

Biosphère

La Fédération de Clouds

de l'IFB

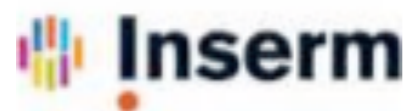
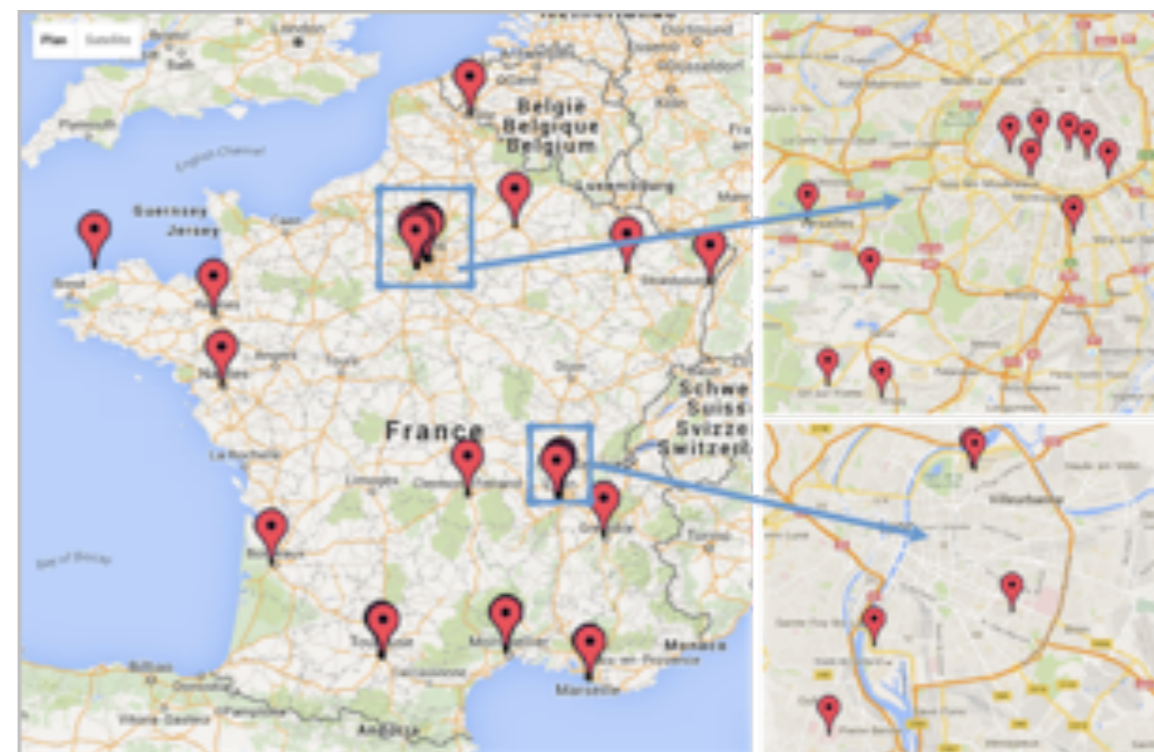


Christophe Blanchet



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

- Infrastructure nationale de support à la recherche.
 - Fondée en 2013 par *Jean-François Gibrat* (INRA).
 - Direction : C. Médigue (CNRS) J. van Helden (AMU)
- Financée par le Programme d'Investissement Avenir (**PIA**).
- **5 tutelles**
 - CNRS, Inserm, INRA, CEA, INRIA
- **32 plateformes régionales**
 - 398 personnes, dont 275 permanents et 123 CDD.
- Un **nœud national**, **IFB-core**, chargé de coordonner les actions.
- Nœud français du **réseau européen ELIXIR** (ESFRI pour la bioinformatique).
- Web: <http://www.france-bioinformatique.fr/>
- Liste des plateformes IFB: <http://www.france-bioinformatique.fr/en/platforms>



- Services courants
 - ❑ Fournir une **infrastructure physique et logicielle** de services en bioinformatique.
 - ❑ **Appui aux programmes de recherche** en biologie, santé, agronomie et environnement, via une expertise et des compétences mutualisées.
 - ❑ **Formations** en bioinformatique pour biologistes et bioinformaticiens.
- Innovation
 - ❑ Développer une **vision stratégique** pour maintenir la France au plus haut niveau d'**expertise** pour l'analyse des données biologiques, et donner accès aux **technologies de pointe** dans le domaine bioinformatique.
 - ❑ Servir de levier pour la conception et la mise en œuvre de **projets de recherche nationaux ambitieux**.
- Réseaux
 - ❑ Actions jointes avec les autres infrastructures nationales.
 - ❑ Assurer la représentation internationale de la communauté bioinformatique française, en particulier au sein du **réseau européen ELIXIR** (<https://www.elixir-europe.org/>).
 - ❑ Relations avec l'industrie.

WP1. Un environnement distribué de services en bioinformatique

Christophe Blanchet & Christophe Caron

A1.1 Calcul et stockage

Christophe Caron & Christophe Blanchet

A1.2 Software and data environment

Gildas Le Corguillé & Jacques van Helden

A1.3 Support aux bases de données

Claudine Médigue & Guy Perrière

A1.4 Catalogue des ressources

Jacques van Helden & Hervé Ménager

A1.5 Accès aux usagers

Christine Gaspin & Christophe Caron

A1.6 Mutualisation des services inter-infra

Christophe Caron & Christophe Bruley

A1.7 Guichet d'orientation et de consultation

Jacques van Helden & Ivan Moszer

WP2. Innovation: bioinformatique intégrative

Claudine Médigue & Jacques van Helden

A2.1 Projets pilotes inter-infrastructures

Claudine Médigue & Etienne Thévenot

A2.2 Appel à défis: lever les verrous scientifiques et technologiques

Ivan Moszer & Jacques van Helden

A2.3 Interopérabilité entre ressources

Sarah Cohen-Boulakia & Jacques van Helden

WP3. Ouverture internationale + industrie

Anne-Françoise Adam-Blondin & Claudine Médigue

A3.1 IFB, nœud français d'ELIXIR

Anne-Françoise Adam-Blondon, Sarah Cohen-Boulakia & C. Médigue

A3.2 Partenariat avec l'industrie

Patrick Durand et Victoria Dominguez Del Angel

WP4. Formation et diffusion

Jacques van Helden & Morgane Thomas-Chollier

A4.1 Formation

Jacques van Helden & Hélène Chiapello

A4.2 Actions jointes avec SFBI + GDR BIM

Morgane Thomas-Chollier (SFBI) & Hélène Touzet (GDR BIM)

A4.3 Communication

Claudine Médigue & Victoria Dominguez Del Angel

A4.4 Valorisation

Claudine Médigue & Victoria Dominguez Del Angel

WP5. Gouvernance

Claudine Médigue, Jacques van Helden

A5.1 Structures de gouvernance et de coordination

Claudine Médigue & Jacques van Helden

A5.2 Système de gestion de qualité

Patricia Laplagne & Claudine Médigue

A5.3 Modèle économique

Christine Gaspin & Patricia Laplagne

- **Réseau national de ressources computationnelles** (*National Network of Computational Resources, NNCR*).
- **Infrastructure physique distribuée.**
- **Environnement logiciel:**
 - middleware cluster + cloud
 - outils et workflows
 - accès aux données
- **Déploiement** de ressources bioinformatiques développées en France (outils, bases de données).
- **Accès aux utilisateurs**
- Appui aux **bases de données à haute valeur d'annotation**
- **Catalogue des ressources françaises** en bioinformatique
- **Partage des services** avec les autres infrastructures nationales de support à la recherche
- **Guichet de consultance et d'orientation**

COMPUTE & STORAGE

Co-construction d'environnement fédératif

Réseau d'infra physique

- Distribution géographique

Réseau de ressources humaines

- Mutualisation expertises
- Homogénéisation des procédures



Calcul

Stockage

Middleware: un environnement mixte

- Fédération de clusters
- Fédération de clouds

TOOLS

Outils et workflows

- Accessibilité
- Portabilité
- Reproductibilité des procédures

Bases de données

- Support logistique
- Déploiement
- Annotation par experts



END-USERS

Accès utilisateurs

- Interfaces conviviales, ou en ligne de commande
- Gestion des ressources par projet
- Modèle économique (en développement)

Portail des ressources

- accessibilité,
- visibilité,
- « trouvabilité »
- distribution des services sur le réseau

Catalogue

- Outils
- Databases
- Formations
- Expertises
- Plateformes

Guichet de consultance et d'orientation

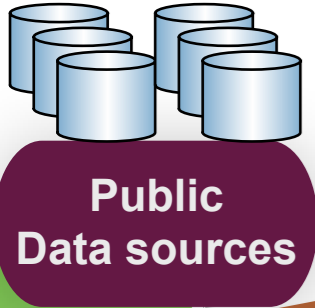
- Support direct aux usagers
- Suivi de projet
- Aide en amont, dès la conception des projets

COMPUTE & STORAGE

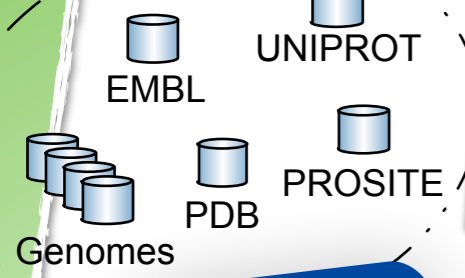
DATA

END-USERS

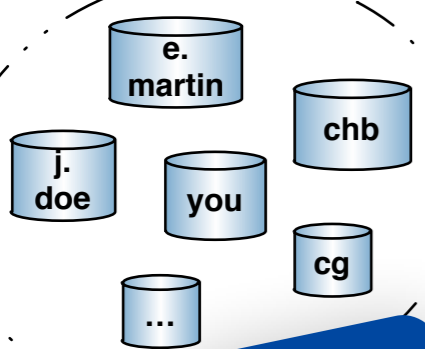
Cloud pour les Sciences Biologiques



BioMAJ

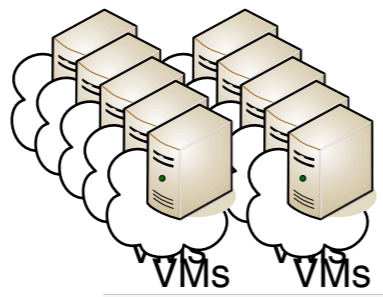


common share



virtual disks

Author.



Regulatory Sequence Analysis Tools (RSAT) interface showing search results and analysis options.

```
[root@vm0007 ~]# fasta36 -0 tutu.fas oneSeq.fasta /ifb/databases/uniprot/uniprot_2014/flat/uniprot_sprot.fasta
# fasta36 -0 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
!>>>apiRACF1FRUK_ECOLI Fructose repressor 05-Escherichia coli (strain K12) GN:fruk PE-1 SV=1 - 334 aa
Library: /ifb/databases/uniprot/uniprot_2014/flat/uniprot_sprot.fasta
@ residues in @ sequences

Statistics: Altschul/Gish parms: m1: 134 Lambda: 0.150 K: 0.019 H: 0.100
@ lotics sampled from 0 (0) to 0 sequences
@ lites: FASTA (3.8 Nov 2011) [optimized]
@ eters: BL50 matrix (15)-5, open/ext: -10/-2
: 2, E-join: 1 (-nan), E-prot: 0.2 (-nan), width: 16
time: 0.000

residues in 1 query sequences
ides in 0 library sequences
!11D [36.3.7a Jan, 2015(preload9)] (2 proc)
t: Thu Jul 2 09:27:23 2015 done: Thu Jul 2 09:27:23 2015
1 Scan time: 0.000 Total Display time: 0.000

len used was FASTA [36.3.7a Jan, 2015(preload9)]
Pum0007 -]#
```

Galaxy interface showing a workflow and data management options.

ID	Accession	Description	Coverage	#peptides	#proteins	MS2 Count	PK
1	Q5L5L8	RAC_HUMAN	10.75	9	1	9	1.00
2	Q5L5L8	RAC_HUMAN	10.75	9	1	9	1.00
3	Q5L5L8	RAC_HUMAN	10.75	9	1	9	1.00
4	Q5L5L8	RAC_HUMAN	10.75	9	1	9	1.00
5	Q5L5L8	RAC_HUMAN	10.75	9	1	9	1.00
6	Q5L5L8	RAC_HUMAN	10.75	9	1	9	1.00
7	Q5L5L8	RAC_HUMAN	10.75	9	1	9	1.00
8	Q5L5L8	RAC_HUMAN	10.75	9	1	9	1.00

Data

Cloud Credentials

Data

Interfaces standards et adaptables

Cloud for Bioinformatics

BIOSPHERE

French Multi-Cloud Infrastructure for Life Science

Node	Compute #vCPU	Storage #TB	RAM #GB	Max. VM vCPU/GB RAM	Federated
ifb-core-cloud <i>Lyon</i>	3936	408	20 408	128 / 3 TB	✓
ifb-core-lab <i>Lyon</i>	320	96	1 024	40 / 128	-
ifb-genouest <i>Rennes</i>	600	150	2 600	48 / 256	✓
ifb-prabi <i>Lyon</i>	448	144	1 500	64 / 768	✓
ifb-bird <i>Nantes</i>	160	50	448	32 / 128	✓
ifb-bistro <i>Strasbourg</i>	480	12	3 000	22 / 512	✓
	5 944	860	28 980		



IFB's Clouds: 5 running, 4 in progress, Openstack

Public clouds: HNSciCloud, Exoscale, AWS, VMware, CloudWatt, OVH, etc.

ifb-core-cloud

Hosted at CNRS CCIN2P3 datacenter, Lyon

- Hosting, cooling, ...
- HW/Sys admin and operation
- Cloud admin and operation

Common Flavors

Name	VCPUs	RAM (GB)	Disk (GB)	Eph.(GB)
ifb.m4.small	1	4	10	50
ifb.m4.large	2	8	20	100
ifb.m4.xlarge	4	16	20	200
ifb.m4.2xlarge	8	32	20	450
ifb.m4.4xlarge	16	64	20	900
ifb.m4.6xlarge	24	128	20	1 300
ifb.m4.8xlarge	32	192	20	1 800
ifb.m4.12xlarge	48	256	20	2 600

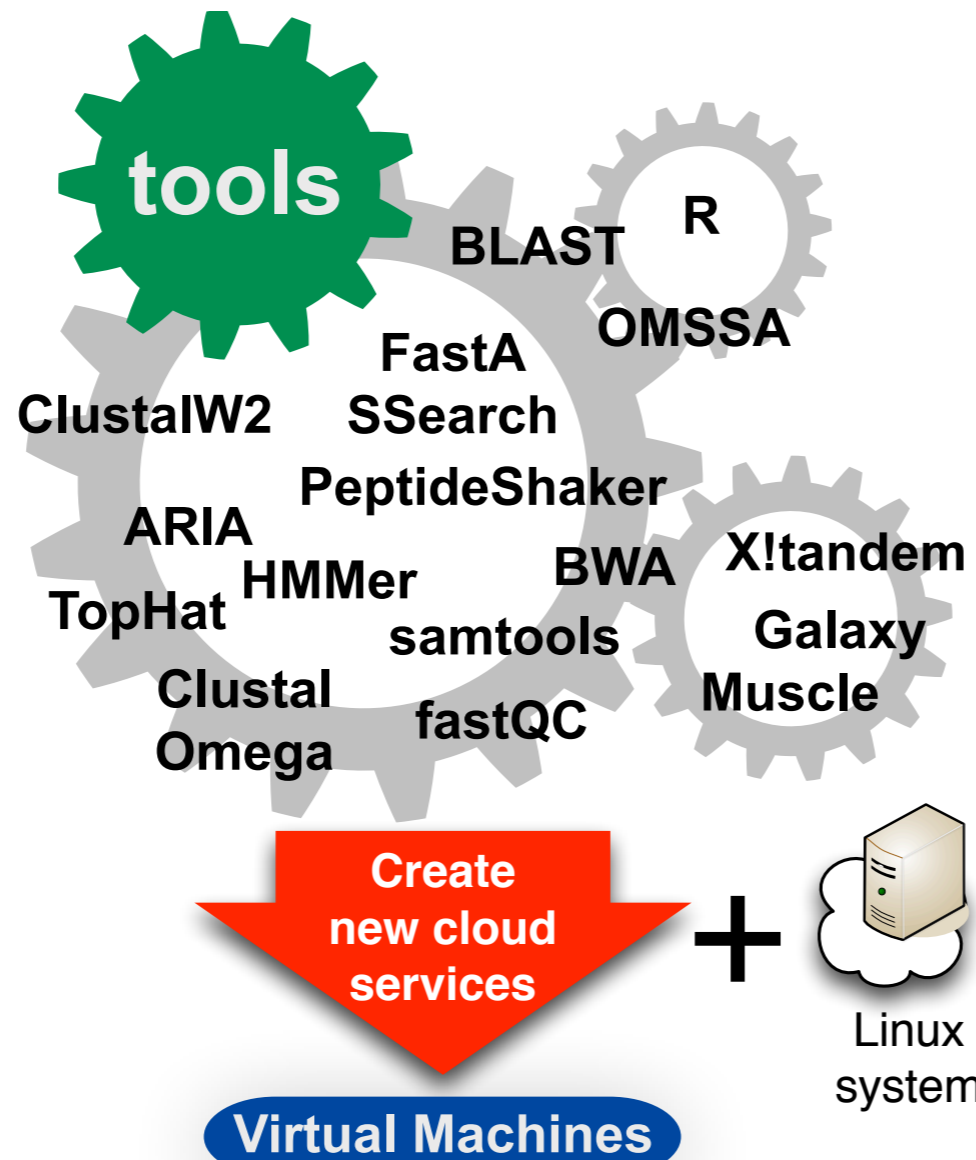
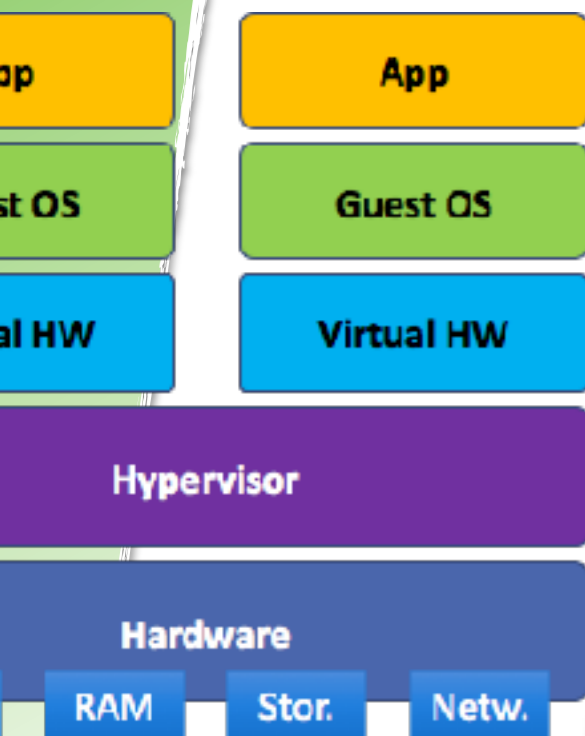


BigMem !

Name	VCPUs	RAM (GB)	Disk (GB)	Eph.
ifb.x1e.large	2	48	20	0
ifb.x1e.xlarge	4	96	20	0
ifb.x1e.2xlarge	8	192	20	0
ifb.x1e.4xlarge	16	384	20	0
ifb.x1e.8xlarge	32	768	20	0
ifb.x1e.16xlarge	64	1 536	20	0
ifb.x1e.32xlarge	128	3 072	20	0

Bioinformatics “Appliance”

Virtualization

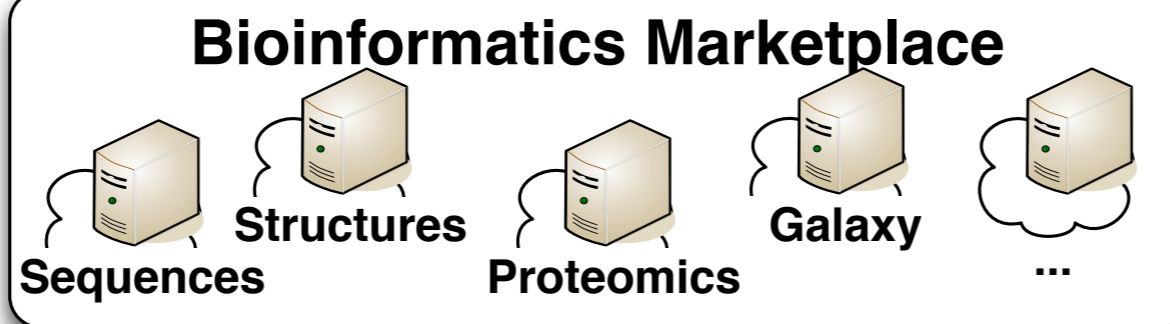


Appliance ?

- predefined virtual machine(s)
- including tools, pipeline, recipes...
- Ready to run

Annotated

- Title
- Description (w. controlled voc.)
 - ★ topics, tools
- Contact, developer(s) and **maintainer(s) !**



RAINBio : Catalogue of Bioinformatics Images

High-level cloud interface to select bioinformatics appliances.

The screenshot shows the RAINBio website interface. At the top, there is a navigation bar with 'Cloud', 'RAINBio', and 'EDAM' tabs. Below this is a header with the title 'RAINBIO - BIOINFORMATICS CLOUD APPLIANCES' and a subtitle 'Catalogue of bioinformatics cloud appliances, which you can browse and filter according to the predefined keywords of the EDAM ontology or with natural language.' The main content area is a grid of appliance cards, each with a title, a list of tools, and a play button icon. A search bar labeled 'enter filter' is located in the top right corner of the grid. A purple arrow points from the search bar to the text '55 Apps' on the right side of the image.

App Store	Appliances	Tools	Topics
Bacterial genomics (Insight)	BIO ComputeNode	BioDataCloud IGV	BioDataCloud RNAseq
BioStruct	Bio Workflow Tools	centos-7@mvla	Cluster swarm12
COURS IZBC Galaxy 2016	COURS Lille-RNAseq 2016	COURS M2 Paris-Saclay 2015	COURS Meet-U IDock 2016
EBA16 Galaxy	Eco Pop	Fast RNASeq Differential	Galaxy
			Galaxy FROGS

55 Apps

Metadata from

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

<http://biosphere.france-bioinformatique.fr/catalogue>

Biosphere Portal

High-level cloud interface to manage multi-cloud deployments.

The screenshot displays the 'Cloud' management interface. At the top, there is a navigation bar with the 'ifb' logo, 'Institut Français de Bioinformatique', and menu items for 'Cloud', 'RAINBio', and 'EDAM'. A user is logged in as 'Admin'.

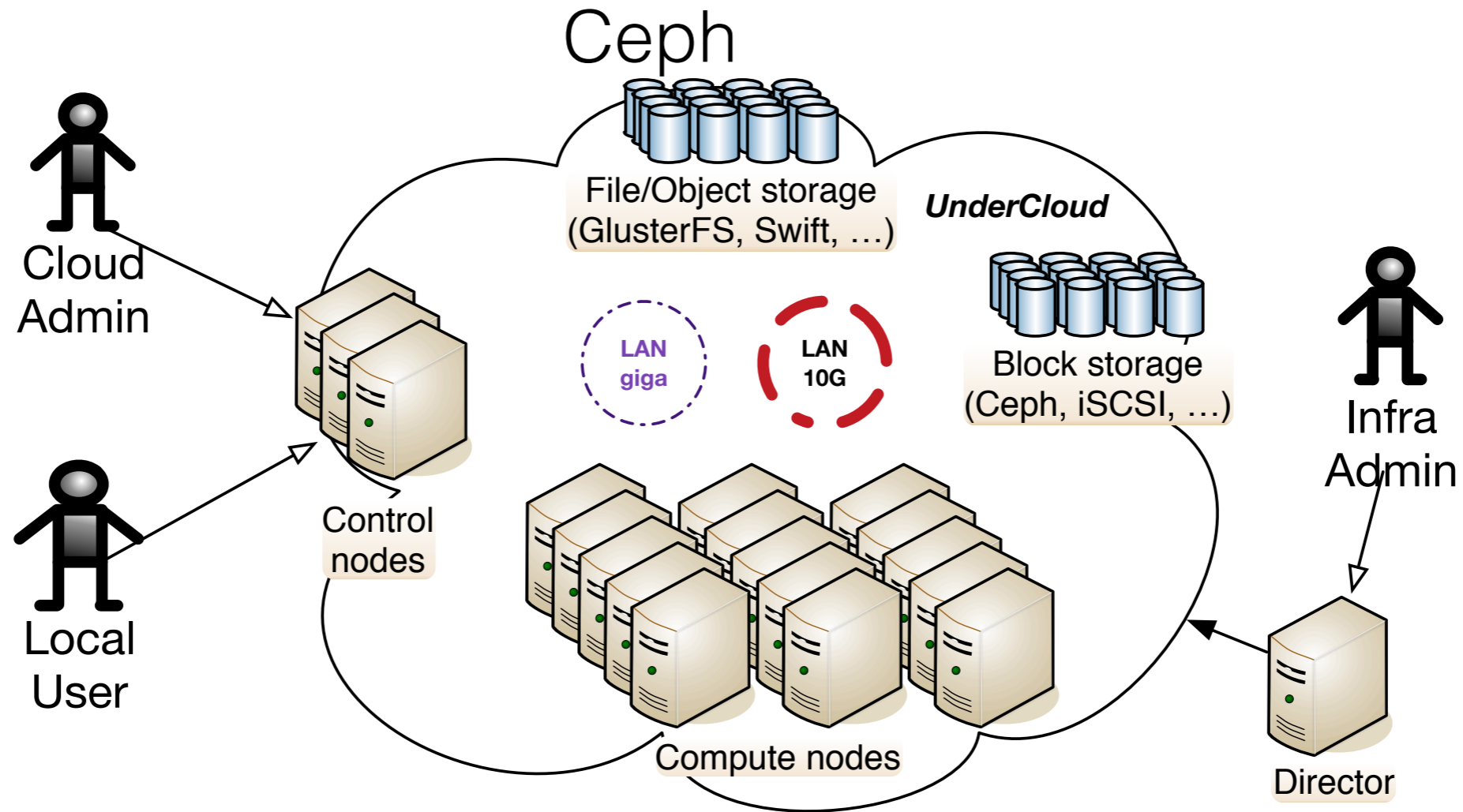
The main content area is titled 'CLOUD' and features a 'Deployments' tab. A table lists various cloud deployments with the following columns: ID, Broker, Name, Start/Stop, User, Spec, and Access. The table contains six rows of deployment data.

ID	Broker	Name	Start/Stop	User	Spec	Access
270	79078009 Muxin	image2 (dev1)	Jan 29 2017, 21h17		1 2 15	13% 02
269	6262d900 Muxin	Jupyter-notebook (dev1)	Jan 29 2017, 21h17		1 2 15	13% 01 ssh- http
256	154d27ec4 SlpSt.FB	Started from Silpstream POC centos7	Jan 29 2017, 20h50		14 14 28 280	19% 28
240	007ea399 Muxin	Bacterial Genomics	Jan 27 2017, 11h55		3 3 6 60	19% 17
169	60c49c2b Muxin	Cluster swarm12 (0.1)	Jan 18 2017, 22h40		2 2 4 90	13% 8
160	60c49c2b Muxin	Started from Nuv.la Cluster swarm12 (0.1)	Jan 18 2017, 22h40		2 2 4 90	13% 8

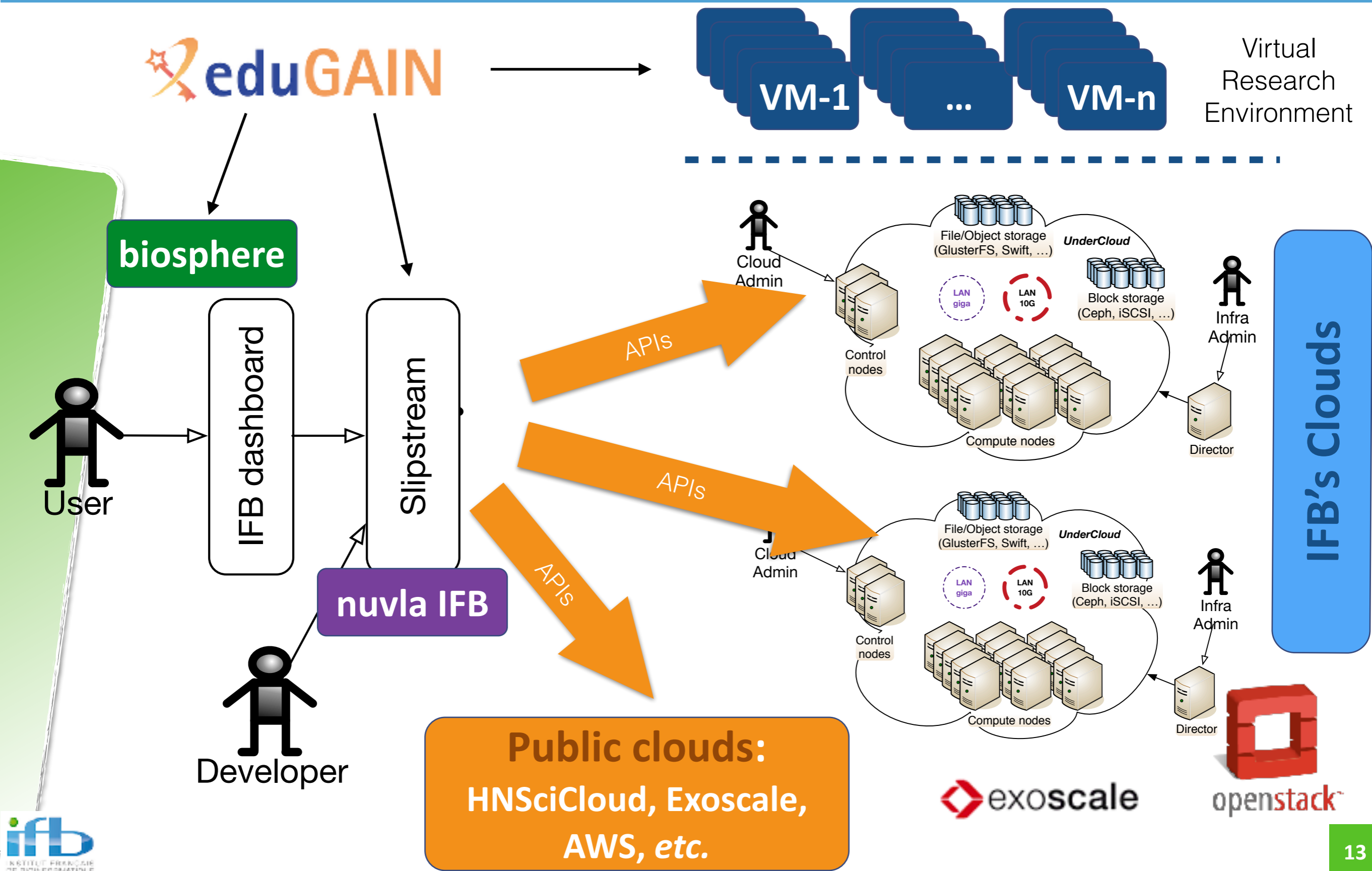
Below the table, there are navigation options: 'Bookmarked appliances and deployments', 'Last terminated deployments', and 'Quota'. A secondary table header is visible at the bottom with columns: ID, Broker, Name, Last start, Last spec, Highest Spec.

<https://biosphere.france-bioinformatique.fr/cloud>

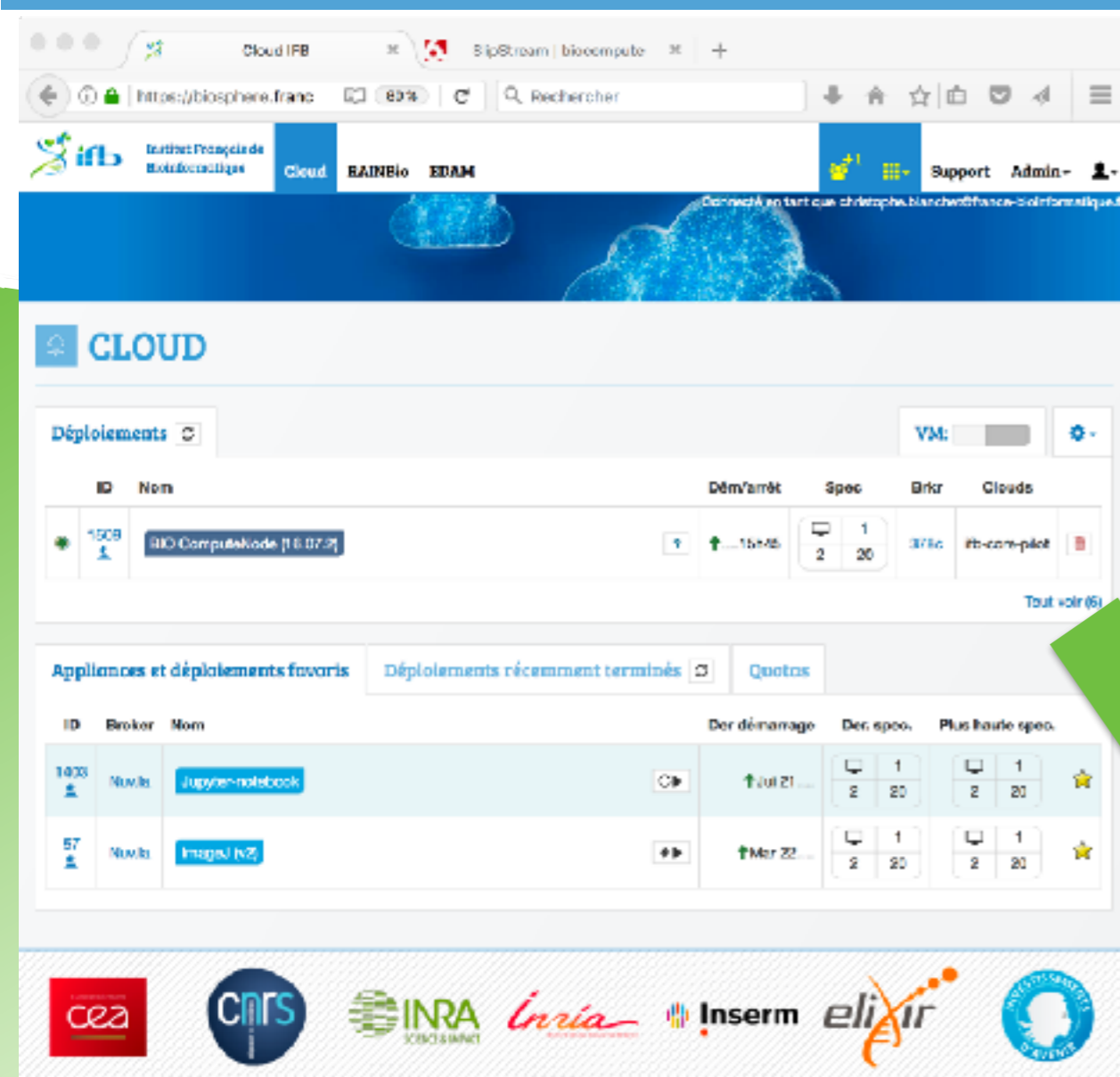
Cloud Site (OpenStack)



Biosphère : Federation of Clouds



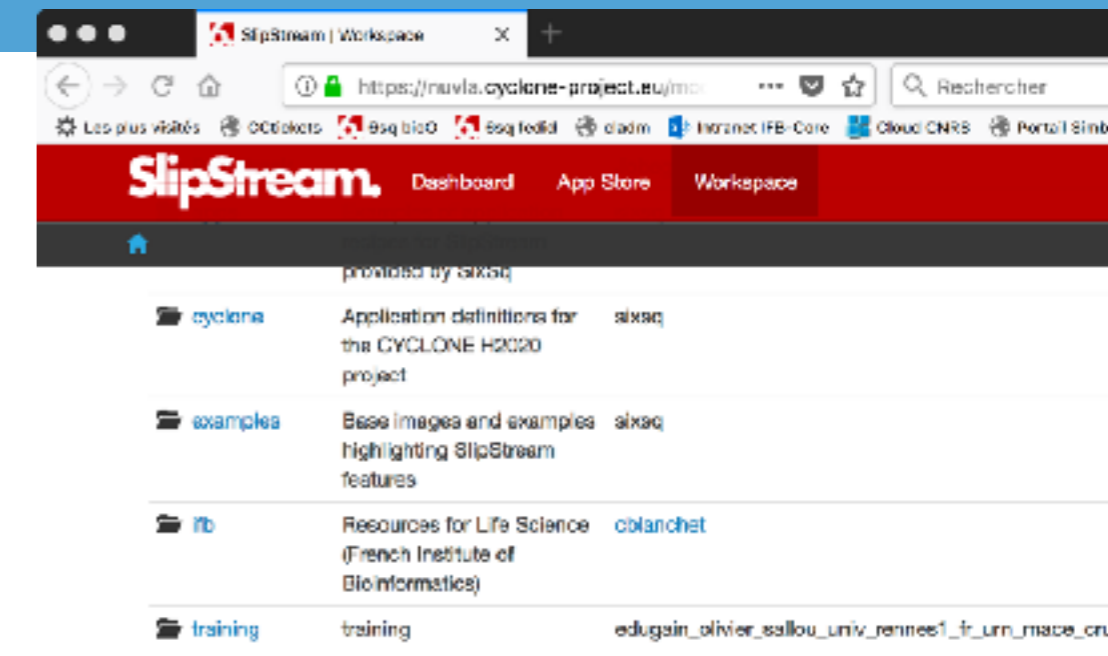
Biosphere Virtual Environment Deployment



The screenshot shows the SlipStream interface on the biosphere.franc website. The top navigation bar includes 'Cloud', 'RAINBio', and 'EDAM'. The main content area is titled 'CLOUD' and features a 'Déploiements' (Deployments) section with a table of active VMs. Below this is a section for 'Appiances et déploiements favoris' (Favorite appliances and deployments) with a table listing 'Jupyter-notebook' and 'ImageJ (v2)'. At the bottom, there is a row of logos for partner organizations: cea, CNRS, INRA, Inria, Inserm, and elixir.

ID	Nom	Dém/arrêt	Spec	Brkr	Clouds
1509	BIO-ComputeNode [1.0.07.2]	↑ 15:45	2 / 20	37/c	ifb-core-pilot

ID	Broker	Nom	Der démarrage	Der. spec.	Plus haute spec.
1433	Nuvla	Jupyter-notebook	↑ Jul 21 ...	2 / 20	2 / 20
57	Nuvla	ImageJ (v2)	↑ Mar 22 ...	2 / 20	2 / 20

The screenshot shows the SlipStream Workspace interface. The top navigation bar includes 'Dashboard', 'App Store', and 'Workspace'. The main content area displays a list of applications available in the workspace, including 'cyclone', 'examples', 'ifb', and 'training'.

Application	Description	Author
cyclone	Application definitions for the CYCLONE H2020 project	sixaq
examples	Base images and examples highlighting SlipStream features	sixaq
ifb	Resources for Life Science (French Institute of Bioinformatics)	cblanchet
training	training	edugain_olivier_sallou_univ_rennes1_fr_urn_mace...

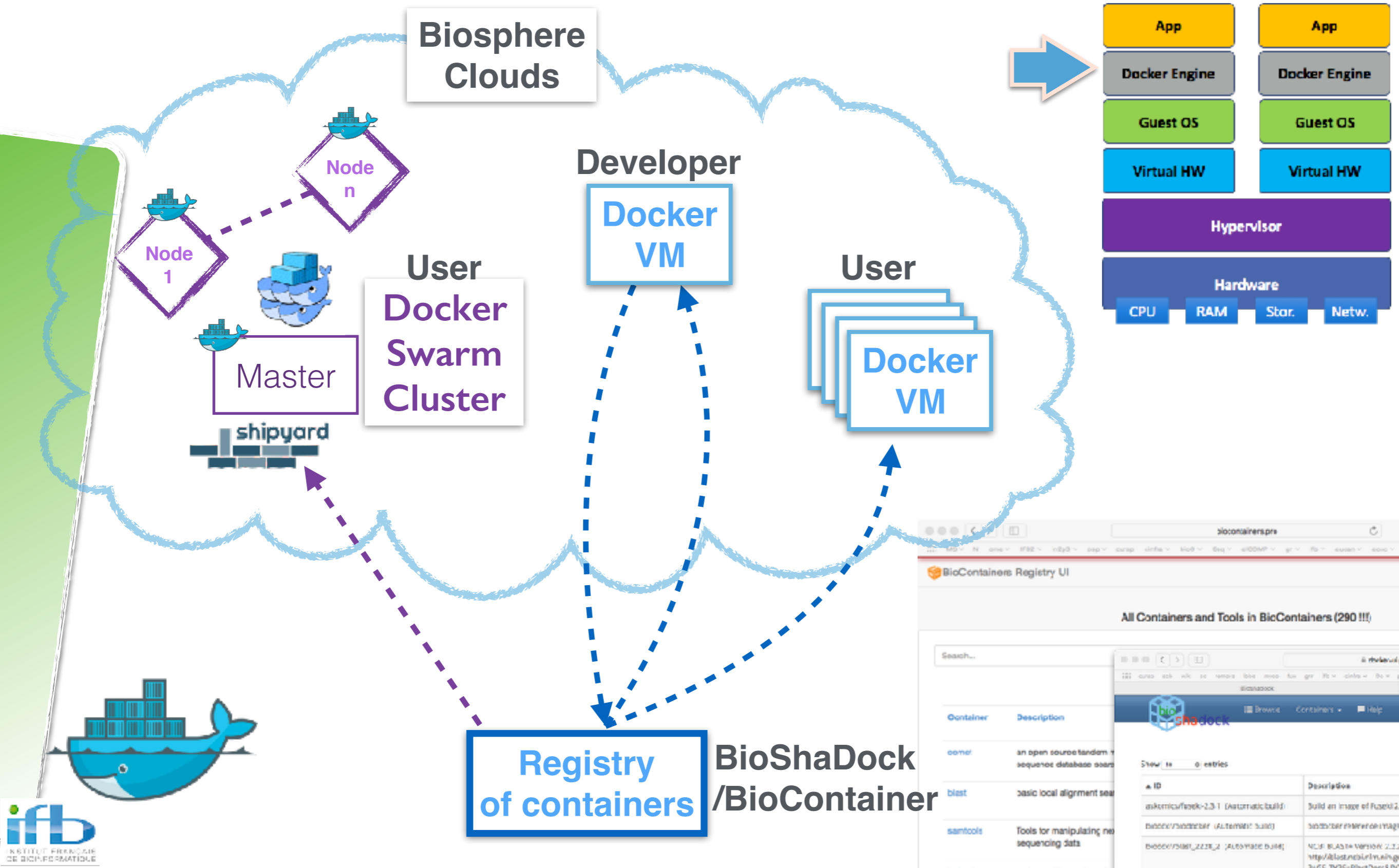


The screenshot shows the Nuvla Dashboard interface. The top navigation bar includes 'Dashboard', 'App Store', 'Workspace', 'Support', and 'Help'. The main content area displays a summary of VMs and a table of active deployments.

Application	VMs
ifb-core-pilot	1
ifb-prabi-girofle	0

ID	Application / Component	Service URL	State	Active VMs	Start Time	Clouds	User
378c0e41	biocompute v12114	ssh://root@192.54.201.207	Ready	1	28 Jun 2017, 13:45:16 UTC	ifb-core-pilot	cblanchet

Tools Integration with Docker



Biosphere Commons

Biosphere portal (internal development IFB)

- Cloud portal for life scientists
- RAINBio, catalogue of VM images

Multi-Cloud deployment

- Slipstream (SixSQ)

Federated identity (CYCLONE)

- eduGAIN, OpenIDconnect, Keycloak, IPA
- Web authorizations

Remote Graphical Desktop

- X2Go

Computing Cluster (intern. dev. IFB)

- 5 schedulers (elasticity)

Tools integration

- BioConda
- Docker, -Compose, Swarm (Images in Bioshadock/Biocontainer catalogues)
- Convergence cloud/container (AtomicHost, Kata,...)

Network-aaS (CYCLONE)

- VPN isolating a whole deployment

Cyclone



Torque/Maui **slurm**
workload manager

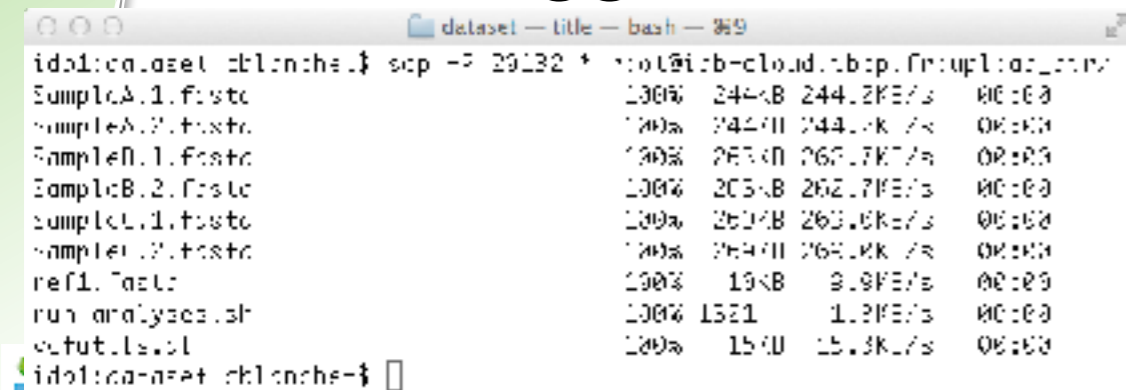
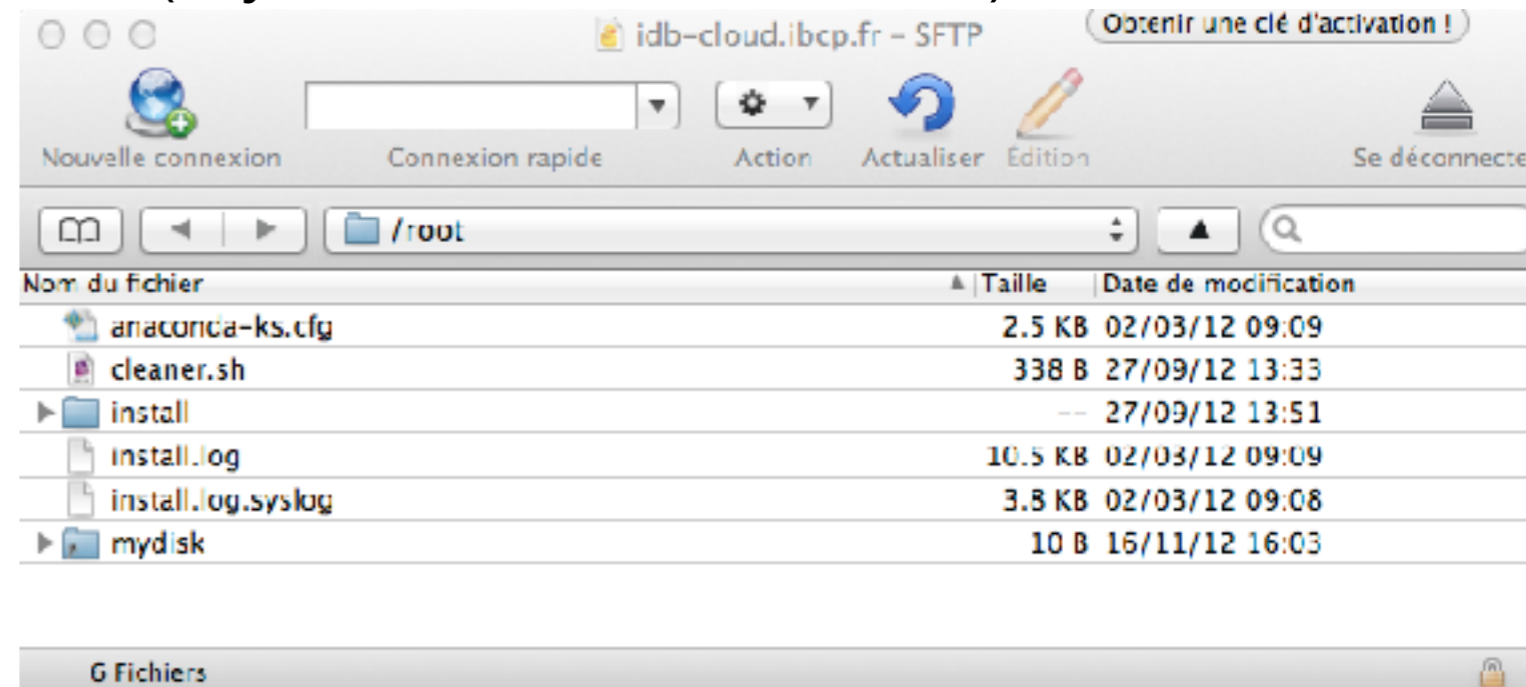
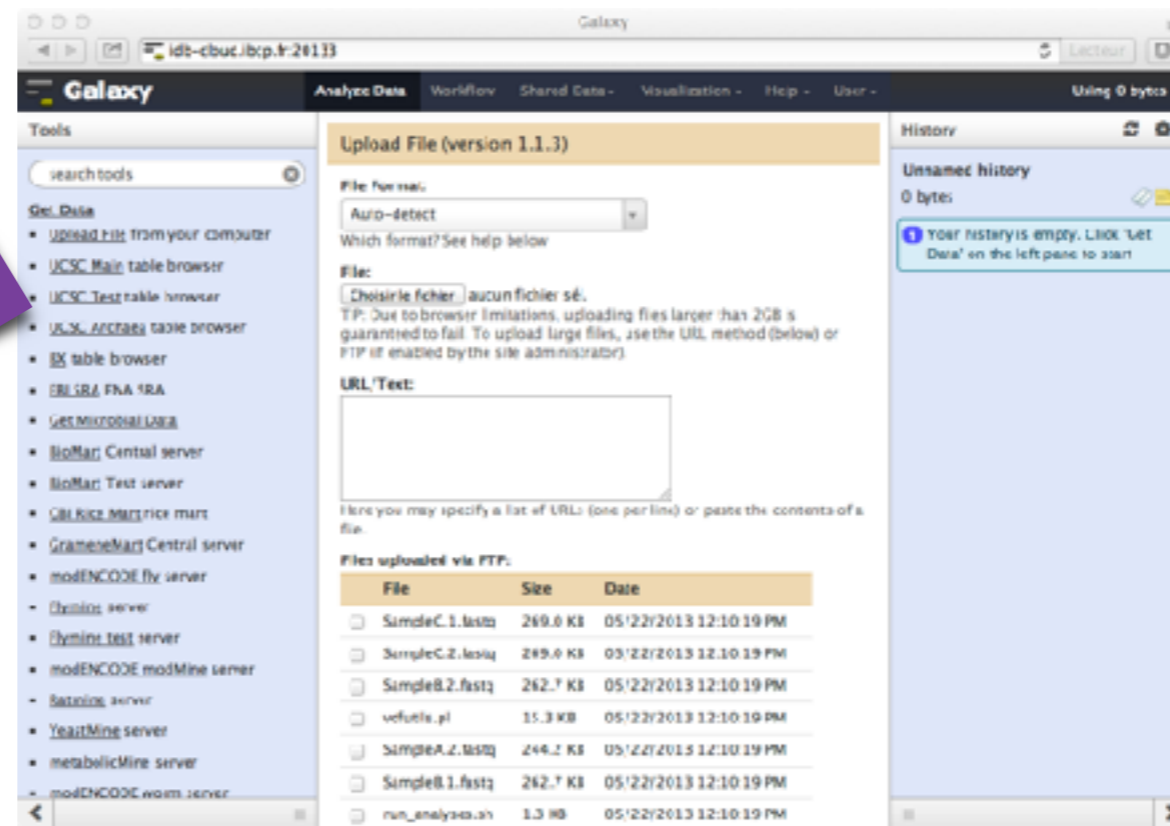
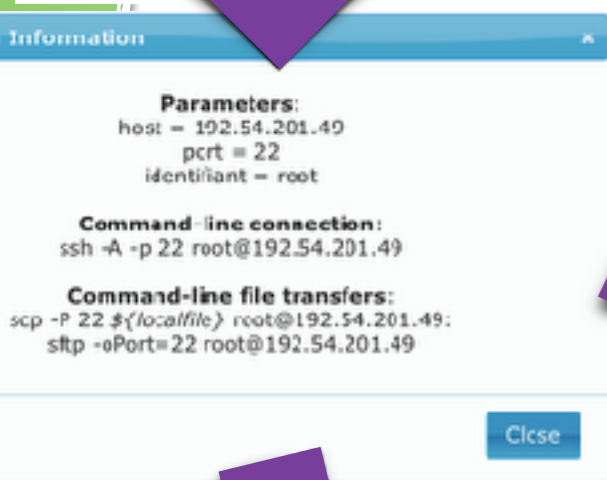
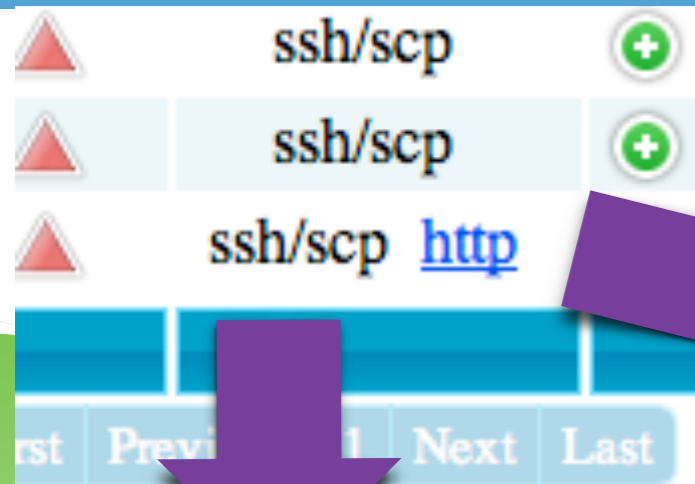
 RabbitMQ™

Echanger les données avec les VMs

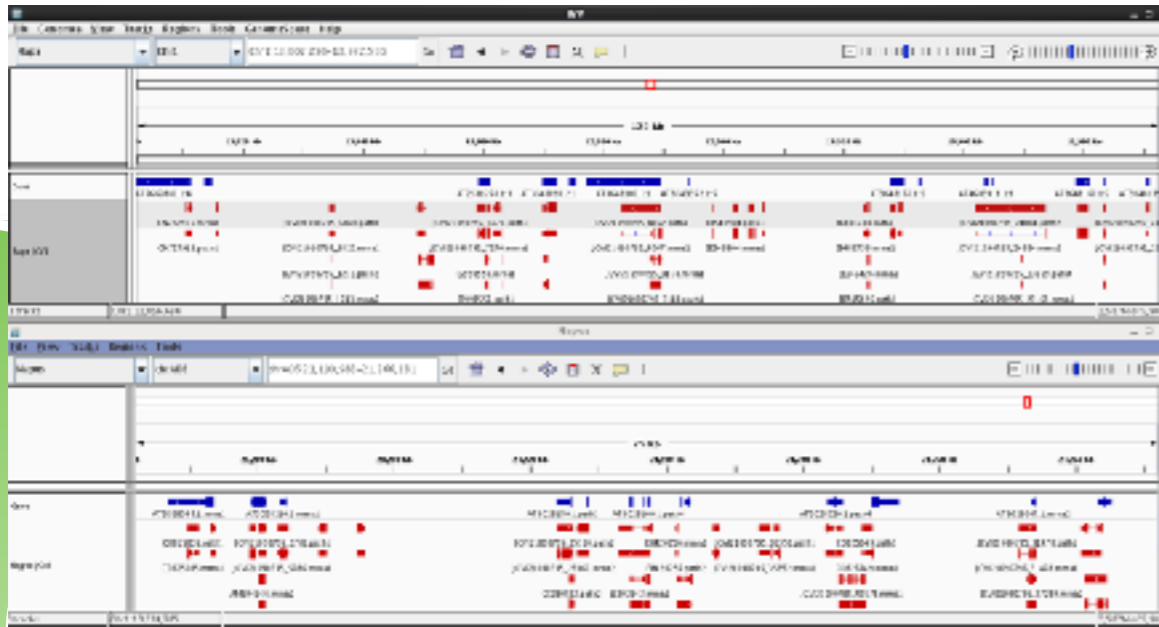
Interface Web dédiée (Galaxy, RSAT...) ou générique (Owncloud/Nextcloud, iRods)

Client GUI (Cyberduck, Filezilla...)

SSH



Quelques Services Actuels



IGV



Welcome

It works properly this appliance need at least 16GB of RAM.
It is strongly recommended to attach a disk before performing any analysis.

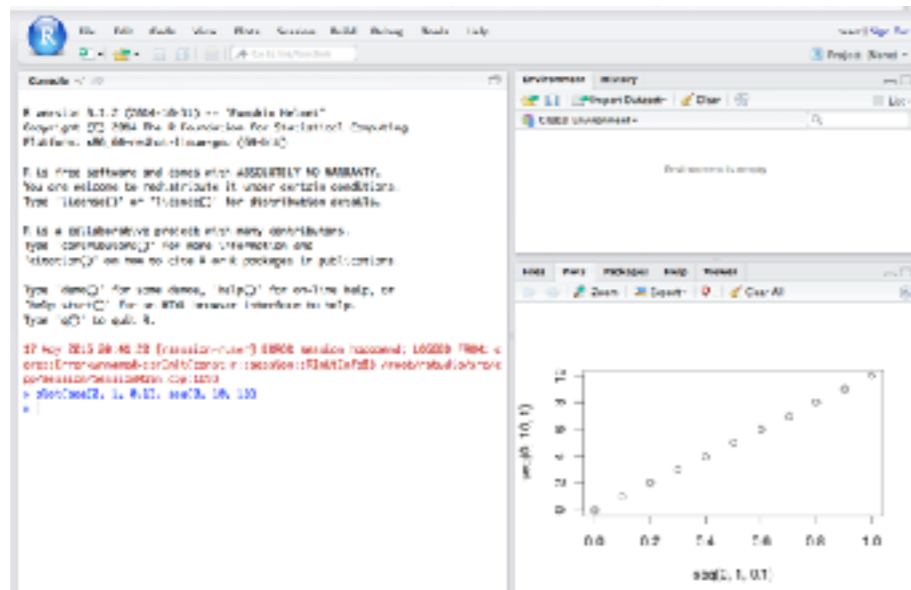
This appliance provides 2 tools dedicated to bacterial/fungal analysis

1. **antiSMASH**: antiSMASH allows the rapid genome-wide identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genomes
2. **CRISPRDetect**: CRISPRDetect as the name suggests let you discover CRISPRs

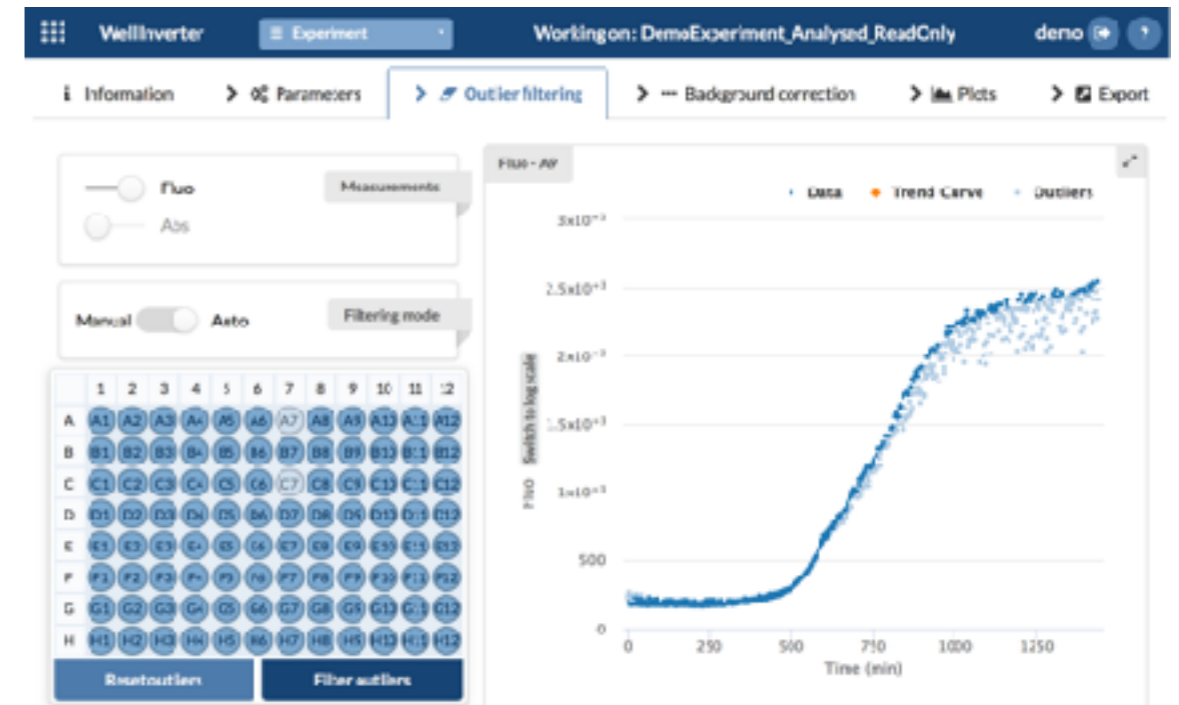
These tools can be run with command lines.

1. use a SSH client: `ssh -i -p 22 root@cpip`
2. put your input file into `/root/ryb15k/microbannot/vebsmash` (to run antiSMASH) or `/root/ryb15k/microbannot/crispr` (to run crispr_detect)
3. connect to docker container: `docker exec -i -t m.crobotinct /bin/bash`
4. run antiSMASH: `run_antiSMASH <input file name> -output directory name- [antiSMASH options]`
5. run crispr detect: `run_crispr_detect <input file name> -output directory`

MicrobAnnot (WPs)



R / Rstudio / Shiny



WellInverter (AAP 2015)

Exemple d'Utilisation Intensive du Cloud IFB

Insyght, analyse de 5 663 génomes bactériens

- Thomas Lacroix, Jean-François Gibrat (IFB-MIGALE, INRA)
 - ★ http://genome.jouy.inra.fr/Insyght_2692_complete_bacteria
- Besoins
 - ★ plus de 16 millions de comparaisons de génomes
 - ★ 52 000 heures de calcul
- Infrastructure virtuelle déployée sur le cloud IFB
 - ★ Calcul (cluster SGE): 130 VM de 32 vCPU
 - ★ Stockage (GlusterFS): 60 VM de 500 Go, volume unique de 30 To
 - ★ Temps de déploiement : 30min+1h (SGE), 33 min (GlusterFS)

● Calculs

- ★ Temps total : 13 heures
- ★ Données produites: 2,4 To
- ★ Aucune erreur !

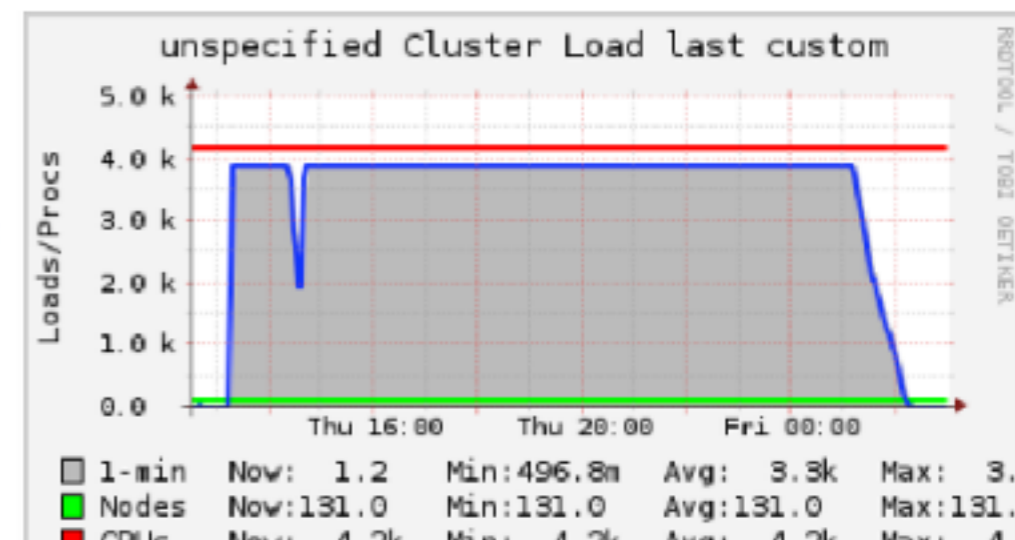
BiosphereCluster Grid > unspecified > --Choose a Node

Overview of unspecified @ 2017-10-13 07:28

CPU's Total: **4168**
Hosts up: **131**
Hosts down: **0**

Current Load Avg (15, 5, 1m):
0%, 0%, 0%

Avg Utilization (last custom):
0%



Conclusion

IFB, une infrastructure nationale de support à la recherche.

- Financée par le Programme d'Investissement Avenir (PIA).
- 5 tutelles: CNRS, Inserm, INRA, CEA, INRIA
- 30 plateformes régionales.
- Un nœud national, IFB-core, chargé de coordonner les actions.
- Nœud français du réseau européen ELIXIR.

Missions

- Fournir une infrastructure physique et logicielle de services en bioinformatique.
- Appui aux programmes de recherche en biologie, santé, agronomie et environnement, via une expertise et des compétences mutualisées.
- Formations en bioinformatique pour biologistes et bioinformaticiens.
- Actions jointes avec les autres infrastructures nationales.

Réseau national de ressources bio-informatique

- Infrastructure physique distribuée.
 - ★ Environnement logiciel cluster + cloud, outils et workflows, données
- Catalogue des ressources françaises en bioinformatique.
- Partage des services avec les autres infrastructures nationales de support à la recherche.
- Guichet de consultance et d'orientation.