

Université Hassan 1^{er}
Ecole Nationale des Sciences Appliquées de Berrechid
Département de mathématique et informatique
Filière : Ingénierie des Systèmes d'Information et BIG DATA
Module : Système Business Intelligence
Semestre : S8

Rapport de Projet

Transport

Equipe de travail :

Yassine EL ANSSARI

Ayoub LOUJA

Encadré Par :

Hamid HRIMISH

Année Universitaire : 2021/2022

Résumé

Ce présent rapport est le résultat du travail sur le projet de fin de semestre que nous avons réalisé pour le module Business Intelligence. Ce projet visait à suivre le processus BI en vue de l'appliquer pour résoudre la problématique que vous allez la trouver ci-dessous. Notre mission a commencé par comprendre le concept du BI, tout en l'exploitant afin de trouver une solution concrète à notre problématique

Ce rapport vise donc à décrire d'une manière bien détaillée toutes les étapes de la solution qu'on a adopté pour résoudre la problématique suivante :

Comment peut-on analyser l'état de transport dans la ville Montréal de Canada à travers les horaires de bus / métro et les différentes stations ?

Ce présent rapport est organisé comme suit : la première section présente une introduction générale, le contexte général, ainsi de la problématique, les outils utilisés et les étapes de la résolution de ce problème sera abordée dans la deuxième section. Enfin, la dernière section contient le résultat de notre solution.

Table des matières

Résumé.....	2
Introduction.....	4
Contexte général	5
Problématique	6
Présentation de la solution	7
I. Outils et logiciels utilisés :.....	7
II. Source des données :	8
III. Conception du datawarehouse :	8
1. Les tables des dimensions :	9
2. Table de faits :	9
IV. Processus ETL :	10
V. Création du Dashboard :	15
Conclusion	18
Références	19

Liste des figures

Figure 1 : Extrait du dataset.....	8
Figure 2 : Schème du Datawarehouse.....	10
Figure 3 : Process de chargement des données dans SQL SERVER.....	11
Figure 4 : Process ETL - alimentation du Datawarehouse	15
Figure 5 : Tableau de bord	16

Introduction

Depuis longtemps, l'analyse de l'état de transport connaît des difficultés à cause de l'accès difficile de l'information, Aujourd'hui et avec la révolution numérique que le monde a connue récemment, un nouveau domaine apparaît, Le Business intelligence ou informatique décisionnel qui désigne l'ensemble des technologies permettant au toute organigramme d'analyser leurs données au profit de leurs prises de décisions. Ce dernier, s'appuie sur les tableaux de bord qui constitue un outil incontournable pour les systèmes d'aide à la décision. Les tableaux de bord assistent les décideurs dans n'importe quel niveau de l'organisme. Ainsi, Ils supportent l'interaction avec les données, comme la fouille dans les détails profonds pour tirer l'information utile. En effet, Ce projet vise à implémenter cet instrument dans le but de créer un système décisionnel tout en partant par l'analyse de besoin, passons par le process ETL jusqu'à l'arrivée à la création du Tableau de Bord.

Contexte général

Ce projet porte sur la création d'un tableau de bord interactif qui permet de visualiser l'état de transport inclus le temps de départ des bus et métro et aussi le nombre de stop par station et d'autres critères . ce travail est dirigé par la société de transport de Montréal (STM) dans les différentes régions de Montréal . les données générées depuis l'API GOOGLE TRANSPORT également connu sous le nom de **GTFS** statique ou de flux statique pour le différencier de l'extension **GTFS-realtime**, définit un format de fichier commun pour les horaires de transports en commun et les informations géographiques associées. et pour faciliter l'analyse il faut que ces données soient stockées, gérées et d'être présentées d'une manière qui facilite l'analyse afin d'extraire la bonne information et prendre par la suite la bonne décision.

Problématique

Tout état à l'intérêt de savoir les stations qui connaissent de gros arrêtes dans les différentes jours de la semaines et les vacances , cela consiste à exploiter les données générées afin de les explorer, faciliter et optimiser l'analyse, cela nous mène à poser la problématique suivante :

Comment peut-on mesurer et analyser l'état de transport dans les différentes régions de Montréal ?

Présentation de la solution

I. Outils et logiciels utilisés :



Microsoft Excel est un logiciel tableur de la suite bureautique Microsoft Office développé et distribué par l'éditeur Microsoft. La version la plus récente est Excel 2019. Il est destiné à fonctionner sur les plates-formes Microsoft Windows, Mac OS X, Android ou Linux....



SQL Server Management Studio (SSMS) est un environnement intégré pour gérer toute infrastructure SQL, de SQL Server à Azure SQL Database. SSMS fournit des outils pour configurer, surveiller et administrer des instances de SQL Server et des bases de données.



SSIS est une plate-forme pour les applications d'intégration de données et de workflow. Il dispose d'un outil d'entrepôt de données utilisé pour l'extraction, la transformation et le chargement des données. L'outil peut également être utilisé pour automatiser la maintenance des bases de données SQL Server et les mises à jour des données de cube multidimensionnel.



Microsoft Power BI est une solution d'analyse de données de Microsoft. Il permet de créer des visualisations de données personnalisées et interactives avec une interface suffisamment simple pour que les utilisateurs finaux créent leurs propres rapports et tableaux de bord.

II. Source des données :

Pour réaliser ce projet, on était censé de trouver des fichiers plats liée avec la thématique de notre sujet, provenant de différentes sources. Les données ont été pris depuis le site web officielle de la société de transport de Montréal (STM) <https://www.stm.info/> .

Après que les données sont transférer a SQL server :

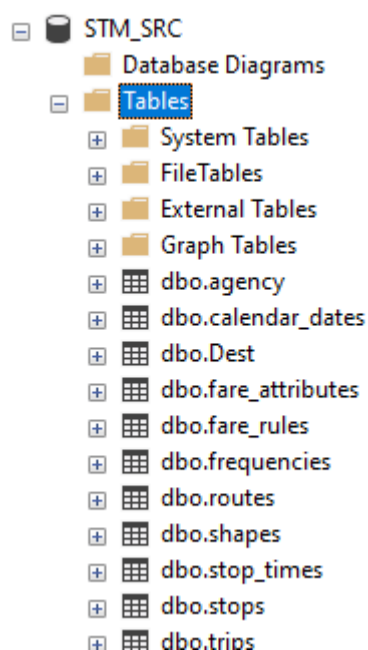






Figure 1 : Extrait du base de données

III. Conception du datawarehouse :

La conception de notre datawarehouse s'appuie sur le schéma en flocon, et qui se compose d'une seule table de fait, et de 4 tables des dimensions .

1. Les tables des dimensions :

-  *Table fare* : Ce table est utilisé comme une dimension afin de définir prix du billet et la catégorie de billet.
-  *Table Line de Buse* : Ce table des information sur la line de bus/métro ,la Direction l'orientation, le lien de bus ...;
-  *Table Stop* : La table contient les différentes stations des bus aussi leurs emplacements .
-  *Table ID service* : ce table il vous permet de définir le jour exacte si il s'agit d'un jour de weekend ou bien un jour de vacance ou n'importe quelle jour de la semaine ainsi la date .

2. Table de faits :

La table des faits est le tableau stop sequence , où se porte l'analyse, il contient les mesures représentées par nombre de stop séquences et temps de départ des bus /métro .

Et voici ci-dessus le schéma de notre entrepôt de données :

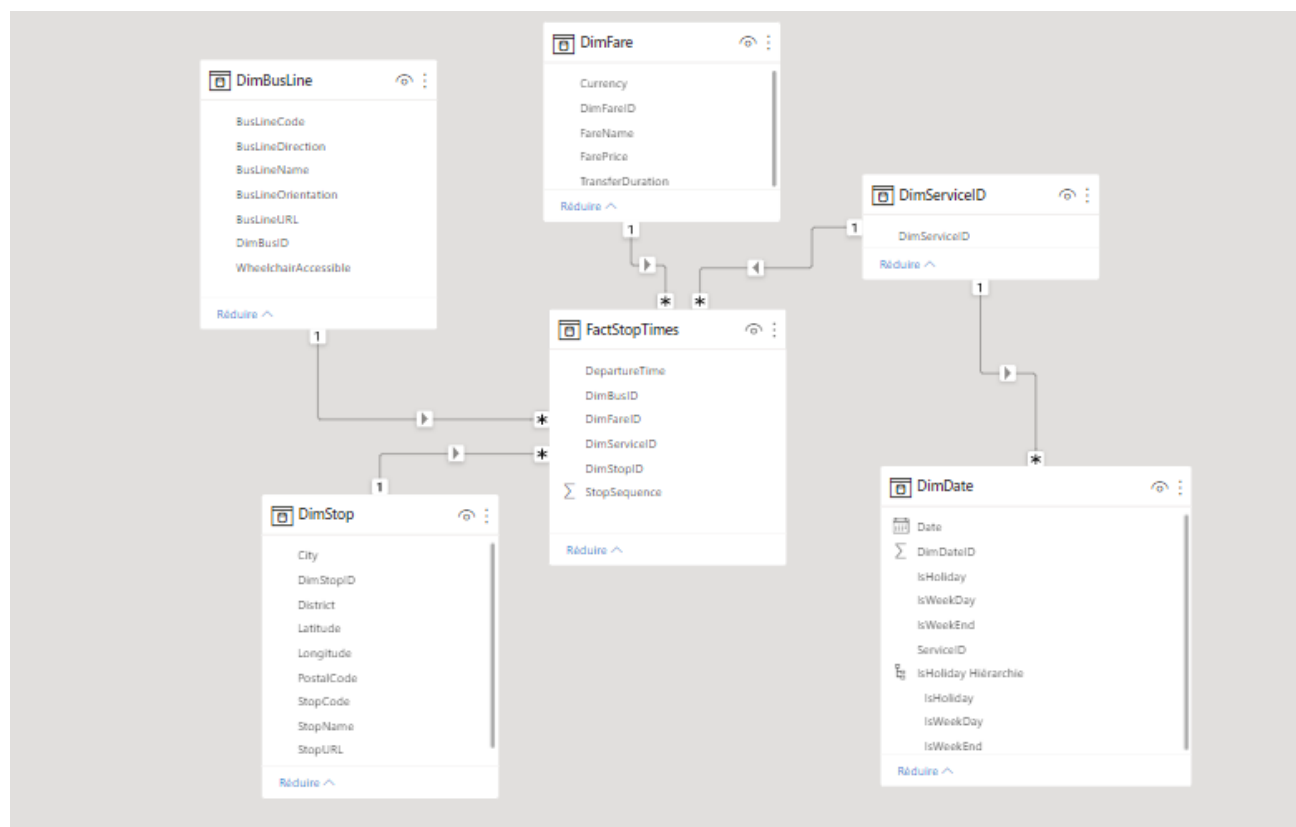


Figure 2 : Schème du Datawarehouse

IV. Processus ETL :

L'objectif de cette étape, est d'alimenter notre entrepôt de données suivant trois étapes, premièrement il faut extraire les données provenant des différentes sources, Ensuite appliquer à ces données les opérations de transformation, et ensuite les charger dans l'entrepôt des données.

Pour notre cas on a commencé la première partie par charger tous les fichiers que nous avons dans une source de données déjà créée dans la base de données s'appelle STM_SRC. Ci-dessous le process ETL qu'illustre le process de chargement des données brutes dans la base de données :

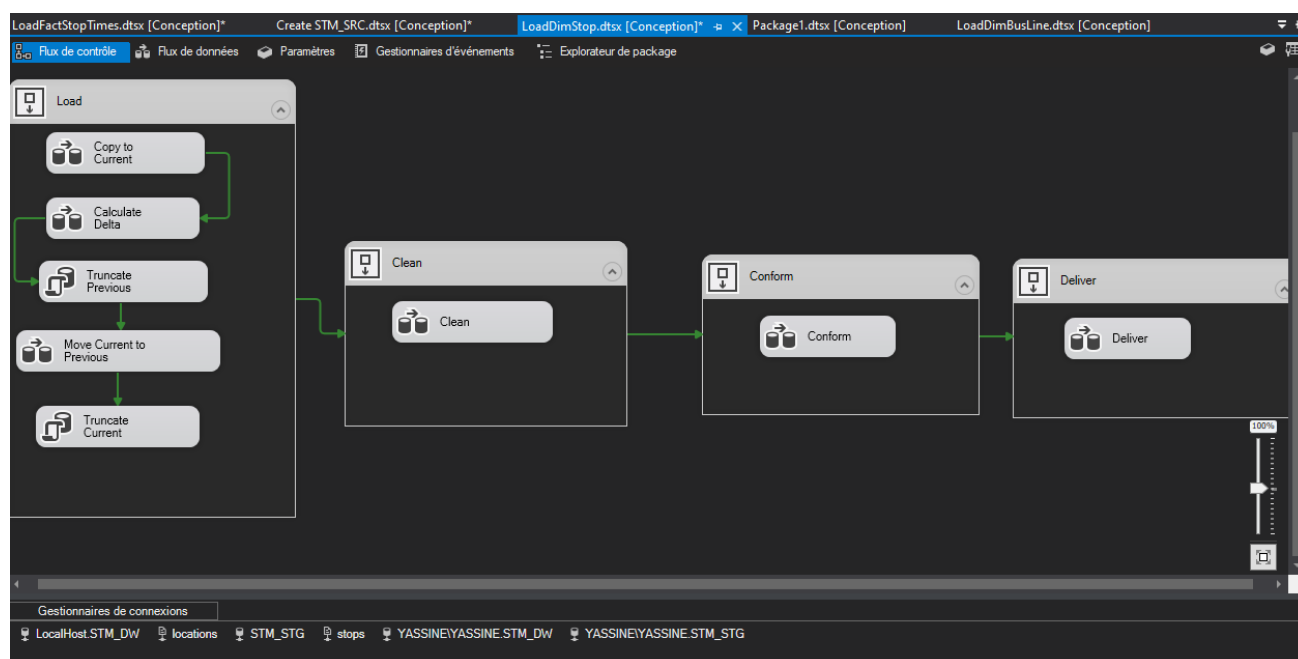


Figure 3 : Process de chargement des données dans SQL SERVER

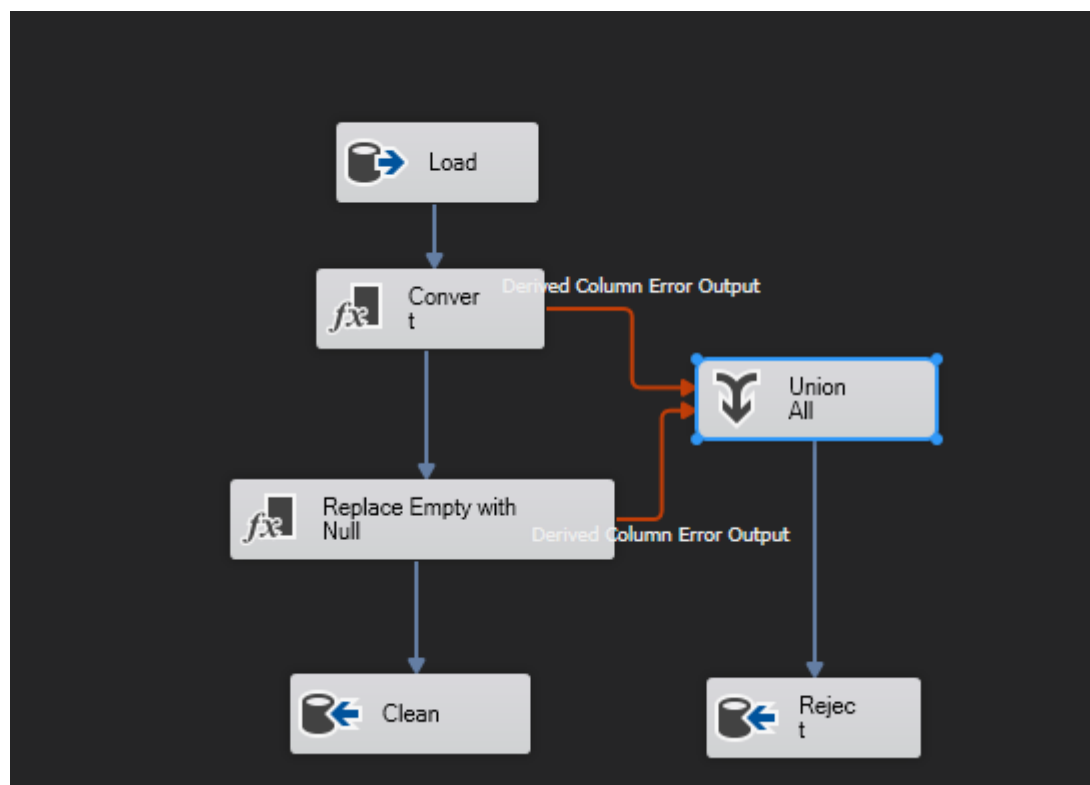
Une fois terminé, la base de données fraîchement créée et alimentée par les données sera exploitée pour alimenter les différentes tables de notre entrepôt.

Après il faut charger les données dans les tables de dimensions mais avant ça puisque nos données n'est pas prêtes alors on a utilisé plusieurs transformations et data cleaning et aussi une étape de conformité pour chaque dimension.

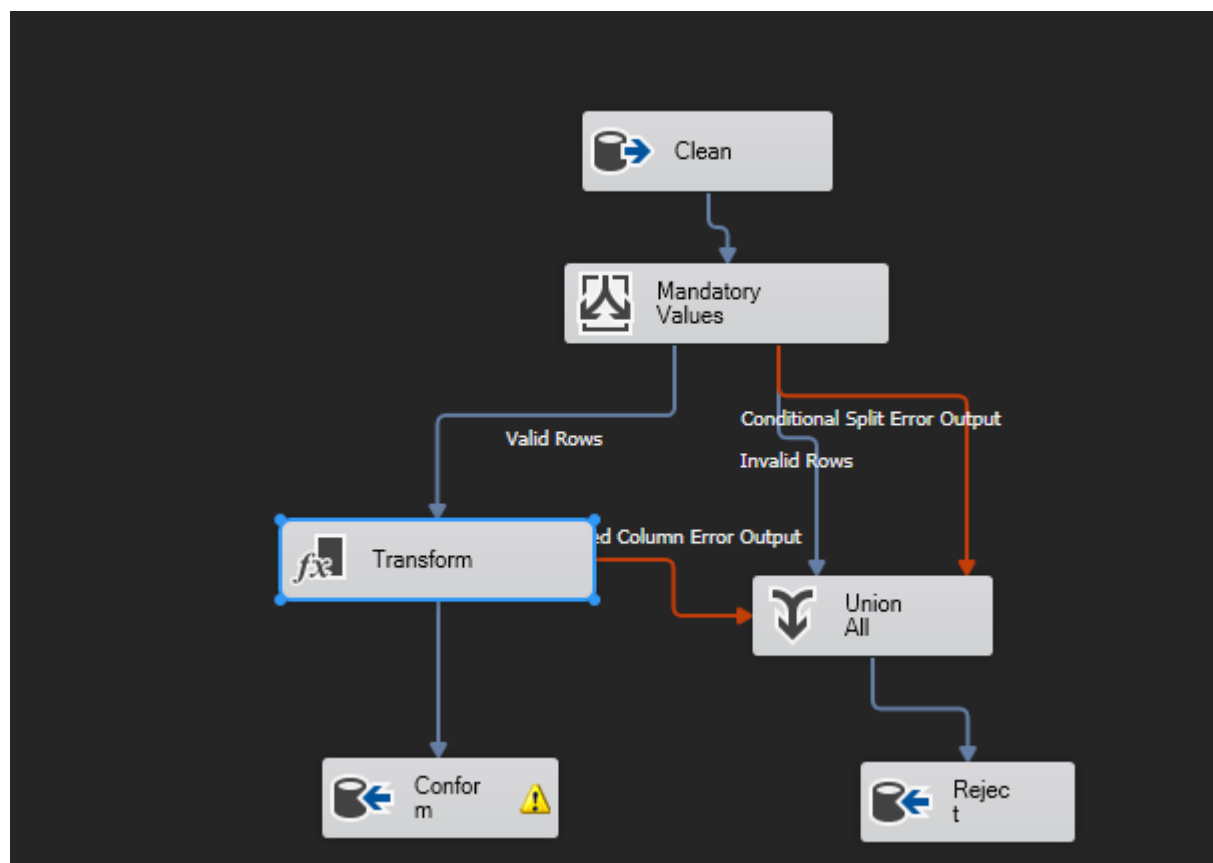
Par exemple pour le chargement de dimension stop voici le flux de contrôle:



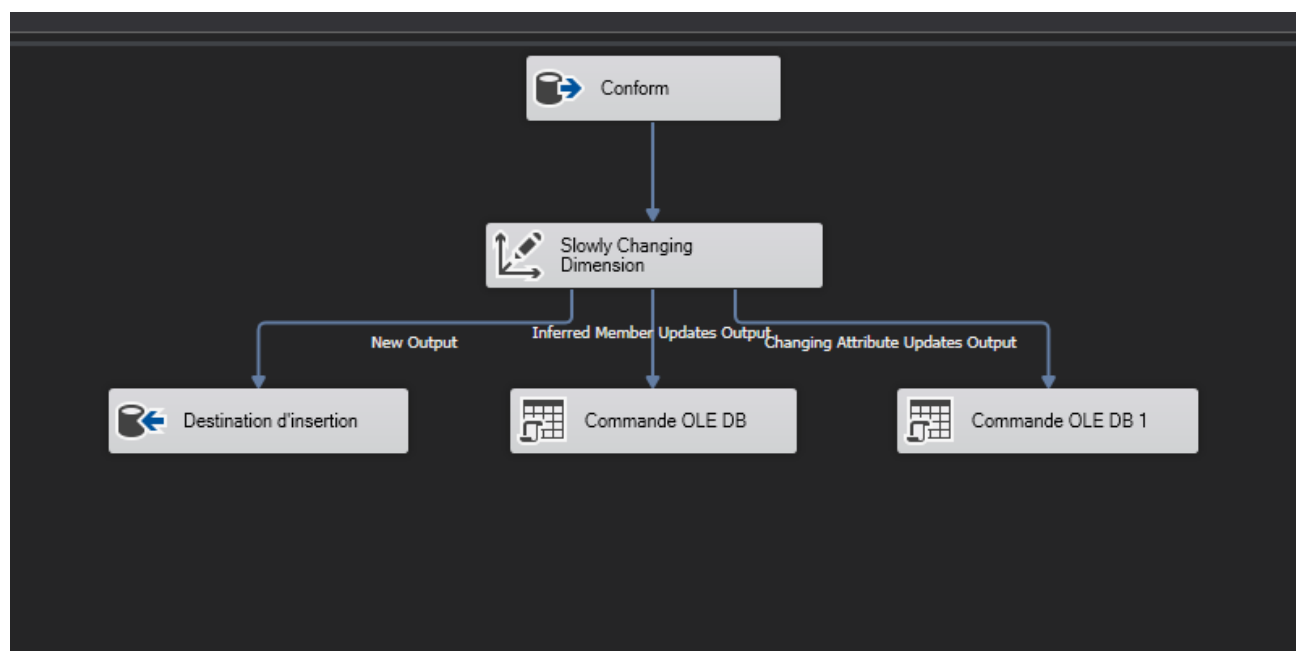
Pour le flux de données clean on a défini un nouveau colonne dérivé et aussi on a remplacer les valeur vides par null pour la bonne décision et en cas d'erreur lors de la transformation un nouveau tableau rejet contient les données refusées va être crée et un autre pour les données après le cleaning (clean stop) :



Après le clean on a le flux de données conform pour d'autres transformation tq la division de champs district en sous champs a travers la fonction substring



Et finalement on a chargé la dimension stop dans le flux de données deliver :



les mêmes flux de données pour toutes les autres tables .

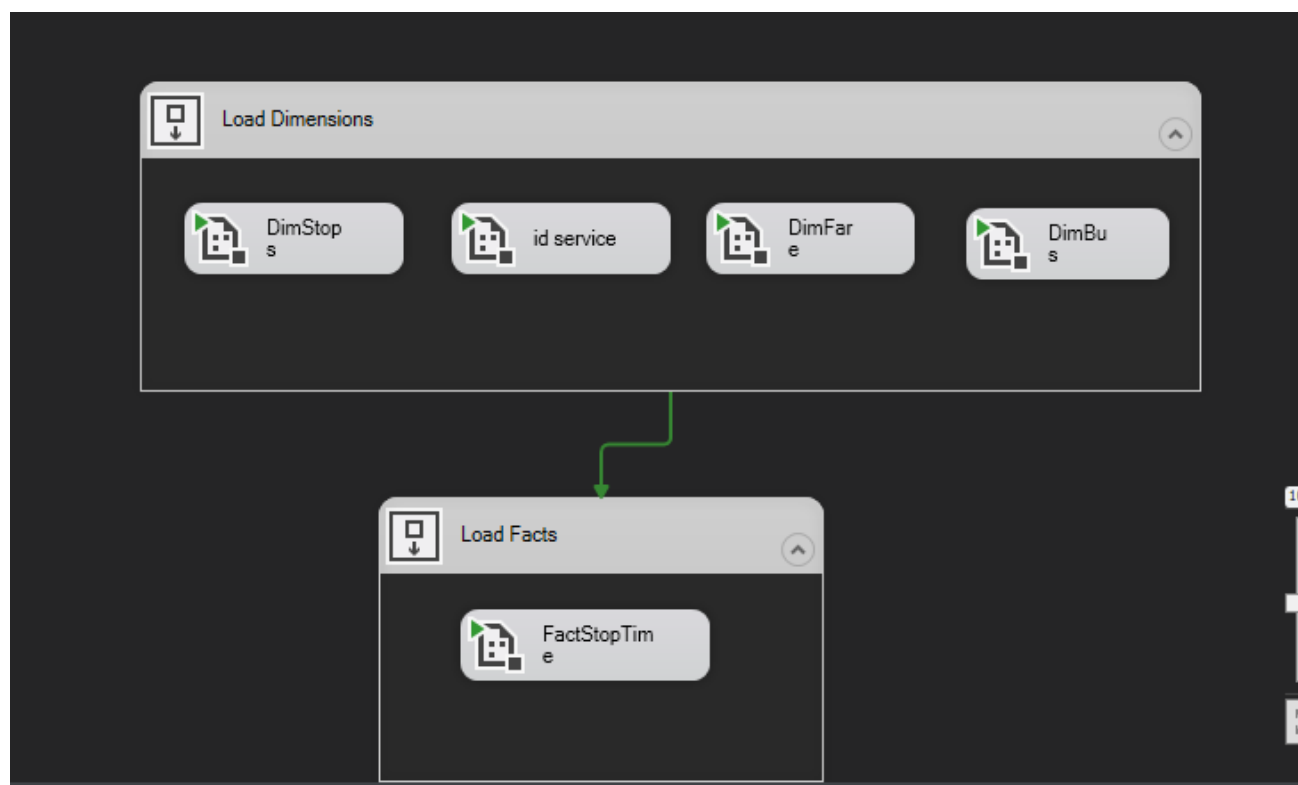


Figure 4 : Process ETL - alimentation du Datawarehouse

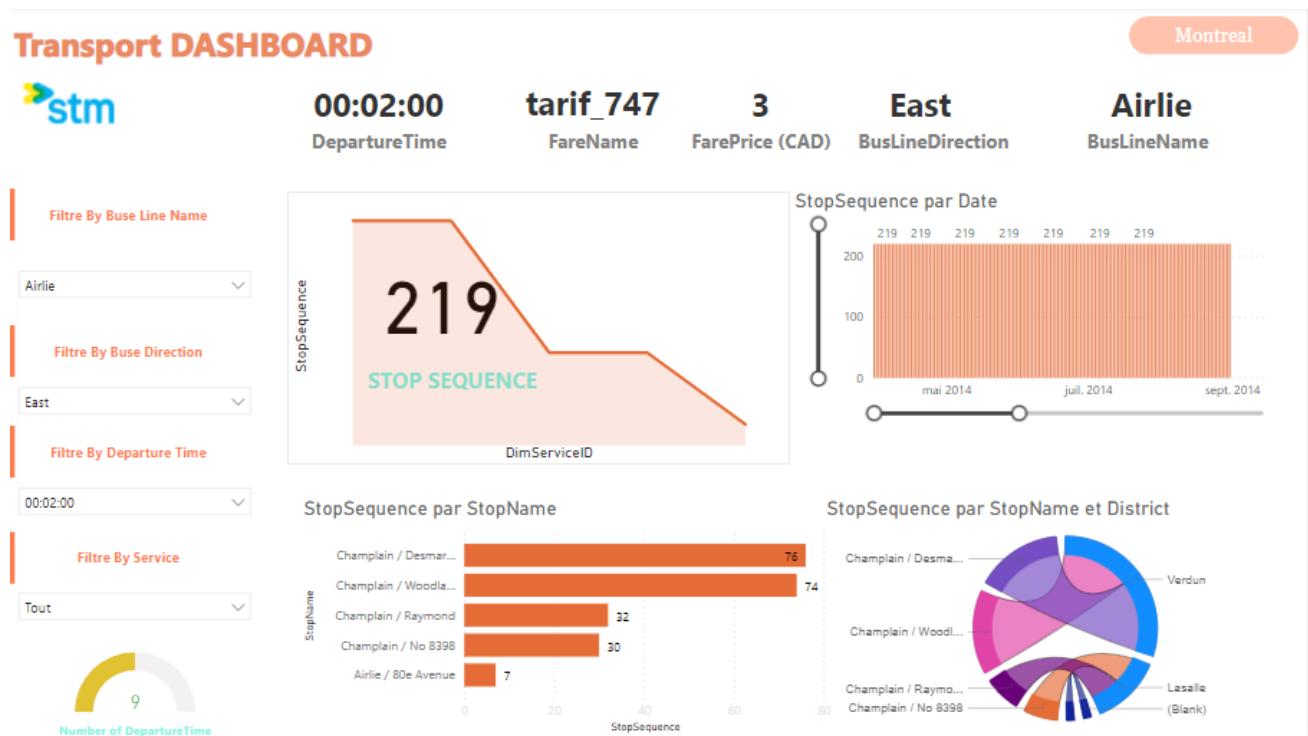
V. Création du Dashboard :

Cette partie consiste à manipuler les données stockées dans notre entrepôt de données, et les mettre en profit pour créer un tableau de bord interactif aidant à bien visualiser les détails qui se trouvent derrière les données pour qu'ils soient bien explorés, exploités d'une manière optimisée. Chose qui facilite la prise des décisions. Pour ce faire, on a adopté la solution de visualisation Power BI. Et voici ci-dessus le résultat :



Figure 5 : Tableau de board

Thème clair :



Conclusion

L'informatique décisionnelle n'est pas juste une nouvelle mode, on la retrouve dans tous les domaines et corps de métiers pourvu qu'il y ait des données, car elle fournit une réelle aide à la décision. En effet, nous avons vu qu'elle permet de déterminer ce qu'il s'est passé, ce qu'il se passe et, encore plus important, ce qu'il se passera.

C'est également une science fiable qui s'appuie sur des algorithmes de recherche (utilisés par le data mining), poussés et approuvés, couplés avec des outils permettant d'extraire et de découvrir de l'information cachée, rendue visible à l'utilisateur grâce aux outils d'analyse et de mesure.

L'informatique décisionnelle permet la transformation de données brutes extraites et transformées en informations par un ETL. L'information est stockée dans un entrepôt de données pour être analysée par des outils d'analyses qui transforment les informations en savoir.

Références

- ✓ <https://www.stm.info/>
- ✓ [Référence | Static Transit | Google Developers](#)