

Reinhard Brandl

Cost Accounting for Shared IT Infrastructures

GABLER EDITION WISSENSCHAFT

Reinhard Brandl

Cost Accounting for Shared IT Infrastructures

With a foreword by Prof. Dr. Martin Bichler

GABLER EDITION WISSENSCHAFT

Bibliographic information published by Die Deutsche Nationalbibliothek
Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at <<http://dnb.d-nb.de>>.

Dissertation Technische Universität München, 2007

1st Edition 2008

All rights reserved

© Betriebswirtschaftlicher Verlag Dr. Th. Gabler | GWV Fachverlage GmbH, Wiesbaden 2008

Editorial Office: Frauke Schindler / Anita Wilke

Gabler-Verlag is a company of Springer Science+Business Media.
www.gabler.de



No part of this publication may be reproduced, stored in a retrieval system or transmitted, mechanical, photocopying or otherwise without prior permission of the copyright holder.

Registered and/or industrial names, trade names, trade descriptions etc. cited in this publication are part of the law for trade-mark protection and may not be used free in any form or by any means even if this is not specifically marked.

Cover design: Regine Zimmer, Dipl.-Designerin, Frankfurt/Main
Printed on acid-free paper
Printed in Germany

ISBN 978-3-8349-0897-1

Für meine Eltern

Foreword

During the past few years, determining the “value of IT” has ranked high on the agenda of IT managers and Chief Information Officers (CIOs). The rather broad and abstract topic has been intensively discussed in the Information Systems literature for many years. It turns into a very tangible problem in the field of IT cost accounting. Nowadays, corporate information systems are distributed systems. A detailed measurement of resource demands of IT services on a distributed IT-infrastructure and respective accounting and cost allocation turns out to be very expensive and impractical in most cases. The large proportion of indirect costs and the difficulty of finding adequate allocation rates are a significant problem in practice, regularly leading to free-rider problems. This problem has largely been ignored in the academic literature so far. Dr. Brandl proposes a method to derive estimators for the resource demand of service requests in a distributed IT infrastructure. This estimator is based on a set of load tests and respective measurements as they are often performed during the deployment phase of new information systems. Cost allocation keys can now be determined based on the number of service invocations per user or per department and the respective estimators.

While these measurements provide a lean method for the determination of usage-based cost allocation keys, it is not obvious that the estimators have sufficient accuracy, in particular concerning different types of services and volatile workloads. Although it is not possible to provide tangible results for all types of information systems, Mr. Brandl performs a large number of experiments for typical multi-tiered information systems as they are in widespread use today. Queuing network models are used to validate the results for different workloads and multiple types of services. The experiments show that resource consumption in heterogeneous environments can be predicted with high accuracy. Overall, this leads to a viable solution for the cost accounting of distributed information systems.

Dr. Brandl makes an important contribution to a largely neglected field. The book provides practical advice for IT managers for a very timely topic. I therefore hope and expect that the book will be well received not only among academics, but in particular among practitioners in cost accounting and IT controlling.

Prof. Dr. Martin Bichler

Preface

The provision of central IT infrastructure components, such as servers, storage, and networking equipment, accounts for a considerable proportion of the IT budgets of larger organizations. Typically, such components are shared among multiple applications and internal customers. Objective measurements of their respective resource consumption are technically difficult and incur high costs. In practice, infrastructure cost allocation is regularly based on simplified allocation keys which cause multiple free-rider problems and discontent among the stakeholders.

This thesis proposes a method to estimate the expected resource consumption of customer-oriented services across the components involved. The estimates are determined in a load test prior to the roll-out of an application system and then combined to so-called resource profiles. By means of these resource profiles, costs can be allocated to services or service invocations. During regular operations, consumption measurements at the different components can be omitted. The estimates therefore need to be unbiased even in cases of varying system workloads and in heterogeneous environments. Furthermore, they should support IT Capacity Planning and bridge the gap between business forecasts and IT resource planning.

The concept was implemented in a software toolkit and evaluated in a set of experiments with multi-tier database applications in a data center of the BMW Group. Queuing Network Models were used to validate the resource profiles under different system workloads. In the experiments, a surprisingly high accuracy of consumption estimates as well as of Queuing Network Model predictions could be determined. Besides the experimental validation, it was analyzed how the approach could be integrated into existing IT processes at the BMW Group.

The work presented in this thesis would not have been possible without the great support of two persons: Prof. Dr. Martin Bichler and Dr. Michael Ströbel. They gave me the

opportunity to work on this topic and provided continuous feedback, inspiration, and encouragement. I am deeply grateful to both of them.

During my time as a doctoral candidate, I was employed by the BMW Group. I greatly appreciated the pleasant working atmosphere and the kind support from all my colleagues. In particular, I would like to thank my managers Harald Raufer, Alexander Pauli, and Bernhard Huber, who gave me the freedom to pursue my research activities and provided me with interesting and responsible tasks to collect valuable business experiences far beyond the dissertation project. At the BMW Group, I was given access to all the relevant information and data center resources. However, without the technical expertise of Volker Smuda and Alexander Pochivalow, it would not have been possible to conduct more than 500 load tests during the development and evaluation phase of the software toolkit. Furthermore, I am particularly thankful to my colleagues Dr. Markus Greunz and Hartmut Liefke for their numerous comments on the viability of my approach.

I would also like to thank Prof. Dr. Helmut Krcmar who readily accepted to act as the second reviewer of the thesis and provided me with valuable advice regarding future steps. Finally, I am deeply indebted to my family for enabling me to go this way and providing continuous support, patience, and understanding throughout my whole life.

Dr. Reinhard Brandl

Contents

List of Figures	XV
List of Tables	XIX
1 Introduction	1
1.1 Problem Statement	1
1.2 Research Approach	5
1.3 Overview of the Thesis	7
2 IT Infrastructure Cost Allocation	11
2.1 IT Cost Accounting and Chargeback	11
2.1.1 Classification	11
2.1.2 Definition of Terms	14
2.1.3 Overview of Processes	16
2.1.4 Cost Allocation for IT Services	18
2.1.5 IT Infrastructure Services	21
2.2 Survey of Literature	23
2.2.1 Scientific Work	23
2.2.1.1 Process and Activity-based Costing	23
2.2.1.2 IT Controlling of Distributed Systems	26
2.2.1.3 Organizational Objectives and Chargeback Systems	27
2.2.2 Empirical Surveys	28
2.2.3 IT Service Management Reference Models	30
2.2.4 Commercial Research	33
2.3 Requirements and Objectives	36
2.4 Practiced Approaches	39
2.5 Discussion	43

3	Cost Allocation based on Resource Profiles	47
3.1	Scope	47
3.2	Concept	49
3.2.1	IT Services from a Customer Perspective	49
3.2.2	Determining Resource Profiles for Services	51
3.2.3	Resources in a Client/Server Environment	52
3.2.4	Cost Allocation by Services and Resource Profiles	53
3.3	Model Hypotheses	55
3.4	Requirements on Resource Profiles	57
3.5	Software Support	58
3.5.1	Tool Categories	58
3.5.2	Profiling Process	60
3.5.3	Software Overview	63
3.5.3.1	Load Test Tool	63
3.5.3.2	Performance Monitors	63
3.5.3.3	Service Profiler	64
3.5.4	Mercury LoadRunner	66
3.5.4.1	LoadRunner Virtual User Generator	66
3.5.4.2	LoadRunner Controller	69
3.5.4.3	LoadRunner Analysis	71
3.5.5	Resource Profile Determination	72
3.6	Overview of Validation Approaches	78
3.6.1	Experiments: Resource Profiles	79
3.6.2	Experiments: Analytical Models	80
3.6.3	Proof of Concept: BMW Group	81
3.6.4	Summary	81
3.7	Related Work	83
3.7.1	Cost Allocation by Services	83
3.7.2	Determination of Resource Profiles	84
4	Experiments: Resource profiles	85
4.1	Experimental Setup	85
4.1.1	Test Infrastructure	85
4.1.2	Example Applications	87

4.1.2.1	PASTA	87
4.1.2.2	Java Pet Store	90
4.2	Overview of Experiments	93
4.3	Experimental Results	96
4.3.1	Background Activities	96
4.3.2	Resource Profiles	99
4.3.3	Repeatability	101
4.3.4	Load-dependent Behavior	104
4.3.5	Linear Regression	108
4.3.6	Linear Regression (cross check)	110
4.3.7	Database Size	111
4.4	Summary	113
5	Experiments: Analytical Models	115
5.1	Motivation	115
5.2	IT Capacity Planning Methods	116
5.3	Queueing Network Theory	119
5.3.1	Elementary Queueing Systems	120
5.3.2	Workload Parameter	122
5.3.3	Queueing Network Models	124
5.3.4	Analysis Techniques	125
5.4	Software Support	129
5.4.1	Software Overview	129
5.4.2	Analysis of Queueing Network Models	130
5.4.3	Validation of Queueing Network Models	131
5.4.4	Interplay of Software Tools	135
5.5	Experimental Setup	136
5.5.1	Overview of Experiments	136
5.5.2	Queueing Network Model	136
5.6	Experimental Results	138
5.6.1	Services including single Client Requests	138
5.6.2	Services including multiple Client Requests	141
5.7	Related Experimental Results	144
5.8	Summary	149

6	Proof of Concept: BMW Group	151
6.1	Motivation	151
6.2	Organizational Context	152
6.2.1	Overview of IT Organization	152
6.2.2	IT Service Management Processes	154
6.2.3	IT Cost Accounting and Chargeback	155
6.2.4	Analysis of the current Situation	158
6.2.5	Customer-oriented Services and Resource Profiles	159
6.3	Feasibility Study: Java/J2EE Application Systems	160
6.3.1	Java/J2EE at the BMW Group	161
6.3.2	Integration into existing Processes	163
6.3.3	Identification of Services	166
6.3.4	Summary of Experiences	167
6.3.5	Other Application Platform Technologies	169
6.4	Summary	171
7	Conclusions	173
7.1	Summary of Results	173
7.2	Outlook	175
	Bibliography	179
A	Appendix	195

List of Figures

1.1	Composition of an average IT Budget	2
1.2	Example: Shared and dedicated Data Center Resources	3
1.3	Structure of the Thesis	7
2.1	Model of Information Management	13
2.2	Example: Organization of IT Service Provision	15
2.3	IT Cost Accounting and Chargeback Processes	16
2.4	Example: Overview of IT Functions	19
2.5	Total Costs of Ownership of an IT System	20
2.6	Information and Communication Technology	22
2.7	Model of IT Chargeback	28
2.8	IT Cost Allocation in Germany	29
2.9	ITIL Publication Framework	31
2.10	Strengths and Weaknesses of Cost Allocation Methods	35
2.11	Stakeholders and Objectives of IT Cost Accounting	37
2.12	Consistent and conflicting Objectives	38
2.13	Classification of practiced Approaches	40
3.1	Classification of Application Systems and Operation Modes	48
3.2	Customer Perspective on Information and Communication Technology	50
3.3	Meta-Model of IT Services	51
3.4	Cost Allocation by Services and Resource Profiles	54
3.5	Profiling Process	61
3.6	Overview of the Service Profiler	65
3.7	Screenshot: LoadRunner Virtual User Generator	67
3.8	Screenshot: LoadRunner Controller	70

3.9 Screenshot: LoadRunner Analysis	71
3.10 Example: Services including single Client Requests	75
3.11 Example: Services including multiple Client Requests	76
3.12 Example: Linear Regression	77
3.13 Example: Consumption Estimates per Service	78
4.1 Overview of the Test Infrastructure	86
4.2 Screenshot: PASTA	88
4.3 Overview of PASTA Use Cases	89
4.4 Screenshot: Java Pet Store	90
4.5 Overview of Java Pet Store Use Cases	91
4.6 Background Processor Utilization at the Database Server	97
4.7 Truncated Means (10%) of Background Processor Utilizations	98
4.8 Resource Profile: PASTA – multiple Client Requests	100
4.9 Resource Profile: Java Pet Store – multiple Client Requests	101
4.10 Repeatability: PASTA – multiple Client Requests	102
4.11 Repeatability: PASTA – single Client Requests	103
4.12 Repeatability: Java Pet Store – multiple Client Requests	103
4.13 Repeatability: Java Pet Store – single Client Requests	104
4.14 Load-dependent Behavior: PASTA – multiple Client Requests	105
4.15 Load-dependent Behavior: PASTA – single Client Requests	106
4.16 Load-dependent Behavior: Java Pet Store – mult. Client Requests	106
4.17 Load-dependent Behavior: Java Pet Store – single Client Requests	107
4.18 Linear Regression: PASTA – multiple Client Requests	109
4.19 Linear Regression: Java Pet Store – multiple Client Requests	109
4.20 Linear Regression (cross check): PASTA – mult. Client Requests	110
4.21 Database Size: Java Pet Store – multiple Client Requests	112
5.1 Overview of IT Capacity Planning Methods	117
5.2 Elementary Queueing System with a single Service Station	119
5.3 Classification and Notation of Elementary Queueing Systems	120
5.4 Open and closed Queueing Network Models	124
5.5 Classification of Analysis Techniques for QN Models	128
5.6 Overview of Software Tools for QN Analysis and Validation	129
5.7 Example: Load Test for QN Model Validation	133

5.8	Example: Mean Processor Utilization during Intervals with constant Numbers of Users	134
5.9	Interplay of the different Software Tools	135
5.10	QN Model of the Test Infrastructure	138
5.11	QN Model Validation: Application Server Processor Utilization – single Client Requests	140
5.12	QN Model Validation: Database Server Processor Utilization – single Client Requests	140
5.13	QN Model Validation: Web Server Processor Utilization – single Client Requests	141
5.14	QN Model Validation: Response Time – single Client Requests	142
5.15	QN Model Validation: Throughput – single Client Requests	142
5.16	QN Model Validation: Application Server Processor Utilization – multiple Client Requests	144
5.17	QN Model Validation: Database Server Processor Utilization – multiple Client Requests	145
5.18	QN Model Validation: Web Server Processor Utilization – multiple Client Requests	145
5.19	QN Model Validation: Throughput – multiple Client Requests	146
6.1	Organizational Structure of the BMW Group IT	153
6.2	Tasks of Central IT, Departmental IT and Business Units	154
6.3	IT Service Management Process Map	155
6.4	IT Chargeback and Budget of BMW Group Central IT	156
6.5	Master Solution Guideline J2EE/EJB	161
6.6	Master Solution Guideline J2EE/Servlet	162
6.7	Integration into the Approval Process	164
6.8	Integration into the IT Service Management Process Map	166
A.1	Background Resource Consumption at the SAN	203
A.2	Background Resource Consumption at the Network	203

List of Tables

2.1	Benefits and Problems of IT Cost Accounting and Chargeback	18
2.2	Overview of Cost Allocation Methods	34
3.1	Rules for Raw Data Normalization	72
3.2	Comparison of Method and Software Toolkit with Nagaprabhanjan and Apte (2005)	84
4.1	Overview of Experiments on Resource Profiles	95
4.2	Experiments on Background Activities: Configuration of Performance Monitors	96
4.3	Average Client Think Times and Load Test Runtimes	108
5.1	Experimental Results of Kounev and Buchmann (2003)	148
6.1	Example: Intervals of expected Resource Consumption	169
A.1	Overview of the Test Infrastructure	196
A.2	Resource Profile: PASTA – multiple Client Requests	197
A.3	Resource Profile: PASTA – single Client Requests	199
A.4	Resource Profile: Java Pet Store – multiple Client Requests	200
A.5	Resource Profile: Java Pet Store – single Client Requests	202
A.6	Repeatability: PASTA – multiple Client Requests	204
A.7	Repeatability: PASTA – single Client Requests	204
A.8	Repeatability: Java Pet Store – multiple Client Requests	205
A.9	Repeatability: Java Pet Store – single Client Requests	205
A.10	Load-dependent Behavior: PASTA – multiple Client Requests	206
A.11	Load-dependent Behavior: PASTA – single Client Requests	206
A.12	Load-dependent Behavior: Java Pet Store – mult. Client Requests	206

A.13 Load-dependent Behavior: Java Pet Store – single Client Requests	207
A.14 Linear Regression: PASTA – multiple Client Requests	207
A.15 Linear Regression: Java Pet Store – multiple Client Requests	207
A.16 Linear Regression (cross check): PASTA – mult. Client Requests	208
A.17 Resource Profile: Java Pet Store – multiple Client Requests – 90% full Tablespace	209
A.18 QN Model Input Parameter: Processors – single Client Requests	210
A.19 QN Model Validation: Processor Utilization – single Client Req.	211
A.20 QN Model Input Parameter: Processors - mult. Client Requests	212
A.21 QN Model Validation: Processor Utilization – mult. Client Req.	213