

# Multidimensional Data Management in Mobile Environments

Dissertation zur Erlangung des akademischen Grades eines Doktors der  
Naturwissenschaften (Dr. rer. nat.) im Fachbereich Mathematik und  
Informatik der Freien Universität Berlin

vorgelegt von

**Ilias Michalarias**

im Oktober 2007

Tag der Disputation: 12. Dezember 2007

Gutachter:

Prof. Dr. Hans-J. Lenz, FU Berlin  
Prof. Dr.-Ing. Jochen H. Schiller, FU Berlin  
Prof. Dr. Timos Sellis, TU Athen



## Abstract

Constantly growing data volumes and accelerating update speeds are fundamentally changing the role of data warehousing in modern business. Data warehousing is increasingly used not only for strategic, but for operative decision making as well. More data, coming in faster and requiring immediate conversion into decisions means that organizations are confronting the need for right-time (active) data warehousing.

In parallel, great advances, both in wireless networks and respective mobile devices functioning within their proximity, enable a wide-scale adoption of mobile information systems. Mobile devices become smaller, cheaper and more powerful, being able to run more sophisticated applications and network services. With ubiquitous data access already being success critical, integrating mobile devices in existing infrastructure is a fundamental requirement.

The research field of mobile online analytical processing (mOLAP) brings the two aforementioned research and application areas together. The term mOLAP encompasses all necessary technologies for mobile information systems that enable multidimensional data access to users carrying a mobile device.

This thesis presents FCLOS, a complete client-server architecture explicitly designed for mOLAP. FCLOS exploits derivability between multidimensional data cubes in conjunction with wireless broadcast in order to become a query efficient, self adaptive and scalable mOLAP information system.



## Acknowledgements

This work was financed by the Berlin-Brandenburg Graduate School on Distributed Information Systems and carried out at the Freie Universität Berlin. Pursuing a Ph.D. in computer engineering is not a trivial task. Fortunately, during this time I had the owner and privilege to work and cooperate with some great persons.

Above all, I would like to express my gratitude to my first advisor, Professor Hans-J. Lenz. He gave me the chance to pursue my ideas in absolute freedom. His genuine interest in science has been and will be a great inspiration for me, not only for this thesis, but for the rest of my life. I also thank my fellows in the institute of Production, Information Systems and Operations Research.

I would also like to sincerely thank my second advisor, Professor Jochen Schiller. Our discussions, especially at the beginning of this work, helped me keep my feet on the ground and stay in a sound scientific direction. I also thank him for letting me participate in his research group. My fellows of the Computer Systems and Telematics group, created an entertaining working environment.

A special thanks goes to Professor Timos Sellis. The fact that he accepted to serve as an external referee of this work, especially bearing in mind his busy schedule, deeply owners me.

Furthermore, I would like to thank all professors of the Berlin-Brandenburg Graduate School on Distributed Information Systems, first for giving me the opportunity to embark on this project, and secondly for their precious feedback during our workshops. I also thank my fellows of the graduate school for our constructive discussions.

I am particularly thankful to the master thesis students Arkadiy Omelchenko and Christian Becker for their invaluable contribution to some parts of this thesis.

I would also like to thank Sarah, Nikos, the BFB and HSEC members (you know who you are) for their individual assistance.

Last but not least, I would like to express my gratitude to my parents. I would have never accomplished anything important in my life, including this thesis, without their incessant support.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem Statement . . . . .	2
1.2	Contribution . . . . .	3
1.3	Thesis Overview . . . . .	5
<b>I</b>	<b>Background Technologies</b>	<b>7</b>
<b>2</b>	<b>Multidimensional Databases</b>	<b>9</b>
2.1	Multidimensional Space . . . . .	9
2.2	Conceptual Multidimensional Models . . . . .	11
2.3	Data Cube . . . . .	12
2.4	Aggregation Lattices . . . . .	14
2.5	Derivability - Subsumption . . . . .	15
2.6	Query Mapping to Aggregation Lattices . . . . .	17
2.7	Querying Multidimensional Data . . . . .	18
2.8	Summary . . . . .	19
<b>3</b>	<b>Data Management in Mobile Environments</b>	<b>21</b>
3.1	Limitations . . . . .	21
3.2	Wireless Information Broadcast . . . . .	22
3.2.1	Broadcast Structure . . . . .	23
3.2.2	Broadcast Architectures . . . . .	24
3.3	Mobile Databases . . . . .	26
3.3.1	Transaction Processing . . . . .	26
3.3.2	Query Processing . . . . .	27
3.3.3	Caching . . . . .	27
3.3.4	Replication . . . . .	27
3.4	Summary . . . . .	28
<b>II</b>	<b>Related Work</b>	<b>29</b>
<b>4</b>	<b>Broadcast Systems</b>	<b>31</b>

4.1	General Data Broadcast Systems . . . . .	31
4.1.1	On-demand . . . . .	31
4.1.2	Push-based . . . . .	34
4.1.3	Hybrid . . . . .	36
4.1.4	Evaluation of Data Broadcast Modes . . . . .	37
4.2	Database Broadcast Systems . . . . .	38
4.3	Mobile OLAP Suitability . . . . .	40
4.4	Summary . . . . .	41
<b>5</b>	<b>Distributed Multidimensional Data</b>	<b>43</b>
5.1	Distributed Data Warehousing . . . . .	43
5.1.1	Distributed Storage and OLAP . . . . .	43
5.1.2	Caching . . . . .	44
5.2	Mobile Data Warehousing . . . . .	45
5.3	Mobile OLAP Architectures . . . . .	46
5.3.1	Criteria . . . . .	46
5.3.2	STOBS . . . . .	48
5.3.3	SBS . . . . .	50
5.3.4	DV-ES . . . . .	51
5.3.5	Evaluation of mOLAP Architectures . . . . .	52
5.4	Summary . . . . .	52
<b>III</b>	<b>The FCLOS Architecture</b>	<b>55</b>
<b>6</b>	<b>Mobile OLAP in Wireless Infrastructure Based Networks</b>	<b>57</b>
6.1	Motivation . . . . .	57
6.2	Requirements . . . . .	58
6.3	System Architecture . . . . .	59
6.3.1	Data Model . . . . .	60
6.3.2	Server Architecture . . . . .	60
6.3.3	Client Architecture . . . . .	63
6.4	The Scheduling Algorithm . . . . .	64
6.4.1	Steps . . . . .	64
6.4.2	Analysis . . . . .	66
6.5	Cost Model . . . . .	68
6.6	Experimental Evaluation . . . . .	69
6.6.1	Simulation Environment . . . . .	69
6.6.2	Basic Evaluation . . . . .	72
6.6.3	Further Evaluation . . . . .	78
6.7	Summary . . . . .	81



---

<b>7</b>	<b>Optimal Query Mapping</b>	<b>83</b>
7.1	Motivation	83
7.2	Derivability in Aggregation Lattices	84
7.3	Query Mapping in mobile OLAP Architectures	84
7.4	An Analytical Model for Subsumption Probabilities	87
7.4.1	Model	87
7.4.2	Ancestor Probabilities in DCL	89
7.4.3	Ancestor Probabilities in hDCL	91
7.5	Experimental Evaluation	93
7.5.1	Exploitation of Subsumption	93
7.5.2	Evaluating mOLAP Dissemination Systems	94
7.6	Summary	96
<b>8</b>	<b>Compressed Aggregations for mobile OLAP</b>	<b>99</b>
8.1	Motivation	99
8.2	Background and Related Work	100
8.2.1	Data Cube Physical Structures	100
8.2.2	The Dwarf Data Cube Physical Structure	101
8.2.3	The mOLAP Tradeoff: Compression vs. Client Processing	103
8.2.4	Coarse-grained Dwarfs	103
8.3	The m-Dwarf Data Cube Physical Structure	104
8.3.1	Design	104
8.3.2	Construction	105
8.3.3	Evaluation	107
8.4	m-Dwarfs in FCLOS	109
8.5	Experimental Evaluation	111
8.6	Evaluation of examined physical structures	113
8.7	Summary	114
<b>9</b>	<b>Ad hoc mobile OLAP</b>	<b>115</b>
9.1	Introduction	115
9.2	Background	116
9.2.1	Architecture	116
9.2.2	Ad hoc vs. Infrastructure based	118
9.2.3	Requirements	118
9.3	Related Work	118
9.3.1	Distributed Data in MANETs	118
9.3.2	Ad hoc mOLAP Suitability	120
9.4	Problem Formalization	121
9.4.1	Generated Traffic	121
9.4.2	Query Access Time	121
9.4.3	Energy Consumption	122
9.5	Query and Disseminate under Global View	122
9.5.1	Data Model and Query Mapping	122

---

9.5.2	View Exchange . . . . .	123
9.5.3	Query Propagation . . . . .	125
9.5.4	Runtime Connection Handling . . . . .	126
9.5.5	Data Dissemination . . . . .	127
9.6	Experimental Evaluation . . . . .	128
9.6.1	Simulation Environment . . . . .	128
9.6.2	Reference Application . . . . .	128
9.6.3	Basic Evaluation . . . . .	129
9.6.4	Further Evaluation . . . . .	132
9.7	Open Issues . . . . .	133
9.8	Summary . . . . .	133
<b>IV</b>	<b>Conclusion</b>	<b>135</b>
<b>10</b>	<b>Conclusion and Future Work</b>	<b>137</b>
10.1	Overview . . . . .	137
10.2	Comparison of mOLAP Architectures . . . . .	138
10.3	Future Work . . . . .	141
	<b>References</b>	<b>143</b>
	<b>Appendices</b>	<b>159</b>

# List of Figures

1.1	mOLAP architecture based on FCLOS and QDGV . . . . .	5
2.1	Multidimensional view of Table 2.1 . . . . .	10
2.2	Star schema . . . . .	12
2.3	Data cube . . . . .	13
2.4	DCL of a 3-dimensional data cube . . . . .	14
2.5	hDCL of a 3-dimensional data cube . . . . .	16
3.1	Indexed broadcast composition . . . . .	23
3.2	Index and data organization of distributed indexing . . . . .	24
3.3	On-demand data dissemination . . . . .	24
3.4	Push-based data dissemination . . . . .	25
3.5	Hybrid data dissemination . . . . .	26
4.1	The $R \times W$ service-queue data structures . . . . .	32
4.2	Chunking model of a relation . . . . .	39
5.1	STOBS OLAP System . . . . .	48
5.2	STOBS: Flexibility in DCL . . . . .	49
5.3	SBS: Flexibility in DCL . . . . .	51
6.1	mOLAP in wireless infrastructure based networks . . . . .	58
6.2	Overview of the FCLOS server architecture . . . . .	60
6.3	Waiting queue before and after query mapping . . . . .	61
6.4	FCLOS bucket . . . . .	62
6.5	Steps of the FCLOS scheduling algorithm . . . . .	66
6.6	Histogram of self-similar distribution . . . . .	70
6.7	Mean query access time ( $T_{all}$ ) . . . . .	74
6.8	Mean query access time without offline answers . . . . .	74
6.9	Percentage of $T_Q$ , $T_C$ and $T_L$ to $T_{all}$ . . . . .	75
6.10	Percentage of offline and online query answers . . . . .	75
6.11	Mean energy consumption overhead ( $E_{all}$ ) . . . . .	76
6.12	Percentage of $E_Q$ , $E_C$ and $E_L$ to $E_{all}$ . . . . .	76
6.13	Mean per query generated traffic ( $Tr_q$ ) . . . . .	77

6.14	Mean per broadcast generated traffic ( $Tr_b$ ) . . . . .	77
6.15	Mean queue length and number of elements served or not per broadcast . . . . .	77
6.16	Total amount of generated traffic ( $Tr_{sum}$ ) . . . . .	77
6.17	Mean stretch . . . . .	78
6.18	Mean query access time ( $T_{all}$ ) for different job classes . . . . .	78
6.19	Mean query access time ( $T_{all}$ ) for different values of the $a$ -optimizer . . . . .	79
6.20	Mean query access time ( $T_{all}$ ) vs. dimensionality . . . . .	79
6.21	Mean query access time ( $T_{all}$ ) vs. query distribution . . . . .	79
6.22	Mean query access time ( $T_{all}$ ) vs. query rate . . . . .	80
6.23	Generated traffic ( $Tr_{sum}$ ) vs. query rate . . . . .	80
6.24	Mean query access time ( $T_{all}$ ) vs. bandwidth . . . . .	81
7.1	Aggregation lattices . . . . .	85
7.2	mOLAP dissemination system and two possible types of query mapping . . . . .	86
7.3	Subsumption probabilities $P(e_a \succeq q^+ \subseteq Q)$ and $P(\exists e : e \succeq q \subseteq Q)$ . . . . .	95
7.4	Mean query access time ( $T_{all}$ ) . . . . .	96
7.5	Percentage of queue elements served per broadcast . . . . .	96
7.6	Mean per broadcast generated traffic ( $Tr_b$ ) . . . . .	96
7.7	Mean per query generated traffic ( $Tr_q$ ) . . . . .	96
8.1	Dwarf of Table 8.1 . . . . .	102
8.2	Fully coarse-grained Dwarf of Table 8.1 . . . . .	105
8.3	m-Dwarf of Table 8.1 . . . . .	106
8.4	An example of the temporary structure rm-Dwarf . . . . .	107
8.5	m-Dwarf storage savings for a semi-synthetic dataset . . . . .	110
8.6	m-Dwarf storage savings for a real dataset . . . . .	110
8.7	Mean per query generated traffic ( $Tr_q$ ) . . . . .	112
8.8	Mean per broadcast generated traffic ( $Tr_b$ ) . . . . .	112
8.9	Total amount of generated traffic ( $Tr_{sum}$ ) . . . . .	112
8.10	Mean query access time ( $T_{all}$ ) . . . . .	112
8.11	Percentage of $T_Q$ , $T_C$ and $T_L$ to $T_{all}$ . . . . .	113
9.1	Extended mOLAP architecture . . . . .	117
9.2	Elementary view for a 3-dimensional schema . . . . .	124
9.3	Example of a QDGV client's local data . . . . .	125
9.4	Mean query access time ( $T_{all}$ ) . . . . .	130
9.5	Mean energy consumption overhead ( $E_{all}$ ) . . . . .	130
9.6	Query propagation traffic ( $Tr_{QProp}$ ) . . . . .	131
9.7	Query propagation traffic ( $Tr_{QProp}$ ) with query interval . . . . .	131
9.8	Discovery success . . . . .	132
9.9	Mean query access time with density ( $T_{all}$ ) . . . . .	133
9.10	Mean query access time with velocity ( $T_{all}$ ) . . . . .	133

---

10.1	Mean query access time ( $T_{all}$ ) for all approaches examined in the thesis . . . . .	140
A.1	Parallel coordinates: <i>FCLOS</i> vs. <i>STOBS</i> optimization of mean query access time ( $T_{all}$ ) . . . . .	162
A.2	Parallel coordinates: <i>FCLOS</i> vs. <i>STOBS</i> optimization of mean energy consumption overhead ( $E_{all}$ ) . . . . .	163
A.3	Parallel coordinates: <i>FCLOS</i> vs. <i>STOBS</i> optimization of mean total amount of generated traffic ( $Tr_{sum}$ ) . . . . .	164
A.4	Parallel coordinates: <i>FCLOS<sub>mD</sub></i> vs. <i>STOBS</i> optimization of mean query access time ( $T_{all}$ ) . . . . .	165
A.5	Parallel coordinates: <i>FCLOS<sub>mD</sub></i> vs. <i>STOBS</i> optimization of mean energy consumption overhead ( $E_{all}$ ) . . . . .	166
A.6	Parallel coordinates: <i>FCLOS<sub>mD</sub></i> vs. <i>STOBS</i> optimization of mean total amount of generated traffic ( $Tr_{sum}$ ) . . . . .	167



# List of Tables

2.1	Fact table example . . . . .	10
2.2	Declared hierarchies of a 3-dimensional data cube . . . . .	14
2.3	Derivability for aggregate functions . . . . .	17
4.1	Evaluation symbols . . . . .	37
4.2	Evaluation of data broadcast modes . . . . .	38
5.1	Evaluation of mOLAP architectures . . . . .	52
6.1	FCLOS scheduler notation . . . . .	64
6.2	A scheduling example: STOBS vs. FCLOS . . . . .	67
6.3	Clusters and broadcast weights in FCLOS . . . . .	68
6.4	Schedules: STOBS vs. FCLOS . . . . .	68
6.5	Real data mart metadata . . . . .	69
6.6	Semi-synthetic data mart metadata . . . . .	70
6.7	Query distribution's characteristics . . . . .	71
6.8	Mobile client technical characteristics . . . . .	73
6.9	Overview of simulation parameters . . . . .	74
7.1	Query mapping tradeoff in mOLAP . . . . .	86
7.2	Notation of the analytical framework . . . . .	88
7.3	Subsumption probability $P(e_a \succeq e_b)$ . . . . .	94
8.1	Base fact table . . . . .	101
8.2	Storage size vs. # of tuples . . . . .	108
8.3	Storage size vs. dimensionality . . . . .	108
8.4	Storage size vs. data distribution . . . . .	109
8.5	Evaluation of examined physical structures . . . . .	113
9.1	Breaking Consequences . . . . .	127
9.2	Simulation parameters . . . . .	129
10.1	Evaluation of mOLAP architectures . . . . .	139
A.1	Dimensions used for parallel coordinates . . . . .	161





# Acronyms

ACID	Atomicity, Consistency, Isolation, Durability	26
BQDA	Broadcast Query and Direct Answer	128
DCL	Data Cube Lattice	14
DW	Data Warehouse	1
fcg-Dwarf	fully coarse-grained Dwarf	104
FCLOS	Force Clustering OLAP Scheduler	3
hDCL	hierarchical Data Cube Lattice	15
m-Dwarf	mobile Dwarf	99
MAC	Medium Access Control	23
MANET	Mobile ad hoc network	115
MDDB	Multidimensional Database	9
mOLAP	mobile Online Analytical Processing	1
P2P	Peer to Peer	44
QDGV	Query and Disseminate under Global View	4
SBS	Subsumption-Based Scheduler	50
ST	Summary Table	12
STOBS	Summary Tables On-Demand Broadcast Scheduler	48

