

TP EDD (Data Warehouse – Part 1: Talend)

Presentation date: November 27, 2025

Presentation duration: 10 minutes per pair

Tasks to be completed :

Expense Tracking File Management

- 1- Retrieve the file **SAISIE_DEPENSE.xlsx** located in the directory:
“Commun – Module EDD > TD3”
- 2- Upload the file to your Google Drive (place it in a Google Drive directory)

Liste détaillée des dépenses					
Numéro dépense	Date dépense	Description dépense	Catégorie dépense	Sous-catégorie dépense	Montant dépense
D00001	01/01/2022	Achat Essence	Dépenses Variables	Carburant	32,41
D00002	02/01/2022	Alimentation	Dépenses Variables	Alimentation	20,37
D00003	06/01/2022	Abonnement Microsoft	Dépenses Fixes	Abonnement Logiciel	6,48
D00004	09/01/2022	Alimentation	Dépenses Variables	Alimentation	20,82
D00005	11/01/2022	Achat Essence	Dépenses Variables	Carburant	32,41
D00006	12/01/2022	Frais Electricité	Dépenses Fixes	Énergie (Électricité & Gaz)	52,08
D00007	12/01/2022	Frais bancaire	Dépenses Fixes	Frais Bancaire	0,58
D00008	13/01/2022	Soins et Hygiène	Dépenses Variables	Soins et Hygiène	12,93
D00009	13/01/2022	Alimentation	Dépenses Variables	Alimentation	0,97
D00010	16/01/2022	Alimentation	Dépenses Variables	Alimentation	26,19
D00011	16/01/2022	Alimentation	Dépenses Variables	Alimentation	7,44
D00012	16/01/2022	Loisirs	Dépenses Variables	Loisirs	4,54
D00013	16/01/2022	Frais bancaire	Dépenses Fixes	Frais Bancaire	0,58
D00014	16/01/2022	Prélèvement Abonnement SFR	Dépenses Fixes	Abonnement Télécommunications	8,43
D00015	27/01/2022	Paiement Loyer	Dépenses Fixes	Loyer	363
D00016	19/01/2022	Alimentation	Dépenses Variables	Alimentation	22,75

Creating a Talend Project

- 1- Create a project on GitLab. The project must be named "**idepende_ingestion**"
- 2- Create a project (same name as on GitLab) in the TMC and link this project to your Git repository
- 3- Access this project via **Talend Cloud API Services**

Creating a Database

- 1- Create a Snowflake database and name it "**idepende**"
- 2- Create an **ODS** schema
- 3- Create a **DWH** schema

- 4- Create the table creation scripts for ODS and DWH
 - a. In the **ODS schema**, you must have the sheets (tables) from the source file **SAISIE_DEPENSE.xlsx**
 - b. In the **DWH schema**, you must have the tables from your dimensional model (dimension and fact tables)

Basic Configuration in Talend

- 1- Create context groups
- 2- Create context variables
- 3- Create a database connection in Talend

Creating the ODS Load Jobs

- 1- Create an **ODS** directory in your Talend project
- 2- In this directory, create the following jobs :
 - a. **jCategorie_FILE_ODS** : This job reads your file stored in Google Drive and retrieves data from the sheet "**CATEGORIE_DEPENDSE**".
Then, the data is inserted into the table **CATEGORIE_DEPENDSE** (table in the ODS schema of your database).
 - b. **jSousCategorie_FILE_ODS** : This job reads your file on Google Drive and retrieves the sheet "**SOUS_CATEGORIE_DEPENDSE**".
Then, the data is inserted into the table **SOUS_CATEGORIE_DEPENDSE**.
 - c. **jDepense_FILE_ODS** : This job reads your file on Google Drive and retrieves the sheet "**DEPENDSE**".
Then, the data is inserted into the **DEPENDSE** table.

After inserting the data into the table, the sheet "**DEPENDSE_EXPORTEE**" must be updated by adding only the identifiers of the exported expenses.

NB :

- For Load data into the ODS tables : **Choose the option to delete existing data and insert new data.**
- You must use **context variables** in your jobs.
Example: database connection parameters, file path, and file name must all be context variables.

Creating the DWH Load Jobs

- 1- Create a **DWH** directory
- 2- Create the following jobs in this directory :
 - a. **jDimSousCategorieDepense_ODS_DWH** : This job loads the Sub-category Expense dimension table.
 - b. **jDimDescriptionDepense_ODS_DWH** : This job loads the Description Expense dimension table.
 - c. **jDimTemp_ODS_DWH** : This job loads the Time dimension table.
 - d. **jFaitDepense_ODS_DWH** : This job loads the Expense fact table.

Submission of the TP:

The TP must be done in pairs. For the submission, you should present your work as a pair. For this :

- You will explain your technical choices
- Do a demo of your integration system using a test expense tracking Excel file.

For the demo:

- o You should create the database
- o You should execute the creation scripts for the ODS and DWH tables
- o Run the .sh or .bat of your master Talend jobs
- o Show the presence of data in the different ODS and DWH tables