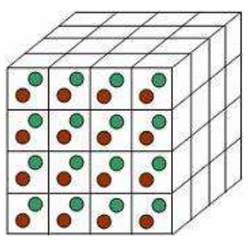
Describing Analytical Sessions Using a Multidimensional Algebra

Jovan Varga
Universitat Politècnica de Catalunya
jovanvarga@yahoo.com

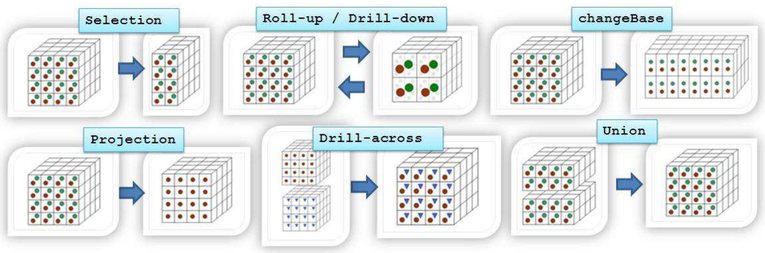
Our proposal is to represent SQL analytical queries in terms of a multidimensional algebra, which better characterizes the analytical efforts of the user

```
SELECT I1.ID, ..., In.ID, [ F ( ] c.Measure1 [ ) ], ...
FROM Cell c, Level1 I1, ..., Leveln In
WHERE c.key1=I1.ID AND ... AND c.keyn=In.ID [ AND Ii.attr Op. K ]
[ GROUP BY I1.ID, ..., In.ID ]
[ ORDER BY I1.ID, ..., In.ID ]
```

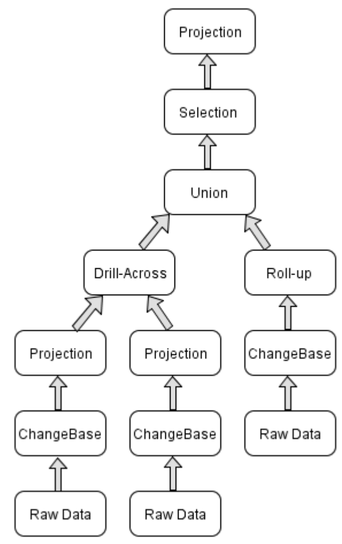
Analytical query



Data cube



Multidimensional algebra

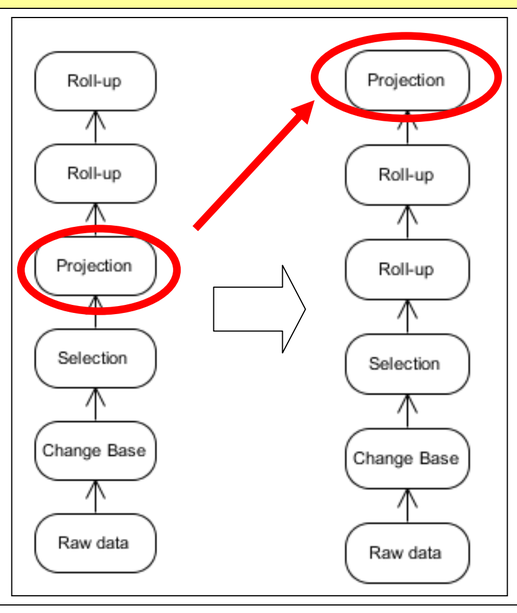


MAC

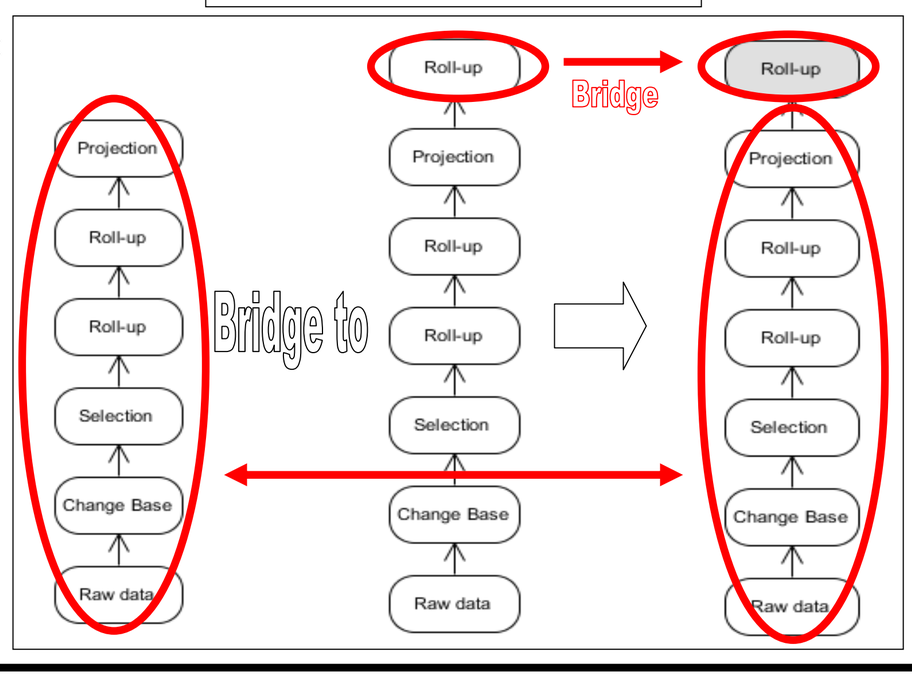
- **In this approach, it is proposed to characterize each issued SQL query (i.e. each query in the query log) by means of the set of MDA operators.**
- **Each operation of this algebraic characterization is associated with corresponding multidimensional schema.**
- **This characterization is giving multidimensional sense to the query and that is what we call Multidimensional Algebraic Characterization (MAC).**
- **MAC forms a tree (due to binary operators such as union or drill-across).**
- **Leafs of this tree are tuples directly retrieved from the database and we refer to them as raw data.**
- **Other nodes in this tree represent multidimensional operations (manipulations) over raw data.**

After identification of MAC remaining phases are:

1) Normalization of MAC:



2) NMAC bridging:



Remaining phases:

1) Normalization of MAC:

- **Once the query has been characterized according to multidimensional algebra the next step normalizes the MAC.**
- **Objective of normalization is to facilitate future manipulation and comparison. To do so, it is compulsory to store each MAC in a normalized form.**
- **We use a set of equivalence rules to pull the multidimensional operators up the algebraic structure. The final product is Normalized MAC (NMAC).**

2) NMAC bridging:

- **Working with algebraic expressions under normal form makes it easier to detect if, syntactically, two expressions are similar to each other.**
- **Similar NMACs may be considered logically related from an analytical point of view, and if two NMACs are close enough to each other, they are considered to belong to the same analytical session.**
- **In that case, they are coalesced into a session and both NMACs are logically related by annotating their bridging operators.**
- **This similarity can be potentially used in future to generate query recommendations for the users.**