

Informal Assignment

Given the DataCube defined in the class notes (with 3 dimensions of Route, Source, and Time; and measures of (Packages, Last)).

1. Describe the cube that is output by the following MDX query?

Select

{ [Source.WesternHemisphere.NorthAmerica], [Source.EasternHemisphere.Europe] } on rows,

{ [Route.NonGround.Air], [Route.Ground.*] } on Columns

{ [Time.*] } on Pages

From DataCube

Where [Measures.Last]

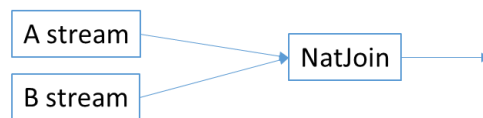
Answer: the output will be a 3D cube of dimensions:

Rows: NorthAmerica, Europe
Columns: Air, Road, Rail
Pages: 1stHalf, 2ndHalf
Measures: Last (aggregations)

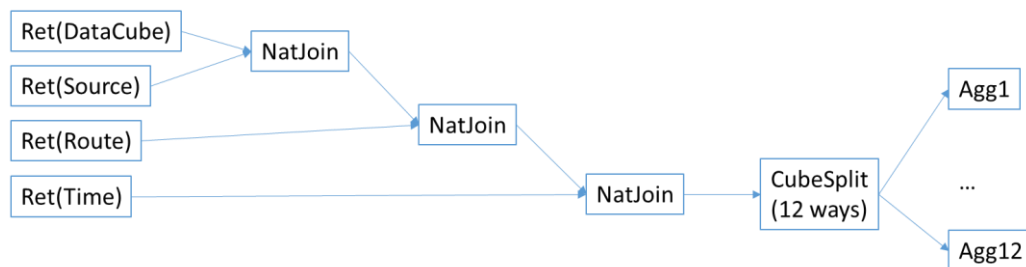
Or 2x3x2 aggregations = 12 aggregations

2. What is the plumbing diagram of the above query?

The DataCube is star schema with spokes Source, Route, and Time dimension tables. The order in which dimension tables are joined to the DataCube are likely to be in decreasing selectivity order. Since we don't know the exact selectivities, let's assume the join order is Source, then Route, then Time. Assume the notation: where if Natural Join is a hash join, the "A" stream would be inhaled in memory.



The plumbing diagram of the query would be:



3. What are the local predicates of each retrieval in (1)

Answer:

DataCube all

Source Continent in {NorthAmerica, Europe}

Route: RouteType in {Air, Road, Rail}

Time: all

4. How would you recommend that the DataCube be stored so that such queries could be parallelized (and be more efficiently processed than non-parallelized processing)

Answer: the dimension tables are small. Their contents can be broadcast for joins. It is the DataCube that needs to be horizontally partitioned. Grep for "Broadcast" join in the DBMachine notes. Basically it would look like this:

