Chapter 9
A Review of RDF Storage in NoSQL Databases

Zongmin Ma
Nanjing University of Aeronautics and Astronautics, China

Li Yan
Nanjing University of Aeronautics and Astronautics, China

ABSTRACT

The Resource Description Framework (RDF) is a model for representing information resources on the Web. With the widespread acceptance of RDF as the de-facto standard recommended by W3C (World Wide Web Consortium) for the representation and exchange of information on the Web, a huge amount of RDF data is being proliferated and becoming available. So RDF data management is of increasing importance, and has attracted attentions in the database community as well as the Semantic Web community. Currently much work has been devoted to propose different solutions to store large-scale RDF data efficiently. In order to manage massive RDF data, NoSQL ("not only SQL") databases have been used for scalable RDF data store. This chapter focuses on using various NoSQL databases to store massive RDF data. An up-to-date overview of the current state of the art in RDF data storage in NoSQL databases is provided. The chapter aims at suggestions for future research.

INTRODUCTION

The Resource Description Framework (RDF) is a framework for representing information resources on the Web, which is proposed by W3C (World Wide Web Consortium) as a recommendation (Manola and Miller, 2004). RDF can represent structured and unstructured data (Duan, Kementsietsidis, Srinivas and Udrea, 2011), and more important, metadata of resources on the Web represented by RDF can be shared and exchanged among application programming without semantic missing. Here metadata mean the data that specify semantic information about data. Currently RDF has been widely accepted and has rapidly gained popularity. And many organizations, companies and enterprises have started using RDF for representing and processing their data. We can find some application examples such as the United

DOI: 10.4018/978-1-4666-9834-5.ch009

States (Data.gov), the United Kingdom (New York Times), New York Times (New York Times), BBC (BBC), and Best Buy (Chief Martec, 2009). RDF is finding increasing use in a wide range of Web data-management scenarios.

With the widespread usage of RDF in diverse application domains, a huge amount of RDF data is being proliferated and becoming available. As a result, efficient and scalable management of large-scale RDF data is of increasing importance, and has attracted attentions in the database community as well as the Semantic Web community. Currently, much work is being done in RDF data management. Some RDF data-management systems have started to emerge such as *Sesame* (Broekstra, Kampman and van Harmelen, 2002), *Jena-TDB* (Wilkinson, Sayers, Kuno and Reynolds, 2003), *Virtuoso* (Erling and Mikhailov, 2007 & 2009), *4Store* (Harris, Lamb and Shadbolt, 2009)), *BigOWLIM* (Bishop *et al*., 2011) and *Oracle Spatial and Graph with Oracle Database 12c (Oracle)*. Here BigOWLIM is renamed to OWLIM-SE and further to GraphDB. Also some research prototypes have been developed (e.g., RDF-3X (Neumann and Weikum, 2008 & 2010), SW-Store (Abadi, Marcus, Madden and Hollenbach, 2007 & 2009) and RDFox (CS Ox).

RDF data management mainly involves scalable storage and efficient queries of RDF data, in which RDF data storage provides the infrastructure for RDF data management and efficient querying of RDF data is enabled based on RDF storage. In addition, to serve a given query more effectively, it is necessary to index RDF data. Indexing of RDF data is enabled based on RDF storage also. Currently many efforts have been made to propose different solutions to store large-scale RDF data efficiently. Traditionally relational databases are applied to store RDF data and various storage structures based on relational databases have been developed. Based on the relational perspective, Sakr and Al-Naymat (2009) present an overview of relational techniques for storing and querying RDF data. It should be noted that the relational RDF stores are a kind of centralized RDF stores, which are a single-machine solution with limited scalability. The scalability of RDF data stores is essential for massive RDF data management. NoSQL (for "not only SQL") databases have recently emerged as a commonly used infrastructure for handling Big Data because of their high scalability and efficiency. Identifying that massive RDF data management merits the use of NoSQL databases, currently NoSQL databases are increasingly used in massive RDF data management (Cudre-Mauroux *et al*., 2013).

This chapter provides an up-to-date overview of the current state of the art in massive RDF data stores in NoSQL databases. We presents the survey from three main perspectives, which are key value stores of RDF data in NoSQL databases, document stores of RDF data in NoSQL databases and RDF data stores in graph databases. Note that, due to the large number of RDF data-management solutions, this chapter does not include all of them. In addition to provide a generic overview of the approaches that have been proposed to store RDF data in NoSQL databases, this chapter presents some suggestions for future research in the area of massive RDF data management with NoSQL databases.

The rest of this chapter is organized as follows. The second section presents preliminaries of RDF data model. It also introduces the main approaches for storing RDF data. The third section introduces NoSQL databases and their database models. The fourth section provides the details of the different techniques in several NoSQL-based RDF data stores. The final section concludes the chapter and provides some suggestions for possible research directions on the subject.