

Contents

List of Figures	v
List of Tables	ix
1 Introduction	1
1.1 Research Context and Contributions	3
1.1.1 Scientific Contributions	4
1.1.2 Out of Scope	5
1.2 Thesis Outline	6
1.3 Publications	7
2 Background and Motivation	11
2.1 Enterprise Applications and Mixed Workloads	11
2.2 Data Aggregation	14
2.2.1 Aggregate Functions	15
2.2.2 Grouping in Data Aggregation	16
2.2.3 Formal Description of Data Aggregation	16
2.2.4 Implementation of Data Aggregation	17
2.2.5 Data Aggregation in Enterprise Applications	18
2.3 Materialized Views	19
2.3.1 Materialized View Selection Problem	19
2.3.2 Materialized View Maintenance Problem	20
2.3.3 Materialized View Exploitation Problem	22
2.4 Application-Maintained Aggregates	22
2.5 In-Memory Database Systems	25
2.5.1 Hardware Trends	25
2.5.2 Row- and Column-Oriented Data Layout	27
2.5.3 Main-Delta Architecture and Merge Process	29
2.6 Benefits of On-the-Fly Aggregation	31
2.6.1 Flexibility	31
2.6.2 Application Simplification	32

2.6.3	Data Footprint Reduction	34
2.7	A Preliminary Workload Analysis	35
2.8	Motivation Summary	38
3	The Aggregate Cache	39
3.1	System Overview	39
3.2	System Architecture	41
3.2.1	Aggregate Cache Manager	41
3.2.2	Aggregate Cache Entries	43
3.2.3	Consistent View Manager	45
3.3	Query Execution Using the Aggregate Cache	47
3.3.1	Supported Query Definitions	50
3.3.2	Aggregate Cache Entry Creation	51
3.3.3	Aggregate Cache Matching	54
3.3.4	Delta Compensation	55
3.3.5	Main Compensation	58
3.4	Aggregate Profit Metrics	63
3.4.1	Buffer Management Metrics	64
3.4.2	Query Result Cache Metrics	66
3.4.3	Aggregate Cache Metrics	67
3.5	Aggregate Selection and Eviction	70
3.5.1	Profit Metrics Management	71
3.5.2	Aggregate Cache Selection	72
3.5.3	Aggregate Cache Eviction	73
3.6	Aggregate Maintenance	74
3.6.1	Delta Aggregate Maintenance	76
3.6.2	Main Aggregate Maintenance	78
4	Join Pruning for Aggregates Based on Multiple Relations	81
4.1	Answering Aggregate Join Queries	81
4.1.1	Answering Aggregate Join Queries without the Aggregate Cache	83
4.1.2	Answering Aggregate Join Queries with the Aggregate Cache	84
4.2	Enterprise Application Characteristics	84
4.2.1	Enterprise Application Schema Design Patterns	85
4.2.2	Enterprise Application Workload Patterns	85
4.3	Partition Pruning and Matching Dependencies	87
4.3.1	Logical Partition Pruning	87
4.3.2	Dynamic Partition Pruning	89
4.3.3	Matching Dependencies	91
4.4	Object-Aware Joins in the Aggregate Cache	92

4.4.1	Dynamic Join Pruning	93
4.4.2	Delta Merge Operation	95
4.4.3	Join Predicate Pushdown	96
4.4.4	Join Pruning for Multiple Partitions	96
5	Experimental Evaluation	99
5.1	Benchmark Description	99
5.1.1	CH-BenCHmark	100
5.1.2	FICO Benchmark	101
5.2	Overhead of Aggregate Caching Algorithm	104
5.3	Aggregate Cache Performance Benefit	105
5.3.1	CH-benCHmark Queries	105
5.3.2	FICO Benchmark Queries	106
5.4	Aggregate Cache Maintenance	108
5.5	Dynamic Aggregate Admission and Eviction	111
5.5.1	Impact of Delta Compensation	111
5.5.2	Aggregate Cache Size	112
5.5.3	Aggregate Cache Eviction During Maintenance	113
5.6	Join Pruning in the Aggregate Cache	115
5.6.1	Memory Consumption Overhead	115
5.6.2	Insert Overhead	116
5.6.3	Join Pruning Benefit	117
5.6.4	Join Predicate Pushdown Benefit	120
5.6.5	Join Pruning Benefit for Multiple Partitions	121
6	Related Work	123
6.1	Query Caching and Materialized Views	123
6.2	Materialized View Maintenance	125
6.2.1	Materialized View Maintenance Timing	125
6.2.2	Incremental View Maintenance	126
6.2.3	Incremental View Maintenance Based on Multiple Relations	128
6.2.4	Maintenance Data Structures	129
6.3	Materialized View Selection and Cache Management	130
6.4	Query Optimization	132
6.4.1	Query Optimization of Joins on Partitioned Tables	132
6.4.2	Modeling of Application Semantics for Query Optimization	133
6.4.3	Indexing	134
7	Conclusions and Future Work	135
7.1	Conclusions	135
7.2	Future Work	137

Appendix	141
Definition of Symbols	141
Abbreviations and Glossary	143
FICO Benchmark Queries	147
Bibliography	151
Eigenständigkeitserklärung (Declaration of Authorship)	169