

Data Transfer and Sharing within Web Service Workflows

Donglai Zhang

September 2013

A dissertation submitted to the School of Computer Science of
The University of Adelaide for the degree of Doctor of Philosophy

Supervisors:

Dr. Andrew L. Wendelborn

Dr. Paul D. Coddington

Contents

Abstract	viii
Statement of Originality	x
Acknowledgement	xi
1 Introduction	1
1.1 Web Service Workflow for Remote Cooperation	3
1.1.1 Wide Area Network and Local Area Network	4
1.1.2 Network Connections for Research Work	5
1.2 Distributed Data Transfer in Workflow	6
1.2.1 Data Transfer Requirements	6
1.2.2 Data Transfer in a Workflow	8
1.3 Distributed Data Transfer and SOAP	9
1.3.1 Distributed Data Transfer	9
1.3.2 Data Transfer Performance Improvement with SOAP	10
1.4 Web Service Data Transfer with Attachment	13
1.4.1 SOAP with Attachment	13
1.4.2 Web Service Attachment Performance	14
1.5 Data Sharing in Web Service Workflow	15
1.5.1 Workflow Classification	15
1.5.2 Workflow Data Sharing	16
1.5.3 Web Service Workflow Data Sharing Review	18
1.5.4 Web Service Data Forwarding (WSDF) Framework	20
1.5.5 WSDF Implementation	20

1.5.6	WSDL Performance Testing	21
1.5.7	WSDL in the Cloud	22
1.6	Thesis Content Outline	23
2	Research Background	24
2.1	Introduction	24
2.2	Distributed Environment Networks	25
2.3	Distributed Resource Sharing	26
2.3.1	Information Sharing	26
2.3.2	Computational Resource Sharing in Distributed Environment	26
2.4	Distributed Computing Infrastructures	28
2.4.1	Grid Infrastructure	29
2.4.2	Cloud Infrastructure	31
2.5	Service Oriented Architecture and Web Service	33
2.5.1	Service Oriented Architecture (SOA)	33
2.5.2	Web Services	34
2.5.2.1	Web Service Architecture	35
2.5.2.2	Web Service Description Language (WSDL)	35
2.5.2.3	Single Object Access Protocol (SOAP)	36
2.5.3	RESTful Approach of Web service	37
2.5.4	Stateful Web Service and WSRF	38
2.6	Distributed Resources Collaboration	39
3	E-Science and Web Service Workflow	40
3.1	E-Science	41
3.1.1	E-Science Projects	41
3.1.2	Instruments in e-Science	44
3.2	Workflow	45
3.2.1	Workflow Definition	46
3.2.2	Workflow System	46
3.3	Data Transfer and Sharing within Workflow	50

3.3.1	Data Transfer with e-Science workflow	50
3.3.2	Web Service Data Transfer	50
3.3.3	Web Service Data Sharing	51
4	Web service Data Transfer with SOAP	53
4.1	Introduction	53
4.2	Research Background and Motivation	54
4.2.1	Synchrotron Project and CIMA	55
4.2.2	Motivation for the Research Work	56
4.3	Web Service Data Transfer with SOAP Message	57
4.3.1	Data Transfer Performance Analysis	58
4.4	SOAP Message Data Transfer Improvements	59
4.4.1	TCP Tuning	59
4.4.2	Push/Pull Model	61
4.4.3	Selection of Communication Protocols	62
4.4.4	Sending Data as Attachment	62
4.4.5	Multiple Data Transfer Connections	63
4.5	Improving Data Transfer in an e-Science Context	63
4.5.1	Experimental Framework	64
4.5.2	TCP Tuning	65
4.5.3	Combination of Pull Model with HTTP	67
4.5.4	Pull Model with Concurrent HTTP Clients	71
4.6	Conclusions	73
5	Web Services Data Transfer with Attachment	76
5.1	Introduction	77
5.2	Web Service with Attachment (WS-Att)	78
5.3	GridFTP	80
5.4	Experiment Environment for WS-Att	80
5.4.1	Experiment Location	80
5.4.2	Multiple Threads WS-Att	81

5.4.3	Facilities	81
5.4.4	File Sizes and APIs	82
5.4.5	GridFTP Experiment Setting	82
5.5	Experiment Result and Analysis for WS-Att vs. GridFTP	83
5.6	Conclusion	90
6	WSDF Definition	91
6.1	Introduction	91
6.2	Web Service Data Forwarding Framework	95
6.2.1	Stateful Workflow	95
6.2.2	Resource Forwarding Information	97
6.2.3	Successor Service	98
6.2.4	WSDF Architecture	99
7	WSDF Implementation	101
7.1	WSDF Server Implementation	102
7.1.1	WSDF as a SOAP engine	103
7.1.2	Attachment Support for Resources	104
7.1.3	Procedures in WSDF engine	105
7.2	WSDF Service	107
7.2.1	Extra Operations	107
7.2.2	Endpoint Reference (EPR)	108
7.3	WSDF Client	109
7.4	Building WSDF workflow with WSDF Framework	110
7.4.1	WSDF Service Operations	110
7.4.2	Workflow Processing Steps	110
8	WSDF Testing	116
8.1	Testing Methodology	117
8.1.1	Basic Service Time Consumption (BST)	117

8.1.2	Distributed Environment for Testing	118
8.1.2.1	WAN environment	118
8.1.2.2	Cloud Environment	119
8.1.3	Latency and Bandwidth Settings for Workflow Environment . .	120
8.1.4	Services	121
8.1.5	Data Size	121
8.2	Experiment Environment	122
8.2.1	Experiments in Emulated Distributed Environment	122
8.2.2	Experiments in Cloud Environment	123
8.3	Theoretical Analysis	123
8.3.1	Theoretical Data Transfer Time Analysis	123
8.4	Results Analysis	127
8.4.1	Data Transfer vs. Resource Management	131
8.4.2	Impact of Data Size	131
8.4.3	Network Connection Between Services and Client	135
8.4.3.1	Number of Services	136
8.4.4	Comparison with Theoretical Results	137
9	WSDF Testing in the Cloud	139
9.1	Introduction	139
9.2	Testing Methodology	141
9.2.1	Cloud Provider	141
9.2.2	Number of Services and Data Size	143
9.3	Experiment Environment	143
9.4	Cloud Experiment Result and Analysis	145
9.4.1	Total Time Consumed	145
9.4.2	Data Size	146
9.4.3	WSDF Performance Improvement Comparison	152
10	Conclusion and Future Work	154
10.1	Data Transfer	154

10.2 Data Forwarding	157
10.2.1 Utilization of Web Service	157
10.2.2 A Generalized Approach	159
10.2.3 WSDF with Decentralized Workflows	160
10.2.4 WSDF and Workflow Execution	161
10.3 Future Work	161
Bibliography	163
Appendix A: Operation Signatures	176

Abstract

With the development of distributed systems, it is more and more common for users to harness different resources to implement a larger task to meet their requirements.

Among the different approaches to distributed resource coordination, workflows based on Service Oriented Architecture (SOA) is an important case, as SOA provides a framework that is designed for loosely coupled applications. This thesis introduces the research work that we have carried out in distributed computing environments to improve the performance of data transfer and sharing in a web service workflow.

In a distributed environment, we explore how to improve the efficiency of data moving between services in a web service workflow. Data movement in a web service workflow can be categorized into two classes: data transfer between any two nodes in the workflow and the intermediate data sharing between different web services. We initially explored ways to improve the data transfer performance between two web service nodes, then improved the data sharing performance via study of the data sharing relationship between applications composed in a workflow.

We carried out the initial part of the research work based on the CIMA (Common Instrument Middleware Architecture) web service interface, which has been used by multiple academic organizations as an interface to distributed scientific instruments and applications. With the related experiments, we explore how the data generated by instruments can be transferred efficiently between different web service nodes. In the rest of the research, we study the data sharing relationships between different web service applications. By proposing the Web Service Data Forwarding (WSDF) framework, we allow intermediate data to be forwarded directly from the data generator to its consumer without going via a third party (the workflow engine).

We have implemented prototype systems for our proposed ideas. We also tested these systems in different environments to demonstrate the performance improvement that is expected from the WSDF approach.

Thesis Declaration

I, Donglai Zhang certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Donglai Zhang

Acknowledgments

Firstly, I would like to thank my wife and my family to support me with my study. Without your help, I will not have a chance to even start this work.

Secondly, great appreciation to my supervisors, Dr. Andrew Wendelborn and Dr. Paul Coddington, for their invaluable guidance and assistance with my study. Both of them are great supervisors. They are very experienced within the area of this research work and provided significant amount of guidance and feedback that was critical to the project. With their help, I not only gained research experiences, but also learned the fundamental methodologies for future research work as a researcher.

Third, I would thank my friends: Kewen, Paul. M, Peter, Wei, and Yidong — thanks for your help during my study.