

Atelier AIDE
EGC 2012

Recommandation Interactive de Requêtes Décisionnelles

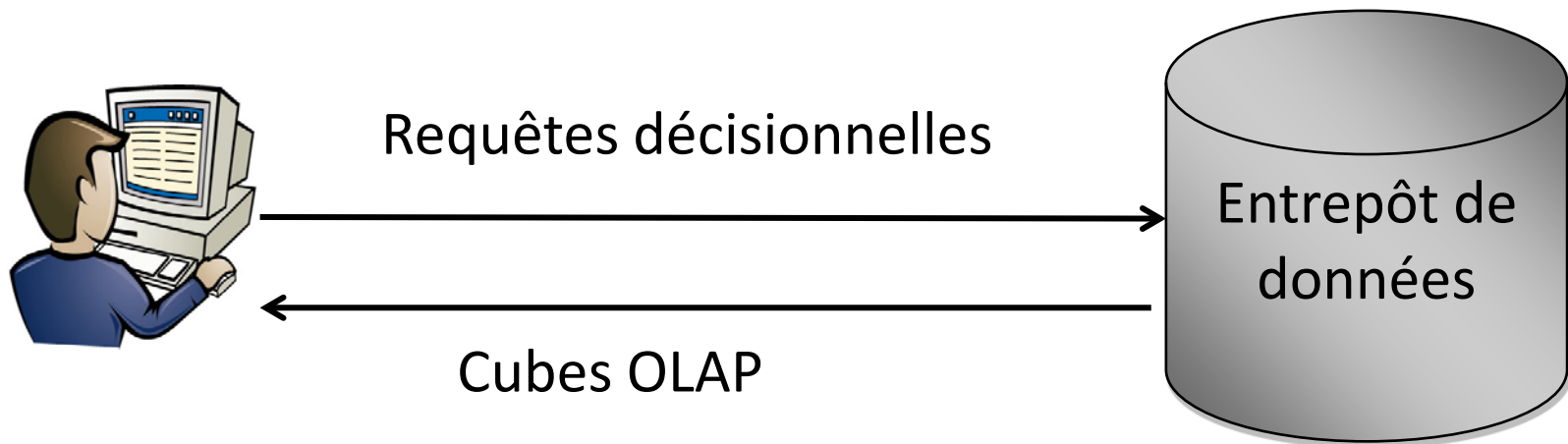
Rym KHEMIRI

Fadila BENTAYEB, Omar BOUSSAID

Laboratoire ERIC
Université Lumière Lyon2

Contexte

- Les entrepôts de données : exploitation de données organisées en fonction d'une vision d'analyse



Contexte

- ▶ Prolifération de données
 - Résultats massifs en réponse aux requêtes des utilisateurs
 - surcharge d'informations
 - Temps de réponse de plus en plus longs
 - passage à l'échelle difficile
- ▶ Requêtes ambigües/incomplètes
 - Réponses non pertinentes

Plan

Problématique

Objectifs

Moyen

Etat de l'art

Approche proposée

Démarche

Exemple

Conclusion

Perspectives

Problématique

- ▶ Comment déterminer une requête qui répond au mieux aux besoins de l'utilisateur ?
- ▶ L'utilisateur est laissé seul lors de la définition d'une requête
- ▶ Difficile de savoir exactement quelles informations rechercher et où les chercher
 - Nécessité de poser plusieurs requêtes
 - Perte de temps

Objectifs

- ▶ Guider l'utilisateur dans son exploration des données
- ▶ Assister l'utilisateur dans la définition de ses requêtes en exploitant les expériences de ses interrogations passées
- ▶ Aider l'utilisateur à formuler sa requête décisionnelle de manière interactive

Recours à la
recommandation

Aide à l'écriture de
requêtes

Requêtes plus
pertinentes

Moyen

- ▶ Exploitation des logs des requêtes
- ▶ Utilisation de la fouille de données
- ▶ Méthode des itemsets fréquents
 - Les attributs/mesures fréquemment posés peuvent être pertinents dans les requêtes futures
- ▶ Recommandation des attributs/mesures

État de l'art

- ▶ **Recommandation dans les bases de données et les entrepôts de données**
 - Enrichissement de requêtes
 - Anticipation de requêtes
 - Sessions utilisateur
 - Log de requêtes/ préférences utilisateur
 - Ensemble de requêtes/ requête entière /fragment de requête

- ▶ **Critères de recommandation**
 - Type de recommandation
 - Moment de la recommandation
 - Objet d'exploitation

Travaux de recommandation : synthèse

Travaux	Type de recommandation		Moment de la recommandation			Objet d'exploitation		
	Cognitif	Collaboratif	Avant	Pendant	Après	Profil	Fichier log	Source externe
Stefanidis et al. (2009)	×	×			×		×	×
Yang et al. (2009)		×		×			×	
Chatzopoulou et al. (2009), Chatzopoulou et al. (2011)		×		×			×	
Khoussainova et al. (2010)		×		×			×	
(Jerbi et al., 2009a)	×		×	×	×	×		
(Giacometti et al., 2009a), Giacometti et al. (2009b)		×	×			×		
(Golfarelli et Rizzi, 2009), Golfarelli et al. (2011)		×	×			×		
Aligon et al. (2011)		×	×			×	×	

Etat de l'art : Bilan

- ▶ La recommandation de requêtes fournit deux fonctionnalités :
 1. Aide à l'écriture de requêtes : pas à pas
 2. Proposition de requêtes complètes
- ▶ Seule la deuxième fonctionnalité a été étudiée par la communauté des entrepôts de données
- ▶ Aucun travail sur l'aide à la formulation de requêtes décisionnelles

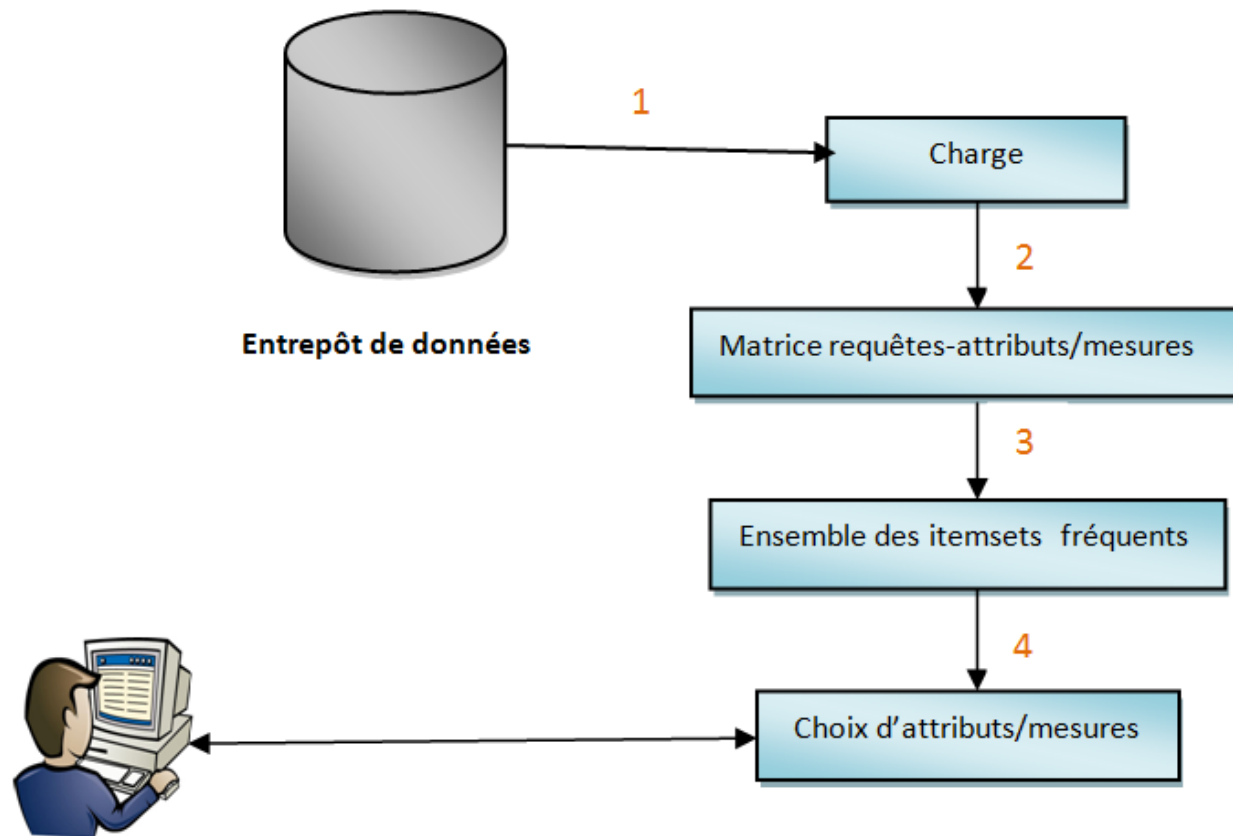
Etat de l'art : Positionnement

- ▶ Notre approche est différente des approches existantes
 - Aide interactive à la formulation de requêtes décisionnelles
 - Requêtes décisionnelles écrites en SQL
 - Recommandation par étapes
 - clause par clause
 - pour chaque clause, recommandation des attributs/mesures
 - Recours à la fouille de données
 - Recommander les attributs/mesures fréquemment utilisés
 - Aide non intrusive

Notre approche

- ▶ Guider l'utilisateur dans sa démarche de construction de sa requête décisionnelle
- ▶ Recherche des itemsets fréquents extraits des requêtes soumises précédemment
- ▶ Construction des usages de l'utilisateur à la volée
- ▶ Les attributs/mesures préfixés par la clause dans laquelle ils apparaissent
- ▶ Interaction du système en temps réel avec l'utilisateur : construction assistée de requêtes

Principe général



1 Extraction de requêtes résolues par le système

2 Extraction des attributs

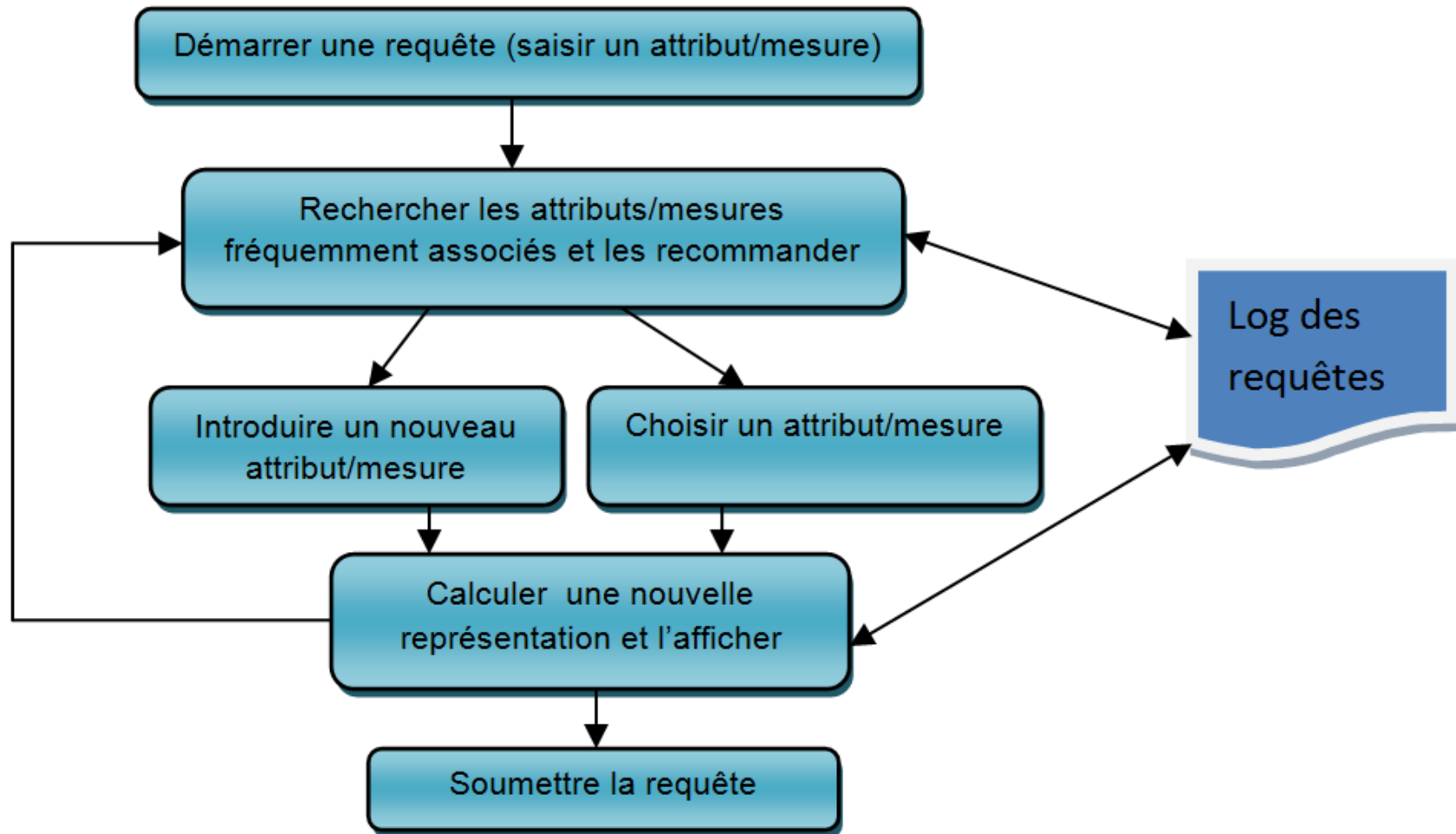
3 Application de l'algorithme Apriori

4 Génération de recommandations à partir des itemsets fréquents

Démarche

- ▶ **Prétraitement des données**
 - Extraction de la charge de requêtes
 - Extraction des clauses et des attributs/mesures
 - Construction du contexte d'extraction:
matrice « requêtes attributs/mesures »
- ▶ **Recherche des itemsets fréquents (Apriori)**
- ▶ **Génération de recommandations**

Scénario de la recommandation



Aide à l'écriture de requête décisionnelle

Select **SALES.PROD_ID**

From **SALES**

Where

Group by

Having

Order by

SALES.TIME_ID
SUM(QUANTITY_SOLD)
SUM(AMOUNT_SOLD)
SALES.PROMO_ID
AVG(AMOUNT_SOLD)
SALES.CUST_ID

Aide à l'écriture de requête décisionnelle

Select SALES.PROD_ID , **SUM(QUANTITY_SOLD)**
From SALES
Where
Group by
Having
Order by

SALES.TIME_ID
SUM(AMOUNT_SOLD)
SALES.PROD_ID

Aide à l'écriture de requête décisionnelle

Select SALES.PROD_ID, SUM(QUANTITY_SOLD), **SALES.TIME_ID**
From SALES, **TIME**

Where

Group by

Having

Order by

SALES.PROD_ID
SUM(AMOUNT_SOLD)

Aide à l'écriture de requête décisionnelle

Select
From
Where
Group by
Having
Order by

SALES.PROD_ID, SUM(QUANTITY_SOLD) , SALES.TIME_ID
SALES, TIME

SALES.TIME_ID=TIME.TIME_ID
TIME.FISCAL_YEAR='2010'
SALES.PROMO_ID=PROMOTION.PROMO_ID
SALES.TIME_ID=TIME.TIME_ID
SALES.PROD_ID=PRODUCT.PROD_ID
PROMOTION.PROMO_CATEGORY='CHEESE'
SALES.CUST_ID=CUSTOMER.CUST_ID
CUSTOMER.CUST_MARITAL_STATUS='SINGLE'
PRODUCT.PROD_CATEGORY='WOMEN'

Aide à l'écriture de requête décisionnelle

Select
From
Where
Group by
Having
Order by

SALES.PROD_ID, SUM(QUANTITY_SOLD) , SALES.TIME_ID
SALES, TIME

TIME.FISCAL_YEAR='2010'

SALES.TIME_ID=TIME.TIME_ID
SALES.PROMO_ID=PROMOTION.PROMO_ID
SALES.TIME_ID=TIME.TIME_ID
SALES.PROD_ID=PRODUCT.PROD_ID
PROMOTION.PROMO_CATEGORY='CHEESE'
SALES.CUST_ID=CUSTOMER.CUST_ID
CUSTOMER.CUST_MARITAL_STATUS='SINGLE'
PRODUCT.PROD_CATEGORY='WOMEN'

Aide à l'écriture de requête décisionnelle

Select
From
Where

SALES.PROD_ID, SUM(QUANTITY_SOLD) , SALES.TIME_ID
SALES, TIME, **PROMOTION**

TIME.FISCAL_YEAR='2010' AND
PROMOTION.PROMO_CATEGORY='CHEESE'

Group by
Having
Order by

SALES.TIME_ID=TIME.TIME_ID
SALES.PROMO_ID=PROMOTION.PROMO_ID
SALES.TIME_ID=TIME.TIME_ID
SALES.PROD_ID=PRODUCT.PROD_ID
SALES.CUST_ID=CUSTOMER.CUST_ID
CUSTOMER.CUST_MARITAL_STATUS='SINGLE'
PRODUCT.PROD_CATEGORY='WOMEN'

Aide à l'écriture de requête décisionnelle

Select SALES.PROD_ID, SUM(QUANTITY_SOLD) , SALES.TIME_ID

From SALES, TIME, PROMOTION

Where TIME.FISCAL_YEAR='2010' AND
PROMOTION.PROMO_CATEGORY='CHEESE'

Group by

Having
Order by

SALES.TIME_ID

Rollup(sales.prod_id,sales.promo_id,sales.time_id)

Cube(sales.cust_id,sales.promo_id,sales.prod_id)

Aide à l'écriture de requête décisionnelle

Select SALES.PROD_ID, SUM(QUANTITY_SOLD), SALES.TIME_ID

From SALES, TIME, PROMOTION

Where TIME.FISCAL_YEAR='2010' AND
PROMOTION.PROMO_CATEGORY='CHEESE'

Group by
Having
Order by

SALES.TIME_ID

Rollup(sales.prod_id,sales.promo_id,sales.time_id)

Cube(sales.cust_id,sales.promo_id,sales.prod_id)

Aide à l'écriture de requête décisionnelle

Select SALES.PROD_ID, SUM(QUANTITY_SOLD), SALES.TIME_ID

From SALES, TIME, PROMOTION

Where TIME.FISCAL_YEAR='2010' AND
PROMOTION.PROMO_CATEGORY='CHEESE'

Group by Rollup(sales.prod_id,sales.promo_id,sales.time_id)

Having

Order by

SALES.TIME_ID

Cube(sales.cust_id,sales.promo_id,sales.prod_id)

Aide à l'écriture de requête décisionnelle

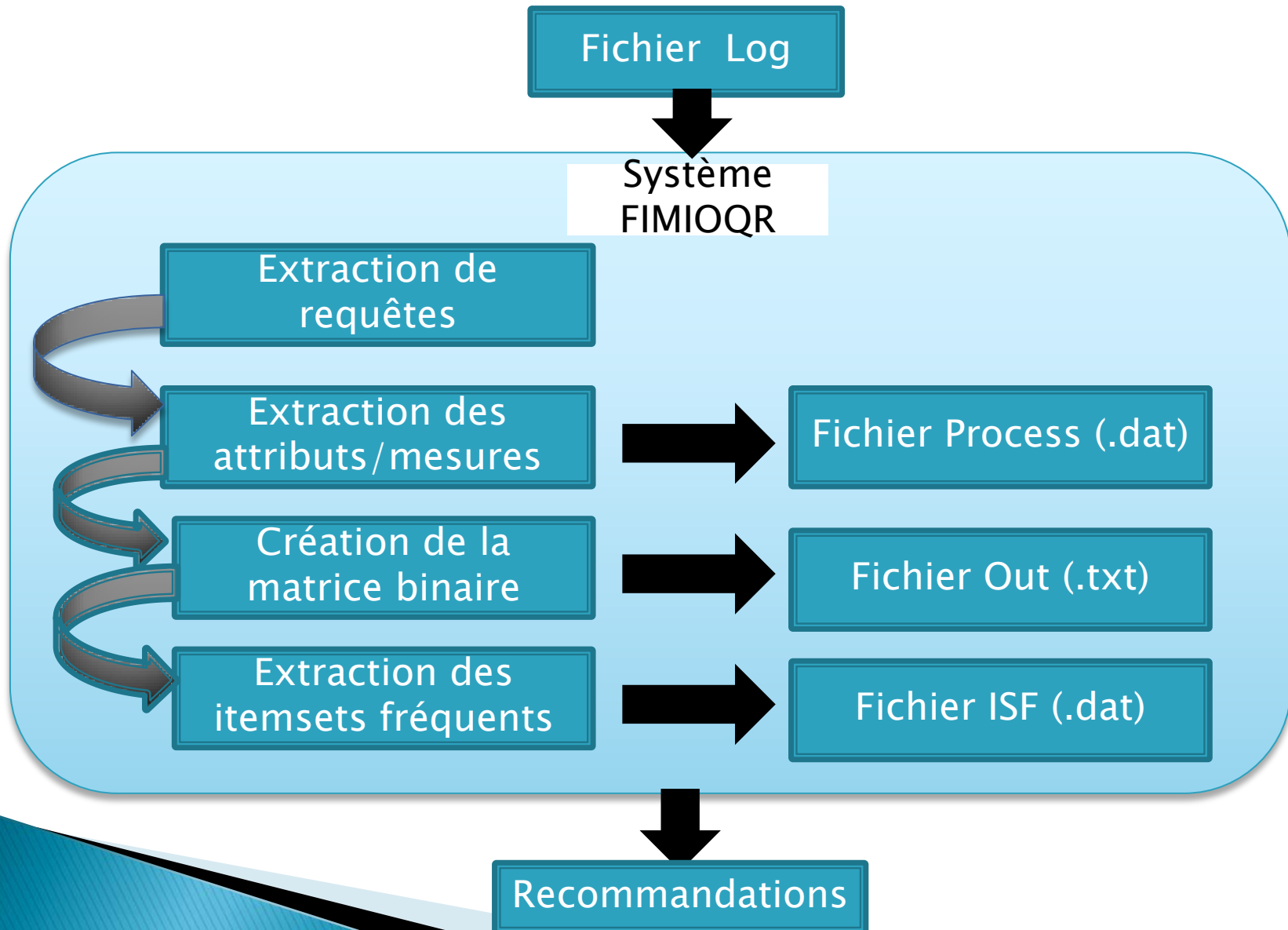
Requête obtenue:

```
SELECT SALES.PROD_ID, SUM(QUANTITY_SOLD), SALES.TIME_ID  
FROM SALES, TIME, PROMOTION  
WHERE TIME.FISCAL_YEAR='2010'  
AND PROMOTION.PROMO_CATEGORY='CHEESE'  
GROUP BY  
Rollup(sales.prod_id,sales.promo_id,sales.time_id)
```

Implémentation

- ▶ Système FIMIOQR (Frequent Itemsets Mining for Interactive OLAP Query Recommendation) en Java dans l'environnement Netbeans
- ▶ <http://eric.univ-lyon2.fr/~bentayeb/logiciels.html>
- ▶ Application sur une charge de 100 requêtes soumises à l'entrepôt de données FoodMart

Processus de recommandation



Extraction des attributs

Item	Clause
SALES.TIME_ID	Select
SUM(QUANTITY_SOLD)	Select
SUM(AMOUNT_SOLD)	Select
SALES.PROD_ID	Select
SALES.PROMO_ID	Select
AVG(AMOUNT_SOLD)	Select
SALES.CUST_ID	Select
SALES	From
TIME	From
PRODUCT	From
PROMOTION	From
CUSTOMER	From
SALES.TIME_ID=TIME.TIME_ID	Where
TIME.FISCAL_YEAR='2010'	Where
SALES.PROD_ID=PRODUCT.PROD_ID	Where
SALES.PROMO_ID=PROMOTION.PROMO_ID	Where
PROMOTION.PROMO_CATEGORY='CHEESE'	Where
SALES.CUST_ID=CUSTOMER.CUST_ID	Where
CUSTOMER.CUST_MARITAL_STATUS='SINGLE'	Where
PRODUCT.PROD_CATEGORY='WOMEN'	Where
SALES.TIME_ID	GroupBy

Recherche des itemsets fréquents

► Application de l'algorithme Apriori

Item Set	Nombre d'item	Support
[Select.SALES.TIME_ID]	1	3
[Select.SUM(QUANTITY_SOLD)]	1	3
[Select.SUM(AMOUNT_SOLD)]	1	3
[Select.SALES.PROD_ID]	1	3
[Select.SALES.PROMO_ID]	1	2
[Select.AVG(AMOUNT_SOLD)]	1	2
[Select.SALES.CUST_ID]	1	2
[From.SALES]	1	3
[From.TIME]	1	3
[From.PRODUCT]	1	2
[From.PROMOTION]	1	2
[From.CUSTOMER]	1	1
[Where.SALES.TIME_ID=TIME.TIME_ID]	1	2
[Where.TIME.FISCAL_YEAR='2010']	1	1
[Where.SALES.PROD_ID=PRODUCT.PROD_ID]	1	2
[Where.SALES.PROMO_ID=PROMOTION.PROMO_ID]	1	2
[Where.PROMOTION.PROMO_CATEGORY='CHEESE']	1	1
[Where.SALES.CUST_ID=CUSTOMER.CUST_ID]	1	1
[Where.CUSTOMER.CUST_MARITAL_STATUS='SINGLE']	1	1
[Where.PRODUCT.PROD_CATEGORY='WOMEN']	1	1
[GroupBy.SALES.TIME_ID]	1	2
[GroupBy.Rollup(sales.prod_id,sales.promo_id,sales.time_id)]	1	1
[GroupBy.Cube(sales.cust_id,sales.promo_id,sales.prod_id)]	1	1
[Select.SALES.TIME_ID Select.SUM(QUANTITY_SOLD)]	2	3
[Select.SALES.TIME_ID Select.SUM(AMOUNT_SOLD)]	2	3

Génération de recommandations

Configuration Processing Item set Frequent **Recommandation**

SELECT FROM

WHERE GROUP BY

HAVING ORDER BY

Nos recommandations :

```
SALES.TIME_ID
SUM(QUANTITY_SOLD)
SUM(AMOUNT_SOLD)
SALES.PROMO_ID
AVG(AMOUNT_SOLD)
SALES.CUST_ID
```

Items du SELECT	Items du WHERE	Items GROUPBY	Items HAVING	Items ORDERBY
<pre>SALES.TIME_ID SUM(QUANTITY_SOLD) SUM(AMOUNT_SOLD) SALES.PROMO_ID AVG(AMOUNT_SOLD) SALES.CUST_ID</pre>	<pre>SALES,TIME_ID=TIME,TIME_I TIME,FISCAL_YEAR='2010' SALES.PROD_ID=PRODUCT.P SALES.PROMO_ID=PROMOTIO PROMOTION.PROMO_CATEGO SALES,CUST_ID=CUSTOMER, CUSTOMER.CUST_MARITAL_S PRODUCT.PROD_CATEGORY=</pre>	<pre>SALES,TIME_ID Rollup(sales,prod_id,sales,pron Cube(sales,cust_id,sales,promo</pre>		

Conclusion

- ▶ Recommandation interactive de requêtes décisionnelles
- ▶ Assistance à la construction de la requête
- ▶ Implémentation du prototype FIMIOQR (Frequent Itemsets Mining for Interactive OLAP Query Recommendation)

Perspectives

- ▶ Intégration de FIMIOQR dans un SGBD
- ▶ Veille sur les recommandations les plus acceptées



classement des recommandations

- ▶ Mettre plus de poids sur les mesures

Merci pour votre attention