

The image features two thick black L-shaped bars. One is located in the top-left corner, and the other is in the bottom-right corner. They are positioned such that they appear to frame the central text area.

БАЗЫ ДАННЫХ

Агрегация данных

Агрегатные функции

Агрегация позволяет выбирать данные для аналитических целей, например, для формирования отчетов или построения аналитических графиков.

SQL предлагает несколько агрегатных функций для получения готовой выборки из таблиц.

Агрегатные функции предназначены для вычисления обобщенного группового значения по имеющимся данным.

Подсчет строк

Функция COUNT() подсчитывает строки в выборке.

COUNT(*colname*)

Принимает в качестве аргумента имя столбца, в этом случае подсчет игнорирует NULL-значения в этом столбце.

COUNT(*)

Может принимать * в качестве аргумента, что позволяет произвести подсчет строк вне зависимости от данных в столбцах.

Подсчет уникальных значений

COUNT(DISTINCT *colname*)

С опцией DISTINCT подсчитываются только различающиеся (уникальные) данные.

Подсчет суммы и средних значений

SUM(*colname*)

Функция SUM() суммирует значения заданного столбца (поля).

AVG(*colname*)

Функция AVG() вычисляет среднее значение по заданному столбцу.

Фактически в этой функции происходит комбинация вызовов SUM(*col*) / COUNT(*col*).

Подсчет предельных значений

MAX(*colname*)

Функция MAX() находит максимальное значение столбца.

MIN(*colname*)

Функция MIN() находит минимальное значение столбца.

Сортировка данных

В аналитических целях часто требуется получить неагрегированные данные, упорядоченные по определенному критерию (столбцу).

ORDER BY *colname* -- в прямом порядке

ORDER BY *colname* **ASC** -- в прямом порядке

ORDER BY *colname* **DESC** -- в обратном порядке

Сортировка данных

Возможна сортировка данных по нескольким столбцам одновременно:

ORDER BY *col_a*, *col_b* **DESC**, *col_c* **ASC**

С помощью функции RAND() мы можем получить выборку в случайном порядке:

ORDER BY RAND() — выборка медленная (!)

Группировка данных

Агрегировать выборку можно с помощью группировки данных по заданному критерию (столбцу).

GROUP BY *colname*

GROUP BY *col_a, col_b, col_c*

При группировке необходимо обязательно выбрать тот столбец или те столбцы, по которым происходит группировка.

Выборка прочих столбцов без агрегатных функций не имеет практического смысла.

Группировка данных

Для группированных данных можно применять агрегатные функции.

```
SELECT `age`, COUNT(*)  
FROM `users`  
GROUP BY `age`;
```

Конкатенация группированных данных

Для получения конкатенированной строки в группированных данных используется функция GROUP_CONCAT().

GROUP_CONCAT(`name`)

По умолчанию данные конкатенируются через запятую. В опции SEPARATOR мы можем задать символ (строку) для конкатенации:

GROUP_CONCAT(`name` **SEPARATOR** ' ; ')

Конкатенация группированных данных

Для получения только различных (уникальных) значений в групповой конкатенации используется опция DISTINCT:

```
GROUP_CONCAT(DISTINCT `name`)
```

Опция ORDER BY позволяет упорядочить данные для конкатенации:

```
GROUP_CONCAT(`name` ORDER BY `name`)
```

Примеры агрегации данных

```
SELECT  
    `age`,  
    COUNT(*),  
    MIN(LENGTH(`name`)),  
    AVG(`balance`)  
FROM `users`  
GROUP BY `age`;
```

Примеры агрегации данных

```
SELECT
    `age`,
    GROUP_CONCAT(
        DISTINCT `name`
        ORDER BY `name`
        SEPARATOR ';'
    )
FROM `users`
GROUP BY `age`
ORDER BY `age` DESC;
```