BITS :: Call for Abstracts 2019 - Oral communication

Type	Oral communication
Session	Big Data: Storage, Analysis and Visualization Biological Databases
Title	Laniakea@ReCaS: an ELIXIR-IT Galaxy on-demand cloud service
All Authors	Tangaro M A(1), Donvito G(2), Antonacci M(2), Chiara M(3), Mandreoli P(3), Pesole G(1,4), Zambelli F (1,3)

Affiliation

(1) Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies - National Research Council (IBIOM-CNR), Bari, Italy
(2) National Institute for Nuclear Physics, Bari Section, Italy
(3) Dept. of Biosciences, University of Milan, Italy
(4) Dept. of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari "Aldo Moro", Italy

Motivation

Galaxy is currently the prevailing workflow manager for bioinformatics thanks to its many useful features and a user-friendly interface. While several Galaxy public services are available to researchers, either general purpose or dedicated to specific research domains, there are still many scenarios where a private Galaxy instance is necessary or preferable, including for example heavy data analysis workloads, data privacy concerns or specific customization needs. However, deploying and maintaining a production-grade Galaxy instance usually requires an adequate IT infrastructure, which is not always readily available or accessible to many researchers or other potential users like bioinformaticians in clinical settings.
Cloud computing offers an unprecedented opportunity to tackle this issue, providing a convenient entry point to robust and scalable computational resources while avoiding the onerous deployment and maintenance of local hardware and software infrastructure. We present Laniakea@ReCaS, an ELIXIR-IT Galaxy on-demand cloud service that comes with the advantages of a public Galaxy server, retaining at the same time the flexibility and control over the data that only a private Galaxy instance can provide.

Methods

The ELIXIR-IT Laniakea@ReCaS Galaxy on-demand service leans on the Laniakea [1] software platform. Laniakea is primarily based on the Platform as a Service (PaaS) layer developed by the INDIGO-DataCloud H2020 project, focused on making cloud e-infrastructures more accessible by scientific communities [2]. The PaaS layer provides the orchestration services and their integration with the Authentication and Authorization Infrastructure, automating the deployment, configuration, software installation, monitoring and update of the virtual Galaxy servers. The repository of reference datasets is shared by all the Galaxy instances hosted by the Laniakea@ReCaS service, to avoid useless and costly data duplication, through the CERN-VM read-only filesystem. LUKS (Linux Unified Key Setup), the current standard for encryption on Linux platforms, is used for the provisioning of encrypted volumes to store users' data. Galaxy instances provided by Laniakea are supported by the full set of auxiliary applications recommended for a Galaxy production-grade environment, i.e., PostgreSQL, NGINX, uWSGI, and Proftpd already configured to work with Galaxy. The ReCaS cloud platform that hosts the service is based on OpenStack (Mitaka), with about 1700 CPU cores, 6.7 TB of RAM and 270 TB of storage (replica 3).

Results

The user interacts with the Laniakea@ReCaS service through a web front-end that allows to configure and launch a production-grade Galaxy instance in a very straightforward way. Through the interface, the user can deploy Galaxy instances over single VMs or virtual clusters, link them to shared reference data volumes, and choose between plain or encrypted volumes for data storage. A selection of "flavours", that is Galaxy instances pre-configured with sets of tools for specific tasks, is available. Once the user is satisfied with the configuration, Laniakea deploys the Galaxy instance over the ReCaS cloud infrastructure, provides a public IP for its access, and hands over to the user full administrative privileges on the new instance and the underlying virtual hardware.

In December 2018 Laniakea@ReCaS announced its closed beta program to test the maturity of the service. The program involved the participation of twelve users from several research institutions and universities and different scientific backgrounds. The users were asked to stress-test Laniakea@ReCaS by deploying, deleting and extensively using one or more virtual Galaxy instance for their daily research activities. During the beta program, we gathered as much information from our users as possible, worked in fixing the juvenile issues of the service and prioritized a list of features for future developments.

The production phase of the ELIXIR-IT Laniakea@ReCaS service is foreseen to launch in the second half of 2019. Access to the service will be offered on a per-project basis through an open-ended call defining terms and conditions in a fashion similar to the ongoing ELIXIR-IT HPC@CINECA service. Each project proposal will be evaluated by a scientific committee and a technical board. Successful proposals will be granted a standard package of computational resources for running Galaxy instances with Laniakea for an amount of time compatible with the project requirements.

Info	
References [1] Tangaro M.A. et al, Laniakea: an open solution to provide Galaxy "on-demand" instances over heterogeneous cloud infrastructures, bioRxiv 472464; doi: https://doi.org/10.1101/472464 [2] Salomoni, D., Campos, I., Gaido, L. et al. J Grid Computing (2018) 16: 381. https://doi.org/10.1007/s10723-018-9453-3	
Figure	
-	
Availability	https://elixir-italy-laniakea.cloud.ba.infn.it
Corresponding Author	
Name, Surname	Marco Antonio, Tangaro
Email	ma.tangaro@ibiom.cnr.it
Submitted on	28.04.2019