Abstract

Data replications and transaction deadlocks can severely affect the performance of distributed database systems. Many current evaluation techniques ignore these aspects, because it is difficult to evaluate through analysis and time-consuming to evaluate through simulation. In this paper, we use a technique that combines simulation and analysis to closely illustrate the impact of deadlock and evaluate performance of replicated distributed database with both shared and exclusive locks.

1. Introduction.

A distributed database system (DDS) is a collection of cooperating nodes each containing a set of data items. A user transaction can access such a system at any of these nodes. The receiving node, often referred to as the coordinating node, undertakes the task of locating the nodes that contain the data items required by a transaction.

In order to maintain database consistency and correctness in the presence of concurrent transactions, several concurrency control protocols have been proposed [1]. Of these, the most commonly used are time-stamping and locking protocols. Locking protocols have been widely used in both commercial and research environments. In static locking, prior to start of execution, a transaction needs to acquire either a shared-lock (for read operations) or an exclusive lock (for update operations) on each of the relevant data items.

Data replication is used to improve the performance of local transactions and the availability of databases. In replicated databases, one data item may have more than one copy in the system. Replica control algorithms are used to maintain the consistency among these copies. One of these is the read-one/write-all protocol. With this protocol, an exclusive lock need to acquire an exclusive lock from every copy of the data item. For a shared lock to succeed, any one copy of the data item has to be shared locked. When transactions with conflicting lock requests are initiated concurrently, they could be possibly blocked due to a deadlock.

There are two major ways to evaluate the performance of distributed systems: simulation and analysis. Simulation is a conceptually tractable technique, but requires large computational time. On the other hand, analysis is computationally faster but may not be tractable for all problems. In [4], Shyu and Li proposed an elegant analysis model to evaluate the response time and throughput of transactions in a non-replicated DDS. Assuming exclusive locking (i.e., only write operations), they model the queue of lock requests at an object as an M/M/1 queue [3]. This results in a closed-form for the waiting time distribution at a node, expressed in terms of the average rates of arrivals of requests and the average lock-holding time. With shared lock and replications added into the picture, it is very difficult to have a close model for it. Because of the limitations of simulation and analysis, we develop a technique that combines simulation and analysis.

This paper is organized as follows. In Section 2, we describe the model used in our performance evaluation. In Section 3, we propose an evaluation technique. In Section 4, we illustrate the results. Finally, Section 5 has the conclusions.

2. Model

Our model has the following parameters:

- There are \( n \) nodes.
- There are \( d \) data items in a DDS.
- A data item may be located at exactly \( c \) number of nodes. The \( dc \) data copies are uniformly distributed across the \( n \) nodes.
- Each transaction accesses \( k \) data items.
- \( r \) is the read ratio. So among \( k \) data items to be accessed, \( rk \) are accessed only for read operations, and the rest are for read-write operations. Due to the read-one/write-all replica control policy, a transaction must procure \( rk \) shared locks for \( rk \) read operations and \( (1 - r)kc \) exclusive locks for the \( (1 - r)k \) read-write operations.
- Each data item is equally likely to be accessed by a transaction.
- Transaction arrivals into the system is a Poisson process with rate \( \lambda \).
- The communication delay between any two nodes is exponentially distributed with mean \( \tau \).
- The average execution time of a transaction, once the locks are obtained, is \( \delta \).
- The deadlock mechanism is invoked every \( T \) seconds.
- After an abortion of a transaction, it takes an average of \( w \) seconds for this transaction to be restarted.
- \( \mu \) is the service rate of transactions.
- \( b \) is the lock-holding time.
- \( \lambda c \) is the arrival rate at each data copy.
3. Performance Evaluation Technique

Our technique consists of two stages. In the first stage, the average transaction response time and throughput are calculated by ignoring the deadlock. This is an iterative step involving simulation and analysis. In the second stage, the probabilities of transaction conflicts and deadlocks are computed by probability models. These probabilities are used, in turn, to compute the response time and throughput in the presence of deadlocks.

Stage 1:
Initially, we assume that there are no lock conflicts between transactions. Each transaction has to procure $r_k$ shared locks on data copies and $(1 - r)k$ exclusive locks on data copies. When a transaction has got all the lock grants from these data objects, it can go ahead with execution.

This procedure is summarized in the following 6 steps.

1. Initialize lock-holding time($b$) to be $1/\mu$.

2. Given the total rate of arrival($\lambda$), the shared lock ratio($r$), the number of data items($d$), the number of data items required by each transaction($k$) and the number of replications($c$), derive the arrival rate at each data copy($\lambda c$).

3. With the arrival rate at each data copy($\lambda c$), the average lock-holding time($b$), and the transmission time($t$) we can simulate the queue at a data copy to arrive wait-time($w$) distribution. With this distribution we can calculate the response time of transactions.

4. With the average service time of transactions($1/\mu$), and the transmission time, we can derive a new lock-holding time($b'$).

5. Set $b$ to this new lock-holding time $b'$.

6. If the old and new lock-holding time are sufficiently close, stop the iteration. Otherwise, go back to step 3.

At the end of stage 1 the response time without the consideration of transaction deadlocks is obtained.

Stage 2:
This stage considers transaction conflicts and computes the deadlock probability. Here the probabilities of transaction deadlock and restart are computed. These are then used to compute response time and throughput in the presence of deadlocks.

Assume there are two transactions $T_1$ and $T_2$. Let $R_S$, $W_S$ be the read and write sets of transactions respectively.

1. Let $f_{S_i}$ be the probability that the read set of $T_1$ has $i$ data items overlapping with the writeset of $T_2$, i.e. $|R_S(T_1) \cap W_S(T_2)| = i$.

2. Let $f_{T_j}$ be the probability that given $|R_S(T_1) \cap W_S(T_2)| = i$, the writeset of $T_1$ has $j$ data items overlapping with the readset and write-set of $T_2$, i.e. the probability that $|W_S(T_1) \cap (R_S(T_2) \cup W_S(T_2))| = j$.

Clearly,

$$f_{S_i} = \frac{(r_k - k)^i (d-k+r+k)^{r_k-i}}{d^{r_k}}$$

$$f_{T_j} = \frac{(r_k - k)^{i/r_k} (d-k+r+k)^{r_k-i/r_k}}{d^{r_k}}$$

It can also be noted that $f_{S_i} f_{T_j}$ is the probability that $\bigl(\text{Read-set(T1)} \cap \text{Write-set(T2)}\bigr) = i$ \& $\bigl(\text{Write-set(T1)} \cap (\text{Write-set(T2)} \cup \text{Read-set(T2)})\bigr) = j$.

If $P_{W_{ij}}$ is the probability that $T_1$ waits for $T_2$,

$$P_{W_{ij}} = p_1 + p_2 - p_1 \cdot p_2$$

Where $p_1$ is the probability that $T_1$ waits for $T_2$ for shared locks in readset and $p_2$ is the probability that $T_1$ waits for $T_2$ for exclusive locks in writeset.

Probability that $T_1$ waits for $T_2$ is now given by

$$P_w = \sum_{i=0}^{\text{max}(d-k+r, 0)} \sum_{j=0}^{\text{max}(r_k-i, 0)} f_{S_i} f_{T_j} P_{W_{ij}}$$

With this probability of waiting and the formulas in [4] we can calculate the probability of a transaction deadlock, the probability of a transaction restart and the probability of a transaction to be blocked by other transactions. And with these probabilities and the time between deadlock detection($r$), we can calculate the response time with consideration of deadlock. (Details are omitted here.)

4. Results

Using this technique, we obtained a number of interesting results that illustrate the effect of deadlocks and number of replications on database performance. These are summarized in Figures 1-5. We make the following observations.

- Transaction response times are quite sensitive to the ratio of shared locks (Figure 1 and 2). Here, we compare the response times when deadlocks are ignored (DI, computed in Stage 1) with those obtained when deadlocks are considered (DC, computed in Stage 2). The effect of deadlocks is more predominant at higher transaction loads and with smaller values of $r$. When $r = 2/3$, the effect of deadlocks is not significant on response time.

- If we compare Figure 1 and 2 with Figure 3 and 4, it can be observed that the increase in replications results in the larger response time when read ratio is smaller than 1/3.

- Fig. 5 shows the response times with different replication numbers. Here we can see that with both cases when read ratio is 2/3 and 1/3, the response time increases as the number of replications increases. But with read ratio equals 1/3, the increasing rate is much smaller than that with read ratio equals 2/3.

5. Conclusions

In [4], Shyu and Li presented an elegant technique to evaluate the performance of distributed database systems in the presence of deadlocks. Their technique assumed only exclusive locks and thus representing the worst-case effects of deadlocks.
In this paper, we have extended their technique to combine simulation and analysis. And with this extended technique we allow both shared and exclusive locking and also replications in our model. We evaluated the effect of number of data items, the number of data items accessed by each transaction, the ratio of read operations on transaction response time and the number of replications. These results show the importance of considering both shared and exclusive lock requests, the deadlock probabilities as well as the number of replications of database for response time evaluations.
References


