

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
Eveleigh, NSW, Australia
Eveleigh, NSW, Australia
Melbourne, VIC, Australia

Liang Zhao
Sherif Sakr
Anna Liu
Athman Bouguettaya

Cloud Data Management

Zhao, L.; Sakr, S.; Liu, A.; Bouguettaya, A.

2014, XIX, 202 p. 86 illus., 50 illus. in color., Hardcover

ISBN: 978-3-319-04764-5