

# АНАЛИТИЧЕСКИЕ ФУНКЦИИ В ORACLE

Е.С. Давыдов

# Предложение **group by**

[GROUP BY выражение группировки]

[HAVING условие отбора групп]

# GROUP BY с ROLLUP

Функция `rollup`, добавляет итоговые данные по промежуточным уровням группировки

Синтаксис:

```
[GROUP BY rollup(поля группировки)]
```

Например:

```
GROUP BY rollup(a,b,c)
```

Это группировки по (a,b,c), (a,b), (a), общий ИТОГ

# GROUP BY с CUBE

Функция cube, добавляет итоговые данные по промежуточным уровням группировки

Синтаксис:

```
[GROUP BY cube(поля группировки)]
```

Например:

```
GROUP BY cube(a,b,c)
```

Это группировки по (a,b,c), (a,b), (a,c), (b,c), (a), (b), (c), общий итог

# Классификация

- Функции ранжирования
- Статистические функции для плавающего интервала
- Функции подсчета долей и итогов
- Статистические функции LAG/LEAD с запаздывающим/ опережающим аргументом
- Статистические функции

# Синтаксис

ИМЯ\_ФУНКЦИИ (<аргумент>, <аргумент>,  
...)

OVER

(<конструкция\_фрагментации>

<конструкция\_упорядочения>

<конструкция\_окна>)

# Аналитические функции

Функция	Назначение
AVG([DISTINCT   ALL] выражение)	Среднее значение в пределах группы или окна
CORR(выражение, выражение)	Коэффициент корреляции для пары выражений, возвращающих числовые значения
COUNT([DISTINCT] *   выражение)	Считает количество значений в выражении или количество строк
COVAR_POP(выражение, выражение)	Ковариация генеральной совокупности пары выражений с числовыми значениями
CAVAR_SAMP(выражение, выражение)	Возвращает выборочную ковариацию пары выражений с числовыми значениями
CUME_DIST	Вычисляет относительную позицию строки в группе. Возвращает число в диапазоне от 0 до 1.

# Аналитические функции

Функция	Назначение
DENSE_RANK	Относительный “плотный” ранг каждой возвращаемой строки по отношению к другим строкам, основываясь на значениях выражений конструкции ORDER BY
FIRST_VALUE	Первое значение в группе
LAG(выражение, <смещение>, <стандартное значение>)	Позволяет ссылаться на строки, предшествующие текущей строке в группе.
LEAD(выражение, <смещение>, <стандартное значение>)	Позволяет ссылаться на строки, последующие за текущей строкой в группе.
<смещение>	1 предыдущая/последующая строка
<стандартное значение>	Возвращается, если индекс выходит за пределы окна
LAST_VALUE	Последнее значение в группе



# Аналитические функции

Функция	Назначение
MAX(выражение)	Максимальное значение выражения в пределах окна или группы
MIN(выражение)	Минимальное значение выражения в пределах окна или группы
NTILE(выражение)	Делит группу на фрагменты по значению выражения
PERCENT_RANK	Аналогична функции CUME_DIST. Вычисляет ранг строки в группе минус 1, деленный на количество обрабатываемых строк минус 1.
RANK	Относительный ранг каждой возвращаемой строки по отношению к другим строкам, основываясь на значениях выражений конструкции ORDER BY

# Аналитические функции

Функция	Назначение
RATIO_TO_REPORT(выражение)	Вычисляет значение отношения выражения к сумме выражений по строкам группы
REGR_XXXXXX(выражение, выражение)	Функции линейной регрессии, 9 различных функций регрессии
ROW_NUMBER	Смещение строки по отношению к началу упорядоченной группы
STDDEV(выражение)	Стандартное (среднеквадратичное) отклонение текущей строки по отношению к группе
STDDEV_POP(выражение)	Стандартное отклонение генеральной совокупности и возвращает квадратный корень из дисперсии генеральной совокупности
STDDEV_SAMP(выражение)	Накопленное стандартное отклонение выборки и возвращает квадратный корень выборочной

# Аналитические функции

Функция	Назначение
SUM(выражение)	Общая сумма выражений для группы
VAR_POP(выражение)	Дисперсия генеральной совокупности для набора числовых значений
VAR_SAMP(выражение)	Выборочная дисперсия для набора числовых значений
VARIANCE(выражение)	Возвращает дисперсию для выражения 0, если количество строк в группе =1 VAR_SAMP, если количество строк в группе >1

# Фрагментация

Задается предложением

PARTITION BY выражение [, выражение ...]

Разбивает результирующее множество на  $N$  групп по критериям, задаваемым выражениями фрагментации.

Если не указать фрагментацию, то результирующее множество считается одной группой.

# Фрагментация: примеры

Sample\_script\_01\_x.sql

# Конструкция упорядочивания

ORDER BY выражение [ASC | DESC]  
[NULLS FIRST | NULLS LAST]

Задаёт упорядочивание в каждой группе

При отсутствии конструкции ORDER BY  
аналитическая функция вычисляется по всей  
группе целиком

# Упорядочивание: примеры

Sample\_script\_02\_x.sql

# Выполнение вычислений по плавающему окну

{ROWS | RANGE}

{{UNBOUNDED | выражение}}

PRECEDING | CURRENT ROW}



# Выполнение вычислений по плавающему окну

{ROWS | RANGE} BETWEEN

{{UNBOUNDED PRECEDING | CURRENT  
ROW | {UNBOUNDED | выражение 1}  
{PRECEDING | FOLLOWING }}

AND

{{UNBOUNDED FOLLOWING | CURRENT  
ROW | {UNBOUNDED | выражение 2}  
{PRECEDING | FOLLOWING }}

Фразы PRECEDING и FOLLOWING задают  
верхнюю и нижнюю границы  
агрегирования

# Окна диапазона

Применяются только для типов дата и число

Синтаксис

RANGE выражение PRECEDING

Sample\_Script\_03\_a

# Окна строк

Задаются в физических единицах, строках

Синтаксис

ROWS выражение PRECEDING

Sample\_Script\_03\_b

# Задание окон

- ▣ UNBOUNDED PRECEDING – окно начинается с первой строки текущей группы и заканчивается текущей строкой
- ▣ CURRENT ROW – окно начинается и заканчивается текущей строкой
- ▣ числовое\_выражение PRECEDING – окно начинается за числовое\_выражение строк до текущей (если по строкам) или со строки, меньшей по значению столбца, упомянутого в ORDER BY, не более чем на числовое\_выражение (если по диапазону)

# Задание окон

Конструкция BETWEEN

CURRENT ROW можно указывать как начальную или конечную строку окна

Числовое\_выражение FOLLOWING – окно заканчивается (или начинается) со строки, через числовое\_выражение строк после текущей (если по строкам) или со строки, большей по значению столбца ORDER BY, не более чем на числовое\_выражение (если по диспазону)

# Пример

Sample\_script\_04\_a.sql