

Présentation Projet I_COMMERCE avec Talend

Contexte

Le projet consiste à intégrer des données de vente avec Talend dans une base de données opérationnelle PostgreSQL. Ces données proviennent d'un site de e-commerce fictif qui édite chaque jour ses données de vente au format csv. Ces données sont stockées dans un dossier ayant pour nom la date du jour au format « YYYYMMDD ».

Les étapes du projet sont la création de la base de données dans PostgreSQL et du dossier à la date du jour avec les données brutes, puis l'élaboration d'un flux d'intégration des données dans la base de données avec Talend.

Ce flux d'intégration comprend pour chacune des tables :

- La connexion à la base de données
- Initialisation des variables globales
- La gestion des valeurs nulles / des valeurs manquantes
- La gestion d'erreurs / gestion des dépendances
- L'insertion/la mise à jour des données dans la base de données

Compétences mises en œuvre

Base de données :

- Création de tables opérationnelles et de contexte
- Requêtes

Talend :

- Création de métadonnées :
 - o Connexion à la base de données
 - o Création de schémas génériques pour les fichiers csv
- Création de groupe de contextes
- Configuration du chargement implicite des variables de contexte
- Intégration de routines Java
- Création de 6 sous jobs et d'un job principal
- Utilisation des variables globales des composants
- Utilisation de divers composants
- Enregistrement du projet et exécution en ligne de commande

Environnement technique

- PostgreSQL
- Talend
- 6 fichiers csv

Liste des composants utilisés sur Talend :

- Orchestration : tPrejob, tPostjob, tFileList, tRunJob
- Log & Errors : tChronometerStart, tChronometerStop, tDie
- Database : tDBConnection, tDBCommit, tDBCclose, tDBInput, tDBOutput
- Custom Code : tSetGlobalVar, tJava
- File : tFileInputDelimited
- Processing : tMap, tFilterRow, tUniqRow







Développement

1) Répertoire de données du jour

Création du répertoire contenant les données brutes : D:/Formation

Talend/Projet/ICOMMERCE/20240117/

20240117 correspond à la date du 17/01/2024.

 ICOM_20240117_CATEGORY.csv	18/01/2024 10:33	Fichier CSV	1 Ko
 ICOM_20240117_CUSTOMER.csv	17/01/2024 13:35	Fichier CSV	1 154 Ko
 ICOM_20240117_PRODUIIT.csv	19/01/2024 14:00	Fichier CSV	246 Ko
 ICOM_20240117_SOUS_CATEGORIE.csv	17/01/2024 13:35	Fichier CSV	1 Ko
 ICOM_20240117_TYPE_CLIENT.csv	17/01/2024 13:37	Fichier CSV	1 Ko
 ICOM_20240117_VENTE.csv	17/01/2024 13:36	Fichier CSV	38 Ko

2) Base de données

- Création de la base de données « COM_INGESTION_DB »
- Création du schéma « VENTE_ODS » et des tables « ODS_CATEGORIE », « ODS_SOUS_CATEGORIE », « ODS_PRODUIIT », « ODS_TYPE_CLIENT », « ODS_CLIENT », « ODS_VENTE » et « ODS_REJET ».
- Création du schéma « PARAMS_LOG » et de la table CONTEXTE, qui contient toutes les variables de contexte, et intégration des variables de contexte.

```
DROP SCHEMA IF EXISTS "PARAMS_LOG" CASCADE;  
CREATE SCHEMA "PARAMS_LOG";
```

```
SET SEARCH_PATH = "PARAMS_LOG";
```

```
--- Création de la table "PARAMS_LOG"."CONTEXTE"  
DROP TABLE IF EXISTS "CONTEXTE" ;  
CREATE TABLE "CONTEXTE"  
(  
  "key" VARCHAR(100) NOT NULL,  
  "value" VARCHAR(255),  
  CONSTRAINT "CONTEXTE_PKEY" PRIMARY KEY("key")  
);
```

```
-- Insertion des données dans la table "CONTEXTE"  
INSERT INTO "CONTEXTE"  
VALUES ('serverName', 'localhost'),  
       ('database', 'COM_INGESTION_DB'),  
       ('port', '5432'),  
       ('schema_ods', 'VENTE_ODS'),  
       ('utilisateur', 'postgres'),  
       ('password', 'mdp'),  
       ('additionalParams', ''),  
       ('schema_param', 'PARAMS_LOG'),  
       ('projectFolder', 'D:/Formation Talend/Projet/ICOMMERCE/'),  
       ('fieldSeparator', '|');
```

```

|-- Création schéma
DROP SCHEMA IF EXISTS "VENTE_ODS" CASCADE;
CREATE SCHEMA "VENTE_ODS";

SET SEARCH_PATH = "VENTE_ODS";

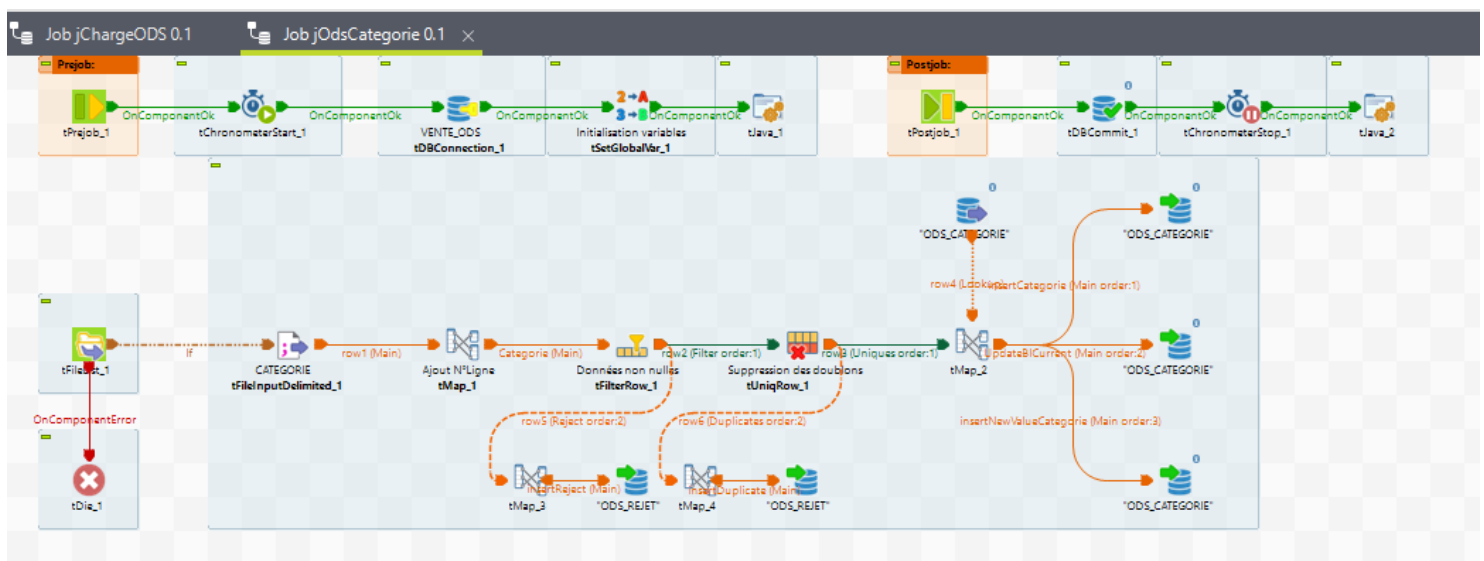
--- Création de la table "VENTE_ODS"."ODS_TYPE_CLIENT"
DROP TABLE IF EXISTS "ODS_TYPE_CLIENT" ;
CREATE TABLE "ODS_TYPE_CLIENT"
(
"CD_TYPE_CLIENT" VARCHAR(50) NOT NULL,
"LB_TYPE_CLIENT" VARCHAR(100) NOT NULL,
"LB_NOM_FICHER" VARCHAR(100) NOT NULL,
"DT_INSERTION" TIMESTAMP NOT NULL,
"BL_LIGNE_ACTIVE" INTEGER NOT NULL,
"LB_JOB_NAME" VARCHAR(100) NOT NULL
);

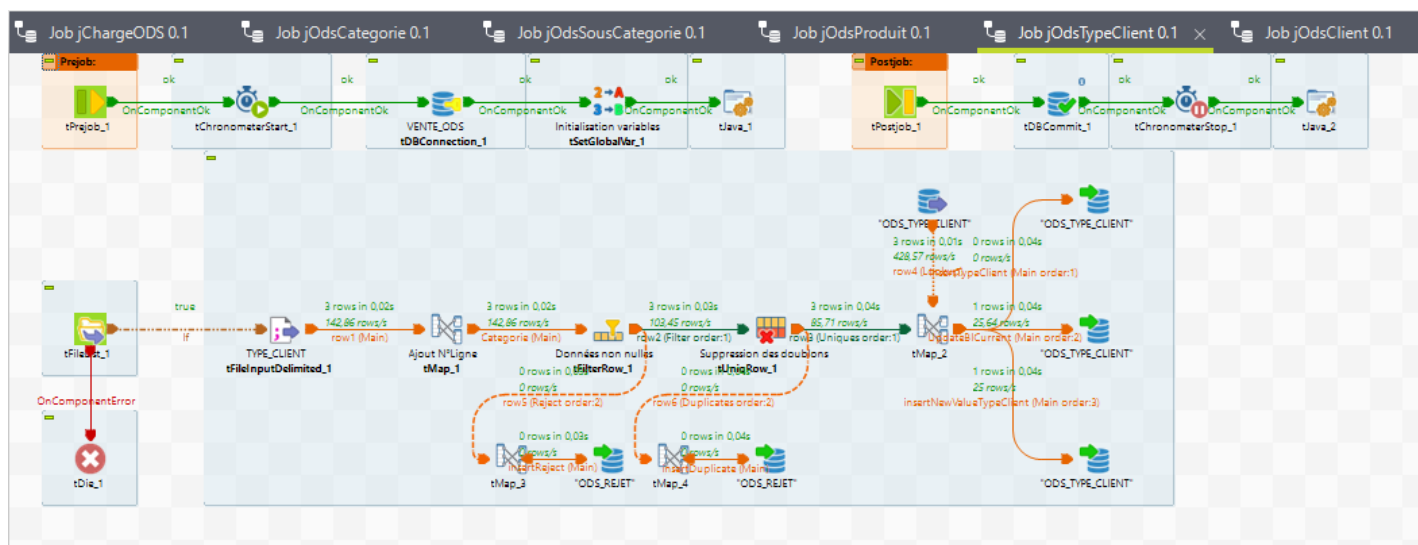
```

3) Processus d'intégration de données avec Talend

- 1- Création du projet Talend : ICOMMERCE
- 2- Création des groupes de contexte
- 3- Création des métadonnées : Connexion à la base de données
- 4- Configuration du chargement implicite des variables de contexte
- 5- Création des différents schémas génériques : Fichiers plats
- 6- Création des job ODS (Transformation et Chargement de la données)
- 7- Création du job principal

4) Présentation des sous jobs





5) Présentation du job principal

