Chapter 2
NoSQL Database Phenomenon

ABSTRACT

The chapter explains that NoSQL databases emerged as an attempt for resolving the limitations of relational databases in coping with Big Data. The issue of Big Data is related to extensive requests for the storage and management of complex, dynamic, evolving, distributed, and heterogeneous data from different sources and platforms. The chapter provides an overview of the technologies, including Google File System (GFS), MapReduce, Hadoop, and Hadoop Distributed File System (HDFS), which were the first responses to Big Data challenges and main driving forces for the development of NoSQL databases. Also, the chapter asserts that NoSQL is an umbrella term related to numerous databases with different architectures and purposes, which can be classified in four basic categories: key-value, column-family, document, and graph stores. The chapter discusses the general features of NoSQL databases, as well as the specific features of each of the four basic categories of NoSQL databases.

INTRODUCTION

The growth of NoSQL databases is inseparable from the Big Data phenomenon. This is related to extensive requests for the storage and management of complex, dynamic, evolving, distributed, and heterogeneous data from different sources and platforms. In turn, all that brought relational databases to their breaking

points. As described in the first chapter, NoSQL databases emerged to resolve relational database limitations in coping with Big Data. To better understand the NoSQL database phenomenon, this chapter will review the concepts and technologies preceding its occurrence. These include: Big Data; Big Data challenges; MapReduce and a Hadoop distributed file system. It will also discuss the main features and types of NoSQL databases.

BIG DATA CHALLENGES

Since 2012, the term *Big Data* has become increasingly mainstream. However, many different (and sometimes unclear) definitions exist. According to Wu, Buyya, and Ramamohanarao (2016), Big Data definitions can be grouped based on:

- **Domains (Vs):** One of the earliest definitions was based on 3Vs: (1) volume; (2) velocity; and (3) variety (Lanely, 2001). The volume represents continuous and cumulative data growth. Velocity is the speed of data transfer from one point to another (real time data streaming, YouTube uploads, social media posts, e-mail, etc.). Variety refers to different types of data formats. IBM (2016) added a fourth domain, veracity, to refer to data uncertainty (e.g., trustworthy and quality data). Microsoft (2016) added variability and visibility, which extended the number of domains to six. Variability relates to data complexity (e.g., the huge number of data attributes). Visibility refers to the required existence of a complete data picture in the decision-making process. The number of domains (Vs) has grown to as many as 11 (Elliott, 2013).
- **Technology:** This type of definition focuses on technological support to Big Data (Hadoop, MapReduce, Spark, etc.).
- **Application:** Applications based on Big Data is the focus of this definition, including machine learning, data mining, social media analytics, etc.).
- **Signals:** Signal definitions relate to application definitions. However, signals focus on timing and finding new signal patterns in Big Data sets.

Related Content

Towards User-Oriented Enterprise Modeling for Interoperability
Kai Mertins, Thomas Knotheand Martin Zelm (2005). *Advanced Topics in Database Research, Volume 4 (pp. 130-141).*
www.irma-international.org/chapter/towards-user-oriented-enterprise-modeling/4371

The Paradoxical Effect of Social Media Usage on Developer Creativity and the Moderating Role of Openness to Experience
Yumeng Miao, Rong Duand Veda C. Storey (2021). *Journal of Database Management (pp. 29-45).*
www.irma-international.org/article/the-paradoxical-effect-of-social-media-usage-on-developer-creativity-and-the-moderating-role-of-openness-to-experience/282443

Blockchain Technology in Digitizing Bancassurance: A Theoretical Perspective of Prospects and Confronts in India
Diksha Vermaand Pooja Kansra (2022). *Applications, Challenges, and Opportunities of Blockchain Technology in Banking and Insurance (pp. 193-203).*
www.irma-international.org/chapter/blockchain-technology-in-digitizing-bancassurance/306462

Fuzzy Classification on Relational Databases
Andreas Meier, Günter Schindlerand Nicolas Werro (2008). *Handbook of Research on Fuzzy Information Processing in Databases (pp. 586-614).*
www.irma-international.org/chapter/fuzzy-classification-relational-databases/20369

Node Partitioned Data Warehouses: Experimental Evidence and Improvements
Pedro Furtado (2009). *Selected Readings on Database Technologies and Applications (pp. 490-498).*
www.irma-international.org/chapter/node-partitioned-data-warehouses/28568