



Université Mohamed Boudiaf de M'sila
Faculté des Mathématique et de l'Informatique
Département d'Informatique

Master : Informatique Décisionnelle et Optimisation

Technologie du décisionnel

Monsieur Akhrouf Samir
samir.akrouf@gmail.com

Intitulé du Master : Informatique Décisionnelle et Optimisation
Semestre :S3

Intitulé de l'UE : UEM 3.1

Intitulé de la matière : Technologie du décisionnel

Crédits : 3

Coefficients : 1

Objectifs de l'enseignement

Développer les compétences de mise en œuvre d'outils permettant la gestion homogène et cohérente des systèmes d'information (SI) des entreprises, en particulier pour la gestion commerciale de la chaîne de production à la vente d'un produit.

Contenu de la matière :

- Ingénierie informatique de la décision
- Panorama des composants logiciels d'un système d'information décisionnel (CRM, Customer Relationship Marketing), ERP (Entreprise Ressource Planning), BI (Business Intelligence).
- Extraction transformation et chargement de données: Techniques et Outils ETL
- Les outils d'analyses multidimensionnels
- Outils de reporting et d'exploration (exemple, Business Object)

Mode d'évaluation :

- Examens de moyenne durée, TP, Mini-projet

Références :

- <http://fablain.developpez.com/tutoriel/presenterp/>
- <http://www.commentcamarche.net/contents/entreprise/erp.php3>
- <http://www.erp-infos.com/article/m/268/erp-ou-plateforme-decisionnelle--qui-pilote-.html>

Une brève histoire de la business intelligence

- Business intelligence,
- intelligence d'affaires,
- ou encore informatique décisionnelle,

Théorisée en 1958 par **Hans Peter Luhn**, un informaticien allemand, cette notion regroupe les solutions IT qui permettent d'exploiter les données d'une entreprise (comptabilité, finances, base clients...) afin d'établir des rapports d'analyse servant à la gouvernance et à la prise de décision.

Un outil devenu incontournable aujourd'hui pour diriger et orienter toutes les business unit d'une entreprise (organisation, ventes, investissements...).

Une solution d'aide à la prise de décision

- Base de données clients,
- tableau Excel des ventes,
- sources web,
- textes,
- informations...

Data (données) !

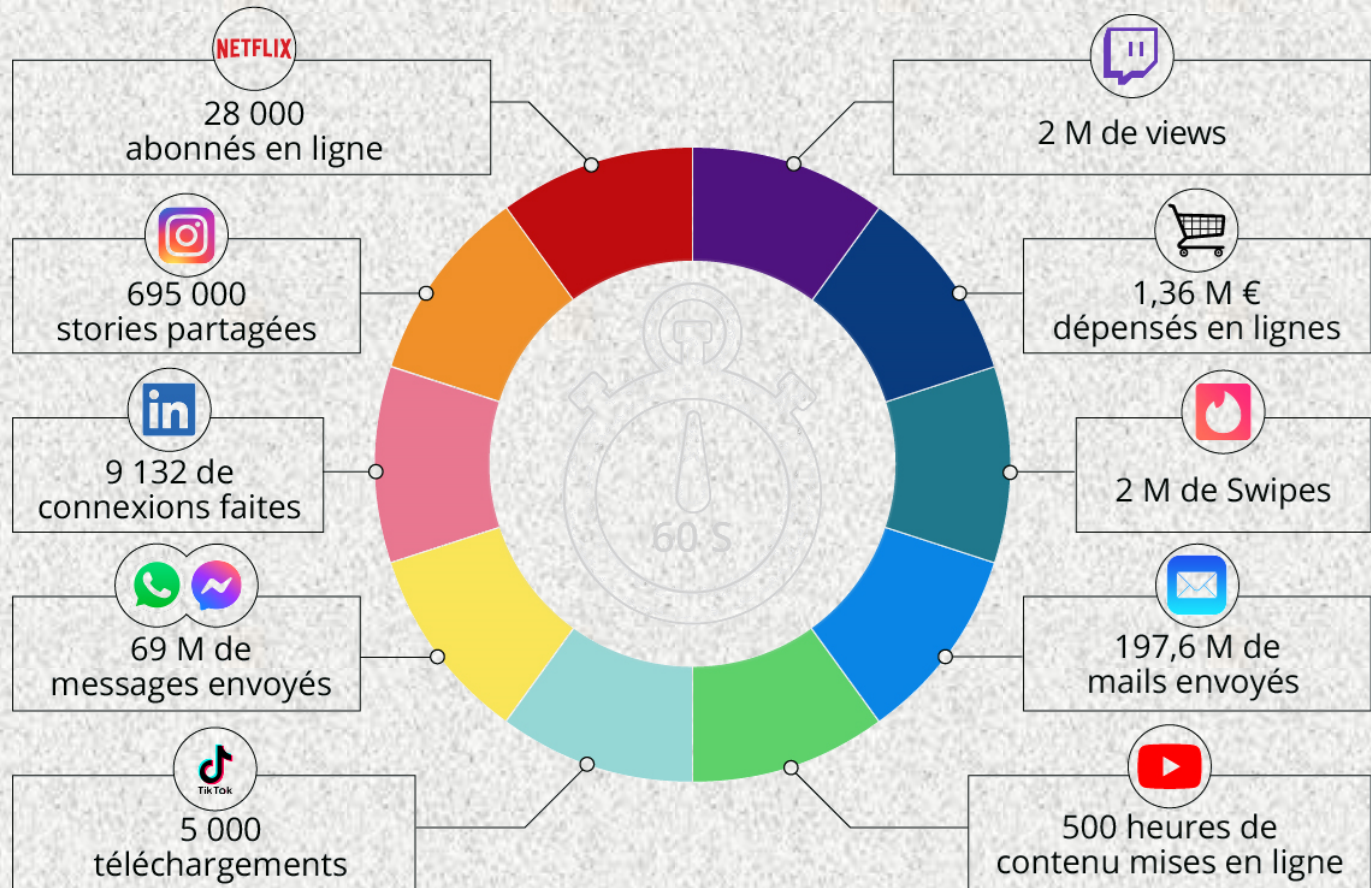
Ce flux de données éparses et multiformes ne cesse d'augmenter :

- En 1992, le monde produisait 100 giga-octets de données par seconde.
- En 2020, on a atteint les 50 000.

le volume de données générées dans le monde devrait dépasser 180 zettaoctets à l'horizon 2025, soit une croissance annuelle moyenne de près de 40 % sur cinq ans.

Une minute sur Internet en 2021

Estimation de l'activité et des données générées sur Internet en l'espace d'une minute



Source : Lori Lewis via AllAccess

La business intelligence (BI) ou l'informatique décisionnelle consiste donc à :

- récupérer ces données (data),
- les organiser,
- les analyser pour en tirer des modèles qui serviront à la prise de décision au sein de l'entreprise.

La récupération des sources se fait via un outil informatique, appelé ETL (Extract, Transform and Load), qui se charge d'extraire et de charger les données dans un « entrepôt » datawarehouse ou datamart.

Des outils informatiques d'analyse décisionnelle permettent ensuite de créer :

- des rapports et
- des tableaux de suivi de l'activité de l'entreprise,

qui sont destinés à guider la direction générale dans ses orientations et ses choix.

La BI conduit à définir des indicateurs pertinents et fiables pour vérifier le bon fonctionnement de l'entreprise et la diriger.

C'est pour cela que l'on parle en français d'intelligence d'affaires ou d'informatique décisionnelle.

Rigueur et prospective :

la BI est un facteur de performance

- L'informatique décisionnelle représente donc le poste de pilotage de l'entreprise. Ou, plus précisément, le tableau de bord qui fait face au pilote et qui lui permet de garder le bon cap, même par temps d'orage ou en cas de faible visibilité.

- Elle a l'avantage de présenter de manière structurée et simple l'ensemble des informations de l'entreprise, et de permettre à tous les décideurs d'une même société (et éventuellement de ses filiales) d'utiliser les mêmes indicateurs pour une meilleure synergie.

La prise de décision est facilitée que ce soit pour donner une nouvelle orientation à l'entreprise (pénétrer de nouveaux marchés...) ou améliorer ses produits et ses services.

D'autre part, l'informatique décisionnelle a cette capacité de mettre en évidence tout écart et toute anomalie dans les statistiques de l'entreprise (production, commerce, finances, organisation...) et facilite la prise de décision pour les corriger rapidement.

Elle permet enfin, à travers l'analyse des big data, de réaliser des projections et des prédictions business.

Des prédictions toujours plus pointues au fur et à mesure des progrès technologiques. Car, à l'image de l'évolution des tableaux de bord des avions depuis le début de l'aéronautique, les performances de la business intelligence ont explosé avec l'arrivée du digital.

La BI de nos jours et ses enjeux

- Actuellement, les données applicatives métier sont stockées dans une base de données relationnelle ou non.
- Ces données sont extraites, transformées et chargées dans un entrepôt de données généralement par un outil de type ETL (Extract-Transform-Load).
- Un entrepôt de données peut prendre la forme d'un data warehouse ou d'un datamart.

- Le reporting est probablement l'application la plus utilisée encore aujourd'hui de l'informatique décisionnelle, il permet aux gestionnaires :
- De sélectionner des données relatives à telle période, telle production, tel secteur de clientèle, etc.
- De trier, regrouper ou répartir ces données selon les critères de leur choix.
- De réaliser divers calculs (totaux, moyennes, écarts, comparatif d'une période à l'autre...)

- De présenter les résultats d'une manière synthétique ou détaillée, le plus souvent graphique selon leurs besoins ou les attentes des dirigeants de l'entreprise.

Un système d'information décisionnel (SID) doit être capable d'assurer quatre fonctions fondamentales :

- la collecte,
- l'intégration,
- la diffusion
- et la présentation des données.

À ces quatre fonctions s'ajoute une fonction d'administration, soit le contrôle du SID lui-même.

La BI futur

La BI est actuellement en train de se réinventer en cette période d'effervescence technologique : le Big Data permet en effet d'explorer de nouvelles possibilités qui vont révolutionner l'informatique décisionnelle. Néanmoins, la BI étant grandement implanté et au vu des coups qu'elle engendre, la solution la plus probable serait soit une amélioration des outils BI pour tourner vers du BIG data, soit une alliance des deux.

Objectifs

Etudiants $\Rightarrow$ Data Integrator $\Rightarrow$ Data Analyst $\Rightarrow$ Data scientists

Donner des clefs de compréhension autour du rôle et de la gestion des données en entreprise

Aborder des problématiques de traitement/intégration de données sur des exemples concrets

Présenter des outils du domaine pro

Et puis...

Développer la créativité autour du traitement de données et de ses applications

Travailler en équipe (expérimental)

Objectif :

L'informatique décisionnelle (en anglais : DSS pour *Decision Support System* ou encore BI pour *Business Intelligence*) désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données, matérielles ou immatérielles, d'une entreprise en vue d'offrir une aide à la décision et de permettre aux responsables de la stratégie d'entreprise d'avoir une vue d'ensemble de l'activité traitée.

Ce type d'application utilise en règle générale un entrepôt de données (ou *datawarehouse* en anglais) pour stocker des données transverses provenant de plusieurs sources hétérogènes et fait appel à des traitements par lots pour la collecte de ces informations

L'informatique décisionnelle s'insère dans l'architecture plus large d'un système d'information. Néanmoins l'informatique décisionnelle n'est pas un concept concurrent du management du système d'information. Au même titre que le management relève de la sociologie et de l'économie, la gestion par l'informatique est constitutive de deux domaines radicalement différents que sont le management et l'informatique. Afin d'enrichir le concept avec ces deux modes de pensées, il est possible d'envisager un versant orienté ingénierie de l'informatique portant le nom d'*informatique décisionnelle*, et un autre versant servant plus particulièrement les approches de gestion appelé *management du système d'information*.

Enjeux de l'informatique décisionnelle

De nos jours, les données applicatives métier sont stockées dans une (ou plusieurs) base(s) de données relationnelle(s) ou non relationnelles. Ces données sont extraites, transformées et chargées dans un entrepôt de données généralement par un outil de type ETL (*Extract-Transform-Load*) ou en français ETC (*Extraction-Transformation-Chargement*).

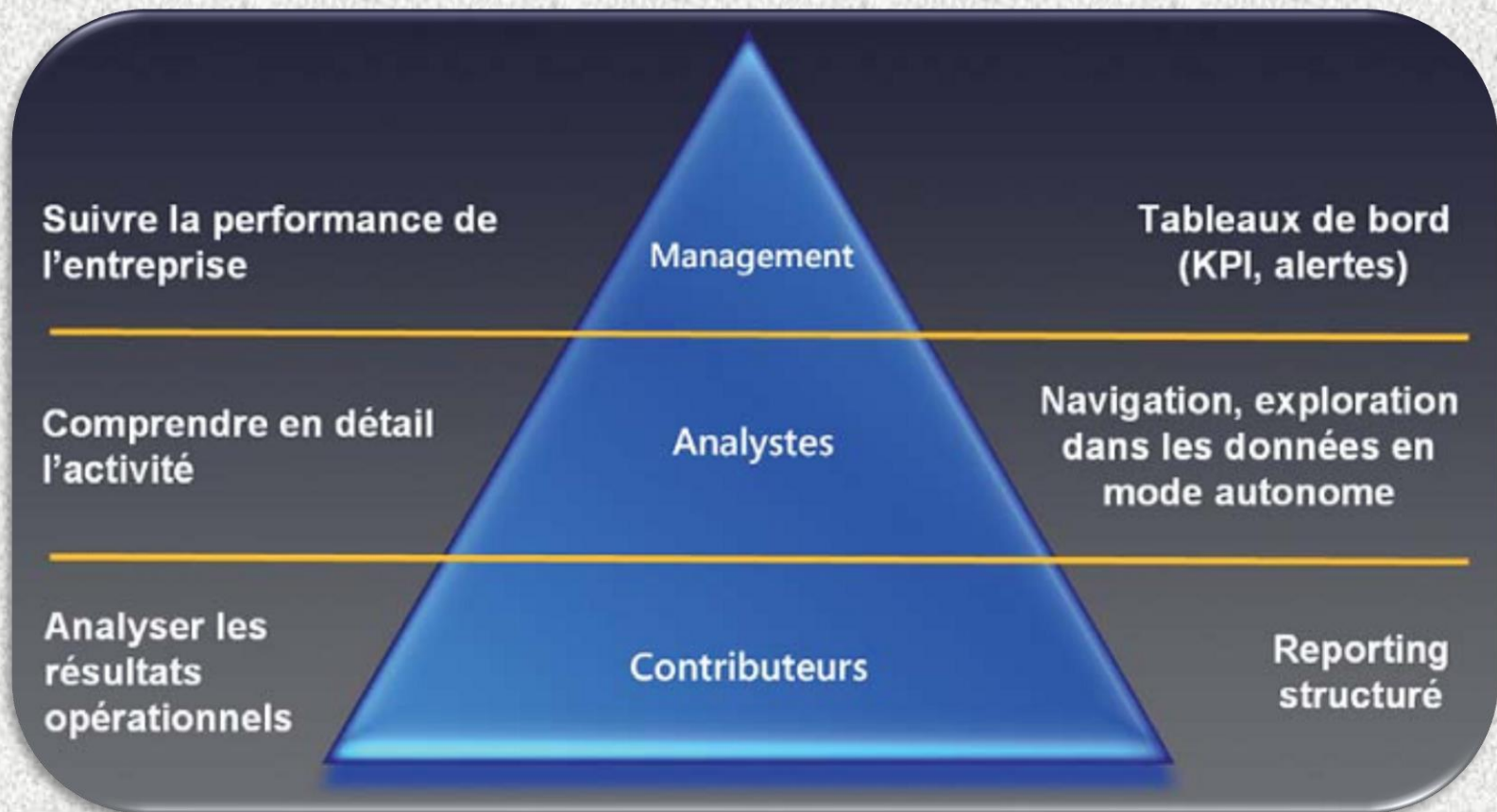
Un entrepôt de données peut prendre la forme d'un **datawarehouse** ou d'un **datamart**. En règle générale, le datawarehouse globalise toutes les données applicatives de l'entreprise, tandis que les datamarts (généralement alimentés depuis les données du datawarehouse) sont des sous-ensembles d'informations concernant un métier particulier de l'entreprise (marketing, risque, contrôle de gestion, ...). Le terme **comptoir de données** ou **magasin de données** est aussi utilisé pour désigner un datamart..

Les entrepôts de données permettent de produire des rapports qui répondent à la question « Que s'est-il passé ? », mais ils peuvent être également conçus pour répondre à la question analytique « Pourquoi est-ce que cela s'est passé ? » et à la question pronostique « Que va-t-il se passer ? ». Dans un contexte opérationnel, ils répondent également à la question « Que se passe-t-il en ce moment ? », voire dans le cas d'une solution d'entrepôt de données actif « Que devrait-il se passer ? ».

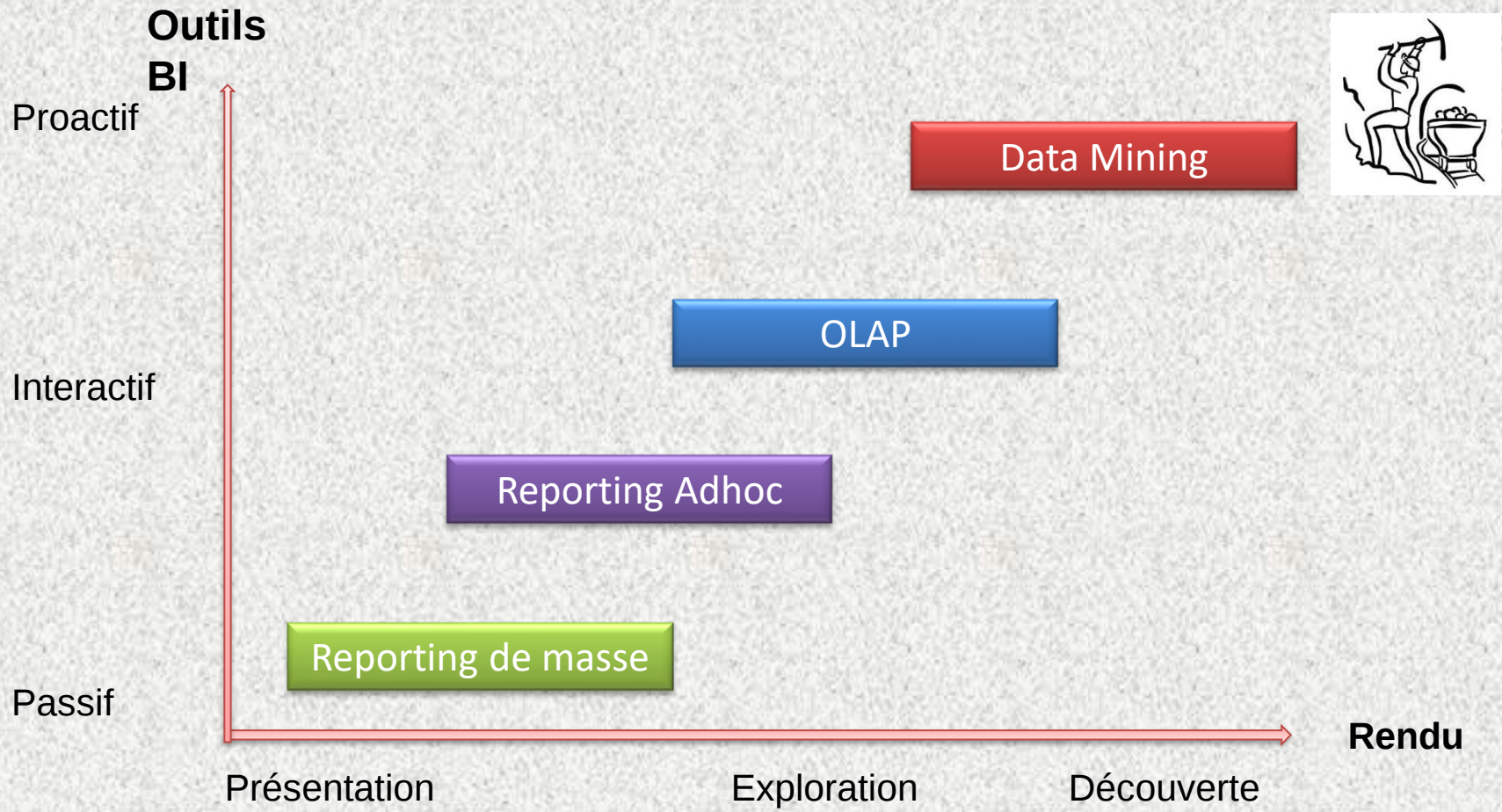
Le reporting est probablement l'application la plus utilisée encore aujourd'hui de l'informatique décisionnelle, il permet aux gestionnaires :

- de sélectionner des données relatives à telle période, telle production, tel secteur de clientèle, etc.,
- de trier, regrouper ou répartir ces données selon les critères de leur choix,
- de réaliser divers calculs (totaux, moyennes, écarts, comparatif d'une période à l'autre, ...),
- de présenter les résultats d'une manière synthétique ou détaillée, le plus souvent graphique selon leurs besoins ou les attentes des dirigeants de l'entreprise.

La BI pour qui ? Pourquoi ?



La chaine décisionnelle



Généralités

La notion d'entreprise

L'entreprise est une unité économique autonome disposant de moyens humains et matériels qu'elle combine en vue de produire des biens et services destinés à la vente. Son environnement interne est composé de trois acteurs principaux : **apporteurs de capitaux, dirigeants, et salariés.**

Ces trois acteurs ont des objectifs et des stratégies individuelles différentes. En général, la finalité d'une entreprise économique est de réaliser des profits, satisfaire le consommateur, et assurer la pérennité dans le marché.

L'entreprise entreprend des relations avec un environnement externe qui comprend plusieurs parties : les clients, les partenaires, les autres entreprises (fournisseurs ou concurrents), les banques, les investisseurs, le public, les associations...etc. Cet environnement exerce sur elle une influence plus ou moins importante d'une façon directe ou indirecte.

Les sous-systèmes composants une entreprise

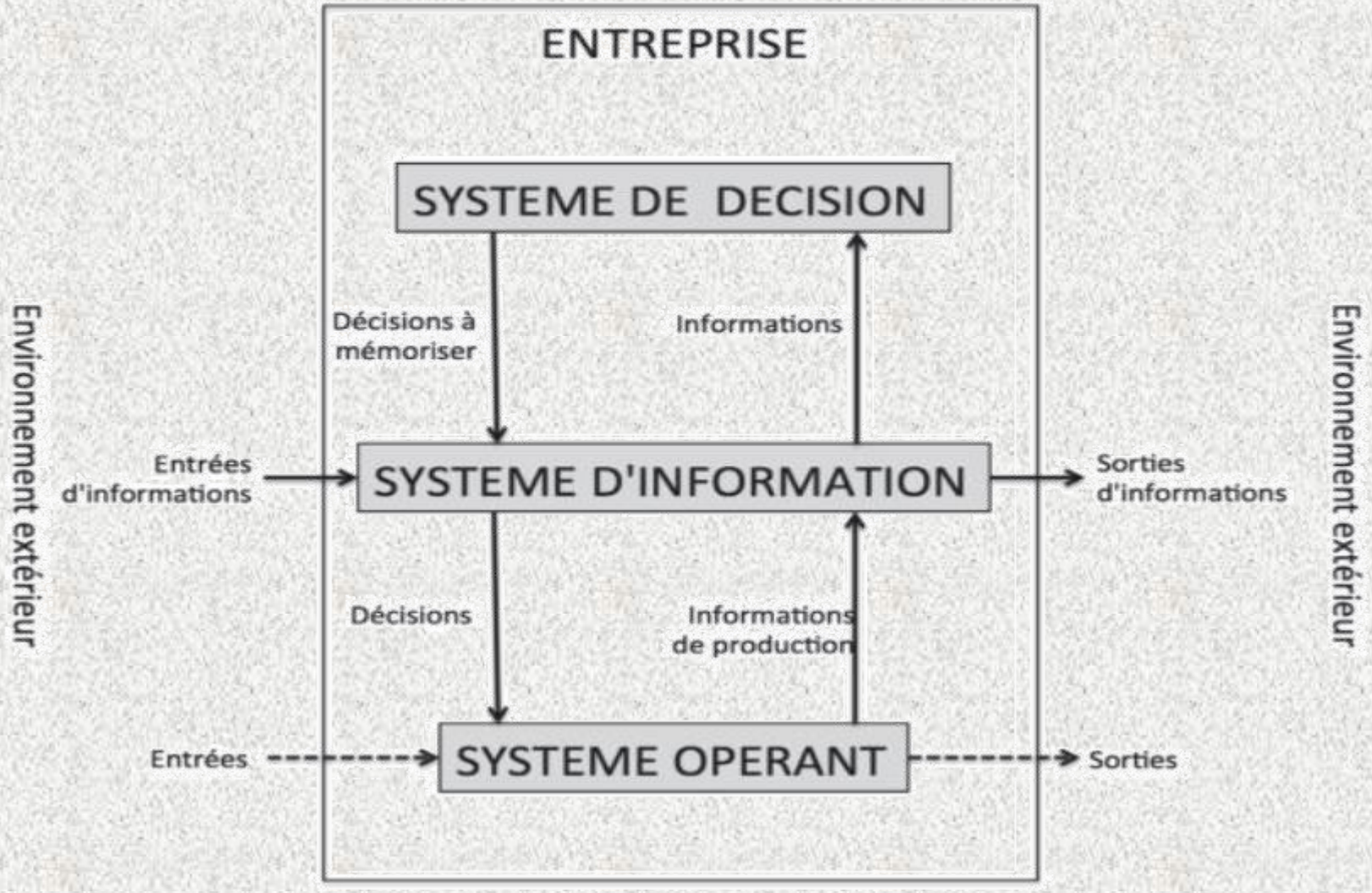
Une entreprise en tant que système complexe est composée de trois sous-systèmes :

- le système de décision (pilotage),
- le système opérant
- et le système d'information.

Le système d'information et le système opérant coopèrent en échangeant des informations provenant de l'environnement interne et externe du système afin d'atteindre les objectifs tracés par le système de décision.

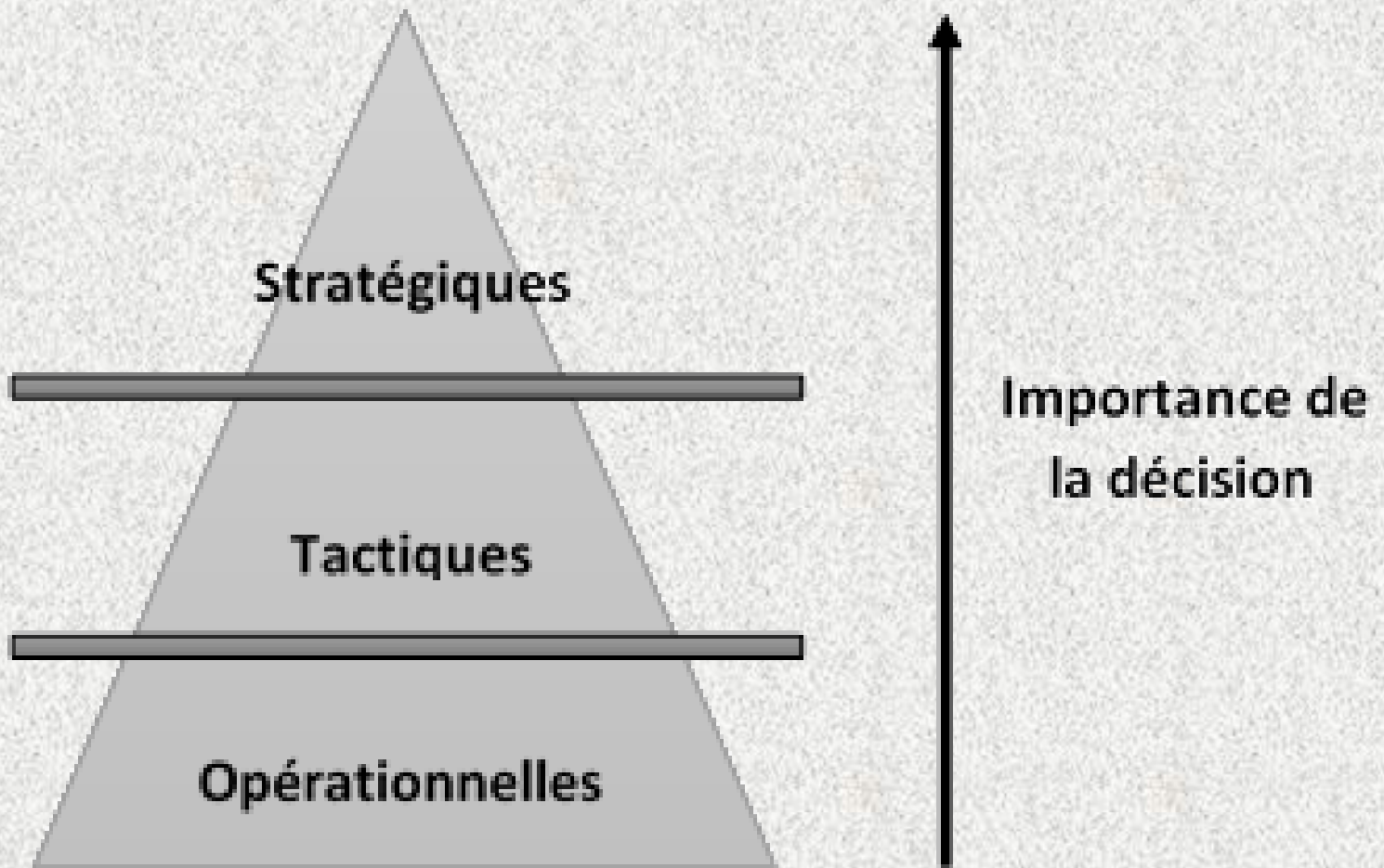
- **Le système opérant** (système de production) réalise la production physique des biens et des services. Son activité est contrôlée par le système de décision. Il est relié à l'environnement par des flux physiques externes et aux autres sous- systèmes par des flux internes d'information.
- **Le système de décision** (système de pilotage ou système de management) contrôle et analyse l'environnement et le fonctionnement interne de l'entreprise. Il est relié aux autres sous-systèmes par des flux internes d'information.

- **Le système d'information** alimente l'entreprise en informations. Il mémorise les informations, les traite et les communique aux deux autres sous-systèmes auxquels il est relié.



Les décisions prises quotidiennement dans l'entreprise ne sont pas toutes de même nature ni de même importance :

- **Les décisions opérationnelles** sont des décisions répétitives à effet immédiat et dont le résultat est connu d'avance. Elles portent sur l'exploitation courante de l'entreprise (ex. : établissement des plannings, décision de réapprovisionnement en pièces détachées...). Elles peuvent être prises à tous les niveaux de responsabilité.
- **Les décisions tactiques** (ou administratives) sont relatives à la gestion des ressources de l'entreprise (ex. : décision d'acquérir un brevet, d'organiser des formations pour les salariés...).
- **Les décisions stratégiques** déterminent l'orientation générale de l'entreprise (la politique générale, les activités, les marchés à pénétrer...etc.). Elles sont prises surtout à partir des informations émanant de l'environnement de l'entreprise. Elles ont un caractère non répétitif et sont très souvent irréversibles (ex. : décision de lancement d'un nouveau produit).



Niveaux d'importance des décisions dans l'entreprise

La prise de décision au sein de l'entreprise

Quel que soit la décision à prendre dans l'entreprise, le processus de la prise de la décision passe par cinq phases essentielles au moins :

- 1) **Définir** le problème à résoudre.
- 2) **Modéliser** les solutions possibles.
- 3) **Choisir** une solution parmi les alternatives envisagées. Le choix doit être en fonction des objectifs fixés, des critères et des expériences passées.
- 4) **Mettre en œuvre la décision** en mobilisant tous les moyens matériels et financiers nécessaires.
- 5) **Evaluer la décision** en comparant les résultats obtenus à ceux anticipés. Des mesures correctives doivent être mises en place en cas d'apparition des écarts significatifs entre le résultat effectif de la décision et l'objectif fixé auparavant.

Le processus de la prise de décision est affecté non seulement par le style de direction et la politique interne et externe de l'entreprise, mais aussi par les facteurs suivants :

- Les caractéristiques de l'entreprise (ses objectifs, son histoire, sa culture, son mode de gouvernance, ...etc.).
- La perception du problème par les décideurs.
- L'incertitude, la dynamique et la complexité de l'environnement.
- Les acteurs sociaux qui sont en relation avec l'entreprise : les **actionnaires, les clients, le personnel et les partenaires.**
- L'image de l'entreprise auprès de l'opinion publique.

Les indicateurs de performance d'une entreprise

Un indicateur clé de performance KPI (Key Performance Indicator), est un paramètre ou une combinaison de paramètres quantitatifs ou qualitatifs qui prennent place dans des tableaux de bord et qui sont utilisés à différents niveaux de l'entreprise. Ils permettent d'analyser l'état d'évolution du système de l'entreprise. Ils permettent également de mesurer, évaluer et suivre les résultats des actions prises par l'entreprise.

Les indicateurs de performance peuvent être utilisés pour exprimer un résultat ou un suivi. Un indicateur de résultat est relié à l'accomplissement ou non de l'objectif fixé. Un indicateur de suivi sert à anticiper l'action et réajuster les ressources en vue d'atteindre l'objectif fixé.

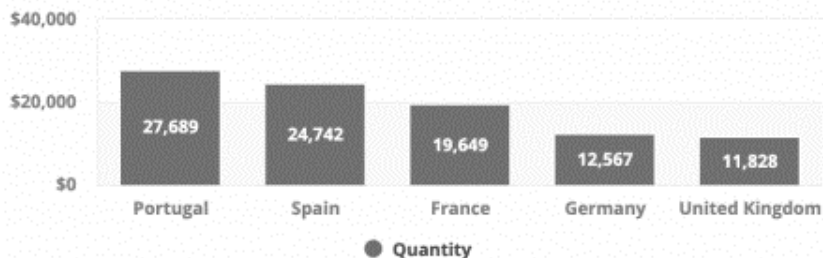
On distingue plusieurs familles d'indicateurs présents dans les tableaux de bord :

- Les indicateurs de production : capacité de production, capacité de production inutilisée, coût de revient par produit, durée moyenne du cycle de production, durée de consommation des stocks critiques.
- Les indicateurs financiers : la valeur ajoutée, la trésorerie prévisionnelle, le coût de revient, retard de paiement, écart sur chiffre d'affaires, marge brute.
- Les indicateurs de rentabilité : résultat net, marge opérationnelle,
- Les indicateurs de qualité : délai de fabrication, satisfaction des clients, réclamation.
- ...etc.

[Asia](#)[Europe](#)[America](#)

Top Sales

by Country



Select Dates:

Last Year To Date

Current Year To Date

From

1/1/2018

To

12/31/2018

\$1,007M

Best Selling Car Model Sales

42.89%

% Change



108,154

Total Sales in Qty (Current Period)

19.89%

% Change



\$840M

Westfield Megabusa

Best Selling Car Model

Total Sales in \$ (Previous Period)

75,692

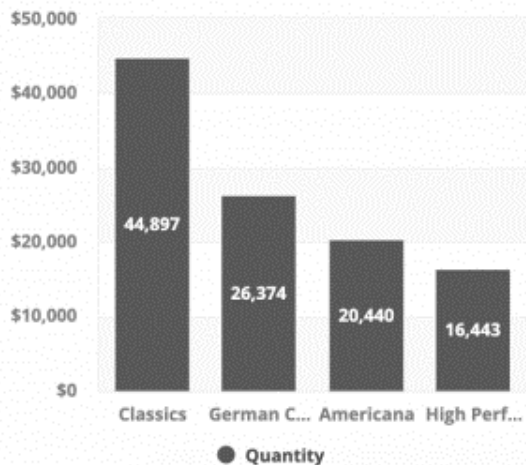
Total Sales in Qty (Previous Period)

\$1,007M

Total Sales in \$ (Current Period)

Top Sales

by Segment

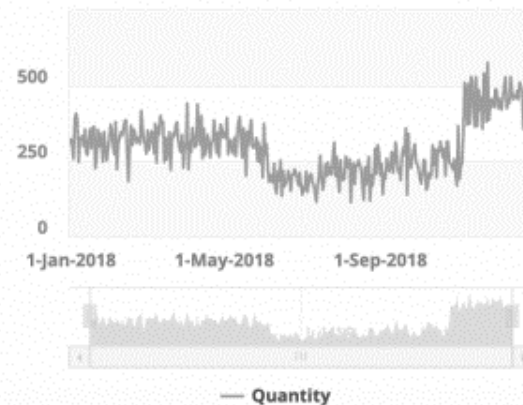


Country Sales



Daily Sales

Zoom 1m 3m 6m YTD 1y All



Définitions – Vocabulaire

- **Infocentre** = Base de données dédiée à l'analyse d'une application
→ A l'origine des "copies" de base de production
- **Staging Area** : Espace de stockage temporaire des données provenant des différentes sources. C'est un lieu où l'on va pouvoir stocker des données qui arrivent à des moments différents. C'est une zone d'attente, une "salle d'embarquement".
- **ODS (Operating Data Store)**
 - Lieu où vont être effectuées les transformations, les croisements, etc. C'est l'étape juste avant l'alimentation du Datawarehouse et il utilise, comme source la Staging Area.
 - Le modèle de données de l'ODS est un modèle relationnel classique assez proche des modèles de production
 - Ce « sas » assure l'isolation des mondes "opérationnel" et "décisionnel" (performances, exploitabilité...)

Définitions – Vocabulaire

Datawarehouse = Entrepôt de données, destiné à centraliser, nettoyer, et uniformiser les données de l'entreprise à des fins de reporting et d'analyse. Il stocke l'historique des données avec la granularité la plus fine,

Data Mart : Entrepôt de données dédié à un métier particulier. Il est situé en aval du Datawarehouse

L'informatique décisionnelle (Business Intelligence)

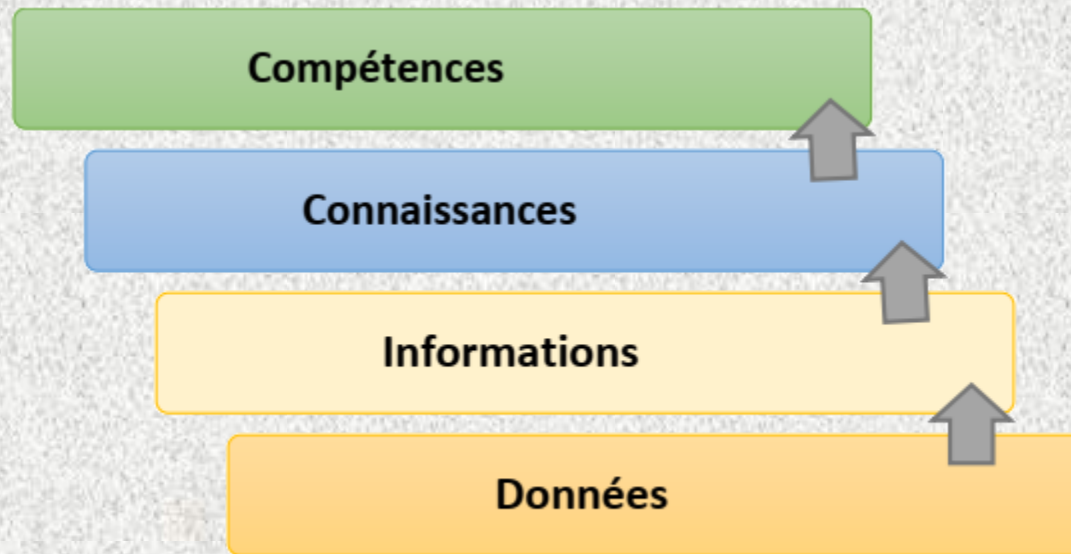
L'Informatique Décisionnelle (**ID**), *en anglais* Business Intelligence (**BI**), est un terme qui a été inventé au milieu des années 1990 pour décrire le processus de la transformation des données brutes à des éléments d'informations utilisables sur lequel les décisions commerciales peuvent être fondées. C'est un terme générique qui relie d'autres disciplines informatiques étroitement liées, notamment l'exploration de données (Data Mining), l'analyse statistique, la prédiction (*forecasting*) et l'aide à la décision.

De point de vue de l'entreprise, la Business Intelligence est une stratégie commerciale visant à comprendre et à anticiper les besoins et les exigences d'une entreprise. C'est la connaissance des clients, des concurrents, des partenaires commerciaux, de l'environnement concurrentiel et de ses propres opérations internes qui permet à la direction de l'entreprise de prendre des décisions commerciales efficaces, importantes et souvent stratégiques.

La Business Intelligence (**BI**) par conséquent, englobe l'ensemble de processus, technologies et outils qui vont permettre de transformer des données en informations, ces informations en connaissances et ces connaissances en plans permettant d'aboutir à une meilleure gestion, plus durable de l'entreprise.

Rappelant qu'une donnée est un élément brut, qui n'a pas encore été interprétée ou mis en contexte. Elle peut être collectée par un outil, par une personne ou juste récupérée via une source annexe. En revanche, une information est une donnée interprétée placée dans un contexte.

La connaissance est une information traitée, utilisée pour aboutir à une action. Elle peut être aussi un ensemble d'informations structuré, orienté sur un sujet, et validé par des règles établies ou des expériences. L'utilisation des connaissances dans le contexte d'activité s'appelle du savoir-faire ou des compétences.



Outils de la Business Intelligence

La Business Intelligence représente une large catégorie d'applications et de technologies qui combinent les données opérationnelles avec les outils analytiques pour aider les décideurs et les dirigeants des entreprises à prendre des meilleures décisions et soutenir leurs stratégies d'affaires. Les outils de BI peuvent aller d'outils très simples à des outils très sophistiqués qui offrent un ensemble très large et complexe de fonctionnalités.

La BI comprend :

- Les systèmes d'aide à la décision
- Les analyses prédictives
- L'entreposage de données
- Les outils d'extraction, transformation et chargement des données (ETL)
- Le traitement analytique en ligne (OLAP)

- L'exploration de données (data mining)
- Les analyses Big Data
- L'analyse statistique
- L'analyse des médias sociaux
- Les outils de restitution de données (Reporting)
- Les rapports
- Les requêtes
- Les tableaux de bord
- ...etc.

Historique de la Business Intelligence

Les systèmes informatiques de « Business Intelligence » remontent, sous une forme ou une autre, à près de cinq décennies (fin des années 1960). On trouve son apparition dans les recherches et les études sur les systèmes d'aide à la décision (DSS, Decision Support System).

Le terme « Business Intelligence » est un terme générique popularisé introduit par Howard Dresner du groupe Gartner en 1989. La Business Intelligence décrit un ensemble de concepts et de méthodes pour améliorer la prise de décision commerciale. Ce terme est parfois utilisé pour désigner entre autres les outils de création des rapports et des requêtes, et les systèmes d'information de direction (EIS, Executive Information System) qui ont évolué à partir de systèmes à modèle mono-utilisateur d'aide à la décision.

À partir de 1990 environ, l'entreposage de données et OLAP ont commencé à élargir le domaine de l'EIS et à définir une catégorie plus vaste des DSS pilotés par les données. D'autre part, le Web et l'Internet ont accéléré les développements dans le domaine de l'aide à la décision et ont fourni un nouveau moyen pour capturer et appliquer les connaissances. Ainsi, le DSS basé sur le Web ou intégrant le Web est devenu réalisable au cours des années suivantes.

Au fur et à mesure que l'analyse commerciale se développait du Reporting périodique de base à l'OLAP de première génération, au ROLAP plus avancé, et plus tard aux systèmes d'entreposage de données, les entreprises ont eu du mal à équilibrer le temps et le coût de mise en œuvre par rapport à leur besoin augmenté d'informations commerciales pertinentes et utilisables. Le domaine de la Business Intelligence a été ainsi exploité pour fournir une large variété d'outils, d'applications et de méthodologies permettant de collecter des données en provenance de systèmes internes et de sources externes, de les préparer pour l'analyse, de les

développer et de lancer des requêtes au sein de ces ensembles de données. Ces outils permettent de créer des rapports, des tableaux de bord et des visualisations de données pour rendre les résultats des analyses disponibles pour les preneurs de décisions.

De temps à autre, le terme de Business Intelligence est remplacé par celui de « Business Analytics », qui désigne plus généralement les technologies analytiques avancées mais peut aussi inclure la Business Intelligence.

L'utilité de la Business Intelligence

L'informatique décisionnelle permet d'effectuer des tâches importantes telles que les prédictions basées sur les données et les expériences passées et courantes de l'entreprise ; les analyses de type « What-If » sur les impacts du changement de l'environnement de l'entreprise et les scénarios alternatifs ; un accès ad-hoc aux données pour répondre à des questions spécifiques et inhabituelles et à des idées stratégiques pour n'en citer que quelques-unes. On peut mentionner aussi les avantages suivants :

- La gestion de la performance d'entreprise ;
- L'optimisation de la relation client ;
- Le suivi de l'activité commerciale et l'aide à la décision ;
- L'amélioration de l'efficacité et de la productivité de l'entreprise ;
- L'amélioration des relations commerciales ;
- La réduction des coûts.

Applications de la BI

Les outils BI sont nécessaires pour presque toutes les secteurs et les domaines.

Aujourd'hui, les dirigeants et les décideurs ont besoin d'accéder aux outils de BI pour disposer des indicateurs à jour sur la prestation et la performance de l'entreprise. Voici quelques domaines d'applications de BI :

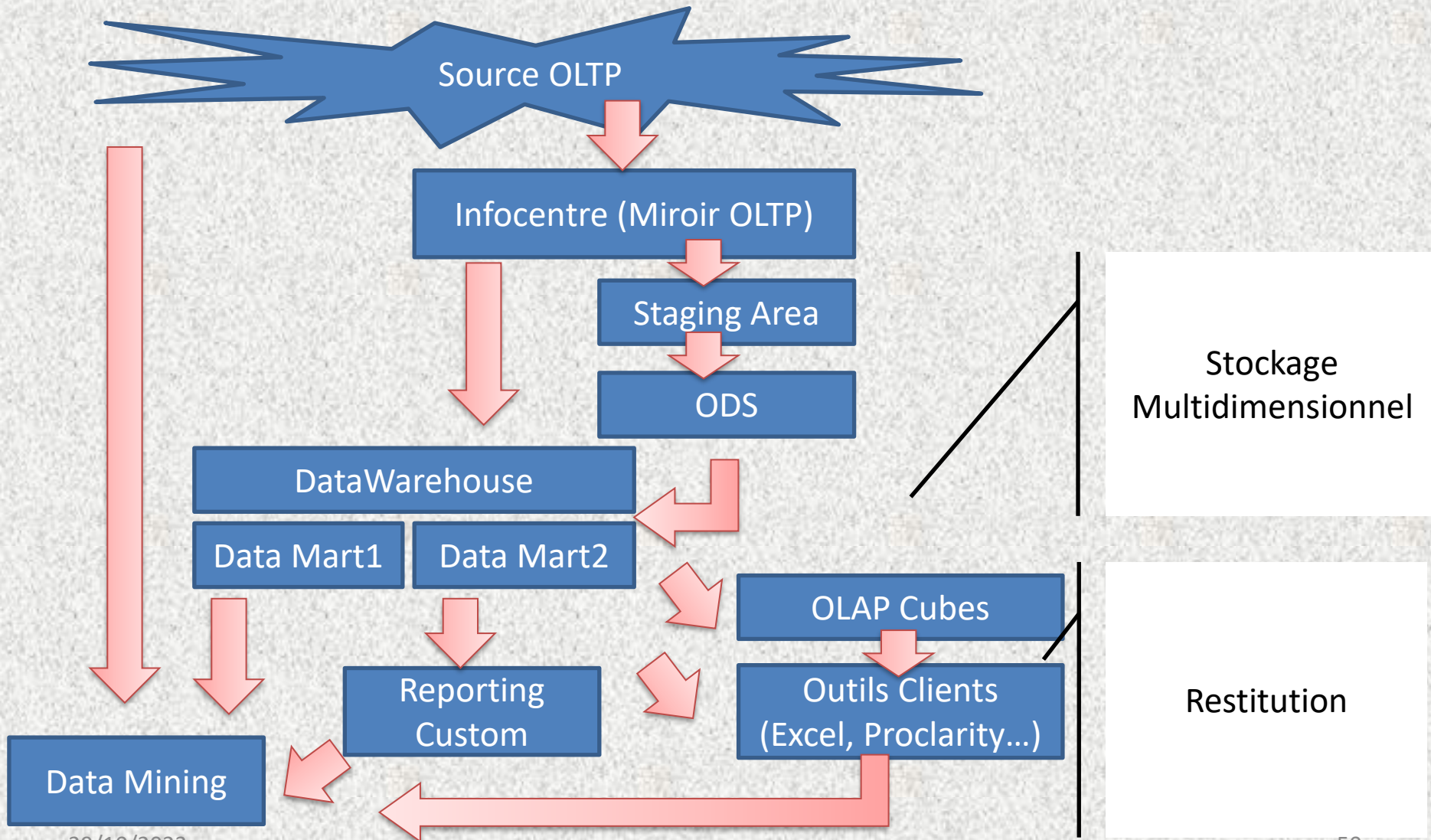
La gestion de la relation client (CRM)

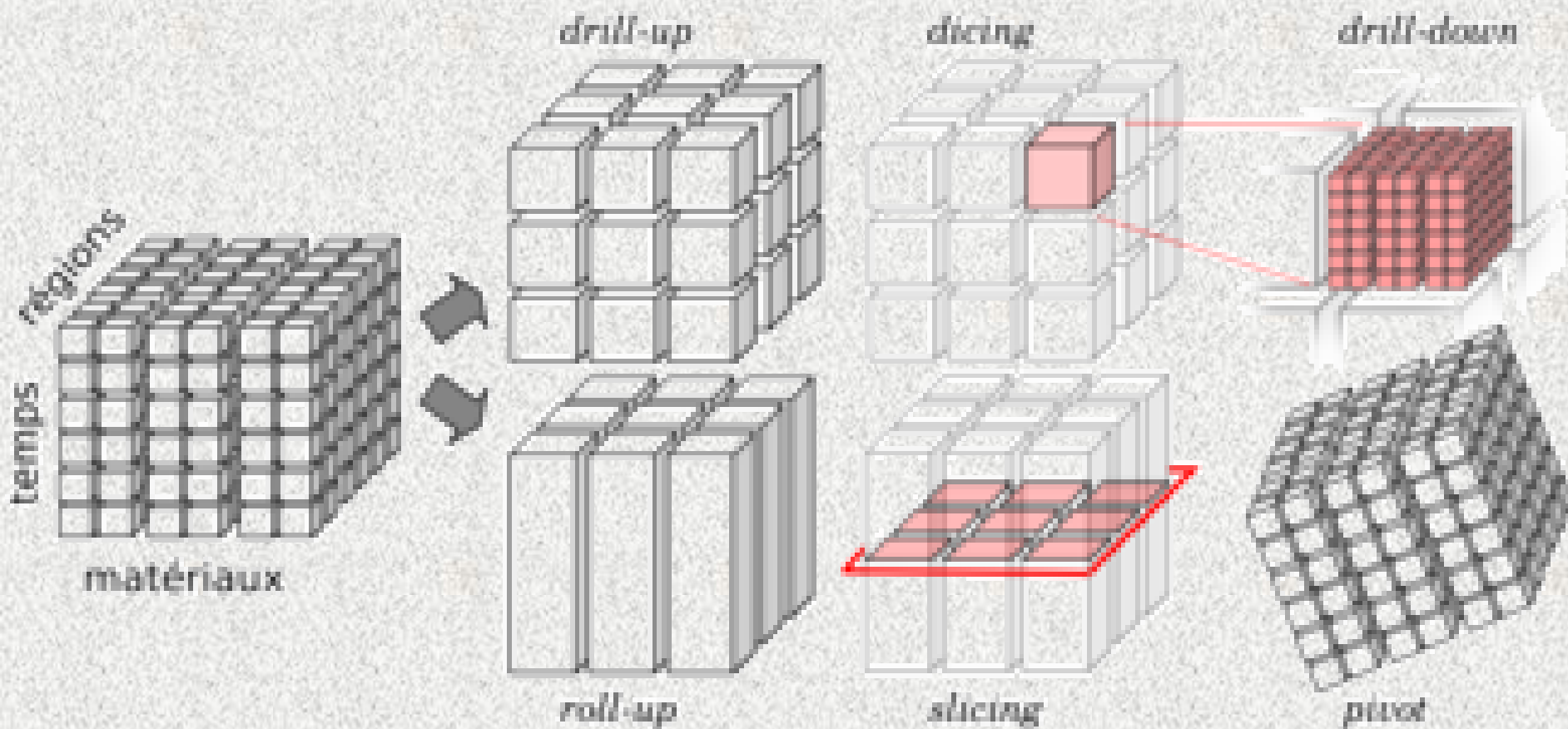
Une entreprise doit comprendre les besoins et les attentes du client, vendre davantage de ses offres aux clients existants et élargir le récipient de clients qu'elle sert. Les applications BI peuvent avoir un impact sur de nombreux aspects du marketing :

- Maximiser les revenus des campagnes de marketing ;
- Améliorer la fidélisation des clients (analyse du taux de désabonnement) ;
- Maximiser la valeur client (vente croisée, vente incitative) ;
- Identifier et satisfaire des clients hautement valorisés ;
- Gérer l'image publique de la marque

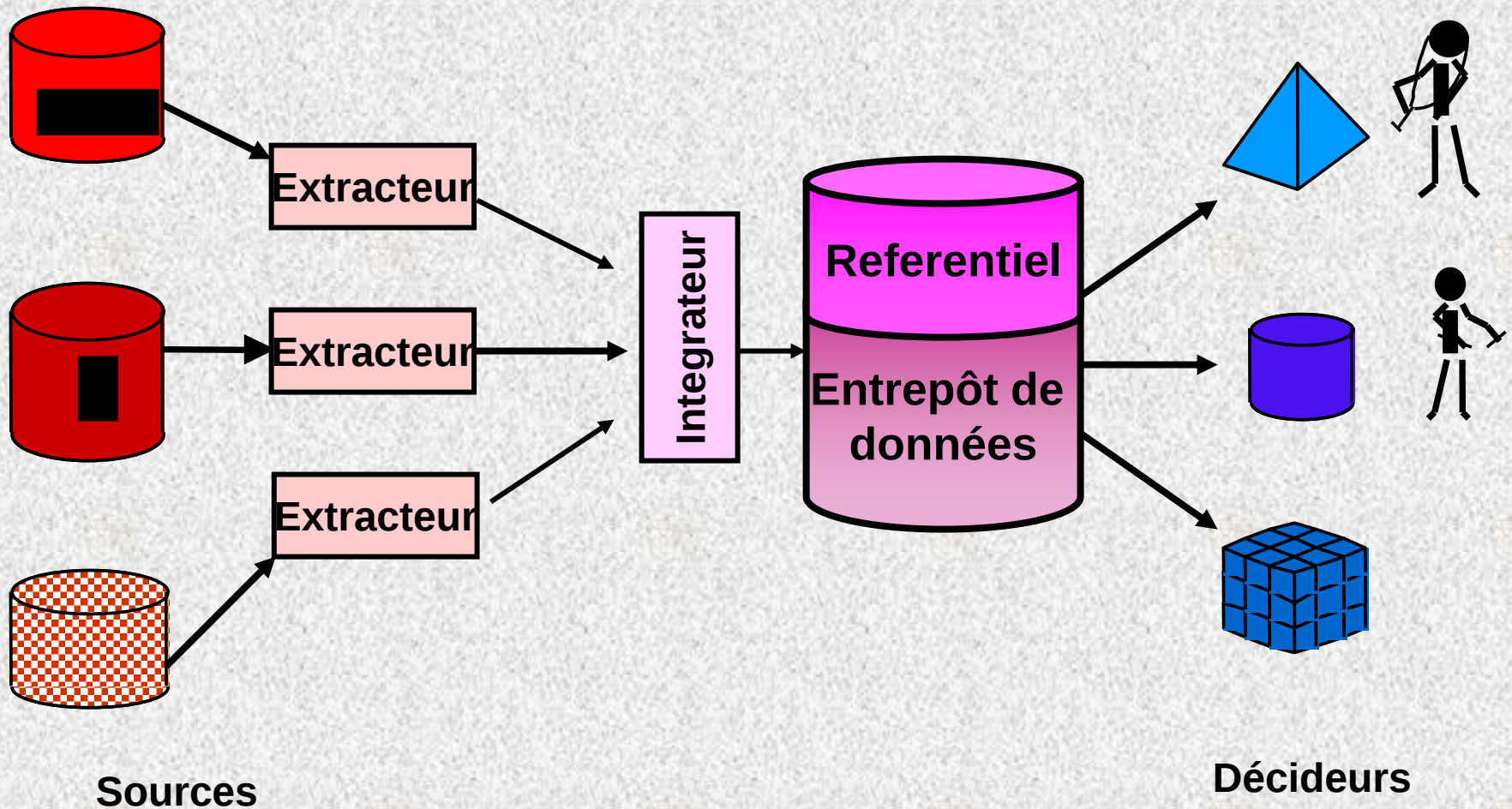
- **La santé**
- **La vente au détail**
- **Les banques**
- **Les assurances**
- **Les télécoms**
- **Les gouvernements**

Architecture cible





Architecture des systèmes décisionnels



Système Décisionnel (1)

- **Un SID (système d'information décisionnel), c'est :** « un système permettant aux utilisateurs de l'entreprise de disposer d'informations pertinentes et d'outils d'analyse puissants pour les aider à prendre les bonnes décisions au bon moment » (Bill Inmon)
- **Un projet décisionnel est réussi quand l'utilisateur dit :** « J'ai l'information, elle est sûre, je la comprends, donc je l'utilise »

Systeme Décisionnel (2)



Les métiers du BI

4 métiers

- Data Integrator
- Data Analyst
- Data Scientist
- Data Steward (Responsable des données)

Fonctions

Data Integration

- Combiner des informations hétérogènes venants de sources différentes

Data Analysis

- Inspection, nettoyage, transformation et modélisation des données.
- Data Mining,
- Data Visualisation
- Rendre la donnée compréhensible
Communiquer à partir de la donnée

- **Data Scientist**

Il s'agit de disposer de compétences de haut niveau en matière d'analyse de données, en combinant à la fois les méthodes statistiques, mais aussi d'autres connaissances telles que la linguistique, la sémantique, utiles notamment pour travailler sur des données non structurées, sans oublier la bonne compréhension du métier sur lequel on travaille, et de mettre en œuvre une démarche d'analyse itérative, en acceptant de tester des hypothèses sans a priori sur le résultat recherché.

- **Data Steward - Responsable des Données**
susceptibles sur un périmètre métier sur lequel ils détiennent une expertise reconnue, de spécifier les exigences sur les données et d'en contrôler la qualité. Ces responsables de données peuvent être positionnés à différents niveaux dans l'organisation, et peuvent être pilotés par des coordinateurs au niveau d'un métier, d'une fonction support ou d'une géographie

Communication

Should be able to convey your insights with confidence.

Data analyst to-do

Understanding the data you will be working on.

Ask Question

Keep asking question until you are confident enough with what is the requirement needs to be fulfilled by you.

Tool you should be good with

- Excel.
- SQL or Query Language.
- BI Tools.
- Python or R Language.

Problem solving

A data analyst must have a good understanding of the question being asked and the problem to be solved

Architecture Générale

Les données opérationnelles sont extraites périodiquement de sources hétérogènes : fichiers plats, fichiers Excel, base de données (DB2, Oracle, SQL Server, etc.), service web, données massives et stockées dans un entrepôt de données.

Les données sont restructurées, enrichies, agrégées, reformatées, nomenclaturées pour être présentées à l'utilisateur sous une forme sémantique (vues métiers ayant du sens) qui permettent aux décideurs d'interagir avec les données sans avoir à connaître leur structure de stockage physique, de schémas en étoile qui permettent de répartir les faits et mesures selon des dimensions hiérarchisées, de rapports pré-préparés paramétrables, de tableaux de bords plus synthétiques et interactifs.

Ces données sont livrées aux divers domaines fonctionnels (direction stratégique, finance, production, comptabilité, ressources humaines, etc.) à travers un système de sécurité ou de datamart spécialisés à des fins de consultations, d'analyse, d'alertes prédéfinies,

Architecture de Base

Extraction des données

- Bases de données
- Autres sources

Structuration des données

- Prétraitements
- Aggrégation
- Interface

Présentation des données

- Visualisation, alertes
- automatiques
- Pour une tâche donnée
- À destination d'un décideur

Stockage

Base de données opérationnelle

- Fonctionnement normal de l'entreprise
- Pas forcément un historique très grand
- Peut changer dans le temps

Datawarehouse

- Stockage pour le BI
- Archivage sur toute l'histoire de l'entreprise
- Format stable dans le temps

Datamart

- Vue métier
- À destination du décideur

Définitions

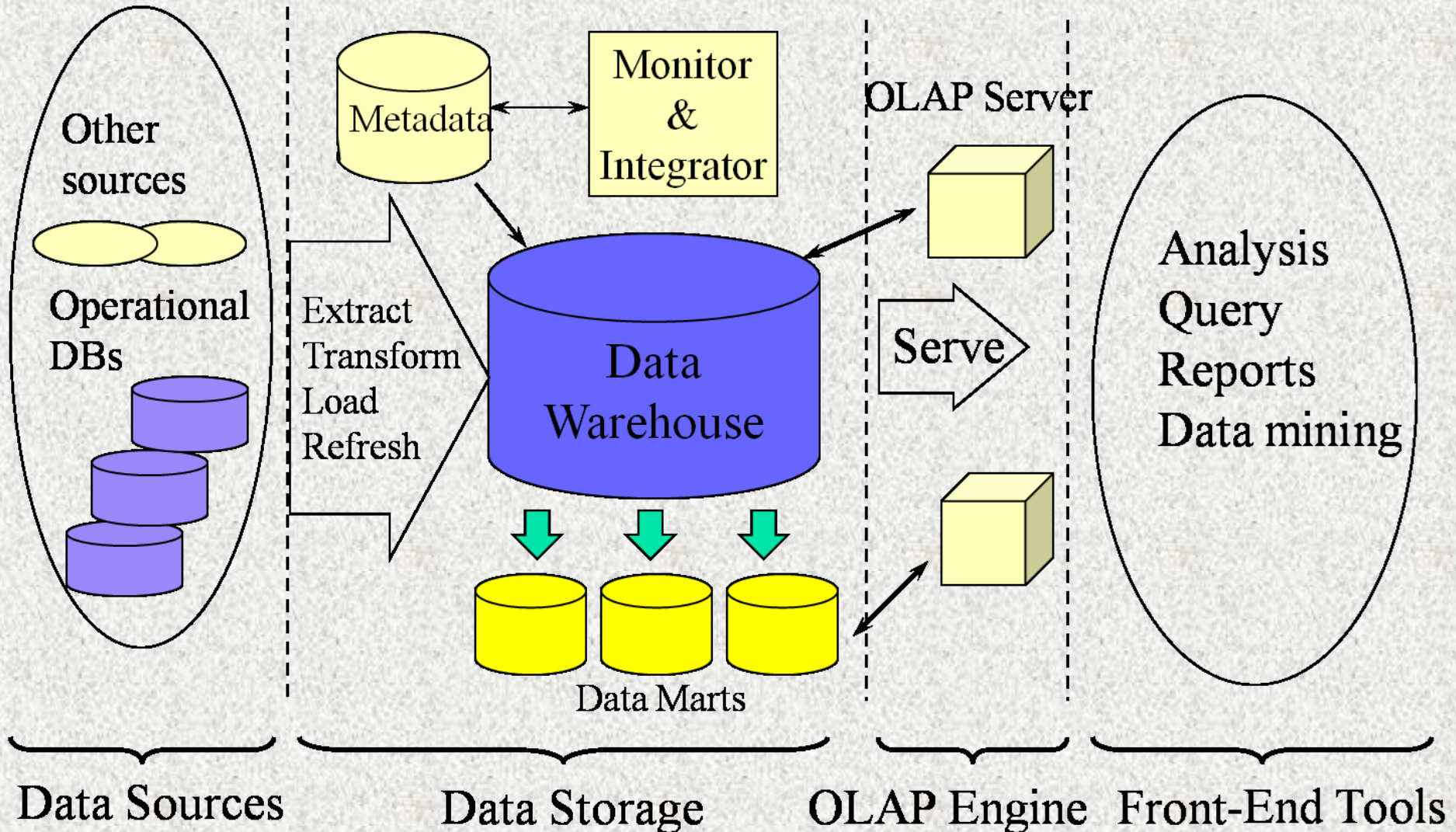
Datawarehouse

Le terme entrepôt de données (ou base de données décisionnelle, ou encore data warehouse) désigne une base de données utilisée pour collecter, ordonner, journaliser et stocker des informations provenant de base de données opérationnelles et fournir ainsi un socle à l'aide à la décision en entreprise.

Datamart

Un DataMart (littéralement en anglais magasin de données) est un sous-ensemble d'un DataWarehouse destiné à fournir des données aux utilisateurs, et souvent spécialisé vers un groupe ou un type d'affaire.

Datawarehouse



Les fonctions

Collecte de données

Intégration

Diffusion (ou distribution)

Présentation

Fonction de collecte

Définition

La fonction collecte (parfois appelée datapumping) recouvre l'ensemble des tâches consistant à détecter, sélectionner, extraire et filtrer les données brutes issues des environnements pertinents

Tâche

- Récupérer les données
- Méthodologie ETL

Données hétérogènes

Plusieurs types de sources

- Fichiers plats
- Fichiers Excel
- Bases de données (SQL)
- Services web
- Systèmes de stockages pour données Massives
- Interfaces exotiques

Plusieurs types de données

- Chiffres, texte, image
- Données statiques, flux
- Données bruitées, manquantes, erronées

Flux de données et données statiques

Données statiques

- Image à un instant donné de l'état de l'entreprise
- Rapports d'activité, bilans, inventaire

Flux de données

- Mise à jour en temps réel
- Compte rendus quotidiens, commandes, livraisons

Recodage

Mise sous forme canonique

- Choix d'une représentation unique
- Indépendante de la représentation en entrée

Stabilité dans le temps

- Un changement dans les formats d'entrées ne doit pas perturber l'analyse.

ETL

Méthodologie et outils

Extract

- Extraire les données de sources hétérogènes

Transform

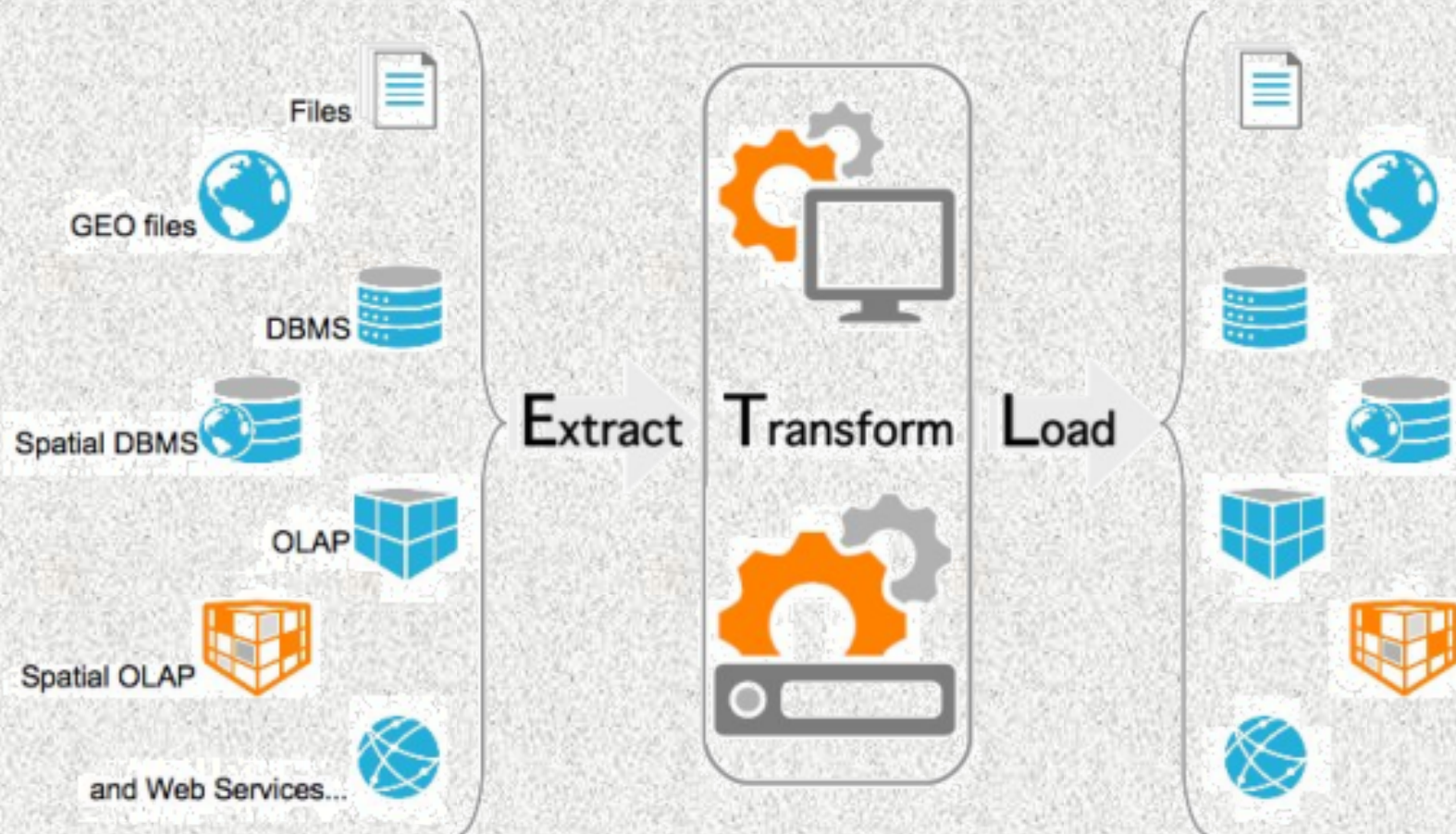
- Transformation des données pour les mettre dans un format acceptable

Load

- Charger les données dans le datawarehouse

ETL

Ensemble de connecteurs



ETL permet aux applications décisionnelles de masquer la diversité de l'origine des données et de bénéficier d'une source d'information commune, homogène, normalisée et fiable, au sein d'un système unique et si possible normalisé.

Logiciels d'ETL

Interfaces graphiques pour le non-spécialiste

- Anatella2
- DataStudio (Data)
- Feature Manipulation Engine (FME)
- Hurence avec un ETL natif Hadoop
- IBM InfoSphere DataStage
- Informatica PowerCenter
- MapReport
- Microsoft SQL Server Integration Services (SSIS)
- OpenText Genio
- Oracle Data Integrator (Sunopsis)
- Oxio Data Intelligence solution ETL
- SAP Data Services
- SAS Data Integration Studio
- Stambia
- STATISTICA ETL (StatSoft)

Pentaho Data Integration

Spoon - rift (changed)

File Edit View Action Tools Help

Perspective: Data Integration

View Design

Steps

- Input
- Output
- Transform
 - Add a checksum
 - Add constants
 - Add sequence
 - Add value fields changing sequence
 - Add XML
 - Calculator
 - Closure Generator
 - Example plugin
 - Number range
 - Replace in string
 - Row denormaliser
 - Row flattener
 - Row Normaliser
 - Select values
 - Sort rows
 - Split field to rows
 - Split Fields
 - String operations
 - Strings cut
 - Unique rows
 - Unique rows (HashSet)
 - Value Mapper
 - XSL Transformation

Welcome! DF_IMPORT_FACTS_back DF_IMPORT_FACTS FEED_ECHOS FEED_YAHOO rift

100%

```
graph LR; A[Get data from XML 2] --> B[Add constants]; C[Get data from XML] --> D[Add constants 2]; B --> E[Append streams]; D --> E; E --> F[User Defined Java Expression]; F --> G[Table output]
```

29/10/2023

74

Fonction d'intégration

Définition

La fonction d'intégration consiste à concentrer les données collectées dans un espace unifié, dont le socle informatique essentiel est l'entrepôt de données. Élément central du dispositif, il permet aux applications décisionnelles de masquer la diversité de l'origine des données et de bénéficier d'une source d'information commune, homogène, normalisée et fiable, au sein d'un système unique et si possible normalisé.

Tâches

- Deuxième passe de filtrage et validation
- Synchronisation
- Certification (liens avec des documents légaux)

Fonction de diffusion

Définition

La fonction de diffusion met les données à la disposition des utilisateurs, selon des schémas correspondant aux profils ou aux métiers de chacun, sachant que l'accès direct à l'entrepôt de données ne correspond généralement pas aux besoins spécifiques d'un décideur ou d'un analyste.

Tâche

- Choisir les données en fonction des besoins des utilisateurs
- Méthodologie OLAP

OLAP

Le modèle multidimensionnel présente une vue statique des données. Il a besoin d'être manipulé pour extraire des informations nécessaires à la prise de décision. L'exploitation des données multidimensionnelles peut se faire par divers outils (OLAP, outils de Reporting, fouille de données).

OLAP (On Line Analytical Processing) est un concept proposé par Codd en 1993. Il peut être défini comme étant l'analyse rapide d'information multidimensionnelle partagée.

Une analyse de type OLAP repose sur les principes suivants :

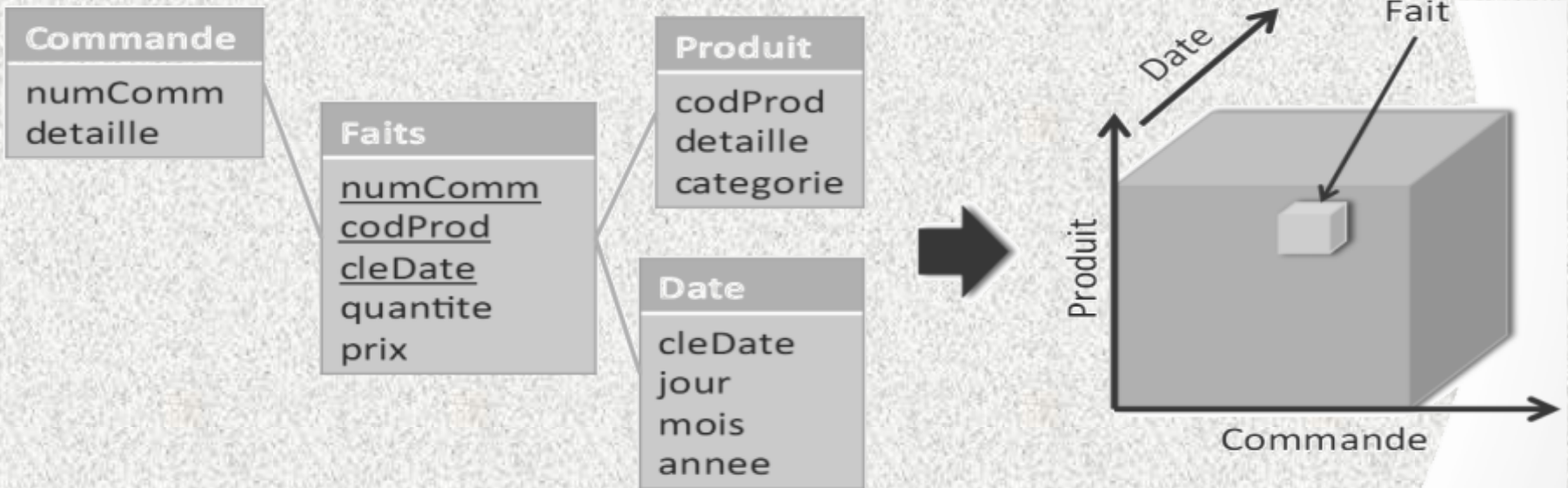
- Analyse en ligne.
- Permettre l'exploration et la navigation dans les données.
- Passage entre différents niveaux d'agrégation.
- Possibilité de sélection des données détaillées ou agrégées.

Notion de cube (hypercube) de données

Les fonctionnalités OLAP s'appliquent sur une structure de données multidimensionnelle qui s'appelle l'**hypercube de données**.

- Un hypercube représente l'information dans un cube à N dimensions.
- Il peut être extrait d'un entrepôt de données ou d'un magasin de données pour effectuer l'OLAP.
- Il peut contenir les données à la base (les plus détaillées) ou les données agrégées (cuboïde)

Exemple



Composantes d'un cube de données

Un cube de données possède les caractéristiques suivantes :

- Chaque **cellule** du cube correspond à une occurrence du fait.
- Chaque cellule contient des **indicateurs** (variables, métriques, ou mesures).
- Les axes d'analyse, également appelés **dimensions**, contiennent un ensemble de valeurs.
- Des **hiérarchies** sont spécifiées sur les dimensions afin de permettre une consolidation des indicateurs.
- Chaque indicateur a **une fonction d'agrégat** afin d'être exploité sur la hiérarchie

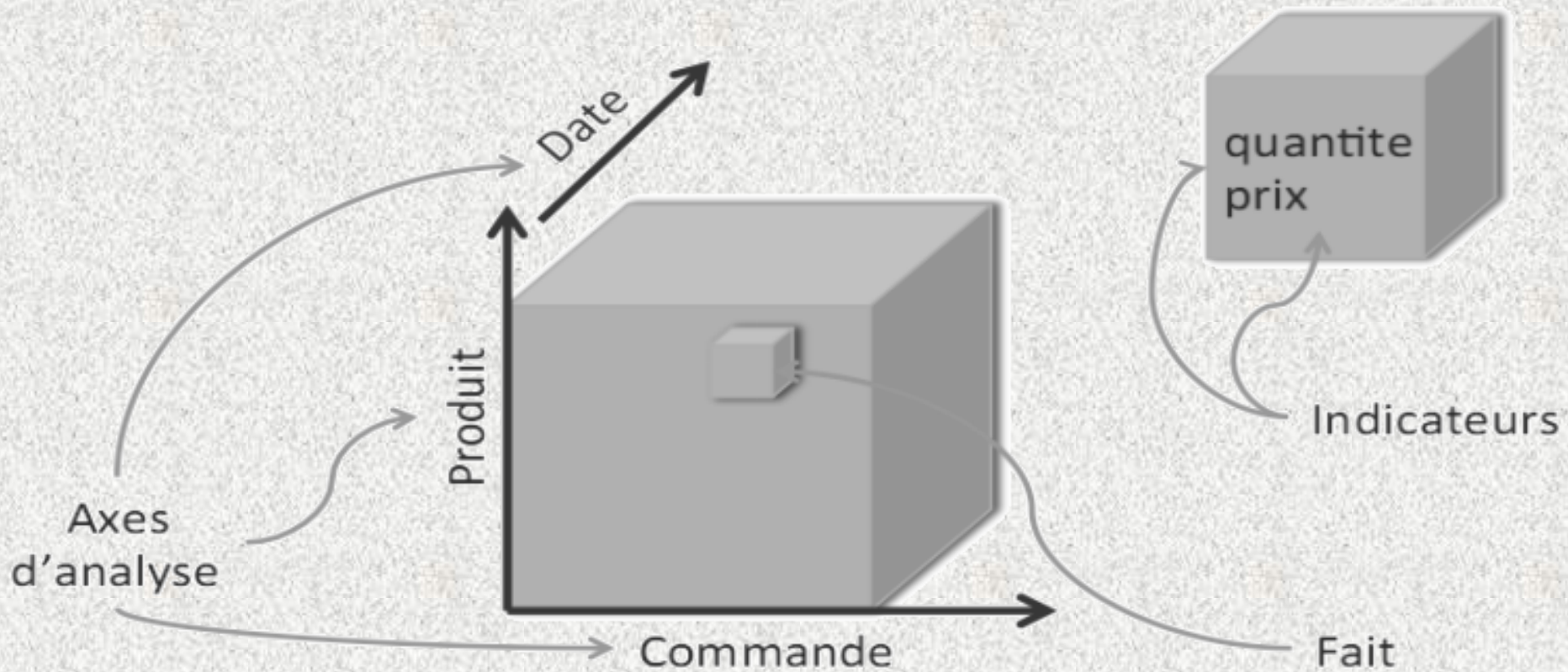


TABLE 3.5 A Comparison Between OLTP and OLAP

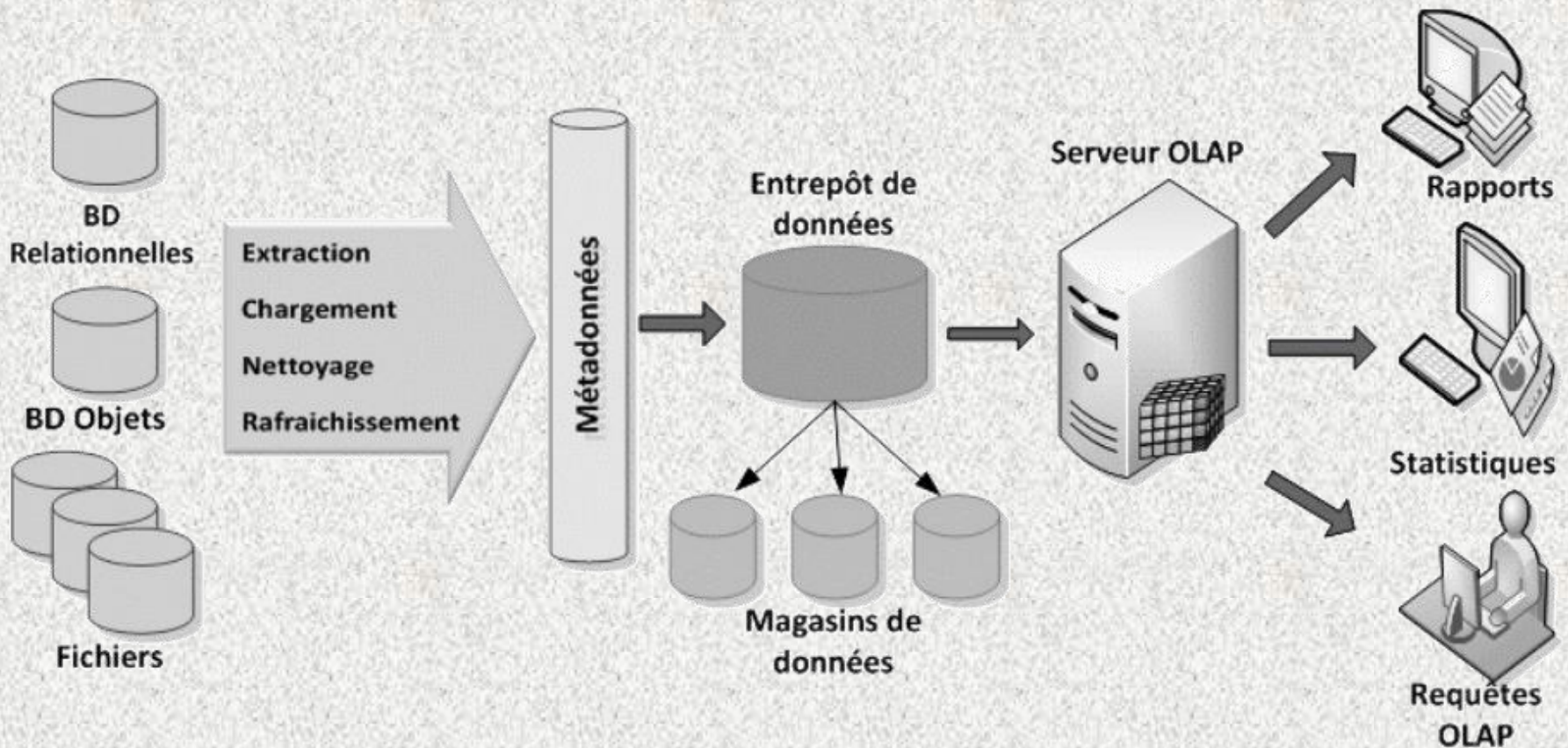
Criteria	OLTP	OLAP
Purpose	To carry out day-to-day business functions	To support decision making and provide answers to business and management queries
Data source	Transaction database (a normalized data repository primarily focused on efficiency and consistency)	Data warehouse or data mart (a nonnormalized data repository primarily focused on accuracy and completeness)
Reporting	Routine, periodic, narrowly focused reports	Ad hoc, multidimensional, broadly focused reports and queries
Resource requirements	Ordinary relational databases	Multiprocessor, large-capacity, specialized databases
Execution speed	Fast (recording of business transactions and routine reports)	Slow (resource intensive, complex, large-scale queries)

Processus du décisionnel

La Business Intelligence est désormais un outil essentiel pour les entreprises. Une BI bien ficelée doit maîtriser le processus décisionnel et atteindre ses objectifs de croissance et d'efficience.

Le processus décisionnel, appelé aussi la chaîne décisionnelle, se compose des processus, des techniques et d'outils permettant le traitement de l'information à des fins décisionnelles. Il se présente souvent en architecture à trois-tiers. Chacun de ces tiers correspond à une phase importante dans la conception d'un projet BI.

Architecture d'un système BI



Le processus décisionnel

Source d'information

La première fonction consiste à collecter et extraire les données applicatives métier de l'entreprise qui sont stockées dans des sources différentes telles que les bases de données relationnelles ou non relationnelles, les fichiers, les feuilles de calcul...etc. Ces données sont hétérogènes en origine et en nature. Elles appartiennent pour la plupart à des systèmes opérationnels, mais peuvent également inclure des documents non structurés, tels que les emails et les données reçues de fournisseurs externes. D'une manière générale, un effort important est requis pour unifier et intégrer ces données qui proviennent de différentes sources moyennant un outil d'intégration de données ETL (Extract-Transform-Load) (en français ETC, Extraction-Transformation-Chargement).

Entrepôts de données et magasins de données

À l'aide d'outils d'extraction, transformation et chargement ETL, les données traitées sont stockées dans des structures de stockage destinées à soutenir les processus des analyses BI. Ces bases de données non classiques sont appelées entrepôts de données et magasins de données.

L'entrepôt de données (base de données décisionnelle, ou encore datawarehouse) est une base de données multidimensionnelle conçue pour stocker et journaliser les données de l'entreprise en provenance de diverses sources de données. Un entrepôt de données peut être modélisé sous forme d'un datawarehouse ou d'un datamart.

En règle générale, le datawarehouse globalise toutes les données de l'entreprise, tandis que les datamarts ; qui sont alimentés depuis les données du datawarehouse (littéralement en anglais magasins de données) ; sont des sous-ensembles d'un datawarehouse souvent spécialisés vers un groupe d'utilisateurs ou un type d'affaire (marketing, achat/vente, contrôle de gestion, ...).

Notant que l'interopérabilité entre les systèmes d'entrepôt de données, les applications informatiques ou de gestion de contenu est réalisée grâce à une gestion des métadonnées.

Reporting

La restitution de données est la fonction de l'informatique décisionnelle la plus importante de point de vue des décideurs de l'entreprise. Elle permet le calcul, la sélection, la visualisation et l'analyse des données finalement interprétées de manière synthétique ou détaillée et selon plusieurs perspectives intéressant les preneurs des décisions.

Dans cette étape, les données intégrées sont finalement extraites des datawarehouses et utilisées pour alimenter des modèles mathématiques et des méthodologies d'analyse destinées à aider les décideurs. Dans un système de Business Intelligence, plusieurs applications d'aide à la décision peuvent être mises en œuvre :

- L'analyse multidimensionnelle OLAP (On-Line Analytical Processing) ;
- L'analyse exploratoire des données ;
- La visualisation des données ;
- L'analyse des séries chronologiques ;
- Les modèles d'apprentissage de fouille de données ;
- Les modèles d'optimisation.
- A noter que les techniques de fouille de données et les outils d'analyse de type OLAP (On-Line Analytical Processing) sont les plus largement utilisés.

Objectifs basés sur les indicateurs clés de performance

Le point de départ des applications BI sont les objectifs d'analyse. Ces objectifs peuvent être formulés de deux manières différentes :

La première formulation est basée sur ce que l'on appelle des indicateurs clés de performance (KPI : Key Performance Indicator). Un indicateur clé de performance KPI relie les activités de l'entreprise à ses objectifs en poursuivant le développement d'une quantité mesurable.

Objectifs analytiques

Les KPIs décrivent l'entreprise à un niveau supérieur et peuvent faire référence à certains aspects de la performance du processus métier ou à l'entreprise dans son ensemble.

Cependant, bien souvent, la BI va au-delà de la formulation des objectifs et de la mesure des performances et vise à comprendre comment les performances de l'entreprise peuvent être affectées par des facteurs dits « influents » qui déterminent et expliquent les valeurs des KPI.

La relation entre les KPIs et les facteurs d'influence peut être utilisée ultérieurement pour l'aide à la décision. Par exemple, dans le cas de la gestion de la relation client, on ne s'intéresse pas uniquement au nombre de clients qui abandonnent la relation mais on veut comprendre les raisons de leurs comportements.

Les objectifs analytiques de la BI visent à comprendre la relation entre les KPI et les facteurs d'influence. On peut distinguer trois grands types d'objectifs analytiques :

1. Les objectifs descriptifs : génèrent une description synthétique des instances du processus décisionnel selon différentes perspectives BI. Trois objectifs principaux peuvent être résumés sous cette rubrique :

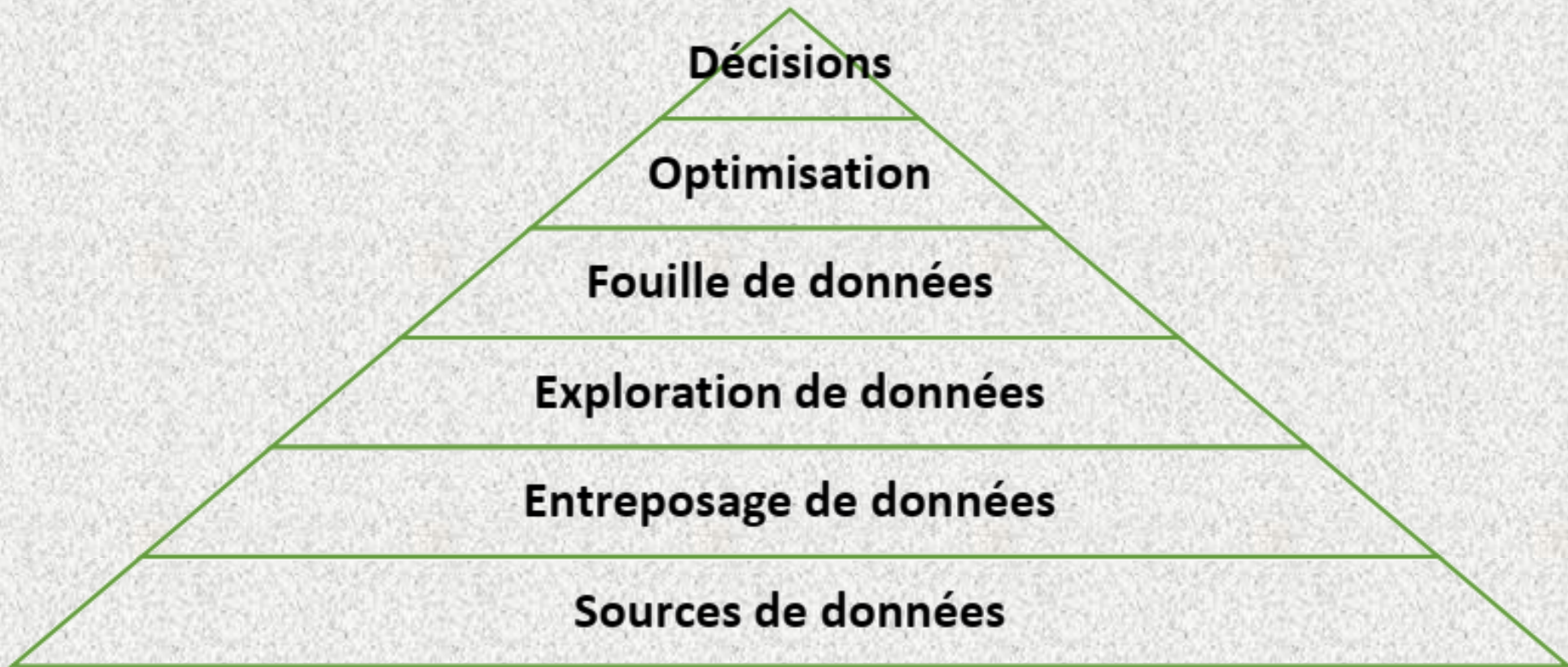
- ❑ **Reporting** : l'objectif descriptif de base est de créer des rapports. Le « rapport » est souvent un objectif supplémentaire pour atteindre d'autres objectifs analytiques.
- ❑ **Segmentation** : consiste à regrouper les informations finales du processus décisionnel selon une mesure de similarité et à trouver des instances représentatives pour ces groupes.
- ❑ **Détection des comportements intéressants** : consiste à identifier les événements et les indices qui permettent d'identifier les aspects importants des informations issues du processus décisionnel.

2. Les objectifs prédictifs : prédisent le comportement des instances du processus BI. On peut distinguer deux types différents de prédiction :

- ❑ **Régression** : consiste à trouver une fonction qui permet la prédiction d'une variable de sortie (généralement un KPI) à partir d'un certain nombre de variables d'entrée (facteurs d'influence).
- ❑ **Classification** : étant donné un ensemble de données classifiées en classes disjointes, la classification consiste à affecter à une nouvelle instance l'une de ces classes préalablement définies.

3. Les objectifs de compréhension : aident les parties prenantes de l'entreprise à comprendre les différents processus métiers de l'entreprise.

Les composants d'un système BI



Les éléments constitutifs d'un système de la Business Intelligence