

Супаков Евгений Сергеевич

Магистрант НАЧОУ ВПО СГА

Направление: Информатика и вычислительной техники

Магистерская программа: Распределенные автоматизированные системы

Исследование информационного состояния корпоративных баз данных

Аннотация. В статье описываются результаты исследования сегмента крупной корпоративной базы данных и предложены решения для улучшения информационного состояния БД. Приводятся примеры типичных ошибок при проектировании БД, и приводится список возможных первопричин их появления. В статье рассматриваются способы сопровождения содержимого корпоративных баз данных, и проводится анализ их эффективности.

Ключевые слова: СУБД, база данных, GUI – языки разработки программного обеспечения, схема БД, стандартизация ввода информации, движение данных.

Технологический рывок в области хранения данных, совершенный в прошлом веке, поднял на качественно новый уровень информатизацию общества. На сегодняшний день фундаментом для обработки, хранения и накопления новой информации служат реляционные базы данных. Их применяют повсеместно, как частные лица, так и коммерческие структуры.

Основной причиной такой популярности реляционных баз данных является факт, что на текущий момент, данный тип баз данных представляет собой удобную и безопасную форму хранения информацию различного типа от текстовых данных до графических изображений и мультимедиа. И что особенно важно с помощью реляционного подхода можно описать практически любую область деятельности общества. В коммерческой деятельности,

используемые реляционные базы данных принято называть корпоративными. Их основным отличием от остальных БД является чрезвычайно сложная структура отношений между таблицами и огромное количество данных содержащихся в них. Информационная структура предприятия может включать не только базы данных, в которой хранится оперативная и постоянно обновляемая и дополняемая информация, но и включать в себя хранилища и витрины данных. Но основой всегда служит база данных, в которой хранятся актуальные данные, и с которыми работают сотрудники компании.

Данные содержащиеся, в таблицах корпоративной БД представляют особую ценность для компании. Успех компании напрямую зависит от того насколько полные, непротиворечивые и структурированные данные содержит корпоративная база данных. Это определяется следующими факторами:

1. Качеством изначальной проектировки базы данных.
2. Способом ввода и изменения информации в БД.
3. Квалификацией сотрудников работающих непосредственно с базой.
4. Оперативностью сопровождения данных.
5. Возможностями СУБД.

Типичными ошибками при проектировании корпоративной базы данных, являются: недостаточная нормализация таблиц, отсутствие таблиц стандартизации ввода данных, отсутствие резервных атрибутов таблиц для будущего расширения базы данных.

Основной причиной появления подобных ошибок является, как правило, сжатые сроки разработки и недостаточно точная формулировка требований к базе данных. Также в процессе коммерческой деятельности требования к базе данных могут меняться.

В данной ситуации необходимо провести анализ структуры базы данных и таблиц. И сформировать новые требования к базе данных можно на основании результатов работы сотрудников с ней. Но для этого сотрудники должны быть достаточно квалифицированы, и иметь представление о структуре базы данных.

Для мониторинга информационного состояния корпоративной базы данных требуются администраторы данных, чьими непосредственными обязанностями будет стандартизация хранимых данных, анализ движения информации в БД и выработка решений для улучшения её информационного состояния.

Для информационного сопровождения корпоративной базы данных требуется достаточно большое количество администраторов данных, что влечет дополнительные издержки для организации. Кроме того, это не решит проблемы с контролем за содержимым базы данных. В процессе деятельности предприятия количество информации в базе данных будет, неуклонно расти, что в свою очередь повлечет трудности в контроле за содержимым таблиц. В определенный момент администраторы уже не будут успевать оперативно, проводить корректировку информации.

Лучшим решением по сопровождению является разработка приложения по автоматизированному сопровождению данных. Это требует детального изучения базы данных и рабочих процессов организации. В качестве языка разработки приложения для сопровождения БД наиболее оптимальным будет использование GUI-языков разработки, по причине простоты, удобства и скорости написания оконных приложений с дружелюбным пользовательским интерфейсом, при сравнительно низких трудозатратах. Путем сравнения наиболее популярных программных средств установлено, что лучшим средством разработки приложения является продукт компании Sybase – PowerBuilder.

В исследовании была проанализирована база данных компании «Сплайн-Центр». Компания Сплайн-центр занимается продажей и сопровождением программы консультант плюс, а также, программных продуктов компании 1С, предназначенных для организаций. За один рабочий день происходит огромное количество операций с базой данных. Непрерывно информация о клиентах и событиях таких как: продажа программы, назначение ознакомительных и учебных семинаров, выезд к клиенту для ознакомления и установки ПО,

уточнение информации о клиентах, нахождение новых клиентов, депонирование клиентов и т. д. обновляется и добавляется. Таким образом, данные в корпоративной БД компании Сплайн-центр постоянно находятся «в движении».

В представленной ниже схеме содержится лишь часть таблиц, с которыми работают сотрудники компании.

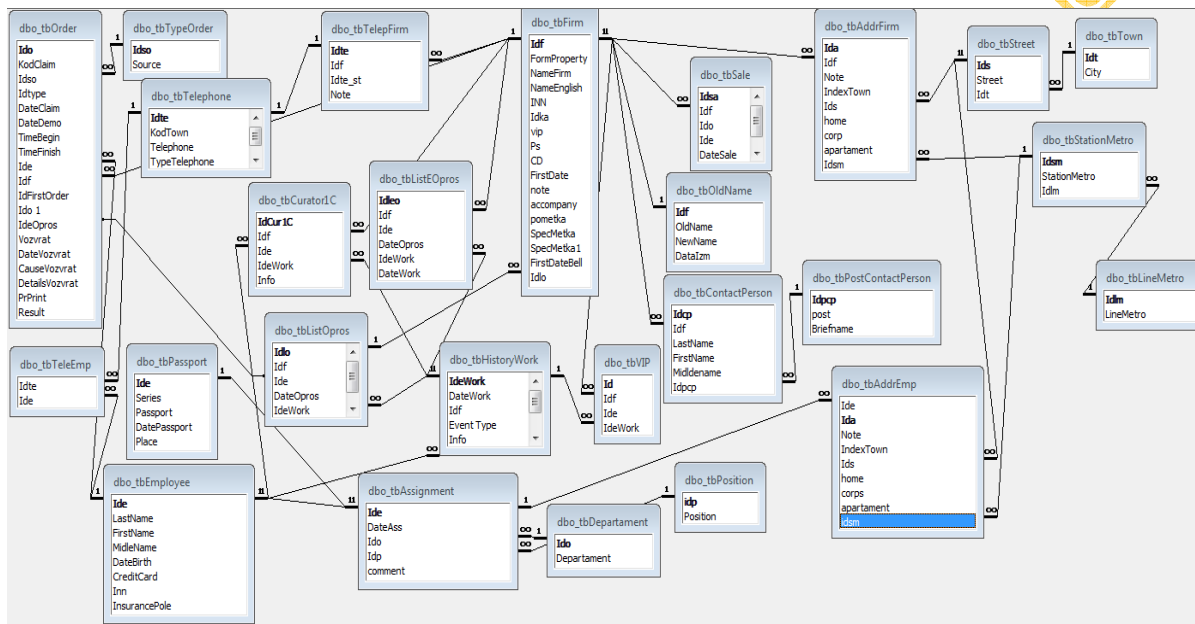


Схема структуры сегмента базы данных информационно аналитического отдела компании «Сплайн-Центр»

Рассматриваемая корпоративная БД содержит более сотни разных таблиц. Часть из них содержит устаревшие данные и почти не используются. Подобные таблицы могут содержать наряду с устаревшей информацией, данные, которые представляют ценность для сотрудников. Поэтому удаление этих таблиц на текущий момент невозможно. Условно таблицы, представленные на схеме, можно условно разделить на четыре группы:

Таблицы данных клиентов.

Таблицы событий.

Таблицы данных сотрудников компании «Сплайн-Центр».

Справочные таблицы.

В результате исследования базы данных компании «Сплайн-Центр» были выявлены ошибки при проектировании. В качестве примера, можно привести справочную таблицу `dbo_tbTown`. В данной таблице хранятся все населенные пункты Московской области, включая Москву. В ней содержатся данные следующего вида:

- Города (Зеленоград, Раменское и т. д.)
- Районы (Ногинский, Одинцовский и т. д.)
- Населенные пункты в районах (д. Донино Раменский р-н, с. Благовещенье Сергиево-Посадский р-н).

Ошибка при построении структуры таблицы `dbo_tbTown`, заключается в том, что данные, содержащиеся в ней, не отвечают свойству атомарности. Это замечание касается третьего пункта списка – населенных пунктов в районах. Данные записи содержат информацию о районе и населенном пункте. При добавлении новой организации в базу и указании адреса, у сотрудников могут возникнуть сложности, по причине наличия отдельной записи с названием района и записей с указанием населенного пункта, а затем района. Так как не все населенные пункты районов отмечены в таблице, то может возникнуть ситуация, когда сотруднику сначала приходится искать `id` населенного пункта, и если его нет в базе, искать `id` района.

Для исключения возникновения подобных ситуаций существуют два решения:

- создание отдельной таблицы населенных пунктов с добавлением столбца с `id` поселков в таблицу городов;
- создание составного ключа таблицы `dbo_tbTown`.

Оба метода позволят упростить процесс присвоения `id` населенного пункта или района фирмы в таблицу `dbo_tbAddrFirm`.

Второй, замеченной при исследовании структуры БД, ошибкой является отсутствие в базе данных «справочной» таблицы для поля «форма собственности» таблицы `dbo_tbFirm`. Так как количество форм собственности

организаций ограничено, есть возможность упорядочить их ввод в базу путем выбора их из списка значений или из отдельной таблицы.

Третьей ошибкой является отсутствие проверки на наличие похожей записи при добавлении нового клиента в базу. Под записью в данном случае подразумевается пространство сущностей содержащих информацию о конкретном клиенте. Пространство сущностей – совокупность сущностей воспринимаемых как единое целое [3, с. 19]. В свою очередь сущность – любой конкретный или абстрактный объект, включая связи между объектами [3, с. 20].

Это недоработка программы ввода, но она напрямую влияет на состояние данных в БД, так как без проверки ввод новых клиентов в базу может породить дубли.

В текущей ситуации решением проблемы с дублями клиентов, может стать разработка приложения, объединяющего записи дублирующихся клиентов по всем таблицам базы данных. Подводя итоги, делается вывод о необходимости информационного сопровождения корпоративной базы данных кроме того к такому же выводу пришли авторы статьи «Решение проблемы комплексного оперативного анализа информации хранилищ данных». Цитата: «Управление любой сложной социально-экономической системой – от предприятия до региона или отрасли – весьма затруднительно без обратной связи, которая заключается в отслеживании и анализе данных, отражающих состояние этой системы и ситуацию вокруг нее. Постоянная доступность актуальной информации дает возможность оценить текущее положение дел, а обзор изменения конкретных характеристик во времени позволяет обнаружить тенденции развития системы и сделать выводы о том, что ожидает ее в будущем. Таким образом, обладая всей полнотой сведений о состоянии системы и ее элементов в статике и динамике, управляющий персонал может принимать грамотные решения по применению мер регулирования. Такое управление основано на знании и потому более эффективно, чем принятие важных решений вслепую» [2].

Наиболее оптимальным решением для информационного контроля корпоративной БД является разработка автоматизированной системы сопровождения данных. Кроме того, в ходе исследования была проанализирована корпоративная база данных и обнаружен ряд недостатков в её структуре и проблемы с сопровождением данных. Предложены решения конкретных недостатков в её структуре.

Литература

1. Кузнецов С.Д. Базы данных. М.: Академия, 2012.
2. Коровкин С.Д., Левенец И.А., Ратманова И.Д., Старых В.А., Щавелёв Л.В. Решение проблемы комплексного оперативного анализа информации хранилищ данных [Электронный ресурс] // Режим доступа: http://citforum.ru/database/articles/art_11.shtml
3. ГОСТ 34.320-96 «Информационные технологии. Система стандартов по базам данных. Концепции и терминология для концептуальной схемы и информационной базы» [Электронный ресурс] // Режим доступа: <http://www.rugost.com/>

© Бюллетень магистранта 2014 год № 4