Windows, Cubes & Rollups
OLAP Extensions in SQL

Dipl.-Inform. Volker Stöffler
Volker.Stoeffler@DB-TecKnowledge.info

DB-TecKnowledge
Dipl.-Inform. Volker Stöffler
Databases - How
Agenda

• From grouped Aggregates to Cubes & Rollups
  ○ Use Cases
  ○ Hierarchies, Dimensions

• Windows
  ○ Use Cases
  ○ Running Totals, Deltas

• Cumulative & non-cumulative metrics
  ○ Count, Total, Average, Median

• With or without Cubes, Rollups & Windows
  ○ Benefits compared to replacements
Basics of Cubes & Rollups

- Based upon Grouped Aggregates
  - Standard SQL
  - Evaluations on Groups of Data Rows sharing Attribute Values
  - GROUP BY a, b, c

- Cubes are used to evaluate independent dimensions
  - GROUP BY CUBE (a, b, c)

- Rollups are used to evaluate levels in a hierarchy
  - GROUP BY ROLLUP (a, b, c)

- GROUPING SETS
  - Arbitrary Subset of CUBE
Cubes

- Cube of $n$ dimensions represents $2^n$ Grouped Aggregations
  - All Dimensions
    - a,b,c
  - All Pairs of Dimensions
    - a,b / a,c / b,c
  - All Single Dimensions
    - a / b / c
  - No Dimension (ungrouped)

  e.g. Product, Geography, Month
Rollups

- Rollup of $n$ Hierarchy Levels represents $n+1$ Grouped Aggregations
  - All Levels
    - a, b, c
  - All Prefixes
    - a / a, b
  - Top Level (ungrouped)

e.g. Region, District, Sales Rep
Non-OLAP Equivalent: Cubes

SELECT a, b, ... GROUP BY CUBE (a, b)

SELECT a, b, ... GROUP BY a, b
UNION ALL
SELECT a, NULL, ... GROUP BY a
UNION ALL
SELECT NULL, b, ... GROUP BY b
UNION ALL
SELECT NULL, NULL, ...
Non-OLAP Equivalent: Rollups

SELECT a, b, c, ... GROUP BY ROLLUP (a, b, c)

SELECT a, b, c, ... GROUP BY a, b, c
UNION ALL
SELECT a, b, NULL, ... GROUP BY a, b
UNION ALL
SELECT a, NULL, NULL, ... GROUP BY a
UNION ALL
SELECT NULL, NULL, NULL, ...
Windows Evaluations

• Evaluations of Rows together with Aggregates on logically adjacent “similar” Rows
• Windows are defined by
  ○ Partitioning Criteria (PARTITION BY)
    ✷ similar to GROUP BY, multiple expressions possible
  ○ Sort Order (ORDER BY)
    ✷ multiple expressions possible
  ○ Range ({ROWS | RANGE} BETWEEN ... AND ...)
    ✷ specified by Value or Number of Rows following / trailing current Row
Windows Use Case: Running Total

select rid, bDate, Subject, Amount, 
  sum (Amount) over RunningTotal Balance
from Bookings
window RunningTotal
  as (order by rid 
      rows between unbounded preceding and current row)
order by rid
Windows Use Case: Delta

select Account, rid, bDate, Item, Balance,
  Balance - last_value (Balance) over Delta Change
from Balances
window Delta
  as (partition by Account order by rid
      rows between 1 preceding and 1 preceding)
order by Account, rid
Benefit: Cubes & Rollups

• For Report Programmers
  ◦ Easier Coding

• During Execution
  ◦ Performance Gain: One Scan for Multiple Aggregations
  ◦ Optimisation is reasonably simple for Cumulative Metrics, but at least difficult / expansive for Non- Cumulative Metrics
[Non-] Cumulative Metrics

- Many Standard Evaluations are Cumulative
  - Count, Total (Sum), Minimum, Maximum
  - A Total Value can be trivially computed from Subtotals
- Average is not Cumulative
  - A Total Average can not be computed from Sub- Averages
  - But the Total Average can be computed from Total Sum and Total Count, which are Cumulative
- Median is not Cumulative
  - A Total Median can not be computed from Sub- Medians
  - The Working Data used to compute the Median can be used for multiple Grouped Aggregations
Benefit: Windows

• It is at least tricky to code Windows Functionality in SQL without OLAP Extensions
• It usually is a Self Join between a Table and a Grouped Aggregate Evaluation on the same Table
  ○ Alternatively a Query against a Table with a Correlated Subquery against the same table
• For Data Sets of non-trivial Size, Performance usually is a Nightmare.
Demo, Questions, Discussion

• Demo on SAP SQL Anywhere
• Queries expected to work on SAP HANA, SAP IQ identically
  ○ DDL, Population Logic need adjustment
Thank you!

Contact information:

Volker Stöffler
DB-Tec\textit{Knowledge}
Independant Consultant
Germany – 70771 Leinfelden-Echterdingen
mailto: Volker.Stoeffler@DB-TecKnowledge.info
http://scn.sap.com/people/volker.stoeffler