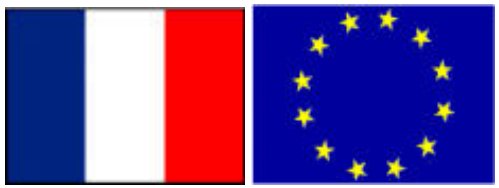


Big Data – TP1 Part 0

Connection to the DCE clusters of CentraleSupelec using *dcejs* or *ssh* or *vscode*

(Data Center for Education)

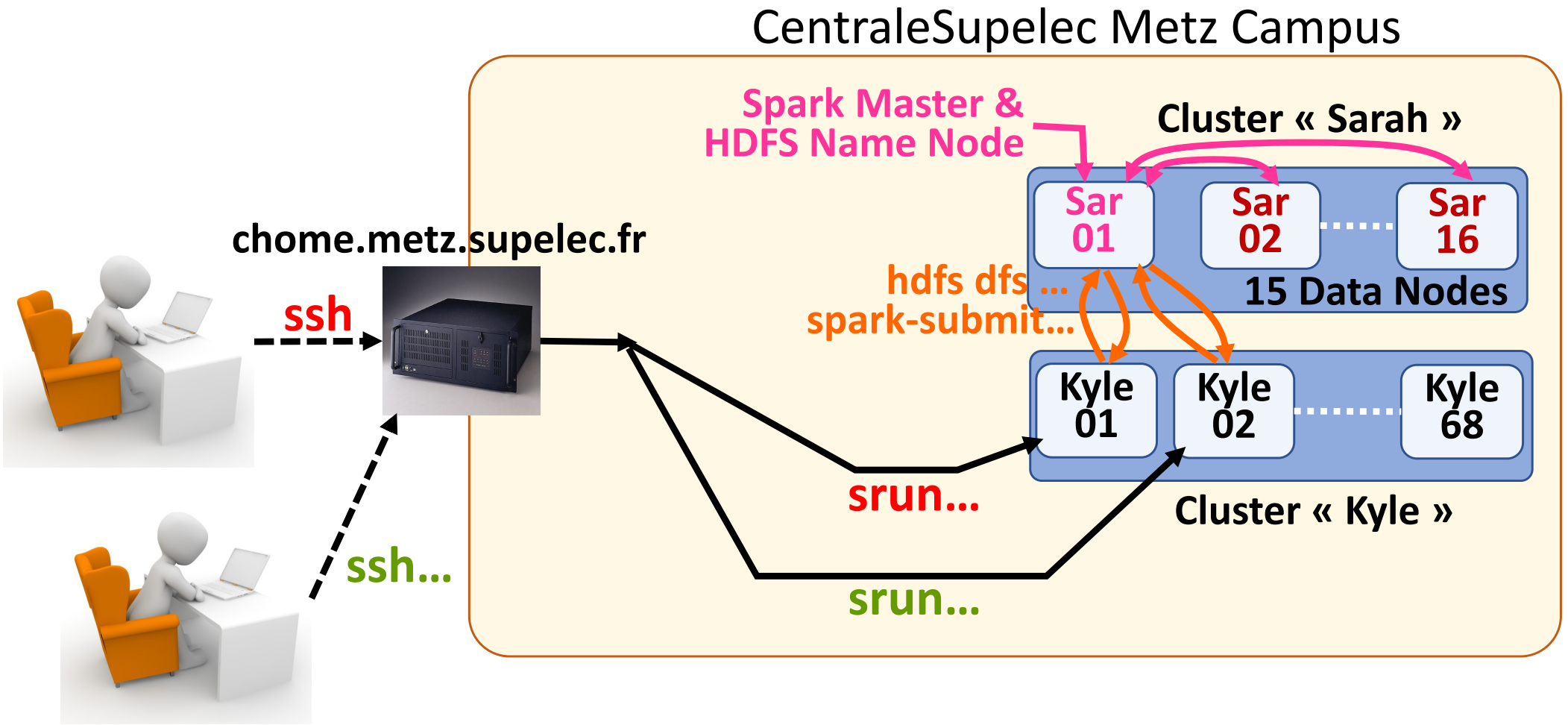
Stéphane Vialle **&** **Gianluca Quercini**

POLYTECH
PARIS-SACLAY

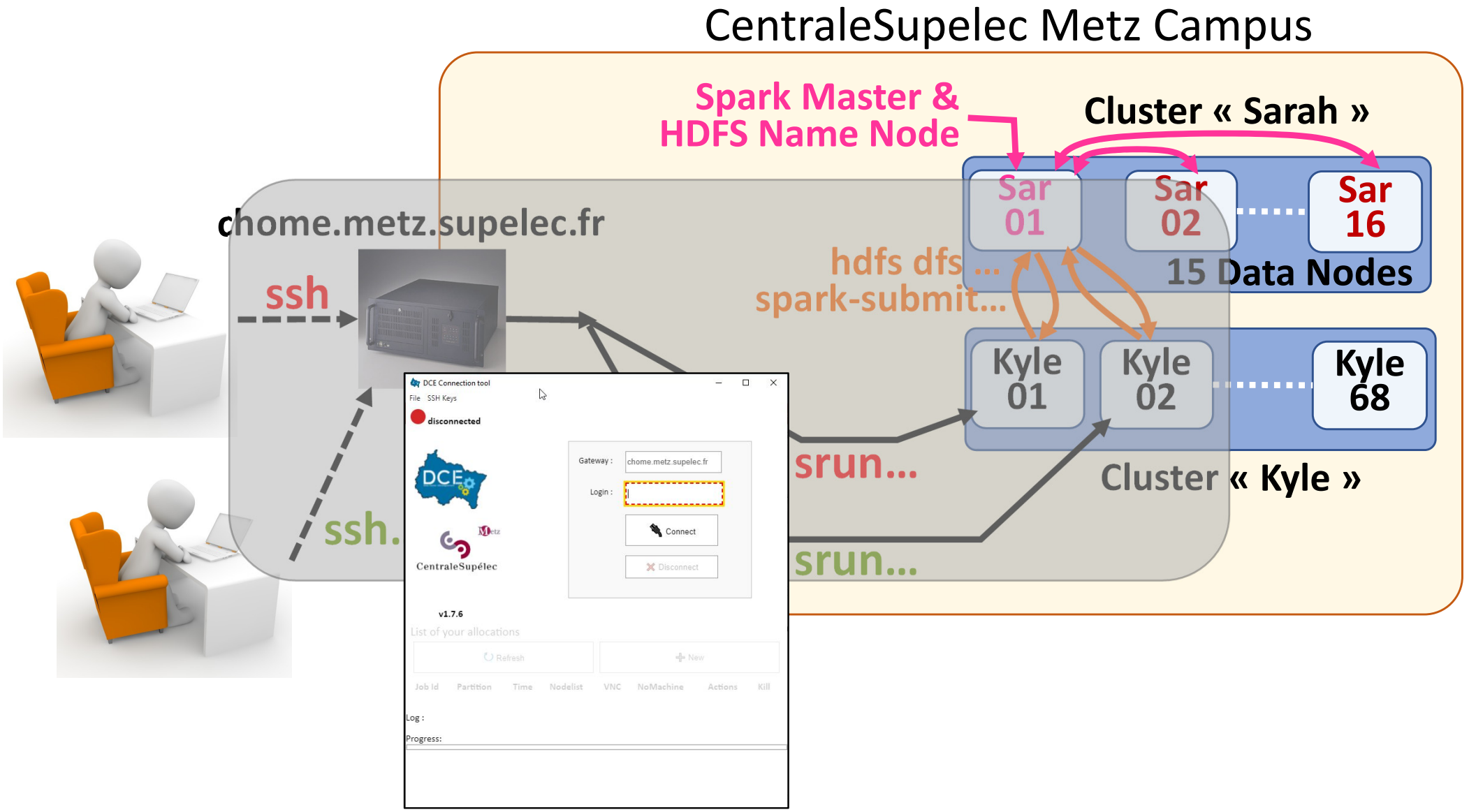
Using Spark cluster of CentraleSupelec DCE

- **DCE architecture and access with dcejs**
- DCE architecture and access with ssh
- DCE access with vscode
- How to kill a zombie session

CPU clusters

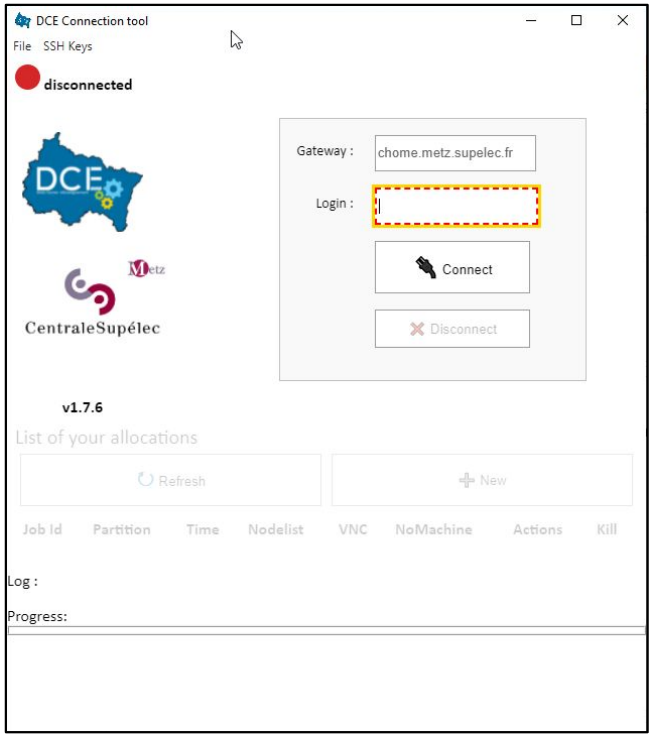


CPU clusters



CPU clusters

- Look at https://tutos.metz.centralesupelec.fr
- Look at **dcejs**
- Download & install the version for your system (Windows, Linux, Mac)



- Install a VNC client/viewer

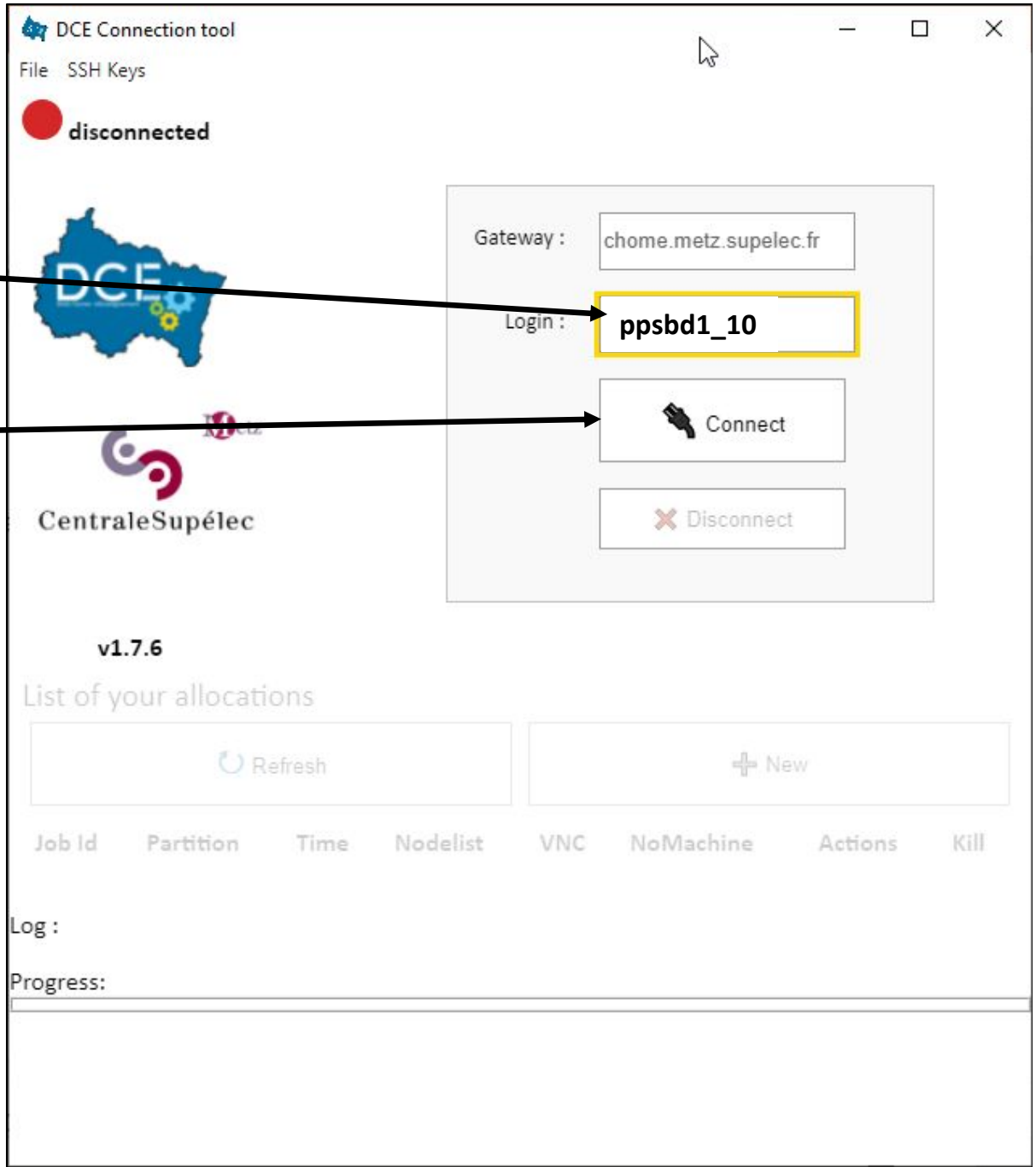
  Ex: TigerVNC on Windows/Mac

  Download vncviewer64-1.12.0.exe (or vncviewer-1.12.0.exe)

  (on https://sourceforge.net/projects/tigervnc/files/stable/1.12.0/)

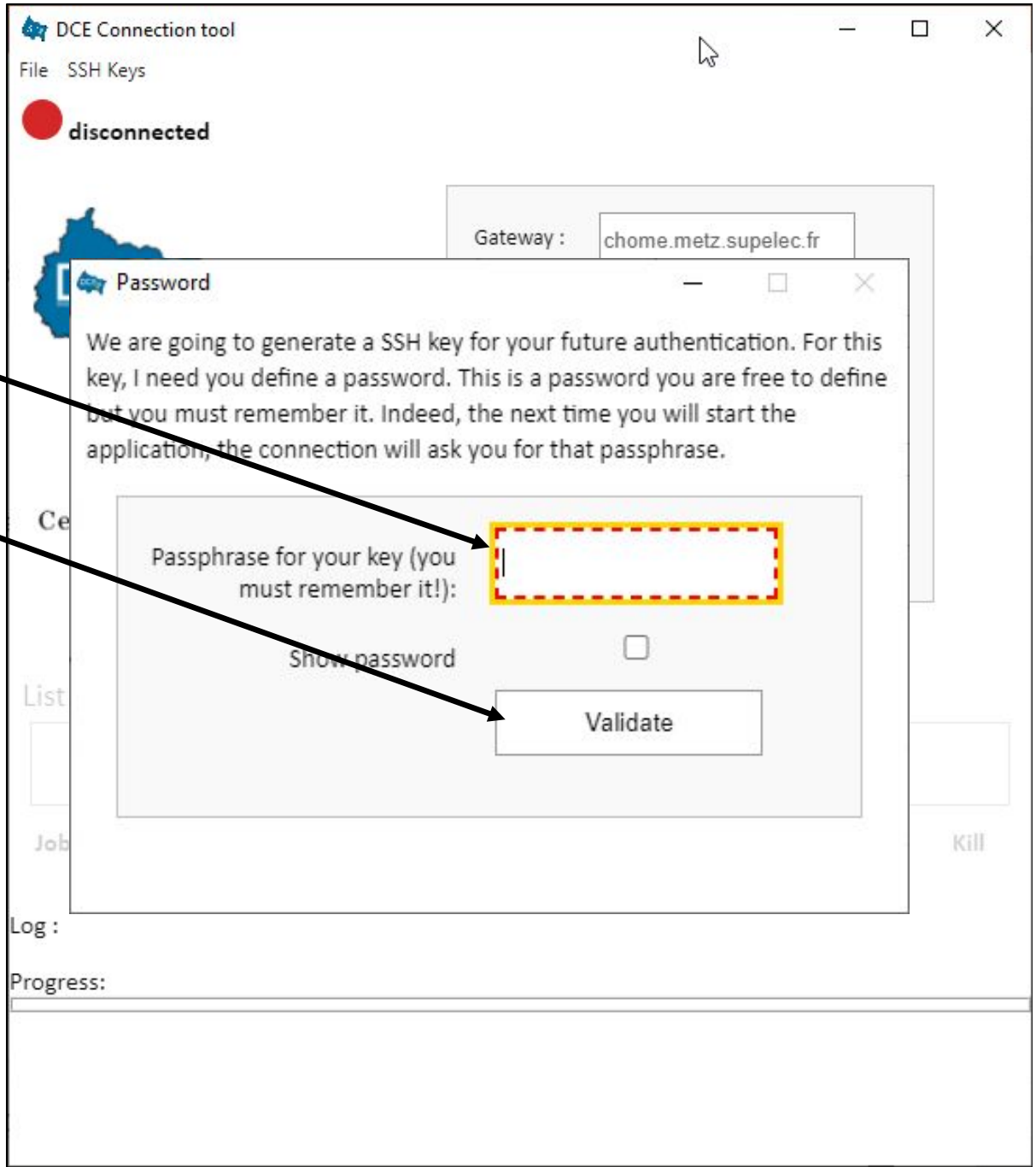
CPU clusters

1. Launch **dcejs**
   - Write your ppsbd*y*_*xx* login
   - Click on « Connect »



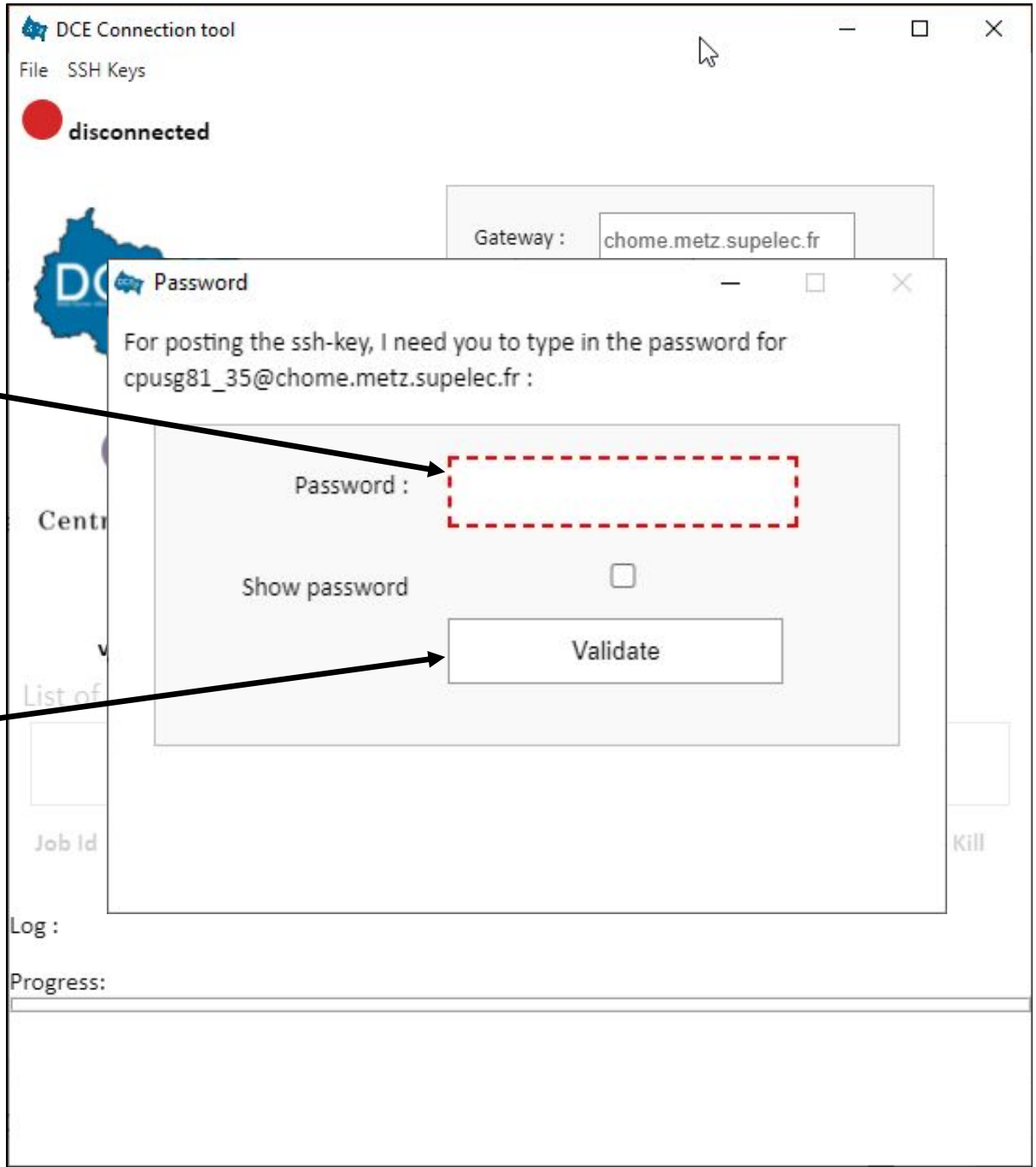
CPU clusters

2. Launch **dcejs**
   - ……
   - Enter a passphrase (choose a basic one)
   - Click on « Validate »



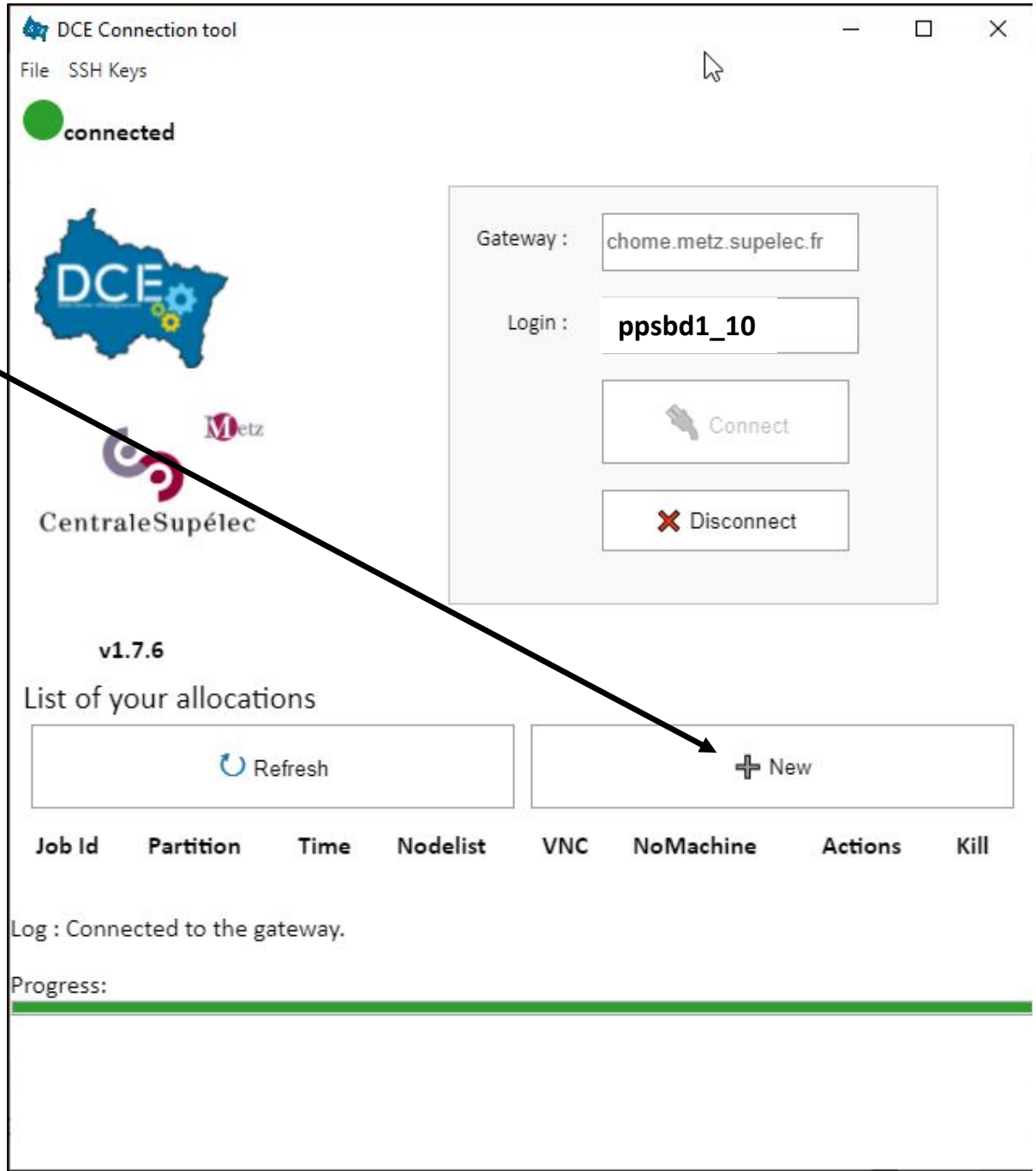
CPU clusters

3. Launch **dcejs**
   - ……
   - During the first connexion to the DCE you will need to enter the passwd of your DCE login
   - Click on « Validate »



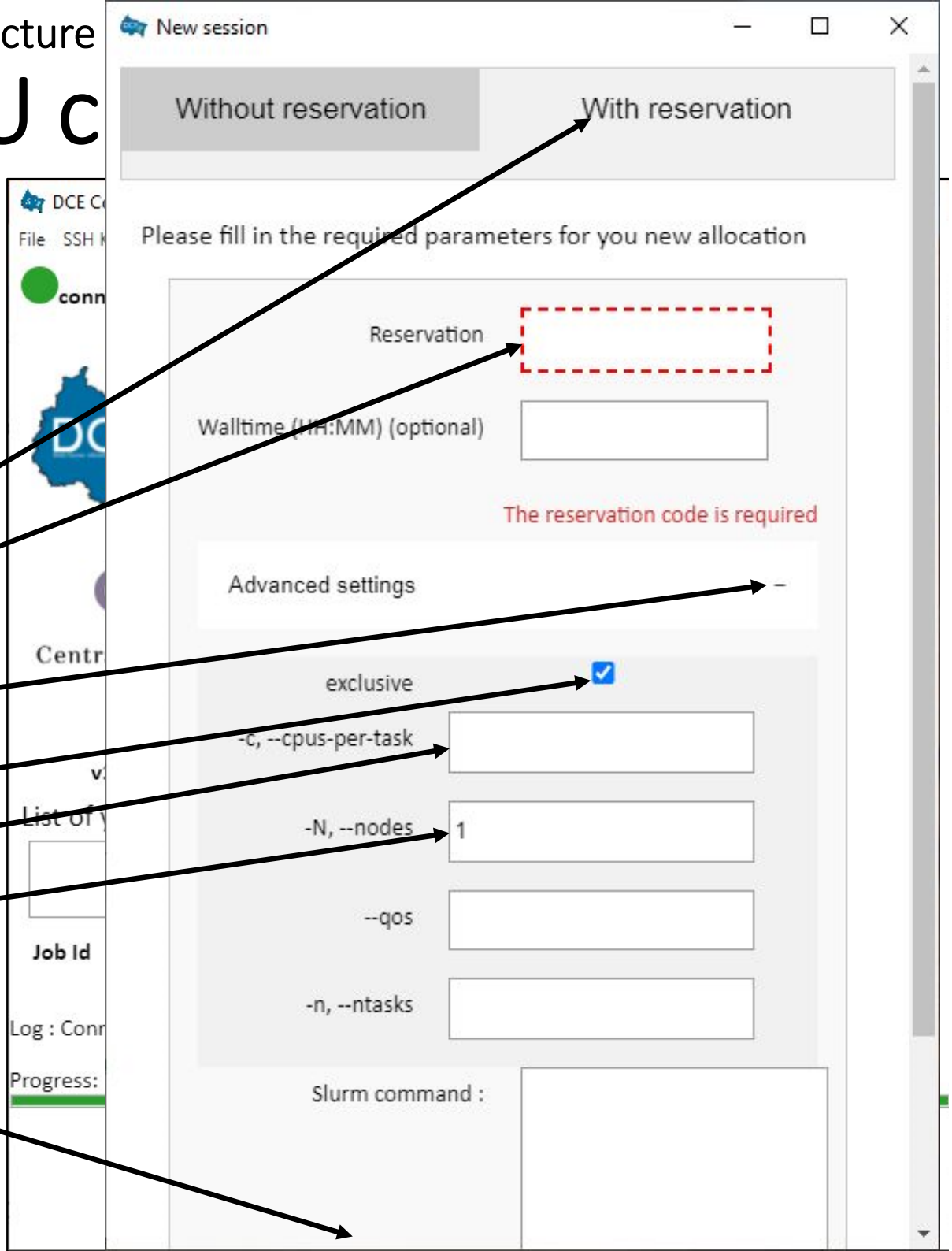
CPU clusters

4. Launch **dcejs**
   - ……
   - Click on « + New » to start to allocate some computing resources



CPU c

5. Launch **dcejs** during the lab
   - ……
   - Click on « With res… »
   - Enter a reservation code (*ask to the Prof)*
   - Open settings
   - Uncheck the box
   - Enter: 2
   - Enter: 1
   - Click on « Validate »

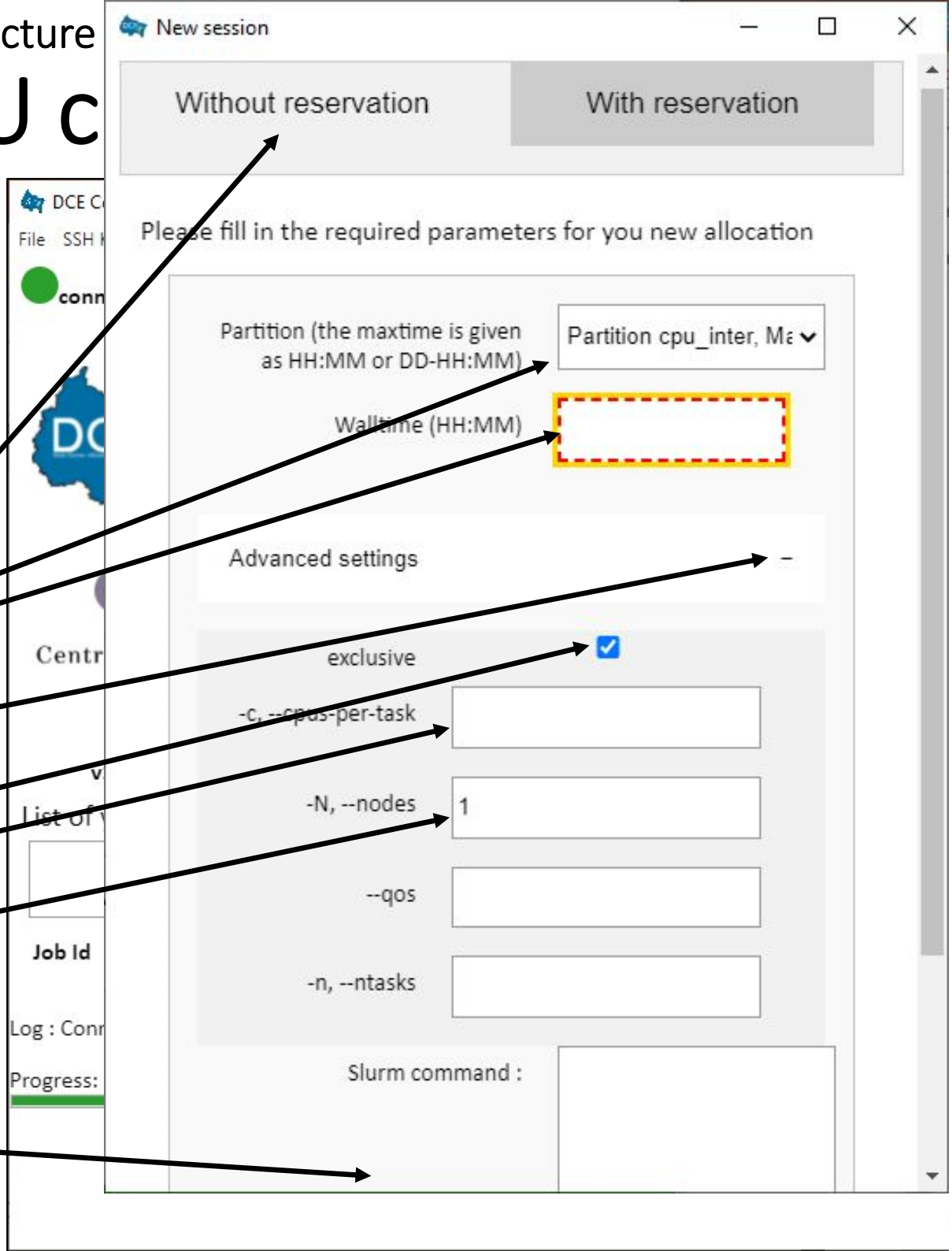


CPU c

5. Launch **dcejs** after the lab

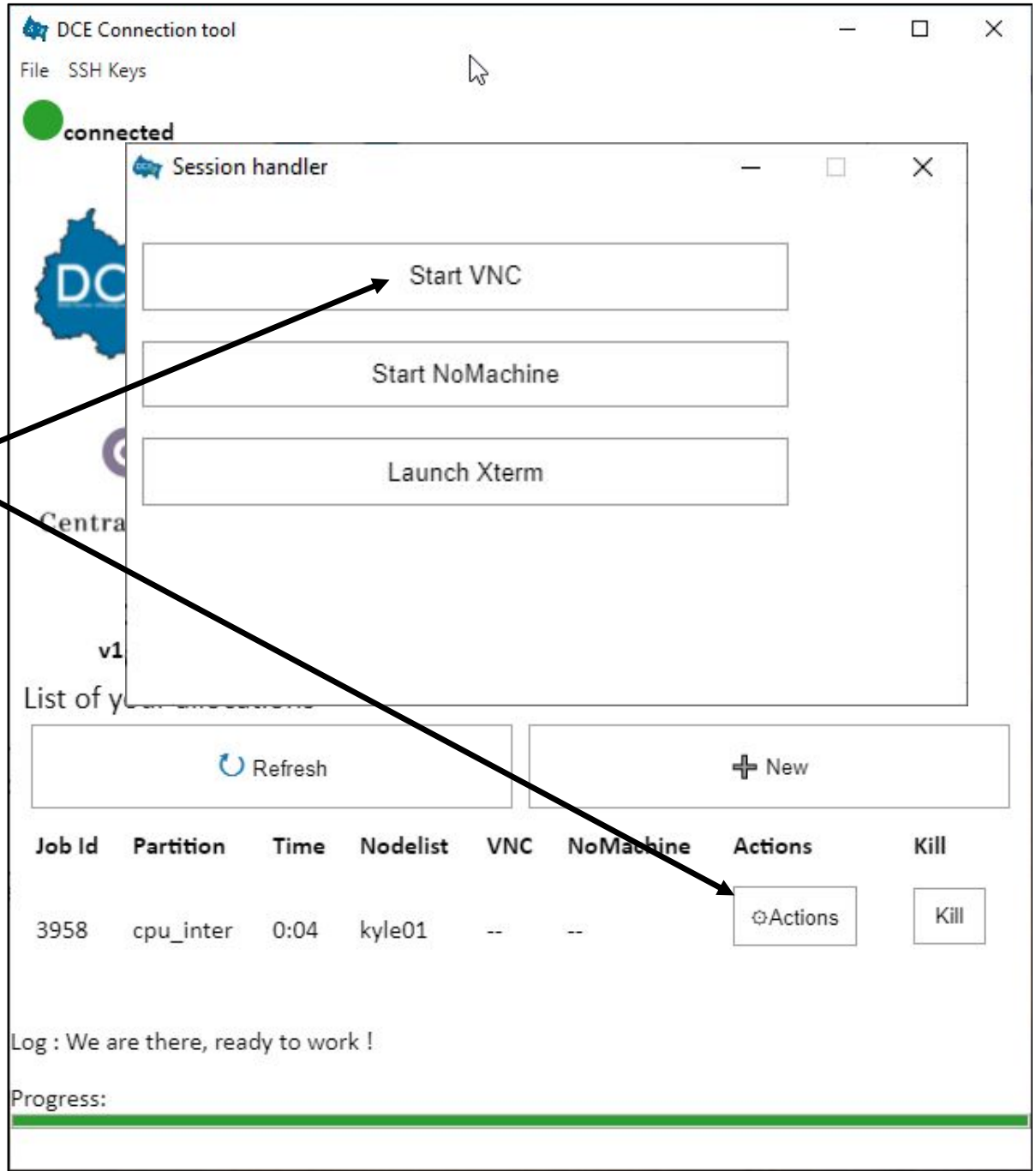
- ……
- Click on « Without… »
- Select partition cpu_tp
- Specify 2:0 (2h)
- Open settings
- Uncheck the box
- Enter: 2
- Enter: 1

- Click on « Validate »



CPU clusters

6. Launch **dcejs**
   - ……
   - Click on « Actions »
   - Click on « Start VNC »

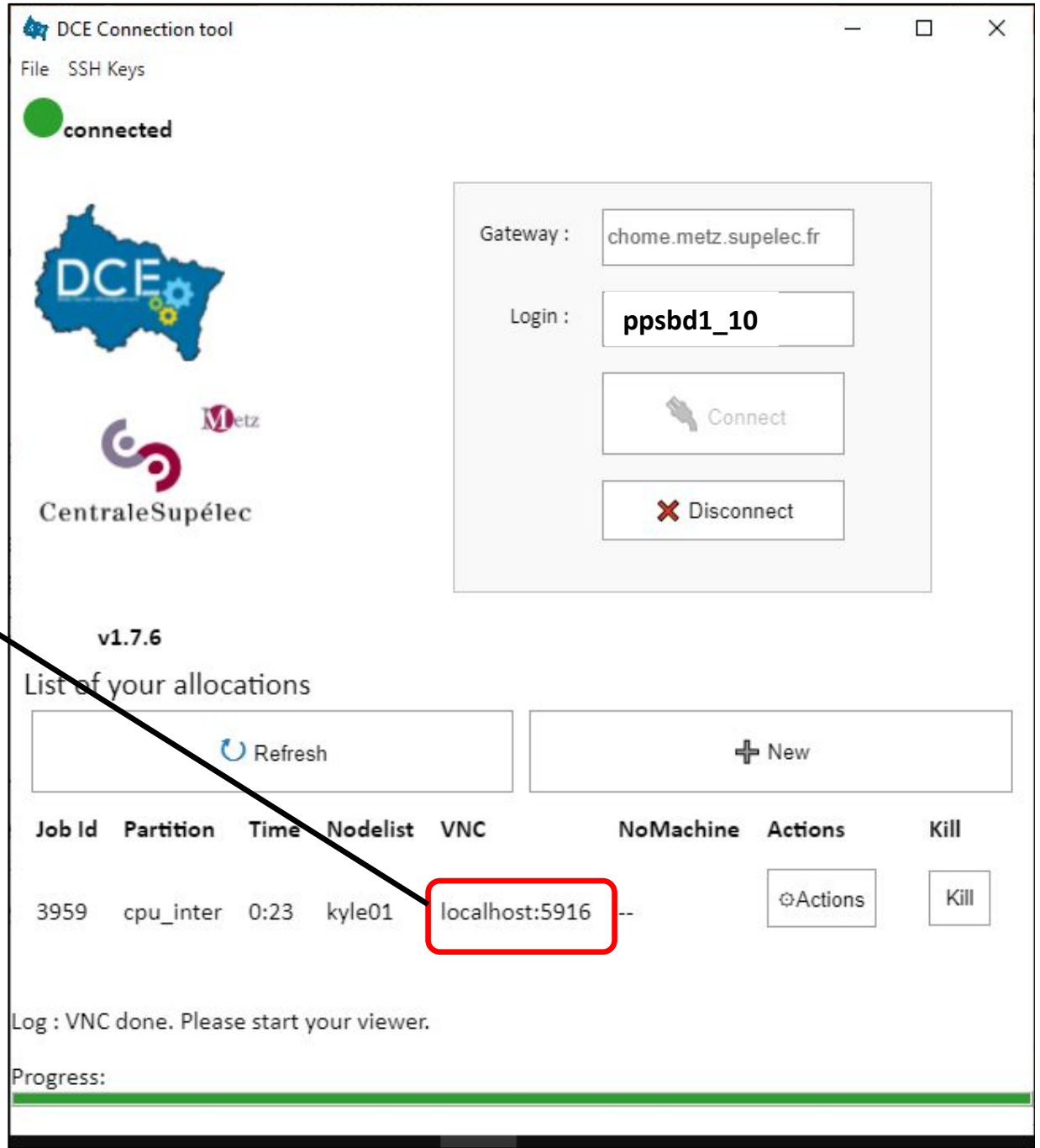


CPU clusters

7. Launch **dcejs**
   - ……
   - Get the local port number

     Ex: 5916
   - Launch your VNC client with all default options

     Ex: TigerVNC

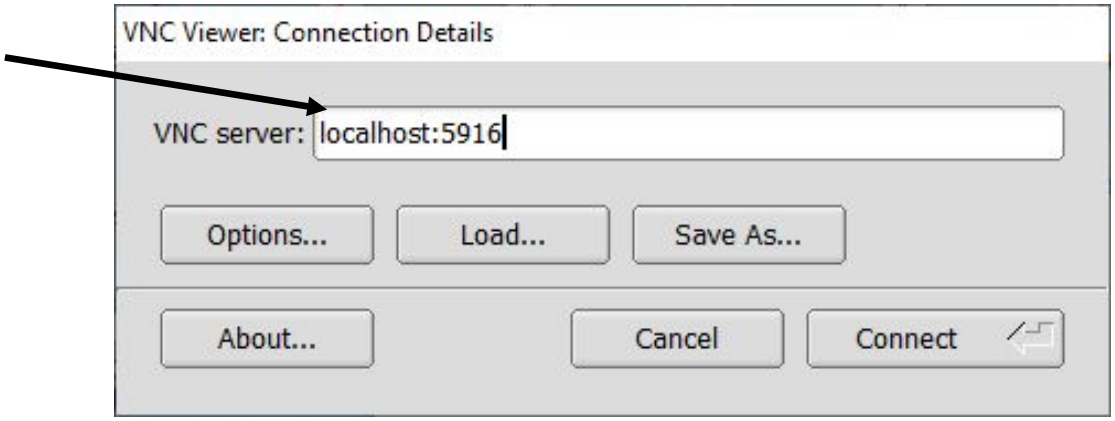


CPU clusters

8. On windows:
   - Launch your VNC client with all default options

     (ex: TigerVNC on Windows)

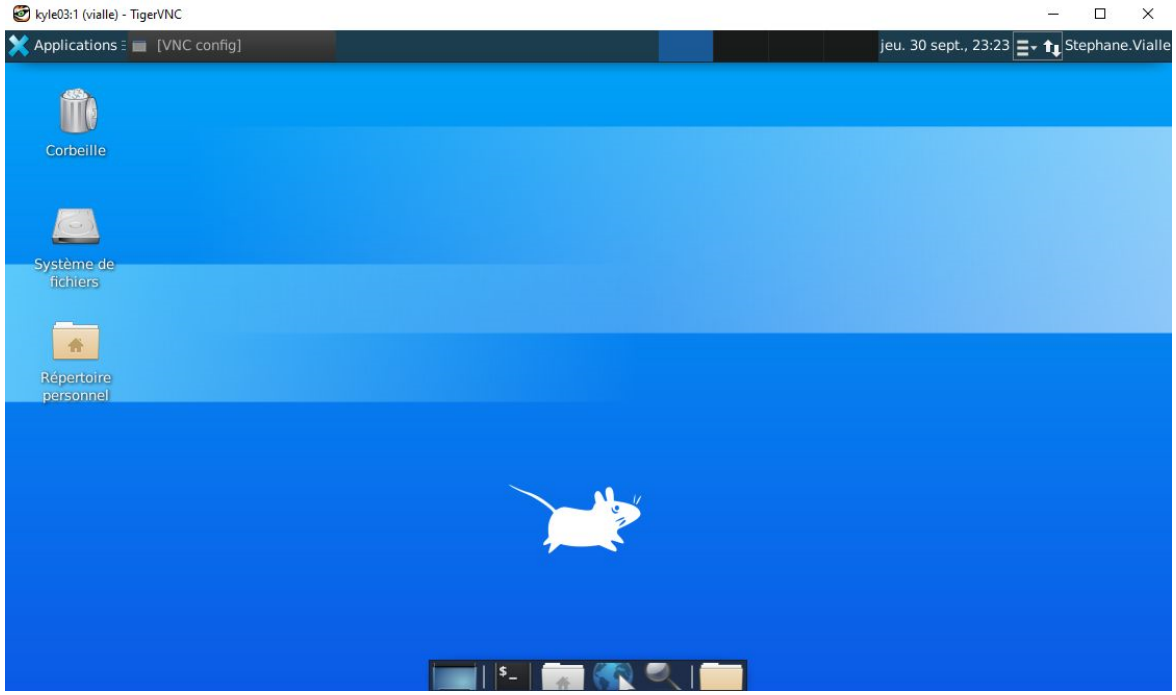
   - Enter the port number returned by dcejs



8. On Linux & Mac :
   - It should be possible to just click on the port number in the dcejs window.

CPU clusters

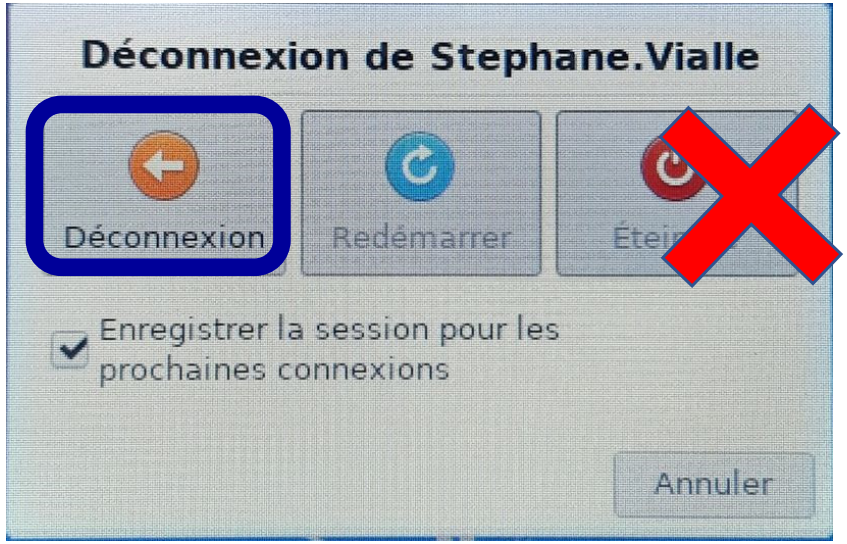
9. The desktop of the remote DCE machine appears
   - You can launch a terminal, and an editor (code, xedit, …)



10. When you deconnect:

    NEVER shut down the machine!

    Use the disconnect button

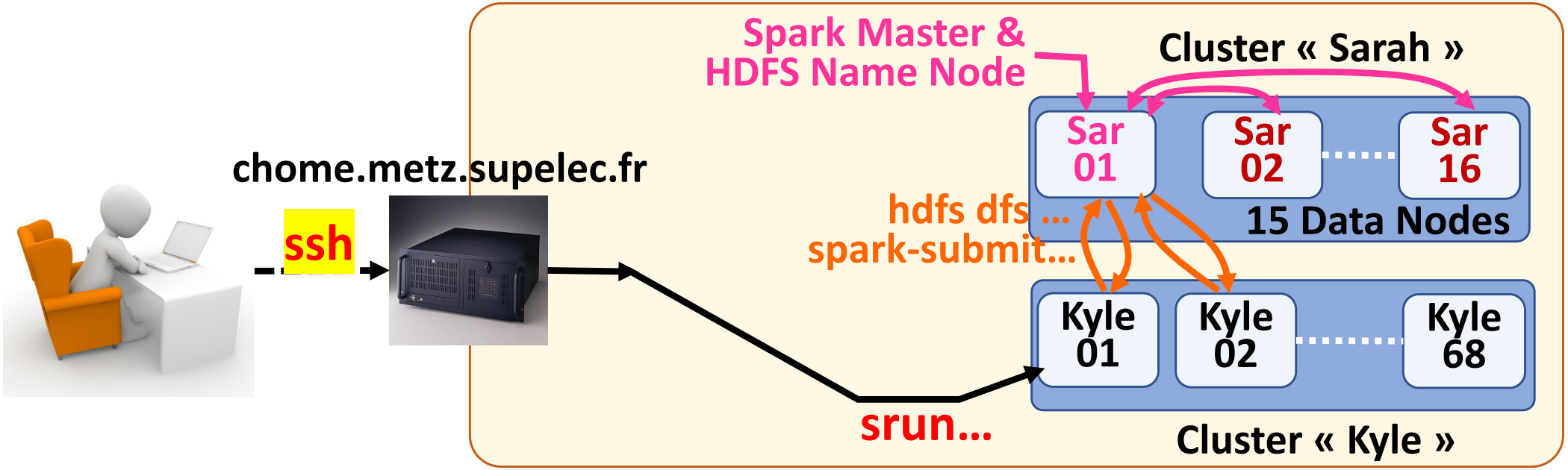


POLYTECH
PARIS-SACLAY

Using Spark cluster of CentraleSupelec DCE

- DCE architecture and access with dcejs
- **DCE architecture and access with ssh**
- DCE access with vscode
- How to kill a zombie session

Cluster connection commands



Linux/Mac `ssh chome.metz.supelec.fr -l ppsbd1_1`

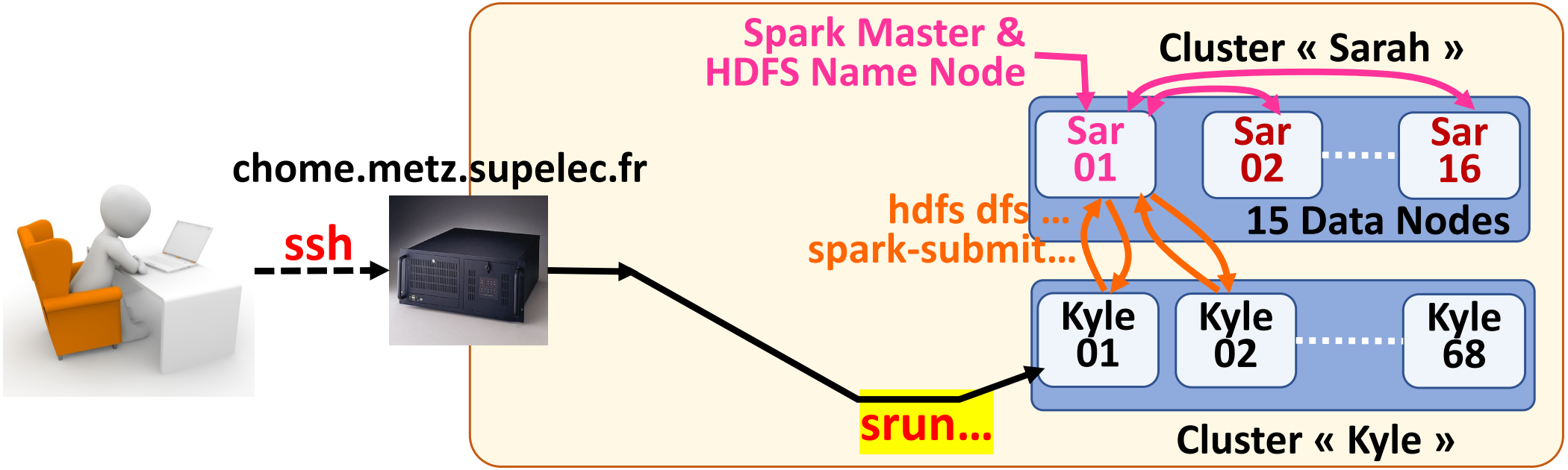
From ppsbd1_1
up to ppsbd1_36

Windows → run a "powershell" and then the above command

Old Windows →

- Download & Instal « putty »
- « Session » menu : **phome.metz.supelec.fr**
  connection type : **ssh** (port 22)
- « Connection » menu: set **Enable TCP keepalives**
  set **30s** between keepalives

Cluster connection commands



On *chome* DURING the lab:

```
srun  --reservation=myCode
      -N 1 --cpus-per-task=2
      --pty bash
```

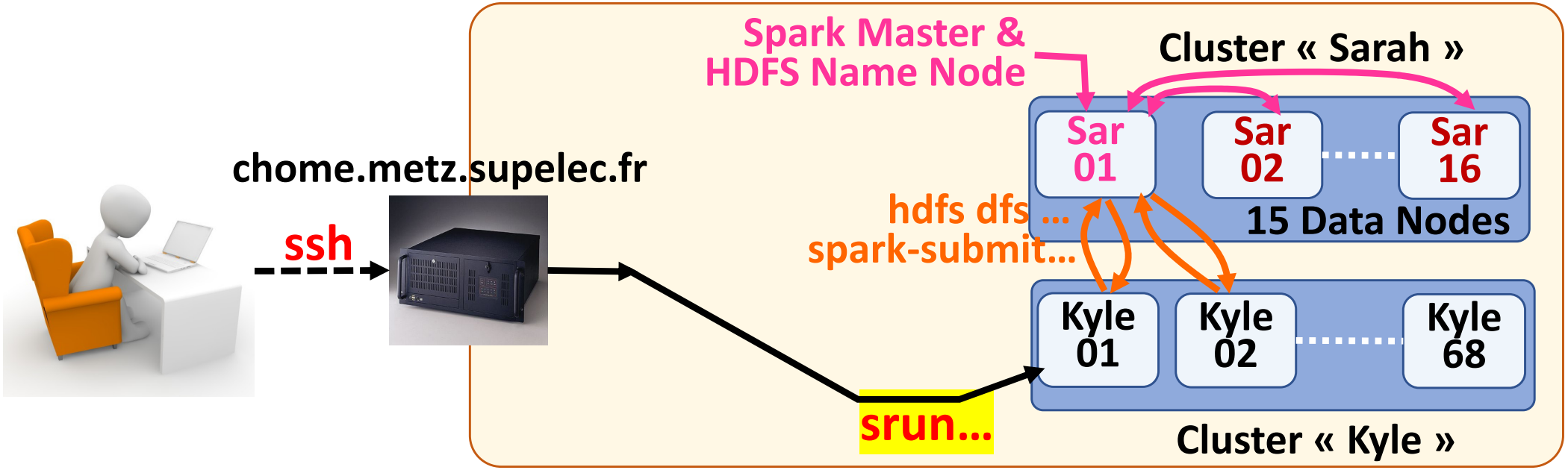
Ask to the teacher

Allocate 2 logical cores on 1 node

Run an interactive session

Write on ONE line!

Cluster connection commands



On *chome* AFTER the lab:

```
srun   -p cpu_tp
       -N 1 --cpus-per-task=2
       --pty bash
```

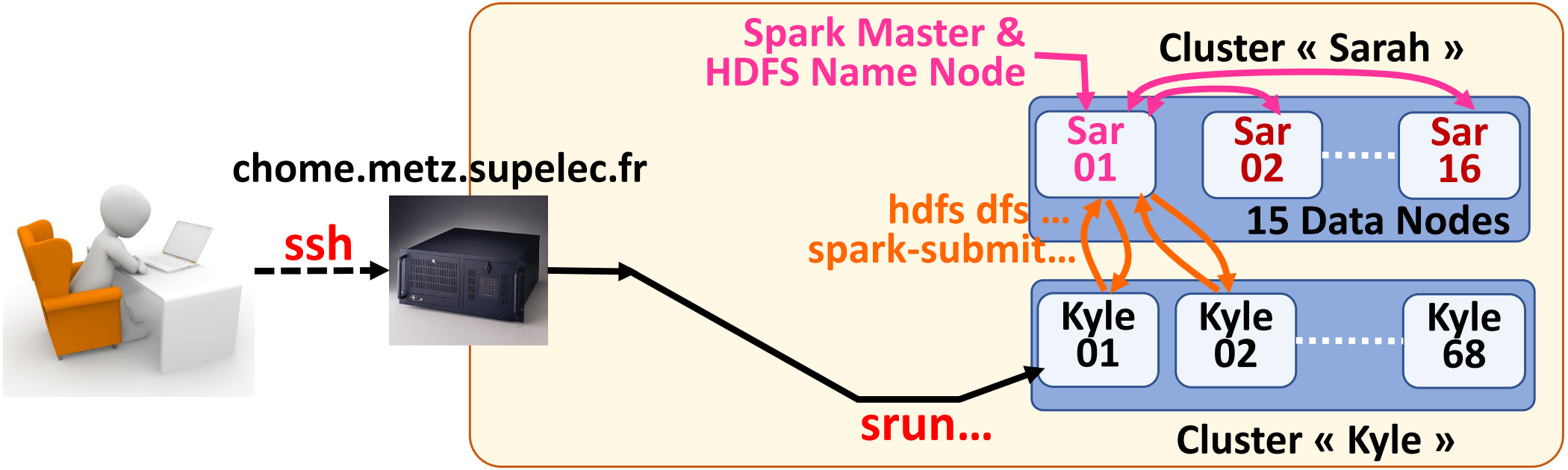
Partition to use

Allocate 2 logical cores on 1 node

Run an interactive session

Write on ONE line!

Cluster connection commands



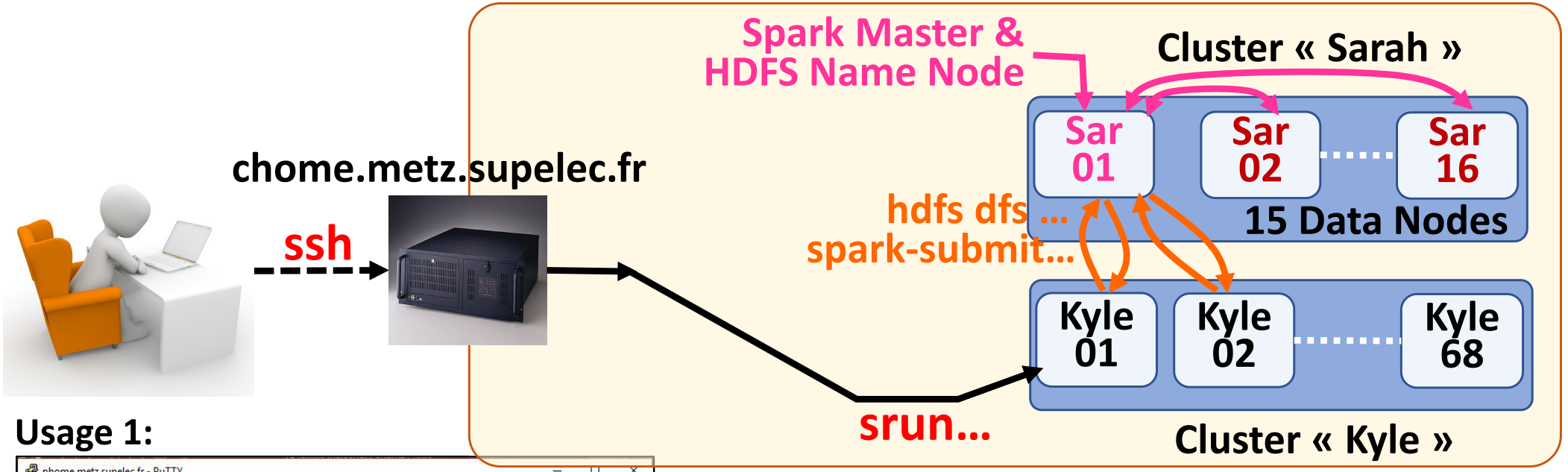
On *the cluster node*:

`sinfo -l` → information on partitions

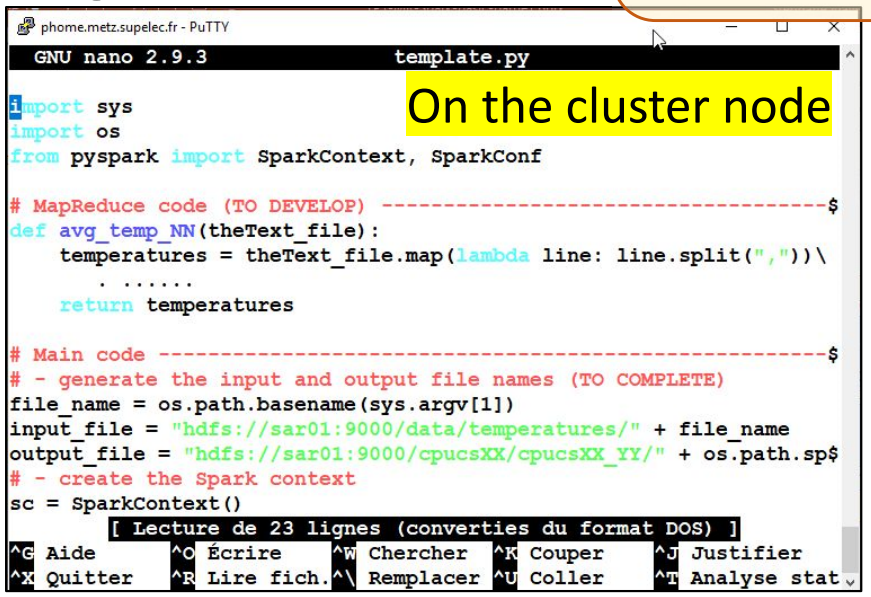
`squeue` → information on job queues

`scontrol show job` → information on running jobs

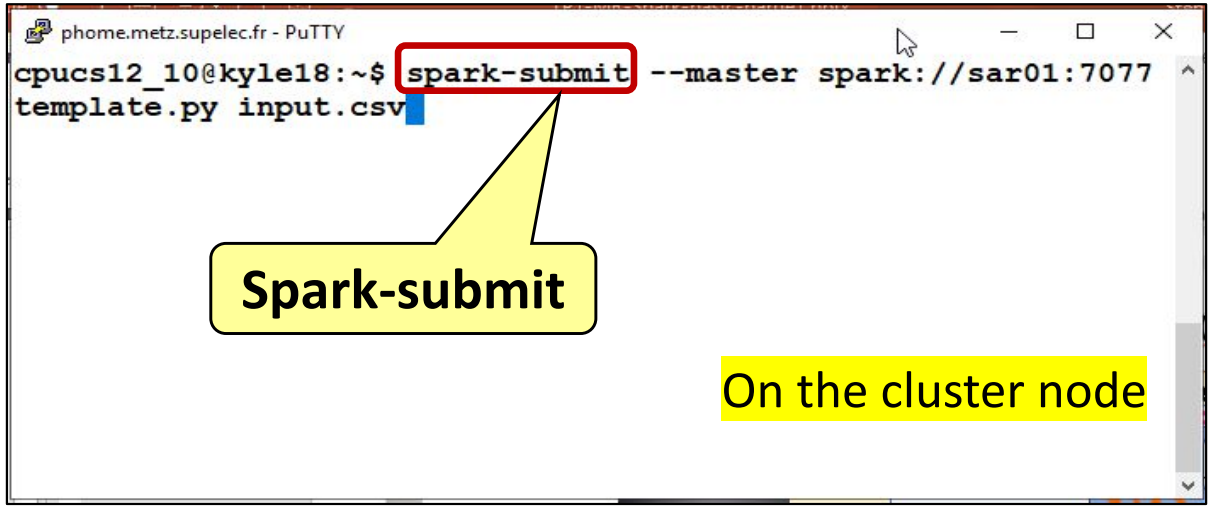
Edition of remote files



Usage 1:

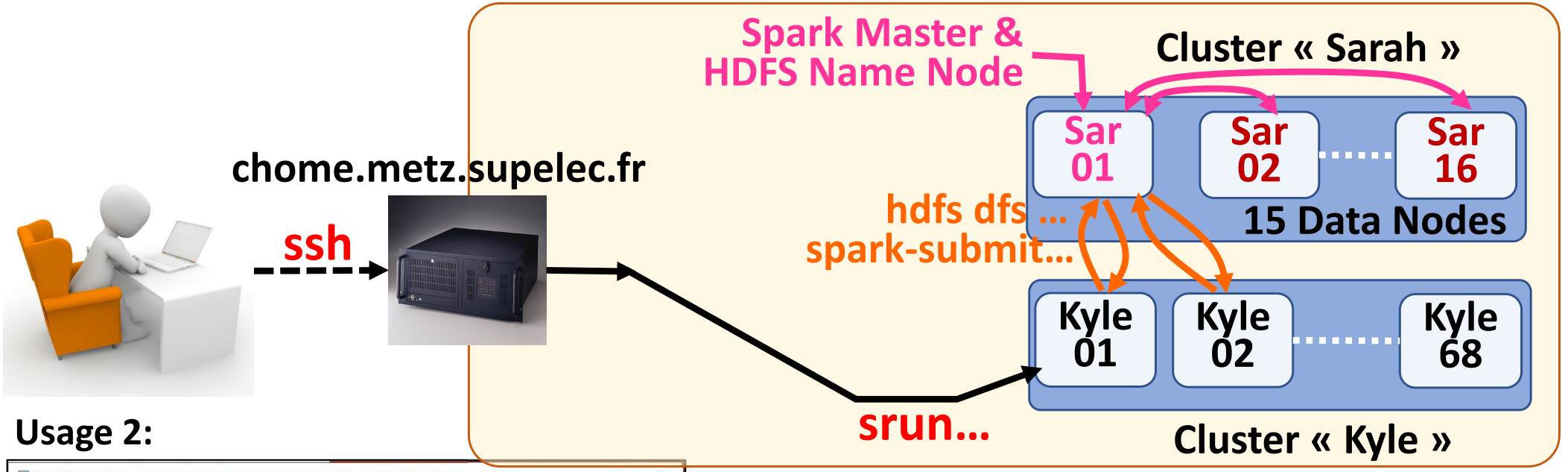


Python-Spark file: remote edition (ex: vi / nano editor)

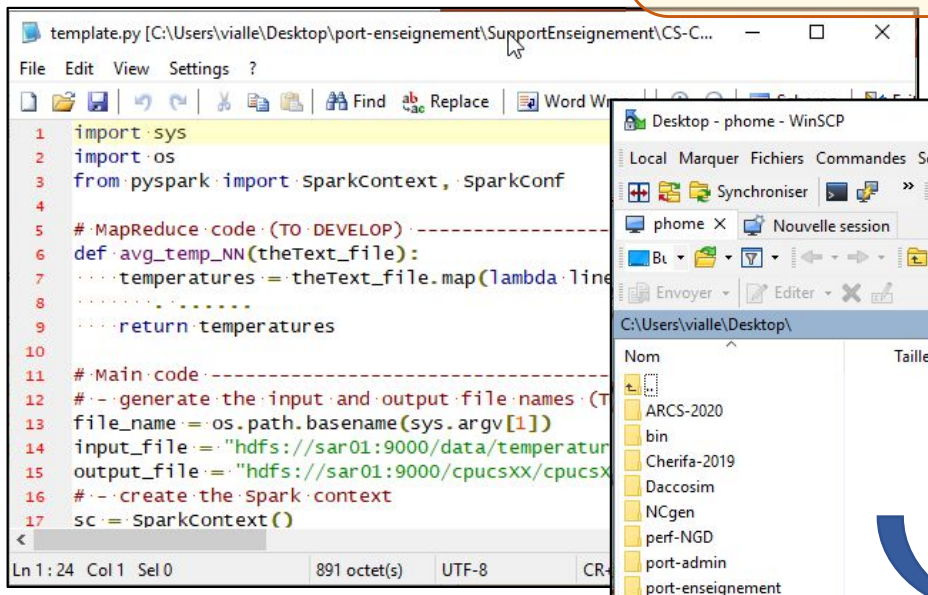


Distributed spark execution (remote spark-submit command)

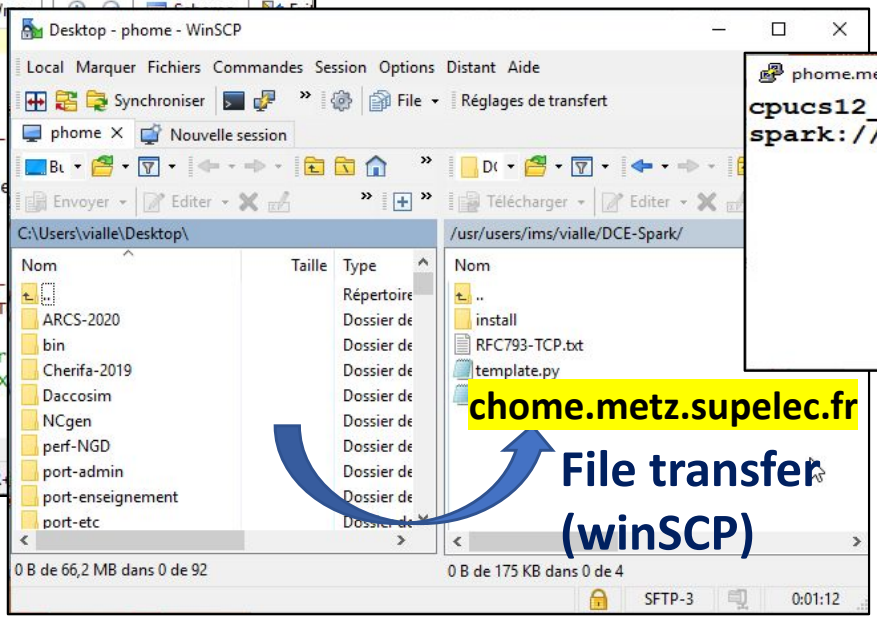
Transfer of remote files

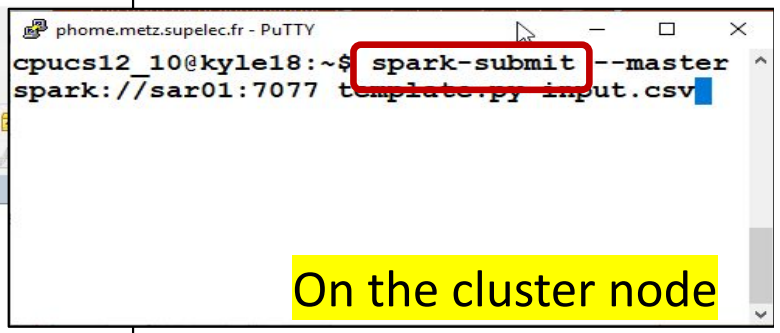


Usage 2:



Python Spark file: local edition





Remote distributed spark execution (spark-submit cmd)

POLYTECH PARIS-SACLAY

Using Spark cluster of CentraleSupelec DCE

- DCE architecture and access with dcejs
- DCE architecture and access with ssh
- **DCE access with vscode**
- How to kill a zombie session

Edition of remote files

Configuration and usage of vscode to reach the DCE:

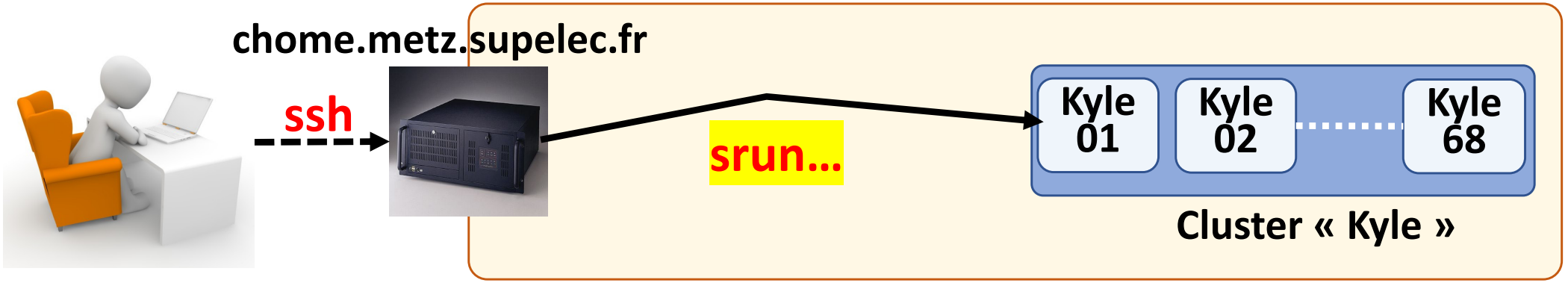
https://webtv.centralesupelec.fr/videos/how-to-connect-to-dce-with-visual-studio-code/

POLYTECH
PARIS-SACLAY

Using Spark cluster of CentraleSupelec DCE

- DCE architecture and access with dcejs
- DCE architecture and access with ssh
- DCE access with vscode
- **How to kill a zombie session**

DCE connection commands



On the *cluster node* or on *chome*:

`mysrun` → information on your running *srun --pty bash*

Ex: **cpu_vialle@chome:~$ mysrun**

506 bash RUNNING 11:08 4:30:00 1 kyle01 k1

`scancel <jobId>` → delete a *slurm job*

Ex: **cpu_vialle@chome:~$ scancel 506**

To remove a zombie job blocking new resource allocations!

POLYTECH
PARIS-SACLAY

Big Data – TP1 Part 0

Connection to the DCE of CentraleSupelec DCE using *dcejs* or *ssh* or *vscode*

(Data Center for Education)

Questions ?