

DTA

Rapport SF : Business Intelligence

Cosette Biolay

Présent : Uka Zotrim, Moreira Flavio, Dasek Joiakim,
David Guillaume, Cardoso Rafael

Table des matières

Définition de la BI	3
Reporting Opérationnel vs Dashboarding Décisionnel.....	3
Reporting Opérationnel.....	3
Dashboarding Décisionnel	3
Data Warehouse	3
Importance de la Donnée	4
Modélisation et Construction de Data Warehouse	4
Modélisation des Données	4
Modèle en Étoile	4
Analyse OLAP, Reporting, Data Mining, et Sécurité	4
OLAP Analysis.....	4
Reporting.....	4
Data Mining	5
Sécurité en BI.....	5
Aspect Juridique.....	5
Modélisation en Étoile, Flocon, et Constellation.....	5
Modélisation en Étoile.....	5
Modélisation en Flocon	5
Modélisation en Constellation	5
Power BI et SAP Univers.....	6
OLAP Cubes.....	6
Langage M.....	6
SQL	6
Clarification.....	6
Tableau de Bord: Contenu et Étapes de Construction	7
Le Contenu	7
Visualisation des Données	7
Aspect Modélisation: Processus et Modélisation DWH.....	8
Processus de Modélisation	8
Composition de la Dimension dans la Modélisation DWH	9
Principe de Jointure Interne dans un Data Warehouse.....	10
Table de Fait: Composants et Caractéristiques	11
Power BI: Visualisations, Nettoyage des Données, et Fonctionnalités Interactives.....	12
Power Query pour le Nettoyage des Données.....	12
Discussion sur les données CIMO :.....	13
Importance de la Conservation des Données Historiques	13

Utilisation de Fichiers Plats et Fabrication de Données.....	13
Optimisation des Performances avec Paramètres et Groupes de Paramètres	14
Structure des Tables de Faits et Gestion du Calendrier	14
Gestion des Paramètres et Unités de Mesure.....	14
Chargement des Données et Gestion des Champs Nuls	14
Dessin de la Table de Faits.....	14
Structure des Différentes Tables.....	15
Relations et Jointures	15
Vérification des Flags.....	15
Table de Conversion.....	16
Conclusion	17

Définition de la BI

La BI est décrite comme étant l'informatique mise au service des décideurs et des dirigeants d'entreprises. Ce domaine englobe les technologies, les applications et les pratiques pour la collecte, l'intégration, l'analyse et la présentation d'informations d'affaires. Son objectif est de soutenir la prise de décision. Elle repose sur des données collectées afin de les consolider, les modéliser et enfin les restituer sous forme de rapports ou de tableaux de bord, offrant ainsi une base solide pour les décisions stratégiques.

Reporting Opérationnel vs Dashboarding Décisionnel

Reporting Opérationnel

Le reporting opérationnel se concentre sur la présentation détaillée des activités quotidiennes de l'entreprise. Il s'agit d'un processus essentiel de la BI qui permet aux managers et aux équipes opérationnelles de surveiller la performance en temps réel, facilitant ainsi la prise de décisions immédiates basées sur des données actualisées. Les rapports opérationnels peuvent inclure des informations telles que les ventes quotidiennes, le niveau des stocks, ou le statut des commandes en cours.

Dashboarding Décisionnel

Le dashboarding décisionnel, en revanche, sert à fournir une vue d'ensemble des performances de l'entreprise à travers des indicateurs clés de performance (KPI) et des résumés de données stratégiques. Ces tableaux de bord sont conçus pour aider les dirigeants et les décideurs à comprendre les tendances, évaluer la réalisation des objectifs et orienter les stratégies à long terme. Ils synthétisent des données complexes en visualisations intuitives, offrant une perspective consolidée sur la santé globale de l'entreprise.

Data Warehouse

Dans l'écosystème de la Business Intelligence (BI), le **Data Warehouse** et le **Data Mart** constituent des fondations essentielles pour le stockage, l'organisation et l'analyse des données. Bien que souvent utilisés de manière interchangeable, ils servent des rôles distincts mais complémentaires dans la gestion des données d'entreprise.

Data Warehouse: Le Répertoire Centralisé de Données

Un **Data Warehouse** est une vaste base de données conçue pour stocker une grande quantité de données historiques provenant de diverses sources au sein de l'organisation. Son objectif principal est de consolider ces données dans un format uniforme pour faciliter l'accès et l'analyse à travers toute l'entreprise. Les caractéristiques d'un Data Warehouse incluent:

- **Intégration de Données:** Fusionne les données de multiples sources pour créer une vue unifiée.
- **Non Volatilité:** Les données sont seulement ajoutées au Data Warehouse; elles ne sont pas modifiées ou supprimées, conservant ainsi un historique complet.

- **Orientation Sujet:** Organisé autour de thèmes clés de l'entreprise (comme les ventes, la finance) plutôt que des opérations quotidiennes.

Importance de la Donnée

Il est primordial de s'assurer de l'existence de la donnée. Le Master Data Management (MDM) joue ici un rôle clé en uniformisant les données à travers l'entreprise, assurant ainsi leur qualité et leur cohérence pour tous les utilisateurs.

Modélisation et Construction de Data Warehouse

Le processus de construction d'un Data Warehouse inclut la définition du sujet, la détermination de la granularité des données, et l'application de la dénormalisation pour améliorer la performance des requêtes. Les tables de fait centralisent les indicateurs clés, tandis que les dimensions fournissent le contexte nécessaire. On va volontairement répéter les données pour éviter les jointures

Modélisation des Données

Modèle en Étoile

Le principe du modèle en étoile est de dénormaliser les données, c'est-à-dire de regrouper les informations pour réduire le besoin de jointures complexes. Cette technique implique une répétition volontaire des données pour améliorer les performances des requêtes.

Processus ETL (Extraction, Transformation, Loading)

Les outils ETL constituent une part significative du travail en BI, représentant environ 80% du temps consacré à extraire les données des bases, les transformer pour garantir leur cohérence, et les charger dans le Data Warehouse.

Analyse OLAP, Reporting, Data Mining, et Sécurité

OLAP Analysis

L'analyse OLAP (Online Analytical Processing) permet aux utilisateurs de manipuler et d'analyser de grandes quantités de données depuis différentes perspectives rapidement et dynamiquement. Cela inclut des opérations telles que le slicing (découpe), dicing (dé découpage), drilling (exploration), et pivoting (pivotement).

Reporting

Le reporting dans le contexte de la BI implique la génération de rapports basés sur les données collectées et traitées, fournissant des insights et des analyses pour la prise de décision. Les rapports peuvent être opérationnels ou décisionnels, en fonction de leur utilisation.

Data Mining

Le data mining est le processus d'exploration et d'analyse de grandes quantités de données pour découvrir des modèles, des tendances ou des relations significatives. Cela peut aider à prédire des tendances futures et à prendre des décisions éclairées basées sur les données.

Sécurité en BI

Dans la BI, la sécurité est primordiale pour protéger les données sensibles. Les mesures de sécurité peuvent inclure des restrictions d'accès aux données au niveau des colonnes ou des lignes, assurant que seules les parties autorisées peuvent accéder aux informations spécifiques.

Aspect Juridique

Il est à noter que les fournisseurs ne peuvent légalement refuser l'exportation des données, soulignant l'importance d'un accès libre aux données pour les opérations de BI.

Modélisation en Étoile, Flocon, et Constellation

Modélisation en Étoile

La modélisation en étoile centre autour d'une table de faits, entourée de tables de dimensions. Chaque table de dimension est connectée à la table de faits par une clé primaire. Cette structure facilite les requêtes analytiques en réduisant le nombre de jointures nécessaires.

Modélisation en Flocon

La modélisation en flocon est une variation de la modélisation en étoile où les tables de dimensions sont normalisées, divisant les données en tables supplémentaires. Cela peut réduire la redondance des données mais augmente le nombre de jointures nécessaires.

Modélisation en Constellation

Une modélisation en constellation comporte plusieurs tables de faits partageant des tables de dimensions. C'est une structure complexe utilisée pour supporter des analyses multi-facettes.

Des tables et des jointures

Au milieu la table des faits et autour les tables de dimensions et il y a toujours une dimension « date »

la table de fait doit toujours rester en CHIFFRE

on va enrichir les données dans le data warehouse, mais jamais réécrire dans le système source.

Power BI et SAP Univers

- **Power BI**: Utilise des DataSet et permet la création de visualisations variées à l'aide de Power Query et d'autres outils d'analytique.
- **SAP Univers**: Propose une abstraction du SQL pour simplifier l'accès et la manipulation des données.

OLAP Cubes

- **Cubes OLAP** (Online Analytical Processing) sont utilisés pour stocker des données de manière multi-dimensionnelle pour supporter des analyses complexes et des requêtes avec une performance élevée. Les cubes OLAP permettent aux utilisateurs de visualiser et de manipuler des données à travers plusieurs dimensions de manière interactive.

Langage M

- **Le langage M** est principalement associé à Power Query, un outil de collecte, de transformation et d'enrichissement des données dans les environnements Microsoft tels que Excel et Power BI. Le langage M est utilisé pour définir des transformations de données de manière déclarative. Bien qu'il puisse être utilisé pour préparer des données pour les analyses OLAP, il n'est pas spécifiquement un langage pour interroger des cubes OLAP.

SQL

- **SQL** (Structured Query Language) est le langage standard pour interroger et manipuler des bases de données relationnelles. Pour les cubes OLAP, une variante du SQL, appelée MDX (Multidimensional Expressions) ou DAX (Data Analysis Expressions pour Power BI), est souvent utilisée. MDX est spécifiquement conçu pour interroger des cubes OLAP et manipuler des données multi-dimensionnelles. DAX est utilisé dans les modèles de données de Power BI et SQL Server Analysis Services pour créer des calculs et des analyses.

Clarification

- Lorsqu'on travaille avec **des cubes OLAP**, le **MDX** ou **DAX** (dans le cas de Power BI et certaines versions de SQL Server Analysis Services) serait le langage de choix pour interroger des cubes et effectuer des analyses multidimensionnelles.
- **Le langage M** est utilisé dans le processus de préparation et de transformation des données avant qu'elles ne soient chargées dans un cube OLAP ou un modèle de données pour l'analyse.
- **SQL** est utilisé pour interroger des bases de données relationnelles mais pas directement les cubes OLAP, pour lesquels MDX ou DAX sont préférés.

Tableau de Bord: Contenu et Étapes de Construction

Le Contenu

Pour créer un tableau de bord efficace, il est crucial d'identifier les objectifs de l'entreprise ou du projet spécifique et de choisir les indicateurs clés de performance (KPI) en conséquence. Ces indicateurs doivent être directement liés à un ou plusieurs inducteurs de performance pour fournir une mesure précise de l'efficacité des différentes stratégies ou opérations. Par exemple, si l'objectif est d'augmenter le revenu, un indicateur pertinent pourrait être le revenu commercial généré à partir de Google Analytics (GA), qui permettrait de mesurer l'impact des campagnes marketing en ligne.

Étapes de Construction

1. Le Choix de la Solution Technique:

- **Sélectionnez la Technologie:** Avant de commencer, il est important de choisir la plateforme ou la technologie qui sera utilisée pour développer et héberger le tableau de bord. Cette sélection se fait généralement une seule fois pour la société, en prenant en compte des facteurs comme l'évolutivité, la facilité d'utilisation, et l'intégration avec les systèmes existants.
- **Identifiez les Sources de Données:** La prochaine étape consiste à déterminer d'où viendront les données. Par exemple, pour un KPI basé sur le revenu commercial, il pourrait être pertinent d'utiliser des données de Google Analytics. D'autres sources de données peuvent inclure des systèmes ERP, des bases de données internes, des feuilles de calcul, etc.
- **Définissez Comment Livrer le Tableau de Bord aux Destinataires:** Il est également important de planifier comment le tableau de bord sera partagé avec les utilisateurs finaux. Cela peut impliquer de le rendre accessible via un portail en ligne, de le distribuer par email, ou même de prévoir des sessions de formation pour aider les utilisateurs à comprendre et à interagir efficacement avec le tableau de bord.

Chaque étape est essentielle pour s'assurer que le tableau de bord final est non seulement techniquement solide mais aussi aligné sur les objectifs de l'entreprise et facilement accessible pour ceux qui en ont besoin.

Visualisation des Données

Les visualisations jouent un rôle crucial dans la restitution des données. Voici quelques principes spécifiques pour différents types de graphiques:

- **Diagramme en camembert:** Éviter plus de 5 tranches, pas de légende ni d'effet 3D/incliné.
- **Jauge:** Idéale pour représenter un pourcentage ou un objectif.
- **Graphique linéaire:** Parfait pour montrer l'évolution dans le temps.
- **Diagramme à barres:** Garder l'échelle à partir de 0 et nommer les axes pour la clarté.
- **Nuages de mots et diagrammes à bulles:** À utiliser avec discernement pour leur impact visuel mais reconnaître leurs limites en termes de précision.

Aspect Modélisation: Processus et Modélisation DWH

Processus de Modélisation

La modélisation de Data Warehouse (DWH) est cruciale pour structurer les données de manière à faciliter l'analyse et le reporting. Le processus de modélisation implique plusieurs étapes clés, décrites ci-dessous :

1. Choisir le Sujet :

- La première étape consiste à déterminer le domaine ou le sujet à analyser. Cela pourrait être lié aux ventes, à la finance, aux opérations, ou à tout autre aspect important de l'entreprise.

2. Granularité :

- La granularité fait référence au niveau de détail ou de finesse des données dans le DWH. Cela détermine la précision des analyses que vous pourrez effectuer. La granularité est choisie en fonction des besoins analytiques ; elle peut varier de très détaillée (comme les transactions individuelles) à plus agrégée (comme les résumés mensuels ou annuels).

3. Dénormalisation :

- La dénormalisation implique le regroupement de données provenant de différentes tables en une seule, réduisant ainsi le nombre de jointures nécessaires lors des requêtes. Cela est contraire à la normalisation en bases de données relationnelles, où l'accent est mis sur la réduction de la redondance des données. La dénormalisation est utilisée dans les DWH pour accélérer les requêtes analytiques en sacrifiant l'espace de stockage pour une meilleure performance de lecture.

4. Table de Fait :

- Les tables de faits sont au cœur de la modélisation en étoile d'un DWH. Elles rassemblent les indicateurs (mesures) du même processus métier ayant une granularité commune. Ces indicateurs sont souvent numériques (comme les quantités vendues, les revenus, les coûts) et peuvent être analysés à travers différentes dimensions (comme le temps, la géographie, les produits).

Modélisation en Étoile

- La modélisation en étoile est une approche populaire pour structurer un DWH. Elle se caractérise par une table de faits centrale, contenant les mesures clés, entourée de tables de dimensions. Les tables de dimensions contiennent les attributs descriptifs relatifs aux mesures dans la table de faits (comme les détails du produit, les informations client, les données temporelles).
- Cette structure simplifie les requêtes et améliore les performances en limitant le nombre de jointures nécessaires pour les analyses. Chaque table de dimension est liée à la table de faits par une clé étrangère, ce qui facilite l'exploration multidimensionnelle des données.

La modélisation DWH est une étape fondamentale pour établir une base solide pour la Business Intelligence et l'analyse de données dans une entreprise. La réussite de ce

processus dépend de la compréhension approfondie des besoins de l'entreprise, de la qualité des données disponibles et de la sélection judicieuse de la technologie et des outils de modélisation.

Composition de la Dimension dans la Modélisation DWH

La structuration des dimensions dans un Data Warehouse est cruciale pour une analyse efficace. Voici les composants clés d'une dimension bien conçue et des recommandations pour leur mise en œuvre :

Clé Datawarehouse (DW Key)

- **Définition** : Un identifiant unique attribué à chaque élément dans la dimension, utilisé pour relier les tables de dimensions à la table de faits.
- **Importance** : Assure l'intégrité référentielle et facilite les jointures rapides avec la table de faits.

Clé Métier (Business Key)

- **Définition** : La clé naturelle qui identifie de manière unique un élément dans le système source d'origine.
- **Importance** : Cruciale pour l'intégration des données, permettant de relier les données du DW aux systèmes sources.

Attributs

- **Définition** : Les caractéristiques descriptives de chaque élément de la dimension, telles que le nom, la description, ou d'autres propriétés spécifiques.
- **Importance** : Fournissent le contexte nécessaire pour les analyses, permettant de filtrer, grouper ou segmenter les données.

Date de Début de Validité et Date de Fin de Validité

- **Définition** : Marquent la période de temps pendant laquelle un enregistrement dans la dimension est considéré comme valide.
- **Importance** : Permettent de gérer les dimensions changeantes dans le temps (SCD - Slowly Changing Dimensions), facilitant l'analyse des tendances et des évolutions.

Flag Current Record

- **Définition** : Un indicateur (souvent un booléen) qui identifie si l'enregistrement est l'instance actuelle et valide de l'élément de la dimension.
- **Importance** : Aide à filtrer rapidement les données actuelles sans avoir à interroger les dates de validité.

Recommandations pour la Composition de la Dimension

- **Le Plus Complet Possible** : Incluez tous les attributs pertinents pour assurer une analyse riche et multidimensionnelle.
- **Nomenclature des Champs** : Adoptez des conventions de nommage claires et cohérentes pour faciliter la compréhension et l'utilisation des données.

- **Granularité** : Déterminez le niveau de détail nécessaire pour vos analyses. Par exemple, pour une dimension temporelle (**dimDate**), décidez si vous avez besoin d'attributs au niveau des heures ou des minutes, ou si des agrégations quotidiennes sont suffisantes.

Exemple Pratique: Dimension Temporelle (**dimDate**)

- **Séparation des Dates et des Heures** : Pour maintenir la clarté et l'efficacité des analyses, il est recommandé de ne pas mélanger les dates et les heures dans la même dimension. Envisagez d'utiliser des dimensions séparées pour le temps si votre analyse nécessite une granularité horaire ou minutée.

Principe de Jointure Interne dans un Data Warehouse

Définition et Utilisation

Dans le contexte d'un Data Warehouse (DWH), les jointures internes sont préférées pour lier les tables de dimensions aux tables de faits. Une jointure interne combine les enregistrements de deux tables quand il y a au moins une correspondance dans les colonnes de jointure. Ce type de jointure est essentiel pour garantir que seules les données correspondant à la fois dans la table de faits et dans les tables de dimensions soient incluses dans les résultats de la requête.

Pourquoi Éviter les Jointures Externes

- **Intégrité des Données** : L'utilisation de jointures externes dans un DWH peut introduire des lignes pour lesquelles il n'y a pas de correspondance directe entre les tables de faits et de dimensions. Cela peut conduire à des résultats analytiques trompeurs ou à l'interprétation incorrecte des données.
- **Performance** : Les jointures internes sont généralement plus performantes que les jointures externes, surtout dans un environnement de grande envergure comme un DWH où le volume de données est conséquent. Les jointures internes permettent des requêtes plus rapides et plus efficaces.
- **Simplicité et Clarté** : Utiliser des jointures internes maintient la structure des requêtes simple et claire. Cela facilite la maintenance et l'optimisation des requêtes dans un DWH.

Bonnes Pratiques

- **Conception de Modèle** : Assurez-vous que votre modèle de données est conçu de manière à éviter la nécessité de jointures externes. Cela peut impliquer une planification soignée des tables de dimensions pour s'assurer qu'elles reflètent tous les scénarios possibles présents dans les tables de faits.
- **Gestion des Valeurs Manquantes** : Dans les cas où des données pourraient être manquantes (par exemple, un produit non classé), prévoyez des enregistrements par défaut dans vos tables de dimensions (comme un produit intitulé "Non Classé") pour garantir que toutes les lignes de la table de faits aient des correspondances dans les dimensions.
- **Validation des Données** : Effectuez régulièrement des contrôles de qualité des données pour identifier et corriger les anomalies où les enregistrements de la table de

faits n'ont pas de correspondances dans les tables de dimensions, ce qui pourrait autrement suggérer l'utilisation de jointures externes pour les inclure.

Table de Fait: Composants et Caractéristiques

Dans un Data Warehouse (DWH), les tables de fait sont au cœur du modèle dimensionnel, servant de point central pour l'analyse et le reporting. Elles contiennent des mesures numériques qui sont les sujets d'analyse, ainsi que des clés étrangères qui les lient aux tables de dimensions. Voici une explication détaillée de leurs composants clés :

Clés Étrangères

- **Définition** : Les clés étrangères dans une table de fait sont des références aux clés primaires des tables de dimensions. Elles permettent de relier chaque enregistrement de la table de fait à des données descriptives spécifiques dans les tables de dimensions (par exemple, date, client, produit).
- **Rôle** : Ces clés jouent un rôle crucial dans l'organisation du modèle de données en étoile ou en flocon. Elles garantissent que les données de mesure peuvent être analysées selon divers axes ou dimensions (temporelle, géographique, par produit, etc.).

Mesures

- **Définition** : Les mesures sont les données quantitatives stockées dans les tables de fait. Elles représentent typiquement les performances ou les résultats d'un processus métier (comme les ventes, les coûts, les quantités).
- **Types** : Les mesures peuvent être des montants bruts, des totaux cumulés, des moyennes, des comptages ou toute autre forme de données numériques qui supportent des calculs et des analyses.

Les Faits Additifs

- **Définition** : Les faits additifs sont des mesures qui peuvent être agrégées (additionnées) à travers toutes les dimensions associées sans perdre leur signification. Ces faits sont fondamentaux pour la plupart des types d'analyse dans un DWH.
- **Exemples** : Les ventes totales, les coûts totaux, et le nombre total d'unités vendues sont des exemples de faits additifs. Ils peuvent être additionnés à travers différentes dimensions, comme le temps (jour, mois, année), la géographie (ville, région, pays), ou les catégories de produits.
- **Importance** : Les faits additifs sont essentiels pour les analyses d'affaires car ils permettent de calculer des totaux, des moyennes, et d'autres agrégats utiles pour le reporting, la planification et l'aide à la décision.

Power BI: Visualisations, Nettoyage des Données, et Fonctionnalités Interactives

Power BI est un outil puissant de Microsoft destiné à la business intelligence qui permet de créer des visualisations de données dynamiques, de nettoyer et transformer des données, et d'intégrer des fonctionnalités interactives telles que les tooltips et les bookmarks. Voici un aperçu de ces capacités :

Visualisations

Power BI offre une large gamme de types de graphiques et de visualisations pour répondre à différents besoins analytiques :

- **Graphiques à barres et colonnes** pour la comparaison de différentes catégories.
- **Graphiques linéaires et à aire** pour afficher les tendances au fil du temps.
- **Graphiques circulaires et en anneau** pour montrer les proportions.
- **Cartes géographiques** pour la visualisation des données par emplacement.
- **Nuages de points et diagrammes à bulles** pour explorer les relations entre les variables.
- **Graphiques en cascade** pour visualiser les décompositions des totaux.
- **Tableaux et matrices** pour afficher des données en format tabulaire avec ou sans hiérarchies.

Power Query pour le Nettoyage des Données

Power Query est l'outil intégré dans Power BI pour la préparation des données. Il permet :

- **Importation des données** de diverses sources.
- **Nettoyage des données** en supprimant les valeurs manquantes ou en corrigeant les erreurs.
- **Transformation des données** par le biais de la pivotement, du groupement, ou de la création de nouvelles colonnes calculées.
- **Fusion et jointure** des tables pour consolider les données de multiples sources.

Contenu, Format, et Analytique

1. **Contenu** : Sélectionnez et organisez les données pertinentes pour vos visualisations. Utilisez des filtres pour affiner le contenu affiché.
2. **Format** : Personnalisez l'apparence de vos visualisations — couleurs, axes, légendes, titres, etc. Utilisez le panneau de format pour modifier les propriétés visuelles et améliorer la lisibilité.
3. **Analytique** : Appliquez des lignes de tendance, des moyennes mobiles, et d'autres outils analytiques directement dans vos visualisations pour une analyse plus profonde.

Tooltip et Bookmark

- **Tooltip (Infobulle)** : Créez des tooltips personnalisés pour fournir des informations supplémentaires ou des contextes spécifiques lorsque l'utilisateur survole une partie de la visualisation.
- **Bookmark (Marque-page)** : Utilisez les bookmarks pour capturer des vues spécifiques ou des états configurés de vos rapports. Vous pouvez :
 - **Créer deux éléments visuels différents** (par exemple, un graphique et un tableau) qui affichent des aspects complémentaires de vos données.
 - **Insérer un bouton bookmark** pour permettre aux utilisateurs de basculer entre ces vues. Cela rend le rapport interactif et engageant, offrant aux utilisateurs la flexibilité d'explorer les données selon leurs préférences.

Ces fonctionnalités de Power BI permettent de créer des rapports et des tableaux de bord interactifs et informatifs, rendant l'analyse de données à la fois plus profonde et plus accessible.

Discussion sur les données CIMO :

Importance de la Conservation des Données Historiques

Transformation des données : Il est crucial de conserver l'historique des anciennes valeurs lors de la transformation des données. Cela permet de toujours pouvoir vérifier les valeurs originales en cas de besoin. La conservation des anciennes valeurs de mesure assure une traçabilité et une vérification des modifications effectuées. En cas de changements futurs, cette approche permet de retrouver les valeurs originales et de comprendre l'évolution des données.

Utilisation de Fichiers Plats et Fabrication de Données

Fichiers plats et Operational Data Store (ODS) : Lors du chargement des données, il est recommandé de privilégier un chargement 1:1 dans un ODS, sans transformation initiale. Cela implique de partir de ces fichiers plats pour transformer les données, plutôt que de se baser directement sur des fichiers Excel. Cette méthode permet de passer d'une table à une autre sans nécessiter de dénormalisation, simplifiant ainsi les opérations de transformation.

Préférence pour les tables : Comme mentionné par Cosette, il est préférable de travailler avec des tables plutôt qu'avec des fichiers. Cette pratique améliore l'organisation des données et facilite leur manipulation.

Optimisation des Performances avec Paramètres et Groupes de Paramètres

Paramètres et Groupe de Paramètres : L'objectif est de minimiser le nombre de jointures pour améliorer les performances. En réduisant le nombre de tables nécessaires aux opérations, on évite des jointures complexes et coûteuses en termes de ressources.

Structure des Tables de Faits et Gestion du Calendrier

Tables de faits : Les tables de faits contiendront à la fois des valeurs numériques et des valeurs texte. Il est également essentiel d'inclure une table Calendrier pour permettre l'agrégation des données par mois, année, etc. PowerBI offre une fonctionnalité AUTO-CALENDAR qui facilite cette étape directement dans l'outil.

Gestion des Paramètres et Unités de Mesure

Paramètres et localisation : Il est crucial de reprendre les lieux de mesure avec leur description correspondante. Pour les unités de mesure, il est recommandé de créer une table de conversion. Les transformations seront gérées par unité et non par paramètre, permettant ainsi de traiter des mesures identiques avec des unités différentes.

Colonnes d'unités : Pour gérer les changements d'unités dans le temps, il est proposé d'avoir deux colonnes : une colonne contenant la valeur historique qui ne change jamais, et une colonne avec la valeur actuelle qui peut être mise à jour. Une table de conversion des unités sera utilisée pour centraliser toutes les unités.

Chargement des Données et Gestion des Champs Nuls

Valeurs par défaut : Lors du chargement des données, il est recommandé de mettre des valeurs par défaut pour éviter les champs nuls, ce qui simplifie les opérations de filtrage et d'analyse ultérieures.

Gestion des nouvelles feuilles Excel : À chaque réception d'une nouvelle feuille Excel, il est conseillé de l'ajouter directement à la suite de la table ODS. Bien que le processus d'agrégation manuelle des six feuilles initiales soit fastidieux, il est préférable de le faire une seule fois pour se concentrer ensuite sur la visualisation des données, ce qui est l'objectif principal du client.

Dessin de la Table de Faits

Stockage des dates : Lorsque la gestion des dates est simple, il est possible de stocker les dates directement dans la table de faits. Cependant, pour des cas où le volume de données est immense, la création de tables calendrier distinctes peut être plus appropriée. PowerBI facilite cette gestion avec la fonctionnalité AUTO-CALENDAR, permettant de créer une table calendrier qui se joint à la table de faits avec une relation 1: *.

Structure des Différentes Tables

Table Date (avec AUTO-CALENDAR) :

- Date
- Mois
- Année
- N° semaine

Table Lieux :

- Lieu de mesure
- Description
- ID

Table Paramètres :

- Nom
- ID
- Unité de base
- Unité actuelle
- Groupe de paramètre

Table de Faits : Mesure :

- Date de prélèvement
- IDLieu
- IDParamètre
- ValeurNumériqueBase
- ValeurNumériqueActuel
- ValeurTexte

Relations et Jointures

Jointures de Date : Dans ce schéma, la table Date se joint à la table de faits via le champ Date de prélèvement. Cela permet de centraliser les informations temporelles et de faciliter les agrégations et analyses par période (jour, mois, année, semaine).

Vérification des Flags

Impact des flags sur la performance : La vérification des flags lors du chargement ou de la transformation des données peut entraîner des ralentissements significatifs. Pour éviter ces ralentissements, il est recommandé de réaliser ces vérifications dans PowerQuery. Cette

méthode permet de décharger les opérations lourdes de l'entrepôt de données, améliorant ainsi la performance globale du système.

Table de Conversion

Création et gestion de la Table de Conversion : Il est conseillé de créer une table de conversion en dehors du DataWarehouse, par exemple dans un Data Lakehouse. Cette table de conversion peut être modifiée et maintenue à partir d'un fichier Excel, offrant une flexibilité et une facilité de gestion. Pour une architecture propre et organisée, il est recommandé de placer ce fichier Excel dans un SharePoint. Cela permet une gestion centralisée et sécurisée des données de conversion.

En suivant ces recommandations, la gestion et la transformation des données seront optimisées, permettant une meilleure performance et une plus grande précision dans l'analyse des données. La conservation des valeurs historiques, l'utilisation de tables, la gestion efficace des unités de mesure et l'optimisation des processus de chargement des données sont des éléments clés pour réussir un projet de business intelligence.

Conclusion

En conclusion, la Business Intelligence (BI) joue un rôle crucial dans le paysage entrepreneurial moderne, en offrant des outils et des méthodologies pour améliorer la prise de décision à tous les niveaux de l'entreprise. Que ce soit à travers le reporting opérationnel, qui met l'accent sur le suivi détaillé et en temps réel des activités quotidiennes, ou via le dashboarding décisionnel, qui vise à fournir une vision globale et stratégique des performances, la BI transforme les données brutes en insights précieux. Ces insights permettent aux entreprises de réagir rapidement aux changements, d'optimiser leurs opérations et de planifier avec plus de précision leur stratégie future. Ainsi, la BI se révèle être un pilier indispensable pour soutenir la croissance et le développement durable des organisations dans un environnement économique de plus en plus compétitif et dynamique.