



Let's discover Jupyter through the IFB cloud

Working environment

What is jupyter book ?

- One of the most popular tool among data scientists to perform data analysis
- Provides a complete environment in which numerous programming languages can be use through a simple web browser

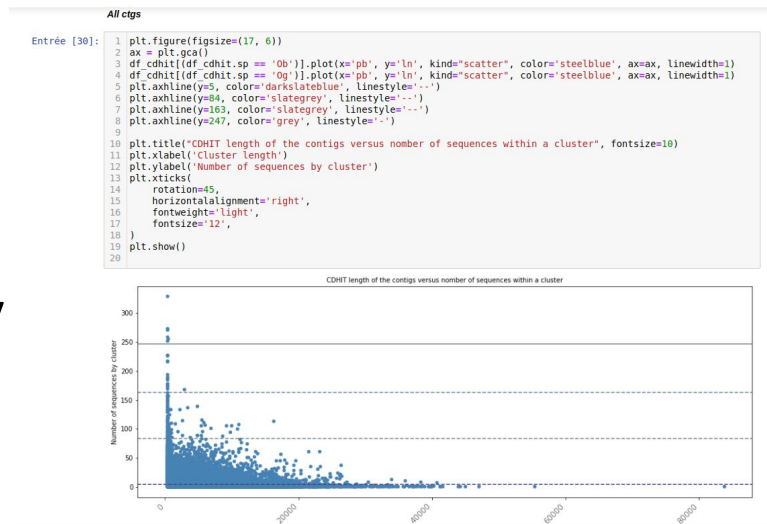
ex : Bash (Linux), Python, Java, R,
Julia, Matlab, Octave, Scheme,
Processing, Scala



Why use jupyter book ?

An unique interface/file where text,code and output codes can be mixed :

- code can be executed inside each cell of the notebook
- code output is directly displayed in the notebook



An unique interface/file where text,code and output codes can be mixed :

- code can be executed inside each cell of the notebook
- code output is directly displayed in the notebook
- explanations, formulas, charts can be added

jupyter parseClstr-Copy1 Dernière Sauvegarde : il y a 8 minutes (auto-sauvegarde) Se déconnecter

Fichier Edition Affichage Insérer Cellule Noyau Widgets Aide Non fiable Python 3

Entrée [1]:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import numpy as np
4
5 csv_cdhit_file = "/home/christine/Documents/These/Data/CDHIT/ALL_CTGS_MERGE/allCtgsIRIGIN_TOG5681.dedup8095.PANDAS.csv"
6 df_cdhit= pd.read_csv(csv_cdhit_file,names=['ctg','sp','ctg-list','sp-list'], header=0)
7 #print(df_cdhit)
```

Anchoring data analysis

1 - CDHIT data analysis before anchoring on genome

1.1 Removing redundancy with CDHIT

- CDHIT Input : 1,306,676 contigs assembled from no mapped reads
- Tests & results

	0.9	0.95
0.00	375.615	484.394
0.05	418.136	531.326
0.90	473.270	598.983
0.95	544.441	659.658

- clusters generated after cdhit analysis : 484,394

1.2 Converting cdhit file into a csv loaded as a dataframe with pandas

The script cdhitVsAnchoring.py create the csv file allCtgsIRIGIN_TOG5681.dedup8095.PANDAS.csv

Load csv file into a dataframe

Lab notebook for science data ?



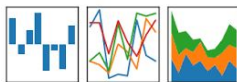
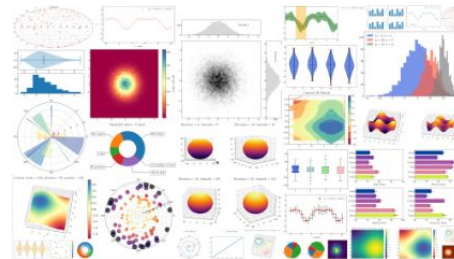
- One file to analyze data and generate reports
- Can be exported to many formats, including PDF and HTML, which makes it easy to share your project with anyone.
- Analysis are more transparent, repeatable and shareable

How to become a super datascientist ?

- easily import tabulated files into dataframes, similar to dataframes in R (and export)
- manipulate data tables / DataFrames
- easily plot graphs from these DataFrames using matplotlib

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



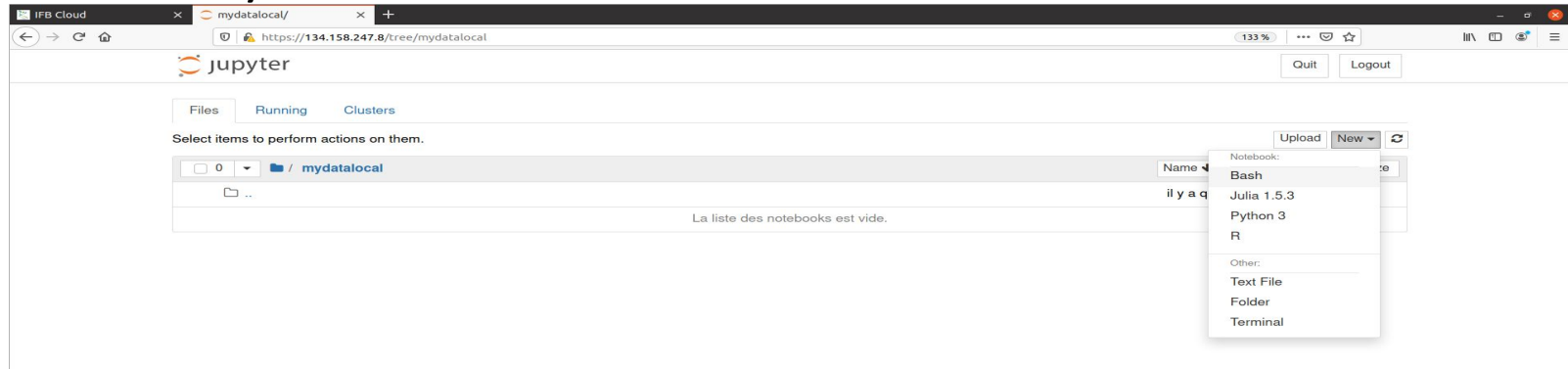
How will you use Jupyter Notebook ?

- Launch our analyses through a jupyter book within a virtual machine launched via the IFB cloud “BIOSPHERE”



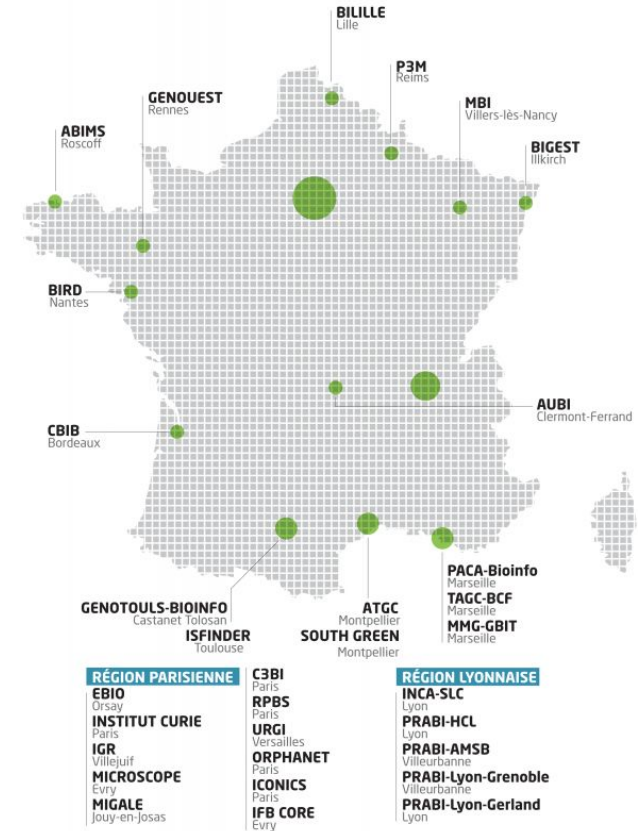
How will you use Jupyter Notebook ?

- Launch our analyses through a jupyter book within a virtual machine launched via the IFB cloud “BIOSPHERE”
- Through this virtual machine, we will create jupyter books and execute all our analysis





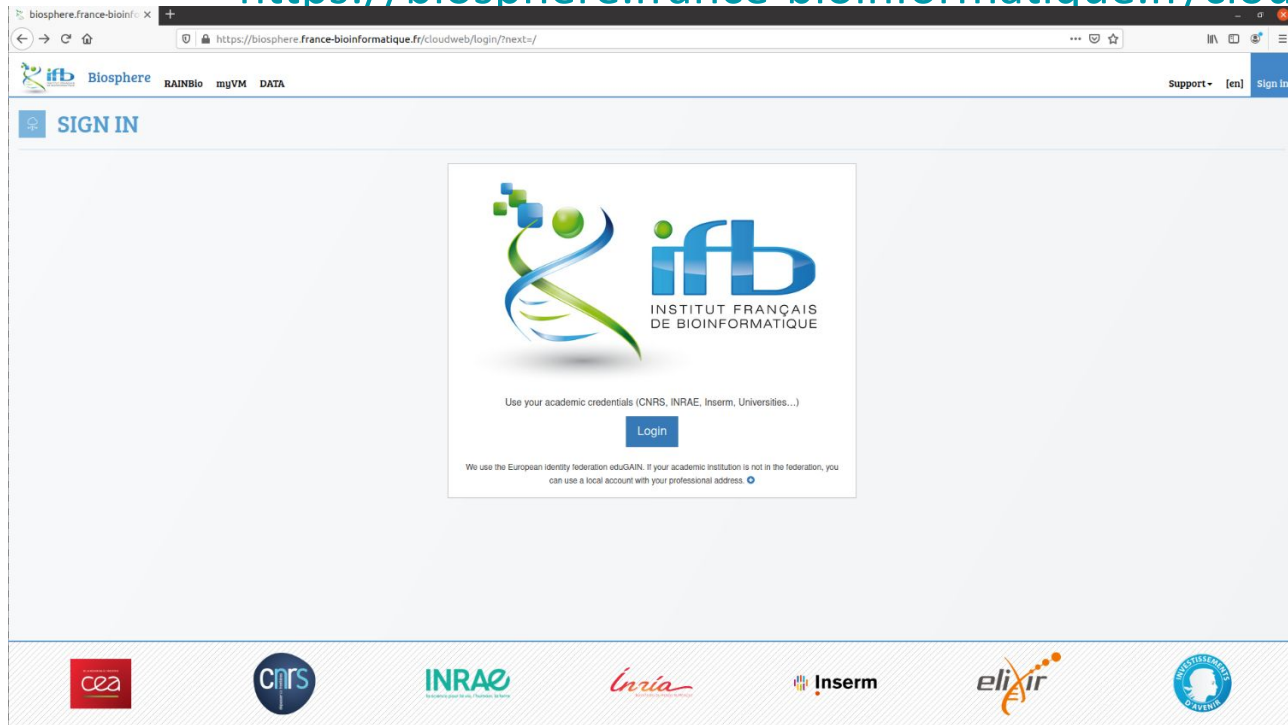
- A federation of clouds, which relies on interconnected IFB's infrastructures, providing distributed services to analyze life science data
- .Access to a large set of virtual machines (computing ressources, bioinformatics tool)



Let's start with biosphere

- Open the biosphere website :

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



Let's start with biosphere

- Select a specific group

The screenshot shows the IFB Biosphere myVM interface. The top navigation bar includes the IFB Biosphere logo, RAINBio, myVM, and DATA tabs. The main content area is titled 'CLOUD' and features a 'Déploiements' (Deployments) section with a refresh icon. Below this, there is a table with columns for ID, Nom, Début, Groupes, and Spécification. A red arrow points to the 'Groupes' link in the user menu, which is located in the top right corner. The user menu also includes options for Languages (English, Français), Paramètres, Groups, Quota, and Se déconnecter.

IFB Biosphere RAINBio myVM DATA

Support

Connecté.e en tant que francois.sabot@ird.fr

Langues
[en] English
[fr] Français

Paramètres
Groupes
Quota
Se déconnecter

Déploiements

ID Nom Début Groupes Spécification

Arrêter les déploiements

Appliances et déploiements favoris Déploiements récemment terminés Quota

ID Broker Nom Der. dém. Paramétrage

Let's start with biosphere

- Ask for joining Variants SG-2023 (Variants (SouthGreen 2023))

 **IFB Biosphere** RAINBio myVM DATA

LISTE DES GROUPES

Mes groupes Rejoindre un groupe Créer un groupe

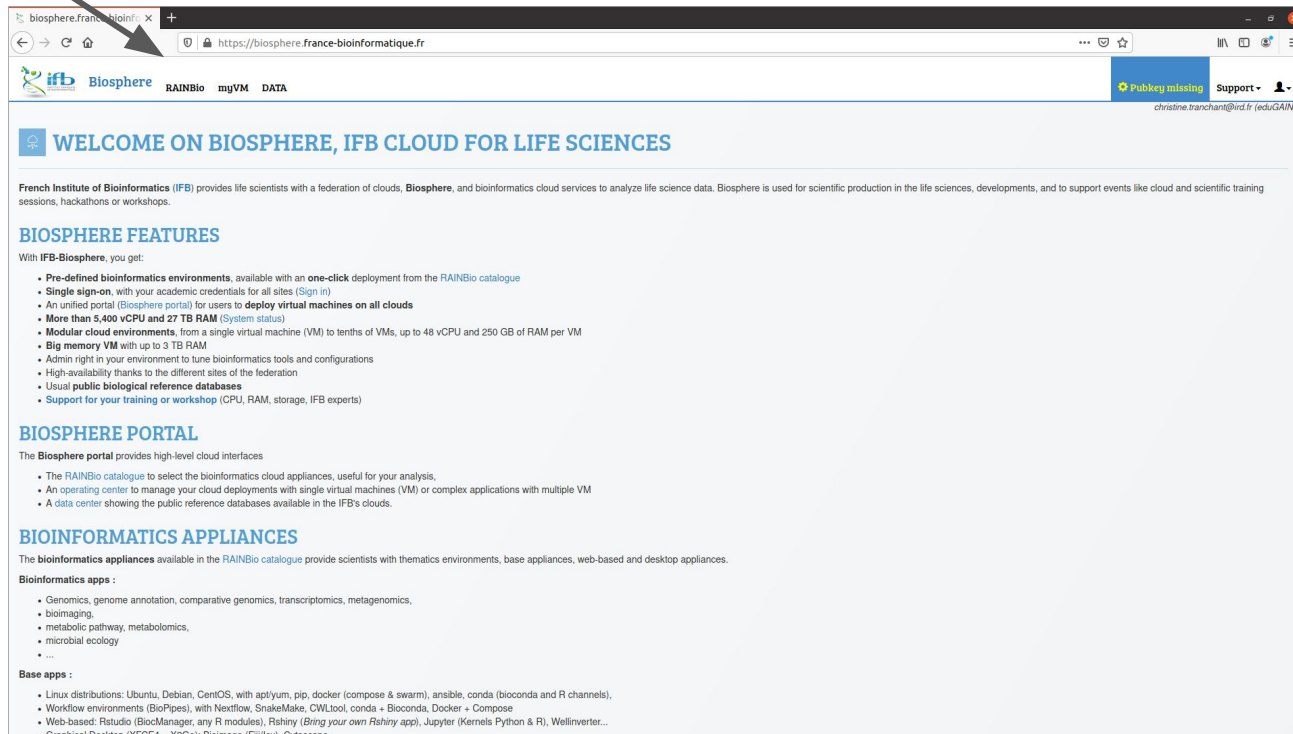
Afficher 10 éléments Recherche

	Nom	Site web	Actions
	2AD (Acquisition et Analyse de Données pour l'Histoire Naturelle), UAR2700		
	AGAP (Amélioration génétique et adaptation des plantes méditerranéennes et tropicales), 1098		
	AgroImpact (UR1158 AgroImpact Agroressources et Impacts environnementaux), UR1158		
	AMAP (botAnique et Modélisation de l'Architecture des Plantes et des végétations), UMR5120		
	ANF_Metabiodiv#2 (Exploration de la Diversité Taxonomique des Ecosystèmes par Metabarcoding)		
	ANSES - Sophia-Antipolis (ANSES - Laboratoire de Sophia-Antipolis), 13001202400043		
	ANSES AVB (ANSES Antibiorésistance et Virulence Bactériennes), -		
	ANSES Fougères (ANSES Laboratoire de Fougères), -		
	ANSES Ploufragan (ANSES Ploufragan-Plouzané), 13001202400118		
	ANSES-MYCO (ANSES - Mycoplasmoses animales), na		

Affichage de l'élément 1 à 10 sur 178 éléments

Précédent 1 2 3 4 5 ... 18 Suivant

RAINBIO catalog to access our Virtual Machine (VM)



The screenshot shows a web browser window with the URL <https://biosphere.france-bioinformatique.fr>. The page header includes the 'iftb Biosphere' logo and navigation links for 'RAINBio', 'myVM', and 'DATA'. A 'Pubkey missing' warning and a 'Support' link are also visible. The main content area features a welcome message and a list of 'BIOSPHERE FEATURES'.

WELCOME ON BIOSPHERE, IFB CLOUD FOR LIFE SCIENCES

French Institute of Bioinformatics (IFB) provides life scientists with a federation of clouds, **Biosphere**, and bioinformatics cloud services to analyze life science data. Biosphere is used for scientific production in the life sciences, developments, and to support events like cloud and scientific training sessions, hackathons or workshops.

BIOSPHERE FEATURES

With IFB-Biosphere, you get:

- Pre-defined bioinformatics environments, available with an one-click deployment from the [RAINBio catalogue](#)
- Single sign-on, with your academic credentials for all sites ([Sign in](#))
- An unified portal ([Biosphere portal](#)) for users to **deploy virtual machines on all clouds**
- More than **5,400 vCPU and 27 TB RAM** ([System status](#))
- Modular cloud environments, from a single virtual machine (VM) to tens of VMs, up to 48 vCPU and 250 GB of RAM per VM
- Big memory VM with up to 3 TB RAM
- Admin right in your environment to tune bioinformatics tools and configurations
- High-availability thanks to the different sites of the federation
- Usual public biological reference databases
- Support for your training or workshop (CPU, RAM, storage, IFB experts)

BIOSPHERE PORTAL

The **Biosphere portal** provides high-level cloud interfaces

- The [RAINBio catalogue](#) to select the bioinformatics cloud appliances, useful for your analysis.
- An [operating center](#) to manage your cloud deployments with single virtual machines (VM) or complex applications with multiple VM
- A [data center](#) showing the public reference databases available in the IFB's clouds.

BIOSPHERES APPLIANCES

The **bioinformatics appliances** available in the [RAINBio catalogue](#) provide scientists with thematic environments, base appliances, web-based and desktop appliances.

Bioinformatics apps :

- Genomics, genome annotation, comparative genomics, transcriptomics, metagenomics,
- bioimaging,
- metabolic pathway, metabolomics,
- microbial ecology
- ...

Base apps :

- Linux distributions: Ubuntu, Debian, CentOS, with apt/yum, pip, docker (compose & swarm), ansible, conda (bioconda and R channels),
- Workflow environments (BioPipes), with Nextflow, SnakeMake, CWLtool, conda + Bioconda, Docker + Compose
- Web-based: Rstudio (BioManager, any R modules), Rshiny (*Bring your own Rshiny app*), Jupyter (Kernels Python & R), Wellinverter...
- Graphical Desktop (XFCE4 + Xfce4-Bioscience, iFBItool, Cytoscape)

vm's name : **analysesSV**

IFB Biosphere RAINBio myVM DATA

Clé publique (PubKey) absente Support christine.tranchant@ird.fr (eduGAIN)

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 (4) Appliances Outils Topics Appliance éditables Ajouter ...

AnalysesSV

- bcftools, BEDTools, BWA, Jupyter, Matplotlib, pandas,
- DNA polymorphism, Genetic variation.

CoursAnalysesNanoporeSG

- bandage, Jupyter
- Data architecture, analysis and design, Mathematics, Statistics and probability

NGSanalysisJupyter

- BEDTools, BWA, Jupyter, SAMtools
- Data architecture, analysis and design, Mathematics, Statistics and probability

REPET

- Repet
- Bioinformatics

Le code couleur reste le même pour une même appliance.

ceza cnrs INRAE Inserm INVESTISSEMENTS D'AVENIR

Appliance AnalysesSV ☆ DEV

Exporter en md

Description

This IFB cloud appliance provides both the Jupyter Notebook and Lab environment (see [explanations](#)) to work on the structural variants detections on short and long reads.

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. In addition, we integrated various tools to perform the SV detection

Tools

- Bash kernel for jupyter
- Pandas
- Matplotlib
- Jupyter notebook/lab
- seqtk
- Minimap2
- BWA-MEM2
- Samtools/BCTools
- BEDTools
- VCFTools
- GATK
- Syri
- BreakDancer
- Sniffles
- Mummer

Contact

- Support Cloud IFB

Developpers

- Francois Sabot SouthGreen Platform
- Julie Orjuela-Bouniol SouthGreen Platform

App data

- Version : 20.04
- OS : Ubuntu
- OS version : 20.04

Licence

Licensed under GPLv3

Site web <https://hub.docker.com/r/francoissabot/trainingontvm>

Outils

bcftools BEDTools BWA Jupyter Matplotlib pandas SAMtools

OS	Ubuntu 20.04
Recette de l'app (git)	https://github.com/SouthGreenPlatform/training_SV_VM
App de base	Jupyter

Caractéristiques

Nom long	Analyses des variants structuraux en short reads, long reads et assemblage
Version	1.0
Cré.e	25 mai 2022 16:53
Dernière mise à jour	8 juin 2022 16:46
Clouds exclus	☒

Crédits

Contact	Francois Sabot Southgreen
Développeurs	Francois Sabot Southgreen Julie Orjuela-Bouniol SouthGreen Platform

LANCER

⚡ LANCER
➡ DÉPLOIEMENT AVANCÉ

Appliance AnalysesSV

Exporter en md

Description

This IFB cloud appliance provides both the Jupyter Notebook and Lab environment (see [explanations](#)) to work on the structural variants detection on short and long reads.

This Jupyter app is based on the Jupyter Docker Stacks (see [details](#)). By default, this Biosphere users can choose any other existing stack with an Advanced deployment in Biosphere portal. In addition, we integrated various tools to perform the SV detection

Tools

- Bash kernel for jupyter
- Pandoc
- Matplotlib
- Jupyter notebook/lab
- seqtk
- Minimap2
- BWA-MEM2
- Samtools/BCFtools
- BETools
- VCFtools
- GATK
- Syri
- BreakDancer
- Sniffles
- Mummer

Contact

- Support Cloud IFB

Developpers

- Francois Sabot SouthGreen Platform
- Julie Orjuela-Bouniol SouthGreen Platform

App data

- Version : 20.04
- OS : Ubuntu
- OS version : 20.04

Licence

Licensed under GPLv3

Site web <https://hub.docker.com/r/francoissabot/trainingontvm>

Outils

LANCER

▶ LANCER
▶ DÉPLOIEMENT AVANCÉ

Configurer le déploiement d'une appliance

Déploiement de l'appliance "AnalysesSV"

Name

CTranchant

Groupe à utiliser

DIADÉ (Diversité, Adaptati

Cloud

ifb-core-cloudbis

Gabarit d'image cloud

ifb.m4.small (1 vCPU, 4Go GB RAM, 25Go GB local disk)

- ifb.m4.small (1 vCPU, 4Go GB RAM, 25Go GB local disk)
- ifb.m4.large (2 vCPU, 8Go GB RAM, 50Go GB local disk)
- ifb.m4.xlarge (4 vCPU, 16Go GB RAM, 100Go GB local disk)
- ifb.m4.2xlarge (8 vCPU, 32Go GB RAM, 200Go GB local disk)
- ifb.m4.4xlarge (16 vCPU, 64Go GB RAM, 400Go GB local disk)**
- ifb.x1e.4xlarge (BigMem) (16 vCPU, 384Go GB RAM, 600Go GB local disk)
- ifb.m4.6xlarge (24 vCPU, 96Go GB RAM, 600Go GB local disk)
- ifb.m4.8xlarge (32 vCPU, 128Go GB RAM, 800Go GB local disk)
- ifb.x1e.8xlarge (BigMem) (32 vCPU, 768Go GB RAM, 600Go GB local disk)
- ifb.m4.12xlarge (48 vCPU, 192Go GB RAM, 1.2To GB local disk)
- ifb.x1e.12xlarge (BigMem) (48 vCPU, 1.1To GB RAM, 50Go GB local disk)
- ifb.m4.14xlarge (56 vCPU, 240Go GB RAM, 1.4To GB local disk)
- ifb.x1e.16xlarge (BigMem) (62 vCPU, 1.5To GB RAM, 1.5To GB local disk)
- ifb.x1e.32xlarge (BigMem) (124 vCPU, 2.9To GB RAM, 2.9To GB local disk)

Annuler

Quelle gabarit d'image doit être utilisé sur ce cloud ?

Loading...


RAINBio
myVM
DATA

Clé publique (PubKey) absente
Support

christine.tranchant@ird.fr (eduGAIN)


CLOUD

Déploiements

	ID	Nom	Début	Groupes	Spécification	Broker	Cloud	Accès
☐	19435	AnalysesSV (1.0) DEV CTranchant	↑ Jul 15 2022, 16h54	DIADÉ	16 64 / 400	1e82	ifb-core-cloudbis	

Arrêter les déploiements
Tout voir (1)

Appliances et déploiements favoris
Déploiements récemment terminés
Quota

ID	Broker	Nom	Der. dém.	Paramétrage
----	--------	-----	-----------	-------------

ready !

CLOUD

Déploiements

ID	Nom	Début	Groupes	Spécification	Broker	Cloud	Accès
19435	AnalysesSV (1.0) DEV CTTranchant	? ↑ Jui 15 2022, 16h54	DIADE	64 16 400	1e82	ifb-core-cloudbis	https Params 134.158.248.237

Arrêter les déploiements

Tout voir (1)

get the url... link "https"

IFB Biosphere RAINBio myVM DATA

Clé publique (PubKey) absente Support

christine.tranchant@ird.fr (eduGAIN)

CLOUD

Déploiements

	ID	Nom	Début	Groupes	Spécification	Broker	Cloud	*Accès*
☐	19435	AnalysesSV (1.0) DEV CTranchant	? ↑ Jui 15 2022, 16h54	DIADÉ	16 64 400	1e82	ifb-core-cloudbis	https 134.158.248.237

Arrêter les déploiements

Tout voir (1)

Get the token identifiant... link “Params”

The screenshot shows the IFB Biosphere RAINBio myVM Cloud interface. A modal window titled "Paramètres" is open, displaying a table with two columns: "nom" and "valeur". The table contains one entry: "JUPYTER_TOKEN" with the value "7b708d2b24d1947f0a48baf37f0986c9ca4774d8157a8825".

The background interface includes a top navigation bar with "IFB Biosphere", "RAINBio", and "myVM". A "Déploiements" section shows a table of deployments. Below this, there are tabs for "Appliances et déploiements favoris", "Déploiements récemment terminés", and "Quota".

ID	Nom	Début	Groupes	Spécification	Broker	Cloud	Accès
19435	AnalysesSV (1.0) DEV CTranchant	Jui 15 2022, 16h54	DIADÉ	64 16 400	1e82	ifb-core-cloudbis	https Params 134.158.248.237

Open your vm (https link) to access to your own jupyter lab

IFB Cloud x Accueil x +

← → ↻ 🏠 🔒 <https://134.158.247.8/tree> 133% ⋮ ☆ 🗂 📄 🔍 ☰

jupyter Quit Logout

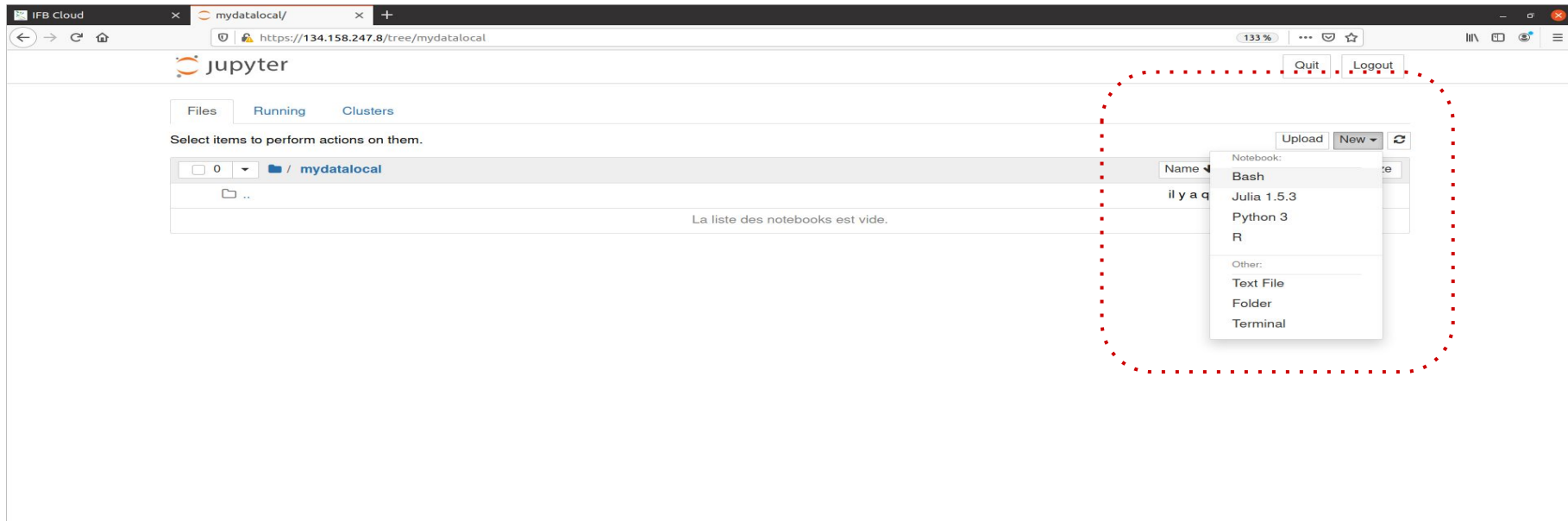
Files Running Clusters

Select items to perform actions on them. Upload New ↺

☐	0 ▾	📁 /	Name ▾	Last Modified	File size
☐		📁 mydatalocal		il y a un jour	
☐		📁 public		il y a 3 mois	

079760dd85a9bbc43e10354 ▾ 🔍 Tout surligner Respecter la casse Respecter les accents et diacritiques Mots entiers Expression non trouvée x

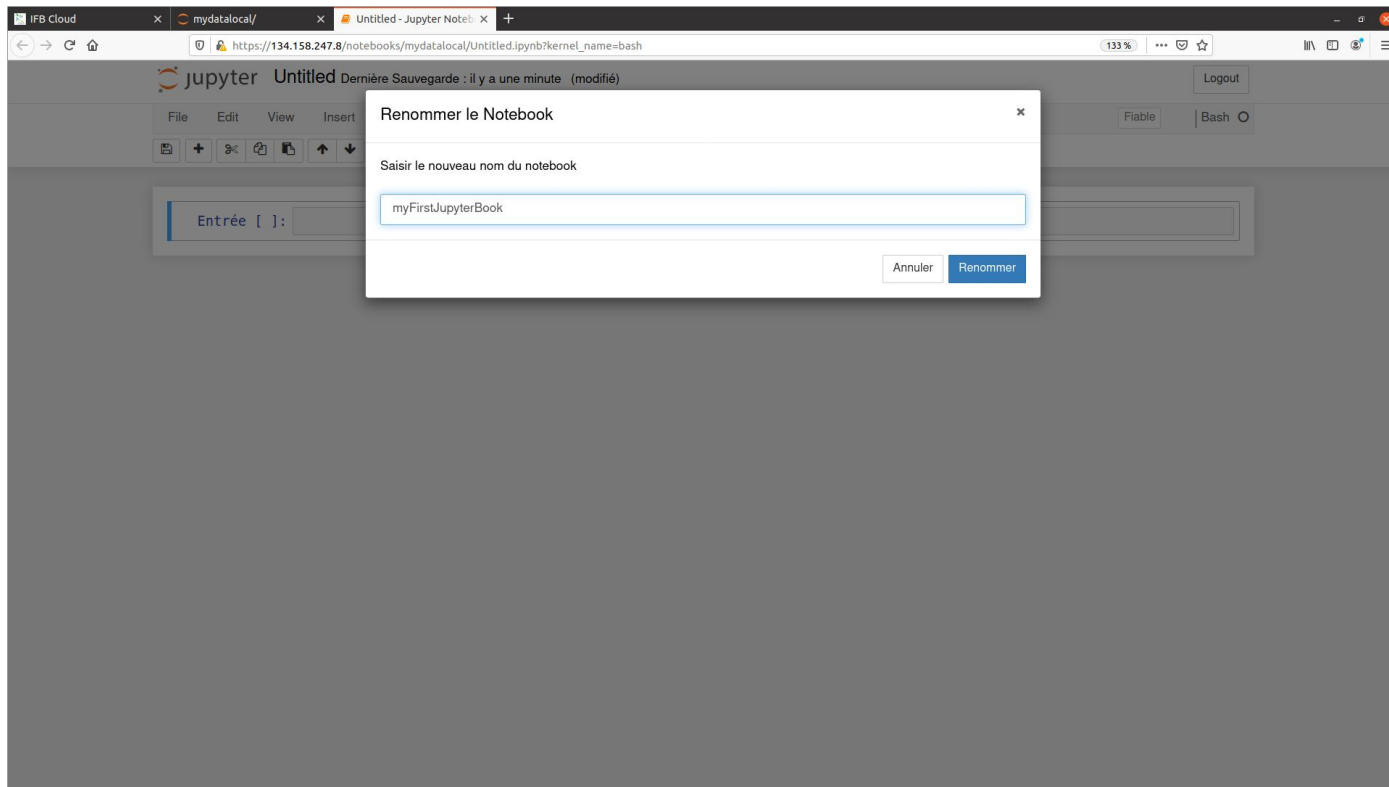
Go into the directory “work” and create a new jupyter book
-> kernel : bash



The screenshot shows the JupyterLab web interface in a browser window. The address bar displays the URL `https://134.158.247.8/tree/mydatalocal`. The interface includes a top bar with the Jupyter logo and tabs for 'Files', 'Running', and 'Clusters'. Below this, a message states 'Select items to perform actions on them.' A file browser shows a directory structure with a folder named 'mydatalocal'. On the right side, there are buttons for 'Upload', 'New', and a refresh icon. The 'New' dropdown menu is open, showing options for 'Notebook' (Bash, Julia 1.5.3, Python 3, R) and 'Other' (Text File, Folder, Terminal). The 'Bash' option is highlighted. A red dashed rectangle encloses the 'New' button and the dropdown menu.

Rename your first jupyter book

myFirstJupyterBook



Run your first bask command - *wget*

- Download all the jupyter books with the command *wget*

`wget https://itrop.ird.fr/trainings/notebook/DAY1-2/Day1_Get_Data.ipynb --no-check-certificate`

The screenshot displays the JupyterLab web interface. On the left, the file browser shows the directory structure: `/ work /`. It lists two files: `training_SV_teaching` (modified 6 minutes ago) and `MyFirstJupyterBook.ipynb` (modified seconds ago). The main area on the right is the code editor, which has a title bar showing `Jovyan@f8b3d6798e27: ~/work`. The editor contains the text `My first Juptyper book - Training SG SV` and `My first linux command - pwd`. Below this, a code cell is shown with the input `[4]: pwd` and the output `/home/jovyan/work`. The interface includes a top navigation bar with tabs for `File`, `Edit`, `View`, `Run`, `Kernel`, `Tabs`, `Settings`, and `Help`. The bottom status bar shows the user `Jovyan` and the current directory `/home/jovyan/work`.