

Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals

Author: Jim Gray, Surajit Chaudhuri, Adam Bosworth, Andrew Layman, Don Reichart, Murali Venkatrao, Frank Pellow, Hamid Pirahesh

Reviewer: Suman Pandey

Summary:

This paper introduces the concept of CUBE and ROLLUP operators in the relational database. It highlights the situations when the GROUP BY fails and the need of the newer concepts. Group By does not allow a direct construction of histograms, Roll-up, drill down and Cross Tabulation. SQL standards do not support these techniques, so the SQL queries look very complicated if we try to do such things.

The cube is defined as a relational operator, with GROUP BY and ROLL UP as degenerate forms of the operator. CUBE is specified by overloading the SQL GROUP BY³. The paper also talks about the Algebra of the CUBE and the proposed Syntax and how it can be computed and maintained.

Comments:

The paper makes a great contribution to the development of CUBE and ROLLUP for data warehousing.

Introducing the dummy value “ALL” for explaining the super aggregation of items is a good way of showing the actual problem.

The main purpose of this paper is to introduce people CUBE and ROLLUP operators, and propose an enhancement to SQL for supporting it.

Shortcomings and Suggestions:

In Database terms the CUBE is defined in combination with the concept of OLAP, by [E.F. Codd & Associates](#) who published a white paper in [1994](#). If the concept of CUBE could have been defined associated with the OLAP as well, then the paper would have looked better for readers who want to see it from OLAP point of view.

“Drill down” could have been a better term to use in the place of “Roll up”, and also if the examples could have been explained in terms of Drill down and not on Roll up then that makes much sense.

For data representation the author has only considered the Microsoft products like Microsoft Excel like pivot tables and spreadsheet , there are many other tools in the market which can present multidimensional data better visually, so the whole explanation looks very much Microsoft oriented.

In this paper the Author tries to give a solution on the top of the Relations Database introduced by Codd, but now a day there are many non relational database approaches, those are not addressed.

In this paper the Tables used are mostly normalized, but what if we use non normalized table, for improving efficiency, then the ROLL UP and CUBE will generate unnecessary data. Authors do not talk much about these situations.