INFN

CYGNO Experiment

in synergy with erc

INITIUM

funded from ERC in Horizon 2020 program (grant agreement 818744)

Exploiting INFN-Cloud to implement a Cloud solution to support the CYGNO computing model

ISGC 2022, 21-25 March 2022, Academia Sinica

I. A. Costa on behalf of CYGNO & INITIUM collaboration and INFN Cloud

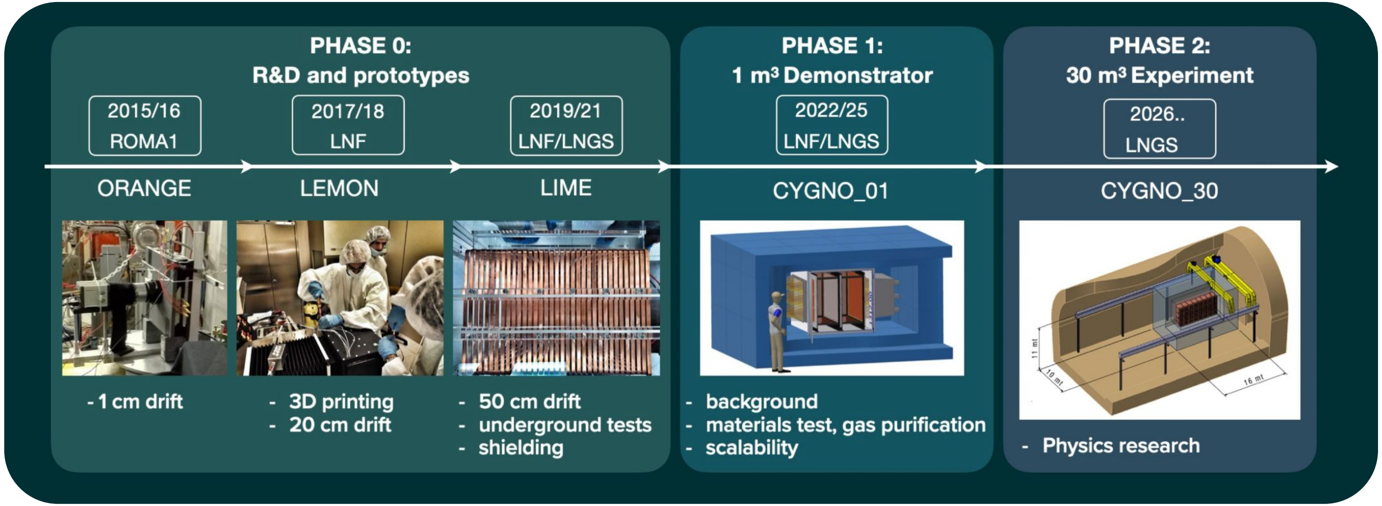
1. Introduction

Brief explanation about CYGNO

CYGNO Experiment

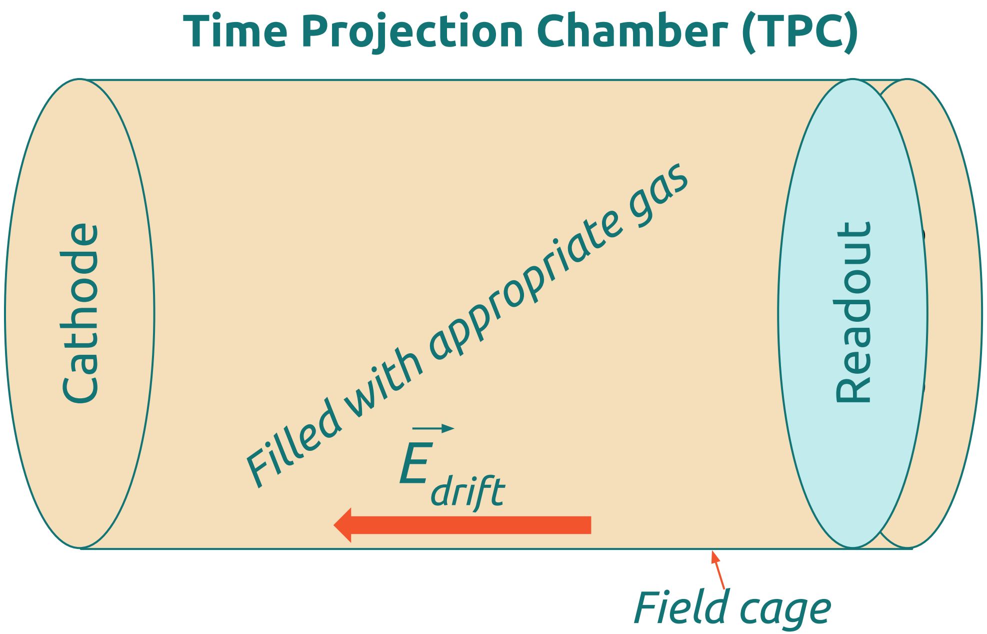
CYGNO is a large TPC for Dark Matter and Neutrino studies;

Exploiting the progress in commercial scientific Active Pixel Sensors (APS) based on CMOS technology to realise a large gaseous Time Projection Chamber (TPC) for Dark Matter and Solar neutrino search.



TPC Detector in a nutshell

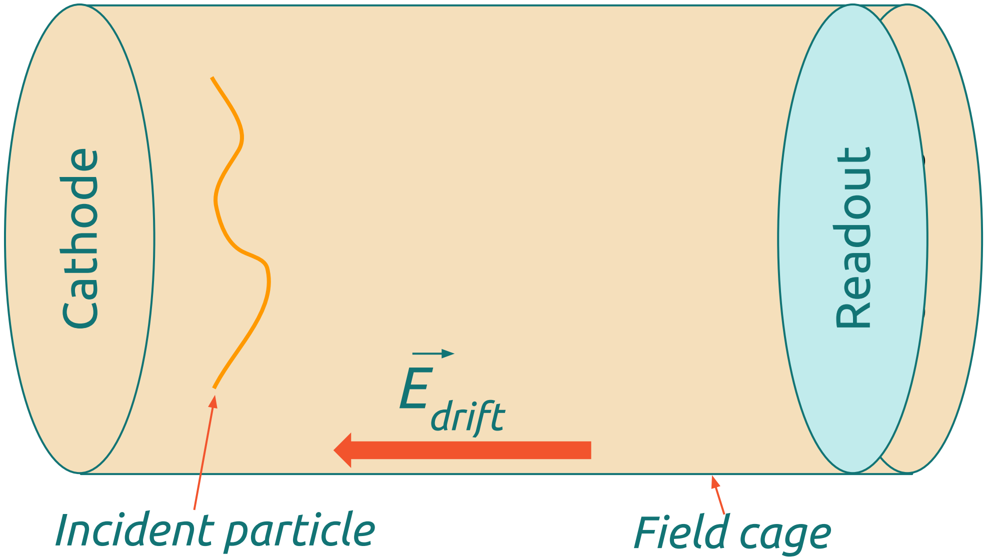
The operation principle of a **TPC gas detector readout** by a high-granularity **sensor** is the following:



TPC Detector in a nutshell

The operation principle of a **TPC gas detector readout** by a high-granularity **sensor** is the following:

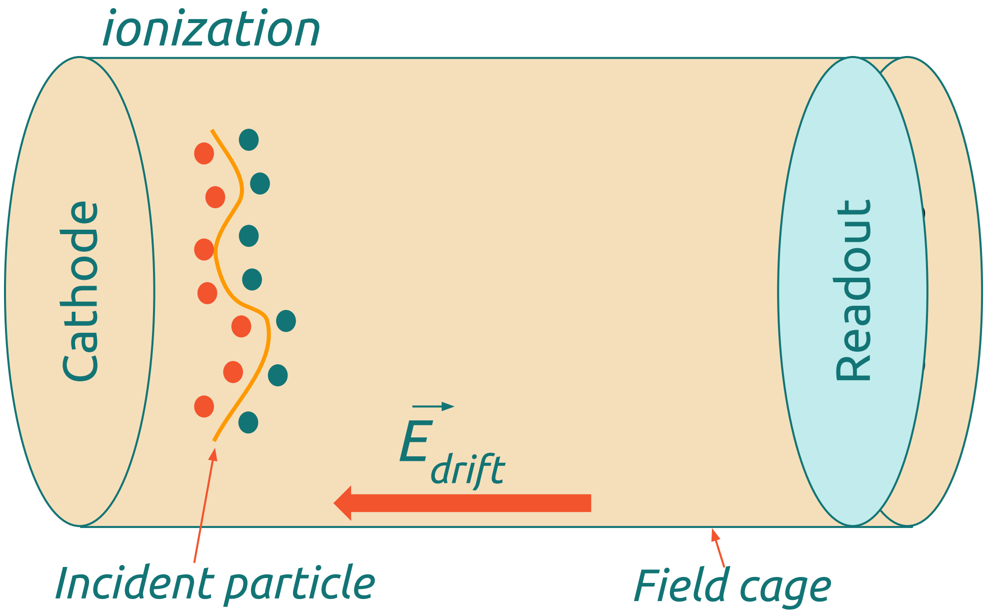
A particle **interacts** with the **gas**



TPC Detector in a nutshell

The operation principle of a **TPC gas detector readout** by a high-granularity **sensor** is the following:

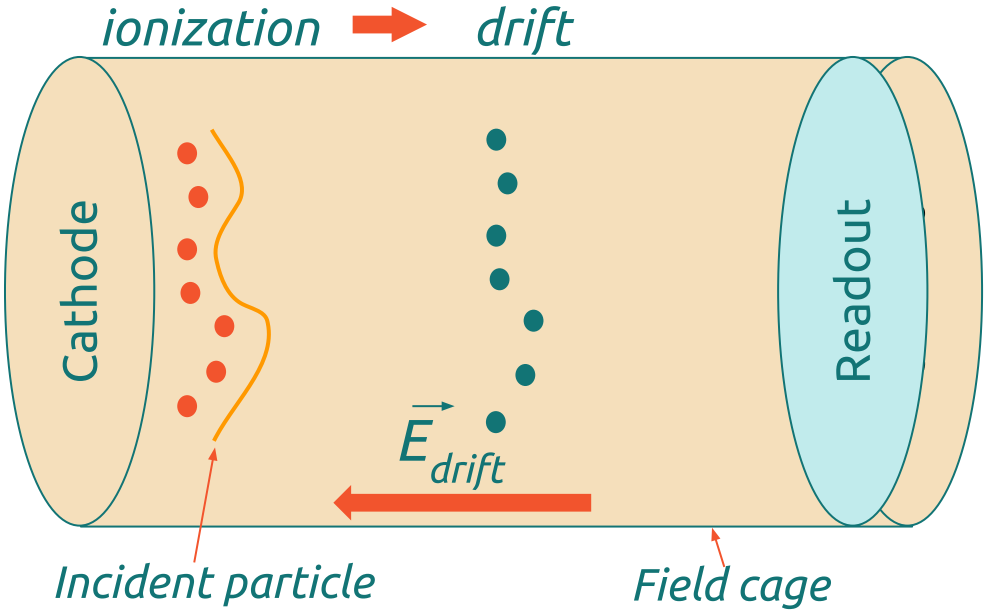
A particle **interacts** with the **gas**, producing **electrons**



TPC Detector in a nutshell

The operation principle of a **TPC gas detector readout** by a high-granularity **sensor** is the following:

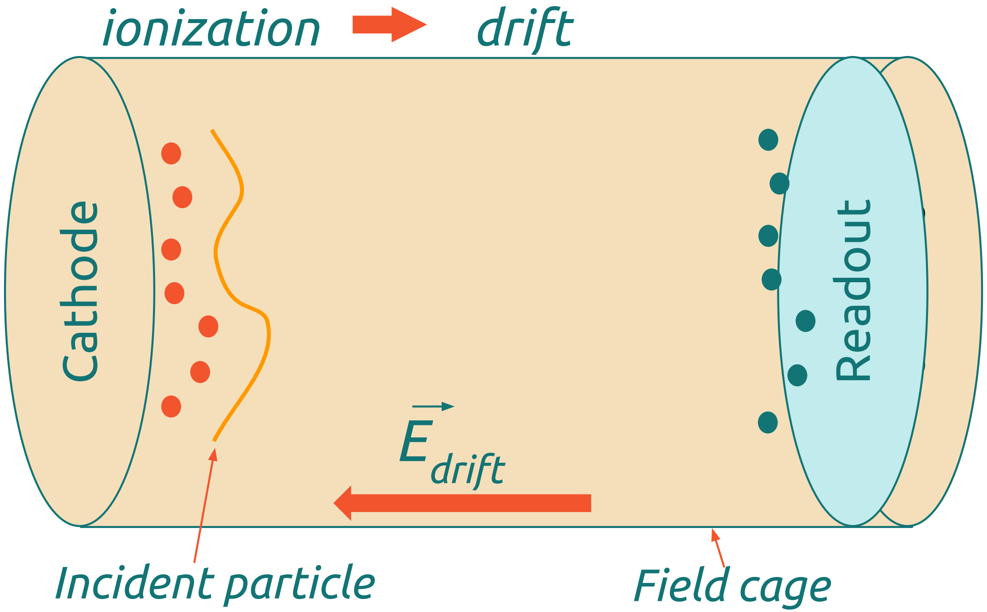
A particle **interacts** with the **gas**, producing **electrons,** that are **drifted** by the **electric field**



TPC Detector in a nutshell

The operation principle of a **TPC gas detector readout** by a high-granularity **sensor** is the following:

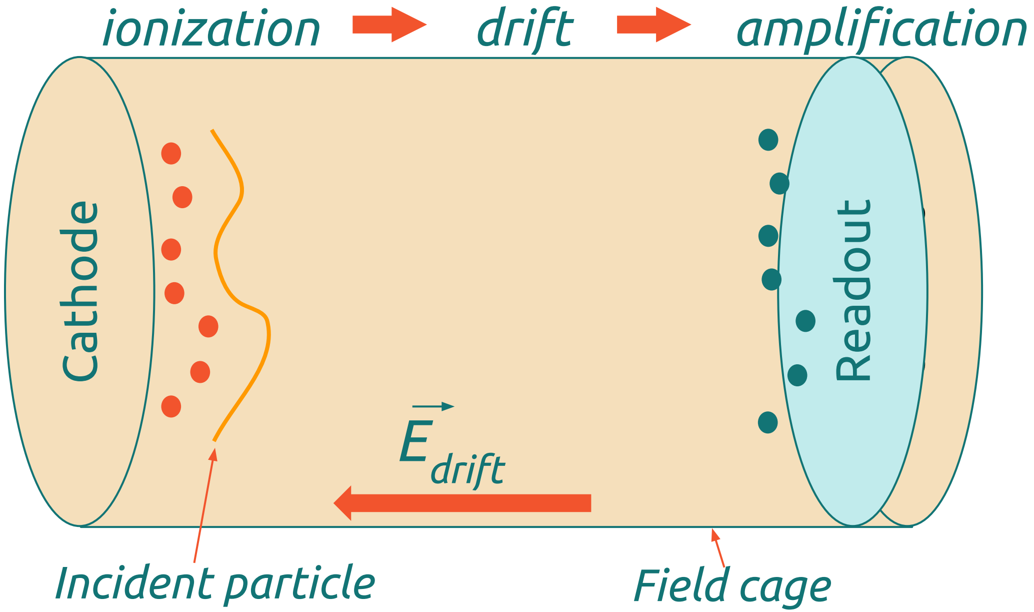
A particle **interacts** with the **gas**, producing **electrons,** that are **drifted** by the **electric field** until **reaching** the **readout**



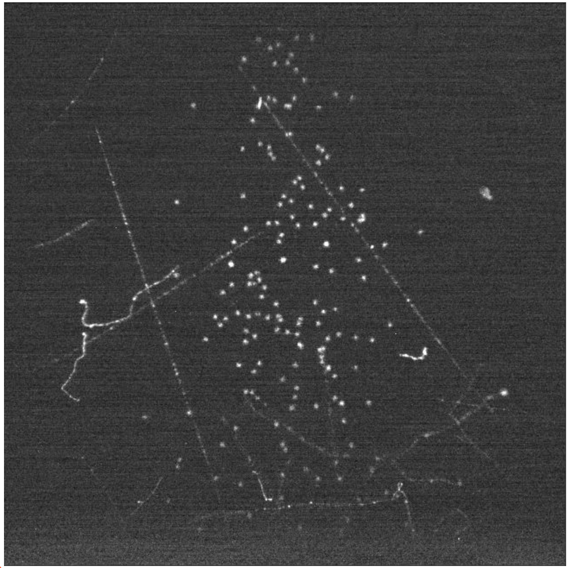
TPC Detector in a nutshell

The operation principle of a **TPC gas detector readout** by a high-granularity **sensor** is the following:

A particle **interacts** with the **gas**, producing **electrons,** that are **drifted** by the **electric field** until **reaching** the **readout** where they can be **amplified** and **collected.**



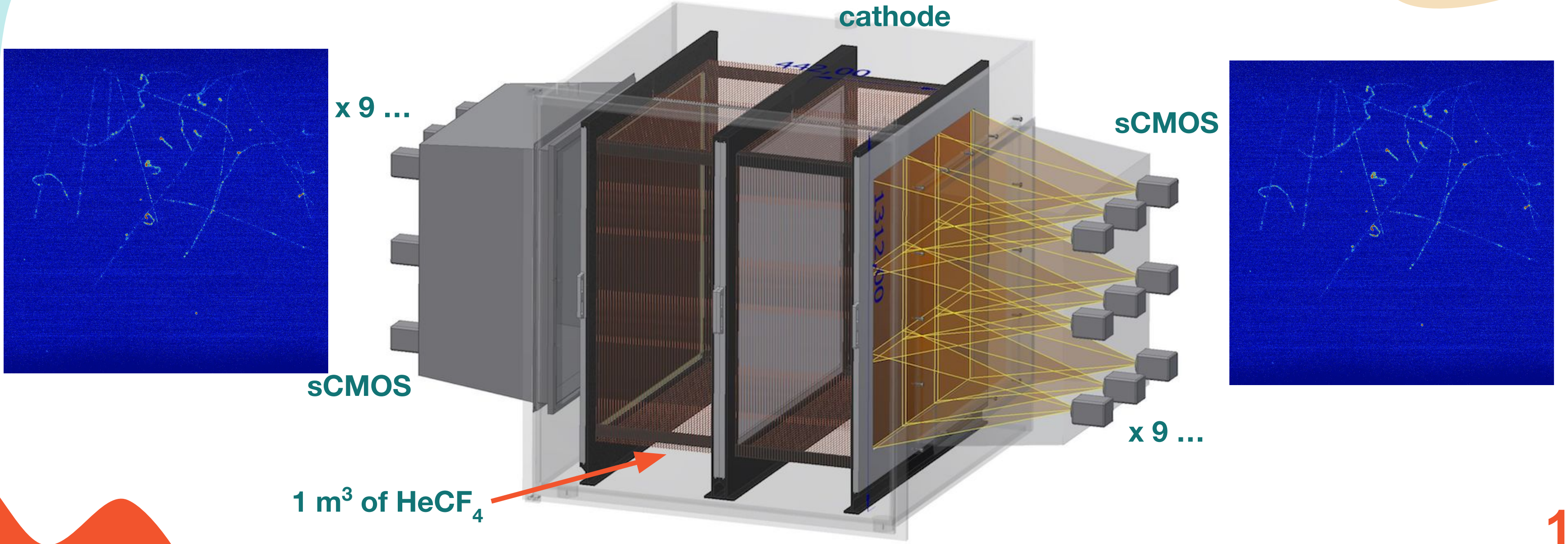
LIME detector setup at Gran Sasso



10^7 readout channels & time signals

Typically the TPC's after the amplification stage, are readout by means of charge amplifier that occupate space, are expensive, radioactive, etc.

The bet of CYGNO experiment is to exploit the APS sensor technology to acquire **18 cameras monitoring 330x330 mm each with 150 μm resolution and a sensitivity of ~ 0.5 ph/eV released in gas.**



2. Computing needs

Computing needs in the world of small and medium physics experiments

Computing needs

- **CYGNO**, like many other **small/medium astroparticle projects,** does **not require** large **computing** and **data storage** resources and, above all, the resources are **not continuously used**, but **can be intense** for **short periods** of time.

Computing needs

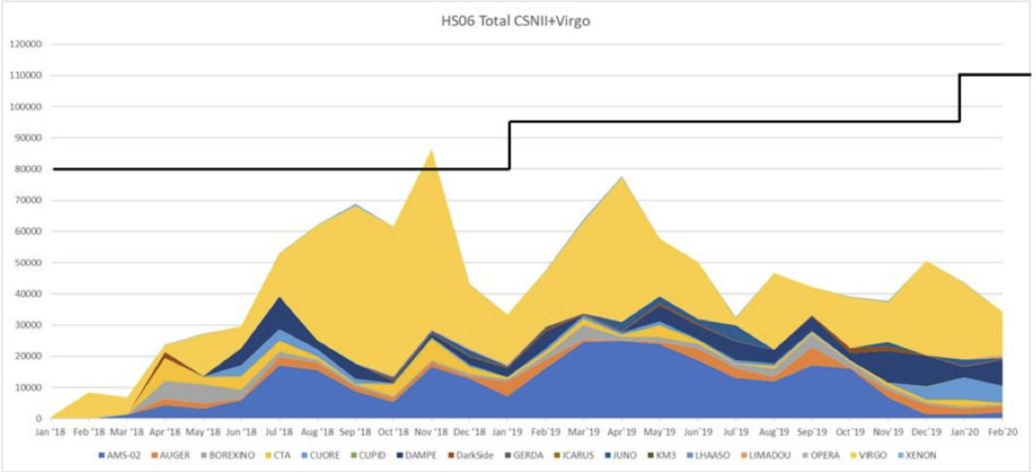
- **CYGNO**, like many other **small/medium astroparticle projects,** does **not require** large **computing** and **data storage** resources and, above all, the resources are **not continuously used**, but **can be intense** for **short periods** of time.
- For example, the **reconstruction and analysis** of CYGNO use **essentially clustering algorithms** and **machine learning** techniques on images, which **today** has a **bigger computational cost** because we are **characterizing the detector** at sea level with **high environmental radioactive,** but **in the future** we expect to have **very little of it** under the **Gran Sasso.**.

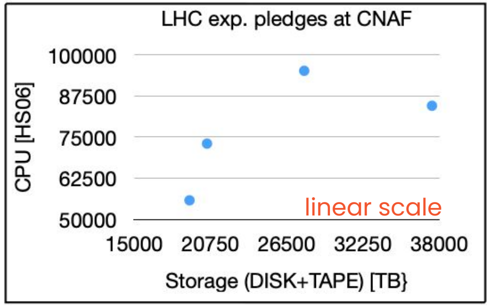
Computing needs

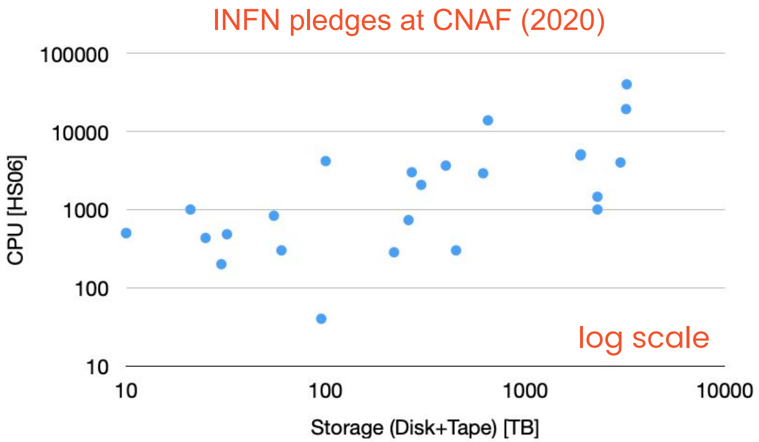
- **CYGNO**, like many other **small/medium astroparticle projects,** does **not require** large **computing** and **data storage** resources and, above all, the resources are **not continuously used**, but **can be intense** for **short periods** of time.
- For example, the **reconstruction and analysis** of CYGNO use **essentially clustering algorithms** and **machine learning** techniques on images, which **today** has a **bigger computational cost** because we are **characterizing the detector** at sea level with **high environmental radioactive,** but **in the future** we expect to have **very little of it** under the **Gran Sasso.**.
- Alongside the analysis, we need a **standard simulation environment** that uses **GEANT4, ROOT, GARFIELD, python,** etc., all of which are standard in our community.

INFN Astroparticle Computing usage

hosts 50 experiments composed of O(5 -> 500) peoples





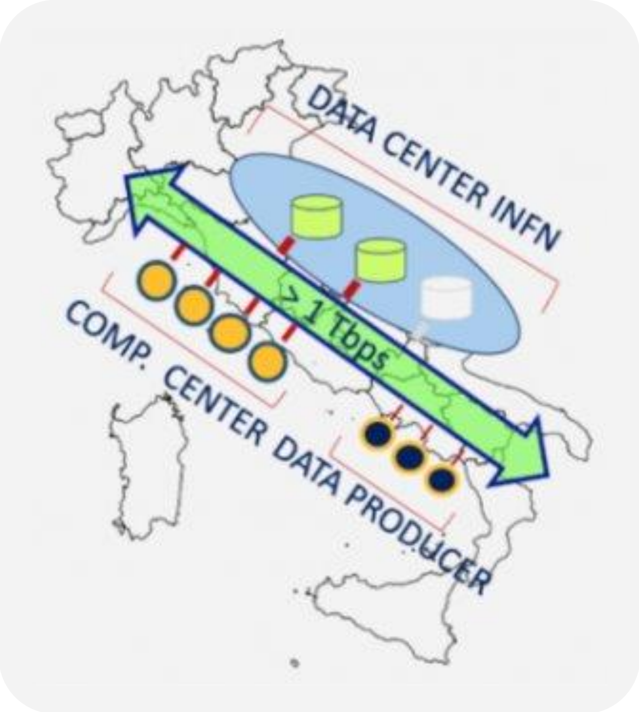


The analysis of the use of the pledges allocated on Astroparticle experiments (data from 2020) demonstrates **inefficiency** and **discontinuity** as well as a considerable **heterogeneity** of requests.

3. INFN Cloud project

The INFN-Cloud project in a nutshell

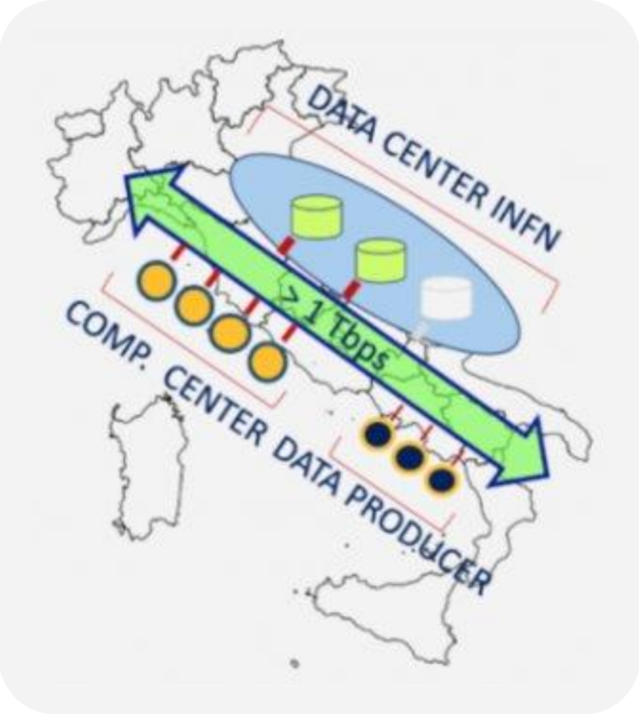
- An **internal effort** at the **INFN** level to **manage** a (large) **fraction of the INFN resources**, in order to **decouple user needs** from the **availability of local and dedicated hardware**.
- An attempt to **rationalize** the **access** to **hardware**, and **optimize** its use
- From "1 GPU on each desk, used 5% of the time" to "**shared resources optimally used**"



For **more details** you are invited to follow the talk of **Davide Salomoni** on **Mar 25 at 12:00 PM**

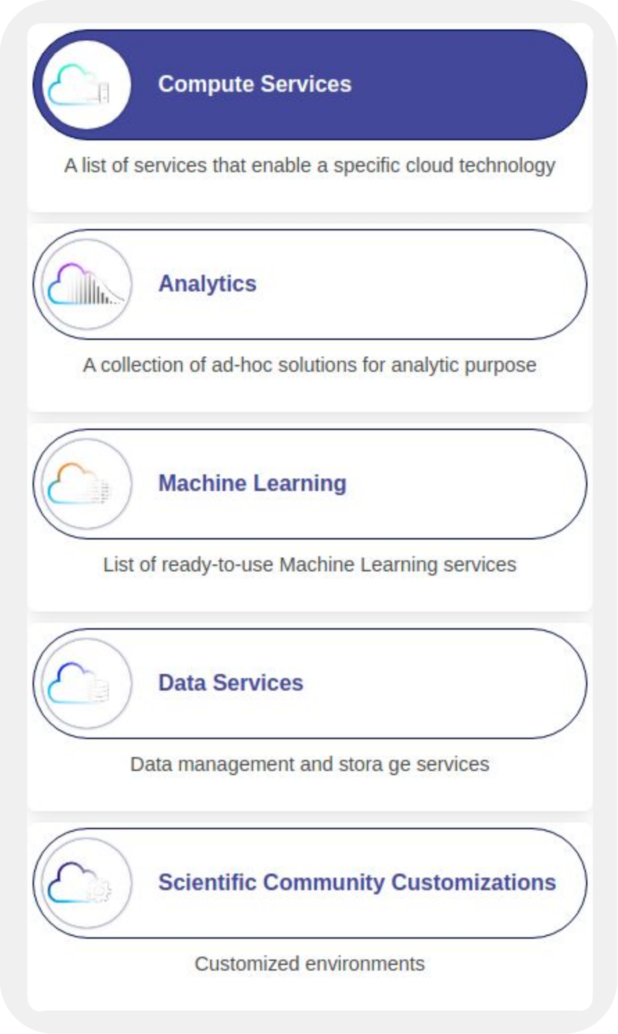
The INFN-Cloud project in a nutshell

- 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.



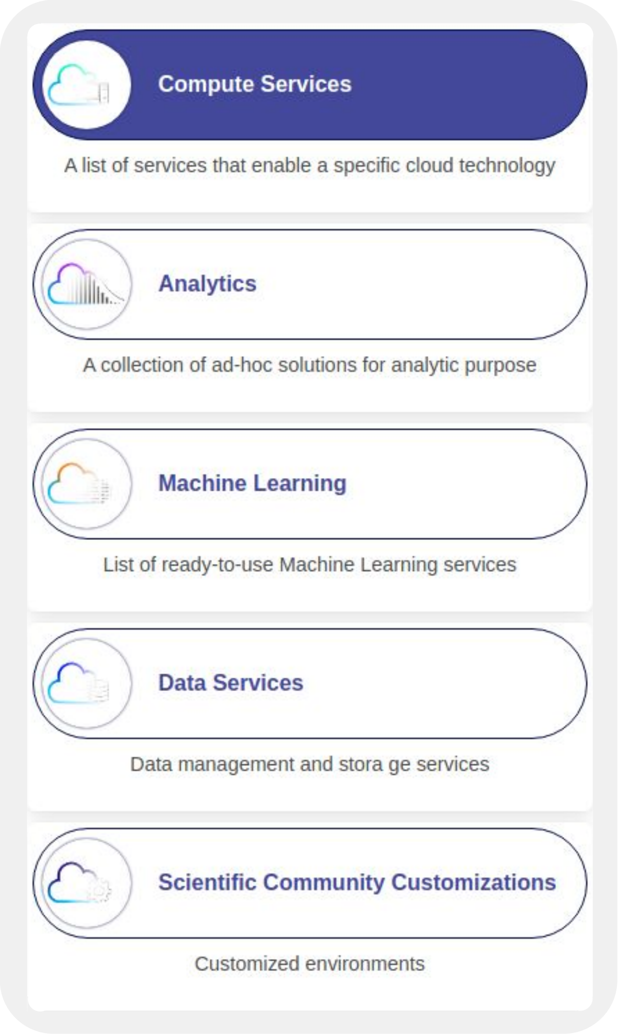
From user’s perspective:
INFN-Cloud services

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



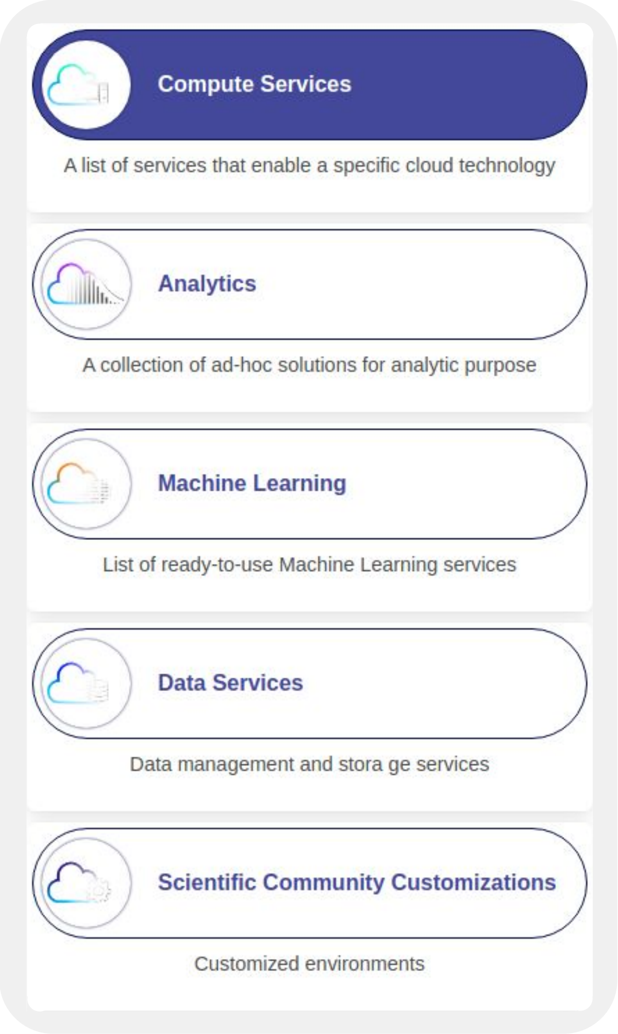
From user's perspective:
INFN-Cloud services

- **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.



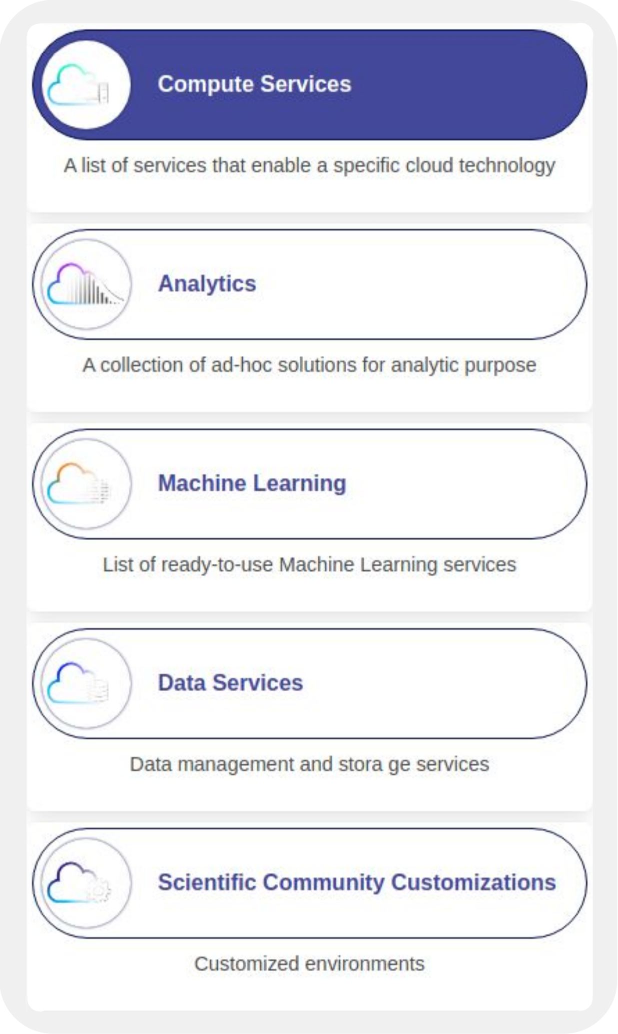
From user’s perspective:
INFN-Cloud services

- **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)



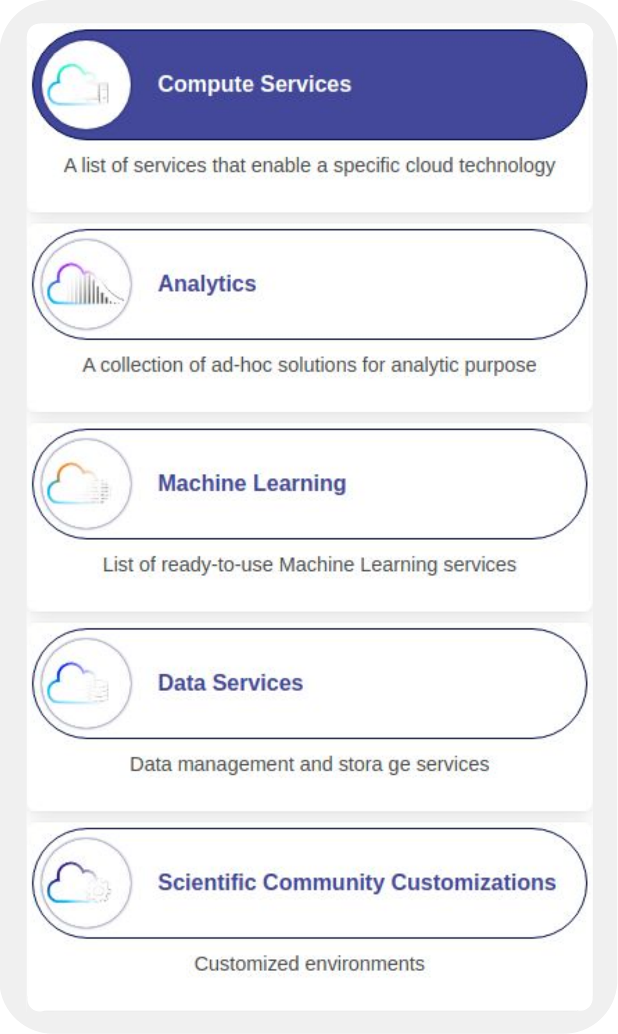
From user’s perspective:
INFN-Cloud services

- **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



From user's perspective:
INFN-Cloud services

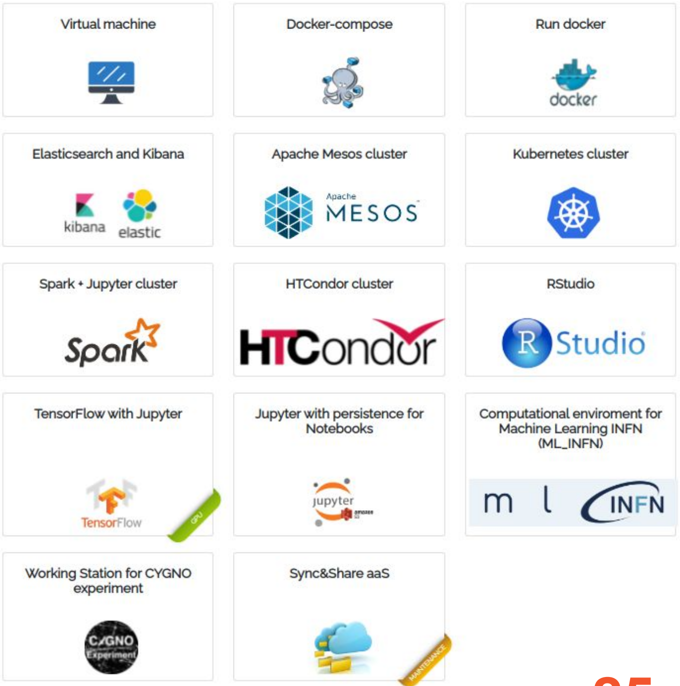
- **Dynamic Clusters even designed and tuned taking into account the specific communities needs;**
  - **HTCondor batch system; environment optimized for ML i.e. equipped with GPUs**
  - Container orchestrators such as K8s and Mesos
- User-level disk encryption **to manage confidential data**
  - **Certified Cloud IEC/ISO 27001 at CNAF**



INFN-Cloud: High Level composable service strategy

- The mantra is: “**No one size fits all solutions! Flexibility** and **composability** are **key** to **support user** tailored configurations”
  - The ultimate objective is to **provide solutions for a wide rage of users/communities**

- **The actual scenario is:**
  - a **set of pre-developed 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.**



… a high level mechanism for services composition

There is a **huge set of tools and solutions available**, but **there is NOT a one-size-fit-all solution**

- INFN-Cloud provides extensible building blocks

Each button of the dashboard is made of 2 main technical pillars :

- TOSCA Template & Types
- Ansible Roles

The INFN-Cloud mechanisms allows to **develop your "preferred" custom types** or, even better, to reuse already developed types

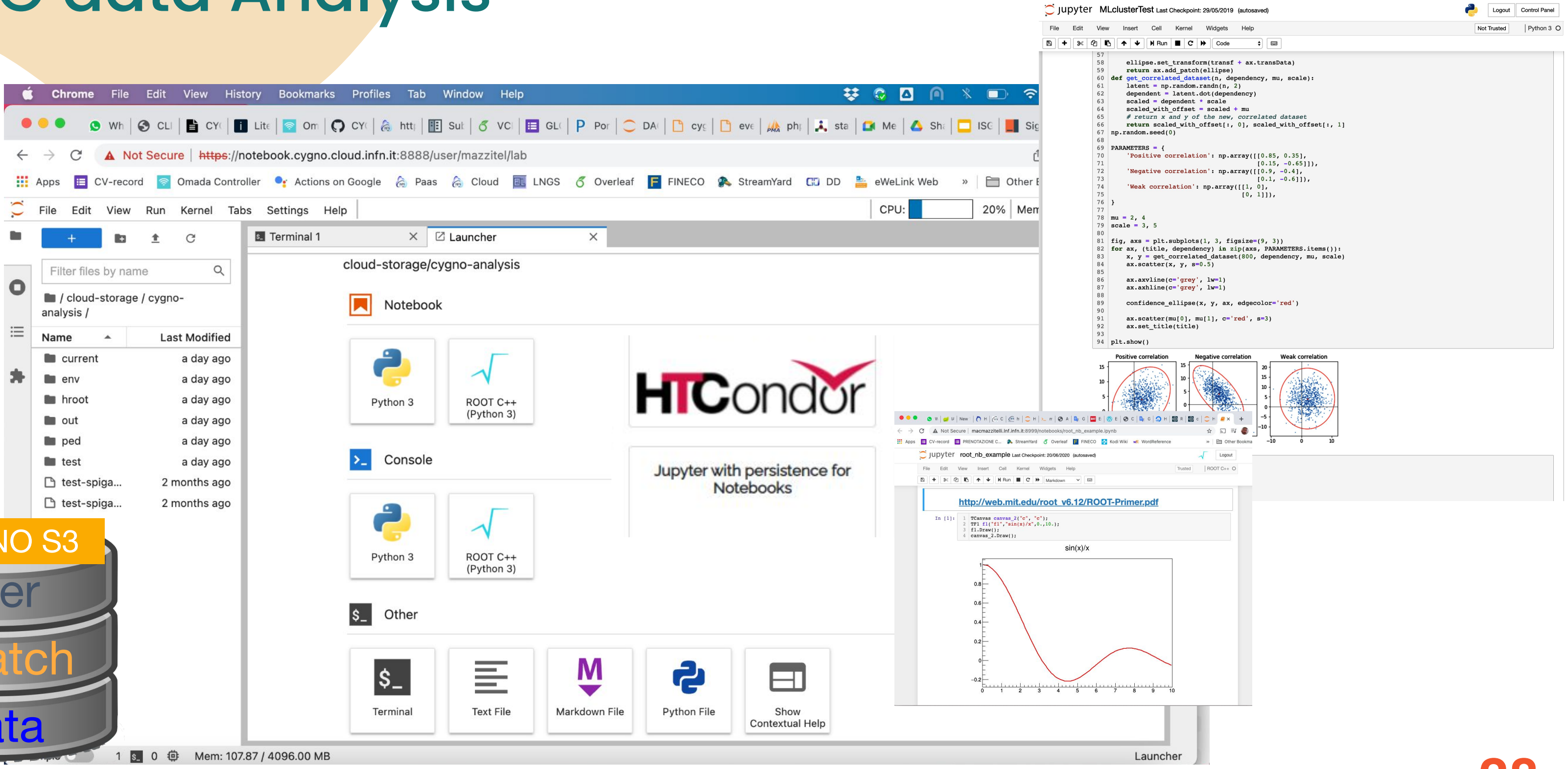
- custom type implement the binding with Ansible.
- you can reuse any available Ansible role, or to develop your own.

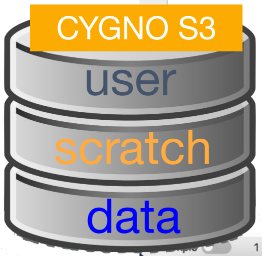
Everything is "embedded" into a **dedicated TOSCA template.**

- This enables a very convenient paradigm for the end user: the declarative approach. This is what allows the user to care about the WHAT instead of the HOW

4. CYGNO use case

SaaS - Jupyter Web App for CYGNO data Analysis





Middleware proposal

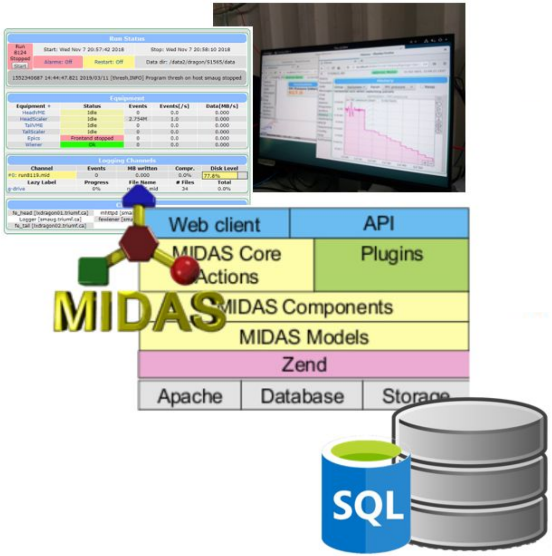
Develop a framework, hosted in the INFN Cloud, capable of **processing the data** acquired by the detector and the sensors, **delivery quality information** about the runs **online**, **send** all the information to a **Database** and show the analyzed information with **Dashboards**.

Middleware proposal

Develop a framework, hosted in the INFN Cloud, capable of **processing the data** acquired by the detector and the sensors, **delivery quality information** about the runs **online**, **send** all the information to a **Database** and show the analyzed information with **Dashboards**.

In other words, the idea is to have the **hole data pipeline hosted in the INFN-Cloud**, in a way that it is **possible** to **scale** the **computational/storage needs** when it will be necessary.

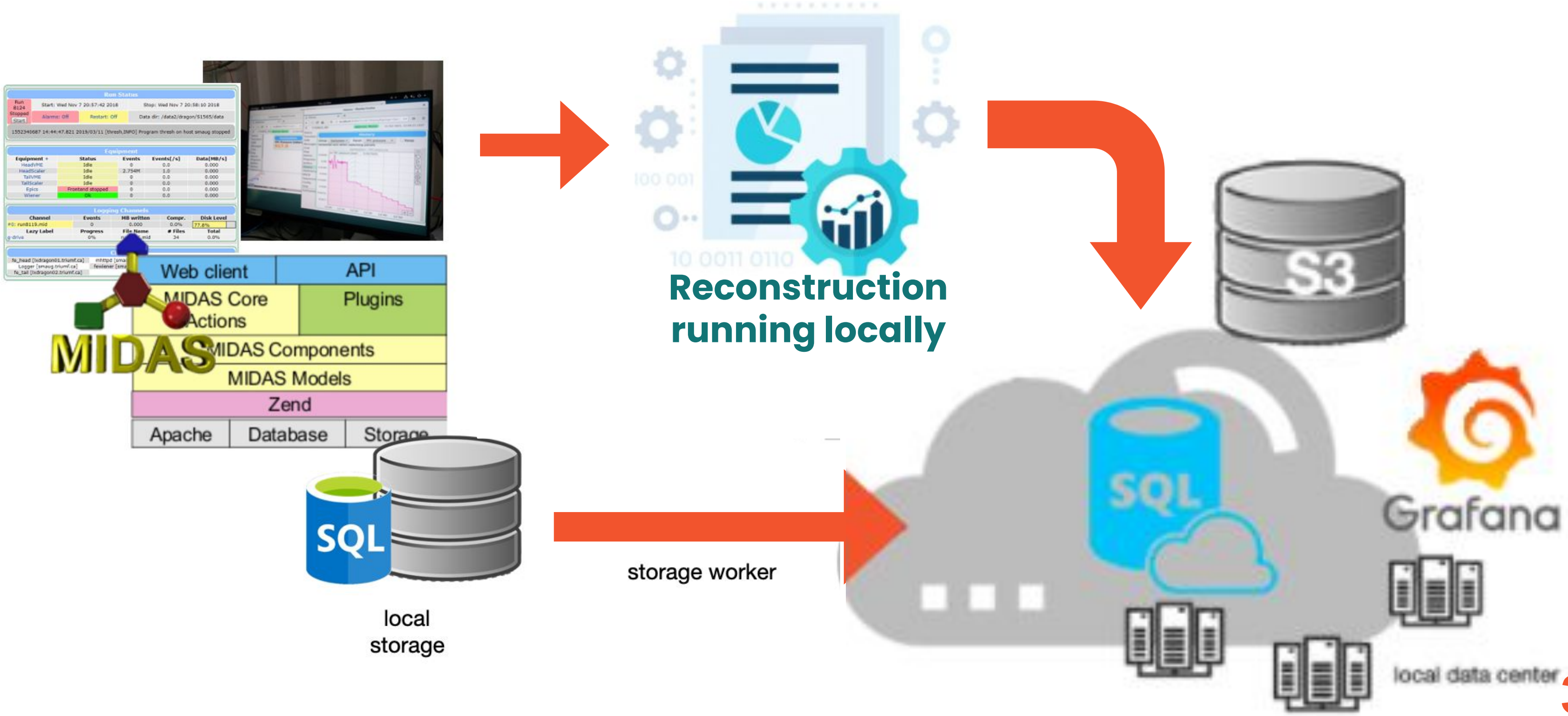
data flow - schematic view
from detector to cloud



DAQ system of the CYGNO detector, that uses MIDAS framework to control it.

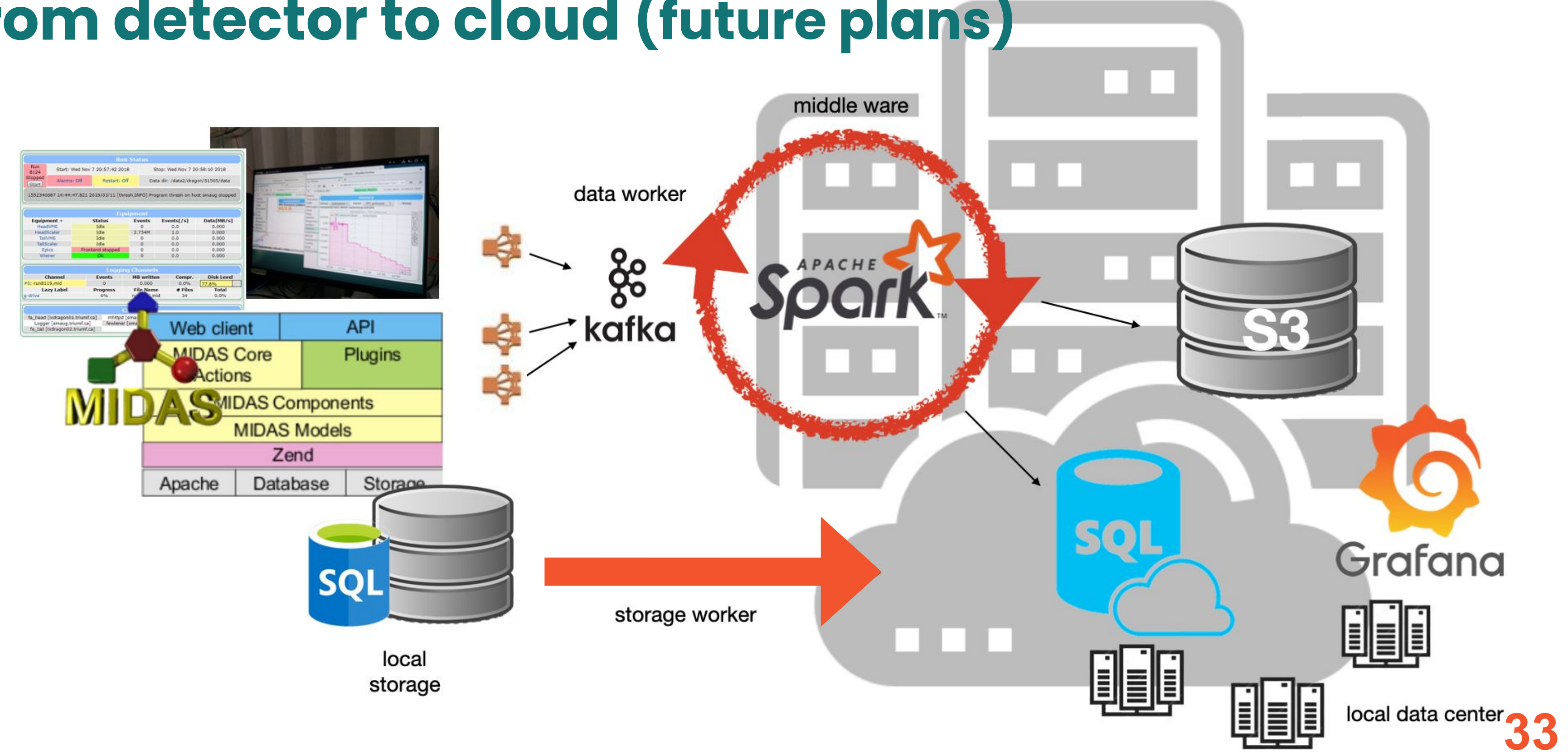
MIDAS is a modular and generic framework for online distributed analysis of data streams, agnostic with regard to the type of data streams.

data flow – schematic view
from detector to cloud (today's status)



data flow - schematic view
from detector to cloud (future plans)

INFN cloud storage
interactive and batch data analyse and simulation



5. Conclusions

Conclusions

- In conclusion, a **Platform as a Service** (PaaS) **was created** inside **INFN Cloud** that **allows** the **experiments** to **access resources** as a **Software as a Service** (SaaS), and **CYGNO** is the **beta-tester of this system.**

Conclusions

- In conclusion, a **Platform as a Service** (PaaS) **was created** inside **INFN Cloud** that **allows** the **experiments** to a**ccess resources** as a **Software as a Service** (SaaS), and **CYGNO** is the **beta-tester of this system.**

- This **implementation** did **not show** any **conceptual** or **practical limits** that the **cloud architecture has not been able to solve**.

Conclusions

- In conclusion, a **Platform as a Service** (PaaS) **was created** inside **INFN Cloud** that **allows** the **experiments** to **access resources** as a **Software as a Service** (SaaS), and **CYGNO** is the **beta-tester of this system.**
- This **implementation** did **not show** any **conceptual** or **practical limits** that the **cloud architecture has not been able to solve**.
- **The cloud**, beyond specific experiments, seems to be **best way** to **efficiently manage** the **resources** available, in **particular** for **small and medium-sized experiments**, now more and more frequent **not only in the world of astroparticles**.

Thank you!