

УДК 681.326
ББК 32.973
Б 92

*Печатается по решению редакционно-издательского совета
Южного федерального университета*

Рецензенты:

Заместитель директора по инновациям МИЭМ НИУ ВШЭ доктор
технических наук, профессор, академик Академии информатизации
образования **Кулагин В.П.**

Заведующий кафедрой информационно-технических систем
безопасности ЮФУ, доктор технических наук, профессор **Целых А. Н.**

Б 92 **Букатов А.А., Пыхалов А.В.**

Методы и средства интеграции независимых баз данных в
распределенных телекоммуникационных сетях: монография / А.А.
Букатов, А.В. Пыхалов; Южный федеральный университет. – Ростов-
на-Дону: Издательство Южного федерального университета, 2013.
– 160 с.

ISBN 978-5-9275-1189-1

Рассматривается задача интеграции изначально разрозненных баз данных (БД), в которых хранятся данные о некоторой общей для этих БД предметной области, а также методы и средства решения этой задачи. На базе анализа степени и качества решения указанной задачи, как известными авторам специализированными системами интеграции (СИД) данных, так и средствами интеграции данных СУБД Oracle, выявлены основные недостатки указанных решений. В монографии предложены методы построения систем интеграции данных, в значительной степени снижающие отрицательный эффект указанных недостатков, а также привносящие в такие системы дополнительные полезные качества. Рассматривается реализация предложенных методов в разработанной авторами СИД DISGO и проводятся демонстрирующие преимущества DISGO результаты экспериментального сравнения этой системы с решениями, основанными на применении других СИД. Предложенная система может применяться для интеграции данных независимых территориально удаленных БД в распределенных телекоммуникационных сетях крупных корпораций и в сетях межкорпоративного взаимодействия.

Монография адресована специалистам в области построения баз данных корпоративных и межкорпоративных информационных систем, а также студентам и аспирантам соответствующих специальностей.

Ключевые слова: базы данных (БД), интеграция данных независимых распределенных БД, локальная и глобальная схемы данных, язык Datalog, оптимизация запросов к БД, неполный ответ на запрос, недоступность части БД.

Публикуется в авторской редакции.

ISBN 978-5-9275-1189-1

УДК 681.326
ББК 32.973

© Букатов А.А., Пыхалов В.П., 2013
© Южный федеральный университет, 2013

СОДЕРЖАНИЕ

	ВВЕДЕНИЕ	6
1.	ЗАДАЧА ИНТЕГРАЦИИ ДАННЫХ И ОБЗОР ИЗВЕСТНЫХ МЕТОДОВ ИНТЕГРАЦИИ ДАННЫХ	10
1.1.	Различные подходы к интеграции данных: GAV, LAV, GLAV	12
1.2.	Модели данных и языки запросов, используемые в области интеграции данных	14
1.3.	Методы обработки и оптимизации запросов в СИД	18
1.3.1.	Методы оптимизации запросов в реляционных СУБД	18
1.3.2.	Методы обработки и оптимизации запросов в распределенных СУБД	21
1.3.3.	Методы борьбы с устаревшей статистикой в СИД	27
1.3.4.	Методы обработки запросов в Oracle Heterogeneous Services	29
1.3.5.	Методы обработки и оптимизации запросов в СИД SIMS	31
1.3.6.	Методы обработки запросов в СИД TSIMMIS	33
1.3.7.	Методы обработки и оптимизации запросов в СИД Information Manifold	34
1.3.8.	Методы обработки запросов в P2P СИД	36
1.4.	Методы работы с неполными и противоречивыми данными	40
1.4.1.	Формальная модель для интеграции данных Multiplex	41
1.4.2.	Методы разрешения противоречий в СИД Fusionplex	48
1.5.	Выводы по результатам анализа существующих СИД и постановка задачи на разработку усовершенствованной СИД	50
2.	МЕТОДЫ ОБРАБОТКИ И ОПТИМИЗАЦИИ ЗАПРОСОВ В РАСПРЕДЕЛЕННОЙ СЕТИ	54
2.1.	Краткое описание предлагаемых методов и реализующих их средств	54

2.2	Метод определения источников данных, используемых в запросе к распределенной совокупности источников данных	57
2.2.1.	Используемая модель данных	57
2.2.2.	Подход к построению отображений между глобальной схемой и локальными схемами	62
2.2.3.	Пример описания отображения между схемами данных источников данных при использовании различных подходов к описанию отображений	66
2.3.	Методы обработки и оптимизации запросов	68
2.3.1.	Общий алгоритм выполнения запросов	69
2.3.2.	Метод непосредственного выполнения запросов	70
2.3.3.	Оптимизированный метод выполнения запросов	73
2.4.	Резюме по разработанным методам обработки и оптимизации запросов в распределенной сети	82
3.	АЛГОРИТМЫ ОБРАБОТКИ ЗАПРОСОВ В СИСТЕМЕ ИНТЕГРАЦИИ ДАННЫХ, ПРЕДНАЗНАЧЕННОЙ ДЛЯ РАБОТЫ В РАСПРЕДЕЛЕННОЙ СЕТИ	84
3.1.	Построение графа взаимосвязанности выражений	84
3.1.1.	Алгоритм унификации предикатов в СИД DISGO	89
3.2.	Алгоритмы генерации выражений РА	91
3.2.1.	Алгоритм генерации выражений РА для не рекурсивных программ	91
3.2.2.	Алгоритм генерации выражений РА для рекурсивных программ	95
3.3.	Алгоритм генерации SQL по выражениям РА	103
3.4.	Алгоритмы оптимизации запросов	106
3.4.1.	Алгоритм оптимизации запросов на основе правил	107
3.4.2.	Алгоритмы сбора и обработки статистики	110
3.5.	Корректность предложенных алгоритмов	116
3.6.	Резюме по разработанным методам выполнения запросов к СИД	117
4.	РЕАЛИЗАЦИЯ МЕТОДОВ И СРЕДСТВ ИНТЕГРАЦИИ ДАННЫХ В РАСПРЕДЕЛЕННОЙ	119

СЕТИ	
4.1. Общее описание СИД DISGO	119
4.1.1. Архитектура СИД DISGO	119
4.1.2. Схема взаимодействия прикладных программ с СИД DISGO	123
4.2. Экспериментальный анализ производительности работы СИД DISGO, реализующей предложенные алгоритмы и методы	125
4.2.1. Анализ производительности СИД DISGO в локальной сети	125
4.2.2. Анализ производительности СИД DISGO в распределенной сети	137
4.2.3. Результаты анализа работы средств интеграции данных DISGO и Oracle	146
ЗАКЛЮЧЕНИЕ	147
ЛИТЕРАТУРА	150
ПРИЛОЖЕНИЕ	156