



HAL
open science

CYCLONE: a Multi-Cloud Federation Platform For Complex Bioinformatics And Energy Applications

D. Gallico, M Biancani, C Blanchet, M Bedri, Jean-Francois Gibrat, J I A
Baranda, D Hacker, D Kourkouli

► **To cite this version:**

D. Gallico, M Biancani, C Blanchet, M Bedri, Jean-Francois Gibrat, et al.. CYCLONE: a Multi-Cloud Federation Platform For Complex Bioinformatics And Energy Applications. Cloud Networking (Cloudnet), 2016 5th IEEE International Conference on, Oct 2016, Pisa, Italy. 10.1109/Cloud-Net.2016.44 . hal-01716613

HAL Id: hal-01716613

<https://hal.science/hal-01716613v1>

Submitted on 2 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CYCLONE: a Multi-Cloud Federation Platform For Complex Bioinformatics And Energy Applications

D. Gallico, M. Biancani
Research & Development, Cloud
Infrastructure Sales
Interoute S.p.A.
Roma, Italy
{domenico.gallico,
matteo.biancani}@interoute.com

C. Blanchet, M. Bedri, J.-F. Gibrat
CNRS - French Institute of
Bioinformatics
Gif-sur-Yvette, France
{christophe.blanchet,
mohamed.bedri, jean-
francois.gibrat}@france-
bioinformatique.fr

J.I.A. Baranda
Internet Architectures and
Services Area (IAS)
i2CAT Foundation
Barcelona, Spain
jose.aznar@i2cat.net

D.Hacker, M.Kourkouli
Smart Utility & Energy Solutions
QSC
Cologne, Germany
{doris.hacker,
maria.kourkouli}@qsc.de

Abstract— Multi-cloud applications delivered across heterogeneous public cloud service providers poses different technical and operational challenges in terms of resource brokering, integration, implementation of security trusts, etc. In this paper we describe the federation approach taken in the H2020 CYCLONE project, which designed a multi-cloud platform integrating and orchestrating different open-source cloud and network management tools (Openstack, OpenNaaS, SlipStream and TCTP). Some real-life use-cases in the scientific area of energy optimization and bioinformatics are in use to validate the CYCLONE architecture and integrated platform, from the viewpoints of service deployment, service availability, elasticity and security. These validation activities are carried out in the CYCLONE shared testbed infrastructure, whose high-level description is also presented in this paper. Initial feedbacks on the validation of the CYCLONE federation approach will be available at the beginning of Q4-2016, when the CYCLONE use-case portfolio will be extended to increase external users' involvement and attempt validation of the federation approach from real world market

Keywords— Cloud computing, cloud federation, complex application deployment, multi-cloud deployment, OpenNaaS, OpenStack, SlipStream, bioinformatic research, renewable energy management, shared testbed infrastructure.

I. INTRODUCTION

Over the last years, the adoption of Cloud services from the IT market has grown continuously as a consolidate business solution for flexible and elastic IT services with reduced CAPEX. As a consequence, many critical workloads are being moved to the Cloud [1]. In cloud market scenario, applications spanning multiple clouds have started becoming more and more attractive for developers particularly focused on avoiding single provider's lock-in, aiming for increased scalability and fault tolerance of their applications and trying to optimize costs for their deployments (i.e. cost reduction of the cloud resources to be adopted). However, the proliferation of commercial offers, pricing schemes and services provided by different cloud service providers is posing new challenges for the developers of multi-cloud applications. The main issues are related to the decision of where to deploy resources, where to

collect all the available information on the accessed cloud infrastructures and how to establish security trusts among the different cloud infrastructure. Sustained by this market demand, Cloud Service Providers have started investing in Cloud federation solutions. Cloud federation is a strategic and commercial alliance between different cloud providers that allows cross-site cooperation regarding the deployment of services and resources. The idea behind federation is related with the sharing of physical resources with the aim of better managing the peaks of resources' requests, so that any of the partners joining the federation will be able to cope with demand variations of clients. The H2020 CYCLONE project [2] has been conceived with the objectives of defining, integrating and improving mature, open-source components such as Openstack [3], OpenNaaS [4], SlipStream [5] and TCTP [6], in order to facilitate the management of the full lifecycle of multi-cloud applications and providing complementary tools that support the configuration of the cloud application, its automated deployment onto different cloud infrastructures, and control of allocated cloud resources.

II. THE MULTI-CLOUD CYCLONE PLATFORM

CYCLONE allows Application Service Providers (ASP) to aggregate resources from multiple cloud providers (private and public) to improve the reliability of their applications through geographic redundancy, to enhance the user experience by placing services near their users, to lower barriers between cloud providers via a cloud deployment engine and to tailor intra- and inter-site networking resources, coupled with application-level security features. Application deployment, cloud networking and security constitutes the main pillars upholding the CYCLONE platform. CYCLONE integrates consolidated cloud solutions for managing software-defined networking, application deployment, cloud security and access management into a coordinated cloud action. With CYCLONE users get access to a platform that allows them to deploy their services on any supported cloud and still be able to manage it uniformly. The CYCLONE platform itself brings together these three concepts: distributed application deployment, cloud networking and cloud security to satisfy ASPs requirements:

- Application Deployment brings to CYCLONE the automated deployment of complex multi-cloud applications as well as “cloud brokering” features that allow policy-based placement and dynamic resource allocation in order to optimize the performance of cloud-based applications.
- Cloud networking, pursues to innovate at federated cloud environment by bridging network service capabilities to cloud based services, while extending the IaaS model from data centres to backbones.
- In terms of Cloud Security and Access Management, the scope is to allow multi-cloud applications to be run securely and with confidence by integrating federated identity management, application-level authentication and authorization, and end-to-end security protocols.

The novelty fostered by CYCLONE tackles some of the drawbacks that multi-cloud deployment environments still have. From the networking management point of view, the tools provided by CYCLONE will overcome the heterogeneity of network services among different cloud infrastructures, introducing fine-grain control on the network features. Application monitoring, deployment manipulation and automatic scale in and out, which are critical aspects for complex application management, will be improved by the development of the CYCLONE solution leveraging on the extension of the Slipstream’s functionalities. On the security side, the CYCLONE Federation Provider will handle the unified identity management, preventing security breaches in the end-to-end management of sensitive data.

A. OpenNaaS

OpenNaaS is an open source platform for provisioning network resources and services. It allows the deployment and automated configuration of dynamic network infrastructures and defines a vendor-independent interface to access services provided by these resources. In the scope of CYCLONE, OpenNaaS has recently incorporated the CNSMO flavour – Cyclone Network Services Manager and Orchestrator (CNSMO) [7] – which provides, configures and runs network services using the OpenNaaS concept and abstractions in overlay topologies and enables to deploy them as part of applications and application providers’ suite of services. Thus, OpenNaaS OpenNaaS-CNSMO enables the deployment, configuration and execution of network services (VPN, Firewalling, Load Balancer, etc.) to be utilized in application deployments as a suite of services available for ASPs.

B. OpenStack

OpenStack is a cloud computing software platform, created primarily to be deployed as an infrastructure as a service (IaaS) solution. The technology consists of a series of interrelated projects that control pools of processing, storage, and networking resources throughout a data center. Due to the open-source nature of OpenStack, there are many components developed by as many developers, but the OpenStack community has collaboratively identified nine key components that constitute the “core” of OpenStack, which are distributed as a part of any OpenStack system and officially maintained by

the OpenStack community. The increasing adoption and support of OpenStack by the computing industry and the academic world, has pushed the Consortium to use it as the main IaaS platform for development and integration. In particular CYCLONE will provide an ad-hoc extension to the Neutron component in order to address the requirements imposed by CYCLONE in terms of cloud networking.

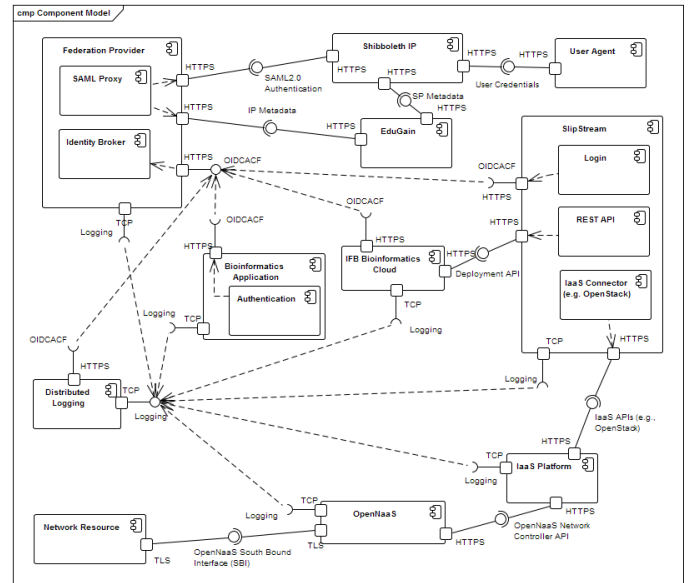


Fig. 1: CYCLONE platform high level architecture

C. SlipStream

SlipStream is a multi-cloud application management platform that is used by CYCLONE to handle the deployment and runtime optimization of (multi-)cloud applications. SlipStream provides a general cloud application model and an associated abstract API to hide differences between cloud service providers from the cloud application developers and operators, making cloud portability and multi-cloud applications a reality. The SlipStream application model consists of “Images” and “Deployments” [8]:

- An image describes a single virtual machine containing information about its operating system, installed packages, configuration, and resource requirements.
- A deployment brings together machine images to create a full cloud application, for example binding a database, business logic, and web frontend into an n-tier web application.

Images (and thus deployments) can be parameterized, making them more generic and allowing them to be customized by the cloud application operator at deployment time.

D. TCTP

The Trusted Cloud Transfer Protocol (TCTP) is an application layer encryption protocol for HTTP which provides true end-to-end security, i.e., from the user agent (e.g. browser) to the origin server (e.g. a single PaaS container), even if the communication is performed through intermediaries acting as

TLS server connection ends, e.g., cloud load balancers. TCTP encrypts and authenticates the HTTP payload using TLS at the application layer, so that in effect all headers are still accessible by cloud computing intermediaries. Encryption keys and cipher suites are negotiated by wrapping the TLS Handshake Protocol into HTTP payload and sending it through the intermediaries. The reliance on this secure handshake minimizes the risk that any intermediary intercepting these messages can act as a man-in-the-middle and compromise TCTP security. Extending an application by TCTP is currently a manual process, but within CYCLONE this process will be facilitated by extending the current TCTP implementation so that its addition to existing cloud solutions can be performed automatically by the deployment management.

III. CYCLONE USE-CASES

CYCLONE has identified two flagship applications: an academic cloud platform and associated services for bioinformatics research and a commercial deployment for smart grids in the energy sector [9]. These applications will guide the initial development of the tools.

IV. BIOINFORMATICS USE-CASES

The Bioinformatics use-cases deal with the collection and efficient analysis of biological data, in particular genomic information from DNA sequencers. Bioinformatics software is characterized by a high degree of fragmentation: literally hundreds of different software packages are regularly used for scientific analyses with an incompatible variety of dependencies and a broad range of resource requirements. For this reason, the bioinformatics community has strongly embraced cloud computing with its ability to provide customized execution environments and dynamic resource allocation.

A. UC1 - Securing human biomedical data

An increasing number of clinicians are including biological results obtained with the genome sequencing technologies in their day-to-day diagnosis practice. In the near future, some of the genomic data processed on the IFB's (Institut Français de Bioinformatique) cloud platform will concern human biomedical data related to patients and thus, will be subject to strict privacy constraints. To guarantee the data security while carrying out the analysis in a federated cloud environment, it is necessary to ensure the security in all involved sites belonging to the federation as well as their integration (especially if the cloud federation involves both public and private cloud infrastructures). This use case is deployed to evaluate the capability of CYCLONE infrastructure to provide users with the deployment of their own cloud infrastructure for the biomedical analyses consisting of a single virtual machine with a web interface in a secured environment. The access to this environment will be based on the identity and authorizations in the federation.

B. UC2 - Cloud virtual pipeline for microbial genomes analysis

This use-case deals with the procedure used by biologist to sequence the bacterial genomes. Currently this procedure involves the use of several bioinformatics software tools with different dependencies, a relational database management system, a web server and servlet container and above all it requires solid skills in system administration. CYCLONE Cloud federation will provide biologists with their own comprehensive platform that can be deployed in one click over one or more cloud infrastructures. To achieve this, new features to automate deployment of complex application will be added to the IFB's cloud portal. Such deployments can be done over several cloud infrastructures with the dynamic allocation of network resources for the isolation of the VMs inside a dedicated network and with the replication of the user data. The Cloud service provider will guarantee access to the federation based on the identity and authorizations, an isolated network for security and confidentiality reasons, the required features in terms of resources availability and elasticity.

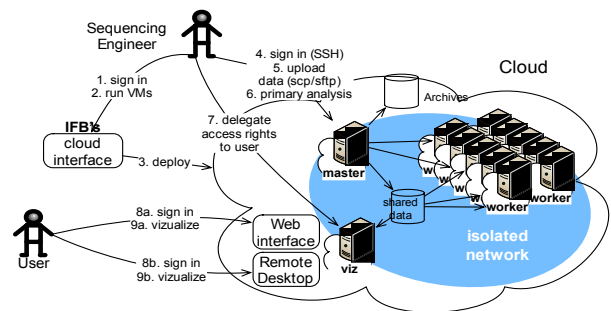


Fig. 2: Functional schema of the Bioinformatics Use-cases

C. UC3 - Live remote cloud processing of sequencing data

Bioinformatics deals with the collection and efficient analysis of biological data, particularly genomic information from DNA sequencers. The terabytes of raw data, produced by the sequencers for each run, require significant computing resources for analysis that may not be available locally. These sequencers are located at a dozen places in France, while the users are distributed throughout the whole country, so most of time the raw data are physically transferred using hard disks sent by express shipping. In this use-case the utilization of federated cloud computing resources will help online data analysis: reducing the data transfers, reducing the need for long-term storage of the raw data in the sequencing center, reducing the time to analyse the raw data, and obviating the need for managing an extensive, local IT infrastructure. Users will need to have access to the secured and isolated workspace, where the complete set of required appliances is made available together with a distributed cluster of VMs in order to run the raw data analysis and the connection towards the shared storage.

V. ENERGY USE CASE

In order to comply with the “202020” [10] climate change mitigation goals of the European Union, the energy generation

has to be moved from fossil, conventional power plants to a sustainable, renewable energy generation, thus the energy production will go from a small number of centralized, big power plants to a huge number of Distributed Energy Resources (DER) located in a wide geographical area. The Virtual Power Plant (VPP) connects several small and medium sized DERs, such that they are manageable as a single larger one. The VPP combines the advantages of various renewable energy sources: wind turbines and solar modules generate electrical energy in accordance to how much wind and sun is available and when it is available. Bio-mass is used to make up the difference: it is converted into electricity as needed in order to balance out short-term fluctuations. To build the Virtual Power Plant based on the ICT platform, the service for distributed data collection gathers the energy generation data from the different DERs and stores them as raw data on the platform. The data authenticity must be preserved and only authorized access to the data will be allowed. The CYCLONE platform will have to ensure trustable cloud computing with an end-to-end secure data management approach, moreover high availability, scalability, and elasticity are needed in order to meet the dynamicity of the computing and storage requirements in a transparent way for the user. This use case will evaluate the capabilities of the CYCLONE features to deploy a complex application on distributed cloud environment. The deployment of the application may require different clouds platforms depending on the availability of cloud resources and requirements of application provider.

VI. THE CYCLONE TESTBED

The CYCLONE testbed [11] is composed by network and IT resources provided by project partners and installed on their premises. This common and shared platform plays a key role both for enabling the full and continuous validation of the software and for demonstrating the concrete realization of the use-cases described above. The testbed is composed by three Infrastructure-as-a-Service (IaaS) platforms spread across different countries around Europe and provided by LAL partner located in France, QSC partner located in Germany and Interoute located in Italy. All three testbeds are deployed with OpenStack using the same configuration in terms of component installed and virtual instance's flavours available. The testbed's user does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications, and possibly control of selected networking components (e.g., firewalls and load-balancers). All the resources made available are managed through a SlipStream service called Nuvla [12] which abstracts the specificities of the different cloud resource providers so that end users can easily choose for each application the best option among the providers. On the networking side, the OpenNaaS-CNSMO component has been integrated with SlipStream so that networking services are deployed together with the application.

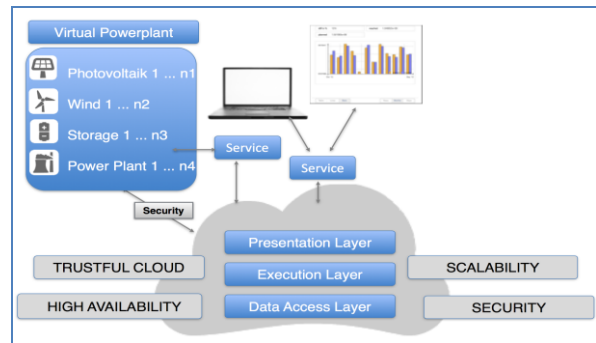


Fig. 3: Functional schema of the Energy use case

VII. CONCLUSION AND FUTURE DEVELOPMENT

This paper has presented the cloud federation approach developed in the CYCLONE project. The platform consists of different mature open-source software and components some of them produced and maintained by consortium's partners other publicly available. Through CYCLONE, the project consortium has created a multi-cloud and multi-provider infrastructure easy to use and compliant with the advanced standards of quality of a product-ready solution. Results from validation are expected in Q4-2016. Once consolidated with its initial set of use cases, CYCLONE will evolve by extending the use case portfolio with new application and cloud scenarios that will further improve the overall functionality of the platform and, above all, increase external users' involvement.

ACKNOWLEDGMENT

The CYCLONE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644925.

The CYCLONE project thanks members of the the bioinformatics platforms Centre Léon Bérard (Lyon, France) and IFB-MIGALE (Jouy-en-Josas, France) to provide their bioinformatics applications as respectively use cases UC1 and UC2.

REFERENCES

- [1] Cisco Global Cloud Index: Forecast and Methodology, 2014–2019
- [2] CYCLONE – Complete Dynamic Multi-cloud Application Management, H2020-ICT-2014-1 Innovation Action, Grant Agreement n. 644925, <http://www.cyclone-project.eu/>
- [3] <https://www.openstack.org/>
- [4] <http://opennaas.org/>
- [5] <http://sixsq.com/products/slipstream/>
- [6] Mathias Slawik, "The Trusted Cloud Transfer Protocol" 2013 IEEE 5th International Conference on Cloud Computing Technology and Science.
- [7] J. Aznar, E. Escalona, I. Canyameres, O. Moya, A. Viñes, "CNSMO: A Network Services Manager/Orchestrator Tool for Cloud Federated Environments", Accepted for publication in proceedings of the 15th IFIP Annual Mediterranean As Hoc Networking Workshop (MedHocNet), June, 2016.
- [8] CYCLONE, Deliverable D6.1
- [9] CYCLONE, Deliverable D3.1
- [10] http://ec.europa.eu/clima/policies/strategies/2020/index_en.htm
- [11] CYCLONE, Deliverable D7.1
- [12] <http://sixsq.com/services/nuvla/>