

IFB-Biosphère : Services cloud pour l'analyse des données des sciences de la vie

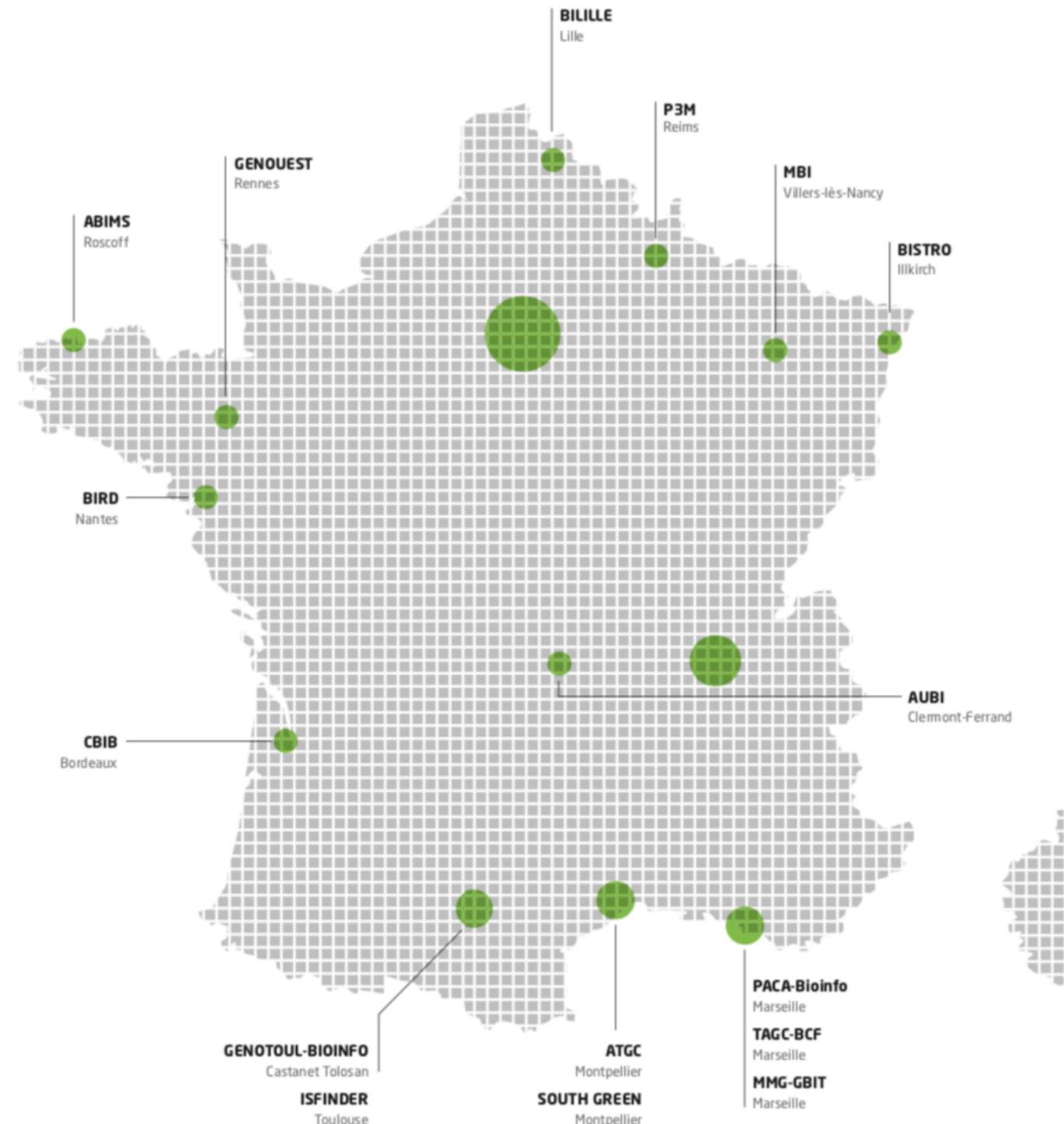
Christophe Blanchet *, Olivier Collin, Matéo Boudet, Stéphane Delmotte, Hervé Gilquin, Jean-François Guillaume, Efflam Lemaillet, Jonathan Lorenzo, Jérôme Pansanel, Olivier Sallou, Bruno Spataro



- Fournir une infrastructure de services en bioinformatique accessible à l'ensemble des usagers (biologistes + bioinformaticiens).
- Appui aux programmes de recherche en biologie, santé, agronomie et environnement, via une expertise, des compétences mutualisées, et des développements méthodologiques.
- Formations en bioinformatique pour biologistes et bioinformaticiens.
- Développer une vision stratégique pour maintenir la France au plus haut niveau d'expertise pour l'analyse des données biologiques.
- Servir de levier pour la conception et la mise en œuvre de projets de recherche nationaux ambitieux.
- Assurer la représentation internationale de la communauté bioinformatique française, en particulier au sein du réseau européen ELIXIR.
- Développer des partenariats avec le monde industriel et les acteurs de la bioinformatique française: SFBI, GDR BIM.

Organisation de l'IFB

- **31 plateformes**
 - Diversité des services
 - Proches des équipes de recherche
 - CNRS, CEA, INRA, INRIA, INSERM, Universités...
- Un noeud **national, IFB-core**
 - CNRS UMS 3601, INRA UMS 1385, Inserm US21
 - Coordination de l'infrastructure
 - Mutualisation des ressources
 - ELIXIR-FR



RÉGION PARISIENNE

EBIO
Orsay
INSTITUT CURIE
Paris
IGR
Villejuif
MICROSCOPE
Evry
MIGALE
Jouy-en-Josas
C3BI
Paris
RPBS
Paris
URGI
Versailles
ORPHANET
Paris
ICONICS
Paris
IFB CORE
Evry

RÉGION LYONNAISE

INCA-SLC
Lyon
PRABI-HCL
Lyon
PRABI-AMSB
Villeurbanne
PRABI-Lyon-Grenoble
Villeurbanne
PRABI-Lyon-Gerland
Lyon

Feuille de route 2018-21

WPI. Un environnement distribué de services en bioinformatique

- A1.1 Réseau national de ressources informatiques (NNCR)
- A1.2 Environnements logiciels et données
- A1.3 Support aux bases de données
- A1.4 Catalogue des ressources
- A1.5 Accès aux usagers
- A1.6 Mutualisation des services inter-infra
- A1.7 Guichet d'orientation et de consultation

WP2. Innovation: bioinformatique intégrative

- A2.1 Projets pilotes inter-infrastructures

- A2.2 Appel à défis: lever les verrous scientifiques et technologiques
- A2.3 Interopérabilité entre ressources

WP3. Ouverture internationale + industrie

- A3.1 IFB, nœud français d'ELIXIR
- A3.2 Partenariat avec l'industrie

WP4. Formation et diffusion

- A4.1 Formation
- A4.2 Actions jointes avec SFBI + GDR BIM
- A4.3 Communication & Valorisation

WP5. Gouvernance

- A5.1 Structures de gouvernance et de coordination
- A5.2 Système de gestion de qualité
- A5.3 Modèle économique

Cloud IFB-Biosphère

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

Fédération de clouds

- 6 sites en production + 4 en préparation
- Premiers clouds en 2011
- Fédération en place depuis Fév. 2017

Un compte unique pour tous les clouds

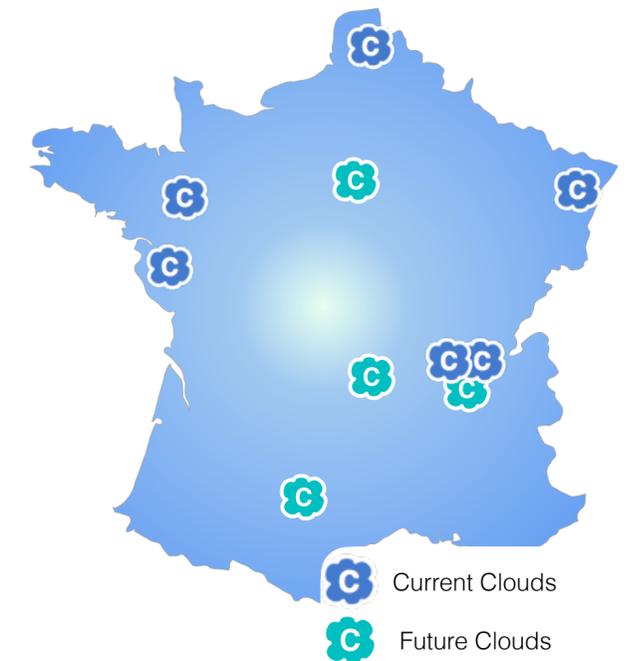
- Basé sur les identifiants officiels des utilisateurs dans leur organisme (fédération européenne eduGAIN)

Un portail unique (Biosphère) pour déployer les machines virtuelles sur tous les clouds

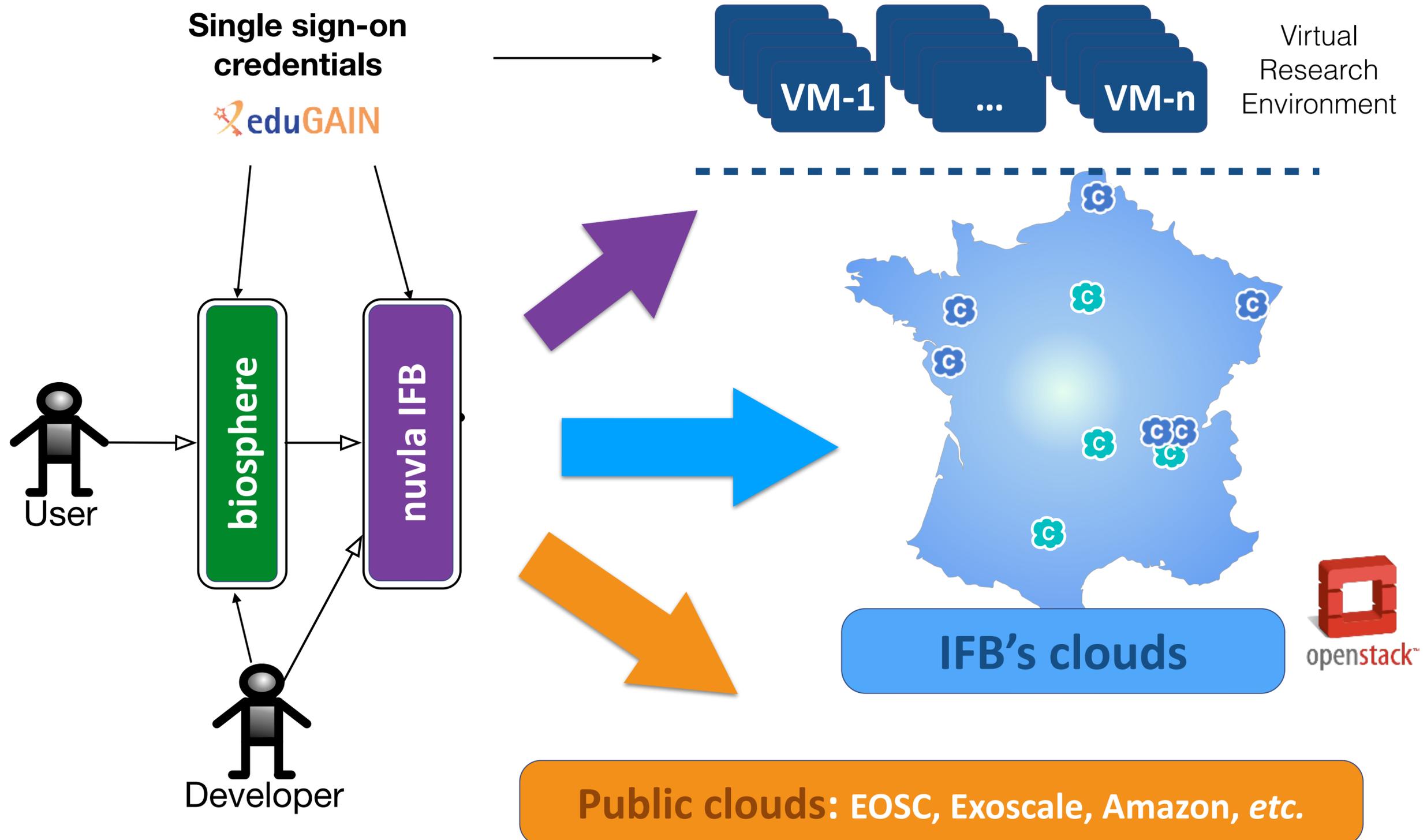
- Documentation en ligne et helpdesk

RAINBio: un catalogue d'appliances (34) prêtes à l'emploi

- Génomique, annotation, bioimage, chemins métaboliques, écologie microbienne...



Architecture du cloud IFB



Ressources du cloud IFB

					Gabarits des VM	Réseau public	Biomaj	vCPU.h (année précédente)	Instances	vCPU	Mémoire (Go)	Stockage (Go)
●	ifb-bilille-thor	Lille	🏠	✉	☰	✔		14 318	8/192	76/192	302/768	na
●	ifb-bird-stack	Nantes	🏠	✉	☰	✖ 🔄		207	0/80	0/80	0/312	na
●	ifb-bistro-iphc	Strasbourg	🏠	✉	☰	✔		442 060	2/100	10/200	20/1024	na
●	ifb-core-cloud*	Lyon	🏠	✉	☰ BigMem	✔	🗄	9 519 166	191/1200	1372/3936	6131/20400	na
●	ifb-genouest-genostack	Rennes	🏠	✉	☰	✖ 🔄	🗄	25 261	2/100	8/120	32/117	254/205000
●	ifb-prabi-girofle	Lyon	🏠	✉	☰	✔		1 669 548	17/150	53/200	106/1024	na

*Cloud par défaut

RAINBio - Appliances bioinformatiques dans le cloud

Appliances, des environnements prêts-à-l'emploi

- Déploiement en 1 clic
- 1 à N machines virtuelles, multi-cloud
- Interfaces usuelles: SSH, HTTP, Desktop
- Définie par les développeurs
- Adaptables par les utilisateurs

Apps disponibles

- Bioinformatique (EDAM), génomique (ChIP-seq, RNA-Seq...), annotation de génome, génomique comparative, bioimageire, chemins métaboliques, métabolomique, écologie microbienne, métagénomique, transcriptomique...

The screenshot shows the RAINBio App Store interface. At the top, it says "RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD" and "Catalogue des appliances bioinformatiques dans le cloud, filtrez-les en utilisant les termes présents dans l'ontologie EDAM, ou en langage naturel." Below this are tabs for "App Store", "Appliances", "Outils", and "Topics". The main area displays a grid of appliance cards, each with a title, a list of tools/technologies, and a play button icon. Some cards include a small video player icon. The appliances shown include: Bacterial genomics (Insyght), COURS LBBE-NGS, Galaxy stable, PathoTRACK, BioPipes, Debian 9, ImageJ-Fiji, Prokka, bistar, EBAME 2018 - Marine Microbiome, integron_finder, RStudio, CentOS, EBAME 2018 - Metabolic Network, Jupyter Notebook, SGE compute cluster, SLURM compute cluster, COURS ENS Lyon NGS 2018, EBAME 2018 - Microbial Ecological Networks, and Maker.



+

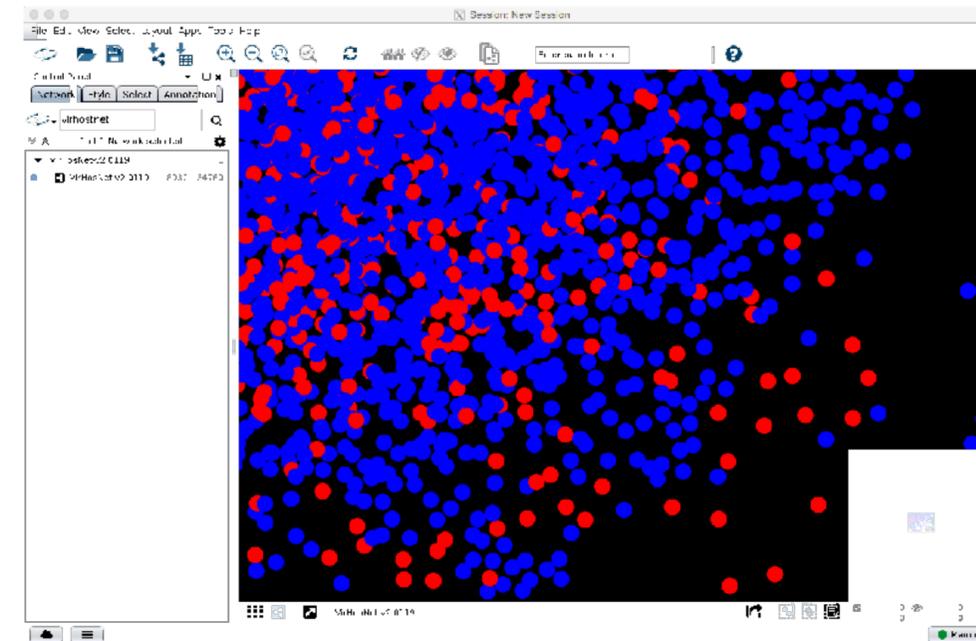


Quelques services actuels

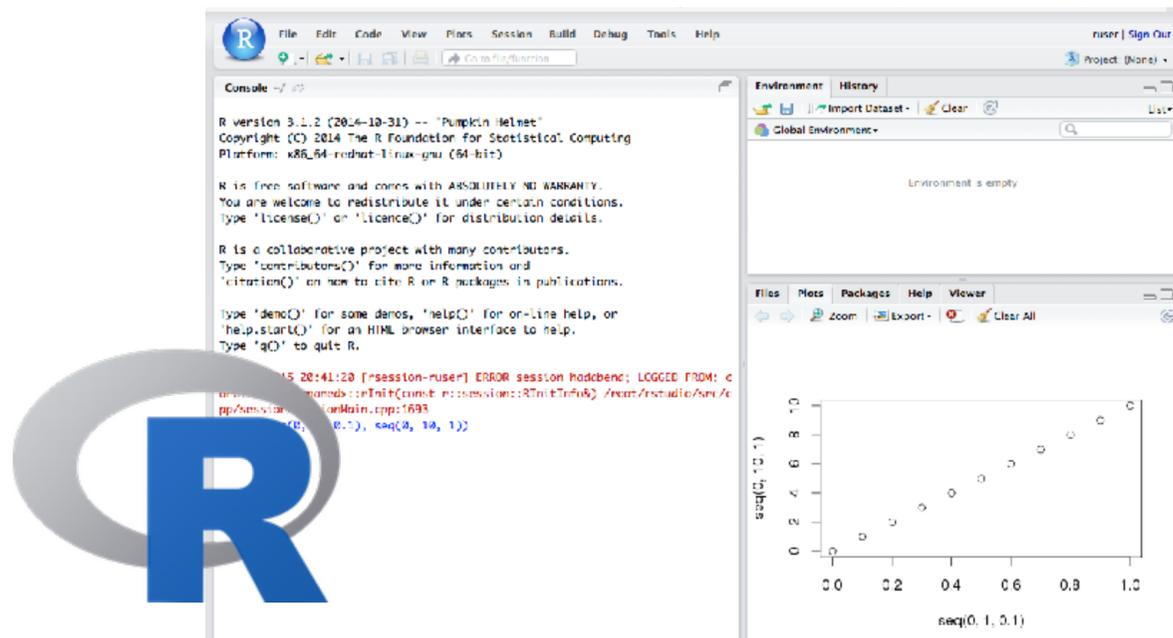
```
ubuntu@machina2fb702a-2b9e-4035-a8e2-1db5734a0dcd:~$ python3.6 --version
Python 3.6.7
ubuntu@machina2fb702a-2b9e-4035-a8e2-1db5734a0dcd:~$ python --version
Python 2.7.15rc1
ubuntu@machina2fb702a-2b9e-4035-a8e2-1db5734a0dcd:~$ ansible --version
ansible 2.7.10
  config file = /etc/ansible/ansible.cfg
  configured module search path = [u'/home/ubuntu/.ansible/plugins/modules', u'/usr/share/ansible/plugins/modules']
  ansible python module location = /usr/lib/python2.7/dist-packages/ansible
  executable location = /usr/bin/ansible
  python version = 2.7.15rc1 (default, Nov 12 2018, 14:31:15) [GCC 7.3.0]
ubuntu@machina2fb702a-2b9e-4035-a8e2-1db5734a0dcd:~$ docker --version
Docker version 18.09.2, build 6247962
ubuntu@machina2fb702a-2b9e-4035-a8e2-1db5734a0dcd:~$ docker compose version
docker-compose version 1.24.0, build 0aa5506
ubuntu@machina2fb702a-2b9e-4035-a8e2-1db5734a0dcd:~$ conda info

active environment : None
shell level : 0
user config file : /home/ubuntu/.condarc
populated config files : /var/lib/miniconda3/.condarc
conda version : 4.5.12
conda-build version : not installed
python version : 3.7.1.final.0
base environment : /var/lib/miniconda3 (writable)
channel URLs : https://conda.anaconda.org/conda-forge/linux-64
               https://conda.anaconda.org/conda-forge/noarch
               https://conda.anaconda.org/bioconda/linux-64
               https://conda.anaconda.org/bioconda/noarch
```

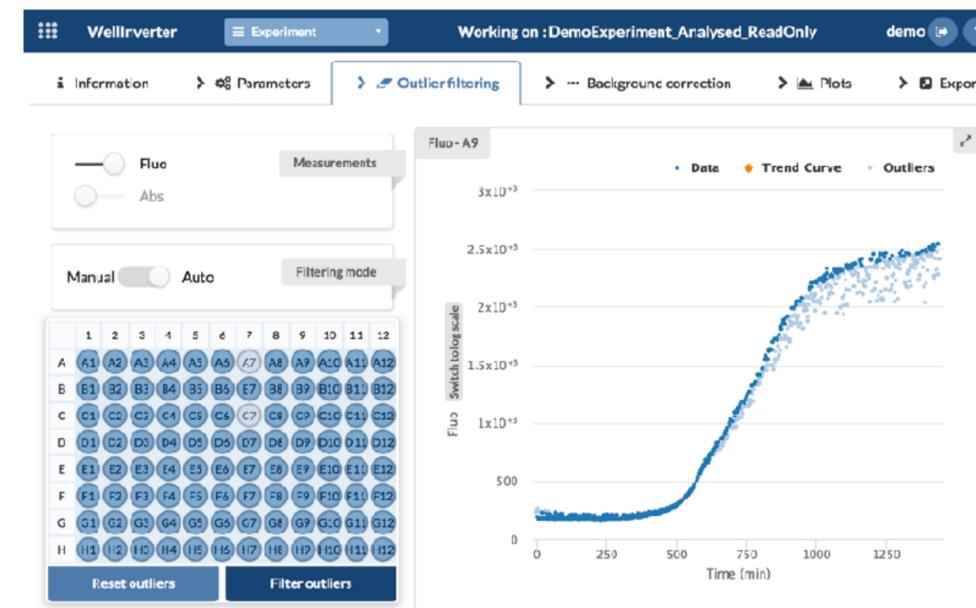
Base Ubuntu 18.04



Cystoscape



Rstudio / Shiny



WellInverter (AAP 2015)

App Jupyter

Description

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

This IFB cloud appliance provides both the Jupyter Notebook and Lab environment (see [explanations](#)). Biosphere's users can switch these environments simply by changing the relative URL of the service from '/tree' to '/lab'.

This Jupyter app is based on the Jupyter Docker Stacks (see [details](#)). By default, this Biosphere app uses the stack `jupyter/datascience-notebook` but users can choose any other existing stack with an Advanced deployment in Biosphere portal.

Main jupyter stacks are:

- `jupyter/minimal-notebook` : Minimally-functional Jupyter Notebook server, Miniconda Python 3.x, Pandoc and TeX Live for notebook document conversion
- `jupyter/r-notebook` : Everything in jupyter/minimal-notebook and popular packages from the R ecosystem.
- `jupyter/scipy-notebook` : Everything in jupyter/minimal-notebook and popular packages from the scientific Python ecosystem.
- `jupyter/datascience-notebook` : Everything in the jupyter/scipy-notebook and jupyter/r-notebook images, plus libraries for data analysis from the Julia, Python, and R communities.
- `jupyter/tensorflow-notebook` : Everything in jupyter/scipy-notebook and popular Python deep learning libraries.

See a detailed list of available Jupyter stacks [there](#).

Domaines associés

Outils

Julia Jupyter R

OS ubuntu 18.04

Caractéristiques

Nom long	Biosphere BASE environment for Jupyter Notebook and Lab			
Version				
Créé.e	16 octobre 2019 17:25			
Dernière mise à jour	15 novembre 2019 14:45			
Temps moyen de déploiement (hh:mm:ss)	jour précédent	semaine précédente	mois précédent	année précédente
Global	--	00:09:15	00:12:11	00:15:55

Crédits

Contact	Support Cloud IFB
Développeurs	Christophe Blanchet IFB-core Jonathan Lorenzo IFB-core

Portail utilisateur multi-cloud

1 2 3 4 5 6 7

 Biosphere RAINBio **myVM** DATA

 Support Admin 

christophe.blanchet@france-bioinformatique.fr (eduGAIN)

CLOUD

Déploiements   

ID	Nom	Début	Groupes	Spéc	Brkr	Cloud	Accès
9451	BioPipes (19.04) reprohack-3	Nov 24	Reprohackathon-3	2 8 120	d5a3	ifb-core-cloud	ssh
9443	Ubuntu 16.04 (16.04.6)	Nov 21	IFB-core	1 4 70	4ed0	ifb-core-cloud	ssh
9421	Ubuntu 18.04 (18.04.2) icli	Nov 21	IFB-core	8 32 470	f25e	ifb-core-cloud	ssh
9386	Ubuntu 18.04 (18.04.2) lega	Nov 19	IFB-core	16 64 920	dc7b	ifb-core-cloud	ssh
9373	Jupyter	Nov 15	IFB-core	1 4 70	e878	ifb-core-cloud	https Params
8189	Ubuntu 18.04 (18.04.2) ada2	Oct 03	IFB-core	4 16 200	4ee1	ifb-core-cloud	ssh

1. Site IFB
2. Accueil Biosphère
3. Catalogue des appliances
4. Tableau de bord
5. Données
6. Informations ponctuelles
7. Liens utiles
8. Paramètres utilisateur

Gestion des usagers et des ressources

Comptes

- **"Se connecter"**
 - ★ Identifiant des organismes (fédération eduGAIN + service Nuvla-IFB)
 - ★ Compte « local » : username/password, interne (django)
 - ★ Compte de service dans nuvla et les clouds (keystone)
- Compte actif si membre d'un groupe valide
 - ★ Permanent (affiliation) ou à durée limitée (formation, projet)

Groupe

- Création et adhésion en ligne
- Actif après validation par les admins
- Permanent (affiliation) ou à durée limitée (formation, projet)

Quota

- Quota utilisateur par défaut : 36 000 vCPU.h
- Allocation supplémentaire
 - ★ CPU/utilisateur : projet moyen, développeur, formateur
 - ★ CPU/groupe : projet lourd
 - ★ BigMem : suivant les sites



Utilisez vos identifiants de votre organisme (CNRS, INRA, Inserm, Universités...)

[Se connecter](#)

Nous utilisons la fédération d'identité européenne eduGAIN. Si votre organisme académique n'est pas dans cette fédération, vous pouvez utiliser un compte local avec votre adresse professionnelle.



Stockage et gestion des données

Stockage disponible pour les utilisateurs

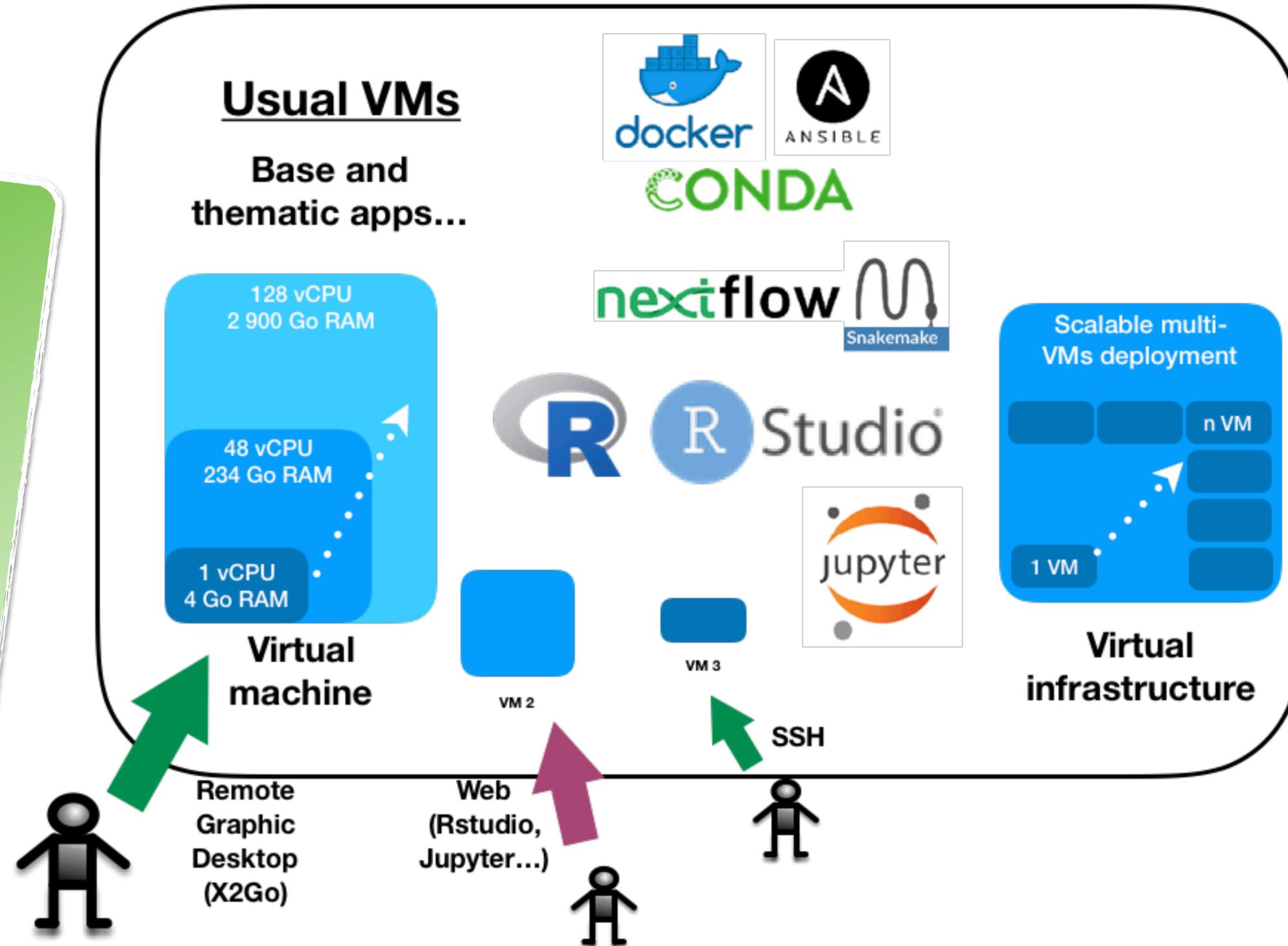
- 4 types (cf. tableau)
- propre à chaque cloud, sans échange inter-sites
- basé sur manila + autofs

Banques publiques

- visibles depuis la page DATA de Biosphère (BioMaJ)

Stockage	Droits	VM	Répertoire	Comm.
Banques publiques	ro	Toutes	/ifb/data/public	BioMaJ
Ephémère	rw	Interne à la VM	/ifb/data/local	Variable par cloud
Partage usager	rw	Toutes VM de l'utilisateur	/ifb/data/mydisk	Manila
Partage groupe	rw	Toutes VM des membres du groupe	/ifb/data/<nom_groupe>	Manila

Environnements modulaires



Gabarits des VM

ifb.m4.small	1 vCPU, 4Go RAM, 70Go local disk
ifb.m4.large	2 vCPU, 8Go RAM, 120Go local disk
ifb.x1e.large (BigMem)	2 vCPU, 48Go RAM, 20Go local disk
ifb.m4.xlarge	4 vCPU, 16Go RAM, 220Go local disk
ifb.x1e.xlarge (BigMem)	4 vCPU, 96Go RAM, 20Go local disk
ifb.m4.2xlarge	8 vCPU, 32Go RAM, 470Go local disk
ifb.x1e.2xlarge (BigMem)	8 vCPU, 192Go RAM, 20Go local disk
ifb.m4.4xlarge	16 vCPU, 64Go RAM, 920Go local disk
ifb.x1e.4xlarge (BigMem)	16 vCPU, 384Go RAM, 20Go local disk
ifb.m4.6xlarge	24 vCPU, 128Go RAM, 1.3To local disk

Déployer une appliance

ifb Biosphere RAINBio myVM DATA

RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD

Catalogue des appliances bioinformatiques dans le cloud, filtrez-les en utilisant les termes présents

App Store Appliances Outils Topics

BioPipes

- 1 ⭐ bioconda, cwltool, Docker, Docker Compose, Nextflow, Snakemake
- 2 ⚡ Bioinformatics, Informatics, Workflows
- 3 🛠️ OWS

1. Mise en favori
2. Lancement rapide
3. Déploiement avancé

Configurer le déploiement d'une appliance

Déploiement de l'appliance "BioPipes"

Name

Cloud

Gabarit d'image cloud

Annuler Lancer

BioPipes

- 1 ⭐ En favori
- 2 ⚡ Lancer
- 3 🛠️ Configurer

Biosphere-commons appliance providing standard tools for bioinformatics pipelines : for example bioconda, cwltool, Docker, Docker Compose, Nextflow, Snakemake

Bioinformatics, Informatics, Workflows

Support Cloud IFB

Se connecter à une VM

Accès

ssh Params

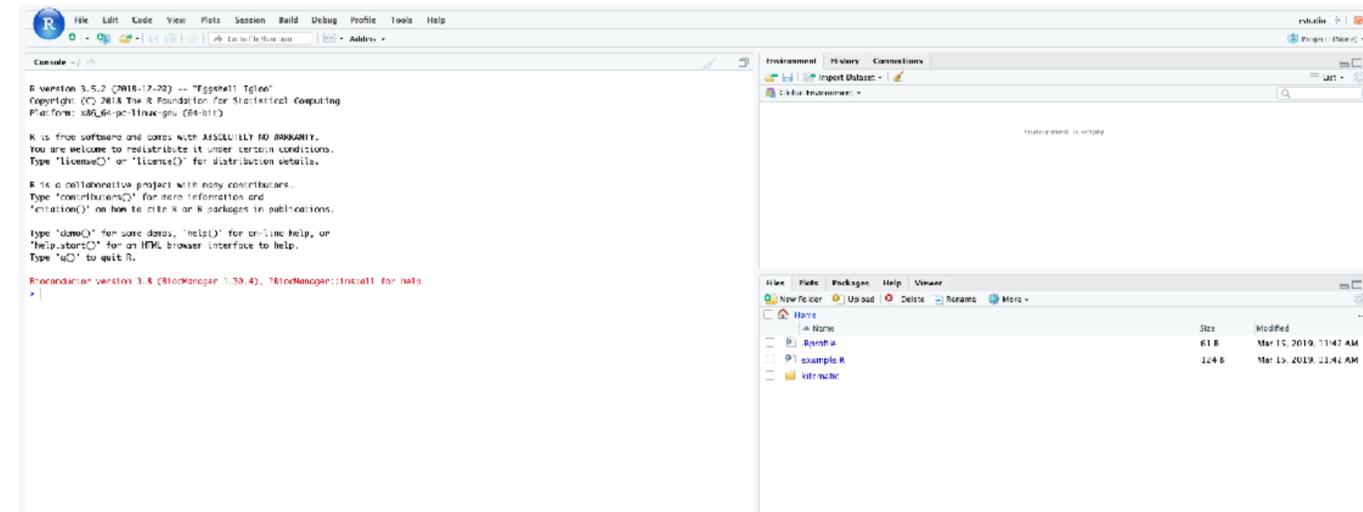
ssh

ssh https

HTTP/S

Bureau X2Go

HSS



Parameters

nom	value
X2Go Username	ubuntu
X2Go Host	134.158.247.87
X2Go Session Type	XFCE

```
lorenzo ~ ubuntu@machine72a5cac6-a20e-4063-b782-12389e891b6a: ~ -- ssh -A -p 22...
ifb-jonathan:~ lorenzo$ ssh -A -p 22 ubuntu@134.158.246.168
Warning: Permanently added '134.158.246.168' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 16.04.5 LTS (GNU/Linux 4.4.0-139-generic x86_64)

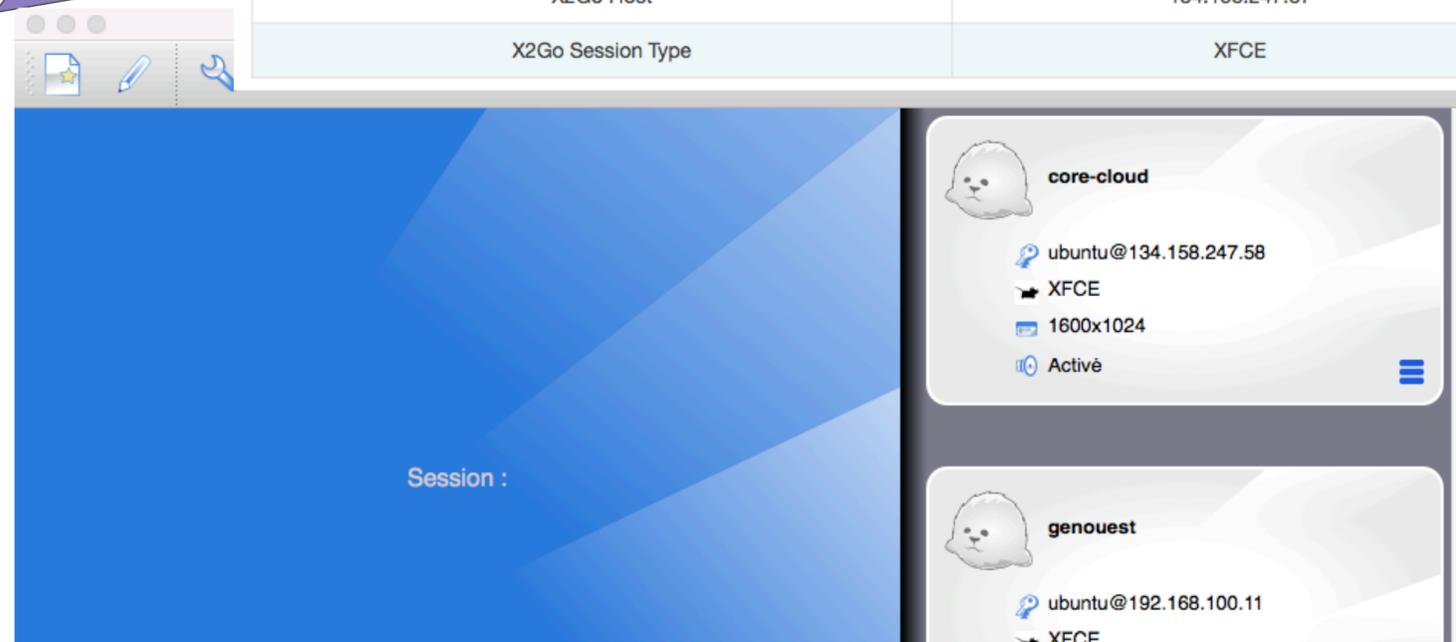
 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud

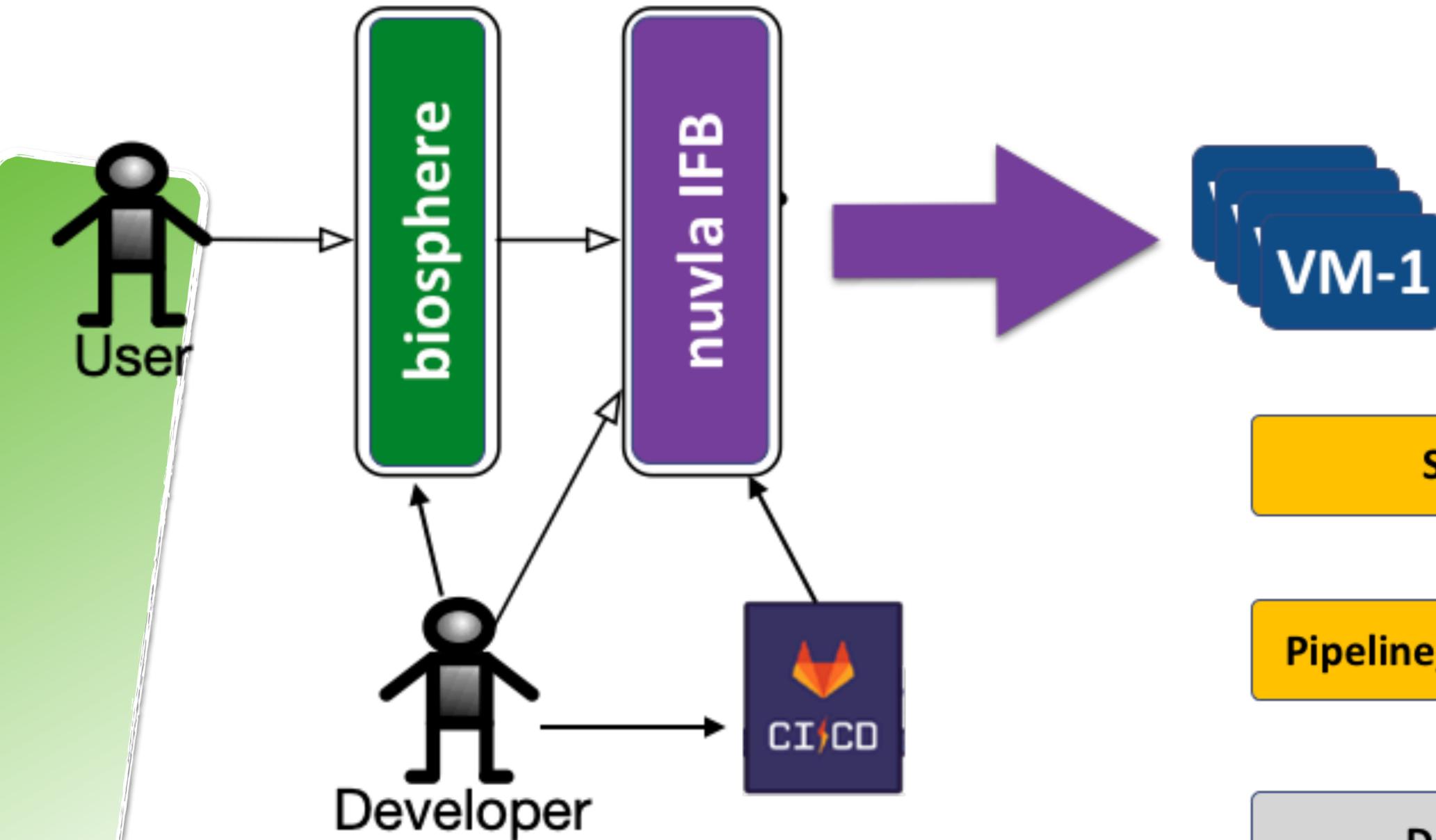
30 packages can be updated.
0 updates are security updates.

New release '18.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

*** System restart required ***
Last login: Fri Mar 15 14:53:28 2019 from 134.214.33.85
ubuntu@machine72a5cac6-a20e-4063-b782-12389e891b6a:~$
```



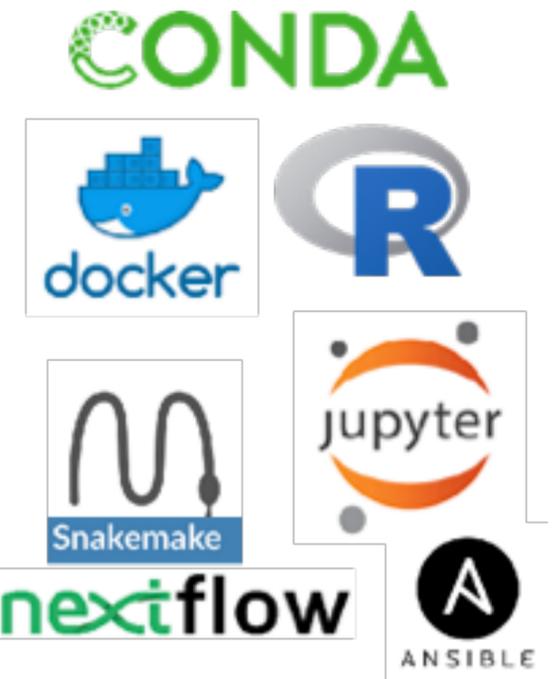
Développement d'apps



Recettes de déploiement

- Dépôts git, integration continue
- Conteneurs docker, référencés dans des catalogues
- Workflows, avec docker
- Jupyter Notebooks
- Scripts R...

Set of tools



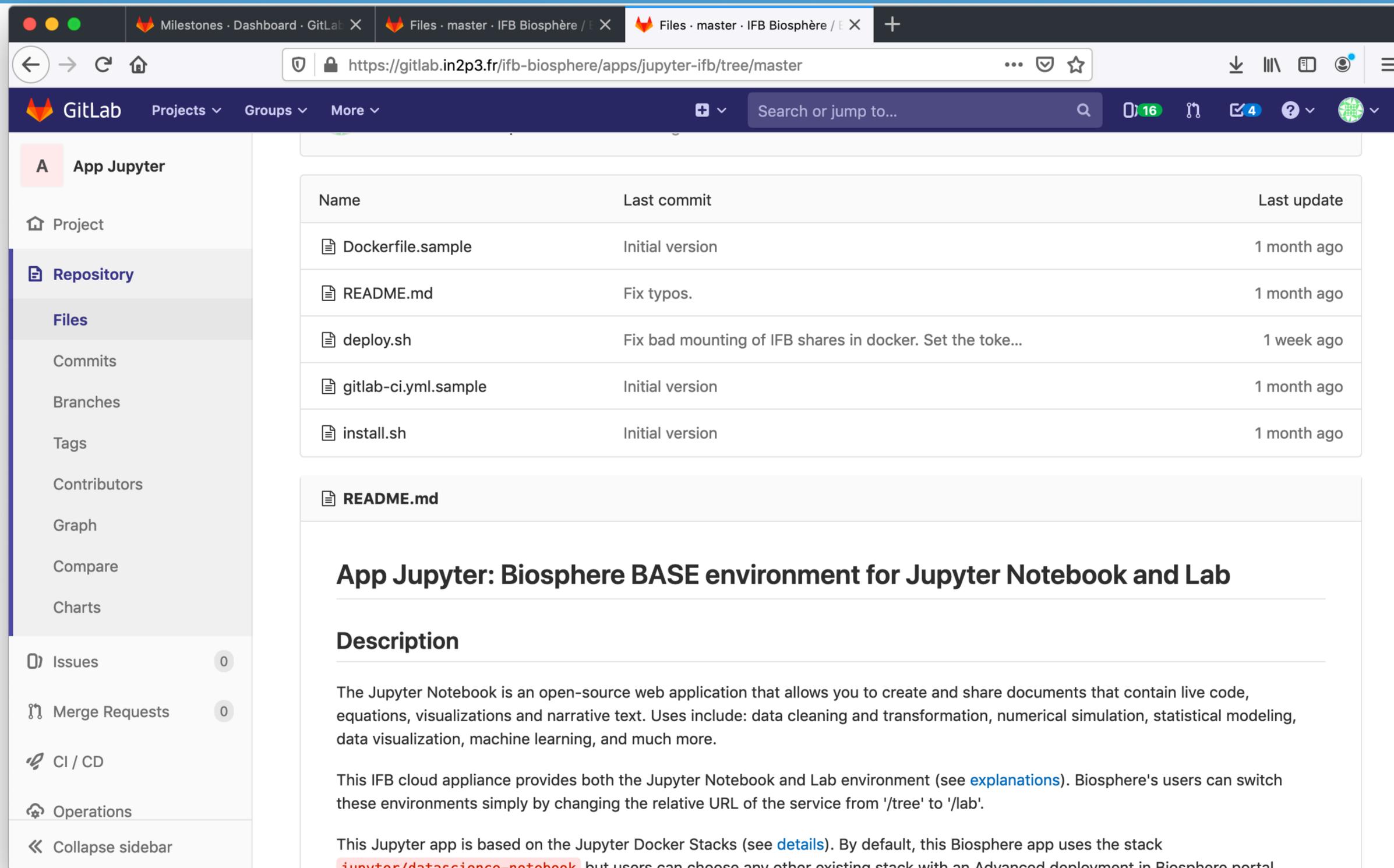
Pipeline, workflow, recipe

Docker (opt.)

Operating System

CentOS, Debian, Ubuntu

App Jupyter, recette git



The screenshot shows a GitLab web interface for a repository named 'App Jupyter'. The browser address bar shows the URL: <https://gitlab.in2p3.fr/ifb-biosphere/apps/jupyter-ifb/tree/master>. The left sidebar contains navigation options: Project, Repository, Files, Commits, Branches, Tags, Contributors, Graph, Compare, Charts, Issues (0), Merge Requests (0), CI / CD, and Operations. The main content area displays a table of files and their commit history.

Name	Last commit	Last update
 Dockerfile.sample	Initial version	1 month ago
 README.md	Fix typos.	1 month ago
 deploy.sh	Fix bad mounting of IFB shares in docker. Set the toke...	1 week ago
 gitlab-ci.yml.sample	Initial version	1 month ago
 install.sh	Initial version	1 month ago

Below the table, the 'README.md' file is selected, showing its content:

App Jupyter: Biosphere BASE environment for Jupyter Notebook and Lab

Description

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

This IFB cloud appliance provides both the Jupyter Notebook and Lab environment (see [explanations](#)). Biosphere's users can switch these environments simply by changing the relative URL of the service from '/tree' to '/lab'.

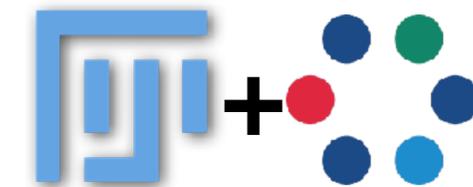
This Jupyter app is based on the Jupyter Docker Stacks (see [details](#)). By default, this Biosphere app uses the stack `jupyter/datascience-notebook` but users can choose any other existing stack with an Advanced deployment in Biosphere portal.

Apps de base

- **Maintenues par l'IFB**
- **Ubuntu, CentOS, Debian**
 - ★ Docker + Compose, Ansible, Bioconda
 - ★ Desktop (MATE)
- **BioPipes**
 - ★ Nextflow, SnakeMake, CWLtool, Conda + Bioconda, Docker + Compose
- **R, Rstudio**
 - ★ BiocManager, HDF5, ...
- **Shiny**
- **Jupyter Notebook**
 - ★ kernel Python & R
- **ImageJ-Fiji**
 - ★ connecteur OMERO
- **Cytoscape**
 - ★ connecteurs NDexBio, PSICQUIC
- **Clusters SGE, Slurm...**



conda



Bonnes pratiques de développement cloud

Installation d'outils

- code source (git)
- paquet linux (apt/yum)
- paquet pip
- paquet conda (bioconda)
- conteneur (docker hub, registres locaux, biocontainer)
- paquet R (CRAN, git)

Configuration d'un environnement complet

- script shell
- environnement conda
- environnement docker (docker-compose)
- playbook ansible
- script R
- Jupyter notebook
- workflow nextflow, snakemake, CWL (+docker/conda)

Les Ateliers du Cloud Biosphère

Usage Avancé et Développement d'Appliances

- Découvrir l'offre de service cloud de l'IFB
- Utilisateurs avancés, développeurs, formateurs

Programme type (2 jours)

- Jour 1, Présentation de la Fédération de Clouds IFB-Biosphère
 - ★ Introduction et présentation de l'architecture de IFB-Biosphère
 - ★ Présentation des bonnes pratiques d'utilisation avancée de IFB-Biosphère
- Jour 2, Hackathon Développement d'Appliances Cloud dans IFB-Biosphère
 - ★ Travail en commun sur les projets des participants
 - ★ Intégration de vos outils et pipelines,

Sessions:

- Session de Lyon, PRABI-LBBE, 12-13 mars 2019, 18 part.
- Session de Rennes, GenOuest, 19-20 mars 2019, 19 part.
- D'autres à venir

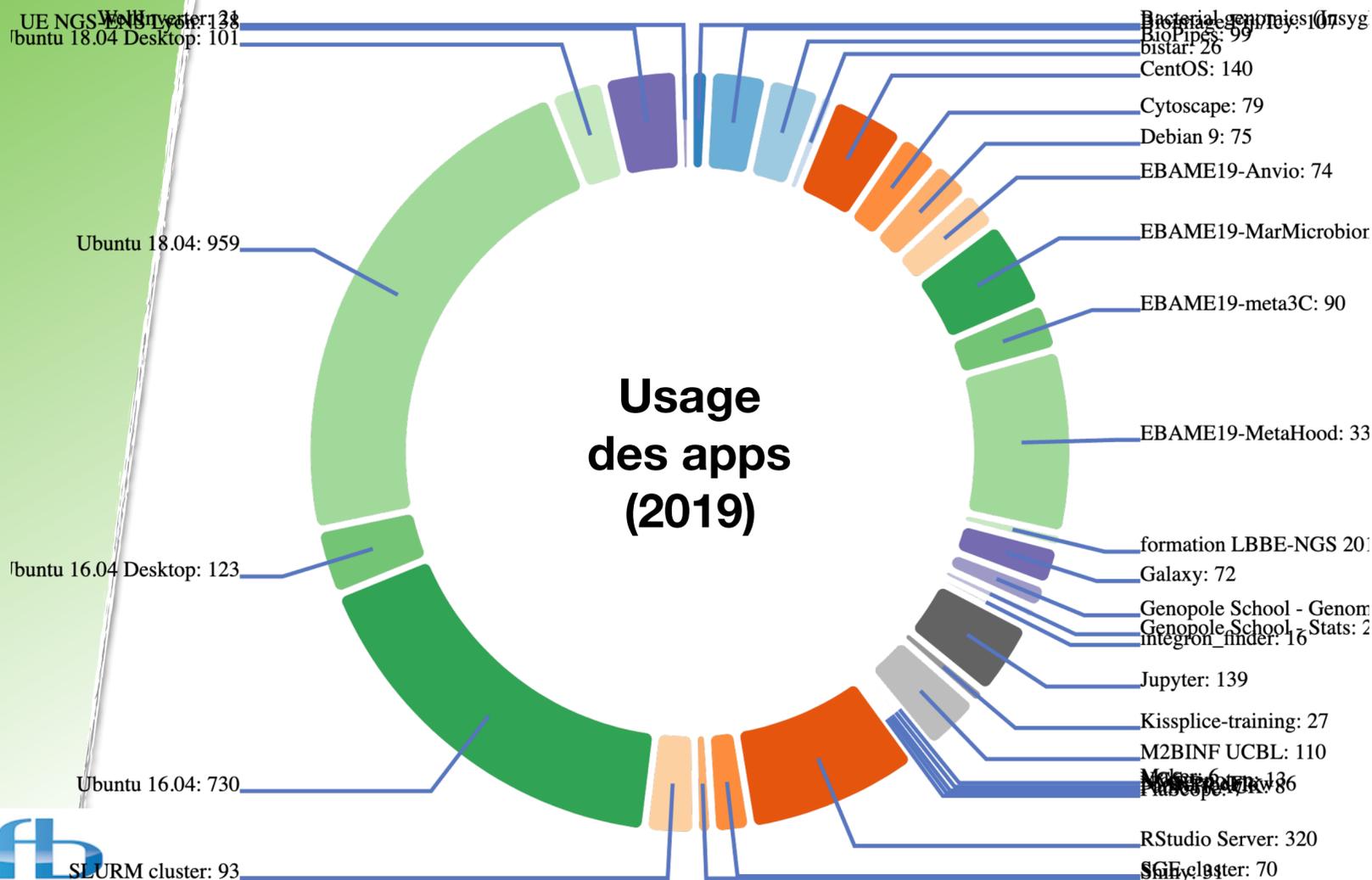
Exemples de formations scientifiques hébergées

<u>Date</u>	<u>Intitulé</u>	<u>Organisateur</u>	<u>Partici-pants</u>	<u>Ancienneté sur le cloud</u>
Sept. 2018	Elixir Training: Plant Genome Assembly and Annotation	IFB-core, CIRAD	25	-
Juin 2019	Summer school Genopole	Genoscope, Evry	20	3 ans
Sept. 2019	Bioinformatique pour le traitement de données de séquençage	LBBE, Lyon & CNRS Entreprise	12	2 ans
Oct. 2019	EBAME Workshop on Computational Microbial Ecogenomics	IUEM, Brest	50	2 ans
Oct. 2019	Master AMI2B	Université Paris Saclay	15	5 ans
Nov. 2019	ReproHackathon-3 (GDR MaDICS)	LRI, IFB-core, CIRAD	20	3 ans
2019-20	Master Biologie ENS-Lyon - Practicals in next generation sequencing	ENS-Lyon	20	4 ans
2019-20	Master 2 de bioinformatique	Université Claude Bernard Lyon1	20	-

Usage

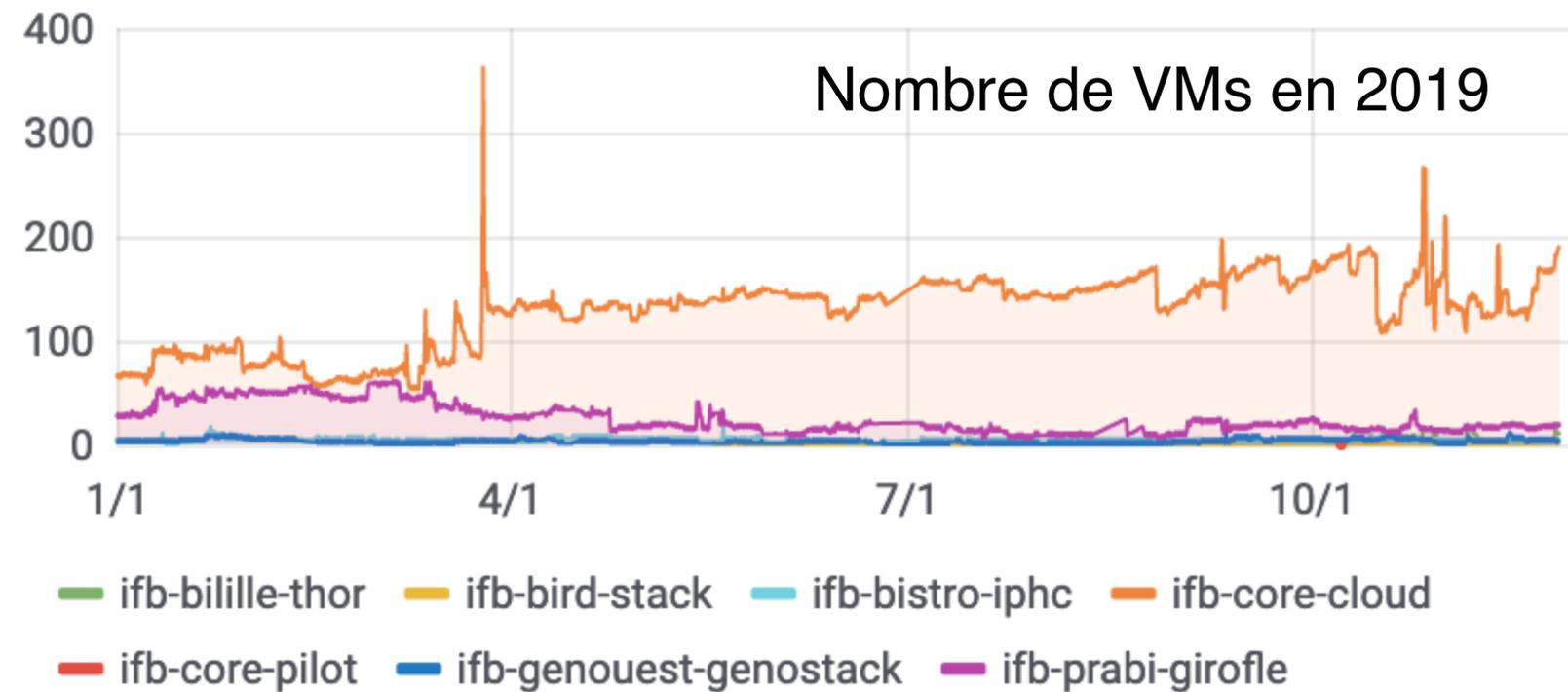
Utilisateurs

- 350 activés
- 95 groupes actifs (+12 expirés)
- 9500+ déploiements
- 200+ outils



Formations hébergées par le cloud

- 10 en 2019 (226 élèves)
- 12 en 2018 (241 élèves)
- 30+ sur 2014-2017
- Certaines récurrentes depuis 5 ans



Conclusions

IFB fournit

- de **nombreux services** de traitements des **données biologiques**,
- une **infrastructure distribuée** (cluster et cloud),
- un accès gratuit pour un premier niveau de besoins (extension possible suivant différents critères).

Le Cloud IFB permet

- aux **scientifiques** d'utiliser les différents clouds d'une manière uniformisée et simplifiée avec leurs identifiants institutionnels,
- aux **développeurs** de construire de nouveaux environnements de traitement avec les outils système communément utilisés (apt/yum, pip, conda, docker, ansible, git...),
- aux **administrateurs** de site des outils de gestion avancés et complémentaires des outils courants Openstack,
- un support de la **reproductibilité** en sciences de la vie.

Remerciements

- IFB, membres, admins des sites et développeurs d'appiances
- CC-IN2P3, hébergement et administration de core-cloud
- Financements IFB
 - ★ French program PIA INBS 2012
 - ★ EU H2020 CYCLONE, EXCELERATE, EMBRIC, EUCANcan...



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

