Future Generation Computer Systems 59 (2016) 63–92

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Expanded cloud plumes hiding Big Data ecosystem

Sugam Sharma

Center for Survey Statistics and Methodology, Iowa State University, Ames, IA, USA

HIGHLIGHTS

- In sum, growing role of cloud in Big Data ecosystem is extensively investigated.
- A novel XCLOUDX classification is proposed for cloud-assisted Big Data models.
- A set of important case studies on cloud Big Data models is discussed.
- The as-a-Service modality of cloud computing is extensively discussed.
- IoT, Smart Data, and Smart Data Lakes are briefly discussed as future directions.

ARTICLE INFO

Article history:
Received 6 September 2015
Received in revised form
25 December 2015
Accepted 11 January 2016
Available online 28 January 2016

Keywords:
Cloud
Big Data
Smart Data and Lakes
IoT
XCLOUDX
as-a-Service

ABSTRACT

Today, a paradigm shift is being observed in science, where the focus is gradually shifting away from operation to data, which is greatly influencing the decision making also. The data is being inundated proactively from several sources in various forms; especially social media and in modern data science vocabulary is being recognized as Big Data. Today, Big Data is permeating through the bigger aspect of human life for scientific and commercial dependencies, especially for massive scale data analytics of beyond the exabyte magnitude. As the footprint of Big Data applications is continuously expanding, the reliability on cloud environments is also increasing to obtain appropriate, robust and affordable services to deal with Big Data challenges. Cloud computing avoids any need to locally maintain the overly scaled computing infrastructure that include not only dedicated space, but the expensive hardware and software also. Several data models to process Big Data are already developed and a number of such models are still emerging, potentially relying on heterogeneous underlying storage technologies, including cloud computing. In this paper, we investigate the growing role of cloud computing in Big Data ecosystem. Also, we propose a novel XCLOUDX {XCloudX, X…X}classification to zoom in to gauge the intuitiveness of the scientific name of the cloud-assisted NoSQL Big Data models and analyze whether XCloudX always uses cloud computing underneath or vice versa. XCloudX symbolizes those NoSQL Big Data models that embody the term "cloud" in their name, where X is any alphanumeric variable. The discussion is strengthen by a set of important case studies. Furthermore, we study the emergence of as-a-Service era, motivated by cloud computing drive and explore the new members beyond traditional cloud computing stack, developed in the past couple of years.

1. Introduction

In last few years, the volume as well as the variety of the data is observed to be grown extensively (Fig. 1a), especially because of the social media such as Facebook, Twitter, multimedia, Internet of Things (IoT), etc. This fast paced heterogeneous, structured or unstructured data generation has overwhelmed the community and as a whole, in modern data science vocabulary it is named as Big Data [1] (Fig. 1b). The birth of Big Data has seriously challenged the widely prevailed existing data management capabilities. The legacy computation techniques are not at par to deal with the Big Data. The rapidly growing volume of the multifarious data has caused an absolute paradigm shift in how the modern data should be computed. This new era of Big Data witnesses a revolutionized evolution of new potential data engineering techniques. Also, the existing infrastructure is required to be sufficiently large, efficient and scalable that is capable to easily accommodate the Big Data challenges. Such infrastructure requires not only high-end hardware and software, but expansive maintenance efforts also. However, several data models to efficiently process Big Data are already developed and a number of such models are still emerging, relying on heterogeneous underlying storage technologies. As

E-mail address: sugamsha@iastate.edu.

http://dx.doi.org/10.1016/j.future.2016.01.003
0167-739X/

Fig. 1a. Data growth vs. years.
Source: walasolutions.com/simplifying-big-data/.

Fig. 1b. Big Data.
Source: mattclarkecto.com/page/2/.

the footprint of Big Data applications grows, an increase in the infrastructure cost cannot be ruled out completely. Consequently, the reliability on cloud environments is also increasing to obtain appropriate services to deal with Big Data stress. However, to construct a suitable, efficient and cost-effective cloud computing environment is also a major challenge.

Today, the cloud computing also has emerged as one of the major shifts in recent information and communication age that promises an affordable and robust computational architecture for large-scale and even for overly complex enterprise applications. It is a powerful and revolutionary paradigm that offers service-oriented computing and abstracts the software-equipped hardware infrastructure from the clients or users. Although, the concept of cloud computing is mainly popular in three flavors—(1) Infrastructure-as-a-Service (IaaS), (2) Platform-as-a-Service (PaaS) and (3) Software-as-a-Service (SaaS), but in this data science age, should be equally expandable to Database-as-a-Service (DBaaS) [2].

One of the other rationales behind the growing popularity of cloud computing over the traditional infrastructure, simple to complex up to enterprise level is the economic feasibility it offers. It is a new shape of business model that is gradually turning into a powerful ubiquitous paradigm to perform massively scaled complex computing. It maintains and delivers the required resources and services on demand with a pay-per-use scheme, especially related to information technology (IT) to the requester business entities. Cloud computing greatly destroys the needs of localization of housing and maintaining the complex and expensive computing infrastructure, hardware and software both, and required dedicated physical space. Some of the other salient features of cloud computing include, comparatively lower upfront investment, risk free environment at client site, and allowable elasticity in requested resources, services, and consequently payment. Today, more and more applications are inclining to avail the benefits of cloud computing and a growing number of such applications are noticed on cloud platforms. Also, observed is the tremendous growth in the scale of data (Big Data) cultivated through the complex computing applications housed in the cloud. Thus, the scalable data management is becoming a critical but integral part of the cloud computing infrastructures, especially for the intensive data-driven applications such as decision support systems. Although, the pursuit to address the Big Data challenges has given rise to number of systems, but the scalability, availability, integrity, transformation, quality, and heterogeneity of the data always remain challenging. In this study, we aim to investigate comprehensively the role and status of cloud computing environment in Big Data ecosystem. This extended investigation will cover a broader spectrum of cloud computing and its growing utility in Big Data. As mentioned above, the quest for efficient solutions for the Big Data challenges has led to a plethora of data models; we propose a novel XCLOUDX {XCloudX, $X \ldots X$} classification, which is solely based upon the scientific name of the data models; the term "X" can be a set of variables and/or numerals. The objective of this novel characterization is to understand whether the name of the cloud-assisted NoSQL Big Data models (XCLOUDX) is intuitive enough to indicate to cloud computing environment underneath or vice versa. The pool of study contains cloud-assisted those data models (XCLOUDX), which (1) consist of the term "cloud" in their name as one of constituents such as XCloudX (one of two children of XCLOUDX), (2) do not include the word "cloud" in their name at all; the latter category is represented by $X \ldots X$ class (another child of XCLOUDX). Additionally, we briefly explore some interesting case studies from different business domains, academia or industry that further strengthen the notion of the growing spectrum of cloud in Big Data ecosystem. Over the last few years, the cloud computing has evolved as scalable, secure and cost-effective solutions to an overwhelming number of applications from diversified areas. The traditional cloud ecosystem delivers the services in three popular flavors—IaaS, PaaS and SaaS. The scientific communities understand the purview of pervasiveness of cloud computing and intent to fan out its technological advantages to benefit every single field of study. The imagination of ubiquitous information access, free from the geography constraints, and supported by the distinguish characteristics of cloud computing has motivated enough the communities further that has extended the cloud ecosystem beyond IaaS, PaaS, and SaaS. This gives rise to the accelerated evolution of "as-a-Service" (aaS) era, permeating through almost all the domains in academia or industry. Moreover, we highlight an extended list of new entrants of aaS family to the cloud computing stack on the top of IaaS, PaaS, and SaaS. We believe that the thorough understanding of the relationship of the cloud computing with Big Data is inevitable to ensure the efficient and successful management of the scalable data (such as Big Data) in the cloud environment, emulating the popularity of relational data management in traditional infrastructure settings.

Rest of the paper is structured as follows. The definition, characteristics, and classification of Big Data are briefly introduced in Section 2. Section 3 provides the brief overview of cloud computing. Section 4 is expanded to explore the role of cloud computing in Big Data. Section 5 discusses the proposed novel XCLOUDX classification and reviews some of the leading Big Data models of XCloudX kind. This section also investigates whether there are some $X \ldots X$ type data models that also leverage cloud computing for data analytics. Section 6 explores some of the interesting and fresh case studies. Section 7 highlights the emergence of as-a-Service era. Finally, the paper concludes in Section 8.

2. Definition, characteristics and classification of big data

The massive amount of digital footprints and sizable traces are being deposited through the leading data outlets such as social media, sensors, power meters, etc. About 2.5 quintillion bytes of