

Chapter 10

Big-Data-Based Architectures and Techniques: Big Data Reference Architecture

Gopala Krishna Behara

Wipro Technologies, India

ABSTRACT

This chapter covers the essentials of big data analytics ecosystems primarily from the business and technology context. It delivers insight into key concepts and terminology that define the essence of big data and the promise it holds to deliver sophisticated business insights. The various characteristics that distinguish big data datasets are articulated. It also describes the conceptual and logical reference architecture to manage a huge volume of data generated by various data sources of an enterprise. It also covers drivers, opportunities, and benefits of big data analytics implementation applicable to the real world.

INTRODUCTION

In Information Age, we are overwhelmed with data, ways to store, process, analyze, interpret, consume and act upon the data. The term Big Data is quite vague and ill defined. The word “Big” is too generic and the question is how “Big” is considered as “Big” and how “Small” is small (Smith, 2013) is relative to time, space and circumstance. The size of “Big Data” is always evolving and the meaning of Big Data Volume would lie between Terabyte (TB) and Zettabyte (ZB) range. The concept of big data is the explosion of data from the Internet, cloud, data center, mobile, Internet of things, sensors and domains that possess and process huge datasets. Cisco claimed that humans have entered the ZB era in 2015 (Cisco, 2017).

Based on social media statistics 2018, the face book claimed that, there are over 300 million photos uploaded to Facebook every day (Nowak & Spiller, 2017). On an average 300 hours of videos are uploaded every minute on You Tube (YouTube, 2017). Approximately, 42 billion texts are sent and 1.6

DOI: 10.4018/978-1-6684-3662-2.ch010

billion photos shared through Whatsapp daily (Stout, 2018). Since 2005, business investment in hardware, software, talent, and services has increased as much as 50 percent, to \$4 trillion (Rijmenam, 2018).

In 2005, Roger Mougallas from O’Reilly Media coined the term Big Data for the first time. It refers to a large set of data that is almost impossible to manage and process using traditional business intelligence tools. During the same year, Yahoo created Hadoop. This was built on top of Google’s MapReduce. Its goal was to index the entire World Wide Web (Rijmenam, 2018).

In 2009, the Indian government decides to take an iris scan, fingerprint and photograph of all of its 1.2 billion inhabitants. All this data is stored in the largest biometric database in the world (Chandra, 2018).

In 2010, at Technomy conference, Eric Schmidt stated, “There were 5 Exabyte’s of information created by the entire world between the dawn of civilization and 2003. Now that same amount is created every two days.” (Schmidt, 2010).

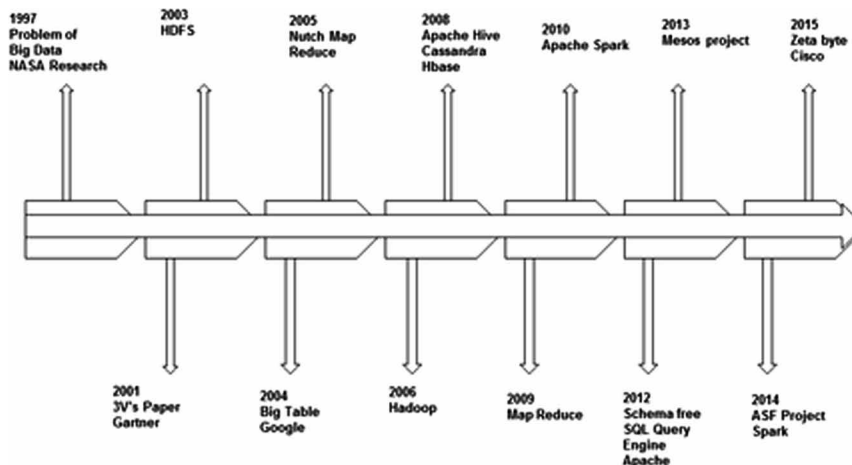
In 2011, McKinsey released a report on Big Data which claimed that, the next frontier for innovation, competition, and productivity, states that in 2018 the USA alone will face a shortage of 140.000 – 190.000 data scientist as well as 1.5 million data managers (Manyika, 2011).

Another detailed review was contributed by Visualizing.org (Hewlett Packard Enterprise, 2017) in Big Data. It is focused on the time line of how to implement Big Data Analytics. Its historical description is mainly determined by events related to the Big Data push by many internet and IT companies such as Google, YouTube, Yahoo, Facebook, Twitter and Apple. It emphasized the significant impact of Hadoop in the history of Big Data Analytics.

In the past few years, there has been a massive increase in Big Data startups, trying to deal with Big Data and helping organizations to understand Big Data and more and more companies are slowly adopting and moving towards Big Data.

Figure 1 shows the history of Big Data and its eco system.

Figure 1. History of Big Data



The data sources and their formats are continuous to grow in variety and complexity. Few list of sources includes the public web, social media, mobile applications, federal, state and local records and databases, commercial databases that aggregate individual data from a spectrum of commercial transac-

29 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-based-architectures-and-techniques/290984

Related Content

A Phenetic Approach to Selected Variants of Arabic and Aramaic Scripts

Osama A. Salman and Gábor Hosszú (2022). *International Journal of Data Analytics* (pp. 1-23).

www.irma-international.org/article/a-phenetic-approach-to-selected-variants-of-arabic-and-aramaic-scripts/297519

For Better Healthcare Mining Health Data

Güney Gürsel (2017). *Intelligent Multidimensional Data Clustering and Analysis* (pp. 135-158).

www.irma-international.org/chapter/for-better-healthcare-mining-health-data/172554

Understanding the Determinants of Big Data Analytics Adoption

Surabhi Verma and Sushil Chaurasia (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications* (pp. 1549-1577).

www.irma-international.org/chapter/understanding-the-determinants-of-big-data-analytics-adoption/291052

Commercial Banks' Digital Paradigm and Customers Responses in the UAE

Muhammad Jumaa (2020). *International Journal of Data Analytics* (pp. 68-79).

www.irma-international.org/article/commercial-banks-digital-paradigm-and-customers-responses-in-the-uae/244170

Churn Analysis Using Selected Structured Analytic Techniques

(2015). *Developing Churn Models Using Data Mining Techniques and Social Network Analysis* (pp. 164-172).

www.irma-international.org/chapter/churn-analysis-using-selected-structured-analytic-techniques/114402