Chapter 6
Towards a Secure, Distributed, and Reliable Cloud-Based Reference Architecture for Big Data in Smart Cities

Jens Kohler
University of Applied Sciences Mannheim, Germany

Thomas Specht
University of Applied Sciences Mannheim, Germany

ABSTRACT

Current revolutions with respect to big data affect people's everyday life more and more. New ways of living, communication, and knowledge sharing have been created under the smart city umbrella. Information and communication technologies (ICT) are used to improve urban services. Examples are car sharing, energy consumption, adaptive traffic management, etc. A closer look at these services reveals that data, produced and consumed by people (or automatically by devices), are the cornerstone for working and reliable urban services. For this, large volumes of data with a great variety have to be processed at a high velocity which is commonly known as big data. This chapter designs a secure, distributed, and reliable cloud-based reference architecture that logically separates confidential data and distributes them to various databases in different clouds. It increases the reliability of the distributed data with high-availability mechanisms to protect data against various threads (e.g., hackers, terrorists, data collecting companies, etc.).

INTRODUCTION

Current revolutions with respect to Big Data, Internet of Things, or Industry 4.0 affect people's everyday life more and more. An example for this is the development of the World Wide Web (WWW 1.0) to the Semantic Web (WWW 3.0). Thus, new ways of how people communicate, interact, and share knowl-

DOI: 10.4018/978-1-7998-5339-8.ch006

edge with each other have been created. This also holds for recent developments under the *Smart City* umbrella, which subsumes urban development strategies towards a sustainable management of resources like space, traffic, and other infrastructural assets, but also services, e.g. municipal, governmental, health care, or industrial ones (Song, Srinivasan, Sookoor, & Jeschke, 2016) (Dameri & Rosenthal-Sabroux, 2014) (Dey, Hassanien, Bhatt, Ashour, & Satapathy, 2018) (Elhayatmy, Dey, & Ashour, 2018) (Bhatt, Dey, & Ashour, 2017) (Sarwar Kamal, Dey, & Ashour, 2017) (Yang, Wang, Song, Yang, & Patnaik, 2018). Here, new Information and Communication Technologies (ICT) architectures are required to implement and improve *urban services*: concrete examples are car sharing, adaptive traffic management, energy consumption, etc. A general goal for these *urban services* is to reach a high degree of automation (without any human interaction) to increase their efficiency and to reduce costs. A closer look at these services reveals that data, produced and consumed by people (or automatically by their devices) are the cornerstones for working and reliable *urban services*. Therefore, large *volumes* of data with a great *variety* (heterogeneity) have to be processed at a high *velocity* which is commonly known as *Big Data*.

In order to address these *Big Data* challenges, NoSQL databases (in contrast to traditional relational databases) provide appealing architectures (i.e. key-value, column, document stores and graph databases) (Sadalage & Fowler, 2013) (Harrison, 2015), especially when data are processed In-Memory, i.e. not stored and analyzed on hard disk drives (HDD), but in the faster random-access memory (RAM) of a system. These non-relational architectures weaken the hard consistency criteria atomicity, consistency, isolation, and durability (ACID) of relational databases and use the basically available, soft-state, eventual consistency (BASE) model instead. This is advantageous in especially distributed systems, as the strong ACID criteria demand a tremendous synchronization overhead, whereas BASE allows several nodes to be inconsistent for a certain time. However, processing huge amounts of heterogeneous data at great pace requires adequate (and expensive) infrastructure capabilities and this is where Cloud Computing can be taken into consideration.

On the one hand, Cloud Computing promises advantages with its five key characteristics: (1) broad network access, (2) on-demand self-service, (3) rapid elasticity, (4) resource pooling, and (5) measured services (Mell & Grance, 2011). These key characteristics are also essential in a *Smart City* infrastructure, as huge data volumes require a high bandwidth to and from the cloud, especially with a great variety of different devices (1). People using urban services should be enabled to use them on-demand without any delaying registration and approval processes (2). The infrastructure should be able to handle dynamic changing loads automatically (3), e.g. when big social events with many people take place). The concrete hardware infrastructure must be abstracted from with the usage of virtualization technologies. This enables *resource pooling*, where different hardware capabilities are jointly used to provide multiple services (4). Last but not least, the provided services must be monitored transparently for the *cloud providers* and *consumers* (5) to create reliable and trustworthy services and to establish adequate billing models.

On the other hand, data processed in a cloud infrastructure are highly sensitive and confidential and thus a very attractive goal for hackers, terrorists, data collecting companies, and the like. Questions regarding data security, privacy, and trust are not trivial to solve, especially in a globalized world where people from different countries move, travel, and visit various *Smart Cities* with different national laws. Basically, security goals with respect to confidential and sensitive data can be formulated as the so-called CIA (confidentiality, integrity, and availability) principle:

- *Confidentiality* states that only authorized entities (e.g. people or devices) are able to access data,
- I*ntegrity* means that data is accurate and complete, and

Related Content