

Context Sensitive Database Summarisation

by

Darin Chan, *B.Eng.(IT&T)(Hons)*
School of Informatics and Engineering,
Faculty of Science and Engineering

21st October 2005

A thesis presented to the
Flinders University of South Australia
in total fulfillment of the requirements for the degree of
Doctor of Philosophy

Table of Contents

Abstract	ix
Certification	x
Acknowledgements	xi
1 Introduction	1
1.1 Mobile Computer Technologies	2
1.2 Mobile Database Issues	4
1.2.1 Power and Bandwidth Capacities	4
1.2.2 Storage Capacity	5
1.3 Problem Statement	6
1.4 Motivating Example	6
1.5 Contribution of this Dissertation	8
1.6 Dissertation structure	9
2 A Survey on Mobile Databases	10
2.1 Mobile Wireless Technologies	11
2.2 Mobile Database Architecture	12
2.2.1 Mobile Operation modes	13
2.3 Mobile Database Functionalities	15
2.3.1 Query Processing and Optimisation	15
2.3.2 Data Replication Schemes	17
2.3.3 Concurrency Controls	18
2.3.4 Transaction Support	20
2.3.5 System Recovery	21
2.4 Summary	22

3	A Survey of Data Summarisation	23
3.1	Schema Modification Techniques	24
3.1.1	Projection (vertical) methods	24
3.1.2	Selection (horizontal) methods	25
3.1.3	Hybrid methods	25
3.2	Domain Modification Techniques	27
3.2.1	Concept Hierarchies	27
3.2.2	Abstraction through aggregation	27
3.2.3	Surrogate	29
3.3	Data Allocation of Fragments	30
3.4	Data Compression	31
3.5	Discussion	32
4	The Uses of Context in Mobile Computing	33
4.1	Location-aware techniques	34
4.2	Information Retrieval	35
4.3	Data Reduction	36
4.4	Sensors	36
4.5	Characteristics of Context	37
4.6	Summary	37
5	COntext Sensitive MOBILE Summarisation (COSMOS)	39
5.1	COSMOS Database Architecture	40
5.1.1	Update Database	41
5.2	A Protocol for Priority-based Data Summarisation (First Stage)	42
5.2.1	Primary Criteria	44
5.2.2	Secondary Criteria	49
5.2.3	Tertiary Criteria	52
5.2.4	User-defined	56
5.3	Summarisation Stage Two	56
5.3.1	Storage Maps	57
5.4	Discussion	59

6	Local Nulls	60
6.1	Previous Research and Motivation	61
6.1.1	Literature Survey	61
6.1.2	Motivation	62
6.2	Notation	63
6.3	Local Nulls and Relational Algebra	65
6.3.1	Set Theory Operations	65
6.3.2	Select and Project Operations	67
6.3.3	Join Operations	69
6.3.4	Division Operation	71
6.4	Local Nulls and SQL	72
6.5	Discussion	74
7	Transaction Management	76
7.1	Transaction Model	76
7.1.1	Transaction Structure	78
7.1.2	Data Consistency	79
7.1.3	Recovery	80
7.1.4	Transaction Relocation	81
7.2	Consistency in Local Operations	82
7.2.1	Correctness Criterion	83
7.2.2	Serialisation Graph	85
7.3	Consistency Reconciliation	86
7.3.1	Correctness Criterion	87
7.3.2	Serialisation Graph	88
7.4	Discussion	89
8	Transaction and System Failure Protocols	91
8.1	Transformation Function	91
8.2	Serialisability	93
8.3	Reconciliation	95
8.4	Local Update Propagation	96
8.5	Update Through Local Item Invisibility	97

8.6	Bounded Divergence	98
8.7	Database Refresh	99
8.8	System Recovery	101
8.9	Summary	103
9	Evaluation of COSMOS	104
9.1	Quantitative Evaluations	104
9.2	Performance Model	105
9.2.1	Resource Contention Analysis	106
9.2.2	Data Contention Analysis	110
9.2.3	Reconciliation	117
9.3	Performance Evaluation	120
9.3.1	Communication Cost	120
9.3.2	Transaction Response Time	121
9.3.3	Reconciliation Costs	122
9.4	System Simulation	124
9.4.1	Summarisation Engine Implementation	124
9.4.2	Simulation Implementation	125
9.4.3	Simulation Results	127
9.5	Discussion	129
10	Conclusion	130
10.1	Contributions	130
10.2	Future Work	133
A	Acronyms and Terminology	134
A.1	Acronyms	134
A.2	Terminology	136
B	Example	137
	Bibliography	143

List of Figures

2.1	Architecture for a Mobile Database	12
2.2	Storage Map	17
3.1	Hybrid Fragmentation Methods	26
3.2	Applicability of Techniques to Different Data Types	30
5.1	Architecture for a COSMOS Database	40
5.2	Conceptual View of the Summarisation Process	43
5.3	A medical E-R scheme, after (Golfarelli et al. 1998)	50
5.4	Example relation with summarisation priorities	57
5.5	View of relation in Figure 5.4 with local nulls	57
5.6	An example of a storage map	58
9.1	Communication Cost	121
9.2	Response time for Local Transactions	121
9.3	Response time for Global Transactions	122
9.4	Response time for Strict Transactions	123
9.5	Probability of abort for Central Database (CDB) and server Summary Database (SDB)	123
9.6	Overview of the Summarisation Process	124
9.7	Response time for Local Transactions	127
9.8	Response time for Global Transactions	128
9.9	Response time for Strict Transactions	128

List of Tables

5.1	Criteria Categories	42
5.2	Example of a criteria protocol	44
5.3	Example of a set of contextual rules.	46
6.1	Example of Local Extension and Local Completion.	64
6.2	Union Compatible Tables	65
6.3	Operations using null substitution principle	65
6.4	Union and Intersection Evaluations	67
6.5	Relation Example	68
6.6	Projection Example	68
6.7	Multiple Conditions Example	69
6.8	Join Example	69
6.9	Conventional Join	69
6.10	Generalised Join	70
6.11	Local Join	70
6.12	Summary of Local Join Operations	71
6.13	Division Example	71
6.14	Local Null Division Example	72
6.15	Summary of Local Join Operations in SQL	73
6.16	Example Results	74
8.1	Storage Map Examples	93
8.2	Locking Matrix 1	94
8.3	Locking Matrix 2	95
8.4	Protocols to maintain bounded inconsistencies	99

9.1	Input parameters	120
B.1	The setting of ρ for each data inclusion criterion.	137
B.2	Enumerated patient's list.	138
B.3	Contextual patient's list.	138
B.4	<i>Patient</i> Table	139
B.5	<i>Town, Physician, Admission</i> Tables	140
B.6	<i>Admission</i> Table after Inductive criterion used.	140
B.7	<i>Patient</i> Table	141
B.8	<i>Admission</i> Tables	142
B.9	<i>Patient</i> and <i>Admission</i> Table	142

Abstract*

The use of databases is a common practice for storing and organising large amounts of data. In a distributed environment, especially widely distributed systems, replication provides a means for storing an entire database locally. However, in an environment where mobile devices are used, resource limitations on both storage and power capacity eliminates the possibility of full replications. Moreover, reliable access to centralised databases is not always possible in these environments. Consequently, there is a need to find ways of storing data on a mobile computer in such a way as to maximise user access while providing answers at an acceptable level of accuracy.

This dissertation argues that it is useful to know and incorporate the user context in which the database will be used in the creation process of the mobile database. It investigates a context sensitive summarisation technique and provides a proof of concept prototype, COSMOS, which maximises data availability based on the context in which the system is being used and the accuracy required by the user. More specifically, a framework for summarising data using the context of the user is introduced.

Local nulls are proposed to the relational algebra and SQL to allow efficient querying of summary databases. Some modifications to the database management system include the handling of transactions in the system, update propagation and system failure protocols. These modifications are proposed to allow the effective use of the summarised data.

*This research was partially supported by Australian Research Council SPIRT Grant C00107533, held in collaboration with HP Labs.

Certification

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

As requested under Clause 14 of Appendix E of the *Flinders University Research Higher Degree Student Information Manual* I hereby agree to waive the conditions referred to in Clause 13(b) and (c), and thus

- Flinders University may lend this thesis to other institutions or individuals for the purpose of scholarly research;
- Flinders University may reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signed

Dated

Darin Chan

Acknowledgements

I would like to thank my supervisor, Professor John Roddick, for his invaluable help in getting me through my candidature.

To my colleagues at the Flinders University of South Australia, School of Informatics and Engineering, I would like to say a grateful thank you. In particular, to Denise de Vries and Aaron Ceglar for their help and encouragement throughout my candidature. I would also like to thank the academics and staff of that school for their continual support over the past few years.

Flinders University of South Australia, in collaboration with HP, had funded my research for the past few years. I would also like to extend my thanks to all the people who were involved there.

Finally, I would like to thank all my friends and family who had helped me during the period of my candidature. In particular, to my parents for their love, support and their confidence in me, and to Adeline Ong for her invaluable help and encouragement.

Darin Chan
October 2005
Adelaide.