



ONYXIA : UNE PLATEFORME OPEN SOURCE ORIENTÉE DATASCIENCE

Cédric Couralet Alexis Guyot Olivier Levitt

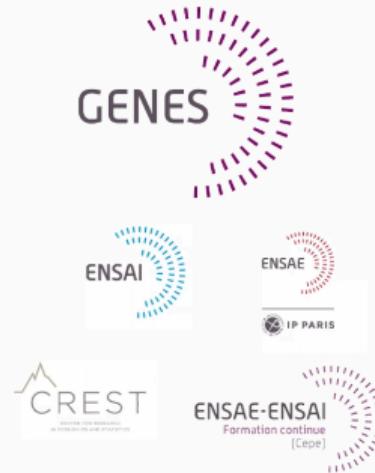
Modernisation de l'offre scientifique du GENES

La plateforme Onyxia

L'instance SSPCLOUD

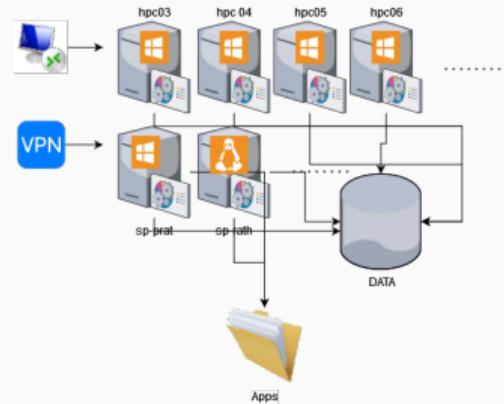
LE GENES

- Groupe des écoles nationales d'économie et statistique
- 2 écoles d'ingénieurs, un centre de formation continue, un centre de recherche (UMR)
- Tutelle Technique : Insee.
- Domaines : Statistique, économie, sociologie...
- **Fort penchant vers la science de la données**



OFFRE SCIENTIFIQUE HISTORIQUE

- Serveurs publics/privés
- Baie de stockage
- Logiciels déployés sur partage réseau



- 😡 Allocation/partage des ressources
- 😡 Gestion des packages (r, python...)
- 😡 Gestion des mise à jour (serveur/application)
- 😑 Performances
- 😑 ...« je veux faire du LLM »
- 😑 Reproductibilité / Science Ouverte

NOUVELLE SOLUTION

- Moins couteuse pour l'exploitation
- Proposant une meilleure allocation des ressources et leur partage
- Favorisant la reproductibilité
- Favorisant les bonnes pratiques de séparation code/data/traitement



 **kubernetes**

 **git**

 **S3**

Comment embarquer les utilisateurs?



kubernetes



git





Un cloud opensource pour la datascience (By Insee)

- Pas d'enfermement dans une solution
- Cloud Native
- 100% Open Source (MIT)
- Déploiement facile

ONYXIA, C'EST QUOI?

The screenshot shows the Onyxia - SSP Cloud DataLab interface. The left sidebar includes links for Réduire, Accueil, Mon compte, Paramètres du projet, Catalogue de services, Mes services (highlighted in red), Mes secrets, Mes fichiers, and Explorateur de Données. The main content area has a header 'Mes services' with a note about launching, visualizing, and managing services. It features sections for 'Services en cours' (with entries for 'Vscode-torch-fastText' and 'Miflow') and 'Enregistrés' (with entries for 'vscode-torch-fastText' and 'vscode-pytorch'). A bottom navigation bar includes links for Français, Conditions d'utilisation, v101.2, and a logo.

- Une application web permettant le déploiement de service sur cluster kubernetes
- Un catalogue de services spécialement construit pour l'intégration de services externes (Stockage Objet, Gestion de secrets, Git...)

DÉMO

Onyxia - SSP Cloud Datalab h4njg personal project

Tutoriels AIML4OS Déconnexion

Mes services

Lancer, visualiser et gérer rapidement vos différents services en cours d'exécution.
Il est recommandé de supprimer vos services après chaque session de travail.

Rafraîchir + Nouveau service Supprimer tous Events

Services en cours

```
$ helm list --namespace user-h4njg
```

Service	Démarré :	
Vscode-torch-fastText	hier	Ouvrir
Mlflow	il y a 2 jours	Ouvrir

Enregistrés

	Développer (2)
vscode-torch-fastText	Lancer :
vscode-pytorch	Lancer :

FONCTIONNEMENT : CONFIGURATION DES SERVICES

- Utilisation de helm pour le lancement des services
- Utilisation du values.schema.json pour l'affichage du formulaire

```
"accessKeyId": {  
    "description": "AWS Access Key",  
    "type": "string",  
    "x-form": {  
        "value": "{{ s3.AWS_ACCESS_KEY_ID }}"  
    },  
    "hidden": {  
        "value": false,  
        "path": "s3/enabled"  
    }  
}
```

FONCTIONNEMENT : MÉCANISME DE DÉCOUVERTE

Utilisation de la fonction lookup de Helm

```
 {{ range $index, $secret := (lookup "v1" "Secret" $namespace) }}  
 {{- if (index $secret "metadata" "annotations") }}  
 {{- if and (index $secret "metadata" "annotations" "onyxia/discovery")  
           ("annotations" "onyxia/discovery" | toString)) }}
```

AUTRES FONCTIONNALITÉS

- Gestion de projet/groupe
- Restriction sur le lancement de service

roles :

```
- roleName: vip
  files:
    - relativePath: nodeSelector-gpu.json
      content: |
        {
          "$schema": "http://json-schema.org/draft-07/schema#",
          "title": "Node Selector",
          "type": "object",
          "properties": {
            "nvidia.com/gpu.product": {
              "description": "The type of GPU",
              "type": "string"
            }
          }
        }
```

L'INSTANCE SSPCLOUD

- Instance d'onyxia maintenue par l'Insee
- Expérimentation sur données non sensible, partage de connaissances (formations)
- Ouverte d'abord au data scientist du service statistique public, étendue à l'ensemble des agents de l'état et aux étudiants