Case Study and INDIGO solution for:

ELIXIR-ITA: Galaxy as a Cloud Service

RIA-653549

Presented by Marco Antonio Tangaro
IBIOM-CNR, ELIXIR-IIB
ma.tangaro@ibbe.cnr.it

INDIGO SUMMIT

EGI-INDIGO workshop on community application support

Catania, 10th May 2017

INDIGO-DataCloud is co-founded by the Horizon 2020Framework Programme

ELIXIR-ITALY: Galaxy as a Cloud Service

MOTIVATION

Galaxy is a workflow manager adopted in many life science research environments in order to facilitate the interaction with bioinformatics tools and the handling of large quantities of biological data.

Currently, it can be deployed in three ways: public servers, local servers and commercial cloud solutions.

Over 2400 Galaxy cloud servers launched in 2015: the demand for cloud solutions is rapidly growing, allowing for the creation of a ready-to-use Galaxy production environment avoiding initial configuration issues, requiring less technical expertise and outsourcing the hardware needs. Relying on commercial cloud providers is quite costly and can pose ethical and legal drawbacks in terms of data privacy.

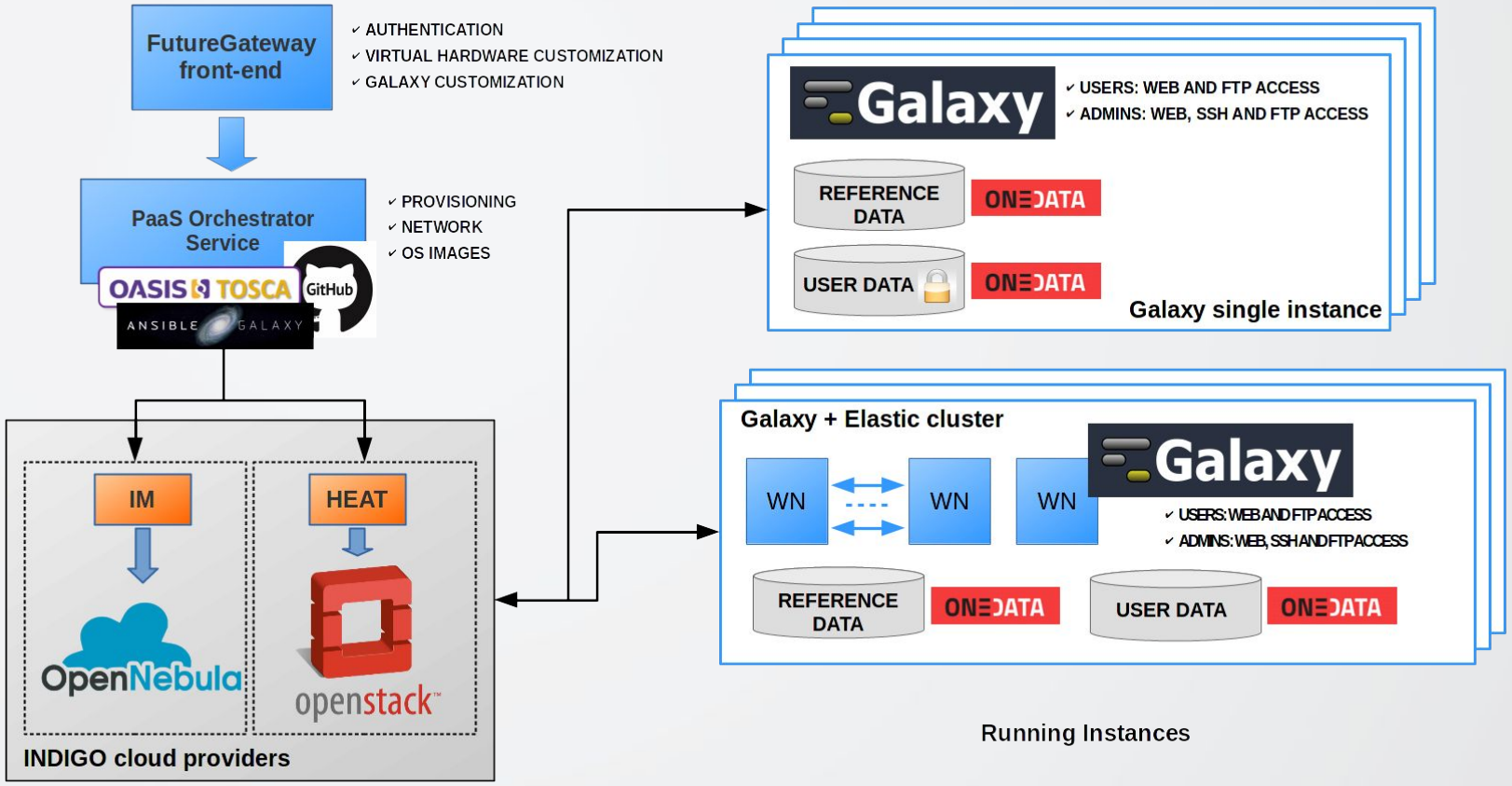
ELIXIR-ITALY in the framework of the INDIGO-DataCloud project is developing a cloud Galaxy instance provider, automating the creation of Galaxy-based virtualized environments exploiting the software catalogue provided by the INDIGO-DataCloud community.

ELIXIR-ITALY: Galaxy as a Cloud Service

- **Research Community: ELIXIR-ITALY**
  - Topic/Area: Bioinformatics & Life Sciences
- **Teams involved**: ELIXIR-ITALY, CNR-IBIOM, INFN-BARI.
- **Objective of the Case Study:**
  - **Develop a Galaxy instance provider platform, allowing to fully customize each virtual instance through a user-friendly web interface, ready to be used by life scientists and bioinformaticians.**
  - Schedule: Features complete version by second quarter of 2017. Production quality by the third quarter 2017.
- **Innovation challenge:**
  - Different software to be installed and configured to deploy the Galaxy production environment.
  - Galaxy customization (tools and reference data availability).
  - Virtual instance isolation.
  - Elastic cluster support.
- **Final user community**: research groups (bioinformatics community), institutions and SMEs.
- **Impact**: Easy access to computational infrastructural resources to end-users and speed up the deployment of complex analysis workflows over large datasets.

Architecture

Integrating different INDIGO-DataCloud technologies to automatically deploy a ready-to-use Galaxy production environment.



- Instance and Galaxy customization web front-end (FutureGateway).
- Full Galaxy production environment deployment (Orchestrator and Infrastructure Manager).
- Persistent storage (Onedata or Iaas block storage with filesystem encryption)
- Tools and reference data availability (Orchestrator and Onedata).
- Automatic elasticity (CLUES).
- Authentication service (FutureGateway and IAM).

INDIGO FutureGateway

FutureGateway is a key component in our solution:

1. *The web front-end is designed to grant user friendly access to this service, allowing to easily configure and launch each Galaxy instance.*
2. *Provides different tabs for each configuration task: virtual hardware configuration, Galaxy configuration, tools configuration and a dedicated section for the elastic cluster support.*
3. *The instance administrator credentials are provided by the users during the configuration phase.*
4. *Each Galaxy instance will then be automatically configured according to the user preferences.*
5. *The service provides support for virtual clusters through a dedicated section of the web front-end and allows to instantiate Galaxy with SLURM as Resource Manager and to customize the number of virtual nodes, nodes and master virtual hardware.*

Each instance thus provides a work environment tailored on each user by the users themselves, supporting research, clinical and educational applications, facilitating the access to computing resources and simultaneously empowering non-bioinformaticians and non-IT experts with the possibility to access, manage and analyse large volumes of biological data.

We are ready to share our experience!

CONTACTS:

- **Marco Antonio Tangaro (INDIGO champion) ma.tangaro@ibbe.cnr.it**
- Giacinto Donvito (INFN) giacinto.donvito@ba.infn.it
- Graziano Pesole (head of ELIXIR-ITA Node) g.pesole@ibbe.cnr.it
- Federico Zambelli (ELIXIR-ITA technical coordinator) f.zambelli@ibbe.cnr.it

https://www.indigo-datacloud.eu

Better Software for Better Science.

Demo video *(3')*

- ***Make it so that it can integrate logically in the chain of a video with slide 2, slide 3, then slide 4 including this DEMO, then finish slide 4.***
- ***Show the user interface***
- ***Show how to launch the application (simplified), and where***
- ***Show how the application is executed if relevant***
- ***Show the output***

Applications vs INDIGO solutions

(x=implemented, x=being implemented, proposed for SUMMIT New/Mature tracks)

INDIGO - DataCloud

	P0a	P0b	P1	P2	P3a	P3b	P4	P5	P6	P7	P8	P9	P10	P11	P12
IAM	X	X	X	X			X		X		X	X	X	X	
Udocker		X					X								X
Orchent		X													
Ophidia	X				X			X				X			
FutureGateway	X			X				X			X				
Kepler								X				X			
Mobile		X													
OneData	X			X	X	X		X	X	X	X	X	X		
QoS	X														
Orchestrator	X		X				X			X	X			X	
Infrastructure Manager			X	X			X			X	X			X	