
Objectifs des entrepôts de données

2/1

Objectifs des entrepôts

Objectifs du cours

- Comprendre l'importance de l'information dans les métiers
- Analyser les limites des systèmes de production existants pour les besoins en traitement de l'information
- Introduire le concept d'entrepôt de données comme une solution alternative
- Introduire les propriétés des entrepôts

2/2

Objectifs des entrepôts

Structure du cours

Importance de l'information

Inadéquation des systèmes OLTP

Concept d'entrepôt

propriétés d'un entrepôt

Comparaisons entrepôt/production

2-3

Objectifs des entrepôts

L'enjeu de l'information

L'information occupe un rôle croissant dans tous les métiers

qualité de service

- traitement personnalité des clients, offres compétitives

Gestion

- réduction des coûts, gestion des profits

Prospective

- analyse des comportements des clients, du marché

Communication

- informer les individus

2-4

Objectifs des entrepôts

Aide à la décision

L'information est à la base du cycle

- information - analyse - prise de décision

La prise de décision nécessite une information

- précise
- fiable
- actualisée
- pertinente

Information vs données

Données

- montant total des ventes pour région Paris
- vendeur ayant réalisé le meilleur chiffre ce mois

Information

- évolution des ventes pour région Paris au cours des 5 dernières années
- sur quels produits faire des offres promotionnelles
- quelle est la rentabilité d'une activité

Quelles différences voyez-vous ?

Gestion des données

Systèmes « Online Transaction Processing » (OLTP)

- comptabilité, achats, réservation, télécommunications, ...
- systèmes stratégiques, haute performance et disponibilité

Multitude de systèmes spécialisés

- fichiers Excel, bases personnelles, documents, ...
- systèmes autonomes, non stratégiques

2.7

Objectifs des entrepôts

Systèmes OLTP

Priorities	Performance, high availability
Processor Use	Predictable
Response Time	Sub-seconds to seconds
Data Store	Hierarchical, network, relational, flat files
Data Content	Organized by application
Nature of Data	Dynamic, constantly changing, current state of the business
Processing	Highly structured, repetitive
End Users	Clerks, data entry, administrators

2.8

Objectifs des entrepôts

Limites des systèmes OLTP

Les systèmes OLTP sont mal adaptés à la gestion d'information pour l'aide à la décision

Problèmes :

- Accès aux données difficile
- Extractions de données non productives
- qualité des données incertaine

Accès aux données

Données structurées pour applications

- tables normalisées (performance transactionnelle)
- valeurs d'attributs codées
- attributs spécifiques pour la production

Données dans des systèmes indépendants

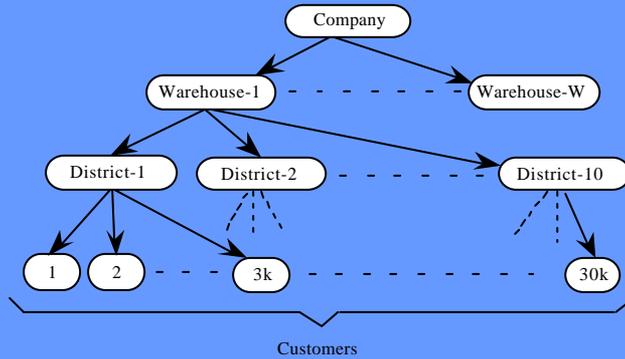
- systèmes hétérogènes (protocoles réseau, systèmes de gestion, modèles de données)

Requêtes simples

- incompatibilité (performances) avec requêtes décisionnelles

Exemple OLTP : base de données TPC-C

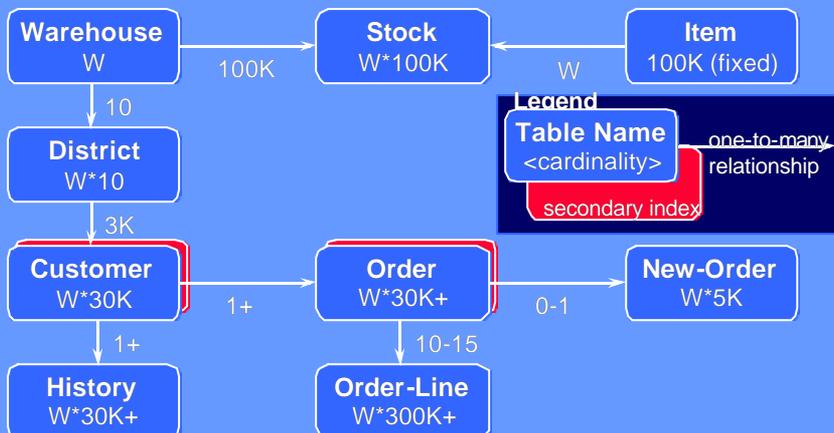
Application: gestion, vente et distribution de produits ou services (www.tpc.org/bench.descrip.html)



2-11

Objectifs des entrepôts

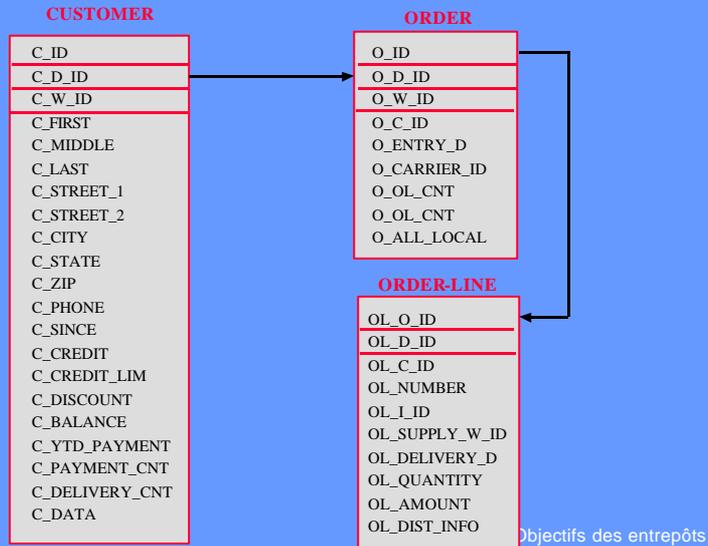
Vue d'ensemble du schéma



2-12

Objectifs des entrepôts

Une vue détaillée du schéma



2-13

Transactions et requêtes

Transactions TPC-C :

- New-order: enter a new order from a customer
- Payment: update customer balance to reflect a payment
- Delivery: deliver orders (done as a batch transaction)
- Order-status: retrieve status of customer's most recent order
- Stock-level: monitor warehouse inventory

Exercice : exprimer des requêtes décisionnelles sur ce schéma et discuter les problèmes rencontrés

2-14

Objectifs des entrepôts

Requêtes décisionnelles

Extraites de TPC-D:

Retrieve the 10 unshipped orders with the highest value.

report the amount of business that was billed, shipped, and returned.

determines how well the order priority system is working and gives an assessment of customer satisfaction. That is, count the number of orders ordered in a given quarter of a given year in which at least one lineitem was shipped by the customer later than its committed date.

The query lists the count of such orders for each order priority sorted in ascending priority order.

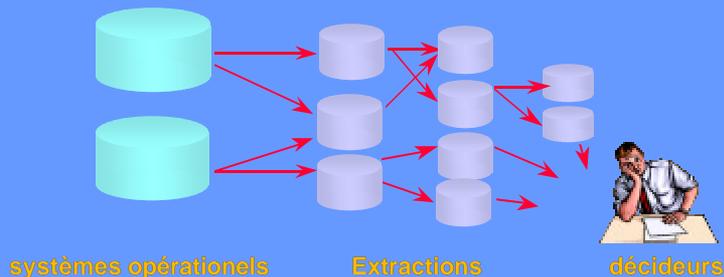
2-15

Objectifs des entrepôts

Extraction de données

Extraire les données pour applis décisionnelles problèmes

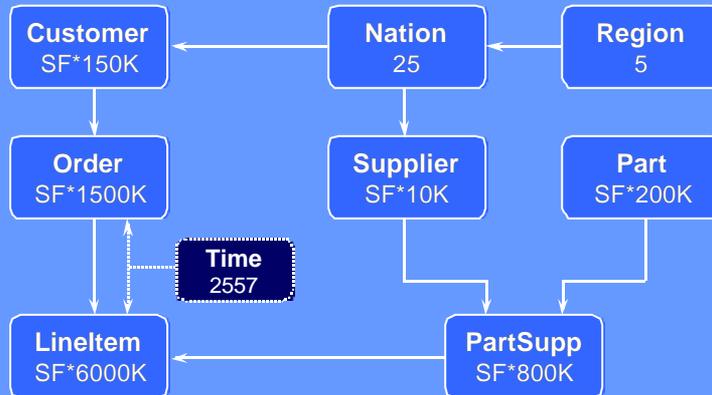
- duplication d'effort dans extractions multiples
- versions incohérentes, obsolètes



2-16

Objectifs des entrepôts

Exemple d'extraction : TPC-D schema



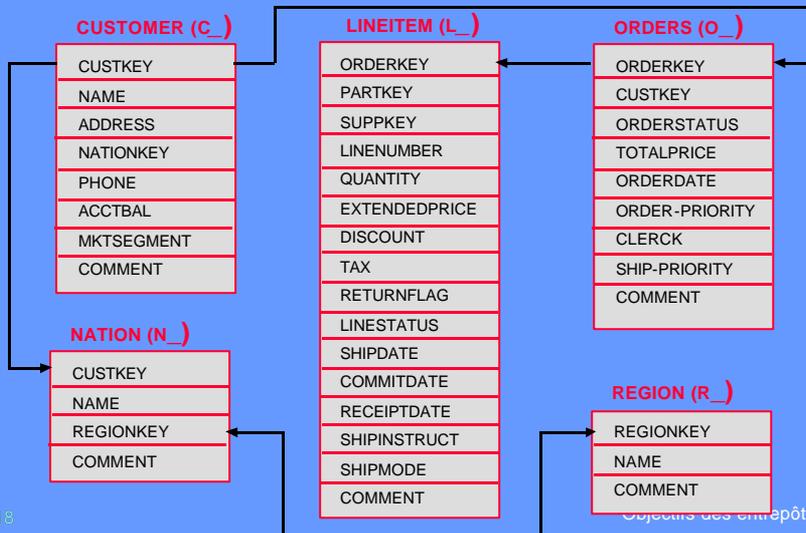
Legend:

- Arrows point in the direction of one-to-many relationships.
- The value below each table name is its cardinality. SF is the Scale Factor.

2:17

Objectifs des entrepôts

Vue détaillée du schéma



2:18

Objectifs des entrepôts

Analyse du schéma

- Quelles informations ont été extraites de TPC-C, quelles transformations ont été apportées ?
- Comparer l'expression des requêtes décisionnelles précédentes avec TPC-C
- Que peut-on dire de ce schéma pour les applications décisionnelles ?

Qualité des données

L'intégration de données extraites de sources multiples et hétérogènes pose des problèmes de qualité des données

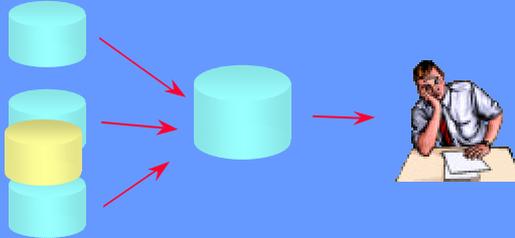
- absence de clé universelle
- standardisation limitée
- données de saisie libre
- champs à valeurs hétérogènes

Quelles notions de « qualité » sont affectées ?

Concept d 'entrepôt de données

Vaste collection centralisée de données

- thématiques
- historisées
- datées
- intégrées



qui offre un niveau de qualité suffisant aux applis décisionnelles

Données thématiques

Les données sont organisées par sujets métier et non par application de production

Exemples :

- client (contrats assurance, prêts, comptes, plans d etc.)
- produit (gamme, ventes, achats, coûts de production, etc.)

Données intégrées

Toutes les données relatives à un sujet métier sont présentées de façon *pertinente, cohérente et non redondante*

L'intégration s'effectue via des processus de transformation des données :

- consolidation
- agrégation
- interprétation

Ces processus doivent être documentés (via métadonnées)

Données datées

Les données de l'entrepôt représentent des clichés successifs du monde réel.

- granularité de temps
- granularité de rafraîchissement
- cohérence des clichés

Données historisées

Les données résident dans l'entrepôt pour une large période de temps.

Ajout successif d'incrément de données

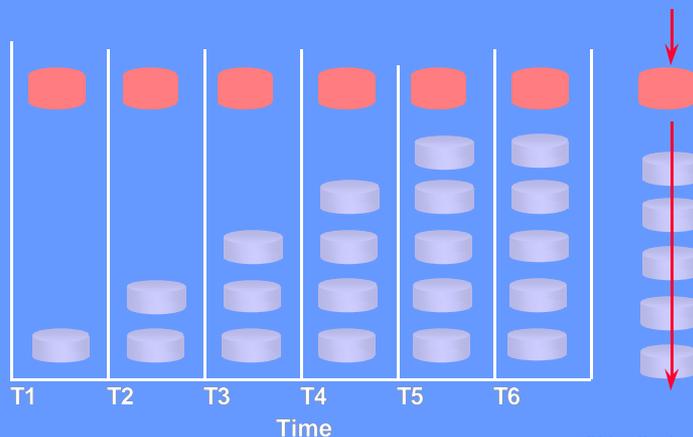
- mises à jour ou suppression rares
- chargements successifs
- archivage des données trop anciennes

2-25

Objectifs des entrepôts

Clichés vs séries chronologiques

- Les systèmes opérationnels donnent des clichés successifs.
- Les entrepôts offrent une série chronologique.



2-26

Objectifs des entrepôts

OLTP vs entrepôt

Property	Operational	Warehouse
Response Time	Sub seconds to seconds	Seconds, to hours
Operations	DML	Read only
Nature of Data	30-60 days	Historical 2-10 years
Data Organization	Application	Subject, time
Size	Small to large	Large to very large
Data Sources	Operational, Internal	Operational, Internal, External
Activities	Processes	Analysis