

EDA'07

3èmes journées francophones
sur les Entrepôts de Données et l'Analyse en ligne
Poitiers, 7 et 8 Juin 2007



Évolution de modèle dans les entrepôts de données : existant et perspectives

Cécile Favre, Fadila Bentayeb, Omar Boussaid

**Laboratoire ERIC
Université Lumière Lyon 2**

Contexte et objectif (1/2)



- Entrepôt de données :
 - consolidation de données provenant de sources de données hétérogènes
 - réponse à des besoins d'analyse en vue d'une prise de décision, en offrant une modélisation multidimensionnelle des données
- Approche mixte de conception prenant en compte :
 - des **sources de données**
 - des **besoins d'analyse**

Contexte et objectif (2/2)



- Évolution des sources de données et/ou des besoins d'analyse
- Nécessité de faire **évoluer le modèle** de l'entrepôt de données
- Différentes solutions proposées dans la littérature : comparaison de celles-ci et perspectives de recherche qui en découlent

Plan

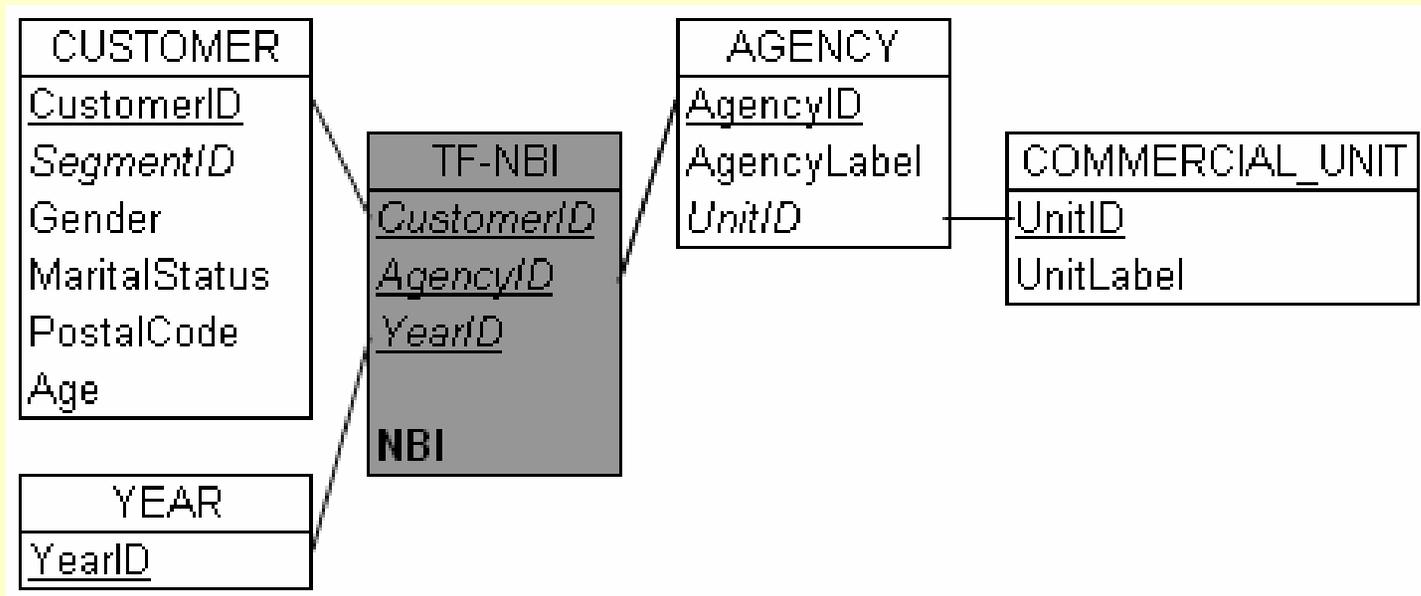


- Exemple
- Évolution de modèle
 - Évolution de schéma
 - Évolution de données
 - Cohérence des analyses
- Travaux existants :
 - Mise à jour de modèle
 - Modélisation temporelle
- Discussion
 - Critères de comparaison
 - Étude comparative
- Perspectives de recherche

Exemple



- Schéma multidimensionnel : analyse du NBI



Évolution de schéma



- Évolutions liées aux concepts (structure) :
 - Ajout / Suppression d'une dimension
 - Ajout / Suppression d'une mesure
 - Ajout / Suppression d'un niveau de granularité
 - ...
- ⇒ **Enrichissement ou appauvrissement** des possibilités d'analyse
- ⇒ **Cohérence des analyses** assurée

Évolution de données

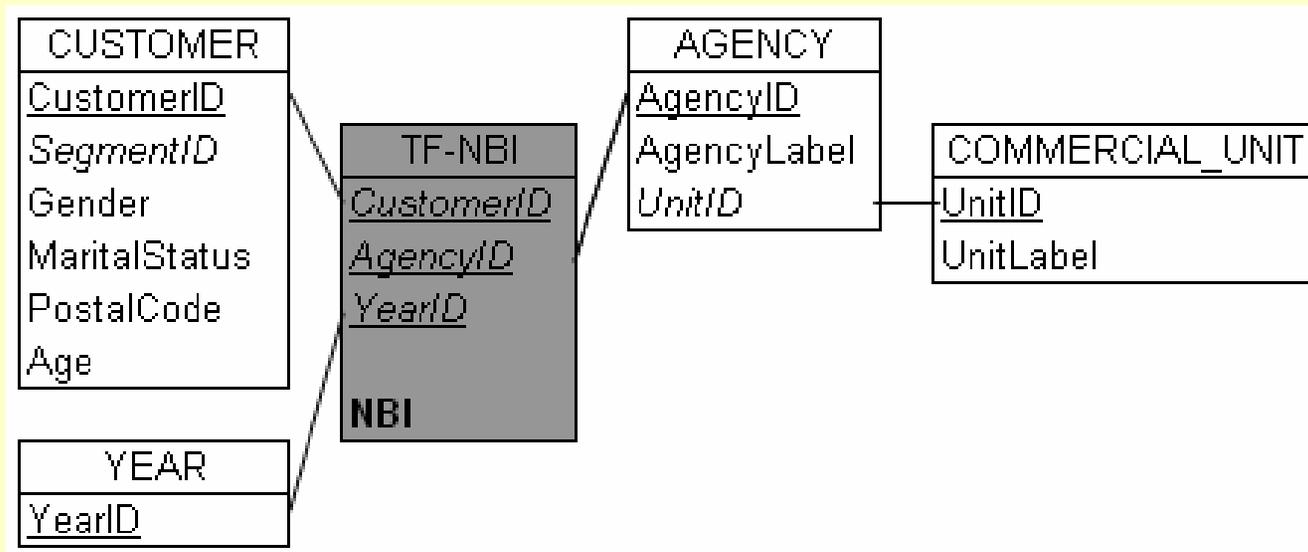


- Insertion / modification / suppression (instances)
- **Non-volatilité** des données (*Inmon, 2002*) : en théorie ni modification, ni suppression :
→ une requête émise sur les mêmes données à différents moments doit donner le même résultat...
- Mais... **évolution** :
 - des **mesures** (*Rizzi et Golfarelli, 2006*)
 - des **dimensions** et de **leurs hiérarchies**
(non indépendance des dimensions avec la dimension temps)

Cohérence des analyses



- «**Slowly Changing Dimension**» (Kimball, 1996) :
 - Écraser l'ancienne valeur
 - Créer un enregistrement valide pour une période donnée
 - Créer un attribut de plus pour conserver l'ancienne valeur



Travaux existants (1/3)



- **Mise à jour de modèle**

→ application des évolutions sans garder trace du précédent modèle

- Opérateurs d'évolution

(*Blaschka et al. 1999, Hurtado et al. 1999, Benitez-Guerrero 2002*)

- Enrichissement de hiérarchie

(*Mazon et Trujillo 2006, Favre et al. 2007*)

- Maintenance de vues matérialisées

(*Bellahsene, 2002*)

Travaux existants (2/3)



- **Modélisation temporelle :**

→recours à des extensions temporelles pour garder trace des évolutions avec des labels de validité sur :

- Instances : schéma en étoile temporel

(Bliujute et al., 1998)

- Liens d'agrégation : modèle qui gère l'évolution des chemins d'agrégation

(Mendelzon et Vaisman, 2000)

- Versions : gestion de différentes versions d'un modèle

(nombreux travaux)

Travaux existants (3/3)



- Versionnement :
 - Modèle (*Eder et Koncilia, 2000 ; Ravat et al., 2006 ; ...*)
 - Analyse avec choix d'une version (*Body et al., 2002-2003*)
 - « what-if analysis » (*Bebel et al., 2004*)
 - Analyse à travers les versions (*Morzy et Wrembel, 2004 ; Golfarelli et al., 2006*)
 - ...

Critères de comparaison



- **Caractéristiques des approches**
 - Historisation des dimensions
 - Cohérence des analyses
 - Approche orientée utilisateur
- **Mise en place**
 - Lors de la conception
 - Complexité
- **Performances**
 - Stockage
 - Temps de réponse aux analyses

Étude comparative



		Mise à jour de modèles			Modélisation temporelle		
		Opérateurs évolution	Enrichissement hiérarchie	Maintenance vues	Instances	Liens agrégation	Versions
Caractéristiques intrinsèques	Historisation des dimensions	-	-	-	+	+	+
	Cohérence des analyses	-	+	-	+	+	+
	Approche orientée utilisateurs	-	+	-	-	-	+/-
Mise en place	Nécessaire dès la conception	+	+	+	-	-	-
	Complexité	+	+	+	-	-	-
Performances	Stockage	+	+	+	-	-	-
	Temps de réponse aux analyses	+	+	+	-	-	-

Perspectives de recherche (1/3)



- **Cadre formel général** pour l'évolution de modèle :
lien avec les sources et les besoins d'analyse
 - S l'ensemble des sources de données
 - B l'ensemble des besoins d'analyse
 - M une structure multidimensionnelle
 - Cadre formel :

$$f(S, B) = M \quad \text{où } f \text{ représente notre cadre formel}$$

- Prise en compte de l'évolution grâce à f :

$$\Delta S \text{ ou } \Delta B \Rightarrow \Delta M ?$$

Perspectives de recherche (2/3)



- **Évolution de l'architecture décisionnelle :**

Au-delà de l'évolution du modèle, impact sur les autres composants de l'architecture décisionnelle : paradigme du « **workflow** » (*Bouzeghoub et al., 1999*)

- Processus ETL
- Magasin de données
- Structures redondantes d'optimisation (index, vues)
- Requêtes (reporting, requêtes «presse-bouton»,...)
- ...

Perspectives de recherche (3/3)

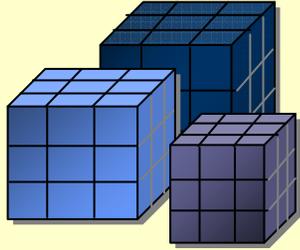


- **Évolution de la charge de requêtes**
 - Au-delà de la maintenance des structures redondantes d'optimisation ...
 - Évolution de la stratégie d'optimisation
 - Propager l'évolution du modèle (schéma et données) sur la charge de requêtes des utilisateurs qui sert dans le choix de l'optimisation (choix d'index, de vues à matérialiser)

Résumé



- Évolution de modèle dans les entrepôts
- Problème de la cohérence des analyses
- Travaux existants :
 - mise à jour de modèle
 - modélisation temporelle
- Critères de comparaison et étude comparative
- Perspectives de recherche



EDA'07

3èmes journées francophones
sur les Entrepôts de Données et l'Analyse en ligne
Poitiers, 7 et 8 Juin 2007

MERCI
DE VOTRE ATTENTION

Contacts :

Cécile Favre : cecile.favre@univ-lyon2.fr

Fadila Bentayeb : bentayeb@eric.univ-lyon2.fr

Omar Boussaid : omar.boussaid@univ-lyon2.fr