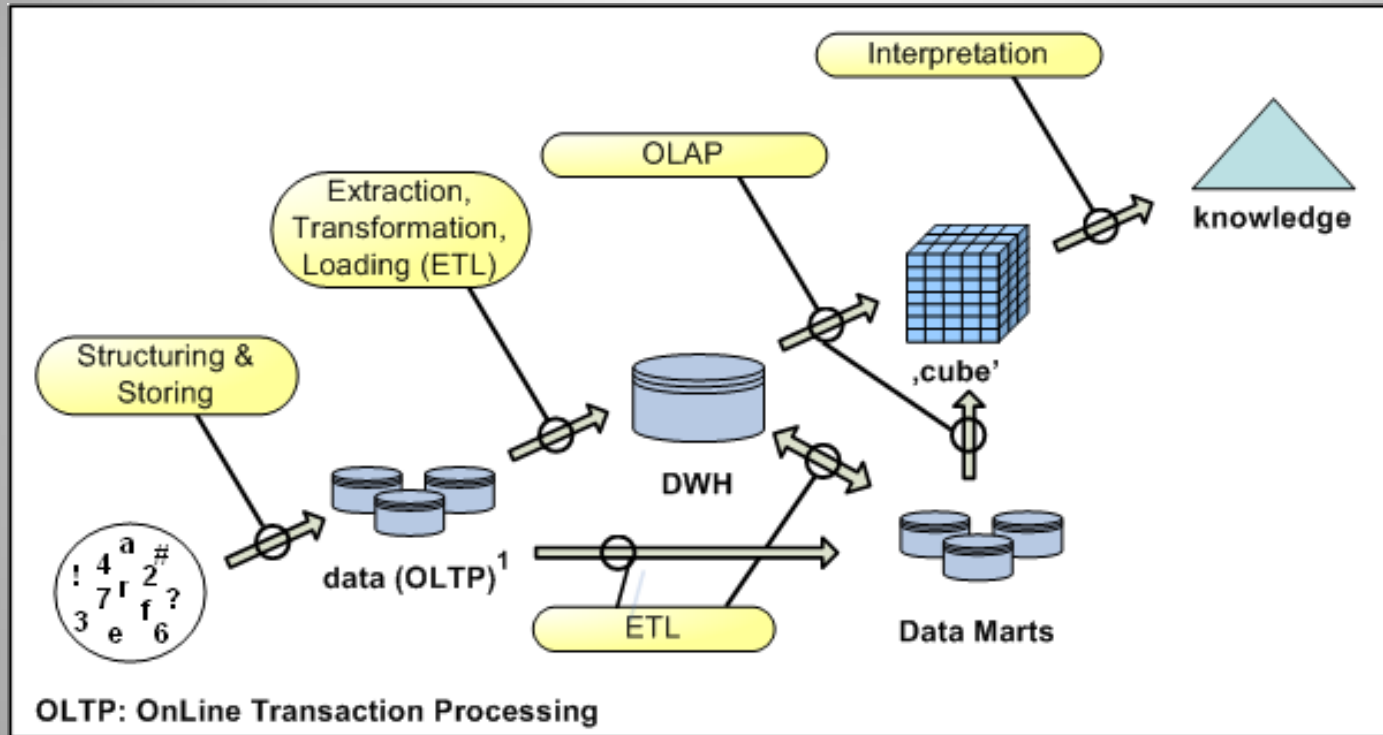


Aide Mémoire 2016 BI –OLAP

André Van Kerrebroeck



Partially based on European Intensive Program : MissLogo 2008-2011

Table of Contents

1. Introduction à l'informatique décisionnelle (BI)
2. La démarche BI avec OLAP
3. Data Warehouse & Data Mart
4. Principes pour la Conception d'un Data Mart
5. Synthèse avant l'étape de la création d'une base de données multidimensionnelle/ HyperCube

1. Introduction à l'informatique décisionnelle (BI)

1. Le but recherché par la technologie BI
2. Secteurs d'activité utilisant l'informatique décisionnelle
3. Les fonctions principales
4. Le focus du cours BI/SQL

1.1. Le but de la Business intelligence/Informatique décisionnelle:

- ❑ ***Mettre à disposition du métier*** par des méthodes automatiques ou semi-automatiques ***des outils d'analyse de masses de données faciles et rapides à utiliser.***
- ❑ ***Faciliter des décisions en fonction de l'analyse de masses de données explorées***

1.2. Branches et domaines dans lesquels est utilisée l'informatique décisionnelle (%)

Aux USA, en 2010 source Wikipedia.

Analyse de la consommation	26,8	Banque	19,2	Santé, ressources humaines	13,1
Détection de fraude	12,7	Finance	11,3	Marketing direct, collecte de fonds	11,3
Télécommunication	10,8	Assurance	10,3	Sciences	10,3
Éducation	9,9	Publicité	9,9	Web	8,9
Médical	8,0	Produits manufacturés	8,0	Commerce de détail	8,0

1.3. L'informatique décisionnelle comporte plusieurs fonctions:

- Exploration de données, prospection de données selon différents axes critères
- Fouille de données, forage de données, *data mining*
 - *Le data mining est une démarche de recherche de corrélations sans nécessairement des hypothèses de départ. Il permet une autre approche statistique tout en utilisant l'apport de la science statistique classique.*
 - Un projet de fouille de données implique d'adopter une bonne méthodologie, il y a plusieurs standards dont SEMMA, CRISP-DM
- Au terme, plus globalement : extraction d'un savoir ou de connaissances à partir de grandes quantités de données.

1.4. Le focus du cours BI/SQL

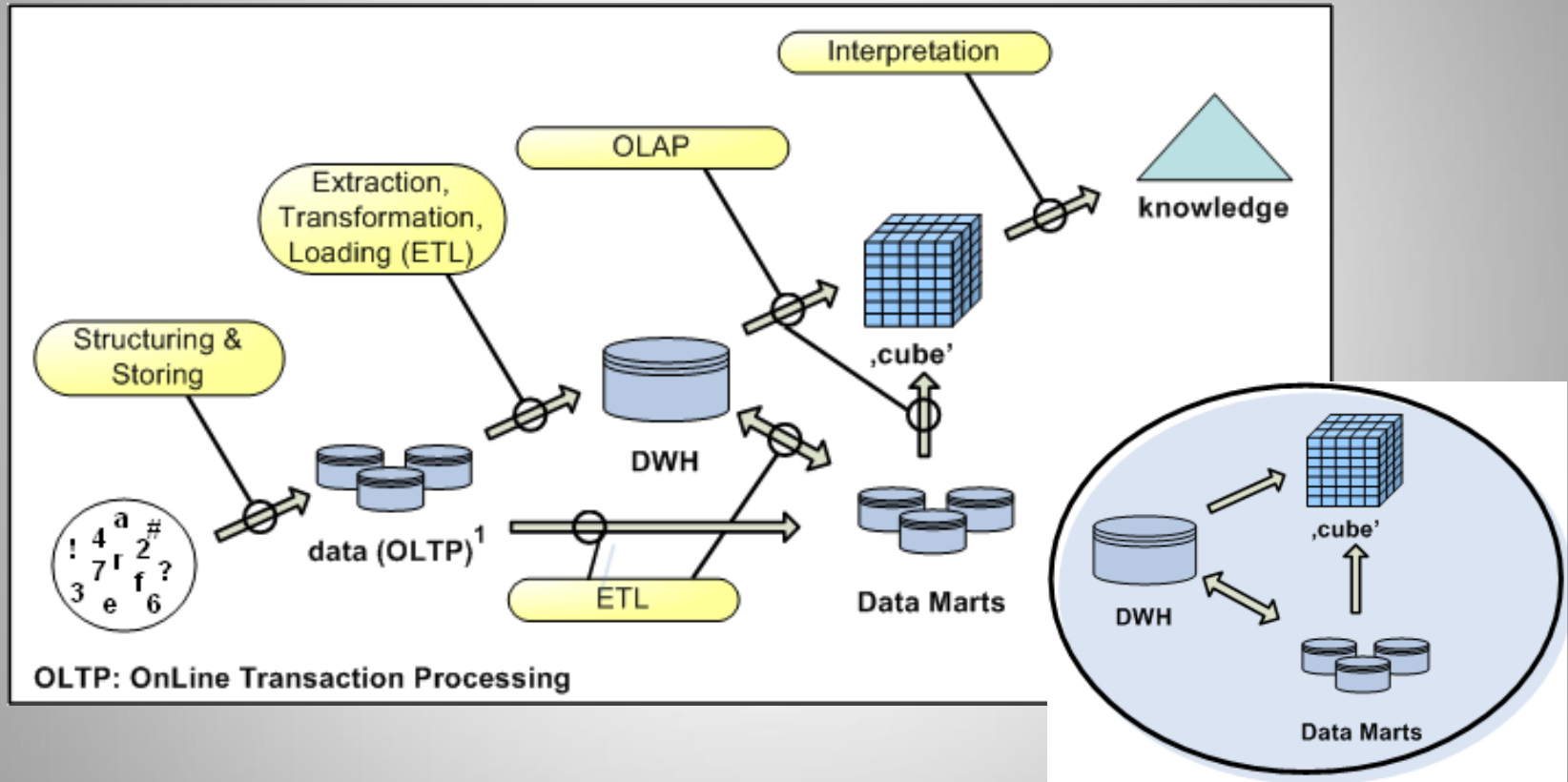
- ❑ Le focus du cours BI/SQL est centré sur la technologie OLAP (Online Analytical Processing) limité à la problématique de
 - ❑ Mise à disposition de données sous forme de DataWareHouse, DataMart
 - ❑ Mise en place de bases de données multidimensionnelles permettant une facilité d'exploration, de prospection
 - ❑ Mise à disposition d'outil d'exploitation des données multidimensionnelles mises en place.
- ❑ Le Data Mining mettant en jeu des compétences statistiques est abordé dans la seconde partie du cours BI.

2. La démarche BI avec OLAP

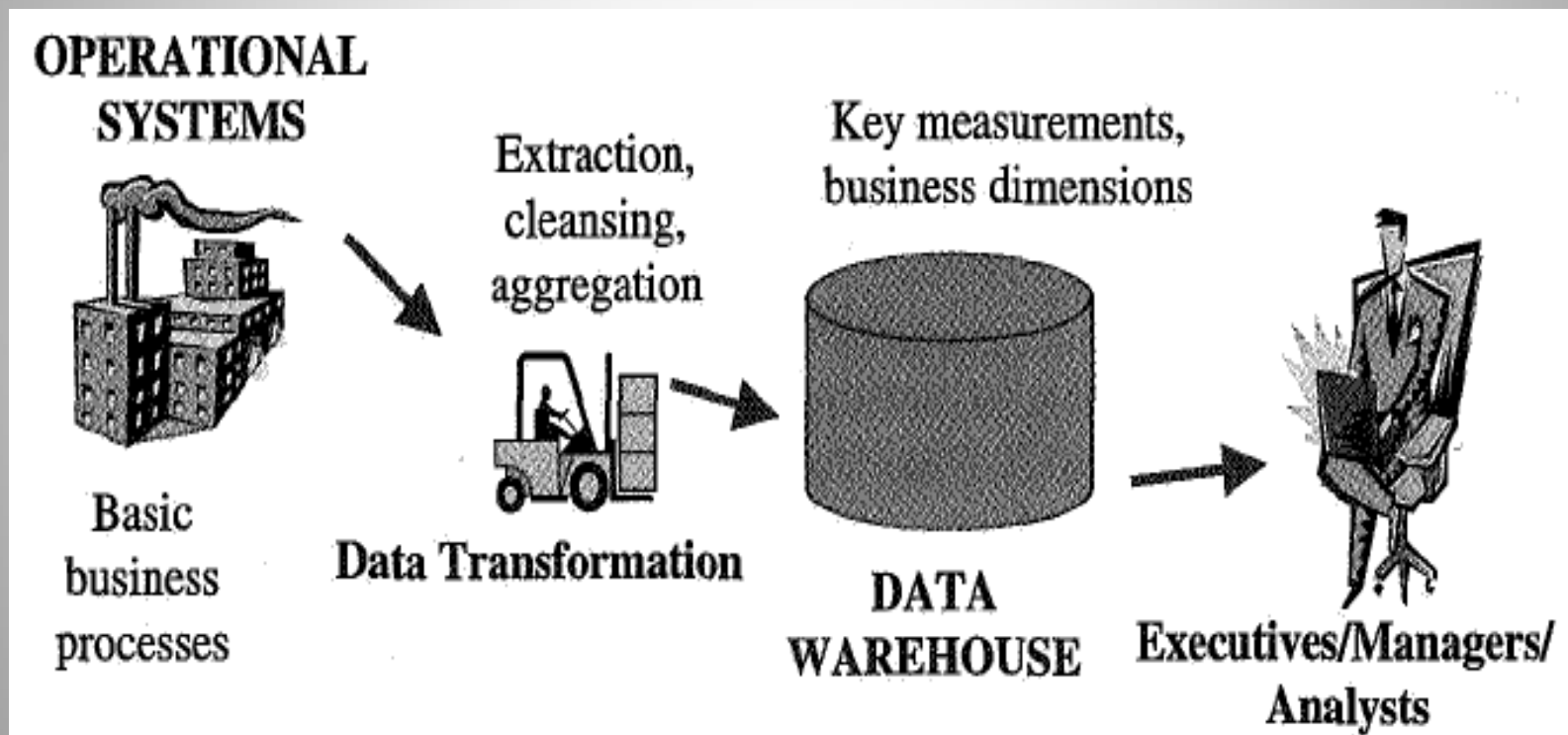
1. Aperçu de la démarche

2. OLAP versus BigData

2.1. Aperçu de la démarche BI/OLAP



OLAP, autre vue:



2.2. Les étapes de la démarches BI/OLAP:

- Identifier les besoins d'information
- Inventorier les sources.
- Construire et implémenter le schéma DataWareHouse/DataMart permettant de répondre aux besoins
- Charger les données dans le DataMart à partir des différentes sources (OLTP et autres sources) Etape **ETL: Extract, Transform and Load**.
- Concevoir et implémenter la base de données multidimensionnelle /HyperCube
- Mettre à disposition des outils d'exploration, de visualisation, d'analyse à partir de l'hypercube.

2.3. DataWareHouse vs BigData:

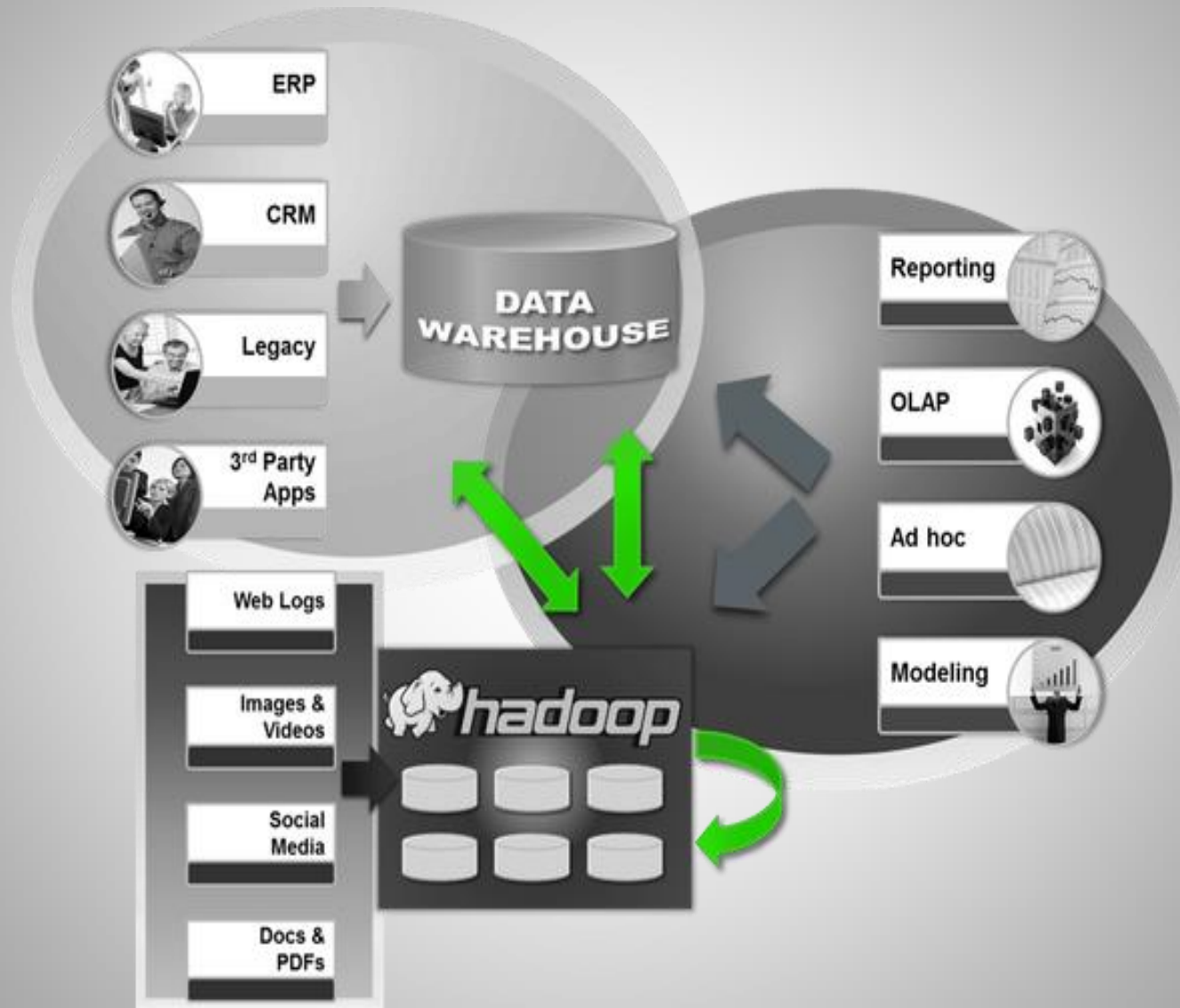
Dans le cas de technologies Big Data, il s'agit de traiter une grande masse de données venant des réseaux sociaux, rapidement et en parallèle. Cela nécessite de grandes ressources (nombre de processeurs, espace de stockage). La technologie OLAP n'est plus suffisante pour traiter les données qui sont fortement non-structurées.

HADHOOP est un des exemples d'implémentation de technologie BIGDATA.

Les entrepôts de données et la technologie OLAP ne sont que des composants de la technologie BIGDATA, ils sont plus directement destinés au domaine métier de l'entreprise ou du service public.

Les technologies OLAP sont implémentées principalement dans le cadre de solutions SGBD: SQL SERVER, ORACLE, SAS,....

Schéma BIGDATA *hadoop*



3. DataWareHouse et DataMart

- 1. Définitions et caractéristiques**
- 2. Approche KimBall vs. Immon:
DatawareHouse ou DataMart?**

3.1. Definitions et caractéristiques

Definition : Data warehouse

“A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format.”

“ A DWH is a pool of data produced to **support decision making**; it is also a repository of **current and historical data** of potential interest to managers throughout the organization.”

“A data warehouse is a **subject-oriented, integrated, time-variant, non-volatile** collection of data in support of management’s decision-making process.”

Data warehouse characteristics

- ☐ Subject-oriented

- ☐ Organized data

- ☐ Integrated

- ☐ Different sources

- ☐ Non-volatile

- ☐ Data can not be changed

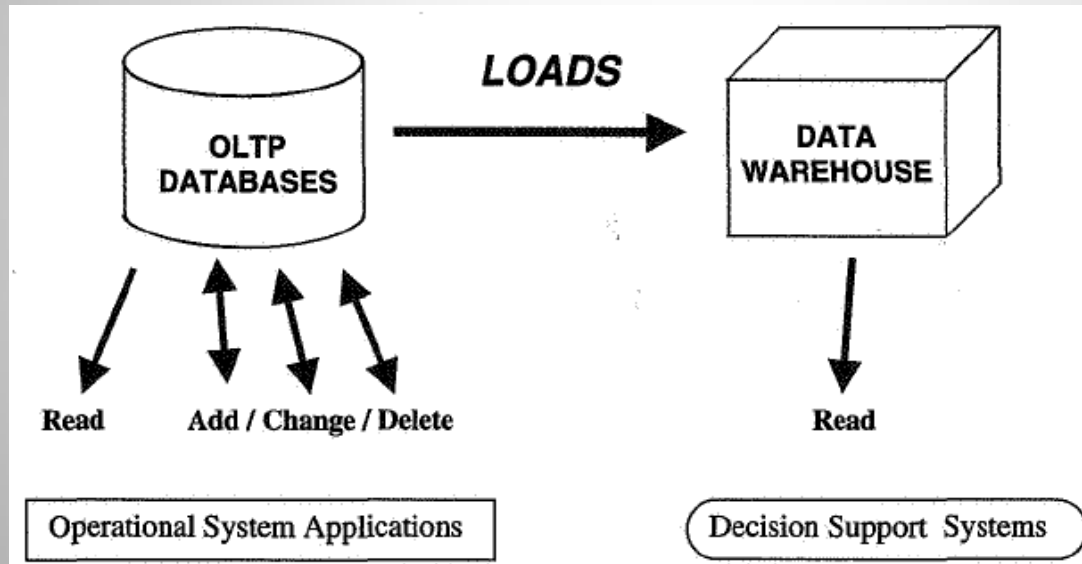
- ☐ Time-variant

- ☐ Historical data

Data warehouse characteristics

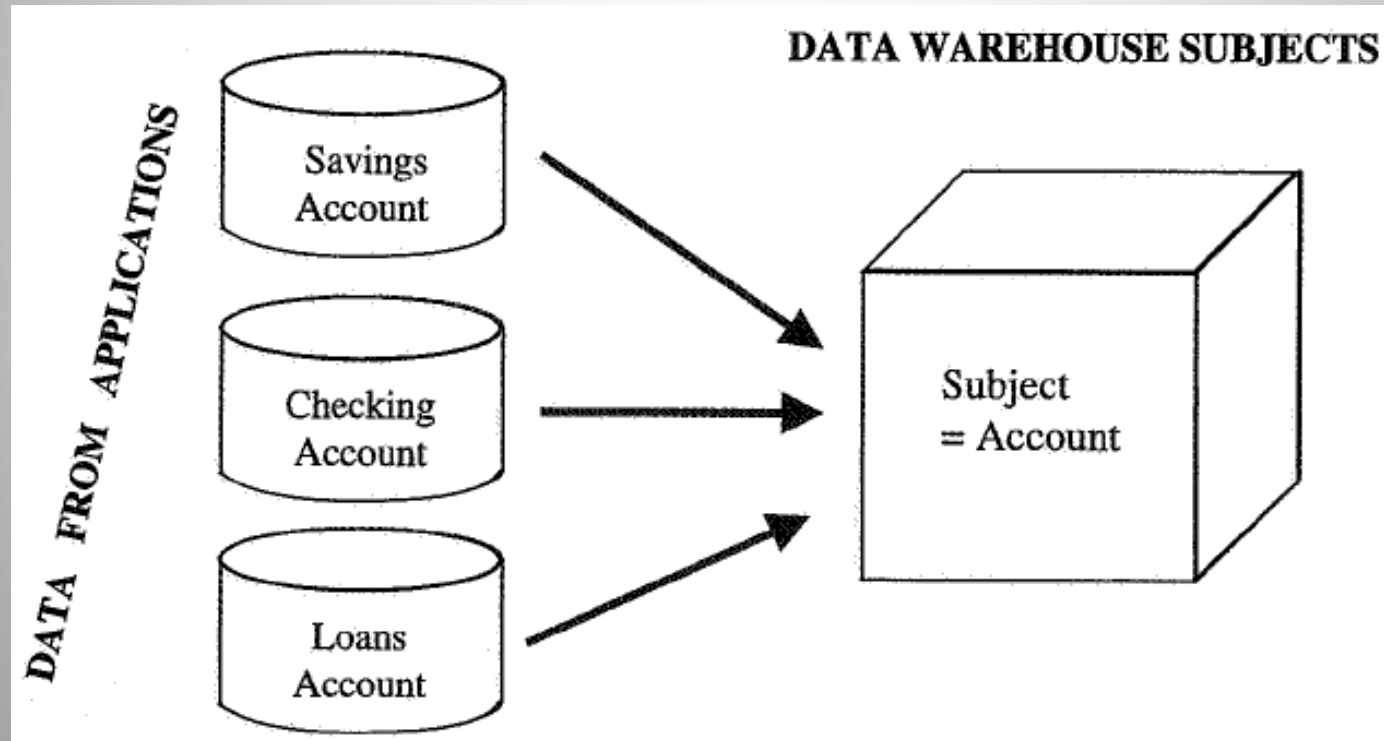
❑ Non-volatile

- ❑ Data can not be changed



Data warehouse characteristics

- ❑ Subject-oriented
- ❑ Organized data



Data Mart

“A *departmental data warehouse* that stores only relevant data.”

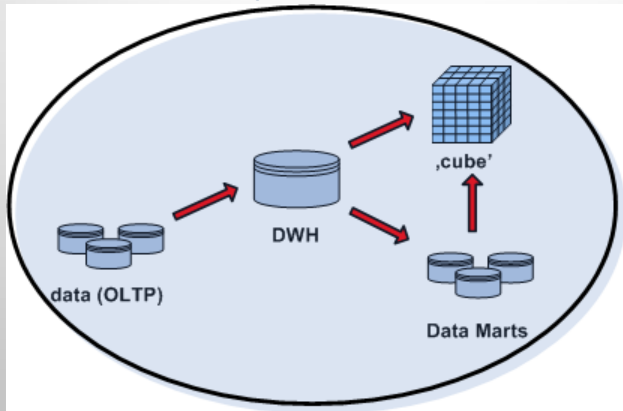
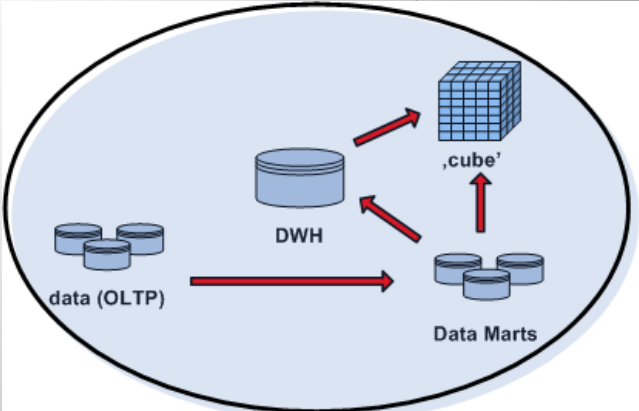
Dependent data mart

“A subset that is created directly from a data warehouse.”

Independent data mart

“A small data warehouse designed for a strategic business unit or a department.”

3.2. Approche KimBall vs. Immon: DatawareHouse ou DataMart?

	William H. Inmon	Ralph Kimball
Movement	Inmonites	Kimballites
DWH model	Entity Relationship (ER) model	Entity Relationship (ER) model
Approach	<p>Top-down</p> 	<p>Bottom-up</p> 

Kimball vs. Inmon

Kimball

'...un Data Warehouse n'est rien d'autre que l'union de tous les Data Marts...'
Défend une approche de construction de sous-ensembles de données pour répondre au fur et à mesure aux questions qui viennent du métier.

Inmon

'Si vous attrapez tous les petits poissons de la mer vous n'obtenez quand même pas une baleine'

Défend une approche où l'on a pensé globalement un entrepôt d'où peuvent venir les réponses à toutes les questions qui viennent du métier.

Kimball vs. Inmon

Inmon Approach:

DWH est généralement utilisé à l'échelle de l'entreprise

La collecte d'information globale à l'échelle de l'entreprise est un processus qui prend du temps

L'approche est peu réaliste sauf dans des solutions ERP largement diffusées

Kimball vs. Inmon

Kimball's approach

Plus compréhensible

Plus facile à concevoir

Mieux pour développer une analyse et obtenir rapidement des résultats.

Nous privilégierons cette approche

Ch 4. Principes pour la conception d'un DataMart

4.1. Les composants d'un DataMart (mini DWH),

4.2. La démarche de la conception

4.3. La conception Star Scheme

4.4. La conception SnowFlake scheme

4.1. les composants d'un DataMart (mini DWH)

Dans la conception d'un DataMart ou DataWareHouse, on doit identifier:

- les faits (Facts)
- les axes (Dimensions) selon lesquels on doit pouvoir interroger les faits
- Les hierarchies dans les dimensions

Fact

A fact is an object of analysis and has both quantifying as well as qualifying aspects. The latter give meaning to the facts, i.e. they provide information

Dimension

A dimension describes the qualifying aspects of facts, i.e. it answers questions like 'Where, when Whom ... did a fact occur?'

Dimension Hierarchy

A (dimension) hierarchy is a ***path of aggregation within a dimension***. The levels of the hierarchy correspond to different degrees of aggregation determining the granularity of the data

Illustration de Faits et Dimensions

Faits : Le nombre d'accidents de la route

Dimensions:

Endroit : dans quels pays, régions,...

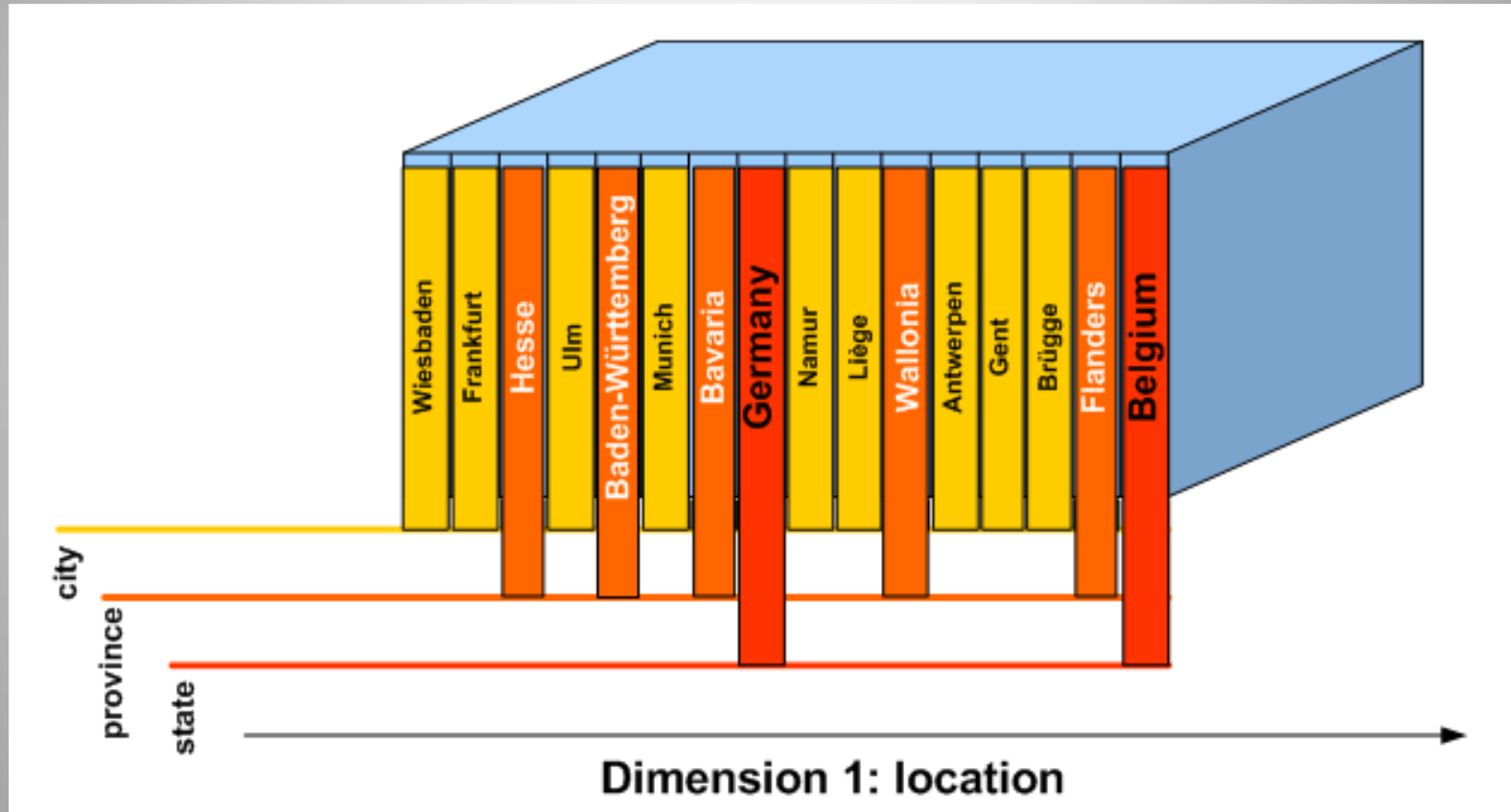
Temps: quand cela s'est passé (année, mois,...)

Cause d'accident: Vitesse, Alcool, Non respect des priorité,...

Categorie d'âge: 18-25, 26-40, ...

.....

Illustration d'une hiérarchie dans la dimension "location"



4.2. La démarche de conception d'un DM/DWH

1. Finding the DM/DWH schema corresponding to the information needs.

There is a choice between two conceptions:

StarSchema

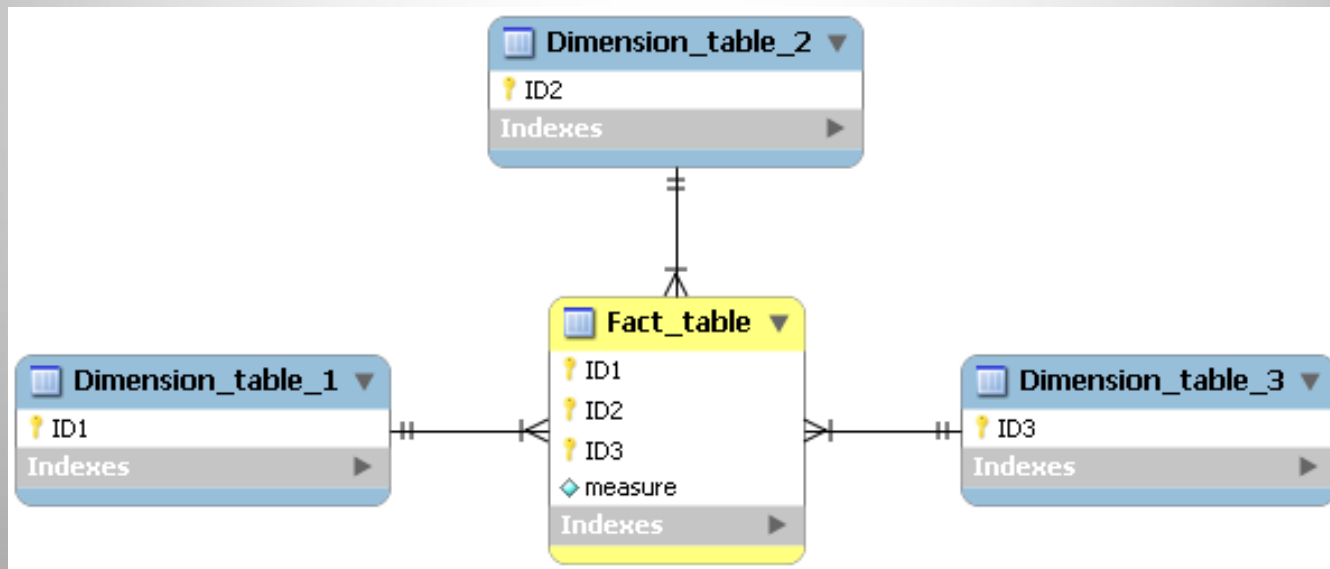
SnowFlakes Schema

2. Locating the dimensions, hierarchies and measures organized in the schema?
3. Asking whether other multidimensional design patterns exist in addition to the schema?

4.3. Star scheme principle

Organizing dimensions and measures

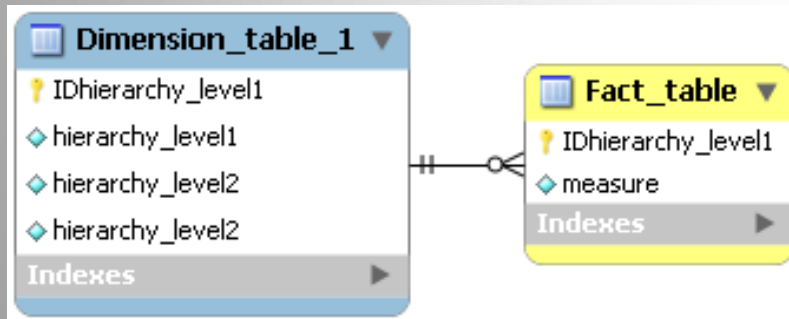
- ❑ Fact table is center of star
- ❑ Dimension tables are tips of star



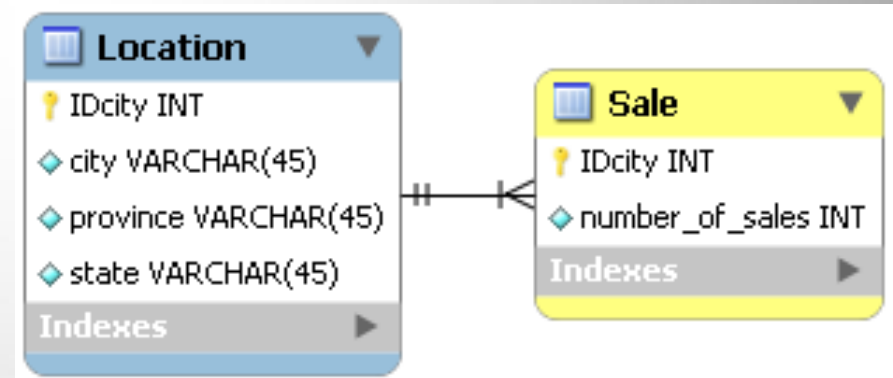
Star scheme principle

Hierarchy levels are either stored in the dimension tables of a star schema...

❑ Dimensions are not in 3NF!



Theory

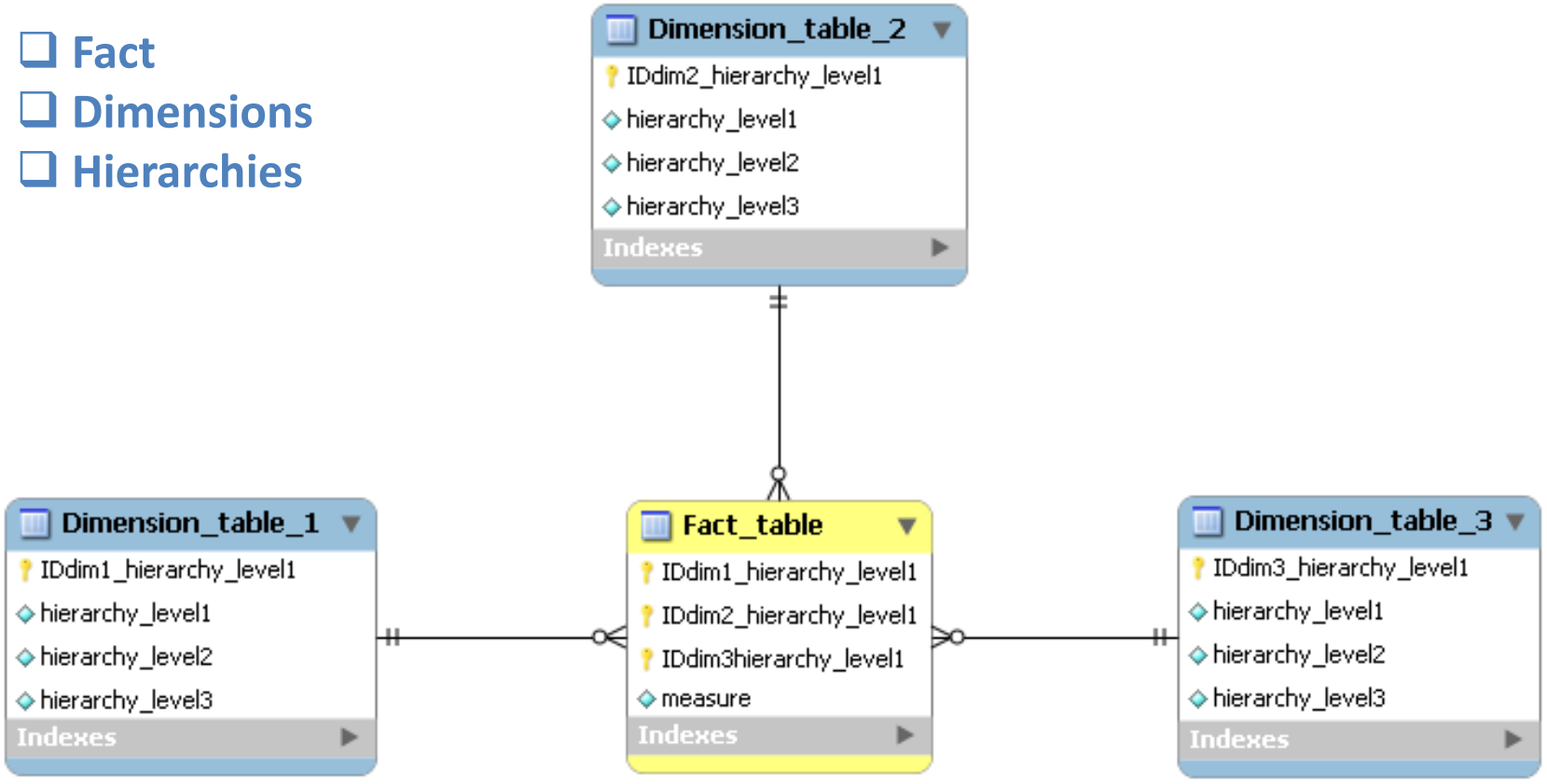


Practice

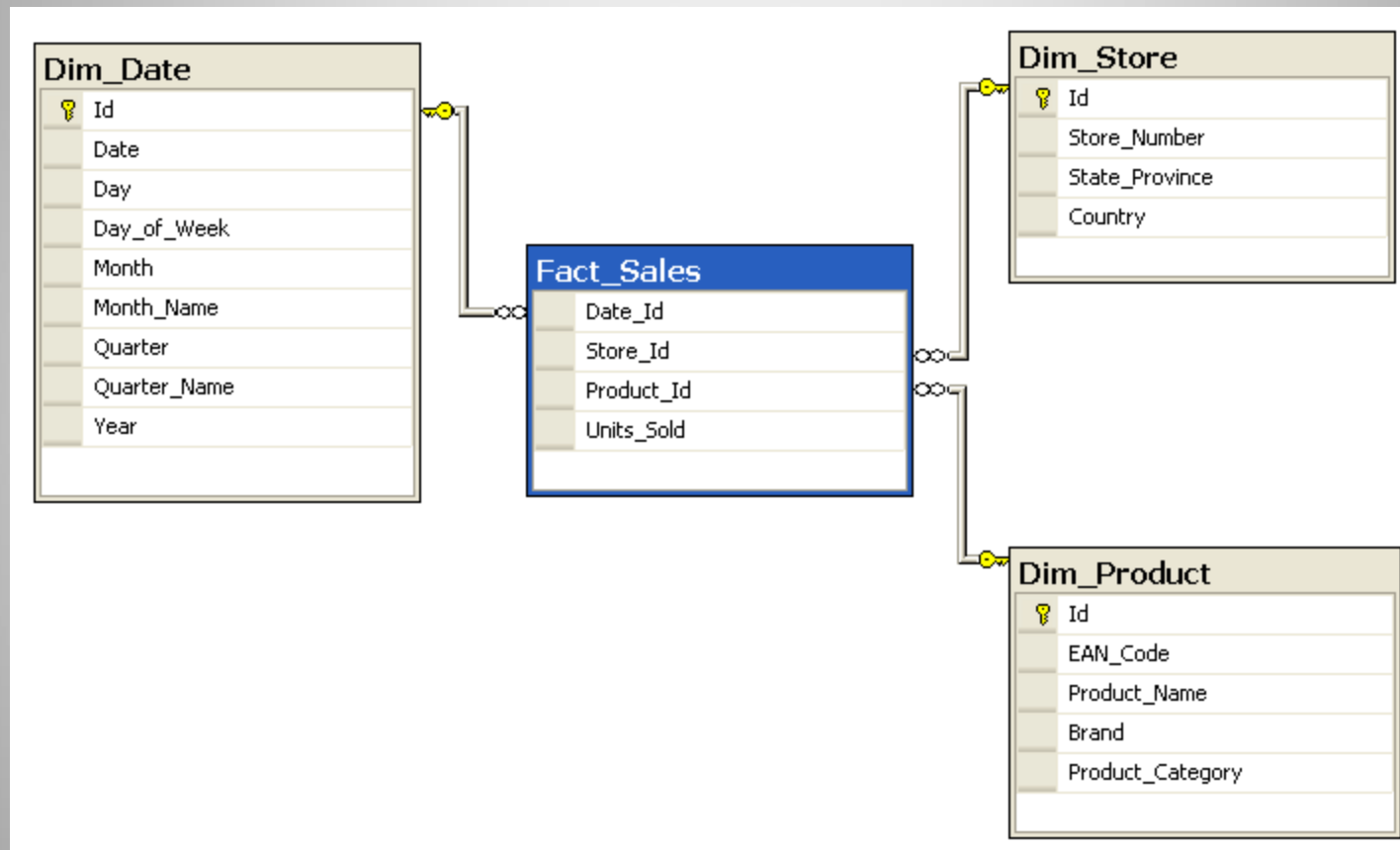
Star scheme

A complete star scheme

- Fact
- Dimensions
- Hierarchies



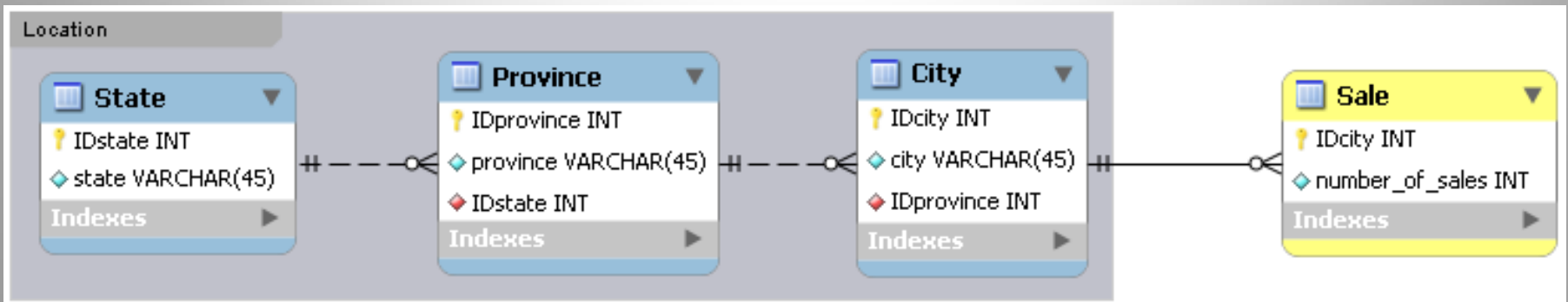
Star scheme example



Snowflake scheme illustration

In a snowflake schema, each hierarchy level is stored in a specific table

- ❑ Now, the dimensions are in 3NF!
- ❑ Hierarchies are fully visible
- ❑ Lowest level of hierarchy (least aggregated) is connected with fact table

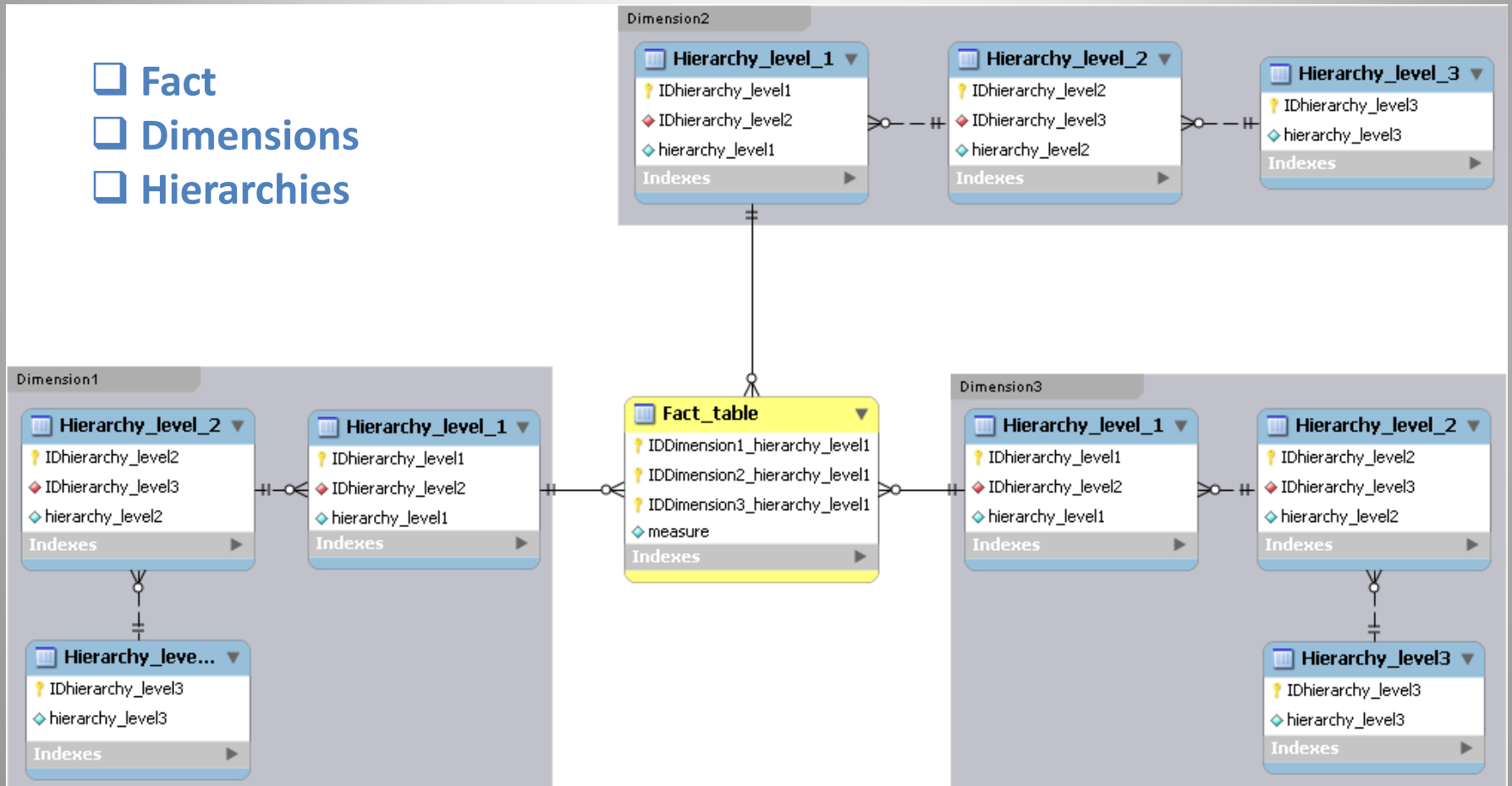


Practice

Snowflake scheme

A complete snowflake scheme

- Fact
- Dimensions
- Hierarchies

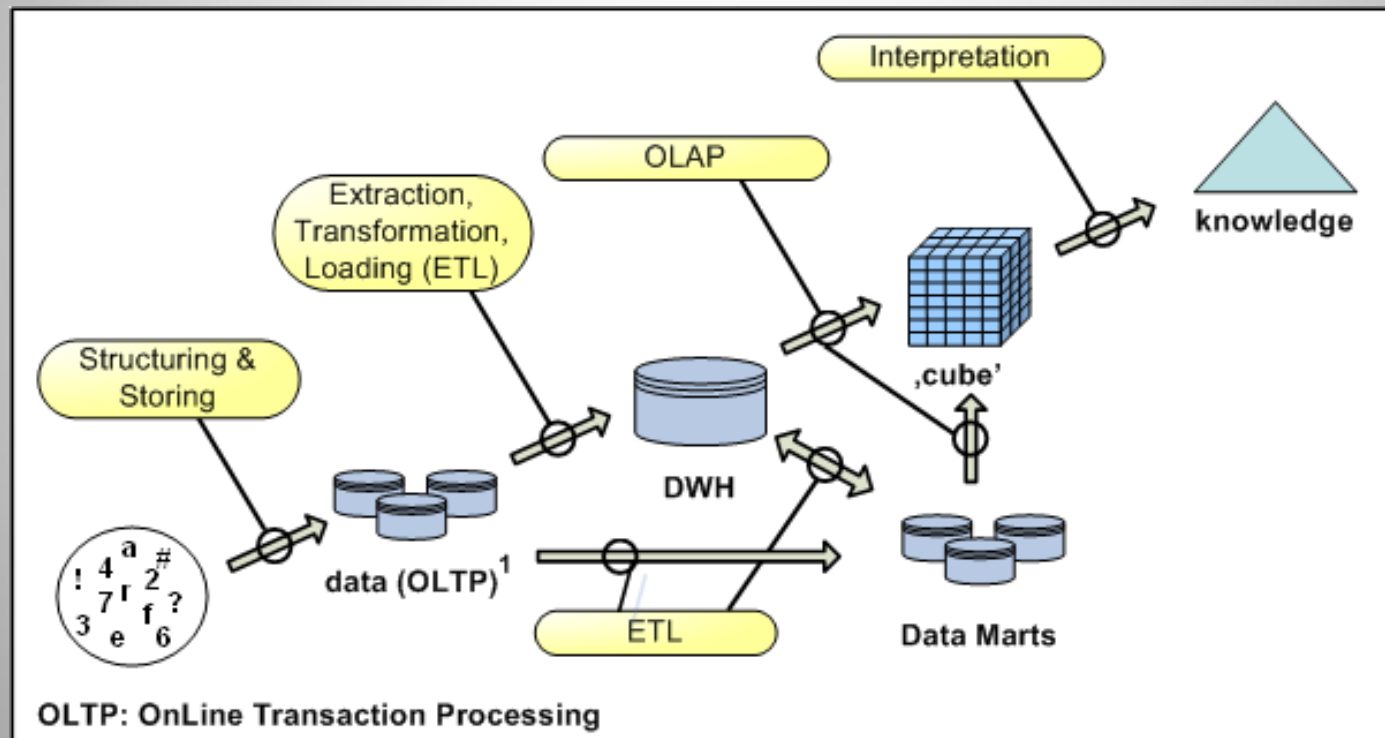


Star scheme or snowflake

	Star scheme	Snowflake scheme
3 rd Normal Form	No	Yes
Amount of data	Redundant information stored	Better structured
Relationships	Simple	Can be quite complex
Queries	Normally very fast, since no JOINS are needed	Can be quite slow due to the necessity of many JOINS

5. Synthèse avant l'étape de la création d'une base de données multidimensionnelle/ HyperCube

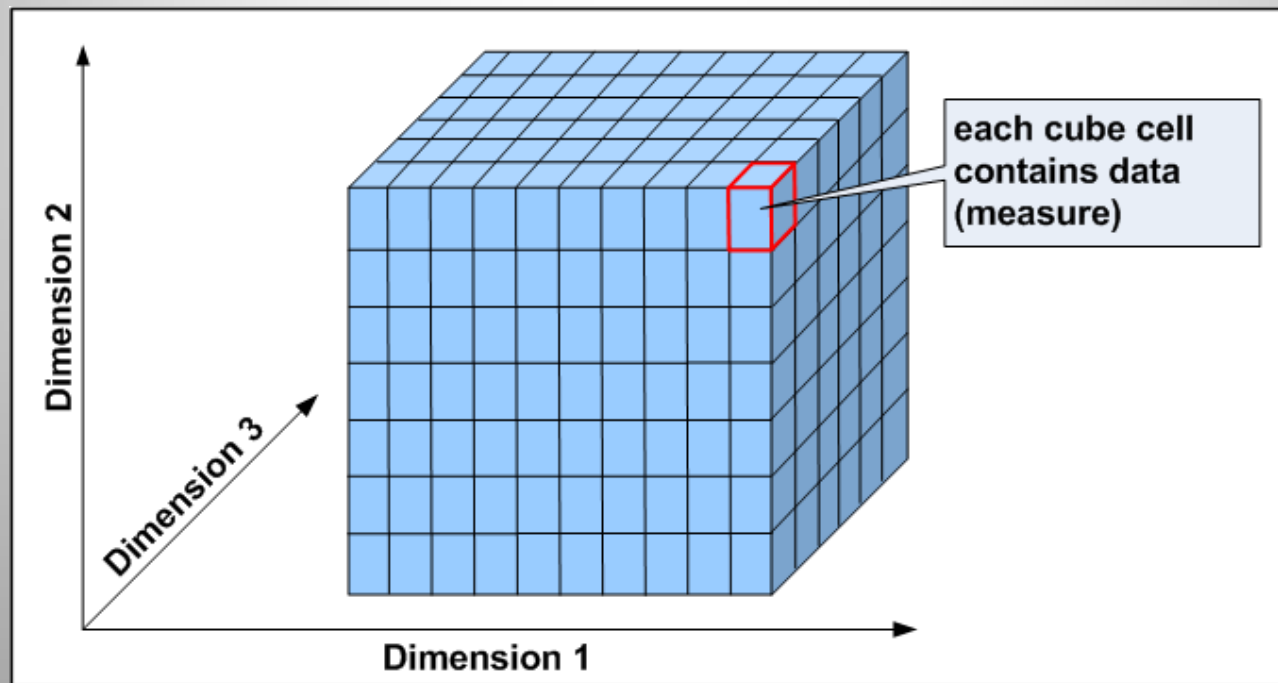
5.1. La prochaine étape, le cube



5.2. L'objectif du processus OLAP à partir d'un Datamart

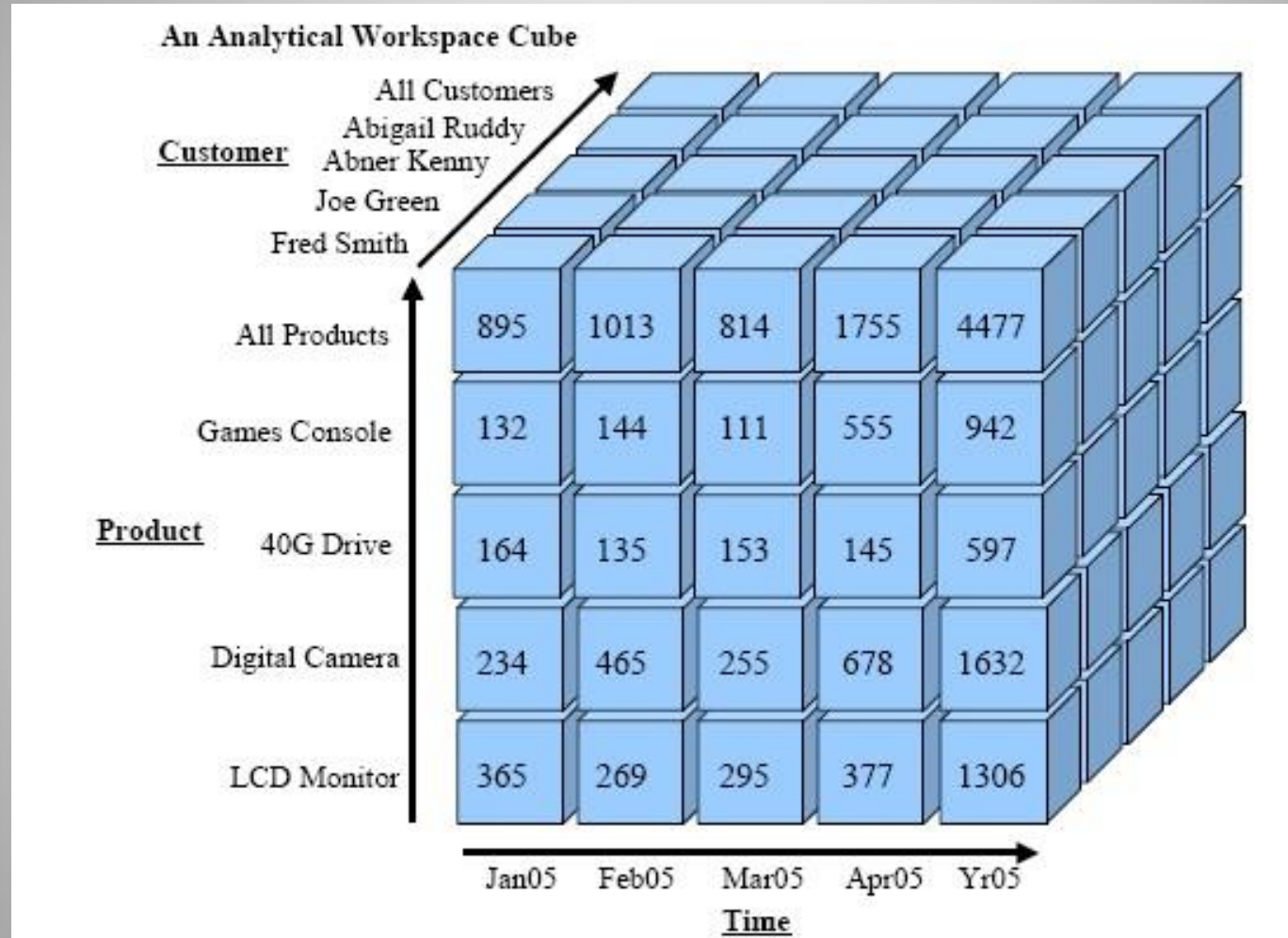
- Créer une base de données multidimensionnelle, elle peut s'illustrer par un cube, si on se limite à 3 dimensions.
- Les axes représentent les dimensions
- Les cases du cube permettront de stocker les valeurs des mesures des faits selon les coordonnées des dimensions

Base de données multidimensionnelle / HyperCube: structure



Il manque peut-être une indication sur le fait que permet de représenter ce cube (ventes?)

Illustration OLAP Cube



5.3. Que pouvons-nous retenir avant d'aborder la construction de “cubes”?

- ☐ BI?
- ☐ OLAP?
- ☐ ETL?
- ☐ OLTP?
- ☐ DM/DWH?
- ☐ Comment concevoir un DM/DWH?
- ☐ Dimensions, Hierarchies and Facts ?
- ☐ HyperCube?