Part II: Storing and Analyzing Big Data

As presented in **Part I**, the drivers behind Big Data adoption are both business- and technology-related. In the remainder of this book, the focus shifts from providing a high-level understanding of Big Data and its business implications to covering key concepts related to the two main Big Data concerns: storage and analysis.

Part II has the following structure:

• **Chapter 5** explores key concepts related to the storage of Big Data datasets. These concepts inform the reader of how Big Data storage has radically different characteristics than the relational database technology common to traditional business information systems.

• **Chapter 6** provides insights into how Big Data datasets are processed by leveraging distributed and parallel processing capabilities. This is further illustrated with an examination of the MapReduce framework, which shows how it leverages a divide-and-conquer approach to efficiently process Big Data datasets.

• **Chapter 7** expands upon the storage topic, showing how the concepts from **Chapter 5** are implemented with different flavors of NoSQL database technology. The requirements of batch and realtime processing modes are further explored from the perspective of on-disk and in-memory storage options.

• **Chapter 8** provides an introduction to a range of Big Data analysis techniques. The analysis of Big Data leverages statistical approaches for quantitative and qualitative analysis, whereas computational approaches are used for data mining and machine learning.

The technology concepts covered in **Part II** are important for business and technology leaders as well as decision-makers who will be called upon to evaluate the business case for Big Data adoption in their enterprises.