



Conception et mise en œuvre d'un Entrepôt de données

PROJET 1

Etudiant : FREDERIC ABDOURAHMANE DIALLO

Enseignant: Mr. IBRAHIM LY

INTRODUCTION

- Le projet de conception et de mise en œuvre d'un entrepôt de données pour l'entreprise MacroBus s'inscrit dans une démarche d'amélioration du pilotage de la performance commerciale.
- Face à un volume croissant de commandes et à une diversité de marchés géographiques, l'objectif était de structurer les données opérationnelles de manière à offrir une base fiable et dynamique aux outils de reporting décisionnel.
- Ce projet a consisté à créer un entrepôt de données robuste à partir d'une base de production, à le charger via différentes approches ETL, et à valider sa pertinence à travers des cas d'usage concrets : suivi des commandes par région, classement des commerciaux, top produits par période, etc.
- L'ensemble du travail aboutit à un tableau de bord interactif Power BI permettant un accès rapide, visuel et synthétique aux indicateurs clés de performance.

Éléments	Détails
Nom du projet	Entrepôt de données MacroBus
Objectif	Centraliser, structurer et exploiter les données commerciales MacroBus
Livrables	Script SQL, entrepôt, chargement ETL, visuels Power BI, rapport final
Outils utilisés	- Montant / nombre de commandes par région - Top 2 commerciaux - Top 3 produits T2 2003 - Montant S2 2003 par filiale & catégorie
Indicateurs validés	PostgreSQL, Power BI, DAX, fichiers .bat, pipelines NiFi ou SQL
Dashboard	Page interactive avec filtres temporels & vues croisées
Durée estimée	~15 jours (selon plan initial)

I. Présentation de l'architecture globale du projet

1. Système source (base de production)

C'est la base transactionnelle opérationnelle utilisée par l'entreprise MacroBus pour gérer les commandes, les produits, les commerciaux, etc. Elle contient des données brutes, orientées processus, organisées en structures complexes optimisées pour la saisie (modèle relationnel) avec postgres.

Exemple : les tables commandes, clients, produits, commerciaux...

Voici des captures des tables présent dans la base de source de donnée

macrobus_production/postgres@PostgreSQL 17

Query Query History

15 SELECT * FROM clients;

Data Output Messages Notifications

Showing rows: 1 to 50 Page No: 1 of 1

	id_client [PK] integer	nom character varying (100)	prenom character varying (100)	email character varying (150)	telephone character varying (20)	region character varying (100)	pays character varying (100)
1	1	Ndiaye	Awa	awa.ndiaye@gmail.com	771234567	Dakar	Sénégal
2	2	Ba	Moussa	moussa.ba@gmail.com	772345678	Saint-Louis	Sénégal
3	3	Diop	Fatou	fatou.diop@gmail.com	773456789	Kaolack	Sénégal
4	4	Fall	Abdou	abdou.fall@gmail.com	774567890	Thiès	Sénégal
5	5	Gueye	Mariama	mariama.gueye@gmail.com	775678901	Ziguinchor	Sénégal
6	6	Sow	Cheikh	cheikh.sow@gmail.com	776789012	Kolda	Sénégal
7	7	Faye	Khady	khady.faye@gmail.com	777890123	Fatick	Sénégal
8	8	Diallo	Ibrahima	ibrahima.diallo@gmail.com	778901234	Tambacounda	Sénégal
9	9	Ba	Mame	mame.ba@gmail.com	770001111	Saint-Louis	Sénégal
10	101	Sow	Mame	mame.sow@gmail.com	771112233	Thiès	Sénégal
11	102	Diop	Cheikh	cheikh.diop@gmail.com	772223344	Dakar	Sénégal
12	103	Sy	Ousmane	ousmane.sy@gmail.com	773334455	Saint-Louis	Sénégal
13	104	Ba	Fatou	fatou.ba@gmail.com	774445566	Ziguinchor	Sénégal
14	105	Fall	Adama	adama.fall@gmail.com	775556677	Louga	Sénégal
15	106	Camara	Aminata	aminata.camara@gmail.com	776667788	Kaolack	Sénégal
16	107	Faye	Ibrahima	ibrahima.faye@gmail.com	777778899	Kolda	Sénégal
17	108	Gueye	Mariama	mariama.gueye@gmail.com	778889900	Dakar	Sénégal
18	109	Seck	Khady	khady.seck@gmail.com	779990011	Matam	Sénégal
19	110	Ndiaye	Awa	awa.ndiaye@gmail.com	770001122	Tambacounda	Sénégal

Total rows: 50 Query complete 00:00:00.788 CRLF Ln 15, Col 24

macrobus_production/postgres@PostgreSQL 17

Query Query History

16 SELECT * FROM commandes;

Data Output Messages Notifications

Showing rows: 1 to 79 Page No: 1 of 1

	id_commande [PK] integer	id_client integer	id_produit integer	id_commercial integer	date_commande date	quantite integer	montant_total numeric (12,2)
1	1	1	1	1	2023-01-15	1	5000000.00
2	2	2	3	2	2023-04-10	1	8000000.00
3	3	3	2	3	2023-06-05	1	12000000.00
4	4	4	4	4	2023-07-20	1	6500000.00
5	5	5	1	5	2023-08-01	2	10000000.00
6	6	6	2	6	2023-09-12	1	12000000.00
7	7	7	3	1	2023-10-22	1	8000000.00
8	8	8	4	2	2023-11-18	2	13000000.00
9	9	1	1	1	2023-06-01	1	18000000.00
10	10001	5	3	2	1998-08-24	10	3286190.74
11	10002	5	2	2	1997-12-14	4	4704000.39
12	10003	6	1	2	1995-10-31	2	2150057.12
13	10004	3	5	4	1999-11-21	3	652761.46
14	10005	5	6	1	2025-02-26	4	6368610.90
15	10006	5	1	2	2012-04-01	8	3645360.25
16	10007	5	1	4	2024-07-21	3	668791.10
17	10008	3	6	3	2023-05-19	1	5438453.85
18	10009	6	2	2	2013-07-11	9	2009249.53
19	10010	5	3	3	1997-01-12	9	1804363.98

Total rows: 79 Query complete 00:00:00.218 CRLF Ln 16, Col 26

The screenshot shows a PostgreSQL query editor interface. At the top, the connection is 'macrobus_production/postgres@PostgreSQL 17'. Below the connection bar, there's a toolbar with various icons. The 'Query' tab is active, showing a SQL query: `SELECT * FROM regions;`. Below the query, the 'Data Output' tab is active, displaying a table with 5 columns: `id_produit` (integer), `nom_produit` (character varying (100)), `categorie` (character varying (50)), and `prix` (numeric (10,2)). The table contains 9 rows of data.

	<code>id_produit</code> [PK] integer	<code>nom_produit</code> character varying (100)	<code>categorie</code> character varying (50)	<code>prix</code> numeric (10,2)
1	1	Toyota Hilux	Utilitaire	18000000.00
2	2	Hyundai Accent	Voiture	9500000.00
3	3	Renault Kangoo	Utilitaire	10500000.00
4	4	Mitsubishi L200	Pick-up	17000000.00
5	5	Kia Rio	Voiture	8900000.00
6	6	Isuzu D-Max	Camion léger	16000000.00
7	7	Toyota Coaster	Minibus	25000000.00
8	8	Nissan Patrol	SUV	27000000.00
9	9	TOYOTA HILUX	UTILITAIRE	18000000.00

2. Extraction / Nettoyage

- Dans cette phase, les données pertinentes sont extraites des systèmes sources via une scrypte sql pour le chargement 1 et via `executesql` sur nifi pour le chargent 3. Elles subissent des transformations essentielles :
 - Formatage des dates
 - Dédoublonnage
 - Conversion de types
 - Vérification d'intégrité (valeurs manquantes, clés étrangères...)
- Cette étape garantit que seules des données **fiables et cohérentes** alimentent les couches suivantes.
- a. Voici le scrypte sql proposer pour cette tache pour le chargement 1 qui va extraire, nettoyer et charger les données en même temps.

```

1  -- Connexion à la base source
2  SELECT dblink_connect('src', 'dbname=macrobus_production user=postgres password=2002 host=localhost');
3
4  -- dim_clients
5  INSERT INTO dim_clients (id_client, nom, prenom, email, telephone, region, pays)
6  SELECT *
7  FROM dblink('src',
8  $$SELECT DISTINCT id_client, nom, prenom, email, telephone, region, pays
9  FROM clients
10 WHERE nom IS NOT NULL AND prenom IS NOT NULL
11 AND email LIKE '%%' AND email NOT LIKE '%@%'
12 AND telephone ~ '[0-9]{7,}$'$$
13 ) AS t(id_client INT, nom TEXT, prenom TEXT, email TEXT, telephone TEXT, region TEXT, pays TEXT);
14
15 -- dim_regions
16 INSERT INTO dim_regions (id_region, nom_region, pays)
17 SELECT *
18 FROM dblink('src',
19 'SELECT DISTINCT id_region, nom_region, pays FROM regions'
20 ) AS t(id_region INT, nom_region TEXT, pays TEXT);
21
22 -- dim_produits
23 INSERT INTO dim_produits (id_produit, nom_produit, categorie, prix)
24 SELECT *
25 FROM dblink('src',
26 'SELECT DISTINCT id_produit, nom_produit, categorie, prix FROM produits WHERE prix IS NOT NULL AND prix > 0'
27 ) AS t(id_produit INT, nom_produit TEXT, categorie TEXT, prix NUMERIC);
28
29 -- dim_commerciaux
30 INSERT INTO dim_commerciaux (id_commercial, nom, prenom, email, id_region)
31 SELECT *
32 FROM dblink('src',
33 $$SELECT DISTINCT c.id_commercial, c.nom, c.prenom, c.email, c.id_region
34 FROM commerciaux c
35 JOIN regions r ON c.id_region = r.id_region
36 WHERE c.nom IS NOT NULL AND c.prenom IS NOT NULL
37 AND c.email LIKE '%%'$$

```

Ce script est accompagné d'un fichier execute.bat qui vas nous permettre de la lancer automatiquement via le cmd. Voir le contenu de l'image si dessous.

```

Fichier  Modifier  Affichage

@echo off
SETLOCAL

REM === Configuration PostgreSQL ===
SET PGUSER=postgres
SET PGPASSWORD=2002
SET PGDATABASE=macrobus_entrepot
SET PGHOST=localhost
SET PGPORT=5432
SET SCRIPT=C:\Users\User\Desktop\Macrobus_project\chargement1\chargement1_filtre.sql



REM === Exécution ===
echo =====
echo Chargement des données en cours...
echo Base cible : %PGDATABASE%
echo Script utilisé : %SCRIPT%
echo =====

psql -U %PGUSER% -h %PGHOST% -p %PGPORT% -d %PGDATABASE% -f "%SCRIPT%"

IF %ERRORLEVEL% NEQ 0 (
    echo ✖ Une erreur est survenue pendant le chargement.
) ELSE (
    echo ✔ Chargement terminé avec succès.
)

```

Maintenant après lancement du fichier execute_charg1.bat

 chargement1_filtre	29/06/2025 23:10	Fichier source SQL	3 Ko
 execute_charg_1	28/06/2025 02:48	Fichier de comma...	1 Ko

Voici l'image du lancement si dessous qui va automatiquement faire le traitement.

```

C:\WINDOWS\system32\cmd. x + v
=====
Chargement des données en cours...
Base cible : macrobus_entrepot
Script utilisé : C:\Users\User\Desktop\Macrobus_project\chargement1\chargement1_filtre.sql
=====
dblink_connect
=====
OK
(1 ligne)

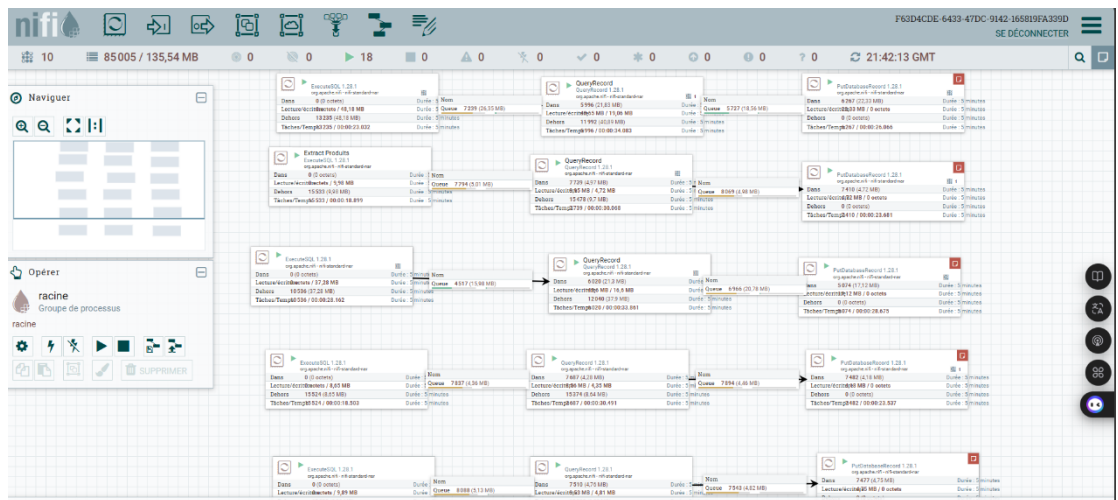
INSERT 0 44
INSERT 0 14
INSERT 0 9
INSERT 0 7
INSERT 0 78
INSERT 0 79
dblink_disconnect
=====
OK
(1 ligne)

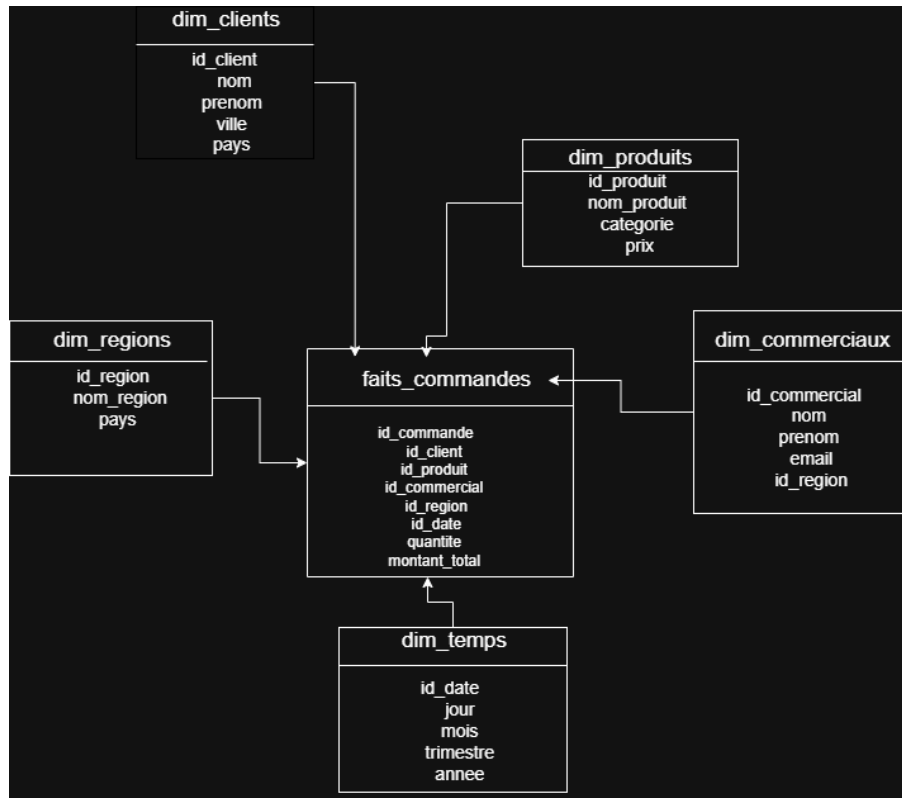
Le chargement terminé avec succès.
Appuyez sur une touche pour continuer...

```

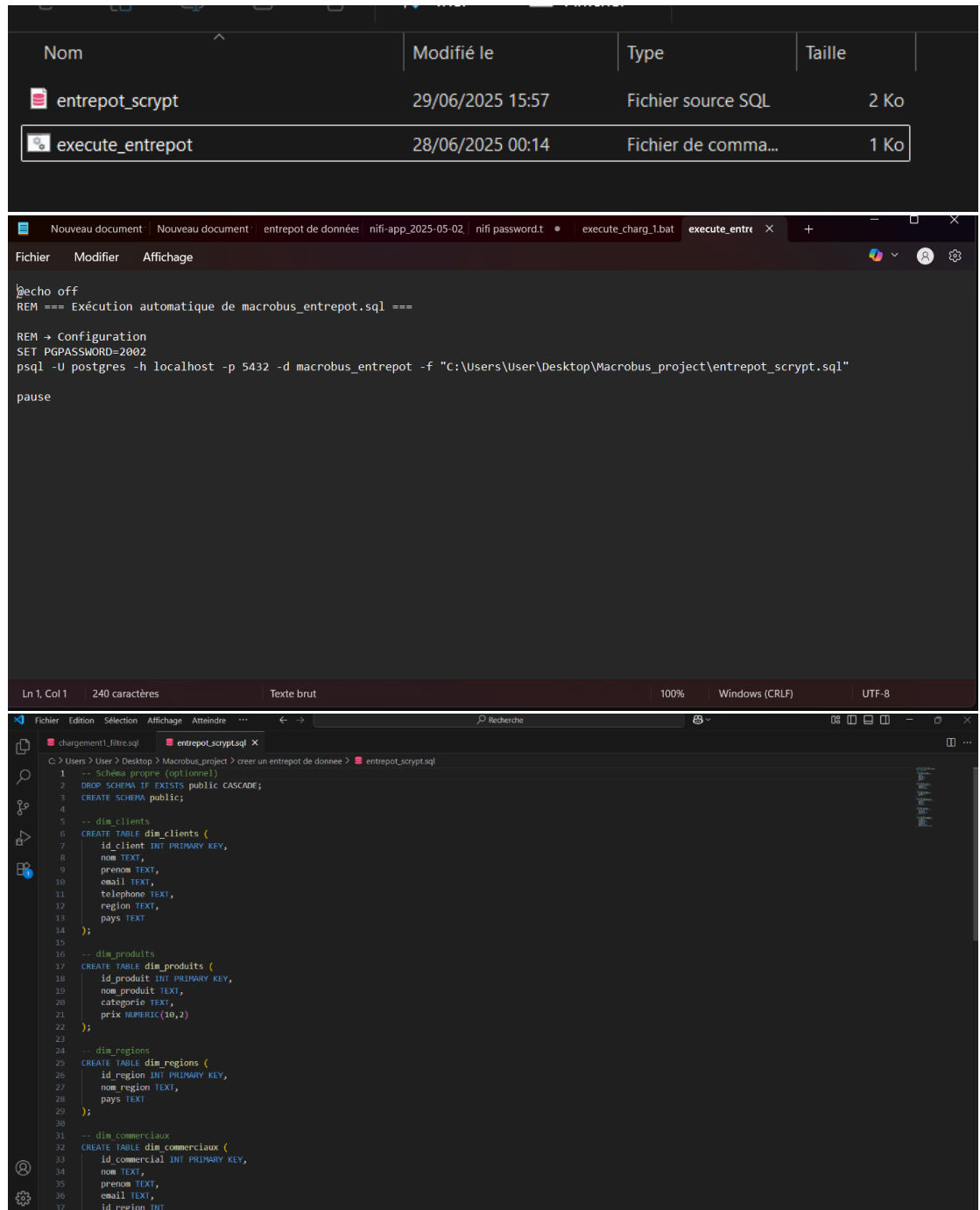
Voici le même procédé avec nifi ici nous voyons des pipelines de données conceptionnel et qui sont opérationnel relié à PostgreSQL.

- b. Voici les relation et pipelines en image avec executesql pour extraire les données, QueryRecord pour filtre et putdatabase pour insérer les données déjà nettoyer dans l'entrepôt de données.





Maintenant la création de l'entrepôt de donne via un scripatesqp avec un fichier execute.bat.



Pour finir voici l'entrepôt de donnée en image avec les données nettoyer et charger.

```
30 SELECT * FROM dim_regions;
```

Data Output Messages Notifications

	id_region [PK] integer	nom_region text	pays text
1	4	Fatick	Sénégal
2	12	Kolda	Sénégal
3	5	Kaolack	Sénégal
4	13	Sédhiou	Sénégal
5	11	Kédougou	Sénégal
6	14	Ziguinchor	Sénégal
7	8	Saint-Louis	Sénégal
8	10	Tambacounda	Sénégal
9	1	Dakar	Sénégal
10	9	Matam	Sénégal
11	3	Diourbel	Sénégal
12	2	Thiès	Sénégal
13	6	Kaffrine	Sénégal
14	7	Louga	Sénégal

Query Query History

28 **SELECT** * **FROM** dim_produits;

Data Output Messages Notifications



	id_produit [PK] integer	nom_produit text	categorie text	prix numeric (10,2)
1	2	Hyundai Accent	Voiture	9500000.00
2	5	Kia Rio	Voiture	8900000.00
3	4	Mitsubishi L200	Pick-up	17000000.00
4	1	Toyota Hilux	Utilitaire	18000000.00
5	9	TOYOTA HILUX	UTILITAIRE	18000000.00
6	7	Toyota Coaster	Minibus	25000000.00
7	3	Renault Kangoo	Utilitaire	10500000.00
8	6	Isuzu D-Max	Camion léger	16000000.00
9	8	Nissan Patrol	SUV	27000000.00






Query Query History

27 **SELECT** * **FROM** dim_commerciaux ;

Data Output Messages Notifications



	id_commercial [PK] integer	nom text	prenom text	email text	id_region integer
1	5	Seck	Pape	pape.seck@macrobus.sn	5
2	3	Fall	Modou	modou.fall@macrobus.sn	3
3	4	Diouf	Aminata	aminata.diouf@macrobus.sn	4
4	7	Diallo	Ibrahima	ibrahima.diallo@macrobus.sn	1
5	2	Sarr	Astou	astou.sarr@macrobus.sn	2
6	6	Sy	Yacine	yacine.sy@macrobus.sn	6
7	1	Ndiaye	Alioune	alioune.ndiaye@macrobus.sn	1

	id_date [PK] date 	jour integer 	mois integer 	trimestre integer 	annee integer 
1	1999-08-07	7	8	3	1999
2	2004-06-14	14	6	2	2004
3	1997-03-17	17	3	1	1997
4	2004-04-04	4	4	2	2004
5	2003-05-16	16	5	2	2003
6	2000-11-18	18	11	4	2000
7	2000-04-06	6	4	2	2000
8	2011-08-24	24	8	3	2011
9	2003-01-13	13	1	1	2003
10	2003-11-23	23	11	4	2003
11	1997-01-12	12	1	1	1997
12	2023-07-20	20	7	3	2023
13	2003-06-13	13	6	2	2003
14	1995-10-31	31	10	4	1995
15	2019-05-12	12	5	2	2019
16	2021-05-08	8	5	2	2021
17	2022-04-07	7	4	2	2022
18	2003-12-19	19	12	4	2003
19	1996-01-10	10	1	1	1996

macrobus_entrepot/postgres@PostgreSQL 17

Query Query History

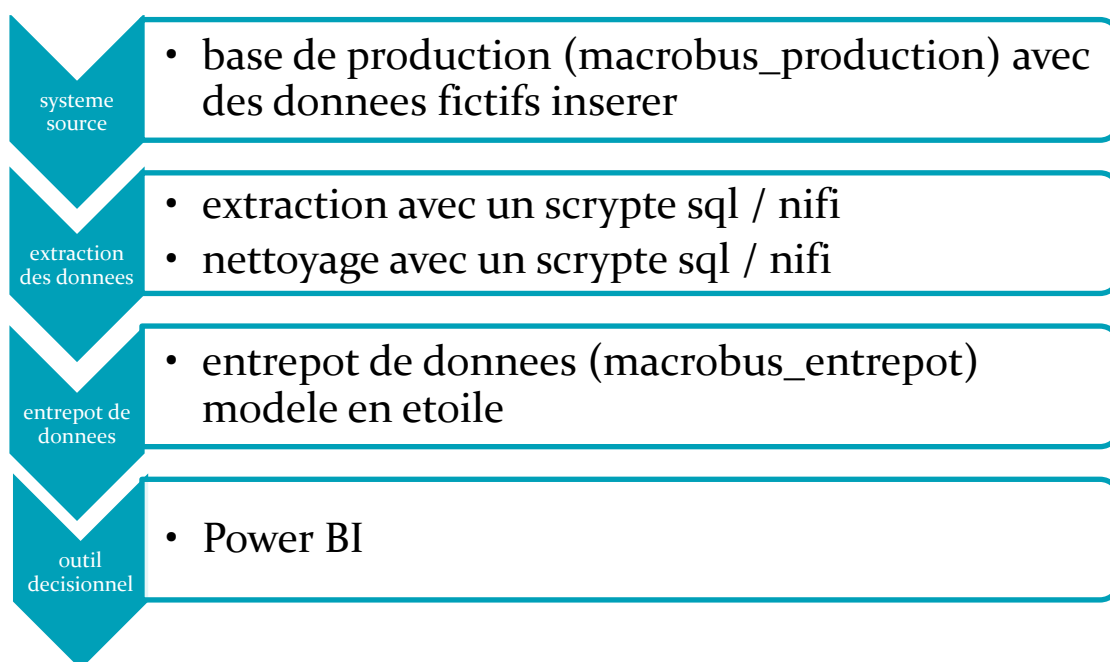
25 **SELECT** * **FROM** faits_commandes;

Data Output Messages Notifications

Showing rows: 1 to

	id_commande [PK] integer	id_client integer	id_produit integer	id_commercial integer	id_region integer	id_date date	quantite integer	montant_total numeric (18,2)
1	30010	1	4	3	1	2003-04-18	7	6164637.55
2	10049	1	4	4	1	2001-07-14	7	426774.42
3	10046	1	2	3	1	2001-08-20	6	4917923.38
4	10043	1	5	4	1	2011-08-24	10	4354503.85
5	10036	1	5	3	1	2015-06-16	5	5615036.75
6	10035	1	3	1	1	2010-06-30	1	2693575.09
7	10015	1	2	3	1	2011-07-10	7	4669806.41
8	9	1	1	1	1	2023-06-01	1	18000000.00
9	1	1	1	1	1	2023-01-15	1	5000000.00
10	30009	4	3	3	2	2003-06-13	3	1389542.34
11	30008	4	3	3	2	2003-05-25	9	2613013.82
12	30007	4	5	2	2	2003-05-06	10	7533870.62
13	30006	4	4	2	2	2003-06-29	2	5691480.14
14	30005	4	5	2	2	2003-05-16	9	2252858.82
15	30004	4	3	4	2	2003-06-27	5	360795.66
16	20010	4	4	3	2	2003-11-23	4	2044742.38
17	20008	4	3	1	2	2003-12-07	6	1663899.65
18	20003	4	2	1	2	2003-11-13	8	3219669.35
19	20002	4	3	3	2	2003-10-05	5	7534057.65

Voici si dessous l'illustration de l'architecture globale du projet avec un petit schéma visuel



4. Validation fonctionnelle de l'entrepôt

Cette étape visait à **évaluer la qualité du modèle décisionnel** en le confrontant à des cas d'usage concrets.

Indicateur métier	Mesure DAX utilisée
Montant & nombre de commandes par région	TotalMontant, NbCommandes
Top 2 commerciaux (commandes)	NbCommandes_Par_commande
Top 3 produits – T2 2003	Montant_T2_2003
Montant par filiale & catégorie – S2 2003	Montant_s3_2003
Commandes par mois et par région	NbCommandes + dim_temps[mois] dim_regions[nom]

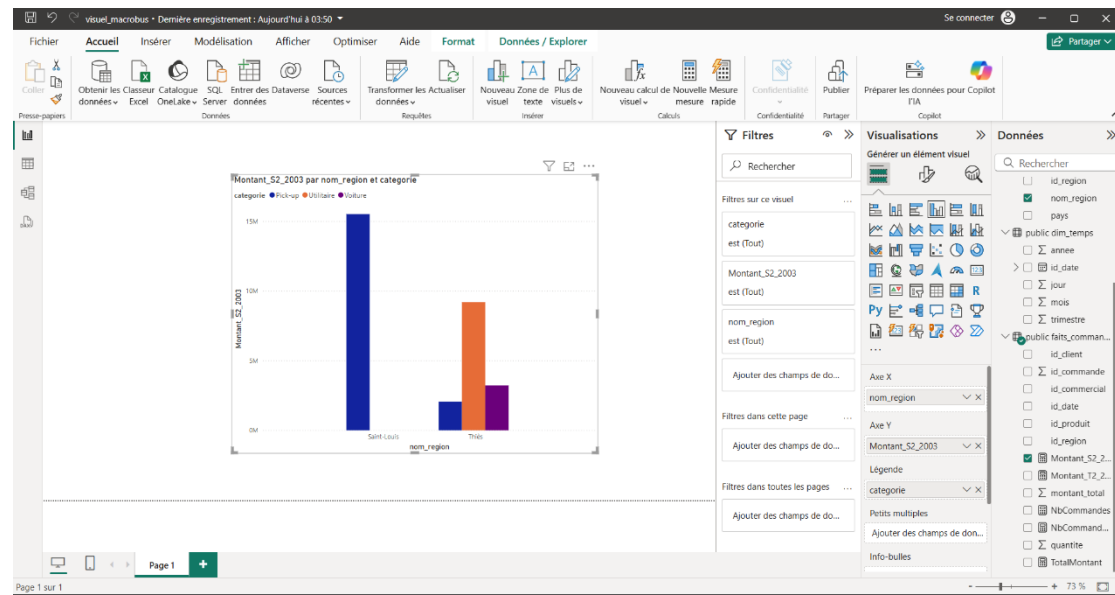
Toutes les mesures ont été vérifiées visuellement et techniquement à travers Power BI.

5. Tableau de bord Power B

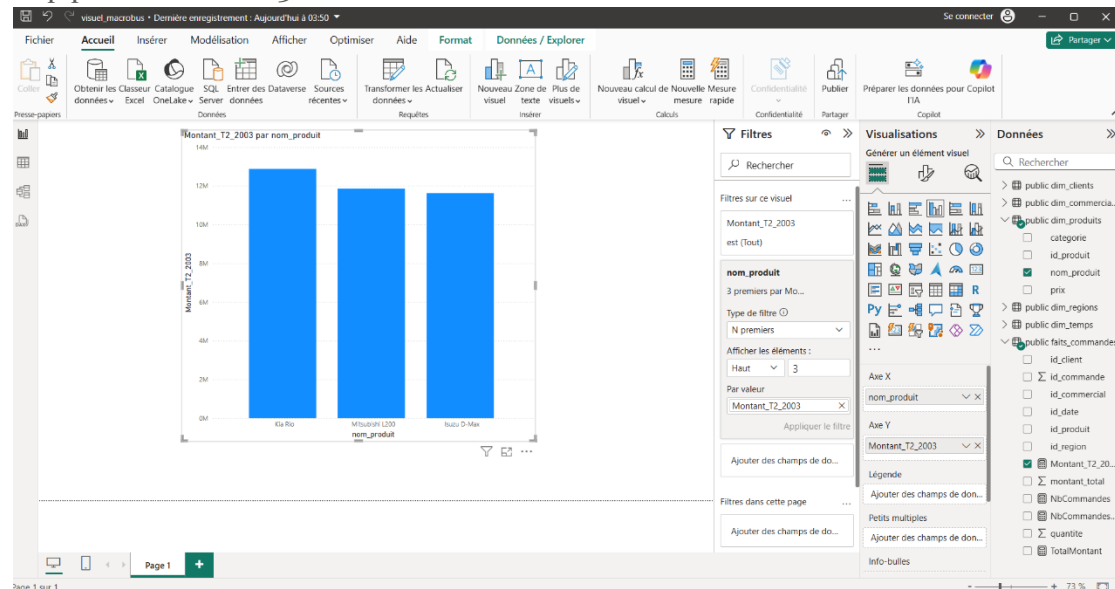
- **Page 1 : Vue globale**
 - Total commandes & montant par région
 - Filtres dynamiques : année, région, produit
- **Page 2 : Performance des commerciaux**
 - Top 2 commerciaux
 - Évolution mensuelle par filiale
- **Page 3 : Analyse produits**
 - Top produits T2 2003
 - Répartition par catégorie
- **Page 4 : Synthèse S2 2003**
 - Montants par filiale & catégorie
 - Indicateurs croisés

Voici les images du Dashboard

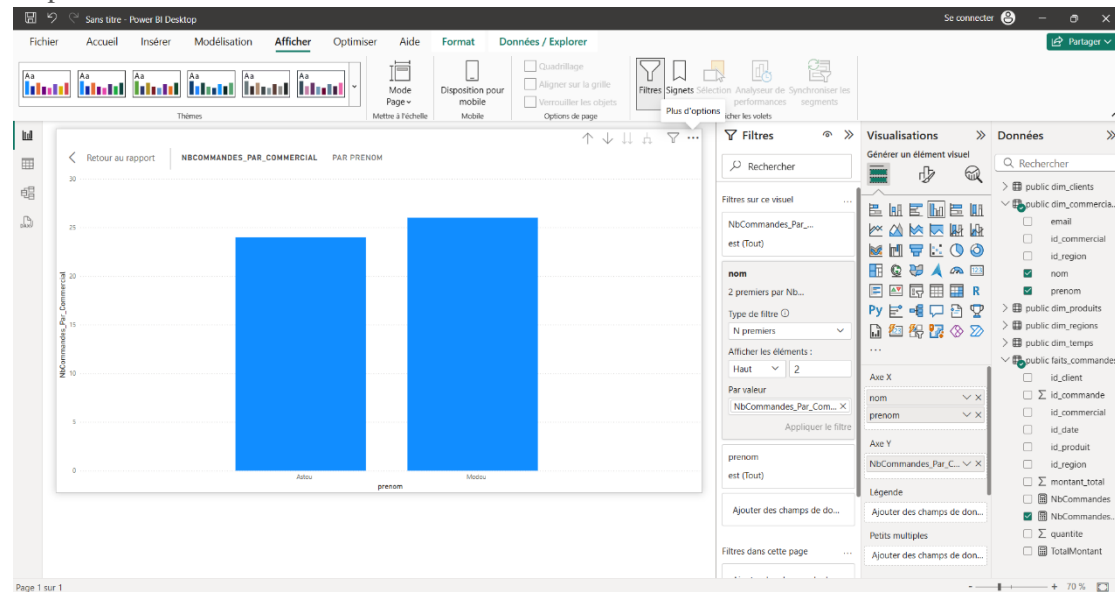
Synthèse S2 2003



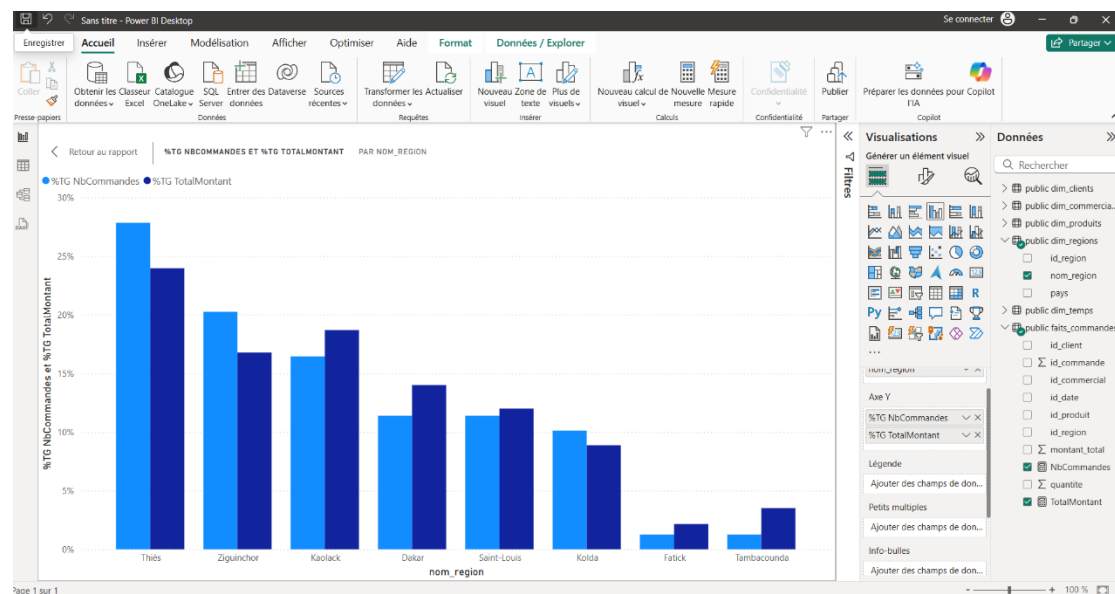
Top produits T2_2003



Top 2 commerciaux



Total commandes & montant par région



6. Conclusion du projet

Le projet d'entrepôt de données MacroBus a permis de :

- Structurer les données opérationnelles dans un **modèle en étoile performant**
- Définir et implémenter un processus **ETL efficace et automatisable**
- Valider la pertinence métier des données à travers des **indicateurs stratégiques**

- Offrir une **interface visuelle dynamique** via Power BI pour le suivi des performances.

En somme cela m'a permis d'acquérir de nouvelles compétences dans le domaine et j'espère qu'elles pourront me servir dans un futur proche.