Chapter 15
Big Data Security Framework for Distributed Cloud Data Centers

Chandu Thota
Infosys Ltd., India

Gunasekaran Manogaran
VIT University, India

Daphne Lopez
VIT University, India

Vijayakumar V.
VIT University Chennai, India

ABSTRACT

The rapid development of data generation sources such as digital sensors, networks, and smart devices along with their extensive use is leading to create huge database and coins the term Big Data. Cloud Computing enables computing resources such as hardware, storage space and computing tools to be provided as IT services in a pay-as-you-go fashion with high efficiency and effectiveness. Cloud-based technologies with advantages over traditional platforms are rapidly utilized as potential hosts for big data. However, privacy and security is one of major issue in cloud computing due to its availability with very limited user-side control. This chapter proposes security architecture to prevent and secure the data and application being deployed in cloud environment with big data technology. This chapter discuss the security issues for big data in cloud computing and proposes Meta Cloud Data Storage architecture to protect big data in cloud computing environment.

INTRODUCTION

Cloud computing has been defined by following characteristics such as Multi-tenancy, Massive scalability, Elasticity, and pay as you go. In cloud computing, Multi-tenancy is an architecture in which a single instance of any application serves multiple customers. Each customer is called a tenant. Elasticity is an ability of a system to automatically provision and deprovision computing resources on demand as

DOI: 10.4018/978-1-5225-7501-6.ch015

workloads change. Scalability is the ability of software and hardware that is being used to continue to function properly with increased workflow volume, in other words to increase or decrease the required resources, meaning you're not paying for resources which you are not utilizing. The data generated from several sources such as mobile devices, sensors, Internet of Things, enterprises, individual archives, software logs, social networks, cameras, etc. Nowadays, Data Explosions' is playing a vital role in current Information and Communication Technology (ICT) era: how to optimally and effectively manage such huge amount of data and discover new techniques to process large amounts of data for unlocking information. It is important to consider the attributes (Multitenancy, Scalability, Elasticity and virtualization) when integrating big data with cloud and deploying the applications in cloud. Big data has been playing a vital role in almost all applications such as Government and the public sector, Nature and natural processes, Health and human welfare, commerce, business and economic systems, social networking and the internet, and computational and experimental processes (Kambatla et al., 2014). Big data plays a vital role in healthcare to reduce the cost and increase the performance. Traditional clinical data are classified as electronic medical records (EMRs), pharmaceutical data, imaging data, data on personal practices and preferences (including exercise patterns, dietary habits, environmental factors), and financial/activity records. The above mentioned data providers are merged together to enhance the performance, delivery, and well-being (Vayena et al., 2015). A recent report from McKinsey Global Institute states that healthcare analytics could produce more than $300 billion in value every year (Mckinsey.com, 2015).

In addition, distributed repositories in healthcare enables user with huge access. For example, imaging data (MRI, fMRI) is stored in distributed manner and frequently accessed by skilled radiologists to make better diagnoses and delivery (Kayyali et al., 2013). In recent years, public sector and government also use big-data analytics to maintain the general services administration data for huge access. For example, AWS GovCloudis constructed to move exhaustive workloads to the cloud. Cloud computing and big data have reduced the execution time (both upload and download) and operational costs (Frost.com, 2016; Kim et al., 2014). Nowadays, Social networking and the internet have been playing a vital role in day-to-day life. Recently, Facebook announced that 2 billion people are actively using social media each month (We Are Social UK, 2015). People emotion monitoring and sentiment analysis have been applied to solve many real time issues (Wang et al., 2012). Big data is very challenging task for computing platforms and organizations. For example, Geospatial modeling (Mhlanga et al., 2015), quantum-mechanical modeling (Pandey et al., 2015), Astro-physical simulations, use computational simulations and scientific instruments to model the real time big datasets to bring in qualitative and quantitative changes in day-to-day life (Reed et al., 2015). Big data is not only used for business and development, but also helps to save nature and natural processes. Natural resource monitoring data is normally collected from satellite imagery, sensors and radars to monitor the extreme weather events, deforestation and urban encroachment. Thus, big data analytics has major impact in, including land and water resources management (Wang et al., 2013), sustainable development (Gijzen et al., 2013), global warming and climate change and environmental impact assessment (Howe et al., 2008), natural resource management (Hampton et al., 2013) and healthcare (Lopez et al., 2014; Lopez et al., 2015; Lopez et al., 2016). There is a need to develop an efficient architecture to process the big data. The intention of this chapter is to propose a security architecture to protect and process the big data in cloud computing.