

Нормализация отношений

Наумов Д.А., доц. каф. КТ

Базы данных и базы знаний, 2020

Содержание лекции

- 1 Типы приложений: транзакционная и аналитическая обработка
- 2 Нормализация

OLTP (On-Line Transaction Processing) — интерактивная транзакционная обработка.

- запросы пользователей выполняются одновременно - обработка идёт в режиме реального или приближенного к реальному времени;
- запросы представляют собой интенсивный поток коротких операций по вставке, изменению и удалению небольшого числа записей в БД;
- большая часть запросов известна на этапе проектирования;
- время выполнения сложных аналитических запросов не является критическим для системы;

Примеры OLTP-приложений:

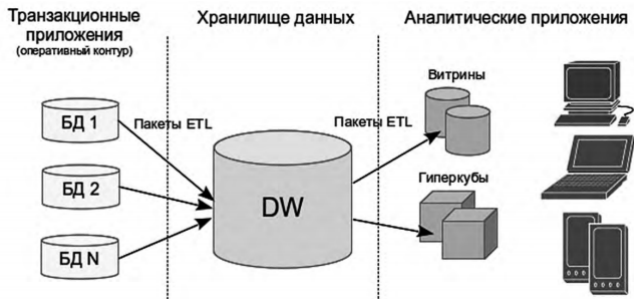
- системы складского учета;
- системы заказа билетов;
- банковские системы;

Данные OLTP-приложений **сильно нормализованы**.

OLAP (On-Line Analytical Processing) - интерактивная аналитическая обработка.

- Данные находятся в **режиме чтения**, за исключением моментов их обновления.
- Выборки представляют собой **одиночные тяжёлые запросы**: поиски и расчёты по множеству **произвольных** критериев.
- **Время отклика системы не регламентировано**.
- Размеры базы данных на порядок и больше транзакционных.

Пример типовой архитектуры OLAP



Нормализация

метод создания набора отношений с заданными свойствами на основе требований к данным, установленными в конкретной предметной области.

Цели нормализации:

- устранение избыточности при хранении данных, приводящей к увеличению размера БД.
- исключение необходимости модификации данных в связанных таблицах для минимизации времени и операций, проводящихся в одной транзакции.

Процесс нормализации:

- 1 исходная точка - представление предметной области в виде одного (или нескольких) отношений
- 2 на каждом шаге создается набор схем отношений, обладающих лучшими свойствами;
- 3 критерий «лучше-хуже» записит от целей проектирования.

Критерии оценки качества логической модели

Критерии оценки качества логической модели

- 1 Адекватность базы данных предметной области.
- 2 Скорость выполнения операций обновления данных (вставка, обновление, удаление кортежей).
- 3 Скорость выполнения операций выборки данных.
- 4 Легкость разработки и сопровождения базы данных.
- 5 Отсутствие неоправданной избыточности данных.

Критерии OLAP и OLTP:

- OLAP: на первый план выходит время отклика системы, данные могут быть избыточны.
- OLTP: на первый план выходит обработка транзакций в режиме реального времени.

Избыточность данных

одни и те же факты можно многократно получить из разных объектов базы данных

- польза: возможность ускорения выполнения запросов;
- избыточность должна быть контролируемой: необходима программная реализация проверок того, что избыточные и базовые данные адекватно согласованы между собой;

Расписание:

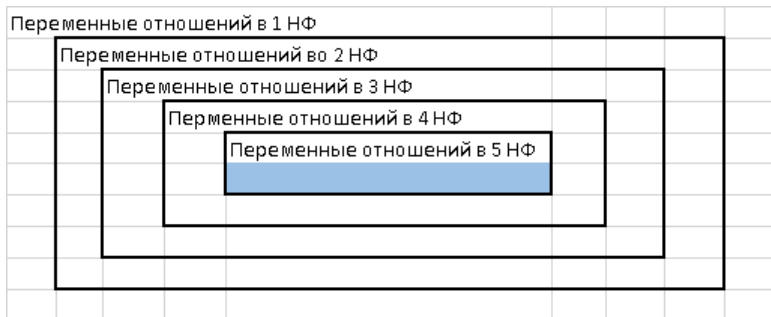
Преподаватель	Таб. №	Должность	День недели	Неделя	Время	Дисциплина	Вид занятия	Группа	Аудитория
Наумов Д.А.	2137	доц. каф. КТ	Понедельник		13:35	Базы данных	Лекция	648	22 БИ
Наумов Д.А.	2137	доц. каф. КТ	Понедельник		15:20	Базы данных	Лаб. раб.	648	22 БИ
Наумов Д.А.	2137	доц. каф. КТ	Понедельник	Числитель	17:05	Мультимедиа технологии	Лаб. раб.	648	22 БИ
Васильев Е.П.	4321	проф. каф. КТ	Понедельник	Знаменатель	9:55	Геоинформационные системы	Лекция	748	23 БИ

Должности:

Преподаватель	Таб. №	Должность	Кафедра	Ставка
Наумов Д.А.	2137	доцент	КТ	0,50
Наумов Д.А.	2137	доцент	ИТГД	0,70
Васильев Е.П.	4321	профессор	КТ	0,80

Основные свойства нормальных форм

- каждая следующая нормальная форма в некотором смысле улучшает свойства предыдущей;
- при переходе к следующей нормальной форме свойства предыдущих нормальных форм сохраняются;



Схемы БД называются эквивалентными

если содержание исходной БД может быть получено путем эквивалентного соединения отношений, входящих в результирующую схему, и при этом не появляется новых кортежей.

Исходная схема отношения:

Расписание(Таб.№, ФИО, Должность, Кафедра, День недели, Неделя, Время, Дисциплина, Вид занятий, Группа, Аудитория)

Таб. №	ФИО	Должность	Кафедра	День недели	Неделя	Время	Дисциплина	Вид занятия	Группа	Аудитория
2137	Наумов Д.А.	доц. каф. КТ	КТ	Понедельник		13:35	Базы данных	Лекция	648	22 БИ
2137	Наумов Д.А.	доц. каф. КТ	КТ	Понедельник		15:20	Базы данных	Лаб. раб.	648	22 БИ
2137	Наумов Д.А.	доц. каф. КТ	КТ	Понедельник	Числитель	17:05	Мультимедиа технологии	Лаб. раб.	648	22 БИ
4321	Васильев Е.П.	проф. каф. КТ	КТ	Понедельник	Знаменатель	9:55	Геоинформационные системы	Лекция	748	23 БИ

Эквивалентная схема, полученная путем декомпозиции на два отношения:

Расписание(Таб.№, День недели, Неделя, Время, Дисциплина, Вид занятий, Группа, Аудитория)

Таб. №	День недели	Неделя	Время	Дисциплина	Вид занятия	Группа	Аудитория
2137	Понедельник		13:35	Базы данных	Лекция	648	22 БИ
2137	Понедельник		15:20	Базы данных	Лаб. раб.	648	22 БИ
2137	Понедельник	Числитель	17:05	Мультимедиа технологии	Лаб. раб.	648	22 БИ
4321	Понедельник	Знаменатель	9:55	Геоинформационные системы	Лекция	748	23 БИ

Преподаватель(Таб.№, ФИО, Должность, Кафедра)

Таб. №	ФИО	Должность	Кафедра
2137	Наумов Д.А.	доцент	КТ
4321	Васильев Е.П.	профессор	КТ

1НФ - первая нормальная форма

выполняется, если все значения атрибутов (колонок таблицы) атомарны, то есть неделимы.

- собственные типы данных СУБД считаются атомарными, исключение могут составлять массивы, в том числе символьные (текстовые) и байтовые.
- атомарность может быть относительно выбранного взгляда со стороны предметной области и контекста.

Примеры:

- телефонный номер (в базе данных маркетинга, у телефонных операторов);
- колонки для хранения комментариев;
- целая и дробная части действительного числа, дата-время;
- фамилия, имя, отчество в одной колонке.

Преподаватель	Таб. №	Должность	Кафедра	Ставка	Телефоны
Наумов Д.А.	2137	доцент	КТ	0.50	+79539000000
			ИТГД	0.70	+79108009090
Васильев Е.П.	4321	профессор	КТ	0.80	+79109009000

Функциональная зависимость

Пусть R является переменной отношения, а X и Y — произвольными подмножествами множества атрибутов переменной отношения R .

$$X \rightarrow Y$$

Y функционально зависимо от X тогда и только тогда, когда для любого допустимого значения переменной отношения R , если два кортежа переменной отношения R совпадают по значению X , они также совпадают и по значению Y .

- Подмножество X - **детерминант**, а Y - **зависимая часть**.
- Функциональная зависимость **тривиальна** тогда и только тогда, когда её правая (зависимая) часть является подмножеством её левой части (детерминанта).
- Функциональная зависимость называется **неприводимой слева**, если ни один атрибут не может быть опущен из её детерминанта без нарушения зависимости.

Переменная отношения находится в 2НФ

тогда и только тогда, когда она находится в первой нормальной форме и каждый неключевой атрибут неприводимо зависит от (каждого) её потенциального ключа

- **Неприводимость:** в составе потенциального ключа отсутствует меньшее подмножество атрибутов, от которого можно также вывести данную функциональную зависимость.
- Если потенциальный ключ является составным, то в отношении не должно быть неключевых атрибутов, зависящих от части составного потенциального ключа.

Должность	Год	Ставка, час	Чтение лекций
ассистент	2019	880	нет
преподаватель	2019	820	нет
старший преподаватель	2019	820	да
доцент	2019	820	да
профессор	2019	760	да
заведующий кафедрой	2019	760	да
директор института	2019	700	да
ассистент	2018	900	нет
преподаватель	2018	840	нет
старший преподаватель	2018	840	да
доцент	2018	840	да
профессор	2018	780	да
заведующий кафедрой	2018	780	да
директор института	2018	780	да

Ставки(Должность, Год, Ставка, Чтение лекций)

Должность	Год	Ставка, час	Чтение лекций
ассистент	2019	880	нет
преподаватель	2019	820	нет
старший преподаватель	2019	820	да
доцент	2019	820	да
профессор	2019	760	да
заведующий кафедрой	2019	760	да
директор института	2019	700	да
ассистент	2018	900	нет
преподаватель	2018	840	нет
старший преподаватель	2018	840	да
доцент	2018	840	да
профессор	2018	780	да
заведующий кафедрой	2018	780	да
директор института	2018	780	да

Ключи и функциональные зависимости:

- ключ: Должность, Год
- функциональная зависимость: Должность, Год \rightarrow Ставка
- функциональная зависимость: Должность \rightarrow Чтение лекций

Нарушение второй нормальной формы: атрибут функционально зависит от части первичного ключа.

Декомпозиция:

- отношение Ставки(Должность, Год, Ставка)
- отношение Должности(Должность, Чтение лекций)

Переменная отношения находится в 3НФ

тогда и только тогда, когда для каждой из её функциональных зависимостей $X \rightarrow A$ выполняется хотя бы одно из следующих условий:

- X содержит A (то есть $X \rightarrow A$ - тривиальная функциональная зависимость);
- X - суперключ (не существует двух кортежей, в которых значения атрибутов X совпадают);
- A - ключевой атрибут (A входит в состав потенциального ключа).

Пример: продажа каждой товарной позиции имеет своим основанием документ (заказ, счёт и т.д.), а её стоимость характеризуется ценой, количеством и валютой.

Транзитивные зависимости:

- Идентификатор продажи \rightarrow Номер документа
- Идентификатор продажи \rightarrow Код валюты
- Номер документа \rightarrow Код валюты

Пример: **Студент**(№ зачетки, ФИО, Группа, Факультет, Специальность, Выпускающая кафедра)

Функциональные зависимости

- № зачетки → ФИО, Группа, Факультет, Специальность, Выпускающая кафедра
- Группа → Факультет, Специальность, Выпускающая кафедра
- Выпускающая кафедра → Факультет

Итоговые отношения

- **Студенты** (№ зачетки, ФИО, Группа)
- **Группы** (Группа, Специальность, Выпускающая кафедра)
- **Выпускающие кафедры** (Выпускающая кафедра, Факультет)

Переменная отношения находится в НФБК

тогда и только тогда, когда каждая её нетривиальная и неприводимая слева функциональная зависимость имеет в качестве своего детерминанта некоторый потенциальный ключ

- НФБК - нормальная форма Бойса-Кодда;
- менее строгое определение: детерминанты всех её функциональных зависимостей являются потенциальными ключами.

Ситуация, когда отношение будет находиться в 3НФ, но не в НФБК:

- отношение имеет два (или более) потенциальных ключа, которые являются составными
- между отдельными атрибутами таких ключей существует функциональная зависимость

Поскольку описанная зависимость не является транзитивной, то такая ситуация под определение 3НФ не подпадает.

Пример: Планирование занятий

Аудитория	Время начала	Время окончания	Планирование занятий
260	08:10	09:45	Аудитория с проектором для студентов младших курсов
260	09:55	11:30	Аудитория с проектором для студентов младших курсов
260	11:40	13:15	Аудитория с проектором для студентов младших курсов
260	13:35	15:10	Аудитория с проектором для студентов старших курсов
260	15:20	16:55	Аудитория с проектором для студентов старших курсов
260	17:05	18:40	Аудитория с проектором для студентов старших курсов
21 БИ	08:10	09:45	Класс ПЭВМ, занятие для младших курсов
21 БИ	09:55	11:30	Класс ПЭВМ, занятие для младших курсов
21 БИ	11:40	13:15	Класс ПЭВМ, занятие для младших курсов
21 БИ	13:35	15:10	Класс ПЭВМ, занятие для старших курсов

Свойства отношения:

- все атрибуты входят в какой-то из потенциальных ключей, неключевых атрибутов в отношении нет (2НФ).
- нет транзитивных зависимостей (3НФ).
- функциональная зависимость Планирование занятий \rightarrow Аудитория, в которой левая часть (детерминант) не является потенциальным ключом отношения, то есть отношение не находится в нормальной форме Бойса — Кодда.

Исходное отношение:

Аудитория	Время начала	Время окончания	Планирование занятий
260	08:10	09:45	Аудитория с проектором для студентов младших курсов
260	09:55	11:30	Аудитория с проектором для студентов младших курсов
260	11:40	13:15	Аудитория с проектором для студентов младших курсов
260	13:35	15:10	Аудитория с проектором для студентов старших курсов
260	15:20	16:55	Аудитория с проектором для студентов старших курсов
260	17:05	18:40	Аудитория с проектором для студентов старших курсов
21 БИ	08:10	09:45	Класс ПЭВМ, занятие для младших курсов
21 БИ	09:55	11:30	Класс ПЭВМ, занятие для младших курсов
21 БИ	11:40	13:15	Класс ПЭВМ, занятие для младших курсов
21 БИ	13:35	15:10	Класс ПЭВМ, занятие для старших курсов

Нормализованные отношения:

Аудитория	Время начала	Время окончания
260	08:10	09:45
260	09:55	11:30
260	11:40	13:15
260	13:35	15:10
260	15:20	16:55
260	17:05	18:40
21 БИ	08:10	09:45
21 БИ	09:55	11:30
21 БИ	11:40	13:15
21 БИ	13:35	15:10

Аудитория	Планирование занятий
260	Аудитория с проектором для студентов младших курсов
260	Аудитория с проектором для студентов старших курсов
21 БИ	Класс ПЭВМ, занятие для младших курсов
21 БИ	Класс ПЭВМ, занятие для старших курсов

Многозначная зависимость

Пусть существует некоторое отношение r со схемой R , а также два произвольных подмножества атрибутов $A, B \subseteq R$.

Пусть $C = R \setminus (A \cup B)$.

В этом случае B многозначно зависит от A , тогда и только тогда, когда множество значений атрибута B , соответствующее заданной паре $[a : A; c : C]$ отношения r , зависит от a и не зависит от c .

4НФ

Переменная отношения R находится в четвёртой нормальной форме, если она находится в НФБК и все нетривиальные многозначные зависимости фактически являются функциональными зависимостями от её потенциальных ключей.

Отношение Дисциплины студентов(Студент, Группа, Дисциплина)

Студент	Группа	Дисциплина
Иванов	648	Экспертные системы и искусственный интеллект
Петров	648	Экспертные системы и искусственный интеллект
Сидоров	648	Экспертные системы и искусственный интеллект
Иванов	648	Базы данных и базы знаний
Петров	648	Базы данных и базы знаний
Сидоров	648	Базы данных и базы знаний
Иванов	648	Операционные системы и системное ПО
Петров	648	Операционные системы и системное ПО
Сидоров	648	Операционные системы и системное ПО
Костин	748	Мультимедиа технологии
Семенов	748	Мультимедиа технологии

- переменная отношения не соответствует 4НФ, так как существует следующая многозначная зависимость:
 - $\{Группа\} \twoheadrightarrow \{Студент\}|\{Дисциплина\}$
- при добавлении нового студента в группу придется ему добавлять список изучаемых его группой дисциплин.

Нормализованные отношения: Группы(Группа, Студент)

Студент	Группа
Иванов	648
Петров	648
Сидоров	648
Костин	748
Семенов	748

Дисциплины(Группа, Дисциплина)

Группа	Дисциплина
648	Экспертные системы и искусственный интеллект
648	Базы данных и базы знаний
648	Операционные системы и системное ПО
748	Мультимедиа технологии

Если к исходной переменной отношения добавить атрибут, функционально зависящий от потенциального ключа, например оценку по дисциплине (Студент, Группа, Дисциплина) → Оценка, то полученное отношение будет находиться в 4НФ и его уже нельзя подвергнуть декомпозиции без потерь.

Декомпозицией отношения R

называется замена R на совокупность отношений $\{R_1, R_2, \dots, R_n\}$ такую, что каждое из них есть проекция R , и каждый атрибут R входит хотя бы в одну из проекций декомпозиции.

Пример: для отношения R с атрибутами (a, b, c) существуют следующие варианты декомпозиции:

- $(a), (b), (c)$
- $(a), (b, c)$
- $(a, b), (c)$
- $(b), (a, c)$
- $(a, b), (b, c)$
- $(a, b), (a, c)$
- $(b, c), (a, c)$
- $(a, b), (b, c), (a, c)$

Рассмотрим отношение R' , которое получается в результате операции естественного соединения (NATURAL JOIN), применённой к отношениям, полученным в результате декомпозиции R .

Декомпозиция R называется декомпозицией без потерь если R' в точности совпадает с R .

Пример: исходное отношение:

Студент	Группа	Является старостой
Иванов	648	да
Петров	648	нет
Сидоров	648	нет
Костин	748	да
Семенов	748	нет

Результат декомпозиции:

Студент	Группа	Группа	Является старостой
Иванов	648	648	да
Петров	648	648	нет
Сидоров	648	648	нет
Костин	748	748	да
Семенов	748	748	нет

Пример: исходное отношение:

Студент	Группа	Является старостой
Иванов	648	да
Петров	648	нет
Сидоров	648	нет
Костин	748	да
Семенов	748	нет

Результат естественного соединения:

Студент	Группа	Является старостой
Иванов	648	да
Иванов	648	нет
Петров	648	да
Петров	648	нет
Сидоров	648	да
Сидоров	648	нет
Костин	748	да
Костин	748	нет
Семенов	748	да
Семенов	748	нет

Пример 2: исходное отношение:

Студент	Группа	Является старостой
Иванов	648	да
Петров	648	нет
Сидоров	648	нет
Костин	748	да
Семенов	748	нет

Результат декомпозиции:

Студент	Группа	Студент	Является старостой
Иванов	648	Иванов	да
Петров	648	Петров	нет
Сидоров	648	Сидоров	нет
Костин	748	Костин	да
Семенов	748	Семенов	нет

Результат естественного соединения:

Студент	Группа	Является старостой
Иванов	648	да
Петров	648	нет
Сидоров	648	нет
Костин	748	да
Семенов	748	нет

Декомпозиция R называется декомпозицией без потерь

Пусть R — переменная отношения, а A, B, \dots, Z — некоторые подмножества множества её атрибутов. Если декомпозиция любого допустимого значения R на отношения, состоящие из множеств атрибутов A, B, \dots, Z , является декомпозицией без потерь, говорят, что переменная отношения R удовлетворяет зависимости соединения $*(A, B, \dots, Z)$

- понятие зависимости соединения определено не для отношения (конкретного значения), а для переменной отношения.
- зависимость соединения определяется не механически по текущим значениям, а следует из внешнего знания о природе и закономерностях данных, которые могут находиться в переменной отношения.

Зависимость соединения $*\{A, B, \dots, Z\}$ является тривиальной

когда по крайней мере одно из подмножеств A, B, \dots, Z является множеством всех атрибутов отношения.

Отношение находится в 5НФ (в проекционно-соединительной нормальной форме)

тогда и только тогда, когда каждая нетривиальная зависимость соединения в нём определяется потенциальным ключом (ключами) этого отношения

- зависимость соединения $*\{A, B, \dots, Z\}$ определяется потенциальным ключом (ключами) тогда и только тогда, когда каждое из подмножеств A, B, \dots, Z множества атрибутов является суперключом отношения.
- суперключ - подмножество атрибутов отношения, удовлетворяющее требованию уникальности: не существует двух кортежей данного отношения, в которых значения этого подмножества атрибутов совпадают (равны).

Пример 3: исходное отношение:

Преподаватель	Должность	Дисциплина	Вид занятия
Наумов Д.А.	доцент	Базы данных	Лекция
Наумов Д.А.	доцент	Базы данных	Упражнения
Шилина О.И.	старший преподаватель	Цифровая обработка изображений	Лекция
Шилина О.И.	старший преподаватель	Цифровая обработка изображений	Упражнения
Наумов Д.А.	доцент	Цифровая обработка изображений	Лекция
Наумов Д.А.	доцент	Цифровая обработка изображений	Упражнения
Наумов Д.А.	доцент	Цифровая обработка изображений	Курсовой проект
Логачева О.А.	ассистент	Рисунок	Упражнения
Корсаков В.В.	профессор	Рисунок	Упражнения

Результат декомпозиции:

Должность	Вид занятия
профессор	Лекция
профессор	Упражнения
профессор	Курсовой проект
доцент	Лекция
доцент	Упражнения
доцент	Курсовой проект
старший преподаватель	Лекция
старший преподаватель	Упражнения
ассистент	Упражнения

Преподаватель	Дисциплина
Наумов Д.А.	Базы данных
Наумов Д.А.	Цифровая обработка изображений
Шилина О.И.	Цифровая обработка изображений
Логачева О.А.	Рисунок
Корсаков В.В.	Рисунок

Преподаватель	Должность
Наумов Д.А.	доцент
Шилина О.И.	старший преподаватель
Логачева О.А.	ассистент
Корсаков В.В.	профессор

Дисциплина	Вид занятия
Базы данных	Лекция
Базы данных	Упражнения
Цифровая обработка изображений	Лекция
Цифровая обработка изображений	Упражнения
Цифровая обработка изображений	Курсовой проект
Рисунок	Упражнения

Отношение находится в 6НФ

когда оно удовлетворяет всем нетривиальным зависимостям соединения (то есть не может быть подвергнуто дальнейшей декомпозиции без потерь)

Сотрудники (Таб.№, Должность, Кафедра, Ставка, Время, Адрес)

Таб. №	Должность	Кафедра	Ставка	Время	Адрес
2137	доцент	КТ	0.25	01.02.2018	г. Рязань, ул. Молодцова, д. X, кв. Y
2137	доцент	ИТГД	0.60	01.09.2018	г. Рязань, ул. Молодцова, д. X, кв. Y
2137	доцент	ИТГД	0.75	01.09.2019	г. Рязань, ул. Бирюзова, д. Z, кв. W

Для хронологических баз данных максимально возможная декомпозиция позволяет бороться с избыточностью и упрощает поддержание целостности базы данных.

Исходное отношение:

Сотрудники (Таб.№, Должность, Кафедра, Ставка, Время, Адрес)					
Таб. №	Должность	Кафедра	Ставка	Время	Адрес
2137	доцент	КТ	0.25	01.02.2018	г. Рязань, ул. Молодцова, д. X, кв. Y
2137	доцент	ИТГД	0.60	01.09.2018	г. Рязань, ул. Молодцова, д. X, кв. Y
2137	доцент	ИТГД	0.75	01.09.2019	г. Рязань, ул. Бирюзова, д. Z, кв. W

Результат декомпозиции:

Сотрудники (Таб.№, Должность, Кафедра, Ставка, Время)				
Таб. №	Должность	Кафедра	Ставка	Время
2137	доцент	КТ	0.25	01.02.2018
2137	доцент	ИТГД	0.60	01.09.2018
2137	доцент	ИТГД	0.75	01.09.2019

Адреса сотрудников (Таб. №, Время, Адрес)		
Таб. №	Время	Адрес
2137	01.02.2018	г. Рязань, ул. Молодцова, д. X, кв. Y
2137	01.09.2019	г. Рязань, ул. Бирюзова, д. Z, кв. W

- теория нормализации является очень ценным достижением реляционной теории и практики, поскольку она даёт научно строгие и обоснованные критерии качества проекта БД и формальные методы для усовершенствования этого качества;
- идеи нормализации полезны для проектирования баз данных, они отнюдь не являются универсальным или исчерпывающим средством повышения качества проекта БД - существует слишком большое разнообразие возможных ошибок и недостатков в структуре БД, которые нормализацией не устраняются;
- во всей сфере информационных технологий практически отсутствуют методы оценки и улучшения проектных решений, сопоставимые с теорией нормализации реляционных баз данных по уровню формальной строгости.