

MÉMOIRE DE FIN D'ETUDES

En vue de l'obtention du diplôme

Master 2

En Technologies de l'Information et Web

**Proof Of Concept Big Data permettant de
mettre en œuvre les technologies et les langages
attendus par un client grand compte du secteur
Retail**

Année universitaire : 2021 - 2022

*Ce document est strictement confidentiel et ne doit pas être distribué, communiqué, ou envoyé
à qui que ce soit.*

Réalisé par :

Sara EL MAHFOUDI

Sous la direction de :

Mr. EVIEUX Jean-Baptiste (Alteca)

Mr. MUNTZER Nicolas (Alteca)

Mme. MOULAI Lila (Carrefour)

Mr. DUCHATEAU Fabien (UCBL)

Soutenu le 02/09/2022

Remerciements

Je tiens à exprimer, au terme de ce travail, mes sincères remerciements et témoigner de ma grande reconnaissance à toutes les personnes qui ont contribué de loin ou de près à sa réalisation.

Je tiens à exprimer ma profonde gratitude envers, Mr. EVIEUX Jean-Baptiste mon tuteur de stage et Mr. MUNTZER Nicolas mon Manager pour leurs encadrements, leurs conseils et leurs encouragements tout au long de mon stage de Master.

J'exprime mon entière gratitude envers les membres de l'équipe Phenix pour leur soutien, collaboration professionnelle, connaissances, conseils et bonne humeur chaque jour.

Mes plus sincères remerciements et reconnaissances vont spécialement à mon encadrant pédagogique Mr. DUCHATEAU Fabien pour les directives et les remarques constructives qu'il m'a prodiguées tout au long de mon stage. Ainsi, que tout le corps professoral de l'université Lyon 1 pour leurs efforts dépensés afin de mener à bien notre formation.

Mes remerciements vont également aux membres du jury qui ont accepté d'évaluer ce travail.

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 Software Factory (SF) du client Carrefour.

L'objectif principal du projet auquel j'ai contribué tout au long de mon stage consiste en la mise en place de nouvelles fonctionnalités dans la plateforme « Phénix » de Carrefour.

PHENIX est la plateforme Big Data et le Data Lake de Carrefour qui a comme but de gérer et de traiter les données en provenance de sources hétérogènes et les exposer à travers d'autres plateformes de prises de décision.

L'enjeu principal du stage était d'apporter de nouvelles briques à des traitements existants sans entraîner de régressions afin de mettre à disposition de Carrefour des applications robustes, performantes, riches en fonctionnalités et qui peuvent apporter de la valeur ajoutée au métier en permettant aux utilisateurs de suivre l'évolution des affaires et d'avoir une visibilité sur les projets à venir.

Mots clés : Carrefour, Phénix, Big Data, Data Lake.

Abstract

This document summarizes the work carried out during my end-of-studies internship at Alteca Lyon within the Software Factory (SF) of the client Carrefour.

The main objective of the project to which I contributed throughout my internship consists of the implementation of new functionalities in the “Phenix” application of Carrefour.

PHENIX is Carrefour's Big Data and Data Lake platform that aims to manage and process data from heterogeneous sources and expose them through other decision-making platforms.

The main challenge of the internship was to bring new features to existing processing without causing regressions in order to provide Carrefour with robust, high-performance applications, rich in features and which can bring benefits to the business by allowing users to follow the evolution of business and to have visibility on upcoming projects.

Keys words: Carrefour, Phénix, Big Data, Data Lake.

Tables des matières

Remerciements	2
Résumé.....	3
Abstract	4
Tables des matières.....	5
Liste des figures.....	7
Liste des abréviations.....	8
1 Introduction générale.....	9
2 Organisme d'accueil et Client.....	10
Introduction	10
2.1 Organisme d'accueil : Alteca, une entreprise d'envergure nationale, dynamique et en pleine expansion.....	10
2.1.1 Présentation	10
2.1.2 Expertise sectorielle	10
2.1.3 Métiers et Pôles	11
2.1.4 Agences	12
2.1.5 Agence de Lyon et Siège Social.....	13
2.2 Présentation du client : Carrefour.....	14
2.2.1 Présentation	14
Conclusion.....	14
3 Cadre général du projet	15
Introduction	15
3.1 Le Big Data, c'est quoi ?.....	15
3.2 Présentation du projet.....	16
3.2.1 Problématique	16
3.2.2 Solution : Phénix, Plateforme Big Data de Carrefour	16
3.2.3 Exemple de cas d'usage de Phénix	19
3.3 Conduite du projet.....	19
3.3.1 Méthodologie agile : Kanban	20
3.3.2 Workflow d'une tâche Phénix.....	20
3.3.3 Les réunions	21
3.3.4 Outils et environnements de développement et de collaboration	22

3.3.5 Environnements.....	23
Conclusion.....	23
4 Réalisation et contribution.....	24
Introduction	24
4.1 Formation et Onboarding sur « Phénix »	24
4.1.1 Formation et première tâche Phénix.....	24
4.1.2 Bilan.....	25
4.2 Service de recherche multicritère sur l’assortiment commandable.....	25
4.2.1 Description générale de l’existant	25
4.2.2 Architecture.....	26
4.2.3 Exemple de cas d’usage applicatif	27
4.2.4 Analyse du besoin	27
4.2.5 Mise en œuvre de la solution.....	28
4.3 Autres réalisations	30
Conclusion.....	30
5 Bilan et conclusion	31
5.1 Bilan professionnel et personnel	31
5.2 Conclusion.....	32
Webographie.....	33
Annexes	34
5.3 Annexe A : Cadre général du projet Phénix.....	34
5.3.1 Architecture Phénix avant la migration.....	34
5.3.2 Migration Dataproc	35
5.3.3 Glossaire de Phénix.....	36
5.4 Annexe B : Contribution et réalisation.....	40
5.4.1 Présentation de l’application Dynacad	40
5.4.2 Cas d’usage applicatifs de Dynacad (suite).....	41
5.4.3 Mise en œuvre de la solution.....	41
5.4.4 Autres réalisations	45

Liste des figures

Figure 1: Logo Alteca	10
Figure 2: Répartition du chiffre d'affaires selon les secteurs	11
Figure 3: Quelques client d'Alteca par secteur.....	11
Figure 4: Domaines d'activités d'Alteca	12
Figure 5: Répartition des agences et des centres de services d'Alteca	13
Figure 6: Logo Carrefour	14
Figure 7: Cible fonctionnelle de Phénix.....	17
Figure 8: Architecture de Phénix	18
Figure 9: Architecture de Phénix après la migration.....	18
Figure 10: Cas d'usage, utilisation d'un moteur de recherche sur données non-structurées	19
Figure 11: Workflow d'une tâche Phénix dans Jira.....	20
Figure 12: Temps réel et Batch PHENIX.....	22
Figure 13: Architecture du traitement de recherche multicritère	26
Figure 14: Architecture Phénix avant la migration	34
Figure 15: Architecture Phénix	36
Figure 16: Processus d'usage de Dynacad.....	40
Figure 17: IHM de l'application web Dynacad	41
Figure 18: Exemple de carte JIRA pour l'ajout des informations de la dernière commande ...	42
Figure 19: Exemple de carte de MEP.....	42
Figure 20: Dynacad avec le niveau de stock de chaque article	43
Figure 21: Barre de filtre sur la date de dernière commande	44
Figure 22: Résultat du filtre sur la date de dernière commande.....	44
Figure 23: Recherche textuelle avec le libellé d'un produit	45
Figure 24: Carte Jira de migration du job "finance-store-invoice"	46
Figure 25: Carte Jira pour la résolution du problème de calcul de la cartographie.....	46

Liste des abréviations

Abréviation	Désignation
API	A pplication P rogramming I nterface
BQ	B ig Q uery
ESN	E ntreprise de S ervice du N umérique
GCP	G oogle C loud P latform
GCS	G oogle C loud S torage
HDFS	H adoop D istributed F ile S ystem
HTTP	H yper T ext T ransfer P rotocol
IDE	I ntegrated D evelopment E nvironment
IHM	I nterface H omme- M achine
K8S	K ubernete S
MEP	M ise E n P roduction
NoSQL	N ot O nly S QL
PO	P roduct O wner
SF	S oftware F actory
SFTP	S SH F ile T ransfer P rotocol
WS	W eb S ervice

Introduction générale

Face à une intensification de la concurrence dans le domaine du Retail, caractérisée par des changements rapides des technologies, d'investissements accrus, de progrès de compétitivité, mais aussi d'une pression soutenue sur les prix de la part des grands groupes du secteur Retail, le Groupe Carrefour exerce une stratégie de croissance axée sur la recherche et l'innovation plus précisément sur une transformation majeure de ses systèmes et technologies d'information. En effet, ces nouvelles technologies de l'information peuvent avoir un impact majeur sur la stratégie de commercialisation et le volume des ventes. Cet impact, rendu possible par des systèmes d'information sophistiqués sont bien exploités par les concurrents de Carrefour. Tout ceci constitue un vrai défi pour Carrefour et le pousse à y faire face.

Dans ce sens, le Groupe fait appel depuis des années maintenant, à des prestataires et partenaires pour le soutien et le développement de l'ensemble de ses applications en matière de Supply Chain, d'assortiment, etc.

De ce fait, en faisant partie de l'équipe « Software Factory », mon projet de fin d'études a été réalisé au sein d'Alteca Lyon qui intervient en tant qu'entreprise de services numériques en qualité de prestataire du projet « Phénix » de Carrefour. Ce projet novateur, a pour but de traiter les différentes données qui transitent entre les différentes applications utilisées par le groupe Carrefour, ma mission donc était d'exploiter et de trier les données du Data Lake afin de consolider les données des différentes activités du groupe, et pouvoir par conséquent les exposer à travers des bases de données et des APIs afin que d'autres applications, Data Analystes ou Data Scientistes puissent les utiliser.

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

2.1.1 Présentation



Figure 1: Logo Alteca

Alteca est une entreprise française de services numériques (ESN). Elle a été créée en 1996 à Lyon par Jean Mougin, qui est son actuel Président Directeur Général. Il a décidé de créer cette entreprise après avoir commencé sa carrière dans un grand groupe informatique. Son objectif était de créer une société de services informatiques indépendante qui placerait ses collaborateurs au cœur de son développement. L'entreprise monte vite en puissance en acquérant au fil des années de nombreuses entreprises dans divers secteurs d'activités. Dédiée aux grands comptes, l'entreprise allie une vocation nationale tout en maintenant une proximité régionale appréciée par ses clients. Alteca se développe fortement et ouvre des agences régionales sur l'ensemble du territoire français dans une logique de proximité locale. Dans sa politique de croissance, plusieurs sociétés font l'objet d'acquisitions par le groupe Alteca, permettant de développer la taille de l'entreprise tout en intégrant des compétences complémentaires dans les équipes. En quelques chiffres Alteca c'est environ 650 collaborateurs, 26 années d'existence, et 8 agences à travers la France.

2.1.2 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.

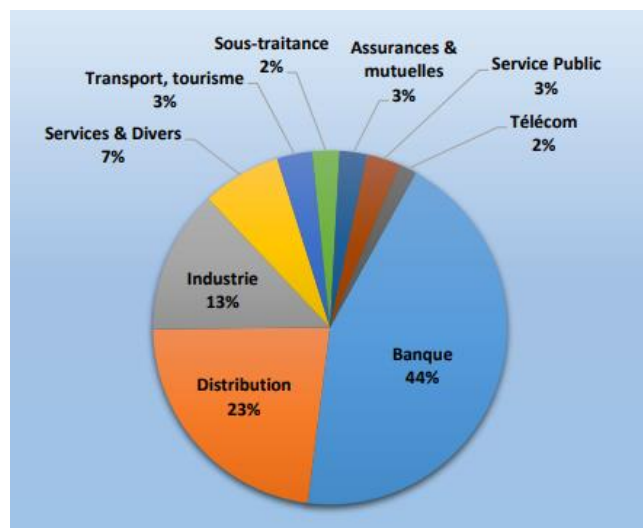


Figure 2: Répartition du chiffre d'affaires selon les secteurs

La clientèle d'Alteca est constituée d'entreprises d'importance majeure dans de multiples domaines comme le montre la figure suivante :



Figure 3: Quelques client d'Alteca par secteur

2.1.3 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.



Figure 4: Domaines d'activités d'Alteca

Au sein de chaque agence, il existe 6 pôles différents regroupant les compétences ensemble:

- ✚ **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.

2.1.4 Agences

Alteca a développé une présence nationale à travers 8 agences en France. Ce maillage géographique permet aux équipes d'évoluer dans des structures à taille humaine avec un management rapproché. La proximité régionale avec ses clients garantit une relation de confiance à long terme, tout en bénéficiant d'une coordination nationale.



Figure 5: Répartition des agences et des centres de services d'Alteca

2.1.5 Agence de Lyon et Siège Social

La première agence d'Alteca fut celle de Lyon, créée en 1996. Dirigée par Pascal Ferrary, elle est basée dans le 6^{ème} arrondissement près du parc de la tête d'or. On y retrouve l'ensemble des activités d'ingénierie informatique. Les activités projets se font au sein du centre de services dans les locaux d'Alteca ou en assistance technique en détachement chez les clients. Les principaux secteurs d'activité sont la grande distribution, l'industrie et le secteur bancaire. Alteca dispose sur Lyon deux sites géographiques différents pour mieux accueillir son effectif en terme de ressources humaines et également son effectif matériel notamment sur le site du 6^{ème} arrondissement, site dédié exclusivement aux activités d'Alteca qui touchent la mise en place, l'installation ou encore la maintenance de matériels informatiques. L'agence de Lyon rapporte aux acteurs de la région Auvergne-Rhône Alpes l'ensemble des offres d'Alteca. Mon stage s'est déroulé donc au sein de cette agence au sein du pôle data.

2.2 Présentation du client : Carrefour

2.2.1 Présentation



Figure 6: 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-discount, 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.

3

Cadre général du projet

Introduction

Dans ce chapitre, nous présenterons le projet dans lequel je suis intervenue, puis nous mettrons le point sur le déroulement du stage, nous expliquerons ensuite le projet et la méthodologie de travail ainsi que les différents environnements de travail.

3.1 Le Big Data, c'est quoi ?

L'explosion quantitative des données numériques a obligé les chercheurs à trouver de nouvelles manières de voir et d'analyser le monde. Il s'agit de découvrir de nouveaux ordres de grandeur concernant la capture, la recherche, le partage, le stockage, l'analyse et la présentation des données. Ainsi est né le « Big Data ».

Le Big Data est composé de jeux de données complexes, provenant essentiellement de nouvelles sources. Ces ensembles de données sont si volumineux qu'un logiciel de traitement de données traditionnel ne peut tout simplement pas les gérer.

L'utilisation du Big Data présente un certain nombre de caractéristiques. Ou plutôt, c'est une affaire de 5 V :

- ✚ La vitesse, très importante, de transmission des données ;
- ✚ Le volume, soit la taille des données dont l'origine peut être très diverse ;
- ✚ La variété, il peut s'agir de différents types de données (données structurées, semi-structurées et non structurées) ;
- ✚ La variabilité, faisant référence aux différentes manières dont le Big Data peut être utilisé et formaté ;
- ✚ La valeur ajoutée par ces données à l'entreprise.

Le Big Data permet aux entreprises de prendre des décisions plus efficaces et plus intelligentes. Son utilisation présente les avantages suivants :

- ✚ La réduction des coûts ;

- ✚ La création de produits et services améliorés ou nouveaux pour répondre aux différents besoins des entreprises ;
- ✚ La possibilité d'avoir des retours en temps réel ;
- ✚ Une meilleure connaissance du marché.

3.2 Présentation du projet

3.2.1 Problématique

Carrefour est un grand groupe qui utilise une multitude d'applications par plusieurs services pour le stock, les factures, etc. Ces applications génèrent un grand nombre de données au quotidien de sources différentes, celles-ci concernent les magasins, l'assortiment, les commandes, les tickets de caisse, les promotions, etc.

Les données transitent entre toutes ces applications, cela implique un problème de taille, d'origine de données, de temps de réponse, etc. Egalement, la récupération des données depuis différentes sources peut entraîner des problèmes de format.

3.2.2 Solution : Phénix, Plateforme Big Data de Carrefour

Présentation générale du programme Phénix

En tenant compte des différents problèmes que Carrefour rencontre lors du traitement de ces données, l'entreprise a donc décidé de créer une Plateforme Big Data qui tire des avantages du concept Big Data, son but est de regrouper toutes ces données au sein d'un même endroit et d'être un intermédiaire entre les sources de données et les applications qui en ont l'utilité. Cette plateforme est nommée Phénix.

Phénix permet de réaliser le stockage et le traitement de gros volumes de données. En assurant le traitement d'une quantité de données considérable, Phénix par sa puissance et sa capacité de stockage, autorise la prise en charge d'une quantité potentiellement infinie de tâches. Aujourd'hui, l'augmentation croissante et la variété des données ainsi que la diversification de leurs sources, imposent des outils de plus en plus perfectionnés pour assurer leur traitement. C'est là le principal atout d'une telle plateforme, la multitude d'outils qu'elle propose et qui offre un traitement rapide et fluide.

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 7 ci-dessous :

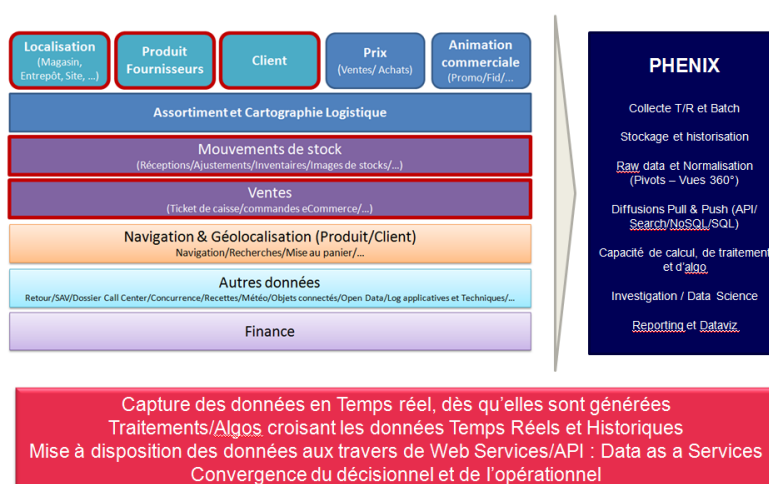


Figure 7: Cible fonctionnelle de Phénix

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 HDFS
- 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 HDFS. HDFS 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). Il est petit à petit remplacé par la suite Google GCS qui 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 fichiers txt ou csv), et comme précédemment on cherche à stocker ces données dans HDFS/ GCS.

La figure 8 représente l'architecture de Phénix, les détails sont représentées dans [l'annexe](#).

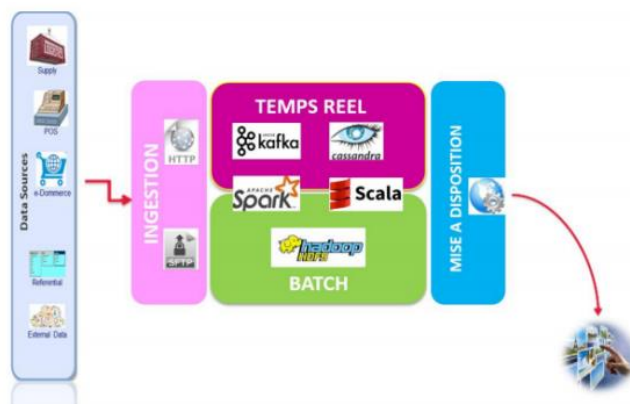


Figure 8: Architecture de Phénix

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, la suite Google permet de remplacer toutes les fonctionnalités Hadoop comme Hive par BigQuery ou bien HDFS par GCS. L'objectif de cette migration est de réduire le Time to market (Délai de mise sur le marché), réduire les coûts liés aux requêtes et stockages des informations. Mais aussi d'avoir un framework de chargement des fichiers présent sur GCS. Plus de détails dans [l'annexe](#).

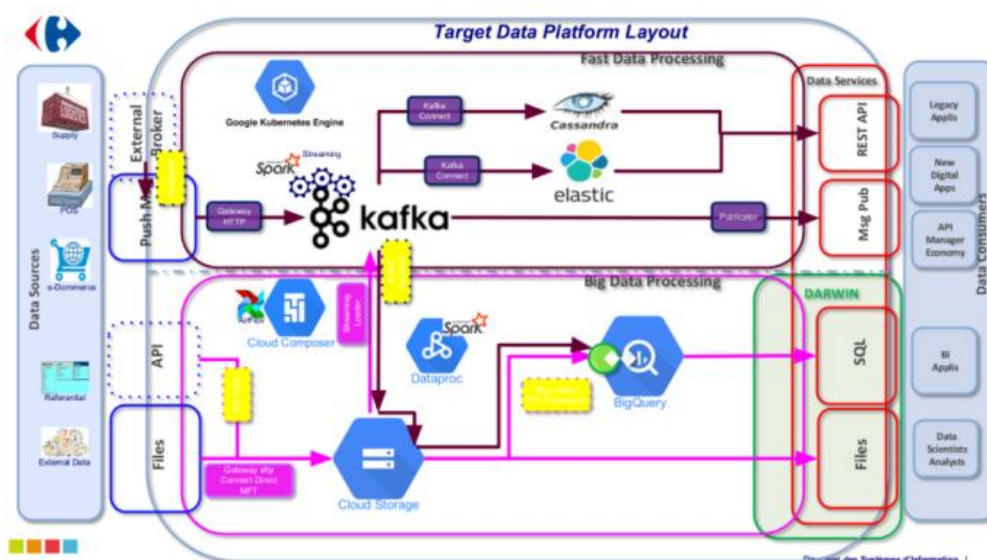


Figure 9: Architecture de Phénix après la migration

Dans le cadre de mon stage, j'ai intégré une équipe de consultants data ingénieurs qui travaillent sur le projet Phénix au sein de la Direction des Systèmes d'Informations de Carrefour

(DSI). Ce groupe composé de consultants proviennent de différentes entreprises de services du numérique (ESN) et d'internes à l'entreprise Carrefour. Ils réalisent plusieurs types de traitements sur les données dans le cadre de projets liés à des problématiques de Supply Chain et d'assortiment des magasins.

3.2.3 Exemple de cas d'usage de Phénix

- ✚ **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, historise 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 applications.

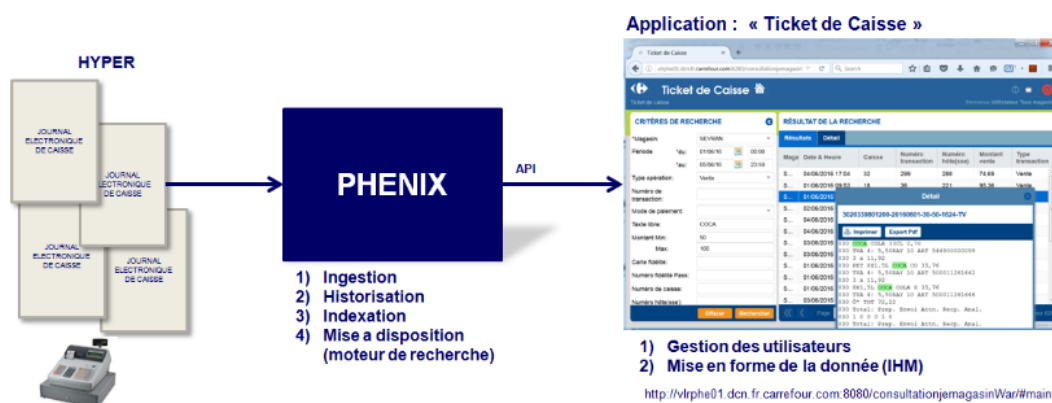


Figure 10: Cas d'usage, utilisation d'un moteur de recherche sur données non-structurées

3.3 Conduite du projet

Tout au long de la période de mon stage, je travaille deux jours par semaine chez le client Carrefour au Danica à la Part-Dieu et trois jours dans les locaux d'Alteca. 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.

3.3.1 Méthodologie agile : Kanban

L'organisation du travail au sein de l'équipe se base sur la méthodologie Kanban, qui fait partie des méthodes agiles. Kanban signifie littéralement « carte de signalisation » en langue japonaise. Le principe est le suivant : chaque développement à effectuer fait l'objet d'une carte. Les cartes attribuées à l'équipe apparaissent dans le Backlog (qui regroupe les différentes tâches à faire), puis chaque carte est d'abord chiffrée puis attribuée à un développeur.

3.3.2 Workflow d'une tâche Phénix

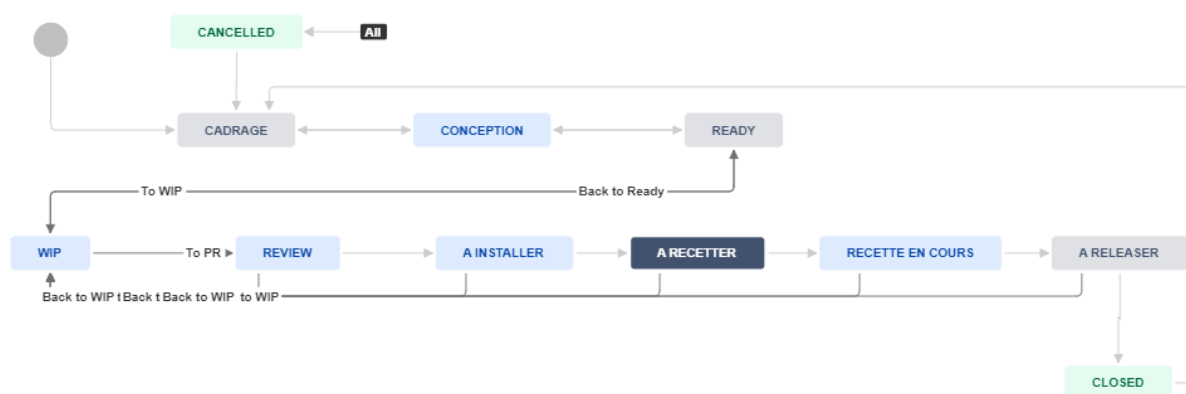


Figure 11: Workflow d'une tâche Phénix dans Jira

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 la 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.

3.3.3 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 où 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.

3.3.4 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.

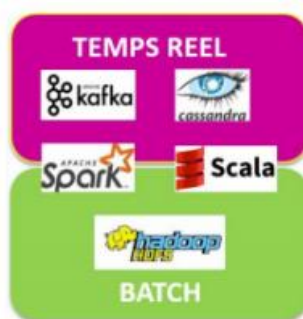


Figure 12: Temps réel et Batch PHENIX

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 artefacts.
- ✚ **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.

3.3.5 Environnements

Les environnements représentent en informatique des plateformes depuis lesquelles l'application est disponible. Les actions et leurs états peuvent différer selon l'environnement. Sur ce projet, il y a trois environnements distincts :

- ✚ **La recette** : ce sont les premiers environnements sur lesquels les nouvelles fonctionnalités ou corrections de bugs sont testées.
- ✚ **La pré-production** : est un environnement qui embarque strictement les mêmes fonctionnalités que la production, dans le même état. Cet environnement n'est pas disponible pour les utilisateurs finaux. Il permet au client de valider les modifications apportées avant de les passer en production.
- ✚ **La production** : représente l'environnement final, celui qu'utilisent les utilisateurs de Phénix.

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.

4

Réalisation et contribution

Introduction

Après avoir présenté le projet « Phénix » d'un point de vue technique et fonctionnel, ce chapitre sera consacré au travail réalisé dans le cadre de mon stage de fin d'études.

4.1 Formation et Onboarding sur « Phénix »

4.1.1 Formation et première tâche Phénix

La première semaine de mon arrivée était consacrée à la découverte de l'équipe du projet, à l'installation du poste, le lancement des différents services du projet ainsi qu'aux formations de la sécurité informatique exigées par Alteca et Carrefour.

Pour commencer, j'étais amenée à consulter quelques documentations Carrefour notamment un kit de bienvenue pour les nouveaux qui décrit l'environnement de travail, les normes de développement, l'organisation de l'équipe, les différents outils, etc.

Après cette semaine, ma contribution s'est basée dans un premier temps principalement sur la correction d'une anomalie remontée à la fois par les consultants fonctionnelles de notre équipe mais aussi par les POs du projet. Mon Scrum Master m'avait affectée une première carte Jira pour la résolution d'une anomalie avec des évolutions mineures afin de prendre en main le projet et de monter en compétences techniques et fonctionnelles.

Cette anomalie concerne deux jobs online, lorsque les jobs étaient démarrés à l'année Y, le job créait en interne une file de traitement pour l'année Y et les X années dans le passé qu'il doit continuer d'accepter et traiter (X étant un paramétrable dans la config du job). Si le job doit traiter 3 années glissantes et qu'il était démarré en 2021, il créait les fils de traitement nécessaires pour traiter les messages Kafka pour 2021, 2020 et 2019 au démarrage. Et lorsqu'on passait à l'année suivante (quand on est passé en 2022) aucune file pour traiter les 2022 n'était créé et la file de 2019 n'était pas supprimée non plus alors que l'on ne doit plus l'accepter.

Pour régler rapidement le souci en production, les jobs ont simplement été relancés. Ma mission était de régler le problème définitivement pour ne pas qu'il se reproduit lors des

prochains changements d'année, la correction a été implémentée et les années de traitements sont maintenant réellement "glissantes".

4.1.2 Bilan

Dans un premier temps, j'ai été affecté à une tâche qui m'avait permis de découvrir l'environnement de travail Phénix, de m'intégrer dans l'équipe en assistant au daily et de communiquer avec les membres de l'équipe. Elle m'a permis aussi de découvrir le workflow d'une tâche Phénix allant du cadrage et arrivant à la mise en production. A travers cette tâche, j'ai pu apporter une première solution à l'une des anomalies de Phénix.

Commencer directement avec une tâche sans passer par des formations théoriques est une méthode de formation par l'action qui m'a permis d'apprendre par la pratique (Learning by doing).

Le début du stage était bon malgré les gros soucis réseaux et configuration VPN Carrefour rencontrés sur le site Alteca qui a ralenti l'installation et la mise en service de mon poste de travail.

4.2 Service de recherche multicritère sur l'assortiment commandable

Le développement d'une solution informatique nécessite au préalable une étude détaillée de l'existant pour bien saisir l'architecture courante, une explication des besoins fonctionnels, non-fonctionnels et techniques des différents acteurs est aussi nécessaire afin d'avoir une analyse exhaustive des attentes du client.

4.2.1 Description générale de l'existant

Carrefour a un grand réseau de magasins de proximité, et chacun de ces magasins a besoin d'accéder aux informations de Cadencier Dynamique afin de préparer leurs commandes.

Un cadencier est un catalogue des produits à l'assortiment permettant aux magasins de passer une commande (suivi des quantités précédemment commandées, des stocks, des prix, etc.)

Et pour pouvoir préparer les commandes, les utilisateurs du Cadencier Dynamique ont besoin de chercher selon des critères différents, par exemple chercher un produit qui n'a jamais été commandé, consulter le stock, chercher par marque ou encore par libellé.

Une recherche classique dans une énorme base de données coûte chère, ce qui rend la tâche difficile aux personnels et cause des pertes pour Carrefour.

« Le Service de recherche multicritère sur l'assortiment commandable » est la solution à la problématique évoquée, c'est est un Web Service qui fait partie de la Plateforme Big Data Phénix, ce service a pour but de :

- ✚ Permettre la recherche de données référentiel produit filtrée à l'assortiment d'un magasin Carrefour :
 - Grâce à une recherche multicritère.
 - Avec la possibilité de regroupements.
 - Recherche en mode full-text pour les champs descriptifs: le libellé du produit, sa marque.
- ✚ Permettre la recherche de produits en fonction de la structure PROXI
 - Il est possible de spécifier des facettes pour compter et regrouper les réponses.
 - Il est possible d'ajouter des filtres pour limiter les réponses.
 - Il est possible de spécifier dans la requête, la liste des attributs produit à renvoyer.
 - Il est possible de trier et de paginer les réponses.

Et ceci grâce à **ElasticSearch** qui a une capacité de recherche avancée et une rapidité de requêtage. En effet, les données insérées dans ElasticSearch depuis BQ sont consultables dès que celles-ci sont stockées et indexées dans la base. La vitesse pendant laquelle ce traitement est exécuté est moins d'une seconde malgré les 98 millions de documents indexés.

Ce Web Service est consommé par l'application Web Dynacad qui n'est autre que le Cadencier Dynamique (Plus de détails dans [l'annexe](#)).

4.2.2 Architecture

Pour y arriver, ces données passent par plusieurs traitements. Tout d'abord, les données sont récupérées dans le Cloud depuis les tables Big Query concernées, ces données sont traitées et utilisées pour faire des calculs afin d'obtenir d'autres données sollicitées par les utilisateurs de Dynacad.

Ces données dans leurs états finaux, sont indexées dans ElasticSearch et exposées via un web service pour la recherche multicritère côté Dynacad.

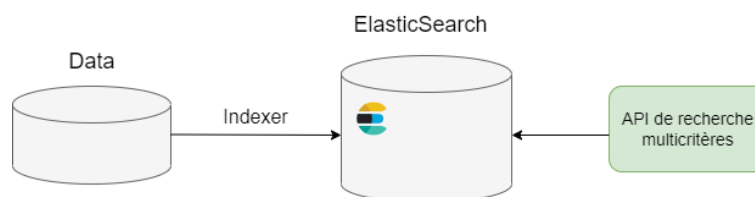


Figure 13: Architecture du traitement de recherche multicritère

4.2.3 Exemple de cas d'usage applicatif

- La recherche de produits à partir d'une chaîne de caractères

L'application cliente passe en paramètre une requête de type "recherche Google". C'est à dire des chaînes de caractères saisies par un utilisateur et qui correspondent à une partie du libellé du produit.

Le phénomène d'auto-complétion perçu par l'utilisateur vient en fait d'une succession d'appels au service, au fur et à mesure que l'utilisateur entre des caractères dans son IHM.

Plus de cas d'usage sont présentés dans [l'annexe](#).

4.2.4 Analyse du besoin

La solution s'adresse en premier lieu aux utilisateurs sièges, aux utilisateurs magasin et aux utilisateurs support gérant le processus des commandes Carrefour, ainsi qu'aux décideurs de l'entreprise. Il s'agit dans l'ensemble des directeurs opérationnels, marketing, méthode et qualité, les coordinateurs enseignes et les employés des magasins proximité. Ils doivent en effet suivre et analyser l'activité.

Cependant, il reste des briques manquantes au service, en effet, il y a toujours un besoin d'enrichissement de la donnée exprimé par le métier afin de renforcer la prise de décision.

Ces briques s'expriment en besoins fonctionnels et en besoins non fonctionnels :

Besoins fonctionnels

Les besoins fonctionnels sont l'expression de ce que le produit ou le service délivré par le projet devrait être ou faire, ils caractérisent la solution proposée. Ces besoins peuvent être présentés comme suit :

Enrichissement du WS multicritère sur l'assortiment commandable

Ce besoin concerne l'ajout de fonctionnalités pour le WS multicritère :

- Mode d'approvisionnement d'un article du cadencier: cette fonctionnalité était très attendue par le client, elle permettra de voir si le flux d'un article est stocké, tendu ou alloti, on doit également pouvoir filtrer dessus dans l'application pour que les utilisateurs puissent voir rapidement les articles qui ont un stock tendu par exemple.
- Niveau du stock de l'entrepôt : cette fonctionnalité permettra de voir directement la jauge de stock d'un article avec des voyants côté IHM. Cette jauge de stock se calcule en se basant sur le stock disponible, sur la prévision du stock et sur le mode d'approvisionnement.



- Informations sur la dernière commande : cette fonctionnalité permettra d'avoir les informations de la dernière commande effectuée, ces informations sont la date de création de la demande, la quantité commandée, la date de livraison prévue, le temps de la préparation de la commande, etc. En plus, il faut que les utilisateurs puissent filtrer sur la date de dernière commande, ils doivent pouvoir voir la commande d'une date précise, les articles qui n'ont pas été commandés, ou bien encore les articles commandés entre deux dates.

Résolution des bugs du Service de recherche multicritère sur l'assortiment commandable

Dynacad avait des bugs qui affectent le bon fonctionnement de l'application, d'où la naissance des nouveaux correctifs à apporter au WS multicritère :

- Problème de la récupération du prix de vente d'un article à cause de l'utilisation de la mauvaise clé.
- La recherche ne fonctionne plus complètement dans Dynacad à cause d'un champ qui n'est plus alimenté, les utilisateurs n'arrivent plus à voir le libellé Marketing d'un article.

Besoins non fonctionnels

-  Conformité : l'ajout des fonctionnalités demandées et la correction des bugs doivent respecter les normes réglementaires, documentaires et pratiques exigés par Carrefour.
-  Le rendement et l'efficacité : durant la réalisation, il faut penser performance et optimisation des ressources.

4.2.5 Mise en œuvre de la solution

La phase de la mise en œuvre de la solution a suivi le Worklow d'une tâche Phénix déjà présenté. Pour chaque besoin, le PO du WS multicritère rédige une carte Jira où il détaille le besoin. Dans ce cas, les fonctionnalités à implémenter et les correctifs à faire étaient sous forme de cartes Jira séparées. ([Annexe : carte JIRA](#))

Méthodologie de travail suivie

- Conception : A la réception de chacune des cartes Jira, je dois me faire expliquer le besoin par le PO en charge afin de m'assurer que tout soit bien compris. Après avoir bien compris le besoin, il faut concevoir la solution en définissant les entrées et les sorties du traitement, les schémas et les croisements de données à mettre en œuvre le cas échéant.
- Développement de la solution : Une fois que c'est clair, je passe à la mise en place de la solution. Chaque fonctionnalité et chaque correctif ont été développés en suivant les bonnes pratiques exigées par Carrefour.

- Test : Chaque fonctionnalité et chaque méthode doit être testée avant d'être livrée, et ceci, par le biais des tests d'intégration pour tester les fonctionnalités en scala et les tests unitaires pour tester les méthodes. Pour pouvoir tester correctement, il faut créer un cas similaire au cas réel, car on ne se connecte pas directement à Big Query en local.
- La mise en recette : à la fin du développement, le travail effectué est installé en recette afin de pouvoir le tester dans un environnement proche de la production et sur les données issues directement de Big Query. Pour valider cette phase du développement, le PO est chargé de tester et de vérifier la bonne réalisation du travail et que ce qui a été livré en recette correspond à ce qui a été demandé dans la carte Jira. Pour certaines cartes techniques, l'aide du développeur est requise, j'ai donc participé aux tests de recette et nous avons validé ensemble mon développement.
- La mise en pré-production : Après la validation de la recette, le PO donne le feu vert pour passer à l'environnement de pré-production. Dans ce cas, le code doit être revu par d'autres développeur pour une éventuelle amélioration. Une nouvelle version du code est générée. Avant l'installation en pré-production, j'étais amené à rédiger des cartes MEP pour installer correctement la nouvelle version. Ces cartes mentionnent les prérequis à l'installation, ses étapes, les vérifications à faire après l'installation et les étapes du retour en arrière au cas où un problème est détecté avec la nouvelle version installée.
- La mise en production : Après la validation de la pré-production, l'installation en production se fait en suivant exactement les étapes d'installation en pré-production mentionnées sur la même carte MEP. ([Annexe : carte MEP](#))

Réalisation

En suivant la méthodologie citée et en communiquant avec les PO, j'ai répondu aux différents besoins du client en respectant les spécifications fonctionnelles et techniques.

Aujourd'hui, tous mes développements sont en production et les utilisateurs de Dynacad profitent de ces nouvelles fonctionnalités et des bugs corrigés pour l'amélioration de leurs services et pour une bonne prise de décision.

- Mode d'approvisionnement d'un article du cadencier: les utilisateurs peuvent dorénavant consulter le mode d'approvisionnement d'un article et filtrer dessus, l'information est donc accessible rapidement grâce à l'utilisation d'ElasticSearch.
- Niveau du stock de l'entrepôt : les utilisateurs peuvent donc regarder directement le niveau de stock d'un article précis, ils peuvent filtrer dessus et anticiper une éventuelle rupture. ([Annexe : présentation dans Dynacad](#))
- Informations sur la dernière commande : A travers cette fonctionnalité, les utilisateurs peuvent donc avoir des informations sur les commandes et surtout filtrer sur les articles qui n'ont pas été commandés afin de prendre la bonne décision. ([Annexe : présentation dans Dynacad](#))

- Problème de la récupération du prix de vente a été résolu en renseignant la bonne clé de jointure, actuellement, les utilisateurs peuvent avoir le prix de vente d'un article sans problème.
- Le champ du libellé Marketing est de nouveau alimenté avec les bonnes valeurs ce qui permettra aux utilisateurs de retrouver tous leurs articles lors d'une recherche textuelle. ([Annexe : présentation dans Dynacad](#))

4.3 Autres réalisations

A côté de ce qui a été cité, j'ai eu l'occasion de travailler sur d'autres sujets et d'autres traitements de Phénix, durant leurs réalisations, j'ai suivi la même méthodologie de travail tout en respectant les besoins fonctionnels et techniques requis par le métier. Pour plus de détails, merci de consulter [l'annexe](#).

Conclusion

Au niveau de ce chapitre, il était question de décrire les traitements auxquels j'ai participé et que j'ai fait évoluer. Il était question aussi de mettre le point sur la méthodologie de travail suivie, et à la fin de ce chapitre, j'ai présenté les réalisations et leurs valeurs ajoutées côté métier.

Bilan et conclusion

5.1 Bilan professionnel et personnel

À l'heure du bilan, je dirais que mon stage a été une expérience pleine d'apprentissage et de la mise en pratique des savoirs acquis lors de ma formation à l'université. J'ai eu la chance de travailler sur plusieurs gros traitements du projet Phénix de Carrefour, notamment le WS multicritère de l'application Dynacad. Mon maître de stage a été très pédagogue puisqu'il a pris le temps de m'expliquer l'ensemble des missions que je devais effectuer en amont. Il est resté avec moi pour me guider au début des missions dont il m'a donné la responsabilité. Il m'a permis de pouvoir prendre des initiatives et de pouvoir donc proposer mes idées.

Alors que je n'étais que stagiaire, l'équipe Phénix m'a donné l'opportunité d'être écoutée et m'a déléguée des missions de plus en plus complexes. Dans ce contexte, j'ai communiqué avec les différents Product Owner, j'ai pu recueillir leurs besoins et travailler dessus, ainsi que participer aux tests de recette de mes développements et la résolution des différents bugs rencontrés lors de la mise en production.

J'ai appris à travailler sous stress et à gérer le temps puisque certaines tâches que l'on m'a confiées étaient urgentes et très attendues par le métier.

Tout au long du stage, je n'ai cessé de monter en compétences, et ceci, grâce à mon maître de stage, son suivi quotidien, ainsi que ces évaluations qui m'ont permis de me concentrer sur mes points forts et fixer des objectifs intermédiaires et concrets.

Les missions que Carrefour me proposaient étaient exactement ce dont j'avais besoin pour appliquer mes connaissances. Grâce à ce stage, j'ai développé mes capacités organisationnelles et mes compétences techniques en Big Data, un domaine qui me tient à cœur et dans lequel j'aimerais continuer mon apprentissage.

J'ai également appris beaucoup sur la communication d'entreprise puisque j'ai dû travailler aux côtés de plusieurs collègues et au sein d'une équipe de jeunes innovateurs. J'ai d'ailleurs réalisé par la même occasion que la communication entre les différentes équipes et divisions de l'entreprise était un facteur clé de la réussite d'un projet et, in fine, de l'entreprise en elle-même.

Ce stage a été pour moi une confirmation de choix de carrière. Il est évident qu'avec ce stage et dans cet environnement agréable, cette réalisation dont je pourrai faire état dans mon curriculum, me rend plus crédible dans le domaine qui continue à me passionner, celui du Big Data. Altea et son client Carrefour sont très satisfaits du travail que j'ai réalisé, j'intègre donc

l'entreprise en CDI à l'issu du stage et je continuerai de travailler pour le compte du client Carrefour.

5.2 Conclusion

Par le biais de ce rapport, j'ai présenté le travail effectué dans le cadre du projets « Phénix » auquel j'ai participé durant mon stage de fin d'étude dans le cadre de ma formation en Master 2 TIW à l'université de Lyon 1.

Au cours de ce stage, j'ai pu découvrir le métier d'ingénieur data sur les parties fonctionnelles et techniques et comprendre d'une manière globale le fonctionnement des ESN ainsi que les difficultés qu'ils rencontrent en satisfaisant les exigences de leurs clients.

Ma mission durant ce stage de fin d'étude a consisté en la montée en compétences sur les différents outils et langages attendus par Carrefour, et à la participation au développement de nouvelles fonctionnalités Phénix afin de satisfaire le besoin du client.

Concernant la conduite du projet, nous avons adopté la méthode Kanban. Cette méthodologie agile nous a permis de bien gérer ce projet à travers l'application de ses règles et ses bonnes pratiques proposées notamment les « Daily Meeting » qui nous ont permis de connaître l'état d'avancement et les blocages des collaborateurs, les réunions d'entraide ou encore la rétrospective avec les membres de l'équipe afin de discuter les points favorables et ceux à améliorer durant les prochains jours.

Pour conclure, ce stage a été pour moi une expérience professionnelle très enrichissante. Sur le plan technique, il m'a permis d'acquérir de nouvelles connaissances, tout en acquérant les bonnes pratiques de codage et la qualité du code. Au niveau relationnel, j'ai dû faire preuve d'initiative, développer mon sens de responsabilité et communiquer d'une manière continue. Enfin, sur le plan personnel, j'ai développé mon autonomie, ma créativité et ma capacité à travailler en équipe et à faire preuve de patience et de persévérance.

Webographie

- **Pages Web :**

[1] : KANBAN Methodology (Agile methodology), [Méthodologie Kanban](#) , consulté le 02.06.2022.

[2] : Chiffres du groupe Carrefour, [Groupe Carrefour — Wikipédia \(wikipedia.org\)](#), consulté le 02.08.2022.

[3] : Histoire de Carrefour, [Histoire | Groupe Carrefour](#), consulté le 02.08.2022

[4] : Google Cloud Dataproc, [Google Cloud Dataproc](#) , consulté le 02.08.2022

5.3 Annexe A : Cadre général du projet Phénix

5.3.1 Architecture Phénix avant la migration

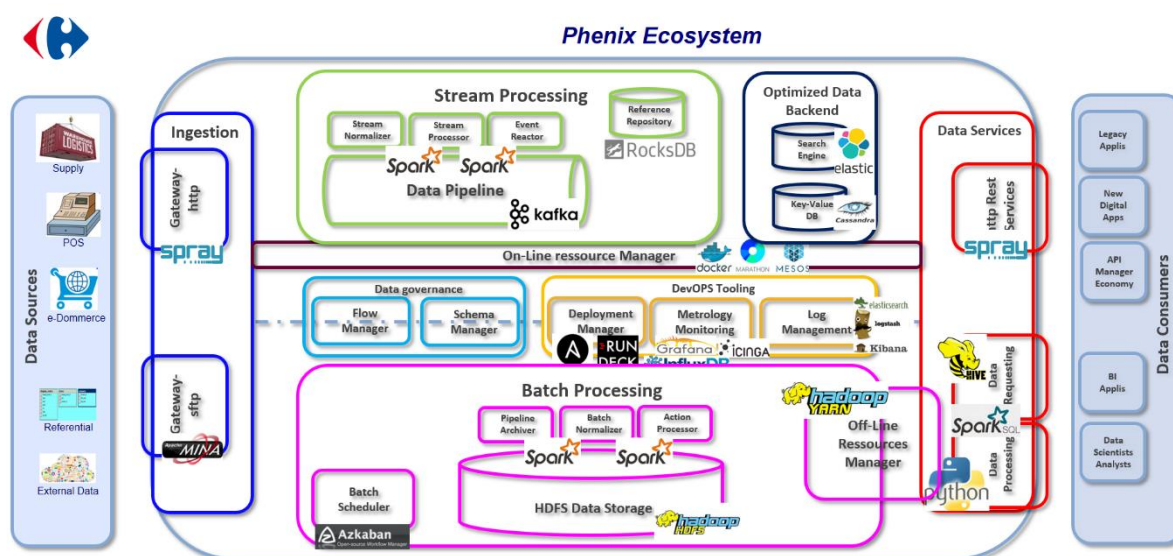


Figure 14: Architecture Phénix avant la migration

Phenix est un système opérationnel centré sur les données, son objectif principal est de capturer des événements en temps quasi réel à partir de sources de données, les nettoyer et les normaliser pour offrir une utilisation efficace aux consommateurs de données.

Les passerelles (HHTTP pour les messages et SFTP pour les fichiers) dans la zone d'ingestion contrôlent toutes les entrées d'événements.

Le traitement de flux est utilisé pour traiter les événements dans NRT : nettoyer, enrichir, normaliser les données.

Le Backend de données optimisé stocke ces données dans des « Business Views » pour pouvoir servir efficacement. Ces « Business Views » sont exposées via l'API REST dans la zone « Data Service ».

Le gestionnaire de ressources en ligne assure la multi-location du système. Les fichiers arrivent dans la passerelle SFTP, alors qu'ils sont directement écrits dans le HDFS, un déclencheur d'événement est enregistré dans le pipeline de données. Le réacteur d'événements démarre les actions planifiées, le travail de normalisation et/ou une action spécifique.

Le lignage des données, le schéma, le SLA sont décrits dans « data governance components » qui gèrent tous les messages entrants et sortants de Phénix. La sérialisation Avro est utilisée pour stocker toutes les données.

Les outils DevOps permettent une installation entièrement automatique, la surveillance de chaque composant du système, la collecte de métriques sur chaque processus et une puissante analyse des logs.

La plupart des composants sont open-source et proviennent de la communauté Apache.

5.3.2 Migration Dataproc

Dataproc est un service géré Apache Spark et Apache Hadoop qui permet de bénéficier d'outils de données Open Source pour le traitement par lot, l'émission de requêtes, le streaming et le machine learning. L'automatisation Dataproc permet de créer des clusters rapidement, de les gérer facilement et de faire des économies en désactivant ceux qu'on n'utilise plus. On consacre moins de temps et d'argent aux fonctions d'administration, ce qui permet de se concentrer sur les tâches et les données. [\[4\]](#)

Dataproc est une fonctionnalité de GCP permettant pour le cas qui nous intéresse ici de créer des clusters Spark à la demande, de lancer des jobs Spark sur ces clusters puis de supprimer ces clusters, le double avantage étant que les machines composant les clusters ne sont facturées que durant l'existence du cluster, soit quelques minutes à quelques heures par jour au lieu de 24 heures sur 24, et les exécuteurs Spark dans chaque cluster ne sont pas partagés avec les autres clusters Dataproc, ce qui nous garantit que les jobs auront toutes les ressources prévues à leur disposition sans être en compétition avec d'autres jobs qui tourneraient au même moment mais dans d'autres clusters.

Un cluster Spark désigne un processus Spark driver et un processus Spark exécuteurs, chacun défini avec son type de machine définissant le nombre de processeurs (au sens machine multi-cœurs), sa mémoire vive et son espace de stockage servant par exemple à persister les Data Frame/RDD entre les étapes de calcul.

Important, sur le cluster Spark instancié sur les machines IAAS, les exécuteurs Spark sont en même temps les datanodes HDFS, ce qui permet aux exécuteurs Spark de lire et écrire de manière efficace sur HDFS car l'exécuteurs peut lire et écrire directement sur l'espace disque de la machine qui l'exécute.

Sur Dataproc le cluster est temporaire : Les machines instanciées pour les exécuteurs font également office de datanode HDFS, ce qui peut permettre d'y écrire des données d'un job à un autre par exemple, mais toute donnée écrite sur cet espace HDFS sera perdue lors de la destruction du cluster. Les jobs exécutés sur Dataproc doivent donc sauver leurs données sur d'autres espaces de stockage : GCS et Big Query en premier lieu, mais aussi les autres services de stockage notamment Cassandra et ElasticSearch sur IAAS.

5.3.3 Glossaire de Phénix

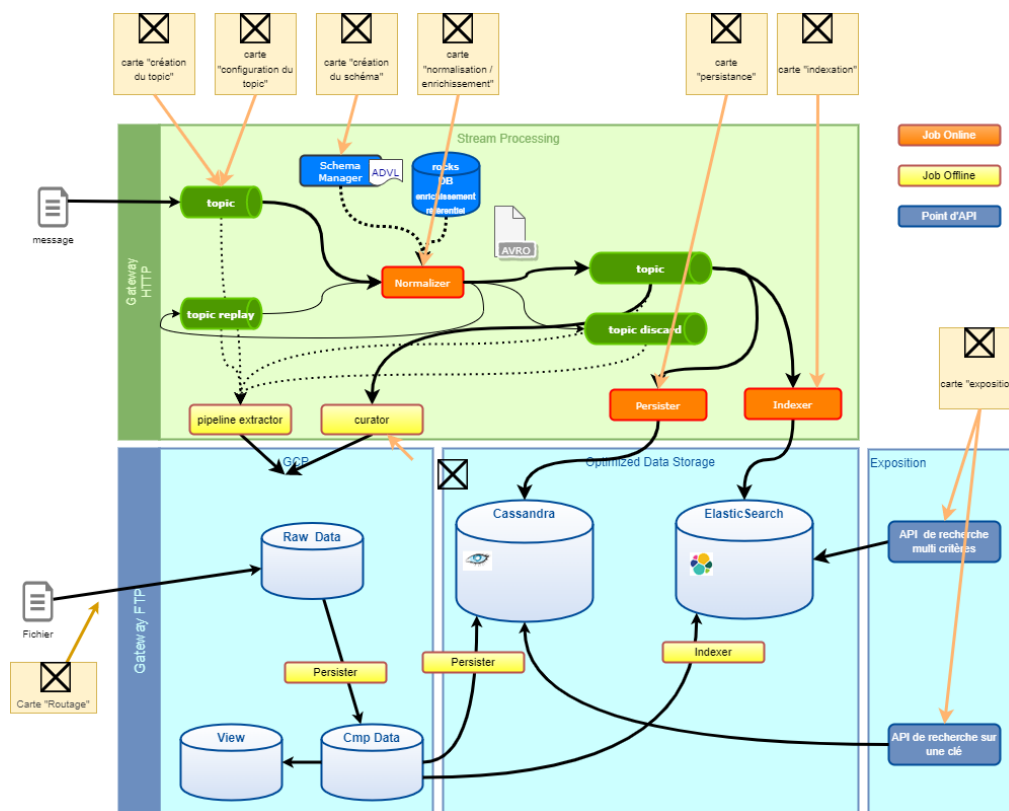


Figure 15: Architecture Phénix

Terme	est un	Signification
Ingestion	Traitement	Traitement général permettant de disposer d'une donnée dans Phénix de manière à ce qu'elle soit utilisable. En fonction de l'utilisation, l'ingestion nécessite un routage, une normalisation, une persistance, une indexation.
Alimentation	Traitement	Traitement général permettant de mettre à disposition de la donnée, quelle que soit la destination. Par exemple, une indexation est l'alimentation d'un Backend de type multicritère.
Normalisation (Normalizer)	Traitement	Traitement détaillé permettant de parser la donnée brute, la transformer vers un schéma de données défini, et potentiellement l'enrichir avec des données référentielles. Le traitement peut être : en temps réel (online) et va lire les données d'un topic et écrire les données normalisées dans un second topic ou en batch (offline) et va lire les données de fichiers et écrire le résultat sur HDFS.

Terme	est un	Signification
Indexation (Indexer)	Traitement	Traitement détaillé permettant de stocker de la donnée dans un Backend de type multicritère (ElasticSearch) pour l'exposition via des web services.
Persister	Traitement	Traitement détaillé permettant de stocker les données dans un Backend de type clé (Cassandra) pour exposition via des web services
Routage	Traitement	Traitement qui permet de recevoir des données dans la passerelle SFTP de Phénix et de le déposer dans un répertoire HDFS spécifique au pattern du fichier.
Injection	Traitement	Traitement détaillé qui permet d'injecter de la donnée dans un topic Kafka.
Archivage (Archiver)	Traitement	Traitement général qui comprend l'extraction et la curation d'un topic Kafka.
Extraction	Traitement	Traitement détaillé qui permet de sauvegarder le contenu d'un topic Kafka sur HDFS.
Curation (Curator)	Traitement	Traitement détaillé qui permet de construire une archive journalière des données à partir des extractions.
Schéma		<p>Métadonnées d'un objet utilisé dans un pivot. Le schéma est composé d'un Namespace, d'un nom et d'un contenu (liste des champs, description et types)</p> <p>Exemples : Le schéma Store permet de fournir le modèle pour la représentation d'une ligne du pivot Store.</p>
Pivot		Les pivots sont un stockage de la donnée dans l'HDFS de Phénix. Ce stockage de données dépend de son usage. Il peut s'agir d'un pivot pour un besoin spécifique ou d'un pivot de manière à fournir une représentation partagée pour tout Carrefour des objets.
Schema Manager	outil	<p>Outil permettant la gestion des schémas. Les schémas sont produits sous forme de JSON.</p> <p>Il existe également une interface Web de recherche de schéma.</p>

Terme	est un	Signification
Cassandra	outil	Base de données NoSQL, structurée en paires clé-valeur. Backend spécifique permettant d'accéder à la donnée selon une clé. Utilisée pour afficher un enregistrement à partir de son identifiant. Exemple : Afficher une transaction à partir de sa date, magasin et identifiant de transaction.
Elastic Search	outil	Moteur de recherche permettant l'indexation et la recherche de données. Backend spécifique permettant de faire des recherches multicritères. Utilisé dans notre cas pour les recherches multicritères comme celle des magasins et de l'assortiment.
HDFS	outil	Système de fichiers distribués de Hadoop.
Gateway SFTP	module	Module du Framework Phénix permettant de gérer la réception de fichiers en SFTP. (Voir routage)
Gateway HTTP	module	Module du Framework Phénix permettant de gérer la réception de messages en temps réel via un appel REST. Les messages reçus sont postés sur Kafka, dans un topic en fonction de la configuration.
Backend	type d'outil	Dans le contexte Phénix, un Backend désigne une zone de stockage optimisé (Optimized Data Storage), Cassandra ou ElasticSearch ou autre (RocksDB a déjà été utilisée à cette fin) pour exposition des données via un web service.
Offline	module	Dans le contexte Phénix, Module logique du Framework Phénix où s'effectue les traitements Batch.
Online	module	Dans le contexte Phénix, Module logique du Framework Phénix où s'effectue les traitements temps réel.
Topic		Composant de Kafka, utilisé pour véhiculer un flux de données.
Topic Replay	topic	Type de topic qui permet de rejouer un traitement si le traitement précédent a échoué. Un job online dont le traitement échoue doit pousser la donnée, soit vers un topic Replay, soit vers un topic Discard.
Topic Discard	topic	Type de topic où sont poussées des données dont le traitement précédent a échoué, mais qui ne doivent pas être rejouées. Un job

Terme	est un	Signification
		online dont le traitement échoue doit pousser la donnée, soit vers un topic Replay, soit vers un topic Discard.
Pipeline		selon le contexte : un topic est un enchainement logique de topics et Normalizers.
Subject (passerelle d'ingestion temps réel)		Identifiant utilisé comme route pour la gateway-http. Le Subject est lié à un topic.
Swagger		Fichier normalisé décrivant le contrat d'interface des API REST.
Avro		Format de données lié à un schéma, permettant ainsi de décrire à la fois la structure de la donnée et la donnée
ORC		Format de fichier pour stocker des données orienté colonne. Un stockage orienté colonne permet des accès en lecture plus rapide que les formats Avro. Il est le format actuel de stockage pour ce type de données dans Phénix. Parquet est également un format de fichier orienté colonne.
Parquet		Format de fichier pour stocker des données orienté colonne. Un stockage orienté colonne permet des accès en lecture plus rapide que les formats Avro. ORC est également un format de fichier orienté colonne.
Queue		File d'attente.
Exposition		L'exposition est un terme générique pour désigner la manière dont des données sont mises à disposition hors de Phénix. Cela peut être un point d'API REST, une business view accessible via Hive, un export de données.
Raw_Data		Répertoire HDFS dans lequel sont stockées les données brutes.
Cmp_Data		Répertoire HDFS dans lequel sont stockées les données calculées (ancienne arborescence).

Terme	est un	Signification
YAML		Langage formel simple qui décrit les données et format de fichier, utilisé dans notre cas pour définir les Swaggers.
Azkaban	Outil	Outil permettant d'assurer l'ordonnancement de jobs.
Enrichissement	Traitement	Traitement détaillé réalisant l'enrichissement d'une donnée à partir d'un référentiel.
Prometheus	outil	Outil permettant de gérer les métriques remontées par les différents composants de la plateforme.
business view		Table Hive présentant une exposition de données. Les business view de Phénix s'appuient sur des données au format ORC et sont préfixées en "bv_".

5.4 Annexe B : Contribution et réalisation

5.4.1 Présentation de l'application Dynacad

Les utilisateurs ont besoin d'accéder aux informations de Cadencier Dynamique afin de préparer leurs commandes. Dynacad permet d'obtenir l'ensemble des informations liées à l'article (tarif, stock, actions commerciales, etc.)

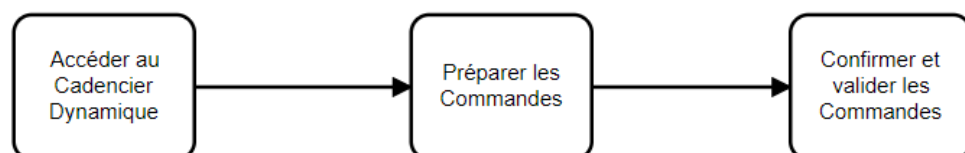


Figure 16: Processus d'usage de Dynacad

CADENCIER DYNAMIQUE Lundi 22 août 2022
CARREFOUR CONTACT - SEPTUILL
Reg comm.: PARIS CENTRE / Reg ass.: 017 - PARIS-CTRE
Anabel : 8677 / NoCII : 12517 / TMB : 63

Recherche par EAN, libellé, etc...

El Mahfoudi Sara

TOUT AJOUTER IMPRIMER / EXPORTER

Total 16633 - Page 5 (1000) - Éléments 10

01 - P. G. C. / 03 - D.P.H. / 101 - ENTRETIEN / 282 - PRODUITS D'ENTRETIEN / 282001 - LIQUIDES DE SYNTHÈSE

	PAIC HYG 3EN1 CIT750ML 01 - P. G. C. / 03 - D.P.H. / 101 - ENTRETIEN / 282 - PRODUITS D'ENTRETIEN / 282001 - LIQUIDES DE SYNTHÈSE / 8718951379770 0.75 L TAN 528 - ST GERMAIN ARP-EP5 Date dem.cde : 02/02/22	Potential de ventes hebdo. 5	Marque PAIC Type Marque Nationale	Prix d'achat Min de cde 12	Prix de vente PCB 12	Marge 10.63%	TVA 20%	Niv. assort. Niv. assort.	Stock Entr. 170815	Commandable le S L M E J V S D S+1 L M E J V S D	AJOUTER
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------	---------------------------------------------------	----------------------------------	----------------------------	-----------------	------------	------------------------------	-----------------------	--------------------------------------------------------	---------

01 - P. G. C. / 03 - D.P.H. / 101 - ENTRETIEN / 282 - PRODUITS D'ENTRETIEN / 282003 - NETTOYANTS MENAGERS

	500G BICARBONATE DE SOUDE B. 01 - P. G. C. / 03 - D.P.H. / 101 - ENTRETIEN / 282 - PRODUITS D'ENTRETIEN / 282003 - NETTOYANTS MENAGERS / 3383482297074 0.5 KG TAN 528 - ST GERMAIN ARP-EP5 Date dem.cde : 20/08/22	Potential de ventes hebdo. 2	Marque BIRCHIN Type Marque Nationale	Prix d'achat Min de cde 10	Prix de vente PCB 10	Marge 20.47%	TVA 20%	Niv. assort. Niv. assort.	Stock Entr. 094631	Commandable le S L M E J V S D S+1 L M E J V S D	AJOUTER
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------	------------------------------------------------------	----------------------------------	----------------------------	-----------------	------------	------------------------------	-----------------------	--------------------------------------------------------	---------

01 - P. G. C. / 03 - D.P.H. / 110 - HYGIENE PAPIER / 284 - HYGIENE / PAPIER / 284001 - PAPIER TOILETTE

	12RLX PAPHYG.COMPA.BLC 2P C. 01 - P. G. C. / 03 - D.P.H. / 110 - HYGIENE PAPIER / 284 - HYGIENE / PAPIER / 284001 - PAPIER TOILETTE / 3550070213627 TAN 528 - ST GERMAIN ARP-EP5 Date dem.cde : >	Potential de ventes hebdo. 5	Marque CRF CDM Type Marque MDO	Prix d'achat Min de cde 9	Prix de vente PCB 9	Marge 18.98%	TVA 20%	Niv. assort. Niv. assort.	Stock Entr. 563738	Commandable le S L M E J V S D S+1 L M E J V S D	AJOUTER
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------	------------------------------------------------	---------------------------------	---------------------------	-----------------	------------	------------------------------	-----------------------	--------------------------------------------------------	---------

01 - P. G. C. / 03 - D.P.H. / 111 - HYGIENE PARFUMERIE / 285 - HYGIENE CORPORELLE / 285007 - DEODORANTS PARFUMANTS

	200ML DEO ATO BLACK AXE 01 - P. G. C. / 03 - D.P.H. / 111 - HYGIENE PARFUMERIE / 285 - HYGIENE CORPORELLE / 285007 - DEODORANTS PA... / 8720181077678 0.2 L TAN 528 - ST GERMAIN ARP-EP5 Date dem.cde : >	Potential de ventes hebdo. 2	Marque AXE Type Marque Nationale	Prix d'achat Min de cde 6	Prix de vente PCB 6	Marge 16.0%	TVA 20%	Niv. assort. Niv. assort.	Stock Entr. 294203	Commandable le S L M E J V S D S+1 L M E J V S D	AJOUTER
--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------	--------------------------------------------------	---------------------------------	---------------------------	----------------	------------	------------------------------	-----------------------	--------------------------------------------------------	---------

01 - P. G. C. / 03 - D.P.H. / 111 - HYGIENE PARFUMERIE / 285 - HYGIENE CORPORELLE / 285008 - HYGIENE BUCCO-DENTAIRE

	75ML DENT SOIN COMPLET 24H ... 01 - P. G. C. / 03 - D.P.H. / 111 - HYGIENE PARFUMERIE / 285 - HYGIENE CORPORELLE / 285008 - HYGIENE BUCCO... / 30545631094620 0.075 L TAN 528 - ST GERMAIN ARP-EP5 Date dem.cde : 08/01/22	Potential de ventes hebdo. 3	Marque SENSODYNE Type Marque Nationale	Prix d'achat Min de cde 12	Prix de vente PCB 12	Marge 13.66%	TVA 20%	Niv. assort. Niv. assort.	Stock Entr. 390496	Commandable le S L M E J V S D S+1 L M E J V S D	AJOUTER
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------	--------------------------------------------------------	----------------------------------	----------------------------	-----------------	------------	------------------------------	-----------------------	--------------------------------------------------------	---------

Figure 17: IHM de l'application web Dynacad

5.4.2 Cas d'usage applicatifs de Dynacad (suite)

La recherche d'une fiche produit à partir d'un barcode

Pour récupérer la fiche produit d'un article dont le magasin est connu, il suffit d'appeler le service en passant en filtre la valeur du « barcode ». Un seul article sera retourné. Il est possible de choisir les champs retournés.

La recherche d'une fiche produit en fonction de l'itemKey

Pour récupérer la fiche produit d'un article dont l'itemKey est connu, il suffit d'appeler le service en passant en filtre la valeur de l'itemKey sur le champ « itemKey ». Un seul article sera retourné. Il est possible de choisir les champs retournés.

5.4.3 Mise en œuvre de la solution

Cartes JIRA

Pour l'ensemble des missions qu'on m'a confiées, elles étaient sous forme de carte JIRA, la figure 18 représente un exemple d'une carte JIRA.

PHENIX / PHX-9593
[DYN] Màj de l'API assortiment commandable - ajout d'infos sur la dernière commande

Details

- Type: User Story
- Priority: Minor
- Component/s: None
- Status: CLOSED (View Workflow)
- Resolution: Done
- Security Level: Phenix Team only (Issues only visible to members of the Phenix Team)
- Labels: DYN
- Epic Link: Enrichissement WS
- Requirement Status: UNCOVERED
- Estimation dev: 5

Description

Partie API

Ajout du champ date de dernière commande de l'article dans le WS multicritère assortiment.
Pouvoir filtrer sur cette date dans Dynacat.
Laisser vide s'il n'y a jamais eu de commande sur cet article.
(En lien avec la carte DYN-36 Gamme active)

cf spec <https://confluence.agilefabric.fr/carrefour.com/pages/viewpage.action?pageId=103722416>
Prévoir un facet pour pouvoir filtrer sur cette date de dernière commande.

Cas d'utilisation :

Je veux pouvoir récupérer tous les articles dont la date de **creationDatetime** (date de création de la commande) est

- > égale à une date donnée.
- > comprise entre 2 dates
- > inférieure à x mois
- > n'ont aucune date de commande (articles n'ayant jamais été commandés).

People

- Assignee: Sara EL MAHFOUDI
- Reporter: Celine JOTTRAS
- Votes: 0
- Watchers: 0

BigPicture - Skills

The issue has not been synced to the scope of any Box yet or you need additional permissions to access it.

Figure 18: Exemple de carte JIRA pour l'ajout des informations de la dernière commande

Cartes MEP

Afin d'installer en pré-production et en production les développements réalisés, il faut passer par des cartes MEP, la figure 19 représente un exemple d'une carte MEP que j'ai réalisée.

PHENIX / PHX-15333
MEP Màj de l'indexer assortiment commandable - Ajout / modification pour le Stock

Details

- Type: Content Deployment
- Priority: Minor
- Component/s: None
- Status: DEPLOYMENT VALID... (View Workflow)
- Resolution: Unresolved
- Security Level: Phenix Team only (Issues only visible to members of the Phenix Team)
- Labels: None
- Release components: assortiments-api, purch-assortments-indexer
- Installation prerequisites: Pré-requis à l'installation

Pré-requis à l'installation

Partie Indexer

Les PR suivantes sur **offers-purchase-assortment-indexer** doivent être mergées :

- https://bitbucket.agilefabric.fr/carrefour.com/projects/PHENIX/repos/offers-purchase-assortment-indexer/pull-requests/6/overview
- https://bitbucket.agilefabric.fr/carrefour.com/projects/PHENIX/repos/offers-purchase-assortment-indexer/pull-requests/7/overview

Le build stable 2.14.0 doit être effectué sur le repository **offers-purchase-assortment-indexer**:

- https://bitbucket.agilefabric.fr/carrefour.com/projects/PHENIX/repos/offers-purchase-assortment-indexer/browse?at=refs%2Ftags%2Fv2.14.0

Partie API

Les PRs suivantes "assortment" et "assortments-api" doivent être mergées :

- https://bitbucket.agilefabric.fr/carrefour.com/projects/PHENIX/repos/assortments/pull-requests/54/overview
- https://bitbucket.agilefabric.fr/carrefour.com/projects/PHENIX/repos/assortments-api/pull-requests/35/overview

La PR k8s d'installation de "assortments-api" en PREPROD doit être prête à être fusionnée:

- https://bitbucket.agilefabric.fr/carrefour.com/projects/PHENIX/repos/k8s/pull-requests/8902/overview

La PR k8s d'installation de "assortments-api" en PROD doit être prête à être fusionnée:

- https://bitbucket.agilefabric.fr/carrefour.com/projects/PHENIX/repos/k8s/pull-requests/8904/overview

Consignes d'installations

Partie Indexer

Jouer le rôle rundeck suivant:

repository	playbook	argansible	exemple	PREPROD	PROD
applications-deployment	applications-offline-jobs	--tags purch-assortments-indexer -e purch_assortments_indexer_version=2.14.0.stable	#38 939	#23 006	#8 535

Partie API

People

- Assignee: Kaouthar DRIIRA
- Reporter: Sara EL MAHFOUDI
- Related Developers: Jean Baptiste EVIEUX, Sara EL MAHFOUDI
- Related Product: Celine JOTTRAS, Nabgha MABROUK
- Owners: Celine JOTTRAS, Nabgha MABROUK
- Votes: 0
- Watchers: 0

BigPicture - Skills

The issue has not been synced to the scope of any Box yet or you need additional permissions to access it.

Dates

- Created: 18/05/2022 3:50 PM
- Updated: 13/06/2022 9:49 AM

Figure 19: Exemple de carte de MEP

Réalisations côté application web Dynacad

- Niveau du stock de l'entrepôt : les utilisateurs peuvent donc regarder directement le niveau de stock d'un article précis, ils peuvent filtrer dessus et anticiper une éventuelle rupture. La figure ci-dessous représente ce calcul de stock côté application web Dynacad.

The screenshot displays the Dynacad application interface. At the top, there is a header with the company name 'CADENCIER DYNAMIQUE', a date 'Lundi 22 août 2022', and a search bar. Below the header, a navigation bar contains buttons for 'TOUT AJOUTER' and 'IMPRIMER / EXPORTER'. The main content area shows a list of products, each with a detailed card. The cards are organized into sections based on product categories. Each product card includes the following information:

- Product Name and Details:** PAIC HYG 3EN1 CIT750ML, 500G BICARBONATE DE SOUDE B..., 12RLX PAPHYG.COMPA.BLC 2P C..., 200ML DEO ATO BLACK AXE.
- Potential de ventes hebdo.:** 5, 2, 5, 2.
- Marque:** PAIC, BRIOCHIN, CRF CDM, AXE.
- Prix d'achat / Prix de vente:** 12 / 12, 10 / 10, 9 / 9, 6 / 6.
- Marge:** 10.63%, 20.47%, 18.98%, 16.0%.
- TVA:** 20%, 20%, 20%, 20%.
- Stock:** 170815, 094631, 563738, 294203.
- Buttons:** 'AJOUTER' and 'A.JOUTER'.

The 'Stock' field for each product is highlighted with a red box, indicating the current stock level.

Figure 20: Dynacad avec le niveau de stock de chaque article

- Informations sur la dernière commande : A travers cette fonctionnalité, les utilisateurs peuvent donc avoir des informations sur les commandes et filtrer dessus. La figure 21 représente la barre de filtre de date de commande et la figure 22 représente le résultat du filtre.

CADENCIER DYNAMIQUE Lundi 22 août 2022
 CARREFOUR CONTACT - SEPTUIL
 Reg comm.: PARIS CENTRE / Reg ass.: 017 - PARIS-CTRE
 Anabel: 8677 / NoCII: 12517 / TMB: 63

Recherche par EAN, libellé, etc...

Mes filtres personnalisés (0)

Choix du Groupe de Familles

Famille

Choix de la Famille

Sous-Famille

Veuillez sélectionner une famille

Date de dernière commande
 * Uniquement commandes hors Mercat

Dans les 6 derniers mois

Entre 22/02/2022 et 22/08/2022

Type de Marque

☐ MDD

☐ Nationale

☐ Premier prix

Libellé Marque

BOUCHERIE LS INDUST / 450 - BOUCHERIE LS / 069 - AUTRES VIANDES LS / 069004 - GAMME BIO

Produit	Marque	Type Marque	Prix d'achat	Prix de vente	Marge	TVA	Commandable le
BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	22.78%	5,5%	617826
Date dem.cde: 13/08/22							
BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	27.36%	5,5%	451208
Date dem.cde: 09/08/22							
BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	28.25%	5,5%	072747
Date dem.cde: 19/08/22							
BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	27.8%	5,5%	072731
Date dem.cde: 19/08/22							
BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	21.27%	5,5%	617848
Date dem.cde: 19/08/22							

BOUCHERIE LS INDUST / 450 - BOUCHERIE LS / 060 - BOEUF LS / 060005 - UVC PREEMBALLEE SOUS VIDE

Produit	Marque	Type Marque	Prix d'achat	Prix de vente	Marge	TVA	Commandable le
BOUCHERIE LS INDUST/450 - BOUCHERIE LS/060 - BOEUF LS/060005 - UVC PREEMBALLEE SOU...	REFS.SANS MARQ...	Nationale	Min de cde 6	PCB 6	18.56%	5,5%	806765
Date dem.cde: 24/06/22							
BOUCHERIE LS INDUST/450 - BOUCHERIE LS/060 - BOEUF LS/060005 - UVC PREEMBALLEE SOU...	CHARAL				26.25%	5,5%	

Figure 21: Barre de filtre sur la date de dernière commande

CADENCIER DYNAMIQUE Lundi 22 août 2022
 CARREFOUR CONTACT - SEPTUIL
 Reg comm.: PARIS CENTRE / Reg ass.: 017 - PARIS-CTRE
 Anabel: 8677 / NoCII: 12517 / TMB: 63

Recherche par EAN, libellé, etc...

TOUT AJOUTER IMPRIMER / EXPORTER

Total 7732 - Page 1 (774) - Éléments 10

17 - BOUCHERIE VOLAILLE LS INDUST / 16 - BOUCHERIE LS INDUST / 450 - BOUCHERIE LS / 069 - AUTRES VIANDES LS / 069004 - GAMME BIO

Produit	Marque	Type Marque	Prix d'achat	Prix de vente	Marge	TVA	Commandable le
1100 STEAK**X2 CRF BIO	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	22.78%	5,5%	617826
17 - BOUCHERIE VOLAILLE LS INDUST/16 - BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/							
0.22 KG TAG 796 - BRIE PF HSPC MSEs							
Date dem.cde: 13/08/22							
KG PC COTE PREM X2 CRF BIO	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	27.36%	5,5%	451208
17 - BOUCHERIE VOLAILLE LS INDUST/16 - BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/							
0.25 KG TAG 796 - BRIE PF HSPC MSEs							
Date dem.cde: 09/08/22							
SKINP 2X125G STK HACHE BIO 5%	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	28.25%	5,5%	072747
17 - BOUCHERIE VOLAILLE LS INDUST/16 - BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/							
0.25 KG TAG 796 - BRIE PF HSPC MSEs							
Date dem.cde: 19/08/22							
SKINP 2X125G STK HACHE BIO 15%	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	27.8%	5,5%	072731
17 - BOUCHERIE VOLAILLE LS INDUST/16 - BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/							
0.25 KG TAG 796 - BRIE PF HSPC MSEs							
Date dem.cde: 19/08/22							
2000 FAUX FILET**X1 CRF BIO	CARREFOUR BIO	MDD	Min de cde 4	PCB 4	21.27%	5,5%	617848
17 - BOUCHERIE VOLAILLE LS INDUST/16 - BOUCHERIE LS INDUST/450 - BOUCHERIE LS/069 - AUTRES VIANDES LS/069004 - GAMME BIO/							
0.2 KG TAG 796 - BRIE PF HSPC MSEs							
Date dem.cde: 19/08/22							

17 - BOUCHERIE VOLAILLE LS INDUST / 16 - BOUCHERIE LS INDUST / 450 - BOUCHERIE LS / 060 - BOEUF LS / 060005 - UVC PREEMBALLEE SOUS VIDE

Produit	Marque	Type Marque	Prix d'achat	Prix de vente	Marge	TVA	Commandable le
KG V8 STEAK** X2 ANGUS	REFS.SANS MARQ...	Nationale	Min de cde 6	PCB 6	18.56%	5,5%	806765
17 - BOUCHERIE VOLAILLE LS INDUST/16 - BOUCHERIE LS INDUST/450 - BOEUF LS/060005 - UVC PREEMBALLEE SOU...							
0.4 KG TAG 796 - BRIE PF HSPC MSEs							
Date dem.cde: 24/06/22							
KG TOURNED.FLT 2X140G CHAR V...	CHARAL		33,15€	47,42€	26.25%	5,5%	
17 - BOUCHERIE VOLAILLE LS INDUST/16 - BOUCHERIE LS INDUST/450 - BOEUF LS/060005 - UVC PREEMBALLEE SOU...							

Figure 22: Résultat du filtre sur la date de dernière commande

- Le champ du libellé Marketing est de nouveau alimenté avec les bonnes valeurs ce qui permettra aux utilisateurs de retrouver tous leurs articles lors d'une recherche textuelle. La figure 23 représente un exemple de recherche en se basant sur le libellé Marketing d'un produit.

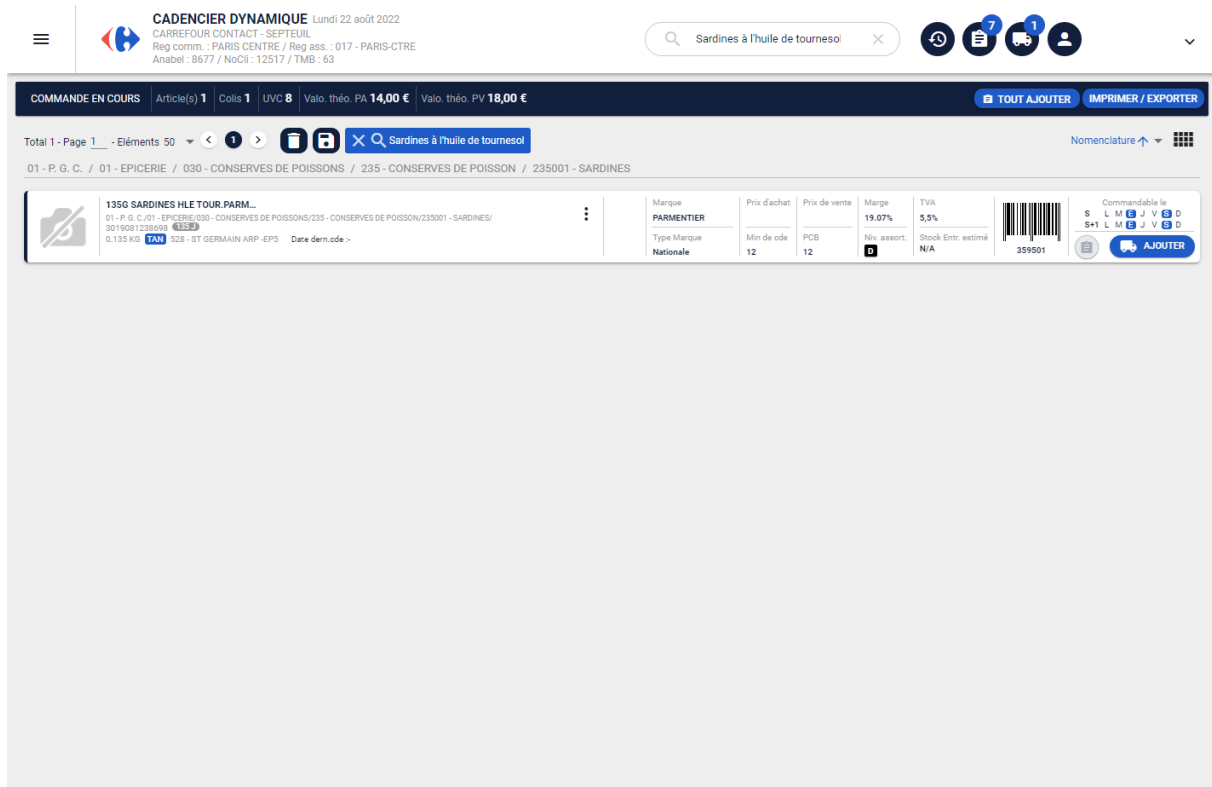


Figure 23: Recherche textuelle avec le libellé d'un produit

5.4.4 Autres réalisations

A côté de ce qui a été cité, j'ai eu l'occasion de travailler sur d'autres sujets et d'autres traitements de Phénix, durant leurs réalisations, j'ai suivi la même méthodologie de travail tout en respectant les besoins fonctionnels et techniques requis par le métier :

- ✚ **Migration Dataproc** : comme déjà cité, Carrefour est en train de faire une migration de tous ses jobs vers GCP, dans ce sens, on m'a affecté la migration du job « finance-store-invoice ». Je travaille toujours dessus.

Migration finances-store-invoice

Type: User Story
Priority: Minor
Component/s: None
Labels: None
Epic Link: Migration GCP Plateforme Entreprise
Requirement Status: UNCOVERED

Status: Unresolved
Resolution: Phenix Team only (issues only visible to members of the Phenix Team)
Security Level: Phenix Team only (issues only visible to members of the Phenix Team)

Description

Solution technique : migration dataprocc

Compte de service : à définir

INPUT HADOOP	INPUT GCP
\$phenixhadoop.hdfs.rootPath.computedData\Referential\referential.bv_warehouse	sites_referential.bv_warehouse
\$phenixhadoop.hdfs.rootPath.computedData\Referential\referential.store.v2.Store	sites_referential.bv_store
ai20: \$phenixjobConfig.input.FwhsP\yyyy\W/d (14) F_WHS_FP AI20: "yyyy\WddHmss.DEL.gz"	raw_finances_store_invoice/AI20/YYYYMM/
ai41: \$phenixjobConfig.input.FwhsP\yyyy\W/d(14) F_WHS_FP AI41 "7.yyyy\WddHmss.DEL.gz"	raw_finances_store_invoice/AI41/YYYYMM/
AI48 Lg Facture Clt avec infos taxes	raw_finances_store_invoice/AI48/YYYYMM/
AI4C Composant Lot lg facture clt inf tax	raw_finances_store_invoice/AI4C/YYYYMM/
AI42 COMPOSANT LOT LIGNE FACT CLT	raw_finances_store_invoice/AI42/YYYYMM/
OUTPUT HADOOP	OUTPUT GCP
/phenix/views/logistics/bv_purchase_invoice_full	finances_store_invoice/purchase_invoice_full

People
Assignee: Sara EL MAHFOUDI
Reporter: Lila MOULAI
Votes: 0
Watchers: 0

Dates
Created: 20/05/2022 3:28 PM
Updated: 06/07/2022 2:17 PM

Development

Figure 24: Carte Jira de migration du job "finance-store-invoice"

✚ **Résolution d'un problème qui concerne un calcul erroné en production :** Le problème concerne un produit qui avait un mauvais calcul au niveau de sa commande future. Ce calcul est important pour pouvoir faire des prévisions de vente et pour calculer les commandes des entrepôts. Pour la résolution de ce problème, j'ai suivi la même méthodologie de travail pour les autres jobs, et mon développement est en recette maintenant.

Carte erronée quand il n'existe pas de condition d'achat future

Type: User Story
Priority: Minor
Component/s: None
Labels: None
Requirement Status: UNCOVERED
Estimation dev: 5

Status: Unresolved
Resolution: Phenix Team only (issues only visible to members of the Phenix Team)
Security Level: Phenix Team only (issues only visible to members of the Phenix Team)

Description

L'itemkey 5721922 n'a pas de condition d'achat future dans productulcamp. Cela empêche de calculer correctement le pivot cartographie et ne permet pas de connaître le commandable futur pour cet article. Le commandable futur permet à Epopé de faire des prévisions de vente et permet de calculer les commandes des entrepôts. Le pivot productulcamp ne contient qu'un seul état daté. L'ajout de plusieurs états datés dans ce pivot a causé des doublons. pas de donnée après le 01/07/2022 alors qu'elle existe dans BCP

```

12 select * from bv_productulcamp
13 where itemkey = '5721922' and itemkey < '00'

```

```

INFO : mapreduce job launched.
INFO : Stage-Stage-1: Map: 2 Cumulative CPU: 18.76 sec HDFS Read: 231293735 HDFS
INFO : Total MapReduce CPU Time Spent: 18 seconds 768 msec
INFO : Completed executing command=queryId=hive_20220627173838_8746eccc-3f85-4d72-8e
INFO : OK

```

People
Assignee: Sara EL MAHFOUDI
Reporter: Rachel DUBOIS
Votes: 0
Watchers: 0

Dates
Created: 20/05/2022 3:28 PM
Updated: 06/07/2022 2:17 PM

Development

Figure 25: Carte Jira pour la résolution du problème de calcul de la cartographie