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-2024/
- **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/ppces2024-ML-DL-labs.tar.gz

# unpack the tar file
tar -xzvf ppces2024-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 the CLAIX HPC Cluster

Accounts

Members of RWTH (or affiliated persons)

- HPC account required:
  https://regapp.itc.rwth-aachen.de/
- Two-factor-authentication (2FA) required

External participants

- Temporary account will be provided

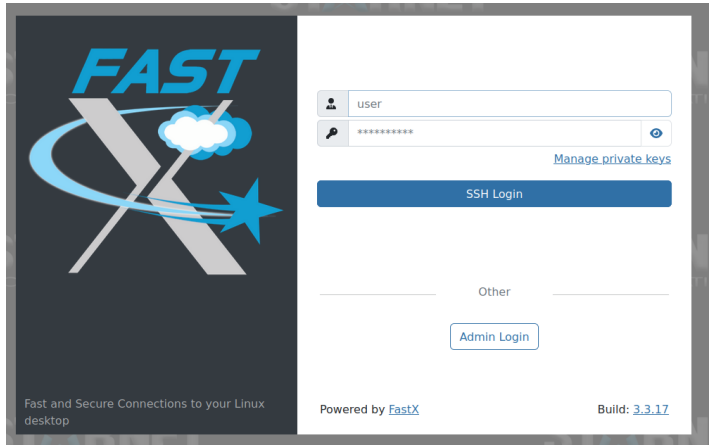
Dedicated hardware for the workshop

	OpenMP days	**MPI days**	**ML day**
Advanced reservation	`PPCES24`	`PPCES24MPI`	`PPCES24GPU`
Compute time project	`lect0113`	`supp0006`	`lect0114`

Example in batch script file:

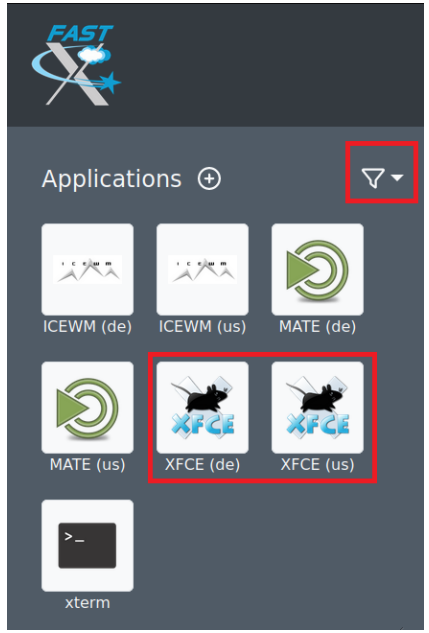
```
#SBATCH -reservation=PPCES24
#SBATCH -account=lect0113
```

Remote Desktop Access



FastX3 (desktop client or browser)

- Browser client: https://login23-x-2.hpc.itc.rwth-aachen.de:3300
- Desktop client: login nodes (see next slide)
- 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



Cluster Access

Login nodes (full list here)

- login23-x-2.hpc.itc.rwth-aachen.de (CLAIX-2023, X-Server)
- login23-2.hpc.itc.rwth-Aachen.de (CLALIX-2023)
- login23-3.hpc.itc.rwth-Aachen.de (CLAIX-2023)
- login18-2.hpc.itc.rwth-aachen.de (CLAIX-2018)
- login18-3.hpc.itc.rwth-aachen.de (CLAIX-2018)

Access via SSH

Access from within the RWTH network / eduroam via ssh:

- ssh -Y ab123456@login23-x-2.hpc.itc.rwth-aachen.de
- GitBash for Windows: https://git-scm.com/download/win

Cluster Access: External Participants Only!

Account

- Username: `hpclab[01-50]`
  - Distribution of accounts by speaker/helpers
- Password: see paper sheet

Access

- Login node: `login18-beta.itc.hpc.rwth-aachen.de`
  - Use desktop client or
  - Web interface: https://login18-beta.hpc.itc.rwth-aachen.de:3300
- Jump from there to a CLAIX-2023 frontend node (see node list on previous slide)

```
ssh -Y hpclab[01-50]@login23-2.hpc.itc.rwth-aachen.de
```

Note: Do not store any private/ sensitive data on the cluster

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

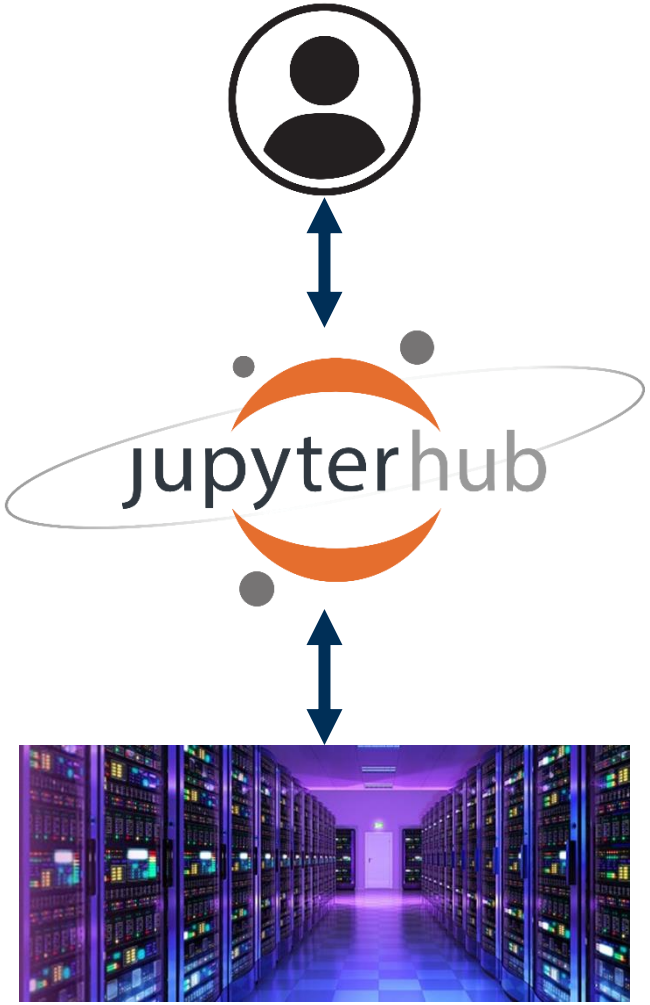
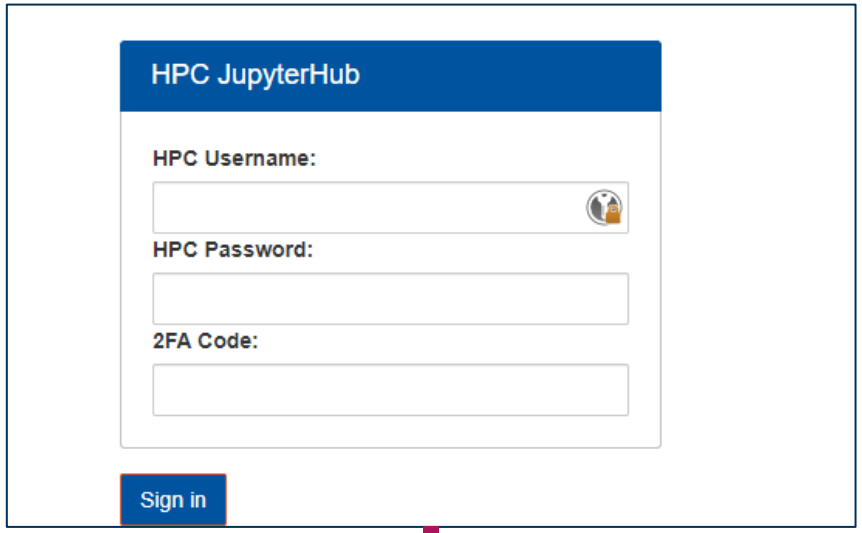


Image: scientific-computing.com

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

HPC.NRW

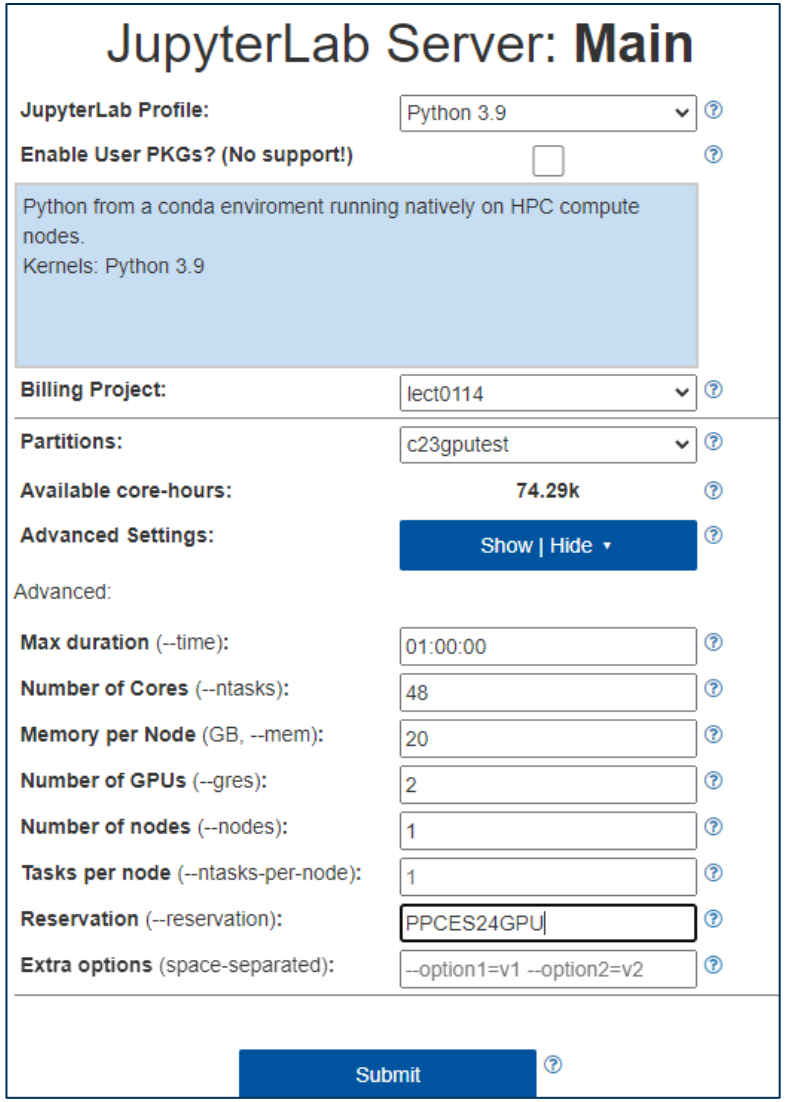
- **Step 1:** Login to JupyterHub HPC
  - Similar to SSH or FastX login
  - You need your HPC 2FA again here



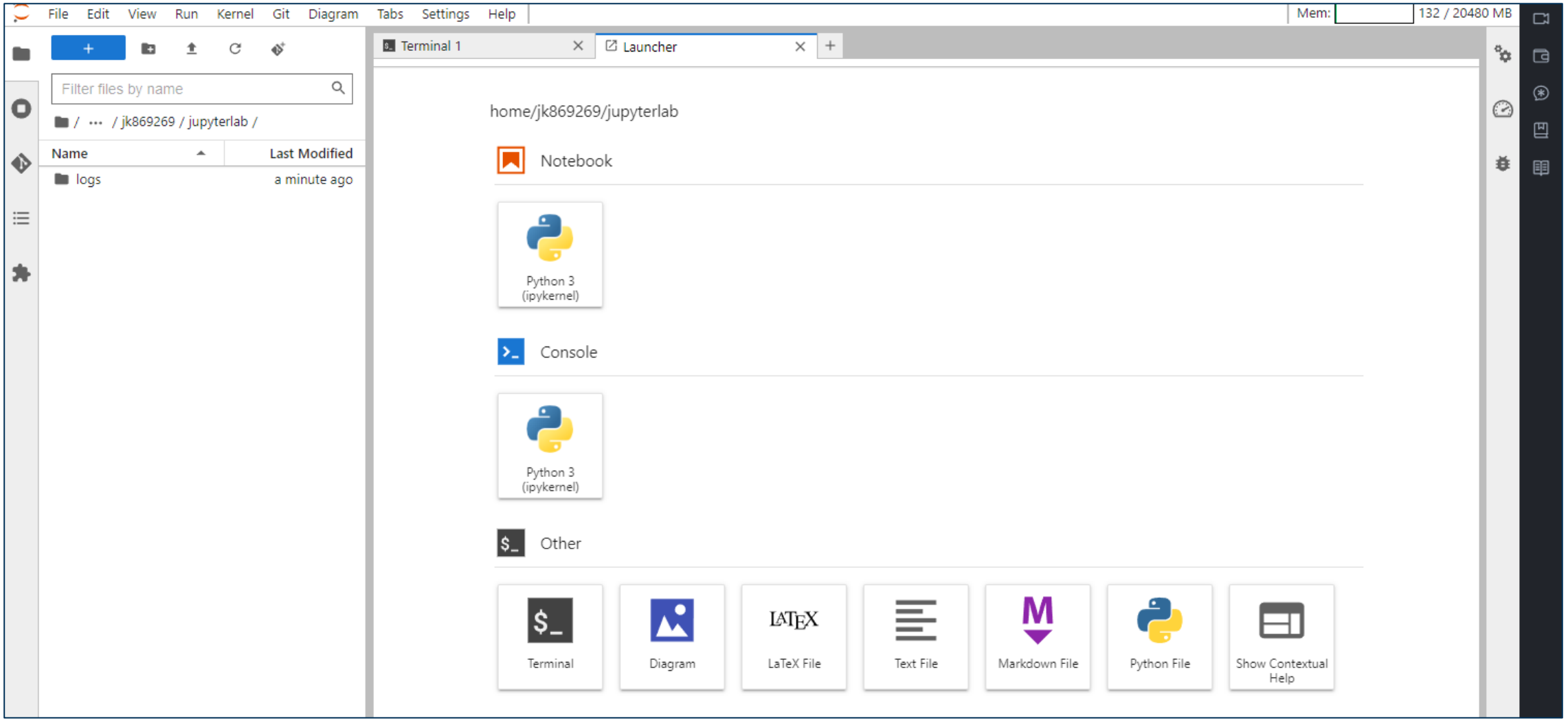
Hands-on: scikit-learn – Getting Started with JuypterHub

HPC.NRW

- **Step 2:** Configure and start your JupyterLab server
  - Request resources on the HPC cluster
  - General Settings
    - JupyterLab Profile Python 3.9
    - Billing Project lect0114
    - Partition c23gputest
  - Advanced Settings
    - Number of cores max. 48
    - Number of GPUs max. 2 (not used in scikit-learn)
    - Max duration max. 1h
    - Reservation PPCES24GPU



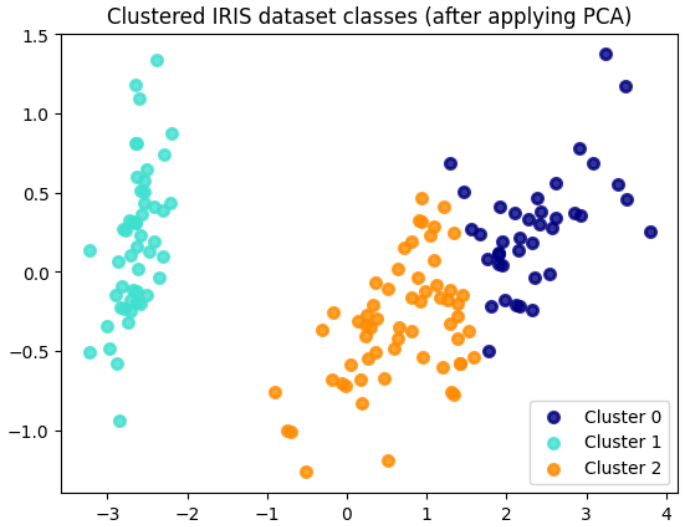
Hands-on: scikit-learn – Getting Started with JuypterHub



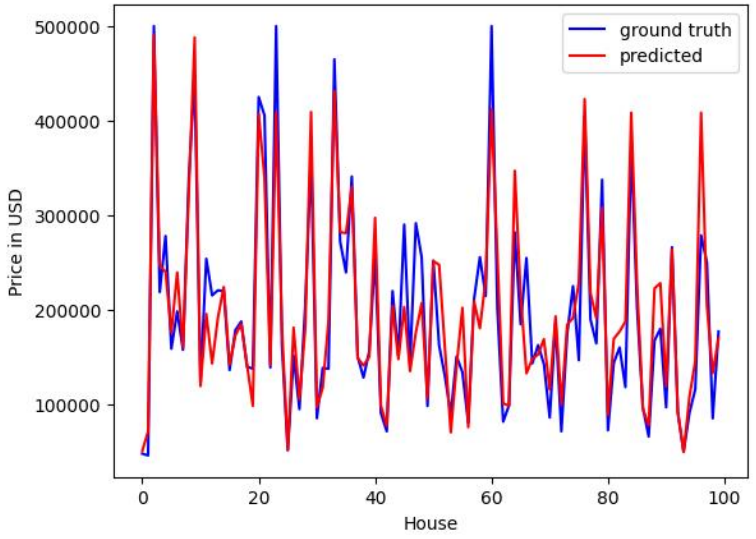
Hands-on: scikit-learn

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



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)

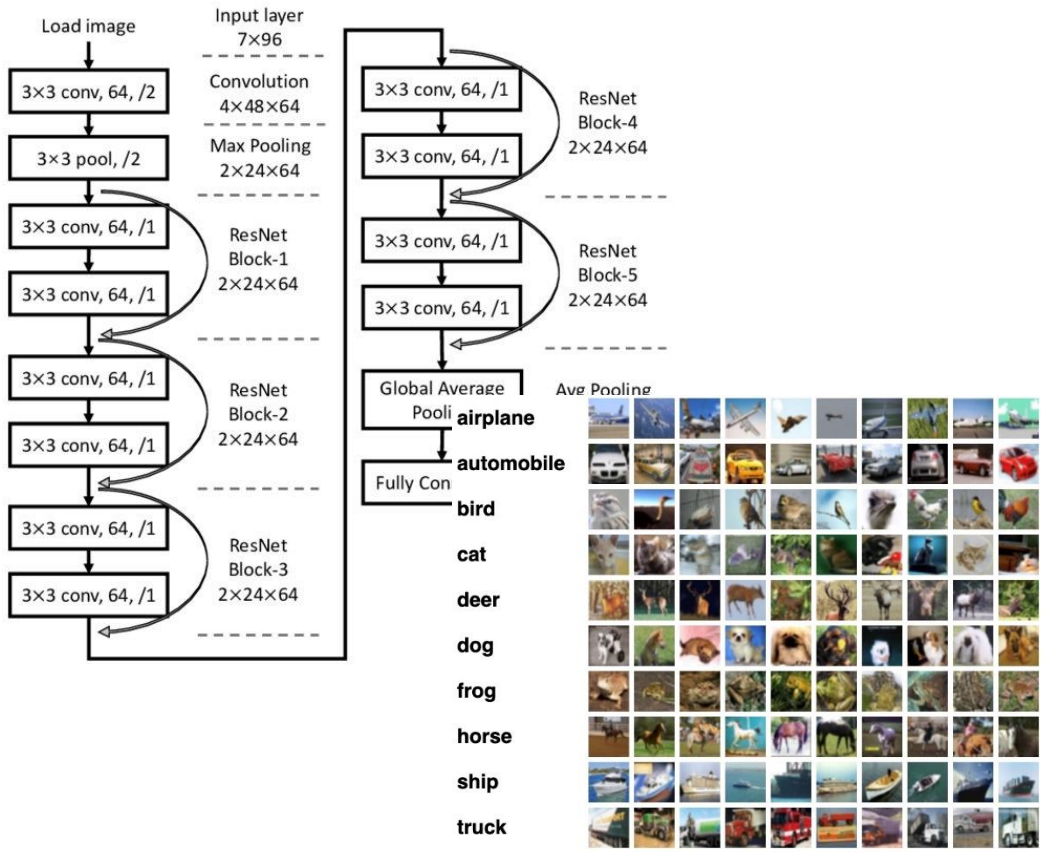


Diagram: Choi, Hyungeun & Ryu, Seunghyoung & Kim, Hongseok. (2018). Short-Term Load Forecasting based on ResNet and LSTM

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`)

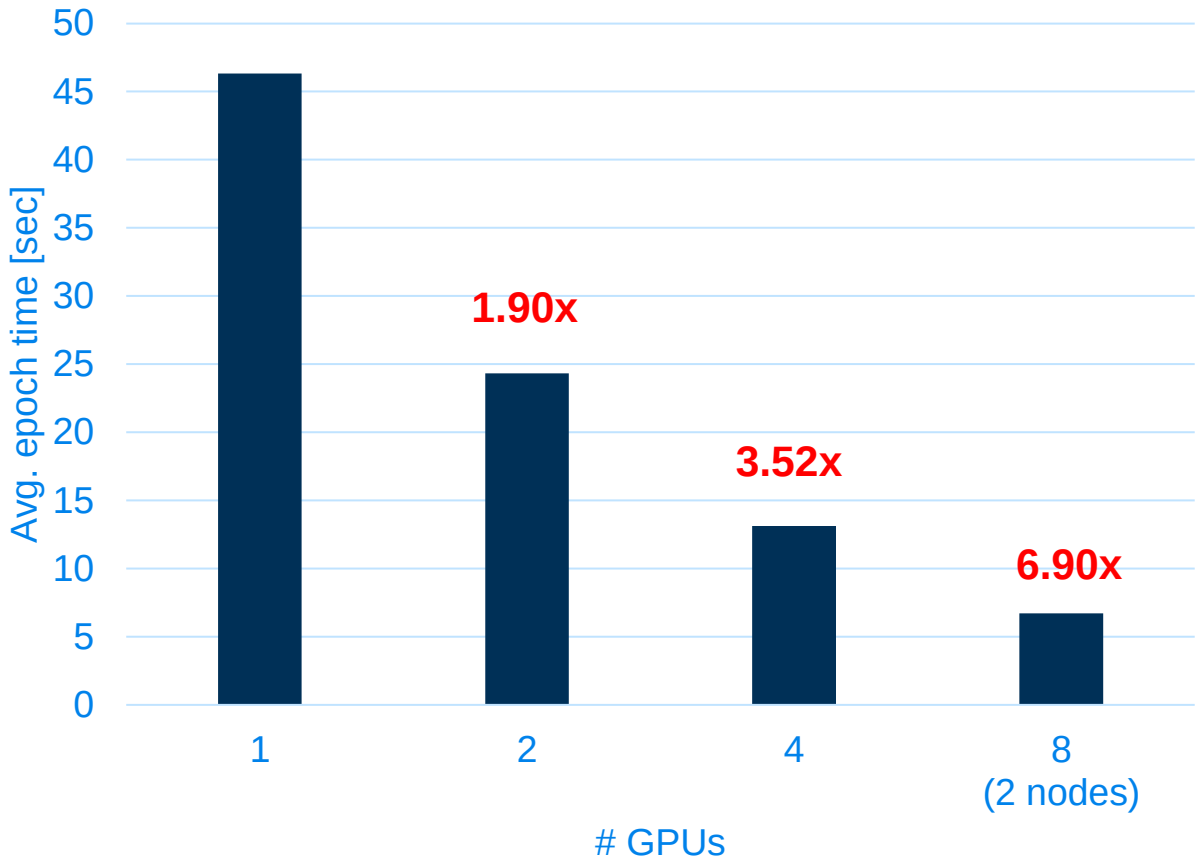
```
# execute with a single GPU
sbatch submit_job.sh

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

# execute with 2 GPUs (distributed)
sbatch --export=RUN_DISTRIBUTED=1 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