Adam Gąsiorowski
Krzysztof Jeznach

Data Warehouse Optimization – report

1. Aim of the laboratory:

The aim of the task is to show issues concerning various physical cube models and aggregation design.

2. Preliminary assumptions:

Size of the database: 1040 MB

Testing environment: SSMS, SQL Server Management Studio

3. Testing:

Testing query execution times for different models, with and without defined aggregations. Testing cube processing times in the same testing settings.

Brief description of the queries:

1. (one with aggregations on dates) -

Show what percentage of total recruited students are those from Erasmus and whether they have a scholarship. Do the same for non-Erasmus students in the 2011 winter.

```
WITH MEMBER [Measures].[Recruitment Percentage] AS
  IIF(
    [Measures].[F Recruitment Count] = 0,
    NULL,
    ROUND(
      DIVIDE(
        [Measures].[F Recruitment Count],
        SUM([Dim Student].[Is Erasmus].[Is Erasmus].ALLMEMBERS * [Dim Student].[Scholarship].[Scholarship].ALLMEMBERS
* [Dim Semester].[SemesterHierarchy].[Semester Year].&[2011].&[Winter], [Measures].[F Recruitment Count])
      ) * 100,
      2
    )
  )
SELECT
  NON EMPTY {[Measures].[Recruitment Percentage]} ON COLUMNS,
  NON EMPTY {(
    [Dim Student].[Is Erasmus].[Is Erasmus].ALLMEMBERS,
    [Dim Student].[Scholarship].[Scholarship].ALLMEMBERS,
    [Dim Semester].[SemesterHierarchy].[Semester Year].&[2011].&[Winter]
  )} DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS
FROM [University Star Schema1]
CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME,
FONT_SIZE, FONT_FLAGS
```

2. (one for a particular dimension attribute) -

Show the average congestion (capacity divided by area) on a given department and growth in the number of the students there. Growth calculated from 2011 winter to 2012 summer.

```
 SELECT NON EMPTY { [Measures].[Difference in students] } ON COLUMNS, NON EMPTY { ([Dim
Department].[Congestion].[Congestion].ALLMEMBERS ) } DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS FROM [University Star Schema1] CELL PROPERTIES VALUE, BACK_COLOR,
FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

3. (general one) -

Show what percentage of all students, in the year with the highest number of students, were erasmus and foreign ones.

```
WITH MEMBER [Measures].[Recruitment Percentage] AS
  ROUND(
    DIVIDE(
      [Measures].[F Recruitment Count],
      SUM([Dim Student].[Is Erasmus].[Is Erasmus].ALLMEMBERS * [Dim Misc].[Most Crowded Semester].[Most Crowded
Semester].&[1], [Measures].[F Recruitment Count])
    ) * 100,
    2
  )
SELECT
  NON EMPTY {[Measures].[Recruitment Percentage]} ON COLUMNS,
  NON EMPTY {(
    [Dim Student].[Is Erasmus].[Is Erasmus].ALLMEMBERS,
    [Dim Misc].[Most Crowded Semester].[Most Crowded Semester].&[1]
  )} DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS
FROM [University Star Schema1]
CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME,
FONT_SIZE, FONT_FLAGS
```

	MOLAP		ROLAP		HOLAP	
	No Aggr.	Aggr.	No Aggr.	Aggr.	No Aggr.	Aggr.
Querying speed (for 3 different queries)	51,4 ms	42 ms	406,4 ms	383,8 ms	370,7 ms	409,2 ms
	33,8 ms	6,6 ms	266,8 ms	237,6 ms	177,4 ms	9,8 ms
	28,8 ms	27,2 ms	422,8 ms	421,6 ms	155,6 ms	282,4 ms
Processing time	9122 ms	7002 ms	5602 ms	5309 ms	5109 ms	5213 ms
Total size	43,40 MB	43,54 MB	36,83 MB	36,86 MB	36,86 MB	36,89 MB

4. Discussion (comparison of the theory with the obtained results):

Our tests for MOLAP, ROLAP and HOLAP without aggregation went according to the theory - MOLAP is the shortest, ROLAP the longest and HOLAP is in the middle.

After aggregation, ROLAP behaved as expected - the results were the same as without aggregation. For MOLAP and HOLAP, only the second query was greatly impacted by aggregation which is deviating from the theory. The reason for that may be that we used wrong aggregations for the first and third query and accurate for the second.

As for processing time, everything is pretty much the same as the theory would suggest except for the small difference between ROLAP and HOLAP. Additionally, the difference between HOLAP with and without aggregation seems a bit small.

As for total size, MOLAP is partially correct. It is the biggest in comparison to the other two but the size for aggregation and no aggregation did not change. After excessive thinking, we can not figure out why that is. ROLAP is slightly smaller than HOLAP which is true according to the theory, however we do not know if this is a significant difference.

We think that the key reason for the differences and deviations from the theory might be a result of poorly working aggregations. We tried our best to make them as efficient as possible and those results were our best tries.