

Maintenance des Entrepôts de Données Hétérogènes

Séminaire LCR - LIPN



Mohamed BADRI

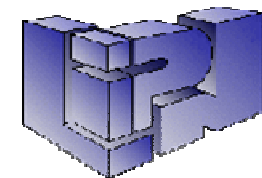


Directeur de Thèse : M. C.F. DUCATEAU (Crip5 – Paris5)

Co-encadré par : M. F. BOUFARES (Lipn – Paris13)



mohamed.badri@univ-paris5.fr



27 Mars 2006

Maintenance des entrepôts de données hétérogènes

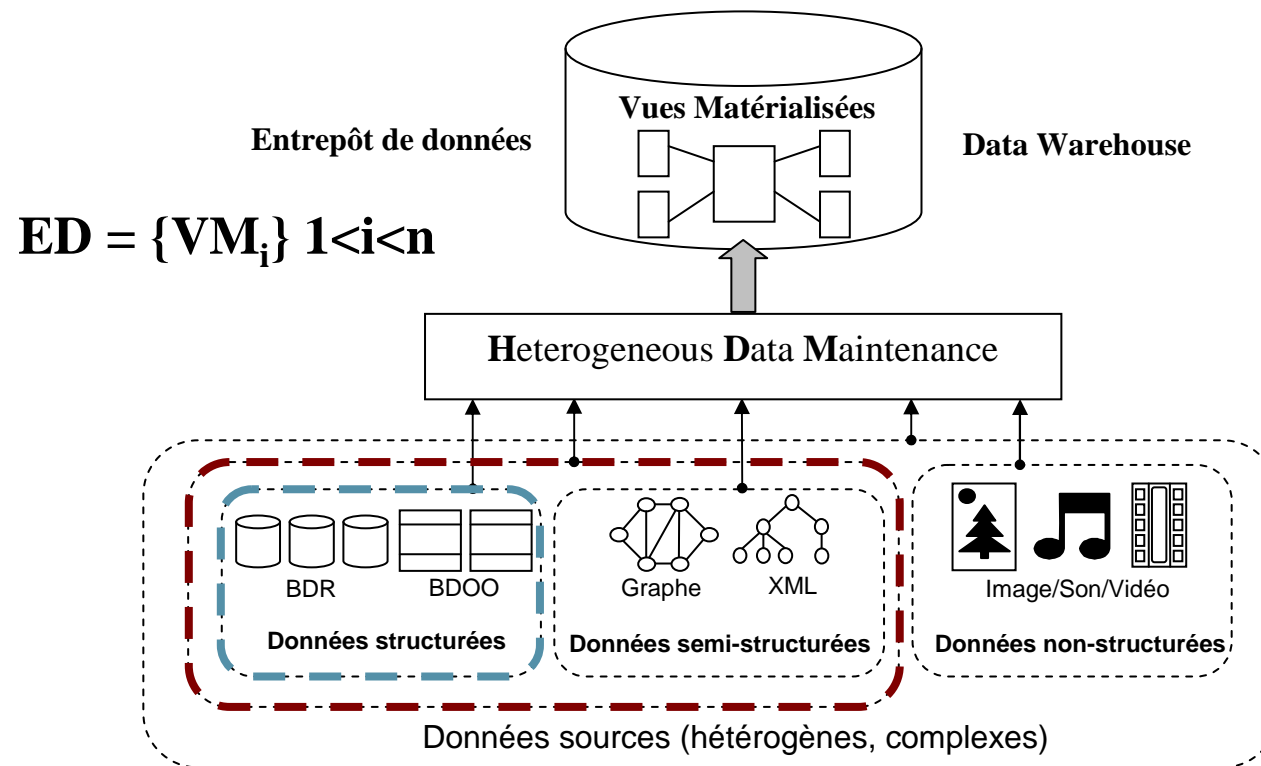


Figure 1 : *Système de maintenance des entrepôts de données hétérogènes*



Maintenance des entrepôts de données

■ Maintenance de données :

- Répercuter les mises à jour des données sources
- En temps réel / Différée / Gestion transactionnelle
- Travaux [Zhuge *et al.*, 1995, Zhuge *et al.*, 1996, O’Gorman *et al.*, 1999, Laurent *et al.*, 2001]

■ Maintenance de structures :

- Répercuter les changements de structures des sources
- Adaptation et réécriture des VM
- Travaux [Nica *et al.*, 1999, Zhang et Rundensteiner, 2000, Velegrakis *et al.*, 2003]

Exemple de maintenance de données

- Le complément canonique d'un ED [Laurent et al. 01]
 - Idée : mettre à jour l'ED sans accéder aux DS (**ED indépendant**)
 - Stocker, en plus des vues matérialisées, des vues auxiliaires
- **Complément d'un ED** : les données que l'entrepôt n'entrepose pas

Sales

Nom	Obj
anne	DVD
anne	TV
paul	PC
paul	DVD
paul	CD
marie	DVD

Emp

Nom	Age
anne	25
paul	45
marie	30

$$C_{Emp} = Emp - \Pi_{(nom, age)} Sold$$

Nom	Age
marie	30

Sold = Emp \bowtie sales

Nom	Obj	Age
anne	DVD	25
anne	TV	25
paul	PC	45
paul	DVD	45
paul	CD	45
marie	DVD	30

Insert (Sales,[marie,DVD])

Insert (Sold,[marie,DVD] \bowtie C_{Emp})

Source

Entrepôt

27-03-2006

M. BADRI

3



Exemple de maintenance de structures

- EVE : Evolvable View Environment
[Rundensteiner et al., 2000]
 - Redéfinition et re-calcul des vues matérialisées
 - Définition des VM à l'aide de E-SQL (Evolvable SQL)
 - définir des préférences d'évolution des vues.
 - paramètres d'évolution : dispensable, remplaçable



Maintenance des ED Homogènes

- Les ED sont issus de sources relationnelles
 - Systèmes : ECA, Strobe, POOSE, Complément Canonique
 - Les travaux ont essentiellement porté sur la maintenance des données [Zhuge *et al.*, 1995, Zhuge *et al.*, 1996, O’Gorman *et al.*, 1999, Laurent *et al.*, 2001]
 - Peu de travaux ont traité la maintenance structurelle [Gupta *et al.*, 1995]
 - Adaptation des VM : Méta données



Maintenance des ED Hétérogènes (1/3)

■ Maintenance des ED Hétérogènes

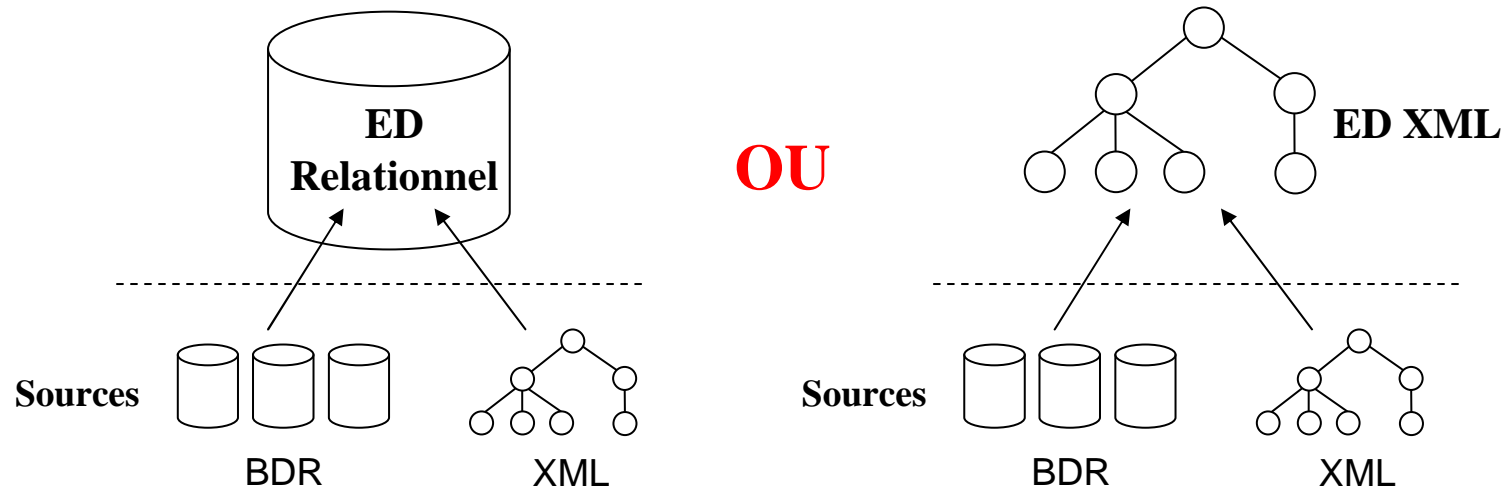
- Hétérogénéité : définition assez ambiguë dans la littérature
- Rappelons la définition de l'hétérogénéité

Données **Hétérogènes**:

1. Elles appartiennent à la **même catégorie** de données mais elles ont des **modélisations différentes**.
2. Elles appartiennent à des **catégories différentes** de données.

Maintenance des ED Hétérogènes (2/3)

- Dans les travaux actuels [Rundensteiner *et al.*, 2000, Engström *et al.*, 2003, Chen *et al.*, 2004], le type et la structure des VM sont fixés à l'avance (Rel./XML)



- Absence de critères justifiant le choix
- Les algorithmes de maintenance dépendent de ce choix



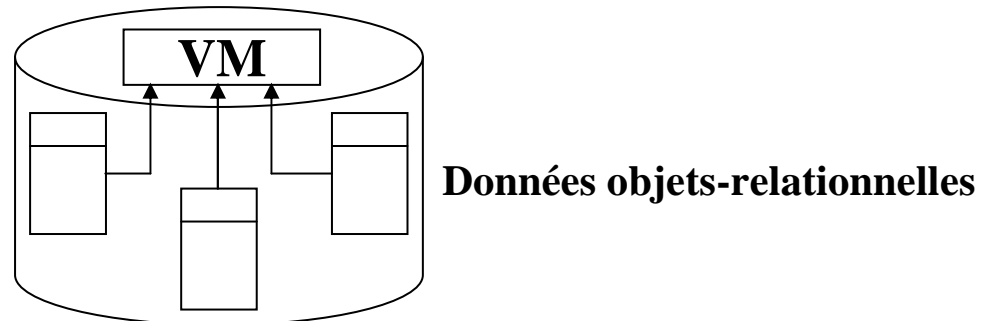
Maintenance des ED Hétérogènes (3/3)

- Maintenance d'ED issus de sources relationnelles / XML :
[Rundensteiner *et al.*, 2000, Engström *et al.*, 2003, Chen *et al.*, 2004, Laurent *et al.*, 2005]
- Utilisation d'une couche de mapping (Wrapper) pour la transformation des données (Relationnel ↔ XML)
- Dans nos travaux le type et la structure des VM ne sont pas fixés à l'avance.
 - Ils sont fonction du type des données sources.

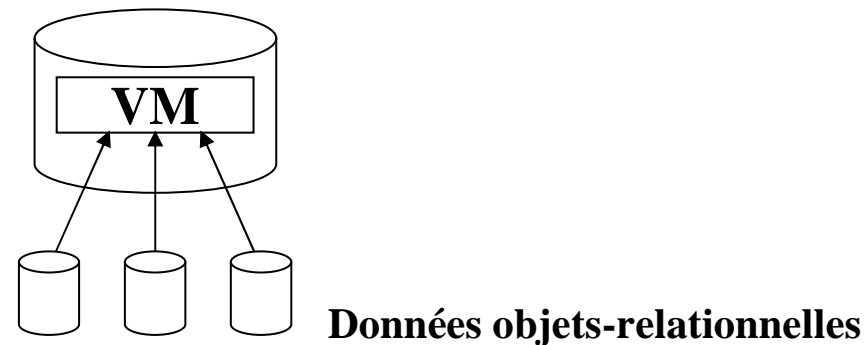
Maintenance de VM issues de données hétérogènes sous Oracle

- Maintenance des vues matérialisées (VM) hétérogènes (complexes)
 - Données relationnelles
 - Données objet-relationnelles

- Cas Mono-Base :



- Cas Multi-Bases



Création de vues matérialisées sous Oracle

Sources		Destination	Type et périodicité de rafraîchissement			
Nombre de bases	Nombre de tables	Composition des vues	Fast On Commit	Fast On Demand	Complete On Commit	Complete On Demand
Mono-Base	Mono-table	VM (att. simples)	Oui	Oui	Oui	Oui
		VM (att. complexes)	—	Oui	—	Oui
	Multi-tables	VM (att. simples)	—	—	Oui	Oui
		VM (att. complexes)	—		—	Oui
Multi-bases	Multi-tables	VM (att. simples)	—	—	—	Oui
		VM (att. complexes)	—		—	—

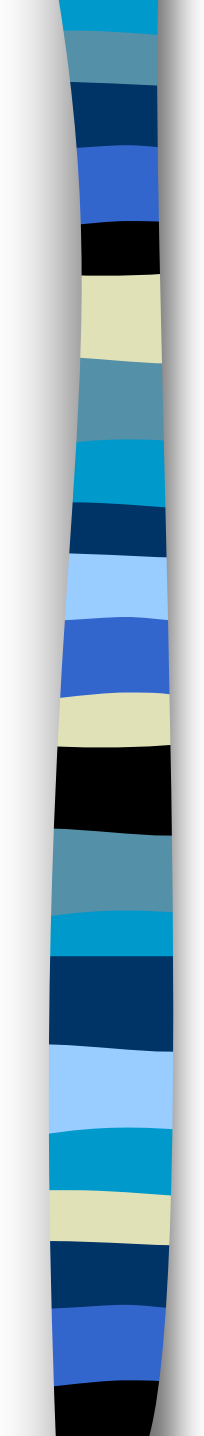
```
CREATE MATERIALIZED VIEW MV_MULTI refresh complete on demand as
SELECT M.MENUM, M.MENOM, M.MEADR, M.MECV, P.PANUM, P.PANOM, P.PAADR, A.ACDATE,
A.ACDESC
FROM MEDECINS@DB1 M, PATIENTS@DB2 P, ACTES@DB2 A
WHERE M.MENUM=A.PRNUM
AND A.PANUM=P.PANUM;
```

ORA-32303: les vues matérialisées avec des types définis par l'utilisateur ne peuvent pas référencer plusieurs sites maîtres



Maintenance de VM issues de données hétérogènes sous Oracle

- Maintenance des VM sous Oracle
 - Recalcule entièrement et à la demande une VM issue de données hétérogènes mono-base multi-tables.
 - Ne permet pas la création de VM multi-bases issues de données hétérogènes.
- Proposition de procédures de maintenance des VM
 - Calculer partiellement la vue matérialisée



Calcul des vues matérialisées complexes (1/2)

■ Données sources (DS)

- \mathbf{R}_k ($k \in [1..n_1]$) les DS relationnelles définies sur l'ensemble $A_k = \{A_{k1}, A_{k2}, \dots, A_{km}\}$.
- \mathbf{C}_p ($p \in [1..n_2]$) les DS objet définies sur l'ensemble $A_p = \{A_{p1}, A_{p2}, \dots, A_{pu}\}$.

■ Vue matérialisée hétérogène

- $V \subseteq \prod_{i \in [1..n]} T_i$; T_i est une relation de type R_i ou C_i

On note V ou $V(B)$ avec $B \subseteq \bigcup_{i \in [1..n]} A_i$; $n = n_1 + n_2$

Calcul des vues matérialisées complexes (2/2)

τ : fonction de transformation Objet \rightarrow Rel.

τ' : fonction de transformation Rel. \rightarrow Objet

■ Mises à jour

- Soit R'_1 (resp. C'_1) : l'ensemble des mises à jour de la relation R_1 (resp. C_1)
- On défini V' comme suit :

- Si toutes les relations sont de type relationnel

$$V'_R = \text{Projection sur B de } [R'_1 \times \prod_{j \in [2..n]} R_j]$$

- Si certaines relations sont de type objet

$$V'_C = \tau \left(\text{Projection sur B de } [T'_1 \times \prod_{j \in [2..n]} T_j] \right); T'_1 \text{ est une relation de type } R'_1 \text{ ou } C'_1$$

$$\text{Maintenance : } V = \begin{cases} V \cup V'_R; \text{ si toutes les DS sont de type relationnel} \\ \text{ou} \\ V \cup \tau'(V'_C); \text{ si au moins une des DS est de type objet} \end{cases}$$



Création de structures pour la maintenance des données complexes

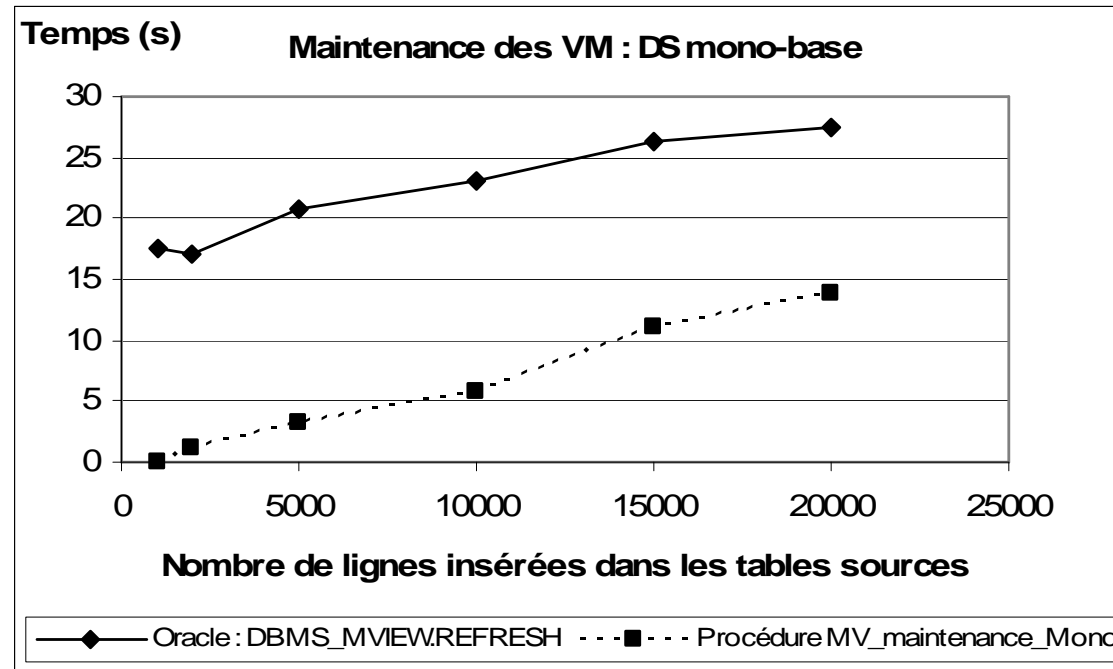
- VM → CREATE TABLE V (B) ;
- CREATE TABLE **META_V** (SOURCE, TABLE, LOG_TAB) ;
- Pour chaque donnée source S qui participe à B
 - Pour chaque Table T de S qui participe à B
 - CREATE TABLE **LOG\$_T** (Att. (T)) ;
 - CREATE TRIGGER **REPLIR_LOG\$_T**
(ensemble des mises à jour) ;
 - INSERT INTO **META_V** VALUES (S, T, LOG\$_T) ;
 - Fin Pour
- Fin Pour



Procédures de maintenance de VM complexes (1/3)

- Procédures de maintenance à la demande : (Calcul de V')
 - MV_maintenance_Mono (Mono-base)
 - MV_maintenance_Multi (Multi-bases)
- $$V = V \cup V'$$
- Tests sous Oracle 10g release 2 (10.2)
 - Vue matérialisée hétérogène
 - Taille de la VM : 160.000 lignes
 - Taille des mises à jour : 1.000 .. 20.000 lignes

Procédures de maintenance de VM complexes (2/3)



Mono-base

Figure. 2 – Comparaison des performances entre les deux procédures de maintenance (Oracle vs. MV_maintenance_mono)

Procédures de maintenance de VM complexes (3/3)

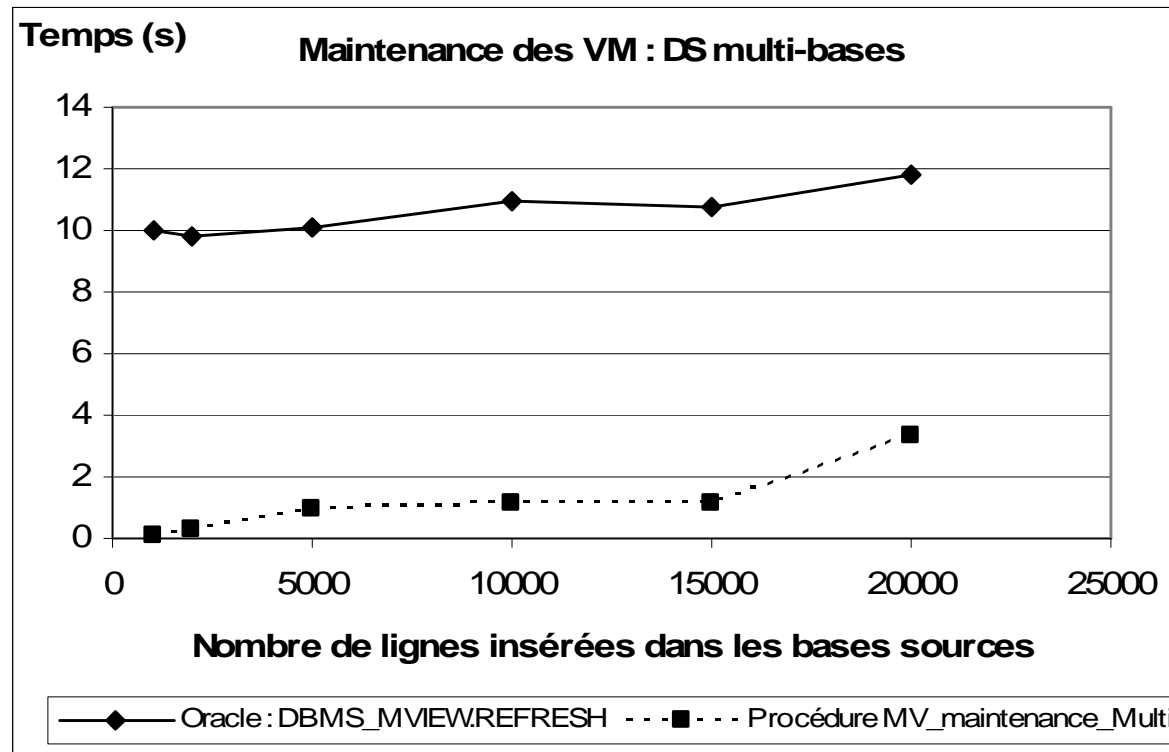


Figure. 3 – Comparaison des performances entre les deux procédures de maintenance (Oracle vs. MV_maintenance_multi)



Maintenance des entrepôts de données hétérogènes

■ Conclusion et perspectives

- Limites du SGBD Oracle à manipuler les données hétérogènes
- Procédures pour maintenir des VM hétérogènes sans les recalculer entièrement
- La prise en compte d'autres types de données (XML, ...)
- Critères pour déterminer le type des VM à créer (relationnelles, objet, XML, ...)



■ Rapports et publications :

– Publiés :

- **Badri M., Boufarès F., Ducateau C.F. et Gargouri F.,**
« Etat de l'art de la maintenance des entrepôts de données issus de systèmes d'information hétérogènes », *Cinquièmes Journées Scientifiques GEI*, Mars 2005, Sousse Tunisie, p 13-18.
- **Badri M., Boufarès F.,**
« Maintenance des vues matérialisées complexes sous Oracle », Rapport interne LIPN, Université Paris 13, Février-Mars 2006.

– Soumis :

- **Badri M., Boufarès F., Ducateau C.F. et Nefoussi R.,**
« Maintenance des Entrepôts de données Hétérogènes : Vues matérialisées complexes sous Oracle », *Congrès INFORSID 2006*.
- **Badri M., Hamdoun S., Boufarès F. et Lellahi K.,**
« Construction et Maintenance des Entrepôts de Données Hétérogènes, sous Oracle », *Journées Entrepôts de données et Analyse en ligne : EDA 2006*.



Merci pour votre attention

Des bases de données vers les entrepôts de données ...

27-03-2006

M. BADRI

20