

Document Complémentaire au chapitre IV

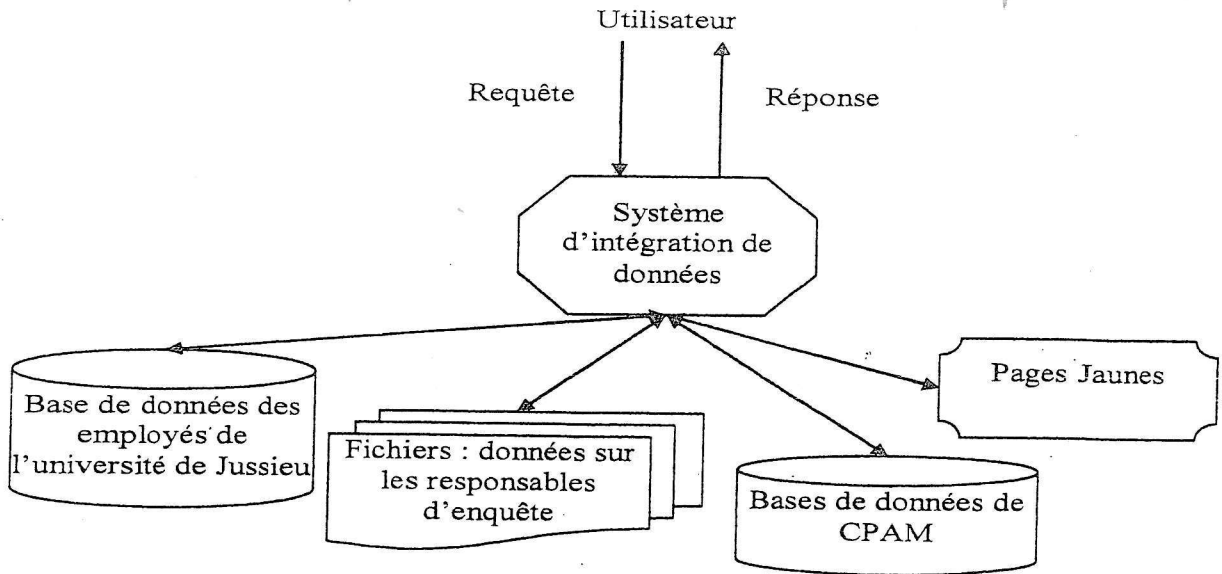


Figure 1 : Exemple d'un système d'intégration de données

Question : Quels sont les numéros de téléphone des médecins traitant des employés de la section dont Durant est le responsable ?

Informatique

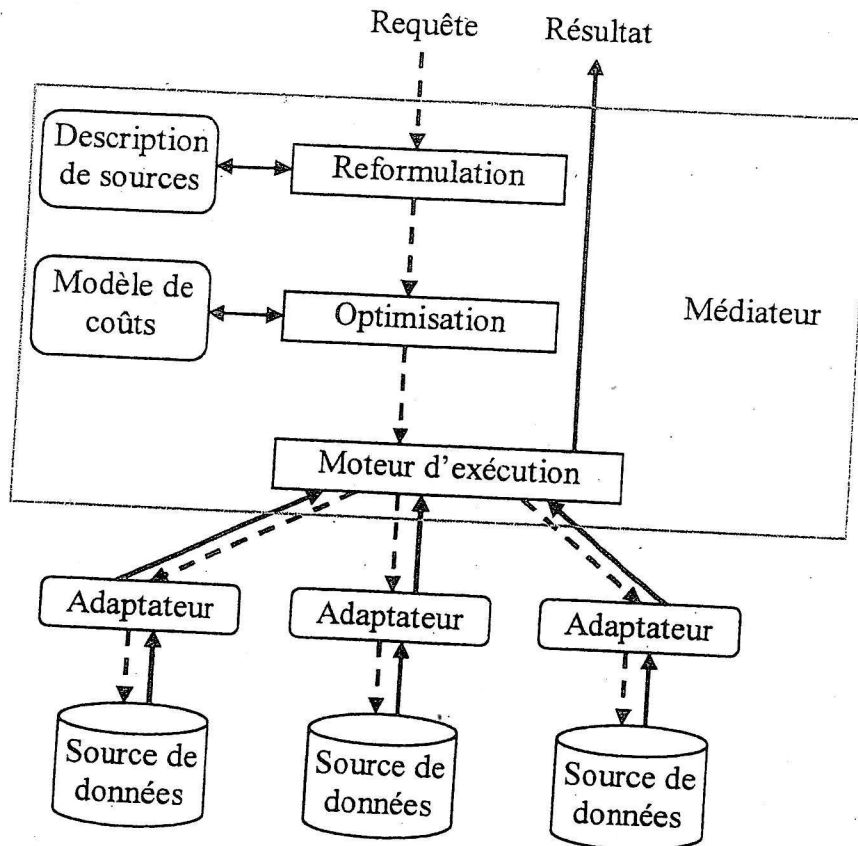
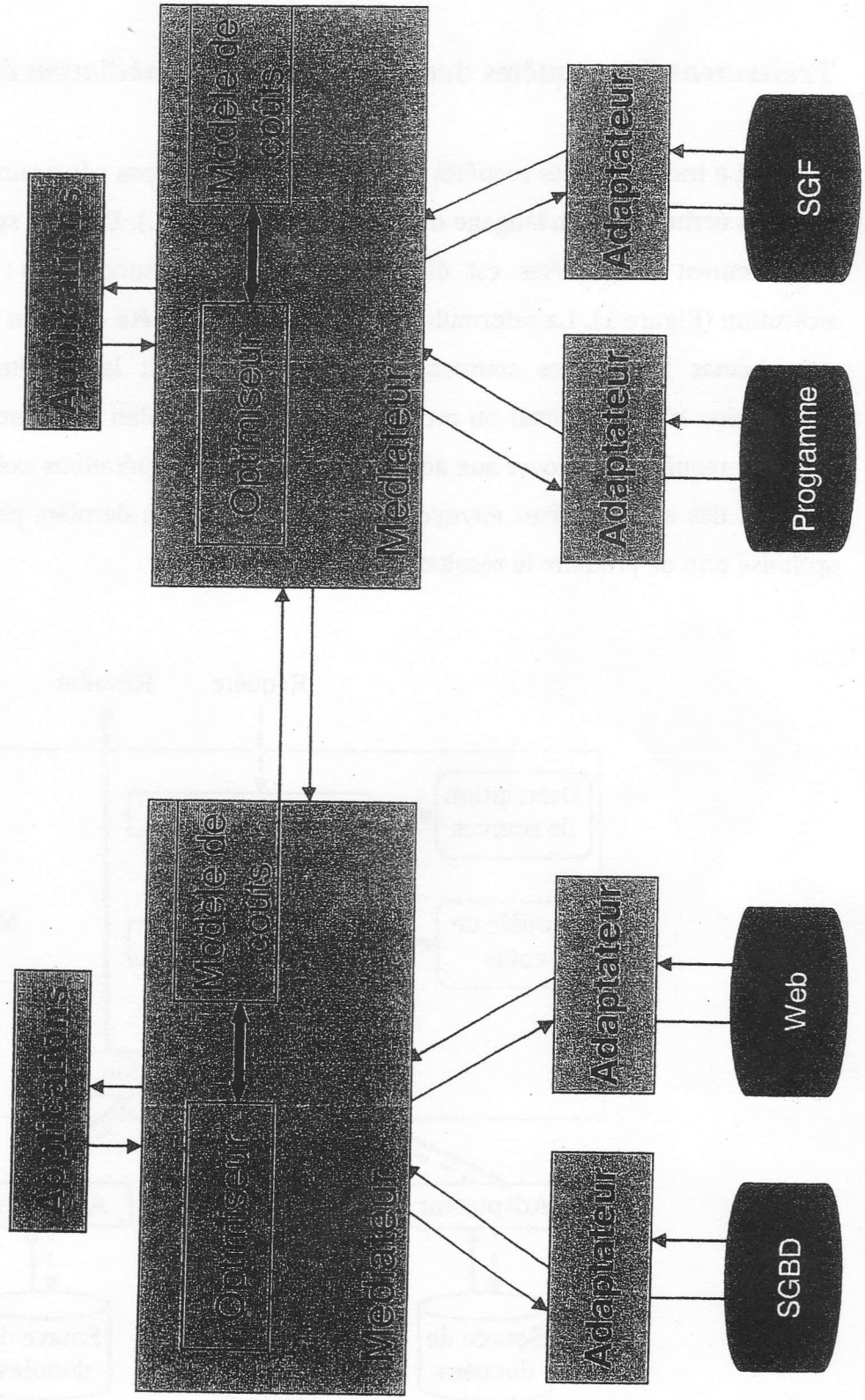


Figure 2 : Etapes de traitement de requêtes dans un système de médiation de données

Contexte : Intégration de données (2)



- Approche médiateurs-adaptateurs



Traitement des requêtes dans les systèmes de médiation de données [Hus 05]

Le traitement des requêtes regroupe toutes les étapes nécessaires pour calculer les résultats des requêtes écrites dans un langage d'interrogation (e.g. SQL). Dans un système de médiation de données, le traitement de requêtes est divisé en trois phases successives : reformulation, optimisation et exécution (Figure 1). La reformulation transforme la requête écrite en schéma global, en requête écrite en schémas locaux des sources. L'optimisation traduit la requête ainsi reformulée en un plan d'exécution réparti optimal ou proche de l'optimal. Ce plan d'exécution est constitué d'un ensemble des sous-requêtes à envoyer aux adaptateurs et quelques opérations exécutées par les médiateurs sur les résultats des sous-requêtes envoyées aux adaptateurs. La dernière phase consiste à exécuter le plan optimisé afin de produire le résultat de la requête.

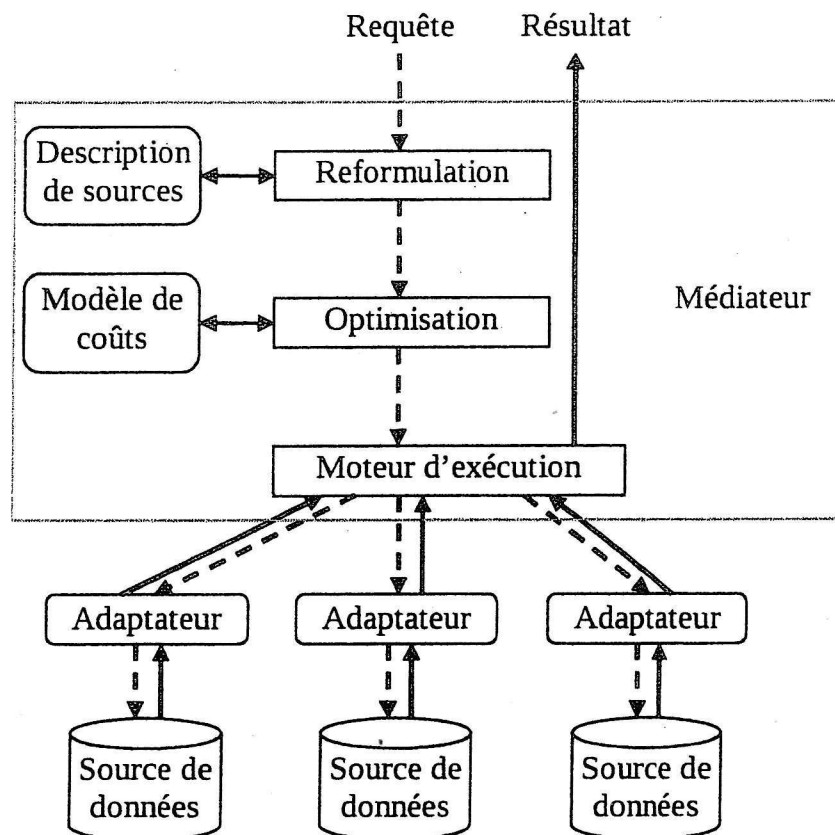


Figure 1 : Etapes de traitement de requêtes dans un système de médiation de données

Processus de reformulation

Les utilisateurs interrogeant les systèmes de médiation de données reformulent leurs requêtes en termes de schémas globaux des médiateurs. Ces requêtes invoquent diverses sources de données qui ont leurs propres schémas locaux. Ainsi, pour calculer le résultat d'une requête, celle-ci doit être reformulée en sous-requêtes écrites en termes de schémas locaux des sources. Par conséquent, deux étapes de reformulation sont distinguées : (i) du schéma global aux schémas exportés par les adaptateurs, et (ii) des schémas exportés par les adaptateurs aux schémas locaux des sources. Ces deux étapes sont effectuées respectivement par le médiateur et les adaptateurs. Dans la suite, nous nous concentrerons sur la reformulation effectuée par les médiateurs. Bien évidemment, les résultats d'une requête reformulée doivent être exactement les mêmes que les résultats de la requête soumise et ils doivent contenir tous les résultats qui peuvent être extraits à partir des sources impliquées. Aussi, la reformulation doit être efficace, c'est-à-dire, éviter l'accès aux sources qui ne contribuent pas à l'élaboration des réponses des requêtes. Pour assurer la reformulation des requêtes, les systèmes de médiation se basent sur un module qui exprime les liens et les associations entre le schéma global et les schémas locaux exportés, ce module est appelé *description de sources*. Il spécifie les contenus des sources et il décrit la capacité de traitement des sources. Dans le cas où le schéma global est construit en s'appuyant sur l'approche GAV, la reformulation est très simple. Il s'agit de remplacer les vues par leur définition. Cependant, dans le cas de LAV, la reformulation est très complexe. Ainsi, il faut réécrire la requête en termes de vues [ABI 98, LEV 00a].

Schémas d'intégration

- Problèmes

- *Intégration de schéma*: comment définir un schéma global d'intégration à partir des schémas des sources?
- *Fusion de données*: comment rendre compatibles, transformer les données en provenance des sources?
- *Mappings/vue d'intégration*: comment décrire le lien entre le schéma global et les schémas des sources?

Page 19

Définition de la vue d'intégration

- Le lien entre schéma global et schémas locaux est défini à travers des vues
 - Mapping entre ces schémas
- Deux façons principales de définir ce lien
 - Le schéma global en fonction des schémas locaux « *global as view* »
 - Approche *ascendante*: on part des sources pour produire le schéma global
 - Les schémas locaux en fonction du schéma global « *local as view* »
 - Approche *descendante*: on fixe le schéma global et on décrit les sources par rapport à ce schéma fixé

Prof. Dan Voelislav (Intégration de données)
Univ. Cergy - Pontoise

5

Page 20

« Global-as-View »

- Le modèle global = vue sur les sources
 - élément global = f(éléments des sources)
$$M = V(S_1, \dots, S_n)$$
- Avantages
 - approche naturelle
 - la traduction de requêtes se fait facilement
- Inconvénients
 - nouvelle source → modification du modèle global
 - il faut considérer l'interaction de la nouvelle source avec les autres

Page 21

« Local-as-View »

- Les sources = vues matérialisées du modèle global
 - une source décrit les données du modèle global qu'elle peut fournir
 - élément source = f(éléments modèle global)
$$S_i \subseteq V_i(M)$$
- Avantages
 - les sources sont décrites indépendamment les unes des autres
 - très simple de rajouter une nouvelle source
- Inconvénients
 - traduction de requêtes plus complexe

Prof. Dan Vodusek (Intégration de données)
Univ. Cergy-Pontoise

⑥

Page 24

Systèmes d'intégration de données

Voici quelques références concernant les systèmes de médiation/d'intégration de données.
 Merci pour la diffusion, si possible, à toute la classe.
 Dans le cadre contraire, merci à Madame Karine Blanchet d'assurer la diffusion.

1. Medience Server *

http://www.oseo.fr/a_la_une/paroles_d_entrepreneurs/autres_temoignages/inria_medience
 + autres

2. Information Integrator (IBM)

http://www.theregister.co.uk/2003/02/05/what_the_hell_is_ibm/
http://www.toolinux.com/news/produits/db2_information_integrator_est_ne_ar3418.html
 + autres

3. OLE-DB.NET (MS)

<http://www.google.fr/search?hl=fr&client=firefox-a&rls=org.mozilla:en-US:official&hs=rF6&q=ole-db.net&start=10&sa=N>
<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=/com.ibm.db2.u>
 + autres

Créée en 2003 par des chercheurs de l'INRIA, Medience a développé une solution technologique innovante de migration et d'intégration de données en temps réel. L'offre de Medience repose sur une technologie d'EII (Entreprise Information Integration), construite autour du concept de virtualisation de données et d'un puissant moteur de requêtes en temps réel sur des sources de données multiples et hétérogènes.

En novembre 2005, Business Objects rachète la société Medience. Elle acquiert, de ce fait, une technologie innovante de migration et d'intégration de données en temps réel basée sur le concept d'EII (Entreprise Information Integration), qui permet de simplifier l'architecture des plates-formes décisionnelles.

Composants logiciels pour l'Intégration de données d'Entreprise

Cette nouvelle technologie va permettre à Business Objects d'étendre son offre d'intégration de données en permettant, notamment, de répondre aux besoins des applications de business intelligence opérationnelle, pour lesquelles l'accès en temps réel à des données intégrées provenant de sources de données disparates devient indispensable.

Intégrant des fonctions de nettoyage de données et un moteur de requête particulièrement puissant, les produits de la société Medience ont bénéficié de plus de 10 ans de recherche effectuée au sein de l'INRIA.

Labellisée "Born of INRIA", Medience s'est développée avec le soutien d'INRIA-Transfert /via/ le Processus IT : processus d'accompagnement et d'évaluation, à l'issue duquel le label Pertinence IT® lui a été attribué en juin 2005.

La société Medience a été lauréate du 5 Concours national d'aide à la création d'entreprises de technologies innovantes (ministère de la Recherche en partenariat avec OSEO), dans la catégorie « Création d'entreprise ».

Prototypes :

Lesselect (2001), DISCO (96), TSIMMIS (97)

GARLIC (99), Infomaster 97 et Manfold 96