

Chapter 4

How NoSQL Databases Work

ABSTRACT

The chapter explains how NoSQL databases work. Since different NoSQL databases are classified into four categories (key-value, column-family, document, and graph stores), three main features of NoSQL databases are chosen, and their practical implementation is explained using examples of one or two typical NoSQL databases from each NoSQL database category. The three chosen features are: distributed storage architecture that comprises the distributed, cluster-oriented, and horizontally scalable features; consistency model that refers to the CAP and BASE features; query execution that refers to the schemaless feature. These features are chosen because, through them, it is possible to describe most of the new and innovative approaches that NoSQL databases bring to the database world.

INTRODUCTION

The first three chapters explained the NoSQL phenomenon, the reasons for NoSQL database emergence, the main characteristics of these databases, and the types of databases. The purpose of this chapter is to explain how NoSQL databases work. Since it is beyond the scope of this book to describe NoSQL databases in detail, as there are specialized books and manuals written for that purpose, a different approach is taken here. Three main features of NoSQL

databases are chosen, and their practical implementation is explained using examples of one or two typical NoSQL databases from each NoSQL database category (see Chapter 2). The three chosen features are:

- Distributed storage architecture that comprises the distributed, cluster-oriented, and horizontally scalable features described in Chapter 2.
- Consistency model that refers to the CAP and BASE feature described in Chapter 2.
- Query execution that refers to the schemaless feature described in Chapter 2.

These features are chosen because, through them, it is possible to describe most of the new and innovative approaches that NoSQL databases bring to the database world.

DISTRIBUTED STORAGE ARCHITECTURES

Data distribution and use of clusters are natural features of NoSQL databases (see Chapter 2.4). Namely, it is precisely the ability to run NoSQL databases on a large cluster that has drawn attention to this type of database. As amounts of data kept increasing, it was more and more difficult and costly for the organization to scale up (to purchase larger and larger servers to run the database on), and the option to scale out (to run the database on a cluster of servers) was becoming increasingly attractive. As already explained (see Chapter), relational databases were developed primarily to run on a single server, so attempts to implement these databases in distributed and cluster environments led to many problems and disrupted the basic postulates of these databases.

Broadly, there are two styles for distributing data (Sadalage & Fowler, 2013):

- Sharding distributes data across multiple servers in a way that every server acts as the only source for the assigned data subset (see Chapter 1).

50 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/how-nosql-databases-work/191982

Related Content

The Impact of Ideology on the Organizational Adoption of Open Source Software

Kris Venand Jan Verelst (2008). *Journal of Database Management* (pp. 58-72). www.irma-international.org/article/impact-ideology-organizational-adoption-open/3385

A Bayesian Belief Network Methodology for Modeling Social Systems in Virtual Communities: Opportunities for Database Technologies

Ben K. Daniel, Juan-Diego Zapata-Rivera and Gordon I. McCalla (2010). *Soft Computing Applications for Database Technologies: Techniques and Issues* (pp. 125-152). www.irma-international.org/chapter/bayesian-belief-network-methodology-modeling/44386

Lock-Based Concurrency Control in Distributed Real-Time Database Systems

Özgür Ulusoy† (1993). *Journal of Database Management* (pp. 3-17). www.irma-international.org/article/lock-based-concurrency-control-distributed/51117

A Template-Based Analysis of GRL

Patrick Heymans, Germain Savaland Gautier Dallons (2006). *Advanced Topics in Database Research, Volume 5* (pp. 124-147). www.irma-international.org/chapter/template-based-analysis-grl/4390

Representation and Storage of Motion Data

Roy Gelbard and Israel Spiegler (2002). *Journal of Database Management* (pp. 46-63). www.irma-international.org/article/representation-storage-motion-data/3283