

UNIVERSITE DE YAOUNDE 1

INF 351
GENIE LOGICIEL
RAPPORT

RAPPORT SUR TALEND

Élèves :

NGONGANG NANA LEA 21T2318
TAMOKOUE SIMO PRINCE
22Y1048
TIOMELA ZANGUE 21U2144
TAGNY TAGNY IDRIS 21T2342

Enseignant :

DR VALERIE MONTHE



8 février 2025



Table des matières

1	Analyse	2
1.1	Description des différents éléments constituant le DW	2
1.2	Dictionnaires des mesures	2
1.3	Dictionnaire des dimensions	2
2	Conception	4
2.1	Choix du Modèle d'Entrepôt	4
3	Implementation : approche ROLAP	4
3.1	Choix du SGBD	4
3.2	Base de Données Opérationnelle	4
3.3	Entrepôt de Données	4
4	Alimentation	5
4.1	Processus ETL	5
4.2	Utilisation de Talend OS-DI	5
4.3	Généralités	5
4.4	Requêtes et Tables de la Base de Données Opérationnelle	5
4.5	Configuration de la Table Affiliation	6
4.6	Création de la Table Année	6
4.7	Agrégation des Données avec tAggregateRow	6
4.8	Création de la Table de Fait	7



1 Analyse

1.1 Description des différents éléments constituant le DW

- **fait** : publication est notre table de fait qui est constituée par les id de : année , auteur , article
- **mesures** : notre mesure c'est le nombre de publication
- **dimensions** :
 - Année : Dimension temporelle qui exprime l'année de publication.
 - Article : Dimension représentant les articles publiés
 - Auteur : Dimension représentant les auteurs des articles .
 - Affiliation : Dimension représentant les institutions des auteurs.
 - Ville : Dimension géographique, avec une granularité au niveau de la ville.
 - Pays : Dimension géographique, avec une granularité au niveau du pays.

1.2 Dictionnaires des mesures

Code	Désignation	Type	Formule d'extraction
M_NB_PUB	Nombre total de publications	Entier	SELECT COUNT(*) FROM publication

1.3 Dictionnaire des dimensions

Code	Désignation	Type	Formule d'extraction
ID_annee	ID de l'année	Entier	SELECT id_annee FROM Annee
annee	Année de la conférence	Entier	SELECT DISTINCT annee FROM Annee

TABLE 2 – Dimension Année

Code	Désignation	Type	Formule d'extraction
ID_ARTICLE	ID de l'article	Entier	SELECT id_article FROM publication
TITRE	Titre de l'article	Texte	SELECT titre FROM Article

TABLE 3 – Dimension Article

Code	Désignation	Type	Formule d'extraction
ID_AUTEUR	ID de l'auteur	Entier	SELECT id_auteur FROM auteur
nom	Nom de l'auteur	Texte	SELECT nom FROM Auteur

TABLE 4 – Dimension Auteur



Code	Désignation	Type	Formule d'extraction
ID_AFFIL	ID de l'affiliation	Entier	SELECT id_affiliation FROM affiliation
NOM_AFFIL	Nom de l'institution	Texte	SELECT nom_institution FROM affiliation

TABLE 5 – Dimension Affiliation

Code	Désignation	Type	Formule d'extraction
ID_VILLE	ID de la ville	Entier	SELECT id_ville FROM ville
NOM_VILLE	Nom de la ville	Texte	SELECT nom_ville FROM ville

TABLE 6 – Dimension Ville

Code	Désignation	Type	Formule d'extraction
ID_PAYS	ID du pays	Entier	SELECT id_pays FROM pays
NOM_PAYS	Nom du pays	Texte	SELECT nom_pays FROM pays

TABLE 7 – Dimension Pays

2 Conception

2.1 Choix du Modèle d'Entrepôt

Le modèle choisi est un modèle en flocon

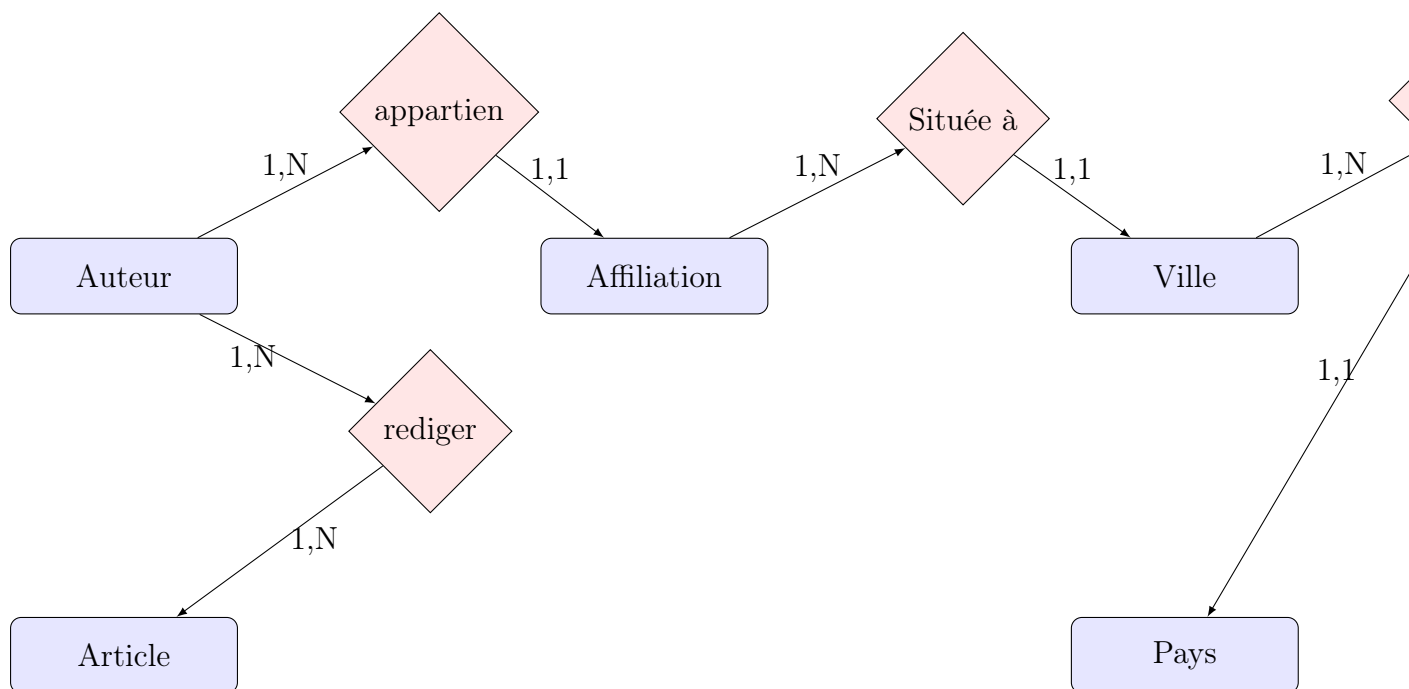


FIGURE 1 – Modèle Conceptuel des Données (MCD)

3 Implementation : approche ROLAP

3.1 Choix du SGBD

Le système de gestion de base de données (SGBD) retenu pour ce projet est **PostgreSQL**. PostgreSQL est un SGBD relationnel open source, robuste et adapté aux entrepôts de données.

3.2 Base de Données Opérationnelle

La base de données opérationnelle, nommée **bd_art**, contient les données brutes des publications, auteurs, affiliations, etc. Voici un exemple de commande SQL pour créer cette base de données :

```
CREATE DATABASE bd_art;
```

3.3 Entrepôt de Données

L'entrepôt de données, nommé **bd_final**, est conçu pour stocker les données transformées et structurées selon le modèle en étoile. Voici un exemple de commande SQL pour créer cette base de données :

```
CREATE DATABASE bd_final;
```

4 Alimentation

4.1 Processus ETL

Le processus ETL (Extract, Transform, Load) est mis en place pour alimenter l'entrepôt de données à partir de la base de données opérationnelle. Voici les étapes principales :

- **Extraction** : Récupération des données brutes depuis **bd_art**.
- **Transformation** : Nettoyage, filtrage et agrégation des données.
- **Chargement** : Insertion des données transformées dans **bd_final**.

4.2 Utilisation de Talend OS-DI

Talend OS-DI est utilisé pour créer des jobs ETL. Les composants principaux incluent :

- **tDBInput** : Pour extraire les données de **bd_art**.
- **tMap** : Pour transformer les données selon les besoins de **bd_final**.
- **tDBOutput** : Pour charger les données dans **bd_final**.

4.3 Généralités

Voici de manière générale comment notre projet réalisé avec Talend se présente :

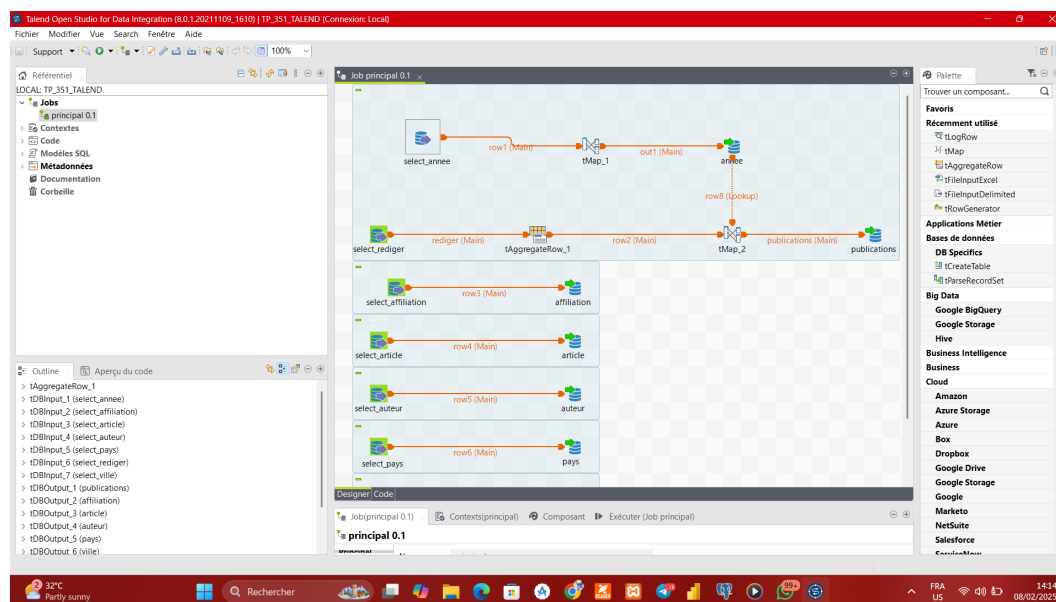


FIGURE 2 – Présentation générale du projet Talend

4.4 Requêtes et Tables de la Base de Données Opérationnelle

Dans cette section, nous pouvons voir les différentes requêtes effectuées sur la base de données opérationnelle (**bd_art**) ainsi que les différentes tables qui la composent.

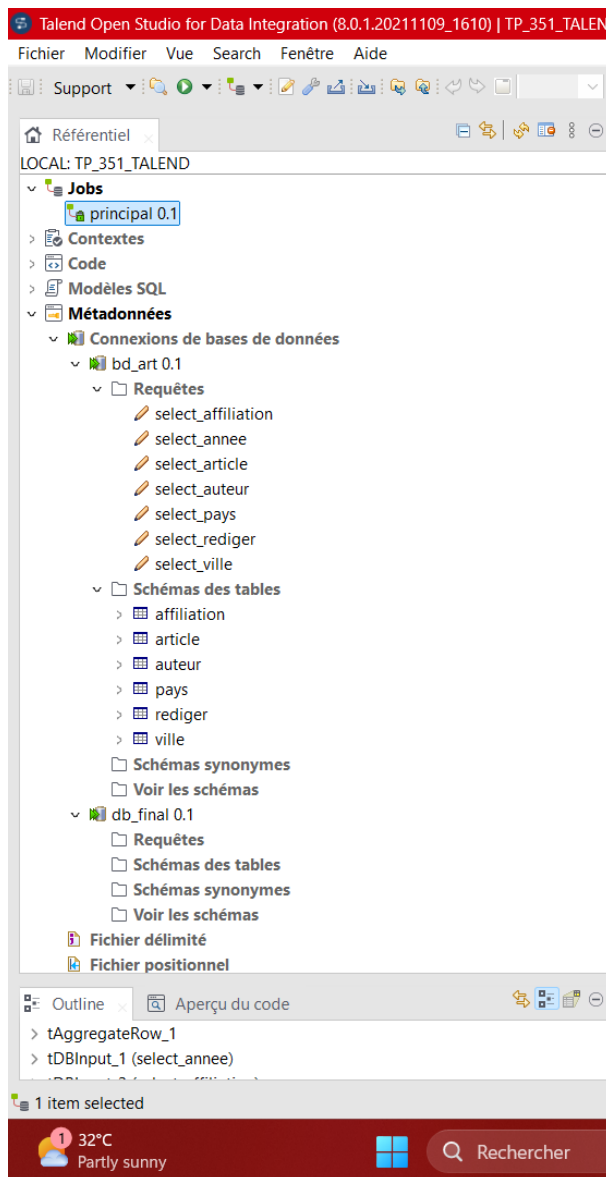


FIGURE 3 – Requêtes et tables de la base de données opérationnelle

4.5 Configuration de la Table Affiliation

Pour la table `affiliation` de notre entrepôt de données (`bd_final`), nous pouvons observer les différentes configurations effectuées.

4.6 Création de la Table Année

Pour créer la table `annee`, nous récupérerons les données de la table `rediriger` en utilisant le composant `tMap`.

4.7 Agrégation des Données avec `tAggregateRow`

Grâce au composant `tAggregateRow`, nous regroupons les données en fonction des auteurs et des années, puis nous comptons le nombre de publications. Cela nous permet

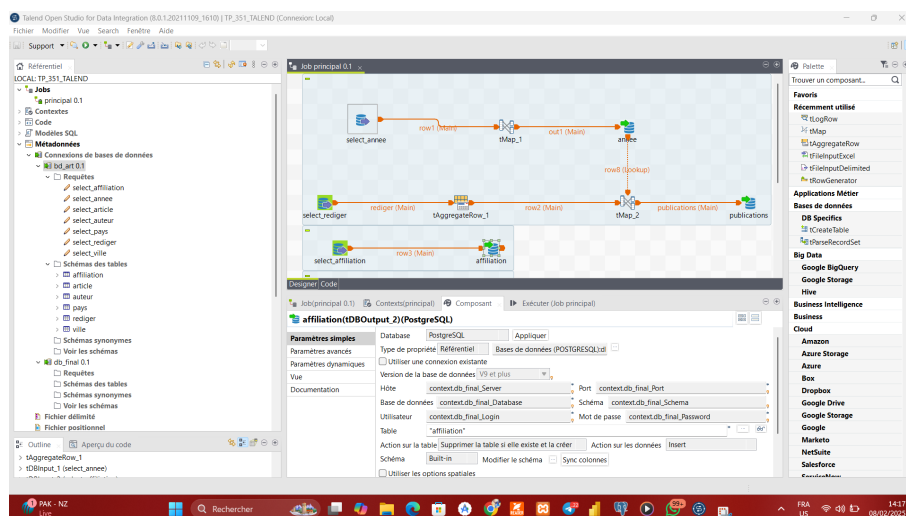


FIGURE 4 – Configuration de la table Affiliation

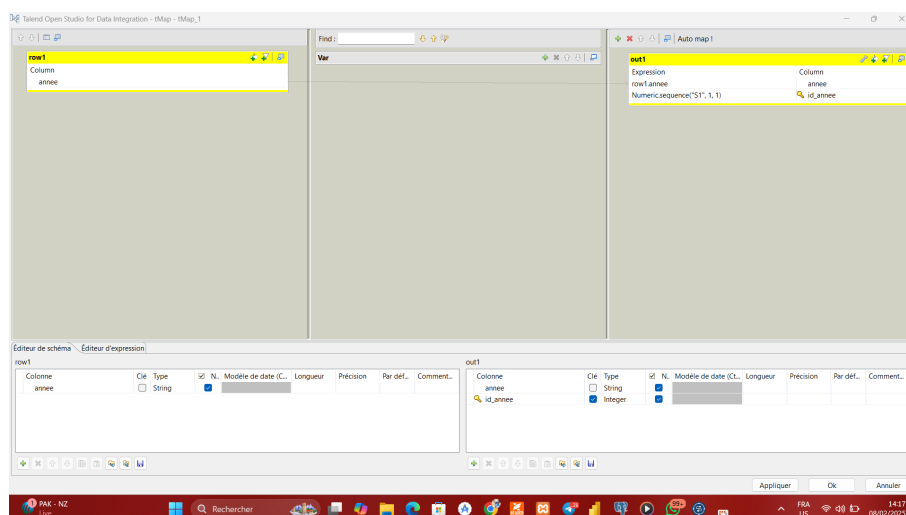


FIGURE 5 – Création de la table Année avec tMap

d'obtenir notre mesure `nb_publication`.

4.8 Création de la Table de Fait

Enfin, nous créons notre table de fait qui va contenir les différents ID (clés étrangères) et une mesure appelée `nombre_publication`.

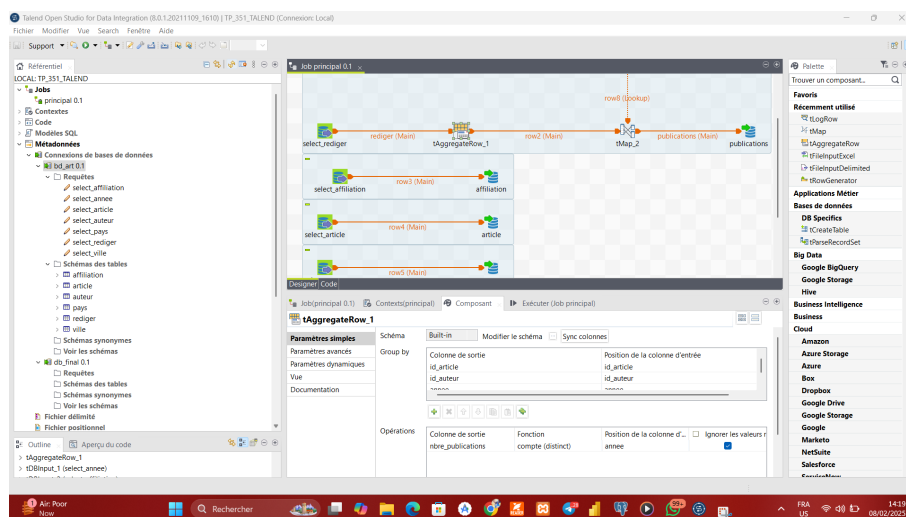


FIGURE 6 – Agrégation des données avec tAggregateRow

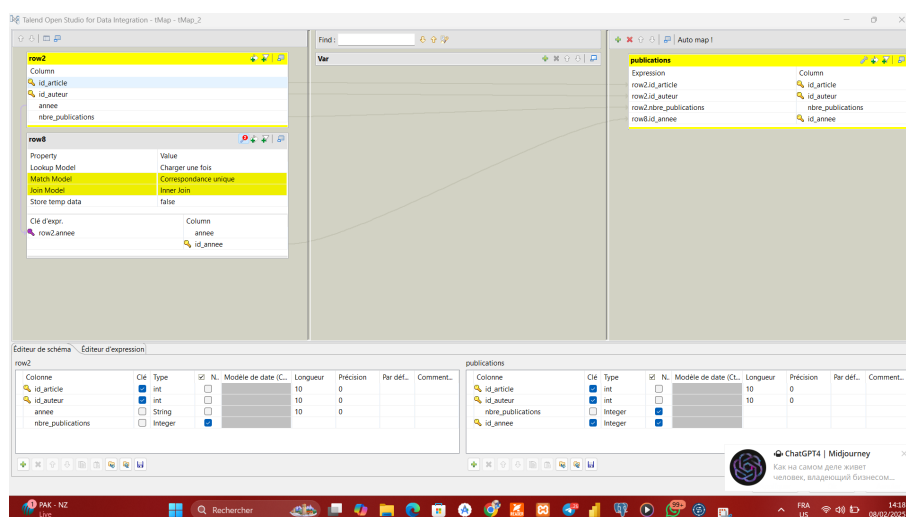


FIGURE 7 – Création de la table de fait