

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Appendix D

MASSEY UNIVERSITY

Application for Approval of Request to Embargo a Thesis  
(Pursuant to AC98/168 (Revised 2), Approved by Academic Board 17/02/99)

Name of Candidate: JEFFREY PAUL BRIDGES ID Number: 12159110

Degree: MASTER OF ENGINEERING Dept/Institute/School: SEAT

Thesis title: Development of a multi-historian  
SCADA data collection middleware

Name of Chief Supervisor: Edmund Lai Telephone Ext: 41580

As author of the above named thesis, I request that my thesis be embargoed from public access until (date) 01/03/2015 for the following reasons:

- Thesis contains commercially sensitive information.
- Thesis contains information which is personal or private and/or which was given on the basis that it not be disclosed.
- Immediate disclosure of thesis contents would not allow the author a reasonable opportunity to publish all or part of the thesis.
- Other (specify): .....

Please explain here why you think this request is justified:

The thesis contains details of the  
implementation of a commercial  
product. Making this information  
public immediately could provide  
an advantage to their competitors.

Signed (Candidate): Bridges - Date: 29/04/2013

Endorsed (Chief Supervisor): Name Date: 29/4/2013

Approved/Not Approved (Representative of VC): [Signature] Date: 2/05/13

Note: Copies of this form, once approved by the representative of the Vice-Chancellor, must be bound into every copy of the thesis.

# Development of a multi-historian SCADA data collection middleware

A thesis presented in partial fulfilment of the  
requirements for the degree of

Master of Engineering  
in  
Computer and Electronic Engineering

at Massey University, Albany, New Zealand

Jeffrey Paul Bridges  
2013

## **Abstract**

This thesis details the implementation of a SCADA middleware package designed to provide a common interface to the most commonly used SCADA historians in use within local government in New Zealand. The middleware integrates with a New Zealand developed cloud based solution that is used throughout the country for compliance monitoring and reporting of water and wastewater treatment facilities.

The middleware connects via the internet to the historians hosted on-site within council networks and retrieves data which is replicated in a local database. This approach provides backup for the data and allows it to be accessed quickly when required for reporting. The middleware database is hosted within a Microsoft SQL Server instance to ensure compatibility and ease of rollout when paired with the commercial cloud solution.

To ensure reliable data collection and resilience to connection interruption all transactions made against the client historians are controlled by a queue stored in the middleware database. The middleware includes an inbuilt mechanism for integrity checking to ensure all data available in the client historian is collected and has been designed to run with minimal intervention of company staff.

To reduce storage requirements the middleware includes a data deduplication system that removes repeated samples from the SQL Server database after integrity checking has been performed. This provides a lossless compression mechanism that does not alter the precision of the data collected.

The middleware implementation has now been in use in a production environment collecting data from councils for compliance purposes for approximately six months. During this time data collection has been very reliable with the middleware handling numerous connection outages without intervention.

## **Acknowledgements**

Thanks to Associate Professor Edmund Lai for providing supervision and guidance over the course of this project.

I am very grateful to Dr Peter Johnson for providing the opportunity to work on this project, and providing his support and guidance over the past year.

This project was partially funded by the Ministry of Science + Innovation. Thanks to Chris Lock for his efforts in securing this funding.

This project would not have been possible without support from the IT and utilities staff at a number of councils around New Zealand who provided access to their operational data.

A special thanks to my partner Alexandra for her support and understanding during the long hours required to complete this project.

Thanks to my family who provided support and motivation throughout the project.

# Contents

Abstract.....	i
Acknowledgements.....	ii
List of Figures .....	vii
List of Tables .....	ix
Chapter 1: Introduction .....	1
1.1 Motivation.....	1
1.2 Objectives.....	1
1.3 Thesis Overview .....	2
Chapter 2: Literature Review .....	4
2.1 Background Information .....	4
2.2 Theory .....	5
2.2.1 SCADA Historians .....	5
2.2.2 Historian Data Access.....	6
2.2.3 Data Transport and Security .....	7
2.2.4 Data Collector Architecture .....	8
2.2.5 Data Storage.....	8
2.2.6 Common Interface .....	9
2.2.7 Development Methodologies .....	10
2.3 Summary .....	12
Chapter 3: SCADA Historians .....	13
3.1 Citect Historian.....	13
3.2 GE Proficy Historian .....	13
3.3 Invensys Wonderware Historian.....	14
3.4 QTech DATRAN .....	14
3.5 Hilltop DataTamer.....	14
Chapter 4: Existing Cloud Solution.....	16
4.1 Production Hardware.....	16
4.2 Production Software .....	16
4.3 Production Virtual Machine Software .....	17
4.4 Production Network.....	17
4.5 Production Schematic .....	18
4.6 Implementation Requirements.....	18
Chapter 5: Investigation and Planning.....	20

5.1	Test Environment Hardware .....	20
5.1.1	Hardware.....	20
5.2	Test Environment Software .....	21
5.2.1	Hypervisor .....	21
5.2.2	Middleware Virtual Machine .....	22
5.2.3	Reporting Virtual Machine .....	23
5.2.4	Citect Historian Virtual Machine .....	24
5.2.5	GE Proficy Virtual Machine .....	25
5.2.6	Invensys Wonderware Virtual Machine.....	25
5.2.7	QTech DATRAN Virtual Machine.....	26
5.2.8	Hilltop Virtual Machine .....	27
5.2.9	Large Dataset Virtual Machine.....	27
5.2.10	Development Workstation.....	28
5.2.11	Network Overview .....	29
5.3	Common Interface .....	29
5.3.1	Tag Listing.....	30
5.3.2	Tag Collection Status.....	31
5.3.3	Tag Data Retrieval .....	31
5.3.4	Tag Data Backfill.....	33
5.4	Data Access Methodologies .....	34
5.4.1	Pass-through .....	34
5.4.2	Collection and Storage .....	35
5.5	Reliable Data Collection .....	37
5.5.1	Data Integrity Checking.....	38
5.5.2	Last Connection Time Tracking .....	39
5.5.3	Queued Data Collection .....	40
5.5.4	Queued Data Collection and Integrity Checking .....	41
5.6	Secure Data Collection .....	42
5.6.1	Source Address Whitelisting .....	43
5.6.2	Authentication .....	43
5.6.3	Source Address Whitelisting with Authentication .....	44
5.6.4	SSL Encryption .....	44
5.6.5	VPN Tunnelling.....	44
5.7	Data Compression .....	46
5.7.1	Sampling Deadband .....	46

5.7.2	Sample Deduplication .....	48
Chapter 6:	Middleware Implementation .....	51
6.1	Application Runtime Environment.....	51
6.2	Middleware Architecture .....	51
6.2.1	Middleware Windows Service .....	52
6.2.2	Middleware Database .....	52
6.2.3	Common Interface Assembly .....	53
6.3	Database Schema .....	53
6.3.1	DataSource Table .....	54
6.3.2	Tag Table .....	55
6.3.3	Sample Table .....	56
6.3.4	Queue Table .....	57
6.3.5	Config Table .....	58
6.4	Middleware Configuration .....	58
6.4.1	Configuration Data Types.....	58
6.4.2	Configuration Parameters.....	59
6.5	Transaction Queue .....	60
6.5.1	Queue Operations.....	60
6.5.2	Operation Status .....	61
6.5.3	Operation Priorities.....	62
6.5.4	Queue Pruning .....	62
6.6	Main Application Loop .....	63
6.7	Worker Application Loop .....	66
6.8	Data Integrity Checking.....	68
6.9	Data Cleaning .....	70
6.10	Historian Data Provider Interface .....	72
6.10.1	Data Retrieval.....	72
6.10.2	Data Integrity Checking.....	72
6.10.3	Tag Verification .....	73
6.10.4	Tag Listing.....	73
6.11	Historian Data Providers .....	74
6.11.1	Citect Historian Data Provider .....	74
6.11.2	GE Proficy Historian Data Provider .....	75
6.11.3	Invensys Wonderware Historian Data Provider.....	77
6.11.4	QTech DATRAN Data Provider .....	78

6.11.5	Hilltop DataTamer Data Provider .....	79
6.12	Connection Security .....	80
6.13	Middleware Deployment .....	80
6.14	Queue Management Tool .....	81
Chapter 7:	Conclusions .....	82
7.1	Conclusions .....	82
7.2	Future Enhancements .....	83
7.2.1	Parallel Worker Threads.....	84
7.2.2	Queue Item Failure Count.....	84
7.2.3	Self-Service Backfill Tools.....	84
7.2.4	Data Collection Processes .....	85
7.2.5	Pass-Through Data Access .....	85
References	.....	87

## List of Figures

Figure 1: The Waterfall method (Shore & Warden, 2008).....	10
Figure 2: The Iterative method (Shore & Warden, 2008).....	11
Figure 3: Agile development (Shore & Warden, 2008).....	11
Figure 4: Production environment schematic .....	18
Figure 5: Hypervisor type comparison .....	22
Figure 6: Development network .....	29
Figure 7: Common interface GetTags response data table schema .....	30
Figure 8: Common interface SetCollectionStatus method inputs .....	31
Figure 9: Common interface GetData method inputs .....	32
Figure 10: Common interface GetData response table schema .....	32
Figure 11: Common interface QueueDataBackfill method inputs.....	33
Figure 12: Pass-through connection data access.....	34
Figure 13: Data collection and storage .....	36
Figure 14: Stored data access .....	36
Figure 15: Simple data integrity checking scheme .....	38
Figure 16: Last connection time tracking scheme .....	39
Figure 17: Example queue showing data sync operations.....	40
Figure 18: Queue item processing .....	41
Figure 19: Example queue showing data sync and data integrity check operations.....	42
Figure 20: Source address whitelisted connection .....	43
Figure 21: VPN tunnelled connection .....	45
Figure 22: Sampling with a 10% deadband.....	46
Figure 23: Sampling with a fixed deadband of 1.....	47
Figure 24: Results of compression with varying deadbands.....	48
Figure 25: Report generation with raw data .....	49
Figure 26: Report generation with deduplicated data .....	49
Figure 27: Results of compression with sample deduplication .....	50
Figure 28: Implemented middleware architecture.....	52
Figure 29: Middleware database schema .....	54
Figure 30: Main program loop .....	65
Figure 31: Middleware worker loop .....	67
Figure 32: Extended data integrity checking scheme .....	69
Figure 33: Data compression algorithm.....	71

Figure 34: Historian data provider GetData request and response .....	72
Figure 35: Historian data provider GetCount request .....	73
Figure 36: Historian data provider CheckTagExists request .....	73
Figure 37: Historian data provider GetTagList request and response .....	74
Figure 38: GE Proficy Connection Methodology.....	75
Figure 39: Middleware Queue Tool .....	81

## List of Tables

Table 1: Citect Tag Mapping .....	74
Table 2: Citect Sample Mapping .....	75
Table 3: GE Proficy Tag Mapping .....	76
Table 4: GE Proficy Sample Mapping .....	76
Table 5: Invensys Wonderware Tag Mapping.....	77
Table 6: Invensys Wonderware Sample Mapping .....	77
Table 7: QTech DATARN Tag Mapping.....	78
Table 8: QTech DATRAN Analog Sample Mapping .....	78
Table 9: QTech DATRAN Digital Sample Mapping .....	79
Table 10: Hilltop DataTamer Tag Mapping .....	80
Table 11: Hilltop DataTamer Sample Mapping .....	80