

Auteur du rapport :

DOUKANI Yacine

Tuteur IUT :

CARRARA Jean-Paul

Dates du stage :

07/04/25 au 13/06/25

RAPPORT DE STAGE

Automatisation des flux de travail assisté par IA



Auteur du rapport :

DOUKANI Yacine

Tuteur IUT :

CARRARA Jean-Paul

Dates du stage :

07/04/25 au 13/06/25

RAPPORT DE STAGE

Tuteur de stage

BALDAN Joris

Remerciements

Au terme de ces dix semaines de stage, je souhaite exprimer ma profonde gratitude à toutes les personnes qui ont contribué à faire de cette expérience un moment aussi enrichissant que formateur.

Je remercie chaleureusement l'ensemble du personnel des différents services d'HABITALYS, que je ne citerai pas tous individuellement, pour leur accueil bienveillant, leur disponibilité, et leur gentillesse, qui ont facilité mon intégration et m'ont permis de me sentir rapidement à l'aise dans cette nouvelle aventure.

Mes remerciements vont tout particulièrement à la **Direction Financière**, au sein de laquelle j'ai été accueilli avec bienveillance et intégré dans des conditions idéales pour mener à bien mon stage.

Je tiens à remercier très sincèrement l'équipe du Système Informatique pour son accompagnement constant et sa patience, notamment face aux difficultés techniques que j'ai pu rencontrer. Leur soutien a été précieux.

Je suis profondément reconnaissant envers **Monsieur Christophe CASADEI**, pour sa pédagogie, sa disponibilité, et sa gentillesse. Grâce à lui, j'ai pu mieux comprendre certaines missions qui m'étaient confiées, et progresser efficacement tout au long du stage.

Je souhaite également adresser un remerciement très particulier à **Monsieur Joris BALDAN**, mon tuteur de stage. Son accompagnement constant, sa bienveillance, et ses conseils techniques m'ont énormément apporté. Il a su m'orienter à chaque étape du développement, me proposer des solutions adaptées, et m'aider à prendre confiance en moi. Merci pour votre gentillesse, votre écoute et votre implication : vous avez grandement contribué à la réussite de mon stage.

Un grand merci également à **Monsieur Mustapha KAAROUR**, responsable des ressources humaines, pour avoir transmis ma candidature au service informatique et pour s'être présenté à moi lors de mon arrivée.

Je tiens à exprimer mon profond respect et mes sincères remerciements à **Monsieur Bruno GUINANDIE**, directeur général d'HABITALYS, pour m'avoir permis d'intégrer l'entreprise et pour s'être personnellement présenté à moi lors de mon arrivée. Merci pour votre professionnalisme et votre bienveillance.

Une reconnaissance toute particulière à **Monsieur Francis GARCIA**, maire de la ville du Passage d'Agen, dont le soutien auprès d'HABITALYS a été déterminant dans l'acceptation de ma candidature. Je vous en serai éternellement reconnaissant.

Je voudrais aussi adresser un **merci du fond du cœur à ma maman**, qui a été mon plus grand soutien durant cette période. Elle m'a accompagné avec amour, force et détermination

dans ma recherche de stage, alors que je traversais des moments de doute. Maman, je te dédie ce rapport. J'espère que tu seras fière de moi, que tu verras dans ce stage une étape vers les objectifs que je me suis fixés pour te rendre un jour un peu de tout ce que tu as sacrifié pour moi. Je suis fier d'être ton fils.

Je remercie également **Monsieur Jean-Paul CARRARA**, mon tuteur IUT, pour son suivi et son implication durant toute la durée de mon stage.

Enfin, je tiens à remercier l'**IUT Paul Sabatier** pour la qualité de l'enseignement reçu, qui m'a permis de vivre cette belle expérience professionnelle chez HABITALYS.

Table des matières

INTRODUCTION.....	3
PARTIE 1 : HABITALYS, UN ACTEUR MAJEUR DE L'HABITAT SOCIAL EN LOT-ET-GARONNE.....	5
A) DES ACTIVITES CENTREES SUR L'HABITAT ET LE DEVELOPPEMENT TERRITORIAL	5
B) UNE EQUIPE AVEC DES ROLES VARIES	6
C) LE TRAVAIL AU SEIN DE L'EQUIPE	7
PARTIE 2 : LES MISSIONS QUI M'ONT ETE CONFIEES.....	8
A) AUTOMATISTION DE L'EXPORTATION DE FICHIERS.....	8
1) Le besoin identifié	8
2) Des outils logiciels pour des rôles variés.....	8
B) AUTOMATISATION DE L'ENVOI DE LA FICHE NAVETTE	10
1) Le besoin identifié	10
2) Des méthodes variés	10
C) CREATION D'UNE RAG AVEC LLM	11
1) Le besoin identifié	11
2) Méthodologie et outils utilisés.....	11
PARTIE 3 : ANALYSE ET METHODOLOGIE.....	12
A) AUTOMATISATION DES EXPORTS DE FICHIERS	12
1) Début de la mission : exportation de fichiers CSV.....	12
2) Récupération des PDF des interventions	13
3) Construction d'un fichier XML.....	14
4) Création d'un fichier CSV pour les PDF des diagnostics manquants	16
5) Problèmes et erreurs rencontrées.....	17
B) AUTOMATISATION DE LA FICHE NAVETTE	19
1) Les débuts de tests avec MCP	19
2) Le test avec JDBC	20
3) Le driver ODBC	20
4) Erreurs rencontrées	21
C) CREATION DE LA RAG.....	22
1) Démarche adoptée	22

2) Choix effectués.....	23
3) Erreurs rencontrées	24
PARTIE 4 : RESULTATS OBTENUS ET EVALUATION	25
A) UNE EXPORTATION COMPLETE DES DIAGNOSTICS	25
B) LA RECUPERATION DES MONTANTS.....	27
C) UN AGENT IA OPTIMISE	27
PARTIE 5 : LA FIN D'UN STAGE, LE DEBUT D'UNE AVENTURE	28
A) UNE ENTREPRISE ENGAGEE ET UN ENVIRONNEMENT DE TRAVAIL FAVORABLE	28
B) UNE EXPERIENCE DETERMINANTE ET UNE VISION CLAIRE DE L'AVENIR	28
CONCLUSION.....	30
GLOSSAIRE	31
WEBOGRAPHIE.....	32

INTRODUCTION

"L'automatisation appliquée à un processus efficace le rend plus efficace. L'automatisation appliquée à un processus inefficace le rend encore plus inefficace." *Bill GATES*
Ce premier pas dans le monde professionnel m'a immergé dans un univers où code et intelligence artificielle transforment les idées en solutions concrètes.

Dans le cadre de ma deuxième année de BUT Informatique, parcours « Déploiement d'applications communicantes et sécurisées » à l'IUT Paul Sabatier de Toulouse, il m'était demandé d'effectuer un stage en entreprise d'une durée de dix semaines afin de valider mon année.

Ce que je croyais être une formalité s'est avéré bien plus exigeant. Trouver une entreprise prête à accueillir un étudiant de deuxième année en informatique n'a pas été chose aisée. Les réponses se faisaient rares, les offres rapidement pourvues, et les refus fréquents, souvent liés à la période demandée ou au fait que les structures ne prenaient pas de stagiaires dans ce domaine.

Ma recherche a débuté par la rédaction de mon CV . Sans réelle expérience professionnelle en dehors de mes projets universitaires, j'ai dû le retravailler plusieurs fois pour qu'il suscite l'intérêt des recruteurs. J'ai également publié mon profil sur LinkedIn, espérant être contacté, mais les semaines passaient sans réponse concrète. Malgré quelques entretiens obtenus, aucun n'a abouti, les postes étant pourvus avant même que je ne sois convoqué.

Face à cette impasse, ma mère m'a suggéré d'élargir mes recherches à notre région de résidence familiale. Elle s'est impliquée activement dans mes démarches, m'a aidé à cibler des entreprises, et s'est même rendue physiquement dans plusieurs d'entre elles pour déposer mon CV et mes lettres de motivation. N'ayant pas la possibilité de le faire moi-même en raison de la distance et de mes cours, son aide a été précieuse.

Parallèlement, une entreprise nommée **HABITALYS** m'a indiqué avoir transmis ma candidature à son service informatique. Voyant que le délai approchait, ma mère a alors pris l'initiative de solliciter un rendez-vous avec le maire de notre ville, **Monsieur Francis GARCIA**, afin de lui exposer ma situation. Sensible à ma démarche, il a accepté de m'apporter son soutien et a contribué à faciliter mon acceptation chez HABITALYS, organisme qui est également le bailleur social de notre logement familial et reconnu dans le département de Lot-et-Garonne en tant que premier Office Public de l'Habitat et acteur majeur dans la gestion et la construction de logements sociaux. Dans le cadre de mon stage, j'ai intégré le service informatique.

Grâce à ces efforts conjoints et au soutien indéfectible de ma mère, j'ai enfin obtenu une réponse positive. Un soulagement immense, car je commençais à perdre espoir et craignais de devoir redoubler mon année faute de stage. Si j'étais soulagé d'avoir trouvé une entreprise, j'étais également plein d'incertitudes : serais-je à la hauteur ? Comment allait se passer l'intégration dans un environnement professionnel encore inconnu pour moi ?

J'ai débuté mon stage le lundi 7 avril 2025 chez HABITALYS, pour une durée de dix semaines, jusqu'au vendredi 13 juin 2025.

L'objectif principal de ce stage était d'automatiser les flux de travail assistés par l'intelligence artificielle, en utilisant des outils tels que N8N, MCP et Mistral. Les missions qui m'ont été confiées visaient à résoudre des problèmes liés à la gestion manuelle et répétitive des données, en automatisant la récupération d'informations, la génération de fichiers, et l'importation de données financières.

L'intitulé exacte du sujet de mon stage était le suivant : « **Analyse et automatisation de flux de travail assistée par IA à l'aide de N8N, MCP, Mistral AI** », avec des technologies telles que **JavaScript, TypeScript, Python, SQL, JSON et les API REST**.

Les missions proposées présentaient un intérêt particulier pour moi, car elles me permettaient de travailler sur des projets concrets ayant un impact direct sur l'efficacité opérationnelle de l'entreprise

Après avoir complété les formalités administratives (convention, définition du sujet, objectifs de mission), j'ai pu entamer cette expérience professionnelle avec curiosité et détermination. Cette première immersion dans le monde professionnel a largement dépassé mes attentes : humaine, formatrice, rassurante, et surtout profondément enrichissante.

Ce rapport présentera mes travaux en commençant par une présentation détaillée de mon lieu de stage au sein d'HABITALYS et du service dans lequel j'ai évolué. Je présenterai ensuite les missions qui m'ont été confiées, les travaux réalisés, ainsi que la méthodologie employée. Enfin, nous terminerons par une synthèse globale afin de dresser un aperçu complet de cette expérience professionnelle.

Mon objectif est de montrer en quoi ce stage a non seulement répondu aux exigences de ma formation, mais a aussi contribué à renforcer ma motivation pour la suite de mon parcours.

PARTIE 1 : HABITALYS, UN ACTEUR MAJEUR DE L'HABITAT SOCIAL EN LOT-ET-GARONNE

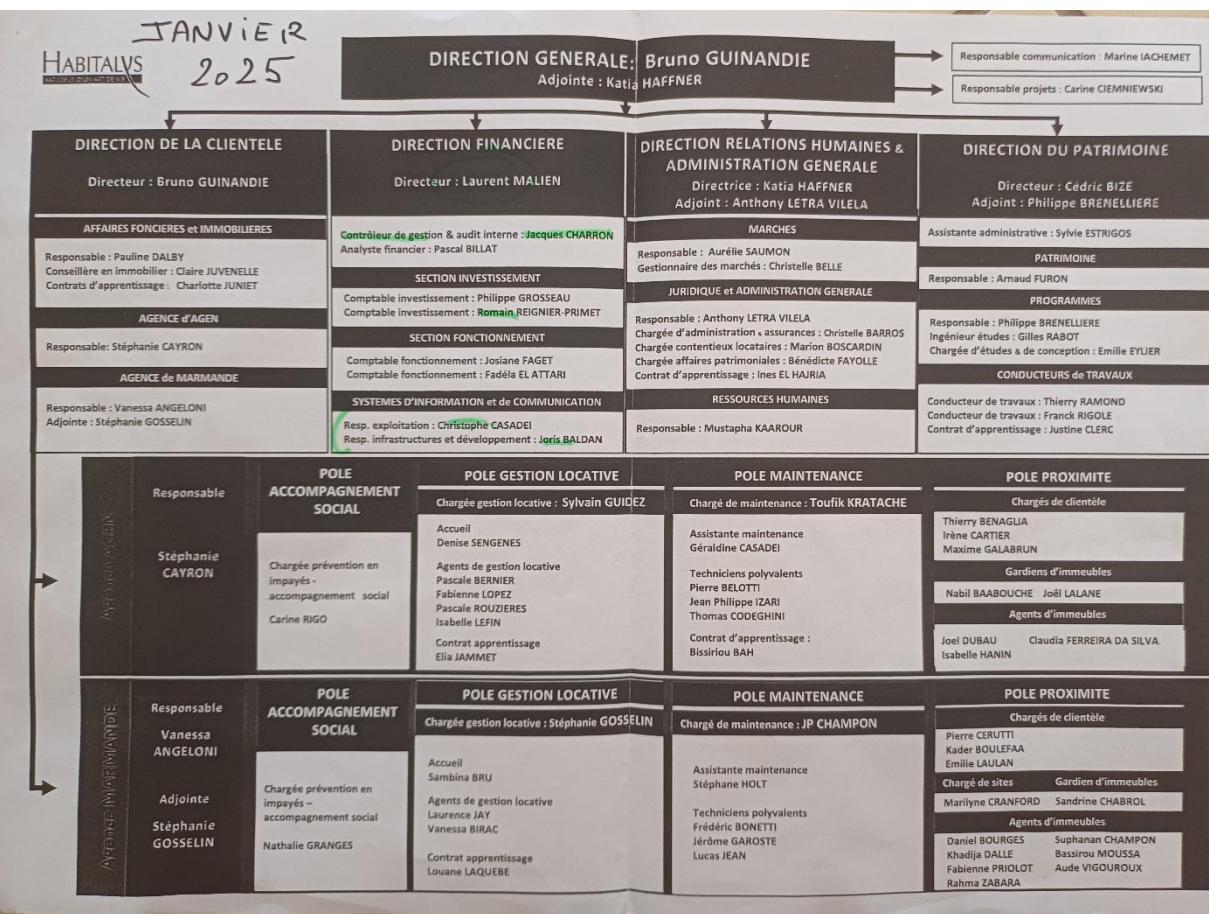
A) DES ACTIVITES CENTREES SUR L'HABITAT ET LE DEVELOPPEMENT TERRITORIAL

HABITALYS, avec un chiffre d'affaires de 42 millions d'euros en 2024, est un Office Public de l'Habitat qui joue un rôle clé dans le département de Lot-et-Garonne. En tant que premier bailleur social du département, HABITALYS gère plus de 4000 logements répartis sur 105 communes, illustrant son ancrage local et son engagement envers la communauté. L'entreprise propose une gamme variée de services incluant la location et la vente de logements, allant des terrains à bâtir aux villas, en passant par divers types d'appartements.

HABITALYS s'engage à améliorer la qualité de vie et le bien-être de ses clients, tout en contribuant à l'aménagement des communes et au développement économique et social de la région. L'entreprise emploie environ 80 personnes, qui œuvrent au quotidien pour offrir des logements de qualité et des services adaptés aux besoins des habitants. Les valeurs de proximité, de qualité de service et d'accompagnement sont au cœur des activités d'HABITALYS, qui incluent la gestion locative, l'accession à la propriété et la maîtrise d'ouvrage.

Afin de mieux comprendre l'organisation interne et la répartition des responsabilités au sein d'HABITALYS, un organigramme général de l'entreprise est présenté ci-après.

JANVIER
HABITALYS
2025



Je me trouve au sein du service "Système d'Information et de Communication" de la "Direction Financière", sous la supervision de Joris BALDAN, responsable infrastructure et développement, et Christophe CASADEI, responsable exploitation.

B) UNE EQUIPE AVEC DES ROLES VARIES

Le service "Système d'Information et de Communication" de la "Direction Financière" est chargé de créer et maintenir un système d'information robuste, ainsi que de gérer tout ce qui concerne les infrastructures informatiques, y compris les ordinateurs et les réseaux.

À cette fin, l'équipe se doit d'être transverse et de rassembler différents savoir-faire. Nous avons d'abord Joris BALDAN, le responsable d'infrastructure qui est chargé de créer, maintenir et sécuriser l'/les infrastructures, gérer le(s) réseau, et assurer son/leur bon fonctionnement(s). Nous avons ensuite Christophe CASADEI, le responsable d'exploitation qui doit s'assurer que les logiciels sont exploités de manière optimale. L'équipe et même toute l'entreprise d'HABITALYS travaille en collaboration avec le groupe ADX fournissant une plateforme informatique permettant de récupérer les détails des diagnostics de chaque intervention.

Elle travaille aussi avec Sopra Steria qui a développé une application qu'utilise l'entreprise pour stocker des informations sur les interventions.

En plus de l'équipe du service informatique. J'ai également collaboré avec Jacques CHARRON, Contrôleur de gestion et audit interne de la Direction Financière, qui m'a confié plusieurs missions.

C) LE TRAVAIL AU SEIN DE L'EQUIPE

Je me trouve sur le site d'Agen, au sein de l'agence d'HABITALYS. Mon environnement de travail a été préparé à l'avance pour faciliter mon intégration et mon efficacité. À mon arrivée, un bureau m'a été attribué et j'ai été équipé d'un ordinateur avec une clé YubiKey pour sécuriser l'accès à mes outils de travail (Annexes 1.1 et 1.2). Un identifiant et un mot de passe m'ont été fournis pour me connecter aux systèmes internes.

Pour faciliter mon travail, un compte Microsoft m'a été créé, me donnant accès à divers logiciels essentiels tels que Word et Teams, ainsi qu'à un disque pour sauvegarder et organiser mon travail. Cela m'a permis de gérer mes documents et de collaborer efficacement avec les autres membres de l'organisation.

En ce qui concerne les horaires de travail, j'avais une certaine flexibilité. Initialement, j'avais choisi de travailler :

- Le matin : 8h30 à 12h30
- L'après-midi : 14h00 à 17h00

Cependant en raison des contraintes de transport, j'ai ajusté mes horaires à partir de la troisième semaine pour travailler :

- Le matin : 9h00 à 12h30
- L'après-midi : 13h30 à 17h00

Cette flexibilité m'a permis de mieux organiser mon temps et de m'adapter aux exigences de mon stage.

Il y avait malgré tout une contrainte à respecter :

- Un volume horaire de 35h par semaine

PARTIE 2 : LES MISSIONS QUI M'ONT ETE CONFIEES

A) AUTOMATISTION DE L'EXPORTATION DE FICHIERS

1) Le besoin identifié

Les locataires d'HABITALYS doivent avoir à disposition les détails des diagnostics qui ont été effectués. HABITALYS doit donc leurs fournir des documents contenant les détails des interventions avec les diagnostics qui ont été effectués. Pour se faire, les employés d'HABITALYS doivent régulièrement télécharger et exporter des diagnostics depuis le site <https://calypso.hyperion-dev.com/extranet> (Annexe 2.1), puis transférer dans un des logiciels de l'entreprise les documents et enfin y rentrer les données adéquates (société, agence, immeuble, etc...).

Le processus actuel nécessite une intervention manuelle à chaque étape : connexion au site, téléchargement des fichiers, et enfin transfert des fichiers sur l'application de l'entreprise en entrant en plus les données. Cette méthode est non seulement chronophage, mais elle est également sujette à des erreurs humaines, telles que des oubli ou des erreurs de manipulation.

L'objectif de cette mission était donc de concevoir une solution permettant d'automatiser ce processus, réduisant ainsi la charge de travail des employés et minimisant les risques d'erreurs. L'automatisation devait permettre de télécharger automatiquement les fichiers depuis le site, de les convertir au format requis, et de les transférer sur l'application de l'entreprise sans intervention manuelle.

Pour réaliser cette mission, j'ai commencé par analyser le processus manuel existant pour identifier les étapes clés à automatiser. Ensuite, j'ai développé un script capable de se connecter au site, de télécharger les fichiers nécessaires, et de les convertir au format approprié. Enfin, j'ai mis en place un système de transfert automatique de fichiers.

Pour pouvoir réaliser cette mission, j'ai eu accès à un ensemble d'outils.

2) Des outils logiciels pour des rôles variés

Pour mener à bien mes missions chez HABITALYS, j'ai utilisé divers logiciels qui ont joué un rôle crucial dans l'automatisation, la gestion des données, le développement et la communication :

- Calypso : Calypso est la plateforme utilisée par l'entreprise pour récupérer les interventions avec les diagnostics sous forme de PDF et en globalité, les interventions sous forme de CSV. J'ai dû utiliser Calypso pour récupérer les données des interventions pour les exploiter.
- Tegia : Logiciel d'entreprise utilisé par Calypso développé par Sopra Steria. Tegia permet de stocker les informations sur les différentes interventions de la société. J'ai dû me servir de Tegia pour comprendre comment intégrer les fichiers, les intégrer et voir quels étaient les différents types de diagnostics et documents, comprendre les informations qu'il fallait rentrer et faire des tests. Evidemment, j'ai utilisé une version de test du logiciel pour ne pas risquer de faire une mauvaise manipulation sur la version PROD. (Annexe 2.2)
- N8N : N8N est un outil d'automatisation de workflows qui permet de connecter différentes applications et services pour automatiser des tâches répétitives. Dans le cadre de mon stage, N8N a été installé sur une machine virtuelle Linux et configuré pour fonctionner en tant que service. Cela m'a permis de créer des flux de travail automatisés pour manipuler des données et interagir avec d'autres services sans intervention manuelle constante. N8N a été le logiciel que j'ai principalement utilisé car le but principal de cette mission était d'automatiser des tâches récurrentes. (Annexe 2.5)
- DBeaver : DBeaver est un outil de gestion de bases de données qui permet de se connecter à diverses bases de données pour exécuter des requêtes SQL, visualiser et manipuler les données. J'ai utilisé DBeaver pour accéder à des bases de données spécifiques et extraire les informations nécessaires pour mes projets d'automatisation. (Annexe 2.3)
- Visual Studio Code (VS Code) : VS Code est un éditeur de code source développé par Microsoft. Il offre des fonctionnalités avancées telles que la coloration syntaxique, le débogage intégré, la complétion intelligente de code (IntelliSense), et des extensions pour supporter divers langages de programmation et outils. J'ai utilisé VS Code pour écrire et tester le code nécessaire à l'automatisation des tâches, ce qui a grandement facilité le processus de développement.
- Microsoft Teams : Microsoft Teams est une plateforme de communication et de collaboration qui permet le chat, les appels vidéo, le partage de fichiers et l'intégration avec d'autres outils Microsoft. Bien que peu utilisé, Teams a été essentiel pour des sessions de partage d'écran et de contrôle à distance avec mon tuteur, facilitant ainsi la collaboration et la résolution de problèmes techniques.
- Outlook : Outlook est un client de messagerie qui fait partie de la suite Microsoft Office. Il permet de gérer les e-mails, les calendriers, les contacts et les tâches. J'ai utilisé Outlook pour recevoir des instructions et des informations de la part de mon tuteur et de mes collègues, ce qui a facilité la communication et le suivi des tâches.
- Google Gmail avec son API : Google Gmail API permet aux développeurs d'intégrer des fonctionnalités de messagerie dans leurs applications. J'ai configuré cette API pour automatiser l'envoi et la réception d'emails dans le cadre de mes projets, ce qui a permis de faire pas mal de tests. (Annexe 2.4)

Ces logiciels ont été essentiels pour créer un environnement de travail efficace et pour mener à bien les missions qui m'ont été confiées.

B) AUTOMATISATION DE L'ENVOI DE LA FICHE NAVETTE

1) Le besoin identifié

Dans les habitations appartenant à la société HABITALYS, des travaux de maintenance et de réparation sont régulièrement nécessaires, engendrant des coûts importants. Afin de gérer ces dépenses de manière efficace, les employés doivent récupérer les montants des travaux depuis un logiciel dédié à l'établissement des états des lieux, qui génère automatiquement les coûts des travaux. Ces montants doivent ensuite être transférés dans un autre logiciel interne, utilisé pour la facturation des locataires.

Cette opération, bien qu'essentielle, est particulièrement répétitive, chronophage et peu stimulante. Elle mobilise du temps pour une tâche à faible valeur ajoutée, tout en comportant un risque d'erreur humaine lors de la saisie manuelle. Les équipes étaient confrontées à ce processus de manière régulière, ce qui rendait l'automatisation d'autant plus pertinente.

Ma mission consistait donc à d'automatiser cette récupération des montants Abyl, puis de les intégrer directement dans le logiciel de facturation de l'entreprise. L'objectif était de fiabiliser le processus tout en réduisant le temps de traitement, pour permettre aux employés de se concentrer sur des tâches plus utiles et moins répétitives.

2) Des méthodes variés

Pour automatiser le transfert des montants des travaux depuis Abyl vers le logiciel de facturation interne, plusieurs outils et technologies ont été mobilisés :

- MCP : Le Model Context Protocol est un protocole standard ouvert conçu pour connecter des modèles d'intelligence artificielle (IA) à des outils, services et sources de données externes.
- Le driver JDBC : (Java Database Connectivity) est un logiciel qui permet à une application Java de se connecter à une base de données. Il sert d'intermédiaire entre l'application et le système de gestion de base de données (SGBD), facilitant l'exécution de requêtes SQL et la récupération des résultats.
- Le driver ODBC : (Open Database Connectivity) est un pilote qui permet à une application de communiquer avec une base de données de manière standardisée, quelle que soit la base utilisée. Il agit comme une interface intermédiaire entre l'application et le système de gestion de base de données (SGBD).

De plus, j'ai aussi utilisé VS Code pour développer et faire certaines configurations et aussi N8N pour tester le MCP.

C) CREATION D'UNE RAG AVEC LLM

1) Le besoin identifié

Dans le cadre de l'amélioration des processus internes et de l'exploitation des données, HABITALYS avait besoin d'une solution pour faciliter la recherche et l'accès à l'information contenue dans ses documents. L'objectif était de créer un système de Recherche Augmentée par Génération (RAG) utilisant des modèles de langage (LLM) et des bases de données vectorielles pour permettre une recherche plus efficace et pertinente.

2) Méthodologie et outils utilisés

Pour répondre à ce besoin, il a fallu utiliser différentes méthodes pour transformer les documents et pour stocker les données. Il y avait plusieurs points et outils à comprendre et à utiliser :

- **LLM** : Un grand modèle de langage (LLM) est un modèle de langage possédant un grand nombre de paramètres (généralement plus d'un milliard). Ce sont des réseaux de neurones profonds entraînés sur de grandes quantités de texte non étiqueté utilisant l'apprentissage auto-supervisé ou l'apprentissage semi-supervisé.
- **RAG** : Le Retrieval Augmented Generation (RAG) est une approche novatrice qui combine le meilleur de deux mondes en IA : la recherche d'informations (retrieval, qui ne génère pas de réponse originale) et la génération de contenu (qui ne s'appuie que sur les données de son entraînement).
- **Embedding** : Représentation vectorielle d'un élément (mot, image, utilisateur, produit, etc.). En d'autres termes, c'est un ensemble de nombres qui résume ce que représente cet élément.
- **Base de données vectorielle** : Permettent de stocker et de récupérer des vecteurs sous forme de points à forte dimensionnalité. Elles ajoutent des fonctionnalités supplémentaires pour une recherche efficace et rapide des voisins les plus proches dans l'espace à N-dimensions.

PARTIE 3 : ANALYSE ET METHODOLOGIE

A) AUTOMATISATION DES EXPORTS DE FICHIERS

1) Début de la mission : exportation de fichiers CSV

La mission qui m'a été confiée consistait à automatiser un processus auparavant réalisé manuellement par les employés : le téléchargement et l'exportation de fichiers PDF et CSV depuis un site web spécifique, suivis de leur transfert vers une autre plateforme. Cette tâche répétitive présentait un risque d'erreurs humaines et mobilisait un temps considérable, d'où l'intérêt d'en assurer l'automatisation.

Dans un premier temps, je me devais de bien comprendre les attentes fonctionnelles du processus, ainsi que les contraintes techniques imposées par les outils en place. Il était essentiel de déterminer les points d'entrée et de sortie du flux de données, les formats utilisés (PDF, CSV, JSON), ainsi que les fréquences d'exécution souhaitées.

L'outil N8N a été identifié comme la solution principale pour répondre à ces besoins. Plateforme d'automatisation low-code, N8N permet d'enchaîner des actions sous forme de workflows en intégrant des nœuds dédiés à des tâches spécifiques (requêtes HTTP, traitement de données, déclencheurs, etc.). Une prise en main de cet outil était donc indispensable pour mettre en œuvre la solution.

Cela impliquait également de :

- Étudier la structure du site source pour comprendre comment s'y connecter automatiquement, notamment en utilisant des requêtes HTTP avec des données d'authentification (dans ce cas, les identifiants de Jacques CHARRON)
- Analyser le format des données à extraire, en particulier le format JSON renvoyé par les requêtes du site, afin d'anticiper leur transformation
- Choisir les nœuds appropriés dans N8N pour traiter, convertir et stocker les données (ex. : « HTTP Request », « Code », « Convert to File », « Write Binary File »)
- Définir un répertoire cible pour l'export des fichiers générés, à savoir un dossier dédié sur un disque monté (n8n\$ (\nas01) (Y:))
- Structurer le workflow de manière logique, en définissant le bon déclencheur, les enchaînements d'étapes et les conditions de vérification éventuelles

Pour se faire, j'ai d'abord regardé des vidéos pour comprendre comment N8N fonctionnait et comment l'installer. Puis avec l'aide de mon tuteur, on a créé une machine virtuelle Linux (VM)

en y mettant Ubuntu puis on a installé N8N. On s'est connecté à un dossier partagé en utilisant CIFS pour pouvoir l'utiliser dans mon travail.

Une fois N8N installé et le dossier partagé connecté, j'ai commencé à créer des petits workflows pour me familiariser avec l'outil. Ces exercices m'ont permis de mieux comprendre comment configurer les nœuds et comment créer des flux de travail automatisés. Cela a été une étape cruciale pour me préparer à travailler sur ma mission principale.

Je me suis attaqué tout d'abord dans l'envoie du CSV. Pour cette mission, j'ai été assez rapide pour comprendre comment automatiser l'export en récupérant les données JSON du site puis en le mettant dans le dossier partagé. Le seul problème a été que je n'avais pas réussi à mettre en forme le CSV car il ne ressortait pas dans le bon format dans mon workflow. J'avais tenté avec un nœud Code mais sans succès et j'ai très vite laissé tomber car il ne m'avait finalement servi à rien pour la suite.

2) Récupération des PDF des interventions

Après avoir travaillé sur les CSV, je me suis ensuite attaqué à la récupération des fichiers PDF des interventions. Cette tâche était cruciale car elle permettait de centraliser et d'archiver les rapports d'intervention pour un accès facile et organisé.

À cette fin, je devais adapter le workflow existant pour inclure le téléchargement de fichiers PDF. Je me devais donc de bien saisir les points d'entrée et de sortie du flux de données, les formats utilisés (PDF, CSV, JSON), ainsi que les fréquences d'exécution souhaitées. Cela impliquait de comprendre les sources de données et les destinations finales.

Je me devais d'identifier les URLs spécifiques des PDF sur le site web pour adapter le workflow existant afin d'inclure le téléchargement de fichiers PDF. Cette étape était cruciale car elle déterminait comment les fichiers PDF pouvaient être accessibles et téléchargés automatiquement. Pour se faire, je devais donc analyser la structure du site web pour comprendre comment les fichiers PDF étaient organisés et accessibles. Je devais naviguer manuellement sur le site pour identifier les sections où les fichiers PDF étaient stockés, ce qui me permettait de comprendre la hiérarchie des dossiers et des sous-dossiers où les PDF étaient situés.

Les outils de développement du navigateur ont été essentiels pour cette tâche. Je me devais d'utiliser l'outil d'inspection pour voir les éléments HTML qui contenaient les liens vers les fichiers PDF. Cela m'a permis de comprendre comment les liens étaient générés et quels attributs étaient utilisés. Je devais également utiliser l'onglet réseau des outils de développement pour voir les requêtes HTTP qui étaient faites lorsque j'accédais aux fichiers PDF. Cela me permettait de voir les en-têtes et les paramètres des requêtes, ce qui était crucial pour configurer correctement les requêtes dans N8N.

Une fois les URLs identifiées, je me devais de configurer mon workflow dans N8N pour télécharger automatiquement les fichiers PDF. Cette configuration devait impliquer plusieurs étapes clés tels que des nœuds HTTP dans N8N pour configurer les requêtes de téléchargement des fichiers PDF. Je me devais de configurer les paramètres de la requête HTTP pour inclure les URLs des fichiers PDF et les en-têtes nécessaires pour accéder aux fichiers. Cela incluait les informations d'authentification et les paramètres spécifiques pour accéder aux fichiers.

Enfin, pour automatiser le processus de téléchargement, je me devais de configurer le workflow pour qu'il s'exécute à des intervalles spécifiques. J'ai donc dû configurer des déclencheurs pour lancer le workflow à des intervalles spécifiques.

Après avoir réalisé cette étape, je me devais de réaliser les XML.

3) Construction d'un fichier XML

La troisième partie de ma mission consistait à construire un fichier XML à partir des données récupérées. L'objectif était de structurer les informations de manière standardisée pour faciliter leur intégration et leur traitement dans le logiciel Tegia.

À cette fin, j'ai dû d'abord comprendre la structure XML requise. Pour se faire, j'ai discuté avec Jacques CHARRON et Christophe CASADEI des informations essentiels à mettre dans le XML et je me suis servi d'un PDF qui expliquait comment réaliser les balises, quoi mettre dedans et quels étaient les informations facultatives à mettre dans le fichier.

Il y avait différents types de balises conformément aux différents types de diagnostics à savoir les balises :

- Diagnostic Liste : Balise « racine ».
- Diagnostic : Elle contient toutes les informations pour un diagnostic technique. Nombre : [1...N].
- DTA : Elle contient les informations propres au DTA.
Cette balise est facultative et ne peut être présente que si l'attribut CategorieDiagnostic a pour valeur 'AMIANTE'. Elle ne peut pas être présente en même temps que les balises DAPP et Plomb. Nombre : [0..1].
- DAPP : Elle contient les informations propres au DAPP.
Cette balise est facultative et ne peut être présente que si l'attribut CategorieDiagnostic a pour valeur 'AMIANTE'. Elle ne peut pas être présente en même temps que les balises DTA et Plomb. Nombre : [0..1].
- Plomb : Elle contient les informations propres au diagnostic Plomb.

Cette balise est facultative et ne peut être présente que si l'attribut CategorieDiagnostic a pour valeur 'PLOMB'. Elle ne peut pas être présente en même temps que les balises DTA et DAPP. Nombre : [0..1].

- Action du diagnostic : Elle contient la liste des actions associées au diagnostic.

Cette balise est facultative. Nombre : [0..N].

- Electricité : Elle contient les informations propres au diagnostic Electricité.

Cette balise est facultative et ne peut être présente que si l'attribut CategorieDiagnostic a pour valeur 'ELECTRICITE'. Nombre : [0..1].

- Gaz : Elle contient les informations propres au diagnostic Gaz.

Cette balise est facultative et ne peut être présente que si l'attribut CategorieDiagnostic a pour valeur 'GAZ'. Nombre : [0..1].

- Constatation : Elle contient la liste des constatations/anomalies associées au diagnostic.

Cette balise est facultative, ne peut être présente que si la balise Electricite ou Gaz est présente. L'utilisation de cette balise sous-entend que la codification utilisée dans l'applicatif est connue par le prestataire qui génère le fichier .xml. Nombre : [0..N].

- RAAT : Elle contient la liste des RAAT associés au DTA ou DAPP selon le contexte.

Cette balise est facultative. Cette balise est facultative et ne peut être présente que si l'attribut CategorieDiagnostic a pour valeur 'AMIANTE'. Nombre : [0..N]. Le patrimoine transmis (de l'attribut Societe à l'attribut Piece) doit au minimum être de même niveau que le DTA ou DAPP auquel le RAAT est rattaché.

- Action du RAAT : Description : Elle contient la ou les actions associées à chacun des RAAT.

J'ai également reçu différents exemples d'XML pour m'aider à le réaliser comme :

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<DiagnosticListe xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="ImportDiagnosticTechnique_v6b.xsd" NbDiagnostic="2" NbElectricite="1" NbGaz="1" NbPlomb="0"
NbDTA="0" NbDAPP="0" NbRAAT="0">
  ▼<Diagnostic TypeDiagnostic="GAZ" CategorieDiagnostic="GAZ" Reference="13246" Realisation="2023-11-29" Societe="01" Agence="01"
  Groupe="32" Batiment="0001" Immeuble="1" Local="1168" Ctiers="0170420" Description="GAZ 1324 01 R0 4 4 5 1002" Document="GAZ-
  1324-02.pdf" Relance="2033-12-01">
    <Gaz InstallationGaz="true" PresenceAnoGaz="true" AnoTypeDGI="true"/>
    <ActionDiagnostic DateAction="2023-11-29" CodeAction="REV"/>
    <ActionDiagnostic DateAction="2023-11-29" CodeAction="RET"/>
    <Constatation CodeConstatation="RAC"/>
    <Constatation CodeConstatation="CATEG"/>
  </Diagnostic>
  ▼<Diagnostic TypeDiagnostic="ELEC" CategorieDiagnostic="ELECTRICITE" Reference="4567" Realisation="2018-12-01" Societe="01"
  Agence="01" Groupe="32" Batiment="0001" Immeuble="1" Local="1168" Ctiers="0170420" Relance="2019-12-05" Description="ELEC 4567
  01 R0 3 3 1 1002" Document="ELEC-4567-02.pdf">
    <Electricite InstallationElec="true" PresenceAnoElec="false"/>
    <Constatation CodeConstatation="B11"/>
    <Constatation CodeConstatation="B4"/>
    <ActionDiagnostic DateAction="2018-12-01" CodeAction="REV"/>
    <ActionDiagnostic DateAction="2018-12-01" CodeAction="RET"/>
  </Diagnostic>
</DiagnosticListe>
```

J'ai donc dû étudier les balises et les attributs nécessaires pour représenter correctement les données. Ensuite, j'ai dû utiliser N8N pour transformer les données téléchargées en un format XML conforme aux spécifications.

4) Création d'un fichier CSV pour les PDF des diagnostics manquants

Dans le cadre de ma mission, une tâche supplémentaire s'est ajoutée : la création d'un fichier CSV pour informer les donneurs d'ordre (DO) des diagnostics manquants pour certaines interventions. Cette tâche n'était pas prévue initialement, mais elle s'est avérée cruciale pour améliorer la gestion des diagnostics sur la plateforme Calypso.

Il a été identifié que certaines interventions sur la plateforme Calypso avaient des diagnostics sans les fichiers PDF qui allaient avec, ce qui posait un problème pour le suivi et la gestion des diagnostics. Par exemple dans une intervention avec un diagnostic de type RAAT :

The screenshot shows a user interface for managing interventions. On the left, a sidebar lists several interventions with their IDs and a small green circle indicating a 'Rapport final reçu' (final report received). In the center, a search interface titled 'Documents de l'intervention MA2409157772' displays a table with columns for 'Nom' (Name) and 'Catégorie' (Category). At the bottom, there are filters for 'Items per page' (10), a page number (0 of 0), and navigation arrows. On the right, a detailed view of the intervention shows a list of documents with their names, sizes (47), and download icons. The documents listed are: ANA-ME DAPP DPE-L ELEC-S ..., RAAT-S RPAVT, ANA-ME PRIVMT RAAT-R RPA..., RAAT-R RPAVT, RAAT-S RPAVT, and RAAT-S RPAVT.

Il n'y avait rien.

L'objectif était donc de créer un fichier CSV qui permettrait d'informer les donneurs d'ordre des diagnostics manquants. Ce fichier devait contenir les informations suivantes :

- L'identifiant de l'intervention (ID)
- La date de l'intervention
- Le type de diagnostic (par exemple, ELEC pour électrique ou GAZ pour gaz)
- Un lien permettant de vérifier le PDF dans la GED (Gestion Électronique de Documents) de Tegia
- Le nom du PDF
- Une indication sur la présence ou l'absence du PDF
- Le donneur d'ordre (DO)

Pour répondre à ce besoin, j'ai dû rajouter dans mon Workflow réalisant les PDF et XML la création du fichier CSV avec les PDF manquants. Cette fonctionnalité de mon Workflow devait être réalisée en plusieurs étapes :

- Récupération des données : J'ai utilisé les nœuds déjà existant pour me connecter à la plateforme Calypso et extraire les informations nécessaires sur les interventions. Cela incluait les identifiants des interventions, les dates, les types de diagnostics et les donneurs d'ordre. J'ai utilisé des requêtes HTTP pour accéder aux données et les récupérer au format JSON.
- Création du fichier CSV : J'ai configuré un nœud pour formater les données récupérées et les écrire dans un fichier CSV. Chaque ligne du fichier CSV correspondait à une intervention et contenait les informations requises : identifiant de l'intervention, date, type de

- diagnostic, lien vers le PDF, nom du PDF, indication sur la présence ou l'absence du PDF, et donneur d'ordre.
- Vérification des PDF : J'ai utilisé des requêtes pour accéder à la GED de Tegia et vérifier la présence des fichiers PDF pour chaque intervention. J'ai récupéré les liens vers les PDF existants pour les inclure dans le fichier CSV.
 - Comparaison des fichiers CSV : Je me devais de comparer le fichier CSV créé aujourd'hui par rapport à celui de la veille pour ne pas envoyer à plusieurs reprises le même CSV.
 - Automatisation du processus : Pour rendre ce processus plus efficace, j'ai automatisé la création du fichier CSV en configurant le workflow pour qu'il s'exécute régulièrement soit une fois par jour. Cela permettait de s'assurer que les donneurs d'ordre avaient toujours accès à des informations à jour sur les diagnostics manquants.

Une fois cette mission terminée, j'ai pu commencer la seconde mission, l'import des montant Abyl.

5) Problèmes et erreurs rencontrées

Au début de la mission, je devais me renseigner sur N8N, ensuite j'ai dû chercher comment l'installer. Après cela avec l'aide de Joris, on a tenté de l'installer sur une VM qu'il avait créé puis de créer un dossier partagé pour que je puisse intégrer les documents de mes travaux dedans. Pour se faire, après avoir installé N8N, on a tenté de créer un point de montage en NFS (Protocole qui permet à un ordinateur d'accéder via un réseau à des fichiers distants.). Nous avons d'abord monté sur la VM le disque « \\nas01 » avec le protocole NFS mais on a eu des problèmes à cause de droits sur le dossier partagé. Nous avons tenté en changeant le fichier de configuration à plusieurs reprises. Même après plusieurs modifications. On n'a pas réussi. On s'est alors rabattu sur le CIFS (Protocole qui permet de partager des fichiers et des imprimantes via Internet ou l'Intranet d'une entreprise.) qui lui a fonctionné.

Après ces péripéties, j'ai fait de petits tests sur N8N pour mieux comprendre comment m'en servir en réalisant quelques workflows de tests (Annexe 5.1). Puis j'ai commencé ma mission.

J'ai facilement pu faire l'export de CSV car il n'était pas très compliqué mais j'ai rencontré quelques problèmes par rapport au format des données dans le CSV (les données étaient mal affichées, les colonnes avaient un mauvais nom, etc ...). J'ai essayé mais je n'ai jamais corrigé le problème (Annexe 4.1).

Ensuite, j'ai travaillé sur les PDF. Sur cette partie, j'ai rencontré des difficultés sur l'export de PDF car je n'arrivais pas à importer plus d'un PDF. Finalement, c'était juste à cause d'une mauvaise utilisation des noeuds que je n'y arrivais pas. Je suis allé jusqu'à utiliser des noeuds non installés sur ma version d'N8N en allant sur le site d'N8N et en les copiant à savoir itemList, Move Binary Data. J'ai également testé pleins de noeuds finalement inutiles à savoir Set, Extract from

File, Convert to File, ... (Annexe 4.2). Au final j'avais juste besoin de faire une boucle sur mon nœud qui écrit les fichiers sur la machine. En plus, j'ai changé à plusieurs reprises la façon de récupérer les PDF. Je les ai d'abord récupérés depuis les commentaires. Puis depuis un payload provenant d'une requête http. Le PDF m'a pris pas mal de temps à faire surtout à cause du nœud Write Files from Disk qui n'écrivait qu'un seul fichier PDF dans le pc et qui nécessitait donc une boucle pour itérer sur tous les documents.

Pour le XML, c'était très compliqué. J'ai passé beaucoup de temps sur le développement du XML avec plusieurs problèmes qui arrivaient à savoir :

- Le format du XML qui était incorrecte. J'ai dû le changer à plusieurs reprises pour arriver à un résultat à peu près cohérent.
- La liste des diagnostics que je n'arrivais pas à mettre à jour. J'ai dû réaliser une fonction pour pouvoir compter les diagnostics avec un compteur.
- Les différents paramètres de l'attribut « estateReference » dans mon JSON qui n'étaient pas conforme à ce que je devais avoir. Tout simplement car j'avais pris des interventions tardives. Il a aussi fallu refaire un peu le format des données récupérées car je devais avoir : Societe, Agence, Groupe, Batiment, Immeuble, Local.
- Je devais aussi trouver Ctiers que je n'ai pas trouvé car oui, dans le JSON récupéré depuis Calypso, il manquait pleins d'informations que je n'avais pas. Donc il y avait des données à rentrer obligatoirement que je ne pouvais pas récupérer car elles n'existaient tout simplement pas dans le JSON. J'ai ainsi dû mettre pas mal de données vides et inventer des données. Il a donc fallu demander à Christophe CASADEI quel était le Ctiers et il me l'a donné en regardant sur une base de données d'HABITALYS dans DBeaver.
- Les dates comme celle de réalisation qui était en format « Unix Time ». Il a donc été nécessaire de la convertir en format YYYY-MM-DD. Puis la date de relance qui n'existe même pas. Il a été nécessaire de donc mettre une relance en dure.
- Mettre à « typeDiagnostic » « AMIA1 » pour RAAT, DAPP, DTA. Puis mettre une section RAAT soit dans un bloc DAPP, soit dans un bloc DTA.

J'ai ensuite dû l'importer sur le logiciel Tegia et là encore, il y avait beaucoup de problèmes. Dès le début, j'ai compris que mon XML était finalement faux et en plus, le PDF qu'on m'avait donné pour faire le XML était plein d'erreurs (j'avais donc une documentation fausse). J'ai dû donc faire des dizaines et des dizaines de tests avec plein d'erreurs qui arrivaient encore et encore (Annexe 4.3). Et même après avoir fini, j'ai dû le corriger plus tard à plusieurs reprises car les données récupérées depuis le JSON du site étaient remplies d'erreurs complètement insensées.

Enfin, il y a eu le CSV contenant les PDF manquants. Là aussi encore, il était compliqué (Annexe 4.4). En effet, il m'a donné beaucoup de mal, pas seulement à cause de la difficulté de la tâche mais tout simplement car les données du JSON provenant d'ADX était complètement mal faites. J'ai donc perdu pas mal de temps, inutilement, à corriger mon travail censé être correct. Au début, je devais dresser une liste de tous les diagnostics avec : la date de l'intervention, sa référence, le nom que le PDF aurait dû avoir s'il manque le PDF, le type et si, oui ou non, il manque

le PDF. Puis j'ai dû rajouter d'autres colonnes comme l'adresse, le donneur d'ordre, etc... Au début, ce n'était pas trop compliqué, jusqu'à ce que je remarque une erreur dans les données JSON provenant de chez ADX. J'ai dû changer encore la logique de mon code dans mon nœud code car il y avait des erreurs sur les données reçus comme « ANA-ME » et « ANA-MO » au lieu de « AMIA1 » => donc « AMIA1 » n'existe pas en réalité que sur Tegia et il y avait la même erreur pour le plomb (PLOMB1). Donc j'ai perdu du temps et j'ai été en difficultés car des professionnels ne sont pas capables d'effectuer leurs travaux correctement. Il m'a également fallu changer d'autres points car il y avait d'autres types de diagnostics que je n'avais jamais vus. De plus, j'avais beaucoup de mal à comprendre comment récupérer les données depuis la base de données d'HABITALYS et récupérer le lien des PDF. Fort heureusement, j'ai reçu de Joris BALDAN l'URL que je devais utiliser pour récupérer les données des diagnostics ce qui m'a permis de gagner du temps, même si au final, j'ai été mis, pas mal de fois, en difficultés sur les données à récupérer, comment les afficher dans le CSV, ne prendre que les nouvelles interventions, et comparer les interventions récupérées par rapport à celles d'avant où pour se faire, j'ai créé un dossier où j'ai téléchargé l'ancien CSV puis je l'ai récupéré dans N8N pour pouvoir le comparer à celui créé aujourd'hui pour voir qu'il l'envoie tous les jours. Il m'a également fallu faire en sorte que le CSV soit généré que s'il y a des PDF manquants dans les nouvelles interventions.

Au final, l'aspect le plus complexe dans le CSV aura été le lien vers le PDF pour pouvoir le consulter. En effet j'ai dû faire énormément de tests pour arriver à un résultat correct.

Après avoir fini, le CSV, il fallait l'envoyer, j'ai donc dû faire l'envoi d'email. Cette partie a été la moins compliquée. Pour commencer, j'avais d'abord fait des tests avec mon propre compte Google en utilisant une API de Google Gmail. Même avec un tutoriel en ligne et des vidéos, j'ai pris un peu de temps, sachant en plus, qu'il y avait un problème dans N8N par rapport à la connexion avec des comptes pour utiliser des Credentials. Fort heureusement, Joris BALDAN l'a corrigé et j'ai réussi à envoyer un message. Ensuite, Joris BALDAN m'a créé un Credential avec Outlook et j'ai pu travailler sur l'envoi du CSV par email qui a été sans trop de difficultés.

B) AUTOMATISATION DE LA FICHE NAVETTE

1) Les débuts de tests avec MCP

Avant de m'attaquer à la récupération des montants Abyl, j'ai d'abord tenté de réaliser un agent IA connecté à un MCP pour voir si je pouvais faire en sorte qu'une IA utilise des outils pour répondre à mes questions. Cela aurait permis plus tard d'utiliser un agent IA pour récupérer les montants de n'importe quel locataire, en quelques instants.

Pour se faire, j'ai d'abord cherché à comprendre comment fonctionnait le MCP en ligne en effectuant des recherches et en regardant des vidéos. Puis j'ai tenté de réaliser un agent IA

utilisant le MCP sur N8N. Pour se faire, j'ai créé un Agent IA avec une IA de Mistral (Mistral small latest) dans un workflow puis dans un autre workflow séparé, j'ai créé un serveur MCP avec une calculatrice en outil. J'ai testé et l'IA a bien utilisé la calculatrice mais seulement lorsque le calcul devenait complexe. Pour des calculs simples (2+2), elle n'utilisait pas la calculatrice. (Annexe 5.2)

Après ces petits tests. J'ai pris un workflow utilisant le MCP en ligne et je l'ai modifié en mettant Mistral en agent IA. Puis en modifiant la configuration du MCP et en ajoutant d'autres outils pour le test tel qu'une commande curl pour avoir le code source d'une page. (Annexe 5.2)

Après avoir testé le MCP, je me suis attaqué au driver JDBC.

2) Le test avec JDBC

Après avoir testé le MCP, je me suis ensuite renseigné sur le JDBC pour pouvoir communiquer avec la base de données d'Abyl pour récupérer le montant IRL et le compte locataire. Après avoir effectué quelques recherches en ligne, je me suis connecté à une base de données d'HABITALYS appelé esopac47 en installant d'abord un driver JDBC puis en utilisant les informations de connexion pour la connexion à la BD. Pour se faire :

- J'ai récupéré et placé un fichier openedge.jar dans un dossier accessible.
- J'ai ensuite créé un fichier de connexion nommé OpenEdgeJDBC.java.
- J'ai compilé le programme avec la commande « javac -cp .:openedge.jar OpenEdgeJDBC.java ».
- Enfin, j'ai exécuté le programme avec la commande « java -cp .:openedge.jar OpenEdgeJDBC ».

Après avoir réussi ma connexion et l'avoir testé sur N8N avec un workflow utilisant le code du fichier OpenEdgeJDBC.java, je me suis attaqué à l'ODBC pour récupérer les montants.

3) Le driver ODBC

Pour récupérer les montants. J'ai essayé d'utiliser un driver ODBC Classic pour HFSQL car la base de données d'Abyl était en HFSQL. Le HFSQL étant une base de données SQL intégrée aux environnements WINDEV, WINDEV Mobile et WEBDEV qui existe en version Locale, Réseau, Cluster, Client/Serveur (Windows et Linux), Cloud et Mobile.

Pour se faire, j'ai suivi les étapes d'installation du site : <https://doc.pcsoft.fr/fr-FR/?9000160>. Je me suis d'abord préparé en installant le package « libiodbc-dev ». Puis après que mon tuteur a cherché dans sa machine le pack « wxpackodbclinux64.zip », je l'ai dézippé puis j'ai exécuté le script « install.sh » se trouvant dans le dossier dézippé. J'ai ensuite avec l'aide de Joris BALDAN configuré le driver et testé l'ODBC.

En dépit de mes efforts, je n'ai pas pu aller plus loin dans la mission. Le problème étant qu'on a dû utiliser une ancienne version de l'ODBC pour tester la connexion et que cette connexion était instable. Après en avoir terminé avec les montants, je me suis attaqué au RAG.

4) Erreurs rencontrées

Dans les différentes parties de ma mission pour la récupération du montant Abyl. J'ai rencontré beaucoup de difficultés et même un imprévu qui m'a empêché de terminer ma mission.

Au début, lorsque j'ai travaillé sur le MCP pour voir si je pouvais faire en sorte qu'un agent IA utilise des outils pour répondre à mes questions, j'ai rencontré quelques problèmes. En effet, lorsque j'ai utilisé l'IA de Mistral pour mon Agent, elle refusait d'utiliser le MCP et après quelques tests, je me suis rendu compte que c'était tout simplement à cause de la version de l'IA qui ne pouvait pas utiliser le MCP. Il a été nécessaire de utiliser la dernière version de Mistral small. Après, lors de tests plus poussés, je n'ai pas rencontré de réels problèmes.

Ensuite, lors de la connexion à la base de données esopac47, j'ai eu du mal à comprendre comment faire. J'avais d'abord commencé sur N8N avec les informations de connexion de me connecter à la base de données sans succès. J'ai ensuite tenté d'utiliser SQL Explorer pour me connecter à la base de données OpenEdge avec un driver JDBC, là encore en vain. Au final, j'ai utilisé le driver `openedge.jar` que Joris BALDAN m'a donné en le mettant dans un dossier accessible. Puis, j'ai installé java et j'ai créé un fichier de connexion `OpenEdgeJDBC.java` que j'ai codé en Java pour réaliser ma connexion.

J'ai rencontré quelques difficultés à réaliser le code et il y a eu pas mal de tests mais après quelques essais, j'ai réussi une connexion et j'ai réussi à l'intégrer dans un workflow sur N8N.

Après avoir réalisé une connexion avec JDBC, j'ai tenté de me connecter à la base de données d'Abyl en HFSQL pour récupérer les montants avec un driver ODBC. Pour se faire, j'ai suivi les étapes d'installation d'un site en installant d'abord le package « `libiodbc-dev` ». Puis j'ai dézippé le pack « `wxpackodbclinux64.zip` » et j'ai exécuté le script « `install.sh` ». Le pack « `wxpackodbclinux64.zip` » était introuvable sur le web puisqu'il n'est accessible que pour les utilisateurs de Windev (il est dans le répertoire "INSTALL\ODBC" du répertoire d'installation de WINDEV, WEBDEV ou WINDEV Mobile). J'ai ainsi dû attendre que Joris BALDAN me le donne. Il l'a cherché puis me l'a envoyé.

J'ai ensuite configuré le driver en créant à la racine du répertoire « `home` » un fichier « `.odbc.ini` » que j'ai configuré comme ceci :

```
[ODBC Data Sources]
<Nom Source> = HFSQL
[<Nom Source>]
Server Name = <Nom du serveur>
Server Port = <Port à utiliser>
Database = <Nom de la base>
UID = <Nom de l'utilisateur>
PWD = <Mot de passe de l'utilisateur>
```

Puis, j'ai tenté différentes approches de test pour l'ODBC. J'ai d'abord tenté d'utiliser l'ODBC depuis un script en python. Pour se faire, j'ai créé un environnement virtuel dans un nouveau répertoire puis je l'ai activé avec « source venv/bin/activate ». J'ai ensuite installé « pyodbc » dans cet environnement et j'ai lancé mon script avec la commande « python test_odbc.py » et il y avait plusieurs erreurs comme :

```
(venv) admini@srvn8n:~$ python test_odbc.py Erreur lors de la connexion ODBC : ('IM002', '[IM002] [unixODBC][Driver Manager]Data source name not found and no default driver specified (0) (SQLDriverConnect)')
```

J'ai essayé de les corriger en exportant des chemins ODBC, en relançant le shell, en vérifiant le DNS, ... Au final, je n'ai pas réussi. Je me suis donc rabattu sur un test avec isql.

J'ai donc fait des tests avec isql avec la commande : isql -v NOM_DSN UTILISATEUR MOTDEPASSE. En dépit de mes efforts, j'avais plein d'erreurs et je ne comprenais pas pourquoi ça ne fonctionnait pas. De plus, je me suis rendu compte que j'utilisais le mauvais type d'HFSQL. Il a été nécessaire de en fait utiliser un driver ODBC pour HFSQL Classic et non pour Client/Server.

J'ai donc corrigé ma configuration en :

```
[ODBC Data Sources]
<Nom Source> = HFSQL
[<Nom Source>]
Driver = /chemin/vers/WD300hfo64.so
RepFic = <chemin des fichiers>
```

Mais là encore, il y avait des erreurs que je n'arrivais pas à corriger. J'ai fait des tests encore et encore sans succès en changeant tout et n'importe quoi, j'ai même créé d'autres fichiers de configuration en plus et cherché d'autres méthodes. (Annexe 5.2)

Au final, avec l'aide de Joris BALDAN, j'ai réussi la connexion tout simplement en utilisant un ancien package pour le driver ODBC. Nous avons dû faire ça car les méthodes pour configurer l'ODBC n'étaient pas à jour depuis plusieurs années. En dépit de mes efforts, la connexion était trop instable pour pouvoir continuer la mission.

C) CREATION DE LA RAG

1) Démarche adoptée

Pour la création de la Recherche Augmentée par Génération (RAG), j'ai commencé par effectuer des recherches approfondies en cherchant d'abord les définitions des techniques que je devais utiliser.

Puis j'ai dû me renseigner pour comprendre comment mettre en place un RAG.

Après cela, j'ai regardé des vidéos pour comprendre plus en détails les concepts de RAG, LLM, Embedding, et des bases de données vectorielles et aussi pour pouvoir le mettre en place dans N8N.

Par la suite, j'ai cherché une base de données vectorielle à utiliser pour le RAG et tenté de l'installer sur ma machine. J'ai choisi d'utiliser une base de données vectorielle Qdrant. Joris BALDAN m'a aidé à l'installer sur la VM et m'a donné un Credential pour pouvoir utiliser la base de données Qdrant sur N8N et aussi un dossier contenant des documents à intégrer dans la base de données pour faire des tests avec l'agent IA.

Il m'a aussi donné un modèle d'exemple de workflow utilisant une RAG pour que je puisse vite avancer dessus. J'ai cherché à comprendre comment fonctionnait le RAG en analysant les différents nœuds puis je me suis mis au travail.

J'ai d'abord changé le modèle d'IA utilisé par Mistral en utilisant le modèle « mistral-small-latest » pour l'agent et « mistral-embed » pour le modèle d'embedding puis la base de données utilisé par Qdrant en y ajoutant les Credentials.

J'ai ensuite changé le workflow en utilisant un nœud Execute Command pour trouver les documents de ma machine puis un nœud Read Files from Disk pour les extraire. Enfin, j'ai utilisé un noeud Code pour formater la réponse.

Le workflow est en 4 étapes (Annexe 3.3) à savoir :

- Etape 1 : Je créer ma collection (à faire une seule fois)
- Etape 2 : Je nettoie puis intègre mes documents dans ma collection
- Etape 3 : Etape permettant à l'Etape 2 d'être plus rapide (elle n'a pas été utilisée)
- Etape 4 : Un agent IA va utiliser la collection pour répondre aux questions

2) Choix effectués

Pour la mise en place de la Recherche Augmentée par Génération (RAG), plusieurs outils ont été utilisés pour assurer le bon fonctionnement et l'efficacité du système :

- Qdrant: Base de données vectorielle open source conçue pour les applications d'intelligence artificielle de nouvelle génération.
- Mistral Small: Modèle de langage étant une version allégée de Mistral Large conçu pour donner des réponses brèves et concises.

Je suis parti sur Qdrant car elle semblait facile à utiliser et on a utilisé le modèle Mistral Small car il fallait choisir un modèle open source contrairement à Chat GPT pour pouvoir l'utiliser sans avoir créé de Credential en créant un compte et qui prenait peu de ressources tout

simplement car il fallait que le modèle soit installé sur une machine de l'entreprise pour être hébergé chez HABITALYS.

3) Erreurs rencontrées

Le développement du RAG sur N8N a été assez rapide et sans trop de difficultés. En effet grâce à mon tuteur qui m'a donné une bonne base, j'ai pu rapidement comprendre comment fonctionnait Qdrant et comment configurer ma base de données pour commencer à travailler sur le RAG. Lors du développement du workflow, j'ai rapidement compris comment fonctionnait le workflow de base et comment récupérer mes documents pour les intégrer dans la base de données vectorielle avec un Execute Command et un Read Files from Disk.

Cependant, j'ai quand même rencontré quelques difficultés. J'ai eu :

- Un problème de taille de données : J'ai rencontré des problèmes liés à la taille des données, ce qui a nécessité des ajustements pour réussir à intégrer les documents. En effet, les documents devaient être correctement formatés et segmentés pour être traités par le système. J'ai dû ajuster les paramètres de traitement pour m'assurer que les documents étaient correctement intégrés dans Qdrant en mettant la taille des vecteurs à 1024.
- Format de réponse de l'Agent IA incorrecte : L'Agent IA répondait initialement dans un format JSON peu compréhensible. Au début je n'avais pas compris pourquoi il répondait dans ce format mais avec l'aide de Joris BALDAN, j'ai finalement compris que c'était à cause des sauts de lignes qu'il faisait mal et on a donc ajouté un nœud de code à la sortie de l'agent pour formater les données correctement. Cela a permis de transformer les réponses en un format plus lisible et utilisable, améliorant ainsi la qualité des résultats générés par le système RAG.

PARTIE 4 : RESULTATS OBTENUS ET EVALUATION

A) UNE EXPORTATION COMPLETE DES DIAGNOSTICS

À la fin de cette tâche, les résultats obtenus ont été très satisfaisants.

J'ai réussi à faire l'export de CSV (Annexe 3.1) avec :

- Un trigger pour déclencher le workflow
- Une HTTP Request pour me connecter au site et récupérer le jeton d'authentification
- Une autre HTTP Request pour récupérer les données en JSON
- Un nœud Code pour formater les données au format voulu
- Un Convert to File qui convertit les données JSON en CSV
- Enfin un Write Files from Disk qui écrit le fichier dans le dossier souhaité

Au final, le CSV ne servait à rien donc je ne l'ai plus utilisé après.

Faire l'export de PDF (Annexe 3.1) avec :

- Un trigger pour déclencher le workflow
- Une HTTP Request pour me connecter au site et récupérer le jeton d'authentification
- Une HTTP Request pour récupérer les données en JSON
- Un nœud Code pour récupérer certaines données parmi celles du JSON d'avant
- Une HTTP Request pour récupérer les PDF de chaque intervention
- Enfin une boucle Loop Over Items qui va boucler sur un Write Files from Disk pour écrire les PDF dans le dossier souhaité

Faire l'export d'XML (Annexe 3.1) avec :

- Un trigger pour déclencher le workflow
- Une HTTP Request pour me connecter au site et récupérer le jeton d'authentification
- Une HTTP Request pour récupérer les données en JSON
- Un nœud Code pour récupérer certaines données parmi celles du JSON d'avant
- Une HTTP Request puis un Code pour récupérer l'URL, l'ID de l'intervention et le nom des PDF de chaque diagnostic
- Un autre Code qui va créer le XML
- Enfin une boucle Loop Over Items qui va boucler sur un Write Files from Disk pour écrire les XML dans le dossier souhaité

Fusionner les 3 exports en 1 (Annexe 3.1).

Une version avec Base de données au lieu de se connecter au site pour économiser les ressources utilisées mais très peu utilisé et vite oublié (Annexe 3.1). Elle fait la même chose mais au lieu de se connecter et de récupérer les données du site, elle les récupère depuis des fichiers sur le disque.

Faire le CSV qui affiche les PDF manquants et l'envoie par email (Annexe 3.1) avec :

- Un trigger pour déclencher le workflow
- Une HTTP Request pour me connecter au site et récupérer le jeton d'authentification
- Une HTTP Request pour récupérer les données en JSON
- Un nœud Code pour récupérer certaines données parmi celles du JSON d'avant
- Une HTTP Request puis un Code pour récupérer l'URL, l'ID de l'intervention et le nom des PDF de chaque diagnostic
- Un Code qui récupère les documents puis une HTTP Request qui récupère les informations des documents dans la GED (Base de données d'HABITALYS)
- Un Code qui crée les CSV puis un Read Files from Disk qui va lire le dernier CSV téléchargé.
- Un Merge qui récupère les informations des deux nœuds précédents puis un Code qui va comparer les deux CSV et ensuite un If qui va voir si les CSV sont identiques ou différents. S'ils sont différents, il envoie aux personnes concerné le CSV par email et remplace le dernier CSV dans le dossier sur le disque par le CSV créé aujourd'hui. S'ils sont identiques, alors il envoie un message à Jacques CHARRON et Christophe CASADEI en disant que tout va bien et qu'il n'y a pas de nouvelles interventions rentrées.

Enfin faire le workflow en 2 temps. A 7h00, l'envoie des PDF et XML. A 8h00, l'envoie des CSV manquants. (Annexe 3.1)

L'automatisation des processus au sein d'HABITALYS a apporté plusieurs avantages significatifs. On peut le voir avec :

- Temps économisé : L'automatisation des processus a permis de réduire considérablement le temps consacré aux tâches manuelles et répétitives. Les employés d'HABITALYS peuvent désormais se concentrer sur des activités à plus forte valeur ajoutée, ce qui a amélioré l'efficacité globale et la productivité de l'entreprise.
- Réduction des erreurs : L'automatisation a également permis de minimiser les risques d'erreurs humaines. Les processus automatisés sont plus fiables et précis, ce qui a réduit les erreurs liées à la manipulation manuelle des données.
- Évaluation par rapport au besoin initial : L'automatisation des exports de fichiers a répondu aux besoins initiaux en réduisant la charge de travail des employés et en minimisant les risques d'erreurs. Les résultats obtenus ont été évalués positivement, avec une amélioration notable de l'efficacité et de la productivité. Les workflows développés ont permis de répondre aux attentes et d'atteindre les objectifs fixés.

B) LA RECUPERATION DES MONTANTS

En dépit de mes efforts, je n'ai pas réussi à aller au bout de cette mission. Fort heureusement, les résultats obtenus ont été pour certains satisfaisants. J'ai réussi à :

- Développement d'un agent IA : L'implémentation d'un MCP sur N8N a permis de vérifier la faisabilité pour une IA d'utiliser des outils afin d'accéder à une base de données, répondant ainsi à un besoin critique d'accès et d'analyse de données.
- Connexion à une base de données : La connexion réussie à une base de données via JDBC a permis d'établir une interaction fonctionnelle avec les systèmes de gestion de données, ouvrant la voie à des intégrations futures.

Cependant, malgré ces avancées, une difficulté majeure est survenue lors de la tentative de création d'un driver ODBC pour HFSQL. L'objectif était de se connecter à une base de données HFSQL (Abyl) afin de récupérer des montants et comptes locataires en temps réel. Cette étape n'a pas pu être menée à bien en raison de méthodes de configuration obsolètes, n'ayant pas été mises à jour depuis plusieurs années. Cela a limité la portée des tests et validations possibles dans ce contexte spécifique et a donc empêché la réussite de la mission.

C) UN AGENT IA OPTIMISE

À la fin de cette tâche, les résultats obtenus ont été très satisfaisants :

- Amélioration de la recherche et de l'accès à l'information : La création de la Recherche Augmentée par Génération (RAG) a permis d'améliorer significativement la recherche et l'accès à l'information au sein d'HABITALYS. Les employés peuvent désormais effectuer des recherches plus précises et obtenir des résultats plus pertinents.
- Évaluation par rapport au besoin initial : La création de la RAG a répondu aux besoins initiaux en facilitant la recherche sémantique et en améliorant la précision des résultats. Les résultats obtenus ont été évalués positivement, avec une amélioration notable de l'efficacité et de la productivité.

PARTIE 5 : LA FIN D'UN STAGE, LE DEBUT D'UNE AVENTURE

A) UNE ENTREPRISE ENGAGEE ET UN ENVIRONNEMENT DE TRAVAIL FAVORABLE

Du point de vue des relations professionnelles, les conditions de travail chez HABITALYS m'ont agréablement surpris par leur qualité. L'ambiance de travail était sereine, et mon intégration au sein de l'équipe a été rapide et fluide. HABITALYS a mis à ma disposition des équipements et des installations qui ont grandement contribué à mon confort et à mon bien-être au sein de l'entreprise.

La flexibilité des horaires a également été très appréciable, me permettant de moduler mes journées de travail selon mes périodes de productivité. Cela m'a permis d'arriver tôt le matin et de partir relativement tôt le soir, ce qui a été particulièrement pratique pour équilibrer vie professionnelle et personnelle.

Concernant ma contribution à l'entreprise à l'issue de ce stage, j'ai pu apporter des solutions concrètes et préparer le terrain pour des améliorations futures :

- Automatisation des exports de fichiers : Les workflows développés sont fonctionnels et pourraient être étendus pour automatiser davantage de processus similaires au sein de l'entreprise.
- Création de la RAG : La Recherche Augmentée par Génération mise en place offre une base solide pour améliorer la gestion de l'information et la recherche sémantique, facilitant ainsi l'accès aux données et la prise de décision.

B) UNE EXPERIENCE DETERMINANTE ET UNE VISION CLAIRE DE L'AVENIR

Avant ce stage, j'étais indécis quant à la voie à emprunter pour la poursuite de mes études (voie classique ou alternance ?). Cette expérience m'a aidé à mieux définir ce que je souhaite pour la suite de mon parcours, que ce soit en termes d'études ou de carrière. Il m'a permis de mieux comprendre l'importance de poursuivre mes études pour acquérir des compétences plus approfondies et spécialisées.

Cette expérience m'a conforté dans l'idée de mener à bien les trois années de mon BUT. Elle m'a également convaincu de l'importance de poursuivre des études supérieures afin de me

spécialiser davantage et d'élargir mes perspectives professionnelles. Ce stage m'a permis de prendre conscience de la nécessité de la formation continue et de l'adaptation constante aux évolutions technologiques et aux nouvelles méthodes de travail.

Mon stage c'est si bien déroulé que l'entreprise HABITALYS m'a proposé de me prendre en alternance.

Sur le plan personnel, ce stage a renforcé ma motivation à poursuivre mes études et à m'engager dans des projets professionnels stimulants. Il m'a également permis de développer des compétences techniques et relationnelles précieuses pour ma carrière future. Je suis désormais résolu à continuer mes études avec un objectif clair en tête : devenir un professionnel compétent et polyvalent.

CONCLUSION

"Ce que l'on apprend avec plaisir, on ne l'oublie jamais." Alfred Mercier

Ce stage chez HABITALYS a représenté bien plus qu'une simple immersion professionnelle : il a été un véritable catalyseur de croissance, de prise de confiance, et de clarification de mes ambitions.

D'un point de vue technique, j'ai eu l'opportunité de me confronter à des problématiques concrètes autour de l'automatisation de processus, en manipulant des outils modernes tels que **N8N**, **Mistral**, ou encore **MCP**, et en interagissant avec des bases de données via **JDBC** et **ODBC**. Ces expériences, parfois semées d'embûches, ont mis à l'épreuve ma rigueur, ma capacité d'analyse et ma persévérance. Malgré les obstacles, j'ai su proposer des solutions fonctionnelles, contribuant à alléger la charge de travail des équipes et à fiabiliser des processus critiques. La création de la **Recherche Augmentée par Génération (RAG)** illustre également ma capacité à intégrer l'intelligence artificielle dans des systèmes complexes et utiles à l'entreprise.

Sur le plan humain, ce stage m'a permis d'évoluer dans un environnement bienveillant, encadré par des professionnels à l'écoute, disponibles, et investis. J'y ai découvert une culture d'entreprise engagée et un sens du collectif qui m'ont profondément marqué. Ces relations de travail harmonieuses ont été un moteur dans ma progression et dans ma motivation quotidienne.

Mais au-delà de la technique et de l'humain, ce stage a été une **expérience de transformation personnelle**. Il a renforcé ma confiance, aiguisé ma curiosité, et confirmé ma volonté de poursuivre mes études vers un parcours spécialisé. Il m'a aussi permis de m'affirmer dans mes choix, en m'ouvrant la voie vers l'alternance, proposée par l'entreprise elle-même, signe de la qualité de mon implication.

En somme, ce stage aura été une étape charnière dans mon parcours. Il m'a apporté bien plus que des compétences : une vision claire, des objectifs consolidés, et la conviction que le travail, lorsqu'il est animé par la passion et la persévérance, finit toujours par porter ses fruits.

GLOSSAIRE

N8N : N8N est une plateforme d'automatisation de workflows en open source. Elle permet la modélisation de processus métier et l'intégration et l'interaction d'applications.

N8N permet de :

- Réceptionner de la donnée, de la traiter, de l'envoyer à différents logiciels, etc...
- Effectuer plusieurs tâches.
- D'automatiser plusieurs tâches.

Workflow : Représentation d'une suite de tâches ou d'opérations effectuées par une personne, un groupe de personnes, un organisme, etc. On peut le définir comme la tâche globale que l'on veut effectuer.

Nœuds : Sous-tâches dans le Workflow.

Agent IA : Entité pilotée par l'intelligence artificielle, capable d'accomplir en autonome les tâches qui lui sont assignées.

Credential : Informations privées émises par des applications et des services pour vous authentifier en tant qu'utilisateur et vous permettre de vous connecter et de partager des informations entre l'application ou le service et le nœud n8n.

MCP : Protocole standard ouvert conçu pour faciliter la transmission de contexte entre les applications et les modèles de langage. Son objectif est de permettre à un LLM (IA) de se connecter facilement à différentes sources de données et outils, tout comme on connecterait divers périphériques à un ordinateur via un unique type de port universel.

NFS : Network File System, en français système de fichiers en réseau, est à l'origine un protocole qui permet à un ordinateur d'accéder via un réseau à des fichiers distants. Il permet de partager des données principalement entre systèmes UNIX.

CIFS : Common Internet File System est un protocole qui permet l'accès aux fichiers, aux imprimantes et aux ports série entre différentes machines d'un même réseau. Grâce à cet accès, les clients peuvent ouvrir, lire et modifier les documents partagés, qui sont automatiquement enregistrés sur les machines locales.

TEGIA : Logiciel utilisé par HABITALYS pour pouvoir faire certaines actions et enregistrer des données.

DBeaver : DBeaver Community est un outil de base de données multiplateforme gratuit pour les développeurs, les administrateurs de bases de données, les analystes et tous ceux qui travaillent avec des données. Il prend en charge toutes les bases de données SQL populaires telles que MySQL, MariaDB, PostgreSQL, SQLite, Apache Family, etc.

WEBOGRAPHIE

Calypso : Site accessible au sein de l'entreprise HABITALYS, utilisé régulièrement

Site Web de N8N : <https://n8n.io/>, utilisé régulièrement au cours des missions.

GitHub : <https://github.com/>, utilisé régulièrement au cours des missions.

Youtube : <https://www.youtube.com/>, utilisé régulièrement au cours des missions.

Wikipédia : https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal, utilisé régulièrement au cours de mes missions.

Site Web pour driver JDBC/ODBC : <https://doc.pcsoft.fr/fr-FR/?9000160>, utilisé régulièrement au cours de la seconde mission.