Machine Learning for Networking

Big Data cluster environment

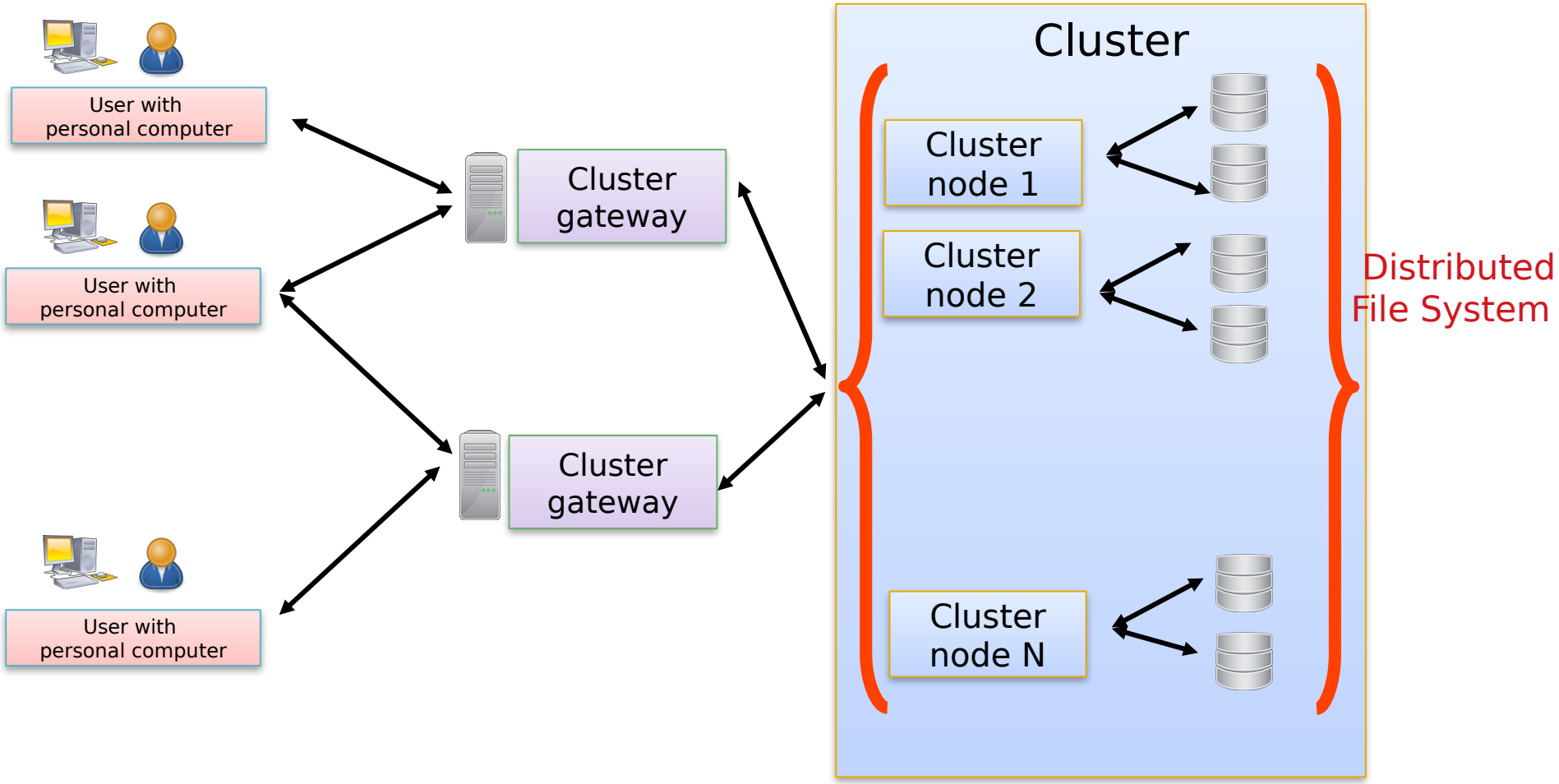
User with personal computer

User with personal computer

User with personal computer

Cluster

Big Data cluster environment



The BigData@Polito environment

- The BigData@Polito cluster has
  - A set of ~80 servers
  - Many **Access Gateway** servers

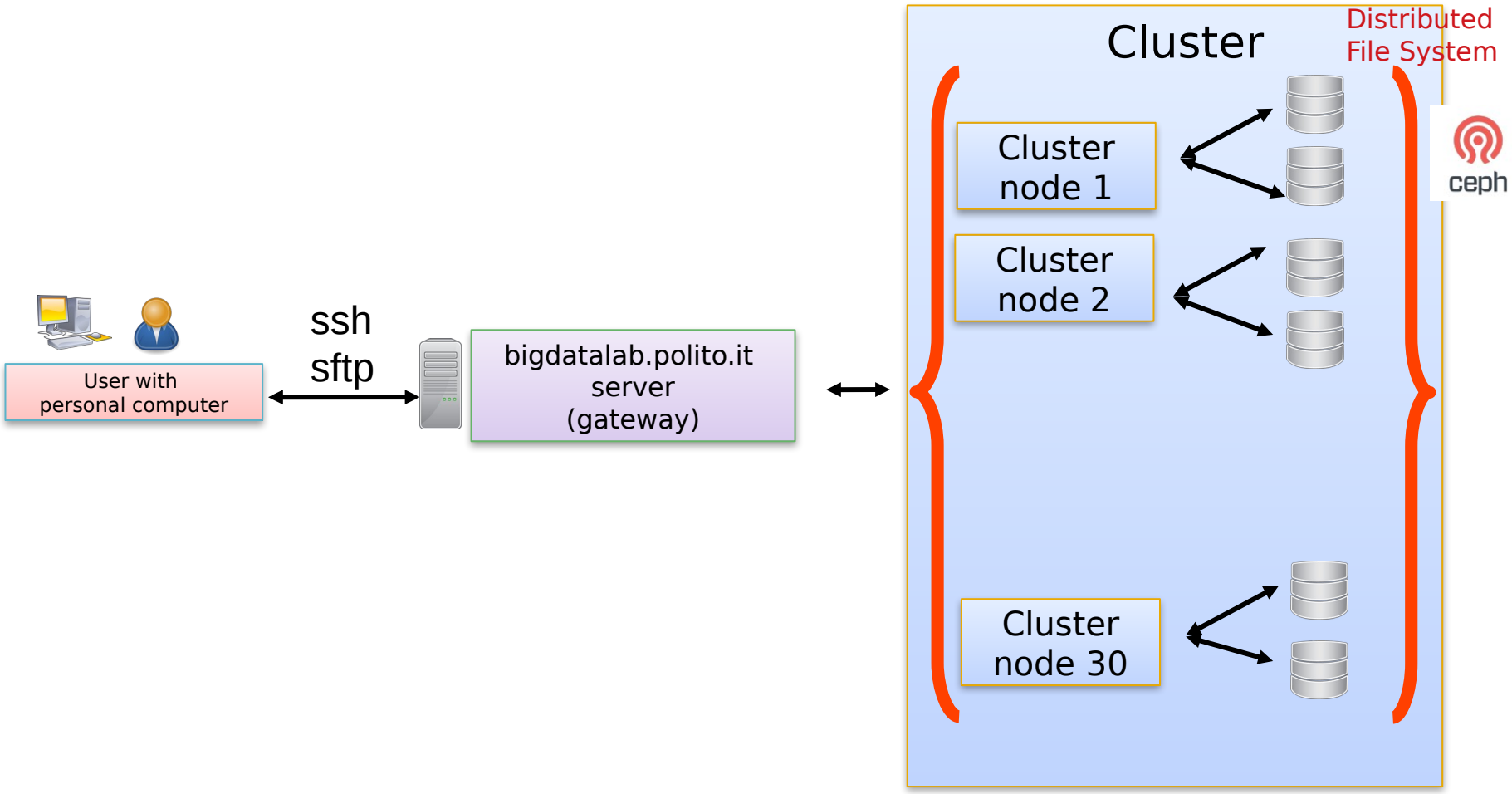
    1)bigdatalab.polito.it

    2)jupyter.polito.it 

The BigData@Polito environment

- Useful links:
  - https://smartdata.polito.it/computing-facilities/
  - https://smartdata.polito.it/cluster-access-instructions/

The BigData@Polito environment



Ssh/sftp access to our BigData cluster

bigdatalab.polito.it is an **Access Gateway** server used to interact with the cluster. Connecting to it through ssh protocol you can:

- Use command line
- ~~Execute python applications~~ – better not
- Transparently interact with the distributed file system (transfer files,...)

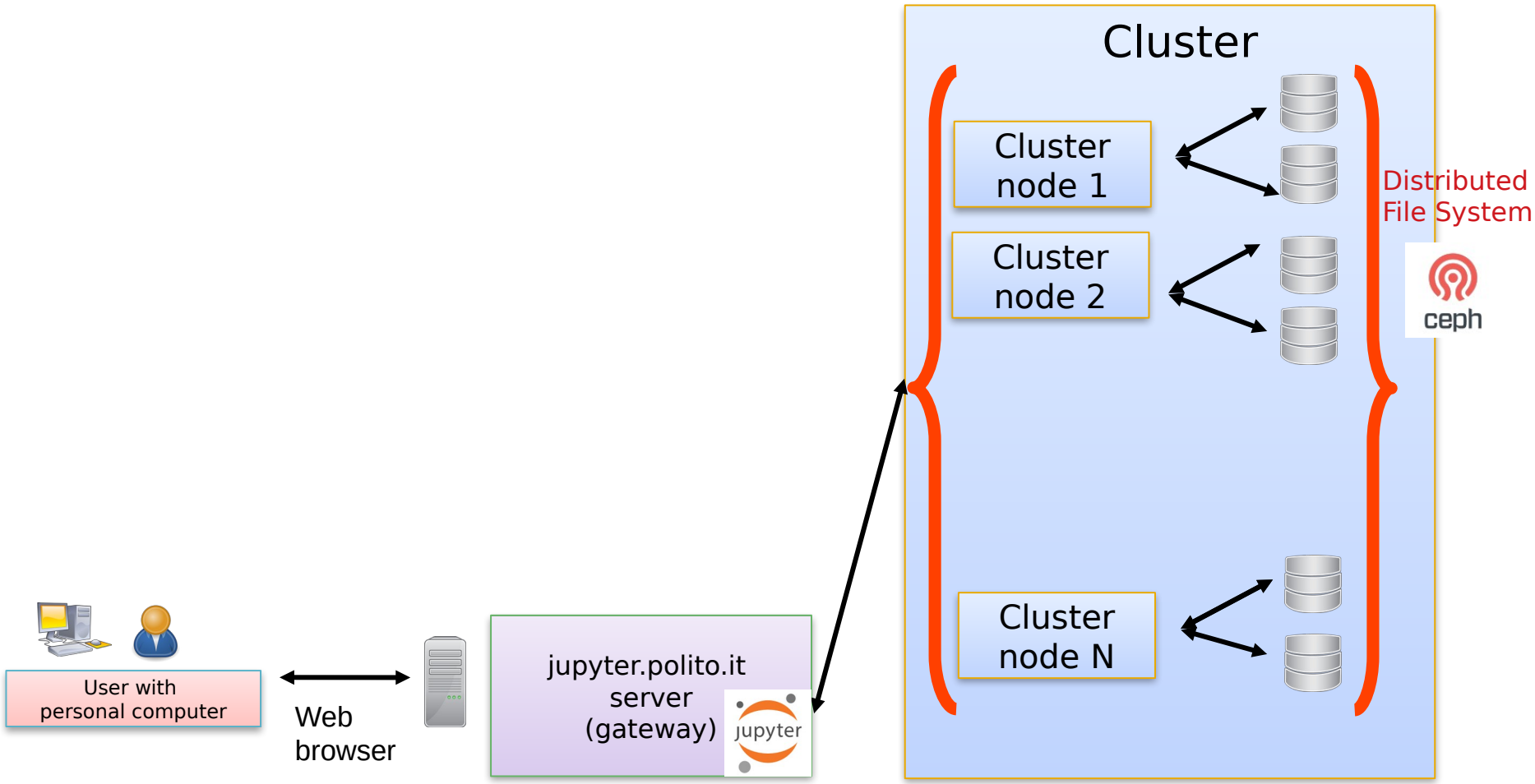
Jupyter notebooks

Jupyter notebook - browser-based interactive IDE. JSON document, containing an ordered list of input/output cells which can contain code, text (using Markdown), ...

Jupyter notebooks

- Can execute just part of a Python code
- Allows the user to include formatted text
- Can mix visualization of the results with comments and code
- Notebooks saved in a way that lets other people open them and execute the code on their own systems
- Ideal for creating reports and doing data science experiments

The BigData@Polito environment



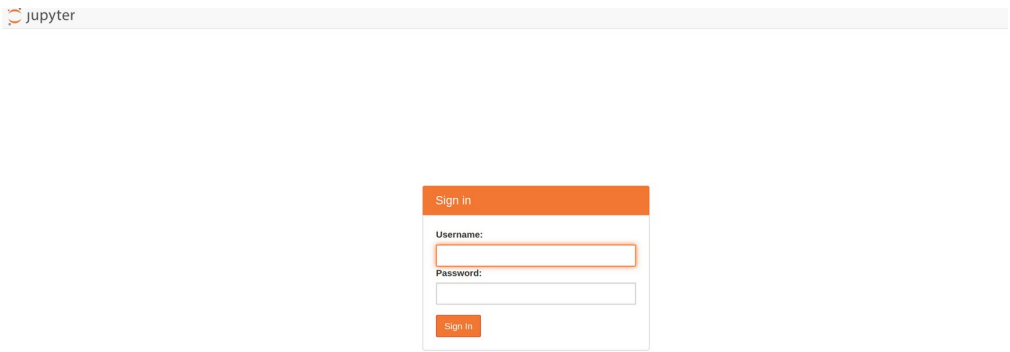
Jupyter web interface to our BigData cluster

jupyter.polito.it is an **Access Gateway** server used to interact with the cluster. Connecting to it through ssh protocol you can:

- **Execute interactive Jupyter notebook**
- Use command line
- **Execute python applications**
- Transparently interact with the distributed file system (transfer files,...)

Jupyter web interface for our BigData cluster

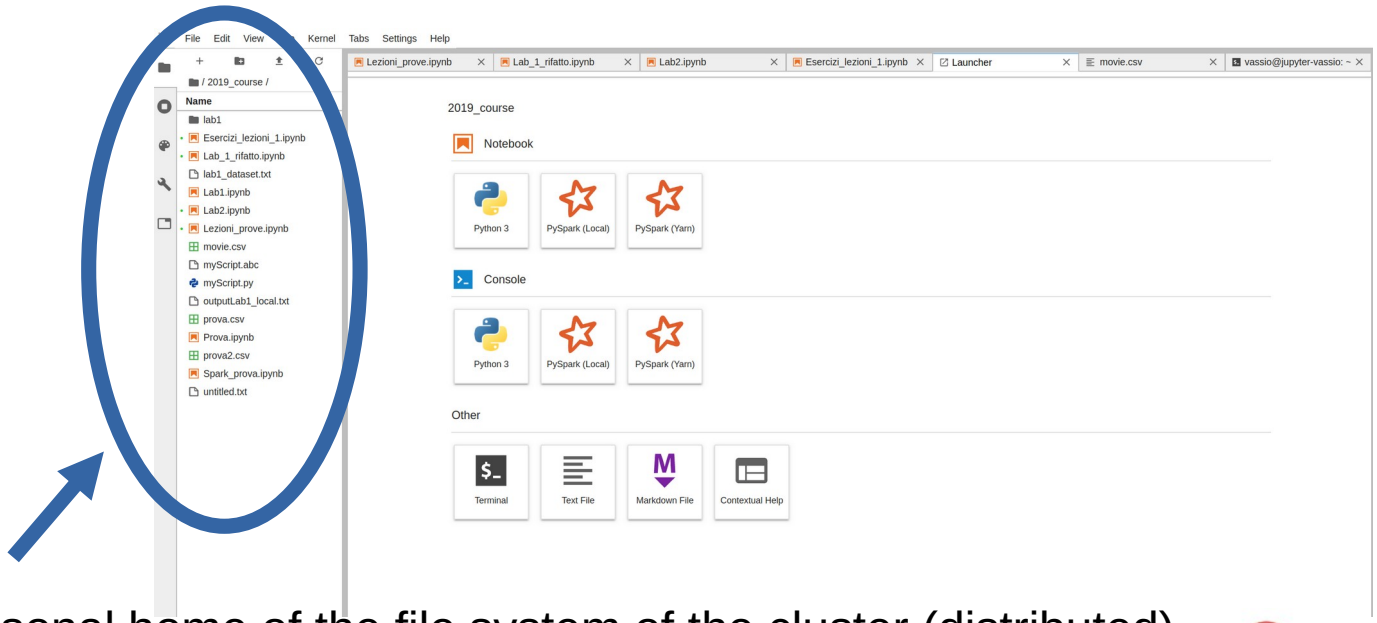
https://jupyter.polito.it/



You should have received credential to your studenti.polito.it email

Jupyter web interface for our BigData cluster

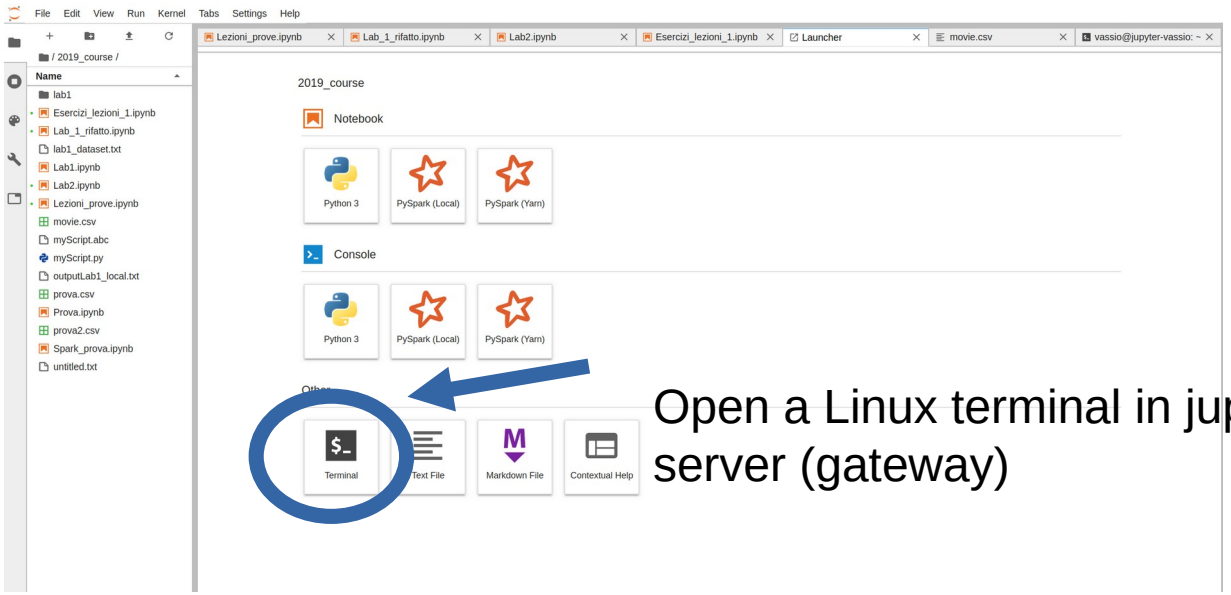
https://jupyter.polito.it/



Personal home of the file system of the cluster (distributed)
Same file system is also mounted on bigdatalab.polito.it

Linux shell in Jupyter web

https://jupyter.polito.it/



Open a Linux terminal in jupyter.polito.it server (gateway)

Jupyter account

Student

- ✓ Max 4 CPU/T and 30GB RAM
- ✓ Easy to saturate with PySpark Local
- ✓ By default, scheduled on old Nodes
- ✓ Python version: 3.7
- ✓ Timeout after 1 hour of inactivity or 7 days of running pod

Storage availability

- Every user has at maximum **10TB** of storage in **/home/<username>**
- Every user can save at maximum **2 millions** of files
- If needed, thanks to **/share** the users can share files among them
- **PAY ATTENTION: all the folder except for /home and /share aren't persistent**
  If you save your stuff in /tmp they will not be saved after Jupyter restart.

The BigData@Polito environment