

УДК 004.65

## МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕННЫХ ХРАНИЛИЩ ДАННЫХ С ПОМОЩЬЮ ОБЪЕКТНО- ОРИЕНТИРОВАННОГО ПОДХОДА

*Николаев С.В., Петров А.В.*

*Донецкий национальный технический университет*

Не секрет, что Украина, Россия и другие страны СНГ, обладая большой территорией, просто обречены на создание и развитие информационных систем на основе распределенных хранилищ данных (РХД). Практически все преуспевающие средние и крупные компании имеют удаленные филиалы. Возникает необходимость во взаимодействии между ними и главным офисом компании, поэтому при управлении такими филиалами и компанией в целом не обойтись без сложных корпоративных распределенных информационных систем.[1,4]

Распределенные хранилища данных создаются для обеспечения пользователям (аналитикам, менеджерам и т.п.) быстрого доступа к распределенным данным. Исходя из этого, оптимальным будем считать такое РХД, в котором время выполнения пользовательских запросов на выборку является минимальным.[2]

В качестве критерия эффективности РХД будем использовать суммарное время выполнения пользовательских запросов на выборку данных.

Время выполнения одного запроса складывается из затрат времени на выборку данных из таблицы фактов, из материализованных представлений и временных затрат на передачу данных.

Суммарное время выполнения запросов пользователей  $\bar{T}_{QS}$  в распределенном хранилище данных рассчитывается как сумма времен их выполнения:

$$T_{QS} = \sum_{i=1}^{QCM} T_{QS_i} \quad (1)$$

где  $QCM$  – общее количество пользовательских запросов на выборку возникших в ходе моделирования;  $T_{QS_i}$  – время выполнения  $i$ -го запроса на выборку, которое определяется выражением:

$$T_{QS_i} = T_{s_i} + T_{tr_i} + \max \left\{ T_{QS_{ij}} \mid j = \overline{1, QCS_i} \right\}, \forall i = \overline{1, QCM} \quad (2)$$

где  $T_{s_i}$  – время выборки при выполнении  $i$ -го запроса;  $T_{tr_i}$  – время передачи данных при выполнении  $i$ -го запроса;  $T_{QS_{ij}}$  – время выполнения  $j$ -го подзапроса при выполнении  $i$ -го запроса;  $QCS_i$  – количество подзапросов возникающих при выполнении  $i$ -го запроса на выборку [5].

Для построения модели РХД целесообразно использовать объектно-ориентированный подход. Объектно-ориентированный подход позволяет получать естественную и модифицируемую модель системы [3].

Рассмотрим особенности построения и функционирования объекта исследований – распределенного хранилища данных. РХД представляет собой множество таблиц фактов и измерений, материализованных представлений, которые хранятся на узлах РХД соединенных между собой каналами связи. Взаимодействие пользователей с хранилищем данных осуществляется путем генерации запросов РХД. Таблицы фактов и измерений, а также материализованные представления, как правило, фрагментированы. Кроме того копии одного и того же фрагмента таблицы или материализованного представления могут храниться на разных узлах РХД.

В ходе проведенного анализа в составе РХД были выделены две группы компонентов: компоненты логической архитектуры и компоненты физической архитектуры.

К компонентам логической архитектуры были отнесены все компоненты хранилища данных, связанные с данными и не связанные с технической стороной функционирования хранилища.

В качестве типовых компонентов логической архитектуры были выделены следующие:

- измерение;
- фрагмент таблицы фактов – части таблицы фактов, хранящиеся независимо от остальных частей и основной таблицы;
- таблица измерения – таблица ХД, хранящая редко изменяемые данные, характеризующие и классифицирующие данные из таблицы фактов;
- материализованное представление – заранее вычисленные результаты запросов;
- запросы на вставку, выборку и обновление.

К компонентам физической архитектуры были отнесены те компоненты хранилища данных, которые отвечают за техническую и организационную сторону функционирования РХД. К типовым компонентам физической архитектуры были отнесены следующие:

- сервер;
- канал связи;
- пользователь;
- узел хранилища данных – структурный компонент хранилища данных, объединяющий в себе серверы, фрагменты таблиц фактов и измерений, материализованные представления, входящие и исходящие каналы связи, территориально находящиеся в одном месте, пользователь.

Были разработаны классы, описывающие каждый из типовых компонентов РХД. Для создания моделей применяется унифицированный язык моделирования UML.

В процессе работы РХД пользователи инициируют запросы: запросы на вставку, на обновление и на выборку данных. Каждый запрос может инициировать один или несколько подзапросов. Кроме того, запросы на вставку могут повлечь за собой пересчет материализованных представлений.

Выполнение запросов происходит следующим образом: запрос возникает на одном из узлов РХД, который передает

его одному из своих серверов. В случае если сервер занят, запрос становится в очередь выполнения данного сервера. При необходимости – генерируются подзапросы, которые передаются по каналам связи соответствующим узлам. Ответы подзапросов возвращаются на узел источник запроса.

Был разработан класс «Запрос» (TDWQuery), моделирующий выполнение запроса в РХД. Были определены свойства и функции данного класса. Свойства класса «Запрос» (TDWQuery) приведены в табл. 1.

Класс «Запрос» содержит следующие функции: «Выполнение запроса» и «Запуск подзапроса». Характеристики функций класса

Таблица 1

Структура класса «Запрос» (TDWQuery)

№	Идентификатор	Наименование	Тип
1	ID	Идентификатор	Целое число
2	Name	Наименование	Строка
3	Tables	Список таблиц	Вектор указателей на объекты класса «Фрагмент таблицы фактов»
4	MatViews	Список представлений	Вектор указателей на объекты класса «Материализованное представление»
5	SubQueries	Список подзапросов	Вектор указателей на объекты класса «Запрос»
6	Cost	Сложность	Вещественное число
7	ResultSize	Объем ответа	Целое
8	Dimensions	Список измерений	Вектор указателей на объекты класса «Измерение»

Таблица 2

Характеристики функций класса «Запрос» (TDWQuery)

№	Наименование	Параметры		Тип результата
		Наименование	Тип	
1	Выполнение запроса (Execute)	Момент запуска	Дата/Время	Дата/Время
2	Запуск подзапроса (RunSubQuery)	Момент запуска	Дата/Время	Дата/Время

«Запрос» представлены в табл. 2.

Время выполнения запроса на выборку  $T_{QS}$  определяется выражением:

$$T_{QS} = t_s + t_{tr} + \max \left\{ T_{QS_i} \mid i = \overline{1, QCS} \right\} \quad (3)$$

где  $t_s$  – время выборки данных из основного фрагмента,  $t_{tr}$  – время передачи данных по каналам связи,  $T_{QS_i}$  – время выполнения  $i$ -го подзапроса,  $QCS$  – количество подзапросов.

Общая модель РХД была построена как система взаимодействующих объектных моделей его типовых компонентов. Схема взаимодействия между компонентами представлена на рис. 1.

Был создан программный комплекс моделирования РХД состоящий из подсистем моделирования РХД и анализа результатов моделирования РХД. Был создан графический пользовательский интерфейс с использованием Borland C++ Builder 6.0. позволяющий просматривать и заполнять БД параметров модели, а также управлять процессом моделирования.

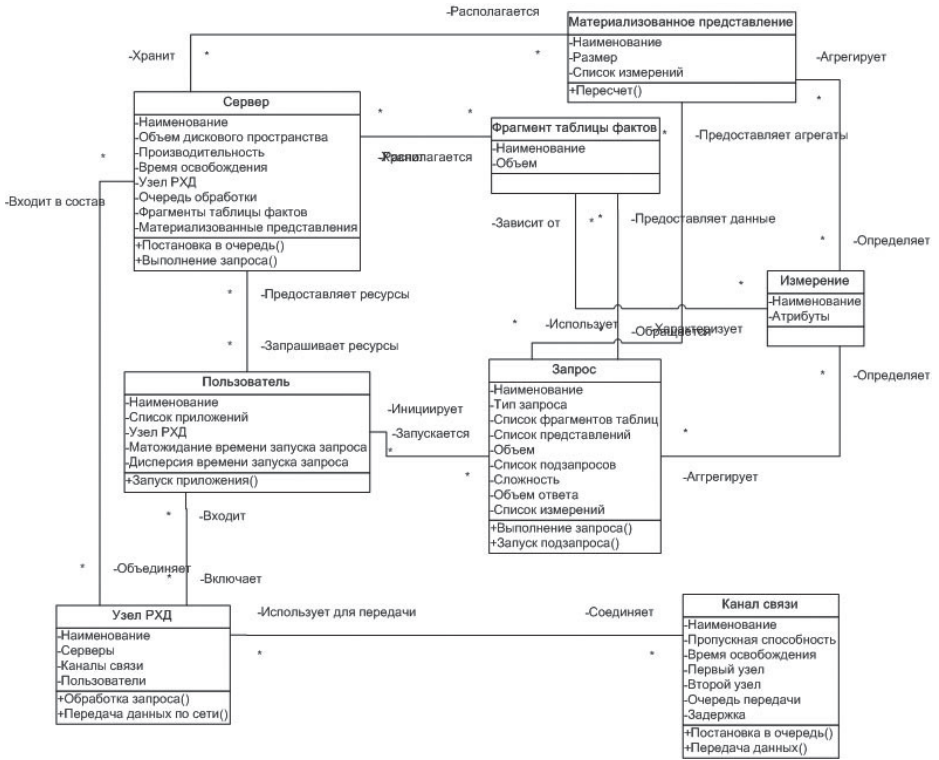


Рисунок 1 – Объектная модель РХД

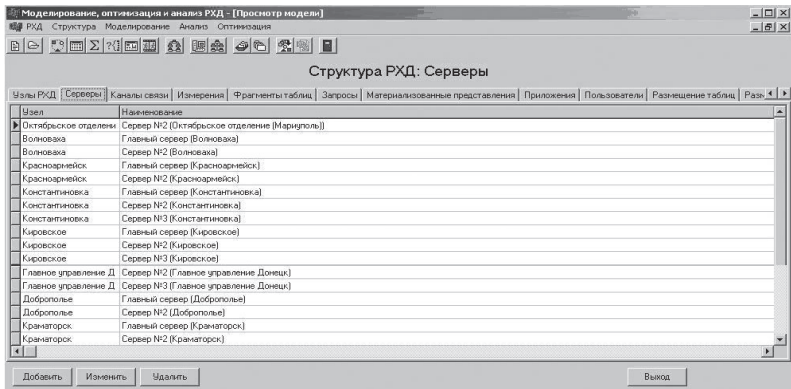


Рисунок 2 – Экранная форма «Моделирование РХД»

## Литература

- [1] Спирли Э. Корпоративные хранилища данных. Планирование, разработка, реализация. Том 1. Пер. с англ. - М.: «Издательский дом Вильямс», 2001 г.
- [2] Лаздинь С.В. Петров А.В. Разработка объектной модели распределенного хранилища данных.//Наукові праці Донецького національного технічного університету. Серія: “Обчислювальна техніка та автоматизація”. Башков Є.О. (голова) та ін. ДонНТУ, 2006 г.
- [3] Буч Г. Объектно-ориентированный анализ и проектирование с примерами приложений на С++. - 2-е изд.: Пер с англ. - М.: “Издательство Бином”, Спб.: “Невский диалект”, 1999 г.
- [4] Роб П., Коронел К. Системы баз данных: проектирование, реализация и управление. – 5-е изд., перераб. и доп.: Пер. с англ. – СПб.: БХВ-Петербург, 2004 г.
- [5] Franconi E., Kamble A. A Data Warehouse Conceptual Data Model.// SSDBM04, 2004 г.