|  |  |  |
| --- | --- | --- |
| http://lavande.cpe.fr/intranet/communication/2019/logo_couleur.jpg | **RAPPORT DE STAGE** | **Date du stage**  du 06/02/2023  au 28/07/2023 |
| Stage d’exécution ❒ Stage élève – ingénieur ❒  Stage année de césure :  12 mois ❒ 1er semestre ❒ 2ème semestre ❒  Stage projet de fin d’études ☒ |

**Nom et Prénom du stagiaire :** NILI Marouane

Spécialité :   ❒ CGP ☒ ETI

**CONFIDENTIALITE du RAPPORT :** ☒ **OUI ❒ NON**

**Titre du rapport en français : Optimisation de gestion des stocks chez Carrefour**

**Titre du rapport en anglais:  Optimizing stock management at Carrefour**

Nom de l’entreprise /organisme : ALTECA Ville : Lyon Pays : France

Nom et Prénom du maître de stage : Jean Baptiste EVIEUX

Fonction : Lead dev en big data Adresse e-mail : jbevieux@alteca.fr

Remerciements

Au terme de ce stage, je souhaite exprimer mes sincères remerciements à toutes les personnes qui ont contribué à la réalisation de ce travail. Mon expérience professionnelle s'est révélée être riche en enseignements et en découvertes grâce aux conseils, aux encouragements et à l'orientation que j'ai reçus de la part de mes différents encadrants. J'ai eu la chance de travailler au sein d'une équipe compétente et bienveillante et je suis conscient que sans leur soutien, la réalisation de ce travail n'aurait pas été possible.

Je tiens à exprimer ma profonde gratitude envers mon tuteur de stage, Monsieur Jean-Baptiste EVIEUX, ainsi que Madame Louciné ASSATRIAN, mon Manager, pour leur encadrement, leurs conseils et leurs encouragements tout au long de mon stage de PFE. Grâce à eux, j'ai pu acquérir de nouvelles compétences, développer mes connaissances et me familiariser avec les rouages du monde professionnel. Leurs encouragements, leur patience et leur soutien ont été des éléments clés dans la réussite de cette expérience et j'en suis très reconnaissant.

Je tiens également à exprimer ma gratitude envers l'équipe de Phenix de Carrefour avec qui j'ai eu la chance de travailler. Leur soutien, leur collaboration professionnelle, leurs connaissances, leurs conseils et leur bonne humeur m'ont permis de me sentir intégré et soutenu tout au long de ce stage. J'ai énormément appris de leur expertise, de leur professionnalisme et de leur attitude positive et je les remercie profondément pour leur contribution à cette expérience.

Enfin, je tiens à exprimer ma reconnaissance envers toutes les personnes qui ont contribué de manière indirecte au succès de ce stage et à la réalisation de ce rapport. Vos conseils, vos encouragements et vos retours ont été d'une grande importance pour moi et m'ont permis d'avancer avec confiance et détermination. Votre soutien a été un véritable moteur pour la concrétisation de ce travail. Je suis profondément reconnaissant(e) d'avoir eu l'opportunité de travailler avec une équipe aussi compétente et bienveillante et j'espère avoir l'occasionde collaborer avec vous à nouveau dans le futur.

# Résumé

Le présent document synthétise le travail effectué dans le cadre de mon stage de fin d’études à Alteca Lyon au sein de la Platform data (PFD) du client Carrefour.

Au cours de ce stage, j'ai eu l'opportunité de travailler sur la mise en place de nouvelles fonctionnalités dans le projet "Phénix" de Carrefour. Ce projet, qui agit comme un Data Lake pour l'entreprise, permet de gérer et de traiter les données provenant de différentes sources en exposant ces données à travers d'autres plateformes de prises de décision. L'objectif était d'optimiser la gestion des stocks et leur allocation aux différents magasins de Carrefour. En utilisant des technologies de big data telles que GCP, Kubernetes, Jenkins et Azkaban, nous avons développé une application robuste, performante et riche en fonctionnalités pour répondre à ce besoin.

Ce rapport de stage met donc en avant les différentes tâches que j'ai pu effectuer durant ma période de stage, avec un focus particulier sur ce projet. J'ai ainsi contribué à la mise en place de nouvelles fonctionnalités en lien avec le traitement des données de manière à améliorer leur gestion et leur allocation.

# Abstract

This document summarizes the work I did during my end-of-studies internship at Alteca Lyon in the Platform Data (PFD) of Carrefour. During this internship, I had the opportunity to work on implementing new features in Carrefour's "Phénix" project, which functions as a Data Lake for the company. This project manages and processes data from various sources by exposing the data through other decision-making platforms. The objective was to optimize stock management and allocation to the different Carrefour stores. Using big data technologies such as GCP, Kubernetes, Jenkins, and Azkaban, we developed a robust, high-performance and feature-rich application to meet this need.

This internship report highlights the various tasks I performed during my internship, with a particular focus on this project. I contributed to the implementation of new features related to data processing to improve their management and allocation.

Table des matières

[1. Résumé 3](#_Toc139383554)

[2. Abstract 4](#_Toc139383555)

[4. Introduction générale 7](#_Toc139383556)

[5. Organisme d’accueil et client 8](#_Toc139383557)

[I. Introduction 8](#_Toc139383558)

[II. 2.1 Organisme d’accueil : Alteca, une entreprise d’envergure nationale, dynamique et en pleine expansion 8](#_Toc139383559)

[a) Présentation 8](#_Toc139383560)

[b) Expertise sectorielle 8](#_Toc139383561)

[III. 9](#_Toc139383562)

[c) Métiers et Pôles 9](#_Toc139383563)

[d) Agences et Siège social 10](#_Toc139383564)

[IV. Présentation du client : Carrefour 12](#_Toc139383565)

[e) Présentation 12](#_Toc139383566)

[V. Conclusion 13](#_Toc139383567)

[6. Cadre général du projet 14](#_Toc139383568)

[VI. Introduction 14](#_Toc139383569)

[VII. La big data : Un atout essentiel pour Carrefour 14](#_Toc139383570)

[VIII. Présentation de Phenix 16](#_Toc139383571)

[f) Problématique 16](#_Toc139383572)

[g) Solution 17](#_Toc139383573)

[IX. Conduite du projet 22](#_Toc139383574)

[h) Méthodologie agile : Scrum 22](#_Toc139383575)

[i) Workflow d’une tache Phénix 23](#_Toc139383576)

[j) Les réunions 24](#_Toc139383577)

[k) Outils et environnements de développement et de collaboration 25](#_Toc139383578)

[X. Conclusion 27](#_Toc139383579)

# Introduction générale

La concurrence dans le secteur de la vente au détail est de plus en plus difficile en raison des changements technologiques rapides, des investissements accrus, et de la pression des grands groupes pour maintenir des prix bas. Pour relever ce défi, le Groupe Carrefour mise sur la recherche et l'innovation, notamment en transformant ses systèmes et technologies d'information.

En effet, l'utilisation de nouvelles technologies peut avoir un grand impact sur la stratégie de commercialisation et le volume des ventes. Les concurrents de Carrefour exploitent déjà ces technologies sophistiquées, ce qui souligne l'importance pour Carrefour de s'adapter. Pour cela, Carrefour a fait appel à des prestataires pour développer de nouvelles applications pour sa chaîne d'approvisionnement et son assortiment. Le projet "Phénix" de Carrefour vise à optimiser la gestion des stocks et de la distribution de produits dans les différents magasins du groupe. Mon rôle au sein de l'équipe "PFD" était d'exploiter et de trier les données du Data Lake pour réserver le stock quotidien fourni aux magasins Carrefour.

Le présent rapport est constitué de quatre chapitres :

* Le premier chapitre sera dédié à la présentation de l’organisme d’accueil et à la présentation du client.
* Le deuxième chapitre donne un aperçu général du projet. Chapitre dans lequel je délimite le périmètre du projet ainsi que sa conduite.
* Le troisième chapitre présente la partie réalisation et contribution.
* Le quatrième chapitre présente le bilan professionnel et personnel à l’issu du stage, ce que j’ai appris, l’impact sur mon orientation ainsi que les perspectives d’emploi

# Organisme d’accueil et client

## Introduction

Dans ce chapitre, nous allons présenter l’organisme d’accueil : Alteca, ses objectifs, sa structuration et ses activités et le client Carrefour au sein duquel j’ai effectué mon stage.

## 2.1 Organisme d’accueil : Alteca, une entreprise d’envergure nationale, dynamique et en pleine expansion

### Présentation

Alteca est une entreprise française de services numériques (ESN) fondée en 1996 à Lyon par son actuel Président Directeur Général Jean Mougin. Après avoir travaillé dans un grand groupe informatique, M. Mougin a souhaité créer une société de services informatiques indépendante qui mettrait ses collaborateurs au centre de son développement. Au fil des ans, l'entreprise a acquis de nombreuses entreprises dans divers secteurs d'activités et s'est développée dans une logique de proximité régionale tout en gardant une vocation nationale. En effet, Alteca est présent sur tout le territoire français avec 8 agences régionales. Elle est également dédiée aux grands comptes et intègre des compétences complémentaires grâce à des acquisitions stratégiques, ce qui lui permet de se développer rapidement. Aujourd'hui, Alteca compte environ 650 collaborateurs avec un chiffre d’affaires de 60 millions d'euros pour l'année 2022.

### Expertise sectorielle

Alteca, experte dans le domaine des nouvelles technologies et de l’informatique, continue son évolution et s’impose dans plusieurs secteurs d’activités. L’activité principale de la société, est tournée vers le conseil, le développement de logiciels en informatique de gestion, la conception d’applications mobiles et l’accompagnement des projets IoT (Internet of Things). Les principaux secteurs d’activités sont la grande distribution, l’industrie, le transport, le service public, les télécoms et les secteurs banque & finance, assurance & mutuelle, La clientèle d’Alteca est constituée d’entreprises d’importance majeure dans de multiples domaines comme le montre la figure suivante :



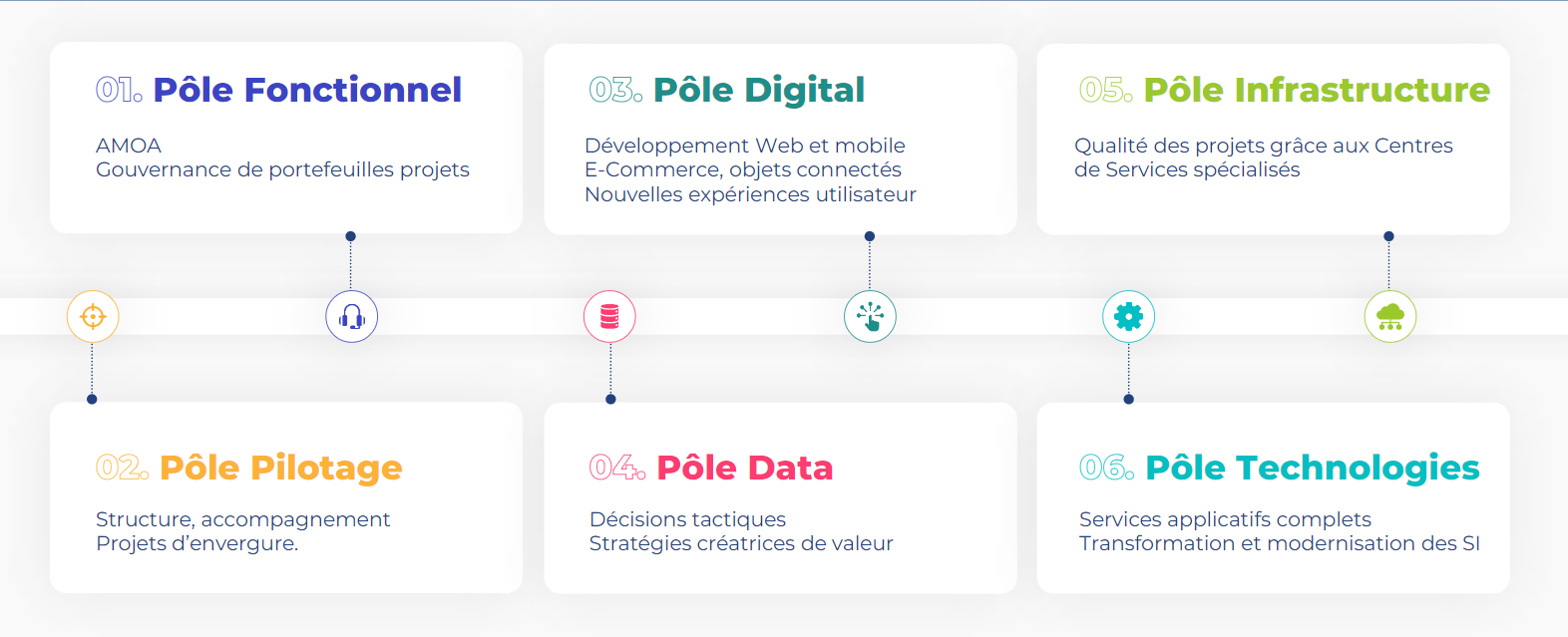
Figure 1: répartition du chiffre d'affaire selon les secteurs et les différents client par secteur

## 

### Métiers et Pôles

Les métiers d’Alteca sont regroupés en plusieurs catégories chacune rattachée à un pôle. Les principaux métiers tournent autour de : Conseiller le client, développer de nouvelles solutions pour répondre aux besoins des clients, maintenir les applications existantes qui est le cœur du métier, assurer la qualité de ses différents services, au sein de chaque agence existe 6 pôles différents regroupant les compétences suivantes :

* Le pôle digital : il propose les meilleurs usages autour du développement Web et mobile, du e-Commerce, des objets connectés, des nouvelles expériences utilisateurs, du Cloud.
* Le pôle data : celui-ci propose des solutions qui permettent de dégager de la valeur ajoutée à partir de données afin d’instruire des décisions tactiques et des stratégies créatrices. Alteca intervient tout au long du cycle de transformation de la donnée de ses clients pour en faire un indicateur métier pertinent qui aide les décideurs à se positionner. Le pôle prend en charge les problématiques de gestion de données structurées ou non structurées comme le Big Data.
* Le pôle technologies : il propose des services applicatifs complets pour la transformation et la modernisation du Système d’Information, quelle que soit sa base technologique.
* Le pôle fonctionnel : son but est d’accompagner la transformation digitale des clients avec des plans d’optimisation de processus, l’amélioration de l’expérience clients, implication des collaborateurs, conformité réglementaire, création de nouveaux services, etc.
* Le pôle pilotage structure : accompagne et délivre des projets d’envergure répondants aux exigences et respectueux des engagements.
* Le pôle infra/déploiement : garantit la qualité de la mise en œuvre des projets grâce aux Centres de Services nationaux dédiés et spécialisés.



### Agences et Siège social

Alteca s'est implantée dans 9 villes différentes en France pour couvrir tout le Pays. Chaque agence dispose d'un nombre restreint de collaborateurs et la direction est proche des équipes locales. Cette approche favorise une relation étroite avec les clients situés dans chaque région, tout en profitant d'une coordination nationale pour garantir une collaboration à long terme.

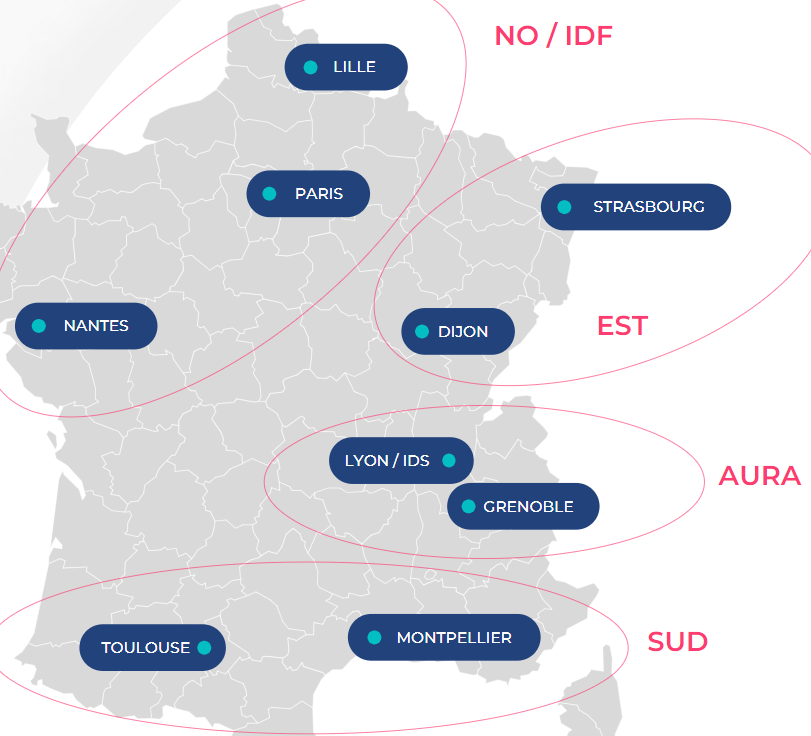


Figure 2 : Répartition des agences et des centres de services d'Alteca suivants les régions

La première agence d'Alteca a vu le jour à Lyon en 1996 sous la direction de Pascal Ferrary. Elle se situe dans le 6ème arrondissement, à proximité du parc de la Tête d'Or, et offre des services informatiques d'ingénierie. Les projets sont menés soit au sein du centre de services chez Alteca soit directement chez les clients en assistance technique. Les principaux secteurs ciblés sont la grande distribution, l'industrie et la banque. Alteca a développé deux sites différents dans la ville de Lyon pour répondre aux besoins de son personnel en termes de ressources humaines et matérielles. L'un de ces sites, situé dans le 6ème arrondissement, est exclusivement dédié à la mise en place, l'installation et la maintenance de matériels informatiques. L'agence lyonnaise est responsable de toutes les offres proposées par Alteca dans la région Auvergne-Rhône-Alpes. J'ai effectué mon stage au sein de cette agence, plus précisément dans l'équipe data.

## Présentation du client : Carrefour

### Présentation



Figure 3 : Logo Carrefour

Le groupe Carrefour spécialisé dans le domaine de la grande distribution, fait partie des leaders dans son domaine, en France, en Europe et dans le monde. Plus précisément, Carrefour occupe la deuxième position après Walmart en chiffre d'affaires. La société Carrefour a été créée en 1959 à Annecy par les familles Fournier, Badin et Defforey. Un premier supermarché s’est ouvert à Annecy en 1960. Carrefour invente ensuite le concept de l'hypermarché en 1963 et inaugure son premier hypermarché dans l'Essonne, à Sainte Geneviève-des-Bois. Ce premier hypermarché possède une surface de vente de 2500 m², 12 caisses et 400 places de parking. Il propose un large assortiment de produits alimentaires et non alimentaires. L'entreprise Carrefour est introduite à la bourse de Paris en 1970. En 1999, elle fusionne avec le groupe Promodès (créé en 1961) et devient ainsi le premier groupe de grande distribution en Europe, et le deuxième au niveau mondial. Le groupe Carrefour se développe à l'international avec l'ouverture de magasins en Belgique, en Espagne et au Brésil dès les années 1970. Il est aujourd'hui présent sur trois grands marchés dans le monde : l'Europe, l'Amérique latine et l'Asie, et ses magasins sont implantés dans plus de 30 pays. Plus de 53% du chiffre d'affaires du groupe se réalise hors de France. Même si l'hypermarché est le format de référence du groupe, celui-ci sait se diversifier. Il développe divers formats de magasins, tels que les supermarchés, les magasins de proximité, les magasins cash & carry (commerce de gros à destination des professionnels), le drive ou encore la livraison à domicile. Le groupe propose également différents services tels que Carrefour banque, assurances, voyages ou encore spectacles. En 2020, le groupe Carrefour réalise un chiffre d'affaires de 78,600 milliards d'euros. Il emploie plus de 321 383 collaborateurs à travers le monde et possède 9 600 magasins répartis en 779 hypermarchés, 2 320 supermarchés, 4 014 maxi-discompte, 2 322 magasins de proximité, 198 points cash & carry.

## Conclusion

Après avoir présenté l’organisme d’accueil et le client et ses différentes caractéristiques, nous allons présenter dans le chapitre suivant l’aperçu général du projet. Nous allons présenter le périmètre du projet ainsi que sa conduite.

# Cadre général du projet

## Introduction

Le cadre général du projet constitue une étape essentielle à aborder dans le cadre d'un rapport de stage. Cette section vise à situer le projet dans son environnement global en mettant en évidence les différents éléments qui ont contribué à sa mise en place. Il s'agit d'exposer les objectifs généraux du projet, de décrire le contexte dans lequel il s'inscrit, ainsi que les acteurs clés impliqués. Cette introduction permettra donc de donner une vision d'ensemble du projet, facilitant ainsi la compréhension des différentes étapes et actions entreprises tout au long du stage.

## La big data : Un atout essentiel pour Carrefour

La big data, ou "grande quantité de données" en français, est un terme qui désigne les ensembles de données massives et complexes qui sont générées au quotidien. Ces données proviennent de différentes sources telles que les réseaux sociaux, les transactions en ligne, les capteurs, les appareils connectés, etc.

La caractéristique principale de la big data est liée aux "5V", qui sont les cinq aspects clés de ces données :

* Volume : La big data se caractérise par des quantités de données massives qui, à l'ère numérique, sont produites à une échelle sans précédent. Ces volumes peuvent atteindre des pétaoctets (millions de gigaoctets) voire des exaoctets (milliards de gigaoctets).
* Vélocité : Les données sont générées à une vitesse élevée et en temps réel. Par exemple, les réseaux sociaux génèrent des milliers de tweets et de publications chaque seconde. L'analyse et le traitement de ces données en temps réel deviennent cruciaux pour en tirer des informations utiles.
* Variété : La big data se compose de différentes variétés de données. Elle peut inclure des données structurées (telles que des tableaux de données provenant de bases de données relationnelles), des données non structurées (comme des textes, des images, des vidéos, etc.) et des données semi-structurées (comme des e-mails et des documents XML).
* Véracité : La big data peut-être de qualité variable. Les données peuvent être erronées, incomplètes ou provenir de sources peu fiables. Il est important de s'assurer de la qualité des données avant de les utiliser pour des analyses ou des prises de décision.
* Valeur : L'objectif de l'analyse de la big data est de tirer des connaissances et des informations exploitables pour les entreprises et les organisations. En comprenant et en traitant ces masses de données, il est possible de prendre des décisions plus éclairées et d'obtenir un avantage concurrentiel.

La big data représente le défi de gérer des quantités massives de données provenant de différentes sources, en temps réel, dans des formats variés. Son exploitation efficace nécessite des compétences en analyse de données, des outils technologiques adaptés et une compréhension approfondie de ses 5V

Carrefour bénéficie de plusieurs avantages stratégiques en exploitant la big data dans ses opérations. Tout d'abord, elle permet à l'entreprise de prendre des décisions basées sur les données en fournissant une vision précise de son activité. L'analyse des tendances de vente, des préférences des clients et des comportements d'achat permet à Carrefour d'optimiser son assortiment de produits, sa logistique, ses prix et ses stratégies marketing.

De plus, la gestion de l'inventaire est grandement améliorée grâce à la big data. En analysant les données d'inventaire en temps réel, Carrefour peut prévoir les fluctuations de la demande, éviter les ruptures de stock, optimiser les niveaux d'inventaire et réduire les coûts de stockage.

La personnalisation de l'expérience client est un autre avantage clé offert par la big data. En comprenant les besoins et les préférences individuelles des clients, Carrefour peut proposer des offres personnalisées, améliorer la satisfaction client et maximiser les opportunités de vente croisée et de vente incitative.

La big data aide également Carrefour dans la gestion de ses relations fournisseurs. En analysant les données de qualité, de livraison et de prix, Carrefour peut sélectionner les fournisseurs les plus fiables, optimiser les négociations et garantir une chaîne d'approvisionnement solide.

Enfin, l'analyse de la concurrence est facilitée par la big data. En surveillant les prix, les promotions, les comportements des concurrents et les changements de tendances, Carrefour peut ajuster ses propres stratégies et rester compétitif sur le marché.

En conclusion, la big data représente un atout essentiel pour Carrefour. Son exploitation efficace permet à l'entreprise de prendre des décisions éclairées, de mieux comprendre ses clients et son marché, d'optimiser ses opérations et de renforcer sa compétitivité globale. Grâce à l'analyse de la big data, Carrefour est en mesure de rester à l'avant-garde de l'industrie de la vente au détail et de continuer à satisfaire les besoins changeants des consommateurs.

## Présentation de Phenix

### Problématique

Carrefour, un groupe d'envergure, utilise une multitude d'applications pour gérer divers aspects tels que le stock, les factures, et bien d'autres. Chacune de ces applications génère quotidiennement un volume considérable de données provenant de sources variées. Ces données englobent les informations relatives aux magasins, à l'assortiment des produits, aux commandes, aux tickets de caisse, aux promotions, et bien plus encore.

Le flux de données entre ces différentes applications pose des défis majeurs, notamment en termes de volumétrie, d'origine des données et de temps de réponse. De plus, la collecte des données à partir de sources différentes peut entraîner des problèmes de formats incompatibles.

### Solution

#### Présentation générale du programme Phénix

Pour résoudre les problèmes rencontrés lors du traitement de ces données, Carrefour a pris la décision de créer une plateforme Big Data appelée Phénix. Cette plateforme tire parti des avantages du concept Big Data en regroupant toutes les données au sein d'un même endroit et en agissant en tant qu'intermédiaire entre les sources de données et les applications qui en ont besoin.

Phénix offre la capacité de stocker et de traiter de gros volumes de données. Grâce à sa puissance et à sa capacité de stockage, elle est en mesure de gérer potentiellement un nombre infini de tâches. À mesure que la quantité et la variété des données augmentent, ainsi que la diversification de leurs sources, des outils de plus en plus sophistiqués sont nécessaires pour assurer leur traitement. C'est là que réside le principal avantage de cette plateforme, qui propose une multitude d'outils permettant un traitement rapide et fluide des données.

#### La cible fonctionnelle et data de Phénix

Phénix consolide la donnée et la met à disposition par API à l’ensemble des objets métiers représentés par la figure ci-dessous

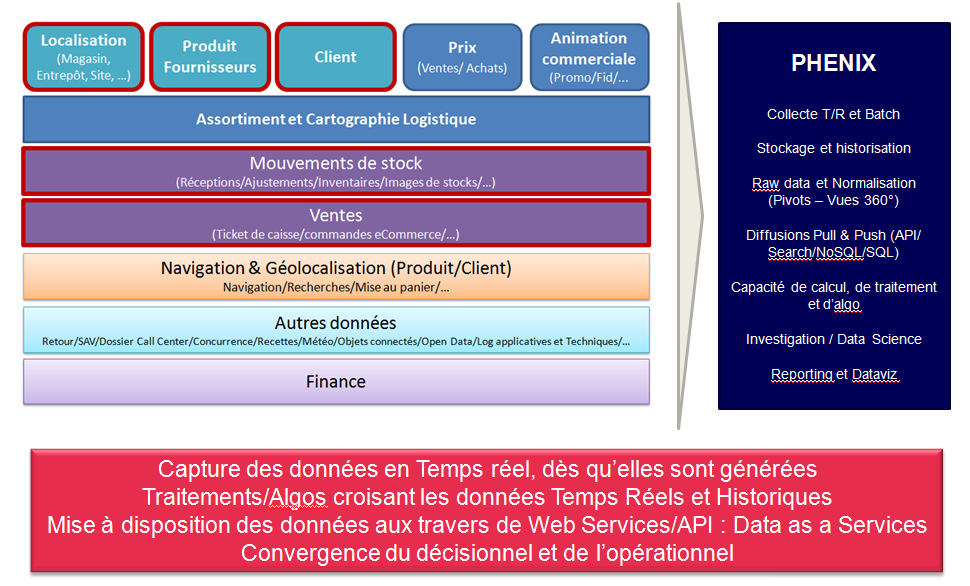


Figure 4 : Cible fonctionnelle de Phénix

#### Les ingestions et exposition phénix

La figure suivante illustre les processus d'ingestion et d'exposition des données dans la plateforme Phénix, en fonction des différentes cibles fonctionnelles. Phénix joue un rôle crucial en permettant la collecte et l'intégration de données provenant de diverses sources, afin de les rendre accessibles et exploitables pour différentes fonctions au sein de l'entreprise Carrefour.

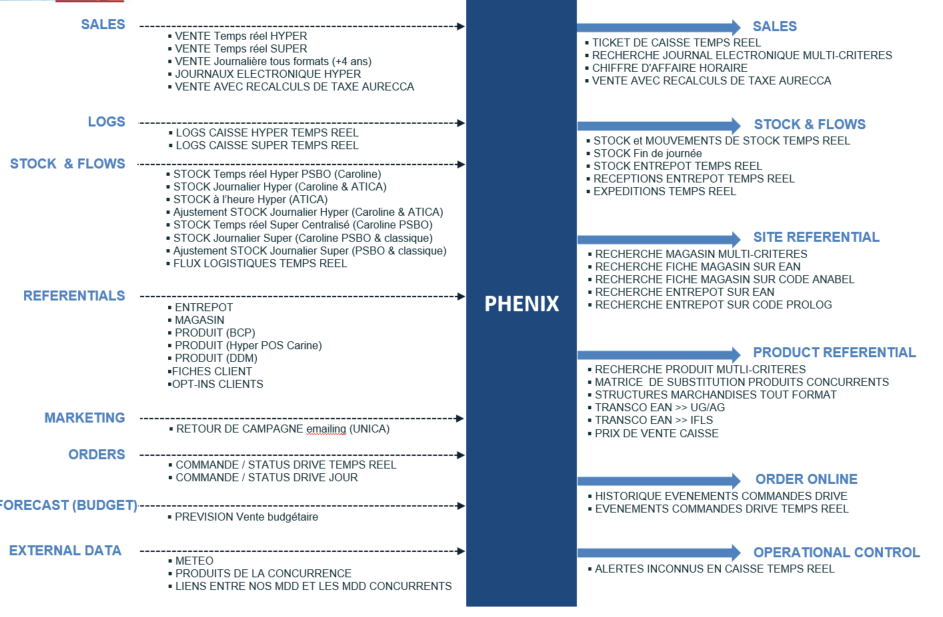


Figure 5: ingestions et exposition phénix

Les cibles fonctionnelles comprennent les ventes, les journaux (logs), les stocks et flux, les référentiels, le marketing, les commandes, les prévisions et les données externes. Chacune de ces catégories représente un domaine spécifique où les données sont utilisées pour des analyses, des prises de décision et des actions pertinentes.

En analysant la figure, on peut observer les flux d'ingestion des données provenant de chaque source spécifique vers la plateforme Phénix. Ces données sont ensuite transformées, nettoyées et préparées pour être exposées aux différentes cibles fonctionnelles. L'exposition des données permet aux différents départements et équipes de Carrefour d'accéder aux informations pertinentes et de les utiliser pour leurs besoins spécifiques, qu'il s'agisse de suivre les ventes, d'analyser les tendances du marché, de gérer les stocks ou de prendre des décisions stratégiques.

La représentation visuelle de ces processus d'ingestion et d'exposition des données dans Phénix aide à visualiser la complexité de la gestion des données au sein de Carrefour et met en évidence le rôle central de cette plateforme dans la consolidation et l'utilisation efficace des informations pour soutenir les différentes activités de l'entreprise.

#### Une idée des volumes

PHENIX en quelques chiffres :

 160,000,000 messages temps réel par mois venant de 1,300 magasins (Hyper et Super).

 10,000 fichiers par jours.

 4,000,000 fichiers stockés dans le data lake.

 200,000 appels par mois aux APIs par les applications métiers.

 200 TB stockés dans GCS

 10 TB stockés dans Cassandra

 5 TB stockés dans Elastic Search

#### Architecture Phénix

Phénix est capable de traiter les données de deux manières différentes : via un traitement batch ou via un traitement en temps réel :

* **Traitements d’ingestion online (temps réel) :** Les données en temps réel arrivent par l’intermédiaire de messages Kafka. A partir de ceux-ci, on stocke ces dernières de manière distribuée grâce au système de stockage GCS. GCS est le système de stockage principal de Phénix car il permet de stocker les données de manière distribuée et il garantit l’immuabilité des données (celles-ci ne peuvent pas être modifiées). Ce système a remplacé HDFS car il permet d’avoir un Framework de chargement des fichiers, réduire le Time To Market pour les expositions SQL et une diminution des coûts.
* **Traitements d’ingestion offline (batch)** : Les données sont cette fois récupérées sous forme de fichiers (par exemple des Tables big query, fichiers txt et csv), et comme précédemment on cherche à stocker ces données dans GCS.

Le Groupe Carrefour a décidé de changer son infrastructure et migre sa partie Hadoop vers les nouvelles solutions Google pour le processing Big Data.

En effet Google Cloud Storage (GCS) et Hadoop Distributed File System (HDFS) sont deux systèmes de stockage de fichiers populaires utilisés dans le domaine du Big Data. Bien qu'ils partagent certaines similitudes, ils présentent également des différences significatives en termes de conception et de fonctionnalités. Voici quelques avantages de GCS par rapport à HDFS :

* **Évolutivité et flexibilité** : GCS offre une évolutivité élevée et une grande flexibilité en termes de capacité de stockage. Il permet de stocker et de gérer des pétaoctets de données, sans nécessiter de configuration complexe ou de dimensionnement préalable. Par rapport à HDFS, GCS peut facilement s'adapter à la croissance des données sans nécessiter d'ajustements matériels.
* **Durabilité et disponibilité** : GCS garantit une durabilité élevée des données grâce à une réplication automatique sur plusieurs emplacements géographiques. Cela réduit les risques de perte de données en cas de défaillance matérielle ou de catastrophe naturelle. De plus, GCS offre une disponibilité élevée, assurant l'accès aux données à tout moment, sans interruption de service significative.
* **Intégration avec les services Google Cloud** : GCS est étroitement intégré à l'écosystème de services Google Cloud, ce qui facilite l'intégration et l'utilisation conjointe avec d'autres services tels que BigQuery, Dataflow, ou encore AI Platform. Cette intégration permet des flux de travail efficaces et une utilisation harmonieuse des données entre les différents services.
* **Sécurité avancée** : GCS offre des fonctionnalités de sécurité avancées, telles que le chiffrement des données au repos et en transit, l'authentification basée sur des clés d'accès, ainsi que des contrôles d'accès granulaires grâce à l'utilisation des rôles et des politiques d'accès. Ces fonctionnalités renforcent la confidentialité et la protection des données stockées dans GCS.
* **Tarification flexible** : GCS propose une tarification basée sur l'utilisation réelle des ressources de stockage et de transfert de données. Il offre des options de stockage à faible coût pour les données moins fréquemment utilisées, ce qui permet d'optimiser les coûts de stockage en fonction des besoins spécifiques.

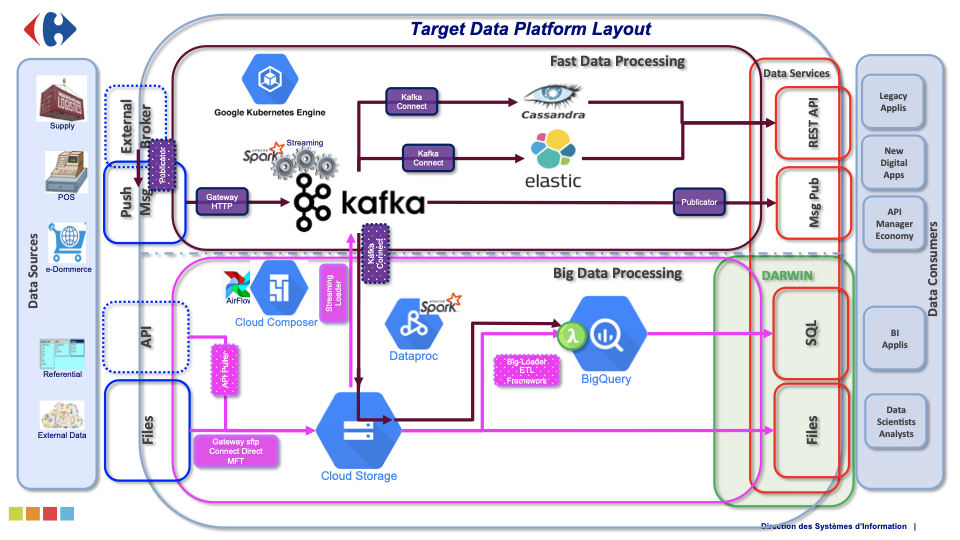


Figure 6 : architecture de phénix après la migration

#### Exemple et cas d’usage de Phenix

* Besoins : Les magasins ont besoin de pouvoir rechercher rapidement des transactions dans les journaux électroniques de caisse basé sur des critères de recherche multiples (recherche libre, montant, moyen de paiement …). L’ancienne application n’est plus maintenue et lourde d’utilisation (basée sur l’archivage en CDROM)
* Solution : Phénix ingère, histories et indexe tous les journaux de caisse des Hypermarchés. Un service de recherche est mise à disposition de l’application « Ticket de caisse ». Le magasin peut maintenant faire des recherches multicritères sur une période donnée (historique d’un an). Ce service de recherche est disponible pour d’autres application.

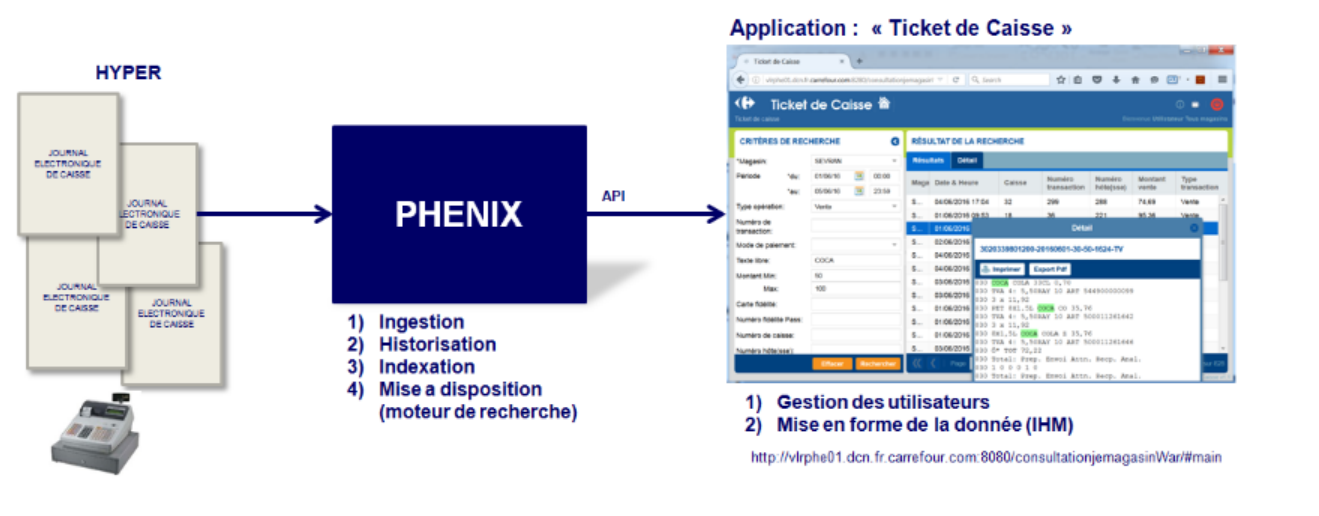


Figure 7: Cas d'usage, utilisation d'un moteur de recherche sur données non-structurée

## Conduite du projet

Tout au long de la période de mon stage, je travaille un jour par semaine chez le client Carrefour au Danica à la Part-Dieu et le reste de la semaine dans les locaux d’Alteca ou en télétravail. Cela m’a permis d’être plus proche du client, de communiquer directement avec les PO et de s’entraider avec les membres de mon équipe. Le projet Phénix de Carrefour suit une méthode de conduite spéciale offrant ainsi plus de visibilité sur l’avancement du travail.

### Méthodologie agile : Scrum

L’organisation du travail au sein de l’équipe se base sur la méthodologie Scrum, qui fait partie des méthodes agiles. La méthodologie agile Scrum est largement utilisée dans le développement de logiciels et de produits complexes. Elle se base sur des principes de transparence, d'adaptabilité et de collaboration pour fournir des résultats de haute qualité de manière itérative et continue. Scrum fonctionne en cycles appelés "sprints" qui durent généralement de deux semaines chez carrefour. Chaque sprint démarre par une réunion de planification où l'équipe définit les objectifs et les tâches à accomplir appelée revue de Spec. Pendant le sprint, l'équipe travaille de manière autonome et se réunit quotidiennement lors des "Daily" pour partager les progrès, les défis et les ajustements nécessaires. À la fin du sprint, une démonstration du travail accompli est présentée aux parties prenantes lors de la réunion de revue du sprint. Une rétrospective est ensuite réalisée pour identifier les points positifs et les améliorations possibles. Cette approche itérative permet à l'équipe de s'adapter rapidement aux changements, de prendre des décisions basées sur des retours d'expérience concrets et de livrer des fonctionnalités utilisables à intervalles réguliers. Scrum favorise également la collaboration étroite entre les membres de l'équipe et les parties prenantes, ce qui facilite la communication, la résolution de problèmes et la prise de décisions conjointes.

### Workflow d’une tache Phénix

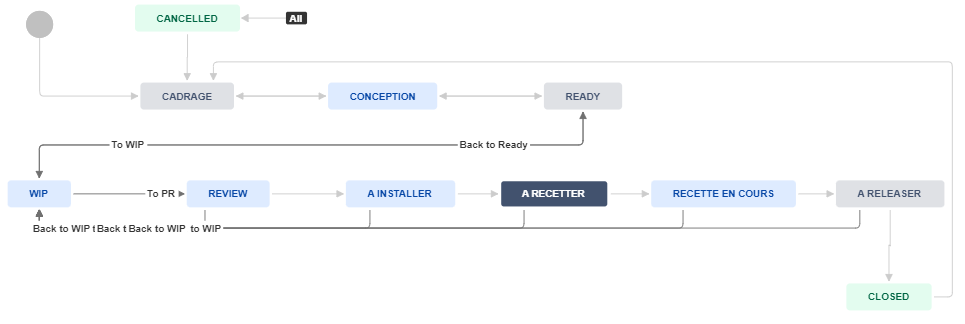


Figure 8 : Workflow d'une tache Phénix dans Jira

Le workflow représente l’avancement d’une tâche Phénix, le développeur fait avancer le statut de sa carte qui concrétise son avancement et qui facilite le suivi de ces tâches par les responsables. Une carte passe par plusieurs états :

1. Le PO présente son besoin fonctionnel à l'équipe de développement.
2. Une Epic de rattachement et une User Story de typologie "Conception" sont créées par le PO et l’User Story est affectée à un développeur, qui deviendra le lead sur le sujet et rentre dans le sprint comme une autre carte. L'affectation dépendra des priorités des cartes en cours réalisées par les différents développeurs et éventuellement de leur connaissance du sujet. S'ils travaillaient sur une carte non prioritaire, celle-ci retournera en haut du backlog ou sera réaffectée à un autre dev.
3. Le lead, seul ou avec l'aide d'autres personnes si nécessaire, travaille sur la conception.  
   Cela inclut notamment de trouver et définir une solution technique au besoin et de lister toutes les actions à réaliser.
4. Le lead présente à d'autres développeurs la solution conçue durant une réunion de revue de spécification ou durant une réunion de conception ad-hoc. Si celle-ci est validée, la carte de conception est clôturée, et les différentes actions définies sont converties en nouvelles User Story, rattachées à la même Epic et qui seront à prioriser à leur tour par le PO.
5. Le développement peut démarrer sur les User Story.
6. Une fois le développement terminé, le code doit être soumis pour le Code Review sous forme d’une PR (Pull Request) et qui doit être validée par un membre de l’équipe qui n’a pas participé au développement.
7. Après, le développement doit être installé dans un environnement de recette, c'est l'étape qui permet de tester les nouvelles fonctionnalités des paquets applicatifs livrés sur des environnements de recette. Après la validation des résultats par les POs dans les environnements de recette, on passe à la pré-production et la préparation de la mise en production.

Le développeur doit surveiller l’avancement de sa carte et doit relancer les personnes concernées si besoin. Ce suivi des cartes est facilité par certains outils tels que l’outil JIRA.

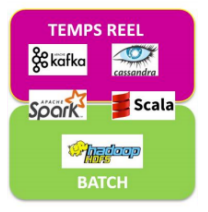
### Les réunions

L’équipe met en place de nombreuses pratiques ainsi que des rituels agiles afin de favoriser l’entraide et le partage de connaissances entre les développeurs.

* **Stand-up (Daily meeting)** : Chaque jour, l’équipe se réunit pendant 15 minutes. Il s’agit d’une réunion pendant laquelle chacun expose rapidement ce qu’il a fait depuis le Stand-up précédent, ce qu’il fera par la suite et s’il rencontre des problèmes particuliers. Cette réunion est l’occasion d’identifier d’éventuels points de blocage et de proposer des solutions à ces problèmes.
* **14-15** : L’équipe a mis en place une réunion appelé « 14-15 » dans laquelle l’équipe se réunit chaque après-midi de 14h à 15h, elle est dédiée à l’entraide, si l’un des membres de l’équipe a des problèmes ou des points bloquants, il les présente lors de cette réunion et les autres membres essaient de l’aider et de le débloquer.
* **La revue de spécifications** : L’équipe a mis en place une méthode de grooming des cartes Jira afin de permettre au développeur de comprendre l’objectif final de la carte. Un grooming est une réunion ou les cartes du backlog sont discutées et possiblement intégrées au prochain sprint. Grâce aux connaissances techniques de l'équipe, elle est capable d'estimer en combien de jours la carte peut être disponible pour une mise en recette. Cette estimation se fait en accord avec toute l’équipe. Mais c’est aussi un moyen de partager les connaissances métier sur les différents sujets à une majeure partie de l’équipe.
* **Rétrospective** : L’équipe se réunit tous les mois pour un événement appelé rétrospective d’une durée d’environ une heure. Cette réunion vise à connaître le ressenti de chacun sur le mois précédent, les points positifs, les points négatifs ainsi que les différents sujets qui ont été exposés à l'équipe.

### Outils et environnements de développement et de collaboration

Le principal enjeu de Phénix est de pouvoir exploiter ses données grâce à différentes technologies / cas d’usage en fonction de la demande. Par exemple, certaines doivent être accessibles par les logiciels en temps réel. Cela signifie que l’on attend l'arrivée d’une nouvelle donnée pour la traiter ou bien nous avons des traitements qui doivent être programmés et déclenchés à la réception d’un fichier.



Temps réel et Batch PHENIX

Et pour y arriver, il existe plusieurs outils nécessaires à la réalisation de ces traitements chez Carrefour. Ces traitements sont représentés avec leurs outils sur le schéma ci-dessus.

* **Kafka** pour la réception des données de partie temps réel de Phénix.
* **Cassandra** est un système de gestion de base données de type NoSQL pour gérer des grosses quantités de données reçues à travers Kafka sur un grand nombre de serveurs.
* **HDFS** permet de stocker et de récupérer des fichiers de façon rapide. Ce système possède une très grande capacité de stockage car il est capable de gérer des milliers de nœuds sans intervention d’un opérateur.
* **Spark** est un moteur de traitement parallèle de données open source qui permet d’effectuer des analyses de grande envergure par le biais de machines regroupées en clusters.
* **Scala** est le langage utilisé pour le développement des différents Job Phénix.
* **Google Cloud Platform** : Carrefour a donc décidé récemment de migrer tous ses jobs vers les nouvelles solutions GCP pour le traitement Big Data.
* **IntelliJ**comme IDE pour le développement.
* **Jenkins** pour l’intégration continue sur les différents environnements.
* **Confluence Carrefour**pour la documentation du projet et le partage d’informations.
* **Jira** pour le suivi des cartes et la gestion du projet.
* **Bitbucket** pour le versioning du code.
* **Nexus** pour la gestion de dépôts et l’hébergement des artéfacts.
* **Azkaban** pour planifier l’exécution des flows.
* **Rundeck** pour l’orchestration.
* **Ansible** pour le déploiement des jobs offline.
* **Kubernetes** pour le déploiement des jobs online.
* **Postman**pour tester les APIs développées.
* **Google Chat et Google Meet et Gmail** pour la communication et l’organisation des réunions entre les membres de l’équipe.

## Conclusion

Durant ce chapitre, j’ai mis le point sur le projet Phénix, ses caractéristiques ainsi que les technologies utilisées pour le mettre en place, ensuite, j’ai parlé de la conduite du projet notamment l’organisation de l’équipe, la méthodologie de travail et les réunions. Le prochain chapitre sera dédié à ma contribution au sein du projet.