



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
DE LA RECHERCHE SCIENTIFIQUE ET DE L'INNOVATION
UNIVERSITÉ SULTAN MOULAY SLIMANE
**ECOLE NATIONALE DES SCIENCES APPLIQUÉS
DE KHOURIBGA**



Mémoire de fin d'études

pour l'obtention du

Diplôme d'Ingénieur d'État

Filière : Informatique et Ingénierie des Données



Présenté par :

Ziyad RAKIB

Participation à la migration d'une architecture monolithique à une architecture microservices

Sous la supervision :

Abdelghani Ghazdali (interne)

Nidal Lamghari (interne)

Bilal Slayki (externe)

Membres du jury :

Nom et prénoms du président	Entité	Président
Nom et prénoms de l'examineur	Entité	Examineur
Nom et prénoms du rapporteur	Entité	Rapporteur
Nom et prénoms du rapporteur	Entité	Rapporteur

Année Académique : 2023

Sommaire

Dédicace	ii
Remerciements	iii
Résumé	iv
Abstract	v
Table des matières	vii
Table des figures	ix
Liste des tableaux	x
Introduction	1
1 Contexte général du projet	2
2 Analyse et Schématisation	13
3 Technologies Utilisées	14
Conclusion	19

Dédicace

“

À ma mère, qui m'a comblée de son soutien et dévouée moi avec un amour inconditionnel. Tu es pour moi un exemple de courage et sacrifice continuuel. Que cet humble travail porte témoignage de mon affection, de mon attachement éternel et qu'il appelle sur moi ta bénédiction continuelle.,

À mon père, aucune dédicace ne peut exprimer l'amour, l'estime, dévouement et le respect que j'ai toujours eu pour vous. Rien dans le monde vaut les efforts faits jour et nuit pour mon éducation et mon bien-être. Ce travail est le fruit de vos sacrifices que vous avez fait pour mon éducation et ma formation,

À mes chers frères, merci pour votre amour, soutien, et encouragements,

À tous mes chers amis, pour le soutien que vous m'avez apporté, je dis Merci encore une fois à tous ceux qui me sont chers, à vous tous

Merci.

”

- Ziyad

Remerciements

Tout d’abord, je remercie le grand Dieu puissant de nous donné la puissance pour continuer et dépasser toutes les difficultés.

J’adresse mes remerciements à, **Monsieur Abdelghani Ghazdali**, chef de filière Informatique et Ingénierie des Données qui a su assurer le bon déroulement des séances d’encadrement du début à la fin.

Je tiens également à adresser mes sincères remerciements à **Professeur Nidal Lamghari**, pour ses précieux conseils, son expertise et sa disponibilité tout au long de cette expérience. Votre encadrement attentionné m’a permis d’approfondir mes connaissances et de progresser dans mes compétences techniques. Je suis reconnaissant d’avoir pu bénéficier de votre expertise et de votre soutien constant.

Au terme de ce travail, je tiens également à exprimer mes sincères remerciements et ma gratitude envers tous ceux qui, par leur enseignement, par leur soutien et leurs conseils, ont contribué au déroulement de ce projet.

Par ailleurs, je tiens à exprimer ma gratitude envers **IZICAP**, où j’ai eu la chance de réaliser mon projet de fin d’études. Je suis reconnaissant envers toute l’équipe de **IZICAP** pour m’avoir accueilli chaleureusement et pour m’avoir permis de mettre en pratique mes connaissances académiques dans un environnement professionnel. Votre expertise, votre mentorat et vos ressources ont été d’une valeur inestimable pour la réalisation de ce projet.

Je tiens à remercier spécialement **M. Bilal SLAYKI**, mon encadrant en entreprise, pour son soutien constant, ses orientations judicieuses et son accompagnement précieux tout au long de ce projet. Votre expérience, vos conseils éclairés et votre bienveillance m’ont inspiré et ont grandement contribué à ma croissance professionnelle. Je vous suis reconnaissant d’avoir partagé votre expertise avec moi et d’avoir été un mentor exceptionnel.

Finalement, mes remerciements s’adressent aussi à l’ensemble du corps professoral et administratif de l’ENSA Khouribga pour l’effort qu’ils fournissent afin de nous garantir une bonne formation et à l’équipe administrative et technique pour tous les services offerts.

Résumé

Le présent rapport de projet de fin d'études (PFE) met en évidence les différentes étapes de notre projet de transformation de l'architecture monolithique en Groovy vers une architecture basée sur des microservices en Java, tout en effectuant une transition de AngularJS vers React et en adoptant une approche de microfrontend. Nous avons entrepris cette démarche afin d'améliorer la flexibilité, la maintenabilité, les performances, la cohérence des données et la modularité de notre application.

Initialement, notre code monolithique était développé en Groovy. La première étape de notre projet a consisté à analyser cette architecture monolithique et à identifier les composants qui pourraient être isolés en microservices indépendants. Parallèlement, nous avons évalué les avantages de migrer de AngularJS, un framework JavaScript obsolète, vers React, un framework plus moderne et performant.

Nous avons procédé à la conception et à la mise en œuvre de ces microservices en utilisant Spring-Boot, en veillant à découpler les fonctionnalités et à assurer une communication efficace entre les services. Dans le même temps, nous avons migré progressivement notre code AngularJS vers React, en réécrivant les fonctionnalités existantes avec les meilleures pratiques de développement React.

Pour assurer la cohérence des données entre les microservices, nous avons mis en place Delta Lake, une technologie de gestion de données incrémentielle. Delta Lake a permis de garantir la fiabilité des données et de simplifier les opérations de lecture et d'écriture sécurisées entre les services.

Les connecteurs en Java ont facilité la communication entre le frontend basé sur React et Delta Lake, assurant ainsi une connexion robuste et sécurisée. Les connecteurs ont également permis des opérations de lecture et d'écriture efficaces, en garantissant l'intégrité et la cohérence des données.

En parallèle, nous avons également adopté une approche de microfrontend pour notre architecture frontend. Cela nous a permis de découpler les fonctionnalités du frontend en modules autonomes, offrant ainsi une plus grande flexibilité et la possibilité de les développer et de les déployer indépendamment.

Enfin, pour faciliter le déploiement et la gestion de notre architecture basée sur des microservices, React et le microfrontend, nous avons adopté l'utilisation de Kubernetes et Docker. Ces outils ont permis l'orchestration et la mise à l'échelle efficaces des services, tout en simplifiant le déploiement dans différents environnements.

Mots-clés: Monolithique, Groovy, architecture microservices, AngularJS, React, Delta Lake, Microfrontend, Kubernetes, Docker.

Abstract

This final year project report highlights the different stages of our project to transform a Groovy-based monolithic architecture into a Java-based microservices architecture, while transitioning from AngularJS to React and adopting a microfrontend approach. Our goal was to enhance the flexibility, maintainability, performance, data consistency, and modularity of our application.

Initially, our codebase was developed as a monolith using Groovy. The first step of our project involved analyzing the monolithic architecture and identifying components that could be isolated into independent microservices. Concurrently, we evaluated the benefits of migrating from the outdated AngularJS framework to the more modern and performant React framework.

We proceeded with the design and implementation of these microservices using Spring Boot, ensuring loose coupling between functionalities and effective communication among services. Simultaneously, we gradually migrated our AngularJS code to React, rewriting existing features according to React's best development practices.

To ensure data consistency among microservices, we implemented Delta Lake, an incremental data management technology. Delta Lake guaranteed reliable data and simplified secure read and write operations between services. Java connectors facilitated communication between the React-based frontend and Delta Lake, ensuring robust and secure connectivity. These connectors also enabled efficient read and write operations, ensuring data integrity and consistency.

In parallel, we adopted a microfrontend approach for our frontend architecture, decoupling frontend features into autonomous modules. This provided greater flexibility and the ability to develop and deploy modules independently.

Finally, to facilitate deployment and management of our microservices, React, and microfrontend architecture, we adopted Kubernetes and Docker. These tools enabled efficient orchestration, scaling of services, and simplified deployment across different environments.

In conclusion, our project resulted in significant improvements in flexibility, maintainability, performance, data consistency, and modularity of our application. The transition from a Groovy-based monolithic architecture to a Java-based microservices architecture, combined with the migration from AngularJS to React, the adoption of Delta Lake, Java connectors, microfrontend, Kubernetes, and Docker, collectively optimized our system.

Mots-clés : Monolithique, Groovy, Microservices, AngularJS, React, Delta Lake, Microfrontend, Kubernetes, Docker

ملخص

يقدم هذا التقرير تحليلاً متعمقاً للمشروع المعقد الذي تم إجراؤه لتوسيع نطاق مصدر بيانات الشركة. يتطلب المشروع إزالة MariaDB وتنفيذ Delta Lake ، وهو حل تخزين ومعالجة بيانات عالي الأداء. بالإضافة إلى ذلك ، تم تقسيم المونوليث الحالي إلى خدمات مصغرة باستخدام Spring Boot كواجهة خلفية مع موصل مناسب لـ Delta

Lake ، وواجهة ReactJS كواجهة أمامية. قدم المشروع العديد من التحديات التي تطلبت تطبيق التقنيات والاستراتيجيات المتقدمة. وشمل ذلك ترحيل البيانات ، وضمان اتساق البيانات ودقتها ، وإدارة تعقيدات الأنظمة الموزعة ، ودمج التقنيات والخدمات المختلفة. يحدد التقرير الحلول المختلفة التي تم تطويرها لمواجهة هذه التحديات ، بما في ذلك استخدام خوارزميات معالجة البيانات المتقدمة ، وبناء الحوسبة الموزعة ، والحاويات. على الرغم من التحديات التي واجهها ، لا يزال المشروع قيد التنفيذ. يقدم التقرير رؤى حول الجهود الجارية لتحسين قابلية المشروع للتوسع والأداء والوظائف العامة

كلمات مفتاحية : المشروع ، مصدر البيانات ، Delta Lake ، المونوليث ، الميكروسيرفس ، Springboot ، ReactJS ، الميكروواجهات ، الخلفية ، الواجهة الأمامية ، البيانات ، القابلية للتوسع ، الأداء. Delta Lake ، Springboot ، ReactJS ، الميكروواجهات.

Table des matières

Dédicace	ii
Remerciements	iii
Résumé	iv
Abstract	v
Table des matières	vii
Table des figures	ix
Liste des tableaux	x
Introduction	1
1 Contexte général du projet	2
Introduction	2
1.1 Introduction	2
1.2 Présentation de l'organisme d'accueil	2
1.2.1 IZICAP	2
1.2.2 Partenariats stratégiques	3
1.3 Organigramme	4
1.4 Cadre du projet	5
1.5 L'architecture du système -Smart Data-	6
1.6 Problématique et besoin fonctionnel	8
1.7 Objectives du stages	9
1.8 Processus de réalisation du projet	10
1.9 Planification du projet	11
Conclusion	12
2 Analyse et Schématisation	13
Introduction	13
Conclusion	13
3 Technologies Utilisées	14
Introduction	14

3.1	Delta Lake	14
3.2	Principaux avantages et caractéristiques de Delta Lake	15
3.3	Trino	17
Conclusion		19

Table des figures

1.1	Partenariats stratégiques	4
1.2	Organigramme de Izicap	4
1.3	Smart data Izicap	6
1.4	Architecture du système actuelle	7
1.5	Méthodologie Scrum	11
1.6	Diagramme de Gantt	11
3.1	Architecture multi-sauts de Delta Lake	15
3.2	Vue d'ensemble de l'architecture Trino avec le coordinateur et les workers	17

Liste des tableaux

1 [List of Abbreviations](#) xi

API	Application programming interface
VSC	Visual Studio Code
DL	Delta Lake
REST	RepresEntational State Transfer
AWS	Amazon Web Services
GCP	Google Cloud Platform
AZR	Microsoft Azure
S3	Simple Storage Service
IAM	Identity and Access Management
MinIO	Minimal Object Storage
SQL	Structured Query Language
DB	Database
ACID	Atomicity, Consistency, Isolation, Durability
OLTP	Online Transaction Processing
OLAP	Online Analytical Processing
ReactJS	React JavaScript
SPA	Single Page Application
JS	JavaScript
CSS	Cascading Style Sheets
HTML	Hypertext Markup Language
UI	User Interface
API	Application Programming Interface
Spark	Apache Spark
Hadoop	Apache Hadoop
HDFS	Hadoop Distributed File System
YARN	Yet Another Resource Negotiator
MapReduce	MapReduce Programming Model
Metadata	Data about Data
ETL	Extract, Transform, Load
ELT	Extract, Load, Transform
CRM	Customer Relationship Management
RDBMS	Relational Database Management System
SPI	Server Provider Interface

TABLE 1 : List of Abbreviations

Introduction Générale

Dans un monde numérique en constante évolution, les entreprises sont confrontées à des défis majeurs pour maintenir leurs applications à la pointe de la technologie et répondre aux attentes croissantes des utilisateurs. L'architecture monolithique présente des limitations en termes de flexibilité, de maintenabilité, de performances et de modularité, ce qui pousse de nombreuses organisations à adopter des architectures basées sur des microservices.

Ce rapport met en évidence la démarche entreprise pour transformer une architecture monolithique en Groovy vers une architecture basée sur des microservices en utilisant Java, tout en migrant de AngularJS vers React et en adoptant une approche de microfrontend. L'objectif principal de cette transformation était d'améliorer la flexibilité, la maintenabilité, les performances, la cohérence des données et la modularité de l'application.

La première étape a été d'analyser l'architecture monolithique existante et d'identifier les composants pouvant être isolés en microservices indépendants. En parallèle, une évaluation des avantages de migrer de AngularJS vers React a été réalisée.

La conception et la mise en œuvre des microservices ont été réalisées en utilisant Spring-Boot, en veillant à découpler les fonctionnalités et à assurer une communication efficace entre les services. Le code AngularJS a été progressivement migré vers React en utilisant les meilleures pratiques de développement React.

Pour garantir la cohérence des données entre les microservices, la technologie de gestion de données incrémentielle Delta Lake a été utilisée. Des connecteurs en Java ont facilité la communication entre le frontend basé sur React et Delta Lake, assurant une connexion robuste et sécurisée.

En parallèle, une approche de microfrontend a été adoptée pour le frontend, permettant de découpler les fonctionnalités en modules autonomes.

Ce rapport détaillera chaque étape de la transformation, mettant l'accent sur les décisions prises, les défis rencontrés et les résultats obtenus. Il soulignera les avantages de l'architecture basée sur des microservices, l'utilisation de React et du microfrontend, ainsi que l'utilisation de Kubernetes et Docker. Cette étude de cas offre une perspective précieuse sur la modernisation des applications et les meilleures pratiques de développement dans un environnement en constante évolution.

Contexte général du projet

1.1 Introduction

L'objectif de cette partie est de présenter le contexte général du projet, en commençant par l'organisme d'accueil, IZICAP, où j'ai effectué un stage de 5 mois. Nous décrirons leurs activités, leur stratégie, leurs valeurs, ainsi que l'organigramme et les profils impliqués dans mon projet. Ensuite, nous aborderons la demande de travaux qui exprime les besoins fonctionnels sur lesquels je me suis basé pour concevoir et réaliser la solution, ainsi que la méthodologie suivie pour atteindre les objectifs fixés.

1.2 Présentation de l'organisme d'accueil

1.2.1 IZICAP

IZICAP est une société unique et novatrice axée sur l'analyse de données, la fidélisation et le marketing numérique pour les petites et moyennes entreprises (PME). Leur solution conviviale et facile à installer repose sur le modèle SaaS, permettant d'exploiter les données transactionnelles des cartes de paiement des clients.

En collaborant avec les banques et les acquéreurs, IZICAP offre une proposition de valeur renforcée et mettons à leur disposition un outil puissant de marketing numérique destiné aux PME. Leur mission est d'accompagner les commerçants au quotidien pour accélérer le développement de leur entreprise, répondant ainsi à leurs besoins spécifiques.

L'histoire d'IZICAP remonte à 2013, lorsque elle a vu le jour dans un paysage des services financiers en constante évolution et de plus en plus complexe. Depuis lors, les modèles bancaires et de paiement traditionnels sont continuellement menacés par de nouvelles entreprises challengers plus innovantes. Une concurrence agressive signifie que la fidélité des commerçants ne peut plus être considérée comme acquise, et souvent les acquéreurs et les fournisseurs de services de paiement sont contraints de réduire leurs frais.

Après s'être solidement implantée en France grâce à des partenariats avec le Groupe BPCE et le Crédit Agricole, Izicap a établi un partenariat avec Nexi, le principal acquéreur et fintech en Italie, et a rejoint le programme StartPath de Mastercard dans le but d'étendre considérablement sa portée à l'échelle mondiale. Cette expansion stratégique renforce Leur position sur le marché et confirme Leur engagement à offrir leur services à un public plus large.

Avec Izicap, les PME ont accès à une solution innovante qui tire parti de l'analyse de données pour optimiser leur stratégie de fidélisation et de marketing. Grâce à leur expertise et à leur présence internationale croissante, IZICAP est déterminée à soutenir la croissance des entreprises et à les aider à prospérer dans un environnement commercial en constante évolution.

1.2.2 Partenariats stratégiques

IZICAP a établi des partenariats stratégiques avec plusieurs acteurs clés de l'industrie financière. Parmi ses partenaires notables figurent :

- **Groupe BPCE** : Izicap a collaboré avec le Groupe BPCE, l'un des principaux groupes bancaires en France, pour fournir sa solution de fidélisation et de marketing numérique aux commerçants.
- **Crédit Agricole** : Izicap a établi un partenariat avec Crédit Agricole, l'une des plus grandes banques françaises, afin d'offrir ses services aux commerçants de leur réseau.
- **Nexi** : Izicap a noué un partenariat avec Nexi, le principal acquéreur et fintech en Italie. Cette collaboration permet à Izicap d'étendre sa présence sur le marché italien et de bénéficier de l'expertise de Nexi dans le domaine des services de paiement.
- **Mastercard StartPath** : Izicap a rejoint le programme StartPath de Mastercard, qui soutient les start-ups et les entreprises innovantes dans le domaine des technologies financières. Cette collaboration offre à Izicap l'opportunité d'accroître sa visibilité et son expansion internationale.
- **BBVA** : Izicap a également établi un partenariat avec BBVA, l'une des principales banques en Espagne et en Amérique latine. Cette collaboration avec BBVA renforce la présence d'Izicap sur le marché espagnol et permet à la société d'offrir ses solutions de fidélisation et de marketing numérique aux commerçants affiliés à BBVA. Le partenariat avec BBVA témoigne de l'engagement d'Izicap à étendre sa portée et à travailler avec des acteurs majeurs de l'industrie financière pour soutenir les petites et moyennes entreprises dans leur croissance.
- **Ingenico** : Izicap a établi un partenariat avec Ingenico, l'un des principaux fournisseurs de solutions de paiement et de services marchands au niveau mondial. Grâce à ce partenariat, Izicap est en mesure d'intégrer sa solution de fidélisation et de marketing numérique aux terminaux de paiement d'Ingenico, offrant ainsi une solution complète aux commerçants utilisant les services d'Ingenico. Cette collaboration renforce la position d'Izicap en tant que fournisseur de solutions de fidélisation et de marketing innovantes, en s'appuyant sur l'expertise et la portée mondiale d'Ingenico dans le secteur des paiements.



FIGURE 1.1 : Partenariats stratégiques

1.3 Organigramme

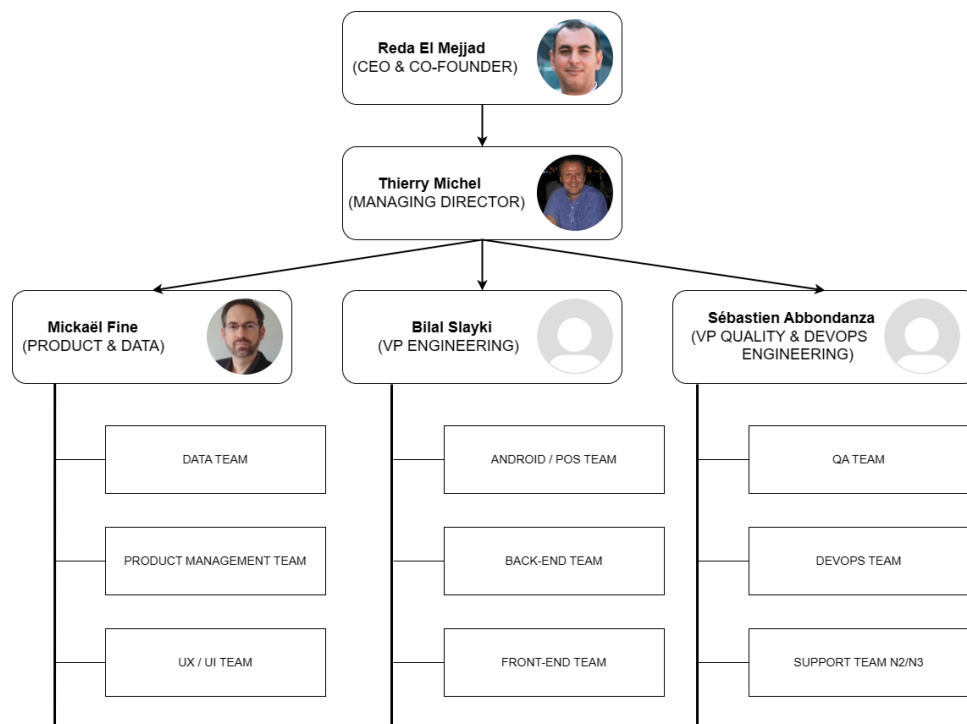


FIGURE 1.2 : Organigramme de Izicap

Au sein de l'équipe de projet, j'ai eu l'opportunité de travailler de manière polyvalente dans différents domaines. J'ai contribué activement au développement du backend en participant à la concep-

tion et à l'implémentation des microservices. J'ai également collaboré avec l'équipe frontend en apportant mon expertise pour l'intégration des microfrontends et l'amélioration de l'expérience utilisateur. De plus, j'ai été impliqué dans la gestion des données en participant à la mise en place de Delta Lake et en contribuant à l'optimisation des flux de données. Cette expérience diversifiée m'a permis d'acquérir une vision globale du projet et de collaborer efficacement avec différentes équipes pour atteindre les objectifs fixés.

1.4 Cadre du projet

Le projet consiste à refondre l'application 'Smart data' qui fait référence au processus d'extraction d'informations précieuses à partir de grandes quantités de données afin de prendre des décisions commerciales éclairées.

La smart data nous permet de :

1. **Collecte des données** : Izicap aide les entreprises à collecter des données clients à partir de différents points de contact tels que les systèmes de point de vente, les programmes de fidélité, les interactions en ligne, etc. Assurez-vous d'intégrer leurs outils de collecte de données dans vos systèmes ou processus existants.
2. **Analyse des données** : La solution de smart data d'Izicap vous permet d'analyser les données collectées pour découvrir des tendances, des modèles et des comportements clients. Utilisez leurs outils d'analyse et leurs algorithmes pour obtenir des informations exploitables à partir des données.
3. **Segmentation des clients** : Segmentez votre base de clients en fonction de leurs préférences, de leur historique d'achat, de leurs caractéristiques démographiques ou d'autres critères pertinents. La solution de smart data d'Izicap peut vous aider à identifier différents segments de clients et à créer des stratégies marketing personnalisées pour chaque segment.
4. **Marketing personnalisé** : Exploitez les informations tirées de la solution de smart data d'Izicap pour créer des campagnes marketing ciblées. Envoyez des offres personnalisées, des promotions ou des recommandations à des segments spécifiques de clients, augmentant ainsi les chances de conversion et de satisfaction client.
5. **Fidélisation de la clientèle** : Utilisez la solution de smart data d'Izicap pour identifier les clients qui risquent de résilier leur abonnement ou de ne plus acheter chez vous. Développez des stratégies de fidélisation en leur offrant des incitations, des programmes de fidélité ou des communications personnalisées afin de les maintenir engagés et fidèles à votre marque.
6. **Suivi des performances** : Surveillez régulièrement les performances de vos campagnes marketing et de vos efforts d'engagement client. La solution d'Izicap peut vous fournir des métriques et des rapports pour évaluer l'efficacité de vos stratégies et apporter des ajustements basés sur les données lorsque nécessaire.

7. **Amélioration continue** : Les solutions de smart data sont les plus efficaces lorsqu'elles sont utilisées de manière itérative. Analysez régulièrement les données, adaptez vos stratégies et peaufinez votre approche en fonction de nouvelles informations et de l'évolution du comportement client.

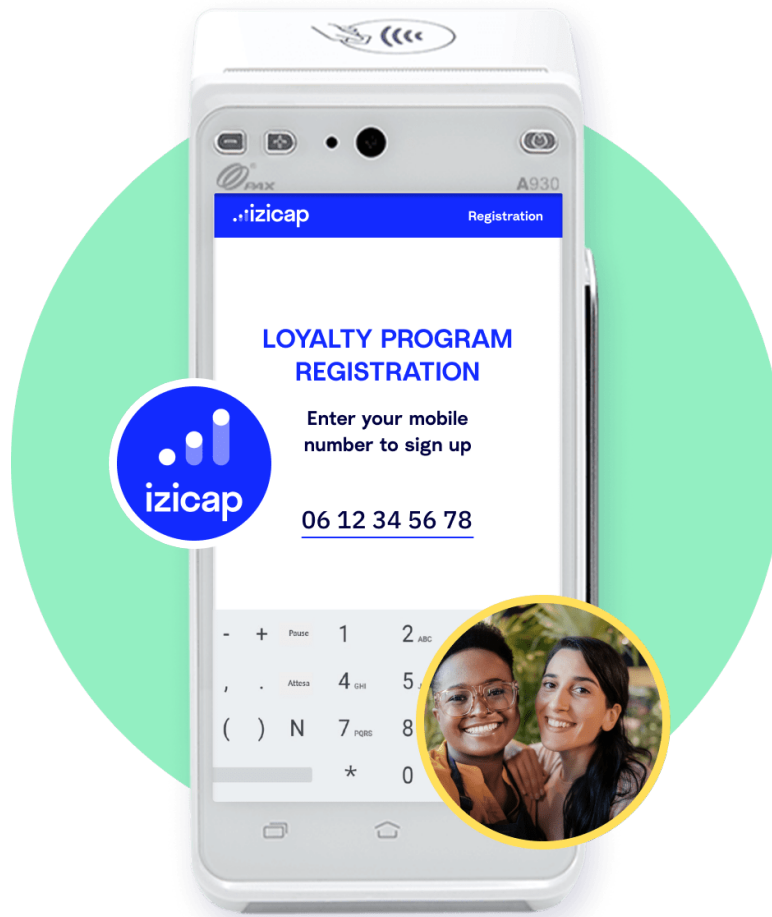


FIGURE 1.3 : Smart data Izicap

1.5 L'architecture du système -Smart Data-

La figure ci-dessous visualise la structure et les composants clés de notre système, ainsi que les interactions entre eux, elle constitue ainsi le fondement de notre système et joue un rôle essentiel dans la fourniture de fonctionnalités et de services à nos utilisateurs.

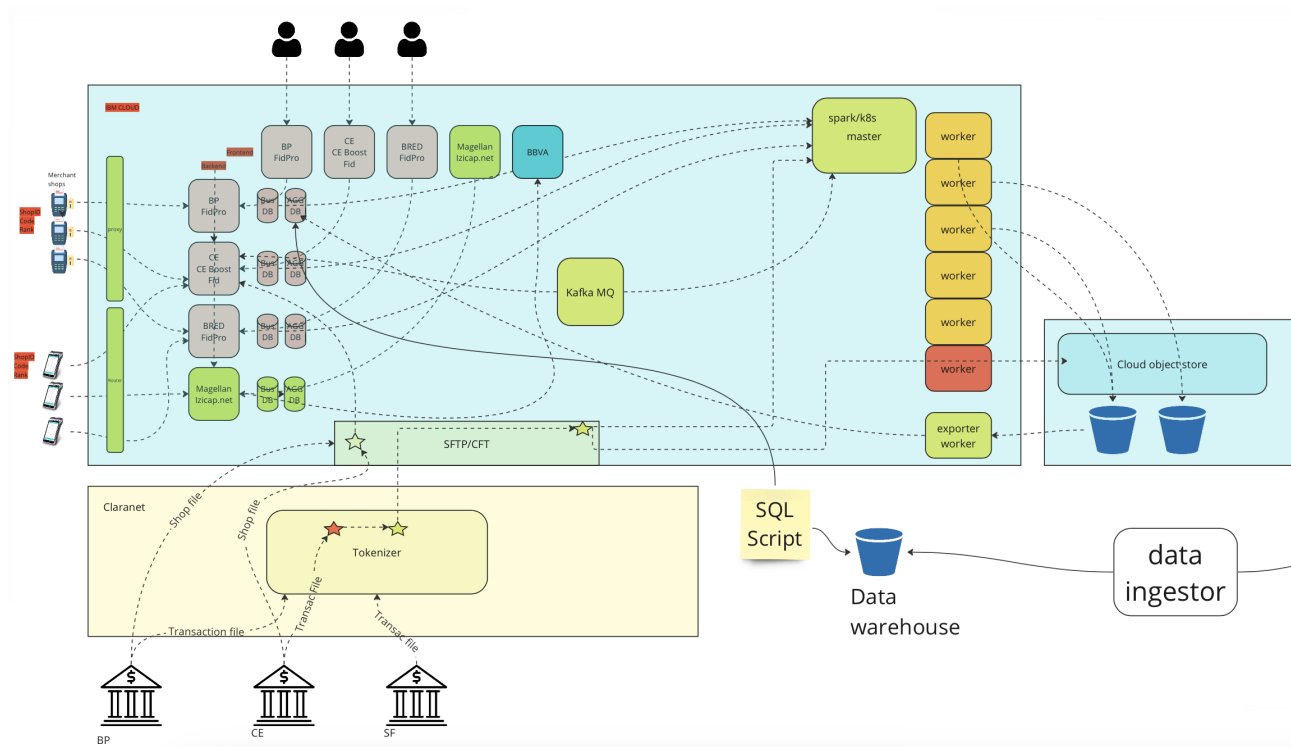


FIGURE 1.4 : Architecture du système actuelle

Ce schéma représente le fonctionnement de l'architecture monolithique actuelle qui présente plusieurs inconvénients qui peuvent entraver l'efficacité, la flexibilité et la maintenabilité du système. Voici une description détaillée des principaux inconvénients :

1. **Complexité et dépendances** : L'architecture monolithique implique que tous les modules et fonctionnalités du système sont regroupés en un seul bloc. Cela crée une forte interdépendance entre les différents composants, rendant la compréhension, la gestion et les mises à jour complexes. Les modifications apportées à une partie du système peuvent avoir des répercussions sur d'autres parties, ce qui rend les tests, les déploiements et les corrections d'erreurs plus difficiles.
2. **Scalabilité limitée** : L'architecture monolithique a souvent des difficultés à s'adapter à une augmentation de la charge ou à une demande croissante. Étant donné que tous les composants sont regroupés dans un seul monolithe, il est difficile de faire évoluer sélectivement une partie spécifique du système. Cela peut entraîner des goulots d'étranglement, des performances réduites et une mauvaise répartition des ressources lorsqu'il s'agit de traiter des volumes de données importants ou de gérer une augmentation du nombre d'utilisateurs.
3. **Déploiements complexes et risques élevés** : Avec une architecture monolithique, les déploiements nécessitent la mise à jour de l'ensemble du système, même pour des modifications mineures. Cela augmente les risques d'erreurs et de régressions, car une seule erreur peut entraîner l'indisponibilité du système tout entier. De plus, les déploiements doivent être soigneusement planifiés et coordonnés, ce qui peut entraîner des temps d'arrêt plus longs et des interruptions de service pour les utilisateurs.

4. **Difficulté de choix technologiques** : Dans une architecture monolithique, les technologies utilisées sont souvent liées et intégrées de manière étroite. Cela peut rendre difficile l'adoption de nouvelles technologies ou l'intégration de composants spécialisés. Les mises à niveau de versions ou les ajouts de nouvelles fonctionnalités peuvent être limités par les choix technologiques initiaux, ce qui peut entraver l'innovation et l'adaptation aux évolutions du marché.
5. **Cohésion et responsabilités** : L'architecture monolithique ne facilite pas une séparation claire des responsabilités et des fonctionnalités. Les différents modules du système sont souvent intimement liés et peuvent partager des fonctionnalités communes. Cela rend difficile l'isolation des problèmes, la maintenance spécifique des modules et la réutilisation de code spécifique à un domaine.

1.6 Problématique et besoin fonctionnel

La problématique qui se pose réside dans l'architecture monolithique actuellement utilisée, laquelle présente des limitations en termes de scalabilité, de flexibilité et de maintenabilité. Les contraintes imposées par cette architecture rendent complexe le déploiement de nouvelles fonctionnalités et entraînent des perturbations potentielles dans l'ensemble du système. De plus, l'utilisation d'une base de données relationnelle traditionnelle, telle que MariaDB, se révèle insuffisante pour traiter efficacement de grands volumes de données transactionnelles, il convient de souligner que l'application Smart Data existe depuis maintenant 10 ans et a été développée en utilisant le langage Groovy. Cette longue période d'existence témoigne de la stabilité et de la maturité de notre système. Cependant, l'utilisation du langage Groovy peut présenter certaines contraintes en termes de maintenabilité et d'évolutivité, notamment lorsqu'il s'agit d'intégrer de nouvelles technologies et de gérer des architectures plus modernes.

Besoin fonctionnel :

Afin de répondre à ces problématiques, différents besoins fonctionnels ont été identifiés :

- **Scalabilité** : Il est nécessaire de disposer d'une architecture qui puisse s'adapter facilement à une croissance du nombre d'utilisateurs, des transactions et des données. Il est essentiel que notre système puisse évoluer harmonieusement et maintenir ses performances, même face à une augmentation significative de la charge de travail.
- **Flexibilité** : Nous devons être en mesure d'introduire de nouvelles fonctionnalités de manière indépendante et de les déployer sans perturber l'ensemble du système. Une approche basée sur des microservices et des microfrontends nous permettra d'atteindre cette flexibilité, en facilitant le développement, les tests et le déploiement isolé de chaque composant.
- **Performances améliorées** : Il est primordial de disposer d'une infrastructure de données capable de gérer efficacement d'importants volumes de données transactionnelles. En remplaçant MariaDB par Delta Lake, nous pourrions tirer parti de fonctionnalités avancées telles que

la gestion des transactions ACID et la compatibilité avec des outils d'analyse performants, ce qui améliorera sensiblement les performances et la fiabilité de notre système.

- **Séparation des responsabilités** : Nous visons une meilleure séparation des responsabilités entre les différents composants de notre système. Les microservices nous permettront de découpler les fonctionnalités, de les attribuer à des équipes spécifiques et de favoriser une gestion du code plus efficace, une maintenance simplifiée et une évolutivité accrue.

1.7 Objectives du stages

Le projet s'inscrit dans le cadre du renouvellement des fonctionnalités de l'application Smart Data, visant à améliorer sa scalabilité, sa flexibilité et ses performances. Dans ce contexte, les objectifs du stage sont les suivants :

1. Étudier et analyser l'architecture actuelle de l'application SMART DATA afin de comprendre les contraintes et les limitations qui entravent sa croissance et son évolution.
2. Proposer et concevoir une stratégie de migration de l'architecture monolithique vers une architecture basée sur des microservices et des microfrontends, permettant ainsi une meilleure modularité et une plus grande flexibilité dans le développement et le déploiement des fonctionnalités.
3. Mettre en œuvre multiples microservices avec Springboot dans le cadre de la nouvelle architecture, en utilisant une technologie appropriée et en veillant à son intégration harmonieuse avec les autres composants du système.
4. Évaluer les avantages et les implications de l'utilisation de Delta Lake en remplacement de la base de données MariaDB, en mettant l'accent sur les performances, la gestion des transactions et l'intégration avec les outils d'analyse.
5. Intégrer Trino dans l'architecture pour permettre l'exécution de requêtes SQL complexes et optimiser le traitement des données, en assurant une collaboration efficace avec l'équipe de développement.
6. Mettre en place une infrastructure de déploiement et de gestion des microservices, en utilisant des outils tels que Kubernetes, Docker et Jenkins, pour faciliter le déploiement, la mise à l'échelle et la gestion des composants.
7. Concevoir et mettre en œuvre des tests unitaires et des tests d'intégration pour assurer la qualité et la fiabilité des nouveaux composants développés dans le cadre de l'architecture basée sur les microservices.
8. Documenter de manière approfondie le processus de migration et les choix technologiques effectués, fournissant des instructions claires pour la maintenance future de l'architecture et la gestion des mises à jour.

9. Collaborer étroitement avec l'équipe existante pour faciliter la transition vers la nouvelle architecture, en offrant un soutien technique, des conseils et des formations sur les nouvelles technologies utilisées.

1.8 Processus de réalisation du projet

Le processus de réalisation du projet s'est déroulé en plusieurs itérations appelées 'sprints', d'une durée généralement fixe de deux semaines. Chaque sprint était axé sur la livraison d'incrément fonctionnels de l'application, permettant ainsi d'obtenir rapidement des résultats tangibles.

Voici les étapes clés du processus de réalisation du projet, basé sur la méthodologie Scrum :

1. **Définition du backlog du produit** : En collaboration avec les parties prenantes et l'équipe de développement, nous avons identifié et priorisé les fonctionnalités à développer et à intégrer dans l'architecture basée sur les microservices.
2. **Planification du sprint** : Au début de chaque sprint, nous avons organisé une réunion de planification pour définir les objectifs spécifiques du sprint, sélectionner les tâches à réaliser et estimer les efforts nécessaires.
3. **Développement itératif** : L'équipe de développement a travaillé de manière itérative sur les tâches assignées, en se concentrant sur la réalisation des fonctionnalités identifiées pour le sprint en cours.
4. **Réunions quotidiennes de stand up** : Chaque jour, l'équipe s'est réunie pour une brève réunion de stand-up afin de partager les progrès, les obstacles éventuels et coordonner les activités.
5. **Revue de sprint** : À la fin de chaque sprint, nous avons organisé une revue de sprint pour présenter les fonctionnalités développées et obtenir des retours des parties prenantes. Cela nous a permis de valider les résultats obtenus et de planifier les prochaines étapes.
6. **Rétrospective de sprint** : Après la revue de sprint, nous avons mené une rétrospective pour évaluer le déroulement du sprint, identifier les points forts et les points à améliorer, et ajuster notre approche en conséquence.
7. **Itérations suivantes** : Le processus de planification, de développement itératif, de revue de sprint et de rétrospective s'est répété pour chaque sprint suivant, permettant ainsi une progression incrémentale vers les objectifs du projet.

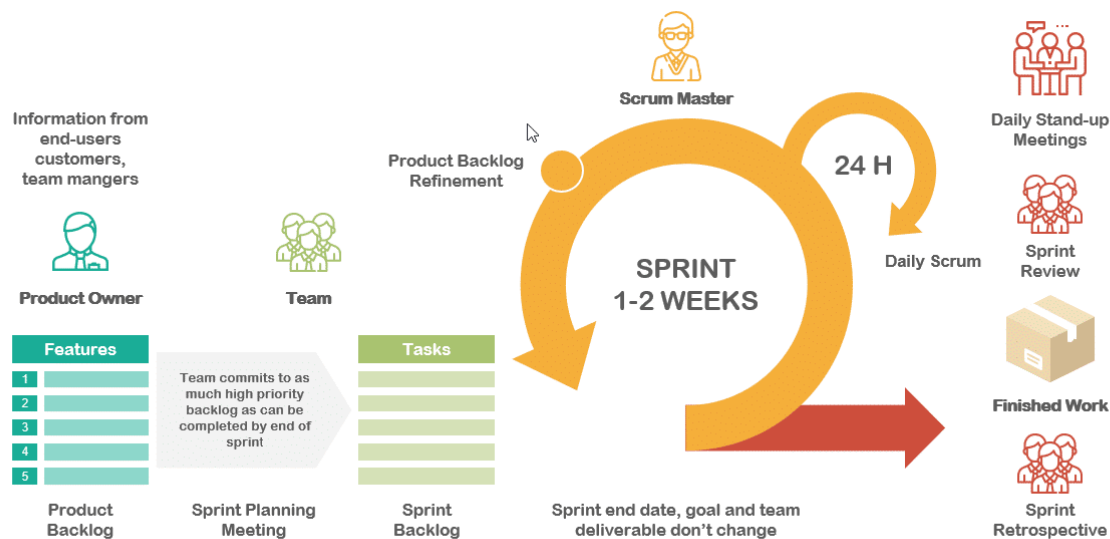


FIGURE 1.5 : Méthodologie Scrum

1.9 Planification du projet

Pour la planification du projet, nous avons utilisé la méthode de modélisation du réseau de dépendance entre les tâches. Cette approche nous a permis de décomposer le travail en différentes tâches structurées et d'établir des relations de dépendance entre elles.

Nous avons également utilisé la technique du diagramme de Gantt pour représenter graphiquement les tâches et les ressources du projet dans le temps. En ligne, nous avons listé les différentes tâches, et en colonne, nous avons défini les jours, les semaines ou les mois. Chaque tâche a été représentée par une barre dont la longueur est proportionnelle à la durée estimée de cette tâche.

Le diagramme de Gantt nous a permis de visualiser la répartition des tâches, leur durée et leur succession. Certaines tâches se sont réalisées en séquence, tandis que d'autres ont pu être réalisées en parallèle, de manière partielle ou totale. Cette représentation claire et visuelle nous a aidés à planifier le projet en déterminant et en organisant les différentes tâches de manière à assurer une gestion efficace du projet.

La figure suivante présente le diagramme de Gantt détaillant la planification du projet, avec les tâches ordonnées dans le temps et leur durée respective. Cela nous a permis de suivre l'avancement du projet, d'identifier les éventuels retards et de prendre les mesures appropriées pour les résoudre.

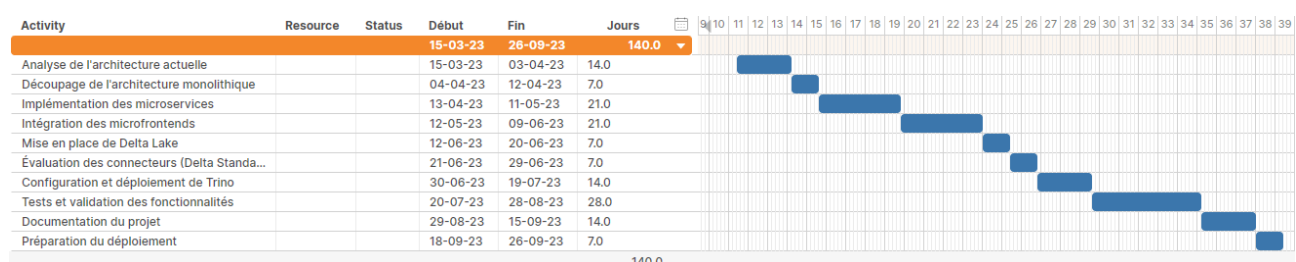


FIGURE 1.6 : Diagramme de Gantt

Conclusion

En conclusion de cette section, nous avons présenté la planification du projet de renouvellement des fonctionnalités de l'application SMART DATA. Nous avons élaboré un planning basé sur la modélisation du réseau de dépendance entre les tâches, en utilisant la technique du diagramme de Gantt. Ce diagramme nous a permis de visualiser les différentes activités, leur durée estimée et les ressources nécessaires.

Au cours de la planification, nous avons identifié les principales activités du projet, telles que l'analyse de l'architecture actuelle, le découpage de l'architecture monolithique, l'implémentation des microservices, l'intégration des microfrontends, la mise en place de Delta Lake, l'évaluation des connecteurs, la configuration et le déploiement de Trino, les tests et la validation des fonctionnalités, la documentation du projet, la préparation du déploiement, ainsi que la formation et la sensibilisation des utilisateurs.

Nous avons également pris en compte la durée totale du projet, qui est estimée à 5 mois, et avons ajusté les durées des activités en conséquence. Cela nous a permis d'avoir une vision plus précise de la planification temporelle du projet.

Analyse et Schématisation

Introduction

Dans ce chapitre, nous explorerons la schématisation détaillée du projet visant à remplacer notre architecture monolithique par une approche basée sur des microservices et des micro-frontends, tout en migrant notre système de gestion de base de données de MariaDB vers Delta Lake. Cette transition marque une étape cruciale dans notre parcours d'évolution technologique, offrant des avantages significatifs en termes de flexibilité, d'évolutivité et de maintenabilité.

Conclusion

Delta Lake est un outil important pour le traitement du Big Data, fournissant une gestion fiable des données et garantissant l'intégrité des données à grande échelle. Ses transactions ACID, l'application des schémas et les fonctionnalités de gestion des versions des données en font un choix populaire pour les entreprises qui doivent traiter de grandes quantités de données avec une grande précision et fiabilité.

En utilisant Delta Lake, les ingénieurs de données et les scientifiques des données peuvent facilement gérer la qualité des données, suivre la lignée des données et collaborer sur des projets d'analyse de données. Grâce à son intégration transparente avec d'autres outils de Big Data, Delta Lake fournit une solution puissante pour le traitement du Big Data qui peut aider les entreprises à mieux comprendre leurs données plus rapidement et plus efficacement.

Technologies Utilisées

Introduction

Dans ce chapitre, nous examinerons en détail les technologies clés qui sont utilisées dans notre solution pour fournir des fonctionnalités avancées et répondre aux besoins spécifiques de notre projet. Les trois principales technologies que nous aborderons sont Delta Lake, Trino et les Microfrontends.

3.1 Delta Lake

Delta Lake est une technologie de gestion des données qui permet de stocker, gérer et analyser des volumes massifs de données de manière efficace et fiable. Il repose sur une architecture basée sur des fichiers parquet et offre des fonctionnalités avancées telles que la gestion des transactions ACID (Atomicité, Cohérence, Isolation, Durabilité) et la compatibilité avec des outils d'analyse populaires. Delta Lake garantit également l'intégrité des données, la cohérence des requêtes et la prise en charge de la réplication et de la récupération en cas de défaillance.

Le concept de 'lakehouse' est rendu possible grâce à Delta Lake. Il s'agit d'une architecture de données qui combine les avantages des entrepôts de données et des lacs de données, en offrant une approche unique et cohérente pour la gestion des données. Les données sont stockées au format Parquet dans le lac de données, permettant ainsi un traitement continu et par lots.

- **Permet une architecture Lakehouse :** Delta Lake permet une architecture de données continue et simplifiée qui permet aux organisations de gérer et de traiter d'énormes volumes de données en continu et par lots sans les tracas de gestion et d'exploitation liés à la gestion séparée du streaming, des data warehouses et des lacs de données.
- **Permet une gestion intelligente des données pour les lacs de données :** Delta Lake offre une gestion efficace et évolutive des métadonnées, qui fournit des informations sur les volumes de données massifs dans les lacs de données. Grâce à ces informations, les tâches de gouvernance et de gestion des données se déroulent plus efficacement.

- **Application du schéma pour une meilleure qualité des données :** Étant donné que les lacs de données n'ont pas de schéma défini, il devient facile pour les données mauvaises/incompatibles d'entrer dans les systèmes de données. La qualité des données est améliorée grâce à la validation automatique du schéma, qui valide la compatibilité DataFrame et table avant les écritures.
- **Permet la transaction ACID :** La plupart des architectures de données organisationnelles impliquent de nombreux mouvements ETL et ELT entrant et sortant du stockage de données, ce qui l'ouvre à plus de complexité et d'échecs aux points d'entrée des nœuds. Delta Lake garantit la durabilité et la persistance des données pendant l'ETL et d'autres opérations de données. Delta Lake capture toutes les modifications apportées aux données pendant les opérations de données dans un journal des transactions, garantissant ainsi l'intégrité et la fiabilité des données pendant les opérations de données.

3.2 Principaux avantages et caractéristiques de Delta Lake

Avec Delta Lake, les données sont stockées dans un format optimisé, tel que Parquet, dans un lac de données. Ce format facilite le traitement efficace des requêtes, quel que soit le mode d'accès aux données, qu'il s'agisse d'un traitement streaming ou par batch.

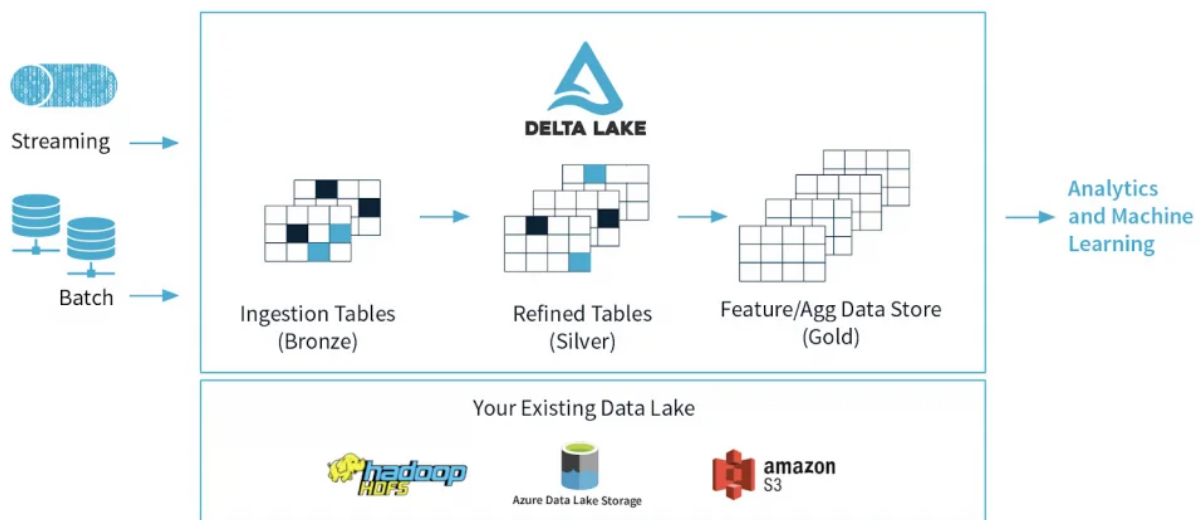


FIGURE 3.1 : Architecture multi-sauts de Delta Lake

- **Pistes d'audit et historique :** Dans Delta Lake, chaque écriture existe en tant que transaction et est enregistrée en série dans un journal des transactions. Par conséquent, toutes les modifications ou validations apportées au journal des transactions sont enregistrées, laissant une trace complète à utiliser dans les audits historiques, la gestion des versions ou à des fins de voyage dans le temps. Cette fonctionnalité de Delta Lake permet de garantir l'intégrité et la fiabilité des données pour les opérations de données d'entreprise.

- **Voyager dans le temps et versionner les données** : Étant donné que chaque écriture crée une nouvelle version et stocke l'ancienne version dans le journal des transactions, les utilisateurs peuvent afficher/restaurer les anciennes versions de données en fournissant l'horodatage ou le numéro de version d'une table ou d'un répertoire existant à l'API de lecture Sparks. À l'aide du numéro de version fourni, Delta Lake construit ensuite un instantané complet de la version avec les informations fournies par le journal des transactions. Les retours en arrière et la gestion des versions jouent un rôle essentiel dans l'expérimentation de l'apprentissage automatique, où les scientifiques des données modifient de manière itérative les hyperparamètres pour former des modèles et peuvent revenir aux modifications si nécessaire.
- **Unifie le traitement par lots et par flux** : Chaque table d'un Delta Lake est un puits de lot et de flux. Avec le streaming structuré Sparks, les organisations peuvent diffuser et traiter efficacement les données de streaming. De plus, grâce à la gestion efficace des métadonnées, à la facilité d'évolutivité et à la qualité ACID de chaque transaction, l'analyse en temps quasi réel devient possible sans utiliser une architecture de données à deux niveaux plus compliquée.
- **Gestion efficace et évolutive des métadonnées** : Delta Lakes stocke les informations de métadonnées dans le journal des transactions et exploite la puissance de traitement distribuée de Spark pour traiter rapidement, lire et gérer efficacement de gros volumes de métadonnées de données, améliorant ainsi la gouvernance des données.
- **transactions ACID** : Delta Lakes garantit que les utilisateurs voient toujours une vue de données cohérente dans une table ou un répertoire. Il garantit cela en capturant chaque modification effectuée dans un journal de transactions et en l'isolant au niveau d'isolation le plus fort, le niveau sérialisable. Au niveau sérialisable, chaque opération existante a et suit une séquence en série qui, lorsqu'elle est exécutée une par une, fournit le même résultat que celui indiqué dans le tableau.
- **Opérations du langage de manipulation de données** : Delta Lakes prend en charge les opérations DML telles que les mises à jour, les suppressions et les fusions, qui jouent un rôle dans les opérations de données complexes telles que la capture de données de modification (CDC), les upserts en continu et la dimension à évolution lente (SCD). Des opérations comme CDC assurent la synchronisation des données dans tous les systèmes de données et minimisent le temps et les ressources consacrés aux opérations ELT. Par exemple, en utilisant le CDC, au lieu d'ETL toutes les données disponibles, seules les données récemment mises à jour depuis la dernière opération subissent une transformation.
- **Schema Enforcement** : Delta Lakes effectue une validation automatique du schéma en vérifiant un ensemble de règles pour déterminer la compatibilité d'une écriture d'un DataFrame vers une table. L'une de ces règles est l'existence de toutes les colonnes DataFrame dans la table cible. Une occurrence d'une colonne supplémentaire ou manquante dans le DataFrame génère une erreur d'exception. Une autre règle est que le DataFrame et la table cible doivent contenir les mêmes types de colonnes, ce qui, sinon, déclenchera une exception. Delta Lake utilise également DDL (Data Definition Language) pour ajouter explicitement de nouvelles colonnes. Cette

fonctionnalité de lac de données permet d'éviter l'ingestion de données incorrectes, garantissant ainsi une qualité élevée des données.

- **Compatibilité avec l'API de Spark :** Delta Lake est basé sur Apache Spark et est entièrement compatible avec l'API Spark, qui permet de créer des pipelines de données volumineuses efficaces et fiables.
- **Flexibilité et intégration :** Delta Lake est une couche de stockage open source et utilise le format Parquet pour stocker des fichiers de données, ce qui favorise le partage de données et facilite l'intégration avec d'autres technologies et stimule l'innovation.

3.3 Trino

Trino, anciennement connu sous le nom de Presto, est un moteur de requêtes SQL distribué et open-source. Il est conçu pour exécuter des requêtes interactives et analytiques à grande échelle sur des données hétérogènes et distribuées. Trino offre une grande polyvalence en permettant l'accès à différents types de sources de données, qu'il s'agisse de bases de données relationnelles, de systèmes de fichiers, de sources de données en temps réel ou de services de stockage cloud. Grâce à sa conception distribuée, Trino permet des performances élevées et une scalabilité horizontale, ce qui en fait un outil essentiel pour l'analyse des données dans notre solution.

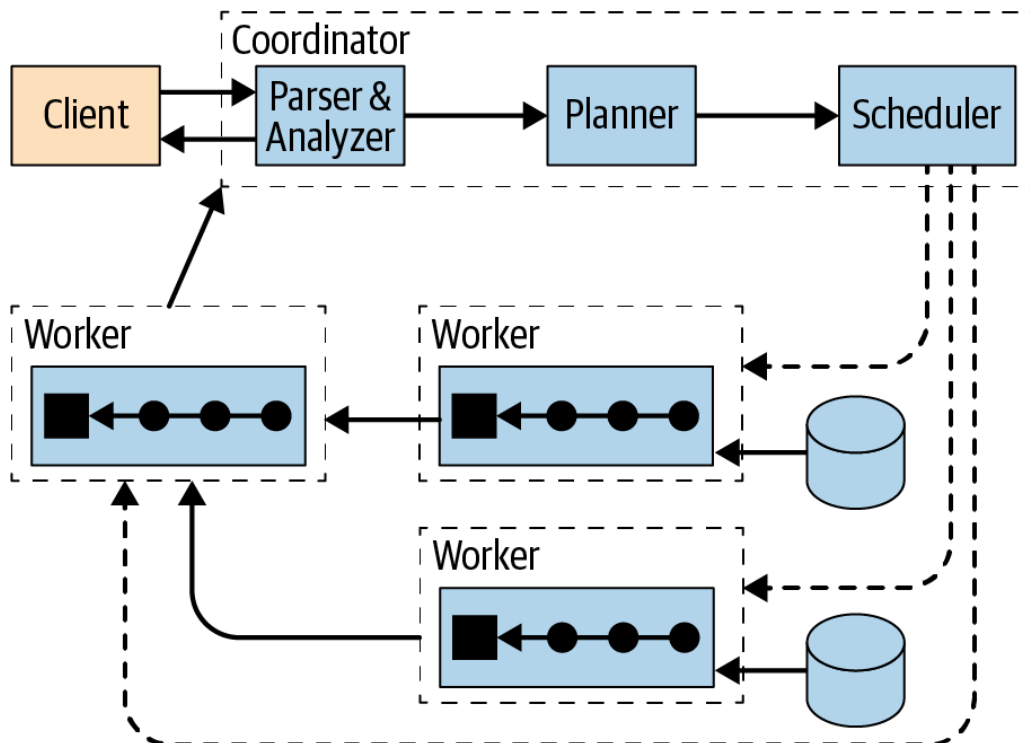


FIGURE 3.2 : Vue d'ensemble de l'architecture Trino avec le coordinateur et les workers

1. Un coordinateur est un serveur Trino qui gère les requêtes entrantes et gère les workers pour exécuter les requêtes.

2. Un worker est un serveur Trino responsable de l'exécution des tâches et du traitement des données.
3. Le service de découverte s'exécute généralement sur le coordinateur et permet aux workers de s'inscrire pour participer au cluster.
4. Toutes les communications et tous les transferts de données entre les clients, le coordinateur et les workers utilisent des interactions basées sur REST sur HTTP/HTTPS.

Conclusion Générale
