

Stage Data : Gestion, modélisation de données et Data visualisation

Entreprise : SIMPLON.CO

Direction : Système d'information (SI)

Équipe : Data

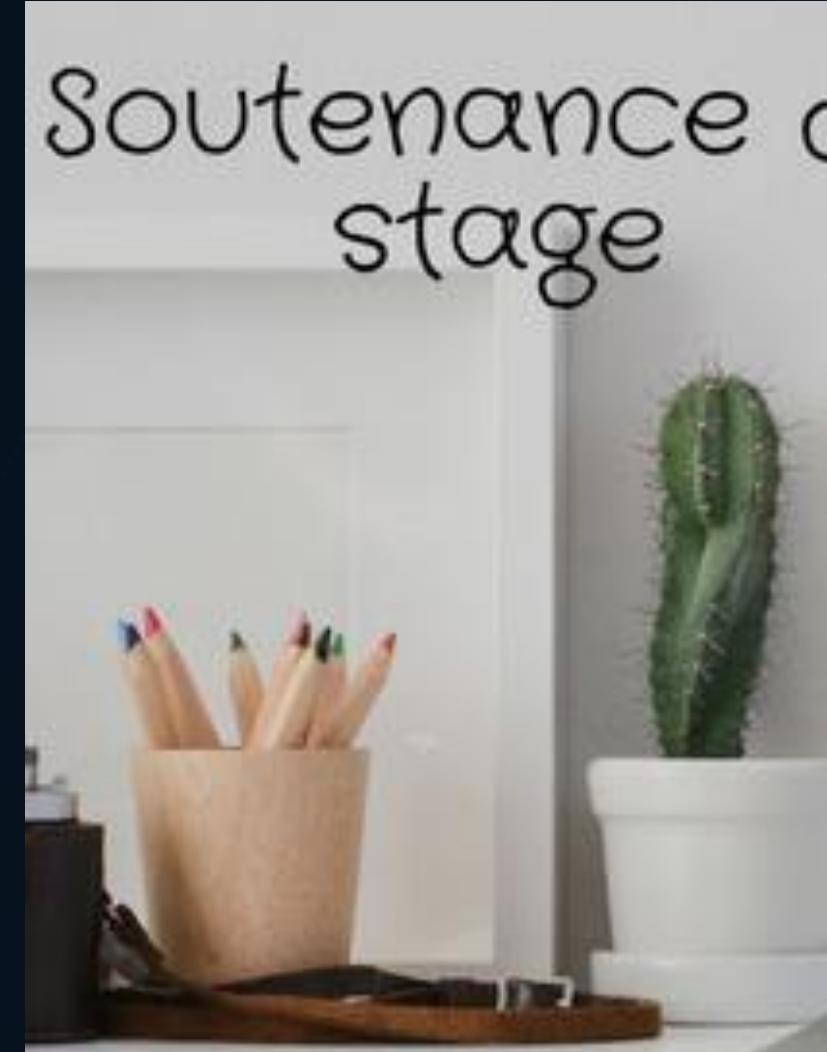
Tutrice entreprise :
Juliana CAVALCANTE

Tutrice universitaire :
Sophie DABO



Plan de présentation

1. **Presentation de l'entreprise**
2. **L'équipe Data & Cycle de la donnée**
3. **Mes missions globales entant que stagiaire Data**
4. **Projet Placement : l'un des projets les plus cruciaux réalisés**
 - **Contexte**
 - **Problème identifié**
 - **Proposition sollicité par le métier**
 - **Solution de l'équipe Data**
 - **Extraction de donnée**
 - **Modélisation des données en suivant les principes de l'architecture Data Warehouse**
 - **Création du tableau de bord et résultat**
5. **Conclusion**
6. **Suite à mon stage : Contrat en CDD**



1. Simplon.co : Qui sommes-nous ?



 Réseau de fabriques numériques inclusives (créé en 2013)

 Bootcamps intensifs : publics éloignés de l'emploi

 +19 937 apprenantes formées dans 34 pays

 Partenariats : Orange Coding Academy, consortiums européens

Simplon.co est un acteur majeur du numérique inclusif, créé en 2013.

C'est un réseau de fabriques numériques qui forme aux métiers du numérique via des bootcamps intensifs.

Aujourd'hui, Simplon.co est présent dans 34 pays et a formé plus de 19 937 apprenantes à travers le monde, avec des pôles comme Simplon Maghreb et Simplon Africa.
L'entreprise travaille avec des partenaires comme :

- les Orange Coding Academies en Jordanie dont plus de 742 apprenantes formées depuis 2019 avec 60% de sortie positive
- ou des consortiums européens financés par l'AFD et l'Union Européenne.

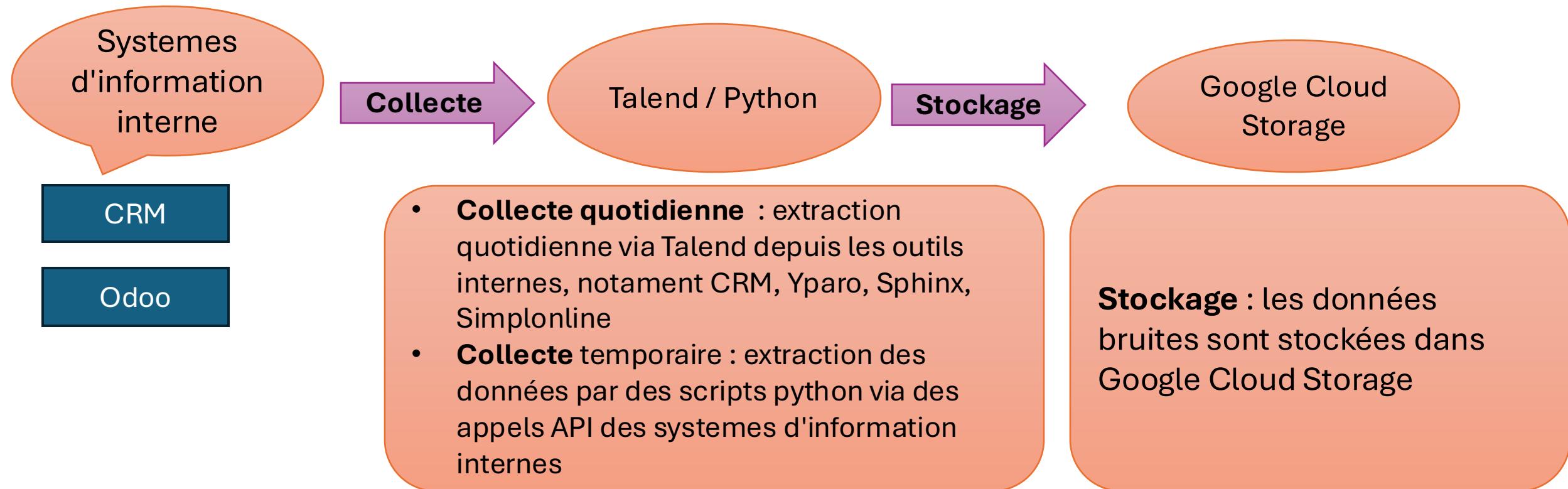
2. L'équipe Data & Cycle de la donnée

Pour piloter toutes ces activités notamment le suivi des sessions de formation, placement des apprenantes en entreprise, indicateurs de performance, la donnée est devenue stratégique. C'est dans ce contexte que s'inscrit mon stage au sein de l'équipe Data.



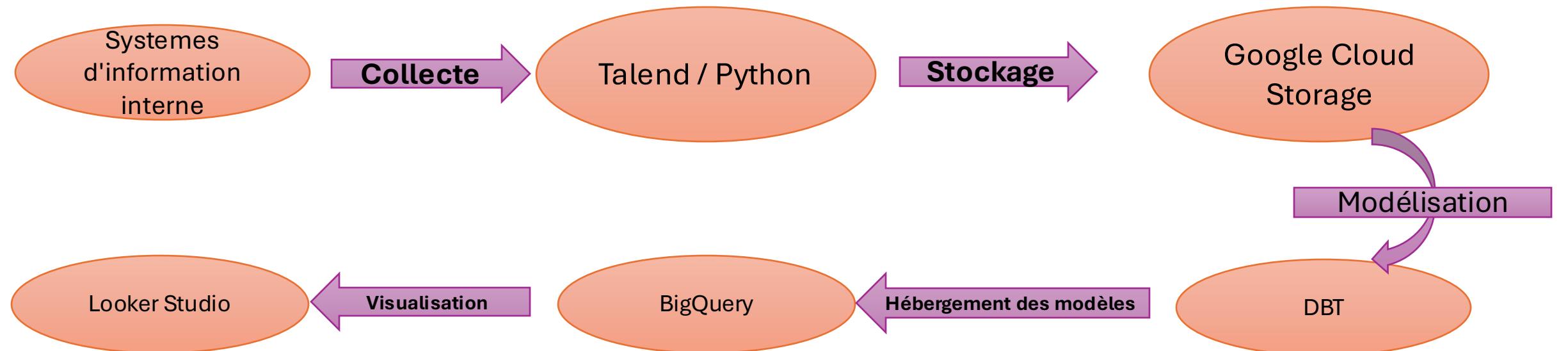
L'équipe Data & Cycle de la donnée

Sous la supervision de Juliana Chef des projets Data, l'équipe Data couvre l'ensemble du cycle de la donnée :



L'équipe Data & Cycle de la donnée

Sous la supervision de Juliana Chef des projets Data, L'équipe Data couvre l'ensemble du cycle de la donnée chez Simplon.co :



Looker Studio : Restitution visuelle par des tableaux des bords pour suivre les indicateurs de pilotage.

BigQuery : pour héberger les modèles DBT et être interrogé.

Modélisation : DBT pour créer des modèles analytiques, suivant le principe d'architecture de la data warehouse.



Stagiaire Data

3. Mes missions globales entant que stagiaire Data

Pendant ces 6 mois, j'ai apporté un soutien concret à l'équipe Data et aux métiers sur plusieurs volets :

Extraction temporaire de données

Data Quality

Modélisation

Visualisation

Support métiers

En dehors des flux de données Talend qui se fait par le Data Engineer, j'occupe les extractions des données temporaire via des script python par des appel API

Mise en place des contrôle des données via des macro SQL pour alerter, au quotidien, tout incohérence de la donnée.

Mise en place des modèles de données dans DBT hébergés dans BigQuery, en suivants le principe d'architecture de la Data Warehouse et refonte des modèles existants pour adapter aux besoin des équipes métiers

Mis en place des tableaux de bords sur Looker Studio.

À la disposition des équipes métiers pour répondre à leurs demandes de ticketing, notamment : des exports de donnée ainsi que des analyses et vérifications spécifiques.

4. Projet Placement

Mon travail a directement servi les équipes commerciales et le pole alternance



simplon

A. Contexte métier

Accompagner les apprenantes vers l'emploi est une priorité chez Simplon. Les chefs de projet alimentaient manuellement un Google Sheets recensant les candidats en recherche d'entreprise : prenom, nom, formation, email, tél, cv, region, ville, qpv, ... etc.

Ce fichier servait aux commerciaux pour matcher rapidement les profils avec les besoins des entreprises partenaires.



B. Problèmes identifiés

La saisie manuelle est source de :

- D'erreur,
- D'oublis de candidats,
- Des données non fiables,
- De perte de temps énorme.



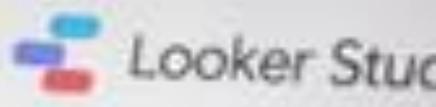
C. Proposition sollicitée

Le pôle alternance distance nous a sollicité pour créer un tableau de bord Looker Studio reprenant ces données avec des filtres spécifique, notamment :

- Region de la session
- Region du candidat
- Ville du candidat
- Nom de la formation
- Parcours de la formation

Tableau de bord Contacts

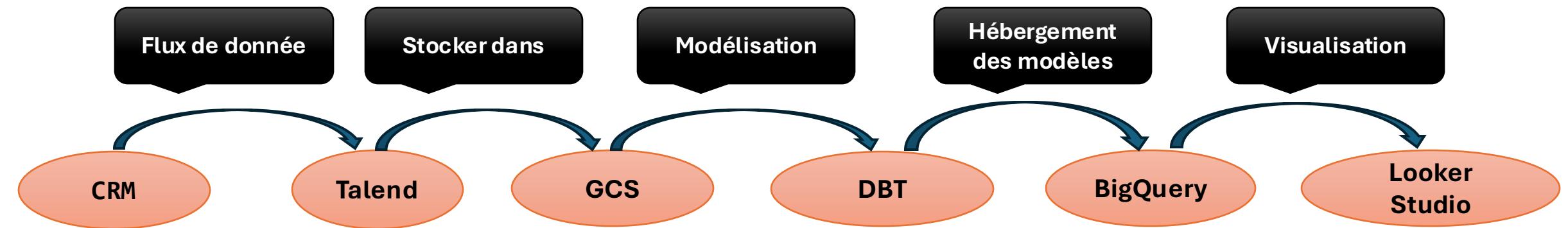
Nom	Prénom	Date de naissance	Email
Sabot	Jean	03/07/2004	sabot@implome.co
Sabot	Sophie	04/12/07/2004	sabot@simplome.co
Sabot	Patrice	03/07/2004	sabot@implome.co
Sabot	Julien	07/07/2004	sabot@implome.co
Sabot	Elisabeth	13/12/2004	sabot@implome.co
Sabot	Mathilde	07/07/2004	sabot@implome.co
Sabot	Pauline	07/07/2004	sabot@implome.co
Sabot	Juliette	07/07/2004	sabot@implome.co
Sabot	Mathilde	07/07/2004	sabot@implome.co
Sabot	Mathilde	07/07/2004	sabot@implome.co



Looker Studio

D. Notre solution technique Data

Avec ma tutrice Juliana, nous avons proposé une solution bien plus robuste : **ne plus partir du Google Sheets, mais exploiter les données déjà existantes dans le CRM Odoo .**

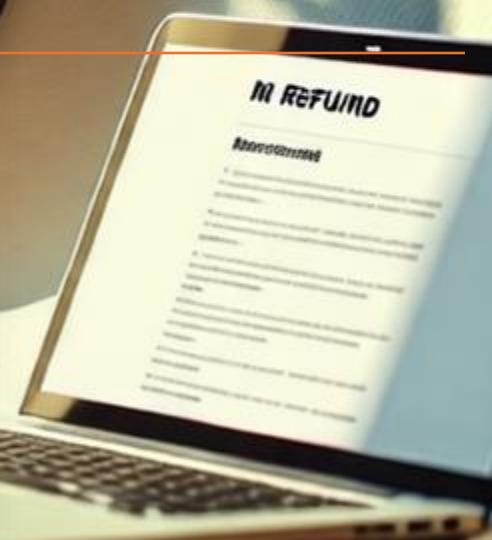


- **Extraction** des données, quotidienne, via Talend
- **Stockage des données brutes** les Google Cloud Storage
- **Modélisation** DBT/BigQuery pour structurer les indicateurs demandées
- **Restitution** via tableau de bord Looker Studio



- Automatisation complète,
- mise à jour quotidienne garantie,
- fiabilité des données.

E. Traitement de la demande (flux + modélisation)



1.

Identification des modèles techniques

La première étape a été d'**identifier précisément quels modèles CRM Odoo** contenaient les informations nécessaires au tableau de bord de placement :

ir_attachment

Contenant les pièces jointes, dont les CV. Champs technique nécessaires :

- Id : clé d'identification de la pièce
- Name : le nom de la pièce jointe

survey_user_input_line

Contenant les réponses des candidats lors de leurs inscriptions. On retrouve certaines informations nécessaires comme :

- Id : clé de la question
- question_id : la question
- display_name : la réponse à la question (ceci permet de filtrer des informations, notamment :

- QPV
- RSA
- France travail
- Nationalité
- RQTH

crm_lead

D'autres informations administratifs, notamment :

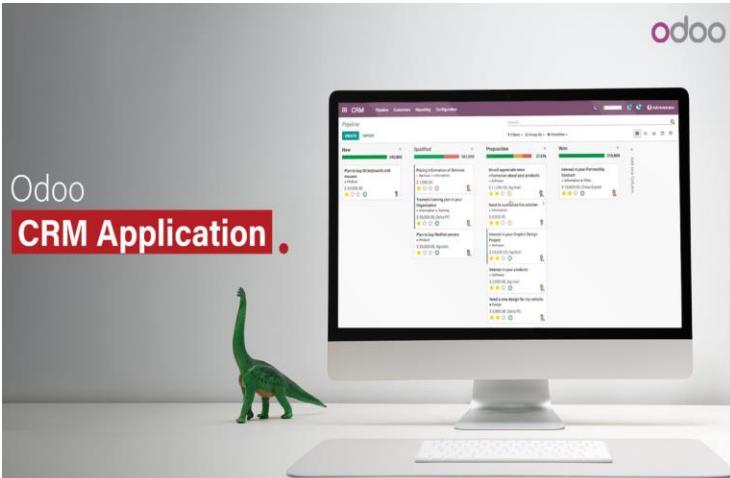
- Conatct_name: nom du candidat
- Contact_firstname : prénom
- Phone : téléphone
- Emai_from : email candidat
- Region : région du candidat
- Ville : ville du candidat
- Description : remarque candidature
- Nom_opportunité : la formation
- Impact_unit : region de la formation
- Stage_id : étape de la candidature (filtrer ceux qui sont en recherche d'entreprise.)

2.

Mise en place des flux Talend

Une fois les modèles identifiés, j'ai **listé les champs techniques** dont j'avais besoin dans chacun d'eux. J'ai transmis cette spécification au **Data Engineer**, qui a mis en place des **flux Talend quotidiens** pour alimenter notre environnement data dans Google Cloud Storage.

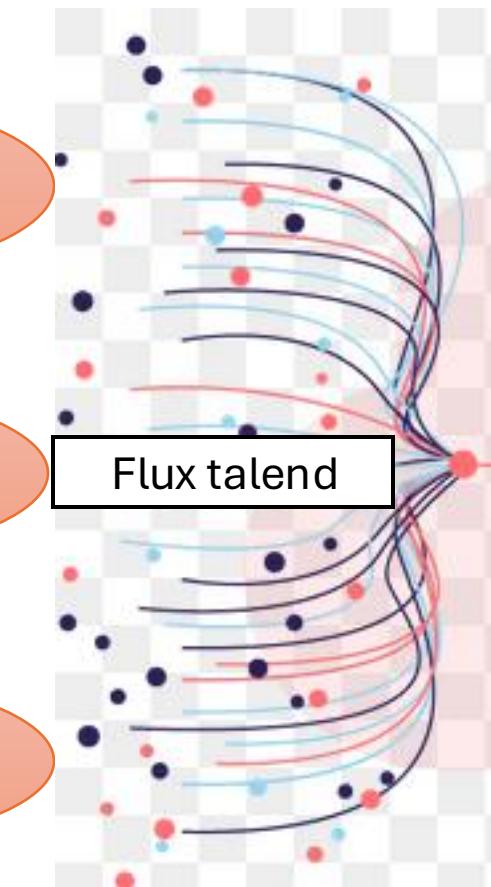
Système d'information : CRM



ir_attachment

survey_user_input_line

crm_lead



Stockage des données brutes



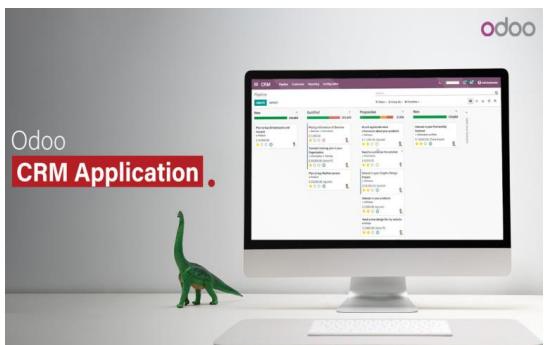
Cloud Storage

3.

Problème de volumétrie et choix de l'incrémental

Sur le modèle `survey_user_input_line`, nous avons rencontré un **problème de volumétrie** : le nombre de réponses de questionnaires accumulé au fil du temps est très important. Recharger **toutes** les réponses tous les jours via Talend, aurait **surchargé le serveur** et rendu les flux instables. Pour contourner cela, on a mis en place une stratégie en **deux temps** :

1. J'ai développé un **script Python personnalisé** qui appelle l'API Odoo pour récupérer **tout l'historique des réponses** une bonne fois pour toutes.



survey_user
_input_line



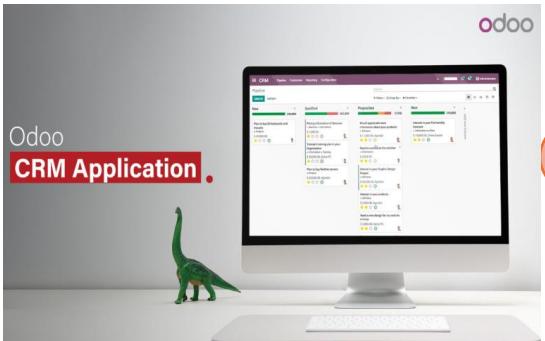
Stockage des données
brutes



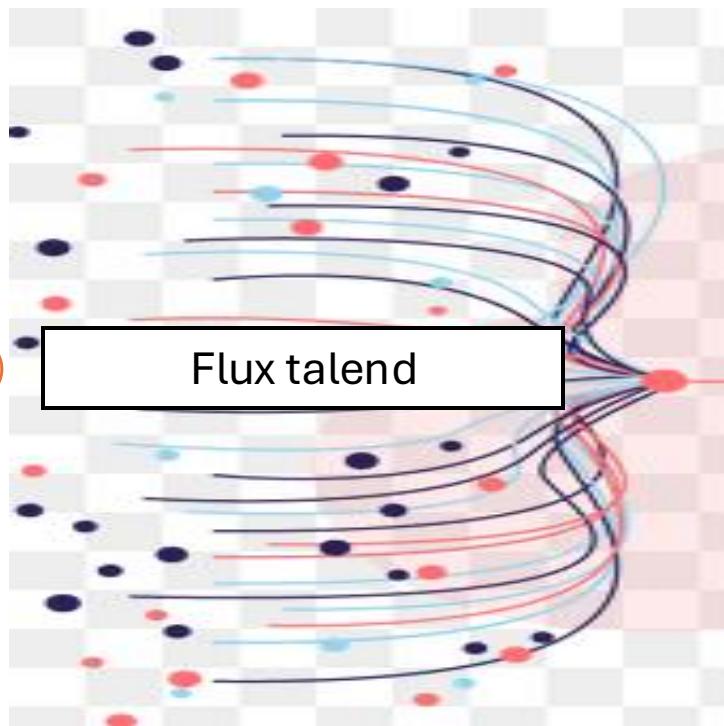
3.

Problème de volumétrie et choix de l'incrémental

2. Ensuite, le Data Engineer a mis en place un **flux Talend quotidien** qui ne récupère que les **nouvelles réponses**.



survey_user
_input_line



Stockage des données brutes



Cloud Storage



4.

Modélisation Data Warehouse

A.

Tables staging

Dans ces modèles de staging, je fais essentiellement du **renommage des colonnes** techniques en noms plus compréhensibles pour les équipes métiers, sans encore appliquer de logique complexe :

stg_odoo_ir_attachment

stg_odoo_survey_user_input_line

Stg_crm_lead

Renommage des champs,
exemple :

- Id = id
- Name = nom_piece_joint

Renommage de champs, exemple

:

- Id = id
- question_id = question
- display_name = reponse

Renommage de champs, exemple :

- Conatct_name = nom
- Contact_firstname = prenom
- Phone = telephone
- Emai_from = email
- Region : region_candidat
- Ville = ville_candidat
- Description = remarque_candidature
- Nom_opportunité = formation
- Parcours
- Impact_unit = region_formation
- Stage_id = etape

B.

Tables ODS

Ensuite, on passe à la couche **ODS**. Là, je commence à intégrer de la **logique métier** et à faire les **jointures** :

ods_odoo_crm_ir_attachment

ods_odoo_crm_survey_question_filtered

ods_odoo_crm_survey_response_pivot

Je filtre les pièces jointes qui contiennent 'CV' et je prends la dernière ajoutée, puis je construis le lien vers ce CV dans le CRM.

Je filtre les questions des questionnaires pour ne garder que celles qui demandent les informations nécessaires, comme: nationalité, QPV, RSA, RQTH, France Travail, etc.

je **reconstruis le profil candidat** à partir de ces réponses filtrées : chaque ligne correspond à un candidat avec ses champs dérivés des réponses, par exemple sa nationalité, son statut QPV, son statut RSA, s'il a une RQTH, etc.

C.

Table Dimension

À partir de tables ODS, j'ai construit une **dimension candidats**, **dim_odoo_crm_reponse_sondage**.
Cette dimension combine :

dim_odoo_crm_reponse_sondage

les informations administratives liées à l'envoi du sondage,

toutes les informations personnelles reconstruites depuis les réponses :
nationalité, QPV, RSA, RQTH, contact France Travail, numéro de téléphone,

ainsi que des éléments pratiques comme : lien vers le CV du candidat.

D.

Table de fait

Cette table existait déjà avant mon arrivée et elle contient les opportunités du CRM, donc la relation entre les candidats et les formations . Enfin, à partir de la table dimension, j'ai enrichi la **table de faits existante** en rajoutant :

fact_odoo_crm_opportunité_table

les informations issues des questionnaires comme : QPV, RSA, RQTH, France travail

les liens CRM vers le CV et la fiche candidat

Se servir des certaines information déjà présentes, comme :

Nom candidat, prénom candidat, téléphone, email, région du candidat, ville, remarque sur la candidature, formation, parcours, région de la formation et étape de la candidature.

5. Restitution visuelle

The image shows a person's hand holding a tablet. On the screen of the tablet, there is a dashboard titled "Tableau de bord Contacts". The dashboard displays a list of eight contacts in a table format. The columns are labeled "Nom", "Prénom", "Date de naissance", and "Email". The data is as follows:

Nom	Prénom	Date de naissance	Email
Dupont	Jean	08.07.2000	jean.dupont@implone.co
Martin	Sophie	04.12.1998	sophie.martin@implone.co
Dumont	François	08.03.2001	francois.dumont@implone.co
Morisseau	Étienne	06.05.2000	etienne.morisseau@implone.co
Dummont	David	18.12.1999	david.dumont@implone.co
Durousseau	Sophia	08.08.2000	sophia.durousseau@implone.co
Pétrigny	Raphaële	03.07.2001	raphaelle.petrigny@implone.co

The Looker Studio logo is visible at the top right of the tablet screen.

Data visualisation



PLACEMENT : CANDIDATS EN RECHERCHE D'ENTREPRISE



Impact Unit session

Région candidat

Parcours

Ville candidat

Ville candidat

Nom campagne

Note importante : Comment lire les remarques ?

Le champ Remarque est affiché de manière abrégée dans le tableau. Pour consulter le texte complet d'une remarque, il suffit de placer le pointeur de la souris sur la cellule correspondante : le contenu intégral apparaîtra alors en info-bulle. Si le texte est trop long, exporter la donnée sinon consulter le candidat directement dans le CRM.

ADMINISTRATIF

CRITERES DE PLACEMENT

Fiche candidat	CV	Prénom	Nom	Formation	Email	Téléphone	Région	Ville	QPV	RSA	RQTH	Contrat alternance	Mobilité	Remarque
Christelle Morello - Test														
				Rémi Dubus - Concepteur Développeur d'Applications DevOps (Alternance Simplon)			Hauts-de-France	La Chapelle d'Armentières		No	Contrat de professionnalisation	Régional		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Noui		Adel Noui - FRB2B_WS_AIS_12M	[REDACTED]@gmail.com	[REDACTED]	Auvergne-Rhône-Alpes	Chanain		No	Contrat de professionnalisation	Régional	Mariane41anscontract proa passé 20 ans dans Kotelle en tant que que responsable logistique mais fermeture	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Boncompain		Agathe Boncompain - Concepteur Développeur d'Applications DevOps	[REDACTED]@boncompain@gmail.com	[REDACTED]	Nouvelle-Aquitaine	Anglet		No	Contrat d'apprentissage	Régional		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	All[REDACTED]	PASTOR	Alexandra PASTOR - Concepteur Développeur d'Applications DevOps	[REDACTED]	[REDACTED]	Auvergne-Rhône-Alpes	HEYRIEUX		No	Contrat de professionnalisation	Régional		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	All[REDACTED]	Billon	CAMPUSDIST_TSSR Surveiller_2025_ALT	[REDACTED]aux-air.fr	[REDACTED]					Contrat d'apprentissage			
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	An[REDACTED]obo	SOW	CAMPUSDIST_TAI_P4_ALT	[REDACTED]mail.com	[REDACTED]	GSO	Bordeaux	Oui	Non	Contrat d'apprentissage		CAP Maintenance informatique légère réalisée en 2020 Renseignements et formations DWWM	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Cieslar		Anais Cieslar - Concepteur Développeur d'Applications DevOps	[REDACTED].com	[REDACTED]	Hauts-de-France	Lille		Yes	Contrat d'apprentissage	Régional		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Diez		Anaïs Diez - Concepteur Développeur d'Applications DevOps	[REDACTED].com	[REDACTED]	Occitanie	Toulouse		No	Contrat de professionnalisation	Régional		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	LAHBIB		CAMPUSDIST_DataEng_2025_ALT	[REDACTED].all.com	[REDACTED]	SUD	92110	Non	Non	Contrat de professionnalisation			

D. Conclusion



BILAN DE MON STAGE SIMPLON.CO

✓ Apports entreprise : Extraction, modélisation, automatisation, fiabilisation des données métiers, restitution visuelle, appuie aux équipes métiers.

✓ Ce que j'ai appris :

- Data Engineering moderne (Extraction de données via un script Python par appel API, Google Cloud Platform, data lakes, data warehouse, modélisation de données (étoile, flocon), versioning et gestion de projets data avec Git/GitHub (branches, PR, workflows)).
- Traduire les besoins métiers en solutions techniques Data
- La modélisation en suivant l'architecture d'un Data warehouse