

This document is the intellectual property of its author. Any reproduction, partial or total, without explicit written permission will be considered a copyright infringement.

Conception of a datawarehouse for project management analysis

Made by: Meradji Anaïs, Anane Nardjes and Lounis Cerine

Table des matières

Introduction générale	5
1. Définition du périmètre du projet	6
2. Définition de la table de faits et des dimensions.....	7
3. Définition des métadonnées	8
4. Schéma du datawarehouse	10
5. Présentation des données brutes	11
6. Processus ETL	14
1. Traitement des données	16
2. Intégration des données	20
3. Datawarehouse final	21
7. Création du cube Olap.....	21
Définition du Data Mart – Analyse de la gestion de projets	22
1. Table de faits	22
2. Tables de dimensions.....	22
3. Relations.....	23
4. Structure du cube OLAP	23
5. Configuration des agrégations	23
Exemples d’analyses permises par ce Data Mart.....	24
Constitution du cube Olap sur Pentaho Schema Workbench.....	24
8. Analyse et visualisation des données	29
Analyses	33
1) Le retard des projets selon les chefs de projets.....	33
2) Dépassement de budget par département	34
3) Retard par Année et Trimestre	35
4) Le retard par nombre de projets menés par les chefs	36
5) Dépassement de budget par chef de projet	37
6) Dépassement de budget par nombre de projets menés par le chef	38
7) Dépassement de budget par année et trimestre	39
Conclusion générale	41
Sources	42

Liste des figures

Figure 1: système d'aide à la décision.....	5
Figure 2: schéma en étoile du datawarehouse	10
Figure 3: fichier CSV des équipes	11
Figure 4: fichier excel des budgets.....	12
Figure 5: fichier json clients	13
Figure 6: fichier JSON clients	13
Figure 7: base de données d'entrée	13
Figure 8: table type de projet.....	14
Figure G: logo Talend	14
Figure 10: ajout du fichier csv, excel, bdd	15
Figure 11: fichier j son ajouté	16
Figure 12: composant t-map sur equipes.csv	17
Figure 13: suppression de la colonne success rate	17
Figure 14: output du traitement du fichier csv	18
Figure 15: filtrage des valeurs non nulles sur budgets excel.....	18
Figure 16: résultats du filtrage des valeurs non nulles.....	19
Figure 17: moyenne de budget initial.....	19
Figure 18: fichier excel budgets sans valeurs nulles	19
Figure 1G: intégration des données	20
Figure 20: datawarehouse final	21
Figure 21: logo Pentaho	21
Figure 22: interface de pentaho schema workbenc	24
Figure 23: connexion au datawarehouse sur Pentaho	25
Figure 24: création du cube et définition des tables.....	25
Figure 25: niveaux de hiérarchies pour la dimension équipe	26
Figure 26: dimension chef de projet	26
Figure 27: dimension temps	26
Figure 28: dimension type de projet.....	26
Figure 2G: dimension client.....	26
Figure 30: dimension équipe	26
Figure 31: mesure retard_jours	27
Figure 32: mesure dépassement budget	27
Figure 33: les mesures de notre table de faits	27
Figure 34: cube olap finalisé.....	28
Figure 35: structure de notre cube Olap	28
Figure 36: sauvegarde du cube Olap	29
Figure 37: logo de Power Bi	29
Figure 38: obtention des données Power Bi	30
Figure 3G: connexion à la bdd sql server.....	30
Figure 40: identifiants de la bdd.....	31
Figure 41: sélection des tables du datawarehouse sur Power Bi.....	31
Figure 42: Chargement des données du datawarehouse sur Power Bi	32
Figure 43: Histogramme empilé retard_jours par Nom chef de projet.....	33
Figure 44: Graphique en anneau représentant la.....	34
Figure 45: Histogramme de la somme des retards par année et trimestre	35

Figure 46: Graphique en aires (area chart) de la somme des retards par nombre de projets menés par le chef de projet.....	36
Figure 47: Graphique en anneau de dépassement budget par chef de projet.....	37
Figure 48: Histogramme de dépassement budget par chef de projet	38
Figure 4G: Graphique circulaire de dépassement budget par nombre de projets menés par le chef de projet	39
Figure 50: Graphique de courbe de dépassement de budget par année (2023) et trimestre	40

Liste des tableaux

Tableau 1: : Structure de la table de faits	8
Tableau 2: : Structure de la table de dimension 'temps'.....	9
Tableau 3: Structure de la table de dimension 'type_projet'	9
Tableau 4: Structure de la table de dimension 'équipe'	9
Tableau 5: Structure de la table de dimension 'chef_projet'	10
Tableau 6: Structure de la table de dimension 'client'	10
Tableau 7: tableau de calcul des mesures de la table de faits	20

Introduction générale

De nos jours, la donnée est devenue une ressource cruciale pour les entreprises. En effet, l'exploitation et l'analyse des données permettent de mieux comprendre les tendances du marché et d'améliorer la prise de décision.

Elles permettent également d'optimiser les processus internes et augmenter l'efficacité opérationnelle. En analysant les données, une entreprise peut identifier les goulots d'étranglement et mettre en place des actions correctives. C'est le cas lors de la gestion de projets.

La gestion de projets est une tâche ardue qui nécessite une coordination rigoureuse des ressources, des délais et des budgets. Il n'est pas rare de constater des dépassements de coûts, des retards ou des écarts entre les objectifs initiaux et les résultats finaux. C'est pourquoi il est essentiel d'analyser ces aspects à la fin de chaque projet et de générer des rapports détaillés afin d'identifier les points d'amélioration pour les futurs projets et de prendre des décisions réfléchies en conséquence. C'est le but de la mise en place de l'entrepôt de données dont il est question dans ce projet.

Pour ce faire, nous allons passer par les différentes étapes nécessaires lors de la mise en place d'un système d'aide à la décision, à savoir :

- Définition du périmètre du projet, identification de la table de fait et des axes d'analyse (dimensions)
- Modélisation multidimensionnelle afin de structurer les données sous la forme de tables de faits et de dimensions.
- Intégration des données dans l'entrepôt de données à travers un processus ETL (**Extract Transform Load**) exercé afin de collecter des données à partir de différentes sources hétérogènes, les nettoyer, puis les charger dans l'entrepôt.
- Exploitation des données du datawarehouse afin d'extraire des informations stratégiques des différents rapports et visualisations générés.

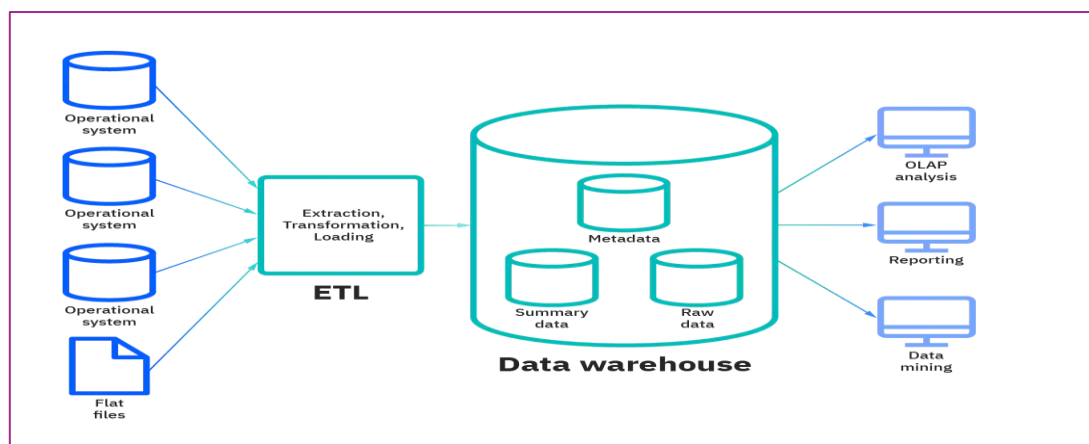


Figure 1: système d'aide à la décision

1. Définition du périmètre du projet

Ce projet consiste en la conception d'un entrepôt de données dédié à l'analyse de la gestion de projets, et ce dans l'optique de la mise en place d'un système d'aide à la décision qui a pour finalité l'optimisation du processus de gestion des différents projets entrepris au sein d'une entreprise.

La réussite d'un projet terminé est définie par plusieurs critères. Dans le cadre de ce projet, nous nous focalisons principalement sur deux critères de réussite : le respect des délais ainsi que le respect du budget (éviter ou minimiser au mieux le dépassement du budget alloué).

Les données collectées et organisées dans l'entrepôt permettront de fournir une réponse aux questions suivantes :

- Combien de projets ont été achevés dans les délais initialement prévus, et combien ont subi des retards ?
- Combien de projets ont respecté le budget alloué, et combien l'ont dépassé ? Quelle est l'ampleur de ces dépassements ?
- Quels sont les types de projets ayant le plus réussi ?
- Quels chefs de projets ont mené à bien les projets dont ils étaient responsables ?
- Quel est le nombre optimal d'employés au sein d'une équipe afin de réaliser un projet réussi ?

Ce projet mettra en œuvre une structure en étoile avec une table de faits centralisant les informations sur les projets et des tables de dimensions permettant des analyses multidimensionnelles.

2. Définition de la table de faits et des dimensions

1. Identification des faits :

- Table des faits : **projet_terminé**

Indicateurs de performances (les mesures) :

- **Nombre de jours de retard** sur le délai prévu
- **Montant** du dépassement budgétaire

2. Identification des dimensions :

- **Dimension temps :**
 - **ID_temps** (clé primaire)
 - Trimestre
 - Semestre
 - Année
- **Dimension type de projet :**
 - **ID_TypeDeProjet** (clé primaire)
 - Libellé
 - Département
 - Catégorie
- **Dimension équipe :**
 - **ID_equipe** (clé primaire)
 - Département
 - Nombre de membres
 - Expérience moyenne des membres
- **Dimension Chef Projet :**
 - **ID_ChefProjet** (clé primaire)
 - Nom Chef Projet
 - Expérience (années)
 - Nombre Projets Menés
 - Spécialisation (ex. IT, Construction, Marketing, etc.)
- **Dimension client :**
 - **ID_client** (clé primaire)

- Nom client
- Type client (particulier/entreprise)
- Localisation

Plus de détails sur les attributs de la table de faits et des dimensions seront fournis au niveau du titre suivant.

3. Définition des métadonnées

Table de faits : projet_terminé

Attribut	Type	Format	Description
Id_temps	INT	Clé étrangère	Identifiant unique du temps
ID_TypeDeProjet	INT	Clé étrangère	Identifiant unique de type de projet
ID_equipe	INT	Clé étrangère	Identifiant unique de l'équipe
ID_ChefProjet	INT	Clé étrangère	Identifiant unique de chef de projet
ID_client	INT	Clé étrangère	Identifiant unique du client
Nombre de jours de retard	INT	Nombre entier positif	Nombre de jours de retard sur le délai prévu
Montant du dépassement budgétaire	FLOAT	Nombre positif	Montant du dépassement budgétaire

Tableau 1: : Structure de la table de faits

Tables de dimensions :

- Table temps :

Attribut	Type	Format	Description
Id_temps	INT	Clé primaire	Identifiant unique du temps
Trimestre	INT	Nombre entier positif	Trimestre du Temps
Semestre	INT	Nombre entier positif	Semestre du Temps
Année	INT	Nombre entier positif	Année du Temps

Tableau 2: : Structure de la table de dimension 'temps'

- Table type_projet :

Attribut	Type	Format	Description
ID_TypeDeProjet	INT	Clé primaire	Identifiant unique de type de projet
Libellé	String	Texte Long	Nom du projet
Département	String	Texte Court	Département du projet
Catégorie	String	Texte Court	Catégorie du projet

Tableau 3: Structure de la table de dimension 'type_projet'

- Table équipe :

Attribut	Type	Format	Description
ID_equipe	INT	Clé primaire	Identifiant unique de l'équipe
Département	String	Texte court	Département de l'équipe
Nombre de membres	INT	Nombre entier positif	Nombre de membres de l'équipe
Expérience moyenne des membres	float	Nombre positif	Expérience moyenne des membres

Tableau 4: Structure de la table de dimension 'équipe'

- Table chef_projet :

Attribut	Type	Format	Description
ID_ChefProjet	INT	Clé primaire	Identifiant unique de chef de projet
Nom Chef Projet	String	Texte court	Nom de Chef Projet
Expérience (années)	INT	Nombre entier positif	Expérience de chef de projet en années
Nombre Projets Menés	INT	Nombre entier positif	Nombre de projets menés par le chef de projet

Spécialisation	String	Texte court	Spécialisation (ex. IT, Construction, Marketing, etc.)
----------------	--------	-------------	--

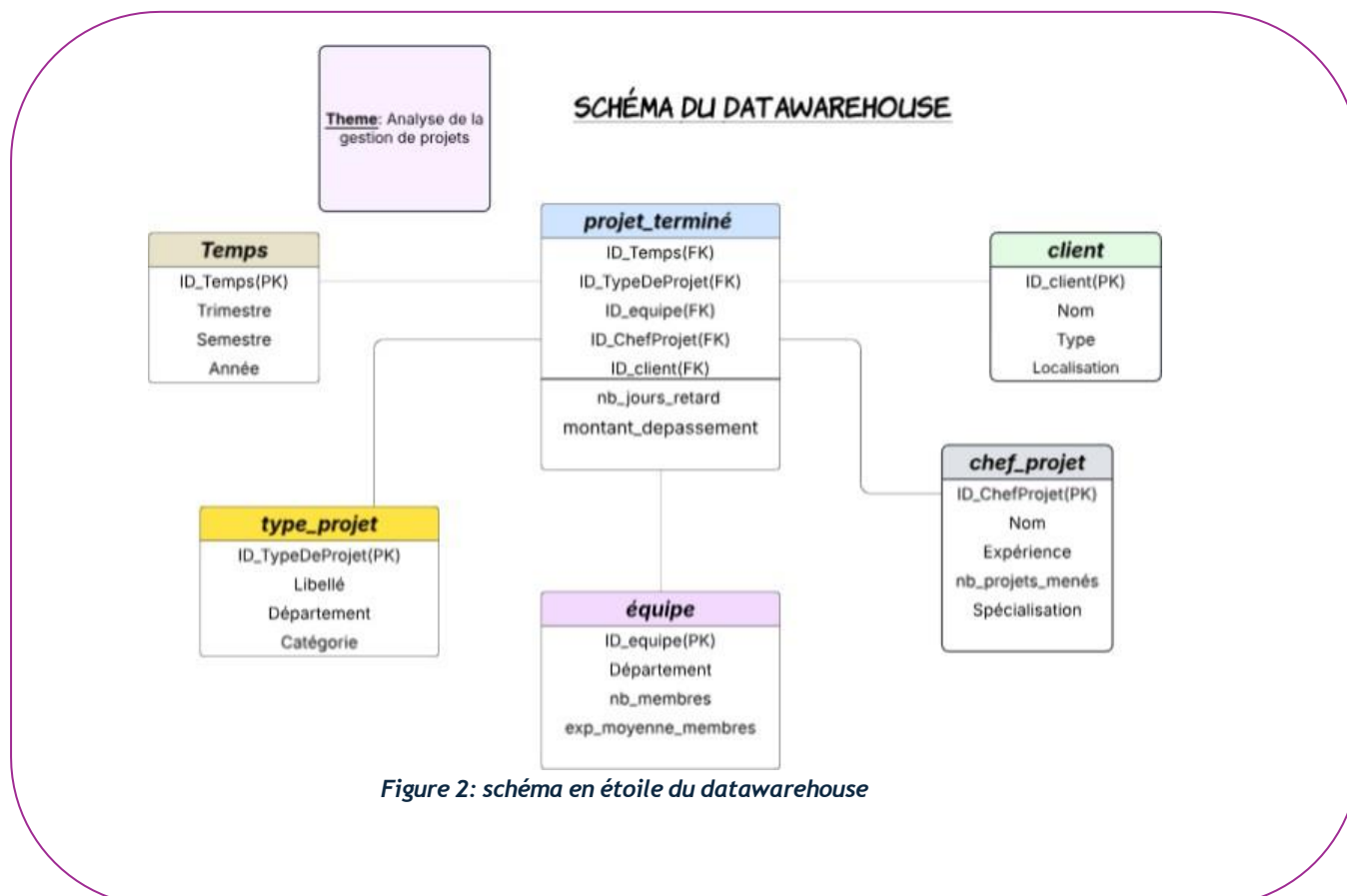
Tableau 5: Structure de la table de dimension 'chef_projet'

- Table client :

Attribut	Type	Format	Description
ID_client	INT	Clé primaire	Identifiant unique du client
Nom client	String	Texte court	Nom du client
Type client	String	Texte court	Type de client (particulier/entreprise)
Localisation	String	Texte court	Adresse du client

Tableau C: Structure de la table de dimension 'client'

4. Schéma du datawarehouse



5. Présentation des données brutes

Nous avons :

1. Un fichier .csv qui contient des informations relatives aux **équipes**

1	Team_ID,Team_Name,Members_Count,Avg_Experience,Success_Rate
2	1,Equipe 1,3,9.3,88%
3	2,Equipe 2,9,5.7,89%
4	3,Equipe 3,12,2.1,69%
5	4,Equipe 4,8,6.4,63%
6	5,Equipe 5,8,3.1,63%
7	6,Equipe 6,6,7.6,67%
8	7,Equipe 7,14,7.5,77%
9	8,Equipe 8,6,8.6,90%
10	9,Equipe 9,15,6.5,74%
11	10,Equipe 10,12,6.6,83%
12	11,Equipe 11,3,7.4,67%
13	12,Equipe 12,3,8.9,99%
14	13,Equipe 13,13,9.0,96%
15	14,Equipe 14,5,3.2,81%
16	15,Equipe 15,12,3.0,94%

Figure 3: fichier CSV des équipes

Ce fichier CSV contient 5 colonnes :

ID équipe, nom de l'équipe, le nombre d'employés dans l'équipe, l'expérience moyenne des membres de l'équipe ainsi que leur pourcentage de succès lors des projets réalisés précédemment.

Et il contient 50 lignes. Donc 50 équipes enregistrées.

- 2 . Un fichier Excel intitulé **budgets** et qui représente les budgets alloués et les budgets réels des projets ainsi que la deadline et la date de fin réelle du projet.

ProjectID	udgetInitia	udgetFina	Deadline	DateFin
1	84352,73	85737	10/01/202	20/01/2026
2	38814,06	39965	31/12/202	04/01/2026
3	64288,49	66011	30/06/202	31/07/2024
4	61169,17	62435	31/12/202	12/01/2026
5	58226,66	59705	31/12/202	31/12/2025
6	64961,23	66145	31/12/202	14/01/2026
7	65630,34	67551	15/03/202	02/04/2025
8	49274,42	50325	15/03/202	15/03/2025
9	52187,65	53587	30/06/202	19/07/2024
10	40676,56	42354	15/03/202	18/03/2025
11	44330,74	45884	15/03/202	10/04/2025
12	ND	52285	15/03/202	21/03/2025
13	42312,3	43633	15/03/202	20/03/2025
14	42044,6	42935	30/06/202	01/07/2024
15	ND	73280	31/12/202	30/01/2026
16	ND	36306	15/03/202	05/04/2025

Figure 4: fichier excel des budgets

Colonnes : id du projet, budget initial alloué, budget final réel, deadline et dateFin.

Nous remarquons que nous avons des valeurs vides. Nous les traiterons plus tard sur [Talend](#).

3. Un fichier .json qui contient nos **clients**

```
[
  {
    "client_id": 1,
    "name": "Entreprise 1",
    "type": "Entreprise",
    "address": {
      "city": "Marseille",
      "country": "France"
    }
  },
  {
    "client_id": 2,
    "name": "Entreprise 2",
    "type": "Entreprise",
    "address": {
      "city": "Nice",
      "country": "France"
    }
  },
  {
    "client_id": 3,
```

Figure C: fichier JSON clients

```
},
  {
    "client_id": 47,
    "name": "Paul Vidal",
    "type": "Particulier",
    "address": {
      "city": "Toulouse",
      "country": "France"
    }
  },
  {
    "client_id": 48,
    "name": "Anaïs Garnier",
    "type": "Particulier",
    "address": {
      "city": "Nice",
      "country": "France"
    }
  },
  {
    "client_id": 49,
```

Figure 5: fichier json clients

Et enfin

4. Une base de données dont les tables décrivent des informations importantes concernant les projets terminées (le temps, les chefs de projet et le type de projet)

Table	Action	Lignes	Type	Interclassement
<input type="checkbox"/> chefs_projet	Parcourir Structure Rechercher Insérer Vider Supprimer	25	InnoDB	utf8mb4_general_ci
<input type="checkbox"/> temps	Parcourir Structure Rechercher Insérer Vider Supprimer	50	InnoDB	utf8mb4_general_ci
<input type="checkbox"/> type_projet	Parcourir Structure Rechercher Insérer Vider Supprimer	15	InnoDB	utf8mb4_general_ci
3 tables	Somme	90	InnoDB	utf8mb4_general_ci

Figure 7: base de données d'entrée

A titre d'exemple voici la table type_projet qui nous renseigne sur les types de projets terminés.

ID_TypeDeProjet	Libellé	Département	Catégorie
1	Développement Logiciel	IT	Technologie
2	Migration Cloud	IT	Infrastructure
3	Sécurité Informatique	IT	Sécurité
4	Bâtiment Résidentiel	Construction	Immobilier
5	Infrastructure Publique	Construction	Génie Civil
6	Rénovation Urbaine	Construction	Aménagement
7	Campagne Digitale	Marketing	Digital
8	Étude de Marché	Marketing	Recherche
9	Stratégie de Marque	Marketing	Branding
10	Optimisation Supply Chain	Logistique	Transport
11	Gestion d'Entrepôt	Logistique	Stockage
12	Automatisation Flux	Logistique	Innovation
13	Audit Comptable	Finance	Contrôle
14	Optimisation Fiscale	Finance	Fiscalité
15	Système ERP Financier	Finance	Technologie

Figure 8: table type de projet

6. Processus ETL

Nous avons nos données hétérogènes. Maintenant il est temps de les **nettoyer**, **transformer** et **homogénéiser** avant de les charger dans notre entrepôt de données.

Pour cela nous allons utiliser **Talend**.



Figure 3: logo Talend

Talend est un outil d'intégration de données open source qui permet de collecter, transformer et intégrer des données provenant de différentes sources vers des destinations cibles .

Il est largement utilisé pour les processus **ETL** (Extract, Transform, Load).

Talend propose une interface graphique pour concevoir des flux de données facilement.

La première étape consiste à **charger les données** que nous avons sur Talend.

Nous nous connectons a notre base de données, nous ajoutons notre fichier CSV en **fichier délimité**, et nous ajoutons notre fichier Excel en **Fichier Excel**.



Figure 10: ajout du fichier csv, excel, bdd

Nous ajoutons également notre fichier JSON.

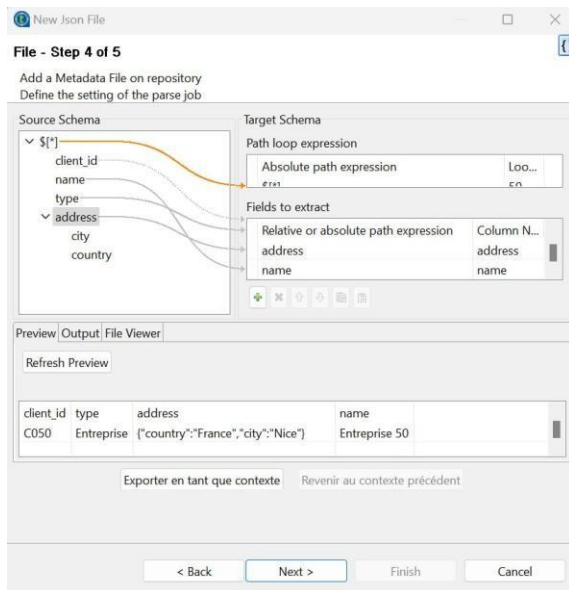


Figure 10: ajout du fichier clients json

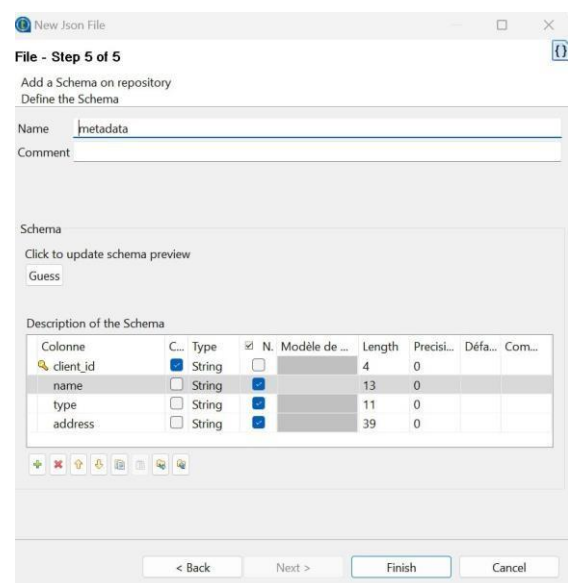


Figure 11: fichier json ajouté

Enfin, nous créons des **jobs** sur Talend afin d'entamer notre processus ETL.

*Un **JOB** sous Talend est un processus automatisé composé de composants enchaînés, permettant d'extraire, transformer et charger des données (ETL).*



Figure 11: création des jobs

1. Traitement des données

- Traitement fichier equipes .csv

Ce fichier contient une colonne success rate dont on a pas besoin, nous allons donc utiliser **Talend** pour ne pas l'intégrer.

Nous allons seulement extraire les colonnes dont nous aurons besoin.

Pour cela nous allons utiliser le composant tMap.

-tMap : le composant tMap dans Talend permet de transformer, filtrer et mapper les données entre des flux d'entrée et de sortie de manière visuelle et personnalisable.

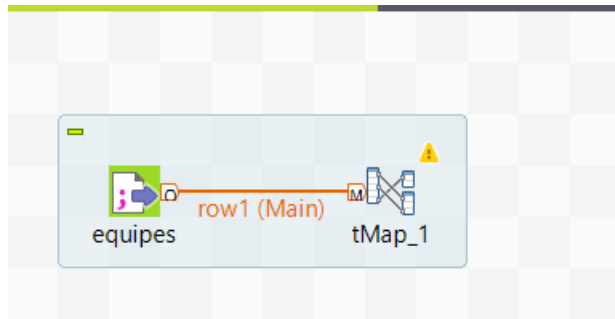


Figure 12: composant t-map sur equipes.csv

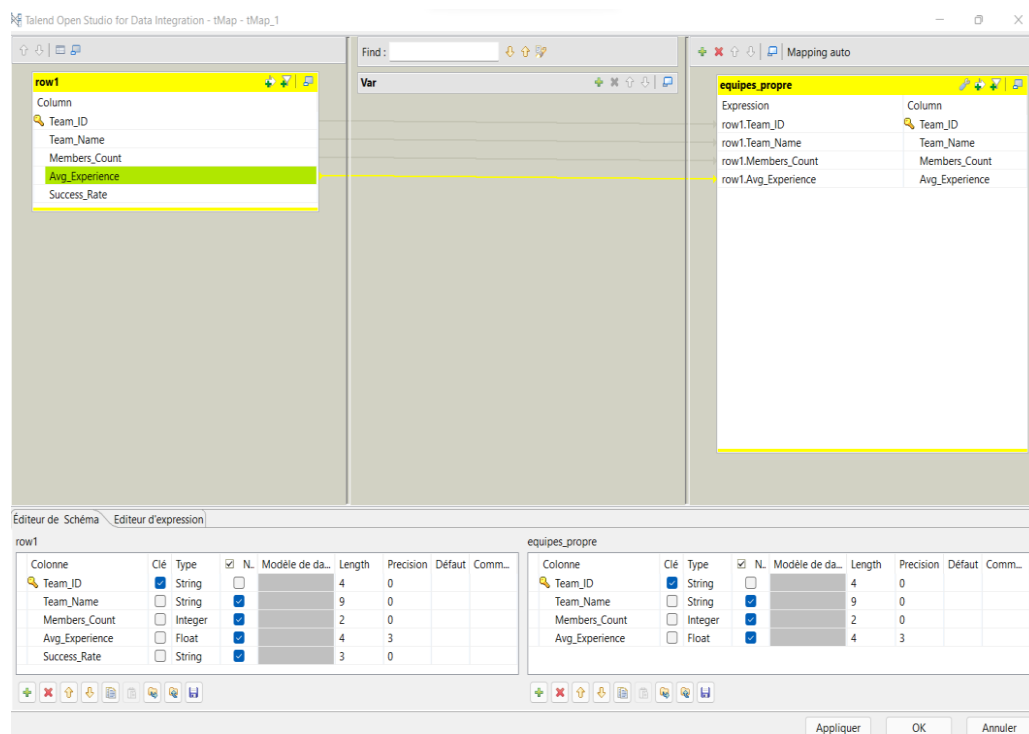


Figure 13: suppression de la colonne success rate

Nous ne glissons pas la colonne Success_Rate.

Nous en avons fini pour le traitement du fichier CSV. Il ne contient pas de valeurs nulles ou incorrectes.

Nous pouvons, par la suite, sauvegarder le résultat de notre traitement dans un nouveau fichier grâce au composant tFileOutputDelimited.

-tFileOutputDelimited : permet de sauvegarder les résultats localement après traitement en générant des fichiers texte délimités, tout en offrant des fonctionnalités comme la gestion des en-têtes, le choix du séparateur, l'encodage, l'écrasement ou l'ajout de données,

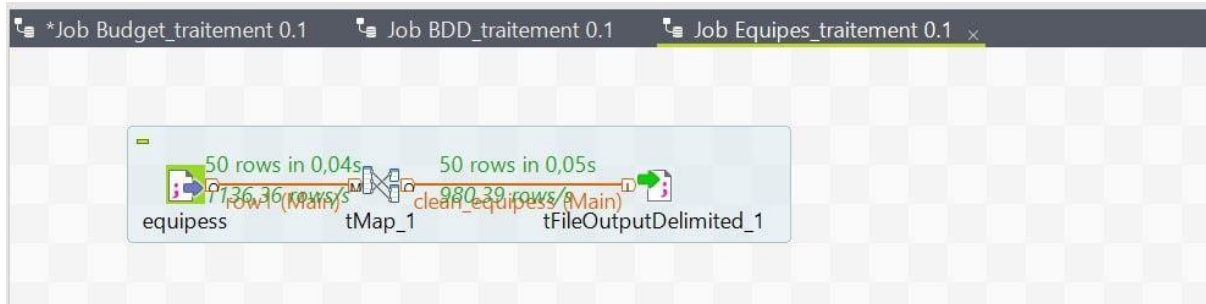


Figure 14: output du traitement du fichier csv

- Traitement du fichier budgets.xlsx :

Problème trouvé :

Le fichier budgets.xlsx contenait des valeurs nulles.

1. Nous utilisons le composant **tFilterRow** afin d'appliquer un filtre sur notre fichier et extraire seulement les lignes qui ne contiennent pas de valeurs nulles pour pouvoir par la suite calculer la moyenne de la colonne budgetInitial.

Le composant **tFilterRow** de Talend permet de filtrer les lignes d'un flux en fonction de conditions définies sur les colonnes. Il redirige les données qui répondent aux critères vers une sortie principale et celles qui ne répondent pas vers une sortie de rejet (si configurée).

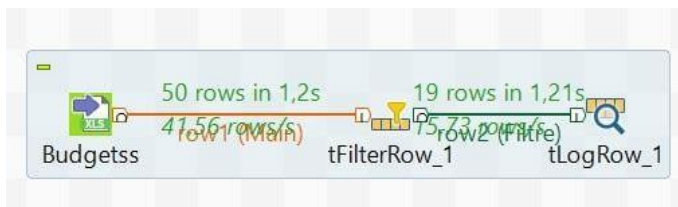


Figure 15: filtrage des valeurs non nulles sur budgets excel

tLogRow_1		
BudgetInitial	BudgetFinal	Deadline
37571	66011	2024/06/
36231	67551	2025-03-
65046	50325	2025-03-

Figure 1C: résultats du filtrage des valeurs non nulles

On fait une **moyenne** de la colonne budget initial.

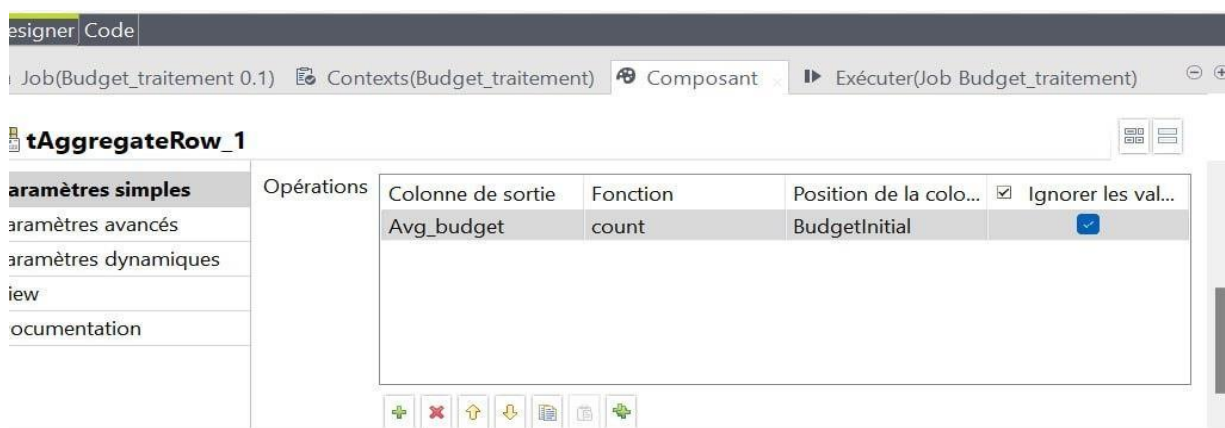


Figure 17: moyenne de budget initial

On remplacera la moyenne de cette colonne au niveau des valeurs NULLES dans la colonne budget initial.

ProjectID	BudgetInitial	BudgetFinal	Deadline	DateFin
1	84352,73	85737	10/01/202	20/01/2026
2	38814,06	39965	31/12/202	04/01/2026
3	64288,49	66011	30/06/202	31/07/2024
4	61169,17	62435	31/12/202	12/01/2026
5	58226,66	59705	31/12/202	31/12/2025
6	64961,23	66145	31/12/202	14/01/2026
7	65630,34	67551	15/03/202	02/04/2025
8	49274,42	50325	15/03/202	15/03/2025
9	52187,65	53587	30/06/202	19/07/2024
10	40676,56	42354	15/03/202	18/03/2025
11	44330,74	45884	15/03/202	10/04/2025
12	57576,14	52285	15/03/202	21/03/2025
13	42312,3	43633	15/03/202	20/03/2025
14	42044,6	42935	30/06/202	01/07/2024
15	57576,14	73280	31/12/202	30/01/2026
16	57576,14	36306	15/03/202	05/04/2025
17	57576,14	89978	30/06/202	02/07/2024
18	57576,14	89695	30/06/202	03/07/2024

Figure 18: fichier excel budgets sans valeurs nulles

- Le fichier de budget sert uniquement **au traitement ETL**, pas à la modélisation du schéma dimensionnel.

Grace a ce fichier, le **dépassement budgétaire** a été calculé à travers une soustraction du budget initial et final et est devenu une mesure que nous avons chargé dans la table de fait par la suite lors de l'intégration des données.

De même pour le **retard**.

Mesure	Source de calcul	Formule Talend
Dépassement budget	BudgetFinal - BudgetInitial	row.BudgetFinal - row.BudgetInitial
Retard (jours)	DateFin - Deadline	daysBetween(row.Deadline, row.DateFin)

Tableau 7: tableau de calcul des mesures de la table de faits

Pour le fichier **Clients.json** et la **base de données**, nous n'avons pas eu de traitement à faire. Maintenant, passons à **l'intégration des données**.

2. Intégration des données :

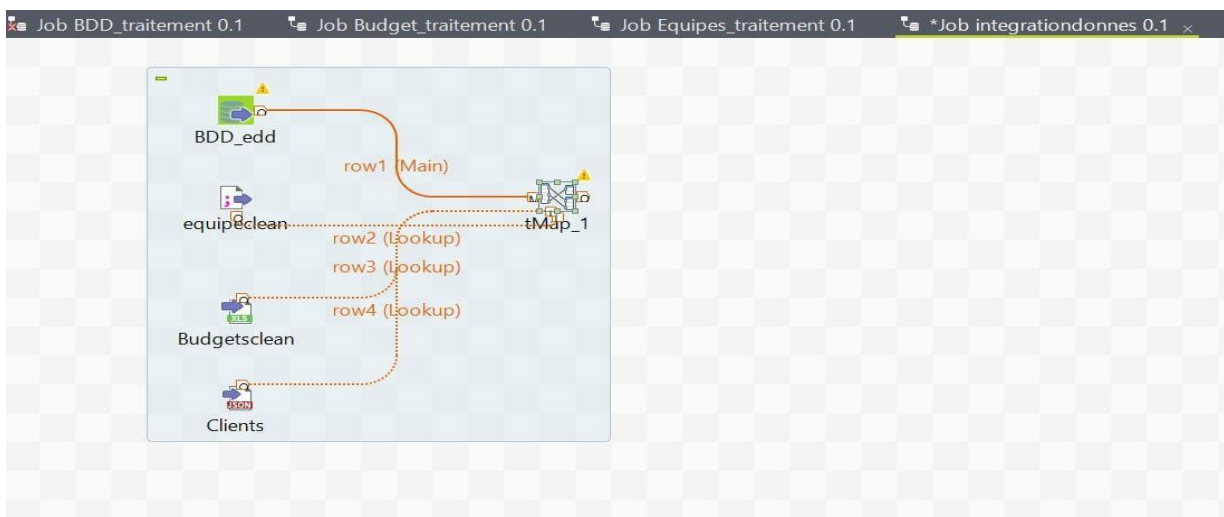
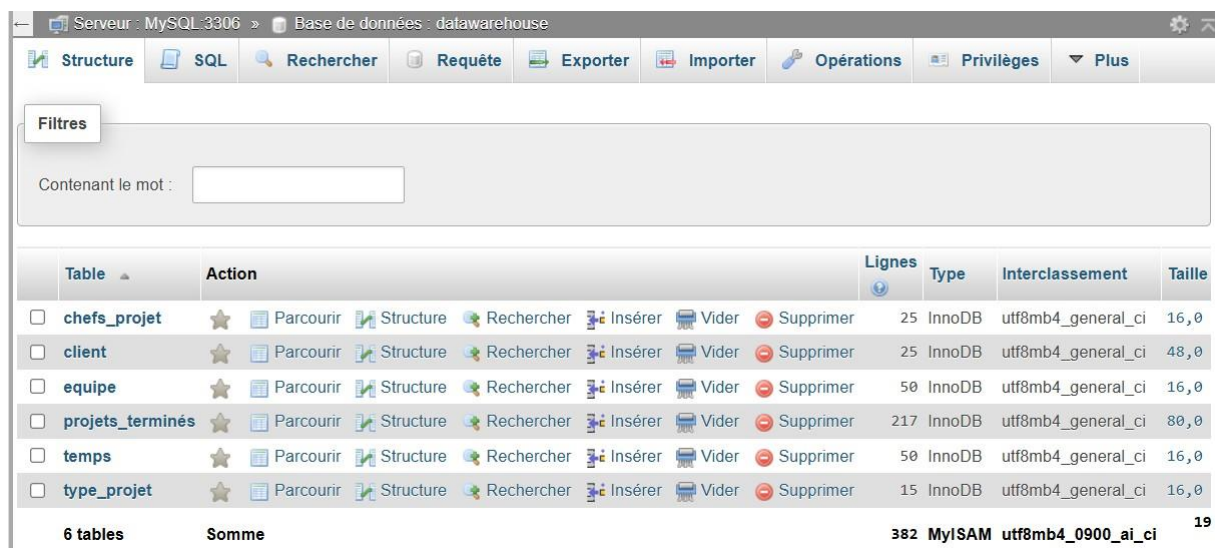


Figure 13: intégration des données

3. Datawarehouse final



The screenshot shows the MySQL Workbench interface for a database named 'datawarehouse'. The 'Structure' tab is active, displaying a list of tables. A search filter is present at the top. The table list includes columns for Table, Action, Lignes, Type, Interclassement, and Taille. The tables listed are chefs_projet, client, equipe, projets_terminés, temps, and type_projet. A summary row at the bottom indicates 6 tables with a total of 382 lines.

Table	Action	Lignes	Type	Interclassement	Taille
<input type="checkbox"/> chefs_projet	★ Parcourir Structure Rechercher Insérer Vider Supprimer	25	InnoDB	utf8mb4_general_ci	16,0
<input type="checkbox"/> client	★ Parcourir Structure Rechercher Insérer Vider Supprimer	25	InnoDB	utf8mb4_general_ci	48,0
<input type="checkbox"/> equipe	★ Parcourir Structure Rechercher Insérer Vider Supprimer	50	InnoDB	utf8mb4_general_ci	16,0
<input type="checkbox"/> projets_terminés	★ Parcourir Structure Rechercher Insérer Vider Supprimer	217	InnoDB	utf8mb4_general_ci	80,0
<input type="checkbox"/> temps	★ Parcourir Structure Rechercher Insérer Vider Supprimer	50	InnoDB	utf8mb4_general_ci	16,0
<input type="checkbox"/> type_projet	★ Parcourir Structure Rechercher Insérer Vider Supprimer	15	InnoDB	utf8mb4_general_ci	16,0
6 tables	Somme	382	MyISAM	utf8mb4_0900_ai_ci	19

Figure 20: datawarehouse final

7. Création du cube Olap

Nous mettons en œuvre notre Data Warehouse préalablement construit en utilisant [Pentaho Schema Workbench](#) pour concevoir un **cube OLAP**. Cette démarche vise à organiser les données de manière à faciliter les analyses multidimensionnelles.



Figure 21: logo Pentaho

Pentaho Schema Workbench est un outil graphique permettant de modéliser des cubes OLAP en définissant les schémas de dimensions et de faits à partir d'un Data Warehouse.

Définition du Data Mart - Analyse de la gestion de projets

Pour rappel, le Data Mart de ce projet est conçu pour analyser les projets réalisés au sein d'une entreprise.

Il est structuré selon un **schéma en étoile**, permettant des analyses multidimensionnelles efficaces sur des indicateurs comme les retards ou les dépassements budgétaires.

1. Table de faits

Nom : projet_terminé

Objectif : Centraliser les données quantitatives liées à l'exécution des projets, telles que le nombre de jours de retard ou le montant du dépassement budgétaire.

Colonnes principales :

- ID_Temps (clé étrangère vers la dimension temps)
- ID_TypeDeProjet (clé étrangère vers type_projet)
- ID_equipe (clé étrangère vers équipe)
- ID_ChefProjet (clé étrangère vers chef_projet)
- ID_client (clé étrangère vers client)
- nb_jours_retard (mesure)
- montant_depassement (mesure)

2. Tables de dimensions

- temps

- Attributs : Trimestre, Semestre, Année

- type_projet

- Attributs : Libellé, Département, Catégorie

- équipe

- Attributs : Département, Nombre de membres, Expérience moyenne des membres

- chef_projet

- Attributs : Nom, Expérience, Nombre de projets menés, Spécialisation

- client

- Attributs : Nom, Type, Localisation

3. Relations

La table de faits projet_terminé est reliée aux dimensions à travers les clés étrangères :

- ID_Temps → temps.ID_Temps
- ID_TypeDeProjet → type_projet.ID_TypeDeProjet
- ID_equipe → équipe.ID_equipe
- ID_ChefProjet → chef_projet.ID_ChefProjet
- ID_client → client.ID_client

4. Structure du cube OLAP

Nom : cubeOlap

Contient :

- Une table de faits : projet_terminé
- Plusieurs dimensions : temps, type_projet, équipe, chef_projet, client

5. Configuration des agrégations

Mesure : **nb_jours_retard**

- **Agrégateur** : SUM
- **Type** : Numérique
- **Objectif** : Calculer le total des jours de retard par projet, équipe, chef de projet, période...

Mesure : **montant_depassement**

- **Agrégateur** : SUM
- **Type** : Numérique
- **Objectif** : Évaluer les dépassements budgétaires totaux selon diverses dimensions (type de projet, client...)

Ce Data Mart permet d'explorer les données sous plusieurs angles d'analyse et de répondre à des besoins décisionnels concrets. Il constitue la base pour la création du **cube OLAP** dans **Pentaho Schema Workbench**, facilitant les requêtes multidimensionnelles.

Exemples d'analyses permises par ce Data Mart :

1. **Évolution annuelle des retards** pour suivre l'amélioration (ou la dégradation) de la gestion des délais.
2. **Performance budgétaire par chef de projet** afin de repérer ceux qui dépassent régulièrement le budget.
3. **Corrélation entre expérience des équipes et dépassement**, pour ajuster l'affectation des projets.
4. **Tendances par type de projet**, pour identifier les catégories à haut risque (en retard ou en dépassement).
5. **Analyse régionale des projets**, via la localisation des clients, pour détecter des zones problématiques.

Constitution du cube Olap sur Pentaho Schema Workbench

- Ci-dessous l'interface de **Pentaho Schema Workbench** :

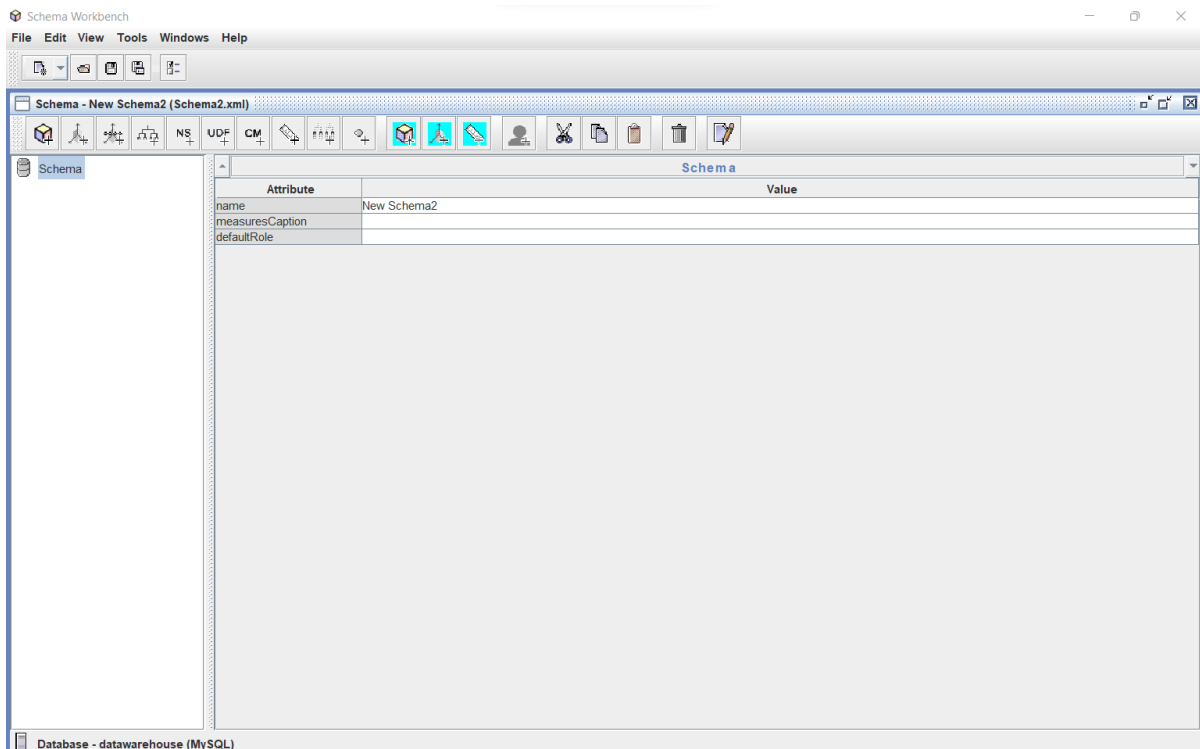


Figure 22: interface de pentaho schema workbench

1. Connexion au datawarehouse :

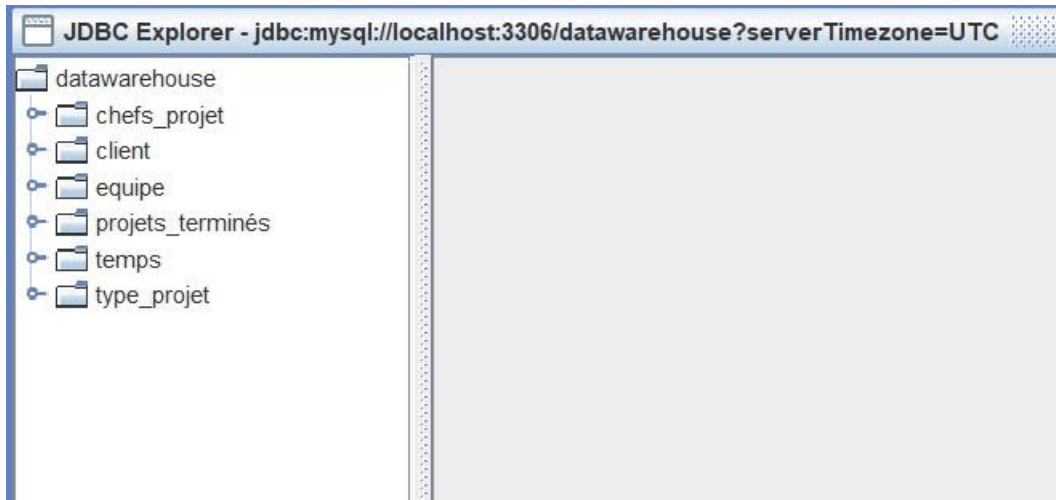


Figure 23: connexion au datawarehouse sur Pentaho

2. Création du cube et définition de la table de faits :

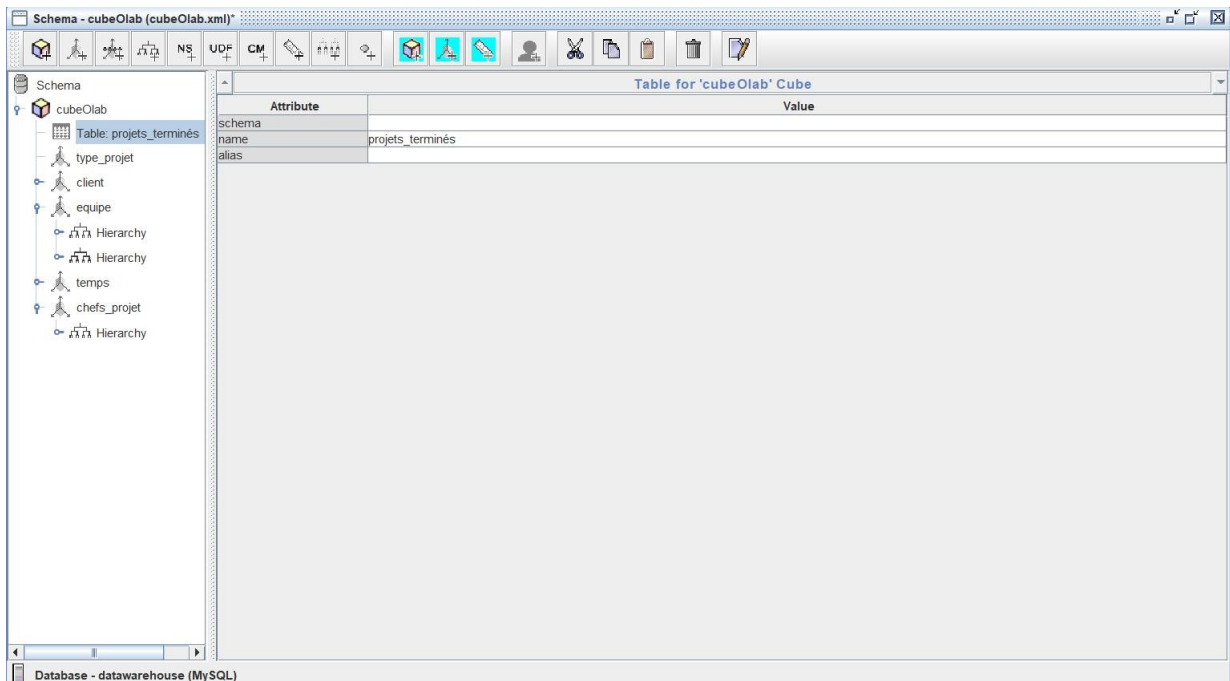


Figure 24: création du cube et définition des tables

3. Définition des dimensions et de leur attributs :

On ajoute plusieurs niveaux a la hiérarchie pour introduire les attributs de la dimension, chaque niveau représente un attribut.

Level for 'equipe' Hierarchy	
Attribute	Value
name	ID_equipe
table	equipe
column	ID_equipe
nameColumn	ID_equipe
parentColumn	
nullParentValue	
ordinalColumn	
type	Integer
uniqueMembers	<input type="checkbox"/>
levelType	Regular
hideMemberIf	
approxRowCount	
caption	
captionColumn	
formatter	

Figure 25: niveaux de hiérarchies pour la dimension équipe

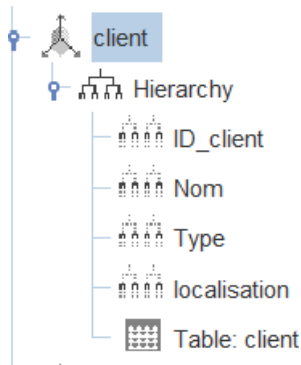


Figure 23: dimension client

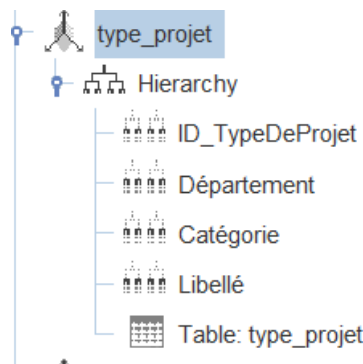


Figure 28: dimension type de projet

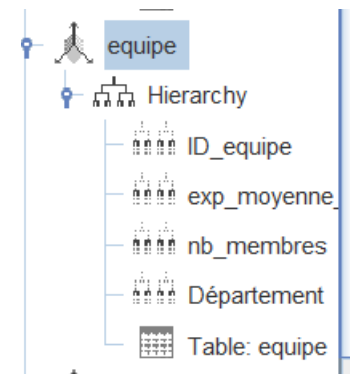


Figure 30: dimension équipe

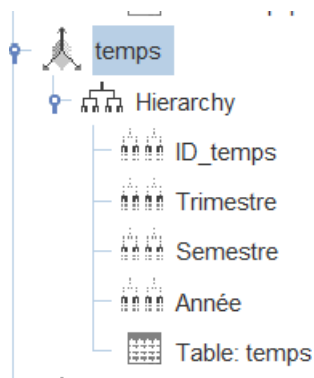


Figure 27: dimension temps



Figure 2C: dimension chef de projet

4. Définition des mesures

Retard jours

Measure for 'cubeOlav' Cube	
Attribute	Value
name	Retard_jours
aggregator	sum
column	Retard_jours
formatString	
visible	<input checked="" type="checkbox"/>
datatype	Integer
formatter	
caption	

Figure 31: mesure retard_jours

Dépassement budget

Measure for 'cubeOlav' Cube	
Attribute	Value
name	dépassement budget
aggregator	sum
column	dépassement_budget
formatString	
visible	<input checked="" type="checkbox"/>
datatype	Numeric
formatter	
caption	

Figure 32: mesure dépassement budget

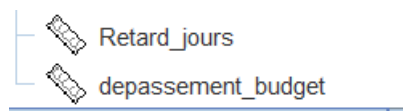


Figure 33: les mesures de notre table de faits

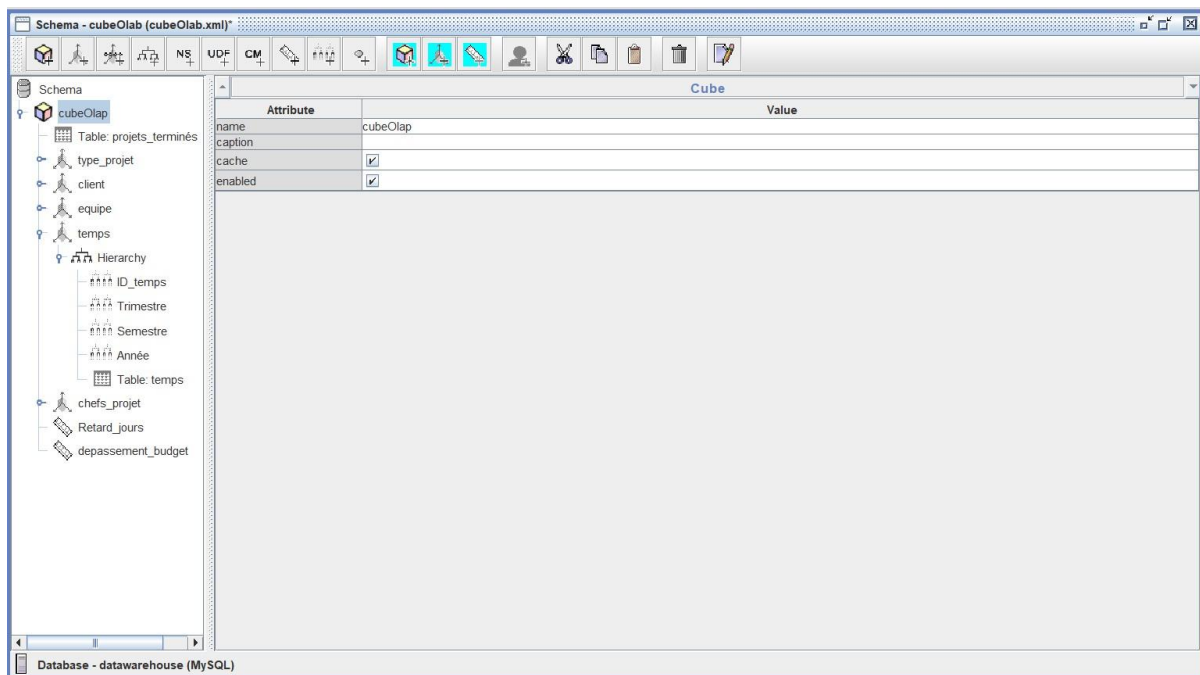


Figure 34: cube olap finalisé

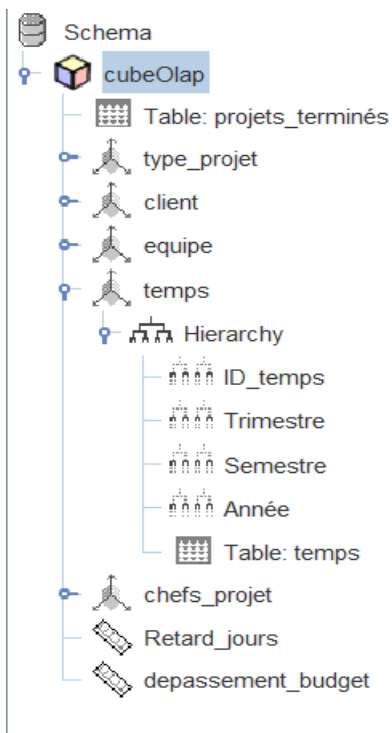


Figure 35: structure de notre cube Olap

Notre **cube olap** est prêt. Plus qu'à le générer et le sauvegarder sous forme de fichier XML.

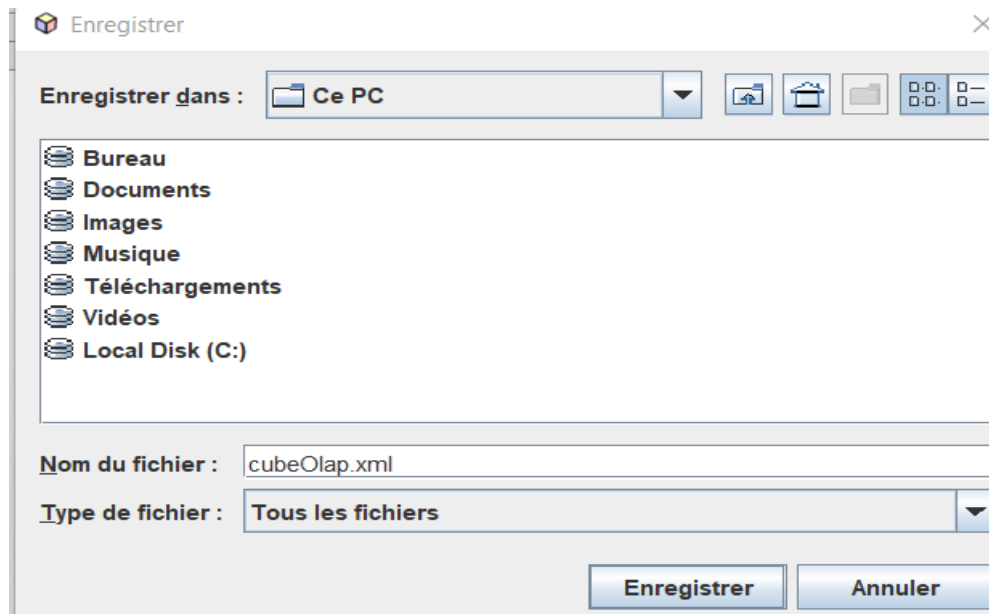


Figure 3C: sauvegarde du cube Olap

8. Analyse et visualisation des données

L'analyse des données issues du Data Mart permet de visualiser les retards et dépassements budgétaires des projets selon plusieurs dimensions (équipe, chef de projet, type de projet, etc.).

Des tableaux de bord interactifs créés avec **Power BI** faciliteront l'interprétation des indicateurs clés car ces visualisations offrent aux décideurs une vue synthétique pour optimiser la gestion et le suivi des projets.

Power BI est un outil de business intelligence de Microsoft qui permet de :

- Analyser et visualiser des données à partir de différentes sources, créant des rapports interactifs.
- Faciliter la prise de décisions grâce à des visualisations claires et des tableaux de bord dynamiques.



Figure 37: logo de Power Bi

Nous allons commencer par établir une **connexion de Power BI à la Base de Données MySQL "datawarehouse"**, ci-dessous les étapes :

1. Ouvrir Power BI Desktop

- Lancer Power BI Desktop pour créer un nouveau rapport.

2. Se connecter à MySQL

- Cliquer sur 'Obtenir les données' dans le ruban.
- Sélectionner MySQL Database comme source de données.

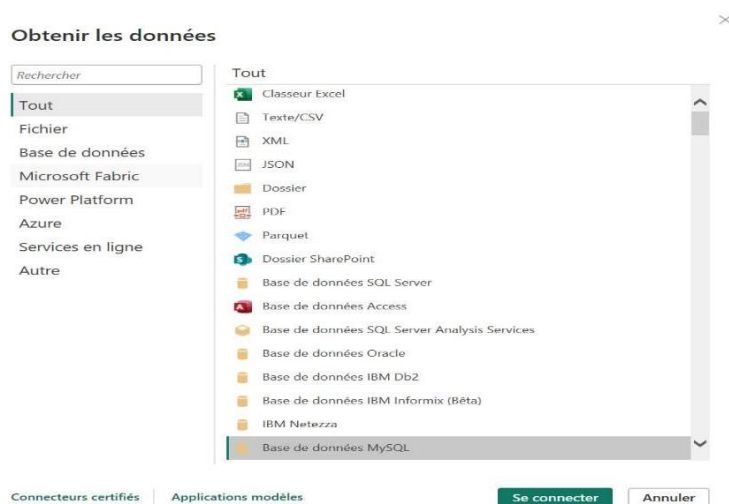


Figure 38: obtention des données Power Bi

3. Entrer les informations de connexion

- Dans la fenêtre qui apparaît, entrer **localhost** comme serveur.
- Entrer **datawarehouse** comme nom de la base de données

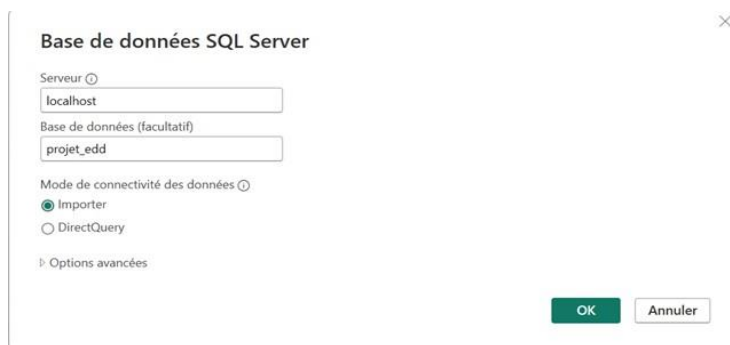


Figure 33: connexion à la bdd sql server

Choisir le mode d'authentification, puis entrer le nom d'utilisateur et le mot de passe.

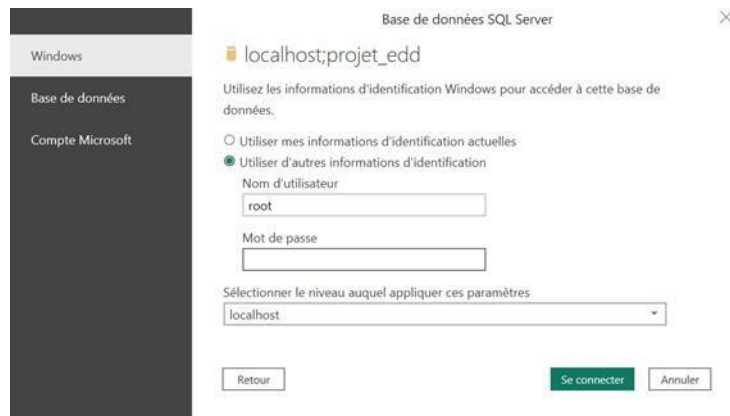


Figure 40: identifiants de la bdd

4. Sélectionner les données

- Une fois connecté, Power BI affiche un aperçu des tables existantes dans la base "datawarehouse".
- Sélectionner les données à importer pour les analyser.

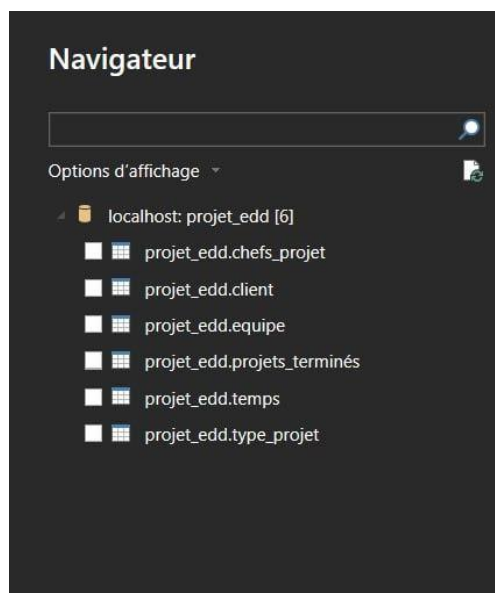


Figure 41: sélection des tables du datawarehouse sur Power Bi

5. Charger les données

- Cliquer sur Charger pour importer les tables choisies dans Power BI.

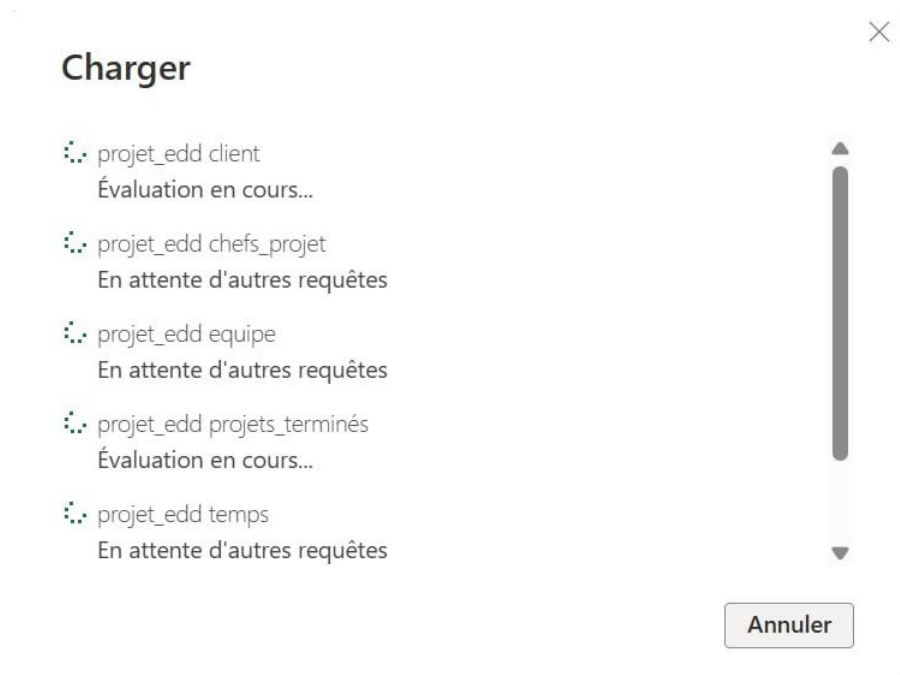


Figure 42: Chargement des données du datawarehouse sur Power Bi

6. Création des visualisations :

Chaque datamart a fait l'objet de visualisations spécifiques permettant de mettre en évidence les indicateurs clés de performance liée aux projets. Ces graphiques interactifs facilitent l'analyse des retards, des budgets, des types de projets et de la répartition géographique des clients.

- **Graphiques en barres et en secteurs** ont été utilisés pour illustrer des répartitions telles que la part des projets par catégorie, le type de client ou encore les chefs de projet les plus sollicités.
- **Graphiques en courbes** permettent de visualiser les tendances annuelles des retards ou des dépassements budgétaires, afin de mieux comprendre l'évolution de la performance des projets dans le temps.
- ...

Analyses :

1) Le retard des projets selon les chefs de projets

Axes de Reporting :

- Chefs_projet[Nom]
- projets_terminés[Retard_jours]

But :

Analyse des **retards cumulés** selon les différents **chefs de projets**. Elle vise à identifier les profils les plus exposés aux retards de livraison des projets réalisés.

L'histogramme présente la **somme des jours de retard** (Retard_jours) pour chaque **chef de projet** (Nom_ChefProjet), ce qui permet une comparaison directe de leur performance en matière de respect des délais.

Graphiques :

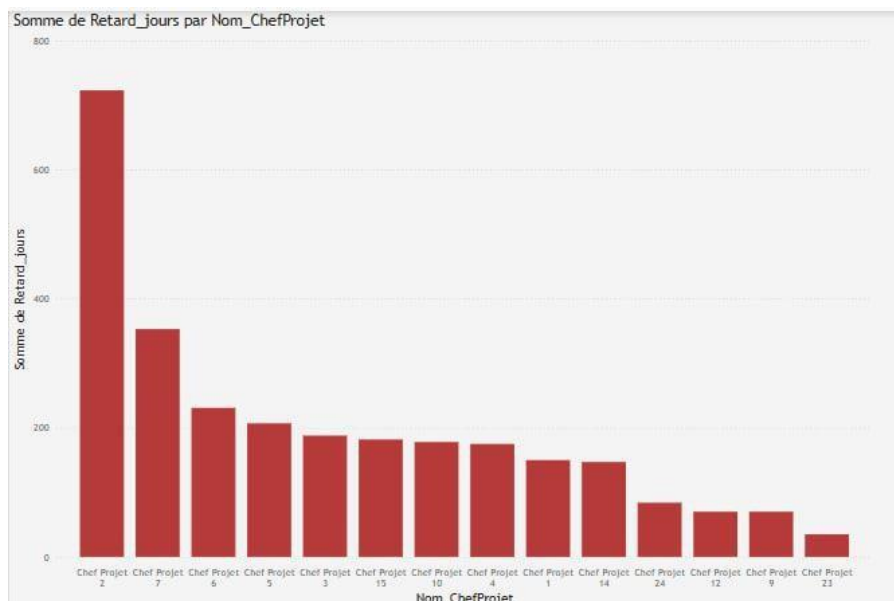


Figure 43: Histogramme empilé retard_jours par Nom chef de projet

Observation :

Disparité importante entre certains chefs : cela peut refléter soit des différences de charge de travail, de complexité des projets, soit des écarts de compétence en gestion de projet ou d'organisation.

On remarque que le chef de projet 2 est celui qui a eu le plus de retards comparés aux autres.

Le chef de projet 7 se démarque aussi. Pour le reste des chefs de projets, ils ont engendré des retards similaires.

2) Dépassement de budget par département :

Axes de Reporting :

- Type_projet[Département]
- projets_terminés[dépassement_budget]

But :

Ce datamart évalue les performances financières par département en mesurant le dépassement budgétaire total des projets terminés.

Graphiques :

Somme de dépassement_budget par Département

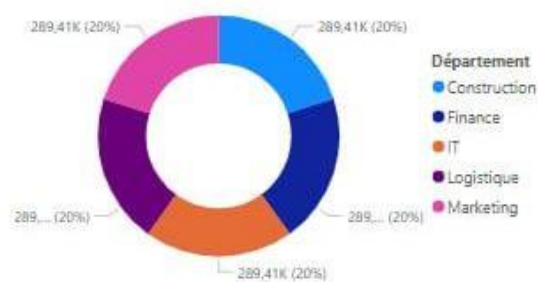


Figure 44: Graphique en anneau représentant la répartition du dépassement de budget par département

Observation :

L'analyse montre une répartition égale du dépassement budgétaire entre les cinq départements — Construction, Finance, IT, Logistique et Marketing — chacun représentant 20 % du total (soit environ **28G,41K**).

Cela indique que le dépassement budgétaire est uniforme entre les départements, suggérant une gestion budgétaire similaire ou une répartition équilibrée des projets à risque.

3) Retard par Année et Trimestre :

Axes de Reporting :

- Temps[Année,Trimestre]
- projets_terminés[Retard_jours]

But :

Cette analyse permet d'évaluer les retards cumulés des projets par période temporelle, afin d'identifier les pics de retard et d'ajuster la planification ou les ressources en conséquence.

Graphiques :



Figure 45: Histogramme de la somme des retards par année et trimestre

Observation :

En 2023, le 1er trimestre concentre à lui seul la majorité des retards (plus de 2000 jours), ce qui indique une **forte perturbation en début d'année**. Les trimestres suivants affichent des retards nettement inférieurs, avec un pic secondaire au 4e trimestre.

Cela suggère soit une amélioration progressive dans la gestion des projets, soit une variation saisonnière des contraintes ou des charges de travail.

4) Le retard par nombre de projets menés par les chefs :

Axes de Reporting :

- Chefs_projet[nb_projet_ménés]
- projets_terminés[Retard_jours]

But :

Mise en relation entre le nombre de projets menés auparavant par le chef chargé du projet en question ainsi que le retard cumulé sur ces projets.

Graphiques :

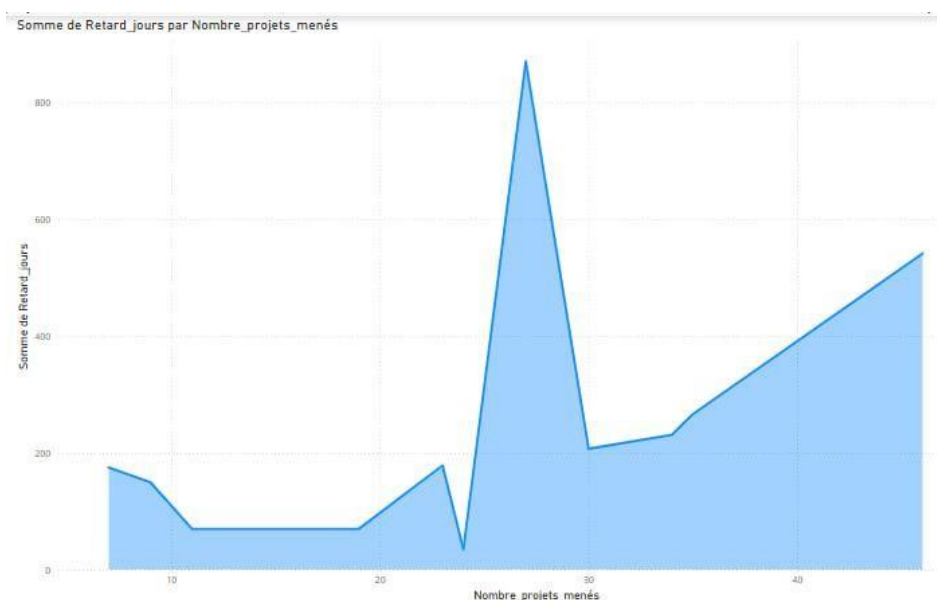


Figure 4C: Graphique en aires (area chart) de la somme des retards par nombre de projets menés par le chef de projet

Observation :

Le graphique montre un pic marqué de plus de **800 jours de retard** pour les chefs de projet ayant mené entre 20 et 30 projets. Cela suggère que cette plage d'activité pourrait correspondre à une surcharge de travail ou à une complexité accrue des projets, entraînant davantage de retards.

5) Dépassement de budget par chef de projet :

Axes de Reporting :

- Chefs_projet[Nom]
- projets_terminés[dépassement_budget]

But :

Analyse comparative des **dépassements budgétaires** attribuables à chaque chef de projet, mettant en lumière les performances individuelles.

Graphiques :

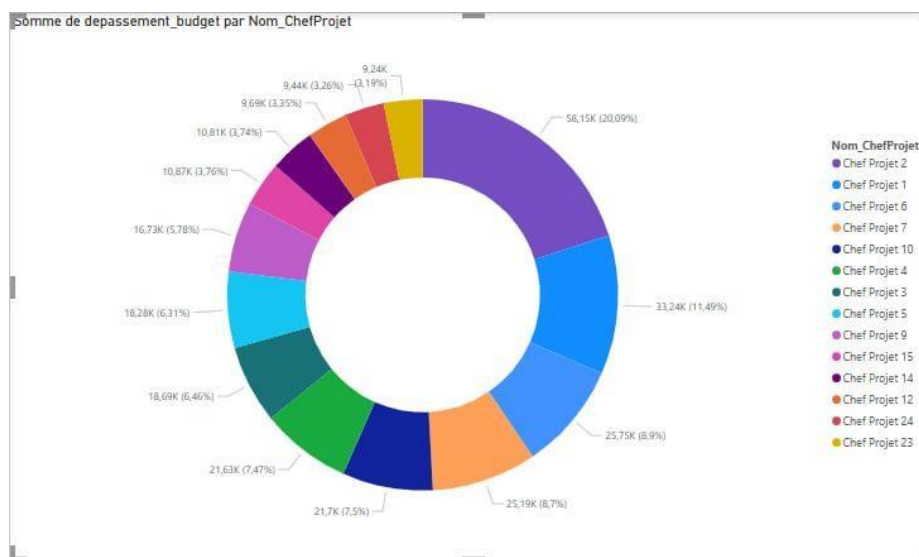


Figure 47: Graphique en anneau de dépassement budget par chef de projet

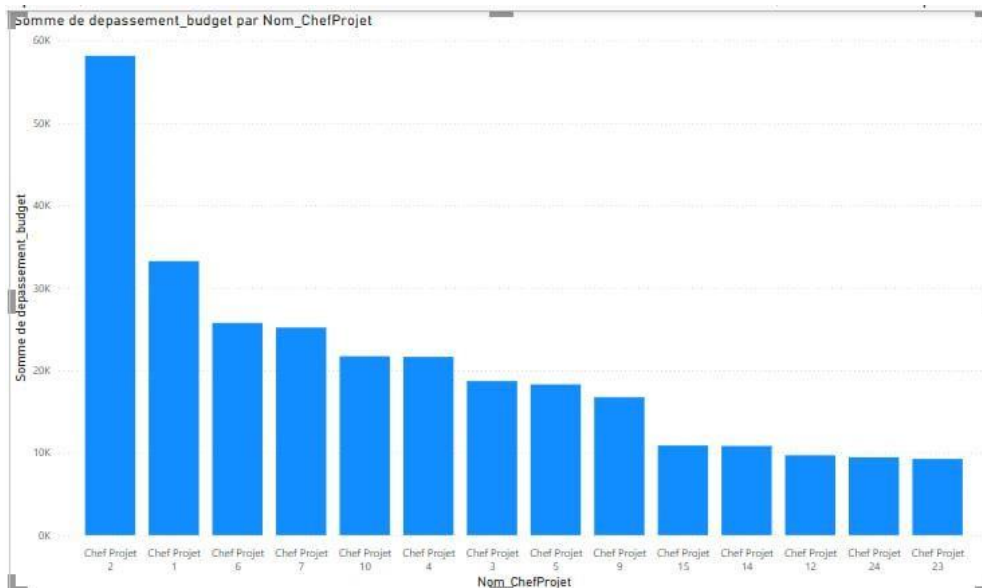


Figure 48: Histogramme de dépassement budget par chef de projet

Observation :

Le **Chef Projet 2** se distingue par un dépassement budgétaire **anormalement élevé (58,15K)** comparé à la moyenne des autres chefs, nécessitant une analyse approfondie des causes spécifiques à ses projets.

6) Dépassement de budget par nombre de projets menés par le chef:

Axes de Reporting :

- Chefs_projet[nb_projet_ménés]
- projets_terminés[dépassement_budget]

But :

Mise en relation entre le nombre de projets menés par le chef de projet ainsi que le dépassement budgétaire sur ces projets.

Graphiques :

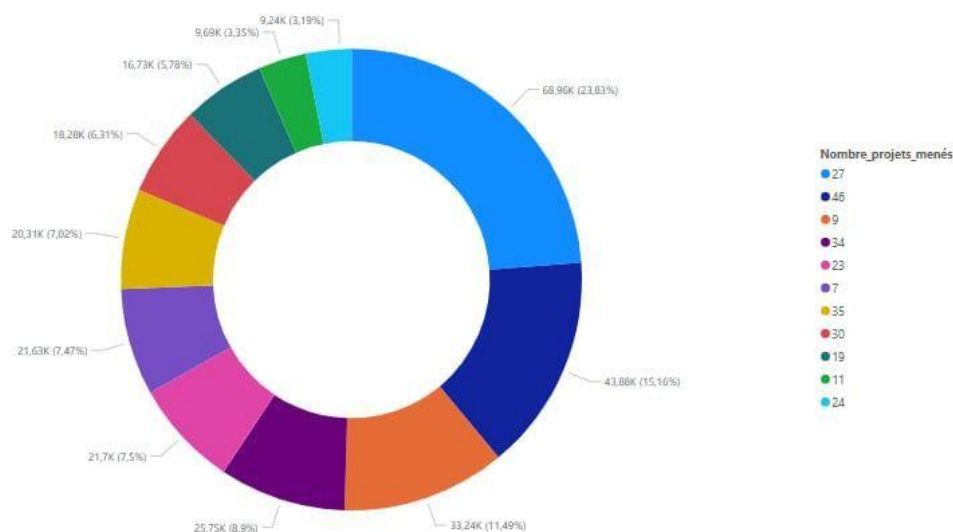


Figure 43: Graphique circulaire de dépassement budget par nombre de projets menés par le chef de projet

Observation :

Le graphique met en évidence une **variation du dépassement budgétaire** en fonction du nombre de projets menés.

Chaque palier de projets montre un niveau différent de dépassement, avec un pic significatif observé pour **27 projets menés**, représentant le plus haut niveau de dépassement budgétaire enregistré.

Cela peut refléter une charge de travail excessive ou une mauvaise planification budgétaire à ce niveau d'expérience.

7) Dépassement de budget par année et trimestre :

Axes de Reporting :

- Temps[Année,Trimestre]
- projets_terminés[dépassement_budget]

But :

Cette analyse permet d'évaluer le **dépassement budgétaire** des projets par année et trimestre, afin d'identifier les périodes critiques en termes de dérapage financier. Elle aide à anticiper les besoins d'ajustement budgétaire, à renforcer le suivi financier, et à améliorer la gestion des coûts au fil du temps.

Graphiques :

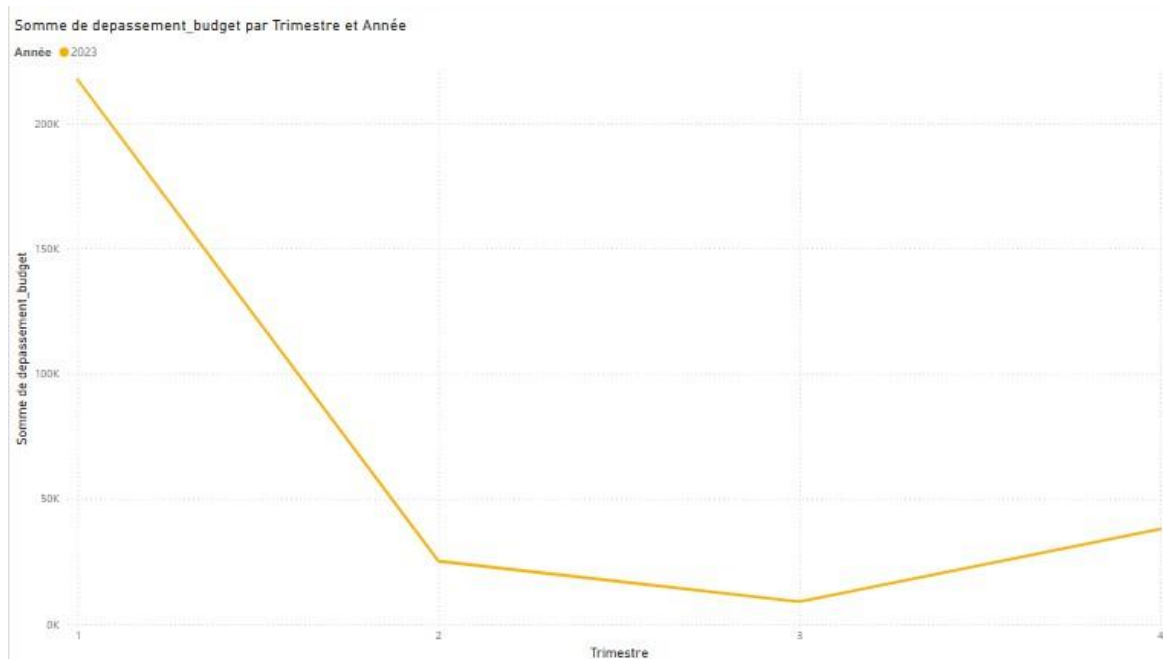


Figure 50: Graphique de courbe de dépassement de budget par année (2023) et trimestre

Observation :

En 2023, Le dépassement budgétaire a atteint un pic important au 1er trimestre (environ **200K**), puis diminue progressivement au fil des trimestres suivants, indiquant une meilleure maîtrise budgétaire au cours de l'année.

Conclusion générale

Ce projet nous a permis de mettre en œuvre de manière concrète les différentes étapes de conception et d'exploitation d'un entrepôt de données dans un contexte réel de gestion de projets. En partant de sources de données hétérogènes (CSV, Excel, JSON, base de données relationnelle), nous avons appliqué un processus ETL rigoureux via **Talend** afin d'unifier, nettoyer et structurer les données autour d'un modèle multidimensionnel cohérent.

L'élaboration d'un schéma en étoile, la définition d'une table de faits pertinente (projets terminés) ainsi que des dimensions adaptées (temps, équipe, chef de projet, client, type de projet) ont permis de construire un Data Mart robuste.

Ce dernier a servi de base à la création d'un cube OLAP via **Pentaho Schema Workbench**, ouvrant la voie à des analyses multidimensionnelles efficaces.

Les visualisations réalisées sous **Power BI** ont mis en évidence des indicateurs clés tels que les retards et les dépassements budgétaires. Elles ont permis de dégager des tendances, d'identifier les points faibles dans la gestion des projets (chefs de projets fréquemment en retard, départements avec des dépassements récurrents, périodes critiques de l'année, etc.) et d'apporter un appui concret à la prise de décision.

Ce projet démontre l'importance stratégique des entrepôts de données et des outils d'aide à la décision dans l'optimisation de la performance organisationnelle. Il constitue également une base solide pour l'extension future du système, notamment par l'ajout de nouveaux indicateurs ou l'intégration d'autres types de projets.

Sources

- Chapitre 1 : Entrepôts de données -Dr Menai
- Chapitre 2 : Modélisation Multidimensionnelle -Dr Menai
- [L'importance de la data pour une entreprise : enjeux et perspectives - Jean Mauferon Conseil](#)
- [Créer son premier Data Warehouse : Le guide complet pour les débutants](#)
- [Approches de mise en place d'un Data warehouse – Decizia](#)
- [How to Build a Data Warehouse from Scratch: Cost + Examples | Airbyte](#)
- [Tutoriel Talend – Qu'est-ce que Talend ETL Tool ?](#)
- [Talend Tutorial](#)
- [Création de job Talend](#)
- [WP EN DI Talend DefinitiveGuide DataIntegration.pdf](#)
- [Data WareHouse Project - Entrepot De Données - YouTube](#)
- [Pentaho Mondrian Documentation](#)
- [Create a Mondrian schema • Pentaho Schema Workbench • Reader • Hitachi Vantara Documentation Portal](#)
- [Schema Workbench](#)
- [\[https://static.packt-cdn.com/downloads/LearningPentahoDataIntegration8CE_ColorImages.pdf\]\(https://static.packt-cdn.com/downloads/LearningPentahoDataIntegration8CE_ColorImages.pdf\)](#)
- [Microsoft Power BI — Wikipédia](#)
- [Power BI : Le Guide Ultime | Tutoriel complet pour débutants](#)
- [Tuto Power BI : apprendre à utiliser l'outil BI de Microsoft](#)
- [<https://thetacdn.blob.core.windows.net/assets/2021:PowerBIGuides/Power%20BI%20for%20Beginners%20-%202020.pdf>](#)
- [Découverte des Visuels dans Power BI de Microsoft](#)