



**Ecole Supérieure Privée des Technologies de l'Information et de  
Management de Nabeul**

**Filière :**

**Cycle Ingénieur en Génie Informatique**

**Rapport de  
Mini Projet ELT avec Talend**

**Sujet :**

**Réalisation d'une Solution BI pour  
analyser les ventes et retour des produits sportif**

**Elaboré par : Ben Hadj Kacem Souha**

**Encadré par : Mr. Ahmed Ben Taleb**

**Année universitaire : 2022-20223**

## Table des matières

<b>1</b>	<b>L'informatique décisionnelle.....</b>	<b>6</b>
1.1	Introduction .....	6
1.2	Concepts généraux du BI .....	6
1.2.1	La Business Intelligence.....	6
1.2.2	Avantage du BI.....	6
1.2.3	Limites du BI.....	7
1.3	Le Principe de L'ELT.....	7
1.4	Entrepôt de données .....	8
1.5	Alimentation de l'entrepôt.....	8
1.5.1	L'extraction des données.....	8
1.5.2	Le chargement des données.....	8
1.5.3	La transformation des données.....	8
1.6	Conclusion.....	9
<b>2</b>	<b>Chapitre 2: Cadre du projet.....</b>	<b>10</b>
2.1	Introduction .....	10
2.2	Cadre général.....	10
2.3	Problématique.....	10
2.4	Solutions Proposé.....	10
2.5	Analyse et spécification des besoins .....	10
2.6	Les spécifications de l'entrepôt.....	10
2.7	Les spécifications de l'application .....	11
2.8	Conclusion.....	11
<b>3</b>	<b>CHAPITRE 3 : Conception .....</b>	<b>12</b>
3.1	Introduction .....	12
3.1.1	Composantes d'un schéma dimensionnel.....	12
3.1.2	Modélisation logique des données :.....	12
3.1.3	Architecture .....	12
3.2	Description de Modele : .....	14
3.3	Identification des indicateurs .....	16
3.4	Conclusion.....	17
	Dance ce chapitre nous avons présenter notre conception de datawarehouse, la description de modèle, de Dataset et des indicateurs.....	17
<b>4</b>	<b>Chapitre 4 : Réalisation .....</b>	<b>18</b>
4.1	Introduction .....	18
4.2	Environnement logiciel .....	18
4.2.1	Talend Open Studio.....	18
4.2.2	WampServer.....	18
4.2.3	Power BI.....	18

<b>4.3</b>	Intégration des données .....	19
4.3.1	Implémentation ELT : .....	19
4.3.2	Etape de la construction de DW.....	20
4.3.3	Composants Talend Utilisées .....	24
<b>4.4</b>	Reporting .....	25

## Table Des Figures :

Figure 1 principe de L'ELT .....	7
Figure 2 modelisation en flocon Fact returns .....	13
Figure 3 modelisation en flocon Fact Sales.....	13
Figure 4 Modélisation de Datawarehouse .....	14
Figure 5 dataset date.....	14
Figure 6 dataset customer.....	15
Figure 7 Dataset Territory .....	15
Figure 8 Dataset Category .....	15
Figure 9 Dataset SubCategory.....	16
Figure 10 DATaset Product.....	16
Figure 11 Job Talend.....	20
Figure 12 création métadonnées.....	20
Figure 13 extraction des données .....	21
Figure 14 Job 1 Extract Product.....	21
Figure 15 Loading des Jobs.....	22
Figure 16 Transform.....	23
Figure 17 alimentation fact sales.....	23
Figure 18 alimentation fact_returns.....	24
Figure 20 TFileInputDelimited.....	24
Figure 21 TMap.....	24
Figure 22 TDBOutput .....	24
Figure 23 TDBInput .....	25
Figure 25 TSortRow .....	25
Figure 26 TUniqRow.....	25
Figure 29 Rapport power BI N1 .....	26
Figure 30 Rapport power BI N2.....	26
Figure 31 Rapport power BI N3.....	<b>Erreur ! Signet non défini.</b>

## Introduction Générale

La Business Intelligence (BI) est un processus technologique d'analyse des données et de présentation d'informations pour aider les dirigeants, managers et autres utilisateurs finaux de l'entreprise à prendre des décisions business éclairées.

La Business Intelligence englobe une grande variété d'outils, d'applications et de méthodologies qui permettent aux organisations de collecter des données à partir de systèmes internes et de sources externes. Ces données sont ensuite préparées pour l'analyse afin de créer des rapports, tableaux de bord et autres outils de Data visualisation pour rendre les résultats analytiques disponibles aux décideurs et aux opérations.

Aujourd'hui, les entreprises s'appuient sur les logiciels de Business Intelligence pour identifier et extraire des informations précieuses des grands volumes de données qu'elles stockent et ce le cas même pour les entreprises d'e-commerce. Ces outils permettent d'en tirer des informations tels que des veilles concurrentielles et les tendances du marché, ainsi que des informations internes tel que trouver les raisons des opportunités perdues.

Au cours de ce projet, nous allons exploiter des données de vente d'un site e-commerce, afin de faire des analyses sur le sujet de vente, la Business Intelligence avec ces outils d'intégration et de reporting peut nous faciliter la réalisation de cet objectif qui est de transformer les données brutes en véritables sources d'informations commerciales.

# 1 L'informatique décisionnelle

## 1.1 Introduction

Ce chapitre sera réservé pour définir l'informatique décisionnelle. Nous présentons dans un premier temps ses avantages et ses limites. Nous abordons, ensuite, ses termes et les concepts clés en détaillant la notion d'ELT et d'entrepôt de données.

## 1.2 Concepts généraux du BI

Pour réussir dans les projets BI, il faut définir l'informatique décisionnelle tout en mettant l'accent sur leurs objectifs et leurs caractéristiques. Ensuite, nous présenterons l'architecture et les différents modèles du système décisionnel.

### 1.2.1 La Business Intelligence

L'informatique décisionnelle, également Business Intelligence ou BI en anglais, désigne les moyens, les méthodes et les outils qui apportent des solutions en vue d'offrir une aide à la décision aux professionnels afin de leur permettre d'avoir une vue d'ensemble sur l'activité de l'entreprise et de leur permettre de prendre des décisions plus avisées à travers des tableaux de bord de suivi et des analyses.

Au cours de leur mission, les chefs d'entreprise sont exposés à de grandes quantités d'informations provenant de sources variées et leurs données sont diverses.

Par conséquent, nous devons automatiser leur collecte et leur distribution.

C'est le rôle des logiciels de business intelligence. Les outils de BI collectent des informations à partir de plusieurs sources, nettoient et organisent les données et les rendent disponibles dans le centre de décision.

En effet l'informatique décisionnelle se caractérise principalement par :

- Plus direct : Un aperçu en temps réel de l'activité.
- Plus cohérent : Un ouvrage de référence sur la gouvernance.
- Plus large : Vue agrandie de votre activité sur votre marché.
- Plus vite : Les outils de décision accélèrent les performances.
- Plus loin : Prédisez les événements et planifiez l'avenir.

### 1.2.2 Avantage du BI

Déployer une solution BI apporte de nombreux avantages :

- Améliorer la visibilité sur les chiffres, les écarts et les anomalies.
- La combinaison de plusieurs sources de données (ERP, systèmes comptable, feuilles de calcul, des

budgets ...).

- La présentation uniforme d'informations fiables.
- L'automatisation permettant l'accélération de la collecte et de la diffusion de l'information.
- La performance dans le calcul d'agrégats sur de gros volume de données.
- La prise de décision grâce à des indicateurs pertinents et à une structure cohérente des informations.
- L'aide à nettoyer les données présentes dans différents logiciels.
- L'anticipation des événements et la projection dans l'avenir.

### 1.2.3 Limites du BI

Parmi les limites de la Business Intelligence :

- La mise en place d'une solution de BI prend beaucoup du temps : de nombreuses entreprises dans le scénario industriel rapide ne sont pas assez patientes pour attendre la mise en place du système décisionnel dans leur organisation.
- Complexité : un autre inconvénient de BI pourrait être sa complexité dans la mise en œuvre des données.
- Erreur : les résultats produits par les systèmes décisionnels sont le résultat de conceptions informatiques et mathématiques complexes, qui peuvent révéler des erreurs, par ailleurs les résultats sont souvent statistiques, donc non déterministes. La possibilité d'une erreur ou d'une approximation inadaptée devra toujours être prise en compte dans les décisions.

## 1.3 Le Principe de L'ELT

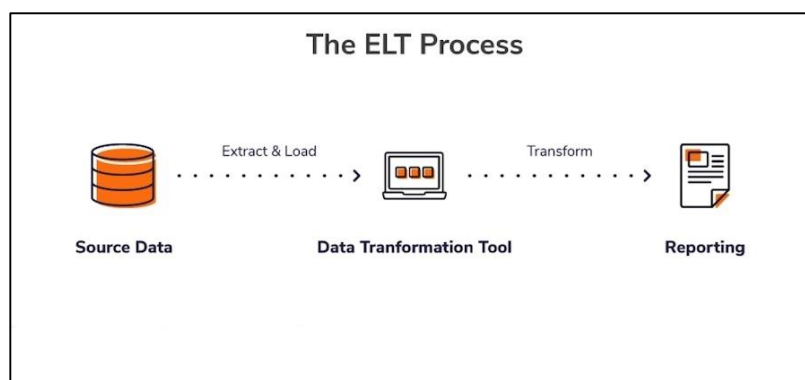


Figure 1 principe de L'ELT

ELT signifie Extract, Load, Transform. L'ELT est une variante de l'Extract, Transform, Load (ETL), un processus d'intégration de données dans lequel la transformation a lieu sur un serveur intermédiaire avant d'être chargée dans la cible. En revanche, les ELT permettent de charger les données brutes directement dans la cible et de les transformer par la suite. Cette capacité est particulièrement

intéressante pour le traitement des grands ensembles de données nécessaires à la Business Intelligence (BI) et à l'analyse de données volumineuses.

## 1.4 Entrepôt de données

Un entrepôt de données est une collection de données thématiques, intégrées, non volatiles et historiées, organisées pour la prise de décision.". D'après cette définition nous distinguons les caractéristiques suivantes :

- Données Orientées sujet : les données des entrepôts sont organisées par sujet et donc triées par thème.
- Données intégrées : les données provenant des différentes sources doivent être intégrées avant leur stockage dans l'entrepôt de données. Un nettoyage préalable des données est nécessaire afin d'avoir une cohérence et une normalisation de l'information.
- Données non-volatiles : à la différence des données opérationnelles, celles de l'entrepôt sont permanentes et ne peuvent pas être modifiées. Le rafraîchissement de l'entrepôt, consiste à ajouter de nouvelles données sans perdre celles qui existent.
- Historiées : les données doivent être datées.

## 1.5 Alimentation de l'entrepôt

Une fois l'entrepôt est conçu, il faut l'alimenter et le charger en données. Cette alimentation s'effectue à travers le processus ELT et se déroule en trois phases :

### 1.5.1 L'extraction des données

Il s'agit de la première étape de récupération des informations dans l'environnement de l'entrepôt de données. L'extraction comprend la lecture et la compréhension de la source de données.

### 1.5.2 Le chargement des données

C'est la phase de l'alimentation d'un entrepôt de données temporaire pour effectuer par la suite les transformations nécessaires

### 1.5.3 La transformation des données

Une fois que les données sont extraites dans l'entrepôt de données temporaire appliquons plusieurs étapes de transformations qui ont pour but de rendre les données cibles homogènes afin qu'elles puissent être traitées de façon cohérente



## 1.6 Conclusion

Dans ce chapitre, nous avons détaillé toutes les notions relatives aux systèmes décisionnels, pour les maîtriser afin de favoriser le bon déroulement du projet.

## 2 Chapitre 2: Cadre du projet

### 2.1 Introduction

Dans ce chapitre, nous, situons le présent travail dans son contexte général, spécifions les besoins et enfin nous proposons la solution aux problèmes soulevés.

### 2.2 Cadre général

Dans le cadre de mini projet Talend. L'objectif premier était de mettre en place une solution Business Intelligence pour les données de ventes des produits sportif. Notre solution s'agit de la Modélisation, intégration et reporting.

### 2.3 Problématique

Les produits sportif sont couteuses de côté qualité et importantes de coté santé, d'où il faut analyse les vente car il s'agit d'un taux de retour si la qualité ne convient pas.

La demande et la diversification des clients et des choix il faut bien exploiter les données disponibles pour établir la bonne stratégie commerciale, marketing et la prise des décisions fiable et efficace

### 2.4 Solutions Proposé

Pour faire face à ces défis et problèmes précédemment cités nous avons proposé une solution BI qui permet de valoriser les données et l'exploiter, dégager les KPIs et mettre en place un rapport de dashboarding bien clair pour l'aide à la prise des décisions.

### 2.5 Analyse et spécification des besoins

Dans notre processus de définition des besoins, nous avons rassemblé les besoins fonctionnels, qui se classifient en deux niveaux majeurs ; la spécification des caractéristiques de l'entrepôt et la spécification des fonctionnalités de l'application

### 2.6 Les spécifications de l'entrepôt

L'entrepôt doit être conçu de manière qu'il permette de :

- Construire une source de données unique et non volatile.
- Mieux contrôler les données : nous pouvons organiser les données selon nos besoins et donc rendre la base de données plus optimale.
- Rendre la base plus évolutive et plus adaptée à nos futurs besoins.
- Garder la traçabilité de chaque donnée : source d'extraction de cette information.
- Garantir une évolution et un rafraîchissement périodique et régulier des données.

## 2.7 Les spécifications de l'application

L'application doit être conçue de manière à permettre :

- La haute disponibilité des données à n'importe quel instant.
- La génération facile des rapports
- La consultation aisée des tableaux de bord fournis.
- L'application facile des différentes opérations sur les données.

## 2.8 Conclusion

Tout au long de ce chapitre, une présentation du contexte de projet à réaliser est mise en place ce qui a mené à déterminer les différentes problématiques et proposer la solution envisagée pour améliorer la situation actuelle.

## 3 CHAPITRE 3 : Conception

### 3.1 Introduction

Dans cette partie, nous allons montrer les composantes d'un graphe multidimensionnel, nous allons ensuite, nous présenterons le modèle logique des données.

#### 3.1.1 Composantes d'un schéma dimensionnel

Ralph Kimball introduit la modélisation multidimensionnelle. Il se compose de deux nouveaux concepts tels que fait et dimension.

- Une table de fait est une table qui contient les données observables (les faits) que l'on possède sur un sujet et que l'on veut étudier, selon divers axes d'analyse (les dimensions).
- Une table de dimensions est une table qui contient les axes d'analyse (les dimensions) selon lesquels on veut étudier des données observables (les faits) qui, soumises à une analyse multi- dimensionnelle, donnent aux utilisateurs des renseignements nécessaires à la prise de décision.

#### 3.1.2 Modélisation logique des données :

Il existe essentiellement trois modèles possibles pour organiser les données stockées dans l'entrepôt de données : le modèle en étoile, le modèle flocon de neige et le modèle en constellation.

**Modèle en étoile** : est un modèle qui centre une table des faits et la relie à chaque table de dimensions ou axe d'analyse. En tenant compte que les dimensions ne sont pas liées entre eux.

#### 3.1.3 Architecture

Notre solution couvre toute la chaîne décisionnelle, de l'extraction des données à la livraison des rapports.

L'architecture fonctionnelle de notre système peut être exprimée comme suit :

- Recueillir des données sources à partir d'un fichier Excel.
- Intégrer les données agrégées dans un entrepôt de données pour obtenir des informations pour la prise de décision.
- L'entrepôt de données chargé se compose d'un seul magasin de données.
- Assurez-vous que diverses données sont renvoyées et décidez comment procéder. Les tableaux de bord sont affichés en fonction des besoins de l'utilisateur final.

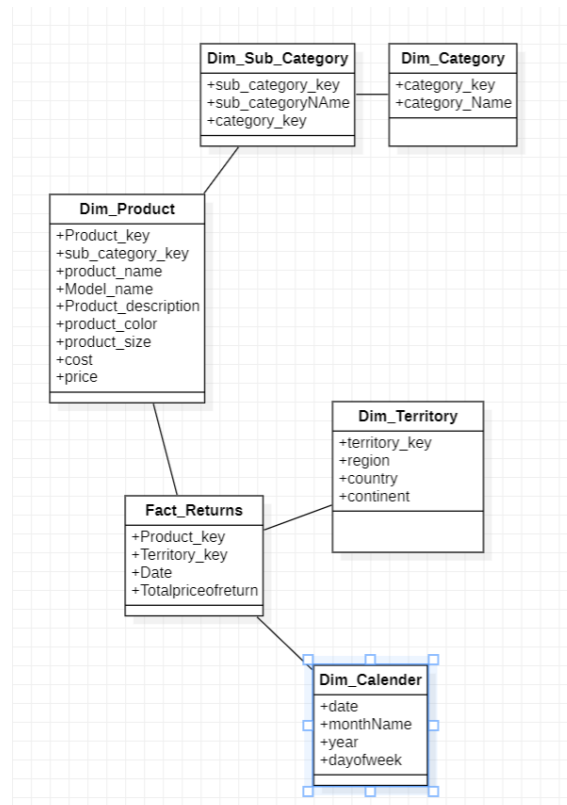


Figure 2 modelisation en flocon Fact returns

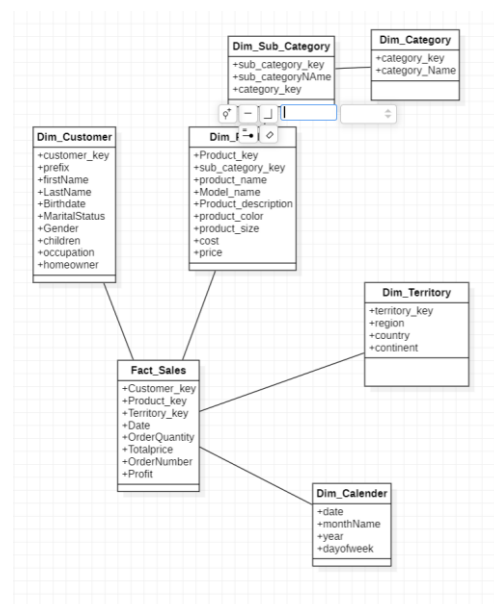


Figure 3 modelisation en flocon Fact Sales

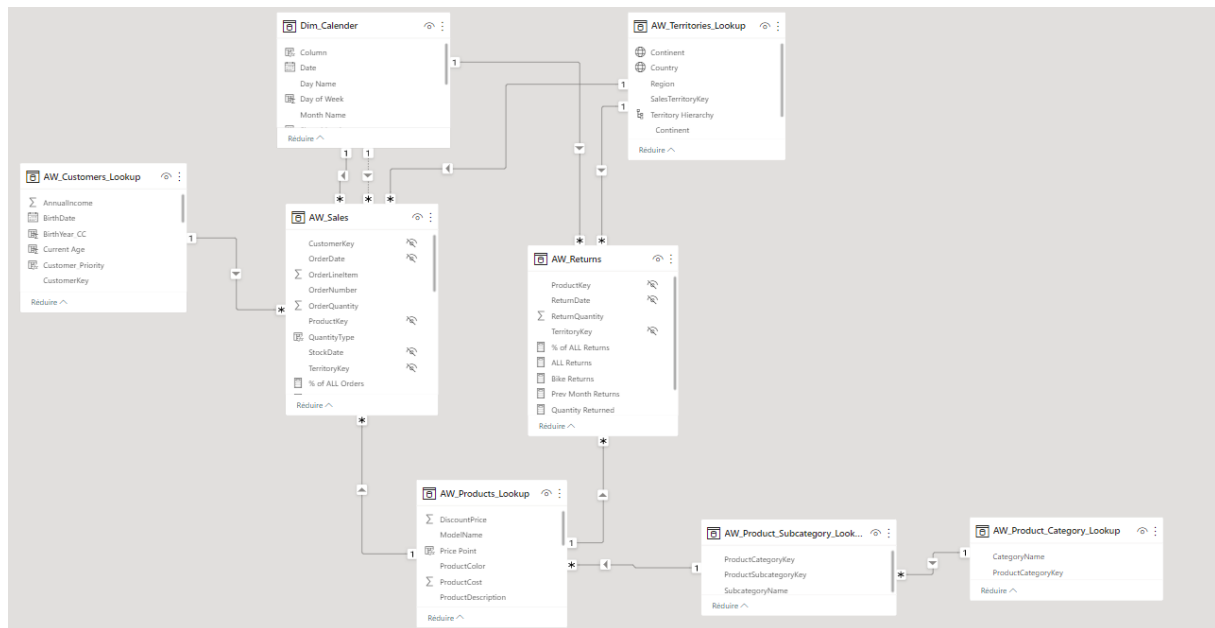


Figure 4 Modélisation de Datawarehouse

## 3.2 Description de Model :

### 3.2.1.1 Dataset

#### 3.2.1.1.1 Calender

	A	B	C
1	Date		
2	1/1/2015		
3	1/2/2015		
4	1/3/2015		
5	1/4/2015		
6	1/5/2015		
7	1/6/2015		
8	1/7/2015		
9	1/8/2015		
10	1/9/2015		
11	1/10/2015		
12	1/11/2015		
13	1/12/2015		
14	1/13/2015		
15	1/14/2015		

Figure 5 dataset date

### 3.2.1.2 Customer

#### 3.2.1.2.1 Columns

CustomerKey	Prefix	FirstName	LastName	BirthDate	MaritalStatus	Gender	EmailAddress	AnnualIncome	TotalChildren	EducationLev	Occupation	HomeOwner
11000	MR.	JON	YANG	4/8/1966	M	M	jon24@adver	\$90,000	2	Bachelors	Professional	Y
11001	MR.	EUGENE	HUANG	5/14/1965	S	M	eugene10@a	\$60,000	3	Bachelors	Professional	N
11002	MR.	RUBEN	TORRES	8/12/1965	M	M	ruben35@adv	\$60,000	3	Bachelors	Professional	Y
11003	MS.	CHRISTY	ZHU	2/15/1968	S	F	christy12@ac	\$70,000	0	Bachelors	Professional	N
11004	MRS.	ELIZABETH	JOHNSON	8/8/1968	S	F	elizabeth5@e	\$80,000	5	Bachelors	Professional	Y
11005	MR.	JULIO	RUIZ	8/5/1965	S	M	julio1@adver	\$70,000	0	Bachelors	Professional	Y
11007	MR.	MARCO	MEHTA	5/9/1964	M	M	marco14@ad	\$60,000	3	Bachelors	Professional	Y
11008	MRS.	ROBIN	VERHOFF	7/7/1964	S	F	rob4@advent	\$60,000	4	Bachelors	Professional	Y
11009	MR.	SHANNON	CARLSON	4/1/1964	S	M	shannon38@	\$70,000	0	Bachelors	Professional	N
11010	MS.	JACQUELYN	SUAREZ	2/6/1964	S	F	jacquelyn20@	\$70,000	0	Bachelors	Professional	N
11011	MR.	CURTIS	LU	11/4/1963	M	M	curtis9@adve	\$60,000	4	Bachelors	Professional	Y
11012	MRS.	LAUREN	WALKER	1/18/1968	M	F	lauren41@ad	\$100,000	2	Bachelors	Management	Y
11013	MR.	IAN	JENKINS	8/6/1968	M	M	ian47@adver	\$100,000	2	Bachelors	Management	Y
11014	MRS.	SYDNEY	BENNETT	5/9/1968	S	F	sydney23@ac	\$100,000	3	Bachelors	Management	N
11015	MS.	CHLOE	YOUNG	2/27/1979	S	F	chloe23@adv	\$30,000	0	Partial Colleg	Skilled Manu	N
11016	MR.	WYATT	HILL	4/28/1979	M	M	wyatt32@adv	\$30,000	0	Partial Colleg	Skilled Manu	Y
11017	MRS.	SHANNON	WANG	6/26/1944	S	F	shannon1@a	\$20,000	4	High School	Skilled Manu	Y
11018	MR.	CLARENCE	RAI	10/9/1944	S	M	clarence32@	\$30,000	2	Partial Colleg	Clerical	Y
11019	MR.	LUKE	LAL	3/7/1978	S	M	luke18@adve	\$40,000	0	High School	Skilled Manu	N

Figure 6 dataset customer

### 3.2.1.3 Territory

#### 3.2.1.3.1 Columns

	A	B	C	D
1	SalesTerritoryKey;Region;Country;Continent			
2	1;Northwest;United States;North America			
3	2;Northeast;United States;North America			
4	3;Central;United States;North America			
5	4;Southwest;United States;North America			
6	5;Southeast;United States;North America			
7	6;Canada;Canada;North America			
8	7;France;France;Europe			
9	8;Germany;Germany;Europe			
10	9;Australia;Australia;Pacific			
11	10;United Kingdom;United Kingdom;Europe			
12				

Figure 7 Dataset Territory

### 3.2.1.4 Product Category

#### 3.2.1.4.1 Columns

	A	B	C
1	ProductCategoryKey;CategoryName		
2	1 Bikes		
3	2 Components		
4	3 Clothing		
5	4 Accessories		
6			

Figure 8 Dataset Category

### 3.2.1.5 Product Sub Category

#### 3.2.1.5.1 Columns

	A	B	C	D
1	ProductSubcategory	SubcategoryName	ProductCategoryKey	
2		1 Mountain Bikes	1	
3		2 Road Bikes	1	
4		3 Touring Bikes	1	
5		4 Handlebars	2	
6		5 Bottom Brackets	2	
7		6 Brakes	2	
8		7 Chains	2	
9		8 Cranksets	2	
10		9 Derailleurs	2	
11		10 Forks	2	
12		11 Headsets	2	
13		12 Mountain Frames	2	
14		13 Pedals	2	
15		14 Road Frames	2	
16		15 Saddles	2	
17		16 Touring Frames	2	
18		17 Wheels	2	

Figure 9 Dataset SubCategory

### 3.2.1.6 Product

#### 3.2.1.6.1 Columns

	A	B	C	D	E	F	G	H	I	J	K	L
1	ProductKey	ProductSubcategory	ProductSKU	ProductName	ModelName	ProductDescription	ProductColor	ProductSize	ProductStyle	ProductCost	ProductPrice	
2	214	31	HL-U509-R	Sport-100 Hel Sport-100	Universal fit, \	Red		0	0	13.0863	34.99	
3	215	31	HL-U509	Sport-100 Hel Sport-100	Universal fit, \	Black		0	0	12.0278	33.6442	
4	218	23	SO-B909-M	Mountain Bike Mountain Bike	Combination \	White	M	U		3.3963	9.5	
5	219	23	SO-B909-L	Mountain Bike Mountain Bike	Combination \	White	L	U		3.3963	9.5	
6	220	31	HL-U509-B	Sport-100 Hel Sport-100	Universal fit, \	Blue		0	0	12.0278	33.6442	
7	223	19	CA-1098	AWC Logo Cap Cycling Cap	Traditional sty	Multi		0	U	5.7052	8.6442	
8	226	21	LJ-0192-S	Long-Sleeve L Long-Sleeve L	Unisex long-sl	Multi	S	U		31.7244	48.0673	
9	229	21	LJ-0192-M	Long-Sleeve L Long-Sleeve L	Unisex long-sl	Multi	M	U		31.7244	48.0673	
10	232	21	LJ-0192-L	Long-Sleeve L Long-Sleeve L	Unisex long-sl	Multi	L	U		31.7244	48.0673	
11	235	21	LJ-0192-X	Long-Sleeve L Long-Sleeve L	Unisex long-sl	Multi	XL	U		31.7244	48.0673	
12	238	14	FR-R92R-62	HL Road Fram HL Road Fram	Our lightest ai	Red		62	U	747.9682	1263.4598	
13	241	14	FR-R92R-44	HL Road Fram HL Road Fram	Our lightest ai	Red		44	U	747.9682	1263.4598	
14	244	14	FR-R92R-48	HL Road Fram HL Road Fram	Our lightest ai	Red		48	U	747.9682	1263.4598	
15	247	14	FR-R92R-52	HL Road Fram HL Road Fram	Our lightest ai	Red		52	U	747.9682	1263.4598	
16	250	14	FR-R92R-56	HL Road Fram HL Road Fram	Our lightest ai	Red		56	U	747.9682	1263.4598	
17	253	14	FR-R38B-58	LL Road Fram LL Road Fram	The LL Frame	Black		58	U	176.1997	297.6346	
18	256	14	FR-R38B-60	LL Road Fram LL Road Fram	The LL Frame	Black		60	U	176.1997	297.6346	

Figure 10 Dataset Product

## 3.3 Identification des indicateurs

Les indicateurs de performance clés, ou KPI, sont un ensemble de mesures quantifiables utilisées pour évaluer la performance d'une entreprise, ont pour objectif d'aider les investisseurs à déterminer les réalisations stratégiques, opérationnelles et financières d'une entreprise, notamment par rapport à d'autres entreprises du même secteur.



Total des ordre par

- category
- pays
- occupation
- genre
- date ( mois – ans )

Revenue = quantité vendue \* prix

Nombre of returns

### 3.4 Conclusion

Dance ce chapitre nous avons présenter notre conception de datawarehouse, la description de modèle, de Dataset et des indicateurs

## 4 Chapitre 4 : Réalisation

### 4.1 Introduction

Dans ce chapitre, nous présentons les étapes suivies pour réaliser notre projet et nous avons exploité les outils présentés ci-dessous.

### 4.2 Environnement logiciel

#### 4.2.1 Talend Open Studio

Talend Open Studio for Data Integration est un outil permettant la gestion des données hétérogènes ou homogènes au sein d'un système d'information.

Talend Open Studio permet de créer des processus complets quelques soient les problématiques autour des données. Les fonctionnalités de Talend Open Studio permettent la mise en place rapide de connexions à un grand nombre de technologies, l'implémentation simplifiée de transformations automatiques, le contrôle de la qualité des données, la communication avec l'ensemble des outils du système d'information par utilisation de standards (API, protocoles standards...). De ce fait, TOS peut intervenir de manière flexible quelle que soit l'architecture du SI choisi (SOA – Architecture Orientée Service, Architecture classique par flux d'intégration point à point...).

#### 4.2.2 WampServer

WampServer est une plate-forme de développement Web sous Windows pour des applications Web dynamiques à l'aide du serveur Apache2, du langage de scripts PHP et d'une base de données MySQL. Il possède également PHPMyAdmin pour gérer plus facilement vos bases de données.

#### 4.2.3 Power BI

Power BI est une solution de Business Intelligence développée par Microsoft pour permettre aux entreprises d'agréger, d'analyser et de visualiser les données en provenance de sources multiples. Découvrez les fonctionnalités, les avantages et les informations techniques (prix et disponibilité) sur cette solution.

L'analyse de données peut s'avérer très utile pour les entreprises. L'on transforme les données en informations exploitables. Cela aide à prendre de meilleures décisions stratégiques

### 4.3 Intégration des données

#### 4.3.1 Implémentation ELT :

Pour réaliser l'intégration des données nous avons opter pour L'ELT Cet axe décrit la phase d'implantation du système d'aide à la décision développé pour cette phase du projet. Les outils Talend automatisent entièrement la création, la maintenance et l'expansion des entrepôts de données. Chaque outil se comporte différemment et a ses forces et ses faiblesses.

##### 4.3.1.1 *L'extraction des données*

Il s'agit de la première étape de récupération des informations dans l'environnement de l'entrepôt de données. L'extraction comprend la lecture et la compréhension de la source de données.

##### 4.3.1.2 *Le chargement des données*

C'est la phase de l'alimentation d'un entrepôt de données temporaire pour effectuer par la suite les transformations nécessaires

Chargement et mis à jour de toutes les tables de dimensions et la table de fait :

- Extraction de données d'une base de données ou autres échantillon (Excel, Access. ).
- Conversion et nettoyage des données.
- Recherche des non-correspondances (abattre les doublant).
- Génération de métadonnées (data mart).
- Ajout d'un nouvel enregistrement.
- Traitement et contrôle des erreurs.
- Alimentation de la table de fait

##### 4.3.1.3 *La transformation des données*

Une fois que les données sont extraites dans l'entrepôt de données temporaire appliquons plusieurs étapes de transformations qui ont pour but de rendre les données cibles homogènes afin qu'elles puissent être traitées de façon cohérente

- Conversion et nettoyage des données.

- Recherche des non-correspondances.
- Échange de clés.
- Génération de métadonnées.
- Ajout d'un nouvel enregistrement.
- Gestion des erreurs et audit

#### 4.3.2 Etape de la construction de DW

Pour construire notre DW en utilisant Talend nous avons créé 3 principaux jobs :

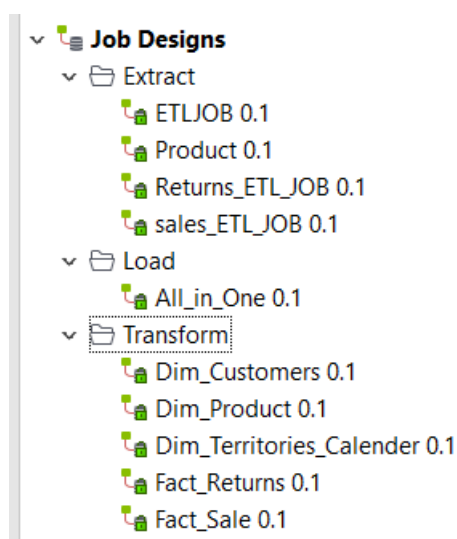


Figure 11 Job Talend

##### 4.3.2.1 JOB 1 :

Extract/Load : pour extraire les données de source Csv vers une datawarehouse temporaire  
A cette étape on a aussi configuré notre connexion de base de données, ainsi que notre métadonnée et nos fichier délimité

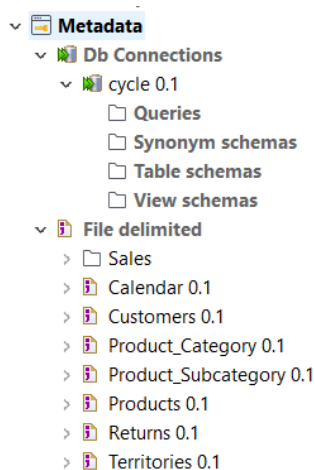


Figure 12 création métadonnées

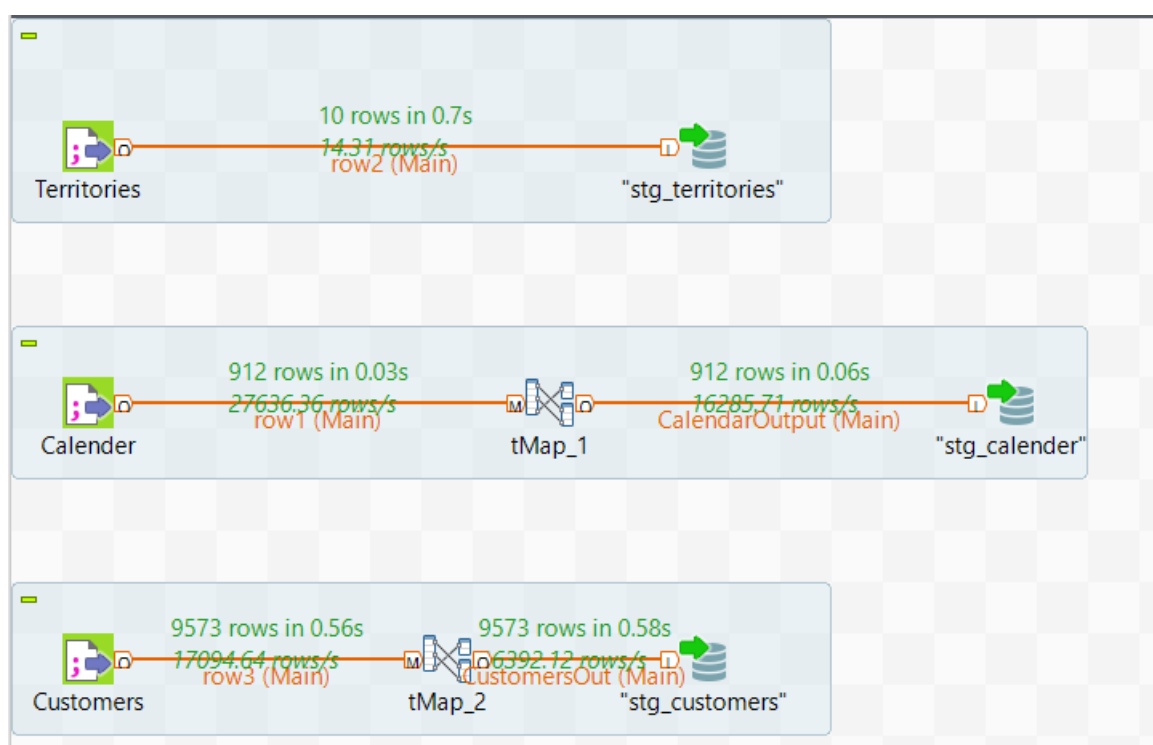


Figure 13 extraction des données

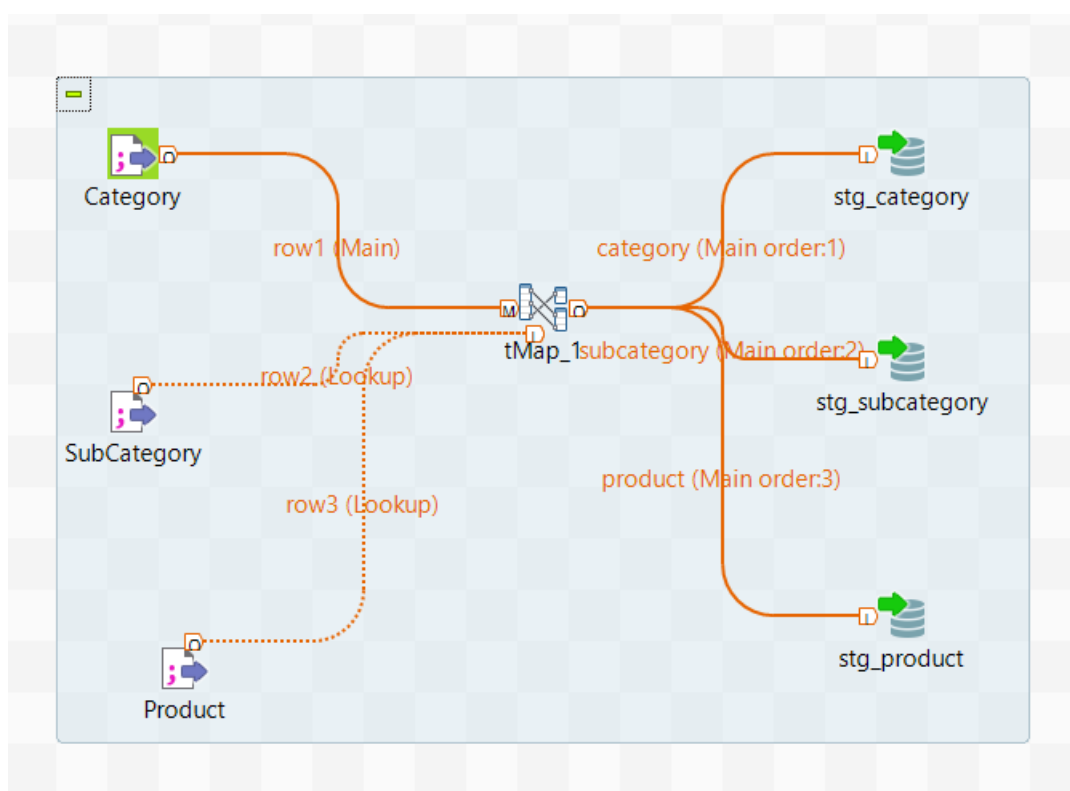


Figure 14 Job 1 Extract Product

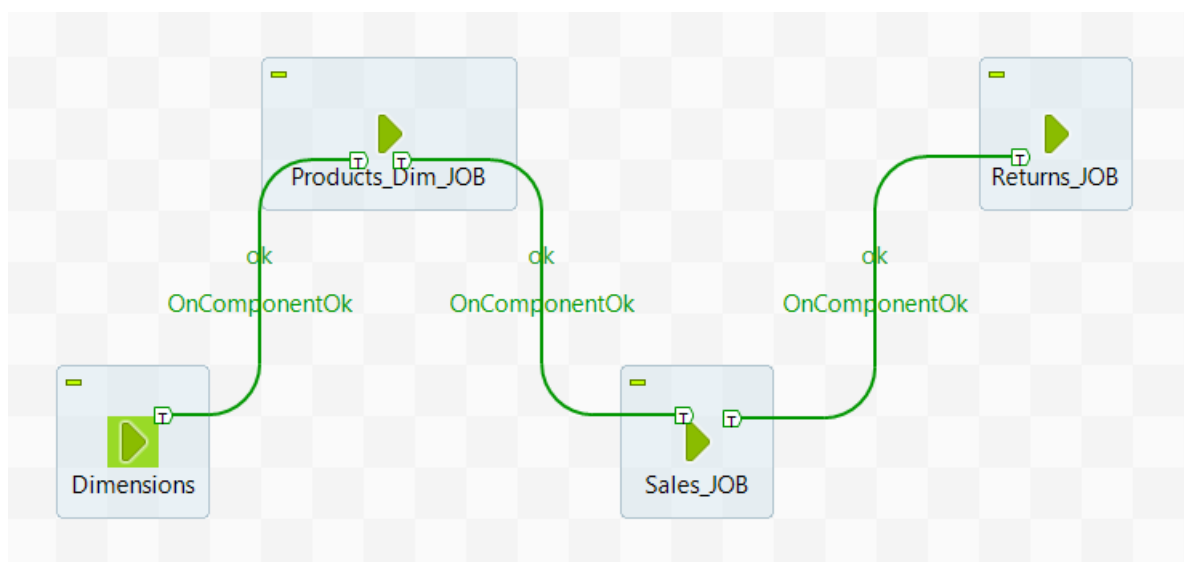
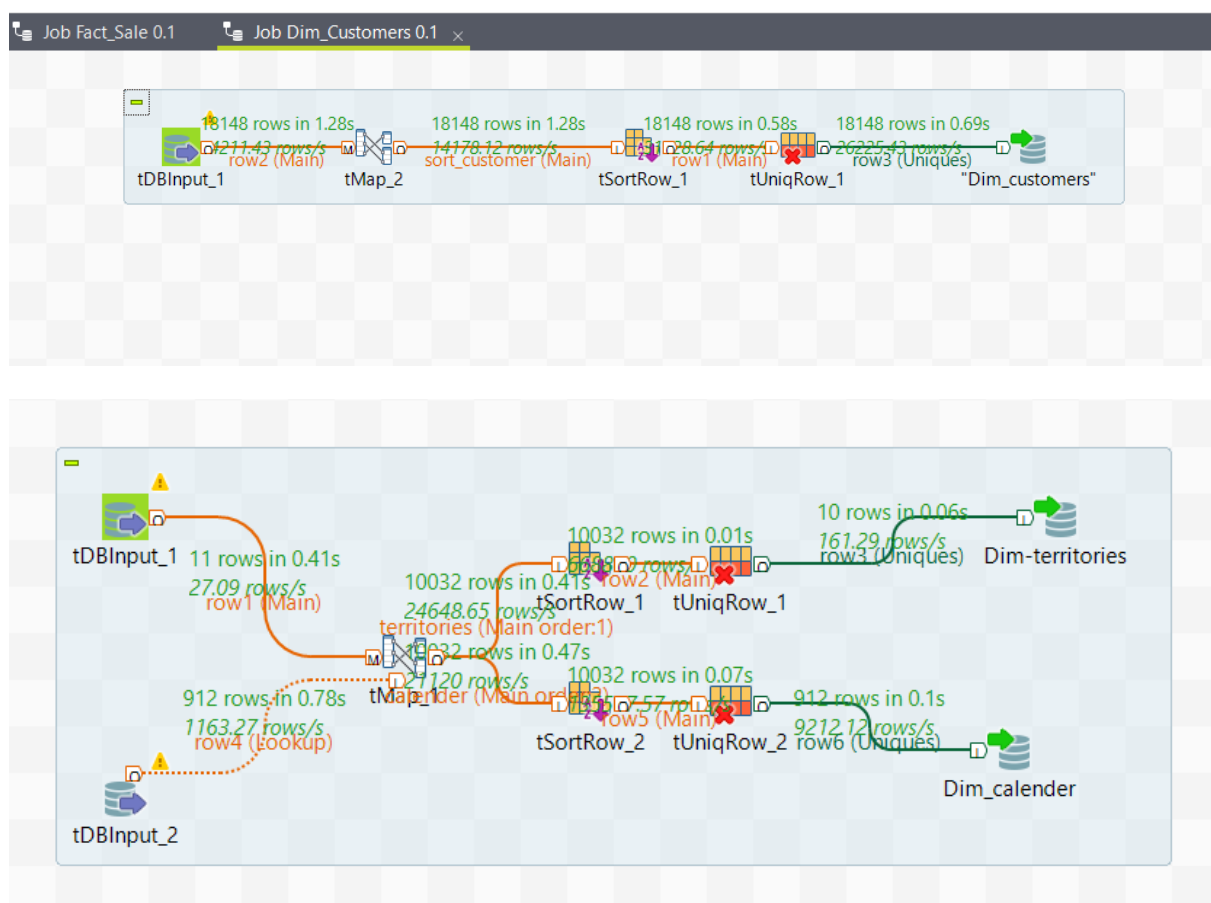


Figure 15 Loading des Jobs

#### 4.3.2.2 Transform

Transform 0,1 : pour effectuer les transformations nécessaires sur les données



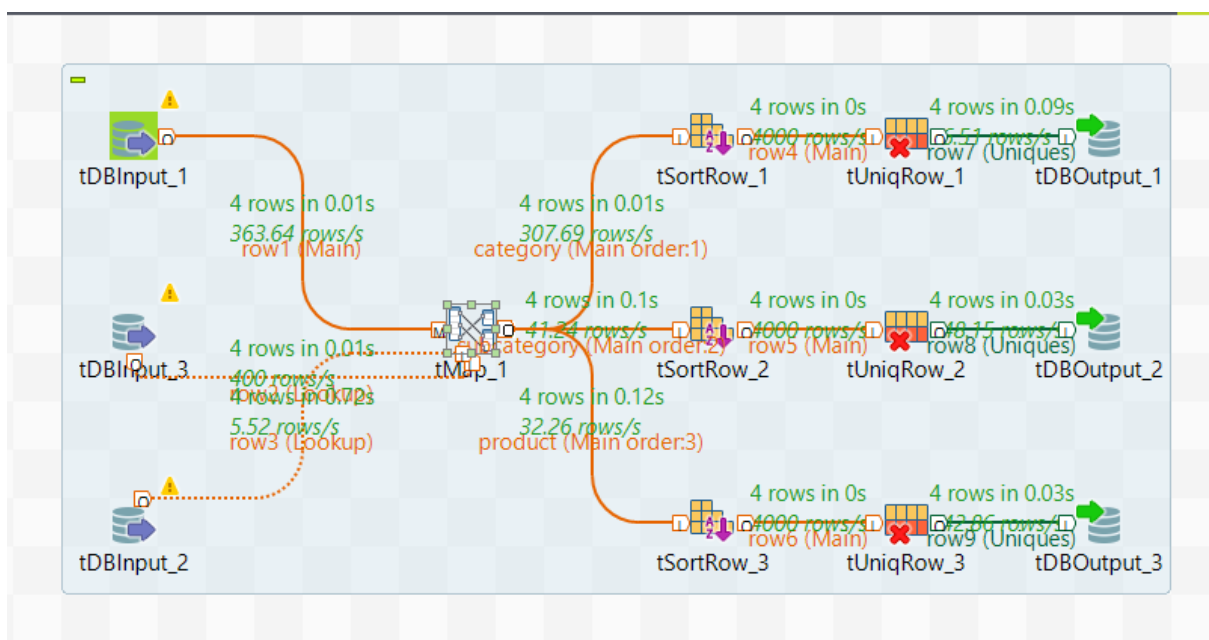


Figure 16 Transform

#### 4.3.2.3 Les Table Fact

Transform 0,2 : pour alimenter la nouvelle table fact pour avoir le nouveau datawarehouse traité, nettoyer et prête pour le reporting

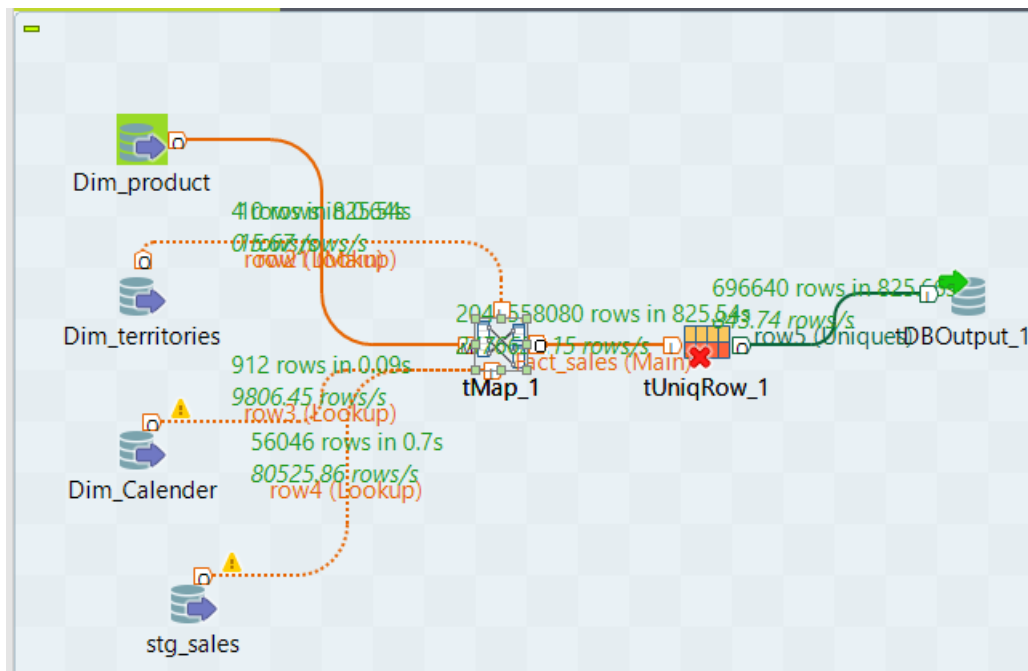


Figure 17 alimentation fact sales

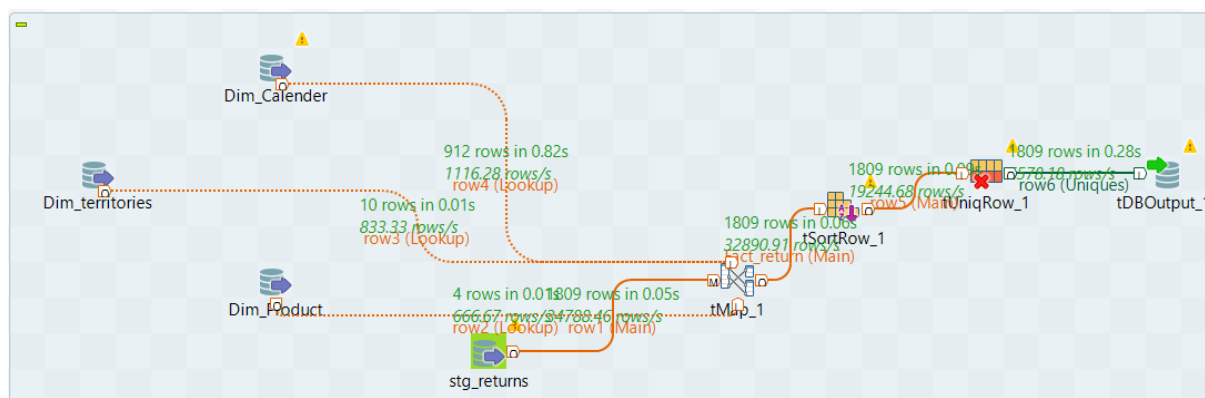


Figure 18 alimentation fact\_returns

### 4.3.3 Composants Talend Utilisées

#### 4.3.3.1 TFileInputDelimited :

Il s'agit d'un composant qui contient le fichier d'entrée du processus et qui est associé aux métadonnées déjà définies (Csv).

Figure 19  
TFileInputDelimited

#### 4.3.3.2 TMap :

Ce composant traduit et achemine les données d'une ou plusieurs sources de données vers une ou plusieurs destinations il permet d'appliquer plusieurs opérations de transformations sur les données



Figure 20 TMap

#### 4.3.3.3 TDBOutput :

Ce composant effectue des opérations définies sur la table et/ou les données de la table en fonction du flux entrant du composant précédent.



Figure 21 TDBOutput

#### 4.3.3.4 TDBInput :

Ce composant lit la base de données et utilise des requêtes pour extraire les champs de la base de données. Les champs récupérés sont transmis au composant suivant via la connexion de diffusion.





Figure 22 TDBInput

#### 4.3.3.5 TSortRow :

Le tSortRow trie les données d'entrée en fonction d'une ou plusieurs colonnes, par type de tri et ordre.



Figure 23 TSortRow

#### 4.3.3.6 TUniqRow

Le tUniqRow compare les entrées et trie les entrées en double du flux d'entrée.

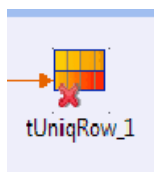


Figure 24 TUniqRow

## 4.4 Reporting

montre le tableau de bord de suivi opérationnel de l'entrepôt. Il aide à connaître la prévision des ventes par année pour l'aide à la prise de décision afin d'avoir une amélioration sur les ventes.

Nous avons utilisé plusieurs graphiques pour la modélisation de notre tableau de bord.

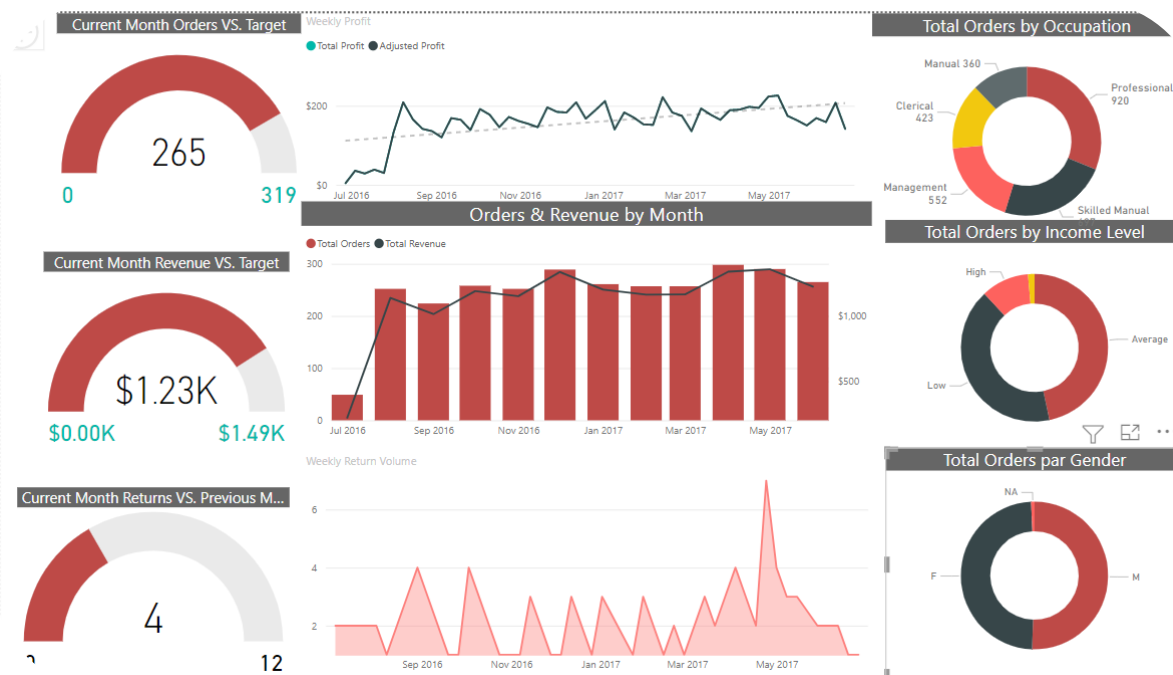


Figure 25 Rapport power BI N1

Dans ce rapport on visualise les mesures suivantes :

- Total des ordres
- Le total de chiffre d'affaire
- Le top des produits par ordre
- Le top des produits par revenue
- Les ordres par pays

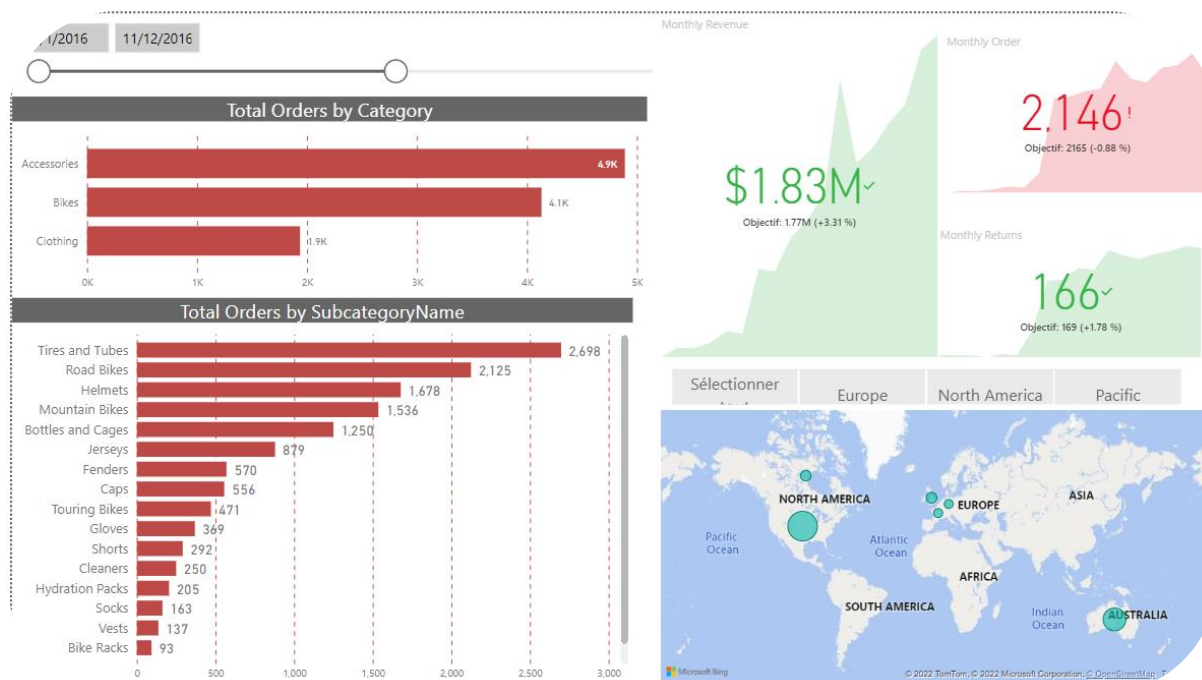


Figure 26 Rapport power BI N2

Dans cette figure on effectue une étude comparative des ordres et de revenue ainsi les retour par mois et par objectif

Dans cette figure on a effectué une analyse sur les ordre par les client

- Total orders pour chaque client
- Total orders par gender , income , occupation , age
- Top customer
- L'évolution des ordre et revenue sur une durée

## Conclusion Générale

L'implémentation de cette solution BI en utilisant un environnement logiciel différent ainsi le concept de l'ELT nous a permis de se développer dans le cadre conceptuel et aussi dans le cadre technique d'intégration et visualisation des données

Cet environnement nous a permis de profiter des atouts de chacun de nous et de mettre notre but commun en premier plan. Enfin, ce projet fut une occasion de grande valeur pour apprendre la gestion du temps de travail et des priorités

En termes de perspective, nous pouvons mentionner que notre application est encore dans la phase d'amélioration, nous sommes en train de travailler sur d'autres détails