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 database (Data Warehouse):

Description:

Estimated size in	1 615 793
Date created:	18.05.2023 15:22:04
Date last updated:	19.05.2023 08:54:42
Date last processed:	19.05.2023 08:55:03
Visible:	True
Language:	English (United States)
Collation:	Polish_CI_AS

Testing environment: SQL Server Profiler 18

3. Testing query execution times for different models, with and without defined aggregations.

1. Compare average grade of courses with the biggest lowest number of hours/

	molap			holap			rolap	
	Aggr,	No aggr.		Aggr,	No aggr.		Aggr,	No aggr.
	5	25		8	22		27	28
	6	16		3	16		26	28
	7	21		3	20		15	15
	3	23		4	17		16	17
	5	18		5	15		17	19
	4	16		5	14		19	16
	5	49		4	18		18	18
	4	16		6	15		15	16
	6	26		5	16		16	17
	8	18		4	19		17	15
mean	5.3	22.8	mean	4.7	17.2	mean	18.6	18.9
median	5	19.5	median	4.5	16.5	median	17	17
deviation	1.494434	9.94205434	deviation	1.49443412	2.529822	deviation	4.351245	4.954235

2. Show the grades from finished courses in shape to the months.

	molap			holap			rolap	
	Aggr,	No aggr.		Aggr,	No aggr.		Aggr,	No aggr.
	5	49		5	85		45	46
	4	21		6	91		39	49
	4	25		4	87		42	47
	4	28		5	89		48	40
	5	25		3	79		41	36
	6	24		6	88		42	45
	4	21		6	83		44	42
	5	19		8	85		42	35
	3	28		7	78		45	34
	5	17		4	82		42	47
mean	4.5	25.7	mean	5.4	84.7	mean	43	42.1
median	4.5	24.5	median	5.5	85	median	42	43.5
deviation	0.849837	8.957306	deviation	1.505545	4.24395	deviation	2.538591	5.546771

3. Which faculty has the highest average grade?

	molap			holap			rolap	
	Aggr,	No aggr.		Aggr,	No aggr.		Aggr,	No aggr.
	4	18		102	469		707	585
	5	19		43	523		504	485
	5	17		43	566		532	544
	4	16		47	543		537	497
	3	17		46	477		514	430
	6	19		37	483		555	425
	7	21		43	525		722	554
	4	15		41	436		634	523
	3	18		42	511		565	543
	5	22		47	498		543	631
mean	4.6	18.2	mean	49.1	503.1	mean	581.3	521.7
median	4.5	18	median	43	504.5	median	549	533
deviation	1.264911	2.149935	deviation	18.82935	38.38822	deviation	78.66674	64.69252

		MOLAP		ROLAP		HOLAP	
		Aggr,	No aggr.	Aggr.	No aggr.	Aggr.	No aggr.
Querying speed (for 3 different queries) in ms	1	5.3	22.8	18.6	18.9	4.7	17.2
	2	4.5	25,7	43	42.1	5.5	84.7
	3	4.6	18.2	581.3	521.7	49.1	503.1
Processing time in ms		11131	4179	2889	2745	2835	2785
Total size		14,2MB	13,99MB	3,39MB	3,41MB	3,47MB	3,42MB

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

MOLAP has much bigger total size of database because it is duplicating the data and aggregations to analytical database. It has also advantage because it decreases MDX query time because it does not have to take data from external source.

HOLAP size is a smaller than MOLAP but bigger than ROLAP. It is because it does not contain duplication of data like MOLAP, so it reaches for information from external sources. There can be seen similarity to ROLAP no aggregation queries.

ROLAP therefore does not copy anything neither aggregation nor data and it is supplied with data from external sources. That is why we can see that the size of database is smaller than the other ones. What is more we can see no difference between aggregation and no aggregation queries, it is because these aggregations are from external sources.