HEALTH DATA HUB

Besoins en services cloud de la plateforme technologique

Juin 2020

Le Health Data Hub a pour mission de permettre l'exploitation des données de santé françaises

- La France dispose d'une **richesse unique au monde : un patrimoine de données de santé** décrivant l'intégralité de sa population
- Ces données, constituées grâce au financement public de la protection sociale, **rassemble des données à la fois administratives et cliniques**

- Néanmoins, **ces données sont actuellement sous-exploitées**, les chercheurs français ont du **mal à y accéder et n'ont pas les moyens techniques** pour mener à bien leurs travaux de recherche
- La **primeur de ces innovations**, que la France pourrait porter, est de plus en plus **captée par des puissances étrangères** comme ce fut le cas pour les premiers algorithmes en imagerie médicale

- Comme annoncé par le Président de la République, **il devient impératif de mettre en place une plateforme sécurisée** pour **faciliter l'accès et le traitement des données de santé** par les porteurs de projets d'intérêt public habilités
- Le **Health Data Hub ambitionne de répondre à ce risque de perte de chance** des citoyens français et de **promouvoir l'innovation et la recherche** dans le domaine de la santé

Le choix des services sur la plateforme technologique du Health Data Hub est porté par trois enjeux

Les enjeux de la plateforme technologique du Health Data Hub sont :

LA SÉCURITÉ

Assurer la protection des données de santé hautement sensibles stockées sur la plateforme technologique

LES FONCTIONNALITÉS & PERFORMANCE

Répondre aux besoins des porteurs de projets d'intérêt public, notamment des projets innovants nécessitant des moyens de calcul de science des données

LES COÛTS & DÉLAIS

Mettre en œuvre rapidement une solution pour répondre au risque de perte de chance des citoyens français en lien avec la stratégie intelligence artificielle nationale

Outre la couverture du besoin fonctionnel, les services privilégiés pour répondre à ces enjeux sont :

- **Les services intégrés et managés contribuant à :**
  - La gestion de la sécurité au niveau de chaque élément de la plateforme technologique (e.g., gestion de l'annuaire, gestion des logs)
  - La robustesse de la plateforme technologique et par conséquent un meilleur niveau de service
  - La réduction de la charge d'intégration et la réduction du temps de maintien en condition opérationnelle pour livrer dans les délais
- **Les services dans un environnement certifié HDS[1] contribuant à :**
  - L'hébergement sécurisé des données
- **Les services éprouvés contribuant à :**
  - La diminution des risques liés à la sécurité
  - La diminution des risques liés à l'indisponibilité des services de la plateforme technologique

De manière transverse, l'**automatisation du déploiement** des ressources et la **gestion fine des droits et des flux** sont deux leviers actionnés pour répondre au besoin de mise à disposition d'environnements de travail, sécurisés et isolés, à la volée.

[1]HDS : Hébergement de données de santé

Cinq grands principes technique de sécurité sont mis en oeuvre sur la plateforme technologique du HDH

ACCÈS

- **Authentification double-facteur** : toute connexion à la plateforme technologique nécessite un mot de passe à usage temporaire en plus des identifiants (nom de compte et mot de passe) de l'utilisateur.
- **Rupture protocolaire** : cette connexion donne accès à la plateforme via un bureau virtuel, entièrement maîtrisé et hébergé sur la plateforme
- **Sécurisation de la connexion** : les échanges entre les postes utilisateur et la plateforme sont chiffrés (HTTPS)
- **Contrôles d'intégrité** : vérification de pré-requis sur les postes sources accédant à la plateforme

PSEUDONYMISATION

- **Pseudonymisation des données** : La plateforme technologique est uniquement faite pour héberger des données de santé au moins pseudonymisées, ainsi, elle ne contiendra aucune information directement identifiante
- **Anonymisation des exports** : Conformément au référentiel de sécurité du SNDS, seules des données anonymisées ont le droit d'être exportée de la plateforme technologique, selon les modalités décrites dans les autorisations CNIL obtenues

CHIFFREMENT

- **Chiffrement des stockages** : tout espace de stockage de la plateforme technologique est chiffré (AES 256 bits)
- **Chiffrement et analyse des flux** : tout flux de données est chiffré (selon la nature des flux : TLS, HTTPS, VPN, etc.), ces flux sont analysés par les pare-feux à l'aide d'un processus de désencapsulation – encapsulation
- **Chiffrement des fichiers** : Un chiffrement au niveau des fichiers (AES 256 bits) est réalisé dans les espaces de stockage contenant une grande quantité et une grande diversité de données (espace opérateur)

GESTION DES CLÉS

- **Maîtrise de la clé maître** : La clé maître est générée dans un HSM dédié, sous contrôle du HDH et est transmise au HSM de la plateforme technologique selon une cérémonie des clés
- **Protection des clés maître** : les clés de chiffrement sont stockées dans des boîtes noires transactionnelles (« HSM »)

SEGMENTATION

- **Segmentation réseau** : La plateforme technologique sera aussi segmentée dans plusieurs sous-réseaux dont le filtrage est assuré par des pares-feux (périmétrique, zone d'administration, zone interne)
- **Segmentation des droits d'opération** : Les droits et accès des opérateurs sont nativement segmentés à l'aide d'un bastion
- **Segmentation des accès** : Les accès opérateurs et utilisateurs sont différents en termes d'interface, de réseau, de base d'authentification et d'outils
- **Segmentation des environnements:** tenants et utilisateurs différents sur les différents environnements

Le Health Data Hub a sélectionné une dizaine de services parmi une cinquantaine

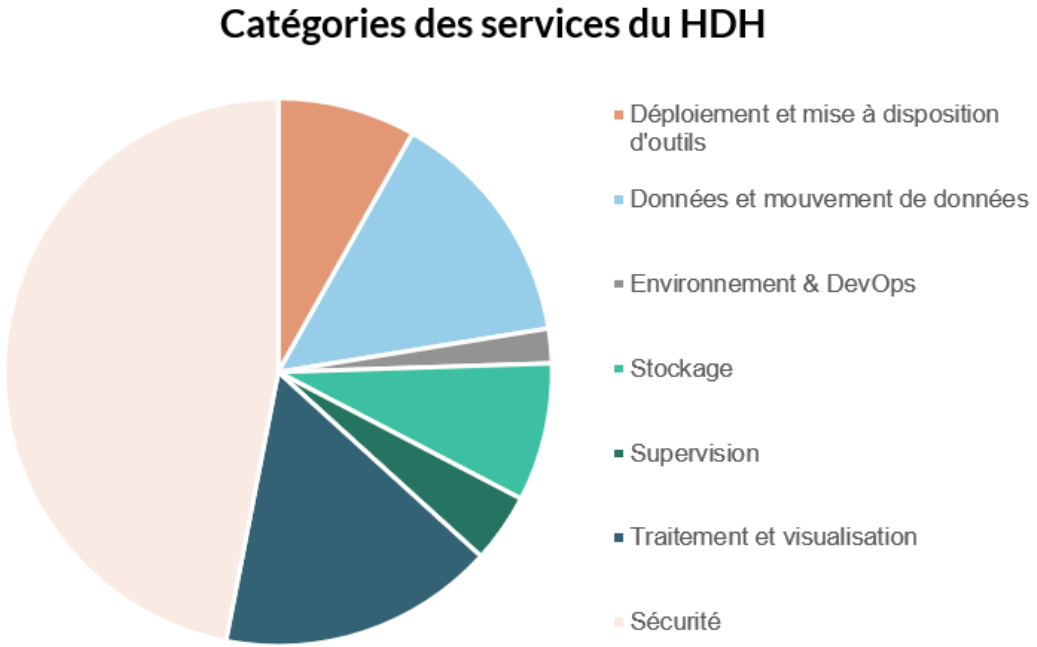
- La plateforme technologique du Health Data Hub représente **une cinquantaine de services**, tous fournisseurs confondus, dont **près de la moitié sont utilisés pour gérer la sécurité**.

- Parmi ces services :

- **Plus de la moitié représentent des services cloud jugés indispensables**[1] dans l'offre d'un fournisseur cloud par les équipes techniques du Health Data Hub, du fait de la complexité d'intégration, et par conséquent du risque de sécurité, que cela engendrerait.

- **Une dizaine sont jugés secondaires**[1] dans l'offre d'un fournisseur cloud du fait de l'existence de solutions externes matures, ne nécessitant pas d'importants efforts d'intégration.

- **Une dizaine de services est déjà externalisée**.



Les services jugés indispensables représentent le strict minimum que le Health Data Hub pourrait attendre d'un fournisseur cloud. Certains services jugés secondaires nécessiteraient pourtant un développement important de la part des équipes du Health Data Hub (e.g., intégration d'un SIEM, d'une interface de supervision ou d'un ETL).

Quatorze services jugés indispensables dans une offre cloud ont été retenus pour base de cette présentation

[1]Avertissement : ce travail a été réalisé sans évaluation des coûts et de la charge réelle d'intégration que représenterait le développement de la plateforme chez un prestataire d'hébergement autre que celui actuellement retenu

Le HDH présente une liste non exhaustive de quatorze services cloud jugés indispensables (1/3)

	BESOIN	DESCRIPTION DU BESOIN	VALEUR ATTENDUE
SERVICES "IAAS"	Virtualisation de machines	**Virtualisation de machines** présentant les **capacités de calcul élastiques** nécessaires à la plateforme et aux projets	▪ Capacité de calcul élastique pour des projets en data science (GPU) ▪ Capacité de résilience au sein d'une région donnée
SERVICES "PAAS"	Accès aux services cloud	**Pilotage et configuration des services** de l'environnement cloud dans une interface graphique	▪ Centralisation des outils de monitoring et de configuration au travers d'une interface unique pour assurer le suivi et l'administration de la plateforme technologique ▪ Facilité d'opération ▪ Politique de sécurité d'accès aux outils d'administration unique et centralisée pour faciliter la protection des informations sensibles y figurant
	Sécurité des accès	**Restriction des accès aux services PaaS** selon des caractéristiques telles que l'adresse IP, la région, les identités, les appareils, etc.	▪ Politique de sécurité d'accès à la console d'administration unique et centralisée pour faciliter la protection des informations sensibles y figurant ▪ Configuration des pare-feu par le fournisseur de services cloud, le HDH ne pouvant implémenter lui-même cette fonctionnalité ▪ En combinaison avec le service d'attribution temporaire des autorisations d'accès, politique de management des interfaces complète
	Gestion des identités et autorisations	**Identification des services, des machines et des utilisateurs et gestion de leurs accès aux éléments d'infrastructure et fonctionnels** de la plateforme technologique	▪ Intégration de l'annuaire avec les différents services pour contrôler les droits d'accès ▪ Gestion des identités de manière sécurisée grâce à une politique de sécurité unique et centralisée ▪ Authentification multi-facteur

Le HDH présente une liste non exhaustive de quatorze services cloud jugés indispensables (2/3)

SERVICES "PAAS"

BESOIN	DESCRIPTION DU BESOIN	VALEUR ATTENDUE
Puits de trace et outils d'analyse	**Centralisation, exploration et analyse des traces** des composants de la plateforme technologique	▪ Génération de logs au niveau de tous les composants de la plateforme technologique ▪ Intégration d'une solution de centralisation des logs avec tous les autres services ▪ Capacité de requêtage standard des logs
Gestion d'événements	**Déclenchement des événements** sur la plateforme technologique et **transports des messages** liés à ces événements	▪ Visibilité au travers d'un bus d'événements du changement d'état des éléments de la plateforme (e.g., cycle de vie des VM, cycle de vie des fichiers) ▪ Evénements recevables et écoutables par des outils pouvant déclencher les actions nécessaires ▪ Écoute des événements bas niveaux de l'infrastructure sans avoir à développer cette fonctionnalité ▪ Récupération des messages de l'ensemble des composants de la plateforme technologique
Réseaux	**Filtrage des flux intégré dans les réseaux et sous réseaux managés**	▪ Intégration d'un filtrage dans la couche réseau de l'offre du fournisseur cloud, le HDH ne pouvant l'implémenter lui-même ▪ Accès privilégié juste-à-temps (JIT), donc limité dans le temps
Gestion des clés de chiffrement	**Protection de l'intégrité et de la confidentialité des clés de chiffrement** au travers d'une boîte noire transactionnelle (HSM)	▪ Respect des standards de sécurité à travers l'utilisation d'un HSM pour protéger les clés de chiffrement ▪ Lien entre les secrets générés par le HDH et les services managés (BYOK)
Configuration	**Mise en place et suivi des critères d'acceptance de sécurité et de santé** pour le maintien en condition de sécurité de la plateforme technologique	▪ Solution permettant d'offrir une visibilité sur l'état de sécurité et de santé de chaque élément de la plateforme ▪ Contrôle de la bonne implémentation des politiques de gouvernance du HDH au niveau technique, concernant, par exemple, la localisation, le nombre et l'état des ressources, les éléments de configuration, etc. ▪ Capacité à créer des alertes en cas de non-conformité aux politiques de gouvernance du HDH (e.g., une zone de stockage non protégée par une clé HDH, une machine non à jour, etc.)

Le HDH présente une liste non exhaustive de quatorze services cloud jugés indispensables (3/3)

	BESOIN	DESCRIPTION DU BESOIN	VALEUR ATTENDUE
SERVICES "PAAS"	Stockage objet	**Structuration des espaces de stockage en mode objet**	▪ Sécurisation des accès utilisateurs et réseau ▪ Capacité de stockage élastique
	Traitement distribué	**Capacité de lancer des traitements distribués**	▪ Intégration en terme d'authentification ▪ Capacité à provisionner un cluster à la demande
	Base relationnelle	**Exposition des données selon le modèle relationnel**	▪ Service de gestion de base de données relationnelle managé
SERVICES "TRANSVERSES"	Automatisation	**Automatisation du déploiement des ressources**	▪ Pilotage de la création des ressources selon une logique d'infrastructure as code (IaC) ▪ Mise à disposition des API appropriées et intégrées à des outils tiers de gestion d'infrastructure (Terraform, ansible, chef, etc.)
	Authentification	**Attribution temporaire d'autorisations d'accès**	▪ Complément au service de restriction des accès aux services ▪ Accès privilégié juste-à-temps (JIT), donc limité dans le temps ▪ Approbation pour l'activation des rôles privilégiés ▪ Justification pour comprendre le motif d'activation (e.g., pour la demande d'export de résultats anonymisés) ▪ Historique de l'élévation de privilège

Merci