



INFN-Cloud and ML_INFN

Daniele Spiga (INFN-PG)
On behalf of DataCloud and ML_INFN teams

The logo for ML_INFN consists of the lowercase letters "m" and "l" in a dark blue, rounded font, followed by the "INFN" logo (a dark blue swoosh underline and the word "INFN" in a dark blue sans-serif font).	<p>Second ML- INFN Hackathon: Starting Level</p> <p>21-23 June 2023</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 cloud at INFN (The vision)

Allow researchers to exploit “free” and open services to manage workflows, build pipelines, data processing and analysis and, of course, to share/to reuse technical solutions

- Allow researchers to focus on science

Technical drivers:

- to enable users to create and provision infrastructure deployments, automatically and repeatedly, with almost zero effort.
- To Implement the *Infrastructure as Code* paradigm based on declarative approach: allows to describe “What” instead of “How”
 - Let the underlying system to deal with technicalities
- To promote (and support) **container-based solutions**
- To grant data sharing among users/infrastructures

...and from user perspective: few pillars

end users should handle just few pillars

- What the user should/might see out of all of the underlying system?

Software management: a central role is played by container. A standard unit of software suitable to create **user tailored environment**, (share and port everywhere).

- Users create containers, the system distribute them via global file systems...

Infrastructure management: in principle user might chose to know “nothing” about infrastructure (SaaS model and above).

- If a researcher need/swants to customize its infrastructure, the system (the Cloud) should offer handles...
through templates [see later]

The context: 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: this applies both to data and compute



Aims at providing solutions for a wide range of user/community needs:


- Computing **Resources optimization**
- **Reuse** of solutions
- Support R&D: **design your computing model**
- A platform for **training**
- ... And of course cover the increasing needs of the community doing AI research (which needs accelerators, large systems, fast access to training data)

**Few
highlights**

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)

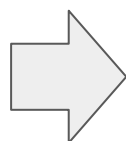
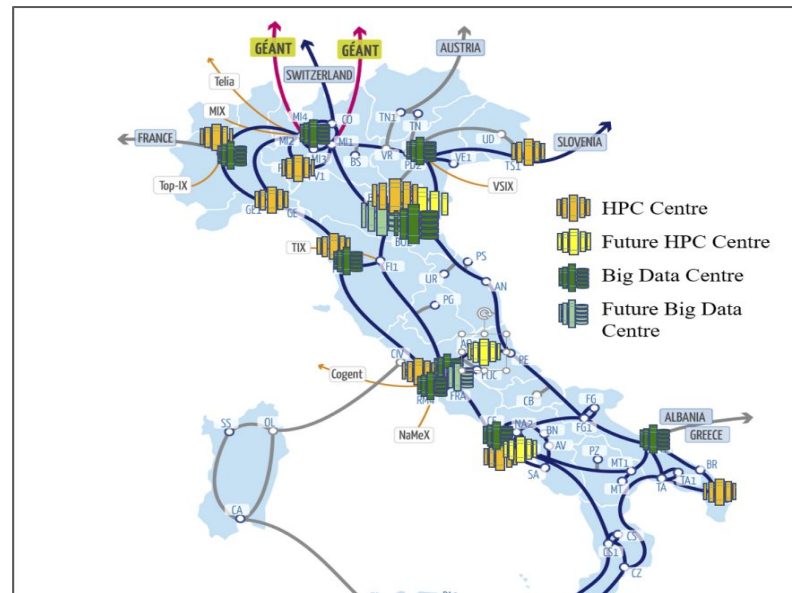
The INFN-Cloud is going to evolve as the Middleware of the National Center ICSC



INFN Computing today



ICSC Tomorrow




Finanziato dall'Unione europea
 NextGenerationEU


Ministero dell'Università e della Ricerca


Italiadomani
 PIANO NAZIONALE DI RIPRESA E RESILIENZA


ICSC
 Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

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.

The service implementation strategy



The employed strategy is based on the **Infrastructure as Code paradigm**. Users describe "**What**" is needed rather than "**How**" a specific service or functionality should be implemented. The adopted technologies enable a Lego-like approach: services can be composed and modules reused to create the desired infrastructure.



TOSCA is used to model the topology of the whole application stack

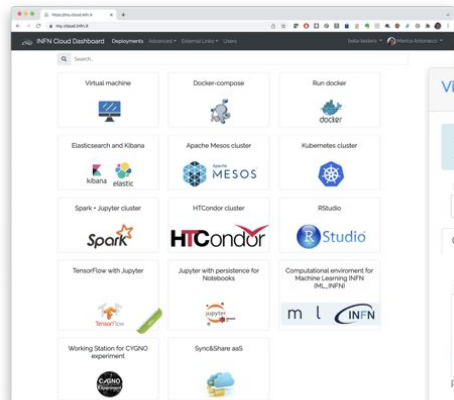


Ansible is used to automate the configuration of the virtual environments



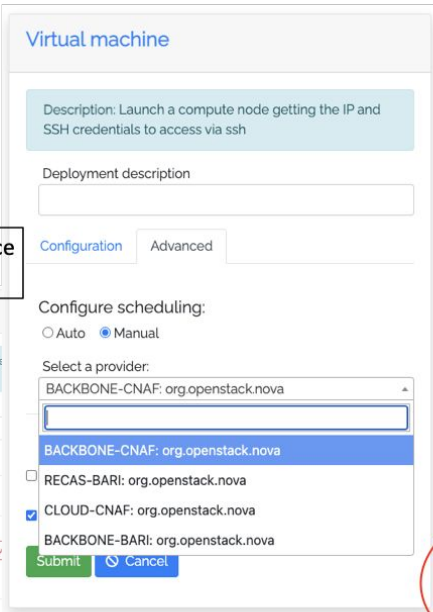
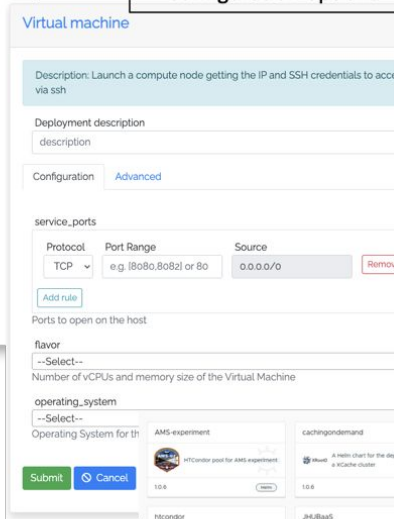
Docker is used to encapsulate the high-level application software and runtime

User interfaces first class citizens

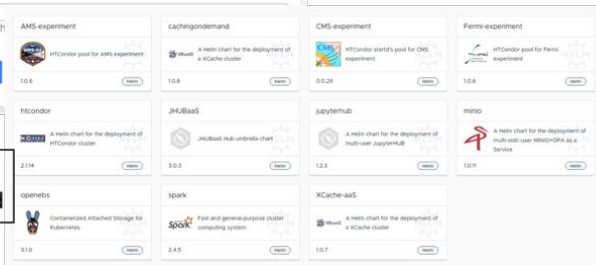


Per-user/per-group
View of the main dashboard

Basic/Advanced per-service
Configuration options



Kubeapp-level
Service dashboard



INFN Cloud Status

This page shows the high level status of the INFN Cloud services.

2022-03-25 -> 2022-03-28 - Power shutdown @ CLOUD-VENETO due to start in about 17 hours Maintenance

1. INFN Cloud
 - Object Storage Operational
 - Backbone - Cloud Compute (Bari) Operational
 - Backbone - Cloud Compute (CNAF) Operational
 - Authentication Operational
2. Federated Cloud - CloudVeneto
 - CloudVeneto - Cloud Compute Operational
3. Federated Cloud - ReCaS-Bari
 - RECAS-BARI - Cloud Compute Operational
4. Federated Cloud - Cloud@CNAF
 - Cloud@CNAF - Cloud Compute Operational
5. PaaS services
 - Infrastructure Manager Operational
 - Orchestrator Operational
 - CPR Operational
 - CMDB Operational
 - Dashboard Operational

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;

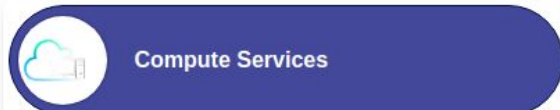
- Jupyter Notebooks with persistent storage (replicated)

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

User-level disk encryption to manage confidential data

- Certified Cloud IEC/ISO 27001 at CNAF



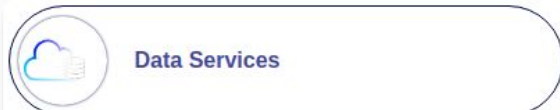
A list of services that enable a specific cloud technology



A collection of ad-hoc solutions for analytic purpose



List of ready-to-use Machine Learning services



Data management and storage services



Customized environments

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

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

You can visit [here](#)

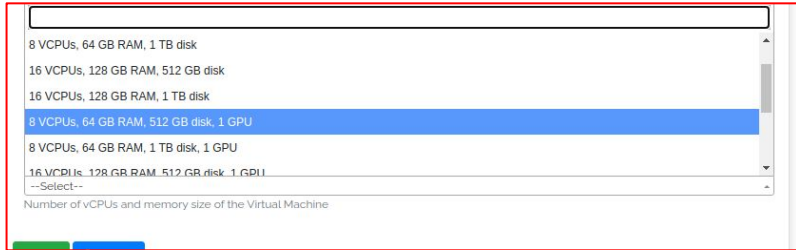
(if you are/once you get authorized)

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, ...



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"



My deployments

Show 10 entries

Refresh + New deployment

Search:

Description	Deployment identifier	Status	Creation time	Deployed at	Actions
Gruppo5	11ec57bo-7583-dfa8-edef-024269g101a7	CREATE_COMPLETE	2021-12-07 17:57:00	CLOUD-CNAF	Details
Gruppo8	11ec5787-28fd-6808-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-ccb6-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

Thanks to Stefano Dal Pra!

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

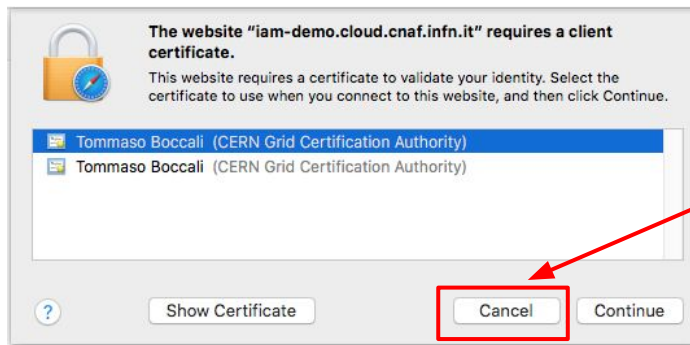
- You can go to the URL we sent you, for example substitute X with your case

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



Sign in with OAuth 2.0

- 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 *jh-client*

Access to :

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

wlwg
wlwg.groups

Remember this decision :

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

Authorizing will redirect to
https://mlinf7.cloud.cnae.infn.it:8888/hub/oauth_callback

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 (landerlini/ml-infn-lab:v1.0.6)
- And then just click **Start** (and wait few seconds)

Server Options

Select your desired image:

Select your desired memory size:

GPU:

Server Options

Select your desired image:

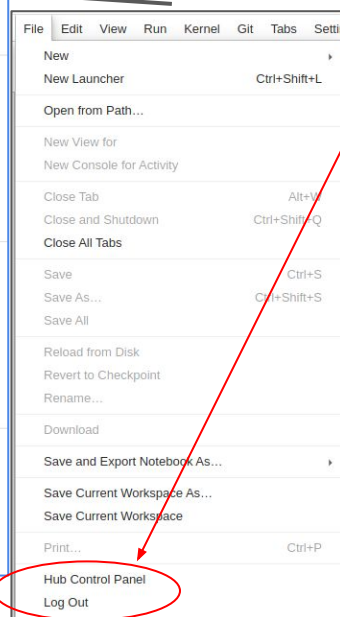
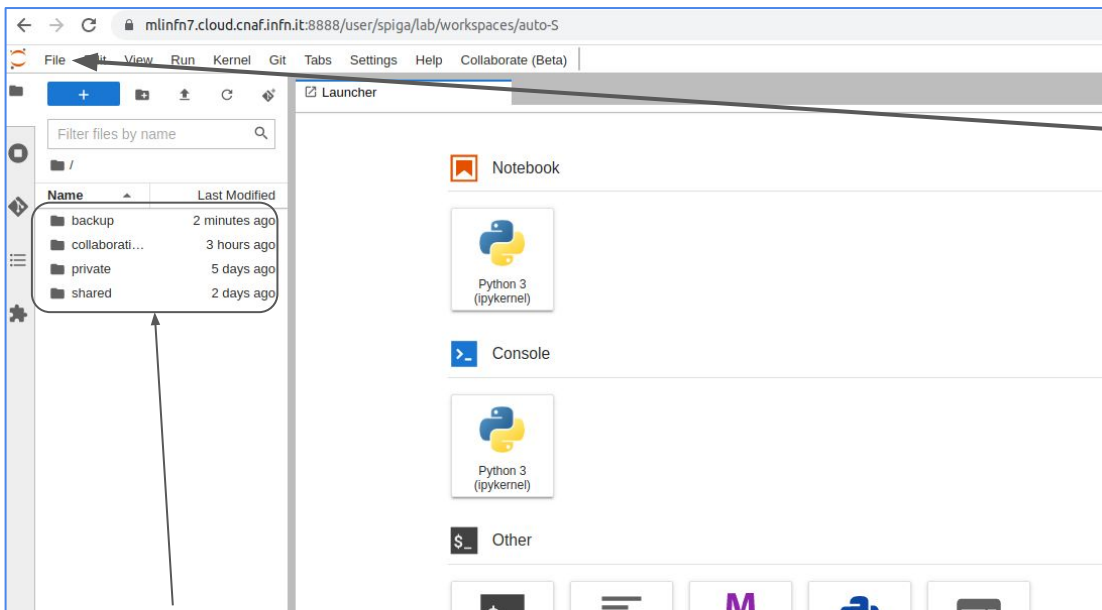
Select your desired memory size:

GPU:

The Container: your JupyterLab interface



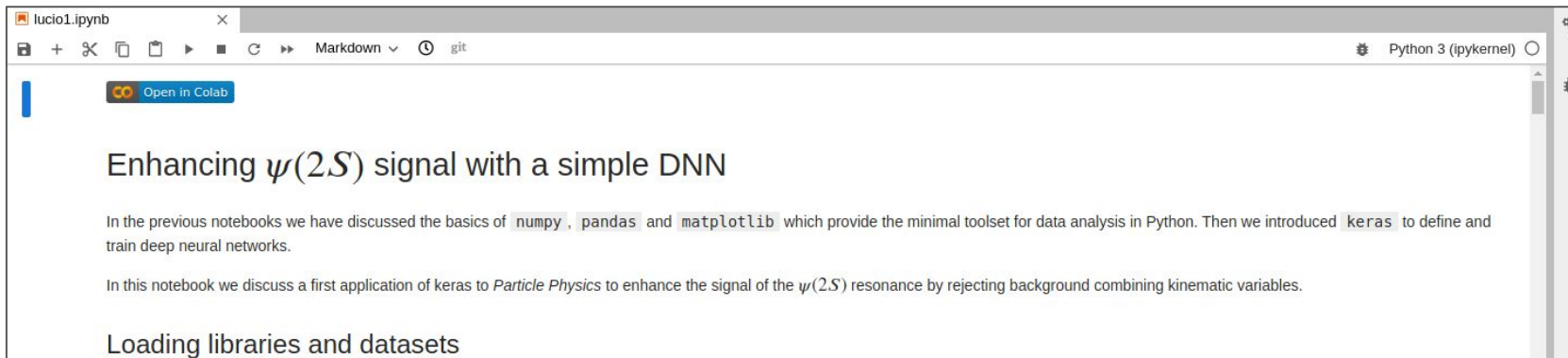
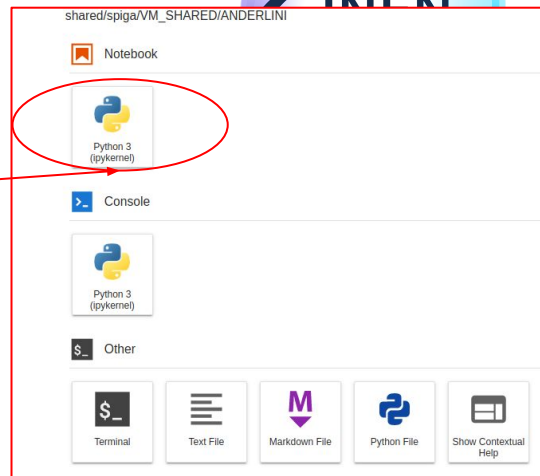
Allows to stop /
recreate the Container,
or to return to it



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

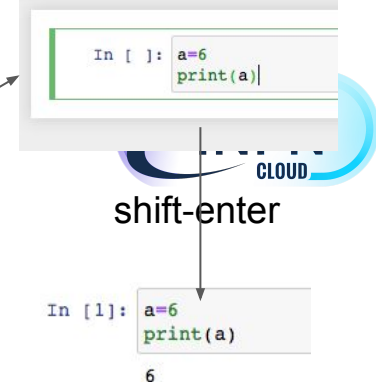
Use python in interactive Mode

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

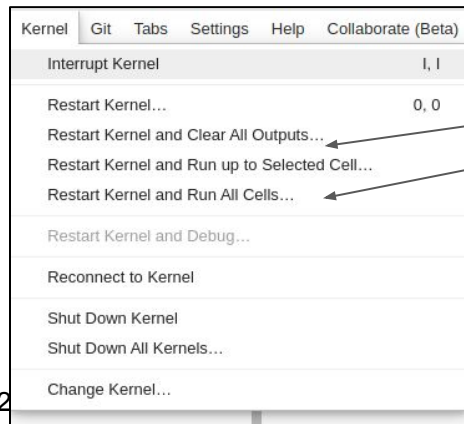


Working with notebooks

- 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!

```
In [1]: a=6
```

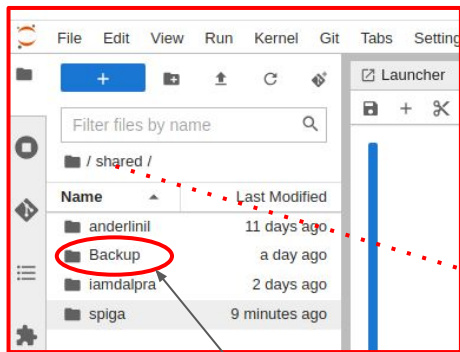
```
In [ ]: print (a)
```

```
In [2]: a=7
```

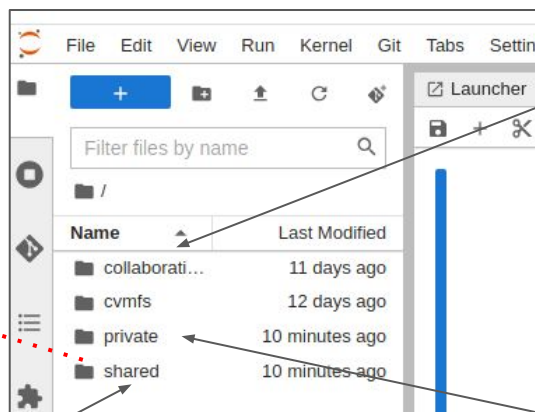
More in the hands on by Lucio this afternoon

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

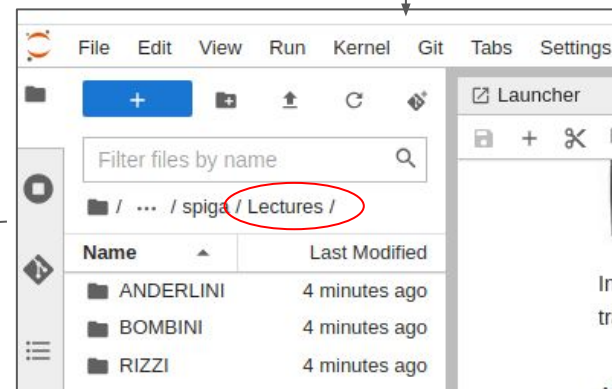
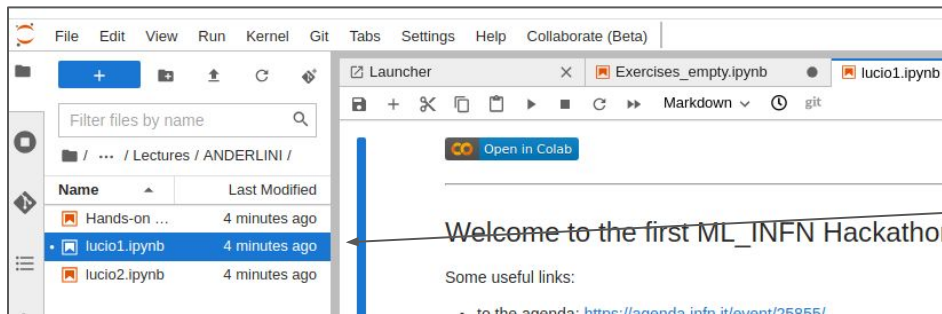
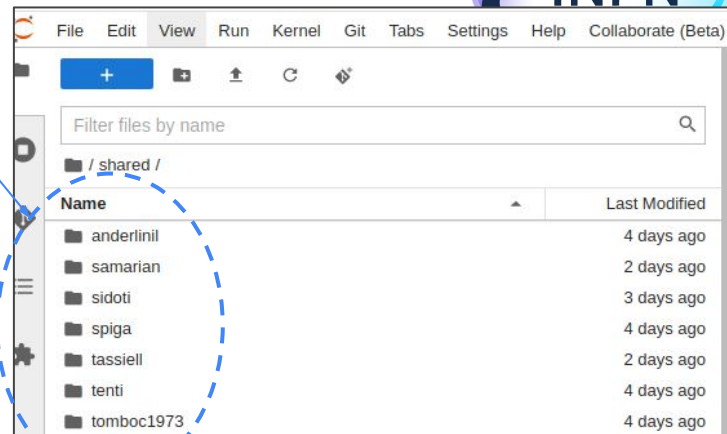
Shared area per work group, to be used for the hackathon.

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

“shared” area:

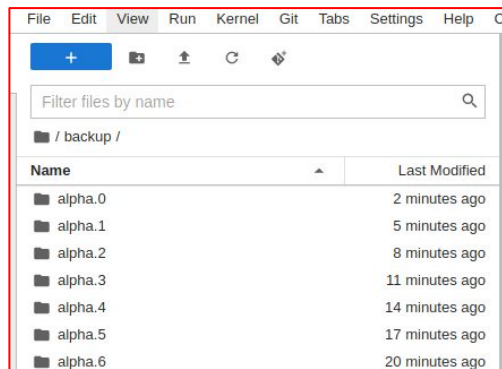
Note: as soon as you login the first time, the system will create your folder automatically

- 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!



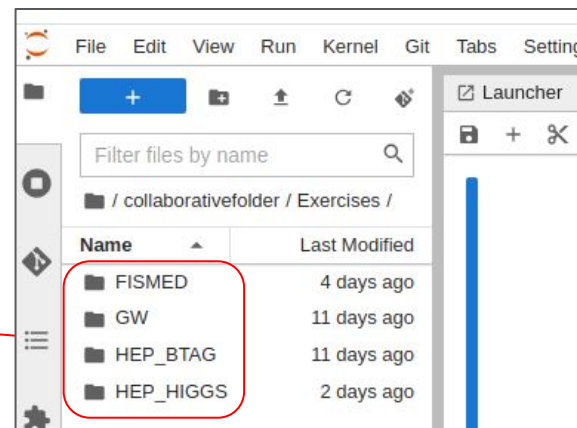
“collaborativefolder” area

- 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,



On many of these topics... **The SOSC School**

If you are interested visit <https://web.infn.it/SOSC23>

To get information about the 2023 edition

The screenshot shows the event page for SOSC 2023. At the top, it says "SOSC 2023 Fifth International School on Open Science Cloud". Below this, the dates "23-27 Oct 2023" and location "INFN Perugia" are listed, along with the "Europe/Rome timezone". A search bar is present in the top right. A navigation menu on the left includes "Overview", "Timetable", "Contribution List", "Registration", "Travel & Accomodation", "Fee Payment", and "School contact - mail" (which is highlighted). The main content area features the theme "Computing Models for Scientific Experiments", start and end times (23 Oct 2023, 08:00 to 27 Oct 2023, 22:00), and the location "INFN Perugia" with the address "Via Alessandro Pascoli, 23c 06123 Perugia". A list of organizers is provided: Daniele Bonacorsi, Daniele Spiga, Davide Salomoni, Diego Ciangottini, Livio Fano', Luca Scrucca, Lucio Anderlini, and Mirko Mariotti. There are also links to "SOSC23 First Bulletin.pdf" and "SOSC23 Poster.png". A detailed paragraph describes the school's focus on multi-disciplinary research and application requirements. At the bottom, a "Registration" section states that registration is currently open and includes a "Register now" button.

23-27 Oct 2023
INFN Perugia
Europe/Rome timezone

Enter your search term

Overview
Timetable
Contribution List
Registration
Travel & Accomodation
Fee Payment

School contact - mail
✉ sosc23-pc@lists.infn.it

The theme of the fifth International School on Open Science Cloud is "Computing Models for Scientific Experiments"

Starts 23 Oct 2023, 08:00
Ends 27 Oct 2023, 22:00
Europe/Rome

INFN Perugia
Via Alessandro Pascoli, 23c
06123 Perugia

Daniele Bonacorsi
Daniele Spiga
Davide Salomoni
Diego Ciangottini
Livio Fano'
Luca Scrucca
Lucio Anderlini
Mirko Mariotti

SOSC23 First Bulletin.pdf
SOSC23 Poster.png

The 5th edition of the international School on Open Science Cloud (SOSC 2023) will be held in **Perugia, from 23 to 27 October 2023**. The school is organized by INFN, Department of Physics and Astronomy "Augusto Righi" of the University of Bologna and the Departments of Physics and Geology of the University of Perugia.

The School is multi-disciplinary and targeted at postgraduate researchers including bachelor degree or equivalent in fields such as physics, statistics, computer science, computer vision, biology, medicine, bioinformatics, engineering, working at any research institute, with some experience and interest in data analysis, in computing or in related fields. Applications by university students (undergraduate) will be considered depending on availability and must be accompanied by a letter of reference from a university professor. **We welcome applications from all nationalities, and encourage all qualified persons to apply.**

Registration
Registration for this event is currently open. [Register now](#)