

Entrepôt de données autour du PMSI

Lama EL SARRAJ

Ingénieur Hospitalier (AP-HM)

2° années de doctorat en informatique (LSIS)



AP-HM

Sommaire

2

- Introduction
- Programme de médicalisation des systèmes d'information
- Aide à la décision
- Entrepôt de données de santé
- Conclusion

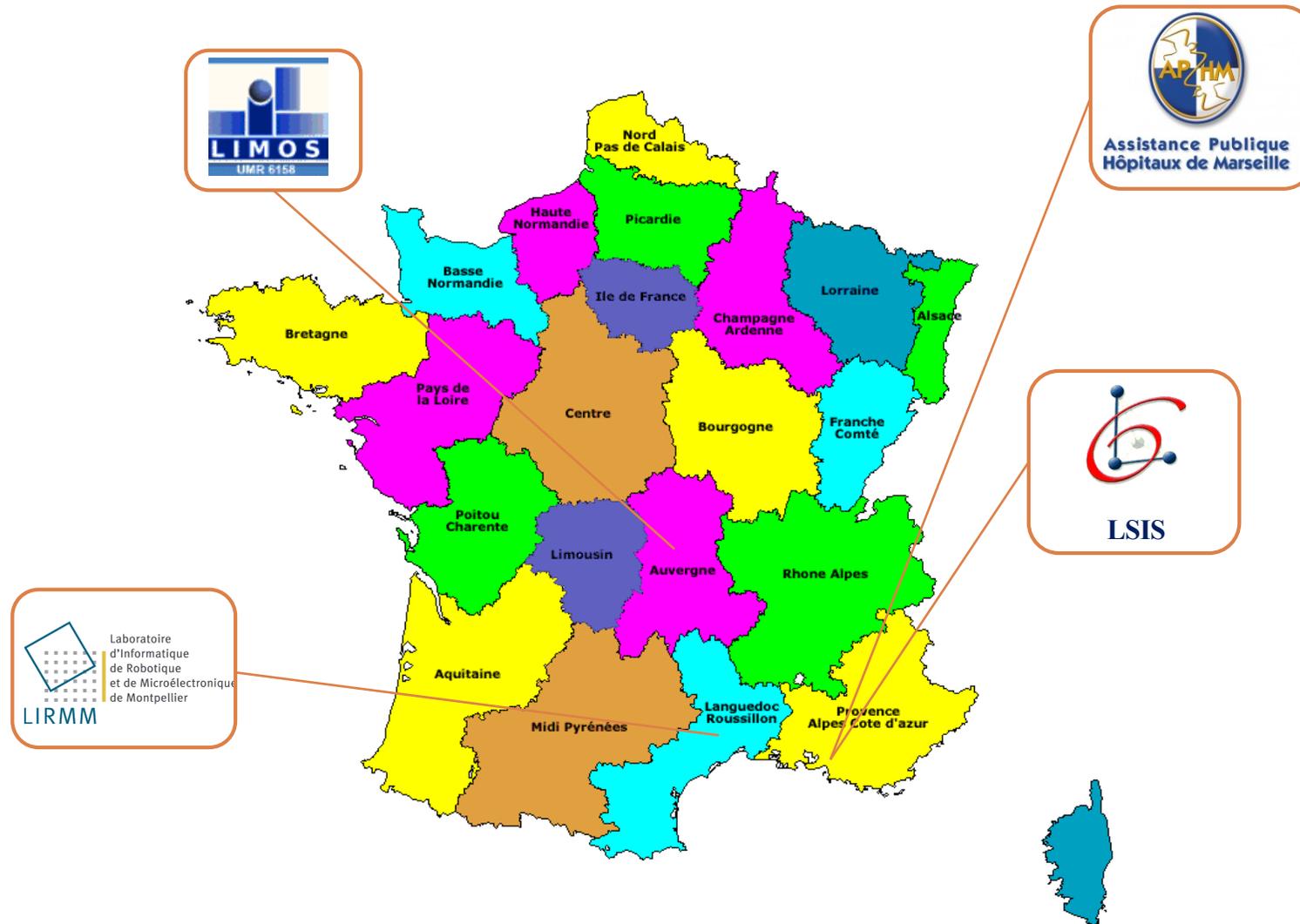
Objectif

3

- Mise en œuvre d'une méthodologie d'implémentation et de mise en œuvre d'un entrepôt de données pour l'aide au pilotage.

Contexte

4



AP-HM

5

- Assistance Public des Hôpitaux de Marseille est un établissement public de santé, regroupant regroupant 5 hôpitaux (Hôpital Nord, la Timone, la Conception, Saint Margueritte et Salvator)
- Budget annuel : ~ 1 milliard d'euros
- Moyens humains: ~ 17 000 agents
- Nombre de lits: ~3 500
- Activité annuelle:
 - ▣ ~ 900 000 consultations externes
 - ▣ ~ 190 00 personnes reçues aux urgences
soit 510 patients reçus en moyenne chaque jour
 - ▣ ~ 66 00 interventions chirurgicales
 - ▣ ~ 5 000 naissances



- Direction des systèmes d'information et de l'organisation
 - ▣ Directeur : Olivier PONTIES

- Service organisation méthodes et recherche appliqué
 - ▣ Chef de service : Sophie RODIER

- Service de l'informatique décisionnel :
 - ▣ Chef de service : Jean-François MIRETTI

Université

7

- Université Paul, Cézanne 2ème année de thèse
 - Ecole doctorale Mathématiques et Informatique de Marseille
n° 184

- LSIS (OASIS) :
 - Directeur de thèse : Bernard ESPINASSE

- LIRMM :
 - Co-directeur de thèse : Thérèse LIBOUREL



8

PMSI

PMSI : Définition

- Programme de *Médicalisation des Systèmes d'Information* (PMSI).
- C'est l'ensemble de systèmes d'information qui tiennent compte des pathologies et des modes de prise en charge.

PMSI : Historique

10

- 1982 :
 - ▣ Début de l'histoire du PMSI en France (initiative du ministère de la santé)
 - ▣ Objectif : recueillir des informations à partir du dossier médical des patients (diagnostics, actes médicaux, date de naissance, date d'entrée, etc.) à des finalités de santé publique et épidémiologique.
- 2005 :
 - ▣ Tarification sur l'activité (T2A)
 - ▣ Objectif : économiques, financement des hôpitaux à la base de l'activité.

PMSI : Codage

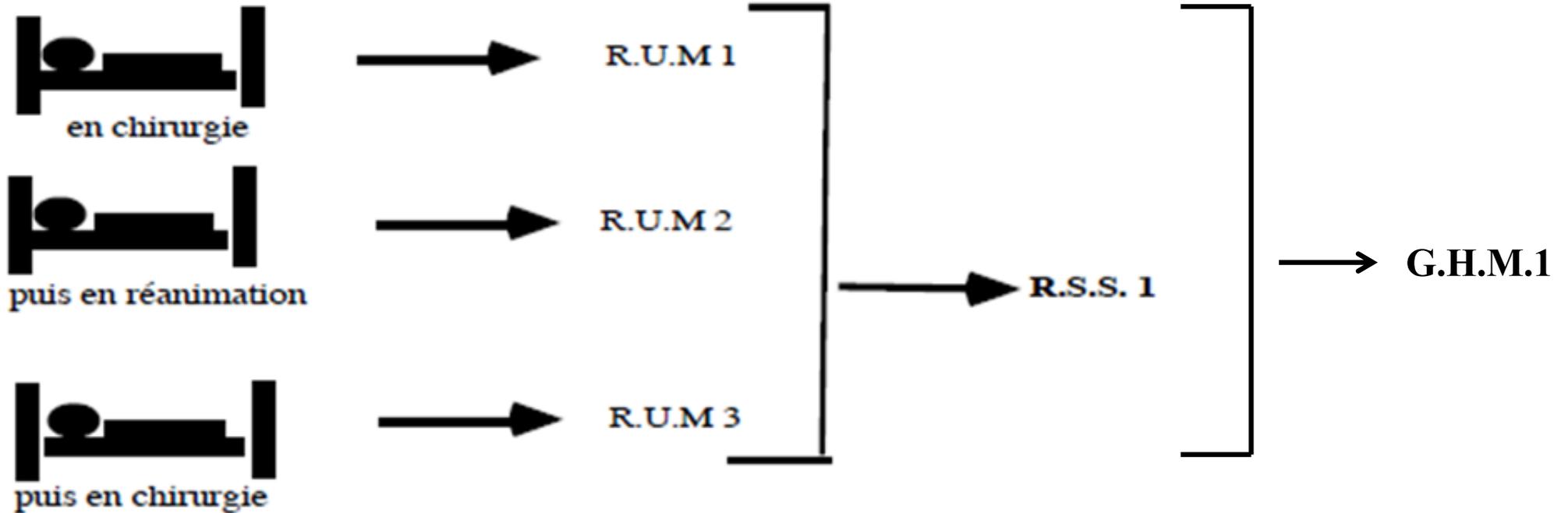
11

- Le codage des diagnostics et des actes médicaux(CIM10, CCAM) est important dans le cadre du PMSI parce que :
 - ▣ un même séjour peut être codé de plusieurs façons différentes
 - ▣ chaque code conduit à un tarif spécifique

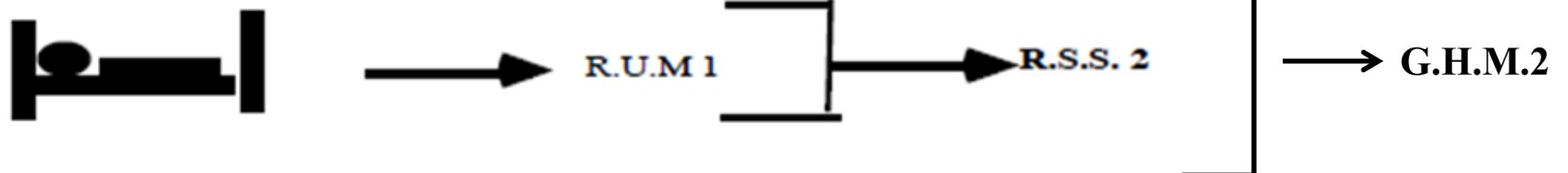
PMSI : Modalité de production

12

Mr Dupont est hospitalisé du 05 janvier au 25 février



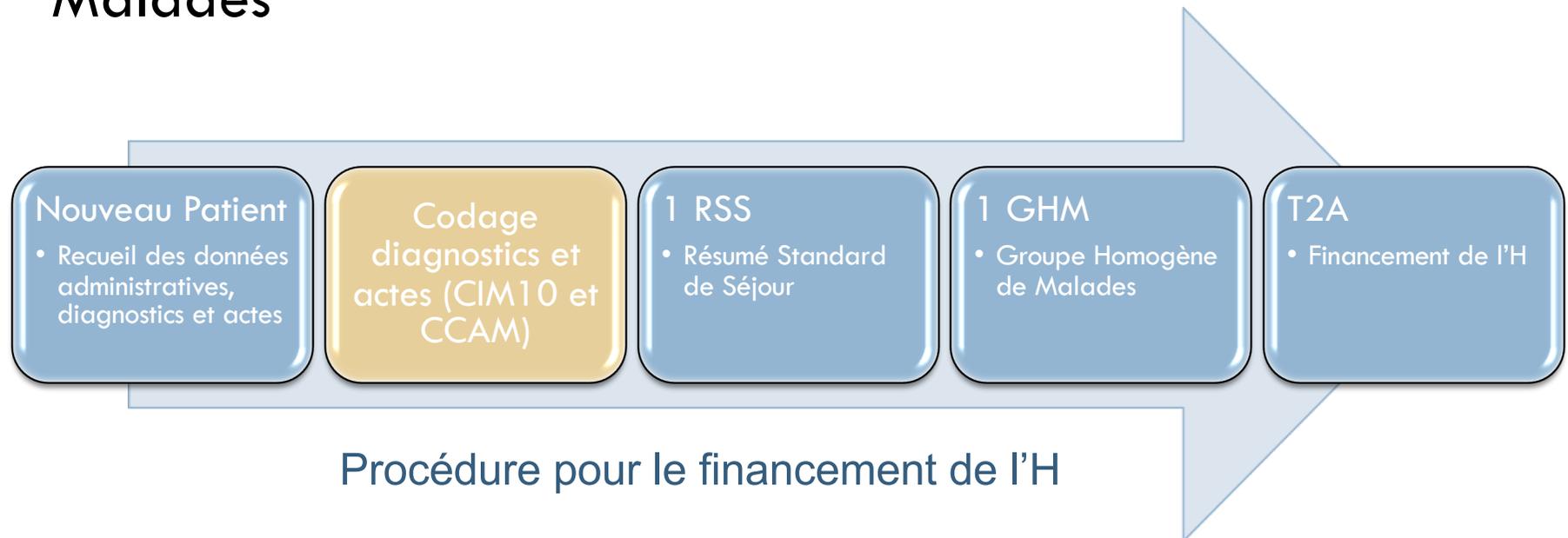
Mr Dupont est hospitalisé du 03 au 10 mai en cardiologie



PMSI : GHM

13

- Groupe Homogène de Malades = G.H.M.
 - ▣ Informations contenues dans les R.U.M. et agrégées en R.S.S. permettent de classer les séjours en Groupes Homogènes de Malades



PMSI : SSPIM

14

- 1 - assistance aux services de soins de court séjour pour les aider à produire et à transmettre les R.U.M.
- 2 - traitement des différents R.U.M. relatifs à un même séjour hospitalier avant groupage (chronologie des séjours multi-unités,...).
- 3 - production des R.S.S. et groupage des R.S.S. en G.H.S.
- 4 - production de statistiques médicales et analyse des résultats à l'usage des services producteurs M.C.O., de la C.M.E. et de la Direction.
- 5 - production des R.S.A. (les RSS rendus anonymes) et transmission de ces R.S.A. à l'Agence Régionale d'Hospitalisation (ARH).
- 6 - contrôle interne de la qualité des R.S.A. et transmission de ce contrôle de qualité à l'ARH.
- 7 - Préparation du contrôle de qualité externe réalisé par l'ARH.

15

Prise de décision

Prise de décision



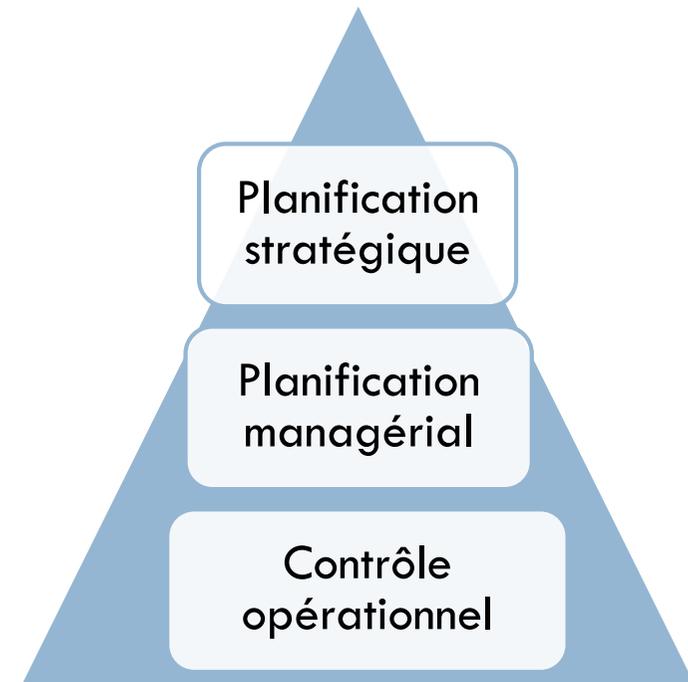
16

- Au quotidiens nos décisions sont souvent prises par intuitions et d'expériences passées.
- Lorsque nous sommes confrontés à des situations nouvelles, la tâche de prise de décision devient beaucoup plus difficile.
- L'environnement des décideurs est de plus en plus complexe et évolue rapidement dû à plusieurs facteurs :
 - ▣ nouvelles technologies de traitement de l'information
 - ▣ complexité structurelle des décisions, plus forte compétition, ...

Prise de décision

17

- Les décisions sont liées à l'activité qui s'y déroulent
- Anthony (1966) distingue 3 niveaux d'activités managériales :



Prise de décision

18

- Régulation (contrôle opérationnel) :
 - ▣ activités concernant principalement des rythmes inférieurs au mois et conduit à des décisions dont les conséquences sont à court ou très court terme
 - ▣ décisions de portée limitée (champ local et limité)
- Pilotage (Planification et le contrôle managériaux) :
 - ▣ activités conduisant à des décisions dont les conséquences sont à moyen terme
 - ▣ décisions de portée intermédiaire entre celles de Régulation et de Planification Stratégique
- Planification Stratégique :
 - ▣ activités engendrant des décisions majeures dont les conséquences sont à long terme.
 - ▣ décisions de portée globale (champ global)

L'informatique décisionnel

19

- L'informatique décisionnelle (Business Intelligence ou BI)
c'est les moyens et les méthodes qui permettent de :
 - ▣ Collecter les données
 - ▣ consolider les données
 - ▣ modéliser les données
 - ▣ restituer les données
- en vue d'offrir une aide à la décision et de donner une vue d'ensemble de l'activité traitée.

20

Entrepôt de données

ED

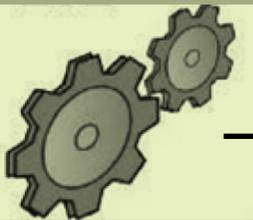
Architecture d'un entrepôt de données (ED)

21

Données opérationnelles



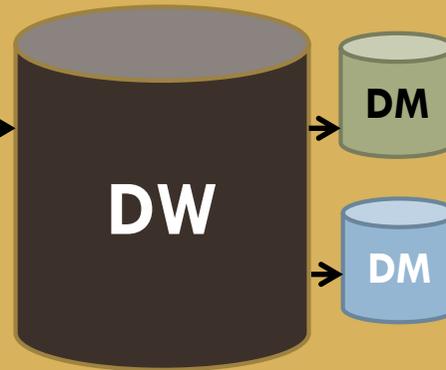
ETL



Structuration

- Homogénéisation
- Nettoyage
- Extraction
- Transformation
- Chargement

ED



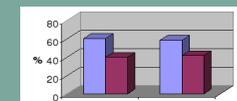
Stockage

Orientation analyses
Données historisées
Non volatile

Data mart:

Filtrer, Résumé, Distribuer

Analyse



Utilisation des données

Rapport statistiques
Data mining
Cube de données

Proportion de la charge du projet

22

➡ Alimenter	50%
➡ Fabriquer	15%
➡ Administrer	15%
➡ Restituer	20%



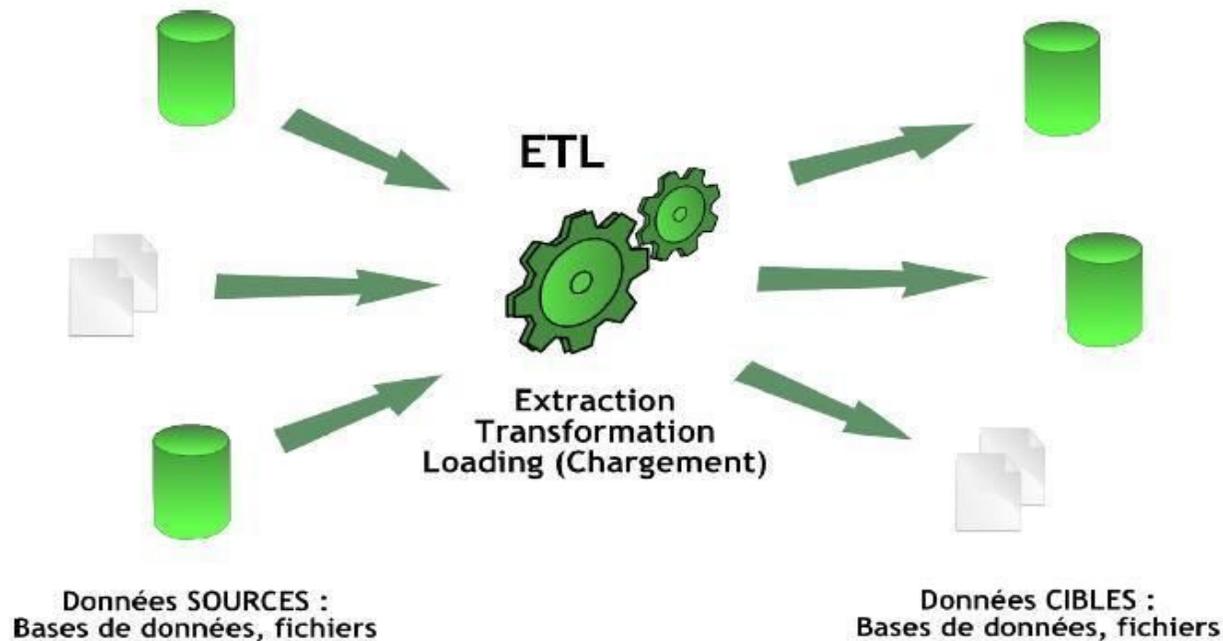
23

Extracting Transforming Loading

ETL

24

- Il s'agit d'une technologie informatique intergicielle permettant d'effectuer une synchronisation massive d'information d'une banque de données vers une autre



ETL, Extraction

25

- L'extraction doit se faire conformément aux règles précises du référentiel.
 - ▣ Ne pas perturber les activités de production
 - ▣ Conserver les informations sur les données (métadonnées)
 - ▣ Gérer les anomalies en traitant et gardant trace
 - ▣ Les données extraites doivent être marquées par « horodatage »

ETL, Transformation

26

- C'est une suite d'opérations qui a pour but de rendre les données cibles homogènes et puissent être traitées de façon cohérente.
- Donnés sources données cibles
 - ▣ Appli 1 : male, Female m, f
 - ▣ Appli 2 : 1, 0 m, f
 - ▣ Appli 3 : Masculin, féminin m, f

ETL, Chargement

27

- C'est l'opération qui consiste à charger les données nettoyées et préparées dans le DW.
- C'est une opération qui risque d'être assez longue. Il faut mettre en place des stratégies pour assurer de bonnes conditions à sa réalisation et définir la politique de rafraîchissement.
- C'est une phase plutôt mécanique et la moins complexe.

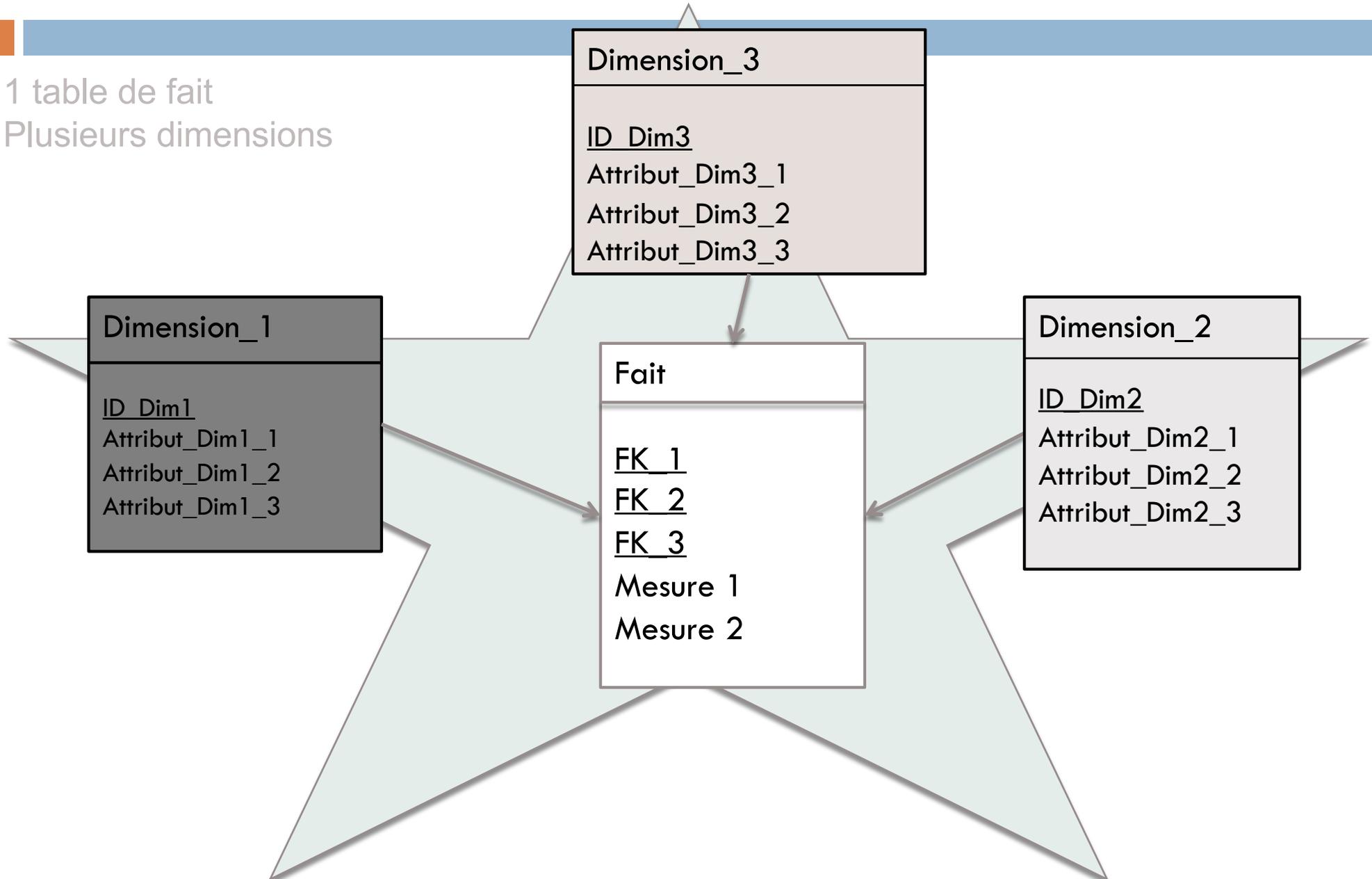
28

Conception

Conception, Modèle en étoile

29

- 1 table de fait
- Plusieurs dimensions



Conception, Modèle en étoile

30

- Une table de faits:
 - ▣ identifiants des tables de dimension;
 - ▣ une ou plusieurs mesures.
- Plusieurs tables de dimension :
 - ▣ descripteurs des dimensions.
- Une granularité définie par les identifiants dans la table des faits.

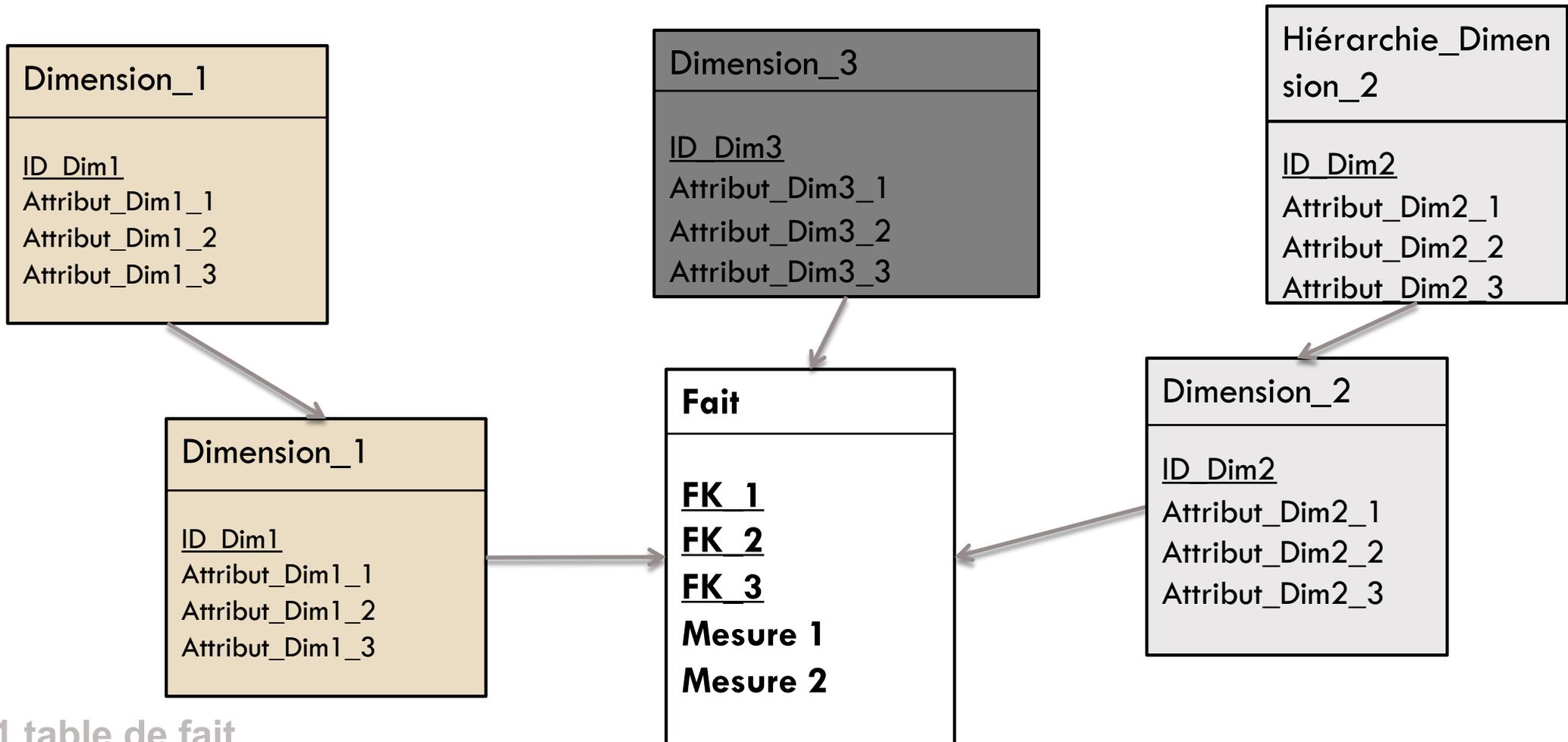
Avantage

- **Nombre de jointure limité**
- **Facilité de navigation**
- **Gestion des agrégats**
- **Fiabilité des résultats**

Inconvénients

- **Alimentation complexe**
- **Toutes les dimensions ne concerne pas les mesures**
- **Redondances dans les dimensions**

Conception, Modèle en flocon de neige



- 1 table de fait
- Plusieurs dimensions
- Des hiérarchies

Conception, Hiérarchies

32

- Le modèle doit être simple à comprendre.
- Les hiérarchies nous aident à augmenter sa lisibilité en regroupant certaines dimensions.
- Les hiérarchies peuvent être géographiques et organisationnelles
- Exemple: Commune, Département, Région, Pays, Continent

Conception, Modèle en flocon de neige

33

- Lorsque les tables sont trop volumineuse

Avantage

- Réduction du volume
- Permet des analyses par pallier (Drill down) sur la dimension hiérarchisée

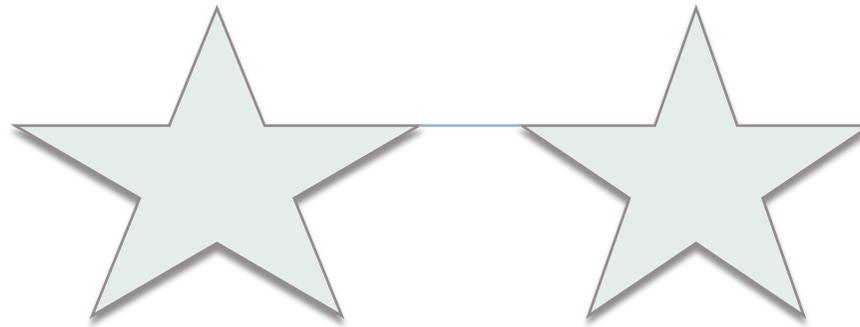
Inconvénients

- Navigation difficile
- Nombreuses jointures

Conception, Modèle d'un ED, en constellation

34

- La modélisation en constellation consiste à la fusion de plusieurs modèles en étoile qui utilisent des dimensions communes.
- 2 tables de faits
- 1 dimension en commun (au minimum)



35

OnLine Analysis Processus

OLAP

OLAP

36

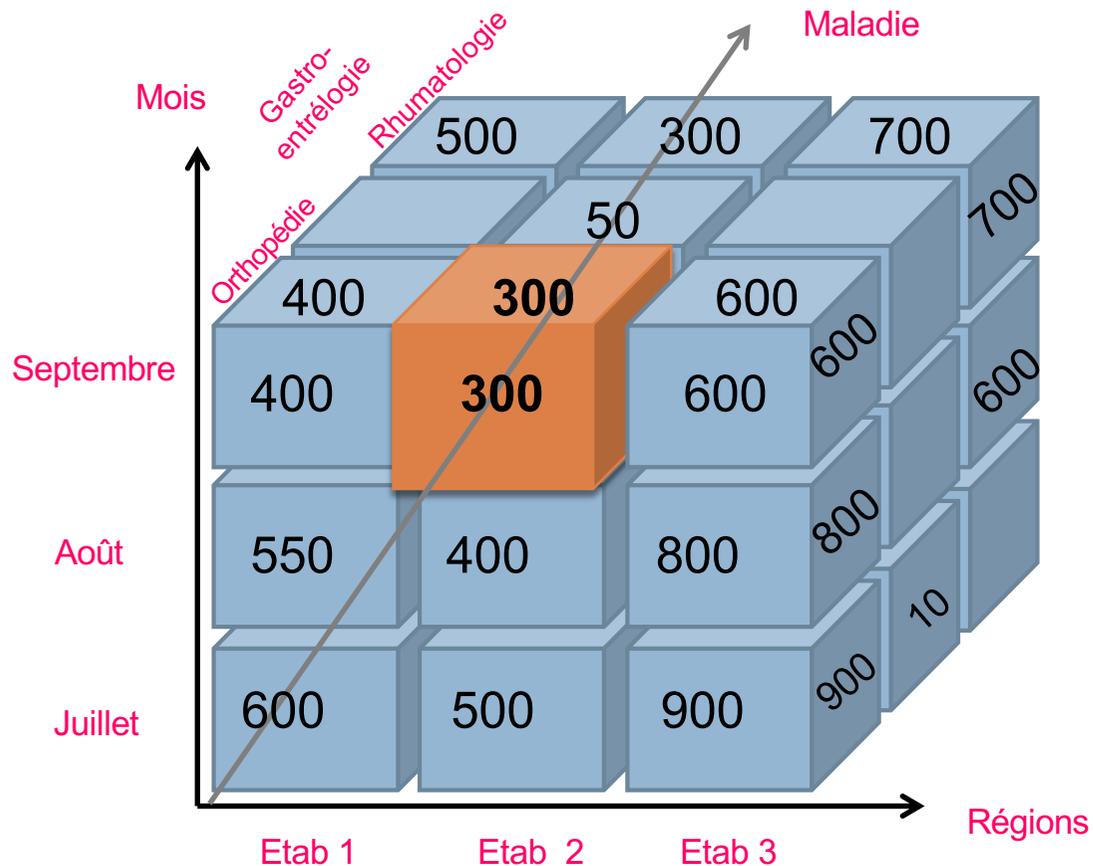
- les requêtes décisionnelles sont de type OLAP exemple :
“Quelle est le nombre de patient par maladie, par établissement par mois ?” On peut calculer de divers totaux.

	système de gestion opérationnel (OLTP)	Systemes décisionnels (OLAP)
Objectif	Dédier au métier et à la production	Dédier à la gestion de l'entreprise (pilotage et prise de décision)
Volatilité	Données volatiles	Données historisées Ex : garder la trace des évolutions des prix, introduction d'une information daté
Historisation	Pas d'historiques	Les données n'ont pas vocation à être supprimer
Granularité des données	Totale, on accèdes directement au données atomiques	Agrégés, niveau de synthèse selon les besoins
Intégrées	Les données proviennent d'une même source de données	Les données proviennent de différentes sources de données (formats, types, etc.)
Performance	<i>on peut traiter quelques centaines de tuples par seconde.</i>	<i>on peut rajouter en lignes et en colonnes plus de 10000 valeurs par seconde.</i>

Cube : exemple

38

Nombre de patients par maladie, dans l'établissement 2 en septembre 2011



cube : nombre de patients

• fait : nombre de patients, maladie orthopédie, établissement 2, septembre 2011, **300**

• mesure : **300**

Différentes opérations OLAP

39

Roll up : Agrégation de données

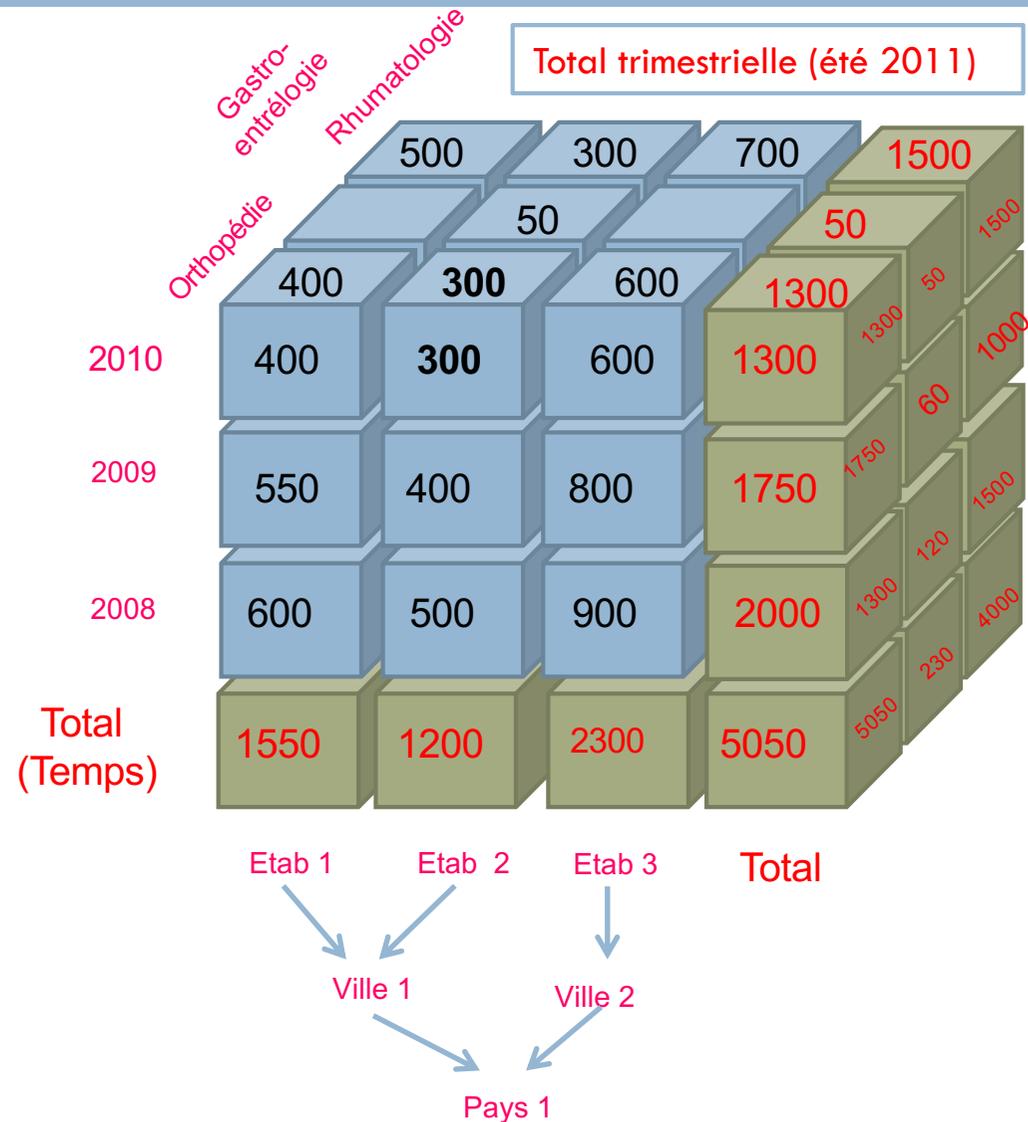
Fonction d'agrégat : Somme

Drill down : forage des données

exemple : du niveau des villes aux régions

Slice and dice : sélection et projection des données, par exemple ne prendre que Etab 1 et 2 pour le mois d'Août et septembre

Rotate, switch, split, etc.



40

Entrepôt de données autour du PMSI

EDS : Besoin de pilotage :

41

- Suivit de l'activité
- Utilisateurs divers
- Besoins évolutifs
- Niveau de détails différents d'un utilisateur à l'autre sur les même données
- Etc.

EDS : définition

42

- Entrepôt de données (ED):

« L'entrepôt de données, ou 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(1990)

- Entrepôt de données de santé =

ED + Appliqué dans le domaine de la santé

EDS: PMSI

43

- Outil quantitatif destiné en priorité à la tarification à l'activité
- Base de dialogue intéressante entre les tutelles et les établissements et entre médecins et administratifs
- Données standards et structurées :
 - ▣ Données du RUM
 - ▣ Données du RSS
 - ▣ Données du GHM
 - ▣ Etc.

EDS : Objectif

44

- Effectuer des analyses multidimensionnelles (OLAP) pour l'aide à la décision et le pilotage
- Création des connaissances
- Organiser les données sur les différents niveaux de granularité et de détails des données
- Indicateurs de pilotage
- Production de tableau de bord
- Etc.

EDS : Problématique de conception

45

- Manque de qualité des données du PMSI, notamment au niveau du codage (données mal codées ou non codées)
- Manque de moyens d'évaluation de ces données
- Évolution des données :
 - ▣ médicales (nouvelles découvertes en permanence, un jargon très vaste, etc.)
 - ▣ référentiel PMSI (codes CIM10, CCAM et groupes d'appartenance GHM)
- Évolution et diversité des besoins

PMSI : Les utilisations

46

- utilisation « diverses » :
 - ▣ comme outil qualitatif : enquête science et Avenir
 - ▣ comme outil de planification
 - ▣ comme outil épidémiologique
 - ▣ ...

- utilisations internes :
 - ▣ analyse des postes de dépenses d'un établissement
 - ▣ analyses des coûts par service
 - ▣ analyses des points ISA par services

- Utilisation externe :
 - ▣ Aide à l'allocation budgétaire

47

Conclusion

Conclusion

48

La nouvelle modalité de financement des hôpitaux

Des données communs à la disposition de tout les hôpitaux (ex : PMSI)

Nécessité de nouvelle moyen d'évaluation de l'activité

Les nouvelles technologies mettent à notre disposition des outils d'aide à la décision

La problématique d'évolutivité des données, des besoins, etc. et alors l'évolution du modèle conceptuel de l'ED

La nécessiter d'une nouvelle méthodologie de conception et d'implémentation d'un ED qui soit adapter aux hôpitaux