



Лекция 4 даталогическое проектирование. Нормальные формы БД.



Составитель: доц. Космачева И.М.

ПРОЕКТИРОВАНИЕ РЕЛЯЦИОННЫХ БД

При проектировании БД решаются 2 основные проблемы:

1. **Оптимальное отображение объектов предметной области в абстрактные объекты модели данных.**
2. **Обеспечение эффективного выполнения запросов к БД в среде конкретной СУБД.**

Необходимо принять решение:

- Из каких отношений должна состоять БД ?
- Какие атрибуты должны быть у этих отношений ?



ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

- Логическая модель данных учитывает особенности выбранной модели организации данных в целевой СУБД (например, реляционная).
- На этом этапе игнорируются остальные характеристики выбранной СУБД, например, любые особенности физической организации ее структур хранения данных и построения индексов.
- Для проверки правильности логической модели данных используется метод *нормализации*.



НОРМАЛИЗАЦИЯ

□ **Нормализация** - процесс реорганизации данных путем ликвидации повторяющихся групп и иных противоречий в хранении данных с целью приведения таблиц к виду, позволяющему осуществлять непротиворечивое и корректное редактирование данных.

□ **Нормальная форма** - совокупность требований, которым должно удовлетворять отношение.

□ Управление данными становится простым, если данные организованы согласно правилам нормализации- **правилам Кодда.**



НОРМАЛИЗАЦИЯ 1 НФ

Таблица 1. Состав семьи (правило 1 нарушается)

п/п	Член семьи	Степень родства	Иждивенцы
1	Борис	Глава семьи	Екатерина, Александра
2	Тамара	Мать	Нет
3	Петр	Сводный брат	Ирина
4	Михаил	Дядя	Анна, Наталья
5	Иван	Дядя	Елена

Таблица 2. Состав семьи (в первой нормальной форме)

п/п	Член семьи	Степень родства	Иждивенец1	Иждивенец2
1	Борис	Глава семьи	Екатерина	Александра
2	Тамара	Мать	Нет	Нет
3	Петр	Сводный брат	Ирина	Нет
4	Михаил	Дядя	Анна	Наталья
5	Иван	Дядя	Елена	Нет

НОРМАЛИЗАЦИЯ 2НФ

Определение 1. Пусть R - отношение. Множество атрибутов Y функционально зависимо от множества атрибутов X (X функционально определяет Y) тогда и только тогда, когда во всех кортежах, имеющих одинаковые значения атрибутов X , значения атрибутов Y также совпадают в любом состоянии отношения R . Символически функциональная зависимость записывается

$$X \rightarrow Y$$

Множество атрибутов X называется детерминантом функциональной зависимости, а множество атрибутов Y называется зависимой частью.

- Таблица содержит ключ
- Все неключевые столбцы зависят от полного ключа

Составной ключ

Курс	Студент	Тип обучения	Оценка	ак. часов
10774	Е. Онегин	очно-заочное	отлично	40
10775	В. Ленский	вебинар	хорошо	40
10776	Т. Ларина	очное	отлично	40
10777	В. Ленский	очное	хорошо	40

НОРМАЛИЗАЦИЯ 3 НФ

- В фокусе внимания – неключевые столбцы
- Нет транзитивных зависимостей
 - Неключевые столбцы не зависят от других неключевых столбцов

№ заказа	Покупатель	Товар	Цена (шт.)	Кол-во	Сумма
1001	Е. Онегин	телефон	100	2	200
1002	В. Ленский	ноутбук	2000	1	2000
1003	Т. Ларина	утюг	50	1	50
1004	В. Ленский	утюг	50	2	100



НОРМАЛИЗАЦИЯ НФ БОЙСА-КОДДА (НФБК)

- Отношение R находится в НФ Б-К тогда и только тогда, когда **детерминанты всех функциональных зависимостей являются потенциальными ключами.**
- Если отношение находится в НФБК, то оно автоматически находится и в ЗНФ.



ПРИВЕДЕНИЕ К 1 НФ

Накладная № 123				
Дата	Покупатель	Адрес		
10.01.2001	ТОО "Суперпулс"	г. Кукуюевск ул. Большая Трубная д.8		
Отпущен товар	Количество	Ед. Изм.	Цена за ед.	Общая стоимость
Банка стеклянная	100	шт.	3,45	345
Стакан граненый	34	шт.	1,34	45,56
Бутылка "чебурашка"	367	шт.	0,45	165,15
Вода минеральная	40	б		314
Водка "Столичная"	25	б		262,5
Пиво "Амур ДВ"	40			182,4

ОТПУСК ТОВАРОВ СО СКЛАДА
Дата
Покупатель
Город
Адрес
Товар
Ед_Измерения
Цена_за_ед
Отпущено_ед
Общая_стоимость
Номер_накладной

Составитель: ЛОП. Космачева И.М.



ПРИВЕДЕНИЕ К 2 НФ

Накладная № 123

<u>Дата</u>	<u>Покупатель</u>	<u>Адрес</u>
10.01.2001	ООО "Суперпулс"	г. Кукуевск ул. Большая Трубная д.6

Отпущен товар	Количество	Ед. Изм.	Цена за ед.	Общая стоимость
Банка стеклянная	100	шт.	3,45	345
Стакан граненый	34	шт.	1,34	45,56
Бутылка "чебурашка"	367	шт.	0,45	165,15
Вода минеральная	40	бутылка	7,85	314
Водка "Столичная"	25	бутылка	10,50	262,5
Пиво "Амур ДВ"	40	банка	4,56	182,4

Составитель: ЛОП. Космачева И.М.

ОТПУСК ТОВАРОВ СО СКЛАДА

Дата
Покупатель
Город
Адрес
Товар
Ед_Измерения
Цена за ед
Отпущено_ед
Общая_стоимость
Номер_накладной



ОТПУСК ТОВАРОВ СО СКЛАДА

Номер_накладной
Товар
Дата
Покупатель
Город
Адрес
Ед_Измерения
Цена за ед
Отпущено_ед
Общая_стоимость



ПРИВЕДЕНИЕ К 2 НФ

ОТПУСК ТОВАРОВ СО СКЛАДА
Номер_накладной
Товар
Дата
Покупатель
Город
Адрес
Ед_Измерения
Цена за ед
Отпущено_ед
Общая_стоимость



Выделение таблицы "Товар"

ОТПУСК ТОВАРОВ СО СКЛАДА
Номер_накладной
Товар (FK)
Дата
Покупатель
Город
Адрес
Отпущено_ед
Общая_стоимость

ТОВАРЫ
Товар
Ед_Измерения
Цена за ед

2 НФ



ПРИВЕДЕНИЕ К 3 НФ



ТОВАРЫ
товар
ед-измерения
цена за ед

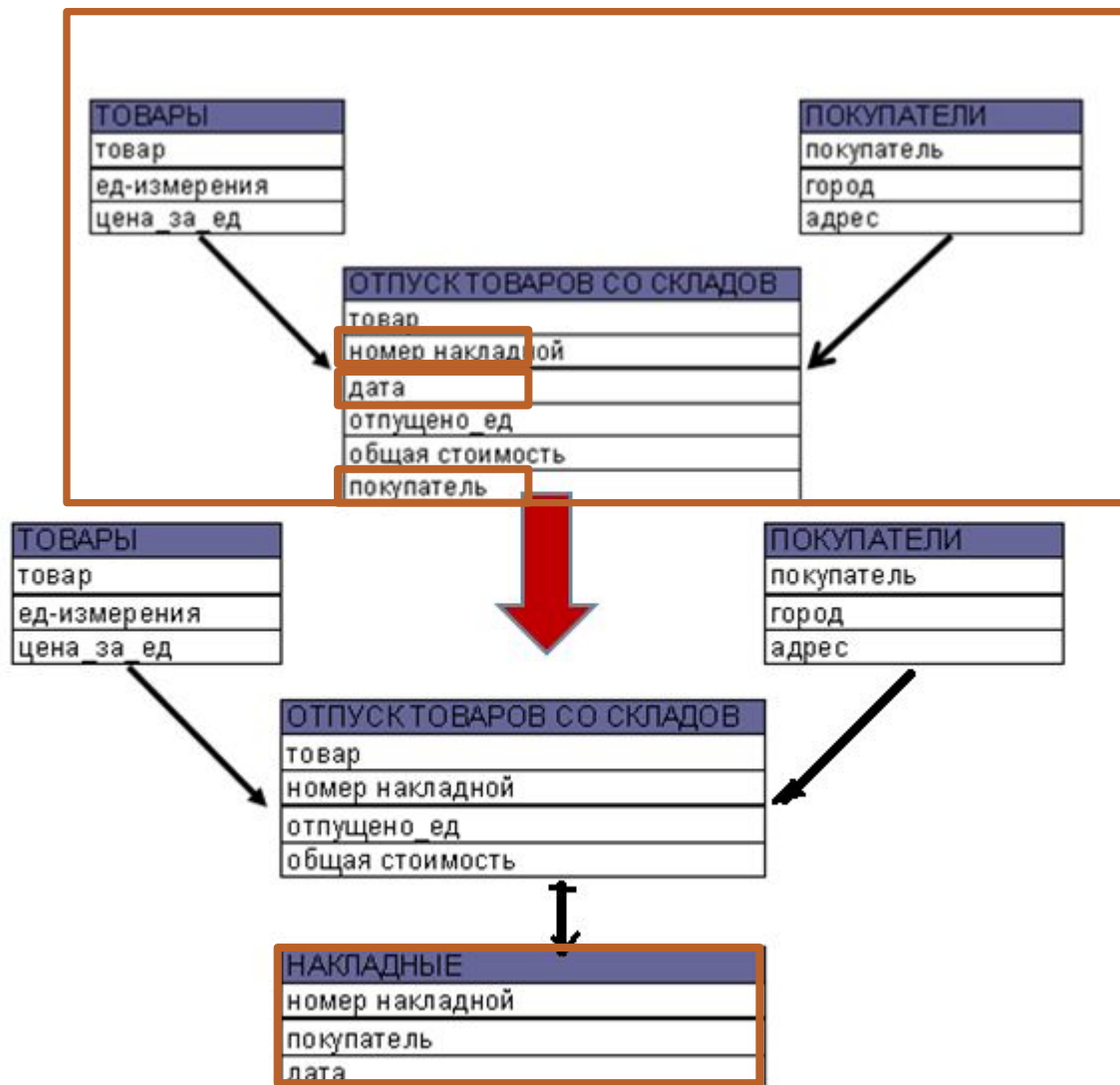


ПОКУПАТЕЛИ
покупатель
город
адрес

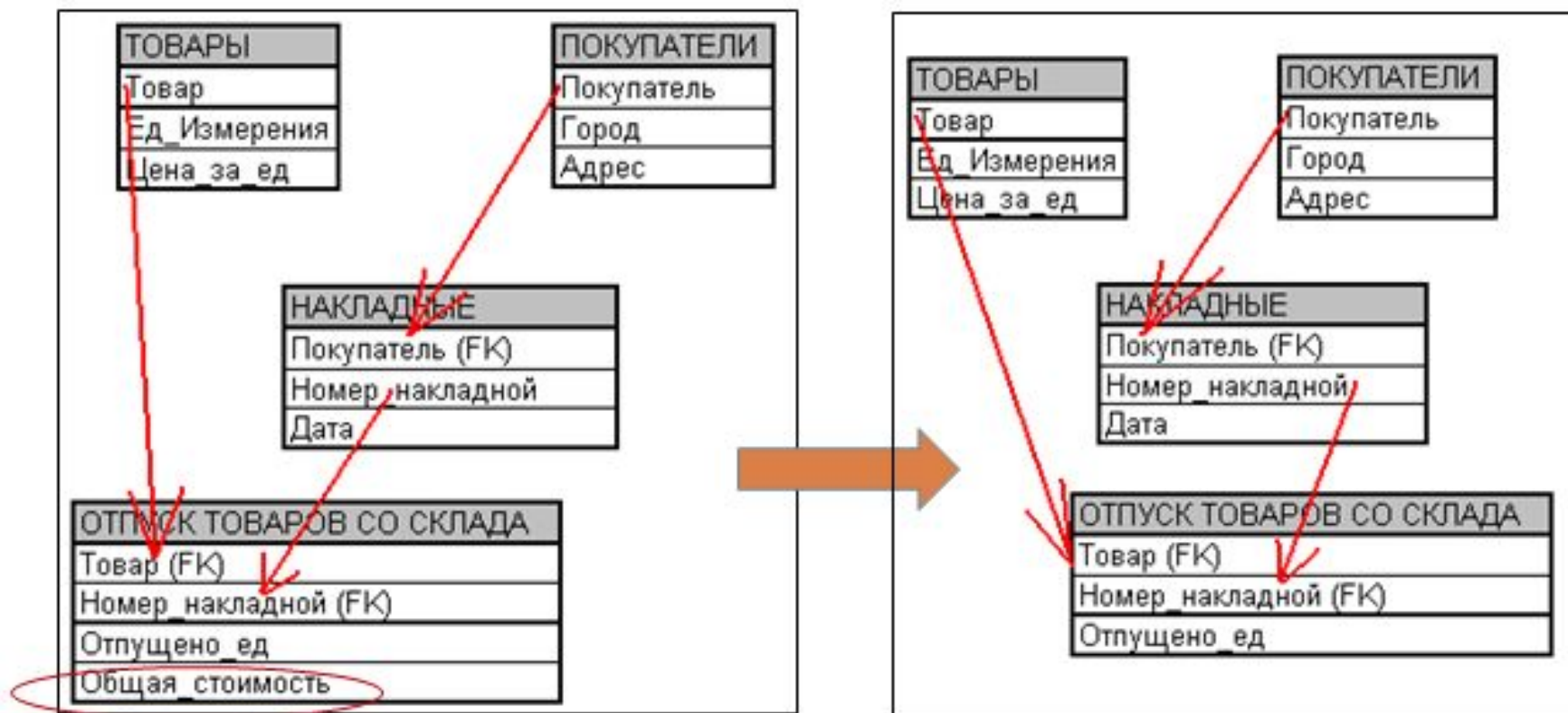
ОТПУСК ТОВАРОВ СО СКЛАДОВ
товар
номер накладной
дата
отпущено_ед
общая стоимость
покупатель



ПРИВЕДЕНИЕ К 2 НФ



ПРИВЕДЕНИЕ К 3 НФ



ДОБРОВОЛЬНОЕ МЕДИЦИНСКОЕ СТРАХОВАНИЕ



НЕДОСТАТКИ НОРМАЛИЗАЦИИ

- Больше количество сущностей БД. Сопровождение и поддержка такой БД сложна.
- Трудности построения запросов к таким БД, так как необходимо связывать несколько таблиц.
- Оперативность **выборки данных низкая для высоко нормализованных БД (3 НФ).**



OLTP И OLAP-СИСТЕМЫ

- Сильно нормализованные модели данных хорошо подходят для OLTP-приложений (**On-Line Transaction Processing (OLTP)**- оперативная обработка транзакций)
- OLAP-приложения (**On-Line Analytical Processing (OLAP)** - оперативная аналитическая обработка данных) используют слабо нормализованные модели данных



ОЛТР-ПРИЛОЖЕНИЯ

- Поддерживает большое число пользователей, работающих параллельно.
- Большое значение имеет время ответа на запрос.
- ОЛТР-системы сопряжены с интенсивными процессами чтения-записи.
- Примеры ОЛТР-приложений - системы складского учета, системы заказов билетов, банковские системы, выполняющие операции по переводу денег, и т.п.



OLAP-ПРИЛОЖЕНИЯ

- Оперировать с большими массивами данных.
- Добавление в систему новых данных происходит относительно редко крупными блоками (например, раз в квартал загружаются данные по итогам квартальных продаж из OLTP-приложения).
- Добавленные данные в систему обычно никогда не удаляются.
- Перед загрузкой данные проходят различные процедуры "очистки", связанные с тем, что в одну систему могут поступать данные из многих источников, имеющих различные форматы.
- Запросы к системе являются нерегламентированными, достаточно сложными. Скорость выполнения запросов важна, но не критична.



ФИЗИЧЕСКОЕ ПРОЕКТИРОВАНИЕ БД

- Проектирование базовых отношений в среде целевой СУБД, отношений, содержащих производные данные.
- Реализация ограничений предметной области.
- Проектирование физического представления БД.
- Анализ транзакций.
- Выбор файловой структуры.
- Определение индексов.



ФИЗИЧЕСКОЕ ПРОЕКТИРОВАНИЕ БД

- Определение требований к дисковой памяти.
- Разработка пользовательских представлений.
- Анализ необходимости введения контролируемой избыточности.
- Организация мониторинга и настройка функционирования ОС.
- Разработка средств и механизмов защиты.



ФИЗИЧЕСКОЕ ПРОЕКТИРОВАНИЕ БД

- **выбор типа носителя, способа организации данных, методов доступа (определение пользователей базы данных, их уровней доступа, разработка и внедрение правил безопасности доступа),**
- **определение размеров физического блока, управление размещением данных на внешнем носителе,**
- **управление свободной памятью, определение целесообразности сжатия данных и используемых методов сжатия,**



ФИЗИЧЕСКОЕ ПРОЕКТИРОВАНИЕ БД

- оценка размеров объектов базы (определение размеров табличных пространств и особенностей их размещения на носителях информации,
- определение спецификации носителей информации для промышленной системы (например, тип RAID-массивов, их количество),

RAID-массивы это высокопроизводительные , устойчивые к отказам подсистемы ввода-вывода, это технология для расширения пропускной способности системы ввода/вывода и обеспечения возможности хранения избыточных данных.

- разработка топологии базы данных в случае распределенной базы данных, определение механизмов доступа к удаленным данным.



СПАСИБО ЗА ВНИМАНИЕ.

Составитель: доц. Космачева И.М.

