Offres de services du pôle
(Infrastructures HPC & VRE)

Gwenaël CAËR, Cyril GERMINEAUD

contact@odatis-ocean.fr | www.odatis-ocean.fr

Les infrastructures HPC

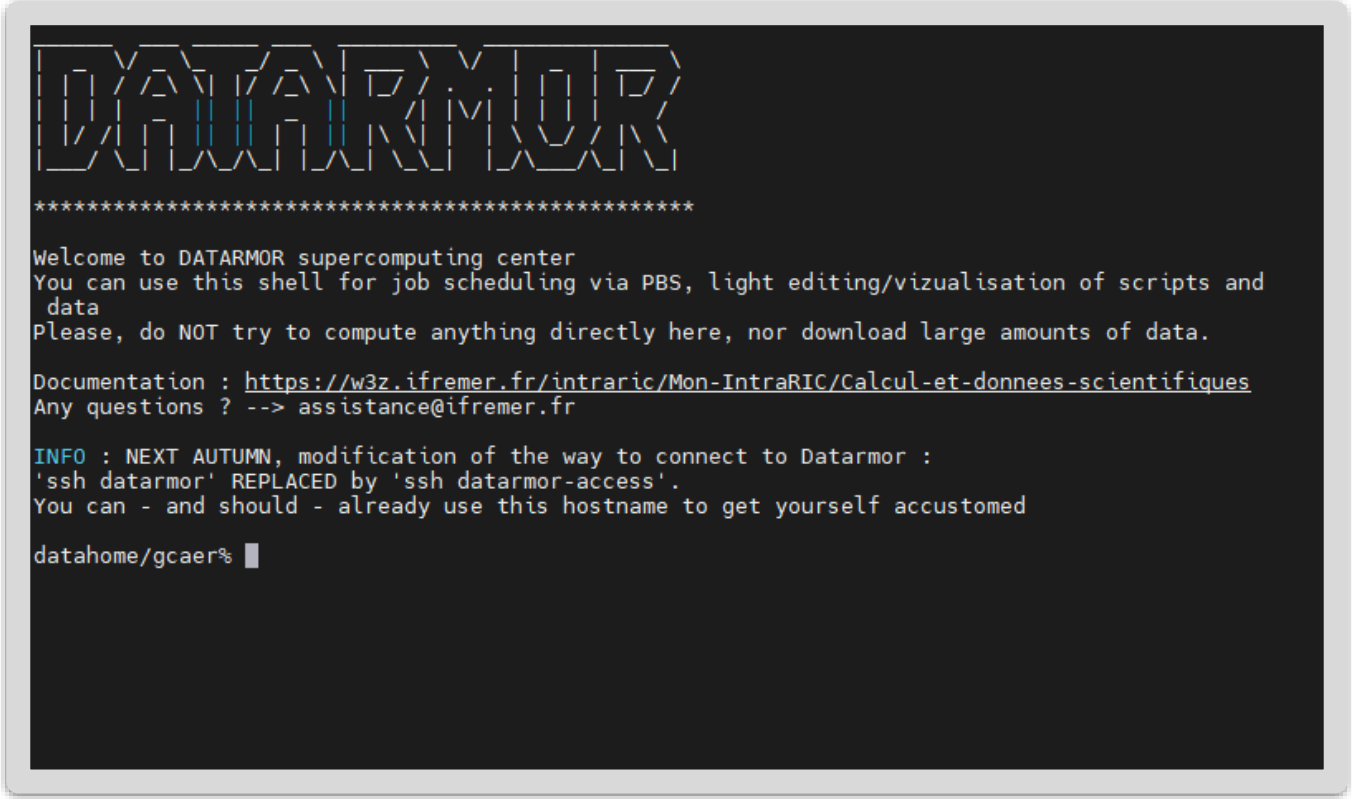
Le pôle ODATIS mets à disposition l'accès aux infrastructures de calcul du CNES (*TREX*) et de l'Ifremer (*DATARMOR*).

Cluster du CNES (TREX)

Cluster de l'Ifremer (DATARMOR)

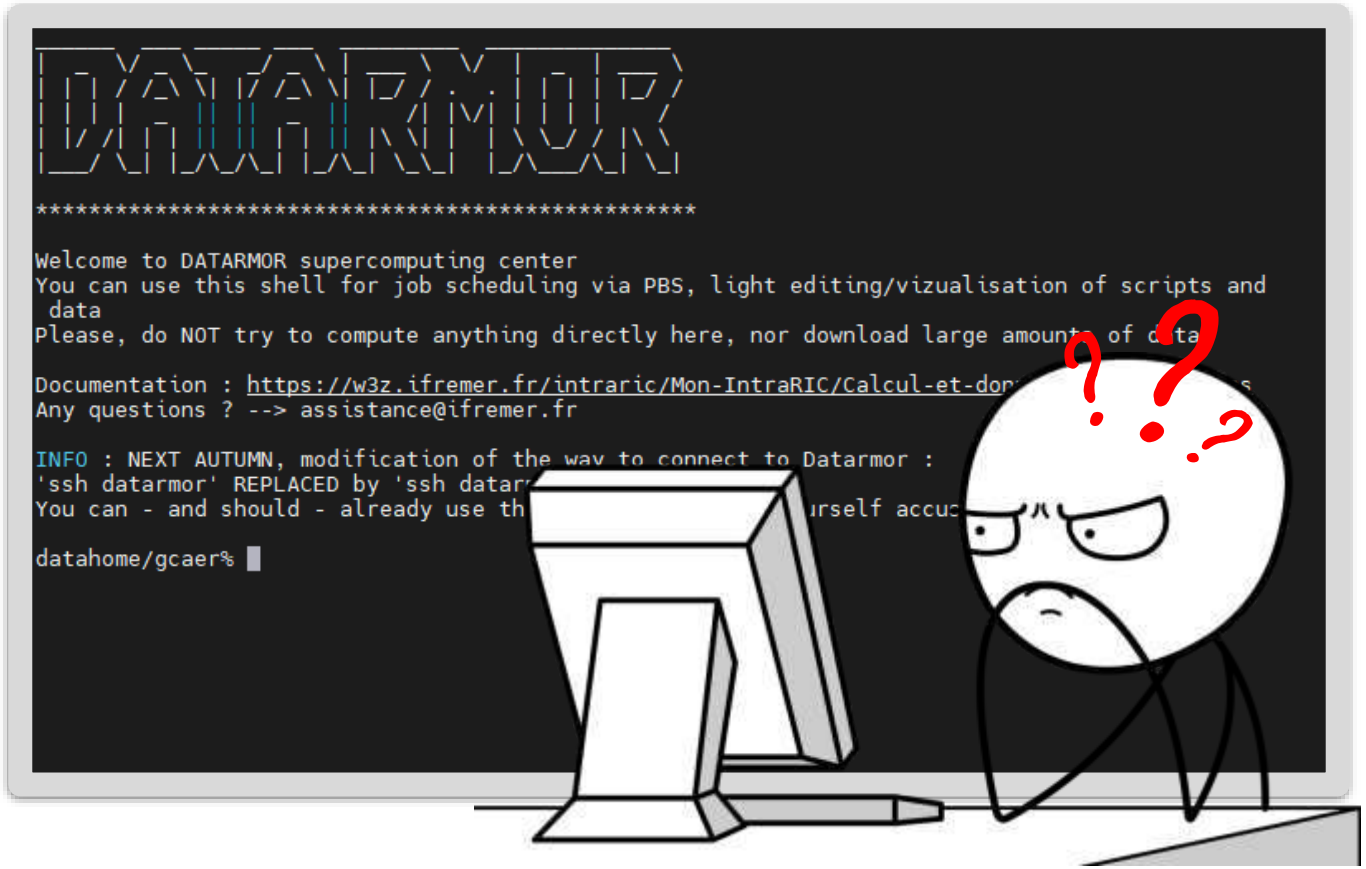
Les infrastructures HPC

Comment y accéder ? Une simple commande SSH, et vous voilà connecté au cluster.



Les infrastructures HPC

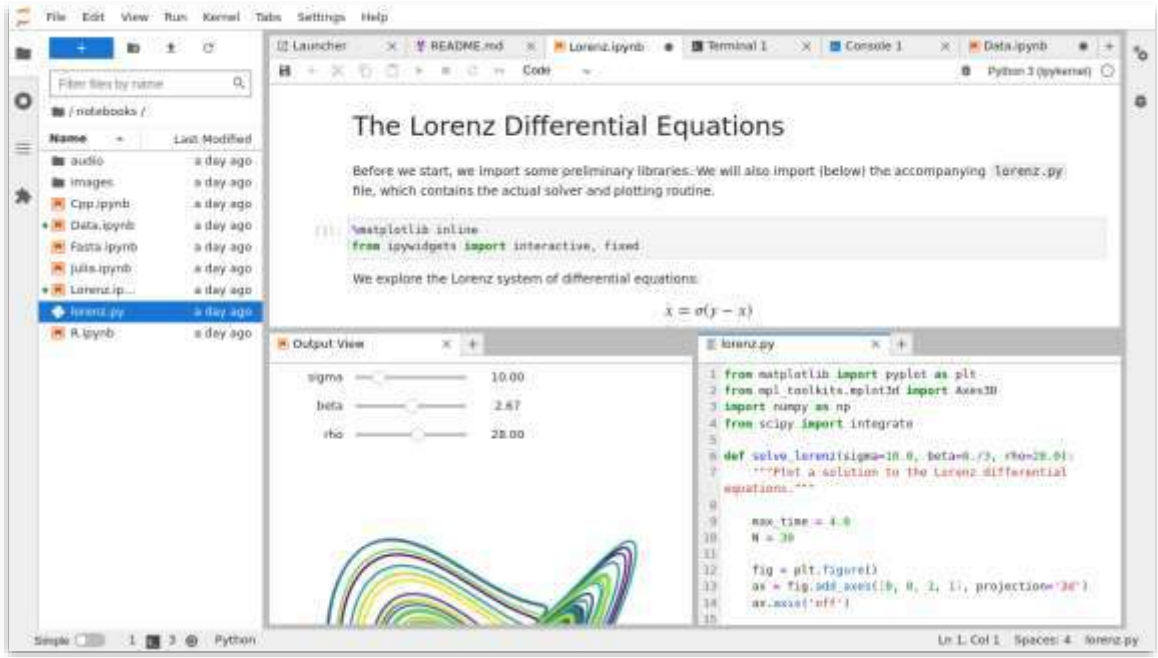
Comment y accéder ? Une simple commande SSH, et vous voilà connecté au cluster.



VRE (Environnement Virtuel de Recherche)

Un VRE pour Environnement Virtuel de Recherche, correspond à une plateforme interactive, permettant de faciliter l'accès aux données et aux ressources des infrastructures (*calculs, logiciels, documentations …*).

jupyterlab



VRE (Concrètement, comment ça fonctionne ?)

Travailler à distance … comme si vous étiez en local.

1 **Accéder à l'URL** *(et se connecter)*

https://jupyterhub.cnes.fr *(ou ifremer.fr)*

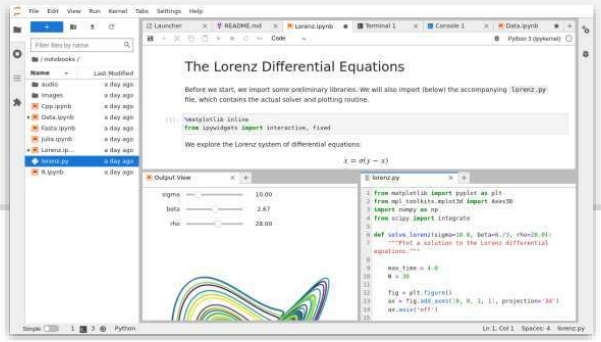
connexion

2 **Sélectionner les ressources** *(CPU, RAM, Durée, ..)*

Sélection des ressources

1 coeur, 4gb, 8 heures
8 coeurs, 32gb, 8 heures
28 coeurs, 115gb, 8 heures

3 **Commencer à l'utiliser la VRE** *(Données, Analyse, Visualisation, …)*



VRE (Concrètement, comment ça fonctionne ?)

Travailler à distance … comme si vous étiez en local.

1 **Accéder à l'URL** *(et se connecter)*

https://jupyterhub.cnes.fr *(ou ifremer.fr)*

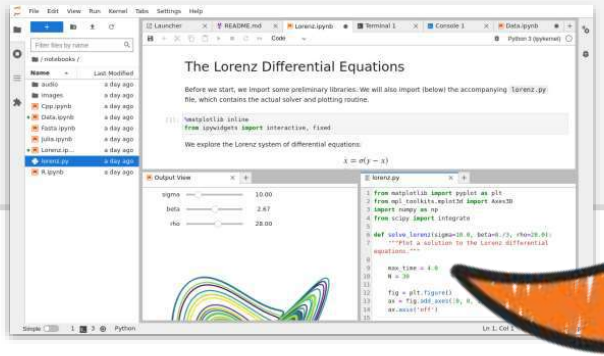
connexion

2 **Sélectionner les ressources** *(CPU, RAM, Durée, ..)*

Sélection des ressources

1 coeur, 4gb, 8 heures
8 coeurs, 32gb, 8 heures
28 coeurs, 115gb, 8 heures

3 **Commencer à l'utiliser la VRE** *(Données, Analyse, Visualisation, …)*



Les données

- **Accès au catalogue** ODATIS
- **Accès distant** *(via S3, HTTPS, OPeNDAP, FTP, …)*

DATA TERRA
ODATIS

VRE (Concrètement, comment ça fonctionne ?)

Travailler à distance … comme si vous étiez en local.

1 **Accéder à l'URL** *(et se connecter)*

https://jupyterhub.cnes.fr *(ou ifremer.fr)*

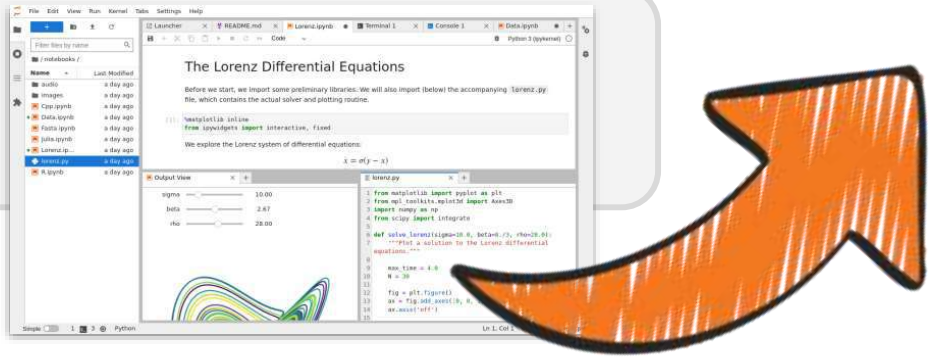
connexion

2 **Sélectionner les ressources** *(CPU, RAM, Durée, ..)*

Sélection des ressources

1 coeur, 4gb, 8 heures
8 coeurs, 32gb, 8 heures
28 coeurs, 115gb, 8 heures

3 **Commencer à l'utiliser la VRE** *(Données, Analyse, Visualisation, …)*



Les données

- **Accès au catalogue** ODATIS
- **Accès distant** *(via S3, HTTPS, OPeNDAP, FTP, …)*

Les ressources

- **Calcul** (CPU/GPU, RAM, …)
- **Langages** (Python, R, Julia)
- **Environnements** (de base *(pangeo)* ou personnalisés)
- **Outils** (QGIS, Git, …)

DATA TERRA
ODATIS

VRE (Concrètement, comment ça fonctionne ?)

Travailler à distance … comme si vous étiez en local.

1 **Accéder à l'URL** *(et se connecter)*

https://jupyterhub.cnes.fr *(ou ifremer.fr)*

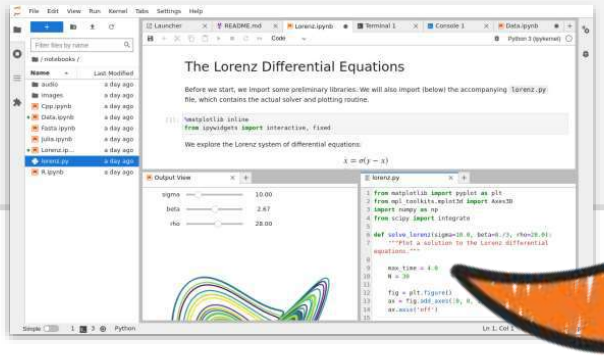
connexion

2 **Sélectionner les ressources** *(CPU, RAM, Durée, ..)*

Sélection des ressources

1 coeur, 4gb, 8 heures
8 coeurs, 32gb, 8 heures
28 coeurs, 115gb, 8 heures

3 **Commencer à l'utiliser la VRE** *(Données, Analyse, Visualisation, …)*



Les données

- **Accès au catalogue** ODATIS
- **Accès distant** *(via S3, HTTPS, OPeNDAP, FTP, …)*

Les ressources

- **Calcul** (CPU/GPU, RAM, …)
- **Langages** (Python, R, Julia)
- **Environnements** (de base *(pangeo)* ou personnalisés)
- **Outils** (QGIS, Git, …)

Les services (Helpdesk)

- **Documentation** (Premiers pas, FAQ, Tutoriels ...) https://odatis-public.gitlab-pages.ifremer.fr/vre/documentation
- **Cas d'usages** (Illustration de l'utilisation de la VRE pour traiter des problématiques scientifiques.)

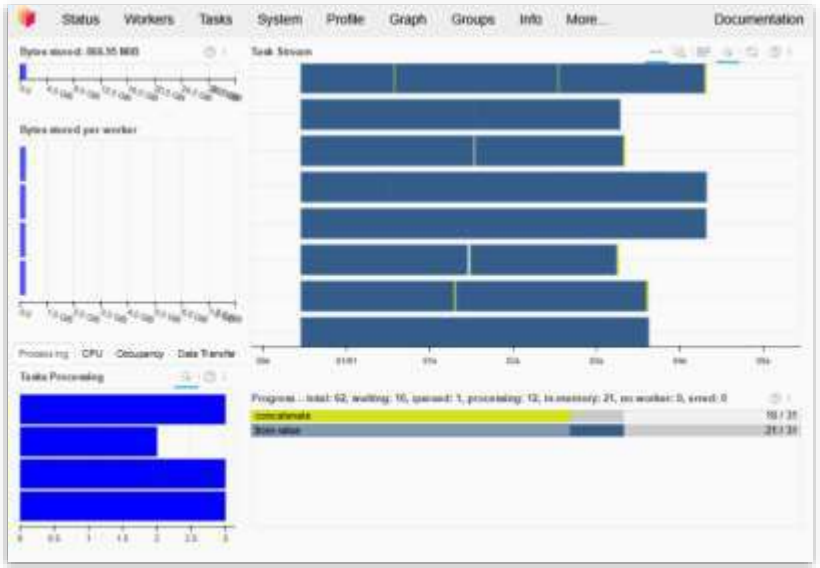
DATA TERRA

ODATIS

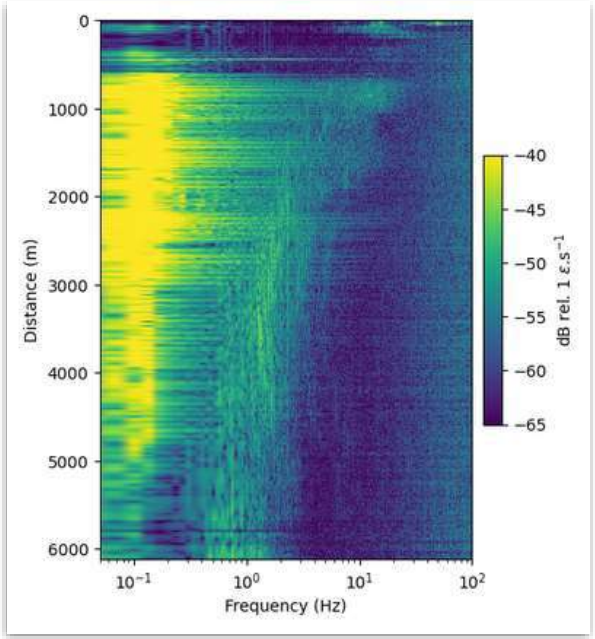
Cas d’usages

Traitement des données de fibre optique (DAS)

Dans le cadre d’une campagne menée en Sicile par une équipe de Geo-Ocean, un volume très important de données a été acquis (environ 20To). L’idée de ce cas d’usages est d’étudier comment optimiser la lecture, le traitement et la visualisation d’un tel volume de données ?



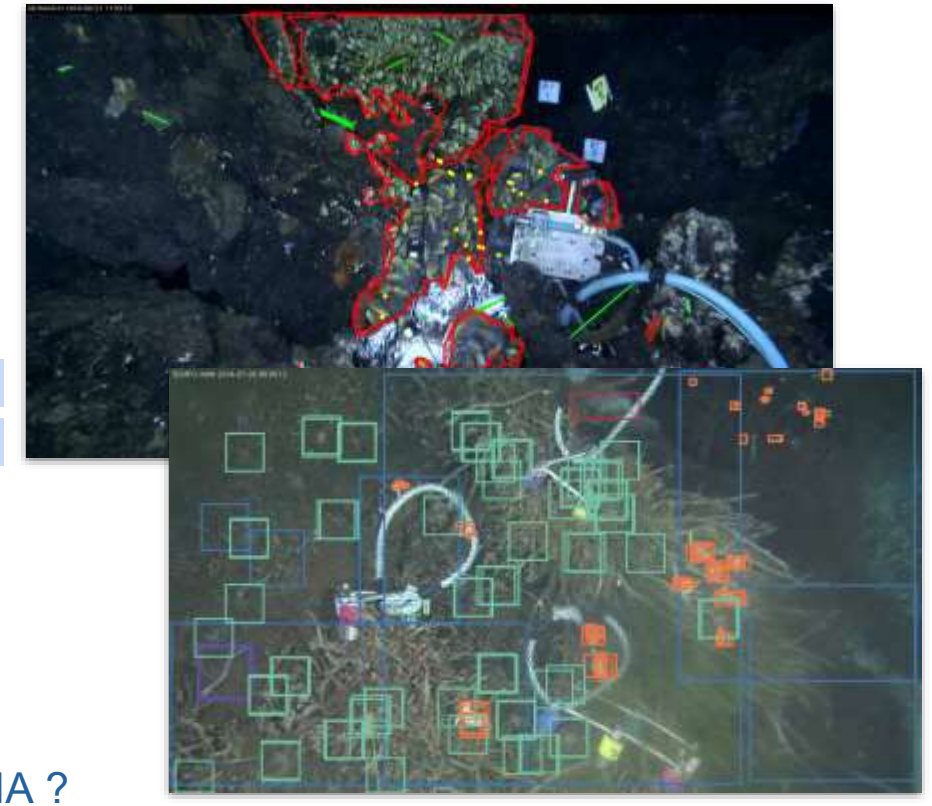
- Les notions abordées:
  - Comment paralléliser la lecture de milliers de fichiers ?
  - Comment appliquer des traitements en parallèle ?
  - Comment visualiser les données de manière interactive et optimisée ?
- Développement d’une librairie en cours.



Cas d’usages

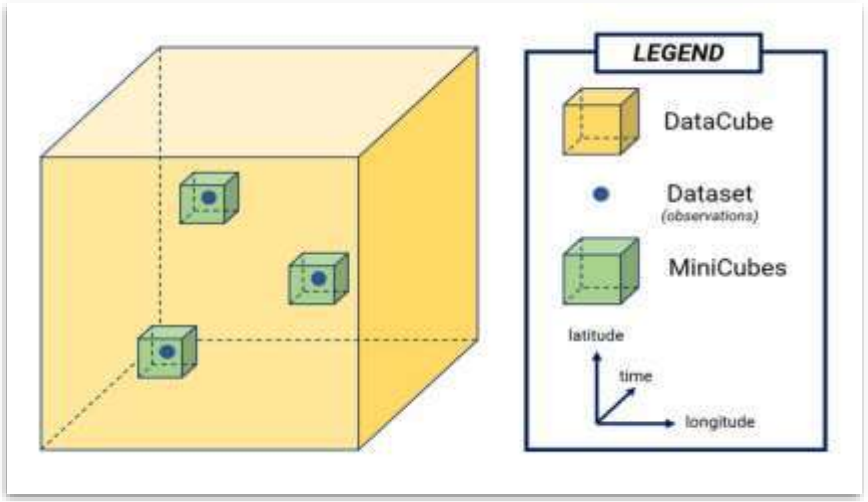
Traitement des données d’imagerie benthique

Dans le cadre du projet DeepSeaSpy mené par le LEP, un grand nombre d’annotation d’espèce provenant de la science participative sont disponible. L’idée de ce cas d’usage est d’étudier comment utiliser des données citoyennes pour entraîner un modèle d’IA, afin d’annoter de futures images de manière automatique ?



- Les notions abordées:
  - Comment accéder aux données sur le HPC ?
  - Comment nettoyer les données ?
  - Comment utiliser les ressources GPU pour entraîner un modèle d’IA ?
  - Comment utiliser le modèle pour annoter de nouvelles images ?

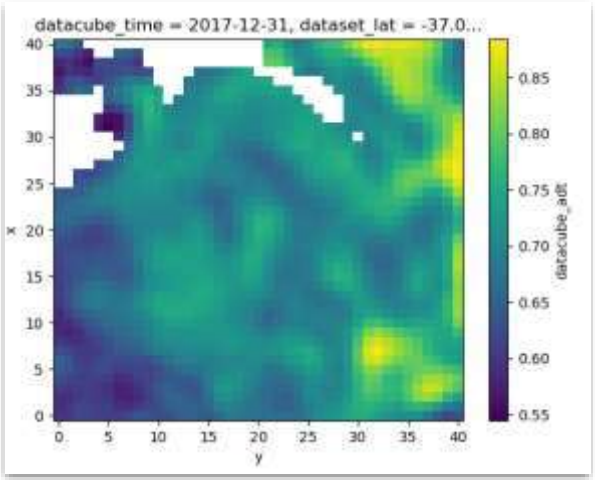
Cas d’usages



Colocalisation des données in situ et satellitaires

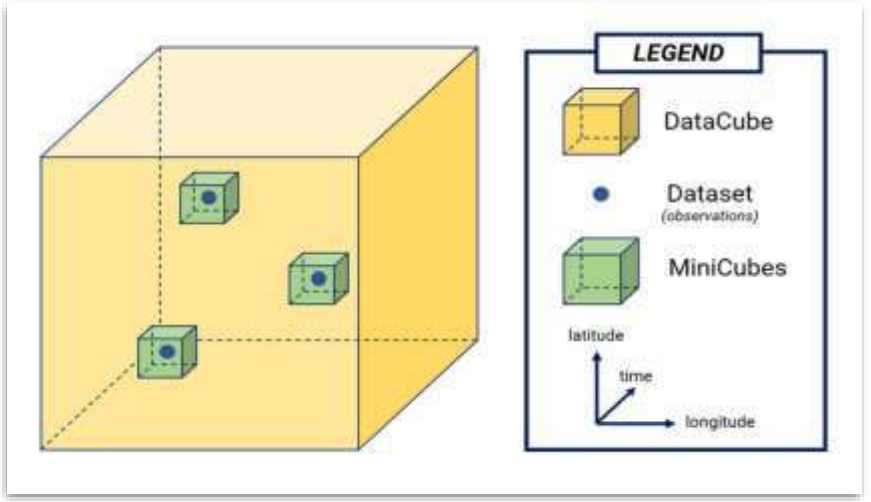
Dans le cadre de nombreux travaux de recherche, une étape importante est la colocalisation des données. L’idée de ce cas d’usage est d’étudier comment utiliser des outils communs aux géosciences (Xarray, Kerchunk, Dask), pour optimiser la colocalisation d’un grand nombre de données ?

- Les notions abordées:
  - Comment utiliser Kerchunk pour émuler un fichier zarr à partir de milliers de fichiers unitaire ?
  - Comment utiliser Xarray pour ouvrir un datacube ?
  - Comment utiliser et optimiser dask pour la lecture et l’extraction des données ?
- Développement d’une librairie de colocalisation : PytCube



PytCube

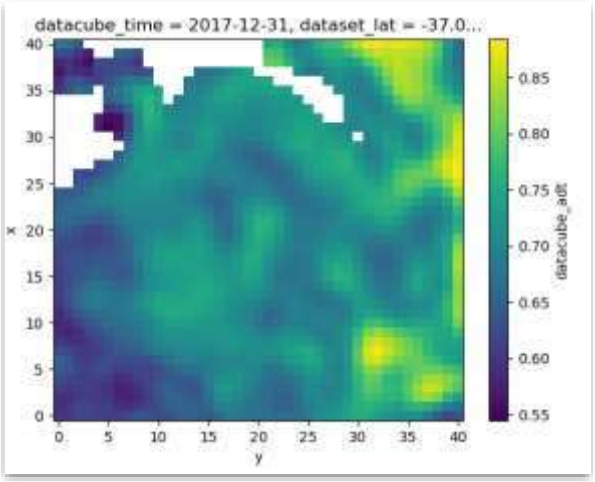
Cas d’usages



Colocalisation des données in situ et satellitaires

Dans le cadre de nombreux travaux de recherche, une étape importante est la colocalisation des données. L’idée de ce cas d’usage est d’étudier comment utiliser des outils communs aux géosciences (Xarray, Kerchunk, Dask), pour optimiser la colocalisation d’un grand nombre de données ?

- Les notions abordées:
  - Comment utiliser Kerchunk pour émuler un fichier zarr à partir de milliers de fichiers unitaire ?
  - Comment utiliser Xarray pour ouvrir un datacube ?
  - Comment utiliser et optimiser dask pour la lecture et l’extraction des données ?
- Développement d’une librairie de colocalisation : PytCube



Vous avez une idée de cas d’usages ? N’hésitez pas à nous contacter.

PytCube

Perspectives

L’offre de service VRE du pôle ODATIS se concentre pour l’instant sur la solution JupyterLab, mais elle a vocation à s’élargir à l’avenir en incluant d’autres solutions, telles que Galaxy.