RDF Query Path Optimization Using Hybrid Genetic Algorithms:
Semantic Web vs. Data-Intensive Cloud Computing

Qazi Mudassar Ilyas, King Faisal University, Saudi Arabia
https://orcid.org/0000-0003-4238-8093

Muneer Ahmad, University of Malaya, Malaysia

Sonia Rauf, COMSATS University Islamabad, Abbottabad, Pakistan

Danish Irfan, COMSATS University Islamabad, Abbottabad, Pakistan

ABSTRACT

Resource description framework (RDF) inherently supports data mergers from various resources into a single federated graph that can become very large even for an application of modest size. This results in severe performance degradation in the execution of RDF queries. As every RDF query essentially traverses a graph to find the output of the query, an efficient path traversal reduces the execution time of RDF queries. Hence, query path optimization is required to reduce the execution time as well as the cost of a query. Query path optimization is an NP-hard problem that cannot be solved in polynomial time. Genetic algorithms have proven to be very useful in optimization problems. The authors propose a hybrid genetic algorithm for query path optimization. The proposed algorithm selects an initial population using iterative improvement, thus reducing the initial solution space for the genetic algorithm. The proposed algorithm makes significant improvements in the overall performance. They show that the overall number of joins for complex queries is reduced considerably, resulting in reduced cost.

KEYWORDS

Cloud Computing, Genetic Algorithm, Information Retrieval, Query Path Optimization, Resource Description Framework, SPARQL

INTRODUCTION

Cloud computing is a relatively new paradigm that provides extensive services to various customers (Ziebell et al., 2019). The data-intensive application running on Cloud may benefit from the machine-understandable Semantic Web technologies (Hossain et al., 2019).

The semantic web technologies have recently gained attention towards proposing viable solutions to Cloud-computing related problems (Elzein et al., 2018). The machine-understandable representation of information has opened new paradigms (Olakanmi & Dada, 2019). This scenario has opened new potentials to researchers and scientists in managing large amounts of data in the best possible and

DOI: 10.4018/IJCAC.2022010101

available ways using the recent archive (Ali, 2019), process and execution mechanism (Bellini et al., 2015; Lee et al., 2013; Silva et al., 2013).

The current challenges related to big data and its emergence have complicated efficient data management due to the exponential growth of data (Acosta et al., 2017; Siow et al., 2017; X. Wang et al., 2015). The current Cloud resources seem insufficient to manage large data repositories and extract knowledge from them (Herzfeldt et al., 2019). Although companies are providing excellent services in terms of data archive, application deployments, and fact findings from existing data, still keeping in mind the current explosion of data, we need more robust and resilient solutions towards better management of Cloud data (Destefano et al., 2016; Wu et al., 2019). Semantic Web technologies provide one possible solution to the problem, Several researchers have used these technologies in solving similar problems (Fang et al., 2016; Niknia & Mirtaheri, 2015; Srinivasulu et al., 2015).

The cyber physical system concepts have introduced new platforms in the form of the industry 4.0 revolution (S. Wang et al., 2016). The current need to integrate Cloud capabilities to align with the industry 4.0 standards is practically realized (AlZu'bi et al., 2020)(Tewari & Gupta, 2020)(H. Wang et al., 2020)(D. Li et al., 2019)(Bhushan & Gupta, 2019). Besides, the intrusion of big data generated from industry-related objects has resulted in new challenges for researchers to devise robust query control and management mechanisms (Liao et al., 2016; Samanthula et al., 2015; Verginadis et al., 2017). Additionally, the smart cities concept enables the smart devices to generate many queries related to each smart city notion. At times, we need to prioritize the queries that demand immediate attention; for instance, the queries related to fire, temperate, explosive detection and natural disaster alarms, etc. Such high priority queries are mixed with low priority queries generated from other objects of smart cities (Liu et al., 2016; Rady et al., 2019; Ye et al., 2018). Since the smart devices are being controlled through network resources and being managed by Cloud applications, the semantic web technologies may offer great help in identifying and addressing the potential queries from low priority queries (Kaoutar et al., 2018; Niknia & Mirtaheri, 2015; Srinivasulu et al., 2015).

Resource Description Framework (RDF) is a simple data model. Its basic building block is a subject-predicate-object triple, called a statement. A statement consists of a resource, property, and value. Values can either be resources or string literals. These statements can be visualized using an RDF graph, also called a semantic network. It is a directed graph with labeled nodes and arcs; the arcs are produced from the resource (subject of the statement) to the value (object of the statement) and known as predicates.

RDF query is used to retrieve and manipulate data stored in RDF format. It can be visualized as a tree where leaf nodes represent inputs (sources), and internal nodes specify relational algebra operations, enabling a user to specify basic retrieval requests on these sources (Stuckenschmidt et al., 2005). To retrieve the required data, the nodes can be ordered in numerous ways, all producing the same result. The different orders in which these possible solutions for retrieval of data needed are executed are known as query plans or query paths. RDF query is solved in the same manner as queries in relational databases, i.e., by dividing the Query into several sub-queries. At the end, results of all sub-queries are combined with the help of join operators and are referred to as chain queries. Most of the time, bushy and right deep query trees are given preferences in an RDF data store (Steinbrunn et al., 1997). The inner relation of each join is a base relation. In bushy trees, base relations can be joined with the results of earlier joins. Right-deep trees, which are a subset of bushy trees, require the left-hand join operands to be base relations. Figures 1 and 2 show an example depicting concepts (R_1, R_2, R_3, R_4, R_5, R_6, R_7) by bushy and right-deep trees.

Query optimization is the process of identifying the access plan with the minimum cost (time taken to get all the answers).

The cost of a query depends on the order of joins of sub-paths. Optimal order of joins will give minimum price by reducing the overall response time.

A query is divided into sub-parts, whose results are then combined with the help of join operators. Join ordering is an NP-hard problem, which cannot be solved with traditional algorithms' support

Related Content

A Framework for Assessing Governmental Websites Quality: The Case of Iranian Free Economic Zones Websites
Mohammad Amin Zareand Ahad Zare Ravasan (2014). *International Journal of E-Services and Mobile Applications (pp. 44-65).*
www.irma-international.org/article/a-framework-for-assessing-governmental-websites-quality/111065

Location-Based Services (LBS) in Micro-Scale Navigation: Shortcomings and Recommendations
Suleiman Almasri, Muhammad Alnabhan, Ziad Hunaitiand Eliamani Sedoyeka (2009). *International Journal of E-Services and Mobile Applications (pp. 51-71).*
www.irma-international.org/article/location-based-services-lbs-micro/37447

An Adaptable Approach to Fault Tolerance in Cloud Computing
Priti Kumariand Parmeet Kaur (2023). *International Journal of Cloud Applications and Computing (pp. 1-24).*
www.irma-international.org/article/an-adaptable-approach-to-fault-tolerance-in-cloud-computing/319032

SCEF: A Model for Prevention of DDoS Attacks From the Cloud
Ganeshayya Ishwarayya Shidaganti, Amogh Shreedhar Inamdar, Sindhuja V. Raiand Anagha M. Rajeev (2020). *International Journal of Cloud Applications and Computing (pp. 67-80).*
www.irma-international.org/article/scef-a-model-for-prevention-of-ddos-attacks-from-the-cloud/256865

Key Challenges for Delivering Web-Based IT Support Services: The Provider Perspective
Vanessa Cooper, Sharman Lichtensteinand Ross Smith (2011). *Developing Technologies in E-Services, Self-Services, and Mobile Communication: New Concepts (pp. 30-50).*
www.irma-international.org/chapter/key-challenges-delivering-web-based/54956