Thibault VIVIER

Université de Lorraine | B.U.T. Science des données – VCod

Analyse et conception d’un outil décisionnel

SAe 501 & 601 – VCOD

# Table des matières

[Table des matières 3](#_Toc190089112)

[1. Contexte 5](#_Toc190089113)

[1.1. Contexte théorique 5](#_Toc190089114)

[1.1.1. Descriptif 5](#_Toc190089115)

[1.1.2. Attendus et objectifs de la SAE. 5](#_Toc190089116)

[1.2. Contexte matériel et environnement 6](#_Toc190089117)

[1.3. Cahier des charges et objectifs 6](#_Toc190089118)

[2. Analyse 7](#_Toc190089119)

[2.1. Veille technique 7](#_Toc190089120)

[2.1.1. Webinaire DataGrandEst 7](#_Toc190089121)

[2.1.2. Conférences Devoxx 11](#_Toc190089122)

[2.1.3. Recherches supplémentaires 13](#_Toc190089123)

[2.1.4. Synthèse 15](#_Toc190089124)

[2.2. Titre 15](#_Toc190089125)

[2.2.1. Définition du projet 15](#_Toc190089126)

[2.2.2. Solutions pratiques envisagées 15](#_Toc190089127)

[2.3. Chaine décisionnelle 16](#_Toc190089128)

[2.3.1. Architecture 16](#_Toc190089129)

[2.3.2. Cas d’utilisations pratiques 18](#_Toc190089130)

[3. Situations d’expérimentation (=partie tech) 20](#_Toc190089131)

[3.1. Gestion du projet 20](#_Toc190089132)

[3.1.1. Méthodologie 20](#_Toc190089133)

[3.1.2. Outils utilisés 20](#_Toc190089134)

[3.2. Stations de collecte et capteurs 20](#_Toc190089135)

[3.2.1. Le Raspberry PI 20](#_Toc190089136)

[3.2.2. Les capteurs 20](#_Toc190089137)

[3.2.3. Architecture logicielle 20](#_Toc190089138)

[3.2.3. Station de collecte : premiers traitements 20](#_Toc190089139)

[3.3. Flux et intégration de données 21](#_Toc190089140)

[3.3.1. Envoi des données en temps réel 21](#_Toc190089141)

[3.3.2. Traitement des images 21](#_Toc190089142)

[3.3.3. Intégration en base de données 21](#_Toc190089143)

[3.4. Tableau de bord 21](#_Toc190089144)

[3.4.1. Outil utilisé : Power BI 21](#_Toc190089145)

[3.4.2. Valorisation des données 21](#_Toc190089146)

[4. Pistes d’amélioration 22](#_Toc190089147)

[4.1. Matérielles 22](#_Toc190089148)

[4.2. Logicielles 22](#_Toc190089149)

[4.3. Organisationnelle 22](#_Toc190089150)

[4.4. Autres ? 22](#_Toc190089151)

[5. Bilan 23](#_Toc190089152)

[6. Bibliographie 24](#_Toc190089153)

[6.1. Référentiel abréviation 24](#_Toc190089154)

[6.2. Veille technique 24](#_Toc190089155)

[6.3. Documentation technique 25](#_Toc190089156)

# 1. Contexte

## 1.1. Contexte théorique

Le programme national de la formation explicite le contexte théorique nécessaire à la compréhension de cette Situation d’Apprentissage et d’Evaluation (SAE). Après un bref descriptif de la SAE, les attendus et les objectifs sont ensuite détaillé.

### Descriptif

Extrait du programme national :

*L’étudiant est mis en situation de développement, de test et de livraison d’un outil décisionnel, selon un cahier des charges imposé. L’étudiant doit savoir travailler avec des approches professionnelles du développement logiciel (méthode agile, gestionnaires de version de code, ...) et se confronter aux exigences d’une gestion de projet devant se solder par la livraison d’un produit logiciel développé à plusieurs, de manière itérative et incrémentale, dont on attend qu’il soit fonctionnel.*

### Attendus et objectifs de la SAE.

Extrait du programme national :

*Après la phase d’analyse des besoins et d’identification des spécifications de l’outil logiciel attendu, le développement d’une solution décisionnelle doit se faire à partir de méthodes de développement adaptées et professionnelles. La validation des programmes, passant par une phase de test et de recette est une démarche fondamentale pour assurer la qualité du programme.*

*En tant que développeur décisionnel, l’étudiant pourra être amené à concevoir, réaliser et maintenir des applications aidant la production, en collectant et stockant les données, puis tester les différents composants de la solution décisionnelle mise en place, et intégrer et mettre en production le développement.*

*Les objectifs de cette SAE sont les suivants :*

* *Mettre en œuvre la démarche décisionnelle dans sa globalité.*
* *Amener l’étudiant à s’approprier les approches professionnelles du développement logiciel.*

Pour répondre à l’objectif : Mettre en œuvre une démarche décisionnelle dans sa globalité, l’étudiant doit concevoir un outil décisionnel en se basant sur l’ensemble du cycle de la donnée. Il doit développer un outil s’appuyant sur la collecte de données, puis déployer des processus de traitement et de stockage ainsi que des outils d’analyse et de valorisation des données.

Afin de répondre au second objectif : S’approprier les approches professionnelles du développement logiciel, l’étudiant doit réaliser ces tâches en utilisant des méthodes de travail professionnelle. Pour cela, il sera amené à utiliser des outils de collaboration et devra mettre en place une démarche de test garantissant la qualité de son travail et de ses développements.

## 1.2. Contexte matériel et environnement

Afin de répondre aux objectifs de cette SAE : la mise en place d’une démarche décisionnelle globale en adoptant une démarche professionnelle, les étudiants ont pu bénéficier de cours théoriques et de séances pratiques sur divers outils et méthodes de travail professionnelles. Des outils matériels et logiciels ont été mis à leur disposition pour leur offrir une grande liberté sur leurs choix technologiques. Les étudiants ont aussi été soumis à un cahier des charges défini par leurs enseignants, précisant les problématiques professionnelles auxquels ils sont confrontés ainsi que les attendus pratiques de la SAE.

Dans le cadre de ce projet, des nano ordinateur Raspberry PI ont été mis à la disposition des étudiants. Confronté à une technologie nouvelle, ils ont dû, dans le cadre de cette SAE, découvrir les outils matériels et logiciels accompagnant ce nouveau matériel afin de répondre à des problématiques professionnelles.

## Cahier des charges et objectifs

Faire le lien entre les objectifs du PN et le contexte matériel => attendus des différents enseignants.

Les objectifs de ce projet défini par le programme national sont les suivants :

* Mettre en œuvre la démarche décisionnelle dans sa globalité.
* Amener l’étudiant à s’approprier les approches professionnelles du développement logiciel.

Afin de répondre à l’objectif : M…, il les étudiants doivent réaliser un outil de monitoring basé sur une station de collecte de données. Cette station de collecte de données se doit d’être déployée sur les Raspberry PI via l’utilisation de différents capteurs. Ce matériel est mis à disposition des étudiants dans le cadre de cette SAE. Ces derniers doivent déployer les stations de collecte et assurer le cycle de vie de la donnée afin de pouvoir valoriser les données avec leur outil de monitoring. La mise en place de diverses stations pour assurer la collecte, le traitement, le stockage, l’analyse t la valorisation des données est un point essentiel dans ce projet.

Concernant la réalisation de l’objectif : R…, les étudiants doivent réaliser le travail demander en utilisant leur savoirs et compétences développés au cours de cette année. Ils ont pu bénéficier de cours sur la mise en place de tests, sur l’utilisation d’outils collaboratifs, et sur des méthodes de travail professionnelles. Aussi, les étudiants doivent répondre à des problématiques nouvelles accompagnant la découverte d’un nouveau contexte : l’Internet des Objets. La compréhension et la réappropriation de ce contexte sont essentielles afin de professionnaliser leur approche.

Afin de familiariser les étudiants au contexte professionnel, des livrables leurs sont demandés. Des rapports détaillant leurs travaux réalisés leur sont demandés. De plus, la mise à disposition du code et des différentes solutions implémentées est essentielle.

Le travail de groupe, l’adaptabilité et l’autonomie sont au cœur de ce projet.

# 2. Analyse

## 2.1. Veille technique

Afin de mener une veille technique et de découvrir les problématiques de l’internet des objets (*Internet of Things* - IoT), nous avons été orienté vers des ressources par nos enseignant.

Nous avons ainsi pu nous renseigner sur divers projets et technologies existants autant sur l’internet des objets que sur le monde de la science des données en générale. Ce fut intéressant de découvrir de nouveaux outils, méthodes et processus au travers d’interview et de conférences. J’ai aussi continué cette veille technique tout au long du projet lors de recherche pour l’implémentation d’idée nouvelle dans nos développements.

### 2.1.1. Webinaire DataGrandEst

Le sujet du webinaire est : « Les capteurs de données en temps réel ». Après une brève introduction par les organisateurs, le webinaire se divise en trois parties qui correspondent chacune à l’intervention de l’un des intervenants professionnels. Chacun de ces intervenants présente son projet pendant une dizaine de minutes avant de répondre à une série de questions sur son projet. Les projets des différents intervenants s’inscrivent tous dans le sujet du webinaire, à savoir une exploitation professionnelle des données collectées par des capteurs en temps réel. Chacun présente un projet très différent avec une approche qui lui est propre.

#### Présentation de DataGrandEst

Pour améliorer la connaissance des territoires, l'Etat et la Région Grand Est ont mis en place une démarche partenariale d'échange de données avec les acteurs publics de l'aménagement du territoire, intitulée « DataGrandEst ». Cette dynamique s'inscrit dans la poursuite de la mise en œuvre de l'infrastructure européenne des données géographiques « INSPIRE » et d'une démarche partenariale *Open Data*. La plate-forme DataGrandEst propose ainsi à ses partenaires et au public des services de recherche, visualisation, téléchargement et transformation de données conformes à INSPIRE et aux principes de l'*Open Data*.

La démarche vise 4 objectifs principaux:

* Piloter: concevoir un nouveau cadre de gouvernance et de confiance de la donnée, en y intégrant notamment les acteurs privés.
* Produire: dynamiser la production de données de référence et enrichir la plateforme de données d’intérêt local.
* Partager: accélérer l’ouverture des données et accompagner les acteurs dans la publication de leurs données
* Valoriser: développer les usages de la donnée, animer la démarche régionale, favoriser les réutilisations.

#### Projet 1 : StrasApp – Mathias Treffot

##### Présentation

StrasApp est une application développée à destination des habitant de l’Europole de Strasbourg. Le but de cette application est de rendre la vie des citoyens de l’Europole de Strasbourg plus simple. Proposer un suivi en temps réel des activités et des évènements présents dans l’Europole.

L’application propose un suivi en temps réel des niveaux de fréquentation de lieux communs (parking, piscines, établissement publiques, …). Elle propose aussi un service de signalement en temps réel des évènements ou défaut urbains présents dans la ville (ex : accidents). D’autre fonctionnalité comme l’agenda ou la possibilité d’avoir un suivi de ses démarches administratives sont intégrées à l’application.

Un système de notification est mis en place pour gérer les évènements en semi temps réels. Ces notifications sont envoyées aux utilisateurs après un signalement ou un évènement particulier (ex : pic de pollution ou encore présence d’un accident). Ces notifications sont catégorisées en familles en l’utilisateur peut choisir les familles de notifications qui l’intéresse.

Si une personne doit aller se faire recenser à la mairie sur les listes électorales de la ville car elle vient d’emménager récemment, elle peut consulter sur l’application l’état du trafic routier et connaître le temps nécessaire pour aller à la mairie tout en ayant connaissance de la présence ou non d’accidents sur la route grâce au système de notifications. Elle peut aussi, avec l’application, connaître le niveau de fréquentation de la mairie et anticiper l’attente à laquelle elle devra faire face si elle décide d’y aller. Si cette personne se décide à ne pas aller à la mairie aujourd’hui, l’application lui rappellera la date des prochaines élections afin que la personne puisse s’organiser en conséquence pour passer à la mairie un autre jour.

##### Aspect technique

La collecte des données se fait par beaucoup de sources différentes. Elles viennent soit de sources officielles (données de la collectivité de Strasbourg), soit de l’*Open Data*, de sources externes (entreprises, association comme Atmo Grand Est).

La multiplicité de ces sources permet de croiser les données et d’offrir une panoplie de fonctionnalité et de services très différents.

Les données sont récupérées avec des capteurs (ex : passage d’un usager dans un tripode) puis elles sont converties en flux JSON avant d’être collectée par StrasApp avec une API et elle est enfin affichée dans l’application.

Cette application ne produit pas de données en *Open Data*, elle ne fait que réadaptée et centraliser les données de diverses sources pour les mettre au service de l’utilisateur.

Remarque : Le code est propriétaire malgré l’utilisation de données en accès libre (provenant notamment de <https://data.strasbourg.eu/pages/accueil/>)

#### Projet 2 : Atmo GrandEst – Éric Herbert

##### Présentation

Atmo GrandEst est une association à but non lucratif agrégée par le ministère de l’environnement. Leurs missions sont diverses, il s’agit principalement de surveiller les niveaux de pollution et d’assurer un suivi de la qualité de l’air à l’aide de données collectées avec des capteurs. Mais Atmo GrandEst sert aussi de référent en matière de pollution et accompagne les entreprises et les collectivités dans la mise en place de projet de dépollution. Enfin, ils sont aussi un acteur de la recherche dans le domaine de la pollution (recherche de solution nouvelle pour faire face aux enjeux environnementaux en matière de pollution).

Voici les atouts développés par Atmo GrandEst :

* Mesures : Atmo GrandEst collecte un grand volume de données avec leurs stations de collectes de données (un total de 98 stations à l’heure de la vidéo). Ces stations sont équipées de divers capteurs leur permettant de mesurer et de prélever un grand nombre d’échantillons (176 préleveur et analyseurs). Les analyses effectuées concernent près de 31 polluants différents.
* Inventaire : Atmo GrandEst propose un inventaire du suivi de la qualité de l’air avec un travail d’archivage des données historique depuis 1990 en matière de pollution. Ce suivi historique se fait selon 72 critères et se base principalement sur l’échelle de l’IRIS (Ilots Regroupés pour l'Information Statistique).
* Modélisation : Atmo GrandEst propose un outil de modélisation qui offre un suivi de la qualité de l’air à un niveau inter-régional. Cet outil de modélisation est retrouvable sur le site d’Atmo Grand Est. Il se présente sous forme de carte interactive permet d’avoir un suivi précis du niveau de qualité de l’air. Il est possible d’obtenir des indicateurs plus précis sur les villes du Grand Est (Metz, Strasbourg, Nancy, …)

##### Aspect technique

Au vu des moyens déployés pour la collecte des données (nombre de stations de collecte), Atmo GrandEst gère de grand volume de données sur des échelles géographiques et temporelles très différentes. D’un point de vue géographique, les données peuvent être collectées à l’échelle d’un quartier jusqu’à un niveau régional voir transfrontalier (Allemagne, Suisse ou encore Luxembourg). Sur la dimension temporelle, l’association propose des données historisées sur 30 ans, les données des observations quotidiennes mais elle propose aussi des modèles prévisionnels en se projetant dans un futur proche. Atmo GrandEst utilise des données de l’ordre des 15 min voir 30 min pour mener ses analyses.

Pour la collecte de ses données, Atmo GrandEst utilise des micros capteurs. Dans la vidéo Éric Herbert détaille le fonctionnement basique d’un micro capteur. Ces micro capteurs sont composés d’un élément sensible qui va réagir en fonction de l’information qu’il capte (ex : présence d’un polluant dans l’échantillon d’air collecté). Cet élément sensible va être associé à une partie électronique qui va transformer l’information perçue par l’élément sensible en signal électrique. Cette association est appelée capteur. Enfin, quand plusieurs capteurs sont installés ensemble, on parle de système capteur.

Ces capteurs peuvent être déployés dans différents contexte : sur du mobilier urbain, à l’intérieur d’un bâtiment, au-dessus d’un véhicule, …

Atmo GrandEst est très concerné par la mise en place, l’utilisation mais aussi le conseil et la réglementation autour des capteurs car ils sont autant utilisateurs (suivi de la qualité de l’air, et recherche), que conseiller sur l’utilisation, la mise en place des capteurs ainsi que sur les différents enjeux accompagnant cette technologie.

Atmo GrandEst a mis en place des procédures de test et de validation des données avec des mesures de références et des tests réalisés sur des atmosphères simulées en laboratoire. La mise en place de ces tests est très importante car ils permettent d’avoir des données fiables et qualitatives. De plus, étant données que l’association est un fournisseur de données en *Open Data*, beaucoup d’autres organisme dépendent de la qualité de leurs données (ex : StrasApp)

Les données récupérées par les capteurs sont envoyées sur un *Cloud* où Atmo GrandEst va aller chercher les données pour les intégrer à sa propre base de données à l’aide d’une API. Ces données vont ensuite être utilisées pour le développement de graphiques et visualisations que les fournisseurs (propriétaire du capteur – ce n’est pas toujours Atmo GrandEst) et Atmo GrandEst vont pouvoir mettre en avant sur leur plateforme respective.

#### Projet 3 : VigieCrues – Fabrice Hery

##### Présentation

VigieCrues est un service public gérée par le Ministère de la transition écologique. Il a pour principal mission d’assurer une surveillance et un suivi du niveau des cours d’eau pour prévenir les futures inondations. Cette surveillance a pour but d’assurer une meilleure prévention de la population, d’une plus grande efficacité des services d’intervention (ex : pompier) et de permettre à chaque citoyen de se tenir informer sur les risque de crues sur diverses échelle géographiques (du département jusqu’à l’ensemble du territoire)

La collecte se fait par des stations de capteur (2 à 3 capteurs par station – 5000 stations sur le territoire français) qui mesure la hauteur d’eau des fleuves et affluents. Ces stations envoient leurs données sur une base de données placée sur un serveur (AQUAREEL) avant d’être transférées sur une base de données centralisée et protégée. Cette base de données (HYDRO 3) sert aux différentes applications comme VigieCrues. L’échelle temporelle des données peut varier en fonction des régions de leur vulnérabilité. Par exemple, les régions à fort régime torrentiel mesurent des données toutes les 5 minutes alors que d’autres régions vont envoyer des données de l’ordre de la demi-heure.

VigieCrues est composé de deux services principaux : l’unité d’hydrométrie et le service de prévision des crues.

La mission principale de l’unité d’hydrométrie de corriger les données quotidiennes récupérée avec les capteurs afin d’avoir une information la plus sûre possible.

Le service de prévision des crues a développé un outil à partir des données collectées afin d’avoir un aperçu de l’état des différents affluents de France en quasi-temps réel, en plus de leurs prévisions. Cet outil propose des cartes interactives et des modèles prévisionnels sur les crues afin de les anticiper et mieux pouvoir prendre des mesures adaptées en fonction de la gravité des crues et de leur localisation. Le service VigieCrues propose une catégorisation des cours d’eau en temps réels en fonction des risques de crues encourus.

### 2.1.2. Conférences Devoxx

Devoxx est une conférence réunissant des professionnels du développement dans divers domaines. Créer en Belgique en 2001 sous le nom de JavaPolis, la conférence changera de nom en 2008 pour devenir Devoxx. Sa version française : Devoxx France, verra le jour en 2011. Lors de cette conférence, de nombreux ateliers et présentations ont lieux, visant à présenter les nouvelles innovations dans le domaine du développement et de l’informatique. La plupart des vidéos présentées ci-dessous viennent de conférences Devoxx France.

#### Conférence 1 : Kit de survie pour l'IoT façon DIY - Laurent HUET

##### Résumé

La présentation a pour but de sensibiliser le public sur la facilité de construction d’un réseau IoT. Le présentateur, Laurent HUET, aborde les différents points clé d’un réseau électronique en commençant notamment par la Raspberry PI et d’autre microcontrôleurs équivalents. Il parle ensuite des coûts des différents composants et de leur rôle électronique (ex : les résistances permettent d’éviter une sur tension des autres composants). Il insiste sur la facilité de mise en place et montre ainsi que l’IoT, loin des laboratoires ou de demander des gros moyens financiers, peut être réalisé très facilement à une échelle locale sans un gros budget.

##### Ressenti

J’ai choisis cette présentation car il me semblait intéressant de découvrir les divers composant d’un réseau électronique basé sur Raspberry PI (ou équivalent) car cela s’inscrit dans le cadre du projet. De plus, la vidéo m’a permis de comprendre que la mise en place de ces circuits est loin d’être inaccessible et que nous pouvions tout à fait être en mesure d’en mettre un en place après quelques heures de travail. Enfin cette vidéo est une bonne introduction à l’aspect pratique et technique de l’IoT.

#### Conférence 2 : Premiers pas avec un microcontrôleur et Google Cloud IoT Core - Gautier MECHLING

##### Résumé

Cette présentation Devoxx Fr nous permet d’avoir un aperçu des bases sur les microcontrôleurs et la récupération des données des capteurs. Gauthier MECHLING commence sa présentation en introduisant les microcontrôleurs, leurs différents composant et en expliquant leur principale fonction : faire tourner du code en continu sur un processeur indépendant. Après cela, il montre comment « flasher » un microcontrôleur (= intégration de code) en live. Il explique ensuite une partie de connectique et d’électronique en créant un capteur à son circuit et récupérant les données du capteur à l’aide du microcontrôleur. Enfin il envoie ces données sur un Cloud et s’en sert pour créer des visualisations en temps réel.

##### Ressenti

J’ai décidé de choisir cette présentation car elle me semblait essentielle pour comprendre les bases de l’IoT et des microcontrôleurs. En effet, notre projet concerne la récupération de données à partir de capteurs et de microcontrôleur, cette vidéo semblait tout à fait adaptée à cette problématique. De plus, le présentateur appuie tous ses points théoriques avec un exemple pratique bien expliqué. Ce qui m’a le plus attiré dans cette présentation était la mise en place des différentes connections (physiques ou numériques) entre les composants (microcontrôleur, capteur, *Cloud*) ainsi que le cheminement de la donnée dans ce circuit.

#### Conférence 3 : La sécurité́ dans l'IoT difficultés, failles et contre-mesures - A. Duque

##### Résumé

Cette présentation Devoxx nous permet d’avoir un aperçu des concepts de la sécurité informatique appliquée à l’IoT. La présentation débute par une brève introduction de l’IoT et la sécurité, notamment en caractérisant des données sécurisées comme étant disponibles, intègre, et avec une confidentialité garantie. Alexis Duque introduit ensuite les dangers auquel est confronté l’IoT, comme l’environnement des objets connecté, la grande variété des appareils dans un réseau IoT et l’étendue du champ d’attaque de ces appareils (physique et numérique) entre autres. Il insiste sur le fait que les utilisateurs et les entreprises sont encore trop peu sensibilisés à ces problématiques due à la nouveauté de l’IoT. Il détaille ensuite toutes les formes les plus courantes d’attaque informatique sur des objets connectés en donnant des exemples. Il cite notamment des attaques d’escalade des privilèges, d’injection de code et d’attaque par Bluetooth (*Blue Born*). Il insiste notamment sur le fait que c’est le rôle des développeurs d’être conscient des problématiques de sécurité dans l’IoT et que parfois, il suffirait d’être aussi prudent que sur le développement des applications web pour éviter certaines vulnérabilités.

##### Ressenti

J’ai décidé de choisir cette présentation car les problématiques de sécurité me semblaient pertinentes bien que cela ne relève pas forcement de notre domaine. Etre sensible aux problématiques de sécurités est quelque chose d’important quel que soit le domaine d’application afin éviter des problèmes, parfois mineurs, qui peuvent souvent avoir de grosses conséquences pour les individus et/ou les entreprises. De plus, nous vivons dans un monde où l’IoT est omniprésent, connaître les failles ou des moyens de se prémunir des attaques, par l’adoption de bonnes pratiques, est toujours utile que cela soit en tant que développeurs ou en tant qu’utilisateur.

#### Conférence 4 : Le Big Data, de la récolte jusqu'à la production à grande échelle ! - Alexis Slawny & Mathias Kluba

##### Résumé

Cette présentation sur le BigData commence par une brève introduction au concept et à ses domaines d’application (marketing, surveillance, santé, aéronautique, etc…). Ensuite les présentateurs détaillent comment fonctionne le BigData ou plutôt ses problématiques fondamentales, à savoir le traitement d’un grand volume de données. Il explique les différentes solutions existantes : scalabilité verticale (augmentation des performances d’une machine) et horizontale (division des traitements de données sur plusieurs machines – cf. Hyper convergence). Ils donnent ensuite un exemple pour expliquer ces concepts fondamentaux. Après cela, ils présentent les différentes technologies utilisées dans le BigData : Hadoop, Spark, NoSQL, et leur fonctionnement basique respectif. Ils insistent beaucoup sur la scalabilité horizontale et sur la répartition de l’exécution des traitements sur plusieurs machines (ressemble au *multi-processing* Python vu en cours, étendu sur plusieurs machines). Enfin, les présentateurs appuient leur propos avec divers exemples de traitement de données en Open Source (meilleure bière de Paris et algorithme de choix de la meilleure route commerciale dans un jeu vidéo).

##### Ressenti

Cette présentation m’a grandement intéressé car elle permet de revoir les fondamentaux de Big Data et elle fait écho à des concepts vus en cours (*multi-processing*). De plus, elle est aussi intéressante dans le cadre de ce projet car au vu des données à traiter (très variées et volumineuses), avoir une connaissance basique du BigData semble essentiel. Enfin le dernier aspect qui m’a intéressé est le fait que les applications BigData présentées ne soient pas restreintes à un seul domaine d’application mais semblent bel et bien utilisable dans de nombreux projets.

### 2.1.3. Recherches supplémentaires

En parallèle de la veille technique que nous avons menée, orientée par nos enseignants, nous avons pu faire des recherches personnelles complémentaires. Si une grande partie de ces recherches en autonomie se sont portées sur des aspects techniques (utilisation d’outils, de logiciels ou de librairie de programmation), nous avons aussi pu nous intéresser aux différents projets et solutions existant.

#### Recherche 1 : Flux Vision – Orange

Flux vision est un outil développé par l’opérateur français Orange. Il s’agit d’un outil visant à étudier les flux humains à partir de données captées par les opérateurs et de données GPS. Le but est de fournir un service qui fournit un ensemble d’indicateurs clés à partir de l’analyse des fréquentations selon les besoins du client. L’idée est de fournir un ensemble d’information très précis sur le type de flux humain en étudiant les profils issus des données captées mais aussi en garantissant une précision géographique maximale par l’utilisation de données GPS. Cet outil est utilisé par de nombreuses industries comme le tourisme ou le retail. L’entreprise a mis un accent très prononcé sur le respect des règlementations et les contraintes légales liées aux données imposées par la Commission Nationale de l’Informatique et des Libertés – CNIL – et le Règlement Général de la Protection des Données - RGPD. Afin d’atteindre des objectifs de fiabilité de la données, Orange pondère les données récupérées par ses parts de marchés locales. En effet, s’ils sont le leader sur le marché des opérateurs mobiles, ils ne possèdent pas le monopole.

#### Recherche 2 : Chasing Your Tail With a Raspberry Pi – Matt Edmondson

Ce projet a été présenté lors de la conférence Black Hat de 2022. Black Hat est une conférence sur la sécurité informatique, réunissant les acteurs du domaine autour de présentations et réunion. Matt Edmondson, a présenté lors de cette conférence un outil portatif, développé sur Raspberry PI permettant de savoir si l’on est suivi. Cet outil scan les appareils environnant afin d’enregistrer leurs adresses MAC. En comparant les diverses adresses MAC enregistrées au cours du temps, il peut ainsi savoir s’il est ou non suivi. Cet outil s’adresse avant tout à un public concerné par les questions de filatures ou de confidentialité, mais il peut être intéressant dans notre cas d’utiliser par exemple, un système de scan des appareils pour réaliser des mesures de fréquentation. L’outil se sert de python et du package Kismet, développé notamment pour scanner les appareils environnant connecter au Wifi et au Bluetooth.

#### Recherche 3 : Horizon Europe

Horizon Europe est un programme d’innovation lancé par l’Union Européenne (UE) visant à maintenir l’UE en tant que l’un des leaders mondiaux de la recherche et de l’innovation. L’un des objectifs de ce programme est de renforcer la base scientifique et technologie de l’UE. Dans le cadre de ce programme, l’UE a adoptée *l’European Data Strategy,* qui a pour but de faire de l’Europe l’un des leaders mondiaux sur le sujet de la donnée. Il s’agit de mettre en place un cadre législatif encadrant le domaine et de mettre en place un marché unique de la donnée, international, qui bénéficierait à tous les acteurs. De plus, dans cette idée de recherche et d’innovation, l’UE a adopté une politique Européenne en matière d’Internet des Objets (IoT – Internet of Things) afin d’accompagner les acteurs du domaine vers le monde de demain.

#### Recherche 4 : Industrie 4.0

L’industrie 4.0 est un terme utilisé pour définir le secteur de l’industrie après la quatrième révolution industrielle. Les trois premières révolutions industrielles désignent les changements majeurs de l’industrie ayant eu lieu à l’émergence et la démocratisation de :

* La machine à vapeur (1769).
* Le pétrole et l’électricité.
* L’électronique et les technologies de l’information.

La quatrième révolution industrielle s’apparente à une transformation numérique du terrain. Ce changement a pour d’offrir une productivité et une flexibilité accrues. Avec l’ère de la personnalisation des produits et de l’instantané, l’industrie doit pouvoir fournir un service répondant aux besoin du client de façon efficace, en temps réel. L’émergence de nouvelle technologie dans les domaines de l’IoT, du BigData, du *Cloud Computing* et de l’Intelligence Artificielle accompagne cette nouvelle révolution en offrant des solutions techniques aux nouvelles problématiques de l’industrie.

Un exemple d’entreprise ayant mis en pratique l’industrie 4.0 est Volkswagen :

Ils ont investi près de 4 milliard d’euro dans des projets de digitalisation. Ils ont ainsi pu moderniser leur chaine de production de véhicule afin d’offrir la possibilité de produire des véhicules personnalisés en plus de leurs véhicules traditionnel produit en série. Le véritable progrès se situe notamment dans le fait que quel que soit le véhicule produit, leur chaine de production reste interrompue. De plus, ils ont su mettre en place des stations de monitoring afin de centraliser les informations sur la chaine de production sur un tableau de bord unique.

D’autre entreprise comme SIEMENS, Toyota ou Bosch sont aussi des acteurs majeurs de l’industrie 4.0.

### 2.1.4. Synthèse

Faire une synthèse des recherches réalisées. Faire un lien clair avec la SAE.

L’ensemble de ces recherches et de cette veille technique réalisée m’a permis de bien comprendre les enjeux professionnels liés à l’IoT et au Big Data. En effet, les nouvelles problématiques professionnelles liées à une société de services, de produits personnalisés, et de réponse quasi-instantanée aux besoins accélère l’innovation technologique et le déploiement de solutions nouvelles. Aucun domaine professionnel n’est épargné, qu’il s’agisse de services publiques, de villes intelligentes, d’industrie, de tourisme, de santé, de sécurité ou encore de banques et assurances, tous ont maintenant besoin de s’adapter et d’utiliser les solutions en terme d’IoT ou de Big Data.

D’un point de vue technique, cette veille technique m’a permis d’appréhender la grande diversité de solutions existantes. Il existe autant de solution qu’il y a de problématique. Les choix technologiques à entreprendre sont dépendant des besoins liés au projet. L’identification des besoins techniques à partir de problématiques professionnelles est le principal enjeu afin d’utiliser les outils appropriés et de pouvoir répondre au mieux à ces problématiques.

## 2.2. Titre

### 2.2.1. Définition du projet

Afin de respecter le cahier des charges et de réponde à l’objectif principal de cette SAE : réaliser un outil de monitoring, nous avons choisi de développer un outil permettant de monitorer les salles de l’I.U.T. à l’aide des Raspberry PI. Pour ce faire, nous avons pour objectif de mettre en place une station de collecte de données, équipées de capteur, puis des stations de traitement de ces données, avant de paramétrer une ou plusieurs stations de monitoring, équipées de tableau de bord permettant de valoriser nos données. L’idée était de pouvoir fournir des indicateurs clés sur des problématiques précises en fonction des évènements ou des périodes de la journée (incendie, nuit, fréquentation des salles, …).

### 2.2.2. Solutions pratiques envisagées

Plus concrètement, il pourrait être intéressant de mettre en place des mesures et calculs de fréquentation des salles de l’I.U.T. en fonction des données captées par le Raspberry PI. L’analyse de la température, du taux de CO2, du bruit, du détecteur de mouvement et d’une photo de la pièce nous permettrait par exemple de calculer le nombre de personnes dans la pièce. Bien que certaines données ne permettent que de faire des approximations sur le nombre de personnes présentes, leur croisement pourrait permettre de gagner en précision. Cela pourrait servir à la mise en place de panneau d’information à destination des étudiants en identifiants les points de passages ou les salles les plus utilisés pour améliorer la transmission de l’informations. Un autre usage possible serait de guider la prise de décision dans les travaux de rénovation : faciliter l’accessibilité des salles très fréquentées ou repenser l’aménagement des locaux moins populaires.

Un autre objectif intéressant serait de repérer des évènements anormaux se déroulant dans les salles de l’I.U.T. à partir des données, par exemple une augmentation anormale de la température sur une durée prolongée pourrait être significative d’un incendie comme la présence d’une personne en pleine nuit dans une salle pourrait être synonyme d’intrusion. Le croisement de plusieurs types de données rend d’autant plus intéressant leur analyse afin d’identifier ces différents évènements. Un incendie sera, par exemple, non seulement repérable par une augmentation permanente de la température, mais aussi par une augmentation significative du taux de CO2 et par la présence d’un bruit de fond assez marqué (crépitement des flammes). Un autre exemple serait la mise en place d’une alerte automatique pour l’aération des pièces : quand le taux d’humidité et le taux CO2 deviennent trop importants, il faut aérer la pièce. Cela pourrait aider dans la prise de décision de mesures de remédiation (en identifiant les causes potentielles d’évènements indésirables) ou d’action immédiate (ouverture de la fenêtre).

Enfin, il pourrait tout à fait être possible, à partir des Raspberry PI de mettre en place un réseau de caméra de surveillance. En captant le flux vidéo fourni par la caméra de différentes stations, il serait possible de monter un système de vidéosurveillance. Il serait théoriquement possible, à partir de ces différents flux vidéo d’automatiser la gestion des absences des étudiants. En effet, la démocratisation des modèles de reconnaissance faciale permet maintenant d’identifier de façon toujours plus précise les personnes présentes sur un flux photo ou vidéo. En enregistrant l’heure à laquelle un visage est identifié par le système, il est possible de notifier un retard ou une absence (cela reste conditionné à la fiabilité du modèle, à sa précision et à sa capacité à détecter tous les visages sur une photo d’ensemble). Malgré les nombreuses contraintes, cela reste théoriquement possible. De plus nous serions confrontés à des problématiques légales (cf. Règlement Général de Protection des Données – RGPD), dû à la collecte, au traitement, à l’analyse et au stockage de données personnelles. Un outil destiné aux étudiants permettant d’être informé de leurs diverses absences, et de leurs conséquences pourrait être développé.

## 2.3. Chaine décisionnelle

### 2.3.1. Architecture

Pour mettre en œuvre une démarche décisionnelle globale, il a fallu définir les différents maillons de notre chaine décisionnelle. Cette chaine décisionnelle repose sur les différentes étapes du cycle de la donnée : la collecte, le traitement, le stockage, l’analyse et la valorisation des données. En partant de ce principe, nous devons ensuite détailler le rôle de nos différentes stations, ainsi que leurs interactions avec le reste du système. Plus simplement, il a fallu définir l’architecture de notre chaine décisionnelle.

Une chaine décisionnelle simple (*cf. Figure 1*) est composée de 4 stations principales :

* La station de collecte : Implémentée sur le Raspberry PI, elle permet de récupérer les données des capteurs et de les envoyer à la prochaine station.
* La station de traitement : Elle applique des calculs et analyses sur les données collectées. C’est aussi sur cette station que nous pouvons insérer les données en base de données.
* La station de stockage : Il s’agit d’une base de données contenant les données captées et/ou leurs dérivations.
* La station de monitoring : Permet d’avoir un suivi et des indicateurs clés sur les données présentes dans la base de données. Réservée à l’utilisateur, c’est le dernier maillon de la chaine et l’objectif final de ce projet.

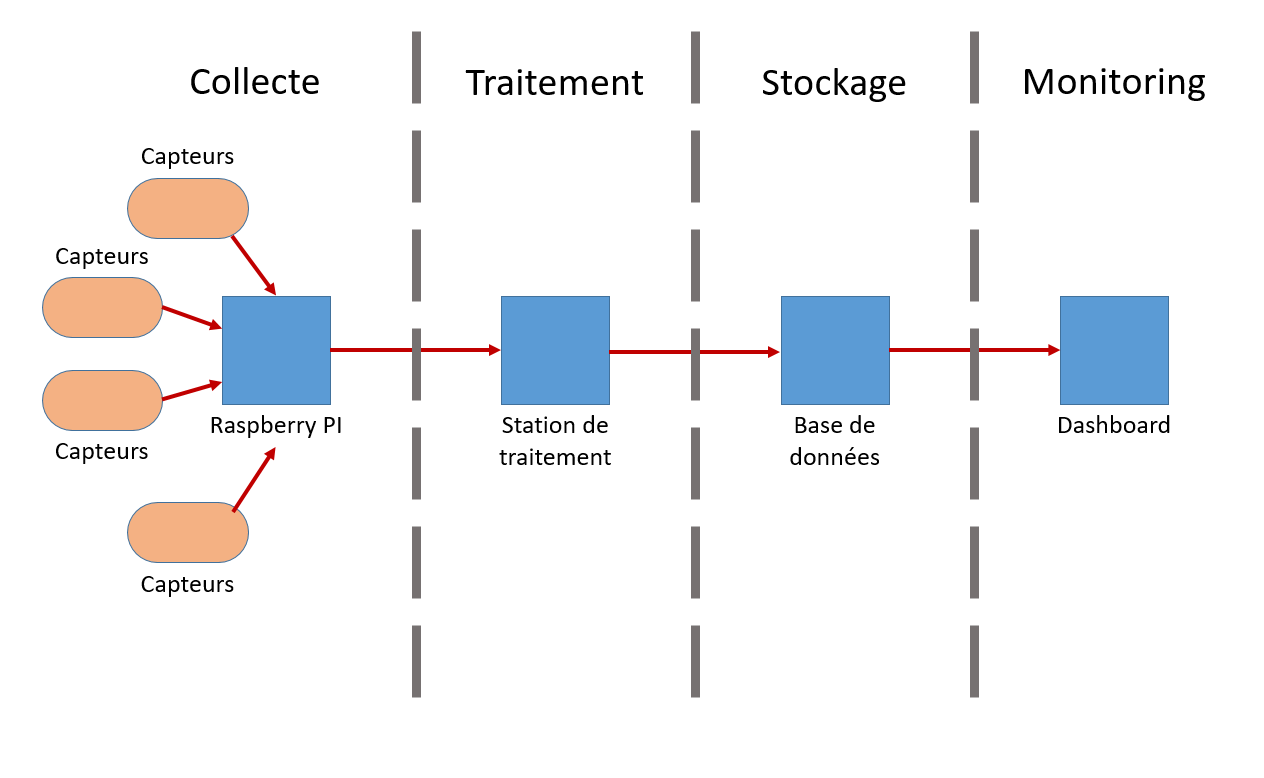


Figure 1

Bien que l’organisation de ces stations puisse différer, cet ordre applicatif permet de simplifier l’organisation et le déploiement pratique du système. Bien que dans le schéma ci-dessus (Figure 1), les stations relatives à chaque tâche soit unique, il est possible d’imaginer de multiples stations de collecte, de traitements, de stockage et de monitoring. Ces stations peuvent être associées à des tâches très différentes ou bien réaliser des tâches similaires selon les besoins et les problématiques auxquelles nous souhaitons répondre (cf. 2.3.2.).

Sur le schéma ci-dessous (*Figure 2*), une architecture plus complexe du système est présentée. Dans cette configuration, les stations de traitements sont associées à des tâches distinctes, parfois indépendantes, parfois complémentaires, qui permettent ensuite de répondre à des problématiques différentes, incarnées par leur station de monitoring respective. Il s’agit ici d’un exemple théorique, bien que nous allons ensuite appliquer un modèle similaire à un cas concret (cf. 2.3.2.).

L’idée est la suivante : nous ne sommes pas obligés de nous limiter à la réalisation d’un unique outil de monitoring, avec un flux linéaire de données. Il est tout à fait possible d’imaginer des traitements multiples, nécessitant plusieurs schémas de base de données indépendants. Il est tout à fait possible d’imaginer divers outils de monitoring, variant selon les besoins, les problématiques mais aussi selon les supports matériels à disposition (ordinateur portable, smartphone, …).

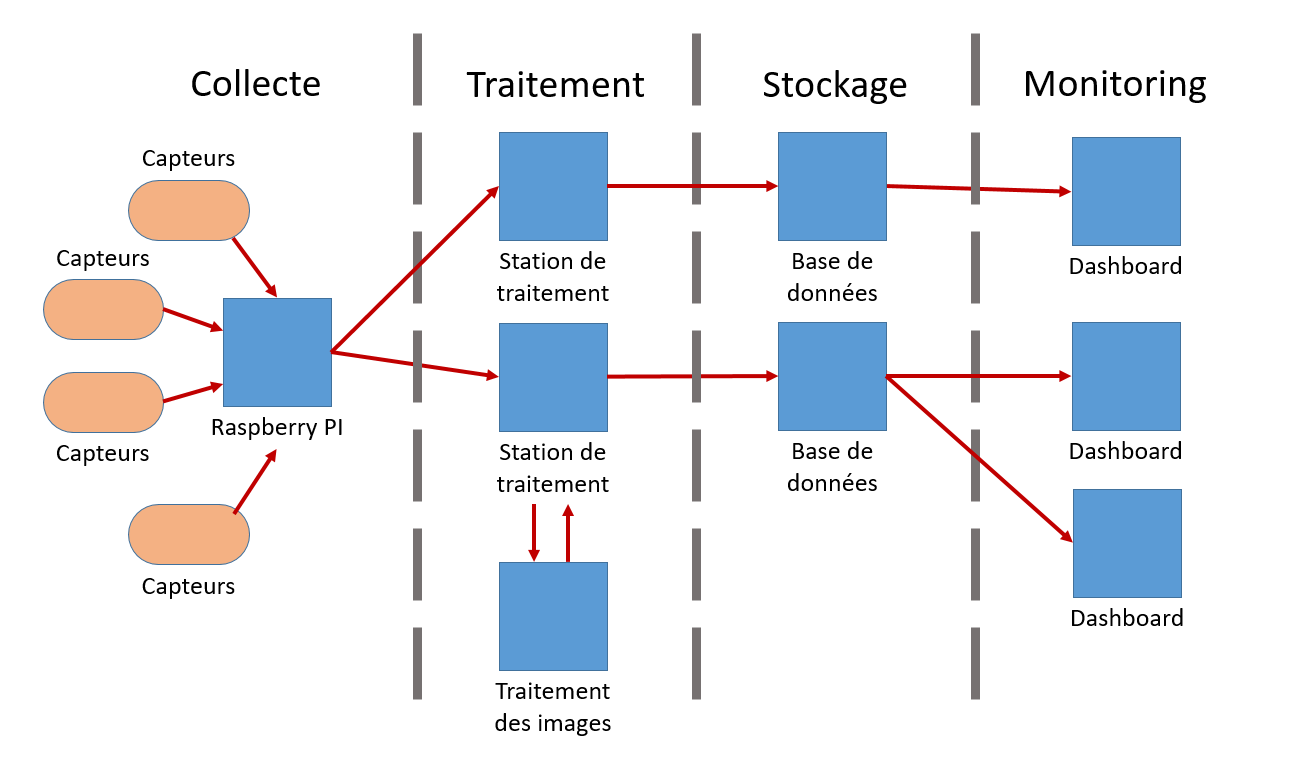


Figure 2

### 2.3.2. Cas d’utilisations pratiques

Cette architecture théorique peut être appliquée aux cas concrets décris précédemment (cf. 2.2.2.).

Dans le cadre de l’estimation de la fréquentation dans les salles de l’I.U.T., nos stations de collecte pourraient être l’ensemble des Raspberry PI installés dans les salles. Avec un système d’intégration de données en temps réel, il est possible de récupérer les données collectées sur une station de traitement : un ordinateur ou un serveur indépendant des Raspberry PI. Ces données peuvent être traitée si besoin, par exemple en calculant le nombre de visages repérés sur une photo prise depuis une station de collecte, puis intégrée dans une base de données. La base de données quant à elle peut être déployée sur un serveur, accessible depuis la station de traitement mais aussi depuis les stations et outils de monitoring. Différents supports peuvent ensuite être utilisés afin de monitorer la fréquentation dans les salles de l’I.U.T. Par exemple, le chef de département des salles associées à un groupe de stations de collecte pourrait être informé en temps réel de l’estimation de la fréquentation dans ces salles, en plus d’être informé de l’état de la salle (fenêtre ouverte, lumière allumée, …). Il pourrait ainsi intervenir pour agir dans ces salles s’il le juge nécessaire. Reposant sur les mêmes informations, un tableau de bord pourrait être mis à disposition des étudiants pour les informer des salles disponibles s’ils souhaitent travailler en autonomie. Un autre outil de monitoring pourrait être par exemple un tableau de bord fournissant des informations et des indicateurs clés à propos de la fréquentation des salles de l’I.U.T. sur une période étendue (mois, trimestre, semestre, …).

Concernant la détection d’évènement anormaux, incendie ou intrusion par exemple, des alarmes spécialisées seraient bien plus efficaces. Cependant il pourrait être envisageable de déployer un système annexe, informant les pompiers ou les personnes responsables en cas d’évènement anormaux. Si l’une des stations de collecte détecte une intrusion, l’intégration de données en temps réels permet à la station de traitement de déclencher un processus de prévention ou d’alarme associé à l’évènement : envoi d’un message ou appel d’un numéro préenregistré par exemple. De plus, malgré l’alerte, la collecte de données devrait continuer. Cela pourrait ensuite permettre, à l’aide d’un tableau de bord ou autre outils de monitoring d’identifier les causes ou les sources d’un évènement : l’intrusion a eu lieu par la fenêtre de telle salle, car un promontoire permet d’y accéder et la fenêtre était resté ouverte. Cela pourrait faciliter la mise en place de mesure de remédiation ou de sécurisation par exemple.

Enfin, dans une problématique de notification des absences et des retards. L’identification des élèves par une station de traitements spécialisées pourrait être reliée à une base de données différentes des cas d’usages précédent. L’un des outils de monitoring associé serait, par exemple, une application destinée aux étudiants pour assurer le suivi de leurs absences et les potentielles pénalités associés à ces dernières.

Grâce à l’architecture détaillée précédemment (cf. 2.3.1.), il serait possible de déployer ces diverses chaines décisionnelles simultanément. Ces diverses chaine décisionnelle reposes toutes sur un ensemble de stations variées. Elles ont pour point communs les stations de collecte de données mais parfois aussi certaine station de traitement voire une base de données. Ce partage de ressources permet de déployer plusieurs chaines décisionnelles, bénéficiant toutes de leur propre architecture, dans une seule architecture globale permettant de répondre à des problématiques variées.

Rajouter un schéma de l’architecture du travail final réalisé (simplifié)

# 3. Situations d’expérimentation (=partie tech)

## 3.1. Gestion du projet

### 3.1.1. Méthodologie

Parler du dépôt Git, de la répartition des tâches, du travail et de l’investissement personnel. Evoquer les méthodes de travail : code expérimental => test => implémentation dans les stations => tests et corrections de bugs (mise en place d’un environnement de tests ?)

### 3.1.2. Outils utilisés

Lister et détailler brièvement tous les outils utilisés dans le projet. Faire des références vers les parties concernées.

## 3.2. Stations de collecte et capteurs

### 3.2.1. Le Raspberry PI

Parler de l’outil en lui-même, ses caractéristiques et spécificités. Evoquer ses cas d’utilisation les plus courants, son système d’exploitation : Linux. Parler de la découverte de Linux : commande de bases, invite de commande, spécificité de l’environnement de travail, etc… Expliquer l’utilisation des gestionnaires de package (apt + pip - python).

Réutiliser documentation technique rapport IOT ? Cf. rapport RPI

### 3.2.2. Les capteurs

Parler des différents capteurs : type de données captées, unités, branchement, contrainte technique, … + Ajouter schéma des branchements de la station + Expliquer et justifier le passage par python (outil préparé par l’enseignant responsable). Expliquer architecture logicielle des stations : main + gateway + autre module (si présent).

Réutiliser documentation technique rapport IOT ? Cf. rapport RPI.

### 3.2.3. Architecture logicielle

Cf. rapport RPI.

### 3.2.3. Station de collecte : premiers traitements

Parler de nos ambitions avec la station de monitoring, les traitements immédiats sur les données, les différentes alertes mises en place, etc… Evoquer le graphe d’état et le rôle des états (a quoi servent les différents composant dans les différents états ?). Détailler la différence entre les différentes inputs et outputs et leurs différents rôles.

Cf. rapport RPI.

## 3.3. Flux et intégration de données

### 3.3.1. Envoi des données en temps réel

Evoquer le choix d’une intégration en temps réel par rapport à une intégration via ETL.

Parler de MQTT et Kafka, discuter du choix technique : avantage/inconvénient et justifier le choix final. Evoquer les différents packages utilisés, la complexité d’implémentation en fonction de la technologie et les contraintes physiques/techniques rencontrées. (Machine virtuelle ?) / Adresse IP variante -> fixe. Evoquer la possible multiplicité des stations émettrice de données. (Plusieurs Raspberry PI). Parler de ‘sudo ifconfig’ et des solutions choisies pour la gestion de l’IP fixe.

#### Système de Queuing : MQTT

Détailler MQTT : particularités, avantages, inconvénients, …

#### Kafka

Détailler Kafka : particularités, avantages, inconvénients, …

### 3.3.2. Traitement des images

Evoquer la particularité des images lors du transfert de données (bytearray). Parler de DeepFace (présentation, implémentation, … => environnement virtuel). Parler ensuite de notre ambition : mesure de fréquence à partir du nombre de visages présents. Evoquer le projet de reconnaissance faciale (piste d’amélioration => /!\ limites légales…). Parler des différents package python utilisés, leurs rôles et particularités dans notre cas d’utilisation.

### 3.3.3. Intégration en base de données

Parler de la récupération des données depuis les différents topics pour les insérer en BD. Parler du package python utilisé. Justifier le choix du SGBD PostgreSQL (pk pas MongoDB ?) => avantage/inconvénients. Contraintes techniques : déploiement en local, passage par le VPN de l’iut impossible => rupture de notre connexion avec la station car changement d’environnement de travail. Solutions ?

## 3.4. Tableau de bord

### 3.4.1. Outil utilisé : Power BI

Justification du choix, parler de la connexion avec la BD + récupération des données en temps réel (cf. Ayoub). Pourquoi temps réel ? Justifier et expliquer choix => parallèle avec l’intégration de données par système de Queuing & Kafka.

### 3.4.2. Valorisation des données

Que voulons-nous représenter ? Comment valoriser les données ? Pourquoi avoir des tableaux de bord différents (si c’est la cas) ? Différencier tableau de bord en fonction des supports (PC, smartphone, …) => problématique d’accès à la BD => limite d’une BD locale.

# 4. Pistes d’amélioration

## 4.1. Matérielles

Parler d’une potentielle collaboration entre département : mise en place des stations dans des salles de classe, dans des boitiers réalisés par des GMP. Contraintes physiques : nécessité de brancher le Raspberry sur batterie ou autre moyen d’alimentation + orientation et positionnement des capteurs (avec/dans le boitier).

## 4.2. Logicielles

Outils utilisables => situation envisagée mais pas mise en place ? Contournement du VPN ou utilisation d’un serveur gratuit pour héberger notre BD => problématique absente en contexte professionnel (les gens préfèrent payer plutôt que de se reposer sur un service instable ou peu fiable).

Parler du Bluetooth => code implémenté mais non fonctionnel => expliquer pourquoi.

## 4.3. Organisationnelle

Comment aurions-nous pu travailler plus efficacement ? Méthode AGILE (SCRUM) ? Meilleure compréhension du sujet et suivi enseignant (nous étions les cobayes)?

## 4.4. Autres ?

Paragraphe facultatif – à compléter si des idées viennent.

# 5. Bilan

Conclusion du projet, notifier les réussites, les objectifs atteints. Revenir brièvement sur les limites. Rajouter un ressenti personnel au bilan ?

# 6. Bibliographie

## 6.1. Référentiel abréviation

Référencer toutes les abréviations utilisées dans le rapport. Sous forme de liste (+définition ?)

SAE : Situation d’apprentissage et d’évaluation

UE : Union Européenne

IoT : *Internet of Things* (Internet des objets)

## 6.2. Veille technique

Liens vers les recherches personnelles, les projets existant, les conférences et articles traitant de sujet en liens avec le projet (voir travail de doc préliminaire + Kismet Raspberry PI – vérifier si l’on est suivi)

Webinaire DataGrandEst : Les capteurs et données en temps réel

<https://www.youtube.com/watch?v=hGrhP5KnqZc>

Flux vision – Orange

<https://www.orange-business.com/fr/solutions/data-intelligence-iot/flux-vision>

<https://www.youtube.com/watch?v=Y0OooJtNYsI>

Projet ‘chasing your tail’ : Kismet – Matt Edmondson

<https://raspberry-pi.developpez.com/actu/335858/Un-outil-anti-pistage-alimente-par-Raspberry-Pi-recemment-concu-permet-de-verifier-si-vous-etes-suivi-en-analysant-les-appareils-environnants-l-outil-serait-efficace-et-facilement-reproductible/>

<https://www.blackhat.com/us-22/briefings/schedule/#chasing-your-tail-with-a-raspberry-pi-26930>

Horizon Europe

<https://www.consilium.europa.eu/fr/policies/horizon-europe/#:~:text=%22Horizon%20Europe%22%20est%20le%20nouveau,projets%20associant%2040%20000%20organisations>.

<https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en>

<https://digital-strategy.ec.europa.eu/fr/policies/internet-things-policy>

Industrie 4.0

<https://merca.team/entreprises-dominent-industrie-40/>

<https://www.ibm.com/fr-fr/topics/industry-4-0>

<https://www.visiativ.com/actualites/actualites/industrie-4-0-definition-et-mise-en-oeuvre-vers-lusine-de-production-connectee/>

## 6.3. Documentation technique

Tous les liens vers les sites officiels de documentations des différents packages. Liens vers les différents sites et forums consultés à des fins techniques.

Kafka python : Data intégrations

<https://kafka-python.readthedocs.io/en/master/>

DeepFace : Reconnaissance faciale

<https://pypi.org/project/deepface/>