



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

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/

1.1 | Quick overview

- Create an account and/or log in on Biosphere web page https://biosphere.france-bioinformatique.fr/
- 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

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

		이 아 ☆
		Support
		cance rey@ensityon.t
Detwi	Specification	Clear Acces
7 #Dec 04 2019, 17122	4 70	its-core-cloud Mitre Params
2 #Dec 04 2019, 16136	2 8 129	Ib-core-cloud ash Parama
9 #Dec 04 2010, 10104	4 70	to-core-cloud Maps seh Parama
P #Dec 04 2010, 10107	2 6 120	Ib-core-cloud Mips Paramo
2 \$CULOB 2019.08546	04 923	the core cloud the parame
		Tout wor
	Der. ders.	Paramétrage
	₱Dic 04 2019, 19837	2 8 120
1	+ Die 03 2019, 14642	44 923
	tou 14 2010, 17928	4 70
		-
🎂 Inserm elixi	-	\cap

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

Biosphere RAINBIO	myVM DATA	
RAINBIO - BIOI	FORMATICS CLOUD APPLIANCES	
Catalog of bioinformatics cloud app	ances, which you can browse and litter according to the predefined keyword	s of the EDAM ontology or
App Store (64) Appliances (64) RStudio Server	Tools Topics Ubuntu 20.04 Ubuntu 22.04	UE NGS-ENS Lyon
Bioconductor, R - base, RSt unio, Web interface.	😭 🗘 Ansible, bioconda, Docker 😭 🍄 Ansible, bioconda, Docker 😭 😋	BEDTools, Bowtie, Bowtie2, nake, R - base, SAMtools

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



1.2 | How to deploy a VM?

 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 a	appliance "BioPipes"			
Name	Super project			
Group to use	CIRI (CIRI, Centre International de recherche en Infectiologie) :	/		
Cloud	meso-psmn-cirrus			
Cloud flavor	ifb.tr.xlarge (8 vCPU, 8GB GB RAM, 80GB GB local disk)	/		
Cancel	Start now!			

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 deployement (grey hourglass)

Dej	ploy	ments	C							¢٠
		ID	Name		Start	Groups	Spec	Cloud	Access	
	×	27574	BioPipes (latest) Super project	?	† 10h03	Pseudogamy ANR	4 4 80	meso-psmn-cirrus		ŵ

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

~ (U							
Deplo	yments [0					-		٥.
	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

Access	
ssh	ŵ
https Params	ŵ

1.2 | How to deploy a VM?

Importar

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



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)?		
	Cancel	Confirm	
t			

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

- They are 2 types of quotas:
 - your personnal 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 extention 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

LIST O	F GROUPS	
My groups Join	a group Add a group	
	Type of group	
	atiliation	~
	affiliation training	
	project	

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

1.2 | Persistent storage



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

CEL ROMONICA	RAINBio myVM	DATA						carine.rey@en	upport + s-lyon.fr (ed
DATA									
plumes Public datab	ases								
Show 10 v entries							Search:		
	Cloud	.l† Si	tart/Stop ⊔†	TB.d ⊥†	Size (Gb) 👘	Max size 🔐	User ⊥†	Group 11	41
pseudogamy_anr	meso-psmn-cirru	is 🕇	May 31 2023, 10h56	162.97	1000	1000	Myself	Pseudogamy ANR	1
showing 1 to 1 of 1 entries								Previous	Next
								+ Ac	ld volume

1. 2 | The different clouds

Platform	Location	Calcul (#CPU HT*)	Storage (#TB)	RAM (#GB)
IFB Core	Lyon (CC-IN2P3)	3 616	180	20 116
AuBl	Clermont-Ferrand (Université)	384	12	1 5 3 6
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 federa	tion 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.

		Documentation	
		Contact	
		System status	
M flavors		Usage	
		ifb-bigest-iphc	
ifb.tr.small	1 vCPU, 1GB RAM, 25GB local disk	ifb-bilille-gaia	
ifb.m4.small	1 vCPU, 4GB RAM, 25GB local disk	ifb-bird-stack ifb-core-cloudbis	
ifb.tr.medium	2 vCPU, 2GB RAM, 50GB local disk	ifb-genouest-genostacl	
ifb.m4.large	2 vCPU, 8GB RAM, 50GB local disk	ifb-prabi-girofle	
ifb.tr.large	4 vCPU, 4GB RAM, 100GB local disk	meao-painte-curus	
	A CRU ACCE RAM ADDCR Intel Set		

1. 2 | The RainBio catalogue

- 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)



IFB cloud (Biosphere)

The RainBio catalogue

1. 3 | The different appliances available : the RAINBio catalogue



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

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



 Need ssh keys configuration before deployement: 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	СВР
Ralf EVE	ERAERS
Loïs TAULELLE Cerasela Iliana CALUGARU Micaël CALVAS	Emmanuel QUEMENER Micaël CALVAS

New since 2023: PSMN + CBP \rightarrow CBPsmn

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

2.1 | Introduction

PSMN users

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

Useful links

- \rightarrow PSMN user documentation
- \rightarrow News feed: server issues, upgrades, maintenance, and so on
- → contact forms: signaling an issue, request a software update/installation, ...

Resources

- \rightarrow LBMC guide for PSMN/CBP
- \rightarrow IGFL Practical introduction for beginners

Common use cases

- Process large volume of data (10²GB \sim 10TB) examples: sequence (huge FASTQ) or image analysis (3D stack \times time)
- Run many tasks example: performing the same task with ≠ 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

Surname *	SURNAME
FirstName(s) *	First name
Birthdate (YYYY-MM-DD) *	
Lab	* Choose a Lab 🗸
Other Lab :	
Team (or Group, Project)	
Email *	
Administrative status	Professor v
Other :	
Person in charge of the account opening st	
End date (YYYY-MM-DD)	
Scientific field description :	
	-
For non-ENS laboratories, justify your reque	st:

Request access to PSMN resources

Prerequisites

1) Request an account through the form

2) An SSH client:



- os : native
- 3) Knowledge of basic terminologies
- 4) Basics of shell scripting (CAN UNIX training or tutorial)
- \Rightarrow PSMN mandatory training (pprox 1h)

Connection servers



- internal gataway ⇔ the ENS network
- external gataway \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

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

2.2 | Overview: Login nodes

Each cluster (except Epyc) contains one or multiple login nodes

login nodes are primarily used to access clusters via SSH

41 / 57

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

2.3 | Guidelines: nodes usage

Compute nodes

🔮 Tip

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

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 10TB)

Remark

Discuss within your lab on how to ${\it properly}$ 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	Ерус	Cascade
JasPer/2.0.33-GCCcore-11.2.0	1	1		1
Java/11.0.2	1	1		1
jbigkit/2.1-GCCcore-11.2.0	1	1		1
Jellyfish/2.3.0-GCC-11.2.0	1	1		1
jemalloc/5.2.1-GCCcore-11.2.0	1	1		1
Julia/1.7.1-linux-x86_64	1	1		1

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 langage services: code completion, error and warning, and so on

 \Rightarrow 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 —	٣
ichier <u>É</u> dition Afficha <u>g</u> e <u>P</u> r	ojets Client LS	P <u>S</u> ignets Session	is Outils Configuration Aide	_
🔘 🔇 🔰 Ċ 🕼 Signets, I	© Q			100 .
sftp: allo-psmn > home >	lgilquin		1 #1/01//2sn	<u> </u>
Nom	 Taille 	Date	3 #SBATCHjob-name= <u>home_size</u> #-name-of-the-job	
> cache		18/10/2023 11:41	4 #SBATCHpartition=E5	
> 💼 .conda		22/02/2023 10:33	5 #SBATCHcpus-per-task=4	
config		16/10/2023 13:29	0 #3DAILT+mem-per-c <u>pu=22007</u>	
> gnupg		29/08/2022 10:47	8 #SBATCHnodes-1	
.gradle		20/07/2023 12:04	9 #SBATCHtime=0-10:00:00	
lmod.d		23/02/2023 09:24	<pre>10 #SBATCHoutput=/home/lgilguin/stdout/%j.%x.out#.standard-output-file.format</pre>	
• 🔳 .nano		06/03/2023 15:09	11 #SBAILHerror=/nome/lgllduln/stderr/%j.%x.err#-error-file-format	
oh-my-zsh		06/11/2023 17:15	3 # • W1(1) 3 # • 9x - = - jobname	
ssh		04/04/2023 14:56	14 #-%j-=-jobid	
.cexiive2020		20/05/2023 10:42		
vim		21/02/2023 10:03	to while-reagr-(ine;	
> bin		30/05/2023 16:41	17 ▼ 00 18	
miniconda3		31/08/2023 14:29	19 ··#·At-most-as-number-of-CPU-cores	
P R		20/07/2023 11:48	20[-\$(-jobs+]-wcl-)ge-\$(-nproc+)-]-&&-wait	
scripts	6 elements	10/11/2023 17:17	<pre>done-<-`\${HOME}/<u>home_list.txt</u>"</pre>	
sortwares		16/10/2023 13:47	22 UNIT	
stderr		19/10/2023 08:59		
Stdout		19/10/2023 08:59	and the second	
	1,0 KI0	23/02/2023 16:05	text editor	
	9,6 Kio	21/03/2023 14:13		
B .Dasnrc	1,7 Ki0	23/02/2023 09:12		
.condarc	26.0	22/02/2023 10:31		÷
iii .gitconrig	29 0	05/06/2023 09:36		-
	793.0	20/05/2023 18:00	[-/IGFL/GitBio/Websites/Quarto/assets/metadata]\$ ssh psmn-e5	^
	2,0 KI0	77/10/2023 18:04	No Slurm jobs found on node.	
B Rodud	48.0	20/05/2025 17:27	No Slurm jobs found on node.	
	210	20/05/2023 17:27	LINUX 05-266/V4Comp2.psmn.ens-lyon.tr 5.10.0-19-amob4 #1 SMP Debian 5.10.149-2 (2022-10-21) X86_64	
iii .viminro	8,9 KI0	22/03/2023 17:18	# #	
Wgeensts	300 0	31/08/2023 13:38		
	295.0	10/05/2023 09:03		
iii .zpronite	30.0	17/05/2025 09:19		
.zsn_nistory	99,2 KIO	10/11/2023 17:17		
iii .zshire	5,4 KIO	18/07/2023 13:47	Debian Bullseye (11)	
.zsmc.pre-on-my-zsn	1,7 Ki0	03/04/2023 09:22		
D have list but	3,8 KIO	03/04/2023 09:37	Whenever possible, thank you to acknowledge the use of PSMN resources:	
in infidite second but	2,3 Kio	10/11/2023 16:33	see nttp://www.ens-tyon.tr/PSMk/doku.pnp?id=Science:accueit	
Igridb_report.txt	3,3 KIO	25/04/2023 15:02	Last Login: Fri Nov 10 16:39:28 2023 from 172.16.1.90 terminal	
file	s tree			
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

 \Rightarrow improves readability and finding error

Example: zsh with Oh-My-Zsh plugin

[~] \$ df -htotal <u>/Xn</u>	fs/igflo	<u>lb</u>			
Sys. de fichiers	Taille	Utilisé	Dispo	Uti%	Monté sur
r730data7:/data/igfldb	40T	34T	6,1T	85%	/Xnfs/igfldb
total	40T	34T	6,1T	85%	
[~] \$ fdisk -1					

Conclusion

- Two complementary tools to use in function of your computational skills and your needs:
 - IFB cloud to start with bioinformatic or for specific analyses
 - $\circ~$ 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."

