



# Корпоративное хранилище данных Росбанка

В конце 2004 года в Росбанке приступили к строительству корпоративного хранилища данных для объединения информации, содержащейся в разных системах, то есть началось создание единого репозитория данных, с которым мог бы работать бизнес

Основное назначение корпоративного хранилища данных — обеспечить руководство банка, его менеджеров и аналитиков финансовой и клиентской информацией, необходимой для принятия решений по управлению рисками, ликвидностью и продажам банковских продуктов. В банке требовалось создать систему корпоративной управленческой отчетности, которая бы помогала централизованно готовить отчеты для акционеров, подраз-

делений банка и внешних организаций, включая отчетность для ЦБ РФ.

Было принято решение о строительстве корпоративного хранилища данных для объединения информации, хранящейся в разных системах, то есть о создании единого репозитория данных, с которым мог бы работать бизнес. От предложения попытаться построить подобное хранилище самостоятельно, используя накопленный опыт, в Росбанке отказались, поскольку крупный банк не может позволить се-

бе риски, связанные с такими «самодеятельными» решениями. После изучения существующих на рынке промышленных систем, а также оценив возможности разных разработчиков, в Росбанке остановились на индустриальной модели хранилища данных финансовой организации IBM Banking DataWarehouse (BDW).

«В решении IBM хорошо реализована работа с самой информационной моделью данных и дано описание банковского бизнеса. Несмотря на то что зарубежный опыт отличается от российской банковской практики, в нем мы нашли все то, что могло бы понадобиться среднему или крупному коммерческому банку. Например, работа в хранилище данных по клиентам, счетам, с ценными бумагами и пр. Это все часть информационной модели, которую можно развивать.

## Интеграция данных — на пути к совершенству



Хранилища данных, системы

управления нормативно-справочной информацией, CRM и многие другие системы нуждаются в данных, однако процессы их получения, очистки и подготовки зачастую весьма нетривиальны, а стоимость создания таких процессов, их поддержки и развития довольно высока.

С другой стороны, немногие специализированные инструменты позволяют максимально автоматизировать и упростить (а значит, и ускорить) все этапы разработки процессов интеграции данных, обеспечивая при этом высокий уровень производительности и масштабируемости с ростом объема данных, а также дают возможность использования всех созданных наработок в рамках других проектов и приложений, в ходе которых необходимо выполнять операции над данными.

Ключевыми аспектами доверия к данным, получаемым с помощью любого средства интеграции, является прозрачность их потоков — от источников до конечных потребителей, а также понимание на уровне бизнес-определений тех ключевых показателей, которые пользователи видят в отчетах и системах. Для реализации этих требований необходимо наличие единой среды метаданных, описывающих источники данных, правила их преобразований, структуры приемников информации, а также те самые бизнес-показатели, которые пользователи анализируют в рамках отчетов.

IBM Information Server является интеграционной платформой, включающей в себя ряд модулей, автоматизирующих практически все этапы процессов интеграции данных (извлечение данных, очистка и преобразование, доставка информации, управление метаданными и т. д.). IBM InfoSphere Data Stage, как один из ключевых модулей этой платформы, предоставляет доступ к огромному разнообразию типов источников данных и обеспечивает высокий уровень производительности на всех стадиях работы с информацией.

— Константин Русаков,  
эксперт по программному обеспечению  
IBM Infospere



началось всестороннее развитие системы, и к проекту подключился бизнес-партнер IBM, компания «Терн», которая помогла реализовать ряд крупных задач в рамках проекта. «Например, с их помощью мы автоматизировали работу call-центра. Эта задача не вполне характерна для хранилища, так как она не связана с формированием аналитической или управленческой отчетности. Но в результате получилась система, которая связала несколько работающих решений, при этом она активно использует информацию, накопленную в хранилище. Система обеспечила реальную работу call-центра в режиме 24×7. Для этого качество выполнения всех процедур загрузки и обработки информации должно было быть очень высоким», — пояснил Полев.

Решение в Росбанке использует систему IBM InfoSphere DataStage для сбора, трансформации и загрузки данных в корпоративное хранилище. Еще одним шагом в развитии системы стал переход к кластерной архитектуре сервера DataStage. Это было необходимо, чтобы увеличить производительность системы и обеспечить ее масштабируемость. Для этой же цели было реализовано распараллеливание на несколько вычислительных узлов конфигурации СУБД DB2, которая лежит в основе хранилища. «Все эти действия пришлось выполнить, чтобы иметь возможность работать ежедневно с объемами данных в несколько миллионов записей», — пояснил Полев.

Средства BDW позволяют создавать витрины на основе готовых шаблонов и изменять их в зависимости от требований банка. При этом под термином «витрины» в терминологии BDW следует понимать следующее. В хранилище данных выгружается вся информация, которая содержится в схеме BDW. Для удобства пользователей и из соображений безопасности отдельно от хранилища организуются специальные базы данных («витрины») по определенным темам, таким как «кредитный портфель», «информация о клиенте» и др. Витрины данных содержат только ту информацию из хранилища, которая необходима для того или иного бизнес-процесса, и возможность работать с этими данными имеют только сотрудники соответствующего подразделения банка.

Многие ответы на текущие задачи по анализу тенденций и управлению банком, требования по управленческой отчетности могут быть реализованы с помощью возможностей IBM BDW. ✘

И это единая модель и единая технология, которая позволяет ничего не переделывать из того, что было сделано ранее, а добавлять к имеющейся информации новую, необходимую на данном этапе развития», — рассказал начальник управления развития корпоративного хранилища данных Виктор Полев.

В IBM BDW использован опыт более чем 180 финансовых организаций по всему миру. В состав решения входят: универсальная логическая модель, позволяющая легко расширять возможности бизнес-анализа и интегрировать данные (поддерживается более 6 тыс. экономических терминов); богатая модель данных для всего спектра задач банка; отработанная методология поэтапного построения решения; более 90 готовых шаблонов для всех основных аналитических задач и отчетности.

Больше года заняла локализация BDW, которая выполнялась силами сотрудников Росбанка. Без этого работать с BDW было бы сложно, а порой и просто невозможно.

Первый этап проекта, который заключался в загрузке первичной информации о клиентах в систему и построении витрин данных, выполнялся также самостоятельно специалистами Росбанка. Но на стадии, когда с системой уже активно работали бизнес-пользователи,