Базы данных

1 лекция

Определение базы данных

What is database?

A collection of information organized to afford efficient retrieval

http://www.usg.edu/galileo/skills/unit04/primer04_01.phtml

Определение базы данных

Jim Grey , “The Fourth Paradigm”

When people use the word database, fundamentally what they are saying is that data should be self-describing and it should have a schema. That’s really all the world database means.

РСУБД

Relational Database Management Systems were invented to let you use one set of data in multiple ways, including ways that are unforeseen at the time the database is built and the 1st applications are written.”

(Curt Monash, anaylst/blogger)

Реляционная модель

Основана на реляционной алгебре (но не полностью)

Реляционная модель данных (РМД) — логическая модель данных, прикладная теория построения баз данных, которая является приложением к задачам обработки данных таких разделов математики, как теория множеств и логика первого порядка

Реляционная модель

- Для лучшего понимания РМД следует отметить три важных обстоятельства:

- модель является логической, то есть отношения являются логическими (абстрактными), а не физическими (хранимыми) структурами;
- для реляционных баз данных верен информационный принцип: всё информационное наполнение базы данных представлено одним и только одним способом, а именно — явным заданием значений атрибутов в кортежах отношений; в частности, нет никаких указателей (адресов), связывающих одно значение с другим;
- наличие реляционной алгебры позволяет реализовать декларативное программирование и декларативное описание ограничений целостности, в дополнение к навигационному (процедурному) программированию и процедурной проверке условий.
- Принципы реляционной модели были сформулированы в 1969—1970 годах Э. Ф. Коддом (E. F. Codd). Идеи Кодда были впервые публично изложены в статье «A Relational Model of Data for Large Shared Data Banks, ставшей классической

Обзор SQL

Рис. 1.1. Применение SQL для обращения к базе данных

Обзор SQL

Функциональные возможности, которые СУБД предоставляет пользователю:

- Определение данных
- Выборка данных
- Обработка данных
- Управление доступом
- Совместное использование данных
- Целостность данных

Обзор SQL

SQL это **не полноценный язык программирования** типа C, C++, Java и т.п.

SQL – это подъязык баз данных, в который входит около 40 инструкций.

Инструкции SQL могут быть встроены в другой язык, такой как C или Java

SQL – **декларативный** язык программирования

SQL – стандарт для работы с РСУБД

Роль SQL

Рис. 1.2. Компоненты типичной системы управления базой данных

Преимущества SQL

- Независимость от конкретных СУБД
- Межплатформенная переносимость
- Наличие стандартов
- Поддержка со стороны вендоров
- Построена на реляционной модели
- Высокоуровневая структура, напоминающая естественный язык

Преимущества SQL

- Возможность выполнения специализированных структурных запросов
- Обеспечение программного доступа к базам данных
- Возможность различного представления баз данных
- Полноценность в качестве языка, предназначенного для работы с БД
- Возможность динамического определения данных

Преимущества SQL

- Поддержка архитектуры клиент/сервер
- Поддержка приложений уровня предприятия
- Расширяемость и поддержка объектно-ориентированных технологий
- Возможность доступа к данным в Интернете
- Интеграция с языком Java (JDBC)
- Поддержка открытого кода
- Промышленная инфраструктура

Развитие СУБД

- Системы управления файлами
- Иерархические базы данных
- Сетевые базы данных
- Реляционная модель данных
- Объектно-ориентированные СУБД
- Объектно-реляционные СУБД
- NO-SQL

Современные СУБД

Основные функции СУБД:

- Управление данными во внешней памяти(на дисках)
- Управление данными в оперативной памяти с использованием дискового кэша
- Журнализация изменений, резервное копирование и восстановление БД после сбоев
- Поддержка языков БД (язык определения данных, язык манипулирования данными DML)

Современные СУБД

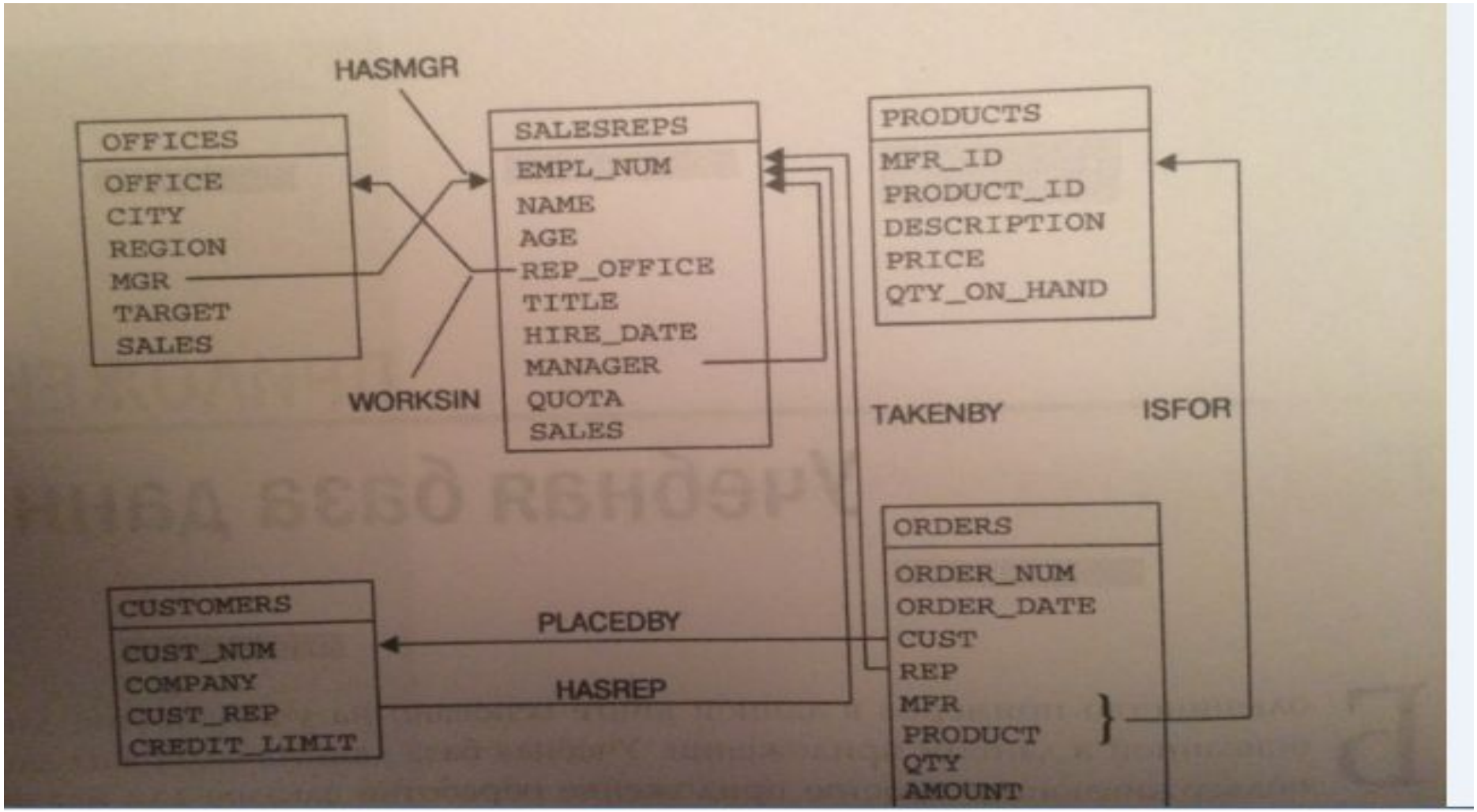
Состав СУБД:

- Ядро; отвечает за управление данными во внешней и оперативной памяти и журнализацию
- Процессор языка БД, обеспечивающий оптимизацию запросов на извлечение и изменение данных и создание, как правило, машинно-независимого исполняемого внутреннего кода
- Подсистему поддержки времени исполнения,

которая интерпретирует программы манипуляции данными, создающий пользовательский интерфейс с СУБД

- Сервисные программы (внешние утилиты), обеспечивающие ряд дополнительных возможностей по обслуживанию информационной системы.

Учебная база данных



Учебная база данных

Таблица SALESREPS

EMPL_NUM – уникальный номер

NAME – полное имя служащего

AGE – возраст

REP_OFFICE – уникальный номер офиса

TITLE – название должности

HIRE_DATE – дата найма

MANAGER – уникальный номер менеджера

QUOTA – план продаж на год

SALES – объем продаж с начала года

Учебная база данных

Таблица PRODUCTS:

MFR_ID – уникальный ключ фабрики

PRODUCT_ID – уникальный ключ фабрики

DESCRIPTION – описание

PRICE – цена

QTY_IN_HAND - количество

Учебная база данных

Таблица ORDERS:

ORDER_NUM – уникальный номер заказа

ORDER_DATE – дата заказа

CUST – уникальный номер заказчика

REP – уникальный номер сотрудника

MFR – уникальный номер фабрики

PRODUCT – уникальный номер продукта

QTY – количество в заказе

AMOUNT – сумма заказа

Учебная база данных

Таблица OFFICES:

OFFICE – уникальный номер офиса

CITY – город

REGION – регион (США)

MGR – уникальный номер менеджера офиса

TARGET – план офиса за год

SALES – продажи офиса за год

Учебная база данных

Таблица Customers:

CUST_NUM – уникальный номер заказчика

COMPANY – название компании

CUST_REP – уникальный номер менеджера для данной компании

CREDIT_LIMIT – лимит кредита для компании

Таблицы

- Первичные ключи
- Взаимоотношения (предок-потомок)
- Внешние ключи

Основы SQL

Инструкции

- Около 40 инструкций

Например:

- SELECT
- INSERT
- UPDATE
- MERGE
- DELETE

Структура инструкции SQL

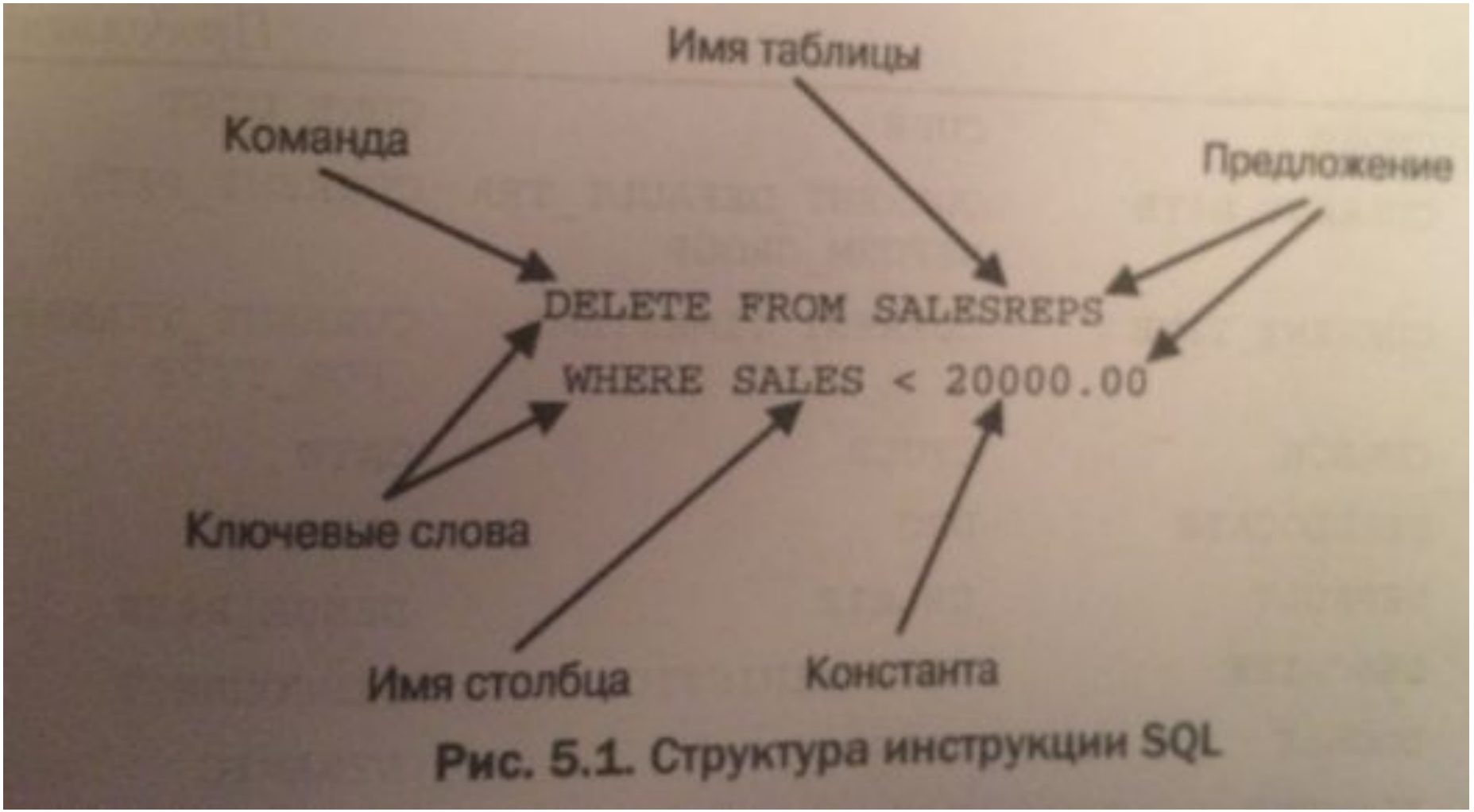


Рис. 5.1. Структура инструкции SQL

Типы данных

- CHAR
- VARCHAR
- NCHAR
- CLOB
- NCLOB
- INT
- SMALLINT
- NUMERIC (точность, масштаб)
- DECIMAL (точность, масштаб)

Типы данных

- FLOAT
- DATE
- TIME
- DATE
- XML

NULL

AND	TRUE	FALSE	UNKNOWN
TRUE	TRUE	FALSE	UNKNOWN
FALSE	FALSE	FALSE	FALSE
UNKNOWN	UNKNOWN	FALSE	UNKNOWN

OR	TRUE	FALSE	UNKNOWN
TRUE	TRUE	TRUE	TRUE
FALSE	TRUE	FALSE	UNKNOWN
UNKNOWN	TRUE	UNKNOWN	UNKNOWN

NOT	
TRUE	FALSE
FALSE	TRUE
UNKNOWN	UNKNOWN

Простые запросы

Минимум:

Стандарт:

SELECT <набор значений>

FROM <название таблицы>

Часть СУБД:

SELECT <набор значений>

Простые запросы

Вывести список офисов с их планами и фактическими объемами продаж

Простые запросы

Вывести список офисов с их планами и фактическими объемами продаж

```
SELECT CITY, TARGET, SALES
FROM OFFICES;
```

Простые запросы

Вычисляемые столбцы

Выдать для каждого офиса список городов, регионов и сумм, на которые был перевыполнен/недовыполнен план

Простые запросы

Вычисляемые столбцы

Выдать для каждого офиса список городов, регионов и сумм, на которые был перевыполнен/недовыполнен план

- Преобразовать в посл-ть действий:

1. Взять таблицу с офисами
2. Выбрать оттуда города, регионы и рассчитать разность сумм

Простые запросы

Вычисляемые столбцы

Выдать для каждого офиса список городов, регионов и сумм, на которые был перевыполнен/недовыполнен план

```
SELECT CITY, REGION,
(SALES - TARGET) AS diff_sales_and_target
FROM OFFICES;
```

Простые запросы

Вычисляемые столбцы:

Показать общую стоимость по каждому товару

Простые запросы

Вычисляемые столбцы:

Показать общую стоимость по каждому товару, продукт и его описание

Шаги:

- Выбрать таблицу по продуктам
- Вывести идентификатор продукта (а чем он является), описание продукта, и общую стоимость (как ее рассчитать)?

Простые запросы

Вычисляемые столбцы:

Показать общую стоимость по каждому товару, продукт и его описание

```
SELECT MFR_ID, PRODUCT_ID, DESCRIPTION,
(PRICE * QTY_ON_HAND) AS "count product"
FROM Products;
```

Простые запросы

Вычисляемые столбцы:

Что получится, если увеличить плановый объем продаж для каждого служащего на 3% от его фактического объема продаж?

Простые запросы

Вычисляемые столбцы:

Что получится, если увеличить плановый объем продаж для каждого служащего на 3% от его фактического объема продаж?

Шаги:

1. Какая таблица нужна?
2. Какие столбцы вывести?
3. Какие столбцы нужны для вычисляемого значения и как его вычислить?

Простые запросы

Константы:

Список объемов продаж для каждого города

```
SELECT CITY, 'has sales of', SALES
FROM OFFICES;
```

Простые запросы

Дополнительные возможности

- SELECT *

Простые запросы

Дополнительные возможности

```
SELECT *
FROM OFFICES;
```

Простые запросы

Дополнительные возможности

- SELECT *
- DISTINCT

```
SELECT MGR
FROM OFFICES;
```

Простые запросы

Дополнительные возможности

- SELECT *
- DISTINCT

```
SELECT DISTINCT MGR
FROM OFFICES;
```

Удаление дублей

Простые запросы

Отбор строк (WHERE):

- Сравнение (=, <>, <, <=, >, >=)
- Проверка на принадлежность диапазону
- Проверка наличия во множестве
- Проверка на соответствие шаблону
- Проверка на равенство значению NULL

Простые запросы

Сравнение:

Найти имена всех служащих, принятых на работу до 2006 года

Простые запросы

Сравнение:

Найти имена всех служащих, принятых на работу до 2006 года

Шаги:

- Выбрать таблицу со служащими
- Применить фильтр на дату приема
- Выбрать нужные нам строки

Простые запросы

Сравнение:

Найти имена всех служащих, принятых на работу до 2006 года

```
SELECT NAME
FROM SALESREPS
WHERE HIRE_DATE < TO_DATE('01.01.2016', 'dd.mm.yyyy');
```

ВАЖНО: Необходимо крайне аккуратно работать с датами

Простые запросы

Сравнение с использованием вычисляемых столбцов:

Вывести список офисов, фактические объемы продаж в которых оставили менее 80 процентов от плановых:

Шаги:

1. Выбрать таблицу с офисами
2. Понять какой фильтр необходим
3. Применить этот фильтр
4. Выбрать необходимые столбцы

Простые запросы

Сравнение с использованием вычисляемых столбцов:

Вывести список офисов, фактические объемы продаж в которых оставили менее 80 процентов от плановых:

```
SELECT CITY, SALES, TARGET
FROM OFFICES
WHERE SALES < (0.8 * TARGET);
```

Простые запросы

Сравнения. Значения NULL.

Сравним

```
SELECT NAME
FROM SALESREPS
```

и два других запроса

```
SELECT NAME
FROM SALESREPS
WHERE SALES <= QUOTA
```

И

```
SELECT NAME
FROM SALESREPS
WHERE SALES > QUOTA
```

Простые запросы

Сравнения. Проверка на принадлежность диапазону.

Найти все заказы, сделанные в последнем квартале 2007 года.

Простые запросы

Сравнения. Проверка на принадлежность диапазону.

Найти все заказы, сделанные в последнем квартале 2007 года.

Шаги:

1. Какая таблица используется?
2. Какие заказы необходимо оставить?
3. Как это перевести на язык сравнений?
4. Применить фильтр
5. Выбрать необходимые поля

Простые запросы

Сравнения. Проверка на принадлежность диапазону.

Найти все заказы, сделанные в последнем квартале 2007 года.

```
SELECT ORDER_NUM, ORDER_DATE, MFR, PRODUCT, AMOUNT
FROM ORDERS
WHERE ORDER_DATE BETWEEN TO_DATE('2007.10.01', 'yyyy.mm.dd') AND TO_DATE('2007.12.31', 'yyyy.mm.dd');
```

Простые запросы

Вывести список служащих, фактические объемы продаж которых не попадают в диапазон от 80 до 120 процентов плана.

Простые запросы

Вывести список служащих, фактические объемы продаж которых не попадают в диапазон от 80 до 120 процентов плана.

```
SELECT NAME, SALES, QUOTA
FROM SALESREPS
WHERE SALES NOT BETWEEN (.8 * QUOTA) AND (1.2 * QUOTA);
```

Простые запросы

Сравнения. Проверка наличия во множестве.

Вывести список служащих, которые работают в Нью-йорке, Атланте или Денвере

Шаги:

1. Какая таблица?
2. Что за условие ?
3. Как его применить?
4. Добавить фильтр
5. Выделить строки

Простые запросы

Сравнения. Проверка наличия во множестве.

Вывести список служащих, которые работают в Нью-йорке, Атланте или Денвере

```
SELECT NAME, QUOTA, SALES
FROM SALESREPS
WHERE REP_OFFICE IN (11, 13, 22);
```

Простые запросы

Проверка на соответствие шаблону. LIKE.

Я помню, что у нас был клиентом компания начинающаяся на Smith и затем со вторым словом Corp. Можешь посмотреть полное название? И еще я хочу знать какой у них кредитный лимит.

Простые запросы

Проверка на соответствие шаблону. LIKE.

Я помню, что у нас был клиентом компания начинающаяся на Smith и затем со вторым словом Corp. Можешь посмотреть полное название? И еще я хочу знать какой у них кредитный лимит.

```
SELECT Company, Credit_limit
FROM Customers
WHERE Company LIKE 'Smith% Corp.';
```

Простые запросы

Проверка на соответствие шаблону. LIKE.

Я помню, что у нас был клиентом компания то ли Smithsen, то ли Smithson. Мне необходимо знать ее кредитный лимит.

Простые запросы

Проверка на соответствие шаблону. LIKE.

Я помню, что у нас был клиентом компания то ли Smithsen, то ли Smithson. Мне необходимо знать ее кредитный лимит.

```
SELECT Company, Credit_limit
FROM CUSTOMERS
WHERE COMPANY LIKE 'Smiths_n Corp.';
```

Простые запросы

Проверка на соответствие шаблону. LIKE.

Найти товары, коды которых начинаются с четырех букв ‘A%BC’

```
SELECT PRODUCT_ID
FROM products
WHERE PRODUCT_ID LIKE 'A$%BC%';
```

Простые запросы

Проверка на соответствие шаблону. LIKE.

Найти товары, коды которых начинаются с четырех букв ‘A%BC’

```
SELECT ORDER_NUM, PRODUCT
FROM ORDERS
WHERE PRODUCT LIKE 'A$%BC%' ESCAPE '$';
```

Простые запросы

Проверка на равенство NULL

Найти служащего, который еще не закреплен за офисом

Простые запросы

Проверка на равенство NULL

Найти служащего, который еще не закреплен за офисом.

```
SELECT NAME
FROM SALESREPS
WHERE REP_OFFICE IS NULL;
```

Также может быть IS NOT NULL

Простые запросы

Составные условия отбора (AND, OR и NOT)

Найти служащих, у которых фактический объем продаж меньше планового и меньше 300 000$

Простые запросы

Составные условия отбора (AND, OR и NOT)

Найти служащих, у которых фактический объем продаж меньше планового и меньше 300 000$

```
SELECT NAME, QUOTA, SALES
FROM SALESREPS
WHERE SALES < QUOTA
  AND SALES < 300000.00;
```

Простые запросы

Составные условия отбора (AND, OR и NOT)

Найти всех служащих, которые работают в Денвере, Нью-йорке или Чикаго (их номера 22, 11, 12) или не имеют менеджера и были приняты на работу после июня 2006 года; или у которых продажи превысили плановый объем, но не превысили 600 000$

Простые запросы

Составные условия отбора (AND, OR и NOT)

Найти всех служащих, которые работают в Денвере, Нью-йорке или Чикаго (их номера 22, 11, 12) или не имеют менеджера и были приняты на работу после июня 2006 года; или у которых продажи превысили плановый объем, но не превысили 600 000$

```
SELECT NAME
FROM SALESREPS
WHERE (REP_OFFICE IN (22, 11, 12))
  OR (MANAGER IS NULL AND HIRE_DATE >=
TO_DATE('2006.06.01', 'yyyy.mm.dd'))
  OR (SALES > QUOTA AND NOT SALES > 600000)
```

Простые запросы

Сортировка результатов запроса

Вывести список офисов, отсортированный по фактическим объемам продаж в порядке убывания.

Простые запросы

Сортировка результатов запроса

Вывести список офисов, отсортированный по фактическим объемам продаж в порядке убывания.

```
SELECT CITY, REGION, SALES
FROM OFFICES
ORDER BY SALES DESC;
```

Простые запросы

Сортировка результатов запроса.

Вывести список всех офисов отсортированных по разности между фактическими и плановыми объемами продаж в порядке убывания.

Простые запросы

Сортировка результатов запроса.

Вывести список всех офисов отсортированных по разности между фактическими и плановыми объемами продаж в порядке убывания.

```
SELECT CITY, REGION, (SALES - TARGET)
FROM OFFICES
ORDER BY 3 DESC; не желательно
```

Или

```
SELECT CITY, REGION, (SALES - TARGET)
FROM OFFICES
ORDER BY (SALES - TARGET) DESC;
```

Или

```
SELECT CITY, REGION, (SALES - TARGET) AS diff
FROM OFFICES
ORDER BY diff DESC;
```

Простые запросы

Группировка. GROUP BY.

SUM([DISTINCT] значение)

AVG([DISTINCT] значение)

MIN(значение)

MAX(значение)

COUNT([DISTINCT] значение)

COUNT(*)

Простые запросы

Группировка.

Какой наибольший процент выполнения плана среди всех служащих?

Простые запросы

Группировка.

Какой наибольший процент выполнения плана среди всех служащих?

```
SELECT MAX( 100 * (SALES/QUOTA)) AS
max_plan_complete
  FROM SALESREPS;
```

Простые запросы

Группировка.

Вычислить среднюю цену товаров от производителя ACI

Простые запросы

Группировка.

Вычислить среднюю цену товаров от производителя ACI.

```
SELECT AVG(PRICE) AS avg_price
FROM Products
WHERE MFR_ID = 'ACI';
```

Простые запросы

Группировка. Статистические функции и значения NULL

```
SELECT COUNT(*), COUNT(SALES), COUNT(QUOTA)
FROM SALESREPS;
```

Статистическая функция COUNT() игнорирует все значения NULL, содержащиеся в столбцах

Простые запросы

Группировка. Статистические функции и значения NULL

```
SELECT COUNT(*), COUNT(SALES), COUNT(QUOTA)
FROM SALESREPS;
```

Статистическая функция COUNT() игнорирует все значения NULL, содержащиеся в столбцах

Простые запросы

Группировка. Статистические функции и значения NULL.

```
SELECT SUM(SALES), SUM(QUOTA), (SUM(SALES) – SUM(QUOTA)), SUM(SALES-QUOTA)
FROM SALESREPS;
```

ВЫВОД: Необходимо следить за обработкой NULL в случаях с статистическими функциями

Простые запросы

Группировка. Запросы с GROUP BY

Какова средняя стоимость заказа для каждого служащего?

Простые запросы

Группировка. Запросы с GROUP BY

Какова средняя стоимость заказа для каждого служащего?

```
SELECT REP, AVG(AMOUNT) AS avg_AMOUNT
FROM ORDERS
GROUP BY REP;
```

Простые запросы

Группировка. Запросы с GROUP BY

Подсчитать общую сумму заказов по каждому клиенту для каждого служащего

Простые запросы

Группировка. Запросы с GROUP BY

Подсчитать общую сумму заказов по каждому клиенту для каждого служащего

```
SELECT REP, CUST, SUM(AMOUNT) AS sum_amount
FROM ORDERS
GROUP BY REP, CUST;
```

Простые запросы

Группировка. Запросы с GROUP BY

Подсчитать общую сумму заказов по каждому клиенту для каждого служащего и каждого клиента

Простые запросы

Группировка. Запросы с GROUP BY

Подсчитать общую сумму заказов по каждому клиенту для каждого служащего и каждого клиента

```
SELECT REP, CUST, SUM(AMOUNT) AS sum_AMOUNT
FROM ORDERS
GROUP BY CUBE (REP, CUST);
```

Простые запросы

Группировка. Запросы с GROUP BY

Какова средняя стоимость заказа для каждого служащего из числа тех, у которых общая стоимость заказов превышает 30000$?

Простые запросы

Группировка. Запросы с GROUP BY

Какова средняя стоимость заказа для каждого служащего из числа тех, у которых общая стоимость заказов превышает 30000$?

```
SELECT REP, AVG(AMOUNT) AS avg_AMOUNT
FROM ORDERS
GROUP BY REP
HAVING SUM(AMOUNT) > 30000;
```