

Using Statistical Functions



Kathi Kellenberger

@aunkathi | aunkathisql.com

Statistical Functions

Top 10!

- PERCENT_RANK – Calculate the rank over the group
- CUME_DIST – Calculate the relative position
- PERCENTILE_CONT – Given a rank, interpolate the value
- PERCENTILE_DISC – Given a rank, find the value

SYNTAX

```
PERCENT_RANK | CUME_DIST()  
OVER([PARTITION BY <expression>] ORDER BY <expression>)
```

Formula

PERCENT_RANK: $(\text{RANK} - 1) / (\text{N} - 1)$

CUME_DIST: RANK / N

Statistical Functions

Top 10!

- PERCENT_RANK – Calculate the rank over the group
- CUME_DIST – Calculate the relative position
- PERCENTILE_CONT – Given a rank, interpolate the value
- PERCENTILE_DISC – Given a rank, find the value

SYNTAX

```
PERCENTILE_CONT | PERCENTILE_DISC(<rank>)  
WITHIN GROUP(ORDER BY <expression>)  
OVER([PARTITION BY <expression>])
```

Median:

```
PERCENTILE_CONT(0.5)...
```

```
PERCENTILE_DISC(0.5)...
```

Real World Examples

Grading Curve

Capacity Planning

Performance



No frame to consider

Index helpful

Worktables in tempdb

Summary



Functions added in 2005

- Ranking functions
- Window aggregates

Functions added in 2012

- Accumulating window aggregates
- Offset functions
- Statistical functions

Summary



Makes writing queries easier

Functions perform over a set of rows

The OVER clause defines the window

- Partition by
- Order by
- Frame

Summary



Covering index

Create worktables on disk

Window aggregates

Statistical functions

Specify frame with ROWS

- Accumulating window aggregates
- FIRST_VALUE and LAST_VALUE

Test, test, test