



Le cloud de l'IFB : une ressource fantastique mais souvent insoupçonnée

Carine Rey

Club Bioinfo #6 - 5 décembre 2019

Table of contents

1. Introduction
2. Les différentes appliances disponibles
3. En pratique
4. Conclusion

Introduction

Qu'est ce que l'IFB ?

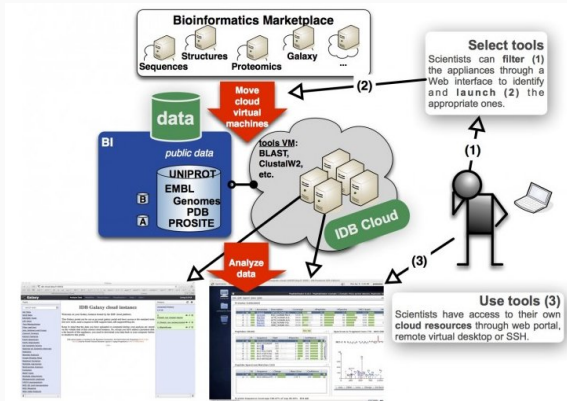
“ L'Institut Français de Bioinformatique (IFB) est l'infrastructure nationale de service en bioinformatique créée dans le cadre du programme national des «Investissements d'Avenir». Elle mutualise, soutient et coordonne le développement des ressources et des activités de support à la recherche de plateformes de bioinformatique dépendant d'organismes publics de recherche (CNRS, INRA, INRIA, CEA et INSERM), des universités, du CIRAD, et des Instituts Pasteur et Curie. ”



Qu'est ce que le cloud de l'IFB ?

=> Parmi d'autres services, il propose un service de calcul :

- “**Le Cloud**, un service offrant une expérience privilégiée pour des travaux liés à des problématiques de calcul séquentiels, d'étude et d'exploitation d'un ensemble d'outils et de données.”



Comment ça marche, en bref.

1. Création d'un compte
2. Déploiement d'un VM basée sur un modèle (une appliance)
3. Accès par internet
 - **Graphique** via une fenêtre (X2GO) ou votre navigateur (http)
 - **Ligne de commande** (ssh)
4. Envoi des données
5. Utilisation (1h -> 1j -> 1 semaine)
6. Récupération des données
7. Extinction (!)

[DEMO]

Documentation :

<https://ifb-elixirfr.github.io/biosphere/>

Le tableau de bord : myVM

The screenshot displays the 'myVM' Cloud dashboard. At the top, there's a navigation bar with 'Biosphere', 'RAINBIO', 'myVM', and 'DATA' tabs. The main content area is titled 'CLOUD' and contains two sections:

- Déploiements**: A table listing various cloud deployments with columns for ID, Nom, Début, Specification, Cloud, and Accès.
- Applications et déploiements favoris**: A section for favorite applications, including sub-tabs for 'Déploiements récemment terminés' and 'Quota'. It lists specific applications like 'HPCLabo Server' and 'GE MG5-EMG Lyon (Nov 2019)'.

At the bottom of the dashboard, there are logos for partner organizations: cea, CNRS, INRA, Inria, Inserm, elixir, and a circular logo for the French Republic.

ID	Nom	Début	Specification	Cloud	Accès
9929	Superb	7 ↑ Dec 04 2019, 17h22	1 4 70	IB-core-cloud	Info Params
9922	Ubuntu 18.04 Desktop (18.04)	7 ↑ Dec 04 2019, 16h36	2 8 120	IB-core-cloud	Info Params
9920	Ubuntu (18.07)	7 ↑ Dec 04 2019, 16h34	1 4 70	IB-core-cloud	Info Info Params
9818	HPCLabo Server	7 ↑ Dec 04 2019, 15h07	2 8 120	IB-core-cloud	Info Params
8205	GE MG5-EMG Lyon (Nov 2019) Analyse Seq	7 ↑ Oct 08 2019, 08h46	16 64 920	IB-core-cloud	Info Params

ID	Nom	Der. dév.	Paramétrage
16	HPCLabo Server	↑ Dec 04 2019, 15h07	2 8 120
58	GE MG5-EMG Lyon (Nov 2019)	↑ Dec 03 2019, 14h42	16 64 920
120	Ubuntu 18.04 (18.04.2)	↑ Oct 14 2019, 17h28	1 4 70

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

Quelles sont les avantages de l'utiliser ?

- Facile d'utilisation (Connexion facile, Gestion facile avec le Tableau de bord)
- Grande capacité (More than 5,200 vCPU and 26 TB RAM)
- Pas de queue, les VMs sont directement disponibles (Pour le moment)
- Accès direct aux bases de données publiques de références (RefSeq, nr, Uniprot, PDB, ...)
- Taille des VMs modulables (up to 48 vCPU and 250 GB of RAM per VM)
- Accès à des VMs avec une grande RAM (up to 3 TB RAM)

Gabarits des VM

lfb.m4.small	1 vCPU, 4Go RAM, 70Go local disk
lfb.m4.large	2 vCPU, 8Go RAM, 120Go local disk
lfb.x1e.large (BigMem)	2 vCPU, 48Go RAM, 20Go local disk
lfb.m4.xlarge	4 vCPU, 16Go RAM, 220Go local disk
lfb.x1e.xlarge (BigMem)	4 vCPU, 96Go RAM, 20Go local disk
lfb.m4.2xlarge	8 vCPU, 32Go RAM, 470Go local disk
lfb.x1e.2xlarge (BigMem)	8 vCPU, 192Go RAM, 20Go local disk
lfb.m4.4xlarge	16 vCPU, 64Go RAM, 920Go local disk
lfb.x1e.4xlarge (BigMem)	16 vCPU, 384Go RAM, 20Go local disk
lfb.m4.6xlarge	24 vCPU, 128Go RAM, 1.3To local disk
lfb.m4.8xlarge	32 vCPU, 192Go RAM, 1.8To local disk
lfb.x1e.8xlarge (BigMem)	32 vCPU, 768Go RAM, 20Go local disk
lfb.m4.12xlarge	48 vCPU, 234Go RAM, 2.6To local disk
lfb.x1e.16xlarge (BigMem)	64 vCPU, 1.5To RAM, 20Go local disk
lfb.x1e.32xlarge (BigMem)	128 vCPU, 2.9To RAM, 20Go local disk

Quelles sont les avantages de l'utiliser ?

- Machines sous linux indépendantes de votre machine (Initiation à la bioinfo facile et sans risque)
- Plus facile à prendre en main qu'un cluster de calcul
- Droits administrateurs sur vos machines
- Pratique pour faire de petits/moyens calculs
- De nombreux environnements bioinformatiques sont déjà pré-installés
- Pratique pour mettre en place des formations

Les différentes appliances disponibles

Les différentes appliances disponibles : le catalogue RainBio

ifb Biosphere RAINBIO API/DB DATA Support 06 Se connecter

RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD

Catalogue des appliances bioinformatiques dans le cloud, filtrez-les en utilisant les termes présents dans l'ontologie EDAM, ou en langage naturel.

App Store (34) Appliances Outils Topics

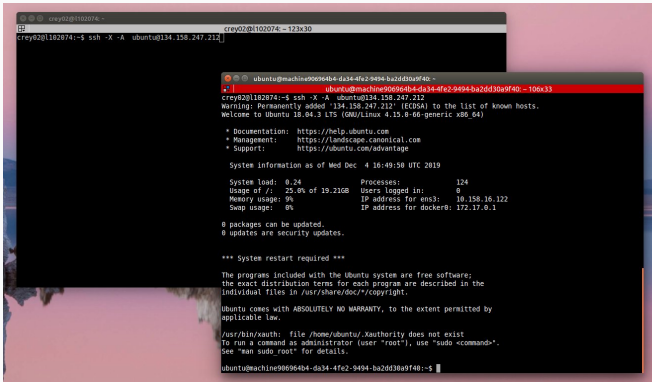
Bioinformatic genomics (Bioright) <ul style="list-style-type: none">BLAST, HMER, R, ImageJ, SGE - GoFEngine, Ubuntu, Web interfaceProtein tools and structural domains, Sequence comparison, Sequence visualization	Bioinformatic FISH/seq <ul style="list-style-type: none">Analyse, bioconda, Bureau virtual, DoFEngine, X2Go, XFCEBioinformatics, Informatics, Data visualization, Visualization	BioPipes <ul style="list-style-type: none">bioconda, cwltool, Docker, Docker CoFEngine, Nextflow, ProkaryoteInformatics, Bioinformatics, Workflow	Bitator <ul style="list-style-type: none">bioconda, bowtie2, FastQC, SnakemakefileBioinformatics, Sequence alignment, Workflow, Sequence visualization	CentOS <ul style="list-style-type: none">Analyse, bioconda, DockerInformatics, Bioinformatics	Cytoscape <ul style="list-style-type: none">Bureau virtual, Cytoscape, X2Go, KXFCEBioinformatics, Data visualization, Visualization, Visualization
Debian 9 <ul style="list-style-type: none">Analyse, bioconda, DockerBioinformatics, Informatics	EBAME19-AncBio <ul style="list-style-type: none">AnalyseGenomics, Omics, Management, PVisualization	EBAME19-MetMicrobiome <ul style="list-style-type: none">Miseeq2, R, RStudio, R - iSymeData architecture, analysis and design Sequence analysis, Sequen	EBAME19-met3C <ul style="list-style-type: none">bioPython, MisaPilo, Numpy, pandasoFEngine, SisyphHygenetics, Molecular genetics, Proteoomics, Sequence analysis	EBAME19-MetFood	formation.IBBE.NDS.2019 <ul style="list-style-type: none">Alysis, BEDTools, Bioinformatics, BUCSD, FastQC, HISAT2, SV - IntegratBioinformatics, Computational
Galaxy <ul style="list-style-type: none">bioconda, Docker, Galaxy portalInformatics, Bioinformatics, CompoFEngine, Visualization, Visualization	Genopole School - Genome Variations <ul style="list-style-type: none">bioTools, BEDTools, BWA, Canvas, FastQC, GATK, IGV - Integrative GenooFEngine, Informatics, Genetic variation, Visualization, Visualization	Genopole School - Stats <ul style="list-style-type: none">DESeq2, RStudio, R - iSyme, Webi InterfaceoFEngine, Data architecture, Data archiecture, Visualization, Visualization	Integron_finder <ul style="list-style-type: none">HMER, Internal, prodigalProtein tools and structural domains, Sequence comparison, Sequence visualization	Jupyter <ul style="list-style-type: none">Julia, Jupyter, RComputer science, Mathematics, Datavisualization, Visualization, Visualization	Kiosplice-training <ul style="list-style-type: none">Bioinformatics, Kiosplice, STARBioinformatics, Computational
MZBINF UCBL <ul style="list-style-type: none">en3, JupyterMathematics, Statistics and probability, Data visualization, Outils	Maker <ul style="list-style-type: none">BLAST, E-annot, RepeatMaskerGenome annotation, Functional genomics, Genomics	NGSgenotyp <ul style="list-style-type: none">bioPython, bowtie2, cap1, etd, FastQC, MultiSeq, muscle, Numpy, phantHygenetics, Protein structure analysis, Data visualization, Visualization	NMRProcFlow <ul style="list-style-type: none">NMRProcFlowMetabonomics	PathTRACK <ul style="list-style-type: none">Centrify, Diamond, FastQC, python, R, SciPost, Web interfaceInformatics, transcriptomics, Bioinformatics, Data visualization, Visualization	PicScope <ul style="list-style-type: none">Centrify, PicScope, SPadesMetagenomics, Sequence analysis, Visualization, Microbiology, Micro
RStudio Server <ul style="list-style-type: none">Bioinformatics, HDF5, R, RStudio, Web interfaceMathematics, Mathematics, Computational Biology, Visualization, Visualization	SGE cluster <ul style="list-style-type: none">SGE - GridEngine, UbuntuInformatics	Shiny <ul style="list-style-type: none">R, ShinyData architecture, analysis and design Bioinformatics, Data visuali	SLURM cluster <ul style="list-style-type: none">SlurmInformatics	Ubuntu 16.04 <ul style="list-style-type: none">Analyse, bioconda, Docker, UbuntuBioinformatics, Informatics	Ubuntu 16.04 Desktop <ul style="list-style-type: none">Analyse, bioconda, Bureau virtual, DoFEngine, X2Go, XFCEBioinformatics, Informatics
Ubuntu 18.04 <ul style="list-style-type: none">Analyse, bioconda, DockerBioinformatics, Informatics	Ubuntu 18.04 Desktop <ul style="list-style-type: none">Analyse, bioconda, Bureau virtual, DoFEngine, X2Go, XFCEBioinformatics, Informatics	UE NGS-ENS Lgen <ul style="list-style-type: none">BEDTools, BLAST, Bowtie, bowtie2, Clustal Omega, cmake, DESeq2, PaSequence alignment, Informatics, Nananotechnology, Omics, Visualization, Visualization	WellInverter <ul style="list-style-type: none">python3, Web interface, WellInverterComputer science, Mathematics, Genevisualization, Visualization, Visualization		

Le code couleur reste le même pour une même appliance.

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

- Filtrage possible par outils disponibles en cliquant sur l'onglet "Outils"

Ubuntu - SSH



```
crey02@1102074 ~
crey02@1102074 ~$ ssh -X -A ubuntu@134.158.247.212
crey02@1102074:~$ ssh -X -A ubuntu@134.158.247.212
Warning: Permanently added '134.158.247.212' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-66-generic x86_64)

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

System information as of Mon Dec 4 16:49:50 UTC 2019

System load: 0.24          Processes:             124
Usage of /:  25.0% of 19.21GB  Users logged in:      0
Memory usage: 9%          IP address for ens3:  10.150.16.122
Swap usage:  0%           IP address for docker0: 172.17.0.1

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

*** System restart required ***

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

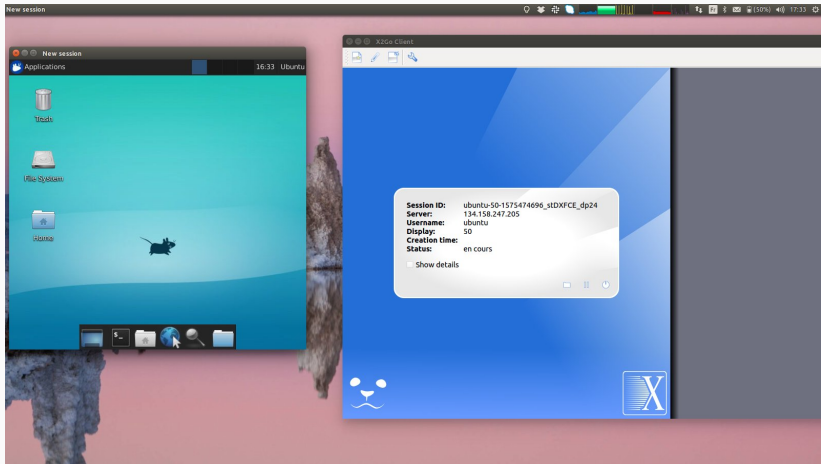
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

/usr/bin/xauth: file /home/ubuntu/.Xauthority does not exist
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@machine906964b4-d34-4fe2-9494-ba2d030a9f40:~$
```

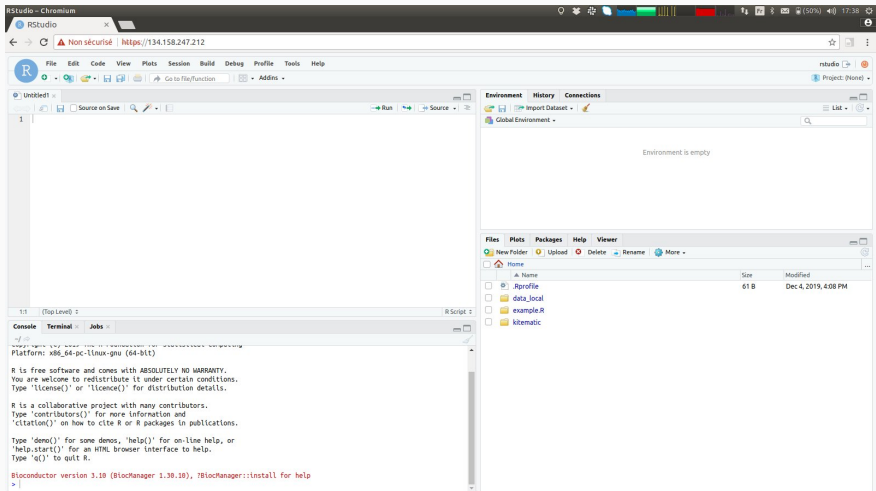
- Besoin de créer des clés ssh :
https://ifb-elixirfr.github.io/biosphere/vm_connect
"See: Annex - Configure your SSH parameters"

Ubuntu - X2GO



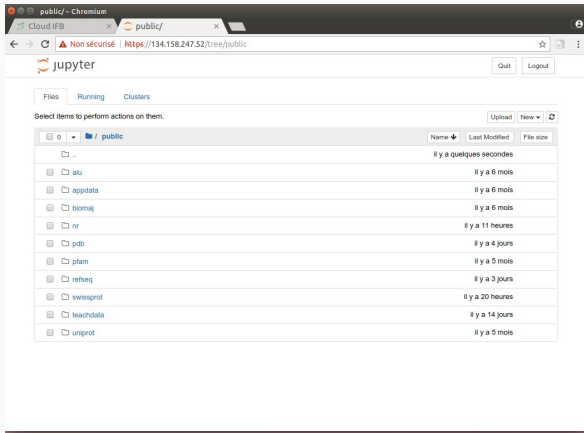
- Besoin de clés ssh et d'installer X2GO (quelque soit votre système d'exploitation, Windows, Mac, Linux):
https://ifb-elixirfr.github.io/biosphere/vm_connect
"See: Connecting with X2Go to a VM"

- De nombreux outils déjà pré-installés (vérifier dans le catalogue)
- Installation de nouveaux outils facilitée
 - Conda
 - Docker
- Portabilité de vos pipelines Nextflow



Accès par votre navigateur (pas besoin de clés ssh) :

- attention firefox, n'est parfois pas compatible (utiliser Chrome)
- il faut accepter l'avertissement de sécurité



Accès par votre navigateur (pas besoin de clés ssh) :

- attention firefox, n'est parfois pas compatible (utiliser Chrome)
- il faut accepter l'avertissement de sécurité

En pratique

Comment s'inscrire ?

1. aller sur <https://biosphere.france-bioinformatique.fr/>
2. en haut à droite cliquer sur **se connecter**
3. cliquer à nouveau sur **se connecter**
4. cliquer sur **accepter les conditions**
5. sélectionner **ENS de Lyon**
6. s'identifier avec votre **login et votre mot de passe ENS**
7. Remplir les informations **Nom, Prénom, ville et code postale** et laisser par défaut les autres informations et accepter
8. Vous aller avoir une nouvelle page, en haut à droite cliquer sur l'**icône petit bonhomme** puis **groupe**
9. Cliquer sur **Rejoindre un groupe** dans les onglets en haut à gauche
10. Chercher le goupe *LBMC (Laboratoire de Biologie et Modélisation de la Cellule - UMR 5239)*, 5239 et cliquer sur le **bouton +** pour candidater
11. et c'est tout bon ! Il faut maintenant que Laurent (?) valide votre candidature.

Comment l'utiliser ?

Documentation très bien faite :

<https://ifb-elixirfr.github.io/biosphere/>

1. Se connecter

<https://ifb-elixirfr.github.io/biosphere/signin>

2. Déployer une VM

https://ifb-elixirfr.github.io/biosphere/vm_deploy

3. Se connecter à une VM

https://ifb-elixirfr.github.io/biosphere/vm_connect

4. Transférer/Récupérer vos données vers/d'une VM

<https://ifb-elixirfr.github.io/biosphere/data>

Des formations sont organisées chaque année à Lyon (à la doua) pour apprendre à utiliser le cloud de l'IFB et/ou aller plus loin (création de ses appliances)

<https://ifb-elixirfr.github.io/biosphere/ateliers>

Quelques conseils

- Bon accès internet (surtout avec X2go)
- Quotas par défaut faibles mais rallonge possible sur projet
- C'est une ressource mutualisée donc bien penser à éteindre les machines inutilisées
- Pas de sauvegarde de vos données donc bien garder des copies en local
- Bien pratiquer les "bonnes pratiques" pour ne pas tout perdre s'il y a un problème technique
- Support technique excellent, ne pas hésitez à les contacter

Conclusion

- Avoir accès à de grandes ressources de calculs
- Tester/apprendre l'environnement linux sans avoir peur de "casser" sa machine
- Très utile pour la formation
- Ressource souvent insoupçonnée qui a changé ma manière de faire de la bioinfo