Chapter 11
Demystifying Big Data in the Cloud:
Enhancing Privacy and Security Using Data Mining Techniques

Gebeyehu Belay Gebremeskel
Chongqing University, China

Yi Chai
Chongqing University, China

Zhongshi He
Chongqing University, China

ABSTRACT

Big data in the cloud are an emerging paradigm for huge and federated data processing, storing and distributing by deploying web applications. Scalability, elasticity, pay-per-use pricing, and an advance of ICT scale from large and dynamic applications and performance are the major reasons for the success and widespread adoption of big data cloud infrastructures. It is 'no secret of the enterprise data', which is challenging for privacy and security. In this chapter, authors deeply discussed and introduce novel approaches and methodologies to easily understood big data phenomenon and technology towards data or web resources privacy and security. Nutshell, big data has a powerful potential to predict cloud risks to develop and deploy corporate security strategies. The chapter's contribution is, in general, to gain a meaningful insight of big data in the cloud and its applications, which is hot issues for today's businesses to make proactive and knowledge-driven decisions.

INTRODUCTION

Computing has become ubiquitous, creating countless new digital puddles, lakes, and oceans of data in its scale, time, type, characteristics, nature …, which is big data in the cloud. Big data is a newly emerging technology and promising to handle this data deluge. However, it also has a big challenge and creates possible risks for the organizations and consumers or users' data security and privacy. "In big data no secret" –if so how does personal or some other sensitive data secure?' Therefore, demystifying big data in the cloud using a Data Mining (DM) approach is a systematic approach to overcoming such potential problems and optimization of cloud performance. Enhancing Cloud Data Privacy and Security (ECDPS) based on DM techniques or technologies are fueling the use of inferential data techniques, for example, spatial data, traffic monitors, magnetic resonance imaging machines, and biological and chemical sensors monitoring … namely few. Moreover, big data as a technology is expected to play a big role to transform competitive dynamics in business and science from root level to the macro level from retail to biotech in tremendously. DM techniques integrated with agents' technologies of searching and/or mining processes for implicit, previously unknown, and potentially valuable information from large-scale data sets are a tactical approach of big data, which support to detect cloud resources risks, such as intrusion patterns, security problems, and so on (Bhavani, Latifur, Mohammad & Kevin, 2008; Lin, Hinke & Marks, 1996).

Security and privacy in the cloud are essential and core components. For example, for consumer's trust and involvement are determined by these factors, such as expanding their online profile and storing personal information in the cloud, which sets up payments and billing for cloud software services, shopping and other activities (Chris & Don, 1996). However, big data is a growing technology that did not well adopt to handle, manage and analytics of massive data that generated by various sources. In such federated and massive data exploration tasks, privacy and security have always been a barrier to adoption. Therefore, understanding big data is the first step in assessing the potential risks need and putting a big data analytics by enhancing DM based security and privacy. The second understands the cloud system and its trends that are affecting organizations looking forward to derive business value, and competitive advantages, from increasingly large and diverse data sets. In this proposed research, the detail of big data in the cloud, DM techniques towards cloud data security and privacy are discussed, which would be a generic inferring paper for scientists, researchers, students and decision makers in general.

In this chapter, authors introduce a novel idea and methodology using DM techniques. It is a systematic approach to demystifying implications of big data in the cloud by knowing and expanding the inferential data analysis. It is a paradigm of empowering the widely adoptions of the new and dynamic data analysis and

Related Content

Map Overlay Problem
Maikel Garma de la Osaand Yissell Arias Sánchez (2009). *Handbook of Research on Geoinformatics (pp. 65-72).*
www.irma-international.org/chapter/map-overlay-problem/20388

Spatial Data Infrastructures
Carlos Granell, Michael Gould, Miguel Ángel Mansoand Miguel Ángel Bernabé (2009). *Handbook of Research on Geoinformatics (pp. 36-41).*
www.irma-international.org/chapter/spatial-data-infrastructures/20384

Investigation of Possible Landslide Precursor Activity in a Small-Scale Laboratory Experiment
Spiridon G. Krokidis, Konstantinos Marmarokoposand Markos Avlonitis (2018). *International Journal of Applied Geospatial Research (pp. 74-86).*
www.irma-international.org/article/investigation-of-possible-landslide-precursor-activity-in-a-small-scale-laboratory-experiment/210153

Utilizing BCF and IFC for Automated Dimensional Compliance Control
Neil Hyland, Shawn E. O'Keeffeand Shane Brodie (2017). *International Journal of 3-D Information Modeling (pp. 16-36).*
www.irma-international.org/article/utilizing-bcf-and-ifc-for-automated-dimensional-compliance-control/216386

Research Commentary: Increasing the Flexibility of Legacy Systems
William L. Garrison, Barry Wellar, Ross MacKinnon, William R. Blackand Arthur Getis (2013). *Emerging Methods and Multidisciplinary Applications in Geospatial Research (pp. 198-214).*
www.irma-international.org/chapter/research-commentary-increasing-flexibility-legacy/68258