

АНАЛИТИЧЕСКИЕ ФУНКЦИИ

Сичная Мария, начальник бюро платежных приложений

Функция LISTAGG объединяет значения `measure_column` для каждой группы на основе `order_by_clause`. Возвращает string значение.

LISTAGG (`measure_column` [, 'delimiter'])
WITHIN GROUP (`order_by_clause`) [OVER `query_partition_clause`]

- `measure_column` столбец, значения которого вы хотите объединить вместе в наборе результатов. Нулевые значения в `measure_column` игнорируются.
- `delimiter` не является обязательным. Этот разделитель используется при разделении значений `measure_column` при выводе результатов.
- `order_by_clause` определяет порядок связанных значений (т.е.: `measure_column`), которые возвращаются.

```
SELECT ENAME  
FROM EMP  
WHERE DEPTNO = 10  
ORDER BY ENAME;
```

ENAME
CLARK
KING
MILLER

```
SELECT DEPTNO, LISTAGG(ENAME, ', ') WITHIN GROUP  
(ORDER BY ENAME) Employees  
FROM EMP GROUP BY DEPTNO;
```

DEPTNO	EMPLOYEES
10	CLARK, KING, MILLER
20	ADAMS, FORD, JONES, SCOTT, SMITH
30	ALLEN, BLAKE, JAMES, MARTIN, TURNER, WARD

[OFFSET offset ROWS]
FETCH NEXT [row_count | percent PERCENT]
ROWS [ONLY | WITH TIES]

Предложение OFFSET указывает количество строк, которые нужно пропустить до начала ограничения строк. Предложение OFFSET не является обязательным. Если вы его пропустите, смещение будет равно 0, и ограничение строки начнется с первой строки.

Смещение должно быть числом или выражением, результатом которого должно быть число. Смещение подчиняется следующим правилам:

- Если смещение отрицательное, оно рассматривается как 0.
- Если смещение равно ПУСТО (NULL) или больше, чем количество строк, возвращаемых запросом, строка не возвращается.
- Если смещение включает дробную часть, дробная часть усекается.

Предложение FETCH определяет количество возвращаемых строк или процент строк.

Для семантической ясности вы можете использовать ключевое слово ROW вместо ROWS, FIRST вместо NEXT. Например, следующие пункты ведут себя одинаково:

- FETCH NEXT 1 ROWS
- FETCH FIRST 1 ROW

ONLY возвращает точно количество строк или процент строк после FETCH NEXT (или FIRST).

WITH TIES возвращает дополнительные строки с тем же ключом сортировки, что и последняя выбранная строка. Обратите внимание: если вы используете WITH TIES, вы должны указать в запросе предложение ORDER BY. Если вы этого не сделаете, запрос не вернет дополнительные строки.

Следующий запрос возвращает 5 сотрудников с наибольшим уровнем оклада:

```
SELECT ENAME, SAL  
FROM EMP  
ORDER BY SAL DESC  
FETCH NEXT 5 ROWS ONLY;
```

ENAME	SAL
KING	5000
SCOTT	3000
FORD	3000
JONES	2975
BLAKE	2850

В следующем запросе используется условие ограничения строки с параметром WITH TIES:

```
SELECT ENAME, SAL  
FROM EMP  
ORDER BY SAL DESC  
FETCH NEXT 2 ROWS WITH TIES;
```

ENAME	SAL
KING	5000
SCOTT	3000
FORD	3000

В следующем запросе используется условие ограничения строки с параметром WITH TIES:

```
SELECT ENAME, SAL  
FROM EMP  
ORDER BY SAL DESC  
FETCH NEXT 2 ROWS WITH TIES;
```

ENAME	SAL
KING	5000
SCOTT	3000
FORD	3000

Несмотря на то, что запрос запрашивал 2 строки, поскольку у него была опция WITH TIES, запрос вернул еще одну дополнительную строку. Обратите внимание, что эта дополнительная строка имеет то же значение в столбце оклада, что и строка 2.

FETCH ВМЕСТО ROWNUM (ORACLE 12C)

Следующий запрос возвращает 35% сотрудников с наибольшим уровнем оклада:

```
SELECT ENAME, SAL  
FROM EMP  
ORDER BY SAL DESC  
FETCH FIRST 35 PERCENT ROWS ONLY;
```

ENAME	SAL
KING	5000
FORD	3000
SCOTT	3000
JONES	2975
BLAKE	2850

Таблица EMP имеет 14 строк, поэтому 35% от 14 составляет 4,9, что округляется до 5 (строк).

FETCH ВМЕСТО ROWNUM (ORACLE 12C)

Следующий запрос пропускает первые 5 сотрудников с наибольшим уровнем оклада и возвращает следующие 5:

```
SELECT ENAME, SAL  
FROM EMP  
ORDER BY SAL DESC  
OFFSET 5 ROWS  
FETCH NEXT 5 ROWS ONLY;
```

ENAME	SAL
CLARK	2450
ALLEN	1600
TURNER	1500
MILLER	1300
WARD	1250

Оконные функции не изменяют выборку, а только добавляют некоторую дополнительную информацию о ней. Т.е. для простоты понимания можно считать, что Oracle сначала выполняет весь запрос (кроме сортировки), а потом только просчитывает оконные выражения.

Окно — это некоторое выражение, описывающее набор строк, которые будет обрабатывать функция и порядок этой обработки.

Причем окно может быть просто задано пустыми скобками (), т.е. окном являются все строки результата запроса.

**имя_функции(<аргумент>, <аргумент>, . . .) over
(<конструкция_фрагментации>
<конструкция_упорядочения> <конструкция_окна>)**

Синтаксис для задания **<конструкции_фрагментации>** выглядит следующим образом:

partition by выражение [, выражение] [, выражение]

Данная конструкция логически разбивает результирующее множество на N групп по критериям, задаваемым выражениями фрагментации. Аналитические функции применяются к каждой группе независимо, – для каждой новой группы они сбрасываются. Если не указать конструкцию фрагментации, все результирующее множество считается одной группой.

```
SELECT ENAME, DEPTNO, SAL,  
       AVG(SAL) OVER (PARTITION BY DEPTNO)  
FROM EMP;
```

ENAME	DEPTNO	SAL	AVG(SAL)OVER(PARTITIONBYDEPTNO)
CLARK	10	2450	2916.6666666666666666666666666667
KING	10	5000	2916.6666666666666666666666666667
MILLER	10	1300	2916.6666666666666666666666666667
JONES	20	2975	2175
FORD	20	3000	2175
ADAMS	20	1100	2175
SMITH	20	800	2175
SCOTT	20	3000	2175
WARD	30	1250	1566.6666666666666666666666666667
TURNER	30	1500	1566.6666666666666666666666666667
ALLEN	30	1600	1566.6666666666666666666666666667
JAMES	30	950	1566.6666666666666666666666666667
BLAKE	30	2850	1566.6666666666666666666666666667
MARTIN	30	1250	1566.6666666666666666666666666667

имя_функции(<аргумент>, <аргумент>, . . .) over
(<конструкция_фрагментации>
<конструкция_упорядочения> <конструкция_окна>)

<Конструкция_упорядочения> имеет следующий синтаксис:

order by выражение [asc|desc] [nulls first|nulls last]

Конструкция **order by** задает критерий сортировки данных в каждой группе (в каждом фрагменте). Это, несомненно, влияет на результат выполнения любой аналитической функции. При наличии (или отсутствии) конструкции **order by** аналитические функции вычисляются по-другому.

SELECT ENAME, SAL, AVG(SAL) OVER ()

FROM EMP;

ENAME	SAL	AVG(SAL)OVER()
SMITH	800	2073.214285714285714285714285714286
ALLEN	1600	2073.214285714285714285714285714286
WARD	1250	2073.214285714285714285714285714286
JONES	2975	2073.214285714285714285714285714286
MARTIN	1250	2073.214285714285714285714285714286
BLAKE	2850	2073.214285714285714285714285714286
CLARK	2450	2073.214285714285714285714285714286
SCOTT	3000	2073.214285714285714285714285714286
KING	5000	2073.214285714285714285714285714286
TURNER	1500	2073.214285714285714285714285714286
ADAMS	1100	2073.214285714285714285714285714286
JAMES	950	2073.214285714285714285714285714286
FORD	3000	2073.214285714285714285714285714286
MILLER	1300	2073.214285714285714285714285714286

```
SELECT ENAME, SAL,  
       AVG(SAL) OVER  
         (ORDER BY ENAME)  
FROM EMP;
```

ENAME	SAL	Avg(Sal) Over (OrderByEname)
ADAMS	1100	1100
ALLEN	1600	1350
BLAKE	2850	1850
CLARK	2450	2000
FORD	3000	2200
JAMES	950	1991.66666666666666666666666666666666666667
JONES	2975	2132.142857142857142857142857142857142857
KING	5000	2490.625
MARTIN	1250	2352.777777777777777777777777777777777778
MILLER	1300	2247.5
SCOTT	3000	2315.9090909090909090909090909090909091
SMITH	800	2189.583333333333333333333333333333333333
TURNER	1500	2136.538461538461538461538461538461538462
WARD	1250	2073.214285714285714285714285714285714286

имя_функции(<аргумент>, <аргумент>, . . .) over
<конструкция_фрагментации>
<конструкция_упорядочения> <конструкция_окна>)

<Конструкция_окна> позволяет задать перемещающееся или жестко привязанное окно (набор) данных в пределах группы, с которым будет работать аналитическая функция.

Можно создавать окна по двум критериям: по диапазону (RANGE) значений данных или по смещению (ROWS) относительно текущей строки. Использование конструкции **range** в некоторых случаях используется неявно, RANGE UNBOUNDED PRECEDING например. Она требует брать все строки вплоть до текущей, в соответствии с порядком, задаваемым конструкцией **order by**.

Окно определяется диапазоном строк, объединяемых в соответствии с заданным порядком. Применять конструкцию range можно либо с числовыми выражениями (NUMBER), либо с выражениями, значением которого является дата (DATE). Еще одно ограничение для таких окон состоит в том, что в конструкции order by может быть только один столбец – диапазоны по природе своей одномерны. Нельзя задать диапазон в N-мерном пространстве.

Пример. Пусть необходимо выбрать зарплату каждого сотрудника и среднюю зарплату всех принятых на работу в течение 100 предыдущих дней, а также среднюю зарплату всех принятых на работу в течение 100 следующих дней. Соответствующий запрос будет выглядеть так:

```
SELECT ENAME, HIREDATE, SAL,  
       AVG(SAL) OVER (ORDER BY HIREDATE ASC RANGE 100  
PRECEDING) AVG_SAL_100_DAYS_BEFORE,  
       AVG(SAL) OVER (ORDER BY HIREDATE DESC RANGE 100  
PRECEDING) AVG_SAL_100_DAYS_AFTER  
FROM EMP ORDER BY HIREDATE DESC
```

[illegible]

Помимо определения окна по диапазону (RANGE), также окна определяются и по количеству строк (ROWS). Для окон по строкам нет ограничений, присущих окнам по диапазону; данные могут быть любого типа и упорядочивать можно по любому количеству столбцов.

Например, пусть нужно вычислить среднюю зарплату для сотрудника и пяти принятых на работу до него и после него. Запрос можно записать следующим образом:

```
SELECT ENAME, HIREDATE, SAL,  
       AVG(SAL) OVER (ORDER BY HIREDATE ASC ROWS 5  
PRECEDING) AVG_5_BEFORE,  
       AVG(SAL) OVER (ORDER BY HIREDATE DESC ROWS 5  
PRECEDING) AVG_5_AFTER  
FROM EMP ORDER BY HIREDATE
```

[illegible]

Зная, как определяются окна (по диапазону или по количеству строк), рассмотрим, как окончательно задаются окна. В простейшем случае, окно задается с помощью одной из следующих взаимоисключающих конструкций:

- **UNBOUNDED PRECEDING**. Окно начинается с первой строки текущей группы и заканчивается текущей обрабатываемой строкой;
- **CURRENT ROW**. Окно начинается (и заканчивается) текущей строкой. Стоит отметить, что окно **CURRENT ROW** в простейшем виде, вероятно, никогда не используется, поскольку ограничивает применение аналитической функции одной текущей строкой, а для этого аналитические функции не нужны. В более сложном случае для окна задается также конструкция **BETWEEN**. В ней **CURRENT ROW** можно указывать в качестве начальной или конечной строки окна;

- Числовое_выражение PRECEDING. Окно начинается со строки за числовое_выражение строк до текущей, если оно задается по строкам, или со строки, меньшей по значению столбца, упомянутого в конструкции order by, не более чем на числовое выражение, если оно задается по диапазону;
- Числовое_выражение FOLLOWING. Окно заканчивается (или начинается) со строки, через числовое_выражение строк после текущей, если оно задается по строкам, или со строки, большей по значению столбца, упомянутого в конструкции order by, не более чем на числовое_выражение, если оно задается по диапазону.


```
SELECT ENAME, HIREDATE,
       FIRST_VALUE(ENAME) OVER (ORDER BY HIREDATE ASC
                                RANGE BETWEEN 100 PRECEDING AND 100 FOLLOWING),
       LAST_VALUE(ENAME) OVER (ORDER BY HIREDATE ASC
                                RANGE BETWEEN 100 PRECEDING AND 100 FOLLOWING)
FROM EMP ORDER BY HIREDATE ASC
```

ENAME	HIREDATE	FIRST_VALUE(ENAME)OVER(ORDERBYHIREDATEASC RANGE BETWEEN 100 PRECEDING AND 100 FOLLOWING)	LAST_VALUE(ENAME)OVER(ORDERBYHIREDATEASC RANGE BETWEEN 100 PRECEDING AND 100 FOLLOWING)
SMITH	17-DEC-80	SMITH	WARD
ALLEN	20-FEB-81	SMITH	BLAKE
WARD	22-FEB-81	SMITH	BLAKE
JONES	02-APR-81	ALLEN	CLARK
BLAKE	01-MAY-81	ALLEN	CLARK
CLARK	09-JUN-81	JONES	TURNER
TURNER	08-SEP-81	CLARK	JAMES
MARTIN	28-SEP-81	TURNER	JAMES
KING	17-NOV-81	TURNER	MILLER
FORD	03-DEC-81	TURNER	MILLER
JAMES	03-DEC-81	TURNER	MILLER
MILLER	23-JAN-82	KING	MILLER
SCOTT	09-DEC-82	SCOTT	ADAMS
ADAMS	12-JAN-83	SCOTT	ADAMS

```
SELECT LEVEL,  
       COUNT(*) OVER (ORDER BY LEVEL ASC ROWS 2  
PRECEDING) ASC_COUNT,  
       COUNT(*) OVER (ORDER BY LEVEL DESC ROWS 2  
PRECEDING) DESC_COUNT  
FROM DUAL CONNECT BY LEVEL <= 10 ORDER BY LEVEL
```

LEVEL	ASC_COUNT	DESC_COUNT
1	1	3
2	2	3
3	3	3
4	3	3
5	3	3
6	3	3
7	3	3
8	3	3
9	3	2
10	3	1

Окно rows 2 preceding, как видно из результата запроса, содержит от 1 до 3 строк (это определяется тем, как далеко текущая строка находится от начала группы). Для первой строки группы имеем значение 1 (предыдущих строк нет). Для следующей строки в группе таких строк 2. Наконец, для третьей и далее строк значение count(*) остается постоянным, поскольку мы считаем только текущую строку и две предыдущие.

```
SELECT N, DAYS,  
       SUM(N) OVER (ORDER BY DAYS RANGE 2 PRECEDING) N_SUM  
FROM  
(SELECT LEVEL N,  
         TO_DATE('10.01.2014','DD.MM.YYYY')+(LEVEL - 1) DAYS  
FROM DUAL CONNECT BY LEVEL <= 10  
)  
ORDER BY DAYS
```

В данном случае идет суммирование в пределах окна, диапазон которого составляет скользящий день и 2 предыдущих дня.

N	DAYS	N_SUM
1	10-JAN-14	1
2	11-JAN-14	3
3	12-JAN-14	6
4	13-JAN-14	9
5	14-JAN-14	12
6	15-JAN-14	15
7	16-JAN-14	18
8	17-JAN-14	21
9	18-JAN-14	24
10	19-JAN-14	27

Он присваивает уникальный номер каждой строке, к которой он применяется (либо каждой строке в разделе, либо каждой строке, возвращаемой запросом), в упорядоченной последовательности строк, указанной в предложении `order_by_clause`, начиная с 1.

```
SELECT DEPTNO, ENAME, EMPNO,  
       ROW_NUMBER() OVER (PARTITION BY DEPTNO ORDER BY  
                           EMPNO) EMP_ID  
FROM EMP;
```

DEPTNO	ENAME	EMPNO	EMP_ID
10	CLARK	7782	1
10	KING	7839	2
10	MILLER	7934	3
20	SMITH	7369	1
20	JONES	7566	2
20	SCOTT	7788	3
20	ADAMS	7876	4
20	FORD	7902	5
30	ALLEN	7499	1
30	WARD	7521	2
30	MARTIN	7654	3
30	BLAKE	7698	4
30	TURNER	7844	5
30	JAMES	7900	6

Функция RANK возвращает ранг значения в группе значений, причем может возвращать не последовательное ранжирование, если тестируемые значения одинаковы.

```
SELECT DEPTNO, ENAME, SAL, RANK () OVER (PARTITION BY  
DEPTNO ORDER BY SAL) EMP_ID  
FROM EMP;
```

DEPTNO	ENAME	SAL	EMP_ID
10	MILLER	1300	1
10	CLARK	2450	2
10	KING	5000	3
20	SMITH	800	1
20	ADAMS	1100	2
20	JONES	2975	3
20	SCOTT	3000	4
20	FORD	3000	4
30	JAMES	950	1
30	MARTIN	1250	2
30	WARD	1250	2
30	TURNER	1500	4
30	ALLEN	1600	5
30	BLAKE	2850	6

Функция DENSE_RANK возвращает ранг строки в упорядоченной группе строк.

```
SELECT DEPTNO, ENAME, SAL, DENSE_RANK () OVER  
(PARTITION BY DEPTNO ORDER BY SAL) EMP_ID  
FROM EMP;
```

DEPTNO	ENAME	SAL	EMP_ID
10	MILLER	1300	1
10	CLARK	2450	2
10	KING	5000	3
20	SMITH	800	1
20	ADAMS	1100	2
20	JONES	2975	3
20	SCOTT	3000	4
20	FORD	3000	4
30	JAMES	950	1
30	MARTIN	1250	2
30	WARD	1250	2
30	TURNER	1500	3
30	ALLEN	1600	4
30	BLAKE	2850	5

Аналитическая функция RATIO_TO_REPORT отображает отношение указанного значения к сумме значений в наборе. Он не поддерживает условия порядка или окон. Отсутствие предложения о секционировании в предложении OVER означает, что весь набор результатов рассматривается как один раздел.

```
SELECT ENAME, SAL,  
       RATIO_TO_REPORT(SAL) OVER () AS SALSHARE  
FROM EMP;
```

ENAME	SAL	SALSHARE
SMITH	800	.0275624461670973298880275624461670973299
ALLEN	1600	.0551248923341946597760551248923341946598
WARD	1250	.043066322136089577950043066322136089578
JONES	2975	.1024978466838931955211024978466838931955
MARTIN	1250	.043066322136089577950043066322136089578
BLAKE	2850	.0981912144702842377260981912144702842377
CLARK	2450	.0844099913867355727820844099913867355728
SCOTT	3000	.1033591731266149870801033591731266149871
KING	5000	.1722652885443583118001722652885443583118
TURNER	1500	.0516795865633074935400516795865633074935
ADAMS	1100	.0378983634797588285960378983634797588286
JAMES	950	.0327304048234280792420327304048234280792
FORD	3000	.1033591731266149870801033591731266149871
MILLER	1300	.0447889750215331610680447889750215331611

Функция SUM вычисляет общую сумму значений выражения для группы.

```
SELECT ENAME, DEPTNO, SAL,  
       SUM(SAL) OVER (PARTITION BY DEPTNO ORDER BY ENAME)  
FROM EMP;
```

ENAME	DEPTNO	SAL	SUM(SAL) OVER (PARTITION BY DEPTNO ORDER BY ENAME)
CLARK	10	2450	2450
KING	10	5000	7450
MILLER	10	1300	8750
ADAMS	20	1100	1100
FORD	20	3000	4100
JONES	20	2975	7075
SCOTT	20	3000	10075
SMITH	20	800	10875
ALLEN	30	1600	1600
BLAKE	30	2850	4450
JAMES	30	950	5400
MARTIN	30	1250	6650
TURNER	30	1500	8150
WARD	30	1250	9400

Функция LAG аналитическая функция, которая позволяет запрашивать более одной строки в таблице, в то время, не имея присоединенной к себе таблицы. Это возвращает значения из предыдущей строки в таблице. Для возврата значения из следующего ряда, попробуйте использовать функцию LEAD.

```
SELECT ENAME, DEPTNO, HIREDATE,  
       LAG(HIREDATE, 1) OVER (PARTITION BY DEPTNO ORDER BY  
HIREDATE) PREVEMPL,  
       LEAD(HIREDATE, 1) OVER (PARTITION BY DEPTNO ORDER BY  
HIREDATE) NEXTEMP  
FROM EMP WHERE DEPTNO = 20;
```

ENAME	DEPTNO	HIREDATE	PREVEMPL	NEXTEMP
SMITH	20	17-DEC-80	-	02-APR-81
JONES	20	02-APR-81	17-DEC-80	03-DEC-81
FORD	20	03-DEC-81	02-APR-81	09-DEC-82
SCOTT	20	09-DEC-82	03-DEC-81	12-JAN-83
ADAMS	20	12-JAN-83	09-DEC-82	-

СПАСИБО ЗА ВНИМАНИЕ !
ВОПРОСЫ ?

www.compassplus.ru