

e-Infrastructure pour les sciences du vivant



Christophe Blanchet

**Institut Français de Bioinformatique - IFB
French Institute of Bioinformatics - ELIXIR-FR**
CNRS UMS3601 - Gif-sur-Yvette - FRANCE

Institut Français de Bioinformatique

Mission générale de l'IFB

- fournir des ressources de base en bioinformatique à la communauté des sciences de la vie

Infrastructure nationale de service en bioinformatique

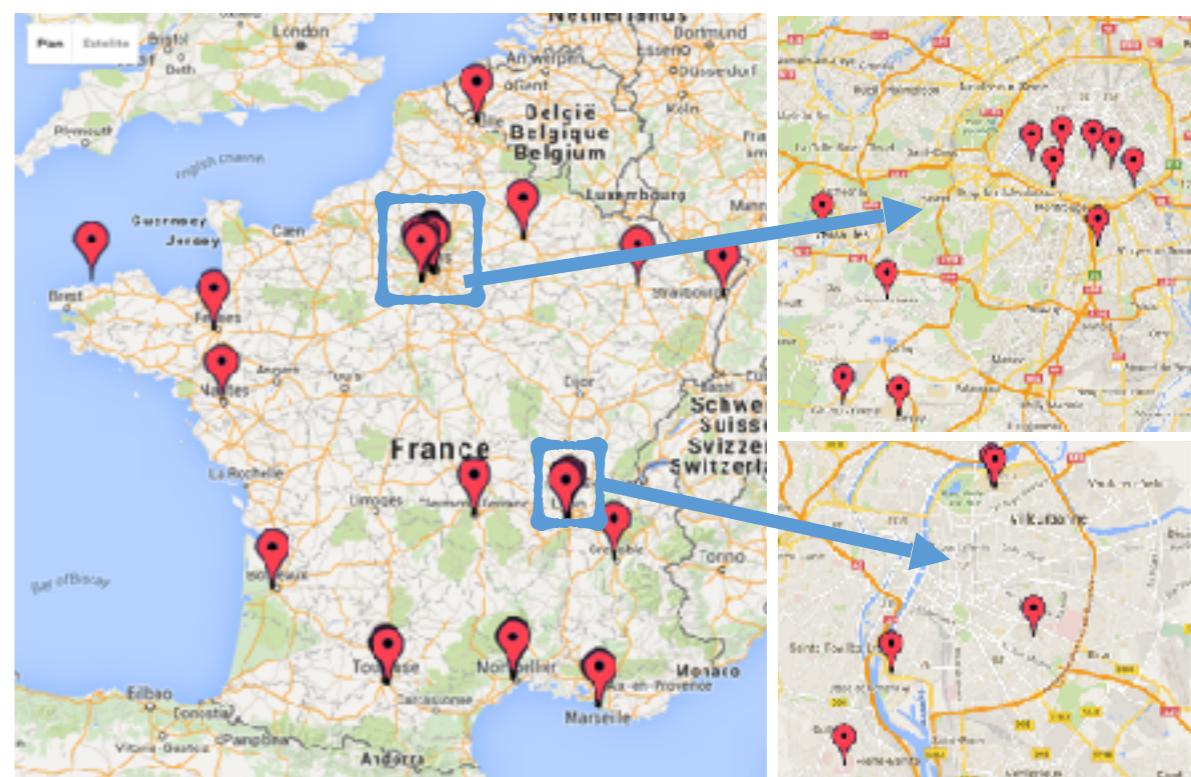
- **Données** : Fournir un accès à des collections de données spécialisées à haute valeur ajoutée issues de l'expertise du laboratoire d'accueil
- **Outils** : Développer et mettre à disposition des outils et services en lignes pour analyser les données correspondant à l'expertise scientifique du laboratoire d'accueil
- **Appui aux projets scientifiques** et hébergement sur une infrastructure informatique
- **Formations**
- **Infrastructure** : Mettre à disposition une infrastructure informatique dédiée à l'analyse des données des sciences du vivant (matériel, données, outils)

Organisation

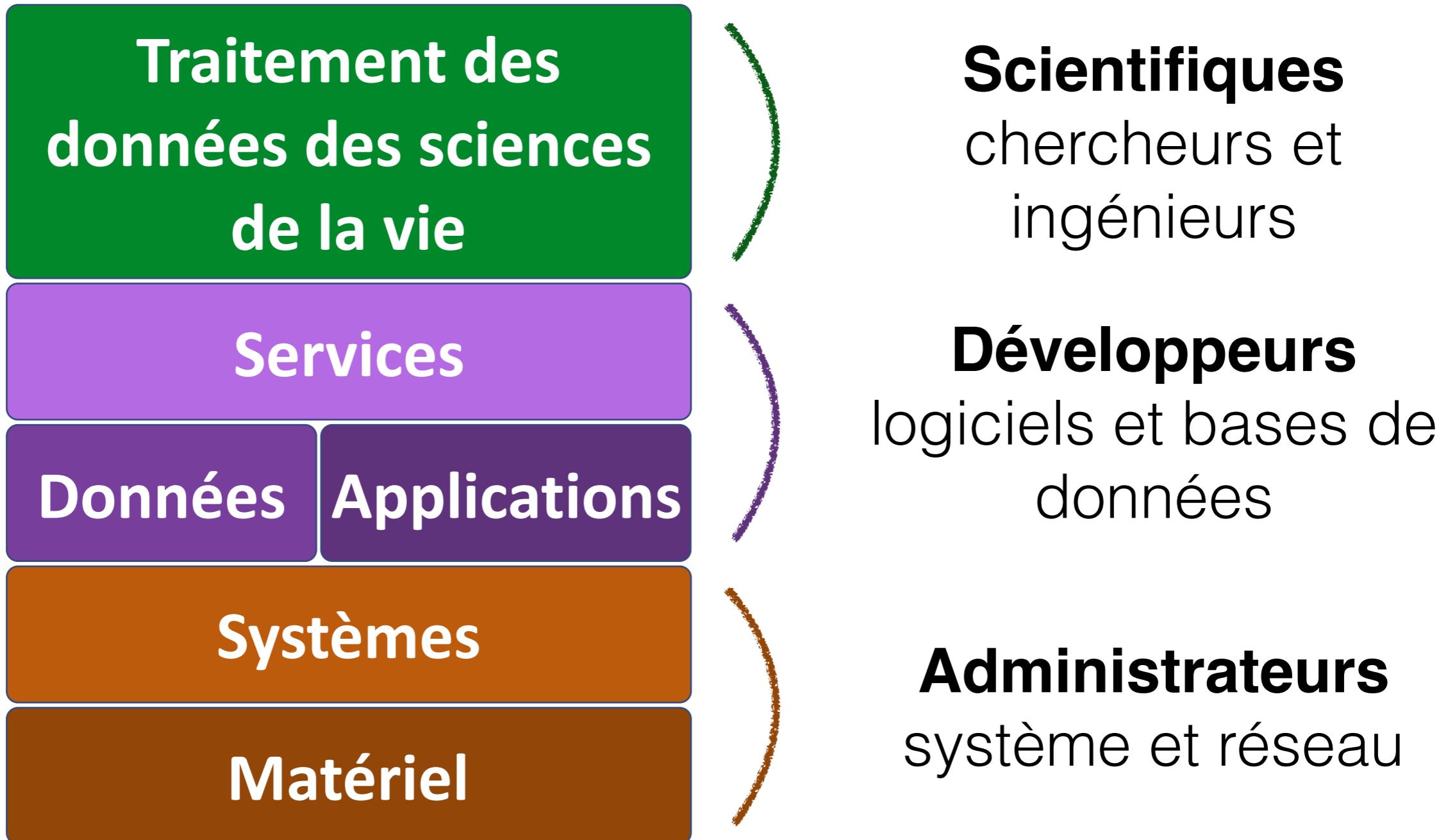
- **36 plates-formes** (6 pôles régionaux)
- **Nœud national IFB-core**, impulser et coordonner la mise en place de l'infrastructure.

Membres et ressources

- 420 ingénieurs et chercheurs (290 ETP)
- Info SI : 17 000+ coeurs, 10 Po



e-Infrastructure



Applications

Données

Plateforme	Outils	Bases de données
AuBI	23	4
GenOuest	19	10
MIGALE	18	2
RPBS	13	2
ATGC	11	0
Bistro	11	10
EBIO	9	7
Pasteur	8	6
Genotoul-bilille	8	0
CBiB	7	1
ABiMS	7	2
TAGC-BCF	6	0
PRABI-Doua	6	2
IGS	5	3
PRABI-Gerland	5	1
URGI	4	4
	4	2

Développeurs

Plateforme	Outils	Bases de données
BIMEPS	4	2
IGR	3	0
MBI	3	1
BiRD	3	0
Curie	2	0
SIGENAE	2	0
GenAtlas	1	0
MicroScope	1	0
PRABI-G	1	2
ISfinder	1	1
Orphanet	1	2
GMGF-GBT	1	3
INCa-SLC	0	1
MMP	0	0
Prabi Lyon-Sud	0	0
PRABI-AMSB	0	0
CAZy	0	1
IMGT	0	0
South Green	0	5
<i>Total</i>	<i>187</i>	<i>74</i>

Outils et bases de données développés par les PF

Developpeurs (2)

Services

This screenshot shows a bioinformatics pipeline interface. At the top, there's a 'DATA STATUS' section with a 'Modify' button and a 'DATA INPUT/FILE' section with a 'Modify' button. Below these are 'Advanced Numerical Options' and a 'COMPUTE & MODEL' section with buttons for 'Compute', 'Compute & Model', 'Compute & Visualize', and 'Compute & Model & Visualize'. The main area displays 'Protein sequence Coverage 100.0% of exp. 600nm - 604 nm' and a 'Peptide-Match Results' table:

SL	Sequence	Charge	Mass Error	Confidence
1	N-Q-D-N-H-L-G-T-E	3	0.08	90
2	N-Q-D-N-H-L-G-T-E	3	1.09	90

```
[root@vm0007 ~]# fastadb -U tutu.fas oneSeq.fasta /ifb/databases/uniprot/uniprot_2014/flat/uniprot_sprot.fasta
# fasta36 -O tutu.fas oneSeq.fasta /ifb/databases/uniprot/uniprot_2014/flat/uniprot_sprot.fasta
FASTA searches a protein or DNA sequence data bank
version 36.3.7a Jan, 2015(preload9)
Please cite:
W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

cannot open library /ifb/databases/uniprot/uniprot_2014/flat/uniprot_sprot.fasta
Query: oneSeq.fasta
  1> ooo-splP0ACP1IFRUR_ECOLI FruCule repressor OS-Escherichia coli (strain K12) GN-TruR PE-1 SV-1 - 334 aa
Library: /ifb/databases/uniprot/uniprot_2014/flat/uniprot_sprot.fasta
  0 residues in 0 sequences

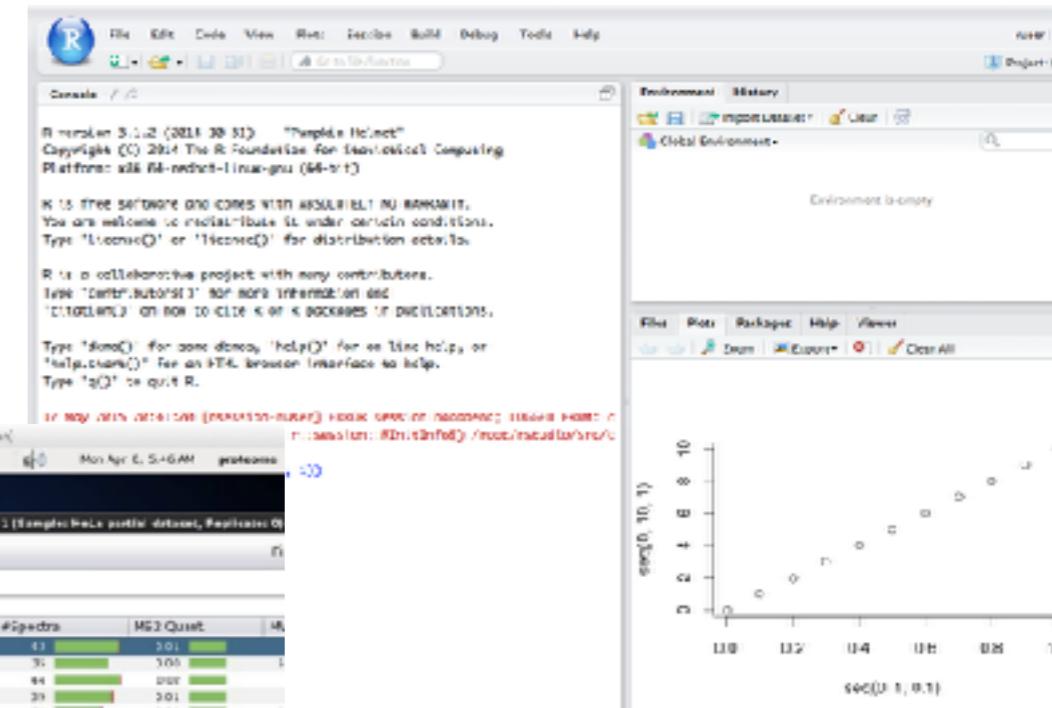
Statistics: Altschul/Gish params: m0: 334 Lambda: 0.158 K: 0.019 H: 0.100
  statistics sampled from 0 (0) to 0 sequences
Algorithm: FASTA (3.0 Nov 2011) [optimized]
Parameters: BL50 matrix (15:-5), open/ext: -10/-2
  ktup: 2, E-join: 1 (-ran), E-opt: 0.2 (-man), width: 36
Scan time: 0.000

334 residues in 1 query sequences
0 residues in 0 library sequences
Tcomplib [36.3.7a Jan, 2015(preload9)] (2 proc)
start: Thu Jul 2 09:27:23 2015 done: Thu Jul 2 09:27:23 2015
Total Scan time: 0.000 Total: Display time: 0.000

Function used was FASTA [36.3.7a Jan, 2015(preload9)]
```

This screenshot shows a peptide viewer application. It has tabs for 'Peptides (100000)' and 'Spectra & Parameters (100 - 600-600)'. The 'Peptides (100000)' tab displays a table of peptides with columns: SL, R, Accession, Description, Coverage, #Peptides, #Spectra, MS2 Quest, and %.

SL	R	Accession	Description	Coverage	#Peptides	#Spectra	MS2 Quest	%
1	*	P1002	COP26_HUMAN	100.0	1	1	201	100
2	*	Q05146	PFC_HUMAN	100.0	1	1	100	100
3	*	Q8B1M4	DETH_HUMAN	100.0	1	1	100	100
4	*	P08261	ACTD_HUMAN	100.0	1	1	201	100
5	*	P01313	F1MA_HUMAN	100.0	1	1	200	100
6	*	P08610	VIM8_HUMAN	100.0	1	1	201	100



This screenshot shows a Galaxy cloud instance. It features a 'Galaxy' logo and a 'Welcome' message: 'Welcome to your Galaxy instance based on the cloud of the French Institute of Bioinformatics - IFB'. Below this is a 'Usage' section and a 'History' list. The history list includes entries for '245.1 MB', 'R-DNA-seq data 1-Raw', 'Data', 'P-1407-04-001-10000', 'Data', and 'L-https://www.galaxyifb.fr/.../galaxy/09996812_1M_2fa'.

Services aux développeurs (64 pers.)

Services

Applications

Données

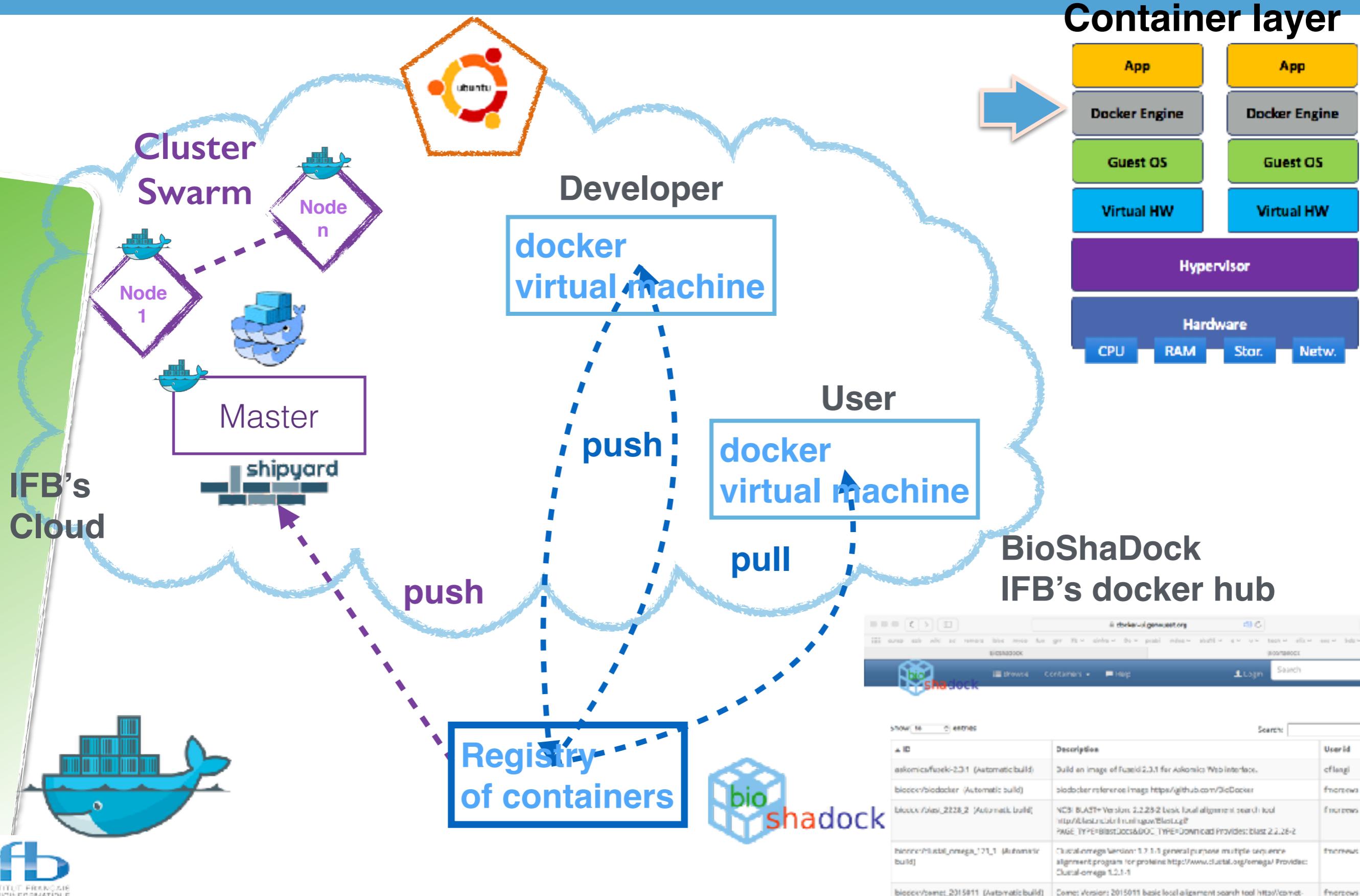
Infrastructure

- Accès au cloud IFB;
- Formations à son utilisation
 - ★ IBI-2 Utilisation avancée du Cloud IFB
 - ★ IBI-3 Développement de machines virtuelles modèles
- Équipe support IFB-core;
- Hackathons.

Développement

- Intégrer et décrire une seule fois leur logiciel;
- Environnement complet pré-configuré;
- Image multi-plateforme de taille légère (VM, conteneur);
- Disposer de sa propre copie de la version stable courante;
- Sans perturber les autres développeurs;
- *Automatisation et intégration continue;*
- Figer une version : démo, congrès, formation...

Intégration simplifiée avec Docker



Mode cluster

Un script de configuration qui crée un cluster à la « volée ».

- 6 gestionnaires de tâches disponibles



GRID ENGINE

Torque/Maui

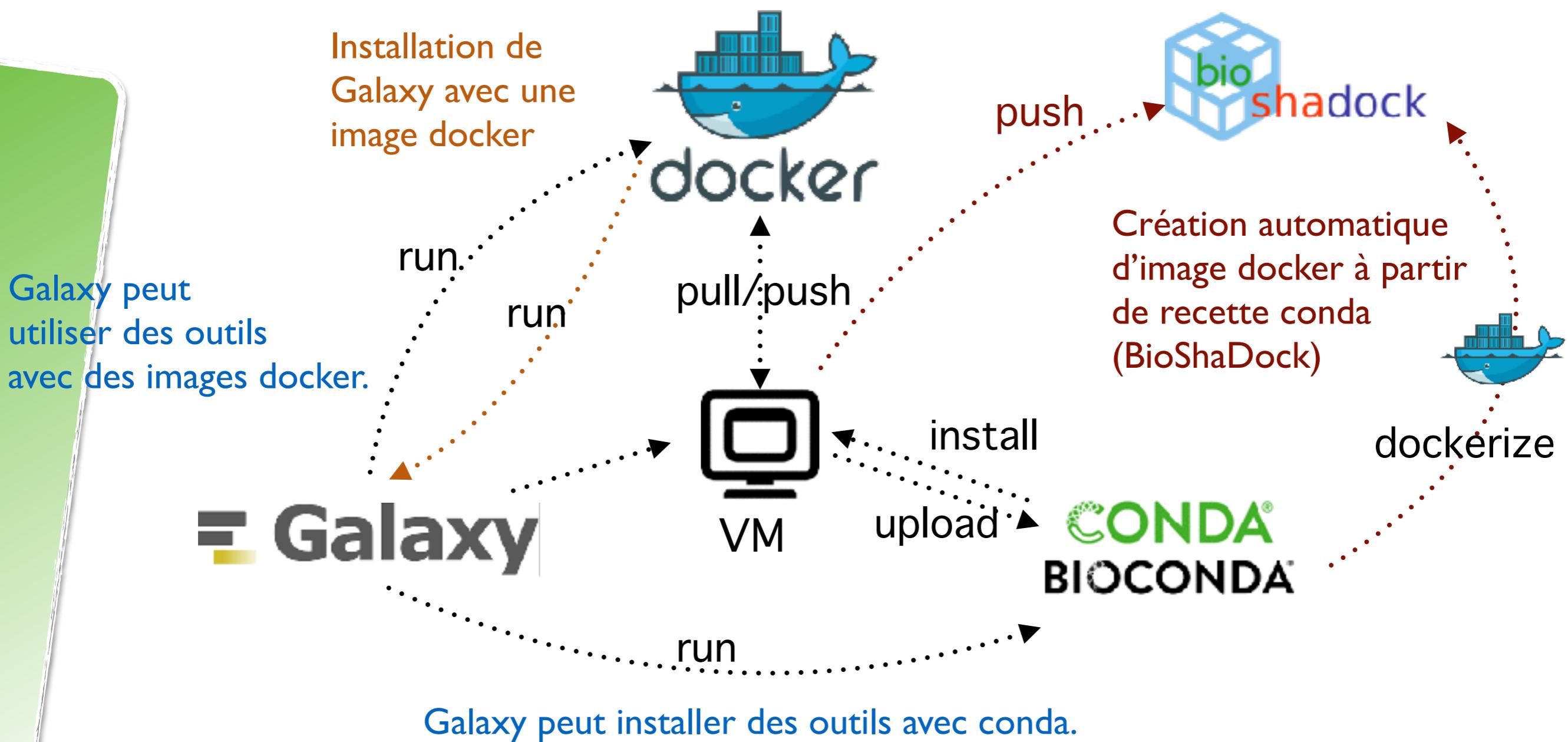
 RabbitMQ™



 Spark



Galaxy : installation simple d'outils



Services aux utilisateurs (504 pers.)

Traitement des données

Infrastructure

- Accès au cloud IFB
- Formations à son utilisation
 - ★ IBI-1 Introduction au Cloud IFB
 - ★ IBI-2 Utilisation avancée du Cloud IFB
- Equipe support IFB-core

Environnements de recherche standards

- utilisation habituelle
- possibilité de personnaliser son environnement
- avec de nombreux outils, pipelines et plateformes pré-configurés
- approve, docker, galaxy toolshed
- sans perturber les autres usagers

Le cloud IFB en soutien aux formations scientifiques

- Soutien en termes de ressources bioinformatiques et d'infrastructure,
- Ecoles scientifiques, tutoriels et formations
- Cycles universitaires : 5 cours en 2014-15 (131 élèves), 5 en 2016 (84 élèves).

Services communs à tous les membres

Un accès à l'infrastructure pour la communauté des sciences de la vie

Des ressources informatiques adaptées

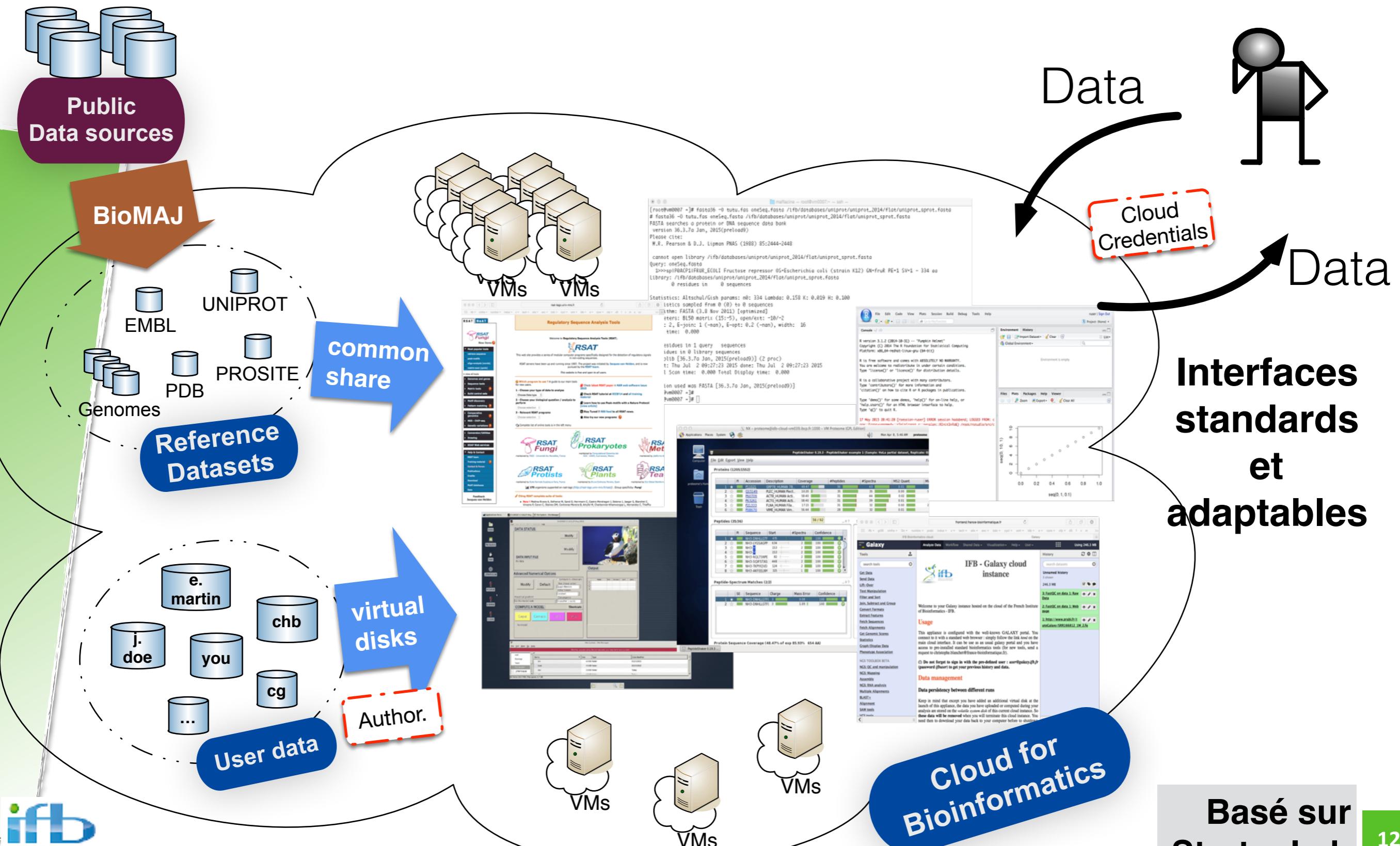
Des environnement de recherche personnalisés

Equipe support IFB-core

Des formations à l'utilisation du cloud IFB pour les analyses et développements méthodologiques

- Cursus "Cloud IFB pour les Sciences du Vivant" »
 - ★ 6 sessions en 2014-2015
 - ★ 7 pour 2016 (142 pers.)
- Documentation en ligne
- Ecole Cumulo Numbio 2015

Un cloud pour les sciences de la vie



Une interface web pour le cloud

IFB Bioinformatics cloud

Rechercher sur Google ou saisir une adresse

ifb tech elixr idb e core bdc cfp d3 m mapred viz r f g tech gbio privé v nimbl stra IRT crit Zr a i ngs

You are signed in as cblanche

News | Dashboard | Monitor | Settings | Administration | Help | Sign out

IFB BIOINFORMATICS CLOUD

DASHBOARD

Hosted at Powered by

News

2014-09-04: The IFB cloud will be demonstrated at ECCB14

2014-08-11: Mise à jour du site web

Rooms

	c2.large	c2.small	c2.xlarge	c3.large	c3.medium	c3.xlarge	m1.medium	m1.large	m1.xlarge
	10 / 36	78 / 144	8 / 18	18 / 34	39 / 70	8 / 16	4 / 6	2 / 2	0 / 2

Showing 1 to 6 of 6 entries

Username	ID	Name	Appliance	CPU%	CPU	Mem.	#Storage	Access
cblanchet	502	Public data source	BIO Data	2%	4	8	1	ssh http
cblanchet	515	compute	BID ComputeNode	0%	4	8	0	ssh
cblanchet	811	ubu 10G	Ubuntu 14.04 (base) - 10GB	0%	4	8	0	ssh
cblanchet	838	RSAT genomes ChB	NFS server	0%	4	8	1	ssh
cblanchet	839	test4	RSAT-ub14 nfs	0%	1	8	0	ssh http
cblanchet	842	testrel	RSAT-ub14-mini	0%	4	8	0	ssh http

STORAGE

Search:

First Previous 1 Next Last

CPU

MEMORY

<http://cloud.france-bioinformatique.fr/cloud>

IFB acknowledges funding by the call "Infrastructures in Biology and Health" in the framework of the French "Investments for the Future" Initiative



IFB is the French ELIXIR node



... qui évolue.

Screenshot of the RAINBio Cloud interface on biosphere.france-bioinformatique.fr:

- Header:** Navigation bar with tabs like to, w, cursp, cinfra, cycI, bioC, gr, eosC, mi, s, ifb, e, Do, prabi, indus, v, elix, elex, bdc, egi, madics, and a search bar for "biosphere.france-bioinformatique.fr".
- Left Sidebar:** "ifb Institut Français de Bioinformatique" logo, "Cloud" tab (highlighted), "RAINBio" and "EDAM" links, and "Admin" user information.
- Right Sidebar:** "Logged in as z_cloud_superuser_test" message.
- Main Content:**
 - CLOUD Section:** Shows a blue background image of a brain slice.
 - Deployments Table:**

ID	Broker	Name	Start/Stop	User	Spec	Access				
270	78078009 Nuvla	ImageJ2 (dev1)	↑ Jan 29 2017, 21h47	sandrine.perin@france-bioinformatique.fr	<table border="1"><tr><td>1</td><td>15</td></tr><tr><td>2</td><td>15</td></tr></table>	1	15	2	15	134.158.74.102 ssh-edu ssh
1	15									
2	15									
269	6282d600 Nuvla	Jupyter-notebook (dev1)	↑ Jan 29 2017, 21h47	sandrine.perin@france-bioinformatique.fr	<table border="1"><tr><td>1</td><td>15</td></tr><tr><td>2</td><td>15</td></tr></table>	1	15	2	15	134.158.74.101 ssh-edu ssh http
1	15									
2	15									
256	154d2ec4 SlipSt IFB	Started from Slipstream POC centos7	↑ Jan 28 2017, 20h50	bryan.brancotte	<table border="1"><tr><td>14</td><td>14</td></tr><tr><td>28</td><td>280</td></tr></table>	14	14	28	280	192.54.201.228 ssh
14	14									
28	280									
240	007ea899 Nuvla	Bacterial Genomics	↑ Jan 27 2017, 11h55	jonathan.lorenzo@france-bioinformatique.fr	<table border="1"><tr><td>3</td><td>3</td></tr><tr><td>6</td><td>60</td></tr></table>	3	3	6	60	192.54.201.217 ssh
3	3									
6	60									
159	50c49c2b Nuvla	Cluster swarm12 (0.1)	↑ Jan 18 2017, 22h40	sandrine.perin@france-bioinformatique.fr	<table border="1"><tr><td>2</td><td>2</td></tr><tr><td>4</td><td>30</td></tr></table>	2	2	4	30	134.158.74.38 ssh
2	2									
4	30									
160	50c49c2b Nuvla	Started from Nuvla Cluster swarm12 (0.1)	↑ Jan 18 2017, 22h40	sandrine.perin@france-bioinformatique.fr	<table border="1"><tr><td>2</td><td>2</td></tr><tr><td>4</td><td>30</td></tr></table>	2	2	4	30	134.158.74.38 ssh
2	2									
4	30									
 - Bottom Navigation:** Buttons for "Bookmarked appliances and deployments", "Last terminated deployments", and "Quota".
 - Table Headers:** "ID Broker Name Last start Last spec Highest Spec" (partially visible).

De nombreuses machines virtuelles (50+)

**Bacterial
genomics
(Insyght)**

**BIO
ComputeNode**

**Bio Workflow
Tools**

**bioDATA repo
(BioMaJ v3)**

**BioDataCloud
DNA-asm**

BioDataCloud IGV

**BioDataCloud
RNAseq**

BioPerl

BioStruct

Centos (3x)

CoreGeneBuilder

**COURS ENS Lyon
NGS 2016**

**COURS I2BC
Galaxy 2016**

**COURS Lille-
RNAseq 2016**

**COURS M2 Paris-
Saclay 2015**

**COURS Meet-U
IDock 2016**

Cytoscape

Docker

**EBA15 Cours-
Unix**

**EBA15 Galaxy
ChIP-seq**

EBA16 Galaxy

Eco Pop

**Fast RNASeq
Differential**

Galaxy

Galaxy FROGS

Galaxy MODAL

Gene regulation

ImageJ

**INCA-NGS BAPT
Normand**

INCA-NGS CLB

Integron Finder

LymAnalyzer

MacSyFinder

MicrobAnnot

ncPRO-seq

NFS server

**NGS PacBio-
Assembly**

PhyML

Proteomics

**R statistical
computing**

RADSeq

REPET

**REPET mysql-
server**

RSAT

SynBioWatch

TAGC Cours-Unix

Ubuntu (3x)

Wellinverter

RAINBio : Catalogue of Bioinformatics Images

Many Virtual Machines (50+)

The screenshot shows the RAINBio Cloud Appliances catalogue. At the top, there's a navigation bar with the IFB logo, Cloud, RAINBio (selected), EDAM, and a search bar. To the right are links for [en], Help, and Sign in. Below the header is a large blue banner with a brain-like image and the text "Genomics, Proteomics, Imaging...". The main content area is titled "RAINBIO - BIOINFORMATICS CLOUD APPLIANCES" and describes it as a catalogue of bioinformatics cloud appliances. It features a grid of cards, each representing a different virtual machine or tool. The cards include:

- Bacterial genomics (Insight)
- BIO ComputeNode
- BioDataCloud IGV
- BioDataCloud RNAseq
- bioDATA repo (BioMag)
- BioPerl
- BioStruct
- Bio Workflow Tools
- centos-7@nuvla
- Cluster swarm12
- CoreGeneBuilder
- COURS ENS Lyon NGS 2016
- COURS I2BC Galaxy 2016
- COURS Lille-RNAseq 2016
- COURS M2 Paris-Saclay 2015
- COURS Meet-U iDock 2016
- Cytoscape
- EBA16 Galaxy
- Eco Pop
- Fast RNASeq Differential
- Galaxy
- Galaxy FROG5

Each card lists the tools included, such as ABiSS, BioPerl, BLAST+, etc., and provides a link to more details.

Genomics,
Proteomics,
Imaging...

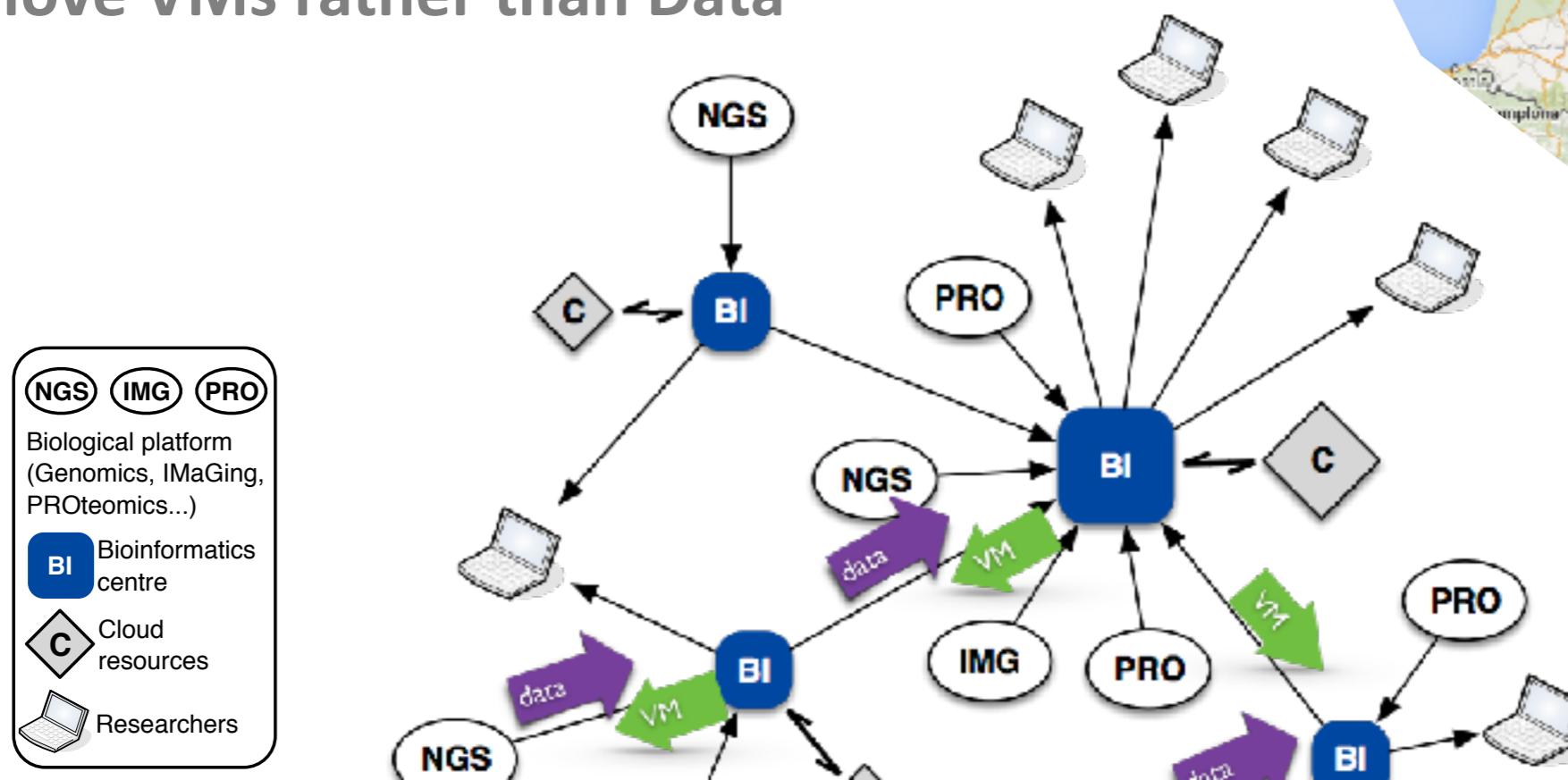
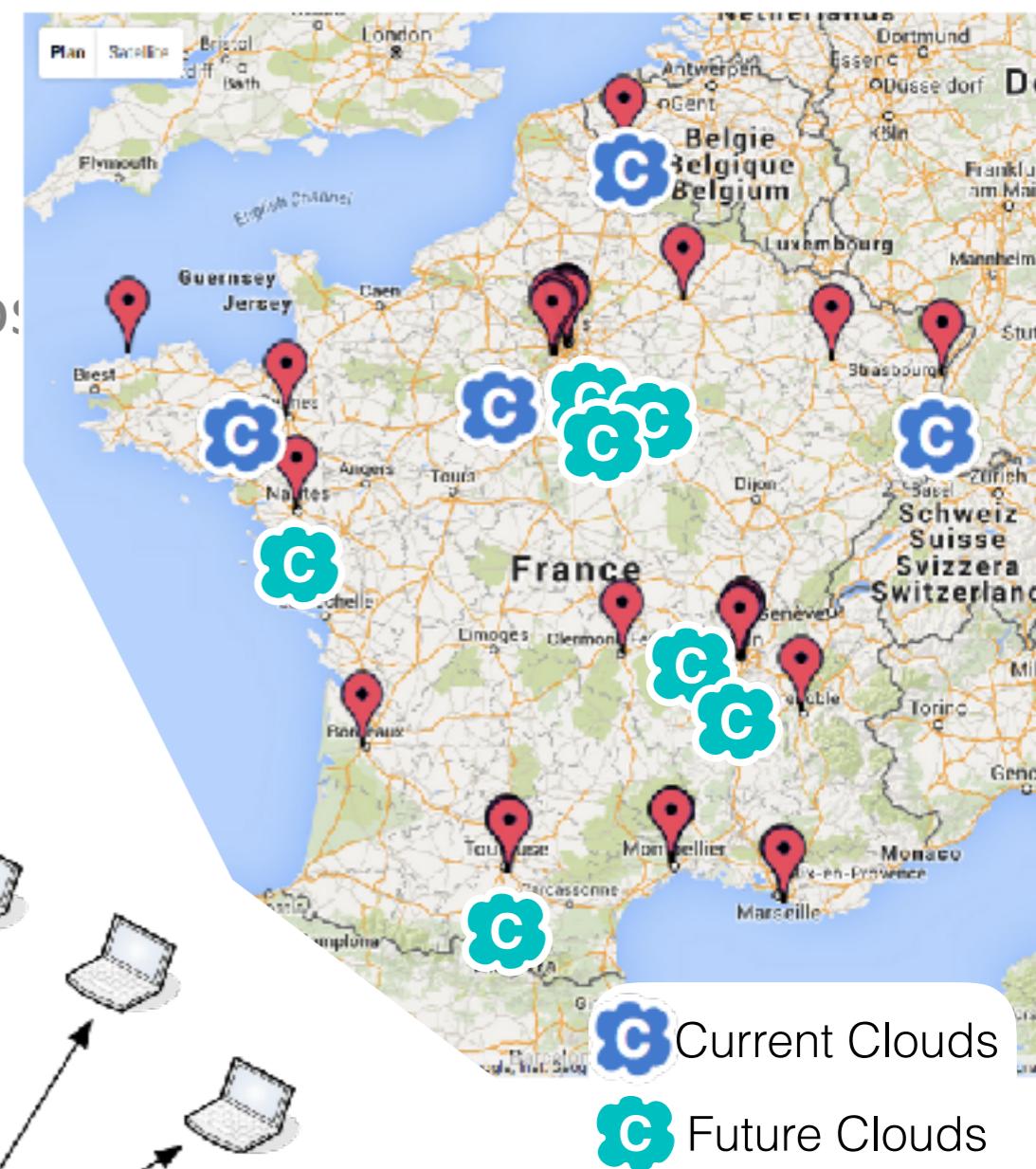
Filter by
Description
Tools
Interfaces
Topics
Contacts

Metadata from

- bio.tools (ELIXIR)
- Cloud Catalogue
- IFB Docker Hub
- BioShadock

Towards a French Multi-Cloud Infrastructure for Life Science

- Biosphere federates IFB's clouds
- 3 running clouds
 - ifb-core-pilot (Orsay), ifb-genouest-genos (Rennes), ifb-bistro-iphc (Strasbourg)
- 3 clouds in progress
 - ifb-core-lamarck (Orsay), ifb-bird-stack (Nantes), ifb-prabi-girofle (Lyon)
- Move VMs rather than Data



Machine nationale 'lamarck'



IFB-core hub	Compute #cores	Storage #TB	RAM #GB	Largest VM	Techno.
2014-08	200	50	2,000	20c 256GB	Stratus Lab
2017-1	5,000	1,000	40,800	128c 3TB	Open Stack
2017	10,000	2,000+			Open Stack

Mise en place:

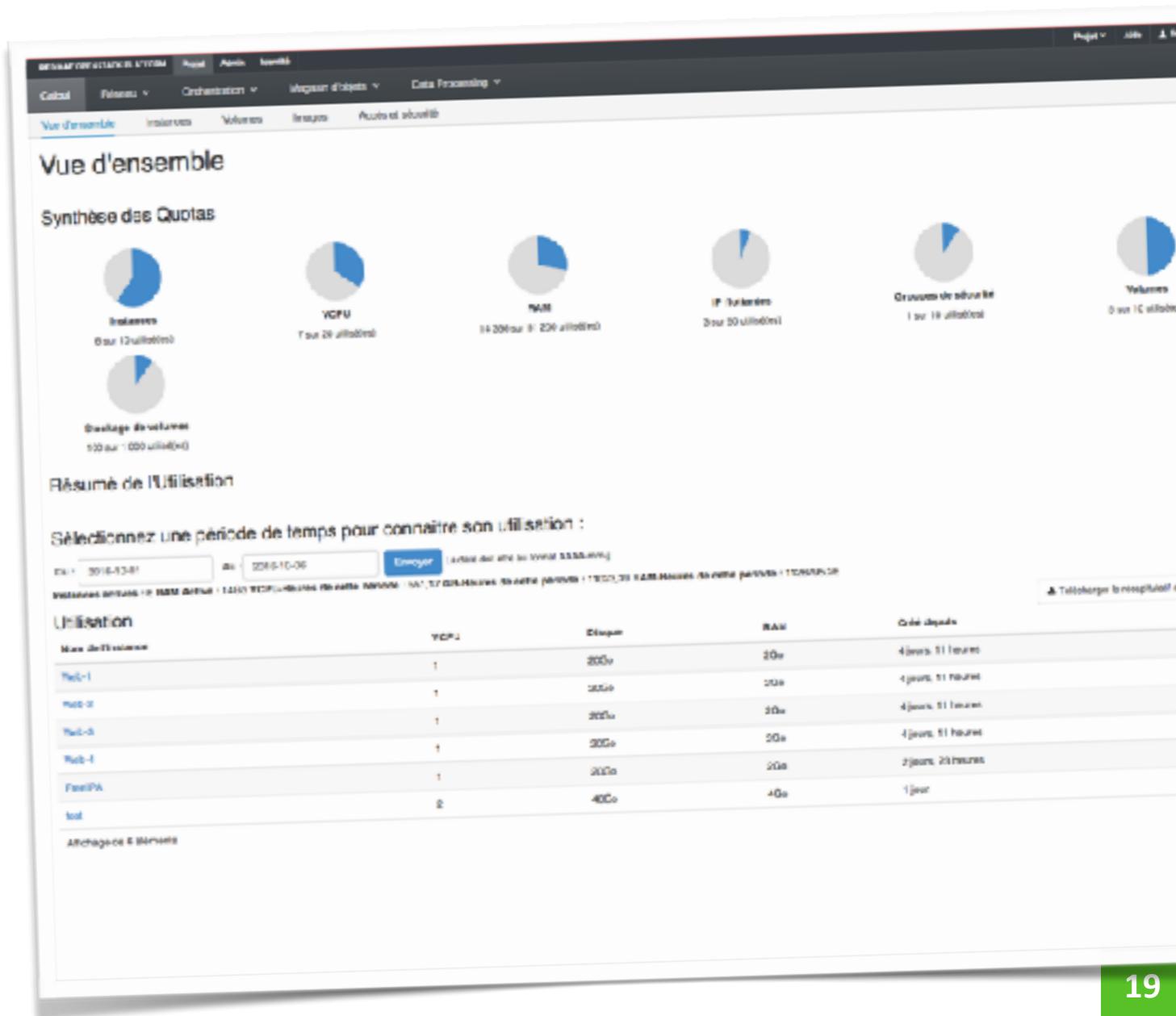
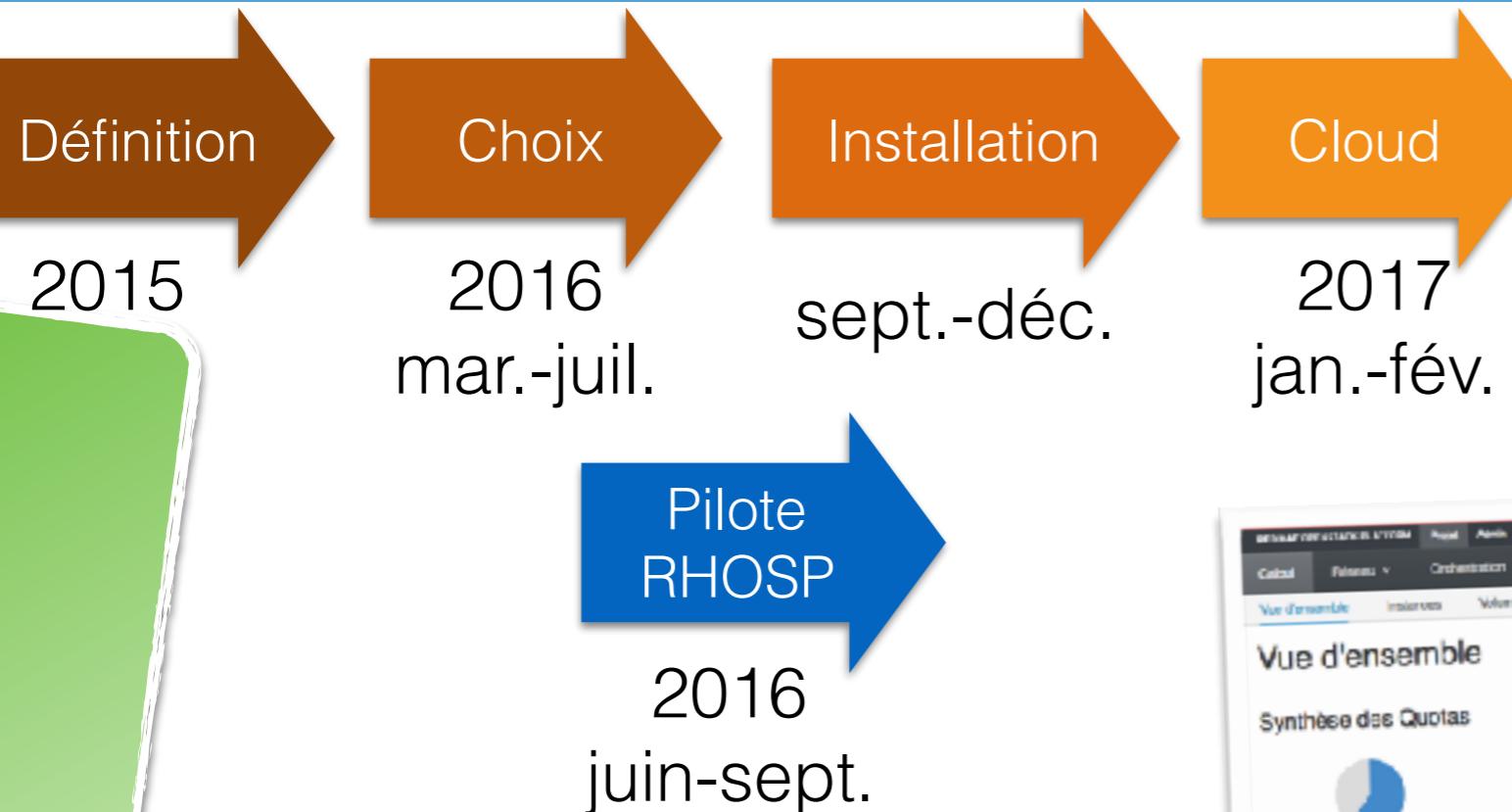
- Définition des besoins
- Choix technologiques
- Installation des matériels, câblage, clim.
- Mise en oeuvre du Cloud

Cellule e-Infrastructure - IFB-core

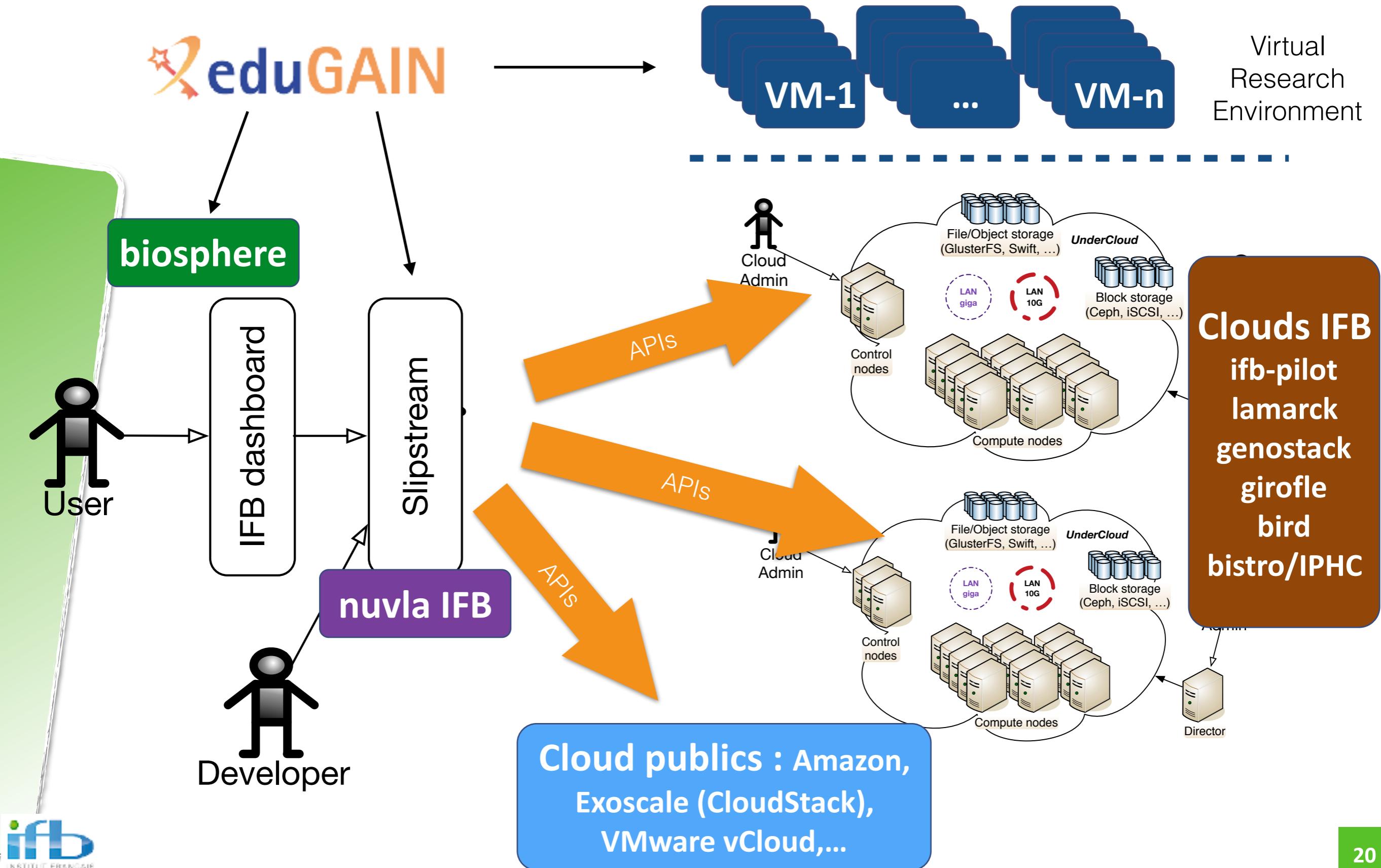
- En appui sur le GT GRISBI/Biosphère



Pilote RHOSP 9.0



Biosphère : Architecture



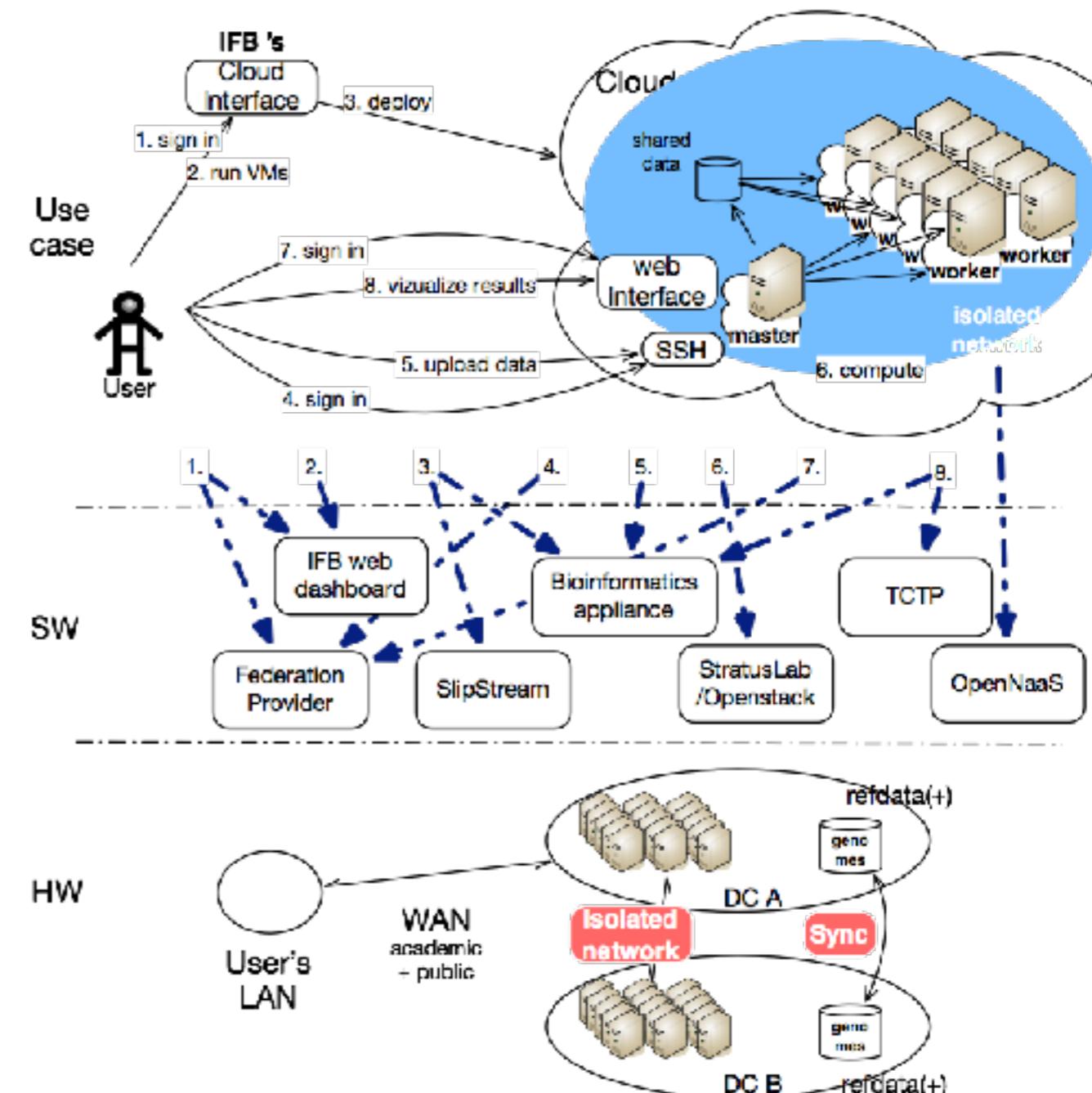
Example: Bacterial genomics - Insyght

Insyght, a synteny-ortholog browser

- Thomas Lacroix (IFB-MIGALE)
- Scientific field : microbial genomics
- Requires a Web portal, a DBMS and a computing cluster

Extended by CYCLONE components

- Nuvla:
 - ★ 1-click deployment (several VMs)
 - ★ Multi-cloud
- Authentication Federation Provider (eduGAIN)
 - ★ Web-access, SSH
- CNSMO-VPN: isolated network



Cyclone

Biosphère : solutions techniques

Gestion commune des utilisateurs

- Fédération d'identité eduGAIN (ou RENATER...)
- Federation Proxy (CYCLONE)
 - ★ Modes Web et SSH

Compatibilité des images (VM/container)

- Déploiements avec OpenStack, Docker, Conda...
- Recette SlipStream/NuvLa (CYCLONE)

Déploiement multi-cloud

- Recette SlipStream/NuvLa (CYCLONE)
- Connecteurs vers infra majeures de cloud (SIXSQ SlipStream/ NuvLa)

Gestion du réseau et de la sécurité sur plusieurs sites

- CNSMO (CYCLONE)
- Métrologie distribuée (CYCLONE)



ReproHackathon - Infrastructure

Appliances

- Docker (ReproHackathon)
- Bio Workflow Tools

Clouds

- ifb-core StratusLab
- ifb-core-pilot
- Genostack (PF GenOuest)
- Exoscale (cloud public)

Types de VM

- 8c-32Go, 4c-16Go, ...
- stockage : 50 Go (ifb-pilot, exoscale), 20 Go (genostack)

ReproHackathon - Support ifb-core

Documentation

- Liens utiles dans le GitHub ReproHackathon/rephackathon /
- par exemple une vidéo YouTube montrant la création d'une VM dans Biosphère

Référents

- docker et appliances ReproHackathon
 - ★ Sandrine et Jonathan
- portail biosphère, comptes cloud et usage
 - ★ Bryan
- cloud ifb-stratuslab et ifb-pilot
 - ★ Awa, Nicole et Frédéric

Questions ?



<http://www.france-bioinformatique.fr>

Acknowledgments

- **ReproHackathon : MaDICS, Psaclay, SIXSQ/Exoscale**
- **IFB members**
- **IFB Appliances developers (60+)**
- **CYCLONE members**
- **StratusLab, CNRS IDRIS**
- IFB is funded by French programs **PIA INBS 2012**
- EU H2020 projects: **CYCLONE (644925), EXCELERATE (676559)** and **EGI-Engage (654142)**

