Chapter 66
Big Data Virtualization and Visualization: On the Cloud

Muhammad Adeel
International Islamic University, Pakistan

ABSTRACT

With the recent explosion of internet usage as well as more and more devices are being hooked up with the cloud, big data is becoming a phenomena to tackle with. Big data management was initially a question of concern for only the big commercial players such as Google, Yahoo, Microsoft and others. But it has now become a concern for others, too. According to recent estimates, big data will continue to grow from terabytes into exabytes and beyond. This data needs to be made available for an organization's own use as well can be made available for scientific and commercial needs to the interested entities. This can include different user segments such as academia, industry etc. Academic use of big data is for further research and enablement of big data over cloud, working with it in containers, usage in virtualized environments etc. This generates a need for a sustainable infrastructure which can hold and maintain big data with opportunities for extended processing.

INTRODUCTION

Big data is an all-encompassing term for any collection of data sets which has become too large and complex to be processed by any conventional means. This implies that such huge data will not be able to be stored in simple conventional storage. Such huge data has to be first of all collected and stored ready to be provided in accessible formats. Such Big Data may be logs, mobile, banking transactions, online user generated content such as blog posts and tweets, online searches and satellite images. The data through its processing and its visualization can result in precious hidden information to be found out from that data.

Moreover, big data refers to a collection of datasets that are huge in size and complex to handle by commonly used data processing tools and applications. These datasets can be structured as well as

DOI: 10.4018/978-1-5225-1837-2.ch066

unstructured and often come from various sources such as social media, social sensors, scientific applications, surveillance cameras, archives, web documents, electronic health records, business applications and web logs. They are mammoth in size with sometimes even more data arriving at a very fast rate a.k.a streaming data. Further, big data should have high value and ensure trust for decision making. Also, these data come from heterogeneous sources and heterogeneity is another important property for this big data besides its variety, volume, velocity, value and veracity.

There are many technologies that support handling big data including parallel processing, distributed computing, cloud computing platforms, large storage systems and MapReduce. There is an urgent need to investigate the challenges of big data computing by leveraging the potential of cloud computing. The chapter will try to bring into focus the concepts related to big data virtualization and its visualization on top of cloud computing.

SIGNIFICANCE OF BIG DATA AND APPLICATIONS

Big data management was initially a question of concern for only the big commercial entities such as Google, Yahoo, Microsoft, big research organizations (such as Nasa which may have to deal with trillions of bytes of satellite imagery, massive capturing of space signals etc), governments and so on. With the recent explosion of internet usage we can see, that more and more devices are being hooked up with the cloud (internet), and there are multiple government, commerical and defence orgranizations which are collecting and/or interested in analysing this big data. According to recent estimates, this data will continue to grow from terabytes into exabytes and possibly even beyond. This data needs to be made available for scientific and commerical needs to the interested parties in addition to its normal use. This can include different user segments such as academia, industry etc. Academic use of big data can result in further research and enablement of big data over cloud, experiments on big data in cloud computing, cloud based big data research, application containers, big data in virtual environments etc. Commercial use of this data can be quite wide ranging such as sentiment analysis, customer behaviour, future trend analysis, pattern mining in big data and the like. This generates a need for a sustainable infrastructure which can hold and maintain big data with opportunities for extended processing, visualization and transformations. Big data has some distinguishing characteristics which makes it unique than traditional data and which are described below in some detail:

- **Volume:** Big data is typically in the order of terabytes, peta bytes and even more. It become obvious that such data cannot be stored in any traditional means easily.
- **Value:** Big data brings value. It may be holding precious gems of information ready to be mined. This information will remain hidden if it cannot be fully processed and analyzed.
- **Velocity:** Big data such as sensors data collected through traffic sensors, home devices sensors, public places sensors, space signals analysis data, flight data of a Boeing 737. It is obvious that such data will be arriving, in fact streaming at a very fast pace. This generates the need to filter the noise out and store the remaining interesting big data quickly.
- **Veracity:** The data validity or truthfulness is important nonetheless. The ultimate value to be derived from big data depends solely on the level of veracity in the big data. The big data has to be truly authentic and should contain only minor defects and errors.