

University of Wollongong - Research Online

Thesis Collection

Title: Synchronizing data stream processing

Author: Mohammad Siddique Fawad Qureshi

Year: 2007

Repository DOI:

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Research Online is the open access repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

2007

Synchronizing data stream processing

Mohammad Siddique Fawad Qureshi
University of Wollongong

Follow this and additional works at: <https://ro.uow.edu.au/theses>

University of Wollongong

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Recommended Citation

Qureshi, Mohammad Siddique Fawad, Synchronizing data stream processing, M.Comp.Sc.-Res. thesis, Information Technology and Computer Science, University of Wollongong, 2007. <http://ro.uow.edu.au/theses/649>

NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Synchronizing Data Stream Processing

A thesis submitted in fulfillment of the requirements for the award of the degree

Master of Computer Science (Research)

from

UNIVERSITY OF WOLLONGONG

by

M. S. Fawad Qureshi

B.Sc Computer Science – University of Sindh
M.Sc Computer Science – University of Sindh

SITACS

School of Information Technology and Computer Science

2007

© Copyright 2007
by
M. S. Fawad Qureshi
All Rights Reserved

Certification

I, M. S. Fawad Qureshi, declare that this thesis, submitted in fulfillment of the requirements for the award of Master of Computer Science, in the School of Information Technology and Computer Science, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

M. S. Fawad Qureshi

Date: 30 March 2007

Table of Contents

Chapter 1

Introduction.....	2
1.1 The Problem.....	3
1.2 Strategy and Objectives.....	4
1.3 Outline of the Thesis.....	5

Chapter 2

Background and Related Work	7
2.1 Data Stream Processing.....	7
2.2 Adaptive Query Processing.....	11
2.3 Continuous Query Processing.....	15
2.4 Synchronization Techniques.....	17

Chapter 3

Data Stream Processing Networks.....	21
3.1 Network Model.....	21
3.1.1 Components.....	22
3.1.2 Computation.....	23
3.1.3 Operations.....	24
3.2 Windows	26

3.2.1	Multiple Operations on a Window.....	27
3.3	Visualization.....	27
3.3.1	Sample Networks.....	27
3.3.2	Translation.....	30

Chapter 4

Synchronization Problems.....	32
4.1	Data Processing Techniques..... 32
4.2	Motivating Examples..... 33
4.2.1	Example 1..... 33
4.2.2	Example 2..... 35
4.2.3	Example 3..... 35
4.3	Transactional Interpretation of Data Stream Network..... 38
4.3.1	Merge and Split 40
4.4	Revised Motivating Examples..... 41

Chapter 5

Synchronization Strategy and Correctness.....	44
5.1	Synchronization Assistants..... 44
5.1.1	Collector..... 44
5.1.2	Regulator..... 45
5.2	Synchronization Rules..... 46

5.3	Solution.....	48
5.3.1	Solution1.....	48
5.3.2	Solution 2.....	49
5.3.3	Solution 3.....	49
5.3.4	Merge and Split.....	50
5.4	Correctness.....	50
5.4.1	Transactional Interpretation of Rules.....	51
5.4.2	Theorem.....	52
 Chapter 6		
	Conclusion.....	55
	References.....	58

List of Figures

Chapter 3

Figure 3.1: Upon arrival, a group of data items s is recorded to a window over an input stream.....	22
Figure 3.2: An edge from a writer to a window represents a write.....	23
Figure 3.3: An edge from a window to an operation represents a read.....	23
Figure 3.4: An operation with a group of data items as input and produces groups of data items as output.....	24
Figure 3.5: A writer records contents of a data item into a window.....	25
Figure 3.6: A situation where merge is performed on two groups of data items.....	26
Figure 3.7: A situation where many operations read from a window.....	27
Figure 3.8: A data stream processing network for relational algebra expression $[(r \bowtie s) - t]$	28
Figure 3.9: A data stream processing network for an arithmetic expression $[(a + b) - (y + z)]$	29

Chapter 4

Figure 4.1: Dataflow processing network for a relational algebra expression of example 1.....	34
Figure 4.2: Dataflow processing network for an arithmetic expression of	

example 1.....	34
Figure 4.3: Dataflow processing network for a relational algebra expression of example 2.....	36
Figure 4.4: Dataflow processing network for an arithmetic expression of example 2.....	36
Figure 4.5: Dataflow processing network for a relational algebra expression of example 3.....	37
Figure 4.6: Dataflow processing network for an arithmetic expression of example 3.....	37
Figure 4.7: Dataflow processing network for a relational algebra expression with sequences of transactional operations.....	39
Figure 4.8: Dataflow processing network for an arithmetic expression with sequences of transactional operations.....	39
Figure 4.9: A situation where a transaction splits into two sub transactions.....	40
Figure 4.10: Schedule for data stream processing network of figure 4.7.....	41
Figure 4.11: Serialization graph for T_i and T_j , where T_i is accessed by T_j ...	42
Figure 4.12: Serialization Graph for schedule of figure 4.10 contains a cycle.....	43

List of Tables

List of Publications

Publications arising from this thesis:

Qureshi, M. S. F. and Getta, J. R. (2007): Synchronizing Data Stream Processing. *Proc. IASTED International Conference on Parallel and Distributed Computing and Networks*, Innsbruck, Austria, 233 – 238.

Peer-reviewed proceedings of an international conference.

Abbreviations

AQP	Adaptive Query Processing
ART	Average Response Time
CQL	Continuous Query Language
CQP	Continuous Query Processing
DBMS	Data Base Management System
DSMS	Data Stream Management System
DSP	Data Stream Processing
DSPN	Data Stream Processing Network
MDR	Maximum Data Rate
PQP	Pipelined Query Processing
QoS	Quality of Service
QP	Query Processor
SPE	Stream Processing Engine
STREAM	Stanford Stream Data Manager
SQL	Structured Query Language
TQL	Tapestry Query Language
XML	Extended Markup Language

to my beloved father
Haji Muhammad Saleem Qureshi

Abstract

Synchronization of data stream processing has a significant impact on performance of systems where processing of long sequences of data items needs to be done simultaneously. In earlier works on stream processing, synchronization has been discussed to a limited extent or has been completely overlooked. This work describes a formal model of synchronization in a data stream processing network. We use a notation of data stream processing networks to identify circumstances that necessitate synchronization. We also express processing of groups of data items in terms of database transactions within a data stream processing network. A technique similar to timestamp ordering of database transactions is used to solve the problems. A solution is presented as a set of rules that govern processing of groups of data items. A proof of correctness has been provided for the strategy used to solve the problems.

Acknowledgments

I would like to thank my supervisor Dr. Janusz R. Getta for his invaluable guidance, support, and patience. He helped me in defining a suitable problem for my thesis, and always provided productive suggestions in the course of the writing process.

My special thanks to Professor John Fulcher and Dr. Heather Jamieson for their encouragement during the revision process.

I would like to thank my family and friends for their continuous support and help in the best way possible.