



INF 735

Entrepôt et Forage de Données

Préface

par Robert J. Laurin



UNIVERSITÉ DE
SHERBROOKE
Faculté des sciences

Centre de formation en technologies de l'information

Conciergerie et règles de la classe

- Adresses courriel : robert.j.laurin@usherbrooke.ca
→ *N'attendez pas une réponse le jour même, ni de confirmation de réception des travaux.*
- Heures de cours *18h45 à 21h45 (ne vous attendez pas à partir tôt)*
- Pauses *de 20 minutes vers 20h00*
- Ordinateurs ...*pour prendre des notes ou voir les notes sur Moodle*
- Cellulaires *silencieux au cour (et rangés aux examens)*
- Enregistrements *audio permis*
- Laboratoires *là où les pneus touchent l'asphalte – avec Michel Hébert...*
- Ma disponibilité: *après les cours,,*
- Livre: *Complément aux présentations et vice versa. Vous rend ra meilleur!*
- *Examens – 2 feuilles à l'intra, 4 au final*

Règles en classe

- Posez vos questions immédiatement ! (main levée SVP)
- Si je ne vois pas votre main levée, interrompez!
- Que vous soyez présent en classe pour ne pas écouter est insultant et démotivant pour moi

- Participez !

(ma mémoire des noms/visages/lieux)

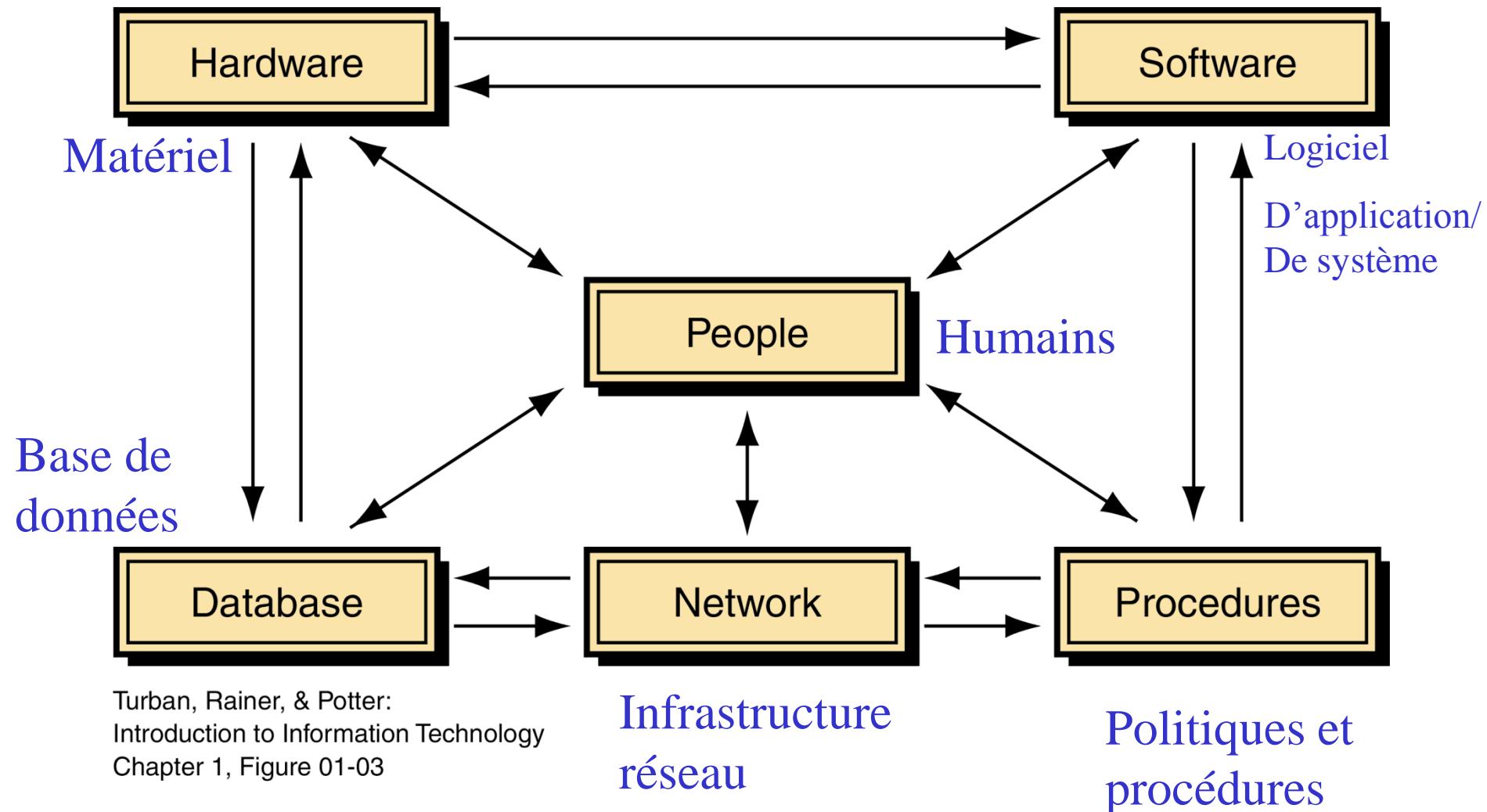
- Ayez du plaisir !



- Apprenez!

Composantes d'un système d'information

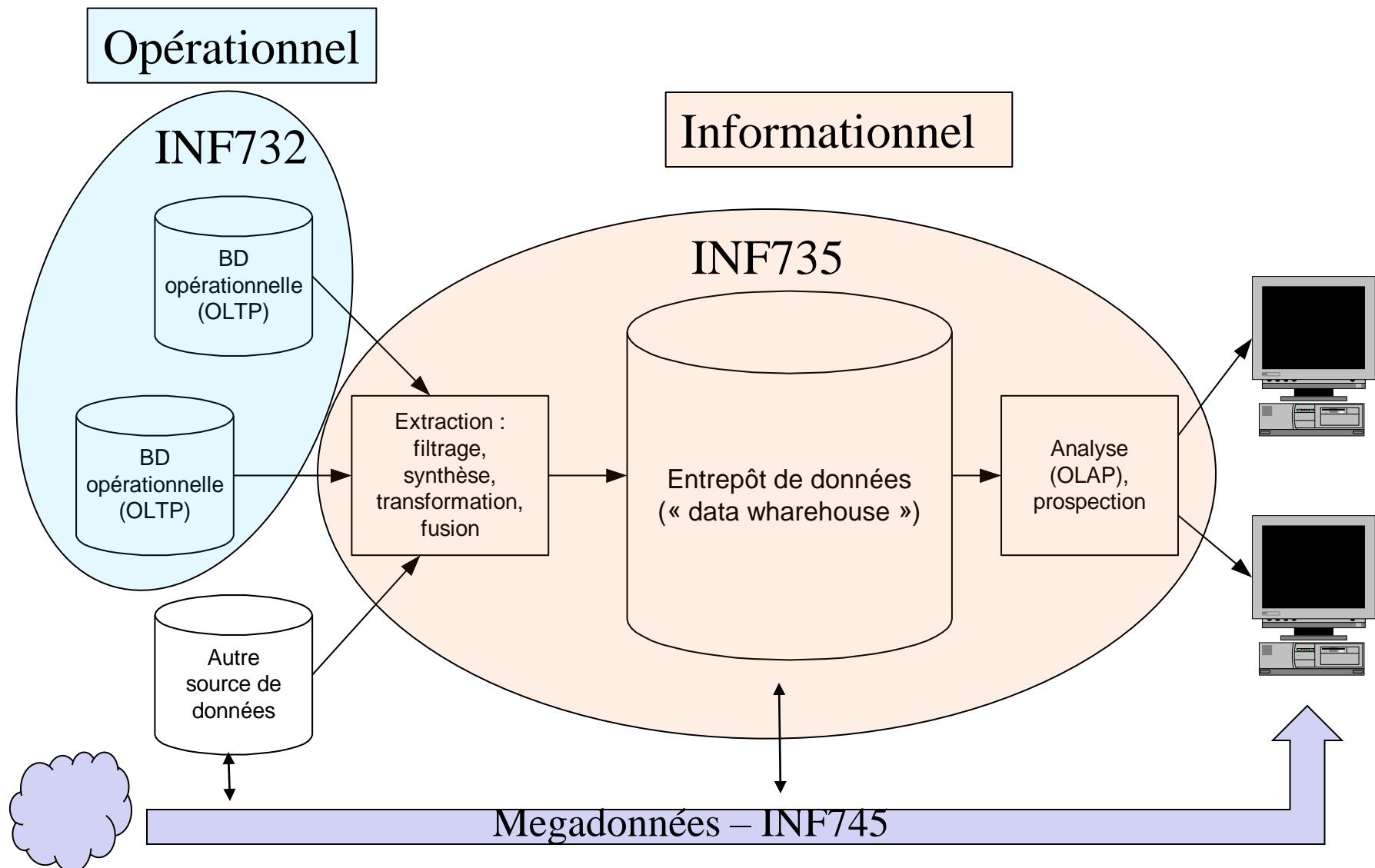
(INF 730)



Turban, Rainer, & Potter:
Introduction to Information Technology
Chapter 1, Figure 01-03

- Le but d'un système d'information est généralement de traiter des données pour obtenir de l'information ou de la connaissance
- Données: une description élémentaire, souvent codée, d'une chose, d'une transaction d'affaires, d'un événement etc.. Conservée et classée sous forme numérique, alphanumérique, d'image, de son etc...
- Information: des données organisées qui permettent à un interlocuteur de décoder la signification d'un phénomène/état décrit par les données. C'est par l'information qu'elles contiennent que les données prennent toute leur valeur, mais elle doivent être traitées pour la révéler
- Connaissance: des données, de l'information, organisées, classées de sorte qu'elles permettent à un interlocuteur une compréhension, une interprétation, un apprentissage renforçant son expérience face à une situation

Entrepôt de données et Cubes



Blocs d'activités pédagogiques obligatoires (30 crédits)

→ Technique

→ Gestion / management

- Processus TI



INF 735

Entrepôt et Forage de Données

Bloc 1
Introduction
(Besoin, concepts et définitions)
par Robert J. Laurin



UNIVERSITÉ DE
SHERBROOKE
Faculté des sciences

Centre de formation en technologies de l'information

Plan du cours – Les blocs

(Bloc1)

Introduction: Le besoin, concepts et définitions

(Bloc 4)

ETC:
Acquisition de
données

(Bloc 2)

Modélisation
(Entrepôt)

(Bloc 3)

Outils de
présentation,
OLAP et Forage

(Bloc 5) Architecture et Méta données

(Bloc 6) Définition des besoins et gestion de projet

(Bloc 7) Techniques de réalisation et opération



UNIVERSITÉ DE
SHERBROOKE
Faculté des sciences

Centre de formation en technologies de l'information

INF735 Entrepôt et forage de données, Bloc 1
Préparé par Robert J. Laurin

- Suggéré:
Data Warehousing Fundamentals, A Comprehensive Guide for IT Professionals,
Paulraj Ponniah
 - Chapitres 1 à 3
- Annexes:
 - How companies learn your secrets
 - la monnaie Canadian Tire en voie de disparition
 - Labatt pours data into holistic management system
 - CRM and the Missing Link
 - Gartner identifies the ‘Fatal Flaws’ of business intelligence and advises organisations on how to avoid them
 - Transport giant drives data warehouse project
 - Case Study: DakotaCare Delivers High Return on Investment With a Cost-Effective BI Program
 - Data Mining – If Only It Really Were about Beer and Diapers
 - SAP Bi Strategy – From ETL thru Data Warehouse to the Front End Layer

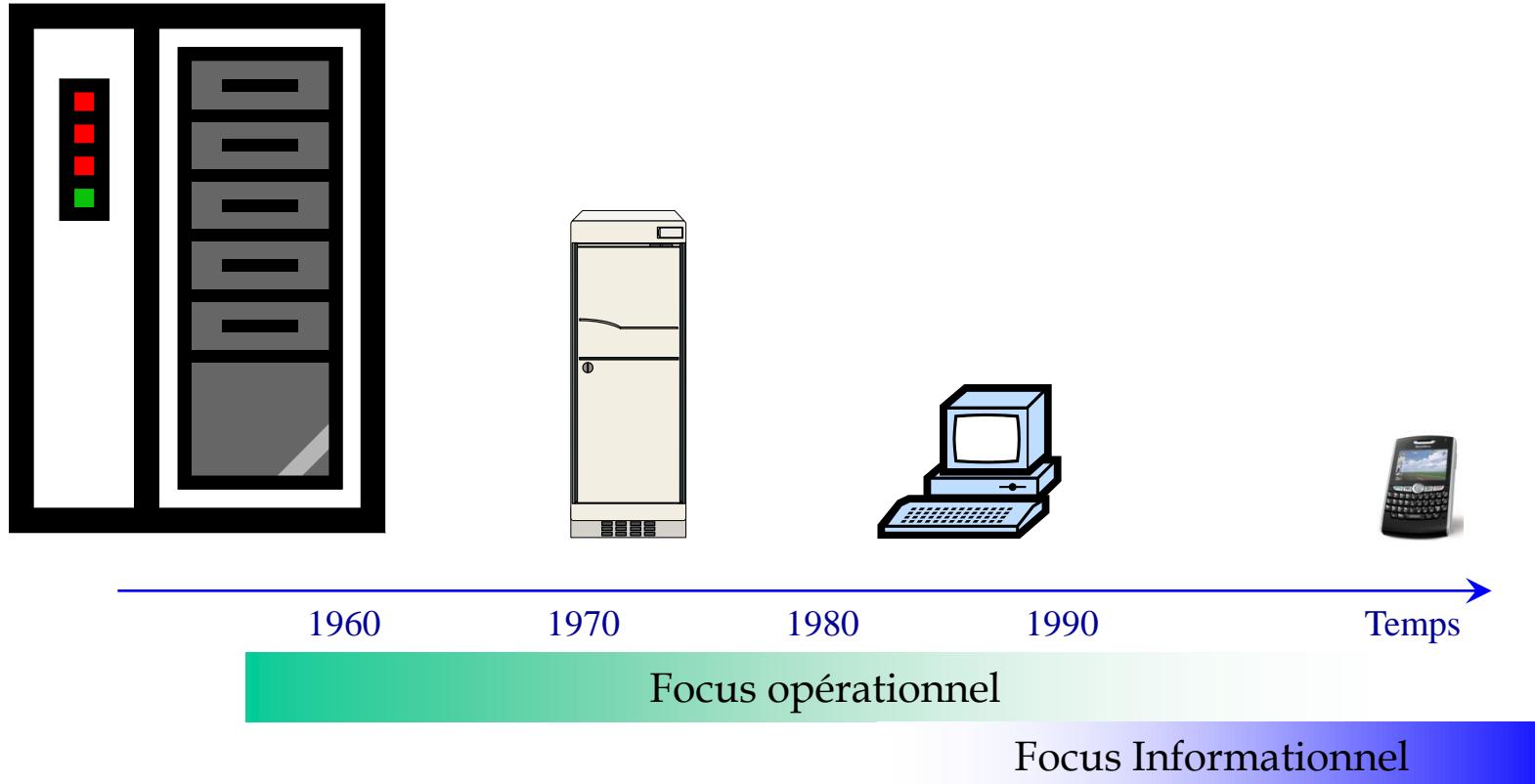
Concepts Clés du Bloc

- L'entreprise a un urgent besoin d'information stratégique
- L'entrepôt de données est un outil Informationnel stratégique à la base de la création de la connaissance
- L'implantation d'un entrepôt s'avère un projet organisationnel important
- Ce n'est pas un produit mais un environnement technologique informationnel

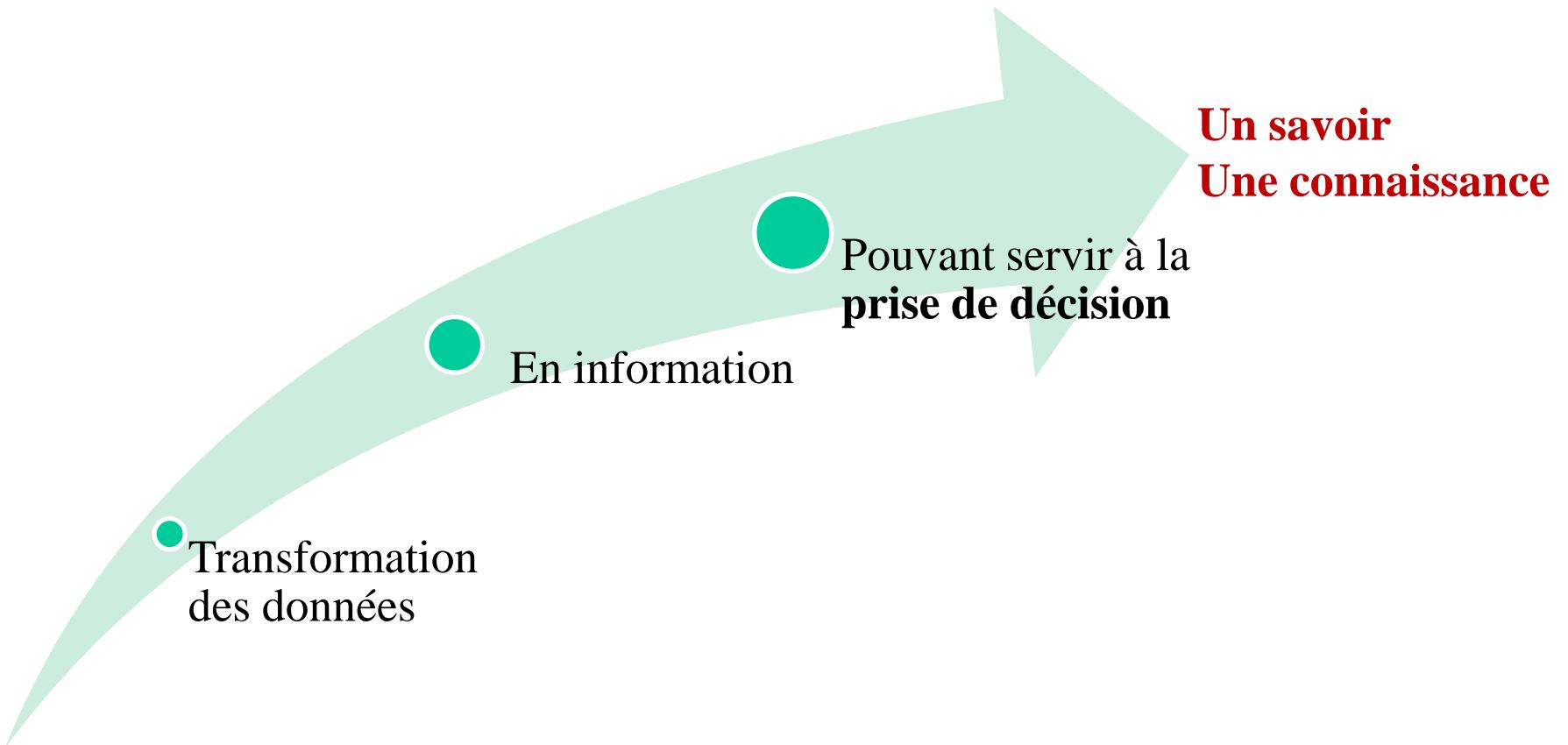
ÉVOLUTION DES SYSTÈMES D'INFORMATION

- Automatisation des opérations (1960 +)
- Pourvoyeur de rapports/information (1960 +)
- Bureautique – Autonomie dans la manipulation/production d'information (1980)
- Révision des processus – Décentralisation des systèmes (Downsizing) (1980 +)
- Passage aux ERP/MRP (1990 +)
- Information stratégique (CRM, DSS, IA) 1995 +

Évolution du focus des T.I.



BESOINS EN INFORMATIONS STRATÉGIQUES

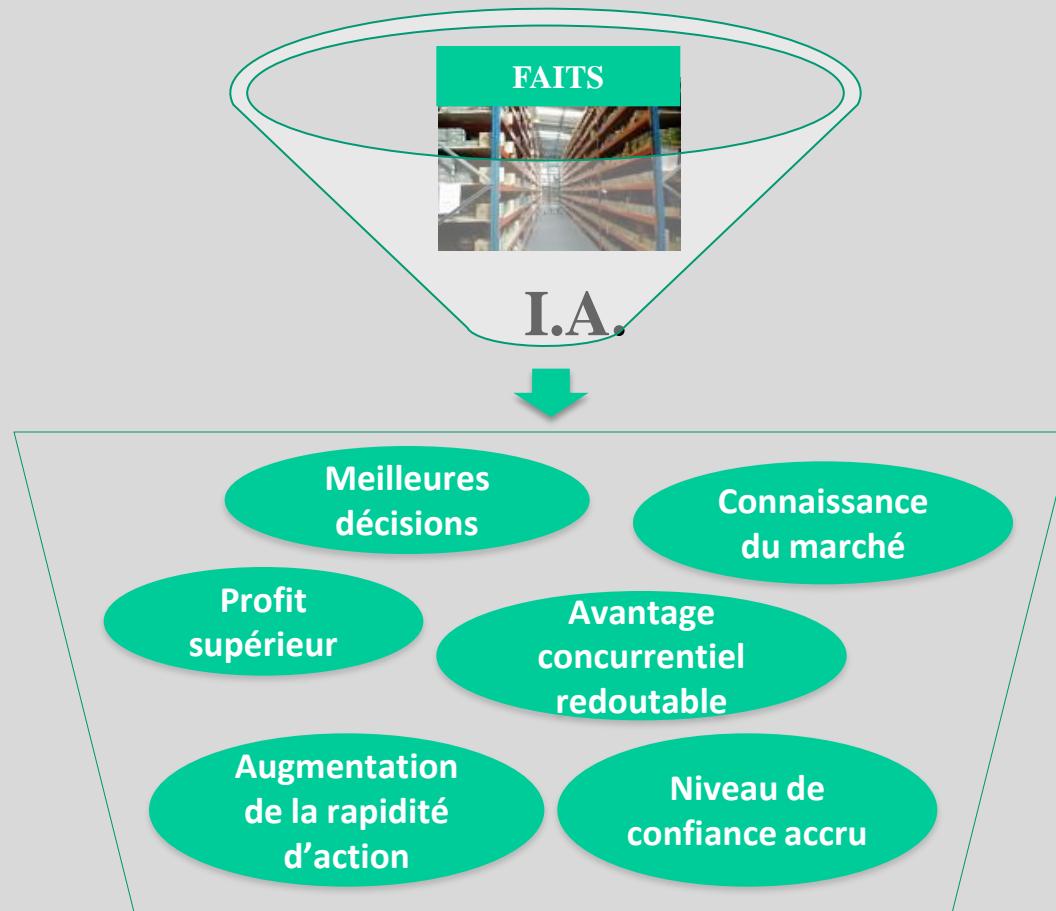


Besoin en information stratégique

- Qualités:
 - Intégré
 - Intègre
 - Claire
 - Accessible
 - Crédible
 - Synchrone/En temps



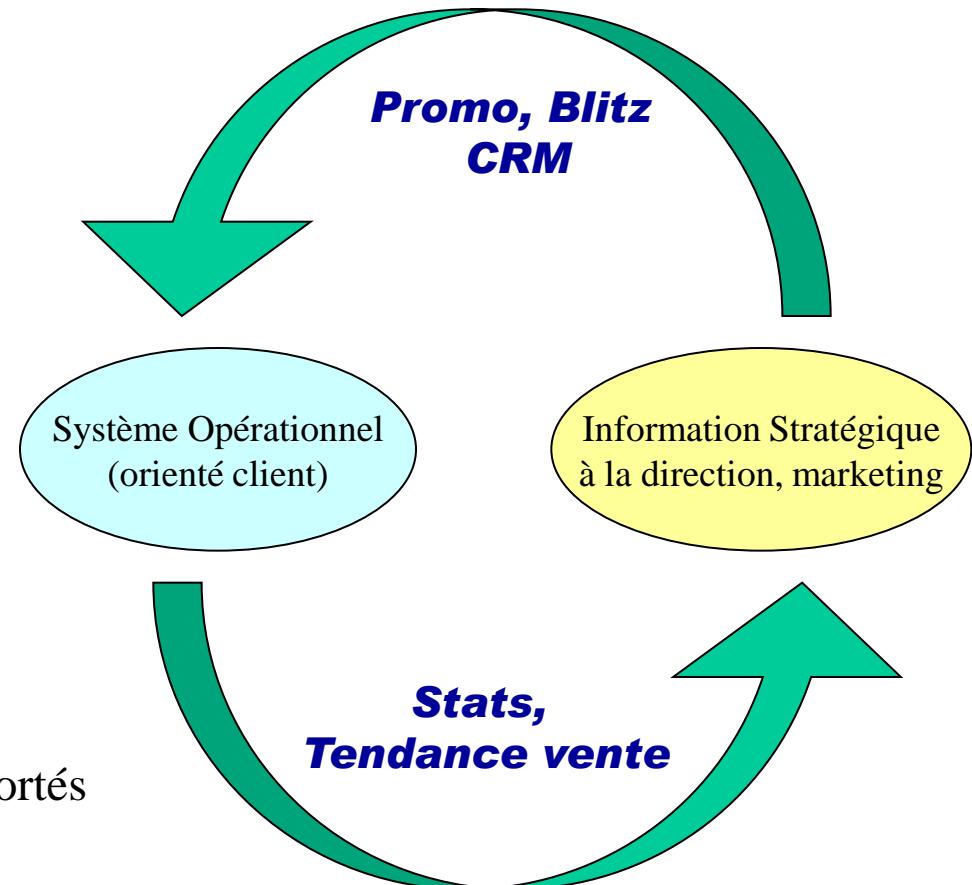
Les bénéfices



Besoin en information stratégique

- NOUVEAUX AVANTAGES COMPÉTITIFS:

- Fidélisation,
- Évaluation des risques,
- Profitabilité par activité,
- Optimisation de processus,
- Gestion des ressources,
- Objectifs documentés et supportés



Besoin en information stratégique

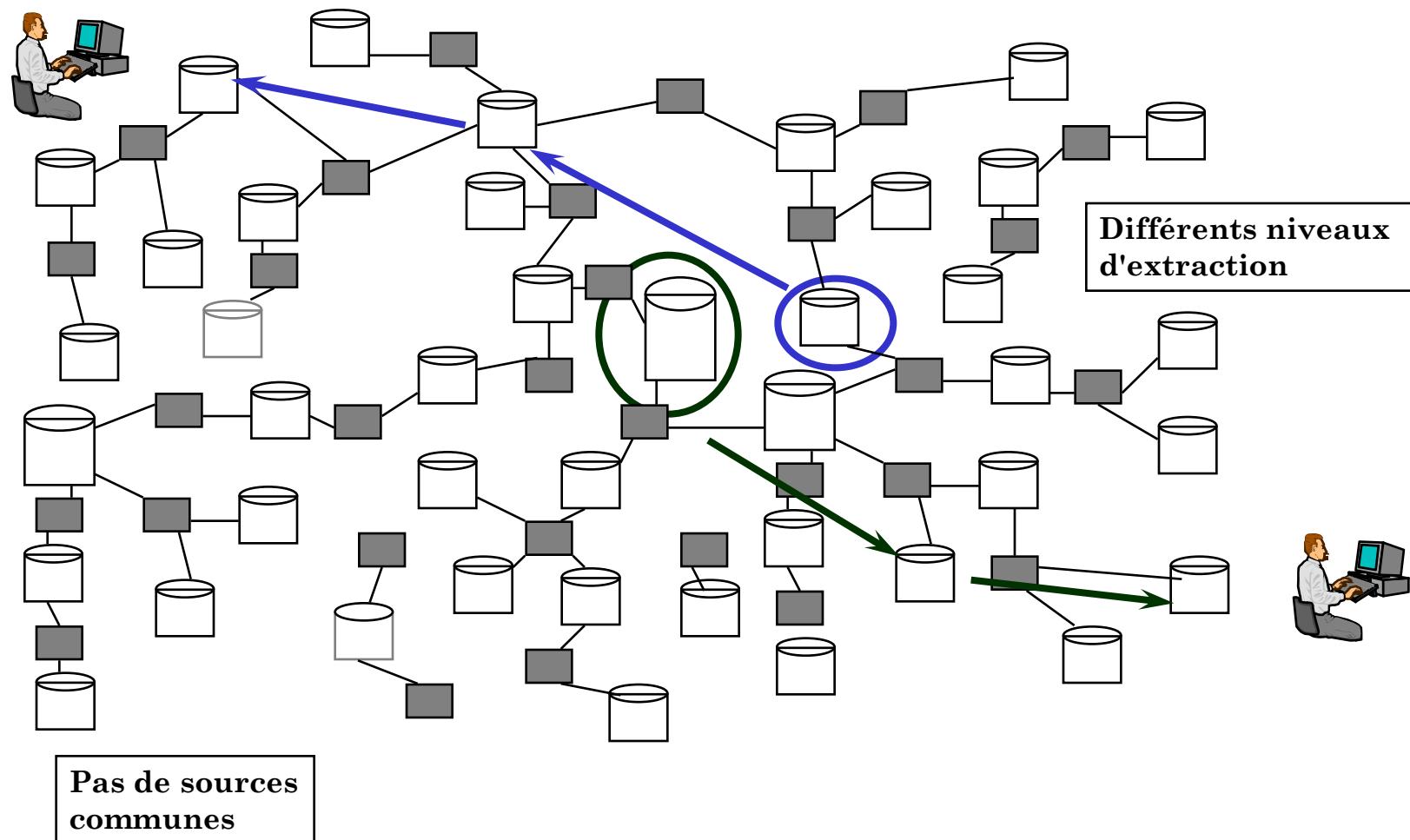
- Obstacles
 - Diversité des systèmes opérationnels
 - Diversité des plate-formes
 - Définitions de donnée différentes d'un système à l'autre
 - Trous d'information
 - Méthodologie de rapports trop long et non adaptée
- ➔ Dépendance des T.I.

+ la politique

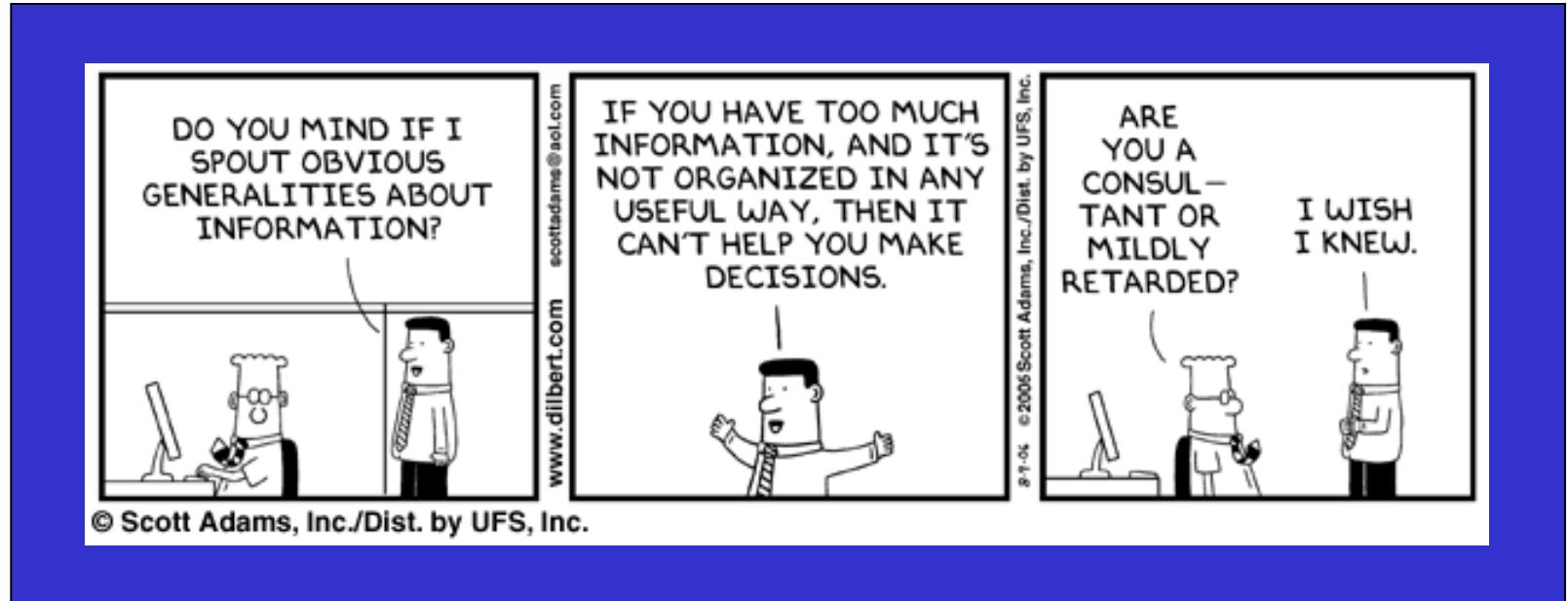


Besoin en information stratégique

- Obstacles



Besoin en information stratégique



Besoin en information stratégique

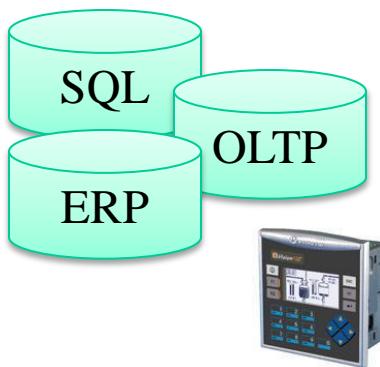


Le rêve d'autonomie des utilisateurs...

La source unique de vérité

Données

Données structurées



Données non structurées



Processus

Rapports



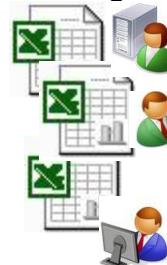
Systèmes



Budget



Budgets



Analyse



Quelle vérité



<http://smartbi.hyperion.com/>

Pour ou contre Wal-Mart?

THE NEW YORK TIMES

La célébration annuelle de l'économie de la consommation aux États-Unis, soit la saison de magasinage des Fêtes, vient tout juste de commencer et il semble que Wal-Mart, le poids lourd du détail, n'a déjà fait une victime. Ainsi, les dirigeants des magasins FAO Schwarz ont annoncé récemment leur intention de déposer le bilan de la compagnie. Se lamentant à l'annonce de cette nouvelle, les analystes ont expliqué que la faute de FAO Schwarz (vendre des jouets à prix forts dans des magasins somptueux) s'est avérée incapable de soutenir l'avance régulière de Wal-Mart dans le domaine des jeux.

« Wal-Mart s'appropriera-t-il Noël ? » demandait un gros titre du magazine Time.

Ce guerre des jouets n'est que la plus récente manifestation de ce que l'on appelle l'effet Wal-Mart. Aux yeux des critiques de la compagnie, Wal-Mart montre la voie à un sombre monde darwinien marqué par la faillite de concurrents, de bas salaires, de chétives prestations de maladie, d'emplois perdus au profit de pays étrangers ainsi que par des centres-villes et des secteurs ruraux dévastés partout aux États-Unis.

Mais il existe une perception plus large, moins partisane de l'entreprise, laquelle marche peut-être plus visiblement que toute autre corporation au pas de l'économie capitaliste mondiale.

« Wal-Mart est l'aboutissement logique et l'avenir de l'économie dans une société dont la valeur préalable consiste à obtenir la meilleure affaire possible », explique Robert B. Reich, ancien Secrétaire américain au travail et professeur de politiques économiques et sociales à l'Université Brandeis.

Un agent de vertu économique

Du point de vue des supporters de la compagnie, Wal-Mart est un agent de vertu économique qui utilise sa puissance sur le marché pour contraindre les fournisseurs à devenir plus efficaces et pour remettre les gains ainsi réalisés aux consommateurs sous forme de prix plus bas. Les enthousiasmes soutiennent que Wal-Mart est un facteur d'importance dans l'inflation presque inexistante aux États-Unis et dans les impressionnantes gains de productivité qu'en observe.



Du point de vue des supporters de la compagnie, Wal-Mart est un agent de vertu économique qui utilise sa puissance sur le marché pour contraindre les fournisseurs à devenir plus efficaces et pour remettre les gains ainsi réalisés aux consommateurs sous forme de prix plus bas.

Il y a beaucoup à dire sur le concept « obtenir la meilleure affaire », soulignent des économistes. Les petits, notent-ils, sont essentiellement une mesure d'efficacité du point de vue du consommateur. Les prix sont concrets et mesurables, tandis que d'autres valeurs de bien-être social et de bien-être des consommateurs (la qualité des produits, par exemple, ou la conservation des emplois) sont souvent difficiles à quantifier ou nécessitent des interventions coûteuses telles des mesures de protectionnisme ou des subventions.

De plus, ajoutent certains économistes, des prix plus avantageux pour des biens essentiels en vente dans les magasins Wal-Mart, tels que les aliments et les vêtements, profitent le plus aux personnes les

moins riches. Par exemple, les prix d'épicerie ont baissé de 10 % à 15 % en moyenne sur les marchés où Wal-Mart a fait son apparition, précisent ces analystes.

« Wal-Mart est la meilleure chose qui soit jamais arrivée aux Américains à faibles revenus », avance W. Michael Cox, économiste principal de la Federal Reserve Bank, à Dallas, « ils peuvent en avoir plus pour leur argent et se permettre des choses qu'ils ne pourraient pas autrement », ajoute-t-il.

Wal-Mart est la plus importante entreprise américaine au chapitre des ventes, soit de 235 milliards de dollars américains l'an dernier. L'entreprise est désormais au premier rang des épiciers, des vendeurs de jouets et des fournisseurs de meubles aux États-Unis. Plus de 30 % des couches jetables achetées aux États-Unis proviennent des magasins Wal-Mart, sous forme de 30 % des produits pour les cheveux, 26 % des dentifrices et 20 % des aliments pour animaux de compagnie. Wal-Mart compte près de 3600 magasins aux États-Unis et la compagnie prévoit en ajouter 1000 au cours des cinq prochaines années. La compagnie exporte de plus en plus sa formule à l'étranger et Wal-Mart est devenu le plus gros employeur privé au Mexique.

La perspective de voir Wal-Mart ravir encore plus de parts de marché n'échappe pas des économistes adeptes du marché libre tels que M. Cox. Malgré les gains de la compagnie, l'industrie du détail n'est pas encore très concentrée, dit-il. Wal-Mart réalise 20 % des ventes des 100 plus importants détaillants. Son succès repose, dit M. Cox, sur sa capacité à maîtriser l'usage de la technologie de l'information pour rationaliser ses opérations, tout comme Dell le fait dans le monde des ordinateurs personnels. Inévitablement, des concurrents moins efficaces vont être éliminés, ajoute-t-il, et ceux qui resteront se battront griffes et ongles pour les dollars des consommateurs.

« Avec l'avènement du nouvel âge de la technologie de l'information », dit M. Cox, nous évoluons vers une nouvelle structure de marché dans de nombreuses industries. Et le nombre optimal de firmes a beaucoup baissé. »

tais des consommateurs.

« Avec l'avènement du nouvel âge de la technologie de l'information », dit M. Cox, nous évoluons vers une nouvelle structure de marché dans de nombreuses industries. Et le nombre optimal de firmes a beaucoup baissé. »

How Companies Learn Your Secrets

By CHARLES DUHIGG

Andrew Pole had just started working as a statistician for Target in 2002, when two colleagues from the marketing department stopped by his desk to ask an odd question: “If we wanted to figure out if a customer is pregnant, even if she didn’t want us to know, can you do that? ”

Pole has a master’s degree in statistics and another in economics, and has been obsessed with the intersection of data and human behavior most of his life. His parents were teachers in North Dakota, and while other kids were going to 4-H, Pole was doing algebra and writing computer programs. “The stereotype of a math nerd is true,” he told me when I spoke with him last year. “I kind of like going out and evangelizing analytics.”

The New York Times®
Reprints

--Charles Duhigg, *The New York Times*, February 16, 2012

Fourni en annexe

Hurricane coming? Get the Pop-Tarts



In general, you'd say hurricanes are bad for business. But apparently they're good for some products—specifically, beer and strawberry Pop-Tarts.

According to an item in Sunday's *New York Times* about *Wal-Mart and its pretty scary knack for learning all about its customers through their buying habits*, the company used some groovy software to figure out what Floridians bought as Hurricane Charley approached—and stocked its stores accordingly for Hurricane Frances.

Survival gear and canned goods weren't the only go-to categories. "We didn't know in the past that strawberry Pop-Tarts increase in sales, like seven times their normal sales rate, ahead of a hurricane," a Wal-Mart rep says. "And the pre-hurricane top-selling item was beer."

Does the next Pop-Tarts commercial not write itself?

--*Posted by Tim Nudd on ADFREAK.COM, November 14, 2004*

Productivité! Labatt



by Neil Sutton

2/17/2005 5:00:00 PM - on ITBusiness.ca

[**Labatt Breweries of Canada**](#) is now serving its sales and marketing personnel the same information by moving to corporate performance management software from [**Cognos Inc.**](#)

Labatt used to divide its data by territory, making it difficult to gain a holistic view of the beer company's performance. Over the last year, Labatt has deployed corporate performance management (CPM) throughout its organization, creating a single view that can be shared across the enterprise.

"From an efficiency perspective, it's a huge change. People aren't looking off into five or 10 different reports. It's all under one house," said Mike Ali, change management and enterprise business intelligence manager at Labatt.

Au tour de l'intelligence d'affaires

Les entreprises ont beaucoup investi afin de mettre en place des plates-formes technologiques qui intègrent les processus d'affaires et renforcent l'efficacité opérationnelle. À présent, bon nombre d'entre elles aimeraient retirer un réel profit des masses de données qu'elles ont emmagasinées. C'est là qu'entre en jeu l'intelligence d'affaires ou *business intelligence* en anglais. Les exemples de succès ne manquent pas, mais il faut s'y prendre correctement.

En collaboration avec HEC Montréal, nous amorçons la publication d'une nouvelle série d'articles rédigés par des professeurs et chercheurs de cette maison d'enseignement. Chaque lundi, ils font le point sur différents aspects des pratiques d'affaires.

MARLEI POZZEBON
et DENIS PARADIS
COLLABORATION SPÉCIALE

Pizza Hut a mis en place un entrepôt de données concernant 40 millions de foyers américains parmi ses clients.

Depuis, Pizza Hut peut segmenter sa clientèle en fonction de ses comportements d'achat, envoyer le meilleur coupon publicitaire à chaque domicile et prévoir à l'avance le succès de chacune de ses campagnes de promotion.

Il n'aura fallu qu'une année pour que l'entreprise puisse récupérer les coûts d'investissement et de formation associés à la mise en oeuvre de cette technologie⁽¹⁾.

Le cas de Pizza Hut n'est pas unique. The Limited (2700 magasins spécialisés en produits de beauté Intimate Brands) a aussi bâti un entrepôt à partir des données de ses 25 millions de clients dans le but premier de segmenter sa clientèle adéquatement.

Ainsi, au cours d'une de ses campagnes de marketing directe, à l'oc-



PHOTO ANDY SHAW, BLOOMBERG NEWS

Pizza Hut peut segmenter sa clientèle en fonction de ses comportements d'achat, envoyer le meilleur coupon publicitaire à chaque domicile et prévoir à l'avance le succès de chacune de ses campagnes de promotion.

casion du lancement d'une nouvelle fragrance, l'entreprise a fait la surprise découverte que les clients utilisant les trois canaux de distribution (magasin, Web et catalogue) dépenseront en moyenne de 3 à 5 fois plus que ceux qui utiliseront un canal unique.

En sélectionnant et ciblant 100 000 de leurs clients, la compagnie a effectué une campagne qui a été un franc succès⁽²⁾.

Au Canada, la Banque de Montréal (BMO) a effectué des investis-

sements importants au cours des dernières années pour mettre au point un outil d'analyse capable d'explorer une base de données d'environ 8 terabytes (18 millions de comptes à travers 32 différentes lignes d'affaires) dans le but d'y découvrir des secrets insoupçonnés chez leurs clients⁽³⁾.

L'expérience de Desjardins va dans le même sens. Avec ses 600 caisses et ses 1000 centres de service, Desjardins a décidé de réduire l'utilisation des outils « maison », à

LA PRESSE MONTRÉAL LUNDI 23 JANVIER 2006

GESTION D'AUJOURD'HUI

Argent Canadian Tire change pour un programme fidélité



(archives)

Le plus vieux programme de récompenses aux consommateurs au pays pourrait bientôt être abandonné au profit d'une carte de fidélité. Au cours des prochaines semaines, Canadian Tire offrira aux clients de 21 succursales de la Nouvelle-Écosse le choix entre la monnaie maison et des points dans un nouveau programme.

Source: Radio-Canada.ca, 16 février 2012

Disney investit 1 milliard de dollars pour suivre à la trace ses visiteurs



Ces bracelets seront disponibles "dans les prochains mois" dans le parc d'Orlando.

La compagnie américaine a développé un programme, baptisé MyMagic+, pour mieux étudier les dépenses des visiteurs fréquentant ses parcs à thème. Un bracelet permet notamment de savoir où et comment les clients ont dépensé leur argent.

L'information a un côté orwellien. Selon [Business Week](#), le géant américain Disney a dépensé pas moins de 1 milliard de dollars dans un programme expérimental baptisé "MyMagic+".

Derrière ce nom "magique", se cache un projet destiné à collecter des données liées aux achats et dépenses des visiteurs des parcs à thème Disney.

Visa

Analyse les patterns d'achats avec des réseaux de neurones et des outils statistiques sophistiqués afin de détecter en temps réel les fraudes

Amazon

Analyse leurs achats antérieurs et ceux de clients similaires afin de proposer des recommandations d'achats personnalisées à chaque client

3M

A développé un entrepôt de données de 5 téraoctets pour stoker et analyser les données financières, de ventes et d'inventaire (e-Business 2.0, p. 361)

Office Depot

A développé un entrepôt de données de plusieurs téraoctets contenant toutes les transactions de ventes. Elles sont utilisées pour analyser les marchés et prédire la demande. (e-Business 2.0, p. 361)

Travelocity

A investi dans un entrepôt de données de plusieurs téraoctets. L'objectif de la compagnie est de mieux gérer le cycle de vie des clients et de mieux cibler les promotions.

Cas:

- DakotaCare is South Dakota's largest statewide health maintenance organization 
- Bax Global, a \$US2 billion transportation company 

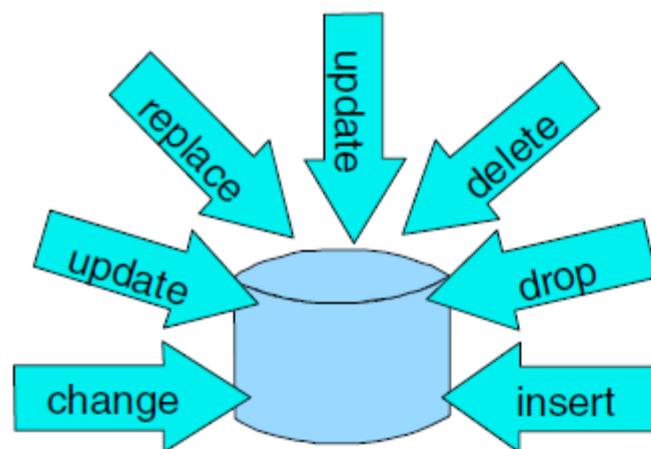
Documents en annexe.

OPÉRATIONNEL CONTRE INFORMATIONNEL

Opérationnel .vs. Informationnel

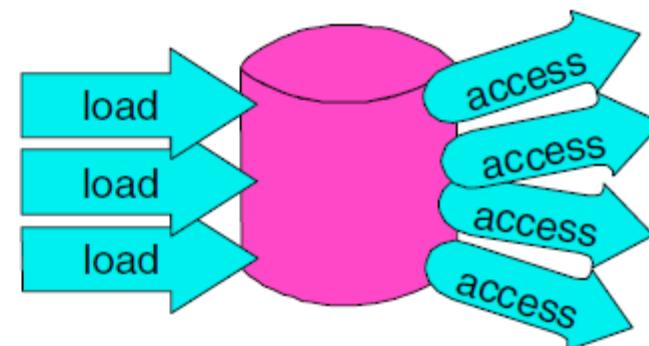
Operational versus Informational Databases

operational



data is regularly updated on a record by record basis

informational



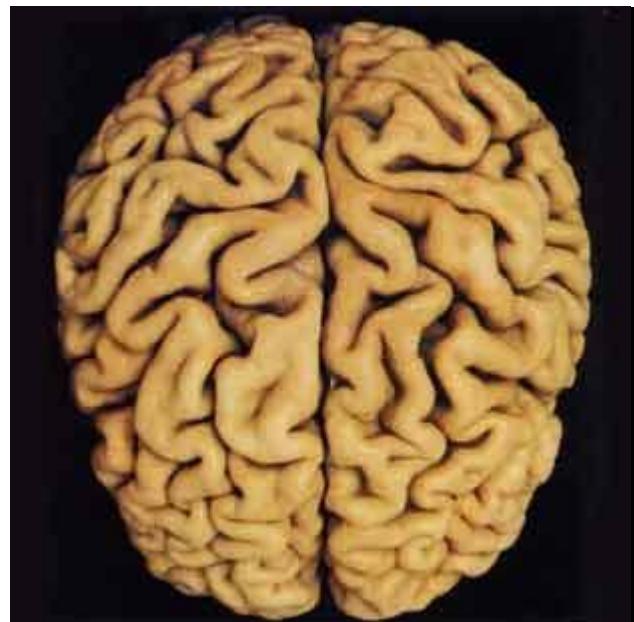
data is loaded into the data warehouse and is accessed there but is not updated

Source: IBM – Business Intelligence Certification Guide – January 2000

Opérationnel .vs. Informationnel

OPERATIONNEL

INFORMATIONNEL



L'opérationnel et l'informationnel remplissent des rôles différents ... par des techniques différentes!

Types de systèmes

	Opérationnel (OLTP)	Informationnel
Utilisation	Opérations de base, au jour-le-jour	Support à la décision
Information	Gestion de l'opération	Gestion stratégique
Données	Valeur courante	Archiv e, épurées, sommarisées
Structure	Optimisé transactions	Orienté querry
"Scope"	Transaction - Département	Entreprise
Usage	Précis, connu et répétitif, toujours	Ad Hoc, semi- prévisible, des fois
Type accès	lecture, écriture, modification, destruction	Lecture
Utilisateurs	Nombreux	Restreints
Données traitées	Peu à la fois	Beaucoup à la fois

Défis des systèmes informationnels

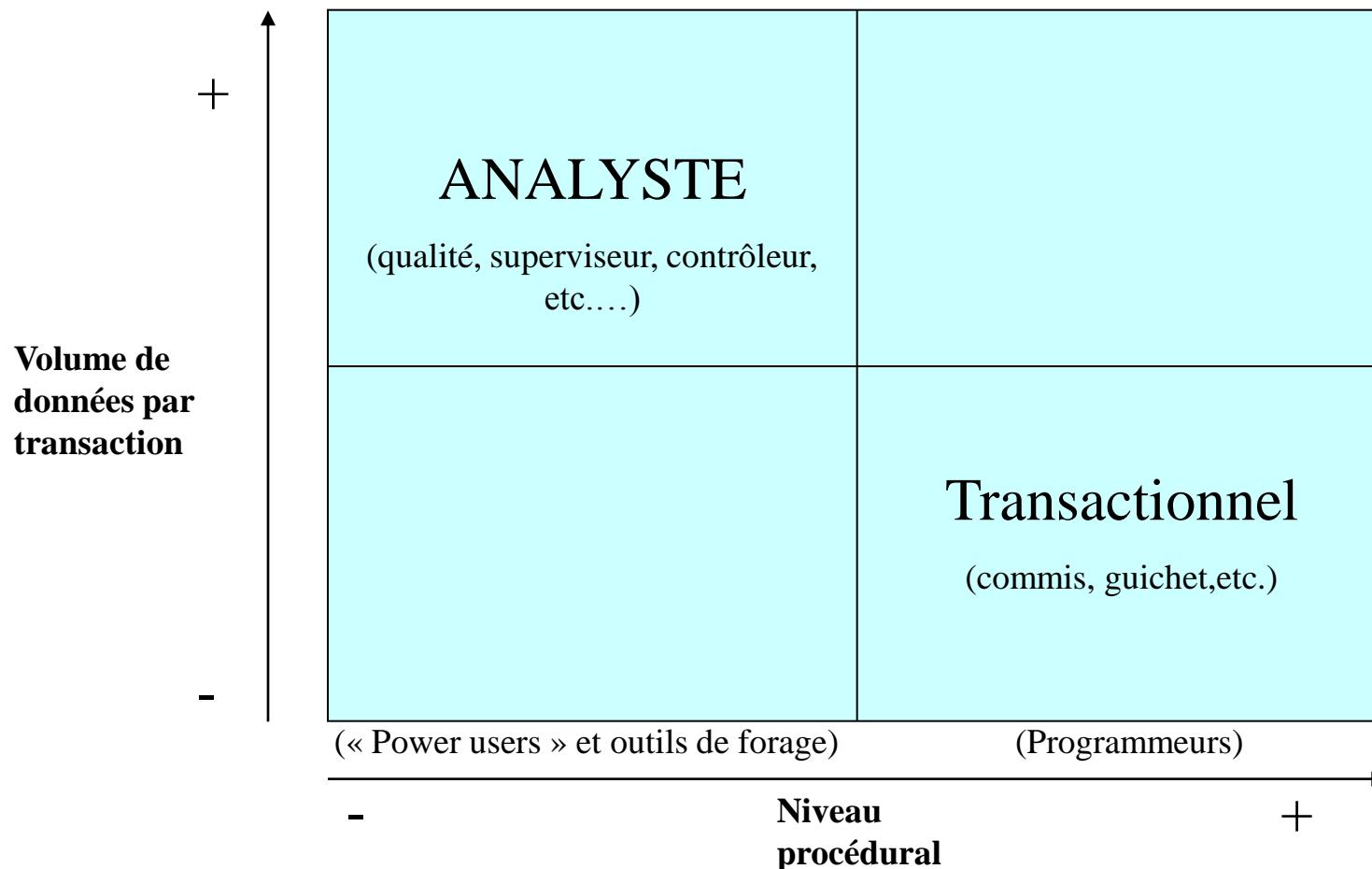
Opérationnel	Informationnel
<ul style="list-style-type: none">•Connaître le comment faire (processus)•Méthode rigoureuse, contrôlée, basée sur des processus connus avec résultats fixes•Valeur mise à jour en direct•Utilisation cadencée, exigences et besoins connus au départ, processus stable•S'adresse au clérical, aux exécutants•Les usagers ont besoin des informaticiens•Temps réponse de 2 secondes	<ul style="list-style-type: none">•Connaître le contenu, ce qui est disponible•Méthode prospective, exploratoire, aléatoire, résultat final qu'on peut manipuler avec valeur ajoutée•Valeur non mise à jour, extraite•Utilisation ad-hoc, besoins non connus au départ, processus variable•S'adresse aux gestionnaires et spécialistes•Les usagers sont dans un libre service pour utiliser l'information•Temps réponse de 2 min, 1 hre, 24 hres, et plus

Opérationnel .vs. Informationnel (stratégique)



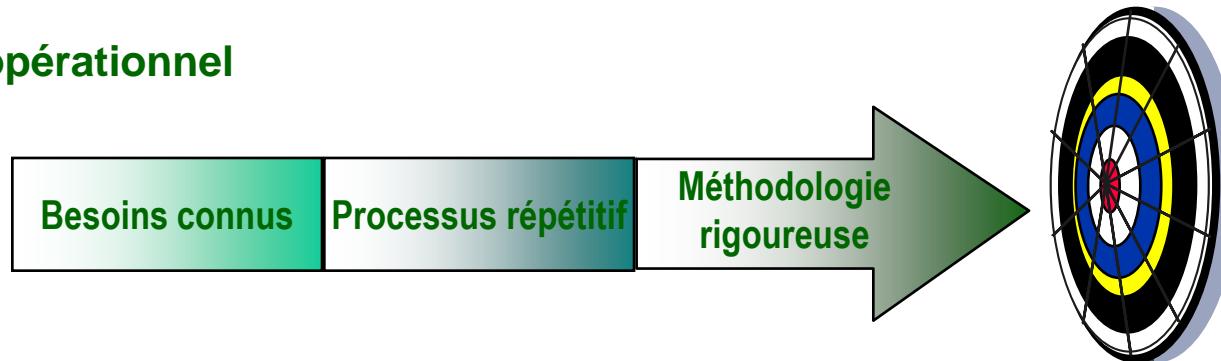
Types de systèmes (Bloc 4 INF-732)

- Besoin des utilisateurs

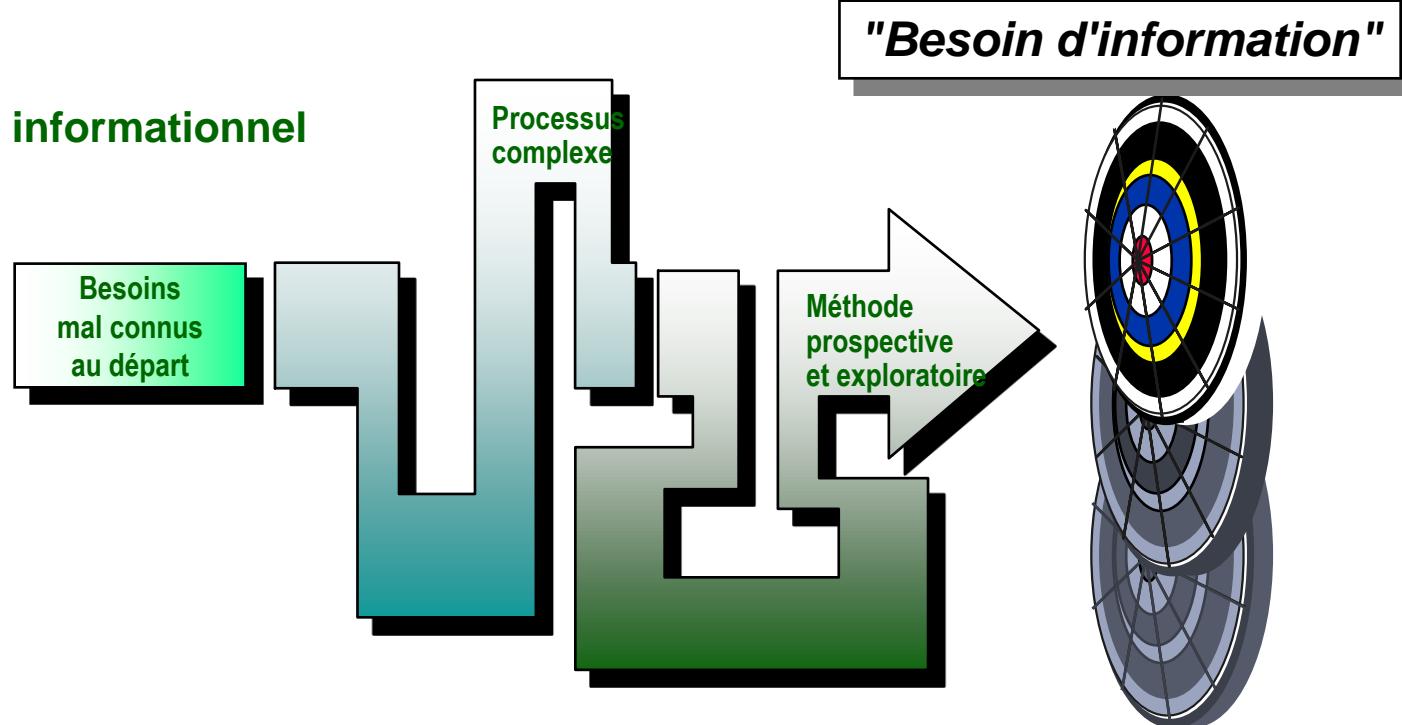


Défis des systèmes informationnels

• Schéma opérationnel



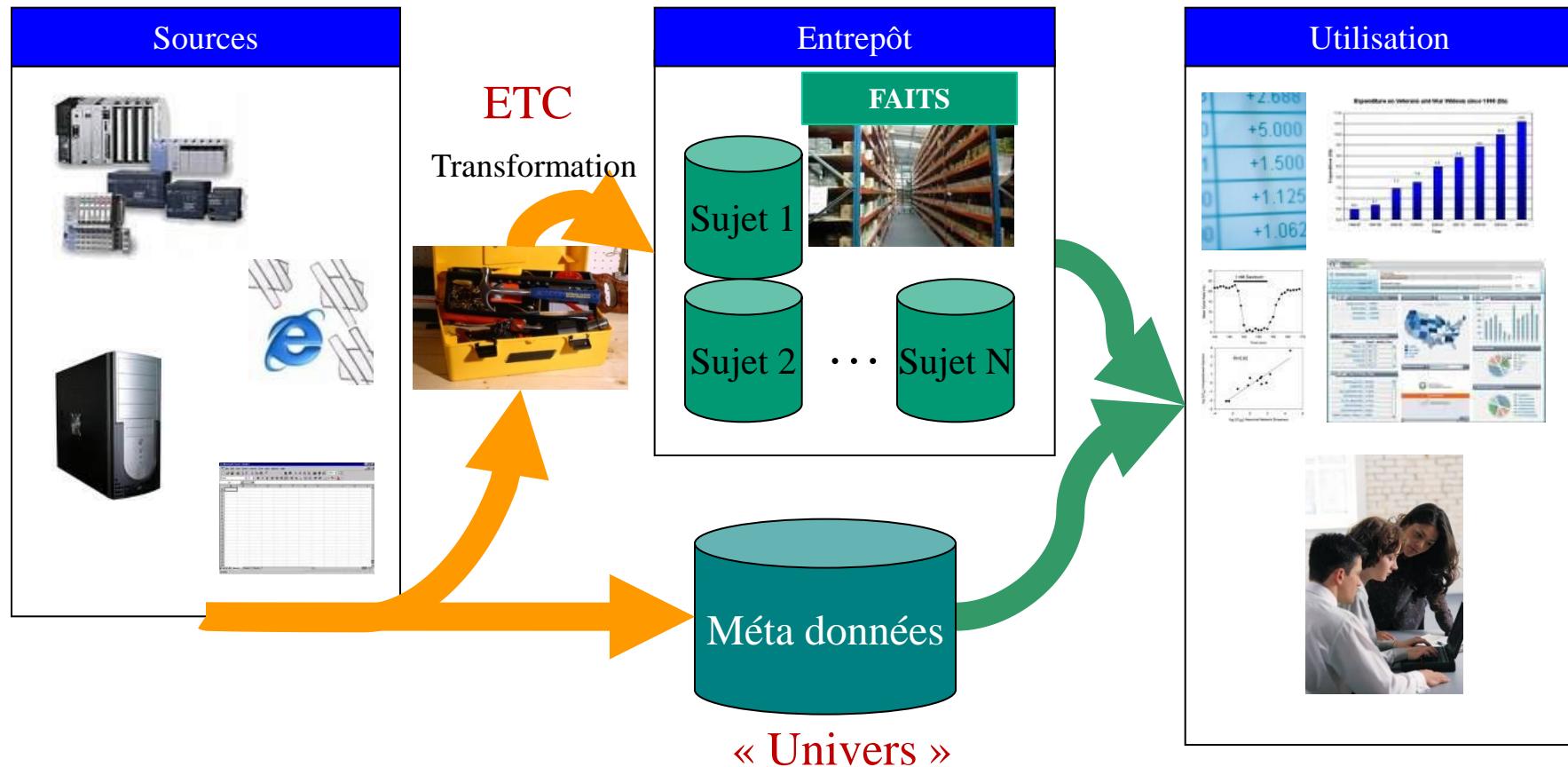
• Schéma informationnel



ENTREPÔT DE DONNÉES = UN ENVIRONNEMENT INFORMATIONNEL

Créer un environnement informationnel

Un mélange de technologies, d'architectures et d'infrastructures



- Vue intégrée et consistante de l'entreprise
- Transforme la donnée opérationnelle en donnée informationnelle disponible à la décision
- Fournit les données pour la prise de décision sans compromettre la performance des systèmes opérationnels
- Source « flexible » d'information stratégique.

Entrepôt de données – pas un produit

- Un environnement informationnel (pas un produit)
 - Plusieurs outils pour des fonctions différentes:
 - + Ravitaillement = Extraction, Transformation, Chargement
 - + BD = Cubes
 - + Présentation = OLAP, EXCEL, Rapports, Tableau de bords, etc..
 - + Méta-information
 - + Monitoring
 - Le modèle dépend du besoin informationnel
 - Cible mouvante, besoin de flexibilité
 - Objectif n'est pas un système mais une source fiable et flexible de données de qualité.

(voir définitions plus loin)

DÉFINITIONS

- « Un Entrepôt de données est une collection de données orientée par sujet, intégrée, non-volatile et variant dans le temps pour l'aide à la décision de gestion »

Bill Inmon



Bill Inmon

Le père de l'entreposage des données

Bill Inmon est reconnu, à juste titre, comme le « père » de l'entreposage de données. Entre la création de l'expression « entrepôt de données » durant une réunion avec Shell Canada à Calgary à la fin des années 80 et la publication de son livre le plus récent sur le sujet, M. Inmon a eu une influence considérable sur la mise en forme de cette technologie. Avec plus de 26 ans d'expérience, 36 livres publiés et plus de 350 articles publiés dans d'importantes revues d'informatique, il a acquis une réputation mondiale grâce à ses séminaires sur l'élaboration d'entrepôt de données. Par l'entremise de sa société Pine Cone Systems Inc., M. Inmon a conseillé nombre des 1000 sociétés les plus importantes selon Fortune.

- Innovation Printemps 1998

Définition (Prise 2)

- La donnée d'un entrepôt est:
 - Orienté sujet
 - Intégré
 - Variable dans le temps
 - Non-volatile
 - Granularités différentes (*ajouté*)

« Les données d'un entrepôt sont:

- Séparées
- Disponibles
- Intégrées
- Datées
- Orientées par sujet
- Non-volatile
- Accessibles »



Sean Kelly

1^{er} ingénieur logiciel européen à implanter un entrepôt complet en 1991

« *He is a regular speaker on the international conference circuit, with particular focus on the areas of customer relationship management (CRM) and information exploitation for business advantage. In addition, he is regularly sought out for his opinions by the media and his perspective on emerging trends regularly appear in print. Seán has completing his latest book, Customer Intelligence, which charts the transition from mass marketing to targeted marketing techniques and is engaged in launching a new start-up software company in Ireland* ».

- CDI-MDM SUMMIT LONDON 2006

Europe's largest conference dedicated specifically to Customer Data Integration and Master Data Management Solutions

« L’information d’un entrepôt est:

- Par sujet
- Datée
- Intégrée
- Non-volatile
- Intègre
- Disponible
- Pertinente
- au moment désigné et utile
- Intelligible
- Complète



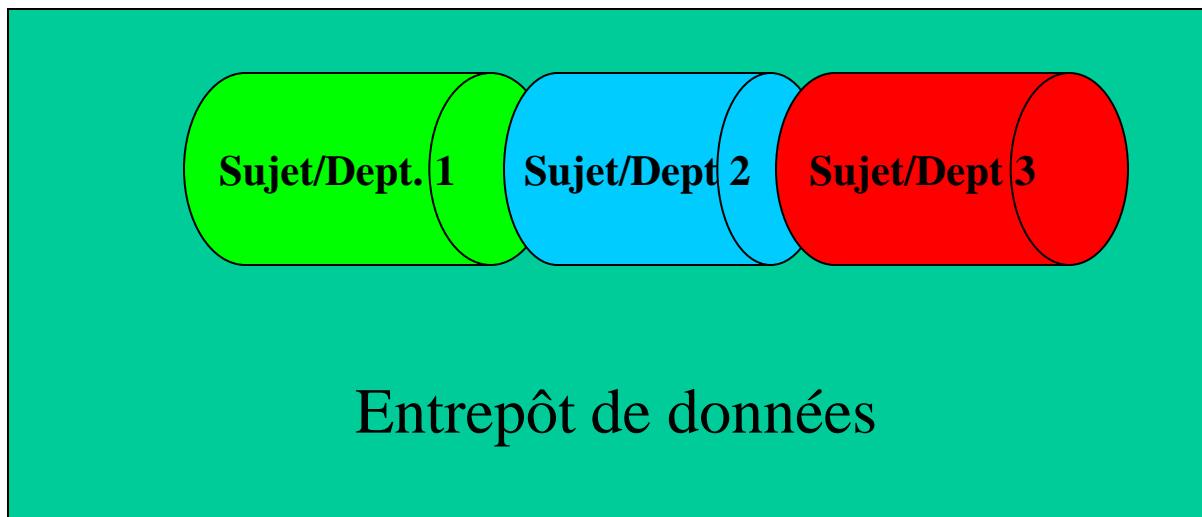
Robert J. Laurin

Robert J. Laurin started his IT career in 1986 with the development of financial applications with Oracle tools. In 1989, he is more attracted by the technical aspect of computing and is now more oriented toward infrastructure design and management. This got him to participate or lead various projects such as data warehousing, review of process control systems, lan/wan design and implementation and many Disaster Recovery Plannings, all for enterprises like Alcan, St-Laurent Cement, Transcontinental distribution, Hershey and Banco de México. M. Laurin is a bachelor from Concordia University in Computer Science, business option. He is now IT Director at Otéra Capital and lecturer at Sherbrooke University for the graduated program in IT.

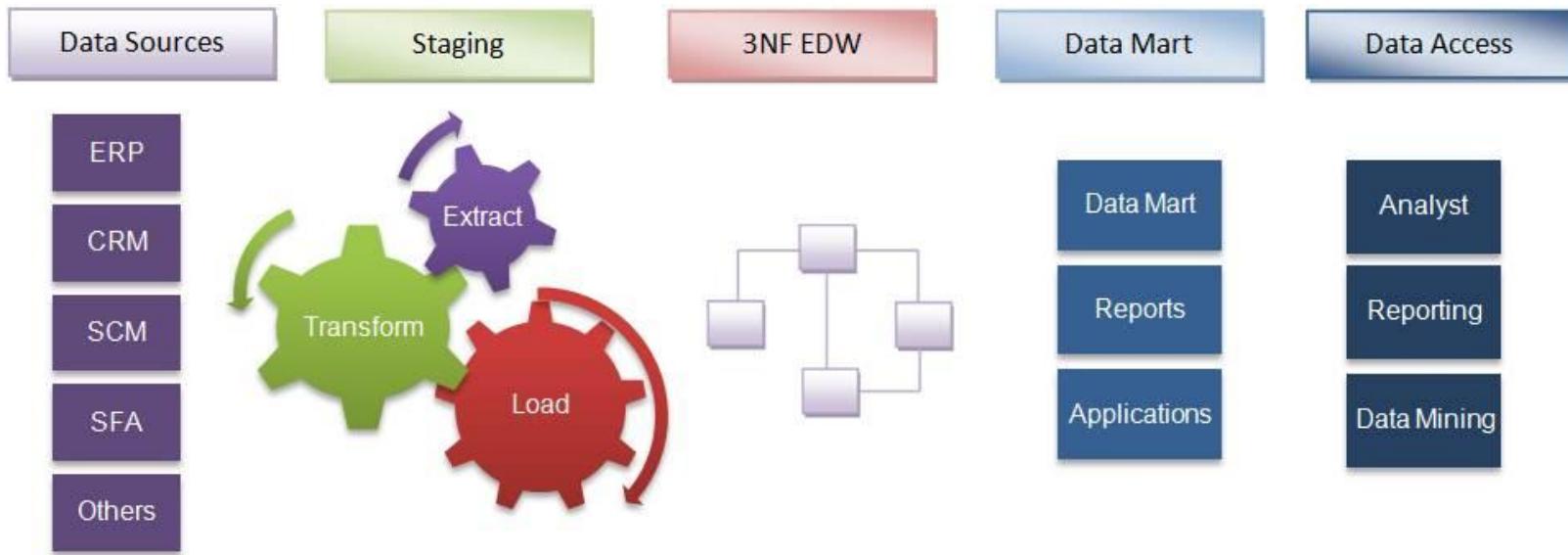
CONSTRUCTION

Entrepôt de données .vs. Comptoirs (« Data marts »)

- Entrepôt peut être découpé en différents Comptoirs



Approche Inmon (« top-down »)

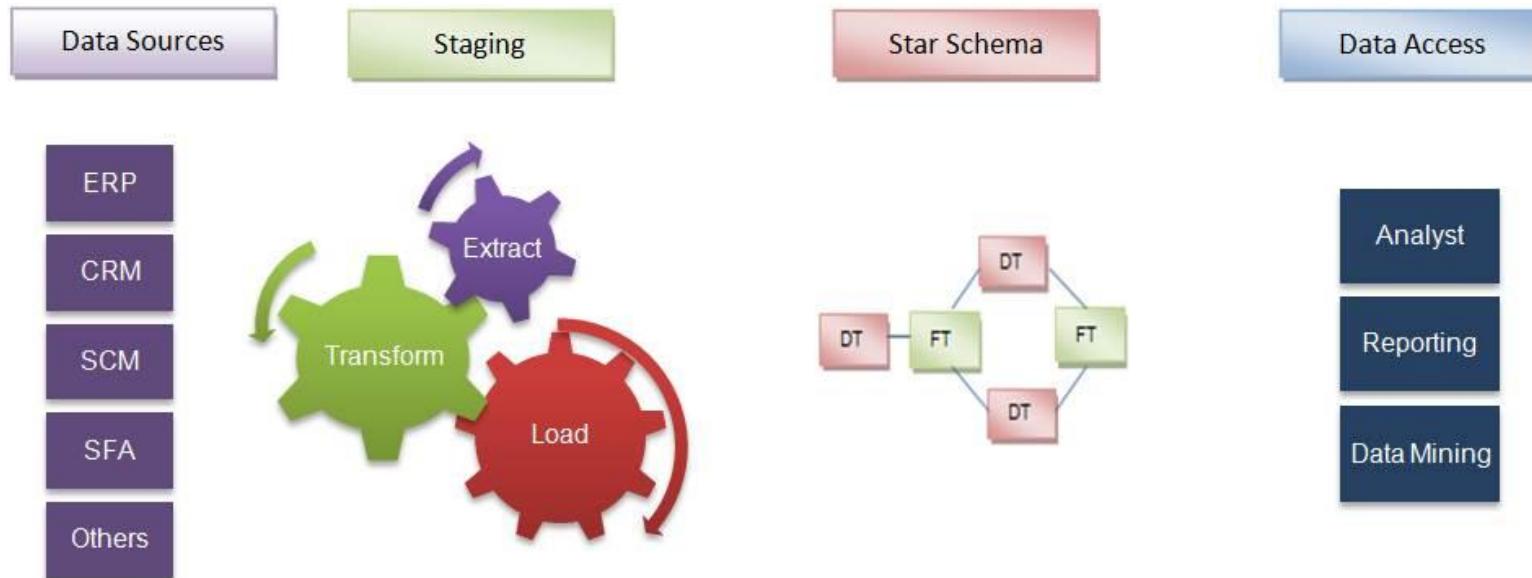


- Amener toutes les données dans un entrepôt normalisé
- Après, créer des comptoirs de données (data marts)
- L'entrepôt global est le point de départ de tous les comptoirs de données

Définitions Entrepôt .vs. Comptoirs

Entrepôt	Comptoirs
<ul style="list-style-type: none">•Englobe l'entreprise•Somme de tous les comptoirs•Toutes les données reçues du ravitaillement•Structuré pour une vue corporative•Sur un SGBD relationnel	<ul style="list-style-type: none">•Départemental•Un sujet précis•Une étoile (partie de la famille/galaxie)•Structuré pour une vue départementale•Technologie adaptée au besoin d'analyse local/spécialisé

Approche Kimball (« Bottom-up »)

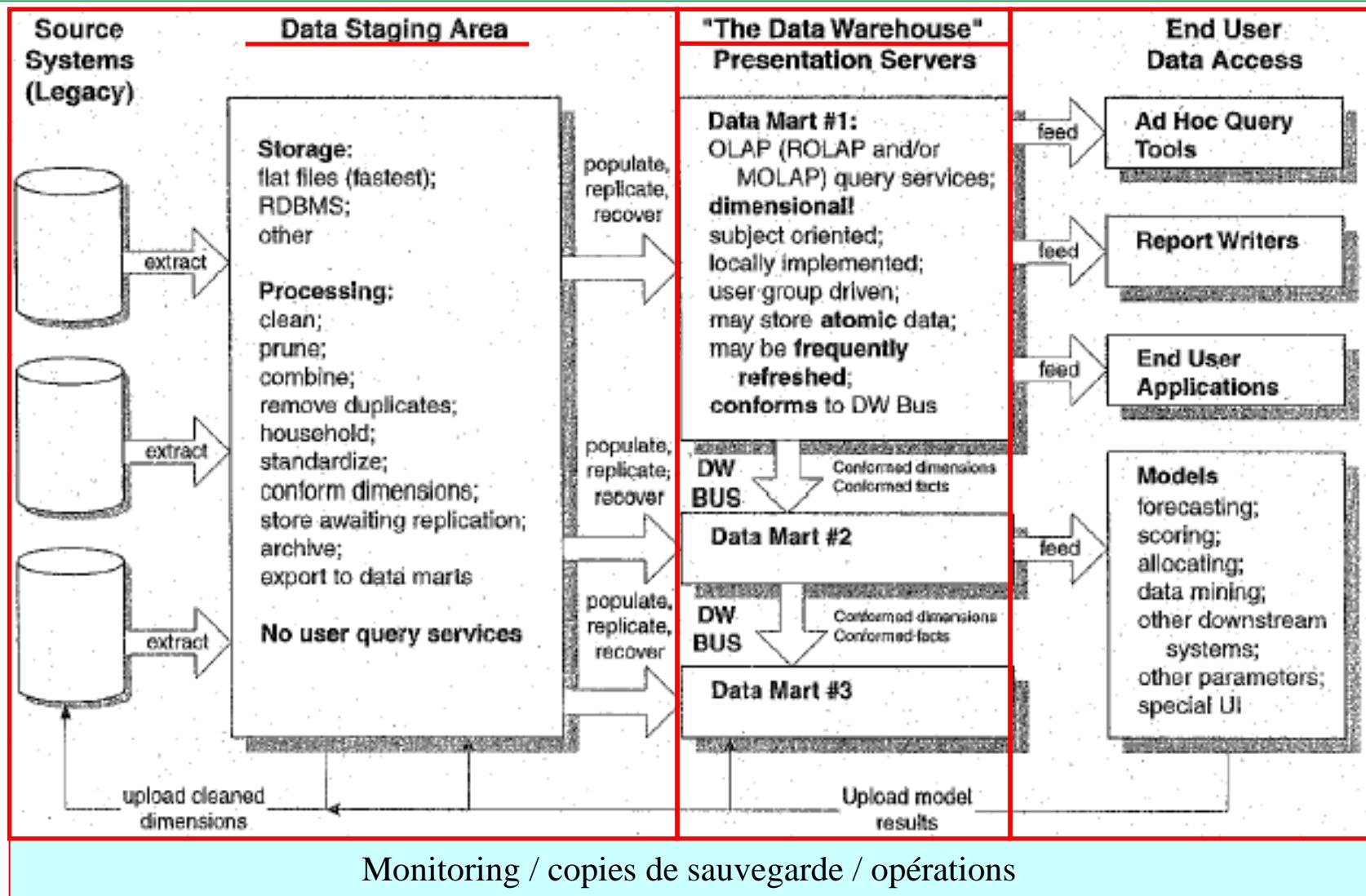


DT – Dimension Table

FT – Fact Table

- Traitements et étoiles par processus d'affaires
- L'entrepôt devient la somme de toutes les étoiles

Étapes/Éléments dans la construction d'un E.D.



[The Data Warehouse Lifecycle Toolkit](#)

Étapes/Éléments dans la construction d'un E.D.

- Sources:
- Zone de ravitaillement des données:
 - Définition :
Zone dans laquelle sont stockées temporairement les données, entre le moment où elles sont extraites des sources avant d'être versées dans un entrepôt de données, et dans laquelle on leur fait subir les différentes opérations de nettoyage et de transformation.
[Office de la langue française, 2000]
- Entrepôt
- Outils de présentation
- Meta Information
- Opérations

1

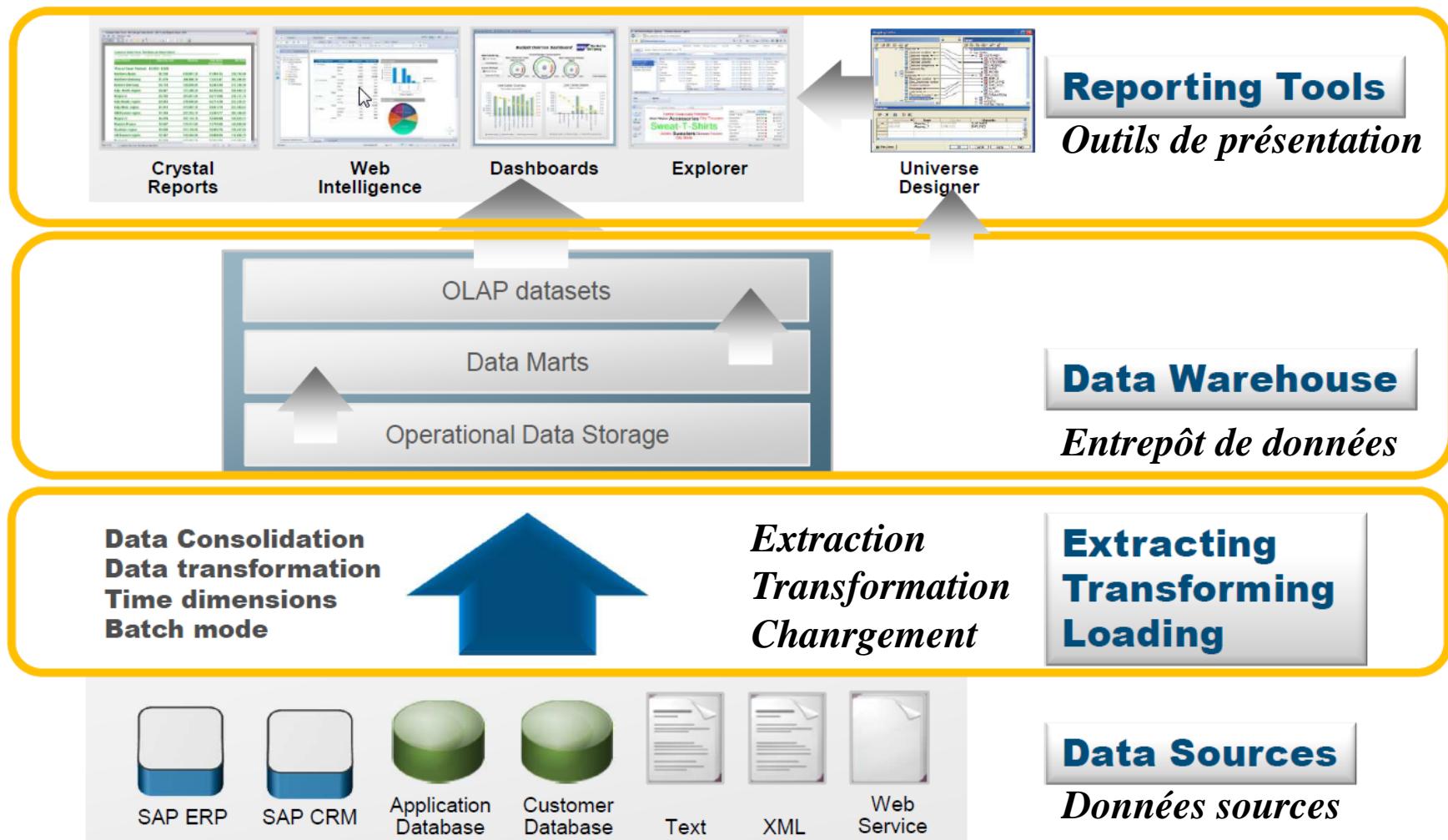
2

3

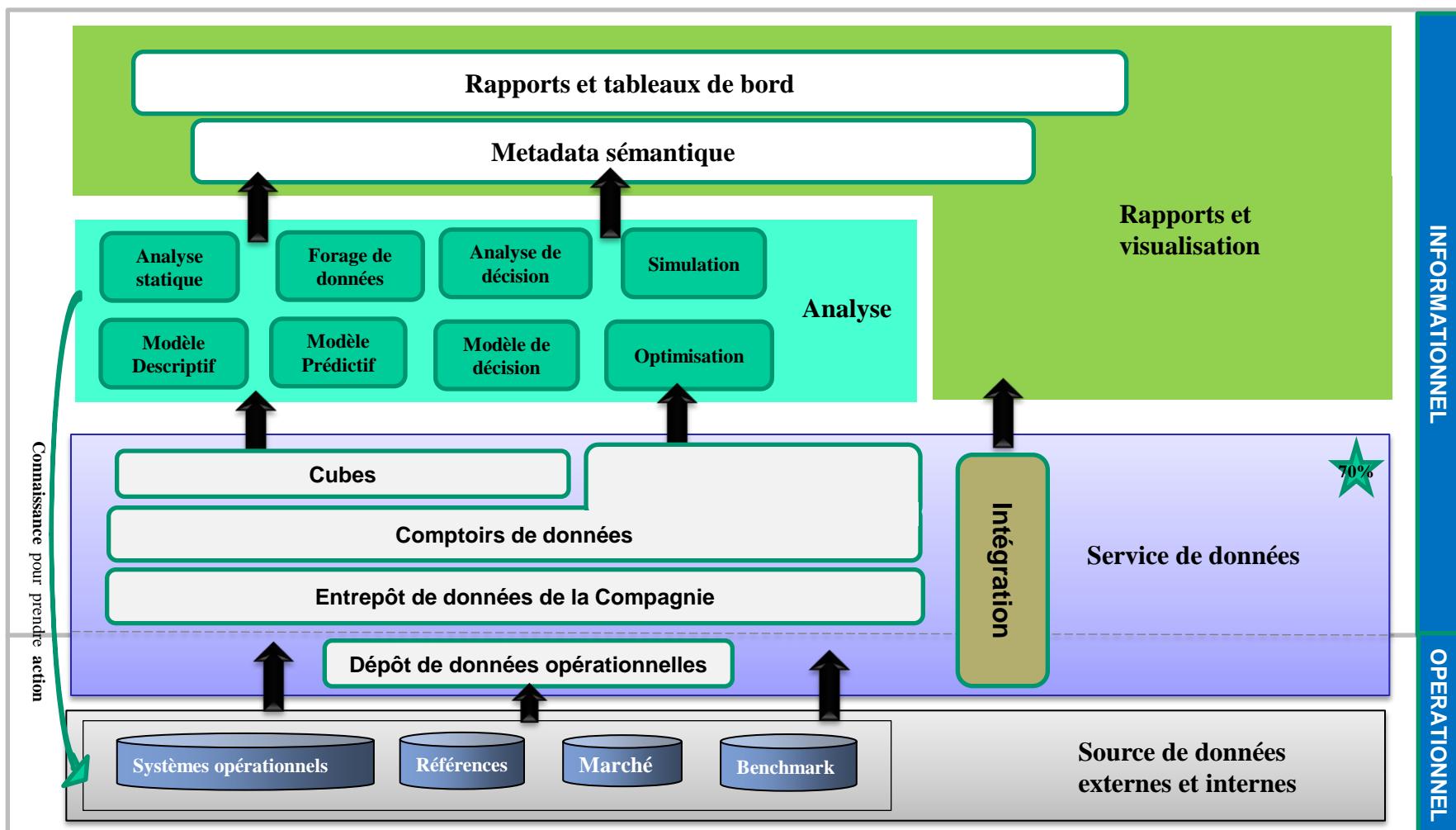
4

5

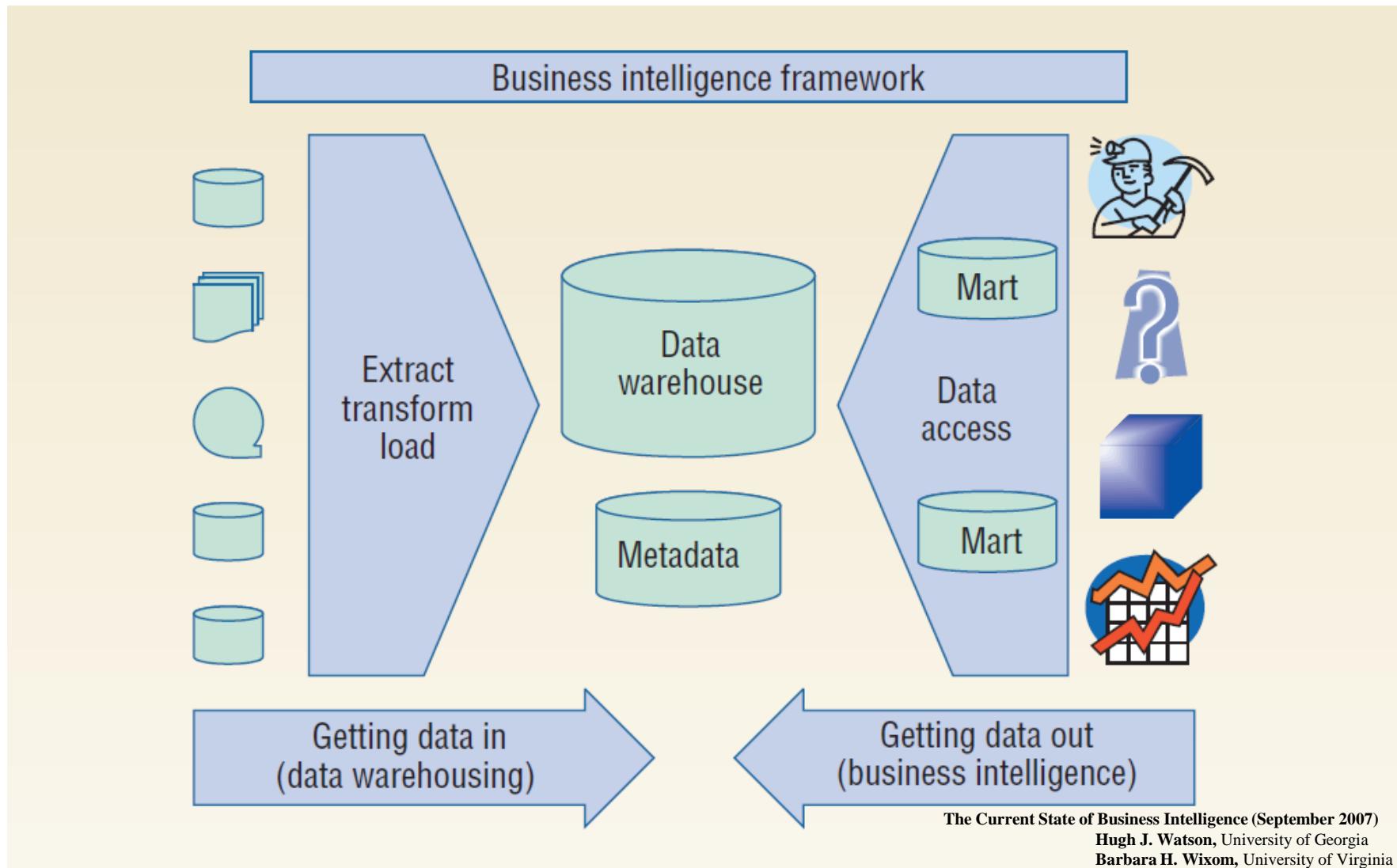
Les grandes étapes



Les éléments d'un environnement informationnel (une autre vision)



Vocabulaire: Entrepôt de données .vs. Intelligence d’Affaire



The Current State of Business Intelligence (September 2007)

Hugh J. Watson, University of Georgia

Barbara H. Wixom, University of Virginia

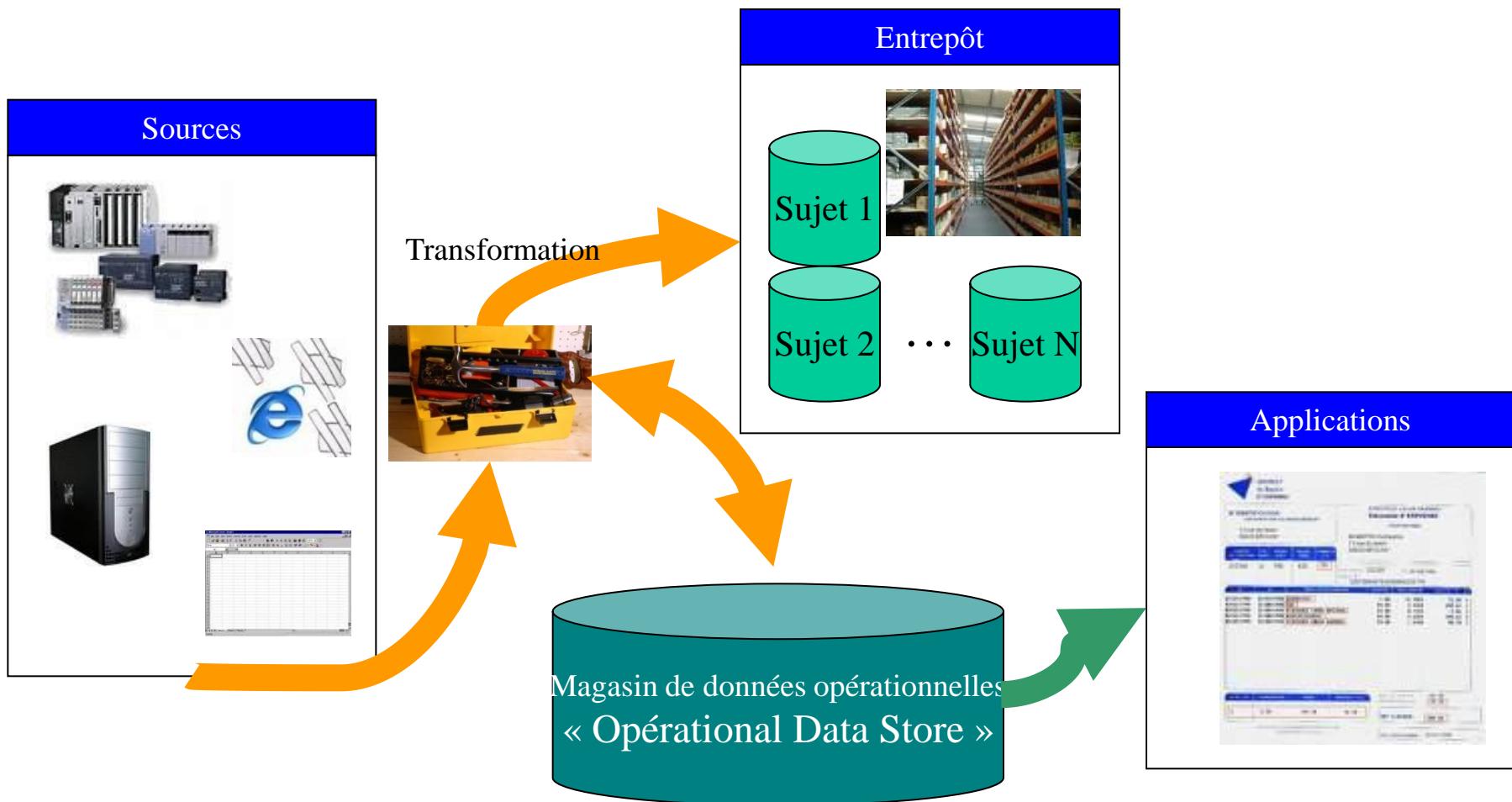
« ODS »

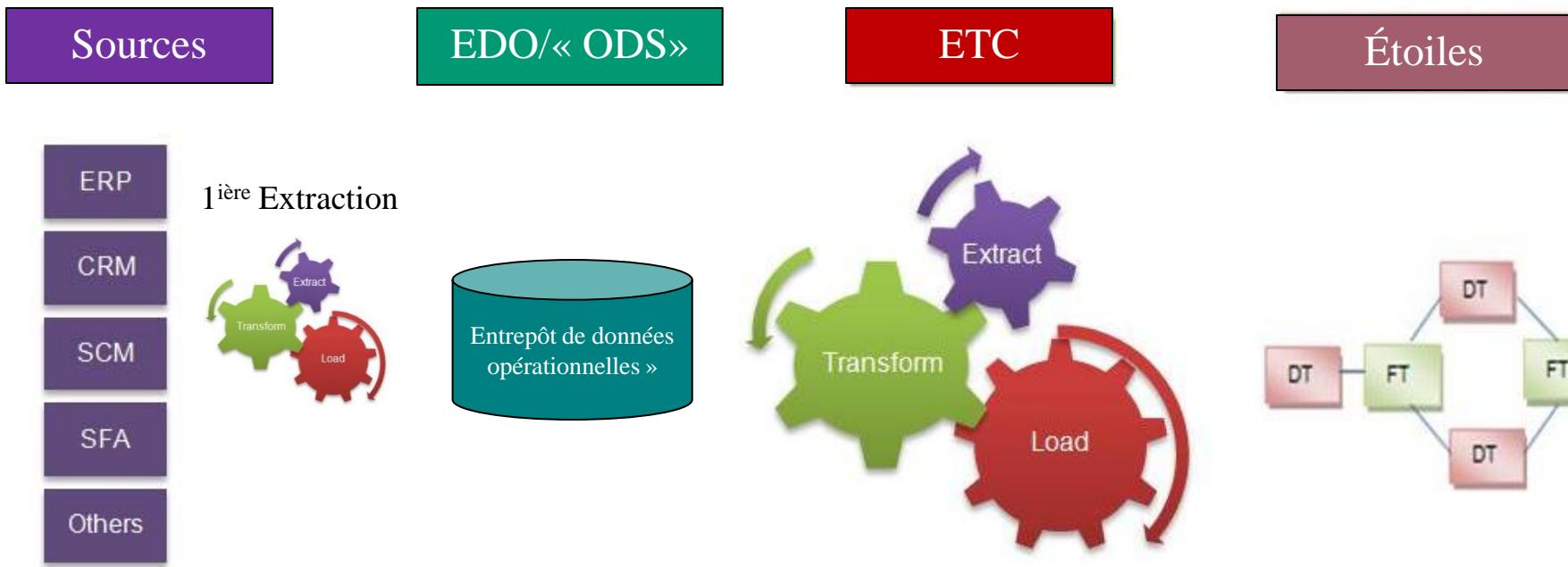
ENTREPÔT/MAGASIN DE

DONNÉES OPÉRATIONNELLES

Entrepôt/Magasin de données opérationnelles

– « ODS »





- 1^{ière} extraction à l'ODS avec transformations « cosmétiques » seulement
 - Normalisé (ou dénormalisé)
 - Nettoyage de base
 - Générique
 - Toutes les données opérationnelles doivent s'y retrouver intégralement → Souhait de Inmon
- Les comptoirs sont créés par sujets/processus
 - Les étoiles sont créées selon les différents besoins avec des modèles adaptés → Souhait de Kimball
 - Les sources de données sont intégrées à l'ODS au fur et à mesure des besoins

EDO / « ODS » - Objectifs

- 1) Conserver les données opérationnelles pour une durée limitée (normalement de 3 à 12 mois)
Mais certains l'utilisent pour « historiser » /conserver pour une longue période (ex: 7 ans) les données opérationnelles et libérer les systèmes de leurs historiques
- 2) Rendre les données opérationnelles « consommables » (intelligibles) pour les rapports opérationnels
- 3) Permettre l'alimentation des sujets de l'entrepôt
- 4) Réutiliser les efforts de l'ETC et la vue unifiée pour permettre de nouvelles opérations
- 5) « Réconciliation » de l'approche Inmon et Kimball

Fait partie de la Zone de ravitaillement et/ou de l'entrepôt.

ENJEUX / DÉFIS

- **41%** des projets d'entrepôt de données ont échoués.
Cutter Consortium: Curt Hall, Corporate Use of Data Warehousing and Enterprise Analytic Technologies, **January 2003**.
Gartner 2003/2007: 50% sera un échec ou pas accepté
- **60-70%** des projets d'entrepôt de données échouent.
TechTarget 2016, by Monique I. Cuvevlier: <http://searchoracle.techtarget.com/tip/Why-data-warehousing-projects-fail>

- La non compréhension de la solution
- Le coût de mise en place
- TI priorise les systèmes opérationnels
- Pas d'expertise dans l'entreprise
- La disponibilité des gestionnaires et des utilisateurs
- Les investissements déjà faits pour rendre l'information disponible
- La résistance au changement
- L'implantation de progiciels qui prétendent répondre à ce besoin
- Le mode de gestion de l'entreprise
- Etc...

Les Enjeux

(11 selon Sanjay Raizada de Syntel, Beaverton, Oregon)

1. Reconnaître que le travail sera probablement plus difficile que prévu
2. Comprendre les données des systèmes existants
3. S'assurer de reconnaître les entités équivalentes
4. Utiliser la méta-donnée pour assurer la qualité des données
5. Sélectionner le bon outil de transformation des données
6. Prendre avantage des sources externes
7. Utiliser les nouvelles méthodes de distribution de l'information
8. Focus sur les applications marketing plus payantes
9. Mettre l'emphase sur les GAINS RAPIDES pour cultiver l'enthousiasme de toute l'organisation
10. Ne pas sous-estimer les besoins hardware
11. Considérer de donner à l'externe le développement et la maintenance de l'entrepôt.

Top 10:

1. La nouvelle technologie réponds à tous les besoins
2. On a parlé aux utilisateurs – mais pas les bons
3. Beaucoup de recherche, beaucoup de recherche, beauc...
4. Un beau projet noyé par la métadonnée
5. Projet détourné par des requêtes «Il serait amusant de savoir... »
6. Adoption d'un DSS sans supporter les décisions
7. TI interne plus cher que les consultants
8. Pas de promotion interne
9. Ne pas reconnaître le cycle de vie d'un outil d'Aide à la Décision
10. Emphase sur le développement et peu sur le déploiement



AUTRES SOLUTIONS INFORMATIONNELLES

Autres outils informationnels

- **Gestion des connaissances (KM)**

- **Définition :**

Gestion, par des moyens informatiques, des informations significatives qui sont acquises par une entreprise et qui y circulent, ainsi que du savoir-faire développé par le personnel, de manière à créer un système interactif de formation maison continue qui débouche sur une meilleure qualité des produits et services, ainsi que sur une plus grande compétitivité de l'entreprise.

- **Gestion Relation Client (CRM)**

- **Définition :**

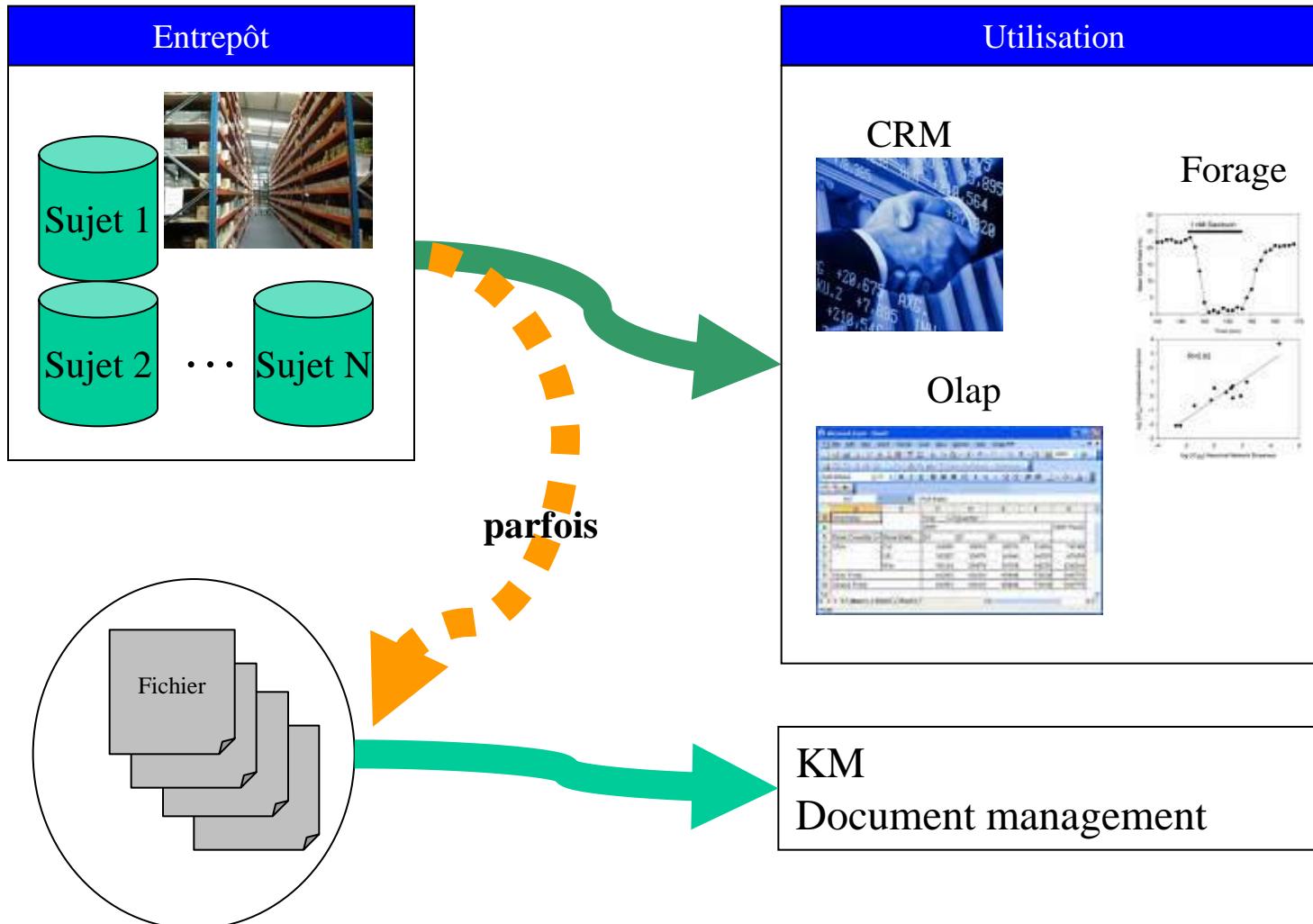
Outil informatique développé expressément pour permettre à une entreprise de fidéliser ses clients et d'accroître sa part du marché, en intégrant la gestion des données relatives aux besoins et aux attentes du client, dans le contexte de la vente et des services après-vente, en ligne ou non.

- **Gestion documentaire – recherche de textes/concepts**

- **Environnement de collaboration**

- **Mégadonnées (« Big data »)**

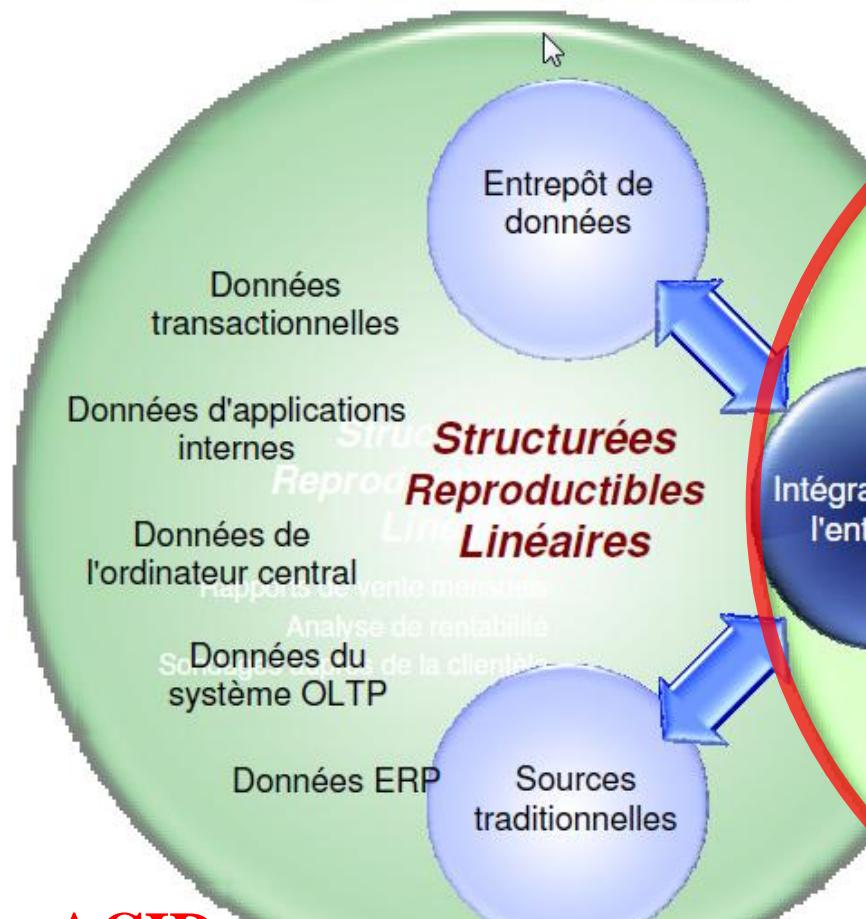
Outils informationnels (structure)



Le « Big Data »

Approche traditionnelle

Structurée, analytique, logique

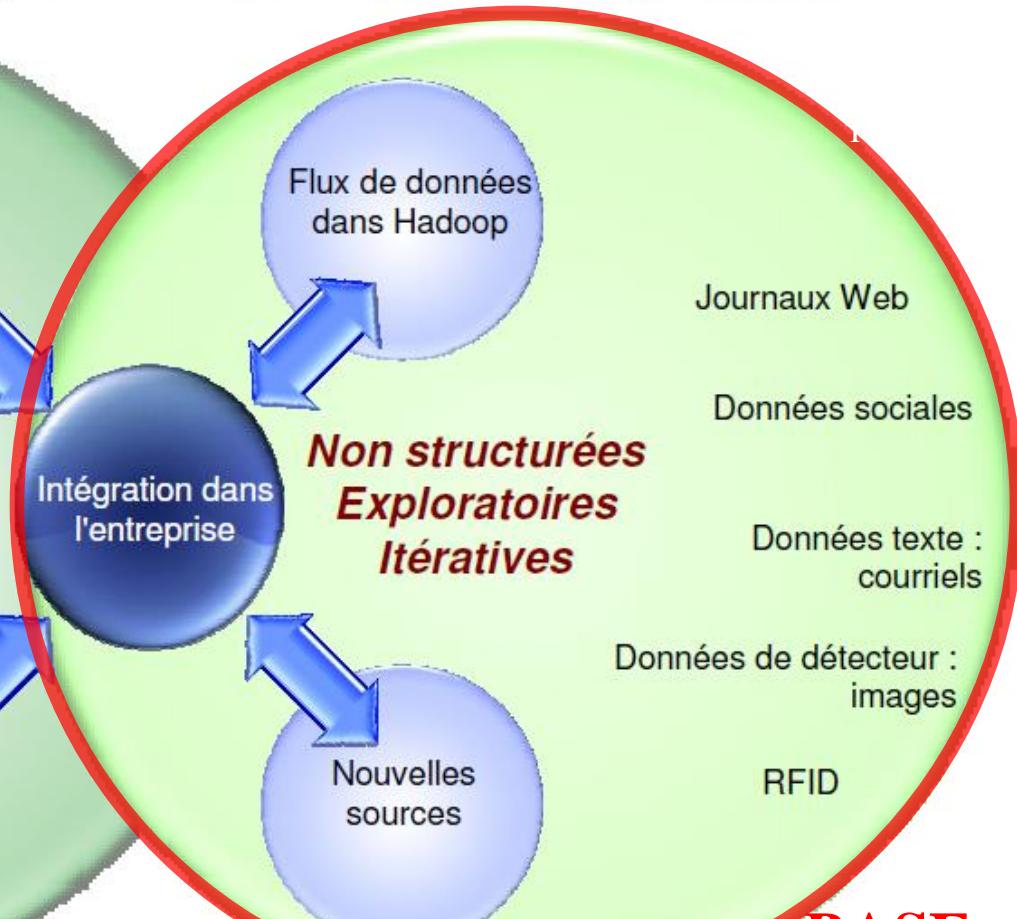


ACID

BD: Relationnel standard

Nouvelle approche

Créative, pensée holistique, intuition



BASE

BD: NoSQL/Hadoop

- Les ERP / MRP offrent des solutions au problème d'entrepôt de données sous forme de B.I.
 - [SAP](#) (document annexé)
 - [Oracle](#)
- Les SGBD, comme Oracle, offrent des outils de réPLICATION et agrégation automatisé vers un zone de données stratégique (entrepôt)