

ЭЛЕКТРОННЫЙ ГОРОД – ЭЛЕКТРОННЫЙ РЕГИОН – ЭЛЕКТРОННОЕ ГОСУДАРСТВО: ВЗАИМОДЕЙСТВИЕ ОБЩЕСТВА, БИЗНЕСА И ВЛАСТИ

Д. В. Горбачев, кандидат технических наук, доцент, декан факультета информационных технологий, ГОУ ВПО «Оренбургский государственный институт менеджмента»
e-mail: gordi47@mail.ru

А. В. Дудко, кандидат педагогических наук, доцент, заведующий кафедрой «Информационные технологии», ГОУ ВПО «Оренбургский государственный институт менеджмента»
e-mail: dedko11111@mail.ru

ПОДХОДЫ К РЕШЕНИЮ ЗАДАЧ ЭЛЕКТРОННОГО ВЗАИМОДЕЙСТВИЯ В СРЕДЕ «ЭЛЕКТРОННЫЙ ГОРОД» НА ОСНОВЕ РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ БАЗ ДАННЫХ

Информатизацию сфер городской жизни представляется целесообразным строить в соответствии принципами обработки распределенных баз данных (БД). Ключевым решением при реализации такого подхода является высокая степень интеграции БД, достичь которую возможно на основе способов обработки федеративных баз. Предлагаемый подход позволяет организовать облачную среду, интегрирующую распределенные системы различных субъектов города.

Ключевые слова: электронный регион, распределенная обработка, федеративные базы данных, облачные вычисления, инфраструктура распределенных систем

Говоря об электронной среде региона, как правило, подразумевается, что эта среда строится на основе информационных порталов и сайтов органов власти, управления и госучреждений (рис. 1).

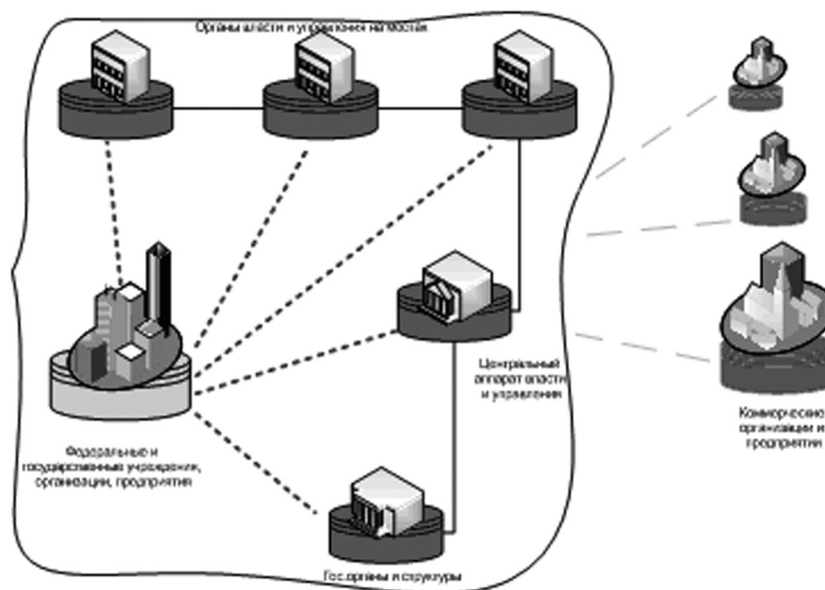


Рис. 1. Упрощенная модель среды «Электронный город»

Действительно Концепция региональной информатизации (одобрена распоряжением Правительства Российской Федерации от 17 июля 2006 г.) в качестве основных целей определяет:

- повышение эффективности управления социально-экономическим развитием субъектов Российской Федерации;
- обеспечение доступа населения и организаций к информации о деятельности органов государственной власти субъектов Российской Федерации и их участия в процессе общественной экспертизы проектов решений в сфере региональной информатизации и эффективности их реализации;
- улучшение качества государственного управления в органах государственной власти субъектов Российской Федерации;
- создание условий для развития информационно-телекоммуникационной инфраструктуры, отвечающей современным требованиям и обеспечивающей потребности населения в информации, а также потребности органов государственной власти субъектов Российской Федерации в информации и информационном взаимодействии;
- обеспечение информационной безопасности региональных и муниципальных информационных систем, информационно-телекоммуникационной инфраструктуры на территории субъектов Российской Федерации.

Далее в Концепции также определяются требования к структуре и содержанию подсистем, реализующих указанные цели.

Однако понятие «Электронный регион» значительно шире. На наш взгляд, «электронный регион» (область, город, поселок) представляет собой некоторую модель, например, города, реализованную в виде распределенных информационных представлений городских субъектов. Традиционными технологиями реализации таких представлений являются Web-технологии и технологии распределенных вычислительных систем. Здесь также можно говорить и о так называемой городской сети, которая может использовать те же коммуникации, что и Web, но без выхода в глобальную сеть. Организация, управление, администрирование, наполнение контента, обслуживание городской сети – в этом случае будут отдельные, скорее всего, достаточно дорогостоящие, процессы, но полученный проект будет более ориентированным на горожан.

Модель «Электронный город» целесообразно рассматривать с позиции горожан – как возможность получать электронные услуги, аналогичные реальным:

- покупка различных товаров;
- развлечения: кино, музыка, путешествия;
- переписка, общение, дружба;
- обучение;
- медицинское обслуживание;

- документационное обеспечение.

Конечно, многие из всех возможных электронных сервисов не заменят реальные в полной мере, однако они могут способствовать их развитию.

Основу перечисленных электронных сервисов составляют информационные базы учреждений, организаций и предприятий, которые условно можно разделить на три группы (рис. 1):

- БД органов власти и управления (являются практически закрытыми);
- БД государственных учреждений и организаций (доступ открыт только к общей информации);
- БД коммерческих предприятий (структуру и содержание можно предположить).

И вот здесь возникают, пожалуй, главные проблемные вопросы:

Как интегрировать все эти БД между собой?

На каком уровне должна проходить эта интеграция?

Как установить уровень и правила доступа к уже интегрированным ресурсам? и т.д.

Сама по себе интеграция разнородных источников данных – фундаментальная проблема, возникшая в последние десятилетия перед сообществом разработчиков БД. Цель интеграции данных состоит в том, чтобы предоставить единый интерфейс к различным источникам и позволить пользователям сосредоточиться на определении того, что они хотят узнать. В результате интеграция должна освободить пользователя от поиска релевантных источников данных, взаимодействия с ними по отдельности, отбора и комбинирования данных из различных источников. Проектирование системы интеграции данных – очень сложная задача. Подходы к ее решению условно делятся на «классические» и онтологические [1].

Классические варианты интеграции предполагают:

- использование *федеративных БД*, которые независимо хранят одну и ту же информацию, периодически синхронизируя свои состояния. Для синхронизации нескольких федеративных БД требуется определить и соответствующее количество связей между БД;

- создание единого *централизованного хранилища* данных. Данные из разнородных источников периодически копируются в хранилище (требуется меньшее, по сравнению с первым случаем количество связей для тех же БД);

- использование технологии создания программных оболочек, или *медиаторов* (mediators, wrappers), обеспечивающих единый интерфейс доступа к различным БД (наиболее эффективный, но и трудоемкий);

Онтологический подход может успешно применяться для решения двух подзадач:

- спецификации содержимого разнородных источников данных в виде онтологии;

- получения ответов на запросы, адресованные интегрирующей системе и основанные на спецификации источников.

Обычно архитектура системы интеграции данных позволяет явно моделировать данные и информационные потребности (т.е. определять те данные, которые система предоставляет пользователю) на различных уровнях.

- **Концептуальный уровень** содержит концептуальное представление источников и согласованных интегрируемых данных вместе с явным декларативным описанием отношений между их компонентами.

- **Логический уровень** содержит представление источников в терминах логической модели данных.

Поиск необходимых сведений в нескольких источниках может быть сопряжен с большими сложностями и значительными затратами времени – в первую очередь из-за того, что в ходе каждой операции поиска запросы пользователей традиционно адресуются лишь одному источнику данных.

Столкнувшись с этой проблемой, некоторые поставщики разработали решения, основанные на технологии так называемого «федеративного» (federative) доступа к нескольким базам данных: Liquid Data компании BEA Systems, программный инструмент IBM DB2 Information Integrator (DB2 II), Microsoft SQL Server 2010, СУБД Oracle 9i. С помощью этих средств пользователи могут одновременно обращаться с запросами сразу к нескольким разнородным источникам данных.

Потребность в технологии федеративного доступа к базам данных проявляется в первую очередь при эксплуатации систем, предназначенных для управления Web-контентом.

Технологии интеграции БД

Традиционно пользователь (например, сотрудник отдела обслуживания клиентов), желающий

получить информацию из нескольких источников, может опрашивать источники лишь в порядке очереди, а это связано с более значительными затратами времени и средств, чем в случае одновременного доступа к множеству источников. В данном случае доступ пользователей к отдельным базам данных осуществляется с помощью программных средств интеграции приложений предприятия (enterprise application integration, EAI) и JDBC.

Информационные системы традиционной архитектуры (рис. 2) в каждый момент обеспечивают доступ лишь к одному источнику данных. Объясняется это тем, что до недавнего времени не существовало ни универсального языка для данных, ни качественных метаданных. Если пользователи желают получить необходимые данные из нескольких источников, они должны располагать точными метаданными, описывающими хранимые данные в форматах, доступных для чтения на различных платформах. Между тем, программные и аппаратные средства не отличались ни достаточной мощностью, ни функциональностью, необходимой для осуществления доступа к распределенным данным.

Раньше разработчики вручную создавали адаптеры для работы с полученными из нескольких источников данными, используя при этом API соответствующих источников. Однако вручную созданные адаптеры не способны автоматически подстраиваться под обновляемые версии API.

Централизованный подход к извлечению данных из различных источников, как правило, предполагает дублирование и последующий сбор скопированных данных в одну базу (или очень небольшое их число). Затем пользователи обращаются к этим агрегированным базам данных, которые иногда называют хранилищами.

Когда данные собраны в одном источнике, пользователи могут быстрее получить требуемую информацию, а система может с большей

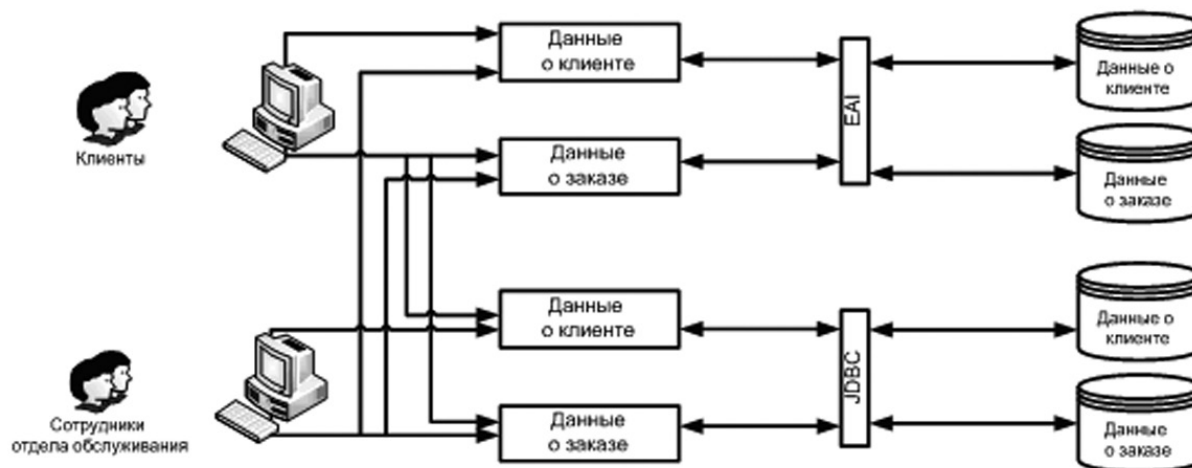


Рис. 2. Интеграция данных на основе традиционной архитектуры

легкостью, чем в случае, если бы весь материал был разбросан по различным системам, нормализовать данные и выполнить другие операции по их обработке. Однако для сбора информации в централизованный источник требуется, чтобы данные, которые часто хранятся в различных форматах, были приведены к одному, а в ходе этого процесса возможно появление ошибок.

Кроме того, хранилищам данных может оказаться сложно взаимодействовать с новыми источниками данных в незнакомых форматах. Наконец, то обстоятельство, что данные нужно дублировать и работать с несколькими их копиями, приводит к повышению издержек.

Федеративный подход (его реализацию средствами Liquid Data иллюстрирует рис. 3) предполагает доступ к данным, находящимся непосредственно в разнородных источниках, и создание единого виртуального хранилища. Разработчики могут писать все свои запросы к федеративной системе данных, играющей роль посредника, роль которого, в сущности, состоит в том, чтобы абстрагироваться от соединений с различными серверными источниками данных [2].

Технология федеративного доступа к базам данных позволяет пользователям (например, сотруднику отдела обслуживания клиентов) одновременно обращаться к данным из нескольких разнородных источников. В соответствии с раз-

работанной компанией BEA Systems технологией Liquid Data архитектор данных формирует стандартизованные представления данных так, чтобы пользователи могли «под одним углом зрения» изучать несхожие материалы. Затем система объединяет данные с тем, чтобы придать материалу логическую стройность; например, все данные о клиенте объединяются в одну категорию, а данные о заказе – в другую. Так пользователи могут с помощью однократных запросов сервера Liquid Data автоматически получать доступ к нескольким разнородным источникам данных.

Однако, как поясняет Фридман из Gartner, процесс сбора данных из различных источников создает дополнительную нагрузку на систему. К тому же обработка «на лету» распределенных запросов, обращенных к различным источникам данных, по его словам, предполагает перемещение по сети значительных объемов данных, что может существенно снизить ее пропускную способность. Наконец, в случае использования федеративного подхода при выполнении нормализации и других операций по обработке данных возникает больше сложностей.

Возможность одновременно направлять запросы нескольким источникам, содержащим разнородные данные, возникла благодаря нескольким технологическим решениям. Так, повышение эффективности оптимизаторов запросов, программ,

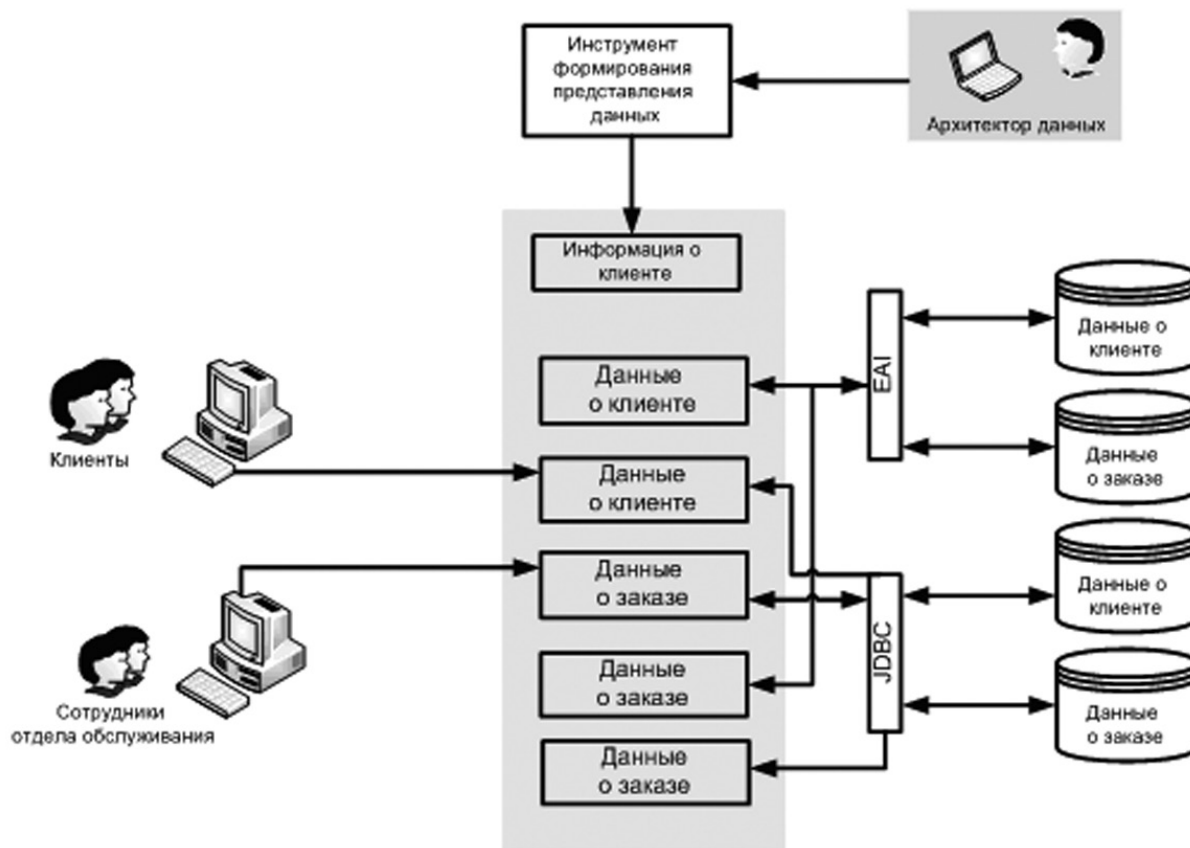


Рис. 3. Интеграция данных на основе федеративного доступа

которые с помощью системы правил осуществляют тонкую подстройку выполнения запросов, стало следствием применения более совершенных алгоритмов. С другой стороны, повысилось быстродействие центральных процессоров и жестких дисков, что привело к росту производительности систем поиска.

Технология федеративного доступа к базам данных позволяет взаимодействовать (в той или иной степени) с данными, написанными практически на любом языке, однако своим существованием она во многом обязана языку XML.

Используемые в языке XML теги представляют семантику данных, которая может быть распознана разнородными системами. Таким образом, XML облегчает выполнение запросов на получение информации, хранящейся на различных платформах в различных структурах данных. XML помогает системе расшифровывать данные из разнородных источников, а сети и оснащенные средствами многопоточковой обработки операционные системы одновременно обрабатывают запросы пользователя на получение данных из нескольких хранилищ.

Стивен О'Грейди аналитик, занимающейся исследованиями рынка компании RedMonk,

полагает, что будущее за технологиями федеративного доступа к базам данных. «Пусть разные поставщики предлагают различные решения проблемы, – отметил он, – но все дело в том, что технические средства в данной сфере наконец-то вышли на уровень потребностей, а значит, что пришло время средств управления федеративными данными. По мере того как предприятия будут проявлять все больший интерес к информации, находящейся на периферии главного поля деятельности предприятия, спрос на эти изделия будет расти».

Федеративная технология, как правило, обходится дешевле, позволяет выполнять задачу быстрее и порождает меньше ошибок, поскольку данные остаются на местах своего постоянного хранения. Кроме того, с помощью этой методики можно с легкостью обращаться к новым источникам данных.

Гипотетическая модель среды «Электронный город» (рис. 4).

Информационное облако, формируемое системой серверов учреждений, организаций и предприятий, объединяется в глобальную информационную среду, к которой любыми способами подключаются горожане.

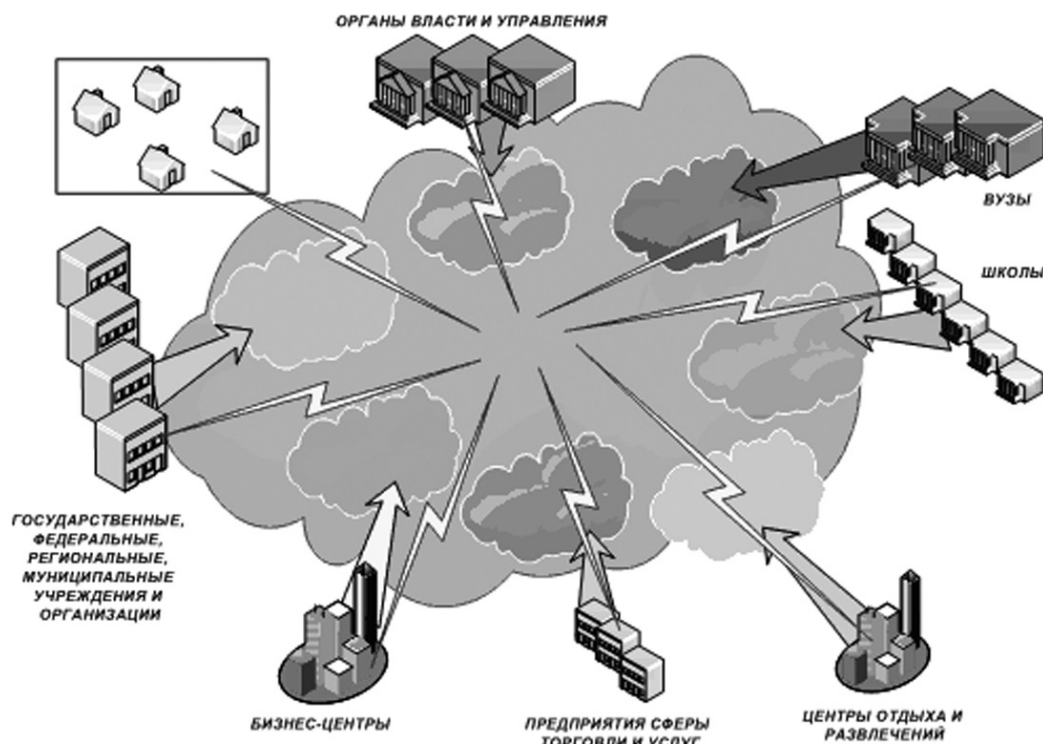


Рис. 4. Гипотетическая модель среды «Электронный город»

Литература

1. Черняк, Л. Интеграция данных: синтаксис и семантика / Л. Черняк // Открытые системы. – № 10. – 2009. [Электронный ресурс] – URL: <http://www.osp.ru/os/2009/10/11170978/> (20.04.2011).
2. David Geer, Federated Approach Expands Database-Access Technology. IEEE Computer. – May 2003. IEEE Computer Society, 2003, All rights reserved. Reprinted with permission.