INFN cloud activities for bioinformatics in Italy

Stefano Nicotri

INFN - Istituto Nazionale di Fisica Nucleare - Sezione di Bari, Italy

ReCaS-Bari Cloud Infrastructure

The IaaS (Infrastructure as a Service) cloud platform **PRISMA-Cloud@ReCaS**, hosted at the **ReCaS Bari** data center, provides **infrastructural computing resources** following the cloud computing paradigm.

Its main features are:

- 1150 CPU core
- 5 TB of RAM
- 10 Gbit/s network
- Layer 2 isolated VLAN with NAT
- Evolved applicative firewall
- 180 TB of replica 3 storage

- Based on OpenStack
- modular
- Highly Available (HA) services

The IaaS cloud platform @ INFN Bari / UNIBA

- Resources (instances, or virtual machines, VM) can be used to develop and deploy software systems;

- It is possible to create **resilient systems** with high-availability using multiple instances (together with services provided by the IaaS infrastructure, as load-balancing and auto-scaling)

- Virtual instances are very similar to traditional hardware servers:
  - They use familiar Operating Systems (OS), as Linux, Windows, etc.
  - Any software compatible with the OS can be executed on them
  - Associating a public IP to the VM it is possible to interact with it through standard methods (ssh, RDP,…)

Image Service and Marketplace

- Pre-configured virtual images (*templates*) can be used to create virtual machines of different kinds (*flavor*) depending on the RAM and CPU required by your application.

- A certain number of templates (software configuration) is already available from the catalog, but the user can upload her/his own (also starting from *snapshots* of her/his own VMs).

Virtual Appliance
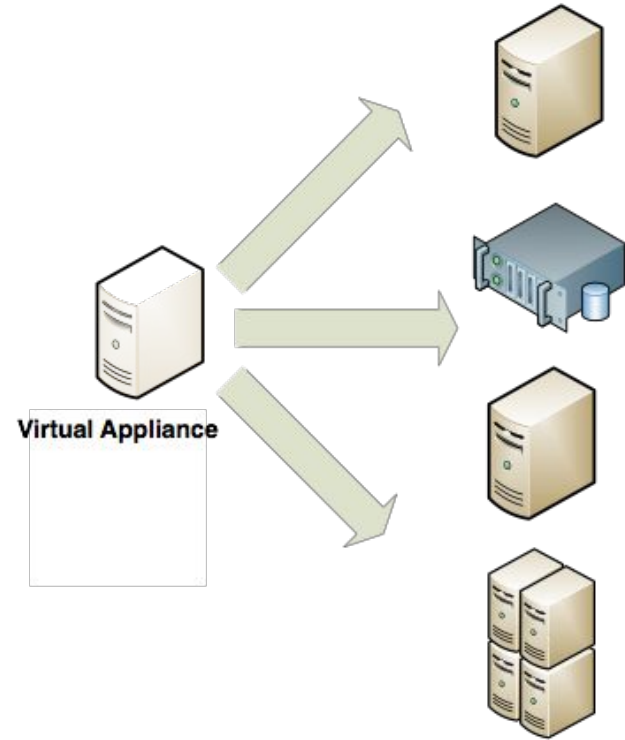
Projects, Groups And Activities
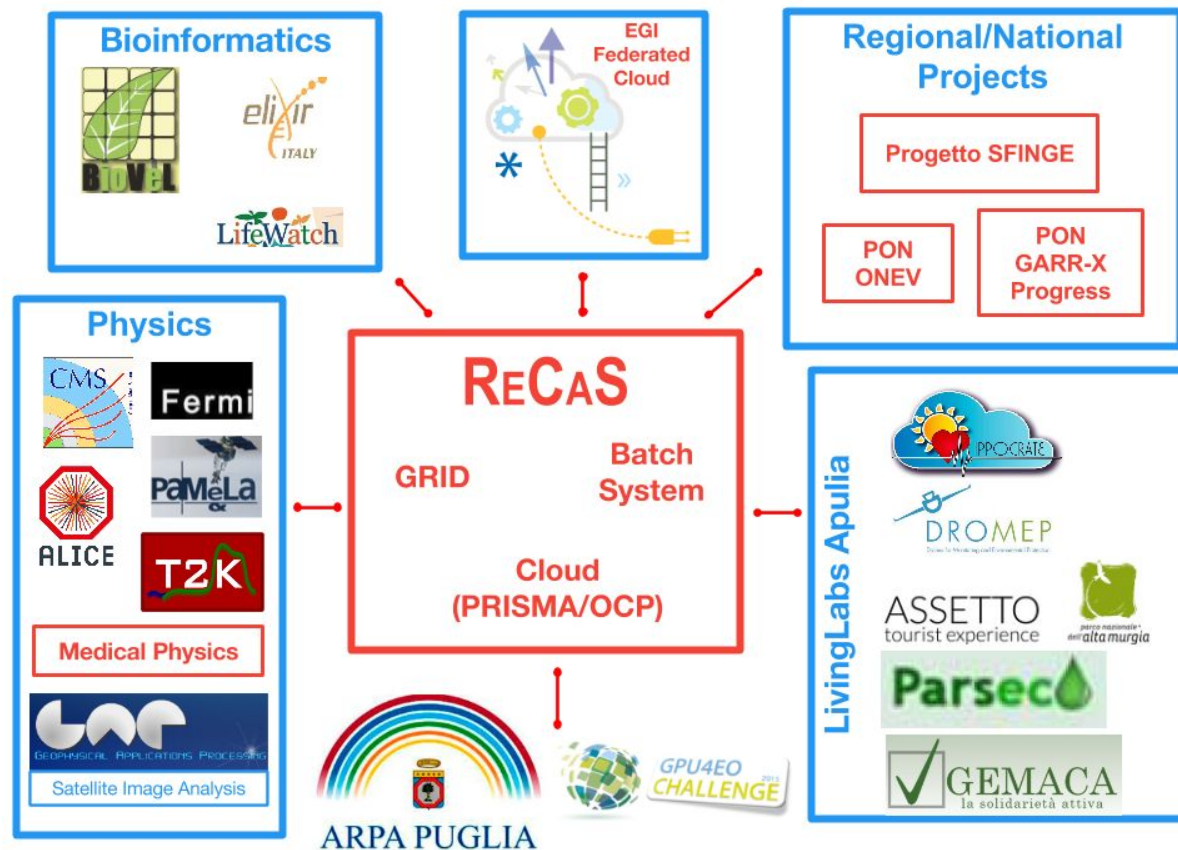
Projects, Groups And Activities
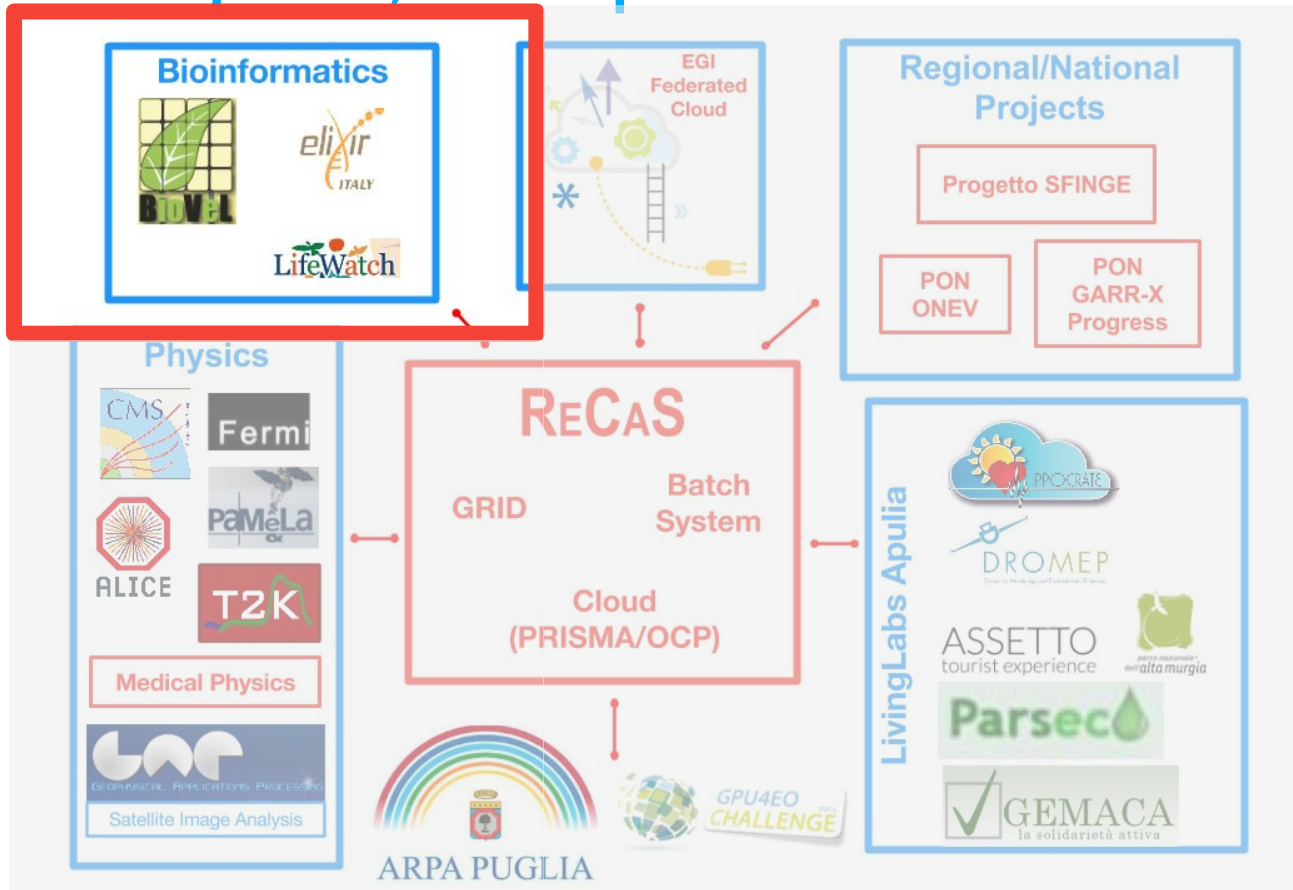
Why Cloud For Bioinformatics And Training ?

- Easier to deploy common tools (SFTP servers, analysis tools)

- Elasticity and scalability

- Easier to manage and share data

- Easier to adapt infrastructure to needs of classroom (number of users, tools, etc)

- Optimized usage of resources

Available Tools / Experience For Bioinformatics

Workflow Management Tools

LONI Pipeline
Taverna
Galaxy (web based)

Analysis Tools

MrBayes, Blast, ITK, FSL, GSNAP, BioPython, R ,Tango, Bowtie …

Applications

BioMaS (Bioinformatic analysis of metagenomic ampliconS)
MSA-PAD (Multiple DNA Sequence Alignment framework)
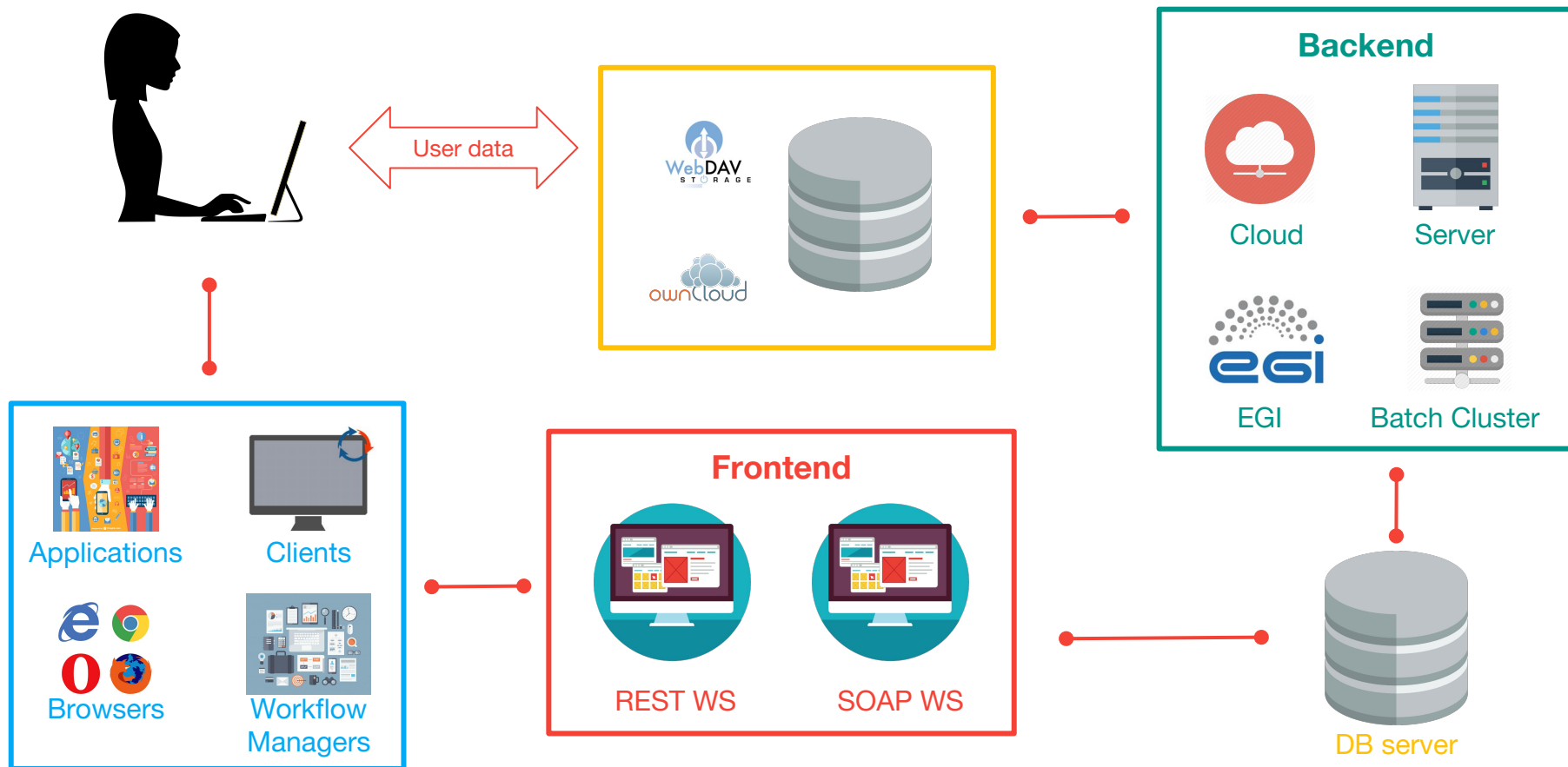
BioVel portal

Evolution models
Phylogenetic Inference
Metagenomics analysis
Analysis chains developed by the project available for users

ReCaS Science Gateway

Simple access to grid/cloud resources and applications (based on Liferay)

All Governed By the Job Submission Tool (JST)

Future Perspectives - INDIGO

The European project **INDIGO DataCloud** is developing an open source data and computing platform targeted at scientific communities, deployable on multiple hardware and provisioned over hybrid, private or public, e-infrastructures:

- flexible data sharing across groups & infrastructures
- multiple sources and storage locations
- transparent network interconnections for distributed computing and storage resources
- dynamic and complex workflow management

Among the supported use-case of INDIGO there is an on demand one-click scalable Galaxy installation.

INDIGO is based on Docker, Apache Mesos and OneData to manage data and application in an easy and flexible way

ONEDATA

Open source storage solution for integrating access to your data from various providers

MY SPACES IN ONEDATA

EXPERIMENT DATA

MY SIMULATIONS

PUBLICATIONS
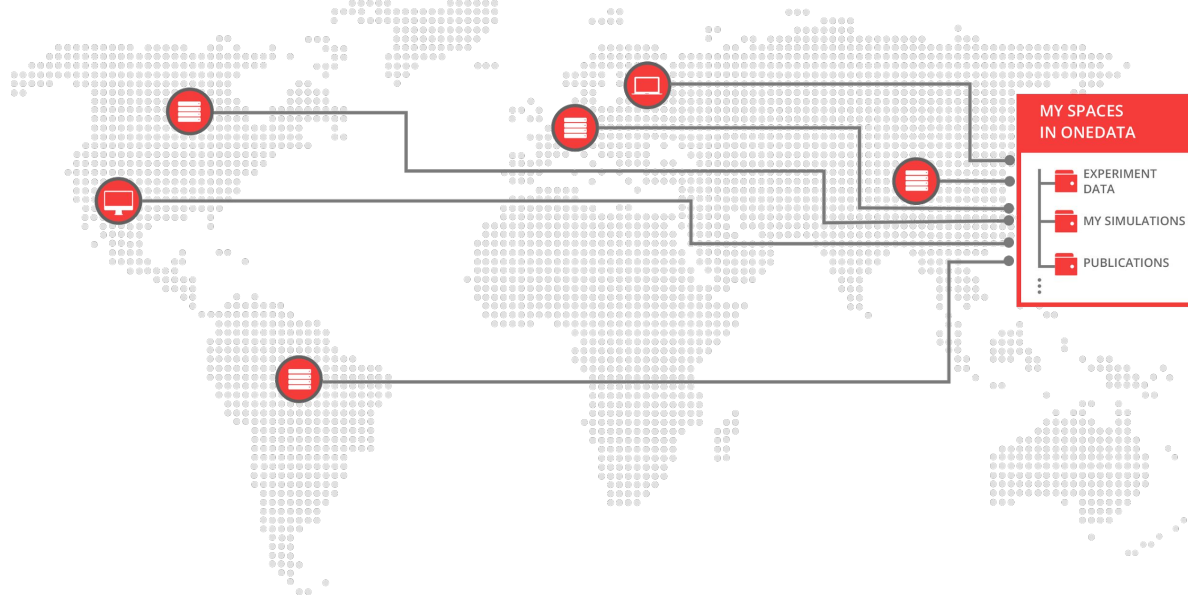
team can easily share and process data on large scale infrastructures with the desired security level

People

M. Antonacci, D. Diacono, G. Donvito,

R. Gallitelli, R. Gervasoni, F. Giannuzzi,

A. Italiano, G. Maggi, A. Monaco,

SN, M. Perniola, V. Spinoso,

M. Tangaro, R. Valentini

Thank you
for
your attention

Links For Demo

Bio Class Heat template

Bio Class stack environment

RStudio class user data

Docker-Mesos Video

Backup

Iaas - Key Elements

IaaS Infrastructure

Public IAAS | Hybrid for Public Administrations | Private

Security and privacy

Storage encryption

Evolved Firewall and VLAN

Geographic and dynamic VPN

Geographic disaster recovery

Performance and reliability

Enterprise Open IaaS infrastructure

Integrated multi-level monitoring infrastructure

IaaS services

Continuous management of services

Deduplicated storage

Geographic High Availability for services

Interoperabilità e federazione

Federation of different IaaS

Federated authentication systems

Interoperability between Open and enterprise platforms (Microsoft/Vmware)

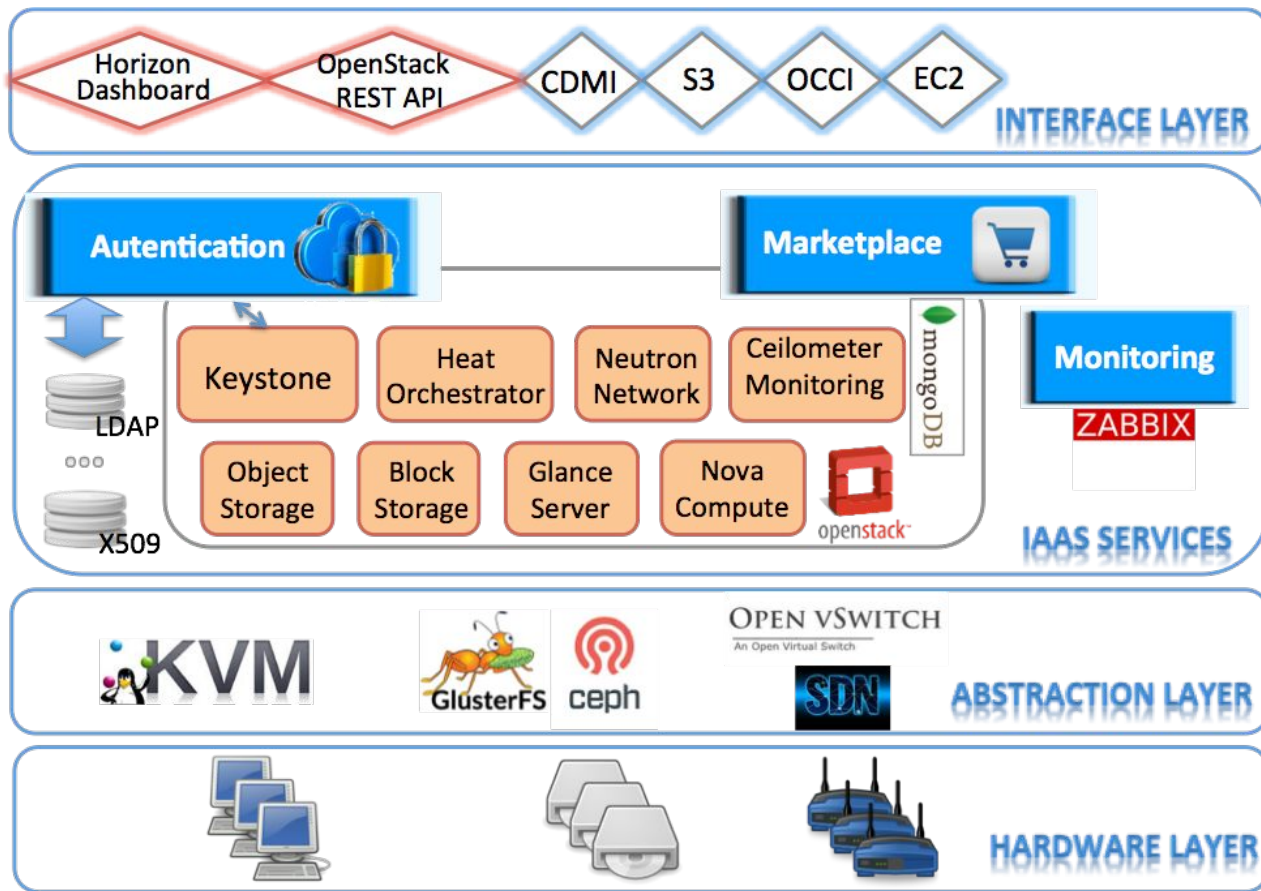Image repository and advanced contextualization of services

Complex Orchestration of IaaS services

Standard IaaS APIs: EC2/S3/OCCI/CDMI

Logical Architecture

Physical Architecture