

Inria

Plateforme « données sensibles » Inria

Antoine Fraboulet
Support Utilisateurs

26 septembre 2023

Objectif : doter Inria d'un environnement de recherche sur les données sensibles

Données sensibles :

- Données de **santé** relevant de la certification HDS
 - Partenariats AP-HP et autres CHU, traitement des données du SNDS
- Données **personnelles sensibles** hors santé
- Données de **partenaires industriels**
- Données d'équipes d'**autres domaines** (cybersécurité)

Exclu du périmètre à ce stade : données Diffusion Restreinte (II 901)

- Partenariats avec la sphère Sécurité/Défense ; avec des acteurs du nucléaire civil

Objectif : doter Inria d'un environnement de recherche sur les données sensibles (2/2)

Offre de service visée :

- **Traitement des données sensibles sur CPU et GPU**
 - Possibilité d'avoir des environnements de développement riches
 - Applications développées par les équipes de recherche
- Procédures d'import/export des données en cohérence avec leur sensibilité
- Stockage des données sur la durée du projet

- **La plateforme n'est pas un entrepôt de données :**
 - Données stockées de manière isolée pour chaque projet, et uniquement pendant la durée nécessaire au projet (destruction en fin de projet)

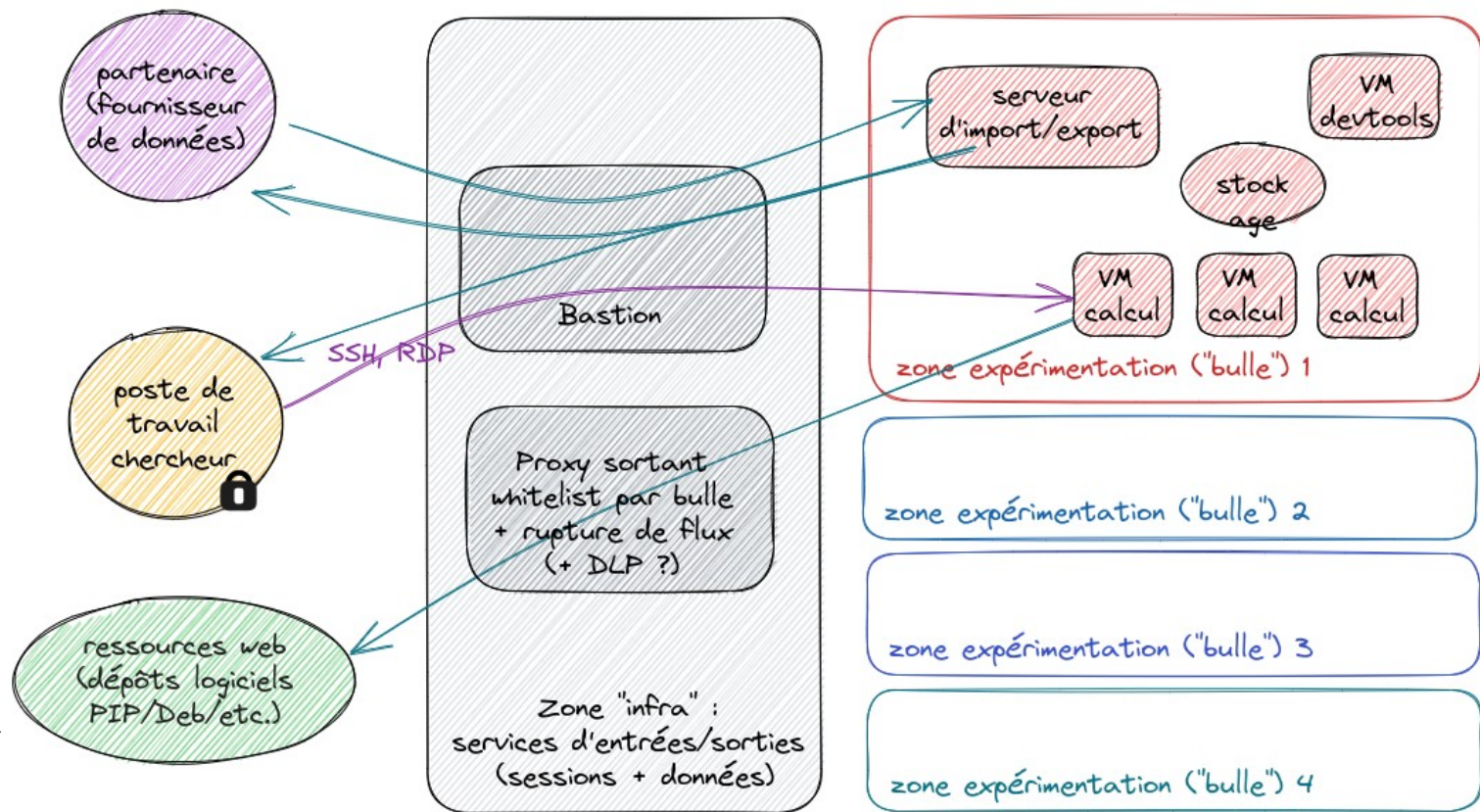
Intérêts : **sécuriser** et **faciliter** les travaux nécessitant la manipulation de données sensibles

- Passer de solutions ad-hoc à une solution commune et bien maîtrisée, présentant les garanties de sécurité nécessaires
- Faciliter le processus d'instruction des projets de recherche
- Permettre aux chercheurs de travailler dans un environnement compatible avec les besoins de la recherche en sciences du numérique

Solution technique prévue

- Plateforme séparée de l'infrastructure Inria
 - Hébergée par un fournisseur de cloud souverain, certifié HDS, et qualifié SecNumCloud
 - Opérée par un prestataire certifié HDS
- Mécanisme de « **bulles sécurisées** » (une bulle par projet, possibilité de plusieurs bulles cohabitant sur la plateforme)
 - Contrôle des flux (sessions, données) en entrée et en sortie des bulles
 - Stockage des données à l'intérieur de chaque bulle
 - Isolation forte entre les bulles

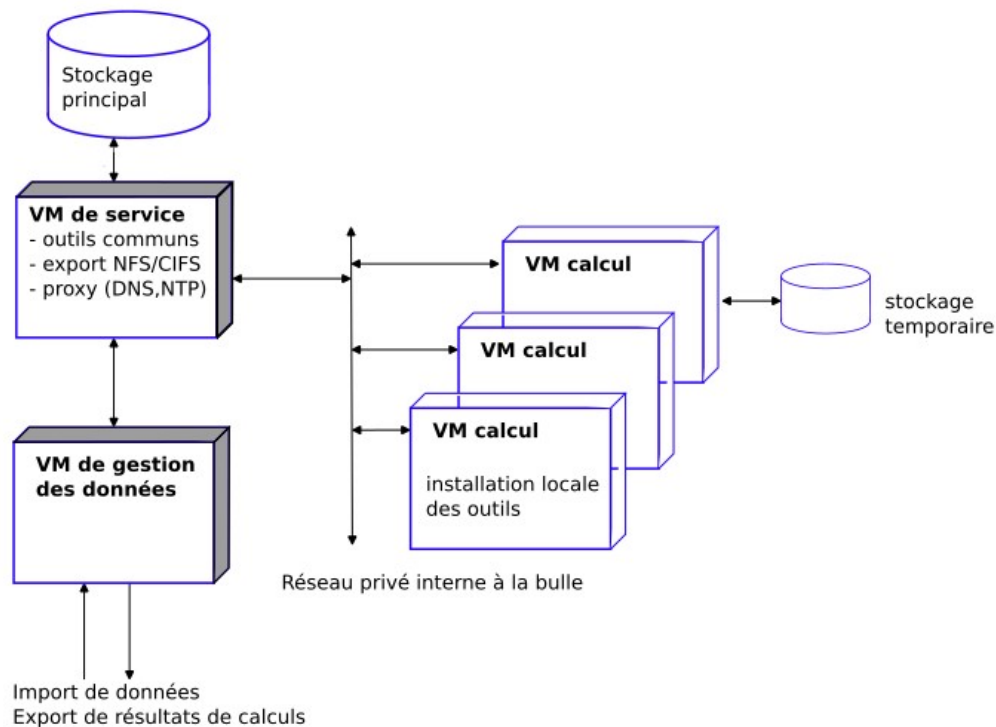
Solution technique prévue (vue fonctionnelle)



Déroulement du projet

- 2020 : Premières réflexions (contexte : crise sanitaire → nombreux projets en lien avec la santé)
- Fin 2020 : réponse à un AAP dans le cadre de France 2030 opéré par l'ANSSI : action « Renforcement du niveau de sécurité de l'Etat » de la stratégie d'accélération Cyber
- Mars 2021 : projet sélectionné par l'ANSSI
- 2021 : Analyse de risques EBIOS-RM
- Mars 2022 : phase 1 validée par l'ANSSI après analyse des documents fournis. Financement débloqué
- 2022 - début 2023 : sélection des prestataires hébergeurs et infogérants, définition des prestations
- En cours / à venir
 - Juillet / août / septembre : définition de l'architecture technique (cible : DAT validé fin septembre)
 - Octobre / novembre : construction de la plateforme
 - **Fin novembre : ouverture progressive aux premiers projets utilisateurs (beta test)**
 - Dès la plateforme stabilisée : nouvelle analyse de risques → nouvelle homologation de sécurité

Bulles de calcul associées aux projets



Bulles de calcul et services proposés (1/2)

- Les VM de **calcul** sontinstanciées à la création du projet
 - Large gamme de VM possibles (CPU, RAM, GPU, stockage local)
 - Elles peuvent être démarrées et arrêtées à la demande
 - Fort impact sur la facturation, notamment sur l'utilisation de GPU
- Les membres du projet peuvent installer des outils externes
 - Python (PIP, Conda), C++, Matlab, R (CRAN)
 - Gestion de groupes d'utilisateurs (administrateurs, utilisateurs)
- Les accès sont réalisés soit par ssh, soit avec déport d'affichage RDP
 - rupture de flux dans tous les cas
 - enregistrement des sessions

Bulles de calcul et services proposés (2/2)

- Les VM gérant le **stockage** et les **imports/exports** ne sont pas autogérées par les membres des projets
 - Les utilisateurs n'ont pas les droits administrateurs sur ces VM
 - Une application de déclaration des imports/exports de données avec validation sera mise à disposition pour la création de canaux spécifiques
- Outils communs mis à disposition
 - Serveur Git pour le développement local (éventuellement une forge?)
 - Partage d'espaces de travail (stockage) entre les VM
 - ...

Support aux utilisateurs

- Recueil des besoins utilisateurs
 - Les projets devront être accompagnés lors de la phase de création des bulles
 - Dimensionnement des bulles avant création
 - Adaptation éventuelle des images pour l'hébergement de projets spécifiques
- Formation et accompagnement des utilisateurs pour la prise en main
 - Mise à disposition du bulle utilisable pour les formations
- Les interventions dans les bulles nécessiteront une accréditations
 - Les personnes du support ne pourront intervenir dans une bulle que pour des actions ponctuelles

Travaux en cours côté support utilisateurs

- Recueil des besoins utilisateurs
 - Un sondage circule auprès des équipes, principalement celles portant des projets en santé numérique dans un premier temps
- Mise en place des premiers environnements compatibles PDS
 - Images dérivées de celles utilisées dans les Moyens de Calculs
 - Continuité entre développement d'applications côté MC et traitement sécurisé dans les bulles

Appel à participation !

- Toutes les participations sont possibles
 - Recueil des besoins des projets
 - Point de contact technique dans les centres
 - Mise en place des environnements (images)
 - Formation, accompagnement des utilisateurs

- Contacts :
 - Lucas Nussbaum : chef de projet PDS
 - Antoine Fraboulet : support utilisateurs

Merci

Inria