Chapter 6

Data Replication Impact on DDBS System Performance

Ali A. Amer
Taiz University, Yemen

ABSTRACT

In distributed database systems (DDBS), the utmost purpose of data distribution and replication aims at shrinking transmission costs (TC), including communication costs, and response time. In this chapter, therefore, an enhanced heuristic clustering-based technique for data fragmentation and replicated based allocation is efficaciously presented. This work is mainly sought to further enhance an existing technique so TC is to be significantly minimized. In fact, the approached enhancement is applied by suggesting different replication scenarios. Of these scenarios, one scenario is to be selected based on competitive performance evaluation process. DDBS performance is measured via its being exposed on objective function (TC). Despite the fact that this work is mildly improved, yet evaluation results show that it has been promising, particularly as TC being the foremost design objective of DDBS System. Experimental results have been analyzed under all presented scenarios as an internal evaluation and are vividly provided to demonstrate the undeniable impact of data replication on DDBS performance.

INTRODUCTION

In most DDBS and cloud-based DDBS applications and services, the storing and retrieving of data are integral essential activities of nature of application at hand. As a matter of fact, with continuous access and retrieval for data, the design of DDBS (that some application uses) can have a great impact (positively/negatively) on DDBS metrics including performance, throughput, and even scalability and availability of DDBS system as a whole. Therefore, the necessity for an effective partitioning technique that is able to find balance between DDBS metrics while applied efficiently in large-scale DDBS is still the hot spot in DDBS research community.

In fact, partitioning data into separate partitions is set to make data more easily to be managed for access and retrieval at the same time. However, the partitioning technique has to be designed critically to maximize the utilities of data being partitioned while minimizing adverse effects that occasionally comes along data partitioning. On the other hand, it is widely known, in literature that data partitioning along with data replication are bound to help enhance DDBS scalability, reduce Transmission Costs (TC),
and promoting performance of system in total (Adel et al., 2017). Nevertheless, it is worth indicating that combining a careful data partitioning and intelligently-designed data replication in one single work can bring several other advantages as follows:

- **To Enhance DDBS Scalability**: As database system is being scaled up, it is set to finally reach a physical hardware limit in terms of capacity, as instance. However, when data is being scattered and replicated across multiple partitions, and each partition allocated to a single site, hardware limitations can be wisely exploited to great extent.

- **To Enhance Performance**: When data access is being frequently accessed as a whole for little information needed, it can be hold a significant bearing on DDBS performance. However, when data is partitioned and properly replicated over network sites, DDBS system is meant to be highly effective. Moreover, parallel access can be leveraged chiefly for activities that reach more than one partition. To minimize latency and TC while maximize response time, each partition must be allocated near the application (or in the site) that uses it greatly frequently.

- **To Enhance Availability**: As a matter of fact, dividing data and placing partitions into multiple sites (data replication is adopted) seeks to circumvent a single point of failure that often might be happening. In the sense that If one site fails, or site is under maintenance, other site(s) is set to have other copy of the same required data (unlike the scenario of non-replication as in this case the data in that partition would surely be unavailable for users’ operations/activities). So, to reduce the possibility of site(s) being unable to provide data (getting into deadlock), the number of partitions have to be cautiously increased (replicated) as needed.

Another benefit can be observed in terms of: enhancing security specifically for data classified as sensitive; maximizing administrative efficiency in terms of data management and monitoring. Finally, data partitioning and replication helps distribute the load over many sites (instead of single one site), which consequently reduces TC and improves performance at the same time. Giving reader(s) a flavour of how all these factors (chiefly data replication) could affect DDBS performance would be highly appreciated, though.

In this chapter, thus, we are trying to expose reader(s) on DDBS design in terms of vertical partitioning of data and replication-based data allocation as well, while listing the come-along benefits which is already drawn above. In section (2), the state of art is being briefly yet usefully discussed. The enhanced work’s methodology for (Adel et al. 2017), with the major aim of significantly reducing TC to most great extent, including heuristics and fragmentation and allocation costs models, is introduced in sections (3, 4, and 6)). In section (7), experimental results have been presented in detail. Brief yet clearly-drawn discussion for results is given in section (8) showing the highly positive impact of data replication, in terms of TC minimization as all data allocation scenarios considered. Finally, conclusions of this chapter are provided in section (9).

**Related Work**

It is a rule of thumb in DDBS that the better the DDBS design is, the higher DDBS performance is to be guaranteed. However, for DDBS design issue the most effective strategies recorded in literature were the data partitioning (Vertical, Horizontal or Mixed) and data allocation, chiefly if replication is adopted intelligently, as discussed in Nashat and Amer, 2018. In our work of (Nashat and Amer, 2018), a
Related Content

Harnessing Semantic Features for Large-Scale Content-Based Hashtag Recommendations on Microblogging Platforms
[www.igi-global.com/article/harnessing-semantic-features-for-large-scale-content-based-hashtag-recommendations-on-microblogging-platforms/172423?camid=4v1a](www.igi-global.com/article/harnessing-semantic-features-for-large-scale-content-based-hashtag-recommendations-on-microblogging-platforms/172423?camid=4v1a)

Towards Process Mediation in Semantic Service Oriented Architecture (SSOA)
[www.igi-global.com/chapter/towards-process-mediation-semantic-service/35757?camid=4v1a](www.igi-global.com/chapter/towards-process-mediation-semantic-service/35757?camid=4v1a)

A Multi-Feature Based Automatic Approach to Geospatial Record Linking
Ying Zhang, Puhai Yang, Chaopeng Li, Gengrui Zhang, Cheng Wang, Hui He, Xiang Hu and Zhitao Guan (2018). *International Journal on Semantic Web and Information Systems* (pp. 73-91).
[www.igi-global.com/article/a-multi-feature-based-automatic-approach-to-geospatial-record-linking/210653?camid=4v1a](www.igi-global.com/article/a-multi-feature-based-automatic-approach-to-geospatial-record-linking/210653?camid=4v1a)

Semantic-Based Bluetooth-RFID Interaction for Advanced Resource Discovery in Pervasive Contexts
[www.igi-global.com/article/semantic-based-bluetooth-rfid-interaction/2846?camid=4v1a](www.igi-global.com/article/semantic-based-bluetooth-rfid-interaction/2846?camid=4v1a)