**Контрольная по пройденному материалу состоится 07.12.17 на Физтехе в 117 гк в 18:35.**

**Задания (про оценку стоимости товара в автомобиле и БД «Книжный магазин») следует обязательно прислать до 10.12.17 23:59 на e-mail nixlab@nix.ru с заголовком "ПБП: задача N". Итоговая оценка будет складываться из домашних заданий и контрольной. Условия заданий в предыдущих лекциях.**

### Лекция 10.

- Использование подзапросов. Предикаты EXISTS, IN, ANY и т.д.

**Практика:**

* Решение задачи про рейтинг при помощи подзапросов.
* Вывод скользящей статистики при помощи подзапросов.
* Поиск минимального пропущенного числа в ряде натуральных чисел.

**Лекция 10.**

### Использование подзапросов:

SQL позволяет использовать результаты одного запроса в другом, то есть вкладывать запросы друг в друга. Таким образом, используя подзапросы, можно конструировать более сложные запросы. Подчиненные запросы могут использоваться для:

* проверки наличия результатов подчиненного запроса (используются зарезервированные слова EXISTS или NOT EXISTS);
* поиска значений в основном запросе, которые равны, превышают или меньше значений, возвращаемых подчиненным запросом (используются зарезервированные слова ANY, IN или ALL);
* создания подчиненных запросов внутри подчиненных запросов (вложенные запросы).

Подзапросы могут быть независимыми (self-contained) и зависимыми (correlated), могут возвращать скалярное значение (single valued), много значений (multi valued) или таблицу (table-valued).

Разберем сначала использование подзапросов в WHERE, хотя подзапросы можно использовать практически во всех предложениях: SELECT (single-valued), FROM, HAVING (single, multi).

**Использование в WHERE:**

Есть таблица Номенклатура с полями: **Товар\_ID**, Наименование, Масса. Ключевое поле – Товар\_ID.

Вы хотите получить все товары, масса которых больше, чем масса товара, задаваемого параметром. Возможный пример такого запроса:

SELECT Товар\_ID, Наименование, Масса

FROM Номенклатура

WHERE Масса > (SELECT Масса FROM Номенклатура WHERE Товар\_ID = @Товар)

Это пример независимого подзапроса (self-contained), возвращающего одно значение (single valued). Независимого – так как подзапрос не ссылается на внешний запрос и может быть выполнен отдельно от внешнего запроса (поэтому независимые запросы легко исследовать на правильность работы (debug)). Одно значение – так как в таблице номенклатура у нас ровно одно значение массы для каждого товара.

В целом, подзапросы, которые возвращают одно значение можно помещать в любую часть запроса, где возможно использование скаляров: SELECT, FROM, WHERE, HAVING.

Следующий запрос выполнить не удастся в том случае, когда параметром будет задано значение массы меньшее, чем значение массы для нескольких артикулов.

SELECT Товар\_ID, Наименование, Масса

FROM Номенклатура

WHERE Товар\_ID = (SELECT Товар\_ID FROM Номенклатура WHERE Масса >@Масса)

В результате выполнения подзапроса выведется более одного артикула, а реляционные операторы задают сравнения строго между двумя значениями.

То есть при использовании реляционных операторов подзапрос должен выбирать одно поле, а тип данных этого поля должен совпадать с тем значением, с которым он будет сравниваться в предикате.

Пример подзапроса выше – это независимый многозначный запрос (self-contained multi valued подзапрос).

Как можно обеспечить единственное значение в подзапросе? Для примера выше можно заменить «=» на in.

Следующие запросы будут написаны к такой схеме данных:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Документы** |  | **Документы\_данные** |  | **Остатки** |
| **ндок** |  | **ндок** |  | **Дата** |
| Дата |  | **Товар\_ID** |  | **Товар\_ID** |
| Покупатель |  | Колво |  | Остаток |
| Сумма |  | Цена |  |  |

1. ***Использование distinct***

Допустим, цена на любой товар устанавливается в начале дня и в течение дня не меняется. Пусть нам необходимо найти товары, цена продажи на которые была меньше, чем цена на товар, заданный параметром, в заданный также параметром день.

SELECT DISTINCT Документы\_данные.Товар\_ID

FROM Документы INNER JOIN Документы\_данные ON

Документы.ндок = Документы\_данные.ндок

WHERE CAST([Документы].[Дата] as date) = @День

AND Документы\_данные.Цена <

(

SELECT DISTINCT Цена

FROM Документы INNER JOIN Документы\_данные ON

Документы.ндок = Документы\_данные.ндок

WHERE Товар\_ID = [Товар] AND

CAST(Документы.Дата as date) = @День

);

1. ***Использование агрегатных функций***

Агрегатные функции автоматически производят единственное значение для любого числа строк.

Пример:

1. Получение документа с максимальным оборотом.

SELECT Покупатель, Сумма, ндок

FROM Документы as Продажи

WHERE Сумма =

(

SELECT MAX(Сумма)

FROM Документы

)

1. Получение документа с максимальным оборотом по каждому покупателю:

Подумайте, почему следующие запросы не подойдут?

|  |  |
| --- | --- |
| SELECT Покупатель, Сумма, ндок  FROM Документы as Продажи  WHERE Сумма =  (  SELECT MAX(Сумма)  FROM Документы  GROUP BY Покупатель  ) | SELECT Покупатель, Сумма, ндок  FROM Документы as Продажи  WHERE Сумма IN  (  SELECT MAX(Сумма)  FROM Документы  GROUP BY Покупатель  ) |

В первом ошибка синтаксиса: так как «равно» предполагает наличие одного значения. Второй запрос синтаксически правильный, но в подзапросе получаем максимальное значение для каждого покупателя, значит, во внешнем запросе выведем всех покупателей, которые купили на такую сумму, но это не значит, что это обязательно их максимальная сумма покупки.

Пример: покупатель 1 купить на 100 и 200 уе, покупатель 2 купил на 100 уе. В результате второго запроса Вы выведете все 3 покупки, а правильно вывести только две.

Правильный запрос:

SELECT Покупатель, Сумма, ндок

FROM Документы as Продажи

WHERE Сумма =

(

SELECT MAX(Сумма)

FROM Документы

WHERE Документы.Покупатель = Продажи.Покупатель

)

Заметим, что если у одного контрагента имеется несколько документов на одну максимальную сумму, то в результат запроса попадут все эти документы. Кроме того, заметим, что в подзапросе можно использовать поле из внешнего запроса. Это пример зависимого (связанного,correlated) подзапроса.

Как понять, какой результат выведет SQL Server при наличии зависимого (связанного) подзапроса? Логическая последовательность обработки следующая:

* Выбирается строка из таблицы, имя которой указано во внешнем запросе.
* Выполняется подзапрос и полученное значение применяется для анализа этой строки в условии предложения WHERE внешнего запроса.
* По результату оценки этого условия принимается решение о включении или не включении строки в состав выходных данных.
* Процедура повторяется для следующей строки таблицы внешнего запроса.

Еще возможный вариант:

SELECT Продажи.Покупатель, Продажи.Сумма, ндок

FROM Документы as Продажи INNER JOIN

(

SELECT Max(Сумма) as MaxSumm, Покупатель as Пок

FROM Документы

GROUP BY Покупатель

) as MaxT ON

Продажи.Покупатель = MaxT.Пок AND

Продажи.Сумма = MaxT.MaxSumm

***Использование подчиненных запросов в предложении HAVING.***

Пример:

Подсчет количества товаров с группировкой по массе, причем таких, что масса больше средней массы по всей номенклатуре.

SELECT Масса, COUNT (Товар\_ID)

FROM Номенклатура

GROUP BY Масса

HAVING Масса > (SELECT Avg(Масса) FROM Номенклатура)

1. ***Использование предикатов***

**Предикат EXISTS().**

Назначение предиката Exists состоит в проверке того, что множество не является пустым.

Предикат генерирует значения TRUE или FALSE в зависимости от выходных данных подзапроса. Используя подзапросы в качестве аргумента, этот оператор оценивает результат выполнения подзапроса как истинный, если этот подзапрос генерирует выходные данные, то есть в случае *существования (возврата)* хотя бы одного найденного значения. В противном случае результат подзапроса ложный.

Пример: поиск товара, который не отпускался в течение месяца.

SELECT Товар

FROM Номенклатура

WHERE not exists

(

SELECT \*

FROM Документы INNER JOIN Документы\_данные ON

Документы.ндок = Документы\_данные.ндок

WHERE Документы\_данные.Товар = Номенклатура.Товар

AND Документы.Дата > ‘20171101’

)

В принципе этот же запрос можно написать другими способами (поэкспериментируйте с большими объемами данных, какая версия сработает лучше?):

1. Запрос с использованием предиката IN.

SELECT Товар

FROM Номенклатура

WHERE Товар not in

(

SELECT Товар

FROM Документы INNER JOIN Документы\_данные ON

Документы.ндок = Документы\_данные.ндок

WHERE Документы.Дата > ‘20171101’

)

1. Запрос с использованием LEFT JOIN

SELECT Номенклатура.Товар

FROM Номенклатура LEFT JOIN

(SELECT distinct Товар

FROM Документы INNER JOIN Документы\_данные ON

Документы.ндок = Документы\_данные.ндок

WHERE Документы.Дата > ‘20171101’

) as Продажи ON Продажи.Товар = Номенклатура.Товар

WHERE Продажи.Товар is NULL

**Пример с EXISTS:**

У Вас есть таблица с натуральными числами T. Некоторые числа пропущены, Вам нужно найти минимальное пропущенное число.

CREATE TABLE #T (num INT NOT NULL PRIMARY KEY CHECK(num > 0));

INSERT INTO #T VALUES (3), (4), (6), (7)

-- Вариант с EXISTS

SELECT CASE WHEN EXISTS(SELECT \* FROM #T WHERE num = 1) THEN missing ELSE 1 END

FROM

(

SELECT MIN(A.num) + 1 as missing

FROM #T AS A

WHERE NOT EXISTS

(

SELECT \* FROM #T AS B

WHERE B.num = A.num + 1

)

) T

-- Вариант с LEFT JOIN

SELECT CASE WHEN (SELECT MIN(num) FROM #T) = 1 THEN missing ELSE 1 END

FROM

(

SELECT MIN(A.num + 1) missing

FROM #T A LEFT JOIN #T B ON

A.num + 1 = B.num

WHERE B.num is NULL

) T

-- Вариант решения проблемы с недостающей единицей

DECLARE @T TABLE (num int)

INSERT INTO @T

SELECT num

FROM #T

INSERT INTO @T

VALUES (0)

SELECT MIN(A.num + 1) missing

FROM @T A LEFT JOIN @T B ON

A.num + 1 = B.num

WHERE B.num is NULL

DROP TABLE #T

**Операторы сравнения с множеством значений IN, ANY, ALL**

|  |  |
| --- | --- |
| IN | Равно любому из значений, полученных во внутреннем запросе |
| NOT IN | Не равно ни одному из значений, полученных во внутреннем запросе |
| = ANY | То же, что и IN. Соответствуют логическому оператору OR |
| >ANY, >=ANY | Больше, чем (либо больше или равно) любое (какого либо) полученное число. Эквивалентно > или >= для самого меньшего полученного числа. |
| <ANY, <=ANY | Меньше, чем (либо меньше или равно) любое (какого либо) полученное число. Эквивалентно < или <= для самого большого полученного числа. |
| =ALL | Равно всем полученным значениям. Эквивалентно логическому оператору AND. |
| >ALL, >=ALL | Больше, чем (либо больше или равно) все полученные числа. Эквивалент > или > = для самого большего полученного числа. |
| <ALL, <=ALL | Меньше, чем (либо меньше или равно) все полученные числа. Эквивалент < или < = для самого меньшего полученного числа. |

Примеры запросов:

DECLARE @T1 TABLE (num int)

INSERT INTO @T1 VALUES (2), (4), (6)

DECLARE @T2 TABLE (num int)

INSERT INTO @T2 VALUES (1), (3), (5)

SELECT \*

FROM @T1

WHERE num > any (SELECT num FROM @T2)

SELECT \*

FROM @T1

WHERE num > all (SELECT num FROM @T2)

**Особенности применения операторов ANY, ALL, Exists при обработке пустых значений (NULL).**

Когда правильный подзапрос не генерирует никаких выходных данных, оператор ALL автоматически принимает значение TRUE, а оператор ANY - FALSE.

**Дополнение**: подумайте, что выведет следующий запрос:

SELECT \*

FROM Товары

WHERE Масса NOT IN (1, 3, NULL)

Как получить сотрудников, которые не являются начальниками?

DECLARE @Empl TABLE (empl\_id int)

INSERT INTO @Empl VALUES (1), (2), (3)

DECLARE @Deps TABLE (empl\_id int, boss\_id int)

INSERT INTO @Deps VALUES

(1, 2), (2, 3), (3, NULL)

Вариант, который требует «допиливания»:

SELECT \*

FROM @Empl E

WHERE empl\_id not in

(SELECT boss\_ID FROM @deps)

Как его нужно исправить, чтобы этот вариант работал верно?

SELECT \*

FROM @Empl E

WHERE NOT Exists (SELECT \* FROM @deps WHERE boss\_id = E.empl\_id)

***Использование single-valued запросов в SELECT.***

Задача: для каждого факта продажи товара проставить остаток товара на этот момент по таблице Остатки:

Возможный вариант:

SELECT Док.ндок, Док.Дата, Док.Покупатель, Док.Сумма, Днн.Товар\_ID, Днн.Колво, Днн.Цена,

(SELECT TOP 1 Остаток FROM Остатки WHERE Остатки.Товар\_ID = Днн.Товар\_ID AND Остатки.Дата <= Док.Дата ORDER BY Остатки.Дата desc) Остаток

FROM Документы Док INNER JOIN Документы\_данные Днн ON

Док.ндок = Днн.ндок

***Комментарий:***

Осторожно обращайтесь с названиями полей. Лучше всего в случае зависимого (связанного) запроса обращаться к названиям полей с указанием таблицы.

**Пример: Что выдаст запрос?**

DECLARE @Покупатели TABLE (Пок\_ID int primary key, ФИО nvarchar(max))

INSERT INTO @Покупатели VALUES (1, 'test1'), (2, 'test2')

DECLARE @Документы TABLE (ндок int primary key, ПокID int)

INSERT INTO @Документы VALUES (1, 1)

SELECT \*

FROM @Покупатели

WHERE Пок\_ID in (SELECT distinct Пок\_ID FROM @Документы)

Дополнительные задачи:

1. Как для каждого акта продажи товара получить, сколько этого товара было продано за месяц до?

SELECT Док.ндок, Док.Дата, Днн.Товар\_ID, Днн.Колво,

(

SELECT SUM(ДннВн.Колво)

FROM Документы ДокВн INNER JOIN Документы\_данные ДннВн ON

ДокВн.ндок = ДннВн.ндок

WHERE ДокВн.Дата >= dateadd(m, -1, Док.Дата) AND ДокВн.Дата < Док.Дата AND ДннВн.Товар\_ID = Днн.Товар\_ID

) ПродажиПредМесяц

FROM Документы Док INNER JOIN Документы\_данные Днн ON

Док.ндок = Днн.ндок

1. Как вывести рейтинг каждого покупателя, используя подзапрос в SELECT?

DECLARE @Покупатели TABLE (id int, s float)

INSERT INTO @Покупатели

VALUES (1, 100), (2, 50), (3, 50), (4, 10)

SELECT id, s, (SELECT COUNT(\*) FROM @Покупатели П2 WHERE П1.s <= П2.s) rk

FROM @Покупатели П1

**Литература.**

1. И.Ф.Астахова, А.П.Толстобров, В.М.Мельников «SQL в примерах и задачах»
2. Т.М.Дадашев, В.Ю.Рубаев, О.Л.Белоусов, Д.Р.Гончар «Введение в реляционные базы данных и язык SQL»
3. Help в Access
4. MSDN