

Chapter 7

Shadow Sensitive SWIFT: A Commit Protocol for Advanced Data Warehouses

Udai Shanker

Madan Mohan Malviya Engineering College, India

Abhay N. Singh

Madan Mohan Malviya Engineering College, India

Abhinav Anand

Madan Mohan Malviya Engineering College, India

Saurabh Agrawal

Madan Mohan Malviya Engineering College,, India

ABSTRACT

This chapter proposes Shadow Sensitive SWIFT commit protocol for Distributed Real Time Database Systems (DRTDBS), where only abort dependent cohort having deadline beyond a specific value ($T_{shadow_creation_time}$) can forks off a replica of itself called a shadow, whenever it borrows dirty value of a data item. The new dependencies Commit-on-Termination external dependency between final commit operations of lender and shadow of its borrower and Begin-on-Abort internal dependency between shadow of borrower and borrower itself are defined. If there is serious problem in commitment of lender, execution of borrower is started with its shadow by sending YES-VOTE message piggy bagged with the new result to its coordinator after aborting it and abort dependency created between lender and borrower due to update-read conflict is reversed to commit dependency between shadow and lender with read-update conflict and commit operation governed by Commit-on-Termination dependency. The performance of Shadow Sensitive SWIFT is compared with shadow PROMPT, SWIFT and DSS-SWIFT commit protocols (Haritsa, Ramamritham, & Gupta, 2000; Shanker, Misra, & Sarje, 2006; Shanker, Misra, Sarje, & Shisondia, 2006) for both main memory resident and disk resident databases with and without communication delay. Simulation results show that the proposed protocol improves the system performance up to 5% as transaction miss percentage.

DOI: 10.4018/978-1-60960-067-9.ch007

INTRODUCTION

Database systems are currently being used as backbone to thousands of applications. Some of these have very high demands for high availability and fast real-time responses. Typically, these systems generate a very large transaction workload against the distributed real time database, and a large part of the workload consists of read, write and update transactions. Unavailability of real time or slow response in processing these transactions used by business applications could, however, be financially devastating and, in worst case, cause injuries or deaths. Examples include telecommunication systems, trading systems, online gaming, sensor networks etc. Typically, a sensor network consists of a number of sensors (both wired and wireless) which report on the status of some real-world conditions. The conditions include sound, motion, temperature, pressure & moisture, velocity etc. The sensors send their data to a central system that makes decisions based on both present and past inputs. To enable the networks to make better decisions, both the number of sensors and the frequency of updates should be increased. Thus, sensor networks must be able to tolerate an increasing load. For applications such as health care in a hospital, automatic car driving systems, space shuttle control, etc., data is needed in real-time and must be extremely reliable as any unavailability or extra delay could result in loss of human lives (Huang, 1991).

Recent years have seen increasing interest in providing support for warehouse-like systems that support fine-granularity insertions of new data and even occasional updates of incorrect or missing historical data; these modifications need to be supported concurrently with traditional updates. Such systems are useful for providing flexible load support in traditional warehouse settings for reducing the delay for real-time data visibility and for supporting other specialized domains such as Customer Relationship Management (CRM) and data mining where there is a large quantity

of data that is frequently added to the database in addition to a substantial number of read-only analytical queries to generate reports and to mine relationships. These “updatable warehouses” have the same requirements of high availability and disaster recovery as traditional warehouses but also require some form of concurrency control, commit protocol and recovery to ensure transactional semantics.

Data mining is the art and science of extracting hidden patterns from the accumulated data for decision-making. It has emerged as a valuable decision support tool with the recognition that:

1. Data Mining and advanced statistical techniques provide insights into data that mere slicing and dicing does not.
2. The human mind’s ability to handle complexity is limited.
3. The advances in computing have made the cost of storing and processing data very affordable.

The three essential requisites of good data mining initiatives are:

1. Domain expertise in the area of business
2. Extensive knowledge of data mining tools, advanced statistics and modeling expertise
3. A data mining vision that includes willingness to commit time and other resources.

Many applications listed above using DRT-DBS require distributed transaction executed at more than one site. Traditional log-based systems require sites force-write log records to disk at various stages of commit processing in order to ensure atomicity. A commit protocol ensures that either all the effects of the transaction persist or none of them persist despite the failure of site or communication link and loss of messages. The Commit processing should add as little overhead as possible to transaction processing. Therefore,

19 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/shadow-sensitive-swift/46894

Related Content

Human Capital and External Knowledge Acquisition in Nonprofit Organizations: Facilitating Strategic Advantage during an Economic Crisis

Eric Kong (2015). *Knowledge Management for Competitive Advantage During Economic Crisis* (pp. 82-99).

www.irma-international.org/chapter/human-capital-and-external-knowledge-acquisition-in-nonprofit-organizations/117844

Opinion Mining with SentiWordNet

Bruno Ohana and Brendan Tierney (2011). *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains* (pp. 266-286).

www.irma-international.org/chapter/opinion-mining-sentiwordnet/46900

Federated Data Warehouses

Stefan Berger and Michael Schrefl (2010). *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications* (pp. 82-107).

www.irma-international.org/chapter/federated-data-warehouses/39589

Optimizing and Managing Digital Telecommunication Systems Using Data Mining and Knowledge Discovery Approaches

Adnan I. Al Rabea and Ibrahim M. M. El Emary (2011). *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains* (pp. 66-84).

www.irma-international.org/chapter/optimizing-managing-digital-telecommunication-systems/46891

Named Entity Recognition for Ontology Population using Background Knowledge from Wikipedia

Ziqi Zhang and Fabio Ciravegna (2011). *Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances* (pp. 79-104).

www.irma-international.org/chapter/named-entity-recognition-ontology-population/53882