

DOAG Nürnberg
Thursday November 19th 2015

Analytic Functions 101

Kim Berg Hansen

KIBEHAN



Who is Kim Berg Hansen?



- Blog: <http://www.kibeha.dk>
- Work: <http://www.thansen.dk>
- Professional geek since 1996
- Oracle developer since 2000
- Uses SQL to the max!
- <http://plsqlichallenge.oracle.com>
- <http://www.otechmag.com>
- Beer enthusiast
- Likes to cook
- Reads sci-fi



What will Kim talk about?

- What is so great about analytic functions?
- How do they work?
- Demo showcasing syntax
- Demo of Top selling items report
- Questions?



So, what *is* so great about analytics?

- Normal SQL functions operate on **one** row
- Aggregates can do more rows but loose detail
- When you need details together with subtotals, ranks, ratios, comparisons, you could do:
 - ☹ Client operations (tool or code with variables/arrays)
 - ☹ Scalar subqueries (multiple access of same data)
 - ☺ Analytic functions (often much more efficient ☺)
- Analytics allow you to operate across the entire resultset, not just a single row



Access data from other rows

DEPTNO	ENAME	SAL
30	JAMES	950
30	MARTIN	1250
30	WARD	1250
30	TURNER	1500
30	ALLEN	1600
30	BLAKE	2850

Diagram illustrating data access from other rows:

- Blue arrows show data flow: from WARD to JAMES, from WARD to MARTIN, from WARD to TURNER, and from WARD to ALLEN.
- A blue box highlights the SAL values for JAMES, MARTIN, and WARD.
- Numbers 1, 2, and 3 are placed to the right of the highlighted rows, corresponding to the first, second, and third rows of the highlighted data.

- LAG()
- LEAD()
- FIRST_VALUE()
- LAST_VALUE()
- ROW_NUMBER()
- SUM()
- SUM()
- SUM()



Choose which rows

DEPTNO	ENAME	SAL	
10	MILLER	1300	8750
10	CLARK	2450	2450
10	KING	5000	7450
20	SMITH	800	10875
20	ADAMS	1100	1100
20	JONES	2975	7075
20	FORD	3000	4100
20	SCOTT	3000	10075
30	JAMES	950	5400
30	WARD	1250	8400
30	MARTIN	1250	6650
30	TURNER	1500	7150
30	ALLEN	1600	1600
30	BLAKE	2850	4450

- SUM() OVER ()
- SUM() OVER (PARTITION BY DEPTNO)
- SUM() OVER (PARTITION BY DEPTNO ORDER BY ENAME ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW)



Syntax diagram

analytic_function ::=



analytic_clause ::=



Case: Top selling items

- Classic task for a programmer:
- Show top three by product group
- Also how big percentage they sold of the total





Case: Top selling items

- What kind of top three do you wish?
 - DENSE_RANK() - ties same rank
 - RANK() - ties same rank olympic style
 - ROW_NUMBER() - unique numbers, ties “random”
- PARTITION BY groups of items
- RATIO_TO_REPORT for percentages



Analytic Functions

AVG
CLUSTER_DETAILS
CLUSTER_DISTANCE
CLUSTER_ID
CLUSTER_PROBABILITY
CLUSTER_SET
CORR
COUNT
COVAR_POP
COVAR_SAMP
CUME_DIST
DENSE_RANK
FEATURE_DETAILS
FEATURE_ID
FEATURE_SET
FEATURE_VALUE
FIRST
FIRST_VALUE
LAG
LAST
LAST_VALUE
LEAD
LISTAGG
MAX
MEDIAN
MIN
NTH_VALUE
NTILE
PERCENT_RANK
PERCENTILE_CONT
PERCENTILE_DISC
PREDICTION
PREDICTION_COST
PREDICTION_DETAILS
PREDICTION_PROBABILITY
PREDICTION_SET
RANK
RATIO_TO_REPORT
REGR_ (Linear Regression) Functions
ROW_NUMBER
STDDEV
STDDEV_POP
STDDEV_SAMP
SUM
VAR_POP
VAR_SAMP
VARIANCE

(As of version 12.1.0.2.0)



The never ending case list

- We use analytic functions all the time
 - FIFO batch picking with warehouse routing
 - Time series forecasting of item sales
 - Forecasting when fireworks stock reaches zero
 - WheelGuide®
 - Replenish shop stock
 - Call Center statistics
 - Spare parts guide
 - Customer count / work schedule / number of orders
 - Booking calendar for mechanics
 - Shop space management
 - Discover idle hands
 - Detect seasonal variations for sales
 - Efficiency of Royal Danish Mail
- We can't imagine living without analytics 😊



You will find your own cases

- Just start using analytics
- The more you do the more often you find cases
- When you start to think you need to process your data procedurally – think again!
- Use the power of SQL to let the database do the hard work processing data
- That's what the database does best
- And you're paying for it so why not use it 😊

Questions?



@kibeha

<http://www.kibeha.dk>

- Download from DOAG:
 - This presentation
 - The scripts used
 - A detailed paper
- Advanced cases:
<http://www.slideshare.net/KimBergHansen/presentations>

Thank you