

Mémoire de fin d'études de Master of Science 2

Titre d'Expert en informatique et Système d'Information

Présenté par Guillaume SAYER

SMARTSTREAM

Une architecture orientée microservices pour la
détection de fraude



Université

SUPINFO INTERNATIONAL UNIVERSITY

lyon@supinfo.com

+33 1 53 35 97 00

16 Rue Jean Desparmet — Lyon (69008)

Entreprise d'accueil

WORLDLINE

53 avenue Paul Kruger

Villeurbanne (69100)

Tuteur entreprise

Virgile DEVAUX

virgile.devaux@worldline.com

Dates de stage de fin d'études

Année 2017

Sign of Success

Remerciements

Je remercie M. Frédéric Oblé, Directeur du département R&D-HPV, et M. Stéphane Valembois, Manager de l'équipe ingénierie, pour m'avoir permis de réaliser mes missions en toute sérénité au cours de mon alternance.

Je tiens également à remercier tout particulièrement mon tuteur, M. Virgile Devaux, pour son aide, sa confiance, ses conseils et le partage de son expertise tout au long de ma dernière année d'alternance. Je remercie aussi M. Christophe Blettry et M. Yannis Mazzer pour leur accueil lors de mon arrivée à Worldline et leurs précieux conseils.

Je veux également exprimer ma gratitude à l'ensemble de mon équipe ainsi qu'à tous les membres du département R&D qui ont largement contribué au succès de mon alternance par leur disponibilité, leur patience et leur bonne humeur.

Enfin, merci à toutes les personnes qui m'ont entouré pendant cette alternance, et qui ont participé d'une façon ou d'une autre à son bon déroulement.



Table des matières

1	Introduction	3
1.1	Présentation du contenu du mémoire	3
1.2	Mon parcours	4
1.3	Argumentation du choix du sujet de mémoire	5
2	Présentation de Worldline: un leader innovant du e-paiement	6
2.1	Worldline le fleuron du groupe ATOS	6
2.2	L'innovation et la RSE au cœur du modèle économique de Worldline	8
2.3	Les domaines d'activité ou « Global Services Lines » de Worldline	9
2.4	La recherche et le développement (R&D) chez Worldline	12
2.4.1	Le management et l'organisation de la R&D	12
2.4.2	Propriété intellectuelle et brevets	13
2.4.3	Les principaux axes de recherche et d'expérimentation	13
2.5	Analyse des risques du secteur et des principaux atouts de Worldline	15
2.5.1	Un environnement complexe et incertain	15
2.5.2	La transformation digitale bouleverse les modèles économiques	16
2.5.3	Une stratégie d'internationalisation bien maîtrisée	17
2.5.4	Des savoir-faire et des capacités technologiques qui font la différence	17
2.6	Mon alternance au sein de l'équipe R&D «High Processing and Volumes»	19
2.6.1	Présentation de l'équipe HPV	19
2.6.2	Un fonctionnement en mode « agile »	19
2.6.3	Mes principales missions pendant l'alternance	20
3	Analyse du contexte: la lutte contre la fraude dans le secteur des e-paiements	21
3.1	La détection de la fraude dans les services numériques de paiement	21
3.1.1	La fraude en ligne explose	21
3.1.2	La lutte contre la fraude est devenue une priorité	22
3.2	Worldline est un spécialiste du « Fraud Risk Management »	22
3.2.1	La détection de fraude est un enjeu stratégique pour Worldline	22
3.2.2	Le projet MI (Machine intelligence)	25
3.3	Un projet porté par les innovations du secteur informatique	27
3.3.1	La démocratisation du Cloud et des modèles As a Service	27
3.3.2	Vers de nouvelles architectures pour le Big Data	28
3.4	Analyse personnelle du contexte de stage de fin d'études	30
4	Problématique: une architecture adaptée à la détection de fraude	31
4.1	Les limites des systèmes de détection de fraude	31
4.2	Vers une nouvelle approche des systèmes de détection de fraude	31
4.3	Les contraintes du Machine Learning	32
4.4	Les contraintes du traitement temps réel	33
4.5	Le manifeste réactif	34
4.6	Les architectures style « microservice »	35
5	Méthodes habituellement utilisées dans les projets Big Data	36
5.1	Les modèles d'architecture adaptés au temps réel	36

5.1.1	Les principaux frameworks utilisés dans les projets Big Data	37
5.1.2	La pile SMACK	37
5.1.3	Cloudera Oryx 2 une architecture Lambda pour le machine learning	39
5.2	Avantages et contraintes d'un modèle d'architecture microservice	40
5.2.1	Des avantages techniques décisifs.....	40
5.2.2	Des changements importants dans la chaîne de production	40
6	Exposé des décisions prises et des interventions dans le cadre du projet SmartStream	42
6.1	Le cadrage du projet.....	42
6.1.1	Objectifs et contraintes	42
6.1.2	La plateforme de préproduction WLAP (Worldline Analytics Platform)	44
6.2	La conception et la mise en œuvre du middleware.....	45
6.2.1	Vue globale de SmartStream	45
6.2.2	La communication « brokerless » avec OMQ (Messaging System)	48
6.2.3	Optimisation des performances : sérialisation avec Protobuf	49
6.3	La mise en place du « Continuous Delivery ».....	53
6.3.1	La conteneurisation avec Docker	53
6.3.2	L'orchestration et la gestion des clusters avec Swarm	53
6.3.3	Le Build & Run avec GitlabCI	54
6.4	Le Datalab avec Jupyter Hub	54
7	Présentation du caractère original de la solution SmartStream.....	56
7.1	Une architecture ouverte, évolutive, élastique et résiliente	56
7.2	Des équipes autonomes et pluridisciplinaires.....	57
7.3	Une approche bien adaptée aux projets R&D	57
7.4	Des changements qui soulèvent des difficultés	58
7.4.1	Une remise en cause des habitudes de travail	58
7.4.2	Le « turn-over » de l'équipe	58
7.4.3	Améliorer le module CaaS (Conteneur as a Service)	58
8	Analyse de l'approche choisie: une solution prometteuse	60
8.1	Résultats obtenus	60
8.2	Analyse du champ d'application de la solution élaborée	61
8.3	Mise en perspective avec d'autres contextes	62
9	Réflexion sur le stage et le mémoire: un parcours riche et stimulant	63
9.1	Auto-évaluation du travail réalisé	63
9.2	Bilan des acquis sur les aspects techniques, stratégiques et managériaux	64
9.3	Perspectives professionnelles en relation avec les compétences acquises.	64
	Conclusion	66
	Bibliographie et Webographie	67
	Liste des annexes	68

1 Introduction

1.1 Présentation du contenu du mémoire

Ce mémoire clôture ma formation à SUPINFO International University, et s'appuie sur mon parcours professionnel au sein du département R&D du groupe Worldline, spécialiste des services transactionnels et d'e-paiement.

J'effectue tout d'abord une brève présentation de mon parcours de formation et des raisons qui m'ont amené à choisir le sujet de ce mémoire. Dans un second temps, je présente l'évolution, le modèle économique et les activités de Worldline de manière à comprendre **comment cette société répond aux nombreux défis, notamment technologiques, auxquels elle doit faire face**. Cette analyse me conduit à détailler ses activités de R&D avec un focus particulier sur l'**équipe HPV** (Hight Process and Volumes) de Lyon où s'est déroulée mon alternance. Un premier bilan m'amène à exposer les nombreux avantages que présente une alternance au sein du département R&D d'un groupe comme Worldline qui a placé l'innovation au cœur de son modèle économique.

Dans la troisième partie, j'aborde les différents **éléments de contexte** nécessaires à la compréhension des enjeux et des moyens mis en place pour lutter contre la fraude dans le secteur des paiements et au sein de Worldline en particulier. Ce qui me conduit ensuite à présenter les travaux R&D de Worldline dans le cadre de la détection de fraude.

La **quatrième et cinquième partie** visent à identifier les **problématiques** auxquelles sont confrontés les systèmes actuels de détection de fraude. Ce sera l'occasion d'examiner comment les technologies issues du Big Data alliées aux techniques d'apprentissage machine (Machine Learning) ouvrent de nouvelles perspectives, mais introduisent de nombreux challenges techniques. Je verrai ensuite dans quelle mesure une architecture « orientée microservice » peut être une bonne alternative aux solutions Big Data habituellement utilisées et un support efficace pour l'amélioration des services de détection traditionnels déjà en place.

La **sixième partie** précise **ma contribution au projet SmartStream**, une architecture orientée microservice pour la détection d'anomalie en temps réel. Cet exposé est scindé en plusieurs parties : la première concerne le cadrage et le rappel des principales exigences auxquelles doit répondre l'architecture alors que la seconde détaille ma participation au choix et à la mise en œuvre des frameworks nécessaires à l'implémentation de ses composants. La troisième partie retrace mes interventions dans le cadre de l'automatisation des processus d'intégration et de déploiement du logiciel dans une perspective de livraison continue (Continuous Delivery). Cette nouvelle organisation du processus de livraison du logiciel impose de nombreux changements, mais pose également des difficultés.

1.2 Mon parcours

Suite à l'obtention de mon baccalauréat Scientifique Option Science de l'Ingénieur, j'ai intégré un BTS SIO (Service Informatique aux Organisations) pour accéder ensuite à un cursus « Master of Science » au sein de SUPINFO International University en 3^e année.

Ma **première expérience significative en entreprise** s'est déroulée en 2014 au sein de la **société V3D** éditeur de logiciels QoS/QoE pour terminaux mobiles. Cette startup d'environ 30 personnes propose aux opérateurs télécoms une solution de monitoring de la qualité de service et de l'expérience des utilisateurs de smartphones et terminaux haut débit. Un CDD de 3 mois, suivi d'une alternance d'un an dans cette société, m'ont permis de développer mon autonomie et mon sens des responsabilités, car j'ai dû prendre en charge progressivement l'évolution de l'infrastructure du système d'information de la société. J'ai été amené à piloter la mise en place, le paramétrage d'un CRM (Customer Relationship Management), un outil stratégique pour la société V3D. J'ai également pris en charge la formation des commerciaux, futurs utilisateurs du CRM. Cette expérience m'a fait prendre conscience de l'importance de l'écoute et de l'accompagnement des utilisateurs pour faciliter l'adoption de nouvelles technologies qui soient adaptées à leurs attentes.

Après cette première expérience enrichissante aussi bien sur le plan humain que technique, j'ai choisi de prolonger ma **professionnalisation en alternance au sein du groupe Worldline**, un leader des services numériques de paiement. Ma première réalisation significative à Worldline s'est déroulée pendant les 6 premiers mois de mon alternance (à partir d'août 2015) dans le département R&D. Associé à un jeune ingénieur débutant, nous nous sommes vu confier la réalisation d'un PoC (Proof of Concept) de reconnaissance vocale et son intégration dans un PDA (Personal Digital Assistant). Ce projet a exigé de mettre en œuvre des compétences très variées (hardware et software) afin de pouvoir embarquer la solution dans le robot réalisé à l'aide d'une imprimante 3D. À cette occasion, j'ai pu mesurer la difficulté, pour une petite équipe de R&D, à aborder un sujet complexe faisant intervenir de multiples technologies « open source » et pas toujours matures. Actuellement, depuis près d'un an, je participe à l'élaboration d'un PoC (Proof Of Concept) de détection d'anomalies utilisant des techniques de Machine Learning. L'objectif est de concevoir une plateforme générique capable d'interpréter des flux de données, afin de prédire des anomalies dans le cadre de cas d'utilisation variés, par exemple, **la détection de fraudes bancaires** ou encore de cellules cancéreuses à partir de clichés médicaux. Cette alternance dans la R&D de Worldline m'a donc permis d'aborder de nombreux domaines et technologies en lien avec les enseignements dispensés à SUPINFO International University.

1.3 Argumentation du choix du sujet de mémoire

Ce mémoire a pour cadre général l'amélioration des services de prédiction de fraude de Worldline. Mon sujet aborde plus particulièrement la conception et l'implémentation d'une architecture qui soit adaptée aux spécificités des applications de Machine Learning.

Bien qu'ayant participé à d'autres projets pendant mon alternance, c'est un sujet qui m'intéresse pour diverses raisons qui tiennent à :

- sa **richesse**, car il mobilise à la fois des connaissances métiers (enjeux de la fraude, stratégie et outils de détection...) et des compétences dans des technologies qui m'intéressent particulièrement (Big Data, Machine Learning ...)
- son **caractère innovant** (architecture microservices) en rupture avec les modèles de conception d'architecture habituellement utilisés
- la **transversalité** des technologies employées, génériques et réutilisables dans d'autres projets Big Data
- les **applications potentielles de l'analyse prédictive** sont très vastes au sein des activités de Worldline

La lutte contre la fraude est un enjeu important pour Worldline et un objectif de long terme pour l'équipe R&D qui m'a accueilli. C'est pourquoi ce travail peut m'ouvrir des **perspectives intéressantes dans le cadre de ma poursuite d'étude à GEM** (Grenoble École de Management) dans le Mastère Spécialisé « Management technologique et de l'innovation » que j'effectuerai en alternance dans le département BCBI (Business Center BI and Big Data) de Worldline à la rentrée 2017.

2 Présentation de Worldline: un leader innovant du e-paiement

2.1 Worldline le fleuron du groupe ATOS

Worldline filiale du groupe ATOS est devenu en quelques décennies un leader européen et un acteur mondial de premier plan dans les secteurs des services de paiement (e-paiement) et des services électroniques transactionnels. Worldline offre à ses clients une nouvelle génération de services numériques qui leur permet d'offrir au consommateur final des solutions innovantes et fluides.

Atos est créé en 1997 par la fusion de Sligos un spécialiste du traitement des transactions bancaires par carte et du groupe Axime, une SSII offrant des prestations dans les domaines de la monétique, télématique et du multimédia, mais également une expertise dans la gestion des opérations sur les marchés bancaires et financiers.

En 2000, suite de l'intégration de la société Origin (la branche informatique de Philips), le groupe s'appelle désormais **Atos Origin Services** et se positionne comme un leader international des services informatiques (ESI). Sa fusion avec Euronext lui permet d'intégrer progressivement les activités d'édition de logiciels de paiement et d'intégration au sein de la filiale Atos Origin Euronext Payment Solutions.

En 2004 Atos Origin Services devient Atos Worldline à l'occasion du regroupement des activités du groupe dans les secteurs des services de paiement et des services électroniques transactionnels.

Dans les années qui suivent, Atos Worldline se positionne en France et par ses filiales en Allemagne puis au Royaume-Uni comme un leader dans les services transactionnels de haute technologie (*High Tech Transactional Services* ou « HTTS »).

En 2006 **Atos Worldline** poursuit son expansion en Belgique en intégrant les sociétés Banksys et Bank Card Company spécialisées dans les solutions et les systèmes de paiement et acquisition commerciale et devient ainsi un acteur majeur sur le marché belge notamment en tant qu'opérateur du schéma de paiement Bancontact/Mister Cash.

Depuis 2009, la stratégie du groupe Atos dans le domaine des services de paiement consiste à déployer ses activités HTTS à l'international. C'est durant cette période qu'elle renforce sa présence en Europe (Belgique, Espagne, Pays-Bas et Royaume-Uni), pour poursuivre son expansion vers l'Asie notamment avec l'acquisition, en 2010, de la société Venture Infotek, leader dans le domaine du paiement sur le **marché indien**.

En juillet **2011**, le groupe Atos a procédé à l'**acquisition de Siemens IT Solutions and Services (SIS)**, une importante SSII européenne appartenant au groupe allemand Siemens AG et issue des apports de plusieurs entités informatiques du conglomerat allemand Siemens.

Cette acquisition a permis au Groupe Worldline d'intégrer principalement les activités de services de mobilité et de transactions électroniques (*Mobility and e-Transactional Services* ou « *MTS* ») de SIS au Royaume-Uni, au Chili et en Argentine.

C'est en **2012** que Atos Worldline fait l'acquisition de la société néerlandaise Quality Equipment BV un acteur majeur sur le marché néerlandais du paiement électronique dans les secteurs de la vente, restauration et du parking.

La période **2013-2014** permet à ATOS de procéder à la filialisation à 100 % de l'ensemble de ses activités de services de paiement et services transactionnels au sein d'une entité unique dénommée **Worldline**. Elle devient une société cotée en bourse à l'Euronext le 27 juin 2014.

2016 est l'année de la fusion avec le groupe Equens pour créer la société **equensWorldline** (filiale contrôlée à hauteur de 64 % par Worldline) et de l'acquisition de Paysquare, filiale d'Equens pour l'activité Acquisition Commerçants. C'est l'occasion pour le groupe d'accéder à une implantation paneuropéenne étendue et consolider son leadership dans les services de paiements. La société est maintenant présente dans 22 pays d'Europe et sur plusieurs marchés émergents en Amérique latine et en Asie-Pacifique où elle se positionne comme un leader dans le secteur des licences de logiciels de paiement (Worldline est également le plus important prestataire de paiement en Inde).

Actuellement, riche de 40 ans d'expérience dans le secteur des paiements, Worldline fait maintenant partie des principaux acteurs du marché du B2B2C. Elle réalise un chiffre d'affaire pro forma estimé à environ 1,5 Md d'euros et emploie plus de 8700 collaborateurs.

Le Groupe fournit actuellement à ses clients des services d'acceptation des paiements, de traitement des transactions cartes comme non-cartes, de support à l'émission des moyens de paiement correspondants, de gestion des relations clients (CRM), de détection des fraudes et gestion des contentieux. Ces services sont opérés par plusieurs plateformes informatiques issues des investissements de Worldline (programme WIPE) et Equens. En Europe, le Groupe exploite un réseau de **huit centres de données** (datacenters) interconnectés, hautement sécurisés et entièrement redondants, situés en France, en Allemagne, en Belgique, aux Pays-Bas et en Italie. Le hub de centres de données européen du Groupe s'étend sur une superficie de plus de 6 000 m² et fait fonctionner environ 19 000 serveurs ayant une capacité de stockage de données d'environ 18 pétaoctets. Ce hub de centres de données européen prend en charge environ 85 % du volume total des transactions du Groupe. Tous les centres de données du hub européen du Groupe, qui sont conçus pour répondre aux besoins spécifiques de ses activités, sont conformes aux normes de sécurité les plus récentes. Worldline respecte les règles élaborées par l'industrie des cartes de paiement Payment Card Industry Data Security Standard (PCI-DSS et 3DS) imposées aux prestataires de services de paiement pour traiter les données des utilisateurs de carte de paiement.

2.2 L'innovation et la RSE au cœur du modèle économique de Worldline

Worldline couvre l'ensemble de la chaîne de valeur étendue des activités de services de paiement. Le Groupe fournit une gamme complète de solutions aux institutions financières, commerçants, entreprises et entités gouvernementales afin de prendre en charge leurs **services de paiement externalisés**. Ces activités s'appuient sur des **contrats de long terme** prévoyant une commission pour la mise en place de la solution, puis selon les volumes ou valeurs des transactions traitées pendant la durée du contrat. Worldline gère ses activités au niveau mondial afin de générer le maximum d'économies d'échelle. Elle a mis en place une combinaison de processus et d'outils standardisés (SOA, ITIL ...), de partage de bonnes pratiques (agilité, DevOps...) afin de proposer des services de haute qualité à un prix compétitif.

❖ « Réinventer les services de paiement »

Worldline propose une nouvelle génération de services qui permet d'améliorer l'expérience utilisateur grâce à des **solutions innovantes, fluides et sécurisées**. Sa capacité technologique et son expertise métier lui permet de mettre sur le marché des solutions complètes et sur mesure « End to End », d'anticiper les besoins de ses clients et les accompagner dans leur transformation numérique. Cet avantage lui permet de se démarquer de ses concurrents et ainsi consolider sa place de leader. Pour Worldline, investir dans l'innovation est donc une priorité stratégique cruciale : il s'agit pour Worldline de « réinventer les services de paiement » autour de l'expérience utilisateur comme aime le rappeler N. Kozakiewicz, le responsable R&D du groupe. Worldline favorise la diffusion d'une culture d'innovation afin d'inciter ses salariés à une plus grande créativité et a encouragé les projets créateurs de valeur grâce à des alliances et des partenaires multiples. La forte culture d'innovation du Groupe « **l'innovation est dans notre ADN...** » lui permet d'améliorer ses services existants et d'exploiter les nouvelles technologies digitales pour créer de nouveaux marchés et services. L'innovation est donc une composante essentielle de son business model.

❖ Être un acteur responsable dans toutes ses activités

La **Responsabilité Sociale d'Entreprise** (RSE) est une démarche qui est au cœur des préoccupations de Worldline et qui lui permet de guider et structurer sa stratégie. Cette démarche vise à **intégrer le développement durable dans l'ensemble de ses activités**. Elle repose sur un dialogue permanent avec de nombreuses parties prenantes : clients (banques, commerçants...), employés, fournisseurs, partenaires, collectivités locales et les autorités publiques, les associations et les investisseurs et analystes financiers. Dans le cadre de ce

dialogue, afin de répondre aux attentes de ses principales parties prenantes, Worldline s'engage à :

- améliorer la **relation de confiance avec ses clients** ce qui implique la mise à disposition de plateformes de traitement de transactions sécurisées (protection des données personnelles...) et disponibles, mais aussi des services innovants sources de création de valeur,
- être **un employeur responsable** qui favorise l'épanouissement personnel et professionnel de ses salariés en améliorant la gestion des talents, la diversité, les conditions de travail et la motivation,
- être en **conformité par rapport au code d'éthique et règles de loyauté en affaire** (non-respect des lois sur la concurrence, les pots-de-vin, la corruption, ainsi que la fraude en général),
- réduire son **empreinte environnementale** avec l'efficacité énergétique de ses datacenters et transition énergétique vers des solutions décarbonées (objectif zéro carbone en 2020).

En 2016, Worldline a pris la décision d'inscrire sa démarche RSE dans une optique de long terme avec le programme « **TRUST 2020 : La confiance au cœur de tout ce que nous faisons** ». Ce programme vise à renforcer le caractère durable des opérations de Worldline pour les années à venir, et consolider la position de leader de Worldline en termes de RSE sur le marché. L'entreprise a identifié les objectifs RSE spécifiques et mesurables à atteindre pour 2020, et s'engage à publier chaque année les progrès réalisés. Les objectifs du programme **TRUST 2020** concernent les activités opérationnelles, les collaborateurs, la chaîne de valeur et l'environnement. Trust 2020 est un atout stratégique de différenciation sur le marché et un **support essentiel pour initier et développer une relation de confiance durable avec l'ensemble de ses parties prenantes**.

2.3 Les domaines d'activité ou « Global Services Lines » de Worldline

Les activités de Worldline sont organisées autour de **trois Lignes de Services Globales**¹ présentant chacune leur propre gamme de services et solutions au sein de marchés spécifiques :



¹ <http://worldline.com/fr/accueil.html>

➤ La Ligne de Services « **Services Commerçants & Terminaux** » offre aux commerçants des solutions de services de paiement numérique destinés au commerce électronique et mobile, quel que soit le canal de vente utilisé. Les services à haute valeur ajoutée proposés par Worldline à ses clients, notamment ceux de la grande distribution, permettent d'améliorer l'expérience d'achat, de **générer des opportunités de vente incitatives et d'optimiser leurs processus commerciaux**. Les principaux services proposés couvrent les domaines suivants :

- **Acquisition commerciale** avec des solutions d'acceptation de paiement dématérialisées, simples à utiliser et sécurisée,
- **E/m-commerce** avec des plateformes comme **SIPS**,
- **Terminaux de paiement** qu'elle conçoit et installe en y intégrant les dernières avancées en matière de cryptologie,
- **Cartes de paiement** privatives et solutions de fidélisation associées.

Le Groupe offre également des périphériques de points de vente sur mobile et des solutions le paiement à partir des tablettes, qui cible les petits commerçants, les startups et les petites entreprises ou encore des secteurs particuliers, comme la restauration ou les cinémas. Que ce soit à travers les programmes de fidélité ou les services de gestion de la relation client (CRM) ou des **solutions exploitant les opportunités de Big Data** ou autres services à valeur ajoutée, le Groupe étend constamment son portefeuille de services innovants. Grâce à ses succès passés en matière d'innovation, le Groupe a commencé à accompagner un grand nombre de **marques B2C**, dont les exemples les plus récents sont McDonald's et Accor (en France), Carrefour (en France et en Belgique), Adidas (par l'intermédiaire du mur virtuel de chaussures innovantes, adiVerse), et Sephora (Sephora Flash) dans la transformation digitale de leurs processus commerciaux.

➤ La Ligne de Services « **Services Financiers** » s'adresse aux banques et autres institutions financières (environ 200 clients). Sa mission est de leur fournir l'ensemble des services de paiement dans un contexte réglementaire complexe et en constante évolution. Les capacités de traitement des paiements à grande échelle de Worldline lui permettent de proposer de **nouvelles formes de paiement et de valoriser les données client**, en associant, par exemple, un portefeuille mobile (m-wallet) à des fonctions d'analyse de données générant ainsi de nouvelles sources de revenus.

L'offre de services financiers de Worldline concerne le(s) :

- Traitement des acquisitions lorsque les banques externalisent leur processus de traitement des transactions d'acquisition,
- Traitement des émissions,
- e-banking (en ligne/mobile),

- Paiement électronique/mobile,
- Licences d'utilisation de logiciel comme sa solution Worldline Pay dont je présente ci-dessous les principales caractéristiques

Worldline Pay est une plateforme moderne de paiement de bout en bout spécialisée dans l'acquisition, le transfert et l'autorisation des paiements de détail. Elle peut traiter les transactions effectuées en point de vente, aux distributeurs automatiques de billets ou encore en ligne, que ce soit en e-Commerce ou en m-Commerce. En outre, elle gère le **cycle de vie complet des cartes** de paiement ainsi que la compensation des transactions, et les règlements relatifs aux cartes de débit, de crédit et prépayées. WL Pay couvre également des fonctionnalités de **détection de fraude en temps réel** ainsi que la gestion des enquêtes pour fraude et le règlement des différends. Enfin, cette solution prend en charge **les technologies émergentes** telles que les E.-wallets, les paiements mobiles NFC avec HCE et la tokenization. Aujourd'hui, WL Pay une solution globale, disponible en Europe, en Amérique Latine et en Asie-Pacifique. Cette solution est conforme au standard 3-D Secure et incorpore un large éventail de méthodes d'authentification, telles que le mot de passe statique, le mot de passe à usage unique envoyé par SMS ou encore la biométrie.

➤ La Ligne de Services « **Mobilité & Services Web Transactionnels** » accompagne les entreprises dans leur transformation digitale en leur faisant bénéficier des **opportunités autour de la mobilité, des objets connectés et de l'analyse des données massives** (Big Data, Machine Learning ...). Cette GBU propose de **nouveaux produits numériques contextualisés et personnalisés**, au service de l'innovation commerciale et de l'efficacité opérationnelle, tels que les solutions de gestion continue des déplacements et les services connectés de la vie quotidienne : **Connected living, E-ticketing, Plateformes e-gouvernement, Relation client et services Cloud**.

Sips est une **Plateforme de paiement multicanal** qui permet aux commerçants d'accepter des moyens de paiement sur leurs sites en ligne et de gérer la transaction tout au long de son cycle de vie. Elle est conforme aux standards de **sécurité PCI** pour la protection des données cartes des porteurs, Sips intègre plus de **50 moyens de paiement**, dont les cartes de débit et de crédit, le virement, le prélèvement, les portefeuilles, électroniques et les cartes privatives. L'offre peut s'interconnecter à différentes plateformes (Web, mobile, tablettes, serveurs vocaux interactifs...). **Sips** intègre de nombreuses fonctionnalités comme le paiement 1clic, les paiements récurrents, **la détection de la fraude**, l'authentification du client et le DCC (Dynamic Currency Conversion).

2.4 La recherche et le développement (R&D) chez Worldline

2.4.1 Le management et l'organisation de la R&D

Le département Recherche et Développement du Groupe Worldline compte environ 350 ingénieurs, la plupart spécialistes des technologies numériques et des nouveaux usages du digital. Les dépenses en Recherche et Développement du Groupe Worldline s'élevaient à 47,6 millions d'euros en 2016.

Les équipes de Recherche et Développement du Groupe sont gérées de manière centralisée depuis le siège. Les ingénieurs affectés à la Recherche et au Développement sont répartis dans l'ensemble des pays dans lesquels le Groupe est implanté. Un grand nombre d'ingénieurs en R&D du Groupe font partie des équipes opérationnelles et sont affectés aux différentes unités d'affaire (Business Units) du Groupe. Ces équipes se concentrent sur **l'innovation progressive ou « incrémentale »** en contact direct avec les clients. Les autres ingénieurs sont directement rattachés au département R&D et se concentrent sur des projets de Recherche et Développement à plus long terme dédiés à **l'innovation de rupture ou « disruptive »**.

Les équipes du département Recherche et Développement du Groupe interagissent naturellement avec les équipes expérimentées de Recherche et Développement des différentes Service Lines du groupe Atos. L'équipe dédiée à la Recherche et Développement de solutions innovantes « disruptives » soutient donc les 6000 ingénieurs de terrain qui accompagne les clients dans la mise en place des solutions numériques innovantes du Groupe. Worldline entretient des liens forts et durables avec des universités européennes de premier rang en matière **d'analyse de données prédictive** et de **cryptographie** avec l'Université de Passau en Allemagne, et le MLG (Machine Learning Group) de l'Université Libre de Bruxelles en Belgique. Ces relations de confiance à long terme entre Worldline et le monde universitaire favorisent à la fois la recherche des universités et l'innovation de Worldline. Les laboratoires de recherche profitent de uses cases concrets pour tester leurs modèles d'algorithmes tandis que Worldline peut orienter sa recherche vers des besoins essentiels nécessitant des connaissances et un savoir-faire pointus par exemple en matière de détection de fraude. Le Groupe travaille également avec des partenaires industriels pour développer des services innovants pour les véhicules et objets connectés utilisant la communication M2M (Machine to Machine).

Ce réseau de partenaires est important, car il favorise l'innovation « disruptive » en permettant à la R&D à Worldline de créer des synergies, mutualiser des expertises et détecter de nouvelles opportunités de développement sur des marchés encore émergents.

2.4.2 Propriété intellectuelle et brevets

Afin de capitaliser et protéger son savoir-faire technologique, Worldline possède un **portefeuille d'environ 80 brevets**, déposés dans les zones géographiques où le Groupe est le plus actif, notamment en Europe, aux États-Unis, au Canada et en Inde. Les dépôts de brevet sont décidés conformément à la politique de propriété intellectuelle du groupe Atos appliquée par Worldline en cohérence avec les orientations du groupe en matière de R&D.

Certaines entités du Groupe ont de plus en plus recours à des **logiciels libres** (Open Source Software), qui peuvent être utilisés de façon gratuite conformément à des licences comportant parfois une obligation de divulguer le code source développé à partir du logiciel libre. La politique interne du groupe Atos en matière de propriété intellectuelle prévoit que ces utilisations doivent être strictement contrôlées tant par les responsables techniques que juridiques du Groupe afin d'éviter les **risques d'utilisations non contrôlées de logiciels libres** et de divulgation des codes sources relatifs.

2.4.3 Les principaux axes de recherche et d'expérimentation

Afin de créer de nouvelles opportunités et de nouveaux services pour ses clients, Worldline concentre ses efforts de Recherche et Développement en matière d'innovation sur trois axes majeurs

- Trusted Services : Sécurité des services ;
- User eXperience : Expérience utilisateur ;
- HPV Volumes et traitements élevés.

➤ Trusted Services

Ce volet est consacré aux innovations conçues pour **sécuriser davantage les transactions et les services** pour le client final. Dans ce domaine, les axes prioritaires sont les solutions innovantes de paiement, les solutions avancées d'authentification, les outils de cryptographie, les solutions de confidentialité des données (telles que les services « Cloud » sécurisés), et des solutions de résilience du réseau. Parmi les innovations majeures de Wordline dans ce domaine, figurent:

- **Les informations contextualisées** : elles permettent au moment de la demande d'authentification de s'adapter à l'utilisateur, ses habitudes, son emplacement immédiat et ses actions récentes afin que le client ait à fournir le moins d'informations possibles, néanmoins suffisantes pour l'authentification au niveau de l'émetteur. Elles peuvent être utilisées pour toute demande d'authentification classique, y compris les porte-monnaie électroniques (e-wallet), les services bancaires en ligne et les services de paiement pour l'e-commerce ;

- **L'authentification multifacteurs** : aucun outil d'authentification ne peut à lui seul fournir une garantie totale. Par conséquent, les services nécessitant une authentification

rigoureuse, comme les paiements, **conjuguent plusieurs outils d'authentification**. En effet appliqués individuellement, ils seraient jugés insuffisants pour renforcer le niveau de confiance. Les technologies disponibles qui reposent sur une authentification rigoureuse sont de plus en plus nombreuses. Elle s'effectue sur la base de la biométrie (empreintes digitales, reconnaissance faciale ou vocale...), d'éléments sécurisés, de mots de passe et de cas d'utilisation. Les services connectés ont renforcé le besoin de solutions d'authentification intégrées. Afin de proposer des solutions novatrices dans ce domaine, Worldline travaille en amont avec des chercheurs sur des « **reconnaisances biométriques continues** ».

➤ **L'expérience utilisateur (User eXperience) :**

Ce volet se concentre sur des innovations qui tirent profit de la technologie pour **améliorer l'expérience de l'utilisateur** dans le large éventail d'activités des clients de Worldline, notamment les services bancaires, les commerçants, l'automobile, les médias et le divertissement. Les axes prioritaires comprennent des solutions de réalité augmentée, d'amélioration du caractère intuitif des interactions, de paiement sur des plateformes multiples, d'analyse d'informations générées par les périphériques connectés, des solutions point à point, des services de recommandations ainsi que de nouveaux périphériques et accessoires. Parmi les exemples d'innovations majeures de ce domaine, figurent:

- **Connected Kitchen.** Elle permet à l'utilisateur de gagner du temps et d'éviter les oublis pendant ses courses. On peut utiliser ce magnet intelligent pour scanner et se rappeler des produits à acheter. Worldline propose le service intégral de bout en bout, du magnet au Cloud. Connected Kitchen lance des pilotes tant pour les clients B2B que pour les clients B2C.

- **PayGgy** est un service bancaire qui prend la forme d'une **tirelire connectée** et qui offre une expérience éducative et ludique tout en favorisant l'image de marque des banques et la fidélisation. L'Internet des objets (IoT) et les services de dématérialisation ont changé les habitudes de consommation et les banques sont donc, elles aussi, concernées par l'apparition de ces nouveaux usages.

➤ **Volumes et traitements élevés (HPV) :**

Cet axe de R&D se concentre sur les solutions innovantes pour le traitement de volumes de données élevés qui proposent de nouvelles générations d'outils à la fois en matière de matériel et d'architecture des infrastructures, mais également en matière d'algorithmes, afin de permettre le développement d'applications « périphérique au cloud ». Les axes prioritaires dans ce domaine comprennent le **calcul de haute performance (HPC)**, des **solutions de liquid computing**, des **services Cloud et de contextualisation**, l'**intelligence artificielle et**

l'apprentissage automatique (Machine Learning). Dans ce domaine de recherche, deux directions semblent particulièrement prometteuses :

- **le Liquid IT** : Worldline développe actuellement la solution Liquid IT, une réalisation informatique « intelligente » capable de s'adapter aux ressources disponibles pour exécuter une application hébergée dans le Cloud. Cette application vise à répartir efficacement les processus logiciels en temps réel entre les composants matériels disponibles les mieux adaptés à leur exécution ;

- **les plateformes d'analyse de données prescriptive**

L'analyse de données prescriptive étant à présent nécessaire pour la quasi-totalité de ses services, Worldline développe des services d'analyse de données structurées, contextuelles ou en temps réel. Worldline prépare actuellement la prochaine génération de logiciels qui devraient bénéficier de ces nouvelles approches grâce au **programme AiDA** (Artificial Intelligence for Data Analytics), un programme de recherche sur trois ans en collaboration avec quatre universités européennes. Afin de compléter les travaux d'analyse de données, Worldline travaille sur la « **Machine Intelligence** » et le **Deep Learning**. Ces travaux sont très importants, car ils permettent aux machines d'apprendre par elles-mêmes, de créer des modèles et des règles grâce auxquelles on va pouvoir interpréter les données pour prendre des décisions plus sûres. Ces travaux sont essentiels pour tous les services de traitement temps réel de l'information de Worldline notamment ceux orientés vers l'analyse prédictive.

2.5 Analyse des risques du secteur et des principaux atouts de Worldline

2.5.1 Un environnement complexe et incertain

Worldline est un PSP (Payment Services Provider) qui évolue dans un environnement complexe et en évolution permanente. Cet environnement est à la fois source de menaces et d'opportunités. Les principaux facteurs d'évolution à intégrer dans ses choix stratégiques sont :

- **Les attentes et le comportement des clients** : la manière dont les consommateurs de services d'e-paiement utilisent au quotidien les nouvelles technologies clés (mobile, Internet, IoT...) a créé de nouveaux usages et des attentes exigeantes,
- **Les changements majeurs et rapides sur le plan technologie** : les nouvelles technologies numériques et la transformation digitale jouent un rôle fondamental dans la modification de l'environnement des paiements et donc dans le positionnement stratégique des acteurs sur les marchés de Worldline (Mobilité, Cloud, Big Data...),

- Les **évolutions réglementaires** : les institutions financières et les prestataires de services de paiement font face à une série de modifications réglementaires (DSP II...). Ces évolutions ont le potentiel de créer de nouvelles opportunités de sous-traitance pour les fournisseurs de services de paiement comme Worldline.
- Les **nouveaux entrants** : les nouvelles « Fintechs », les opérateurs mobiles et GAFA (Google, Apple, Facebook, Amazon, etc.) font désormais partie de l'écosystème de paiement et risquent de prendre de parts de marché aux acteurs existants comme Worldline,
- Les **produits de substitution** : les moyens de paiement alternatifs vont également augmenter de façon considérable et menacer de désintermédiariser les institutions financières et les prestataires de service de paiement en place comme Worldline (Blockchain, paiement direct en per to per...).

2.5.2 La transformation digitale bouleverse les modèles économiques

Worldline accompagne ses clients dans leur processus de transformation digitale, car celle-ci est devenue un enjeu déterminant pour le secteur du commerce dans son ensemble. Selon **IDC Financial Insights** *«Disruptive technologies including cognitive, robotic process automation, and blockchain will be in use at 50% of banks worldwide by 2020, accelerating digital transformation by 30%.»* La transformation digitale, que l'on appelle parfois aussi transformation numérique, désigne le processus qui permet aux entreprises d'intégrer toutes les technologies digitales disponibles au sein de leurs activités. La transformation digitale n'est pas un concept nouveau puisqu'il existe depuis l'essor d'Internet. 75 % de la population mondiale dispose aujourd'hui d'un téléphone portable. La **mobilité est donc devenue un axe majeur** de la transformation digitale des entreprises. IDC s'attend à ce que d'ici la fin de l'année, les deux tiers de 2000 plus grandes entreprises du monde l'aient placée au centre de leur stratégie d'entreprise (1). **L'étude révèle d'autre part les services informatiques considèrent la virtualisation de l'infrastructure et l'adoption de modèles de distribution dans le cloud privé comme public comme des moteurs majeurs de la croissance.** Les organisations financières doivent donc mettre en place les infrastructures informatiques capables de s'intégrer dans le nouvel écosystème digital : cloud, mobiles, réseaux sociaux, objets connectés. Pour atteindre ces objectifs, elles revoient leurs priorités et passent à une gestion de projets plus agile, à une meilleure utilisation des ressources, à une infrastructure plus résiliente et sécurisée, et à l'informatique normalisée selon l'IDC.

Les entreprises de services financiers comme Worldline sont donc face à un défi majeur. Elles doivent s'adapter aux évolutions rapides des attentes des clients et des conditions du marché ainsi qu'à des réglementations de plus en plus contraignantes, dans un monde de plus en plus numérique.

Comment Worldline répond à ces nouveaux défis et menaces potentielles ?

2.5.3 Une stratégie d'internationalisation bien maîtrisée

Un élément clé de la stratégie du Groupe est la consolidation et le développement de sa présence internationale, à la fois sur les marchés européens et au-delà, avec un focus particulier sur les marchés émergents. En effet Worldline a su consolider les activités de traitement des transactions en Europe et pénétrer de nouveaux marchés émergents par le biais d'**acquisitions ou d'alliances**. Dans le cadre de sa stratégie d'acquisitions, le Groupe étudie les entreprises et les technologies qui offrent des synergies potentielles avec ses propres activités afin de renforcer ses propositions de valeur aux clients. La croissance du Groupe est également renforcée par le **fort soutien du groupe Atos**, l'une des premières ESI en Europe bénéficiant d'une présence importante à l'échelle mondiale.

Cette dimension internationale permet au Groupe de stimuler l'innovation, de proposer des prix compétitifs (grâce aux économies d'échelle) et d'attirer des clients multinationaux importants. Le Groupe dispose d'une **base de clients étendue** pour toutes ses Lignes de Service caractérisée par des relations de long terme et diversifiées. Son avantage concurrentiel premier vient de sa **capacité à accéder et tirer profit localement d'infrastructures technologiques sécurisées et conformes**, de sa connaissance du terrain local dans les pays où il est actif et d'un large éventail de produits et services qui couvrent toute la chaîne de valeur des services de paiement.

2.5.4 Des savoir-faire et des capacités technologiques qui font la différence

Si Worldline est devenu le leader européen du secteur des paiements et des services transactionnels, ce n'est pas uniquement grâce à une stratégie d'internationalisation efficace. Selon moi, plusieurs facteurs internes expliquent également cette croissance rapide. Dans ce travail, je me limiterai mon analyse à deux atouts qui me semblent essentiels pour permettre à Worldline de conserver son leadership : le **management des RH** et sa **capacité en matière d'innovation technologique**.

➤ Worldline veut attirer les talents et soigner son « Wellbeing »

Pour que Worldline puisse faire face à la concurrence et poursuivre son développement, elle doit **attirer, fidéliser, recruter et développer ses talents** pour qu'ils aient les compétences nécessaires pour répondre aux besoins de ses clients. Compte tenu de l'activité de Worldline, la capacité à attirer et recruter des collaborateurs qualifiés et à favoriser la diversité culturelle est essentielle à la réussite de l'entreprise. Les équipes opérationnelles de Worldline interviennent dans un contexte qui évolue rapidement sur les plans technologique, social, économique et réglementaire, nécessitant à la fois agilité intellectuelle, esprit d'équipe ainsi que des expertises pointues et diversifiées. Afin de répondre à ces défis, Worldline met en

place des **plans de développement et des actions de formation individualisées** visant à acquérir les compétences techniques et relationnelles essentielles pour que Worldline puisse rester leader sur ses marchés. En 2016, Worldline a lancé un programme RSE important (TRUST 2020) pour **développer sa marque employeur** et améliorer sa notoriété. Ce programme repose sur quatre axes : attraction des talents, développement des talents, wellbeing@worldline et la Fierté d'appartenance.

En 2016, Worldline a également reçu le label « **Happy Trainees** » qui récompense les entreprises qui portent une attention particulière à l'intégration et l'encadrement de leurs stagiaires. Cette première récompense est le résultat des plans d'action lancés dans le cadre du programme TRUST 2020.

➤ ***Des capacités technologiques de pointe grâce à une R&D performante***

Le marché des services de paiement évolue rapidement et le Groupe assure sa croissance à long terme en développant son offre sur les segments de marché des services de paiement à forte croissance et en constante évolution. Worldline doit également exploiter les nombreuses opportunités qu'offre la transformation digitale de notre société. Cet objectif requiert des **capacités technologiques de pointe** en matière de paiement pour intégrer rapidement et de manière sécurisée de nouveaux services innovants dans sa chaîne de valeur. Il est donc important que Worldline puisse continuer à **investiguer et expérimenter les nouveaux usages** permis par les avancées des technologies de l'informatique et du numérique en général.

Dans ce contexte, **la R&D de Worldline est un élément clé de sa capacité à innover** en lui permettant d'anticiper les risques et de profiter des opportunités générées par ces marchés émergents en pleine évolution. En effet, l'**offre de services nouvelle génération** de Worldline va au-delà des solutions traditionnelles de paiement. Ses solutions omnicanal de bout en bout couvrent des domaines tels que le **Connected Living, l'Internet des objets et l'expérience d'achat digitale ou encore l'analyse de données massives** (Big Data Analytics) qui sont des domaines prometteurs. Sa **capacité d'intégration** est également un atout déterminant, car la plupart de ces solutions sont intégrées dans des plateformes Cloud hébergées au sein de ses propres datacenters.

Malgré un environnement complexe et incertain, **le Groupe Worldline semble donc bien positionner pour capitaliser sur la croissance de la prochaine génération des services de paiement**, car il peut à la fois s'appuyer à la fois sur son expertise métier client, ses succès en matière d'innovation et sa performance en matière de recherche et développement.

2.6 Mon alternance au sein de l'équipe R&D «High Processing and Volumes»

Durant les deux années de mon alternance chez Worldline, j'ai été intégré au sein de l'équipe **HPV** (High Processing & Volumes) de Lyon qui est rattachée au département Recherche & Développement du Groupe.

2.6.1 Présentation de l'équipe HPV

La mission première de l'équipe HPV est d'explorer de nouveaux marchés ou perspectives de croissance pour l'entreprise. L'équipe s'intéresse plus particulièrement aux perspectives « business » offertes par les avancées technologiques en matière de **traitement et d'analyse de données massives** (Machine Learning, Big Data Analytics) et de Calcul Haute Performance (High Performance Computing). Au-delà de ses activités de veille (technologies, marchés...) et de détection d'opportunités, elle a également pour mission de communiquer en interne sur les projets incubés et les savoir-faire acquis, pour en faire profiter les autres équipes opérationnelles du groupe. Cette équipe à taille humaine est composée de profils « techno » variés : architectes logiciel, ingénieurs de recherche, seniors/juniors développeurs et de profils orientés « recherche » (PhD, Doctorants, Datascientists) pour un effectif total de 17 personnes. L'équipe est dirigée par Frédéric Oblé qui impulse et oriente les travaux de recherche et communique en interne et avec les partenaires sur les projets R&D de l'équipe. Les ingénieurs sont sous la responsabilité de Stéphane Valembois qui anime et gère les activités R&D au quotidien.

2.6.2 Un fonctionnement en mode « agile »

Une fois une opportunité qualifiée, l'équipe **incube et préindustrialise les travaux des chercheurs en Data sciences** sous forme de POC (Proof Of Concept), avant de les transmettre aux autres unités d'affaires (BU). Il est important de souligner que l'équipe ne fonctionne pas comme une équipe traditionnelle : tandis que les BU de Worldline ont des clients et fonctionnent en B2B, notre équipe n'a aucun client externe direct et **est considérée comme une unité d'investissement**. Ceci impose de vérifier que les choix technologiques des projets sont en cohérence avec la politique R&D globale de l'entreprise.

En interne l'équipe HPV met en œuvre les principes de « **feature team** » pour fonctionner sur un mode de type agile. En effet, n'ayant pas d'engagement contractuel direct avec des clients externes, elle est régulée par des « deadlines » projets ou des événements (salon, manifestations internes...). Elle doit en permanence s'auto-organiser, de façon à être capable de changer ses priorités en fonction du contexte et des opportunités. Il faut donc être réactif afin de répondre aux évolutions rapides des besoins des BU (Business Units) cibles de nos projets ou des nouvelles orientations managériales. Cette **recherche d'agilité dans nos projets R&D** nous conduit à utiliser les démarches, méthodes et outils adaptés (DevOps,

Scrum...) afin d'accélérer le rythme de prototypage de nos innovations. Cet environnement de travail agile nous permet également d'abandonner un projet assez tôt pour ne pas perdre trop de temps ou de ressources.

2.6.3 Mes principales missions pendant l'alternance

Mon rôle en tant qu'alternant Ingénieur R&D au sein de l'équipe HPV (High Processing and Volume) de Lyon est de participer à la préindustrialisation des travaux effectués par les chercheurs en data science de l'équipe. Durant mes deux années d'alternance, j'ai participé à des projets innovants, à travers la réalisation de POC (Proof Of Concept) portant essentiellement sur le traitement analytique de données massives (Big Data) et utilisant des technologies issues de l'intelligence artificielle comme le Machine Learning.

3 Analyse du contexte: la lutte contre la fraude dans le secteur des e-paiements

3.1 La détection de la fraude dans les services numériques de paiement

3.1.1 La fraude en ligne explose

Dans les secteurs de la banque, de l'assurance ou encore des télécommunications, les comportements frauduleux sont à l'origine de **pertes financières considérables**. En France en 2014, les fraudes à la carte bancaire s'élevaient à 500 millions d'euros. Les pays les plus affectés par la fraude aujourd'hui sont ceux dont le marché des cartes de paiement est mature et dont l'activité e-commerce est développée (ex. la France et le Royaume-Uni). Selon la BCE, 60 % de la valeur totale de la fraude cartes proviennent de paiements effectués en ligne (Rapport sur la fraude à la carte bancaire, février 2014).



Source: Wordline

À l'avenir les volumes et les types de paiements électroniques vont poursuivre leur augmentation et leur diversification entraînant des risques de fraude accrus notamment dans le cas des paiements à distance.

Les paiements à distance sont des paiements réalisés hors du point de vente par Internet, téléphone ou courrier sans que le commerçant puisse voir le porteur de la carte. Il est donc impossible pour le commerçant de détecter la fraude :

Ce type de fraude intervient principalement de 3 manières :

- vol de la carte bancaire
- piratage à distance du numéro de carte, de la date d'expiration et du cryptogramme (par exemple lors d'un paiement sur un site pas suffisamment sécurisé)
- récupération, chez un commerçant, des informations concernant une CB.

En cas de fraude, les contrats de vente à distance (VAD) stipulent généralement que les sommes en cause seront débitées du compte du l'accepteur. C'est donc l'accepteur

(Worldline) qui supporte la fraude et non la banque du porteur comme on pourrait le croire. On constate donc aujourd'hui une **augmentation de la fraude** qui risque de s'amplifier si de mesures de sécurité forte ne sont pas implémentées

3.1.2 La lutte contre la fraude est devenue une priorité

La fraude s'est déportée du monde physique aux transactions en ligne et l'économie numérique en pleine évolution a créé de nouvelles opportunités en matière de fraude : les fraudeurs s'organisent et rivalisent d'ingéniosité.

D'après une étude réalisée par Ovum, l'investissement dans la protection contre la fraude augmentera de 6,5 % par an sur la période 2012-2020 (selon les prévisions d'Ovum concernant les dépenses liées aux technologies de paiement).

Pour l'ensemble des acteurs du secteur des services de paiement, la détection et la prévention des fraudes sont donc devenues un combat permanent à mener sur tous les canaux et pour tous les moyens de paiement. L'enjeu principal pour les acteurs du secteur des services de paiement étant d'**identifier les fraudes plus efficacement et rapidement** pour en limiter les conséquences **sans nuire à l'expérience client** (ralentissement de la transaction, blocage intempestif du moyen de paiement...).

De manière générale, les solutions de lutte contre la fraude disponibles sur le marché s'appuient sur plusieurs approches utilisant :

- l'analyse comportementale en renfort d'une solution d'authentification,
- des outils de cryptologie à installer sur les terminaux clients,
- des moteurs de règles à intégrer directement aux sites transactionnels
- l'analyse de corrélation d'événements adaptée à la détection de fraude
- et plus récemment sur les techniques de machine learning

3.2 Worldline est un spécialiste du « Fraud Risk Management »

Le Groupe Worldline se différencie de ses concurrents par la fiabilité de ses plateformes, le niveau de sécurité de ses solutions, mais aussi par ses capacités d'innovation en matière de gestion et détection des fraudes bancaires.

3.2.1 La détection de fraude est un enjeu stratégique pour Worldline

En matière de lutte contre la fraude, l'**enjeu pour Worldline** est double :

- **financier** (réduire les coûts directs et indirects liés à la fraude notamment en tant qu'acquéreur du paiement),
- **stratégique**, en terme de préservation de son image de PSP fiable et innovant auprès de ses clients, notamment ceux du secteur bancaire.

Aujourd'hui Worldline est l'un des rares acteurs du marché à offrir des solutions transverses capables de collecter et d'analyser des données pertinentes en provenance de plusieurs canaux, permettant d'avoir une **stratégie globale de lutte antifraude**.

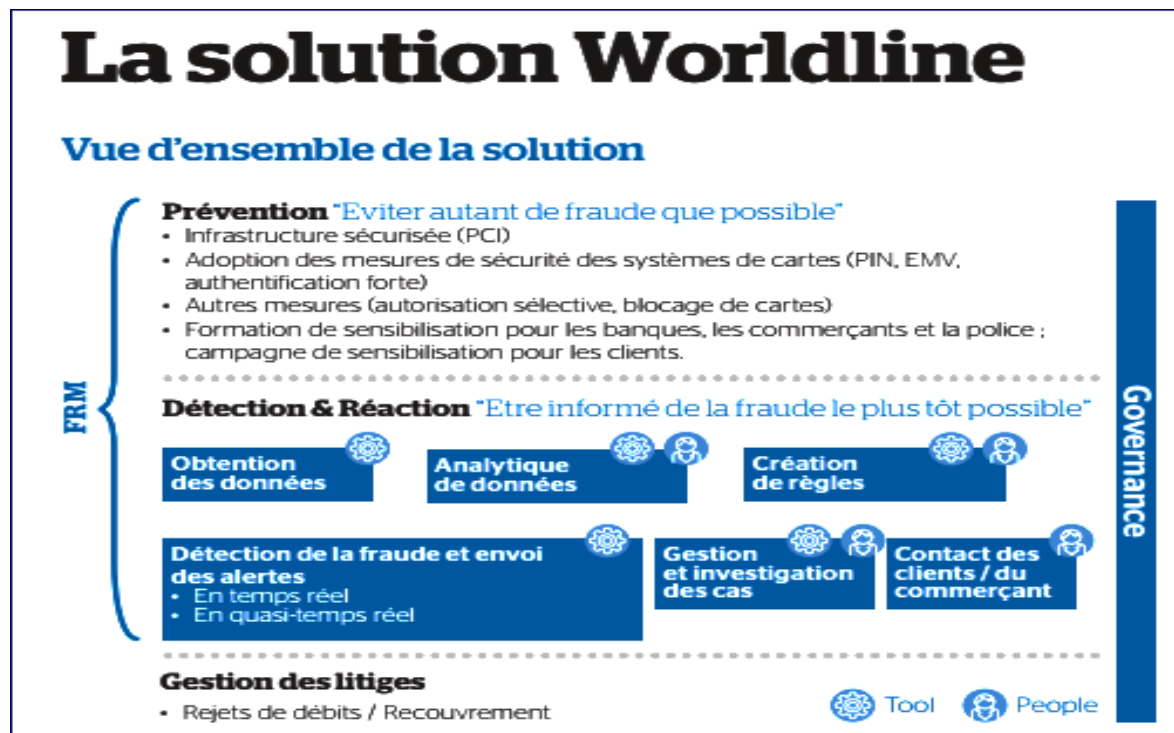


Source: *Worldline*

Worldline est reconnue actuellement comme un expert du FRM (**Fraud Risk Management**) dans le secteur des services de paiement. Afin de conserver son avance dans ce domaine, la **R&D de Worldline travaille depuis de nombreuses années à l'amélioration de ses stratégies de lutte contre la fraude**. L'entreprise propose en effet des services, des outils et des experts qui couvrent la totalité de la chaîne de valeur de gestion du risque de fraude. Worldline possède des offres, des outils et des personnes capables de couvrir toute la chaîne de valeur de la gestion du risque de fraude : prévention, détection et réaction (de la détection de fraude jusqu'au contact du client) et gestion des litiges. Au sein de ses plateformes de paiement (SIPS, WPay ...), Worldline allie des technologies récentes de détection de la fraude, en temps réel et quasi-temps réel, des experts métier pour la définition des règles, une équipe d'investigateurs pour gérer les alertes et des capacités de workflow permettant l'automatisation des procédures de contrôle.

Les principales **solutions de détection et de prévention** intégrées par Worldline au sein de ses plateformes de paiement sont :

➤ **Worldline Fraud Risk Management (FRM)** pour les E.-marchands est une **solution omnicanal et omniproduit** qui s'accompagne de l'assistance d'experts humains. C'est une solution permettant de détecter les risques de fraude avant, pendant et après le paiement. Chaque transaction est analysée et évaluée à travers de nombreux critères afin d'établir un score. En fonction du niveau de tolérance fixé, la transaction sera acceptée ou rejetée. L'outil peut être ajusté pour affiner les contrôles exercés et mieux répondre aux spécificités de l'activité du client.

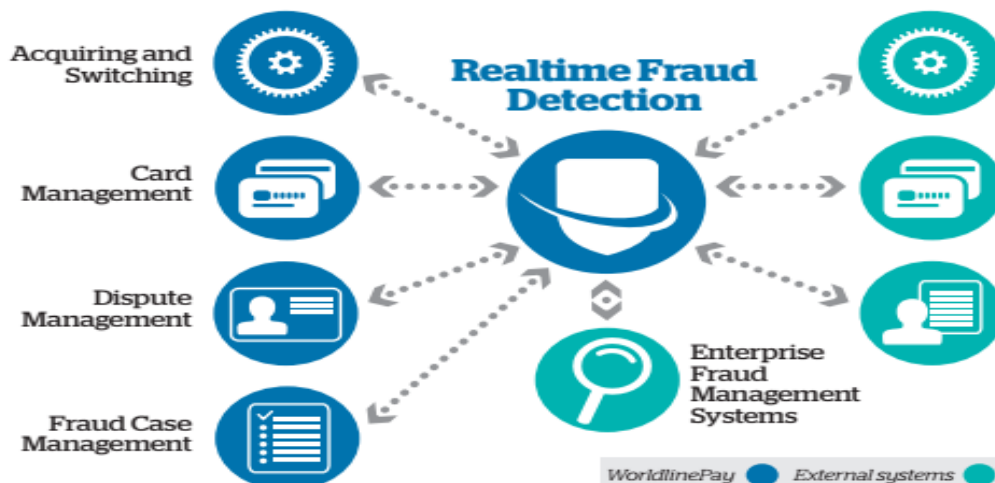


Présentation Worline **FRM**

Dans son offre E/m-Commerce, Worldline combine deux autres **solutions intégrées à sa plateforme WL Pay** : WL Access Control Server et WL Online Watcher.

WL Online Watcher permet la détection et la prévention de fraude, grâce à une série de règles, en évaluant le risque de chaque transaction en temps réel. Un score de risque est ensuite fourni à **WL Access Control Server** qui gère le système d'alerte. Les règles de notation sont basées sur les données transactionnelles ainsi que sur l'appareil de l'utilisateur, l'environnement du marchand et l'analyse de données historisées.

Integration Capabilities



3.2.2 Le projet MI (Machine intelligence)

Comme nous venons de le voir, l'amélioration de la détection des fraudes est l'une des priorités de Worldline. L'objectif est d'améliorer le processus de détection en le rendant plus fiable et adaptatif face aux nouvelles formes de fraude de plus en plus sophistiquées. Dans cette perspective l'équipe HPV de Lyon qui m'a accueilli travaille sur des sujets relevant du domaine de la « **Machine Intelligence** ». Ce domaine de recherche est au carrefour de nombreuses technologies « disruptives » et implique un recours intensif aux techniques d'apprentissage automatique issues du **Machine Learning**.

L'une des étapes majeures du projet est de concevoir une plateforme « intelligente » capable de distribuer à la demande des modèles d'interprétation **en temps réel**, de flux de données hétérogène (transactions de paiement, données captées, images...) et à forte volumétrie, afin de **prédire des anomalies**. Les use-cases potentiels sont variés, comme, par exemple, la détection de fraudes bancaires, de cellules cancéreuses à partir de clichés médicaux ou encore la maintenance prédictive à partir de données émises par des objets connectés dans un contexte industriel.

Le projet est ambitieux, car il aborde de nombreuses problématiques liées :

- Au développement de **nouveaux modèles d'apprentissage** en mode supervisés et/ou non supervisés impliquant le recours à des techniques issues du Machine Learning. Cet axe de recherche vise notamment à **améliorer les classifications et les scores obtenus avec les méthodes actuellement utilisées par Worldline** (cf. FRM présenté précédemment) **en matière de détection de fraude**.
- À la conception et l'expérimentation des composants d'une **architecture distribuée avec des exigences élevées** en termes de performance, disponibilité, sécurité, scalabilité permettant :
 - des traitements temps réel sur des flux de données (streaming), hétérogènes à haute volumétrie,
 - des traitements batch sur de gros volumes de données (Big Data) stockées dans des bases NoSQL (Hadoop Hortonwork, Cassandra ...).
- Le recours à un **CaaS** (Container as a Service) pour un déploiement automatisé d'applications de Machine Learning sous forme de container.
- L'infrastructure cloud sous-jacente nécessaire aux déploiements industrialisés des applications conteneurisées pour des traitements massivement parallèles avec des processeurs adaptés au calcul intensif (HPC, multi-GPU ...).

Le projet implique l'alliance de **nombreux partenaires** qui se partagent les différents lots de travaux de recherche et d'expérimentation :

- R&D HPV Worldline pour la partie « Direction Projet et Ingénierie »
- Université/Recherche avec le MLG de l'Université Libre de Bruxelles (Aspects recherche nouvelles stratégies de prédiction Machine Learning/Deep Learning)
- Business Unit BCBI (production) + plateforme WLAP (pre-production)
- BU FRM de Frankfort (Use case — détection de fraude —)
- BULL Grenoble (High Process Computing)
- Atos & Siemens (Use case industriel : projet de réalité augmentée dans le cadre de la maintenance prédictive)

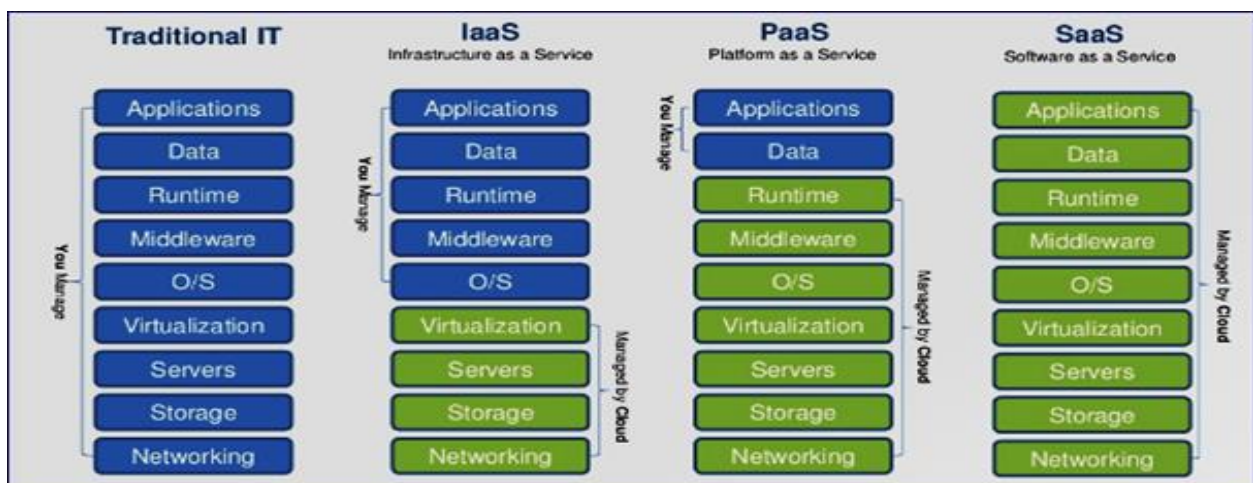
3.3 Un projet porté par les innovations du secteur informatique

Le sujet de ce mémoire s'inscrit dans un projet qui mobilise de nombreuses technologies « disruptives » qui ont bouleversé le monde de l'IT et ouvert de nouvelles perspectives pour la R&D chez Worldline. Dans le cadre de ce travail, je me limiterai à évoquer :

- les évolutions récentes dans les principaux domaines technologiques couverts par le projet MI (Cloud Computing, BigData, Machine learning, IoT), ainsi que,
- les nouvelles architectures, frameworks et méthodes utilisés dans les entreprises.

3.3.1 La démocratisation du Cloud et des modèles As a Service

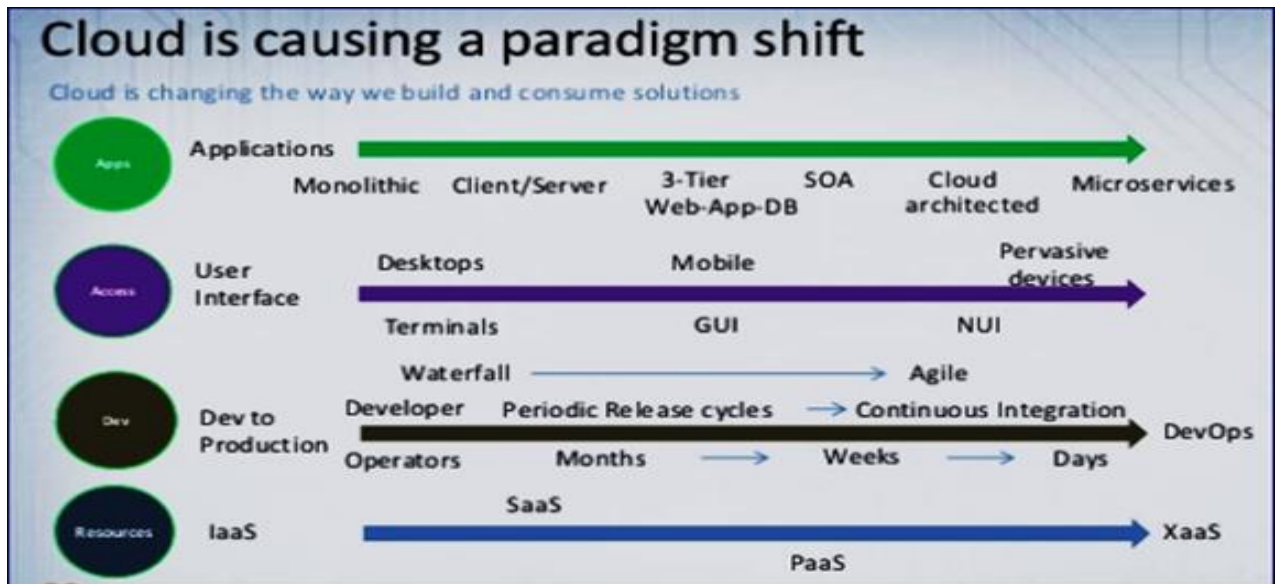
Le terme « Cloud » renvoie aujourd'hui de manière générale à un **nouveau modèle à la fois de distribution et de consommation de l'informatique** qui consiste à mettre à disposition via les réseaux de communication et à la demande (= « as a service »), un ensemble de « ressources » (puissance de calcul, stockage de données, applications, etc.) et de « services » (gestion, administration, etc.).



Source : <http://www.loglan.com.br/microsoft-azure/>

Le recours de plus en plus fréquent aux technologies de conteneurisation comme **Docker** qui accélère la migration des applications du SI traditionnel vers le Cloud en facilitant la création de conteneurs, mais aussi en permettant l'industrialisation de leur déploiement sur des architectures distribuées à grande échelle, favorisant ainsi la livraison continue des applications (**Continuous Delivery**) et les pratiques **DevOps**.

Le cloud implique donc une profonde rupture au niveau des modèles IT en place dans les entreprises aussi bien au niveau de la manière de « consommer » le logiciel que de le produire.

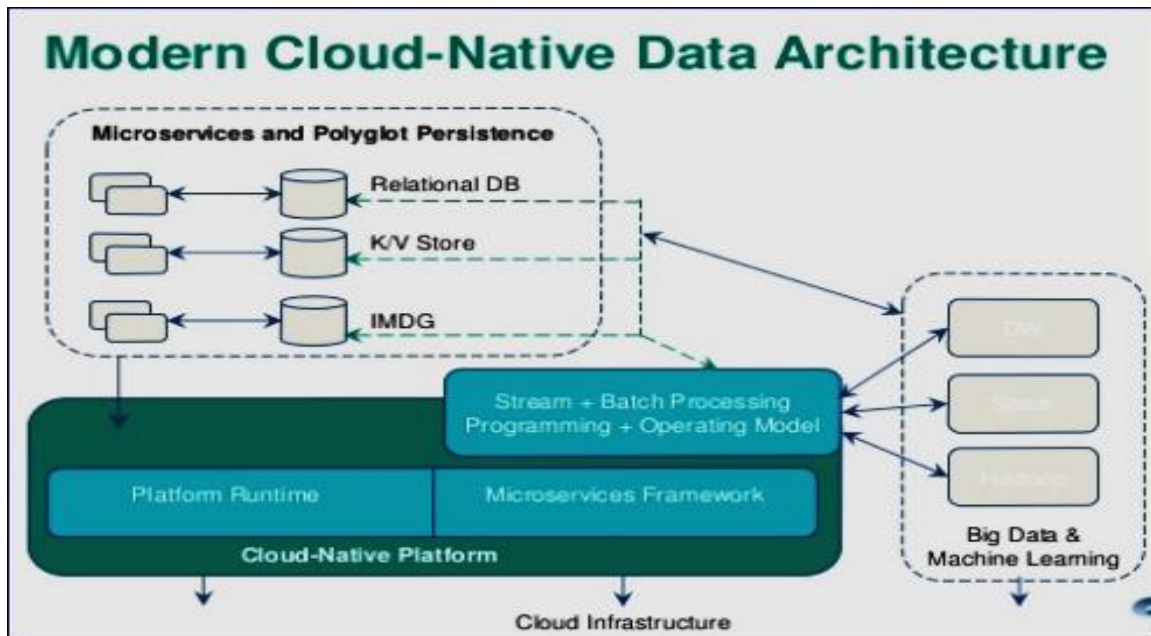


<https://www.slideshare.net/AmazonWebServices/hybrid-cloud-solutions-to-transform-your-organization>

3.3.2 Vers de nouvelles architectures pour le Big Data

En 2020,² l'univers numérique devrait atteindre 44 zettabytes. Un zettabyte est égal à un milliard de téraoctets. Les données provenant de l'Internet des Objets ou de sources non traditionnelles devraient augmenter en taille absolue et relative. Or ces données représentent de précieux gisements de valeur pour les entreprises. La nécessité de traiter des données volumineuses et hétérogènes a conduit à l'élaboration de solutions Big Data open source telles que Hadoop. La plateforme **Hadoop** est capable de traiter des volumes importants de données, mais avec une latence importante. C'est pourquoi sont apparus des systèmes (quasi) temps réels tels que Spark, Storm, Flink, Druid, ou encore Tez et il existe de nombreux composants qui répondent à des besoins variés tels Sqoop, Kafka, Flume, HBase, ZooKeeper... L'un des sujets les plus importants dans la mouvance Big Data aujourd'hui est l'architecture des plateformes, en partie parce qu'il y a de nombreux composants open source disponibles, et les entreprises commencent à s'intéresser à la façon de construire **des architectures Big Data**, en combinant ces nombreuses technologies.

² source IDC "The digital universe of opportunities: Rich data and the increasing value of the Internet of Things,"



Dans le cadre de ces projets, les techniques de machine Learning sont de plus en plus utilisées. Elles permettent d'exploiter les grands volumes de données stockés dans Hadoop et autres bases noSQL à des fins d'**analyses prédictives** (recommandations, promotions, fraudes...). Les nouvelles **architectures orientées microservice** apportent des solutions aux challenges techniques que représente l'intégration de ces nouveaux services d'analyse prédictive aux plateformes existantes souvent monolithiques.

3.4 Analyse personnelle du contexte de stage de fin d'études

Mon alternance s'est déroulée dans une société de dimension internationale qui a placé l'innovation technologique au cœur de sa stratégie. Cet environnement m'a conduit à mieux **comprendre les problématiques liées à l'innovation technologique** et comment une équipe de R&D, comme celle de Lyon, pouvait les traiter.

Le sujet choisi a pour cadre le **projet MI** (Machine Intelligence), un projet d'envergure qui représente un enjeu important pour Worldline. Il mobilise un nombre important de chercheurs et d'ingénieurs du département R&D mais également de **nombreuses autres parties prenantes** en France et à l'étranger. Cette dimension internationale a ses contreparties et je me suis vite rendu compte de la difficulté pour un alternant, à en appréhender toutes les dimensions stratégiques, managériales et techniques. J'ai donc dû faire preuve de curiosité pour comprendre les enjeux de nos recherches, mais aussi d'une certaine souplesse afin de m'adapter rapidement au mode de travail de l'équipe.

C'est un également un **projet complexe sur le plan technologique**, car il mobilise des connaissances et des compétences dans de nombreux domaines. Je me suis particulièrement investi au niveau des aspects liés à l'architecture de la solution et ma contribution a été appréciée par l'équipe. Ce travail m'a permis de capitaliser sur mes expériences professionnelles précédentes et **d'acquérir de nouvelles compétences dans des domaines technologiques passionnants** comme le Big Data ou encore le Machine Learning. Cette période de professionnalisation m'a également offert l'opportunité de découvrir, sur le terrain, comment une équipe pouvait fonctionner en mode agile. Je pense donc avoir progressé non seulement sur le plan technique, mais également sur le plan relationnel.

Travailler dans le département R&D d'un groupe comme Worldline a été, pour moi, à la fois stimulant et enrichissant. Découvrir de nouvelles problématiques, utiliser des technologies récentes sont de réelles **sources de motivation** pour moi. Mais ce que j'ai apprécié tout particulièrement, c'est l'**état d'esprit** qu'il est nécessaire de développer pour travailler et s'épanouir dans ce milieu.

4 Problématique: une architecture adaptée à la détection de fraude

4.1 Les limites des systèmes de détection de fraude

La sophistication croissante des fraudes et de la cybercriminalité met les solutions de lutte contre la fraude de première génération à rude épreuve. La fraude se complexifiant de plus en plus, **une approche standardisée est insuffisante** pour la neutraliser efficacement et les acteurs du secteur des E.-paiements ont des exigences accrues en matière de flexibilité et de niveau d'alerte. En effet, les systèmes antifraudes actuels basés sur des règles métier automatisées présentent souvent une **rigidité** dans la prise de décision qui a pour conséquence de bloquer des transactions saines (faux positifs) et donc générer un manque à gagner. Les taux élevés de **faux positifs** et **l'inefficacité des processus d'investigation** augmentent les coûts opérationnels, les banques sont contraintes de multiplier les investissements pour faire face à ces attaques de plus en plus sophistiquées.

De plus la conception monolithique des applications de détection de fraude actuelles limite :

- les possibilités d'évolution des modèles prédictifs basés sur des règles figées devant faire face à des formes de fraude de plus en plus variées et sophistiquées
- leur capacité à réagir en temps réel, car tout retard dans l'alerte peut être synonyme de lourdes pertes financières.

4.2 Vers une nouvelle approche des systèmes de détection de fraude

Les systèmes actuels de détection de fraude sont comme nous l'avons vu trop figés et difficiles à faire évoluer dans un contexte d'applications encore souvent monolithiques qui n'ont pas intégré les possibilités offertes par les techniques de Machine Learning associées aux technologies Big Data

Le Big Data associé au Machine Learning a introduit, en effet, une triple rupture par rapport aux stratégies déployées dans les services numériques de détection de fraude classique, basés sur de règles expertes préétablies et statiques, car :

- le volume de données disponible à analyser est beaucoup plus grand et permet **d'affiner les analyses**. Avec le Big Data les modèles prédictifs peuvent s'**entraîner en continu** et donc améliorer la détection
- on dispose des outils nécessaires à l'analyse **en temps réel** de flux de données continus pour améliorer la réactivité des systèmes d'alerte,
- jusqu'à présent les règles étaient déterminées à l'avance : « si 2 achats dans 2 pays différents avec moins de 2 heures d'écart alors alerte ». Les modèles statistiques de **Deep learning** détectent des comportements anormaux sans qu'ils soient connus au préalable.

- on a la **possibilité d'intégrer dans l'analyse des données variées** issues, par exemple, des réseaux sociaux ou encore la géolocalisation qui peuvent indiquer si vous êtes en vacances en Chine ou non (et si oui l'alerte à la fraude sera relativisée).

De nouvelles approches plus innovantes fondées sur des algorithmes et technologies d'analyse de larges volumes de données générées et traitées à grande vitesse sont donc apparues. Ces techniques d'apprentissage automatique (Machine Learning) et les plateformes qui les implémentent permettent de détecter des schémas de fraude connus, mais également d'être efficace face à des situations inconnues susceptibles d'être frauduleuses.

Il faut donc intégrer ces modèles prédictifs aux systèmes actuels de détection pour les rendre plus adaptatifs et donc plus fiables. La mise en œuvre de ces nouvelles approches prédictives exige des architectures adaptées. Récemment de nouveaux modèles d'architecture plus évolués ont émergé, afin de s'adapter aux exigences du traitement temps réel dans les projets Big Data et le Machine Learning. Ces architectures apportent des réponses efficaces en permettant :

- l'ingestion en continu de flux de données volumineux et variés (transactions, réseaux sociaux, IOT...)
- un stockage permanent des données dans des bases NoSQL (Data Lake) permettant de construire et d'entraîner des modèles prédictifs de Machine Learning
- le traitement en temps réel des données en mouvement par les modèles prédictifs
- l'amélioration et l'adaptation continues des modèles prédictifs (boucle de rétroaction)

Ces modèles d'architectures plus « réactives » (Event Driven) peuvent être construits sur le pattern (modèle) des « **microservices** ».

Comment une architecture orientée microservices peut-elle contribuer à l'amélioration des solutions actuelles de détection de fraude ?

Pour **répondre à cette problématique**, j'ai dû à la fois mobiliser mes compétences (réseaux, système, programmation, conteneurisation...) et en approfondir d'autres comme les technologies Big Data mais aussi découvrir de nouvelles approches comme les microservices et le Machine Learning.

4.3 Les contraintes du Machine Learning

Les projets de Machine Learning se déroulent en général dans des environnements Big Data, et doivent intégrer les contraintes liées au 3 V du Big Data. Pour commencer de tels projets, il faut donc disposer d'une base historisée de données la plus large possible

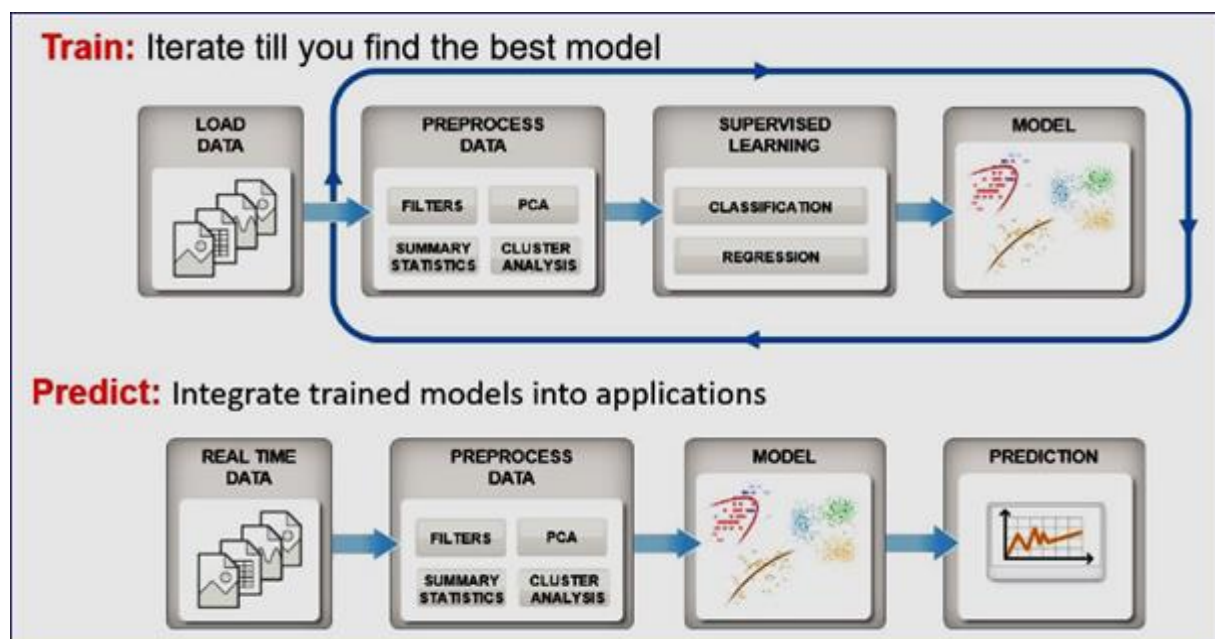
(données de transaction, données issues des d'objets connectés, etc.). La qualité des modèles prédictifs de Machine Learning est liée aux 3 V du Big Data

- **Volume** des données disponibles (transactionnelles ou issues d'objets connectés),
- **Variété** des données (structurées ou non).
- **Vélocité**, car les prédictions doivent s'effectuer en traitement réel ou quasi-temps réel

Dans le cas où l'on dispose de modèles suffisamment représentatifs de la cible à détecter, les data scientists s'orientent plutôt vers une approche supervisée et semi-supervisée.

Dans le cas où la cible n'est pas bien définie (pour des incidents non répertoriés, de nouvelles formes de fraude...), une **approche non supervisée** pourra être utile pour permettre d'inférer de nouveaux modèles qui pourront compléter/affiner les modèles existants.

Ces deux approches se complètent et fournissent tout un arsenal de techniques (statistiques, PCA, réseaux de neurones...) qui offre de solides bases d'amélioration des stratégies de détection traditionnelles.



<https://insidebigdata.com/2016/06/03/how-to-use-analytics-driven-embedded-systems-to-drive-smart-technology-development/>

Ces nouvelles stratégies de détection d'anomalies nécessitent donc que les applications de Machine Learning puissent **analyser les données « au fil de l'eau »** (donc en temps réel), **en s'appuyant sur l'analyse des données de long terme**, stockées dans des bases de données NoSQL ou relationnelles, afin de déclencher aussitôt l'alerte appropriée.

4.4 Les contraintes du traitement temps réel

Lorsqu'on réalise un calcul sur des données massives, on a parfois besoin du résultat dans la journée, dans l'heure grâce à des traitements de type « batch », mais dans un contexte de

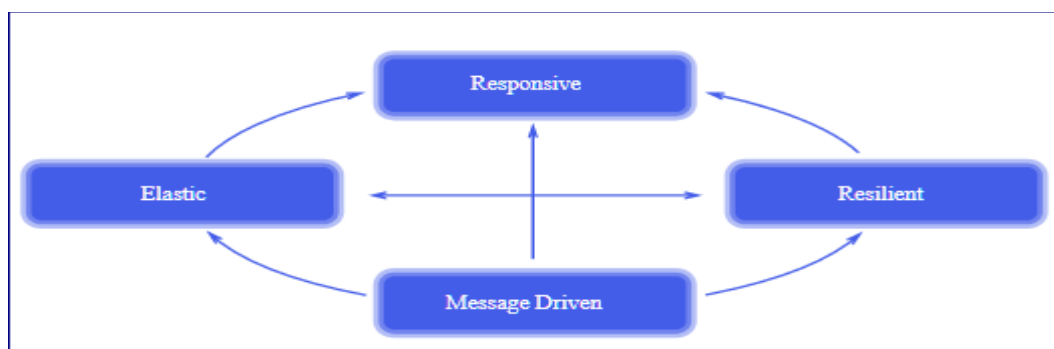
système critique comme la détection de fraude bancaire (et dans bien d'autres...), la réaction doit être immédiate. C'est tout l'enjeu **du traitement temps réel de flux de données en continu**.

En étudiant la séquence de code d'un programme, on s'aperçoit qu'il est souvent possible d'anticiper certaines tâches (thread impliquant des I/O coûteuse en temps) non dépendantes de la tâche en cours pour lancer leur exécution en parallèle. La communication entre les composants des applications « temps réel » doit être **asynchrone** afin de favoriser leur parallélisation. Si, par exemple, le langage utilisé pour le développement des modèles de ML est R, il est important de noter que la parallélisation des algorithmes n'est pas implémentée de manière native, ce qui peut conduire à des problèmes de lenteur des calculs lors de l'entraînement des modèles. Il est donc important que la solution intègre de manière native la parallélisation des traitements afin de proposer aux data scientists de Worldline une architecture qui favorise la parallélisation de leurs algorithmes afin de créer des services temps réel avec une latence la plus faible possible. Face à ce problème, il est recommandé de coder une application en partant d'un modèle d'architecture dont l'un des principes de base est que toute source de données est par essence un flux à traiter en temps réel : on parle de « **programmation réactive** » conforme aux principes du manifeste réactif.

4.5 Le manifeste réactif

Un modèle réactif est un modèle de développement basé sur les principes prônés par le manifeste réactif. Selon ce dernier, un système réactif (architecture, application, composant...) doit être conforme à 4 principes directeurs

:



<http://www.reactivemanifesto.org/>

Une application réactive est :

- **dirigée par les événements**, car les données sont poussées vers les traitements dès qu'elles sont disponibles,
- **responsive** ou capable d'offrir une **expérience utilisateur optimale** en étant disponible en toute circonstance,

- **élastique** grâce à la **scalabilité** des ressources qui s'ajustent automatiquement aux variations de charge de traitement,
- **résiliente** ou robuste afin de mieux tolérer les erreurs et les pannes grâce à l'isolation des composants du système (couplage faible) et leur réplication.

4.6 Les architectures style « microservice »

Un style d'architecture est un ensemble de contraintes qui permettent, lorsqu'elles sont appliquées aux composants d'une architecture, d'optimiser certaines exigences propres au cahier des charges du système à concevoir.

Les architectures microservices répondent à des problèmes récurrents que posent les architectures monolithiques traditionnelles en terme évolutivité, maintenabilité et scalabilité.

Le modèle microservices améliore l'élasticité et la résilience des applications grâce au **couplage lâche entre les composants microservices d'une application**.

Les **principes fondamentaux** des microservices sont les suivants :

- un service ne couvre qu'un **périmètre restreint** et clairement défini de fonctionnalités (techniques ou métiers),
- chaque service est **autonome**, s'exécute sur son propre serveur embarqué et ses données
- chaque service est développé avec **le langage et la base de données les plus adaptés** pour répondre efficacement,
- un service doit être **sans état** (stateful)
- les microservices utilise le plus souvent le protocole HTTP via des **API REST** (application Programming Interface/Representational State Transfer) pour communiquer entre eux au lieu du protocole SOAP utilisé dans les architectures style SOA (Service Oriented Architecture). REST est fréquemment utilisé pour la construction de tout ou partie d'une architecture microservices, car ce style d'architecture n'introduit pas de couplage entre les composants.

Une architecture style microservices repose sur un vaste ensemble de fonctionnalités allant de la découverte de services à la gestion des identités, en passant par la messagerie, le routage réseau, la détection des pannes, la journalisation, le stockage que nous ne pouvons pas toutes détailler dans le cadre de ce mémoire.

5 Méthodes habituellement utilisées dans les projets Big Data

5.1 Les modèles d'architecture adaptés au temps réel

Dans cette partie, je présente les principaux modèles d'architectures et frameworks susceptibles de fournir un support adapté aux exigences de notre projet de traitement de flux de données en temps réel.

L'architecture Lambda

Elle sert à résoudre des problématiques complexes mélangeant traitement temps réel et batchs. Elle est adaptée à la mise en place des projets Big Data en offrant scalabilité et tolérance aux pannes. L'architecture Lambda permet de stocker et de traiter de larges volumes de données (batch) tout en intégrant dans les résultats des batchs des analyses issues des données plus récentes (temps réel).

Une architecture Lambda est composée de 3 couches :

❖ Couche batch (Batch Layer) :

Elle effectue des traitements massifs et réguliers à partir de l'ensemble des données (BD relationnelles, Datawarehouse, NoSQL...) afin de produire des vues consultables par les utilisateurs. La fréquence des traitements ne doit pas être trop importante afin de minimiser les tâches de fusion des résultats afin de constituer les vues.

❖ Couche temps réel (Speed Layer) :

Elle ne traite que les données récentes en réalisant le calcul des vues incrémentales qui vont compléter les vues batch afin de fournir des résultats plus actualisés avec la suppression des vues temps réel obsolète (antérieur à un traitement batch).

❖ Couche de service (Serving Layer) :

Elle sert à stocker et exposer aux clients les vues créées par les couches batch et temps réel. Elle permet aussi de créer dynamiquement des vues par fusion des vues batch/temps réel.

Une alternative de l'architecture Lambda est l'**architecture Kappa**, qui est née en réaction à la complexité de son aînée : elle améliore le système de traitement de flux de données et permet la persistance des données grâce à l'apport de **middleware orienté flots d'événements** (Kafka présenté en 5.2.1). Partant d'un constat que la plupart des solutions sont capables de faire à la fois des traitements temps réel (streaming) et traitements batch, l'architecture Kappa, permet de simplifier l'architecture Lambda, en fusionnant la couche batch et la couche temps réel.

5.1.1 Les principaux frameworks utilisés dans les projets Big Data

L'architecture Lambda est générique, elle est indépendante des technologies à mettre en œuvre car n'importe quel framework de développement ou base NoSQL peut convenir. Il n'y a donc pas de solutions ou de frameworks dédiés à cette architecture, mais un choix de plus en plus large :

- **Stockage**: NoSQL surtout, mais aussi JMS, Kafka, HDFS.
- **Couche Batch**: Hadoop MapReduce, Spark, Flink,
- **Couche Temps réel**: Storm, Spark streaming, Flink, Samza, Tez, etc.
- **Couche de service**: Druid, Cassandra, Hive, HBase, ElasticSearch, etc.

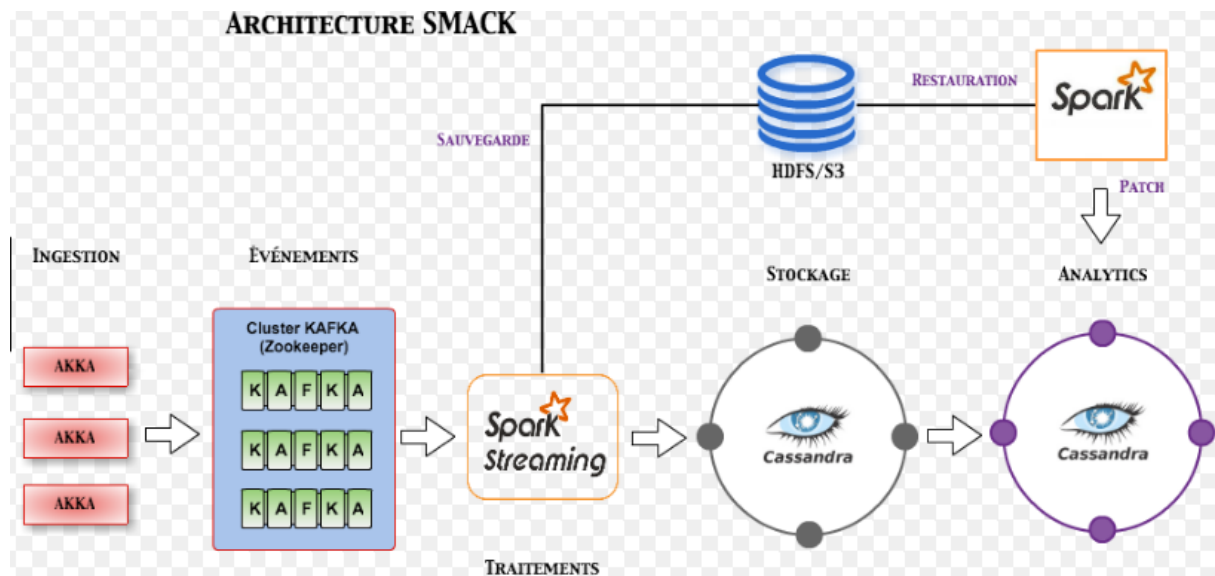
En ce qui concerne le middleware supportant le concept de flot d'événements, de nombreux outils existent désormais, parmi lesquels le leader **Apache Kafka**, RxNetty (basé sur la librairie de communication Netty) ou encore des solutions comme **Spring Cloud Data Flow**. En ce qui concerne les outils de développement, des frameworks de développement de services réactifs existent désormais, issus de la communauté open source. On notera surtout l'existence d'un projet de normalisation du concept de flot d'événements : le **projet Reactive Stream**. Java 9 a annoncé le support de ce standard ainsi que les principaux frameworks de microservices. Un langage de programmation orienté flots d'événements existe et supporte ce standard : **Akka**

5.1.2 La pile SMACK

L'architecture SMACK (pour **Spark-Mesos-Akka-Cassandra-Kafka**) est une solution d'implémentation pour des architectures Kappa. Elle s'appuie sur une pile de solutions plutôt que sur des principes et patterns. Toutefois chacune des solutions est dédiée à une tâche particulière, c'est donc une alternative souple et modulaire. Les frameworks sur lesquels repose cette solution sont matures et répondent aux exigences des 3 V du Big Data :

- **Spark** : Framework de traitement des données (batch) et micro-batch pour Spark streaming
- **Mesos** : Gestion des ressources du cluster (CPU/RAM), haute disponibilité grâce à Zookeeper.
- **Akka** : Implémentation du paradigme acteurs pour la JVM (ingestion des données dans Kafka).
- **Cassandra** : Solution NoSQL (Stockage des données brutes, mais aussi pour des analyses temps réel des données).
- **Kafka** : Stockage des événements/messages échangés (afin d'assurer leur persistance).

Dans le cadre de notre projet, il est intéressant de présenter plus en détail la partie middleware de cette solution, avec les frameworks Akka pour le développement de l'architecture et Kafka pour l'échange des messages entre ses composants.



Source : <https://blog.ippon.fr/2016/03/31/big-data-panorama-des-solutions-2016/>

❖ **Akka est un framework de programmation concurrente « orienté message » hautement distribué et résilient** construit sur la JVM. Akka est conçu de manière à permettre le développement d'applications réactives sans avoir à se préoccuper de la gestion des éléments d'infrastructure sous-jacents (pool de thread, mutexes, deadlocks...). Il repose sur le **modèle d'acteurs**, un modèle de programmation concurrente, dans lequel la charge de travail est répartie entre des entités s'exécutant en parallèle, les acteurs. C'est un modèle dans lequel il n'y a pas d'état partagé, les **acteurs sont isolés** et l'information ne peut circuler que sous forme de messages. Les acteurs reçoivent ces messages et ne peuvent réagir qu'en manipulant les données dans le message (effectuer un calcul ou une transformation sur la donnée), en envoyant un message à d'autres acteurs ou en créant de nouveaux acteurs. La philosophie adoptée en termes de **tolérance de panne** pour ces systèmes est « let it crash ». En effet, les acteurs sont des éléments légers, qu'il est facile d'instancier et de relancer en cas de crash (du fait qu'ils ne partagent pas leur état) et qui tolèrent bien la mise à l'échelle (scalabilité). Pour cela, **Akka impose une supervision** afin de pouvoir gérer le crash d'acteurs : il permet donc la création de systèmes résilients qui s'auto-réparent.

❖ **Apache Kafka** : est un système de messagerie distribué, originellement développé chez LinkedIn, et maintenu au sein de la fondation Apache depuis 2012. Son adoption n'a cessé de croître pour en faire un quasi de-facto standard dans les pipelines de traitement de données temps réels actuels. Kafka est bien adapté à la distribution de **volumes élevés**

d'événements et de données de trace à latence faible. Il fonctionne selon un mode publication-souscription, persistant les données qu'il reçoit. Il est conçu pour monter facilement en charge et supporter des débits de données très importants. Kafka conserve les messages qu'il reçoit dans des topics, correspondant à des catégories de données. On nomme les systèmes qui publient (publish) des données dans des **topics Kafka** des **Producers**. Les **Consumers** sont les systèmes qui vont s'abonner (subscribe) aux données publiées dans les topics Kafka. Tous les composants de la plateforme peuvent être mis à l'échelle et distribués :

- plusieurs producers peuvent émettre des données vers un même topic.
- Les données d'un topic sont partitionnées et répliquées au sein des différents brokers qui sont les nœuds d'un cluster Kafka afin de garantir la tolérance à la panne
- un consumer d'un topic peut lui même être un système distribué (Hadoop, Spark, ...) batch ou temps réel.

Les messages sont publiés par les producers dans un topic Kafka sous la forme d'une séquence de bytes ordonnée et immuable (key bytes[], value bytes[]), ce qui permet de transmettre et de persister les messages sur disque en faisant abstraction de leur format (JSON, Avro,...). Les messages possédant une clé commune seront automatiquement envoyés à la même partition d'un topic. Enfin pour envoyer des messages aux consommateurs, Kafka va s'appuyer sur un mécanisme très performant en zero copy (commande linux sendfile). Les données des messages ne passeront pas par la couche applicative du broker Kafka, passant « directement » de la lecture sur disque à la socket réseau.

5.1.3 Cloudera Oryx 2 une architecture Lambda pour le machine learning

*Le projet **Cloudera Oryx** implémente une architecture Lambda **dédiées au machine learning**.*

Oryx 2 est une plateforme de traitement pour le machine learning appliquée à de vastes ensembles de données. L'architecture de la solution est de type Lambda et utilise les frameworks **Apache Spark** and **Apache Kafka**. Cette solution permet aussi bien de construire des applications de ML que d'utiliser des composants prêts à l'emploi pour filtrer, classer ou encore effectuer des régressions ou du clustering.

Les messages échangés via les topics Kafka ici sont des modèles de Machine Learning sérialisés par PMML. Le Predictive Model Markup Language (PMML) utilise le format XML pour représenter les modèles de datamining (fouille de données).

La couche batch sert **au test et à l'entraînement des modèles** sur les données historiques (stockage HDFS). Le modèle, une fois validé, peut être déployé via Kafka sur la couche

temps réel puis mis à jour (Model Updates) afin de prendre en compte les résultats obtenus à partir des flux de données récentes. La couche de service gère la mise à jour des modèles de Machine Learning des topic Kafka et expose sous la forme d'API REST des services de requêtes sur le modèle persisté en mémoire.

Source: Cloudera Oryx 2 (<http://oryx.io/>).

5.2 Avantages et contraintes d'un modèle d'architecture microservice

Les conséquences de l'adoption d'un modèle microservice sont multiples. Elles sont à la fois techniques et stratégiques, car l'adoption des microservices impacte aussi bien les technologies, le processus de production/livraison du code que la structure et l'organisation des équipes.

5.2.1 Des avantages techniques décisifs

Sur le plan technique les avantages de l'approche microservices pour le projet se situent au niveau de :

- ❖ l'amélioration de la **scalabilité** (et donc l'élasticité), car ce sont des composants autonomes et légers qui sont plus faciles à conteneuriser et à répliquer. Les microservices sont donc bien adaptés à des applications critiques à une forte charge et devant assurer un service continu.
- ❖ L'amélioration de la **résilience**, car :
 - chaque service est hébergé et administré indépendamment.
 - des services ayant un **couplage lâche** entre eux
 - l'**expérience utilisateur** non impactée par une erreur dans un service : les applications doivent pouvoir automatiquement prendre des actions correctives quand une de leurs dépendances est défectueuse.
 - la **communication asynchrone** privilégiée, car elle permet d'accroître le découplage et le parallélisme des traitements, mais aussi facilite la mise en place d'une communication des messages par abonnements entre les microservices.

5.2.2 Des changements importants dans la chaîne de production

Pour créer et exécuter des microservices, les équipes doivent rompre avec l'approche et les outils traditionnels utilisés pour la conception et le développement de logiciels monolithiques. De plus, la gestion et la distribution des microservices et leur nombre introduit de nouvelles formes de complexité:

- ❖ l'approche nécessite un **effort de conception et d'analyse fonctionnelle** (cf. notion de **contexte délimité** ou Bounded Context issu de l'approche *Domain Driven Design*). La

granularité (la taille) des microservices doit permettre la définition d'**unités fonctionnelles autonomes et fortement cohésives**.

❖ il est nécessaire d'**utiliser des mécanismes de découverte automatique et d'indexation des microservices** (registres), et des services centralisés de configuration. La règle étant que les microservices doivent à priori être conçus indépendamment de la connaissance de cette infrastructure, et ne doivent individuellement pas dépendre de ces services pour fonctionner.

❖ **la gestion opérationnelle des microservices est plus complexe**, notamment en ce qui concerne le déploiement, les tests et le monitoring. La mise en œuvre d'une architecture microservices démultiplie le nombre de livrables à déployer. La gestion des livrables devient une tâche complexe qui doit être intégralement automatisée. Cette automatisation passe par la maîtrise de la chaîne de production logicielle, des tests et des déploiements. L'approche microservices est donc indissociable d'une démarche de **livraison continue** (Continuous Delivery) incluant :

- La gestion de configuration
- La construction des livrables
- La virtualisation des environnements et le « provisionning » des infrastructures
- L'automatisation des modifications de schéma, des tests et des déploiements
- le suivi (monitoring) de l'ensemble de la chaîne de production.

Dans ce contexte, **Docker et son écosystème** (Swarm, Kubernetes...) **sont des outils indispensables** ils permettent d'isoler efficacement les microservices, de fiabiliser la gestion de l'environnement d'exécution des applications et d'automatiser les tests et le déploiement.

6 Exposé des décisions prises et des interventions dans le cadre du projet SmartStream

6.1 Le cadrage du projet

6.1.1 Objectifs et contraintes

Mon sujet de mémoire s'inscrit dans le cadre des recherches en « Machine Intelligence » de Worldline présenté précédemment. Il a notamment pour objectif de faciliter et améliorer les processus de détection de fraude déjà intégrés à son offre de services FRM (Fraud Risk Management). L'essentiel de ma réflexion et de mon travail ont porté sur la conception et l'implémentation d'une architecture distribuée « orientée microservices » qui soit adaptée à la intégration de nouveaux modèles de Machine Learning, dans les offres de services numériques de Worldline afin d'en augmenter la valeur ajoutée.

Ce projet soutenu par l'équipe HPV de Lyon est une architecture qui doit être capable d'orchestrer des pipelines d'application « microservices ». Ces applications « microservices » devront pouvoir effectuer des *traitements* « *temps réel* » sur des flux de données à haute volumétrie (datastream), mais aussi des traitements en mode « batch » sur les entrepôts de données de type « Big Data » de Worldline. Cette architecture doit notamment permettre aux data scientists de Worldline d'intégrer en continu de nouveaux modèles de Machine Learning permettant **d'enrichir l'analyse « temps réel »** des flots de données.

Dans un premier temps ma démarche a consisté à mieux cerner :

- les enjeux et contraintes du projet évoqués précédemment (performance, nature des données et mode de traitement, intégration dans les environnements de production existant...),
- leurs conséquences sur le choix d'une architecture qui soit adaptée aux exigences techniques et aux problématiques métiers de Worldline,
- les implications des choix technologiques sur les principes d'organisation et les méthodes de travail utilisées dans l'équipe (DevOps et Agile/Scrum...).

Notre **cahier des charges** comporte des contraintes fortes inhérentes à toute application « réactive » (selon le manifeste réactif) :

- **orientée message** grâce à un middleware supportant les échanges de message à latence faible entre les composants de l'application
- **la résilience**, car la solution doit reposer sur une architecture robuste et tolérante aux pannes grâce à l'isolement des composants microservices (couplage lâche) et leur réplication

- **l'élasticité** par le dimensionnement automatique des ressources en fonction de la charge et l'automatisation de leur déploiement en mode Container as a Service (CaaS)
- **la disponibilité** (i.e. **responsive**) avec un temps de **latence faible** lors de l'exécution des composants microservice de l'application sur le flot de données, ce qui impose :
 - au plus un passage par événement/message
 - temps de traitement par événement court
 - ressources CPU et mémoire utilisées par événement faibles
 - applicable à l'ensemble du flux ou juste une portion de ce dernier (windowing)

Notre middleware pour satisfaire ces contraintes, doit comporter deux éléments essentiels : tout d'abord, nous avons besoin d'un composant dont la tâche va être de recevoir des messages, de les stocker puis de les redistribuer à la demande. Ce composant est une **file d'attente de messages** (*message queue*). Le second composant récupère les données stockées sous forme de messages dans la file et les traite en réalisant les tâches correspondantes : c'est un **système de traitement de flux de données** (*stream processing system*).

La **file d'attente de messages** doit elle-même répondre à plusieurs contraintes :

- 1) L'ajout de messages dans la file doit se faire avec une latence faible et qui ne dépend pas du nombre de messages déjà présents dans la file.
- 2) La lecture des messages doit pouvoir se faire en respectant l'ordre dans lequel les messages sont arrivés.
- 3) Pour stocker les données de manière fiable, une file de messages doit pouvoir stocker ses données de manière redondée et distribuée.

De son côté, le **système de traitement des flux de données** doit également satisfaire à des contraintes :

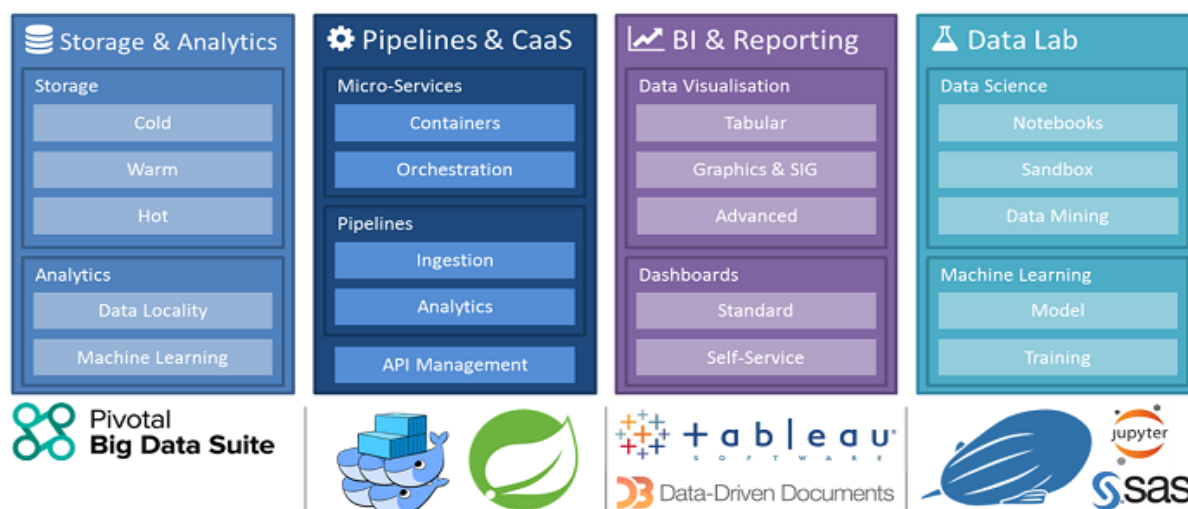
- 1) Chaque étape de traitement des données doit pouvoir se faire de manière distribuée, de manière à passer à l'échelle avec les variations du volume de données ingéré.
- 2) Le système de traitement de données doit avoir accès à un système de stockage (In memory, base de données, stockage de fichiers...) pour stocker le résultat des traitements.
- 3) Il faut pouvoir détecter l'échec de traitement d'un message pour adopter une réaction appropriée ; par exemple, il peut être souhaitable de tenter plusieurs fois de traiter les messages...

Dans un second temps j'ai participé au choix et l'expérimentation de certains des composants middleware de la solution SmartStream. Pour la partie ingestion des flots de données et gestion des pipelines de microservices, nous avons utilisé le framework Spring

Cloud Data Flow. La solution s'intègre à l'environnement et aux outils fournis par la plateforme de préproduction (WLAP) *décrite ci-dessous*. Elle permet de se connecter aux services extérieurs fournis par la plateforme de manière **asynchrone (connexion via des API REST)**.

6.1.2 La plateforme de préproduction WLAP (Worldline Analytics Platform)

Dans le cadre de nos expérimentations, nous utilisons la plateforme de préproduction WLAP gérée par la **division « Business Center BI & Big Data »** de Worldline. Elle nous permet de tester les possibilités d'intégration de la solution SmartStream avec les technologies et les outils utilisés en production par Worldline. Cet environnement met à disposition à la fois des technologies Big Data (Hadoop, Spark ...), un service d'orchestration de conteneur (Docker Swarm), des services de Data Analytics, des outils pour la BI (Business Intelligence) et un DataLab utilisé par les data scientists pour construire et entraîner leurs modèles prédictifs.



Notre utilisation de WLAP comprendra principalement :

- The **Pivotal Big Data Suite**, pour la gestion et le suivi de la plateforme Hadoop et incluant **Pivotal Greenplum**, un data warehousing (entrepôt de données d'entreprise pour la BI accessible par SQL ANSI) et **Pivotal GemFire**, une base de données NoSQL en mémoire (In Memory) scalable et bien adapté à gérer de fortes charges à haut débit et faible latence.

- Le **CaaS** repose principalement sur **Docker**, un outil conçu pour faciliter la création, déploiements et l'exécution des applications à l'aide de conteneur. Plutôt que de créer un système d'exploitation virtuel complet, Docker permet aux applications d'utiliser le même noyau Linux que le système sur lequel elles s'exécutent. Cela augmente significativement la performance et réduit la taille de l'application. Docker est accompagné d'un orchestrateur **Docker Swarm**, grâce auquel nous pouvons déployer sous forme de services la solution de microservices SmartStream conteneurisée pour un traitement des flux de données.

➤ Un **DataLab** qui est un environnement d'analyse à la demande d'une grande variété de sources de données et qui permet ainsi tester des opportunités commerciales immédiates. Cet environnement « **bac à sable** » (sandbox) isolé est réservé aux data scientists pour explorer les données et créer des modèles prédictifs qui peuvent ensuite être poussés sur le CaaS

6.2 La conception et la mise en œuvre du middleware

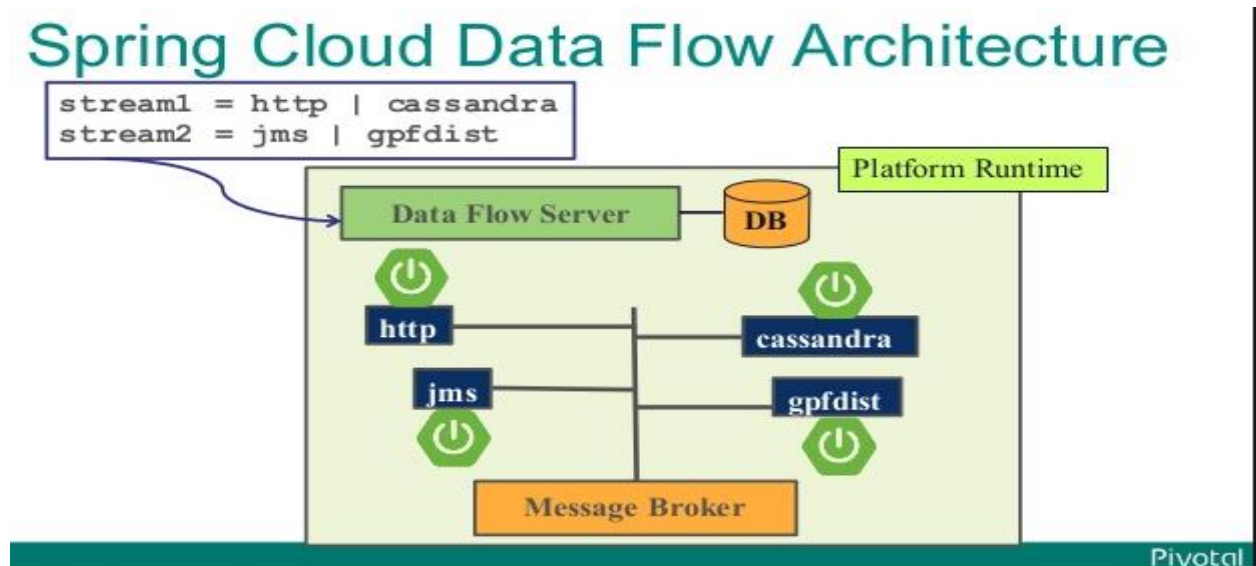
6.2.1 Vue globale de SmartStream

SmartStream s'appuie sur le Framework Spring Cloud Data Flow qui est une solution idéale pour créer des architectures pour le traitement temps réel de flots d'événements et l'analyse prédictive.

[Annexe 1. Vue globale de l'architecture SmartStream](#)

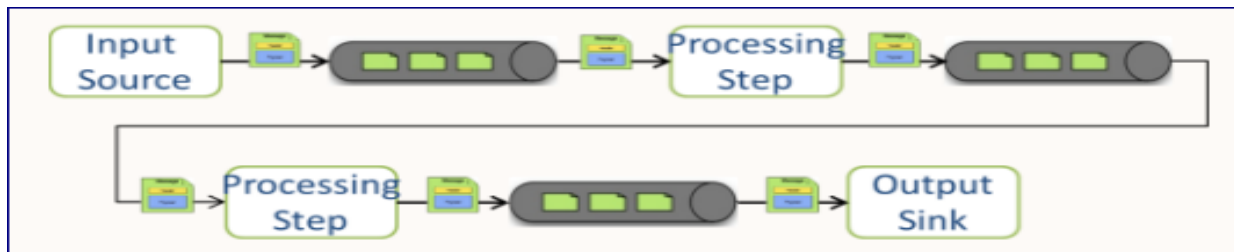
Ce framework offre des fonctionnalités évoluées d'intégration et de traitement de pipelines de microservices temps réel et batch. Les pipelines sont des applications Spring Boot construites avec **Spring Cloud Stream** (traitement temps réel) ou Spring Cloud Task (batch). Dans le premier cas, un flot continu de données est échangé via un middleware de messagerie, alors que dans le second il s'agit de traitements délimités dans le temps et appliqués à des ensembles de données statiques. **Spring Cloud Stream** s'appuie sur Spring Boot pour créer des **middlewares** « **message-driven** » avec de nombreuses options possibles pour les brokers de message Kafka et RabbitMQ

L'implémentation des applications microservice développées peut s'effectuer sur de nombreuses plateformes (serveur local, Cloud Foundry, Kubernetes, Mesos, Yarn ...). Le **Data Flow Server** est le composant responsable de l'exécution des stream DSL qui décrivent la logique de traitement streaming et batch des pipelines.



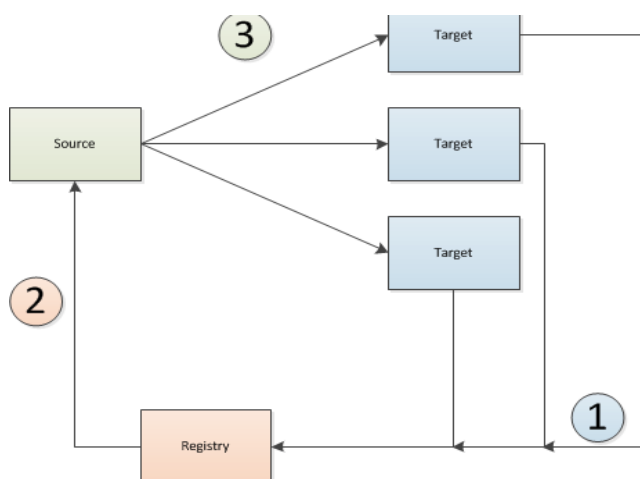
Un **pipeline de microservices** Spring XD est composé de trois différents types d'application :

- **Source** est une application qui ingère/reçoit/injecte la donnée à l'intérieur du pipeline
- **Processor** est l'application qui transforme/effectue le traitement sur les données
- **Sink** est l'application qui conserve le résultat du traitement



❖ Fonctionnement du registre d'application

Le *Registry* (Spring Cloud Data Registry³) est une application Spring boot livrée avec Spring Cloud Data flow permettant de maintenir un annuaire de l'ensemble des applications avec leurs coordonnées réseau (host, port) permettant ainsi de les administrer.



• Initialisation de la topologie réseau

Au démarrage, chaque application s'inscrit au sein du *Registry* d'application (1). Lorsqu'une application *Source* démarre, elle envoie une demande au *Registry* pour recevoir la liste de ses cibles, et se connecte à chacune d'elles directement (2). Finalement, la *Source* peut envoyer des messages directement à ses applications cible, par exemple l'application *Processor* (3).

• Événement au sein de la topologie réseau

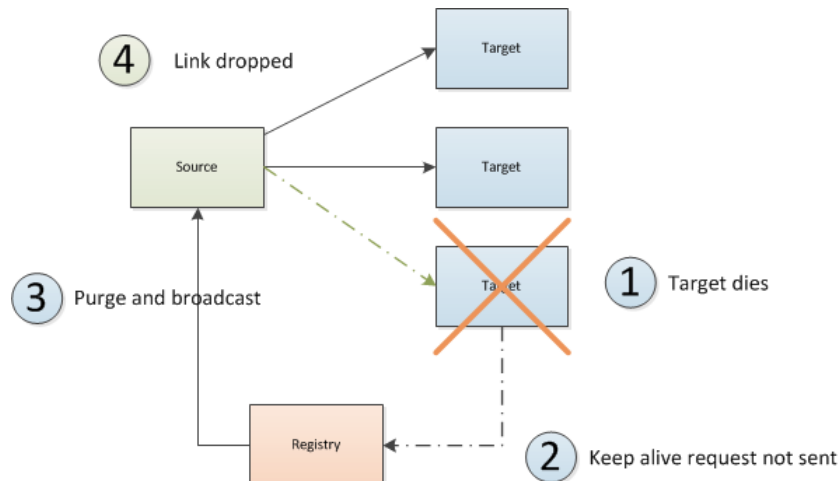
Après cette initialisation de la topologie réseau, les applications cibles enverront régulièrement des demandes d'inscription au *Registry*, donc, si le *Registry* décède, il peut acquérir instantanément la connaissance de toutes les applications vivantes (et donc la topologie réseau) quand il redémarrera. Pendant ce temps, le *Registry* conserve les traces

de la dernière demande de chaque cible. Si elle ne reçoit pas une demande d'une cible, après un certain temps, elle la considérera comme défaillante et la supprimera de son registre. Chaque purge ou ajout de cible (application) du registre est diffusé sur les applications *Source* concernées afin qu'elles puissent intelligemment rediriger leur flux de donnée sur les cibles vivantes.

Cycle de vie d'une application cible :

• Décès d'une cible

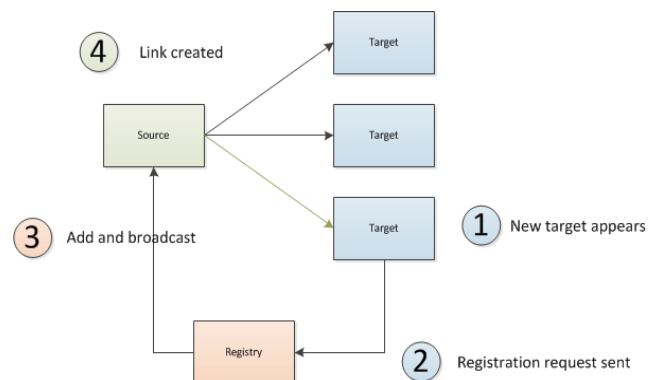
1. Une cible meurt ou son lien réseau dysfonctionne.
2. La cible morte ne sera plus en mesure d'envoyer des demandes d'inscription au *Registry*.
3. Le *Registry* considérera que cette cible a dépassé le délai pour être encore vivante.
4. Le *Registry* supprime la cible de son registre de cibles connues et diffuse l'information sur les applications *Source* présentes dans la topologie.



• Naissance d'une cible

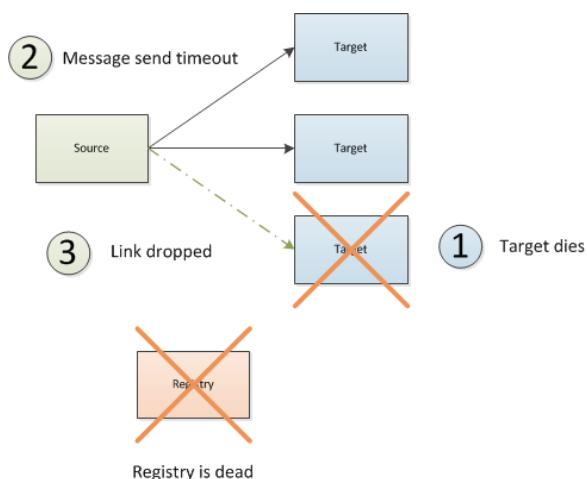
Lors d'une mise à l'échelle, de nouvelles cibles peuvent être créées. Les anciennes cibles considérées comme décédées peuvent réapparaître.

1. Une nouvelle cible apparaît.
2. Elle envoie immédiatement une demande d'inscription au *Registry*.
3. Le registre l'ajoute à sa table cible et diffuse/broadcast l'événement.
4. La *Source* concernée reçoit l'adresse de la nouvelle cible et crée un lien direct avec elle.



• Décès du registre

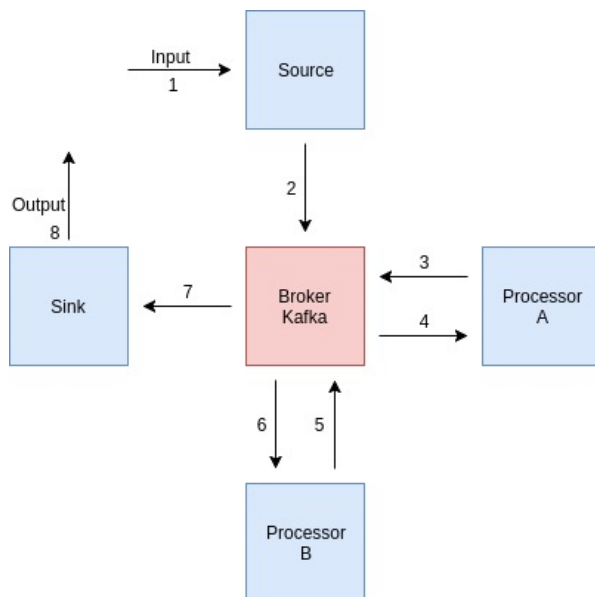
Dans le cas particulier où l'application *Registry* décède, le réseau établi continue de fonctionner, car les applications communiquent directement entre elles. Cependant, les nouvelles cibles ne pourront pas s'inscrire. Le *Registry* ne sera plus capable de diffuser des événements tels que les décès des cibles, cependant la *Source* possède un mécanisme intégré qui lui permet de supprimer les cibles qui ne répondent pas.



6.2.2 La communication « brokerless » avec OMQ (Messaging System)

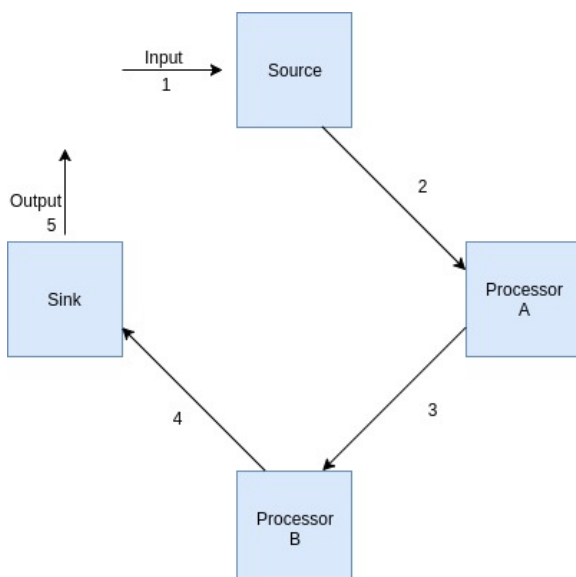
La connexion entre nos applications est un point critique puisque nous opérons dans une architecture Cloud qui inévitablement introduit une certaine latence. Spring Cloud Data Flow intègre différentes solutions de 'broker' qui opèrent les communications entre les différents microservices, telles que Kafka (présenté en 5.2.1) ou encore RabbitMQ.

Ce sont habituellement des **architectures classiques en « étoile » avec un broker** qui centralise et orchestre les échanges de messages. Aucune application ne communique directement avec les autres, toutes les communications passent par le broker.



Ce schéma représente le parcours des transactions à travers un pipeline SmartStream, l'application *Source* ingère les données (1) pour ensuite les transmettre au broker Kafka (2), ce dernier distribue les messages aux applications *Processor A* et *B* (4-6) qui après avoir effectué les traitements sur les données vont transmettre leur résultat au broker (3-5). Finalement le broker envoie les résultats finaux à l'application *Sink* (7). On comptabilise donc **6 échanges** entre applications passant par le broker sur les 8 transmissions totales.

Le problème de ce type d'architecture est le risque de surcharge du broker central qui peut générer des latences de lors de la transmission des messages. **Afin d'optimiser la latence de l'architecture, nous avons choisi d'éliminer le broker: une architecture sans broker est appelée architecture brokerless.**

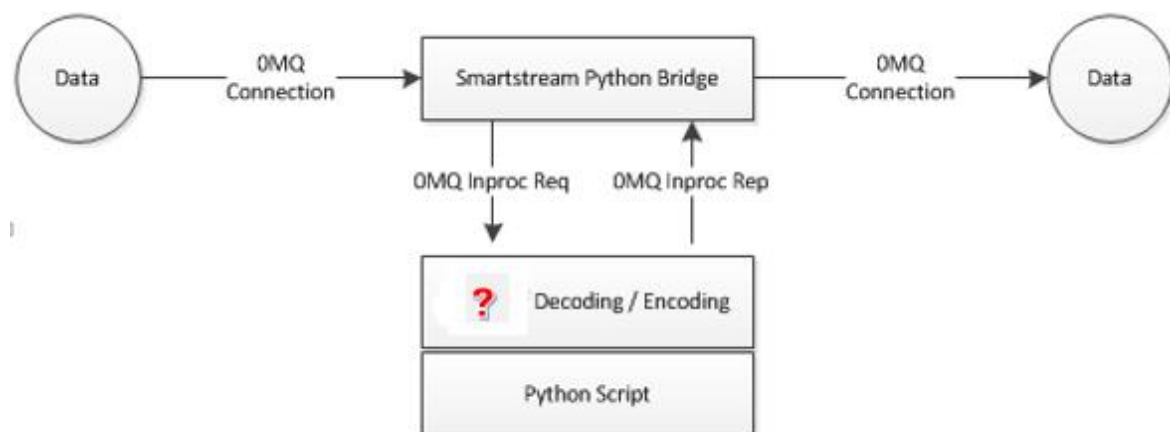


Ce type d'architecture de communication est idéal pour les applications nécessitant une faible latence et/ou un taux de transaction élevé. Bien sûr ce concept nécessite un composant se chargeant de connaître l'adresse réseau de chacune de ces applications, afin que chaque application puisse se connecter directement à l'application avec laquelle elle doit échanger. Comme on peut le voir, le nombre d'échange entre applications a drastiquement diminué et il n'y a plus de risque de goulet d'étranglement au sein de l'architecture.

Nous avons choisi de supprimer le broker Kafka habituellement employé et d'utiliser le système de messagerie **ZeroMQ**.⁴

Notre solution implémente une **connexion tcp directe avec les sockets ZeroMQ**. L'intérêt de ces sockets est d'échanger des données entre des threads d'une manière élégante et sans passer par toutes les contraintes classiques liées aux communications inter-threads ou inter-processus.: **la rapidité des échanges de messages entre les applications Spring est donc augmentée.**

Dans le cadre du projet SmartStream, les modèles prédictifs construits par nos datascientists sont majoritairement écrits en **langage Python**. ZeroMQ est compatible avec la plupart des langages, nous avons donc pu échanger des données sérialisées d'une application Java à une application Python, sans avoir à supporter des étapes de désérialisation intermédiaires. Pour cela nous avons dû **choisir un framework de sérialisation qui soit adapté à nos besoins.**



6.2.3 Optimisation des performances : sérialisation avec Protobuf

❖ Pourquoi la sérialisation est-elle importante ?

Dans un contexte d'analyse de données à haute volumétrie, il est courant que le réseau et l'interconnexion entre les applications agissent comme des goulets d'étranglement. En outre, l'optimisation de la taille de millions de messages constituerait une énorme différence en termes de performance. L'utilisation de **schémas d'objet permet de représenter et structurer les données** transitantes de manière simple et compréhensible,

⁴ **ZeroMQ (OMQ)** est une bibliothèque permettant de créer des applications client-serveur. ZeroMQ met à disposition de nombreux types de connexion.:client-serveur classique ; envoi de données asynchrones avec load-balancing, publish-subscribe, connexion multiple simultanée, inter-threads (au sein d'un même processus), inter-processus (au sein de la même machine) et/ou inter-machines (au sein du même réseau).

et ainsi savoir quel objet est obtenu ou attendu par n'importe quel utilisateur du microservice. On peut spécifier des fonctions de définition de contraintes à appliquer sur chaque champ du message. Ces contraintes portent sur plusieurs aspects comme le type, le nombre d'occurrences ou les cardinalités et les restrictions de champs obligatoires. On peut ainsi définir des modèles de données structurées et les valider au runtime **sans avoir à écrire des lignes de code dans chaque application (sérialisation/désérialisation des données)**. On évite l'écriture de code de validation et on s'abstrait encore un peu plus du langage de programmation pour permettre à n'importe quel service tiers de consommer ou produire le microservice.

❖ Le choix d'un framework de sérialisation

Pour une plateforme microservice, par essence multi-langages, il est important de trouver un système de sérialisation capable d'échanger des données de manière optimale entre différents langages de programmation sans perte, ni altération des messages.

Nous avons choisi trois métriques pour comparer différentes solutions de sérialisation :

- La taille du message une fois qu'il a été sérialisé
- La durée de sérialisation
- La durée de désérialisation

Nous avons cherché à comparer **5 solutions de sérialisation** :

1— Sérialisation Java est la plus simple à utiliser, essentiellement parce qu'elle est intégrée par défaut en Java. C'est aussi la solution la plus flexible en termes de structuration de données en raison de sa capacité à sérialiser tout type d'objet. Cependant, cette solution ne prend pas en charge la sérialisation en plusieurs langages de programmation, mais elle constitue une bonne référence pour nos mesures. Les autres solutions peuvent être considérées comme étant multi-langages.

2— XML (eXtensible Markup Language) encore utilisé dans le cadre d'architecture style SOA (SOAP impose XML), mais trop lourd (verbeux) pour des systèmes critiques (temps réel) comme le nôtre. C'est un format **lisible par l'homme**. Il ne nécessite pas de compilation

3- JavaScript Object Notation (JSON) est un format de données textuelles dérivé de la notation des objets du langage JavaScript. Il permet de représenter de l'information structurée comme le permet XML par exemple. C'est la **solution la plus utilisée en production** en raison de sa syntaxe claire, qui est flexible et lisible par l'homme. Il ne

nécessite pas de compilation et Json est indépendant de tout langage de programmation. Le JSON est un format léger qui permet des échanges rapides de données.

4— Protobuf⁵ est une solution open source développée par Google, basée sur un langage neutre. Cette sérialisation est décrite comme « lisible par machine ». Le principal avantage de Protobuf est son **ouverture**, car il est capable de générer des fichiers de classe pour toutes les langues requises, cette classe peut être générée grâce à un fichier « .proto » qui décrit la structure des données qui doit être sérialisée. Il permet de définir des structures de message ou schémas et de les utiliser pour créer du contenu binaire optimisé.

L'application cliente devra disposer du schéma pour récupérer le contenu original. Protobuf peut être utilisé avec ou sans compilation. Il est nettement **plus léger que XML et plus consistant qu'une description JSON**. Protobuf génère les encodeurs/décodeurs des données sérialisées alors qu'avec JSON, il faut les coder à la main.

5— Avro⁶ est un projet très similaire à Protobuf, il a été développé par Apache et est intégré par défaut dans des projets bien connu comme Hadoop ou Kafka. Avro, à l'instar de Protocol Buffer, utilise des schémas pour convertir le texte en contenu binaire (le plus léger possible). Ce protocole est maintenu et développé par la fondation Apache et connaît un grand essor. Il possède la même philosophie que Protobuf. Contrairement à ce dernier, il permet néanmoins de se passer de la génération de code pour les langages non typés et il permet de schématiser des appels RPC. Il est utilisé par divers acteurs importants, par exemple par la solution Confluent encapsulant le broker Kafka. Avro est donc également un bon candidat en raison de sa flexibilité sur les structures de données qu'il peut traiter.

❖ Analyse comparative et test des sérialisateurs Protobuf et Avro

Il n'y a pas de solution miracle, il faut utiliser le bon outil en fonction du besoin. Protobuf et Avro permettent de s'affranchir du langage et du format pour la transition des données indispensable pour la communication dans des systèmes distribués. Ce sont des **outils qui permettent de mutualiser les schémas des protocoles les rendant utilisables par différentes équipes, avec différents langages de programmation**. Ils intègrent les technologies d'enregistrement de schémas et de validation des applications de services. Ainsi on structure et valide les données transitant dans le flot et on évite les incohérences lors des évolutions de contenu (les schema-registry assurent la rétrocompatibilité).

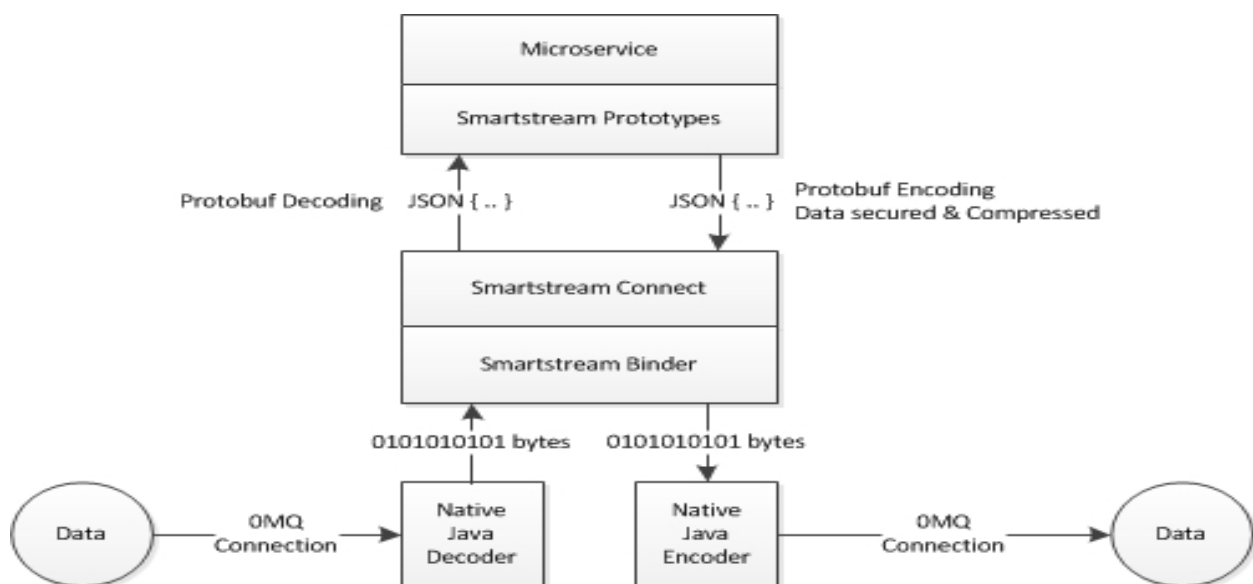
Afin de comparer les deux sérialisateurs, nous avons retenu les indicateurs qualité suivants :

⁵ <https://developers.google.com/protocol-buffers/docs/proto3>

⁶ <http://docs.confluent.io/3.0.0/avro.html>

- **utilisabilité**, Protobuf n'est pas simple à utiliser sur un environnement Microsoft Windows (utilisé sur les laptops des développeurs de micro services), car conçu pour un environnement UNIX, contrairement à Avro, qui a été très facile à utiliser.
- **robustesse**, les données sans schéma spécifié, les données sont bien sérialisées/désérialisées, il n'y a pas de perte d'informations par contre pour des **structures complexes** avec des objets hérités, Avro semblait avoir un problème pour créer un code Java qui fonctionne correctement.
- **performance**, nous avons testé un stream portant sur dix millions de transactions de paiement: chaque message pèse 1 kb, passe à travers 50 microservices, sur un réseau de 10 Gb/s. Avec la sérialisation native de Java, nous avons estimé que la désérialisation, la sérialisation et l'envoi prendront au moins 9 heures, contrairement à la sérialisation de Protobuf qui n'a pris qu'environ 1 heure
- **support et maintenabilité**, à ce jour la communauté **Protobuf** semble être plus active avec environ 250 contributeurs contrairement à **Avro** qui n'a que 33 contributeurs. Le dernier commit sur le projet Avro a été poussé il y a 3 mois et Protobuf est mis à jour tous les jours.

Suite à ces expérimentations, nous avons décidé de retenir Protobuf, en raison de sa performance et de son extensibilité, mais également de la faible robustesse d'Avro sur les structures complexes. Grâce à ces tests, nous avons constaté que l'utilisation d'un framework de sérialisation moderne peut avoir un impact très important sur la performance d'une architecture microservice destinée à traiter des flots de données à forte volumétrie. Le schéma de conception ci-dessous montre l'intégration de Protobuf à SmartStream.



Cette conception nous permet de désérialiser les données au niveau applicatif, ce qui signifie que seul le microservice concerné sait comment interpréter les données. Cette solution présente un double avantage : elle est à la fois performante en terme de latence et sécurisée au niveau des données. Nous avons dû garder le codage natif Java pour les en-têtes des messages en raison de l'incapacité de protobuf à sérialiser et désérialiser des 'immutable Maps'.

[Annexe 2. Benchmarking des solutions de sérialisation](#)

6.3 La mise en place du « Continuous Delivery »

6.3.1 La conteneurisation avec Docker

SmartStream est construit à partir de Spring Cloud Data Flow, développé par Pivotal. Les microservices déployés sont des applications Spring Boot Application qui sont déployées sous forme de conteneurs Docker, ce qui permet d'embarquer tout l'environnement nécessaire au bon fonctionnement du microservice.

6.3.2 L'orchestration et la gestion des clusters avec Swarm

L'orchestration des conteneurs est pour l'instant prise en charge par **Docker Swarm** (service fourni par la plateforme de préproduction WLAP). Docker Swarm est **une solution de clustering** qui transforme un pool d'hôtes Docker en un serveur virtuel unique.

L'**orchestration** des conteneurs permet de passer du déploiement de conteneurs individuellement sur un seul hôte, au déploiement d'applications multi-conteneurs complexes (comprenant les aspects stockage et réseau) sur de nombreuses machines réparties sur différents nœuds. Le composant Service Provider Interface (SPI) de Spring Cloud Data Flow permet de déployer des microservices via le composant **Spring Cloud Deployer** sur différentes plateformes : Cloud Foundry, Kubernetes, Mesos. Le **CaaS** (Container as a Service) fourni par la plateforme WLAP de BCBI est basé sur **Docker Swarm**. Spring Cloud Deployer ne supporte pas nativement cette plateforme d'orchestration de conteneur. J'ai donc travaillé en binôme avec un Ingénieur R&D au développement d'un module « **Service Provider Interface** » (SPI) capable d'interagir avec l'orchestrateur Docker Swarm.

[Annexe 3. L'implémentation de Spring Cloud Deployer Docker Swarm](#)

Tandis que mon collègue s'est focalisé sur les aspects développements du module avec Spring Boot Application, pour ma part je me suis principalement concentré sur la compréhension du fonctionnement du Docker Swarm intégré à WLAP et la façon dont on pouvait faire interagir ce dernier avec notre futur module SPI.

Depuis la version 1.12 de Docker, l'orchestrateur de conteneur SWARM est directement intégré dans l'outil de conteneurisation et permet de déployer des conteneurs sur les différents nœuds du réseau.

Ces nœuds sont gérés par un nœud maître exposant une API REST pouvant déclencher des actions comme créer de nouveaux conteneurs, les supprimer ou plus généralement modifier leur façon de fonctionner. Cette API est le moyen idéal de **faire communiquer l'orchestrateur SWARM avec des services externes comme notre module Service Provider Interface** qui sera ensuite capable de piloter Spring Cloud Deployer.

Pour pouvoir implémenter la communication entre notre SPI et Docker Swarm, il est nécessaire d'utiliser l'API HTTP de Docker Swarm. Nous avons utilisé une librairie Java développée par Spotify qui permet des requêtes sur cette API.

[Annexe 4. Connexion de Spring Cloud Deployer avec Docker Swarm](#)

6.3.3 Le Build & Run avec GitlabCI

Le déploiement et l'intégration sont effectuées en continu à l'aide d'un outil fournit par la plateforme WLAP, **GitLab CI**, qui est une extension à la célèbre forge GitLab. Le principe est relativement simple : l'ingénieur développe sur sa machine physique (le plus souvent un ordinateur fournit par l'entreprise), et chaque modification, suppression ou création de code sur le serveur de versionning GitLab est considéré comme un événement déclenchant un « **pipeline de process** ». Ce pipeline est ni plus ni moins qu'un enchaînement d'actions comprenant la compilation du logiciel prenant en compte les dernières modifications (**build**) puis le déploiement (**deploy**) de ce logiciel sur les infrastructures de production pour exécution (**run**).

Pour indiquer à GitLabCI les différentes étapes de construction puis de déploiement, j'ai rédigé des **fichiers standardisés « .gitlab-ci.yml »** qui sont placés à la racine des projets. Les « .gitlab-ci.yml » sont interprétés par des « runners » (agents exécutifs) qui se chargent d'effectuer les actions nécessaires au déploiement et à l'intégration en continu des nouvelles versions. [Annexe 5. Fonctionnement de GitLab CI](#)

6.4 Le Datalab avec Jupyter Hub

J'ai également travaillé sur l'optimisation du workflow des modèles de machine learning utilisés par les datascientists dans SmartStream. Avant mes travaux les datascientists effectuaient leurs expérimentations de modèle prédictif sur leur machine **en local**, directement dans leur IDE. Cette méthode de travail les contraignait à installer et maintenir plusieurs environnements de travail pour accéder à tous les outils nécessaires à leurs projets.

Notre choix s'est orienté vers **Jupyter Notebook**, une application Web qui permet de créer et de partager les documents qui contiennent les « live code », équations, visualisations ainsi que des textes explicatifs pour les data-scientists. Ils permettent le nettoyage et la transformation des données, les simulations numériques, les modélisations statistiques et le machine learning. J'ai choisi de déployer ces environnements dans des **conteneurs Docker** afin d'isoler efficacement les utilisateurs avec des inconvénients mineurs en performance. J'ai utilisé **JupyterHub**⁷, un composant intermédiaire entre l'utilisateur et le notebook Jupyter qui permet de gérer l'accès multi-utilisateurs.

[Annexe 6. Fonctionnement de JupyterHub](#)

La communauté Jupyter rend disponible plusieurs systèmes d'authentification.⁸ Les datascientists bénéficient d'un accès sécurisé par authentification via un annuaire **LDAPauthenticator**⁹ que j'ai adapté afin de permettre un accès personnalisé à partir de l'annuaire de Worldline. Le nom de l'utilisateur est transmis depuis le système d'authentification à un spawner¹⁰ et permet une **authentification personnalisée**. J'ai utilisé **DockerSpawner**¹¹ afin de permettre à JupyterHub de créer et déployer les conteneurs. Ces conteneurs Jupyter vont permettre aux data scientists d'accéder à divers kernels de développement (Scala, Java and Python) et de nombreuses librairies comme Numpy, Ggplot, ou encore Tensorflow contenus dans le notebook. La **solution est à la fois flexible et robuste**, car elle permet une isolation complète des processus utilisateurs.

⁷ Utilisé pour le déploiement d'environnement de travail collaboratif dans les entreprises, les universités, ainsi que les laboratoires de recherche.

⁸ <https://github.com/jupyterhub/jupyterhub/wiki/Authenticators>

⁹ <https://github.com/jupyterhub/ldapauthenticator>

¹⁰ Dans JupyterHub, un spawner peut être vu comme un générateur d'instances de notebook Jupyter.

¹¹ <https://github.com/jupyterhub/dockerspawner>

7 Présentation du caractère original de la solution SmartStream

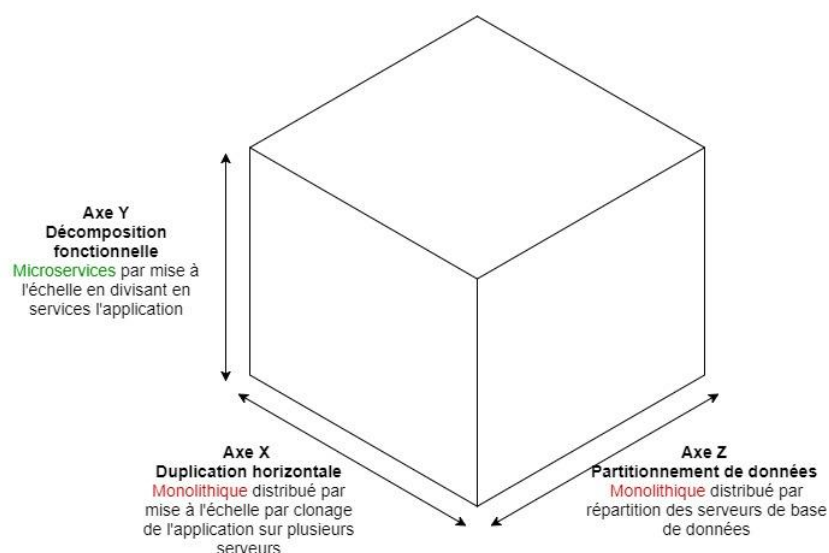
L'utilisation conjointe des techniques de conteneurisation et des microservices dans SmartStream améliore considérablement les capacités de la plateforme. Microservices et conteneurs peuvent fonctionner de façon indépendante, mais leur association améliore la fréquence des livraisons réduisant les cycles d'intégration et de déploiement des composants logiciels. Les microservices sont évolutifs et réutilisables, tandis que les conteneurs apportent une solution efficace en termes d'élasticité et d'industrialisation de la chaîne de production.

7.1 Une architecture ouverte, évolutive, élastique et résiliente

SmartStream est ouvert à la connexion aux outils externes (API REST) contenus dans la plateforme de production WLAP afin de profiter de l'écosystème Big Data de Worldline.

SmartStream est une solution également **résiliente** : les microservices simplifient la surveillance du fonctionnement et garantissent la robustesse de l'application, car les différents composants sont isolés : si un microservice échoue, les autres continueront de fonctionner.

SmartStream est une architecture élastique, car elle permet une mise à l'échelle (scalabilité) de ses composants selon l'axe Y du **cube de scalabilité** :



En effet la granularité des pipelines de microservices de SmartStream (Service *Source*, Service *Processor*, Service *Sink*) permet d'effectuer des mises à l'échelle automatique par type de microservice sans compromettre l'intégrité de l'application. L'utilisation des technologies de conteneurisation apporte d'autre part la scalabilité horizontale en permettant

le clonage des applications et des composants par conteneurisation. Les composants d'une même application SmartStream peuvent être déployés automatiquement sous forme de conteneur sur plusieurs serveurs ou plusieurs centres de données.

7.2 Des équipes autonomes et pluridisciplinaires.

Dans un modèle microservice, des équipes autonomes gèrent le cycle de vie complet d'un service (développement, lancement, exécution, maintenance...). Les architectures microservices conduisent à un découpage en **unités fonctionnelles ayant chacune leur propre pile technologique**, plutôt que le découpage en couches des applications monolithiques. Cela impose donc de disposer d'équipes capables de résoudre des problèmes tant au niveau du middleware que de la base de données ou du front end. Il faut donc adapter l'organisation.

Cela augmente l'autonomie des développeurs, car les idées peuvent être mises en œuvre et déployées sans avoir à se coordonner avec les process de livraison longs, lourds et coûteux des applications monolithiques. Une meilleure réactivité est alors possible au niveau des modifications du code et du déploiement de nouvelles versions en production. Cette nouvelle organisation permet également une **montée en compétence** plus efficace des équipes, car elles peuvent se concentrer sur un périmètre fonctionnel plus restreint. Ce modèle offre donc des avantages certains pour la production et la **gestion de services critiques et complexe comme la détection de fraude**. Ils sont plus simples à comprendre qu'une application traditionnelle et seront donc plus **faciles à faire évoluer** pour les développeurs, ils peuvent donc aider un nouveau membre de l'équipe à devenir rapidement opérationnel.

7.3 Une approche bien adaptée aux projets R&D

Traditionnellement dans le cadre d'une solution monolithique, une équipe de développement travaille sur deux ou trois versions majeures par an. L'innovation est donc rythmée par la sortie de ces versions contrairement à notre système qui permet d'intégrer de nouvelles fonctionnalités en continu après les avoir construites et testées. Les microservices privilégient une **démarche itérative à cycle court** et permettent davantage d'agilité dans les projets. Ils sont donc bien adaptés à des **contextes comme la R&D fortement innovants et/ou les besoins métiers sont très évolutifs**. Nous avons donc mis en place un workflow très efficace au niveau des outils/process avec des déploiements les plus automatisés possible des modèles de machine learning. SmartStream grâce à sa connexion à un orchestrateur comme Docker Swarm, permet le déploiement automatisé et l'orchestration d'applications de microservices conteneurisés. Cette industrialisation de la livraison permet de **réduire les erreurs et les délais améliorant ainsi le « time to market »**. **Ce type d'environnement est**

primordial en R&D pour les tests et les POC (Proof of Concept) où l'on a besoin d'environnements « jetables » comme les sandbox. Réduire les cycles de livraison de version et ainsi **réduire le « time to market »** est un enjeu stratégique pour Worldline et un des fondements du DevOps. Finalement le DevOps accélère considérablement notre innovation au sein des projets et banalise la mise en production, qui était dans le passé, un événement pouvant être laborieux et difficile.

7.4 Des changements qui soulèvent des difficultés

7.4.1 Une remise en cause des habitudes de travail

Les pratiques DevOps bien qu'elles soient actuellement reconnues comme un facteur de performance sont difficiles à mettre en place au sein d'une organisation, car elles remettent en cause les habitudes de travail. On passe d'une vision projet (à long terme) à une vision « produits » de plus court terme. Elles imposent également aux membres des équipes de production une **montée en compétence** car ils doivent pouvoir intervenir sur toutes les couches technologiques du projet. Lorsque nous nous sommes tournés vers le pôle d'exploitation de Worldline (Production Services), notre méthode de fonctionnement « DevOps » (avec GitLabCI) et les piles technologiques utilisées (Conteneurisation, Spring Cloud Dataflow...) ont parfois créé des réticences. Heureusement, nous avons pu collaborer avec le pôle BCBI et bénéficier des outils fournis par la Worldline Analytics Platform.

7.4.2 Le « turn-over » de l'équipe

Une autre difficulté rencontrée durant le projet fut le « **Turn-over** » important au sein de l'équipe R&D. De nombreux départs et arrivées dans l'équipe nous ont contraints à effectuer très régulièrement des « passages de connaissance ». Cette situation entraîne des pertes de temps au début, mais peut ensuite se révéler positive, car les nouveaux collègues apportent de nouvelles compétences et des points de vue originaux.

7.4.3 Améliorer le module CaaS (Conteneur as a Service)

En m'appuyant sur mon expérience des orchestrateurs de conteneurs et sur les retours de collègues, je pense que **Docker Swarm n'est pas forcément le meilleur choix** pour la mise en œuvre de nos pipelines de déploiement automatisé. En effet Docker Swarm apporte la facilité du management des nœuds, services et conteneurs, mais reste une « boîte noire » avec une simple API unifiant tous les nœuds du cluster Docker Swarm. Cette API souffre d'un manque de fonctionnalités pour certaines tâches comme l'auto-scaling. Nous avons également constaté un manque de robustesse engendrant des instabilités quelques fois inexplicables.

Aujourd'hui, peu d'entreprises se tournent vers Docker Swarm pour gérer le déploiement et le management de leurs conteneurs en production, et préfèrent d'autres solutions, notamment **Kubernetes**, un produit développé par Google.

8 Analyse de l'approche choisie: une solution prometteuse

8.1 Résultats obtenus

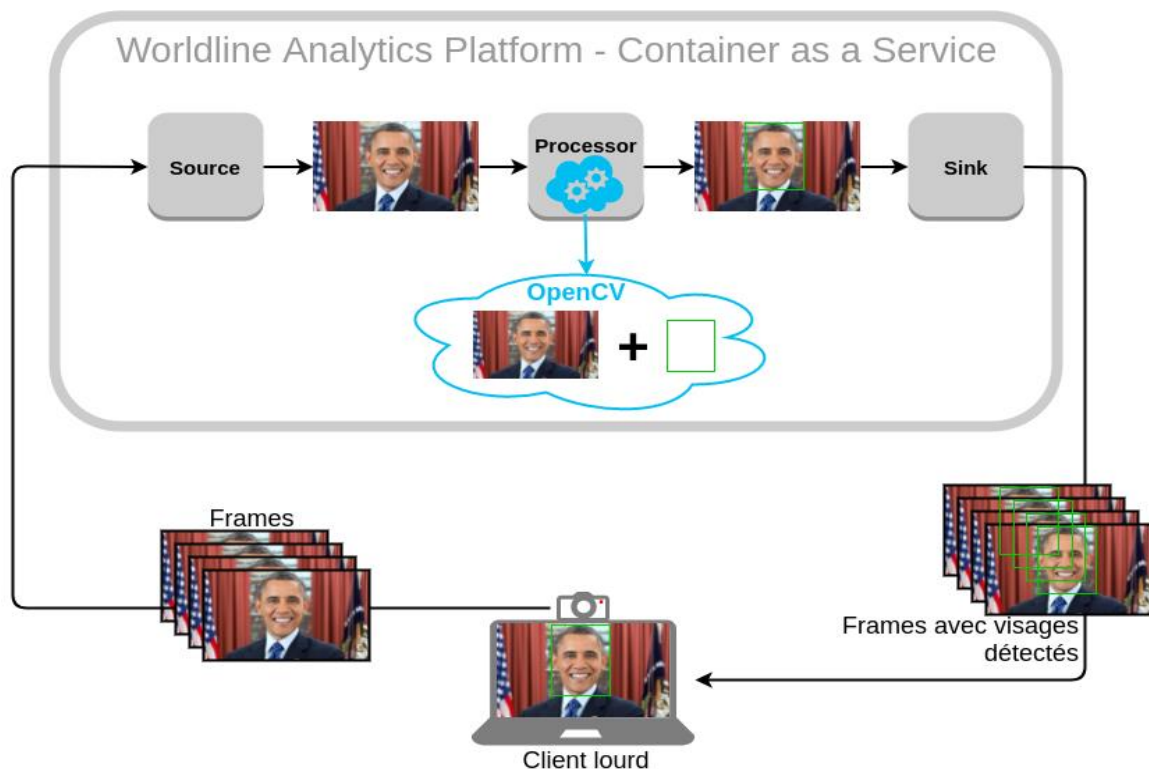
❖ Smart Stream Coré terminé avec deux brevets

Actuellement l'équipe R&D HPV a terminé le développement de l'ensemble des composants qui sont au cœur de SmartStream:

- la communication inter- microservices avec ZeroMQ,
- le déploiement automatisé des microservices sur l'environnement de production (CaaS)
- l'intégration du DataLab permettant l'utilisation des outils nécessaires à la création des modèles de Machine Learning par les datascientist. Une première release¹² de SmartStream est aujourd'hui disponible en interne.

❖ Demo247 fonctionnant sur un CaaS

Afin de montrer l'efficacité de SmartStream dans le cadre de traitements de flux de données en streaming, nous avons créé une démonstration d'analyse de flux vidéo mettant en œuvre un modèle de détection de visage humain.



¹²Release : étapes de développement d'un logiciel informatique : allant de son développement initial à sa version finale.

L'utilisateur installe un client lourd accédant à la webcam de l'ordinateur, pour ensuite envoyer des frames d'images (un nombre fixe d'images par seconde assez important pour constituer un flux vidéo) à un microservice *Source* qui est le point d'entrée du pipeline SmartStream. Une fois les frames transmises au microservice *Processor*, ce dernier effectue la **détection de visage humain présent dans les images** à l'aide d'une librairie de traitement d'image, **OpenCV** (Open Computer Vision). Le résultat du traitement (on appose un rectangle vert autour de chaque visage détecté) est transmis du microservice *Processor* au microservice *Sink* qui comme l'application *Source*, communique avec le client lourd de l'utilisateur. L'utilisateur pourra ainsi recevoir l'image traitée et visualiser les détections de visages dans l'interface de son client. Dans le cas du traitement de flux vidéo, c'est généralement le *Processor* qui exige le plus de ressource. Notre architecture orientée microservice, nous donne la possibilité de multiplier le nombre de *Processor* (scaling up) autant que nécessaire afin d'éviter des latences.

❖ DataLab avec entraînement sur GPU

À la fin de mon alternance, le Datalab permettant l'expérimentation en environnement isolé de la production est opérationnel, il est actuellement hébergé au côté de SmartStream sur le CaaS de WLAP. À l'avenir il migrera sur des machines dédiées (afin que les besoins en ressources soient uniquement consacrés à l'expérimentation des modèles prédictifs).

Le Datalab permet maintenant d'authentifier les datascientist à l'aide du module relié au serveur LDAP de Worldline ([Annexe 7. Fonctionnement de JupyterHub - Figure 1](#)).

Après l'authentification, selon son besoin, l'utilisateur choisi entre plusieurs environnements ayant des librairies et outils différents ([Annexe 7. Fonctionnement de JupyterHub - Figure 2](#)).

Le conteneur est ensuite déployé permettant une utilisation simple et immédiate de son environnement de travail. ([Annexe 7. Fonctionnement de JupyterHub - Figure 3](#)).

La fonctionnalité la plus utilisée est le notebook, qui se présente comme un IDE embarquant toutes les librairies nécessaires à la construction des modèles de Machine Learning ([Annexe 7. Fonctionnement de JupyterHub - Figure 4](#)).

Jupyter permet également aux utilisateurs d'accéder au terminal en ligne de commande bash (shell Unix) sans quitter l'interface Web ([Annexe 7. Fonctionnement de JupyterHub - Figure 5](#)).

8.2 Analyse du champ d'application de la solution élaborée

Dans SmartStream, le code des microservices peut être écrit dans différents langages de programmation au sein d'une même application. Cela permet de ne pas se limiter à une technologie, mais de choisir en fonction de la nature du problème, l'outil le mieux adapté.

L'architecture SmartStream apporte également aux développeurs la liberté de déployer des services indépendamment des services déjà existant. L'approche microservice est bien adaptée à faire évoluer à moindre coût certains services monolithiques. Les API REST, très flexibles et quasi universellement supportées, **facilitent l'intégration dans les applications existantes**. À l'avenir, il est probable que les services prédictifs seront de plus en plus présents dans les offres de Worldline ce qui nécessitera le recours à des architectures comme SmartStream.

8.3 Mise en perspective avec d'autres contextes

L'équipe R&D HPV accueille des doctorants en CIFRE¹³ mais également des docteurs diplômés ayant des connaissances pointues dans leur domaine de prédilection. Un chercheur de l'équipe, spécialisé dans le traitement d'image par ordinateur, investit sur la **détection prédictive de cancer** à partir de clichés médicaux. Nous pourrions dans un futur proche intégrer son modèle de détection de cellules cancéreuses au sein de SmartStream, et ainsi créer un service de détection prédictive à la demande qui pourra être utilisé par les hôpitaux.

L'utilisation d'une architecture sans broker permet de déployer les microservices de façon décentralisée, **Siemens** prévoit d'utiliser SmartStream associé à de la réalité augmentée pour **la maintenance prédictive** de ses équipements industriels.

De manière plus générale, le caractère hautement distribué et scalable de SmartStream en fait également un bon middleware pour le déploiement de **microservices de machine learning pour les objets connectés (IoT)** dans le cadre du « Fog Computing »

¹³ En France, une convention industrielle de formation par la recherche (CIFRE) est un dispositif de financement de thèse qui aide les entreprises pour le recrutement de jeune chercheur-doctorant.

9 Réflexion sur le stage et le mémoire: un parcours riche et stimulant

9.1 Auto-évaluation du travail réalisé

J'ai beaucoup appris du travail **en équipe** que ce soit dans le cadre de mes études ou pendant mes stages ou alternances. Au cours de **mes études à SUPINFO International University**, j'ai effectué de nombreux projets de groupe qui m'ont permis de bénéficier de la mise en commun des connaissances et des savoirs-faire. Ces projets informatiques consistaient à répondre à des cahiers des charges de spécifications fonctionnelles et techniques nécessitant la coordination de compétences techniques variées. J'ai donc été amené à coopérer avec les autres membres de l'équipe en prenant en charge, après découpage, certaines parties du projet. Dans de nombreux cas, j'ai eu à **argumenter les choix de conception et de production des solutions** au sein de l'équipe, pour ensuite les argumenter face aux différents jurys. Cela m'a permis de mieux comprendre les difficultés liées à la prise de décision dans un groupe où souvent les points de vue divergent. Je me suis également rendu compte de l'importance de la cohésion de l'équipe dans la réussite du projet quand, parfois, il faut « trancher » pour avancer.

Au cours de mon parcours professionnalisant, j'ai dû m'adapter à des contextes très différents. Je suis passé avec succès d'un poste d'Administrateur Infrastructure au sein d'une startup à celui d'alternant ingénieur R&D dans un grand groupe international. Au sein de Worldline, j'ai également su faire preuve d'**adaptabilité** lors de ma mission de trois mois à Aachen, en Allemagne, afin de m'intégrer à une équipe multiculturelle. Je suis également souvent **force de proposition** au sein de mon équipe en suggérant à mes collègues d'autres alternatives sur différents aspects de nos projets, par exemple l'utilisation de nouveaux outils ou technologies. Enfin mes managers successifs ont apprécié mon engagement et mon implication au travail.

La rédaction de ce mémoire a été l'occasion de **prendre du recul sur mes deux années passées au sein de Worldline**. Cela m'a permis de mieux comprendre les activités de Worldline, sa stratégie et la place de la R&D dans son business model. J'ai également mené une réflexion enrichissante sur les problématiques liées à la lutte contre la fraude dans le secteur des services de paiement.

9.2 Bilan des acquis sur les aspects techniques, stratégiques et managériaux

Dans le cadre d'une **mission en Allemagne** (Worldline Aachen) de 3 mois, j'ai eu pour rôle de **piloter la mise en place d'une plateforme d'outils pour data scientists** permettant la conception et l'expérimentation d'algorithmes de prédiction de comportement, sur la base de données massives (Big Data Analytics). Je devais étudier les besoins des data scientists pour ensuite intégrer ou créer des outils adaptés. Il fallait ensuite coordonner leur mise en production en lien avec les différentes entités de Worldline. Ce travail m'a appris à recueillir, analyser et structurer des besoins, mais aussi **travaillé en mode agile**. Grâce à ce projet, j'ai eu l'opportunité de piloter un projet du début à la fin en quasi totale autonomie. Au sein de l'équipe, j'ai su être initiateur de changement en étant **force de proposition** sur l'adoption de technologies innovantes, notamment en matière de conteneurisation de logiciel utilisant l'écosystème Docker. Savoir convaincre est également important dans le cadre de projet nécessitant l'adoption de nouvelles technologies impliquant des changements dans l'organisation et des méthodes de travail. En effet, j'ai dû procéder à un **travail d'évangélisation** auprès des différents membres de l'équipe, afin qu'ils perçoivent l'intérêt pour eux des transformations que je leur proposais.

Actuellement en tant qu'alternant ingénieur R&D à Worldline, je me situe à l'interface entre l'équipe chargée de l'infrastructure Big Data de la plateforme et les data scientists. Je participe également au travail collectif de développement de l'architecture de la solution. Cette expérience m'enseigne au quotidien qu'au-delà de l'expertise technique nécessaire, il est important que l'équipe soit soudée et partage la même vision du projet

9.3 Perspectives professionnelles en relation avec les compétences acquises.

Évoluant dans un contexte où l'innovation joue un rôle central, je souhaite enrichir mon profil technique orienté vers l'ingénierie informatique, de compétences en business développement et management de l'innovation technologique. Ce projet de formation, mûrement réfléchi depuis un an, s'inscrit dans le prolongement de mes deux années d'expérience professionnelle au sein du département R&D de la société Worldline. J'ai donc postulé et été admis au **Mastère Spécialisé « Management Technologique et de l'Innovation »** (Ms MTI) proposé par Grenoble École de Management qui propose cette formation en alternance. Ce programme est, pour moi, l'opportunité de capitaliser sur l'expérience acquise en R&D pour mieux comprendre les enjeux et les problématiques liées à l'innovation et acquérir les compétences nécessaires au développement d'offres de services numériques innovantes. Le programme du Ms MTI correspond également parfaitement à ma volonté d'évoluer, à plus long

terme, vers des postes à responsabilité où la double compétence techno-managériale est essentielle.

Ce programme est le prolongement cohérent des deux années d'alternance passées au sein du département R&D de Worldline Lyon. Intégrer un mastère spécialisé en alternance est une étape importante dans la réalisation de mon projet professionnel, car c'est l'opportunité de mettre en pratique les enseignements dispensés dans le programme dans un contexte professionnel riche et stimulant. Le programme correspond donc à la fois à mon projet professionnel à court terme au sein de Worldline ainsi qu'à ma **volonté d'élargir mon horizon professionnel** et me faciliter l'accès à des postes à responsabilité à plus long terme.

Dans l'optique d'affiner mon projet professionnel, j'ai eu de nombreux échanges avec des professionnels impliqués dans l'innovation au sein et en dehors de ma société. J'ai ainsi discuté avec des Business Developers autour des futurs projets de Worldline, comme l'intégration de chatbots dans les offres Big Data, des innovations dans le domaine du Digital Banking ou encore de l'E.-ticketing. J'ai également rencontré des responsables innovation appartenant au groupe Atos qui m'ont exposé leur point de vue sur le programme du Ms MTI proposé par GEM et son intérêt, notamment dans le cadre du développement du réseau Atos BTIC (Business Technology & Innovation Centers). À l'issue de ces rencontres, j'ai reçu plusieurs propositions d'alternance intéressantes et cohérentes avec le MS. **J'ai finalement retenu une proposition émanant de la Business Center BI & Big Data de Worldline Lyon.**

Ma future mission consistera à participer à la mise en place des projets autour de la Data chez Worldline à différents niveaux : modèles économiques, business développement, organisation et assistance au suivi de l'innovation technologique entre les différentes entités (R&D, DataLab, Production...). L'alternance dans ce service en tant que Product Manager junior me permettrait notamment de **continuer à accompagner le projet de « plateforme intelligente pour la détection d'anomalies »** après avoir participé à son développement en tant qu'alternant ingénieur R&D.

Conclusion

Ce travail m'a permis de constater l'ampleur du phénomène de la fraude et des risques qu'elle entraîne dans le secteur des E.-paiements avec la multiplication des transactions électroniques. La lutte contre la fraude est devenue un enjeu de plus en plus important pour Worldline. La fraude se complexifiant, les moyens de détection en place doivent s'adapter et Worldline se trouve confronté à la nécessité de faire évoluer les solutions en place. Les avancées en matière d'apprentissage automatique (Machine Learning/ Deep learning) et les possibilités offertes les technologies Big Data offrent de nouvelles perspectives en matière d'analyse prédictive.

Ces nouvelles approches imposent la mise en œuvre d'**architectures cloud réactives**¹⁴ (« event driven ») similaires à celle du **projet SmartStream** que je présente dans ce mémoire. Dans cette optique le choix des composants et l'implémentation d'un **middleware à base de microservices** ont été guidés par les contraintes qu'introduisent les **traitements analytiques en temps réel de flots continus de données à haute volumétrie**.

J'ai plus particulièrement participé aux travaux d'optimisation du système de communication entre les microservices de SmartStream grâce à une **approche « brokerless »** avec le framework OMQ associé à la mise en œuvre d'un **mécanisme de sérialisation sûr et performant** (Protobuf). J'ai également contribué de manière significative à la mise en place de nouveaux process dans une optique de « Continuous Delivery ». Dans cette perspective, j'ai collaboré à l'adaptation du module **Spring Cloud Deployer** de SmartStream pour qu'il puisse communiquer avec l'orchestrateur de conteneur fourni par notre environnement d'intégration (WLAP). J'ai ensuite automatisé l'intégration et le déploiement des travaux d'ingénierie dans des pipelines de process avec **GitLab CI**.

J'ai également été chargé de la mise en place d'un **JupyterHub** pour les data scientists. J'ai configuré et adapté différents modules (**DockerSpawner, LDAPAuthenticator**) afin de créer un véritable service de déploiement d'environnement de travail à la demande. Le datalab permet donc maintenant aux datascientists d'effectuer leurs travaux sur les données de manière sécurisée (isolée) grâce aux conteneurs *notebook Jupyter*. Enfin la solution est générique, car elle couvre un large spectre d'applications potentielles et ouvre donc de nombreuses perspectives.

¹⁴ en accord avec les principes du manifeste réactif

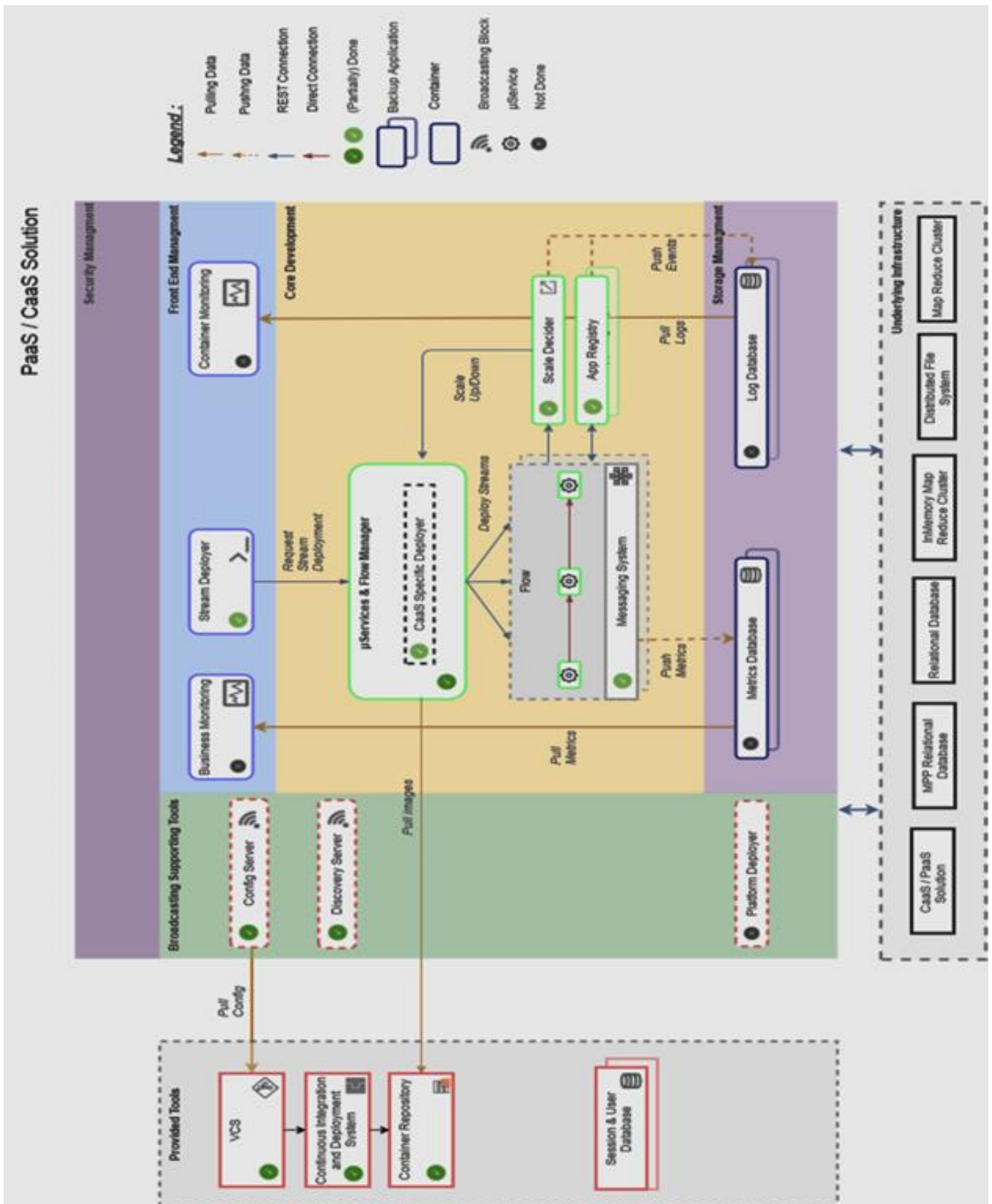
Bibliographie et Webographie

- Worldline Document de Référence 2016:
<https://worldline.com/content/dam/worldline/documents/investors/amf-regulated-information/registration/worldline-document-de-reference-2016.pdf>
- Gartner Trend Insight Report - Top 10 Strategic Technology Trends for 2017 Published: 21 March 2017 (<https://www.gartner.com/doc/3645332?srcId=1-8362264702#a-349750886>)
- Tech Trends 2017 «The kinetic enterprise» Deloitte
- WaveStone « Les nouvelles méthodes de lutte contre la fraude bancaire en ligne »
- SOA, microservices et API management - 4e éd, Le guide de l'architecte des SI agiles – Xavier Fournier-Morel, Pascal Grojean, Guillaume Plouin, Cyril Rognon -
- What are Microservices? Martin Fowler, James Lewis:
<https://martinfowler.com/microservices>
- Guide de référence Spring Cloud Stream: <http://docs.spring.io>
- Site officiel Spring Cloud DataFlow: <http://cloud.spring.io/spring-cloud-dataflow/>
- <https://blog.ippon.fr/2016/09/26/formats-et-methodes-de-serialisation-rest/>
- Site officiel de Jupyter Notebook: <http://jupyter.org/>
- Documentation de JupyterHub: <https://jupyterhub.readthedocs.io/>
- Deploying JupyterHub for Education (Berkeley University):
<https://developer.rackspace.com/blog/deploying-jupyterhub-for-education/>
- Documentation officielle de Docker: <https://docs.docker.com/>

Liste des annexes

<u>Annexe 1. Vue globale de l'architecture SmartStream</u>	page 69
<u>Annexe 2. Benchmarking des solutions de sérialisation</u>	page 70
<u>Annexe 3. L'implémentation de Spring Cloud Deployer - Docker Swarm</u>	page 73
<u>Annexe 4. Interaction de Spring Cloud Deployer avec Docker Swarm</u>	page 73
<u>Annexe 5. Fonctionnement de GitLab CI</u>	page 74
<u>Annexe 6. Fonctionnement de JupyterHub</u>	page 74
<u>Annexe 7. Interface DataLab</u>	page 77

Annexe 1 : Vue globale de l'architecture SmartStream



Annexe 2 : Benchmarking des solutions de sérialisation

Ces tests ont été réalisés sur un ensemble de 500 000 objets, remplis au hasard avec une structure aléatoire. Chaque solution a été testée sur les mêmes objets.

Tableau 1 : Durée nécessaire à la sérialisation d'un objet en nanoseconde

	Médiane	Moyenne
Protobuf	1520	5380,193882
JSON	3421	6522,744906
XML	7602	12227,012
Native	8743	21501,71154

Durée nécessaire à la sérialisation

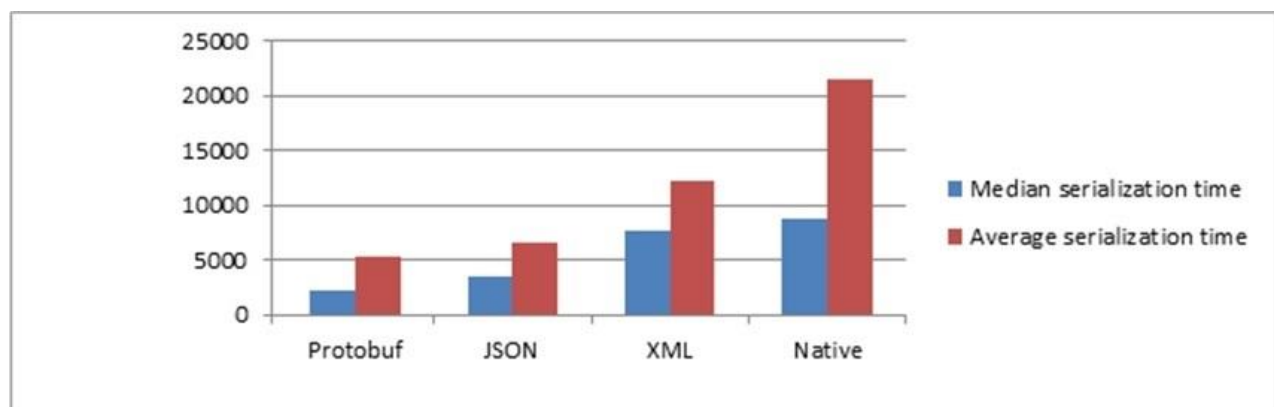


Tableau 2 : Durée nécessaire à la désérialisation d'un objet en nanoseconde

	Médiane	Moyenne
Protobuf	1521	1826,640792
JSON	3041	3709,577588
XML	19767	24949,39496
Native	35732	43490,44696

Durée nécessaire à la désérialisation

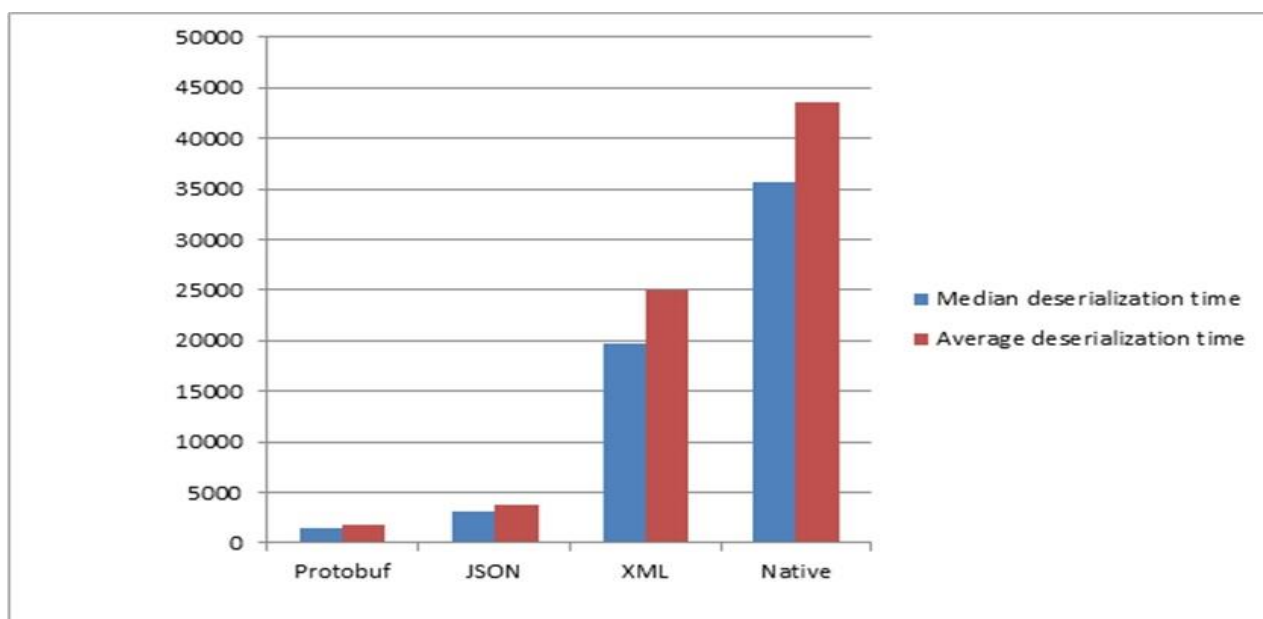
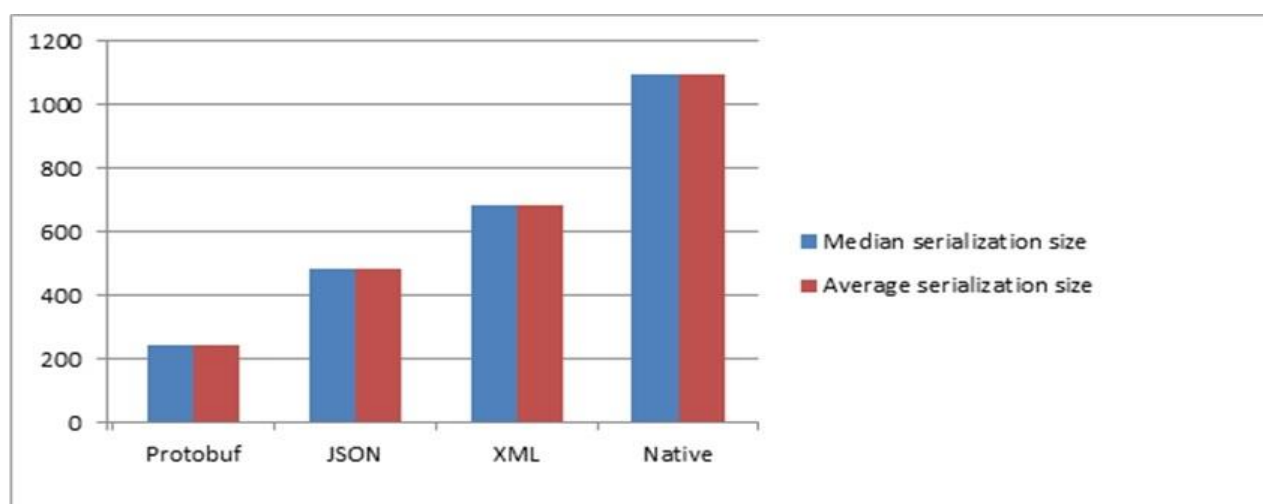


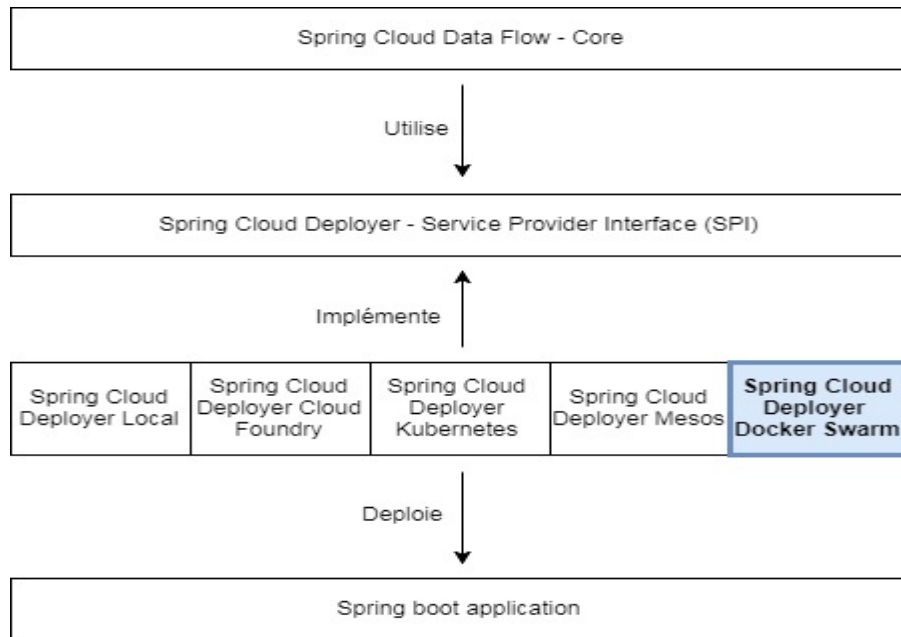
Tableau 3 : Taille de l'objet après sérialisation en bytes

	Médiane	Moyenne
Protobuf	247	246,8
JSON	484	484,33
XML	686	686,33
Native	1096	1095,72

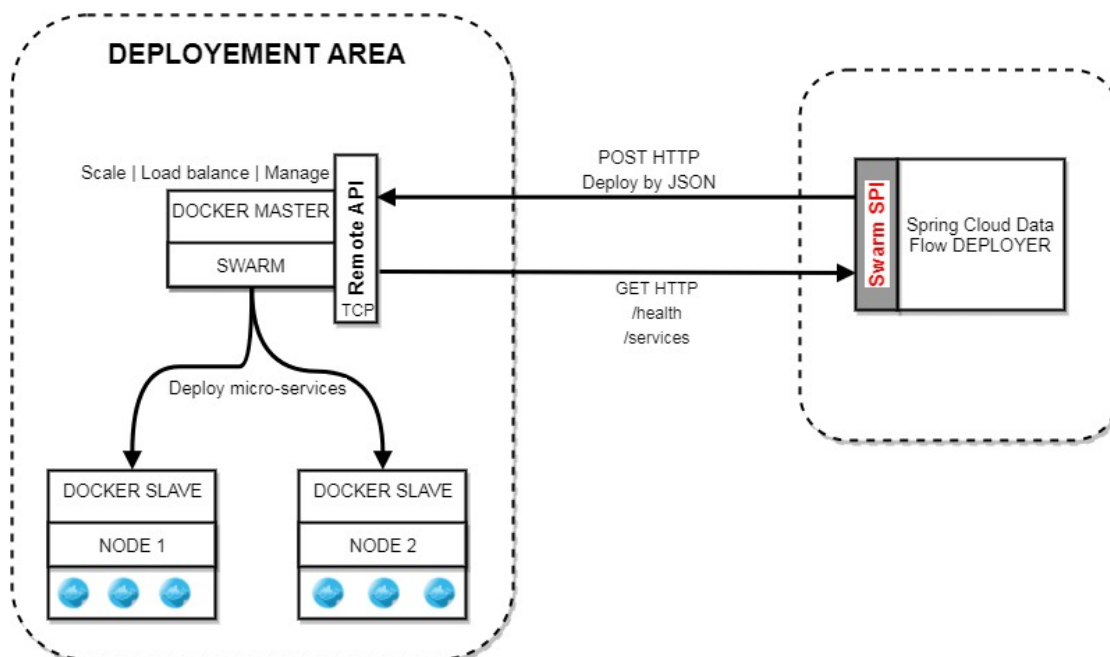
Taille de l'objet après sérialisation



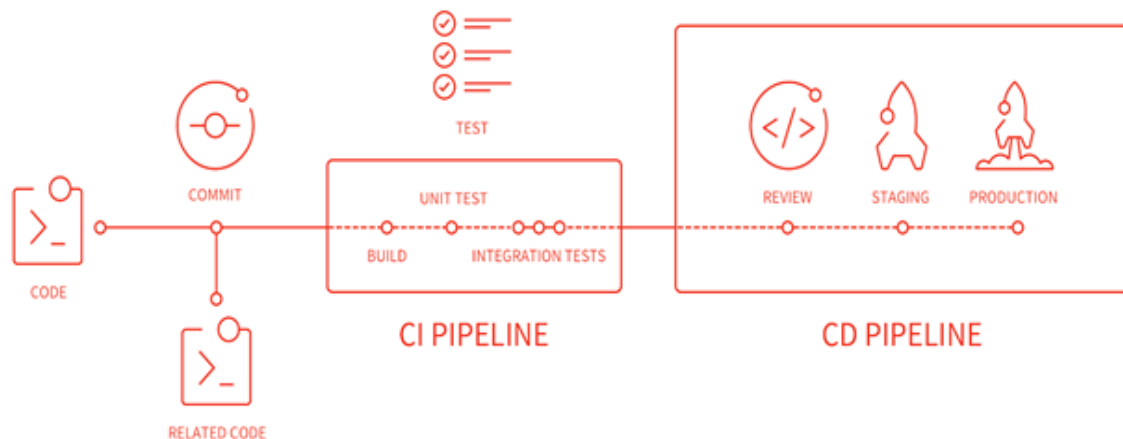
Annexe 3 : L'implémentation de Spring Cloud Deployer Docker Swarm



Annexe 4 : Interaction de notre Spring Cloud Deployer avec Docker Swarm



Annexe 5 : Fonctionnement de GitLab CI



Annexe 6 : Fonctionnement de JupyterHub

Il y a trois catégories de processus démarrés par JupyterHub :

- Plusieurs « Single-User Notebook Server » (Python/IPython/tornado) : Pour chaque utilisateur identifié, un Notebook Jupyter est démarré. Le module qui démarre ces processus est un appelé un Spawner.
- Un proxy HTTP configurable (node-http-proxy) : l'utilisateur accède à une partie du serveur qui utilise un proxy dynamique pour acheminer les requêtes HTTP vers le Hub et les Single User Servers
- Multi-user Hub (tornado process) : gère les comptes utilisateurs, authentifie et coordonne les Single-Users Servers utilisant un Spawner.

Les principes de fonctionnement de JupyterHub sont :

- Hub instancie un proxy
- Par défaut le proxy transfère toutes les requêtes vers le Hub
- Le Hub traite l'authentification (login) et instancie des single-user servers sur demande
- Le Hub configure le proxy pour transférer le préfixe de l'url aux single-user servers
- JupyterHub fournit aussi une API¹⁵ REST

¹⁵ <http://jupyterhub.readthedocs.io/en/latest/api/index.html>

Le proxy est le seul processus qui écoute sur une interface publique. Le Hub est accessible derrière le proxy via `/hub`. Le Single-users servers est accessible derrière le proxy via `/user/[username]`.

Accès à un Notebook Jupyter (via l'authentification par JupyterHub)

Pour un nouvel utilisateur :

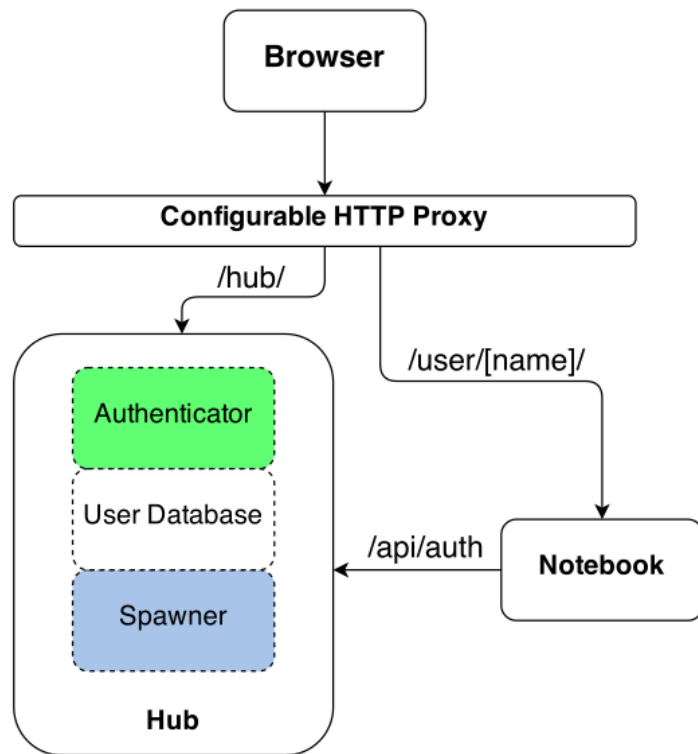
1. Les informations d'identifications sont vérifiées par le module « Authentification » afin d'être validées.
2. Le module « Authentification » retourne le nom d'utilisateur si les informations sont valides.
3. Une instance single-user server est instanciée pour l'utilisateur identifié
4. Quand le serveur est démarré, le proxy est notifié de transmettre le préfixe `user/[username]/*` au single-user server
5. Deux cookies ¹⁶sont créés, un pour `/hub` et un autre pour `/user/[username]`, contenant un token ¹⁷encrypté.
6. Le navigateur est redirigé vers `/user/[username]`, qui correspond au bon single-user server.

Pour un utilisateur déjà identifié sur JupyterHub :

1. Sur demande, le single-user server transmet le cookie chiffré au Hub pour vérification.
2. Hub répond avec le nom de l'utilisateur s'il valide le cookie.
3. Si l'utilisateur est le bon alors il a accès au Notebook Jupyter.
4. Si l'utilisateur n'est pas le bon ou que le cookie n'est pas valide, le navigateur est redirigé vers `/hub/login`.

¹⁶ Un cookie (ou témoin de connexion) est défini par le protocole de communication HTTP comme étant une suite d'informations envoyée par un serveur HTTP à un client HTTP, que ce dernier retourne lors de chaque interrogation du même serveur HTTP sous certaines conditions.

¹⁷ Token est un mot anglais signifiant jeton. En informatique, cet anglicisme est utilisé pour désigner un identificateur.



Processus de Connexion

Annexe 7 : Interface DataLab

Figure 1 : Module d'authentification via LDAP

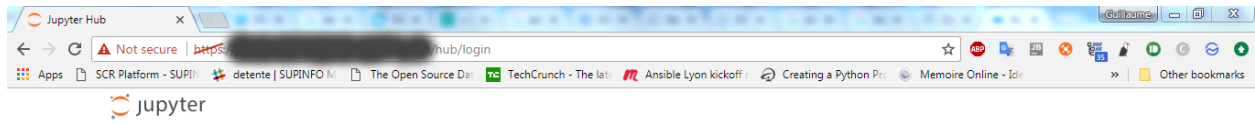


Figure 2 : Module de sélection de l'environnement utilisateur

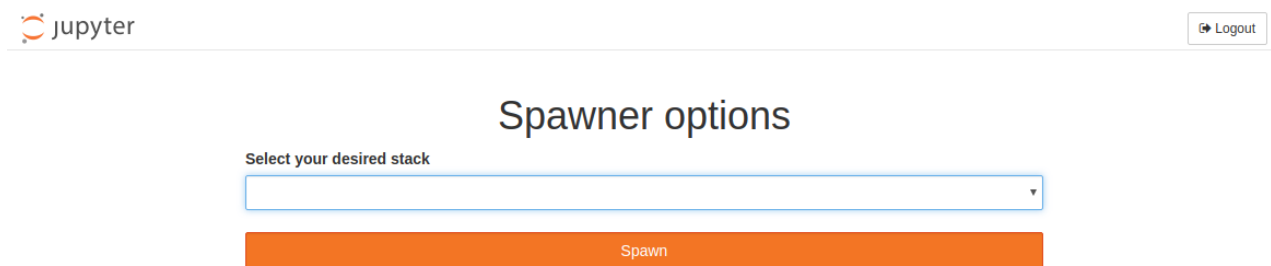


Figure 3 : Écran d'accueil d'un utilisateur



Figure 4 : Exemple d'un notebook python

The screenshot shows a Jupyter Notebook titled "Untitled" with the last checkpoint at 07/26/2017. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Help) and a toolbar with icons for file operations, code execution, and output viewing. The notebook content is as follows:

Linear regression with one variable

In the first part of the exercise, we're tasked with implementing linear regression with one variable to predict profits for a food truck. Suppose you are the CEO of a restaurant franchise and are considering different cities for opening a new outlet. The chain already has trucks in various cities and you have data for profits and populations from the cities.

Let's start by importing some libraries and examining the data.

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:

```
import os
path = os.getcwd() + '\data\ex1data1.txt'
data = pd.read_csv(path, header=None, names=['Population', 'Profit'])
data.head()
```

Out[2]:

	Population	Profit
0	6.1101	17.5920
1	5.5277	9.1302
2	8.5186	13.6620

Figure 5 : Exemple d'un terminal depuis l'interface Web de Jupyter

The screenshot shows a Jupyter Notebook terminal window with the following output:

```
jovyan@d416fb670ce8:~$ ls -al
total 12
drwxr-xr-x. 6 jovyan users 113 Aug 31 14:46 .
drwxr-xr-x. 3 root root 20 Aug 26 22:52 ..
-rw-rw-r--. 1 jovyan users 220 Aug 31 2015 .bash_logout
-rw-rw-r--. 1 jovyan users 3771 Aug 31 2015 .bashrc
drwxr-xr-x. 3 jovyan users 18 Aug 26 22:53 .conda
drwxr-xr-x. 5 jovyan users 67 Aug 31 14:46 .ipython
drwx--S---. 3 jovyan users 19 Aug 31 13:47 .local
-rw-rw-r--. 1 jovyan users 655 May 16 12:49 .profile
drwxrwxr-x. 3 root root 138 Aug 31 14:48 .ssh
jovyan@d416fb670ce8:~$
```