

HEALTH DATA HUB

Présentation de l'étude de réversibilité

7 novembre 2019

DOCUMENT ADMINISTRATIF COMMUNIQUE PAR LE HEALTH DATA
HUB – PUBLICATION EN LIGNE (article L.312-1-1 du CRPA)



- **Contexte**
- **Etude de réversibilité**
- **Capitalisation sur les travaux menés**



- **Contexte**
- **Etude de réversibilité**
- **Capitalisation sur les travaux menés**

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 nécessaire de **mettre en place d'une plateforme** pour **faciliter l'accès et le traitement des données de santé** par les personnes habilitées

Les enjeux de cette plateforme sont :



SÉCURITÉ

- Solution agréée ou certifiée « Hébergeur de données de santé »
- Arrêté relatif au référentiel de sécurité du Système National des Données de Santé



FONCTIONNALITÉ & PERFORMANCE

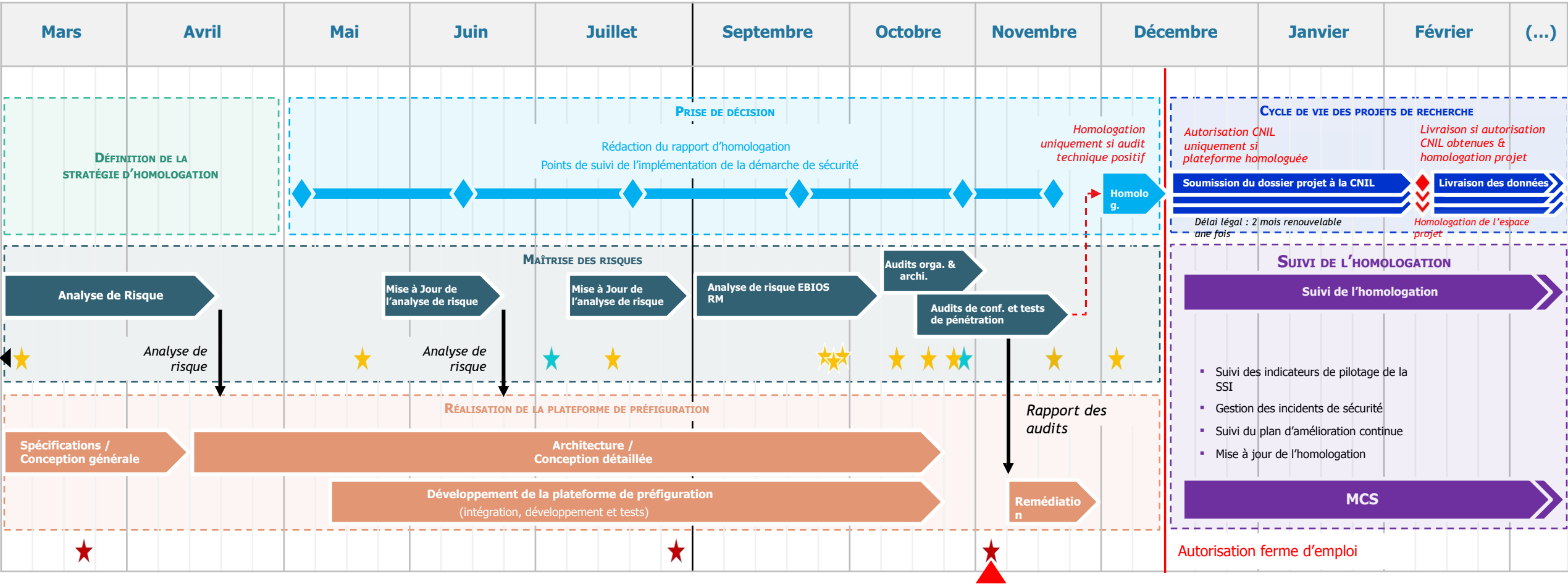
- Service mature, se positionnant comme une référence internationale sur le marché
- Offre couvrant les besoins fonctionnels du Health Data Hub



COÛTS & DÉLAIS

- Existence d'un véhicule contractuel directement accessible dans l'attente de la stratégie Cloud française
- Solution permettant la mise en place rapide de la plateforme de préfiguration

La mise en place du Health Data Hub suit une feuille de route technique



- Légende :**
- Atelier réalisé avec la CNIL
 - Atelier réalisé avec l'ANSSI
 - Atelier réalisé avec la DINSIC

Une étude de l'écosystème technique a été conduite pour identifier la solution d'hébergement

	SÉCURITÉ	FONCTIONNALITÉ & PERFORMANCE	Coûts & Délais
AZURE		Solution sélectionnée	
OVH		Pas de GPU dans l'environnement certifié HDS	Pas de véhicule contractuel existant
SANTÉOS			Coûts de réalisation chiffrés à 2,5 M€ Durée estimée entre 5 et 8 mois, pas de véhicule contractuel
AWS	Pas agréé / certifié HDS		Pas de véhicule contractuel
GCP	Pas agréé / certifié HDS		Pas de véhicule contractuel
DOCAPOSTE		Pas de CPU / GPU dimensionné pour la science des données	-
PADAWAN	Pas agréé / certifié HDS	Dimensionnement ne permettant pas de répondre aux besoins du HDH	Complexité de contractualisation élevée (institut de recherche)
TERALAB	Pas agréé / certifié HDS	Dimensionnement ne permettant pas de répondre aux besoins du HDH	-
CASD	Pas agréé / certifié HDS	Pas de GPU, peu de latitude pour ajouter des fonctionnalités	Tarif par utilisateur élevé
OBS		Pas de GPU dans l'offre souveraine	Offre IaaS, délai d'intégration important Pas de véhicule contractuel existant
OUTSCALE	Pas agréé / certifié HDS	Pas de GPU	Offre IaaS, délai d'intégration important

Etude terminée en février 2019, lors du choix de la solution d'hébergement

La DREES reste consciente des enjeux de réversibilité et de sécurité de la plateforme

LA PLATEFORME DE PRÉFIGURATION A ÉTÉ PENSÉE POUR ÊTRE LA PLUS RÉVERSIBLE POSSIBLE



- Les **données** stockées pourront être facilement transférées : les coûts sont maîtrisés et aucun format propriétaire n'est utilisé (coût d'export de données : ~0,07€ / Go soit ~7 k€ pour 100 To)
- Les **algorithmes** seront développés avec des technologies de référence, indépendantes de la solution d'hébergement
- Les **éléments de la plateforme** seront en partie réutilisables (architecture, scripts de déploiement, etc.)

ELLE A ÉTÉ CONÇUE POUR OFFRIR UN CONTEXTE DE SÉCURITÉ ADAPTÉ À L'HÉBERGEMENT DE DONNÉES DE SANTÉ



- La plateforme n'hébergera que des **données de santé pseudonymisées**
- Elle repose sur un environnement **certifié « hébergeur de données de santé »**, selon les normes proposées par l'ASIP-Santé
- Les **briques de sécurité active sont externalisées** et reposent sur des technologies **certifiées** par l'ANSSI ou Critères Communs pour répondre aux exigences des textes applicables et des bonnes pratiques

Cette plateforme repose sur l'utilisation de services « PaaS », simplifiant fortement son intégration

SERVICES
"IAAS"

Cat.	Brique	Description
Infrastructure	Stockage	Espaces pour stocker les données et éléments de la plateforme
	Sauvegarde (stockage)	Espaces de stockage économique sans haute disponibilité
	Calcul	Puissance de calcul nécessaire à la plateforme et aux travaux de recherche
	Réseaux	Capacités de transfert de données
Général	Gestion des identité	Identification et gestion des accès des éléments de la plateforme
	Supervision	Maintien en conditions opérationnelles des ressources, suivi des coûts
Sécurité	Infrastructure de bureaux virtuels	Mise à disposition de bureaux virtuels maîtrisés pour les utilisateurs
	Pare-feu	Filtrage des flux réseaux
	Anti-virus	Protection contre les éléments malveillants
	Puits de traces et outil d'analyse	Centralisation, exploration et analyse des traces
	Gestion de l'info. de sécu. (SIEM)	Analyse avancée des traces
	Scellement des traces	Conservation de l'intégrité des traces
	Gestion des services info.	Gestion des tickets utilisateurs
	Gestion des clés de chiffrement	Protection de l'intégrité et de la confidentialité des clés de chiffrement
	Réseau virtuel privé (VPN)	Mise en place de connexions sécurisées

SERVICES
"PAAS"

SERVICES
"PAAS"

Cat.	Brique	Description
Données	Stockage objet	Structuration des espaces de stockage en mode objet
	Stockage fichier	Structuration des espaces de stockage en mode fichier
	Import plateforme	Ingestion des données venant de l'extérieur de la plateforme
	ETL	Déplacement et transformation des données (chiffrement, etc.)
	Evénements	Gestion des événements générés par les éléments de la plateforme
	Workflow	Séquençage et orchestration d'actions
	Chiffrement données	Protection de la confidentialité des données
Projet	Gouvernance	Organisation globale des données présentes sur la plateforme
	Travail collaboratif	Gestion de projet et automatisation de l'intégration de codes
	Traitement non distribué	Langage de programmation pour lancer des traitements non distribués
	Traitement distribué	Langage de programmation pour lancer des traitements distribués
	Base relationnelle	Structuration des données selon le modèle relationnel
	Base NoSQL	Structuration des données selon les modèles semi ou non relationnels
	Visualisation	Visualisation graphique des données ou résultats produits
	Import / export projet	Import ou export de données mis à disposition aux projets



- **Contexte**
- **Etude de réversibilité**
- **Capitalisation sur les travaux menés**

Une étude de réversibilité a été lancée en parallèle de la mission

1

Etude comparative des solutions d'hébergement



- **Définition des critères de comparaison** des solutions d'hébergement
- **Evaluation de chaque solution** selon les critères définis
- **Sélection d'une solution d'hébergement** la plus adaptée à l'étude de réversibilité

2

Etude de réversibilité



- **Définition du scénario d'étude** selon la solution d'hébergement sélectionnée
- **Chiffrage du scénario** défini
- **Comparaison des coûts** avec le scénario de la plateforme de préfiguration
- **Evaluation** des économies réalisées.

3

Mise en œuvre de la réversibilité

- **Rédaction du plan de réversibilité**
- **Mise en œuvre du plan de réversibilité** en adéquation avec la stratégie Cloud nationale

Critères de comparaison



Niveau de contribution à la stratégie Cloud de l'Etat



Présence d'une certification / agrément « Hébergeur Données de Santé »



Offre de service adaptée pour des traitements en science des données



Efforts à fournir pour intégrer la couche applicative

Référence de comparaison

La plateforme de préfiguration sert de référence pour élaborer le scénario de réversibilité

Note sur le lancement du plan de réversibilité

La mise en œuvre du plan de réversibilité s'effectuera **dans un second temps**, selon les **retours d'expérience de la mission** du Health Data Hub et **l'évolution des offres françaises** pour répondre aux différentes contraintes de la plateforme

Le scénario de réversibilité retenu présente certaines limites

Le scénario de réversibilité défini sur OVH est une adaptation du scénario de la plateforme de préfiguration...

- Le **scénario de réversibilité a été construit en capitalisant sur le scénario de la plateforme préfiguration**, autrement dit, la plateforme doit couvrir :
 - Les **mêmes fonctionnalités** (e.g., traitements distribués)
 - Les **mêmes besoins matériel** (e.g., mise à disposition de processeurs graphiques)
 - Les **mêmes exigences de sécurité** (e.g., certification HDS)
 - Les **mêmes principes d'architecture** (e.g., le principe de défense en profondeur)
- Certaines adaptations ont néanmoins dû être réalisées pour fournir valider ces exigences :
 - Combinaison des offres **Cloud privé** et **serveurs dédiés** d'OVH pour fournir un environnement « IaaS / PaaS » (Cloud privé) et donner accès à des GPU (serveurs dédiés)
 - **Intégration de nombreuses solutions tierces** pour combler le manque de maturité de l'offre PaaS d'OVH

... qui présente encore aujourd'hui certaines limitations aussi bien au niveau de la réalisation que du fonctionnement

- **Expérience utilisateur dégradée :**
 - L'utilisation de serveurs dédiés introduit un manque d'élasticité des ressources (e.g., processeurs graphiques) limitant la performance lorsque la demande dépasse les capacités allouées en phase de fonctionnement
- **Surcoût :**
 - Le manque de maturité de l'offre PaaS d'OVH induit d'important coûts d'intégration en phase de réalisation
 - Le provisionnement de ressources non exploitées induit des coûts « non utiles » en phase de fonctionnement
- **Délai supplémentaire :**
 - L'intégration des nombreuses solutions tierces induit des délais supplémentaires en phase de réalisation
 - Le déploiement de ressources supplémentaires (e.g., serveurs), pour satisfaire la demande, allonge les délais de mise en fonctionnement
- **Complexité de gestion de la sécurité :**
 - La multitude de solutions tierces intégrées limite la cohérence des traces et impacte la traçabilité des actions réalisées

L'offre PaaS d'OVH présente un niveau de maturité plus faible que celui d'Azure (sept. 2019)

Cat.	Brique	Scénario Azure		Scénario OVH*
Infra.	Stockage	Azure blob storage		Ceph / OpenStack Swift / Scality
	Sauvegarde (stockage)	Azure Archive		Ceph
	Calcul	Azure VM		Cloud privé (OpenStack / Vmware)
	Réseaux	Azure Virtuel Network		Vrack + NSX
Général	Gestion des identité	Azure AD		OpenIAM / AD / Shibboleth
	Supervision	Azure Monitor		VMware operation manager
Sécurité	Infrastructure de bureaux virtuels	Citrix Cloud		OVH Cloud Desktop
	Pare-feu	Fortinet & Palo Alto		Fortinet & Palo Alto
	Anti-virus	Trend Micro		Trend Micro
	Puits de traces et outil d'analyse	Azure Log Analytics		Logs Data Platform
	Gestion de l'info. de sécu. (SIEM)	Azure Sentinel		Splunk
	Scellement des traces	Tiers archiveur de confiance		Tiers archiveur de confiance
	Gestion des services info.	Service Now (cible)		Service Now
	Gestion des clés de chiffrement	HSM	Azure Key Vault	HSM
	Réseau virtuel privé (VPN)	Azure VPN Gateway		Gateway NSX

* Exemples de technologies

Cat.	Brique	Scénario Azure		Scénario OVH*
Données	Stockage objet	Azure Blob Storage		Ceph / OpenStack Swift / Scality
	Stockage fichier	Azure File / Azure Data Lake Storage		Vmware
	Import plateforme	Azure Gateway + Azure Data Factory		SFTP
	ETL	Azure Data Factory		Talend
	Événements	Azure Data Factory		RabbitMQ / Kafka / etc.
	Workflow	Azure Data Factory		Node red / BPM
	Chiffrement données	SDK Azure		GPG
	Gouvernance	Azure Data catalog		Informatica / Collibra
Projet	Travail collaboratif	Azure DevOps		gitLab (hébergé et sécurisé)
	Traitement non distribué	Azure DSVM + patch		Master ad-hoc
	Traitement distribué	Azure HDInsights		Spark
	Base relationnelle	PostgreSQL managée		PostgreSQL / Maria DB / MySQL
	Base NoSQL	Azure CosmosDB		Clé valeur, Graphe, Document
	Visualisation	PowerBI		QlikView, Tableau, Kibana, Grafana
	Import / export projet	Azure Explorer		SFTP

LÉGENDE

Brique PaaS / IaaS **disponible** dans l'offre de l'hébergeur et **sélectionnée**

Brique PaaS / IaaS **disponible** dans l'offre de l'hébergeur mais **non sélectionnée**

Brique PaaS / IaaS **non disponible** nécessitant d'être entièrement **intégrée**

L'intégration d'une brique technologique **non présente dans l'offre PaaS** de l'hébergeur implique une **importante charge d'intégration** (coûts et délais)

Les chiffrages d'OVH et d'Azure reposent sur les hypothèses de croissance identiques

Projets « de recherche » - type 1

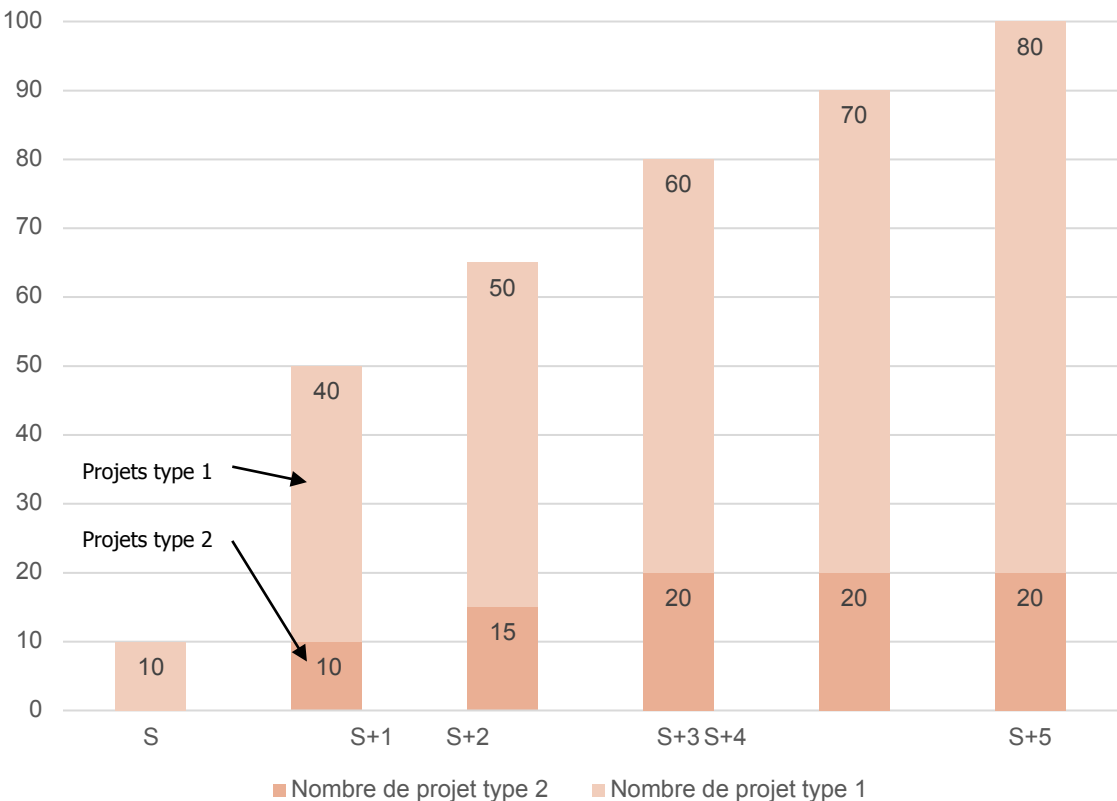
- Les projets de recherche de type 1 ont des besoins de **performance élevés** pour réaliser des traitements avancés en science des données
- Ces projets sont donc caractérisés par des **besoins plus élevés** en capacité de stockage et puissance de calcul
- Il s'agit de projets de taille réduite avec une moyenne de **7 utilisateurs** par projet

Projets « accès permanent » - type 2

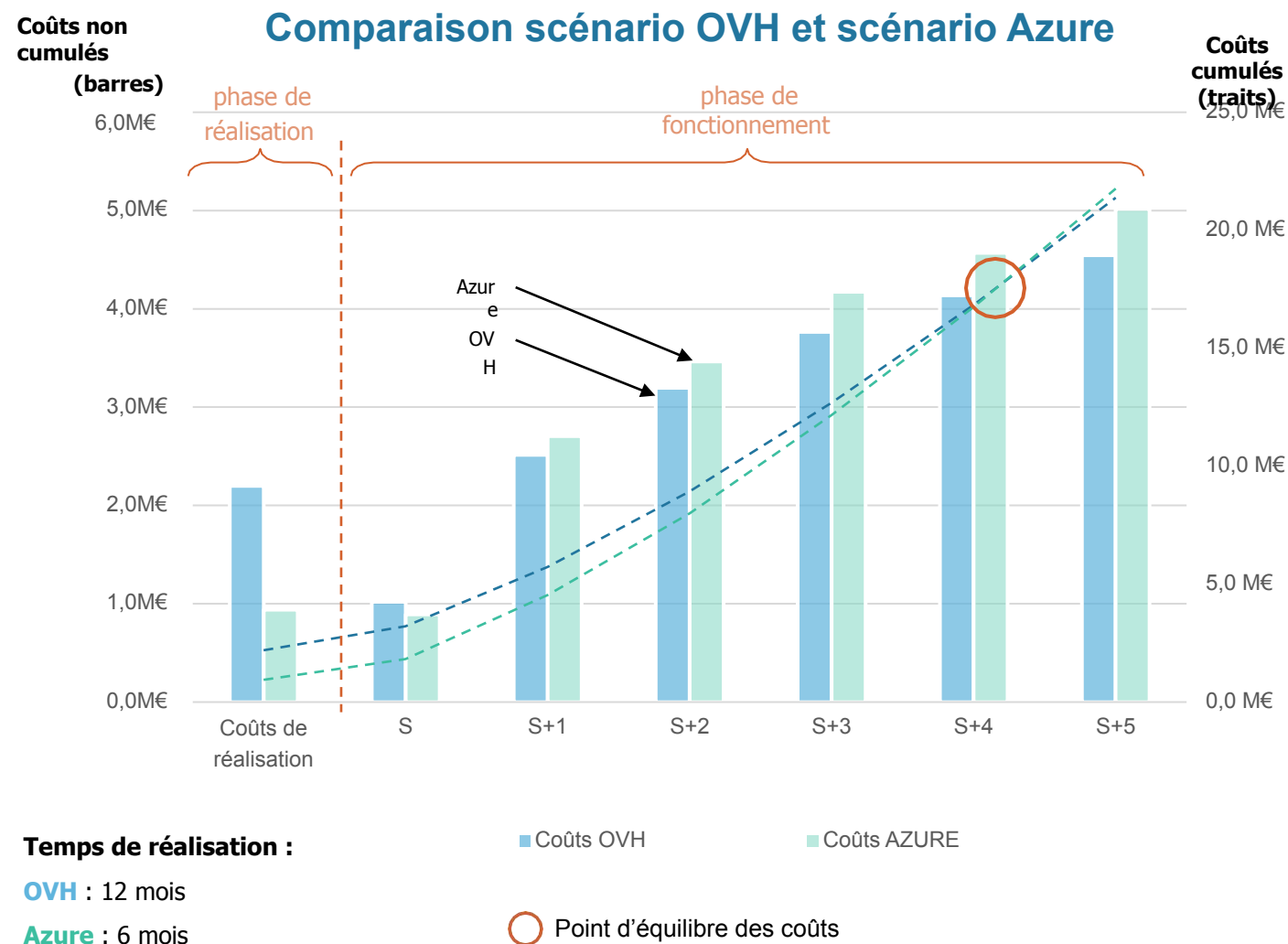
- Les projets de recherche de type 2 ont des besoins de **performance intermédiaires voire faibles**
- Ces projets sont donc caractérisés par des **besoins intermédiaires** en capacité de stockage et puissance de calcul
- Il s'agit de projets de taille plus grande avec une moyenne de **20 utilisateurs** par projet

Projection projet de référence

Hypothèses de croissance du nombre de projets type 1 et type 2



Les résultats de l'étude de réversibilité justifient la stratégie en deux vitesses du Health Data Hub



Temps de réalisation :

OVH : 12 mois

Azure : 6 mois

(hypothèse : équipe de 8 personnes)

Résultat S

Phase de réalisation :

- Les coûts de réalisation du scénario OVH sont supérieurs à ceux du scénario Azure (**2,2M€** vs. **0,9M€**)
- La durée de la phase de réalisation du scénario OVH est supérieure à celle du scénario Azure (**12 mois** vs. **6 mois**)

Phase de fonctionnement :

- Les coûts de fonctionnement du scénario de réversibilité OVH sont en moyenne inférieurs de **275 k€/S** par rapport à ceux d'Azure

Point d'équilibre des coûts :

- Le point d'équilibre des coûts se situe autour du semestre S+4

Le chiffrage de la plateforme est fortement dépendant du nombre de projets lancés

COMPARAISON

Définition de la « projection projet atténuée » :

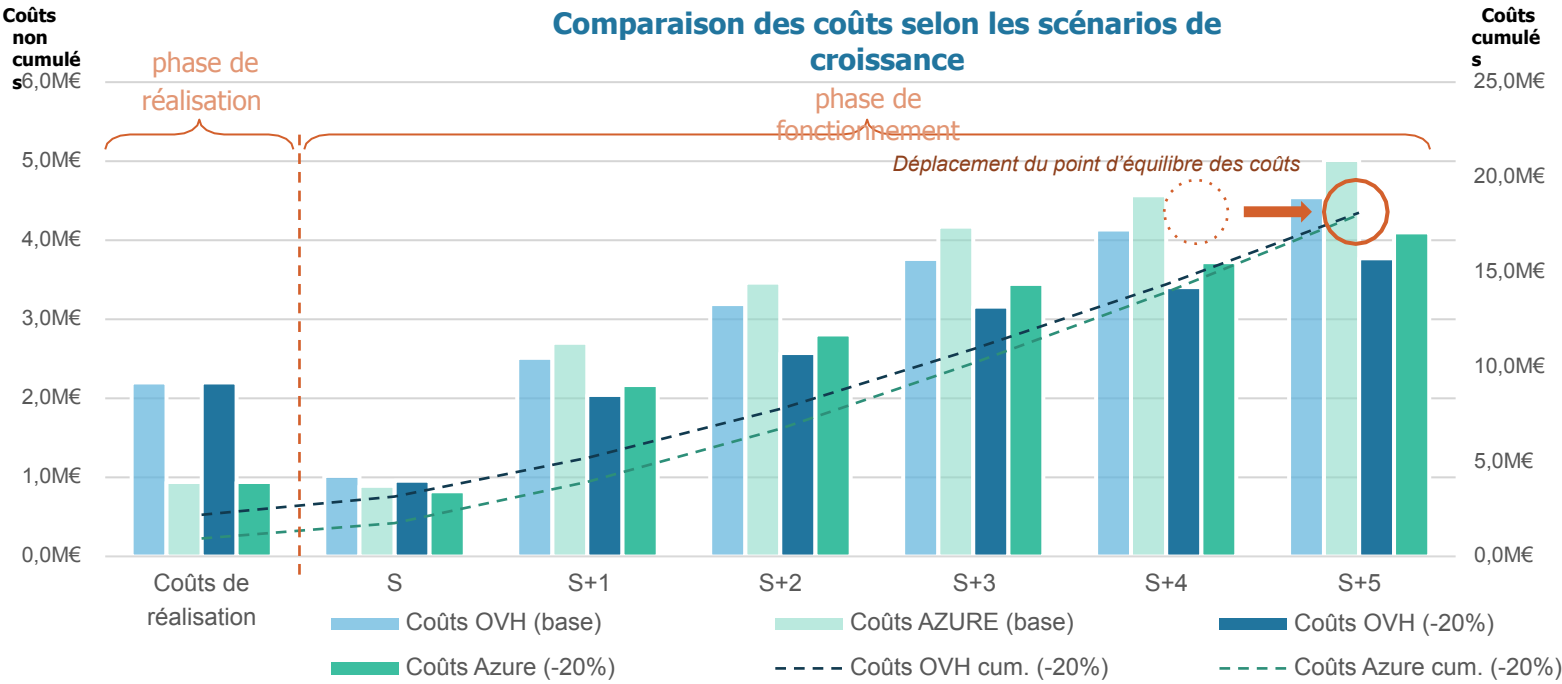
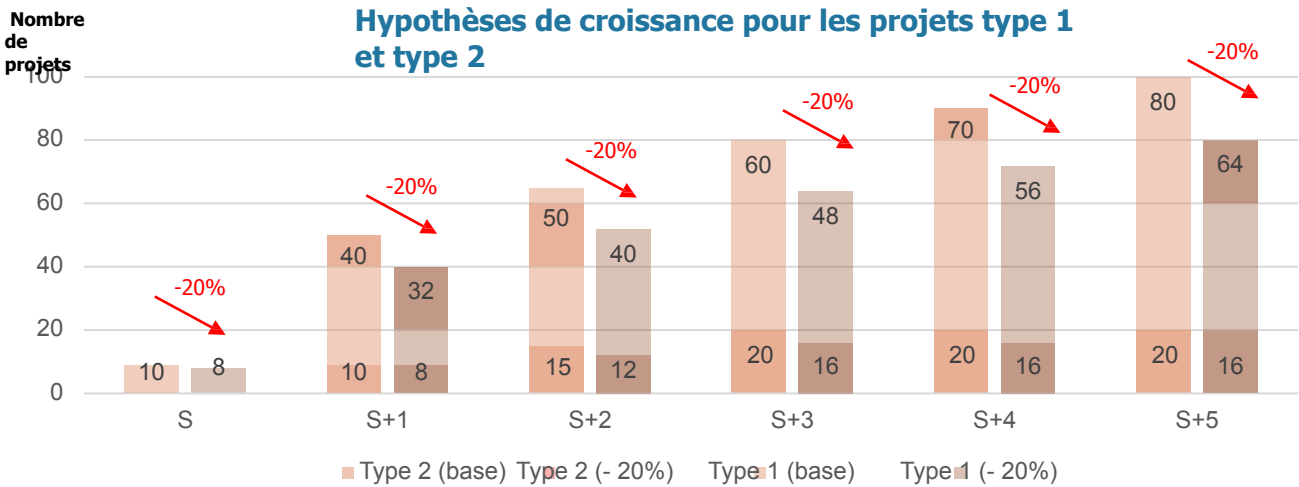
- La projection projet atténuée représente la projection projet de référence diminuée de 20%

Phase de fonctionnement, projection de référence vs. projection atténuée :

- Les coûts moyens de fonctionnement du scénario de réversibilité OVH baissent de **75 k€** par rapport à ceux issus de la projection de référence (**275 k€/S** contre **200 k€/S**)

Point d'équilibre des coûts, projection de référence vs. projection atténuée :

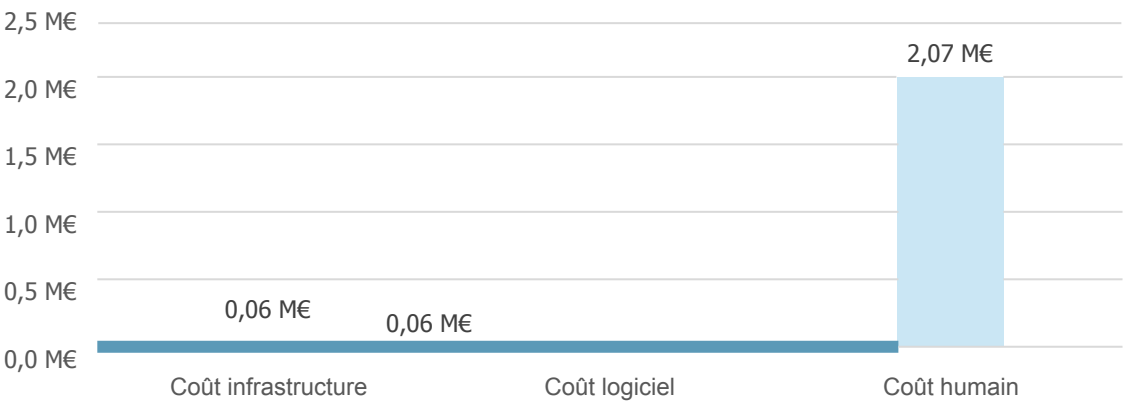
- Le point d'équilibre des coûts issu de la projection atténuée est **retardé d'un semestre** par rapport à celui issu de la projection de référence



Le chiffrage du scénario de réversibilité repose sur des coûts d'infrastructure, logiciels et humains

Scénario de réversibilité OVH

Présentation des coûts de réalisation



Coûts d'infrastructure¹ :

- Les coûts d'utilisation de l'infrastructure sont proportionnels à la durée de réalisation et sont basés sur la tarification publique
- Les coûts des matériels physiques (e.g., HSM) sont basés sur les coûts du scénario de la plateforme de préfiguration

Coûts logiciels² :

- Les coûts des licences logiciels et les coûts d'installation sont basés sur ceux du scénario de la plateforme de préfiguration

Coûts humains³ :

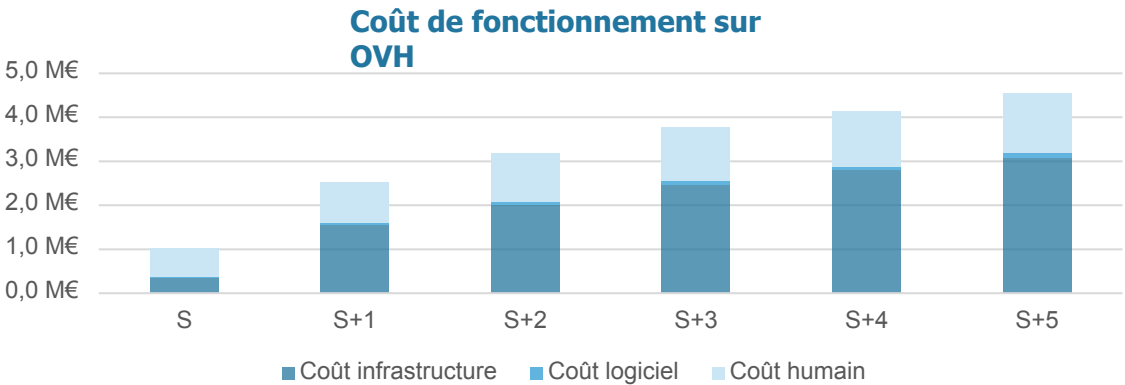
- Les coûts d'intégration sont calculés suivant l'identification des composants de l'architecture, l'analyse de la complexité de leur intégration et l'association d'une quantité de jours homme à chaque niveau de complexité

1 – Coûts infrastructure : coûts des ressources « IaaS » (calcul, stockage, réseau)

2 – Coûts logiciels : coûts des ressources type « PaaS » (licences & support)

3 – Coûts humains : coûts d'intégration et d'opération de la plateforme

Présentation des coûts de fonctionnement



Coûts d'infrastructure¹ :

- Les coûts d'utilisation de l'infrastructure sont calculés sur la base d'une prévision d'utilisation de la plateforme et de la tarification publique

Coûts logiciels² :

- Les coûts de support sont basés sur ceux du scénario de la plateforme de préfiguration

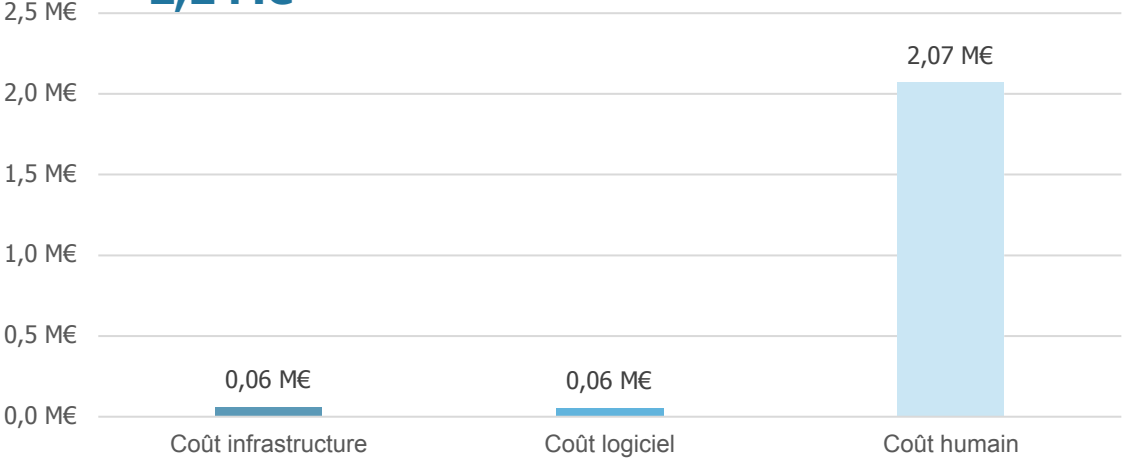
Coûts humains³ :

- Les coûts de gestion fonctionnelle et de MCS⁴ de la plateforme sont calculés en fonction du nombre d'utilisateurs
- Les coûts de MCO⁵ relatifs à la gestion des machines virtuelles sont calculés selon le nombre de machines qu'un opérateur peut administrer
- Les coûts de maintien à jour sont calculés sur la base des coûts d'intégration

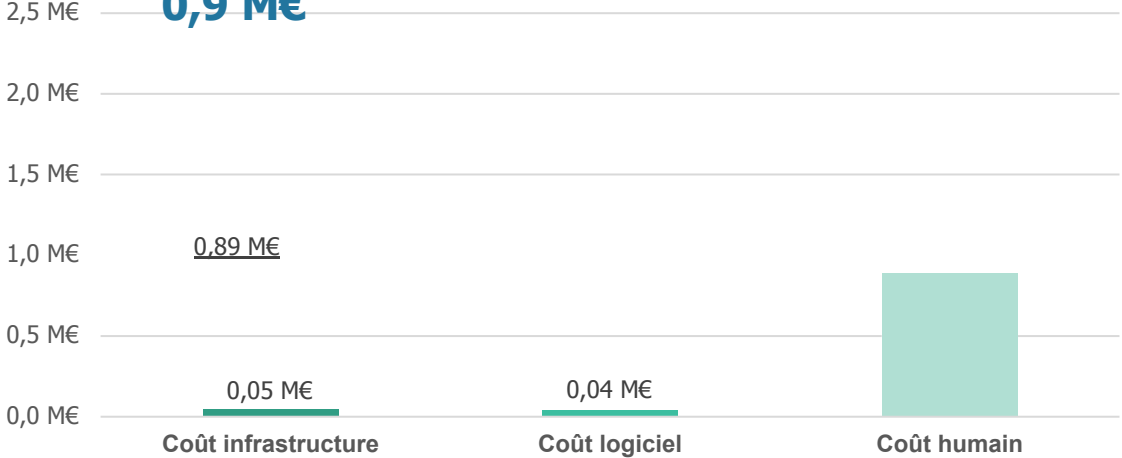
4 – MCS : maintien en condition de sécurité
5 – MCO : maintien en condition opérationnelle

Les coûts du scénario de réversibilité peuvent être comparés aux coûts du scénario de la plateforme préfiguration

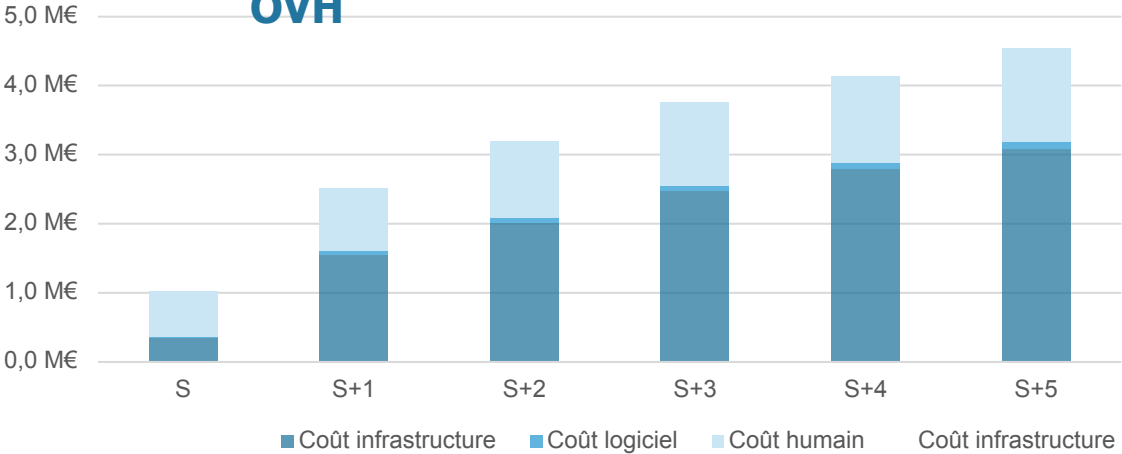
Coût total de réalisation sur OVH :
2,2 M€



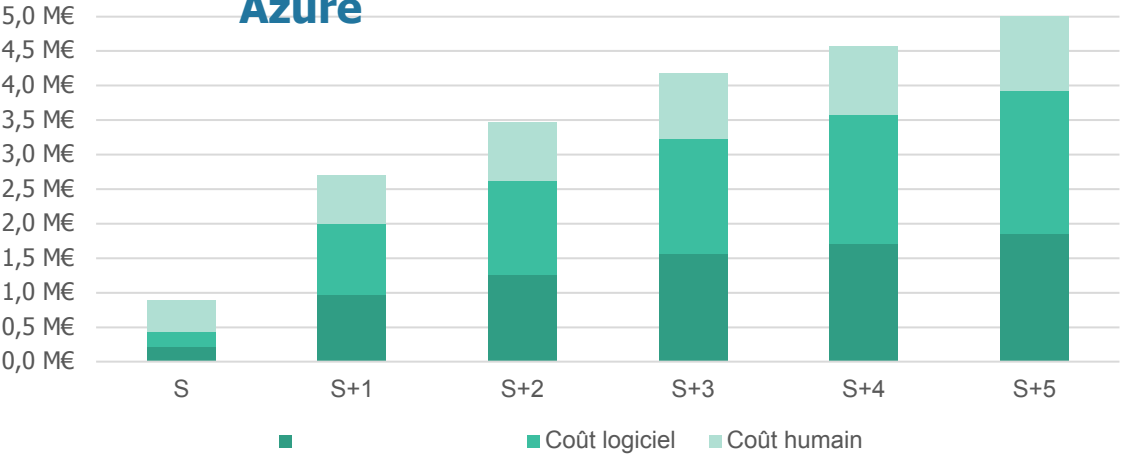
Coût total de réalisation sur Azure :
0,9 M€



Coût de fonctionnement sur
OVH



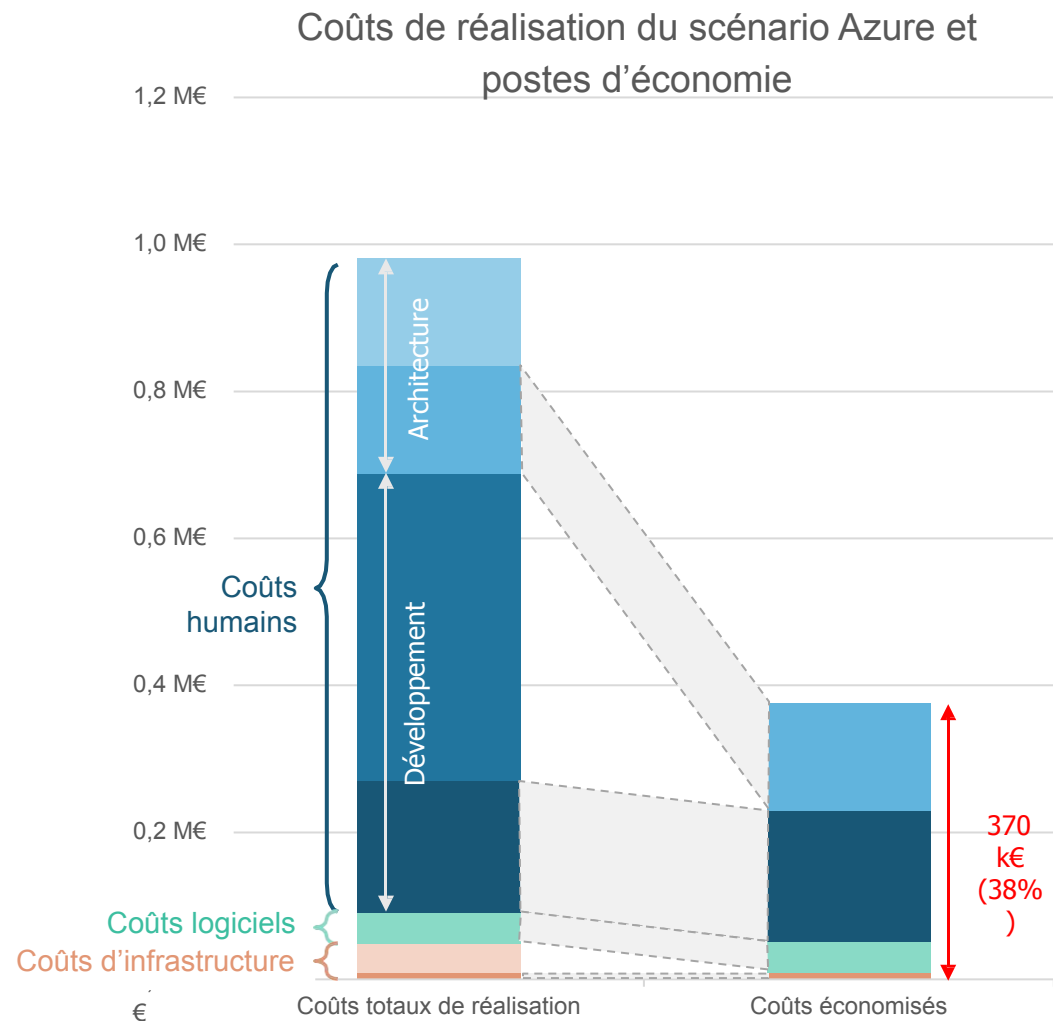
Coût de fonctionnement sur
Azure





- **Contexte**
- **Etude de réversibilité**
- **Capitalisation sur les travaux menés**

Le travail mené lors de la mission du Health Data Hub sert la mise en place de la plateforme cible



La plateforme de préfiguration **assure la réversibilité des données stockées et des travaux de recherche** qui y sont conduits

Lors d'un portage vers un autre hébergeur, des **économies de transfert importantes (environ 38%)** sont permises grâce à :

- Coûts humains – architecture : la conservation d'une grande partie des **travaux d'architecture** réalisés (estimation : 50% d'économie)
- Coûts humains – développement : la réutilisation d'une partie des **codes de déploiement** des ressources non Azure (Citrix, Fortinet, Palo Alto, etc.)
- Coûts logiciels : le portage **des licences logicielles** achetées
- Coûts d'infrastructure : le portage de la **boîte noire transactionnelle** (HSM)

Ces économies sont néanmoins à relativiser étant donné la **faible importance des coûts de réalisation par rapport à ceux de fonctionnement** (avantage du cloud par rapport aux infrastructures dédiées)

Il est cependant important de noter que, même si cette réversibilité est possible, elle offrira actuellement un **niveau de service inférieur à celui proposé par Azure** :

- Une expérience utilisateur dégradée
- Des coûts et des délais de réalisation plus importants sans compter les procédures de marché
- Des processus de fonctionnement plus lourds
- Une gestion de la sécurité plus difficile à mettre en place

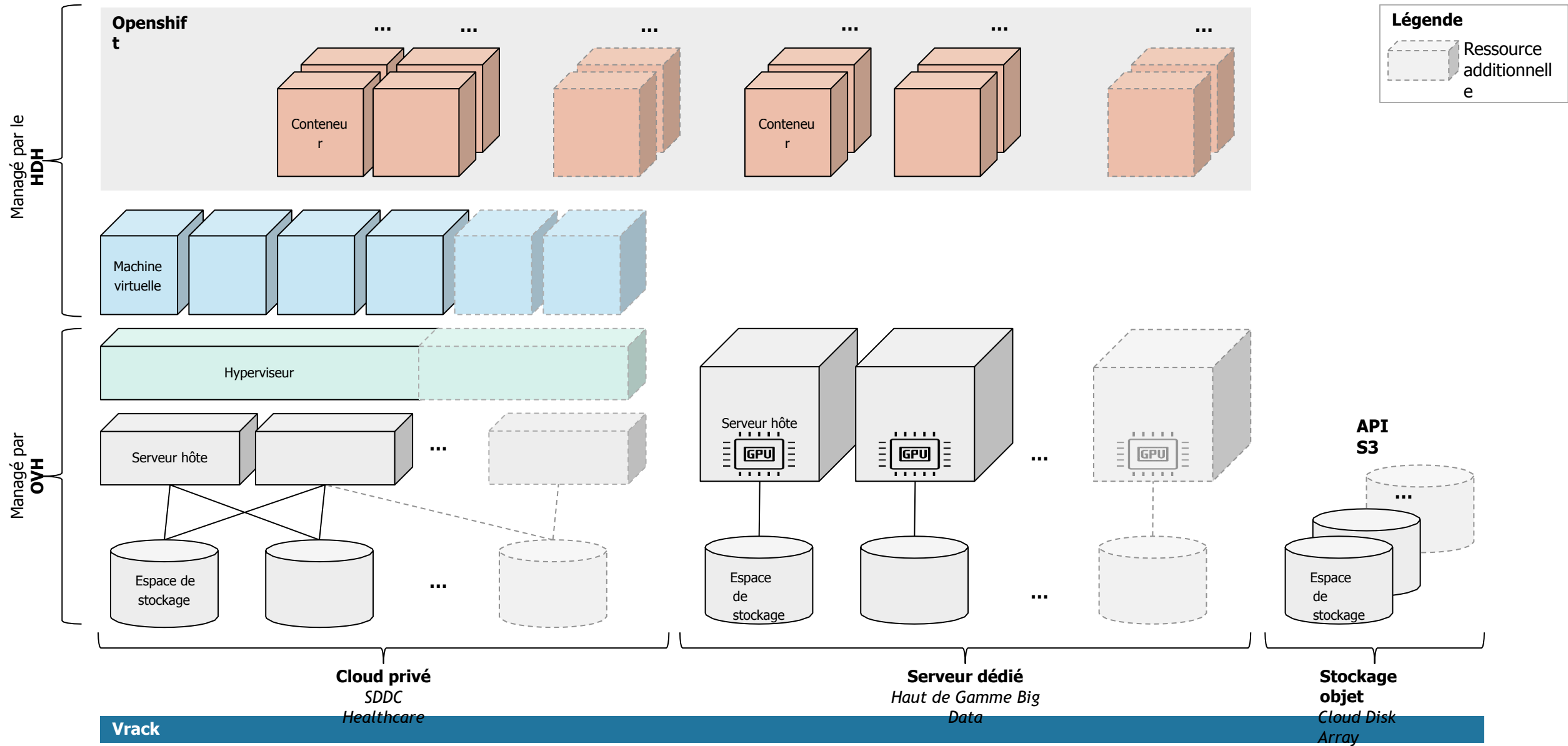
Annexe

DOCUMENT ADMINISTRATIF COMMUNIQUE PAR LE HEALTH DATA
HUB – PUBLICATION EN LIGNE (article L.312-1-1 du CRPA)

OVH a été retenu parmi les solutions identifiées dans la stratégie Cloud nationale

			
STRATÉGIE NATIONALE 	Acteur identifié dans la stratégie Cloud de l'Etat	Acteur identifié dans la stratégie Cloud de l'Etat	Acteur identifié dans la stratégie Cloud de l'Etat
Coût 	Certification HDS sur les Cloud privé et serveurs dédiés en 2019	Agrément HDS sur l'ensemble des centres de données parisiens en 2019	Renouvellement de l'agrément HDS en cours Qualification SecNumCloud en cours pour l'offre Cloud Secteur Public
SÉCURITÉ 	Qualification SecNumCloud en cours pour l'offre Cloud privé	Processeur graphique disponible	Processeur graphique disponible
FONCTIONNALITÉ 	Offre principalement IaaS évoluant sur du PaaS minimisant les coûts d'intégration	Offre principalement IaaS induisant des coûts d'intégration	Offre principalement IaaS induisant des coûts d'intégration

Annexe – Architecture envisagée chez OVH



Annexe – Caractéristiques de l’offre OVH et hypothèses retenues sur le scénario associé

Thèmes	Caractéristiques du l’offre / Hypothèses du scénario
OFFRE SERVEURS DÉDIÉS	
PROCESSEUR GRAPHIQUE DE L’OFFRE SERVEUR DÉDIÉ	▪ Les serveurs dédiés Big Data ne peuvent embarquer qu’un processeur graphique dans l’offre OVH, à date ▪ Des cartes graphiques NVIDIA P100 ont été choisies pour les serveurs dédiés Big Data
CONTENEURISATION	▪ L’architecture du scénario OVH repose sur l’utilisation du service PaaS Openshift qui embarque Kubernetes, un logiciel d’orchestration de conteneurs ▪ L’hypothèse d’un nombre de 5 conteneurs par machine virtuelle permet d’optimiser les coûts
STOCKAGE OBJET	▪ L’exploitation du stockage objet, sur le scénario OVH retenu, repose notamment sur l’API S3 qui n’est pas active dans l’offre OVH, à date
TARIFICATION	▪ Le scénario OVH a été construit en s’efforçant d’offrir une fonctionnalité comparable à la plateforme de préfiguration ▪ Les prix utilisés dans le chiffrage du scénario OVH sont les prix publiques affichés sur le site, avant le 28/10/2019 (Sources : Stockage objet, Cloud privé, Espace de stockage additionnel Cloud privé, Serveur hôte additionnel Cloud privé, OVH Cloud Connect, Serveur dédié Big Data, Cloud desktop infrastructure, Openshift)

Annexe – Hypothèses sur les coûts de réalisation et coûts de fonctionnement du scénario OVH

Thème	Hypothèses / Caractéristiques
COÛTS D'INTÉGRATION	- Les coûts d'intégration ont été calculés en fonction d'un TJM de 1000 € et d'un nombre de jours nécessaires pour l'intégration des différentes briques - Les jours nécessaires ont été déterminés selon la complexité d'intégration : Simple, 5 jours ; Moyen 10 jours ; Complexe, 20 jours; Très complexe, 40 jours - Les coûts totaux d'intégration prennent en compte un facteur multiplicateur (3x) afin de modéliser l'ensemble des étapes de réalisation, à savoir : la spécification du besoin, la conception technique, les coûts de développement, les tests unitaire, les tests d'intégration, l'optimisation de la performance, la période de recette, et la gestion de projet (NB : les coûts liés au maintien en condition opérationnelle et maintien en condition de sécurité ne sont pas inclus)
NOMBRE D'ÉQUIVALENT TEMPS PLEIN EN PHASE D'INTÉGRATION	Une moyenne de 8 équivalents temps plein a été retenue pour le développement de la plateforme en phase d'intégration
COÛTS SUPPLÉMENTAIRES LIÉS À LA MCO ET MCS	- Des coûts supplémentaires sont ajoutés au chiffrage d'OVH pour modéliser la différence de complexité de la maintenance des systèmes OVH par rapport à Azure - L'hypothèse retenue est que ces coûts annuels représentent 20% des coûts totaux de réalisation
NOMBRE D'INFRASTRUCTURE DE BUREAU VIRTUEL	Chaque utilisateur et opérateur de la plateforme utilise une infrastructure de bureau virtuel dédiée associée à une machine virtuelle
ESPACE OPÉRATEUR	17 machines virtuelles d'administration et 3 conteneurs d'administration ont été dénombrés, en plus des machines virtuelles lié au nombre d'opérateurs - Un besoin d'une capacité de stockage de 100 TB a été retenu - Les traitements réalisés sur l'espace opérateur ne nécessitent pas de recourir à un serveur Big Data
ESPACE PROJET	2 machine virtuelles et 5 conteneurs ont été dénombrés par espace projet, en plus des machines virtuelles lié au nombre d'utilisateurs - Un besoin d'une capacité de stockage de 10 TB a été retenu - Les caractéristiques des projets de type 1 sont : une moyenne de 7 utilisateurs, une quantité moyenne de 5 GPU allouées à l'heure, une quantité moyenne de 2 GPU allouée au mois (NB : l'offre OVH ne permet pas, à date, une tarification des serveurs dédiés Big Data à l'heure) - Les caractéristiques des projets de type 2 sont : une moyenne de 20 utilisateurs, une quantité moyenne de 3 GPU allouées à l'heure, une quantité moyenne de 0 GPU allouée au mois (NB : l'offre OVH ne permet pas, à date, une tarification des serveurs dédiés Big Data à l'heure)
CROISSANCE DES COÛTS DES OPÉRATEURS	- Le salaire annuel d'un opérateur est estimé à 100k€ - Un seul superopérateur est nécessaire - Une redondance de chaque type d'opérateur est prise en compte (hors superopérateur) - Un opérateur plateforme peut gérer 350 machines virtuelles (contre 1000 machines virtuelles sur le scénario Azure) - Un opérateur données peut gérer un maximum de 20 projets - Un opérateur projet peut gérer un maximum de 150 utilisateurs - Un opérateur sécurité peut gérer un maximum de 150 utilisateurs

Annexe – Etude de réversibilité : externalisation des services de la plateforme de préfiguration

Catégorie	Brique	Actuel	Cible	Valeur/Criticité	Complexité*
				*	
Général	Serverless	Azure Function	Application custom	2	2
	Batch	Azure Batch	Application custom	2	3
	Sauvegarde	Azure Archive	Application custom	1	2
Supervision (MCO)		Azure Monitor	Centréon	3	2
	Accès VDI	Citrix Cloud	Citrix Cloud	0	5
	Droits / Authentification	Azure AD	AD/LDAP	5	3
Sécurité	Réseau / Firewall	Fortinet / Palo Alto	Fortinet / Palo Alto	0	0
	Anti-virus / EDR	Trend Micro	Trend Micro	0	0
	SIEM	Sentinel	Splunk / ??	5	2
	Logs	Log Analytics	Splunk / ELK	5	4
	Scellement des logs	Tiers archiveur de confiance	Tiers archiveur de confiance	0	0
	ITSM	e-mail	ServiceNow	4	3
	SSO	Azure AD & environnement Azure	AD & écosystème cible	5	5
	Gestion des clés / HSM	Azure Key Vault + nCipher	nCipher & écosystème	5	5
VPN		Azure VPN gateway	Cisco, FS, etc.	4	1
	Stockage Objet	Blob Storage	Solution hébergeur	0	N/A
Données	ETL	Data Factory	Talend, etc.	4	3
	File Storage	Azure File / ADLS	HDFS	5	4
	Workflow	Logic App	Node red / BPM	1	3
	Chiffrement données	SDK Azure	Lib Open source	2	2
Gouvernance		Azure data catalog	Informatica / Collibra	4	1
	DevOps	Azure DevOps	GitLab	2	1
	Execution non distribué	Azure VMDS + patch	Master ad-hoc	1	2
Projet	Execution distribué	HDInsights	Hortonworks	0	4
	Base SQL	Postgres managée	Postgres	0	3
	Base NoSQL	CosmoDB	Clé valeur, Graphe, Document	à la demande	à la demande
	Visualisation	PowerBI	QlikView, Tableau	2	2
	Ingestion données	Azure explorer	SFTP	1	2
	CI / CD plateforme	gitLab (git) (SaaS)	gitLab (hosted + secured)	5	1

* Note de 0 à 5

Merci

DOCUMENT ADMINISTRATIF COMMUNIQUE PAR LE HEALTH DATA
HUB – PUBLICATION EN LIGNE (article L.312-1-1 du CRPA)