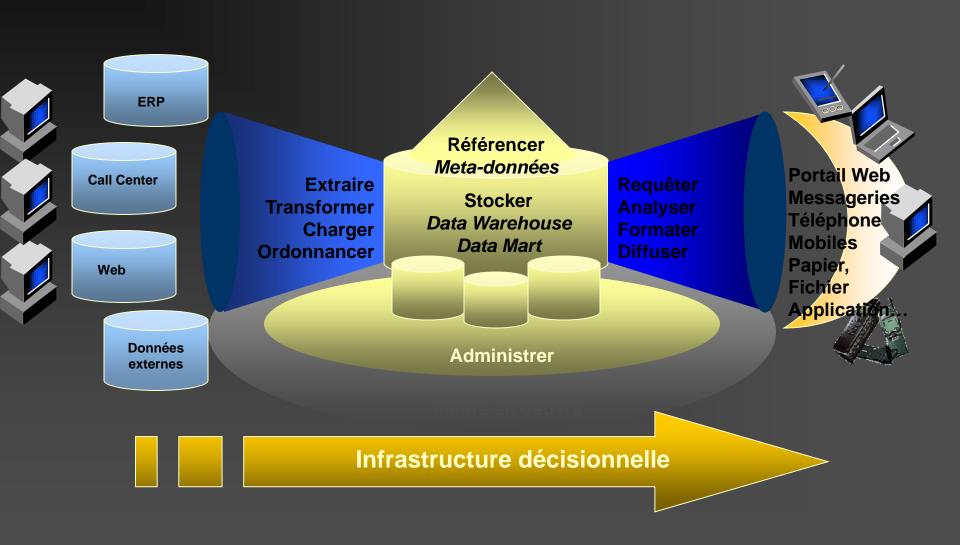
Modelisation des Systèmes Décisionnels 2ème Partie Julien DUFOUR

- Rappels
- Métadonnées
- ETL
- Exemple et Réflexion
- Conclusion

Rappels

- Grand Principe : Stocker l'information à des fins d'Analyse
- Privilégier les performances des requêtes

Architecture



Définition du DWH

 « Le DataWareHouse est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision »

Bill Inmon, « Using the Data Warehouse »

- Orientées Sujet
- Intégrées
- Non Volatiles
- Historisées
- Organisées
- Support
- Objectif : Aide à la décision

Mots clé

- Orientées sujet :
 - Données structurées par thème (orientées métier)
 - Notion de datamart : approche incrémentale contient des données du DWH pour un secteur particulier de l'entreprise (Ex : datamart Marketing)
- Intégrées
 - Définition
 - Sources / Mapping
 - Unification : tables de correspondance ?
 - Qualification : données manquantes ?

Mots clé

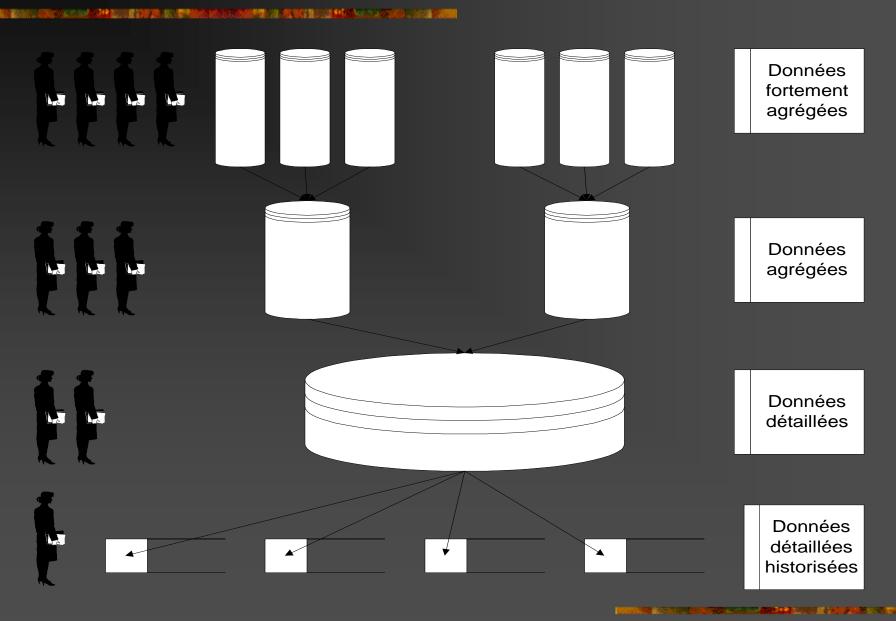
Historisées

- Système décisionnel : les données de détail ne sont jamais mises à jour et on conserve l'historique des changements.
- Associer un référentiel de temps : les données sont datées

Non volatiles

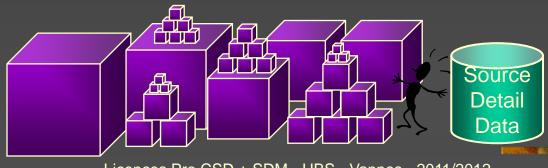
- Conséquences de l'historisation
- Les données déjà stockées dans le DWH ne sont jamais modifiées

Les 4 classes de données : Agrégation



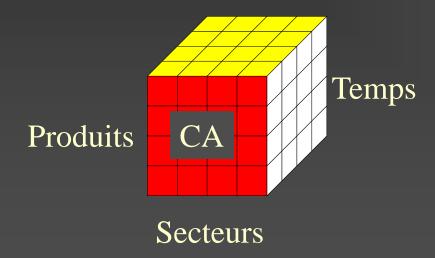
Le cube

- Sur l'ensemble des dimensions, on pré-calcule à tous les croisements la valeur des indicateurs (NWAY)
- Outils spécifiques pour créer et accéder aux données
- Intégration de la notion de hiérarchie
- Structure spécifique
- A quoi ça sert ?

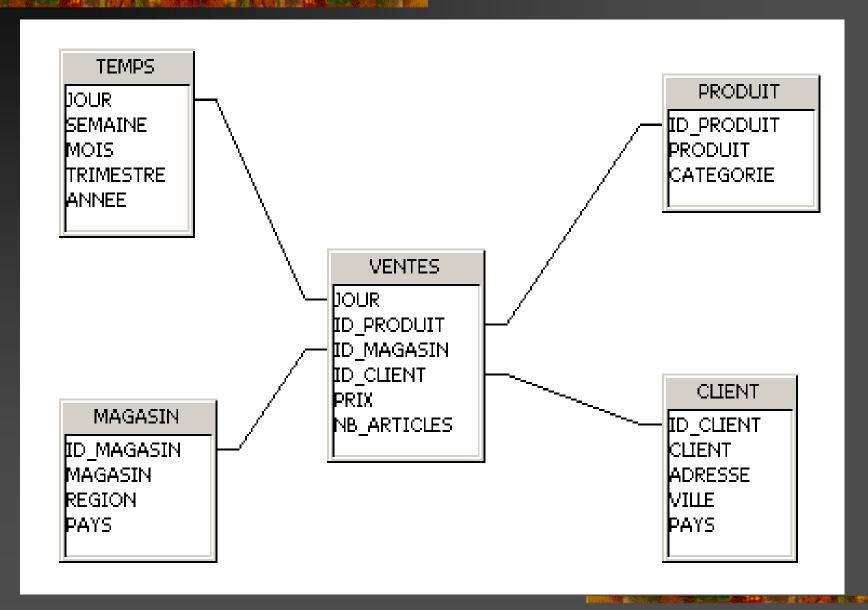


Mots clés Cube

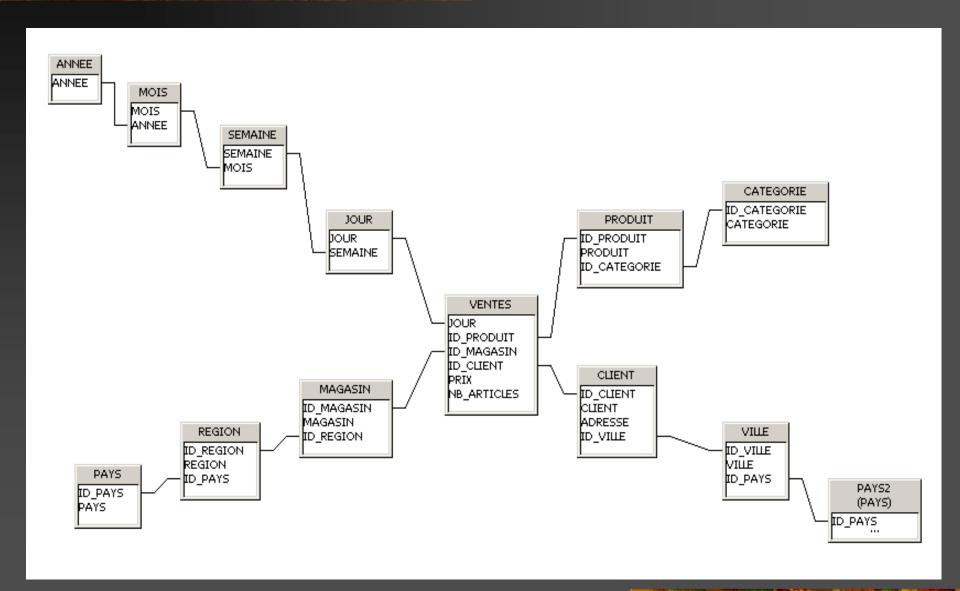
- OLAP
- Dimension et hiérarchie
- Nway
- Cellule



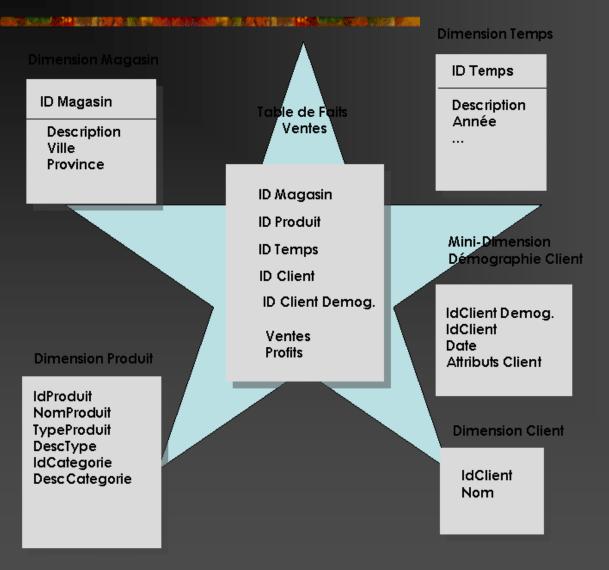
Modélisation Décisionnelle : schéma en Etoile



Modélisation Décisionnelle : schéma en Flocon

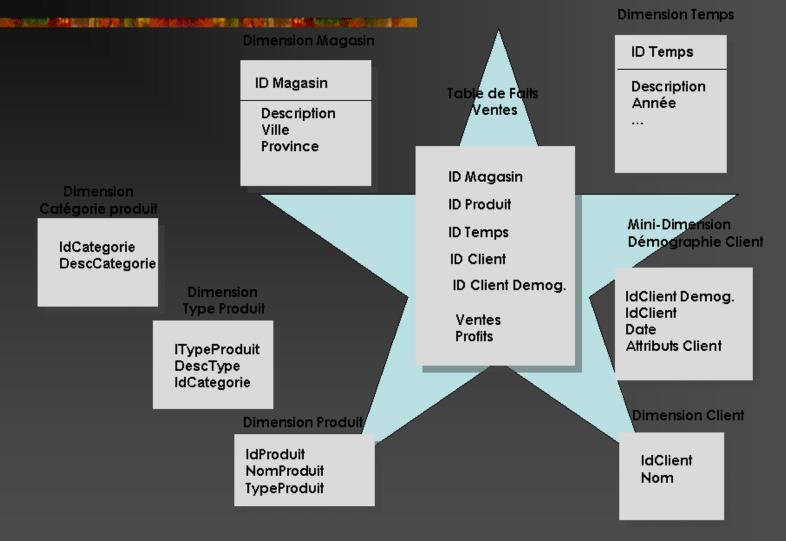


Modélisation Décisionnelle : schéma en Etoile





Modélisation Décisionnelle : schéma en Flocon





Différences entre les schémas : Etoile vs Flocon

	Etoile	Flocon
Dimensions	dénormalisées	normalisées
Lecture	simple	complexe
Relations		Si relation 1 à n entre les différents niveaux des dimensions Obligatoirement si relation n à n entre table de faits et dimensions
Besoins d'affaires		Banque (relation client-compte) ou Assurances (relation client-police)
Performance requêtes	le plus performant (moins de jointures)	moins performant
Ratio Attributs maître dans les dimensions		beaucoup d'attributs et ratio faible (par exemple dimension commandes avec 25 attributs communs par commande et une commande qui contient en moyenne 50 produits (ratio de 1:50)
Technologie	Cognos exclusif	MicroStrategy exclusif

Codification et normalisation des libellés

Type d'information stockée

	Denomination	Description
PK		Clé primaire.
FK		Clé étrangère (issue de la jointure avec une autre table de données du modèle).
IF		Information.
MS		Mesure.
His		Information qui engendre une historisation de la ligne

Nom

Normalisation des noms des tables

· · · · · · · · · · · · · · · · · · ·	
Table de Faits	FAIT_xxxx
Table de Faits aggrégés	AGG_xxxx
table de dimension	DIM_xxxx
Table de référence	REF_xxxx

Normalisation des noms des variables

Туре	Nom
Code	Code_xxxx
Libellé	Lib_xxxx
Nombre	Nb_xxxx
Montant	Mt_xxxx
Date	Dat_xxxx
Heure	Hr_xxxx
Score	Sc_xxxx
Booléen (ex:Oui/Non)	B_xxxx
Numéro	Num_xxxx

Les Méta Données

- Définition
- Contenu

Définition des Métadonnées

- Information décrivant une donnée
- C'est une donnée sur une donnée
- On parle aussi de dictionnaire de données
- Elle qualifie une donnée en précisant
 - Sa sémantique
 - Les règles de gestion associées
 - Sa source
 - Son format
 - Son utilisation

Information sémantique

- Que signifie cette donnée ?
 - Chiffre d'affaires, Client
 - Taux de sinistralité
- A quoi sert-elle ?
- Qui en a besoin ?

 Informations destinées aux Directions Fonctionnelles

Information d'Origine

- D'où vient cette donnée ?
- Comment et par qui est-elle mise à jour ?
- Sa date de création ...

Mapping des Informations

Règles de Gestion

- Les règles de calcul et les règles d'agrégations
 - Comment calcule-t-on cet indicateur ?
 - Quel est le périmètre de consolidation ?
- Il faut donc avoir bien fait les 2 étapes précédentes : Sémantique et Origine
- Exemples :
 - Comment calcule-t-on la marge brute ?
 - Pour calculer le CA Européen : quelle est la liste des pays (CEE, Zone Euro?)

Le format

- Sous quel format la donnée est-elle restituée ?
 - Échelle (€, K€…)
 - Unité (€, \$...)
 - Domaine de valeurs (1=F, 2=H...)
- Si l'indicateur est utilisé par différentes entités fonctionnelles :
 - Comment la stocke-t-on?
 - Quelles sont les règles de correspondance ?

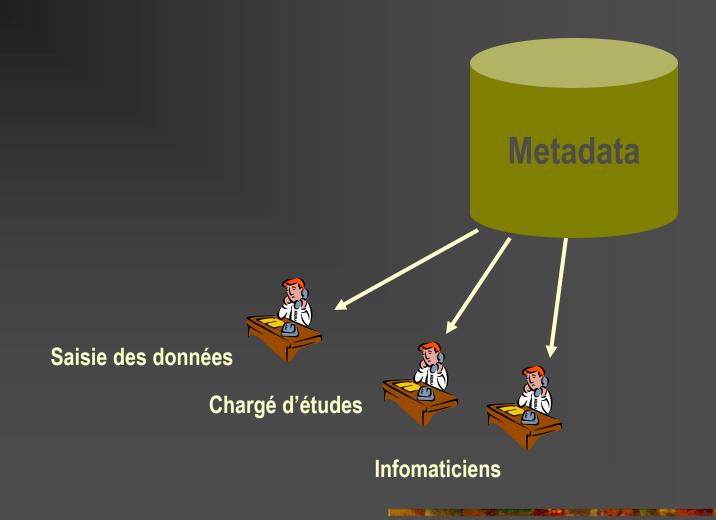
Utilisation

 Utilisation dans le cadre du DWH (et non « utilisateurs »)

- Permet une maintenance plus aisée
 - Liste des programmes
 - Liste des autres variables l'utilisant
 - Notion d'Impact

Les utilisateurs des métadonnées

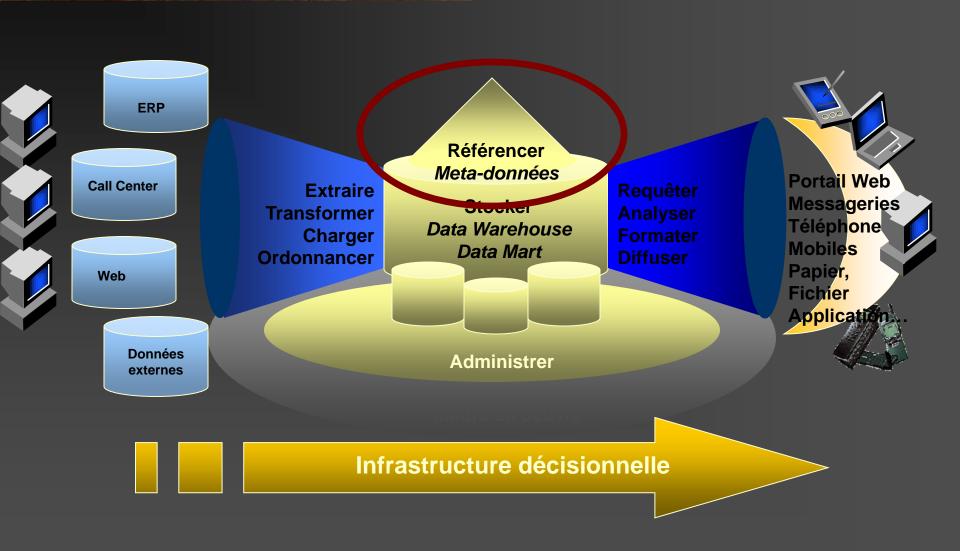
Qui voit les métadonnées ?



Le stockage

- Sur fiche/papier.
 - Peu utilisable, peu de diffusion
 - Difficilement maintenable
- Sous formes de bases de données.
 - Elles sont souvent exprimées via un MCD/MLD.
 On obtient ainsi une vision synthétique de l'information stockée et donc du DWH.
- Dans la définition même du DataWarehouse

Architecture

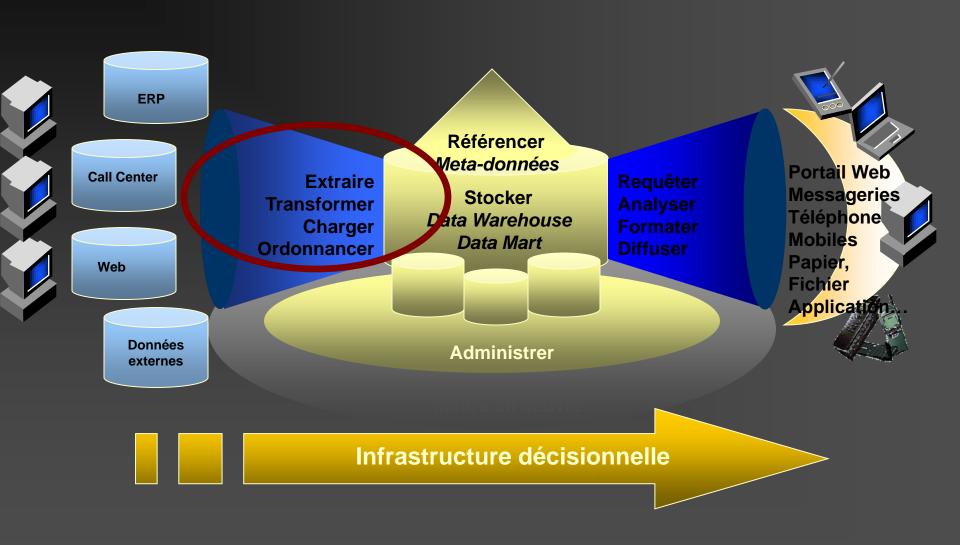


ETL

La Phase

Les Produits

Architecture



ETL

- Phase d'extraction, transformation et de chargement
 - E : Extract
 - T: Transform
 - L: Load
- Lien avec les métadonnées :
 - Les règles de gestion associées
 - Les sources
 - Le format
 - L'utilisation

ETLs: les produits

- SAS / WareHouse Administrator (SAS/WA)
- Oracle WareHouse Builder
- Sunopsis
- DataStage
- Informatica

Exemple et Réflexions

Définition du besoin

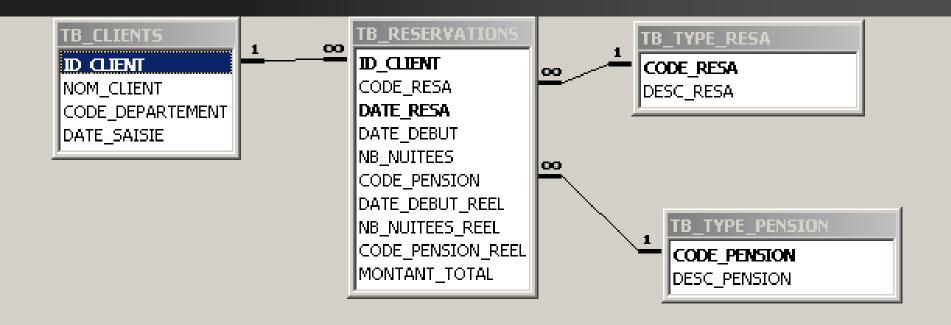
 Un hôtel de luxe possède 3 chambres. Son but est d'avoir au moins 10 nuitées par mois.

- Le responsable de l'hôtel aimerait mettre en place une campagne marketing, mais ne sait pas :
 - Où (national, régional, local)
 - Comment (type de médias)
 - Quand

Pourquoi un exemple ...

- L'exemple nous servira de mise en évidence des notions suivantes :
 - Métadonnées
 - Mapping
 - Intégration
 - Restitutions

Modèle Opérationnel



Conclusion

- Analyse des besoins
- Définitions des indicateurs
- Mapping : Intégration
- Métadonnées
- Privilégier les performances des requêtes