Astronomical Data Analysis Software and Systems: XXIV
ASP Conference Series, Vol. 495
A. R. Taylor and E. Rosolowsky, eds.

Cloud Computing for Astronomers on Top of EGI Federated Cloud

Giuliano Taffoni, Claudio Vuerli, and Fabio Pasian

INAF - Osservatorio Astronomico di Trieste - via Tiepolo 11, 34131 Trieste Italy

Abstract. EGI Federated Cloud offers a general academic Cloud Infrastructure. We exploit EGI functionalities to address the needs of representative Astronomy and Astrophysics communities through clouds and gateways while respecting commonly used standards. The vision is to offer a novel environment empowering scientists to focus more on experimenting and pitching new ideas to service their needs for scientific discovery.

1. Introduction

In the last few years, the term "Cloud Computing" (CC) has been mentioned in relation to services or infrastructural resources, which can be contracted over a network. CC is a solution that provides users with services that can be drawn upon on–demand and invoiced as and when used.

CC has a variety of characteristics, with the main ones being: *shared Infrastructure* (CC uses a virtualized software model, enabling the sharing of physical services, storage, and networking capabilities), *dynamic provisioning* (CC allows for the provision of services based on current demand requirements), *network access* (CC needs to be accessed across the internet from a broad range of devices such as PCs, laptops, and mobile devices, using standards-based APIs), *managed metering* (CC uses metering for managing and optimising the service and to provide reporting and billing information).

Deploying a CC infrastructure can differ depending on requirements, they exist four deployment models, each with specific characteristics that support the needs of the services and users of the clouds in particular ways:

- Private Cloud. The cloud infrastructure has been deployed, and is maintained and operated for a specific organization.
- Community Cloud. The cloud infrastructure is shared among a number of organizations with similar interests and requirements.
- Public Cloud. The cloud infrastructure is available to the public on a commercial basis by a cloud service provider.
- Hybrid Cloud. The cloud infrastructure consists of a number of clouds of any type, but the clouds have the ability through their interfaces to allow data and/or applications to be moved from one cloud to another.

Currently, Public Clouds account for the majority of the marketed resources, but their infrastructures are not always a viable or desirable choice for scientists and other researchers. In a research context, sensitive data may have to stay within institutional boundaries, big data may be too expensive to move, or the speed and quality of long distance data transfer may be unsatisfactory. As a consequence, many research institutions are studying ways provisioning cloud and traditional services by means of virtualised resources to satisfy the evolving requirements of their researchers.

The European astronomy and astrophysics community (A&A) though the Astronet project[1], identified CC infrastructures as a crucial resource for Astronomers that offers highly flexible, scalable, on-demand computational storage and computing (Robson 2014). In particular, A&A communities are increasingly recognizing that an approach along the lines of a distributed Cloud Infrastructure is useful particularly for satellite data analysis where well-defined computing resources and software are necessary (such as specific OS vendor and version and library applications).

Europe is currently missing a cloud infrastructure dedicated to A&A, while there are some important implementations in other countries as Canada (Gaudet et al. 2010) or China (Cui 2014).

In this paper we discuss the work in progress to construct, operate, sustain and promote a unique environment supporting A&A cloud based environment through European Grid Infrastructure (EGI)[2] services and resources.

2. Federated Cloud for Research in Europe

The European Grid Infrastructure is a federation of independent national and community resource centers supporting specific research communities and international collaborations worldwide. EGI integrates 352 resource centers across 56 countries worldwide. The integrated e–Infrastructure capacity is $\sim$ 270000 CPU cores of computing power and $\sim$ 140 PB of online disk space available for $\sim$ 20000 researchers.

EGI build a federated, standards-based IaaS Cloud platform, federating together the Cloud resources provided by research institution across Europe. EGI chose a federation-based governance model, balancing out the individual freedom of participating suppliers, and the cost benefits of providing common services once instead of over and over again.

EGI has chosen the following set of standards as being part of the mandate to participate in the Cloud infrastructure federation (Figure 1):

- OCCI: a family of specifications defining access and management operations for IaaS Clouds. Highly extensible hence not limited to IaaS only.
- CDMI: A specification defining a generic management interface, metadata management, and data access protocol negotiation for Cloud storage services.
- GLUE2 and UR2 Information model specification. This information system was originally defined for academic Grid resources, and extended for Cloud resources.

[1] http://www.astronet-eu.org

[2] http://www.egi.eu

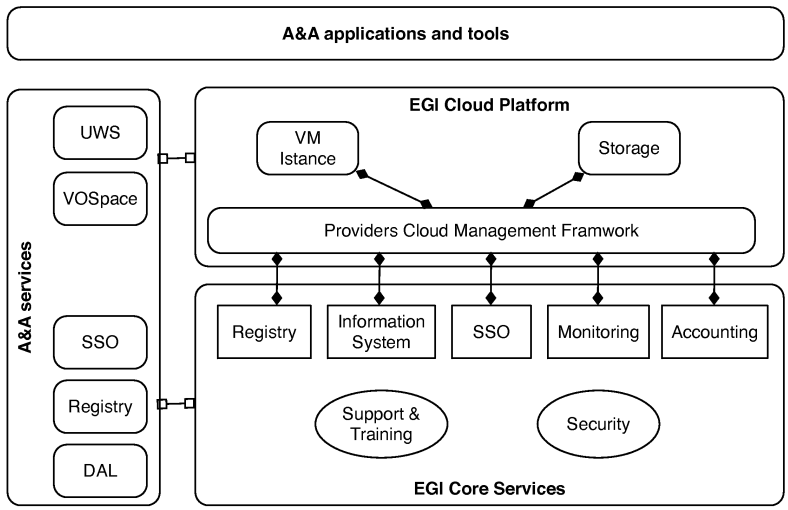

Figure 1. Description of EGI federated Cloud services and layers.

- SAML: Specification regulating details of authentication and authorisation.
- OVF: A specification for a container structure that includes binary VM images and basic deployment and contextualisation instructions. Starting point for managing virtual appliances.

EGI Federated Cloud (EFC) provides a generic infrastructure available for all the scientific communities, that offers generic services to customise to specific community needs.

3. Prospectives for Astronomy and Astrophysics

EFC can be used as the foundation for a A&A specific cloud environment. At threshold level a set of core functionalities together with flexible customization tools can be implemented and subsequently offered to A&A communities to allow them to easily construct an environment for collaborations and international projects wrapped around EGI infrastructures and services. This environment will build upon clouds and science gateways and will respect commonly used IVOA[3] standards (see Figure 1).

In particular our development activities focus on three main aspects: Authentication and Authorization, Storage Management and Computational Tasks Management.

For what regards access to storage systems, we plan to develop a thin layer on top of the EGI CDMI service to interface the EFC storage manager with a VOSpace RESTful service. In this way, IVOA application and services will be able to store, access and exchange data within the EFC.

EFC services have restricted access. The access to the service is secured by the X.509 Certificates with VOMS extensions or by SAML specifications (e.g. Shibboleth).

[3]http://www.ivoa.nte

Moreover when a user makes use of a remote resource (e.g. a VObs resource that query data from a database) that resource may in turn need to access third-party resources (e.g. data repositories in the Cloud environment) on behalf of the user in order to complete the task. Such access may possibly span across multiple security domains. In this scenario, it is necessary to delegate (parts of) the users rights to the remote resource such that it in turn can access the necessary third parties. To benefit of the use of a federated cloud environment IVOA services and applications should adopt an authentication and authorisation mechanism based on SAML specifications and a credential delegation protocol to allows a client programme to delegate a user's credentials to a service such that that service may make requests of other services in the name of that user.

We use science gateways technology based on gUSE/WS-PGRADE to develop web access to scientific applications. Deploying portals and science gateways can significantly facilitate transparent access to Cloud Infrastructures providing scientists with high-level environments that hide appropriately technical details exposing only the science-specific parts of the applications to be executed on Cloud resources. gUSE/WS-PGRADE is a framework to build science gateways and web applications based on workflow engines that can be natively configured to access different kind of computing infrastructures including CC infrastructures. Science gateways will also guarantee access to a VM market place that Astronomers can use to upload and deploy their own VM Images.

3.1. Discussion

We present an on-going activity to develop a A&A cloud environment in Europe. This environment allows to support scientific workflows and operational procedures related to data processing, code execution, data analysis and semi-analytical simulations. This environment is particularly useful for Data Centers and it can be used to ensure long term preservation of software and applications (any software that cannon be installed on a modern Operating System but still used by the A&A community).

We are also planning to federate EFC with the CANFAR cloud for A&A, effectively bridging EGI with the A&A virtual observatory and gaining direct access to a rich toolset of resources that can drive future developments within Europe towards catering for virtual observatories in A&A and beyond.

Acknowledgments. The research leading to these results has received funding from the FP7 project under contract no 312579 ER-flow (Building anEuropean Research Community through Interoperable Workflows and Data) .

References

Cui, C. 2014, in ADASS XXIV, edited by A. R. Taylor, & E. Rosolowsky (San Francisco: ASP), vol. TBD of ASP Conf. Ser., TBD

Gaudet, S., Hill, N., Armstrong, P., Ball, N., & et al. 2010, vol. 7740 of Proc. SPIE, 77401I

Robson, I. 2014, Astronet infrastructure raodmap update v0.4. URL `http://www.astronet-eu.org/IMG/pdf/ASTRONET_IR_v0_5.pdf`