

# Rapport de stage

*ING 5 SI – M2 Data sciences  
2022/2023*

ECE Paris  
Institut Polytechnique de Paris  
AXA Investment Managers

# Remerciements

Je tiens chaleureusement à exprimer ma gratitude envers AXA Investment Managers pour m'avoir accueilli au sein de leur équipe et m'avoir permis d'effectuer mon stage de fin d'études.

Je souhaite également adresser des remerciements tout particuliers à mon tuteur de stage, Ruddy Haddad, pour m'avoir accueilli au sein de son équipe et pour avoir été mon guide tout au long de cette expérience.

Un grand merci également à Monsieur Grégoire Bessau, Technical Leader, pour m'avoir encadré sur tous les aspects techniques de mon stage. Son expertise et ses bonnes pratiques dans le domaine de la technologie me seront utiles tout au long de ma carrière professionnelle.

Un grand merci aussi à tous les autres membres de la Squad finance, Pierre Nguyen et Julie, pour avoir partagé leurs connaissances dans tout ce qui est en lien avec le monde de l'asset management.

Je remercie également mes proches, ma famille et mes amis, pour leur encouragement et leur soutien inconditionnel tout au long de cette période.

# Glossaire

- AUD
- AUM
- Asset
- Segmentation

# Table des matières

<b>I.</b>	<b>INTRODUCTION .....</b>	<b>6</b>
<b>II.</b>	<b>PRESENTATION GENERALE DU CADRE DU STAGE .....</b>	<b>7</b>
A.	LE GROUPE AXA.....	7
B.	AXA INVESTMENT MANAGEMENT .....	9
1.	<i>Présentation</i> .....	9
2.	<i>AXA IM en chiffres</i> .....	9
3.	<i>Le marché des gestionnaires de fonds</i> .....	10
4.	<i>Organigramme</i> .....	11
5.	<i>Business unit</i> .....	11
C.	POLITIQUE RSE D'AXA INVESTMENT MANAGERS : .....	15
D.	FINANCIAL DATA TRANSPARENCY : .....	16
1.	<i>Enjeux du projet</i> .....	16
2.	<i>Financial Dashboard</i> .....	16
E.	PRESENTATION DE LA MISSION DE STAGE : .....	18
1.	<i>Définition et enjeux de la mission</i> :.....	18
2.	<i>Gestion de projet – Méthode Agile</i> .....	18
3.	<i>Planification prévisionnelle du projet</i> .....	19
4.	<i>Cahier des charges</i> .....	21
<b>III.</b>	<b>REALISATION DES MISSIONS .....</b>	<b>22</b>
A.	ASPECT TECHNIQUE GLOBAL DE MES MISSIONS : .....	22
1.	<i>La dataplatform</i> .....	22
2.	<i>Data flows</i> .....	24
B.	TECHNICAL ASPECT OF DATA TRANSPARENCY.....	25
1.	<i>Schéma architecture</i> .....	25
2.	<i>Data procession for Dashboard</i> .....	27
C.	ESSBASE – HISTORICAL DATA INDUSTRIALIZATION .....	28
D.	FINANCIAL - CUSTOM INPUTS INDUSTRIALISATION .....	29
1.	<i>Présentation des Template d’ingestion</i> :.....	31
2.	<i>Présentation des transformation silver to silver reporting</i> :.....	31
E.	ANONYMISATION OF THE DATA .....	32
F.	FINANCE REPORTING TO UNITY CATALOG .....	33
G.	EXPORT FROM DP .....	34
<b>IV.</b>	<b>BILAN ET PERSPECTIVES .....</b>	<b>35</b>

## Table des figures

Aucune entrée de table d'illustration n'a été trouvée.

## I. Introduction

Cette année, j'ai eu la possibilité de façonner ma formation académique en effectuant un double diplôme avec l'Institut Polytechnique de Paris au sein du master 2 en data science. Cette formation est ainsi venue répondre à mon besoin d'en apprendre davantage sur le monde de la data. En combinaison avec ma formation initiale d'ingénieur à l'ECE Paris et avec une base solide acquise lors de mon précédent stage dans le domaine du software engineering, j'ai ainsi eu la volonté d'effectuer un stage de fin d'études en tant que Data Engineer, ce métier qui pour moi est à la croisée du système d'information et de la Big Data.

J'ai eu l'opportunité de rejoindre AXA Investment Managers pour un stage de fin d'études avec une double position de Data Engineer et de Data Analyst.

Ce stage a été une véritable révélation pour moi, car j'ai pu rejoindre l'un des plus grands gestionnaires d'actifs en France, et il m'a permis de découvrir ce monde de la gestion d'actifs, mais il m'a aussi permis de développer grands nombres de compétences.

En quelques mots, on définit l'asset management comme l'action d'une société gérant les fonds de ses clients dans le but de les fructifier à travers différents placements financiers.

Nous verrons au travers de ce rapport de stage, l'ensemble des missions qui m'ont été confiées en tant que Data Engineer et dont la finalité est la mise en place de du projet Data Transparency au sein de AXA Investment Management.

## II. Présentation générale du cadre du stage

### A. LE GROUPE AXA

L'aventure du groupe AXA a commencé il y a 60 ans, d'une petite société mutuelle d'assurance en Normandie, qui occupe aujourd'hui le rang de leader mondial dans le monde de l'assurance. Depuis sa création, le parcours d'AXA a été marqué par un grand nombre d'acquisitions et par une forte volonté de s'étendre à l'international.

Aujourd'hui, AXA compte près de 93 millions de clients à travers 51 pays et 145 000 collaborateurs. Son cœur de métier, l'assurance s'adresse aux particuliers et aux entreprises. Opérant au travers de 5 hubs : France, l'Europe, l'Asie, le Moyen-Orient et l'Afrique, elle y exerce diverses activités :

#### **Assurance dommages : 48,73 milliards d'€ (50,38 % du C.A 2020)**

- Assurances auto : contrats d'assurance pour les véhicules automobiles et utilitaires.
- Assurances habitation : axa offre divers contrats d'assurance habitation, ainsi que des assurances emprunteur en cas de prêt pour l'achat immobilier.
- Assurances voyages et loisirs : assistance 24h/24, 7j/7, protection contre d'éventuels dommages matériels ou corporels lors de voyages, de vacances, de la chasse ou d'activités de plaisance.

#### **Produit financier d'épargne : 31,52 milliards d'€ (32,59 % du C.A 2020)**

- Assurance retraite et PER : contrats d'épargne retraite et des Plans d'Épargne Retraite, offrant des avantages fiscaux aux souscripteurs.
- Assurance vie : c'est le produit d'épargne préféré des Français, et Axa y propose une large gamme de contrats d'assurance vie.
- Gestion privée : Axa met à disposition des experts en gestion de patrimoine pour les clients intéressés par la gestion privée.

#### **Assurances santé : 14,71 milliards d'€ (15,21 % du C.A 2020)**

- Santé : contrats de complémentaire santé.
- Prévoyance : gammes de contrats tels que la protection contre les accidents de la vie, l'assurance obsèques.

#### **Activités bancaires : 0,49 milliards d'€ (0,51 % du C.A 2020)**

- Banque : services bancaires tels que des comptes bancaires, des prêts personnels et immobiliers, des livrets d'épargne, ainsi que des conseils.
- Axa XL, filiale du groupe Axa, se spécialise dans l'assurance et la réassurance pour les entreprises de tailles intermédiaires, les multinationales et les grandes fortunes.

**Gestion d'actifs : 1,27 milliards d'€ (1,31 % du C.A 2020)**

- AXA Investment Managers (AXA IM) est une filiale d'AXA spécialisée dans la gestion d'actifs. Elle gère des portefeuilles d'investissements pour le compte de clients institutionnels, de particuliers fortunés et d'autres entités (Third Party) mais aussi pour le compte de AXA (Main Fund).

Au niveau international, AXA Groupe est en concurrence direct avec les autres plus grandes compagnies d'assurance au monde tel que l'allemand Alliance, ou et le chinois « china life insurance », mais aux vues de la diversité de ces activités sur le secteur financier peut également être confronté à la concurrence des banques, de gestionnaires d'actifs et d'autres institutions financières tel que Amundi.



## B. AXA Investment Management

### 1. Présentation

AXA Investment Managers (AXA IM) représente la filiale spécialisée dans la gestion d'actifs au sein du groupe AXA. Établie en 1994, elle opère de manière complètement indépendante du groupe AXA. L'objectif central d'AXA IM est de se positionner en tant que leader international de la gestion d'actifs responsable. Cela se réalise grâce à des investissements actifs à long terme, orientés vers des produits visant à contribuer à la construction d'un avenir durable pour la planète.

*'Investing responsibly for what matters to our clients and the world we live in.'*

#### **Une mission fondamentale :**

AXA IM agit en faveur du progrès humain en investissant dans ce qui compte. Cette démarche est guidée par une mission allant au-delà des simples profits financiers. Elle vise à créer un impact positif et durable en contribuant au bien-être de la société et à l'évolution de l'humanité.

#### **Une conviction forte :**

L'approche d'AXA IM est fortement imprégnée par leurs convictions, cela lui permet de dénicher des opportunités d'investissement mondiales dans les domaines divers tel que les obligations, des actions, des multi-assets et des investissements alternatifs.

#### **Le groupe AXA :**

AXA IM bénéficie du solide soutien de son groupe, cette base solide lui permet de renforcer ses compétences en gestion d'actifs et de proposer des solutions d'investissement de haute qualité tant le nom de Axa est reconnu mondialement pour sa fiabilité.

L'ensemble de ces valeurs décrit parfaitement la philosophie d'AXA IM, marquée par un engagement envers le progrès humain et une vision d'investissement responsable et durable.

### 2. AXA IM en chiffres

- **Actifs sous gestion (AUM) :** 824 milliards d'euros
- **Fonds et mandats :** 1355
- **Présence internationale :** Opérant dans 19 pays
- **Effectif :** 2 500 employés
- **Business Unit :** Core, Alts (Alternatives), Architas, Prime
- **Résultat net :** 348 millions d'euros
- **Chiffre d'affaires :** 1 425 millions d'euros

AXA Investment Managers (AXA IM) pour assurer sa rentabilité, l'entreprise doit générer des revenus en appliquant divers frais liés à la gestion des actifs.

- Les frais de souscription sont perçus au moment de la souscription à un fonds.
- Les frais de gestion annuels, représentant un pourcentage des actifs sous gestion.
- Les frais de maintenance sont associés aux coûts des transactions lors de l'achat et de la vente d'actifs au sein du portefeuille.
- Les frais de rachat sont perçus lors du retrait des fonds par les clients.

En cas de performances exceptionnelles, des frais au rendement sont imposés pour récompenser les gestionnaires. Ces divers mécanismes financiers de frais sont essentiels pour soutenir les opérations d'AXA IM et assurer sa viabilité économique dans le secteur de la gestion d'actifs.

### 3. Le marché des gestionnaires de fonds

Le monde de la gestion d'actifs est hautement concurrentiel, les compagnies peuvent être comparées sur plusieurs critères tels que la performance, mais l'un des critères les plus importants est le montant des actifs sous gestion (AUM).

Rang	Société de gestion	AUM en milliards d'euro 2022
1	Amundi	1420
2	Natixis Investment Managers	830
3	AXA Investment Managers	745
4	BNP Paribas Asset Management	416

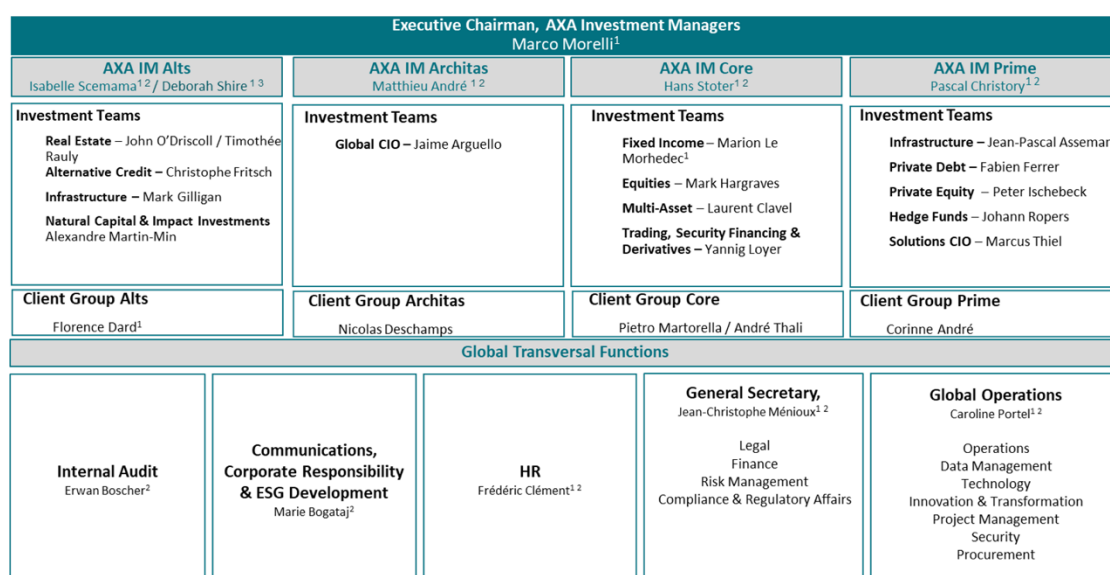
En 2022, Malgré la crise actuelle qui touche le marché à cause de l'augmentation des taux mais aussi de l'ensemble des conflits géopolitique en cours, AXA IM a su figure parmi les grands en enregistrant de la gestion d'actifs avec une de collecte nette en 2022, 17 milliards d'euro.

Sur un plan plus international, dans le monde de la gestion d'actifs, les compagnies américaines mènent la course bien loin devant les sociétés françaises. Amundi, leader français, se classe quant à lui parmi les cinq premiers au niveau mondial.

Rang	Société de gestion	AUM en milliards d'euro 2023
1	BlackRock	9425
2	Vanguard	7250
3	Fidelity Management & Research	3880

#### 4. Organigramme

La structure générale d'AXA s'est organisée autour de deux grands business unit, CORE et Prime, depuis le début de son histoire. Cependant, avec la récente restructuration d'AXA IM, deux nouveaux business unit, Alts et Architas, ont été ajoutées. Tous ces domaines corps de métiers sont soutenus par des fonctions transversales, tel que la communication, l'audit interne, et global opérations.



Sous la supervision de Maroc Morelli, évoluer indépendamment les quatre différents business avec un support des fonctions transversal composé par global opérations, audit interne et autre.

#### 5. Business unit

##### a) **CORE :**

- 496 milliards € AUM
- 720 collaborateurs
- 3900 clients

AXA IM Core se distingue en excellent dans la gestion active des investissements de base, privilégiant des résultats à long terme et la conviction que les investissements

peuvent apporter des bénéfices doubles, tant sur le plan financier que social. Cela le place en tant qu'acteur majeur dans le domaine des stratégies d'investissement. L'expertise distinctive d'AXA IM Core s'étend sur une variété de stratégies d'investissement, couvrant différentes classes d'actifs.

#### **Fixed Income :**

- *High Yield* : Solutions axées sur les obligations à haut rendement, offrant des opportunités de rendement supérieur.
- *IG Credit* : Excellence dans les investissements en crédit à la qualité d'investissement, adapté aux investisseurs recherchant un profil de risque plus stable.
- *Emerging Markets* : Opportunités d'investissement dans les marchés émergents, participant à la croissance de ces économies dynamiques.
- *Govt Bonds* : Obligations gouvernementales, fournissant des options de placement dans des titres émis par des gouvernements.

#### **Equities :**

- *Global* : Stratégies d'investissement axées sur des opportunités globales en actions.
- *Regional and Country* : Stratégies ciblées sur des régions spécifiques ou des pays.
- *Thematic* : Distinguée par des stratégies thématiques, permettant aux investisseurs de participer à des thèmes spécifiques qui façonnent les marchés.
- *Sustainable & Impact* : Accent mis sur les investissements durables et à impact, intégrant des considérations ESG dans ses stratégies.

AXA IM Core propose une gamme complète de stratégies d'investissement, permettant aux investisseurs de personnaliser leurs portefeuilles selon leurs objectifs financiers spécifiques.

#### **ALTS :**

- 186 milliards € AUM
- 950 collaborateurs
- 1200 clients

AXA IM ALTS se distingue comme la seule grande plateforme purement dédiée aux alternatives en Europe et l'une des plus vastes à l'échelle mondiale, détenant une position de leadership établie et une expertise reconnue dans toutes les catégories alternatives d'actifs.

Elle permet aux investisseurs de diversifier leurs portefeuilles et de participer à des opportunités de croissance dans des secteurs en leur offrant des stratégies d'investissement et des fonds alternatifs.

- *Real estate and infrastructure Debt* : AXA IM Alts excelle dans les investissements de dette immobilière, offrant des opportunités dans ce secteur pour générer des rendements.
- *Real estate equity* : La société excelle également dans les investissements en actions immobilières, offrant des opportunités pour participer à la croissance du marché immobilier.

- Impact private equity : Alts met l'accent sur le private equity à impact, démontrant son engagement envers des investissements générant un impact social et environnemental positif.
- Alternative crédit : AXA IM Alts offre des stratégies de crédit alternatives pour les start up.

#### *Prime :*

- 36 milliards € AUM

Prime offre aux investisseurs une plus grande flexibilité et une liquidité accrue par rapport aux investissements traditionnels pour les investisseurs cherchant à diversifier leur portefeuille avec des actifs privés.

- Private Equity : offrant des opportunités d'investissement dans des entreprises non cotées, avec un accent particulier sur la création de valeur à long terme.
- Infrastructure Equity : L'entreprise se distingue également dans le domaine de l'infrastructure equity, investissant dans des projets liés aux infrastructures pour générer des rendements attractifs tout en contribuant au développement économique.
- Private Debt : Des solutions de dette privée, permettant aux investisseurs d'accéder à des opportunités de financement privé avec des rendements ajustés au risque.
- Nav Financing : L'entreprise offre des solutions de financement de la valeur nette d'inventaire (Nav financing), permettant aux investisseurs de tirer parti de la valeur de leurs actifs tout en maintenant une approche flexible.

#### *Architas :*

- 30 milliards € AUM

Architas se positionne comme un acteur majeur offrant une gamme diversifiée de services dans le domaine de la gestion d'actifs.

- Funds of Funds : AXA IM Architas excelle dans la création et la gestion de fonds de fonds, offrant aux investisseurs une approche diversifiée en investissant dans une sélection de fonds provenant de différentes classes d'actifs depuis les autres business unit CORE et ALTS.
- Investment Advisory : L'entreprise propose des services de conseil en investissement, mettant à profit son expertise pour guider les clients dans leurs décisions d'investissement en fonction de leurs objectifs et de leur tolérance au risque.
- Discretionary Portfolio Management : AXA IM Architas offre des services de gestion discrétionnaire de portefeuille, permettant aux investisseurs de déléguer la prise de décision à des gestionnaires expérimentés pour une approche plus personnalisée.

Dans l'ensemble, AXA IM Architas se positionne comme un partenaire complet pour les investisseurs, offrant une combinaison de services qui va de la constitution de portefeuilles diversifiés à des solutions personnalisées et à la gestion discrétionnaire.

### **C. Politique RSE d'AXA Investment Managers :**

La Responsabilité Sociétale des Entreprises (RSE) est définie comme la contribution volontaire des entreprises aux enjeux du développement durable. Pour la Commission européenne, il s'agit de l'intégration volontaire par les entreprises de préoccupations sociales et environnementales à leurs activités commerciales et à leurs relations avec les parties prenantes.

AXA Investment Managers (AXA IM) illustre cet engagement en poursuivant une volonté forte de placer la RSE au cœur de sa stratégie. La société s'engage activement pour et avec ses collaborateurs à travers l'association du groupe AXA Atout Cœur, qui mobilise plus de 28 350 actes de bénévolat en 2019, couvrant des domaines tels que l'exclusion, la prévention des risques et la protection de l'environnement.

AXA IM excelle dans l'Investissement Responsable, avec plus de 20 ans d'expérience. Ses spécialistes dédiés influencent l'ensemble des classes d'actifs sous gestion, intégrant les critères et enjeux ESG (Environnementaux, Sociaux et de Gouvernance) dans les processus d'investissement en mettant à disposition de ces gérants de portefeuille et analyste financier un ensemble de données ESG.

L'approche de AXA IM, en termes de IR :

- ☐ Éviter tous les actifs qui ne respectent pas les critères de IR
- ☐ Identifier les risques et les opportunités en lien avec les ESG
- ☐ Investir en s'appuyant sur les critères ESG
- ☐ Impacter positivement directement la société et/ou l'environnement
- ☐ Influencer les états et les entreprises pour promouvoir la durabilité des décisions

Dans son propre fonctionnement, AXA IM adopte un cadre de responsabilité d'entreprise reflétant ses priorités environnementales, sociales et de gouvernance. Le respect de critères ESG est au cœur de ses processus, que ce soit dans la gestion des émissions de CO2 ou dans la protection des droits de ses collaborateurs.

Au-delà des performances financières, AXA IM s'engage activement dans des actions à enjeux durables, la société vise à devenir « Net Zéro » d'ici 2050 et investit dans des marchés verts, socialement responsables et durables. Cette approche active démontre une volonté de construire une économie durable et résiliente.

## **D. Financial Data Transparency :**

### **1. Enjeux du projet**

Le projet Financial Data Transparency repose sur plusieurs grands enjeux répartis en trois grands axes de travail, communément appelés workstreams.

La principale features du projet Financial *Data Transparency* est la mise en place d'un sourcing, en autre termes récupères de les données de façons automatique depuis plusieurs sources externes de données telles que des cubes multidimensionnels, ou de la donnés depuis des outils tierces tel Adjuto et Dali asset. L'objectif du projet est de centraliser au sein d'une même plateforme, la Data Platform toutes la donnés issue de ces sources.

Cela permettra à AXA IM d'améliorer la qualité de la gestion de ses actifs, car toutes les données seront centralisées en un seul endroit. La mise en place de cette centralisation permettra également d'atteindre une granularité accrue et de fournir une source de données de qualité supérieure pour tout ce qui concerne les besoins financiers.

L'un des autres grande features de ce projet est la mise en place d'un nouveau système de reporting, le but est de remplacer un ancien système vieux de plus de 7 ans. Cela permettra l'amélioration de la qualité des reportings, la mise en place de nouveau reporting plus complexe, ainsi que la génération de reporting de manière plus régulière, plus granulaire et plus précise, étant donné que le processus sera géré automatiquement se basera sur les sources gold mises en place en amont.

Une autre volonté du projet est de fournir un système de reporting dit dynamique ad hoc qui permettra à l'ensemble des métiers d'interagir de manière totalement libre et flexible avec les données.

### **2. Financial Dashboard**

Le Financial Dashboard est une collection de reports tel que des graphiques et ou tableau qui seront utilisés par toutes les business unit d'AXA IM, présenté en amont et pour la prise de décisions stratégiques. Ce reporting peut aussi être utilise par des commerciaux pour présentes AXA IM au client potentielle.

Il comprend des tableaux de bord tels que la répartition entre les actifs sous gestion et les actifs en distribution, ainsi que d'autres informations relatives aux performances par pays, ou encore des indicateurs plus fondamentaux tels que les dix meilleurs portefeuilles de la compagnie.



Traditionnellement construit et fourni par l'équipe financière chaque mois, ce tableau de bord avait une durée de réalisation d'environ deux semaines. Aujourd'hui, grâce au projet Data Transparency, il peut être généré en moins de cinq minutes.

Une autre valeur ajoutée du projet réside dans la granularité plus détaillée dont les contrôleurs financiers ne disposaient pas auparavant, offrant ainsi à l'équipe de contrôle financier un niveau de détail plus approfondi.

(Photo du Dashboard qlq graphique)

## **E. Présentation de la mission de stage :**

### **1. Définition et enjeux de la mission :**

La mission de mon stage de fin d'études visait principalement à contribuer en tant que Data Engineer à diverses missions au sein de l'équipe, tout en intégrant une perspective métier. Cette approche métier impliquait une compréhension approfondie des besoins métiers pour répondre de manière précise à leurs exigences.

Ma mission principale en tant que Data Engineer chez AXA IM consistait à mettre en place des pipelines d'ingestion et de distribution de données au sein de la Data Platform. Ces pipelines étaient conçus pour répondre aux besoins métiers, notamment dans la gestion de toutes les données externes à la dataplatform, qu'il s'agisse de fichiers métiers à mettre à disposition dans la dataplatform ou de données historiques à ingérer issues d'anciens systèmes qui seront supprimés.

Les données ainsi traitées permettaient le calcul d'un grand nombre d'indicateurs utilisés ultérieurement dans le reporting, principalement dans le Financial Dashboard.

Une autre mission à fort impact durant mon stage était la veille technologique, dans le but de répondre à divers besoins du métier, tels que la mise en place de l'anonymisation des données sensibles liées au client suisse à la suite des demandes des clients. Cela incluait également d'autres besoins tels que la mise en place d'une technologie pour répondre au besoin d'extraction de la donnée depuis la dataplatform.

Ces missions liées à la veille technologique vont ainsi permettre à l'équipe finance de répondre à des besoins spécifiques du métier grâce à de nouvelles technologies, tout en restant toujours à jour au niveau des dernières nouveautés en termes de Big Data.

Un troisième volet de mon stage sera laissé à mon choix, me permettant de découvrir plus en détail le métier de data analyst.

### **2. Gestion de projet – Méthode Agile**

Pour mieux comprendre mon rôle au sein de l'équipe et la gestion global du projet, je vais vous présenter l'ensemble des collaborateurs que j'ai côtoyés au quotidien ainsi que leurs différents rôles au sein du projet et de l'équipe.

Pendant mon stage, j'ai principalement évolué au sein du Workstream 1 et 2. Les membres de l'équipe ont tous des profils différents qui se complètent, ce qui permet d'avoir une bonne ambiance de travail et de faire avancer le projet.

**Bernard Eidel**, Project Manager du projet Data Transparency, chargé de définir la vision globale du projet et de coordonner l'ensemble des parties prenantes, qu'elles soient techniques, métiers des différents Product Owner.

**Ruddy Haddad**, mon tuteur de stage, il occupe principalement le rôle de Data Analyst et Data Engineer. En tant que Product Owner des Workstreams 1 et 2, il a une vision claire du produit final à livrer et coordonne les différents membres travaillant sur ces workstreams à travers la planification et l'établissement du cahier des charges.

**Pierre NGUYEN BUU** : Application Analyst, ayant de solide connaissant métiers, il est principalement en chargé de gérer la décommissions des cubes multidimensionnels qui seront remplacé par la dataplatform, sa compréhension du monde financier permet à l'équipe de facilité la mise en œuvre technique des différents besoin métiers

**Gregory BESSAU** : Technical Leader, il évoluer au sein de plusieurs équipes, et joue le rôle de référent technique pour la mise en place du produit, il veille au respect des normes définies par AXA IM

J'ai également côtoyé au quotidien avec d'autres membres du projet, notamment dans le Workstream 3, principalement **Mohamed EL MABROUK** et **Taha KHARROUBI**, tous deux Software Engineer.

Nous avons évolué au sein d'une méthodologie de type agile, avec Mohamed EL MABROUK en tant que Scrum Master. Il animait quotidiennement des réunions (daily meetings) permettant à chaque membre de l'équipe de présenter ses avancements de la veille et son programme pour la journée aux travers des user stories qui on était assigné à chacun. Les user stories sont de petites missions auxquelles on attribue des story points, avec un ratio de 1 story point équivalent à 1 jour de travail.

L'équipe évoluait ainsi dans une dynamique de sprint de 2 semaines, avec un sprint review réalisée avec l'ensemble de l'équipe technique finance, mais aussi avec l'ensemble de l'équipe métiers. Cela permettait aux métiers de voir l'avancement du projet, de faire des remarques et des critiques dans le but d'éviter l'effet tunnel.

Ainsi, au travers de cette évolution en sprint, chaque sprint était préalablement défini avec les missions à accomplir. Les user stories, combinées entre elles, permettaient la mise en place d'une features, et l'assemblage de différentes features permettait d'obtenir le produit final.

### 3. Planification prévisionnelle du projet

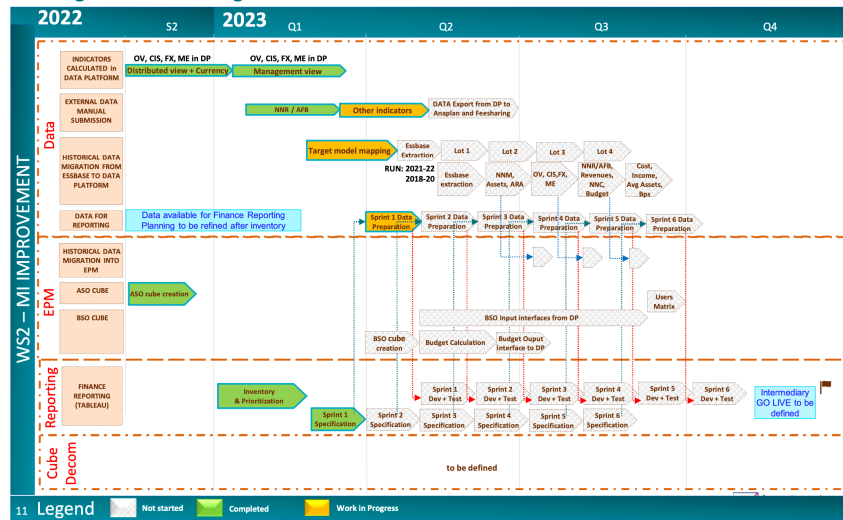
Chaque mois, un comité de suivi est tenu et présenté par le product manager et les product owners face à un comité composé des différentes parties prenantes, ainsi que des heads de la section dans laquelle nous évoluons.

Le planning est en constante évolution, en fonction des nouveaux besoins définis par les métiers, mais aussi de l'importance d'une features par rapport à une autre.

Le planning prévisionnel est ainsi converti par le product manager et les product owner en un ensemble de user stories et de features définies dans un back log, puis elles sont assignées avant chaque sprint lors du sprint planning au membre de l'équipe le mieux positionné pour leur réalisation.

## WS2 – MI Improvement

### High-Level Planning



C'est le planning prévisionnel du Workstream 2 pour la présentation au comité du mois de mars. On peut ainsi y voir les features et l'état d'avancement pour chaque tâche ainsi que leurs enchaînements.

Ainsi, dès mon arrivée au sein de l'équipe en juillet 2023 et en vue d'un deadline fixée pour la mise en production vers la fin d'octobre, mon planning prévisionnel était déjà préparé et convenu avec mon tuteur de stage ainsi que le product manager.

Au cours de cette période, près de 6 sprints de 2 semaines étaient représentés, m'amenant à la première moitié de mon stage. Nous verrons ainsi plus en détail le contenu de mes sprints dans la section des cahiers des charges, mais pour vous donner une vision globale, j'ai eu à ma charge avec Pierre Nguyen, l'Application Analyste dans la partie data, les features :

- External Data Manual Submission.
- Historical Data Migration from Essbase to Data Platform.

Ainsi, mon travail repose sur l'industrialisation de l'ensemble du travail réalisé par Pierre en amont. Pierre, ayant une connaissance plus approfondie du métier, effectue des contrôles financiers, applique les règles définies par le business, et effectue des corrections sur des données issues des cubes de production.

## 4. Cahier des charges

Pour mieux définir le cahier des charges qui m'est confié je vais vous expliquer plus en détails le contenu d'une user story.

Une user story a 3 composante principale,

- Un titre.
- Une description.
- Un critère d'acceptation.

La description me donner ainsi une idée de la tâche à réaliser, mais en fonction des tâches qui m'ont été confiées on est souvent amené à aller chercher les informations nécessaires auprès des métiers.

Le critère d'acceptation représente ainsi mon cahier de charge personnel, et nous allons ainsi voir en détails, pour chaque sprint, le cahier des charges qui a été défini.

Filtering 1000 out of 1794 items. Filter the next 794 items									
Order	Work Item Type	Title	State	Effort	Busin...	Value Area	Tags		
	User Story	Export data from DP to REAL ASSETS - Industrialization in DP	*** Closed			Business	MVP WS2		
	User Story	HISTORIC - Industrialize the silver to gold data from FMR/FML/FMLC	Closed			Business			
	User Story	HISTORIC - Industrialize the silver to silver technic data from FMR/FML/FMLC - Testing	Closed			Business	WS2		
	User Story	HISTORIC - Industrialize the raw to silver data from FMR/FML/FMLC	Closed			Business			
	User Story	FIX Industrialization - HISTORY - ESSBASE FMR Loader - Sep 22 - Ingestion in DP - Fix scope NNM/MF/France - Copy	Closed			Business	MVP WS2		
	User Story	MANUAL FILE - DISTRIBUTED REVENUES - Automate Transformation - Silver to Gold	Closed			Business	WS2		
	User Story	NNR MANUAL - Target - Transformation Silver to Gold	Closed			Business	MVP WS2		
	User Story	MANUAL FILE - DISTRIBUTED REVENUES - Automate Ingestion in the DP - Raw/Bronze/Silver	Closed			Business	WS2		
	User Story	ANAPLAN NNR - Automate Transformation Silver to Gold - Histo	Closed			Business	MVP WS2		
	User Story	NNR MANUAL - Target - Ingestion Raw to Silver	Closed			Business	MVP WS2		
	User Story	ANAPLAN NNR - Automate Ingestion in DP Raw/silver - Histo	Closed			Business	MVP WS2		
	User Story	EPM MGT - COST/CASH BALANCE/REVENUES - Industrialize Transformation Silver to Silver Reporting	Closed			Business	WS2		
	User Story	Repos - Finance Reporting - Migrate to Unity Catalog - Cascade Table	Closed			Business	WS2		
	User Story	BUDGET - Industrialize Transformation in DP Bronze to Silver to Gold	Closed			Business			
	User Story	BUDGET - Industrialize Ingestion in DP Raw/Bronze	Closed			Business	EPM WS2		
	User Story	EPM MGT - COST/CASH BALANCE/REVENUES - Industrialize Ingestion in DP Raw/Silver	Closed			Business			
	User Story	Repos - Finance Reporting - Migrate to Unity Catalog - Other table (histo/epm)	Closed			Business	WS2		
	User Story	Distributed Revenues - Silver Reporting to Gold Reporting	Closed			Business	MVP WS2		
	User Story	NNR ANAPLAN - Silver Reporting to Gold Reporting process - Target	Closed			Business	MVP WS1 WS2		
	User Story	NNR ANAPLAN - Gold to Reporting Gold process - Historic data	Closed			Business	MVP WS1 WS2		

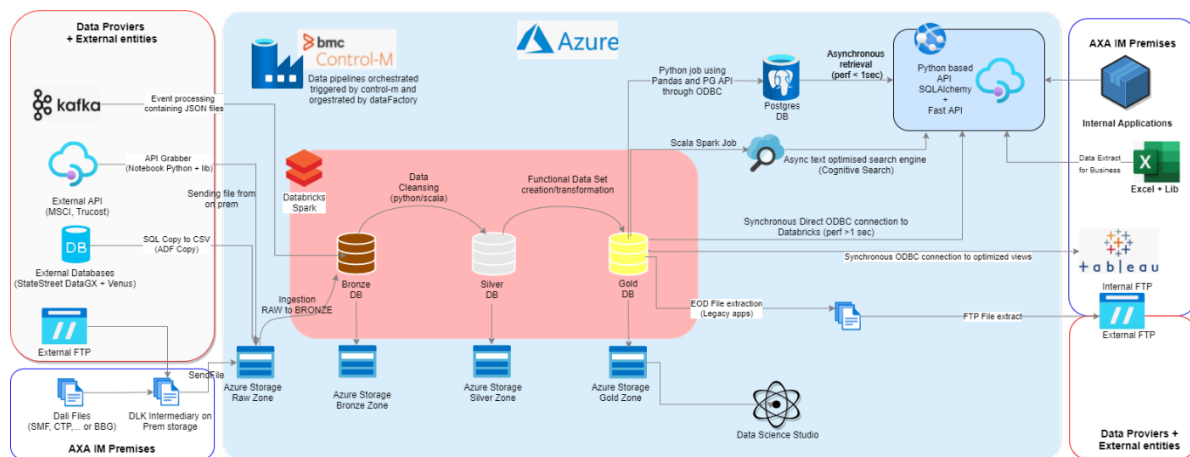
### III. Réalisation des missions

#### A. Aspect technique global de mes missions :

Pendant mon stage de fin d'études au sein d'AXA IM et dans le cadre du projet Finance Data Transparency, j'ai eu l'opportunité d'utiliser un ensemble de compétences techniques acquises auparavant. J'ai également eu à utiliser divers outils, que je vais détailler dans cette section.

##### 1. La dataplatform

La dataplatform permet à l'équipe de AXA IM de dispose des capacités de traitement des données. Elle permet faciliter toutes les opérations liées à la gestion des données, couvrant ainsi les connecteurs vers différents types de sources de données pour faciliter l'ingestion de données, une certaine capacité de transformation, ainsi qu'un endroit pour interroger les données et les visualiser.



La nouvelle plateforme de données d'AXA IM a été mise en place pour répondre à un besoin croissant, causé par plusieurs lacunes des anciennes plateformes.

Parmi celles-ci :

1. La duplication des flux de données : chaque application client s'appuie sur un sous-ensemble de Warehouse, copiant les données dans sa propre base de données et effectuant des transformations. D'autres applications client, basées sur les mêmes sources de données, copieront eux aussi les mêmes données, effectuant les mêmes transformations causant ainsi une duplication des flux de données
2. Manque de ownership de la data : comme la création des données et leur consommation finale sont très éloignées, la propriété se dilue dans le processus.

Ce sont les principaux points de faiblesse par conséquent entraînent des coûts de maintenance et des risques potentiels, car si un problème survient dans un warehouse cela peut déclencher des incidents dans les applications suivantes quelques jours après.

Pour palier à ces soucis, AXA IM, c'est tourné vers la solution d'utiliser databricks, Cette solution est, représentée dans le schéma au-dessus avec différentes couches (raw, bronze, silver, gold) elle permet de répondre aux différents besoins de manière plus efficace.



Databricks se présente comme une plateforme Lakehouse, conçue pour unifier les données au sein de l'entreprise, simplifiant ainsi leur exploitation. Elle se positionne entre les data lakes et les data warehouses, facilitant la gestion des données pour axa im cette l'architecture permet à la dataplatform d'en tirer grand nombre d'avantages :

- Connecteurs de données intégrés
- Template de pipeline d'ingestion et de distribution en python et scala
- Un accès SQL aux données
- L'organisation des données sous forme de catalogue
- Des outils de visualisation tel que Tableau, Lakeview Dashboard, powerbi
- Une distribution de données via des API
- Une facilitation de l'accès aux données

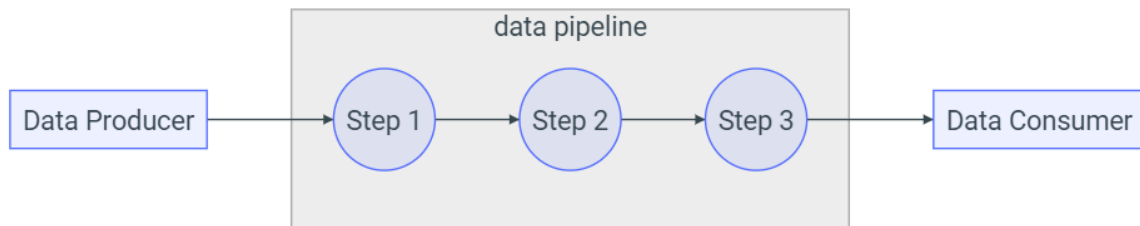
Cette nouvelle architecture permet aussi une concentration différentes sources de données dans un seul référentiel, ce qui permet de une mutualisation des transformations de une mutualisation des contrôles de qualité des données (effectuer les transformation et les contrôle une fois pour tous les utilisateurs de la Dataplatforme)

## 2. Data flows

Un pipeline de données est constitué de toutes les étapes allant d'un producteur de données à un consommateur de données.

Le cœur du pipeline de données consiste généralement en une étape de traitement des données en mode ETL (Extract Transform Load) ou ELT (Extract Load Transform).

- ETL : (Extract - transform - Load)
- ELT : (Extract - Load - transform)



Au sein de la DP, le modèle utilisé est l'EtLT (Extract - transform - Load - Transform).

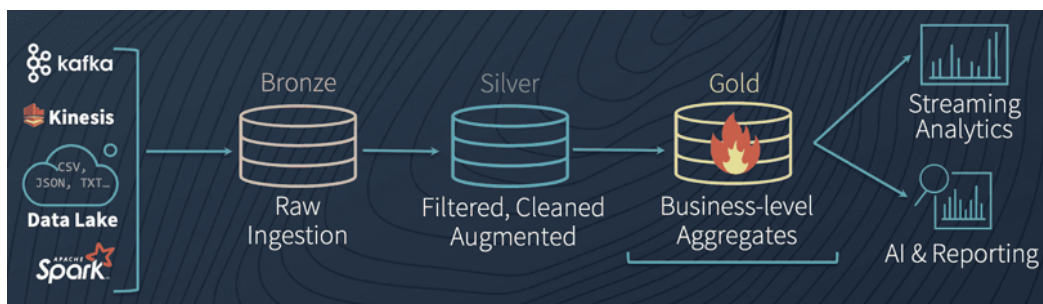
Le concept est le suivant, l'idée est de gérer d'abord les petites transformations et de réserver les transformations plus complexes pour la fin du processus.

La première transformation est appliquée à une seule source à la fois. Ces transformations sont rapides et simples car elles modifient chaque source indépendamment des autres. Elles doivent également être génériques, car tout le monde s'appuiera sur le résultat pour construire les transformations ultérieures.

En détail, la première partie de la transformation peut comprendre :

- Format des données : analyse syntaxique
- Masquage / chiffrement / déchiffrement des données
- Nettoyage des données : qualité technique ou données limitées
- Modélisation des données : mise à jour / insertion - ajout - troncature / insertion

L'application de cette notion de EtLT a la databricks correspond a l'architecture standards représente dans ce schema





La couche brute (raw layer) :

- C'est l'étape d'extraction de notre processus.
- La couche brute stockera une copie des données provenant du fournisseur.

La couche bronze :

- Il s'agit de notre niveau de transformation légère
- La couche bronze est une étape technique destinée à gérer l'analyse syntaxique, la qualité des données, le masquage et d'autres opérations par source. Le petit t

La couche argent (silver layer)

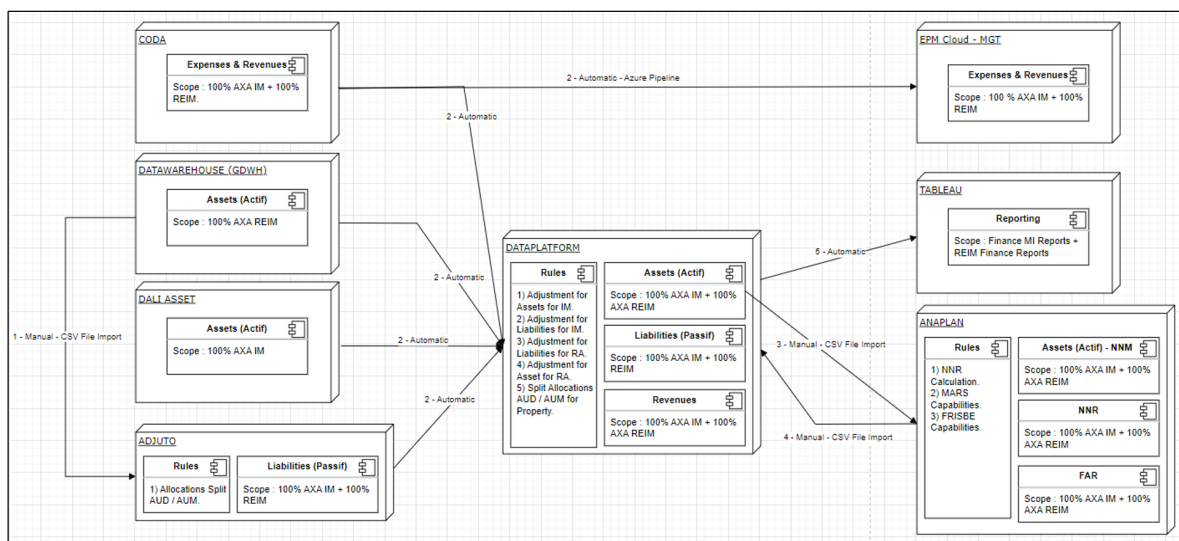
- Principalement utilisée pour stocker les données par source à la fin des transformations légères.
- Peut également servir à stocker des modèles techniques non conçus pour être distribués comme des modèles métier.

La couche or (gold layer)

- Intervient après la phase ultérieure de transformation.
- Destinée à stocker des modèles métier finaux et fortement transformés.

## B. Technical aspect of Data Transparency

### 1. Schéma architecture



Le schéma d'architecture cible qui sera livré à la fin du projet est le suivant. Au cœur de ce projet, on retrouve la dataplateforme. Le projet Data Transparency vise donc à centraliser un grand nombre de données issues de plusieurs sources dans le but d'améliorer la qualité des données financières et de permettre une meilleure gestion des actifs tout en utilisant pleinement la puissance de la dataplateforme.

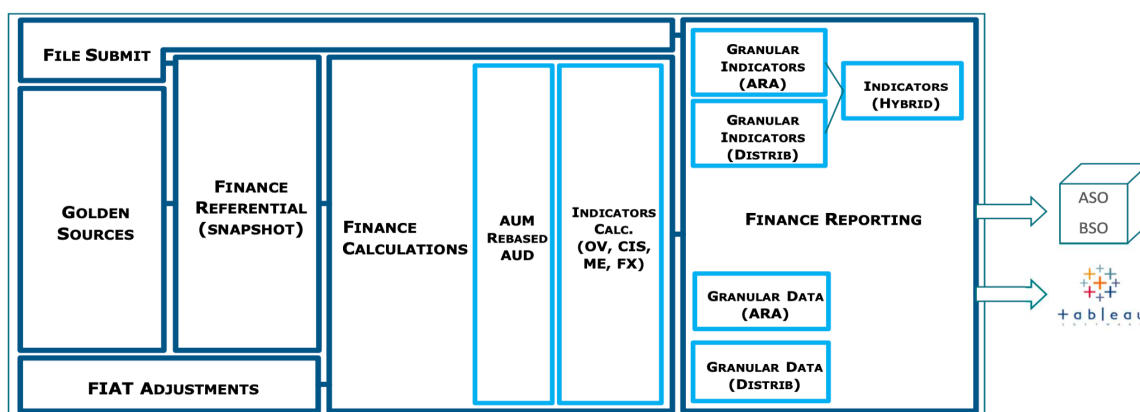
Les principales sources utilisées par ce projet sont des sources Gold, issues de pipelines externes au projet, qui nous fournissent des données telles que Dali Asset pour les positions et Adjuto pour les transactions. Les données peuvent aussi venir de systèmes externes tels que Coda. Des pipelines d'ingestion sont ainsi mis en place pour acheminer l'ensemble de ces sources de données vers la dataplateforme et les y centraliser.

Une fois ces données centralisées au sein de la dataplatforme, le projet Data Transparency donne la main aux métiers pour effectuer ce que l'on appelle des ajustements via l'outil FIAT. Une fois ces ajustements effectués principalement sur les positions ou les transactions et que ces données sont déployées, elles peuvent ensuite être distribuées à d'autres services.

Maintenant que ces données sont centralisées et harmonisées sur la dataplatforme, elles peuvent être distribuées à plusieurs clients. Une partie de ces données est distribuée au système Anaplan. Cet outil Anaplan reçoit ainsi des données depuis la dataplatforme et procède au calcul d'un ensemble d'indicateurs. Une fois ces indicateurs calculés, ils sont de nouveau ingérés dans la dataplatforme pour les rendre disponibles à l'exploitation.

Maintenant que ces données financières sont centralisées, elles peuvent être utilisées à des fins diverses telles que l'extraction, l'audit et surtout pour le reporting.

Une vision plus centrée sur le reporting

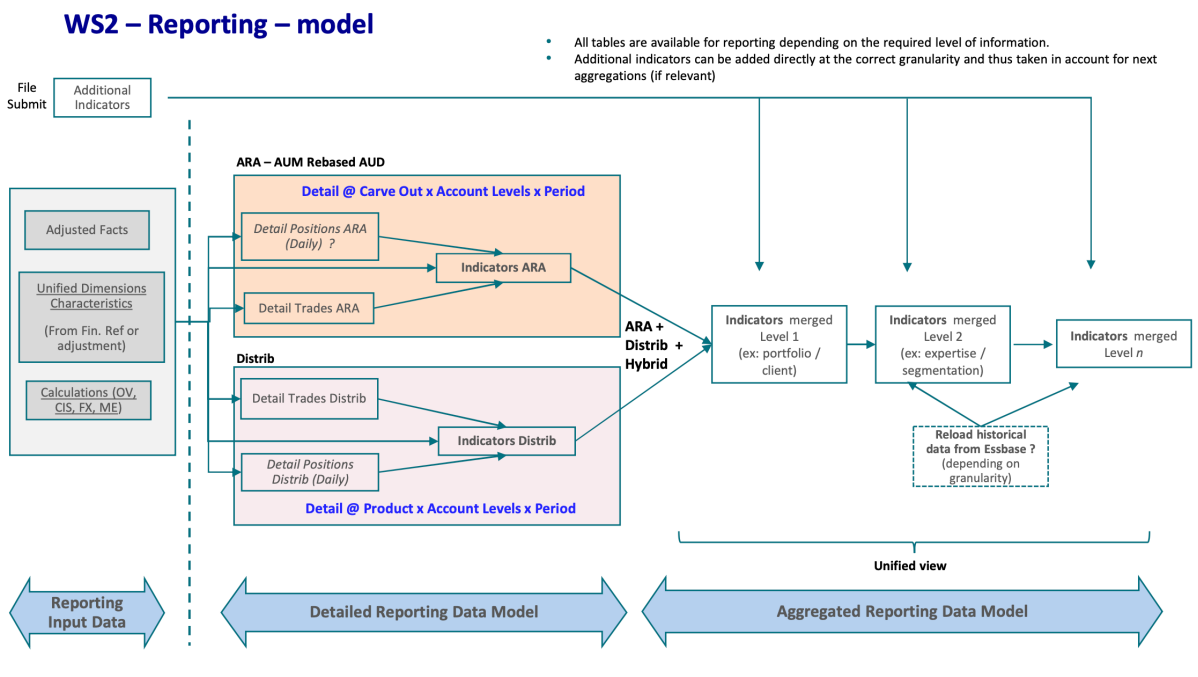


Le reporting du financial dashboard défini auparavant se base ainsi sur un datamodel que nous verrons dans la section suivante. Ce data model permet ainsi de combiner et d'agréger les données dans un ensemble de tables permettant de rendre les données accessibles pour tous.

Toutes les connexions qui permettent de lier les données externes à la dataplatforme reposent sur des pipelines d'ingestion que l'on verra plus tard étant la principale composante de mes missions de stage,

Et l'ensemble des pipelines utilisés au sein même de la dataplatforme pour transformer les données y applique l'ensemble des règles fonctionnelles sont des pipelines de distribution.

## 2. Data procession for Dashboard



Le Financial Dashboard est une véritable finalité du projet, car il permet de rendre tout le travail effectué un peu plus tangible.

Maintenant que toutes les données sont disponibles au sein de la dataplatforme, il faut mettre en place un data model pour pouvoir les distribuer, en particulier pour le reporting input data.

La modélisation des données est le processus de cartographie d'un système d'information et de représentation des connexions entre plusieurs tables. Les modèles de données sont généralement illustrés dans un diagramme entité-relation pour les bases de données relationnelles.

Le Details Reporting Model est composé de deux visions importantes à explorer : la vision distribution et la vision management.

La vision distribution correspond à la vision qui reflète l'AUM (Asset Under Management) en distribution. En d'autres termes, c'est une vision qui permet de savoir quels clients ont investi dans quels portefeuilles, ainsi que les transactions en cours. La vision management, quant à elle, permet de savoir où l'argent est investi.

Le carveout est une notion importante. Un client achète un placement au sein d'un produit, mais ce produit peut être de type fixed income ou multi asset. Dans le cas du multi-asset, l'argent sera donc investi dans différents autres produits tels que le real estate, le fixed income. Cette reallocation permet d'avoir une vision de gestion et est gérée par un algorithme ARA AUM rebased aud.

Ces différentes tables sont ensuite agrégées dans une vision hybride, qui permet ainsi d'avoir une meilleure vision pour le reporting. Pour cela, on élimine certaines informations telles que le portefeuille et le carveout, laissant seulement le produit en fonction de diverses granularités.

Au niveau 1, portefeuille, client, au sein de ces tables de reporting agrégé, d'autres données externes viennent s'ajouter, occupant ainsi mes deux principales missions de stage : les données historiques depuis Essbase et les données externes qui sont un ensemble d'indicateurs issus du business et délivrées par le processus Filesubmit.

### **C. Essbase – Historical Data Industrialization**

Comme vu précédemment, les données historiques d'Essbase doivent être ajoutées au niveau des tables d'agrégation de reporting.

Essbase est un système qui sera décommissionné début 2024, mais pour des besoins de reporting, spécialement pour le type de graphique "5 YTD", j'ai été chargé de mettre en place des pipelines d'ingestion et de distribution pour assimiler 7 ans de données issues de ce cube.

Pierre Nguyen m'a ainsi fourni les fichiers CSV que nous avons déposés dans un conteneur dans le cloud Azure.

La mise en place de la pipeline d'ingestion et de distribution est la suivante :  
Notre objectif est de charger l'ensemble de ces données au sein de la dataplatform de manière automatique, puis de les formater et de les transformer pour s'intégrer au datamodel.

La première étape de cette pipeline est l'ingestion, qui est gérée par l'outil interne de la dataplatforme, l'autoloader. Via des fichiers de configuration en YAML, l'outil autoloader sait quels sont les fichiers CSV à lire.

La task 1 va donc simplement prendre les fichiers CSV tels qu'ils sont ingérés et les stocker au niveau du bronze layer.

La complexité technique repose sur la bonne configuration du fichier et sur la compréhension de cet outil, assez peu compris au sein de l'équipe. Grâce à mes compétences, j'ai rapidement pris en main cet outil utilisé pour l'ingestion de données.

Une fois que les données sont chargées au niveau du bronze layer, au sein de la database...

La deuxième task est le passage du layer bronze au silver.

Ce passage va comprendre un ensemble de légères modifications, principalement des modifications de type structurel par rapport au nom des colonnes et des types. Il peut aussi y avoir des corrections sur des données de prod, telles que des renommages à effectuer sur ces données issues de la ligne.

La troisième task : silver to silver-reporting.

Lors de cette étape, un ensemble de modifications est effectué. On effectue des mappings pour récupérer des dimensions manquantes, on applique des calculs par rapport au taux de conversion sur les actifs, on fait aussi des modifications au niveau du client level intermédiaire final. Tout cela dans le but de s'aligner parfaitement au sein du datamodel défini pour le reporting.

Cette mission était la première qui m'a vraiment plongé dans le bain, car je suis parti d'un nouveau repo et j'ai ainsi mis en œuvre l'ensemble de l'architecture générale des pipelines.

Trois cubes différents étaient gérés par Essbase, FMLH/FMR/FML, et étaient liés à cette historique des tables de mappings. Les tables de mappings permettent ainsi, en ayant quelques informations telles que des clés primaires, d'obtenir des caractéristiques d'un élément. Par exemple, au travers du `portfolio_id`, nous pouvons ainsi récupérer l'ensemble des caractéristiques du portefeuille depuis ces tables de mappings.

#### D. Financial - Custom inputs Industrialisation

Pour cette mission, la problématique était simple : comment permettre aux métiers de fournir un ensemble d'indicateurs depuis des fichiers externes à la data plateforme et que ces fichiers, ainsi que via un seul reporting et un seul catalogue de données, nous devions trouver la solution pour gérer des pipelines d'ingestion et de distribution totalement indépendantes les unes des autres.

Cette mission m'a ainsi poussé à développer des compétences très poussées au niveau de Databricks et de la gestion des outils associés.

Le répertoire Finance Custom Inputs permet ainsi la gestion de 5 indicateurs différents :

- La donnée historique d'Anaplan
- Le NNR
- EPM MGT
- Distributed Revenue
- Le budget

☆	ingestion_financebusiness_custominputs_budget_v1	DLK_Ingestion.FinanceBusiness.CustomInputs	Finance	f536f424-e816-4a6f-	🟢🟢🟢🟢🟢▶⋮
☆	ingestion_financebusiness_custominputs_epm_v1	DLK_Ingestion.FinanceBusiness.CustomInputs	Finance	f536f424-e816-4a6f-	🟢🟢🟢🟢🟢▶⋮
☆	ingestion_financebusiness_custominputs_histo_anaplan_v1	DLK_Ingestion.FinanceBusiness.CustomInputs	Finance	f536f424-e816-4a6f-	🟢🟢🟢🟢🟢▶⋮
☆	ingestion_financebusiness_custominputs_nnr_indicators_v1	DLK_Ingestion.FinanceBusiness.CustomInputs	Finance	f536f424-e816-4a6f-	🟢🔴🔴🟢🟢▶⋮
☆	ingestion_financebusiness_custominputs_revenue_indicators_v1	DLK_Ingestion.FinanceBusiness.CustomInputs	Finance	f536f424-e816-4a6f-	🟢🟢🔴🟢🟢▶⋮

L'ensemble de ces indicateurs est ainsi fourni par les métiers via des fichiers de template définis en concertation.

Cela va ainsi suivre le même processus cité précédemment. Pour chaque indicateur, nous allons suivre les étapes suivantes :

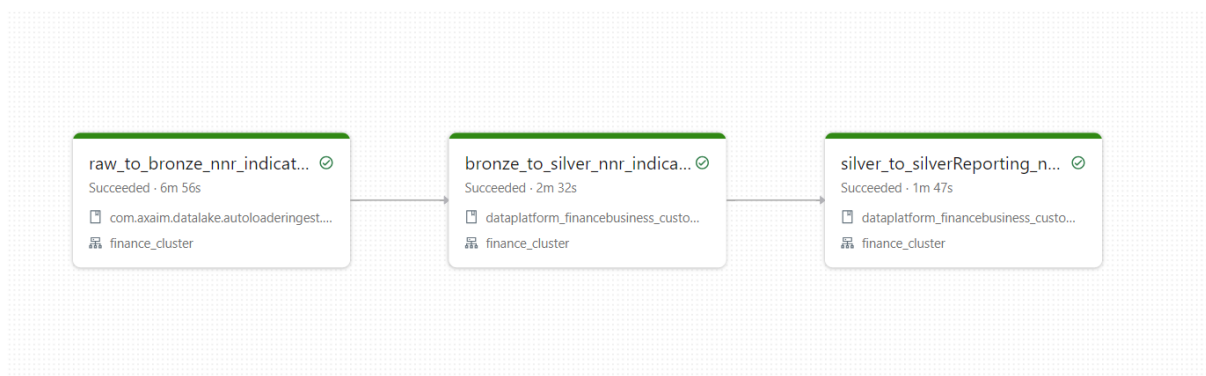


## Raw, Bronze, Gold

Une fois que ces 5 pipelines sont mises en place, pour vraiment gérer le processus de manière automatique, nous allons utiliser un outil appelé Filesubmit.

Filesubmit est un outil développé en interne qui permet aux équipes métiers de déposer des fichiers de différents types au sein de notre conteneur iCloud. Cet outil permet également, via un job Contrôle M, de lancer automatiquement les pipelines d'ingestion.

De ce fait et de manière automatique grâce à mon travail, les équipes métiers peuvent déposer chaque jour, chaque semaine ou chaque mois un fichier respectant le format défini au préalable via le Filesubmit. Ce fichier de type CSV sera identifié, et via un job Contrôle M, il lancera le job Databricks correspondant à l'indicateur en question, le tout de manière automatique. Ce fichier sera alors intégré au sein de la dataplatforme, subira un ensemble de transformations et de formatage, puis sera de nouveau acheminé de manière automatique vers une table de reporting. Cette table sera ensuite automatiquement prise en compte dans les rapports de Tableau.



Pour aller dans le details nous allons ainsi commenter le code utilise pour la task silver silver reporting

## 1. Présentation des Template d'ingestion :

La structure d'une template d'ingestion utilisée par AXA IM représente ainsi la disposition suivante. L'utilisation de ces templates vise à standardiser et faciliter la maintenance de diverses pipelines de données développées par les équipes centrales d'AXA IM.

Le dossier "metadata" contient l'ensemble des fichiers de configuration utilisés par l'autoloader. Ces fichiers de configuration permettent à l'autoloader de savoir quels fichiers CSV il doit lire. Un fichier de metadata se comporte ainsi de la manière suivante :

L'autre dossier présent est le dossier d'orchestration. Ce dossier permet de créer la pipeline d'ingestion et de distribution. On y définit ainsi l'enchaînement des tâches et quel élément doit être appelé pour chaque tâche. Une tâche peut appeler un script Python ou un JAR développé en Scala. D'autres paramètres peuvent être définis dans ce fichier, tels que le cluster qui sera utilisé, la version, mais aussi les dépendances entre les différentes tâches, comme suit :

On trouve également le dossier spécifique à notre cas, "financecustominputs". Ce dossier contient l'ensemble du code que j'ai pu réaliser, composant ainsi les scripts qui peuvent être appelés au sein de nos workflows sur Databricks. Grâce à ce template, ce code sera compilé en un package .wheel que l'on pourra appeler.

## 2. Présentation des transformation silver to silver reporting :

## **E.     Anonymisation of the data**

Un besoin pressant du métier était la possibilité d'anonymiser les données pour les clients suisses. Le cahier des charges était simple : lorsque l'entité correspond à Zurich, le nom du client doit être anonymisé en "client suisse".

Pour répondre à cette problématique, l'équipe en charge de la dataplatform a fortement recommandé à notre équipe de migrer vers un nouveau type de catalog sur Databricks : Unity Catalog. Le principal avantage d'Unity Catalog réside dans ses fonctionnalités centralisées de contrôle d'accès, d'audit, de traçabilité et de découverte des données dans les espaces de travail Azure Databricks. Unity Catalog est fortement mis en avant par Databricks, car toutes les futures nouveautés de Databricks seront uniquement supportées par Unity Catalog.

Parmi ces nouveautés, on retrouve les fonctions de masquage et de filtrage des données. J'ai ainsi été en charge de tout cela. J'ai commencé tout d'abord par faire un proof of concept sur l'une des tables Unity Catalog du catalog financecustominputs.

Le proof of concept était le suivant : en fonction du groupe d'appartenance d'une personne, cette personne pouvait ou non voir l'information. Ainsi, j'avais à ma disposition deux groupes, le groupe "user" dont je faisais partie et le groupe "admin". J'ai appliqué l'anonymisation sur une table, et l'anonymisation a parfaitement marché au niveau de la table.

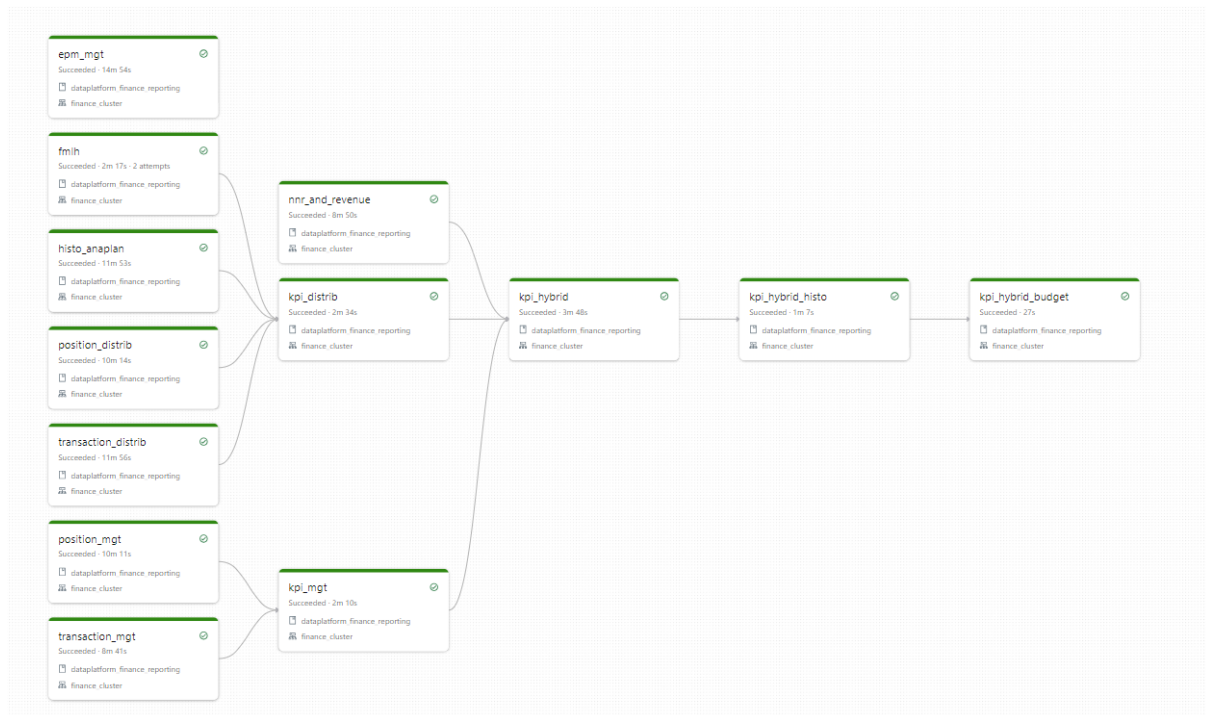
Mais pour réaliser ce proof of concept, le chemin était long, car il fallait être sur les dernières versions proposées par Databricks. Ainsi, ma première partie de stage a consisté à être une beta-testeuse d'un cluster Databricks de la version 13.3, qui était utilisé par AXA IM pour toutes les pipelines d'ingestion, alors que le runtime était toujours sur la version 10.2.2. Cela a déjà constitué une première étape de mon stage : tester et déployer des pipelines sur une dernière version de cluster de Databricks, ainsi que mettre à jour la version de Python en 3.10. Mon travail permet ainsi à AXA d'être à jour et à la pointe des dernières technologies sur Databricks.

Une fois le proof of concept réalisé et présenté aux membres de l'équipe et validé pour le business, le plus gros du travail reste à faire. J'ai ainsi été en charge de la migration de la pipeline de distribution s'occupant du reporting, car l'anonymisation devait intervenir à ce niveau. La migration vers Unity Catalog est alors intervenue.

Je mets ci-dessous les requêtes SQL pour des captures d'écran, ainsi que le code SQL utilisé pour le masquage, avec explications également.



## F. Finance Reporting to unity catalog



Si vous êtes attentive, vous pourrez remarquer que ce workflow peut être positionné au datamodel du reporting expliqué précédemment. L'une des étapes les plus difficiles à réaliser pendant mon stage était le fait de migrer ce workflow qui représente le travail de plus d'un an de code et de tests vers une nouvelle technologie, Unity Catalog.

Cette migration vers Unity Catalog se divise en plusieurs grandes parties. Pour effectuer cette migration, nous sommes obligés d'utiliser une bibliothèque mise en place par la team core d'AXA IM, la bibliothèque AXA-IM PySpark.

La migration vers Unity Catalog a commencé par une étape de sourçage. Je devais sourcer l'ensemble des tables utilisées par le reporting et les référencer au sein d'une classe, CatalogRepository.

Code ici

Cette étape permet ainsi de centraliser et de référencer l'ensemble des tables qui peuvent être utilisées pour le reporting. Ces tables peuvent être les fameuses golden sources, provenant des sources Adjuto ou Dali PP. D'autres tables proviennent des tables finance-custom-inputs, mais aussi des tables de référentiel pour pouvoir effectuer des mappings et compléter les données.

Ensuite, l'autre problématique de cette migration était la vérification de la qualité des données. Pour chaque cut-off date, je devais ainsi réaliser des vérifications sur les données pour m'assurer que je conserve bien la donnée, que je ne duplique pas de lignes, ou que je n'arrondisse pas les montants étant trop grands. Le fait d'arrondir un chiffre peut représenter plusieurs dizaines de milliers d'euros de pertes.

Pour réaliser ces checks de la data sur les tables de reporting que je suis en train de migrer, j'ai ainsi pu utiliser mes compétences en SQL pour requêter rapidement la donnée et la vérifier. Mais à force de vérifier la somme, j'ai décidé de mettre en place une fonction qui me permet de comparer deux tables de données, de vérifier la somme des actifs, et de me renvoyer les lignes qui ne matchent pas. Cela me permet ainsi d'investiguer et de résoudre le problème. Il faut vraiment faire attention, cela m'a vraiment mis en avant mes compétences de data analyst, puisque je devais comprendre l'ensemble des règles business appliquées pour pouvoir retomber sur le même actif à chaque fois, ce qui devait être de l'ordre de 800 milliards d'euros entre chaque étape.

Une fois que j'ai réussi à migrer l'ensemble de ces tables de reporting vers Unity Catalog, nous avons ainsi pu appliquer l'anonymisation des clients suisses et migrer aussi l'ensemble du reporting vers Tableau.

Mon autre travail sur le reporting a été l'intégration des indicateurs suivants au sein des tables de reporting.

Prenons comme exemple l'intégration de "histo\_anaplan". Cette intégration repose ainsi sur le fait de faire arriver de la donnée ingérée et disponible sur les tables du finance-custom-histo-anaplan-reporting dans le reporting. Pour ce faire, j'ai dû ainsi transformer et prendre seulement les colonnes pour matcher avec le datamodel.

Le revenu, par exemple, j'ai dû sommer l'ensemble des indicateurs pour les faire descendre au niveau des tables de reporting.

La table la plus importante est la "kpi\_hybrid".

## **G. Export from DP**

Le projet Data Transparency va décommissionner un ensemble de systèmes tels qu'Essbase, mais d'autres systèmes seront maintenus, tels qu'Anaplan, Mars, Magnitude.

Le business a ainsi besoin d'extraire des données depuis la dataplatform, maintenant que tout est centralisé. Pour ce faire, j'ai réalisé un benchmark, car nous devons répondre à plusieurs problématiques telles que le nombre de lignes à extraire, les règles à appliquer, la facilité d'accès et la maintenance.

- Tableau
- FIAT
- Power BI
- Datalake View

La problématique ad hoc est également à considérer.

## IV. Bilan et perspectives

Dès le début du stage, mon tuteur m'a annoncé que ce serait un défi, et que je ne serais pas simplement un stagiaire, mais un membre à part entière de l'équipe. Rapidement, j'ai pu démontrer des compétences techniques solides, ce qui m'a donné l'opportunité d'apporter une valeur ajoutée à l'équipe. Actuellement, au sein de l'équipe finance, je suis devenu un collaborateur formé et spécialisé dans la dataplatform.

J'ai pris en main la dataplatform et contribué à son amélioration en testant de nouvelles technologies, maîtrisant ainsi en avant-première un ensemble de technologies au sein de mon équipe et dans toute AXA IM. Ma valeur ajoutée s'est concrétisée au niveau des KPI des plannings, en gérant entièrement les fonctionnalités liées à l'ingestion de toutes les données externes à la dataplatform, ainsi que la gestion complète de l'ingestion des données historiques.

Lors de la mise en production, ce sera mon travail qui sera utilisé quotidiennement pendant des années par les métiers, ce qui représente une grande satisfaction d'avoir contribué à un projet aussi vaste et stimulant.

En ce qui concerne les difficultés rencontrées, la compréhension du monde des asset managers a été un défi quotidien. Chaque jour, j'apprends davantage sur ce secteur, ses termes techniques et les règles financières appliquées. Mieux comprendre le business m'a motivé à approfondir mes connaissances, malgré les défis de compréhension et de vérification des données.

La communication a également été un point de friction. Travailler sur mes compétences de communication est essentiel, car expliquer des concepts techniques de manière simple est une valeur ajoutée. Je m'efforce d'améliorer cette capacité.

Ce stage a eu un impact significatif sur mes perspectives professionnelles. Il m'a ouvert les portes du monde financier, suscitant un vif intérêt pour le métier de data engineer. Cette expérience enrichissante, combinée à mon expérience précédente dans une startup, m'a convaincu de poursuivre ma carrière dans le domaine de la finance.

En réfléchissant sur le métier d'ingénieur, je retiens ma première définition : "L'ingénieur est un magicien qui résout des problèmes de la vie." Cette vision persiste et me motive à continuer à résoudre des défis complexes.

Quant à mes perspectives futures, je souhaite accumuler davantage d'expérience en tant que data engineer et poursuivre ma carrière dans le domaine financier, explorant les opportunités qui se présentent.