and SSH key was sent to you by e-mail
    - Account: hpclab[01-20]
    - Key: Linux format (pem) or PuTTY/pageant format (ppk)
    - If you didn't get an e-mail, please contact the IT Center staff
  - Password: will be provided during the lab sessions
  - Hpclab accounts have all permissions for dedicated hardware

  - See **Handout: HPC Cluster Access** for further details



HPC Cluster Access for External Users

Links
[1] FastX: https://www.starnet.com/download/fastx-client
[2] pageant: comes, e.g., with PuTTY [3] or WinSCP [4]
[3] PuTTY: https://www.putty.org/
[4] WinSCP: https://winscp.net/eng/download.php

Basic Information
- SSH (private) keys: attached in previous e-mail (in *pem*-format (= no file extension!) and *ppk*-format)
- HPC account name: `hpclab<no>` where `<no>` is part of the SSH key filename

Linux Users
- Access via SSH

```
ssh -Y -i <SSH key filename> hpclab<no>@login23-x-2.hpc.itc.rwth-aachen.de
```

with `<SSH key filename>` and `hpclab<no>` see above
- More login nodes can be found on the slides or IT Center help.

Windows Users
You have different options how to access the HPC cluster. Choose one of the following.

Option 1: Remote Desktop Access with the FastX Desktop Client [1]
- SSH Key Agent: pageant [2]
  - Start pageant
  - Right click on its system tray icon (computer with hat)
  - Choose "Add Key"
  - Select the private key (see "Basic Information") in ppk-format
  - Enter the passphrase as given during the event
- Remote Desktop Access: FastX Desktop Client [1]
  - Start FastX desktop client
  - Add a new "Connection"
  - Upper drop down menue: ssh
  - Host: login node, see slides, e.g.,: login23-x-2.hpc.itc.rwth-aachen.de
  - User: hpclab<no>
  - Advanced tab: make sure that "Forward

Accessing CLAIX – Access (SSH)

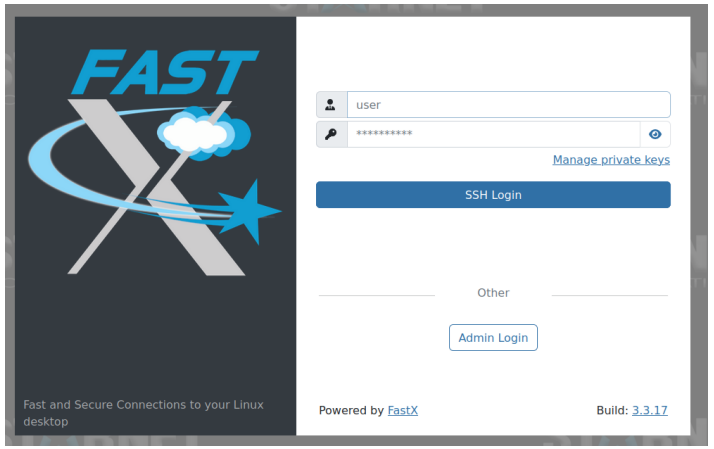
HPC.NRW

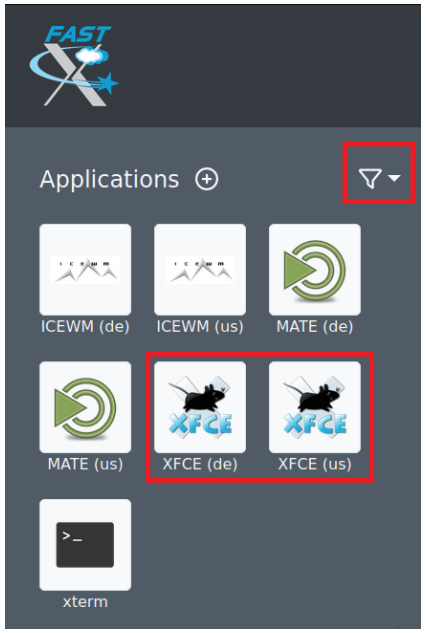
- **Access via SSH**
  - Access from within the RWTH network / eduroam via ssh:
    - ssh -Y ab123456@login23-2.hpc.itc.rwth-aachen.de
  - Tools:
    - Linux: Bash/Zsh/… + SSH
    - PuTTY: https://www.putty.org/
    - GitBash for Windows: https://git-scm.com/download/win

Accessing CLAIX – Access (FastX)

HPC.NRW

- **FastX3 (desktop client or browser)**
  - Browser client: https://login23-x-2.hpc.itc.rwth-aachen.de:3300
  - Desktop client: login nodes (see slides before)
  - Select the correct application
    - Recommended: XFCE
    - May need to select “all applications” in filter at the top right corner of the Application box
  - Start the session
    - Click on the session (may have to click twice)
  - Open up a terminal
    - Bottom of the screen
    - Click the “terminal” icon





Part 1: Hands-on: scikit-learn

Hands-on: scikit-learn – Getting Started with JuypterHub

HPC.NRW

- **Large Jobs / high demand → JupyterHub HPC**
  - Special JupyterHub instance
  - Starts containers on HPC cluster nodes
  - https://jupyterhub.hpc.itc.rwth-aachen.de:9651

- **Alternatives**
  - Another RWTH JupyterHub
    - Cheaper standard servers under the hood
    - https://jupyter.rwth-aachen.de/hub/login
    - Profiles / kernles might differ
  - Install JupyterLab in your local virtual environment
    - Execution happens on your local machine
    - Visual Studio Code integration available

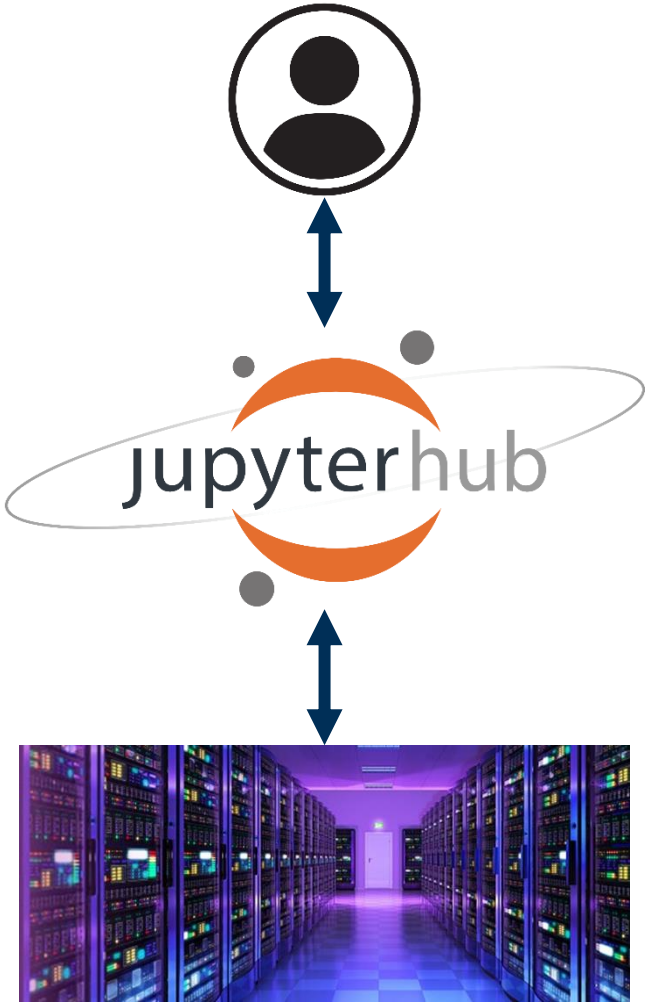
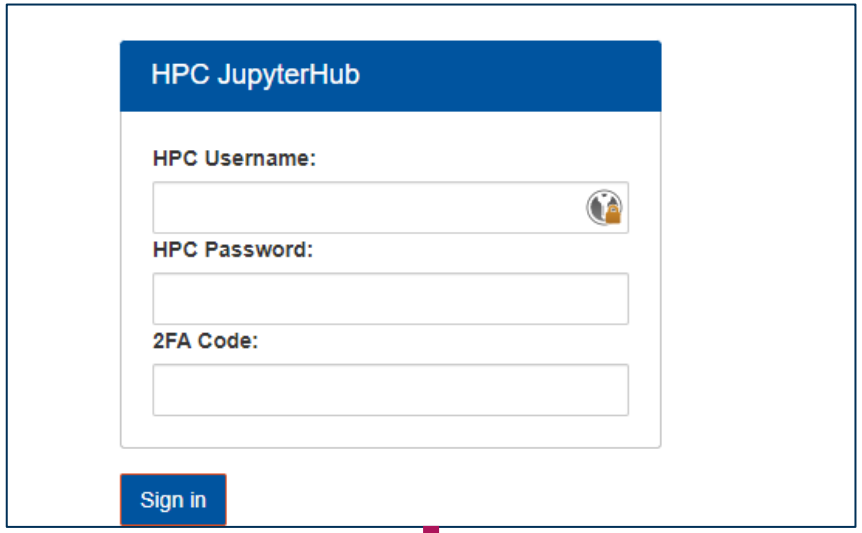


Image: scientific-computing.com

Hands-on: scikit-learn – Getting Started with JuypterHub

HPC.NRW

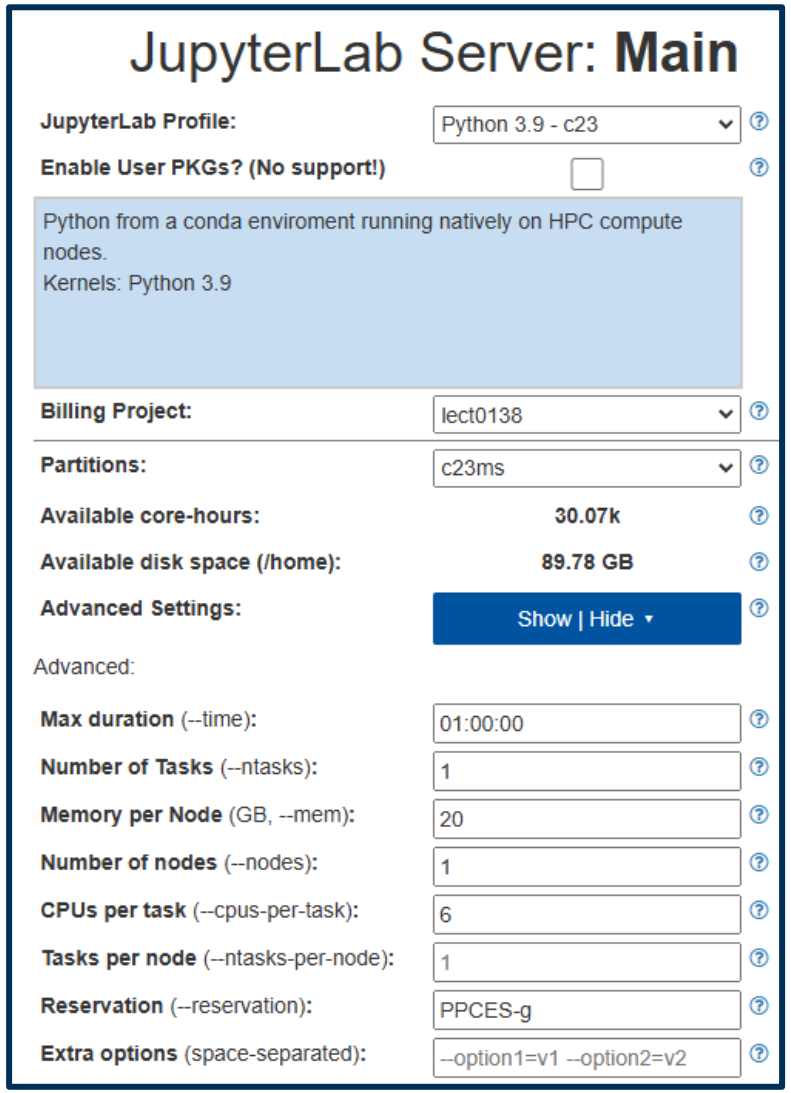
- **Step 1:** Login to JupyterHub HPC
  - Similar to SSH or FastX login
  - You need your HPC 2FA again here



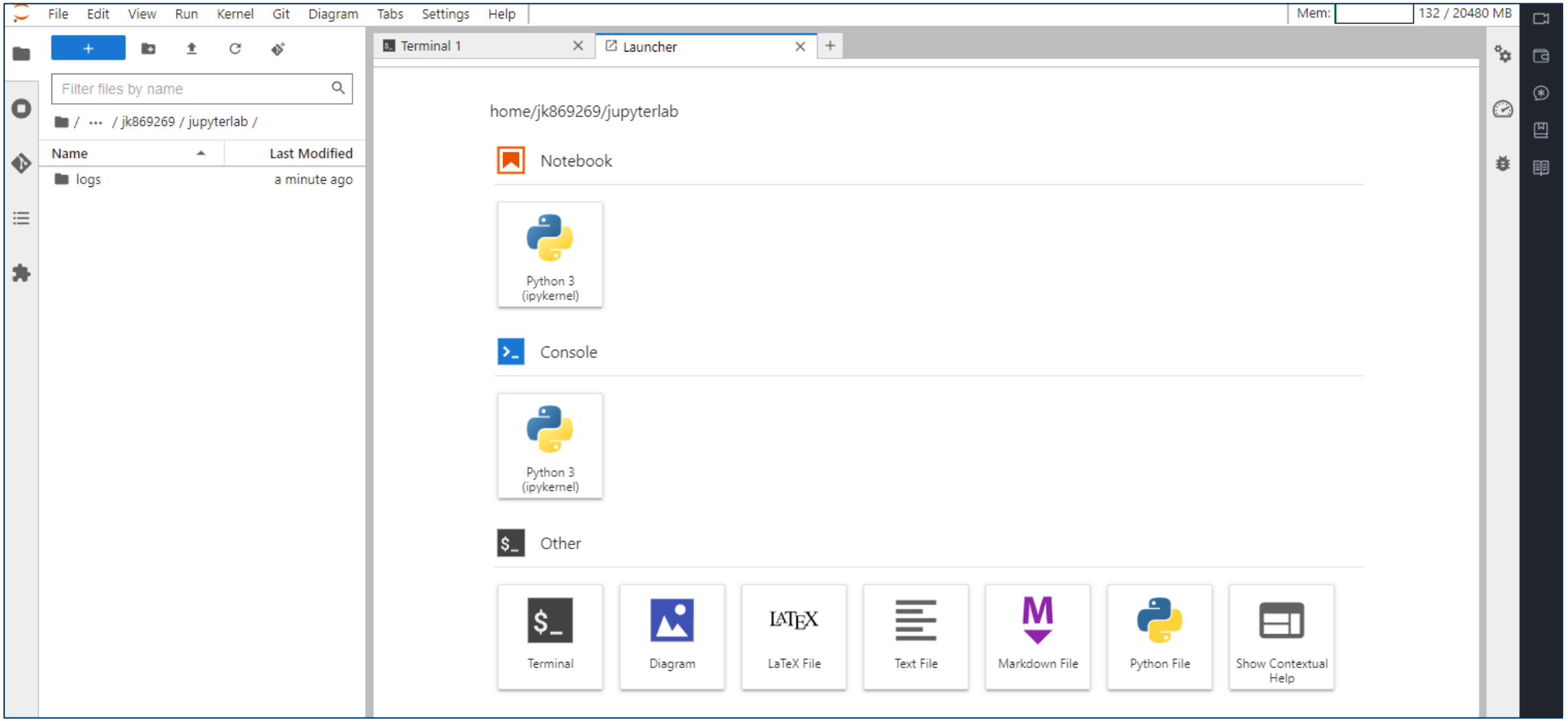
Hands-on: scikit-learn – Getting Started with JuypterHub

HPC.NRW

- **Step 2:** Configure and start your JupyterLab server
  - Request resources on the HPC cluster
  - General Settings
    - JupyterLab Profile Python 3.9 – c23
    - Billing Project lect0138
    - Partition c23ms
  - Advanced Settings
    - Number of Tasks 1
    - CPUs per task 6
    - Memory per Node: 20 GB
    - Max duration max. 1h
    - Reservation PPCES-g



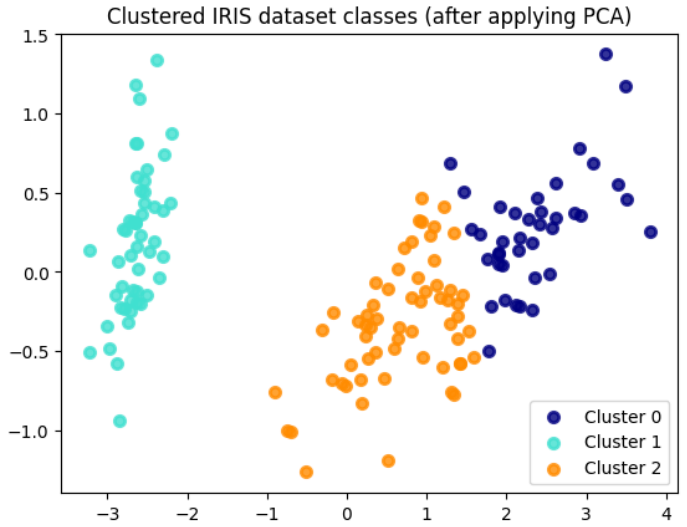
Hands-on: scikit-learn – Getting Started with JuypterHub



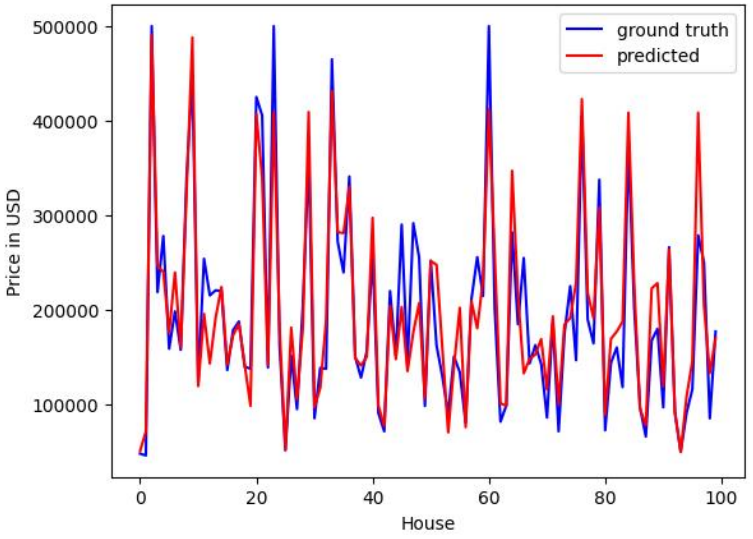
Hands-on: scikit-learn

HPC.NRW

- **Exercise 1:** Clustering with Iris Dataset
  - Load dataset
  - Train clustering model (here: Kmeans clustering)
  - Apply dimensionality reduction (PCA)
  - Visualize and compare against true labels



- **Exercise 2:** Regression with California Housing Dataset
  - Load dataset
  - Apply preprocessing techniques (Standardization)
  - Train regression model (here: RandomForest)
  - Model performance evaluation and visualization



Part 2: Hands-on: PyTorch

Hands-on: PyTorch (with classical SSH)

HPC.NRW

- **Exercise:** Train ResNet model with CIFAR-10 dataset
  - Model: ResNet
    - Popular model for image classification
    - Winner of ILSVRC 2015
      (ImageNet Large Scale Visual Recognition Challenge)
    - Tackles vanishing gradient problem
  - Dataset: CIFAR-10
    - Several images of 10 different classes
    - Airplane, automobile, bird, cat, dog, frog, …
  - Single GPU (Task 1), Distributed (Task 2)

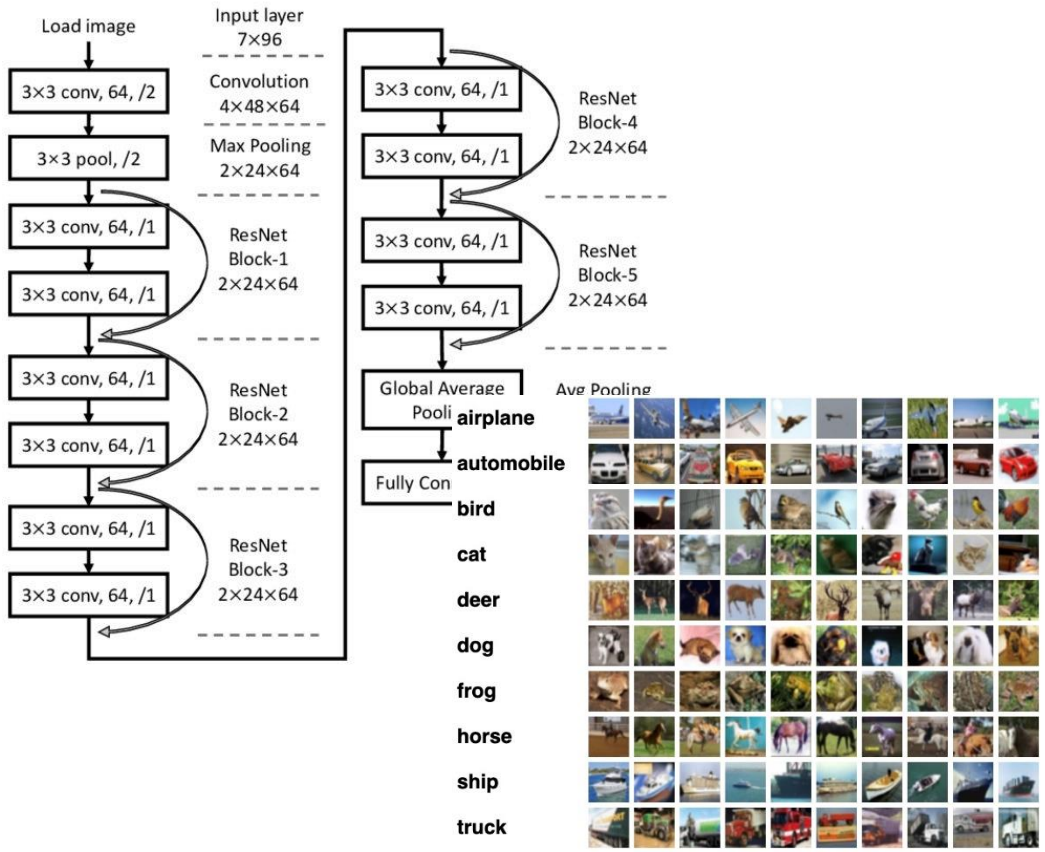


Diagram: Choi, Hyungeun & Ryu, Seunghyoung & Kim, Hongseok. (2018). Short-Term Load Forecasting based on ResNet and LSTM

Hands-on: PyTorch (with classical SSH)

HPC.NRW

– **Note:** We will use a container for this exercise!

	File	Description
Provided by us	`set_vars.sh`	Shell script that sets environment variables for DDP and container
Provided by us	`submit_job.sh`	Submits a batch job to SLURM, which loads the required container module and executes the Python code
Your job	`train_model.py`	Python code that is responsible for training and testing the model

– **About: `submit_job.sh`**
  – Option to run with 1 GPU or 2 GPUs (distributed)
  – Option to enable explicit monitoring (`nvidia-smi`)

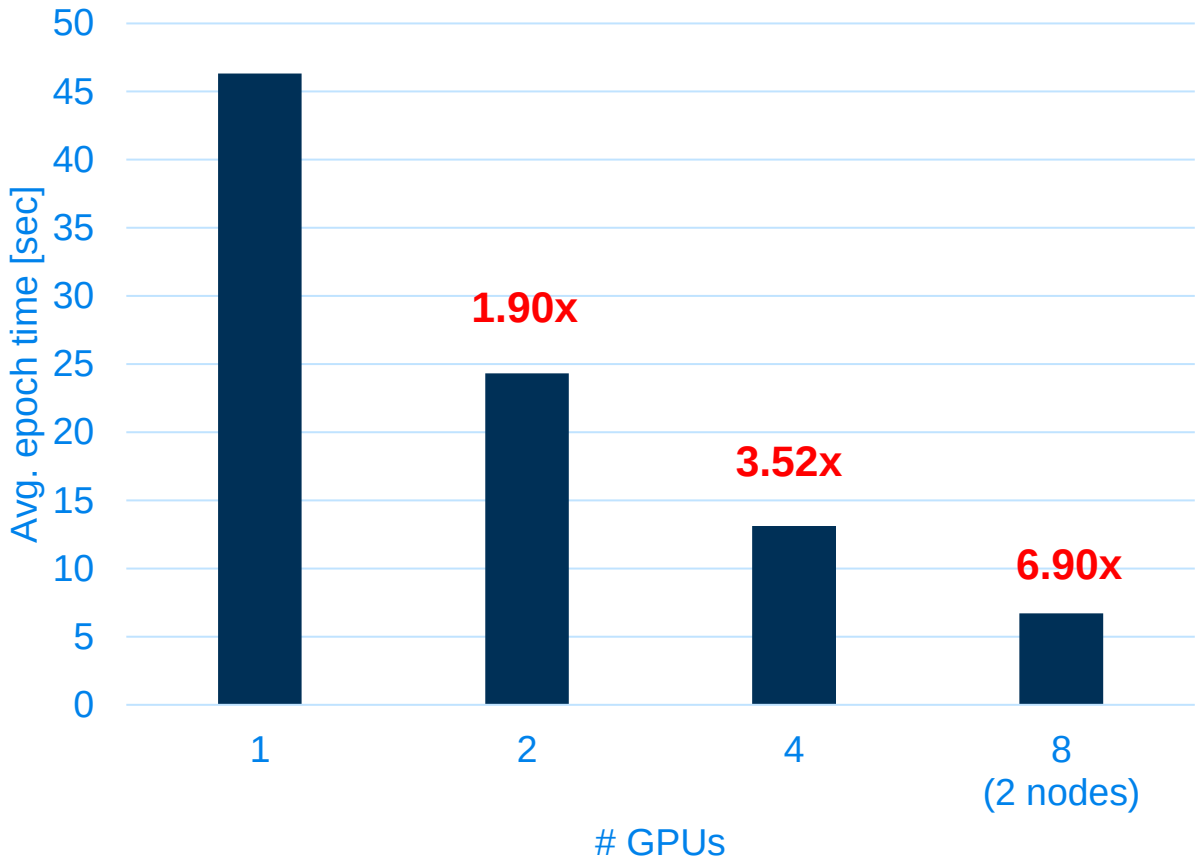
```
# execute with a single GPU
sbatch submit_job.sh

# execute with nvidia-smi monitoring
sbatch --export=ENABLE_MONITORING=1 submit_job.sh

# execute with 2 GPUs (distributed)
sbatch --ntasks-per-node=2 --gres=gpu:2 submit_job.sh

# check for status
squeue --me -a
```

Example

HPC.NRW

– Output and scaling behavior

```
...
Epoch 2/2  Step    0 /  391
Epoch 2/2  Step   20 /  391
Epoch 2/2  Step   40 /  391
Epoch 2/2  Step   60 /  391
Epoch 2/2  Step   80 /  391
Epoch 2/2  Step  100 /  391
Epoch 2/2  Step  120 /  391
Epoch 2/2  Step  140 /  391
Epoch 2/2  Step  160 /  391
Epoch 2/2  Step  180 /  391
Epoch 2/2  Step  200 /  391
Epoch 2/2  Step  220 /  391
Epoch 2/2  Step  240 /  391
Epoch 2/2  Step  260 /  391
Epoch 2/2  Step  280 /  391
Epoch 2/2  Step  300 /  391
Epoch 2/2  Step  320 /  391
Epoch 2/2  Step  340 /  391
Epoch 2/2  Step  360 /  391
Epoch 2/2  Step  380 /  391
Epoch 2/2  Elapsed: 47.857 sec   Acc: 0.642
Epoch 2/2  Test Acc: 0.648
```



Troubleshooting Section

HPC.NRW

- **Problem:** PyTorch throws errors when executing distributed variant
  - “Port already in use”
  - **Reasons:** You will potentially be working on a shared machine (with only 2 of 4 GPUs)

- **Solution:** Select a different port

Troubleshooting Section

HPC.NRW

- **Problem:** Be carful with using `$(pwd)` or `$PWD` inside a container
  - This might return /rwthfs/… which is not mounted by default
  - What's interesting:
    - **/work/<user-id>/** is mounted by default
    - **/rwthfs/rz/cluster/work/<user-id>/** is not although pointing to the same path
  - **Cause:** Can occur when working with multiplexers like tmux on the host system

- **Solution:**
  - Map the directories that you want to use in the container
  - Also see hands-on examples