

Chapter 17

A Security–By–Distribution Approach to Manage Big Data in a Federation of Untrustworthy Clouds

Jens Kohler

University of Applied Sciences Mannheim, Germany

Christian Richard Lorenz

University of Applied Sciences Mannheim, Germany

Markus Gumbel

University of Applied Sciences Mannheim, Germany

Thomas Specht

University of Applied Sciences Mannheim, Germany

Kiril Simov

Bulgarian Academy of Sciences, Bulgaria

ABSTRACT

In recent years, Cloud Computing has drastically changed IT-Architectures in enterprises throughout various branches and countries. Dynamically scalable capabilities like CPUs, storage space, virtual networks, etc. promise cost savings, as huge initial infrastructure investments are not required anymore. This development shows that Cloud Computing is also a promising technology driver for Big Data, as the storage of unstructured data when no concrete and defined data schemes (variety) can be managed with upcoming NoSQL architectures. However, in order to fully exploit these advantages, the integration of a trustworthy 3rd party public cloud provider is necessary. Thus, challenging questions concerning security, compliance, anonymization, and privacy emerge and are still unsolved. To address these challenges, this work presents, implements and evaluates a security-by-distribution approach for NoSQL document stores that distributes data across various cloud providers such that every provider only gets a small data chunk which is worthless without the others.

DOI: 10.4018/978-1-5225-8176-5.ch017

INTRODUCTION

No other trend has changed the entire Information Technology during the last decade as Cloud Computing actually has done. Slowly, the hype about this buzzword abates and enterprises recognize the true added value of renting computing resources from the cloud. Cloud Computing in the context of this work is defined by the five essential characteristics listed in Mell and Grance (2011), with the on-demand self-service where customers are able to rent computing capabilities by themselves whenever they need them, followed by the requirement of a broad network bandwidth access. Furthermore, resources from the cloud are pooled together with the usage of virtualization from a provider perspective, which enables a rapid elasticity to provide requested resources. Finally, all provided resources are monitored (i.e. measured) by both, the providers and the consumers to have a provable accounting model. For most enterprises, the essential benefit is the dynamic scalability of computing resources along with cost advantages from the pay-as-you-go billing models (Furht et al., 2010). But also other benefits like working independently from any location, the fast deployment of resources and the development of new business models and markets with high-dynamic (i.e. elastic) IT infrastructures were key drivers for the development of Cloud Computing (Gens & Shirer, 2013).

Moreover, a new business case regarding Cloud Computing is now emerging: Big Data. Here, huge amounts of unstructured data at a great velocity must be managed, i.e. stored, analyzed, interpreted, corrected, etc. The notion of Big Data was firstly introduced by Pettey and Goasduff (2011) in 2011 where the three above-mentioned properties volume, variety and velocity are explained in greater detail. With respect to this, Cloud Computing is able to offer dynamically scalable resources to address these three challenges: instead of huge initial or new investments in better hardware, the required capabilities can be rented on-demand. Then, they can be used for a certain time, be dynamically scaled according to the data volume and velocity, and finally just turned off when they are not required anymore. Thus, two of the three Big Data properties are addressed, but variety is still a challenging issue. Here, NoSQL (not only SQL) databases offer promising features to efficiently store unstructured data. These new kinds of databases are considered in more detail in this chapter and are therefore defined in the following section.

Additionally, it is a fact that the more data are collected, the more important becomes privacy and security for enterprises as well as for end-customers. This becomes even more challenging, if data are managed in the cloud at an external provider. Therefore, the increasing usage of cloud services is accompanied by concerns regarding security, compliance, and privacy and customers depend on the security measures of the service providers. Moreover, for customers it is not transparent, which security measures are implemented by the provider Neves, Correia, Bruno, Fernando, and Paulo (2013), Sood (2012), and Cloud Security Alliance (2013). Hence, challenges of data privacy and compliance still are the most significant obstacles for an increased usage of Cloud Computing (Gens & Shirer, 2013).

To address these challenging security concerns and to ensure data privacy, several different approaches exist, e.g. encryption with digital signatures. A different approach is developed with SeDiCo at the University of Applied Sciences in Mannheim. SeDiCo (A Secure and Distributed Cloud Data Store) is a framework for distributed and secure data storage. With this framework, sensitive data can be segregated from non-sensitive data and stored at physically different places such as different cloud providers. Thus, the actual place of the data is disguised, as every chunk of data is worthless without the others. For example, bank accounts can be stored separately from the owners' names and thus, even if an attacker gets access to one of the partitions, the compromised data does not contain useful information. A simplified example that illustrates this basic principle is shown in Figure 1.

28 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/a-security-by-distribution-approach-to-manage-big-data-in-a-federation-of-untrustworthy-clouds/224582

Related Content

Security for Cross-Tenant Access Control in Cloud Computing

Pramod P. Pillai, Venkataratnam P. and Siva Yellampalli (2020). *Modern Principles, Practices, and Algorithms for Cloud Security* (pp. 44-78).

www.irma-international.org/chapter/security-for-cross-tenant-access-control-in-cloud-computing/238902

Novel Taxonomy to Select Fog Products and Challenges Faced in Fog Environments

Akashdeep Bhardwaj (2018). *International Journal of Fog Computing* (pp. 35-49).

www.irma-international.org/article/novel-taxonomy-to-select-fog-products-and-challenges-faced-in-fog-environments/198411

Digital Library and Its Requirements in the Global World

Surbhi Saini (2015). *Cloud Technology: Concepts, Methodologies, Tools, and Applications* (pp. 1119-1139).

www.irma-international.org/chapter/digital-library-and-its-requirements-in-the-global-world/119900

Legal Issues Surrounding Connected Government Services: A Closer Look at G-Clouds

Mariam Kiran (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications* (pp. 1787-1808).

www.irma-international.org/chapter/legal-issues-surrounding-connected-government-services/224657

Exploiting Semantics to Improve Classification of Text Corpus

Hammad Majeed and Firoza Erum (2016). *Managing and Processing Big Data in Cloud Computing* (pp. 23-36).

www.irma-international.org/chapter/exploiting-semantics-to-improve-classification-of-text-corpus/143337