Management of Time Series Data

by

Abel Matus Castillejos

A thesis submitted to the
School of Information Sciences and Engineering
University of Canberra

in partial fulfilment of the requirements
for the award of the degree of
Doctor of Information Technology

December 2006

Canberra, Australia
Abstract

Every day large volumes of data are collected in the form of time series. Time series are collections of events or observations, predominantly numeric in nature, sequentially recorded on a regular or irregular time basis. Time series are becoming increasingly important in nearly every organisation and industry, including banking, finance, telecommunication, and transportation. Banking institutions, for instance, rely on the analysis of time series for forecasting economic indices, elaborating financial market models, and registering international trade operations. More and more time series are being used in this type of investigation and becoming a valuable resource in today’s organisations.

This thesis investigates and proposes solutions to some current and important issues in time series data management (TSDM), using Design Science Research Methodology. The thesis presents new models for mapping time series data to relational databases which optimise the use of disk space, can handle different time granularities, status attributes, and facilitate time series data manipulation in a commercial Relational Database Management System (RDBMS). These new models provide a good solution for current time series database applications with RDBMS and are tested with a case study and prototype with financial time series information. Also included is a temporal data model for illustrating time series data lifetime behaviour based on a new set of time dimensions (confidentiality, definitiveness, validity, and maturity times) specially targeted to manage time series data which are introduced to correctly represent the different status of time series data in a timeline. The proposed temporal data model gives a clear and accurate picture of the time series data lifecycle. Formal definitions of these time series dimensions are also presented. In addition, a time series grouping mechanism in an extensible commercial relational database system is defined, illustrated, and justified. The extension consists of a new data type and its corresponding rich set of routines that support modelling and operating time series information within a higher level of abstraction. It extends the capability of the database server to organise and manipulate time series into groups. Thus, this thesis presents a new data type that is referred to as GroupTimeSeries, and its corresponding architecture and support functions and operations. Implementation options for the GroupTimeSeries data type in relational based technologies are also presented.
Finally, a framework for TSDM with enough expressiveness of the main requirements of time series application and the management of that data is defined. The framework aims at providing initial domain know-how and requirements of time series data management, avoiding the impracticability of designing a TSDM system on paper from scratch. Many aspects of time series applications including the way time series data are organised at the conceptual level are addressed. The central abstraction for the proposed domain specific framework is the notions of business sections, group of time series, and time series itself. The framework integrates comprehensive specification regarding structural and functional aspects for time series data management. A formal framework specification using conceptual graphs is also explored.
Acknowledgements

The true roots for this research began with my professional experience at Banco de México (Mexico’s Central Bank). The challenges that I encountered when I dealt with time series data while working there motivated me to undertake this research.

My sincere gratitude goes to my two main supervisors, Dr Ric Jentzsch and Dr Masoud Mohammadian. I am grateful to Dr Jentzsch for accepting me as doctoral student at the School of Information Sciences and Engineering (ISE) and also for his invaluable guidance, ideas, and support throughout my research study. Thanks are due to Dr Masoud Mohammadian for accepting the role as my main supervisor when Dr Jentzsch moved to a new position at Compucat Research Pty Ltd. Discussions with Dr Mohammadian were also very important in improving my research work through constructive suggestions at different phases of this thesis. In addition, I appreciate his time and valuable observations during multiple readings of the thesis. I would also like to thank the other member in my supervision panel committee, Assoc Prof Dharmendra Sharma, Head of the School of ISE, whose inputs to my work gave me a completely different view, especially in the latter stages of the research.

I would like to express my appreciation to Dr Joelle Vandermensbrugghe, convenor of the Research Education Program at the University of Canberra, for her generous time in reading and correcting my research writings throughout this project. Special gratitude goes to Sue Prentice for her professional editing advice according to academic standards prior to submission of the thesis.

Many thanks also go to many people in the School of Information Sciences and Engineering (ISE) at the University of Canberra, Australia.

My greatest thanks are reserved for my beloved wife Berthita ‘Esposita Bonita’, to whom this thesis is dedicated. I hope she will be proud of this work; somewhere behind each line of this document is part of our Matus & Ordaz Adventure Team project. We were jointly on a journey during which we were both students. During the time we lived here in Canberra she undertook an MBA (Master of Business Administration) program at the Australian National University.
I hope my mother (Victoria) and my father (Faustino) will also be proud of this work; the achievement is also theirs. The greatest gifts I received from them are my education and determination. I am proud to be their son and I thank them for their endless love. I also give thanks to my brothers and sisters, nieces and nephews for cheering me on to finish this degree and “Paty”, my sister-in-law, for her help and support.

I would also like to express my special gratitude to my parents-in-law and sister-in-law for their support and motivation during this journey. Mirnita, many thanks for being a witness and a part of this project. I love you all.

Finally, I appreciate the support of the Mexican National Council of Science and Technology (CONACYT) in sponsoring this doctoral study, scholarship number 129475/192300.

Abel Matus Castillejos
Canberra, Australia
December 2006
Dedication

To my lovely wife:

Berthita Bonita

Her love and patience, her constant encouragement and support provided a solid foundation for me throughout this work, and steered me along a road that was sometimes difficult.

My love, this is only for the sake of our family.

I am sure we will be able to spend more time together now.

I love you more now and I always will.
Table of Contents

Abstract .................................................................................................................................................. ii
Acknowledgements ........................................................................................................................ iv
Dedication ............................................................................................................................................ vi
Certificate of Authorship of Thesis .................................................................................................. vii
List of Figures ..................................................................................................................................... xii
List of Tables ....................................................................................................................................... xv
List of Acronyms ............................................................................................................................... xvi
List of Publications from Current Research ....................................................................................... xvii

Chapter 1: Introduction ......................................................................................................................... 1
1.1 Background .................................................................................................................................. 1
1.2 Motivation and Goals .................................................................................................................. 3
1.3 Contributions ............................................................................................................................... 5
1.4 Methodology ............................................................................................................................... 7
1.5 Delimitations of Scope ................................................................................................................ 8
1.6 Thesis Outline .............................................................................................................................. 9
1.7 Chapter Summary ......................................................................................................................... 11

Chapter 2: Time Series Data Management in the Literature .............................................................. 12
2.1 Introduction .................................................................................................................................. 12
2.2 Data Types Universe ................................................................................................................... 12
2.3 Temporal Model .......................................................................................................................... 15
2.4 Calendar for Time Series ............................................................................................................ 15
2.5 Temporal Data ............................................................................................................................ 16
2.6 Overview of Temporal Database ................................................................................................ 18
2.6.1 Notions of Time Dimension .................................................................................................. 20
2.6.2 Temporal Dimensions Reported in the Literature ................................................................. 21
2.6.3 Temporal Dimensions for Specific Domains ......................................................................... 26
2.7 Overview of Time Series Database ............................................................................................. 29
2.7.1 Time Series Properties .......................................................................................................... 29
2.7.2 Basic Concepts ....................................................................................................................... 30
2.8 Overview of Existing Solutions for Managing Time Series Data ............................................. 33
2.8.1 Relational Model ................................................................................................................... 33
2.8.2 Object-Relational Model: Major Vendors (Oracle, Informix, Sybase, IBM) Solutions .......... 34
2.8.3 Object-Oriented Model: Vision & Calanda .......................................................................... 35
2.8.4 Proprietary Approach: Fame, TimeIQ, & DBank ................................................................. 36
2.8.5 Academic Studies and Prototypes ......................................................................................... 37
2.9 Challenges in Time Series Data Management ............................................................................. 38
2.9.1 Mapping Time Series Data to Relational Database ............................................................ 38
2.9.2 Formal Definition of Time Series Dimensions ..................................................................... 39
2.9.3 On the Need for a Time Series Grouping Mechanism .......................................................... 41
2.9.4 Framework for Time Series Data Management .................................................................... 41
2.10 Chapter Summary ....................................................................................................................... 42
Chapter 3: Research Methodology ........................................................................................................... 43

3.1 Introduction .......................................................................................................................................... 43
3.2 Research Philosophy ............................................................................................................................. 44
3.2.1 Research Approach – Underlying Ontology .................................................................................. 44
3.2.2 Research Approach – Underlying Epistemology .......................................................................... 45
3.2.3 Research Approach – Underlying Axiology .................................................................................. 48
3.3 Research Methodology .......................................................................................................................... 48
3.4 Information System Research Methodology .......................................................................................... 50
3.5 Design-Science as a Research Methodology .......................................................................................... 51
3.5.1 The Research Process of Design-Science Research Methodology .................................................. 58

1. Construct Conceptual Framework ......................................................................................................... 58
2. Develop System Architecture ................................................................................................................ 59
3. Analyse and Design the System ............................................................................................................ 59
4. Build the System ....................................................................................................................................... 60
5. Observe and Evaluate the System .......................................................................................................... 60
3.5.2 Evaluation Criteria .............................................................................................................................. 60
3.5.3 Strengths and Limitations of the DSRM ............................................................................................ 62
3.6 Research Process Proposed in this Research .......................................................................................... 63
1. Develop Conceptual Framework ............................................................................................................ 63
2. Establish Framework Formalisation ......................................................................................................... 63
3. Develop Systems Architecture ................................................................................................................. 63
4. Analyse and Design the System ............................................................................................................ 63
5. Build the System ....................................................................................................................................... 63
6. Observe and Evaluate the System .......................................................................................................... 63
3.7 Chapter Summary .................................................................................................................................. 63

Chapter 4: Models for Mapping Time Series Data to Relational Databases ............................................. 65

4.1 Introduction .......................................................................................................................................... 65
4.2 Models for Mapping Time Series Data to Relational Databases .......................................................... 67
4.2.1 Modelling One Relational Table per Time Series .......................................................................... 67
4.2.2 Modelling One Relational Table per Time Series Type .................................................................... 69
4.2.3 Modelling Time Series Metadata .................................................................................................... 71
4.3 New Models for Mapping Time Series Data to Relational Databases .................................................. 72
4.3.1 Modelling Univariate Time Series .................................................................................................... 73
4.3.2 Modelling Multivariate Time Series .................................................................................................. 75
4.3.3 Modelling Irregular Time Series ...................................................................................................... 76
4.4 Comparisons amongst the Models ........................................................................................................ 77
4.5 Validation .............................................................................................................................................. 78
4.5.1 Case Study: Daily Stock Prices ......................................................................................................... 79
4.5.2 System Implementation ..................................................................................................................... 82
4.5.3 Evaluation Criteria ............................................................................................................................ 83
4.6 Chapter Summary .................................................................................................................................. 86
Chapter 7: A Time Series Data Management Framework .............................................. 125

7.1 Introduction ........................................................................................................ 125
7.2 Characteristics of Time Series Applications ...................................................... 126
7.3 Time Series Data Management .......................................................................... 128
    7.3.1 The Framework .......................................................................................... 128
    7.3.2 The Framework Architecture ...................................................................... 132
    7.3.3 General Principles ..................................................................................... 133
    7.3.4 Additional Requirements .......................................................................... 135
    7.3.5 Management Aspects ................................................................................. 135
7.4 Fundamentals of Conceptual Graphs ................................................................ 138
7.5 Framework Formalisation ................................................................................. 139
    7.5.1 General Rules ........................................................................................... 139
7.6 Framework Implementation Option ................................................................. 144
7.7 Chapter summary ............................................................................................... 145

Chapter 8: Conclusion and Future Work ................................................................ 146

References ................................................................................................................ 150