

Projet tutoré :

BUSINESS INTELLIGENCE

-MASTER : Big Data & Cloud
Computing

Présenté à :

TKATEK Said

Réalisé par :

BOURACHED Douha

BENLAMHAIRA Oumaima

Sommaire :

Abstrait.....	3
Introduction :	4
Chapitre 1 : Notions et Techniques	4
1 Architecture décisionnelle :	4
1.1 Business Intelligence :	4
1.1.1 Avantage du BI :	5
1.1.2 Limites du BI :	5
1.2 Les principes des systèmes décisionnels :	5
1.3 Les étapes du processus de la business intelligence :	6
1.3.1 La collecte.....	6
1.3.2 L'intégration.....	7
1.3.3 La diffusion	8
1.3.4 La restitution.....	8
2 Outils et définitions :	9
2.1 ETL :	9
2.2 Data warehouse :	9
2.3 Table de fait et dimension :	10
2.3.1 Différences clés entre la table de faits et la table de dimensions :	10
2.4 Modèles étoile, flocons, constellation	11
2.4.1 Modèles en étoile :	12
2.4.2 Modèles en flocon :	13
2.4.3 Modèle en constellation :	14
2.4.4 Conclusion :	14
2.5 Datamart :	14
2.6 Multidimensionnel expressions MDX :	15
2.7 OLAP :	15
2.7.1 Fonctionnement des systèmes OLAP	16
2.7.2 Cinq opération d'analyse OLAP	16
2.7.2.1 Roll-up (forage vers le haut).	17
2.7.2.2 Drill-down (forage vers le bas).	17
2.7.2.3 Slice (le découpage).	17

2.7.2.4	Dice.....	17
2.7.2.5	Pivot.....	17
2.7.3	Dérivés d'OLAP :	17
2.7.3.1	La technologie de stockage des données OLAP :	18
2.7.3.2	Les techniques de traitements des données OLAP :	18
2.8	OLTP :	18
2.9	Différence entre OLAP et OLTP :	19
2.9.1	Conclusion :	20
Chapitre2 : Environnement technique du projet		21
3	Introduction :	21
4	Environnement matériel :	21
5	Environnement logiciel :	21
5.1	SQL SERVER :	21
5.2	SQL SERVER MANAGEMENT STUDIO :	22
5.3	VISUAL BASIC :	23
5.4	SQL SERVER DATA TOOLS :	23
5.5	MICROSOFT POWER BI :	24
5.6	24
Chapitre 3 : Réalisation		25
6	Introduction :	25
7	Implementation de la datawarehouse	25
7.1	Table de fait sales (Fact Sales) :	25
7.2	Tables de dimensions :	25
7.2.1	DIM PRODUCT :	25
7.2.2	DIM COSTUMER :	26
7.2.3	DIM LOCATION :	26
7.2.4	DIM TIME :	26
7.2.5	DIM SHIPMENT :	27
7.3	MODELE EN ETOILE :	27
8	SSIS :	27
8.1	Etapas d'importation de donnes :	29
9	SSAS :	33
9.1	Principale caractéristique de ssas :	33
9.2	Fonctionnalité des clés ssas :	34
9.3	Développement de cube OLAP :	34
9.4	Power BI :	42
Conclusion Générale.....		43

Abstrait

Le sujet de ce projet tutoré s'inscrit dans le domaine très vaste de l'informatique décisionnelle, en anglais Business Intelligence. Cette discipline récente s'appuie sur un ensemble de processus, de technologies, de compétences et d'application utilisés pour transmettre des informations pertinentes au responsables et managers dans les sociétés.

Plus précisément nous avons été placés face à des problèmes d'analyse de données, d'apprentissage automatique, le tout avec l'aide de l'outil informatique.

Nous allons détailler dans ce rapport comment nous avons fait face à ces défis et quelles méthodes nous avons employées.

Introduction :

Pour comprendre n'importe quel projet, il est nécessaire de comprendre les architectures typiques et les bonnes pratiques pour ce genre de projet. Alors nous commencerons par comprendre l'informatique décisionnelle (BI) afin d'approfondir nos connaissances sur les points clés du projet.

Chapitre 1 : Notions et Techniques

Dans ce chapitre, nous allons présenter les notions d'informatique décisionnelle que nous utilisons ; à savoir l'architecture décisionnelle, le processus ETL et le reporting. Les outils engagés au cours du projet sont spécifiés.

1 Architecture décisionnelle :

1.1 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 leurs permettre d'avoir une vue d'ensemble sur l'activité de l'entreprise et de leurs permettre de prendre des décisions plus avisées à travers des tableaux de bord de suivi et des analyses.

La Business Intelligence (BI) combine analytique métier, data mining, visualisation de données, outils et infrastructure de gestion des données et meilleures pratiques pour aider les entreprises à prendre des décisions basées sur les données. En pratique, on parle de BI lorsque vous disposez d'une vue d'ensemble des données de votre entreprise et utilisez ces données pour favoriser le changement, gagner en efficacité et vous adapter aux évolutions du marché.

Il est important de noter qu'il s'agit là d'une définition très moderne de la BI, un terme qui a connu une histoire très mouvementée. La Business Intelligence traditionnelle est née dans les années 1960 sous la forme d'un système qui partageait des informations entre organisations. Elle s'est développée au cours des années 1980, dans le sillage des modèles informatiques pour la prise de décisions et la transformation de données en informations exploitables, avant de devenir un service à part entière dispensé par des équipes BI avec des solutions s'appuyant sur l'IT. Les solutions BI modernes donnent la priorité à la flexibilité de l'analyse en libre-service, aux données sur des plates-formes fiables dans un cadre de gouvernance et à la rapidité d'accès aux informations, tout en donnant davantage de moyens aux utilisateurs métier.

Bien loin d'être spécifique, le terme « Business Intelligence » désigne tout à la fois les processus et les méthodes utilisés pour collecter, stocker et analyser les données produites par les opérations ou les activités dans le but d'optimiser les performances. Tout cela se combine pour fournir une visibilité totale sur l'activité d'une entreprise, ce qui facilite et améliore la prise de décision et l'action.

1.1.1 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.

1.1.2 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.

1.2 Les principes des systèmes décisionnels :

Le système décisionnel est architecturé de la façon suivante :

- Plusieurs sources de données en lecture.
- Un entrepôt de données fusionnant les données requises.
- Un ETL permettant d'alimenter l'entrepôt de données à partir des données existantes.
- Des magasins de données permettant de simplifier l'entrepôt de données.

- Des applications d'exploitation de données pour présenter l'étude aux utilisateurs finaux et décideurs.

1.3 Les étapes du processus de la business intelligence :

Un système d'information décisionnel assure quatre fonctions qui sont la collecte, l'intégration, la diffusion et la restitution des données.

Le principe de fonctionnement d'un SID peut être résumé par le schéma suivant :

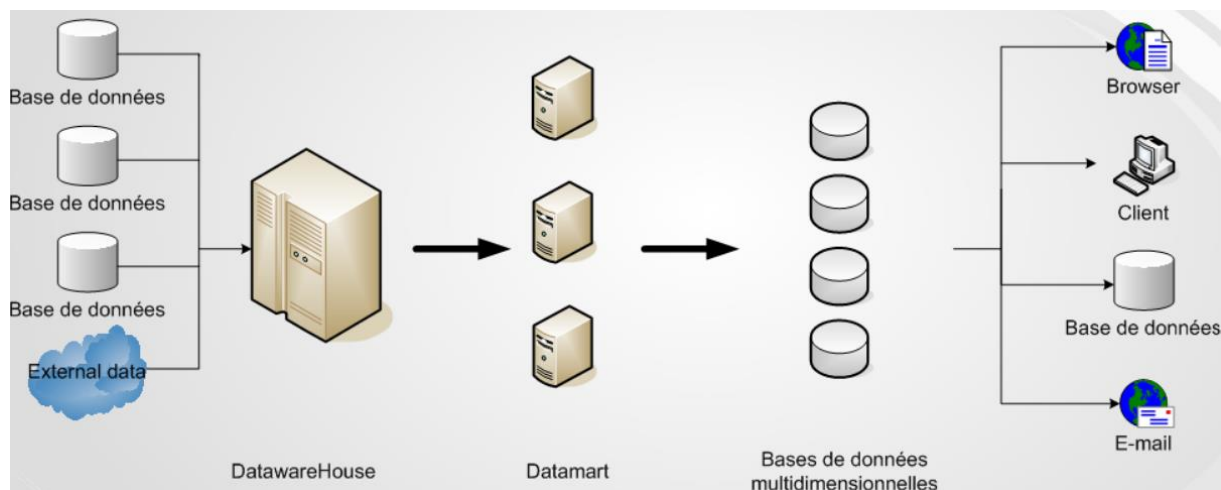


Figure : Architecture générale d'un système *décisionnel*

Diverses sources de données d'entreprise sont utilisées en entrée pour être mises en commun dans un Datawarehouse. Ces données peuvent ensuite être réparties en divers Datamarts, chacun physiquement représenté par une base de données multidimensionnelle. Ces données sont ensuite mises à disposition des utilisateurs et décideurs par divers moyens.

Généralement elles sont mises en forme dans des tableaux de bords et rapports avant de leur être présentées via un serveur web.

1.3.1 La collecte

La première étape de collecte des données va permettre à terme de produire les indicateurs nécessaires au périmètre du SID. Pour cela il convient d'aller chercher les données où elles se trouvent. Les données applicatives métier sont stockées dans une ou plusieurs bases de données correspondant à chaque application utilisée.

La collecte est donc l'ensemble des tâches consistant à détecter, à sélectionner, à extraire et à filtrer les données brutes issues des environnements pertinents pour obtenir des

indicateurs utiles dans le cadre de l'aide à la décision. Les sources de données internes et/ou externes étant souvent hétérogènes tant sur le plan technique que sur le plan sémantique, cette fonction est la plus délicate à mettre en place dans un système décisionnel complexe.

Ces données applicatives sont donc extraites, transformées et chargées dans un entrepôt de données (ou datawarehouse) par un outil de type ETL (Extract-Transform-Load) ou en français ETC (Extraction-Transformation-Chargement).

Ainsi un outil d'ETL permet la synchronisation de données de tous types dans notre datawarehouse et nos datamarts

Un ETL repose sur des connecteurs permettant l'extraction ou l'importation des données de types divers (bases de données de tout type, fichiers XML ou autres formats, ...) et sur des transformateurs qui manipulent les données : agrégations, filtres, conversions, mises en correspondance. Ainsi un tel outil va aller chercher les données d'une entreprise, les transformer pour les mettre en commun et les rendre utilisables dans le cadre de l'aide à la décision pour enfin les injecter dans un entrepôt de données.

Ces actions peuvent être effectuées de manière périodique via l'utilisation de batchs. Par exemple l'action d'extraction, de transformation et de chargement peut être lancée tous les soirs à minuit lorsque plus aucune application n'est utilisée.

1.3.2 L'intégration

Cette deuxième étape est l'intégration des données. Une fois les données centralisées par un outil d'ETL, celles-ci doivent être structurées au sein de l'entrepôt de données. Cette étape est toujours faite par un ETL grâce à un connecteur permettant l'écriture dans le datawarehouse. L'intégration est en fait un prétraitement ayant pour but de faciliter l'accès aux données centralisées aux outils d'analyse.

Ainsi l'intégration consiste à concentrer les données collectées dans un espace unifié, dont le socle informatique essentiel est l'entrepôt de données. Ce dernier est l'élément central du dispositif dans le sens où il permet aux applications d'aide à la décision de bénéficier d'une source d'information homogène, commune, normalisée et fiable. Cette centralisation permet surtout de s'abstraire de la diversité des sources de données.

Lors de cette étape les données sont transformées et filtrées en vue du maintien de la cohérence d'ensemble (les valeurs acceptées par les filtres de l'outil d'ETL de la fonction de collecte mais qui peuvent introduire des incohérences dans les données centralisées sont soit rejetées, soit intégrées après une phase d'adaptation)

Enfin, c'est aussi durant cette étape que sont effectués les éventuels calculs et agrégations communs à l'ensemble du SID.

1.3.3 La diffusion

Cette étape de diffusion met les données à la disposition des utilisateurs. Elle permet la gestion de droits d'accès et respecte donc des schémas correspondant au profil ou au métier de chacun. Ainsi l'accès direct à l'entrepôt de données n'est pas autorisé. En effet ce genre de pratique ne correspond généralement pas aux besoins des décideurs ou analystes. L'objectif principal de l'étape de diffusion est de segmenter les données collectées en contextes qui soient cohérents, simples à utiliser et qui correspondent à une activité décisionnelle particulière (par exemple aux besoins d'un service particulier). En comparaison de l'entrepôt de données peut héberger de nombreuses variables ou indicateurs, un contexte de diffusion n'en présente que quelques dizaines pour rester simple d'exploitation. Chaque contexte peut correspondre à un datamart, bien que le stockage physique ne soit pas sujet à des règles particulières. Généralement un contexte de diffusion est multidimensionnel : il est modélisable sous la forme d'un hypercube et peut donc être mis à disposition via un outil OLAP.

Enfin les différents contextes d'un même SID n'ont pas forcément tous besoin du même niveau de détail selon la cible visée. En effet de nombreux agrégats n'intéressent que certaines applications et ne sont alors donc pas considérés comme des agrégats communs. Ces cumuls ne sont donc pas gérés par la fonction d'intégration mais par la diffusion. Ils peuvent être soit calculés dynamiquement soit stockés de manière persistante.

1.3.4 La restitution

Cette dernière étape, également appelée reporting, se charge de présenter les informations à valeur ajoutée de telle sorte qu'elles apparaissent de la façon la plus lisible possible dans le cadre de l'aide à la décision. Les données sont principalement modélisées par des représentations à base de requêtes afin de constituer des tableaux de bord ou des rapports via des outils d'analyse décisionnelle.

Cette quatrième fonction, la plus visible pour l'utilisateur assure le fonctionnement du poste de travail, le contrôle d'accès aux rapports, la prise en charge des requêtes et la visualisation des résultats sous quelque forme que ce soit.

Le reporting est l'application la plus utilisée dans l'informatique décisionnelle, il permet aux décideurs de :

- Sélectionner des données par période, production, secteur de clientèle, etc.,
- Trier, regrouper ou répartir ces données selon des critères de choix,
- Réaliser des calculs (totaux, moyennes, sommes, pourcentages, écarts, comparatif, ...),

- Présenter les résultats de manière synthétique ou détaillée, généralement sous forme de graphiques.

Les programmes utilisés pour le reporting permettent de faire varier certains critères pour affiner l'analyse. Des instruments de type tableau de bord équipés de fonctions d'analyses multidimensionnelles de type Olap sont aussi utilisés pour cette dernière étape du SID.

2 Outils et définitions :

2.1 ETL :

L'ETL (Extract, Transform, Load) est un processus d'intégration des données qui permet de transférer des données brutes d'un système source, de les préparer pour une utilisation en aval et de les envoyer vers une base de données, un entrepôt de données ou un serveur cible. Dans ce processus la transformation des données intervient sur un serveur intermédiaire avant le chargement sur la cible

2.2 Data warehouse :

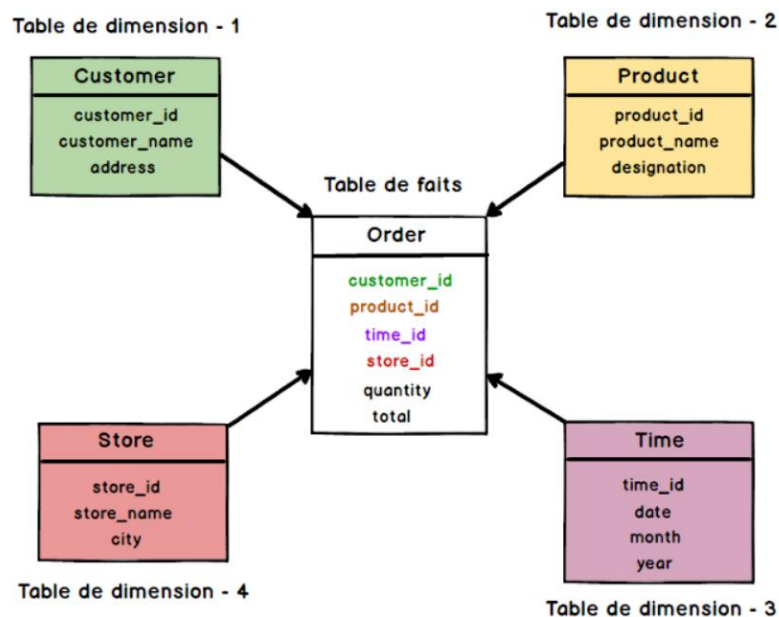
Un Data Warehouse est une base de données relationnelle hébergée sur un serveur dans un Data Center ou dans le Cloud. Il recueille des données de sources variées et hétérogènes dans le but principal de soutenir l'analyse et faciliter le processus de prise de décision. En matière d'intégration dans le système de données existant, le fonctionnement du Data Warehouse est basé sur le processus ETL (Extract, Transform, Load) permettant de charger les données issues des différentes applications.

D'un point de vue plus technique, un Data Warehouse est défini comme un ensemble de données orientées sujet, intégrées, variables dans le temps et non volatiles.

- Orienté sur le sujet : Organisé par thème, il est possible d'utiliser le Data Warehouse pour analyser n'importe quel secteur particulier de l'entreprise.
- Intégré : Avant toute utilisation, les données récupérées de sources hétérogènes internes ou externes sont intégrées au Data Warehouse. Pour cela, il est nécessaire de les mettre en forme et de les unifier pour garantir une certaine cohérence. Les données proviennent principalement de traitement transactionnel en ligne (OLTP).
- Variante temporelle : Les données passées sont également conservées dans le Data Warehouse, contrairement à certains des systèmes transactionnels traditionnels où seules les données les plus récentes sont stockées. Cela permet de visualiser l'évolution dans le temps des différentes données.
- Non volatile : Les données une fois stockées dans l'entrepôt de données ne peuvent jamais être modifiées.

En business intelligence la modélisation d'un data warehouse utilise la notion de table de fait et table de dimension

2.3 Table de fait et dimension :



La table de faits et la table de dimensions sont utilisées pour créer des schémas. L'enregistrement d'une table de faits est une combinaison d'attributs de différentes tables de dimension. La table des faits aide l'utilisateur à analyser les dimensions de l'entreprise, ce qui aide à prendre des décisions pour améliorer son activité. D'ailleurs, les tables de dimensions aident à rassembler les dimensions avec lesquels les mesures doivent être prises.

2.3.1 Différences clés entre la table de faits et la table de dimensions :

	Table de faits	Table de dimensions
Clé primaire	La table de faits contient une clé primaire qui est une concaténation de clés primaires de toutes les tables de dimensions.	Chaque table de dimension contient sa clé primaire.
Signification	La table de faits contient les mesures avec des attributs d'une table de dimension.	La table de dimension contient les attributs avec lesquels la table de faits calcule la métrique.
Taille de la table	La table de faits se développe verticalement.	La table de dimensions se développe horizontalement.
Attribut & Records	La table de faits contient moins d'attributs et plus d'enregistrements.	La table de dimension contient plus d'attributs et moins d'enregistrements.
Création	La table de faits peut être créée uniquement lorsque les tables de dimensions sont complétées.	Les tables de dimension doivent être créées en premier.
Schéma	Un schéma contient moins de tables de faits.	Un schéma contient plus de tables de dimension.
Les attributs	La table de faits peut contenir des données au format numérique et au format textuel.	La table de dimension contient toujours des attributs au format textuel.

figure : table différence entre la table de faits et la table de dimensions

- **La table de faits** contient des mesures sur les attributs d'une **table de dimension**.
- **La table de faits** contient plus d'enregistrements et moins d'attributs par rapport à la **table de dimension**, tandis que la **table de dimensions** contient plus d'attributs et moins d'enregistrements.
- La taille de **table de faits** augmente verticalement tandis que la taille de **table de dimension** augmente horizontalement.
- Chaque **table de dimension** contient une clé primaire pour identifier chaque enregistrement de la table, tandis que la **table de faits** contient une clé concaténée qui est une combinaison de toutes les clés primaires de toutes les **tables de dimension**.
- La **table de dimension** doit être enregistré avant la création de la **table de faits**.
- Un schéma contient moins de **tables de faits** mais plus de **tables de dimension**.
- Les attributs de la **table de faits** sont aussi bien numériques que textuels, mais les attributs de la **table de dimension** ont uniquement des attributs textuels.

2.4 Modèles étoile, flocons, constellation

Dans un Data Warehouse les données et leurs relations sont organisées suivant un modèle de données spécifique. Le choix du modèle de données structure et définit le design du Data Warehouse. Nous avons vu qu'il existait trois modélisations possibles :

- La modélisation en étoile.
- La modélisation en flocons.
- La modélisation en constellation.

2.4.1 Modèles en étoile :

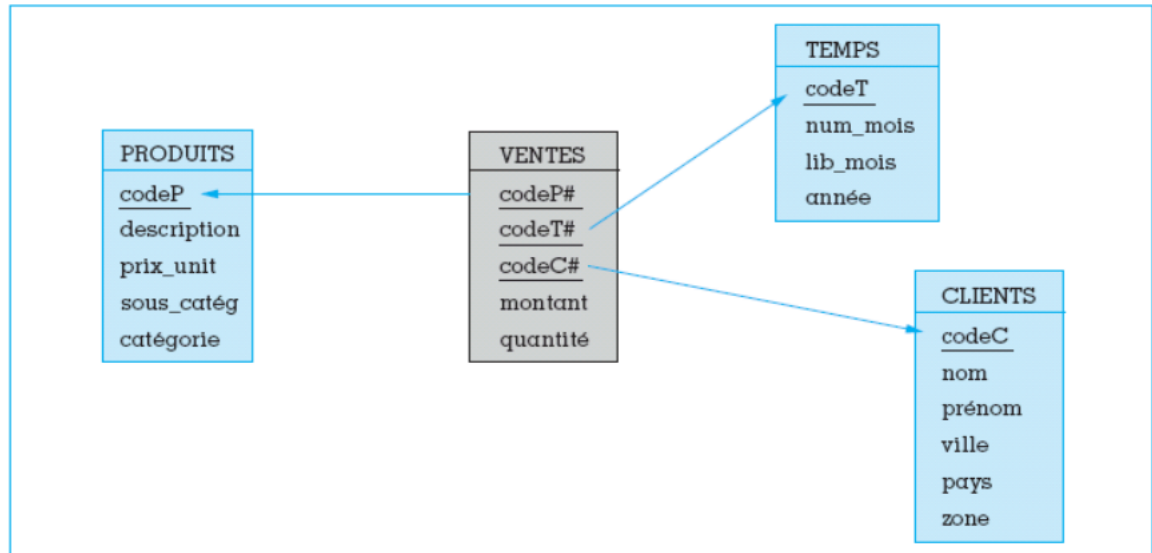


Figure : Exemple de schéma en Etoile

La modélisation en étoile est le modèle le plus simple et celui qui est le plus couramment utilisé dans le design des Data Warehouses. Le schéma ci-dessus permet de comprendre l'origine de cette appellation « en étoile ».

Dans ce modèle, la table des faits est au centre du schéma et est entourée par des tables de dimensions. Cela ressemble visuellement à une étoile – surtout lorsqu'il y a 5 branches ! Lorsque la table des faits est entourée par un grand nombre de petites tables de dimensions, on parle de schéma centipède. C'est une variante du modèle en étoile. Pour filer la métaphore, les tables de dimensions peuvent être comparées à des planètes.

Les schémas en étoile sont très utilisés pour concevoir les Data Marts. La modélisation en étoile peut être directement associée à l'approche Top-Down que nous décrivons plus haut. La modélisation est obtenue par agrégation de tables de dimensions extraites du Data Warehouse.

Cette modélisation a une orientation très métier dans la mesure où chaque table des faits correspond à un objet d'étude : les ventes, les achats, la logistique, la production, etc. La table des faits contient tous les faits et toutes les mesures associées à l'objet d'étude – l'essentiel des données qu'elle contient sont des données chiffrées : montants, quantités, taux... Dans l'exemple ci-dessus, ces faits sont au nombre de deux : prix et quantités. Au-dessus de ces faits figurent les clés étrangères renvoyant vers les 5 tables de dimensions.

Les tables de dimensions correspondent aux différents axes d'analyse utilisés pour étudier l'objet d'étude. Les mesures observables dans la table des faits sont issues de l'agrégation des données stockées dans les tables de dimensions. Visuellement parlant, les lignes de la table des faits correspondent aux faits mesurés et les colonnes aux différentes dimensions.

Comme nous le disions en introduction, le principal avantage de la modélisation en étoile est qu'il permet d'optimiser la performance et la simplicité des requêtes effectuées sur de grands volumes de données. Nous savons maintenant pourquoi. Le schéma en étoiles, par sa structure essentiellement redondante, permet d'agréger la table de faits avec n'importe quelle dimension en une unique opération de jointure.

2.4.2 Modèles en flocon :

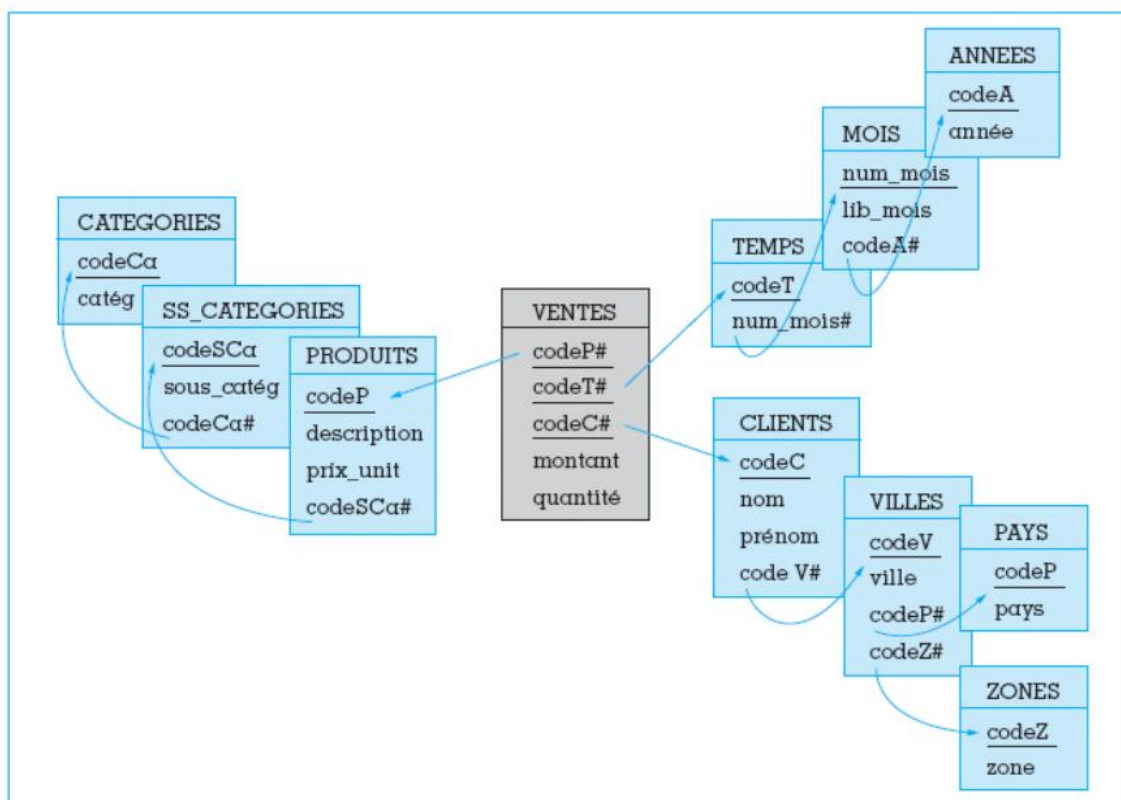


Figure : Exemple de schéma en flocon

Le modèle en flocon de neige est le modèle en étoile avec une normalisation des dimensions. Les dimensions ne sont pas liées entre elles dans le modèle en étoile or au niveau du modèle en flocon de neige il peut exister une hiérarchie entre les dimensions. Au niveau du modèle en étoile les dimensions sont dénormalisées contrairement au modèle en flocon de neige où les dimensions sont normalisées. Le modèle en flocon de neige est complexe avec de nombreuses jointures souvent coûteuses contrairement au modèle en étoile qui évite les jointures.

2.4.3 Modèle en constellation :

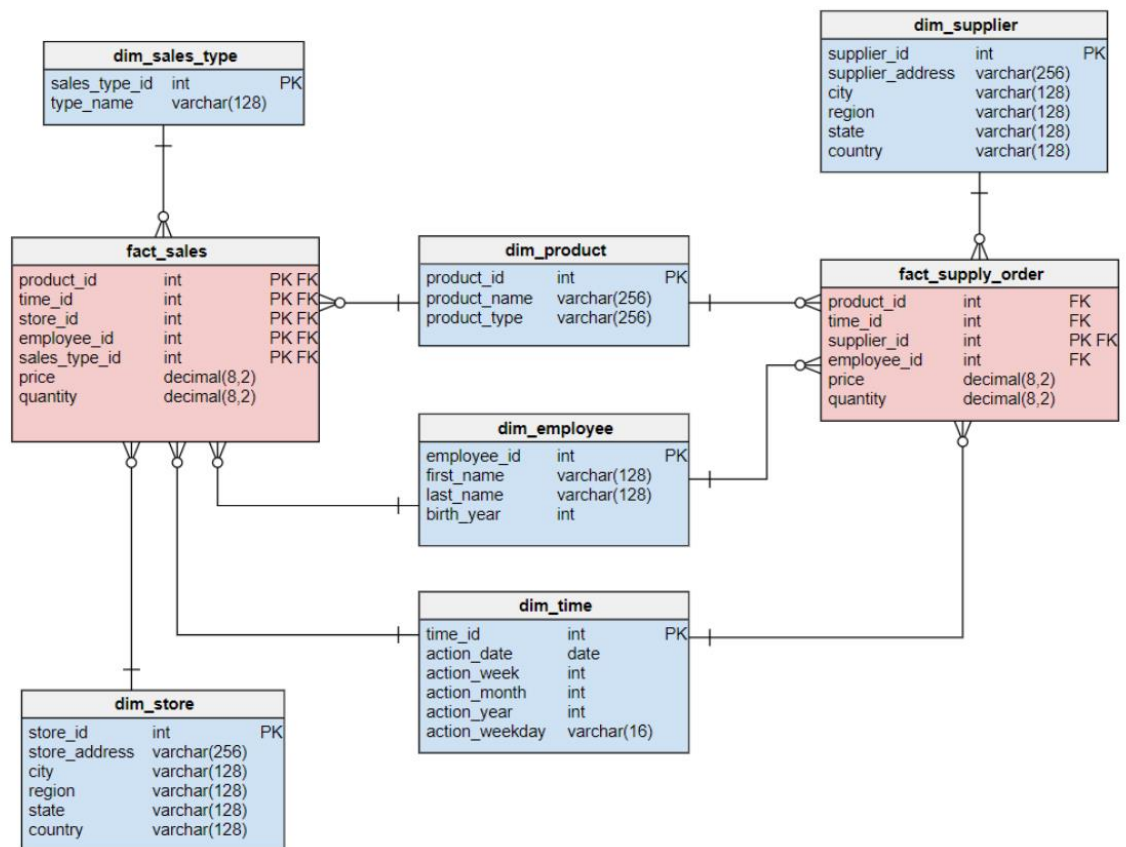


Figure : Exemple de schéma en constellation

Lorsque plusieurs schémas en étoile partagent des dimensions (voir dimensions conformes) on parle de **schéma en constellation**.

2.4.4 Conclusion :

La modélisation en étoile est une approche pour organiser un Data Warehouse. Elle a l'avantage de la simplicité. C'est pour cette raison qu'elle reste la plus communément utilisée. Si vous n'êtes pas limité en matière d'espace de stockage et que vous prenez soin de veiller à l'intégrité de vos données, la modélisation en étoile est un très bon choix, idéal pour commencer.

2.5 Datamart :

Un Data Warehouse est un système utilisé pour stocker les données de l'entreprise en vue de faire de l'analyse et du reporting (Business Intelligence). Les données du Data Warehouse sont importées dans le système en batch à partir des sources de données dynamiques issues des systèmes transactionnels/opérationnels de l'organisation. Les données du Data Warehouse sont largement « statiques » : elles ne font pas l'objet de mise à jour.

Les Data Marts sont des compartiments du Data Warehouse conçus pour répondre à un besoin métier spécifique. Un Data Mart regroupe et organise toutes les données

nécessaires au traitement d'une problématique précise. Par exemple : toutes les données relatives aux ventes, ou bien toutes celles relatives à la logistique, au marketing, à la production, etc.

Pour prendre une image, le Data Warehouse peut être comparé à un immeuble de bureaux et les Data Marts aux différents bureaux qui le constituent.

Les Data Marts ont donc une utilité du point de vue des utilisateurs mais aussi en terme de sécurité. Au lieu que tous les utilisateurs aient accès à toutes les données stockées, l'organisation en Data Marts permet de contrôler et de restreindre les accès aux données en fonction des besoins des utilisateurs.

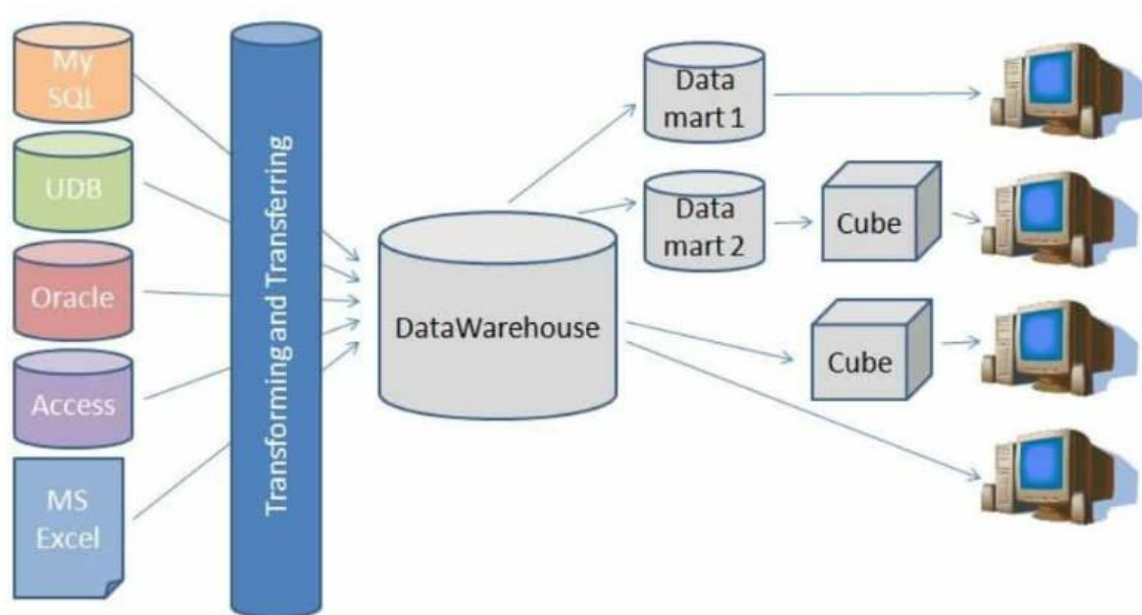


Figure : Datawarehouse architecture

2.6 Multidimensionnel expressions MDX :

Ce langage est l'équivalent de SQL mais pour les cubes au lieu des bases de données. Les applications de Business Intelligence l'utilisent largement pour permettre aux utilisateurs finaux d'analyser les données financières ou de production d'une entreprise donnée.

En d'autre mot MDX est un langage de requêtes pour les bases de données OLAP, analogue au rôle de SQL pour les bases de données relationnelles. C'est aussi un langage de calcul avec une syntaxe similaire à celle des tableurs.

2.7 OLAP :

OLAP (Online Analytical Processing) permet aux utilisateurs d'analyser des données présentes de plusieurs systèmes de base de données en même temps. Alors que

les bases de données relationnelles sont considérées comme bidimensionnelles, les données OLAP sont multidimensionnelles, ce qui signifie que l'information peut être comparée de nombreuses façons différentes. Par exemple, une entreprise peut comparer ses ventes d'ordinateurs en juin avec ses ventes en juillet, puis comparer ces résultats avec les ventes d'un autre endroit, qui pourraient être stockées dans une base de données différente.

Un serveur OLAP est nécessaire pour organiser et comparer les informations. Les clients peuvent analyser différents ensembles de données à l'aide des fonctions intégrées au serveur OLAP. En raison de ses puissantes capacités d'analyse de données, le traitement OLAP est souvent utilisé pour le data mining, qui vise à découvrir de nouvelles relations entre différents ensembles de données.

L'OLAP fournit aux analystes, aux gestionnaires et aux cadres supérieurs l'information dont ils ont besoin pour prendre des décisions efficaces sur les orientations stratégiques d'une organisation. L'OLAP peut fournir des informations précieuses sur la performance de leur entreprise, ainsi que sur les améliorations qu'ils peuvent apporter.

Les outils OLAP sont optimisés pour les requêtes et le reporting. Ceci contraste avec les applications OLTP (Online Transactional Processing), qui s'occupent principalement des tâches basées sur les transactions. Les outils OLAP permettent aux utilisateurs d'analyser des données multidimensionnelles de manière interactive à partir de perspectives multiples. OLAP peut être utilisé pour trouver des tendances et obtenir une vue d'ensemble des données. Il peut également être utilisé pour des calculs complexes et pour créer des scénarios "et si" pour la planification prévisionnelle.

Les applications OLAP typiques incluent les rapports d'affaires pour les ventes, le marketing, les rapports de gestion, la gestion des processus d'affaires, la budgétisation et les prévisions, les rapports financiers, et plus encore.

2.7.1 Fonctionnement des systèmes OLAP

Pour faciliter ce type d'analyse, les données sont recueillies à partir de multiples sources de données et stockées dans des Data warehouse, puis nettoyées et organisées en cubes de données. Chaque cube OLAP contient des données classées par dimensions (telles que les clients, la région géographique de vente et la période de temps) dérivées par tables dimensionnelles dans les Data Warehouse. Les dimensions sont ensuite complétées par les membres (tels que les noms de clients, les pays et les mois) qui sont organisés de manière hiérarchique. Les cubes OLAP sont souvent pré-résumés dans toutes les dimensions afin d'améliorer considérablement le temps de requête par rapport aux bases de données relationnelles.

2.7.2 Cinq opération d'analyse OLAP

Les systèmes OLAP sont conçus pour repérer les intersections entre ces multiples dimensions. Les analystes peuvent ensuite effectuer cinq types d'opérations d'analyse

OLAP à partir de ces bases de données multidimensionnelles :

2.7.2.1 *Roll-up (forage vers le haut).*

Aussi appelée consolidation ou éclatement, effectue l'agrégation sur un cube de données soit en montant dans la hiérarchie, soit en réduisant les dimensions.

2.7.2.2 *Drill-down (forage vers le bas).*

Cette opération permet d'obtenir plus de détails soit en descendant dans la hiérarchie, soit en ajoutant une dimension. Cela permet aux analystes de naviguer plus profondément parmi les dimensions, par exemple en descendant de "période de temps" à "années" et "mois" pour représenter la croissance des ventes d'un produit.

2.7.2.3 *Slice (le découpage).*

L'opération de découpe sélectionne une dimension particulière à partir d'un cube donné et fournit un nouveau sous-cube, tel que "les ventes en 2017".

2.7.2.4 *Dice*

Cette opération est similaire à une « Slice ». La différence dans le « Dice » est que vous sélectionnez 2 dimensions ou plus qui permet de créer un sous-cube. Cela permet à un analyste de sélectionner des données à partir de plusieurs dimensions à analyser, telles que "les ventes de ballons de plage bleus en Iowa en 2017".

2.7.2.5 *Pivot*

L'opération de pivotement est également connue sous le nom de rotation. Il fait pivoter les axes de données en vue de fournir une présentation alternative des données. Les analystes peuvent obtenir une nouvelle vue des données en faisant tourner les axes de données du cube.

Le logiciel OLAP localise alors l'intersection des dimensions et les affiche. Le résultat est la "mesure" ; chaque cube OLAP possède au moins une à peut-être des centaines de mesures, qui sont dérivées des informations stockées dans les tables de fait de l'entrepôt de données.

2.7.3 Dérivés d'OLAP :

- MOLAP = Multidimensional OLAP
- ROLAP = Relational OLAP
- HOLAP = Hybrid OLAP
- DOLAP = Desktop OLAP.

Il existe deux critères pour classer une solution OLAP en tant que MOLAP, ROLAP, HOLAP ou DOLAP :

2.7.3.1 La technologie de stockage des données OLAP :

1. Base de données relationnelles : les données sont stockées dans un SGBD relationnel. Il permet un stockage presque infini des données OLAP. (Rolap)
2. Base de données dimensionnelle (Cube) : Les données sont stockées dans une base de données multidimensionnelles le plus souvent propriétaires. Cependant il y'a des limitations quant à la quantité de données ! (Molap)
3. Fichiers sur le poste client : Une petite quantité de données est stockée directement sur le poste client de l'utilisateur (Dolap)

2.7.3.2 Les techniques de traitements des données OLAP :

1. SQL : SQL est utilisé pour effectuer les différents traitements sur les données OLAP. On réalise des roll-ups, des drills-down... en utilisant des requêtes en général très complexes et très exigeantes en terme de ressources et de temps d'exécution. (Rolap)
2. Serveur de traitement OLAP : Il s'agit de l'approche la plus adaptée aux traitements de données OLAP. Un serveur, conjointement avec la base de données, est alors dédié à effectuer les différents traitements de données OLAP. Les performances sont excellentes en général ! (Molap)
3. Client de traitement OLAP : Une approche qui perd beaucoup de terrain vu l'émergence des postes clients légers. Un nombre limité de traitement OLAP se font sur le poste client de l'utilisateur. (Dolap).

En se basant sur ces deux critères il devient donc clair que :

Molap = Base de données dimensionnelle + Serveur de traitement OLAP

Rolap = Base de données relationnelle + SQL avancé

Dolap = Fichier sur le poste client + Client de traitement OLAP

Holap = Molap pour les données sommaires + Rolap pour les données détaillées

2.8 OLTP :

OLTP (Online Transaction Processing) est une classe de logiciels capables de prendre en charge les applications orientées transactions sur Internet. Généralement, les systèmes OLTP sont utilisés pour la saisie des commandes, les transactions financières, la gestion de la relation client (CRM) et les ventes au détail. Ces systèmes comptent un grand nombre d'utilisateurs qui effectuent des transactions courtes. Les requêtes de base de données sont généralement simples, nécessitent des temps de réponse inférieurs à la

seconde et retournent relativement peu d'enregistrements. Un attribut important d'un système OLTP est sa capacité à maintenir la simultanéité. Pour éviter les points de défaillance uniques, les systèmes OLTP sont souvent décentralisés.

2.9 Différence entre OLAP et OLTP :

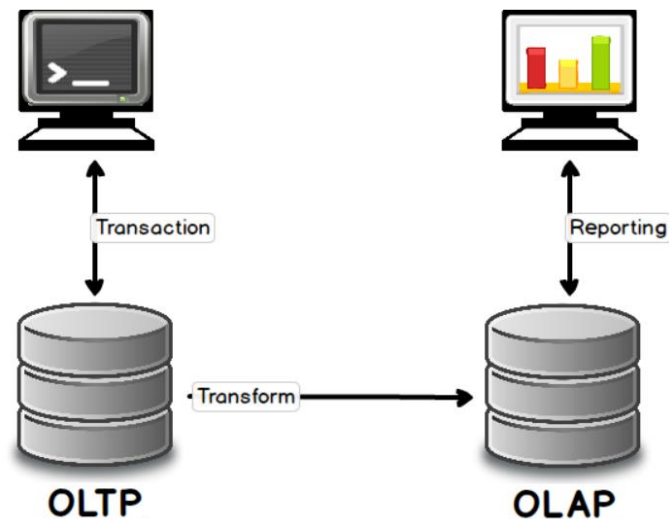


Figure : Différence entre OLTP et OLAP

OLTP et **OLAP** sont les deux systèmes de traitement en ligne. **OLTP** est un traitement transactionnel alors que **OLAP** est un système de traitement analytique. **OLTP** est un système qui gère des applications orientées vers les transactions sur Internet, par exemple ATM. **OLAP** est un système en ligne qui répond à des requêtes analytiques multidimensionnelles telles que le reporting financier, les prévisions, etc. La différence clé entre **OLTP** et **OLAP** est que **OLTP** est un système qui sert à effectuer des modifications de base de données en ligne.

	OLTP	OLAP
Définition	C'est un système transactionnel en ligne qui sert à effectuer des modifications dans une base de données.	C'est un système de récupération de données et d'analyse de données en ligne.
Transaction	OLTP a des transactions courtes.	OLAP a des transactions longues.
Les données	OLTP et ses transactions constituent la source originale de données.	Différentes bases de données OLTP deviennent la source de données pour OLAP.
Intégrité	La base de données OLTP doit maintenir la contrainte d'intégrité des données.	La base de données OLAP n'est pas fréquemment modifiée. Par conséquent, l'intégrité des données n'est pas affectée.
Normalisation	Les tables dans la base de données OLTP sont normalisées (3NF).	Les tables dans la base de données OLAP ne sont pas normalisées.
Requêtes	Des requêtes plus simples.	Des Requêtes plus complexes

figure : Table de comparaison

2.9.1 Conclusion :

- **OLTP** est un système de transaction en ligne, tandis que **OLAP** est un système de récupération et d'analyse de données en ligne.
- Les données transactionnelles en ligne deviennent la source de données pour **OLTP**. Toutefois, la base de données **OLTP** différente devient la source de données pour **OLAP**.
- Les principales opérations d'**OLTP** sont l'insertion, la mise à jour et la suppression alors que l'opération principale d'**OLAP** consiste à extraire des données multidimensionnelles pour les analyser.
- **OLTP** a des transactions courtes mais fréquentes, tandis que **OLAP** effectue des transactions longues et moins fréquentes.
- Le temps de traitement de la transaction **OLTP** est plus long que celui de **OLAP**.
- Les requêtes **OLAP** sont plus complexes en ce qui concerne les **OLTP**.
- Les tables de la base de données **OLTP** doivent être normalisées (3NF) alors que les tables de la base de données **OLAP** peuvent ne pas être normalisées.
- Comme les transactions **OLTP** exécutent fréquemment des transactions dans une base de données, au cas où une transaction échoue au milieu, cela peut nuire à l'intégrité des données et doit donc assurer l'intégrité des données. Bien que dans **OLAP**, la transaction soit moins fréquente, elle ne gêne pas beaucoup l'intégrité des données.

Chapitre2 : Environnement technique du projet

3 Introduction :

Pour mettre en place notre système, nous avons utilisé un environnement de développement qui a assuré le bon déroulement de la phase implémentation. Cet environnement comporte des outils matériels ainsi que logiciels.

4 Environnement matériel :

Durant ce présent projet, tout le travail a été réalisé sur deux ordinateurs Dell qui se caractérisent par :

- Processeur : Intel (R) Core(TM) i7-6600U CPU @ 2.80GHZ.
- Mémoire installé (RAM) : 8,00 Go.
- Type de système : système d'exploitation 64bits, processeur x64.
- Edition Windows : Windows 10 Professionnel.

5 Environnement logiciel :

Nous avons utilisé les outils suivants :

5.1 SQL SERVER :



Figure : Logo Microsoft SQL Server

Microsoft SQL Server est un système de gestion de base de données (SGBD) en langage SQL incorporant entre autres un SGBDR (« SGBD relationnelle ») développé et commercialisé par la société Microsoft. Il fonctionne sous les OS Windows et Linux (depuis mars 2016), mais il est possible de le lancer sur Mac OS via Docker, car il en existe une version en téléchargement sur le site de Microsoft.

5.2 SQL SERVER MANAGEMENT STUDIO :



Figure: logo SQL server management studio

SQL Server Management Studio (SSMS) est une application logicielle lancée pour la première fois avec Microsoft SQL Sever 2005 qui est utilisée pour configurer, gérer et administrer tous les composants de Microsoft SQL server. C'est le successeur d'**Enterprise Manager** dans SQL 2000 ou avant. L'outil comprend à la fois des éditeurs de scripts et des outils graphiques qui fonctionnent avec les objets et les fonctionnalités du serveur.

SSMS est l'un des outils de gestion de SQL Server, quel que soit votre emplacement, utilisé pour concevoir des requêtes et gérer des bases de données et des entrepôts de données via un ordinateur personnel ou le Cloud.

En réalité, SSMS est un environnement intégré qui fournit des outils de configuration, de surveillance et d'administration des instances et des bases de données SQL Server.

Une fonctionnalité centrale de SSMS est l'Explorateur d'objets, qui permet à l'utilisateur de parcourir, sélectionner et agir sur n'importe lequel des objets du serveur.

Il a également livré une édition Express distincte qui peut être téléchargée gratuitement, mais les versions récentes de SSMS sont entièrement capables de se connecter et de gérer n'importe quelle instance SQL Server Express. Microsoft a également intégré la rétrocompatibilité pour les anciennes versions de SQL Server, permettant ainsi à une version plus récente de SSMS de se connecter aux anciennes versions des instances de SQL Server. Il est également livré avec Microsoft SQL Server Express 2012, ou les utilisateurs peuvent le télécharger séparément.

À partir de la version 11, l'application était basée sur le shell visual studio 2010 utilisant wpf pour l'interface utilisateur. Les versions 18 et ultérieures sont basées sur visual studio 2017 isolated Shell.

En juin 2015, Microsoft a annoncé son intention de publier les futures versions de SSMS indépendamment des versions du moteur de base de données SQL Server.

5.3 VISUAL BASIC :



Figure : Logo Visual Basic

Visual Basic (VB) est un langage de programmation événementielle de troisième génération ainsi qu'un environnement de développement intégré créé par Microsoft pour son modèle de programmation COM. Visual Basic est directement dérivé du BASIC et permet le développement rapide d'applications, la création de l'interface utilisateur graphique, l'accès aux bases de données en utilisant des technologies DAO, ADO et RDO ainsi que la création de contrôles ou objets ActiveX. Les langages de script tels que Visual Basic for Applications et VBScript sont syntaxiquement proches de Visual Basic, mais s'utilisent et se comportent de façon sensiblement différente.

Un programme en VB peut être développé en utilisant les composants fournis avec Visual Basic lui-même. Les programmes écrits en Visual Basic peuvent aussi utiliser l'API Windows, ceci nécessitant la déclaration dans le programme des fonctions externes.

Dans une étude conduite en 2005, 62 % des développeurs déclaraient utiliser l'une ou l'autre forme de Visual Basic. Selon la même étude, les langages les plus utilisés dans le domaine commercial sont Visual Basic, C++, C, et Java.

La dernière mise à jour de Visual Basic est la version 6.0, sortie en 1998. Le support étendu Microsoft a pris fin en 2008. À partir de la version 7, le Visual Basic subit des changements substantiels le rapprochant de la plate-forme « .NET », et qui amènent Microsoft à le commercialiser sous le nom de Visual Basic.NET.

5.4 SQL SERVER DATA TOOLS :

SQL Server Data Tools (SSDT) est un outil de développement moderne permettant de créer des bases de données relationnelles SQL Server, des bases de données SQL Azure, des modèles de données Analysis Services (AS), des packages Integration Services (IS) et des rapports Reporting Services (RS), qui est intégré à Visual Studio. Avec SSDT, vous pouvez concevoir et déployer n'importe quel modèle de données SQL Server avec la même facilité que vous développeriez une application dans Visual Studio.

L'avantage par rapport à l'utilisation de SSMS est qu'avec SSDT dans Visual Studio et Azure DevOps, vous obtenez :

- Code source de la base de données archivé dans le référentiel avec les identifiants de validation
- Travaillez sur les branches, appliquez les politiques de branche avec examen par les pairs et enregistrement contrôlé du code (avec des versions de demande d'extraction)
- Construction continue et déploiements continus

5.5 MICROSOFT POWER BI :

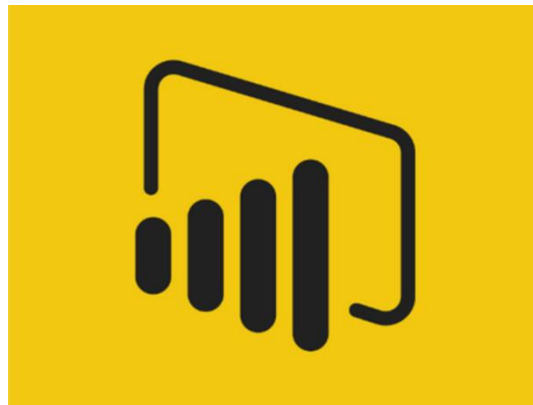


Figure : Logo Power BI

5.6

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

Power BI est un ensemble de services logiciels, d'applications et de connecteurs qui fonctionnent ensemble pour transformer différentes sources de données en informations visuelles, immersives et interactives. Plusieurs sources de données peuvent être utilisées telles que des fichiers Excel, des sources SQL, ou des entrepôts de données hybrides locaux ou sur le cloud. Les données sont personnalisées et interactives avec une interface suffisamment simple pour que les utilisateurs finaux créent leurs propres rapports et tableaux de bord. L'objectif est de faciliter la création des tableaux de bord afin d'améliorer les moyens de communications et de collaboration proposés par Microsoft. Il permet donc de collecter, construire et exposer les données au travers d'indicateurs. Son ergonomie permet par la suite d'animer des présentations interactives qui aideront à la prise de décision.

Chapitre 3 : Réalisation

6 Introduction :

Cette partie contient le dernier volet de ce rapport, elle a comme but d'exposer les étapes suivies pour achever ce travail

7 Implementation de la datawarehouse

Bien évidemment cela commence par la création des tables de fait et table de dimension pour pouvoir modéliser notre étoile on a créé :

7.1 Table de fait sales (Fact Sales) :

Fact Sales	
	[Row Id]
	[Order Id]
	[Customer Id]
	[Product Id]
	Sales
	Quantity
	Discount
	Profit

7.2 Tables de dimensions :

7.2.1 DIM PRODUCT :

Dim Product	
🔑	[Product Id]
	Category
	[Sub-Category]
	[Product Name]

7.2.2 DIM COSTUMER :

Dim Customer	
🔑	[Customer Id]
	[First Name]
	[Last Name]
	Segment

7.2.3 DIM LOCATION :

Dim Location	
🔑	[Customer Id]
	Country
	City
	State
	[Postal Code]
	Region

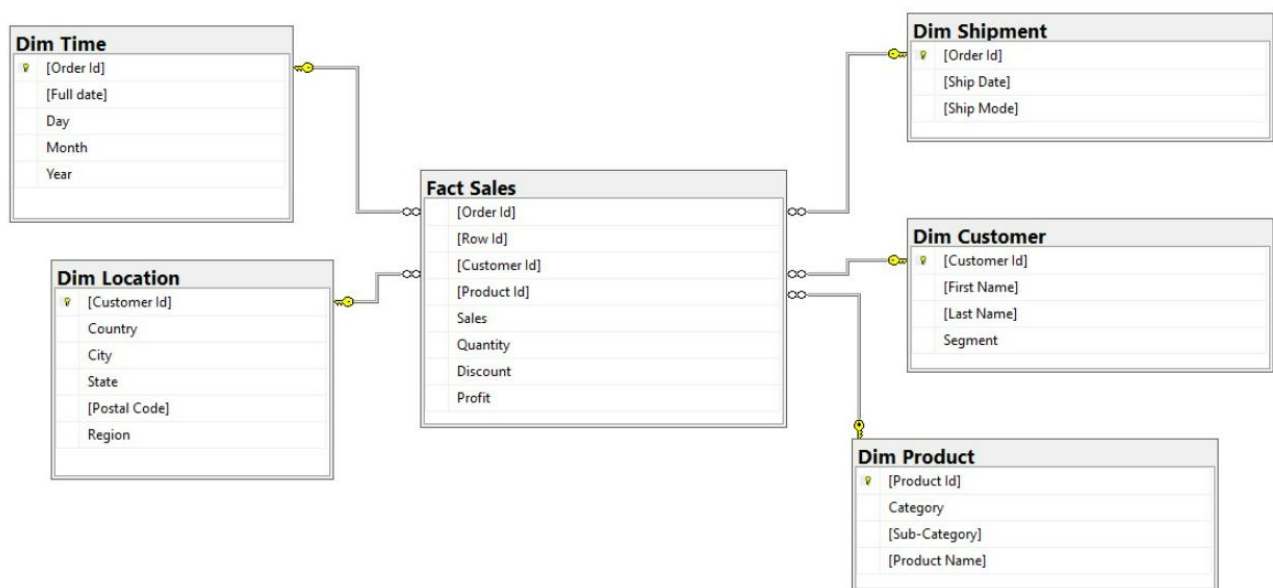
7.2.4 DIM TIME :

Dim Time	
🔑	[Order Id]
	[Full date]
	Day
	Month
	Year

7.2.5 DIM SHIPMENT :



7.3 MODELE EN ETOILE :



Après vient la deuxième étape : importation de data

8 SSIS :

La forme complète de SSIS est SQL Server Integration Services, il s'agit essentiellement d'un composant du logiciel de base de données Microsoft **SQL Server** qui est utilisé pour

effectuer la migration des données à grande échelle. En tant que plate-forme, SSIS améliore l'intégration des données et les applications de flux de travail.

SQL Server Integration Services est une plate-forme de développement de solutions d'intégration et de transformation de données au niveau de l'entreprise. Utilisez Integration Services pour gérer des bases de données, copier des fichiers et télécharger des données, charger des entrepôts de données, nettoyer et analyser des données et gérer des objets SQL Server.

Integrating Services extrait et transforme des données à partir d'un large éventail de sources, telles que des fichiers XML, des fichiers plats et des sources de données relationnelles, et charge ces données sur une ou plusieurs destinations.

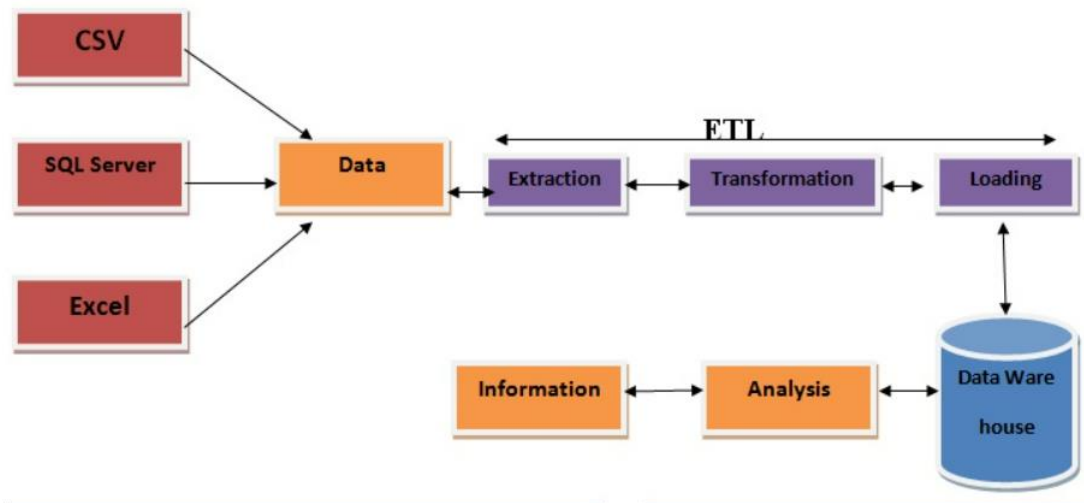
De nombreuses tâches et transformations sont intégrées à Integration Services pour créer des packages, et vous pouvez les gérer dans la base de données de catalogue Integration Services.

Avançons dans ce didacticiel SSIS et apprenons quels sont les composants de SSIS :

- Flux de contrôle (pour le stockage des conteneurs et des tâches)
- Flux de données (source, destination et transformations)
- Gestionnaire d'événements (pour gérer les messages et les e-mails)
- Package Explorer (pour offrir une vue tout-en-un)
- Paramètres (pour favoriser l'interaction de l'utilisateur)

Nous pouvons utiliser SSIS pour implémenter [ETL](#) ou extraire, transformer et charger des données dans un entrepôt de données.

L'acronyme ETL signifie Extraction, Transformation et Chargement. Ce processus implique l'extraction de données, leur transformation et leur chargement dans le référentiel final. ETL est le processus de chargement des données dans l'entrepôt à partir du système source



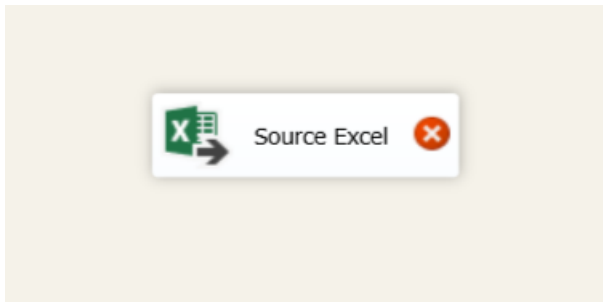
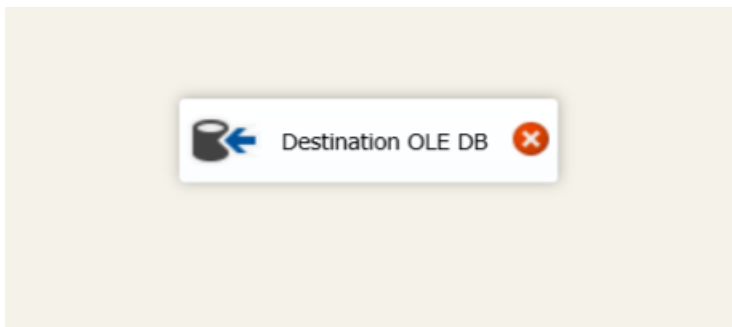
Ensuite, dans le didacticiel SSIS, jetez un œil aux fonctionnalités clés de SSIS -

- Nettoyage et profilage des données pour une meilleure qualité des données
- Intégration fluide des données à partir de sources de données disparates
- Intégration transparente avec d'autres composants des produits Microsoft SQL
- Environnement de studio enrichi et outils/assistants graphiques
- Fonctionnalités de flux de travail telles que le protocole de transfert de fichiers
- API pour la modélisation d'objets SSIS
- Mise en œuvre efficace de la connectivité/intégration des données à haut débit
- Connecteurs de source de données packagés
- Requête d'exploration de données organisée et transformation de recherche
- Gestion du master et des métadonnées

8.1 Etapes d'importation de données :

Etape1 : Créer un flux de données transférer les données d'une source Excel vers une destination ole db Il aide à définir le flux de données entre la source et la destination.



ETAPE2 : créer une source excel**ETAPE3** : créer une destination**Etape4** : configurer la feuille Excel

Éditeur de source Excel

Configurez les propriétés qui permettent à la tâche de flux de données d'obtenir des données à partir du fournisseur Excel.

Spécifiez un gestionnaire de connexions, une source de données ou une vue de source de données pour la source Excel. Ensuite, sélectionnez le mode utilisé pour accéder aux données à l'intérieur de la source. Une fois que vous avez choisi le mode d'accès aux données, faites votre choix parmi les autres options d'accès proposées.

Gestionnaire de connexions Excel :

Gestionnaire de connexions Excel

Nouveau...

Mode d'accès aux données :

Table ou vue

Nom de la feuille Excel :

Aperçu...

Etape5 : configurer la destination :

Éditeur de destination OLE DB

Configurez les propriétés utilisées pour insérer des données dans une base de données relationnelle en utilisant un fournisseur OLE DB.

Gestionnaire de connexions
Mappages
Sortie d'erreur

Spécifiez un gestionnaire de connexions OLE DB, une source de données ou une vue de source de données, puis sélectionnez le mode d'accès aux données. Si vous utilisez le mode d'accès par commande SQL, spécifiez la commande SQL soit en la tapant, soit en utilisant le Générateur de requêtes. Pour un accès aux données par chargement rapide, définissez les options de mise à jour

Gestionnaire de connexions OLE DB :

Mode d'accès aux données :

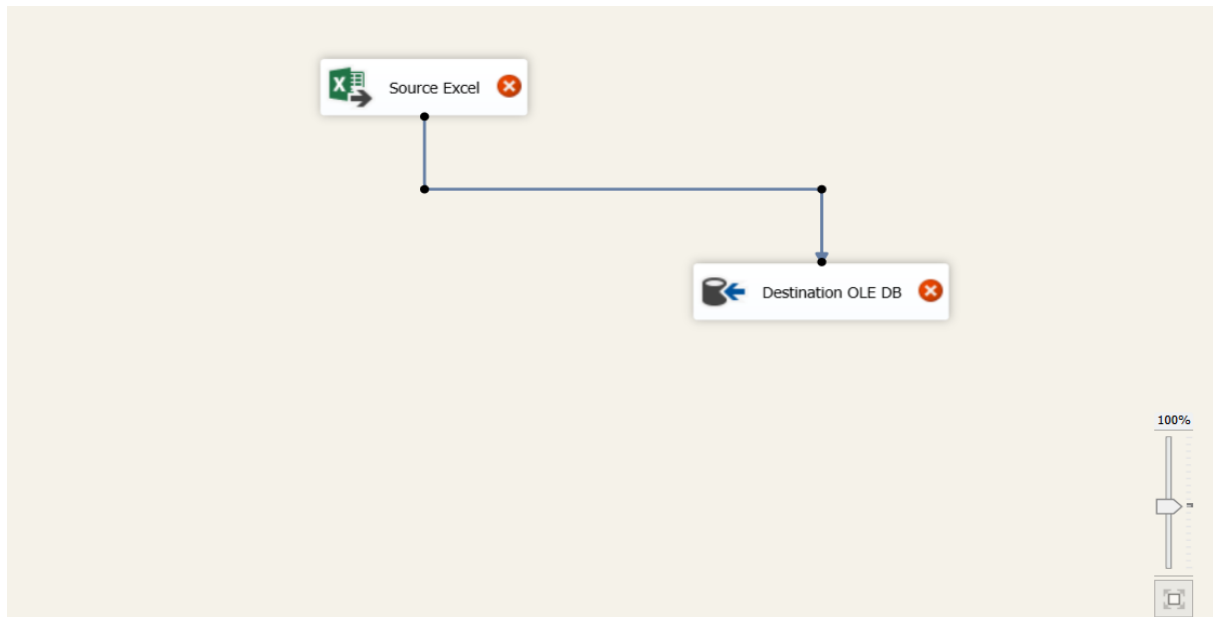
Nom de la table ou de la vue :

Nouveau...

Table ou vue

Afficher les

Etape6 : Attacher la source a la destination



Etape7 : Exécuter le package



9 SSAS :

SSAS est une technologie de la pile Microsoft Business Intelligence, qui est utilisée pour développer des solutions de traitement analytique en ligne (OLAP). Il peut également être utilisé pour créer des cubes à l'aide de données provenant de magasins de données/d'entrepôts de données pour une analyse de données plus rapide et efficace.

9.1 Principale caractéristique de ssas :

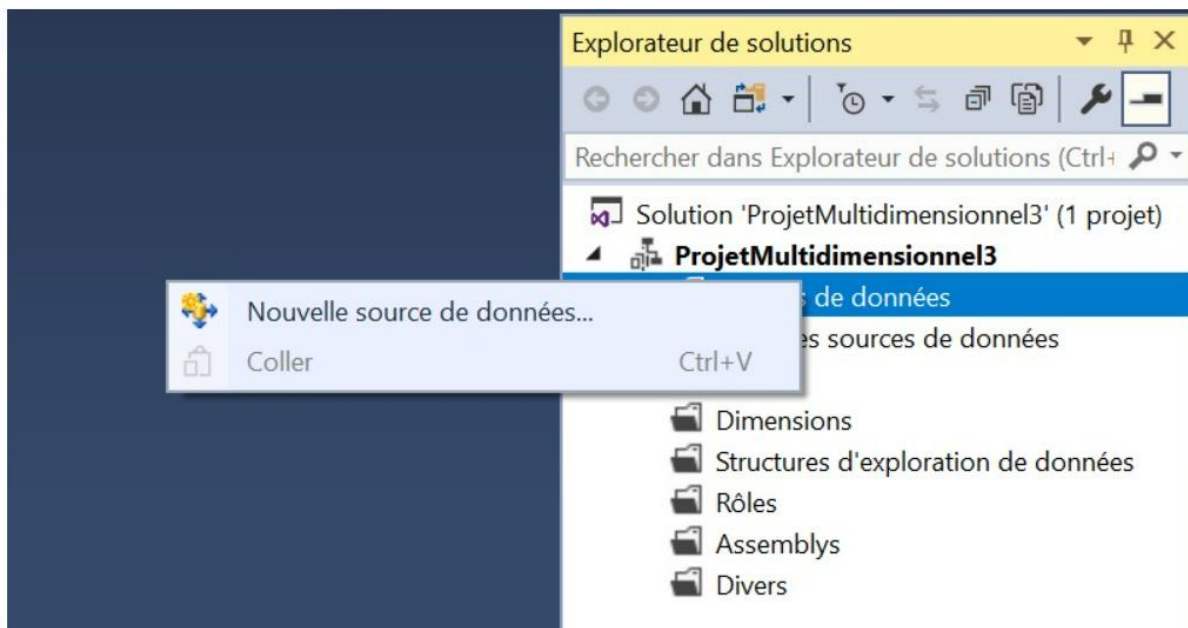
- Facilité d'utilisation avec de nombreux assistants et concepteurs

- La création et la gestion des modèles de données sont flexibles
- Gestion OLAP à partir d'une architecture évolutive
- Personnalisez l'application à partir d'un support étendu

9.2 Fonctionnalité des clés ssas :

- Rapidité : Il faut moins de temps pour répondre à une requête en raison de l'agrégation des faits
- Analyse des données : permet une analyse multidimensionnelle facilitée par les cubes
- Lien et affichage automatiques : il offre la possibilité de lier et d'afficher automatiquement le rapport
- Bon modèle de données : pour de meilleurs rapports et analyses d'entreprise, un bon modèle de données peut être créé

9.3 Développement de cube OLAP :



Data Source Wizard

Select how to define the connection
You can select from a number of ways in which your data source will define its connection string.

☐ Create a data source based on another object
☒ Create a data source based on an existing or new connection

Connexions de données :

LocalHost
LocalHost.project4

Propriétés des connexions de données :

Propriété	Valeur
Data Source	.
Initial Catalog	project4
Integrated Sec...	SSPI
Provider	SQLNCLI11.

Nouveau... Supprimer

< Précédent Suivant > Terminer >>| Annuler

Data Source Wizard

Welcome to the Data Source Wizard

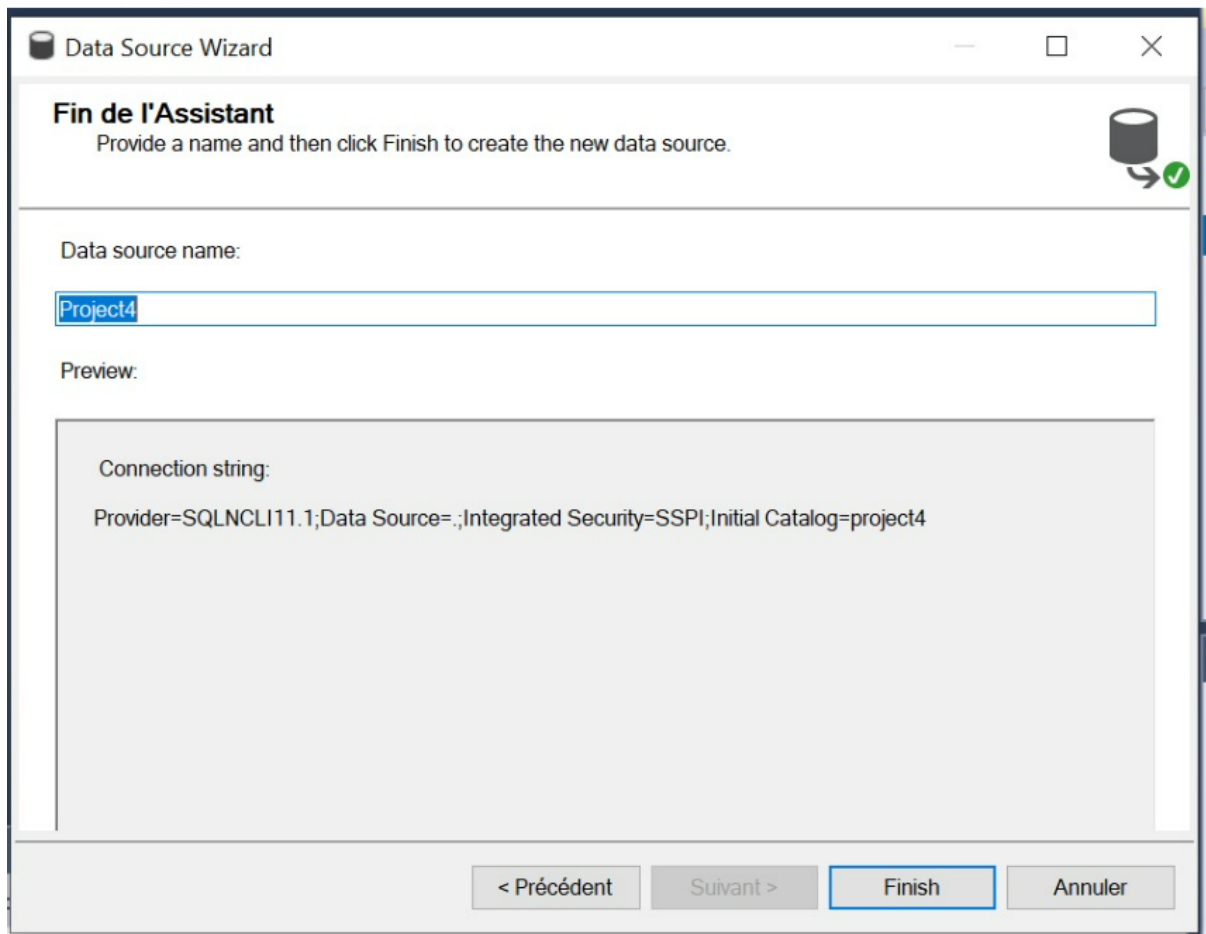
Use this wizard to create a new data source.

A data source represents a connection to your data.

A data source does not provide features such as caching metadata, adding relationships, adding calculations, and adding annotations. To apply these features to a data source, use this wizard to create the data source, and then use Data Source View Wizard to create a view that includes the appropriate features.

☐ Ne plus afficher cette page

< Précédent Suivant > Terminer >>| Annuler



Data Source Wizard

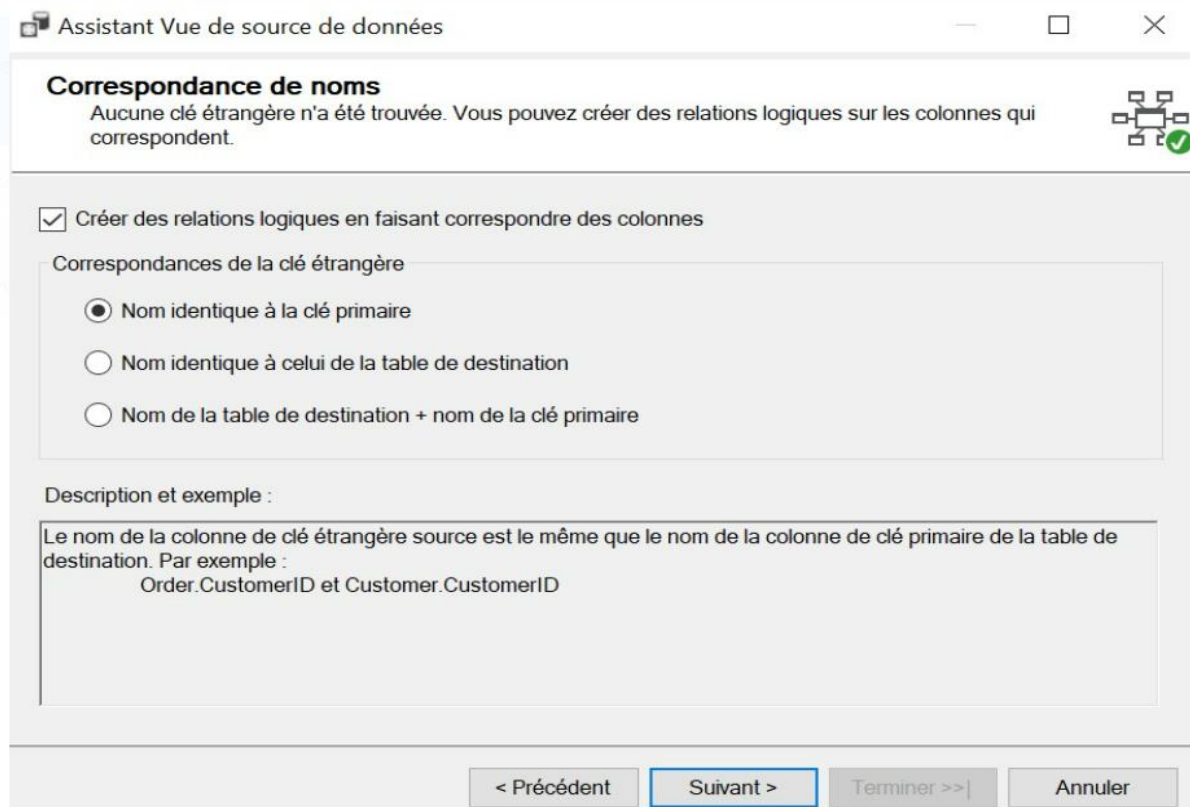
Fin de l'Assistant
Provide a name and then click Finish to create the new data source.

Data source name:

Preview:

Connection string:
Provider=SQLNCLI11.1;Data Source=.;Integrated Security=SSPI;Initial Catalog=project4

< Précédent Suivant > **Finish** Annuler



Assistant Vue de source de données

Correspondance de noms
Aucune clé étrangère n'a été trouvée. Vous pouvez créer des relations logiques sur les colonnes qui correspondent.

☒ Créer des relations logiques en faisant correspondre des colonnes

Correspondances de la clé étrangère

☒ Nom identique à la clé primaire

☐ Nom identique à celui de la table de destination

☐ Nom de la table de destination + nom de la clé primaire

Description et exemple :

Le nom de la colonne de clé étrangère source est le même que le nom de la colonne de clé primaire de la table de destination. Par exemple :
Order.CustomerID et Customer.CustomerID

< Précédent **Suivant >** Terminer >> | Annuler

Assistant Vue de source de données

Sélectionner des tables et des vues

Sélectionnez dans la base de données relationnelle les objets à inclure dans la vue de source de données.

Objets disponibles :

Nom	Type
Destination OLE DB (dbo)	Table
sysdiagrams (dbo)	Table

Filtre :

☐ Afficher les objets système

Objets inclus :

Nom	Type
Fact Sales (dbo)	Table
Dim Time (dbo)	Table
Dim Location (dbo)	Table
Dim Customer (dbo)	Table
Dim Shipment (dbo)	Table
Dim Product (dbo)	Table

Ajouter des tables associées

< Précédent Suivant > Terminer >> | Annuler

Assistant Vue de source de données

Fin de l'Assistant

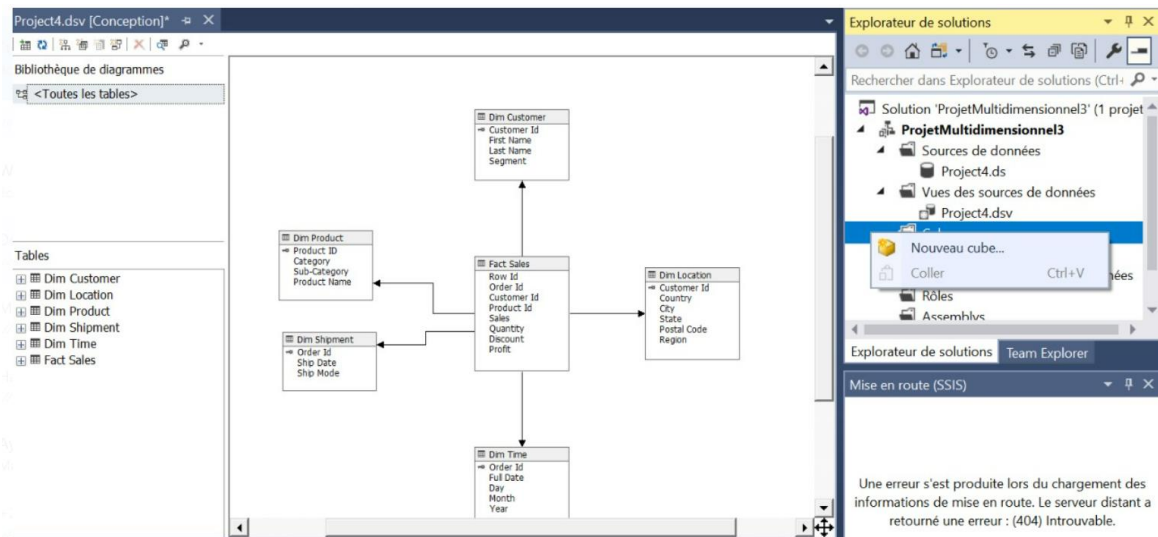
Fournissez un nom, puis cliquez sur Terminer pour créer la nouvelle vue de source de données.

Nom:

Aperçu :

- Project4
 - Fact Sales (dbo)
 - Dim Time (dbo)
 - Dim Location (dbo)
 - Dim Customer (dbo)
 - Dim Shipment (dbo)
 - Dim Product (dbo)

< Précédent Suivant > Terminer Annuler



Assistant Cube

Sélectionner la méthode de création
Les cubes peuvent être créés en utilisant des tables existantes, en créant un cube vide ou en générant des tables dans la source de données.

Comment souhaitez-vous créer le cube ?

☒ Utiliser des tables existantes
☐ Créer un cube vide
☐ Générer des tables dans la source de données

Modèle :

(Aucune)

Description :

Créez un cube en fonction d'une ou de plusieurs tables dans une source de données.

Assistant Cube

Sélectionner les tables de groupes de mesures

Sélectionnez une vue de source de données ou un schéma, puis sélectionnez les tables à utiliser pour les groupes de mesures.

Vue de source de données :
Project4

Tables de groupes de mesures : Suggérer

- ☒ Fact Sales
- ☐ Dim Time
- ☐ Dim Location
- ☐ Dim Customer
- ☐ Dim Shipment
- ☐ Dim Product

< Précédent **Suivant >** Terminer >>| Annuler

Assistant Cube

Sélectionner les mesures

Sélectionnez les mesures à inclure dans le cube.

Mesure

- ☒ Fact Sales
 - ☐ Row Id
 - ☒ Sales
 - ☒ Quantity
 - ☐ Discount
 - ☐ Profit
 - ☐ Fact Sales Nombre

< Précédent **Suivant >** Terminer >>| Annuler

Assistant Cube

Sélectionner de nouvelles dimensions

Sélectionnez les nouvelles dimensions à créer, en fonction des tables disponibles.

- ☒ Dimension
 - ☒ Dim Time
 - ☒ Dim Time
 - ☒ Dim Shipment
 - ☒ Dim Shipment
 - ☒ Dim Location
 - ☒ Dim Location
 - ☒ Dim Customer
 - ☒ Dim Customer
 - ☒ Dim Product
 - ☒ Dim Product

< Précédent **Suivant >** Terminer >> | Annuler

Assistant Cube

Fin de l'Assistant

Attribuez un nom au cube, revoyez sa structure, puis cliquez sur Terminer pour l'enregistrer.

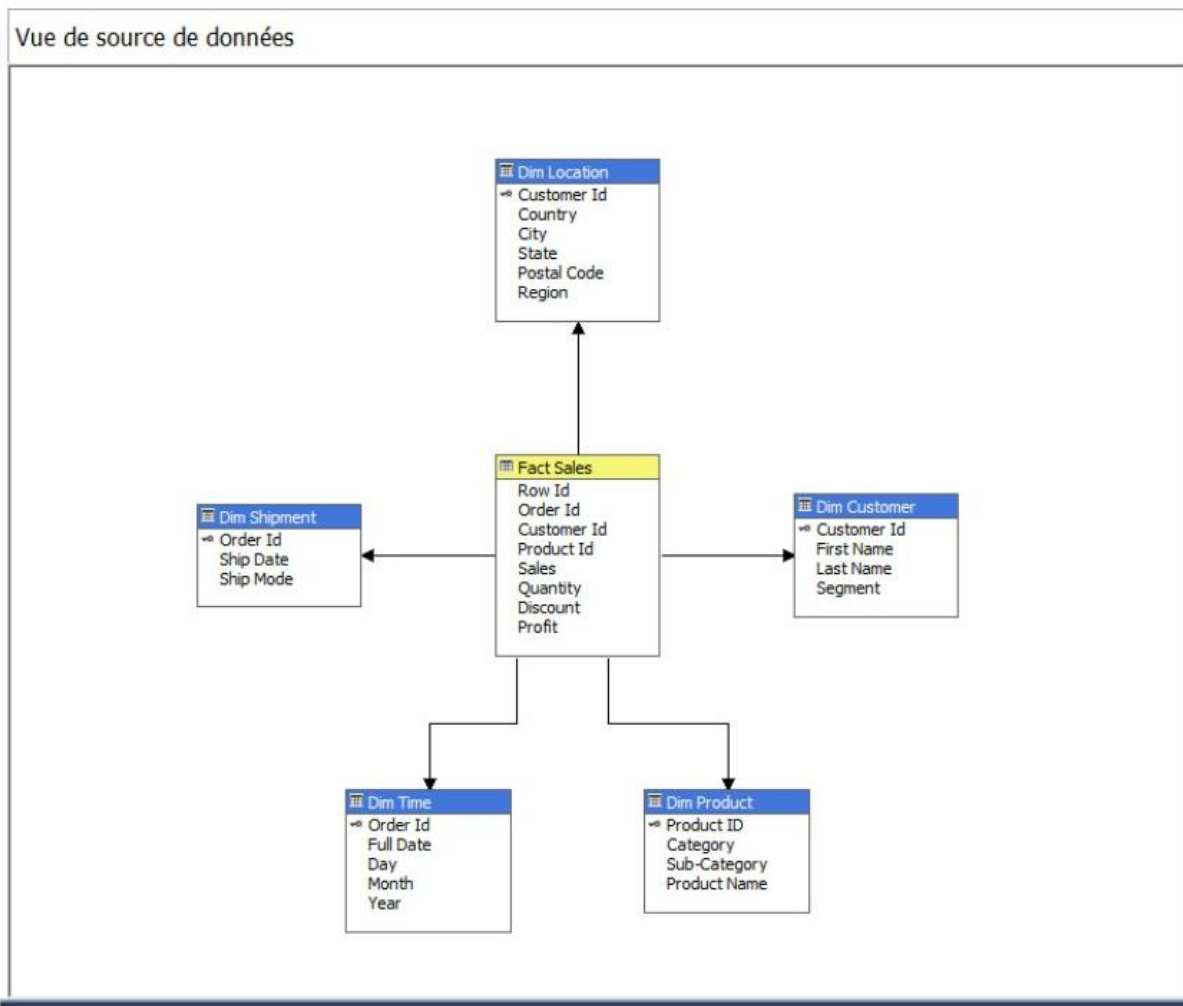
Nom du cube :

Project4

Aperçu :

- Groupes de mesures
 - Fact Sales
 - Sales
 - Quantity
- Dimensions
 - Dim Time
 - Dim Shipment
 - Dim Location
 - Dim Customer
 - Dim Product

< Précédent Suivant > **Terminer** Annuler



Dimensions

Project4

[-] [🔍] Dim Time

[🔍] [Modifier Dim Time](#)

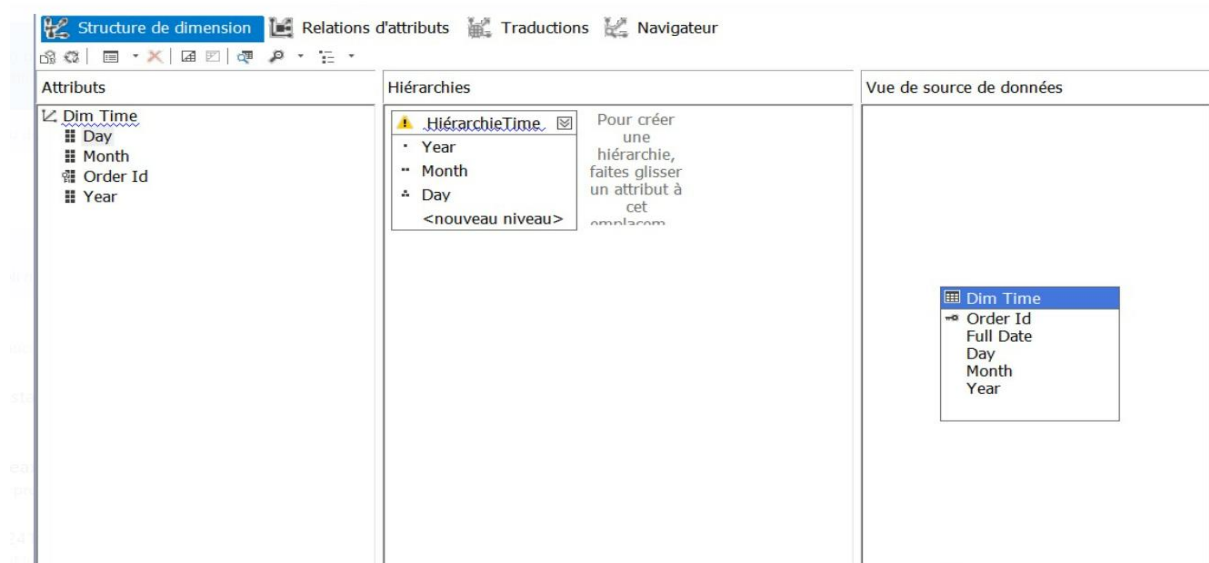
[+] [📁] Attributs

[+] [🔍] Dim Shipment

[+] [🔍] Dim Location

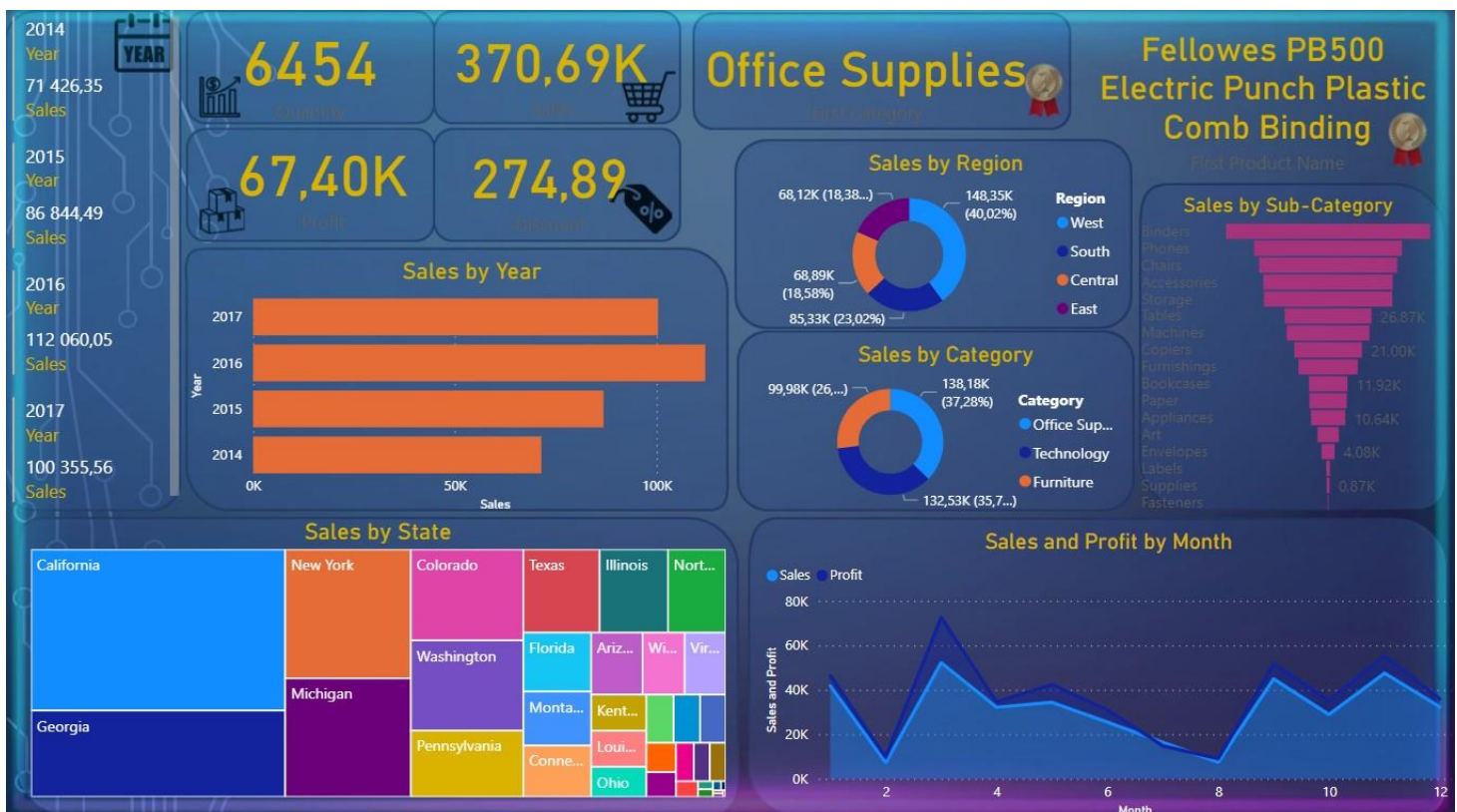
[+] [🔍] Dim Customer

[+] [🔍] Dim Product



9.4 Power BI :

Puis viens la dernière étape qui consiste à créer les tableaux et rapports à l'aide de l'outil Power BI



Conclusion Générale

La BI nous a permis de prendre des décisions plus éclairées en présentant des données actuelles et historiques dans leur contexte métier. Les analystes peuvent tirer parti de la BI pour réaliser des benchmarks sur les performances et la concurrence, dans le but d'optimiser les opérations dans l'entreprise. Ils peuvent également détecter plus facilement les tendances du marché pour augmenter les ventes ou le chiffre d'affaires. Utilisées efficacement, les données vous aideront à optimiser aussi bien votre mise en conformité que votre recrutement.

C'est la visualisation des données qui incarne le plus souvent la BI. La nature a doté les êtres humains de capacités visuelles qui les rendent particulièrement aptes à détecter des différences ou les similitudes dans les couleurs. La visualisation de données est la manière la plus accessible et la plus directe de présenter des informations.

En compilant des visualisations dans des tableaux de bord, il est souvent possible de raconter une histoire ou de mettre en évidence des tendances ou des modèles plus rapidement qu'en analysant manuellement des données brutes. Cette accessibilité favorise également la discussion autour des données, ce qui étend encore les bénéfices de la BI.

La BI évolue en permanence en fonction des besoins métier et de la technologie, et chaque année nous identifions les nouvelles tendances pour aider les utilisateurs à rester au fait des dernières innovations. Vous devez prendre conscience que l'intelligence artificielle et le machine learning vont continuer à se développer, et que les entreprises peuvent intégrer les insights de l'IA dans une stratégie BI à plus grande échelle. Le partage de données et la collaboration vont continuer à se développer à mesure que les entreprises chercheront à s'appuyer davantage sur les données. La visualisation de données jouera un rôle encore plus essentiel pour la collaboration entre équipes et services.

La BI facilite le suivi des ventes quasiment en temps réel, et permet aux utilisateurs de découvrir des insights dans le comportement des clients ou d'effectuer des prévisions pour le chiffre d'affaires. Bon nombre de secteurs différents, comme la vente au détail, les assurances ou le secteur pétrolier ont adopté la BI, et d'autres secteurs leur emboîtent le pas chaque année. Les plates-formes BI s'adaptent aux nouvelles technologies et aux innovations sur lesquelles s'appuient les utilisateurs.