

ЛАБОРАТОРНАЯ РАБОТА № 7 НОРМАЛИЗАЦИЯ БАЗЫ ДАННЫХ

Цель работы: изучить нормальные формы и привести существующую базу данных к третьей нормальной форме.

Теоретические сведения *Аномалии операций с данными*

Аномалии операций с данными (data operation anomalies) – некорректное выполнение операций с данными или возникновение побочных эффектов операций с данными, ставшее результатом нарушения требования адекватности базы данных предметной области.

Выделяют аномалии вставки, изменения и удаления. Рассмотрим на примерах каждую из них.

Аномалии вставки возникают в том случае, когда при добавлении информации в значение поля нам необходимо достраивать всю запись данными либо устанавливать вместо них значения NULL.

Добавляя в таблицу нового режиссера (name), нам необходимо заполнить поля с его местом и годом рождения либо написать там NULL. Очевидно, что запись, содержащая место и/или год рождения без указания режиссера, вообще теряет смысл, т. е. значение NULL в поле name неприемлемо (таблица 1).

Таблица 1 – Пример аномалии вставки в отношении Directors

id_d	name	place_of_birth	year_of_birth
1	Степанов Б.	Петропавловск	1927
2	Гайдай Л.	Свободный	1923
...
8	Тарковский А.	?	?
9	???????	?	1920

Аномалии удаления возникают в том случае, когда при ликвидации одних данных исчезают другие, которые изначально не предполагалось удалять.

Предположим, что все архивы начинают оцифровывать, информацию переносить на сервер, а записи на пленках удаляют из ячеек архивов. Например, оцифровали фильмы (и перенесли на сервер), созданные режиссером Э. Рязановым. Следовательно, из базы данных исчезнет информация (например, год выпуска) о всех картинах, снятых этим режиссером, а также информация о том, что в данном архиве есть полка под определенным номером, на которой расположена ячейка с определенным кодом (таблица 2).

Таблица 2 – Пример аномалии удаления в отношении Soviet_movies

archive	shelf	cod	film	director	year
1	782	10	Альпийская баллада	Степанов Б.	1965
1	782	15	Иван Васильевич меняет профессию	Гайдай Л.	1973
2	782	10	Служебный роман	Рязанов Э.	1977
3	310	4	Москва слезам не верит	Меньшов В.	1979
1	513	10	В бой идут одни «старики»	Быков Л.	1973
4	205	43	Обыкновенное чудо	Захаров М.	1978
2	205	43	Собачье сердце	Бортко В.	1988
2	782	72	Гараж	Рязанов Э.	1979

Аномалии изменения возникают в том случае, когда при обновлении значения некоторого атрибута в одной записи появляется необходимость его обновления и в других записях.

Например, архивы теперь не нумеруются, а обозначаются буквами латинского алфавита. То есть архив 1 становится архивом А, архив 2 – архивом В и т. д. Теперь нам необходимо обновить названия архивов во всех записях, не пропустив и не перепутав их новые заглавия (таблица 3):

Таблица 3 – Пример аномалии изменения в отношении Soviet_movies

archive	shelf	cod	film	director	year
A	782	10	Альпийская баллада	Степанов Б.	1965
1	782	15	Иван Васильевич меняет профессию	Гайдай Л.	1973
B	782	10	Служебный роман	Рязанов Э.	1977
C	310	4	Москва слезам не верит	Меньшов В.	1979
1	513	10	В бой идут одни «старики»	Быков Л.	1973
...

Получение отношений, не подверженных аномалиям вставки, удаления и изменения данных, реализуется, в частности, в процессе их нормализации.

Теория зависимостей

Заметим, что в теории баз данных можно встретить такие понятия, как «переменная отношения» и «отношение». Это различные понятия, которые часто отождествляют. Разница в том, что первое – это именно таблица в СУБД, второе – данные, хранящиеся в ней.

Функциональная зависимость. В значении переменной отношения r атрибут Y функционально зависит от атрибута X в том и только в том случае, если каждому значению X соответствует в точности одно значение Y . В этом случае говорят также, что атрибут X функционально определяет атрибут Y . Обозначается такая зависимость следующим образом: $r.X \rightarrow r.Y$.

Функциональную зависимость Y от X будем обозначать как $X \rightarrow Y$. X и Y могут представлять собой как единичные атрибуты, так и составные, полученные из нескольких атрибутов одного отношения.

Рассмотрим некоторые виды функциональных зависимостей, которые будут использоваться далее.

Полная функциональная зависимость – функциональная зависимость $X \rightarrow Y$, при которой Y не зависит функционально от любого подмножества X .

Например, в отношении Soviet_movies название фильма можно определить по полному адресу ячейки, в которой лежит пленка; если исключить, например, из адреса ячейки номер архива, то сказать с уверенностью, какой фильм содержится в ячейке хранения, не получится:

$$\{\text{archive, shelf, cod}\} \rightarrow \text{film}$$

Частичная функциональная зависимость – функциональная зависимость $X \rightarrow Y$, при которой Y зависит функционально от некоторого подмножества X .

Например, в отношении Soviet_movies для определения года выпуска фильма можно знать название фильма и режиссера. Если определить, что фильмы должны сниматься под уникальным названием, то год выпуска не зависит от имени режиссера, достаточно знать только название картины:

$$\{\text{film, director}\} \rightarrow \text{year}$$

Транзитивная функциональная зависимость – зависимость $X \rightarrow Y$, при которой существует такой атрибут Z , что имеются функциональные зависимости $X \rightarrow Z$ и $Z \rightarrow Y$ и отсутствует функциональная зависимость $Y \rightarrow X$. Например, в отношении Soviet_movies название фильма в ячейке можно узнать по адресу ячейки, а год выпуска фильма – по его названию:

$$\{\text{archive, shelf, cod}\} \rightarrow \text{film} \rightarrow \text{year}$$

Выделяют также многозначную зависимость, избыточные полную и частичную зависимости и др.

Нормальные формы

Вспомним одну из специальных реляционных операций – проекция. Рассмотрим ее следующим образом: допустим, что отношение R состоит из n атрибутов и k кортежей; если из данного отношения составить новое, взяв только m атрибутов ($m < n$) и все кортежи без дубликатов, то полученное отношение R_1 будет являться **проекцией** исходного отношения.

Нормализация – процесс декомпозиции переменной отношения R на набор проекций R_1, R_2, \dots, R_n , таких, что:

а) объединение проекций R_1, R_2, \dots, R_n позволяет гарантированно получить исходную переменную отношения R ;

б) каждая из проекций R_1, R_2, \dots, R_n является необходимой для выполнения условия «а»;

в) как минимум одна из проекций R_1, R_2, \dots, R_n находится в более высокой нормальной форме, чем исходная переменная отношения R .

Нормализация позволяет устранить дублирование данных в таблице, исключить аномалии вставки, удаления, изменения данных, борется с избыточностью данных.

Нормализация может рассматриваться как процесс поэтапного выполнения определенных условий до тех пор, пока дальнейшая декомпозиция получаемых на каждом этапе отношений станет невозможной.

Для реализации процесса нормализации отношения приводят к нормальным формам. В общей теории баз данных выделяют шесть нормальных форм, нормальную форму Бойса – Кодда и доменно-ключевую нормальную форму. Однако часто достаточно нормализовать отношения базы данных до третьей нормальной формы.

Сформулируем определения нормальных форм.

Первая нормальная форма (1НФ) предполагает приведение переменной отношения к виду, в котором все атрибуты отношения будут атомарны. Атомарность может определяться особенностями предметной области, через исключение многозначности данных, составных данных. Здесь следует исключить дублирование кортежей отношения (повторения строк (записей) в таблице).

Вторая нормальная форма (2НФ). Переменная отношения находится во 2НФ, если оно находится в 1НФ и не содержит частичных функциональных зависимостей неключевых* атрибутов от первичного ключа: неключевые атрибуты должны функционально полно зависеть от первичного ключа.

Очевидно, что 2НФ в этом случае будет достигнута, если первичный ключ – простой. Если он составной, то целесообразно добавить в отношение искусственный первичный ключ – счетчик (идентификатор (id)), а составной первичный ключ (который становится альтернативным ключом) вынести в отдельное отношение. Если же добавление id не предполагается по каким-то причинам, то следует добиться соответствующей полной функциональной зависимости.

Третья нормальная форма (3НФ). Переменная отношения находится в 3НФ, если оно находится во 2НФ и не содержит транзитивных функциональных зависимостей неключевых атрибутов от первичного ключа.

Например, нарушением 3НФ будет наличие вычисляемого столбца в таблице.

Нормальная форма Бойса – Кодда (НФБК) (частная форма 3НФ). Рассмотрим ситуацию, когда отношение содержит более одного потенциального ключа, два и более потенциальных ключа являются составными и имеют хотя бы один общий атрибут. Переменная отношения находится в нормальной форме Бойса – Кодда (НФБК) тогда и только тогда, когда она находится в 3НФ и не содержит пересекающихся потенциальных ключей.

* Неключевой атрибут – атрибут отношения, который не входит в состав ни одного потенциального (возможного) ключа этого отношения. К. Дейт определяет 2НФ и 3НФ при условии наличия в отношении только одного потенциального ключа, который и является первичным ключом отношения, а также неключевой атрибут, как атрибут, который не входит в состав первичного ключа рассматриваемой переменной отношения.

Порядок выполнения лабораторной работы

1. Определите вариант для выполнения заданий (приложение А) и выполните нормализацию до 3НФ соответствующего отношения (возможна нормализация и до высших нормальных форм – по желанию).
2. Изобразите графически структуру (диаграмму) полученной базы данных в любом редакторе (используйте правила наименования таблиц и их полей, сформулированные в лекции 6 - Даталогическое проектирование).
3. Оформите отчет.

Содержание отчета

1. Выполненные задания лабораторной работы.
2. Поэтапное описание нормализации исходной таблицы до 3НФ.
3. Графическое изображение структуры (диаграмма) итоговой базы данных.
4. Выводы по работе.

ПРИЛОЖЕНИЕ А

ЗАДАНИЯ ДЛЯ ЛАБОРАТОРНОЙ РАБОТЫ № 7

Таблица А.1 – Определение варианта для выполнения заданий

Номер варианта задания	Номер по списку в подгруппе		
1	1	6	11
2	2	7	12
3	3	8	13
4	4	9	14
5	5	10	15
Задание	Комментарий к заданию		
Задание 1. Рассмотреть аномалии при работе с соответствующим отношением	Привести пример аномалии вставки в отношении, предложить способы его модификации по устранению этой аномалии	Привести пример аномалии удаления в отношении, предложить способы его модификации по устранению этой аномалии	Привести пример аномалии обновления в отношении, предложить способы его модификации по устранению этой аномалии
Задание 2. На каждом этапе нормализации определить первичные ключи			
Задание 3. Нормализовать отношение до 3НФ:	Пояснить, каким образом производилась нормализация к 1НФ, 2НФ, 3НФ		
Задание 3.1. Нормализация до 1НФ	Указать многозначные и составные поля. Пояснить, что происходит с ними при нормализации		
Задание 3.2. Нормализация до 2НФ	Указать существующие частичные и полные функциональные зависимости		
Задание 3.3. Нормализация до 3НФ	Указать существующие транзитивные функциональные зависимости		
Задание 4. Проверить условия НФБК			

*Варианты выполнения заданий**

Вариант 1

Таблица А.2 – Студенты

ФИО преподавателя	Информация о преподавателе (должность, стаж работы, кафедра)	Дисциплина	Группа	ФИО старосты	Контакты старосты (номер телефона, e-mail)
Маленя И. С.	Ассистент, 4, философии	Философия	882513	Иванов И. И.	+555(71)2222222, 882513@mail.com
		Логика	962145	Петров П. Л.	962145@mail.com
Белкин А. М.	Доцент, 8, физики	Физика	202015	Сидорова Н. Г.	+872(54)42020152
		Физика	202015	Никулич П. Р.	+375(99)2222222, 202015@gmail.com
Зайцевич С. С.	Ассистент, 10, информатики	Базы данных	215636	Иванов И. М.	8(071)2222223, 215636@mail.by

Вариант 2

Таблица А.3 – Сессия

ФИО студента	Информация о студенте (группа, факультет, год поступления)	Академические задолженности по предметам	ФИО преподавателя, предмет, дата пересдачи	ФИО старосты	Контакты старосты (номер телефона, e-mail)
Булков А. П.	202015, ФКТ, 2020	Философия, ООП	Захаров П. К., философия, 12.02.2022; Крупеня А. Л., ООП, 18.02.2022	Сидорова Н. Г.	+872(54)42020152
Зимцевич С. С.	202015, ФКТ, 2018	ООП	Крупеня А. Л., ООП, 18.02.2022	Иванов И. М.	8(071)2222223, 215636@mail.by
Медведев Е. В.	708015, МТФ, 2017	Нет	Нет	Обухопич Н. В.	+742(59)2222223, 708015@mail.by

*Информация в таблицах не относится к действительности, все совпадения случайны.

Вариант 3

Таблица А.4 – Группы и кураторы

ФИО студента	Информация о студенте (группа, факультет, год поступления)	Оценки по всем предметам за сессию	Средний балл за сессию	ФИО куратора	Кафедра, рабочий телефон	Стаж работы в вузе
Новиков А. М.	202015, ФКТ, 2020	8, 6, 2, 3	4,75	Сидорова Н. Г.	ПОИТ, 223015	15
Купрович С. М.	202015, ФКТ, 2018	4, 4, 4, 2	3,5	Иванов И. М.	ВМСиС, 223018	12
Муштитч Е. В.	708015, МТФ, 2017	5, 6, 7, 8	6,5	Обухопич Н. В.	ПДиУ, 223014	7

Примечание – Средний балл за сессию для конкретного студента рассчитывается как среднее арифметическое его оценок по всем предметам за сессию.

Вариант 4

Таблица А.5 – Кафедры

Факультет	ФИО декана	Кафедры	ФИО зав. кафедрой	Количество преподавателей на кафедре	Количество остепененных преподавателей	Доля остепененных преподавателей на кафедре
ФКП	Никропов А. Н.	КИиКП	Иванов И. Д.	22	10	45 %
		ИТ	Петрович С. М.	21	9	43 %
		БД	Сидоренок П. Р.	15	5	33 %
ФКСиС	Зарембо Л. П.	ПИиКП	Казаков П. Д.	18	9	50 %
		ИТвЛ	Унсович Р. Т.	24	12	50 %
		ОИТ	Шпунчик П. Л.	17	7	41 %

Примечание – Доля остепененных преподавателей на кафедре – это отношение преподавателей, имеющих ученую степень, к их общему количеству. Выражается в процентах.

Вариант 5

Таблица А.6 – Товары

Наименование товара	Поставщики	Количество товара, шт.	Себестоимость, ден. ед.	Стоимость поставки, ден. ед.	Склад хранения (адрес, тип склада), ФИО зав. складом
Товар 1	ООО «Мороз-рыба»	500	30	15 000	Орловская обл., тип А, Кукушкина П. Р.
	ООО «Океан»	650	27	17 550	Челябинская обл., тип А, Звенюк Р. Д.
	ЗАО «Рыбхоз №28»	320	31	9920	Минская обл., тип В, Хлобич В. К.
Товар 2	ООО «Сладости для радости»	321	10	3210	Владимирская обл., тип А, Казаков П. Д.
	ЗАО «Вкусно есть»	850	9	7650	Гомельская обл., тип В, Унсович Р. Т.
	ООО «Конфетка»	1000	12	12 000	Пинский р-н, тип А, Шпунчик П. Л.

Примечание – Стоимость поставки рассчитывается как произведение количества товара на себестоимость.