

<input type="checkbox"/>	cle_abonnement	cle_facture	date_emission_facture	conso_facture	date_releve_index_pr...	date_releve_index_fa...	nb_fac
<input type="checkbox"/>	GN_1180780700356501	GN 1180780700356501 2...	2022-09-06	60	2022-02-17	2022-08-24	5
<input type="checkbox"/>	GN_1180780700356901	GN 1180780700356901 2...	2022-09-06	1	2022-02-17	2022-08-24	5
<input type="checkbox"/>	GN_1180780700377406	GN 1180780700377406 2...	2022-09-06	70	2022-02-17	2022-08-24	5
<input type="checkbox"/>	GN_1180780700455901	GN 1180780700455901 2...	2022-09-06	19	2022-02-17	2022-08-24	5
<input type="checkbox"/>	GN_1180780700470401	GN 1180780700470401 2...	2022-09-06	408	2022-02-17	2022-08-24	2
<input type="checkbox"/>	GN_1180880800109301	GN 1180880800109301 2...	2022-10-06	0	2022-03-16	2022-09-20	2
<input type="checkbox"/>	GN_1180880800118501	GN 1180880800118501 2...	2022-10-06	546	2022-03-16	2022-09-20	5
<input type="checkbox"/>	GN_1180880800281801	GN 1180880800281801 2...	2022-10-06	0	2022-03-16	2022-09-20	5
<input type="checkbox"/>	GN_1180880800284601	GN 1180880800284601 2...	2022-10-06	0	2022-03-16	2022-09-20	5
<input type="checkbox"/>	GN_1183183100168201	GN 1183183100168201 2...	2022-10-12	1	2022-03-23	2022-09-27	5
<input type="checkbox"/>	GN_1183183100244601	GN 1183183100244601 2...	2022-10-12	28	2022-03-23	2022-09-27	5
<input type="checkbox"/>	GN_1183183100248601	GN 1183183100248601 2...	2022-10-12	37	2022-03-23	2022-09-27	5
<input type="checkbox"/>	GN_1181881800075801	GN 1181881800075801 2...	2022-11-17	164	2022-05-03	2022-11-07	5
<input type="checkbox"/>	GN_1181881800212401	GN 1181881800212401 2...	2022-11-17	650	2022-05-03	2022-11-07	5

***15 000 collaborateurs dépendent de données critiques...  
mais beaucoup du temps des équipes data est perdu à traquer des erreurs.***



# Introduction

## **Problème clé**

Qualité des données non scalable, contrôles manuels chronophages, risques d'erreurs opérationnelles

## **Souhait**

Automatiser la génération de contrôles qualité via un agent IA

# WZ-Solutions

Notre agent IA détecte automatiquement les anomalies dans les données de Veolia pour déterminer leur qualité, générant des contrôles et requêtes SQL optimisés.



Prosper WANG

ESILV - M1 Datascience & IA



Kylie WU

ESILV - M1 Cybersécurité



Mark-Killian ZINENBERG

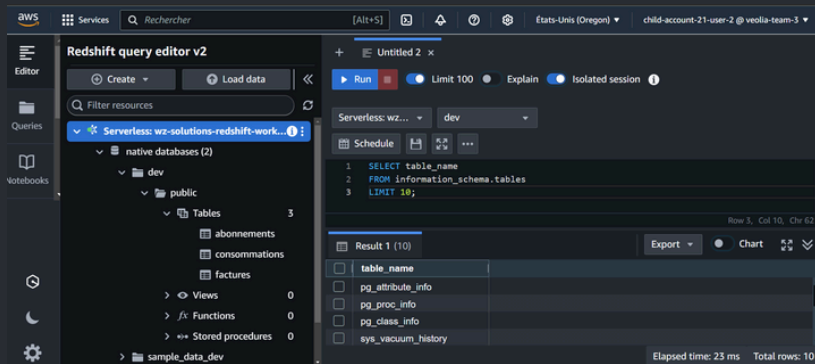
ESILV - M1 Ingénierie  
Financière



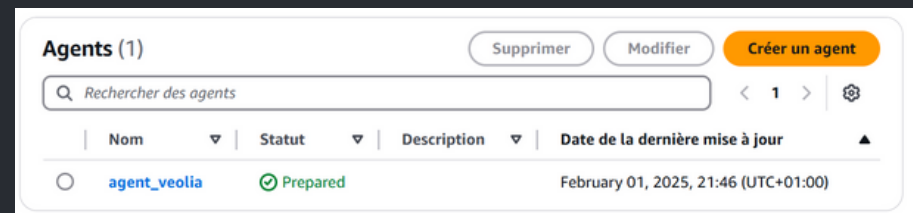
Ilan ZINI

ESILV - M1 Datascience & IA

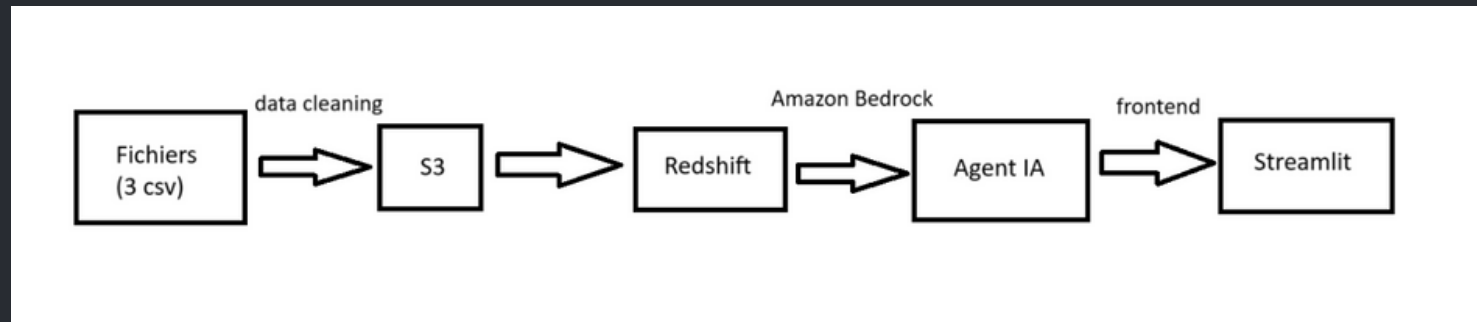
# Notre solution



Redshift



Agent IA



# Démo



Analyse des données



Détection des anomalies



Génération de requêtes SQL

Veolia Data Quality AI

localhost:8501

RUNNING... Stop Deploy

## Veolia - Analyse et Qualité des Données

Description des Tables Détection d'Anomalies Analyse IA avec Mistral Exécution des Requêtes SQL

### Structure des Tables Redshift

Table abonnements

- cle\_abonnement (character varying)
- date\_entree\_local\_abonnement (date)
- date\_resiliation\_abonnement (date)
- date\_souscription\_abonnement (date)

Voir un extrait (abonnements)

Table consommations

# Notre solution

## Analyse des Tables Redshift

### abonnements

1. Gère les informations sur les abonnements des clients.
2. cle\_abonnement : identifiant unique de l'abonnement
  - date\_entree\_local\_abonnement : date d'entrée en vigueur locale de l'abonnement
  - date\_resiliation\_abonnement : date de résiliation de l'abonnement
  - date\_souscription\_abonnement : date de souscription de l'abonnement


### consommations

1. Enregistre les données de consommation des clients.
2. annee\_conso : année de la consommation
  - mois\_conso : mois de la consommation
  - diametre\_nominal : diamètre nominal de la conduite
  - volume\_mois : volume consommé dans le mois
  - type\_abaque : type d'abaque utilisé pour le calcul de la consommation
  - libelle\_categorie\_abonne : libellé de la catégorie de l'abonné
  - code\_contrat : code du contrat
  - libelle\_territoire : libellé du territoire
  - libelle\_region : libellé de la région
  - cle\_pds : identifiant unique du point de livraison
  - date\_conso\_mois : date de la consommation mensuelle

### factures

1. Gère les informations sur les factures émises aux clients.
2. num\_fac\_par\_pds : numéro de facture par point de livraison
  - nb\_jours\_connus : nombre de jours connus pour la facturation
  - nb\_factures\_par\_pds : nombre de factures par point de livraison
  - conso\_facture : consommation facturée
  - cle\_facture : identifiant unique de la facture
  - cle\_abonnement : identifiant de l'abonnement associé à la facture
  - date\_emission\_facture : date d'émission de la facture
  - date\_releve\_index\_facture : date du relevé d'index de la facture
  - date\_releve\_index\_precedent\_facture\_composite : date du relevé d'index précédent de la facture composite

# Notre solution

 \*\*Rapport d'Anomalies dans les Jointures\*\* 

Type d'anomalie : Factures sans abonnement actif  
Description : Des factures ont été émises pour des abonnements qui ne sont pas actifs.  
Requête SQL :  
```\n`

```
SELECT f.num_fac_par_pds, f.date_emission_facture, a.cle_abonnement
FROM factures f
LEFT JOIN abonnements a ON f.cle_abonnement = a.cle_abonnement
WHERE a.cle_abonnement IS NULL;
```\n`
```

Type d'anomalie : Factures sans date d'émission  
Description : Des factures ont été enregistrées sans date d'émission.  
Requête SQL :  
```\n`

```
SELECT f.num_fac_par_pds
FROM factures f
WHERE f.date_emission_facture IS NULL;
```\n`
```

Type d'anomalie : Factures sans relevé d'index  
Description : Des factures ont été émises sans relevé d'index.  
Requête SQL :  
```\n`

```
SELECT f.num_fac_par_pds
FROM factures f
WHERE f.date_releve_index_facture IS NULL;
```\n`
```

Type d'anomalie : Factures sans abonnement  
Description : Des factures ont été enregistrées sans abonnement associé.  
Requête SQL :  
```\n`

```
SELECT f.num_fac_par_pds
FROM factures f
LEFT JOIN abonnements a ON f.cle_abonnement = a.cle_abonnement
WHERE a.cle_abonnement IS NULL;
```\n`
```

# Bénéfices & Limitations

1

## **Gain de temps**

Réduction de X heures/jour  
pour les professionnels data

2

## **Fiabilité**

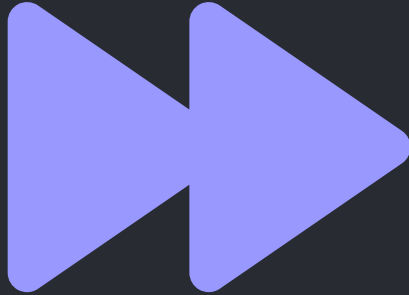
Moins d'erreurs dans la  
valeur des données

3

## **Scalabilité**

Facilité à intégrer dans les flux existants de Veolia





## Next steps

1

### **Amélioration de l'architecture solution**

Utilisation du multi-agents

2

### **Faciliter les mises à jour et les tests en continu**

Pipeline CI / CD pour tests réguliers

3

### **Automatisation des analyses**

Analyses périodiques via AWS Lambda ou AWS Step Function

# Conclusion

1

**Impact**  
immédiat

2

**Tech**  
maîtrisée



3

**Solution**  
réplicable

**Merci pour votre attention !**