Chapter 45

The Network Infrastructures for Big Data Analytics

Pethuru Raj
IBM India Pvt Ltd, India

ABSTRACT

The most delectable factor here is that the stability and maturity of networking and communication technologies enable the seamless and spontaneous interconnectivity of diverse and distributed consumer electronics, electrical, mechanical, and manufacturing devices at ground level and a bevy of services (Web, enterprise, cloud, embedded, analytical, etc.) at cyber level. Any tangible artefact and article gets connected with another to get the right and relevant empowerment, which in turn facilitates more data generation and transmission. Regulated interactions amongst digitalized entities have put a stimulating foundation for hitherto unforeseen and creative new capabilities and competencies. In short, data has grandly acquired the status of an asset not only in business organizations but also in personal lives, and hence, the data gathering, storage, and leverage tasks are fast-growing. With the data explosion happening feverishly, the discipline of big data computing and analytics has become a much-discoursed and deliberated domain of study and research. In this chapter, the authors discuss the emerging and evolving network infrastructures and architectures for big data analytics.

INTRODUCTION

In the days that have passed by, worldwide businesses have mostly focussed on productivity-related activities such as faster transactions, higher throughput, etc. with the limited set of business-centric data (transaction, operations, sales and marketing, customer, product, region, etc.). Nowadays, the quantity of data getting generated, captured, and capitalised is growing up exponentially and expediently. The data sources are also varying and many. With the Internet being established as the world's largest information superhighway and open, public, and affordable communication infrastructure, the focus is on faster delivery of data-driven insights to executives, and other decision-makers for ensuring customer delight, taking informed decisions, embarking on infrastructure optimization, developing next-generation services and applications, to close down the gap between business and IT, etc. The real challenge for organizations in the forthcoming era of knowledge will

DOI: 10.4018/978-1-4666-6539-2.ch045

be to find proven ways to better collect, analyze, monetize, and capitalize on all data heaps emanating from different and distributed sources in multiple formats. In a nutshell, big data computing represents a growing collection of robust and resilient technologies, methodologies, architectures, and tools to economically and elegantly extract value from very large volumes of a wide variety of data by enabling high-velocity capture, smart processing, and cognitive analysis towards real-time knowledge discovery and dissemination. It is expected that organizations that are best able to make real-time business decisions using derived intelligence will gain a distinct competitive advantage over those that are unable to embrace it appropriately.

Big data analytics however mandates newer software frameworks, versatile platforms, and optimized infrastructures. It requires new system designs, administrative skillsets and data management capabilities. The most pragmatic approach for most organizations is to jump-start their efforts with smaller deployments built on existing IT resources to gain the much-needed confidence and clarity. As the scope and business value of the big data discipline in transitioning enterprises to be smarter in their deeds and decisions, the importance of innovative networking technologies, topologies, and tools cannot be taken lightly. This chapter is allocated for expounding network architectures, infrastructures, platforms and practices for simplifying and streamlining big data analytics. One area where big data will have a direct impact on enterprise networks is in the area of network intelligence. The ability to automatically reconfigure the network for changing network loads and failed links (requiring zero administration for the addition of switch infrastructure) makes the network more agile. The result is a significantly reduced administration burden on the operations team, which minimizes the risk of errors and improves resiliency. The aspects on which big data differ from traditional data are shown in Table 1.

By the end of this decade(Das, Lumezanu, Zhang, Singh, Jiang, Yu, 2013), IDC expects the number of virtual and physical servers worldwide to grow by 10 times. During the same time period, the amount of data/information managed by enterprise datacenters would grow by 50 times and the number of files handled by datacenters by 75 times. These unprecedentedly higher growth rates along with increasing demands for rapid real-time processing of all kinds of data illustrate and insist for disruptive and innovative solutions. Especially the current processes, applications, platforms and infrastructures need to go through a series of empowerments in order to be right and relevant for the big data era. For example, since relational databases have limitations on the number of rows, columns and tables, the data must be spread across multiple databases using a process called sharding. Non-relational databases are being prescribed for storing, processing, mining and analyzing big data. Applications and middleware also have scale and performance limitations typically requiring a distributed architecture to support big data. Even operating systems have inherent limitations. Getting beyond 4 gigabytes of addressable RAM

Table 1. Component wise comparison between traditional data and big data

Components	Traditional Data	Big Data
Architecture	Centralised	Distributed
Data Volume	Terabytes	Petabytes, Exabytes and to Zettabytes
Data Types	Structured	Multi-Structured
Data Affinity	Known Relationships	Unknown & Complex Relationships
Data Model	Schema-centric	Schema-less