

# The IFB Core Cluster Infrastructure

**EBAII 2021**

**Gildas Le Corguillé, Thomas Denecker & Julien Seiler**  
IFB Core Cluster taskforce

# IFB, une offre nationale

*L'IFB Core Cluster n'est pas la seule infrastructure*

NNCR - National Network  
Computing Resources

Cluster	Localisation du Data center	Coeurs	RAM (Go)	Stockage (To)
IFB Core	IDRIS - Orsay	5 042	26 542	2 000
Genotoul	Toulouse	6 128	34 304	3 000
ABiMS	Roscoff	2 608	10 600	2 500
GenOuest	Rennes	1 824	7 500	2 300
Migale	Jouy en Josas	1 084	7 000	350
BiRD	Nantes	560	4 000	500

# IFB Core Cluster - Billing



## Pourquoi ?

- Les financeurs de l'IFB demandent 10% d'autosuffisance
- Une garantie de la pérennité de l'infrastructure (*non, l'informatique n'est pas "virtuel"*)
- Discipliner les utilisateurs pour garantir une utilisation raisonnable des ressources

## Comment ?

### Forfaits / Bundle

		Disk Quota		
		< 250 GB	250 GB < < 1,5 TB	1,5 TB < < 5 TB
CPU Quota	< 10 k hours	0 €	450 €	950 €
	10 k hours < < 65 k hours	1 700 €	2 000 €	2 500 €
	65 k hours < < 100 k hours	3 000 €	3 375 €	4 000 €

## Tarifs Académiques

Provisoire

### Au-delà des forfaits / Beyond bundles

Stockage / Storage	250 €	per TB
Calcul / Compute	0.05 €	per h CPU

# Besoin d'aide ?

Il manque un outil !

J'ai plus de place !

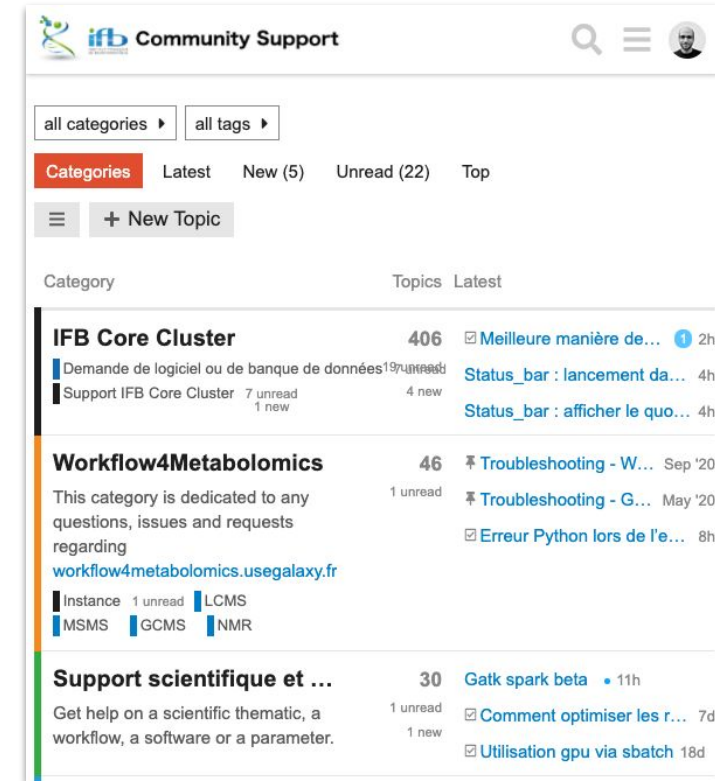
HEEEELP !

Je ne trouve pas un index...

Rejoignez la communauté IFB

Rendez-vous sur :

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



# High Performance Computer

# Votre ordinateur peut-il faire de la bioinformatique ?



## **Un ou deux microprocesseurs**

*Un microprocesseur est chargé de l'exécution des instructions élémentaires demandées par le logiciel*

## **4 à 8 Go de mémoire vive (RAM)**

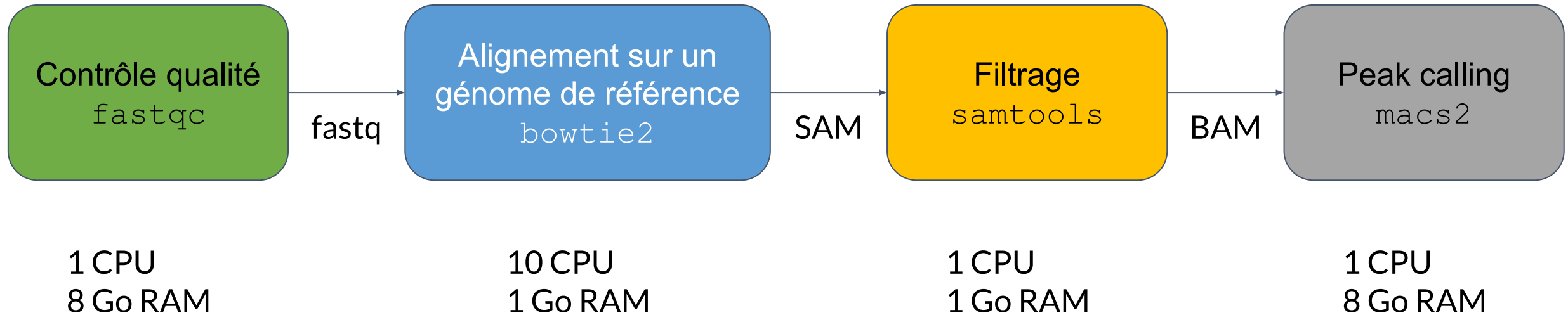
*La mémoire vive est utilisée par le microprocesseur pour traiter les données*

## **≈ 1 To d'espace de stockage**

*L'espace de stockage est utilisé pour conserver de grandes quantités de données de manière plus permanente*



# Votre ordinateur peut-il faire de la bioinformatique ?

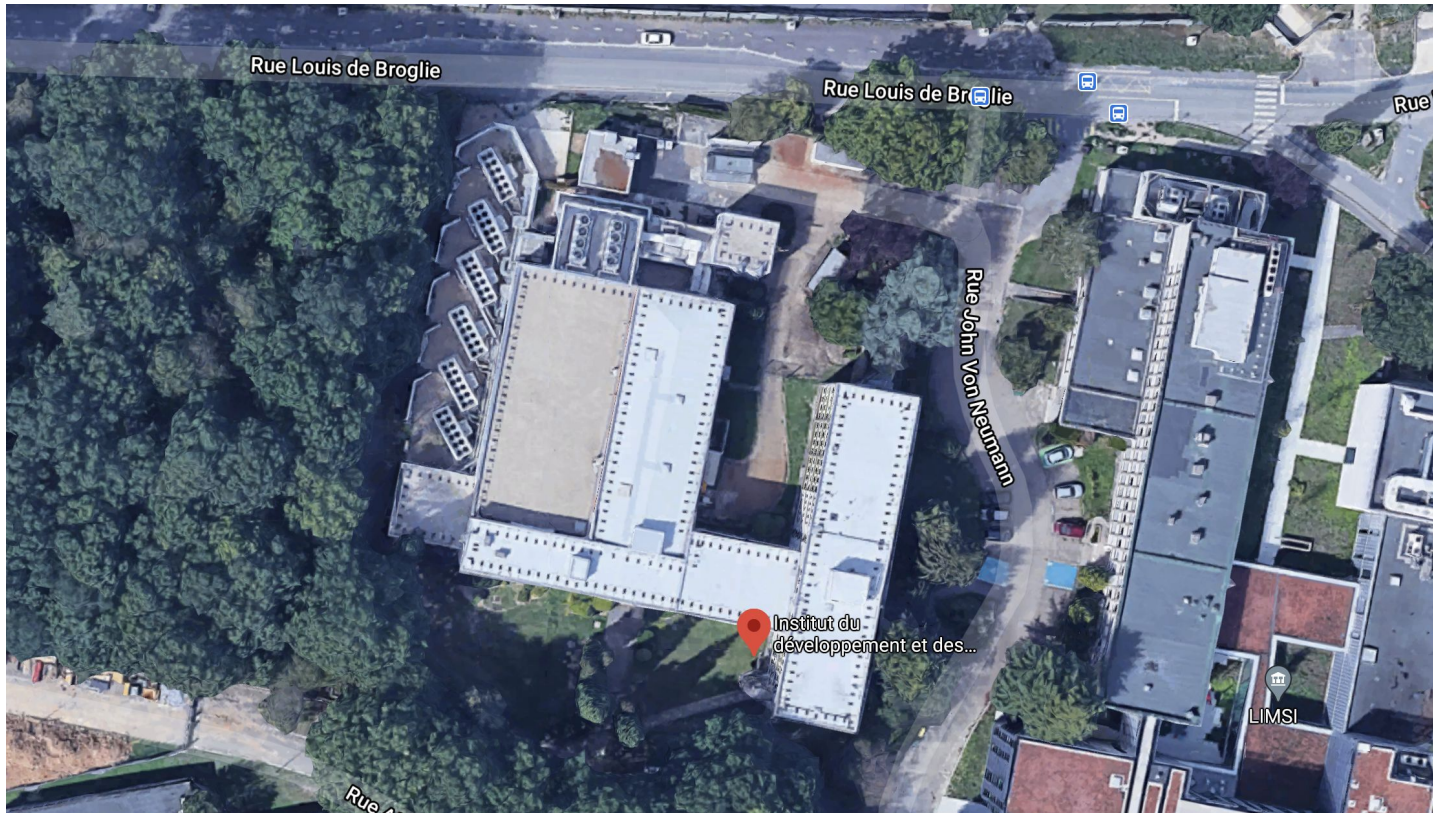


L'exécution de ce workflow nécessite au minimum toutes les ressources d'un ordinateur de bureau pendant plusieurs heures et ceci seulement pour 1 seul fichier fastq.

**Pour faire ce type d'analyse nous avons besoin d'ordinateurs plus puissants !**



# Du data center au coeur



**Le Data Center de l'IDRIS**  
Un bâtiment conçu pour  
accueillir des infrastructures  
informatiques



# Du data center au coeur

**Groupe froid**  
Pour refroidir les  
équipements



# Du data center au coeur

**Groupe électrogène**  
Pour garantir l'alimentation  
électrique



# Du data center au coeur



## **Les armoires de l'IFB**

Chaque armoire peut contenir  
80 super-ordinateurs

# Du data center au coeur



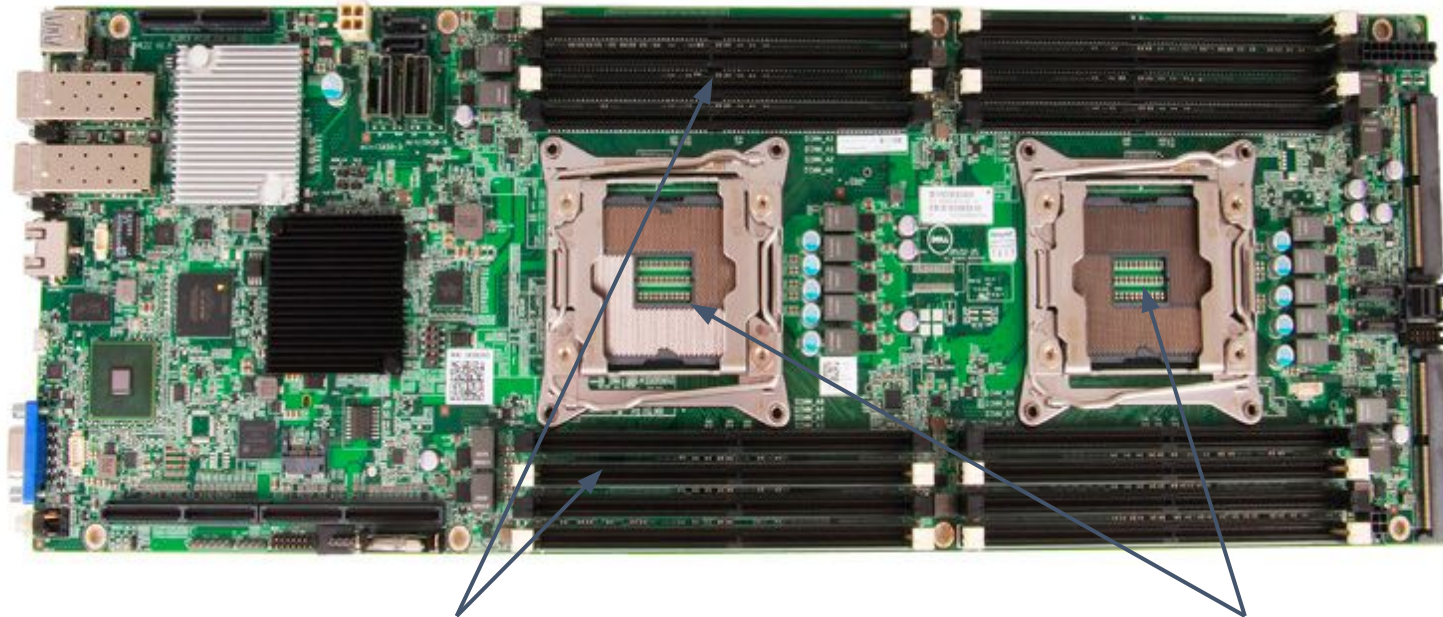
ordinateurs de calcul

Baies de stockage



# Du data center au coeur

Un ordinateur ou **noeud** de calcul



Mémoire vive

Supports processeurs

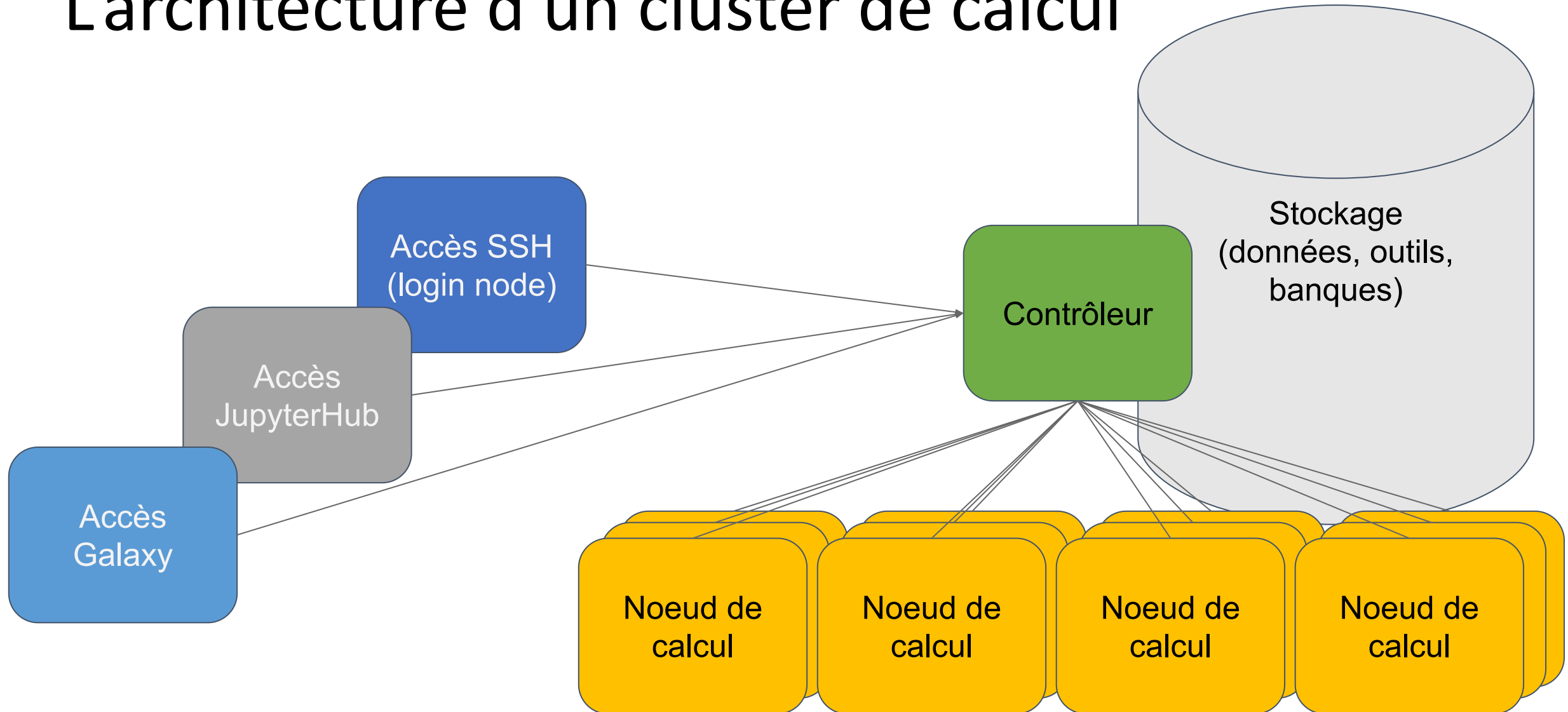
# Du data center au coeur

Un microprocesseur



Un microprocesseur contient plusieurs **coeurs**  
Chaque coeur se comporte comme un microprocesseur unique.

# L'architecture d'un cluster de calcul





# Tools

# Where is my tools?



# Software environment



`module load fastqc/0.11.9`



**BIOCONDA<sup>®</sup>**

par défaut



si une licence doit être acceptée  
si ce n'est pas dans Bioconda et difficile à intégrer ou urgent  
si un conteneur Docker existe

Pour toute demande d'installation : <https://community.france-bioinformatique.fr/>

# Besoin d'un outil ?

## Demande d'installation de FusionInspector

IFB Core Cluster Demande de logiciel ou de banque de données

This is the first time Camille has posted — let's welcome them to our community! 20d

**C** Camille

Bonjour,

Je me permets de vous contacter car dans le cadre de mon projet de master 2, j'aurais besoin de l'outil FusionInspector pour lequel mon ordinateur personnel n'est pas assez puissant. C'est un outil permettant d'analyser les fusions en RNAseq. Ne sachant pas comment procéder, serait-il possible de vous l'installer ? Voici l'adresse <https://github.com/FusionInspector/FusionInspector/wiki/installing-FusionInspector>.

Je vous serais très reconnaissante de votre aide.

Cordialement,

Camille Baron encadrée par Audrey Gros à l'université de Bordeaux

✓ Solved by [gildaslecorguille](#) in [post #4](#)

Hop: module load fusion-inspector/2.2.1

created 20d last reply 19d 4 replies 25 views 2 users 2 links

## Installation (nonpareil) @team.software

IFB Core Cluster Demande de logiciel ou de banque de données

**E** echase 3 Feb

Bonjour @team.software

Serait-il possible d'installer nonpareil sur le cluster, SVP ?

GitHub

**Imrodriguezr/nonpareil**

Estimate metagenomic coverage and sequence diversity - Imrodriguezr/nonpareil

Merci pour votre temps,

Emily

✓ Solved by [julien](#) in [post #3](#)

nonpareil 3.3.4 est à présent disponible sur le cluster : module load nonpareil/3.3.4 Bonne journée, Julien

created 3 Feb last reply 4 Feb 2 replies 28 views 2 users 2 likes 1 link

## Installation pyslim @team.software

IFB Core Cluster Demande de logiciel ou de banque de données

**G** Guillaume\_Lan-Fong 24d

Bonjour,

Faisant suite à mon post pour l'installation de SLiM, j'aimerais savoir s'il était possible d'installer le package python "pyslim" (<https://pyslim.readthedocs.io/en/latest/installation.html>) permettant de travailler sous python sur les .trees générés par SLiM.

L'installation devrait être possible via un simple : `python3 -m pip install pyslim`

En vous remerciant par avance,

Lan-Fong Guillaume

✓ Solved by [gildaslecorguille](#) in [post #7](#)

Hop: module load pyslim/0.501

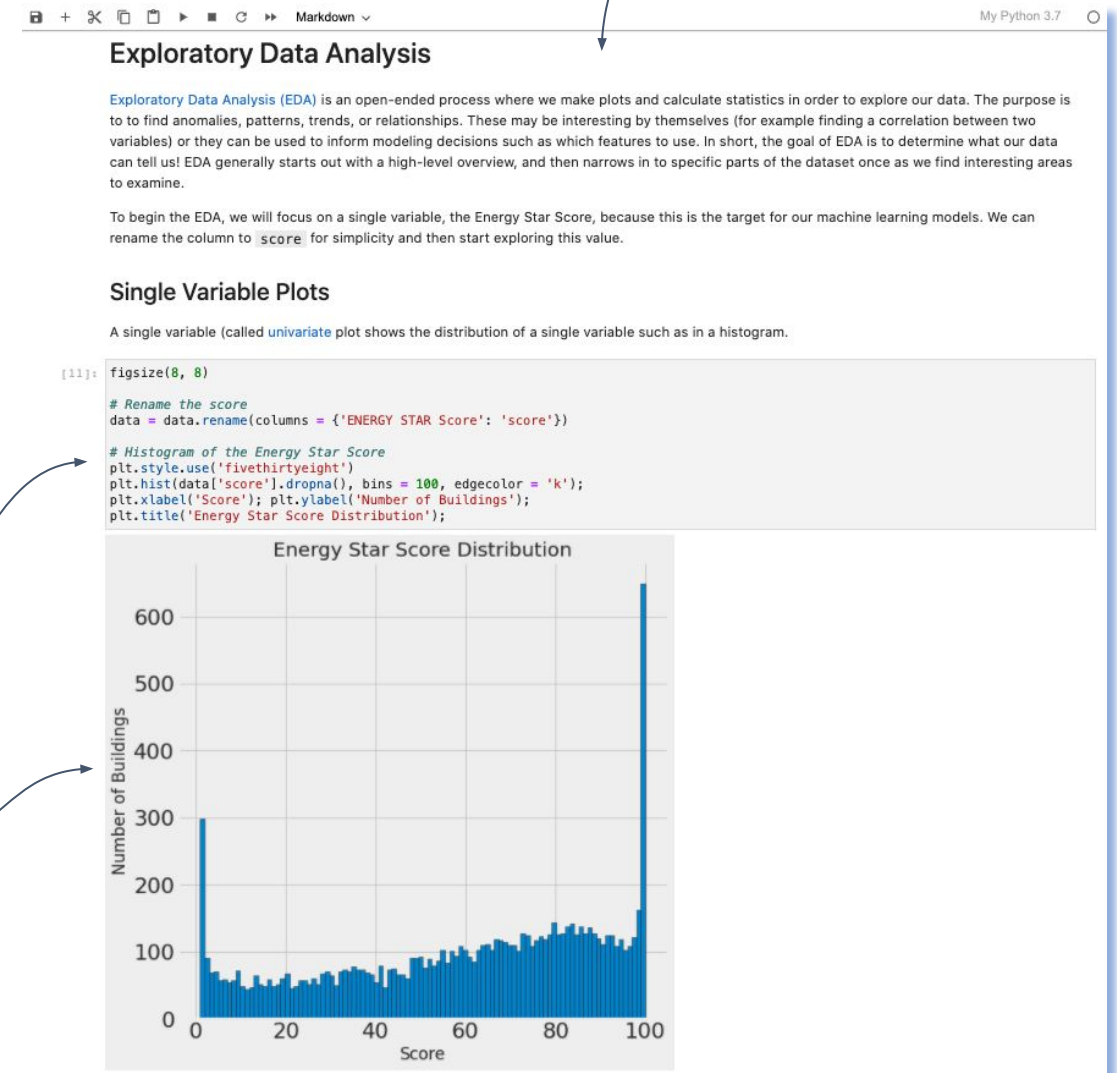
created 24d last reply 11d 7 replies 38 views 3 users 1 like 1 link

# Jupyter

# Qu'est-ce qu'un Notebook Jupyter ?

- Un fichier spécial avec l'extension `.ipynb`
- Combinaison de **Markdown** et de **code**
- Le code peut être exécuté à l'intérieur du notebook
- La sortie du code est intégrée directement dans le notebook

Cellule  
Markdown



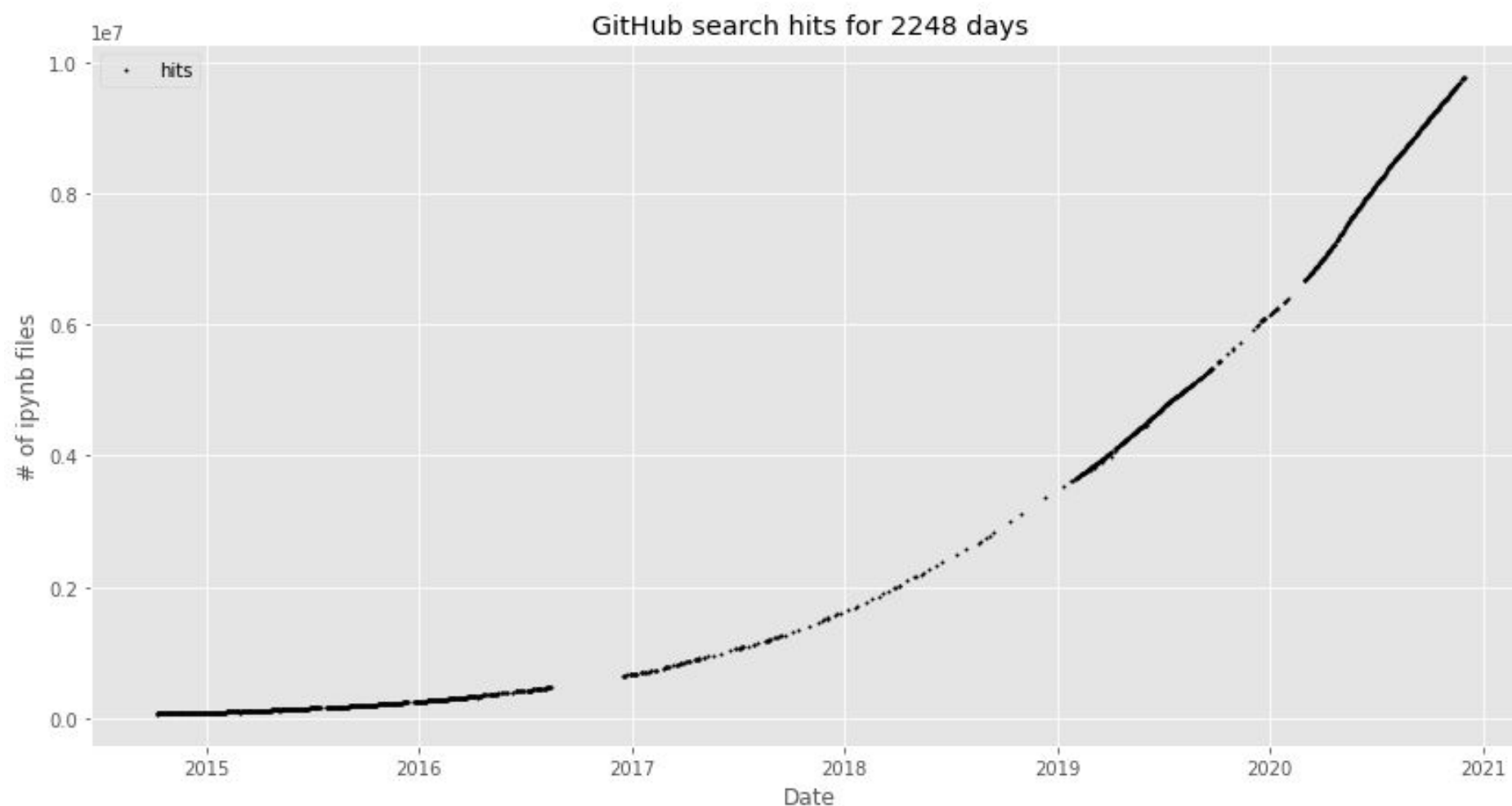
Cellule  
de code

Sortie du  
code

# Qu'est-ce qu'un Notebook Jupyter ?



# Les notebooks sont populaires



Source : Nbviewer - <https://nbviewer.jupyter.org/github/parente/nbestimate/blob/master/estimate.ipynb>



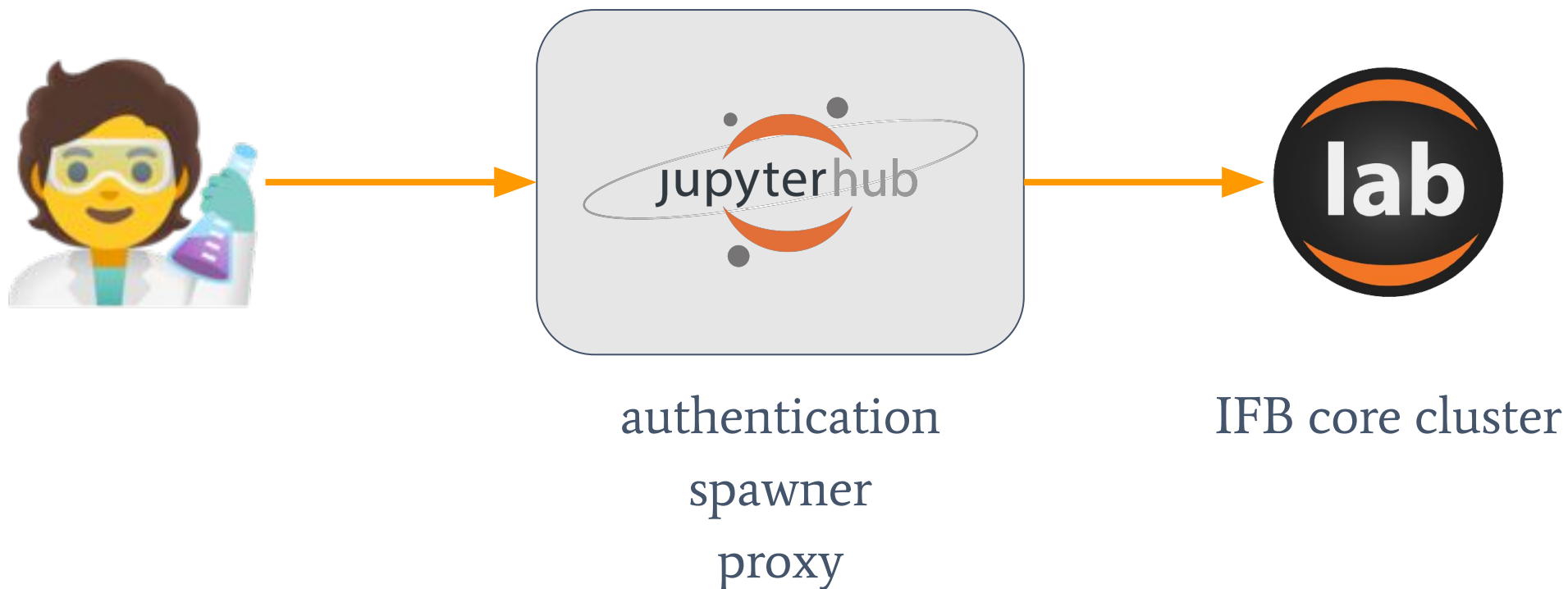
# Pourquoi s'intéresser aux notebooks ?

- Les notebooks vous permettent d'analyser des données et de rédiger des rapports en un seul endroit
- Ils prennent en charge la visualisation des données en temps réel
- Vous pouvez facilement inclure une section interactive dans un notebook
- En accord avec le principe de la science reproductible

**Les Notebooks Jupyter sont  
le cahier de laboratoire pour la science des données.**

# Qu'est-ce que JupyterHub

JupyterHub est une application web qui vous permet de créer des serveurs JupyterLab sur un cluster ou une infrastructure en nuage.



# SLURM

Présentation et démonstration ->

[https://gitlab.com/ifb-elixirfr/cluster/tutoriel\\_slurm](https://gitlab.com/ifb-elixirfr/cluster/tutoriel_slurm)