

# **RAPPORT DE STAGE DE FIN D'ETUDES**

*5ème Année en Ingénierie Informatique et Réseaux  
Option MIAGE*

---

## **Mise en place d'une Infrastructure de Données Multi-sources avec Plateforme de Visualisation et Business Intelligence**

---

***Réalisé par :***

*MAKROUM Ahmed*

***Tuteur (s) :***

*Encadrant Professionnel : Omar BENTATA.*

*Encadrant Pédagogique : Badr HIRCHOUA.*

*Au sein de (Allianz Maroc) :*



# Dédicace

Ce travail est dédié, avant tout, à mes parents adorés, **Mohamed MAKROUM** et **Fatima Ez-zahra NAJAH**, qui m'ont offert leur amour, leurs sacrifices et leur confiance sans limites. Leur présence bienveillante et leurs valeurs restent à jamais le socle de ma vie.

J'adresse également une pensée toute particulière à ma future fiancée, **Fatima Ez-zahra AZIM**, qui occupe déjà une place immense dans mon présent et mon avenir. Merci d'être à mes côtés, de croire en moi, de m'encourager à donner le meilleur de moi-même, et de rendre chaque étape de ce parcours plus douce, plus forte et plus belle par ta présence.

Je dédie aussi ces pages à **ma famille et à mes amis**, pour leur affection, leur soutien et tous les moments partagés, qui ont fait de ce chemin une aventure humaine unique et précieuse.

Je n'oublie pas **mes enseignants et mentors**, qui ont su éveiller en moi la passion de l'apprentissage et m'ont guidé avec sagesse. Leur enseignement a été une source d'inspiration constante, et je leur en suis profondément reconnaissant.

Enfin, je dédie ce travail à ceux qui, de près ou de loin, ont contribué à faire de moi la personne que je suis aujourd'hui, par leur aide, leurs conseils, ou simplement leur bienveillance tout au long de mon parcours.

**À tous ceux qui ont croisé ma route et ont contribué, ce moment vous appartient également.**

# Remerciements

La rédaction de ce mémoire marque l'aboutissement d'un parcours universitaire riche et intense, jalonné de rencontres, d'expériences et d'échanges qui ont profondément contribué à ma formation et à mon évolution personnelle.

Je tiens à adresser mes plus sincères remerciements à mon encadrant professionnel, **Mr Omar BENTATA**, pour la confiance qu'il m'a témoignée, pour l'opportunité d'intégrer l'équipe d'Allianz Maroc, et pour la richesse de ses conseils et de ses orientations tout au long de mon stage. Son accompagnement bienveillant et ses remarques avisées ont été d'une grande valeur pour la réussite de ce travail.

J'exprime également toute ma reconnaissance à mon encadrant académique, **Mr Badr HIRCHOUA**, dont l'exigence, la disponibilité et la rigueur scientifique ont été une véritable source de motivation. Son soutien constant, ses suggestions constructives et son regard critique ont largement contribué à l'amélioration de ce mémoire et à mon développement académique.

Je remercie chaleureusement le département IT d'**Allianz Maroc**, et tout particulièrement **Mme Layla BARRIT**, **Mr Mohamed BENABED** et **Mr Taha LOULIDI**, pour leur accueil, leur écoute et leur esprit d'équipe. Leur soutien et leur implication ont fait de mon stage une expérience professionnelle et humaine des plus enrichissantes.

Mes remerciements s'adressent également aux **membres du jury** pour avoir bien voulu consacrer de leur temps à l'évaluation de ce travail et pour la pertinence de leurs remarques et questions. Leur expertise a été précieuse.

Je n'oublie pas de remercier toute la communauté académique et administrative de l'**École Marocaine des Sciences de l'Ingénieur** de Casablanca, pour la qualité de l'enseignement dispensé, l'excellence de leur accompagnement et la fierté d'appartenir à cette institution.

Enfin, mes pensées vont à **ma famille et à mes amis**, pour leur soutien indéfectible, leur patience et leurs encouragements tout au long de cette aventure.

# Résumé

Ce projet de fin d'études a été réalisé au sein d'**Allianz Maroc**, une structure qui centralise et exploite quotidiennement un volume considérable d'informations issues de sources hétérogènes : systèmes transactionnels, applications internes, bases de données diverses, etc. Malgré cette richesse informationnelle, l'entreprise ne disposait jusque-là que de très peu d'outils dédiés à la Business Intelligence ou à la visualisation de données, ce qui limitait fortement la capacité des équipes à analyser, exploiter et valoriser la donnée au service des métiers.

Face à ce constat, la problématique principale de ce projet a été d'imaginer, de concevoir et de mettre en place une infrastructure moderne et évolutive de gestion de données, intégrant un *Data Warehouse* et des processus d'ETL automatisés pour centraliser, transformer et fiabiliser les données de l'ensemble des systèmes d'**Allianz Maroc**. L'objectif était double : d'une part, garantir la qualité, la cohérence et la disponibilité de l'information ; d'autre part, offrir aux utilisateurs métiers des outils conviviaux pour l'exploration, l'analyse et la visualisation des données, favorisant ainsi la prise de décision et l'innovation.

Pour répondre à ces enjeux, plusieurs solutions technologiques ont été déployées. **Apache NiFi** a été utilisé pour l'ingestion et l'orchestration des flux de données depuis les différentes sources, tandis que **Apache Spark** a permis d'assurer le traitement et la transformation de grandes volumétries de données de manière performante. Le stockage centralisé et structuré a été confié à une base de données relationnelle robuste, **PostgreSQL**, choisie pour sa fiabilité et sa capacité à gérer des environnements analytiques complexes. Enfin, la plateforme **Metabase** a été mise en place afin de permettre la visualisation dynamique, l'analyse exploratoire et le partage de tableaux de bord interactifs auprès des différents utilisateurs.

Ce mémoire retrace l'ensemble des étapes du projet, depuis l'analyse des besoins et des contraintes métiers jusqu'à la conception technique, la mise en œuvre des flux d'intégration et la valorisation des données via des outils de BI. Il met en lumière les défis rencontrés, les choix technologiques opérés, les solutions apportées ainsi que les perspectives d'évolution pour une exploitation optimale de la donnée chez **Allianz Maroc**.

**Mots-clés** : Infrastructure de données, Data Warehouse, ETL, Business Intelligence, Allianz Maroc, Apache NiFi, Apache Spark, PostgreSQL, Metabase, Visualisation de données, Gouvernance des données.

# Abstract

This end-of-study project was carried out within **Allianz Maroc**, an organization that manages and leverages a significant volume of information from heterogeneous sources on a daily basis, including transactional systems, internal applications, and various databases. Despite this wealth of data, the company previously lacked dedicated Business Intelligence and data visualization tools, which greatly limited the ability of teams to analyze, exploit, and enhance data in support of business operations.

Given this context, the main challenge of this project was to design and implement a modern, scalable data management infrastructure, integrating a *Data Warehouse* and automated ETL processes to centralize, transform, and ensure the reliability of data from all Allianz Maroc systems. The objective was twofold : first, to guarantee the quality, consistency, and availability of information ; and second, to provide business users with user-friendly tools for data exploration, analysis, and visualization, thus promoting better decision-making and innovation.

To address these challenges, several technological solutions were deployed. **Apache NiFi** was used for data ingestion and orchestration from various sources, while **Apache Spark** ensured efficient processing and transformation of large data volumes. Centralized and structured storage was entrusted to a robust relational database, **PostgreSQL**, chosen for its reliability and ability to manage complex analytical environments. Finally, the **Metabase** platform was implemented to enable dynamic data visualization, exploratory analysis, and the sharing of interactive dashboards with different users.

This report details all stages of the project, from needs and business requirements analysis to technical design, implementation of integration flows, and data enhancement through BI tools. It highlights the challenges encountered, the technological choices made, the solutions provided, and the prospects for further optimizing data usage at Allianz Maroc.

**Keywords :** Data infrastructure, Data Warehouse, ETL, Business Intelligence, Allianz Maroc, Apache NiFi, Apache Spark, PostgreSQL, Metabase, Data Visualization, Data Governance.

# Table des matières

<b>Dédicace</b>	<b>II</b>
<b>Remerciements</b>	<b>III</b>
<b>Résumé</b>	<b>IV</b>
<b>Abstract</b>	<b>V</b>
<b>Liste des sigles et acronymes</b>	<b>XI</b>
<b>Introduction Générale</b>	<b>1</b>
<b>1 Contexte général du Projet</b>	<b>3</b>
Introduction . . . . .	3
1 Présentation de l'organisme d'accueil . . . . .	4
1.1 Présentation générale . . . . .	4
1.2 Produits d'assurance d'Allianz . . . . .	7
1.3 Organisation . . . . .	8
1.4 Équipe de Management d'Allianz Maroc . . . . .	8
1.5 Principes fondamentaux d'Allianz . . . . .	9
1.6 Compagnies concurrentes d'Allianz Maroc . . . . .	10
2 Cadrage du projet . . . . .	11
2.1 Étude de l'existant . . . . .	11
2.2 Problématique . . . . .	12
2.3 Solution proposée . . . . .	13
2.4 Objectifs du projet . . . . .	14
3 Concepts fondamentaux : ETL, Big Data et l'Architecture Décisionnelle Moderne	15
3.1 Relation avec le projet Allianz Maroc . . . . .	16
4 Conduite du projet . . . . .	16
4.1 Diagramme de Gantt . . . . .	17
4.2 Diagramme de cas d'utilisation . . . . .	19
5 Conclusion . . . . .	20
<b>2 Analyse et Modélisation</b>	<b>21</b>

Introduction . . . . .	21
1 Cahier des charges . . . . .	21
2 Analyse des sources de données . . . . .	23
3 Modélisation du Data Warehouse . . . . .	25
4 Qualité des données et gouvernance . . . . .	27
5 Conclusion . . . . .	30
<b>3 Étude technique</b>	<b>32</b>
Introduction . . . . .	32
1 Analyse comparative et choix technologiques . . . . .	32
1.1 Ingestion et orchestration : Apache NiFi . . . . .	32
1.2 Traitement des données : Apache Spark . . . . .	33
1.3 Stockage : PostgreSQL . . . . .	33
1.4 Visualisation : Metabase . . . . .	34
2 Dataflow de la solution . . . . .	34
3 Architecture finale . . . . .	35
4 Conclusion . . . . .	36
<b>4 Réalisation et déploiement</b>	<b>38</b>
Introduction . . . . .	38
1 Création du PoC . . . . .	38
2 Création de la pipeline de données . . . . .	39
2.1 Ingestion des données avec Apache NiFi . . . . .	39
2.2 Transformation et chargement avec Apache Spark . . . . .	42
2.3 Stockage des données dans PostgreSQL . . . . .	44
2.4 Visualisation des données avec Metabase . . . . .	46
3 DevOps et déploiement . . . . .	52
3.1 Dockerisation des composants . . . . .	52
3.2 Orchestration avec Docker Compose . . . . .	55
3.3 Déploiement sur serveur local . . . . .	57
3.4 Tests et validation de la solution . . . . .	58
4 Conclusion . . . . .	58
<b>Conclusion Générale</b>	<b>59</b>
<b>Webographie</b>	<b>62</b>

# Liste des figures

1.1	Logo d'Allianz . . . . .	4
1.2	Présence d'Allianz en Afrique . . . . .	5
1.3	Chiffres clés sur les ressources humaines chez Allianz . . . . .	5
1.4	Équipe de management d'Allianz Maroc . . . . .	9
1.5	Les valeurs fondamentales d'Allianz . . . . .	10
1.6	Principales compagnies concurrentes d'Allianz Maroc sur le marché marocain . .	11
1.7	Architecture de la solution décisionnelle proposée . . . . .	13
1.8	Diagramme de Gantt du projet d'infrastructure de données Allianz Maroc . . . .	19
1.9	Diagramme de cas d'utilisation du système décisionnel Allianz Maroc . . . . .	20
2.1	Exemple d'extraction de données sinistres – flux orienté métier . . . . .	24
2.2	Schéma conceptuel du Data Warehouse – Sinistres Automobiles . . . . .	28
3.1	Schéma du dataflow de la solution . . . . .	35
3.2	Architecture finale de la solution . . . . .	36
4.1	Flow d'ingestion NiFi pour la base Oracle . . . . .	40
4.2	Flow d'ingestion NiFi pour la base AS/400 . . . . .	41
4.3	Flow d'ingestion NiFi pour les fichiers plats reçus par email . . . . .	42
4.4	Extrait du script PySpark de transformation et chargement . . . . .	43
4.5	Extrait du script PySpark de transformation et chargement . . . . .	43
4.6	Vue d'ensemble dans pgAdmin . . . . .	45
4.7	monitoring performance . . . . .	46
4.8	Ajout de la connexion PostgreSQL dans Metabase . . . . .	47
4.9	Configuration de la connexion PostgreSQL dans Metabase . . . . .	48
4.10	Interface de création d'une question/requête dans Metabase pour l'extraction des données . . . . .	48
4.11	Interface de création d'un dashboard . . . . .	49
4.12	Interface de création d'un utilisateur . . . . .	49
4.13	Envoi du mot de passe à l'utilisateur . . . . .	50
4.14	Gestion des groupes et des permissions dans Metabase . . . . .	51
4.15	Interface de connexion Metabase avec authentification sécurisée . . . . .	51
4.16	Dashboard de visualisation des règlements par sinistre (données floutées) . . . .	52



4.17 Conteneurisation de NiFi . . . . .	53
4.18 Conteneurisation de Spark . . . . .	54
4.19 Conteneurisation de PostgreSQL . . . . .	54
4.20 Conteneurisation de Metabase . . . . .	55
4.21 Configuration Docker Compose pour NiFi . . . . .	56
4.22 Configuration Docker Compose pour Spark . . . . .	56
4.23 Configuration Docker Compose pour PostgreSQL . . . . .	56
4.24 Configuration Docker Compose pour Metabase . . . . .	57
4.25 Configuration des volumes et réseaux Docker Compose . . . . .	57

# Liste des tableaux

1.1	Principaux produits d'assurance proposés par Allianz . . . . .	8
3.1	Comparatif des outils d'ingestion de données . . . . .	33
3.2	Comparatif des outils de traitement de données . . . . .	33
3.3	Comparatif des solutions de stockage . . . . .	34
3.4	Comparatif des outils de visualisation . . . . .	34

# Liste des sigles et acronymes

<b>API</b>	Application Programming Interface
<b>AS/400</b>	Application System/400 (IBM i)
<b>BI</b>	Business Intelligence
<b>CI/CD</b>	Continuous Integration / Continuous Deployment
<b>CSV</b>	Comma-Separated Values
<b>DWH</b>	Data Warehouse (entrepôt de données)
<b>dashboard</b>	Tableau de bord interactif
<b>EMSI</b>	École Marocaine des Sciences de l'Ingénieur
<b>ETL</b>	Extract, Transform, Load
<b>KPI</b>	Key Performance Indicator
<b>MAD</b>	Dirham marocain
<b>Metabase</b>	Plateforme de datavisualisation open-source
<b>NiFi</b>	Apache NiFi (outil d'orchestration de flux de données)
<b>PFE</b>	Projet de Fin d'Études
<b>PoC</b>	Proof of Concept
<b>PostgreSQL</b>	Système de gestion de base de données relationnelle
<b>Spark</b>	Apache Spark (moteur de traitement distribué de données)
<b>SQL</b>	Structured Query Language

# Introduction Générale

La donnée est aujourd'hui au cœur des leviers de transformation des organisations. Elle n'est plus seulement un sous-produit des systèmes transactionnels, mais une ressource stratégique, source de valeur, d'efficacité et de différenciation. Dans un contexte de digitalisation accélérée, la capacité à exploiter cette donnée de manière structurée, fiable et proactive devient un facteur critique de compétitivité, en particulier dans les secteurs fortement réglementés et sensibles à la performance opérationnelle, comme celui de l'assurance.

Allianz Maroc, filiale d'un groupe mondialement reconnu, génère quotidiennement une masse importante de données à travers ses activités : gestion des contrats, traitement des sinistres, suivi des indemnisations, reporting financier, conformité réglementaire, etc. Ces données sont issues d'écosystèmes techniques variés — mainframe IBM i (AS/400), bases relationnelles Oracle, fichiers plats générés par des systèmes tiers — souvent conçus dans des logiques silotées et historiquement orientées production plutôt que valorisation.

Cette diversité des sources et cette hétérogénéité des formats, combinées à l'absence d'un référentiel unifié de traitement et de visualisation, ont rendu complexe l'accès rapide et fiable à l'information pertinente. Les métiers exprimaient un besoin fort : disposer d'une plateforme agile, sécurisée, automatisée et évolutive, capable d'offrir une vision consolidée des données critiques, tout en restant accessible sans dépendance forte à la DSI.

C'est dans ce contexte qu'a été initié le projet objet de ce mémoire. L'objectif principal : concevoir et mettre en œuvre une infrastructure complète de Business Intelligence "self-service", capable de couvrir toute la chaîne de traitement des données — de l'ingestion brute jusqu'à la visualisation dynamique — dans une architecture cohérente, documentée, et déployée en environnement réel.

Ce projet s'inscrit dans une démarche volontairement pragmatique : éviter les usines à gaz technologiques, privilégier des outils open-source matures, assurer la lisibilité du code et des pipelines, faciliter la maintenance future, et maximiser l'appropriation métier. Les choix techniques ont été faits non seulement sur la base de benchmarks fonctionnels, mais aussi sur la capacité d'interfaçage avec l'existant et la robustesse dans un environnement de production réel.

L'architecture finale repose sur quatre piliers complémentaires :

- **Apache NiFi**, pour l'orchestration des flux de données : extraction automatisée depuis les systèmes AS/400, Oracle et les fichiers plats, normalisation minimale et déclenchement des traitements.
- **Apache Spark**, pour la transformation et le nettoyage à grande échelle : contrôle de la qualité des données, jointures inter-sources, enrichissement et préparation analytique.
- **PostgreSQL**, pour le stockage centralisé sous forme de *Data Warehouse* modélisé en étoile : structuration des faits et des dimensions, optimisation des requêtes analytiques.

- **Metabase**, pour la restitution visuelle et l'autonomisation des utilisateurs métier : dashboards interactifs, filtres dynamiques, gestion des rôles et des accès.

L'ensemble du projet a été conduit dans un esprit d'industrialisation. Chaque composant est scripté, conteneurisé lorsque pertinent, et documenté. La chaîne CI/CD a été pensée pour permettre des itérations rapides et sûres. Un soin particulier a été apporté à la traçabilité des traitements (loggings, audits), à la transparence des transformations (data lineage explicite), et à la conformité avec les politiques de sécurité internes.

Ce rapport propose une restitution détaillée du projet, structuré en quatre chapitres complémentaires :

- Le **Chapitre 1** pose le cadre du projet : présentation d'Allianz Maroc, problématique métier, objectifs stratégiques, périmètre fonctionnel, organisation du travail.
- Le **Chapitre 2** expose les travaux d'analyse et de modélisation : identification des sources, recensement des besoins utilisateurs, conception du modèle de données cible.
- Le **Chapitre 3** présente l'étude technique : exploration des alternatives possibles, justification des choix d'outils, définition de l'architecture logique et physique.
- Le **Chapitre 4** détaille la réalisation : mise en œuvre des pipelines d'ingestion et de transformation, implémentation du Data Warehouse, création des dashboards métiers.

Enfin, ce mémoire se conclut par une **Conclusion Générale** qui synthétise l'ensemble des résultats obtenus, évalue les objectifs atteints et détaille les apports du projet — aussi bien sur les plans techniques, organisationnels que personnels. Elle revient également sur les principales difficultés rencontrées et les solutions apportées, avant d'ouvrir sur des perspectives concrètes d'évolution à court, moyen et long terme.

Ce projet a démontré qu'il est possible, même dans un cadre contraint, de construire une infrastructure data robuste, cohérente et scalable, capable de répondre à des enjeux métiers complexes. Il s'inscrit dans une logique plus large de modernisation du système d'information d'Allianz Maroc, en posant les fondations d'une gouvernance des données structurée et en favorisant une culture orientée "data-driven" dans les processus de décision. L'impact produit dépasse le périmètre technique : il offre une base tangible sur laquelle bâtir de futures initiatives de transformation numérique au sein du groupe.

**Ce projet représente bien plus qu'un simple exercice académique : il est le fruit d'un investissement personnel intense, mêlant rigueur technique, questionnements constants et adaptation continue face aux contraintes réelles du terrain. J'y ai consacré mon énergie, ma curiosité, et ma volonté d'apprendre, parfois au prix de longues heures de recherche, d'expérimentations infructueuses, de doutes techniques, mais toujours avec l'exigence de livrer un résultat concret, utile et pérenne. Ce mémoire est le reflet de cet engagement : il incarne le produit final d'un travail de fond, mené avec sincérité et détermination.**

# Chapitre 1

## Contexte général du Projet

### Introduction

Ce premier chapitre constitue le socle introductif du mémoire, en offrant une immersion détaillée dans le contexte global qui a motivé et encadré la réalisation de notre projet de fin d'études. Il s'agit d'un passage essentiel, car il permet de comprendre non seulement l'environnement dans lequel s'inscrit le projet, mais aussi les enjeux stratégiques, organisationnels et technologiques qui le sous-tendent.

Nous débuterons par une présentation approfondie d'Allianz Maroc, l'organisme d'accueil, en retraçant son histoire, en détaillant sa structure organisationnelle, ses domaines d'activité, ainsi que sa position de leader sur le marché marocain de l'assurance. Cette introduction ne se limite pas à une simple description institutionnelle : elle met en lumière la dynamique de transformation digitale engagée par l'entreprise, illustrant sa volonté constante d'innovation et d'adaptation face aux nouveaux défis du secteur de l'assurance.

La suite du chapitre est consacrée à un cadrage rigoureux du projet. Nous procédons à une analyse exhaustive de l'existant, afin d'identifier les forces, les faiblesses et les axes d'amélioration du système actuel. Cette démarche analytique permet de cerner avec précision la problématique à laquelle notre intervention entend répondre, en tenant compte des besoins opérationnels, des contraintes techniques et des attentes des différentes parties prenantes. Nous définissons ensuite les objectifs spécifiques, mesurables et alignés avec la stratégie globale d'Allianz Maroc. Enfin, nous exposons la méthodologie adoptée et la planification détaillée des différentes phases du projet, depuis l'étude préliminaire jusqu'à la mise en œuvre des solutions proposées.

Tout au long de ce chapitre, notre objectif est d'offrir au lecteur une compréhension claire, argumentée et contextualisée des enjeux du projet : pourquoi il a été initié, comment il sera mené, et pour qui il est destiné. Chaque choix méthodologique est replacé dans son contexte, et les défis rencontrés sont illustrés par des exemples concrets issus de l'environnement professionnel d'Allianz Maroc. Nous mettons ainsi en lumière la contribution de ce projet à la stratégie d'innovation et de digitalisation de l'entreprise, en soulignant son impact potentiel sur la performance et la compétitivité d'Allianz Maroc dans un secteur en pleine mutation et en constante évolution.

# 1 Présentation de l'organisme d'accueil

## 1.1 Présentation générale

**Allianz Maroc : Un Acteur de Référence en Pleine Transformation** Le groupe **Allianz SE**, fondé en 1890 à Munich (Allemagne), s'impose aujourd'hui comme l'un des **leaders mondiaux** dans les domaines de l'**assurance** et de la **gestion d'actifs**. Fort de plus de **130 ans d'expérience**, Allianz a su bâtir une réputation internationale fondée sur la solidité financière, l'innovation et la qualité de service. Son rayonnement mondial est illustré par une présence dans plus de **70 pays** répartis sur les quatre continents – Europe, Afrique, Amérique et Asie-Pacifique – où il accompagne environ **128 millions de clients**, qu'il s'agisse de particuliers, de professionnels ou d'entreprises de toutes tailles.

L'organisation du groupe repose sur trois pôles stratégiques majeurs : l'**assurance Vie/-Santé**, l'**assurance Non-Vie** et la **gestion d'actifs**. Chacun de ces pôles est structuré selon des segments géographiques et sectoriels, permettant à Allianz d'adapter ses offres aux spécificités de chaque marché et de répondre efficacement aux besoins variés de ses clients. Cette organisation matricielle favorise la synergie entre les différentes entités du groupe et assure une gestion optimale des risques, tout en maintenant une proximité avec les marchés locaux.

Allianz se distingue également par sa capacité d'innovation et sa volonté d'intégrer les nouvelles technologies dans ses processus métiers. Le groupe investit massivement dans la transformation digitale, la cybersécurité et le développement de solutions d'assurance connectées, afin d'anticiper les évolutions du secteur et de proposer des services toujours plus performants et personnalisés.



FIGURE 1.1 – Logo d'Allianz

Sur le continent africain, Allianz est solidement implanté dans **15 pays**, notamment à travers ses filiales **Allianz Africa**, **Allianz Égypte**, **Euler Hermes Acmar** et **Allianz Global Corporate & Specialty**. Cette présence régionale permet au groupe de couvrir un large éventail de besoins en assurance et en gestion de risques, aussi bien pour les particuliers que pour les grandes entreprises et les institutions publiques. Allianz Africa joue un rôle clé dans le développement de solutions adaptées aux spécificités du marché africain, en tenant compte des enjeux économiques, sociaux et réglementaires propres à chaque pays.

L'acquisition de **Zurich Assurances Maroc** en 2016, désormais intégrée à **Allianz Maroc**, a marqué une étape stratégique majeure pour le groupe. Cette opération a permis à Allianz de renforcer sa position sur le marché marocain, d'élargir son portefeuille de clients et de services, et d'accroître sa capacité d'innovation locale. Allianz Maroc bénéficie ainsi du savoir-faire international du groupe, tout en développant des offres spécifiques répondant aux attentes du marché national. Grâce à cette intégration, Allianz Maroc dispose aujourd'hui d'une structure organisationnelle solide, d'un réseau commercial étendu et d'une expertise reconnue dans la gestion des risques et la transformation digitale.

En résumé, la stratégie d'expansion internationale d'Allianz, combinée à une politique d'acquisitions ciblées et à une capacité d'adaptation aux marchés locaux, lui permet de consolider sa position de leader et d'accompagner durablement ses clients dans la gestion de leurs risques et la valorisation de leur patrimoine.

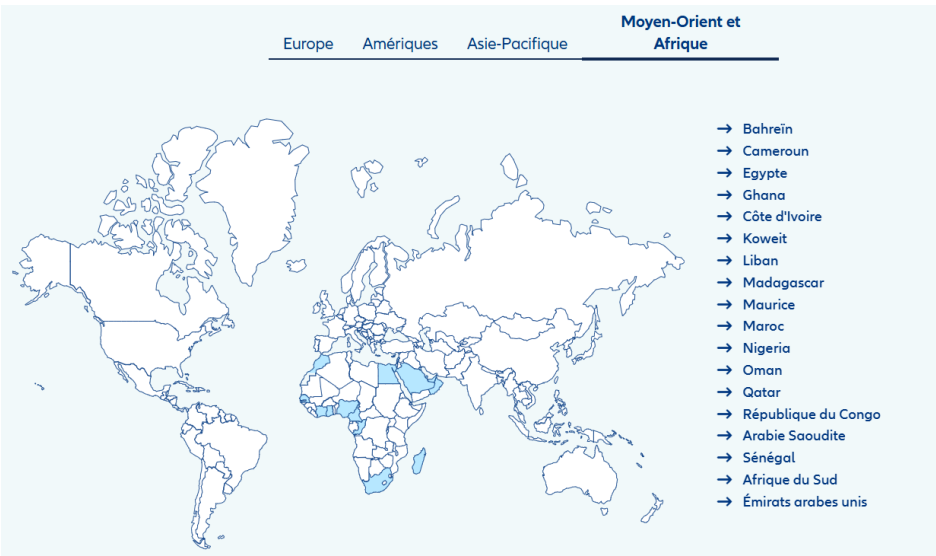


FIGURE 1.2 – Présence d'Allianz en Afrique

À l'échelle mondiale, Allianz emploie plus de **157 000 personnes**, dont **25 128 nouveaux collaborateurs** recrutés en 2023, témoignant de la vitalité et de l'attractivité du groupe. En 2024, le groupe a enregistré un **chiffre d'affaires de 179,8 milliards d'euros** et un **résultat d'exploitation de 16,0 milliards d'euros**. Il gère environ **776 milliards d'euros d'actifs** pour le compte de ses clients assureurs, tandis que ses filiales spécialisées, **PIMCO** et **Allianz Global Investors**, administrent près de **1 900 milliards d'euros** pour le compte de tiers, ce qui illustre la puissance et la diversité de ses activités financières.



FIGURE 1.3 – Chiffres clés sur les ressources humaines chez Allianz



L'un des piliers majeurs de la stratégie d'Allianz réside dans l'intégration systématique de **critères écologiques et sociaux** à tous les niveaux de ses décisions stratégiques et opérationnelles. Cette démarche volontariste, initiée depuis plusieurs années, positionne Allianz comme l'un des **leaders incontestés du secteur en matière de développement durable et de responsabilité sociétale**. Concrètement, Allianz s'engage à évaluer l'impact environnemental et social de chacune de ses activités, que ce soit dans la gestion quotidienne de ses opérations, dans la conception et la commercialisation de ses produits, ou encore dans la sélection de ses investissements financiers. L'entreprise met en œuvre des politiques rigoureuses visant à réduire son empreinte carbone, à promouvoir l'efficacité énergétique, à soutenir la transition vers une économie verte et à favoriser l'inclusion sociale.

L'engagement d'Allianz en faveur du développement durable est reconnu à l'échelle internationale, comme en atteste son classement régulier dans l'indice de référence **Dow Jones Sustainability**. Cet indice prestigieux évalue la performance des entreprises selon des critères environnementaux, sociaux et de gouvernance (**ESG**), et distingue Allianz pour la qualité de sa gouvernance, la transparence de ses pratiques, ainsi que pour ses initiatives innovantes en matière de responsabilité sociétale. Allianz participe activement à des programmes mondiaux tels que le *Global Compact des Nations Unies*, et publie chaque année un rapport détaillé sur ses actions et ses progrès en matière de développement durable, illustrant ainsi sa volonté de rendre compte de ses engagements auprès de l'ensemble de ses parties prenantes.

Par ailleurs, Allianz accorde une importance capitale au bien-être, à la motivation et à l'engagement de ses collaborateurs, considérés comme le principal moteur de sa performance et de son innovation. L'entreprise a mis en place un ensemble de politiques et d'initiatives visant à améliorer la qualité de vie au travail, à promouvoir la diversité et l'inclusion, et à favoriser le développement professionnel de chacun. Son **indice WorkWell**, qui mesure le niveau de satisfaction, de motivation et de bien-être au travail, a atteint un niveau record de **76 %** en 2023, soit l'un des meilleurs scores du secteur. Ce résultat témoigne de l'efficacité des dispositifs internes, tels que les programmes de formation continue, les actions de prévention des risques psychosociaux, les mesures d'accompagnement à la parentalité, ou encore les politiques de flexibilité et de télétravail.

De plus, la politique d'actionnariat salarié constitue un levier puissant de fidélisation et d'implication des équipes. Allianz encourage activement ses collaborateurs à devenir parties prenantes du groupe à travers le plan d'achat d'actions (**ESPP** – Employee Share Purchase Plan), qui offre des conditions avantageuses et un accompagnement personnalisé. Aujourd'hui, **70 %** des employés d'Allianz sont devenus **actionnaires du groupe**, ce qui favorise un sentiment d'appartenance fort et un engagement durable. Ce dispositif permet d'aligner les intérêts des collaborateurs avec ceux de l'entreprise et de ses parties prenantes, tout en renforçant la cohésion interne et la culture d'entreprise.

En résumé, la stratégie d'Allianz en matière de développement durable et de gestion des ressources humaines repose sur une approche globale et intégrée, visant à concilier performance économique, responsabilité sociale et respect de l'environnement. Ces engagements structurants

constituent un véritable avantage concurrentiel, permettant à Allianz de bâtir une réputation solide, de fidéliser ses talents et de répondre aux attentes croissantes de ses clients et partenaires en matière d'éthique et de durabilité.

### **1.2 Produits d'assurance d'Allianz**

Allianz propose une offre particulièrement diversifiée et complète de solutions d'assurance et de gestion financière, conçue pour répondre de manière précise et adaptée aux besoins spécifiques de différents segments de clientèle : particuliers, entreprises de toutes tailles et institutions. Cette gamme couvre l'ensemble des risques majeurs auxquels peuvent être confrontés les clients, qu'il s'agisse de la protection des biens, de la santé, de la prévoyance, de la gestion d'actifs ou encore de la couverture des risques professionnels et commerciaux. Le tableau ci-dessous présente de façon synthétique les principaux produits proposés par Allianz, accompagnés d'une brève description de chacun afin d'illustrer la richesse et la complémentarité de l'offre :

Produit	Définition générale
<b>Assurance de biens et de dommages</b>	Protection des biens personnels ( <i>maison, automobile, etc.</i> ) et couverture de la <b>responsabilité civile</b> . Offre une gamme de produits et services pour particuliers et entreprises afin de les protéger contre divers risques de la vie quotidienne, de la naissance à la retraite.
<b>Assurance santé et vie</b>	Solutions évolutives répondant aux besoins de <b>protection santé</b> , bien-être, invalidité et <b>assurance-vie</b> des particuliers, familles, organisations et partenaires. Permet de prendre soin de ses proches avec confiance.
<b>Gestion d'actifs</b>	Gestion de portefeuilles d'investissement pour clients privés et institutionnels, couvrant un large éventail d'instruments ( <i>obligations, actions, placements alternatifs</i> ). Vise à développer et protéger le patrimoine avec l'appui d'un réseau mondial d'experts.
<b>Assurance entreprise</b>	Assurance spécialisée et conseils en <b>gestion des risques</b> pour entreprises de toutes tailles (PME aux multinationales), incluant la protection des actifs, des chaînes d'approvisionnement, etc. Présence mondiale dans plus de 160 pays.
<b>Assurance-crédit commercial et cautions</b>	Solutions de gestion du <b>crédit interentreprises</b> incluant l'assurance-crédit, la caution, le recouvrement de créances, la prévention de la fraude et la couverture contre les risques politiques. Permet aux entreprises de sécuriser leurs transactions et de faire face aux imprévus grâce à un réseau de surveillance mondiale.

TABLE 1.1 – Principaux produits d'assurance proposés par Allianz

### 1.3 Organisation

### 1.4 Équipe de Management d'Allianz Maroc

L'équipe de management d'Allianz Maroc est composée de directeurs responsables de différentes directions stratégiques de l'entreprise. Chacun d'eux pilote un domaine clé, tel que la finance, les marchés des particuliers et des entreprises, le marché vie, la distribution, l'indemnisation, la transformation digitale, les ressources humaines, le secrétariat général ou encore la technique assurance et les risques. Cette diversité de compétences permet d'assurer une gestion cohérente et efficace, tout en garantissant l'alignement des opérations avec les objectifs globaux d'Allianz Maroc.



FIGURE 1.4 – Équipe de management d'Allianz Maroc

J'ai effectué mon stage de fin d'études au sein du **département IT** d'Allianz Maroc, sous la supervision de Monsieur **Omar Bentata**, Responsable de l'**Architecture d'Entreprise**, de la **Stratégie** et de la **Gouvernance IT**. Cette expérience s'est déroulée dans un contexte **stimulant**, au cœur des enjeux de **transformation numérique** du groupe. Elle m'a permis de contribuer à un projet structurant autour de la mise en place d'une solution de **Business Intelligence en libre-service**, tout en développant mes **compétences techniques** et ma compréhension des **problématiques IT** à l'échelle d'une grande entreprise.

## 1.5 Principes fondamentaux d'Allianz

Allianz Maroc s'appuie sur un socle solide de **valeurs fondamentales** qui orientent l'ensemble de ses actions, décisions et relations, tant en interne qu'en externe. Ces principes ne sont pas de simples slogans, mais constituent la véritable colonne vertébrale de la culture d'entreprise d'Allianz. Ils sont intégrés à tous les niveaux de l'organisation et servent de boussole pour chaque collaborateur, favorisant ainsi un environnement de travail où la collaboration, l'excellence et l'épanouissement professionnel sont constamment recherchés. Ces valeurs sont également un levier stratégique, permettant à Allianz de se distinguer sur le marché et de bâtir une réputation durable auprès de ses clients, partenaires et parties prenantes.

### 1. Entrepreneuriat

L'esprit entrepreneurial occupe une place centrale dans l'ADN d'Allianz. L'entreprise encourage activement ses collaborateurs à adopter une attitude proactive, à prendre des initiatives et à sortir des sentiers battus. L'innovation est valorisée à tous les niveaux, et chaque employé est incité à proposer de nouvelles idées, à expérimenter et à apprendre de ses erreurs. Cette culture de l'audace s'accompagne d'un soutien managérial qui favorise la prise de risque réfléchie, tout en assurant un cadre sécurisant pour transformer les échecs en opportunités d'apprentissage et de progrès collectif.

### 2. Excellence client et marché

La recherche de l'excellence, tant dans la relation client que dans la performance sur le

marché, est une priorité constante pour Allianz. L'entreprise s'efforce d'offrir un service irréprochable, personnalisé et à forte valeur ajoutée à chacun de ses clients. Cela se traduit par une écoute attentive des besoins, une réactivité exemplaire et une volonté permanente d'anticiper les attentes du marché. Allianz investit dans le développement de l'expertise de ses équipes, la simplification des processus internes et l'amélioration continue de ses offres, afin de garantir une expérience client optimale et de consolider sa position de leader.

### 3. **Confiance**

La confiance constitue le socle des relations qu'Allianz entretient avec ses clients, ses collaborateurs et l'ensemble de ses partenaires. Cette confiance se construit au quotidien grâce à une intégrité sans faille, une transparence dans la communication et un respect rigoureux des engagements pris. Allianz attache également une grande importance à la diversité, à l'inclusion et à la responsabilité sociétale, considérant que la richesse des profils et l'engagement citoyen sont des atouts majeurs pour bâtir une entreprise solide, respectée et durable.

### 4. **Leadership collaboratif**

Le leadership chez Allianz ne se limite pas à la prise de décision individuelle, mais repose avant tout sur la capacité à fédérer, à écouter et à développer les talents. L'entreprise promeut un management participatif, où l'échange de bonnes pratiques, le partage d'expériences et le feedback constructif sont encouragés. Le bien-être au travail est également une priorité, car Allianz est convaincue que la performance collective passe par l'épanouissement de chacun. Ce leadership collaboratif favorise l'innovation, la cohésion d'équipe et l'atteinte des objectifs communs dans un climat de confiance et de respect mutuel.



FIGURE 1.5 – Les valeurs fondamentales d'Allianz

## 1.6 **Compagnies concurrentes d'Allianz Maroc**

Le marché marocain de l'assurance est **fortement concurrentiel**, avec la présence de plusieurs acteurs majeurs. Allianz Maroc fait face à une concurrence intense de compagnies bien établies telles que **Wafa Assurance** (leader du secteur), **RMA**, **AXA Assurance Maroc**, **San-**

lam, Mutuelle Taamine Chaabi et AtlantaSanad. Ces entreprises disposent d'un **réseau commercial étendu**, d'une **forte notoriété** et d'une **large gamme de produits**.

Dans ce contexte, **Allianz Maroc** se positionne comme un acteur en croissance, cherchant à renforcer sa part de marché grâce à l'**innovation**, l'**amélioration de ses services** et des **partenariats stratégiques**.



FIGURE 1.6 – Principales compagnies concurrentes d'Allianz Maroc sur le marché marocain

## 2 Cadrage du projet

### 2.1 Étude de l'existant

#### Le Défi Actuel chez Allianz Maroc : Des Données Enfermées

Aujourd'hui, Allianz Maroc fait face à une **réalité complexe** concernant ses données. L'entreprise jongle avec une **multiplicité de sources d'information**, allant des systèmes historiques comme l'**AS/400** aux bases de données plus modernes comme **Oracle**, sans oublier une quantité significative de **fichiers plats**. Cette hétérogénéité, loin d'être un atout, crée de véritables **silos de données**.

**Un Accès Pénible et une Exploitation Limitée** : l'impact de cette fragmentation est **direct et palpable**. L'accès aux données est un **processus manuel, lent et totalement décentralisé**. Pour obtenir la moindre information, les équipes métier sont contraintes de passer par des **demandes par e-mail, des appels téléphoniques ou des extractions ponctuelles**. Ce mode de fonctionnement est non seulement **inefficace**, mais il engendre également des **délais considérables** et un **risque d'erreur accru**.

**Absence de Vision Centralisée** : il n'existe **aucune plateforme de visualisation unifiée et conviviale** à la disposition des équipes métier. Cela signifie que malgré l'immense **volume et la richesse des données** qu'Allianz Maroc possède, leur **potentiel reste largement inexploité**. Les équipes peinent à avoir une vue d'ensemble, à identifier des tendances ou à prendre des

décisions basées sur des informations consolidées et à jour. Les données sont là, mais elles sont **inaccessibles et invisibles** pour ceux qui en ont le plus besoin pour piloter l'activité.

## 2.2 Problématique

### Fragmentation des Données : Un Frein Majeur à la Performance d'Allianz Maroc

Dans un environnement où les **sources de données sont multiples, cloisonnées et hétérogènes**, Allianz Maroc fait face à une problématique structurelle : l'incapacité à exploiter pleinement son capital informationnel. Cette situation génère des inefficacités critiques, limitant aussi bien la **réactivité métier** que la **capacité décisionnelle**.

#### 1. Absence de Vision Globale et Fiable

Le principal enjeu réside dans l'**impossibilité d'obtenir une vue consolidée de l'activité**. Les données, issues de multiples systèmes non interconnectés, restent **fragmentées et non harmonisées**. Cela empêche :

- La production rapide de **rapports analytiques cohérents**,
- La **traçabilité fiable** des indicateurs de performance,
- Une prise de décision **fondée sur des données unifiées**.

Ce manque de convergence nuit à la lisibilité opérationnelle et accroît la complexité des analyses transverses.

#### 2. Surcharge Manuelle et Faible Valeur Ajoutée

La collecte de données repose encore largement sur des **processus manuels et redondants**, mobilisant des ressources qualifiées sur des tâches à faible valeur ajoutée :

- **Extraction manuelle de fichiers**, fusion de données sous Excel,
- **Traitements répétitifs** non automatisés,
- **Temps considérable** consacré à la mise en forme plutôt qu'à l'analyse.

Ce fonctionnement freine la capacité des équipes à générer de l'insight stratégique.

#### 3. Dépendance aux Experts et Risques Opérationnels

La connaissance des jeux de données critiques est souvent détenue par quelques **personnes référentes**, ce qui crée des risques majeurs :

- **Goulots d'étranglement** lors des absences ou rotations d'effectifs,
- **Perte de continuité** dans la compréhension des flux métiers,
- **Non-scalabilité** des pratiques actuelles, difficilement transposables ou industrialisables.

L'absence de documentation formelle ou de partage des savoirs renforce ces vulnérabilités.

#### 4. Impact Global : Inefficacité, Lenteur et Manque d'Agilité

Les effets concrets sont multiples :

- **Allongement des délais** de reporting et d'analyse,
- **Réduction de la capacité à répondre aux besoins métiers en temps réel**,
- **Frein à la scalabilité** des processus analytiques et décisionnels.

Ces limites structurantes entravent la transformation data-driven de l'organisation et réduisent sa compétitivité à moyen terme.

**En résumé** : la fragmentation des données, couplée à une faible automatisation et une dépendance humaine forte, expose Allianz Maroc à un **risque opérationnel réel** et freine sa capacité à bâtir une intelligence décisionnelle robuste.

### 2.3 Solution proposée

Face à la fragmentation et à l'inaccessibilité des données, la solution proposée repose sur la mise en place d'une **infrastructure décisionnelle unifiée**, conçue pour automatiser l'ingestion, le traitement, le stockage et l'exploitation des données à l'échelle de l'entreprise. L'architecture cible s'articule autour d'un pipeline structurant : **données sources** → **processus ETL automatisés** → **entrepôt de données centralisé** → **visualisation et exploration interactive**.

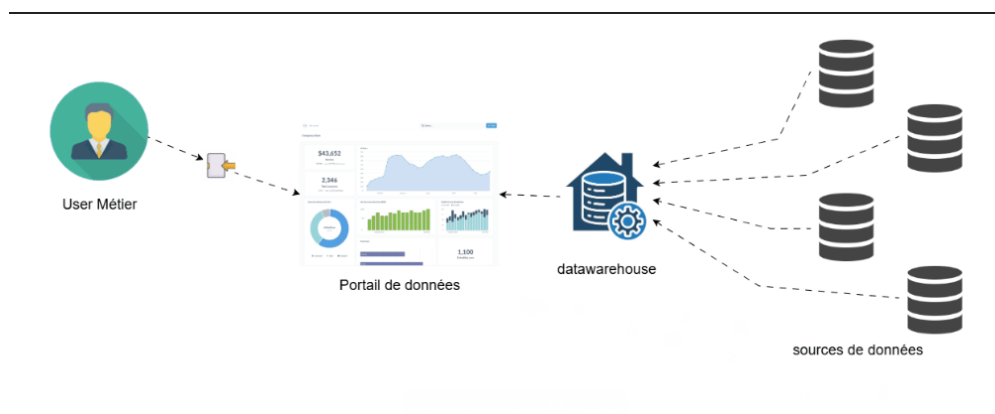


FIGURE 1.7 – Architecture de la solution décisionnelle proposée

Dans cette approche, les données issues de différents systèmes (applications métiers, bases de données, fichiers plats, etc.) sont collectées de façon sécurisée et automatisée. Un processus d'intégration des données (ETL) est mis en place pour extraire les informations pertinentes, les transformer (nettoyage, enrichissement, harmonisation des formats) et les charger dans un entrepôt de données unique. Ce référentiel centralisé permet de regrouper l'ensemble des données métiers dans une structure cohérente, durable et facilement exploitable.

L'entrepôt de données joue un rôle fondamental : il garantit la qualité, la traçabilité et la fiabilité des informations, tout en facilitant leur accès pour les différents utilisateurs. Les données y sont organisées selon des modèles adaptés aux besoins d'analyse et de reporting, permettant de croiser les informations issues de sources hétérogènes et de produire des indicateurs consolidés.

Pour exploiter ce patrimoine data, la solution prévoit la mise à disposition d'outils de visualisation et d'exploration interactifs. Ces outils permettent aux utilisateurs métiers de créer et consulter des tableaux de bord, d'effectuer des analyses ad hoc et d'exporter des rapports sans dépendre des équipes techniques. L'objectif est de démocratiser l'accès à l'information, d'accroître l'autonomie des utilisateurs et de favoriser une prise de décision rapide et éclairée.



En résumé, la solution proposée vise à transformer une architecture fragmentée et manuelle en une plateforme **automatisée, évolutive et pilotée par la donnée**. Elle permet de centraliser, fiabiliser et valoriser les données d'Allianz Maroc, tout en offrant aux équipes métiers les moyens d'explorer et d'analyser l'information de manière autonome et pertinente.

### 2.4 Objectifs du projet

Le projet a été conçu avec des objectifs clairs et stratégiques visant à transformer la gestion et l'exploitation des données chez Allianz Maroc. Au-delà de la simple mise en place d'une infrastructure, l'ambition était de créer une véritable culture de la donnée au sein de l'entreprise.

**Renforcer l'autonomie des utilisateurs dans l'exploration de la donnée** : L'un des buts fondamentaux était de permettre aux équipes métiers de **naviguer, d'analyser et d'extraire des insights** par elles-mêmes, sans dépendre constamment des services informatiques. Cela inclut la mise à disposition d'outils intuitifs et de tableaux de bord interactifs pour une **analyse self-service**.

**Réduire drastiquement les dépendances aux équipes IT pour les reportings** : En automatisant les processus de collecte et de préparation des données et en fournissant des outils de visualisation simples d'utilisation, le projet visait à **diminuer la charge des équipes IT** liée aux demandes de rapports récurrentes, leur permettant ainsi de se concentrer sur des tâches à plus forte valeur ajoutée et sur l'innovation.

**Assurer une qualité et une traçabilité des données centralisées** : Plutôt que de jongler avec des informations fragmentées et potentiellement incohérentes issues de multiples sources, l'objectif était de consolider toutes les données pertinentes dans un **Data Warehouse unique et fiable**. Cela garantit une **source unique de vérité**, améliorant considérablement la **fiabilité et la cohérence** des analyses et des décisions. Chaque donnée serait traçable jusqu'à sa source originale.

**Valoriser un patrimoine data aujourd'hui sous-exploité** : Allianz Maroc dispose d'un volume considérable de données transactionnelles et opérationnelles. Le projet visait à transformer ces **données brutes en un actif stratégique** en les rendant analysables et exploitables, permettant ainsi d'identifier de nouvelles opportunités, d'optimiser les processus existants et de mieux comprendre le comportement des clients.

**Passer d'un fonctionnement manuel à une culture de la donnée structurée** : L'ambition ultime était d'opérer un **changement de paradigme**, passant d'un environnement où les données sont traitées manuellement et de manière ad hoc, à une approche où l'analyse des données devient une **composante intégrale et systématique** de la prise de décision à tous les niveaux de l'organisation. Cela implique l'adoption de processus standardisés, d'outils performants et d'une mentalité axée sur la donnée.

### 3 Concepts fondamentaux : ETL, Big Data et l'Architecture Décisionnelle Moderne

Dans un monde où les entreprises génèrent et collectent chaque jour des quantités croissantes d'informations, savoir exploiter ces données devient un levier stratégique majeur. Cette problématique est particulièrement vraie pour Allianz Maroc, dont les données sont aujourd'hui **dispersées, peu accessibles, et sous-utilisées**. Pour répondre à ce défi, le projet s'inscrit dans une approche dite de **Big Data**, c'est-à-dire une manière de gérer les données à plus grande échelle, de manière automatisée, structurée et exploitable.

**Mais que signifie réellement "Big Data" dans ce contexte ?** Il ne s'agit pas simplement de volume, mais aussi de **variété** (multiplicité des formats et sources), de **vélocité** (vitesse de génération et d'accès) et surtout de **valeur**. L'objectif est de transformer des données brutes et dispersées en **information structurée et utile pour la prise de décision**.

**La réponse technologique proposée s'articule autour de quatre piliers fondamentaux :**

1. **L'ingestion des données (collecte automatisée)**

Les données d'Allianz Maroc proviennent de multiples systèmes, fichiers ou bases. Pour les collecter et les faire circuler automatiquement, nous utilisons **Apache NiFi**, un outil conçu pour **orchestrer les flux de données** sans effort manuel. Il agit comme un "chef de trafic", capable de gérer en continu les données en provenance de différentes sources.

2. **Le traitement des données (nettoyage et structuration)**

Une fois collectées, les données doivent être nettoyées, transformées et rendues cohérentes. Ce travail est réalisé par **Apache Spark**, un moteur puissant de traitement de données qui travaille en parallèle sur des volumes importants. Spark permet de **préparer les données** pour les rendre fiables et exploitables.

3. **Le stockage structuré (organisation et centralisation)**

Les données transformées sont stockées dans une base relationnelle appelée **PostgreSQL**, qui constitue le **Data Warehouse**. C'est un espace centralisé, organisé, où toutes les données métiers sont accessibles de manière claire, sécurisée et durable.

4. **La visualisation des données (compréhension et pilotage)**

Une fois centralisées, les données doivent pouvoir être comprises et utilisées par les équipes métiers. C'est ici que **Metabase** entre en jeu. Cet outil permet de **créer facilement des tableaux de bord**, des graphiques, des analyses sans écrire une seule ligne de code. Il démocratise l'accès à l'information.

**Ainsi, le projet met en place une architecture de type Big Data, qui permet à Allianz**

Maroc de passer :

Données brutes → Collecte automatisée → Traitement intelligent  
→ Stockage structuré → Décision éclairée

Ce parcours est souvent appelé **pipeline de données**, ou encore **ETL** pour *Extraction, Transformation, Chargement*. Il constitue l'épine dorsale de toute stratégie de valorisation de données. Grâce à cette architecture, Allianz Maroc pourra **centraliser l'information, réduire la dépendance aux tâches manuelles**, et surtout **outiller ses équipes métiers pour qu'elles exploitent la donnée de façon autonome et intelligente**.

En résumé, cette démarche vise à bâtir une fondation data solide, évolutive et ouverte, à la fois pour répondre aux besoins actuels et pour accompagner la croissance future de l'entreprise dans un monde piloté par la donnée.

### 3.1 Relation avec le projet Allianz Maroc

Ce projet s'inscrit parfaitement dans cette logique en mettant en place :

**Une fondation ETL robuste** avec Apache NiFi pour l'orchestration des flux et Apache Spark pour les transformations complexes, capable de gérer l'évolution des volumes et la complexité croissante des traitements.

**Une architecture évolutive** conçue pour absorber de nouveaux types de données et s'adapter aux exigences futures en termes de performance et de scalabilité.

**Une approche Data-Driven** qui transforme Allianz Maroc d'une organisation où les données sont subies en une organisation où elles deviennent un actif stratégique pour la prise de décision, l'innovation produit, et l'optimisation opérationnelle.

Cette infrastructure constitue donc le **socle technique nécessaire** pour accompagner la transformation digitale d'Allianz Maroc et préparer l'entreprise aux défis data de demain.

## 4 Conduite du projet

La réalisation de ce projet s'est appuyée sur une **démarche structurée, séquencée et itérative**, combinant analyse des besoins, expérimentations progressives et validations continues. Cette approche a permis une adaptation dynamique aux contraintes du terrain et aux retours des parties prenantes, garantissant ainsi la pertinence et la robustesse de la solution mise en œuvre.

### 1. Étude de l'existant

La première étape a consisté à analyser en profondeur les **systèmes d'information existants**, les flux de données actuels, ainsi que les pratiques métiers en place. Cette phase a permis d'identifier les principaux **points de friction**, les sources de données critiques et les dysfonctionnements liés

à la dispersion et à la manipulation manuelle de l'information.

### 2. Recueil des besoins métiers

Un ensemble d'échanges ciblés avec les utilisateurs finaux et les équipes fonctionnelles a permis de **formaliser les besoins métiers en matière de reporting, de visualisation et d'accès aux données**. L'objectif a été de clarifier les attentes en termes de périmètre fonctionnel, de fréquence d'analyse et de niveau d'autonomie souhaité par les équipes métiers.

### 3. Choix technologiques

Sur la base des exigences identifiées, une sélection rigoureuse d'outils a été effectuée. Les technologies retenues — **Apache NiFi, Apache Spark, PostgreSQL et Metabase** — ont été choisies pour leur **complémentarité**, leur **maturité industrielle** et leur **capacité à s'adapter au contexte technique et organisationnel d'Allianz Maroc**.

### 4. Conception et réalisation d'un Proof of Concept (PoC)

Avant de généraliser la solution, un **prototype fonctionnel** a été développé. Ce PoC a permis de **valider les choix technologiques**, de simuler des cas d'usage concrets et de **recueillir les premiers retours utilisateurs**. Cette étape a été cruciale pour **ajuster les orientations techniques** avant un passage à l'échelle.

### 5. Réalisation de la solution finale et déploiement progressif

Sur la base du PoC, la solution a été implémentée de manière **itérative et incrémentale**, en intégrant progressivement les différents modules techniques et en assurant à chaque itération :

- des tests de bon fonctionnement,
- des démonstrations aux parties prenantes,
- et des ajustements continus en fonction des retours utilisateurs.

Cette logique cyclique a permis de **minimiser les risques**, de **maximiser l'adhésion utilisateur** et de garantir une montée en charge maîtrisée.

**En résumé**, cette approche — *étude de l'existant → analyse des besoins → choix techniques → PoC → déploiement par itérations* — a permis de construire une solution **adaptée, évolutive et immédiatement opérationnelle**, en phase avec les attentes métiers d'Allianz Maroc.

## 4.1 Diagramme de Gantt

Le suivi de l'avancement du projet a été rigoureusement encadré à l'aide d'un **diagramme de Gantt détaillé**, couvrant l'ensemble de la période du **3 mars 2025 au 7 juillet 2025**. Ce diagramme a constitué un véritable fil conducteur, permettant de visualiser de façon synthétique et structurée toutes les **étapes majeures du projet**, les **jalons critiques**, ainsi que les **délais associés à chaque phase**. Grâce à cette planification, il a été possible d'anticiper les éventuels points de blocage, d'optimiser la répartition des ressources et de garantir une coordination efficace entre les différents acteurs impliqués. Le diagramme de Gantt a ainsi joué un rôle central pour **assurer le respect des échéances, suivre l'avancement des tâches, et faciliter la com-**

**munication** entre les membres de l'équipe projet, tout en offrant une vision globale et partagée de la feuille de route.

Le projet s'est articulé autour de plusieurs étapes clés, chacune correspondant à une phase bien définie du cycle de vie du projet :

- **Phase 1 : Analyse des besoins** (3 mars 2025 - 14 mars 2025) : Cette première phase a consisté à identifier précisément les problématiques métiers, à recueillir les attentes des parties prenantes et à analyser en détail la situation existante. Elle a permis de dresser un état des lieux complet et de poser les bases fonctionnelles du projet.
- **Phase 2 : Recherche technologique** (17 mars 2025 - 21 mars 2025) : Durant cette étape, une veille approfondie a été menée afin d'identifier et d'évaluer les différentes solutions technologiques disponibles sur le marché. L'objectif était de sélectionner la stack la plus adaptée aux besoins identifiés, en tenant compte des contraintes techniques et organisationnelles.
- **Phase 3 : Conception de l'architecture** (24 mars 2025 - 4 avril 2025) : Cette phase a porté sur la définition de l'architecture du Data Warehouse, la modélisation des données, ainsi que le choix des technologies à mettre en œuvre. Un effort particulier a été consacré à la conception d'une architecture évolutive, robuste et alignée avec les objectifs stratégiques d'Allianz Maroc.
- **Phase 4 : Développement ETL** (7 avril 2025 - 23 mai 2025) : Il s'est agi ici de configurer Apache NiFi pour l'orchestration des flux, de développer les jobs Spark pour les traitements complexes, et de mettre en place les pipelines de transformation des données. Cette étape a également inclus des tests unitaires et d'intégration pour garantir la fiabilité des traitements.
- **Phase 5 : Visualisation avec Metabase** (26 mai 2025 - 13 juin 2025) : Cette phase a consisté à configurer Metabase, à développer les tableaux de bord interactifs et à former les utilisateurs finaux à l'utilisation de l'outil. L'accent a été mis sur la restitution claire et pertinente des indicateurs métiers.
- **Phase 6 : Tests et déploiement** (16 juin 2025 - 27 juin 2025) : L'ensemble de la solution a été soumis à une batterie de tests fonctionnels et techniques, suivis d'une validation finale avec les parties prenantes. Le déploiement en production a été réalisé de manière progressive afin de limiter les risques et d'assurer une transition en douceur.
- **Phase 7 : Documentation** (30 juin 2025 - 4 juillet 2025) : Enfin, une documentation complète a été produite, incluant le rapport de projet, la préparation de la présentation finale, ainsi que la rédaction de la documentation technique destinée à faciliter la maintenance et l'évolution future de la solution.

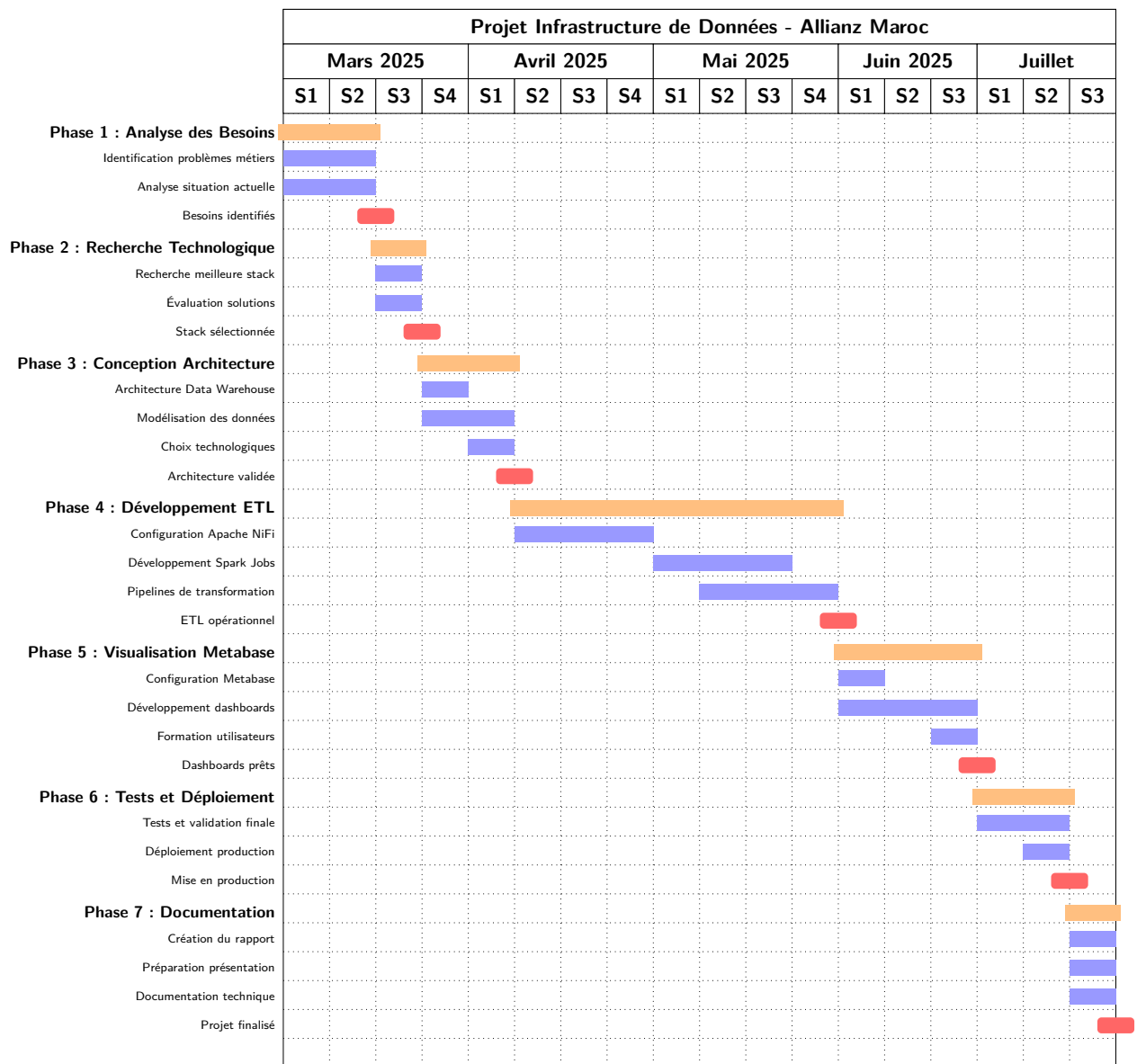


FIGURE 1.8 – Diagramme de Gantt du projet d'infrastructure de données Allianz Maroc

## 4.2 Diagramme de cas d'utilisation

Pour illustrer les interactions entre les différents acteurs et le système décisionnel proposé, le diagramme de cas d'utilisation ci-dessous présente les principaux scénarios d'utilisation du Data Warehouse et des outils de visualisation chez Allianz Maroc.

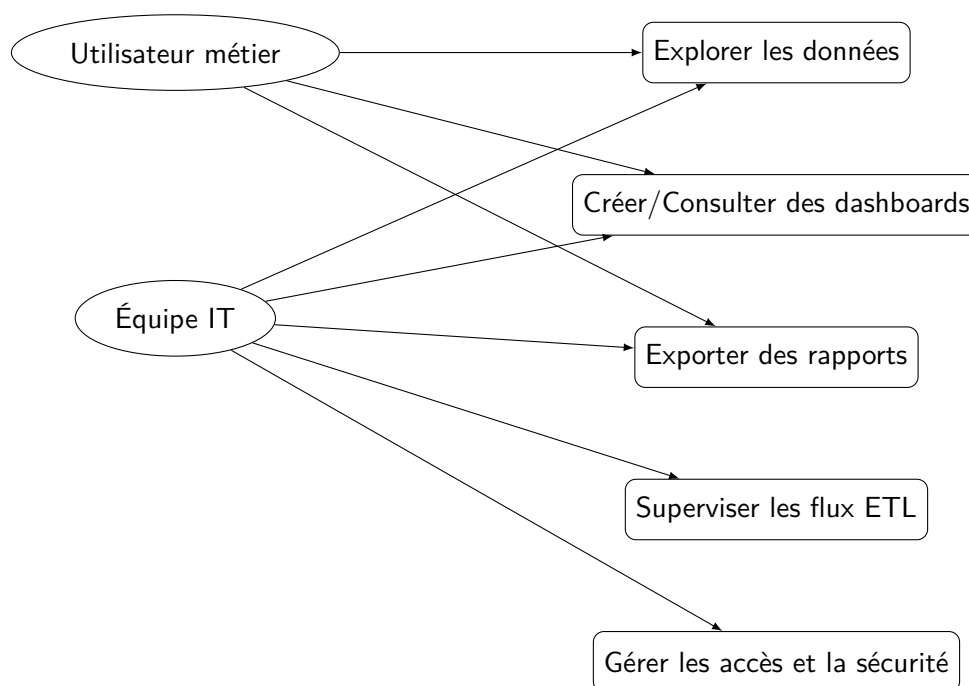


FIGURE 1.9 – Diagramme de cas d'utilisation du système décisionnel Allianz Maroc

## 5 Conclusion

Ce chapitre a permis de présenter le contexte général du projet, en décrivant l'organisation d'Allianz Maroc, ses principaux produits, ses valeurs et son environnement concurrentiel. Nous avons mis en évidence les difficultés rencontrées dans la gestion et l'exploitation des données, notamment la fragmentation des sources, l'accès manuel et l'absence de vision consolidée.

La problématique centrale réside dans la valorisation insuffisante d'un patrimoine de données pourtant important, ce qui limite la réactivité et la capacité décisionnelle de l'entreprise. Pour y répondre, une solution basée sur une architecture décisionnelle moderne a été proposée, intégrant l'automatisation des flux, le traitement distribué, la centralisation des données et la mise à disposition d'outils de visualisation accessibles à tous les métiers.

La démarche suivie, structurée autour de l'analyse de l'existant, du recueil des besoins, du choix des technologies et d'un déploiement progressif, vise à garantir l'adéquation de la solution aux attentes des utilisateurs et à assurer la réussite du projet.

Ce cadre organisationnel et technique servira de fondation pour les développements détaillés dans les chapitres suivants, où seront abordées la conception, la mise en œuvre et l'évaluation de l'infrastructure décisionnelle proposée.

# Chapitre 2

## Analyse et Modélisation

### Introduction

Ce chapitre constitue le socle fondamental de la conception du projet, en posant les bases méthodologiques et techniques qui guideront l'ensemble de la démarche. L'objectif principal est de traduire une ambition métier forte — à savoir l'optimisation et la valorisation de l'exploitation décisionnelle des données de sinistres — en une solution technique à la fois cohérente, robuste, pérenne et pleinement exploitable par les utilisateurs métiers.

Pour atteindre cette finalité, l'approche adoptée s'articule autour de trois grandes étapes structurantes, qui s'enchaînent de manière logique et progressive :

- **L'élaboration d'un cahier des charges détaillé** : cette première étape vise à formaliser de façon exhaustive l'ensemble des exigences fonctionnelles et des contraintes techniques qui pèseront sur le système cible. Il s'agit de recueillir, clarifier et prioriser les attentes des parties prenantes, afin de disposer d'un référentiel partagé qui guidera toutes les décisions ultérieures.
- **L'analyse approfondie des sources de données internes** : il s'agit ici d'identifier, de cartographier et de qualifier les différents gisements de données disponibles au sein du système d'information d'Allianz Maroc, dans le but d'en extraire les jeux de données les plus pertinents à intégrer et à modéliser pour répondre aux besoins métiers.
- **La conception du schéma du Data Warehouse** : cette étape consiste à traduire les besoins d'analyse identifiés en une architecture de données structurée, performante et évolutive, en veillant à garantir la lisibilité métier, la performance des traitements et la capacité d'évolution du modèle dans le temps.

L'ensemble de ces démarches vise à poser des fondations solides pour une architecture de données durable, capable de soutenir les usages métiers d'Allianz Maroc sur le long terme, tout en anticipant les évolutions futures du contexte et des besoins.

### 1 Cahier des charges

Le **cahier des charges** occupe une place centrale dans la réussite du projet. Il s'agit d'un document de référence qui formalise, de manière structurée et argumentée, les objectifs stratégiques et opérationnels du projet, ainsi que les besoins exprimés par l'ensemble des parties prenantes. Cette formalisation permet de traduire les attentes métiers en spécifications techniques précises,



exploitables par les équipes de développement et d'intégration. Le cahier des charges joue ainsi le rôle de fil conducteur tout au long du projet, conditionnant les choix d'architecture, d'outillage, de modélisation et de gouvernance.

**Les besoins majeurs identifiés se regroupent autour des axes suivants :**

— **Centralisation des données hétérogènes**

Le système cible doit être en mesure d'agréger et de consolider des données issues de multiples sources internes, telles que les systèmes AS/400, les bases de données Oracle ou encore des fichiers plats produits ponctuellement. L'objectif est de parvenir à unifier l'information relative aux sinistres dans un référentiel unique, cohérent, fiable et facilement interrogeable, afin de faciliter les analyses transverses et la prise de décision.

— **Automatisation de l'alimentation (ETL)**

L'industrialisation et l'automatisation des flux d'alimentation constituent un enjeu clé. Les processus d'extraction, de transformation et de chargement (ETL) doivent être orchestrés de manière automatique, avec une journalisation systématique des traitements. Cette automatisation vise à limiter les interventions manuelles, à réduire les risques d'erreur humaine et à garantir la régularité ainsi que la traçabilité des mises à jour.

— **Qualité et sécurité des données**

La fiabilité des analyses repose sur la qualité des données. Il est donc indispensable de mettre en place des mécanismes robustes de validation, de nettoyage, de contrôle de cohérence inter-systèmes et d'anonymisation des données sensibles. Ces dispositifs doivent permettre d'atteindre un niveau de confiance élevé sur l'ensemble des données exploitées, tout en respectant les exigences réglementaires et les standards du groupe.

— **Modèle décisionnel robuste**

Le cœur de la solution réside dans la conception d'un Data Warehouse bâti selon les meilleures pratiques de la modélisation dimensionnelle. Ce modèle doit faciliter les analyses exploratoires, permettre la restitution rapide et fiable d'indicateurs clés, et offrir une grande flexibilité pour l'évolution des besoins métiers.

— **Valorisation via des outils accessibles**

La donnée ne prend toute sa valeur que si elle est facilement accessible et exploitable par les utilisateurs métiers. Il est donc essentiel de proposer une interface de datavisualisation intuitive et conviviale (telle que Metabase), permettant aux utilisateurs de consulter, d'explorer et d'analyser les données en toute autonomie, sans dépendance vis-à-vis des équipes IT.

— **Scalabilité et évolutivité de l'architecture**

Enfin, la solution doit être conçue pour absorber de nouveaux flux de données, gérer l'augmentation des volumes et s'adapter à l'émergence de nouveaux cas d'usage, le tout sans nécessiter de refonte majeure de l'architecture existante. Cette capacité d'évolution est un gage de pérennité et d'agilité pour l'organisation.

**En résumé**, ce cahier des charges pose un cadre ambitieux et exigeant : il s'agit de construire une infrastructure de données **fiable, automatisée, sécurisée et évolutive**, capable de répondre efficacement aux besoins analytiques d'Allianz Maroc, tant à court qu'à moyen terme, tout en anticipant les évolutions futures du contexte métier et technologique.

**Note sur la notion de sinistre** : Dans le secteur de l'assurance, un **sinistre** désigne un événement dommageable (accident, vol, incendie, etc.) impliquant un assuré et donnant lieu à une déclaration auprès de l'assureur. La gestion d'un sinistre englobe l'ensemble des opérations liées à la prise en charge, l'évaluation, le règlement et la clôture du dossier correspondant. Dans ce rapport, le terme "sinistre" fait principalement référence aux dossiers d'accidents automobiles traités par Allianz Maroc.

## 2 Analyse des sources de données

La première étape opérationnelle du projet a consisté à réaliser une analyse approfondie des différentes sources de données internes disponibles au sein d'Allianz Maroc. Cette phase d'investigation a pour objectif d'identifier, de qualifier et de prioriser les flux de données les plus pertinents à intégrer dans le futur Data Warehouse. L'accent a été mis, dans le cadre de ce projet pilote, sur les données relatives aux **sinistres automobiles**, considérées comme un domaine stratégique et prioritaire pour l'organisation.

**Sélection orientée métier** : La sélection des extractions de données à intégrer n'a pas été réalisée de manière arbitraire. Elle a fait l'objet d'une collaboration étroite avec les équipes métier en charge de la gestion des sinistres, afin de s'aligner sur les besoins réels et récurrents en matière de reporting et d'analyse. Cette démarche a permis de s'appuyer à la fois sur les besoins exprimés par les utilisateurs et sur les extractions déjà mobilisées dans les processus d'analyse manuelle existants (règlements, réouvertures, ouvertures, clôtures, réserves, etc.), garantissant ainsi la pertinence et l'utilité des données sélectionnées.

**Les principales sources de données identifiées sont les suivantes :**

— **Systèmes IBM i (AS/400)**

Ces systèmes historiques hébergent l'essentiel de l'historique transactionnel des sinistres. L'accès aux données s'effectue via des **API internes sécurisées**, qui exposent des endpoints spécifiques pour l'extraction des flux critiques, tels que :

- Les règlements de sinistres automobiles
- Les ouvertures et clôtures par garantie
- Les sinistres en cours et les réouvertures

Les données issues de ces systèmes sont particulièrement riches, mais également complexes à manipuler, en raison de la multiplicité des champs métier (code garantie, type de recours,



```

1 Extraction reglement sinistre auto:
2 CODEINTE,INTERMEDIAIRE,VILLE_INTERMEDIAIRE,SINISTRE,POLICE,USAGE,CODNATSI,NATURE_SIN,DATE_SURVENANCE,DATEDECL,DATECCO,DATE_OUVREURE,
3 OUVREURE,DATE_CLOTURE,ASSURESP,TAUX_RESP_ASSURE,FLAG_HP,CANAL,LIEUSINI,VILLE_SINISTRE,PAYS,IMMATRICULATION_ASSURE,IDECONBA,LTB_BAREME,
4 FAIT_GENERATEUR,CAUSE_CIRCSTANCE,TYPE_RECOURS,COMPAGNIE_ADVERSE,NBR_VICTIME,GARANTIES,ETAPE,COEGARA,GARANTIE,RUBRIQUE,MNT_REGLEMENT,
5 MODE_PAIEMENT,TYPE_REGLEMENT,NATURE_REGLEMENT,DATESAIS,UTILSAIS,DATEVALI,UTILVALI,DATESIGN,UTILSIGN,STATUT_REG,STATUT_DOSS,
6 NOM_BENE,Extraction Reouvertures sinistres Auto
7 CODEINTE,INTERMEDIAIRE,VILLE_INTERMEDIAIRE,SINISTRE,COMMENTAIRE,POLICE,USAGE,CODNATSI,NATURE_SIN,DATE_SURVENANCE,DATEDECL,DATECCO,
8 DATE_OUVREURE,OUVERT_PAR,DATE_CLOTURE,ASSURESP,TAUX_RESP_ASSURE,FLAG_HP,CANAL,LIEUSINI,VILLE_SINISTRE,PAYS,IMMATRICULATION_ASSURE,IDECONBA,
9 LTB_BAREME,FAIT_GENERATEUR,CAUSE_CIRCSTANCE,TYPE_RECOURS,COMPAGNIE_ADVERSE,NBR_VICTIME,GARANTIES,ETAPE,COEGARA,GARANTIE,RUBRIQUE,
10 MNT_REGLEMENT,MODE_PAIEMENT,TYPE_REGLEMENT,NATURE_REGLEMENT,DATESAIS,UTILSAIS,DATEVALI,UTILVALI,DATESIGN,UTILSIGN,STATUT_REG,
11 STATUT_DOSS,NOM_BENE
12
13 Extraction Reouvertures sinistre auto par GARANTIE:
14 NUMESINI,CODNATSI,NUMPOLICE,DATESURV,ASSURESP,LIEUSINI,CODEINTE,INTERMEDIAIRE,ASSURE,IMMATRICULATION_ASSURE,DATE_OUVREURE,DATE_CLOTURE,
15 DATE_REOUVERTURE,USER_REOUVERTURE,COEGARA,MNT_RESERVE,MNT_REGLEMENT,MOTIF_CLOTURE
16
17
18 Extraction ouvertures sinistre Auto:
19 CODEINTE,INTERMEDIAIRE,VILLE_INTERMEDIAIRE,SINISTRE,SIN_GS,ASSURE,VILLE_ASSURE,IMMATRICULATION_ASSURE,MARQUE_VEHICULE,TYPE_VEHICULE,POLICE,
20 DATE_EFFECT,DATE_FIN,DATE_CLOTURE,USAGE,PRODUIT,CODNATSI,NATURE_SIN,DATE_SURVENANCE,DATEDECL,DATECCO,DATE_CREATION_OUVREURE,
21 DATE_VALIDATION_OUVREURE,USER_CREATION_OUVREURE,USER_VALIDATION_OUVREURE,DATE_CLOTURE,ASSURESP,TAUX_RESP_ASSURE,FLAG_HP,CANAL,LIEUSINI,
22 VILLE_SINISTRE,PAYS,IMMATRICULATION_ASSURE,IDECONBA,LTB_BAREME,FAIT_GENERATEUR,CAUSE_CIRCSTANCE,TYPE_RECOURS,COMPAGNIE_ADVERSE,NBR_VICTIME,
23 GARANTIES,COEGARA,MNT_REGLEMENT,P_MNT_REGLEMENT,H,ETAPE,FLAGICA,TELE,ATTESTATION,CIN,TYPE_SINISTRE,DOULEUX,DAGISTON,FINAL_INVESTIGATION,
24 MONTANT_RESERVE_DEBUT_P,MONTANT_RESERVE_DEBUT_H,MONTANT_RESERVE_P,USERGEST,LIBE_DERNIER_ACTE,DATE_DERNIER_ACTE,
25 MODE_PAIEMENT,SOUHAITE,FRANCHISE,DOSSCONT,MOTIF_CLOTURE
26
27
28 Extraction Cloture sinistre Auto:
29 NUMESINI,CODNATSI,NUMPOLICE,DATESURV,ASSURESP,LIEUSINI,CODEINTE,INTERMEDIAIRE,ASSURE,IMMATRICULATION_ASSURE,DATE_OUVREURE,DATE_CLOTURE,
30 USER_CLOTURE,COEGARA,MNT_RESERVE,MNT_REGLEMENT,MOTIF_CLOTURE
31
32
33 Extraction Cloture Sinistres Auto par garantie:
34 CODNATSI,NUMPOLICE,DATESURV,ASSURESP,LIEUSINI,CODEINTE,INTERMEDIAIRE,ASSURE,IMMATRICULATION_ASSURE,DATE_OUVREURE,DATE_CLOTURE,USER_CLOTURE,
35 COEGARA,MNT_RESERVE,MNT_REGLEMENT,MOTIFNUMESINI

```

FIGURE 2.1 – Exemple d'extraction de données sinistres – flux orienté métier

cause, immatriculation, etc.), de la faible documentation et de la nécessité d'échanges réguliers avec les experts fonctionnels pour en garantir la bonne interprétation.

#### — Bases de données Oracle

Ces bases contiennent des référentiels structurés, tels que les clients, les véhicules ou les produits, qui sont utilisés en appui des processus métier. L'extraction des données s'effectue par le biais de requêtes SQL ou de vues spécifiques. Ces sources permettent d'enrichir les données de sinistres avec des informations complémentaires, telles que :

- L'usage et la nature du véhicule
- Le profil du client
- La structure des polices d'assurance

#### — Fichiers plats (CSV / TXT)

Il s'agit d'extractions ponctuelles, générées par les équipes IT à des fins de contrôle interne, d'audit ou de reporting. Ces fichiers couvrent des flux variés, tels que :

- Les avances assurés
- Les réserves historiques depuis 2014
- Les sinistres clôturés non indemnisés

L'intégration de ces fichiers nécessite une étape préalable de **normalisation des formats**, de **détection des doublons** et de cartographie des colonnes métier, qui sont parfois non typées ou non nommées, ce qui complexifie le travail de préparation.

## Synthèse sur la diversité des sources de données

Il est important de souligner que la multiplicité des sources de données n'est pas le fruit du hasard, mais résulte directement de la complexité et de la diversité des informations nécessaires à la gestion complète d'un sinistre. En effet, chaque aspect ou détail d'un sinistre est souvent géré par un système ou une application métier spécifique, ce qui explique la dispersion des données :

- **Les systèmes transactionnels (AS/400)** centralisent l'historique des mouvements fi-

nanciers et opérationnels liés aux sinistres (règlements, ouvertures, clôtures, réouvertures). Ces systèmes sont optimisés pour la gestion quotidienne des dossiers et la traçabilité des opérations.

- **Les bases de données Oracle** hébergent les référentiels clients, véhicules et produits, qui permettent d'enrichir chaque sinistre avec des informations contextuelles (profil du client, caractéristiques du véhicule, type de contrat, etc.).
- **Les fichiers plats** sont utilisés pour des besoins ponctuels de reporting, d'audit ou de contrôle interne, et contiennent des extractions spécifiques (avances assurés, réserves historiques, sinistres non indemnisés) qui ne sont pas toujours disponibles dans les systèmes principaux.

Ainsi, pour reconstituer une vision complète et fiable d'un sinistre, il est indispensable de croiser et d'intégrer ces différentes sources. Chaque source apporte une pièce essentielle du puzzle : les mouvements financiers, les informations contractuelles, les données techniques du véhicule, ou encore les historiques d'événements. Cette approche garantit une analyse exhaustive et pertinente, tout en permettant de répondre aux exigences métiers les plus fines (calculs de KPI, analyses par segment, suivi des délais, etc.).

### **Problématiques rencontrées lors de l'analyse des sources :**

- *Granularité variable* : les extractions disponibles présentent des niveaux de granularité différents (certaines sont au niveau du sinistre, d'autres au niveau du mouvement ou de la garantie), ce qui impose un travail d'harmonisation pour permettre des analyses cohérentes.
- *Temporalité désynchronisée* : les différentes sources ne sont pas toujours actualisées à la même fréquence, ce qui peut générer des écarts temporels et nécessite un alignement précis des flux pour garantir la fiabilité des analyses.
- *Qualité inégale des données* : selon les sources, on observe des valeurs aberrantes, des formats incohérents (par exemple, des dates inversées ou des champs codifiés non documentés), ce qui impose la mise en place de contrôles et de traitements de qualité rigoureux.

Malgré ces contraintes et ces défis, l'approche méthodique adoptée a permis de **cibler les flux prioritaires**, de poser les bases d'une ingestion automatisée des données via NiFi, et de préparer dans de bonnes conditions la modélisation d'un entrepôt de données décisionnel dédié à la gestion des sinistres.

## **3 Modélisation du Data Warehouse**

La phase de modélisation du Data Warehouse constitue une étape clé du projet, car elle conditionne la performance, la lisibilité et l'évolutivité de l'ensemble du dispositif décisionnel.

Le choix s'est porté sur un **schéma en étoile** (*Star Schema*), reconnu pour sa robustesse et son efficacité dans les systèmes décisionnels. Ce type de modélisation permet une lecture simplifiée des données, un accès optimisé aux mesures clés, et une intégration fluide avec les outils de visualisation tels que **Metabase**. Le périmètre modélisé couvre exclusivement les **sinistres automobiles**, en s'appuyant sur un sous-ensemble pertinent d'extractions issues des systèmes d'information d'Allianz Maroc.

### 1. Choix du modèle en étoile

Le modèle en étoile a été privilégié pour plusieurs raisons fondamentales, qui répondent aux attentes des utilisateurs métiers et aux exigences techniques du projet :

- Simplicité fonctionnelle : la structure du modèle permet aux utilisateurs métiers d'interroger les données selon des axes d'analyse clairs et intuitifs (temps, lieu, garantie, véhicule), facilitant ainsi l'exploration et la compréhension des informations.
- Performance des requêtes : en limitant les jointures à une table centrale (table de faits) et à ses dimensions, le modèle réduit la complexité des traitements et optimise les temps de réponse lors des analyses.
- Adaptabilité analytique : le modèle en étoile facilite la création d'agrégats, de regroupements et de filtres, ce qui est essentiel pour la production de KPI et la réalisation d'analyses exploratoires ou récurrentes.

### 2. Description du modèle conceptuel

Le Data Warehouse s'articule autour d'une table centrale, `Faits_Sinistres`, qui regroupe l'ensemble des indicateurs clés issus de plusieurs extractions (règlements, réserves, ouvertures, clôtures, réouvertures). Cette table de faits est reliée à un ensemble de dimensions métier, qui permettent de contextualiser et d'enrichir les analyses.

Table de faits : `Faits_Sinistres`

La table de faits enregistre les mesures atomiques associées à chaque sinistre (ou mouvement), parmi lesquelles :

- `montant_reglement` (extrait de `MNT_REGLEMENT`)
- `montant_reserve` (extrait de `MNT_RESERVE`)
- `taux_responsabilite`
- `statut_dossier`
- `nb_victimes`
- `delai_traitement` (calculé entre la date d'ouverture et la date de clôture)
- `date_ouverture`, `date_cloture`, `date_reouverture`
- `id_client`, `id_garantie`, `id_vehicule`, `id_canal`, `id_zone`, `id_date`

Dimensions principales :

- Dim\_Date : dimension temporelle structurée avec des hiérarchies jour / mois / année, permettant des analyses à différents niveaux de granularité.
- Dim\_Client : regroupe les informations sur l'assureur ou l'assuré, incluant l'immatriculation, l'identité de l'assuré, l'usage du véhicule et le produit souscrit.
- Dim\_Garantie : basée sur les champs GARANTIE et CODEGARA, elle recense les différents types de garanties souscrites (vol, tous risques, etc.).
- Dim\_Véhicule : extraite des champs MARQUE\_VEHICULE, TYPE\_VEHICULE, IMMATRICULATION, elle permet d'analyser les sinistres selon les caractéristiques du véhicule impliqué.
- Dim\_ZoneGeographique : correspond aux informations géographiques telles que VILLE\_SINISTRE, LIEUSINI, PAYS, pour des analyses spatiales.
- Dim\_Canal : identifie le canal de déclaration du sinistre (ex. AGENCE, INTERMEDIAIRE, WEB, COURTIER), utile pour l'analyse des modes de contact.

Mapping avec les sources réelles :

- Faits\_Sinistres : construite à partir des extractions de règlements, réserves, ouvertures, clôtures, réouvertures.
- montant\_reglement  $\leftarrow$  MNT\_REGLEMENT
- montant\_reserve  $\leftarrow$  MNT\_RESERVE
- taux\_responsabilite  $\leftarrow$  TAUX\_RESP\_ASSURE
- Dim\_Garantie  $\leftarrow$  GARANTIE, CODEGARA
- Dim\_Véhicule  $\leftarrow$  TYPE\_VEHICULE, MARQUE\_VEHICULE
- Dim\_ZoneGeographique  $\leftarrow$  VILLE\_SINISTRE, LIEUSINI, PAYS
- Dim\_Canal  $\leftarrow$  CANAL, INTERMEDIAIRE

Visualisation du modèle conceptuel :

Ce modèle dimensionnel a été implémenté dans une base **PostgreSQL** et interfacé avec **Metabase**, afin d'assurer une restitution visuelle fluide, accessible en mode self-service. Cette architecture garantit à la fois la performance des analyses, la scalabilité du dispositif et l'accessibilité des données pour les utilisateurs métiers, tout en facilitant l'évolution future du modèle.

## 4 Qualité des données et gouvernance

Dans un contexte où la donnée s'impose comme un actif stratégique majeur, il est impératif d'accorder une attention particulière à sa **qualité**, à sa **traçabilité** et à la mise en place d'une **gouvernance rigoureuse**. Ces dimensions sont essentielles pour garantir la fiabilité des analyses, la conformité réglementaire et la confiance des utilisateurs dans les résultats produits. Allianz Maroc, pleinement consciente de ces enjeux, a intégré ces préoccupations au cœur du projet dès

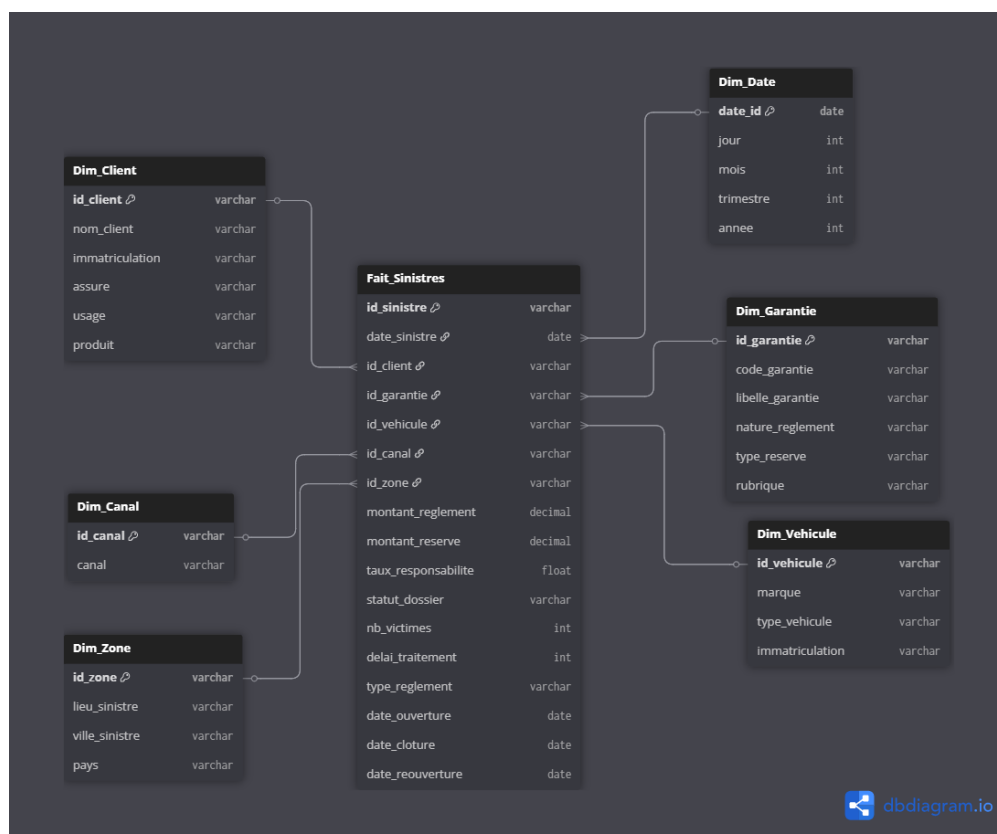


FIGURE 2.2 – Schéma conceptuel du Data Warehouse – Sinistres Automobiles

ses premières phases.

## 1. Démarche de gestion de la qualité des données

La gestion de la qualité des données s'est articulée autour de trois axes prioritaires, qui ont guidé l'ensemble des travaux d'intégration et de préparation des données :

- **Nettoyage et validation des données en amont** : chaque flux de données entrant a été soumis à une série de traitements automatiques et manuels visant à éliminer les **doublons**, à corriger les **formats incorrects** (notamment pour les dates et les montants), et à identifier les **valeurs manquantes** ou aberrantes. Cette étape est cruciale pour garantir la fiabilité des analyses et la pertinence des indicateurs produits.
- **Tests de cohérence inter-systèmes** : une attention particulière a été portée à la **réconciliation des données** issues de sources multiples (AS/400, Oracle, fichiers plats). Des contrôles croisés ont permis de détecter et de corriger des **incohérences de clés**, des **référentiels désalignés** ou encore des **écarts de volumétrie** entre les différents systèmes, assurant ainsi l'intégrité globale du référentiel.
- **Traçabilité des traitements** : grâce à l'utilisation d'**Apache NiFi**, chaque étape des pipelines de données (extraction, transformation, chargement) est systématiquement journalisée. Cette traçabilité garantit une visibilité complète sur le cycle de vie de la donnée, permettant de remonter à l'**origine exacte de chaque enregistrement** présent dans le

Data Warehouse, et de répondre ainsi aux exigences de transparence et d'auditabilité.

### 2. Sécurité et protection de l'information

La **sécurité des données** constitue un pilier fondamental du dispositif, et repose sur une combinaison de mesures techniques et organisationnelles complémentaires :

- Mise en place d'un contrôle des accès basé sur des **profils utilisateurs spécifiques** (lecture seule, contributeur, administrateur), afin de garantir que chaque utilisateur dispose uniquement des droits nécessaires à l'exercice de ses missions.
- Déploiement de politiques de **sécurisation réseau et d'authentification** renforcée pour l'ensemble des outils utilisés (Metabase, PostgreSQL, accès API), afin de prévenir tout accès non autorisé ou toute fuite de données.
- Application d'une **anonymisation sélective** des données sensibles (noms, numéros de police, etc.), en conformité avec les principes du RGPD et les standards du groupe Allianz, pour protéger la vie privée des assurés et limiter les risques juridiques.
- Réalisation de **sauvegardes régulières et contrôlées** des données et des configurations, afin d'assurer la résilience du système et la capacité de reprise en cas d'incident.

### 3. Fondations d'une gouvernance de la donnée

Au-delà de la qualité opérationnelle, une **stratégie de gouvernance** a été amorcée pour structurer durablement la gestion des données et garantir leur valorisation sur le long terme. Cette gouvernance s'appuie sur plusieurs leviers complémentaires :

- **Gestion des habilitations dans Metabase** : l'accès aux dashboards, aux requêtes SQL et aux collections de données est strictement **restreint par rôle métier**, ce qui permet de garantir la confidentialité des données stratégiques tout en favorisant l'autonomie des analystes et des utilisateurs avancés.
- **Standardisation du modèle de données** : des conventions de nommage strictes et homogènes ont été adoptées pour les tables, les colonnes, les mesures et les dimensions, facilitant ainsi la compréhension, la documentation et l'évolutivité du modèle, tout en réduisant les risques d'ambiguïté ou d'erreur d'interprétation.
- **Documentation unifiée et partagée** : un espace dédié au sein de Metabase regroupe l'ensemble des définitions métiers, des règles de gestion, des mappings de sources et des descriptions des transformations ETL. Cet espace fait office de **catalogue de données initial**, accessible à toutes les parties prenantes, et constitue un socle de référence commun pour l'ensemble des utilisateurs.
- **Processus de gestion du changement** : toute modification du modèle de données, ajout de nouvelles sources ou évolution d'un indicateur fait l'objet d'une **validation croisée**



entre les équipes IT et métiers, avec un système de **versioning** et de suivi dans GitHub, garantissant ainsi la traçabilité et la maîtrise des évolutions.

### 4. Vers une gouvernance orientée self-service

Cette première implémentation marque une étape structurante vers une **gouvernance des données moderne et décentralisée**, pleinement alignée avec les principes de la **démocratisation de la donnée**. L'objectif est de **réduire la dépendance vis-à-vis de l'IT** pour l'accès à l'information, et de favoriser l'émergence d'une culture **data-driven** fondée sur l'autonomie et la responsabilisation des équipes métiers.

- La plateforme Metabase offre aux utilisateurs la possibilité d'une **exploration libre et intuitive des données**, via des interfaces ergonomiques ne nécessitant pas de compétences techniques avancées. Chacun peut ainsi construire ses propres requêtes, visualisations et tableaux de bord, en fonction de ses besoins spécifiques.
- Des **règles de sécurité et d'habilitation fines** assurent que chaque profil dispose d'un accès strictement adapté à ses responsabilités, tout en garantissant la confidentialité et la protection des données sensibles.
- La documentation intégrée à la plateforme (définitions métiers, règles de gestion, description des indicateurs) constitue un **socle de référence partagé**, qui facilite la compréhension, l'appropriation et l'usage autonome des données par l'ensemble des utilisateurs.
- À moyen terme, il est envisagé de renforcer ce dispositif par l'introduction d'un **catalogue de données collaboratif** (de type Amundsen, DataHub), qui facilitera la découverte, la classification et la réutilisation des jeux de données en mode self-service, tout en favorisant la collaboration entre les différents acteurs de la donnée.

En favorisant l'**autonomie des utilisateurs métiers** tout en maintenant un cadre structurant et sécurisé, cette approche ouvre la voie à une exploitation plus fluide, agile et réactive de la donnée au sein d'Allianz Maroc, et contribue à faire de la donnée un véritable levier de performance collective.

Cette démarche vise à faire de la donnée un **actif maîtrisé, sécurisé et valorisé**, au service de la performance opérationnelle et stratégique d'Allianz Maroc, et à inscrire l'organisation dans une dynamique d'amélioration continue et d'innovation.

## 5 Conclusion

Ce chapitre a permis de poser et de structurer l'ensemble des fondations nécessaires à la mise en place d'une solution décisionnelle adaptée aux besoins spécifiques d'Allianz Maroc. Après avoir défini un cahier des charges précis et ambitieux, l'analyse approfondie des sources de données internes a mis en lumière la diversité, la richesse mais aussi la complexité des flux à intégrer. La

démarche adoptée a insisté sur la qualité, la traçabilité et la gouvernance des données, afin de garantir la fiabilité, la sécurité et la pérennité des analyses futures.

La modélisation du Data Warehouse, centrée sur les sinistres automobiles, offre un cadre structuré, performant et évolutif pour l'exploitation des données. Ce socle analytique facilite l'accès à l'information, accélère la prise de décision et prépare le terrain pour les étapes suivantes du projet, notamment la mise en œuvre technique, l'intégration des outils et la valorisation des premiers résultats.

Le chapitre suivant abordera en détail la mise en œuvre technique de cette architecture, en explicitant les choix d'outils, les processus d'intégration, les bonnes pratiques adoptées et les premiers bénéfices observés, afin d'illustrer concrètement la valeur ajoutée de la démarche engagée.

# Chapitre 3

## Étude technique

### Introduction

Ce chapitre propose une exploration approfondie de l'étude technique qui a guidé la conception et la mise en œuvre de la solution. Il s'agit d'un parcours structuré, allant de l'analyse comparative des différentes technologies disponibles sur le marché jusqu'à la définition détaillée de l'architecture finale retenue. L'objectif principal de cette démarche est de justifier, de manière argumentée et transparente, chacun des choix technologiques opérés, en tenant compte des besoins spécifiques d'Allianz Maroc. La cohérence globale, la performance opérationnelle et l'évolutivité de la solution ont été placées au cœur de la réflexion, afin de garantir la pérennité et l'adaptabilité du système face aux évolutions futures.

## 1 Analyse comparative et choix technologiques

La sélection de la stack technologique s'est appuyée sur une analyse méthodique et rigoureuse des solutions existantes, en les comparant selon leur catégorie fonctionnelle (ingestion, traitement, stockage, visualisation). Chaque décision a été motivée par des critères objectifs, étroitement liés aux exigences du projet et aux contraintes du contexte métier. Cette démarche comparative a permis de mettre en lumière les avantages et les limites de chaque outil, afin de retenir les solutions les plus pertinentes et les mieux adaptées à l'écosystème d'Allianz Maroc.

### 1.1 Ingestion et orchestration : Apache NiFi

Pour la gestion de l'ingestion et de l'orchestration des flux de données, le choix s'est porté sur **Apache NiFi**. Ce dernier s'est distingué par sa capacité à offrir une interface graphique intuitive pour la conception et la gestion des workflows, ce qui facilite grandement la prise en main et la maintenance des processus d'intégration. Sa flexibilité, ainsi que sa robustesse dans le traitement de volumes importants de données en temps réel, en font un outil de référence. Contrairement à Apache Airflow, qui nécessite une définition des workflows en Python et s'adresse davantage à des profils techniques, NiFi permet une modélisation visuelle accessible et rapide. Talend, bien que performant, reste une solution propriétaire, impliquant des coûts de licence non négligeables et une dépendance vis-à-vis de l'éditeur.

Critère	Apache NiFi	Apache Airflow	Talend
Interface graphique	Oui, intuitive	Non, code Python	Oui, propriétaire
Temps réel	Excellente	Limitée	Moyenne
Facilité d'utilisation	Très bonne	Difficile	Bonne
Coût	Gratuit	Gratuit	Payant
Communauté	Très bonne	Excellente	Limitée
Scalabilité	Très bonne	Bonne	Bonne

TABLE 3.1 – Comparatif des outils d'ingestion de données

## 1.2 Traitement des données : Apache Spark

En ce qui concerne le traitement des données, **Apache Spark** a été retenu comme la solution la plus adaptée. Spark se distingue par ses performances élevées en traitement distribué, sa capacité à gérer de grands volumes de données et sa compatibilité avec une large variété de formats. Il permet d'exécuter des traitements complexes, aussi bien en mode batch qu'en temps réel, tout en offrant un écosystème riche et mature. Par rapport à Hadoop, Spark se montre nettement plus rapide et flexible, tandis que Flink, bien qu'efficace pour le streaming, reste moins répandu et dispose d'une communauté plus restreinte. Le choix de Spark s'inscrit donc dans une logique de performance, de scalabilité et de pérennité.

Critère	Apache Spark	Apache Hadoop	Apache Flink
Performance	Très élevée	Moyenne	Élevée
Temps réel	Excellent	Non	Excellent
Facilité d'utilisation	Bonne	Difficile	Moyenne
Écosystème	Très riche	Riche	En développement
Compatibilité formats	Excellente	Bonne	Bonne
Courbe d'apprentissage	Moyenne	Élevée	Élevée

TABLE 3.2 – Comparatif des outils de traitement de données

## 1.3 Stockage : PostgreSQL

Pour la persistance et l'historisation des données, le choix s'est naturellement orienté vers **PostgreSQL**. Ce système de gestion de base de données relationnelle open source est reconnu pour sa robustesse, sa fiabilité et la richesse de ses fonctionnalités avancées, telles que la gestion des transactions, le support natif du format JSON ou encore la possibilité d'effectuer des requêtes analytiques complexes. PostgreSQL se démarque également par sa forte intégration avec les outils analytiques et sa communauté très active. Comparativement, MySQL présente des capacités analytiques plus limitées, tandis qu'Oracle, bien que très performant, implique des coûts de licence importants et une complexité de gestion accrue.

Critère	PostgreSQL	MySQL	Oracle
Fonctionnalités avancées	Excellentes	Bonnes	Excellentes
Performance analytique	Très bonne	Moyenne	Excellente
Coût	Gratuit	Gratuit/Payant	Payant
Support JSON	Natif	Basique	Bon
Communauté	Très active	Très active	Professionnelle

TABLE 3.3 – Comparatif des solutions de stockage

## 1.4 Visualisation : Metabase

Pour la restitution et la valorisation des données, **Metabase** a été privilégié. Cet outil open source se distingue par son interface utilisateur intuitive, sa simplicité d'installation et sa capacité à rendre la donnée accessible à un large public, y compris aux utilisateurs non techniques. Metabase permet de créer rapidement des tableaux de bord interactifs et des visualisations pertinentes, favorisant ainsi la démocratisation de l'accès à l'information. À l'inverse, Tableau et Power BI, bien que très puissants et riches en fonctionnalités, nécessitent l'acquisition de licences et une courbe d'apprentissage plus importante, ce qui peut constituer un frein à l'adoption rapide.

Critère	Metabase	Tableau	Power BI
Facilité d'utilisation	Excellente	Bonne	Bonne
Coût	Gratuit/Payant	Payant	Payant
Installation	Simple	Complexe	Moyenne
Connecteurs	Nombreux	Très nombreux	Nombreux
Performance	Bonne	Excellente	Bonne
Personnalisation	Moyenne	Excellente	Bonne

TABLE 3.4 – Comparatif des outils de visualisation

## 2 Dataflow de la solution

Le *dataflow* représente le cheminement des données à travers les différentes étapes du système, depuis leur collecte initiale jusqu'à leur restitution finale sous forme d'indicateurs ou de tableaux de bord. Il s'agit d'une succession logique de phases, indépendantes des technologies sous-jacentes, qui structurent le traitement de la donnée : extraction, ingestion, transformation, stockage et visualisation. Cette vision globale permet de mieux appréhender la circulation de l'information et d'identifier les points clés de contrôle et d'optimisation.

Le schéma ci-dessus illustre les principales étapes du parcours des données au sein de la solution :

- **Sources de données** : Les données sont extraites depuis différentes sources, qu'il s'agisse de bases de données, d'API externes ou de fichiers plats.

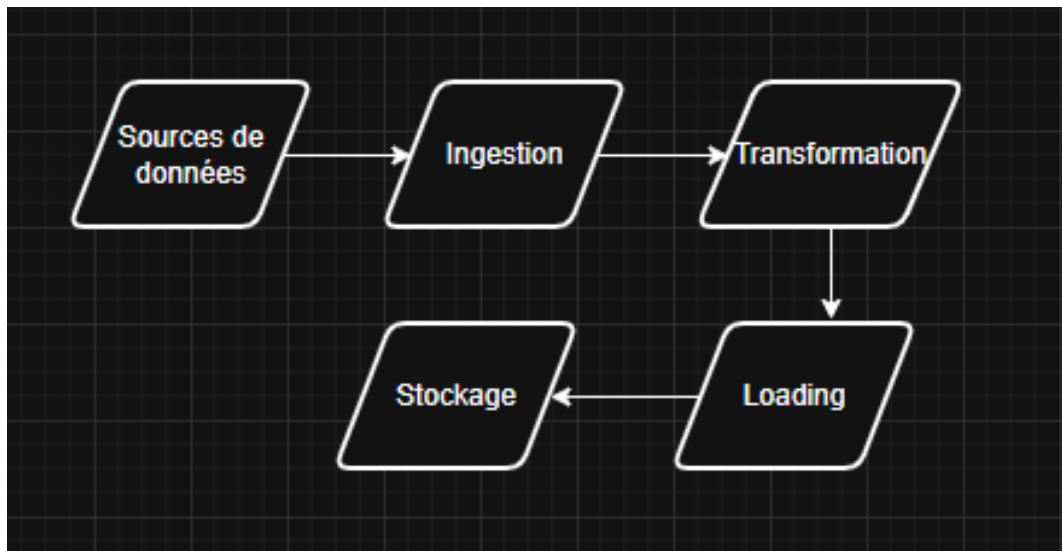


FIGURE 3.1 – Schéma du dataflow de la solution

- **Ingestion** : Ces données sont ensuite centralisées et orchestrées à l'aide d'Apache NiFi, qui assure la fluidité et la traçabilité des flux entrants.
- **Traitement** : Une fois ingérées, les données sont nettoyées, transformées et enrichies grâce à Apache Spark, afin de garantir leur qualité et leur pertinence pour les usages métiers.
- **Stockage** : Les données traitées sont stockées dans PostgreSQL, ce qui permet leur historisation, leur sécurisation et leur exploitation analytique.
- **Visualisation** : Enfin, les indicateurs et tableaux de bord sont restitués via Metabase, offrant ainsi une visibilité claire et synthétique aux utilisateurs finaux.

### 3 Architecture finale

L'architecture finale de la solution s'articule autour de quatre couches complémentaires, chacune jouant un rôle déterminant dans la chaîne de valeur de la donnée. Cette structuration en couches permet d'assurer une séparation claire des responsabilités, de faciliter la maintenance et l'évolution du système, et d'optimiser la performance globale.

- **Ingestion** : Cette couche assure la centralisation des données issues de sources hétérogènes (bases de données, API, fichiers plats) à l'aide d'**Apache NiFi**. Elle garantit la flexibilité, la traçabilité et la sécurité des flux de données entrants, tout en permettant une adaptation rapide aux évolutions des sources.
- **Traitement** : La transformation, le nettoyage et l'enrichissement des données sont réalisés avec **Apache Spark**. Cette couche permet d'exécuter des traitements distribués, performants et évolutifs, adaptés aux volumes croissants et à la diversité des besoins métiers.
- **Stockage** : L'historisation et la sécurisation des données sont assurées par un entrepôt **PostgreSQL**, qui offre robustesse, fiabilité et compatibilité avec les outils analytiques.

Cette couche constitue le socle de la valorisation de la donnée.

- **Visualisation** : Enfin, la création de tableaux de bord interactifs et la restitution des indicateurs sont réalisées avec **Metabase**. Cette couche facilite la prise de décision et démocratise l'accès à l'information pour l'ensemble des métiers de l'entreprise.

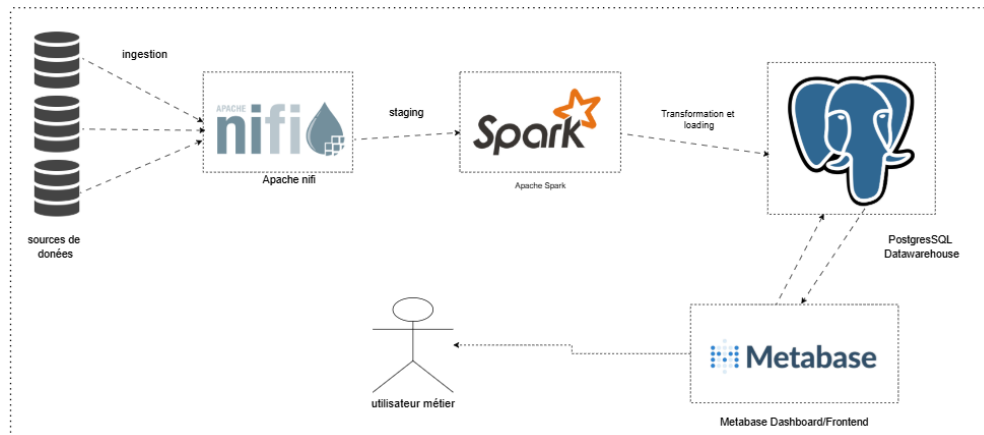


FIGURE 3.2 – Architecture finale de la solution

Cette architecture, à la fois modulaire et ouverte, offre une grande souplesse d'évolution. Elle permet d'ajouter facilement de nouveaux modules fonctionnels, de connecter des sources de données supplémentaires ou d'adapter les traitements en fonction des nouveaux besoins métiers. Par ailleurs, elle intègre nativement des mécanismes de sécurité essentiels, tels que le contrôle d'accès, le chiffrement des données et la traçabilité des opérations. La performance est également assurée grâce à la distribution des traitements et à la robustesse de la couche de stockage, garantissant ainsi la disponibilité et la fiabilité du système.

## 4 Conclusion

L'étude technique menée a permis de sélectionner, de manière argumentée et structurée, une stack technologique cohérente, performante et parfaitement alignée avec les enjeux stratégiques d'Allianz Maroc. L'association d'Apache NiFi, Apache Spark, PostgreSQL et Metabase constitue une solution modulaire, évolutive et sécurisée, capable de répondre efficacement aux besoins actuels tout en anticipant les évolutions futures du contexte métier. Chaque choix technologique a été motivé par des critères précis, garantissant la robustesse, la scalabilité et la facilité de maintenance de l'ensemble du système d'information.

Cette architecture constitue ainsi une base solide pour la mise en œuvre opérationnelle du projet. Elle facilite l'intégration de nouvelles sources de données, l'automatisation des traitements et la valorisation de l'information à travers des outils de visualisation adaptés aux différents profils d'utilisateurs. De plus, elle permet d'assurer la conformité aux exigences de sécurité et de gouvernance des données, qui sont des enjeux majeurs pour une entreprise telle qu'Allianz Maroc.

Le chapitre suivant viendra détailler la phase de réalisation, en présentant de manière concrète

les étapes de mise en place de la solution, les choix d'implémentation retenus ainsi que les principaux défis rencontrés lors du développement. Cette approche progressive permettra de mieux comprendre la logique de construction du projet et d'apprécier la pertinence des choix techniques opérés.



# Chapitre 4

## Réalisation et déploiement

### Introduction

Ce chapitre constitue une étape clé du mémoire, puisqu'il expose en détail la phase de réalisation concrète du projet, en mettant en lumière l'ensemble des étapes techniques et organisationnelles qui ont permis de passer de la conception à la mise en œuvre effective de la solution. L'objectif de cette partie est de présenter, de manière structurée et exhaustive, la démarche suivie pour transformer les choix d'architecture et les spécifications fonctionnelles en un système opérationnel, robuste et sécurisé, parfaitement adapté aux besoins spécifiques d'Allianz Maroc.

Plus précisément, ce chapitre s'articule autour de plusieurs axes majeurs : la création d'un prototype fonctionnel (Proof of Concept, ou PoC), le développement et l'industrialisation de la pipeline de données, ainsi que la mise en place des pratiques DevOps pour le déploiement, l'automatisation et la maintenance de l'ensemble de la solution. Chaque section mettra en avant les outils, technologies et méthodes employés, tout en explicitant les raisons des choix effectués et les bénéfices attendus en termes de performance, de fiabilité et de sécurité.

L'ensemble de ces éléments vise à démontrer comment la solution technique a été pensée, construite et déployée pour répondre aux exigences d'Allianz Maroc, tout en anticipant les besoins d'évolutivité, de scalabilité et de gouvernance des données à long terme.

### 1 Création du PoC

Avant d'entamer le traitement de données sensibles ou réelles, il a été primordial de valider la faisabilité technique du projet à travers la réalisation d'un prototype fonctionnel, ou Proof of Concept (PoC). Cette étape préliminaire a permis de simuler l'ensemble de la chaîne de traitement, d'anticiper les éventuels points de blocage, et d'identifier les dépendances critiques susceptibles d'impacter le bon déroulement du projet.

Pour ce faire, la première action a consisté à générer un jeu de données synthétiques, mais réalistes, en s'appuyant sur la structure d'extraction fournie par l'équipe métier. À l'aide d'un script Python développé sur mesure et reposant sur la librairie `Faker`, il a été possible de créer des fichiers CSV mimant fidèlement la structure et les principaux champs des extractions réelles. Les tables simulées incluaient notamment les règlements, les ouvertures, les clôtures et les réouvertures de sinistres, ce qui a permis de couvrir l'ensemble des cas d'usage métier visés.

Ces données factices ont ensuite été injectées dans Apache Spark, via un job de transformation initial spécialement conçu pour cette phase de test. Ce job avait pour objectif de nettoyer les données (suppression des valeurs nulles ou incohérentes), d'homogénéiser les formats (notamment les dates et les types de champs), et d'enrichir les jeux de données par des mappings de statuts ou des normalisations de champs. Cette étape a permis de valider la capacité de Spark à traiter efficacement des volumes de données similaires à ceux attendus en production.

À l'issue de la transformation, les données ont été chargées dans une instance PostgreSQL de test, structurée selon un schéma simplifié de data warehouse. Cette base de données a ensuite été connectée à Metabase, ce qui a permis de construire un premier tableau de bord de test, centré sur un profil utilisateur fictif. Cette démarche a permis de valider l'intégration de bout en bout, depuis la génération des données jusqu'à la visualisation.

Le PoC a ainsi permis de valider plusieurs points essentiels :

- Le bon enchaînement des différentes étapes du pipeline : ingestion, transformation, stockage et visualisation des données.
- La cohérence des formats de données et la compatibilité entre les différents composants techniques (Spark, PostgreSQL, Metabase).
- L'automatisation complète d'un pipeline simple, démontrant la faisabilité d'une solution industrialisée.

Les enseignements tirés de cette phase de prototypage ont constitué une base solide pour aborder sereinement le traitement des vraies données et le déploiement en environnement contrôlé, tout en limitant les risques techniques.

## 2 Création de la pipeline de données

Suite à la validation du PoC et à sa présentation à l'équipe IT d'Allianz Maroc, le projet est entré dans une phase de développement sur les données réelles issues des systèmes internes de l'entreprise. L'objectif principal était alors de concevoir et de mettre en œuvre une pipeline de données robuste, automatisée et maintenable, capable de traiter l'ensemble des flux de données depuis les sources jusqu'à la visualisation finale dans Metabase.

### 2.1 Ingestion des données avec Apache NiFi

L'ingestion des données a été confiée à Apache NiFi, un outil reconnu pour sa capacité à automatiser et orchestrer les flux de données de manière graphique et modulaire. Deux grandes familles de sources de données ont été prises en charge dans le cadre du projet :

- **Base de données Oracle** : L'ingestion des données depuis la base Oracle s'appuie sur un flow NiFi périodique, orchestré comme suit :
  1. `GenerateFlowFile` sert de minuteur pour déclencher le processus à intervalles réguliers.

2. ExecuteSQL interroge la base Oracle et extrait les enregistrements requis (sinistres, contrats, clients, etc.).
3. PutFile stocke temporairement les résultats dans une zone de staging locale, assurant la traçabilité et la reprise en cas d'incident.
4. ExecuteStreamCommand lance un script Spark pour le nettoyage, l'enrichissement et le chargement des données dans PostgreSQL.
5. Des LogAttribute sont placés à chaque étape pour journaliser les événements et faciliter le suivi du flux.

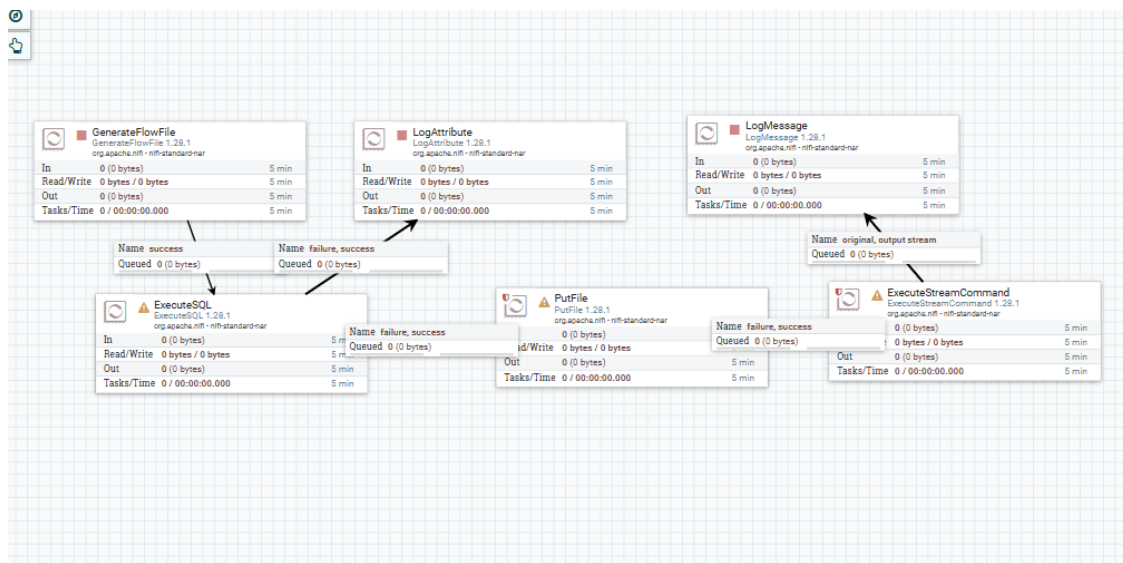


FIGURE 4.1 – Flow d'ingestion NiFi pour la base Oracle

— **Base de données AS/400** : L'ingestion depuis le système AS/400 repose sur un flow NiFi dédié, structuré ainsi :

1. InvokeHTTP effectue une requête vers l'API interne pour récupérer les données brutes au format JSON.
2. SplitJson fragmente la réponse pour traiter chaque enregistrement individuellement.
3. EvaluateJsonPath extrait les champs pertinents (identifiants, montants, dates, etc.).
4. ExecuteStreamCommand déclenche un script Spark pour le nettoyage, l'enrichissement et le chargement dans PostgreSQL.
5. Des LogAttribute assurent la supervision et la traçabilité à chaque étape.

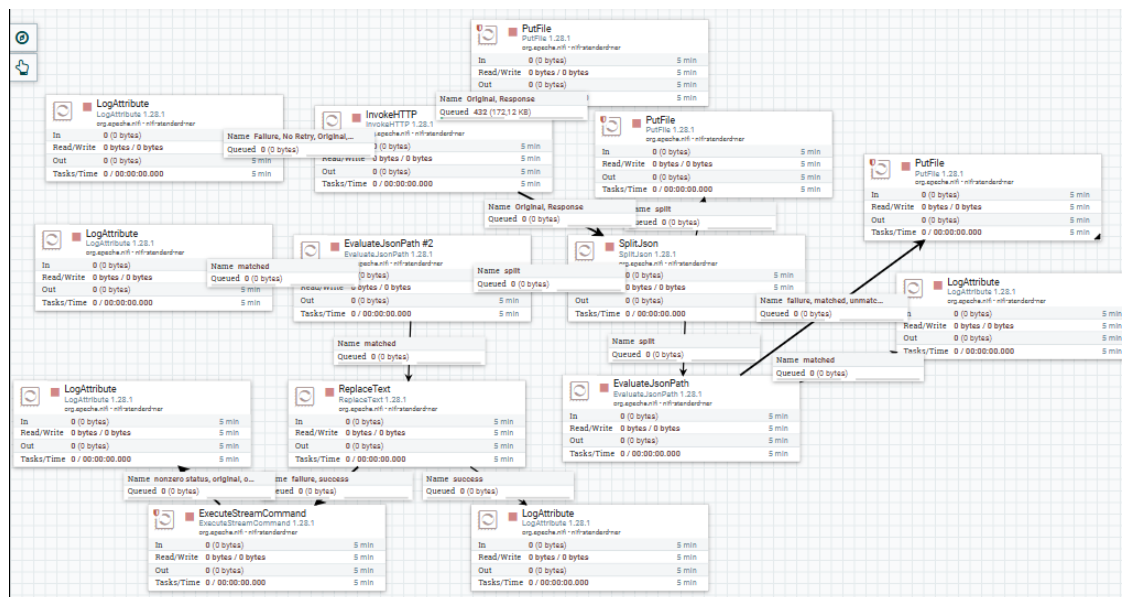


FIGURE 4.2 – Flow d'ingestion NiFi pour la base AS/400

— **Fichiers plats (CSV/TXT par email)** : L'ingestion des fichiers plats reçus par email est automatisée via le flow suivant :

1. ConsumeIMAP se connecte à une boîte mail dédiée et télécharge automatiquement les pièces jointes contenant les données.
2. PutFile enregistre les fichiers dans un répertoire de staging local.
3. Un déclencheur surveille ce répertoire pour activer ExecuteStreamCommand, qui lance un script Spark chargé du parsing, de la normalisation et du chargement dans PostgreSQL.
4. Des LogAttribute sont intégrés pour garantir la supervision, le contrôle qualité et la maintenance.

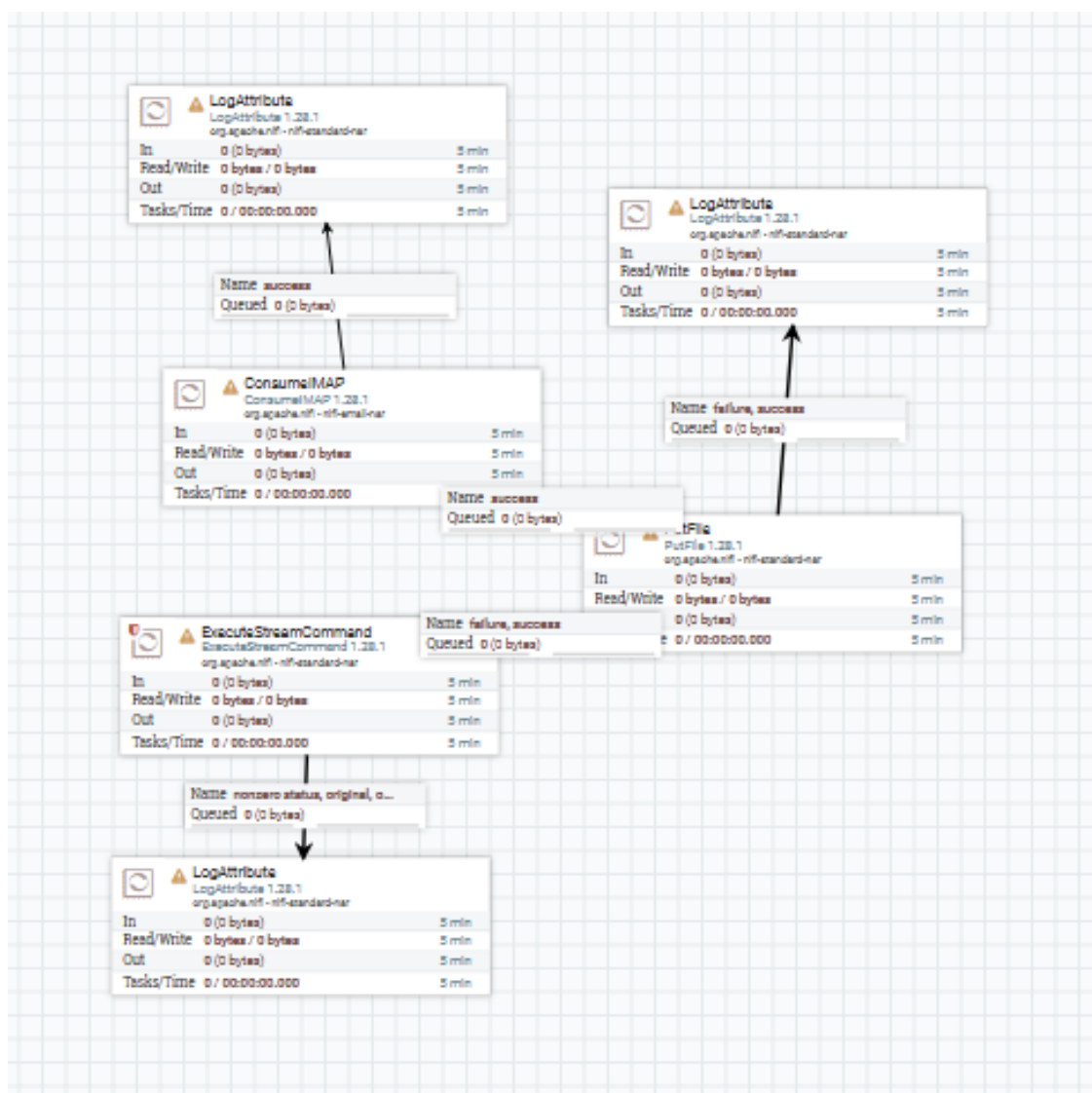


FIGURE 4.3 – Flow d'ingestion NiFi pour les fichiers plats reçus par email

Cette approche, avec un workflow NiFi détaillé et adapté à chaque source, garantit une ingestion fiable, traçable et flexible, tout en facilitant l'ajout de nouvelles sources ou l'évolution des traitements selon les besoins métier.

## 2.2 Transformation et chargement avec Apache Spark

La phase de transformation et de chargement des données s'appuie sur un enchaînement automatisé, orchestré entre Apache NiFi et Apache Spark. Après l'ingestion des fichiers et extractions, des jobs PySpark sont déclenchés pour appliquer les traitements nécessaires avant le chargement dans le data warehouse.

Les scripts PySpark ont été structurés en plusieurs étapes successives, chacune répondant à un objectif précis :

- **Pré-nettoyage automatique** : chaque jeu de données est soumis à un nettoyage systématique, incluant la suppression des valeurs nulles ou incohérentes, l'homogénéisation des formats de date, et l'encodage des champs textuels pour garantir la qualité des données

en entrée.

- **Jointures métier** : les différentes sources (ouvertures, règlements, clôtures, réouvertures) sont croisées à travers des clés communes (identifiants sinistres, contrats) afin de produire une vue consolidée et enrichie, conforme aux besoins analytiques des utilisateurs.
- **Chargement** : les résultats finaux, une fois validés, sont insérés dans les tables PostgreSQL correspondantes, en respectant un schéma en étoile prédéfini pour optimiser les requêtes analytiques.

```

169 fait_df = fact_base \
170     .join(dim_assures_db.select("id", "codeentre"), ["codeentre"], "left") \
171     .withColumnRenamed("id", "id_intermediaire") \
172     .join(dim_assures_db.select("id", "immatriculation"), \
173           fact_base["immatriculation_assure"] == dim_assures_db["immatriculation"], "left") \
174     .withColumnRenamed("id", "id_assure") \
175     .join(dim_vehicules_db.select("id", "immatriculation"), \
176           fact_base["immatriculation_assure"] == dim_vehicules_db["immatriculation"], "left") \
177     .withColumnRenamed("id", "id_vehicule") \
178     .join(dim_garanties_db.select("id", "codegara"), ["codegara"], "left") \
179     .withColumnRenamed("id", "id_garantie") \
180     .join(dim_sinistres_nature_db.select("id", "codnatsi"), ["codnatsi"], "left") \
181     .withColumnRenamed("id", "id_sinistre_nature") \
182     .join(dim_compagnies_adverses_db.select("id", "nom"), \
183           fact_base["compagnie_adverse"] == dim_compagnies_adverses_db["nom"], "left") \
184     .withColumnRenamed("id", "id_compagnie_adverse") \
185     .join(dim_canaux_db.select("id", "canal"), ["canal"], "left") \
186     .withColumnRenamed("id", "id_canal") \
187     .join(dim_localisations_db.select("id", "ville"), \
188           fact_base["ville_sinistre"] == dim_localisations_db["ville"], "left") \
189     .withColumnRenamed("id", "id_localisation")
190
191 # Join with dates for different date milestones
192 fait_df = fait_df \
193     .join(dim_dates_db.alias("d1").select("id", "date_full"), \
194           to_date(fait_df["date_surveillance"]) == col("d1.date_full"), "left") \
195     .withColumnRenamed("id", "date_surveillance_id") \
196     .join(dim_dates_db.alias("d2").select("id", "date_full"), \
197           to_date(fait_df["datedecl"]) == col("d2.date_full"), "left") \
198     .withColumnRenamed("id", "date_declaration_id") \
199     .join(dim_dates_db.alias("d3").select("id", "date_full"), \
200           to_date(fait_df["date_ouverture"]) == col("d3.date_full"), "left") \
201     .withColumnRenamed("id", "date_ouverture_id")

```

FIGURE 4.4 – Extrait du script PySpark de transformation et chargement

```

2 .filter(col("assure").isNotNull()) \
3 .distinct() \
4 .withColumnRenamed("assure", "nom") \
5 .withColumnRenamed("ville_assure", "ville") \
6 .withColumnRenamed("immatriculation_assure", "immatriculation")
7
8 dim_assures.write.jdbc(url, "sh_sinistres.dim_assures", "append", props)
9 print(f"dim_assures created with {dim_assures.count()} records")
10
11 # 3. DIM_VEHICULES (immatriculation, marque, type_vehicule)
12 print("Building dim_vehicules...")
13 dim_vehicules = all_data.select("immatriculation_assure", "marque_vehicule", "type_vehicule") \
14     .filter(col("immatriculation_assure").isNotNull()) \
15     .distinct() \
16     .withColumnRenamed("immatriculation_assure", "immatriculation") \
17     .withColumnRenamed("marque_vehicule", "marque")
18
19 dim_vehicules.write.jdbc(url, "sh_sinistres.dim_vehicules", "append", props)
20 print(f"dim_vehicules created with {dim_vehicules.count()} records")
21
22 # 4. DIM_GARANTIES (codegara, libelle, rubrique, garanties)
23 print("Building dim_garanties...")
24 dim_garanties = all_data.select("codegara", "garantie", "rubrique", "garanties") \
25     .filter(col("codegara").isNotNull()) \
26     .distinct() \
27     .withColumnRenamed("garantie", "libelle")
28
29 dim_garanties.write.jdbc(url, "sh_sinistres.dim_garanties", "append", props)
30 print(f"dim_garanties created with {dim_garanties.count()} records")
31

```

FIGURE 4.5 – Extrait du script PySpark de transformation et chargement

Cette organisation modulaire des traitements permet d'assurer la maintenabilité du pipeline,

tout en facilitant l'ajout de nouvelles règles métier ou l'adaptation des transformations en fonction de l'évolution des besoins.

### 2.3 Stockage des données dans PostgreSQL

Le stockage des données transformées s'effectue dans un data warehouse PostgreSQL, conçu selon un modèle en étoile, particulièrement adapté à l'analyse et aux requêtes OLAP. Cette architecture a été définie en étroite collaboration avec l'équipe data et les utilisateurs métier, afin de garantir la cohérence des données, la performance des requêtes et la facilité d'évolution du schéma.

#### Architecture du schéma de données

Le schéma du data warehouse comprend plusieurs éléments structurants :

- **Une table de faits principale** (`fact_sinistres`) : cette table centralise l'ensemble des montants de règlements, les dates clés et les métadonnées associées à chaque sinistre, tout en assurant la liaison avec les différentes dimensions via des clés étrangères.
- **Des tables de dimensions** :
  - `dim_contrats` : regroupe les informations contractuelles, les produits et les garanties associées.
  - `dim_types_sinistres` : décrit la typologie des sinistres, les catégories et sous-catégories.
  - `dim_dates` : propose un calendrier détaillé avec des hiérarchies temporelles (jour, mois, trimestre, année).
  - `dim_clients` : contient des données clients anonymisées et segmentées pour respecter la confidentialité.
  - `dim_statuts` : retrace le workflow des statuts de sinistres.
  - `dim_geographie` : précise la localisation des sinistres (région, ville, etc.).
- **Des clés étrangères** : garantissent l'intégrité référentielle et permettent des jointures efficaces lors des analyses.

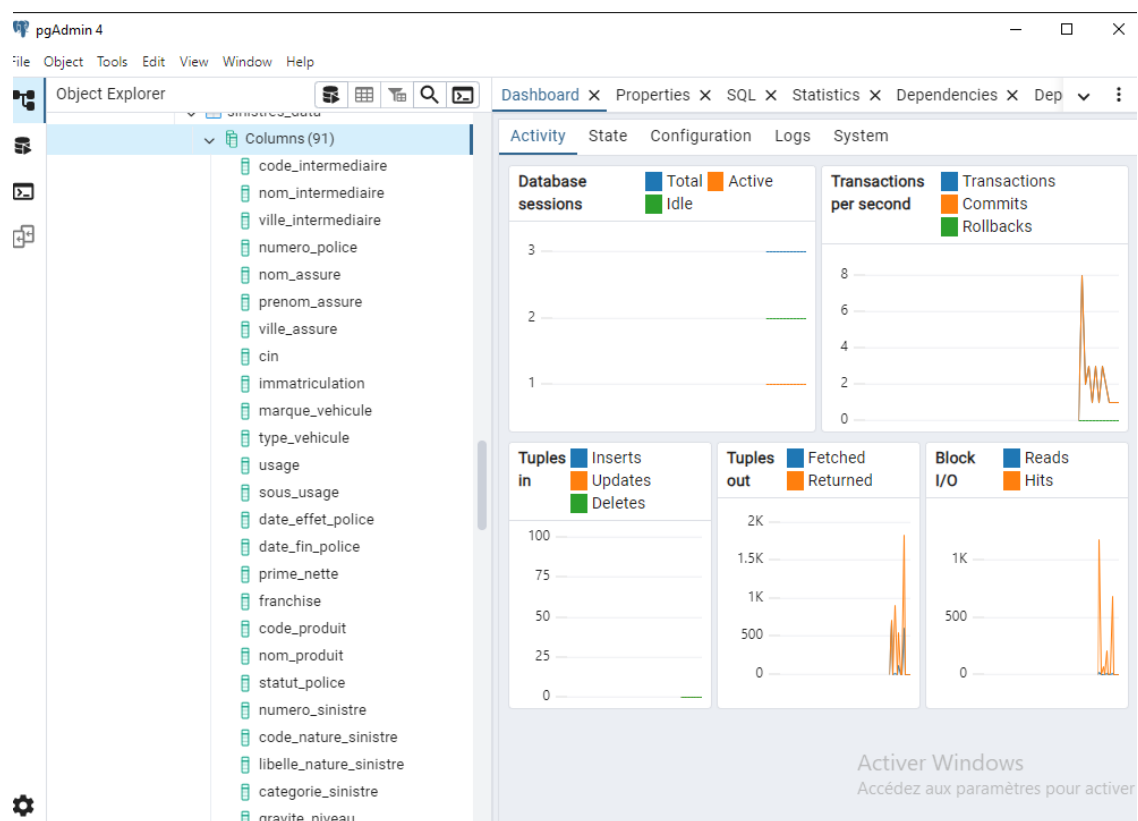


FIGURE 4.6 – Vue d'ensemble dans pgAdmin

## Optimisations de performance

Pour garantir des performances optimales lors des requêtes analytiques, plusieurs optimisations ont été mises en œuvre :

- **Indexation stratégique** : création d'index sur les clés étrangères, les dates de sinistres et les montants, afin d'accélérer les jointures et les filtres les plus courants.
- **Vues matérialisées** : mise en place de vues matérialisées pour pré-calculer les agrégations les plus fréquemment utilisées (par exemple, les totaux par mois ou par type de sinistre), ce qui réduit considérablement les temps de réponse.
- **Statistiques optimisées** : configuration de la mise à jour automatique des statistiques PostgreSQL, afin d'améliorer la qualité des plans d'exécution générés par le moteur de base de données.



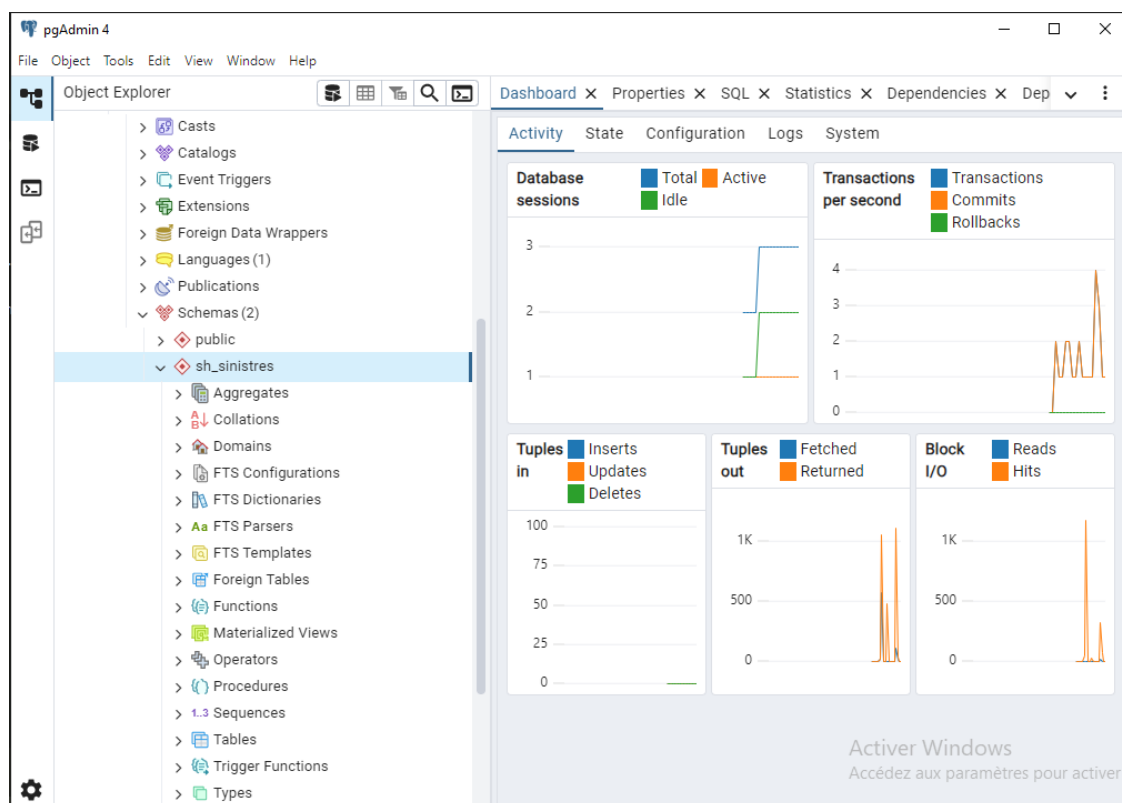


FIGURE 4.7 – monitoring performance

Cette architecture de stockage constitue le socle technique indispensable pour alimenter efficacement Metabase, tout en assurant la cohérence, la performance et la traçabilité des données à destination des utilisateurs métier d'Allianz Maroc.

## 2.4 Visualisation des données avec Metabase

Une fois les données chargées et structurées dans PostgreSQL, la base a été connectée à Metabase, une plateforme de Business Intelligence open source, afin de permettre une visualisation claire, interactive et personnalisée des indicateurs clés à destination des utilisateurs métier.

La conception des tableaux de bord s'est appuyée sur une analyse fine des besoins, réalisée en collaboration avec les responsables métier. Les requêtes SQL utilisées dans Metabase ont été construites à partir des extractions les plus pertinentes, en veillant à optimiser à la fois la performance et la lisibilité des résultats.

L'interface intuitive de Metabase a été configurée pour offrir une grande autonomie aux utilisateurs, qui peuvent :

- Créer leurs propres “questions” (requêtes visuelles) sans avoir besoin de compétences techniques avancées.
- Construire des dashboards personnalisés à partir des jeux de données partagés, en fonction de leurs besoins spécifiques.
- Filtrer, trier et explorer les données en temps réel grâce à une interface graphique conviviale.

La gestion des rôles et des permissions a été soigneusement paramétrée pour garantir un accès sécurisé et conforme aux exigences de confidentialité :

- Attribution de rôles personnalisés selon les départements (finance, gestion, actuariat, etc.).
- Masquage des colonnes sensibles (montants, informations personnelles) pour limiter l'exposition des données critiques.
- Segmentation des données selon les droits d'accès, afin de respecter les principes de moindre privilège.

**Note :** Les captures d'écran présentées dans cette section sont volontairement floutées afin de préserver la confidentialité des données réelles utilisées en production.

L'intégration de la base de données PostgreSQL dans Metabase représente une étape cruciale, car elle permet d'établir la connexion entre le data warehouse et l'interface de visualisation. Cette configuration définit les paramètres d'accès, les permissions et les modalités d'interrogation des données par Metabase.

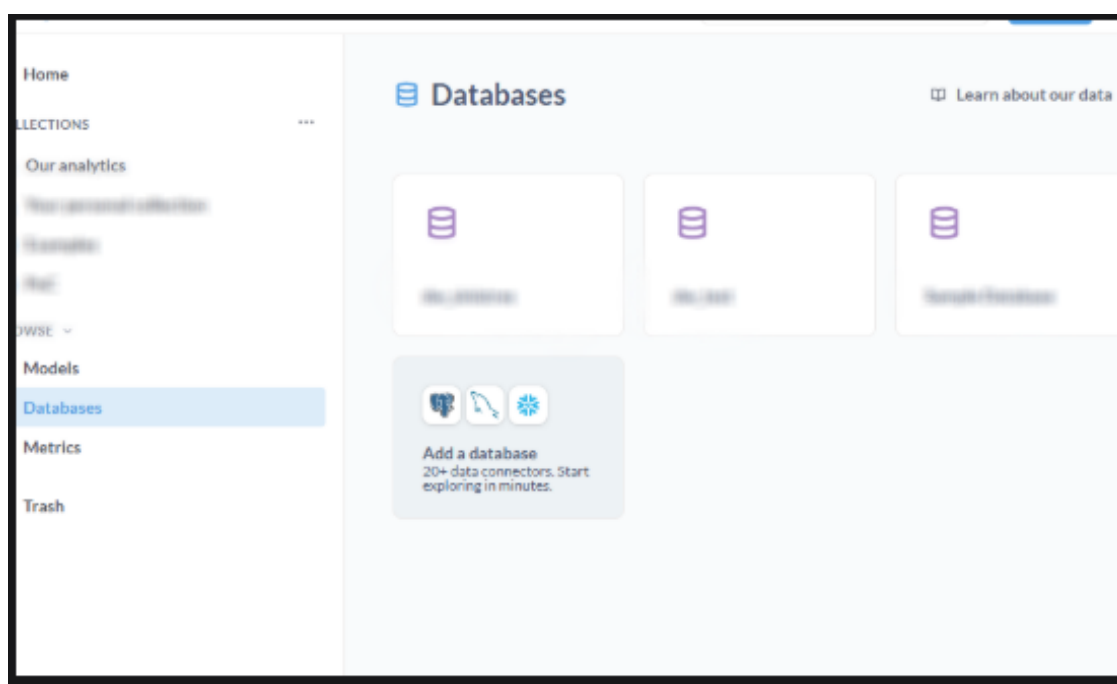


FIGURE 4.8 – Ajout de la connexion PostgreSQL dans Metabase

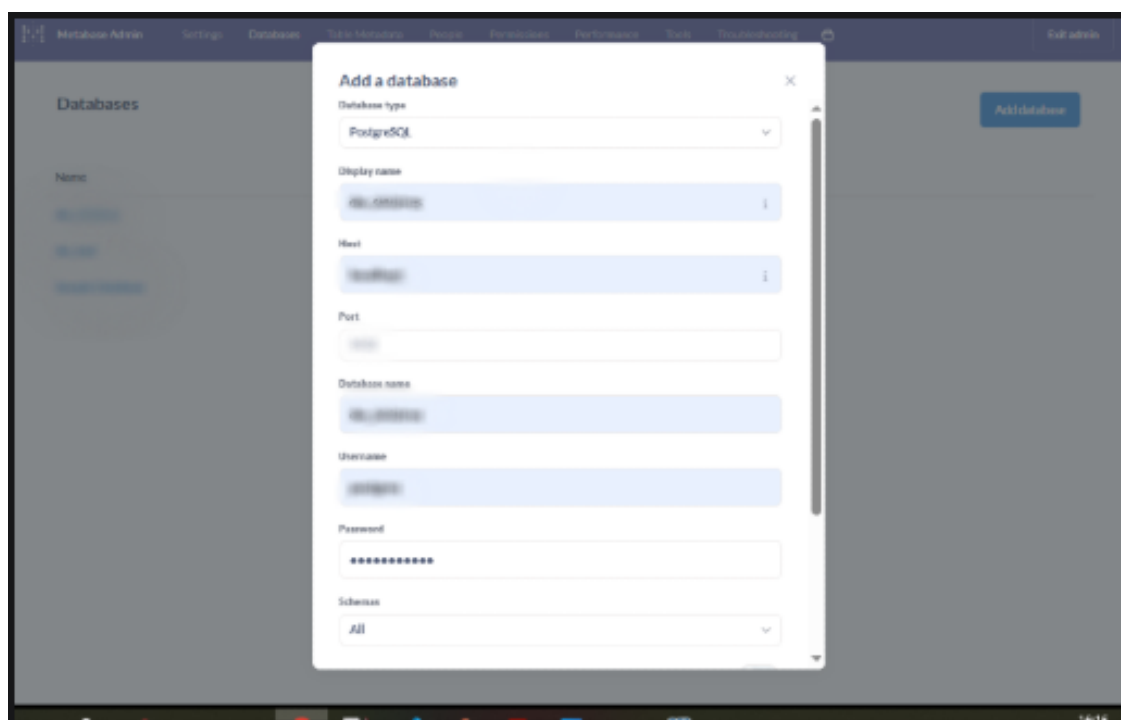


FIGURE 4.9 – Configuration de la connexion PostgreSQL dans Metabase

Une fois la connexion établie, la création des tableaux de bord s'effectue en définissant des “questions”, c'est-à-dire des requêtes SQL ou des requêtes visuelles, qui extraient et présentent les données selon les besoins métier. Ces questions servent de briques de base pour la construction de dashboards personnalisés, adaptés à chaque profil utilisateur.

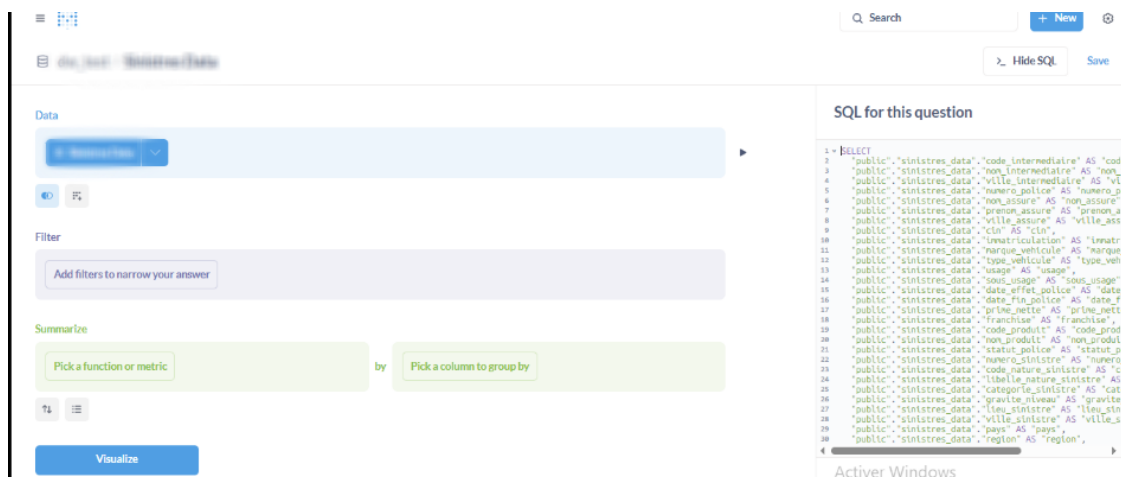


FIGURE 4.10 – Interface de création d'une question/requête dans Metabase pour l'extraction des données

Les dashboards sont ensuite assemblés en combinant plusieurs questions, ce qui permet d'offrir une vue d'ensemble cohérente, interactive et dynamique. L'interface de création de Metabase permet de personnaliser la disposition des graphiques, d'ajouter des filtres et de configurer les interactions entre les différents éléments, afin de répondre précisément aux besoins de chaque département.

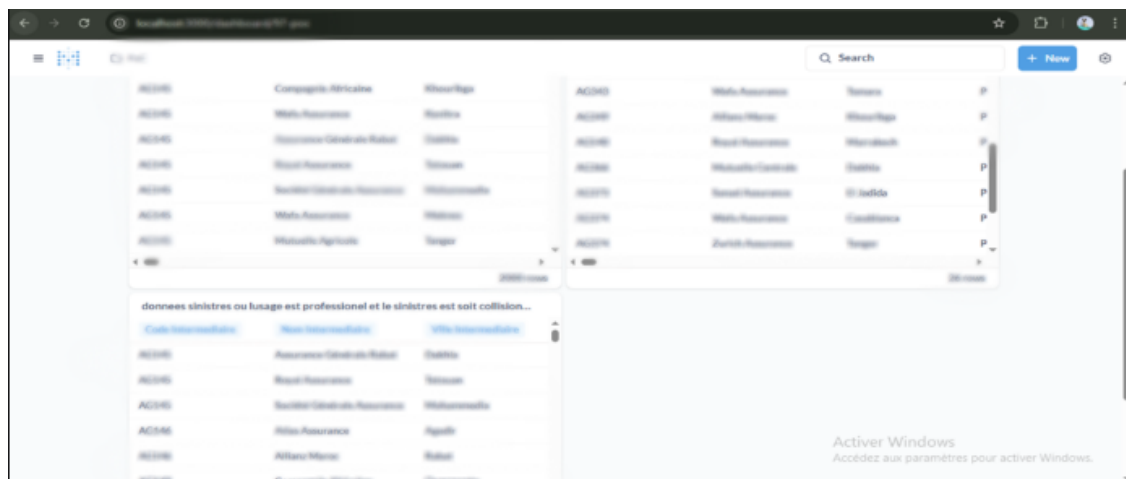


FIGURE 4.11 – Interface de création d'un dashboard

La gestion des utilisateurs dans Metabase constitue un pilier fondamental de la sécurité et de la gouvernance des données. L'interface d'administration permet de créer des comptes utilisateurs avec des permissions spécifiques, garantissant ainsi un accès contrôlé et personnalisé aux tableaux de bord, en fonction des profils métier de chaque utilisateur.

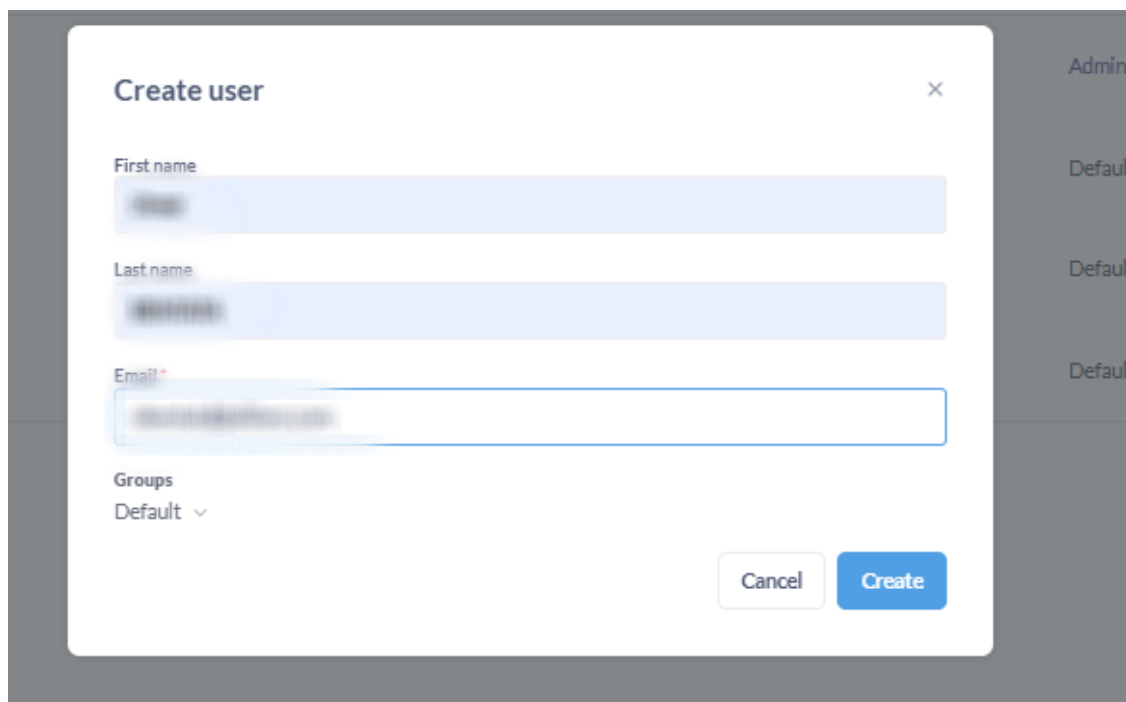


FIGURE 4.12 – Interface de création d'un utilisateur

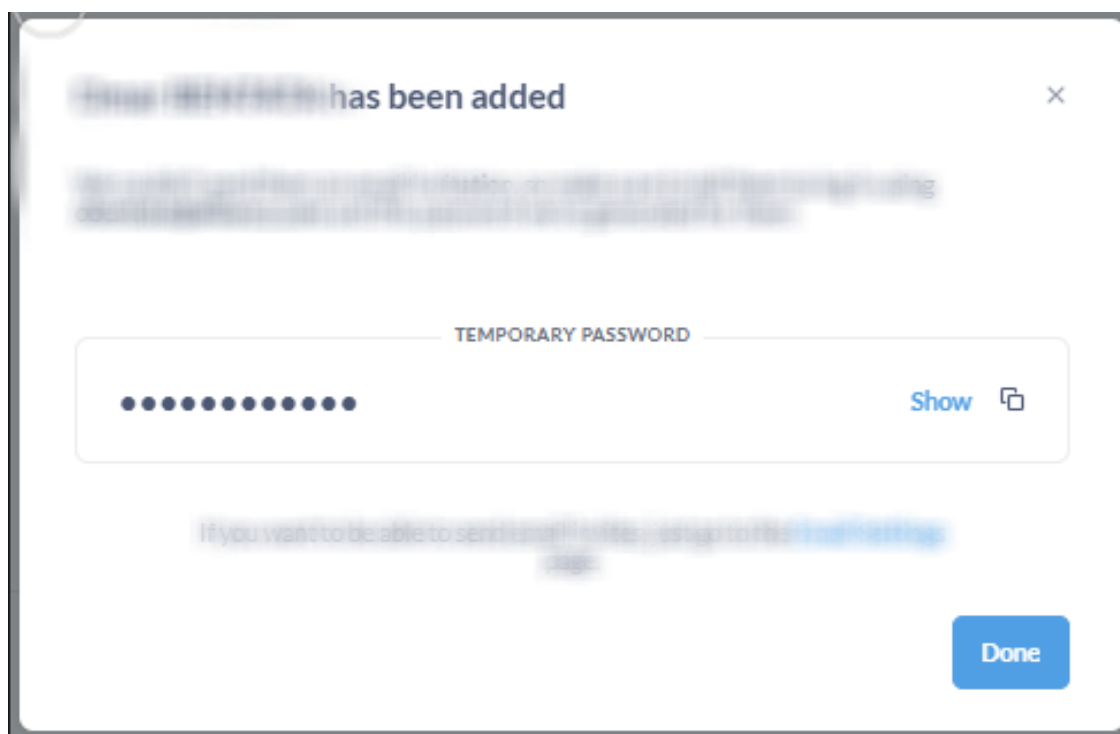


FIGURE 4.13 – Envoi du mot de passe à l'utilisateur

La gestion fine des droits d'accès dans Metabase repose sur la notion de **groupes d'utilisateurs**. Chaque utilisateur est rattaché à un ou plusieurs groupes, lesquels se voient attribuer des permissions spécifiques sur les bases de données, les collections de questions et les tableaux de bord. Cette organisation permet de moduler précisément les accès selon les rôles métier (ex. : analyste, manager, administrateur) et de garantir le respect des politiques de sécurité et de confidentialité.

L'ajout d'un utilisateur à un groupe s'effectue via l'interface d'administration, où il est possible de :

- Créer de nouveaux groupes correspondant aux différents profils ou départements de l'entreprise (ex. : Direction, Sinistres, Finance, IT).
- Associer chaque utilisateur à un ou plusieurs groupes selon ses besoins d'accès.
- Définir les permissions de chaque groupe sur les différentes ressources (lecture, écriture, administration, etc.).

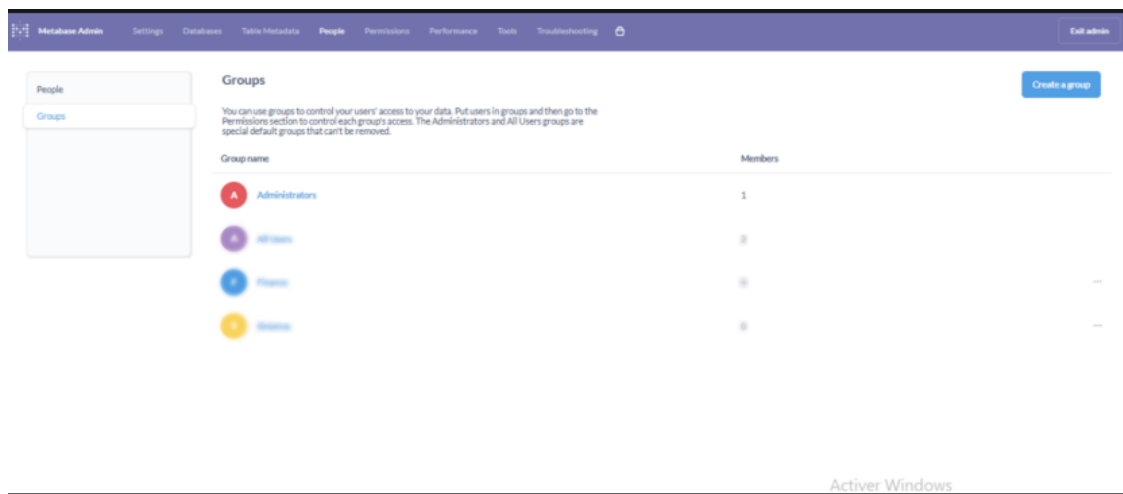


FIGURE 4.14 – Gestion des groupes et des permissions dans Metabase

Cette approche par groupes facilite la maintenance des droits d'accès, l'intégration de nouveaux collaborateurs et l'évolution des responsabilités au sein de l'organisation. Elle constitue un levier essentiel pour assurer la conformité, la traçabilité et la sécurité des usages de la plateforme de datavisualisation.

L'accès à Metabase est protégé par un système d'authentification robuste, qui contrôle l'accès aux tableaux de bord selon les profils utilisateurs. Cette interface de connexion constitue le point d'entrée principal pour tous les utilisateurs métier souhaitant consulter et analyser les données.

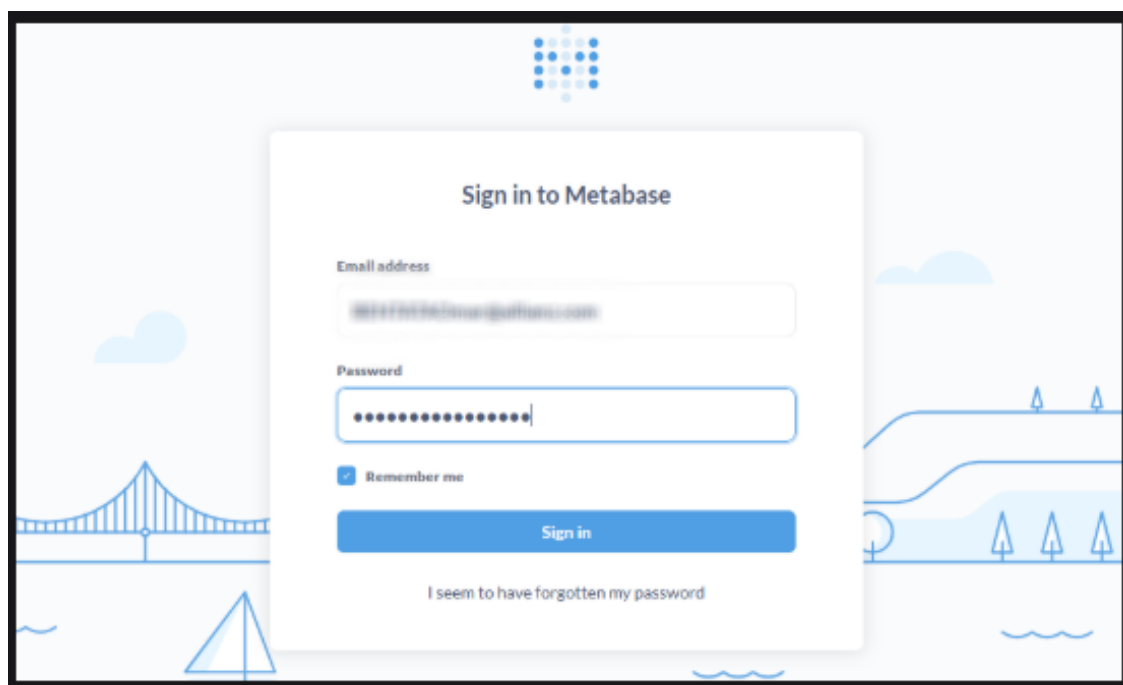


FIGURE 4.15 – Interface de connexion Metabase avec authentification sécurisée

Une fois connecté, l'utilisateur final accède à son tableau de bord personnalisé, où il peut consulter, analyser et explorer les données selon ses besoins métier spécifiques. L'ergonomie de

l'interface permet une navigation fluide entre les différents graphiques et indicateurs, avec de nombreuses possibilités de filtrage, de tri et d'exploration interactive.

The image shows a dashboard with four tables of accident data. Each table has columns for 'Code Intermédiaire', 'Nom Intermédiaire', and 'Ville Intermédiaire'. The tables are titled 'Sinistres Data', 'Sinistres Data, Filtered by Ville Assure is Casablanca and Marque Veh...', 'données sinistres ou usage est professionnel et le sinistres est soit collision...', and 'Sinistres Data, Filterer ou usage professionnel et accident collision ou tout...'. The tables contain rows of accident data with various codes and names. The interface includes filters and sorting options for each table.

FIGURE 4.16 – Dashboard de visualisation des règlements par sinistre (données floutées)

### 3 DevOps et déploiement

Dans une optique de portabilité, de maintenabilité et de standardisation, l'ensemble des composants de la solution a été conteneurisé et orchestré à l'aide de Docker Compose. Ce choix technologique a permis de garantir la reproductibilité des environnements, de simplifier le déploiement sur les infrastructures internes, et de faciliter la gestion des dépendances entre les différents services.

#### 3.1 Dockerisation des composants

Chaque composant du projet – à savoir Apache NiFi, Apache Spark, PostgreSQL et Metabase – a été encapsulé dans un conteneur Docker distinct, avec une configuration personnalisée adaptée aux besoins du projet.

- **Apache NiFi** : utilisation de l'image officielle, avec exposition des ports nécessaires et montage d'un volume pour assurer la persistance des flows et des configurations.

```
nifi > Dockerfile > FROM
1 FROM apache/nifi:1.23.2
2
3 # Copy JDBC drivers
4 COPY drivers/*.jar /opt/nifi/nifi-current/lib/
5
6 # Copy NiFi flows and templates
7 COPY flows/ /opt/nifi/nifi-current/flows/
8 COPY templates/ /opt/nifi/nifi-current/templates/
9
10 # Create necessary directories
11 RUN mkdir -p /opt/nifi/data \
12     && mkdir -p /opt/nifi/nifi-current/conf/templates \
13     && chown -R nifi:nifi /opt/nifi/data \
14     && chown -R nifi:nifi /opt/nifi/nifi-current/flows \
15     && chown -R nifi:nifi /opt/nifi/nifi-current/templates
16
17 # Set environment variables
18 ENV NIFI_WEB_HTTP_PORT=8080
19 ENV NIFI_CLUSTER_IS_NODE=false
20 ENV NIFI_CLUSTER_NODE_PROTOCOL_PORT=8082
21 ENV NIFI_ELECTION_MAX_WAIT=30sec
22
23 EXPOSE 8080 8443 8082
24
```

FIGURE 4.17 – Conteneurisation de NiFi

- **Spark (job standalone)** : déploiement d'une image basée sur PySpark, avec montage du dossier des scripts pour permettre l'exécution des jobs de transformation.



```

spark > Dockerfile > FROM
1  FROM bitnami/spark:3.4
2
3  # Copy Spark jobs and libraries
4  COPY jobs/ /opt/bitnami/spark/jobs/
5  COPY libs/ /opt/bitnami/spark/libs/
6  COPY config/ /opt/bitnami/spark/conf/
7
8  # Install additional Python packages for Spark jobs
9  USER root
10 RUN pip install --no-cache-dir \
11     pandas \
12     numpy \
13     psycopg2-binary \
14     sqlalchemy \
15     pyspark
16
17 # Create necessary directories
18 RUN mkdir -p /opt/bitnami/spark/data \
19     && chown -R 1001:1001 /opt/bitnami/spark/jobs \
20     && chown -R 1001:1001 /opt/bitnami/spark/libs \
21     && chown -R 1001:1001 /opt/bitnami/spark/data
22
23 USER 1001
24
25 EXPOSE 8080 7077 4040
26

```

FIGURE 4.18 – Conteneurisation de Spark

- **PostgreSQL** : configuration du schéma du data warehouse via un script SQL exécuté au démarrage du conteneur, assurant ainsi la cohérence et la reproductibilité de la structure de la base.

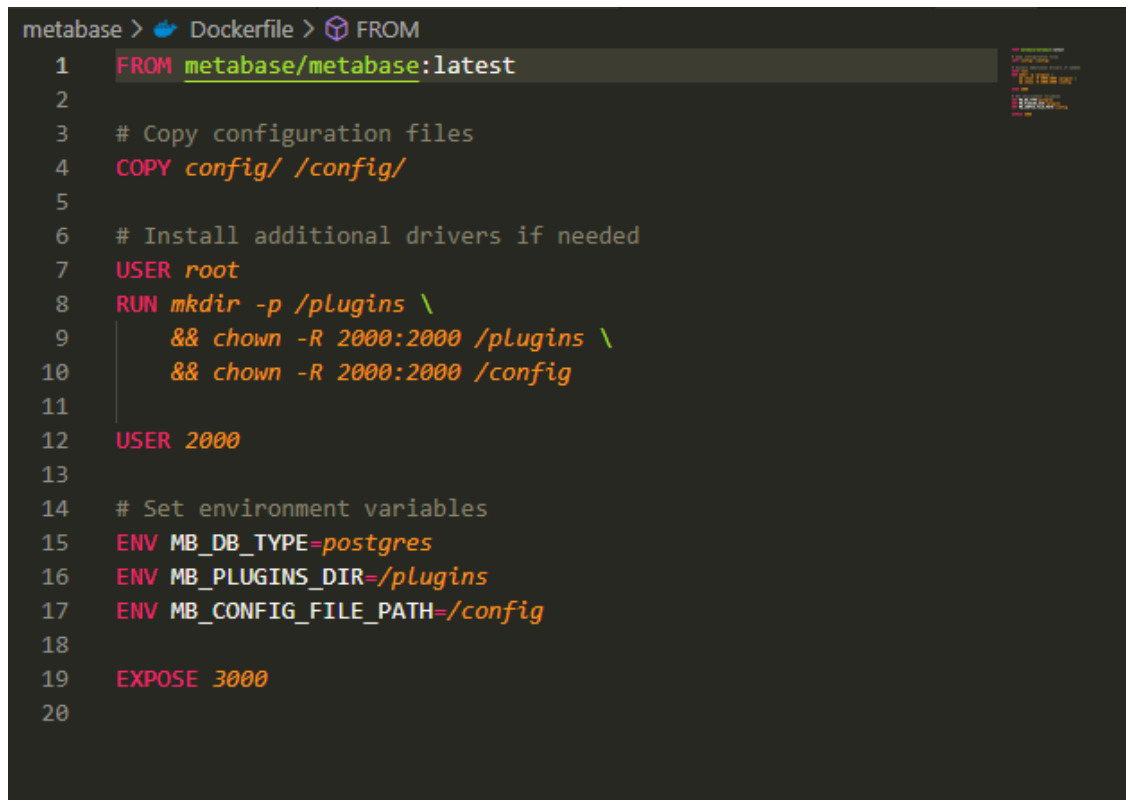
```

postgres > Dockerfile > ...
1  FROM postgres:15-alpine
2
3  # Install additional extensions
4  RUN apk add --no-cache postgresql-contrib
5
6  # Copy initialization scripts
7  COPY init-scripts/ /docker-entrypoint-initdb.d/
8
9  # Set proper permissions
10 RUN chmod +x /docker-entrypoint-initdb.d/*.sql
11
12 # Create backup directory
13 RUN mkdir -p /var/lib/postgresql/backup
14
15 EXPOSE 5432
16

```

FIGURE 4.19 – Conteneurisation de PostgreSQL

- **Metabase** : utilisation de l'image officielle, avec configuration de la connexion à PostgreSQL pour permettre l'accès aux données et la création des dashboards.

A screenshot of a Dockerfile for Metabase. The file is edited in a dark-themed editor. The content of the Dockerfile is as follows:

```
metabase > Dockerfile > FROM
1 FROM metabase/metabase:latest
2
3 # Copy configuration files
4 COPY config/ /config/
5
6 # Install additional drivers if needed
7 USER root
8 RUN mkdir -p /plugins \
9     && chown -R 2000:2000 /plugins \
10    && chown -R 2000:2000 /config
11
12 USER 2000
13
14 # Set environment variables
15 ENV MB_DB_TYPE=postgres
16 ENV MB_PLUGINS_DIR=/plugins
17 ENV MB_CONFIG_FILE_PATH=/config
18
19 EXPOSE 3000
20
```

FIGURE 4.20 – Conteneurisation de Metabase

Cette approche modulaire permet de déployer, mettre à jour ou remplacer chaque composant indépendamment, tout en assurant une cohérence globale de l'environnement.

## 3.2 Orchestration avec Docker Compose

L'orchestration de l'ensemble des services a été centralisée dans un fichier `docker-compose.yml`, qui définit la configuration, les dépendances, les volumes et les réseaux pour chaque conteneur. Cette organisation permet de lancer l'intégralité de la solution en une seule commande, tout en assurant l'isolation et l'interconnexion des services au sein d'un réseau bridge dédié.

- Lancement simplifié via la commande `docker compose up -d`, permettant de démarrer tous les services en mode détaché.
- Utilisation de fichiers `.env` pour la gestion des variables d'environnement et la personnalisation des configurations, offrant ainsi une grande flexibilité.
- Centralisation des logs pour faciliter le debugging et le suivi des opérations.

```

22  ▷ Run Service
23  nifi:
24    build:
25      context: ./nifi
26      dockerfile: Dockerfile
27    container_name: nifi
28    environment:
29      NIFI_WEB_HTTP_PORT: 8080
30      NIFI_CLUSTER_IS_NODE: false
31      NIFI_CLUSTER_NODE_PROTOCOL_PORT: 8082
32      NIFI_ZK_CONNECT_STRING: zookeeper:2181
33      NIFI_ELECTION_MAX_WAIT: 30 sec
34      NIFI_SENSITIVE_PROPS_KEY: ${NIFI_SENSITIVE_PROPS_KEY}
35    ports:
36      - "8080:8080"
37      - "8443:8443"
38    volumes:
39      - nifi_data:/opt/nifi/nifi-current/conf
40      - nifi_flowfile_repo:/opt/nifi/nifi-current/flowfile_repository
41      - nifi_content_repo:/opt/nifi/nifi-current/content_repository
42      - nifi_provenance_repo:/opt/nifi/nifi-current/provenance_repository
43      - ./nifi/flows:/opt/nifi/nifi-current/flows
44      - ./nifi/drivers:/opt/nifi/nifi-current/drivers
45      - ./data:/opt/nifi/data
46    networks:
47      - data-pipeline
48    depends_on:
49      - postgres

```

FIGURE 4.21 – Configuration Docker Compose pour NiFi

```

50  ▷ Run Service
51  spark-master:
52    build:
53      context: ./spark
54      dockerfile: Dockerfile
55    container_name: spark-master
56    environment:
57      - SPARK_MODE=master
58      - SPARK_RPC_AUTHENTICATION_ENABLED=no
59      - SPARK_RPC_ENCRYPTION_ENABLED=no
60      - SPARK_LOCAL_STORAGE_ENCRYPTION_ENABLED=no
61      - SPARK_SSL_ENABLED=no
62    ports:
63      - "8081:8080"
64      - "7077:7077"
65    volumes:
66      - ./spark/jobs:/opt/bitnami/spark/jobs
67      - ./spark/libs:/opt/bitnami/spark/libs
68      - ./spark/config:/opt/bitnami/spark/conf
69      - ./data:/opt/bitnami/spark/data
70    networks:
71      - data-pipeline

```

FIGURE 4.22 – Configuration Docker Compose pour Spark

```

1  docker-compose.yml
2  ↓ Pull Images | ▷ Compose Up
3  version: '3.8'
4  ▷ Run All Services
5  services:
6    ▷ Run Service
7    postgres:
8      build:
9        context: ./postgres
10       dockerfile: Dockerfile
11       container_name: postgres-db
12       environment:
13         POSTGRES_DB: ${POSTGRES_DB}
14         POSTGRES_USER: ${POSTGRES_USER}
15         POSTGRES_PASSWORD: ${POSTGRES_PASSWORD}
16       ports:
17         - "5432:5432"
18       volumes:
19         - postgres_data:/var/lib/postgresql/data
20         - ./postgres/init-scripts:/docker-entrypoint-initdb.d
21         - ./postgres/backup:/backup
22       networks:
23         - data-pipeline

```

FIGURE 4.23 – Configuration Docker Compose pour PostgreSQL

```

D:\Run Service
metabase:
  build:
    context: ./metabase
    dockerfile: Dockerfile
  container_name: metabase
  environment:
    MB_DB_TYPE: postgres
    MB_DB_DBNAME: ${METABASE_DB_NAME}
    MB_DB_PORT: 5432
    MB_DB_USER: ${POSTGRES_USER}
    MB_DB_PASS: ${POSTGRES_PASSWORD}
    MB_DB_HOST: postgres
  ports:
    - "3000:3000"
  volumes:
    - metabase_data:/metabase-data
    - ./metabase/config:/config
  networks:
    - data-pipeline
  depends_on:
    - postgres

```

FIGURE 4.24 – Configuration Docker Compose pour Metabase

```

117 volumes:
118   postgres_data:
119   nifi_data:
120   nifi_flowfile_repo:
121   nifi_content_repo:
122   nifi_provenance_repo:
123   metabase_data:
124
125 networks:
126   data-pipeline:
127     driver: bridge
128

```

FIGURE 4.25 – Configuration des volumes et réseaux Docker Compose

La configuration des volumes assure la persistance des données entre les redémarrages des conteneurs, ce qui est essentiel pour PostgreSQL (stockage des données du data warehouse), NiFi (flows et repositories), et Metabase (configuration des dashboards et des utilisateurs). Le réseau bridge data-pipeline garantit la communication fluide entre les différents conteneurs, tout en maintenant une isolation stricte vis-à-vis du système hôte.

### 3.3 Déploiement sur serveur local

Le déploiement de la solution a été réalisé sur une machine locale (serveur Linux) mise à disposition par l'équipe IT d'Allianz Maroc. L'environnement a été préparé avec soin, en suivant plusieurs étapes :

- Installation de Docker et Docker Compose, pour permettre la gestion des conteneurs et l'orchestration des services.
- Configuration des ports réseau, notamment pour exposer Metabase sur l'intranet de l'entreprise et permettre l'accès aux utilisateurs métier.
- Mise en place d'un lancement automatique des services au démarrage du serveur, via des cron jobs, afin de garantir la disponibilité continue de la solution.

Cette démarche a permis d'assurer un déploiement fiable, reproductible et facilement maintenable, tout en facilitant les opérations de mise à jour ou de restauration en cas d'incident.

### 3.4 Tests et validation de la solution

La phase de tests a constitué une étape cruciale pour garantir la qualité, la fiabilité et la conformité de la solution par rapport aux attentes métier. Plusieurs scénarios de validation fonctionnelle ont été définis et exécutés sur l'ensemble de la chaîne de traitement :

- Vérification du mapping des champs entre les sources de données et le data warehouse, afin de s'assurer de la cohérence et de l'exhaustivité des informations transférées.
- Contrôles de cohérence sur les jointures et les agrégations réalisées par Spark, pour valider l'exactitude des calculs et des consolidations.
- Tests de performance sur Metabase, incluant le temps de chargement des dashboards, la réactivité des filtres dynamiques et la robustesse de l'interface en conditions réelles d'utilisation.

Ces tests ont permis d'identifier et de corriger rapidement d'éventuelles anomalies, tout en apportant des garanties supplémentaires quant à la robustesse et à la pérennité de la solution déployée.

## 4 Conclusion

L'implémentation de la solution technique décrite dans ce chapitre a permis de construire une pipeline de données complète, automatisée et sécurisée, parfaitement alignée avec les besoins d'Allianz Maroc. Grâce à l'intégration harmonieuse d'Apache NiFi pour l'ingestion, d'Apache Spark pour la transformation, de PostgreSQL pour le stockage et de Metabase pour la visualisation, chaque étape du processus a été optimisée pour garantir la performance, la scalabilité et la fiabilité du système.

L'approche modulaire et conteneurisée adoptée facilite grandement la maintenance, le déploiement et l'évolution de la solution, tout en assurant une intégration fluide avec les systèmes existants de l'entreprise. Ce projet pose ainsi les bases d'une infrastructure de données moderne, évolutive et résiliente, capable de s'adapter aux futurs besoins d'Allianz Maroc en matière de Business Intelligence, d'analyse et de visualisation des données. Il constitue également un socle solide pour l'extension de la gouvernance des données et l'industrialisation de nouveaux cas d'usage à venir.

# Conclusion Générale

## Synthèse du projet

Ce projet de fin d'études, réalisé au sein du service IT d'Allianz Maroc, a consisté en la conception et le déploiement d'une architecture complète de traitement de données, dédiée à la valorisation du patrimoine informationnel de l'entreprise. L'objectif : permettre une exploitation fluide, fiable et sécurisée des données issues de multiples systèmes hétérogènes (AS/400, Oracle, fichiers plats), tout en fournissant aux utilisateurs métier un accès libre-service aux indicateurs critiques via une plateforme de BI moderne.

La solution repose sur une chaîne de traitement full-stack :

Ingestion automatisée des données avec Apache NiFi

Préparation, jointure et enrichissement via Apache Spark

Modélisation et stockage dans un Data Warehouse PostgreSQL

Visualisation interactive via Metabase

Cette stack, contenue dans des conteneurs Docker, est orchestrée via Docker Compose et déployée sur un serveur interne, avec un pipeline CI/CD assurant un cycle de vie fluide du code.

## Objectifs atteints

Les objectifs initiaux ont été atteints, voire dépassés, sur les volets techniques, organisationnels et fonctionnels :

- **Fiabilisation de la donnée** : qualité contrôlée à chaque étape du pipeline (nulls, formats, typage).
- **Automatisation** : traitements quotidiens déclenchés sans intervention humaine.
- **Modélisation pertinente** : schéma en étoile adapté aux besoins d'analyse sinistre.
- **Accessibilité** : interface Metabase intuitive pour utilisateurs non techniques.
- **Sécurité et gouvernance** : rôles, segmentation et masquage des données critiques.

## Apports techniques et professionnels

Ce projet m'a permis de développer une expertise concrète dans l'écosystème moderne de la donnée, notamment sur :

- L'ingestion et l'orchestration de flux complexes avec Apache NiFi.
- Le traitement distribué avec Apache Spark (PySpark).
- La modélisation d'un entrepôt de données PostgreSQL conforme aux principes BI.
- Le paramétrage avancé d'un outil de visualisation orienté utilisateur (Metabase).

— La mise en place d'une chaîne DevOps avec GitHub Actions et Docker.

Au-delà de l'aspect technique, j'ai pu acquérir des compétences clés en gestion de projet data, communication avec les métiers, cadrage des besoins, arbitrages technologiques et adaptation à un environnement complexe (secteur assurance, contexte réglementaire, exigences sécurité).

## Difficultés rencontrées et résolutions

### Techniques

La diversité des formats sources (API, SQL, fichiers plats) et leur structuration hétérogène ont nécessité une architecture flexible.

Le traitement initial de la volumétrie a exigé une optimisation des scripts Spark et des jointures.

Le passage de données de production, sensible, a demandé une coordination fine avec les équipes métier et sécurité.

### Organisationnelles

La planification du projet a dû être ajustée en fonction des disponibilités des équipes métier.

Le temps d'apprentissage de certains outils (notamment NiFi) a été absorbé en s'appuyant sur une documentation technique rigoureuse et des tests itératifs.

## Perspectives d'évolution

### Court terme

Extension à d'autres modules métier (santé, habitation, fraude).

Intégration d'alerting et de monitoring (Prometheus / Grafana).

Ajout de tests automatisés plus fins sur la qualité des données.

### Moyen terme

Gouvernance avancée avec data catalog (ex : Amundsen, DataHub).

Mise à disposition d'un environnement sandbox pour requêtes ad hoc.

Développement de connecteurs supplémentaires (API externes, fichiers SFTP...).

### Long terme

Généralisation du modèle à d'autres filiales d'Allianz.

Passage à une architecture orientée "data-as-a-service".

Intégration de modules prédictifs ou d'analyse avancée (ML, scoring sinistres).

## Retour d'expérience

Ce projet de fin d'études a représenté bien plus qu'une simple application technique ou un exercice académique. Il a été une véritable aventure humaine et professionnelle, marquée par l'exigence, la découverte et la remise en question permanente. Travailler au sein d'Allianz Maroc, dans un environnement où la donnée est au cœur des enjeux stratégiques, m'a permis de mesurer l'importance de la rigueur, de l'écoute et de l'adaptabilité.

J'ai eu la chance de participer à toutes les étapes du cycle de vie d'un projet data, de la réflexion initiale à la mise en production, en passant par la documentation, la qualité, le déploiement et l'accompagnement des utilisateurs. Cette immersion m'a confronté à la complexité du réel : la diversité des besoins métiers, la gestion des imprévus, la nécessité de communiquer efficacement avec des interlocuteurs variés, et l'importance de la pédagogie pour embarquer l'ensemble des parties prenantes.

Au fil des mois, j'ai compris que la réussite d'un projet data ne repose pas uniquement sur la maîtrise des outils ou des technologies, mais avant tout sur la capacité à créer du lien entre les équipes, à comprendre les attentes profondes des utilisateurs, et à traduire des problématiques métier en solutions concrètes et pérennes. J'ai appris à valoriser l'écoute, la curiosité et l'humilité, tout en développant une exigence de qualité et de fiabilité à chaque étape.

Ce projet a renforcé ma conviction que la donnée, bien au-delà de sa dimension technique, est un formidable levier de transformation pour l'entreprise. Il m'a donné envie de poursuivre dans cette voie, d'approfondir mes compétences et de continuer à apprendre, au service de projets porteurs de sens et d'impact. Je ressors grandi de cette expérience, avec la certitude que l'excellence technique doit toujours s'accompagner d'une compréhension fine des enjeux humains et organisationnels.

Je tiens à remercier chaleureusement l'ensemble des équipes d'Allianz Maroc pour leur accueil, leur confiance et leur accompagnement tout au long de cette aventure. Ce stage restera pour moi une étape fondatrice, tant sur le plan professionnel que personnel.



# Webographie

## Sites web consultés

1. **PostgreSQL Documentation**

<https://www.postgresql.org/docs/>

Consulté le : **8 mars 2025**

Description : Documentation officielle et complète de PostgreSQL.

2. **Docker Documentation**

<https://docs.docker.com/>

Consulté le : **19 mars 2025**

Description : Guide officiel pour la containerisation, les images et les déploiements.

3. **Kubernetes Documentation**

<https://kubernetes.io/docs/>

Consulté le : **5 mars 2025**

Description : Documentation et meilleures pratiques pour l'orchestration de conteneurs.

4. **Apache NiFi Documentation**

<https://nifi.apache.org/documentation/>

Consulté le : **13 mars 2025**

Description : Guides et référence de NiFi, outil d'orchestration de flux de données.

5. **GitHub Actions Documentation**

<https://docs.github.com/en/actions>

Consulté le : **22 mars 2025**

Description : Documentation officielle sur l'automatisation CI/CD via GitHub Actions.

6. **DevOps Best Practices (Azure DevOps)**

<https://docs.microsoft.com/azure/devops/>

Consulté le : **10 mars 2025**

Description : Guide Microsoft pour les pratiques DevOps incluant CI/CD.

7. **ChatGPT (OpenAI)**

<https://chatgpt.com/>

Consulté le : **3 mars 2025**

Description : Assistant IA conversationnel basé sur GPT, pour la rédaction, la recherche et le codage.

8. **Claude (Anthropic)**

<https://claude.ai/>

Consulté le : **3 mars 2025**

Description : Assistant IA alternatif axé sur la sécurité, la précision et la productivité.

9. **GitHub Copilot**

<https://github.com/features/copilot>

Consulté le : **3 mars 2025**

Description : Assistant IA de programmation pour l'autocomplétion et la génération de code.

10. **Metabase Documentation**

<https://www.metabase.com/docs/>

Consulté le : **12 mars 2025**

Description : Documentation officielle de Metabase pour la visualisation et l'analyse de données.

11. **Apache Spark Documentation**

<https://spark.apache.org/docs/latest/>

Consulté le : **15 mars 2025**

Description : Référence technique pour le traitement distribué et l'analytique big data.

12. **OWASP Top Ten**

<https://owasp.org/www-project-top-ten/>

Consulté le : **21 mars 2025**

Description : Guide des risques majeurs de sécurité applicative, utile pour la sécurisation des APIs et des plateformes BI.

13. **Data Engineering Podcast**

<https://www.dataengineeringpodcast.com/>

Consulté le : **18 mars 2025**

Description : Podcast sur les architectures, outils et retours d'expérience en data engineering.

14. **Allianz Group**

<https://www.allianz.com/en.html>

Consulté le : **2 mars 2025**

Description : Site institutionnel du groupe Allianz, pour la compréhension du contexte métier et des enjeux sectoriels.

## Tutoriels et articles techniques

1. **Building Scalable Data Pipelines**

<https://medium.com/data-engineering>

Consulté le : **14 mars 2025**

Description : Articles techniques sur la conception de pipelines de données évolutifs.

2. **Start Data Engineering (blog)**

<https://www.startdataengineering.com/>

Consulté le : **25 mars 2025**

Description : Blog spécialisé sur les pratiques, outils et architectures en data engineering.

3. **Kubernetes Best Practices**

<https://cloud.google.com/kubernetes-engine/docs/best-practices>

Consulté le : **7 mars 2025**

Description : Bonnes pratiques pour la mise en œuvre de Kubernetes sur Google Cloud.

4. **Database Design Patterns (PostgreSQL)**

<https://www.postgresql.org/docs/current/ddl.html>

Consulté le : **17 mars 2025**

Description : Patterns et bonnes pratiques pour la conception de schémas relationnels.

5. **API Design Guidelines (REST)**

<https://restfulapi.net/>

Consulté le : **11 mars 2025**

Description : Directives pour la conception, la normalisation et la sécurisation des APIs REST.

6. **Modern Data Stack (Fivetran Blog)**

<https://fivetran.com/blog/modern-data-stack>

Consulté le : **16 mars 2025**

Description : Analyse des tendances et outils de la data stack moderne.

7. **DataOps Manifesto**

<https://www.dataopsmanifesto.org/>

Consulté le : **20 mars 2025**

Description : Principes et bonnes pratiques pour l'industrialisation des pipelines de données.

8. **YouTube – Apache NiFi Tutorials (DataCouch)**

<https://www.youtube.com/playlist?list=PLf0swTFhTI8q18p6U6QvFqQnQKp1K0QdC>

Consulté le : **9 mars 2025**

Description : Tutoriels vidéo pour la prise en main et l'automatisation de flux avec NiFi.

9. **Towards Data Science (Medium)**

<https://towardsdatascience.com/>

Consulté le : **23 mars 2025**

Description : Articles de vulgarisation et d'approfondissement sur la data science, l'IA et l'ingénierie des données.

10. **Awesome Data Engineering (GitHub)**

<https://github.com/igorbarinov/awesome-data-engineering>

Consulté le : **27 mars 2025**

Description : Liste collaborative de ressources, outils et lectures recommandées pour l'ingénierie des données.

11. **Le Monde Informatique – Dossiers Data**

<https://www.lemondeinformatique.fr/dossiers/data/>

Consulté le : **28 février 2025**

Description : Dossiers et actualités sur la transformation digitale, la gouvernance et la valorisation des données en entreprise.

12. **Qwen (Alibaba Cloud)**

<https://qwen.alibaba.com/>

Consulté le : **4 mars 2025**

Description : Modèle d'IA générative open source développé par Alibaba Cloud, utilisé pour la génération de texte et l'assistance à la rédaction technique.

13. **Gemini (Google AI)**

<https://deepmind.google/technologies/gemini/>

Consulté le : **8 mars 2025**

Description : Modèle d'IA multimodal de Google, utilisé pour la recherche, la génération de code et l'analyse de données complexes.

14. **Google Cloud Copilot**

<https://cloud.google.com/ai/copilot>

Consulté le : **12 mars 2025**

Description : Assistant IA de Google Cloud pour l'aide à la programmation, la génération de code et l'automatisation des workflows cloud.