**9. Проектирование БД. Критерии оценки качества логической модели данных.**

**Аномалии при работе с таблицами. Нормализация и ее необходимость**

## Критерии оценки качества логической модели данных

**Критерии качественной базы данных:**

-Адекватность базы данных предметной области;

-Легкость разработки и сопровождения базы данных;

-Скорость выполнения операций обновления данных (вставка, обновление, удаление кортежей);

-Скорость выполнения операций по работе с данными данных.

**Адекватность базы данных предметной области**

**База данных должна адекватно отражать предметную область**. Это означает, что должны выполняться следующие условия:

1. Состояние базы данных в каждый момент времени должно соответствовать состоянию предметной области.

2. Изменение состояния предметной области должно приводить к соответствующему изменению состояния базы данных

3. Ограничения предметной области, отраженные в модели предметной области, должны некоторым образом отражаться и учитываться базе данных.

**Легкость разработки и сопровождения базы данных**

Практически любая база данных, за исключением совершенно элементарных, содержит некоторое количество программного кода в виде триггеров и хранимых процедур, о которых речь пойдет ниже, когда перейдем непосредственно к вопросам изучения языка SQL.

**Одно из назначений базы данных - предоставление информации пользователям**. Информация извлекается из реляционной базы данных при помощи оператора SQL - SELECT. Одной из наиболее дорогостоящих операций при выполнении оператора SELECT является операция соединение таблиц. Таким образом, чем больше взаимосвязанных отношений было создано в ходе логического моделирования, тем больше вероятность того, что при выполнении запросов эти отношения будут соединяться, и, следовательно, тем медленнее будут выполняться запросы.

## Нормализация и ее необходимость

**Основной пример:**Рассмотрим в качестве предметной области, как и ранее, «Университет».

Первоначальная таблица отношения «Университет» (универсальная таблица)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Caf (pk)** | **Phone** | **Disp (pk)** | **Cat (pk)** | **FIO** | **Tab\_N\_S** |
| ВТ | 211718 | САПР | ЛК | Ланцов В.Н. | 1111 |
| ВТ | 211718 | САПР | ПР | Мамаев А.А | 2222 |
| ВТ | 211718 | АрхЭВМ | ЛК | Буланкин В.Б. | 3333 |
| ВТ | 211718 | АрхЭВМ | КП | Буланкин В.Б. | 3333 |
| ИСИМ | 344248 | ЗИ | ЛК | Алешкин А.А. | 4444 |
| ИСИМ | 344248 | ЗИ | ЛБ | Алешкин А.А. | 4444 |
| ИЗИ | 317442 | Инф-ка | ПР | Алешкин А.А. | 4444 |

Очевидно, что такое представление не является самым эффективным. Т.к. скажем телефон и название кафедры надо будет вводить от раза к разу. Не исключено, что при этом будут возникать ошибки (опечатки) при вводе. Например, «ИСИМ»–«ИЗИМ» и т.п. С точки зрения пользователя это будут одинаковые записи, или по крайней мере он так будет думать. Однако при обработке данных (таблиц) это уже будут разные кафедры (телефоны, сотрудники и т.д.). Кроме того, при удалении какой либо записи (кортежа)возможно будут потери информации. Например, при увольнении «Алешкина», его необходимо удалить из таблицы. Очевидным образом потеряются и сведения о соответствующих дисциплинах. Если же, этот человек перейдет на другую кафедру, то о нем надо будет вводить всю информацию заново.

Т.о. при работе с универсальными таблицами есть существенная проблема - **Проблема избыточности данных.** Т.е. данные практически всех столбцов многократно повторяются. Это видно;

*Как следствие этой проблемы выделяется 3 аномалии*

**Аномалии вставки (INSERT)**

В отношение «Университет**»** нельзя вставить данные о сотруднике, который пока не ведет ни одну дисциплину (занятие). Действительно, если, например, на кафедре «ИЗИ» появляется новый сотрудник, Пушкин, и он пока не читает ни одну дисциплину, то мы должны вставить в отношение кортеж <ИЗИ, 317442, , null, null, Пушкин>. Это сделать невозможно, т.к. атрибуты Caf, Disp, Cat входит в состав потенциального ключа, и, следовательно, не может содержать null-значений.

Точно также нельзя вставить данные о новой кафедре.

**Причина аномалии** - хранение в одном отношении разнородной информации (и о сотрудниках, и о дисциплинах, и о видах занятий).

**Вывод** - логическая модель данных неадекватна модели предметной области. БД, основанная на такой модели, будет работать некорректно.

**Аномалии обновления (UPDATE)**

Фамилии сотрудников, наименования дисциплин, номера телефонов повторяются во многих кортежах отношения. Поэтому если сотрудник меняет фамилию, или дисциплина меняет наименование, или меняется номер телефона, то такие изменения необходимо *одновременно* выполнить во всех местах, где эта фамилия, наименование или номер телефона встречаются, иначе отношение станет некорректным (например, одна и та же дисциплина в разных кортежах будет называться по-разному). Т.о., обновление БД одним действием реализовать невозможно.

**Причина аномалии** - избыточность данных, также порожденная тем, что в одном отношении хранится разнородная информация.

**Вывод** - увеличивается сложность разработки базы данных. База данных, основанная на такой модели, будет работать правильно только при наличии дополнительного программного кода.

**Аномалии удаления (DELETE)**

При удалении некоторых данных может произойти потеря другой информации. Например, если закрыть кафедру "ИСИМ" и удалить все строки, в которых она упоминается, то будут потеряны все данные о дисциплине «ЗИ» и частично о сотруднике «Алешкине» и т.п.

**Причина аномалии** - хранение в одном отношении разнородной информации (и о сотрудниках, и о дисциплинах, и о кафедрах).

**Вывод** - логическая модель данных неадекватна модели предметной области. БД, основанная на такой модели, будет работать неправильно.

**Нормализация таблиц** это формальный аппарат ограничений на формирование таблиц, описывающий разбиение таблиц на две или более части и обеспечивающий применение лучших методов добавление, изменения и удаления данных.

Можно также сказать, что нормализация – это процесс представления данных в виде простых двумерных таблиц, который позволяет устранить дублирование этих данных и обеспечивает непротиворечивость хранимой в базе информации. **>>***Цель нормализации* – получение проекта БД, в котором любая часть логически законченной информации хранится в одном месте, т.е. исключается избыточность информации.

Основа нормализации – аппарат нормализации отношений.

Всего существует 6 форм нормальных отношений. Но, как правило, необходимо и достаточно привести БД к третьей нормальной форме, чтобы исключить указанные аномалии при работе с БД.

Таблица считается нормализованной на определенном уровне, когда она удовлетворяет условиям, накладываемым соответствующей формой нормализации.

Формы нормализации:

-первая нормальная форма (First Normal Form – 1NF);

-вторая нормальная форма (Second Normal Form – 2NF);

-третья нормальная форма (Third Normal Form – 3NF);

-нормальная форма Бойса-Кодда (Boice-Codd Normal Form – BCNF);

-четвертая нормальная форма (Fourth Normal Form – 4NF);

-пятая нормальная форма или нормальная форма проекции-соединения (Fifth Normal Form – 5NF или PJ/NF);