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 D3.2

Consolidated Evaluation of Use Cases

Due date: 31/12/2016

Submission date: 17/02/2017

Milestone leader: 7 - 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 Services)
☐	CO:	Confidential, only for members of the consortium (including the Commission Services)
☐	WO:	Working document, only for members of the consortium.

List of Contributors

Participant	Short Name	Contributor
Interoute S.P.A.	IRT	D. Gallico, M. Biancani
SixSq Sàrl	SIXSQ	C. Loomis
QSC AG	QSC	R. Fischer
Technische Universitaet Berlin	TUB	M. Slawik, I. Zilci
Fundacio Privada I2CAT, Internet I Innovacio Digital A Catalunya	I2CAT	J. Aznar, E. Escalona, I. Canyameres
Centre National De La Recherche Scientifique	CNRS	C. Blanchet, B. Brancotte, J.-F. Gibrat
Universiteit van Amsterdam	UvA	F. Turkmen

Change history

Version	Date	Partners	Description/Comments
ToC	23/11/2016	CNRS	Table of contents
V1	9/12/2016	CNRS, i2CAT, QSC, SIXSQ, TUB	Initial version with current use cases description and evaluation.
V2	13/01/2017	CNRS	Revised version
V3	24/01/2017	CNRS, i2CAT, IRT, SIXSQ, TUB, UvA	Revised version with contributions from partners leading use cases.
V4	06/02/2017	CNRS, IRT, QSC	Revised version with contributions from partners. Editorial work.
V5	10/02/2017	CNRS, IRT, QSC, UVA	First internal review. Revised version
V6	15/02/2017	CNRS, i2CAT, QSC, SIXSQ	Second internal review. Revised final version.

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 and health, and energy developments. During the first months of year 2, 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 fulfill the needs. To face the growth of the number of use cases, the consortium has developed a web application to describe the use cases and publish them online: the CYCLONE use case portal.

This document provides a consolidated evaluation of all CYCLONE use cases, linked to the CYCLONE key technical areas. Focusing on the multi-cloud perspective, we provide a clarification of the multi-cloud requirements for current use cases. The multi-cloud deployments of these use cases clearly take advantage of the CYCLONE components such as SlipStream/Nuvla for the universal one click deployment of a complex application, the Federation proxy to manage authentication both for Web and CLI access and CNSMO to isolate the deployed applications from other cloud users.

This document also describes the purpose and developments related to the CYCLONE use case portal and the results of the investigation of CYCLONE industrial partners related to market analyses and business models. Finally, we propose a new use case in relation with the current roadmaps for personalized medicine that are being initiated in many European countries.

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.

Since the beginning of the project, the partners have elaborated the use cases that were initially presented in the DOA: a bioinformatics use case proposed by CNRS, split into 3 micro use cases (UC1 to UC3), and an energy use case proposed by QSC (UC4). CNRS and QSC have provided for all these scenarios an overall description of the use cases, and a set of associated requirements. 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. This work was described in deliverable D3.1.

During the first months of year 2, the consortium has identified several new use cases (UC5 to UC16) related to different areas: teaching, data management, energy, bioinformatics, etc. For each new use case, partners have provided a detailed description of the application, its objectives, the overall relation to CYCLONE and the expected impact in the domain of the application. To facilitate this work, the consortium has developed a web interface, the CYCLONE use case portal, to help new application providers to express their needs in terms of cloud resources, and to describe their application for a further analysis and selection by the CYCLONE partners.

For all the initial and new use cases (16), CYCLONE partners have maintained a balance between academic and industrial uses of the CYCLONE software. Industrial partners have also performed market analyses and business model investigations.

This document provides a consolidated evaluation of all CYCLONE use cases, focusing on the multi-cloud perspective. Section 2 contains a consolidated evaluation in terms of achievements and integrations with the CYCLONE key technical areas software developments. In section 3, we describe the purpose and developments related to the CYCLONE use case portal. Section 4 provides a clarification of the multi-cloud requirements for both initial and new CYCLONE use cases, and describes our proposal of a flagship use case for the next phase of the project. Finally, in section 5, we update our preliminary market analyses and business models investigations. The appendix summarizes all CYCLONE use cases with a brief description extracted from the CYCLONE use case portal.

2. Evaluation and achievements of all use cases

CYCLONE use cases come from several domains. Table 1 gives an overview of these use cases, their domains, the partner who is the principal investigator and the priority assigned by the consortium. Short descriptions of these use cases are available in the Annex of this document. For detailed descriptions, the reader should refer to the CYCLONE use case portal [1].

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. Moreover, the French Institute of Bioinformatics (CNRS unit, [2]) has a model for cloud that relies heavily on self-service and on-demand features where users are deploying their own resources without any intervention of sysadmins. Thus, the application deployments are very volatile, require high adaptability to the current state of the available clouds and must be fully automated.

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

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

2.1. Use cases requirements and implications for the CYCLONE software

2.1.1. CYCLONE Key technical areas and related software developments

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 which aims to bring typical network services and integrate them on-demand and automatically with complex applications, adding value and features to them. The CYCLONE component that supports these services and developments is the OpenNaaS CNSMO software framework that is integrated in the SlipStream App Store to facilitate the network services.

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).

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
Network: **4) Dynamic network configuration and management**	CNSMO-VPN CNSMO-FW

The SlipStream Service Catalog provides the means for collecting cloud provider information (through “service offers”) and for selecting appropriate offers based on application requirements (or “policies”) manually or automatically. To date, the cloud resources targeted by the matchmaking, brokering, and mediation features have been solely computing resources. However, many cloud applications and several of the defined CYCLONE use cases require matchmaking, brokering, and mediation based on storage resources, or specifically the data within the storage resources. During this last year of the project, data management features will be added to SlipStream, expanding the Service Catalog to cover data resources on the cloud.

2.1.2. Use cases requirements regarding the key technical areas

For all CYCLONE use cases, we identified and defined precisely the links between their 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 already developed by CYCLONE partners. It clearly shows strong needs from most use cases for automated cloud deployment, for complex applications and on multi-cloud infrastructure. This table also shows many requirements for computer security, from reliable federated identity management, both in Web and command line environment, to isolation of the virtual environment with a private network filtered with firewall rules.

Table 3: Details of use cases requirements and software implications

	Key Tech. Areas	Cloud deployment	Complex App	Multi-cloud	Federation Proxy	Web authn	SSH authn	CNSMO-VPN	CNSMO-FW
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	✔	–	✔	✔	✔	-	-	--
UC14	3,4	-	✔	✔	-	✔	✔	-	✔
UC15	**1,2,3**	✔	✔	✔	✔	✔	✔	-	-
UC16	**1,3**	✔	✔	✔	✔	✔	✔	-	-

2.2. Achieved integration of use cases

The four initial use cases (UC1-4) have been deployed during the first year of the project, and extended with multi-cloud features during the second year. Table 4 summarizes the current status of the different use cases regarding their integration with the CYCLONE software tools (according to the three main domains defined in the Table 2).

The deployment of three of the new use cases of Year 2 was also initiated during the hackathon for CYCLONE use cases organized by CNRS. This hackathon brought together both use case and CYCLONE component developers during two days. The expected outcomes have been to have i) validated deployment recipes in SlipStream/Nuvla for the use cases, and ii) a good integration of the applications with the CYCLONE components. The goal has also been to reach a better understanding between developers of the use case applications and of the CYCLONE software components.

Five CYCLONE use case developers attended the hackathon (UC2, UC3, UC11, UC12, UC13), all from the bioinformatics domain, with tools developers for a total of 12 attendees. Several CYCLONE tools were presented by the CYCLONE partners: SlipStream cloud broker, IFB bioinformatics cloud portal, Federation proxy and CNSMO. A GitHub repository was created to manage the presentations and the use cases integration (https://github.com/cyclone-project/usecases-hackathon-2016). The infrastructure used during the hackathon was based on the SlipStream/Nuvla CYCLONE instance, the IFB bioinformatics cloud portal and three clouds sites (CNRS LAL CYCLONE testbed site, CNRS IFB Pilot, Exoscale) relying on two different cloud frameworks (OpenStack and CloudStack).

Thanks to the hackathon, most of the represented use cases (UC2, UC3, UC12, UC13) are now ready for deployment with the SlipStream cloud broker on the CYCLONE testbed. Some of them were also integrated with the Federation Proxy and CNSMO components. Developments were continued after the hackathon days and will be extended during CYCLONE Year 3.

Table 4: 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	**In progress**	**In progress**	-
UC7	Open Scientific Data	**In progress**	-	-
UC8	Benchmark Driven Placement	**In progress**	-	-
UC10	Smart Utility 4.0	✔	-	✔
UC11	Assembling genomes from sequencing reads	**In progress**	**In progress**	**In progress**
UC12	Metagenomics	✔	**In progress**	**In progress**

UC13	Shared environment between cloud Galaxy portals	✔	**In progress**	**-**
UC14	WebRTC video-conference solution	-	-	-
UC15	Interactive Authorization Policy Management for Multi-cloud Applications	Planned Y3	Planned Y3	Planned Y3
UC16	Attribute-based Authorizations with XACML	**In progress**	**In progress**	**In progress**

3. Use case portal

With the growth of the number of use cases and to facilitate their analysis, the consortium has developed a web application to describe the use cases and publish them online: the CYCLONE use case portal [1]. Using the CYCLONE use case portal, an application investigator can describe finely his application according to different topics (description, list of requirements, user stories, workflow, etc.), and submit it as a proposal of new use case to the CYCLONE consortium.

3.1. Description of the CYCLONE use case portal

The CYCLONE use case portal is a web application developed with Python and with the Django Web framework. The data model is centered on a use case as an object composed of the following elements: Actors, Implementation tasks, KPIs, Requirements, Responsible People, User stories and User story abilities.

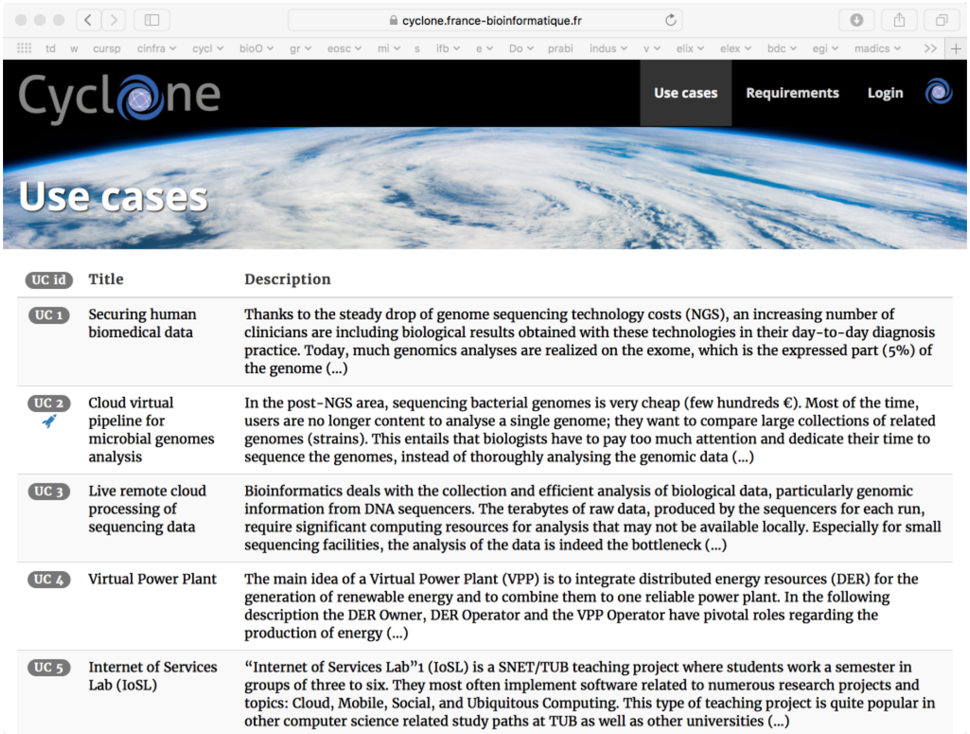


Figure 1: Main view of the CYCLONE use case portal.

The portal presents from the main page different views:

- the main view listing the public use cases (see Figure 1) with the use case ID, the title, a brief description and in some case a direct link to the cloud deployment recipe in CYCLONE Nuvla,
- the use case detailed view (see below for details),
- the Requirements view listing the main elements of each requirement (ID, title and description),
- the Requirement detailed view including the implementation tasks related to a requirement,

- the sign-in interface proposing both eduGAIN and local accounts (see Figure 2). The portal uses the standard Django's user management system, extended with the CYCLONE Federation Proxy.
- the Stories and Actors view (if signed in)
- the Tasks view (if signed in),
- the Summaries menu (if signed in), providing different tables cross-linking the use cases, requirements, stories, workflows plans of deployment, test scenario, implementation tasks and software components.

The use case detailed view includes:

- the whole description of the use case,
- the workflow description and figure,
- the user stories (ID and description),
- the list of requirements related to that use case.

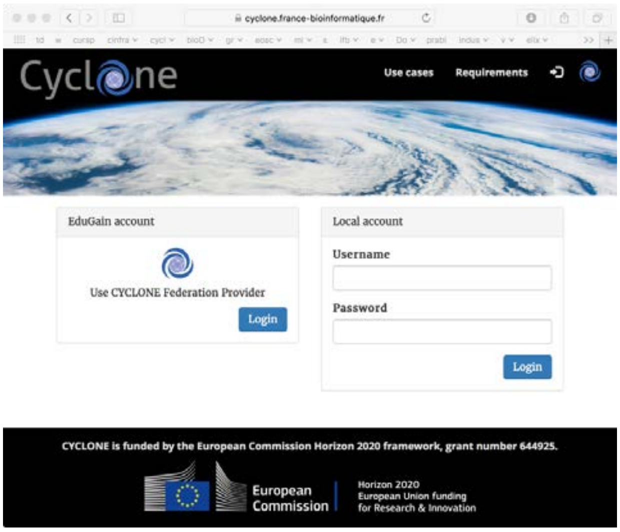


Figure 2: Sign-in interface of the CYCLONE use case portal.

3.2. Usage by scientists

The CYCLONE use case portal allows scientists to have a clear description of applications being or already integrated with multi-cloud functionalities provided by the CYCLONE software tools. A scientist willing to propose a new use case to CYCLONE can ask to sign in to the portal with his eduGAIN identity. Once his account is validated by the admin, he can describe his use case according to the following topics:

- title,
- demonstrator's URL,
- description,
- workflow description and figure,
- requirements (all the previously identified ones are available),
- actors and user stories,
- test scenarios.

4. Multi-cloud implementations

CYCLONE initial use cases (UC1 to UC4, see Table 1 for an overview and also D3.1 [3]) have been integrated in the CYCLONE testbed during the first period of the project. They have been initially selected for their relevance to the goals of CYCLONE project and so, have requirements in relation with a deployment on a multi-cloud infrastructure.

The main examples of these initial use cases are UC2 and UC4 that have the more important needs for multi-cloud. UC2 is a good example of a bioinformatics application requiring multi-cloud features such as high-scalability of computing resources. Indeed, this use case deals with genomic data that can be very large and numerous, and so requires large and elastic cloud resources. UC4 must bring together many energy production and monitoring centers to face the future 20-20-20 landscape. To be able to satisfy all the related challenges to connect different platforms in a secure environment, a multi-cloud deployment is required.

Another example is UC1 that deals with the security of human biomedical data. It requires high-availability to satisfy guaranteed services to hospital staff, but it also needs to be deployed on secure and certified cloud infrastructure. This can be guaranteed by the ability to deploy the application on different clouds selected against their certification and availability at the time of the request. If a cloud cannot provide resources at the time of the required treatment, another one can take over. UC3, regarding the size of genomic data to analyze and the quality of service to provide to their client, needs both scalability and high-availability features.

The next section presents more details about the two use cases UC2 and UC4, and proposes a new one (UC17) to reinforce the relevance of the developments done in CYCLONE.

4.1. Current multi-cloud achievements for the analysis of Microbes

As stated previously in deliverable D3.1, in the use case UC2 description [4], sequencing bacterial genomes is now very cheap (few hundreds €), and most of the time, users are no longer content to analyze a single genome; they want to compare large collections of related genomes (strains). In this context, Lacroix et al., *Nucleic Acids Research*, 2014, have developed Insyght, a comparative genomic visualization tool that tightly integrates three complementary views. Insyght consists of 3 components: a pipeline of Perl scripts to compute the required data, a relational database to store these data and the visualization tool itself that queries the relational databases and presents these data in a user-friendly way. The platform can automatically launch a set of bioinformatics tools (e.g. BLAST, PLAST) to analyze the new experimental data. These tools require several public reference data collections to perform their analyses.

To provide a more comprehensive coverage of microbial diversity, genomic comparisons are carried out with all the complete, non-redundant, bacterial genomes available so far in the public collections. A careful selection among the latter leads to a set of about 4500 non-redundant, complete genomes representative of the different branches of the microbial tree of life known today (large portions of this tree have not yet been explored).

4.1.1. Achieved multi-cloud deployment for microbial analysis

From a computing standpoint, this represents about 10 million genome pair comparisons to carry out. Previous computations with about 2700 bacterial genomes, generated 1.2 TB of raw, compressed data. These comparisons are independent and thus can be distributed in batches over many cores. Typically, 20 genome pair comparisons take 5 minutes, hence in total the comparisons require about 42,000 CPU hours. Considering that many CPUs can be obtained from several cloud infrastructures, the time of computation can be reduced to days with thousands of CPUs, or to hours with tens of thousands of CPUs.

The Insyght platform comprises two kinds of component: (i) a master to run the workflow and schedule the genome comparisons, (ii) nodes to perform the genome comparisons. A complete deployment requires one master and several nodes sharing the same storage. The number of nodes varies with the size of the genome dataset to analyze. Previously, they were both deployed with a single image, which moreover required to be imported in the targeted cloud site. As many clouds neither allow the import of custom VM images nor the utilization of a very generic deployment recipe, these two deployment components have been described as separate recipes in SlipStream/Nuvla [5].

This automated deployment of a complex application has been tested on two CYCLONE testbed sites, CNRS cyclone-fr2 and Exoscale [10]. With the CNSMO-VPN tool, the whole virtual infrastructure is isolated for security and operation purposes (for example to ease the exchanges of the biological data between the nodes) from the others cloud users and VMs. The multi-cloud deployment of the UC2 then benefits from the following CYCLONE components:

- SlipStream for the generic deployment recipe compliant with several flavors of clouds
- SlipStream for the one click deployment of a complex application
- Federation proxy to manage authentication both for Web and CLI accesses
- CNSMO to isolate the whole deployment form other cloud users and VMs.

4.1.2. Further developments and perspectives for microbial analysis

Such comparison of reference bacterial genomes is done once to define the reference dataset for the Insyght platform. After this initial step, it must be regularly updated with new bacterial genomes that are sequenced and published every month in scientific journals. However, such a comparison workflow can also be used by researchers to analyze their own extended set of bacterial genomes, requiring each time a multi-cloud infrastructure.

CYCLONE Cloud federation can provide life science researchers with their own comprehensive genome comparison platform that can be deployed in one click over one or more cloud sites. Further developments will also add a high-availability feature to the master and scalability and elasticity to the nodes deployment.

4.2. Current multi-cloud achievements for the energy use case

The Climate change goals of the European Union [6] and the goals of the German energy transition, the so called “Energiewende”, lead to a tremendous change in the architecture and processes of energy supply and power generation as already described in D3.1 [3]. Within this picture the approach to integrate Big Data and Cloud Computing with embedded systems and interactive services can be combined to the term “Smart Core Interworks” – platform (SCI, UC10 [7]).

Virtual Power Plants (VPP, described in UC4[8]) comprise two ideas. Firstly, the integration of distributed energy resources (DER) like windmills/wind plants and photovoltaic plants into one reliable power plant. Secondly the necessity to exchange countless and heterogeneous data (like power, voltage of solar panel, etc.) for energy production management and demand side management on the consumer side. This directly involves a huge number of different parties leading to numerous clouds with a remarkable diversity of ownership.

Hence, to balance energy demand and production with the major requirement of real time capability in the energy domain, real time reaction on changing operational and/or environmental conditions is mandatory. Otherwise it will impossible to avoid instabilities or even blackouts in the energy grid. Therefore, the presence of numerous clouds and the need of stability within the VPP and the grid leads to the importance of integrated and consolidated multi-cloud management.

4.2.1. Achieved multi-cloud deployment for Energy use cases

The SCI platform comprises two main components:

- A stack of uniform clouds running the workflow, scheduling the virtual power plant operation and data aggregation, ensuring the real-time and autonomous operation.
- A stack of heterogeneous clouds performing the prediction of power production and managing demand response forming virtual utilities, ensuring that the energy services can be accounted.

As for the microbial use case [4], there was a single image, deploying both components, which also required to be pre-imported in the targeted cloud sites. Now, the two SCI components are available as SlipStream/Nuvla recipes [9]. So far, deployments (requiring many VMs of these two flavours) have been tested on the AWS and the QSC clouds. The system scales with the number of DERs forming the VPP. The whole virtual infrastructure is isolated, for security and operation purposes, from other cloud users and VMs using the VPN offered by OpenNaaS-CNSMO.

The multi-cloud deployment of the UC4 and UC10 are benefitting from the following CYCLONE components:

- SlipStream/Nuvla image creation at deployment time – actuality of cloud images. Definition of application specifics in VMs using these base components
- SlipStream/Nuvla for the one click deployment of a complex application
- OpenNaaS-CNSMO isolated network, integrated image, Multi-cloud VPN
- SlipStream/Nuvla access security and documentation beginning at the definition level of applications
- OpenNaaS-CNSMO Configurable Network security.

4.2.2. Further developments

Further developments will add high-availability and real-time features to the VPP VMs and scalability and elasticity to the overall node deployment.

4.3. Proposal of a new multi-cloud use case: Genomic Variant Analysis for Cancer and Rare Disease Diagnosis (UC17)

4.3.1. Overall description

In the context of personalized medicine, an increasing number of patients suffering from a cancer or a rare disease will have their genome sequenced. The objective of this sequencing is to determine which variants (point mutations, gene copy number variation, insertions/deletions, DNA rearrangements) are observed in the genome. These variants are then annotated, which means that an expert tries to infer which regulatory motif on DNA or gene or metabolic/regulatory pathway is affected by the observed variants and determine whether the observed modifications in this motif, gene or pathway could be involved in the disease. Based on the expert's findings, a specific (personalized) treatment, if one exists, is prescribed to the patient.

From a bioinformatics point of view, this analysis starts with the mapping of the sequencing reads on a reference human genome and the determination of variants. This first stage, named variant calling, is carried out by a pipeline that has the reads as input, chains together several analysis programs and outputs a VCF file (i.e., a file, in a standardized format, that contains the location of the different types of variants having been found in the genome). The next stage consists in annotating the variants, as explained above. This stage is less automatic and involves running several tools locally, using web services, and accessing several databases. These two stages can be run on a cloud infrastructure that is not certified for biomedical data as far as the studied data have been anonymized before.

A third step is particularly important with the ability to combine the results of this annotation stage with the patient clinical data to filter out non-relevant variants. But patient clinical records are kept in hospital and cannot leave the premises of the hospital.

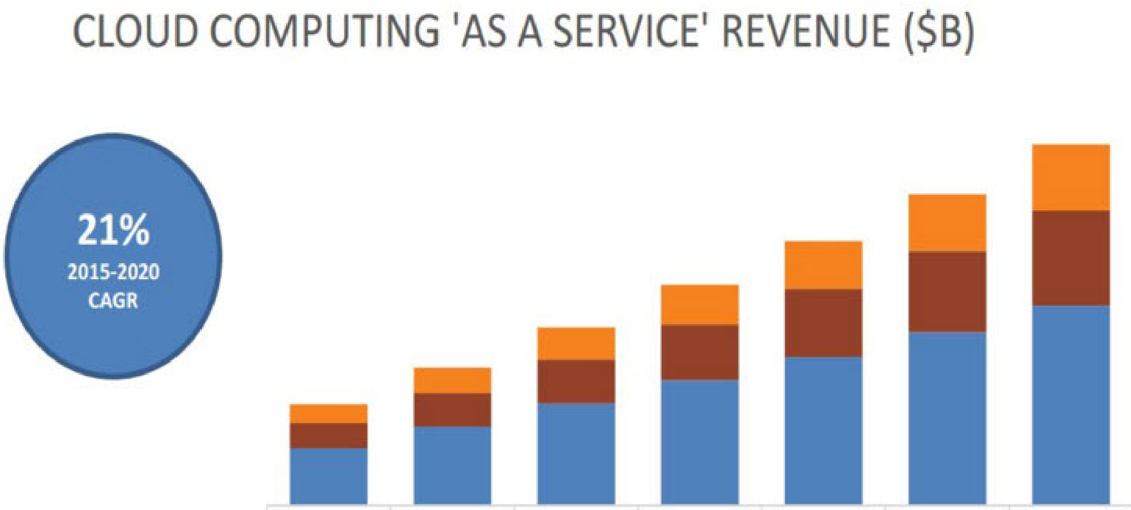
4.3.2. Multi-cloud requirements

To perform this type of analysis would involve a master VM that runs the pipelines and possibly executes workers on the main Cloud to perform the first two stages (variant calling and annotation). The master would also execute, asynchronously, mediators on VMs running on hospital clouds. These mediators endowed with web interfaces (possibly different based on the credential of the user) would allow the user to query the local hospital database without having access to all the patient records. All the data handled in this type of analysis are sensitive, the master and the mediator VMs need to be isolated within a VPN.

5. Market analysis and business model investigations

Enterprises heavily investing in cloud computing across multiple service providers do understand the tremendous effort it takes to manage services and policies in a unified fashion. Moreover, an increased move to public infrastructures with well-defined interfaces for delivery and service provision elevates the necessary skillset within the respective IT organization. Possibly the number of admins to keep scope of the daily work must be increased. Gaining and keeping the knowledge is hard to achieve and therefore many CIOs are looking to multi-cloud management platforms as the solution to their problems. Additionally, the growing impact of so-called IoT services brings another grade of complexity into the operating models. To keep within given limits in operational expenditures, again multi-cloud services will be preferred over single cloud or even legacy architectures. This also leads to the need of adequate management toolsets.

The following diagram gives a forecast on “cloud as a service revenues” and is underpinned with references provided by e.g. IDC.



	2014	2015	2016	2017	2018	2019	2020
PaaS	$2.3	$3.1	$4.0	$4.9	$5.9	$6.9	$8.1
Infrastructure SaaS	$3.1	$4.1	$5.4	$6.8	$8.4	$10.0	$11.6
IaaS	$7.0	$9.7	$12.5	$15.4	$18.2	$21.2	$24.5

Figure 3: Cloud Computing revenue forecast up to 2020 (Forbes)

- By 2020, over 50% of all new applications developed on PaaS will be IoT-centric, disrupting conventional architecture practices [11].

- By 2020, penetration of software as a service (SaaS) versus traditional software deployment will be over 25%. Packaged software will shrink to 10% of new enterprise installations. In 2012, 60% of SaaS providers delivered services from do-it-yourself datacenters. By 2020, that number will be less than 50%. SaaS will begin to "live" in the cloud [12].
- Worldwide spending on public cloud services will grow at a 19.4% compound annual growth rate (CAGR) from nearly $70B in 2015 to more than $141B in 2019. IDC predicts Software as a Service (SaaS) will remain the dominant cloud computing type, capturing more than two-thirds of all public cloud spending through most of the forecast period. Worldwide spending on Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) will grow at a faster rate than SaaS with five-year CAGRs of 27.0% and 30.6%, respectively [13].

Today cloud landscape is dominated by four big players, i.e. Amazon, Microsoft, Google and IBM. They account for more than 50% of worldwide cloud landscape revenues. Their portfolio ranges from bare CPU usage (e.g. Amazon EC2), via storage solutions and application hosting up to the provision of complete software service environments (e.g. MS Office 365).

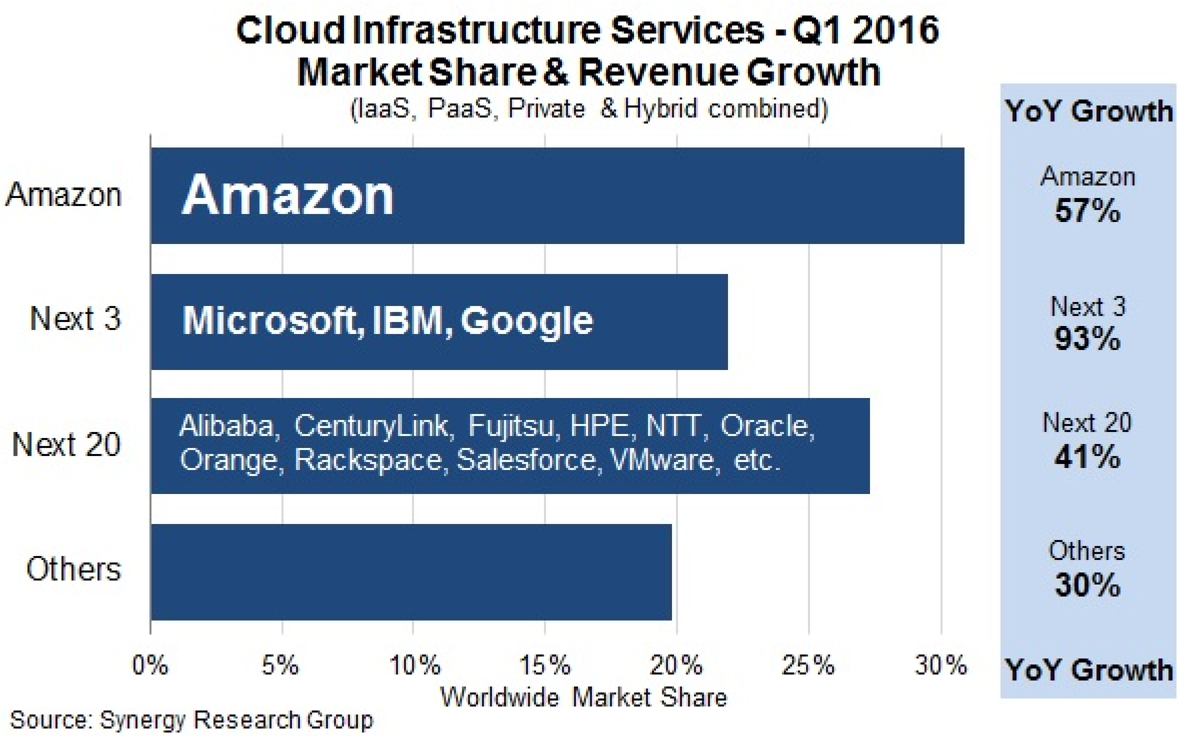


Figure 4: Cloud worldwide landscape and major players

In that sense, Multi-cloud represents the opportunity for smaller players, orchestrating different CSP[1] to provide services as a MCSP[2] to position them in a market which is dominated by big firms. All that without huge upfront investments in infrastructure while proposing its own unique selling points that cannot be provided solely by one of the "big four" mentioned above. Let us describe this picture in more detail on the following pages.

[1] Cloud Service Providers
[2] Multi-Cloud Service Provider

A part from the obvious opportunities for the large plethora of minor cloud providers, the multi-cloud framework can provide a large set of benefits:

- **Autonomy**. Applications deployed on a multi-cloud environment, provides the advantage of reducing the dependence on a single provider. This relative independence can help the end-user to negotiate more favorable service-level agreements, pricing or both.
- **Avoid Vendor Lock-in**. The multi-cloud customer is offered a wider range of Cloud services with fewer constraints.
- **Flexibility**. Customers can choose to keep some applications on-premises while migrating others to one or more public clouds. Multi-cloud solutions are also much more resilient during outages and maintenance.
- **Advanced cloud capabilities**. Different cloud providers support different solutions and offer constantly evolving arrays of features and capabilities. Some offerings will not be supported by any other cloud provider. Multi-cloud gives you the ability to select your favorite data services from any provider to create a customized solution that best fits your organization's needs.

Multi-cloud framework is an emerging trend which has been investigated in the past few years, but has not yet been fully applied to the actual cloud market. In that sense, valid and market-proof business models need to be carefully designed and consolidated. Nevertheless, we can try to sum up some requirements which can drive the design and the adoption of new models:

- The first and most crucial challenge for a service provider is to create transparency throughout the entire service chain, across all partners and layers. Cloud-based service items are managed on a higher abstraction level, based on the relevant service parameters and service levels, in conjunction with the cloud service provider. Creating transparency means structuring all the items to deliver a holistic product portfolio and service model. This requires a sophisticated methodology and appropriate software tools to cope with the challenges of the multi-cloud environment.
- Careful management and integration of SLA (Service level agreement) with OLA (Operational Level Agreement) and FLA (Federation Level Agreement) at any levels (technical, business, horizontal and vertical levels)
- Implementation of standardized software tools and methodology ensures reliable products and high service quality. A service must be defined, controlled, and guaranteed with respect to customer delivery. To provide such services, a provider needs an environment where all guaranteed functions can be assured in a reliable and reproducible way. This includes planning and engineering, as well as service delivery and verification of the services within a service architecture framework.
- Integration at OSS/BSS level for service management, with proper tool/framework solution is among the most important elements of an integrated approach for multi-cloud service management.

From the technical perspective, the central act of simplifying transactions between a multitude of cloud consumers and cloud providers is cloud federation, and cloud brokerage is one necessary technological capability of accomplish this. Depending on how the activities relating to federation are divided between different organizations or actions, a variety of business models may emerge. One way of categorizing these business models is to view them as "broker-driven" or "federation-driven".

- Federation-driven business models: implies a deeper collaboration between the involved cloud suppliers (that share cloud resources and hence form the federation). For example, members of a cloud federation may sign a joint federation-level-agreement (FLA) instead of bilaterally signing a multitude of contracts between the cloud suppliers in the federation. The FLA generally creates a longer-term and more reciprocal relationship between its members than could be expected from a purely free market transaction perspective. The technology that the cloud federation uses internally to share resources can vary between peer-to-peer and centralized approaches.

- Broker-driven business models: in this case, there are no constraints on how the cloud broker communicates with each separate cloud provider, or whether there are special contractual relationships between the cloud broker and any of the cloud suppliers. Notably, also in broker-driven business models, no assumptions are made about the ownership or control of the entity carrying out the brokering activities. It is hence possible for a group of cloud providers to set up a common (external) cloud broker to participate in the cloud market more efficiently, or the broker may be a separate business entity from the cloud suppliers.

More in detail, a **self-managed federation** is a federation that is in completely controlled by the federating cloud providers. In this case, no activities are outsourced to a third party.

In the cooperative model, a group of CSPs federate cloud resources, and establish joint customer-facing marketing capabilities that seek to package offerings responding to cloud consumers' needs, dividing the tasks of fulfilment and revenue sharing between CSPs. This approach creates synergies in marketing compared to each CSP doing marketing separately. However, the SLA remains between the cloud consumer and the CSP, and each CSP maintains its own brand identity. The operation and maintenance of the common marketing interface (e.g. common sales portal) is done by one of the cloud service providers.

In the context of an **externally managed federation,** a third party is taking over more and more responsibilities from the cloud service providers depending on the level of integration to be achieved. At the end, a third party can be in full control of the federation and the cloud providers are only supplier of their cloud resources.

The cooperative model can also be carried out by involving a third party entity to do promotion and the matchmaking between cloud providers and cloud consumers by establishing a marketplace (fig. 4.3). This marketplace offers only the functionality of a shopping window. This model may be used to create service partner networks to increase the marketing reach of cloud service providers. This approach may be viable, if the CSPs themselves have limited marketing capabilities possibly due to their small size or if they jointly have poor reach in a particular foreign market.

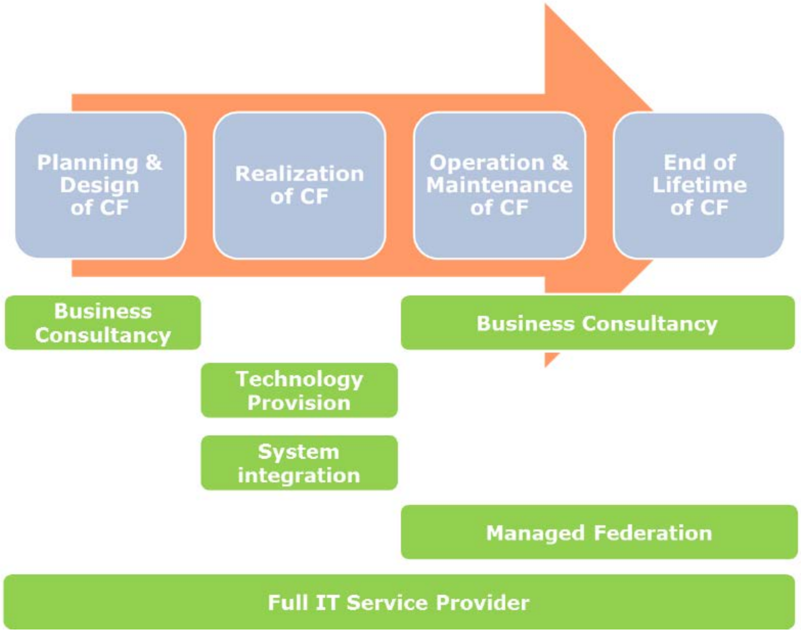


Figure 5: Cooperative model

Finally, the adoption of a multi-cloud panorama can provide different business opportunities to the players of the cloud landscape, providing the emerging of new demands and new roles which can satisfy them. In the following we have tried to depict some of the new roles and related business models.

The "**Business Consultancy Service**" model is mainly relevant in the planning phase, in the operational phase of a cloud federation and, in addition, at the end of a cloud federation's lifetime. This form of business consultancy can be offered in different directions. Business consultancy services can be offered to a single cloud provider, a cloud federation, cloud consumers and IT companies.

The **"Technology Provision"** model is only relevant in the realization phase of multi-cloud framework. It addresses cloud providers to establish their own federation network or IT companies to establish their own federation offering by providing them with the relevant technologies and solutions. Furthermore, IT companies with the purpose to become a cloud federator/cloud broker can be provided with the relevant technologies and solutions as well. The final integration work is either done by the cloud providers/IT companies themselves or a third-party company (system integrator).

The **"System Integration**" model is also addressing the needs of the realization phase. Instead of the pure provision of technology components and solutions to create a cloud federation, the system integrator is realizing the cloud federation by combining the various solution elements provided by technology providers.

The **"Managed Federation"** model is mainly relevant in the operational phase of a cloud federation. It is all about the operation of the technical infrastructure, the federation/brokerage portal and the execution of the contractual aspects by providing additional services. Depending on the scope of a federation (lightweight federation with no integrated offerings to fully integrated federation) the managed federation model can differentiate various business models. In case of a lightweight federation with no integrated offerings, the managed federation model could just come up with a portal where cloud providers can highlight/describe their offerings and potential cloud consumers can search for those offerings.

The **"Full IT Service Provider**" model covers the complete lifecycle by accompanying cloud providers on their roadmap to set up and run a cloud federation. There exist two main service delivery concepts for the "full IT service provider" model.

On the one hand cloud federation can be offered exclusively to the handful of cloud providers that want to federate their offering. In this case, individual adaptions are possible based on the providers' requirements. The sustainability of this kind of cooperative federation arrangement is however nontrivial, and is a matter of economic utility to the various players involved. Each firm in the cooperation needs to compare the utility of participating in a federation to its opportunity costs. For small cloud providers, a cooperation arrangement with other firms with roughly equal financial resources and complementary offerings may be feasible, as any member of the cooperation would lack the resources to obtain exclusive access to a significant share of the other member's resources.

On the other hand, if market uncertainty (demand volatility) is high, any single player (with possibly greater financial resources) would be less willing to make fixed investments to obtain exclusive access to the resources of any of the players. These factors reduce the opportunity costs of participating in a federation. If a large player is involved, the setting is different. The large firm may see higher risks and costs in collaborating with smaller players compared to expanding its own cloud resources (for example, through building new capacity or acquiring competitors) and retaining full control on their management and development over time.

6. Conclusions

In this document, we presented a consolidated evaluation of the CYCLONE use cases. There are currently 16 use cases. Four (4) were included in the initial project proposal and twelve (12) new ones were proposed during the project by the CYCLONE partners (sometime in collaboration with third-party). These new use cases were selected according to the relevance or their requirements to the project goals in terms of cloud deployment, security and network features. Particularly, we put the focus in this document on two bioinformatics (UC2) and energy (UC4) that we have identified as having the more important needs in terms of multi-cloud deployment.

The initial use cases (UC1 to UC4) have been deployed on the CYCLONE testbed and have demonstrated the benefits of the different CYCLONE tools for multi-cloud deployments. All use cases benefit from the simplified deployment of complex applications in SlipStream/Nuvla. Some also take advantage from the security features provided by the Federation Proxy and the associated eduGAIN authentication both at the web and SSH levels. Finally, three of them (UC2, UC3 and UC4) demonstrate the benefit of running in an isolated network provided by the CNSMO components. Further integration of the new use cases is in progress, however these use cases have already demonstrated the benefit of CYCLONE tools to help their deployment on cloud facilities.

Regarding the multi-cloud needs, the bioinformatics use case UC2 has a strong requirement about the scalability of computing resources. Due to the microbial genomic data that are treated, the required computing resources can be large, and therefore require the scalability of the cloud resources. For its part, the energy use case UC4 must gather many energy production and monitoring centers to face the future 20-20-20 landscape. To be able to satisfy all the related challenges to connect different platforms in a secure environment, a multi-cloud deployment is required. The multi-cloud deployment of these two use cases benefit from the CYCLONE components such as SlipStream/Nuvla for the universal one-click deployment of a complex application, the Federation proxy to manage authentication both for Web and CLI accesses and CNSMO to isolate the current applications from other cloud users.

Market analysis and business model investigations show that worldwide spending on public cloud services will grow at a high steady compound annual growth rate (roughly 20%). Multi-cloud also represents the opportunity for smaller players to challenge the current four (4) main players, providing benefits like autonomy, flexibility, advance capabilities and avoiding vendor lock-in. From the technical perspective, the central act of simplifying transactions between a multitude of cloud consumers and cloud providers is cloud federation, and cloud brokerage is one necessary technological capability of accomplish this. Also, the adoption of a multi-cloud panorama can provide different business opportunities to the players of the cloud landscape, providing the emerging of new demands and new roles which can satisfy them.

Finally, we propose further developments and integrations focusing on a new use case about the Genomic Variant Analysis for Cancer and Rare Disease Diagnosis. This new use case is proposed in the context of the roadmaps for personalized medicine being initiated in many European countries. The purposes are to provide mechanisms to solve the difficult requirements of such health applications and to demonstrate the relevance of the multi-cloud developments realized by the CYCLONE consortium in terms of cloud elasticity, security and network features.

Note that a detailed comparison between SlipStream and market alternatives has been carried out. The details of that analysis can be found in the deliverable “Solutions for Non-functional Aspects of Cloud Applications” (D6.3) [14].

7. References

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

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

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

[4] Description of CYCLONE use case 2, https://cyclone.france-bioinformatique.fr/usecases/view/125

[5] Cloud deployment recipe for CYCLONE use case 2, https://nuv.la/module/cyclone/UC2_Microbial_genomes_analysis/UC2-Demo-Brussels

[6] http://ec.europa.eu/clima/policies/strategies/2020/index_en.htm

[7] Description of CYCLONE use case 10, https://cyclone.france-bioinformatique.fr/usecases/view/163

[8] Description of CYCLONE use case 4, https://cyclone.france-bioinformatique.fr/usecases/view/157

[9] Cloud deployment recipe for CYCLONE use case 4, https://nuv.la/module/cyclone/Energy-UC_VPP/5190

[10] https://exoscale.ch

[11] Predicts 2016: PaaS Innovation Continues Unabated 10 December 2015 G00293230. Analysts: Yefim V. Natis, Lydia Leong, Benoit J. Lheureux, Paul Vincent. https://www.gartner.com/doc/3177118/predicts--paas-innovation-continues

[12] IDC 50th Anniversary, Transformation Everywhere. https://www.idc.com/IDC_50anniversary/download/IDC50thAnniversary.pdf

[13] Worldwide Public Cloud Services Spending Forecast to Double by 2019, According to IDC. https://www.idc.com/getdoc.jsp?containerId=prUS40960516

[14] "Solutions for Non-functional Aspects of Cloud Applications" (D6.3), M. Zivkovic, C. Loomis et al. See http://www.cyclone-project.eu/documents.html#deliverables.

8. 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

9. Annex – Description of current use cases

The following description is an extract of the CYCLONE use case portal. The reader should refer to this portal for the updated and detailed use cases description at https://cyclone.france-bioinformatique.fr/usecases.

9.1.1. UC1 - Securing human biomedical data

Thanks to the steady drop of genome sequencing technology costs (NGS), an increasing number of clinicians are including biological results obtained with these technologies in their day-to-day diagnosis practice. Today, much genomics analyses are realized on the exome, which is the expressed part (5%) of the genome. However, the full genome sequencing is being envisaged and will be soon included in daily medical practices.

In the near future, some of the genomic data processed on the IFB cloud platform will concern human biomedical data related to patients and thus, will be subject to strict privacy restrictions. To ensure 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 and ensure their integration (especially if the cloud federation involves both public and private cloud infrastructures).

9.1.2. UC2 - Cloud virtual pipeline for microbial genomes analysis

In the post-NGS area, sequencing bacterial genomes is very cheap (few hundreds €). Most of the time, users are no longer content to analyse a single genome; they want to compare large collections of related genomes (strains). This entails that biologists have to pay too much attention and dedicate their time to sequence the genomes, instead of thoroughly analysing the genomic data. Thus, this brings light to the increasing need for automating the annotation of bacterial genomes.

The team of the IFB-MIGALE platform (one of the bioinformatics platforms of the IFB) developed an environment for the annotation of microbial genomes and a tool for the visualization of the synteny (local conservation of the gene order along the genomes). The platform automatically launches a set of bioinformatics tools (e.g. BLAST and PLAST) to analyse the data and stores the results of the tools in a relational database (PostgreSQL). These tools use several public reference data collections. A web interface allows the user to consult the results and perform the manual annotation (manual annotation means adding manually metadata and biological knowledge to the genome sequence). Installing the platform requires solid skills in system administration since many bioinformatics tools with different dependences, a relational database management system, a web server and servlet container, etc. must be installed. Thus, performing the analysis of collections of genomes requires large computing resources that can be found with the distribution of the jobs over several computers, generally the computing nodes of a cluster.

CYCLONE Cloud federation will provide life science researchers with their own comprehensive annotation 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.

9.1.3. 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. Especially for small sequencing facilities, the analysis of the data is indeed the bottleneck.

These sequencers are located at a dozen places in France, while the users are distributed throughout the country and possibly further afield via international collaborations. Most of time the raw data are transferred from the sequencing centers to the researchers on hard disks sent by express shipping (e.g. FedEx, UPS) or carried by the researcher travelling by train or plane. The researchers may then not have sufficient on premise computing resources to analyse the raw data and will transfer them again to a remote computing center.

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 (from the researcher's point of view).

9.1.4. UC4 - Virtual Power Plant

The main idea of a Virtual Power Plant (VPP) is to integrate distributed energy resources (DER) for the generation of renewable energy and to combine them to one reliable power plant.

In the following description, the DER Owner, DER Operator and the VPP Operator have pivotal roles regarding the production of energy. These roles will be analysed in the specified workflow sections.

The Virtual Power Plant connects several small and medium sized power plants, DERs, 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, to provide reliable electricity in accordance with demand. 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, or is temporarily stored. Combining the different kind of DERs gives the benefit of increasing the usage of renewable energy.

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. The combination of the energy generation resources to a VPP enables to access a wider range of markets for selling energy. One single DER can only provide the small amount of energy it is generating. In combining the advantages of the various energy resources the Virtual Power Plants makes it possible to access energy markets usually only available for power plants which are generating a higher amount of energy and even to sell their flexibility to providers of system services.

9.1.5. UC5 - Internet of Services Lab (IoSL)

"Internet of Services Lab"1 (IoSL) is a SNET/TUB teaching project where students work a semester in groups of three to six. They most often implement software related to numerous research projects and topics: Cloud, Mobile, Social, and Ubiquitous Computing. This type of teaching project is quite popular in other computer science related study paths at TUB as well as other universities.

There are different cloud-related innovation areas where CYCLONE provides benefits within the IoSL:

1. Rapid provisioning of resources for student projects: Students require resources for conducting their projects, e.g., Virtual Machines. In the current set-up, it is a manual procedure to provision those resources.

By leveraging an infrastructure platform as well as deployment tools we minimize the required effort from the student supervisors considerably.

2. Utilization of SlipStream modules for reproducible application deployments: After students finish their course it is often problematic for other students and their supervisors to pick up their work. Most often, documentation is lacking and software versioning is not reliable, if available. By building upon SlipStream modules we create application deployments that can be easily reproduced, extended, and scaled. Also, students will learn how to structure their applications to leverage cloud characteristics, e.g., how to create immutable application deployments.

3. Federation Provider Integration: From our experience, every built demonstrator has its own user management. Furthermore, they often do not follow security best practices. By integrating the Federation Provider into each demonstrator, students learn about federated identity and are also liberated from implementing their own user management, as all students and supervisors will be able to login to the demonstrators via eduGAIN.

9.1.6. UC6 - Multi-cloud deployment and federated identity integration of ENtRANCe components

ENtRANCe (ExteNsible and geneRic AuthorizatioN for Cloud resources and personal files) has been chosen as one of the most innovative IT projects in 2014 to be funded within the Software Campus initiative. It aims at developing a user-centric and enhanced access rights management for a secure and practical privacy aware data exchange in distributed networks. The project focusses on the development of an efficient encryption system including the provision of management functionalities, such as efficient revocation.

The ENtRANCe project aims to define and deliver an Open Source platform for a secure and privacy aware exchange of data. Today, especially common Internet users utilize several services (eMail, Facebook, Flickr…) in order to share whatever data or information with friends, colleagues or family members. Normally this is accompanied by a loss of control over this “data”. ENtRANCe focuses on this disadvantage by using and state of the art technologies and standards in order to give control and management capabilities back into the hands of the data owner.

ENtRANCe is meant to support secure and privacy aware exchange of data. To ease the deployment of ENtRANCe components, we will create SlipStream modules to be used on the SlipStream App Store. This should enable rapid and multi-cloud deployment of those components by other SlipStream users and therefore advance their use considerably. Furthermore, the reliance of ENtRANCe on modern cloud deployment tools strengthens its innovation capacity and uptake by other users.

Especially, the Attribute Authority and Management Component which handles data encryption and management of access control rules empowers data owners. AA is an open source project, data owners can and are encouraged to deploy and operate the AA themselves. In an optimal setup, data owner has the full control over the AA, therefore it should be made easy for data owners to deploy and operate their own AAs with scripts for the most common scenarios.

Moreover, ENtRANCe needs a mechanism to identify new data receivers. Currently, data receivers cannot identify themselves using their eduGAIN accounts. Especially in an academic data sharing scenario, eduGAIN accounts are the most suitable option for identification. The integration of ENtRANCe with CYCLONE Federated Identity Provider will address this provisioning phase problem.

Finally, to manage access control to their resources, data owners maintain a list of [userId, dataId, access policy]. Access policies include constraints on certain attributes of data receivers. To define access policies based on eduGAIN attributes, JWT claims returned by CYCLONE Federated Id Provider will be utilized in ENtRANCe access control management.

9.1.7. UC7 - Open Scientific Data

Publicly funded research creates an immense amount of data that has academic and commercial value both inside and outside of the original research domain. The taxpayers through funding agencies are increasing the pressure to make these datasets generally available through "open data" programs. A major obstacle has been finding viable business models that keep the maintenance costs for the original funding agency reasonable by monetizing the datasets for commercial use. SixSq is exploring solutions with ESA in which public data is hosted on European cloud infrastructures.

To make such a solution viable on the technical side would require: 1) detailed knowledge of the storage locations of dataset components, 2) means of placing analysis applications near the data of interest, and 3) ranking of multiple providers based on price or other characteristics. All potential features of the CYCLONE brokering and matchmaking components. There may also be a need for remote access to the public datasets to, for example, perform a combined analysis with private datasets that could not be moved to the cloud. In this case, the CYCLONE networking features may be useful.

9.1.8. UC8 - Benchmark Driven Placement

A significant part of the recent design discussion for the CYCLONE brokering and matchmaking components dealt with benchmarks, both general benchmarks and application-specific benchmarks. At the recent UCC 2015 conference [http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7430473], there was a presentation from a bioinformatics group in Cardiff that used their own benchmarks for common bioinformatics tools on a number of community and public clouds to drive placement of the virtual machines [http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7431434&searchWithin%3Dbenchmark%26filter%3DAND%28p_IS_Number%3A7431374%29]. The benchmarks are collected continuously from applications being run as well as directed probes of particular infrastructures.

Their work would be an interesting direct validation of the CYCLONE brokering and matchmaking design. CYCLONE would benefit by reusing their benchmarks for this validation. They would possibly benefit by replacing their ad hoc deployment engine with SlipStream, freeing them from the maintenance of their engine and allowing them to expand their benchmarks from single virtual machines to full application pipelines.

SixSq is currently responding to a tender that would involve continuous monitoring for applications in the domains of high-energy physics, earth observation, and bioinformatics. If the offer is accepted, this would be an additional scenario for this use case.

9.1.9. UC9 - On-Demand Bandwidth

At the NetCloud 2015 workshop as part of the UCC 2015 conference [http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7430473], there was a presentation of a "zero touch" configuration project to allow automated, on-demand network bandwidth allocation across administrative domains [http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7431457&refinements%3D4225654905%26filter%3DAND%28p_IS_Number%3A7431374%29]. There are a number of National Research and Education Networks (NRENs) around Europe involved in this activity, one of which is SWITCH (Swiss NREN) that has connected CloudSigma and Exoscale [https://exoscale.ch] to Géant. SixSq has a good relationship with Exoscale and it may be possible to get them involved in any on-demand bandwidth activities/demonstrations.

9.1.10. UC10 - Smart Utility 4.0

On first year the energy use cases focused on the energy generation by the distributed renewable energy resources. The aggregation in a Virtual Power Plant in order to contribute to efficient energy management was considered also during the first year of the project. Given that the future energy management concerns the sectors of energy production, energy consumption and the transportation, it is essential to extend the energy use cases in the field of consumption.

New innovative services that offer an easier and more useful way of managing energy consumption are required. Because of the volatile availability of renewable energy, the focus is on a demand side management and a balanced demand response system. Demand side management gives the ability to consumers to make better informed decisions about their energy consumption, adjusting the timing and the quantity of their consumption. Demand response provides the consumers' energy management with incentives to motivate them to schedule their consumption in such a way that the operation and cost of the power system is improved. The consolidation of consumption schedules for industrial consumers and the intelligent connection with the VPPs lead to an optimal, sustainable usage of renewable energy resources. The connection with the VPPs is achieved via cloud-based services.

Moreover, the extension of the energy use case will consider developing an energy management system for balancing energy production and consumption considering technical and legal aspects and always ensuring the security and privacy of data. Both the collected data from the distributed renewable energy resources aggregated by the VPP and the data from consumers are stored in the platform. Services have to be deployed to balance the generation and the consumption in order to ensure the stability of the grid.

9.1.11. UC11 - Assembling genomes from sequencing reads

To decipher the genome of new organisms, biologists split it into a very large number of redundant fragments that are subsequently sequenced (shotgun sequencing). For large eukaryotic genomes, this generates billions of reads (partial sequence of the fragments). The genome is reconstructed by piecing up overlapping reads. Algorithmically, this is done by searching for an Eulerian path in a de Bruijn graph built from the reads. For large eukaryotic genomes, this graph is particularly large and requires computers with a sizeable amount of memory (sometimes more than 1TB of RAM).

The PRABI AMSB bioinformatics platform is currently implementing a technical partnership with another platform of the University Lyon 1 (DTAMB), which is equipped with next-generation sequencers. The goal is to provide an automated service proposing primary bioinformatics analyses for academic clients of the sequencing facility. Bacterial genome assembly is a key application in this context, representing more than half of the demand. This project would highly benefit from the possibility of instantiating virtual machines with large memory (256-512GB), especially in the perspective of extending the service to eukaryotic genome and transcriptome assembly.

9.1.12. UC12 - Metagenomics

Metagenomics is the study of the genomes present in communities of organisms found in environmental samples. Metagenomics is a generic term describing a range of studies that use high-throughput DNA sequencing (reads) to characterize microbial systems, including whole-genome shotgun (WGS) sequenced metagenomic and metatranscriptomic studies, as well as amplicon-based approaches that target specific marker genes (often rRNA genes). The analysis pipeline of WGS metagenomics consists of 4 broad steps: i) quality control of the reads, ii) sequence reconstruction and grouping, iii) gene and regulatory elements predictions, iv) function prediction for the protein-coding genes. The amplicon metagenomics pipeline comprises mostly 3 steps: i) quality control of the reads, ii) OTU clustering and picking followed by their taxonomic assignment, iii) statistical analysis. Both approaches, WGS and amplicon, involve the comparison of reads (or assembled reads) with databases of protein sequences for the former and rRNA genes for the latter. Metagenomics studies generate large amount of reads and these comparisons can be extremely computer-intensive since they involve matching this massive number of reads with large protein sequence or rRNA gene databases. Hence the need to distribute these computations over several Cloud facilities.

9.1.13. UC13 - Shared environment between cloud Galaxy portals

Galaxy is an open and web-based platform for bioinformatics research that gathers multiple bioinformatics tools and allows chaining their execution in a workflow. Users of the Galaxy environment need to share histories, workflows and data although they run different instances of the same VM (most bioinformatics programs have been designed to work with 'regular' file systems). Once shared with other users or groups, they can use the workflow or the history as is or edit it and thus import it in its own environment. A Galaxy server is generally dimensioned to be able to handle large computation, and thus pre-empting a large amount of resources in the form of a computing cluster as the backend. The goals of having Galaxy in the CYCLONE cloud federation is to allocate on-demand adapted amount of resources through appliance, while keeping the user appreciated feature that is workflow sharing and re-using inside the Galaxy environment.

Within this use case we will also explore using the ENtRANCe attribute-based encryption (ABE) components to manage the sharing of Galaxy Users' data. When using ENtRANCe, all shared data can be saved within a public storage backend which would ease the management of this data considerably while keeping the confidentiality through ABE. ABE would also allow fine-grained access rules, based on lab affiliation or designated work area, for example. We'll also issue keys based on the federated identity to allow traceability for such processes.

9.1.14. UC14 - WebRTC video-conference solution

WebRTC is an open-source project enabling plugin-free, Real Time Communications (RTC) in the browser. It includes the fundamental building blocks for high-quality communications such as network, audio, and video components used in voice and video chat applications.

In WebRTC the signaling plane is on a separate channel from the media plane (where the actual data streams pass through). In most cases (roughly 80%) the endpoints communicate directly in true peer-to-peer fashion. A STUN server may be necessary for endpoints that are behind a NAT or firewall. The purpose of the STUN server is to report the public IP address of an endpoint so that the other endpoint may be able to address it directly. Additionally, in the event that a direct connection could not be established between the endpoints, a TURN server can be used to relay the media packets. This incurs more overhead but allows most peers to communicate despite having restrictive network configurations.

I2CAT is currently developing a WebRTC Solution is based on EasyRTC (https://github.com/priologic/easyrtc), a backend signaling service. It can be integrated with 3rd party authentication and authorization systems or provide standalone user and password based credentials system. The WebRTC components comprise:

- The clients
- The signaling server.
- Multipoint control unit (MCU).
- The STUN server(s) (optional).
- The TURN server(s) (optional).

The scope of the use case in the context of CYCLONE federated environment is to provide 1: N videoconferencing for RAS and IBUS as well as to provide support to the features of the application by deploying it in cloud federated environments.

Thus the idea behind, consists of the possibility to put the WebRTC in the cloud federation and provide "WebRTC as a service": Enterprises wishing to deploy WebRTC applications have two options: either deploy their own WebRTC servers or outsource these servers to someone else. By offering WebRTC as a service, a service provider would basically host the WebRTC gateway for the enterprises. WebRTC calls destined to the enterprise would be handled by the WebRTC gateway of the service provider. Incoming WebRTC calls would be translated into SIP calls and routed to the enterprise. The enterprise would not have to change anything in its infrastructure, as it will still be only handling SIP calls. In order to show the potential benefits that utilizing the CYCLONE solution may comport for this use case, it is next proposed the following practical example situation: The "OnlyPayIfUseIt" is an application provider company specialized on the provisioning of a variety of general communication services to their customers. One of these services, consist of a WebRTC based communication system. OnlyPayIfUseIt, does not own infrastructure to deploy the services and rents infrastructure to one or several IaaS. In order to offer competitive fees to their clients, there is a service category that enables to deploy the service per daily basis, so that the service can be configured to be only deployed (and thus available) during the working hours. The Web-RTC is automatically deployed and removed according to the user configuration (the informatics department of the client company), thus, saving the use of resources during the "out of office" periods. This modality enables to the client to have available the Web-RTC service during the working hours and to the OnlyPayIfUseIt service provider, to save IaaS resources. The OnlyPayIfUseIt leverages on the CYCLONE solution to enable the automatic deployment of the components and modules of the Web-RTC system in a distributed, redounded and reliable way. It also may take advantage of the security and networking features that the CYCLONE software enables.

The concrete use case requirements are:

Signaling and Media: Signaling is related to of the bandwidth and CPU usage. It is necessary to ensure that both work well while deploying the application. The possibility to split them between different providers in a cloud federated scenario is a good alternative rather than having all in the same CSP.

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. 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 signaling servers would also aid system reliability.

Dedicated servers: Currently, virtualization technologies applied to CPS PaaS leads to the discussion on whether it´s needed dedicated resources, or if virtualized ones are already enough.

Multi-point control unit: Cloud federated deployments enable the possibility of having the multipoint control unit be hosted separated locations. The performance and availability of an MCU is crucial to the perceived quality of an n-to-n videoconferencing call.

Security: For maximum security service providers are expected to secure the signaling 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.

Ports open in the Firewall: Web ports (443 or 50001 TCP) and RTP ports (9256-9500 UDP).

Automation requirements: It should be enabled some automation procedures to enable the configuration of the periods of time when the service should be deployed and when it should be removed.

9.1.15. UC15 - Interactive Authorization Policy Management for Multi-cloud Applications

Multi-cloud applications span multiple cloud providers (CP) from which different components of an application are made available. The developer of the application can devise a deployment plan such that the components of her application are installed on resources that best fit her requirements. While the proposed technique in the use-case can be used for a wide variety of requirements, we focus on the ones pertinent to security and privacy that are represented in the form of policies. In this use-case, a cloud provider imposes restrictions on the use of its computational resources to certain user groups based on their attributes. Similarly, cloud consumers strictly regulate access to their data in the object repository and applications they offer in the market place. We call a consumer as data owner (D) if her main asset in the cloud is her data. A data owner may impose restrictions on access to her data, its replication or geographical location of its storage. Other cloud consumers called clients (C) in the same application ecosystem use data owner's data and produce (commercial/scientific) value from it with their business logic (e.g. application). Both types of cloud consumers specify their security requirements in a security policy. We assume that some of the rules in these policies are hard rules and need to be satisfied while some others can be relaxed.

9.1.16. UC16 - Attribute-based Authorizations with XACML

Attribute-based access control allows for flexible management of permissions by enabling the use of user/resource attributes when specifying security policies.

It can be used to capture the existing authorization models such as role-based access control (RBAC) where core concepts are simply treated as attributes (e.g. role as a user attribute). In this use-case, we will demonstrate the use of XACML policies for managing permissions of CYCLONE users over available resources. The attributes provided by EDUGAIN will be used to generate the policies. For instance, the following attributes: name, surname, organization type, affiliation, e-mail and username can be used to specify authorization policies. Certain types of policies can be automatically generated (i.e. from ad hoc group-based policies to XACML) from existing authorization management methods/systems of use-case owners.

In the same use-case, there will also be the demonstration of multi-tenant access control system presented in [6] where two imaginary cloud providers will consume the authorization services developed by UvA to manage the permissions over their resources.

<END OF DOCUMENT>