Searchable Storage in Cloud Computing

Yu Hua · Xue Liu

Searchable Storage in Cloud Computing

Springer

Yu Hua
Huazhong University of Science and Technology
Wuhan, Hubei, China

Xue Liu
McGill University
Montreal, QC, Canada

ISBN 978-981-13-2720-9 ISBN 978-981-13-2721-6 (eBook)
https://doi.org/10.1007/978-981-13-2721-6

Library of Congress Control Number: 2019930156

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Exabyte- or Zettabyte-scale storage systems are one of the major challenges facing cloud computing because it is nontrivial to store and manage huge amount of data stemming from many data-intensive applications such as business transactions, scientific computing, social network webs, mobile applications, and information visualization, in which cloud computing serves as an important infrastructure. As a matter of fact, we are generating and storing much more data than ever before and this trend continues at an accelerated pace. What is more, the data in storage systems need to be searchable, which becomes an important function and the foundation of many operations in cloud computing. Therefore, researchers and engineers must find ways to quickly and accurately find the right information, query the data, extract the knowledge, and at the same time, reduce computational and transmission overheads.

Yu Hua and Xue Liu have been the leading contributors to searchable storage research. This book provides the backgrounds of searchable storage in cloud computing systems. It covers the basic concepts and implementation techniques such as hash functions, semantic namespace, exact and approximate queries, data analytics, multicore computation, and data cube in the networks in a chapter-by-chapter presentation. Furthermore, these topics are presented with examples and real-world applications, and they are readable and informative.

Researchers and engineers in the fields of cloud computing and storage systems will find it interesting to learn about the searchable techniques due to the ease of use and simplicity with the aid of many open-source codes. Readers will also obtain the needed backgrounds and gain new insights and implementation experiences in data structures and computer systems.

Personally, I found this book very informative and stimulating, and believe that readers will gain lots of new insights from this book.

Hong Kong
December 2018

John C. S. Lui

Preface

In the era of cloud computing and big data, storage systems are very important and have become the infrastructure of many real-world applications. However, existing storage systems fail to offer cost-efficient searching service due to overlooking the semantics behind massive data. This book introduces and presents new schemes for exploring and exploiting the searchable storage via cost-efficient semantic hashing computation. The contents in this book include basic hashing structures (Bloom Filters, Locality Sensitive Hashing, Cuckoo Hashing), semantic storage systems, and searchable namespace, which support multiple applications, such as cloud backups, exact and approximate queries, and image sharing. More importantly, all these mentioned structures and techniques have been really implemented to support real-world applications, some of which offer open-source codes for public use.

This book consists of eight chapters to present the core ideas and methodology for searchable storage in cloud computing.

Chapter 1 introduces the backgrounds of searchable storage, which requires multiple queries, including point query, range query, cover query, and bound query. The proposed Bloom filter based R-tree can efficiently support the above queries in a cost-efficient manner.

Chapter 2 describes the scalable metadata management in the context of cloud computing via efficient hash computation. A real system prototype, i.e., the Group-Hierarchical Bloom Filter Array, is used as a case in point to comprehensively demonstrate the significant performance improvements and scalability.

Chapter 3 presents an important component, i.e., the semantic-aware namespace, which serves as the infrastructure of large-scale cloud computing platforms. This namespace exploits the semantic correlation behind the multi-dimensional attributes of files to efficiently support not only namespace construction but also dynamic evolution.

Chapter 4 summarizes the exact-matching queries for metadata via efficient semantic grouping from the views of both users and systems. This scheme is able to provide fast and exact-matching query services. In the meantime, the consistency can be guaranteed by using versioning techniques.

Chapter 5 provides the approximate queries by using locality sensitive Bloom filters. Unlike conventional Bloom filters, the proposed space-efficient filters offer approximate query services with the constant-scale complexity.

Chapter 6 introduces the near real-time searchable analytics for images. This technique well explores and exploits the features of images to configure semantic-aware methodology. The analytics are accurate and in the meantime, the memory systems deliver high performance.

Chapter 7 shows computing infrastructure of searchable storage in cloud computing. By leveraging the data similarity, we can accurately place suitable data within the hierarchical caches and memory for multicore processors, thus obtaining significant performance improvements.

Chapter 8 presents semantic-aware data cube to provide fast online queries for the cloud. Correlation-based representation is used to support partial materialization and efficient incremental updates.

This book is for computer scientists, computer engineers, and others who are interested in the cloud computing from the perspective of searchable storage.

Acknowledgement

This book is supported in part by National Natural Science Foundation of China (NSFC) under Grant No. 61772212.

Wuhan, China — Yu Hua
Montreal, Canada — Xue Liu
November 2018

Contents