Кафедра ИВС

ЛЕКЦИЯ

Тема: «Основные понятия реляционных баз данных. Реляционная алгебра»

Дисциплина «Базы данных в ИС»

ОП «Информационные системы»

Авторы: ст.преп.каф.ИВС Саданова Б.М.

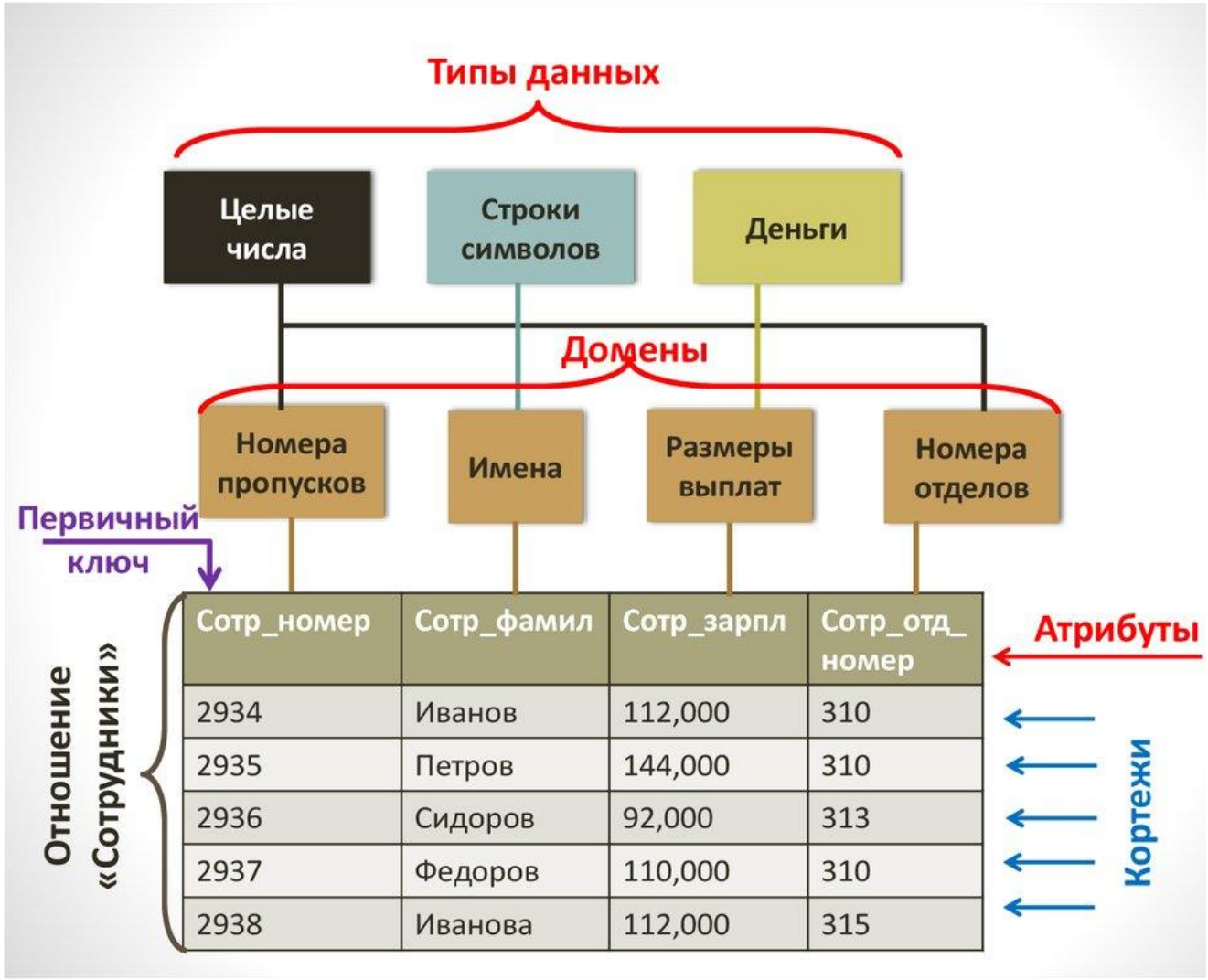
ст.преп.каф.ИВС Олейникова А.В.

План лекции:

1. Базовые понятия реляционных БД
2. Фундаментальные свойства отношений
3. Реляционная алгебра
4. Теоретико-множественные операции
5. Специальные реляционные операции

Цель лекции: рассмотреть формализованное описание реляционной модели и операций манипулирования данными как основу для использования математических методов проектирования баз данных и основу создания языков запросов к базе данных.

Базовые понятия реляционных БД



Базовые понятия реляционных БД

Тип данных

- Понятие *тип данных* в реляционной модели данных полностью адекватно понятию типа данных в языках программирования.
- В современных реляционных БД допускается хранение символьных, числовых данных, битовых строк, специализированных числовых данных (таких как "деньги"), а также специальных "темпоральных" данных (дата, время, временной интервал).

Базовые понятия реляционных БД

Домен

- *Домен* - это семантическое понятие, подмножество значений некоторого типа данных, имеющих определенный смысл.
- Домен определяется заданием *базового типа данных*, к которому относятся элементы домена, и произвольного *логического выражения*, применяемого к элементу типа данных.
- Если вычисление этого логического выражения дает результат "истина", то элемент данных является элементом домена.

Базовые понятия реляционных БД

Домен

- Домен имеет *уникальное имя* (в пределах базы данных).
- Домен несет определенную *смысловую нагрузку*.
- Пример: домен D, имеющий смысл "возраст сотрудника" описан как подмножество множества натуральных чисел:

$$D = \{n \in \mathbb{N} : n \geq 18 \text{ and } n \leq 60\}$$

- Домен - потенциальное множество значений данного типа. Домен "Имена" определен на базовом типе строк символов, но в число его значений могут входить только те строки, которые могут изображать имя.

Базовые понятия реляционных БД

Домен

- *Семантическая нагрузка понятия домена*: данные считаются сравнимыми только в том случае, когда они относятся к одному домену.
- Пример: значения доменов "Номера пропусков" и "Номера групп" относятся к типу целых чисел, но не являются сравнимыми.
- Пример: домены "Вес детали" и "Имеющееся количество" можно одинаково описать как множество неотрицательных целых чисел, но смысл этих доменов будет различным, и это будут различные домены.
- Основное значение доменов в том, что *домены ограничивают сравнения*. Некорректно сравнивать значения из различных доменов, даже если они имеют одинаковый тип.

Базовые понятия реляционных БД

Кортеж

- *Кортеж, соответствующий данной схеме отношения* - это множество пар {имя атрибута, значение}, которое содержит одно вхождение каждого имени атрибута, принадлежащего схеме отношения.
- "Значение" является допустимым значением домена данного атрибута.
  - **({A1:Val1}, {A2:Val2},... {An:Valn})**
- *Степень или "арность" кортежа*, т.е. число элементов в нем, совпадает с "арностью" соответствующей схемы отношения.
- *Кортеж* - это набор именованных значений заданного типа.

Базовые понятия реляционных БД

Отношение

- *Отношение* - это множество кортежей, соответствующих одной схеме отношения. Схему отношения называют *заголовком отношения*, а отношение как набор кортежей - *телом отношения*.
- Житейским представлением отношения является таблица, заголовком которой является схема отношения, а строками - кортежи отношения-экземпляра; в этом случае имена атрибутов именуют столбцы этой таблицы.
- *Реляционная база данных* - это набор отношений, имена которых совпадают с именами схем отношений в схеме БД.

Базовые понятия реляционных БД

Отношение

- Отношение R, определенное на множестве доменов D1, D2,... Dn, содержит две части: *заголовок и тело*.
- Заголовок отношения содержит фиксированное количество атрибутов отношения:
- **(<A1:D1>, <A2:D2>,... <An:Dn>)**

Базовые понятия реляционных БД

Схема отношения

- *Схема отношения* - это именованное множество пар {имя атрибута, имя домена}. Степень или "арность" схемы отношения определяет мощность этого множества.
- Степень отношения СОТРУДНИКИ равна четырем, то есть оно является 4-арным.
- *Схема БД* (в структурном смысле) - это набор именованных схем отношений.

Термины в реляционной модели данных имеют соответствующие "табличные" синонимы

Реляционный термин	Соответствующий "табличный" термин
База данных	Набор таблиц
Схема базы данных	Набор заголовков таблиц
Отношение	Таблица
Заголовок отношения	Заголовок таблицы
Тело отношения	Тело таблицы
Атрибут отношения	Наименование столбца таблицы
Кортеж отношения	Строка таблицы
Степень (-арность) отношения	Количество столбцов таблицы
Мощность отношения	Количество строк таблицы
Домены и типы данных	Типы данные в ячейках таблицы

Фундаментальные свойства отношений

Отсутствие кортежей-дубликатов

- Отношения не содержат кортежей-дубликатов. Это свойство следует из определения отношения как множества кортежей и, как всякое множество, не может содержать неразличимые элементы.
- наличие у каждого отношения *первичного ключа* - набора атрибутов, значения которых однозначно определяют кортеж отношения.
- при формальном определении первичного ключа требуется обеспечение его *"минимальности*", т.е. в набор атрибутов первичного ключа не должны входить такие атрибуты, которые можно отбросить без ущерба для основного свойства - однозначно определять кортеж.

Фундаментальные свойства отношений

Отсутствие упорядоченности кортежей

- это свойство является следствием определения отношения-экземпляра как множества кортежей, а множество не упорядочено.
- отсутствие требования к поддержанию порядка на множестве кортежей отношения дает дополнительную гибкость СУБД при хранении баз данных во внешней памяти и при выполнении запросов к базе данных.
- это не противоречит тому, что при формулировании запроса к БД, например, на языке SQL можно потребовать сортировки результирующей таблицы в соответствии со значениями некоторых столбцов.

Фундаментальные свойства отношений

Отсутствие упорядоченности атрибутов

- атрибуты отношений не упорядочены (слева направо), поскольку каждый атрибут имеет уникальное имя в пределах отношения, то порядок атрибутов не имеет значения.
- для ссылки на значение атрибута в кортеже отношения всегда используется имя атрибута.
- по определению схема отношения есть множество пар
  - {имя атрибута, имя домена}

Фундаментальные свойства отношений

Атомарность значений атрибутов

- значения всех атрибутов являются *атомарными*.
- это следует из определения домена как потенциального множества значений простого типа данных, т.е. среди значений домена не могут содержаться множества значений (отношения).
- в реляционных базах данных допускаются только *нормализованные отношения* или отношения, представленные в первой нормальной форме.

Реляционная алгебра

- Т.к. отношения являются *множествами*, то средства манипулирования отношениями могут базироваться на традиционных *теоретико-множественных операциях*, дополненных некоторыми *специальными операциями*, специфичными для баз данных.

- Начальный вариант алгебры был предложен Коддом.

- Набор основных алгебраических операций состоит из восьми операций, которые делятся на два класса:

1. теоретико-множественные операции
2. специальные реляционные операции.

Реляционная алгебра

Теоретико-множественные операции:

- объединение отношений;
- пересечение отношений;
- разность отношений;
- прямое произведение отношений

Специальные реляционные операции :

- ограничение отношения;
- проекция отношения;
- соединение отношений;
- деление отношений

Реляционная алгебра

Дополнительные операции:

- операция *присваивания*, позволяющая сохранить в базе данных результаты вычисления алгебраических выражений,
- операция *переименования атрибутов*, дающая возможность корректно сформировать заголовок результирующего отношения.

Реляционная алгебра

- Реляционная алгебра представляет собой набор операторов, использующих *отношения* в качестве *аргументов*, и возвращающие *отношения* в качестве *результата*.

- Реляционный оператор f - функция с отношениями в качестве аргументов:

$$R = f(R1, R2, ..., Rn)$$

Реляционная алгебра

- Реляционная алгебра является замкнутой, т.к. в качестве аргументов в реляционные операторы можно подставлять другие реляционные операторы, подходящие по типу:

R = f(f1(R11, R12 ,..), f2(R21, R22 ,..),...)

Реляционная алгебра

- Каждое отношение обязано иметь *уникальное имя* в пределах базы данных. Имя отношения, полученного в результате выполнения реляционной операции, определяется в левой части равенства.
- *Не требуется* наличия имен от отношений, полученных в результате реляционных выражений, если эти отношения подставляются в качестве аргументов в другие реляционные выражения. Такие отношения называют **неименованными отношениями.**
- Неименованные отношения реально не существуют в базе данных, а только вычисляются в момент вычисления значения реляционного оператора.

Реляционная алгебра

- Отношения называют **совместимыми по типу**, если они имеют идентичные заголовки, а именно:
  - ✓ отношения имеют одно и то же множество имен атрибутов;
  - ✓ атрибуты с одинаковыми именами определены на одних и тех же доменах.

Оператор переименования атрибутов

- Синтаксис оператора переименования атрибутов :

R **RENAME** ***Atr1, Atr2*****,... AS** ***NewAtr1,New Atr2*****, ...**

где

R – отношение;

Atr1, Atr2 - исходные имена атрибутов;

NewAtr1,New Atr2 - новые имена атрибутов.

Реляционная алгебра

Объединение

- **Объединением** двух совместимых по типу отношений A и B называется отношение с тем же заголовком, что и у отношений A и B, и телом, состоящим из строк (кортежей), принадлежащих или A, или B, или обоим отношениям.

- Синтаксис операции объединения: **A UNION B**

Таблица 1. Отношение A

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Петров	2000
3	Сидоров	3000

Таблица 2. Отношение B

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Пушников	2500
4	Сидоров	3000

Таблица 3. Отношение A UNION B

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Петров	2000
3	Сидоров	3000
2	Пушников	2500
4	Сидоров	3000

Реляционная алгебра

Пересечение

- **Пересечением** двух совместимых по типу отношений A и B называется отношение с тем же заголовком, что и у отношений A и B, и телом, состоящим из кортежей, принадлежащих одновременно обоим отношениям A и B.
- Синтаксис операции пересечения: **A INTERSECT B**

Таблица 1. Отношение A

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Петров	2000
3	Сидоров	3000

Таблица 2. Отношение B

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Пушников	2500
4	Сидоров	3000

Таблица 4. Отношение A INTERSECT B

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000

Реляционная алгебра

Вычитание (разность)

- **Вычитанием** двух совместимых по типу отношений A и B называется отношение с тем же заголовком, что и у отношений A и B, и телом, состоящим из кортежей, принадлежащих отношению A и не принадлежащих отношению B.

- Синтаксис операции вычитания: **A MINUS B**

Таблица 1. Отношение A

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Петров	2000
3	Сидоров	3000

Таблица 2. Отношение B

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Пушников	2500
4	Сидоров	3000

Таблица 5. Отношение A MINUS B

Табельный номер	**Фамилия**	**Зарплата**
2	Петров	2000
3	Сидоров	3000

Реляционная алгебра

Декартово произведение

- **Декартовым произведением** двух отношений *A(A1, A2,...An)* и *B(B1, B2,...Bn)* называется отношение, заголовок которого является сцеплением заголовков отношений *A* и *B*: *(A1,A2,...,An,B1,B2,...Bn)* , а тело состоит из кортежей, являющихся сцеплением кортежей отношений *A* и *B*: *(a1,a2,...,an,b1,b2,...,bn)* , таких, что
  - $(a1,a2,...,an) \in A$,
  - $(b1,b2,...,bn) \in B$

- Синтаксис операции декартового произведения: **A TIMES B**

Таблица 6. Отношение A (Поставщики)

Номер поставщика	**Наименование поставщика**
1	Завод 1
2	Завод 2
3	Завод 3

Таблица 7. Отношение B (Детали)

Номер детали	**Наименование детали**
1	Болт
2	Гайка
3	Винт

Декартово произведение отношений *A* и *B* будет иметь вид:

Таблица 8. Отношение A TIMES B

Номер поставщика	**Наименование поставщика**	**Номер детали**	**Наименование детали**
1	Завод 1	1	Болт
1	Завод 1	2	Гайка
1	Завод 1	3	Винт
2	Завод 2	1	Болт
2	Завод 2	2	Гайка
2	Завод 2	3	Винт
3	Завод 3	1	Болт
3	Завод 3	2	Гайка
3	Завод_3	3	Винт

Реляционная алгебра

Выборка (ограничение, селекция)

- **Выборкой** на отношении A с условием z называется отношение с тем же заголовком, что и у отношения A, и телом, состоящем из кортежей, значения атрибутов которых при подстановке в условие z дают значение ИСТИНА.
- «Z» представляет собой логическое выражение, в которое могут входить атрибуты отношения A и скалярные выражения.
- В простейшем случае условие z имеет вид XΘY, где Θ - один из операторов сравнения, а X и Y - атрибуты отношения A или скалярные значения. Такие выборки называются Θ-выборки (тэта-выборки) или Θ-ограничения, Θ-селекции.
- Синтаксис операции выборки:
- **A WHERE z или A WHERE XΘY**

Пример. Пусть дано отношение *A* с информацией о сотрудниках:

Таблица 9. Отношение A

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Петров	2000
3	Сидоров	3000

Результат выборки *A* WHERE *Зарплата <3000* будет иметь вид:

Таблица 10. Отношение A WHERE Зарплата<3000

Табельный номер	**Фамилия**	**Зарплата**
1	Иванов	1000
2	Петров	2000

Реляционная алгебра

Проекция

- **Проекцией** отношения A по атрибутам X,Y,...,Z, где каждый из атрибутов принадлежит отношению A, называется отношение с заголовком (X,Y,...,Z) и телом, содержащим множество кортежей вида (x,y,. . .,z), таких, для которых в отношении A найдутся строки (кортежи) со значением атрибута X равным x, значением атрибута Y равным y, …, значением атрибута Z равным z.

- Синтаксис операции проекции: **A[X,Y,...,Z]**

Пример. Пусть дано отношение *A* с информацией о поставщиках, включающих их наименование и месторасположение:

Таблица 11. Отношение A (Поставщики)

Номер поставщика	**Наименование поставщика**	**Город поставщика**
1	Завод_1	Уфа
2	Завод_2	Москва
3	Завод_3	Москва
4	Завод_4	Челябинск

Проекция *A* [Город поставщика] будет иметь вид:

Таблица 12. Отношение A[Город поставщика]

Город поставщика
Уфа
Москва
Челябинск

Реляционная алгебра

Соединение

- Соединением отношений A и B по условию z называется отношение ***(A* TIMES *B)* WHERE z**

  где z представляет собой логическое выражение, в которое могут входить атрибуты отношений A и B и скалярные выражения.

Таблица 13 Отношение A (Поставщики)

Номер поставщика	**Наименование поставщика**
1	Иванов
2	Петров
3	Сидоров

Таблица 14. Отношение B (Детали)

Номер детали	**Наименование детали**
1	Болт
2	Гайка
3	Винт

Декартово произведение отношений *A* и *B* будет иметь вид:

Таблица 15. Отношение A TIMES B

Номер поставщика	**Наименование поставщика**	**Номер детали**	**Наименование детали**
1	Иванов	1	Болт
1	Иванов	2	Гайка
1	Иванов	3	Винт
2	Петров	1	Болт
2	Петров	2	Гайка
2	Петров	3	Винт
3	Сидоров	1	Болт
3	Сидоров	2	Гайка
3	Сидоров	3	Винт

Если к нему приложить условие «Какие поставщики поставляют болты» то есть «Какие поставщики поставляют детали с номером 1», то в результате получим отношение

Таблица 16. Отношение A TIMES B WHERE Номер детали = 1

Номер поставщика	**Наименование поставщика**	**Номер детали**	**Наименование детали**
1	Иванов	1	Болт
2	Петров	1	Болт
3	Сидоров	1	Болт

Реляционная алгебра

Тэта-соединение

- Пусть отношение A содержит атрибут X, отношение B содержит атрибут Y, а Θ - один из операторов сравнения.
- Тогда Θ-соединением отношения A по атрибуту X с отношением B по атрибуту Y называют отношение
  - **(A TIMES B) WHERE XΘY**.

- Синтаксис операции Θ-соединения : **A [XΘY] B**

Тэта-соединение

Пусть отношение *A* содержит атрибут *X*, отношение *B* содержит атрибут *Y*, а Θ - один из операторов сравнения (и т.д.).

Тогда *Θ-соединением* отношения *A* по атрибуту *X* с отношением *B* по атрибуту *Y* называют отношение

(A TIMES B) WHERE *XΘY*.

Синтаксис операции Θ-соединения :

A [XΘY] B.

Таблица 17. Отношение A (Поставщики)

Номер поставщика	**Наименование поставщика**	**Статус поставщика (X)**
1	Иванов	4
2	Петров	1
3	Сидоров	2

Таблица 18. Отношение B (Детали)

Номер детали	**Наименование детали**	**Статус детали (Y)**
1	Болт	3
2	Гайка	2
3	Винт	1

Ответ на вопрос "какие поставщики имеют право поставлять какие детали?" дает Θ-соединение *A [X>=Y]*

Таблица 19. Отношение A [X>=Y] B

Номер поставщика	**Наименование поставщика**	**X (Статус поставщика)**	**Номер детали**	**Наименование детали**	**Y (Статус детали)**
1	Иванов	4	1	Болт	3
1	Иванов	4	2	Гайка	2
1	Иванов	4	3	Винт	1
2	Петров	1	3	Винт	1
3	Сидоров	2	2	Гайка	2
3	Сидоров	2	3	Винт	1

Реляционная алгебра

Экви-соединение

- Наиболее важным частным случаем Θ-соединения является случай, когда Θесть просто равенство.

- Синтаксис операции экви-соединения : **A [X=Y] B**

Таблица 20. Отношение A (Поставщики)

Номер поставщика	**Наименование поставщика**	**Статус поставщика (X)**
1	Иванов	4
2	Петров	1
3	Сидоров	2

Таблица 21. Отношение B (Детали)

Номер детали	**Наименование детали**	**Статус изделия (Y)**
1	Болт	3
2	Гайка	2
3	Винт	1

Таблица 22. Отношение A [X=Y] B

Номер поставщика	**Наименование поставщика**	**X (Статус поставщика)**	**Номер детали**	**Наименование детали**	**Y (Статус изделия_1**
3	Сидоров	2	2	Гайка	2

Реляционная алгебра

Естественное соединение

- Пусть даны отношения *A(A1,A2,…,An,X1,X2,…Xn)* и *B(X1,X2,…Xn,B1,B2,. ..,Bn),* имеющие одинаковые атрибуты *X1,X2,…Xn.*
- Тогда **естественным соединением** отношений A и B называется отношение с заголовком *(A1,A2,…,An,X1,X2,…Xn, B1,B2,…,Bn)* и телом, содержащим множество кортежей *(a1,a2,…,an,x1,x2,…,xn,b1,b2,…,bn),* таких, что
  - (a1,a2,...,an,x1,x2,...,xn) $\in$ A и
  - (x1,x2,...,xn,,b1,b2,...,bn) $\in$ B.
- Синтаксис: **A JOIN B**

Таблица 23. Отношение P (Поставщики)

Номер поставщика (PNUM)	**Наименование поставщика (PNAME)**
1	Иванов
2	Петров
3	Сидоров

Таблица 24. Отношение D (Детали)

Номер детали (DNUM)	**Наименование детали (DNAME)**
1	Болт
2	Гайка
3	Винт

Таблица 25. Отношение PD (Поставки)

Номер поставщика (PNUM)	**Номер детали (DNUM)**	**Поставляемое количество VOLUME**
1	1	100
1	2	200
1	3	300
2	1	150
2	2	250
3	1	1000

Ответ на вопрос "какие детали поставляются поставщиками", можно записать в виде естественного соединения трех отношений *P JOIN PD JOIN D*):

Таблица 26. Отношение P JOIN PD JOIN D

Номер поставщика PNUM	**Наименование поставщика PNAME**	**Номер детали DNUM**	**Наименование детали DNAME**	**Поставляемое количество VOLUME**
1	Иванов	1	Болт	100
1	Иванов	2	Гайка	200
1	Иванов	3	Винт	300
2	Петров	1	Болт	150
2	Петров	2	Гайка	250
3	Сидоров	1	Болт	1000

Реляционная алгебра

Деление

- Пусть отношение A определено на множестве атрибутов X, а отношение B - на множестве атрибутов Y, причем Y является подмножеством X.
- Пусть Z= X- Y, то есть Z является множеством атрибутов отношения A, которые не являются атрибутами отношения B.
- Результатом операции **деления** является набор кортежей отношения A, определенных на множестве атрибутов Z, которые соответствуют комбинации всех строк отношения B.

- Синтаксис операции деления: **A DEVIDBY B**

Пример. В примере с поставщиками, деталями и поставками ответим на вопрос, "какие поставщики поставляют *все* детали?".

В качестве делимого возьмем проекцию *X= PD [PNUM, DNUM]*, содержащую номера поставщиков и номера поставляемых ими деталей:

Таблица 27. Проекция X=PD[PNUM,DNUM]

Номер поставщика PNUM	**Номер детали DNUM**
1	1
1	2
1	3
2	1
2	2
3	1

В качестве делителя возьмем проекцию *Y=D[DNUM]*, содержащую список номеров *всех* деталей (не обязательно поставляемых кем-либо):

Таблица 29. Проекция Y=D[DNUM]

Номер детали DNUM
1
2
3

Деление *X DEVIDEBY Y* дает список номеров поставщиков, поставляющих *все* детали:

Таблица 29. Отношение X DEVIDEBY Y

Номер поставщика PNUM
1

Оказалось, что только поставщик с номером 1 поставляет все детали.

Контрольные вопросы:

1. Дайте определение основным понятиям реляционных баз данных. Поясните каждое из понятий на примере своей разработанной базы данных
2. На чем основано свойство «Отсутствие кортежей-дубликатов»?
3. На чем основано свойство «Отсутствие упорядоченности кортежей»?
4. На чем основано свойство «Атомарность значений атрибутов»?
5. Какие операции относятся к теоретико-множественным? Приведите примеры
6. Какие операции относятся к специальным реляционным операциям? Приведите примеры

Список литературы:

1. Т. Конноли. Базы данных. Проектирование, реализация и сопровождение. Теория и практика.: Пер.с англ. – М.: Изд.дом «Вильямс», 2012. – 1440 с.
2. К.Дейт. Введение в системы БД. изд.6-е. :Пер.с англ. - М.: Изд.дом "Вильямс", 2014. – 1200с.
3. Д.Ульман. Введение в системы баз данных. – М.: Издательство «Лори», 2015. – 853с.

Спасибо за внимание!