INFN-Cloud and ML_INFN

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

First ML-INFN Hackathon

07-09 June 2021

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]

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 (ore) 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 and ML_INFN,

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 rage 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

- **Scientific Computing**
- **Development and R&D, testing of new services**
- **Training activities**
- Support to INFN data centers (for example for backups of services, etc)

INFN-Cloud and ML_INFN,

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

Compute Services
A list of services that enable a specific cloud technology

Analytics
A collection of ad-hoc solutions for analytic purpose

Machine Learning
List of ready-to-use Machine Learning services

Data Services
Data management and stora ge services

Scientific Community Customizations
Customized environments

INFN-Cloud and ML_INFN,
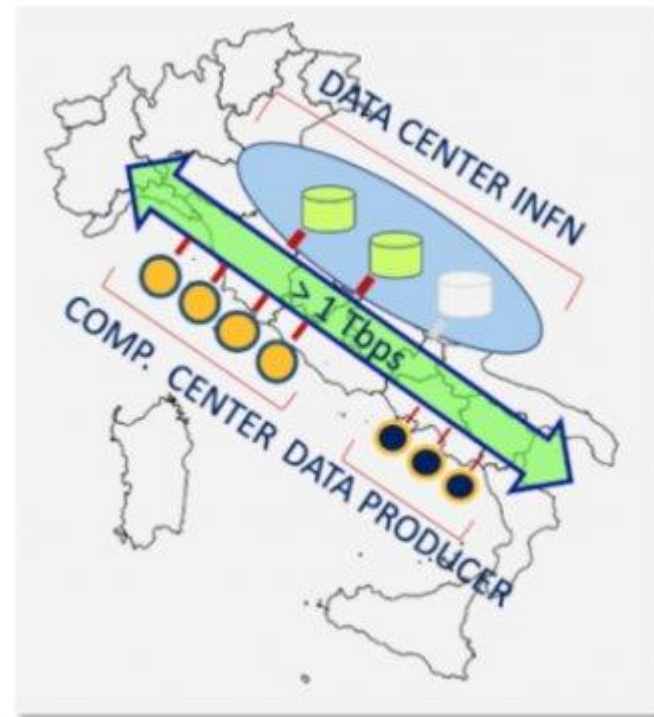
INFN-Cloud physical implementation

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
    - **P**latform **as a S**ervice tools, DNS servers, monitoring and logging services, ...
- It is also used to host mission critical serviced 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 could is where you (as a Physicist) are expected to work one day!

INFN-Cloud and ML_INFN,

From user perspectives: INFN-Cloud in practice

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 enviroment for Machine Learning INFN (ML_INFN)

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

Deployment description

> description

| General | IAM integration | Advanced |

jupyter_images

> dodasts/mlinfn-base:v1 dodasts/mlinfn-conda-base:v2

Default image for jupyter server

cvmfs_repos

> cms.cern.ch sft.cern.ch atlas.cern.ch

CMFS repositories to mount

ports

> Add rule

Ports to open on the VM

flavor

> --Select--

> 16 VCPUs, 128 GB RAM, 512 GB disk, 1 GPU
> 16 VCPUs, 128 GB RAM, 1 TB disk, 1 GPU
> 8 VCPUs, 64 GB RAM, 512 GB disk, 2 GPU
> 8 VCPUs, 64 GB RAM, 1 TB disk, 2 GPU
> 16 VCPUs, 128 GB RAM, 512 GB disk, 2 GPU
> 16 VCPUs, 128 GB RAM, 1 TB disk, 2 GPU

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

9 VMs for the 9 groups -- you have received an email on "which is yours"

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

My deployments

Refresh   + New deployment

Show [10] entries   Search:

Description	Deployment identifier	Status	Creation time	Deployed at	Actions
hackaton-gruppo9	11ebc390-4568-1d5d-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 10:49:00	CLOUD-CNAF	Details
hackaton-gruppo8	11ebc38e-5cba-7d0a-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 10:36:00	CLOUD-CNAF	Details
hackaton-gruppo7	11ebc38e-3b67-4cb3-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 10:35:00	CLOUD-CNAF	Details
hackaton-gruppo6	11ebc38e-22cb-6768-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 10:34:00	CLOUD-CNAF	Details
hackaton-gruppo5	11ebc38e-0825-f68f-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 10:33:00	CLOUD-CNAF	Details
hackaton-gruppo4	11ebc38d-cd73-1b6b-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 10:32:00	CLOUD-CNAF	Details
hackaton-gruppo3	11ebc38d-afab-8347-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 10:31:00	CLOUD-CNAF	Details
hackaton-gruppo2	11ebc38d-8dd4-f180-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 10:30:00	CLOUD-CNAF	Details
hackaton-gruppo1	11ebc383-2535-ffd6-a7b8-0242699101a7	CREATE_COMPLETE	2021-06-02 09:15:00	CLOUD-CNAF	Details
testfinal	11ebb26c-a284-3f73-a7b8-0242699101a7	CREATE_COMPLETE	2021-05-11 15:22:00	CLOUD-CNAF	Details

Showing 1 to 10 of 14 entries   Previous 1 2 Next

Let's move to the actual setup for this Hackaton

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 Jupyter Notebooks (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://mlinfn**X**.cloud.cnaf.infn.it:8888/
  - Tested with Chrome, Safari, Firefox, Opera

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

INFN-Cloud and ML_INFN,
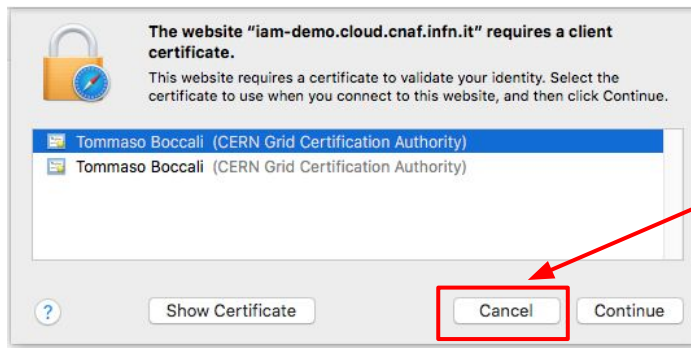
Creating / accessing your container #2

- Login to **IAM-DEMO**

- The consent step, this is required only the first time, just **authorize**

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 **DODAS V5** (the first)
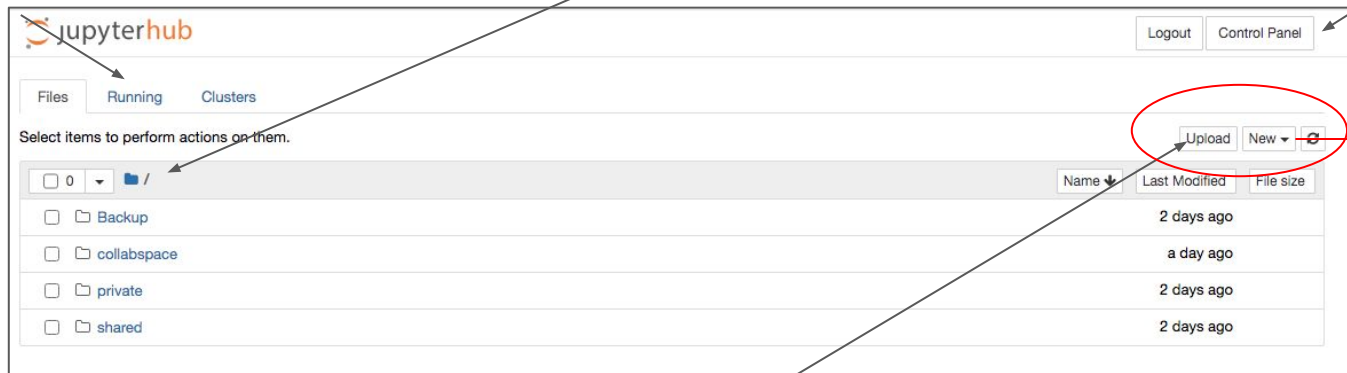
  - And then click **Start**

The Container: your Jupyter interface

Running processes tab (you can kill them from here)

Filesystem (private / shared, … see later)

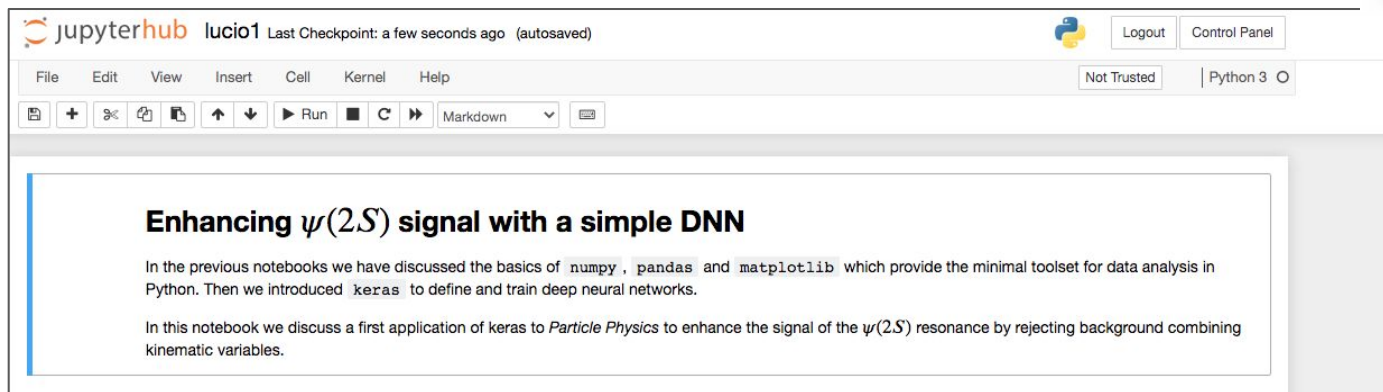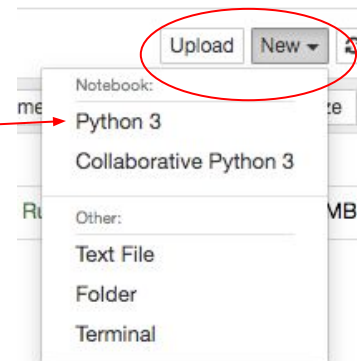**Allows to stop / recreate the Container, or to return to it**

Upload files

INFN-Cloud and ML_INFN,

Use python in interactive Mode

- Double click on a **.ipynb** file (**i**nteractive **py**thon **n**ote**b**ook)
- OR create a new one:
- In all the cases the env is like below

Upload | New ▾

Notebook:
Python 3
Collaborative Python 3

Other:
Text File
Folder
Terminal

Enhancing $\psi(2S)$ signal with a simple DNN

In the previous notebooks we have discussed the basics of `numpy`, `pandas` and `matplotlib` which provide the minimal toolset for data analysis in Python. Then we introduced `keras` to define and train deep neural networks.

In this notebook we discuss a first application of keras to *Particle Physics* to enhance the signal of the $\psi(2S)$ resonance by rejecting background combining kinematic variables.

Working with notebooks

shift-enter
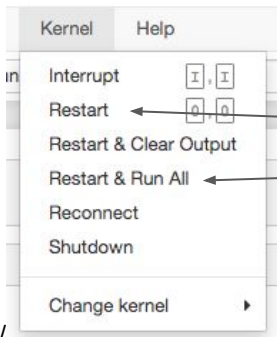
- Every cell is a (series of) python command(s)
- Evaluate the cell with the python interpreter → **Shift-Enter**
- The order of execution, not of writing. What happens if here you do shift-enter on the print cell?
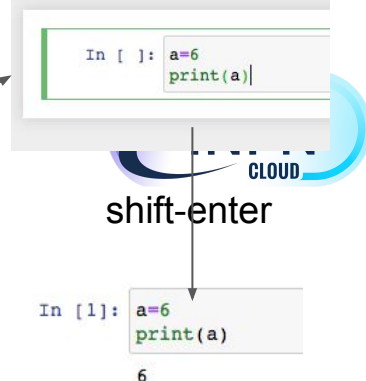
Kernel menu is what you are probably going to use many tim

Reset the interpreter
Reset the interpret and evaluate all the cells in order
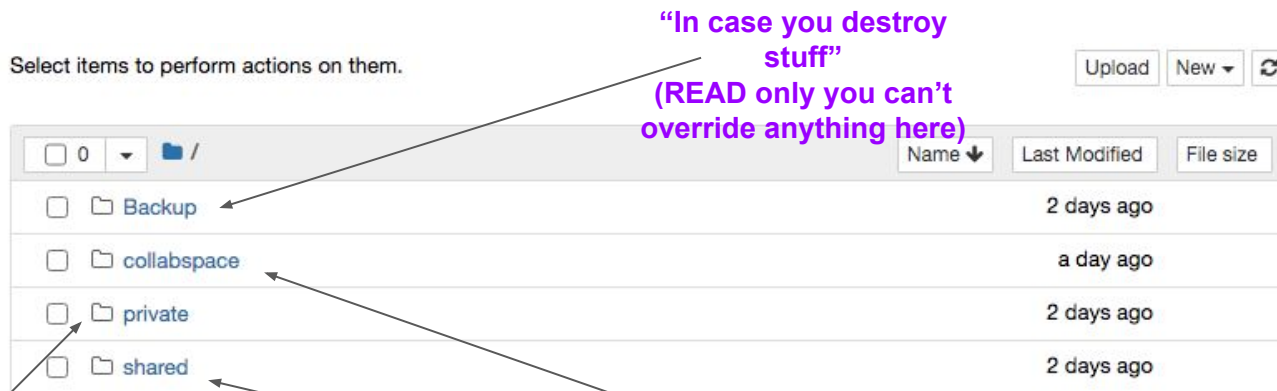
More in the hands on
by Lucio this afternoon

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)

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

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.

d and ML_INFN,

"shared" area:

- 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!

INFN-Cloud and ML_INFN,

"collabspace" area

- it contains one directory per hackathon use case (**GW**, **HEP**, **FisMed**) + an additional "rolling" backup
- 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
- In backup, all the notebooks are saved every 3 min
  - (as an additional safety)

L_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 "Running tab"

Files	Running	Clusters

Currently running Jupyter processes

Terminals ▾

There are no terminals running.

Notebooks ▾

📓 private/Untitled.ipynb   Python 3   Shutdown   seconds ago

Control Panel → Stop My Server   My Server

Kill if un-needed