Chapter 10
Accessing Big Data in the Cloud Using Mobile Devices

Haoliang Wang
George Mason University, USA

Wei Liu
University of Rochester, USA

Tolga Soyata
University of Rochester, USA

ABSTRACT

The amount of data acquired, stored, and processed annually over the Internet has exceeded the processing capabilities of modern computer systems, including supercomputers with multiple-Petaflop processing power, giving rise to the term Big Data. Continuous research efforts to implement systems to cope with this insurmountable amount of data are underway. The authors introduce the ongoing research in three different facets: 1) in the Acquisition front, they introduce a concept that has come to the forefront in the past few years: Internet-of-Things (IoT), which will be one of the major sources for Big Data generation in the following decades. The authors provide a brief survey of IoT to understand the concept and the ongoing research in this field. 2) In the Cloud Storage and Processing front, they provide a survey of techniques to efficiently store the acquired Big Data in the cloud, index it, and get it ready for process-ing. While IoT relates primarily to sensor nodes and thin devices, the authors study this storage and processing aspect of Big Data within the framework of Cloud Computing. 3) In the Mobile Access front, they perform a survey of existing infrastructures to access the Big Data efficiently via mobile devices. This survey also includes intermediate devices, such as a Cloudlet, to accelerate the Big Data collection from IoT and access to Big Data for applications that require response times that are close to real-time.

INTRODUCTION

The amount of data generated annually over the Internet has exceeded the zetabyte levels. Process-ing data with such high volume far exceeds the computational capabilities of today's datacenters and computers, giving rise to the term *Big Data*. Although the growth rate of supercomputers that are capable of processing such explosive amount of data is also breathtaking (TOP500, n.d.), the

rate of data growth far surpasses the capabilities of even the fastest supercomputers available today. Even though the top supercomputers are able to handle Big Data analysis, their highly-specialized designs are not affordable for commercial use. Instead, large commodity computer clusters are used, where faults are common and interconnect speeds are limited. Also the storage and management of Big Data poses different unique challenges: While the storage has to be performed by high-availability and high-performance distributed file systems, it must also be done in a way to allow application of efficient data analytics later. Being able to perform analytics on this data is crucial: It has been reported that, performing analytics on Big Data can save the government 14% all across their budget (Big Data, 2013). This specific example shows the importance of manipulating Big Data while keeping both phases of usage in mind concurrently: storage and computation.

By today's standards, considering the utility computing (termed *Cloud Computing*), is unavoidable for any organization, regardless of its size. While it is possible for different organizations to build their own datacenters, it is an expensive business proposition to do so, since the economies of scale for organizations such as Amazon (AWS, n.d.), Google (Google, n.d.), and Microsoft (Microsoft, n.d.), will allow them to build these datacenters for a fraction of the price. Furthermore, while an organization that is building its own datacenter must size it for the worst case, cloud operators offer much more favorable pricing options, such as, per-hour usage pricing. This allows corporations to rent much higher peak amounts of computational power with zero upfront investment. To make cloud computing even more appealing, the responsibility of continuously upgrading the underlying computational infrastructure is shifted to the cloud operators, thereby permitting access to modern high performance resources whenever they are available without any investment.

Due to the wide scope of Big Data and cloud computing, we restrict our focus to futuristic concepts involving Big Data in this chapter. Specifically, we will investigate one emerging source of Big Data, called *Internet of Things* (IoT). IoT, introduced in 1999, conceptualizes a network of numerous data-generating devices (things) such as home energy meters, wireless sensors, and other sensory devices. For IoT to be realized, a unique Internet addressing scheme for each device, called IPv6, is necessary that significantly expands what used to be the standard a decade ago (IPv4). With the widespread use of IPv6, each device (i.e., thing) can be assigned its unique address to globally identify it over the Internet. The acceptance of IPv6 is accelerating for desktop PCs and is expected to expand over to IoT within the following decade.

Cloud computing, as a new model for delivering computing resources on demand, provides a powerful, flexible and elastic platform which enables collection, analytics, processing and visualization of Big Data. Storage of Big Data is performed by file systems that are drastically different than traditional file systems such as NT File System (NTFS). One such user-level distributed file system – Google File System (GFS) allows not only the distributed storage of Big Data, but also its access with high availability (and fault-tolerance) due to the built-in redundancy in GFS. This file system also dictates how the processing should be performed: Standardized methods, such as MapReduce, ease the handling of Big Data and provide a tool for cloud operators to make their platform more accessible. Cloud computing service providers have already releases of the public platforms for Big Data analysis (Amazon Elastic MapReduce and Google BigQuery).

Access to Big Data in the cloud through mobile devices (termed *Mobile-Cloud Computing*) significantly expands the reach of Big Data due to the widespread availability of smartphones and tablets. While multiple definitions are available in the literature (Dinh et al., 2011; Fernando et al., 2013), mobile-cloud computing can be defined as the "co-execution of a mobile application within

Related Content

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

Development of Community Based Intelligent Modules Using IoT to Make Cities Smarter
Jagadish S. Kallimani, Chekuri Sailusha, Pankaj Latharand Srinivasa K.G. (2019). *International Journal of Fog Computing (pp. 1-12).*
www.irma-international.org/article/development-of-community-based-intelligent-modules-using-iot-to-make-cities-smarter/228127

An IoT-Based Framework for Health Monitoring Systems: A Case Study Approach
N. Sudhakar Yadav, K. G. Srinivasaand B. Eswara Reddy (2019). *International Journal of Fog Computing (pp. 43-60).*
www.irma-international.org/article/an-iot-based-framework-for-health-monitoring-systems/219360

Detection and Classification of Dense Tomato Fruits by Integrating Coordinate Attention Mechanism With YOLO Model
Seetharam Nagesh Appe, G. Arulselviand Balaji G. N. (2023). *Handbook of Research on Deep Learning Techniques for Cloud-Based Industrial IoT (pp. 278-289).*
www.irma-international.org/chapter/detection-and-classification-of-dense-tomato-fruits-by-integrating-coordinate-attention-mechanism-with-yolo-model/325947

Security Model for Mobile Cloud Database as a Service (DBaaS)
Kashif Munir (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications (pp. 760-769).*
www.irma-international.org/chapter/security-model-for-mobile-cloud-database-as-a-service-dbaas/224604