

# CHAPTER 10



## Big Data

Traditional applications of relational databases are based on structured data, and they deal with data from a single enterprise. Modern data management applications often need to deal with data that are not necessarily in relational form; further, such applications also need to deal with volumes of data that are far larger than what a single enterprise would generate. We study techniques for managing such data, often referred to as Big Data, in this chapter.

### Bibliographical Notes

[Davoudian et al. (2018)] provides a nice survey of NoSQL data stores, including data model querying and internals. More information about Apache Hadoop, including documentation on HDFS and Hadoop MapReduce, can be found on the Apache Hadoop homepage, [hadoop.apache.org](http://hadoop.apache.org). Information about Apache Spark may be found on the Spark homepage, [spark.apache.org](http://spark.apache.org). Information about the Apache Kafka streaming data platform may be found on [kafka.apache.org](http://kafka.apache.org), and details of stream processing in Apache Flink may be found on [flink.apache.org](http://flink.apache.org). Bulk Synchronous Processing was introduced in [Valiant (1990)]. A description of the Pregel system, including its support for bulk synchronous processing, may be found in [Malewicz et al. (2010)], while information about the open source equivalent, Apache Giraph, may be found on [giraph.apache.org](http://giraph.apache.org).

### Bibliography

- [Davoudian et al. (2018)] A. Davoudian, L. Chen, and M. Liu, “A Survey of NoSQL Stores”, *ACM Computing Surveys*, Volume 51, Number 2 (2018), pages 2–42.
- [Malewicz et al. (2010)] G. Malewicz, M. H. Austern, A. J. C. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski, “Pregel: a system for large-scale graph processing”, In *Proc. of the ACM SIGMOD Conf. on Management of Data* (2010), pages 135–146.

[Valiant (1990)] L. G. Valiant, “A Bridging Model for Parallel Computation”, *Communications of the ACM*, Volume 33, Number 8 (1990), pages 103–111.