



INFN-Cloud and ML_INF

Daniele Spiga (INFN-PG)
On behalf of INFN-Cloud and ML_INF teams

The logo consists of the lowercase letters "m" and "l" in a dark blue, rounded font, followed by the "INFN" logo (a dark blue swoosh underline above the word "INFN" in a dark blue sans-serif font).	<p>Second ML- INFN Hackathon: Starting Level</p> <p>13-15 December 2021</p>
---	---

What is a Cloud?

Cloud computing

From Wikipedia, the free encyclopedia

Cloud computing^[1] is the on-demand availability of **computer system resources**, especially data storage (**cloud storage**) and **computing power**, without direct active management by the user.^[2]



Google Cloud





Why a INFN-Cloud?

- An **internal effort** at the INFN level in order to manage a (large) fraction of the INFN resources, in order to decouple user needs from the availability of local and dedicated hardware
- An attempt to rationalize the access to hardware, and optimize its use
 - From “1 GPU on each desk, used 5% of the time” to “shared resources optimally used”
 - It is the same direction we saw in the change “buy me 1000 dedicated computers” to “let’s build a GRID and use it with definite priority settings”
- A way to “equalize” INFN users in the access of resources, regardless from the (richness of the) experiment, the vicinity to a powerful computing centre, the capability to administer a complex resource such as those with GPUs etc

INFN-Cloud in a (more technical) nutshell


With INFN-Cloud project we have build (and we are building) :

- ☑ A multi-site Federated Cloud infrastructure
 - ☐ resources from all INFN Structures can appear as a single entity
- ☑ **A set of services that can be used through a portal, from a terminal or with a set of APIs.**
- ☑ **A "high-level" mechanism for the adaptation and evolution of the service portfolio according to the needs and requests of users.**
- ☑ A fully distributed intra-INFN organization for the support and management of infrastructure and services.
- ☑ A series of rules for access and management policies of INFN Cloud resources that incorporate INFN regulations and the more general national ones.

Objectives of INFN-Cloud

To provide solutions for a wide range of user/community needs :

- a set of distributed computing solutions, from the simplest (“I need a Linux PC for some uses, I do not want to buy one”) to open source composable components that allow INFN users to use, build and develop modern computing models and related resources.
- For example: ability to leverage a mix of public / private Cloud infrastructures, distributed POSIX / object storage solutions, CPU / GPU resources, reusable ML models

- 
- A large, hollow red arrow pointing to the right, positioned to the left of the boxed list.
- **Scientific Computing**
 - **Development and R&D, testing of new services**
 - **Training activities**
 - Support to INFN data centers (for example for backups of services, etc)

The INFN-Cloud services

Virtual Machines (VM) possibly with external volume for storing data.

Docker containers

Pre-configured environment for **data analytics**

- Spark e/o Elasticsearch e Kibana, R, etc..

Storage solutions: Object storage / posix, possibly connected to high level application layers;

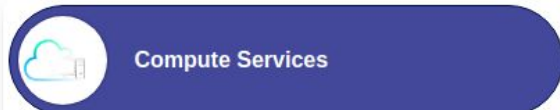
- Jupyter Notebooks with persistent storage (replicated)

Dynamic Clusters even designed and tuned taking into account the specific communities needs;

- HTCondor batch system; environment optimized for ML i.e. equipped with GPUs
- Container orchestrators such as K8s and Mesos

User-level disk encryption to manage confidential data

- Certified Cloud IEC/ISO 27001 at CNAF



A list of services that enable a specific cloud technology



A collection of ad-hoc solutions for analytic purpose



List of ready-to-use Machine Learning services



Data management and storage services



Customized environments

INFN-Cloud physical implementation

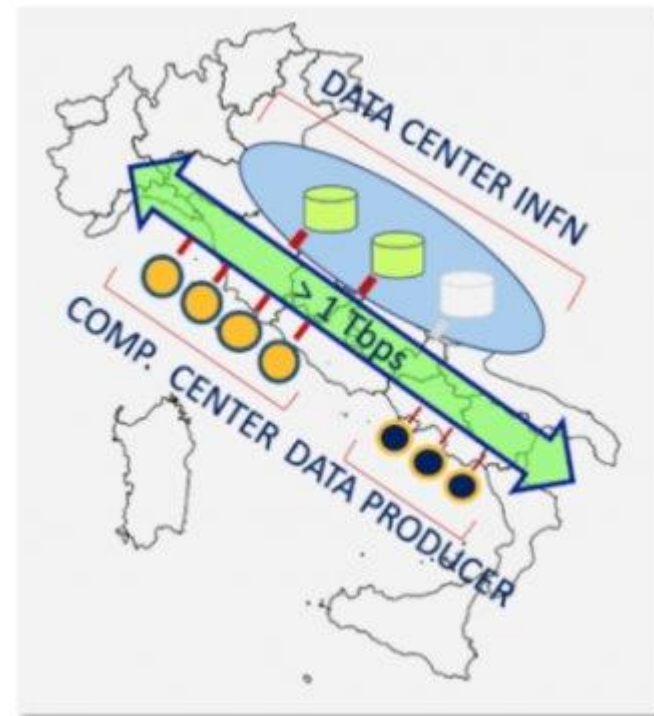
A guaranteed “backbone” which connects and includes the two largest INFN data centers (CNAF and Bari)

- The backbone is used for INFN core and certified services
 - Platform as a Service tools, DNS servers, monitoring and logging services, ...
- It is also used to host mission critical services from users (high availability, redundancy, disaster recovery)

A possibly large and diverse list of distributed Cloud Infrastructures, connected with the backbone and “federated” no replication,

- One day, possibly all INFN resources
- Not replicated / less redundant But eventually many more!
- **Currently other slices of CNAF and Bari + Cloud Veneto**

The federated cloud is where you (as a Physicist) are expected to work one day!



From user perspectives: INFN-Cloud in practice



INFN Cloud Dashboard Deployments Advanced External Links Users ml-infn Daniele Spiga

Search...

Virtual machine	Virtual machine	Docker-compose
Run docker	Elasticsearch and Kibana	Apache Mesos cluster
Kubernetes cluster	Spark + Jupyter cluster	RStudio
TensorFlow with Jupyter	(Experimental) Jupyter with persistence for Notebooks	Computational environment for Machine Learning INFN (ML_INFN)
Computational environment for Machine Learning INFN (ONLY FOR TEST)	Galaxy	Working Station for CYGNO experiment

You can visit [here](#)

(if you are/once you get authorized)

ml-infn Daniele Spiga

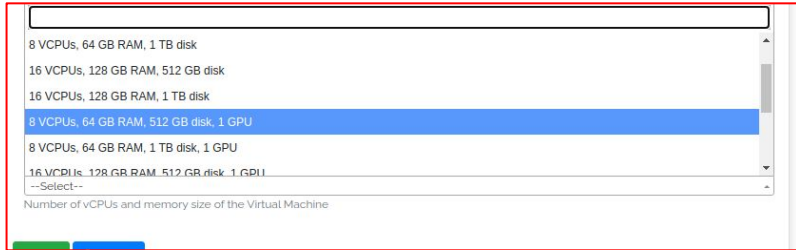
- infn-cloud-catchall
- ml-infn
- orchestrator-admin
- beta-testers

These are the resources you will use in these 3 days ...

If you are authorized ... you can create your own machine!

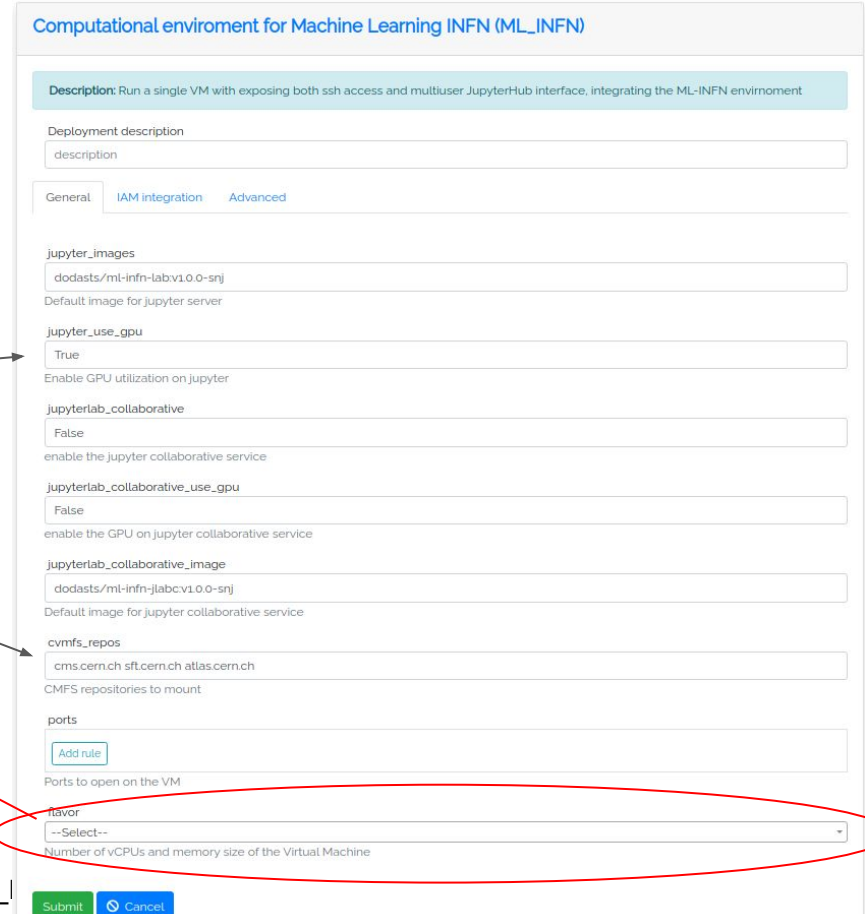
Simple high-level configuration template to create your personal environment

- Either for single user and multi users (group activities)
- Ask for CVMFS areas, GPUs, ...



8 VCPUs, 64 GB RAM, 1 TB disk
 16 VCPUs, 128 GB RAM, 512 GB disk
 16 VCPUs, 128 GB RAM, 1 TB disk
8 VCPUs, 64 GB RAM, 512 GB disk, 1 GPU
 8 VCPUs, 64 GB RAM, 1 TB disk, 1 GPU
 16 VCPUs, 128 GB RAM, 512 GB disk, 1 GPU
 --Select--

Number of vCPUs and memory size of the Virtual Machine



Computational environment for Machine Learning INFN (ML_INFN)

Description: Run a single VM with exposing both ssh access and multiuser JupyterHub interface, integrating the ML-INFN environment

Deployment description
 description

General | IAM integration | Advanced

jupyter_images
 dodasts/ml-infn-lab.v1.0.0-snj
 Default image for jupyter server

jupyter_use_gpu
 True
 Enable GPU utilization on jupyter

jupyterlab_collaborative
 False
 enable the jupyter collaborative service

jupyterlab_collaborative_use_gpu
 False
 enable the GPU on jupyter collaborative service

jupyterlab_collaborative_image
 dodasts/ml-infn-jlab.v1.0.0-snj
 Default image for jupyter collaborative service

cvmfs_repos
 cms.cern.ch sft.cern.ch atlas.cern.ch
 CVMFS repositories to mount

ports
 Add rule

Ports to open on the VM

flavor
 --Select--
 Number of vCPUs and memory size of the Virtual Machine

Submit Cancel

Today we've done all this for you in advance:

10 VMs for the 10 groups -- you have received an email on "which is yours"



INFN Cloud Dashboard Deployments Advanced External Links Users ml-inf Daniele Spiga

My deployments

Show 10 entries Search:

Description	Deployment identifier	Status	Creation time	Deployed at	Actions
Gruppo5	11ec57bo-7583-dfa8-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 22:53:00	CLOUD-CNAF	Details
Gruppo8	11ec5787-28fd-6806-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 17:57:00	CLOUD-CNAF	Details
Gruppo6	11ec5786-eb92-23ef-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 17:55:00	CLOUD-CNAF	Details
Gruppo10	11ec577c-4ca7-cfee-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 16:39:00	CLOUD-CNAF	Details
Gruppo9	11ec577c-374c-6fcb-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 16:39:00	CLOUD-CNAF	Details
Gruppo7	11ec577c-0d14-0675-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 16:38:00	CLOUD-CNAF	Details
Gruppo4	11ec5720-9a11-f222-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 05:43:00	CLOUD-CNAF	Details
Gruppo3	11ec5720-7ad6-ccb-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 05:42:00	CLOUD-CNAF	Details
Gruppo2	11ec5720-49c9-4574-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 05:41:00	CLOUD-CNAF	Details
Gruppo1	11ec5720-2659-01e5-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 05:40:00	CLOUD-CNAF	Details

Let's move to the actual setup for this Hackathon



Brief description of the tools

- What you should know

Description of the working areas

- Where you need to work

What about if we've a major disaster:

- What should we do i.e if we cancel our work by mistake?
- A list of best practices

Finally:

→ A quick live walkthrough -- you can try it with me!

The Jupyter based environment

Everything is [Jupyter](#) based. As you will see we are working with JupyterHub because we want a **multi-users environment**

- In the end you will just work with your Notebooks via JupyterLab (which is ~ **interactive python**)
- Each user has his personal environment (see it just a distinct personal computer).
To some extent user A cannot interfere with user B
 - But A and B can share files indeed you are expected to share notebooks...

The access requires authentication via INDIGO-IAM (the [iam-demo](#) instance), the **standard INFN tool for Authentication and Authorization**

- You should be aware of the instructions [here](#)

How to access my jupyter environment?

- Let see it in the next slide



Creating / accessing your container #1

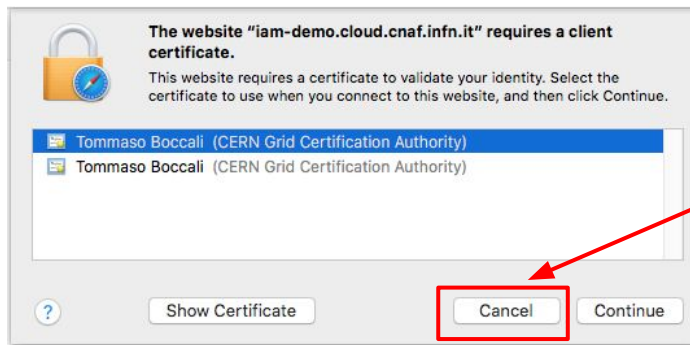


- You can go to the URL we sent you, for example substitute X with your case

- <https://mlinfnx.cloud.cnaf.infn.it:8888/>
- Tested with Chrome, Safari, Firefox, Opera

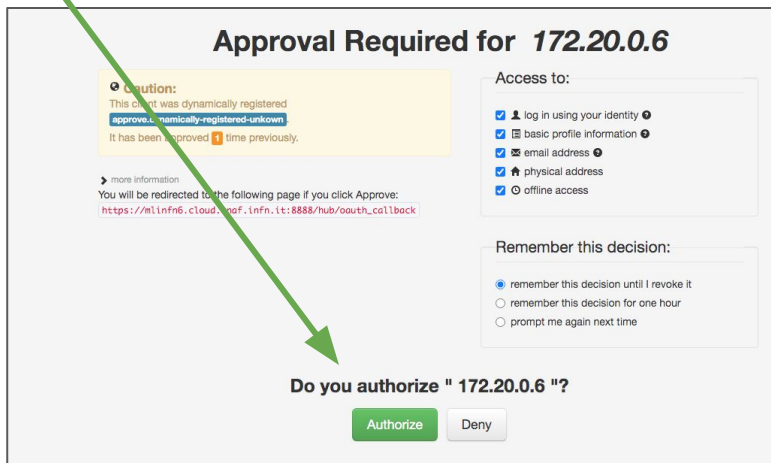
An orange rectangular button with the text 'Sign in with OAuth 2.0' in white. An arrow points from the URL above to this button.

- Your browser can ask you to select a certificate, just hit **cancel** (we are not using certificates ...)



Creating / accessing your container #2

- Login to [IAM-DEMO](#)
- The consent step, this is required only the first time, just **authorize**



Approval Required for 172.20.0.6

Caution:
This client was dynamically registered.
approved dynamically-registered-unknown
It has been approved 1 time previously.

[more information](#)
You will be redirected to the following page if you click Approve:
https://mlinf6.cloud.opf.infn.it:8888/hub/oauth_callback

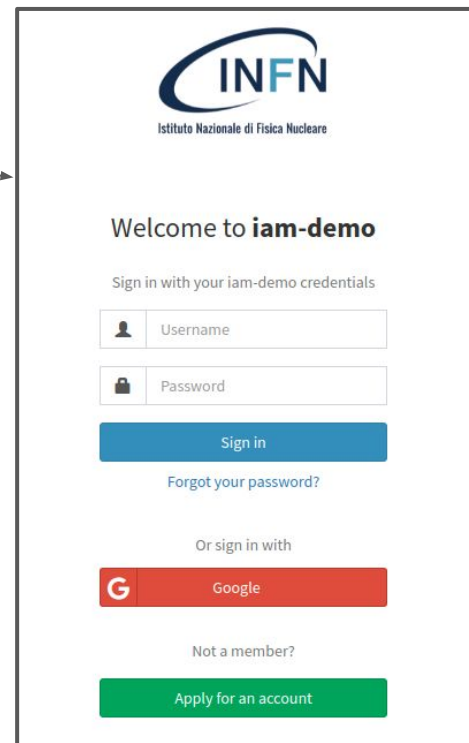
Access to:

- log in using your identity
- basic profile information
- email address
- physical address
- offline access

Remember this decision:

- remember this decision until I revoke it
- remember this decision for one hour
- prompt me again next time

Do you authorize " 172.20.0.6 "?




INFN
Istituto Nazionale di Fisica Nucleare

Welcome to **iam-demo**

Sign in with your iam-demo credentials

[Forgot your password?](#)

Or sign in with



Not a member?

Creating / accessing your own environment #3



- Each user works in a separate environment (although user A and B can share)
 - This is implemented using **docker containers**, the system spawns a docker image per user
- If you do not have your working environment (a container running) either it is the first time or you just deleted it.

Let's create one

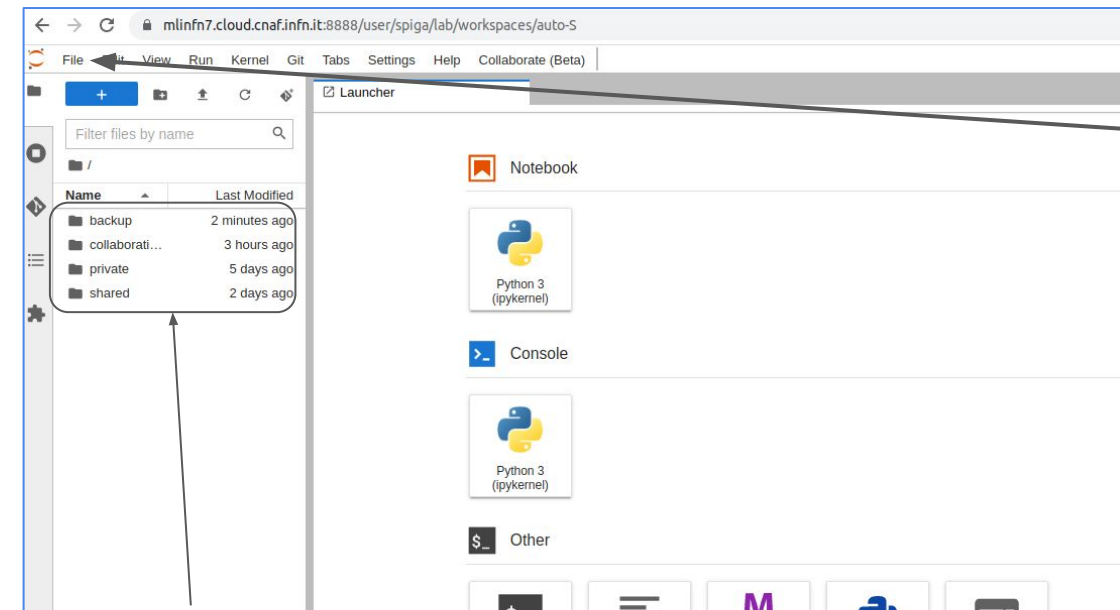
- Click on the image box, choose **the default one** (dodasts/ml-infn-lab:v1.0.0-snj)
- And then just click **Start** (and wait few seconds)

The image shows two identical screenshots of a 'Server Options' form. Each form has a title 'Server Options' at the top. Below the title, there are two input fields: 'Select your desired image:' with a text box containing 'dodasts/ml-infn-lab:v1.0.0' and 'Select your desired memory size:' with a dropdown menu showing '8GB'. At the bottom of each form is a large orange button labeled 'Start'. A blue arrow points from the first 'Start' button to the second one. Another blue arrow points from the text 'the default one' in the list to the image input field in the first form. The text 'INFN-Cloud' is visible at the bottom left of the second form, and the number '5' is at the bottom right.

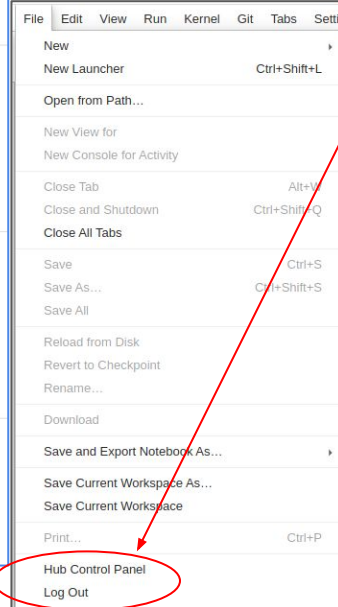
The Container: your JupyterLab interface



Allows to stop /
recreate the Container,
or to return to it

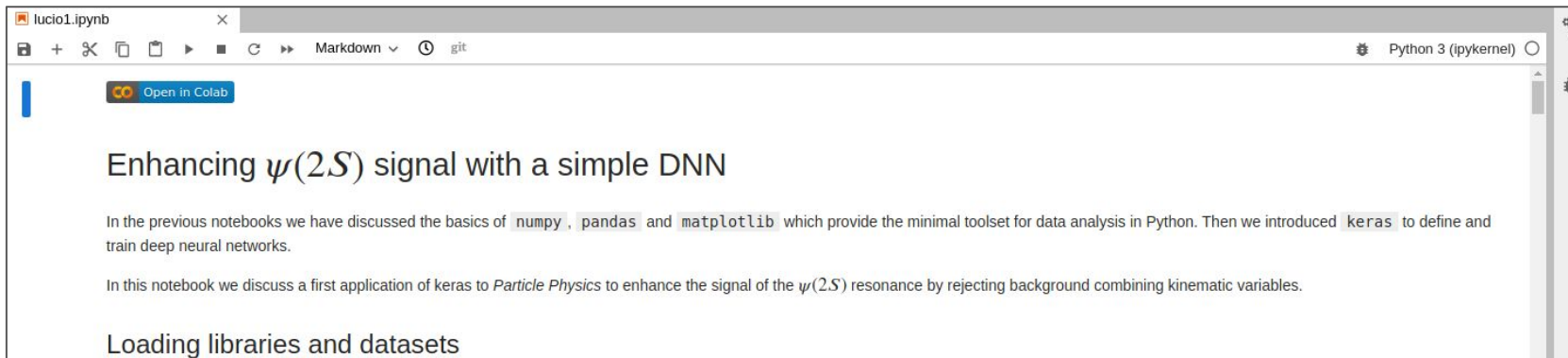
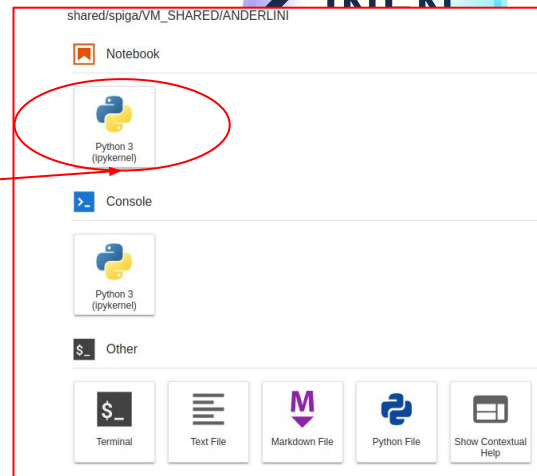


Filesystem (private /
shared, ... see later)



Use python in interactive Mode

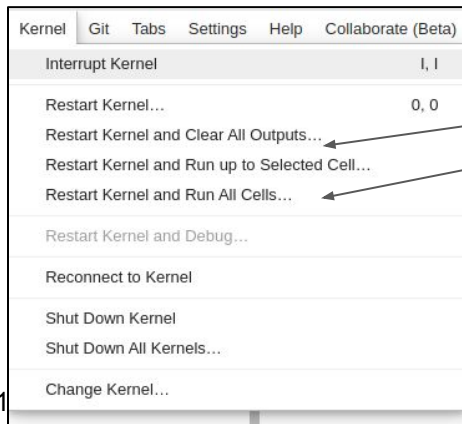
- Double click on a `.ipynb` file (**interactive python notebook**)
- OR create a new one:
- In all the cases the env is like below



Working with notebooks

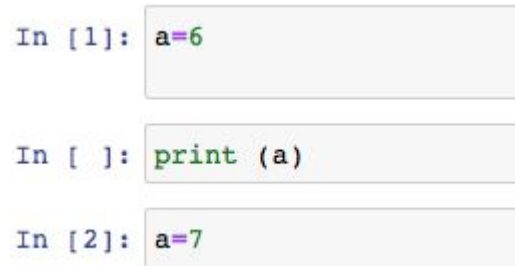
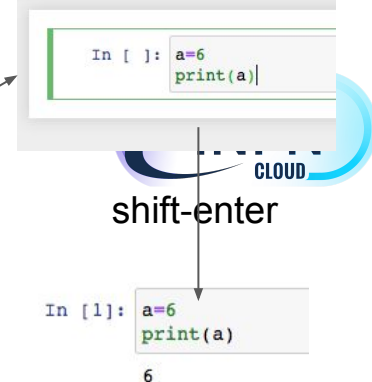
- Every cell is a (series of) python command(s)
- Evaluate the cell with the python interpreter → **Shift-Enter**
- The order of execution, not of writing. What happens if here you do shift-enter on the print cell?

Kernel menu is what you are probably going to use many times:



Reset the interpreter
Reset the interpret and evaluate all the cells in order

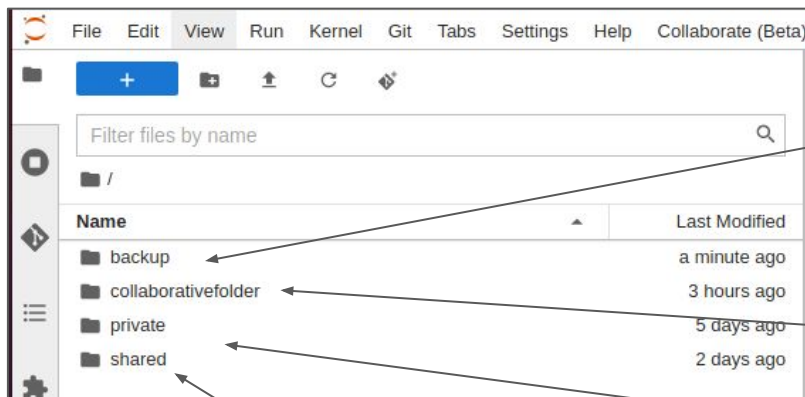
Also: The mouse right click is your friend!



More in the hands on by Lucio this afternoon

Storage Working areas

- Aka where are my/my group files? where should we work and run our code?
- We prepared 4 basic storage areas, these are SURVIVING if you destroy the container / recreate it. **All the other areas are not**



“In case you destroy stuff”
(READ only you can't override anything here)

Shared area per work group, to be used for the hackathon.

Shared area, visible by everyone. Inside, **one directory per user** (so that you can show files etc)

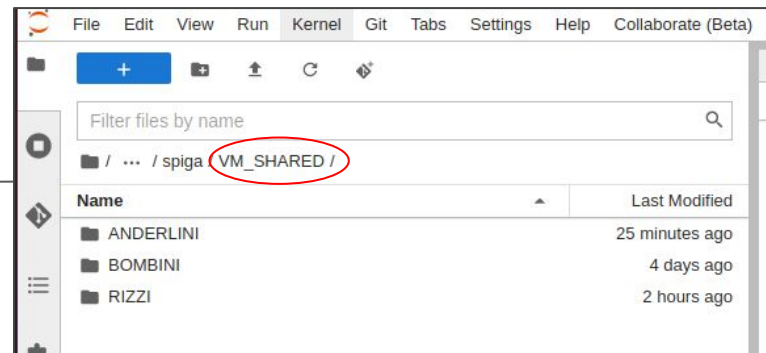
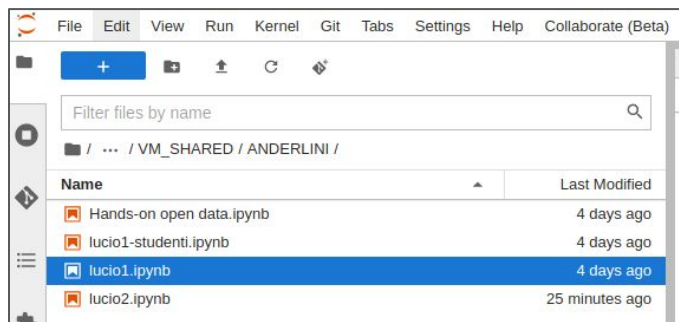
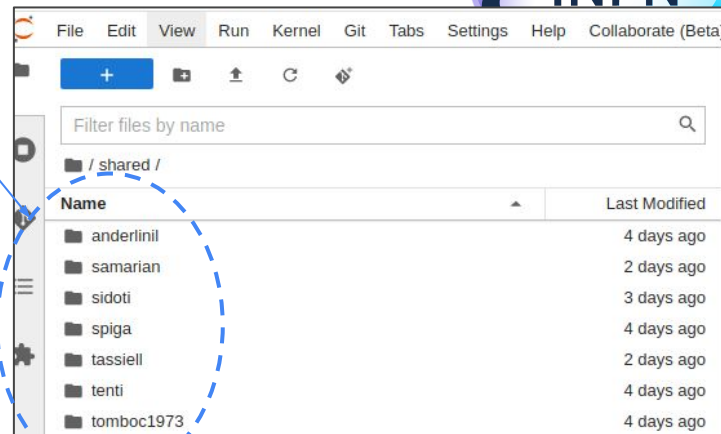
Private area, associated to your username; the other users cannot see it

“shared” area:

Note: as soon as you login the first time, the system will create your folder automatically

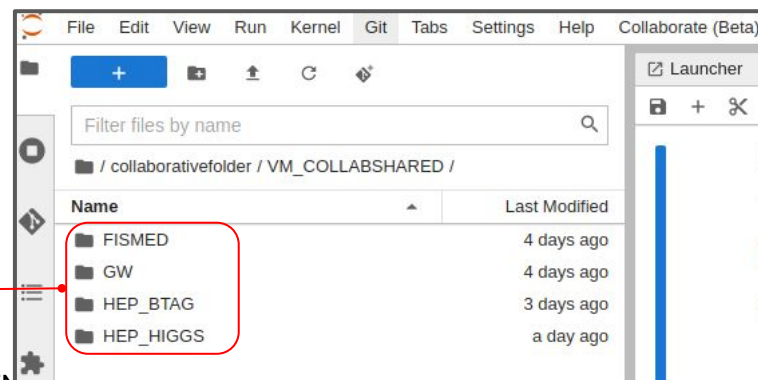
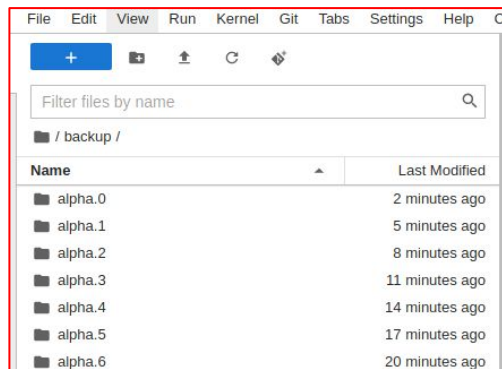


- It contains one directory per user (userid = your iam-demo credentials)
- In that directory, 3 directories for the hands-on, with the names of the Teacher
- They will tell you what to use at the start of the lesson!



“collaborativefolder” area

- it contains one directory per hackathon use case (**GW**, **HEP_BTAG**, **HEP_Higgs**, **FisMed**)
- Beware that **everyone can write there**, so please make sure you follow the instructions on Wednesday morning: only one user opens it in turn, and shares it with the others via Zoom
- The “rolling” backup take care of saving all the notebooks every 3 min
 - (as an additional safety, in case of a disaster)

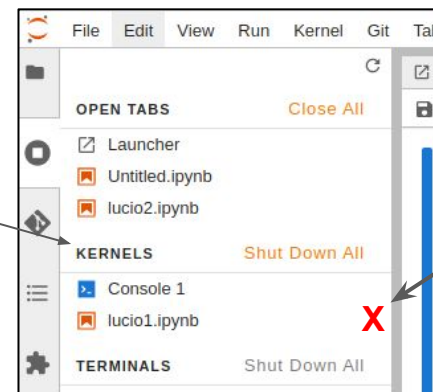
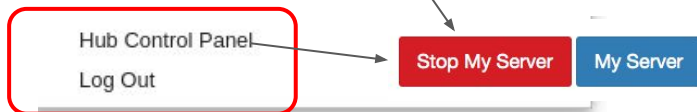


Backup

INFN-Cloud and ML_INFN,

Best practices for these three days

- On the hackathon (**wed morning**) we are trying to use a shared FS to simulate a group work
 - Ideally we would be in front of the same screen
 - We need discipline .. follow what your tutor says!
- (**in every moment**) you have limited resources (CPU, RAM, ...), please make sure you do not have notebooks hanging: you can see and kill from the “Kernel tab”
 - In case, you can even destroy the container and restart it



Kill if un-needed