Лекция 8.1.

Операторы работы с наборами. Объединение результатов двух запросов в один результирующий набор.

Оператор UNION [ALL].

Пересечение и разность результатов двух запросов.

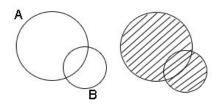
Операторы NTERSECT, EXCEPT.

1. Oператор UNION [ALL]

Объединением множеств A и B называется множество, обозначаемое $A \cup B$, состоящее из всех тех и только тех элементов, которые принадлежат или A или B, то есть

$$A \cup B = \left\{x \mid (x \in A) \lor (x \in B)\right\}.$$
Например, если $A = \{1, 2, 3, 4\}, B = \{1, 3, 5\}, C = \{5, 6\},$ **то** $A \cup B = \{1, 2, 3, 4, 5\}, A \cup C = \{1, 2, 3, 4, 5, 6\}, B \cup C = \{1, 3, 5, 6\}.$

Операции над множествами удобно иллюстрировать фигурами, называемыми диаграммами Венна (другое название - круги Эйлера). На рисунке ниже слева большим и малым кругами обозначены соответственно множества A и B, а справа - результат объединения этих множеств (заштрихованная фигура).



На основе теории множеств создана концепция реляционных баз данных, а на основе операций над множествами - реляционная алгебра и её операции - используемые в языке запросов к базам данных SQL. Операция объединения есть в реляционной алгебре.

Повтор лекции 2. Раздел 2 Модели данных

1) Объединением двух совместимых по типу отношений A и B (A UNION B) называется отношение c тем же заголовком, как и b отношениях b и b и b и b или обоим отношениям.

Примеры: Исходные отношения:

A		
CityNo	CityName	RgNo
1	Темиртау	1
2	Сарань	1
3	Балхаш	1

В		
CityNo	CityName	RgNo
2	Сарань	1
3	Балхаш	1
4	Алматы	2

Объединение: A UNION B

CityNo	CityName	RgNo
1	Темиртау	1
2	Сарань	1
3	Балхаш	1
4	Алматы	2

Объединение — бинарная операция над односхемными отношениями результатом является отношение, которое включает в себя все кортежи обоих исходных отношений без повторов; причём, имена полей односхемных отношений могут быть разными, достаточно, чтобы совпадало количество полей и типы данных соответствующих полей.

Пример. Для отношений R(A,B,C) и S(A,D,E), объединение R U S будет таким как на рис.:

Отношение К		
A	В	C
a	b	С
С	a	d
С	b	d

Отн	Отношение S		
A	D	E	
C	a	е	
С	b	d	

Объе	Объединение R U S		
A	D	F	
а	b	С	
С	а	d	
С	b	d	
С	а	е	

Объединением двух множеств (назовем их A и B) является множество, содержащее все элементы из A и B. Другими словами, в результирующий набор попадают элементы, принадлежащие хотя бы одному из множеств. Закрашенная область представляет результат, возвращаемый оператором работы с наборами.

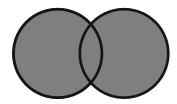


Рис. 6.1. Объединение двух множеств А и В

В языке T-SQL <u>оператор UNION</u> объединяет результаты двух входящих запросов. Строка попадает в результирующий набор даже в том случае, если присутствует только с одной стороны. Этот оператор в T-SQL поддерживает оба параметра — DISTINCT (по умолчанию) и ALL.

Для объединения результатов двух или более запросов в одну таблицу используется команда UNION. Команда UNION объединяет вывод двух или более запросов в единый набор строк и столбцов и имеет вид:

Первый запрос UNION [ALL]

Второй запрос

. . .

Для объединения результатов нескольких запросов с помощью UNION, они должны соответствовать следующим требованиям:

- содержать одинаковое количество столбцов;
- типы данных столбцов должны совпадать во всех запросах;
- <u>– в промежуточных запросах нельзя использовать сортировку ORDER BY.</u>

Чтобы отсортировать результат объединения, в конце запроса добавляется ORDER BY. Названия столбцов в запросах могут отличаться. Поэтому в команде ORDER BY указывается название столбцов с первого запроса.

Если объединяемые наборы содержат в строках идентичные значения, то при объединении повторяющиеся строки удаляются.

Если необходимо при объединении сохранить повторяющиеся строки, то для этого **используется параметр ALL.**

Все запросы выполняются независимо друг от друга, а уже их вывод объединяется.

В объединяемых запросах можно использовать одну и ту же таблицу.

Результатом объединения двух множеств (отношений) А и В ($^{A \cup B}$) будет такое множество (отношение) С, которое включает в себя те и только те элементы, которые есть или во множестве А или во множестве В. Говоря упрощённо, все элементы множества А и множества В, за исключением дубликатов, образующихся за счёт того, что некоторые элементы есть и в первом, и во втором множестве.

	R1			R2	
A1	A2	A3	A1	A2	A3
Z7	aa	w11	X8	pp	k21
B7	hh	h15	Q2	ee	h15
X8	pp	w11	X8	pp	w11

SELECT A1, A2, A3 from R1 UNION SELECT A1, A2, A3 from R2

	R	
A1	A2	A3
Z7	aa	w11
B7	hh	h15
X8	pp	w11
X8	pp	k21
Q2	ee	h15

Оператор UNION подобно inner join или outer join позволяет соединить две таблицы. Но в отличие от inner/outer join объединения соединяют не столбцы разных таблиц, а два однотипных набора в один. Формальный синтаксис объединения:

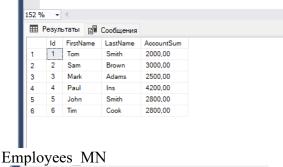
```
SELECT_выражение1
UNION [ALL] SELECT_выражение2
[UNION [ALL] SELECT_выражениеN]
```

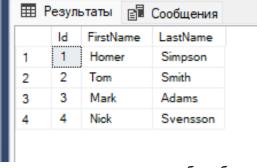
Пусть в базе данных будут две отдельные таблицы для клиентов банка (таблица Customers MN) и для сотрудников банка (таблица Employees MN):

```
use master;
| create database testbasa_MN;
    use testbasa_MN;

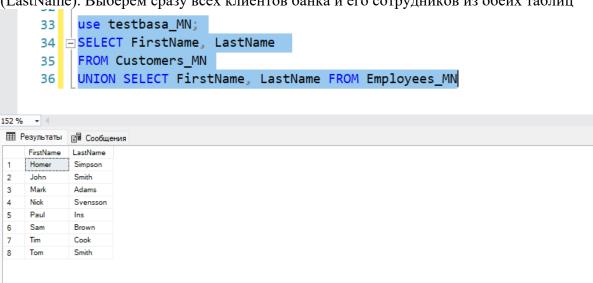
□ CREATE TABLE Customers_MN
           Id INT IDENTITY PRIMARY KEY,
           FirstName NVARCHAR(20) NOT NULL,
           LastName NVARCHAR(20) NOT NULL,
          AccountSum MONEY
     CREATE TABLE Employees_MN
11
          Id INT IDENTITY PRIMARY KEY,
FirstName NVARCHAR(20) NOT NULL,
 13
14
 15
          LastName NVARCHAR(20) NOT NULL,
16
     21
       INSERT INTO Employees_MN VALUES
('Homer', 'Simpson'),
('Tom', 'Smith'),
('Mark', 'Adams'),
('Nick', 'Svensson')
```

Customers MN





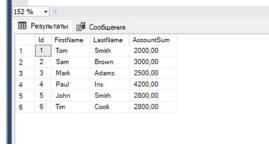
Здесь можем заметить, что обе таблицы, несмотря на наличие различных данных, могут характеризоваться двумя общими атрибутами - именем (FirstName) и фамилией (LastName). Выберем сразу всех клиентов банка и его сотрудников из обеих таблиц



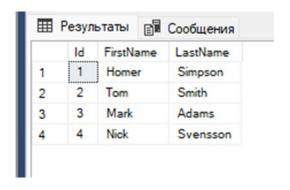
В данном случае из первой таблицы выбираются два значения - имя и фамилия клиента. Из второй таблицы Employees также выбираются два значения - имя и фамилия сотрудников. То есть при объединении количество выбираемых столбцов и их тип совпадают для обеих выборок.

Оператор работы с множествами UNION (с параметром DISTINCT умолчанию) объединяет результаты выполнения двух запросов, удаляя дубликаты. Если строка содержится в обоих наборах, в конечный результат она попадет только один раз; другими словами, итоговый набор является полноценным множеством.

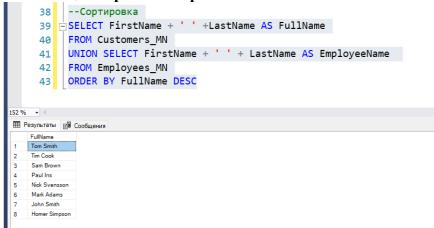
Customers MN



Employees MN



При этом названия столбцов объединенной выборки будут совпадать с названия столбцов первой выборки. И если мы захотим при этом еще произвести сортировку, то в выражениях ORDER BY необходимо ориентироваться именно на названия столбцов первой выборки



Если же в одной выборке <u>больше столбцов, чем в другой</u>, то они не смогут быть объединены. Например, в следующем случае объединение завершится с ошибкой:

SELECT FirstName, LastName, AccountSum

FROM Customers

UNION SELECT FirstName, LastName

FROM Employees

Также соответствующие **столбцы должны соответствовать по типу**. Так, следующий пример завершится **с ошибкой из-за не соответствия по типу данных:**

SELECT FirstName, LastName

FROM Customers

UNION SELECT Id, LastName

FROM Employees

B данном случае первый столбец первой выборки имеет тип NVARCHAR, то есть хранит строку. Первый столбец второй выборки - Id имеет тип INT, то есть хранит число.

Если оба объединяемых набора содержат в строках идентичные значения, то при объединении повторяющиеся строки удаляются.

Например, в случае с таблицами Customers_MN и Employees_MN сотрудники банка могут быть одновременно его клиентами и содержаться в обеих таблицах. При объединении в примерах выше всех дублирующиеся строки удалялись.

Если же необходимо при объединении сохранить все, в том числе повторяющиеся строки, то для этого необходимо использовать оператор ALL

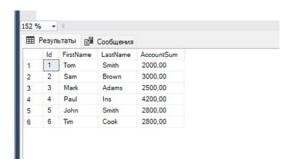
```
44
45
— -Выберем сразу всех клиентов банка и его сотрудников из обеих таблиц
46
47
48
use testbasa_MN;
49
— SELECT FirstName, LastName
FROM Customers_MN
UNION ALL SELECT FirstName, LastName FROM Employees_MN

152

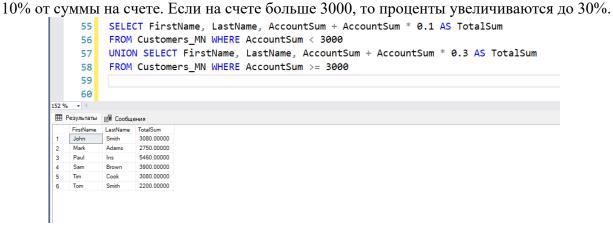
■ Pesymetato

| FirstName | LastName | | | |
| Form | Simit |
| Same | Sowin |
| Same | Same | Sowin |
| Same | Same | Same | Same | Same |
| Same | Same
```

Поскольку оператор UNION ALL не удаляет дубликаты, возвращаемый им результат является мультимножеством. Одна и та же строка может присутствовать в результирующем наборе в нескольких экземплярах



Объединять выборки можно и из одной и той же таблицы. Например, в зависимости от суммы на счете клиента нам надо начислять ему определенные проценты: В данном случае если сумма меньше 3000, то начисляются проценты в размере



Пример 1. В базе данных фирмы есть таблица Staff, содержащая данные о сотрудниках фирмы. В ней есть столбцы Salary (размер заработной платы), Job (должность) и Years (длительность трудового стажа). Первый запрос возвращает индивидуальные размеры заработной платы, упорядоченные по должностям:

```
SELECT Name, Job, Salary
FROM STAFF ORDER BY Job
```

Результатом выполнения запроса будет следующая таблица:

Name	Job	Salary
Sanders	Mgr	18357.5
Marenghi	Mgr	17506.8
Pernal	Sales	18171.2
Doctor	Sales	12322.4
Factor	Sales	16228.7

Второй запрос вернёт суммарную заработную плату по должностям. Мы уже готовим этот запрос для соединения с первым, поэтому будем помнить, что условием соединения является равное число столбцов, совпадение их названий, порядка следования и типов данных. Поэтому включаем в таблицу с итогами также столбец Name с произвольным значением 'Z-TOTAL':

Результатом выполнения запроса будет следующая таблица:

Name	Job	Salary
Z-TOTAL	Mgr	35864.3
Z-TOTAL	Sales	46722.3

Теперь объединим запросы при помощи оператора UNION и применим оператору ORDER BY к результату объединения. Группировать следует по двум столбцам: должность (Job) и имя (Name), чтобы строки с итоговыми (суммарными) значениями, в которых значение имени - 'Z-TOTAL', находились ниже строк с индивидуальными значениями. Объединение результатов запросов будет следующим:

Результатом выполнения запроса с оператором UNION будет следующая таблица, в которой каждая первая строка в каждой группе должностей будет содержать суммарную заработную плату сотрудников, работающих на этой должности:

Name	Job	Salary
Marenghi	Mgr	17506.8
Sanders	Mgr	18357.5
Z-TOTAL	Mgr	35864.3
Doctor	Sales	12322.4
Factor	Sales	16228.7
Pernal	Sales	18171.2
Z-TOTAL	Sales	46722.3

..

<u>Команда UNION объединяет данные из нескольких таблиц в одну при</u>выборке.

<u>При объединении количество столбцов во всех таблицах должно совпадать,</u> иначе будет ошибка

<u>Имена столбцов будут такие же, как в основной таблице, в которую</u> добавляются данные из других таблиц.

Внимание: если не используется ключевое слово ALL для UNION, все возвращенные строки будут уникальными, так как по умолчанию подразумевается distinct, который удаляет неуникальные значения.

Чтобы отменить такое поведение - нужно указать ключевое слово ALL UNION AL

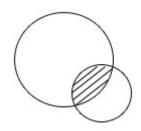
2. Пересечение результатов двух запросов. Оператор NTERSECT

Пересечением множеств A и B называется множество, обозначаемое $A \cap B$ и состоящее из всех тех и только тех элементов, которые принадлежат каждому из множеств A и B, то есть

$$A \cap B = \{x \mid (x \in A) \land (x \in B)\}$$

Например, если
$$A = \{1, 2, 3, 4\}$$
, $B = \{1, 3, 5\}$, $C = \{5, 6\}$, то $A \cap B = \{1, 3\}$, $A \cap C = \emptyset$, $B \cap C = \{5\}$.

На рисунке ниже - результат пересечения множеств A и B - заштрихованная фигура.



Повтор лекции 2. Раздел 2 Модели данных

Пересечение:

Исходные отношения:

CityNo	CityName	RgNo
1	Темиртау	1
2	Сарань	1
3	Балхаш	1

CityNo	CityName	RgNo
2	Сарань	1
3	Балхаш	1
4	Алматы	2

A INTERSECT B

CityNo	CityName	RgNo
2	Сарань	1
3	Балхаш	1

Пересечение — бинарная операция над односхемными отношениями. Пересечение односхемных отношений R и S есть подмножество кортежей, принадлежащих обоим отношениям; причём, имена полей односхемных отношений могут быть разными, достаточно, чтобы совпадало количество полей и типы данных соответствующих полей.

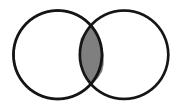
Пересечение можно выразить через разность так: $R \cap S = R - (R - S)$.

Отн	Отношение R		
Α	В	C	
a	b	С	
С	а	d	
С	b	d	

Отношение S		
Α	D	E
g	h	а
a	b	С
С	a	е

Пере	Пересечение R ∩ S		
Α	D	F	
а	b	С	

Пересечением двух множеств (назовем их A и B) является множество всех элементов, которые принадлежат как A, так и B.



В языке T-SQL оператор INTERSECT возвращает пересечение результирующих наборов, полученных при выполнении двух запросов; в результат попадают только те строки, которые содержатся в обоих входящих наборах. Сначала мы рассмотрим оператор INTERSECT (с параметром DISTINCT по умолчанию), затем перейдем к его аналогу, INTERSECT ALL, который возвращает мультимножество

Множества (оператор INTERSECT)

Первым делом оператор INTERSECT убирает дубликаты из двух входящих мультимножеств и затем возвращает строки, которые содержатся в обоих наборах. Таким образом, если строка попадает в результат, это означает, что она как минимум по одному разу входит в каждое мультимножество.

Команда **INTERSECT** позволяет **найти общие строки в результатах двух запросов**, то есть данный оператор выполняет операцию пересечения множеств. Команда INTERSECT имеет следующий вид:

Первый запрос INTERSECT Второй запрос.

Требования к использованию команды INTERSECT такие же, как к командам UNION

Теперь посмотрим, что получится в результате выполнения этой операции реляционной алгебры и соответствующего ей запроса SQL. Вновь даны два отношения R1 и R2:

	R1			R2	
A1	A2	A3	A1	A2	A3
Z 7	aa	w11	X8	pp	k21
B7	hh	h15	Q2	ee	h15
X8	pp	w11	X8	pp	w1

SELECT A1, A2, A3 from R1 INTERSECT SELECT A1, A2, A3 from R2

Просматриваем все записи в двух отношениях, и обнаруживаем, что и в первом, и во втором отношении есть одна строка - та, которая является третьей и в первом, и во втором отношении.

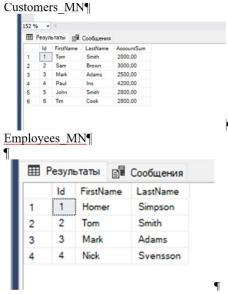
Получаем новое отношение

	R	
A1	A2	A3
X8	pp	w11

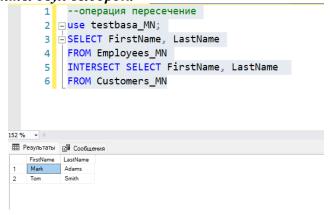
Оператор INTERSECT позволяет найти общие строки для двух выборок, то есть данный оператор выполняет операцию пересечения множеств.

Для его использования применяется следующий формальный синтаксис:

SELECT_выражение1 INTERSECT SELECT выражение2



В таблице Customers хранятся все клиенты банка, а в таблице Employees - все его сотрудники. Но сотрудники могут быть одновременно и клиентами банка, поэтому их данные могут храниться сразу в двух таблицах. Найдем всех сотрудников банка, которые одновременно являются его клиентами. То есть нам надо найти общие элементы двух выборок:



INTERSECT (пересечение) — это оператор Transact-SQL, который выводит одинаковые строки из первого, второго и последующих наборов данных. Другими словами, он выведет только те строки, которые есть как в первом результирующем наборе, так и во втором (третьем и так далее), т.е. происходит пересечение этих строк.

Данный оператор очень полезен, например, тогда, когда необходимо узнать какие строки есть и в первой таблице и во второй (*к примеру*, <u>повтор данных</u>).

Как и у оператора UNION, у INTERSECT есть правила, например, то, <u>что количество</u> полей во всех результирующих наборах должно быть одинаковым, также как и их тип данных.

3. Разность результатов двух запросов Операторы EXCEPT

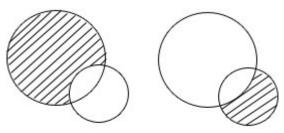
Разность множеств

Разностью множеств A и B называется множество, обозначаемое A-B и состоящее из всех тех и только тех элементов множества A, которые не являются элементами множества B, то есть

$$A-B=\left\{x\left|\left(x\in A\right)\wedge\left(x\not\in B\right)\right\}$$

Например,
$$A = \{1, 2, 3, 4\}$$
, $B = \{1, 3, 5\}$, $C = \{5, 6\}$, $B = \{2, 4\}$, $A - C = A$, $B - C = \{1, 3\}$, $C - B = \{6\}$, $B - A = \{5\}$.

Hа рисунке ниже слева - результат разности множеств A и B, а справа - результат разности множеств B и A.



Повтор лекции 2. Раздел 2 Модели данных

Вычитанием двух совместимых по типу отношений A и B (A MINUS B) называется отношение с тем же заголовком, как и в отношениях A и B, и с телом, состоящим из множества всех кортежей, принадлежащих отношению A и не принадлежащих отношению B.

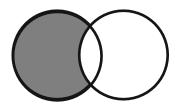
A		
CityNo	CityName	RgNo
1	Темиртау	1
2	Сарань	1
3	Балхаш	1

В		
CityNo	CityName	RgNo
2	Сарань	1
3	Балхаш	1
4	Алматы	2

A	MINUS B	
CityNo	CityName	RgNo
1	Темиртау	1

B N	IINUS A	
CityNo	CityName	RgNo
4	Алматы	2

Разностью множеств A и B (A – B) является множество элементов, принадлежащих A, но не принадлежащих B. Можете считать, что это все элементы A, кроме тех, которые также входят в B.



В языке T-SQL разность множеств реализована в виде оператора работы с на борами под названием EXCEPT; он принимает два входящих запроса и возвращает строки, которые присутствуют в первом из них, но отсутствуют во втором.

Paccмотрим оператор EXCEPT (с параметром DISTINCT по умолчанию)

Множества (оператор ЕХСЕРТ)

Первым делом оператор EXCEPT убирает дубликаты из двух входящих мультимножеств (превращая их тем самым в настоящие множества), после чего возвращает строки, которые содержатся в первом наборе, но которых нет во втором.

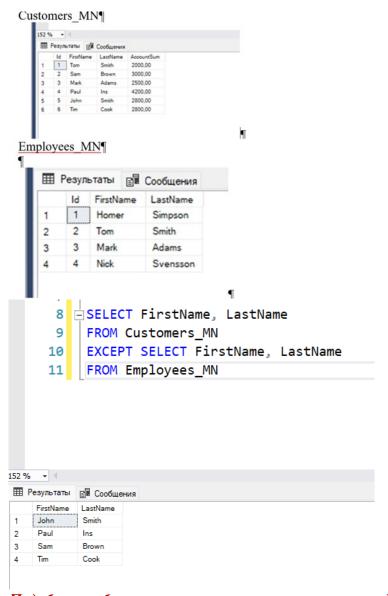
SELECT_выражение1 EXCEPT SELECT выражение2

Другими словами, строка, попавшая в конечный результат, встречается в первом мультимножестве как минимум в одном экземпляре, а во втором ее точно нет.

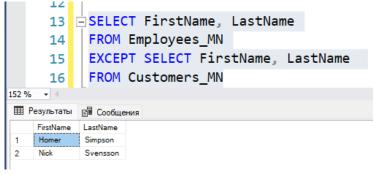
Оператор ЕХСЕРТ, в отличие от предыдущих двух, является ассиметричным, то есть учитывает порядок следования входящих множеств.

Здесь те же правила, что и у оператора INTERSECT, т.е. количество столбцов (и их тип) должно быть одинаковым.

Таблица Employees содержит данные обо всех сотрудниках банка, а таблица Customers - обо всех клиентах. Но сотрудники банка могут также быть его клиентами. И допустим, нам надо найти всех клиентов банка, которые не являются его сотрудниками:



Подобным образом можно получить всех сотрудников банка, которые не являются его клиентами:



EXCEPT полезен тогда, когда необходимо сравнить две таблицы и вывести только те строки первой таблице, которых нет в другой таблице.