

**JUILLET
2020**

-

**SEPTEMBRE
2020**

RAPPORT DE STAGE

BNP PARIBAS

La banque d'un monde qui change

NOBIAL PAULINE E3



SOMMAIRE



1

REMERCIEMENTS

2-4

I- PRESENTATION DE L'ENTREPRISE

5-11

II- L'ORGANISATION DE LA SOCIÉTÉ

- II-1. Les Fonctions de Contrôle.....p6*
- II-2. Les Pôles opérationnels et les Métiers.....p6-7*
- II-3. Les Territoires.....p8*
- II-4. Les Fonctions transversales.....p8-11*

12-19

III- MA MISSION

- III-1. Le contexte.....p13-14*
- III-2. Le Data Mining.....p15-17*
- III-3. Le Travail réalisé.....p18*
- III-4. Présentation des outils utilisés.....p19*

20-33

IV- LES ETAPES DU PROJET

- IV-1. Les règles d'associations.....p22-25*
- IV-2. Le code.....p26-30*
- IV-3. Les fonctions.....p31*
- IV-4. JSON en quelques mots.....p32*
- IV-5. Exemple de mise en application.....p33*

34

BILAN DE COMPÉTENCES ET D'EXPÉRIENCE

35

CONCLUSION

REMERCIEMENTS

J'aimerais tout d'abord remercier mon maître de stage Nicolas Raffalli ainsi que le manager de l'équipe pour leur accueil, et le partage de leurs expériences tout au long de mon stage.

J'aimerais également remercier les alternants avec qui j'ai pu collaborer durant mes trois mois de stage.

Enfin, je voudrais également remercier les collaborateurs de l'équipe de productions réseaux et sécurité (équipe WIN) qui n'ont pas hésité m'accorder de leur temps lorsque j'avais besoin d'aide sur certains points notamment techniques.



II

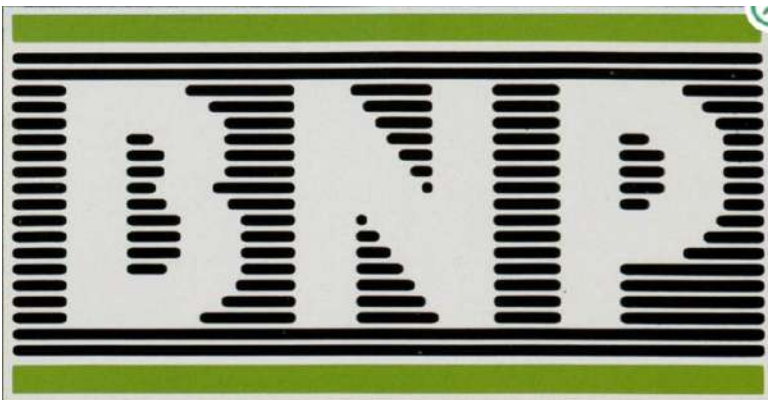
PRÉSENTATION DE L'ENTREPRISE



L'HISTOIRE

1872

En 1872, la Banque de crédit et de dépôt des Pays-Bas, ainsi que la Banque de Paris fusionnent et deviennent « la Banque de Paris et des Pays-Bas ». Cette fusion créée par des banquiers européens, a pour objectif de prendre des participations dans les entreprises ainsi que de s'insérer dans les marchés de capitaux.



Logo de la Banque nationale de Paris (BNP).



Affiche publicitaire suite à l'annonce de la fusion du CNEP et de la BNCP, 1966



L'HISTOIRE

1966 – 1999

En 1966, la Banque nationale de Paris (BNP) née de la fusion de la Banque nationale pour le commerce et l'industrie (BNCI) et du Comptoir national d'escompte de Paris (CNEP). Cette fusion, orchestrée par le gouvernement français, a pour but de créer une banque moderne au service de l'économie nationale, et d'élever Paris dans le monde financier international. Petit à petit, on assiste une bancarisation de la société. La BNP fait partie des six banques à lancer la carte bleue en 1967. Après avoir passé le cap de la crise du début des années 1990 avec succès, René Thomas, l'actuel président de la BNP, quitte sa place et la cède à Michel Pébereau en 1993. Ce dernier va privatiser la BNP, puis sa fusion avec la banque d'affaires Paribas en 2000. C'est ainsi que se forme la BNP Paribas.

AUJOURD'HUI

Aujourd'hui, la BNP Paribas est présente dans 71 pays avec 198 816 collaborateurs (chiffres de Mars 2020, source : bnp paribas). Elle fait partie des pionniers des services bancaires et financiers en Europe.

The background image is a photograph of a modern office building's interior. The upper portion shows a multi-story structure with glass walls and curved balconies, reflecting the interior lights. The lower portion shows a large, open-plan office space with a polished floor, a curved wooden reception area, and several people walking through the space, creating a sense of movement. The overall atmosphere is professional and contemporary.

II

L'ORGANISATION DE LA SOCIÉTÉ



Le grand Groupe européen de services financier s'articule autour de différentes fonctions. On retrouve les Fonctions de Contrôle, les Pôles opérationnels et leurs Métiers, ainsi que les Fonctions transversales. Pour assurer leur bon fonctionnement, il garantit une coordination géographique sur chaque Territoire.

II-1. LES FONCTIONS DE CONTRÔLE :

Pour pouvoir gérer et anticiper les différents risques auxquels est confronté la BNP Paribas, quatre Fonctions de Contrôle indépendantes et intégrées ont été créés :

- **Conformité**
- **Légal**
- **RISK**
- **Inspection Générale**

Ces quatre fonctions sont indépendantes des unes des autres et dispose par ce fait d'une pleine autorité sur son budget et la gestion de ses ressources humaines.

II-2. LES PÔLES OPÉRATIONNELS ET LES MÉTIERS :

La BNP Paribas son activité autour de deux domaines d'activités : **Retail Banking & Services** (RBS), et **Corporate & Institutional Banking**

RETAIL BANKING & SERVICES (RBS)

Le premier regroupe tous les réseaux des banques de détail ainsi que les services financiers du Groupe en France et à l'International. Il est composé par :

Domestic Markets (DM) qui regroupe les banques de détail comme les agences d'Europe.

68 119

COLLABORATEURS
DANS 20 PAYS

+ 17 millions

CLIENTS PARTICULIERS

+ 850 000

CLIENTS PROFESSIONNELS,
ENTREPRENEURS ET ENTREPRISE

International Financial Services (IFS) qui regroupe notamment les activités du Groupe hors zone euro (Etat-Unis, Turquie, Pologne,...)

35 417

COLLABORATEURS

57

PAYS

13 000

CLIENTS ENTREPRISES ET
INSTITUTIONNELS

CORPORATE & INSTITUTIONAL BANKING (CIB)

La CIB s'occupe d'une part de la clientèle entreprises et d'autre part la clientèle institutionnelle. Elle a pour rôle d'apporter des solutions sur mesures dans les différents domaines des marchés.

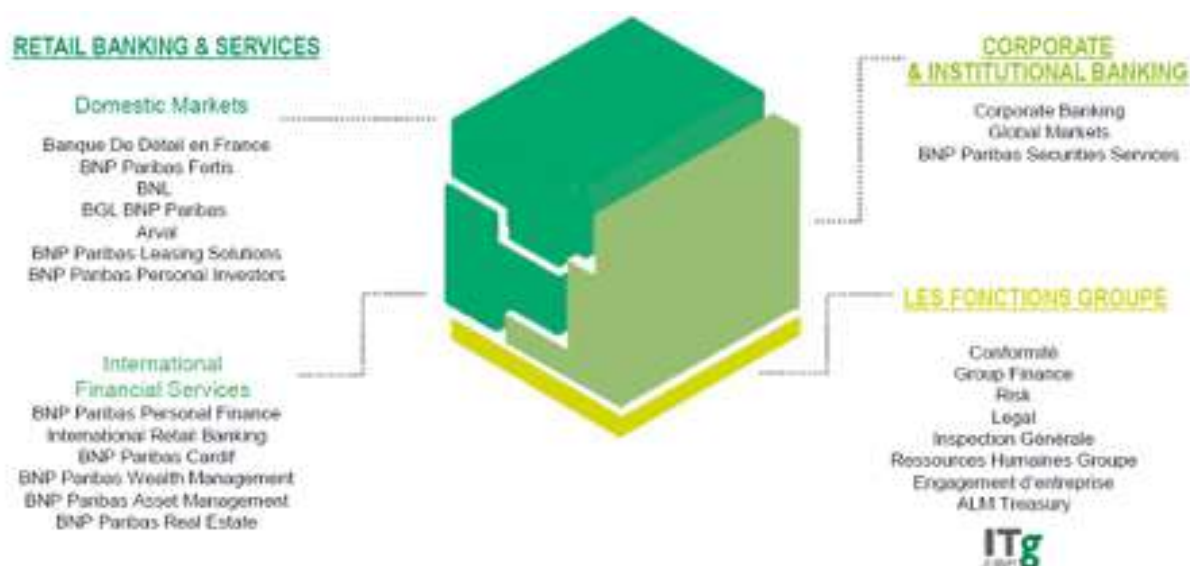


Schéma regroupant les branches Retail Banking & Services ainsi que Corporate & CIB

II-3. LES TERRITOIRES :

Chaque territoire où BNP Paribas exerce une activité, est rattaché à un Pôle. Ces territoires sont par la suite dirigés par un Directeur de Territoire, qui a été préalablement nommé par la Direction Générale ou un délégataire. Ce Directeur de Territoire dépend également hiérarchiquement du Pôle de rattachement.



Carte des différentes branches de BNP Paribas

II-4. LES FONCTIONS TRANSVERSALES :

Il existe au total quatorze fonctions transversales au sein de BNP Paribas. Chaque une de ces fonctions effectuent des missions de support et conseil, ainsi que de contrôle des activités opérationnelles.



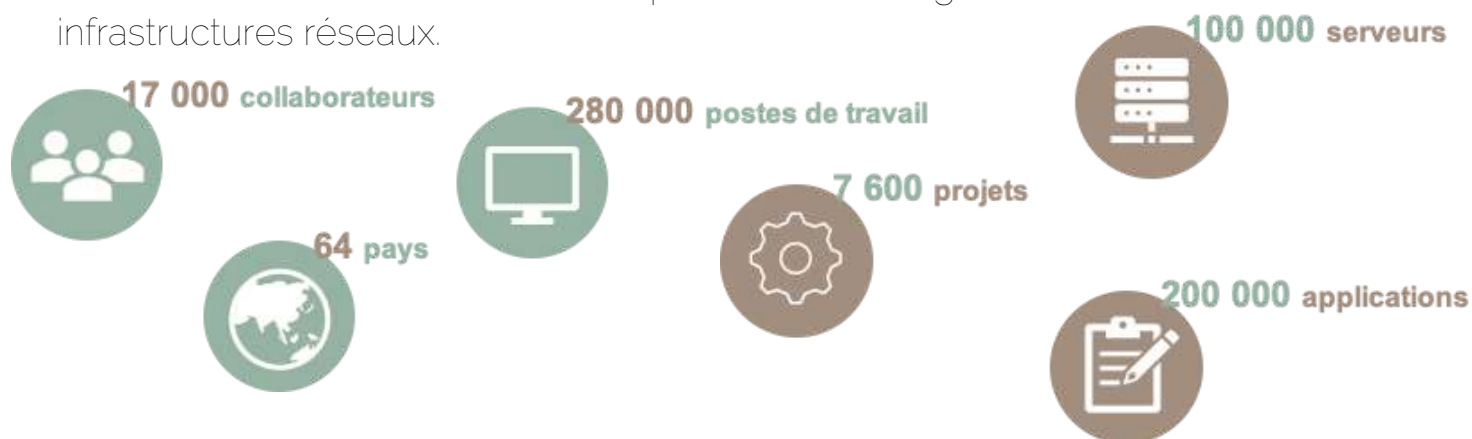
Les fonctions d'Engagement d'Entreprise

Fonction support du Groupe

La fonction à laquelle j'ai été rattaché durant mon stage, est la fonction IT-Groupe. C'est notamment grâce à cette fonction que BNP Paribas a pu se développer au cours du temps tout en restant compétitif. En effet, pour une banque aussi importante que la BNP Paribas, l'informatique est un actif très stratégique, car elle comporte l'usage du matériel informatique, tels que les ordinateurs, qui permettent de stocker, transmettre, manipuler, sécuriser et échanger des données au sein d'un réseau.

A) La fonction IT. GROUPE

La fonction ITG-IT Groupe est créée en mai 2011. Elle permet d'assurer la mise à disposition et le maintien en condition opérationnelle des infrastructures informatiques pour les différents métiers de la banque. Elle est articulée autour de 3 enjeux. Le premier enjeu est de savoir communiquer, répondre, accompagner et mettre en oeuvre les besoins des différents métiers de la banque. Le deuxième est de maintenir la stabilité des systèmes informatiques notamment au sein du data center (centre de données). Et pour finir, il s'agit d'assurer la sécurité des infrastructures réseaux.



Pour répondre aux différents objectifs que s'est fixée la fonction Groupe ITG-IT, plusieurs périmètres d'activités ont été mises en place et réparties à travers différentes sections qui sont : La Gouvernance IT - Les Risques Informatiques - Les Infrastructures Réseaux - Les Services à l'utilisateur et la fonction RH.

GOVERNANCE IT

Assure la cohérence entre les Systèmes d'Information et de Production puis établit des processus informatiques afin de définir la politique des ressources informatiques.

RISQUES INFORMATIQUES

Gère le risque informatique et la conformité. Une politique de sécurité informatique est mise en place afin d'assurer une continuité des services informatiques et télécoms.

INFRASTRUCTURE & RÉSEAUX

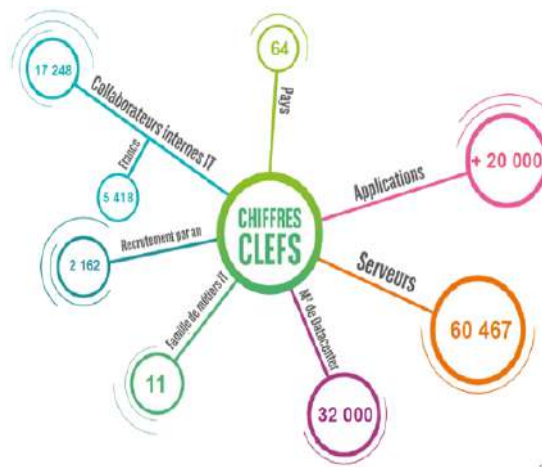
Gère le Cloud Privé, l'hébergement, le stockage, le Big Data, les Télécoms et Flux, la sécurité du réseau BNPP et le Datacenter. Le tout au travers de la production mutualisée et du partenariat avec IBM.

SERVICES UTILISATEURS

Distribue et gère les postes de travail, mène l'assistance utilisateurs, prend en charge le support technique et développe des solutions collaboratives.

CONSTRUCTION ET DÉVELOPPEMENT DU SYSTÈME D'INFORMATION

Etudie, développe et maintient les applications informatiques de la banque. Elle construit l'architecture du Système d'Information et sécurise les applications des clients et des utilisateurs.

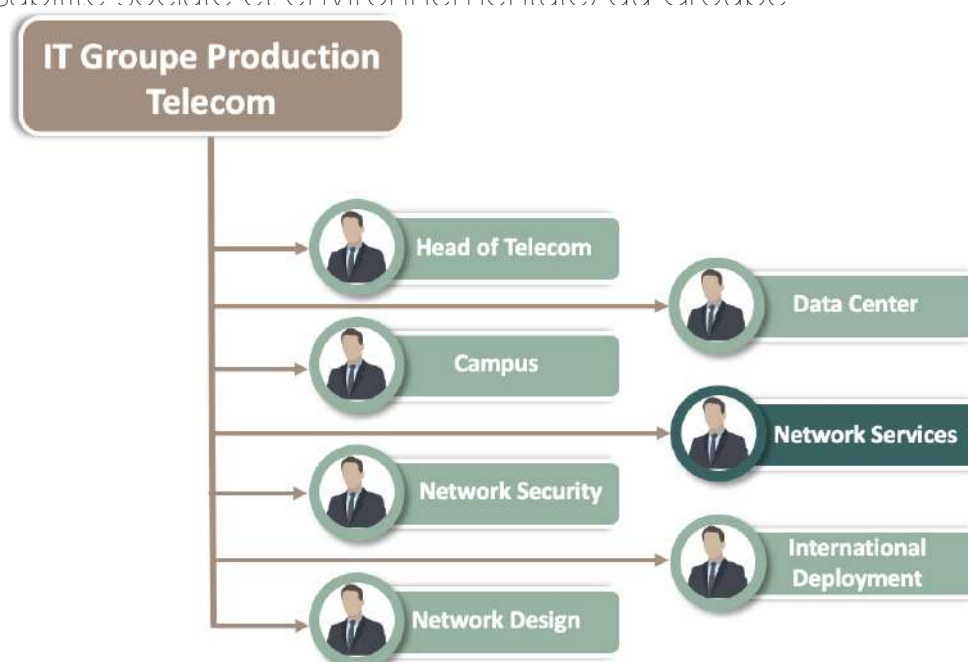


Chiffres clefs du groupe IT de BNP Paribas de 2017

B) La section ITG Infrastructure, Production & Services

La section ITG Infrastructure Production & Service (dite ITG Productions) permet de :

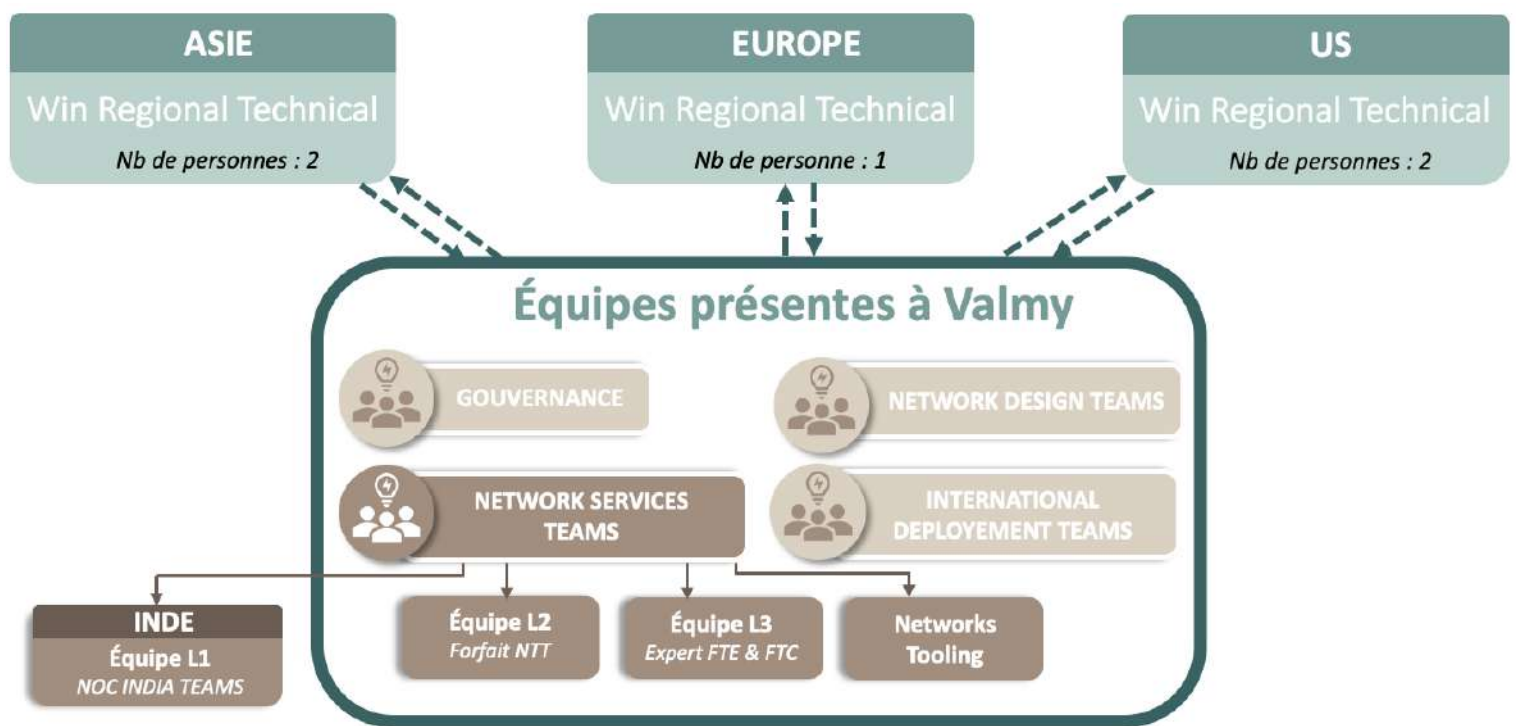
- Assurer un niveau de qualité et de services, en accord avec les besoins des métiers « Retail & Transverse »
- Améliorer l'aisance, l'adaptation et la réactivité de la production
- Optimiser les coûts grâce à un modèle de services adaptés
- Concevoir et mettre en œuvre des solutions innovantes et maîtrisées
- Déployer les actions nécessaires à la maîtrise des risques
- Conduire toutes ces actions dans le respect des engagements RSE (responsabilité sociale et environnementale) du Groupe



Organisation du groupe ITG Productions Télécoms

Durant mon stage, j'ai travaillé au sein de l'entité ITG Production Télécoms, dans le domaine Networks Services. Ce domaine est en charge de la mise en oeuvre des infrastructures réseaux de transport mondiale de BNP Paribas. Il supervise aussi ces infrastructures pour assurer leur bon fonctionnement. Il contribue notamment au développement de l'architecture du Système d'Information et à sa sécurisation et son évolution en fonction des évolutions technologiques. Le domaine Networks Services est divisé en plusieurs équipes pour assurer les fonctions ci dessous:

- Une équipe de niveau 01
- Une équipe de niveau 02
- Une équipe de niveau 03 et d'expertises.



Networks Services une organisation internationale

Le domaine Networks Services est situé principalement sur la ville de Montreuil (lieu de mon stage) mais aussi possède une extension en Europe, Asie, aux Etats-Unis et en Inde.. Ces extensions permettent d'avoir des relais à des endroits stratégiques dans le monde pour offrir un service de maintien en condition opérationnelle des infrastructures 24h/24h et 7j/7j . En Europe il y a qu'un seul relais du fait du créneau horaire similaire à Montreuil, alors qu'en Asie et aux Etats-Unis, il y en a deux. L'Inde assure la fonction de Niveau 01 qui supervise les infrastructures réseaux et sécurité.

Lorsqu'une alerte sur un appareil des infrastructures réseaux d'un pays donné est émise notamment durant la nuit en France, l'équipe en Inde s'appuie sur les relais avant de solliciter les équipes de Montreuil qui effectuent une astreinte durant les heures non ouvrées.



III MA MISSION



III-1 CONTEXTE :

J'ai réalisé un stage de trois mois au sein de l'entité ITG Production et principalement dans le domaine Network Services de la société BNP Paribas. Mon stage s'est déroulé dans les bureaux de Valmy 1 à Montreuil. J'ai eu la chance de collaborer avec l'équipe en charge des infrastructures réseaux et de transports à l'échelle internationale. Mon responsable de stage, Nicolas Raffalli est responsable de l'équipe projet en charge des déploiements des changements nécessaires à la mise en place des services au sein des infrastructures réseaux.

Pour mieux comprendre la mission qui m'a été confiée, il est nécessaire d'expliquer comment travaille cette entité. Comme nous l'avons évoqué plus haut, BNP Paribas dispose d'un réseau international qu'il se doit de gérer à travers différents pôles répartis dans le monde entier, notamment en Amérique du Sud et Nord, en Asie Pacifique, en Europe et aussi en Afrique et au Moyen Orient.



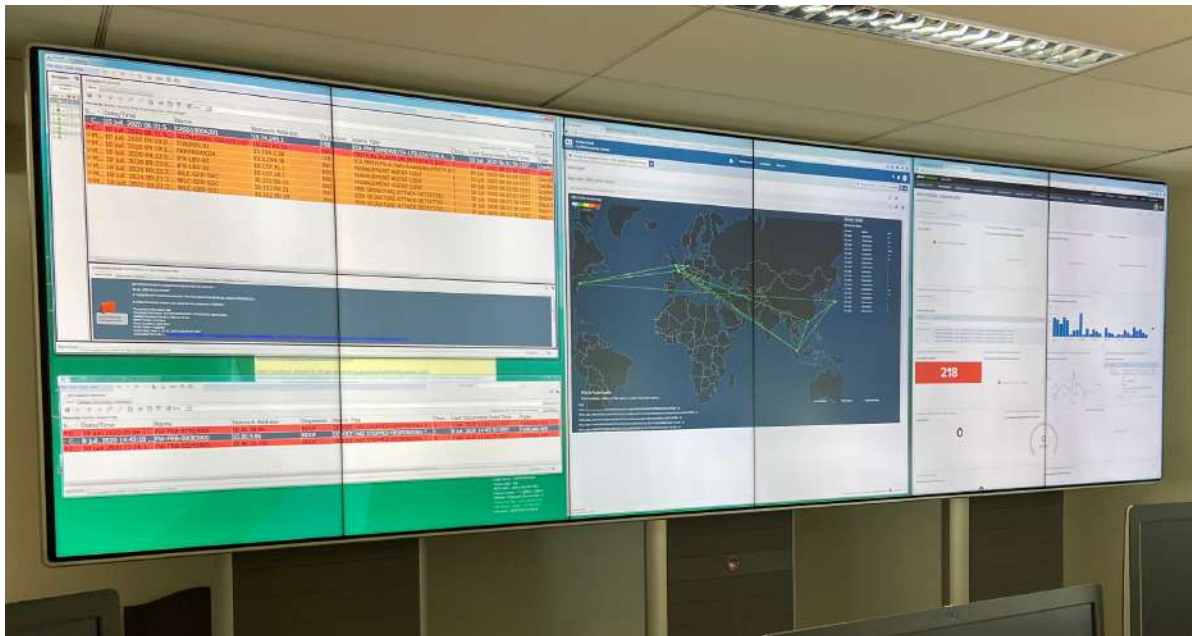
Chiffres Carte Rapport intégré 2018

Pour offrir un service de réseaux et sécurité stable et opérationnel au sein de ces différents territoires, il est nécessaire d'avoir une organisation rigoureuse, afin de maintenir les différentes infrastructures réseaux. C'est pourquoi, un environnement de production est établi pour délivrer du service de réseau et transport sur le périmètre international. Différents domaines d'activités collaborent pour délivrer ce service : Les équipes d'administration des infrastructures, de projets et de relations clientes et d'architectures.



Pour que ces réseaux de transports soient toujours fonctionnels, il est nécessaire de régler et corriger les différents problèmes qui sont détectés en temps réel sur les appareils réseaux. Le domaine Networks Services dans laquelle j'ai travaillé reçois en permanence des alertes à traiter. Ces différentes alertes sont émises par les appareils qui composent le réseau de transport

Pour une meilleure réactivité au niveau de la résolution des alertes, une supervision temps réelle est mise en oeuvre depuis un mur d'image comme ci dessous:



Écran d'alarmes présent dans l'open space

Par exemple sur l'écran de gauche, nous retrouvons des alertes reçues (avec en rouge celles à traiter en priorité). Sur l'écran du milieu, nous avons une visualisation en temps réel de l'utilisation des tuyaux qui relient les différents pays au sein du réseau de transport. L'objectif est de détecter en temps réel des saturations éventuelles (supervision de la bande passante), que nous pouvons assimiler au trafic info sur les routes. Et enfin, sur l'écran le plus à droite, ce sont les différents types d'évènements reçus et leur nombre (exemple 218 en rouge).

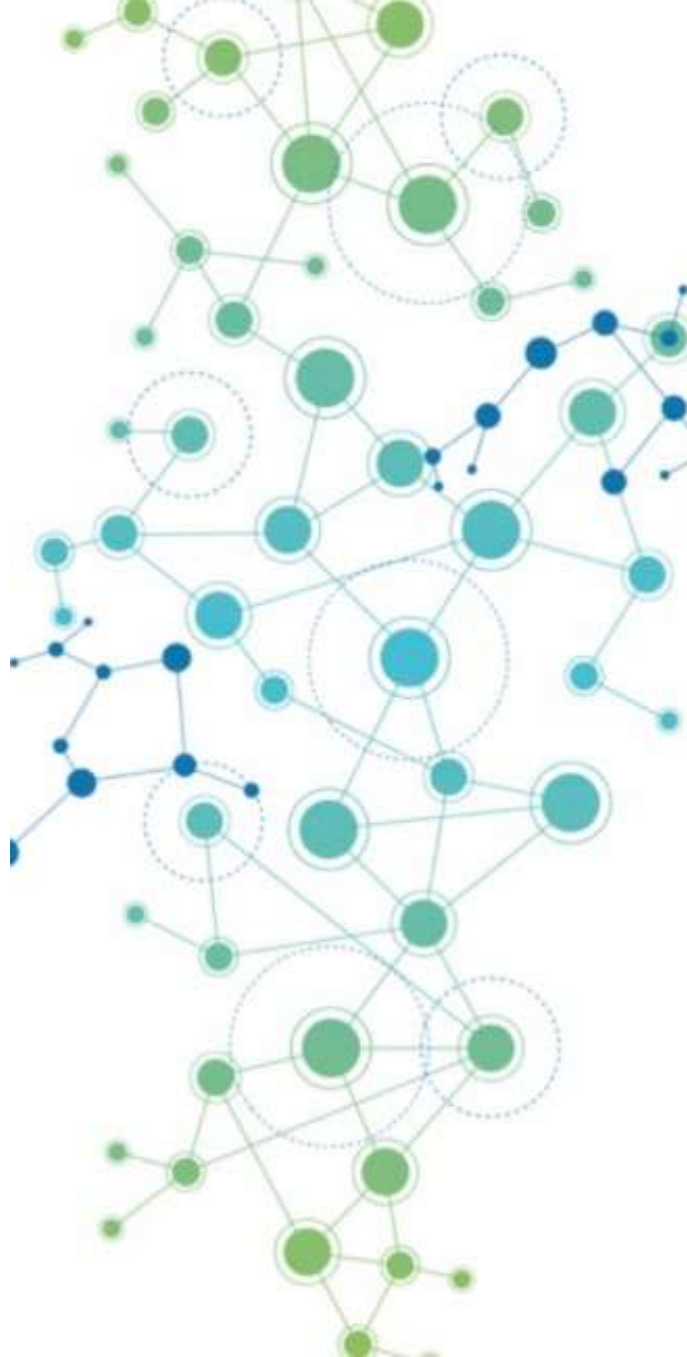
Cependant, toutes ces évènements ne sont pas directement reliés entre eux. Mon stage consistait à étudier un modèle de corrélation permettant de trouver des relations entre eux, pour faciliter le travail d'analyse des ingénieurs.

III-2 LE DATA MINING

Aujourd'hui, on vit dans un monde où la donnée prend de plus en plus de place chaque jour. Elle est tellement omniprésente, qu'on commence à se noyer sous toutes les informations qu'on reçoit au quotidien (actualités, notifications, objets connectés, ect).

Le « Data Mining », qui signifie en français « Exploration de données », regroupe l'ensemble des techniques ou outils utiliser pour analyser cette donnée qui peut être parfois trop volumineuse. C'est grâce au data mining qu'un pré tri est effectué, pour distinguer les informations dites pertinentes du reste. Pour cela, on fait appel à une suite d'algorithmes complexes qui permettent d'exploiter au mieux cette donnée. L'objectif final est de faire apparaitre des corrélations entre différents évènements en apparence distincts, afin d'anticiper des tendances. Le data mining peut être appliqué à une multitude de domaines. Dans mon cas, son application concerne le domaine technique.

Au sein de mon unité, la donnée représente un atout majeur pour les ingénieurs. Ils utilisent notamment pour améliorer le troubleshooting également appelé « dépannage », qui est un processus de recherche et de résolution de problèmes appliqué aux ordinateurs, aux systèmes logiciels, ect.



Plusieurs éléments doivent être pris en compte, dans l'analyse des données pour une meilleure restitution des corrélations. Ils sont décrites ci dessous:

1) La diversité de l'origine des informations

Au sein d'un système d'informations comme à BNP Paribas, l'information au niveau des données à analyser peut provenir des messages syslogs mais aussi des alarmes sur seuils, des indicateurs de performance ou encore de directives externes en cas d'anomalies. C'est pourquoi les formats, les structures et les volumes des données, peuvent varier d'une donnée à une autre.

2) Les différents niveaux de sévérités

Les données transmises par le système d'information se matérialisent par différents niveaux d'impact. L'équipe avec laquelle j'ai travaillé utilise un code couleur mettant en avant le degré d'urgence. Cela permet de prioriser les directives à effectuer.

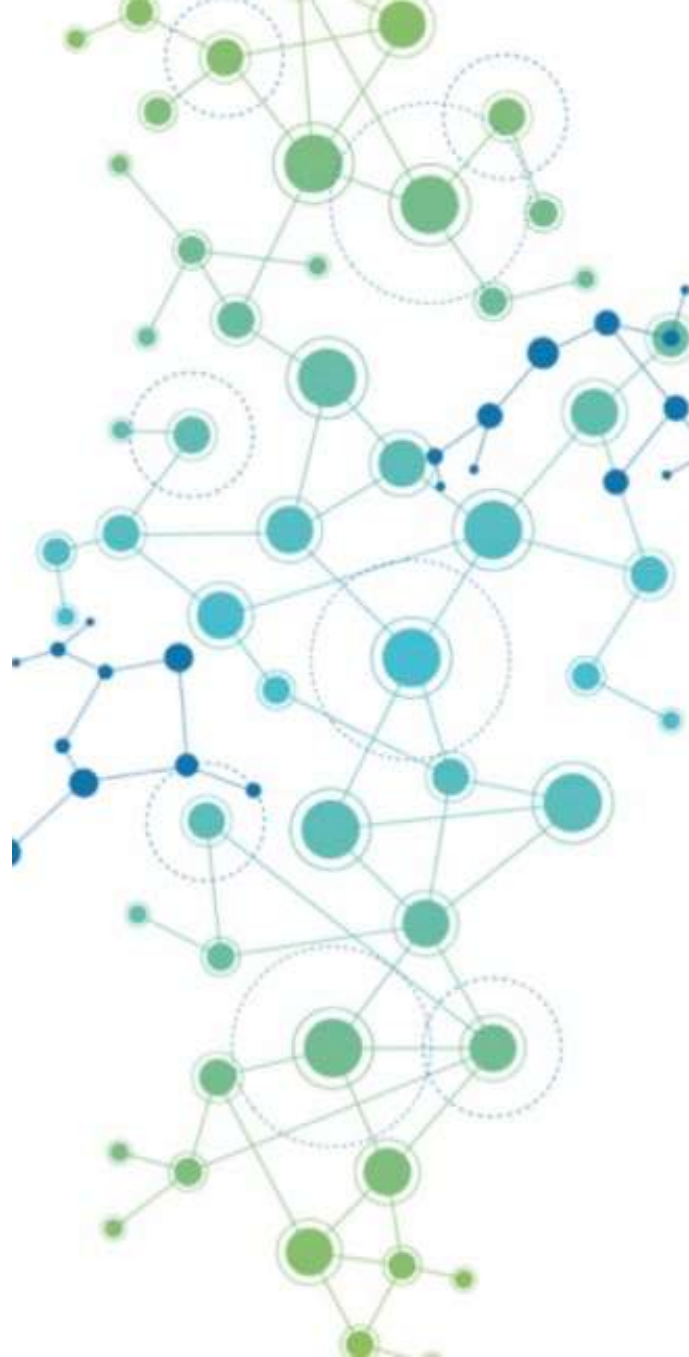


CRITIQUES

MAJEURS

MINEURS

INFORMATIONNEL



3) Les différents périmètres techniques

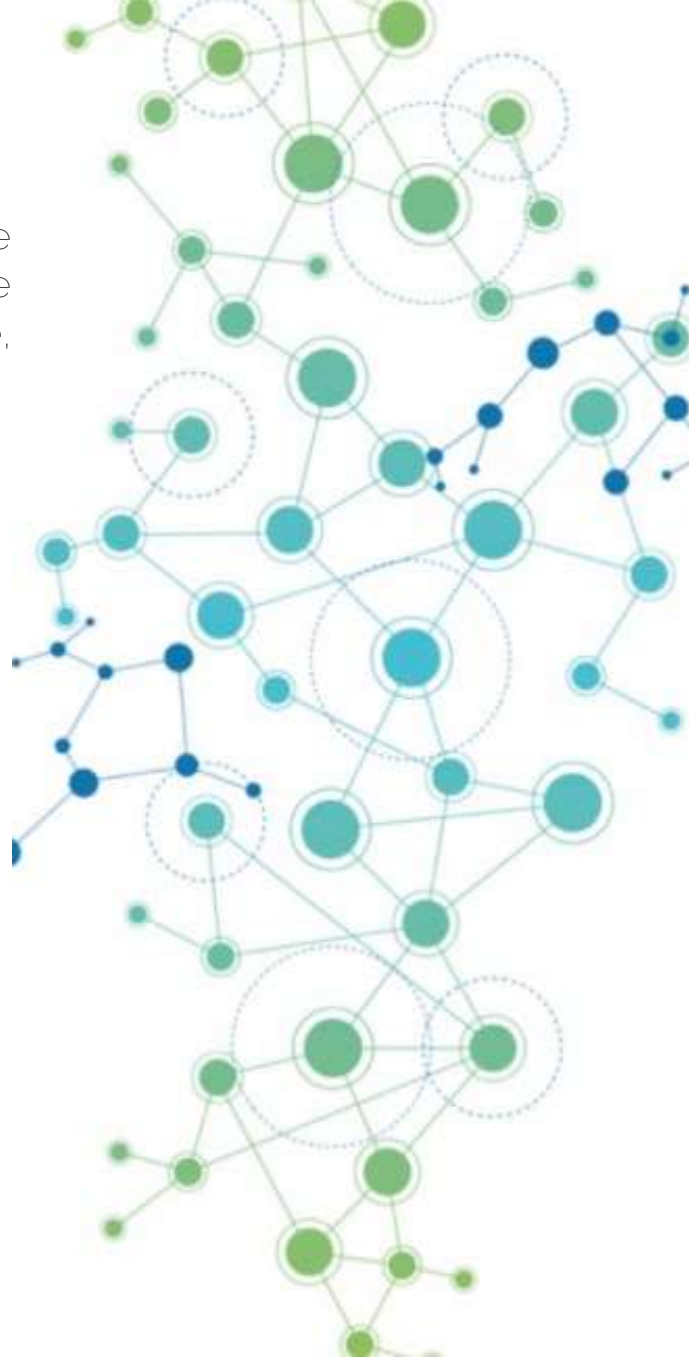
Pour la mise en place d'un système d'informations au sein d'une entreprise comme BNP Paribas avec une portée internationale, plusieurs volets sont à prendre en compte:

- la répartition géographique des infrastructures techniques nécessaires pour délivrer les services de transport.
- Les éléments techniques pour offrir le services réseaux aux utilisateurs finaux comme les connexions WIFI dans les bâtiments ou les connexions réseaux locaux pour connecter les postes de travail.
- Les liaisons de transports de opérateurs télécoms, comme Orange permettant de relier les bâtiments ou régions (New York vers Paris ou Londres vers PARIS)
- Les centres de calculs dits "Datacenter" qui hébergent les serveurs qui hébergent les applications utilisées par les clients de la banque ou les collaborateurs de BNP Paribas.

Tous ces éléments constituent le socle du systèmes d'informations de l'entreprise. Ils sont mis en oeuvres à l'aide d'équipements réseaux comme des routeurs , switchs ou Firewall, ou des liaisons télécoms.

Le Challenge est équipes d'exploitations est de garantir le bon fonctionnement de tous cet ensemble et d'être proactif dans leur analyse en cas d'un élément défaillant .

PROACTIVITE = CORRELATION + PRIORITISATION + ANTICIPATION





III-3 TRAVAIL RÉALISÉ

Durant ce stage, j'ai donc travaillé sur la probabilité et la prédiction d'éléments à venir. L'objectif était de faire corrélérer les données ensemble, afin d'obtenir un résultat. Tout d'abord, je me suis penchée sur un exemple de la vie de tous les jours, exemple les courses au sein d'un supermarché. L'objectif était de me familiariser avec une technique de corrélation des données, en utilisant des termes techniques familiers. Le principe d'analyse de la donnée est identique quelque que soit le domaine technique.

Au sein d'un supermarché, la diversité des données peut être représentée par les différentes courses effectuées par chacun des clients, et la diversité des composants techniques, représentée par les clients du supermarché.

"Il s'agissait donc d'analyser les courses achetées par jour et par client, et de trouver des relations entre elles, afin d'en ressortir une analyse puis une interprétation. Par exemple, si plusieurs clients achètent souvent des articles similaires (exemple des pommes et de l'eau), nous pouvons donc définir une relation et proposer une action comme une promotion sur ces deux produits pour attirer plus de clients.

Au niveau de la sécurité des réseaux, le principe est le même. Dans un premier temps, mon travail a consisté à récupérer les alarmes et événements (syslogs) émis par les différents appareils réseaux. Puis, analyser ces données et les faire corrélérer entre elles, afin d'interpréter au mieux les différents risques émis.

Dans un second temps, l'objectif était de rendre visuel cette analyse de données, afin que ce soit plus rapide et intuitif. En effet, l'équipe dans laquelle j'ai travaillé reçoit régulièrement des alertes avec un volume élevé concernant les appareils réseaux.

III-4 PRÉSENTATION DES OUTILS UTILISÉS



L'équipe Networks Services, en charge de la production sur les infrastructures réseaux de transports de BNP Paribas, utilisent plusieurs outils réseaux pour effectuer cette exploitabilité du réseau. Ces outils génèrent un volume d'évènements (données) dont la corrélation est généralement très difficile à réaliser pour définir des actions techniques correctives. C'est pourquoi, à travers ce stage j'ai travaillé sur le principe du data mining afin de comprendre au mieux leur quotidien.

Mon stage s'est articulé autour de deux volets. En premier lieu, l'objectif était de trouver un moyen pour définir des règles d'associations qui ne sont pas trivial, afin de référencer un processus d'analyse, pour ensuite prédire les situations à venir. Secondement, il a fallu se structurer pour mieux utiliser les messages syslog et définir des directives temporaires de façon proactives.

Le portail WEB sur lequel j'ai travaillé a permis de faciliter la vie quotidienne de l'équipe de production. Il s'articule autour de 2 parties majeurs. Une partie collecte des données via des API, qui permettent de rassembler des données distantes via un lien web (une URL). Et un autre partie interface d'automatisation du programme, pour que la mise en corrélation des données se fasse automatiquement. Pour cela, j'ai développé un programme python.



IV LES ETAPES DU PROJET

L'Objectif en créant des règles d'associations, est de pouvoir prédire les événements à venir. Par exemple, lorsqu'une personne présente de la fièvre et une toux sèche. On va lui demander si elle se sent également fatiguer. Si c'est le cas, on peut supposer qu'elle est atteinte du covid-19 et prendre les mesures nécessaires. Pour arriver à cette conclusion, il a fallu étudier un lot de personnes malades.

On regarde à quelle fréquence les symptômes fièvre et la toux apparaissent dans le lot de malades. Si on remarque que ces deux symptômes, en plus d'une sensation de fatigue, apparaissent de manière fréquentes chez les malades atteints du Covid-19. On peut dans ce cas établir plus rapidement un diagnostic et prendre des mesures si nécessaire.



Mais comment créer ces règles d'associations ?

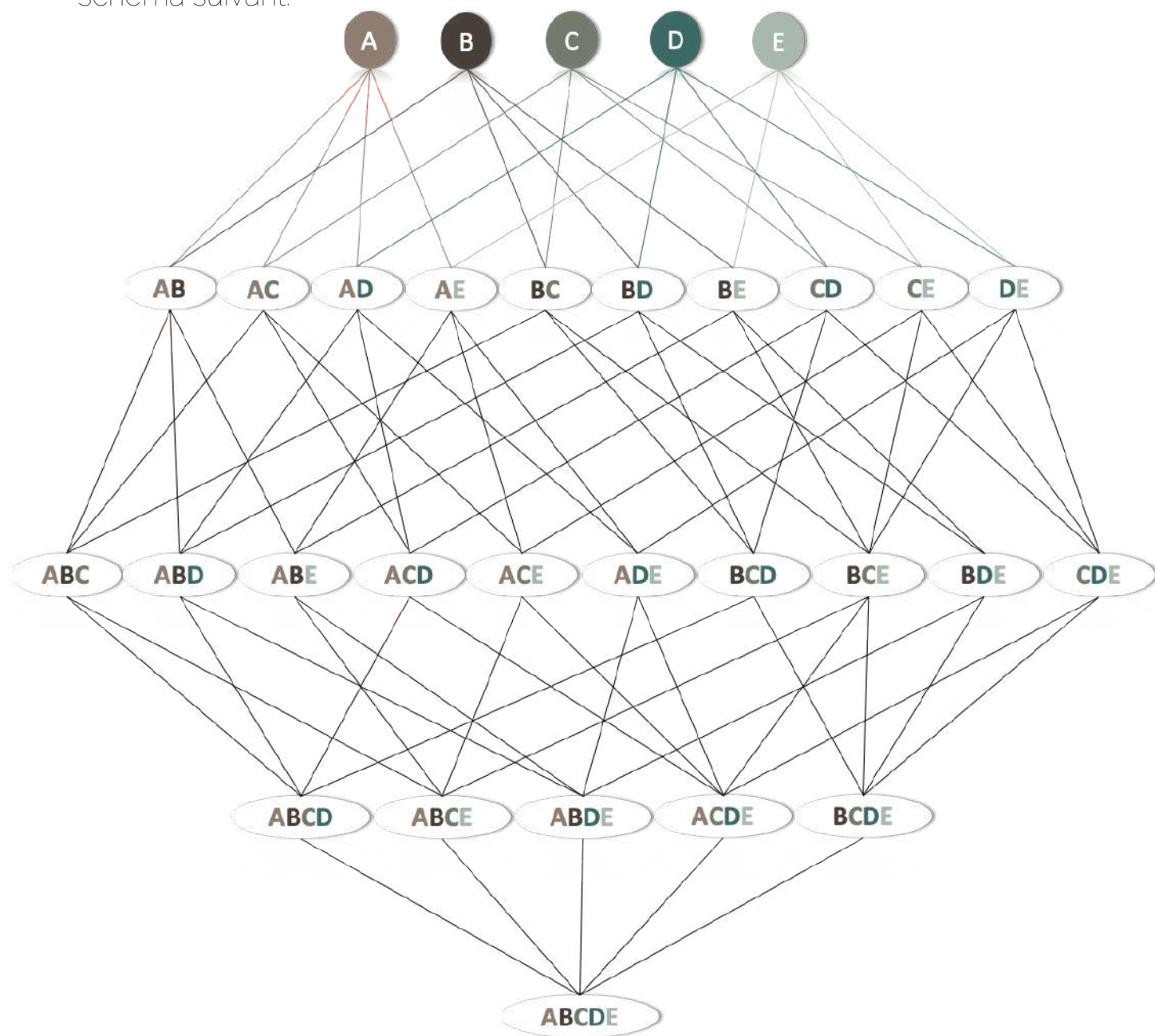
On a vu que l'utilisation de règles d'association permet de rechercher les relations entre les éléments d'un ensemble de données. Pour trouver les règles d'association d'un ensemble, il suffit d'étudier l'occurrence d'un élément en fonction des occurrences des autres éléments présents dans l'ensemble. Pour cela, il est nécessaire d'établir toutes les combinaisons possibles, puis de garder seulement celles qui ressortent le plus souvent.



IV-1 LES RÈGLES D'ASSOCIATIONS :

Exemple :

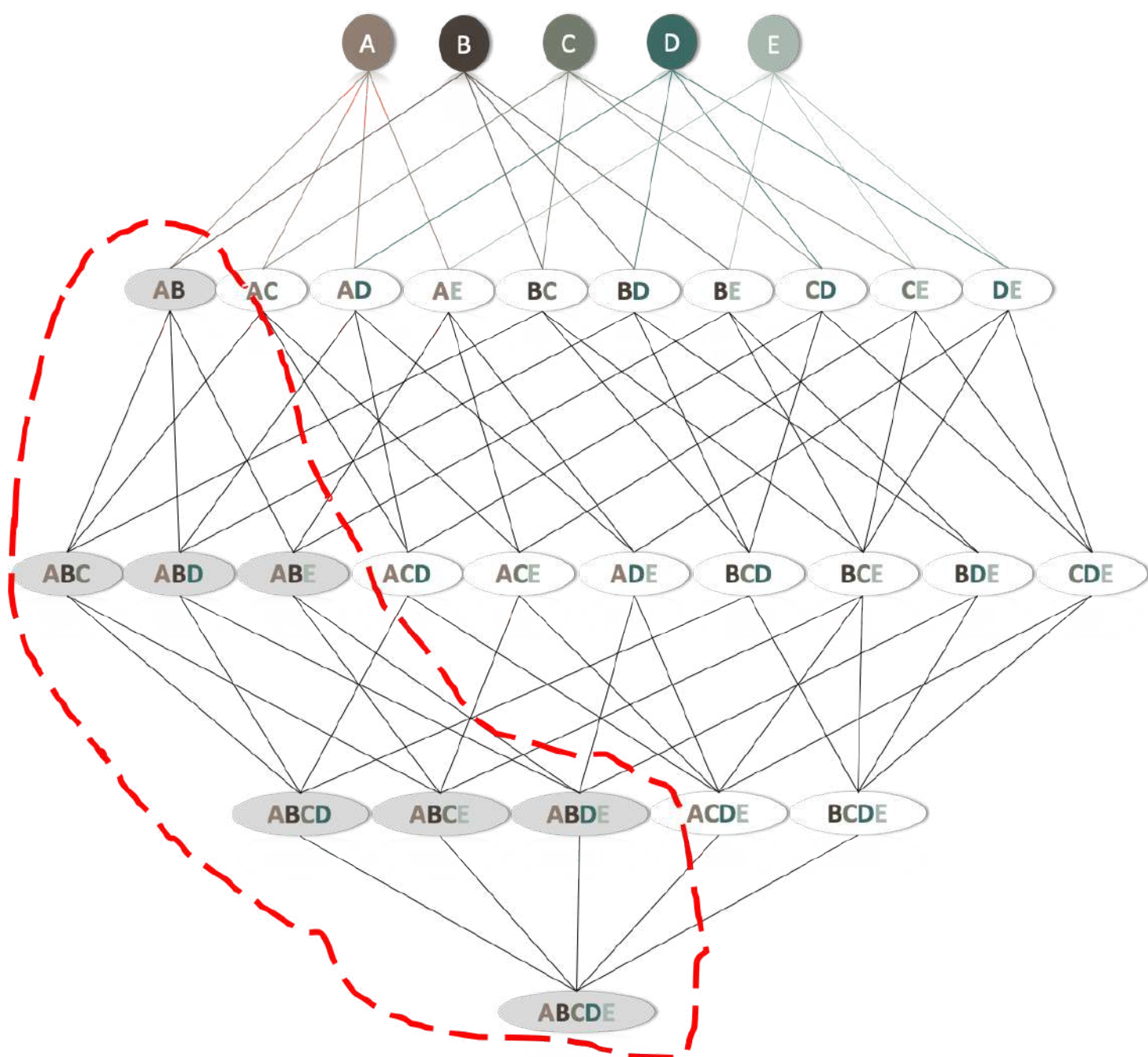
Soit A,B,C,D et E, des éléments qu'on fait correspondre entre eux. Si on cherche à établir tous les ensembles qu'on peut créer, on obtient le schéma suivant.





Exemple :





Maintenant, on se rend compte que la combinaison AB est tout de même peu fréquente. On supprime donc tous les combinaisons comprenant AB.





1) L'exemple d'un supermarché

Prenons un exemple concret pour mieux comprendre. Considérons plusieurs cadis de courses :

	Courses	Articles fréquents	Règles
 n°1	{ <i>Pain, Lait, Muselis,</i> <i>Kinder ...</i> }	→ (Pain, Lait)	→ Pain → Lait
 n°2	{ <i>Pain, Pomme,</i> <i>Jambon, Oeufs, ...</i> }	→ (Pain, Lait)	→ Pain → Lait
 n°3	{ <i>Lait, Banane, Pain,</i> <i>Kinder, ...</i> }	→ (Pain, Lait)	→ Pain → Lait
 n°4	{ <i>Pain, Pomme,</i> <i>Coca, Pain</i> }	→ (Pain, Lait)	→ Pain → Lait

On remarque, que sur ces quatre cadis de courses, les clients ont acheté du pain ont également acheté du lait. On peut donc créer une règle qui associe le pain avec le lait.

La prochaine étape consiste à se demander si les clients ayant acheté ce couple d'aliments en commun, ont aussi d'autres articles en commun, et ainsi de suite. Bien sûr, cet exemple est trop incomplet pour pouvoir établir une véritable règle d'association entre le pain et le lait.



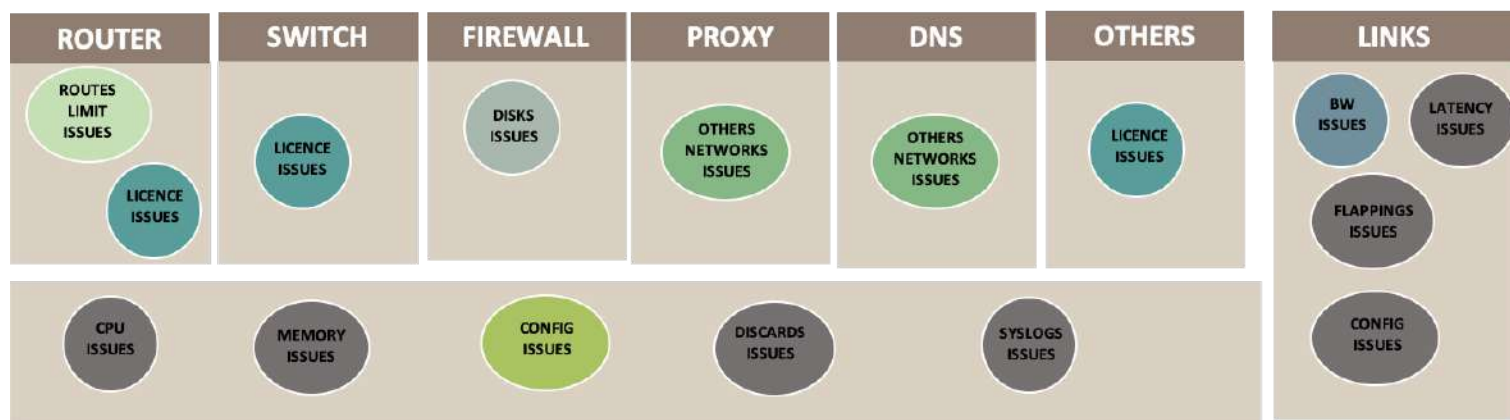
2) Application au domaine technique

Pour déterminer les règles d'associations, il faut d'abord regrouper tous les événements générés au sein des différentes infrastructures. Pour cela, je me suis concentrée sur deux types de typologies d'événements.

- Les alarmes sur seuil générées par les outils de performances
- Les messages syslog

En créant ces règles d'associations, l'objectif est de définir des relations amenant des consignes de productions plus pertinentes.

Voici-ci-dessous un exemple regroupant différents événements auxquelles l'équipe de production peut être confortés.



Dans le monde du système d'informations; Les alarmes n'ont pas la même portée. Il existe des relations plus ou moins intéressantes, qu'on peut par exemple numéroter. Plus le numéro de lot est élevé plus on va s'intéresser à eux. Si on fait le parallèle avec le supermarché, on peut considérer qu'on s'intéresse seulement aux relations entre les articles coûteux. C'est pourquoi l'utilisation d'un paramètre seuil est nécessaire si on veut filtrer d'avantages les relations établies.



IV-2 LE CODE

Comme expliqué plus haut, sur les conseils de mon maitre de stage, je me suis d'abord penchée sur un exemple de la vie de tous les jours pour construire le code. Cela me permettait de mieux appréhender la problématique, car je n'étais pas familier avec les termes techniques au sein des réseaux. La première étape consistait à créer un fichier CSV comprenant une liste de courses. Pour générer une liste de produits de manière aléatoire, j'ai utilisé la fonction ALEA présente sur Excel. Une fois la liste créée, il suffit d'exporter le fichier excel en fichier .CSV.

Jour	Client	Produit
Jour05	Client09	Fromage
Jour02	Client07	Ketchup
Jour05	Client05	Carotte
Jour02	Client03	Eponge
Jour02	Client04	Nutella
Jour05	Client10	Mache
Jour02	Client04	Ketchup
Jour05	Client06	Moutarde
Jour04	Client08	Fromage
Jour02	Client01	Nutella
Jour05	Client07	Nutella
Jour05	Client05	Pates
Jour05	Client03	Moutarde
Jour03	Client01	Nutella
Jour03	Client10	Sardine

L'extension .CSV permet transformer un fichier informatique de type tableur, par un fichier dont les valeurs sont séparées par des virgules.

```
Jour;Client;Produit
Jour05;Client09;Fromage
Jour02;Client07;Ketchup
Jour05;Client05;Carotte
Jour02;Client03;Eponge
Jour02;Client04;Nutella
Jour05;Client10;Mache
Jour02;Client04;Ketchup
Jour05;Client06;Moutarde
Jour04;Client08;Fromage
Jour02;Client01;Nutella
Jour05;Client07;Nutella
Jour05;Client05;Pates
Jour05;Client03;Moutarde
Jour03;Client01;Nutella
Jour03;Client10;Sardine
```

Nous obtenons donc un fichier sous cette forme. Une fois le fichier CSV crée, il a fallu passer au code. Pour cela, j'ai structuré le code de la manière suivante :



MA STRATÉGIE:

Etape 1 : Chargement des données : Lecture du fichier CSV

Etape 2 : Préparation des données : Stocker toutes les données du fichier CSV dans un premier dictionnaire Python

Etape 3 : Traitement des données : Ordonner les données par jour et par client

Etape 4 : Extraction des itemsets fréquents : ici ce sera les aliments les plus achetés

Etape 5 : Recherche des sous-ensembles d'itemsets comportant des items particuliers

Etape 6 : Déduction des règles d'association à partir des itemsets fréquents

Dans ce programme, je prend en paramètre trois critères, qui sont :

- Un paramètre temporel :

permet de savoir sur quel échantillon de temps je travaille.

- Un paramètre de seuil :

permet de savoir au bout de combien de répétitions je veux être alerté

- Un paramètre de corrélation :

permet de savoir si je veux étudier un item ou un set-item précis (ensemble items)

Au total, j'ai utilisé trois dictionnaires différents. Mais qu'est-ce qu'un dictionnaire en Python ? C'est une liste qui va utiliser des clés alphanumériques au lieu d'utiliser des index. Pourquoi avoir utiliser des dictionnaires ? Tout simplement car contrairement aux listes, les dictionnaires sont des objets pouvant en contenir d'autres. Néanmoins, au lieu de stocker des données dans un ordre bien précis, les dictionnaires vont venir associer une clé avec une valeur. Par exemple, nous pouvons avoir un dictionnaire qui contient toutes les adresses répertoriées dans une ville, et pour accéder à ces adresses, il suffit d'entrer un nom. La clé sera le nom et la valeur sera son adresse.



Exemple :

Nous voulons créer un dictionnaire qui permet d'accéder à l'adresse de chaque habitant d'une ville. Prenons comme exemple l'habitant "Patric Dubois", et supposons qu'il habite au "1 rue de Paris 93100 Montreuil". On aura un dictionnaire de la forme :

`dic_des_adresses(clé,valeur)`

avec comme `clé` : Patric Dubois et comme `valeur` = 1 rue de Paris 93100 Montreuil

Voici ci-dessous, les trois dictionnaires utilisés pour ce code.

dic_courses : est un dictionnaire qui nous donne la liste des courses achetées par jour (Key_jour) et pour chaque client (Key_client)

Ce dictionnaire pointe d'abord sur `key_jour` puis sur `key_client` puis sur `listecourses`

Exemple :

dic_courses	→	JOUR1	→	Client1	→	{Fraise,Peché,Haricots,Coco}
dic_courses	→	JOUR1	→	Client2	→	{Casis,Sardine, Fraise}

dic_liste_courses : est un dictionnaire où je mets toutes les courses d'une journée. Ce dictionnaire pointe d'abord sur `key_jour` puis sur `listcourses` (la liste des aliments)

Ce dictionnaire pointe d'abord sur `key_jour` puis sur `listcourses` puis sur la liste article.

Exemple :

dic_liste_courses	→	JOUR1	→	{Fraise,Framboise,Mangue,Casis,Kiwi}
dic_liste_courses	→	JOUR1	→	{Fraise,Framboise,Mangue,Banane}

dic_items : est un dictionnaire qui nous donne la liste des clients par aliments achetés. Ce dictionnaire pointe d'abord sur `key_jour` puis sur `key_aliment` et sur la `liste client`.

Exemple :

dic_items	→	JOUR1	→	Fraise	→	{Client1,Client2,Client3,Client4}
dic_items	→	JOUR1	→	Kiwi	→	{ Client2,Client3,Client5}



J'utilise à chaque fois des dictionnaires car je veux pointer sur une journée donnée (Key-jour) avant d'accéder aux données souhaitées.

Pour mieux comprendre, voici un exemple. Disons que le dictionnaire dict_Items nous renvoie sur ce tableau et que je choisis de garder seulement les cas où il y a deux occurrences.

Kiwi	{Client1,Client2,Client5}
Mangue	{Client2,Client5}
Fraise	{Client5}
Abricot	{Client1,Client3,Client5}
Cassis	{Client3}
Pêche	{Client2,Client4,Client5}

1) Je décide donc de supprimer Fraise et Cassis car ils ont qu'un seul client. Le nouveau tableau sera :

Kiwi	{Client1,Client2,Client5}
Mangue	{Client2,Client5}
Abricot	{Client1,Client3,Client5}
Pêche	{Client2,Client4,Client5}

2) Les articles restant sont **Pêche, Abricot, Kiwi, Mangue** . Je vais pouvoir créer 6 couples :

{Pêche, Abricot}	{Abricot, Mangue}
{Pêche, Kiwi}	{Abricot, Kiwi}
{Pêche, Mangue}	{Mangue, Kiwi}

3) Le nouveau tableau est : (Je regarde quel client a dans son panier chaque couple d'articles)

{Pêche, Abricot}	{Client5}
{Pêche, Kiwi}	{Client2,Client5}
{Pêche, Mangue}	{Client2,Client5}
{Mangue, Kiwi}	{Client2,Client5}
{Abricot, Mangue}	{Client5}
{Abricot, Kiwi}	{Client1, Client5}



4) Je garde seulement les éléments avec 2 clients ou plus car l'occurrence choisit est de 2. Le tableau devient donc :

{Pêche, Kiwi}	{Client2,Client5}
{Pêche, Mangue}	{Client2,Client5}
{Mangue, Kiwi}	{Client2,Client5}
{Abricot, Kiwi}	{Client1, Client5}

5) Je recrée des couples avec 3 articles. Il reste Pêche, Abricot, Kiwi, Mangue

{Pêche, Abricot, Kiwi}
{Pêche, Abricot, Mangue}
{Abricot, Kiwi, Mangue}
{Kiwi, Mangue, Pêche}

{Pêche, Abricot, Kiwi}	{Client5}
{Pêche, Abricot, Mangue}	{Client5}
{Abricot, Kiwi, Mangue}	{Client5}
{Kiwi, Mangue, Pêche}	{Client5}

6) Le nouveau tableau est :

7) Il n'y a pas de trio ayant deux clients minimum, donc j'affiche ce que nous a trouvé à l'étape n-1., soit les résultats de l'étape 4 :

{Pêche, Kiwi}
{Pêche, Mangue}
{Mangue, Kiwi}
{Abricot, Kiwi}

Je peux conclure que quand lorsqu'un client achète une Pêche il y a plus de chances qu'il achète Mangue ou Kiwi.

Une fois que le principe du code été bien assimilé, je suis passée à l'application de ce dernier dans le domaine technique. En effet les « clients » sont devenus les devices (appareils) et les « courses » les alertes.



IV-3 LES FONCTIONS

Pour structurer au mieux le programme, j'ai défini plusieurs fonctions. En effet, cela permet d'appeler de manière récursive la même fonction, mais cela sera surtout plus pratique quand on appliquera le code au domaine informatique (sur les différents devices).

➤ **Fonction 1 :** Lecture du fichier .csv comportant les différentes données.

Jour	→	Time
Client	→	Device
Courses	→	Alertes

```
"_time","Device_Name",typologie,AlarmTitle,"Count of SpectrumLog"
"2020-08-18T00:00:00.000+0200",DOHTTCE04,"#Bandwidth#", "INTERFACE BANDWIDTH UTILIZATION ABOVE 100% for Gi0/2/0",1
"2020-08-18T00:00:00.000+0200",DOHTTCE04,"#Bandwidth#", "INTERFACE BANDWIDTH UTILIZATION ABOVE 100% for Gi0/2/0.731",1
```

Exemple extrait d'un fichier csv contenant les différentes alarmes

➤ **Fonction 2 :** Affichage des données lues.

➤ **Fonction 3 :** Création du dictionnaire avec ajout des données lues.

➤ **Fonction 4 :** Analyse de ces données (si le nombre de client ayant acheté le même article est supérieur à l'occurrence entrée, alors on crée un nouveau dictionnaire).

➤ **Fonction 5 :** Création de combinaisons. Par exemple, si **Alerte_1** **Alerte_3** ont été détectées sur les appareils **Device_5** **Device_7** et **Device_9** alors on crée la combinaison suivante :

(Alerte_1 ; Alerte_3) : [Devices_5 ; Devices_7 ; Devices_9].

➤ **Fonction 6 :** Une fois les associations créées, on les fait corrélérer entre elles.

➤ **Fonction 7 :** Ecrire du résultat obtenu dans un fichier JSON



IV-4 JSON EN QUELQUES MOTS

Le format JSON (JavaScript Objet Notation) est un format léger permettant l'échange de données. Ce format est à l'initialement basé sur du code Javascript, mais peut-être utilisé avec tous les langages modernes. C'est pourquoi je l'ai utilisé en Python. Pourquoi convertir les données en format JSON ? Tout d'abord, on extrait les résultats pour pouvoir les exploiter. On décide de les convertir au format JSON, pour pouvoir par la suite les exporter sur des logiciels de visualisation tel que Splunk ou Grafana.

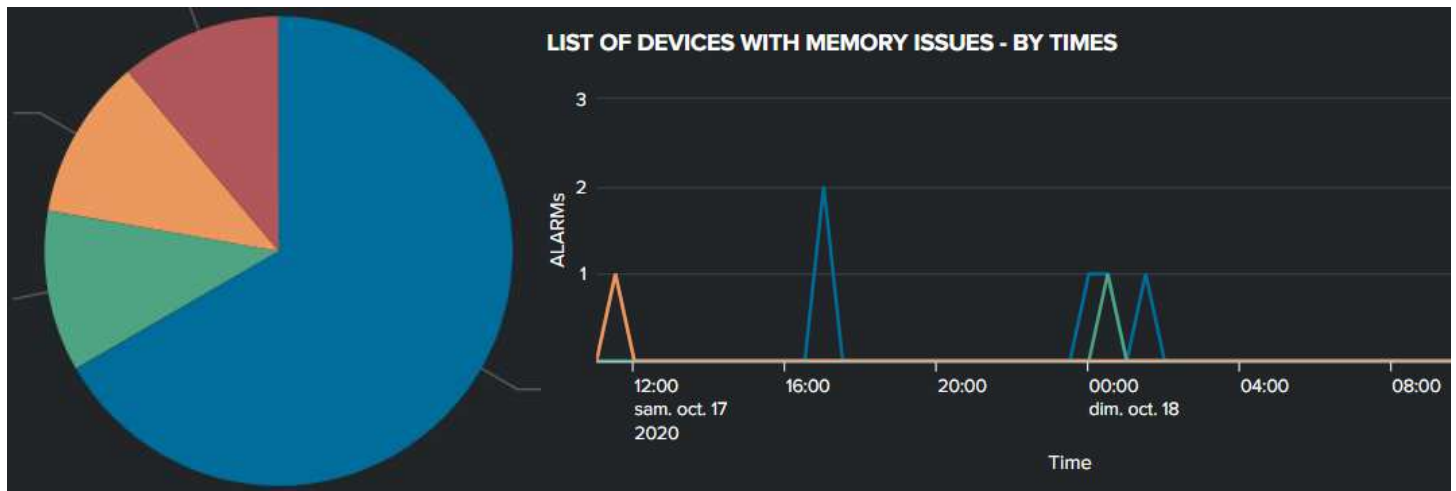
```
"Jour04": {
  "_Nutella_Coca_Ketchup": [
    "Client05",
    "Client03",
    "Client08"
  ]
},
"Jour05": {
  "_Riz_Pates_Salade": [
    "Client04",
    "Client10",
    "Client08"
  ]
},
"Jour02": {
  "_Mais_Pates_Ketchup": [
    "Client10",
    "Client07",
    "Client06"
  ]
},
```

```
"Jour04": {
  "_Nutella_Coca_Ketchup": [
    "Client05",
    "Client03",
    "Client08"
  ]
},
"Jour05": {
  "_Riz_Pates_Salade": [
    "Client04",
    "Client10",
    "Client08"
  ]
},
"Jour02": {
  "_Mais_Pates_Ketchup": [
    "Client10",
    "Client07",
    "Client06"
  ]
},
```

Extrait du fichier JSON

IV-5 EXEMPLE DE MISE EN APPLICATION

Liste des événements reçus sur une fenêtre de 24 heures :



Extrait du logiciel SPLUNK.

- Alarme CPU** : problèmes de performance
- Alarme Latency** : problèmes de latence sur l'appareil
- Alarme Bandwidth** : problèmes liés à la capacité de véhiculer beaucoup d'informations
- Alarme DROP** : problèmes de pertes d'informations



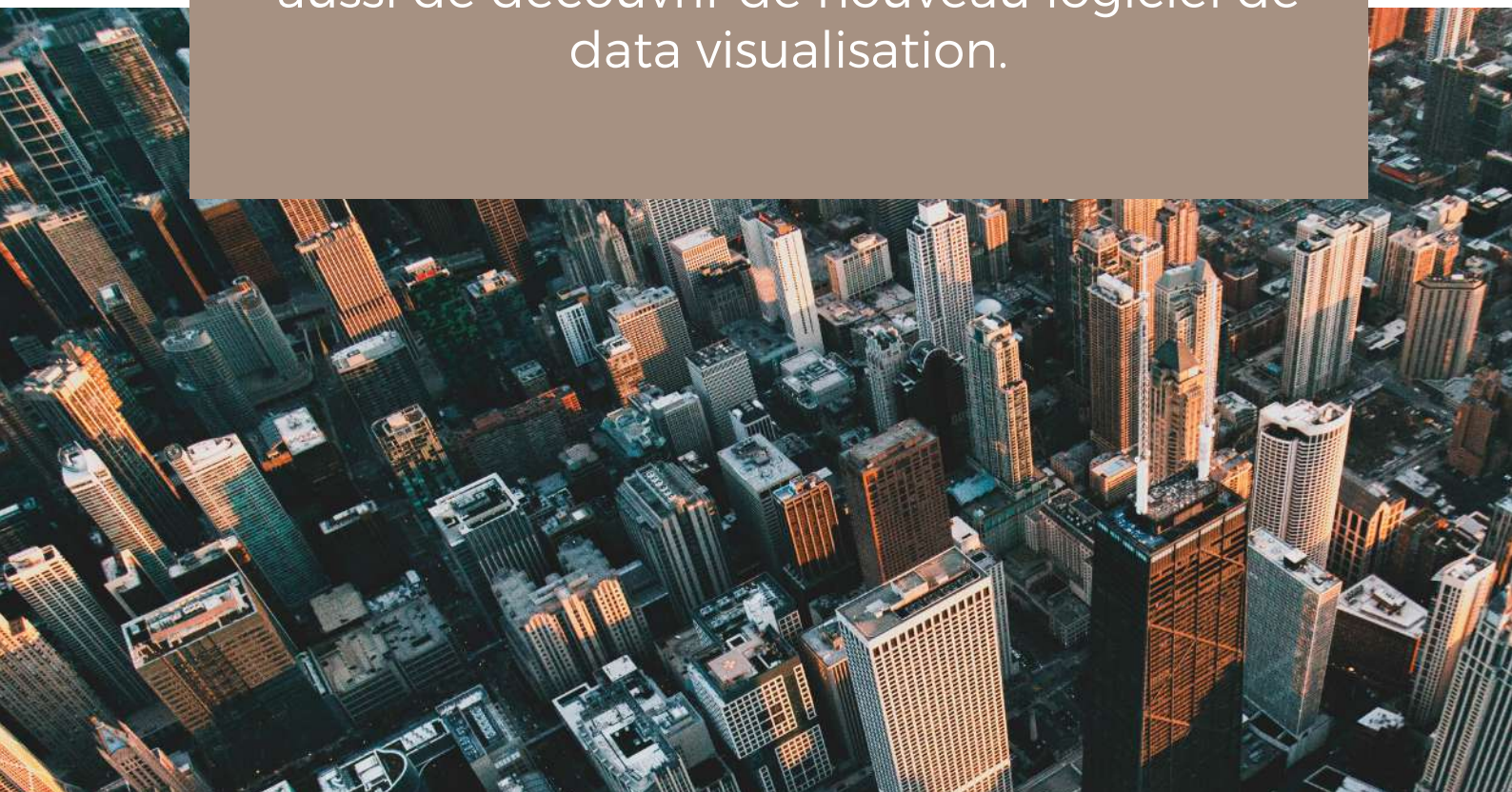
Si on veut faire corréler les événements entre eux, il suffit de regarder l'heure d'arrivée des alarmes. On remarque qu'à 19h, on a reçu une alarme type CPU puis deux alarmes type BANDWIDTH. On peut donc émettre l'hypothèse que ces trois alarmes sont liées entre elles.

Créer des corrélations entre les alarmes, permet d'anticiper les événements à venir, et par conséquent de prendre plus rapidement des mesures pour les régler.

BILAN DE COMPÉTENCES ET D'EXPÉRIENCE

Ce stage m'a permis de découvrir le monde du travail pour la première fois. J'ai pu tester mes compétences en informatique, en passant du cas théorique au cas pratique. De plus, j'ai également gagné en autonomie et en esprit d'équipe.

Quant à la mission sur laquelle j'ai travaillé, elle m'a permis d'améliorer mes compétences en langage python, mais aussi de découvrir de nouveau logiciel de data visualisation.



CONCLUSION

Pour conclure, j'ai pu travailler dans une banque mondialement connue. En effet, travailler dans une banque qui rayonne à l'international à travers ces différents domaines d'activités, m'a permis de découvrir la manière dont elle réfléchit et déploie ces infrastructures réseaux à travers le monde.

J'ai également pu en apprendre plus sur le domaine des réseaux, notamment au niveau de leur configuration mais aussi au niveau de leur gestion.

J'ai pu comprendre le réel besoin de ces automatisations lorsqu'il est question d'importantes opérations sur le réseau ou de configurations longues.