

# **Découverte de la Business Intelligence**

**28/01/2021**

# Tonight's Guest Stars



**Miguel Angel Puente**  
BI Business Analyst



**Maria Giovanna Zilli**  
Regulatory Intelligence Oper. Expert



# Plan

1. Histoire de la Business Intelligence
2. BI – Overview
3. Techniques BI
4. Modes de Restitution
5. Le BA dans un projet de BI
6. Une nouvelle ère pour l'analyse de données

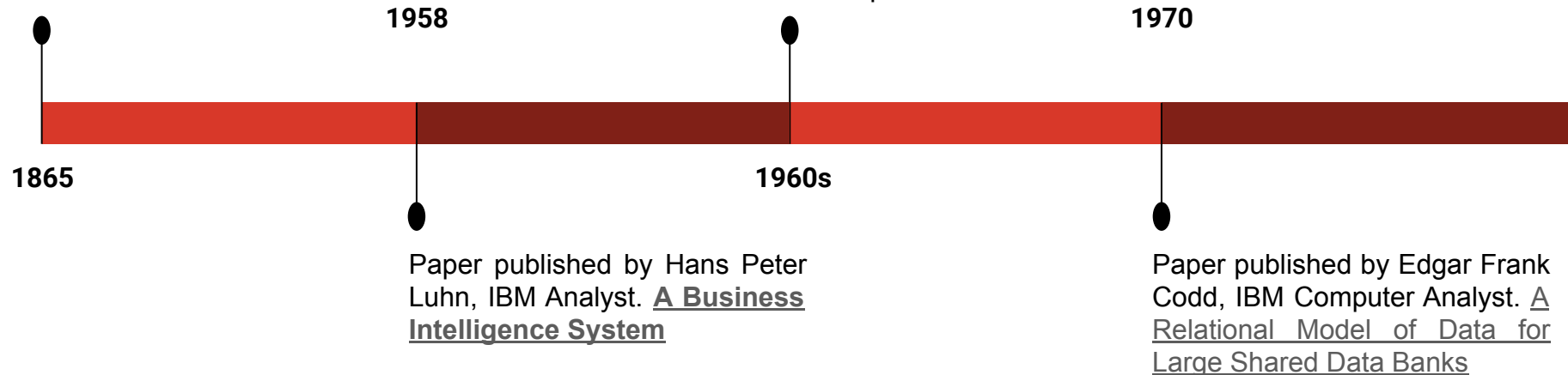


# Histoire de la Business Intelligence

# History of Business Intelligence (I)

Richard Millar Devens presented the phrase “Business Intelligence” (BI) in the *Cyclopædia of Commercial and Business Anecdotes*.

Data from multiple sources was normally stored in silos, and research was typically presented in a fragmented, disjointed report that was open to interpretation.



## History of Business Intelligence (II)



### Ralph Kimball

- Founder of the Kimball Group
- DW/BI industry's thought leader since the mid-1980s
- Dimensional approach trainer and evangelizer
- Co-Author of the Toolkit Series:

**The Data Warehouse Toolkit**

**The Data Warehouse Lifecycle Toolkit**

**The Data Warehouse ETL Toolkit**



# BI - OVERVIEW

# Quelques chiffres

**23 milliards de \$** → Investissement mondial dans le secteur BI pour 2020

**70%** → Pourcentage des projets de BI en échecs

**80%** → Niveau de précision des statistiques

**3,8** → Nombre moyen de solutions BI dans une entreprise

**610 zettaoctets** → Volume mondiale de données en 2020

\*1 zettaoctet (Zo) = 10<sup>21</sup> octets = 1 000 Eo = 1 000 000 000 000 000 000 000 octets





# Définition et Objectif

La **business intelligence (BI)** = **informatique décisionnelle**

Placer un ensemble de ressources technologiques, méthodes et outils informatiques au service de la prise de décision au sein d'une entreprise.

**Objectif** : Présenter les informations disponibles au sein des bases de données de l'entreprise à des fins de compréhension, d'analyse et de décision.

**Valeur ajoutée** : Données -> Information

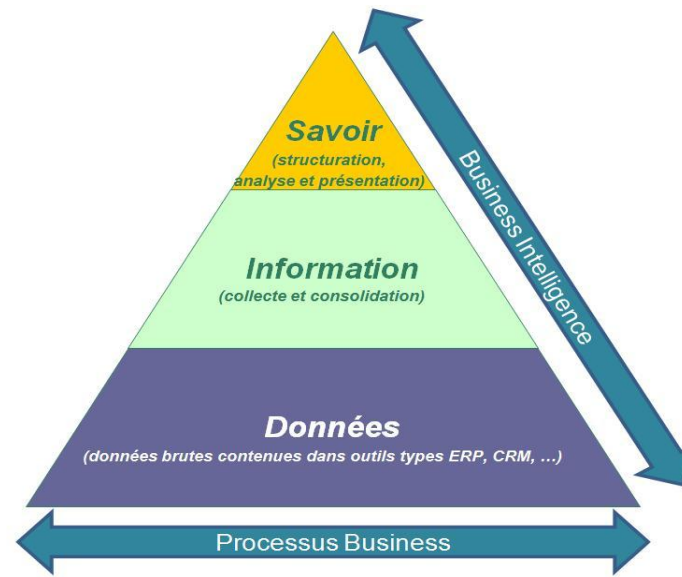
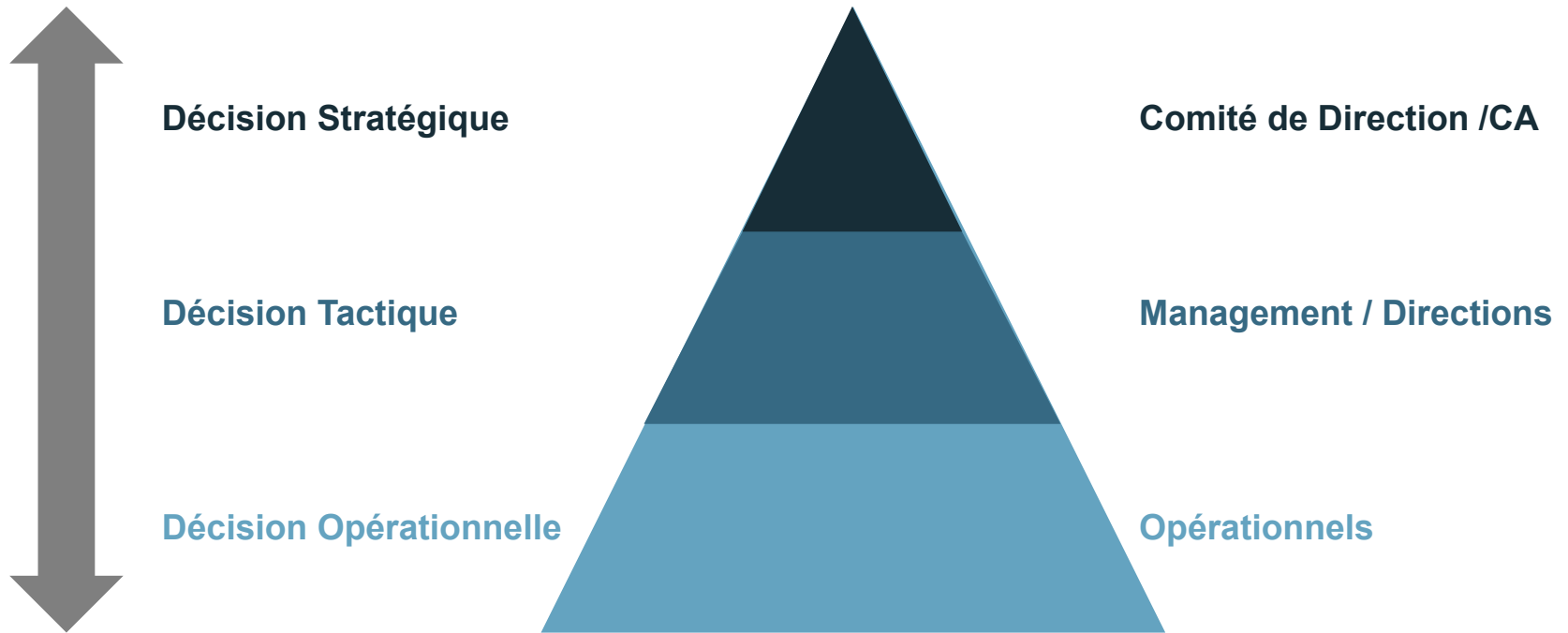


Figure 2: Flux informationnel lié au processus BI (piloter.org)

# Domaines d'application

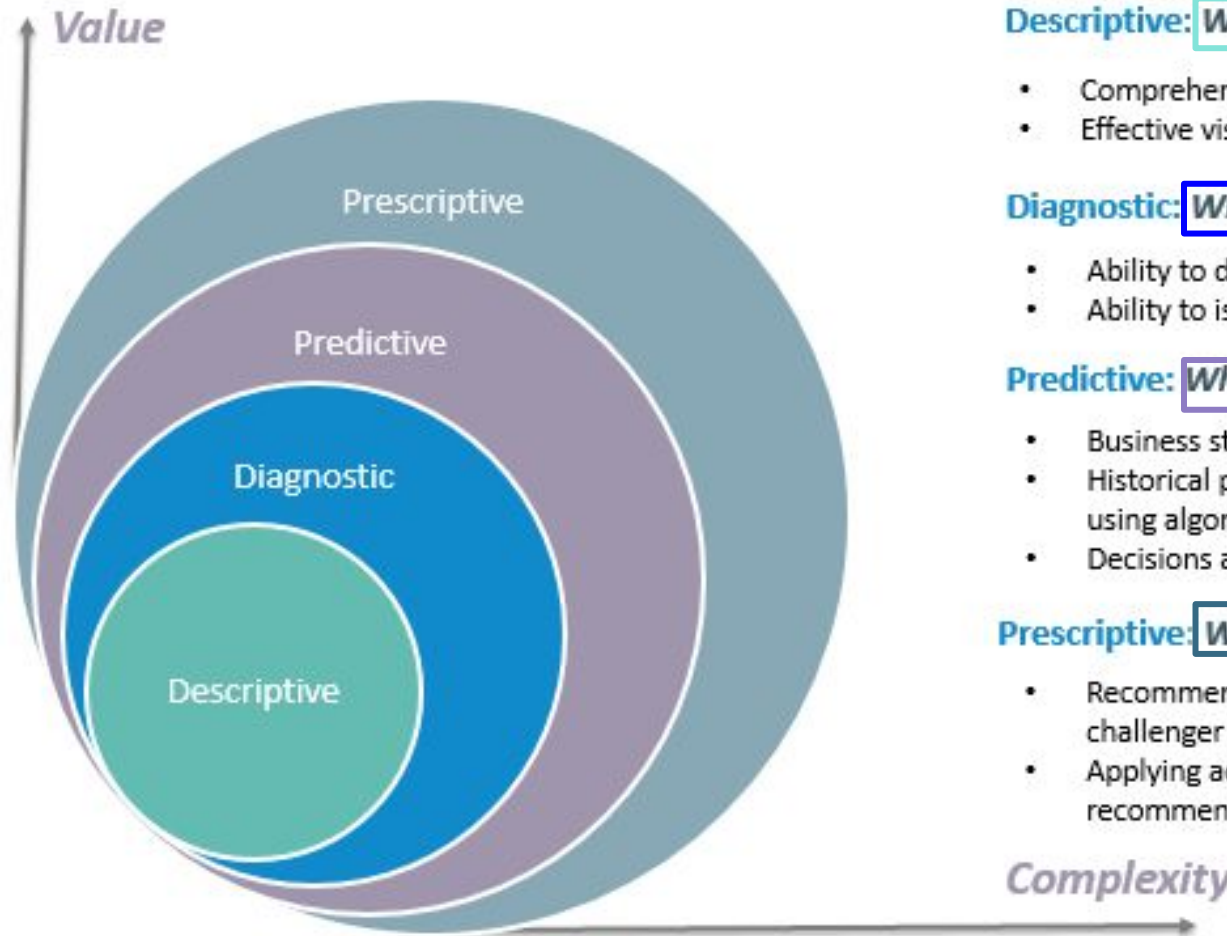


# Niveaux de décision



# Types d'analyses

## 4 types of Data Analytics



### What is the data telling you?

**Descriptive:** *What's happening in my business?*

- Comprehensive, accurate and live data
- Effective visualisation

**Diagnostic:** *Why is it happening?*

- Ability to drill down to the root-cause
- Ability to isolate all confounding information

**Predictive:** *What's likely to happen?*

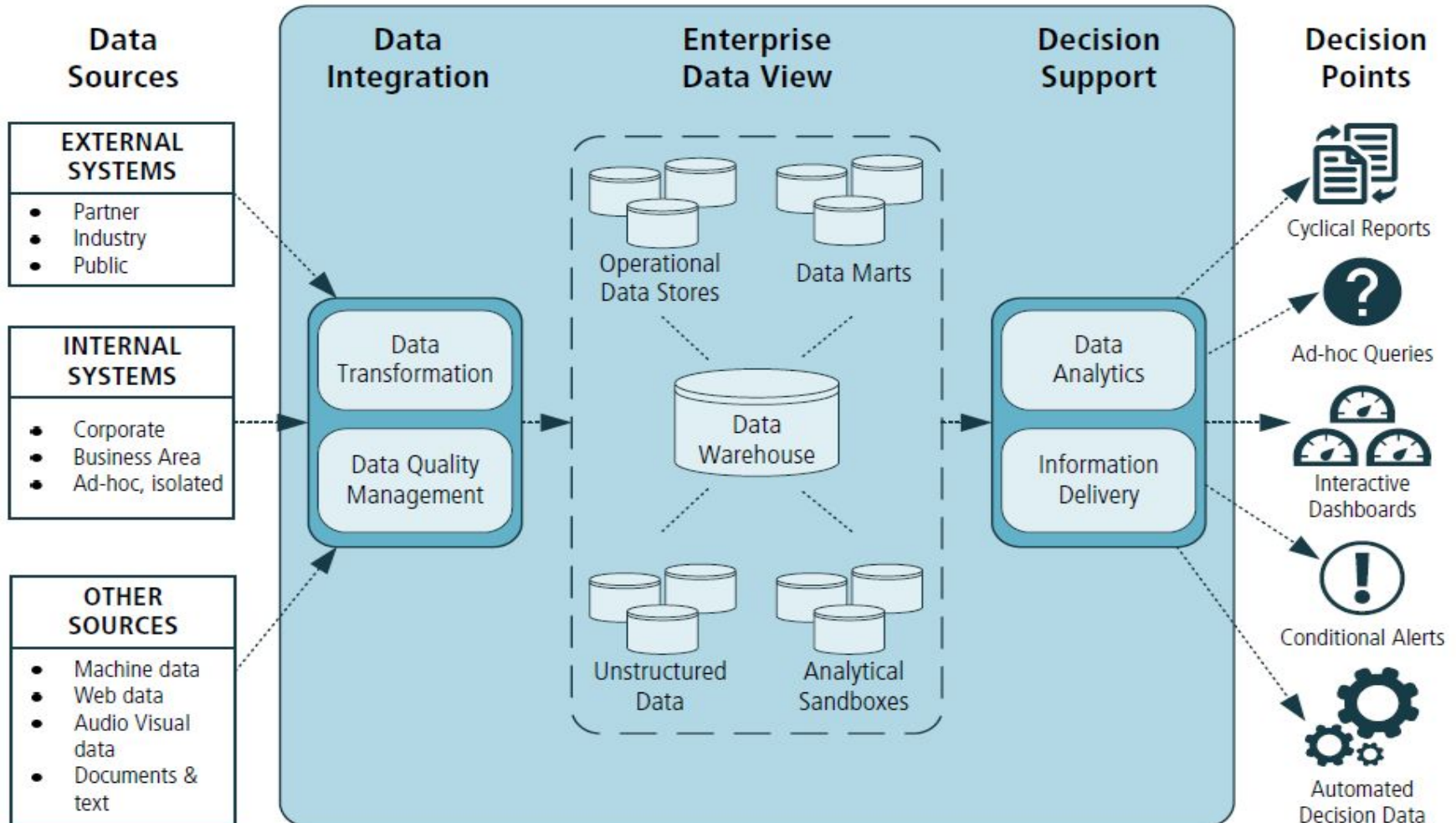
- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

**Prescriptive:** *What do I need to do?*

- Recommended actions and strategies based on champion / challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations

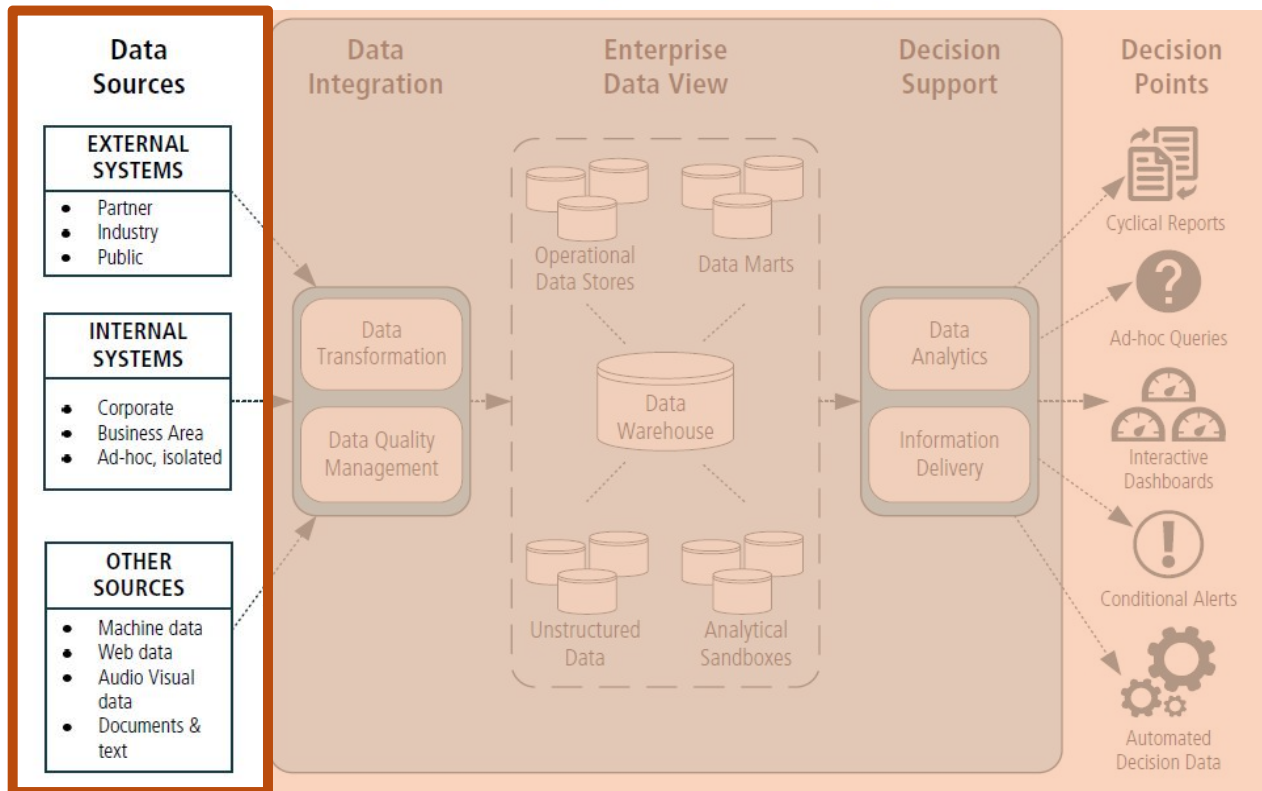
# Cadre Conceptuel

Figure 11.2.1: Business Intelligence Solution - Conceptual Framework



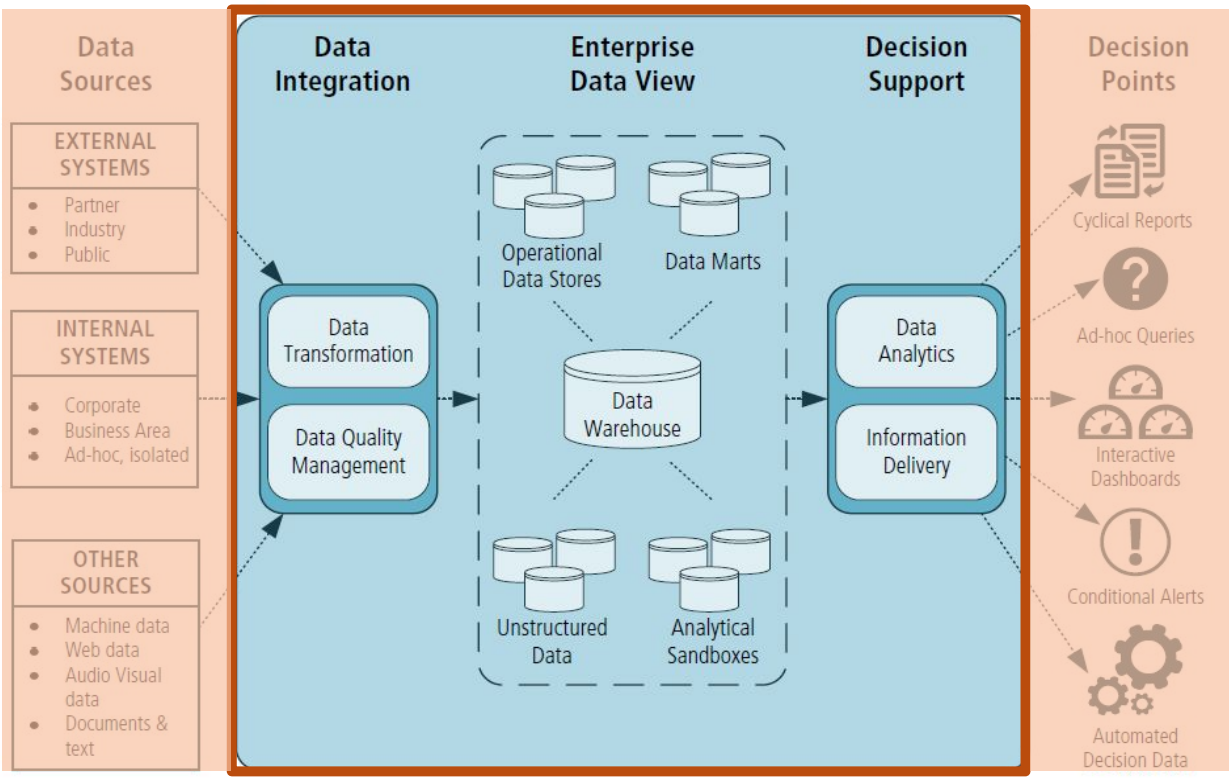
# Etape 1 – Identifier les sources

Figure 11.2.1: Business Intelligence Solution - Conceptual Framework



# Etape 2 : Collecte, Transformation et Chargement

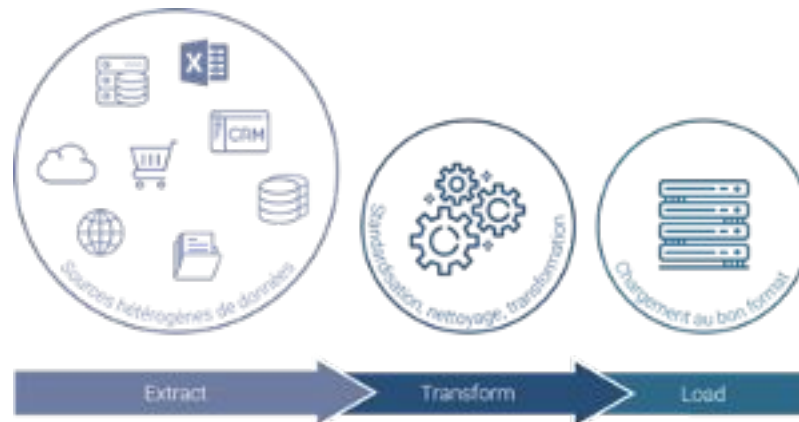
Figure 11.2.1: Business Intelligence Solution - Conceptual Framework



# Etape 2 : Collecte, Transformation et Chargement

**ETL** : (Extract, Transform, Load)

Il s'agit d'une technologie informatique permettant d'effectuer des synchronisations massives d'informations d'une base de données vers une autre. Les processus sont industrialisés et automatisés



**E**xtract Connexion aux différentes sources et collecte des données

**T**ransform Application d'ensembles de règles sur les données extraites pour répondre aux exigences techniques et métier de la base cible



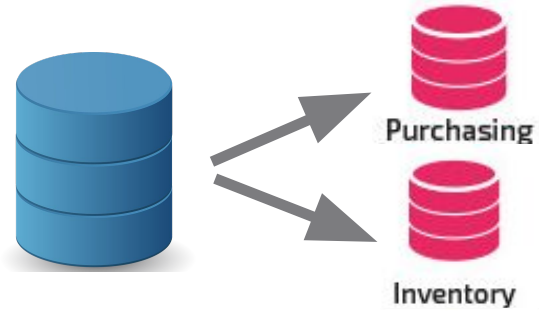
# Etape 2 : Collecte, Transformation et Chargement

**Load** Chargement des données dans les bases cibles

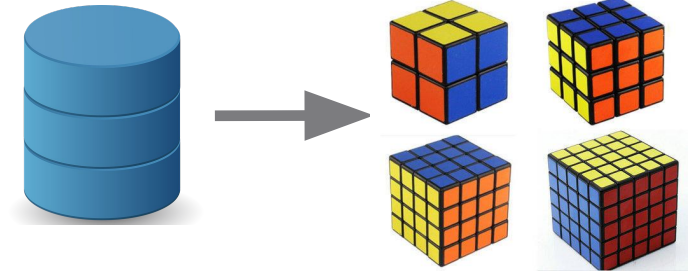
- **Data Warehouse** (Entrepôt de données)  
Couverture de l'ensemble des besoins d'analyse de l'entreprise  
Respect de règles spécifiques



- **Data Mart** (Magasin de données)  
Couverture des besoins spécifiques d'un département  
Conçu avec et pour les collaborateurs de l'entité  
Sélection de données, calculs, agrégation...



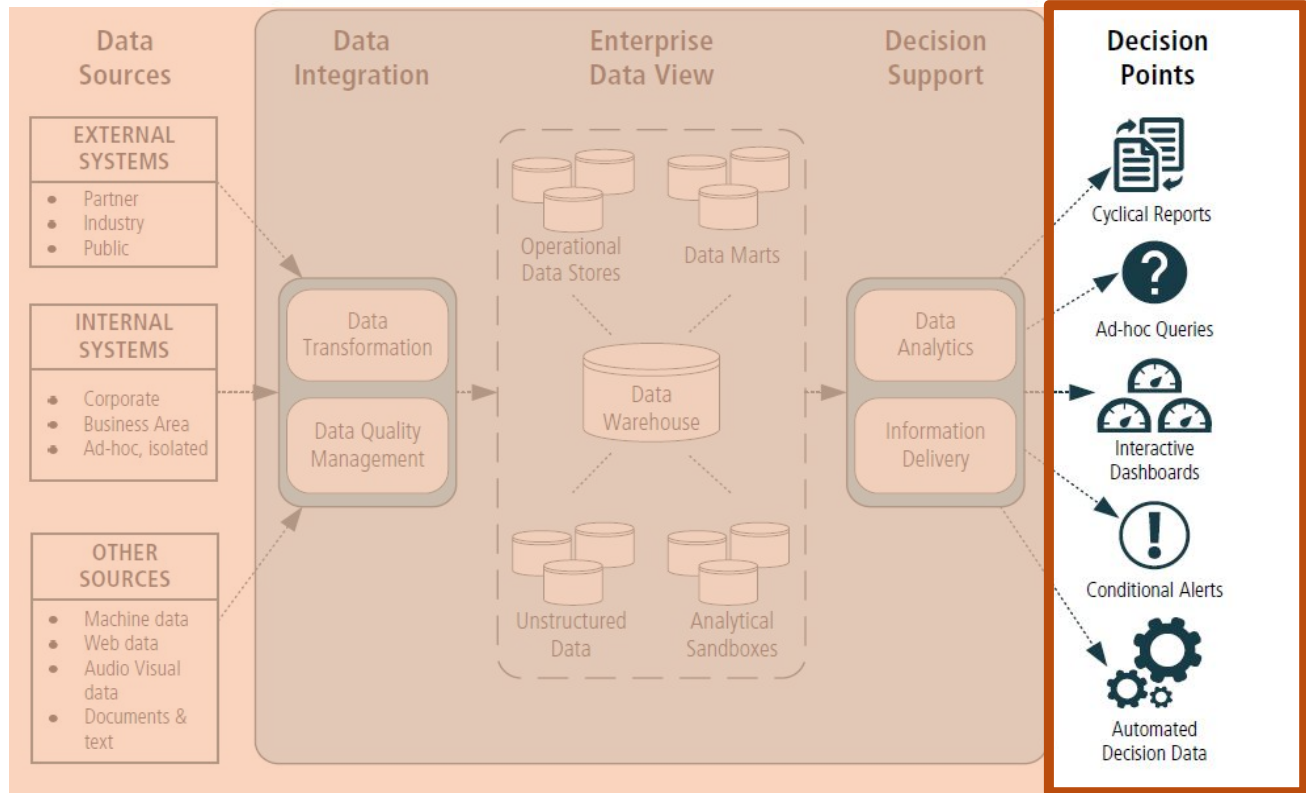
- **Cube OLAP** (Online Analytical Processing)  
Couverture de besoins spécifiques d'analyse  
Analyse multidimensionnelle



# Etape 3 : Restitution

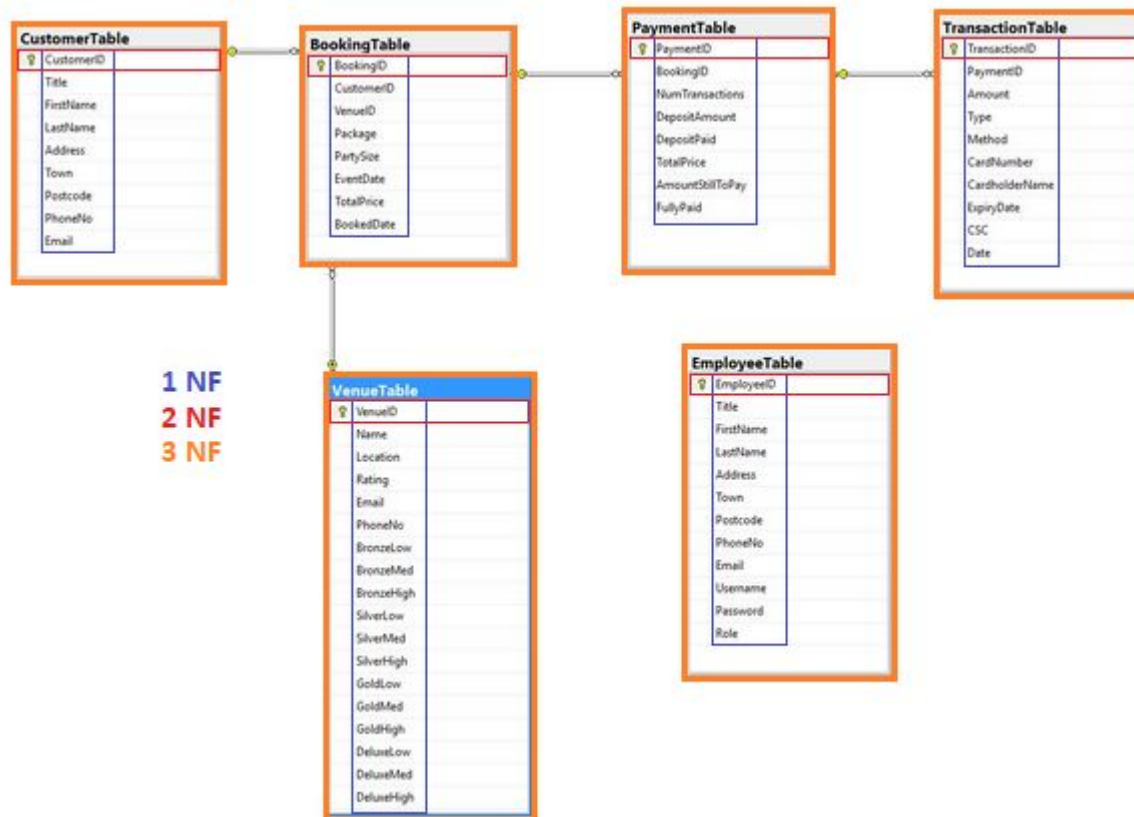
L'offre en matière d'outils de restitution est aujourd'hui arrivée à maturité  
Simplicité d'usage des applications et mobilité

Figure 11.2.1: Business Intelligence Solution - Conceptual Framework



# Techniques BI

# Data Normalization and 3 Normal Forms - Example



## Techniques BI (I)

### Kimball Techniques

#### Gather Business Requirements and Data Realities

- Understanding the needs of the business, as well as the realities of the underlying source data.
- **Requirements** are uncovered via sessions with business representatives to understand their objectives.
- Data realities are uncovered by meeting with source system experts.



## Techniques BI (II)

### Data Denormalization and Kimball Techniques

- Operational Data are stored in transactional normalized databases. Operational databases are hard to query.
- Dimensional methods and data denormalization are key to create Dimensional Data Warehouses, Kimball's core method.
- Data are stored in and queried from central / massive fact tables. Dimension tables give perspective and context.



## Techniques BI (III)

### Kimball Techniques

#### Four-Step Dimensional Design Process

The four key decisions made during the design of a dimensional model include:

1. Select the business process.
2. Declare the grain.
3. Identify the dimensions.
4. Identify the facts.



## Techniques BI (IV)

### Kimball Techniques

#### Business Processes

- **Business processes** are the operational activities performed by the organization.
- **Business process events** generate metrics that translate into facts in a fact table.
- A **Business process** determines a design target , the grain, dimensions, and facts to be declared
- Each **Business process event** corresponds to a row in the enterprise data warehouse.





## Techniques BI (V)

### Kimball Techniques

#### Dimensions for Descriptive Context

- **Dimensions** provide the context surrounding a business process event.
- **Dimension tables** contain the descriptive attributes used by BI applications.
- **Dimension tables** contain the entry points and descriptive labels that enable the BI system to be leveraged for business analysis.



## Techniques BI (VI)

### Kimball Techniques

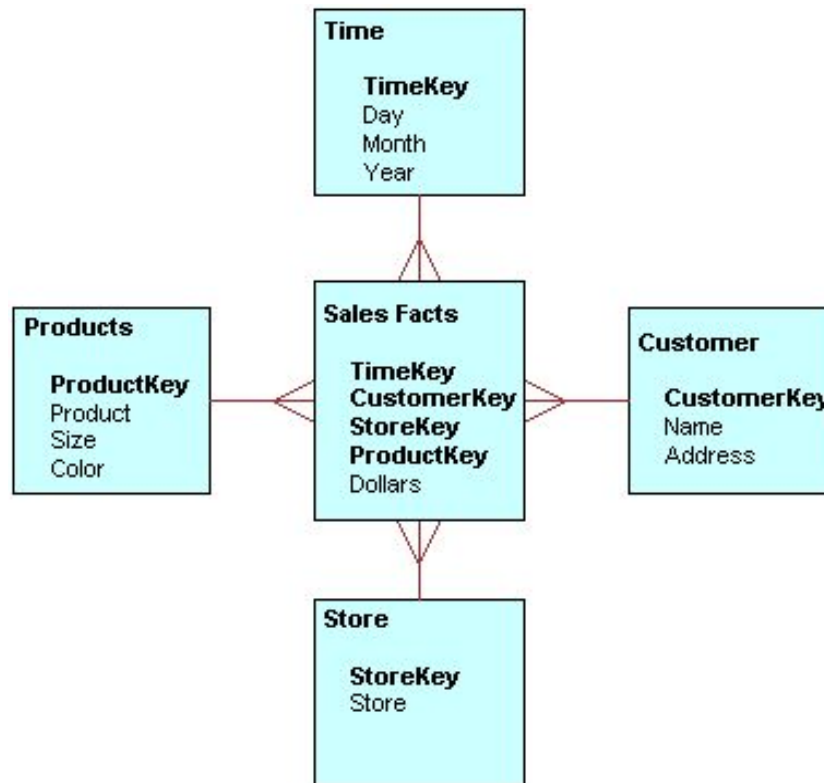
#### Facts for Measurements

- **Facts** are the measurements that result from a business process event.
- **A single fact table** row has a one-to-one relationship to a measurement event as described by the fact table's grain.
- Within a **fact table**, only facts consistent with the declared grain are allowed.



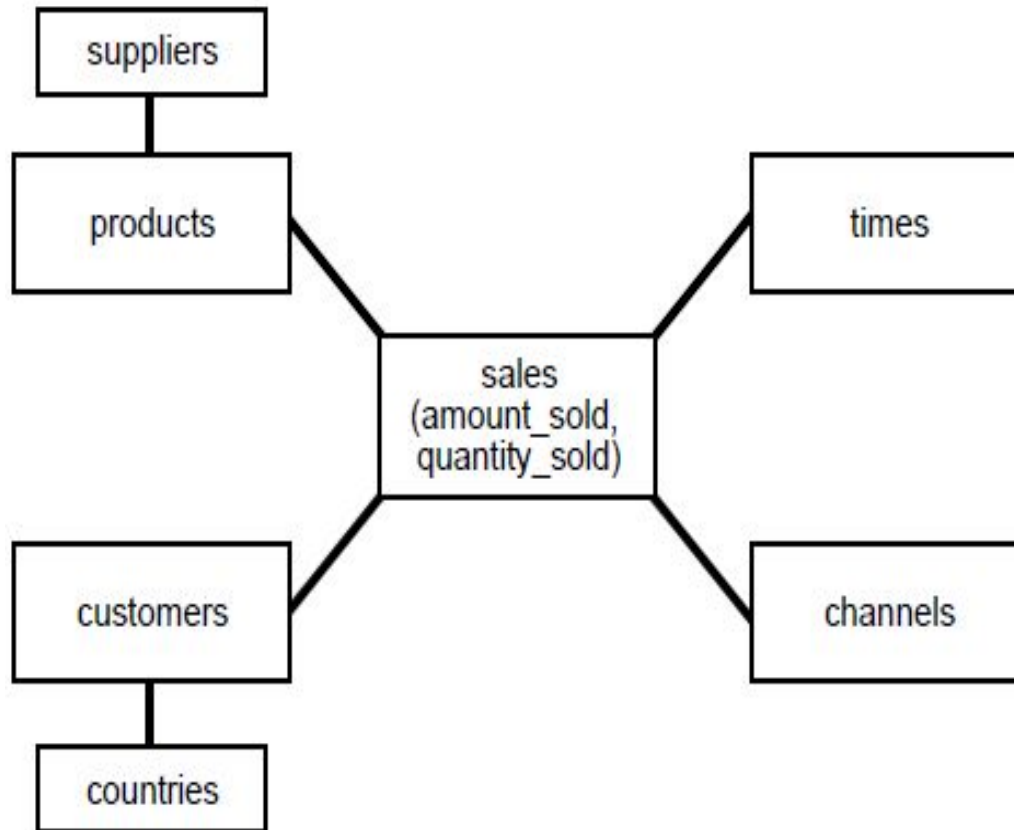
# Data Normalization and Kimball Dimensional Modeling : Star Schemas

Star schemas consist of fact tables linked to associated dimension tables via primary/foreign key relationships



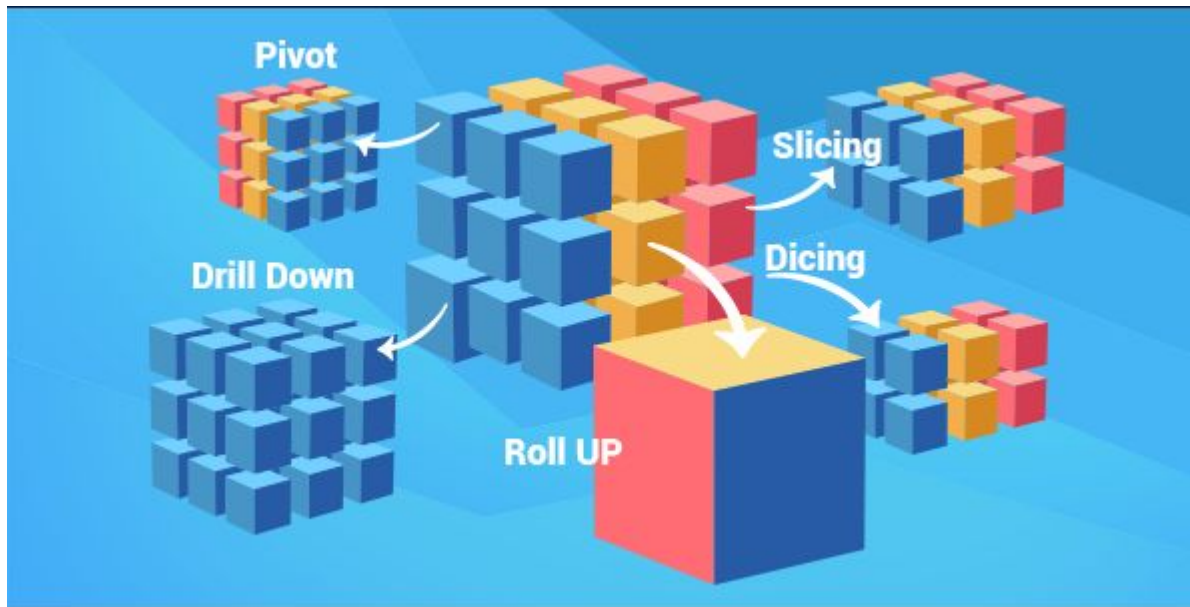
# Data Normalization and Kimball Dimensional Modeling : Snowflake Schemas

Snowflake schemas consist of fact tables linked to associated not-fully-denormalized dimension entities via primary/foreign key relationships.



## OLAP Cubes

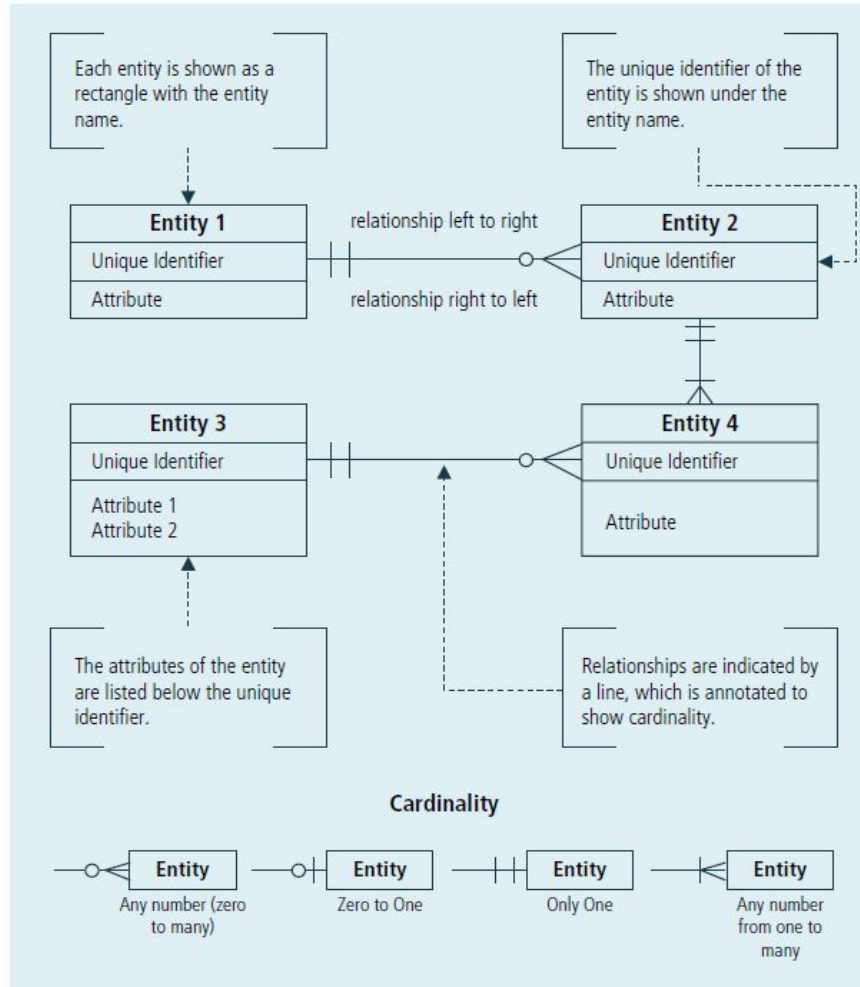
An **OLAP (OnLine Analytical Processing) cube** is a multi-dimensional array of data consisting of pre-calculated numeric facts (**measures**) categorized by dimensions organized in **hierarchies**.



# Data Modelling

A data model describes the entities, classes or data objects relevant to a domain, the attributes that are used to describe them, and the relationships among them.

Figure 10.15.1: Entity-Relationship Diagram (Crow's Foot Notation)

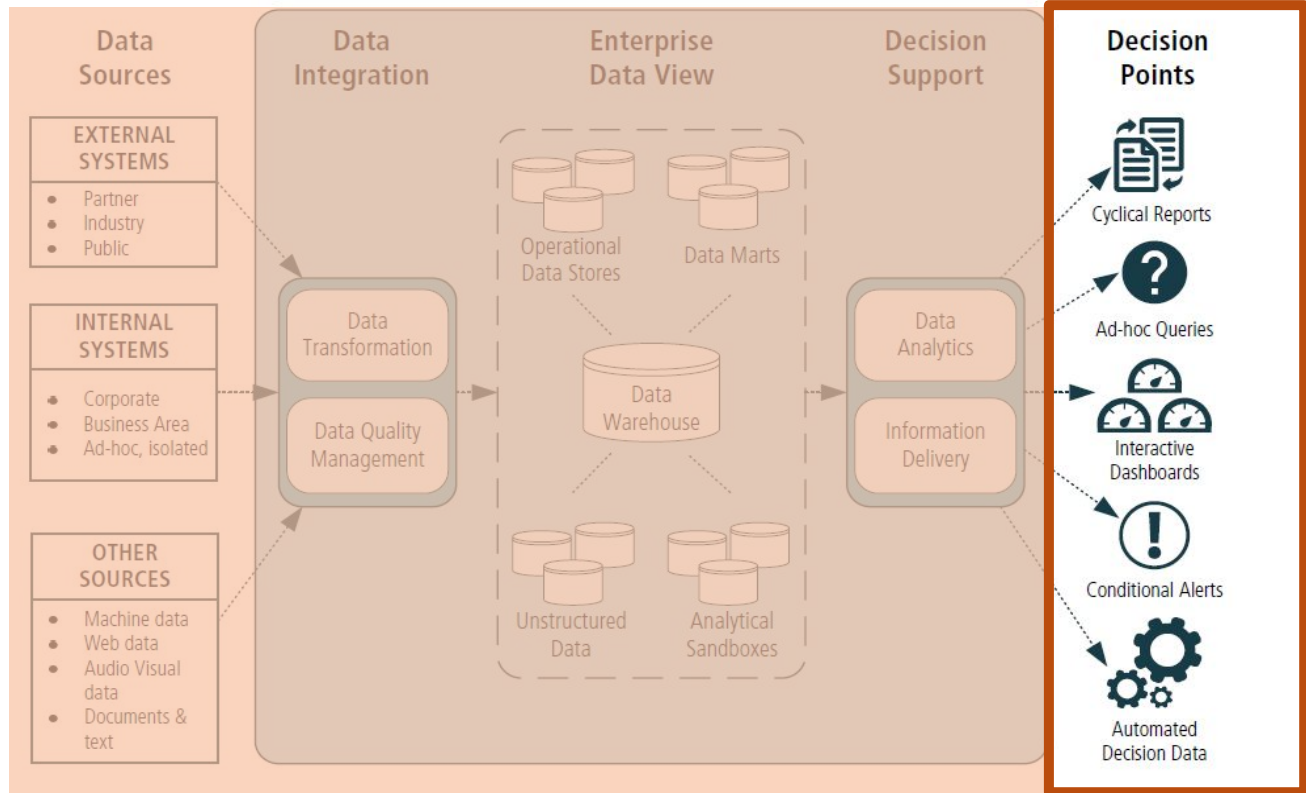


# Modes de Restitution

# Etape 3 : Restitution

L'offre en matière d'outils de restitution est aujourd'hui arrivée à maturité  
Simplicité d'usage des applications et mobilité

Figure 11.2.1: Business Intelligence Solution - Conceptual Framework





## Requête Ad-Hoc

**Requête ad hoc (« pour cela »)** → Répondre  
Question non récurrente, Résultat brut



# Report

Report (rapport) → Communiquer et Comprendre

Photo d'une situation à l'instant T. Info brute et sans relief (listes, TCs, Graph)

Live Reporting > Rg02 - P&L - Act vs Fcst

Notifications Feedback Help Nils Rasmussen

Data Warehouse

Report parameters

Auto-refresh: Off

**soiver Profit and Loss - Act vs Fcst Variance**

Corporate US  
Forecast Small Change  
Last Actual Period: Mar-2021

Account Descriptions	Jan-21 Actual	Feb-21 Actual	Mar-21 Actual	Apr-21 Forecast	May-21 Forecast	Jun-21 Forecast	Jul-21 Forecast	Aug-21 Forecast	Sep-21 Forecast	Oct-21 Forecast	Nov-21 Forecast	Dec-21 Forecast	2021 Actual + Forecast	2020 Actual	Dollar Variance	Per Vari
<b>Revenue</b>																
40010 Product Revenue	843,763	787,830	782,411	796,713	685,961	1,056,434	986,226	892,347	730,880	846,514	827,366	1,157,279	10,393,724	9,757,863	635,862	
40020 Services Revenue	450,808	382,793	417,447	367,154	380,392	506,989	490,540	559,682	540,925	456,505	498,953	605,371	5,657,560	5,300,527	357,032	
40030 Maintenance Revenue	75,856	84,414	79,133	76,916	70,032	102,926	98,778	98,099	59,664	81,419	87,278	126,410	1,040,925	1,004,036	36,890	
40040 Other Revenue	42,274	35,501	38,304	35,894	32,725	44,601	52,294	50,521	31,677	46,165	47,238	61,449	518,645	494,789	23,855	
<b>Total Revenue</b>	<b>1,412,700</b>	<b>1,290,538</b>	<b>1,317,295</b>	<b>1,276,677</b>	<b>1,169,111</b>	<b>1,710,950</b>	<b>1,627,837</b>	<b>1,600,649</b>	<b>1,363,145</b>	<b>1,430,604</b>	<b>1,460,836</b>	<b>1,950,510</b>	<b>17,610,854</b>	<b>16,557,215</b>	<b>1,053,639</b>	
<b>Cost of Goods Sold</b>																
50010 COGS - Product Sales	281,476	245,562	270,768	258,414	235,415	342,409	328,226	346,062	254,515	287,393	329,859	412,901	3,593,001	3,442,916	(150,085)	
<b>Cost of Goods Sold</b>	<b>281,476</b>	<b>245,562</b>	<b>270,768</b>	<b>258,414</b>	<b>235,415</b>	<b>342,409</b>	<b>328,226</b>	<b>346,062</b>	<b>254,515</b>	<b>287,393</b>	<b>329,859</b>	<b>412,901</b>	<b>3,593,001</b>	<b>3,442,916</b>	<b>(150,085)</b>	
<b>Profit Margin</b>	<b>1,131,224</b>	<b>1,044,976</b>	<b>1,046,527</b>	<b>1,018,263</b>	<b>933,696</b>	<b>1,368,542</b>	<b>1,299,612</b>	<b>1,254,588</b>	<b>1,108,630</b>	<b>1,143,211</b>	<b>1,130,977</b>	<b>1,537,609</b>	<b>14,017,853</b>	<b>13,114,299</b>	<b>903,554</b>	
Profit Margin %	80.1%	81.0%	79.4%	79.8%	79.9%	80.0%	79.8%	78.4%	81.3%	79.9%	77.4%	78.8%	79.6%	79.2%		
<b>Departmental Expenses</b>																
<b>Administration</b>																
60010 Full Time - Salary	80,731	72,159	74,542	64,151	65,597	90,613	87,634	86,864	65,175	84,941	80,548	112,752	965,708	974,489	8,780	
60020 Full Time - Commission	6,055	6,103	5,836	5,239	5,206	7,869	7,421	7,803	5,268	5,636	6,238	8,115	76,789	71,648	(5,140)	
60030 Full Time - Bonus	20,375	19,383	22,117	16,394	15,922	23,369	21,673	21,477	20,068	18,989	21,450	28,758	249,975	244,190	(5,784)	
61010 Part Time - Salary	7,689	8,137	8,847	6,985	6,878	9,634	8,481	9,355	8,027	7,351	8,930	10,250	100,563	98,999	(1,564)	
61040 Part Time - Overtime	1,768	1,862	2,191	1,853	1,624	2,313	2,544	2,386	1,872	2,144	2,386	2,705	25,648	24,919	(729)	
61050 Part Time - Bonus	3,001	2,939	2,952	2,568	2,199	3,294	3,855	3,762	3,013	3,096	3,549	4,574	38,802	36,252	(2,550)	
61500 Payroll Taxes - FICA	14,801	13,568	14,335	13,347	11,479	15,190	15,171	18,041	14,723	14,722	15,935	19,732	181,043	168,149	(12,894)	
61510 Payroll Taxes - FUTA	1,884	2,034	2,027	1,907	1,449	2,432	2,167	2,195	1,987	2,226	2,057	2,733	25,100	24,063	(1,037)	
61520 Payroll Taxes - SUTA	1,807	2,034	2,212	1,657	1,449	2,385	2,262	2,243	2,103	2,042	2,101	3,104	25,398	24,787	(611)	
61530 Payroll Taxes - Workers Compensation	1,826	1,823	2,191	1,764	1,433	2,575	2,167	2,315	2,065	2,185	2,320	2,876	25,540	25,084	(456)	
61540 Benefits	398	409	424	344	352	475	483	489	423	411	467	521	5,196	4,967	(229)	
EXP01 Salaries and Benefits	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
<b>Salaries and Benefits</b>	<b>140,334</b>	<b>130,452</b>	<b>137,674</b>	<b>116,210</b>	<b>113,588</b>	<b>160,148</b>	<b>153,858</b>	<b>156,931</b>	<b>124,724</b>	<b>143,742</b>	<b>145,981</b>	<b>196,119</b>	<b>1,719,762</b>	<b>1,697,547</b>	<b>(22,215)</b>	
62010 Marketing	167,612	161,207	157,275	149,686	123,551	194,580	194,115	192,819	162,086	176,417	183,860	225,505	2,088,712	1,923,948	(164,764)	
62020 Conference and seminars	1,399	1,249	1,576	1,210	1,159	1,585	1,731	1,653	1,256	1,343	1,424	1,853	17,437	17,649	212	
62030 Advertising	427	431	484	421	392	593	529	628	442	443	481	742	6,012	5,881	(131)	
62040 Gift and donations	44	50	53	44	35	62	60	54	43	50	55	68	618	581	(36)	

PL Trending

# Metrics et Key Performance Indicators

## Éléments

- **INDICATORS:**
  - mesures specific et numérique pour évaluer le progrès vers un objectif stratégique
- **METRICS:**
  - quantification de le progrès dans un instant T - avec un target, les metrics devient un objectif à atteindre dans un temps specific
- **STRUCTURE:**
  - basé sur les procédures de collection des données, analyse, reporting pour assurer la gestion des KPIs
- **REPORTING:**
  - définir la fréquence, assurer la validité et la confiance des KPIs pour les utilisateurs



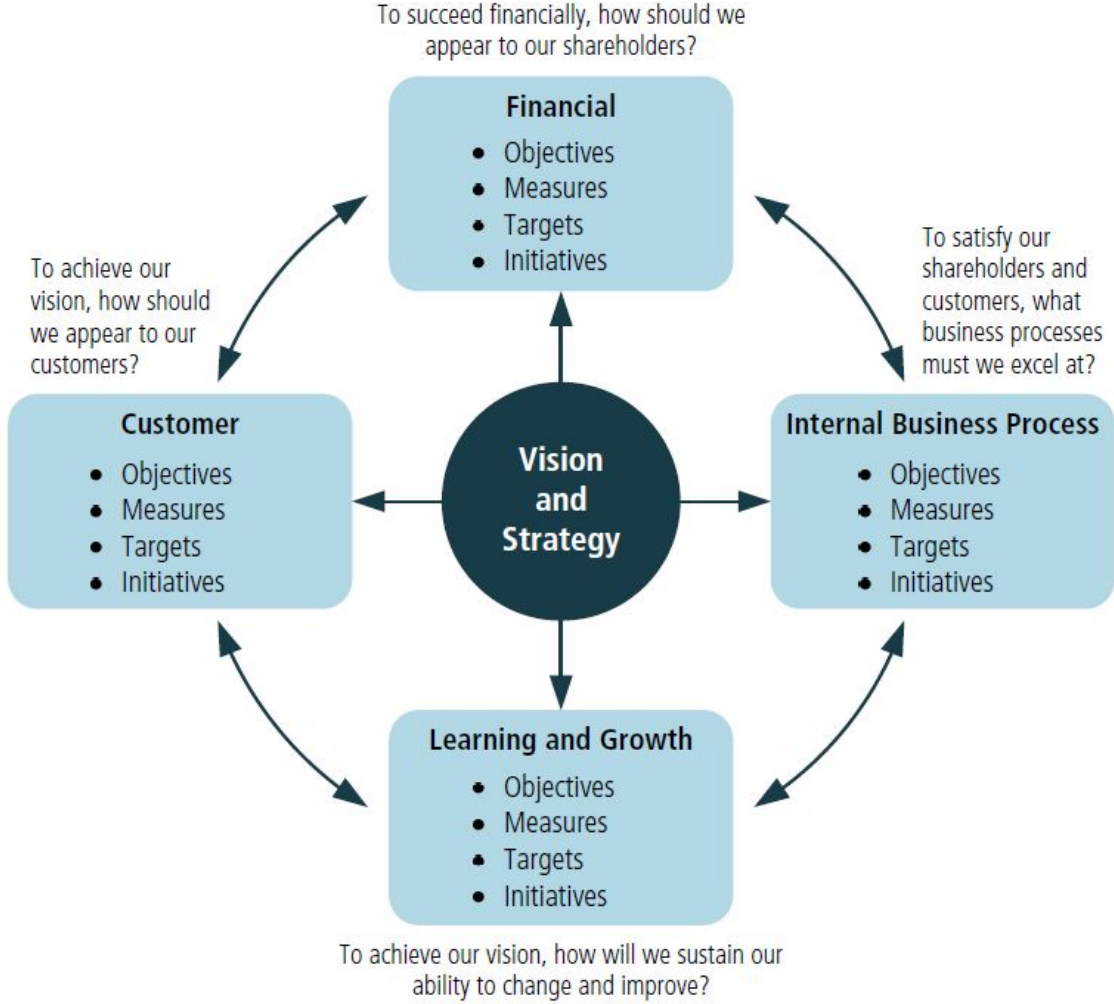
# Dashboard

**Dashboard (Tableau de bord)** → Analyser et Piloter une activité  
Présentation de KPIs pour le suivi d'un service, d'une activité - Rapidité d'analyse -



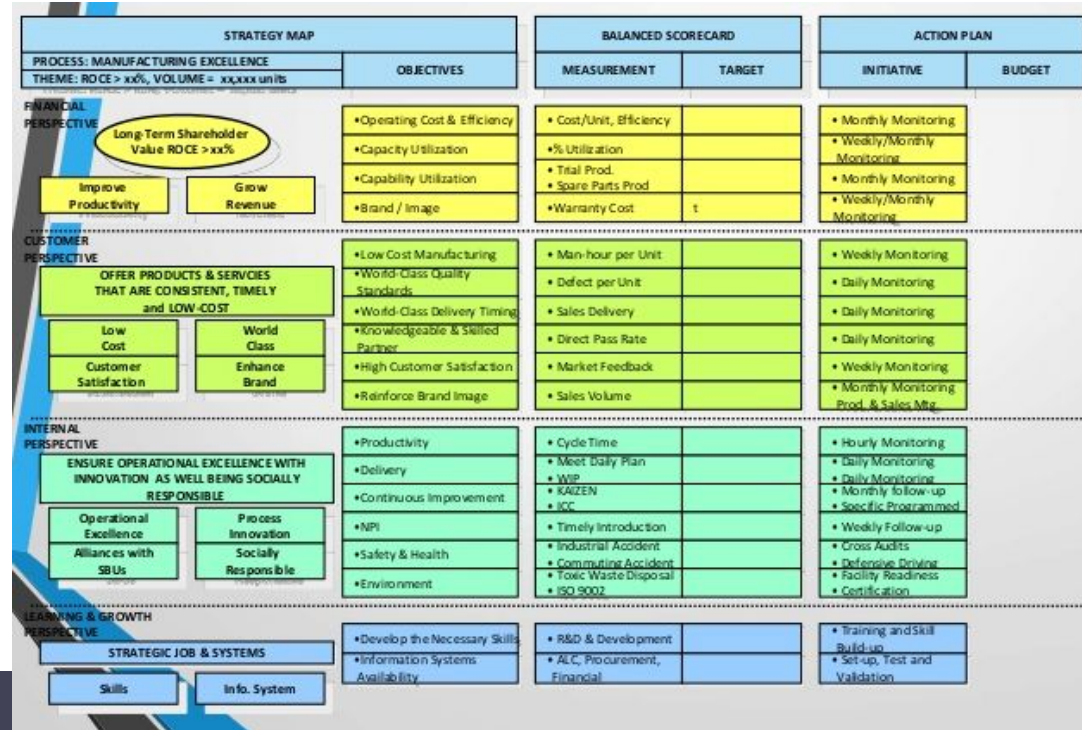
# Balanced Scorecard

Figure 10.3.1: Balanced Scorecard



# Balanced Scorecard

**Balanced Scorecard**  
 → Piloter l'entreprise  
 Présentation de KPIs pour  
 évaluer les performances de  
 l'organisation.



## Balanced Scorecard Indicators Dashboard



# Data Mining

**Data Mining (Forage de données)** → Découvrir des opportunités  
Il s'agit de trouver des corrélations informelles entre les données



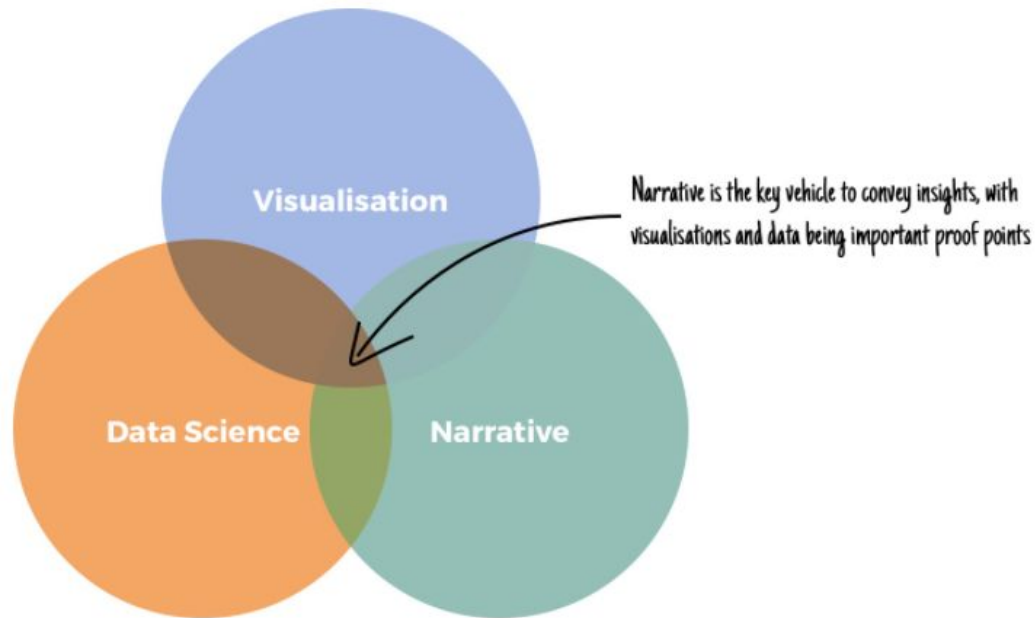
<https://theappsolutions.com/blog/development/data-mining-guide/>

## Techniques:

1. Descriptif: grouper les données pour mieux connaître la situation
2. Diagnostique: créer des segmentations pour évaluer des corrélations existant
3. Prédicatif: e.g. régression statistique pour trouver des tendances futures

# Storytelling

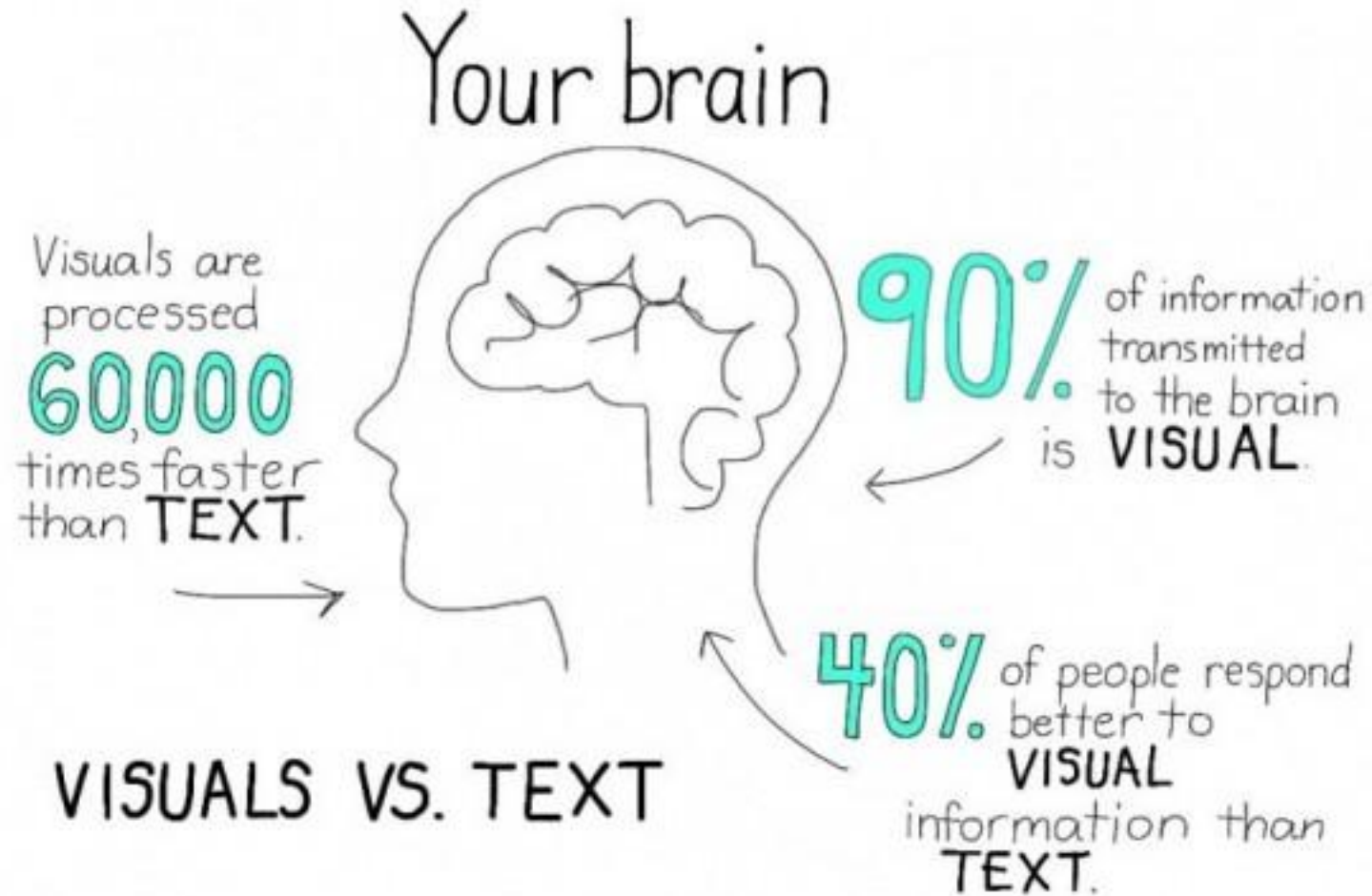
Storytelling (Mise en récit) → Communiquer et Scénariser les résultats



Hans Rosling VIDEO (5.12)



# Data Visualisation



[annemacoll.com](http://annemacoll.com)

# Data Visualization

The best way to learn sometimes is by playing a funny game!

So, please use this link <https://app.quizwhizzer.com/> to start playing :)

And now, two nice examples to compare with what we have just learned:

<https://public.tableau.com/fr-fr/gallery/lighthouses-england-and-wales>

<https://public.tableau.com/fr-fr/gallery/covid-19-white-house>

Finally, some “theory”:

<https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

<https://thenextweb.com/dd/2015/05/15/7-most-common-data-visualization-mistakes/>



**LET'S PLAY**



International Institute  
of Business Analysis™

# Le BA dans un projet de BI

# Le BA consommateur de BI



## Pour les projets de réingénierie des processus ou implémentation de solution IT

- ❑ Connaissance de la stratégie de l'organisation
- ❑ Evaluation de la situation actuelle
- ❑ Validation de la solution apportée



# Le BA acteur de projet BI



## Objectif : création de valeur

Fournir des informations fiables et utilisables, au bon moment, et aux bonnes personnes

## Rôles du BA :

- Définition des exigences et évaluation des solutions potentielles
- Conception et/ou création de représentations
- Modélisation des données

## Les résultats de l'activité du BA :

- Couverture du processus business
- Modèles décisionnels
- Modèles de données sources et cibles et dictionnaires de données
- Evaluation de la qualité des données
- Règles de transformation
- Spécification des extrants
- Architecture de la solution
- Cartographie des droits des utilisateurs



# Le BA acteur de projet BI



## Deux approches :

- Axée sur l'offre - Objectifs techniques

*Amélioration des systèmes de présentation existants (Coût donné -> Quelle Valeur ?)*

Intégrer des bases de données existantes à l'architecture de la solution

Remplacer ou améliorer les dérivables existants

Explorer de nouvelles perspectives pouvant être obtenus des données existantes

- Axée sur la demande – Objectifs Business

*Amélioration du processus de décision (Valeur donnée -> Quel coût)*

Analyse de la faisabilité d'une nouvelle demande des dérivables

## Les compétences spécifiques :

- Base de données. Terminologie et règles
- Définition des KPIs
- Modélisation des décisions
- Techniques d'analyses de données
- Concept et architecture du DWH
- Modélisation de base de données
- Pratique ETL
- Pratique d'outils de restitution



# Projet BI – Les erreurs à éviter



## Les raisons des échecs de ces projets BI (piloter.org)

- Commencer par choisir l'outil
- Ne pas solliciter les utilisateurs
- Ne pas étudier précisément les besoins, Le progiciel configurable !
- Ne pas impliquer la direction
- Ne pas suivre une méthode
- Ne pas s'intéresser à la stratégie de l'entreprise
- Ne pas réfléchir sur le choix des indicateurs
- Ne pas nettoyer les données
- Mettre toutes les informations dans un Data Warehouse. Les utilisateurs trouveront leur bonheur !
- Ne pas définir précisément les rôles et les responsabilités



Intégrer un **Business Analyst** dans l'équipe



**Une nouvelle ère  
pour l'analyse de données**

# Business Intelligence. Present & Future (I)

## Big Data

- **Big data** is a field that treats ways to deal with data sets that are too large or complex to be dealt with by traditional data-processing application software.
- **Big Data sources:** Unstructured data like Internet of things devices, software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks.



- By **2025**, IDC predicts there will be **163 zettabytes** of data.
- **A transaction (centralized)**. This approach moves all management and updating of master data to the MDM hub, which publishes data changes to each source system.



# Business Intelligence. Present & Future (II)

## Hadoop as large scale data processing tool

- Relational database management systems and software packages used to visualize data often have difficulty handling big data.
- **Apache Hadoop** is an open source software framework for storage and large scale processing of data-sets on clusters of commodity hardware.
- All the modules in **Hadoop** are designed with a fundamental assumption that hardware failures are common and thus should be automatically handled in software by the framework
- **HDFS (Hadoop Distributed File System)** is a distributed file system that handles large data sets running on commodity hardware. It is used to scale a single Apache Hadoop cluster to thousands of nodes. HDFS, MapReduce and YARN are the major components of Apache Hadoop.



# Business Intelligence. Present & Future (III)

## Artificial Intelligence

- Massive data volumes and data processing tools ease AI systems learning. Two main AI system development approaches.
- **Neural Networks** for pattern recognition - supervised and unsupervised learning.  
In supervised learning the training data has the answer key with it so the model is trained with the correct answer
- **Reinforcement Learning.** There is no answer but the reinforcement agent decides what to do to perform the given task. In the absence of a training dataset, it is bound to learn from its experience.



**MERCI!**