Usecases Hackathon 2016

Christophe Blanchet, CNRS

IFB-core CNRS UMS3601

Gif-sur-Yvette, FR, November 15t-16th, 2016

H2020-ICT-644925 – CYCLONE

Complete Dynamic Multi-cloud Application Management

Agenda

Day 1, 15 November 2016

9:30 Welcome coffee

10:00-12:30

General introduction (C. Blanchet, 15')

Presentation of applications (10' each)

- T. Lacroix, new developments of CYCLONE UC2
- P. Veber, CYCLONE UC11
- S. Blanck (in link with CYCLONE UC13)
- A. Josso, App PathoTRACk
- S. Perrin, CYCLONE UC3

Presentation of CYCLONE components (20' each)

- Slipstream, C. Loomis
- Federation Proxy and tools, M. Slawik and Ö. Malik
- CNSMO, I. Canyameres
- New developments of IFB portal, B. Brancotte

12:30-14:00 Lunch

14:00-18:00 Hackathon

Day 2, 16 November 2016

9:00-12:30 Hackathon

12:00-13:30 Lunch

13:30-16:00 Hackathon

CYCLONE Usecases

- Two flagship applications:
  - an academic cloud platform and associated services for bioinformatics research (Usecases 1-3)
  - a commercial deployment for future energy management (Usecase 4)
- These applications guide the initial development of the tools.
- Additional use cases, selected during the course of the project:
  - highlight missing, critical features
  - guide the further evolution of the software.

Use Cases: Initial Ones

ID	Title	Domain	Key Tech. Areas *	Resp.
UC1	Securing human biomedical data	Bioinformatics	**1, 2, 3**	CNRS
UC2	Cloud virtual pipeline for microbial genomes analysis	Bioinformatics	**1, 2, 4**	CNRS
UC3	Live remote cloud processing of sequencing data	Bioinformatics	**2, 4**	CNRS
UC4	Virtual Power Plant	Energy	**2, 4**	QSC

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

Use Cases: New Ones

ID	Title	Domain	Key Tech. Areas *	Resp.
UC5	Internet of Services Lab (IoSL)	Teaching	**App Deploy, 3**	TUB
UC6	ENTRANCE	App mgmt.	**1, 3**	TUB
UC7	Open Scientific Data	Data mgmt (Earth obs.)	**2, 1**	SixSq**
UC8	Benchmark Driven Placement	Bioinformatics	**2**	SixSq**
UC9	On-Demand Bandwidth	Network Provisioning (Géant)	**4**	SixSq**
UC10	Smart Utility 4.0	Energy	**2,3,4**	QSC
UC11	Assembling genomes from sequencing reads	Bioinformatics	**2**	IFB**
UC12	Metagenomics	Bioinformatics	**2, 3, 4**	IFB**
UC13	Shared environment between cloud Galaxy portals	Bioinformatics	**1, 2**	IFB
UC14	WebRTC video-conference solution	Network	**3,4**	I2CAT

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

**** in coll. with 3rd party partners**

Cyclone

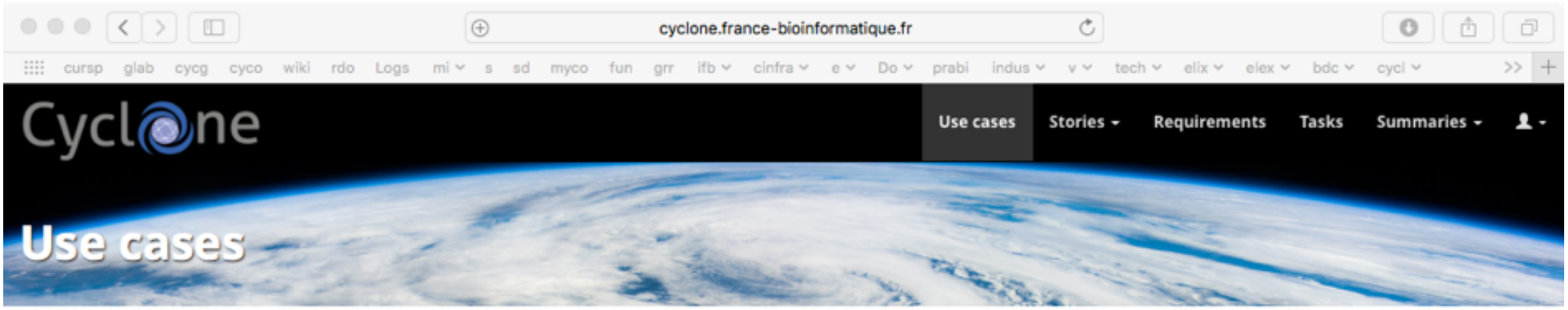
Use cases @ a Glance

- 14 use cases
- 93 user stories ; 48 common requirements
- 39 implementation tasks (for Year1 UCs)

CYCLONE extension of Usecases	UC1	UC2	UC3	UC4
SlipStream	✔	✔	✔	✔
Complex App deployment		✔	✔	✔
Multicloud deployment				✔
Federation Proxy	✔	✔	✔	
Web authentication	✔	✔		
SSH authentication		✔	✔	
CNSMO-VPN		✔	✔	✔
CNSMO-FW				✔

Use Cases Portal

https://cyclone.france-bioinformatique.fr/usecases



UC id	Title	Description
UC 1	Securing human biomedical data	Thanks to the steady drop of genome sequencing technology costs (NGS), an increasing number of clinicians are including biological results obtained with these technologies in their day-to-day diagnosis practice. Today, much genomics analyses are realized on the exome, which is the expressed part (5%) of the genome (...)
UC 2	Cloud virtual pipeline for microbial genomes analysis	In the post-NGS area, sequencing bacterial genomes is very cheap (few hundreds €). Most of the time, users are no longer content to analyse a single genome; they want to compare large collections of related genomes (strains). This entails that biologists have to pay too much attention and dedicate their time to sequence the genomes, instead of thoroughly analysing the genomic data (...)
UC 3	Live remote cloud processing of sequencing data	Bioinformatics deals with the collection and efficient analysis of biological data, particularly genomic information from DNA sequencers. The terabytes of raw data, produced by the sequencers for each run, require significant computing resources for analysis that may not be available locally. Especially for small sequencing facilities, the analysis of the data is indeed the bottleneck (...)
UC 4	Virtual Power Plant	The main idea of a Virtual Power Plant (VPP) is to integrate distributed energy resources (DER) for the generation of renewable energy and to combine them to one reliable power plant. In the following description the DER Owner, DER Operator and the VPP Operator have pivotal roles regarding the production of energy (...)
UC 5	Internet of Services Lab (IoSL)	"Internet of Services Lab"1 (IoSL) is a SNET/TUB teaching project where students work a semester in groups of three to six. They most often implement software related to numerous research projects and topics: Cloud, Mobile, Social, and Ubiquitous Computing. This type of teaching project is quite popular in other computer science related study paths at TUB as well as other universities (...)
UC 6	Multi-cloud deployment and federated identity integration of ENtRANCe components	ENtRANCe (ExteNsible and geneRic AuthorizatioN for Cloud resources and personal files) has been chosen as one of the most innovative IT projects in 2014 to be funded within the Software Campus initiative. It aims at developing a user-centric and enhanced access rights management for a secure and practical privacy aware data exchange in distributed networks (...)
UC 7	Open Scientific Data	Publicly funded research creates an immense amount of data that has academic and commercial value both inside and outside of the original research domain. The taxpayers through funding agencies are increasing the pressure to make these datasets generally available through "open data" programs. A major (...)
UC 8	Benchmark Driven Placement	A significant part of the recent design discussion for the CYCLONE brokering and matchmaking components dealt with benchmarks, both general benchmarks and application-specific benchmarks. At the recent UCC 2015 conference, there was a presentation from a bioinformatics group in Cardiff (...)

Use Cases Portal : Useful to manage all elements

Cyclone

Use cases | Stories | Requirements | Tasks | Summaries

Use cases and associated requirements

UC id	Title	Reqs.	Stories	Workflow	Plan of deployment	Test scenario
UC 1	Securing human biomedical data	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 21, 26, 32, 43	1, 2, 3, 4, 5, 6, 7	✓	✓	✓
UC 2	Cloud virtual pipeline for microbial genomes analysis	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 32, 34	8, 9, 10, 11, 12, 13, 14, 15, 16	✓	✓	✓
UC 3	Live remote cloud processing of sequencing data	1, 2, 3, 4, 5, 9, 10, 11, 12, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 34	17, 18, 19, 20, 21, 22, 23, 24, 25, 26	✓	✗	✗
UC 4	Virtual Power Plant	8, 10, 13, 27, 33	27, 28, 29, 30, 31, 32	✓	✓	✓
UC 5	Internet of Services Lab (IoSL)	4, 6, 16, 17, 20, 22, 36, 37	59, 60, 61, 62, 63, 64, 65	✓	✓	✓
UC 6	Multi-cloud deployment and federated identity integration of ENtRANCe components	4, 5, 6, 8, 16, 17, 20, 22, 23, 31, 32, 36, 37, 41, 42	66, 67, 68, 69, 70	✓	✓	✓
UC 7	Open Scientific Data		76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86	✓	✓	✓
UC 8	Benchmark Driven Placement		87, 88, 89, 90, 91, 92	✗	✗	✗
UC 9	On-Demand Bandwidth		93	✓	✓	✓
UC 10	Smart Utility 4.0	10, 27, 47, 48	33, 34, 35, 36, 37, 38	✓	✓	✓
UC 11	Assembling genomes from sequencing	1, 2, 4, 5, 6, 8, 22, 27, 35, 36, 37, 42	71, 72, 73	✓	✓	✓

Use Cases Deployed on the Testbed

nuvla.cyclone-project.eu

Nuvla. Dashboard | App Store | Workspace | Service Catalog | Support | Help | cblanchet

/ cyclone / 4484 | New project | New component | New application

CycLoneDACI	Security Services (Auth, Context, Token services and more)	fturkmen	3611
Distributed_Logging		ilke	2496
Energy-UC_VPP	Energy Use Case Virtual Power Plant	qsmart1	5190
Federation_Provider		ilke	4462
IFB_UCs_Integration_With_CNSMO		mbedri	3810
IRT-WebServer	Deployed by Domenico	domenico	3689
OpenStack_Base_Images		mbedri	3608
StratusLab_Base_Images		ilke	2750
UC1_Securing_human_biomedical_data		bryanbrancotte	5286
UC2_Microbial_genomes_analysis		bryanbrancotte	5310
UC3_NGS_mapping		sandrineperrin	5267
mapping	cluster to run mapping step in workflow to analyse	edugain_a18cecb36339a0396c814d591003b8fa4671dfa9	4448

Support

Cyclone

IFB – French Institute of Bionformatics

http://www.france-bioinformatique.fr

CNRS UMS3601. Avenue de la Terrasse,
Bât 21. 91190 Gif-sur-Yvette

- **Distributed infrastructure**
  - **National hub** : IFB-core
  - **36 platforms** (e.g. MIGALE, INCA-SLC and PRABI-AMSB)
- **IT resources**
  - 15,000 cores - 9 PB
  - 4 running clouds

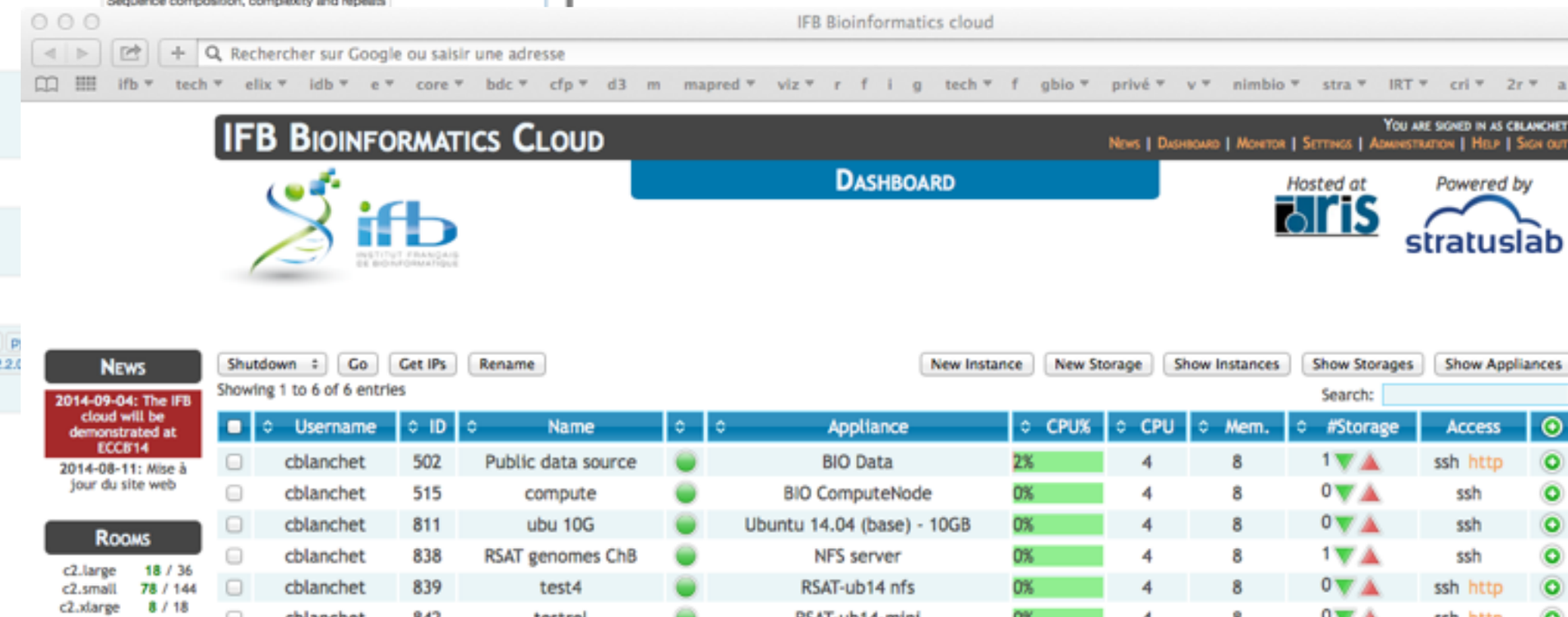
- **Mission** : to make available core bioinformatics resources to the life science research community (academic and private).
  - To provide support for national biology programs
  - To provide an IT infrastructure devoted to management and analysis of biological data
  - To act as a middleman between the life science community and the bioinformatics/computer science research community
- IFB is the **ELIXIR-FR Node**
  - Especially involved in the construction of the Compute platform based on a cloud federation

IFB's Bioinformatics services on Cloud

- Pilote infrastructure running since 2014
  - 48 bioinformatics appliances already available
  - Scientific usage: 430+ users, 12,000+ VMs
  - Training (2014-16): 13 cloud sessions, 4 scientific trainings, 5 Master courses

Services registry

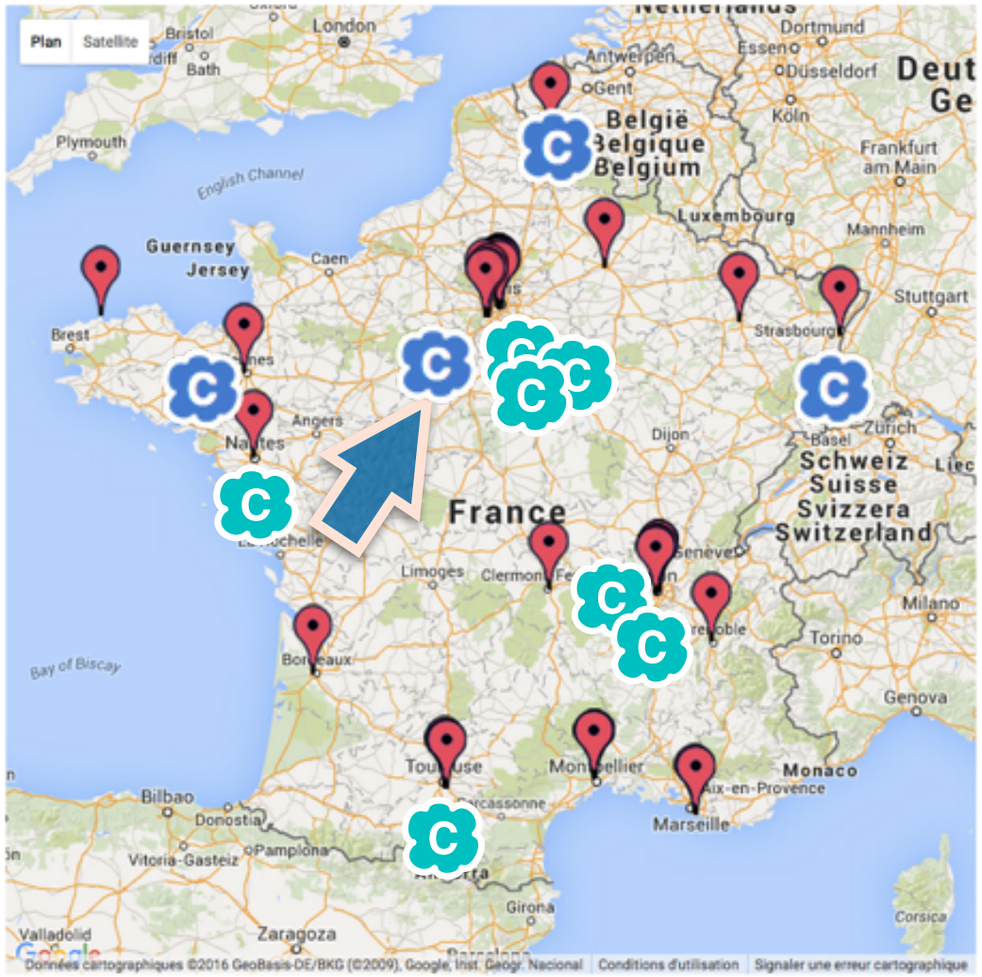
Cloud dashboard



Cyclone

Towards a multi-cloud infrastructure for Life Science

Current clouds Future clouds



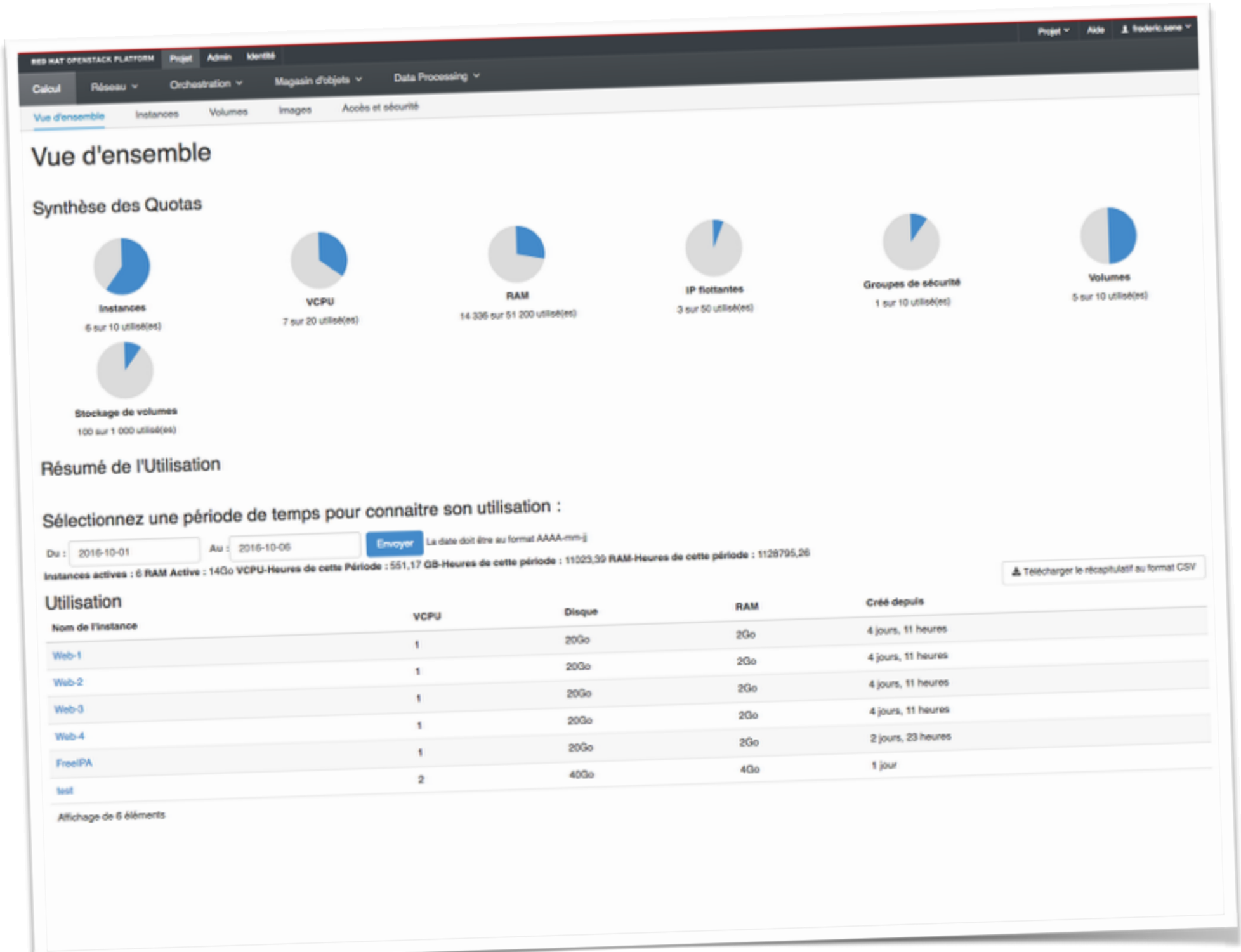
- **Federate IFB's cloud**
  - The four running clouds of IFB
  - Seven future clouds to be deployed on 2016-2017
- **Requirements**
  - common identities and authorizations management based on community standards (EduGAIN)
  - interoperability of virtual images (VM/container)
  - tools for multi-cloud deployment (e.g. the SlipStream/ NuvLa broker)
  - network management, security and propagation over several datacenters
- **Solutions expected** from
  - ELIXIR/EXCELERATE and CYCLONE projects

IFB-core hub	Compute #cores	Storage #TB	RAM #GB	Largest VM	Technology	Location
2014-08	**200**	**50**	**2,000**	**20c 256GB**	**StratusLab**	**CNRS-IDRIS, Paris**
2016-11	*5,000*	*1,000*	*40,800*	*128c 3TB*	*OpenStack*	*CNRS-IDRIS, Paris*
2017	*10,000*	*2,000+*			*OpenStack*	*CNRS-IDRIS, Paris*

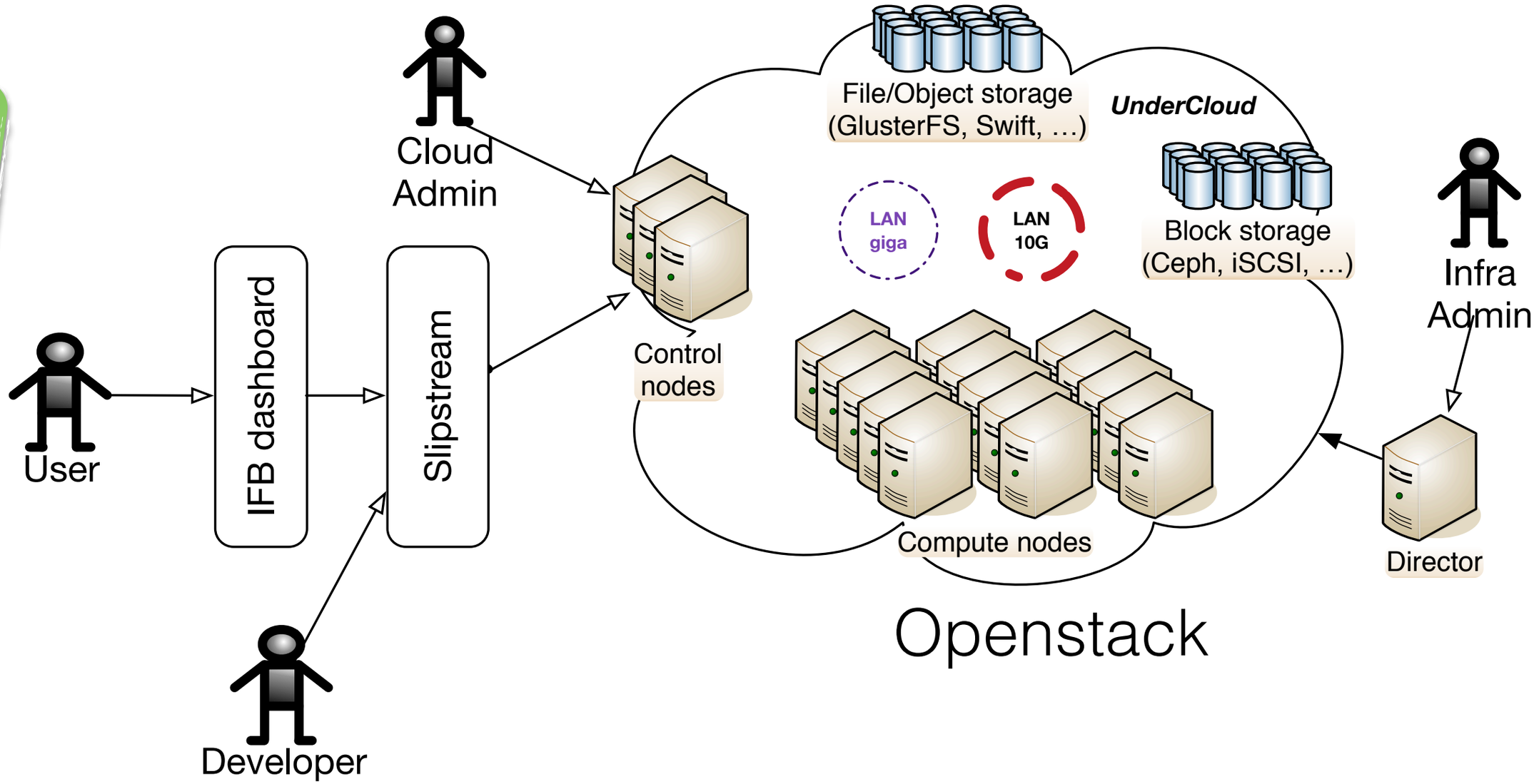
Cyclone

Lamarck

Pilote RHOSP 9.0

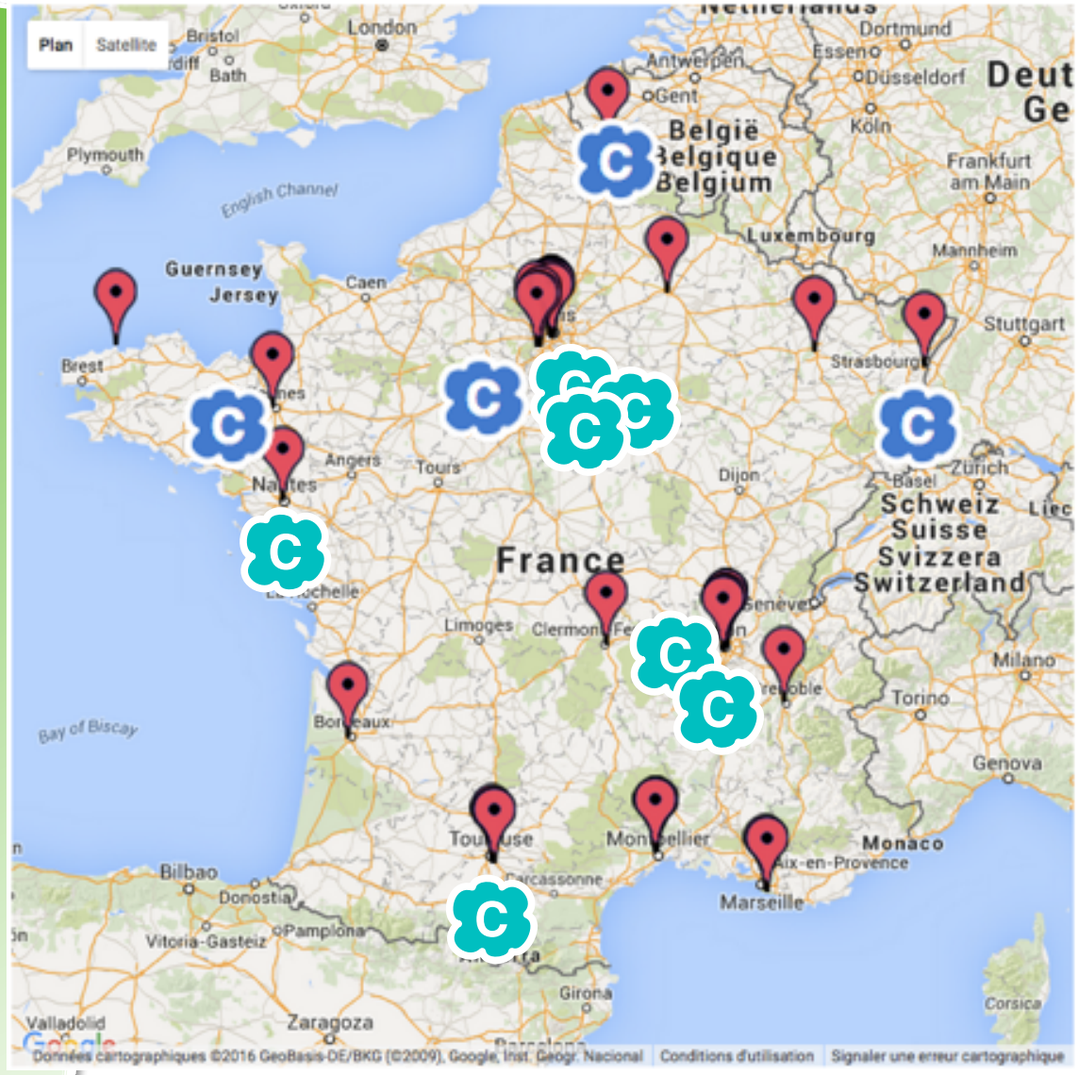


Architecture IFB



Biosphere

Current clouds Future clouds



The four running clouds of IFB

- IFB-core
- Genouest
- IPHC-BISTRO
- Bilille

Seven future clouds to be deployed on 2016-2017

- PRABI-LBBE
- Genotoul
- RPBS
- eBio
- BiRD
- Pasteur
- PSMN

Requirements for a Federation

Common identities and authorizations management based on community standards (EduGAIN)

Interoperability of virtual images (VM/container)

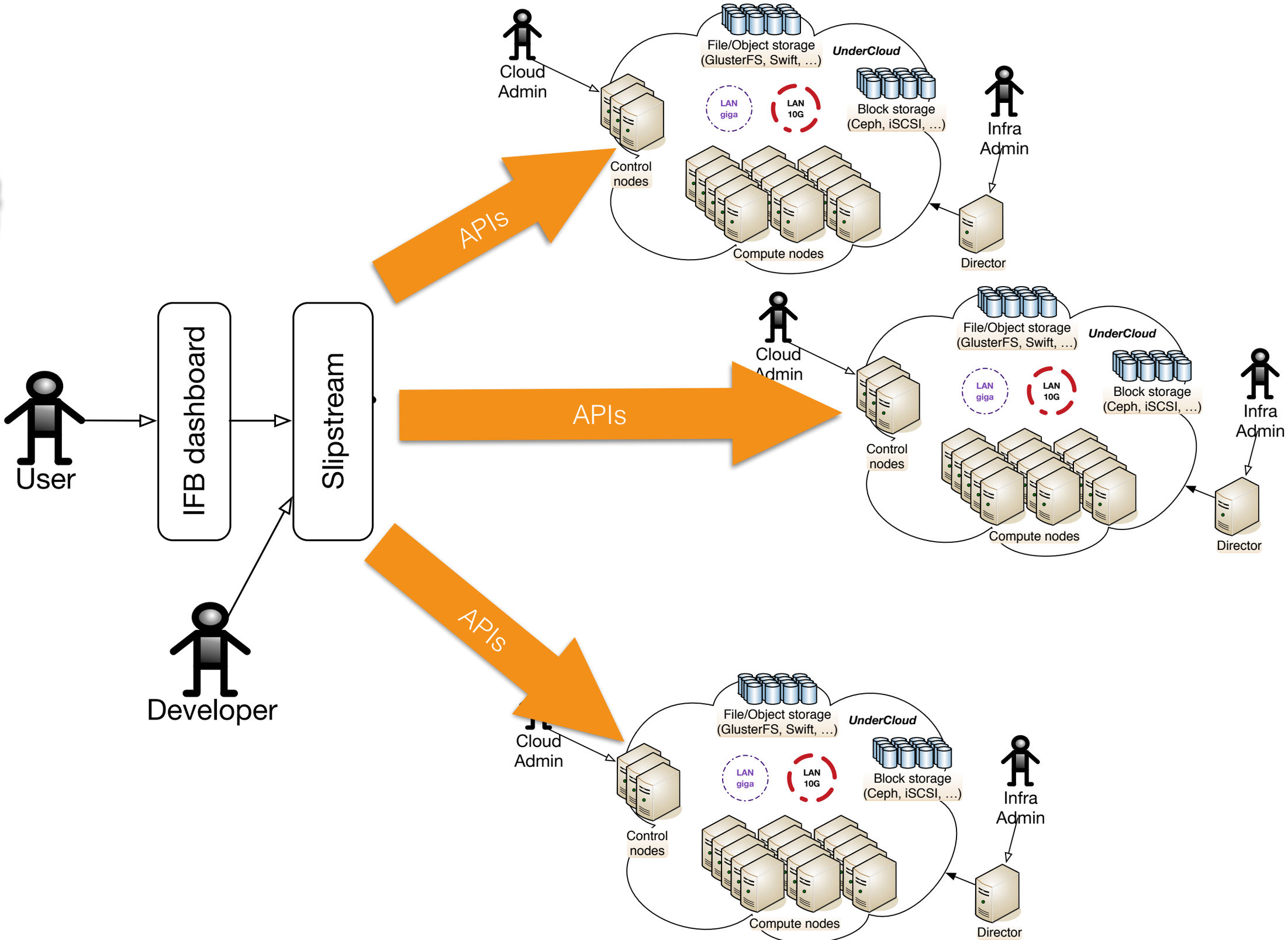
- standard format and best practices
- Registry of images

Tools for multi-cloud deployment (e.g. SlipStream/NuvLa broker)

Network management, security and propagation over several datacenters

Solutions expected from ELIXIR/EXCELERATE and CYCLONE projects

Biosphère : Architecture



Agenda

Day 1, 15 November 2016

9:30 Welcome coffee

10:00-12:30

General introduction (C. Blanchet, 15')

Presentation of applications (10' each)

- T. Lacroix, new developments of CYCLONE UC2
- P. Veber, CYCLONE UC11
- S. Blanck (in link with CYCLONE UC13)
- A. Josso, App PathoTRACk
- S. Perrin, CYCLONE UC3

Presentation of CYCLONE components (20' each)

- Slipstream, C. Loomis
- Federation Proxy and tools, M. Slawik and Ö. Malik
- CNSMO, I. Canyameres
- New developments of IFB portal, B. Brancotte

12:30-14:00 Lunch

14:00-18:00 Hackathon

Issues

Check

Day 2, 16 November 2016

Check

9:00-12:30 Hackathon

12:00-13:30 Lunch

Demo

13:30-16:00 Hackathon

Cyclone

Goals

- 5 applications fully integrated with the Cyclone components
- Validated recipes of deployment
- A validated 3-tier infrastructure with Cloudweb, Slipstream and Openstack
- A clearer understanding between bioinformatics applications and Cyclone components developers