

Preface

The efficient management of data is a core asset of almost every organization. Data is collected, stored, manipulated, and analysed to drive the business and derive support for the decision making process. Establishing and running the data management platform within a larger organization is a complex, time- and budget-intensive tasks. Not only do data management systems have to be installed, deployed, and populated with different sets of data. The data management platform of an organization has to constantly maintained on a technical as well as on a content level. Changing applications and business requirements have to be reflected within the technical setup, new software versions have to be installed, hardware has to be exchanged etc. Although the benefit of an efficient data management platform can be enormous not only in terms of a direct controlling of the business or a competitive advantage but also in terms of strategic advantages, the technical challenges are enormous to keep a data management machinery running.

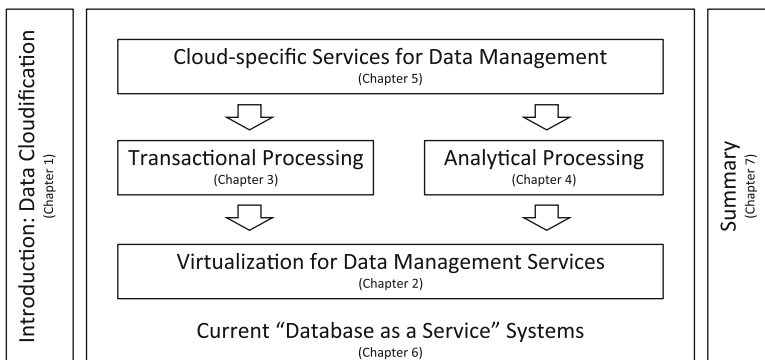
In the age of “Everything-as-a-Service” using cloud-based solutions, it is fair enough to question on-premise data management installations and consider “Software-as-a-Service” an alternative or complementary concept for system platforms to run (parts of) a larger data management solution. Within this book, we tackle this question to give answers to some of the huge number of different facets, which have to be considered in this context. We will therefore review the general concept of the “Everything-as-a-Service” paradigm and discuss situations where this concept may fit perfectly or where the concept is only well-suited for specific scenarios. The book will also review the general relationship between cloud computing on the one hand and data management on the other hand. We will also outline the impact of being able to perform data management on a web-scale as one of the key challenges and also requirements for future commercial data management platforms.

In order to provide a comprehensive understanding of data management services in the large, we will introduce the notion of virtualization as one of the key techniques to establish the “Everything-as-a-Service”-paradigm. Virtualization can be exploited on a physical and logical layer to abstract from specific hardware software environments providing an additional indirection layer which can be

used to differentiate between services or components running remotely in cloud environments or locally within a totally controlled system and administration setup. Chapter 2 will therefore discuss different techniques ranging from low-level machine virtualization to application-aware virtualization techniques on schema and application level.

In the following two chapters, we dive into detail while discussing different methods and technologies to run web-scale data management services with special emphasis on operational or analytical workloads. Operational workload on the one hand usually consists of highly concurrent accesses to the same database with conflict potential resulting in a requirement of transactional behavior of the data management service. Analytical workloads on the other hand usually touch hundreds of million of rows within a single query feeding complex statistical models or produce the raw material for sophisticated cockpit applications within larger Business Intelligence environments. Due to the completely different nature of using a data management system, we take a detailed look with respect to the individual scenarios. The first part (Chap. 3) addressing challenges and solutions for transactional workload patterns focuses on consistency models, data fragmentation, and data replication schemes for large web-scale data management scenarios. The second part (Chap. 4) outlines the major techniques and challenges to efficiently handle analytics in big data scenarios by discussing pros and cons of the MapReduce data processing paradigm and presenting different query languages to directly interact with the data management platform.

Cloud-based services do not only offer an efficient data processing framework by providing a specific degree of consistency but also comprise non-functional services like considering quality of service agreements or adhere to certain security and privacy constraints. In Chap. 5, we will outline the core requirements and provide insights into some already existing approaches. In a final step, the book gives the reader a state-of-the-art overview of currently available “everything-as-a-service”-implementations or commercial services with respect to infrastructural services, platform services or within the context of software as a service.



The book closes with an outlook on challenges and open research questions in the context of providing a data management solution following the “as-a-service”-paradigm. Within this discussion, we reflect the different discussion points and outline future directions for a wide and open field of research potential and commercial opportunities.

The overall structure of the book is reflected in the accompanying figure. Since the individual sections are – with the exception of the general introduction – independent of each other, readers are encouraged to jump into the middle of the book or use a different path to capture the presented material. After having read the book, the readers should have a deep understanding of the requirements and existing solutions to establish and operate an efficient data management platform following the “as-a-service”-paradigm.

Dresden, Germany
Ilmenau, Germany

Wolfgang Lehner
Kai-Uwe Sattler

<http://www.springer.com/978-1-4614-6855-4>

Web-Scale Data Management for the Cloud

Lehner, W.; Sattler, K.-U.

2013, XV, 193 p., Hardcover

ISBN: 978-1-4614-6855-4