|  |  |
| --- | --- |
| *Аналитические функции Oracle* | *Назначение* |
| AVG([DISTINCT | ALL] выражение) | Используется для вычисления среднего значения выражения в пределах группы и окна. Для поиска среднего после удаления дублирующихся значений можно указывать ключевое слово **DISTINCT**. |
| CORR (выражение, выражение) | Выдает коэффициент корреляции для пары выражений, возвращающих числовые значения. Это сокращение для выражения:  COVAR\_POP<выражение1, выражение2> / STDDEV\_POP(выражение1) \* STDDEV\_POP(выражение2)  В статистическом смысле, *корреляция* – это степень связи между переменными. Связь между переменными означает, что значение одной переменной можно в определенной степени предсказать по значению другой. Коэффициент корреляции представляет степень корреляции в виде числа в диапазоне от -1 (высокая обратная корреляция) до 1 (высокая корреляция). Значение 0 соответствует отсутствию корреляции. |
| COUNT ([DISTINCT] [\*] [выражение]) | Эта функция считает строки в группах. Если указать \* или любую константу, кроме **NULL**, функция count будет считать все строки. Если указать выражение, функция count будет считать строки, для которых выражение имеет значение не **NULL**. Можно задать модификатор **DISTINCT**, чтобы считать строки в группах после удаления дублирующихся строк. |
| COVAR\_POP (выражение, выражение) | Возвращает *ковариацию генеральной совокупности* (population covariance) пары выражений с числовыми значениями. |
| COVAR\_SAMP (выражение, выражение) | Возвращает *выборочную ковариацию* (sample covariance) пары выражений с числовыми значениями. |
| CUME\_DIST | Вычисляет относительную позицию строки в группе. Функция **CUME\_DIST** всегда возвращает число, большее 0 и меньшее или равное 1. Это число представляет «позицию» строки в группе из N строк. В группе из трех строк, например, возвращаются следующие значения кумулятивного распределения: 1/3, 2/3, и 1 (3/3). |
| DENSE\_RANK | Эта функция вычисляет относительный ранг каждой возвращаемой запросом строки по отношению к другим строкам, основываясь на значениях выражений в конструкции **ORDER BY**. Данные в группе сортируются в соответствии с конструкцией **ORDER BY**, а затем каждой строке поочередно присваивается числовой ранг, начиная с 1. Ранг увеличивается при каждом изменении значений выражений, входящих в конструкция **ORDER BY**. Строки с одинаковыми значениями получают один и тот же ранг (при этом сравнении значения **NULL** считаются одинаковыми). Возвращаемый этой функцией «плотный» ранг дает ранговые значения без промежутков. Сравните с представленной далее функцией **RANK**. |
| FIRST\_VALUE | Возвращает первое значение в группе. |
| LAG (выражение, <смещение>, <стандартное значение>) | Функция **LAG** дает доступ к другим строкам результирующего множества, избавляя от необходимости выполнять самосоединения. Она позволяет работать с курсором, как с массивом. Можно ссылаться на строки, предшествующие текущей строке в группе. О том, как обращаться к следующим строкам в группе, см. в описании функции **LEAD**.  Смещение – это положительное целое число со стандартным значением 1 (предыдущая строка).  Стандартное значение возвращается, если индекс выходит за пределы окна (для первой строки группы будет возвращено стандартное значение). |
| LAST\_VALUE | Возвращает последнее значение в группе. |
| LEAD (выражение, <смещение>, <стандартное значение>) | Функция **LEAD** противоположна функции **LAG**. Если функция **LAG** дает доступ к предшествующим строкам группы, то функция **LEAD** позволяет обращаться к строкам, следующим за текущей.  Смещение – это положительное целое число со стандартным значением 1 (следующая строка).  Стандартное значение возвращается, если индекс выходит за пределы окна (для последней строки группы будет возвращено стандартное значение). |
| MAX (выражение) | Находит максимальное значение выражения в пределах окна в группе. |
| MIN (выражение) | Находит минимальное значение выражения в пределах окна в группе. |
| NTILE (выражение) | Делит группу на фрагменты по значению выражения. Например, если выражение = 4, то каждой строке в группе присваивается число от 1 до 4 в соответствии с фрагментом, в которую она попадает. Если в группе 20 строк, первые 5 получают значение 1, следующие 5 – значение 2 и т.д. Если количество строк в группе не делиться на значение выражения без остатка, строки распределяются так, что ни в одном фрагменте количество строк не превосходит минимальное количество в других фрагментах более чем на 1, причем дополнительные строки будут в группах с меньшими номерами фрагмента. Например, если снова выражение = 4, а количество строк = 21, в первом фрагменте будет 6 строк, во втором и последующих – 5. |
| PERCENT\_RANK | Аналогична функции **CUME\_DIST** (кумулятивное распределение). Вычисляет ранг строки в группе минус 1, деленный на количество обрабатываемых строк минус 1. Эта функция всегда возвращает значения в диапазоне от 0 до 1 включительно. |
| RANK | Эта функция вычисляет относительный ранг каждой строки, возвращаемой запросом, на основе значений выражений, входящих в конструкцию **ORDER BY**. Данные в группе сортируются в соответствии с конструкцией **ORDER BY**, а затем каждой строке поочередно присваивается числовой ранг, начиная с 1. Ранг увеличивается при каждом изменении значений выражений, входящих в конструкция **ORDER BY**. Строки с одинаковыми значениями выражений получают, входящих в конструкцию **ORDER BY**, получают одинаковый ранг, следующее значение ранга пропускается. Если две строки получили ранг 1, строки с рангом 2 не будет, следующая строка в группе получит ранг 3. В этом отличие от функции **DENSE\_RANK**, которая не пропускает значений. |
| RATIO\_TO\_REPORT (выражение) | Эта функция вычисляет значение: выражение / (sum(выражение)) по строкам группы.  Это дает процент, который составляет значение текущей строки по отношению к sum(выражение). |
| REGR\_xxxxxxx (выражение, выражение) | Эти функции линейной регрессии применяют стандартную линейную регрессию по методу наименьших квадратов к паре выражений. Предлагается 9 различных функций регрессии. |
| ROW\_NUMBER | Возвращает смещение строки по отношению к началу упорядоченной группы. Может использоваться для последовательной нумерации строк, упорядоченных по определенным критериям. |
| STDDEV (выражение) | Вычисляет *стандартное (среднеквадратичное) отклонение* (standard deviation) текущей строки по отношению к группе. |
| STDDEV\_POP (выражение) | Эта функция вычисляет *стандартное отклонение генеральной совокупности* (population standard deviation) и возвращает квадратный корень из дисперсии генеральной совокупности (population variance). Она возвращает значение, совпадающее с квадратным корнем из результата функции **VAR\_POP**. |
| STDDEV\_SAMP (выражение) | Эта функция вычисляет *накопленное стандартное отклонение выборки* (cumulative sample standard deviation) и возвращает квадратный корень выборочной дисперсии (sample variance). Она возвращает значение, совпадающее с квадратным корнем из результата функции **VAR\_SAMP**. |
| SUM (выражение) | Вычисляет общую сумму значений выражения для группы. |
| VAR\_POP (выражение) | Эта функция возвращает дисперсию генеральной совокупности для набора числовых значений (значения NULL игнорируются). Функция VAR\_POP вычисляет значение:  (SUM(выражение\*выражение) –  SUM(выражение)\*SUM(выражение) /  COUNT(выражение)) / COUNT(выражение) |
| VAR\_SAMP | Эта функция возвращает выборочную дисперсию для набора числовых значений (значения NULL игнорируются). Она вычисляет значение:  (SUM(выражение\*выражение) –  SUM(выражение)\*SUM(выражение) /  COUNT(выражение)) / (COUNT(выражение) – 1) |
| VARIANCE (выражение) | Возвращает дисперсию для выражения. Сервер Oracle вычисляет дисперсию как:   * 0, если количество строк в группе = 1; * VAR\_SAMP, если количество строк в группе > 1. |