上海科技大学
ShanghaiTech University

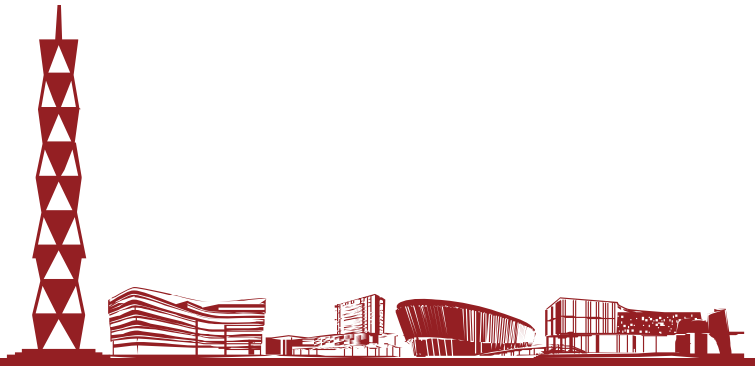
AI cluster Introduction

Chuyang Xiao
2025/4/15

立志成才报国裕民

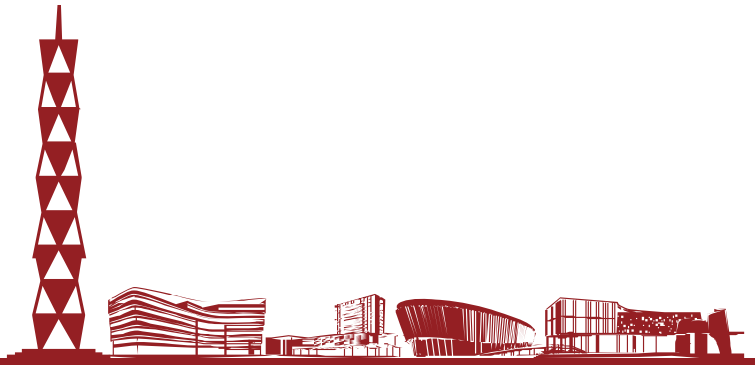
Outline

- Conda environment
- Login, debug, compute nodes & how to connect
- Some tools
- Demonstration



Conda environment

- A package manager
  - Install, update, and remove packages (Python/R/etc.)
  - Resolve dependency conflicts
  - Create isolated environments (e.g., conda create -n myenv python=3.9)
- AI cluster already has Anaconda installed -> can use conda directly



AI cluster login

上海科技大学
ShanghaiTech University

An **AI Cluster** is a high-performance computing system designed for accelerating AI workloads.

- Login nodes:
  - 10.15.89.191
  - 10.15.89.192
  - 10.15.89.41

❌ run code
❌ GPU
✅ Internet -> pip install

- Debug nodes:
  - 10.15.88.73
  - 10.15.88.74

✅ debug code
✅ already have GPU

- Compute nodes:
  - access Via CS182 queue

✅ run code
✅ GPU
❌ Internet -> pip install

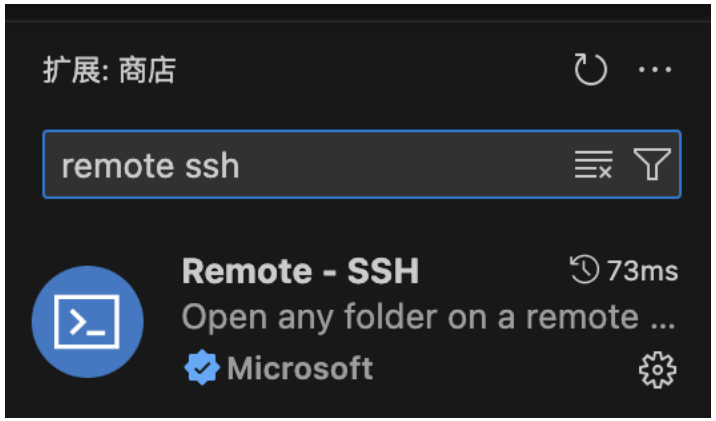
Guide:
http://10.15.89.177:8889/app/index.html

```
(robodiff) [xiaochy@debug01 ~]$ nvidia-smi
Mon Apr 14 23:16:08 2025
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 530.41.03              Driver Version: 530.41.03    CUDA Version: 12.1     |
|-------------------------------+----------------------+----------------------+
| GPU  Name                 Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf           Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                                        |                      |               MIG M. |
|===============================+======================+======================|
|   0  NVIDIA GeForce GTX 1080        On | 00000000:02:00.0 Off |                  N/A |
| 28%   25C    P8               7W / 180W|    734MiB /  8192MiB |      0%      Default |
|                                        |                      |                  N/A |
+-------------------------------+----------------------+----------------------+
|   1  NVIDIA GeForce GTX 1080        On | 00000000:03:00.0 Off |                  N/A |
| 28%   25C    P8               7W / 180W|      4MiB /  8192MiB |      0%      Default |
|                                        |                      |                  N/A |
+-------------------------------+----------------------+----------------------+
|   2  NVIDIA GeForce GTX 1080        On | 00000000:83:00.0 Off |                  N/A |
| 28%   34C    P2              39W / 180W|   1994MiB /  8192MiB |      0%      Default |
|                                        |                      |                  N/A |
+-------------------------------+----------------------+----------------------+
|   3  NVIDIA GeForce GTX 1080        On | 00000000:84:00.0 Off |                  N/A |
| 27%   25C    P8               7W / 180W|      4MiB /  8192MiB |      0%      Default |
|                                        |                      |                  N/A |
+-------------------------------+----------------------+----------------------+
```

SSH ->login node

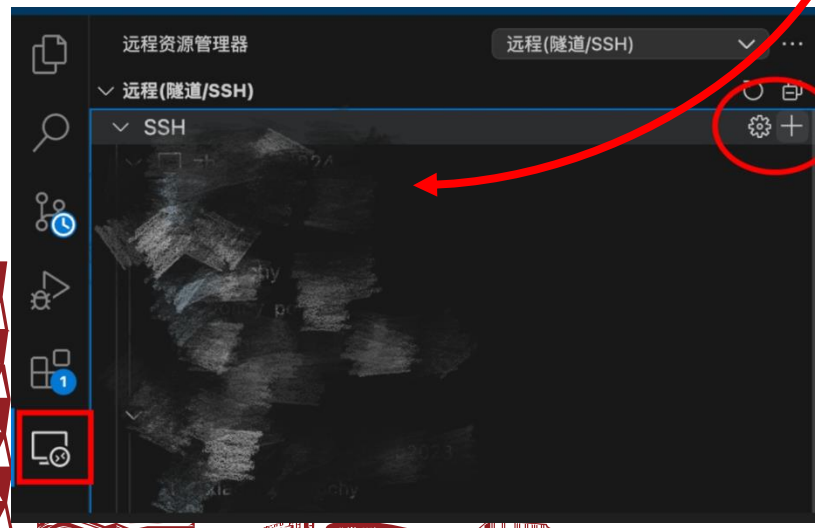
上海科技大学
ShanghaiTech University

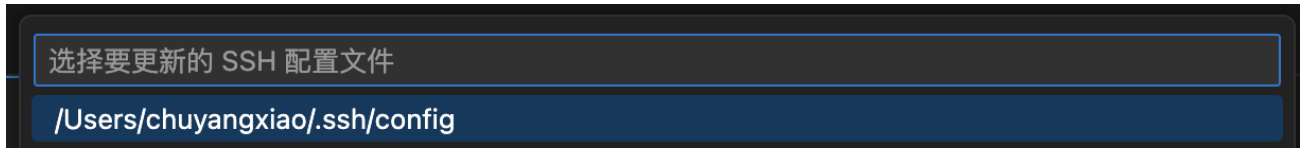
- VSCode extension



Can be changed by yourself

```
Host 10.15.89.191
  HostName 10.15.89.191
  Port 22112
  User zhang3-cs182
```





- User name: email prefix-cs182
- Password: sist -> change via yppasswd

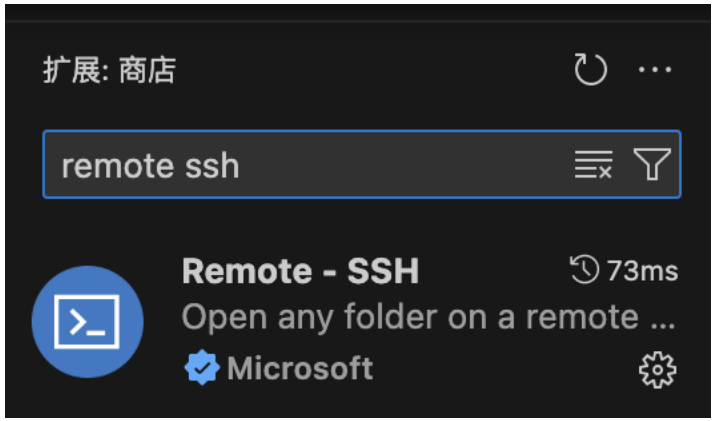
Will fill the config file automatically

立志成才报国裕民

SSH ->login node

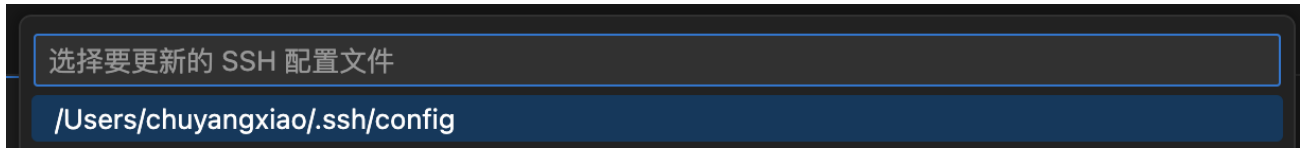
上海科技大学
ShanghaiTech University

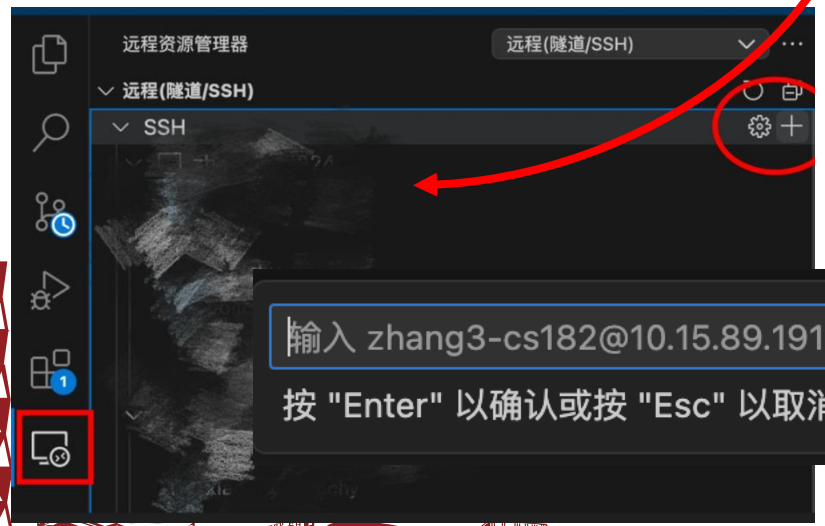
- VSCode extension



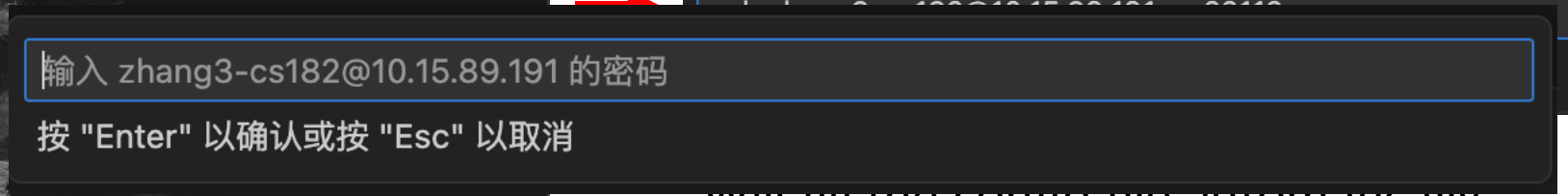
Can be changed by yourself

```
Host 10.15.89.191
  HostName 10.15.89.191
  Port 22112
  User zhang3-cs182
```





- User name: email prefix-cs182
- Password: sist -> change via yppasswd



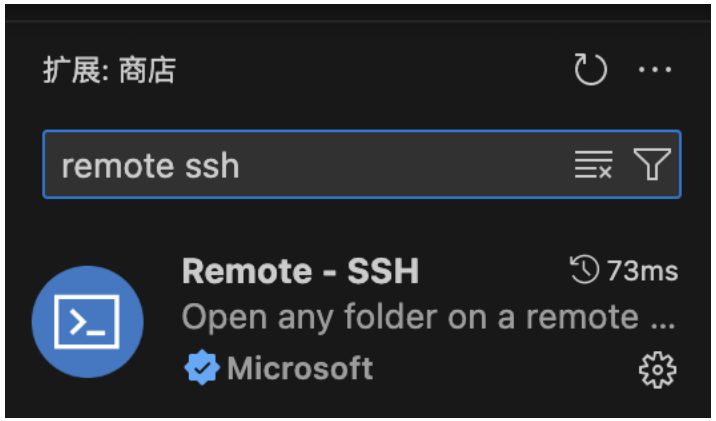
Will fill the config file automatically

立志成才报国裕民

SSH ->login node

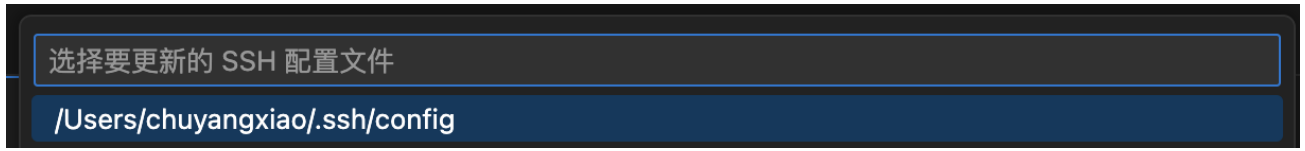
上海科技大学
ShanghaiTech University

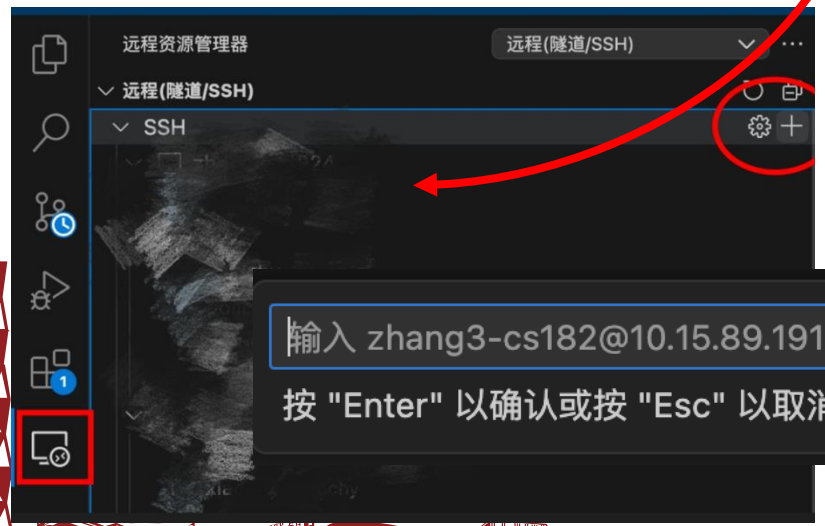
- VSCode extension



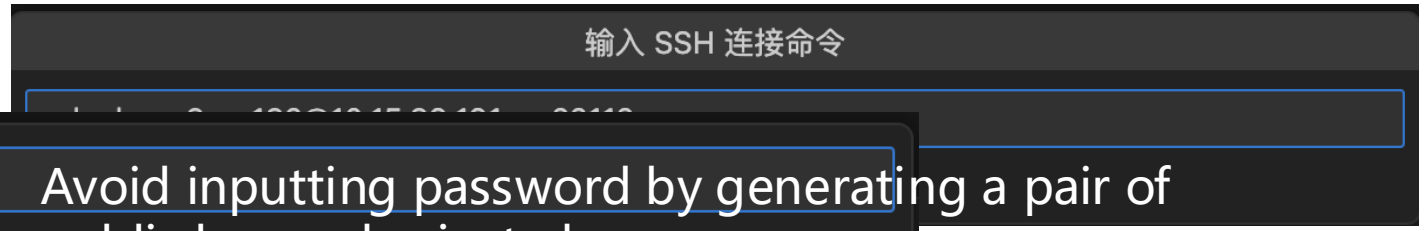
Can be changed by yourself

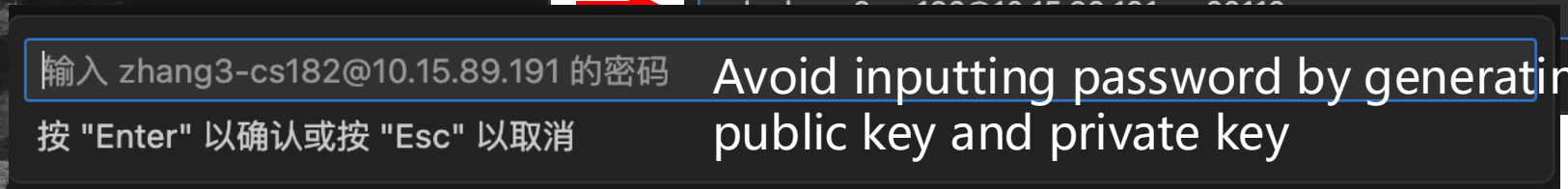
```
Host 10.15.89.191
  HostName 10.15.89.191
  Port 22112
  User zhang3-cs182
```





- User name: email prefix-cs182
- Password: sist -> change via yppasswd



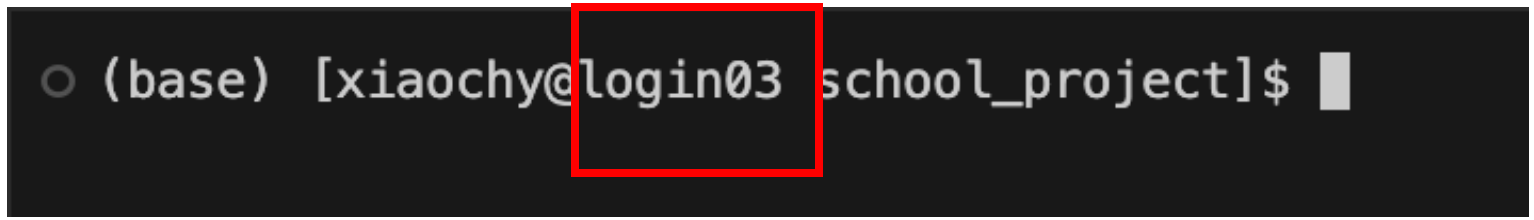


Avoid inputting password by generating a pair of public key and private key

Will fill the config file automatically

立志成才报国裕民

SSH ->login node

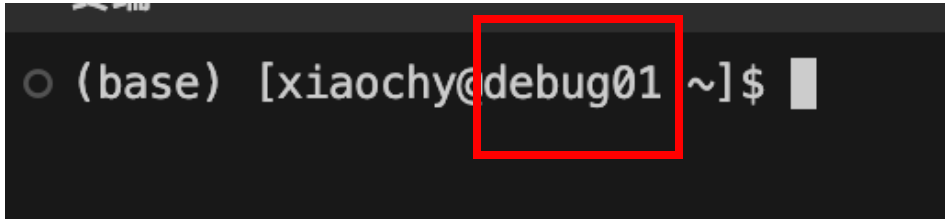


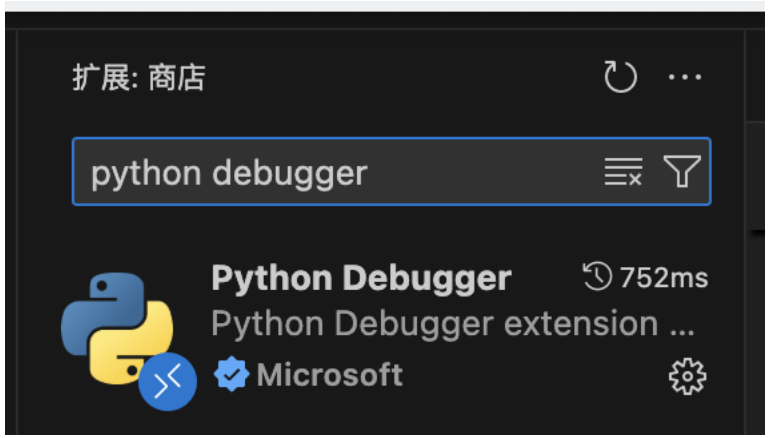
Success!

SSH ->debug node

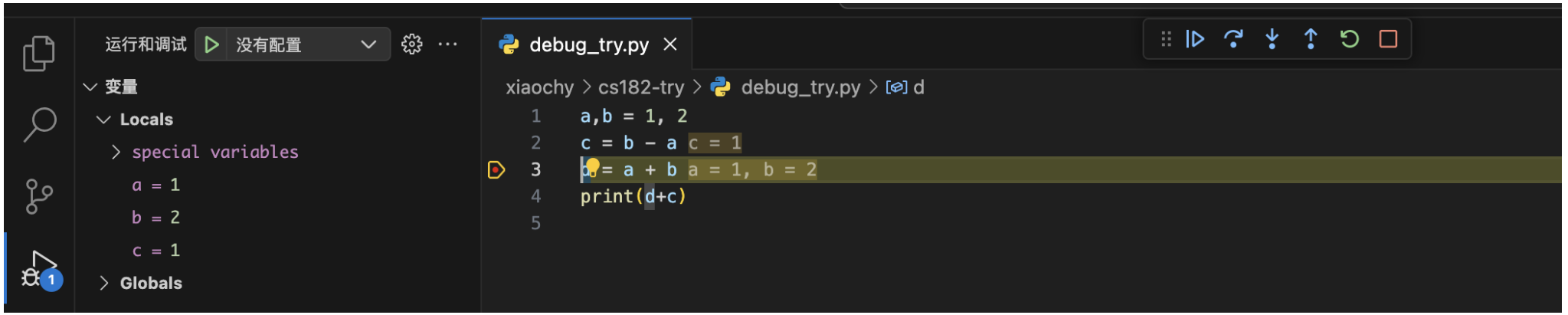
上海科技大学
ShanghaiTech University

- Same as the login node -> only change the ip address
- But without port number
- Vscode extension





ssh zhang3@10.15.88.73



SSH ->compute node

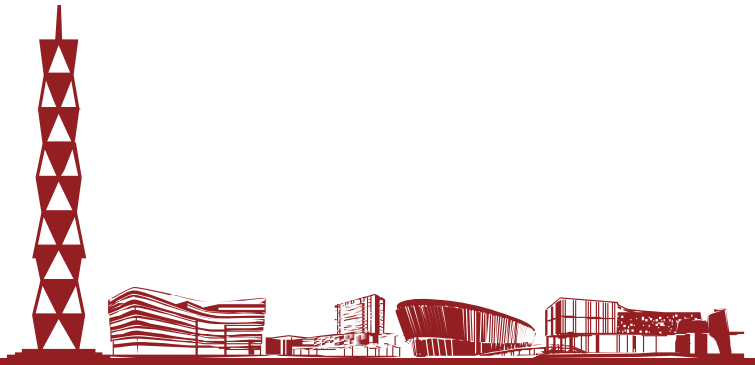
上海科技大学
ShanghaiTech University

- Write a .slurm script:
  - E.g. 2080_120G.slurm
- Submit the task:
  - E.g. sbatch 2080_120G.slurm

```
#SBATCH -J 2080_GPU          Task name
#SBATCH -p CS182             queue name
#SBATCH --cpus-per-task=12
#SBATCH -N 1
#SBATCH -t 2-00:00:00
#SBATCH --output=%j.out
#SBATCH --error=%j.err
#SBATCH --mail-type=ALL
#SBATCH --mem-per-cpu=10240
#SBATCH --gres=gpu:NVIDIAGeForceRTX2080Ti:1
#SBATCH --mail-user=zhang3@shanghaitech.edu.cn
# sleep 9999999
```

Your AI cluster account

Field	Description
Username	Your email prefix + "-cs182" (e.g. zhang3-cs182)
Initial Password	`sist` (Change immediately after login using `yppasswd`)
Queue	CS182
GPU Limits	Max 2 GPUs, each 2-day maximum runtime



SSH ->compute node

上海科技大学
ShanghaiTech University

```
#SBATCH -J 2080_GPU
#SBATCH -p CS182
#SBATCH --cpus-per-task=12
#SBATCH -N 1
#SBATCH -t 2-00:00:00
#SBATCH --output=%j.out
#SBATCH --error=%j.err
#SBATCH --mail-type=ALL
#SBATCH --mem-per-cpu=10240
#SBATCH --gres=gpu:NVIDIAGeForceRTX2080Ti:1
#SBATCH --mail-user=zhang3@shanghaitech.edu.cn
conda activate your_conda_env_name
cd /the_path_to_force_gpu.py
python force_gpu.py
```

force_gpu.py

```
import torch
from time import sleep
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
if not torch.cuda.is_available():
    print("CUDA is not available. Running on CPU...")

# You can change N here
N = 500
# Ensure persistent computation to maintain GPU occupancy
while True:
    a = torch.randn(N, N, device=device)
    b = torch.randn(N, N, device=device)
    c = torch.matmul(a, b)
    d = torch.matmul(a, b)
    sleep(0.01)
```

立志成才报国裕民

SSH ->compute node

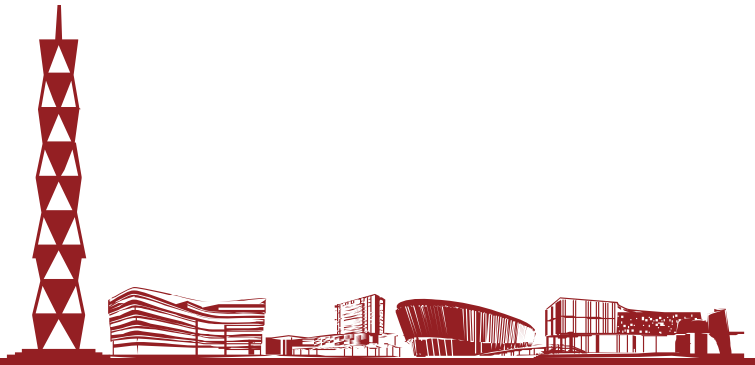
上海科技大学
ShanghaiTech University

```
(base) [xiaochy@login03 school_project]$ squeue
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
            736343    normal staction  xiaochy  R    3:20:44      1 sist_gpu70
            736344    normal staction  xiaochy  R    3:20:44      1 sist_gpu70
            736345    normal staction  xiaochy  R    3:20:44      1 ai_gpu21
```

Queue name

Task name

```
(base) [xiaochy@login03 school_project]$ ssh sist_gpu70
xiaochy@sist_gpu70's password:
Last login: Mon Apr 14 23:10:04 2025 from login03
-bash: ulimit: open files: cannot modify limit: Operation not permitted
(base) [xiaochy@sist_gpu70 ~]$
```

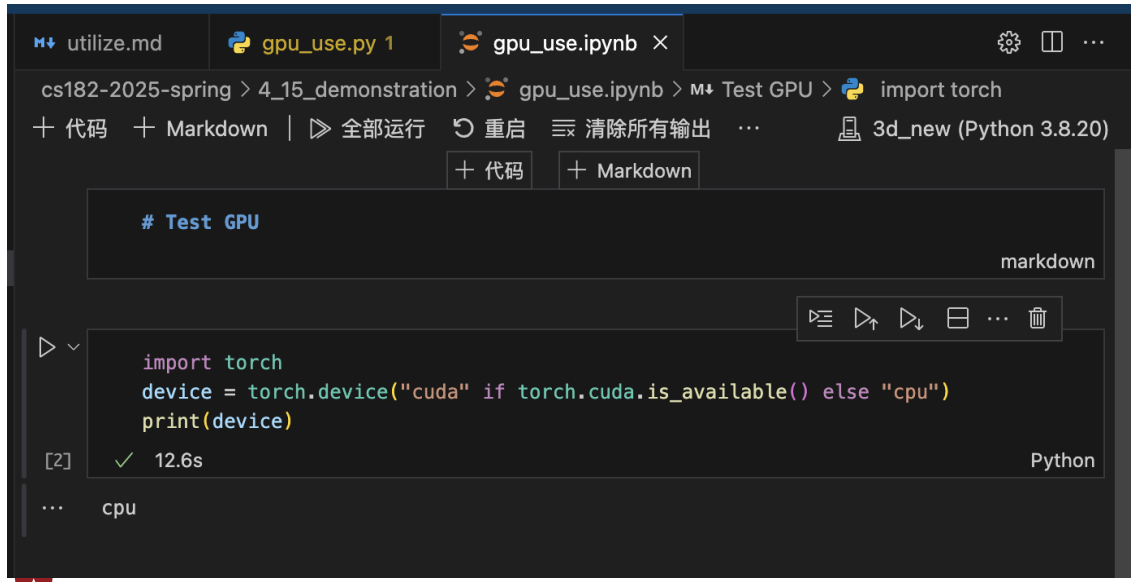


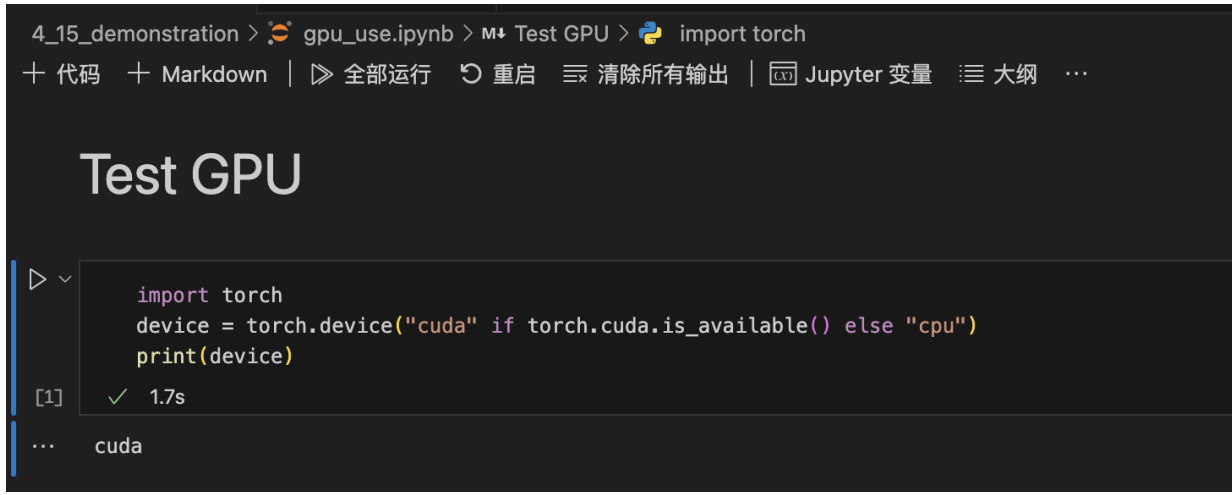
立志成才报国裕民

Jupter Notebook - how to use GPU

上海科技大学
ShanghaiTech University

- Can not ssh gpu through the jupter notebook code block
- ssh -J zhang3@10.15.89.191:22112 zhang3@sist_gpu70

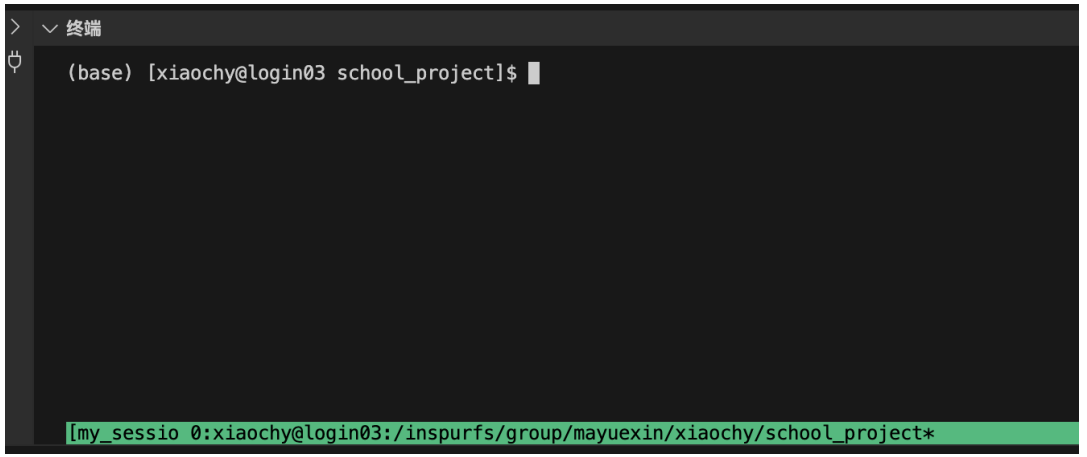




Tmux

- Keep processes running on servers after SSH logout.
- Already installed on login node, no installation on compute node
- Some commands:
  - tmux ls
  - tmux new-session -s my_session_name
  - tmux attach-session -t my_session_name

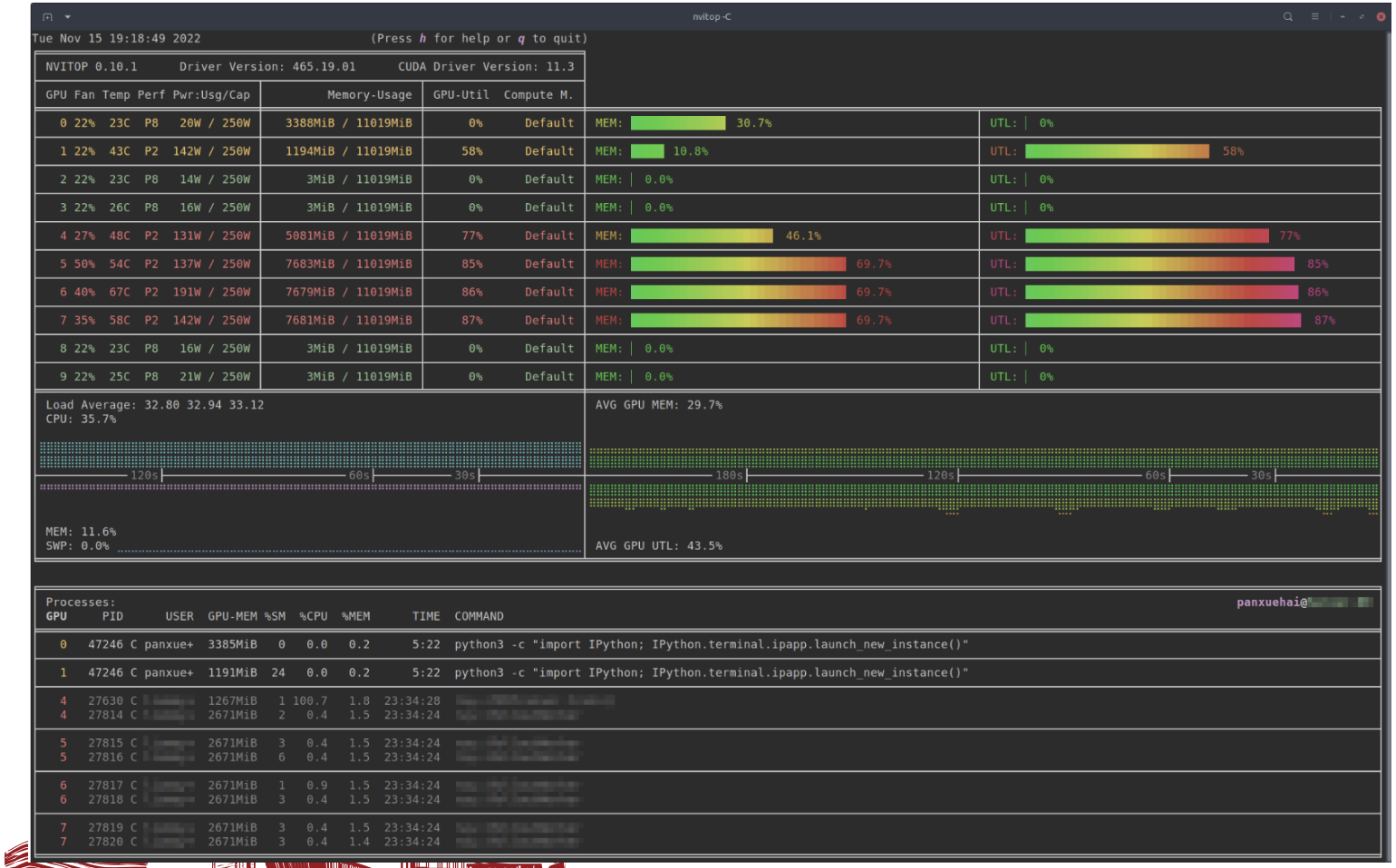
```
(base) [xiaochy@login03 school_project]$ tmux ls
2080_1: 1 windows (created Mon Mar  3 09:06:28 2025) [80x29]
2080_2: 1 windows (created Mon Mar  3 09:13:03 2025) [80x29]
2080_3: 1 windows (created Mon Mar  3 09:13:40 2025) [80x29]
2080_4: 1 windows (created Mon Mar  3 09:14:22 2025) [80x29]
2080_5: 1 windows (created Mon Mar  3 09:15:14 2025) [80x29]
2080_6: 1 windows (created Mon Mar  3 09:16:20 2025) [80x29]
2_28_h=8_a=4: 1 windows (created Fri Feb 28 09:55:45 2025) [141x7]
eval_3_1_v=3: 1 windows (created Sun Mar  2 17:12:17 2025) [80x29]
eval_3_1_v=4: 1 windows (created Sun Mar  2 17:13:00 2025) [80x29]
force_08: 1 windows (created Tue Mar  4 10:44:17 2025) [141x7]
force_gpu_3_3: 1 windows (created Sun Mar  2 23:35:25 2025) [80x29]
```



nvitop

上海科技大学
ShanghaiTech University

- An interactive NVIDIA-GPU process viewer and beyond, the one-stop solution for GPU process management.



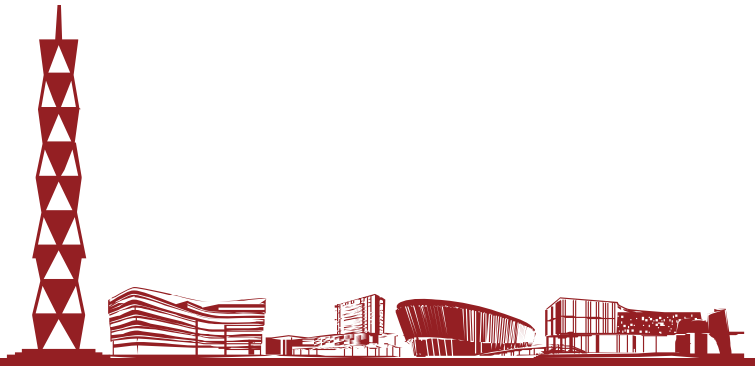
https://github.com/XuehaiPan/nvitop

Demonstration

Contact

上海科技大学
ShanghaiTech University

- Piazza
- Email: xiaochy@shanghaitech.edu.cn
- Markdown and video will be uploaded to Piazza soon!



立志成才报国裕民