Specification and Analytical Evaluation of Heterogeneous Dynamic Quorum-Based Data Replication Schemes

Christian Storm

Specification and Analytical Evaluation of Heterogeneous Dynamic Quorum-Based Data Replication Schemes

Foreword by Prof. Dr.-Ing. Oliver Theel

Springer Vieweg

RESEARCH

Christian Storm
Oldenburg, Germany

Dissertation University of Oldenburg, 2011

ISBN 978-3-8348-2380-9 ISBN 978-3-8348-2381-6 (eBook)
DOI 10.1007/978-3-8348-2381-6

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Springer Vieweg

Cover design: KünkelLopka GmbH, Heidelberg

Printed on acid-free paper

Springer Vieweg is a brand of Springer DE. Springer DE is part of Springer Science+Business Media.
www.springer-vieweg.de

Foreword

Our modern life depends more and more on the correct and continuous functioning of computer systems. This fact bears benefits and risks. Traveling from Hanover to Frankfurt am Main, both in Germany, by train in just a little more than two hours can only be achieved by the cooperation and correct functioning of many computer systems embedded in a high speed train. Parts of them form, for example, a sophisticated brake system or a communication system meant for issuing warnings to the train driver enabling him or her to stop the train in time prior to a potentially deadly collision.

Unfortunately, correct functioning of all these components at all times is impossible: in the absence of perfect components all systems are built out of components that are subject to fail to some extent at some time. Thus, there is always a non-zero probability that a system, such as a high speed train's braking system, fails at a time when its correct functioning is dearly needed. But fortunately, the probability that a system does not work at a particular point in time can be controlled by means of redundancy: as a rule-of-thumb, the more redundancy spent for a system, the higher is its availability and the higher are the costs for building and operating it. Thus, a reasonable trade-off must be found.

In his research work, Christian Storm presents a new universal framework for the specification and implementation of heterogeneous data replication strategies. Data replication is some form of redundant resources, and data replication can be exploited for implementing highly available services, such as information services for the train driver in the high speed train scenario described earlier. Data replication strategies correctly handle the redundant resources, called data replicas. A particular data replication strategy exhibits one particular trade-off between availability and costs. The framework presented allows to model and realize all known, relevant data replication strategies from literature but also goes way beyond. It does so by introducing and exploiting the powerful concept of tree-shaped voting structures. Based on this unifying modeling abstraction, Christian Storm presents a new and efficient analysis technique for heterogeneous, dynamic data replication strategies (that also covers all homogeneous and static cases) based on Petri-Net modeling, reachability and steady-state analyses. It allows to

efficiently analyze and customize data replication strategies helping to identify systems that exhibit the highest availability possible for a certain cost budget. Since there is no reason to strive for less, researchers and students should know about the underlying concepts and techniques. This book describes them in a very systematic fashion and is accompanied by carefully chosen examples and evaluations.

Oliver Theel

Acknowledgments

During the time of my doctorate, too many people to mention them individually have contributed in personal and scientific respects in one or another way. While I owe sincere thankfulness to all, I would like to name a few of them explicitly.

First and foremost, I would like to express my sincerest gratitude and special thanks to my advisor, Prof. Dr.-Ing. Oliver Theel, for his enduring encouragement, support, and his confidence in me. Not only has he given me the freedom to explore and pursue my own ideas but he also provided a pleasant and friendly place to work in. His huge repertoire of anecdotes and stories will be remembered.

Furthermore, I would like to thank my second supervisor Prof. Dr. Wilhelm Hasselbring for refereeing this dissertation.

I am also grateful to the members of the System Software and Distributed Systems Group as well as to the members of the TrustSoft graduate school at the University of Oldenburg. I appreciate the pleasant and productive atmosphere as well as the constructive discussions that developed on various occasions. In particular, I would like to acknowledge Timo Warns, Kinga Kiss-Iakab, Jens Happe, Roland Meyer, Henrik Lipskoch, Heiko Koziolek, and Eike Möhlmann for working together, sharing an office, co-organizing workshops, friendship, and simply having a great time.

Finally, and most importantly, I would like to thank my family for their absolute support and confidence in me throughout my life.

Christian Storm

Abstract

Data replication by employing quorum systems is a well-established concept to improve operation availability on critical data objects in distributed systems that have strong consistency demands. It is therefore an important base concept for constructing dependable distributed systems. Modern distributed systems have become dynamic in nature with processes arriving and deliberately departing at run-time. Traditional data replication schemes are either static, that is, they use a fixed quorum system and cannot adapt to varying numbers of processes in the system, or their dynamics in adapting the quorum system is usually constrained by an upper bound on the number of processes that is predetermined at design-time. These dynamic data replication schemes are homogeneous in the sense that for each set of processes, the quorum system is constructed using the same scheme-inherent quorum system construction strategy. Like there is no single data replication scheme superior for every application scenario, there is also none superior for every set of processes. Motivated by this fact, heterogeneous dynamic data replication schemes are free to use a particular quorum system construction strategy per set of processes.

The first contribution of the thesis is a uniform data replication scheme specification method that combines the potential of heterogeneous dynamic data replication schemes with an unbounded flexibility in the number of processes at run-time. This method provides advanced means to design specifically tailored data replication schemes that can utilize the respective best-option selection of quorum system construction strategies for a specific application scenario.

The choice of a data replication scheme in the design space spawned by static and dynamic, unstructured and structured, and homogeneous and heterogeneous data replication schemes has a strong impact on the performance and dependability of a system and therefore needs a careful evaluation. In light of constantly evolving modern distributed systems, this choice cannot be definitely made at design-time but has to be repeatedly revised and adapted to a changing environment at run-time. For this purpose, evaluation methods based on simulation are inadequate because of their massive time complexity or their approximate nature of results un-

der time constraints. Contrarily, evaluation methods based on analysis are fast and accurate but require a careful crafting of the system model for it to be tractable and to provide meaningful results. To date, the analytical evaluation of dynamic data replication schemes is limited to a subset of the specific subclass of unstructured homogeneous dynamic data replication schemes. The existing approaches are customized to one single data replication scheme and are therefore inapplicable to the evaluation of other schemes.

The second contribution of the thesis is a general and comprehensive approach to the analytical evaluation of data replication schemes that supports unstructured as well as structured homogeneous and moreover heterogeneous dynamic data replication schemes, with static ones being a simple special case. Different data replication schemes are reflected in the system model by merely varying the data replication scheme specification. Furthermore, the system model allows quality measures besides operation availability, such as operation costs, to be evaluated for the write operation as well as for the read operation.

Zusammenfassung

Datenreplikation unter Verwendung von Quoren-Systemen ist ein weit verbreitetes Konzept, um die Operationsverfügbarkeit auf kritischen Datenobjekten in verteilten Systemen mit hohen Anforderungen an die Datenkonsistenz zu erhöhen. Es ist daher ein wichtiges Basiskonzept zur Konstruktion zuverlässiger verteilter Systeme. Moderne verteilte Systeme sind naturgemäß dynamisch in ihrer Struktur, da Prozesse das System zur Laufzeit verlassen und neue hinzukommen können. Bisherige Replikationsverfahren sind allerdings entweder statisch, d.h., sie benutzen ein feststehendes und nicht an variierende Prozessanzahlen anpassbares Quoren-System, oder ihre Dynamik in der Anpassung des Quoren-Systems an variierende Prozessanzahlen ist meist durch eine zur Entwicklungszeit festgelegte obere Grenze beschränkt. Diese dynamischen Replikationsverfahren sind homogen in dem Sinne, dass zur Konstruktion des Quoren-Systems die gleiche verfahrensspezifische Quoren-Konstruktionsvorschrift für jede Prozessanzahl benutzt wird. Ebenso wie es kein einzelnes Replikationsverfahren gibt, das für alle Einsatzszenarien optimal ist, gibt es auch keines, das für alle Prozessanzahlen optimal ist. Daher erlauben heterogen-dynamische Replikationsverfahren die Benutzung verschiedener Quoren-Konstruktionsvorschriften für verschiedene Prozessanzahlen.

Der erste Beitrag der Dissertation ist eine uniforme Spezifikationsmethode für quoren-basierte Replikationsverfahren, die das Potenzial heterogen-dynamischer Replikationsverfahren mit einer unbeschränkten Flexibilität in der Anzahl der Prozesse zur Laufzeit kombiniert. Dadurch bietet diese Methode die Möglichkeit, spezifisch angepasste Replikationsverfahren für ein bestimmtes Einsatzszenario unter Verwendung der jeweils besten Kombination von Quoren-Konstruktionsvorschriften zu entwickeln.

Die Auswahl eines Replikationsverfahrens im Entwurfsraum, der durch statische und dynamische, unstrukturierte und strukturierte sowie durch homogene und heterogene Replikationsverfahren aufgespannt wird, ist eine wichtige Entscheidung im Hinblick auf die Performanz und die Zuverlässigkeit eines Systems, weshalb sie sorgfältig evaluiert werden muss. Angesichts sich ständig wandelnder moderner verteilter Systeme kann diese Entscheidung nicht abschließend zur Entwurfszeit getroffen werden, sondern

muss wiederholt zur Laufzeit überprüft und auf ein sich änderndes Einsatzszenario angepasst werden. Evaluationsmethoden, die auf Simulation basieren, sind aufgrund ihrer großen Zeitkomplexität bzw. ihrer approximativen Ergebnisse unter Laufzeitbeschränkung dafür unzureichend. Im Gegensatz dazu sind Evaluationsmethoden, die auf Analyse basieren, schnell und präzise. Allerdings benötigen diese ein sorgfältig erstelltes Systemmodell, das handhabbar ist und sinnvolle Ergebnisse liefert. Bisher ist die analytische Evaluation dynamischer Replikationsverfahren beschränkt auf eine bestimmte Untermenge unstrukturierter homogen-dynamischer Verfahren. Existierende Ansätze sind speziell auf ein Replikationsverfahren ausgerichtet und daher nicht anwendbar auf andere Replikationsverfahren.

Der zweite Beitrag der Dissertation ist ein allgemeiner und umfassender Ansatz zur analytischen Evaluation von Replikationsverfahren, der sowohl statische als auch unstrukturierte und strukturierte homogen-dynamische und darüber hinaus heterogen-dynamische Replikationsverfahren unterstützt. Verschiedene Replikationsverfahren werden im Systemmodell abgebildet, indem lediglich ihre Spezifikation ausgetauscht wird. Das Systemmodell erlaubt die Evaluation weiterer Qualitätsmaße neben der Operationsverfügbarkeit, wie z.B. der Operationskosten, sowohl für die Schreiboperation als auch für die Leseoperation.

Contents