

Base de données
Multidimensionnelle : Etude
de cas



Année académique : 2023-2024

Membres du Groupe :

- Samuel S.R.C. AÏSSI
- Tobî-Logo J.M. Marc A. ALE
- Tèkiyath A. F. E AMOUSSA
- Rimbé G. S. KAFANDO
- Boladé A.M. MAMADOU

Professeur :

Dr BESSANE

Table des matières

I. Analyse Décisionnelle.....	3
A. Identification des Problèmes.....	3
B. Propositions de Solutions.....	3
C. Identification des Faits, Mesures, Dimensions et Hiérarchies.....	4
II. Modélisation Multidimensionnelle et modèle conceptuel de données	9
A. Modèles multidimensionnels (en étoile ou en flocon)	9
B. Modèle entité- relationnel (principale source de données)	12
C. Utilisation de « Data Generate » pour créer un jeu de données pour la base de données source	13
III. Intégration des données sur Talend Data Intégration	15
A. Connection de la base de données source et l'entrepôts de données avec Talend	15
B. Application du processus ETL pour alimenter la base de données décisionnelle Processus ETL :	17
C. Étapes pour calculer chaque mesure à partir de la base de données source	30
Conclusion.....	32

I. Analyse Décisionnelle

A. Identification des Problèmes

1. Défaillance des Systèmes de Test

- Insuffisance des kits de test.
- Inaccessibilité des centres de test dans certaines régions

2. Problèmes d'Approvisionnement en Vaccins

- Pénurie de vaccins
- Multitude de vaccins

3. Manque de Personnel Médical

- Insuffisance de personnel pour gérer les cas COVID-19
- Problèmes de formation et de compétence

4. Problèmes de Communication

- Informations erronées ou insuffisantes par rapport mesures sanitaires
- Difficultés à atteindre certaines zones géographiques

5. Systèmes de Gestion de l'Information

- Difficulté à centraliser et partager les données COVID-19
- Problèmes de mise à jour des données en temps réel

6. Problèmes d'Infrastructure

- Manque de lits en réanimation
- Infrastructures de santé insuffisantes

B. Propositions de Solutions

1. Renforcement des Systèmes de Test

- Améliorer l'accès aux centres de test dans toutes les régions
- Investir dans de nouveaux équipements de test

2. Amélioration de l'Approvisionnement et de la Distribution des Vaccins

- Optimiser les chaînes d'approvisionnement

- Établir un plan de distribution équitable

3. Recrutement et Formation du Personnel Médical

- Recruter du personnel médical supplémentaire
- Former le personnel sur les nouvelles procédures COVID-19

4. Renforcement de la Communication

- Établir des canaux de communication clairs et efficaces
- Améliorer l'accès à l'information dans les zones rurales
- Utiliser les médias locaux et les réseaux sociaux pour diffuser des informations précises et à jour
- Créer des campagnes de sensibilisation ciblées pour les différentes régions et populations

5. Centralisation et Amélioration des Systèmes de Gestion de l'Information

- Développer une plateforme centralisée pour les données COVID-19
- Assurer la mise à jour et la vérification des données en temps réel

6. Amélioration des Infrastructures de Santé

- Investir dans la construction de nouvelles infrastructures
- Améliorer les équipements existants

C. Identification des Faits, Mesures, Dimensions et Hiérarchies

1. Faits

- Suivi_des_patients
- Suivi_vaccination
- Suivi_Tests
- Suivi_Matériels
- Suivi_Personnel_Médical
- Suivi_Taux_satisfaction

2. Mesures

- **Suivi_des_patients**
 - Temps d'hospitalisation
 - Nombre de cas confirmés

- Nombre de patients guéris
- Nombre de patients décédés
- Nombre de patients confinés
- **Suivi_vaccination**
 - Nombre de vaccins administrés
 - Nombre de personnes vaccinées
- **Suivi_Tests**
 - Nombre de tests effectués
 - Résultats des tests (positifs/négatifs)
- **Suivi_Matériels**
 - Nbr Matériels disponibles
 - Nbr Matériels défectueux
 - Nbr Matériels réparés
 - Nbr Matériels perdus
 - Nbr Durée d'utilisation
- **Suivi_Personnel_Médical**
 - Nombre d'absence
 - Nbr retards
- **Taux satisfaction**
 - Nbr de satisfaction

3. Dimensions

- Temps
- Zone
- Test
- Matériel
- Vaccin
- Personnel Médical
- Patient

- Centre Hospitalier

4. Hiérarchies

- **Temps**

- Jour
- Mois
- Saison
- Année

- **Zone**

- Pays
- Département
- Commune
- Région

- **Test**

- **Nom :**
 - **PCR (Polymerase Chain Réaction)**
 - Détection de l'ARN viral
 - Considéré comme le test de référence (gold standard)
 - Haute précision et sensibilité
 - Prélèvement nasopharyngé, oropharyngé ou salivaire
 - **Test Antigénique Rapide**
 - Détection des protéines virales (antigènes)
 - Résultats rapides (environ 15-30 minutes)
 - Moins sensible que le PCR, mais utile pour le dépistage rapide
 - Prélèvement nasopharyngé ou nasal
 - **Test Sérologique (Anticorps)**
 - Détection des anticorps (IgM, IgG) produits en réponse à l'infection
 - Utilisé pour déterminer une infection passée ou la réponse immunitaire après vaccination
 - Prélèvement de sang (échantillon sanguin)
 - **Test de Salive**
 - Alternative aux tests nasopharyngés pour PCR ou antigénique
 - Moins invasif
 - Détection de l'ARN viral dans la salive
 - **Test LAMP (Loop-mediated Isothermal Amplification)**

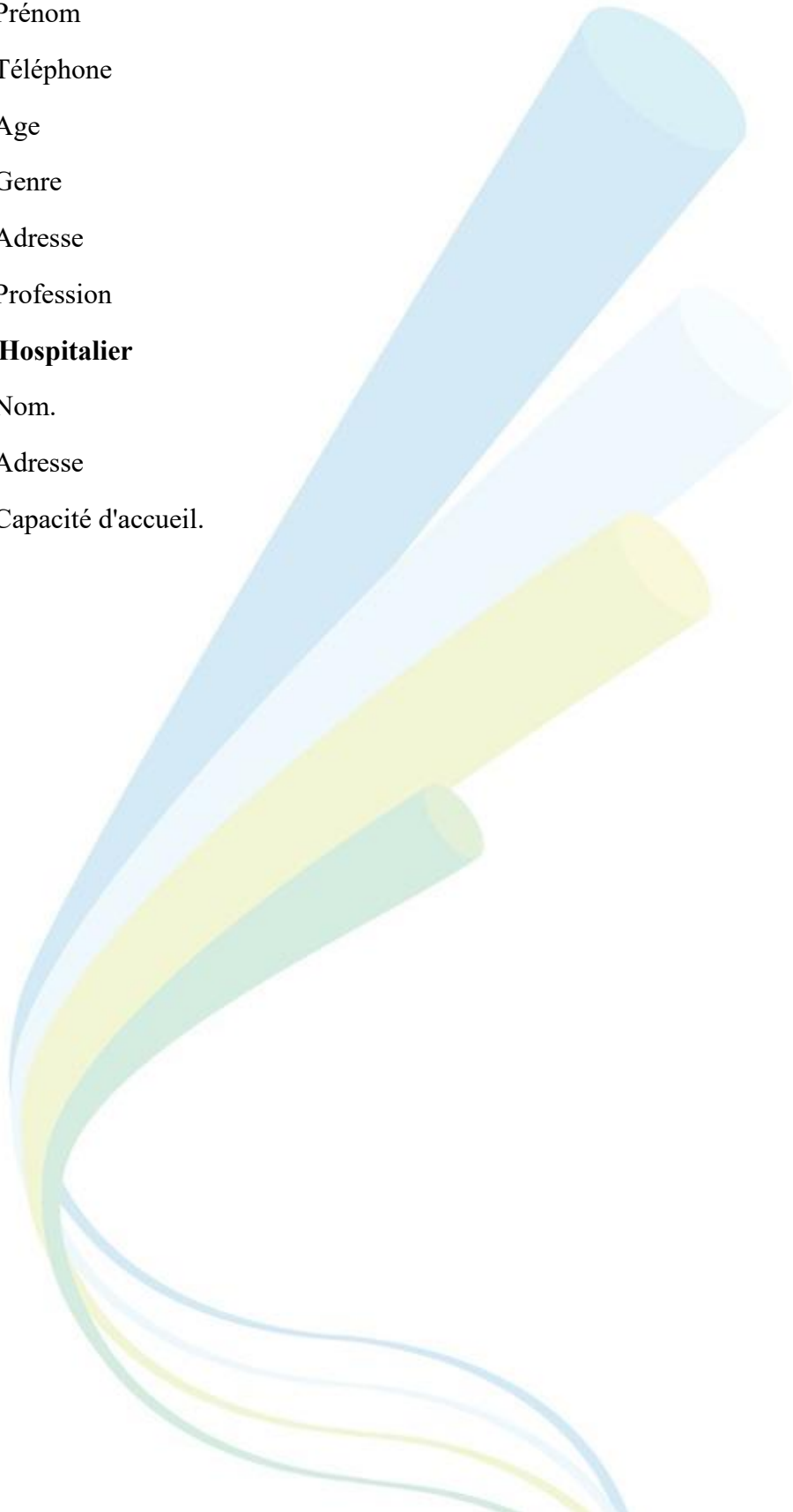
- Amplification de l'ARN viral à une température constante
 - Rapide et peut être utilisé dans des environnements à faibles ressources
 - Prélèvement nasopharyngé ou salivaire
- **Test RT-LAMP (Reverse Transcription LAMP)**
 - Combine la transcription inverse et l'amplification isotherme
 - Rapide et précis
 - Prélèvement nasopharyngé ou salivaire
- **Test TMA (Transcription-Mediated Amplification)**
 - Amplification de l'ARN viral
 - Haute sensibilité similaire au PCR
 - Prélèvement nasopharyngé ou salivaire
- **Test de Fluorescence**
 - Utilisé pour détecter les antigènes viraux avec un signal fluorescent
 - Rapide et précis
 - Prélèvement nasopharyngé ou nasal
- **Type de Vaccin**
 - Nom du vaccin
 - Fabricant
- **Catégorie de Cas**
 - Cas Confirmés
 - Cas Guéris
 - Cas Décédés
- **Matériel**
 - Type de matériel
 - État du matériel
 - Durée d'utilisation
- **Personnel Médical**
 - Nom
 - Prénom
 - Spécialité
 - Zone d'intervention
 - Genre
 - Adresse

- **Patient**

- Nom
- Prénom
- Téléphone
- Age
- Genre
- Adresse
- Profession

- **Centre Hospitalier**

- Nom.
- Adresse
- Capacité d'accueil.



II. Modélisation Multidimensionnelle et modèle conceptuel de données

A. Modèles multidimensionnels (en étoile ou en flocon)

Ci-dessous nous avons les images illustrant les modèles multidimensionnels que nous avons créé en nous basant sur notre analyse décisionnelle et grâce à l'outil looping.

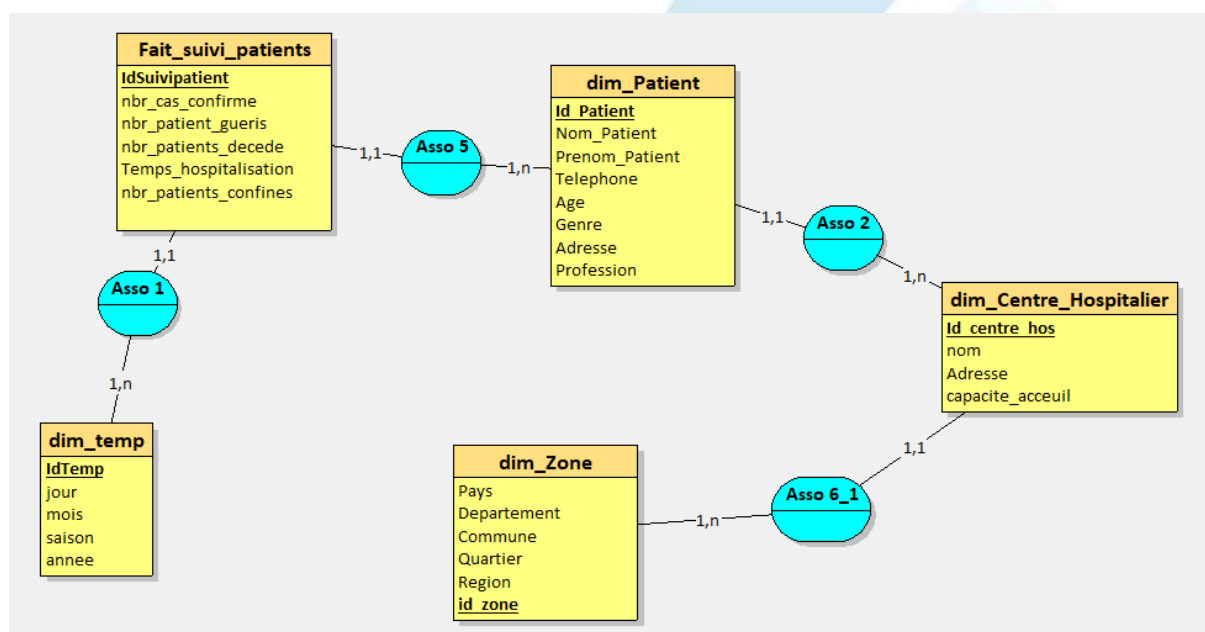


Fig. : Suivi_des_patients

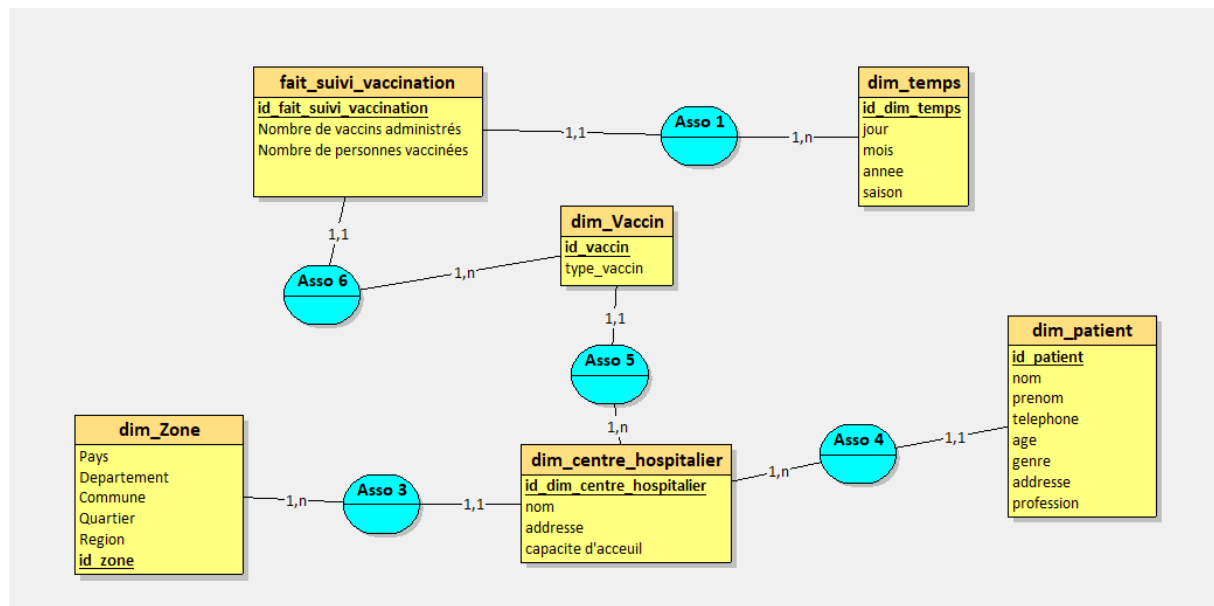


Fig. : Suivi_vaccination

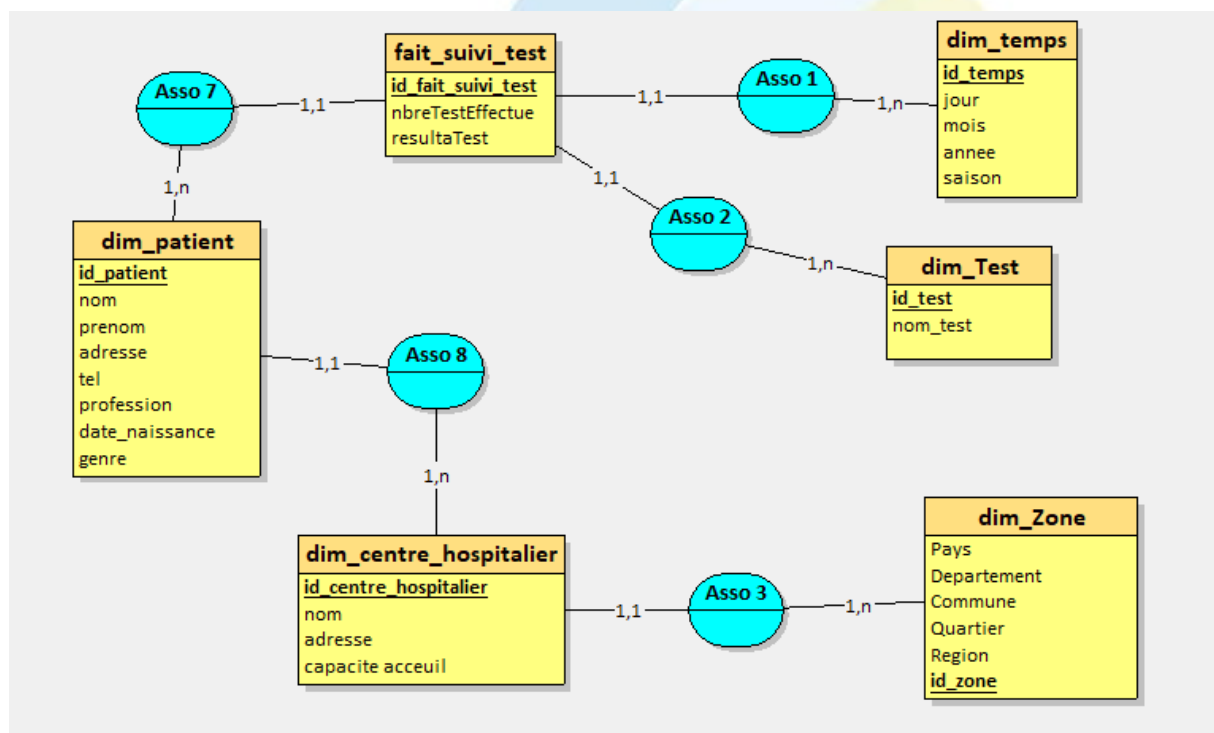


Fig. : Suivi_Tests

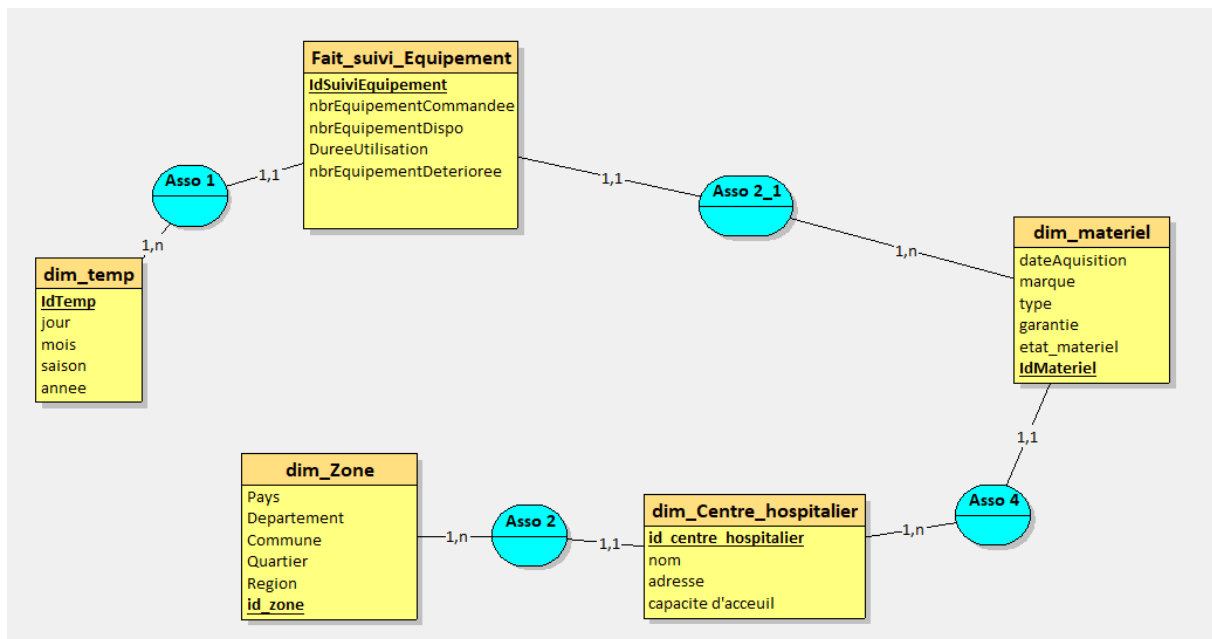


Fig. : Suivi_Matériels

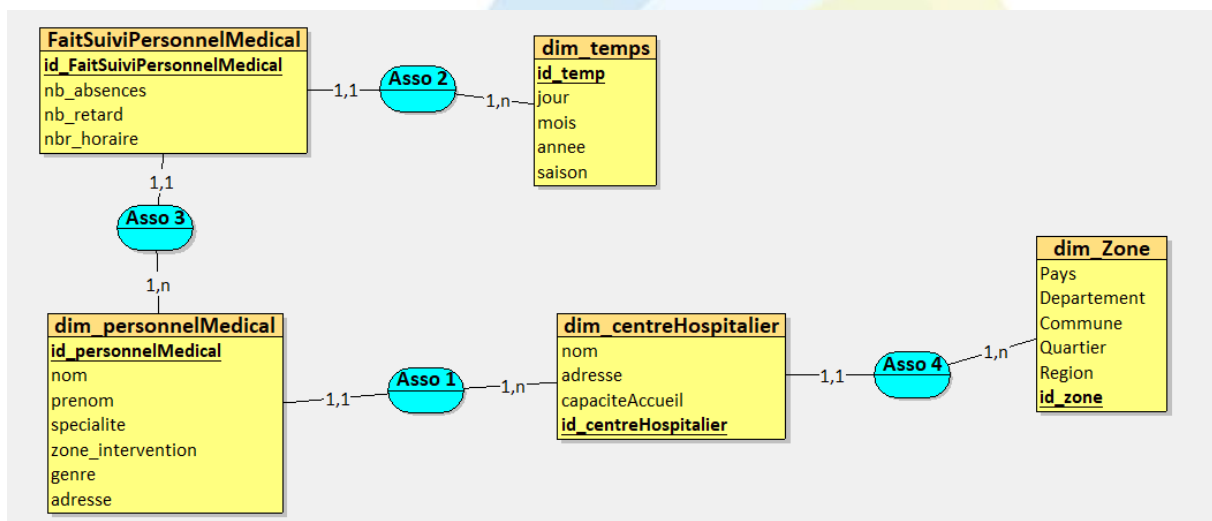


Fig. : Suivi_Personnel_Médical

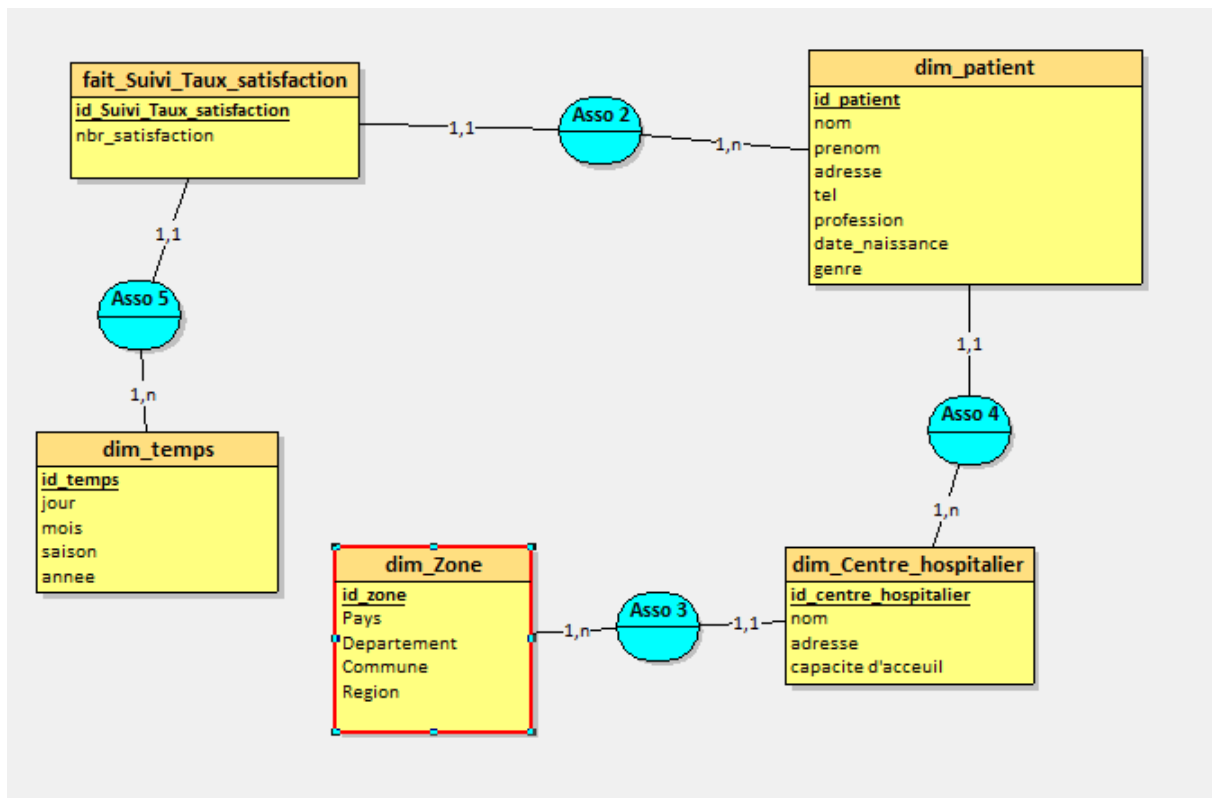
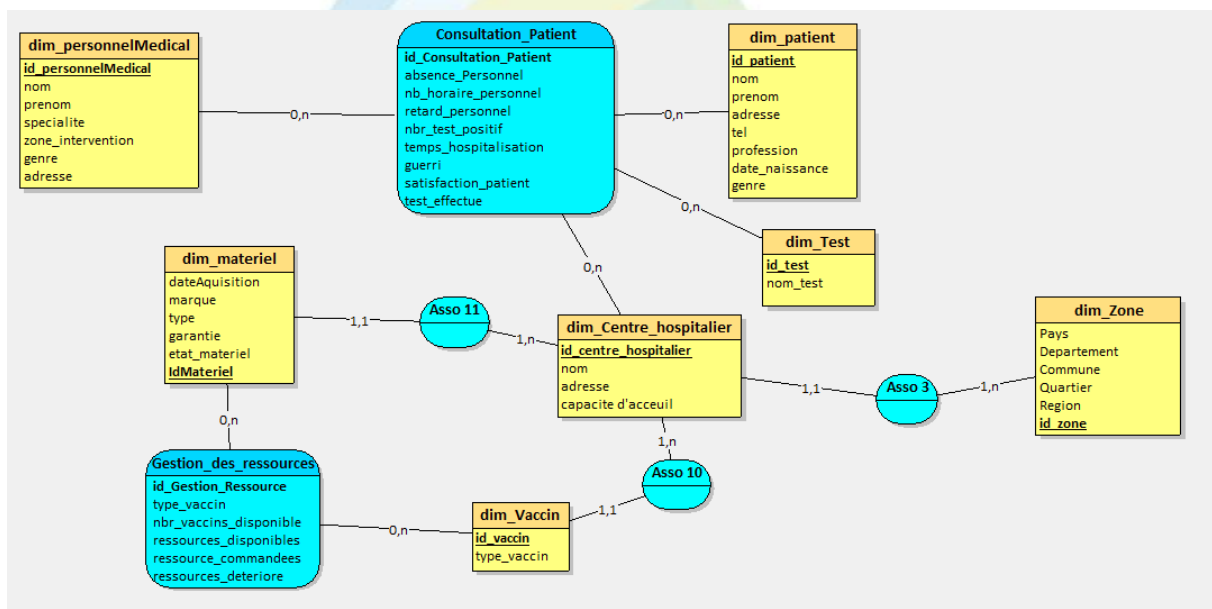
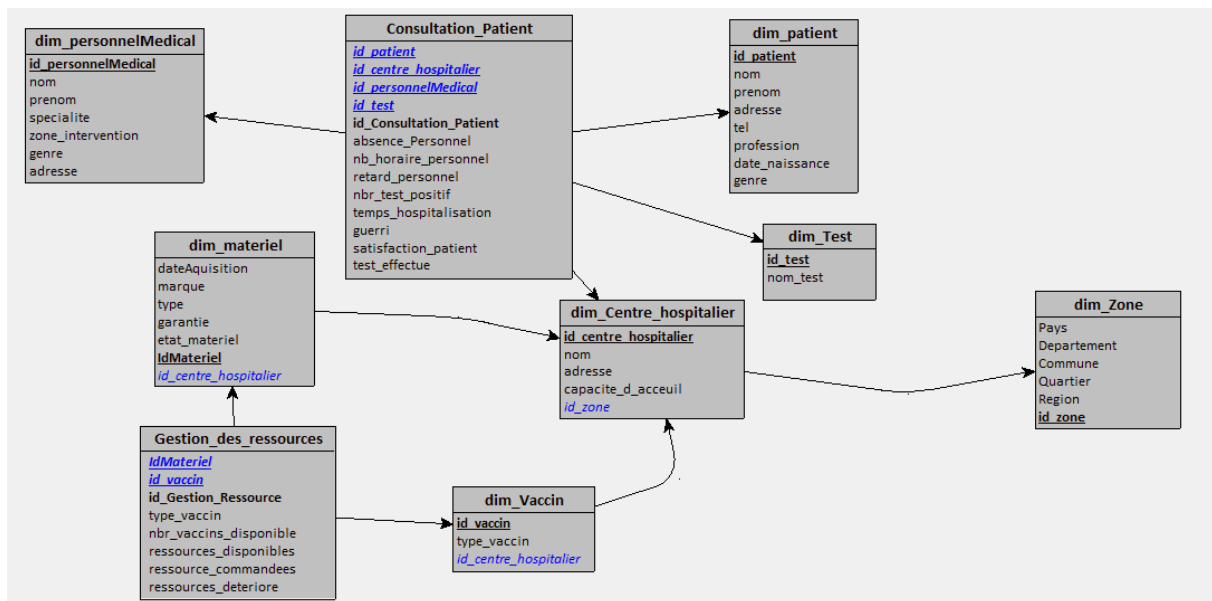


Fig. : Suivi_Taux_satisfaction

B. Modèle entité- relationnel (principale source de données)





C. Utilisation de « Data Generate » pour créer un jeu de données pour la base de données source

Field Name	Type	Options
id	Row Number	blank: 0 % Σ ×
first_name	First Name	blank: 0 % Σ ×
last_name	Last Name	blank: 0 % Σ ×
adresse	Email Address	blank: 0 % Σ ×
date_naissance	Datetime	02/07/1955 to 06/11/2022 format: SQL datetime blank: 0 % Σ ×
profession	Job Title	blank: 0 % Σ ×
genre	Gender	blank: 0 % Σ ×
telephone	Phone	format: ###-###-#### blank: 0 % Σ ×

+ ADD ANOTHER FIELD GENERATE FIELDS USING AI...

Rows: 1000 Format: SQL Table Name: Patients ☒ Include CREATE TABLE

GENERATE DATA PREVIEW SAVE AS... DERIVE FROM EXAMPLE... MORE

Importation des jeux de données créés avec « Data Generate » dans la base de données « source_covid » sur « PhpMyAdmin ».

phpMyAdmin

Seigneur : MySQL:3306 > Base de données : source_covid

Structure SQL Rechercher Requête Exporter Importer Opérations Privileges Procédures stockées Événements Plus

Filtres

Contenant le mot :

Table	Action	Lignes	Type	Interclassement	Taille	Perte
<input type="checkbox"/> centre_hospitalier_final	★ Parcourir Structure Rechercher Insérer Vider Supprimer	100	InnoDB	utf8mb4_0900_ai_ci	16,0 kio	-
<input type="checkbox"/> consultation_patient	★ Parcourir Structure Rechercher Insérer Vider Supprimer	1 000	InnoDB	utf8mb4_0900_ai_ci	96,0 kio	-
<input type="checkbox"/> gestion_ressources	★ Parcourir Structure Rechercher Insérer Vider Supprimer	1 000	InnoDB	utf8mb4_0900_ai_ci	96,0 kio	-
<input type="checkbox"/> materiel_hospitalier	★ Parcourir Structure Rechercher Insérer Vider Supprimer	1 000	InnoDB	utf8mb4_0900_ai_ci	96,0 kio	-
<input type="checkbox"/> patient	★ Parcourir Structure Rechercher Insérer Vider Supprimer	1 000	InnoDB	utf8mb4_0900_ai_ci	128,0 kio	-
<input type="checkbox"/> personnel_medical	★ Parcourir Structure Rechercher Insérer Vider Supprimer	200	InnoDB	utf8mb4_0900_ai_ci	48,0 kio	-
<input type="checkbox"/> test	★ Parcourir Structure Rechercher Insérer Vider Supprimer	8	InnoDB	utf8mb4_0900_ai_ci	16,0 kio	-
<input type="checkbox"/> vaccins	★ Parcourir Structure Rechercher Insérer Vider Supprimer	11	InnoDB	utf8mb4_0900_ai_ci	16,0 kio	-
<input type="checkbox"/> zone	★ Parcourir Structure Rechercher Insérer Vider Supprimer	300	InnoDB	utf8mb4_0900_ai_ci	48,0 kio	-
9 tables	Somme	4 619	InnoDB	utf8mb4_0900_ai_ci	560,0 kio	0 o

☐ Tout cocher
 Avec la sélection :

III. Intégration des données sur Talend Data Intégration

A. Connection de la base de données source et l'entrepôts de données avec Talend

Connexion base de données

Mise à jour de la connexion à la base de données - Étape 1/2

Mettre à jour les propriétés

Nom: connexion_source_covid

Objectif:

Description:

Créé par: user@talend.com

Verrouillé par: user@talend.com

Version: 0.1 M m

Statut:

Chemin d'accès:

< Back Next > Finish Cancel

Connexion base de données

Mise à jour de la connexion à la base de données - Étape 2/2

Vous devez cliquer sur le bouton Vérifier pour vérifier les paramètres de la base de données

Type de BdDMySQL

Version de la base de donnéesMySQL 8

Chaîne de caractères de connexionjdbc:mysql://localhost:3306/source_covid?serverTimezone=UTC

Connexionroot

Mot de passeroo

Serveurlocalhost

Port3306

Base de donnéessource_covid

Paramètres supplémentairesserverTimezone=UTC

Tester la connexion

v

Exporter en tant que contexte

Revenir au contexte précédent

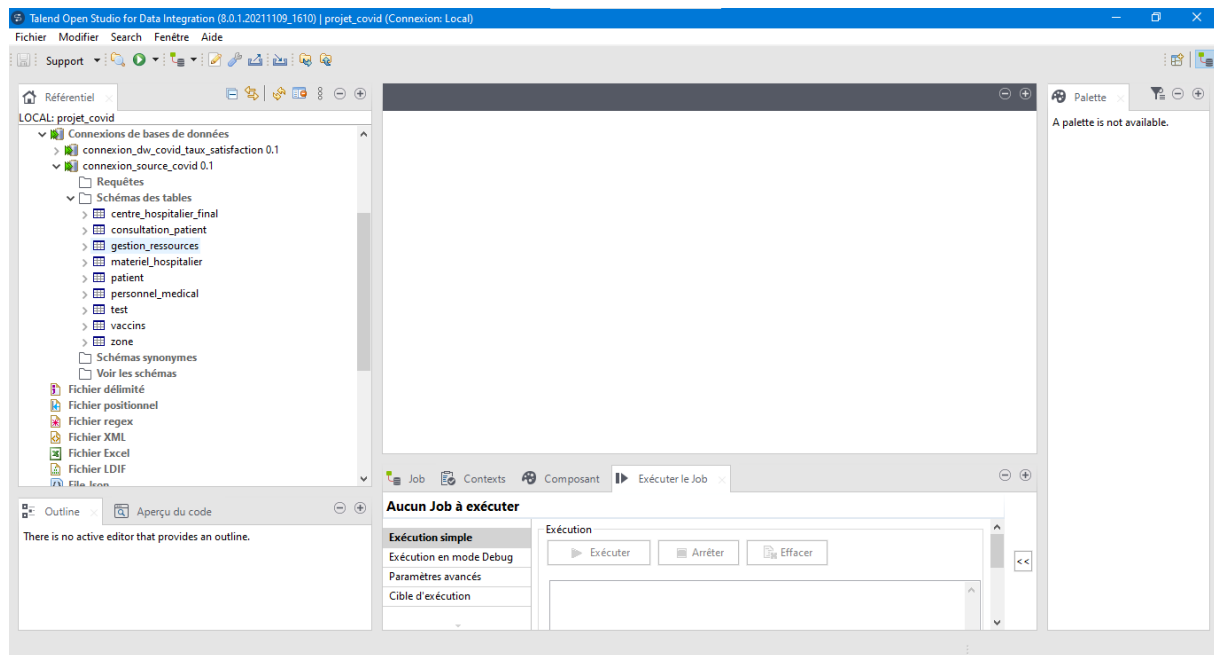
[Installer un pilote](#)

< Back

Next >

Finish

Cancel

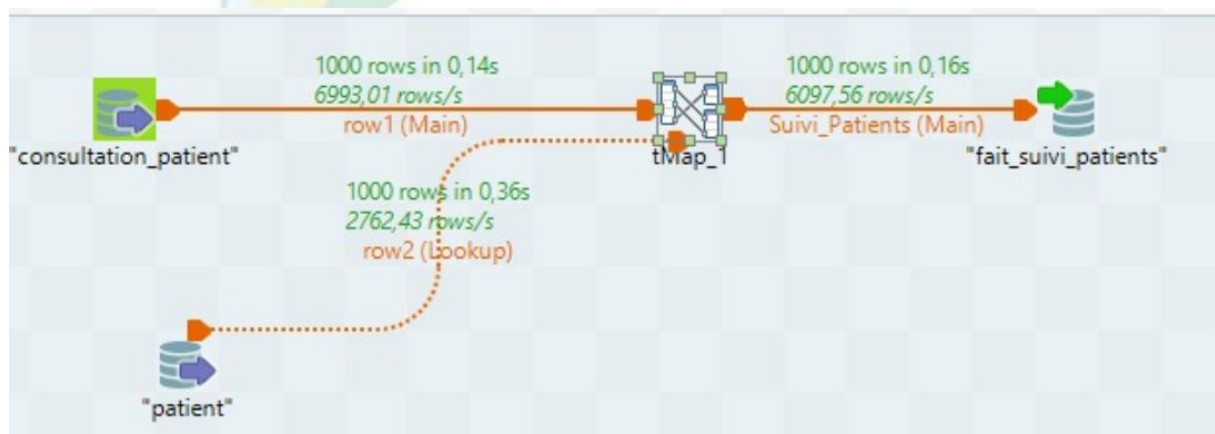
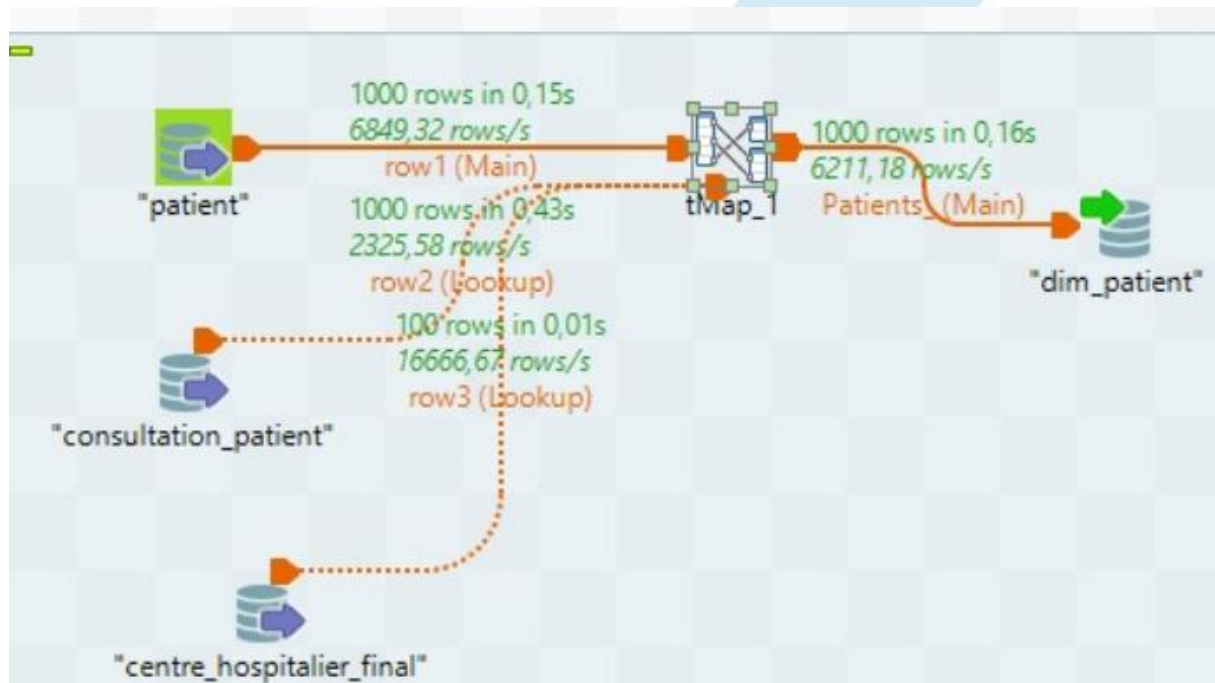
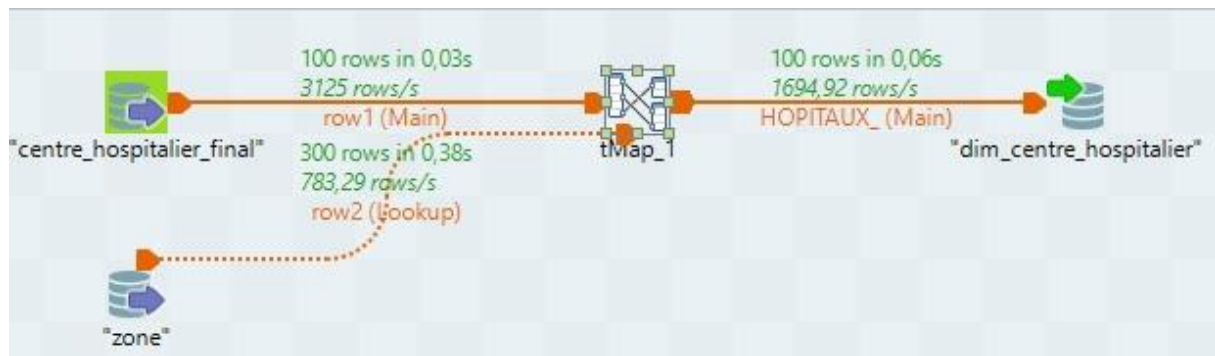


B. Application du processus ETL pour alimenter la base de données décisionnelle Processus ETL :

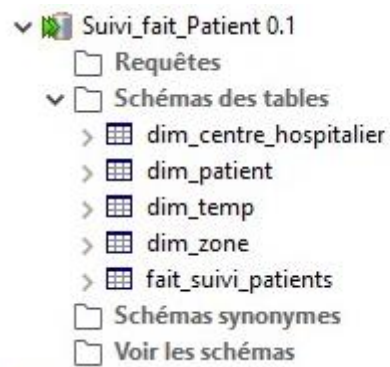
Cette section présente les étapes et les résultats de l'alimentation d'une base de données multidimensionnelle via un processus ETL (Extract, Transform, Load) réalisé avec Talend Data Intégration, illustré par des captures d'écran de la création et de l'exécution des workflows.

- Suivi_des_patients
 - Jobs
 - Jobs
 - Suivi_Pateint_alimenter_dim_Temps 0.1
 - Suivi_Pateint_alimenter_dim_Zone 0.1
 - Suivi_Pateint_alimenter_Fait_suivi_Patient 0.1
 - Suivi_Patient_alimenter_dim_Centre_hospitalier 0.1
 - Suivi_Patient_alimenter_dim_Patient 0.1
 - Contextes



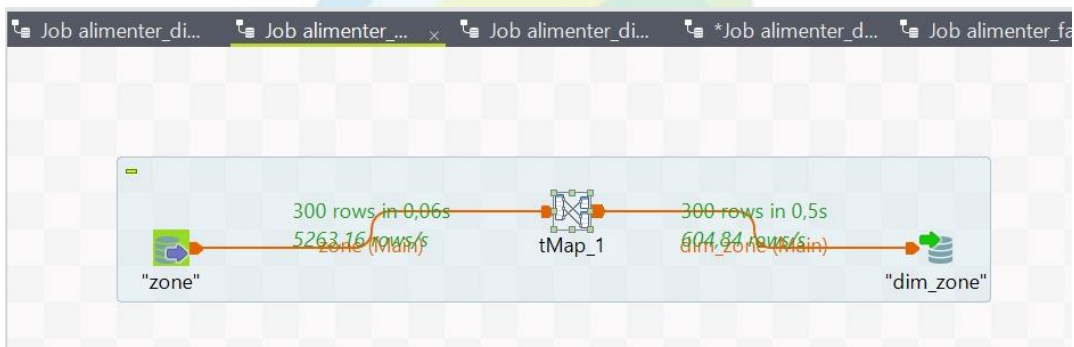
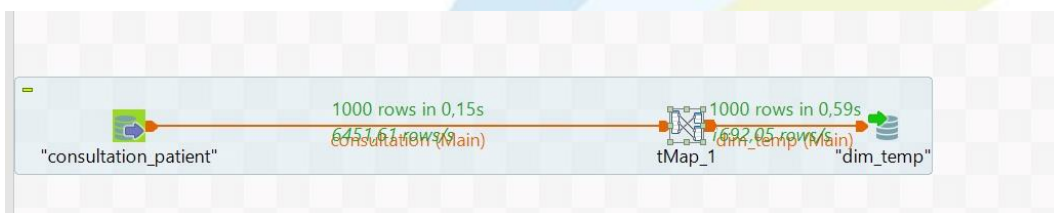
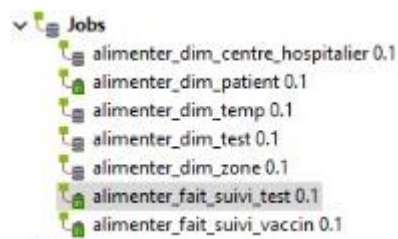


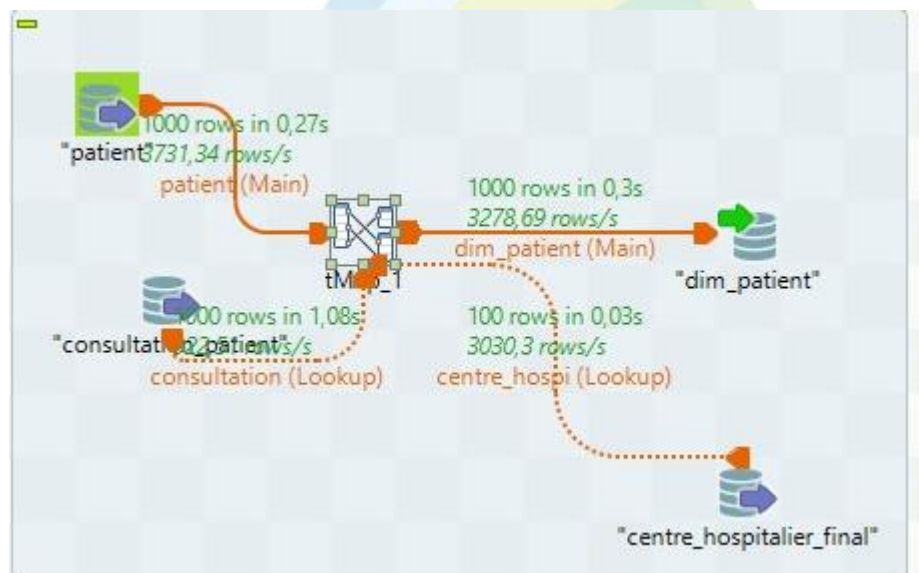
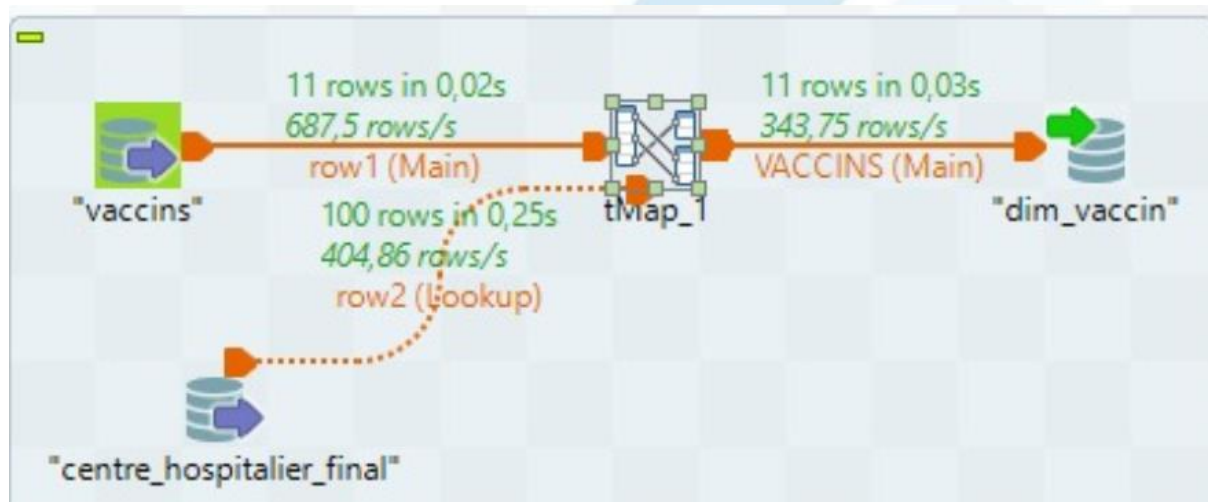
- Connexion à la base de données dw_covid_Suivi_des_patients

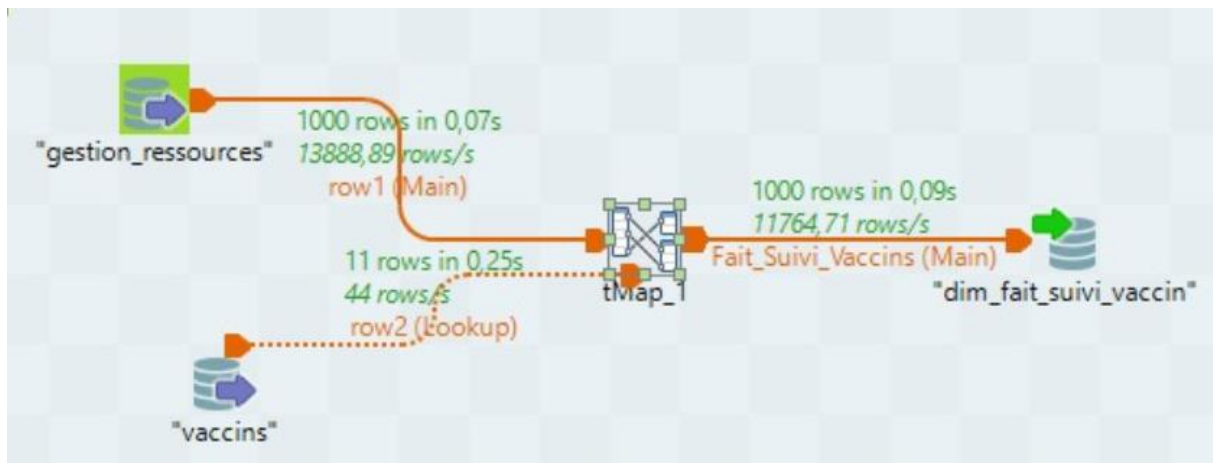


- Suivi_vaccination

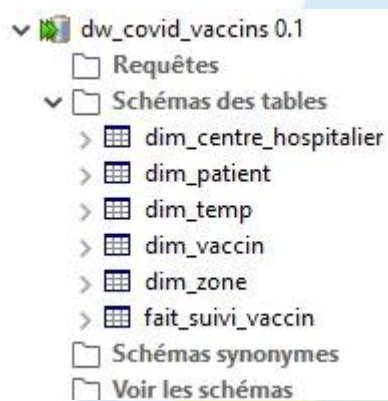
- Jobs





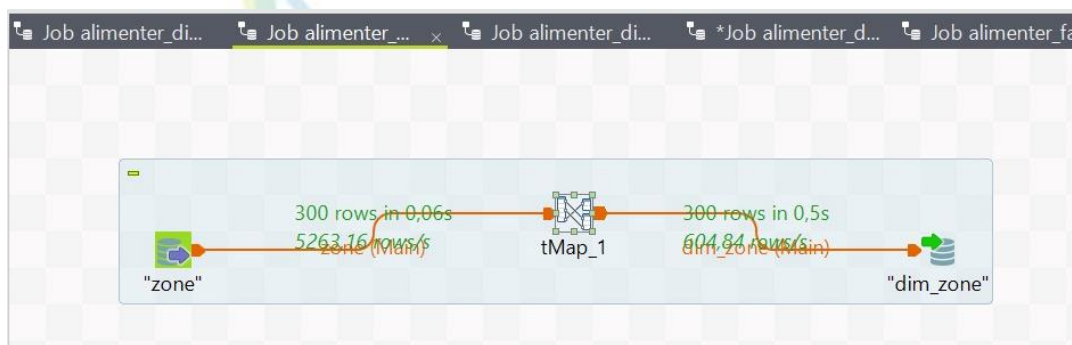
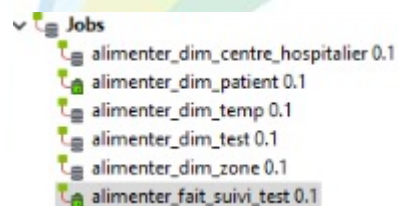


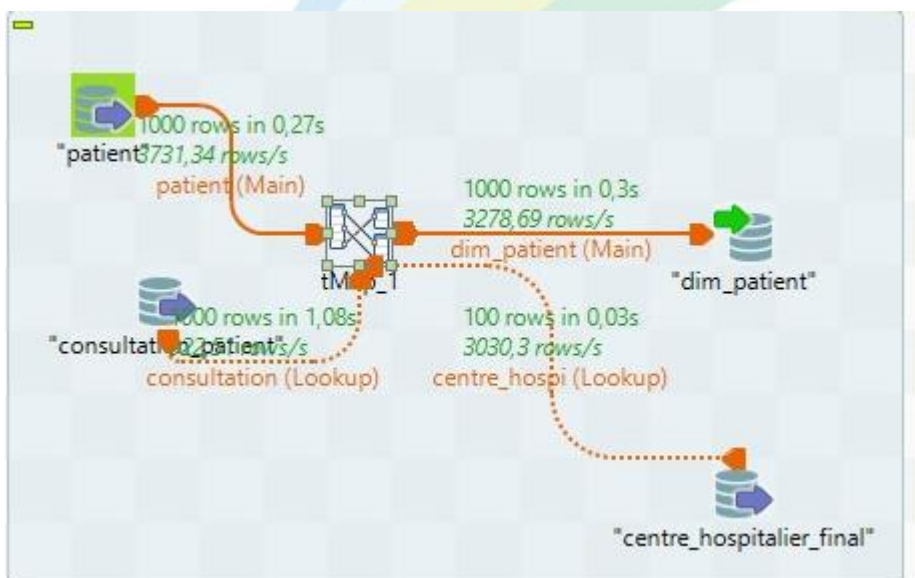
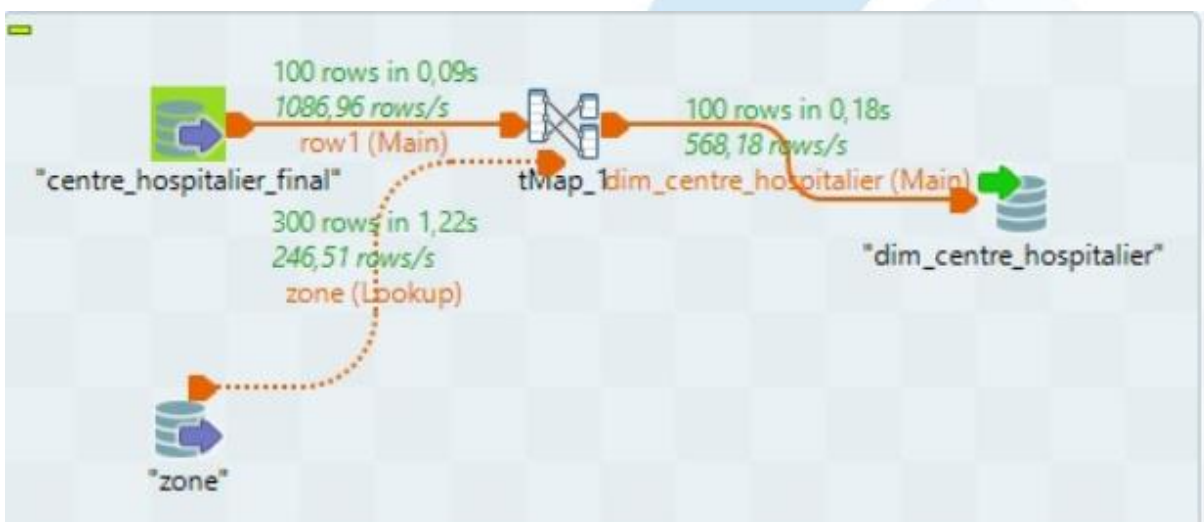
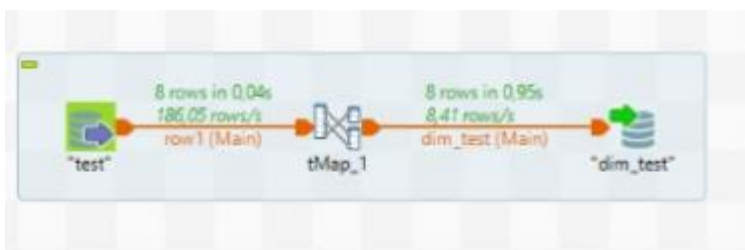
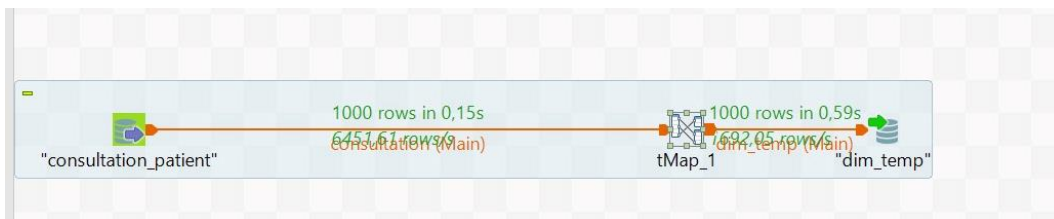
- Connexion à la base de données dw_covid_Suivi_vaccination

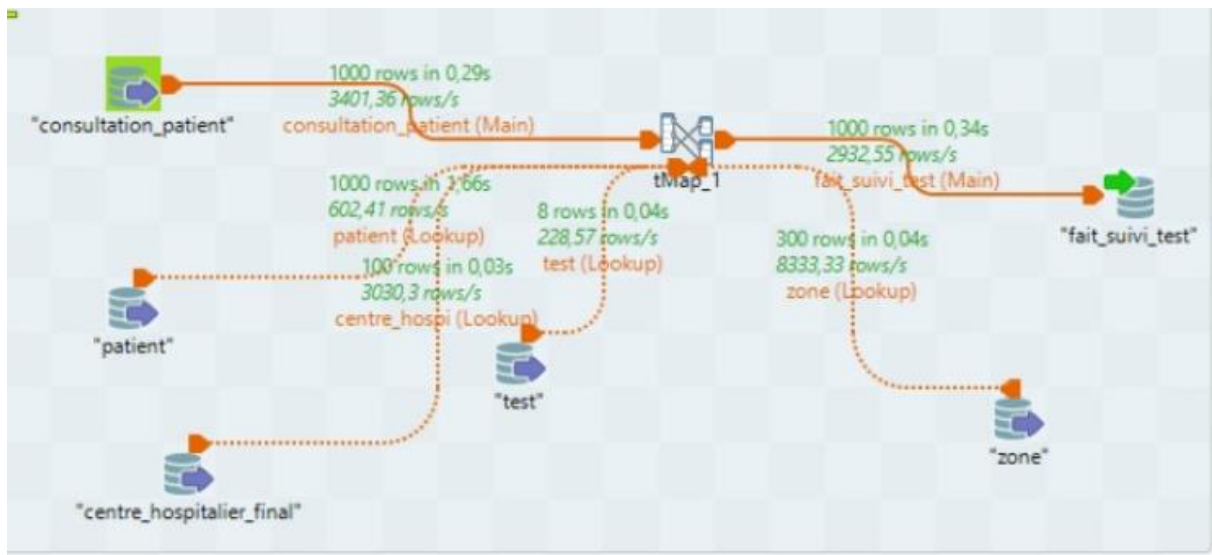


- Suivi_Tests

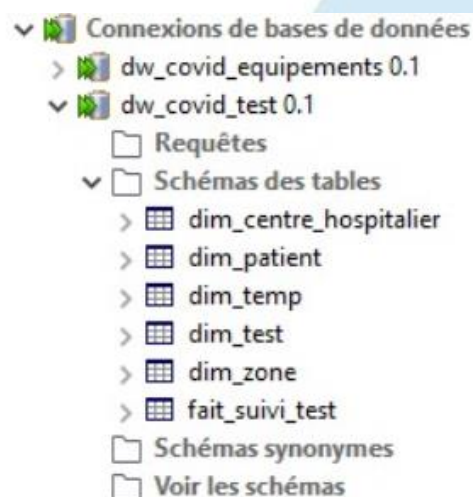
- Jobs





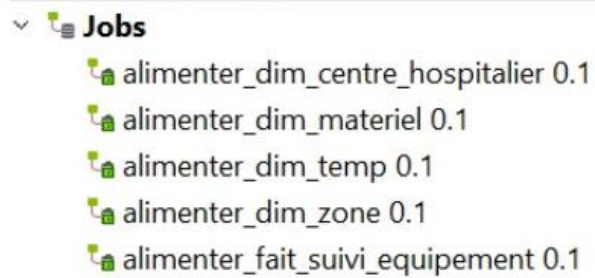


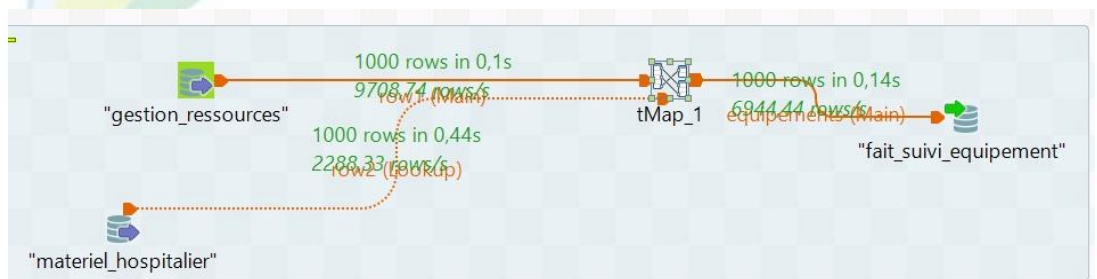
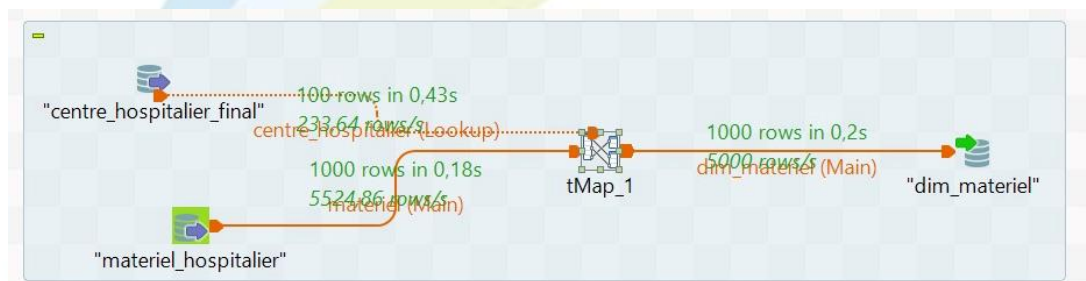
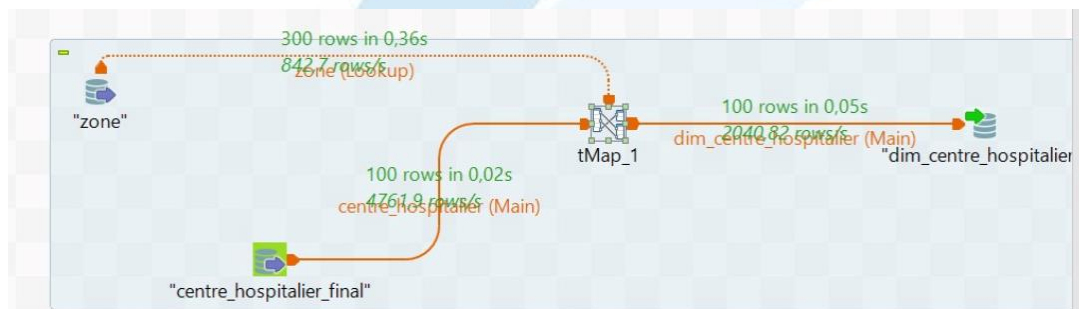
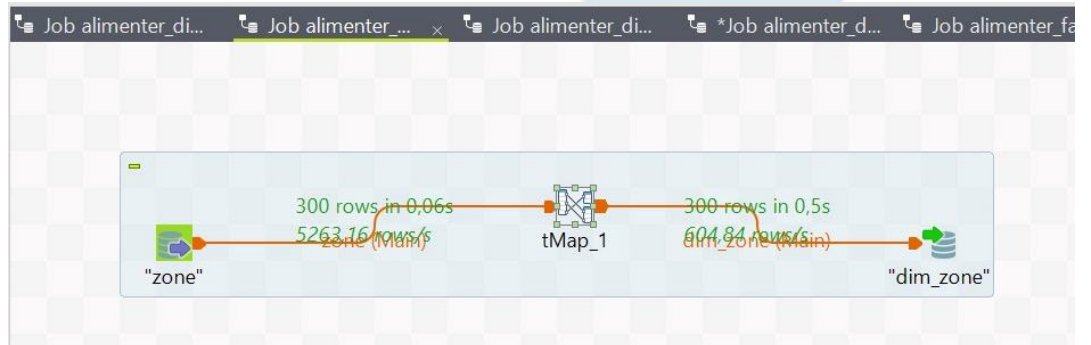
- Connexion à la base de données dw_covid_Suivi_Test



- Suivi_Matériels

- Jobs





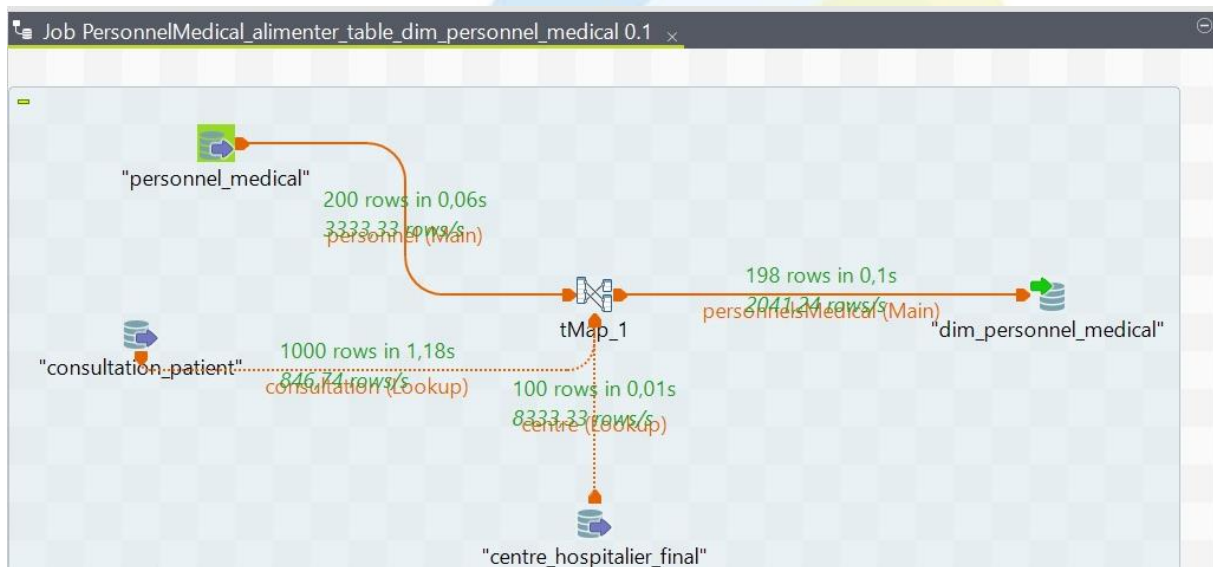
- Connexion à la base de données dw_covid_Suivi_Matériels

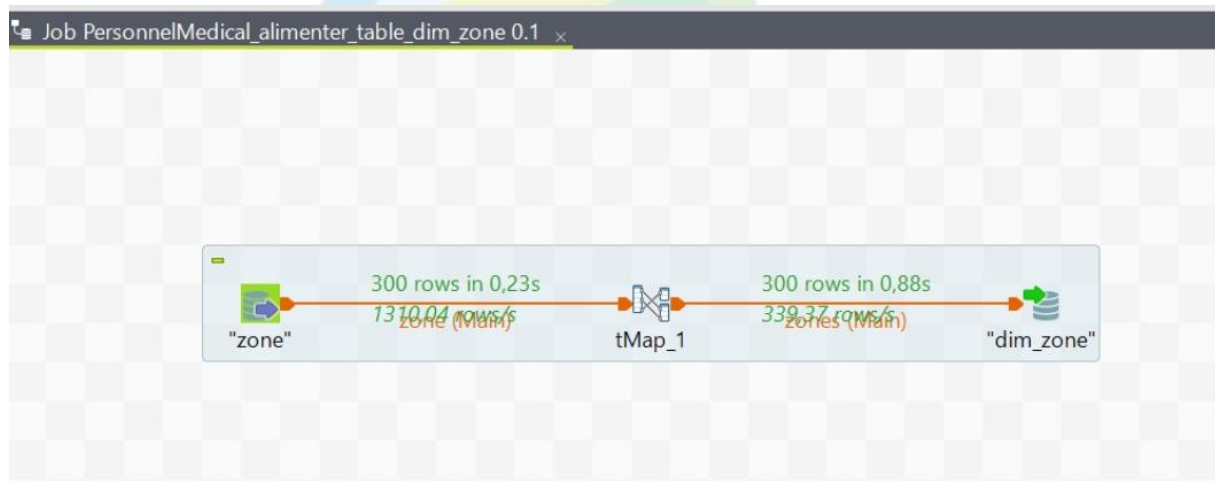
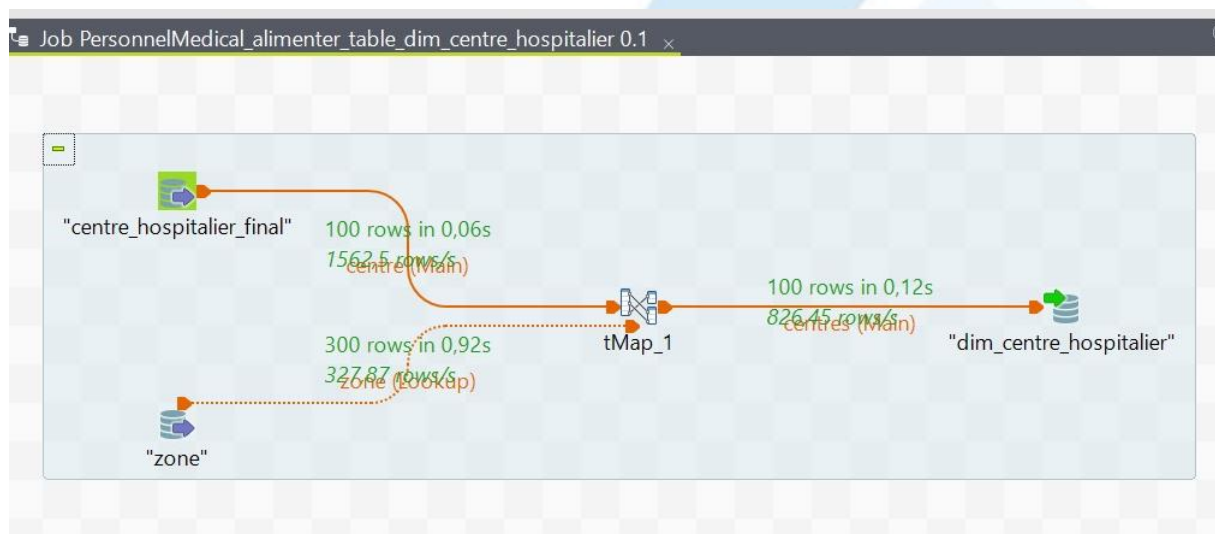
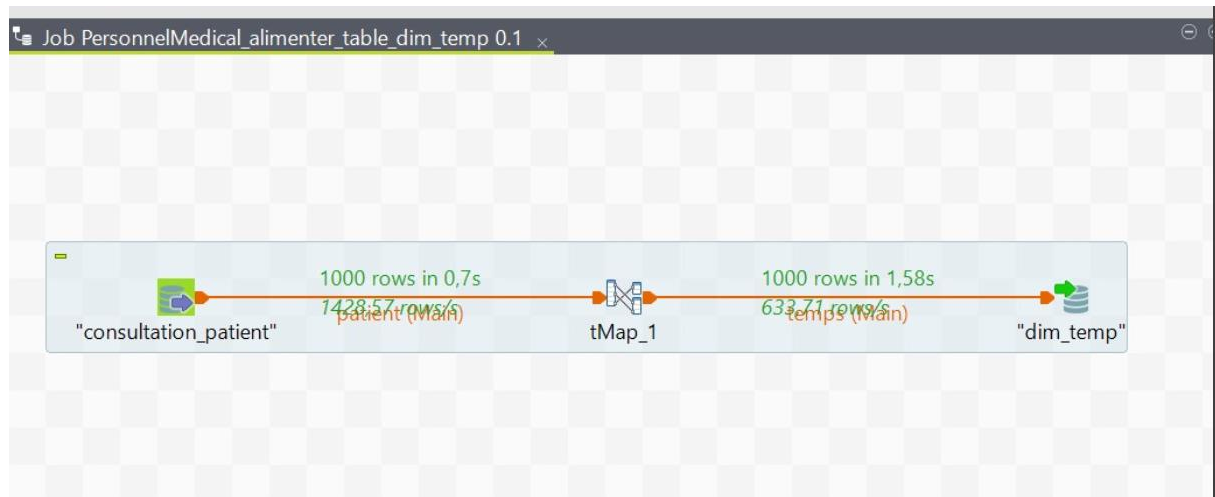
- Connexions de bases de données
 - dw_covid19_equipement 0.1
 - Requêtes
 - Schémas des tables
 - dim_centre_hospitalier
 - dim_materiel
 - dim_temp
 - dim_zone
 - fait_suivi_equipement

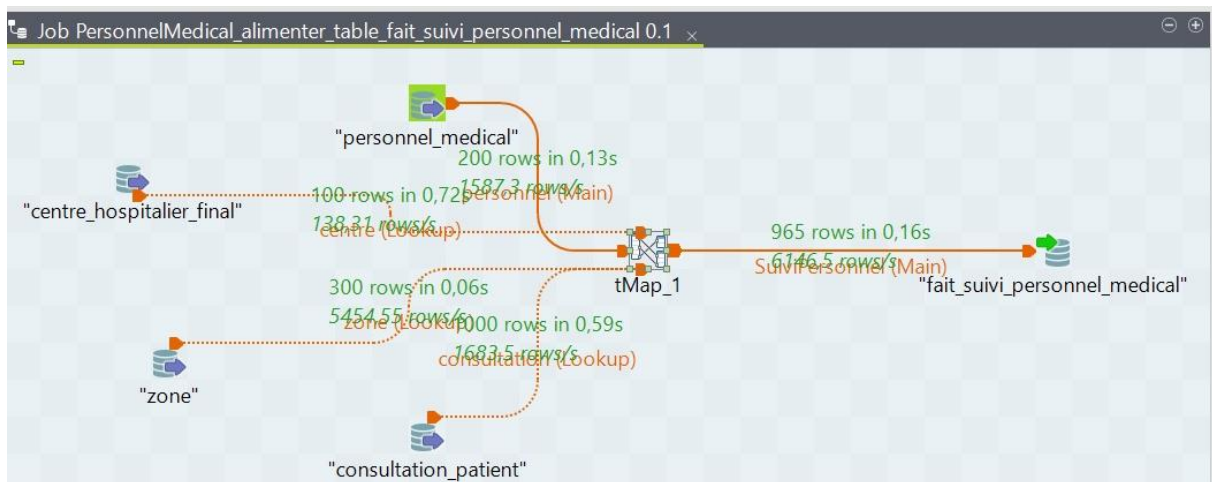
- Suivi_Personnel_Médical

- Jobs

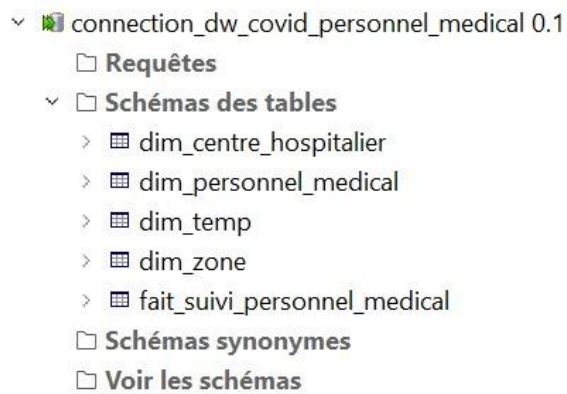
- Jobs
 - PersonnelMedical_alimenter_table_dim_centre_hospitalier 0.1
 - PersonnelMedical_alimenter_table_dim_personnel_medical 0.1
 - PersonnelMedical_alimenter_table_dim_temp 0.1
 - PersonnelMedical_alimenter_table_dim_zone 0.1
 - PersonnelMedical_alimenter_table_fait_suivi_personnel_medical 0.1





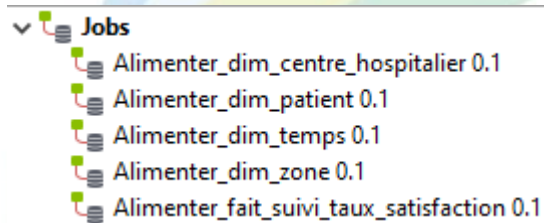


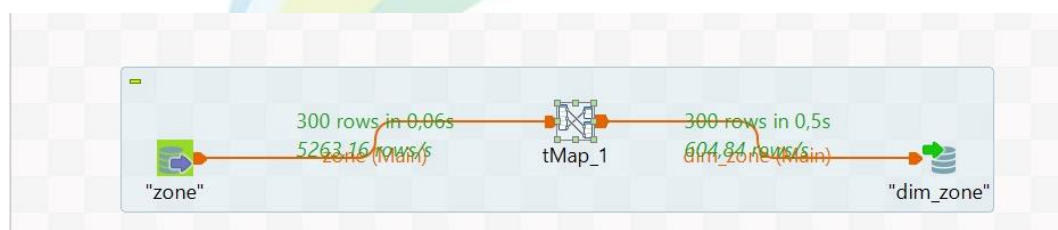
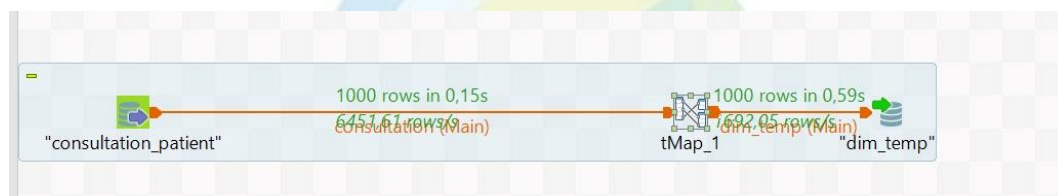
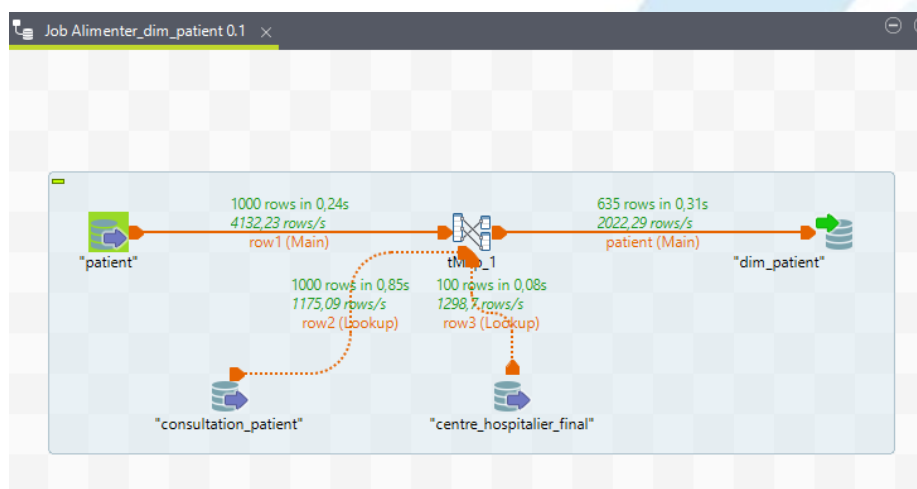
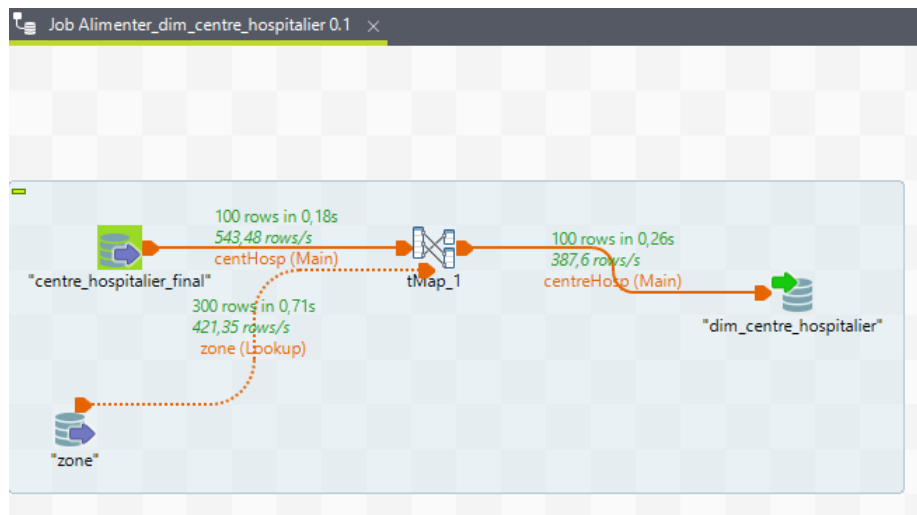
- Connexion à la base de données dw_covid_Suivi_Personnel_Médical

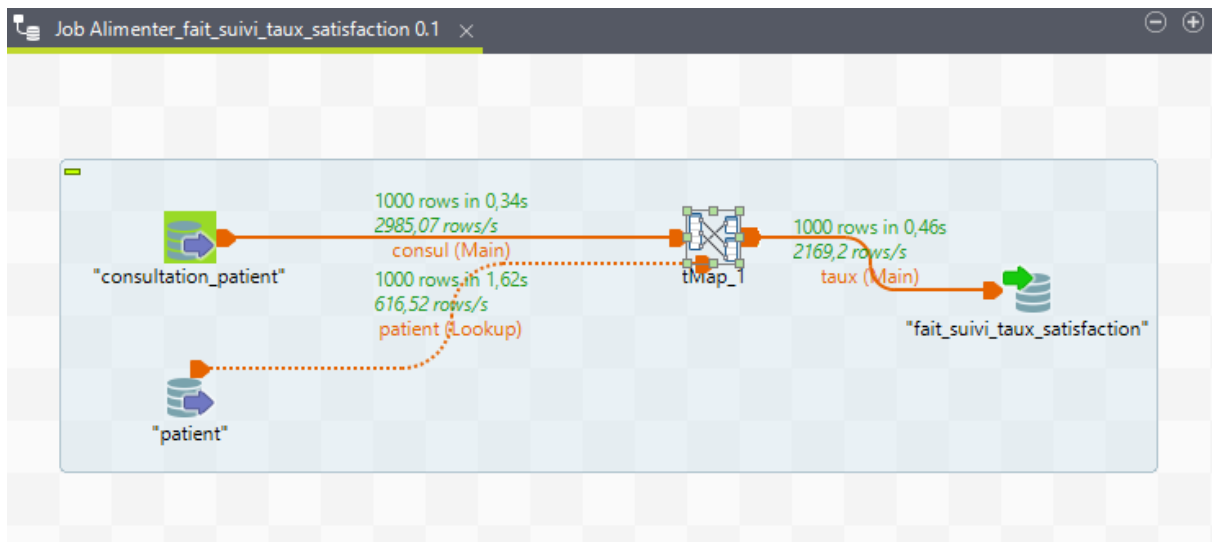


- Suivi_Taux_satisfaction

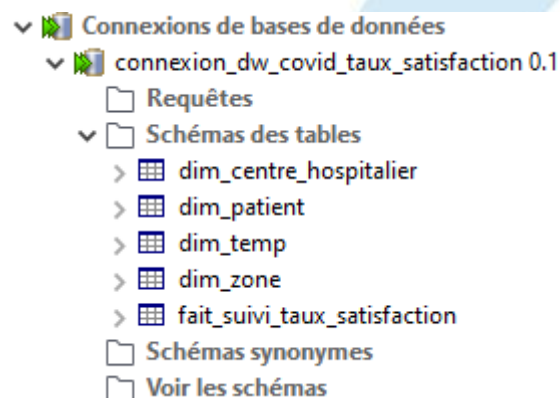
- Jobs







- Connexion à la base de données dw_covid_Suivi_Taux_satisfaction



Dans Talend nous avons utilisé les composants **tMap**, **tFilterRow** et **tJoin**.

- ✚ **tMap** : Le composant **tMap** est utilisé pour effectuer des transformations et des **mappings** de données complexes entre différentes sources ou cibles. Il offre une interface graphique conviviale pour configurer les transformations en utilisant des règles de mapping, des filtres et des fonctions. Vous pouvez mapper des colonnes d'entrée vers des colonnes de sortie, appliquer des expressions, des filtres et des jointures, et effectuer des opérations de conversion de données. Le **tMap** est un composant essentiel pour la manipulation et la transformation de données dans les jobs Talend.
- ✚ **tFilterRow** : Le composant **tFilterRow** est utilisé pour filtrer les lignes de données en fonction de conditions prédéfinies. Il permet de spécifier des expressions ou des règles de filtrage qui déterminent quelles lignes de données doivent être incluses dans le flux de sortie. Vous pouvez utiliser des opérateurs logiques (ET, OU, NON), des opérateurs de comparaison (égal à, supérieur à, inférieur à, etc.) et des fonctions pour créer des conditions de filtrage avancées. Le **tFilterRow** est utile lorsque vous souhaitez exclure certaines données qui ne répondent pas à certains critères de votre flux de données.

- ✚ **tJoin** : Le composant **tJoin** dans Talend est utilisé pour joindre deux flux de données en fonction de colonnes communes. Il permet de combiner les données de deux sources en une seule, en utilisant des règles de jointure spécifiées. Cela peut être utile pour joindre des tables, enrichir des données ou filtrer des enregistrements. Le **tJoin** nécessite la spécification des flux d'entrée, des colonnes de jointure et des règles de jointure. Il offre différents types de jointure (**INNER**, **LEFT**, **RIGHT**, **FULL**) pour gérer les enregistrements non correspondants.

En résumé, le composant **tMap** est utilisé pour les transformations et les **mappings** de données complexes, le **tFilterRow** permet de filtrer les lignes de données en fonction de conditions prédéfinies, le **tAggregateRow** permet d'effectuer des agrégations de données sur des groupes de lignes et le **tJoin** permet de faire des jointures entre plusieurs tables. Ces composants sont largement utilisés dans les jobs Talend pour manipuler, transformer et analyser les données de manière flexible et puissante.

C. Étapes pour calculer chaque mesure à partir de la base de données source

1. Suivi des patients

- **Temps d'hospitalisation :**
 - **Étapes** : Soustraire la date d'admission à la date de sortie pour chaque patient.
 - **Formule** : `Patient.DateSortie - Patient.DateAdmission`
- **Nombre de cas confirmés :**
 - **Étapes** : Compter le nombre de patients avec un test positif.
 - **Formule** : `Count(Patient.TestPositif = 1)`
- **Nombre de patients guéris :**
 - **Étapes** : Compter le nombre de patients dont le statut est "guéri".
 - **Formule** : `Count(Patient.Statut = 'Guéri')`
- **Nombre de patients décédés :**
 - **Étapes** : Compter le nombre de patients dont le statut est "décédé".
 - **Formule** : `Count(Patient.Statut = 'Décédé')`
- **Nombre de patients confinés :**
 - **Étapes** : Compter le nombre de patients dont le statut est "confiné".
 - **Formule** : `Count(Patient.Statut = 'Confiné')`

2. Suivi vaccination

- **Nombre de vaccins administrés :**
 - **Étapes :** Compter le nombre total de doses administrées.
 - **Formule :** Count(Vaccination.DoseAdministree = 1)
- **Nombre de personnes vaccinées :**
 - **Étapes :** Compter le nombre unique de personnes ayant reçu au moins une dose.
 - **Formule :** Count(DISTINCT Vaccination.PersonneID)

3. Suivi Tests

- **Nombre de tests effectués :**
 - **Étapes :** Compter le nombre total de tests effectués.
 - **Formule :** Count(Test.TestEffectue = 1)
- **Résultats des tests (positifs/négatifs) :**
 - **Étapes :** Compter séparément les résultats positifs et négatifs.
 - **Formule :**
 - Positifs: Count(Test.Resultat = 'Positif')
 - Négatifs : Count(Test.Resultat = 'Négatif')

4. Suivi Matériels

- **Nombre de matériels disponibles :**
 - **Étapes :** Compter le nombre total de matériels non défectueux.
 - **Formule :** Count(Materiel.Etat = 'Disponible')
- **Nombre de matériels défectueux :**
 - **Étapes :** Compter le nombre total de matériels défectueux.
 - **Formule :** Count(Materiel.Etat = 'Défectueux')
- **Nombre de matériels réparés :**
 - **Étapes :** Compter le nombre total de matériels réparés.
 - **Formule :** Count(Materiel.Etat = 'Réparé')
- **Nombre de matériels perdus :**

- **Étapes** : Compter le nombre total de matériels perdus.
- **Formule** : Count (Materiel.Etat = 'Perdu')
- **Durée d'utilisation** :
 - **Étapes** : Calculer la durée d'utilisation pour chaque matériel.
 - **Formule** : Materiel.DateFinUtilisation - Materiel.DateDebutUtilisation

5. Suivi Personnel Médical

- **Nombre d'absences** :
 - **Étapes** : Compter le nombre total de jours d'absence.
 - **Formule** : Count(PersonnelMedical.Absence = 1)
- **Nombre de retards** :
 - **Étapes** : Compter le nombre total de retards enregistrés.
 - **Formule** : Count(PersonnelMedical.Retard = 1)

6. Taux de satisfaction

- **Nombre de satisfaction** :
 - **Étapes** : Compter le nombre total de retours satisfaisants.
 - **Formule** : Count(Satisfaction.Evaluation = 'Satisfait')

Conclusion

Ce projet a mis en évidence les faiblesses critiques des systèmes de santé face à la pandémie de COVID-19, tout en proposant des solutions concrètes telles que l'amélioration des systèmes de test, l'optimisation de l'approvisionnement en vaccins, le renforcement du personnel médical et des infrastructures de communication. En utilisant Talend Data Intégration pour centraliser et transformer les données, nous avons pu calculer des mesures clés nécessaires à une prise de décision éclairée. Les recommandations formulées visent à renforcer la résilience des systèmes de santé, permettant ainsi une meilleure gestion des crises actuelles et une préparation efficace pour les défis futurs.