Distribution Window Functions: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2024

Syntax

- The CUME_DIST() Function:

```
CUME_DIST() OVER ( [PARTITION BY expression, ... ]
                   ORDER BY expression [ASC | DESC], ... )
```

- The PERCENT_RANK() Function:

```
PERCENT_RANK() OVER ( [PARTITION BY expression, ... ]
                      ORDER BY expression [ASC | DESC], ... )
```

- The PERCENTILE_CONT() Function:

```
PERCENTILE_CONT(percent) WITHIN GROUP(ORDER BY expression)
```

- The PERCENTILE_DISC() Function:

```
PERCENTILE_DISC(percent) WITHIN GROUP(ORDER BY expression)
```

- The WINDOW Clause:

```
WINDOW window_name AS
(
    [partition_definition]
    [order_definition]
    [frame_definition]
)
```

- the WITHIN GROUP Clause:

```
function_name() WITHIN GROUP (ORDER BY column_expression [ ASC | DESC ])
```

Concepts

- There are two types of distribution window functions in SQL:
    - **Rank distribution functions** are used to calculate the relative rank of a specific row in a window partition based on the ordering specified in the `OVER()` clause.
        - `PERCENT_RANK()` calculates the relative rank of a specific row within a window partition as a percentage value between 0 and 1, with 0 representing the first row and 1 representing the last row in the partition.
        - `CUME_DIST()` computes the cumulative distribution of a specific row within a window partition, which is the proportion of rows with values less than or equal to the current row's value.
    - **Inverse distribution functions** are used to calculate the value at a specified percentile in a group based on the ordering specified in the `WITHIN GROUP()` clause.
        - `PERCENTILE_CONT()` is a continuous inverse distribution function that calculates the value at a specified percentile within a group using linear interpolation.

    - `PERCENTILE_DISC()` is a discrete inverse distribution function that finds the value at a specified percentile within a group, returning the first value with a cumulative distribution greater than or equal to the specified percentile.
- `WITHIN GROUP` is a clause used to specify an ordering expression for aggregate and analytic functions, allowing them to operate on sorted rows within a group instead of the unsorted rows created by the `GROUP BY` clause. It is particularly useful for inverse distribution functions, which require an ordering specification within a group and return a single result per group, such as `PERCENTILE_CONT()` and `PERCENTILE_DISC()`.

References

- PostgreSQL Window Functions
- PostgreSQL Aggregate Functions

Takeaways by Dataquest Labs, Inc. - All rights reserved © 2024