Cloud Data Management

Liang Zhao • Sherif Sakr • Anna Liu
Athman Bouguettaya

Cloud Data Management

Foreword by Albert Y. Zomaya

Springer

Liang Zhao
NICTA Kensington
NSW, Australia

Anna Liu
NICTA, Eveleigh
NSW, Australia

Sherif Sakr
Software Systems Research Group
NICTA, Eveleigh, NSW
Australia

Faculty of Computers and Information
Cairo University, Egypt

Athman Bouguettaya
School of Computer Science
and Information Technology
RMIT University, Melbourne
VIC, Australia

ISBN 978-3-319-04764-5 ISBN 978-3-319-04765-2 (eBook)
DOI 10.1007/978-3-319-04765-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014932980

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my parents, Dianchi Zhao and Lijuan Chu, and my wife, Huan Wang, for their love, understanding, and encouragement.

Liang Zhao

To my wife, Radwa, and my lovely daughter, Jana, for their love, encouragement, and support.

Sherif Sakr

Lifelong gratitude to my family.

Anna Liu

To my wife, Malika, and lovely sons, Zakaria, Ayoub, and Mohamed-Islam.

Athman Bouguettaya

Foreword

The rapidly expanding generation of Internet-based services such as e-mail, blogging, social networking, search, and e-commerce has substantially redefined the behavior and trends of web users when it comes to creating, communicating, accessing content, sharing information, and purchasing products. Information technology professionals are witnessing a proliferation in the scale of the data generated and consumed because of the growth in the number of these systems; this ever increasing need for scalability and new application requirements has created new challenges for traditional relational database management systems (RDBMS). Currently, the apparent goal of the system and tool manufacturers is to facilitate the job of implementing every application as a distributed, scalable, and widely accessible service on the web (e.g., services from Facebook, Flickr, YouTube, Zoho, and LinkedIn).

Cloud computing technology is a relatively new model for hosting software applications. The cloud model simplifies the time-consuming processes of hardware provisioning, hardware purchasing, and software deployment; therefore it revolutionizes the way computational resources and services are commercialized and delivered to customers. In particular, it shifts the location of this infrastructure to the network in order to reduce the costs associated with the management of hardware and software resources. This means that the cloud represents the long-held dream of envisioning computing as a utility, a dream in which the economy of scale principles help to effectively drive down the cost of the computing infrastructure. In practice, cloud computing promises a number of advantages for the deployment of software applications such as pay-per-use cost model, short time to market, and the perception of (virtually) unlimited resources and infinite scalability.

The rise of the cloud technology has been somewhat disruptive. The advantages of the cloud computing model open up new avenues for deploying novel applications that were not economically feasible in a traditional enterprise infrastructure setting. Therefore, the cloud has become an increasingly popular platform for hosting software applications in a variety of domains such as e-retail, finance, news, and social networking. The proliferation in the number of applications also delivers a tremendous increase in the scale of the data generated and consumed by

these applications. This is why a cloud-hosted database system powering these applications forms a critical component in the software stack of these applications.

To meet the challenges posed by hosting databases on cloud computing environments there are a plethora of systems and approaches. This book is the first that approaches the challenges associated with hosting databases on cloud computing environments from different but integrated perspectives; it connects the dots. The authors deal with the problems that may be encountered in every cloud-based data hosting solution: NoSQL storage services, database-as-a-service (DaaS), virtualized database servers in addition to batch-based processing systems for big data. The book is useful for many database researchers or practitioners because the inherent change in hosting database in cloud environment is fundamental on many perspectives as it originates from new foundations and models of thinking.

I found the book to contain a lot of timely and useful information. The book has many gems that inspire the readers as they go through the different chapters which are covering an area that is currently changing the data management field in a fundamental way. It covers an impressive array of topics with great clarity that will excite any reader wishing to understand this emerging technology. It also provides extensive references which will help the interested reader find out more information about the discussed topics. All in all, this is an impressive piece of work and an invaluable source of knowledge for advanced students and researchers working in or wishing to explore this exciting field.

Darlington, NSW, Australia Albert Y. Zomaya

Preface

Cloud computing technology represents a new paradigm for the provisioning of computing resources. Cloud computing is with us and for the foreseeable future. This paradigm shift allows for the outsourcing of computing resources to reduce the ownership costs associated with the management of hardware and software. Cloud computing simplifies the time-consuming processes of hardware provisioning, hardware purchasing, and software deployment.

Cloud computing is not a passing trend but a stubborn reality that is rooted on an emerging trend leading computing into a technological quantum leap. It builds on decades of research in virtualization, autonomic computing, grid computing, and utility computing, and ubiquity of the web as the network and delivery medium.

Central to the success of cloud computing is the ability to provision data using different quality of service requirements, including latency, performance, and reliability. Unfortunately, most cloud providers do not guarantee, and let alone, provide information about actual quality of service for data access. This is a complex exercise that depends on many factors, including the location of the data store, type of data, network congestion and data store platforms.

This book fills a gap in that it provides an in-depth analysis of major data cloud platforms using an exhaustive series of tests and experiments to unlock the unanswered questions surrounding the performance of each cloud data platform that is considered. The work presented in this book focuses on evaluating cloud databases in the presence of very little information from cloud providers. This can also be interpreted as reverse-engineering the performance of cloud databases with its own risks in interpretation.

The data cloud platforms considered in this book include the leaders in the field, including, Amazon, Microsoft, and Google. Amazon offers a collection of services, called Amazon Web Services, which includes Amazon Elastic Compute Cloud (EC2) as cloud hosting server, offering infrastructure as a service and Amazon SimpleDB and Simple Storage Service (S3) as cloud databases.

Microsoft Azure is recognized as a combination of infrastructure as a service and platform as a service. It features web role and worker role for web hosting tasks and computing tasks, respectively. It also offers a variety of database options including

Windows Azure Table Storage and Windows Azure Blob Storage as the NoSQL database options and Azure SQL Database as the relational database option.

Google App Engine supports a platform as a service model, supporting programming languages including Python and Java and Google App Engine Datastore as a Bigtable-based, non-relational, and highly sharable cloud database.

We propose a performance evaluation framework of cloud platforms as a uniform testing environment for all the cloud data environments. We describe novel frameworks and architectures to address the following issues: (1) the performance characteristics of different cloud platforms, including cloud hosting servers and cloud databases, (2) availability and reliability characteristics that cloud platforms typically exhibit, (3) type of faults and errors that may be encountered when services are running on different cloud platforms under high request volume or high stress situations, (4) reasons behind the faults and errors, (5) the architecture internal insights that may be deduced from these observations, and (6) the software engineering challenges that developers and architects could face when using cloud platforms as their production environment for service delivery.

Kensington, NSW, Australia — Liang Zhao
Eveleigh, NSW, Australia — Sherif Sakr
Eveleigh, NSW, Australia — Anna Liu
Melbourne, VIC, Australia — Athman Bouguettaya

Acknowledgements

I would like to thank my parents, Lijuan Chu and Dianchi Zhao, for their constant love, support, and encouragement during my Ph.D. study. I would also like to express my appreciation to my wife, Huan Wang, for her patience, understanding, and encouragement. The work would not have been possible without their support and care.

Liang Zhao

I would like to thank my parents, Ali Sakr and Amira Awad, for their encouragement and support. I want to thank my daughter, Jana, for the happiness and enjoyable moments she is always bringing to my life. My most special appreciation goes to my wife, Radwa Elshawi, for her everlasting support and deep love.

Sherif Sakr

I would like to sincerely thank the excellent hard work of my coauthors, without their dedication, commitment, and persistence, this book would not be here. I would also like to thank the wider NICTA Dependable Cloud Computing research team, who have provided wonderful companionship through our research journey. Lastly, many thanks go to our industry partners, who have provided much of the use inspiration for our work.

Anna Liu

I would like to thank my family for their love and understanding during my work on this book.

Athman Bouguettaya

The authors of this book would like to extend their sincere gratitude and appreciation to their collaborators for the contribution to this book; in particular, we would like to mention Alan Fekete, Jacky Keung, Kevin Lee, Hiroshi Wada, Xiwei Xu, Zhen Ye, Xiaofang Zhou, and Liming Zhu. Thank you all!

Contents

List of Figures

List of Tables