Replication Strategies for High Availability and Disaster Recovery

Robert Breton
Director
Sybase Replication Server Engineering

1 The Business Problem

Businesses are faced with the critical need to ensure the availability of information and continuous operation of their mission-critical systems in the face of potential failures ranging from hardware losses such as disk crashes and CPU failures to catastrophic losses to their computing facilities and communications networks. For example, the Federal Reserve required banks by 1998 to have appropriate disaster recovery procedures in place that would ensure that business interruptions would not exceed four (4) hours. Most of today's disaster recovery strategies are not well suited to meeting such stringent business requirements without suffering significant loss of business information.

While solutions exist to provide tolerance to component failure, the issue of site loss is often more challenging, with potentially dire consequences due to both business interruption and loss of information. While remote backup storage and disaster recovery sites have traditionally satisfied the requirements for disaster recovery for batch-oriented business systems, they are generally inadequate for protecting the information in on-line transaction processing (OLTP) systems. With the introduction of database replication, continuous duplication of critical OLTP information to off-site backup facilities without the high latency inherent in tape backup approaches is now available. Once established, such an environment can be automated to ensure that information is replicated in a timely manner and the switch to backup systems is accomplished with minimal business interruption.

This paper will look at the evolution of such a replication system, Sybase's Replication Server, in supporting disaster recovery and high availability business requirements.

2 Replication Architectures

While the concept of a single federated database offers the advantage of centralized control, the reliance on a central site often requires compromises in accessibility, performance and local autonomy over the business information as dependency on the network and central site availability and scalability grows. In decentralized businesses, the proliferation of distributed computing platforms often means a lack of consistent, timely consolidated information at the corporate level. While the technology exists to support real-time concurrent updates to multiple systems, the performance impact and availability risks do not often justify its use except in the most critical business transactions.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

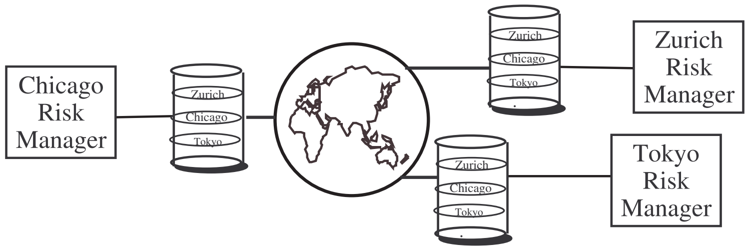

Figure 1: Peer-to-peer replication.

The widespread adoption of clientserver technologies has driven the adoption of more distributed systems architectures. These distributed systems introduce a number of management challenges requiring new strategies for ensuring accurate and timely availability of business information. As the same information is often required in diverse business locations, asynchronous distribution and synchronization satisfies the need to minimize the performance and response time impact of capturing and distributing replicated data while ensuring reliable, timely delivery to replicate sites. This deferred delivery model ensures that the originating application does not wait for the replicated data to be delivered, and any interruption in delivery to one location does not interfere with deliver to other business locations. This provides a high level of information currency, availability and performance, while lowering the cost of delivery. By supporting customization of information as it is delivered, business information delivery can often be tailored to the needs of the each business location.

The following example shows a typical peer-to-peer replication architecture enabling each major business center to manage local information while acting as a backup site to the other locations (see Figure 1). This architecture ensures $7x24$ availability of information throughout the world by eliminating reliance on any single site for information access, while eliminating the dependence on network availability by employing an asynchronous database transaction replication facility.

3 Replication Server: the Publish and Subscribe Model

In 1993, Sybase introduced the first generation of technology providing asynchronous, reliable distribution of database changes. Based on a publish and subscribe model, Replication Server utilizes a highly efficient ”push” architecture, capturing database transactions directly from database recovery logs in near real-time, distributing to subscribing databases through a multi-tiered routing scheme for efficient network utilization, and reliable delivery protocols. Selective distribution using content-based constraints enables subscribers to receive only database changes of interest. By using a ”push” model, latency times between transaction commit at the primary database and delivery to the replicate are minimized. Through intermediate message queue storage, database change messages were logged to ensure reliable delivery and recoverability in the event of interruption in the distribution network. As illustrated in Figure 2 below, database transactions flow through a multi-tiered routing scheme for network efficiency. Each site may select all or a subset of the rows updated at the primary location, based on the content of one or more fields within the row.

4 Warm Standby

The goal of Replication Server is to provide publish and subscribe services for reliable delivery of database transactions. Many customers saw the additional advantages in replication to mirror databases for high availability. Many of the features required to support complex hierarchical distribution and selection publication and subscrip-

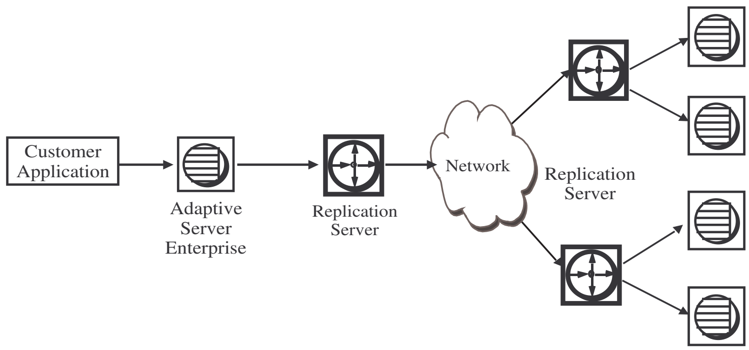

Figure 2: Multi-tiered routing scheme.

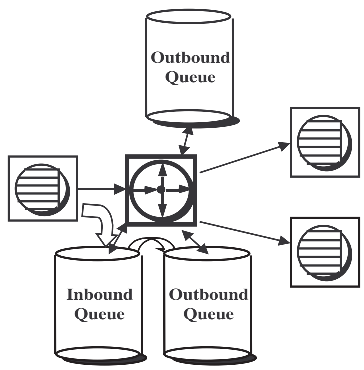

Figure 3: Publish and subscribe model.

tions are less critical in this environment. Because of the simplicity of this requirement, it is possible to simplify internal processing to enable distribution that is more efficient.

In some environments, it is desirable to support this database redundancy for availability combined with standard publish/subscribe services. By creating a special implementation in Replication Server 11 to support this requirement called Warm Standby, administration is simplified as it is no longer necessary to explicit subscribe to replicated tables. Since the Warm Standby implementation appears as a single logical entity to other replication participants, the publish and subscribe environment is not impacted by the additional of Warm Standby.

To contrast publish and subscribe replication vs. warm standby, figure 3 outlines the queues utilized in a standard publish and subscribe model. Transactions from the primary database are first committed to the inbound queue. Inbound queue transactions are filtered based on destinations and subscription criteria, then reordered and delivered to the outbound queues in transaction commit order. Finally, outbound queue transactions are delivered to the replicate databases. In the Warm Standby configuration, there is no need to order transactions or filter subscription criteria. In this model, transactions flow directly from the inbound queue to the standby configuration as shown in figure 4. The administrator marks tables to be replicated in the primary database. Subscriptions are not required to direct transactions to the standby database. The active and standby databases may exchange roles. This is transparent to other publish and subscribe activity against the active/standby database pair.

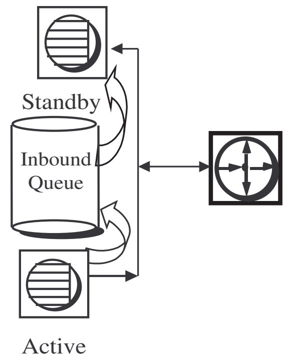

Figure 4: Warm Standby configuration.

5 Schema Replication for Warm Standby

The next evolution of Warm Standby addressed the database administration requirements. Replication Server distributed changes to user-defined tables only. Changes to the structure of tables, security permissions or the maintenance of other objects in the database (stored procedures, rules, triggers, etc.) remained a manual administration process.

In Replication Server 11.5, the concept of Schema Replication was introduced. Schema Replication enables the configuration of Warm Standby to encompass distribution of any operation performed on objects in the database. It also provides a database level configuration command, eliminating the requirement to mark individual tables as replicated objects. Any new objects created are immediately replicated to the standby site. By mirroring all database operations, the administration requirements for Warm Standby are drastically reduced.

6 High Availability Alternatives

There are a number of alternatives for high availability database architectures. This section outlines the major alternatives and the risks and benefits of each approach.

6.1 Hardware-based High Availability Solutions

Hardware and software solutions are available on most platforms to support high availability in a clustered environment. High availability clusters allow multiple CPUs to share common resources, such as disk drives, and provide automatic fail-over in the event of a CPU failure. Cluster solutions generally have severe distance limitations that preclude separating the hardware components adequately to mitigate risks associated with geographic proximity. These include natural disasters, such as fire, floods, tornadoes and earthquakes, and related services outages such as electricity and telecommunications networks.

Distributed disk arrays are emerging that mitigate the site protection issues by providing asynchronous mirroring to physically separate devices for better site protection. While synchronous mirroring satisfy the requirement for committed transactions to be preserved upon a disk or CPU failure, they do present various limitations. Asynchronous mirroring, while extending protection from many site-related failures, has additional risks associated with the potential loss of data. A discussion of benefits and limitations of both synchronous and asynchronous mirroring are presented in Table 1.

Benefits	**Limitations**
Synchronous disk mirroring with clusters: No loss of committed database transactions on a CPU failure.	Cluster fail-over may requires a significant downtime to perform appropriate restart and recover operations.
Applications generally do not require awareness of physical resource changes, such as network addresses.	Synchronous remote disk mirroring may impact application performance due to network overhead to remote disks.
Asynchronous disk mirroring can provide physical protection by supporting extended physical distances.	Most database systems require special handling to properly recover potential I/O loss in Asynchronous distributions.
Clustered CPU's may be used for other activities when not providing backup services.	Clustered CPU's are not available for use against the same database image.
	No protection from data corruption introduced by the hardware/software.

Table 1: Clusters with high availability disks (RAID/mirroring)

6.2 Software High Availability Solutions

Software high availability solutions are typically support physically separate hardware to provide protection against site loss. For example, a simple implementation of Cold Standby can be accomplished by periodically restoring database backup images and rolling transaction logs to a standby site.

For Hot Standby, applications could update multiple systems simultaneously (synchronous updates), using two-phase commit (2PC) protocols to guarantee transaction integrity. The risk of this type of solution is the potential business interruption if there is a failure in a participating system. Generally, 2PC applications are appropriate for very high-value transactions where consistency across multiple systems is more critical than availability.

Warm Standby offers the protection afforded by redundancy, without the availability constraints of synchronous updates or the time delays of batch-oriented cold backup methods. By providing asynchronous reliable delivery, applications are minimally impacted by the operation of the Warm Standby software system and the availability of the standby system. Benefits and limitations of software-level Warm Standby such as Sybase's Replication Server are presented in Table 2.

7 Warm Standby Considerations

Since a two-phase commit protocol is not being used to synchronize the standby site with the active site, there is a possibility that a failed active site has committed a transaction, but the standby site has not received that transaction. The major advantage replication offers over cold standby methods (such as tape backups) is that transaction loss due to latency is minimized due to the continuous ”push” architecture.

New procedures may be required to support the switch over to the standby system, since applications may not be transparently connected to the standby server when a failure occurs. The application can be designed to re-establish its connection to the appropriate database, or switching middleware may be imposed to support this function. It makes economic sense to maximize use of the Warm Standby database during non-failover periods for other business purposes. Warm Standby offers that flexibility as it maintains an online database system. Using the standby site for applications that generate reports, ad-hoc queries, or perform decision support functions is ideal. These applications would normally affect the performance of the active site so this offers the potential to reduce primary site overhead, with some potential impact to the standby. Applications running against the

Benefits	Limitations
Ability to quickly swap to the standby system in the event of failure, as standby database is online.	Warm standby system may be ”out-of-date” by transactions committed at the active database that have had sufficient time to reach the standby database.
Warm standby systems can be configured over a Wide Area Network, providing protection from site failures.	Client applications may need to be aware of a swap to the Warm Standby if the active system fails.
Data corruption is not replicated as transactions are logically reproduced rather than I/O blocks mirrored.	Logical transaction reproduction incurs more overhead than I/O block copying.
Originating applications are minimally impacted as replication takes place asynchronously after commit of the originating transaction.	Some overhead is introduced in the application server for transaction capture and delivery to the replication system.
The warm standby database is available for read-only operations such as DSS systems, improving utilization of standby databases.	Network bandwidth, hardware resource requirements and software overhead may limit scalability of software based solutions.

Table 2: Software-level Warm Standby systems.

standby site do need to be cognizant that the data may not be current, but decision-support applications often tolerate this latency.

In a peak volume OLTP environment, the effects of the Warm Standby database falling farther behind the active database's transaction stream must be considered. If the standby site must be operational within five minutes, and the standby system falls 30 minutes behind, then replication may not be the optimal solution. If the standby site is expected to fall 30 minutes behind during peak operations, yet the standby site has one (1) hour to recover, then using Replication Server to maintain the standby site is feasible. Capacity planning is an important consideration in the architecture of a standby system to ensure that not only average performance loads can be achieved, but that peak loads and typical down-time situations can be adequately recovered.

8 Summary

Warm Standby with Replication Server is a significant tool in providing distributed high availability services, affording protection against site failures through asynchronous, wide-area delivery of database transactions. By enabling multiple usage of standby systems, replication provides a cost-effective alternative to redundant hardware that is only utilized for recovery operations. In compliment with traditional hardware high availability solutions, Replication Server extends support for disaster recovery requirements beyond component failure recoverability to significantly improve site protection and reduce potential information loss of cold standby solutions.