

Computing resources: IFB cloud and PSMN

Description, Usefulness, Guidelines

Laurent Gilquin (IGFL) & Carine Rey (CIRI)

November 14th 2023



1. IFB cloud (Biosphere)
 1. 1. Introduction
 1. 2. In practice
 1. 3. The RainBio catalogue
 1. 4. Conclusion

2. PSMN
 2. 1. Introduction
 2. 2. Overview
 2. 3. Guidelines
 2. 4. Utilities

IFB cloud (Biosphere)

Introduction

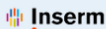
1.1 | What are the IFB and the IFB cloud?

The IFB: French Institute of Bioinformatics

“It is the National Bioinformatics Infrastructure that provides support, deploys services, organizes training and carries out innovative developments for the life sciences communities.”

The IFB cloud: This is one of the services provided by the IFB

“IFB is providing cloud services to analyze life science data. These services rely on a **federation of clouds** - **Biosphere** - built on interconnected IT infrastructures of some IFB’s platforms.”



Documentation: <https://www.france-bioinformatique.fr/en/ifb-cloud/>

1. Create an account and/or log in on **Biosphere** web page
<https://biosphere.france-bioinformatique.fr/>
2. Deploy a VM based on a template (an **appliance**) and a configuration (a **flavor**)
3. Connect to your VM according the appliance
 - with a web interface (HTTPS)
 - with the SSH command line interface through a terminal window
 - with a remote graphical desktop (X2Go)
4. Upload your data, scripts, ...
5. Analyse your data (1hour -> 1 day -> 1 week -> ...)
6. Download results
7. Shutdown the VM (!)

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

IFB cloud (Biosphere)

In practice

1.2 | To create your IFB cloud account:

- Go to <https://biosphere.france-bioinformatique.fr/>
- At the top right, click on **login**.
- Click again on **login**
- Click on **accept conditions**
- Select **ENS de Lyon|CNRS|Inserm|...** according your employer
- Identify yourself with your institutional **login and password**
- Fill in **your name, first name, city and postal code**, leaving all other information as default and accept.
- You'll get a new page. In the top right-hand corner, click **on the little man icon**, then on group.
- Click on **Join a group** in the tabs at top left.
- Search for the group corresponding to your laboratory *LBMC/IGFL/CIRI/RDP/...* and click on the **+** button to apply.
- And you're all set! Now , your application need to be validated by the IFB

1.2 | The dashboard:

- Once, your account is approved you can access to your dashboard:

The screenshot displays the 'Cloud' dashboard for a user named 'Biosphere'. The main section, 'Déploiements', lists several deployments with columns for ID, Name, Date, Specifications, Cloud provider, and Access. Below this is a section for 'Applications et déploiements favoris' (Favorite applications and deployments), which lists three specific applications with their IDs, names, dates, and configurations.

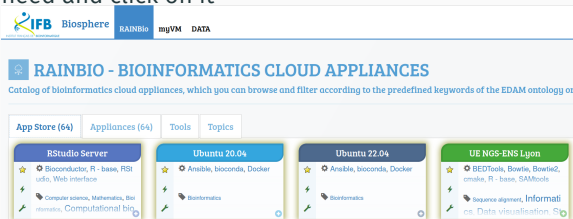
ID	Nom	Début	Spécification	Cloud	Accès
9028	Server	7 Dec 04 2016, 17h22	1 4 70	ib-core-cloud	https Params
9022	Ubuntu 16.04 Desktop (16.04)	7 Dec 04 2016, 16h36	2 8 120	ib-core-cloud	ssh Params
9020	Galaxy (19.01)	7 Dec 04 2016, 16h34	1 4 70	ib-core-cloud	https ssh Params
9018	Ubuntu Server	7 Dec 04 2016, 15h57	2 8 120	ib-core-cloud	https Params
9209	UC-MSD-4760 Lysol (Nov 2017)	7 Oct 08 2016, 08h48	16 64 920	ib-core-cloud	https Params

ID	Nom	Des. date	Paramétrage
16	Ubuntu Server	7 Dec 04 2016, 15h57	2 8 120
18	UC-MSD-4760 Lysol (Nov 2017)	7 Dec 03 2016, 14h42	16 64 920
120	Ubuntu 16.04 (16.04.0)	7 Oct 14 2016, 17h28	1 4 70

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

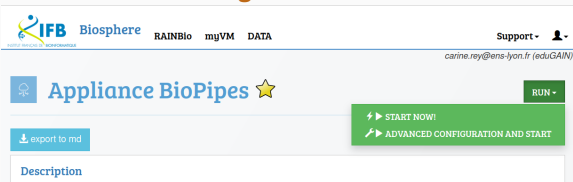
1.2 | How to deploy a VM ?

- 1) Click on **RainBio** in the head bar to select an appliance according your need and click on it



The screenshot shows the RainBio website interface. At the top, there is a navigation bar with the IFB Biosphere logo and links for RAINBio, myVM, and DATA. Below the navigation bar is the main heading "RAINBIO - BIOINFORMATICS CLOUD APPLIANCES" and a sub-heading "Catalog of bioinformatics cloud appliances, which you can browse and filter according to the predefined keywords of the EDAM ontology of". There are four tabs: "App Store (64)", "Appliances (64)", "Tools", and "Topics". Below the tabs, there are four appliance cards: "RStudio Server", "Ubuntu 20.04", "Ubuntu 22.04", and "UE NGS-ENS Lyon". Each card displays a star icon, a list of keywords, and a "Run" button in the top right corner.

- 2) Check the description and click on **Run** in the top right corner
- 3) Chose **Advanced configuration and Start**



The screenshot shows the configuration page for the "Appliance BioPipes" appliance. At the top, there is a navigation bar with the IFB Biosphere logo and links for RAINBio, myVM, and DATA. On the right side, there is a "Support" link and a user profile icon. Below the navigation bar, there is a heading "Appliance BioPipes" with a star icon. There is a "RUN" button in the top right corner. Below the heading, there is a "export to md" button. In the center, there is a green button with two options: "START NOW!" and "ADVANCED CONFIGURATION AND START". Below the button, there is a "Description" section.

- 4) Configure your VM by specifying: a **name**; a **group** from which take quotas; a **cloud**; a **flavor**

Customize the deployment of an appliance ✕

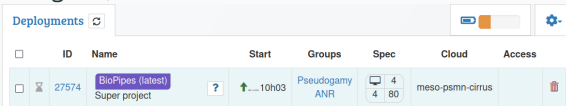
Deployment of appliance "BioPipes"

Name	<input type="text" value="Super project"/>
Group to use	<input type="text" value="CIRI (CIRI, Centre International de recherche en Infectiologie) : ▾"/>
Cloud	<input style="border: none; background-color: #f0f0f0; padding: 2px 5px; width: 100%;" type="text" value="meso-psmn-cirrus"/>
Cloud flavor	<input type="text" value="ifb.tr.xlarge (8 vCPU, 8GB GB RAM, 80GB GB local disk) : ▾"/>

- 5) Click on **Start Now!**, you will be redirected to your dashboard.

1.2 | How to deploy a VM ?

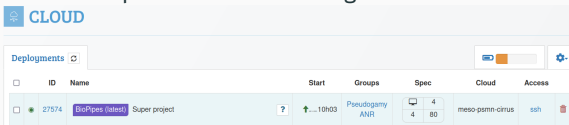
- 6) You can see on your dashboard your new VM in deployment (grey hourglass)



The screenshot shows the 'Deployments' page in Google Cloud Platform. A table lists a deployment with ID 27574, named 'BioPipes (latest) Super project'. The 'Start' column shows a grey hourglass icon and '10h03'. The 'Groups' column lists 'Pseudogamy ANR'. The 'Spec' column shows '4' instances and '80' GB of memory. The 'Cloud' column lists 'meso-psmn-cirrus'. The 'Access' column has a red trash icon.

ID	Name	Start	Groups	Spec	Cloud	Access
27574	BioPipes (latest) Super project	10h03	Pseudogamy ANR	4 4 80	meso-psmn-cirrus	

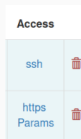
- 7) Wait for the complete deployment (15-20min; green light) and get the connexion parameters on the right



The screenshot shows the 'Deployments' page with a 'CLOUD' tab selected. The deployment with ID 27574 is now in a green state. The 'Access' column now shows 'ssh' and a red trash icon.

ID	Name	Start	Groups	Spec	Cloud	Access
27574	BioPipes (latest) Super project	10h03	Pseudogamy ANR	4 4 80	meso-psmn-cirrus	ssh

- 8) Connexion parameters will differ according your appliance



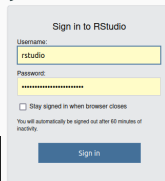
The 'Access' column is expanded to show two options: 'ssh' and 'https Params', each with a red trash icon to its right.

Access
ssh
https Params

1.2 | How to deploy a VM ?

- 9) Sometime the deployment may go wrong (red light), try again.
- 10) Using the connexion parameters, you can log in on your VM

```
(base) admincarine@bibs-3571-cr:~$ ssh ubuntu@192.33.153.182
Welcome to Ubuntu 22.04.3 LTS (GNU/Linux 5.15.0-69-generic x86_64)
```



Sign in to RStudio

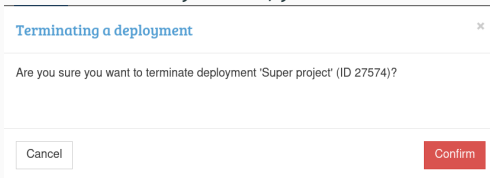
Username:

Password:

Stay signed in when browser closes

You will automatically be signed out after 60 minutes of inactivity.

- 11) When you have finished with your VM, you will click on **the trash**.



Terminating a deployment ✕

Are you sure you want to terminate deployment 'Super project' (ID 27574)?

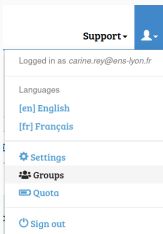
! Important

This action is irreversible, all the data on the VM will be destroyed.

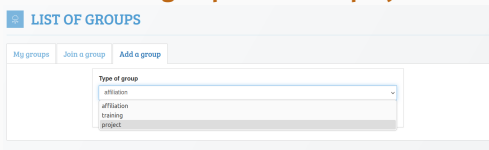
- They are 2 types of quotas:
 - your personal quota (small: by default 30.000 vCPU.h)
 - quotas for the groups you belong to (may be big)
- Quotas are counted in **vCPU.h**, which means that a VM with 8 CPUs will cost 8 vCPU.h each hour.
- If you use group quotas, the quota total is shared between all the people who belong to the group.

1.2 | Create a group

- You can get an extension of quotas by creating your group
- Click on the **little man icon** on the top right corner and **Group**



- Click on **Add a group** and select **project**



- Fill the **form** and wait for its acceptance (15 days)



Tip

Using a group is the only way to have persistent and shared storage mounted on each VM (of the group).
(For the moment only for meso-psmn-cirrus)

- You can view your persistent storage in the **Data** tab in the head bar
- Warning: A volume is attached to a cloud

The screenshot shows the IFB Biosphere interface. At the top, there are navigation tabs for 'RAINBio', 'myVM', and 'DATA'. The 'DATA' tab is selected. Below the navigation, there is a 'Support' link and a user profile icon. The main content area is titled 'DATA' and contains a 'Volumes' section. The 'Volumes' section has a 'Public databases' sub-tab. Below this, there is a search bar and a table of volumes. The table has columns for 'Volume name', 'Cloud', 'Start/Stop', 'TB.d', 'Size (Gb)', 'Max size', 'User', and 'Group'. One volume is listed: 'pseudogamy_anr' on the 'meso-psmn-cirrus' cloud, with a size of 162.97 TB.d and 1000 Gb. The 'User' is 'Myself' and the 'Group' is 'Pseudogamy ANR'. There are also 'Previous', 'Next', and 'Add volume' buttons.

Volume name	Cloud	Start/Stop	TB.d	Size (Gb)	Max size	User	Group
pseudogamy_anr	meso-psmn-cirrus	↑ May 31 2023, 10h56	162.97	1000	1000	Myself	Pseudogamy ANR

1.2 | The different clouds

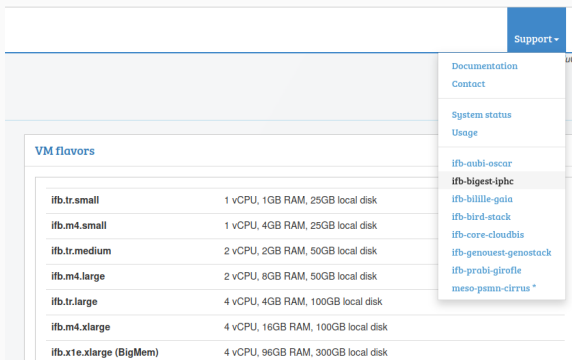
Platform	Location	Calcul (#CPU HT*)	Storage (#TB)	RAM (#GB)
IFB Core	Lyon (CC-IN2P3)	3 616	180	20 116
AuBI	Clermont-Ferrand (Université)	384	12	1 536
GenOuest	Rennes (IRISA)	600	350	2 600
PRABI	Lyon (Université)	536	144	3 300
BiRD	Nantes (Université)	860	150	2 500
BIGEst	Strasbourg (IPHC)	1024	500	4 000
BILILLE	Lille (Université)	192	0	768
CBP-PSMN	Lyon (ENS-Lyon)	3 408	24	7 552
Total federation of clouds		10 620	1 360	43 372

* CPU HT = Hyperthreading CPU

- You can use any of them (If one of them is undergoing maintenance, you can use another)
- Each of them will provide different VM sizes called **flavors**

1.2 | Each cloud provide different flavors

- You can view the different **flavors** available by clicking on "Support" and choosing a cloud.
- If you require a specific configuration (flavor), do not hesitate to contact the support.



The screenshot shows a web interface for VM flavors. A 'Support' dropdown menu is open, listing various support options. The 'ifb-bigest-iphc' flavor is highlighted in the table below.

VM flavors	
ifb.tr.small	1 vCPU, 1GB RAM, 25GB local disk
ifb.m4.small	1 vCPU, 4GB RAM, 25GB local disk
ifb.tr.medium	2 vCPU, 2GB RAM, 50GB local disk
ifb.m4.large	2 vCPU, 8GB RAM, 50GB local disk
ifb.tr.large	4 vCPU, 4GB RAM, 100GB local disk
ifb.m4.xlarge	4 vCPU, 16GB RAM, 100GB local disk
ifb.x1e.xlarge (BigMem)	4 vCPU, 96GB RAM, 300GB local disk

Support -

- Documentation
- Contact
- System status
- Usage
- ifb-aubi-oscar
- ifb-bigest-iphc**
- ifb-billie-gala
- ifb-bird-stack
- ifb-core-cloudbis
- ifb-genouest-genostack
- ifb-prabi-girofle
- meso-psmn-cirrus *

- You can use each appliance in all VM flavors / cloud
- If you require a specific appliance, you can built it, do not hesitate to contact the support. (Require advanced skills)

The screenshot shows the RainBio cloud appliances catalogue. At the top left is the IFB Biosphere logo. Navigation tabs include RAINBio, myVM, and DATA. The main heading is "RAINBIO - BIOINFORMATICS CLOUD APPLIANCES" with a subtext: "Catalog of bioinformatics cloud appliances, which you can browse and filter according to the predefined keywords of the EDAM ontology or". Below this are filter tabs: "App Store (64)", "Appliances (64)", "Tools", and "Topics". Four appliance cards are displayed:

- RStudio Server** (blue header): Includes Bioconductor, R - base, RStudio, Web interface. Keywords: Computer science, Mathematics, Bioinformatics, Computational bio.
- Ubuntu 20.04** (light blue header): Includes Ansible, bioconda, Docker. Keyword: Bioinformatics.
- Ubuntu 22.04** (dark blue header): Includes Ansible, bioconda, Docker. Keyword: Bioinformatics.
- UE NGS-ENS Lyon** (green header): Includes BEDTools, Bowtie, Bowtie2, cmake, R - base, SAMtools. Keywords: Sequence alignment, Informatics, Data visualisation, S.

IFB cloud (Biosphere)

The RainBio catalogue

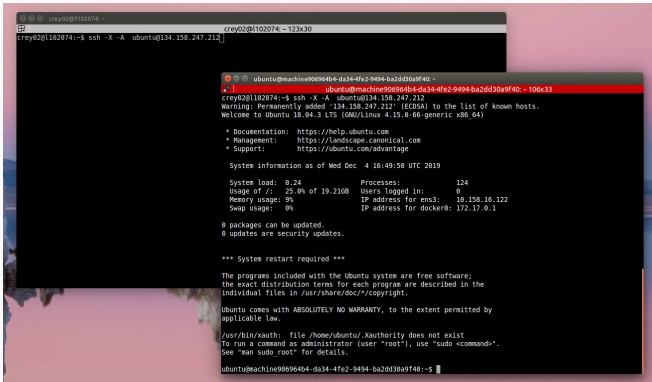
1.3 | The different appliances available : the RAINBio catalogue

The screenshot displays the RAINBio catalogue interface. At the top, there is a navigation bar with the 'ifb Biosphere' logo, 'RAINBio' branding, and 'ajgym DATA' text. On the right, there are links for 'Support', 'IR', and 'Se connecter'. Below the navigation is the main title 'RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD' and a subtitle 'Catalogue des appliances bioinformatiques dans le cloud, filtrez-les en utilisant les termes présents dans l'ontologie EDAM, ou en langage naturel.' Below this is a search bar and a filter menu with tabs for 'App Store (34)', 'Appliances', 'Outils', and 'Topics'. The main content area is a grid of 48 appliance cards, each with a colored header, a play button icon, a list of supported tools, and associated keywords. The appliances are arranged in 8 rows and 6 columns. The bottom right corner of the grid contains the text 'Le code couleur reste le même pour une même appliance.'

Appliance	Tools	Keywords
Bioinformatics (Insght)	BLAST+, HMMER, Inqth, SGE - G nEngine, Ubuntu, Web interface	Microbe tools and structural variants, Sequence comparison, Sequence visualization
Blotimage Fiji/ky	Ansible, bioconda, Bureau virtuel, Docker, Inqth3 Fiji, X2Go, XFCE	Bioinformatics, Informatics, Data visualization, Bioinformatics
BioPipes	bioconda, cellob, Docker, Docker Compose, Nextflow, Scalapack	Informatics, Bioinformatics, Workflow
Metar	bioconda, bowtie2, FastQC, Snake make	Bioinformatics, Sequence alignment, Workflow, Sequence analysis
CentOS	Ansible, bioconda, Docker	Informatics, Bioinformatics
Cytoscape	Bureau virtuel, Cytoscape, X2Go, XFCE	Bioinformatics, Data visualization, Molecular interaction, workflow
Debian 9	Ansible, bioconda, Docker	Bioinformatics, Informatics
EBAME19 Astra	Anvi's	Genomics, omics, metagenomics, P
EBAME19-MaxMicrobiome	Minimap2, R, RStudio, R - tidyverse	Data architecture, analysis and design, Sequence analysis, Bioinformatics
EBAME19-metaJC	bioPython, Mapioff, Numpy, pandas, system, Scipy	Pygenetics, Population genetics, Microbiome analysis, Sequence analysis
EBAME19-Metaflow		
formation LBEE NGS 2019	Alysis, BEDTools, Bioconductor, B LSCIO, FastQC, HBBT3, IGV - Integrat	Bioinformatics, Computation
Galaxy	bioconda, Docker, Galaxy portal	Informatics, bioinformatics, Compa
Genepipe School - Genome Variations	bedtools, BEDTools, BWA, Canvas, T asQC, GATK, IGV - Integrative Gene	avi, Informatics, Genetic variat
Genepipe School - Stats	DESeq2, RStudio, R - tidyverse, We b interface	transomics, Data architecture, visualization, Workflow, Proteomics
Integrat_finder	HMMER, Internal, prodigal	Microbe tools and structural variants, Sequence comparison, Sequence visualization
Jupyter	Julia, Jupyter, R	Computer science, Mathematics, Data visualization, workflow, workflow
Kiospice-training	Bioconductor, Kallisto, STAR	Bioinformatics, Computation
M2BINF UCBL	ncs3, Jupyter	Statistics, Statistics and prob ability, Proteomics, Bioinformatics
Maker	BLAST+, Emaxima, RepeatMasker	Genome assembler, Functional gen omics, Bioinformatics
NGSgenotag	bioPython, bowtie2, cap3, vnc, Fast QC, Mapioff, muscle, Numpy, ohmy	Pygenetics, Microbe tools and structural variants, Data visualization, workflow, workflow
NMRProcFlow	NMRProcFlow	Metabolomics
PathoTRACK	Centrifuge, Diamond, FastQC, pytho n, R, SortMeRNA, Web interface	Informatics, taxonomics, Bioinfo matics, Proteomics, Bioinformatics
PicoScope	Centrifuge, PicoScope, SPades	Genomics, Sequencing analysi s, Proteomics, Microbiology, omics
RStudio Server	Bioconductor, HDF5, R, RStudio, Web interface	Statistics, Mathematics, Comput ational biology, Computation, workflow
SGE cluster	SGE - GridEngine, Ubuntu	Informatics
Shiny	R, Shiny	Data architecture, analysis and design, Bioinformatics, Data visualis ation
SLURM cluster	SLURM	Informatics
Ubuntu 16.04	Ansible, bioconda, Docker, Ubuntu	Bioinformatics, Informatics
Ubuntu 16.04 Desktop	Ansible, bioconda, Bureau virtuel, Docker, Ubuntu, X2Go, XFCE	Bioinformatics, Informatics
UE NGS-ENS Lgen	BEDTools, BLAST+, Bowtie, bowtie 2, Clustal Omega, smake, DESeq2, Fa stQC	Sequence alignment, Informatics, tax onomics, omics, bioinformatics, omics
Weillvertier	pytho3, Web interface, Weillvertier	Computer science, Mathematics, Gene omics, Data visualization, workflow, workflow

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

- Filter by tools available by clicking on the "Tools" tab



The image shows two overlapping terminal windows. The top window is a terminal on a host named 'crey02@1102074'. The bottom window is a terminal on a host named 'ubuntu@machine906964b6-d34-4fe2-9494-ba2dd30a9f40'. The bottom terminal shows the output of an SSH command, including a warning about permanent IP addition, a welcome message for Ubuntu 18.04.3 LTS, system documentation links, system information as of Dec 4 16:49:58 UTC 2019, and a system restart requirement.

```
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 Wed Dec  4 16:49:58 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: 10.158.16.122
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.

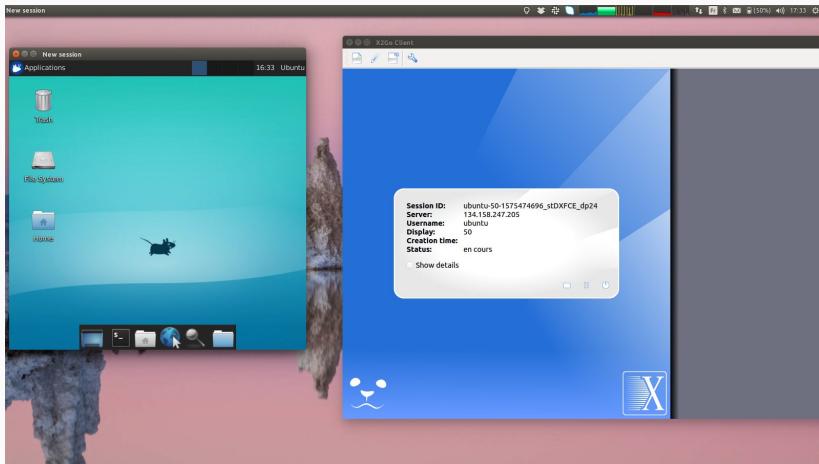
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

/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-ba2dd30a9f40:~$
```

- Need ssh keys configuration **before** deployment : https://ifb-elixirfr.github.io/biosphere/vm_connect
"See: Annex - Configure your SSH parameters"

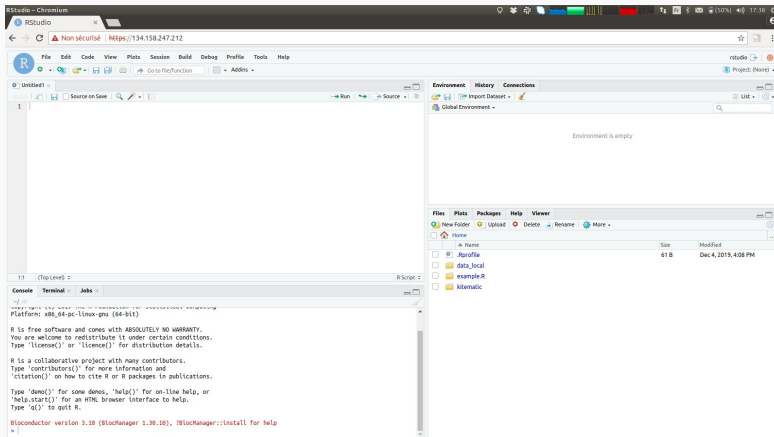
1.3 | Ubuntu Desktop - X2GO



- Need ssh keys and install X2GO (whatever your operating system, Windows, Mac, Linux): https://ifb-elixirfr.github.io/biosphere/vm_connect "See: Connecting with X2Go to a VM"

- Many tools already pre-installed (check the catalogue)
 - For example: BioPipes appliance is a Biosphere-commons app providing usual bioinformatics pipeline tools (nextflow, snakemak, cwltool)
- Easy installation of new tools
 - Conda
 - Docker

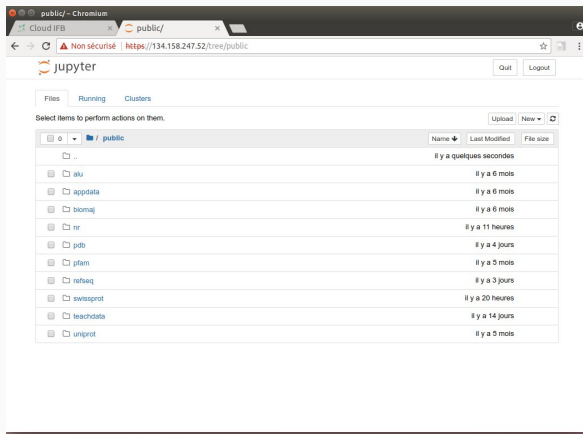
1.3 | Rstudio



Access via your browser (no ssh keys required) :

- you must accept the safety warning
- sometime, you must also modify your firewall rules

1.3 | Jupyter



Access via your browser (no ssh keys required) :

- you must accept the safety warning
- sometime, you must also modify your firewall rules

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

Useful links:

- Sign in
<https://ifb-elixirfr.github.io/biosphere/signin>
- Deploy your VM
https://ifb-elixirfr.github.io/biosphere/vm_deploy
- Connecting to your VM
https://ifb-elixirfr.github.io/biosphere/vm_connect
- Transferring data with your VM
<https://ifb-elixirfr.github.io/biosphere/data>

IFB cloud (Biosphere)

Conclusion

- Access to large-scale computing resources and/or specific ready to use environments
- Testing/learning the Linux environment without fear of breaking your machine
- Administrator rights on your machines
- No queue, VMs are directly available (For the moment)
- Direct access to public reference databases (RefSeq, nr, Uniprot, PDB, ...)
- Very useful for training
- No backup of your data, so be sure to keep copies locally
- An often unknown resource that has changed the way I do bioinfo

PSMN

Introduction

PSMN: Scientific Pole of Numerical Modelling

“The High Performance Computing (HPC) facility of ENS Lyon that hosts a large computational power and provide a fast and flexible access to it.”

Organization chart

PSMN	CBP
Ralf EVERAERS	
Loïs TAULELLE Cerasela Iliana CALUGARU Micaël CALVAS	Emmanuel QUEMENER Micaël CALVAS

New since 2023: PSMN + CBP → CBPsmn

- administrative change
- platform & tools identical (for now)

PSMN users

- ENS Lyon biology labs: LBMC, RDP, CIRI, IGFL
- other fields: Chemistry, Geology, Physics, Informatics, Astrophysics

Useful links

- [PSMN user documentation](#)
- [News feed](#): server issues, upgrades, maintenance, and so on
- [contact forms](#): signaling an issue, request a software update/installation, ...

Resources

- [LBMC guide for PSMN/CBP](#)
- [IGFL Practical introduction for beginners](#)

Common use cases

- Process large volume of data ($10^2\text{GB} \sim 10\text{TB}$)
examples: sequence (huge FASTQ) or image analysis (3D stack \times time)
- Run many tasks
example: performing the same task with \neq input file
- Long-running time task (several days)
- Exploit parallelism (single-node) or distributed computing (multi-nodes)
- Share files:
 - (i) databases and genomes
 - (ii) collaborative project requiring several TB

Prerequisites

- 1) Request an account through the form

Request access to PSMN resources

Ouverture de compte au PSMN

Surname *	<input type="text" value="SURNAME"/>
FirstName(s) *	<input type="text" value="First name"/>
Birthdate (YYYY-MM-DD) *	<input type="text"/>
Lab *	<input type="text" value="Choose a Lab"/>
Other Lab :	<input type="text"/>
Team (or Group, Project)	<input type="text"/>
Email *	<input type="text"/>
Administrative status *	<input type="text" value="Professor"/>
Other :	<input type="text"/>
Person in charge of the account opening *	<input type="text"/>
End date (YYYY-MM-DD)	<input type="text"/>

Scientific field description :

For non-ENS laboratories, justify your request:

Fields marked with a start (*) are mandatory

Prerequisites

1) Request an account through the form

2) An SSH client:



: native



: native



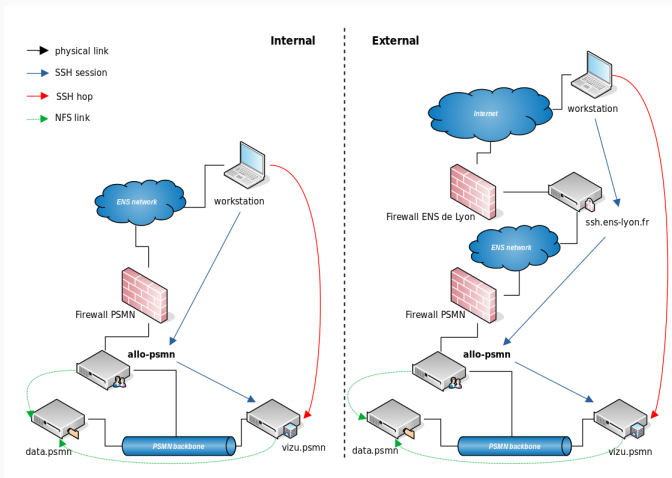
: OpenSSH (native since Win10 v.1803), Putty

3) Knowledge of [basic terminologies](#)

4) Basics of shell scripting ([CAN UNIX training](#) or [tutorial](#))

⇒ PSMN mandatory training (≈ 1h)

Connection servers



- internal gateway \Leftrightarrow the ENS network
- external gateway \Leftrightarrow outside the ENS network

Job submission

i Job

A **shell** script that contains a series of directives and instructions to be run

On a **shared** cluster, you submit a job to a **scheduler**, here **SLURM**, that will decide if and where the job can be run

sbatch myjob.sh

```
1  #!/bin/bash
2  #
3  #SBATCH --job-name=test ← SLURM directives
4
5  hostname -s ← instructions
6  sleep 60s
7
```


PSMN

Overview

The PSMN is composed of 4 **clusters**

i Cluster

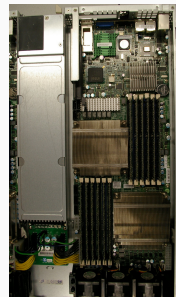
A set of computers linked to each other so that their **resources** can be **pooled** together



Each cluster contains several **compute nodes**

i Compute node

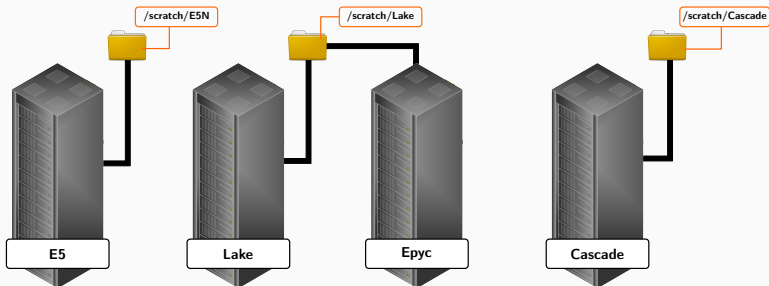
A unit of a cluster acting as a provider of **computing resources** (CPU, RAM) for **jobs** to run on



Each cluster possesses a **scratch** volume to store data **during** a job

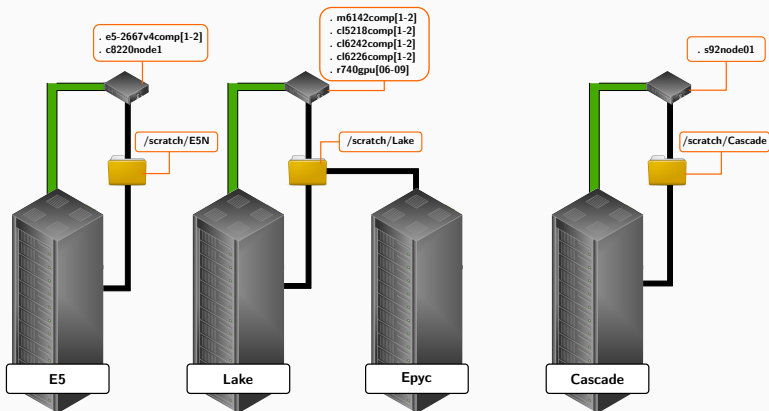
! Important

Scratch data are often erased to make space



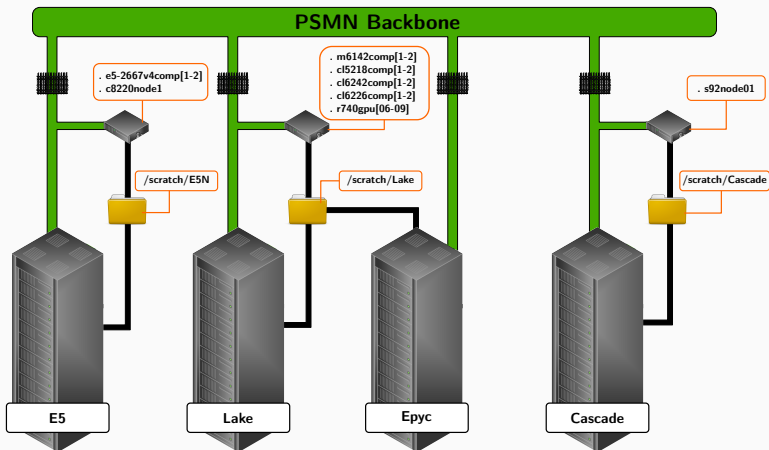
2.2 | Overview: Login nodes

Each cluster (except Epcy) contains one or multiple **login nodes**



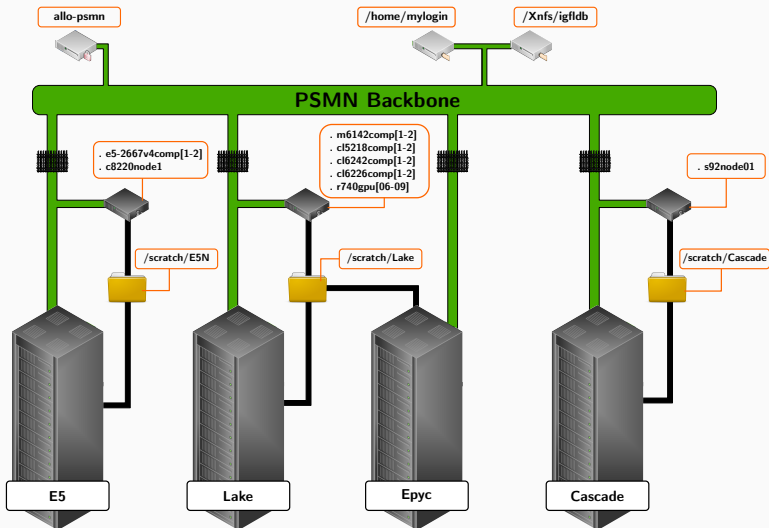
2.2 | Overview: Login nodes

login nodes are primarily used to access clusters via SSH



2.2 | Overview: Data spaces

Each user has access to two spaces: its **home** and the lab **Xnfs** volume



PSMN

Guidelines

Login nodes

Front-end servers on which you can:

- edit and transfer files,
- compile and build software,
- test scripts or software on small datasets,
- submit and monitor jobs.

! Important

Do not run **heavy** calculations directly on the login nodes

! allo-psmn

allo-psmn is a connexion server and should only be used to transfer files or access one of the login nodes

Compute nodes

- sole purpose is to run submitted jobs
- you can only connect to it if you have a job running on it



Tip

You can display information about the compute nodes of a cluster

PARTITION	CORES	CPUS	MEMORY(MB)	NODES	NODELIST
E5*	8	16	128872(MB)	4	c8220node[31-34]
E5*	8	16	128828(MB)	24	c6320node[1-24]
E5*	8	16	257852(MB)	24	c6320node[101-124]

“Cluster **E5** contain 4 nodes (c8220node31, c8220node32, c8220node33, c8220node34) that each possesses 8 cores, 16 CPUs and 128GB of RAM.”

Job scheduler (SLURM)

Multiple purposes:

- match and allocate compute nodes with user requests
- provide a framework to execute and monitor requests
- manage and order the queue of pending requests

SLURM directives

Placed **at the head** of the job script to specify computing resources:

- resources allocation (CPU, RAM)
- estimated run time
- constrains (cluster, node, GPU, ...)

Two ways to submit a job:

(i) through a script: **sbatch** (ii) interactively: **srun**

Available spaces

- your home folder: **`/home/{username}`**
- your lab Xnfs folder: **`/Xnfs/{lab}db`**
- cluster scratch: **`/scratch/{cluster}/{username}`**
- folder for Nextflow job: **`/Xnfs/abc`** (maintained by LBMC)

i Project space

It is possible to get a temporary volume on **`/Xnfs`** for project requiring a decent amount of space ($\approx 10\text{TB}$)

Remark

Discuss within your lab on how to **properly** organize your lab space (avoid duplicata, keep it tidy, ...)

SSH-based tools

- **SCP:** command line tool limited to file transfer
- **SFTP:** CLI or GUI for file transfer, access and management
- **rsync:** command line tool to synchronize a set of directories
- **SSHFS:** mount your PSMN home on your workstation

Modules

PSMN provides pre-installed software through **Lua Environment Modules**.

3-steps process:

- initialize the environment
- choose a list of modules (one per cluster)
- load one or several modules

J	E5	Lake	Epyc	Cascade
JasPer/2.0.33-GCCcore-11.2.0	✓	✓		✓
Java/11.0.2	✓	✓		✓
jbigkit/2.1-GCCcore-11.2.0	✓	✓		✓
Jellyfish/2.3.0-GCC-11.2.0	✓	✓		✓
jemalloc/5.2.1-GCCcore-11.2.0	✓	✓		✓
Julia/1.7.1-linux-x86_64	✓	✓		✓

Local install

You can download and compile an external program/software on your PSMN home.

Requirements:

- the program must be available for Debian 11
- the libraries must be already installed on PSMN

Warning

You have no admin rights on PSMN \Rightarrow no **sudo** commands

Alternatively, you can use containers: Singularity, Charliecloud (not Docker)

PSMN

Utilities

Shellcheck

Command line tool to check for syntax errors and sometimes bugs:

```
shellcheck myjob.sh
```

Warning

- **shellcheck** cannot detect incorrect paths
⇒ check them at the login node or through an interactive job
- **shellcheck** cannot detect syntax errors in SLURM directives
⇒ see [Caution block](#)

A web version is available here: <https://www.shellcheck.net>

Text editor

Available on PSMN: Nano, Vim, Emacs

Limitations

- juggling with terminal and editor
- keep track of the file tree
- no language services: code completion, error and warning, and so on

⇒ use an IDE that support SSH connexion to remote server:
VSCodium, NeoVIM(+plugins), KDevelop, ...

2.4 | Utilities: SSH-based IDE

Quarto Website: home_size.sh — Kate

Fichier Édition Affichage Projets Client LSP Signets Sessions Outils Configuration Aide

Explorateur de systèmes de fichiers

Liste des symboles

sfpt: allo-psmn > home > Igitquin

Nom	Taille	Date
.cache		18/10/2023 11:41
.conda		22/02/2023 10:33
.config		16/10/2023 13:29
.gnupg		29/08/2022 10:47
.gradle		20/07/2023 12:04
.lmod.d		23/02/2023 09:24
.nano		06/03/2023 15:09
.oh-my-zsh		06/11/2023 17:15
.ssh		04/04/2023 14:56
.texlive2020		20/05/2023 10:42
.vim		21/02/2023 10:03
.bin		30/05/2023 16:41
.miniconda3		31/08/2023 14:29
.R		20/07/2023 11:48
scripts	6 éléments	10/11/2023 17:17
softwares		16/10/2023 13:47
stderr		19/10/2023 08:59
stdout		19/10/2023 08:59
.bash_aliases	1,0 Kio	23/02/2023 16:05
.bash_history	9,6 Kio	21/03/2023 14:13
.bashrc	1,7 Kio	23/02/2023 09:12
.condarc	26 o	22/02/2023 10:31
.gitconfig	29 o	05/06/2023 09:36
.profile	793 o	20/05/2023 18:00
.python_history	2,0 Kio	17/10/2023 18:04
.RData	48 o	20/05/2023 17:27
.Rhistory	21 o	20/05/2023 17:27
.viminfo	8,9 Kio	22/03/2023 17:18
.wget-hsts	300 o	31/08/2023 15:58
.Xauthority	295 o	10/05/2023 09:03
.zprofile	36 o	17/05/2023 09:19
.zsh_history	99,2 Kio	10/11/2023 17:17
.zshrc	5,4 Kio	18/07/2023 13:47
.zshrc-pre-oh-my-zsh	1,7 Kio	03/04/2023 09:22
.zshrc.save	3,8 Kio	03/04/2023 09:37
home_list.txt	2,3 Kio	10/11/2023 16:33
igfidb_report.txt	3,3 Kio	25/04/2023 15:02

per_read_mean.py polars_per_read_mean.py polars_eventalign_WT20B1_noNN_processed_perpos_mean.csv home_size.sh

```
1 #!/bin/zsh
2
3 #SBATCH --job-name=home_size .....#name-of-the-job
4 #SBATCH --partition=E5 .....#name-of-the-partition
5 #SBATCH --cpus-per-task=4 .....#number-of-CPU-per-task
6 #SBATCH --mem-per-cpu=250M .....#memory-required-per-CPU, -total: 2*2048=4G18
7 #SBATCH --ntasks=1 .....#number-of-tasks-of-the-job
8 #SBATCH --nodes=1 .....#number-of-nodes-requested
9 #SBATCH --time=0-10:00:00 .....#day-hours:minutes:seconds
10 #SBATCH --output=/home/igitquin/stdout/%j.%x.out .....#standard-output-file-format
11 #SBATCH --error=/home/igitquin/stderr/%j.%x.err .....#error-file-format
12 #-with:
13 #.%*.*.jobname
14 #.%j.*.jobid
15 |
16 while read -r line;
17 do
18   .du -sh "/home/${line%.*}" 2->>[grep -v "Permission non accordée"]>>"${HOME}/home_size.txt"&
19   ..#At-most-as-number-of-CPU-cores
20   ..[ ${jobs} | wc -l ] -ge ${nproc} ] -66 wait
21 done << "${HOME}/home_list.txt"
22
23 wait
24
```

text editor

```
~/IGFL/GitBio/Websites/Quarto/assets/metadata] $ ssh psmn-e5
No Slurm jobs found on node.
No Slurm jobs found on node.
Linux e5-2667v4comp2.psmn.ens-lyon.fr 5.10.0-19-amd64 #1 SMP Debian 5.10.149-2 (2022-10-21) x86_64

      #
     ##
    ###
   ####
  #####
 #####
#####

Debian Bullseye (11)

Whenever possible, thank you to acknowledge the use of PSMN resources:
see http://www.ens-lyon.fr/PSMN/doku.php?id=science:accueil

Last login: Fri Nov 10 16:39:28 2023 from 172.16.1.90
[~]$
```

terminal

Rechercher

Sortie Chercher et remplacer Projet en cours Panneau de terminal Client LSP

Syntax highlighting

Highlights source code in different colors and fonts according to the category of terms

⇒ improves readability and finding error

Example: zsh with Oh-My-Zsh plugin

```
[~]$ df -h --total /Xnfs/igflldb
Sys. de fichiers      Taille Utilisé Dispo Uti% Monté sur
r730data7:/data/igflb 40T    34T  6,1T  85% /Xnfs/igflb
total                 40T    34T  6,1T  85% -
[~]$ fdisk -l
```

- Two complementary tools to use in function of your computational skills and your needs:
 - IFB cloud to start with bioinformatic or for specific analyses
 - PSMN for advanced users and high computing analyses
- Both tools require users to manage their environment and data
- Do not hesitate to get help from your bioinformatics hub

They are mutualized tools

Take care to respect the other users by:

- managing your storage
- tidying shared spaces
- shutting down useless VM / removing intermediate files

“The Institut Français de Bioinformatique (IFB) is funded by the Programme d’Investissements d’Avenir (PIA), grant Agence Nationale de la Recherche, number ANR-11-INBS-0013.”



“We gratefully acknowledge support from the PSMN (Pôle Scientifique de Modélisation Numérique) of the ENS de Lyon for the computing resources.”

