

# 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:

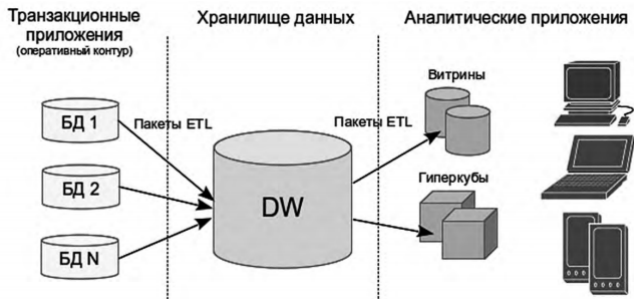
- системы складского учета;
- системы заказа билетов;
- банковские системы;

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:

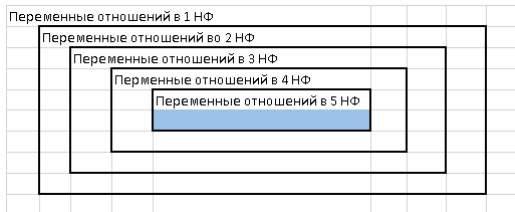
Преподаватель	Таб. №	Должность	День недели	Неделя	Время	Дисциплина	Вид занятия	Группа	Аудитория
Наумов Д.А.	2137	доц. каф. КТ	Понедельник		13:35	Базы данных	Лекция	648	22 БИ
Наумов Д.А.	2137	доц. каф. КТ	Понедельник		15:20	Базы данных	Лаб. раб.	648	22 БИ
Наумов Д.А.	2137	доц. каф. КТ	Понедельник	Числитель	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.



## 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.

Пример 1: телефонный номер (в базе данных маркетинга, у телефонных операторов), колонки для хранения комментариев, целая и дробная части действительного числа, дата-время.

Пример 2: фамилия, имя, отчество в одной колонке.

## Пример

Препода- ватель	День недели	Номер пары	Название дисциплины	Тип занятий	Группа
Петров В. И.	Понед.	1	Теор.выч.проц.	Лекция	4906
	Вторник	1	Комп.графика	Лаб.раб.	4907
	Вторник	2	Комп.графика	Лаб.раб.	4906
Киров В. А.	Понед.	2	Теор.информ.	Лекция	4906
	Вторник	3	Пр-е на C++	Лаб.раб.	4907
	Вторник	4	Пр-е на C++	Лаб.раб.	4906
Серов А. А.	Понед.	3	Защита информ.	Лекция	4944
	Среда	3	Пр-е на VB	Лаб.раб.	4942
	Четверг	4	Пр-е на VB	Лаб.раб.	4922

## Пример

Препода- ватель	День недели	Номер пары	Название дисциплины	Тип занятий	Группа
Петров В. И.	Понед.	1	Теор.выч.проц.	Лекция	4906
Петров В. И.	Вторник	1	Комп.графика	Лаб.раб.	4907
Петров В. И.	Вторник	2	Комп.графика	Лаб.раб.	4906
Киров В. А.	Понед.	2	Теор.информ.	Лекция	4906
Киров В. А.	Вторник	3	Пр-е на C++	Лаб.раб.	4907
Киров В. А.	Вторник	4	Пр-е на C++	Лаб.раб.	4906
Серов А. А.	Понед.	3	Защита информ.	Лекция	4944
Серов А. А.	Среда	3	Пр-е на VB	Лаб.раб.	4942
Серов А. А.	Четверг	4	Пр-е на VB	Лаб.раб.	4922

## 2НФ - вторая нормальная форма

означает, что выполнены требования 1НФ, при этом все атрибуты целиком зависят от составного ключа и не зависят ни от какой его части.

- в определении говорится о ключах вообще, а не только о первичных.
- в отношении может быть несколько ключей, и некоторые из них могут являться составными

Пример 1: Описание продаж товара.

Пример 2: Сдача сессии (ФИО, №Зачетки, Группа, Дисциплина, Оценка).

Пример 3: Ключ - атрибут не выделенной еще сущности.

## 3НФ - вторая нормальная форма

означает, что выполнены требования 2НФ, при этом в между атрибутами отношения нет транзитивных зависимостей.

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 → Номер документа
- Идентификатор продажи → Код валюты
- Номер документа → Код валюты

Результатом нарушения 3НФ является избыточность хранения и необходимость обновления данных в связанной таблице.

Пример: (ФИО, Номер зачетки, Группа, Факультет, Специальность, Выпускающая кафедра)

## Пример

Номер зачетки - > ФИО

Номер зачетки - > Группа

Номер зачетки - > Факультет

Номер зачетки - > Специальность

Номер зачетки -> Выпускающая кафедра

Группа - > Факультет

Группа -> Специальность

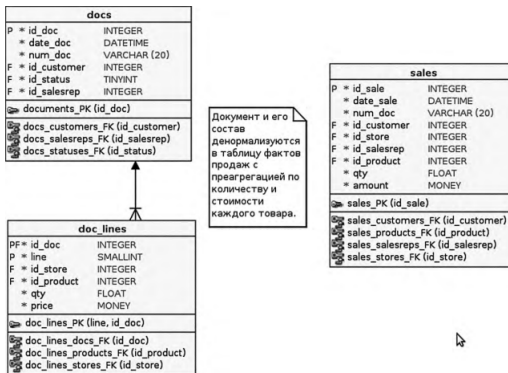
Группа - > Выпускающая кафедра

Выпускающая кафедра - > Факультет

## Допустимые схемы для OLAP: "снежинка" "звезда":

- центральным элементом являются таблицы фактов, содержащие события, транзакции, документы и др.
- в таблице фактов одному документу (каждой его строке), соответствует одна запись.

## Денормализация документов в таблицу фактов



## Пример SQL-запроса в транзакционной СУБД

```
SELECT
    SUM(dl.qty) AS total qty, SUM(dl.price) AS total amount,
    c.name
FROM
    docs d
    INNER JOIN doc_lines dl ON d.id_doc = dl.id_doc
    INNER JOIN customers c ON d.id_customer = c.id_customer
    INNER JOIN products p ON dl.id_product = p.id_product
WHERE
    c.name IN ('ООО Пирожки', 'ЗАО Ватрушки') AND
    p.name = 'Мука' AND
    d.date BETWEEN '2014-01-01' AND '2014-02-01'
GROUP BY c.name
```

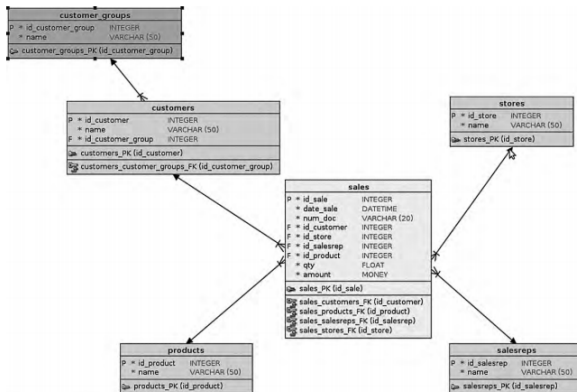
## Пример SQL-запроса в OLAP-СУБД

```
SELECT
    SUM(s.qty) AS total_qty, SUM(s.amount) AS total_amount,
    c.name
FROM
    sales s
    INNER JOIN customers c ON d.id_customer = c.id_customer
    INNER JOIN products p ON dl.id_product = p.id_product
WHERE
    c.name IN ('ООО Пирожки', 'ЗАО Ватрушки') AND
    p.name = 'Мука' AND
    s.date BETWEEN '2014-01-01' AND '2014-02-01'
GROUP BY c.name
```



Если таблица фактов ссылается на таблицы-измерения, имеющие ссылки на другие измерения, то такая схема называется "снежинка".

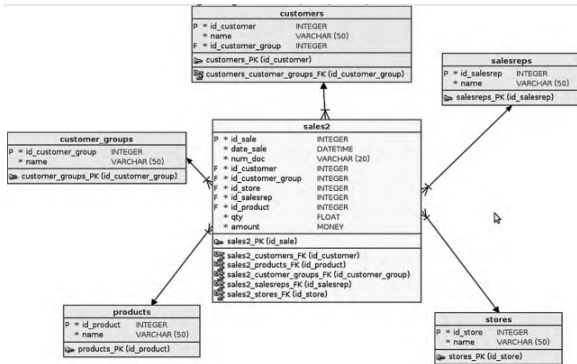
Таблица фактов в схеме "снежинка"



Для запросов, включающих фильтрацию по группам клиентов, приходится делать дополнительное соединение.

Схема "Звезда" полностью исключает иерархию измерений и необходимость соединения соответствующих таблиц в одном запросе.

Таблица фактов в схеме "звезда"



Обратной стороной денормализации всегда является избыточность, являющаяся причиной увеличения размера БД как в случае транзакционных, так и аналитических приложений.

## Пример SQL-запроса в схеме "снежинка"

```
SELECT sum(amount)
FROM sales s
WHERE s.id_customer_group IN (1, 2, 10, 55)
```

## Пример SQL-запроса в схеме "звезда"

```
SELECT sum(amount)
FROM sales s
      INNER JOIN customers c ON s.id_customer = c.id_customer
WHERE c.id_customer_group IN (1, 2, 10, 55)
```

## Задача

Посчитаем, примерную дельту на приведённом выше примере преобразования "снежинки" в "звезду" если таблица продаж не использует компрессию данных и содержит около 500 миллионов строк, а количество групп покупателей порядка 1000.

- **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 - что 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.