



DATA
TERRA



ODATIS

Atelier Accès aux Données 2024 CDS-AVISO

10-11/10/2024 – Plouzané, Ifremer
Cyril GERMINEAUD (CNES) pour l'équipe AVISO



10/10/2024

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

LE CDS-AVISO en quelques chiffres

« ARCHIVAGE, VALIDATION ET INTERPRÉTATION DES DONNÉES DES SATELLITES OCÉANOGRAPHIQUES »



TYPES DE PRODUITS

~60

- De la mesure satellite aux produits à valeur ajoutée
- Le long de la trace et grillés
- Temps réel, temps différé,...

EQUIPES DE PRODUCTION

27

CNES
CLS
CTOH
Magellium, ...

MISSIONS SATELLITES

21

Jason-1/2/3
Sentinel-3A/B/6 MF
Saral

Cryosat-2
CFOSAT
SWOT

SERVICES DE DISTRIBUTION

7

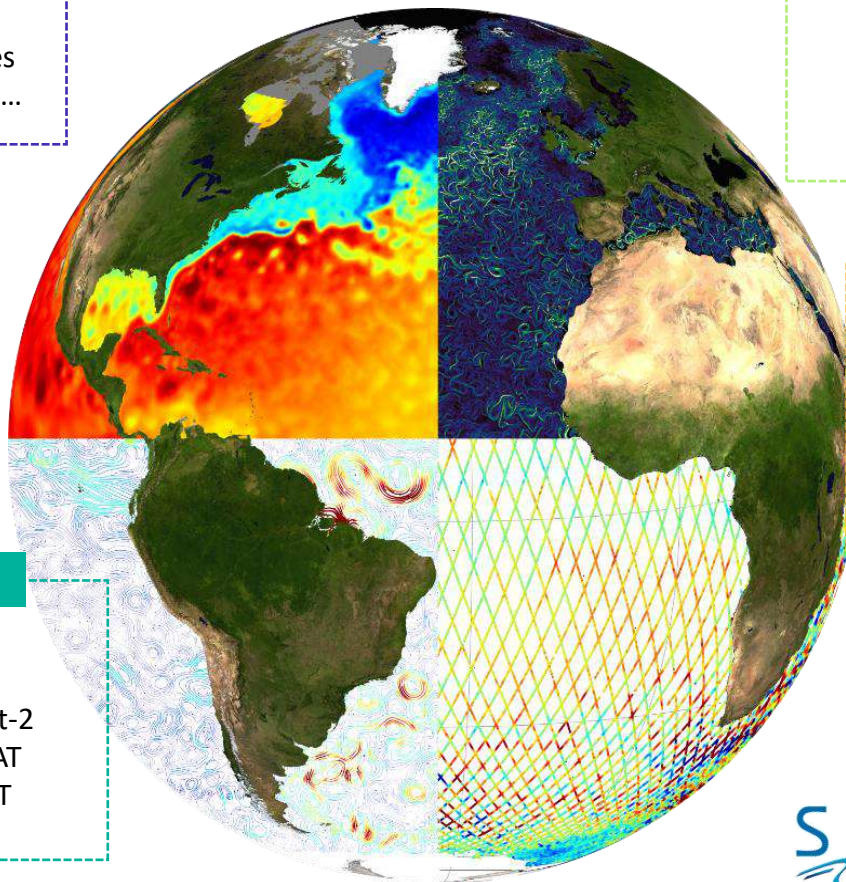
AVISO CNES Data
Center
FTP/SFTP
OPeNDAP

HTTP
WMS
NetCDFSubset,
...

VISUALIZATION SERVICES

3

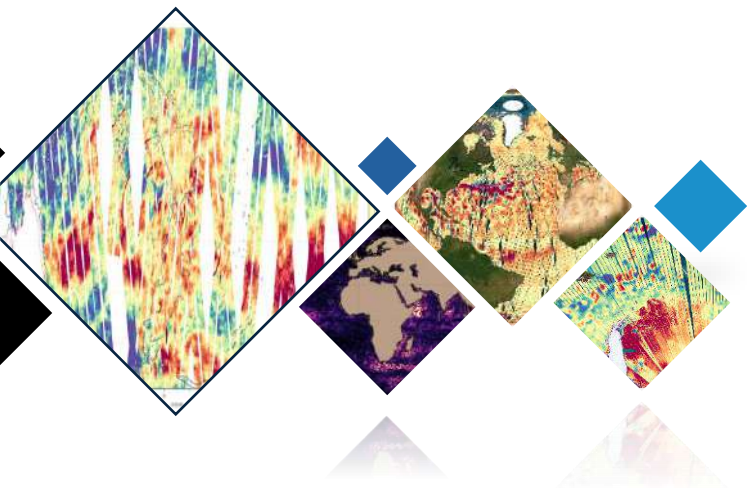
Seewater
Ocean Virtual Lab (ESA/CNES)
Indicateurs Climatiques (e.g.,
GMSL)



SERVICE
ALTIMÉTRIE
&
LOCALISATION
PRECISE



Services d'accès aux données



Protocols/APIs for exposing/querying metadata
and have access to related data sets based on FAIR
(Findable, Accessible, Interoperable and Reusable) principles



Plateformes et Services :

- AVISO CNES Data Center (catalogue d'archivage)
- THREDDS (OPeNDAP, HTTP & WMS, NetCDFSubset)
- FTP/SFTP,...



Offre d'hébergement type cloud (HPC CNES)

SWOT OCEAN DATA ACCESS & SERVICES IN A NUTSHELL

SWOT

JUPYTER
NOTEBOOKS

DATA ACCESS

- SWOT ocean data (L2, L3 & L4 products)
- Easy access to other data sets
- Catalog requests & download tools...

ENVIRONMENT

- Research-orientated Python libraries
- SWOT-dedicated toolbox (community-driven GitHub)
- Free hosting of SWOT Projects on CNES HPC

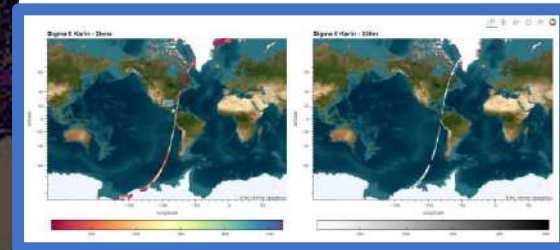
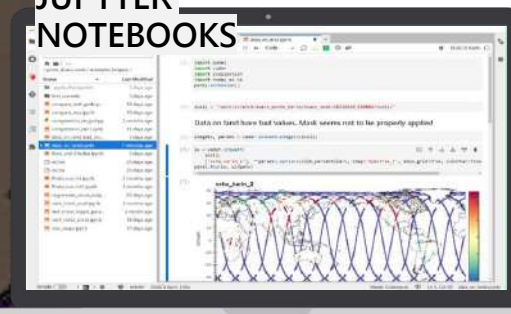
DOCUMENTATION

- Dedicated use case examples
- User guides, FAQ,...

USER SUPPORTS

- SWOT data training
- Help for code optimization
- Helpdesk...

HELPDESK & TECHNICAL SUPPORT
aviso-swot@altimetry.fr



INTERACTIVE VISUALISATION TOOL





DATA
TERRA



ODATIS

Merci pour votre attention
Questions ?



10/10/2024

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

Fiches de métadonnées – GeoNetwork

Catalogue Sextant d'édition des fiches AVISO



Catalogue d'accès à la donnée du pôle Odatis



Futur système de gestion

Moissonnage journalier
des fiches AVISO

Fiches produits,
handbooks et services
d'accès à la donnée

Catalogues hors IR Data Terra



Moissonnage et Mise à
disposition des fiches AVISO

Instance AVISO



Produits/données AVISO

<https://www.aviso.altimetry.fr/en/data/products.html>



Accès aux Notebooks – l'approche galerie

Par exemple :

- <https://swot-community.github.io/SWOT-galleries/index.html>



DATA
TERRA



ODATIS

Atelier Accès aux Données 2024 DataLake CNES / Protocole S3

10-11/10/2024 – Plouzané, Ifremer
Cyril GERMINEAUD (CNES)



10/10/2024

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

Le Datalake du CNES : Protocole S3

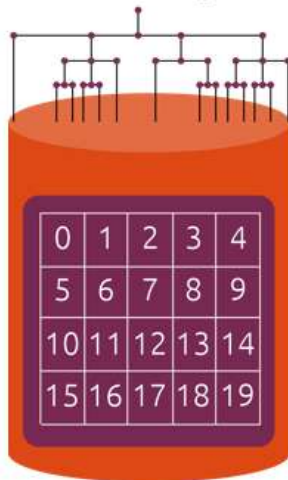
Utiliser le Datalake => utiliser le protocole S3

système de fichiers : **POSIX**

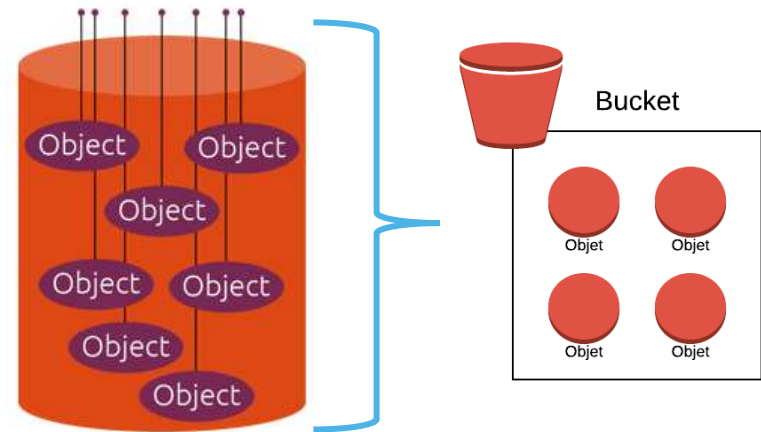
VS

AWS S3 (Simple Storage Service) : stockage **objet**

File Storage



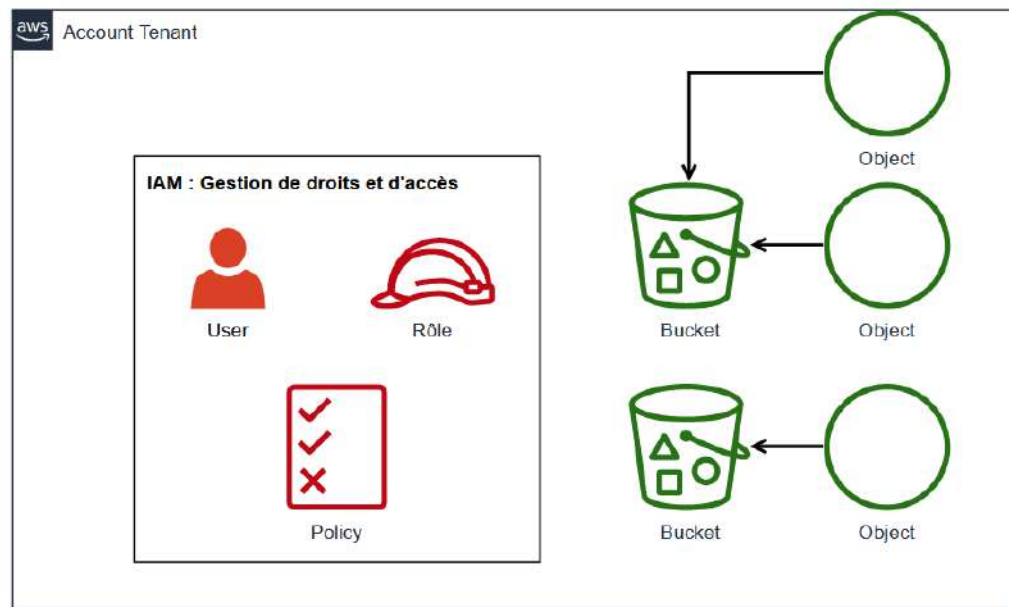
Object Storage



Le Datalake du CNES : Protocole S3

Concepts Importants

- Organisation des données :
 - Tenant / Projet (Account au sens AWS)
 - Buckets à l'intérieur des tenants
 - Des objets placés à plat au sein des buckets
- Les accès via entités IAM :
 - Policy : les politiques d'accès
 - User/Group : utilisateurs/groupes IAM authentifiés avec une paire de clé
 - Rôle : Entité détenant des droits d'accès à des ressources (bucket/objet) et pouvant être endossé temporairement



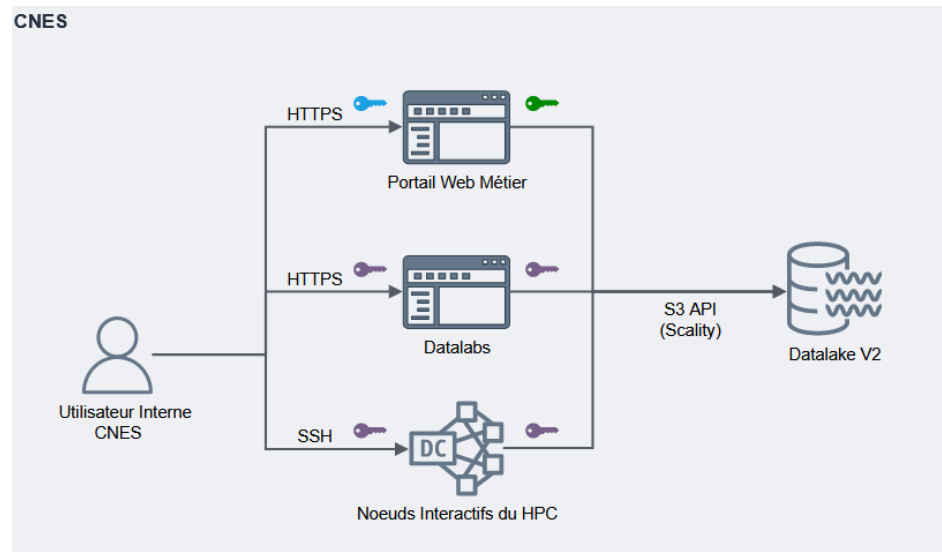
Le Datalake du CNES : Protocole S3

Credentials S3 == paire de clés

- `aws_access_key_id`
- `aws_secret_access_key`
- `aws_session_token` (si utilisation d'un rôle)

Différents modes d'authentification

- Via credentials permanents :
 - IAM User (pas lié au compte IPA)
 - À destination des projets / producteurs de données
- Via credentials temporaires (max 12h) :
 - En endossant un rôle grâce à son compte IPA
 - À destination de tout les utilisateurs



Légende :

- Utilisateurs gérés par l'application métier. Inconnu du Datalake V2.
- Utilisateurs gérés par l'annuaire IPA CNES. A accès aux services du SIS. Utilisateurs disposants des groupes IPA permettant un accès au Datalake V2.
- Utilisateurs techniques disposants des groupes permettant un accès au Datalake V2.



Les services AWS

Uniquement 3 services AWS sont disponibles sur le Datalake (gestion stockage S3)



Gestion de ressources
[bucket et object]



Gestion de entités IAM
[user, group, role et policy]
Chaque account possède son propre IAM



Gestion de credentials temporaires avec une durée de 12h

Le Datalake respecte les APIs standards AWS

la documentation AWS S3 est donc une excellente base pour commencer



Le Datalake du CNES : Protocole S3

Accès au Datalake

- Un endpoint : <https://s3.datalake.cnes.fr>
- Une région : us-east-1

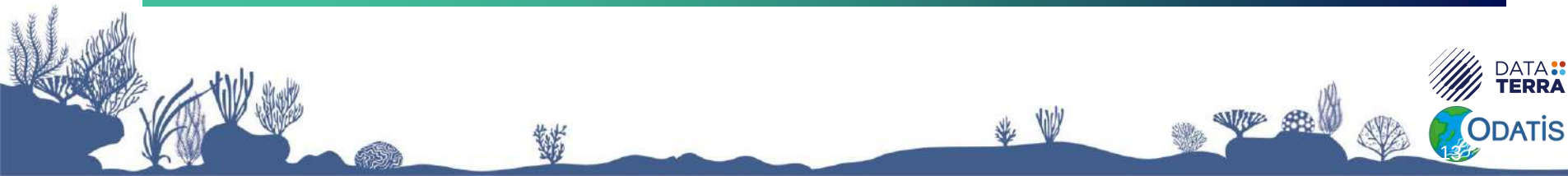
S'authentifier au Datalake

- Paire de clés `aws_access_key_id` et `aws_secret_access_key`
- Token `aws_session_token` (si on utilise un rôle)

Configuration des credentials

- Variables d'environnement `AWS_ACCESS_KEY_ID` `AWS_SECRET_ACCESS_KEY` `AWS_SESSION_TOKEN`
- Profils *en configurant les fichiers `~/.aws/credentials` et `~/.aws/config`*
- Directement dans le client (NON recommandé)

Les credentials permettent d'accéder et d'avoir la visibilité aux ressources d'un seul Account.
Les accès sont limités aux autorisations fournis via les politiques (IAM policy).
Par défaut, aucune autorisation (implicit deny).



Le Datalake du CNES : Usage du S3

Quel est le périmètre d'un utilisateur ?

Dans son périmètre



- Connaître les principes de la gestion des droits
- Pouvoir demander si son besoin le nécessite des clés permanentes (Read-only, Read/write)



AWS STS (alternate)

- Connaître les principes
- Utiliser les scripts présents dans /softs/tools/datalake/ ou librairie assumeroles



Amazon S3

- Choisir son client S3
- Paramétrer ses authentifications (paire de clé)
- Manipuler les objets
- **Tier 3** : Restaurer les objets et superviser le tier de l'objet

Hors périmètre (équipe Datalake)

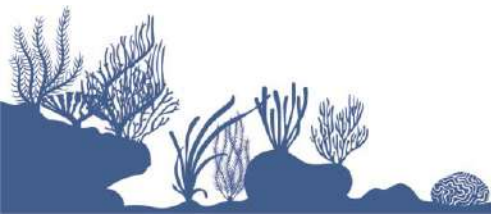
Création de IAM User, Rôle ou policy

Paramétrer les rôles

- en gestion d'accès (Policy)
- en gestion d'autorisation pour les comptes IPA

Création des buckets

Tier 3 : Paramétrer les règles ILM



Clients S3 et particularités du Datalake

Toujours utiliser un client S3 reconnu ou librairies de traitements

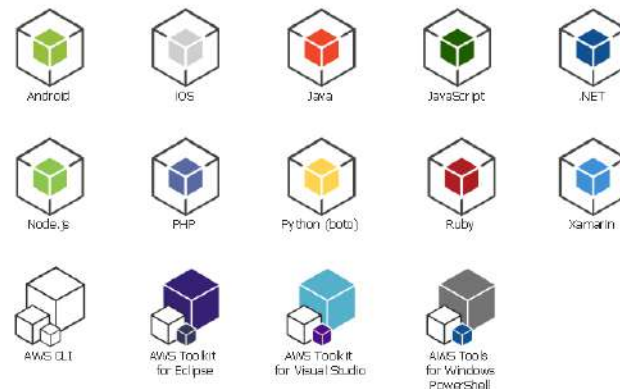
- **Utiliser un client S3 qui permet de manipuler les données**

- Des APIs haut-niveaux :
 - Simuler les commandes POSIX (ls, cp ...)
 - Uploads/downloads partitionnés
- Des APIs bas niveaux : get-object, put-object, head-object, list-object...

- **Prendre en compte les caractéristiques du stockage objet**

- très performant en lecture mais comparativement peu performant en écriture.
- organisation des données : metadonnées personnalisable, tags ...
- formats de données plus ou moins adaptés (en terme de performance) pour le S3

SDKs



Les opérations autorisées (GetObject, PutObject ...) dépendent des autorisations acquises par l'entité IAM (credentials) utilisée

