



Maintenance de charge pour l'optimisation des entrepôts de données évolutifs : aide à l'administrateur

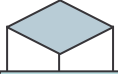
C. Favre, F. Bentayeb, O. Boussaid
Université de Lyon (ERIC-Lyon 2)

Contexte

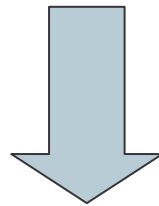


- Entrepôt de données évolutif :
 - Modèle d'entrepôt défini en fonction des sources de données et des besoins d'analyse
 - Évolution des sources et des besoins => évolution du modèle de l'entrepôt
- Gestion des performances de l'entrepôt de données par l'administrateur :
 - Nécessité d'optimisation des temps de réponse pour l'OLAP (On Line Analytical Processing)
 - Sélection des structures d'optimisation basée sur une charge (ensemble de requêtes)

Problème et solution proposée



Comment aider l'administrateur
à gérer de façon pro-active
les performances de l'entrepôt qui évolue ?



Aider l'administrateur
dans la maintenance de la charge
qui sert à l'évaluation des performances

Plan



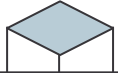
- État de l'art : utilisation et évolution de charge
- Aide à la maintenance de charge
 - Principe
 - Typologie des changements
 - Traitement des requêtes de la charge
- Étude de cas LCL-Le Crédit Lyonnais
- Éléments d'implémentation
- Conclusion et perspectives

État de l'art : utilisation de charge



- Temps de réponse aux requêtes des entrepôts de données (contexte relationnel) : structures d'optimisation telles que les index et les vues matérialisées [Rizzi & Saltarelli, 2003]
- Adaptation de la performance du système en fonction de son utilisation [Gallo, 2002]
- Sélection des structures d'optimisation : souvent basée sur une charge (ensemble de requêtes utilisateurs) [Aouiche *et al.*, 2006]

État de l'art : évolution de charge



- Évolution de modèle : évolution de schéma et de données [Rizzi & Golfarelli, 2006]
- Évolution de charge proposée précédemment :
 - Changement dans les probabilités des requêtes [Lawrence & Rau-Chaplin, 2006]
 - Ajout de requêtes [Theodoratos & Sellis, 2000]
 - => MAIS supposition de la consistance des requêtes
- Travail préliminaire [Favre *et al.*, 2007] : tentative de maintenance automatique de la charge

Aide à la maintenance de charge : principe

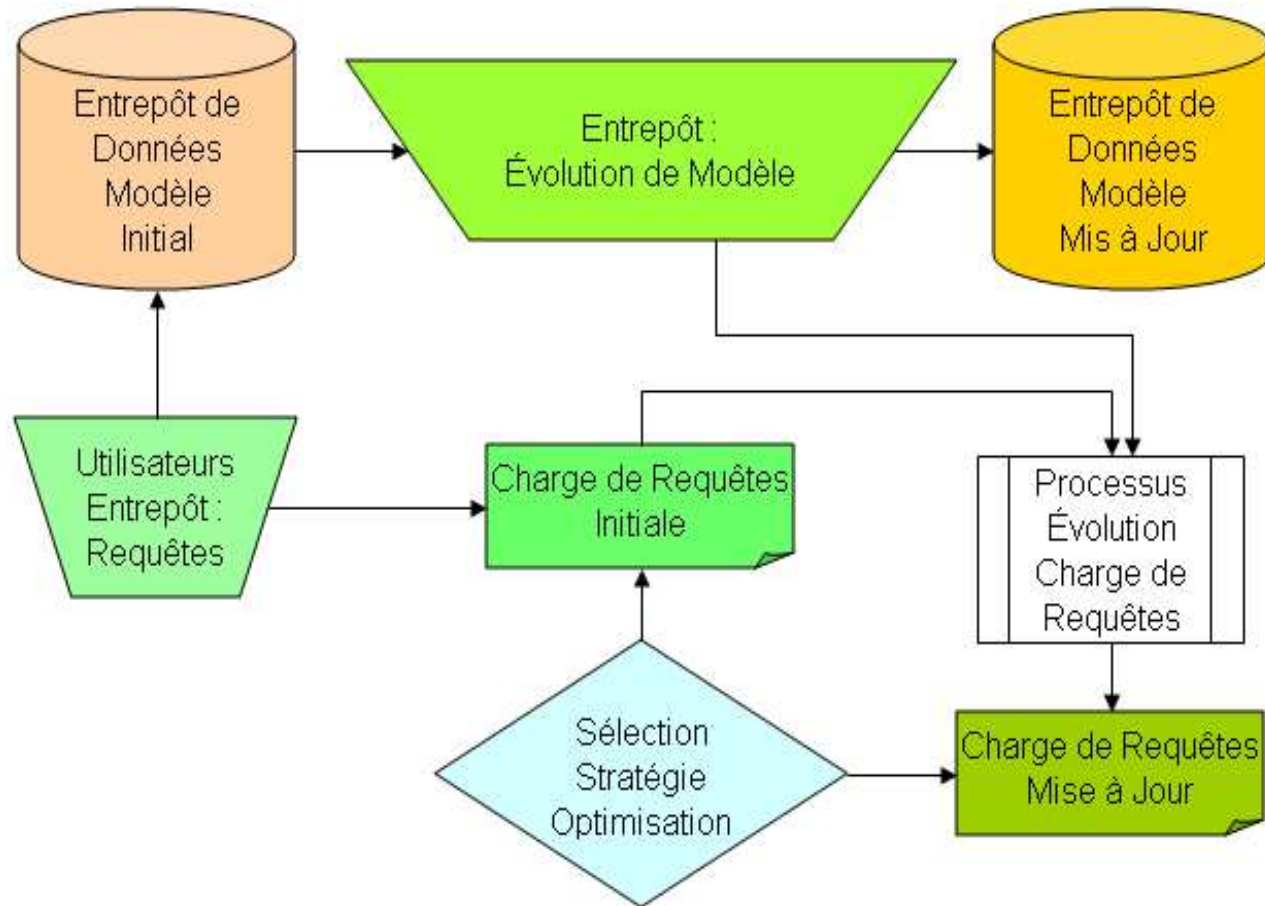
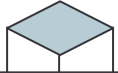


Fig 1. Architecture pour l'évolution de la charge

Aide à la maintenance de charge : principe



- Pro-activité de l'administrateur en évitant une nouvelle charge pour sélectionner la configuration d'optimisation
- Maintenance de la charge :
 - Maintenir consistantes les requêtes de la charge en fonction de l'évolution du modèle
 - Créer de nouvelles requêtes pour prendre en compte les nouvelles possibilités d'analyse

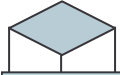
Aide à la maintenance de charge : typologie des changements



- Typologie des changements pour l'évolution de schéma

Rôle dans le modèle	Type	Opération	Mise à jour requête(s)	Suppression requête(s)	Création requête(s)
Dimension et Niveau	Table	Création			X
Dimension et Niveau	Table	Suppression	X	X	
Dimension et Niveau	Table	Mise à jour	X		
Fait	Table	Mise à jour	X		
Mesure	Attribut	Création			X
Mesure	Attribut	Suppression	X	X	
Mesure	Attribut	Mise à jour	X		
Descripteur Dimension	Attribut	Création			X
Descripteur Dimension	Attribut	Suppression	X	X	
Descripteur Dimension	Attribut	Mise à jour	X		

Aide à la maintenance de charge : traitement des requêtes de la charge

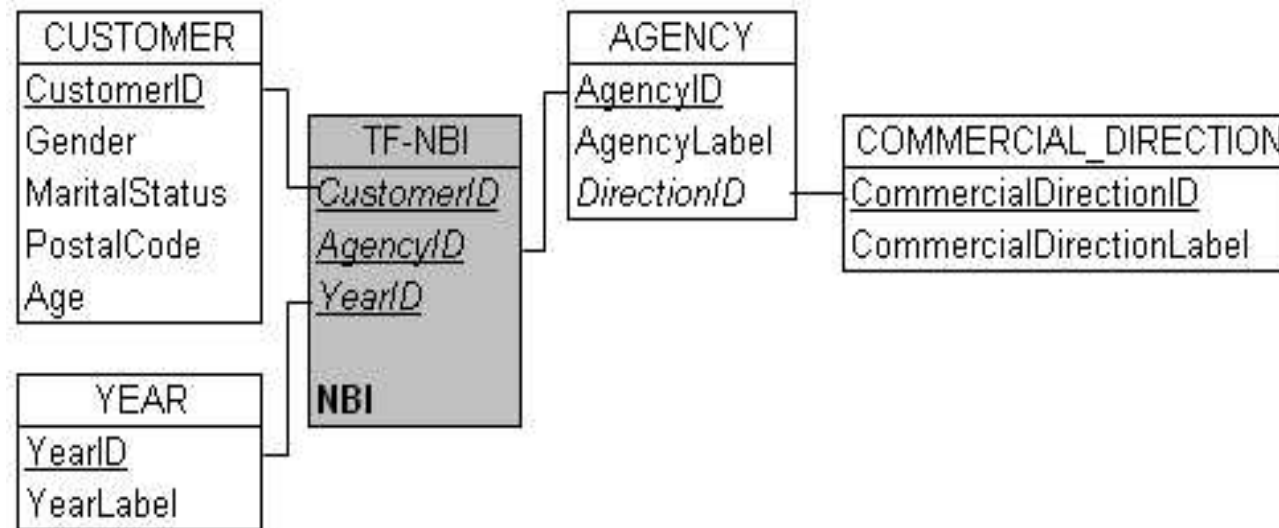


- Traitement des requêtes
 - Maintenance de requêtes
 - Contexte de détection des requêtes impactées par les changements :
 - Matrice requêtes-tables
 - Matrice requêtes-attributs
 - Utilisation de la typologie pour suggérer les modifications à apporter
 - Création de requêtes
 - Suggestion des éléments pouvant faire l'objet de création de requête(s)
 - Mise à jour des matrices du contexte de détection

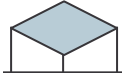
Étude de cas LCL-Le Crédit Lyonnais (1/10)



- Schéma initial pour l'analyse du NBI



Étude de cas LCL-Le Crédit Lyonnais (2/10)



- Extrait de la charge initiale

Q1: **SELECT** Gender, Agency Label, YearLabel, AVG (NBI) **FROM** AGENCY, YEAR, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND YEAR.YearID=TF-NBI.YearID **GROUP BY CUBE** (Gender, Agency Label, YearLabel);

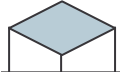
Q2: **SELECT** CommercialDirectionLabel, YearLabel, AVG (NBI) **FROM** AGENCY, COMMERCIAL_DIRECTION, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND COMMERCIAL_DIRECTION.CommercialDirectionID= AGENCY.DirectionID **GROUP BY CUBE** (CommercialDirectionLabel, YearLabel);

Q3: **SELECT** MaritalStatus, YearLabel, SUM (NBI) **FROM** CUSTOMER, TF-NBI WHERE CUSTOMER.CustomerID=TF-NBI.CustomerID **GROUP BY CUBE** (MaritalStatus, YearLabel);

Q4: **SELECT** CommercialDirectionLabel, AgencyLabel, SUM (NBI) **FROM** AGENCY, COMMERCIAL_DIRECTION, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND COMMERCIAL_DIRECTION.CommercialDirectionID=AGENCY.DirectionID AND YearLabel='2000' **GROUP BY ROLLUP** (CommercialDirectionLabel, AgencyLabel);

Q5: **SELECT** AgencyLabel, AVG (NBI) **FROM** AGENCY, COMMERCIAL_DIRECTION, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND COMMERCIAL_DIRECTION.CommercialDirectionID=AGENCY.DirectionID AND COMMERCIAL_DIRECTION.CommercialDirectionID=8 **GROUP BY** AgencyLabel;

Étude de cas LCL-Le Crédit Lyonnais (3/10)



- Contexte de détection de changements :
matrice requêtes-tables

	t1	t2	t3	t4	t5
q1	1	1	1	0	0
q2	1	1	0	0	1
q3	1	0	0	1	0
q4	1	1	0	0	1
q5	1	1	0	0	1

t1	TF-NBI
t2	AGENCY
t3	YEAR
t4	CUSTOMER
t5	COMMERCIAL_DIRECTION

Étude de cas LCL-Le Crédit Lyonnais (4/10)



- Contexte de détection de changements :
matrice requêtes-attributs

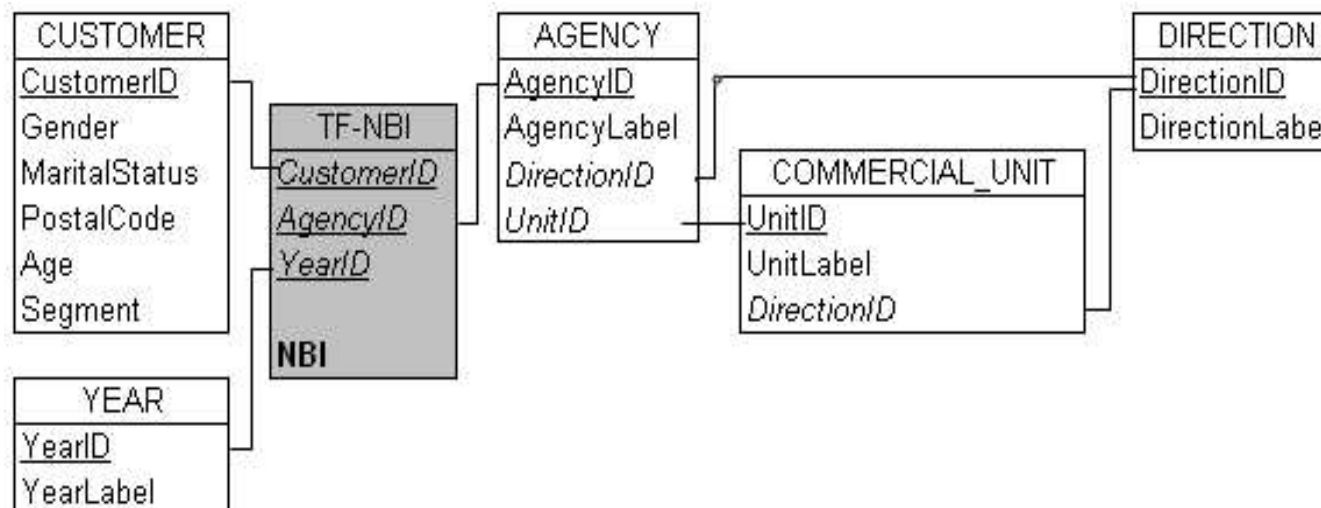
	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14
q1	1	1	0	0	0	0	1	0	1	0	1	1	1	1
q2	1	0	1	1	1	0	0	0	1	0	1	0	0	1
q3	0	0	0	0	0	1	0	1	0	1	1	0	0	1
q4	1	1	1	1	1	0	0	0	1	0	1	0	0	1
q5	1	1	1	1	0	0	0	0	1	0	1	0	0	0

a1	AGENCY.AgencyID
a2	AGENCY.AgencyLabel
a3	AGENCY.DirectionID
a4	COMMERCIAL_DIRECTION.CommercialDirectionID
a5	COMMERCIAL_DIRECTION.CommercialDirectionLabel
a6	CUSTOMER.CustomerID
a7	CUSTOMER.Gender
a8	CUSTOMER.MaritalStatus
a9	TF-NBI.AgencyID
a10	TF-NBI.CustomerID
a11	TF-NBI.NBI
a12	TF-NBI.YearID
a13	YEAR.YearID
a14	YEAR.YearLabel

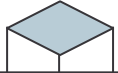
Étude de cas LCL-Le Crédit Lyonnais (5/10)



- Évolution du schéma pour l'analyse du NBI
 1. Renommage du niveau COMMERCIAL_DIRECTION en DIRECTION
 2. Renommage des attributs CommercialDirectionID et CommercialDirectionLabel en DirectionID et DirectionLabel
 3. Ajout de l'attribut Segment dans la dimension CUSTOMER
 4. Insertion du niveau COMMERCIAL_UNIT entre les niveaux AGENCY et DIRECTION

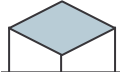


Étude de cas LCL-Le Crédit Lyonnais (6/10)



- Modifications à apporter d'après la typologie :
 1. Mise à jour de(s) requête(s) incluant la table COMMERCIAL_DIRECTION (DIRECTION)
 2. Mise à jour de(s) requête(s) incluant les attributs CommercialDirectionID et CommercialDirectionLabel (DirectionID et DirectionLabel)
 3. Création de requête(s) portant sur l'attribut Segment de la dimension CUSTOMER
 4. Création de requête(s) portant sur le niveau créé COMMERCIAL_UNIT

Étude de cas LCL-Le Crédit Lyonnais (7/10)



- Détection dans la matrice requêtes-tables :
Changement 1. : table COMMERCIAL_DIRECTION (DIRECTION)

	t1	t2	t3	t4	t5
q1	1	1	1	0	0
q2	1	1	0	0	1
q3	1	0	0	1	0
q4	1	1	0	0	1
q5	1	1	0	0	1

t1	TF-NBI
t2	AGENCY
t3	YEAR
t4	CUSTOMER
t5	COMMERCIAL_DIRECTION

Étude de cas LCL-Le Crédit Lyonnais (8/10)



- Détection dans la matrice requêtes-attributs :
Changement 2. : attributs CommercialDirectionID (DirectionID) et CommercialDirectionLabel (DirectionLabel)

	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14
q1	1	1	0	0	0	0	1	0	1	0	1	1	1	1
q2	1	0	1	1	1	0	0	0	1	0	1	0	0	1
q3	0	0	0	0	0	1	0	1	0	1	1	0	0	1
q4	1	1	1	1	1	0	0	0	1	0	1	0	0	1
q5	1	1	1	1	0	0	0	0	1	0	1	0	0	0

a1	AGENCY.AgencyID
a2	AGENCY.AgencyLabel
a3	AGENCY.DirectionID
a4	COMMERCIAL_DIRECTION.CommercialDirectionID
a5	COMMERCIAL_DIRECTION.CommercialDirectionLabel
a6	CUSTOMER.CustomerID
a7	CUSTOMER.Gender
a8	CUSTOMER.MaritalStatus
a9	TF-NBI.AgencyID
a10	TF-NBI.CustomerID
a11	TF-NBI.NBI
a12	TF-NBI.YearID
a13	YEAR.YearID
a14	YEAR.YearLabel

Étude de cas LCL-Le Crédit Lyonnais (9/10)



- Suggestions communiquées à l'administrateur en fonction des requêtes :
 - Maintenance de requêtes :
 - Q2 : table COMMERCIAL_DIRECTION, attributs CommercialDirectionID et CommercialDirectionLabel
 - Q4 : table COMMERCIAL_DIRECTION, attributs CommercialDirectionID et CommercialDirectionLabel
 - Q5 : table COMMERCIAL_DIRECTION, attributs CommercialDirectionID
 - Création de requêtes :
 - Niveau COMMERCIAL_UNIT et descripteur de dimension Segment

Étude de cas LCL-Le Crédit Lyonnais (10/10)



- Extrait de la charge mise à jour

Q1: **SELECT** Gender, AgencyLabel, YearLabel, AVG (NBI) **FROM** AGENCY, YEAR, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND YEAR.YearID=TF-NBI.YearID **GROUP BY CUBE** (Gender, AgencyLabel, YearLabel);

Q2: **SELECT** *DirectionLabel*, YearLabel, AVG (NBI) **FROM** AGENCY, *DIRECTION*, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND *DIRECTION.DirectionID*=AGENCY.DirectionID **GROUP BY CUBE** (*DirectionLabel*, YearLabel);

Q3: **SELECT** MaritalStatus, YearLabel, SUM (NBI) **FROM** CUSTOMER, TF-NBI WHERE CUSTOMER.CustomerID=TF-NBI.CustomerID **GROUP BY CUBE** (MaritalStatus, YearLabel);

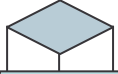
Q4: **SELECT** *DirectionLabel*, AgencyLabel, SUM (NBI) **FROM** AGENCY, *DIRECTION*, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND *DIRECTION.DirectionID*=AGENCY.DirectionID AND YearLabel='2000' **GROUP BY ROLLUP** (*DirectionLabel*, AgencyLabel);

Q5: **SELECT** AgencyLabel, AVG (NBI) **FROM** AGENCY, *DIRECTION*, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND *DIRECTION.DirectionID*=AGENCY.DirectionID AND *DIRECTION.DirectionID*=8 **GROUP BY** AgencyLabel;

Q6: **SELECT** *AgencyLabel*, *UnitLabel*, AVG (NBI) **FROM** AGENCY, *COMMERCIAL_UNIT*, TF-NBI WHERE AGENCY.AgencyID=TF-NBI.AgencyID AND *COMMERCIAL_UNIT.UnitID*=AGENCY.UnitID **GROUP BY ROLLUP** (*UnitLabel* , *AgencyLabel*);

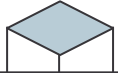
Q7: **SELECT** *Segment*, AVG (NBI) **FROM** CUSTOMER, TF-NBI WHERE CUSTOMER.CustomerID=TF-NBI.CustomerID AND **GROUP BY** *Segment*;

Éléments d'implémentation



- Implémentation en cours d'un outil d'aide à la maintenance sous Oracle 10g avec interface Web en php
 - Détection des évolutions du schéma de l'entrepôt pour lancer le processus de maintenance de charge grâce à des triggers sur le méta-modèle de l'entrepôt
 - Matrices du contexte de détection implémentées sous forme de tables relationnelles
 - Parseur de la charge pour la construction du contexte de détection
 - Requêtes sur les matrices pour détecter les requêtes impactées

Conclusions



- Évolution de charge en réponse à l'évolution du modèle de l'entrepôt
- Prise en compte de l'évolution de schéma
- Aide à la maintenance de charge basée sur un contexte de détection
- Possibilité d'une gestion pro-active des performances par l'administrateur

Perspectives



- Prise en compte de l'évolution des données : impact dans les clauses « WHERE »
- Vers une version semi-automatique :
 - Maintenance de requêtes automatique validée par l'administrateur
 - Génération paramétrée de requêtes soumises à l'administrateur
- Étude de performances de notre approche dans un contexte réel (charge comportant un grand nombre de requêtes, fréquences de changements du modèle, volumétrie des données,...)

4èmes Journées Francophones sur les Entrepôts de Données et l'Analyse en ligne
EDA'08 Toulouse – 5 et 6 juin 2008

Merci de votre attention !

Contact :
cecile.favre@univ-lyon2.fr