Cyclone

Complete Dynamic Multi-cloud Application Management

Project no. 644925

Innovation Action

Co-funded by the Horizon 2020 Framework Programme of the European Union

Call identifier: H2020-ICT-2014-1

Topic: ICT-07-2014 – Advanced Cloud Infrastructures and Services

Start date of project: January 1st, 2015 (36 months duration)

Deliverable D3.3

Final Evaluation of Use Cases

Due date:	31/10/2017
Submission date:	15/12/2017
Deliverable leader:	CNRS
Editors list:	C. Blanchet (CNRS)

Dissemination Level

☒	PU:	Public
☐	PP:	Restricted to other programme participants (including the Commission Services)
☐	RE:	Restricted to a group specified by the consortium (including the Commission
☐	CO:	Confidential, only for members of the consortium (including the Commission Services)

List of Contributors

Participant	Short Name	Contributor
Interoute S.P.A.	IRT	Domenico Gallico, Matteo Biancani
SixSq Sàrl	SIXSQ	Charles Loomis
QSC AG	QSC	Ralf Fischer, Adelheid Weinert
Technische Universitaet Berlin	TUB	Dirk Thatmann
Fundacio Privada I2CAT, Internet I Innovacio Digital A Catalunya	I2CAT	Eduard Escalona
Centre National De La Recherche Scientifique	CNRS	Christophe Blanchet
Universiteit van Amsterdam	UvA	Fatih Turkmen, Yuri Demchenko

Change history

Version	Date	Partners	Description/Comments
0.1	02/11/2017	CNRS	Initial version.
0.2	20/11/2017	CNRS, QSC, TUB	Revised version with contributions from partners leading use cases
0.3	30/11/2017	CNRS, i2CAT, QSC, SIXSQ, TUB, UvA	Revised version with contributions from partners.
0.4	08/12/2017	CNRS, i2CAT, QSC, SIXSQ	Revised version, ready for internal review
0.5	12/12/2017	CNRS, QSC	Revised version after first internal review
0.6	13/14/2017	IRT	Final Review

Table of Contents

List of Figures

List of Tables

Executive Summary

Today scientific and industrial applications are increasingly complex and require advanced computing infrastructures. With the development of cloud computing, cloud providers are obvious partners for deploying these applications that requiring large and adaptable computing resources, as well as adequate security services associated with network features. The main goals of CYCLONE consist of integrating and extending open source software to create such a unified cloud application management solution for application service providers, DevOps and researchers.

One of the main added values of the CYCLONE developments is the ability to deploy legacy and new applications on multi-cloud infrastructures and to face needs like deployment of complex applications, security and access management.

The partners have evaluated the proposed solutions based on initial use cases related to life science, medicine, health, and energy developments. During the time of the project, the consortium has also identified several new use cases related to different areas: teaching, data management, energy, bioinformatics, etc.

Based on all use cases' requirements, CYCLONE partners have defined the software implications related to the CYCLONE key technical areas and developed new software components to fulfil the needs. The overall analysis of all CYCLONE use cases provided common requirements, and their level of achievements at the end of the project, demonstrates that these goals are reached. Among the 48 identified common requirements, 38 were satisfied, with almost all relevant requirements.

Most of the use cases identified during the project were successfully integrated with CYCLONE software to achieve multi-cloud deployments. Those interrupted or closed were due to minor re-scoping of some activities of the project related to the components developments and the target audience. They are available to the deployment as Nuvla recipes, and for the bioinformatics use cases, also as an entry in the IFB's Biosphere catalogue.

Finally, an example of the reuse of CYCLONE developments is the IFB cloud federation and the Biosphere portal that relies on the CYCLONE component to federate several IFB's cloud to provide life science scientists and engineers tools to deploy multi-cloud bioinformatics virtual environments.

1. Introduction

CYCLONE integrates and extends open source software to create a unified cloud application management solution for application service providers, DevOps, and researchers. One of the main added values of the CYCLONE developments is the ability to deploy legacy and new applications on multi-cloud infrastructures and to meet related needs such as deployments of several services of a complex application in many virtual machines for scalability or availability, identity and authorizations management, or network filtering and isolation.**Error! Not a valid filename.**

At the beginning of the project, the partners have elaborated the use cases that were initially presented in the DOA: bioinformatics and energy use cases. New use cases were also selected to complete this portfolio and to help to identify new requirements. This work was described in deliverable D3.1 [1]. During year 2, the consortium deployed the new uses cases and provided a consolidated evaluation of all CYCLONE use cases, focusing on the multi-cloud perspective. Based on these requirements, CYCLONE partners have defined the software implications related to the CYCLONE key technical areas and developed new software components to meet the needs (see deliverable D3.2 [2]).

This document describes the final evaluation of all CYCLONE use cases.

In chapter 2, we provide an overall analysis of all use cases including the identified common requirements of all use cases, their links to the CYCLONE key technical areas and related software developments, and their level of achievements at the end of the project.

In chapter 3, we detail the achieved integration of CYCLONE use cases. We provide below both details for the recent achievements and a summary for the previously achieved integrations.

And finally, in chapter 4, we describe the motivation and developments done in the context of the CYCLONE project for the Biosphere portal to hide to life scientists and engineers some complexities, but without losing the power, of the cloud, applying the developments and technology of CYCLONE to setup the IFB's cloud federation.

2. Evaluation and analysis of all use cases

CYCLONE use cases come from several domains, *e.g.* bioinformatics, energy, teaching, security. Table 1 gives an overview of the CYCLONE use cases, their domains and the partner who is the principal investigator. We also assigned previously a priority of treatment. For detailed descriptions, the reader should refer to the CYCLONE use case portal [3]. It is important to notice that some of the use cases rely on legacy applications, especially in the bioinformatics fields, where the proposed applications where previously deployed on local computing clusters or web portal.

Table 1: List of CYCLONE use cases

ID	Title	Domain	Resp.	Priority
UC1	Securing human biomedical data	Bioinformatics	CNRS	high
UC2	Cloud virtual pipeline for microbial genomes analysis	Bioinformatics	CNRS	high
UC3	Live remote cloud processing of sequencing data	Bioinformatics	CNRS	high
UC4	Virtual Power Plant	Energy	QSC	high
UC5	Internet of Services Lab (IoSL)	Teaching	TUB	high
UC6	ENTRANCE	App mgmt.	TUB	medium
UC7	Open Scientific Data	Data management (Earth Observation)	SixSq	medium
UC8	Benchmark Driven Placement	Bioinformatics	SixSq	high
~~***UC9***~~	~~*On-Demand Bandwidth*~~	~~*Network Provisioning*~~	~~*SixSq*~~	~~*low*~~ *
UC10	Smart Utility 4.0	Energy	QSC	high
UC11	Assembling genomes from sequencing reads	Bioinformatics	CNRS	high
UC12	Metagenomics	Bioinformatics	CNRS	medium
UC13	Shared environment between cloud Galaxy portals	Bioinformatics	CNRS	high
~~***UC14***~~	~~*WebRTC video conference solution*~~	~~*Real Time Communications*~~	~~*I2CAT*~~	~~*low-medium*~~ **
~~***UC15***~~	~~*Interactive Authorization Policy Management for Multi-cloud Applications*~~	~~*Security*~~	~~*UvA*~~	~~*medium*~~ ***
UC16	Attribute-based Authorizations with XACML	Security	UvA	high
UC17	Genomic Variant Analysis for Cancer and Rare Disease Diagnosis	Bioinformatics	CNRS	high

* Due to the rescoping of the networking activities this cannot be achieved.

** External resources to implement this use case were eventually not available and it was kept as a study only use case.
*** The use case has been found out of scope of CYCLONE's target audience

2.1. CYCLONE Key technical areas and related software developments

At the initial steps of the project, the CYCLONE consortium has defined four key technical areas:

1) Cloud Access Management through cloud proxies,
2) Matchmaking, Brokering, and Mediation of Cloud Resources,
3) End-to-end Security for HTTP-based Applications,
4) Dynamic network configuration and management.

In relation with these key technical areas, CYCLONE partners have identified and realized several software developments to enhance existing cloud tools or to create new specific components (See Table 2).

During the last period of the project, new requirements were raised by last identified use cases (see Table 3), and the consortium developed three new components to satisfy them: CNSMO-LB, CNSMO-DNS and DACI. These developments are related to the network and security domains, and detailed below.

CNSMO-LB and –DNS

CNSMO-LB is the CNSMO component that deploys and enables a network Load Balancer as part of the application deployment process. It allows distributing traffic across multiple servers within a cluster increasing the fault tolerance of application components. CNSMO-LB is also designed to elastically scale by adding or removing target servers without disrupting the flow of requests to the application.

CNSMO-DNS is the CNSMO component that deploys and enables a Domain Name System server as part of the application deployment process. It acts as a local DNS in the application's network to resolve host names into IP addresses. New entries in the DNS server can be added dynamically and is automatically integrated with the CNSMO-VPN component in order to configure all VPN clients with the DNS server address.

DACI

Dynamic Access Control Infrastructure (DACI) is a flexible access control service that allows flexible authorizations by employing eXtensible Access Control Markup Language (XACML) as the underlying language. DACI can be dynamically provisioned as a component of a cloud application to regulate access to sensitive resources. DACI includes four core services (see deliverables D4.2, D4.4 and D4.5 for more details): Authorization Service, Context Management Service Token Service and Tenant Management service. Figure 1 presents the dependency relations between these subservices.

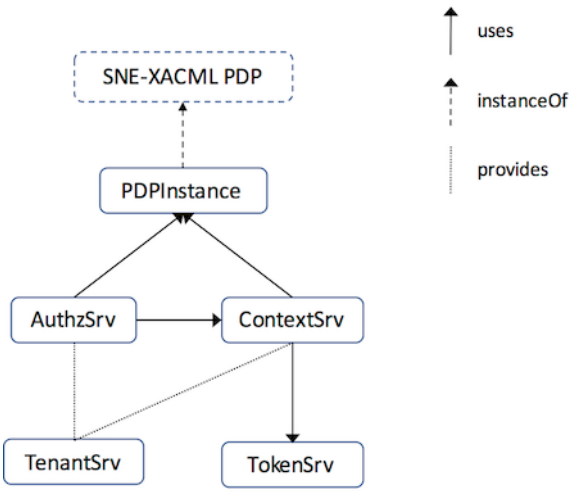

Figure 1: DACI Dependency Relations

The DACI sub-services can be hosted on the same node or deployed over distributed nodes in a cloud infrastructure. They communicate with each other over mostly REST interfaces. The tenant information and their access control policies are stored in in-memory database (i.e. Redis).

Table 2: CYCLONE key technical areas and associated software developments

Key technical areas	Software developments
Deployment: **1) Cloud Access Management** **2) Matchmaking, Brokering, and Mediation of Cloud Resources**	Cloud deployment
	Complex App deployment
	Multi-cloud deployment
Security: **3) End-to-end Security for HTTP-based Applications**	Federation Proxy
	Web authentication
	SSH authentication
	DACI
Network: **4) Dynamic network configuration and management**	CNSMO-VPN
	CNSMO-FW
	CNSMO-LB
	CNSMO-DNS

2.2. Software implications according to the use cases requirements

For all use cases, the CYCLONE consortium identified common requirements (see the Annex of this document for a detailed a description of them). We defined precisely the links between the requirements, the key technical areas, and the implications related to the software components developed by CYCLONE.

Table 3 shows a detailed view of these links between the requirements of CYCLONE use cases and software tools developed in the context of the CYCLONE project.

Table 3: Details of use cases requirements and related CYCLONE software implications

	Key Tech. Areas	Cloud deployment	Complex App	Multi-cloud	Federation Proxy	Web authn	SSH authn	DACI	CNSMO-VPN/SDN	CNSMO-FW	CNSMO-LB	CNSMO-DNS
UC1	1, 2, 3	✔	-	✔	✔	✔	-	-	-	-	-	-
UC2	1, 2, 4	✔	✔	✔	✔	✔	✔	–	✔	-	-	✔
UC3	2, 4	✔	✔	✔	✔	-	✔	–	✔	-	-	-
UC4	2, 4	✔	✔	✔	-	-	-	-	✔	✔	-	-
UC5	1, 3	✔	✔	-	✔	✔	✔	-	-	-	-	-
UC6	1,3	✔	-	-	✔	✔	-	-	-	-	-	-
UC7	2, 1	✔	✔	✔	✔	✔	-	-	-	-	-	-
UC8	2	✔	✔	✔	-	-	-	-	-	-	-	-
UC10	1,3,4	✔	✔	✔	-	-	-	✔	✔	✔	-	-
UC11	2	✔	✔	✔	✔	✔	✔	–	✔	✔	-	-
UC12	2, 3, 4	✔	✔	✔	✔	✔	✔	–	✔	✔	-	✔
UC13	1, 2	✔	-	✔	✔	✔	-	-	-	-	✔	-
UC16	**1,3**	✔	✔	✔	✔	✔	✔	✔	-	-	-	-
UC17	**1,2,3 ,4**	✔	✔	✔	✔	✔	-	–	✔	✔	-	✔

2.3. Achievements of use case common requirements

During the development of the project, the use cases helped to defined **48 common requirements**. They were attributed a level of relevance according to the following scale: **MAY be satisfied (7 requirements), SHOULD (15) and MUST be (26)**. At the end of the project, these requirements are mostly fulfilled with the new multi-cloud components developed by CYCLONE partners: **38 are satisfied, 4 partially and 6 were finally not implemented**. It is important to notice that almost all requirements qualified as “MUST be

satisfied" have been fulfilled. The Table 4 summarizes the achievements of these common requirements. And the reasons why some of them were not, or partially, satisfied are explained below the Table 4.

Table 4: Level of satisfaction of the use cases requirements

ID	Title	Relevance	Achievement
1	Cloud User Interface for service access based on web technology	MUST	✔
2	Cloud User Interface adapted to community usage	SHOULD	✔
3	Cloud User Interface adapted to mobile devices	SHOULD	✔
4	Federated identity management	MUST	✔
5	Federated authorization management	MUST	✔
6	One-click deployment of simple application	MUST	✔
7	VM web interface secured by the identity federation	MUST	✔
8	End-to-end secure data management	MUST	✔
9	VM isolated network	MUST	✔
10	VPN connectivity services	SHOULD	✔
11	FedId Access to storage	MUST	✔
12	Community reference datasets	MUST	✔
13	Access to public Cloud Services	MUST	✔
14	Cloud deployment according to medical data treatment certification	MUST	✔
15	Data erasure	MUST	-
16	Deployment of complex application	MUST	✔
17	One-click deployment of Complex application	SHOULD	✔
18	High-throughput storage	MUST	✔
19	Multi-VMs shared volume	MUST	✔
20	Multi-clouds deployment of complex application	SHOULD	✔
21	Multi-clouds distribution of community reference datasets	SHOULD	✔
22	FedId SSH connection to the VM	MUST	✔
23	Elastic management of complex applications	MAY	✔
24	Dynamic Network resource allocation	MUST	✔
25	Multi-clouds distribution of user data	SHOULD	✔
26	Ensure WAN High bandwidth links	MUST	~
27	Archiving of raw data	MAY	-
28	Configuration of volatile disks	MUST	~

29	X11 remote display	SHOULD	✔
30	FedId access for X11 service	MAY	✔
31	Group authorizations	MUST	✔
32	User-defined authorizations	MUST	✔
33	Deployment of applications to non-dedicated cloud infrastructure	MUST	✔
34	Guaranteed network performance (QoS)	SHOULD	~
35	VM with large memory	MUST	✔
36	Deployment of complex workflows using Docker containers.	MUST	✔
37	Software deployment based on a Docker container	MUST	✔
38	Public storage backend	SHOULD	✔
39	Define authorizations to access user data on a public storage	SHOULD	✔
40	Map a public storage as a filesystem in a VM	SHOULD	✔
41	Horizontal elasticity	MUST	✔
42	Vertical elasticity	MAY	-
43	Manage academic license for the tools deployed in the images/VMs	SHOULD	✔
44	Load balancing for Scalability, reliability	SHOULD	~
45	Ensure WebRTC ports in the firewall	MAY	-
46	WebRTC signalling server securized	MAY	-
47	Redundant Storage	MAY	-
48	Access Documentation	SHOULD	✔

The reasons that some requirements were partially or not achieved are the following:

- **Req. 3**: The SlipStream user interface is responsive and does adapt to smaller devices (mobile and tablet). It works fine for monitoring applications, but any significant modifications to applications or deploying them is rather awkward (although possible).
- **Req. 15**: The data erasure cannot be guaranteed on all clouds, but it can be alternatively satisfied as in Req. 14 by allowing the deployment only on compliant clouds.
- **Req. 26**: Guarantying high bandwidth links requires access to the physical infrastructure. The solution implemented can monitor the used bandwidth to check against the agreed SLAs with the service provider.
- **Req. 27**: Not all clouds are providing archive-enabled storage
- **Req. 28**: The availability of volatile disks is related to cloud configuration
- **Req. 34**: Guarantying QoS requires access to the physical infrastructure. The solution implemented can use QoS guarantees over virtual resources only and check them against the agreed SLAs with the service provider.

- **Req. 42**: The vertical elasticity requires a restart of the instance which is not possible for the related use cases.
- **Req. 45**: This requirement was not implemented due to the lack of external resources to develop UC14.
- **Req. 46**: This requirement was not implemented due to the lack of external resources to develop UC14.
- **Req. 47**: In the project it was decided not to include developments regarding storage.
- **Req. 48**: The logging mechanism was developed for the Federated Identity Provider.

3. Achieved integration of CYCLONE use cases

Most of the use cases identified during the project were integrated with CYCLONE software to achieve multi-cloud deployments. Table 5 shows an overview of the achieved integration of all CYCLONE use cases. All the use case deployment recipes are available in the CYCLONE restricted area of the Nuv.la portal (https://nuv.la/module/cyclone), and for the bioinformatics ones also in the IFB's Bioinformatics portal (https://biosphere.france-bioinformatique.fr/catalogue). When useful, the related links are provided below in the use case related paragraphs. Most of the use cases were integrated previously in the CYCLONE testbed, but we provide below both details for the recent achievements and a summary for the previously achieved integrations.

Table 5: Summary of use cases integration with the different CYCLONE components

ID	Title	Deployment	Security	Network
UC1	Securing human biomedical data	✔	✔	–
UC2	Cloud virtual pipeline for microbial genomes analysis	✔	✔	✔
UC3	Live remote cloud processing of sequencing data	✔	✔	✔
UC4	Virtual Power Plant	✔	-	✔
UC5	Internet of Services Lab (IoSL)	✔	✔	-
UC6	ENTRANCE	✔	✔	-
UC7	Open Scientific Data	✔	-	-
UC8	Benchmark Driven Placement	✔	-	-
UC10	Smart Utility 4.0	✔	✔	✔
UC11	Assembling genomes from sequencing reads	✔	✔	✔
UC12	Metagenomics	✔	✔	✔
UC13	Shared environment between cloud Galaxy portals	✔	✔	-
UC16	Attribute-based Authorizations with XACML	✔	✔	✔
UC17	Genomic Variant Analysis for Cancer and Rare Disease Diagnosis	✔	✔	✔

3.1. Recent achievement

UC6 ENTRANCE

The Entrance System has been extended to support **Cyclone's Federation Provider** for user authentication. This addresses the fundamental question of how to provision and authenticate users in a secure way. Entrance took advantage of the new German ID card (nPA) but this approach limits the number of possible data receivers to owners of the German ID-card. With the CYCLONE Federation Provider integration new groups of data receivers can be easily provisioned by an ABE system owner.

In case a company wants to deploy their own Entrance system, e.g. to provide a data-centric and generic encryption for file exchanges for its employees, the use of **Cyclone's deployment** simplifies the rollout.

Workflow

Figure 2 depicts the provision process of a new user using Cyclone's Federation Provider. Beside an automatic rule-based user provision (steps: 5 -> 6 -> 9) the system owner can approach pending registrations manually (steps: 5 -> 6 -> 7 -> 8).

Users' stories

Actors

- Bob is the System owner and want to setup an Entrance System to encrypt and share files with recipients. The list of recipients and the attributes are maintained, and the private key generation is performed by Bob acting as system owner.
- Enterprises Companies want to setup an Entrance instance to enable their employees Alice and Eve to encrypt and share files with a set of recipients. Recipients are maintained by the Enterprise Administrator Bob who is in charge of maintaining the identities.
- A cloud provider supplies the computing and storage resources, with the required features permitting the automatic deployment of an application
- Rita the data receiver registers to the system. After being granted by the system she receives her secret key. She is now able to decrypt certain ciphertexts, depending on whether the attributes attached to her key can solve the access structure embedded to the ciphertext or not.

Stories

- As a private individual, I can setup an Entrance system for a self-sovereign file exchange for ABE encrypted data.
- As a private individual, I have access to the Entrance Dashboard to manage attributes, access structures, files and data recipients, such as friends, family member or colleagues.
- As an enterprise, I can setup an Entrance system to offer each of my employees a specific private key. Hereby attributes can reflect a group or department affiliation, individual security levels or other differences.
- As an enterprise, I have access to the Entrance Dashboard to manage attributes, access structures, files and data recipients, such as friends, family member or colleagues.
- As an enterprise, I would like to have an easy and straight forward system deployment.

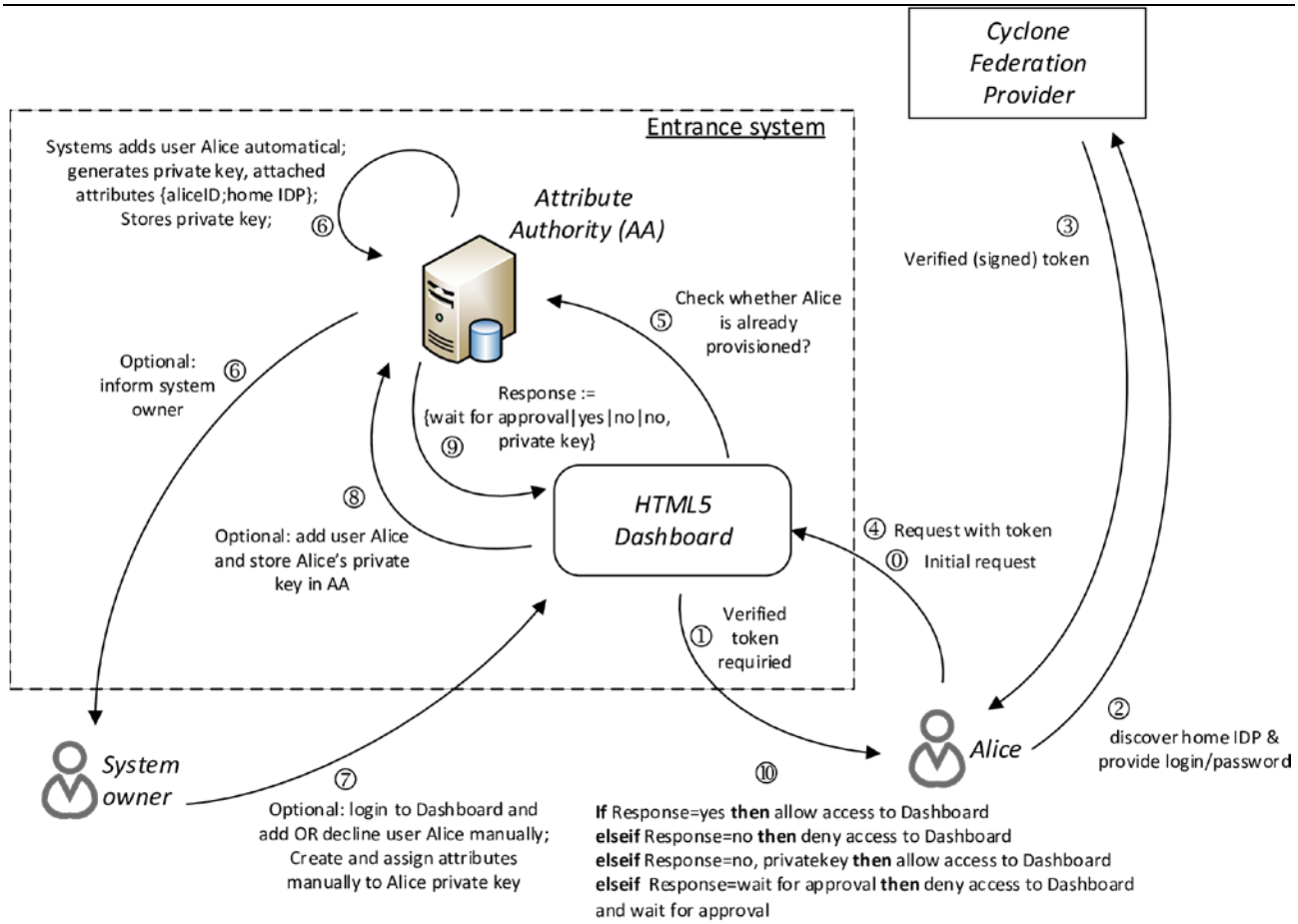

Figure 2: New user provisioning using Cyclone's Federation Provider.

UC7 Open Scientific Data

A SixSq intern created proof of concept service for generating and updating an Earth Observation data product. It performs satellite image processing according to the client's requests, while meeting execution time and cost constraints. The architecture of the developed service is comprised of multiple entities including a benchmark system to enable cost predictions of the image processing unit. The benchmarks, cloud service offers, and resulting data products were all stored within the SlipStream Service Catalog. This proof of concept validated the general architecture for generating and maintaining an open scientific data set on the cloud using the SlipStream Service Catalog and cloud application automation.

UC8 Benchmark Driven Placement

SixSq created a specific CIMI resource within SlipStream to allow users to register benchmarks associated with specific cloud services and cloud resources. This resource integrates well with the overall SlipStream Service Catalog, allowing users to take into account their benchmarks (or benchmarks published by others) when selecting cloud resources. Currently, the selection must be managed directly by the application, using the information in the Service Catalog. Eventually, the SlipStream deployment engine will directly support benchmark-driven placement.

UC10 Smart Utility 4.0

The energy use case Smart Utility 4.0 comprises the first use case UC4, the Virtual Power Plant (VPP). In UC4 distributed renewable energy sources were aggregated into one virtual power plant, the VPP. To consider the requirements of future energy management which include also the areas of energy

consumption and transport, that use case is extended by including energy consumers and the ICT platform services. The inclusion provides an energy management that enables the partners based on the aggregated energy data collected within the platform to balance the generation and consumption of energy - thus forming the Smart Utility.

The use case combining renewable energy generation with industrial consumers to achieve the optimal usage of the generated energy, has been defined on Nuvla integrating the DACI and CNSMO tools developed by CYCLONE. The DACI authorization is used for the flexible specification of policies for provider and tenants for data access. The tenants are the operators of the energy sources, the VPP and the consumers. The advantage provided by the package is that it allows separating the policy specification from the application code. The aspect oriented method of integration was used and consequently no modification of the application code was necessary. Integrating the DACI tool into the application provides a flexible method to specify access policies and quickly react to changes like e.g. adding or blocking a tenant, thus adding a layer of data access security.

Furthermore, the CNSMO network tools were included in the application to provide dynamic VPN and FW services for the data transfer between the VPP partners and clouds, i.e. within the Smart Utility. It provides a centralized network provisioning, presenting an abstracted view of the cloud for the users of the platform and services. The VPN service allows the user to dynamically add and remove partners from the private network.

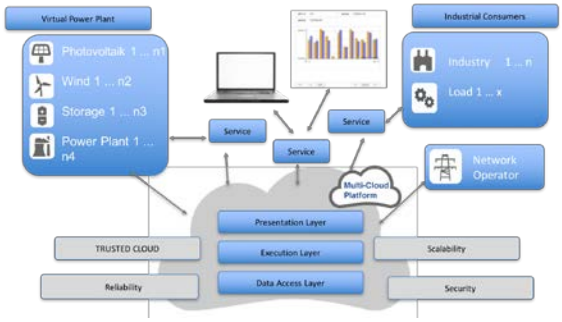

Figure 3: Smart Utility Roles and Datatransfer

The use case includes renewable energy resources and industrial consumers in the region near Cologne (3). The Smart Utility application enables the VPP operators and the consumers to make informed decisions, this way contributing to the stability of the grid and profiting on the energy market.

Figure 4: Energy Generation

Figure 5: Energy Load

The deployment via the CYCLONE platform provides a reliable and time effective method to set up and manage the Smart Utility application. The added value provided by DACI and CNSMO for security features and network management within the application enables the application provider to save the effort for development and maintenance of these services. The provider of the Smart Utility application can focus on the core problem having not to provide the know-how for the added services.

UC11 Assembling genomes from sequencing reads

https://nuv.la/module/cyclone/UC11_Assembling_genomes

https://biosphere.france-bioinformatique.fr/catalogue/appliance/111

The goal in use case 11 was to provide a pipeline for bacterial genome assembly that could be deployed and used easily from a simple web interface. Our implementation is based on *bistro*[1], an OCaml library for specifying and running scientific workflows. *bistro* features cached, parallel and resumable (on failure) execution, and is able to use Docker containers to run external programs. In addition, it provides an extension implementing a lightweight web server to run a pipeline, follow its execution and browse through its results. Notably, this web server is able to automatically derive an HTML form to input the data and parameter values.

For deployment, we wrote a Slipstream recipe that specifies the configuration of a virtual infrastructure running the pipeline. This infrastructure is composed of a light front-end VM that runs the web server and a virtual cluster operated via Docker Swarm. The workload required by the bioinformatics computing analysis can be then easily distributed in case of heavy requirement on a scalable bunch of VM thank to the CYCLONE developments.

UC12 Metagenomics

https://nuv.la/module/cyclone/UC12_metagenomics_pathotrack

https://biosphere.france-bioinformatique.fr/catalogue/appliance/100

The PathoTRACK application targets the design of software that will screen reads (*i.e.* DNA fragments) obtained by massive sequencing against specific or generalist databases for the identification of threat agents or pathogens produced by synthetic biology. The expected specifications must provide:

- The detection and identification of pathogenic organisms in complex samples using high throughput screening data
- An automated analysis (i.e. computations) management
- The analyses history
- An easy installation/deployment
- A friendly user interface

The application is now available as a complex application recipe (with several components) in NuvLa and registered in the bioinformatics Biosphere portal. The user data are uploaded through the web interface, and compared to several reference databases shared among the components. The current stable version consists of different components that are deployed in one-click as separate VM:

- A web user interface,
- A SQL service relying on the mariadb database,
- A job manager that uses the Snakemake workflow system to distribute the computing workload over a devoted cluster,
- Several computing nodes
- A REST API linking the different components.

The PathoTrack application has multi-cloud requirements to permit:

- To keep the private reference datasets on premises

[1] https://github.com/pveber/bistro

- To distribute the required computing workload on available large-scale cloud resources

The application was demonstrated at a life science summer school about metagenomics (June 2017) and validated by attending bioinformaticians.

UC13 Shared environment between cloud Galaxy portals

https://nuv.la/module/cyclone/UC13-Galaxy

https://biosphere.france-bioinformatique.fr/catalogue/appliance/68

https://biosphere.france-bioinformatique.fr/catalogue/appliance/79

Galaxy portal aims to provide biologists with an open and web-based platform for bioinformatics research that gathers multiple bioinformatics tools and allows chaining their execution in workflows. In the UC13 recipe, the galaxy portal and the bioinformatics tools are deployed automatically with Docker containers. The application now uses the CYCLONE Federation Proxy to authenticate users in Galaxy portal. Biologists can now access to their own portal according to their eduGAIN identity. The Federation proxy also provides the owner of the cloud instance a simple way to allow shared access with other biologists. It was presented (as a poster) at the international Galaxy 2017 conference.

UC15 Interactive Authorization Policy Management for Multi-cloud Applications

This use case has been considered relevant at the early stages of CYCLONE and inspired from [4].However, the evolution of the project rendered its realization moot since the authorization functionality provided by DACI (see Section 2.1 for more details) has been found suitable for the energy related use cases (Table 3) where the access control requirements are different. More specifically, no conflicts arise in the policies of tenants as the rules of access are established during the collaboration setup and enforced at the level of interfaces (i.e. endpoints). To this end the use case has been discussed within the consortium meetings and closed.

UC16 Attribute-based Authorizations with XACML

The use case UC16 has been implemented (as part of the component DACI) and integrated with some of the energy related use cases (e.g. UC4. The architectural details of the components employed in the use case have been extensively discussed in deliverables D4.2, D4.4 and D4.5. Example access control policies for the use case have been specified according to requirements of various use cases including UC10.

All the code related to the use case is stored on the project repository [6] and the relevant cloud deployments are available at [7]. Figure 6 shows the deployment parameters of an application that employs the components developed in the use case. Since all subservices of DACI work on the same VM node in this application, the services are assigned different port numbers.

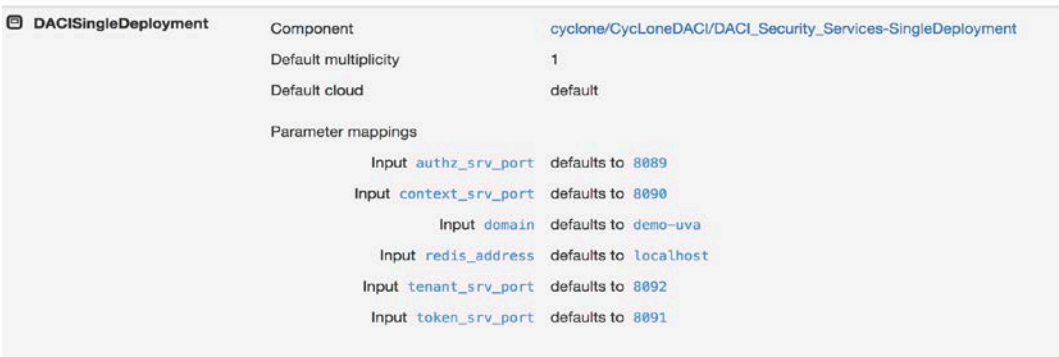

Figure 6: DACI Single Deployment Parameters

One interesting requirement for the integration of the use case with UC10 was related to the enforcement of policies. The enforcement has been implemented in such a way that the UC10 code base has been left

intact thanks to the use of the aspect oriented programming techniques. The cloud application that integrates this use case with UC10 is available at [5].

UC17 Genomic Variant Analysis for Cancer and Rare Disease Diagnosis

https://nuv.la/module/cyclone/UC17_VCF_for_Diagnosis/uc17-wf-nte
https://biosphere.france-bioinformatique.fr/catalogue/appliance/109

The application relies on the Snakemake workflow engine and several bioinformatics tools (bwa, samtools, picard, GATK, VEP). The bioinformatics components are deployed with the Conda framework, on a base cloud image of CentOS 7. At the first step, the workflow (See Figure 7) ensures that the reference data set is available (VEP/homo_sapiens_merged_vep_86_GRCh37, 8+ GB) or downloads it. The workflow can be coupled to a Grid Engine cluster distributed over an isolated network (CNSMO-VPN, coupled with CNSMO-DNS) for large workloads.

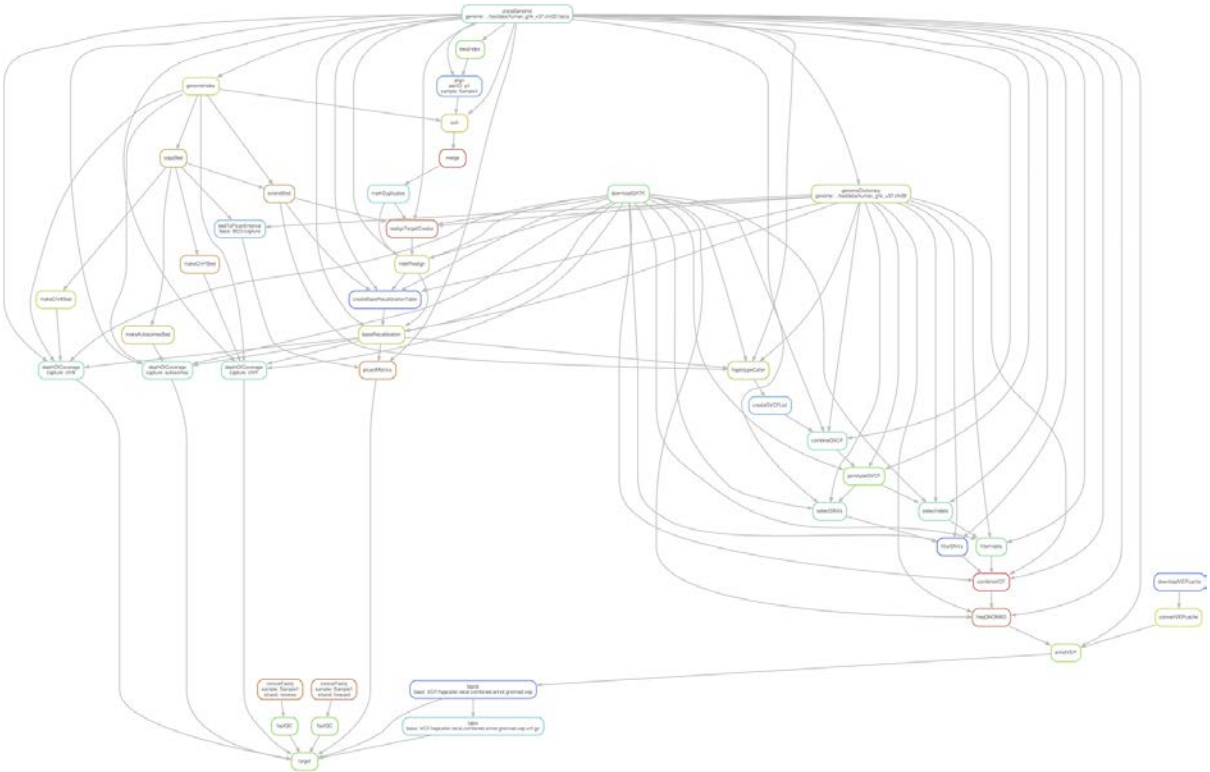

Figure 7: Workflow of the UC 17 Genomic Variant Analysis for Cancer and Rare Disease Diagnosis

3.2. Summary of previous integrations since the beginning of the project

The following use cases were already deployed with the CYCLONE components.

UC1 Securing human biomedical data

https://nuv.la/module/cyclone/UC1_Securing_human_biomedical_data

The VM was instrumented with the Federation Provider and deployed in the CNRS IFB's cloud infrastructure.

UC2 Cloud virtual pipeline for microbial genomes analysis

https://nuv.la/module/cyclone/Bacterial_Genomics

https://biosphere.france-bioinformatique.fr/catalogue/appliance/19

This use case was adapted to multi-cloud deployments. A successful validation was done on 2 sites of CYCLONE testbed (CNRS LAL and Exoscale).

UC3 Live remote cloud processing of sequencing data

https://nuv.la/module/cyclone/UC3_NGS_mapping

https://biosphere.france-bioinformatique.fr/catalogue/appliance/13

New features were added, the application recipe includes:

- A scalable virtual cluster with Docker-Swarm
- An isolated network with CNSMO-VPN
- A web interface to monitor Docker containers with ready-to-use bioinformatics tools
- And authentication mecanisms using the academic federation eduGAIN for web, SSH and virtual desktop access.

UC4 Virtual Power Plant

The Virtual Power Plant connects several small and medium sized power plants such that they are manageable as a single larger one. This comprises resources like wind, solar and bio-mass which are distributed geographically. The Virtual Power Plant provides a solution through joint control of these small and decentralized renewable energy resources. The Virtual Power Plant optimally 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 control the VPP as a tailored energy supply, a new approach of ICT for managing distributed energy resources is necessary. Each Virtual Power Plant provides a single operating profile to the energy system and can react in a flexible way.

In the use case the energy generation of the renewable energy resources is collected and aggregated by the VPP management to provide the basis for the operation of the VPP.

The components and applications of the use case have been defined and deployed on Nuvla using the capabilities of SlipStream to create a base component on which several of the more specialized components are based. This way the features needed in all components only need to be defined once are included using the base component. From the CYCLONE tools the CNSMO component is integrated into the use case to form the isolated network for secure data transfer between the VPP management and the DER (Distributed Energy Resource) cloud.

UC5 Internet of Services Lab (IoSL)

Student projects usually consist of three tier web applications with complementary apps which use the same backend. Some of these applications are data-intensive and strongly benefit from the one-click deployment of complex application with clusters and auto scaling mechanisms of Nuv.la. The students are encouraged to use containerization i.e. Docker to simplify the shipment of their software. The details of the deployment depend on respective projects. However, there are two basic setups. For the simple three tier web applications, first step is to deploy web, data and business logic as containers on one VM with one-click. Second step is to use one-click deployment of complex application by distributing one container on one VM and linking them together using Nuv.la. For data-intensive applications, deployment of complex application is more suitable as the analysis is run distributed on a cluster of data nodes.

4. Bioinformatics Cloud Federation Portal

The French Institute of Bioinformatics – IFB [8] consists of 31 bioinformatics platforms (PF) grouped into 6 regional centers spanning the entire French territory, and a national hub, the "UMS 3601–IFB-core", which is the representative of CNRS in this project. IFB has set up a federation of clouds, **Biosphere**, which relies on interconnected IT infrastructures of some IFB's platforms, providing distributed services to analyze life science data. Biosphere already gathers five academic clouds deployed in regional bioinformatics platforms (see Figure 8). Biosphere is used for scientific production in the life sciences, developments, and also to support events like cloud and scientific training sessions, hackathons or workshops.

Figure 8: IFB's academic clouds deployed in regional bioinformatics platforms.

Biosphere aims to provide multi-cloud deployments that help for example to combine larger CPU resources, to use different data sources, or to guarantee the availability of cloud resources. All the available bioinformatics cloud environments are registered in the Biosphere portal [9], more precisely in the appliance catalogue [10], like the ones related to CYCLONE bioinformatics use cases (see Figure 9). All these cloud recipes are ready to be deployed on one or several clouds thank to the CYCLONE technology.

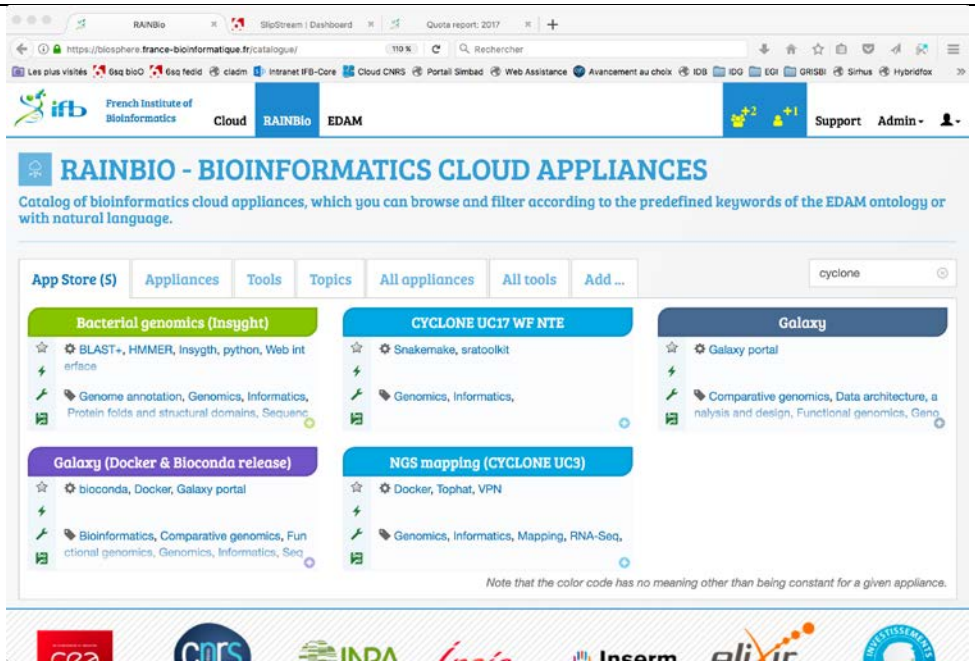

Figure 9: CYCLONE bioinformatics use cases referenced in CNRS IFB bioinformatics cloud portal.

IFB developed the Biosphere portal to hide to life scientists and engineers some complexities, but without losing the power, of the cloud. To do so, CNRS has applied the developments and technology of CYCLONE to setup its cloud federation. The biosphere portal is directly connected to SlipStream through a python API [11] that CNRS and SixSQ developed in collaboration (see Figure **10**). Then the IFB's users take benefit of being able to deploy personalized environment over private (IFB) and public clouds. The user accesses are managed with the CYCLONE Federation Proxy for the Web authentication, at the level of the Biosphere portal and inside the virtual machines, and for the SSH authentication of the users to their VM. The virtual research environments deployed by the biologists and bioinformaticians have secured network features relying on the CNMSO components (FW, VPN, LB and DNS).

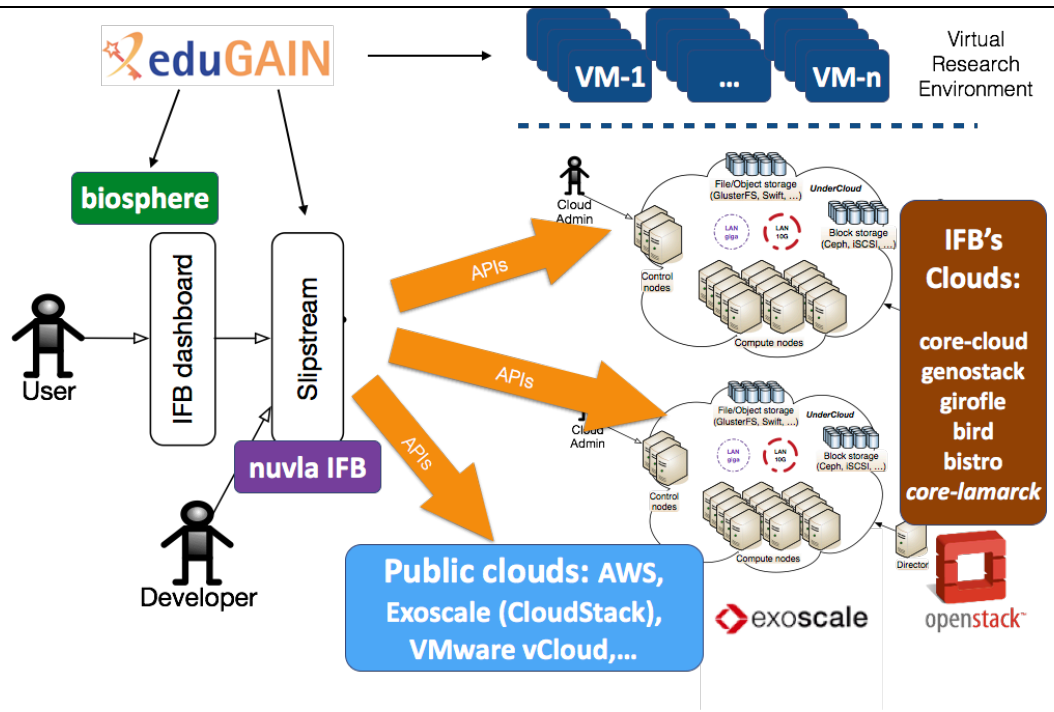

Figure 10**: Technical architecture of the IFB cloud federation relying on CYCLONE components.**

5. Conclusions

In this document, we provided the final evaluation of all CYCLONE use cases. The main motivations of these use cases were to help to identified relevant requirements, and to evaluate their satisfaction with newly-developed components, for the multi-cloud deployment of legacy and new applications.

The overall analysis of all CYCLONE use cases including the identified common requirements, and their level of achievements at the end of the project, demonstrates that these goals are reached. Among the 48 identified common requirements, 38 were satisfied, with almost all requirements qualified as “MUST be satisfied”.

We also detailed the achieved integration of CYCLONE use cases with both details for the recent achievements and a summary for the previously achieved integrations. Most of the use cases identified during the project were integrated with CYCLONE software to achieve multi-cloud deployments. There are available to the deployment as Nuvla recipes, and for the bioinformatics use cases, also as an entry in the IFB’s Biosphere catalogue.

An example of the reuse of CYCLONE developments is the IFB cloud federation and the Biosphere portal that relies on the CYCLONE component to federate several IFB’s cloud to provide life science scientists an engineer’s tools to deploy multi-cloud bioinformatics virtual environments.

6. References

[1] CYCLONE Deliverable D3.1, http://www.cyclone-project.eu/assets/images/deliverables/Evaluation%20of%20Use%20Cases.pdf

[2] CYCLONE Deliverable D3.2, http://www.cyclone-project.eu/assets/images/deliverables/Consolidated%20Evaluation%20of%20Use%20Cases.pdf

[3] CYCLONE use case portal, https://cyclone.france-bioinformatique.fr/usecases

[4] Fatih Turkmen, Simon N. Foley, Barry O'Sullivan, William M. Fitzgerald, Tarik Hadzic, Stylianos Basagiannis, Menouer Boubekeur, "Explanations and Relaxations for Policy Conflicts in Physical Access Control", International Conference on Tools with Artificial Intelligence (ICTAI), 2013.

[5] UC16 Integration with the Energy usecase (UC4), https://nuv.la/module/cyclone/CycLoneDACI/Demo_DACI_Finesce_Application

[6] DACI source code, https://github.com/cyclone-project/cyclone-DACI

[7] UC16 Slipstream Deployments, https://nuv.la/module/cyclone/CycLoneDACI

[8] CNRS French Institute of Bioinformatics, https://www.france-bioinformatique.fr

[9] CNRS Bioinformatics Cloud Federation Portal, https://biosphere.france-bioinformatique.fr/cloud

[10] CNRS Catalogue of Bioinformatics Cloud Appliances, https://biosphere.france-bioinformatique.fr/catalogue

[11] SlipStreamPythonAPI source code, https://github.com/cyclone-project/SlipStreamPythonAPI

7. Acronyms

B2B	Business to Business
CSP	Cloud Service Providers
DC	Data Center
DNA	DeoxyriboNucleic Acid
E2E	End to End
FLA	Federation Level Agreement
IaaS	Infrastructure-as-a-Service
IPR	Intellectual Property Rights
IT	Information Technology
MaaS	Metal as a Service
MCSP	Multi-Cloud Service Provider
NaaS	Network-as-a-Service
Net-HAL	Network Hardware Abstraction Layer
NFV	Network Function Virtualization
NGS	Next-Generation Sequencing
OLA	Operational Level Agreement
PaaS	Platform-as-a-Service
PC	Project Coordinator
PMB	Project Management Board
PoP	Point of Presence
SaaS	Software-as-a-Service
SCI	Smart Core Interworks
SDN	Software Defined Networks
SLA	Service level agreement
SP	Service Provider
TC	Technical Coordinator
TCTP	Trusted Cloud Transfer Protocol
TMB	Technical Management Board
VCF	Variant Call Format
VM	Virtual Machine
WP	Work Package
WPL	Work Package Leader

8. Annex – Description of use case common requirements

ID	Requirement	Use cases	Description
1	Cloud User Interface for service access based on web technology	1, 2, 3, 11, 12, 13, 17	Users are not familiar with low level interfaces like CLIs. The cloud resources management (VMs, virtual disks) can be done through a Web interface.
2	Cloud User Interface adapted to community usage	1, 2, 3, 11, 12, 13, 17	Users are not familiar with all the parameters and functionalities of a cloud. The web interface of the cloud should provide users with simple forms requiring only a few parameters and mouse clicks to launch the VM. Advanced forms with all the cloud parameters cans also be available but not by default.
3	Cloud User Interface adapted to mobile devices	3	Users can connect through a dedicated application on a mobile device.
4	Federated identity management	1, 2, 3, 5, 6, 11, 12, 13, 17	A common identity management system (based on RENATER/eduGAIN) across the cloud federated resources ensures consistent authentication of users.
5	Federated authorization management	1, 2, 3, 6, 11, 12, 13, 15, 17	Community administrators can define groups and associated authorizations to access cloud resources: VMs, storage volumes, etc.
6	One-click deployment of simple application	1, 2, 5, 6, 11, 12, 13, 17	The user should be able to run a VM by simply clicking on its entry in the appliance marketplace.
7	VM web interface secured by the identity federation	1, 2, 12, 13, 17	The access to the web interface of a scientific service provided by a VM is done according to the identity federation and authorizations.
8	End-to-end secure data management	1, 2, 4, 6, 11, 12, 13	The data must be secured throughout an application to avoid inappropriate access to data in transit or in storage. The securing mechanisms should be embedded with the VM creation process and totally transparent to the users. The connection between the cloud infrastructure and the user's computer need to be secured by default. The Trusted Cloud Transfer Protocol (TCTP) can be used to enable such end-to-

			end data transport.
9	VM isolated network	1, 2, 3, 12, 13, 17	All the VMs of one user need to be in the same local network isolated from other users.
10	VPN connectivity services	1, 2, 3, 4, 10, 12	The user can access to all of its VMs through a single point of entry based on VPN mechanisms.
11	FedId Access to storage	1, 2, 3, 12, 13	The user has access to the cloud storage according to its federated identity.
12	Community reference datasets	1, 2, 3, 12, 13, 17	Some scientific applications require common collections of data to compare the new experimental data to. For example, in bioinformatics the reference human genome hg19 (Human Genome version 19 or GRCh37) is required for analysing human biomedical data. Such community reference datasets consist of many files containing the reference data. It needs to be previously deployed by the cloud providers as public datasets available to all users of the considered community.
13	Access to public Cloud Services	4, 17	Running the VPP management service for prediction of energy generation, it is necessary to combine the publicly available data, such as weather forecasts, with data from the Big Data Analysis of DER data.
14	Cloud deployment according to medical data treatment certification	1, 17	In France, medical data can be stored and analysed only on a certified infrastructure. This requirement must be applied to cloud deployment as well.
15	Data erasure	1	A guarantee of the erasure of the data is required, to ensure that data subject to strict privacy restrictions is not persistent in the cloud infrastructure storage after the VM is terminated.
16	Deployment of complex application	2, 3, 5, 6, 12, 17	Running a complex application requires the deployment of several VMs copied from appliances with different complementary features.
17	One-click deployment of Complex application	2, 3, 5, 6, 12, 17	The user should be able to run a complex application with one click.
18	High-throughput storage	2, 3, 12	Some life science applications require high-throughput access to large datasets from many VMs.
19	Multi-VMs shared volume	2, 3, 12, 13	User VMs need to share the same volume.
20	Multi-clouds deployment of complex application	2, 3, 5, 6, 12, 17	A complex application can require to be deployed over two or more cloud infrastructures to obtain the necessary computing resources.
21	Multi-clouds distribution of community reference datasets	1, 2, 3, 12	The deployment of a scientific workflow over two or more cloud infrastructures requires that relevant community

			reference datasets used during the treatment are available in all of these clouds
22	FedId SSH connection to the VM	2, 3, 5, 6, 11, 12	The access to the VM is allowed according to the user identity and roles in the identity federation.
23	Elastic management of complex applications	2, 3, 6, 12	Running a complex application requires resource scale-up and scale-down options (in terms of computing and data)
24	Dynamic Network resource allocation	2, 3, 12	Deploying complex application can require to integrate ad-hoc network resource configuration into these applications, and to provide network connectivity reallocation according to dynamic modifications of the application architecture.
25	Multi-clouds distribution of user data	2, 3, 12, 13	Deploying a complex application requires the distribution of the user data in a secure way in several cloud infrastructures. The user data can be files, relational or NoSQL databases.
26	Ensure WAN High bandwidth links	1, 2, 3, 12, 14	Some applications require the dynamic allocation of high-bandwidth links between the platform producing the experimental data and the cloud storage. The dynamic network management will enable "live data processing" where the scientific instrument can be connected to the necessary computing resources to fully analyse the data as it is produced.
27	Archiving of raw data	3, 4, 10, 11, 12	The raw data will be stored as archives, and also put on the shared workspace to be treated.
28	Configuration of volatile disks	2, 3, 12	Some of the tools used to treat the raw sequencing data require intensive IO, for example because they are using the Hadoop paradigm, and then they require that volatile disks can be allocated on the local storage of the hypervisor hosts and plugged to the VMs for the duration of the process.
29	X11 remote display	2, 3, 12	Tools to analyse genomics data require graphical display, for example IGV - Integrative Genomics Viewer. Associated QoS should be enabled to the link between the user LAN and the DC.
30	FedId access for X11 service	2, 3, 12	The access to the VM providing the visualization features is done using the NX protocol and the X2Go software. The authentication is SSH-based and relies on the user identity and roles in the identity federation.
31	Group authorizations	3, 6, 12, 13, 15, 16, 17	Different cloud users may need to share their cloud resources (images, running VMs, storage...). The authorization relies on the static roles/groups to which the user belongs in the identity federation.

32	User-defined authorizations	1, 2, 3, 6, 12, 13, 15, 16	Cloud users can customize the authorization rules to access their own cloud resources: VMs, storage volumes, and more.
33	Deployment of applications to non dedicated cloud infrastructure	4	The main requirement for the cloud infrastructure for running the Energy use case is to run on Unix-like systems on non-dedicated infrastructure.
34	Guaranteed network performance (QoS)	2, 3, 12, 14	The transfer of large datasets (upload or download) between the user desktop (or a server storing locally the data) and the cloud storage, or the use of remote virtual desktop to connect to the cloud VMs, require guaranteed network performance.
35	VM with large memory	11	Some tasks like genome, metagenome and transcriptome assembly typically require large amounts of memory (from 64GB to 1TB).
36	Deployment of complex workflows using Docker containers.	5, 6, 11, 12	Analysis workflows rely on various applications that may not easily be installable on the same OS (for instance if they assume incompatible versions of a library). This problem can be overcome using containers like those provided by Docker.
37	Software deployment based on a Docker container	5, 6, 11, 12, 13	Some software tools of different scientific communities are already available as Docker containers (for example in life science). Developers and users should be able to deploy them in a Docker environment (that can be built upon a OS-minimal virtual machine like Alpine linux)
38	Public storage backend	13	Enable a public storage reachable from several cloud infrastructure.
39	Define authorizations to access user data on a public storage	13, 15, 16	In order to define the access to user data stored on a public storage, the authorizations based on the CYCLONE identity federation can be define by the owner of the data (an individual user or a group data manager).
40	Map a public storage as a filesystem in a VM	13	Some software tools require a POSIX filesystem to store data and parameters, like user data or preferences. To make these data and parameters available in several clouds, the local filesystem need to be be mapped on a public storage.
41	Horizontal elasticity	6, 12, 13	Vary the cloud resources allocated to the user by increasing and decreasing the number of virtual machines.
42	Vertical elasticity	6, 11	Vary the cloud resources allocated to the user by increasing and decreasing the size of a virtual machine (CPU and memory).
43	Manage academic license for	1	In academia, some tools are free of use for academic people

	the tools deployed in the images/VMs		but submitted to a commercial license for for-profit usage. But the license is most of time not manage with a license managent server but on a declarative manner. So the access to the concerned images and the capability to launch a VM of such image should be managed at the level of the images repository and of the clouds broker.
44	Load balancing for Scalability, reliability	14	Scalability, reliability and global footprint: In most 1-1 videoconferencing cases the endpoints communicate in true peer to peer fashion, so there is no system load during a video call. Peers connect to the signalling server during call setup and teardown, but need not be connected during the call. Therefore the impact on the signalling server of a large number of clients in minimal. However if the peers couldn't communicate directly and relaying has to be involved, then having multiple load-balanced TURN servers is important for scalability and reliability. Having multiple signalling servers would also aid system reliability.
45	Ensure WebRTC ports in the firewall	14	Ports open in the Firewall: o Web ports (443 or 50001 TCP) o RTP (9256-9500 UDP).
46	WebRTC signalling server securized	14	secure the signalling server and offer identity provider services to its clients. These could be provided by the CSP as a value-add to the basic WebRTC offering.
47	Redundant Storage	10	In the energy use case it is mandatory to ensure the availability of data that the operator requests for them.
48	Access Documentation	10	Any change in the system it is necessary to be recorded.

<END OF DOCUMENT>