

Chapter 37

Big Data Computing Strategies

Malavika Jayanand

M. S. Ramaiah Institute of Technology, India

K. G. Srinivasa

M. S. Ramaiah Institute of Technology, India

M. Anil Kumar

M. S. Ramaiah Institute of Technology, India

G. M. Siddesh

M. S. Ramaiah Institute of Technology, India

ABSTRACT

In order to store these huge bulks of data, organizations have to buy servers and scale it according to their need. One of the solution for storage would be Cloud Environment as there is no need to scale up storage by aggregating more physical servers, installing, updating, or to run backups, which also cuts down system hardware and makes application integration easier. This chapter discusses key concepts of cloud storage as storage as a solution for Big data, different big data domains, and its application with case studies of how organizations are leveraging Big data for better business opportunities.

1. INTRODUCTION

1.1 Big Data: A Revolutionary Technology

Big Data and Cloud Computing has brought about huge transition to computing infrastructure, traditional computing techniques cannot be used for storage and processing of very large quantities of digital information. Cisco cited an estimate that, by end of 2015, global internet traffic will reach 4.8 zettabytes a year that is 4.8 billion terabytes which indicates both the Big Data challenge as well as Big Data opportunities on the horizon.

1.2 Exploding Data Volumes

The amount of digital data being generated is growing exponentially for a number of related reasons. First, every organization is becoming increasingly aware of value locked in massive amounts of data. E-commerce, retailers are starting to build up vast databases of recorded customer activity. Organizations working in many sectors including healthcare, retail, social media are generating additional value by capturing more and more data. New data collectors such as sensors, geo-data, internet click tracking

DOI: 10.4018/978-1-4666-9840-6.ch037

have created a world where everything around us where right from automobiles to mobile phones are collecting massive amount of data which may potentially be mined to generate valuable insights.

1.3 How to Avoid Data Exhaust?

The more an organization recognizes the lucrative role of Big Data more data they seek to capture and utilize. However, because of big data's volume, velocity and variety few organizations have ignored huge quantities of very valuable potential information. In effect most of the data that surrounds organizations today is ignored. Massive amount of data that they gather is lost unprocessed, with a significant quantity of useful information passing straight through them as "data exhaust".

Traditional large-scale computing solutions rely on expensive server hardware which are highly fault tolerant towards hardware failures or other system problems at the application level because of which there is a high level of service continuity to be delivered from clusters server computers, each of which may be prone to failure. The most feasible premise would be to process vast quantities of data across large low-cost distributed computing infrastructures.

Big Data requires new technological solutions. The need for more dependable, scalable, distributed computing systems capable of handling the Big Data deluge led to use of more flexible technologies like Hadoop, database virtualization, storage virtualization, network virtualization, and more in order to avoid single device barrier, since they impede scaling. To spread analytical processing across bunch of commodity servers, Hadoop uses MapReduce. Elasticity and agility were needed to scale and address Big data

Big Data is characterized by its volume and velocity of data and for convenient processing requires decentralization. Due to lot of limiting factors such as resources, expertise in the domain etc, organizations are highly unlikely to implement their own solutions. Nonetheless, big pioneers in the market like Amazon, NetApp and Google, allow organizations of all sizes to start benefiting from Big Data processing capabilities. Big Data sets need to be utilized, since data and processing need will change and change may be rapid and disruptive a more flexible platform such as Cloud computing services which makes it more manageable and agile (James Taylor, 2011). As new types of computer processing power become available, Big Data will progress in leaps and bounds as technology advances.

2. CLOUD COMPUTING FOR BIG DATA

Cloud computing transforms the way big data has been dealt with for all these years – it has emerged as a cost-effective and elastic computing paradigm that facilitates large scale data storage and analysis. The cloud facilitates Big Data processing for wide spectrum of diverse enterprises of various sizes. Getting Business insights from Big Data is prevailing and Cloud is becoming an ideal choice for that. Cloud complements big data since cloud computing provides boundless capabilities on demand. Having these capabilities made available now, any organization can work with unstructured data at a huge scale (Gartner Report, 2013).

Relational Database management systems or a Traditional data management can handle only limited data which are either structured or semi-structured. But, they do not have the capabilities to handle the "Big Data" which consists of massive amount of data sets of varied data types (John, 2007) A new variety

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/big-data-computing-strategies/150194

Related Content

Preserving Privacy in Time Series Data Mining

Ye Zhu, Yongjian Fu and Huirong Fu (2013). *Developments in Data Extraction, Management, and Analysis* (pp. 325-346).

www.irma-international.org/chapter/preserving-privacy-time-series-data/70804

HASTA: A Hierarchical-Grid Clustering Algorithm with Data Field

Shuliang Wang and Yansen Chen (2014). *International Journal of Data Warehousing and Mining* (pp. 39-54).

www.irma-international.org/article/hasta/110385

Spatio-Temporal OLAP Queries Similarity Measure and Algorithm

Olfa Layouni and Jalel Akaichi (2019). *International Journal of Data Warehousing and Mining* (pp. 22-41).

www.irma-international.org/article/spatio-temporal-olap-queries-similarity-measure-and-algorithm/225805

Ontologies, Repository, and Information Mining in Component-Based Software Engineering Environment

Rajesh Kumar Bawa and Iqbaldeep Kaur (2019). *Extracting Knowledge From Opinion Mining* (pp. 104-122).

www.irma-international.org/chapter/ontologies-repository-and-information-mining-in-component-based-software-engineering-environment/211555

Conclusions and Future Work

Wynne Hsu, Mong Li Lee and Junmei Wang (2008). *Temporal and Spatio-Temporal Data Mining* (pp. 262-265).

www.irma-international.org/chapter/conclusions-future-work/30270