

# Создание пользовательских витрин в банковской сфере с использованием корпоративного хранилища данных

А. О. Сынкова, email: an.synkova@gmail.com

М. В. Матвеева, email: marie.matveeva@gmail.com

Воронежский государственный университет

***Аннотация.** Данная работа посвящена процессу проектирования и реализации корпоративного хранилища данных. Проведен анализ существующих решений, рассмотрены процесс загрузки данных от систем-источников до конечных витрин и типовая архитектура хранилища. Также проведен анализ конкретной задачи, поставленной в рамках развития КХД.*

***Ключевые слова:** база данных, корпоративное хранилище данных, Oracle, SQL.*

## Введение

С началом эры BigData у многих предприятий появилась необходимость в том, чтобы хранить и анализировать большие объемы информации, которая возникает в процессе их деятельности. Используя эту информацию, компании могли бы формировать различные отчеты, анализировать уровень эффективности своей работы, выявлять слабые места и улучшать качество обслуживания клиентов.

Для того, чтобы реализовать указанные цели бизнес-пользователям необходимо видеть нужную информацию в удобном для себя виде. Но как правило, данные поступают из систем-источников, которые имеют абсолютно разную структуру. Кроме того, часто компаниям требуется не просто собирать, агрегировать и анализировать информацию из различных источников, но и проводить анализ по данным за длительный период времени.

Все это привело к появлению корпоративных хранилищ данных, которые позволили загружать большие объемы информации из систем-источников, приводить полученные данные к унифицированному виду и агрегировать их в соответствии с нуждами конечных пользователей. Однако КХД в большинстве случаев не могут рассматриваться как коробочные продукты, поэтому возникает необходимость в разработке хранилищ для конкретных предприятий на базе уникальных моделей данных.

В рамках данной работы будет рассмотрена разработка корпоративного хранилища данных для Банка и в частности, реализация задачи «Формирование заданий на обзвон по неисполненным заявкам с целью увеличить количество оформленных договоров на кредиты».

## **1. Анализ существующих решений**

Поскольку основной задачей КХД является загрузка данных из различных систем-источников и преобразование этих данных в удобный для конечных пользователей вид, хотелось бы привести описание процесса загрузки данных в хранилище и его типовую архитектуру.

### **1.1. Процесс загрузки**

ETL (Extract, Transform, Load) – системы, которые применяются, чтобы привести данные из разных учетных систем к единым справочникам и загрузить их в КХД. Включают в себя следующие этапы:

1. Extract – извлечение данных из внешних систем-источников.
2. Transform – их трансформация и очистка, приведение к унифицированному виду. Целью этого этапа является подготовка данных к загрузке в хранилище и приведение их к более удобному для последующего анализа формату. В процессе преобразования данных чаще всего выполняется преобразование структуры, агрегирование данных, перевод значений и создание новых данных.

3. Load – перенос данных из промежуточных таблиц в структуру КХД. При этом, при очередной загрузке в хранилище переносится не вся информация из источников, а только та, которая была изменена в течение промежуточного времени, прошедшего с предыдущей загрузки. В процессе загрузки выделяется 2 потока:

- Поток добавления – в хранилище переносится новая, ранее не существовавшая информация;
- Поток обновления (дополнения) – в хранилище передается информация, которая существовала ранее, но была изменена или дополнена.

С точки зрения процесса ETL архитектуру хранилища можно представить следующим образом:

1. Системы-источники: содержат данные в виде таблиц, совокупности таблиц или файла, в котором данные разделены символами разделителями.

2. Промежуточная область: содержит вспомогательные таблицы, которые создаются временно для процесса выгрузки.

3. Получатель данных: хранилище или база данных, в которую помещаются извлеченные данные.

Общая схема загрузки представлена на рис. 1.

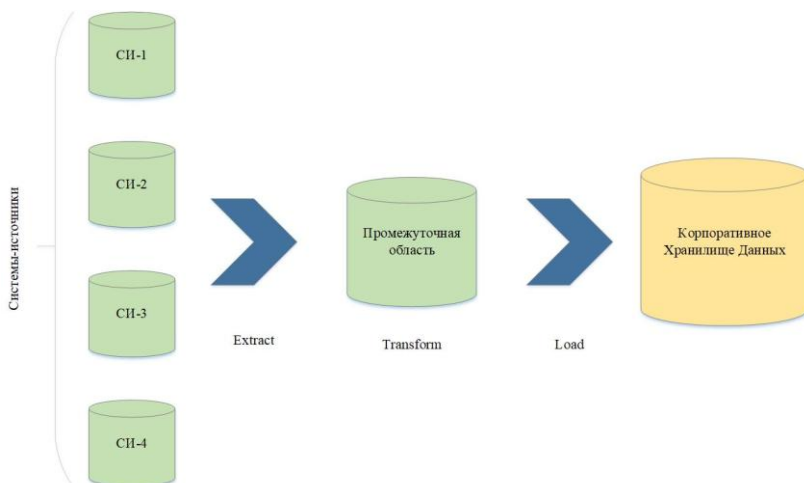


Рис. 1. Общая схема загрузки данных в КХД

### 1.2. Основные слои КХД

Прежде чем попасть к конечным пользователям, данные проходят по слоям самого хранилища. В каждом слое они претерпевают определенные преобразования. Рассмотрим их подробнее.

1. ODS (Operational Data Store) – область оперативного хранения данных. Данные в оперативной области полностью соответствуют данным в системах-источниках. При загрузке данных не производится никаких изменений, кроме обогащения служебной информацией (системной датой загрузки, идентификацией загрузочного процесса и пр.).

2. STG (Staging) – область подготовки данных. На этом слое данные преобразуются к структуре области детальных данных. Также происходит отбор инкремента исходных данных относительно уже загруженных в детальный слой.

3. DDS (Data Detail Store) – область детальных данных. В этом слое данные хранятся на том же уровне детализации, что и в системе источнике. Они не должны агрегироваться под нужды пользователей, никаких расчетов не производится.

4. PRE\_DM – область, предназначенная для предварительных расчетов сложных показателей. Расчеты выносятся в этот слой в двух случаях:

- когда сложный расчет витрины лучше разделить на несколько промежуточных этапов с точки зрения производительности;
- для сохранения предварительных расчетов в тех случаях, когда какая-то часть алгоритма расчета является общей для нескольких витрин, т.е. для исключения ситуации вычисления одного и того же алгоритма в каждой витрине отдельно.

5. DM (Data Mart) – область пользовательских агрегатов и витрин. Предназначена для конечных пользователей – аналитиков и других сотрудников банка. В витринах данные из детального слоя агрегируются согласно требованиям заказчика.

Более подробная схема слоев, из которых состоит КХД представлена на рис. 2.



Рис. 2. Слои КХД

### 1.3. Основные типы сущностей

Далее рассмотрим основные типы сущностей (таблиц), которые используются для организации данных в детальном слое КХД банков.

1. Факты (Fact) – основные таблицы хранилища данных. Отражают какие-либо события. Например, проводки – движение денежных средств по счетам (перечисление заработной платы, оплата покупок картой банка и т.д.). Поскольку в крупном банке ежедневно производится огромное количество операций с денежными средствами, для таблиц этого типа характерен большой объем данных (около миллиона/миллиарда записей каждый день) и частые изменения.

2. Измерения (Dimension) – таблицы, которые хранят данные о таких понятиях, как договор, сделка, клиент, счет и т.д. Примерный объем данных таких таблиц – миллион записей за все время.

3. Справочники (Dictionary) – используются для расшифровки каких-либо значений. Например, справочник типа клиента: 1 – физ.лицо, 2 – юр.лицо, -1 – не определено. Такие таблицы хранят меньше всего записей и редко изменяются.

## 2. Общая структура и функциональность

Далее рассмотрим реализацию конкретной задачи. Заказчиком была поставлена задача реализовать витрину данных, в которой бы формировались задания на обзвон по неисполненным заявкам на кредит наличными (т.е. заявкам, которые не были доведены до оформления договора). Целью данной задачи являлось увеличение количества

продаж путем обзвона клиентов и закрытия сделки оформления кредита по совершенным ранее заявкам.

### 3. Модель данных

Ниже описана часть модели данных детального слоя КХД, которая была задействована при решении поставленной задачи. Логика конечной витрины строится соединениями представленных ниже таблиц:

1. Заявка. Таблица содержит информацию по заявкам клиентов, в том числе заявки на кредиты наличными.

2. Статус заявки. Таблица является справочником статусов заявок. В данной задаче необходимо отбирать заявки по всем статусам, кроме статусов «Отменена» и «Просрочена».

3. Канал заведения заявки. Таблица является справочником каналов, по которым клиенты банка могут заводить заявки. В данной задаче необходимо отбирать заявки, оформленные по каналам «Фронт» и «Интернет».

4. Макропродукт. Таблица является справочником макропродуктов. По условию задачи необходимо отбирать заявки на кредит наличными.

5. Договор. Таблица содержит данные о заключенных договорах. Поскольку целью задачи является оформление договоров по неисполненным заявкам, нужно отбирать те заявки, для которых еще нет соответствия в таблице договоров.

6. Решение по заявке. Таблица содержит данные о решениях по заявке, которые принимают различные инстанции. По условию задачи необходимо отбирать заявки, которые прошли полную проверку и по итогу последней проверки получили статус «Одобрена».

7. Клиент. Содержит основную информацию о клиентах банка. В рамках задачи в конечную витрину добавляется ID клиента, с которым необходимо произвести коммуникацию.

8. Структурная единица банка.

9. Тип структурной единицы. Таблица является справочником. В рамках задачи необходимо отбирать заявки, оформленные в структурных единицах с типами: 'Центр Обслуживания Клиентов', 'Объект Почтовой Связи' и 'Стойка'.

10. Центр обслуживания клиентов.

11. Подтип центра обслуживания клиентов. в рамках задачи необходимо отбирать заявки, оформленные в структурных единицах с типами: 'Центр Обслуживания Клиентов', 'Объект Почтовой Связи' и 'Стойка'.

12. Статус центра обслуживания клиентов. Таблица является справочником. В рамках задачи отбираются центры обслуживания со статусом 'Действующий Центр Обслуживания Клиентов или Стойка'.

13. Часовой пояс центра обслуживания клиентов. Справочник содержит информацию о часовых поясах точек относительно московского времени (в формате MSK+xx). В рамках задачи используется с целью вычисления возможного времени проведения коммуникаций с клиентом с учетом его часового пояса.

На рис. 3 изображена диаграмма базы данных на логическом уровне.

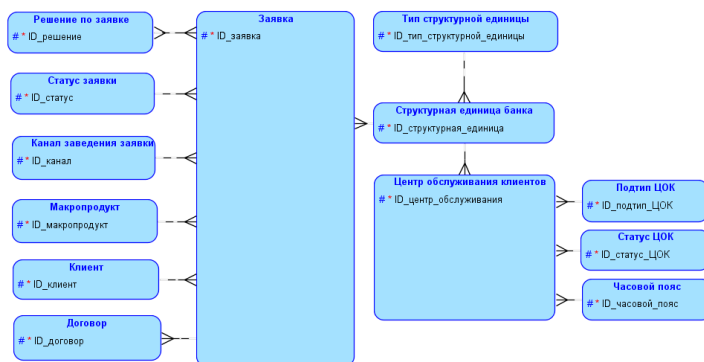


Рис. 3. Диаграмма базы данных на логическом уровне

### Заключение

По итогам работы был проведен анализ существующих решений в области разработки корпоративных хранилищ данных, описаны процесс загрузки данных из систем-источников и архитектура хранилища в общем виде, а также рассмотрена конкретная задача, поставленная в рамках развития КХД. Проведен анализ требований бизнес-пользователей и выявлены условия, по которым необходимо построить витрину данных для использования сотрудниками банка.

### Литература

1. Дейт, К. Дж. Введение в системы баз данных. 8-е издание/ К. Дж. Дейт: – Москва: Издательский дом «Вильямс», 2005. – 1328 с.
2. Хоббс Л. Разработка и эксплуатация хранилищ баз данных/ Лириан Хоббс, Сьюзан Хилсон, Шилпа Лоуенд: – Москва: Издательство «КУДИЦ-Образ», 2004. – 596 с.