Dędza Kamil, Agnieszka Kuleta

name, surname, studID

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): 463.23 MB

Tested cube has 6 dimensions and 2 fact tables.

Testing environment: Visual Studio 2022, Microsoft Server 2019

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)

Average number of additional lessons selected to each facility during the current year.

2. (one for a particular dimension attribute) = particular dimension attribute => age

We are looking for a number of applications to each age group in May in current year compare to previous one .

3. (general one)
   We are looking for number of applications in each month during current year.

	MOLAP		ROLAP		HOLAP	
	Aggr,	No aggr.	Aggr.	No aggr.	Aggr.	No aggr.
Querying speed (for 3 different queries)	108	98	425	383	355	323
	48	37	547	324	317	289
	38	40	157	224	115	158
Processing time	4125	5479	956	1076	1589	3304
Total size	436.76 MB	**463.23MB**	382.14MB	383.14MB	387.58MB	386.25MB

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

- Obtained results agree with the theory connected the design.
    - MOLAP – really quick query executing, but uses the most place and has the highest time processing the cube. We can clearly see the main disadvantage of this model- storing of duplicate data in the analytical database. As a result, the analytical database and data warehouse store the same large amounts of data, and changes to the source data are not visible when viewing the cube.

    - ROLAP- the shortest time processing the cube for sure, also the less space used, but the longest time of executing the queries. We can clearly see the main disadvantage of the ROLAP model - the low performance of viewing and processing cubes: if the data displayed is not in the SSAS server buffer or the client application.

    - HOLAP- slower cube processing, but relatively quick queries' execution. The small storage is needed.
      Viewing detailed data in the analytical cube is slower because it requires reading from the warehouse. The advantage of the HOLAP model is the lack of storing copies of the same data in two databases (warehouse and analytical base).

- Aggregations, which we have created, weren't designed correctly, so the execution time did not improve. Our aggregations were created on the wrong attributes, so in fact it

decreased the performance. To increase performance we should select attributes that we aggregate on, not all possible.