Martyna Aszyk, 193524

Martyna Majewska, 193360

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):

Table Name	# Records	Reserved (KB)	Data (KB)	Indexes (KB)	Unused (KB)
dbo.BUS	220	72	24	16	32
dbo.BUS_OFFICE	23	72	8	8	56
dbo.BUS_TRAVEL	1,023,369	59,400	58,568	520	312
dbo.DATE	519	144	32	32	80
dbo.JUNK	16	144	8	24	112
dbo.ROUTE	72	144	8	24	112
dbo.TIME	1,440	144	48	64	32

Testing environment:

- CPU: AMD Ryzen 5 5600H 3.30 GHz
- Memory (RAM): 16.0 GB (15.4 GB usable)
- Disc: 475GB (173GB used)
- System: Win 11

Software:

- Visual Studio 2019
- SQL Server Management Studio 19
- SQL Server Profiler 19

3. Testing

Testing query execution times for different models, with and without defined aggregations. Processing times are obtained as the mean of 10 processing executions for each physical cube structure. There were no outliers.

Aggregations: *Route Number* instead of *Route Name* in ROUTE dimension table.

Code for clearing cash before each execution:

```
<ClearCache xmlns="http://schemas.microsoft.com/analysisservices/2003/engine">
<Object>
<DatabaseID>BusTravel_DW</DatabaseID>
</Object>
</ClearCache>
```

Brief description of the queries:

1. Compare the Total Number of Passengers per Route on Weekdays and Weekends in Current and Previous Month

This query calculates the total number of passengers for each route, distinguishing between weekdays and weekends, and compares these totals between the current month and the previous month. It uses the PARALLELPERIOD function to reference the previous month and retrieves the total number of passengers for both periods. The results are displayed side by side, allowing for a detailed comparison.

3.1. Query Execution Times

Experiment no.	MOLAP		ROLAP		HOLAP	
	AGGR	NO AGGR.	AGGR	NO AGGR.	AGGR	NO AGGR.
1	6	3	2	6	2	6
2	5	5	2	7	2	4
3	6	9	2	7	2	6

4	9	5	2	6	3	8
5	9	8	2	8	2	8
6	6	9	3	9	2	9
7	4	5	2	5	3	5
8	7	9	2	6	2	3
9	8	4	2	4	2	5
10	4	1	2	4	2	7
Mean	6.4	5.8	2.1	6.2	2.2	6.1
St. Dev.	1.74	2.68	0.3	1.54	0.4	1.81

2. Identify routes with the lowest bus occupancy (<15% of total bus capacity).

This query identifies the bus routes with the lowest occupancy rates, specifically those with an occupancy of less than 15% of the total bus capacity – those from occupation category "very low".

3.2. Query Execution Times

Experiment no.	MOLAP		ROLAP		HOLAP	
	AGGR	NO AGGR.	AGGR	NO AGGR.	AGGR	NO AGGR.
1	5	8	2	4	1	7
2	3	1	2	3	1	6
3	3	1	3	2	1	4
4	5	5	2	5	2	8
5	4	9	2	0	1	2
6	1	3	1	4	1	1
7	5	7	2	6	1	3
8	3	4	1	2	1	3
9	6	3	3	5	2	4
10	1	5	2	4	1	3
Mean	3.6	4.6	2	3.5	1.2	4.1
St. Dev.	1.62	2.62	0.63	1.69	0.4	2.12

3. Do certain routes are associated with more frequent passenger feedback?

This query identifies the top 10 routes with the highest number of satisfaction surveys completed by passengers. It orders the routes by the number of surveys in descending order and selects the top 10 routes to show which ones receive the most feedback.

3.3. Query Execution Times

Experiment no.	MOLAP		ROLAP		HOLAP	
	AGGR	NO AGGR.	AGGR	NO AGGR.	AGGR	NO AGGR.
1	4	1	1	4	2	7
2	2	2	1	4	1	2
3	3	1	1	6	1	4
4	6	4	1	2	2	8
5	4	4	1	2	1	8
6	3	1	1	3	1	5
7	2	2	1	1	3	1
8	8	4	1	4	1	1
9	3	1	3	5	1	1
10	4	2	1	4	1	1
Mean	3.9	1.9	1.2	3.5	1.4	3.6
St. Dev.	1.76	1.58	0.6	1.43	0.66	3.07

3.4. Results

	MOLAP		ROLAP		HOLAP	
	AGGR	NO AGGR.	AGGR	NO AGGR.	AGGR	NO AGGR.
Querying speed (for 3 different queries)	1) 6.4ms 2) 3.6ms 3) 3.9ms	1) 5.8ms 2) 4.6ms 3) 1.9ms	1) 2.1ms 2) 2ms 3) 1.2ms	1) 6.2ms 2) 3.5ms 3) 3.5ms	1) 2.2ms 2) 1.2ms 3) 1.4ms	1) 6.1ms 2) 4.1ms 3) 3.6ms
Processing Time	32234ms	8251ms	1127ms	2047ms	6137ms	2593ms
Total Size	119.80MB	12.21 MB	1.04 MB	1.04 MB	108.63MB	1.04 MB

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

- Theory – BLUE
- Obtained Results – GREEN

	MOLAP		**ROLAP**		**HOLAP**	
	Theory	Results	Theory	Results	Theory	Results
Querying Time	Short	Moderate (designed aggregations do not have a big impact)	Long	Moderate (with aggregations time is shorter)	Moderate (short with well designed aggregations)	Moderate – time is shorter when aggregations are applied
Processing Time	Long	Long	Short	Short	Moderate (if no aggregations are designed it will be short)	Moderate – when aggregations are applied processing time is short)
Total Size	Big (size of the measure group is much smaller if no aggregations are designed for them)	Big – size is much smaller without applied aggregations	Small	Small	Moderate	Small – without aggregations Big – with aggregations

MOLAP:

- Querying time was moderate, designed aggregations don't have a big impact.
- Processing time was long so agreed with theory
- Size is the largest of all structures, so agreed with theory and without applied aggregations the size is much smaller

ROLAP:

- Querying Time was moderate when in theory it should be long
- Processing Time and Total Size agreed with theory

HOLAP:

- Querying Time was moderate and agreed with the theory (time is shorter when aggregations are applied)
- Processing Time was moderate and agreed with the theory (if no aggregations are designed it will be short)
- Total Size was smaller without aggregations and bigger with aggregations, in theory it should be moderate