

ИСПОЛЬЗОВАНИЕ НОВЫХ ТЕХНОЛОГИЙ В ИНФОРМАЦИОННЫХ КОРПОРАТИВНЫХ СИСТЕМАХ

САЕНКО В.И., КЛИМЕНКО А.В.,
МОХАММАД АЛЬ РАВАДЖБА

Актуальным вопросом любой информационной корпоративной системы является повышение эффективности управления базами данных, администрирования, повышение гибкости системы и развития сервисных функций. Наилучшим решением является использование новых технологий, развитых в последние несколько лет. К ним относятся Data Mining, Multidimensional Database, Data Analysis (Multidimensional Analysis, On-Line Analysis Process), Intelligent Agent, Push-технология, Hypertext, Intranet.

Появление новых информационных технологий позволяет переосмыслить общие концепции построения информационных систем. Использование этих технологий в существующих информационных сервисных системах (диспетчерских, справочных, поисковых) переводит их на новую качественную ступень, порой не предоставляя видимого дополнительного экономического эффекта от внедрения. Это в первую очередь касается информационных корпоративных систем с уже сложившейся информационной инфраструктурой.

Особенности использования новых технологий. Современные тенденции информационных систем привели к формированию трех типов транспортных систем: локальных, корпоративных, общего пользования. Наиболее интенсивно развиваются корпоративные системы. При этом, с одной стороны, наблюдается переход локальных сетей в корпоративные без потери их индивидуальности (рис. 1), с другой – отдельные части сетей общего пользования становятся также частями корпоративных сетей (серверы, шлюзы (gate), маршрутизаторы (router), отдельные домены (domen), и даже магистрали (backbone)).

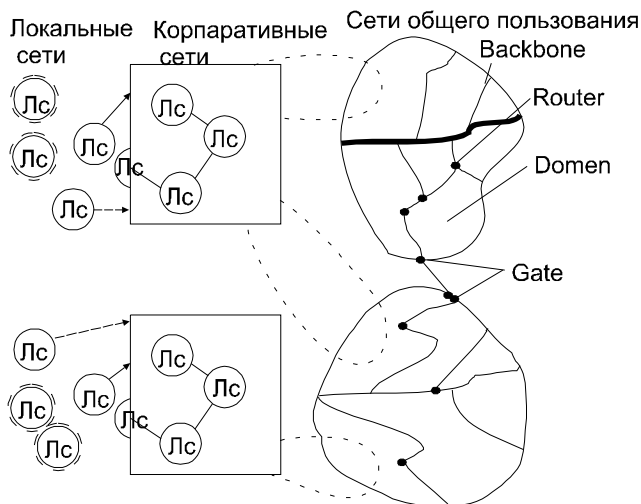


Рис. 1. Переход локальных сетей в корпоративные

Естественность таких тенденций очевидна, так как только корпоративные сети имеют достаточно средств для интенсивного их развития. Такой переход влечет за собой также смещение в тенденциях развития новых технологий, при которых особенно актуальными становятся вопросы защиты информации, администрирования и повышения эффективности системы в целом. Корпоративность ресурсов позволяет наиболее целесообразно распределить эти ресурсы между доменами, пользователями и функциональными процессами, а также осуществить перераспределение функциональных процессов при внедрении новых информационных технологий. В качестве наиболее актуальных для корпоративной системы будем рассматривать проблемы, связанные с повышением эффективности управления базами данных, администрированием, повышением гибкости системы, развитием ее адаптивности и развитием сервисных функций.

Анализ существующих технологий позволил выявить технологии, оказывающие наибольшее влияние на эффективность функционирования информационной системы в целом. К ним относятся технологии Data Mining, Hypertext, технология PUSH и технология intranet.

Технология Data Mining. Повышает эффективность организации баз данных и позволяет провести полный анализ накапливаемой информации. Основным в Data Mining является получение результата анализа корпоративных данных для пользователя. Data Mining представлена в свою очередь тремя технологиями: Multidimensional Database (MDDB), Data Analysis (Multidimensional Analysis (MDA) or On-Line Analysis Process (OLAP)), Intelligence Agent (IA) [1]. Рассмотрим кратко эти технологии.

Multidimensional Database (MDDB). В соответствии с существующими технологиями базы данных могут быть представлены в виде реляционных, суперреляционных, многомерных (логических и физических) и объектных реляционных баз [2].

Реляционные БД (NRDB – normalized relation database – обычная реляционная база данных) – базы данных в виде таблиц (отношений), описываемых схемами с атрибутами и содержащих какое-то число кортежей заданной арности [3]. Операции над базами данных – это операции над компонентами, заданные m -местными отношениями. Это наиболее простая и удобная форма хранения информации, используется в большинстве СУБД.

Суперреляционные базы данных. (SRDB). Это те же базы NRDB, но поддерживающие расширенные форматы данных, индексные схемы и позволяющие хранить формулы (правила) m -местных отношений как компоненты базы данных.

Объектные реляционные базы данных (ORDB). Базы данных, подчиняющиеся правилам NRDB с точки зрения их организации, но оперирующие с нетрадиционными данными – изображениями, видео-, аудио-информацией, допускающие различные операции над этими данными.

Для перечисленных баз данных NRDB, SRDB, ORDB характерна некоторая ограниченность, связанная с тем, что они отражают статические или локальные характеристики описываемого объекта

и, как правило, имеют высокую разреженность (sparse data). Это сказывается на неэффективном использовании компьютерных ресурсов и затруднении в проведении многомерного анализа. Разрешить эти проблемы помогает технология MDDB.

MDDB имеет различные представления и способы организации. Можно выделить четыре: многомерные гиперкубы (HCDB), плотные базы с радиальной структурой (DRDB-DIS - dense relation database with Data in Star), многомерные массивы (MWA-multiway array), базы данных типа хранилища данных (DWDB - Data Warehouse database). Рассмотрим каждую из них.

HCDB основана на представлении сложных информационных объектов в виде абстрактной многомерной структуры (гиперкуба), орты (размерности) которого образованы атрибутами NRDB. Элементы этой базы - фиктивные переменные, соответствующие коду координат (рис. 2).

В этом случае каждая переменная - значение вида $k_1k_2k_3k_4k_5$, при этом хранится ключ разрядности каждой переменной, соответствующей числу значений, которое она принимает в NRDB. Значение 000_000 означает отсутствие этой пере-

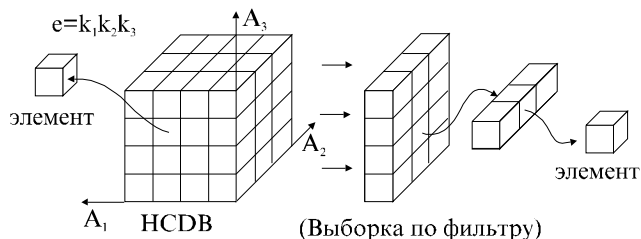


Рис. 2. Структура многомерной базы типа гиперкуб

менной. Запросы формируются в терминах значений координат гиперкуба. База данных сильно разреженная, физического представления база не имеет, существует только в логическом виде. Хранение осуществляется в плотно упакованном виде. Назначение - обеспечение целостности данных при построении распределенных систем.

DRDB-DIS основана на технологии Data In Star (DIS) [4], когда из NRDB (разреженной) формируются DRDB (dense relation database - плотные реляционные базы данных) меньшей арности. Каждая DRDB представляет собой таблицу NRDB и имеет свое имя. В итоге образуется куб 2, 5D (рис. 3). Запросы формируются в терминах новых атрибутов $\{E_1E_2...E_n\}$. База данных физическая, плотная. Наиболее эффективна в системах, в которых для отдельных (не для всех) компонентов

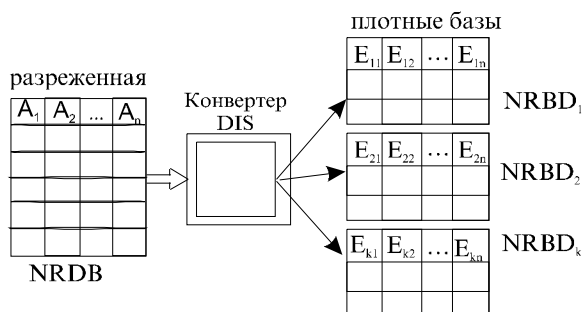


Рис. 3. Структура плотной базы

приводится поясняющая информация, структурированная в виде разветвленного дерева. Формирование запросов по новой размерности бессмысленно. Куб отражает только упакованность базы.

MWA основана на формировании многомерного массива, состоящего из страниц (dimension - размерностей) [5]. Каждая страница имеет свой заголовок, в котором указываются имя, номер страницы (размерности), номер позиции (переменной) (рис. 4).

Переменные на одной странице разного типа и соответствуют компонентам кортежей NRDB. Например, $NAM_1(1, 34) = Alex$, $NAM_1(1, 35) = 67$, $NAM_1(1, 34) = TRUE$. В результате получаются физические базы данных высокой плотности, кото-

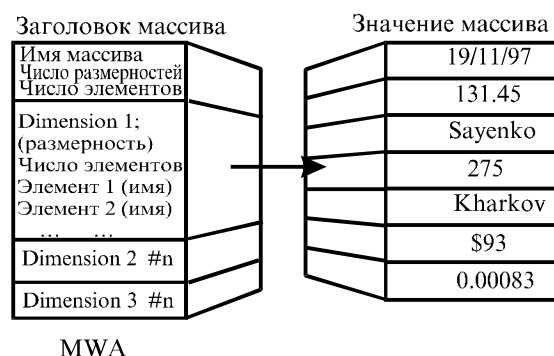


Рис. 4. Структура многомерного массива

рые используются для уплотнения разреженных баз данных высокой арности.

DWDB основана на объединении NRDB в группы, например месяцы, кварталы, регионы, города, которые соответствуют новой размерности. Выборка данных может при этом производиться не только в пределах одной таблицы, но и по новой размерности для выявления тенденций какого-либо рассматриваемого показателя. (рис. 5). База

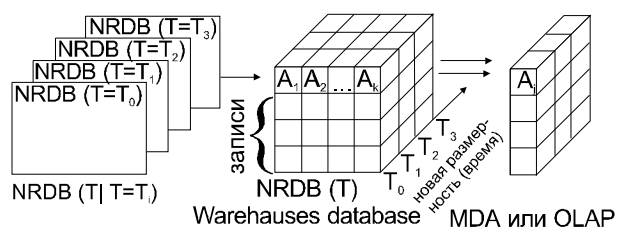


Рис. 5. Структура базы типа DWDB

данных физическая, разреженная. Используется при проведении MDA or OLAP. Эти базы - основной инструмент для систем типа Data Warehouses [2].

Все запросы формируются в системе в терминах MDDB, конвертируемые в запросы физической базы данных: $F: vm - vd$, F - функция, переводящая запросы MDDB в запросы физической базы данных $Bd = Bm(s)$, где s - переменная кортежа; $s = (A_1, A_2, ..., A_k)$; A_i - атрибут базы данных; k - арность.

Выбор той или иной базы данных определяется прежде всего задачами, решаемыми в системе и

выбранными критериями оптимизации. Не существует структуры базы, удовлетворяющей всем критериям (если базы плотные, то имеют либо плохие характеристики надежности и достоверности, либо низкие показатели скорости выполнения запросов; если базы разреженные, то они не экономны в расходовании ресурсов компьютеров). Таким образом, весовые коэффициенты критериев оптимизации в первую очередь определяют выбор той или иной структуры базы.

Использование MDDB позволяет построить более эффективные средства контроля достоверности, полноты и целостности данных, осуществить неравномерное распределение базы между отдельными локальными серверами и сформировать гибкую систему обработки запросов.

Data Analyses (DA) представлен Multidimensional Analysis (MDA) и On-Line Analysis Process (OLAP). – многомерный анализ или анализ процессов в реальном масштабе времени [2]. На сегодняшний день это основной аппарат для описания динамических объектов и состояния объектов, представляемых в виде NRDB. Реализация MDA и особенно OLAP требует высокого быстродействия обрабатывающих станций и станций DBMS (database management system).

Анализ данных делится на четыре уровня:

1) анализ, обеспечивающий реализацию простых запросов и ответов о произошедших событиях, при этом выдаются статические справки;

2) анализ, обеспечивающий выполнение запросов в реальном масштабе времени, дополняющих простые сообщения сведениями о причинах произошедших событий, при этом выдаются динамические формы (диаграммы);

3) анализ по формированию сообщений исполнительных систем, дополнительно оповещающих о том, что необходимо знать в текущий момент времени о системе, фильтруя информацию о ее состоянии по степени важности, при этом выдаются детальные отчеты;

4) анализ по формированию сообщений, дополняющих сведения о состоянии системы, выявляющий влияние побочных факторов, трендов, межбазовых связей, формирующий сообщения о вероятностных переходах состояния системы.

Этот аппарат является основным в технологии Data Warehouses (хранилищ информации) для анализа эффективности реализации маркетинговых программ компаний.

Технология интеллектуальных агентов. IA – это технология резидентных программ, размещаемых на клиентских станциях, составляющих единое целое с некоторой удаленной специализированной системой (RIS remote information system). IA образует независимый виртуальный канал с сервером (если RIS находится на нем) и позволяет осуществить полный мониторинг работы станции клиента. Используя IA, можно также управлять нагрузкой станции на сеть и на сервер, осуществлять сбор информации об активности клиента, вести дополнительное администрирование. IA также обеспечивает контроль целостности распределения MDDB, собирает статистику для реализации OLAP и является обязательным элементом при реализации технологии Data Mining. В насто-

ящее время эта технология является самостоятельной, а не составной частью Data Mining.

Технология PUSH или webcasting. Представляет собой широковещательную рассылку информации по заранее сформированному графику [6]. (Альтернатива этой технологии – технология pull, при которой пользователь сам ищет требуемую информацию в базах данных). Технология PUSH предполагает поддержку программ интеллектуальных агентов, с использованием которых администратор формирует запрос на выборочную информацию для пользователей, а система автоматически поставляет результат запроса (форма, документ или ссылка на документ) либо периодически, либо при очередном изменении состояния данных в наблюдаемой системе. При реализации такой технологии изначально предполагается, что высоких требований на ограничение ресурсов не налагается, потому что эта технология несколько повышает трафик в сетях и загрузку процессоров. Использование технологии PUSH значительно увеличивает возможности системы с точки зрения ее удобства, особенно при реализации корпоративных информационных систем, обеспечивает гарантированную доставку документов всем исполнителям и способствует повышению производительности труда. Однако необходимо учитывать, что обслуживать модули push должен только администратор системы, в противном случае формирование запросов через модули непосредственно самими пользователями приводит к перегрузкам в сети. Это состояние соответствует настоящему времени, так как с появлением специальных анализаторов запросов пользователь сам будет настраивать эти модули.

Технология Hypertext. На сегодняшний день становится одной из основных для информационных систем. Причина – широкие дизайнерские возможности, гибкость в изменении форм, совместимость с большинством информационных систем и с Web-серверами сети Intranet [7]. Технология Hypertext основана на нескольких положениях (документ, приложение, страница), описываемых в терминах HTML и содержащих соответствующие команды. Документ представляет собой гипермедиа документ, т.е. текст, в котором содержатся ссылки на другие документы (текстовые, аудио, видео). Страница – отображаемый экранный фрагмент документа. Приложение – исполняемые модули, которые можно выполнить, находясь в конкретной странице. Преимущества такого представления – независимость от платформы, простота интеграции в любую систему. Использование технологии гипертекста в сочетании с Web-серверами в информационной корпоративной системе приводит к значительному сокращению общего трафика в сети и гибкости всей системы в целом за счет использования браузеров в качестве основного интерфейсного модуля клиента.

Технология Intranet. Как указывалось выше, информационная система может эксплуатироваться в разных условиях транспортной сети: локальной, корпоративной и общего пользования. Технология Intranet используется исключительно для корпоративной сети и представляет собой технологию Internet [8] в замкнутой корпоративной сети (т.е.

технологии Intranet). Эта технология позволяет перенести любую информационную систему из категории ограниченного числа пользователей в категорию массового пользователя, чем обуславливает большую эффективность общего управления организационным объектом с возможностью администрирования распределения ресурсов и доступа к ним. Для пользователей, входящих в сеть Internet, реализовать строгое администрирование практически невозможно, поэтому для них вводятся анонимные имена типа (anonymous (FTP), IUSER_system (WEB)). Intranet-сеть позволяет не только выполнить строгое администрирование (такое же как и для корпоративной сети), но и обязать пользователя следовать правилам строгого регламента работы [8]. Intranet-сеть позволяет также создать дополнительную индексную базу данных, соответствующую ресурсам Web-сервера, т.е. страницам (Web DB). Элемент этой базы самостоятелен и может быть взаимосвязан с любыми другими элементами по уникальному связанному коду, например URL соответствующей страницы.

Несомненное преимущество Intranet сети состоит в представлении системы в виде уже существующих модулей (сетевая операционная система, программы middleware по согласованию всех стеков протоколов, Web-серверы, средства администрирования, браузеры). Такой подход освобождает пользователя от выполнения наиболее трудоемких частей работы по согласованию компонентов и разработке интерфейсных оболочек пользователя, позволяя сосредоточить усилия собственно на организации баз данных, выполнении анализа информации, повышении отказоустойчивости системы и т.п.

Основные выводы. Структурная схема технологий, рассмотренных в работе, представлена на рис. 6.

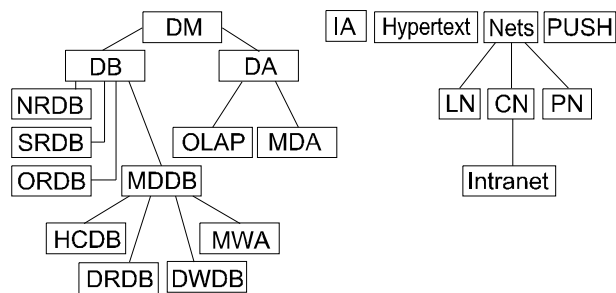


Рис. 6. Структурная схема технологий

Наиболее актуальным вопросом любой информационной корпоративной системы является повышение ее эффективности. Как показано в работе, решение этой проблемы может быть достигнуто путем использования новых информационных технологий: Data Mining, Multidimensional Database, Multidimensional Analysis, On-Line Analysis Process, Intellegent Agent, Push-технология, Hypertext, Intranet. Корпоративная сеть – наилучший полигон для внедрения и развития новых информационных технологий.

Литература: 1. Hedberg S.R. The Data Golden Rush. BYTE, okt., 1995. – P. 83-88. 2. Warehouse Cornerstones. BYTE, jan., 1997. p.85-90. 3. Ульман Дж. Основы систем баз данных. – М.: Финансы и статистика. – 1983. – 334с. 4. Linthicum D.S. Power Tools for Date Drilling. BYTE, jan., 1997. – P.143-144. 5. Dejesus E.X. Dimensions of Data. BYTE, 1995, apr. – P. 139-143. 6. Fontana J. Apps Make Push for Desktop Space. CW.-N20.-1997. – P.19. 7. Пайк М. Интернет в подлиннике. – Санкт-Петербург. – BHV. – 1996. – 640 с. 8. Liebman L. Taming the Intranet. – CW.-N8.-1997. – P. 22.

Поступила в редколлегию 21.12.97

Саенко Владимир Иванович, канд. техн. наук, доцент кафедры ИУС ХТУРЭ. Научные интересы: администрирование, мониторинг и управление процессами в компьютерных сетях. Увлечения: коллекционирование марок. Адрес: 310086, Харьков-86, ул. 23 Августа, 18, кв. 21, тел.40-94-51.

Клименко Александр Васильевич, ст. препод. кафедры ИУС ХТУРЭ. Научные интересы: управление распределенными информационными системами. Увлечения: путешествия. Адрес: Харьков, ул. Архитекторов, 30, кв. 232, тел.40-94-51.

Мохаммад Альраваджба, аспирант кафедры ИУС ХТУРЭ. Научные интересы: управление процессами в компьютерных сетях, сетевые технологии. Увлечения: туризм. Адрес: 310058, Харьков-58, ул. Гвардейцев Широнинцев, 40, кв. 229, тел.40-94-51.