Zwischenpräsentation in Programmiermodel Seminar

Finding Hot Aggregates in Application Workload

Idee

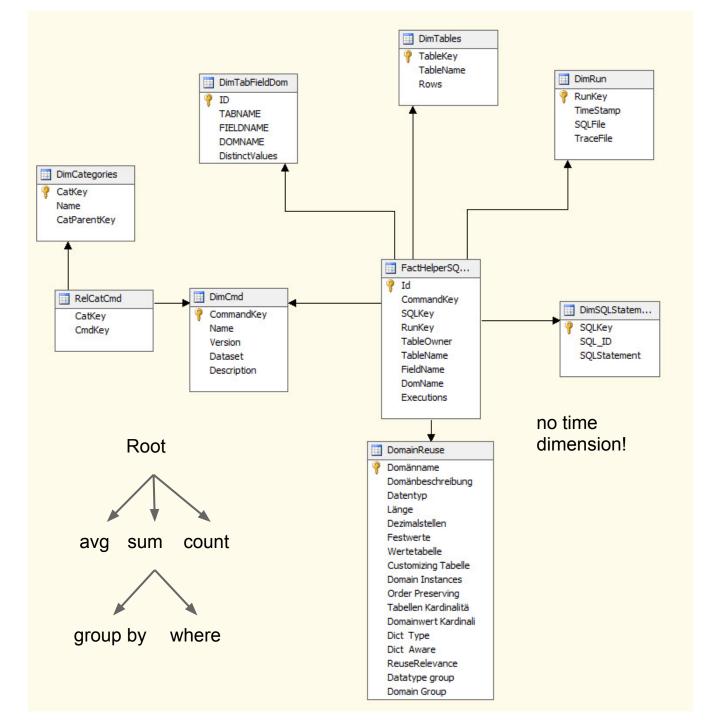
based on "statistical db performance log" (workload) analysis, aggregate functions shall be identified, evaluated and materialized according to their relevance and the change frequency of the underlying data. => IO and CPU usage shall be reduced for repetitive calculation of unchanging aggregates.

- Evaluation of historical relevance and change frequency
- Determine a optimal level of granularity for materialization
- Derive a data schema for materializing aggregate result
- Find important data regions (hot spots) in workload
- Query matching strategies for aggregate functions

Structure

- Introduction
- Background
 - Related Work
 - Basic Therms ()
- Workload Analysis
 - Workload data
 - Command (extractions)
 - Star Schema / Dimensions
 - Data Mining Analysis
- Hot Aggregates
 - Metrics
 - Characteristics
 - Results
- Future Work

Schema



Facts

$\underline{https://docs.google.com/spreadsheet/ccc?key=0AptTKF7XREOBdFlkZlZtSzd5VmNfYm43eU5DRHhXMVE\#gid=0}$

CmdKey	SQLKey	Text	Table	Field	Exec	Rows effected	1
sum	1		rev	net	300	7000000	54
sum	54		rev	net	10	5000	151
sum	151		rev	net	874		
sum			rev	net	1184		
grpby_sum	1		rev	prodID	300		
grpby_sum	1		rev	arealD	300		
grpby_sum	54		rev	custID	10		
grpby_sum	151		rev	custID	874		
grpby_sum	151		rev	prodID	874		
grpby_sum	151		rev	arealD	874		
whr_sum	54	"= schoki"	rev	prodID	10		
whr_sum	151	"= 2012"	rev	year	874		
grpby_sum			rev	prodID	1174		
grpby_sum			rev	arealD	1174		
grpby_sum			rev	custID	884		
whr_sum		"= schoki"	rev	prodID	10		
whr_sum		"= 2012"	rev	year	874		1
insert	3		rev		5	5	4
update	4		rev	//	1	50	
delete	5		rev				2
							2 3 5
			12	9			5
				100			
				100			5.1
							5.2

Related Work - Papers

[1] A. Bog and J. Kruger, "A Composite Benchmark for Online Transaction Processing and Operational Reporting,"[2] J. Zhou, P.-A. Larson, J. Goldstein, and L. Ding, "Dynamic Materialized Views,"

[3] R. Cole, F. Funke, L. Giakoumakis, W. Guy, A. Kemper, S. Krompass,

H. Kuno, R. Nambiar, T. Neumann, M. Poess, and Others, "The mixed

workload CH-benCHmark,"

[4] A. Dan, P. S. Yu, and J.-Y. Chung, "Characterization of database

access pattern for analytic prediction of buffer hit probability,"

[5] P. M. Deshpande and J. F. Naughton, "Aggregate Aware Caching for

Multi-Dimensional Queries,"

[6] J. Goldstein and P.-a. k. Larson, "Optimizing queries using materialized views: a practical, scalable solution,"

[7] C.-S. Park, M. H. Kim, and Y.-J. Lee, "Usability-based caching of

query results in OLAP systems,"

[8] P. Roy, B. Laboratories, and M. Hill, "Don't Trash your Intermediate

Results, Cache'

[9] P. Roy, K. Ramamritham, and S. Seshadri, "ON WORKLOAD CHARACTERIZATION OF RELATIONAL DATABASE ENVIRONMENTS,"