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 D3.1

Evaluation of Use Cases

Due date:	31/10/2015
Submission date:	01/12/2015
Deliverable 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)

List of Contributors

Participant	Short Name	Contributor
Interoute S.P.A.	IRT	D. Gallico, M. Biancani
SixSq SARL	SIXSQ	C. Loomis, R. Branchat
QSC AG	QSC	D.Hacker, M.Kourkouli
Technische Universität Berlin	TUB	M. Slawik
Fundacio Privada I2CAT, Internet I Innovacio Digital A Catalunya	I2CAT	J. Aznar, E. Escalona
Universiteit Van Amsterdam	UVA	-
Centre National De La Recherche Scientifique	CNRS	C. Blanchet, J.-F. Gibrat

Change history

Version	Date	Partners	Description/Comments
0.1	21/09/2015	CNRS	Table of Contents
1.0	20/10/2015	CNRS	First version
2.0	28/10/2015	I2CAT, IRT, SixSq, TUB	Revised version with the integration of comments and contributions from partners
3.0	16/11/2015	CNRS, QSC, TUB, I2CAT	Revised version of the use cases description and new use cases identification.
4.0	26/11/2015	CNRS, I2CAT, IRT, QSC, SixSq, TUB,	Revision of the Use cases requirements and implications for the CYCLONE software, Contribution to the Preliminary market analysis and business model investigations. Ready for review
4.1	27/11/2015	SixSq	General review
4.2	30/11/2015	CNRS	Last fixes
vFINAL	01/12/2015	IRT	Final review and submission

Table of Contents

List of Figures

List of Tables

Executive Summary

The goal of this document is to provide a concise summary of selected use cases, explain how they have been deployed within CYCLONE, and provide formal feedback on the CYCLONE software. The document focuses on the analysis of use cases presented in the DOA, derives functional requirements to be considered by the CYCLONE federated framework, and analyses their deployment and testing on the CYCLONE testbed.

The application developers detailed the use cases in Chapter 2 with a list of user stories and requirements. In Chapter 3, a set of common requirements was derived from the use case descriptions, from which a list of implementation tasks was derived. Chapter 4 describes the evaluation of the use cases with the deployment of the use case, a list of test scenarios, and the current results. Chapter 5 describes a list of potential new use cases to extend current ones, brings in use cases from other areas, and shows how users can integrate the CYCLONE solution. Finally, a preliminary market analysis and business model is investigated in Chapter 6.

Two bioinformatics use cases were already deployed in M10. This deployment was done incrementally, beginning with a one-VM use case integrating security features (UC1 Securing human biomedical data). Afterwards, we then deployed a complex application requiring the coordinated deployment of several VMs (UC2 Cloud virtual pipeline for microbial genomes analysis). The UC1 benefits from enhanced security features and the appliance was extended with the integration of the Federation Provider. The UC2 Cloud virtual pipeline for microbial genomes analysis was extended with SlipStream deployment recipes to solve the requirement of a complex application deployment with several virtual machines in one deployment.

1. Introduction

The project identified two flagship applications: an academic cloud platform (and its associated services) supporting bioinformatics research and a commercial platform for future energy management. Each provides 3 “micro” use cases (6 in total), highlighting different needs.

Taking full advantage of cloud technologies in general and the CYCLONE services in particular will require changes to the architecture and implementation of CYCLONE components, specifically to make them more scalable, dynamic, and secure.

The partners of CYCLONE have worked with the owners of the selected use cases to analyse their needs in relation to the features provided (or to be provided) by CYCLONE and help them adapt their applications to maximize their benefits from using CYCLONE.

The goal was to provide a set of requirements retrieved from the use cases, and work together with CYCLONE components developers to develop the CYCLONE framework to adapt it to the use cases and be able to satisfy and deploy them in a dynamic, multi-cloud environment. According to the requirements and the solutions provided by the software components as well as their availability in the testbed infrastructure, the partners proposed a plan of implementation with an associated timeline to validate all of the critical CYCLONE features for the application, exercising the scalability, robustness, and completeness of those features.

The CYLONE partners have also described a list of potential new use cases to extend the scope, bring use cases of other areas, and show how users can take the CYCLONE solution. A preliminary market analysis and business model is also investigated.

2. Use case descriptions

2.1. Bioinformatics use cases

2.1.1. Overall description

Bioinformatics deals with the collection and efficient analysis of biological data, particularly genomic information from DNA sequencers. The capability of modern sequencers to produce terabytes of information coupled with low pricing (less than US$1000 for a human genome) that makes parallel use of many sequencers feasible causes a "data deluge" that is being experienced by researchers in this field [1].

Bioinformatics software is characterized by a high degree of fragmentation: literally hundreds of different software packages are regularly used for scientific analyses with an incompatible variety of dependencies and a broad range of resource requirements. For this reason, the bioinformatics community has strongly embraced cloud computing with its ability to provide customized execution environments and dynamic resource allocation.

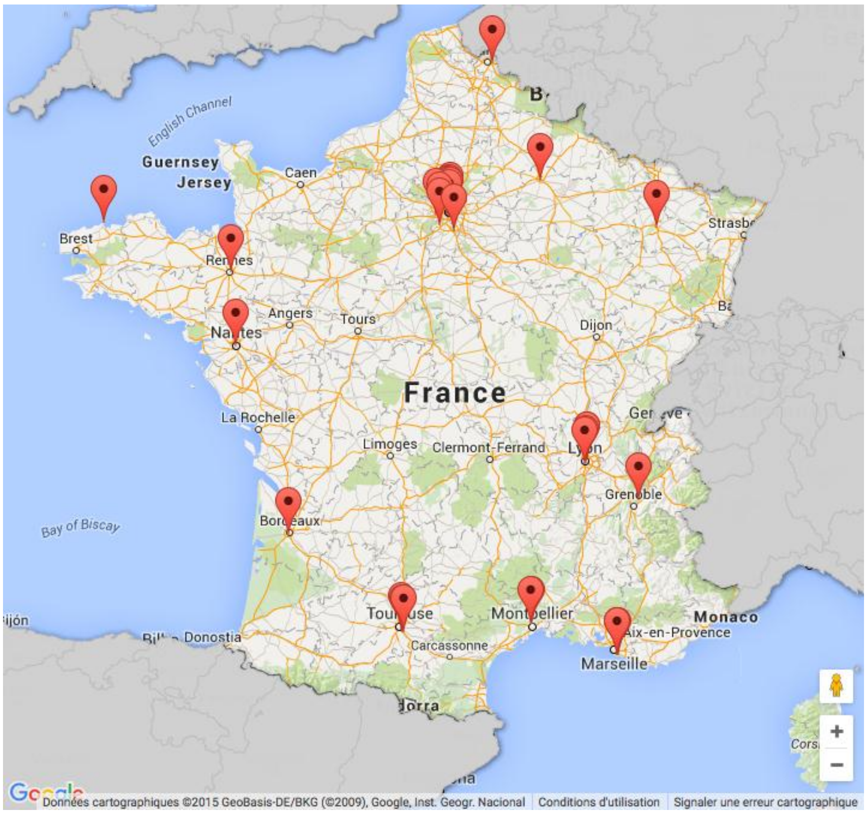

Figure 2-1: The distributed infrastructure of the IFB (32 sites)

The French Institute of Bioinformatics - IFB [2] consists of 32 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. The IFB has deployed a cloud infrastructure on its own premises at IFB-core, and aims to deploy a federated cloud infrastructure over the regional PFs. This cloud infrastructure is devoted to the French life science community, research and industry, with services for the management and analysis of life science data.

Naturally, the infrastructure users, for example biomedical staff analysing human biomedical data or researchers analysing genome sequences, will benefit from these features while accessing the system and carrying out their analyses.

Thus, the CYCLONE consortium has identified three concrete bioinformatics use cases that aim to address some specific well-identified limitations while carrying out the analyses starting from general cloud federation requirements and specific use cases features to enhance current cloud infrastructure processes and services provisioning mechanisms.

2.1.2. UC1 - Securing human biomedical data

2.1.2.1. Description

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).

Let us emphasize that tests within the CYCLONE project will be carried out **only on non-personally identifiable data** (either anonymized benchmark data or simulated data) to avoid exposing any personal data during development and testing. These data will be extracted from existing public reference datasets such as the European Nucleotide Archive [3] or the Ensembl Genomes resource [4].

2.1.2.2. Workflow

The use case workflow to ensure data security is as follows:

As in Figure 2-2, 1) a biomedical user connects to the cloud through the IFB‘s web authenticated dashboard, uses it to 2) run an instance of the appliance containing the relevant pre-configured analysis pipeline. Then 3) the VM is deployed on the testbed. Then 4) the user signs into the web interface of the VM, 5) uploads the patient’s biomedical data to the VM, and 6) runs the analysis in a secure environment. Finally, 7) the user gets the results.

The bioinformatics treatment generally relies on a comparison with the current release of the reference human genome hg19 (Human Genome version 19 or GRCh37 [5]). The hg19 is a database consisting of many files containing the public genomics data. It needs to be previously deployed by the cloud providers as a public data set available to all users.

2.1.2.3. Users’ stories

Actors

- Biomedical staff analyse the human biomedical data.
- A cloud provider supplies the computing and storage resources, with the required features permitting the automatic and elastic deployment of an application.

- A cloud provider supplies the required level of security for the VMs environment, in terms of authentication, authorizations management and end-to-end security.

Stories

- As a biomedical staff member, I can have a personal cloud account, based on my identity in the academic federation (the national one provided by RENATER or international by eduGAIN)
- As a biomedical staff member, I have access to the authenticated web dashboard to manage my secured cloud resources
- As a biomedical staff member, I have access to the required appliances, published in the marketplace, to help me to treat the human biomedical data
- As a biomedical staff member, I can run a VM to analyse the biomedical data
- As a cloud provider, I can guarantee that the VM will be deployed in a user-devoted network, isolated from other users
- As a cloud provider, I can guarantee that the connection from the user desktop to the VM is secured
- As a cloud provider, I can guarantee that the VM data (disk image and data storage) will be erased at the end of the execution

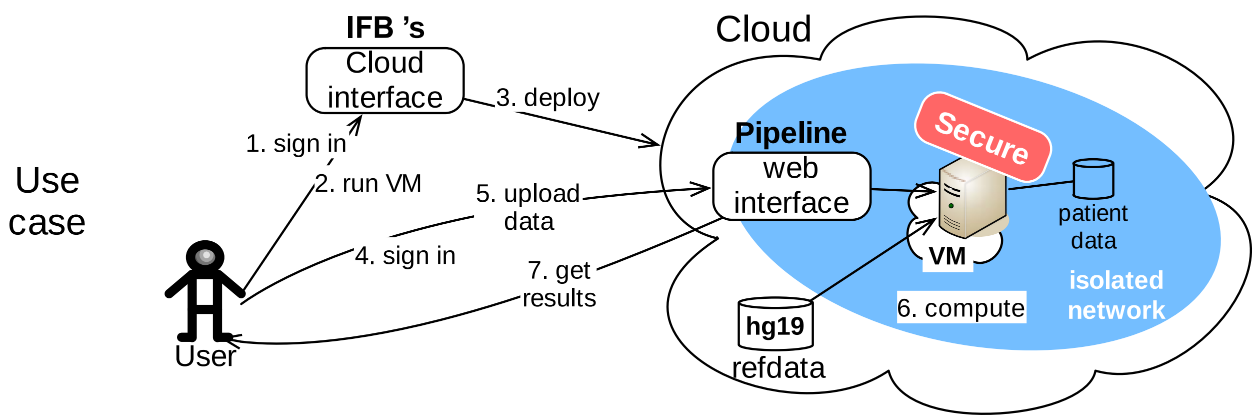

Figure 2-2: Functional schema of the use case "Securing human biomedical data". *The figure shows the application components and describes the different steps of the workflow.*

Evaluation

This use case will evaluate the capability of CYCLONE infrastructure to provide biomedical staff as cloud users with the deployment of their own cloud infrastructure for the biomedical analyses consisting of a single virtual machine with a web interface in a secured environment. The access to this environment will be based on the identity and authorizations in the federation.

2.1.3. UC2 - Cloud virtual pipeline for microbial genomes analysis

2.1.3.1. Description

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 [6], PSI-BLAST [7], INTERPROScan [8]) 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.

2.1.3.2. Users' stories

Actors

- A bioinformatician analyses the experimental data and visualizes the annotated data.
- A cloud provider supplies the computing and storage resources, with the required features permitting the automatic and elastic deployment of an application.

Stories

- As a bioinformatician, I can have a personal cloud account, based on my identity in the academic federation (the national one provided by RENATER or international by eduGAIN)
- As a bioinformatician I have access to the authenticated web dashboard to manage my cloud resources
- As a bioinformatician, I have access to the required appliances, published in the marketplace, to help me to annotate my bacterial genomics data
- As a bioinformatician, I can deploy and synchronize the required public reference data collections in the targeted clouds
- As a bioinformatician, I can deploy a distributed cluster of VMs to annotate my unknown microbial genomes
- As a bioinformatician, I can run a VM to visualize my genomic data
- As a cloud provider, I can guarantee that the VM will be deployed in a user-devoted network, isolated from other users.
- As a cloud provider, I can provide users with the complete deployment of a distributed cluster of VMs sharing a common workspace.
- As a cloud provider, I can provide users with the deployment of a VM with a visualization environment, connected to the shared storage.

2.1.3.3. Workflow

As in Figure 2-3, a bioinformatician 1) connects to the cloud web dashboard, uses it to 2) run and 3) deploy with one click a genomes annotation platform consisting of many VMs, comprising of a master node of the virtual cluster that provides also the visualization interface (Web), associated with several computing nodes. Then the user 4) uses SSH to connect to the master and 5) uploads the raw microbial genomic data (MB) to the cloud storage. SCP/SFTP protocols are used from a command line tool or a GUI, to ensure AuthN/Z for the data transfer, and to overcome the performance issues of HTTP for large datasets. Still in command line interface, the user 6) runs the computation to annotate the new microbial genomes. The first step consists of many data-intensive jobs performing the comparisons between the new genome and the reference data

The results are stored in a relational database (provided by a cloud service or a VM deployed within the platform). Then the scientist 7) signs in the annotated data visualization environment provided by the Insygth Web interface to 8) navigate between the abundant homologues, syntenies and gene functional annotations in bacteria genomes.

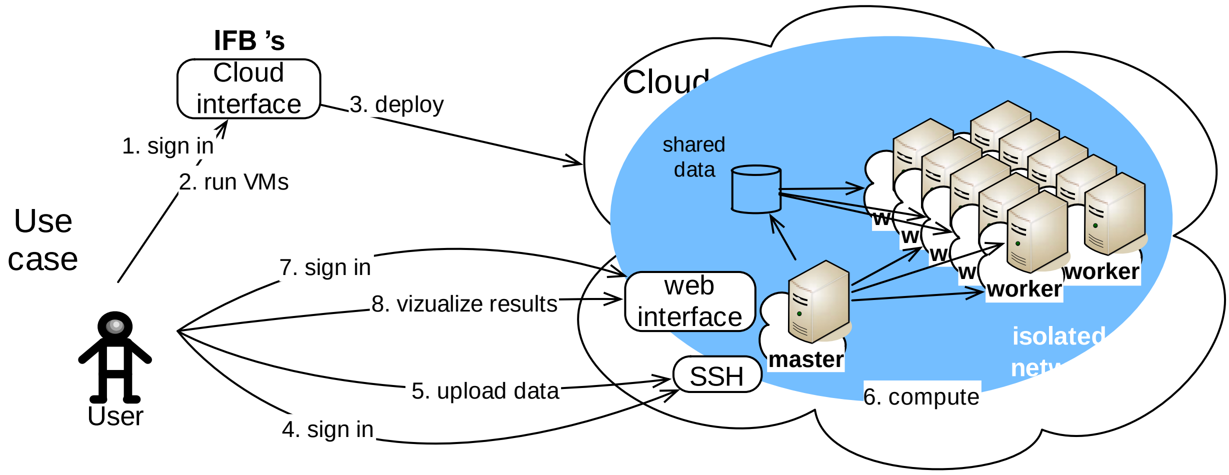

Figure 2-3: Functional schema of the use case "Cloud virtual pipeline for microbial genomes analysis".
The figure shows the application components and describes the different steps of the workflow.

Evaluation

This use case will evaluate the capability of CYCLONE infrastructure to provide a bioinformatician with a one-click deployment of a complex application. This deployment needs to be done in an isolated network for security and confidentiality reasons. The access to this environment will be based on the identity and authorizations in the federation. The application may require to be deployed over several clouds because it could require large computing resources not available in one place or - for functional reasons - some data or tools are only available in certain environments.

2.1.4. UC3 - Live remote cloud processing of sequencing data

2.1.4.1. Description

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).

2.1.4.2. Users' stories

Actors

- A sequencing engineer produces the experimental data in sequencing centers.
- A bioinformatician analyses the experimental data.
- A cloud provider supplies the computing and storage resources, with the required features permitting the automatic and elastic deployment of an application.

Stories

- As a sequencing engineer/bioinformatician, I can have a personal cloud account, based on my identity in the academic federation (the national one provided by RENATER or international by eduGAIN).
- As a sequencing engineer/bioinformatician, I have access to the authenticated web dashboard to deploy an instance of such VM.
- As a sequencing engineer/bioinformatician, I have access to the required appliances, published in the marketplace, to help me to analyse and visualize my genomics data.
- As a sequencing engineer, I can upload the genomic raw data (reads) to the cloud storage to be archived and stored in a shared workspace.
- As a sequencing engineer, I can run a distributed cluster of VMs to do the first analysis of the raw data (mapping).
- As a sequencing engineer, I can manage the authorizations rules to provide a bioinformatician with the data access rights.
- As a bioinformatician, I can run a VM containing the useful tools (e.g. IGV [9]) to visualize the genomics data.
- As a cloud provider, I can provide a sequencing engineer/bioinformatician with the complete deployment of a distributed cluster of VMs sharing a common workspace.
- As a cloud provider, I can guarantee that the VMs will be deployed in a user-devoted network, isolated from other users.
- As a cloud provider, I can provide users with the deployment of a VM with a visualization environment, connected to the shared storage.

2.1.4.3. Workflow

As in Figure 2-4, a staff member of the sequencing center 1) connects to the IFB's cloud interface, uses it to 2) run and 3) deploy in one click a Next Generation Sequencing data analysis platform consisting of many VMs, comprising a master node associated with several computing nodes, and a virtual machine providing the visualization environment based on a Web interface (a) or on a software tool with a GUI (b).

The sequencing engineer 4) signs in the master VM and 5) uploads the terabytes of raw data (it requires a high-bandwidth connection between the sequencing center and the cloud infrastructure). The protocols scp/sftp is used to ensure AuthN/Z for the data transfer, but can induce performance issues. The raw data consist of hundreds of thousands of short nucleotide sequences (150-200 long) called the 'reads' that the sequencing engineer 6) maps on a reference genome of the same organism than the sequenced one (the primary analysis). Once the analysis is done, the sequencing engineer 7) provides the bioinformatician with the access rights to the data and the visualization platform.

Then the bioinformatician 7) connects to the visualization environment to analyse the results, for example determining the quality of the initial sequencing step. This can be done with a web-based visualization platform like Galaxy (a) or with a GUI-based tool like IGV (b). In both cases, the bioinformatician 8) connects with its federation credentials (it requires either an HTTP identity check or a SSH one), and used the environment to 9) visualize the results associated to the sequenced genome (with either a web display or a remote desktop based on X11/NX protocols).

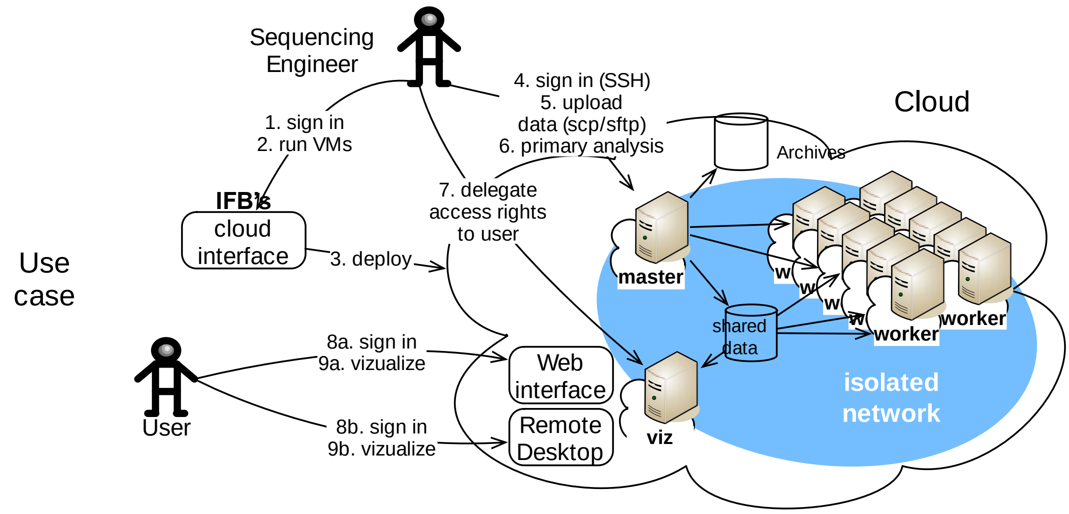

Figure 2-4: Functional schema of the use case "Live remote cloud processing of sequencing data". *The figure shows the application components and describes the different steps of the workflow.*

Evaluation

This use case will evaluate the capability of CYCLONE infrastructure to provide several actors, a sequencing engineer and bioinformatician, a one-click deployment of a complex application. This deployment needs to be done in an isolated network for security and confidentiality reasons. The access to this environment will be based on the identities and authorizations in the federation of both users connecting from different places: the sequencing center and the bioinformatician laboratory. The sequencing engineer will be the user uploading the data to the cloud and deploying the analysis infrastructure and should be able to permit access to the bioinformatician, both to the data and the VMs.

2.1.5. Overall relation of the bioinformatics use case to CYCLONE ambition and main objectives

The bioinformatics use case is strongly related to the ambition of the CYCLONE project, and most of its objectives will help to answer the bioinformatics needs. Regarding the key technical areas of the CYCLONE project, ***Cloud Access Management through cloud proxies*** and ***Matchmaking, Brokering, and Mediation of Cloud Resources*** will provide the IFB with features to access other cloud infrastructures than the current national IFB's one, to integrate the future cloud infrastructures that will be deployed by the regional IFB's platform, and also to access external cloud infrastructures from academic or commercial providers. These developments will also provide services for the multi-cloud deployment of complex applications as in UC2 and UC3. **The developments related to the *End-to-end Security for HTTP-based Applications* will bring a**

solution to the secured web access that the bioinformatics use cases require. The *Dynamic network configuration and management will additionally help satisfying the isolation of the virtual machine of a user from unauthorized access.*

The bioinformatics use cases match the following objectives specifically:
- Integration of a common federated identity scheme (e.g. EduGAIN) to ease the federation of cloud infrastructures
- Complete integration of advanced features of OpenNaaS
- Providing a policy-based cloud brokering engine
- Providing isolated network service delivery
- Network resource allocation
- Multi-cloud extension of location-based access control and distributed logging

In terms of impact, the CYCLONE developments will improve the quality of experience through a more powerful interface (complex application deployment) to federated heterogeneous cloud environments (both IFB's and external ones). They will increase the innovation for bioinformatics service providers by allowing them to integrate more features into their Cloud applications. For example, the incorporation of common federated identity schemes (e.g. EduGAIN) will facilitate the federation of cloud infrastructures that IFB is planning to deploy in the next years, and simplify the work of bioinformatics application developers thanks to a common identity management regardless of the targeted cloud infrastructure.

These three use cases will lead to demonstrators that will be made publicly available to the French and European community (through the ELIXIR collaboration). The list of requirements and solutions that have been identified will help us to define clearly the related future applications to deploy in the IFB's cloud

2.2. Energy use case

2.2.1. Overall description

In order to comply with the "202020" climate change mitigation goals of the European Union, as well as the goals of the German energy transition, the "Energiewende", the energy generation has to be moved from fossil, conventional power plants to a sustainable, renewable energy generation.

Focusing on the climatic change, the EU leaders defined ambitious goals to reduce CO2 emission and change to a more sustainable usage of energy.

The "202020" goals are summarized in the 2020 Climate & package[1] to ensure the EU meets the climate and energy targets for the year 2020.

The package sets three key targets:

- 20 % cut in **greenhouse gas** emissions (from 1990 levels)
- 20 % of EU energy from **renewables**
- 20% improvement in **energy efficiency**

To achieve these goals, more and more renewable energy sources have to be included in the overall energy system. This leads to the necessity to restructure the energy supply system. It has to be transformed from a centralized system with large power plants to an increasingly decentralized system of power generation. In

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

the actual situation, conventional power plants are adjustable within a more or less fixed range, so that the energy management is done in a centralised top down approach.

The integration of CO2 free, volatile, and decentralized, renewable energy resources into the energy system changes the characteristics and the overall need for the energy management system. Coming from a centralized system where energy is produced when its needed, it changes to a system where energy is produced when sun or wind are available. According to these constraints, a secure energy supply is only possible through the combination of energy resources. The huge amount of decentralized energy components included into the grid forces the grid to become smarter. Integrating the latest Information and Communication Technology (ICT) for managing the energy components in the grid is becoming more and more essential. ICT from distributed embedded control to Big Data and Cloud Computing is essential for the transition.

The decentralized energy production leads to the necessity to collect measurement data of energy production and consumption in real time all over the grid. Concerning the different energy components coming with their own smart metering technologies, the data management has to deal with heterogeneous data.

The aim is to use every renewable produced kWh.

In Germany in 2013 there were ~1.4 Mio installed photovoltaic plants and ~25.000 Windplants. The location of distributed energy resources that are combined to a virtual power plant can be widely distributed, even over different countries, as it is depicted in the following figures.

Figure 2-5 **Distributed Energy Resources in Germany and in Belgium**

2.2.2. Energy Management System

The future energy management concerns to the sectors of energy production, energy consumption and the transportation.

Regarding the critical role of a secure energy supply system, the derived requirements for the Energy management system based on ICT are challenging.

2.2.2.1. Energy Production – Virtual Power plant

Regarding the change of the energy production from conventional power plants to renewable energy production, the energy production goes from a small number of centralized, big power plants to a huge number of distributed energy resources (DER) located in a wide geographical area.

To cope with the new situation of sustainable energy production, the energy management has to react to:

- Decentralized energy production.
  The energy is generated in several small units/sources which are geographically distributed
- Volatile energy production.
  The regenerative energy generation cannot be scheduled, not in relation to place, time, or amount. External circumstances, such as the weather can lead to irregularities in the energy production.

We mainly have three different types of Distributed Energy Resources (DER) included in the energy system:

- Windmills
- Photovoltaic
- Bio gas plants

Every type of DER included in the system has its special characteristics in power generation that has to be taken into account for the integration into the energy system.

2.2.2.2. Windmill

Windmills are mostly combined to windparks, containing several windmills located outside of urban areas. The energy generation relies on the actual amount of wind and the wind speed. The energy production is not steerable and cannot be guaranteed. Based on different weather information, there can be a prediction of generated energy for the next day. The windparks can be switched off if there is too much energy in the overall energy system.

2.2.2.3. Photovoltaic

Photovoltaics convert solar energy into electrical energy. They generate electrical energy in accordance to how much sun is shining and when sun is shining. Most photovoltaics are related to households and come up with a small amount of generated energy, but there is a huge number of single installations. The generated energy is not really predictable, even on weather forecasts. The energy production is not steerable and cannot be guaranteed. The photovoltaics can be switched off if there is too much energy in the overall energy system.

2.2.2.4. Bio gas plant

The bio gas plant is running on biological processes. The Bio gas plants are located both in urban and rural environments. The energy generation is steerable and can be switched on and off.

2.2.3. Involved Roles

2.2.3.1. DER Owner

The DER Owner, as the name declares, is the owner of the resource. The DER Owner connects to a service on the ICT-Platform to get the information about his DER. He is able to have an overview of his components (meters, etc.). The DER Owner has concerns about the privacy of the data. Thus, the DER Owner needs to trust the cloud provider, he requires secure transmission and storage of the data in the cloud. As the owner of raw data he specifies who will have access to the authenticated data. This role is responsible to provide access to the DER Operator and to the Virtual Power Plant (VPP) Operator and prevents unauthorized access to his data providing access to different applications. The data owner authorizes specific services to access his data. Furthermore, the DER owner needs to know that his data is not accessible by other data owners sharing the storage facilities of the same cloud. Moreover, using the services the DER Owner has access to the aggregated data from energy generation of the DER that tailored his needs. For every DER owner it is possible to get the aggregated information about his resource, for example the actual generated amount of energy, as well as the hourly or monthly statistics and analysis, the amount of energy brought to customers via the VPP, the working status of the VPP that he has a contract. This role can connect to the platform after authentication, for example over a web interface or with a dedicated application on a mobile device and views the revenues.

The described requirements concerning the DER Owner are summarized in the following table.

	Requirements
DER Owner	Authenticated connection to the platform (web interface or mobile device)
	Overview of his components
	Access to Aggregated Data and to data tailored his needs
	Secure transmission and storage
	Privacy for his data
	Prevent unauthorized access to his data
	He authorizes specific services to access his data
	Data must not be accessible by other data owners sharing the storage facilities of the cloud
	View statistics, revenues, status of VPP

Table 2-1 Requirements for DER Owner

2.2.3.2. DER Operator

The DER Operator, can also be the DER Owner, connects to the ICT-Platform and through the available services he gets information about the DERs that he manages. The DER Operator requires to have access to the information concerning the operation of DER as well as the historical data, the hourly or monthly statistics and analysis, the amount of energy brought to customers via the VPP, and the working status of the VPP that he has a contract. This role keeps track the status of the DERs, for example which are working or are under maintenance. Furthermore, he schedules the maintenance plan for the available DERs.

The DER Operator has access to the abstracted raw data that the DER Owner made accessible. In this data, which constitutes the first level of abstraction of raw data, he is able to add the suitable metadata that are significant for the operation of DERs, such as an ID, a name, and the type of energy. The dedicated application enables the editing and the updating of the metadata. Moreover, through the abstracted data the DER Operator requires to have overview of his component.

Over a web interface or with a dedicated application on a mobile device this role is able to use an authenticated connection to the platform and gets a visualization of the data. The DER Operator has his own view on the aggregated data for monitoring. The DER Operator has the same security requirements for the data as the DER Owner, such as access control, anonymity, and privacy for the data.

The DER Operator is responsible for verifying that the data is stored on the cloud, thus he needs to be informed if there exist constraints that prevent uploading, such as network failures. The DER Operator, connected to a service, gets all relevant data, raw data, as well as metadata, for an overview of the distributed capacity of the DERs as the energy production is dependent on the weather conditions and is volatile. Based on the collected data from his DER he gets a clear view of the amount of energy that is brought to customers through the VPP.

The following table summarizes the requirements of the DER Operator:

	Requirements
DER Operator	Authenticated connection to the platform (web interface or mobile device)
	Status of DER
	Distributed Capacity
	Integrate metadata/update
	Be informed for constraints to upload data
	View historical data, statistics
	Authenticated access to raw and metadata
	Overview of his component
	Security requirements for the data
	Information for Operation of DER

Table 2-2 Requirements for DER Operator

2.2.3.3. VPP Operator

A Virtual Power Plant is an aggregation platform for distributed energy resources. In this platform the capacity and flexibility of different DERs are aggregated. So, a single operating profile for dealing in the energy market is provided. The VPP Operator runs the Virtual Power Plant. The connection of DERs to a VPP gives the VPP Operator the right to access and aggregate data coming from the DER. Thus, he is able to manage the VPP. A VPP Operator over a web interface or with a dedicated application on a mobile device connects to a service on the ICT-Platform to get the relevant information about all the DERs connected to the VPP. The VPP Operator interprets the raw data to extract the specifications of the energy management. As a VPP Operator, he can collect provided data from different DERs and visualize them. The VPP Operator needs to be able to start different calculation processes on demand to make an aggregated view out of the raw data provided by the connected DERs in order to decide for the provided amount of energy to the market. The VPP Operator, considering the schedule of generation based on the weather conditions and the demand based on the needs of his customers, provides the balancing diagram of energy. The communication with the Distribution System Operator (DSO) is essential for the energy balance and the prediction scenario. He can make a prediction to the schedule for the energy generation of the next days, which is quite important in the energy management system based on volatile energy generation. He adjusts the available resources in response to the demand and he recommends actions based on the defined rules and conditions to fulfil the set objectives. Moreover, the VPP Operator uses ancillary services for the management of energy, such as the comparison among the energy production of different DERs or the consumption among the customer. The VPP Operator is involved in the task to plan, control and monitor the energy generation and the energy consumption. Apart from the distributed collection of energy data the VPP contributes also to the demand side management.

The platform stores both raw data and the data that originates from various calculation services made by the VPP Operator for the VPP management. The VPP Operators requires that the necessary metadata for the operation of a VPP is integrated and periodically refreshed. Furthermore, the VPP Operator must be able to create and manage groups according to his needs, such as “same DER Owner”, “DER Operator”, and “regional DERs”. He is responsible for providing the access rights to the respective groups and their members. In order to run efficiently, the VPP Operator requires a detailed overview of all components. It is crucial for the operation of a VPP to provide authenticated access to a combination of aggregated, anonymized and abstracted, raw or metadata, data and public data, such as weather forecasts. In each step the security requirements concerning the dedicated private data need to be fulfilled.

	Requirements
Vpp Operator	Authenticated data access
	Security requirements concerning the data
	Integrate and update metadata
	Access to public data
	Communication with DSO
	Compare DERs-consumers
	Component Overview
	Organize user groups responsible for providing access to involved members
	Schedule of generation and demand
	Plan-control-monitor energy production/consumption
	Access to raw metadata, aggregated, anonymized, abstracted, and public data
	Timeline requirements

Table 2-3 **Requirements for VPP Operator**

2.2.4. Energy use case with focus on energy production

Within the project, the objective of the Energy Use Cases can be summarized as the improvement of energy management considering the integration and the highest possible usage of renewable energy sources. The power distribution grid must become smarter to efficiently incorporate resources that are both decentralized and volatile. Coming from the more general description, the first described use case will focus on energy production.

The proposed approach to achieve the mission of including renewable energy produced by small and medium sized energy generators is to aggregate the different types of energy production (windmill, photovoltaic, biogas) and create a Virtual Power Plant (VPP).

The adopted scenario for the energy use case is summarized as follows:

1. Initially the Distributed Energy Resources (DERs) are connected to VPP.
2. Then the Raw Data of energy generation is gathered and stored continuously.
3. The collected energy data is aggregated and distributed to the VPP for further usage.
4. The VPP Operator is responsible for the energy management.

2.2.5. UC4 - Virtual Power Plant

2.2.5.1. Description

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.

In Figure 2-6 some of the connected DERs are shown.

Figure 2-6 Distributed Energy Resources

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.

2.2.5.2. Scenarios

For the scenario of the VPP use case, the Figure 2-7 shows an overview over the VPP system concerning the ICT- Platform. To build the Virtual Power Plant based on the ICT platform the service for distributed data collection gathers the energy generation data from the different DERs and stores them as raw data on the platform.

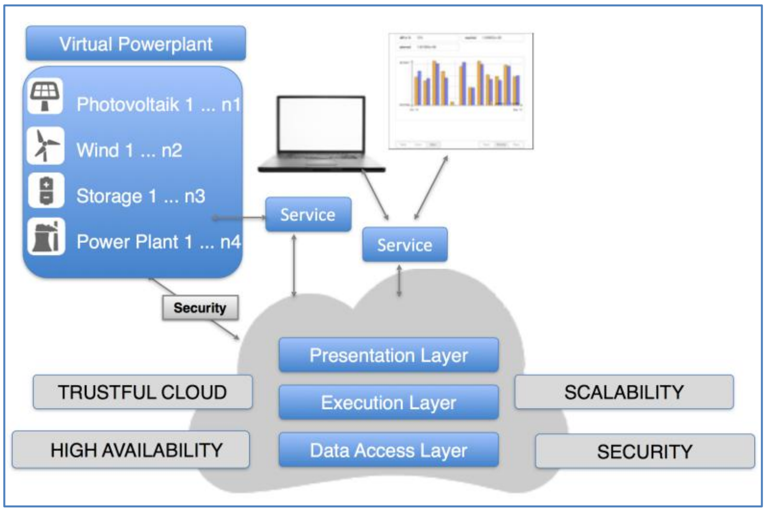

Figure 2-7: Virtual Power Plant built on ICT-platform

Given the sensitivity of the energy system the use case requires trustable Cloud computing with an end-to-end secure data management approach. The data must be secured during the whole processing from transferring the data to the platform during processing and in the storage. The authenticity of the data must be guaranteed and there is only authorized access to the data allowed. Other requirements for the ICT Cloud are high availability and scalability, i.e., an elastic cloud computing platform. The use case processes an increasing amount of data, so there is the need to dynamically adjust the resources available. It would be desirable that the request for more resources such as performance and storage are as transparent as possible for the user to provide the necessary scalability for the use case.

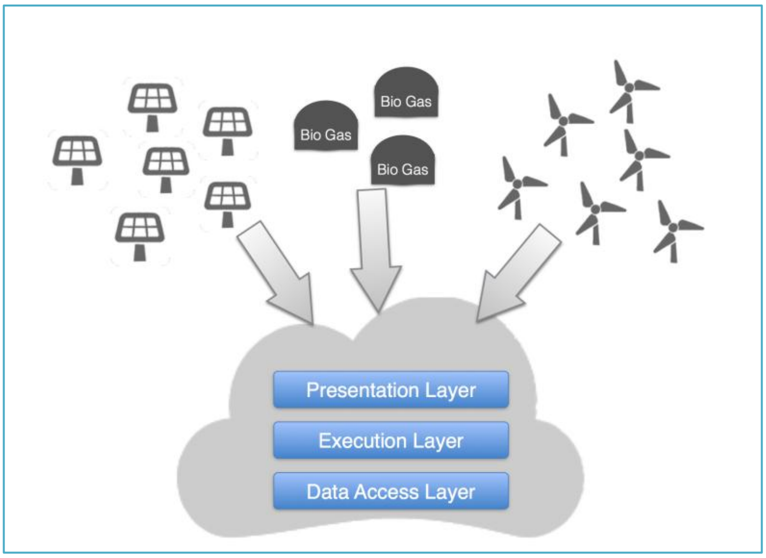

Figure 2-8: Connecting renewable energy resources

In Figure 2-8 the connection of the DERs to the ICT-Platform is shown. The location of the DERs can be geographically widely distributed, even over different countries. The DERs are connected via an Internet connection to the platform and send their data to the "Data Collection Service" of the ICT-Platform. With the "Data Collection Service" the data from DERs are stored as authentic raw data.

For bringing up the VPP system on top of the collected data, it is necessary to bring various aggregation, calculation and visualization services on top of the collected DER energy generation data. The challenge for future ICT Systems is to provide rules and methods for security and access control to the distributed multi-cloud environment. Real-time capability is a major requirement in the energy domain due to the need to react to events and changing conditions in real-time.

2.2.5.3. Requirements for the platform

The described energy use case focuses on the integration of distributed renewable energy resources and increasing the amount of energy used that was produced by DERs. This implies the integration of Big Data Analysis and Cloud Computing realized through an ICT-Platform.

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.

This platform requires a data processing system, which collects data from the individual producers and consumers in the market and then analyses and processes them. Each actor requires authenticated access to his private data. Given the sensitivity of the energy supply system, there are requirements and strict constraints that the platform must fulfil.

This leads to the following constraints and requirements for the platform:

- Big Data management: Large, heterogeneous data sets are collected, stored, analysed and visualized. Following from the scalability and the availability of the SCI-platform it becomes possible to handle a large amount of automatically generated data.
- Object Security: The raw data is stored in a file-based architecture. It is necessary to provide a mechanism that guarantees a trustable and secure storage.
- End-to-End Security: Dedicated, heterogeneous sensor/actor systems are connected to the SCI-Platform. The system needs to allow the collection of data as well as the control of the sensor/actor systems. The combination of private, public and machine-generated data sets and the processing on the cloud-based platform require highest security standards. The data needs to be secured during the transmission as well as within the storage. Moreover, a careful authenticity management, and identity management together with encryption mechanisms is required in a trustful cloud computing. Access control, access documentation, and data abstraction must be some of the main functionalities for the data access layer.
- Real time: The automated change of a state by means of control mechanisms, based on the SCI-platform, takes place in real-time.
- Autonomy: The smart data platform has to adapt to quickly changing circumstance autonomously. In order to do so, intelligent algorithms, which react to impulses from data entry need to be integrated and improved.
- Certification: Throughout the whole process, the certification of the data needs to be established. Starting from the collection of the data to the storage in the cloud. This collected data will be the basis for innovative applications in the energy economy, for example energy brokering or price based decision finding for energy usage. The aggregation platform needs to provide the knowledge for the required analysis and controlling tools for the energy grid as well as for the energy market. The foundation for billing and accounting as well as the compilation of the usage of services to the invoicing are challenges for the platform.
- Interactive services: Through the interactive services the end user demands to have access to the stored data and be able to visualize them. Moreover, he expects to have the control of the involved components.

The above requirements should be fulfilled by the cloud-based Smart Core Interworks platform. The concept of this SCI-platform is depicted in Figure 2-9.

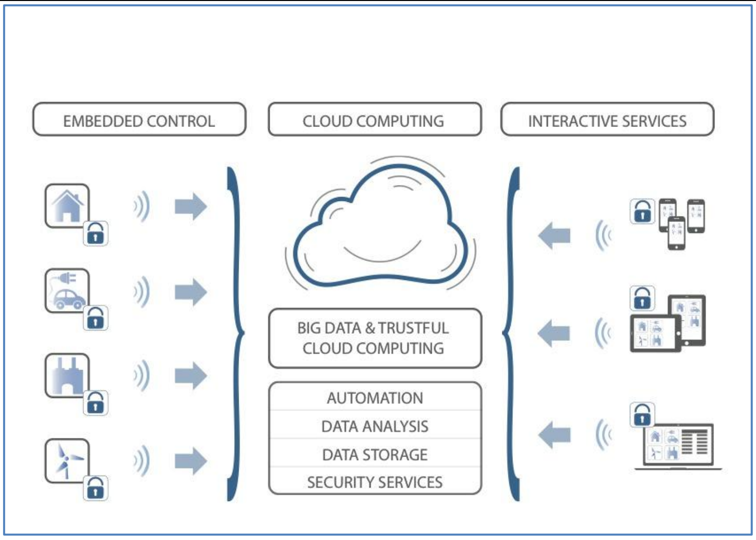

Figure 2-9 Smart Core Interworks ICT Platform

The SCI Platform has to ensure a secure and stable solution for usage in distributed systems based on the three pillars: Embedded Control, Cloud Computing and Interactive Services.

2.2.6. Workflow DER Owner – DER Operator

A DER Owner connects to a Service on the SCI-Platform to get the information about his DER. For every DER owner it is possible to get the aggregated information about his resource, for example the actual generated amount of energy, as well as the hourly or monthly summations and analysis. He can connect to the platform after authentication, for example over a web interface or with a dedicated application on a mobile device. Based on the collected data from his DER, he gets a tailored view on the generated energy and the amount of energy brought to market via the VPP.

The DER Operator, who can also be the DER Owner, connects to the SCI-Platform to get information about his DER. This includes raw data as well as metadata from his DER which contains the necessary information for this role to operate the DER.

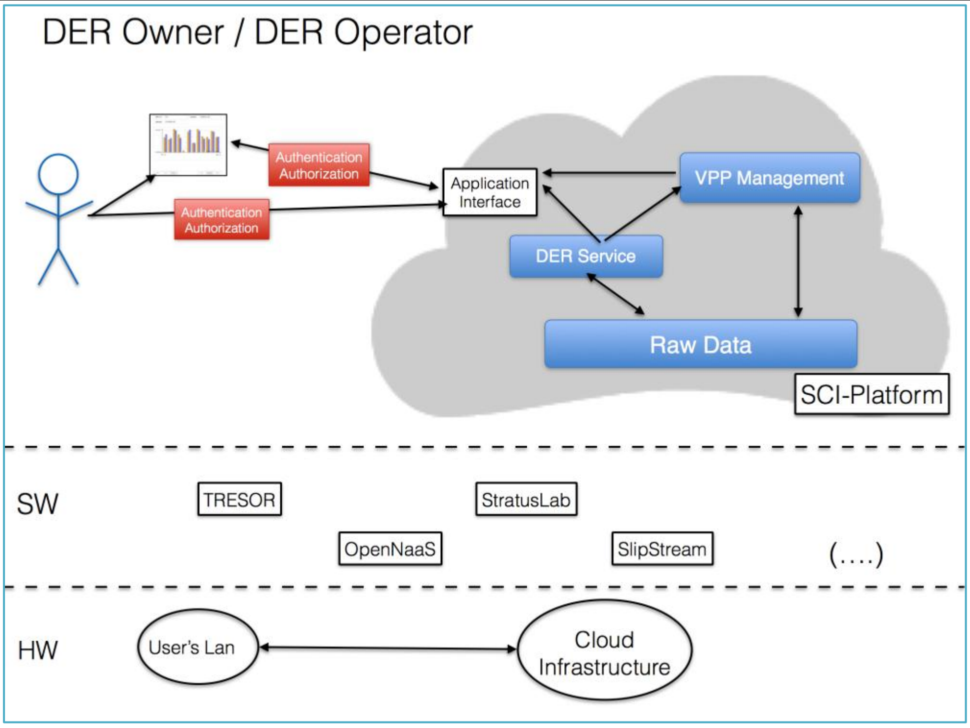

Figure 2-10 **Workflow DER Owner/DER Operator**

2.2.7. Workflow VPP Operator

A VPP Operator connects to a Service on the SCI-Platform to get the relevant information about all the DERs connected to the VPP. He can start different calculation processes on demand to make an aggregated view out of the raw data provided by the connected DERs. This calculation processes can be started directly by the VPP Operator or automatically with a time schedule. Running the calculation processes, the VPP Operator gets the aggregated information, for example the actual generated energy as a summation, so that he can make decisions for bringing this amount of energy to the market.

Besides the actual generated energy data, following the Big Data approach, a huge amount of heterogeneous data coming out of the different calculation services for the VPP management are stored on the platform. The VPP operator has access to this stored data and derived from them and in combination with for example weather forecasts (which are publicly available) he can make a prediction for the schedule of energy generation in the coming days, which is quite important for an energy management system based on volatile energy generation. The VPP Operator can connect to the platform after authentication, for example over a web interface or with a dedicated application on a mobile device.

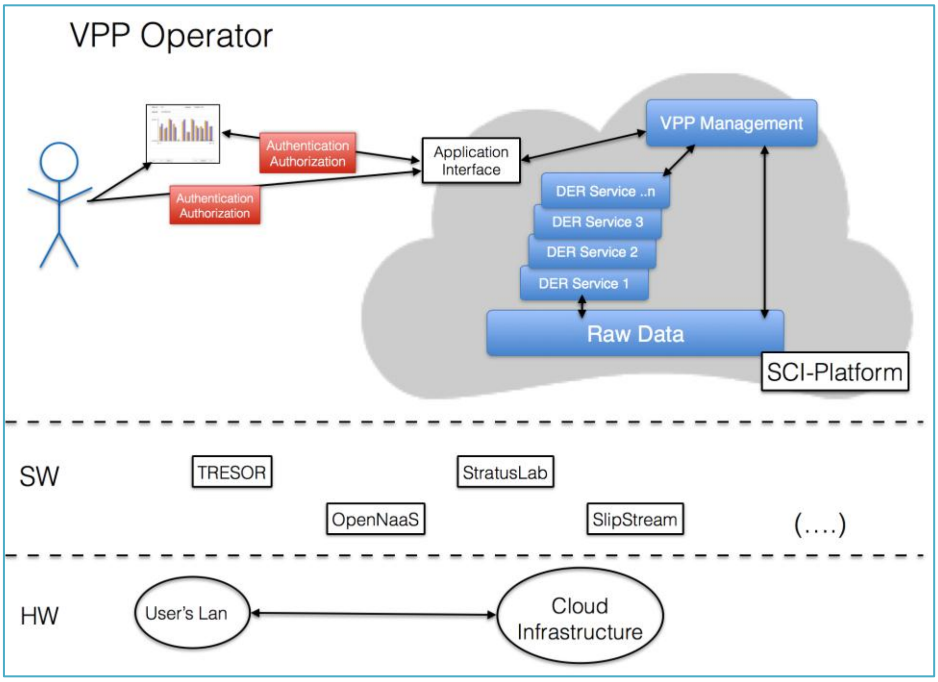

Figure 2-11 **Workflow VPP Operator**

Evaluation:
This use case will evaluate the capabilities of the CYCLONE features to deploy a complex application on distributed cloud environment. The deployment of the application may require different clouds depending on availability of cloud resources and requirements of application provider. Storing and archiving the gathered heterogeneous raw data will be done with respect to the end-to-end security requirements.

Role based data access control and data access documentation is also part of the evaluation.

To give access to the different abstraction of gathered raw data, a Cloud user interface for data access service is evaluated. The data access will be provided by APIs based on web technologies respectively by dedicated mobile applications.

2.2.8. Overall relation of the energy use case to CYCLONE ambition and main objectives

As an application provider, CYCLONE components are used to deploy the modules of the SCI-platform over a multi-cloud environment. The different cloud sites are connected via public internet connection. The requirement for the different, non-dedicated cloud infrastructure is to be able to run on unix-like systems.

In this context, security plays a major role in the system:

- Data access control and access documentation
- Data security
    - Availability – 24/7
    - Authenticity
    - Secured storage

- Network security
  - Communication security
  - Transmission security

3. Use case requirements and implications for the CYCLONE software

The detailed description of the use cases, including the actors and users' stories, led to the definition of a set of requirements for each of the six use cases. From all these requirements, a common list has been compiled. This list provides the basis for the identification of implementation tasks that need to be carried out in the project in order to implement those requirements.

3.1. Common requirements list of the use cases

Although the cloud infrastructure has proven to be a very useful framework for analysing genomic information and for other bioinformatics studies, moving to a federated infrastructure is the next obvious step, for example coordinating the regional and national services between the IFB-core and the hospitals or the sequencing centers and providing the community with larger and fault-tolerant resources.

Regarding the Virtual Power Plant use case, we have identified a list of initial requirements for the usage of ICT in energy management systems. The increasing number of participants and controllable components in the energy grid lead to a coordinated cloud environment. Associated with this situation of an increasing number of different data owners, secured data management such as giving access to different layers of abstractions to the active partners of the energy system is one of the main requirements.

However, from both Bioinformatics and Energy use cases a number of identified features are required to definitely enhance the cloud infrastructures' current performance, mainly in terms of security, management, automation of processes and dynamicity. These common requirements have been identified and are listed in the following table.

Table 3-1 describes the common list of use case requirements.

Table 3-1 Common list of use case requirements

ID	Title	Description
1	Cloud User Interface for service access based on web technology	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	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	Users can connect through a dedicated application on a mobile device.
4	Federated identity management	A common identity management system (based on RENATER/eduGAIN) across the cloud federated resources ensures consistent authentication of users.
5	Federated authorization management	Community administrators can define groups and associated authorizations to access cloud resources: VMs, storage volumes, etc.

6	One-click deployment of simple application	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	The access to the web interface of a bioinformatics service provided by a VM is done according to the identity federation and authorizations.
8	End-to-end secure data management	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	All the VMs of one user need to be in the same local network isolated from other users.
10	VPN connectivity services	The user can access to all of its VMs through a single point of entry based on VPN mechanisms.
11	FedId Access to storage	The user has access to the cloud storage according to its federated identity.
12	Community reference datasets	The reference human genome hg19 (Human Genome version 19 or GRCh37) is required for analysing human biomedical data. It is a database consisting of many files containing the public genomics data. It needs to be previously deployed by the cloud providers as a public data set available to all users
13	Access to public Cloud Services	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	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	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	Running a complex bioinformatics application requires the deployment of several VMs copied from appliances with different complementary features.
17	One-click deployment of Complex application	The user should be able to run a complex application with one click.
18	High-throughput storage	Some life science applications require high-throughput access to large datasets from many VMs.
19	Multi-VMs shared volume	User VMs need to share the same volume.
20	Multi-clouds deployment of complex application	A complex bioinformatics application requires to be deployed over two or more cloud infrastructures to obtain the necessary computing resources.
21	Multi-clouds distribution of community reference datasets	The deployment of a bioinformatics workflow over two or more cloud infrastructures requires that the collections of public reference data used during the treatment is available in all of these clouds
22	FedId SSH connection to the VM	The access to the VM is allowed according to the user identity and roles in the identity federation.
23	Elastic management of complex applications	Running a complex bioinformatics application requires resource scale-up and scale-down options (in terms of computing and data)

24	Dynamic Network resource allocation	Deploying a complex bioinformatics 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.
25	Multi-clouds distribution of user data	Deploying a complex bioinformatics 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	UC3 requires the dynamic allocation of high-bandwidth links between the sequencing centers and the cloud storage. The dynamic network management will enable "live data processing" where the sequencer can be connected to the necessary computing resources to fully analyse the data as it is produced.
27	Archiving of raw data	The raw data will be stored as archives, and also put on the shared workspace to be treated.
28	Configuration of volatile disks	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	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	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	The sequencing engineer and the life science researcher are both cloud users and need to share the raw genomics data and the annotation environment. The authorization relies on the static roles/groups to which the user belongs in the identity federation.
32	User-defined authorizations	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	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)	The upload of large datasets (Bioinformatics UC3) and the use of remote virtual desktop require guaranteed network performance (Bioinformatics UC2).

3.2. Deriving Implementation Tasks

Based on both the list of use cases and the list of requirements, the following implementation tasks have been identified, which CYCLONE needs to implement in order to create the desired CYCLONE functionality to meet the requirements of the users.

The following table provides a numbered list of Use Cases in this document, which are referred to by the table of implementation tasks. As only the bioinformatics use cases are detailed enough to derive implementation tasks, only they are represented in this list and the list of implementation tasks.

Table 3-2 List of user stories

ID	Use Case	Description
1	UC1	As a biomedical staff member, I can have a personal cloud account, based on my identity in the academic federation (the national one provided by RENATER or international by eduGAIN).
2	UC1	As a biomedical staff member, have access to the authenticated web dashboard to manage my secured cloud resources.
3	UC1	As a biomedical staff member, I have access to the required appliances, published in the marketplace, to help me treat the human biomedical data.
4	UC1	As a biomedical staff member, I can run a VM to analyse the biomedical data.
5	UC1	As a cloud provider, I can guarantee that the VM will be deployed in a user-devoted network, isolated from other users
6	UC1	As a cloud provider, I can guarantee that the connection from the user desktop to the VM is secured.
7	UC1	As a cloud provider, I can guarantee that the VM data (disk image and data storage) will be erased at the end of the execution.
8	UC2	As a bioinformatician, I can have a personal cloud account, based on my identity in the academic federation (the national one provided by RENATER or international by eduGAIN).
9	UC2	As a bioinformatician I have access to the authenticated web dashboard to manage my cloud resources.
10	UC2	As a bioinformatician, I have access to the required appliances, published in the marketplace, to help me to annotate my bacterial genomics data.
11	UC2	As a bioinformatician, I can deploy and synchronize the required public reference data collections in the targeted clouds.
12	UC2	As a bioinformatician, I can deploy a distributed cluster of VMs to annotate my unknown microbial genomes.
13	UC2	As a bioinformatician, I can run a VM to visualize my genomic data.
14	UC2	As a cloud provider, I can guarantee that the VM will be deployed in a user-devoted network, isolated from other users.
15	UC2	As a cloud provider, I can provide users with the complete deployment of a distributed cluster of VMs sharing a common workspace.
16	UC2	As a cloud provider, I can provide users with the deployment of a VM with a visualization environment, connected to the shared storage.
17	UC3	As a sequencing engineer/bioinformatician, I can have a personal cloud account, based on my identity in the academic federation (the national one provided by RENATER or international by eduGAIN).
18	UC3	As a sequencing engineer/bioinformatician, I have access to the authenticated web dashboard to deploy an instance of such VM.
19	UC3	As a sequencing engineer/bioinformatician, I have access to the required appliances, published in the marketplace, to help me to analyse and visualize my genomics data .
20	UC3	As a sequencing engineer, I can upload the genomic raw data (reads) to the cloud

		storage to be archived and stored in a shared workspace.
21	UC3	As a sequencing engineer, I can run a distributed cluster of VMs to do the first analysis of the raw data (mapping).
22	UC3	As a sequencing engineer, I can manage the authorizations rules to provide a bioinformatician with the data access rights.
23	UC3	As a bioinformatician, I can run a VM containing the useful tools (e.g. IGV [9]) to visualize the genomics data.
24	UC3	As a cloud provider, I can provide a sequencing engineer/bioinformatician with the complete deployment of a distributed cluster of VMs sharing a common workspace.
25	UC3	As a cloud provider, I can guarantee that the VMs will be deployed in a user-devoted network, isolated from other users.
26	UC3	As a cloud provider, I can provide users with the deployment of a VM with a visualization environment, connected to the shared storage.

The following table contains all of the identified implementation tasks. Each task is relevant for one or more components. The responsible persons for the implementation of the tasks are those responsible for the components. We tried to assign Story Points to each implementation task to estimate the time it will take to complete it. Those use the equivalence of 1 SP = ½ Day of work.

Table 3-3 List of identified implementation tasks

ID	UC	Stories	Reqs.	Task	Components	Responsible	Status	Stry Pnts
1	1,2,3	1,8,17	4,5,31,32	Integrate Federation Provider with Bioinformatics Portal	Bioinformatics Portal	IFB	In Progress	
2	1,2,3	1,8,17	1,2,3,4	Offer Federation Provider	Federation Provider	TUB	Done	20
3	1,2,3	2,9,18	4,5,31,32	Rely on user attributes for authentication within the Portal	Bioinformatics Portal	IFB	In Progress	
4	1,2,3	2,9,18	4	Offer user attributes within the JSON Web Token	Federation Provider	TUB	In Progress	4
5	1,2,3	3,10,19	1,2,3,4	Offer a marketplace functionality for Bioinformatics appliances	Bioinformatics Portal	IFB	In Progress	
6	1	3	4,7	Integrate OpenID Connect into HTTP server in UC1 appliance	Bioinformatics Appliances	IFB	Done	
7	1	3		Create appliance for UC1	Bioinformatics Appliances	IFB	Done	
8	1	4	6	Create SlipStream deployment recipe for UC1	Bioinformatics Appliances	IFB	In Progress	
9	1,3	5,25	9	Create isolated Network for a deployed VM	OpenNaaS	I2CAT		
10	1	6	8	Integrate TCTP into UC1 appliance	TCTP	TUB	Planned	6
11	1,2,3	7	15	Discuss approach for secure data erasure	Bioinformatics Appliances, IaaS	IFB/LAL	Planned	
11	1,2,3	7	15	Ensure data erasure of destroyed VMs and data	StratusLab, OpenStack	IFB/LAL	Planned	
12	2	11	12	Integrate shared public reference data into a cloud deployment	Multiple	IFB/LAL/SixSq	Planned	?/?/10

ID	UC	Stories	Reqs.	Task	Components	Responsible	Status	Stry Pnts
13	2	12	16,17	Create cluster deployment recipe for UC2	Bioinformatics Appliances	IFB	Done	
14	2	12	4,8,22	Allow federated SSH login for Bioinformaticians to upload data to UC2 appliance	Federated SSH Login	TUB	Planned	10
15	2	13	4,22,29,30	Integrate X2Go with the federated SSH login	Federated SSH Login	TUB	Planned	6
16	2	5, 14, 25	9	Create isolated Network for a set of deployed VMs (partitioning and isolation mechanism)	OpenNaaS	I2CAT		
18	2	16	11,16	Connect VM clusters to shared reference dataset	Multiple	IFB/LAL/SixSq	Planned	?/?/6
19	3	20	8,18	Provide Cloud Storage for archiving raw data (reads) and energy data	Multiple	IFB/LAL/SixSq	Planned	?/?/?
20	3	20	4,5,11,18, 19,27	Allow sharing the raw data (reads)	Multiple	IFB/LAL/SixSq	Planned	?/?/6
21	3	21,23	16	Create cluster deployment recipe for UC3	Bioinformatics Appliances	IFB	Planned	
22	3	22	1,2,3,4,5, 31,32	Integrate federated authorization into the architecture	Security Architecture	TUB	Planned	10
23	3	24,26	16,17	Allow a cluster deployment to include access to a shared VM	Multiple	IFB/LAL/SixSq	Planned	?/?/6
25	2	13	4,7	Integrate OpenID Connect into HTTP server in UC2 appliance	Bioinformatics Appliances	IFB	Planned	
26	3	23	4,7	Integrate OpenID Connect into HTTP server in UC3 appliance	Bioinformatics Appliances	IFB	Planned	
27	2,3	-	20	Create demonstrators for multi-cloud deployment	Bioinformatics Appliances	IFB	Planned	

ID	UC	Stories	Reqs.	Task	Components	Responsible	Status	Stry Pnts
28	2,3	-	21	Create demonstrators for multi-cloud "sharing" aspect	Bioinformatics Appliances	IFB	Planned	
29	2,3	-	23	Elastic scaling of UC2 and UC3 clusters	Bioinformatics Appliances	IFB	Planned	
30	2,3	-	24	Integrate ad-hoc network resource configuration into UC2 and UC3	Bioinformatics Appliances	IFB	Planned	
31	2,3	-	25	Enable multi-cloud data sharing	Multiple	IFB/LAL/SixSq	Planned	?/?/10
32	1,2,3, 4	11, 15, 20, 23, 24, 5, 11, 12, 13, 14, 15, 16, 20, 21, 23, 24, 25, 26	26	Enable configuration of high-bandwidth DC links	OpenNaaS, Multiple	I2CAT		
33	2,3	-	28	Integrate volatile disks into the use case deployments	Multiple	IFB/LAL/SixSq	Planned	?/?/6
34	1, 2, 3	3, 10, 19, 11, 15, 20, 23, 24	8, 21	Enable firewalling mechanism for application deployment	OpenNaaS, Multiple	I2CAT	Planned	
35	1, 2, 3	3, 10, 19, 6, 20	8, 10	Enable VPN mechanism for application deployment	OpenNaaS, Multiple	I2CAT	Planned	

ID	UC	Stories	Reqs.	Task	Components	Responsible	Status	Stry Pnts
36	1, 2, 3, 4	5,11, 12, 13, 14, 15, 16, 20, 21, 23, 24, 25, 26	24	Provide Network connectivity reallocation	OpenNaaS, Multiple	I2CAT	Planned	
37	1, 2, 3, 4	5,11, 12, 13, 14, 15, 16, 20, 21, 23, 24, 25, 26	34	Enable QoS (BW, Latency) while provisioning connectivity	OpenNaaS, Multiple	I2CAT	Planned	

We suspect the requirements of the Energy Use Case to be similar, therefore expect our contributions to be applicable to both use cases.

4. Evaluation of use cases

4.1. Deployment of the use cases

The deployment of the Bioinformatics use case was done in a progressive manner, beginning with a one-VM use case integrating security features (UC1 Securing human biomedical data). Afterwards, we then deployed a complex application requiring the coordinated deployment of several VMs (UC2 Cloud virtual pipeline for microbial genomes analysis).

4.1.1. Deployment of the UC1 Securing human biomedical data

The first developments are related to the UC1 Securing human biomedical data, which can be seen as a single-VM deployment but with enhanced security features. In Figure 4-1, the upper part describes the use case workflow. In the middle, the steps of the workflow (1, 2...7) are linked to the related CYCLONE software components and services. The bottom part shows the testbed (HW) infrastructure. The VM was instrumented with the Federation Provider and deployed in the IFB's cloud infrastructure.

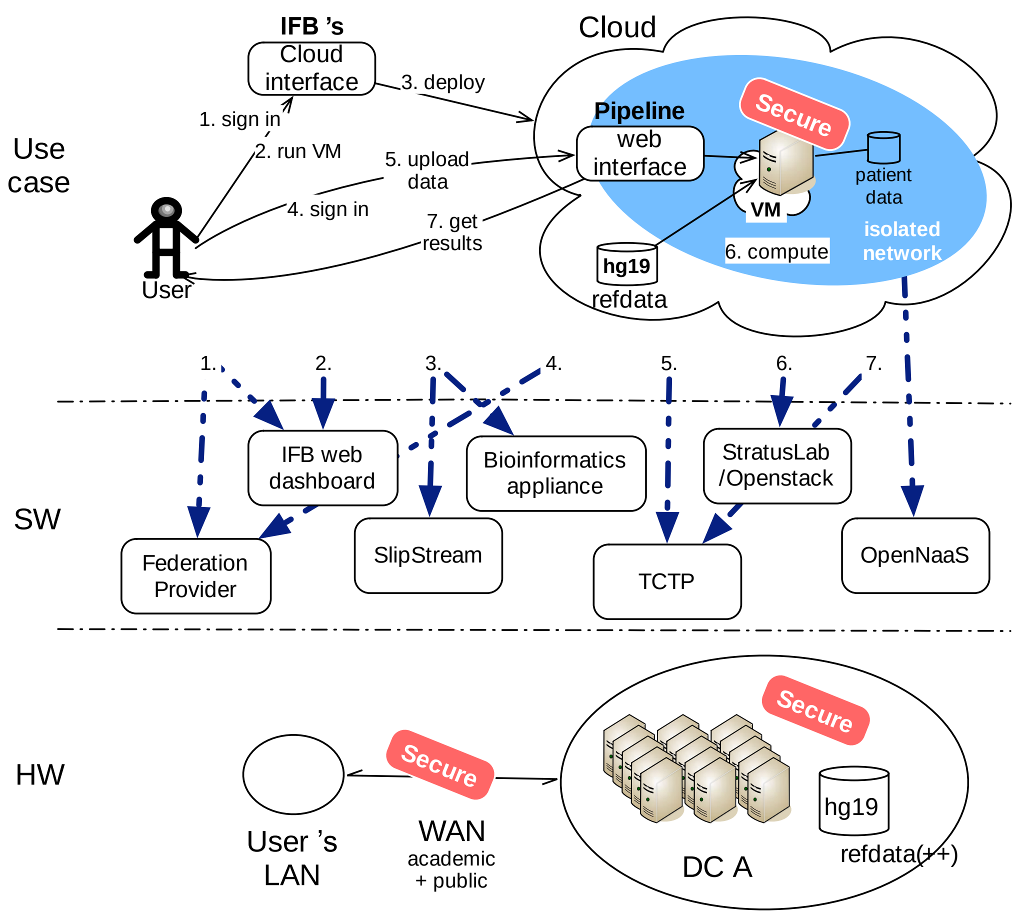

Figure 4-1: Functional relations between the use case "Securing human biomedical data" and the Cyclone components.

4.1.2. Deployment of the UC2 Cloud virtual pipeline for microbial genomes analysis

The second deployment was about the UC2 Bacterial genomics that requires features related to the deployment of a complex application. The first tests were done with a generic bioinformatics appliance built previously by IFB (the Biocompute node), and requiring a cluster mode. The image was exported from the IFB's cloud and registered in the StratusLab Marketplace. Then a deployment recipe based on SlipStream instantiates the complete application with all the required VMs on the testbed infrastructure in CNRS LAL's site. In Figure 4-2, the upper part describes the use case workflow. In the middle the steps of the workflow (1, 2...8) are linked to the related CYCLONE software components and services. The bottom part shows the testbed (HW) infrastructure. This was demonstrated in the F2F meeting in Berlin. Then the appliance containing the whole Insyght pipeline was deployed according to the recipe defined in the previous step.

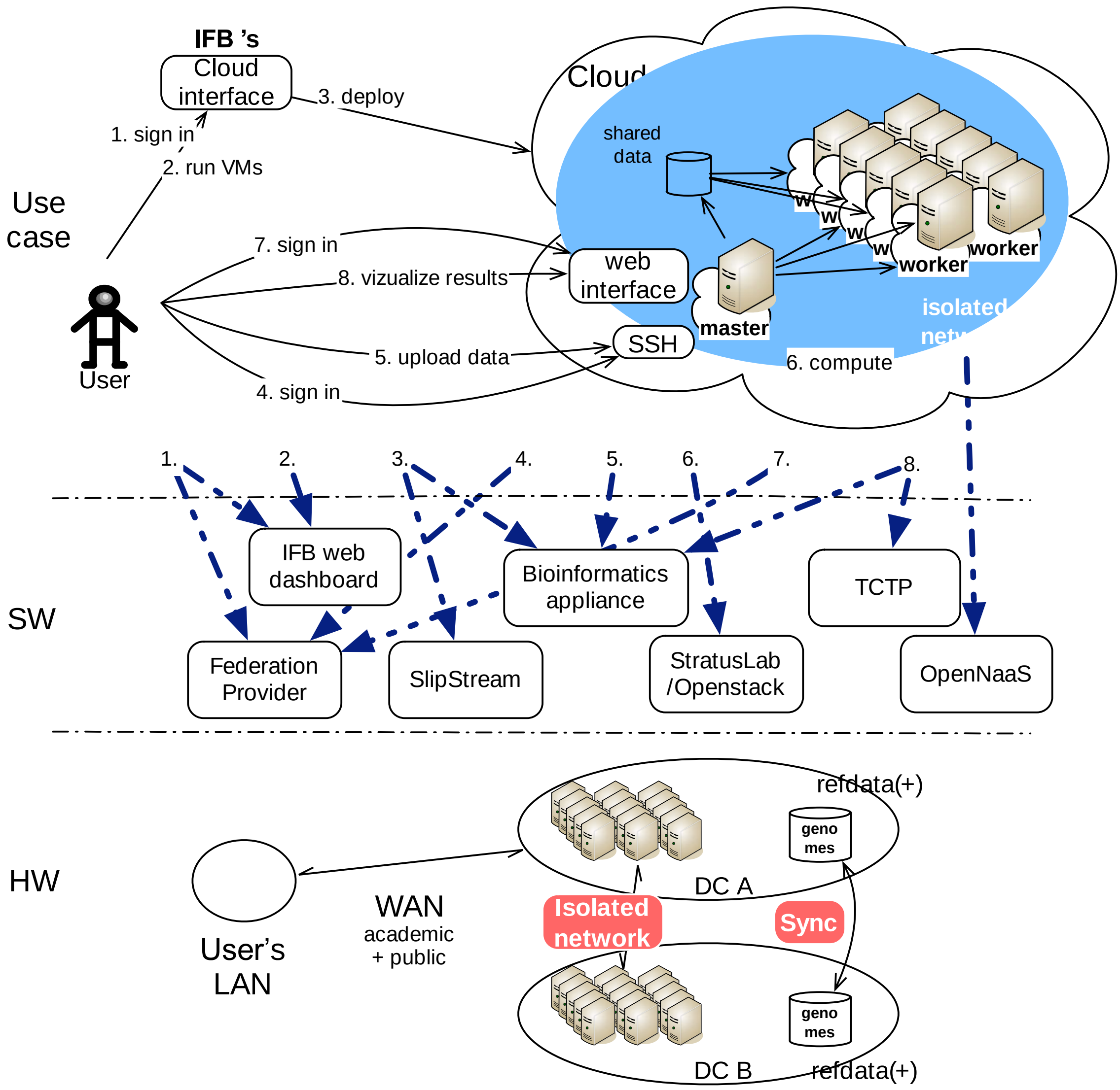

Figure 4-2: Functional relations between the use case "Cloud virtual pipeline for microbial genomes analysis" and the Cyclone components.

4.2. Test scenarios

The test scenarios for the Bioinformatics and Energy use cases are the following:

UC 1 NGS biomedical analysis

- Deploy a simple application from a web interface in one click
- Use the federation provider to access the VM web interface
- End-to-end secured upload of user data
- Get access to community reference dataset (HG19)
- Analyse human genomics data (public)

UC 2 Bacterial genomics

- Deploy a complex application from a web interface in one click
- Isolated network for the VMs (based on *iptables*)
- Upload of user data
- Annotate Bacterial genomes (intensive multi-VMs run)

UC 4 Virtual Power Plant

- Connect DER to VPP and with this give the VPP Operator access rights and aggregate data coming from DER
- The VPP operator starts an aggregation process over all connected DERs and visualizes the results. Therefore, he starts the aggregation application.
- The VPP operator can connect to the Cloud User Interface to see the presentation of the results.
- The DER Owner and the DER operator connect to the DER application service and get a visualization of data and metadata from the DER

The following table contains an estimation of the completion of the implementation tasks, scheduled for M10, M14, and M18.

Table 4-1: Timeline for the deployment of use cases and associated tests

Description	Component	Resp.	Status	-M10	M10 - M14	M14 - M18	Later
Integrate Federation Provider with Bioinformatics Portal	Bioinformatics Portal	IFB	In Progress		X		
Offer Federation Provider	Federation Provider	TUB	Done	X			
Rely on user attributes for authentication within the Portal	Bioinformatics Portal	IFB	In Progress		X		
Offer user attributes within the JSON Web Token	Federation Provider	TUB	In Progress		X		
Offer a marketplace functionality for	Bioinformatics Portal	IFB	In Progress		X		

Description	Component	Resp.	Status	-M10	M10 - M14	M14 - M18	Later
Bioinformatics appliances							
Integrate OpenID Connect into HTTP server in UC1 appliance	Bioinformatics Appliances	IFB	Done	**X**			
Create appliance for UC1	Bioinformatics Appliances	IFB	Done	**X**			
Create SlipStream deployment recipe for UC1	Bioinformatics Appliances	IFB	In Progress		**X**		
Create isolated Network for a deployed VM	OpenNaaS	I2CAT	Planned			**X**	
Integrate TCTP into UC1 appliance	TCTP	TUB	Planned		**X**		
Discuss approach for secure data erasure	Bioinformatics Appliances/IaaS	IFB/LAL	Planned		**X**		
Ensure data erasure of destroyed VMs and data	StratusLab/OpenStac k	IFB/LAL	Planned			**X**	
Integrate shared public reference data into a cloud deployment	Multiple	IFB/LAL/SixS q	Planned		**X**		
Create cluster deployment recipe for UC2	Bioinformatics Appliances	IFB	Done	**X**			
Allow federated SSH login for Bioinformaticians to upload data to UC2 appliance	Federated SSH Login	TUB	Planned			**X**	
Integrate X2Go with the federated SSH login	Federated SSH Login	TUB	Planned			**X**	
Create isolated Network for a set of deployed VMs (partitioning and isolation mechanism)	OpenNaaS	I2CAT	Planned			**X**	
Connect VM clusters to shared reference dataset	Multiple	IFB/LAL/SixS q	Planned			**X**	
Provide Cloud Storage for archiving raw data (reads) and energy data	Multiple	IFB/LAL/SixS q	Planned			**X**	
Allow sharing the raw data (reads)	Multiple	IFB/LAL/SixS q	Planned			**X**	
Create cluster deployment recipe for UC3	Bioinformatics Appliances	IFB	Planned		**X**		
Integrate federated	Security Architecture	TUB	Planned		**X**		

Description	Component	Resp.	Status	-M10	M10 - M14	M14 - M18	Later
authorization into the architecture							
Allow a cluster deployment to include access to a shared VM	Multiple	IFB/LAL/SixSq	Planned			X	
Integrate OpenID Connect into HTTP server in UC2 appliance	Bioinformatics Appliances	IFB	Planned		X		
Integrate OpenID Connect into HTTP server in UC3 appliance	Bioinformatics Appliances	IFB	Planned		X		
Create demonstrators for multi-cloud deployment	Bioinformatics Appliances	IFB	Planned			X	
Create demonstrators for multi-cloud "sharing" aspect	Bioinformatics Appliances	IFB	Planned			X	
Elastic scaling of UC2 and UC3 clusters	Bioinformatics Appliances	IFB	Planned			X	
Integrate ad-hoc network resource configuration into UC2 and UC3	Bioinformatics Appliances	IFB	Planned			X	
Enable multi-cloud data sharing	Multiple	IFB/LAL/SixSq	Planned			X	
Enable configuration of high-bandwidth DC links	OpenNaaS, Multiple	I2CAT	Planned			X	
Integrate volatile disks into the use case deployments	Multiple	IFB/LAL/SixSq	Planned			X	
Enable firewalling mechanism for application deployment	OpenNaaS, Multiple	I2CAT	Planned			X	
Enable VPN mechanism for application deployment	OpenNaaS, Multiple	I2CAT	Planned			X	
Provide Network connectivity reallocation	OpenNaaS, Multiple	I2CAT	Planned			X	
Enable with QoS (BW, Latency) while provisioining connectivity	OpenNaaS, Multiple	I2CAT	Planned			X	

4.3. Results

Two bioinformatics use cases were deployed in M10. They are described in details in D7.2 Overlay with focus on Component Manager, with a summary below, and demonstrated at the F2F meeting in Berlin.

The UC1 - Securing human biomedical data is a single-VM deployment integrating enhanced security features. The cloud appliance NGS-Unicancer [10] is developed by the bioinformatics platform of the Centre Léon Bérard (Lyon, France) in the context of the project NGS-Clinique (INCA - Institut National du Cancer). It provides a simple web interface to launch the analysis pipeline. The appliance was enhanced by the Federation Provider and deployed on the IFB's cloud infrastructure. The user deploys the appliance NGS-Unicancer through the IFB's web interface in one click (steps 1-3 in Figure 2-2) and uses the CYCLONE federation provider to get access to the VM web interface based on its identity in the federation (step 4). The user can then upload its data, run the analysis and get the results (steps 5-6-7).

The UC2 Cloud virtual pipeline for microbial genomes analysis was extended by SlipStream deployment recipes. The first tests were done with a generic bioinformatics appliance built previously by IFB (the appliance BIO compute node [11]), and requiring a cluster mode. BIO compute node is an image, based on Centos, which provides a variety of bioinformatics tools highly used by life science researchers. The image was exported from the IFB's cloud and registered in the StratusLab Marketplace. Afterwards, a deployment recipe based on SlipStream instantiates the complete application with all the required VMs on the testbed infrastructure at the CNRS LAL site.

5. Potential new use cases

The purpose of new use cases is to:

- Extend current use cases
- Bring in use cases of other areas
- Provide some additional features and value to the CYCLONE framework.
- Bring use cases that show how users can take the CYCLONE solution, deploy and make use of the functionality. Perhaps we could Involve (if available) SlipStream users that could be addressed as use cases.

We have identified different possibilities described below.

TUB proposals are:

- Provisioning IoSL and thesis resources
    - There are many chairs at TUB which conduct something similar to our (SNET chair) own "Internet of Services Lab", and also conduct student supervision, where we constantly face the challenge how we – for example – allow students to rapidly provision VMs, let other students access these VMs and our own tools, at best with their own TUB account. We could maybe leverage SlipStream, StratusLab and the FP to offer a better and more rapid provisioning of student resources.
- Brokering & Matchmaking
    - This is a common requirement for most applications and can be used for placement of application services for the existing use cases and we will search for more advanced use cases in addition.
    - ENTRANCE[2] use case
    - TUB is working on an attribute-based encryption system in the ENTRANCE project which we could maybe integrate into CYCLONE for secure data sharing scenarios.
- Integration of Federation Provider into IoSL prototypes
    - There are many prototypes students create at the chair. Integration of the Federation Provider into these use cases could allow worldwide access to these demonstrators with local university accounts. We already integrated the FP successfully into a demonstrator this year.[3]

From a SlipStream perspective, the most likely candidates for applications are in the list below.

- Utility Distribution: We have interest in the utilities sector on using cloud platforms to provide a more flexible infrastructure for monitoring analysis and equipment control. They might potentially

[2] http://entrance.snet.tu-berlin.de/
[3] https://github.com/IoSL-INav/backend

be interested in the CYCLONE components for encrypting data in transit and integration of federated identity capabilities in the platform. Given that these systems run on purchased networks, they probably would be less interested in the networking aspects of the project. This comes from another H2020 project (SCISSOR) that SixSq is involved in.

- Scientific Data Distribution and Analysis: A recurring need for ESA datasets is making them available to the public but at the same time monetize them in some way (either directly or indirectly). The analysis model and architecture are not really clear at this point, but they might be interested in something like the “live analysis” in bioinformatics where large bandwidth pipes are opened between a client and a dataset on demand. They might also be interested in adding features to SlipStream/CYCLONE that would handle data services more efficiently (access, duplication, co-location of computing, etc.).
- Automated Placement: We have a few initiatives around the SlipStream Service Catalog and automated placement of application components based on location, certifications, pricing, etc. The need for this is coming from another European project that is representing public administrations.

From an energy perspective, following the demands of the future energy supply system leads to extending use cases in the surroundings of the energy management system. To increase the usage of renewable produced energy, the consumption of energy has to be managed in a more direct way. Because of the volatile availability of renewable energy, the focus is led to the demand side management and a balanced demand response system. Billing and accounting requirements for the Utility 4.0 should also be a part of the system for the sector of the energy consumption. In this use case CYCLONE components for security in multi-cloud environment will be interesting, as well as the dynamic application deployment for an elastic usage of cloud computing.

- Energy Consumption
  - Demand Side Management
  - Demand Response
  - Utility 4.0

In bioinformatics, the various analyses that can be performed with Next Generation Sequencing (NGS) data are good candidates for CYCLONE “use cases”. In particular workflows for:

- Mapping reads on known genomic sequences
- Assembling genomes from reads
- Detecting variants in genomics sequences (SNPs, gene copy number, indels, structural rearrangements)
- Transcriptomics (RNA-seq): differential gene expression, alternative splicing and mRNA structure
- Studying gene regulation: ChIP-seq (transcription factor binding sites), sRNA-seq (small non-coding RNA), epigenetics (genome methylation)
- Annotating large eukaryotic genomes
- Metagenomics analyses (amplicons and whole genomes)

Most of these workflows have been ported to the Galaxy environment (a web interface that allows users to run programs without using the command line interface that has become very popular in the life science community). Some pipelines are also run from (virtual) desktops. NGS analyses are characterized by large amounts of data that need to be transferred via the network before being processed. Some analyses, such as the genome assembly of large eukaryotic species, require access to large amounts of RAM (>1TB), others access to important distributed computer capacities. Processed data and results are often stored in relational databases. Increasingly, NoSQL solutions are adopted to store the data. Users, in particular those using 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). For some data, such as biomedical data, confidentiality is of prime importance. Authentication and authorization infrastructures are thus crucial as well as the possibility of isolating VMs of a particular user within a private virtual network.

Imaging (cellular and biomedical) is another domain of life sciences that can provide relevant use cases for CYCLONE. Although the data handled is different and does not have the same characteristics, images vs. sequences, the life science community faces many similar issues starting with the storage and processing of large amounts of raw data.

6. Preliminary market analysis and business model investigations

This section provides a brief overview on the actual Cloud market and an outlook on the foreseeable mid-term and long-term period based on market trends analysis.

Over the last few years, Cloud service adoption by SMEs and LSEs (Large Scale Enterprise) has been evolving from an emerging phenomenon to an established IT and business solution that is gaining widespread acceptance and deployment so that even mission-critical workloads are being moved to the cloud. On the end-user side, cloud services offer ubiquitous access to content and applications on multiple devices, that can be used almost anywhere.

Many telecommunication and IT operators provided several product and service offers targeting this increasing market demand. Year over year this opportunity has become stronger and thus more profitable, generating conspicuous revenues that enabled an always increasing effort and investments in IT facilities and infrastructures.

Interoute delivered VDC products on January 2012 after 6 months of testing with more than 200 top business customers using 3 big data centers in London, Amsterdam and Geneva. At the moment the global cloud services platform encompasses 12 data centres, 14 Virtual Data Centres and 31 colocation centres relying on an IP/MPLS backbone (See Figure 6-1). This remarkable leap can give us an idea of what “increasing cloud service adoption” means. Currently, more than 60% of Interoute's business is generated from enterprises on VDC and managed hosting services, coupled with professional services. The remaining 40% is generated by traditional wholesale customers (connectivity).

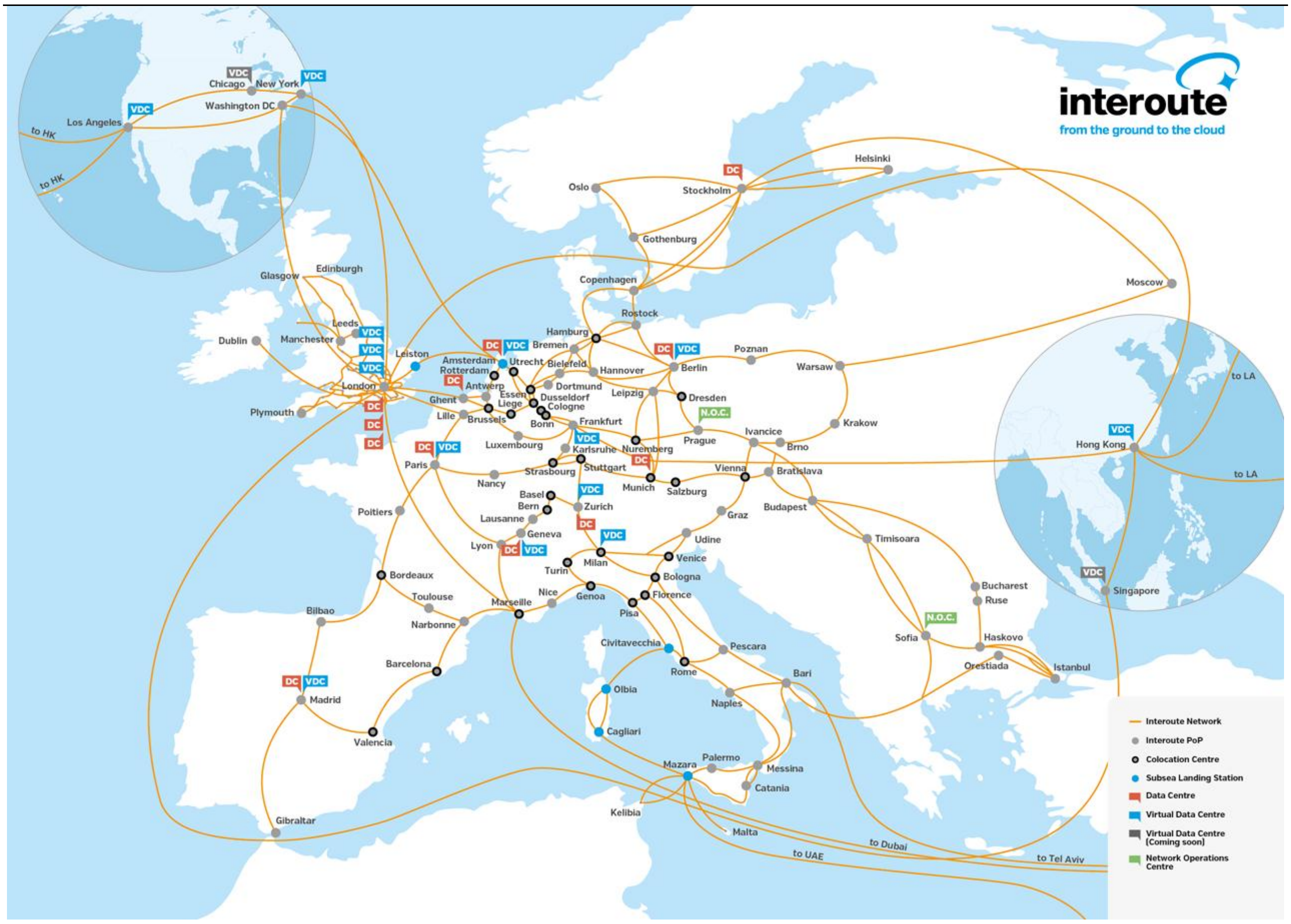

Figure 6-1: The actual footprint of Interoute's Unified ICT Platform.

Leveraging on the success of this Cloud service offer, mostly based on IaaS, Interoute has observed a rapid increase of intra- and inter-DC IP traffic with a +70% Year-over-Year traffic increase rate also confirmed by public reports and forecasts. In this scenario efficiency, flexibility and programmability of the DC and Cloud infrastructures are key elements of a business proposition, with major emphasis on:

- Delivery, operation and maintenance procedures for cloud applications and virtualized services capable to respond to the request for dynamicity and elasticity of customers' application scenarios
- Possibility to interface cloud services with legacy networks (customer VPNs) or other public clouds (e.g. Amazon WS).

Interoute VDC2.0 APIs has been designed exactly for these purposes allowing for IaaS programmability, options for deploying storage, disaster recovery, etc.

The second point above pinpoints a very important fact of the current cloud market. While big and small companies continue to invest in their own IT infrastructures to some extent, the demand for public cloud services rapidly increases and so does the number of players and offers in the field. Therefore, in the context of CYCLONE we set up to offer a product that not only works flawlessly with Interoute's cloud infrastructure, but also allows the end user to target their existing resources as well as other public clouds. The core component of CYCLONE SlipStream, SixSq's flagship product, allows automated deployment of applications in multiple clouds and provides the means for creating sector specific "app stores". SlipStream in conjunction with StratusLab, an open source IaaS cloud distribution, power the SixSq NuvlaBox providing a personal cloud computing platform that can scale out to other cloud infrastructures.

Enhanced, policy-based management of running multi-cloud applications as foreseen in CYCLONE will make the SixSq products even more attractive to end users as a platform for managing their production services, by reducing the effort required to maintain them and giving more control on optimizing their deployment. At the same time, the inclusion of networking, the advanced policy-based application optimization, and the "app store" enhanced with CYCLONE components make SlipStream and the surrounding tools extremely attractive for cloud service providers as an internal tool for managing their resources and providing their users with new PaaS and SaaS offerings.

Apart from facilities and equipment, virtualization is playing a disruptive role in innovating the DC market landscape. Virtualization technologies and software have concretely allowed data centre operators to reduce the operational and capital expenditure associated to managing data centre environments. The automation of resource provisioning and monitoring, service recovery, intra- and inter-DC workload migration or network management is leading to a more efficient allocation of virtual instances among geographically dispersed data centers.

Regarding the Cloud data center market it's worth mentioning the CISCO Global Cloud Index (CGI) [10] stating that the data center IP traffic has reached the zettabyte era and over the 2014-2019 period this value will triple up to 10.4 ZB annually. In particular the cloud data center traffic will rapidly grow until reaching a quadruple value over the forecast period and will represent more than 80 percent of all data center traffic by 2019 (See Figure 6-2). An important enabler for this notable expansion is the increasing of data center virtualization which provides flexibility, efficiency and easy deployment of services and applications. According to the market trends analysis, by 2019, more than four-fifths of all workloads will be processed in the cloud.

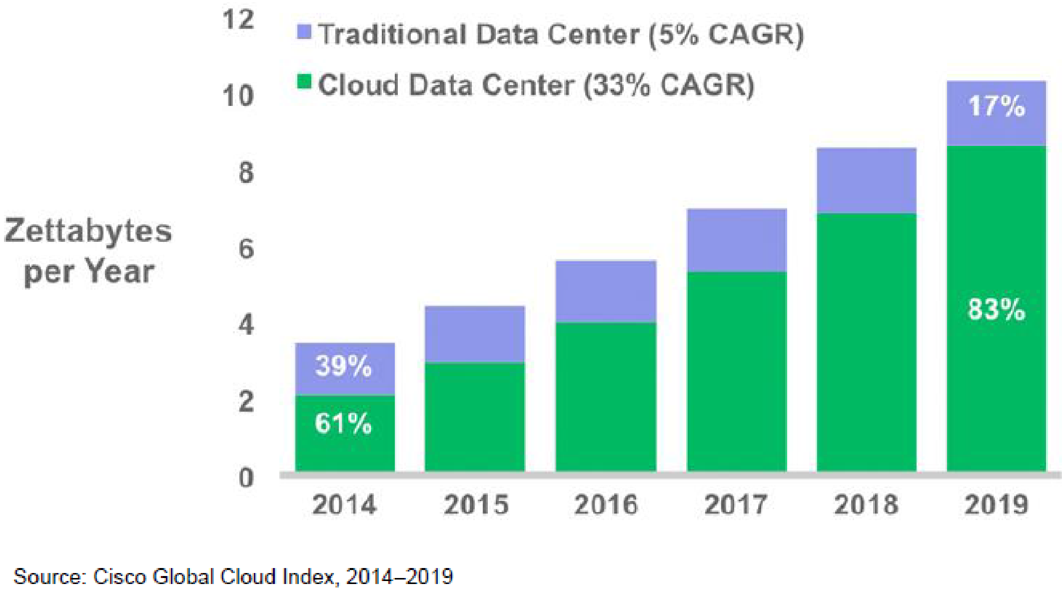

Figure 6-2: DC traffic growth rate vs. the Cloud DC traffic growth rate.

The cloud traffic growth is being and will be promoted by the rapid adoption and migration to cloud architecture data centers which can handle significantly higher traffic and computational loads due to the use of virtualization support, real-time provisioning and resiliency ensuring better performance, higher capacity and throughput.

The importance of virtualization is confirmed also by Gartners' 2015 Magic Quadrant for Data Center Networking [11], where is reported that SDN-based products are becoming more visible and available, although wide-scale adoption has yet to occur (See Figure 6-3).

Figure 6-3: The Magic Quadrant for Cloud-Enabled Managed Hosting, related to the European market.

The topics targeted by CYCLONE project are closely related to the market needs mentioned before, since the management and federation of virtual resources distributed among different and multi-tenant VDCs has a strong relationship with all the emerging technologies involved in the landscape of cloud services.

Conclusions

This document provides a concise summary of selected use cases, explain how they have been deployed within CYCLONE, and provide formal feedback on the CYCLONE software. The document focuses on the analysis of use cases presented in the DOA, derives functional requirements to be considered by the CYCLONE federated framework, and analyses their deployment and testing on the CYCLONE testbed.

The project identified two flagship applications: an academic cloud platform (and its associated services) supporting bioinformatics research and a commercial platform for future energy management. Each provides 3 “micro” use cases (6 in total), highlighting different needs. These use cases have been described and reveal several kinds of workflows.

The bioinformatics use cases identify the following objectives specifically:

- Integration of a common federated identity scheme (e.g. EduGAIN) to ease the federation of cloud infrastructures
- Complete integration of advanced features of OpenNaaS
- Providing a policy-based cloud brokering engine
- Providing isolated network service delivery
- Network resource allocation
- Multi-cloud extension of location-based access control and distributed logging

For the Energy use cases, security plays a major role in the system:

- Data access control and access documentation
- Data security: Availability – 24/7, authenticity, secured storage
- Network security: communication security, transmission security

From both Bioinformatics and Energy use cases a number of identified features are required to definitely enhance the cloud infrastructures’ current performance, mainly in terms of security, management, automation of processes and dynamicity. That analysis conducted to the definition of 34 common requirements and 26 user stories. From these two lists, we defined a list of 36 implementation tasks, for one or more components, and put them in a roadmap for the first period of the project with milestones at M10, M14 and M18. Some of them have been done to achieve the following results, the other ones are in progress or planned.

Two bioinformatics use cases were already deployed, and linked with test scenarios for the three use cases. These developments consist first of the UC1 Securing human biomedical data deployment, consisting of a VM integrating security features. Afterwards, we deployed the UC2 Cloud virtual pipeline for microbial genomes analysis, a complex application requiring the coordinated deployment of several VMs. The UC1 benefits from enhanced security features and the appliance was extended with the integration of the Federation Provider. The UC2 Cloud virtual pipeline for microbial genomes analysis was extended with SlipStream deployment recipes to solve the requirement of a complex application deployment with several virtual machines in one deployment.

New use cases are foreseen for the next period of the project. They are related to general technical requirements such as an attribute-based encryption system or automated placement, but still driven by the applications requirements from the Energy and bioinformatics domains. However other use cases will also be considered.

Finally, a brief overview on the actual Cloud market and an outlook on the foreseeable mid-term and long-term period based on market trends analysis concluded that the topics targeted by CYCLONE project are closely related to the market needs, since the management and federation of virtual resources distributed among different and multi-tenant VDCs has a strong relationship with all the emerging technologies involved in the landscape of cloud services.

References

[1] http://www.elixir-europe.org/about/data-deluge

[2] http://www.france-bioinformatique.fr

[3] European Nucleotide Archive – ENA, http://www.ebi.ac.uk/ena

[4] Ensembl Genomes public human datasets, http://www.ensembl.org/Homo_sapiens/Info/Index

[5] http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/

[6] http://blast.ncbi.nlm.nih.gov/Blast.cgi

[7] http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE=Proteins&PROGRAM=blastp&RUN_PSIBLAST=on

[8] https://www.ebi.ac.uk/interpro/interproscan.html

[9] Integrative Genomics Viewer, https://www.broadinstitute.org/igv

[10] http://marketplace.france-bioinformatique.fr:8081/metadata/IqDPi0GvoJOhFo00BqoHlqIVH40/Christophe.BLANCHET@france-bioinformatique.fr/2015-09-25T10:22:30Z

[11] https://marketplace.stratuslab.eu/marketplace/metadata/KpuPQj5xVmBbMVf3rfflJXRLlh2/Christophe.BLANCHET@france-bioinformatique.fr/2015-09-17T13:22:15Z

7. Abbreviations and Definitions

7.1. Definitions

No specific definitions are introduced in this document.

7.2. Abbreviations

B2B	Business to Business
DC	Data Center
DER	Distributed Energy Resources
E2E	End to End
ENA	European Nucleotide Archive
IaaS	Infrastructure-as-a-Service
ICT	Internet and Communication Technology
IPR	Intellectual Property Rights
IT	Information Technology
MaaS	Metal as a Service
NaaS	Network-as-a-Service
Net-HAL	Network Hardware Abstraction Layer
NFV	Network Function Virtualization
NGS	Next-Generation Sequencing
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
SP	Service Provider
TC	Technical Coordinator
TCTP	Trusted Cloud Transfer Protocol
VPP	Virtual Power Plant
WP	Work Package

<END OF DOCUMENT>