

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**Institut National de formation en Informatique
(INI)
Oued Smar – Alger**

Mémoire de Fin D'Etudes

Pour l'obtention du diplôme d'Ingénieur d'Etat en Informatique

Option : Systèmes d'information

Thème :

Intégration de bases de données relationnelles hétérogènes.

Réalisé par :

Bilal AMROUCHE
Ferhat MAKHLOUFI

Encadré par :

Mr SALHI Nadir
M^{elle} BENNA Amel

Organisme d'accueil

Centre de Recherche sur l'Information Scientifique et Technique (CERIST)

Promotion : 2006/2007

Résumé

Aujourd'hui, les entreprises sont constituées de plusieurs systèmes d'information autonomes, évoluant dans des environnements hétérogènes et distribués. Dans ces nouveaux environnements l'information est devenue une ressource importante et stratégique. En particulier le partage des informations informationnelles devient primordial pour des raisons de productivité, de compétitivité et d'efficacité.

Pour que ces entreprises deviennent performantes, elles doivent être capables de gérer et d'intégrer toutes les sources informationnelles hétérogènes, avec la prise en compte de l'aspect sémantique. L'intégration permet la création des systèmes globaux de partage d'informations. Ces systèmes de partage d'informations regroupent les schémas des différentes bases de données, afin de produire un schéma global de toutes les ressources informationnelles disponibles.

L'objectif de ce travail est de faire un état de l'art exhaustif sur l'intégration de bases de données relationnelles hétérogènes et sur les différentes approches qui sont utilisées pour solutionner le problème d'intégration. Enfin choisir une approche parmi celles existantes et implémenter un prototype.

Mots clés : Bases de données, Ontologie, Intégration, Médiateur, Adaptateur, Réécriture de requête.

ABSTRACT

Today, the companies consist of several autonomous information systems, evolving in heterogeneous and distributed environments. In these new environments information became an important and strategic resource. In particular the sharing of informational information becomes essential for reasons of productivity, competitiveness and effectiveness.

So that these companies become powerful, they must be able to manage and integrate all the heterogeneous informational sources, with the taking into account of the semantic aspect. Integration allows the creation of the total systems of division of information. These systems of division of information gather the diagrams of the various data bases, in order to produce a total diagram of all the informational resources available.

The objective of this work is to make an exhaustive state of the art on the integration of heterogeneous relational data bases and on the various approaches which are used to solve the problem of integration. Finally to choose an approach among those existing and to implement a prototype.

Key words: Data bases, Ontology, Integration, Mediator, Wrapper, Rewriting of request.

TABLE DES MATIERES

<i>INTRODUCTION GENERALE</i> -----	1
<i>PROBLEMATIQUE</i> -----	2
<i>STRUCTURE DU DOCUMENT</i> -----	3

PARTIE I: Etat de l'art

1. INTRODUCTION -----	4
2. BASES DE DONNES REPARTIES -----	5
2.1 Définition -----	5
2.2 Objectifs de la répartition des données -----	6
2.3 Avantages de la distribution des données par rapport à l'approche centralisée -----	7
2.4 Principe de base d'une BDR -----	7
2.5 Conception de BD Réparties -----	8
2.5.1 Approche descendante -----	8
2.5.2 Approche ascendante -----	8
3. SYSTEMES D'INTEGRATION DE SOURCES DE DONNEES HETEROGENES --	10
3.1 Composition d'un système intégré -----	10
3.2 Caractéristiques des sources de données -----	11
3.2.1 L'autonomie -----	11
3.2.2 L'hétérogénéité -----	12
3.2.3 La distribution -----	12
3.2.4 Interopérabilité -----	12
3.3 Conflits existants dans les sources hétérogènes -----	13
3.3.1 Les conflits syntaxiques -----	13
3.3.2 Les conflits logiques de données -----	14
4. APPROCHES D'INTEGRATION DE BASES DE DONNEES HETEROGENES -----	16
4.1 Approche multi-base -----	16
4.2 Approche par fédération -----	17
4.2.1 Processus d'intégration dans les BDF -----	17
4.3 Approche entrepôt de données -----	19
4.3.1 Définition -----	19
4.3.2 Concepts de base d'un entrepôt de données -----	20
4.3.2.1 Les données détaillées -----	20

4.3.2.2	Les données agrégées -----	20
4.3.2.3	Les données historiées -----	20
4.3.2.4	Les méta-données -----	20
4.3.3	Les étapes de construction d'un entrepôt de données -----	21
4.3.3.1	Extraction-----	21
4.3.3.2	Le stockage-----	22
4.3.3.3	L'accès-----	22
4.3.4	Exploitation et extraction de données-----	22
4.4	L'approche médiateur -----	23
4.4.1	Classification selon la résolution des conflits de données -----	24
4.4.1.2	La médiation de schéma-----	24
4.4.1.2	La médiation de contexte-----	25
4.4.2	Classification selon la localisation du médiateur -----	25
4.4.2.1	Les médiateurs centralisés -----	26
4.4.2.2	Les médiateurs décentralisés -----	29
4.5	Comparaison des différentes approches -----	31
4.6	Analyse générale des approches d'intégration -----	32
5.	<i>APPROCHES DE MISES EN CORRESPONDANCE DANS LES MEDIEATEURS</i>	
	<i>CENTRALISES -----</i>	33
5.1	Approche GAV -----	33
5.2	Approche LAV -----	35
5.3	Comparaison entre les approches GAV et LAV -----	37
6.	<i>LES ONTOLOGIES DANS LES SYSTEMES MEDIEATEUR -----</i>	39
6.1	Définition -----	39
6.2	Composantes d'une ontologie -----	39
6.3	Types d'ontologies -----	40
6.4	Représentation d'une ontologie -----	40
6.5	Processus de construction d'une ontologie -----	41
6.6	Rôles de l'ontologie dans les systèmes de médiation -----	42
6.7	Approches de mise en correspondances basées sur l'ontologie -----	42
6.7.1	Approche avec une ontologie -----	42
6.7.2	Approche avec plusieurs ontologies -----	43
6.7.3	Approche hybride -----	43
6.7.4	Comparaison des trois approches -----	44
7.	<i>TRAITEMENT DE REQUETES DANS UN SYSTEME DE MEDIATION -----</i>	45
7.1	Analyse de la requête -----	45
7.2	Décomposition et réécriture de la requête -----	45
7.3	Exécution des sous-requêtes -----	45
7.4	Recomposition des résultats -----	45
	<i>CONCLUSION -----</i>	46

PARTIE II : Conception

1. INTRODUCTION	47
2. DESCRIPTION DE LA SOLUTION PROPOSEE	48
2.1 Niveau utilisateur	49
2.2 Niveau médiateur	50
2.2.1 Ontologie de domaine partagée	50
2.2.2 Base de description des schémas	50
2.2.3 Module de traitement de la requête	50
2.3.1 Les adaptateurs	51
2.3 Niveau sources de données	51
3. MISE EN CORRESPONDANCE ONTOLOGIE-SCHEMA SOURCE	51
3.1 Base de description des schémas	53
3.1.1 Liste des tables	53
3.1.2 Liste des attributs	54
3.1.3 Liste des associations	55
4. TRAITEMENT DE LA REQUETE GLOBALE	56
4.1 Analyse de la requête globale	57
4.2 Localisation des sources pertinentes	57
4.3 Décomposition et réécriture de la requête globale en sous requêtes locales	59
4.3.1 Algorithme de décomposition	68
4.3.2 Algorithme de réécriture	70
4.4 Recomposition du résultat	71
4.4.1 Algorithme de recomposition	74
4.5 Algorithme général de la solution	74
5. MISE À JOUR DANS LE SYSTEME DE MEDIATION	75
5.1 Ajout d'une source	75
5.2 Suppression d'une source	75
5.3 Modification d'une source	75
CONCLUSION	76

PARTIE III : Mise en œuvre

1. INTRODUCTION	77
------------------------	-----------

2. EXPERIMENTATION	78
2.1 L'ontologie	78
2.2 Base de description des schémas	79
2.3 Les sources de données	80
3. ENVIRONNEMENT DE DEVELOPPEMENT	80
3.1 Le langage JAVA	80
3.2 Les JSP (Java Server Pages)	81
3.3 L'API JENA	81
3.4 L'EDI Eclipse 3.2	81
3.5 Le serveur d'applications Apache Tomcat 5.5.20	81
3.6 Le serveur de déploiement Axis2	81
3.7 Les SGBD des sources de données	82
4. ARCHITECTURE DETAILLEE DU SYSTEME	82
4.1 Composants du système	83
4.2 Opérations effectuées par le système	84
4.2.1 Traitement d'une requête de médiation	84
4.2.2 Opération de mise à jour	85
5. IMPLEMENTATION DES COMPOSANTS DU SYSTEME	85
5.1 L'ontologie	85
5.2 La base de description des schémas	86
5.3 Le module de traitement de requêtes	86
5.4 Les adaptateurs	87
6. INTERFACES	89
CONCLUSION	93
<u>CONCLUSION ET PERSPECTIVES</u>	
CONCLUSION GENERALE	95
PERSPECTIVES	95
BIBLIOGRAPHIE	96
Annexe A : Les services web	100
Annexe B : Quelques systèmes d'intégration	107