

Как войти в норму

Другой пример избыточности данных – дублирование записей. С ним бороться гораздо сложнее. Приведем пример различного написания имени одного и того же лица в таблице базы данных.

**Перестукин Иван Иванович
Перестукин Иван Иванович
Иван Иванович Перестукин
Иван Иванович Перестукин
Иван Иванович Перестукин
Перестукин Иван Иванович
Перестукин Иван Иванович
Перестукин Иван Иванович
Перестукин Иван Иванович
Перестукин Иван Иванович
Перестукин Иван Иванович**

В данном примере приведено 11 вариантов, на самом деле их может быть значительно больше. И каждая из этих записей при анализе данных будет воспринята как другое лицо. Теперь представьте себе картину, которую вы получите, попытавшись получить из базы данных сведения о страховании, например, по водителю Перестукину Ивану Ивановичу. Трудно

себе представить, что анализ этой информация будет достоверным. Такая нормализация базы данных осуществляется методологическими способами. Основными средствами, позволяющими предотвратить повторы, являются программные средства, такие как уникальные индексы. Но уникальный индекс должен базироваться на некотором наборе параметров, которые запрещается повторять. Чем шире набор параметров, описывающих индекс, тем ниже степень уникальности, а значит и вероятность повторов. С другой стороны, жесткое значение уникального индекса может заблокировать ввод. Поэтому аналитику в каждом случае надо искать оптимальный вариант. Одной из самых сложных аналитических задач методической нормализации является именно борьба с повторами лиц, в то время как повторы записей о лицах делают весьма затруднительной такую важную часть аналитики, как история страхователя. Предположим, мы накладываем условие недопустимости повторения сочетания

Фамилия-Имя-Отчество. Это либо заблокирует ввод при первом же появлении полного тезки, либо побудит оператора обойти этот индекс, например, изменить букву или ввести лишний пробел. Значительно более приемлемым в данном случае является административное установление единого формата ввода в сочетании с добавлением дополнительного параметра уникальности, например даты рождения. То есть в тех случаях, где не удается добиться уникальности исключительно с помощью программных средств, применяют методологические средства в сочетании с программными. Но этого, как правило, бывает недостаточно.

Для того чтобы база данных функционировала стабильно, работы по нормализации учетных данных (содержания записей справочников и рабочих таблиц) рекомендуется производить не реже 1 раза в квартал накануне формирования квартальной отчетности. Мероприятия по нормализации структуры лучше производить после формирования отчетности, не реже 1 раза в год, но лучше 2-3 раза в течение календарного года. Проведение мероприятий по нормализации учетных данных без остановки основной работы в системе возможно, хотя и не желательно. Проведение мероприятий по нормализации структуры БД без остановки работы в системе может привести к весьма нежелательным последствиям. Поэтому при проведении мероприятий по нормализации и других «профилактических и ремонтных» работ в базе данных и в системе управления необходимо предусматривать технологические перерывы в их работе.

Для достижения корректности учетных записей необходимо систематически производить мероприятия по их нормализации, которые состоят в выполнении серии запросов на корректность и исправлении учетных ошибок.

Нормализация структуры и нормализация данных является одним из основных мероприятий по обслуживанию базы данных. Необходимость этого связана с тем, что любая база данных в процессе эксплуатации «засоряется» и нуждается в «уборке». Предположим, вы построили дом. Затратили на это немало сил и средств. Представьте себе, что случится с ним через несколько лет, если вы не будете его убирать, ремонтировать, вывозить мусор.

Процесс нормализации осуществляется квалифицированными IT-специалистами. Он состоит из 6 уровней, достаточно сложен, но хорошо описан в специальной литературе и не является предметом рассмотрения в нашей книге. Здесь мы коснемся лишь мероприятий по нормализации учетных данных.

Для некоторых электронных журналов оптимальным решением вопроса уникальности записей, неизменным условием формирования которых является абсолютная невозможность их повторения в базе данных, когда бы то ни было.

Присвоение шифров без внесения соответствующей записи в таблицу базы данных допускается только при абсолютном и объективном отсутствии технической возможности исполнения этого действия, как-то: нарушение электропитания, длительный сбой работы сети, и т. д. В этом случае выданные шифры заносятся на бумажный носитель до момента восстановления работы базы данных, после чего пропущенные записи подлежат немедленному вводу в соответствующую таблицу. Специалисты по учету в этом случае обязаны принять все необходимые меры для исключения дублирования уникальных номеров шифров.

