

# Projet Pole Santé

## Remerciements



Je remercie profondément tous ceux qui m'ont aidé de près ou de loin à la réalisation de ce travail et sans lesquels ce travail n'aurait pu arriver à terme.

Je remercie en particulier :

Mon encadrante à la Cooperl Mme Élodie CORMAND Chef d'équipe BI (Business Intelligence) pour sa confiance aussi la qualité de l'encadrement qu'elle m'a assuré, et surtout pour ses encouragements continuels qui m'ont permis d'avancer.

Mes remerciements s'adressent également à tous les membres de l'équipe BI de la Cooperl pour leurs encouragements et pour toute l'aide qu'ils m'ont fourni durant la phase du développement.

Finalement, je tiens à remercier les membres du jury qui m'ont fait l'honneur de juger ce travail.

## Résumé

Ce travail a été effectué dans le cadre du projet de fin d'études pour l'obtention du diplôme en MBA Data Engineering au sein de l'institut F2i

L'objectif de ce projet consiste à développer une application CGE Pôle Santé - Suivi mensuel de l'activité commerciale au sein de la coopérative Cooperl .

Cette application permettra aux responsables de gérer au sein du groupement, le suivi mensuel de l'activité commerciale du pôle santé, en fournissant des outils de collecte, d'analyse et de visualisation des données pertinentes pour prendre des décisions stratégiques informées.

## Table des matières

### Chapitre 1 : Étude Préliminaire

#### 1.1 Introduction

#### 1.2 Présentation de l'organisme d'accueil

#### 1.3 Présentation du projet

##### 1.3.1 Problématique

#### 1.4 État de l'art et étude préalable

##### 1.4.1 Etude de l'existant

##### 1.4.2 Critique de l'existant

### **1.4.3 Solution proposé** [↗](#)

## **1.5 Méthodologies** [↗](#)

### **1.5.1 Méthodologie Kanban**

### **1.5.2 Les avantages de la méthodologie Kanban**

### **1.5.3 Etude comparative des méthodologies** [↗](#)

## **1.6 Conclusion** [↗](#)

# **Chapitre 2 : L'informatique décisionnelle** [↗](#)

## **2.1 Introduction**

## **2.2 Concepts généraux du BI** [↗](#)

### **2.2.1 La Business Intelligence**

### **2.2.2 Avantages du BI**

### **2.2.3 Les limites du BI** [↗](#)

## **2.3 Les principes des systèmes décisionnels** [↗](#)

### **2.3.1 Sources de données**

### **2.3.2 Entrepôt de données**

### **2.3.3 Extract-Transform-Load**

### **2.3.4 La différence entre OLTP et OLAP**

### **2.3.5 Conclusion** [↗](#)

# **Chapitre 3 : Les Différentes Plateformes Cloud** [↗](#)

## **3.1 Introduction au Cloud Computing**

## **3.2 Modèles de Service Cloud**

## **3.3 Avantages du Cloud**

## **3.4 Inconvénients du Cloud**

## **3.5 Modèles de Déploiement Cloud**

## **3.5 Comparaison entre AWS, Azure et Oracle Cloud**

## **3.6 Conclusion**



## Chapitre 4 : Mise en Œuvre de la Solution sur AWS Athena

### 4.1 Introduction

### 4.2 Architecture de la solution

#### 4.2.1 Schéma de l'architecture

#### 4.2.2 Composants utilisés (Athena, S3, Glue)

### 4.3 Conception de la base de données

#### 4.3.1 Modélisation des données

#### 4.3.2 Création des tables dans Athena

### 4.4 Processus ETL (Extract, Transform, Load)

### 4.5 Optimisation des requêtes sur Athena

#### 4.5.1 Indexation et partitionnement

#### 4.5.2 Bonnes pratiques pour les requêtes

### 4.6 Conclusion



## Chapitre 5 : Conception et Déploiement du Tableau de Bord sur Amazon QuickSight

### 5.1 Introduction

### 5.2 Présentation d'Amazon QuickSight

#### 5.2.1 Fonctionnalités principales

#### 5.2.2 Avantages de QuickSight

### 5.3 Connexion de QuickSight à Athena

#### 5.3.1 Configuration des sources de données

#### 5.3.2 Sécurisation des accès

### 5.4 Conception du tableau de bord

#### 5.4.1 Identification des indicateurs clés de performance (KPI)

#### 5.4.2 Création des visualisations

#### 5.4.3 Intégration des filtres et des contrôles interactifs

### 5.5 Déploiement et partage du tableau de bord

#### 5.5.1 Publication et mise à jour du tableau de bord

## 5.5.2 Gestion des permissions et partage avec les utilisateurs

## 5.6 Conclusion

### Introduction

Le présent rapport constitue le fruit d'un projet de fin d'études réalisé en collaboration avec l'entreprise Cooperl. Ce projet s'inscrit dans le cadre d'un MBA Data engineering et vise à aborder le suivi mensuel de l'activité commerciale.

Cooperl, fondée en 1966 sous le nom de Coopérative des éleveurs de la région de Lamballe se positionne comme un acteur incontournable dans le secteur de l'agroalimentaire, spécifiquement dans l'élevage porcin et la transformation de viande. La coopérative opère sur l'ensemble de la chaîne de valeur, depuis la production des matières premières jusqu'à la distribution des produits finis, en mettant l'accent sur des pratiques durables et une approche éthique de l'élevage.

Fortement ancrée dans le tissu économique français, Cooperl a su développer une expertise reconnue tant au niveau national qu'international. Avec plus de 3000 de membres coopérateurs et une présence sur 17 sites de production, l'entreprise joue un rôle central dans la dynamique régionale, contribuant à l'emploi et au développement économique des territoires où elle est implantée.

La collaboration avec Cooperl offre une opportunité unique d'allier théorie et pratique dans un contexte industriel concret. En s'engageant aux côtés de cette entreprise, ce projet de fin d'études vise à explorer le suivi mensuel de l'activité commerciale du pôle santé de la Cooperl tout en répondant aux besoins et défis actuels du secteur de l'agroalimentaire.

Lors de la mise en œuvre de notre solution, nous avons suivi les méthodes approuvées de développement agile. La solution recommandée est décrite dans ce rapport qui est organisé comme suit :

- Le premier chapitre commence d'abord par une présentation générale de la société ,puis une mise en place de la problématique, proposition des solutions ainsi que la méthodologie de travail utilisée.
- Le deuxième chapitre : c'est la phase d'analyse des besoins : fonctionnels et non fonctionnels, ainsi, on va faire une étude de l'art avec une critique, pour aboutir enfin à l'élaboration de la solution proposée, et comme résultat de cette phase, on va détailler les diagrammes de cas d'utilisation ainsi que leurs scénarios.
- Le troisième chapitre sera concerné par la conception générale et la conception détaillée de la solution proposée. La première partie présentera l'architecture générale du projet et la deuxième partie illustrera le Schéma en étoiles .
- Le dernier chapitre sera consacré à la mise en œuvre de la partie conversion de la modélisation effectuée précédemment en une partie exécutable. En fait, ce chapitre est composé de deux parties : la première est la justification de choix technologique, et la deuxième est la phase d'implémentation : elle regroupe les différentes fonctionnalités du système avec des captures d'écran.

Nous terminerons ce projet par une conclusion générale et une ouverture à d'autres perspectives en vue d'améliorer et d'élargir ce travail.

## Chapitre 1 : Étude Préliminaire

### 1.1. Introduction

Notre projet, intitulé 'Cooperl Groupement d'Éleveurs Pôle Santé', vise à moderniser et optimiser notre infrastructure de Business Intelligence au sein de Cooperl.

Actuellement, nous utilisons Amazon Redshift pour stocker et interroger nos données, mais nous cherchons à évoluer vers une solution plus agile et économique.

Notre objectif principal est de décommissionner l'entrepôt de données RedShift au profit d'Athena, en reconstruisant les tableaux de bord existants sous QuickSight, tout en les complétant si nécessaire. Nous souhaitons ainsi exploiter les capacités de requêtage sans serveur d'AWS Athena pour accéder à nos données stockées sur Amazon S3 et créer des tableaux de bord dynamiques et interactifs sur QuickSight. Cette transition nous permettra non seulement de réduire nos coûts d'infrastructure, mais aussi d'améliorer la flexibilité et la rapidité de notre processus d'analyse de données.

Nous sommes convaincus que cette modernisation renforcera notre capacité à prendre des décisions éclairées et à répondre aux besoins croissants de nos utilisateurs finaux.

## 1.2. Présentation de l'organisme d'accueil

Cooperl est une coopérative agricole française majeure opérant dans le secteur de l'élevage porcin et de la transformation de viande. Fondée en 1966 en Bretagne, la coopérative est devenue au fil des décennies un acteur incontournable de l'agroalimentaire français et européen.

Cooperl trouve ses racines dans le terroir breton, où un groupe d'éleveurs de porcs a décidé de s'unir pour mieux valoriser leur production. Cette initiative coopérative a donné naissance à Cooperl, qui a débuté ses activités avec quelques membres et un site de production modeste.

Au fil des années, Cooperl a su élargir son réseau de producteurs et moderniser ses installations pour répondre aux demandes croissantes du marché. Aujourd'hui, la coopérative compte des milliers de membres coopérateurs et exploite plusieurs sites de production à travers la France.

Cooperl opère sur l'ensemble de la chaîne de valeur de l'élevage porcin et de la transformation de viande. Ses principales activités comprennent :

- **L'élevage porcin** : Grâce à son groupement d'éleveurs, Cooperl collabore étroitement avec ses éleveurs partenaires pour assurer la production de porcs de qualité, tout en respectant des normes rigoureuses en matière de bien-être animal, de traçabilité et de durabilité environnementale. Le groupement d'éleveurs porcs (CGE-P) est au cœur de cette activité, favorisant des pratiques agricoles responsables et innovantes.
- **L'élevage bovin** : En plus de l'élevage porcin, Cooperl a élargi son activité avec le groupement d'éleveurs bovins (CGE-B), garantissant une approche similaire de qualité, de bien-être animal et de traçabilité dans le secteur de l'élevage bovin.
- **La nutrition animale** : Cooperl Nutrition (CNU) est spécialisée dans la formulation et la production d'aliments destinés aux animaux, offrant une expertise en nutrition pour optimiser la santé et la croissance des élevages, tout en respectant les standards de durabilité.
- **La transformation de viande** : Avec des unités de transformation à la pointe de la technologie, Cooperl transforme la viande porcine en une vaste gamme de produits frais et transformés, allant des viandes découpées aux produits de salaison. Cooperl Viandes (CVI) et Cooperl Salaisons (CSA) sont les divisions responsables de ces activités, mettant l'accent sur la qualité, l'innovation et la sécurité alimentaire.
- **La distribution** : La coopérative commercialise ses produits sous des marques reconnues, répondant aux besoins des consommateurs via un réseau étendu de partenaires, aussi bien au niveau national qu'international. La division Cooperl Distribution (CDI) assure la diffusion de ses produits dans différents segments de marché, que ce soit en grande distribution, en restauration ou à l'exportation.
- **Les équipements et services** : En plus de ses activités de production, Cooperl propose à ses éleveurs des équipements et solutions techniques adaptés aux besoins de l'élevage et de la transformation. Cooperl Équipements (CEQ) fournit ainsi des infrastructures et technologies de pointe pour optimiser la production. Par ailleurs, Cooperl Environnement (CEN) se concentre sur des solutions innovantes pour gérer les impacts environnementaux de l'élevage, avec un accent mis sur la gestion des déchets, le traitement de l'eau et l'énergie renouvelable.
- **Fonctions support** : Toutes ces activités sont soutenues par des fonctions support solides, allant des services financiers aux ressources humaines, en passant par l'innovation technologique et la qualité, permettant à Cooperl de maintenir sa position de leader dans l'industrie agroalimentaire. Cooperl s'engage à promouvoir une agriculture durable et responsable, en mettant l'accent sur le bien-être animal, la sécurité alimentaire et la préservation de l'environnement. La coopérative valorise également l'innovation et la recherche pour améliorer constamment ses pratiques et ses produits.

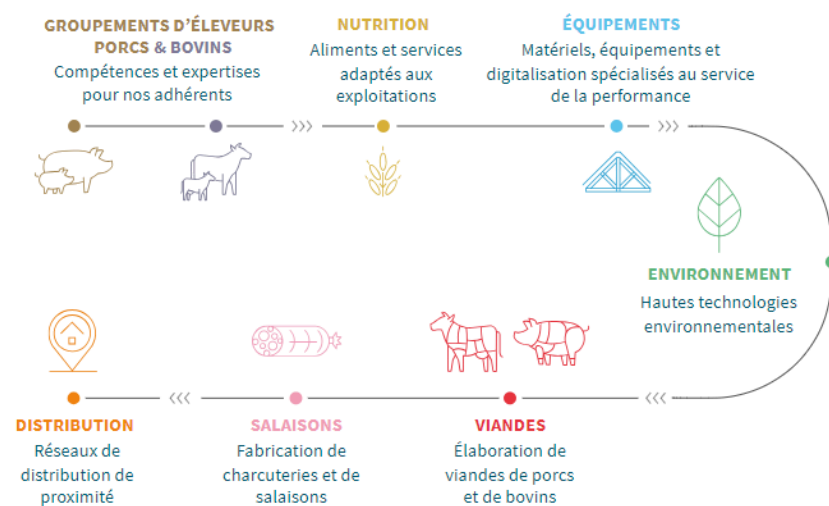
- **L'abattage** : En intégrant ses propres unités d'abattage, Cooperl maîtrise les normes sanitaires et de traçabilité à chaque étape de la production, garantissant ainsi transparence et qualité aux consommateurs.
- **La nutrition animale** : La coopérative développe des solutions de nutrition animale pour ses éleveurs, permettant une alimentation équilibrée des porcs qui favorise leur santé et la qualité de la viande, tout en respectant les normes environnementales.
- **L'environnement** : Dans un souci de durabilité, Cooperl investit dans des technologies pour la valorisation des sous-produits et le traitement des effluents, limitant ainsi son empreinte écologique. Elle s'engage également dans des projets d'économie circulaire pour recycler les déchets issus de la production.

En tant qu'entreprise ancrée dans le tissu économique et social français, Cooperl joue un rôle significatif dans la création d'emplois, le développement des territoires ruraux et la vitalité de l'industrie agroalimentaire. Sa présence sur le marché contribue également à renforcer la souveraineté alimentaire et la compétitivité de la filière porcine française.

En conclusion Cooperl incarne l'exemple d'une réussite coopérative dans le domaine de l'agroalimentaire. Grâce à son engagement envers la qualité, l'innovation et la durabilité, la coopérative continue de prospérer tout en répondant aux défis et aux exigences d'un marché en constante évolution.

# Cooperl

## COOPERL C'EST...



Supply Chain

Origine

Française

Fondée en	1966
Siège	<a href="#">7 Rue de la Jeannaie Maroué, 22400 Lamballe-Armor</a>
Secteur	COOPÉRATIVE AGRICOLE ET AGROALIMENTAIRE
Taille de l'entreprise	7 700 (2023) employés
E-mail	<a href="mailto:communication@cooperl.com">communication@cooperl.com</a>
Téléphone	+33 2 96 30 70 00
Site web	<a href="#">Cooperl, coopérative agricole et agroalimentaire du Grand Ouest organisée en filière</a>

### 1.3 Présentation du projet

#### 1.3.1 Problématique

Cooperl est une entreprise leader dans le secteur de l'agroalimentaire, spécialisée dans l'élevage et la transformation de porcs. Avec une présence internationale et un engagement envers l'innovation, Cooperl cherche constamment à améliorer ses processus et ses outils pour rester compétitif sur le marché.

Dans ce contexte, le département de Business Intelligence de Cooperl joue un rôle crucial en fournissant des insights stratégiques basés sur les données pour guider les décisions de l'entreprise. Cependant, l'infrastructure existante basée sur Amazon Redshift présente des limitations en termes de coûts, de performances et de flexibilité.

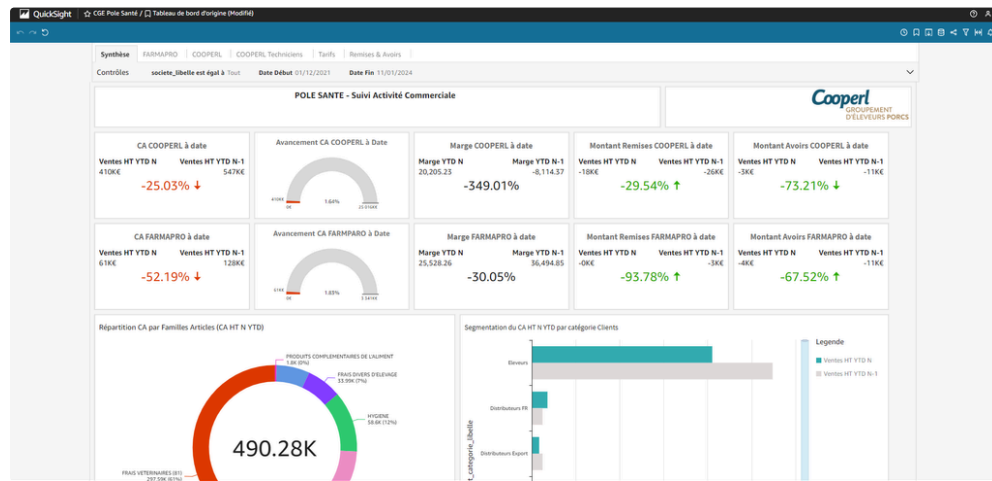
Le projet "Cooperl Groupement d'Éleveurs Pôle Santé" émerge de la nécessité de moderniser cette infrastructure et pour décommissionner Amazon Redshift qui coûte très cher. En passant à QuickSight pour la visualisation des données et en adoptant AWS Athena pour l'interrogation des données stockées sur Amazon S3, Cooperl vise à créer un écosystème de Business Intelligence plus agile, économique et évolutif. Cette transition permettra à Cooperl de mieux répondre aux besoins analytiques de ses équipes métier, d'accélérer le processus de prise de décision et de maintenir son avantage concurrentiel dans un marché en constante évolution.

### 1.4 État de l'art et étude préalable

#### 1.4.1 Etude de l'existant

- Dans le cadre de notre projet visant à créer un tableau de bord de suivi mensuel de l'activité commerciale du pôle santé de Cooperl, il est essentiel de réaliser une analyse approfondie de l'existant pour comprendre le paysage technologique actuel et les défis rencontrés.
- **Infrastructure Actuelle :** Actuellement, nous utilisons Amazon Redshift pour le stockage et l'analyse des données commerciales du pôle santé. Redshift offre des capacités puissantes en matière de traitement des données, mais les coûts associés à son utilisation sont élevés et le volume de données exploité pour le cas d'usage ne justifie pas l'utilisation du service d'entrepôt de données RedShift.
- **Analyse des Coûts :** Une analyse détaillée des coûts associés à Amazon Redshift révèle que les frais de stockage et de traitement des données représentent une part importante du budget. Cette situation nécessite une réévaluation de notre infrastructure technologique pour optimiser les coûts tout en maintenant des performances élevées.
- **Évaluation des Besoins :** Les besoins des utilisateurs finaux incluent la capacité de visualiser et d'analyser les données commerciales de manière intuitive et en temps réel. La convivialité et la rapidité d'accès aux informations sont des facteurs clés pour assurer l'efficacité opérationnelle.

- **Alternatives Technologiques** : Nous avons identifié Athena et QuickSight comme des alternatives potentielles à Amazon Redshift. Athena offre une solution de requête SQL sur les données stockées dans Amazon S3, tandis que QuickSight permet la création de tableaux de bord interactifs et visuels.
- **Benchmarking** : Une comparaison entre Amazon Redshift, Athena et QuickSight révèle que chaque solution présente des avantages et des inconvénients uniques. Athena offre une flexibilité et des économies potentielles en termes de coûts de stockage, tandis que QuickSight propose des fonctionnalités avancées de visualisation des données.



TDB Existant

### 1.4.2 Critique de l'existant [🔗](#)

- **Coûts Élevés** : Les coûts liés à l'utilisation d'Amazon Redshift sont significatifs et peuvent peser lourdement sur le budget du projet. La tarification basée sur la capacité et l'utilisation peut entraîner des dépenses imprévues, ce qui rend la gestion budgétaire plus complexe.
- **Complexité de la Gestion** : La gestion et l'administration d'une infrastructure Redshift peuvent être complexes, nécessitant des compétences techniques avancées et des ressources dédiées. Cela peut entraîner des retards dans le développement et la maintenance du tableau de bord.
- **Limitations de Scalabilité** : Bien que Redshift offre une scalabilité horizontale, atteindre des performances optimales avec de grandes quantités de données peut nécessiter des ajustements coûteux de la configuration et de l'infrastructure.
- **Expérience Utilisateur Limitée** : Malgré ses fonctionnalités avancées, le tableau de bord actuel peut présenter des limitations en termes d'expérience utilisateur. Les visualisations peuvent ne pas être assez intuitives ou interactives, ce qui peut limiter l'efficacité opérationnelle et la prise de décision.
- **Dépendance Technologique** : En se basant exclusivement sur Amazon Redshift, le projet est exposé au risque de dépendance technologique. Les changements dans les politiques de tarification ou les évolutions du marché pourraient avoir un impact négatif sur la viabilité à long terme du projet.
- **Besoins Évolutifs Non Satisfaits** : En raison des limitations de l'infrastructure actuelle, il est possible que certains besoins évolutifs des utilisateurs finaux ne soient pas pleinement satisfaits. Cela peut inclure des demandes de fonctionnalités supplémentaires ou une évolution des exigences en matière de performances et de convivialité.

### 1.4.3 Solution proposé [🔗](#)

- **Migration vers Amazon Athena** : Remplacez Amazon Redshift par Amazon Athena pour le stockage et l'interrogation des données. Athena permet d'exécuter des requêtes SQL sur les données stockées dans Amazon S3 sans nécessiter de provisionnement ni de gestion de l'infrastructure.
- **Utilisation de QuickSight pour le Tableau de Bord** : Intégrez Amazon QuickSight comme outil de visualisation pour créer et gérer le tableau de bord de suivi de l'activité commerciale. QuickSight offre des fonctionnalités avancées de création de visualisations interactives et de partage de données.
- **Architecture Serveless** : Adoptez une architecture Serveless en utilisant Athena et S3 pour stocker et interroger les données. Cette approche permet de réduire les coûts opérationnels et de s'adapter plus facilement aux variations de charge et de volume de données.



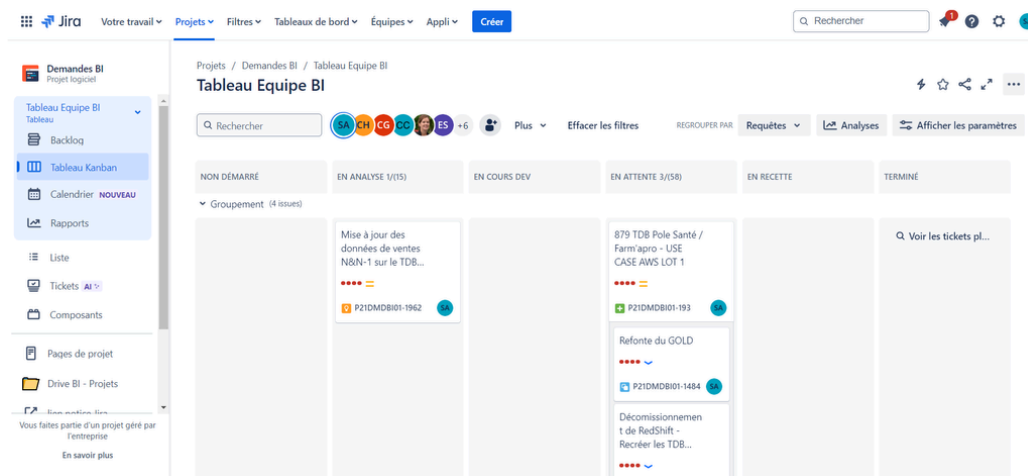
- **Optimisation des Performances** : Configurez Athena pour optimiser les performances des requêtes SQL en utilisant des partitions, des indexes et des techniques de compression de données. Cela garantira des temps de réponse rapides même avec de grandes quantités de données.
- **Formation et Support** : Fournissez une formation approfondie aux utilisateurs finaux sur l'utilisation de QuickSight, ainsi qu'un support continu pour résoudre les problèmes et répondre aux questions.
- **Planification de la Migration** : Élaborez un plan détaillé pour la migration de l'infrastructure actuelle vers la nouvelle solution, en identifiant les étapes clés, les ressources nécessaires et les échéanciers. Assurez-vous d'inclure des tests approfondis pour valider la performance et la fiabilité de la nouvelle architecture.
- **Communication et Gestion du Changement** : Communiquez de manière transparente avec les parties prenantes sur la transition vers Athena et QuickSight, en mettant en avant les avantages en termes de coûts, de performances et de convivialité. Impliquez les utilisateurs finaux dans le processus pour garantir leur adhésion et leur satisfaction.

## 1.5 Méthodologies

Dans le cadre de la gestion de projet, différentes méthodologies peuvent être utilisées pour organiser et suivre le travail de manière efficace. Parmi celles-ci, la méthodologie Kanban offre une approche visuelle et itérative, axée sur l'optimisation du flux de travail et la livraison continue.

### 1.5.1 Méthodologie Kanban avec Jira

La méthodologie Kanban peut être pleinement intégrée et gérée à travers l'utilisation de l'outil Jira, une plateforme de gestion de projet populaire développée par Atlassian. Voici comment la méthodologie Kanban peut être implémentée à travers Jira :



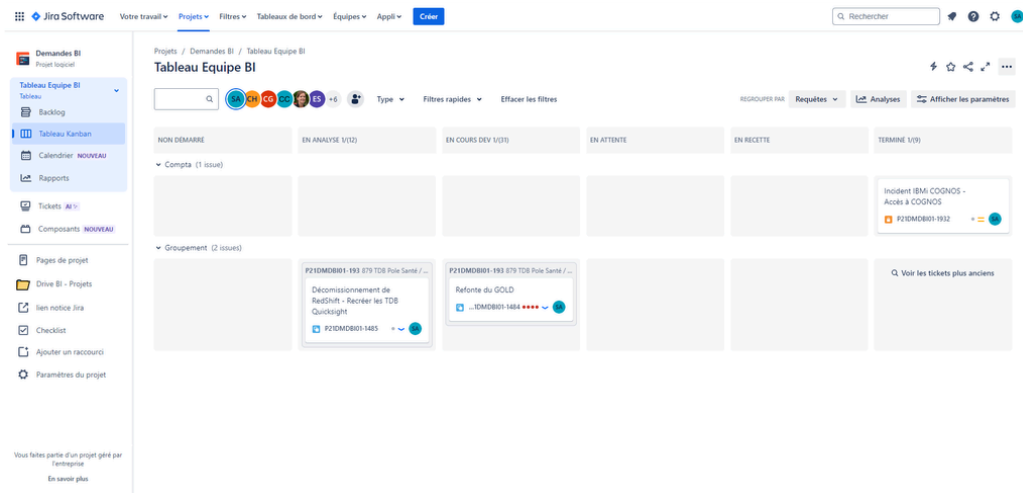


Tableau Jira

En utilisant Jira avec la méthodologie Kanban, votre équipe peut bénéficier d'une gestion de projet efficace, axée sur la visualisation du travail, la limitation du travail en cours, la gestion des priorités et l'amélioration continue. Cela permettra à votre équipe de travailler de manière plus productive et de livrer des résultats de meilleure qualité.

### 1.5.2 Les avantages de la méthodologie Kanban [🔗](#)

La méthodologie Kanban offre plusieurs avantages significatifs pour la gestion de projet. Voici quelques-uns des principaux avantages :

- **Visibilité du Travail** : Kanban offre une visualisation claire et en temps réel du travail en cours grâce à son tableau Kanban. Cela permet à l'équipe de comprendre rapidement ce sur quoi elle travaille actuellement et de prendre des décisions éclairées sur les priorités et les ressources nécessaires.
- **Limitation du Travail en Cours (WIP)** : En limitant le nombre de tâches autorisées dans chaque colonne du tableau Kanban, Kanban permet de maintenir un flux de travail régulier et d'identifier rapidement les goulots d'étranglement. Cela aide à réduire les retards et à optimiser les délais de livraison.
- **Flexibilité** : Kanban est une méthodologie flexible qui peut s'adapter à différents types de projets et environnements de travail. Il permet aux équipes de s'ajuster rapidement aux changements de priorités et de besoins du client, tout en maintenant un flux de travail régulier et prévisible.
- **Amélioration Continue** : Kanban encourage l'amélioration continue en identifiant et en résolvant les problèmes dès qu'ils surviennent. En utilisant des métriques telles que le temps de cycle et le taux de complétion, les équipes peuvent identifier les opportunités d'optimisation et mettre en œuvre des ajustements pour améliorer leur efficacité.
- **Collaboration Efficace** : Kanban favorise une communication claire et transparente au sein de l'équipe en utilisant un tableau visuel pour représenter le travail en cours, les priorités et les blocages éventuels. Cela facilite la collaboration et la résolution rapide des problèmes.
- **Réduction des Temps de Cycle** : En identifiant et en éliminant les gaspillages et les inefficacités dans le processus de travail, Kanban permet de réduire les temps de cycle et d'optimiser les délais de livraison. Cela permet à l'équipe de livrer plus rapidement et plus efficacement des résultats de haute qualité.

En résumé, la méthodologie Kanban offre une approche simple et efficace pour la gestion de projet, en mettant l'accent sur la visualisation du travail, la limitation du travail en cours, la gestion visuelle et l'amélioration continue. Ses nombreux avantages en font un outil précieux pour les équipes cherchant à optimiser leur productivité et leur efficacité dans la livraison de projets.

### 1.5.3 Etude comparative des méthodologies [🔗](#)

Dans le domaine de la gestion de projet, plusieurs méthodologies sont disponibles, chacune avec ses propres caractéristiques, avantages et inconvénients. Voici une étude comparative des méthodologies Kanban, Scrum et Agile :

Caractéristiques	Kanban	Scrum	Agile
<b>Principes Clés</b>	Visualisation du travail, limitation du WIP, gestion visuelle, amélioration continue	Itérations de travail ("sprints"), réunions quotidiennes de stand-up, rôles définis (Product Owner, Scrum Master, Équipe de Développement)	Collaboration avec le client, livraison continue de logiciels, adaptation aux changements, travail en équipe autoorganisée
<b>Flexibilité</b>	Grande flexibilité, adapté aux exigences changeantes	Flexible mais suit une structure formelle, avec des sprints définis	Favorise la flexibilité en encourageant l'adaptation aux changements des exigences du client
<b>Visibilité du Travail</b>	Offre une bonne visibilité du travail en cours grâce à un tableau visuel	Offre une bonne visibilité du travail en cours grâce à des réunions régulières et des tableaux visuels	Met l'accent sur la transparence et la communication ouverte pour assurer une visibilité maximale du travail
<b>Structure</b>	Moins structuré mais offre une approche souple	Offre une structure formelle avec des rôles définis et des itérations de travail régulières	Philosophie qui peut être mise en œuvre de manière flexible en fonction des besoins du projet et de l'organisation
<b>Avantages</b>	Flexibilité, visibilité du travail en cours, réduction des temps de cycle	Structure formelle, prévisibilité des délais, transparence dans le processus de développement	Réactivité aux changements des exigences, engagement du client, communication ouverte
<b>Inconvénients</b>	Risque de devenir trop flexible, nécessite une discipline pour maintenir les limites de WIP	Risque de surcharge de réunions, peut être trop rigide pour certains projets	Nécessite une forte implication du client, risque de manque de documentation

Tableau comparatifs des méthodologies

Ce tableau offre une comparaison claire des caractéristiques, des avantages et des inconvénients de chaque méthodologie, ce qui peut aider à prendre une décision éclairée sur celle qui convient le mieux à votre projet et à votre équipe.

## 1.6 Conclusion

Dans cette section, nous avons contextualisé le projet en présentant d'abord l'organisme d'accueil, puis en examinant l'existant ainsi que les motivations sous-jacentes au projet. Dans la section à venir, nous explorerons en détail les fonctionnalités offertes par notre solution, approfondissant ainsi notre compréhension du sujet.

# Chapitre 2 : L'informatique décisionnelle

## 2.1 Introduction

Ce chapitre sera réservé pour définir l'informatique décisionnelle. Nous présentons dans un premier temps ses avantages et ses limites. Nous abordons, ensuite, ses termes et les concepts clés en détaillant la notion d'ETL et d'entrepôt de données.

## 2.2 Concepts généraux du BI

### 2.2.1 La Business Intelligence

L'informatique décisionnelle, également Business Intelligence ou BI en anglais, désigne les moyens, les méthodes et les outils qui apportent des solutions en vue d'offrir une aide à la décision aux professionnels afin de leur permettre d'avoir une vue d'ensemble sur l'activité de l'entreprise et de leur permettre de prendre des décisions plus avisées à travers des tableaux de bord de suivi et des analyses.

### 2.2.2 Avantages du BI

Déployer une solution BI apporte de nombreux avantages :

- Améliorer la visibilité sur les chiffres, les écarts et les anomalies.
- La combinaison de plusieurs sources de données (ERP, systèmes comptable, feuilles de calcul, des budgets ...).
- La présentation uniforme d'informations fiables.
- L'automatisation permettant l'accélération de la collecte et de la diffusion de l'information.
- La performance dans le calcul d'agrégats sur de gros volume de données.
- La prise de décision grâce à des indicateurs pertinents et à une structure cohérente des informations.
- La prise de décision grâce à des indicateurs pertinents et à une structure cohérente des informations.

### 2.2.3 Les limites du BI

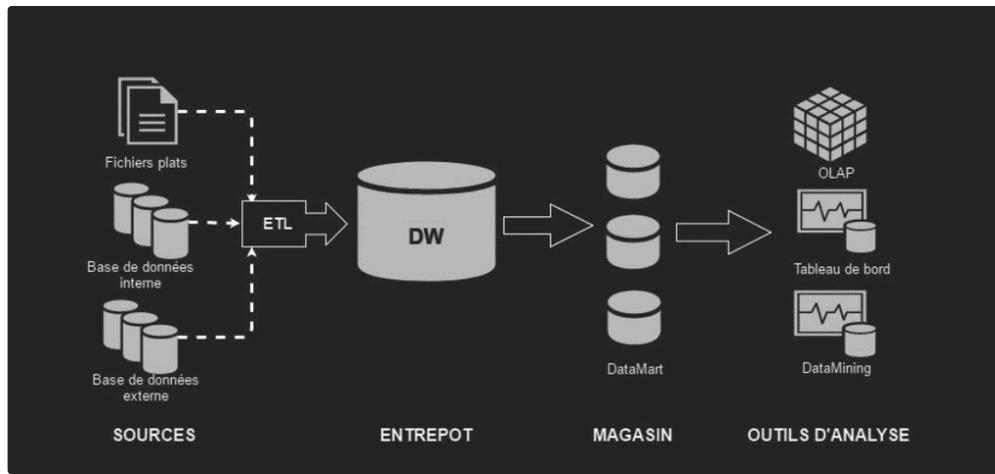
Parmi les limites de la Business Intelligence :

- La mise en place d'une solution de BI prend beaucoup du temps : de nombreuses entreprises dans le scénario industriel rapide ne sont pas assez patientes pour attendre la mise en place du système décisionnel dans leur organisation.
- Complexité : un autre inconvénient de BI pourrait être sa complexité dans la mise en œuvre des données.
- Erreur : les résultats produits par les systèmes décisionnels sont le résultat de conceptions informatiques et mathématiques complexes, qui peuvent révéler des erreurs, par ailleurs les résultats sont souvent statistiques, donc non déterministes. La possibilité d'une erreur ou une approximation inadaptée devra toujours être prise en compte dans les décisions.

## 2.3 Les principes des systèmes décisionnels

Le système décisionnel est architecturé de la façon suivante :

- Plusieurs sources de données en lecture
- Un entrepôt de données fusionnant les données requises.
- Un ETL permettant d'alimenter l'entrepôt de données à partir des données existantes.
- Des magasins de données permettant de simplifier l'entrepôt de données
- Des applications d'exploitation de données pour présenter l'étude aux utilisateurs finaux et décideurs.



*Architecture générale d'un système décisionnel*

### 2.3.1 Sources de données 🔗

Afin d'alimenter l'entrepôt, les informations doivent être identifiées et extraites de leurs emplacements originaux. Il s'agit des sources de données hétérogènes qui peuvent comporter des données internes à l'entreprise, stockées dans les bases de données de production des différents services.

Elles peuvent être aussi des sources externes, récupérées via des services distants et des web services ou des sources qui peuvent être sous format de fichier plats comme le cas de notre projet.

### 2.3.2 Entrepôt de données 🔗

D'après **BILL Inmon** : "Un entrepôt de données est une collection de données thématiques, intégrées, non volatiles et historiées, organisées pour la prise de décision"

D'après cette définition nous distinguons les caractéristiques suivantes :

- Données Orientées sujet : les données des entrepôts sont organisées par sujet et donc triées par thème.
- Données intégrées : les données provenant des différentes sources doivent être intégrées avant leur stockage dans l'entrepôt de données. Un nettoyage préalable des données est nécessaire afin d'avoir une cohérence et une normalisation de l'information.
- Données non-volatiles : à la différence des données opérationnelles, celles de

l'entrepôt sont permanentes et ne peuvent pas être modifiées. Le rafraîchissement de l'entrepôt, consiste à ajouter de nouvelles données sans perdre celles qui existent.

- Historiées : les données doivent être datées.

### 2.3.3 Extract-Transform-Load 🔗

ETL acronyme d'extraction transformation loading est un processus d'intégration des données. Il permet de transférer des données brutes d'un système source, de les préparer pour une utilisation et de les envoyer vers l'entrepôt de données.

Ce système doit faire passer les données par un tas de processus pour les dénormaliser, les nettoyer, les contextualiser, puis de les charger de la façon adéquate.

Cependant, la réalisation de l'ETL est une étape très importante et très complexe parce qu'il constitue 70% d'un projet décisionnel en moyenne.

### 2.3.4 La différence entre OLTP et OLAP : 🔗

Les systèmes informatiques peuvent se subdiviser en deux catégories : **Les systèmes transactionnels OLTP** (Online Transaction Processing) et **les systèmes analytiques OLAP** (Online Analytical Processing) :

- Les systèmes OLTP sont dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes et donc directement opérationnels, le mode de travail est transactionnel, l'objectif est de pouvoir insérer, modifier et interroger rapidement et en sécurité la base. Ces actions doivent pouvoir être effectuées très rapidement par de nombreux utilisateurs

simultanément. Il est proposé essentiellement pour les applications gérant des opérations commerciales comme les opérations bancaires, ou l'achat de bien divers.

- Les systèmes OLAP sont dédiés au management de l'entreprise pour l'aider au pilotage de l'activité. C'est un outil de reporting dont la couche d'analyse permet de générer des résultats en fonction du contenu d'un entrepôt de données. Les programmes consultent une quantité importante de données pour procéder à des analyses. Les objectifs principaux sont : regrouper, organiser des informations provenant de sources diverses, les intégrer et les stocker pour permettre à l'utilisateur de retrouver et analyser l'information facilement et rapidement.

Bien que les systèmes d'informations OLTP et OLAP aient le point commun de regrouper les données de l'entreprise dans un SGBD et d'en fournir l'accès aux utilisateurs, ils présentent de profondes différences :

Aspect	OLTP	OLAP
Objectif	Gérer et enregistrer les transactions en temps réel	Analyser et interroger les données historiques
Type de données	Données opérationnelles en temps réel	Données agrégées et historiques
Utilisation principale	Applications transactionnelles	Applications décisionnelles et analytiques
Structure de données	Base de données relationnelle	Base de données multidimensionnelle
Modélisation	Modèle en étoile ou en flocon	Modèle dimensionnel
Requêtes	Simple et transactionnelles	Complexes et analytiques
Performance	Haute vitesse pour de nombreux utilisateurs	Moins critique pour la latence, mais nécessite une réponse rapide pour les requêtes analytiques
Volume de données	Gère de grandes quantités de transactions individuelles	Traite de grandes quantités de données agrégées
Mises à jour	Fréquentes et en temps réel	Occasionnelles et planifiées
Exemples d'applications	Systèmes de gestion des ventes, des achats, des réservations	Tableaux de bord, rapports, analyses de données
Exigences matérielles	Performances rapides, mise à l'échelle horizontale	Puissance de calcul élevée, capacité de stockage importante
Systèmes populaires	Oracle, MySQL, SQL Server	Microsoft SQL Server Analysis Services, SAP BW

### 2.3.5 Conclusion:

Dans ce chapitre, nous avons détaillé toutes les notions relatives aux systèmes décisionnels, pour les maîtriser afin de favoriser le bon déroulement du projet et puis nous revenons sur nos recherches à propos des différentes plateformes du cloud computing

## Chapitre 3 : Les Différentes Plateformes Cloud

### 3.1. Introduction au Cloud Computing

Le cloud computing révolutionne la manière dont les entreprises gèrent leurs ressources informatiques en permettant l'accès à des services via Internet. Il offre une flexibilité, une évolutivité et une efficacité opérationnelle sans précédent.

### 3.2. Modèles de Service Cloud

Les principaux modèles de service cloud incluent l'Infrastructure as a Service (IaaS), le Platform as a Service (PaaS) et le Software as a Service (SaaS). Ces modèles permettent aux entreprises de choisir le niveau de contrôle et de responsabilité qui leur convient.

### 3.3 Avantages du Cloud

Le cloud computing présente de nombreux avantages, notamment la réduction des coûts initiaux, la flexibilité pour ajuster les ressources selon les besoins, et la possibilité de se concentrer sur l'innovation plutôt que sur la gestion de l'infrastructure.

### 3.4. Inconvénients du Cloud

Cependant, le cloud computing comporte également des inconvénients, tels que la dépendance à l'égard de la connectivité Internet, les problèmes de sécurité et de confidentialité des données, ainsi que les risques de pannes de service.

### 3.5. Modèles de Déploiement Cloud

Il existe plusieurs modèles de déploiement cloud, notamment le cloud public, le cloud privé et le cloud hybride. Chaque modèle offre des avantages et des inconvénients uniques en fonction des besoins et des contraintes de l'entreprise.

### 3.6. Comparaison entre AWS, Azure et Oracle Cloud

#### AWS (Amazon Web Services)

- Leader du marché avec une vaste gamme de services cloud.
- Large présence mondiale avec de nombreux centres de données.
- Large écosystème de partenaires et de communautés.
- Fortement axé sur l'innovation et la rapidité de mise sur le marché.

#### Azure de Microsoft

- Intégration étroite avec les outils et les technologies Microsoft.
- Forte présence dans les entreprises utilisant déjà des produits Microsoft.
- Large gamme de services cloud, y compris des solutions PaaS et SaaS.
- Mises à jour fréquentes et intégration avec les outils de développement.

#### Oracle Cloud

- Spécialisé dans les solutions cloud pour les applications d'entreprise et les bases de données.
- Intégration étroite avec les produits et les solutions Oracle.
- Axé sur la sécurité et la conformité, en particulier pour les entreprises réglementées.

- Offre des services PaaS et IaaS ainsi que des solutions SaaS dans certains domaines.

### 3.7.Conclusion [↗](#)

Le cloud computing a révolutionné la manière dont les entreprises gèrent leurs ressources informatiques, offrant une flexibilité et une évolutivité sans précédent. Les plateformes cloud telles qu'AWS, Azure et Oracle Cloud offrent une gamme de services et de solutions pour répondre aux besoins variés des entreprises. Chacune de ces plateformes a ses propres forces et faiblesses, et le choix entre elles dépend des besoins spécifiques de chaque entreprise, de ses exigences en matière de sécurité et de conformité, ainsi que de sa stratégie technologique globale. En tirant parti des avantages du cloud computing et en tenant compte des considérations spécifiques à chaque plateforme, les entreprises peuvent réaliser des gains d'efficacité, stimuler l'innovation et rester compétitives dans un marché en constante évolution.

## Chapitre 4 : Mise en Œuvre de la Solution sur AWS Athena [↗](#)

### 4.1 Introduction & Objectifs [↗](#)

Voici les objectifs principaux pour le projet "Cooperl Groupement d'Éleveurs Pôle Santé" :

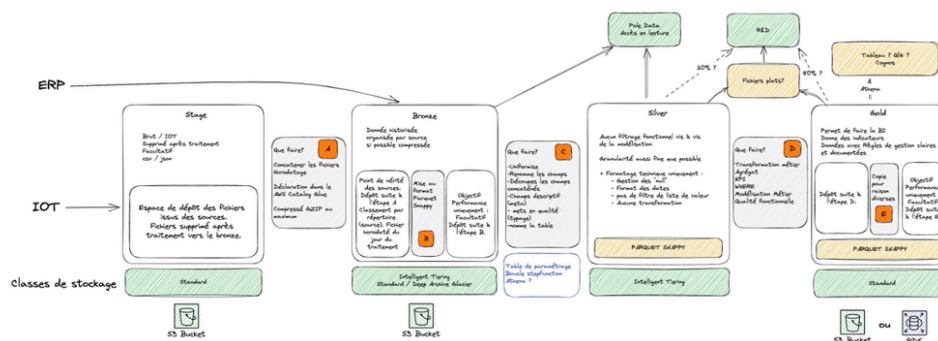
1. Migration vers QuickSight et AWS Athena : Le premier objectif est de migrer avec succès l'infrastructure de Business Intelligence de Cooperl de Amazon Redshift vers QuickSight pour la visualisation des données et AWS Athena pour l'interrogation des données stockées sur Amazon S3. Cette migration devrait être réalisée de manière fluide, en minimisant les interruptions pour les utilisateurs finaux.
2. Réduction des coûts : Le projet vise à réduire les coûts associés à l'infrastructure de Business Intelligence en abandonnant Amazon Redshift au profit de QuickSight et d'AWS Athena. Cette transition devrait permettre des économies significatives en termes de frais de stockage, de requêtage et de gestion de base de données.
3. Amélioration de la performance et de la scalabilité : En adoptant QuickSight pour la visualisation des données et AWS Athena pour l'interrogation des données sans serveur, le projet vise à améliorer la performance et la scalabilité de l'infrastructure de Business Intelligence de Cooperl. Cela permettra de répondre plus efficacement aux besoins croissants en analyse de données de l'entreprise.
4. Facilitation de l'analyse des données : L'objectif est de fournir aux équipes métier de Cooperl des outils puissants et conviviaux pour l'analyse des données, leur permettant d'explorer facilement les données, de créer des visualisations personnalisées et de prendre des décisions éclairées basées sur les insights obtenus.
5. Renforcement de la compétitivité : En modernisant son infrastructure de Business Intelligence, Cooperl renforce sa capacité à rester compétitif sur le marché en accélérant le processus de prise de décision, en améliorant la qualité des insights fournis et en répondant de manière agile aux évolutions du marché et des besoins des clients.

### 4.2 Architecture de la solution [↗](#)

#### 4.2.1 Schéma de l'architecture [↗](#)

Cette image illustre l'architecture d'un Data Warehouse (DWH) au sein du pôle Data qui utilise le théorème de couche sémantique , en utilisant des technologies AWS pour la gestion et le traitement des données provenant de différentes sources:





Entrepôt de données au sein du pole DATA COOPERL

## Sources de Données :

**ERP et IoT :** Les données brutes issues des systèmes ERP et des dispositifs IoT sont déposées dans des buckets S3 en format CSV ou JSON.

## Zone de Stage :

**Stage :** Les données brutes sont stockées temporairement dans une zone de stage sur S3. Ces fichiers sont supprimés après traitement. Le stockage utilise la classe standard.

## Zone Bronze :

**Données historisées :** Les données sont organisées par source et sont souvent compressées.

**Étape de traitement :** Les fichiers sont vérifiés et nettoyés avant d'être transformés. Les données sont mises en format Parquet ou Snappy et peuvent être déplacées vers un stockage à long terme, comme Deep Archive Glacier.

**Actions :** Normalisation des champs, découpage des fichiers, et ajout de métadonnées descriptives.

## Zone Silver :

**Filtrage fonctionnel :** Les données sont filtrées pour répondre aux besoins de modélisation avec une granularité aussi fine que possible.

**Transformation :** Les données sont formatées techniquement, les nulls sont gérés, et les formats des dates sont standardisés.

**Stockage :** Utilisation de Parquet Snappy avec Intelligent Tiering pour le stockage.

## Responsabilités des Équipes :

**Équipe Data Engineering :** L'équipe de data engineering est responsable de toutes les étapes jusqu'à la zone Silver incluse. Cela comprend la collecte, le nettoyage, la transformation, et le stockage des données dans les zones de Stage, Bronze, et Silver.

**Équipe BI :** Pour l'équipe de Business Intelligence (BI), la source des données est la zone Silver. Leur objectif est de transformer ces données Silver en données Gold, prêtes pour l'analyse et la visualisation. La couche visée par l'équipe BI est donc la couche Gold, où les données sont agrégées, enrichies, et optimisées pour des outils de BI comme QuickSight ou Tableau.

## Zone Gold :

**Analyse et BI :** Les données sont prêtes pour les analyses et la Business Intelligence. Elles sont agrégées, transformées et modélisées pour la qualité métier.

**Actions :** Application d'agrégats, filtres, et transformations métiers. Les données sont stockées en format Parquet Snappy.

**Sortie :** Les données transformées sont disponibles pour des outils de BI comme QuickSight ou Tableau via Athena.

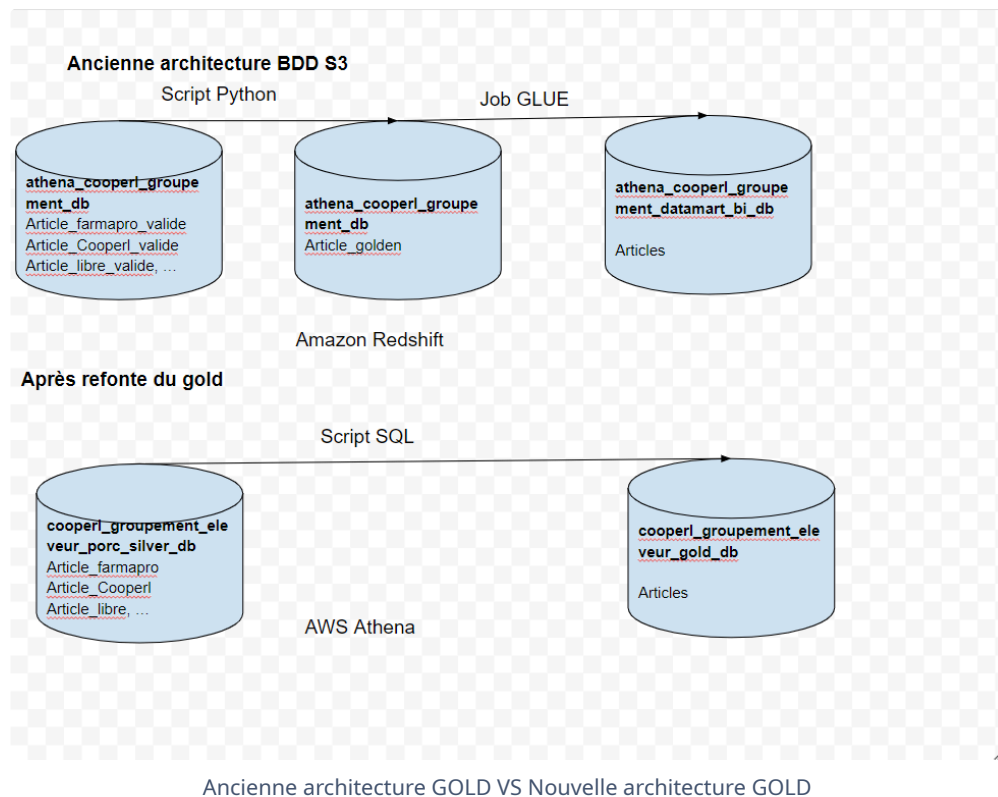
## Stockage et RDS :

Les données peuvent être stockées soit dans S3 pour des analyses ultérieures via Athena ou directement dans RDS pour une utilisation plus structurée.

## Flux des Données :

- Les données passent par plusieurs étapes de transformation (Stage, Bronze, Silver, Gold) pour assurer leur qualité et pertinence pour l'analyse.

- À chaque étape, des transformations spécifiques sont appliquées pour préparer les données à l'étape suivante, culminant dans un ensemble de données prêtes pour l'analyse approfondie et la visualisation par des outils BI.

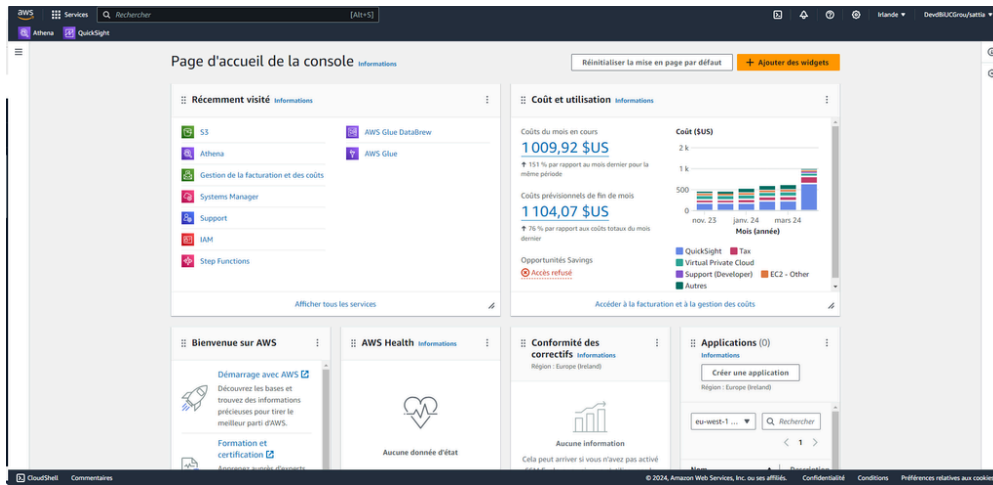


Cette architecture permet une gestion efficace des données, en assurant leur qualité et leur disponibilité pour des analyses décisionnelles au sein du pôle Data. L'équipe de data engineering se concentre sur le traitement initial et intermédiaire des données, tandis que l'équipe BI se concentre sur l'exploitation des données finales pour des analyses et des visualisations détaillées.

#### 4.2.2 Composants utilisés (Athena, S3, Glue) [🔗](#)

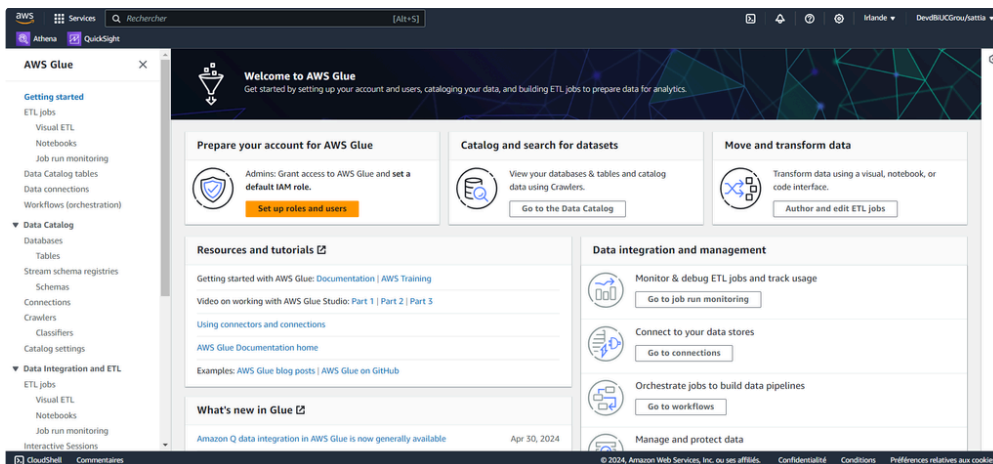
##### 1. Console AWS :

Pour faciliter la gestion et le suivi de notre infrastructure cloud, nous nous appuyons sur la Console AWS, une interface conviviale et intuitive qui offre une vue unifiée de tous les services et ressources AWS que nous utilisons. Grâce à la Console AWS, notre équipe peut facilement provisionner, surveiller et gérer nos services QuickSight, Athena et Amazon S3, ainsi que d'autres services AWS pertinents pour notre projet. Cette interface centralisée nous permet de naviguer efficacement entre les différents services, d'accéder rapidement aux fonctionnalités de gestion et de configuration, et de visualiser en temps réel les métriques et les indicateurs de performance essentiels. En utilisant la Console AWS, nous renforçons notre capacité à orchestrer harmonieusement notre environnement cloud, à optimiser nos ressources et à garantir le bon fonctionnement de notre infrastructure de Business Intelligence chez Cooperl.



## 2. AWS Glue :

est un service entièrement managé d'extraction, transformation et chargement (ETL) qui facilite la préparation et le chargement des données pour l'analyse. Il permet aux utilisateurs de découvrir, préparer et combiner des données à des fins d'analyses, de création de rapports et d'apprentissage automatique.



Console AWS Glue

### Principales fonctionnalités d'AWS Glue :

#### Catalogue de données :

Un méta-référentiel centralisé pour stocker les métadonnées des données.  
Permet de découvrir et d'organiser les données provenant de diverses sources.

#### Moteur ETL :

Exécution de tâches ETL pour transformer, nettoyer et enrichir les données.  
Supporte les transformations basées sur Python et Scala.

#### Planificateur de tâches :

Permet de planifier et d'exécuter des workflows ETL de manière automatisée.

#### Détection automatique du schéma :

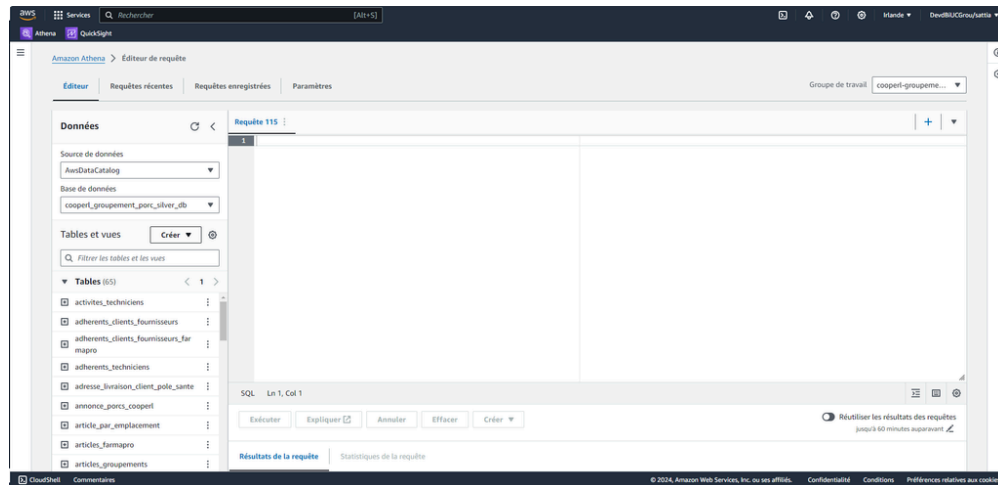
AWS Glue peut découvrir automatiquement le schéma des données stockées dans Amazon S3, les bases de données et d'autres sources.

#### Flexibilité et scalabilité :

Ajuste automatiquement les ressources nécessaires pour exécuter les tâches ETL, assurant ainsi une performance optimale et un coût maîtrisé.

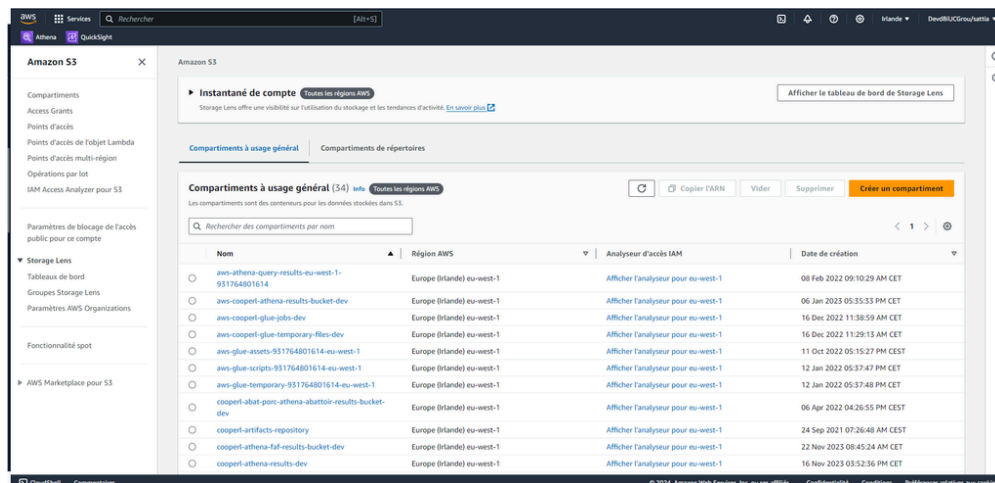
### 3. AWS Athena :

Pour l'interrogation de nos données stockées dans Amazon S3, nous avons choisi AWS Athena, un service de requêtage sans serveur. Cette technologie nous permet d'exécuter des requêtes SQL standard sur nos données, sans avoir à gérer d'infrastructure sous-jacente. En adoptant Athena, nous bénéficions de la scalabilité automatique et de la tarification à l'utilisation, ce qui nous permet de réduire les coûts tout en garantissant des performances optimales pour nos besoins en analyse de données.



### 4. Amazon S3 :

Au cœur de notre architecture de données se trouve Amazon Simple Storage Service (S3), un service de stockage d'objets cloud évolutif et hautement disponible. Amazon S3 constitue le référentiel central où sont stockées toutes nos données brutes, qu'elles soient structurées ou non. Sa fiabilité, sa durabilité et sa capacité à stocker des volumes massifs de données en font un choix idéal pour notre infrastructure de données. En utilisant S3 comme notre lac de données principal, nous pouvons stocker efficacement des données de différents types et tailles, prêtes à être exploitées par nos outils d'analyse tels que QuickSight et AWS Athena. De plus, l'intégration transparente entre S3, QuickSight et Athena nous permet de tirer pleinement parti de nos données pour obtenir des insights stratégiques et prendre des décisions éclairées.



### 4.3.1 Modélisation des données

L'objectif principal de la modélisation des données dans le contexte de la BI est de fournir une représentation logique et cohérente des données de l'entreprise, en les structurant de manière à refléter les processus métier, les interactions entre les entités et les dimensions pertinentes pour l'analyse. Cette modélisation permet aux utilisateurs de la BI de naviguer facilement à travers les données, de les interroger de manière intuitive et de générer des rapports et des analyses significatifs.

**Modèle conceptuel des données (MCD) :** Il s'agit d'une représentation abstraite des concepts métier et des relations entre eux. Le MCD fournit une vue globale des données sans se soucier des détails techniques de leur stockage.



Dans le cadre de notre solution BI sur AWS, nous opérons avec deux environnements distincts : développement et production. Nous entamons notre processus de création des tables dans l'environnement de développement (dev). Une fois que toutes les

tables sont correctement conçues, testées et validées dans l'environnement de développement, nous procédons à leur déploiement vers l'environnement de production (prod).

Nous adoptons une approche itérative et sécurisée, où nous rassemblons tous les scripts de création de tables dans un fichier JSON, accompagné d'un fichier de test d'acceptance des tables. Ce fichier global sera ensuite transmis à l'équipe de data engineering, qui est responsable de l'infrastructure AWS Glue pour l'ETL (Extract, Transform, Load). Une fois que l'équipe de data engineering a validé les scripts et les tests d'acceptance, nous procédons à la bascule vers l'environnement de production. Dans la section suivante (4.4 Processus ETL), nous détaillerons le processus d'ETL, y compris l'interaction avec l'équipe de data engineering pour assurer une transition fluide des données de l'environnement de développement vers l'environnement de production.

#### Listes des scripts des tables créées :


On a créé un fichier json où on a stocké tous les scripts des tables créés à fin de l'envoyer à l'équipe data engineering pour la migration en production. Le Fichier est sur ce format :

```
"Tables" :[
  {"Nom_de_la_table": [
    "Script du drop Table",
    "Script du create Table ",
    "Script de l'insert Table " ]},
  le reste des tables ]
```


Fichier json :  Cooperl\_groupement\_porc\_gold\_db.json

## 4.4 Processus ETL (Extract, Transform, Load)

Dans l'architecture de l'entreprise sur le cloud on peut pas développer ou exécuter des grandes requêtes sur l'environnement de production en tant qu'ingénieur BI puisque c'est l'équipe de data engineering qui en est responsable ce qui fait que la démarche va être de cette façon :

On développe la couche de la base de données sur le serveur de développement sur le bon périmètre, on prend une plage de temps pour faire une comparaison entre le développement et la production, on valide nos tests d'acceptance de script de tables, on crée le fichier d'acceptance :  CGP\_Pole\_sante\_acceptance.xlsx, on met à disposition nos scripts dans le fichier json et après on les envoie à l'équipe data engineering pour qu'elle puisse faire la migration en utilisant lambda/Ariflow et après reconstituer le TDB existant en utilisant la nouvelle couche de BDD créer et finir avec décommissionner l'ancienne couche + architecture ancienne de la BDD et décommissionner l'ancien TDB.

## 4.5 Optimisation des requêtes sur Athena

 Amazon Athena est un service interactif de requête qui permet d'analyser des données directement dans Amazon S3 en utilisant SQL standard. Pour optimiser les requêtes sur Athena, il est crucial de se concentrer sur l'indexation, le partitionnement et les bonnes pratiques de requêtage.

### 4.5.1 Bonnes pratiques pour les requêtes

Pour  écrire des requêtes performantes sur Athena, voici quelques bonnes pratiques :

- **Sélectionner uniquement les colonnes nécessaires** : Éviter d'utiliser `SELECT *`, et préférer spécifier uniquement les colonnes nécessaires pour réduire le volume de données scannées.
- **Utiliser la clause `WHERE`** : Réduire le nombre de lignes traitées en filtrant les données avec des clauses `WHERE`.
- **Optimisation des jointures** : Lors de l'utilisation de jointures, s'assurer que les colonnes jointes sont de types compatibles et éviter les jointures sur des colonnes non indexées.
- **Exploiter les fonctions de répartition** : Les fonctions comme `GROUP BY`, `ORDER BY`, et `DISTINCT` peuvent être coûteuses. Utiliser des sous-requêtes ou des vues matérialisées pour optimiser ces opérations.

- **Utiliser les tables externes de manière efficace** : Structurer les tables externes avec des partitions et stocker les données dans des formats compressés et colonnes pour optimiser les performances des requêtes.
- **Utiliser les CTE au lieu des sous requêtes** : Une Common Table Expression (CTE) est une expression nommée temporairement dans le cadre d'une requête SQL. Elle est définie au début d'une requête et peut être référencée plusieurs fois dans cette même requête. Voici un exemple de syntaxe CTE :

```
WITH cte_name AS (
    SELECT column1, column2
    FROM table_name
    WHERE condition
)
SELECT *
FROM cte_name;
```

Une CTE peut être référencée plusieurs fois dans la même requête, évitant ainsi la répétition de code et améliorant l'efficacité du développement.

Les CTE permettent de diviser une requête complexe en morceaux plus petits et plus faciles à déboguer. Cela simplifie l'identification des problèmes et améliore le processus de développement.

## 4.6 Conclusion [↗](#)

En conclusion, l'optimisation des requêtes sur Amazon Athena est une tâche cruciale pour améliorer les performances et réduire les coûts d'analyse des données. En mettant en œuvre des techniques d'indexation et de partitionnement, ainsi que des bonnes pratiques de requêtage, il est possible de minimiser le temps de traitement et d'optimiser l'utilisation des ressources. Durant mon stage, j'ai appris à maîtriser ces aspects, ce qui m'a permis de contribuer significativement à l'efficacité des analyses de données au sein de l'entreprise. Cette expérience m'a apporté une compréhension approfondie des outils et des techniques nécessaires pour exceller en tant qu'ingénieur BI.

# Chapitre 5 : Conception et Déploiement du Tableau de Bord sur Amazon QuickSight [↗](#)

## 5.1 Introduction [↗](#)

Dans le cadre de ce projet de Business Intelligence (BI), l'outil de visualisation Amazon QuickSight a été choisi pour créer des tableaux de bord interactifs, permettre l'exploration des données et fournir des insights essentiels aux utilisateurs finaux. QuickSight est une solution de BI basée sur le cloud, intégrée dans l'écosystème AWS, qui facilite l'accès en temps réel aux données et offre une large gamme de fonctionnalités analytiques, adaptées aux besoins de l'organisme d'accueil.

Ce chapitre présente d'abord les principales fonctionnalités et avantages de QuickSight, avant d'aborder le processus de connexion aux sources de données via AWS Athena. Enfin, il détaillera les étapes de conception du tableau de bord et les procédures de déploiement et de partage de celui-ci avec les utilisateurs finaux.

## 5.2 Présentation d'Amazon QuickSight [↗](#)

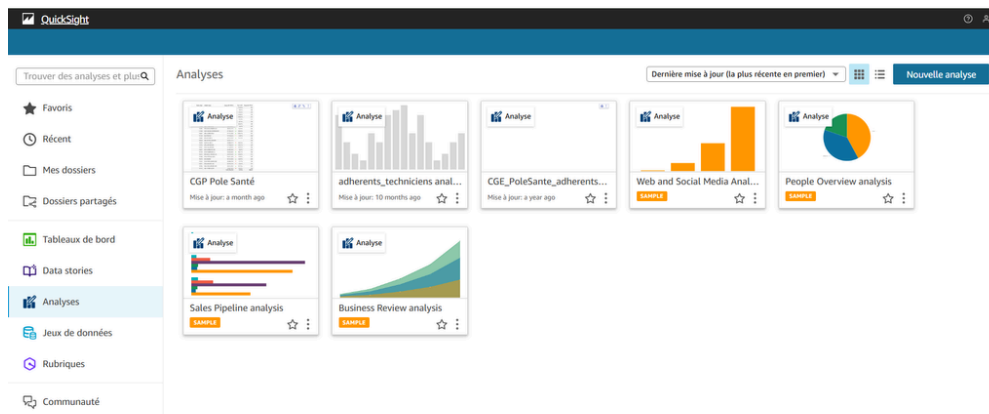
### 5.2.1 Fonctionnalités principales [↗](#)

Amazon QuickSight est un outil de visualisation qui permet aux utilisateurs de créer des tableaux de bord dynamiques et interactifs. Il intègre plusieurs fonctionnalités puissantes pour répondre aux exigences d'analyse des données, notamment :

- **Analyse rapide des données** : Grâce à SPICE (Super-fast, Parallel, In-memory Calculation Engine), QuickSight peut charger des données en mémoire pour des analyses rapides sans nécessiter de requêtes répétées vers la source de données.
- **Personnalisation des visualisations** : QuickSight offre une variété de graphiques (barres, lignes, cartes, etc.), permettant de choisir les visualisations les plus adaptées aux différents indicateurs de performance (KPIs).



- **Intégration native avec AWS** : QuickSight s'intègre facilement aux services AWS comme Athena, S3, et Redshift, simplifiant ainsi la gestion et l'analyse des données stockées sur AWS.
- **Fonctionnalités d'apprentissage automatique** : QuickSight propose des fonctionnalités de détection d'anomalies et des prévisions, permettant aux utilisateurs de repérer les tendances et anomalies dans les données.



### 5.2.2 Avantages de QuickSight [↗](#)

QuickSight présente plusieurs avantages pour les organisations recherchant des solutions de BI performantes et économiques :

- **Évolutivité et performance** : Avec SPICE, QuickSight est capable de gérer de larges volumes de données tout en offrant une performance optimale, même pour des analyses complexes.
- **Flexibilité et coûts ajustables** : Le modèle de tarification flexible de QuickSight (par session ou utilisateur) est avantageux, surtout pour les organisations ayant des besoins ponctuels en termes de visualisation.
- **Accès multiplateforme et interactivité** : Les tableaux de bord peuvent être consultés à partir de différents appareils (ordinateurs, tablettes, smartphones) et peuvent inclure des fonctionnalités interactives (filtres, sélecteurs de date, etc.), offrant ainsi une grande flexibilité aux utilisateurs.

## 5.3 Connexion de QuickSight à Athena [↗](#)

### 5.3.1 Configuration des sources de données [↗](#)

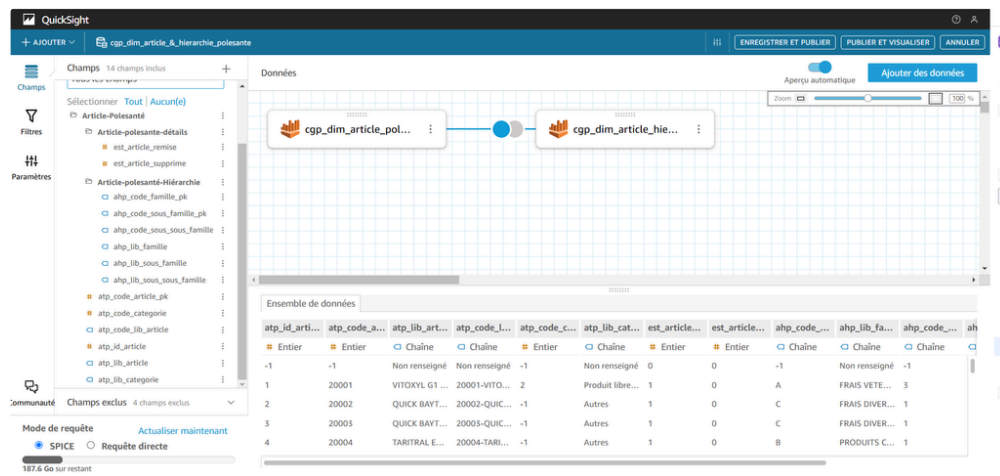
Pour tirer parti des données hébergées dans Amazon S3 et analysées via AWS Athena, il est nécessaire d'établir une connexion sécurisée entre Amazon QuickSight et Athena. Ce processus inclut les étapes suivantes :

1. **Création de l'ensemble de données dans QuickSight** : Dans l'interface de QuickSight, on sélectionne AWS Athena comme source de données. QuickSight récupère ensuite les tables définies dans le catalogue de données AWS Glue, ce qui permet de sélectionner les tables et champs nécessaires pour le tableau de bord.

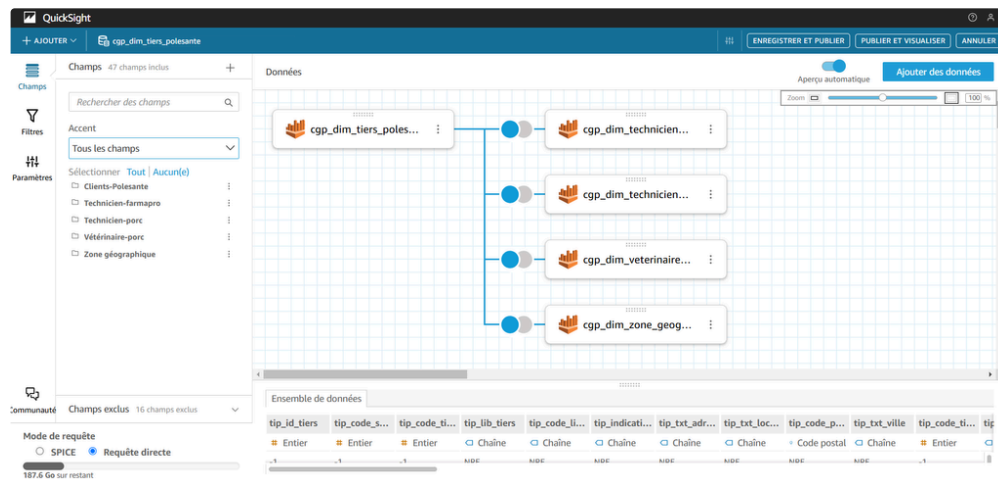
Notre ensemble de données finale contient 3 ensembles de données ( pour ne pas faire plusieurs jointures dans un seul ensemble de données .

**CGP\_DIM\_ARTICLE\_&\_HIERARCHIE\_POLESANTE :**

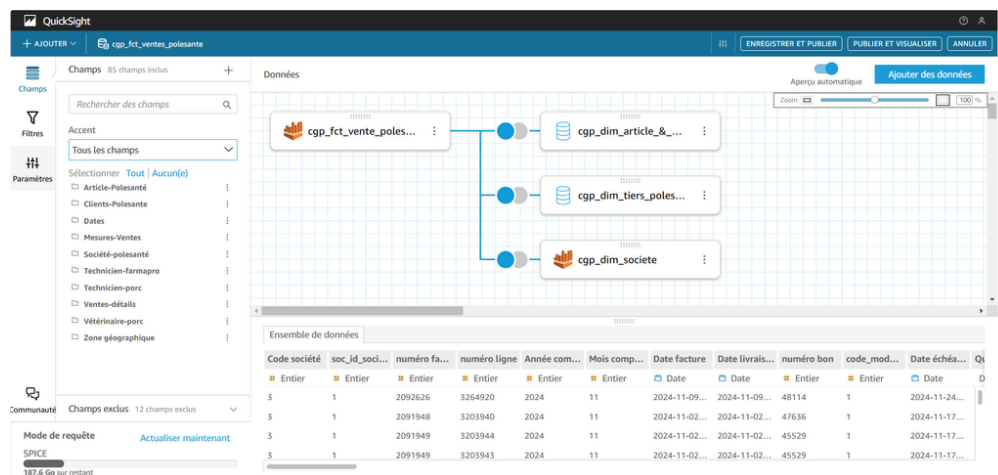




## CGP\_DIM\_TIERS\_POLESANTE :



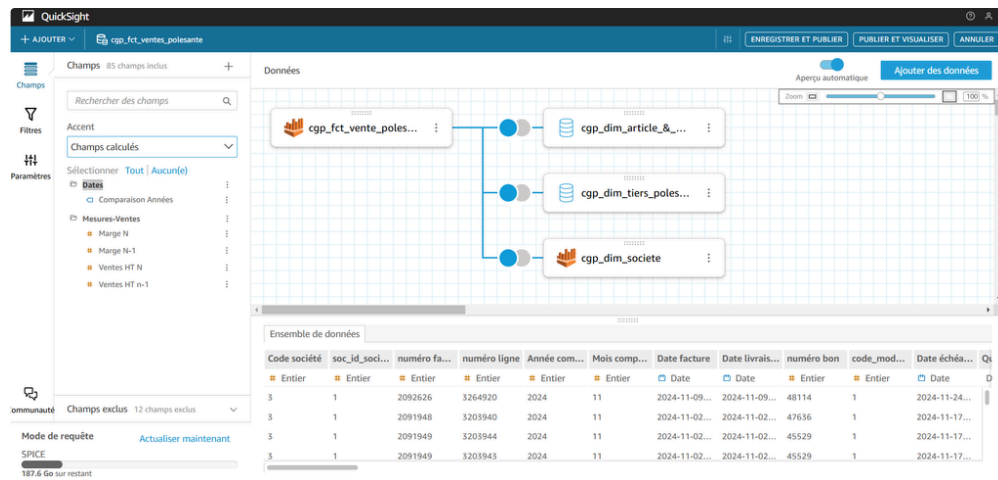
## CGP\_FCT\_VENTES\_POLESANTE :



On peut remarquer sur les 3 jeu de données dans la barres des champs , qu'on a renommé tous les champs pour but analyse (simplifier + les mettre dans des sous dossier pour ne pas passer trop de temps à les chercher en construisant les différents KPI)

2. **Préparation des données** : QuickSight propose des options de transformation de données, comme le filtrage, l'agrégation et la création de champs calculés pour préparer les données avant leur visualisation.

On a créer plusieurs champs calculées dans l'analyse pour but d'optimisation de calcul et pour les utiliser après :



3. **Chargement des données avec SPICE** : Pour des performances optimales, les données peuvent être chargées dans SPICE, ce qui permet des temps de réponse rapides et réduit les requêtes directes vers Athena.

### 5.3.2 Sécurisation des accès

La sécurité des données étant un enjeu crucial, QuickSight intègre plusieurs mécanismes de contrôle d'accès pour protéger les données lors de leur visualisation :

- **Gestion des permissions IAM** : Les autorisations IAM (Identity and Access Management) contrôlent l'accès aux sources de données dans AWS. Des rôles et des politiques spécifiques sont configurés pour restreindre les accès aux seules personnes autorisées.
- **Sécurisation au niveau des objets QuickSight** : QuickSight permet de définir des permissions fines au niveau des tableaux de bord, des analyses et des ensembles de données. Ces paramètres de partage garantissent que seuls les utilisateurs finaux validés peuvent accéder aux rapports.

## 5.4 Conception du tableau de bord

### 5.4.1 Identification des indicateurs clés de performance (KPI)

Les indicateurs clés de performance (KPIs) sont les métriques essentielles permettant de suivre l'efficacité et la performance de l'organisme. Dans le cadre de ce projet, les KPIs choisis incluent, par exemple, Avancement du CA, Marge, Montant remises & Avoirs, les 10 articles les plus vendus, les 10 clients les plus importants de l'année, etc. Ces indicateurs sont identifiés en concertation avec les parties prenantes pour s'assurer qu'ils reflètent les objectifs stratégiques de l'organisation.

### 5.4.2 Création des visualisations

Ce TDB Contient plusieurs pages : Synthèse, Farmapro, Cooperl, Cooperl Techniciens, Tarifs, Remises & Avoirs et on va élaborer chaque KPI de ces pages

Chaque KPI est représenté à l'aide de visualisations adaptées pour offrir une interprétation rapide et claire des données :

**Synthèse** : [Synthèse\\_\\_2024-11-18T14\\_02\\_24.pdf](#)

**CA Cooperl à date** : Indicateurs clé de performance qui nous montre la comparaison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1 avec des champs calculés qu'on a créer :

**Ventes HT YTD N** : `periodToDateSum({Ventes HT},{Date facture},YEAR)`

**Ventes HT YTD N-1** : `periodToDateSum({Ventes HT},{Date facture},YEAR, addDateTime(-1,'YYYY',{DATEFIN}))`

Filtres : (code\_societe = 3)

**Avancement CA Cooperl à Date** : Calibre qui nous montre la comparaison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1

Montant remises Cooperl à Date : Indicateurs clés de performances qui compare les ventes qui sont en remises à date vs ventes en remises à date N-1 (filtres : remises = oui )

Montant Avoir Cooperl à Date : Indicateurs clés de performance qui compare les ventes à date avoir vs ventes à date avoir N-1 (filtres : est\_un\_avoir )

CA Farmapro à date : Indicateurs clé de performance qui nous montre la comparaison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1 (filtre : code\_societe = 9 )

Avancement CA Farmapro à Date : Calibre qui nous montre la comparaison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1

Montant remises Farmapro à Date : Indicateurs clés de performances qui compare les ventes qui sont en remises à date vs ventes en remises à date N-1 (filtres : remises = oui )

Montant Avoir Farmapro à Date : Indicateurs clés de performance qui compare les ventes à date avoir vs ventes à date avoir N-1 (filtres : est\_un\_avoir )

Répartition CA par familles Articles (CA HT YTD): Graphique en anneau qui nous montre la répartition du CA par les familles d'articles et on peut descendre deux niveaux sous famille et sous sous famille d'articles

Segmentation du CA HT N YTD par catégories de clients: Graphique à barres horizontales qui répartit le chiffres d'affaires to date par catégorie clients

Quels sont les 10 articles les plus vendus cette année Un tableau qui illustre les 10 articles les plus vendus cette année

Qui sont les 10 clients les plus importants cette année ? : une table qui illustre les 10 clients qui ont générés le plus de chiffres d'affaires

**Farmapro :**  FARMAPRO\_2024-11-18T14\_02\_39.pdf

CA Farmapro à date : Indicateurs clé de performance qui nous montre la comparaison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1 (filtre : code\_societe = 9 )

Avancement CA Farmapro à Date : Calibre qui nous montre la comparaison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1

Nombre de clientes Actifs (derniers 6 mois): indicateur clés de performances qui montres le nombres de clients actifs les derniers 6 mois

Comparaison des ventes HT YTD N & Ventes HT YTD N-1 & Var % par catégorie client :

CA HT Farmapro YTDN&N-1 par mois : Comparaison des chiffres d'affaires hors taxe year to date N&N-1 par mois

Ventes HT YTD N par libellé famille : ventes HT YTD N de Farmapro par famille articles en descendant jusqu'à la sous sous famille d'articles

Croisement du CA YTD N entre les catégories clients & familles de produits : un diagramme de sankey qui illustre les client qui ont achetés des produits et plus précisément quels types de produits

CA Filiales : Tableau qui illustre les filiales Farmapro qui ont générés du chiffres d'affaires


CA d'affaires export filiales : Tableau qui illustre les filiales export qui ont générés du chiffres d'affaires

CA distributeurs france : Tableau qui illustre les distributeurs Français qui ont générés du chiffres d'affaires

Détails des ventes : Tableau croisé dynamique qui illustre les détails de ventes de l'entreprise Farmapro

**Cooperl :**

 COOPERL\_2024-11-18T12\_26\_25.pdf

CA Cooperl à date : Indicateurs clé de performance qui nous montre la com  COOPERL\_2024-11-18T14\_02\_41.pdf paraison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1

Avancement CA Cooperl à Date : Calibre qui nous montre la comparaison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1

Nombre de clientes Actifs (derniers 6 mois) : indicateur clés de performances qui montres le nombres de clients actifs les derniers 6 mois

CA par catégories de produits : Tableau qui indique le CA par catégories de produits

CA HT YTD N&N-1 par mois : Histogramme qui illustre la comparaison du chiffres d'affaires HT YTD N&N-1 par mois

CA par familles de produits : Graphique en anneau qui illustre le chiffres d'affaires par familles de produits pour l'entreprise Cooperl

CA Zone géographiques : Tableau croisé dynamique qui affiche le CA de la société Cooperl par zone géographiques

Détails des ventes : Tableau croisé dynamique qui montre les détails des ventes de la Cooperl

CA par client par Année : Tableau qui montre le chiffres d'affaires par client par année de la Cooperl

### Cooperl Techniciens:

 COOPERL\_Techniciens\_2024-11-18T13\_14\_13.pdf

CA Techniciens à date : Indicateurs clé de performance qui nous montre le CA générés par les techniciens N&N-1

Avancement CA techniciens à date : Calibre qui nous montre la comparaison entre le chiffre d'affaires à date par rapport au chiffre d'affaires à date N-1 générés par les techniciens

Nombre de clientes Actifs (derniers 6 mois) : indicateur clés de performances qui montres le nombres de clients actifs les derniers 6 mois

CA par catégories de client : Tableau qui indique le CA par catégories de produits


Part du CA (YC remises) YTD N par TC(en%) : graphique à secteurs qui représente le CA par techniciens y compris remises

Ventes HT YTD N (YC remises) par famille de produit : graphique à secteurs qui représente le CA par familles de produits y compris remises

Suivi mensuel du CA HT YTD N&N-1 YC(remises) : Histogramme qui illustre la comparaison du chiffres d'affaires HT YTD N&N-1 par mois

Détails des Qté vendues Techniciens Farmapro : Tableau qui illustre les quantités vendues par les techniciens Farmapro

Détails du CA techniciens : Tableau croisé dynamique qui illustre les détails de ventes des techniciens Farmapro

Tarifs :  Tarifs\_2024-11-18T13\_32\_20.pdf

Suivi des remises YTD N : Graphique linéaire qui affiche le montant de remises YTD N par mois comptable et libellé d'article

Suivi des avoirs YTD N : Graphique à barres verticales qui affiche le montant des avoirs YTD N par mois comptable et libellé d'article

Remises & Avoirs :  Remises\_&\_Avoirs\_2024-11-20T14\_36\_56.pdf

Détails des remises : Tableau croisé dynamique qui affiche le montant de remises YTD N par mois comptable et libellé d'article

Détails des avoirs : Tableau croisé dynamique qui affiche le montant des avoirs YTD N par mois comptable et libellé d'article

### 5.4.3 Intégration des filtres et des contrôles interactifs

Les **filtres** permettent de limiter l'affichage des données dans les visualisations en fonction de critères définis. Ils agissent directement sur les données avant leur affichage et Les **paramètres** permettent aux utilisateurs de personnaliser leur expérience en agissant comme des variables dynamiques. Contrairement aux filtres, les paramètres ne modifient pas directement les données, mais influencent le contenu des visualisations ou des calculs.

Voici une présentation des filtres et des paramètres qu'on a créer :

Contrôles

Libellé société

Tout

Date début

01/01/2023

Date fin

18/11/2024

Contrôles

Code libellé article est égal à	Date début	Date fin	Code libellé client est égal à
<input type="text" value="Tout"/>	<input type="text" value="01/01/2023"/>	<input type="text" value="18/11/2024"/>	<input type="text" value="Text"/>

Contrôles

Code libellé client

Libellé famille

Code libellé article

Date début

Date fin

Tout

Tout

Tout

01/01/2023

18/11/2024

Contrôles

Date début: 01/01/2023 
 Date fin: 18/11/2024 
 Libellé catégorie article est ég...: Tout 
 Code libellé article est égal à: Tout 
 Technicien Farmapro Nom: Tout 
 Technicien Porc nom: Tout 

▼ Paramètres d'exploration

- 📅 DATEDEBUT
- 📅 DATEFIN
- # TOP

Données

Jeu de données

SPICE

cgp\_fct\_vente\_pol...

Article-Polesanté

Clients-Polesante

Dates

Mesures-Ventes

Société-polesanté

Technicien-farmapro

Technicien-porc

Ventes-détails

Vétérinaire-porc

Zone géographique

Marge YTD N

Marge YTD N-1

Quantité YTD N

Quantité YTD N-1

Remises (hors remise RFA)

Variation std Ventes HT N & ...

Ventes HT YTD N

Ventes HT YTD N (hors remises)

Ventes HT YTD N-1

Filtres

AJOUTER

# Code article

Haut - TOP

Cette feuille (personnalisé)

# Code client

Exclure - 30074

Feuille croisée (personna...

Libellé catégorie client

Inclure - tout

Cette représentation visu...

Libellé société

Inclure - tout

Cette feuille

Date facture

Entre - DATEDEBUT et DATEFIN

Feuille croisée (personna...

Données

Jeu de données100%

SPICEcgp\_fct\_vente\_pol...

Article-Polesanté

Clients-Polesante

Dates

Mesures-Ventes

Société-polesanté

Technicien-farmapro

Technicien-porc

Ventes-détails

Vétérinaire-porc

Zone géographique

Marge YTD N

Marge YTD N-1

Quantité YTD N

Quantité YTD N-1

Remises (hors remise RFA)

Variation std Ventes HT N & ...

Ventes HT YTD N

Ventes HT YTD N (hors remises)

Ventes HT YTD N-1

Filtres

+ AJOUTER

# Code client

Exclure - 30074

Feuille croisée (personnalisé)

ET

📅 Date facture

Entre - DATEDEBUT et DATEFIN

Feuille croisée (personnalisé)

ET

🔍 Code libellé article

Inclure - tout

Cette feuille

ET

🔍 Code libellé client

Inclure - tout

Cette feuille

ET

# Code société

Inclure - 9

Cette feuille

ET

= Ventes HT YTD N

Supérieure à - 0

Cette feuille (personnalisé)

ET

= Ventes HT YTD N

Supérieure à - 0

Cette feuille (personnalisé)

ET

= Ventes HT YTD N-1

Supérieure à - 0

Cette feuille (personnalisé)

ET

🔍 Libellé catégorie client

Inclure - Distributeur FR

Cette représentation visuelle uniquement

Données

Jeu de données100%

SPICEcgp\_fct\_vente\_pol...

Article-Polesanté

Clients-Polesante

Dates

Mesures-Ventes

Société-polesanté

Technicien-farmapro

Technicien-porc

Ventes-détails

Vétérinaire-porc

Zone géographique

Marge YTD N

Marge YTD N-1

Quantité YTD N

Quantité YTD N-1

Remises (hors remise RFA)

Variation std Ventes HT N & ...

Ventes HT YTD N

Ventes HT YTD N (hors remises)

Ventes HT YTD N-1

Filtres

+ AJOUTER

Date facture

Entre - DATEDEBUT et DATEFIN

Feuille croisée (personnalisé)

ET

Code libellé client

Inclure - tout

Cette feuille

ET

Libellé famille

Inclure - tout

Cette feuille

ET

Code libellé article

Inclure - tout

Cette feuille

ET

Code société

Inclure - 3

Cette feuille

Données

Jeu de données100%

SPICEcgp\_fct\_vente\_pol...

Article-Polesanté

Clients-Polesante

Dates

Mesures-Ventes

Société-polesanté

Technicien-farmapro

Technicien-porc

Ventes-détails

Vétérinaire-porc

Zone géographique

Marge YTD N

Marge YTD N-1

Quantité YTD N

Quantité YTD N-1

Remises (hors remise RFA)

Variation std Ventes HT N & ...

Ventes HT YTD N

Ventes HT YTD N (hors remises)

Ventes HT YTD N-1

Filtres

+ AJOUTER

Date facture

Entre - DATEDEBUT et DATEFIN

Feuille croisée (personnalisé)

ET

Code société

Inclure - 3

Cette feuille

ET

Code libellé client

Inclure - tout

Cette représentation visuelle uniquement

ET

Libellé famille

Inclure - tout

Cette représentation visuelle uniquement

ET

Code libellé article

Inclure - tout

Cette représentation visuelle uniquement

ET

Code société

Inclure - 3

Cette représentation visuelle uniquement

Données

Jeu de données

100%

SPICE

cgp\_fct\_vente\_pol...

Article-Polesanté

Clients-Polesante

Dates

Mesures-Ventes

Société-polesanté

Technicien-farmapro

Technicien-porc

Ventes-détails

Vétérinaire-porc

Zone géographique

Marge YTD N

Marge YTD N-1

Quantité YTD N

Quantité YTD N-1

Remises (hors remise RFA)

Variation std Ventes HT N & ...

Ventes HT YTD N

Ventes HT YTD N (hors remises)

Ventes HT YTD N-1

Filtres

+ AJOUTER

Date facture

Dernier 6 mois

Cette représentation visuelle uniquement

ET

Ventes HT

Supérieure à - 0

Cette représentation visuelle uniquement

ET

tcf\_lib\_technicien

Inclure - tout

Cette feuille

ET

tcp\_lib\_technicien

Inclure - tout

Cette feuille

ET

Libellé catégorie article

Inclure - tout

Cette feuille

ET

Code libellé article

Inclure - tout

Cette feuille

Les filtres qu'on voit en haut c'est les filtres interactifs qu'on peut modifier par rapport au besoin (si on veut un technicien bien précis ou une date bien précises ou un article ex....)

Les paramètres sont des fonctionnalités QuickSight :

Date début : c'est un paramètre qu'on a créer pour fixé une date relative dans l'année précédente :



en

Modifier le paramètre

×

Utilisez des paramètres pour contrôler dynamiquement les valeurs dans vos champs, vos filtres et votre feuille.

Nom

DATEDEBUT

Type de données (non modifiable une fois créé)

Date-heure

Granularité temporelle

Jour

▼

Date par défaut

Glissante: Au début de l'année précédente, 2023/01/01

☐ Date fixe

☒ Date relative

☐ Aujourd'hui

☐ Hier

☒ Condition de filtre

Plage

Période

Début de

Précédent(e)(s)

Année

Valeur dynamique par défaut

[Définir une valeur dynamique par défaut](#)

Annuler

Mettre à jour

Date fin : c'est un paramètre qu'on a créé pour fixer une date relative qui est aujourd'hui

Avec ces deux paramètres là le TDB va toujours pointer sur Aujourd'hui et une année passée à partir d'aujourd'hui .

Modifier le paramètre

Utilisez des paramètres pour contrôler dynamiquement les valeurs dans vos champs, vos filtres et votre feuille.

Nom

DATEFIN

Type de données (non modifiable une fois créé)

Date-heure

Granularité temporelle

Jour

Date par défaut

Glissante: Aujourd'hui, 2024/11/18

Date fixe

Date relative

Aujourd'hui

Hier

Condition de filtre

Plage

Période

Début de

Ce/Cette/Ces

Jour

Valeur dynamique par défaut

[Définir une valeur dynamique par défaut](#)

Annuler

Mettre à jour

TOP : c'est un paramètres qu'on a créer pour sélectionner le TOP 10 d'articles ou clients par exemple

Modifier le paramètre

Utilisez des paramètres pour contrôler dynamiquement les valeurs dans vos champs, vos filtres et votre feuille.

Nom

TOP

Type de données (non modifiable une fois créé)

Entier

Valeurs (non modifiable une fois créé)

Valeur unique ⓘ

Valeur statique par défaut

10

Valeur dynamique par défaut

[Définir une valeur dynamique par défaut](#)

Paramètres avancés

▼

Annuler

Mettre à jour

## 5.5 Déploiement et partage du tableau de bord 🔗

### 5.5.1 Publication et mise à jour du tableau de bord 🔗

#### Étape 1 : Finalisation du tableau de bord 🔗

Avant la publication, il est crucial de :

- Valider que toutes les visualisations (graphiques, cartes, indicateurs clés de performance) sont correctement configurées.
- Vérifier que les filtres, paramètres et contrôles interactifs fonctionnent comme prévu.
- Effectuer des tests pour s'assurer que les données affichées sont exactes et cohérentes.

#### Étape 2 : Publication dans QuickSight 🔗

1. Accéder à l'analyse dans QuickSight.
2. Cliquer sur l'option "**Publier le tableau de bord**".
3. Attribuer un nom descriptif au tableau de bord pour qu'il soit facilement identifiable par les utilisateurs finaux.
4. Configurer les autorisations d'accès :
  - Définir les utilisateurs ou groupes ayant accès au tableau de bord.
  - Choisir le type d'accès (lecture seule, modification, etc.).

Publier un tableau de bord

Publier un nouveau tableau de bord en tant que

Nouveau tableau de bord 1

Remplacer un tableau de bord existant

CGP Pole Santé

Renommer

TOUTES LES FEUILLES SONT SÉLECTIONNÉES

Data story

Allow sharing data stories

Generative capabilities

Allow executive summary

No topic linked yet. [Link topic](#)

Options de publication avancées

Publier un tableau de bord

Tarifs\_2024-11-18T13\_32\_20.pdf COOPERL\_Techniciens\_2024-11-18T13\_14\_13.pdf COOPERL\_2024-11-18T12\_26\_25.pdf Synthèse\_2024-11-18T14\_02\_24.pdf FARMAPRO\_2024-11-18T14\_02\_39.pdf COOPERL\_2024-11-18T14\_02\_41.pdf

### Étape 3 : Distribution aux utilisateurs finaux

- Partager le tableau de bord via un lien sécurisé généré par QuickSight.
- Les utilisateurs peuvent accéder au tableau de bord directement depuis leur navigateur ou l'application QuickSight sur leurs appareils mobiles.

Partager un tableau de bord

Copier le lien

Copier le code incorporé

Inviter des utilisateurs et des groupes dans le tableau de bord

Rechercher par e-mail d'utilisateur ou d'ID

Activer l'accès pour

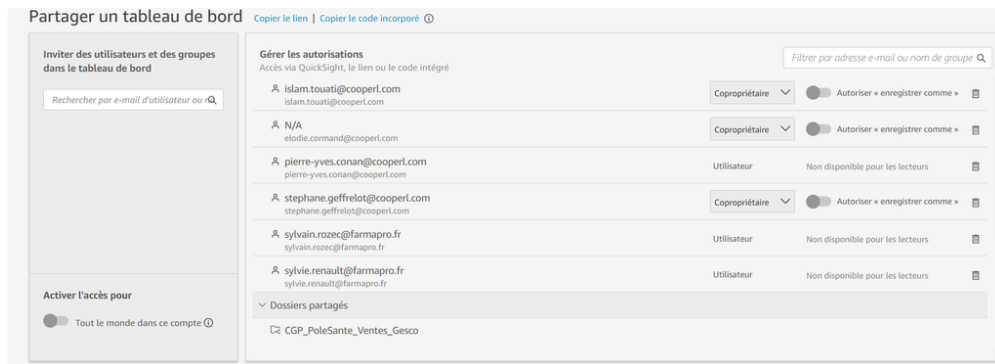
Tout le monde dans ce compte

Gérer les autorisations

Accès via QuickSight, le lien ou le code intégré

Filter par adresse e-mail ou nom de groupe

Nom	Autorisation	Enregistrer en tant qu'analyse
Utilisateurs et groupes		
AWSReservedSSO_AWSAdministratorAccess_5232f35c6c700b3d/rania.zerrouki@coopert.com	Utilisateur	<div><div></div>Autoriser « enregistrer comme »</div>
AWSReservedSSO_AWSAdministratorAccess_5232f35c6c700b3d/sgeffrelot	Copropriétaire	<div><div></div>Autoriser « enregistrer comme »</div>
AWSReservedSSO_ProdBIUCNU_b4967812cf995530/ecormand	Copropriétaire	<div><div></div>Autoriser « enregistrer comme »</div>
AWSReservedSSO_ProdBIUCGrou_069886fdccfd6e03/ecormand	Copropriétaire	<div><div></div>Autoriser « enregistrer comme »</div>
AWSReservedSSO_ProdBIUCGrou_069886fdccfd6e03/sattia	Propriétaire	<div><div></div>Autoriser « enregistrer comme »</div>
benjamin.lesage@farmapro.fr	Utilisateur	Non disponible pour les lecteurs



### 5.5.2 Gestion des permissions et partage avec les utilisateurs

La gestion des permissions et le partage des tableaux de bord dans Amazon QuickSight assurent un accès sécurisé et contrôlé aux données. Les rôles utilisateur (administrateur, auteur, lecteur) permettent de définir des niveaux d'accès adaptés, tandis que des outils comme le **Row-Level Security (RLS)** restreignent les données visibles selon les utilisateurs. Les tableaux de bord peuvent être partagés directement via QuickSight, des liens sécurisés, ou des intégrations avec AWS IAM et Active Directory pour simplifier la gestion à grande échelle. En appliquant des bonnes pratiques, telles que le principe du moindre privilège et des audits réguliers, les organisations garantissent une collaboration efficace et une sécurité optimale pour leurs analyses décisionnelles.

### 5.6 Conclusion

Ce projet a permis de démontrer comment Amazon QuickSight, associé à des outils tels qu'AWS Athena, S3 et Glue, peut être utilisé pour concevoir une solution robuste d'informatique décisionnelle. En partant d'une problématique liée à la gestion et à l'analyse de données volumineuses, nous avons mis en œuvre une architecture cloud performante et évolutive, intégrant des processus ETL optimisés et une visualisation interactive des données.

Grâce à l'utilisation des meilleures pratiques en matière de BI et à l'application de méthodologies comme Kanban, nous avons structuré le projet en étapes claires, de la conception des bases de données à la publication des tableaux de bord. Le déploiement de QuickSight a permis de transformer les données en informations exploitables, offrant ainsi des indicateurs clés de performance (KPI) pertinents pour les décideurs.

En outre, la mise en place de filtres interactifs, de contrôles avancés et d'une gestion rigoureuse des permissions a garanti que les tableaux de bord répondent aux besoins variés des utilisateurs tout en respectant les contraintes de sécurité et de confidentialité des données. Ce projet illustre donc comment une stratégie cloud bien pensée peut améliorer la prise de décision et l'efficacité organisationnelle.

En conclusion, cette expérience a non seulement résolu les enjeux initiaux mais a également démontré le potentiel des technologies cloud pour transformer les données brutes en leviers stratégiques, tout en restant flexible et prêt à évoluer face aux besoins futurs.