Chapter 1
Adapting Big Data Ecosystem for Landscape of Real World Applications

Jyotsna Talreja Wassan
University of Delhi, India

ABSTRACT

Big data is revolutionizing the world in the internet age. The wide variety of areas like online businesses, electronic health management, social networking, demographics, geographic information systems, online education, etc. are gaining insight from big data principles. Big data is comprised of heterogeneous datasets which are too large to be handled by traditional relational database systems. An important reason for explosion of interest in big data is that it has become cheap to store volumes of data and there is a major rise in computation capacity. This chapter gives an overview of big data ecosystems comprising various big data platforms useful in today's competitive world.

INTRODUCTION

Big data is revolutionizing world in the age of Internet. The wide variety of areas like online businesses, electronic health management, social networking, demographics, geographic information systems, online education etc. are gaining insight from *big data principles*. Big data is comprised of heterogeneous datasets which are too large to be handled by traditional relational database systems. An important reason for explosion of interest in big data is that it has become cheap to store volumes of data and there is a major rise in computation capacity.

To extract valuable patterns from big data, one needs to choose a right platform for capturing, organizing, searching and analyzing the context of voluminous data.

Data Management sytems adhering to big data, aim to add computing nodes to cater to increasing data volumes, automatic balancing of data between various nodes, and reducing the operational cost for functioning of distributing data over various nodes (Patel, 2016).

DOI: 10.4018/978-1-5225-7598-6.ch001

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Various NoSQL data stores like Cassandra, MongoDB and Hadoop HBASE etc. are in use today to acquire, manage, store and query big data. NoSQL databases are inherently schema-less and permit records to have variable number of fields, making them distinct from other non-relational databases like hierarchical databases and object-oriented databases. These are highly scalable and well suited for dynamic data structures. NoSQL data is characterized by being basically available and eventually consistent.

The frameworks like MapReduce, Dryad etc. support processing of large amounts of data in parallel and hence the management of big data (Singh & Reddy, 2015). The technologies like GNU R and Apache MAHOUT are useful in exploring and analyzing big data for finding relevant valuable patterns. This article aims at giving an overview of Big Data Ecosystem comprising various big data platforms useful in today's competitive world.

BACKGROUND

In 1970's big meant megabytes, subsequently with the increasing data needs, it grew to gigabytes and terabytes and further to zettabytes with the increase in digital information. The traditional world of relational database systems like Oracle RDBMS etc. faced challenges in storing large quantities of data and needed to scale databases to data volumes beyond the storage and/or processing capabilities of a single large computer system. Many efforts have been made to store and manage data being generated from everywhere on the web. Several database management systems were proposed on the basis of master/slave, cluster computing or partitioning architecture like IBM DB2 partitioning, VoltTB etc.

However, the problems in reliance on shared facilities and resources (CPU, Disk, and Processors), scalability and complex administration limitations, augmented by lack of support for critical requirements, led to development of SHARED NOTHING architectures (Strauch, 2011; Lee, 2011) in 1980's. These systems focused on parallel and distributed data computation and solved big data problems using parallel computations. By 90's, even these solutions faced challenges in running OLTP and queries due to data overload. To provide solutions to these problems, Google responded with its GFS (Dean & Ghemawat, 2004), followed by a powerful programming paradigm of MapReduce (Dean & Ghemawat, 2004) .Thereafter a spectrum of new technologies emerged as the NoSQL movement stating a broad class of database management system to support increasing data storage and analytical requirements.

MAIN FOCUS

Major real world applications like health care, business analytics etc., operational on big data, cannot store or process all of the data on just one machine. The data must be stored, distributed or processed in parallel manner for computations to be completed efficiently. Various platforms are making *big data* management and processing more effective, forming the basis of current research theme in the era of *Big Data (*Gandomi & Haider, 2015). The main focus of this article is to discuss about NoSql Movement, big data platforms which could support processing of futuristic massive volumes of data in parallel and their applications.

12 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/adapting-big-data-ecosystem-for-landscape-of-real-world-applications/214600

Related Content

Packet Dropping Counter Measures in a MANET Through Reliable Routing Protocol Leveraging a Trust Management Framework
Shirina Samreen (2018). *International Journal of Mobile Computing and Multimedia Communications (pp. 60-75).*
www.irma-international.org/article/packet-dropping-counter-measures-in-a-manet-through-reliable-routing-protocol-leveraging-a-trust-management-framework/209390

Performance Evaluation of 2-Wavelength Cognitive Wireless Network for V2R and V2V Communication
Akira Sakuraba, Yoshitaka Shibata, Goshi Satoand Noriki Uchida (2020). *International Journal of Mobile Computing and Multimedia Communications (pp. 84-101).*
www.irma-international.org/article/performance-evaluation-of-2-wavelength-cognitive-wireless-network-for-v2r-and-v2v-communication/273170

Machine Learning Based Prediction and Prevention of Malicious Inventory Occupied Orders
Qinghong Yang, Xiangquan Hu, Zhichao Chengand Kang Miao (2014). *International Journal of Mobile Computing and Multimedia Communications (pp. 56-72).*
www.irma-international.org/article/machine-learning-based-prediction-and-prevention-of-malicious-inventory-occupied-orders/144445

Design of an Enhanced 3G-Based Mobile Healthcare System
Julián Fernández Navajas, Antonio Valdovinos Bardají, Robert S.H. Istepanian, José García Moros, José Ruiz Masand Eduardo Antonio Viruete Navarro (2009). *Mobile Computing: Concepts, Methodologies, Tools, and Applications (pp. 419-431).*
www.irma-international.org/chapter/design-enhanced-based-mobile-healthcare/26518

Aligning iPad Applications with Evidence-Based Practices in Inclusive and Special Education
Therese Cumming, Cathi Draper Rodríguezand Iva Strnadová (2013). *Pedagogical Applications and Social Effects of Mobile Technology Integration (pp. 55-78).*
www.irma-international.org/chapter/aligning-ipad-applications-evidence-based/74905