

Автор:

Тарантин Дмитрий Алексеевич

магистрант

ФГОБУ ВО «Финансовый университет
при Правительстве Российской Федерации»
г. Москва

ПРЕИМУЩЕСТВА РАЗРАБОТКИ ВИТРИН ДАННЫХ ДЛЯ ВЕДЕНИЯ ОТЧЁТНОСТИ В БАНКЕ

Аннотация: для автоматизации процессов предоставления финансовой и аналитической отчетности в настоящее время актуальным и действенным решением является построение витрины данных. В статье рассматриваются актуальность, основные принципы и преимущества от использования витрин данных с целью совершенствования процессов построения отчетности в крупных банках.

Ключевые слова: витрина данных, хранилище данных, автоматизация отчетности.

В настоящее время крупные банки используют такой инструмент для построения отчетности как витрины данных. Витрины данных являются неотъемлемой частью систем поддержки принятия решений (DSS) и BI (Business Intelligence) в современных крупных компаниях. Витрины представляют собой срез хранилищ данных, ориентированных на определенный круг пользователей, например, витрина данных пассивов юридических лиц разрабатывается для корпоративного подразделения организации, а витрина данных системы управления персоналом для HR-подразделения.

Идея витрин данных (Data Mart) возникла, когда стало ясно, что разработка корпоративного хранилища является длительным и дорогостоящим процессом, требующим значительных усилий для анализа деятельности организации и переориентации ее на новые технологии. Появились витрины данных, чтобы избежать трудностей при разработке и внедрении хранилищ [1].

Витрина данных – это специализированное хранилище, которое служит, как правило, единственным направлением деятельности организации, например, учет запасов. Построение витрин данных – менее дорогостоящий процесс, чем создание хранилищ данных, поскольку бизнес-процессы, происходящие в одной из функций организации, лучше изучаются и сложнее, чем процессы во всей организации.

Концепция витрин данных была предложена Forrester Research в 1991 году. По мнению авторов, витрины данных представляют собой набор тематических баз данных, содержащих информацию, относящуюся к определенным аспектам деятельности организации.

Эта концепция имеет ряд несомненных преимуществ:

- 1) аналитики видят и работают только с теми данными, которые им действительно нужны;
- 2) целевая база данных находится как можно ближе к конечному пользователю;
- 3) витрины данных обычно содержат тематические подмножества предварительно агрегированных данных, их легче разрабатывать и настраивать;
- 4) для реализации витрин данных не требуются высокопроизводительные вычисления.

На современном уровне развития информационных технологий витрина подразделения может быть организована за 2–3 месяца. Установлено, что наиболее оптимальным вариантом использования витрин является обслуживание 10–15 человек. Следует отметить, что успех небольшого проекта (стоимость которого мала по сравнению с расходами на создание корпоративного хранилища), приводит к быстрой окупаемости затрат и способствует продвижению новых технологий [2].

Преимущества использования витрин данных – это физическое разделение данных между группами аналитиков, а также относительная простота семантики данных в рамках одного витринного экрана. Успех внедрения витрин привел к появлению концепции замены корпоративного хранилища коллекцией витрин

данных. Тем не менее, использование витрин данных показало, что с увеличением числа витрин в организации увеличивается сложность их взаимодействия, поскольку невозможно сделать витрины полностью независимыми от хранилищ данных. Поэтому чаще всего разработка корпоративного хранилища идет параллельно с разработкой и внедрением витрин данных.

При построении схемы взаимодействия между корпоративным хранилищем и витринами данных рекомендуется определить некоторую специальную структуру для хранения исторических данных и дополнительно развернуть ряд витрин, заполненных данными из этой структуры. Таким образом, можно выделить два процесса: накопление исторических данных и их анализ. Фактическим стандартом структуры данных при разработке хранилищ данных и витрин данных является «звезда», основанная на одной таблице фактов и наборе таблиц измерений.

Формирование аналитических отчетов диктует бизнесу необходимость сбора информации из нескольких баз данных транзакций разной структуры и содержания. Это неудобно и требует серьезных усилий. Главная проблема такого подхода – несогласованность и непоследовательность различных баз данных. Соответственно, говорить о высоком качестве аналитической информации в данном случае не приходится. Размыщляя о предоставлении надежной информации, необходимой для оперативного анализа и принятия решений, бизнес осознает необходимость создания корпоративных информационных хранилищ.

Такой репозиторий представляет собой предметно-ориентированную базу данных, расположенную отдельно от производственных систем и объединяющую данные из промышленных баз. Информация в нём поступает с помощью определенного инструмента (ETL-инструмента).

Среди многих аспектов строительства и эксплуатации хранилищ можно выделить обеспечение доступа к корпоративным данным. Его можно реализовать как напрямую, так и через витрины данных. При «прямом доступе» аналитики для отчетности получают доступ непосредственно к самому хранилищу данных. Преимуществом такого подхода является отсутствие дополнительного шага в

процедуре ETL. При изменении бизнес-задач нужно только перенастроить интерфейсы доступа к хранилищу.

Недостаток прямого доступа можно назвать самим хранилищем данных, поскольку он не ориентирован на решение каких-либо функциональных аналитических задач [3].

В большинстве случаев для выполнения функционально закрытых аналитических задач было бы правильно создать макет данных, который представляет собой отдельную базу данных (возможно, многомерную). Витрины данных – это более совершенная система, которая позволяет детально, градуировано, оптимизировать и обновлять информацию по расписанию и загружать данные в Интернет [4].

Витрины, как правило, работают с данными одной системы учета (в отличие от прямого доступа), что положительно влияет на скорость работы в системе. Чаще всего информация поступает из хранилища, что можно считать недостатком витрины данных, поскольку для ее реализации требуется дополнительный шаг процедуры ETL. Другим недостатком витрины можно считать ее привязанность к бизнес-процессам. Когда меняется бизнес-процесс, часто нужно изменить процедуру ETL, возможно, даже полностью перестроить витрину. Среди недостатков этого подхода можно выделить более высокую стоимость сбора и обновления источников данных, а также проблему консолидации данных в конкретном источнике.

Витрины данных – один из самых больших источников информации для современных аналитиков. И, по словам Gartner, в ближайшее время останется одним из ключевых компонентов ИТ-инфраструктуры предприятий.

Аналитики прогнозируют, что в ближайшие годы на российском финансово-рынке увеличится количество проектов по созданию хранилищ данных. И эти прогнозы подтверждаются текущей деятельностью в этом сегменте рынка. Это связано, с одной стороны, с началом размораживания ИТ-бюджетов после кризиса, с другой стороны, происходящим переходом банков от эксплуатации

хранилищ данных собственной разработки к промышленным решениям. Практически каждый банк, который сегодня выбирает решение для построения хранилища данных, уже имеет некоторый опыт создания хранилища самостоятельно и знает, с какими ограничениями в разработке и обслуживании решения приходится сталкиваться. Для обновленных кредитных учреждений внедрение хранилищ данных – лучший способ добиться прозрачности в управлении растущим бизнесом и внедрить нормативные требования для обязательной отчетности [5].

Курс, принятый банками для автоматизации отчетности, значительно повысил приоритет обеспечения качества данных в хранилище. Оказалось, что полнота, согласованность и точность данных, собранных из модулей учета, обычно оставляют желать лучшего. Чтобы выявить и устранить эту проблему, поставщики предлагают механизмы обогащения и даже ввод недостающих данных в хранилище. Но главным инструментом здесь должны быть организационные изменения. Прежде всего, речь идет о формировании, документировании и реализации правил ввода и обработки учетных данных. Сложность решения задачи мониторинга и обеспечения качества данных сегодня, пожалуй, является самым болезненным моментом в создании хранилищ данных в российских банках.

Список литературы

1. Туманов В.Е. Проектирование хранилищ данных для систем бизнес-аналитики // Бином. – 2014. – С. 178.
2. Чубукова И.А. Data Mining // Бином. – 2014. – С. 109.
3. Григорович А.Г. Архитектура хранилищ данных с ненормализованными отношениями. – Новосибирск: СибАК, 2013.
4. Туманов В.Е. Проектирование реляционных хранилищ данных / В.Е. Туманов, С.В. Маклаков // Диалог-МИФИ. – 2013. – С. 90.
5. Коровкин С.Д. Решение проблемы комплексного оперативного анализа информации хранилищ данных / С.Д. Коровкин, И.А. Левенец, И.Д. Ратмanova [и др.] // СУБД. – 2013. – №5–6. – С. 47.