Chapter 8
Semantic Web Data Partitioning

Trupti Padiya
DA-IICT Gandhinagar, India

Minal Bhise
DA-IICT Gandhinagar, India

Sanjay Chaudhary
DA-IICT Gandhinagar, India

ABSTRACT

Semantic Web database is an RDF database due to increased use of Semantic Web in real life applications; one can find heavy growth in RDF database. As there is a tremendous increase in RDF data, performance and scalability issues are of main concern. This chapter discusses improving and scaling up query performance for increasingly growing Semantic Web. It discusses current Semantic Web data storage techniques, which have been found to scale poorly and have poor query performance. It discusses the partitioning techniques vertical and horizontal partitioning to improve query performance. To further improve the query performance, along with these partitioning techniques, various compression techniques can also be used. Relational data offers faster execution of queries as compared to RDF data. To demonstrate these ideas, semantic data is converted to relational data and then query performance improvement techniques are applied. The scaling up of Semantic Web data is also discussed.

INTRODUCTION

The Semantic Web is an effort by the W3C to enable integration and sharing of data across different applications and organizations. It is a contemplation of combination of different content, information applications, and systems across Internet. There is a tremendous growth in semantic Web data due to its application in real world. Application which relies on semantics in real life has perceived the importance of RDF (Resource Description Framework) data that is spread across and as a result we can find a rapid growth in semantic Web data. Mostly all semantic applications require fetching the data efficiently which is spread over Internet and hence different Semantic Web applications will often retrieve information dynamically from URLs and merge

them into a storage system to make the data available. Therefore, data manipulation has to be carried out many times, on the scale of the Web, and it should be done in efficient manner. Researchers are investigating usefulness of RDBMS tools for this purpose. Efficient management of RDF data is an important factor in realizing the semantic Web vision. There is need for efficient query processing as performance and scalability issues are becoming increasingly important as semantic Web technology is applied to real-world applications.

ISSUES IN ACCESSING SEMANTIC WEB DATA

The Semantic Web data, which is in RDF format <Subject, Property, Object> known as an RDF triple can be stored in database table format. To retrieve data efficiently we can store data tables in two ways: row-by-row or column-by-column. The former approach will keep all information about an entity together. For example in a vendor table, it will store all information about the first vendor and then all information about the second vendor and so on. The later approach will keep all attribute information together: the entire vendor names will be stored consecutively, then the entire vendor addresses and so on. Now as these designs are even handed the choice is based on performance expectations. If it is expected the result should be based the granularity of an entity for, e.g., find a vendor, add a vendor, delete a vendor, etc., then the row-by-row storage is preferable as all of the required information will be stored together. On the other hand, if the expected result tends to read per query only a few attributes from many records will be the result, for, e.g., a query that finds the most common e-mail address domain, then column-by-column storage is preferred as other attributes which are not required for a particular query need not have to be accessed. This is what is said to be partitioning or division

of a logical database or its constituting elements into distinct independent parts, which is done for manageability, performance or availability. These are our basic requirements in order to scale up Semantic Web data.

Scalability

As the Semantic Web evolves, scalability becomes increasingly important. Triples are used for resource description on the Web. All resources are described with many triples and this makes a complex graph of relationships, which includes references to other resources and the relationships in semantic Web data also have types. It is therefore necessary that triple stores can deal with large numbers of triples in real life. To fetch and store triples also impacts scalability, it brings in the issues of timeliness, caching and other general problems from a distributed system, which can fail or be delayed.

Performance

Manipulation of data in a triple store is required as triples can be added, modified, and/or removed. This needs support for identifying triples, querying the graph as well as administering the graph— i.e. creating and deleting it and other operations for transferring data to and from the network. Indexes should be created for efficient searching, or for providing specialized searches such as text-based search, or other searches based on properties, sub-properties, logical inference, etc. It is always possible that data from multiple sources will be merged into single graphs of data, and the relationships between them will connect them up and this needs triple-store support in order to handle such merging when the graphs are large enough and also should be possibly separated again later. Such manipulations by applications using semantic Web data will be occurring many times for even simple systems and thus these need

Related Content

Ontology Management for Large-Scale Enterprise Systems

Juhnyoung Lee, Richard Goodwinand Rama Akkiraju (2006). *Web Semantics & Ontology (pp. 91-114).*

www.irma-international.org/chapter/ontology-management-large-scale-enterprise/31199

Mobile Service Oriented Architecture (MSOA) for Businesses in the Web 2.0 Era

Ming-Chien Wuand Bhuvan Unhelkar (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications (pp. 178-191).*

www.irma-international.org/chapter/mobile-service-oriented-architecture-msoa/39170

Overview of MERA: An Architecture to Perform Record Linkage in Music-Related Databases

Daniel Fernández-Álvarez, José Emilio Labra Gayo, Daniel Gayo-Avelloand Patricia Ordoñez de Pablos (2019). *Semantic Web Science and Real-World Applications (pp. 219-245).*

www.irma-international.org/chapter/overview-of-mera/215067

Semantic Decision Internal-Attention Graph Convolutional Network for End-to-End Emotion-Cause Pair Extraction

Dianyuan Zhang, Zhenfang Zhu, Jiangtao Qi, Guangyuan Zhangand Linghui Zhong (2023). *International Journal on Semantic Web and Information Systems (pp. 1-21).*

www.irma-international.org/article/semantic-decision-internal-attention-graph-convolutional-network-for-end-to-end-emotion-cause-pair-extraction/325063

Linked Data: The Story so Far

Christian Bizer, Tom Heathand Tim Berners-Lee (2011). *Semantic Services, Interoperability and Web Applications: Emerging Concepts (pp. 205-227).*

www.irma-international.org/chapter/linked-data-story-far/55046