

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 (data warehouse): 1 000 000

Testing environment: SQL Server Profiler

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)

```
SELECT
{ [Measures].[Claim Processing Count] } ON COLUMNS,
{ ( [Ref Submission Date].[Year].[2023] ),
  ( [Ref Submission Date].[Year].[2022] ) } ON ROWS
FROM [Insurance DW]
```

2. (one for a particular dimension attribute)

```
SELECT { [Measures].[Survey AVG L] } ON COLUMNS,  
Filter([Insurance Agent].[Name And Surname].[Name And Surname].ALLMEMBERS,  
[Measures].[Survey AVG L] > [Measures].[AvgSatisfaction] ) ON ROWS FROM  
[Insurance DW]
```

3. (general one)

```
WITH MEMBER [Measures].[Claim Count] AS  
If( IsEmpty( [Measures].[Claim Processing Count] ), 0, [Measures].[Claim Processing  
Count])  
SELECT  
{ [Measures].[Claim Count] } ON COLUMNS,  
{ ([Ref Submission Date].[Month].Children, BOTTOMCOUNT ( EXCEPT ({[Insurance  
Agent].[Name And Surname].Children}, {[Insurance Agent].[Name And  
Surname].[All].UNKNOWNMEMBER}), 1, [Measures].[Claim Processing Count])),  
([Ref Submission Date].[Month].Children, TOPCOUNT ( EXCEPT ({[Insurance  
Agent].[Name And Surname].Children}, {[Insurance Agent].[Name And  
Surname].[All].UNKNOWNMEMBER}), 1, [Measures].[Claim Processing Count]))  
} ON ROWS  
FROM [Insurance DW]
```

	MOLAP		ROLAP		HOLAP	
	Aggr.	No aggr.	Aggr.	No aggr.	Aggr.	No aggr.
Querying speed (for 3 different queries, in ms)	24	71	770	730	38	754
	24	23	25	16	25	34
	89	64	171	174	191	172
Processing time	18554	15966	778	5372	4449	4124
Total size	5,41 MB	5,42 MB	25,32 MB	24,10 MB	17,10 MB	16,32 MB

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

For the 1st and 3rd query ROLAP and HOLAP have slower querying speed than MOLAP. It might be because MOLAP is storing pre-aggregated data in a multidimensional format and thanks to that the results are coming faster.

ROLAP and HOLAP are also bigger than MOLAP in size - so for MOLAP there is no need to store raw data, only after some aggregations.

ROLAP on the other hand stores data in a relational database which involves more complex SQL queries.

With aggregations MOLAP and HOLAP have similar results in 1st and 2nd queries. It may come from the fact that HOLAP combines features of MOLAP and ROLAP.