

SECRÉTARIAT GÉNÉRAL

OFFRE DE SERVICE

Qu'est-ce que le Datalab Onyxia du Groupe ?

Département des Systèmes d'Information et de Télécommunication



DESCRIPTION DE L'OFFRE

Cette offre présente le Datalab Onyxia du Groupe ENSAE-ENSAI, une application web permettant d'accéder à des environnements de travail dédiés à la data, sans avoir à gérer le déploiement de l'infrastructure technique.

Le Datalab met à disposition plusieurs services liés au traitement, à l'analyse et à la science des données, tels que Jupyter, RStudio, MongoDB, Kafka, MLflow, ainsi que d'autres outils utiles aux travaux de data science.

Au-delà de la mise à disposition d'outils modernes, le Datalab favorise les bonnes pratiques en matière de traitement des données et de programmation. Il encourage notamment la reproductibilité des résultats grâce à des environnements facilement configurables, enregistrables, restaurables et partageables.

Pour les élèves, chercheurs, enseignants et collaborateurs, le Datalab constitue également un espace d'expérimentation, d'autoformation et de travail collaboratif, accessible via une interface intuitive.

L'utilisation du Datalab est soumise à des conditions générales d'utilisation précisant notamment les catégories de données pouvant être traitées, les limites de ressources par utilisateur et les responsabilités associées.

<https://documentation-dsit.lab.groupe-genes.fr/Services/Datalab/Onyxia/terms/>



POINT DE CONTACT

Alexis GUYOT

SG > DSIT

Email : alexis.guyot@ensae-ensai.fr

Téléphone : 01.70.26.67.63



PUBLIC CIBLE

Cette offre s'adresse aux utilisateurs du Groupe ENSAE-ENSAI ayant besoin d'accéder à des environnements de travail data modernes, reproductibles et collaboratifs.

Elle concerne notamment :

- les élèves ;
- les chercheurs ;
- les enseignants ;
- les collaborateurs travaillant sur des projets data ;
- les équipes projet ayant besoin d'environnements techniques partagés ;
- les utilisateurs souhaitant expérimenter, apprendre ou produire des analyses avec des outils de data science.



ACCÈS À L'OFFRE

Procédure

i. Accès au Datalab Onyxia

Le Datalab est accessible depuis l'adresse suivante :

<https://onyxia.lab.groupe-genes.fr/>

Il permet de lancer des services de data science en libre-service, sans avoir à déployer soi-même l'infrastructure.

ii. Choix d'un service dans le catalogue

L'utilisateur peut accéder au catalogue des services et choisir l'environnement dont il a besoin, par exemple Jupyter, RStudio ou d'autres services liés à la data :

<https://onyxia.lab.groupe-genes.fr/catalog/ide>

iii. Configuration des ressources

Lors du lancement d'un service, l'utilisateur peut définir les ressources nécessaires, notamment la RAM, le CPU et, lorsque disponible, le GPU.

Les limites indiquées par utilisateur sont les suivantes :

- 10 services ;
- 20 CPU ;
- 50 Go de RAM ;
- 1 GPU ;
- 200Go de stockage.

Ces limitations peuvent être différentes dans le cadre de groupes de projets.

iv. Utilisation du répertoire Git et du stockage S3

Chaque utilisateur dispose d'un répertoire Git et d'un espace de stockage de type S3 automatiquement connectés aux services créés.

Git permet de synchroniser un projet local avec un serveur distant, de limiter le risque de perte de code et de conserver un historique complet des modifications.

Le stockage S3 permet de stocker des fichiers accessibles depuis différents services et de les partager plus facilement via une URL.

v. Partage et collaboration

Le Datalab permet de partager l'accès à des services et à des ressources avec d'autres collaborateurs, notamment dans le cadre de projets communs.

vi. Scripts d'initialisation et reproductibilité

L'utilisateur peut spécifier un script d'initialisation personnalisé, exécuté au lancement du service.

Il est également possible d'enregistrer, restaurer et partager la configuration des services avec d'autres collaborateurs, afin de faciliter la reproductibilité des environnements de travail.

vii. Gestion des secrets

Pour les informations sensibles complémentaires (clés d'API, identifiants tiers, etc.), le Datalab met à disposition un gestionnaire de secrets basé sur Vault. Les secrets sont stockés sous forme de paires clé-valeur, chiffrées, et organisables en dossiers depuis l'interface "Mes secrets". Ils sont ensuite injectés dans les services sous forme de variables d'environnement via l'onglet Vault de la configuration du service.

viii. Demande d'ajout de nouveaux services

Si un service nécessaire n'est pas disponible dans le catalogue, une demande peut être adressée à la DSIT du Groupe ENSAE-ENSAI via le canal Teams Datalab - GENES - Teams ou par email à l'adresse :

support.informatique@ensae.fr



ENGAGEMENTS

Le Département des Systèmes d'Information et de Télécommunication met à disposition un environnement Datalab permettant aux utilisateurs de créer, configurer et partager des environnements de travail data en libre-service.

L'offre permet notamment :

- l'accès à une application web dédiée aux usages data ;
- la mise à disposition de services comme Jupyter, RStudio, MongoDB, Kafka ou MLflow ;
- l'accès à un catalogue de services déployables en libre-service ;
- la possibilité de configurer les ressources nécessaires aux services ;
- la mise à disposition d'un répertoire Git ;
- la mise à disposition d'un espace de stockage S3 ;
- la connexion automatique de Git et du stockage S3 aux services créés ;
- la possibilité de partager des services et ressources avec des collaborateurs ;
- la possibilité d'enregistrer, restaurer et partager des configurations ;
- l'intégration de secrets via Vault ;
- la suppression automatique des services inactifs depuis plus de 15 jours, volumes de données associés inclus, afin de garantir une utilisation optimale des ressources mutualisées ; il est recommandé d'utiliser l'espace de stockage S3 et les dépôts de code pour éviter toute perte de données ;
- la mise à disposition d'un explorateur de données permettant de prévisualiser des fichiers aux formats Parquet et CSV directement depuis le navigateur, sans recours à un service de data science ;
- l'accompagnement des utilisateurs pour les demandes d'ajout de nouveaux services au catalogue.



PÉRIMÈTRE DE L'OFFRE

L'offre couvre l'accès et l'utilisation du Datalab Onyxia du Groupe ENSAE-ENSAI, notamment :

- l'accès à la plateforme Onyxia ;
- l'accès au catalogue de services ;
- le lancement de services data en libre-service ;
- la configuration des ressources RAM, CPU et GPU ;
- l'utilisation de Git pour la gestion et l'historisation du code ;
- l'utilisation du stockage S3 pour les fichiers et données ;
- le partage de services et de ressources avec d'autres utilisateurs ;
- l'utilisation de scripts d'initialisation ;
- l'enregistrement, la restauration et le partage de configurations ;
- l'utilisation de secrets sous forme de variables d'environnement ;
- la prévisualisation de fichiers de données (Parquet, CSV) via l'explorateur de données intégré ;
- les demandes d'ajout de nouveaux services au catalogue ;
- la gestion du cycle de vie des services, incluant la suppression automatique des services inactifs au-delà de 15 jours.

Le Datalab vise à faciliter le travail sur des environnements stables, reproductibles et collaboratifs, adaptés aux besoins d'enseignement, de recherche, d'expérimentation et de projets data.