the Availability Digest

www.availabilitydigest.com

Choosing a Business Continuity Solution
Part 4 – Choosing an Availability Architecture

October 2011

In the first three parts of this series,[1] we explored various data replication techniques and the highly available architectures that can be implemented with them. In this final part, we look at the considerations that will lead you to the choice of the proper architecture to meet your business needs.

The Replication Quadrant

In our discussion of highly- and continuously available systems, we have reviewed asynchronous and synchronous replication in unidirectional and bidirectional configurations. This gives us four distinct combinations to consider; and they are reflected in the Replication Quadrant, shown in Figure 1. In this figure, we depict typical RPOs and RTOs (Figure 1a), typical configurations (Figure 1b), and typical applications (Figure 1c) for the four combinations. Representative characteristics for systems implemented with each of these combinations are as follows:

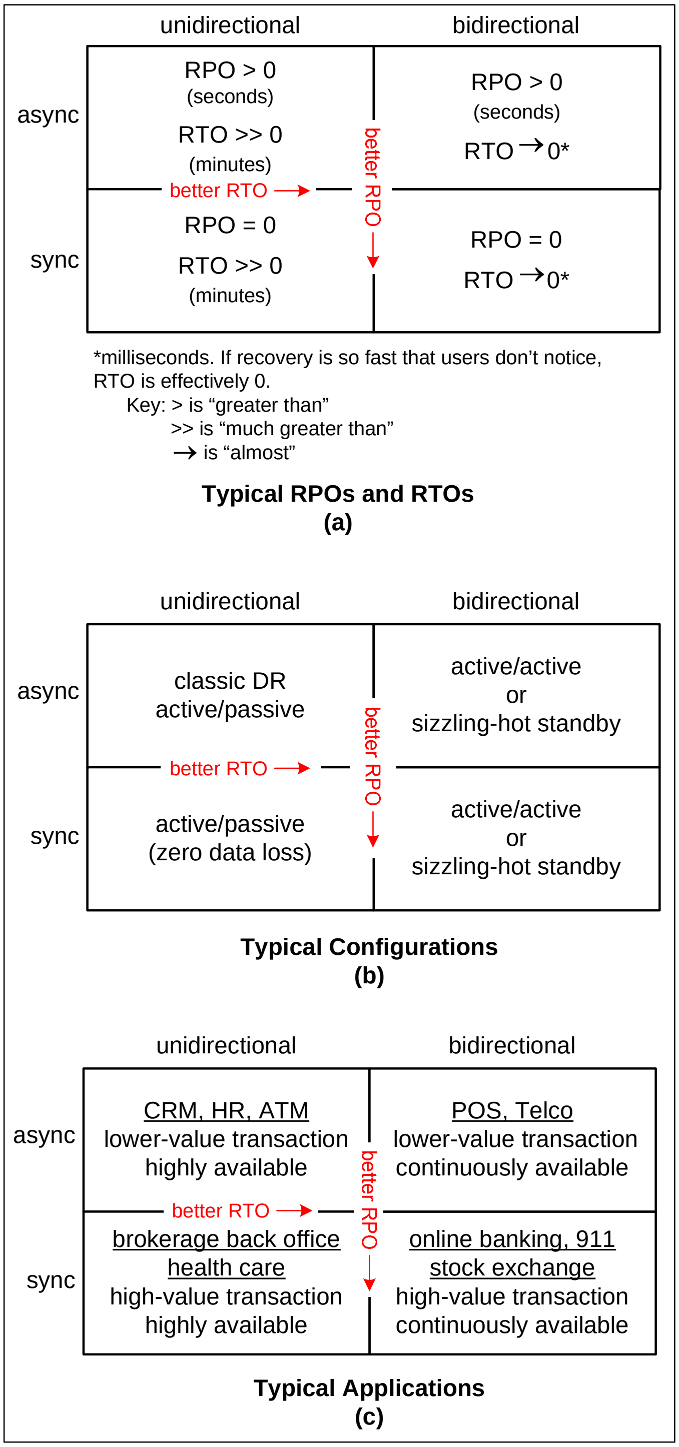
- A unidirectional asynchronous system has an RPO measured in seconds (the replication latency of the data-replication channel). Its RTO is measured in minutes or longer as applications are started following a failure of the active node, the databases are mounted, and the network is reconfigured. Additional recovery time is typically required for the management decision time to fail over to the backup system and for testing to ensure that the backup is performing properly.

  This replication method is used for classic disaster-recovery, active/passive configurations. It supports applications that must be highly available but for which some small data loss is tolerable.[2] CRM (customer-relationship management) and HR (human resources) corporate applications are examples of this class of application, as are ATM transactions. ATM transactions have a low value; and if the ATM machine is down, the customer can often go to a different ATM machine serviced by a different bank.

- A unidirectional synchronous system has the same recovery (RTO) characteristics as the unidirectional asynchronous system. However, it suffers no data loss following a node failure (its RPO is zero). Consequently, it is often referred to as a *zero data loss system (ZDL)*.

[1] This series of articles is a reprint of a Gravic, Inc., white paper and is published with the permission of Gravic. See the Gravic web site for their other white papers.

[2] Actually, there are many systems today that use asynchronous replication in production and that process high-value transactions. In these systems, there must be a way to recover lost transactions, such as manual reentry from printed reports.



	unidirectional	bidirectional
async	RPO > 0 (seconds) RTO >> 0 (minutes)	RPO > 0 (seconds) RTO → 0*
sync	RPO = 0 RTO >> 0 (minutes)	RPO = 0 RTO → 0*

*milliseconds. If recovery is so fast that users don't notice, RTO is effectively 0.

Key: > is "greater than"
>> is "much greater than"
→ is "almost"

Typical RPOs and RTOs
(a)

	unidirectional	bidirectional
async	classic DR active/passive	active/active or sizzling-hot standby
sync	active/passive (zero data loss)	active/active or sizzling-hot standby

Typical Configurations
(b)

	unidirectional	bidirectional
async	CRM, HR, ATM lower-value transaction highly available	POS, Telco lower-value transaction continuously available
sync	brokerage back office health care high-value transaction highly available	online banking, 911 stock exchange high-value transaction continuously available

Typical Applications
(c)

Figure 1: Choosing a Replication Method

It is suitable for applications that require high availability while processing high-value transactions that cannot be lost. Back-office applications for brokerage firms and banks as well as funds-transfer applications are examples. The current state of health care is heading here.

- A bidirectional asynchronous system has an RPO measured in seconds (the replication latency) and can recover in subseconds or seconds in an active/active or sizzling-hot standby configuration.

  It is suitable for applications that require continuous availability but for which some small data loss is acceptable. Typical applications include telco applications (many call-related transactions worth pennies). Point-of-sale (POS) transactions are another example. Like ATM transactions, they generally have low value. However, should a POS application go down, retailers cannot service customers using credit cards or debit cards.

- A bidirectional synchronous system is the ultimate in system availability. It suffers no data loss following a node failure (RPO = 0), and it can recover in subseconds or seconds in an active/active or sizzling-hot standby configuration. In fact, if the failover is sufficiently fast so that users don't realize that there has been a failure, in effect, there has been no failure. An RTO of zero has been effectively achieved.

  This configuration supports applications that must be continuously available and in which transaction value is high. Typical applications with these characteristics include online financial applications such as online banking and stock-market trading as well as some 911 applications.
  Online health care is another application that is heading here.

Risk Assessment

In order to make effective use of the Replication Quadrant, a company must know the availability requirements for each application. In an organization, some applications are typically mission-critical or even safety-critical and may require continuous availability. Business-critical and task-critical applications may require high availability. Some applications may be non-critical and can be down for hours or days without causing any great problem.

The critical ranking of applications may change over time. A good example of this is email. A few years ago, one could live without email for a few days. Today, however, email often forms the communication backbone between a company and its employees and customers. Company operations may grind to a halt if the email system is lost. In many companies, email has become a mission-critical or business-critical application that should have continuous availability.

Risk assessment is one of the steps in generating a proper Business Continuity Plan (BCP), and the procedures for risk assessment are well-documented in references describing how to create a proper BCP.[3] The result of the risk assessment should include the costs of system downtime and of lost transactions.

[3] See the *Disaster Recovery Journal* at www.drj.com.

Cost Factors

Along with the risk assessment should be an analysis of cost factors for the various architectures. Additional costs may include:[4]

- duplicate systems
- redundant networks
- bidirectional replication engine
- additional maintenance costs
- additional software licenses
- application modifications
- multiple sites
- additional staffing
- distributed management tools
- system testing

Cost/Benefit Analysis

Knowing the costs of downtime and of data loss and the costs of protecting against these, one can calculate the required RTO and RPO. For instance, a large brokerage firm finds that its average trade is about $25,000. If the firm earns an average of 2% commission on each trade, a trade is worth $500 in revenues to the firm. If it is doing 10 transactions per second, its cost of downtime is $5,000 per second, or $300,000 per minute. Its order-processing system currently runs on a UNIX cluster, which the firm estimates will be down about five minutes per year. Its total cost of downtime is therefore $1,500,000 per year.

The firm has determined that it will cost $1,350,000 to add enough hardware, software, and networking to eliminate all single points of failure. It wants to get a return on its investment in one year. An RTO of 30 seconds for the new system will save 4.5 minutes of downtime per year, yielding the required $1,350,000 savings. Furthermore, the brokerage firm cannot lose any transactions without being in violation of security regulations. Therefore, it needs an RPO of zero. From the Replication Quadrant of Figure 1, it is clear that the firm should seriously consider an active/active configuration using synchronous replication.

In other cases, it is not the monetary cost of downtime that influences the choice of an availability architecture but rather a required level of service. Consider the Home Location Registers (HLR) that mobile telecommunication providers use to track mobile users and to place calls. An HLR typically handles 10,000 customers. If an HLR should fail, none of the mobile users that it is servicing can make or receive mobile calls.

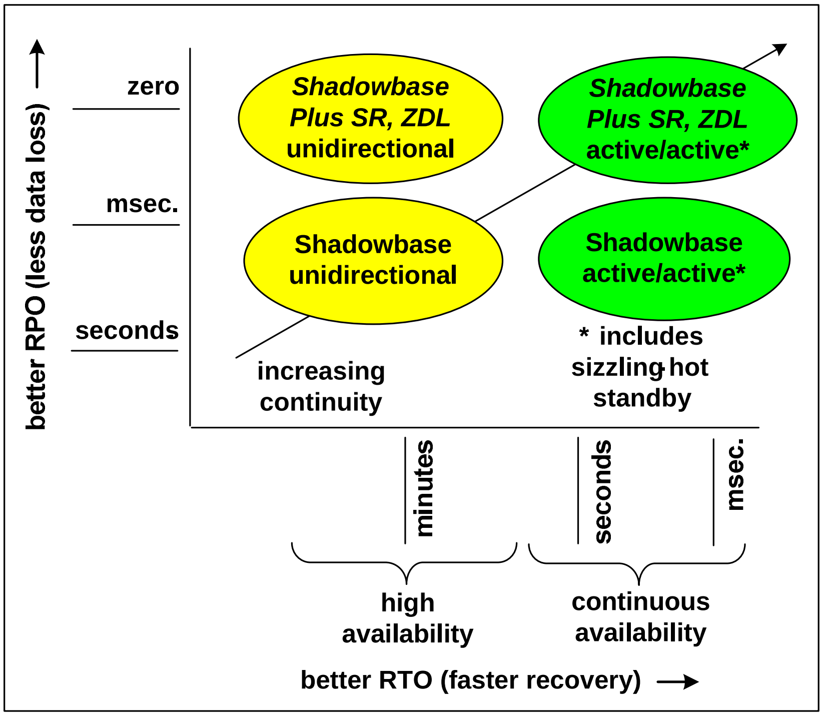
During a busy period, an HLR may be placing 1,000 calls per minute. If each call has a value of $1.00, the cost associated with the downtime of an HLR is $1,000 per minute. If an HLR is down for five minutes each year, the mobile provider will lose $5,000 per year due to the downtime of an HLR. This most likely would not justify the cost of an active/active system. However, the market and regulatory requirements for uninterrupted mobile service even in the face of a disaster argue strongly for the continuous availability of active/active systems.[5]

[4] Total Cost of Ownership, *Breaking the Availability Barrier II: Achieving Century Uptimes with Active/Active Systems*, AuthorHouse; 2007.
Achieving Century Uptimes Part 2: What Will Active/Active Cost Me? *The Connection*; January/February 2007.

[5] HP's Active/Active Home Location Register, *Availability Digest*; November 2006.
Telecom Italia's Active/Active Mobile Services, *Availability Digest*; March 2007.

The Shadowbase Replication Suite

The Shadowbase Replication Suite from Gravic, Inc., (www.gravic.com) comprises a set of products that support unidirectional and bidirectional data replication between homogeneous and heterogeneous systems. Shadowbase currently supports asynchronous replication, and synchronous replication is on the product roadmap.[6] The range of availability architectures supported by Shadowbase is shown in Figure 2.



better RPO (less data loss)
zero
msec.
seconds
Shadowbase Plus SR, ZDL unidirectional
Shadowbase Plus SR, ZDL active/active*
Shadowbase unidirectional
Shadowbase active/active*
increasing continuity
* includes sizzling-hot standby
minutes
seconds
msec.
high availability
continuous availability
better RTO (faster recovery)

Figure 2: The Shadowbase Business Continuity Continuum

Shadowbase replication products include:

- ***Shadowbase***, a low replication latency asynchronous replication engine using efficient process-to-process data replication. Shadowbase supports both unidirectional and bidirectional data replication.

- ***Shadowbase Plus SR***, a synchronous data-replication engine using *coordinated commits* for low application latency over long distances. ***Shadowbase Plus SR*** eliminates data loss and data collisions.

- ***Shadowbase ZDL***, a synchronous data-replication engine using coordinated commits similar to ***Shadowbase Plus SR.*** In addition to eliminating data loss, ***Shadowbase ZDL*** reduces application latency relative to ***Shadowbase Plus SR*** by safe-storing rather than applying replicated data during the source transaction. This may lead to data collisions in certain active/active configurations for some applications.

[6] Achieving Century Uptimes - Part 17: HP Unveils Its Synchronous Replication API for TMF, *The Connection*; July/August 2009.
Chapter 4, Synchronous Replication, *Breaking the Availability Barrier: Survivable Systems for Enterprise Computing*, AuthorHouse; 2004.
Contact Gravic for the future availability of its synchronous-replication products.

The Shadowbase product suite covers the upper-right quadrant of the Business Continuity Continuum depicted in our previous part in this series.

Shadowbase Asynchronous Replication Engine

The Shadowbase asynchronous replication engine uses process-to-process transactional replication to move data from a source database to a target database. Since Shadowbase introduces no disk-queuing points, replication is highly efficient; and replication latency is minimized.

A unidirectional Shadowbase asynchronous replication engine is useful to maintain synchronization of a backup database with its primary database. In this case, the backup system is typically idle so far as update processing is concerned. However, it can be used for query processing since Shadowbase provides for the target database to be a consistent copy of the source database, though delayed by the replication latency.

By extending this architecture to bidirectional data replication, both nodes can be actively processing transactions since each has an up-to-date copy of the application. Shadowbase accomplishes bidirectional replication by using two replication engines, one in each direction, that share information about the data being replicated to avoid ping-ponging.[7]

As described earlier, a serious problem encountered with bidirectional asynchronous replication is data collisions. The Shadowbase asynchronous data-replication engine supports collision-avoidance methods. If collisions cannot be avoided, Shadowbase provides the mechanisms necessary for detecting and resolving data collisions by embedding application logic into the replication engine.

Shadowbase replication engines have been used for several decades in many high-demand, mission-critical installations.

Shadowbase Plus SR Synchronous Replication

The upcoming ***Shadowbase Plus SR*** synchronous replication engine provides both unidirectional and bidirectional synchronous replication.[8]

Shadowbase Plus SR is a future technology based on Gravic's patented coordinated-commit technology,[9] which uses asynchronous replication of database transactions coupled with synchronous commit. It extensively leverages the existing Shadowbase asynchronous architecture to replicate data. In addition, it joins the transaction at the source system and votes on the outcome of the transaction. If at commit time it has successfully applied all of the transaction updates to the target database, it votes to commit. Otherwise, it votes to abort.

Shadowbase Plus SR continues the best-in-class features of Shadowbase, such as low latency, high availability, and communication efficiency.

[7] As described earlier, "ping-ponging" or data oscillation is the re-replication of replicated data back to the source database. See the section in this white paper entitled Bidirectional Replication and Active/Active Systems on page 11.
G. E. Strickler, H. W. Knapp, B. D. Holenstein, P. J. Holenstein, Bidirectional database replication scheme for controlling ping-ponging, *U.S. Patent 6,122,630*; September 19, 2000.

[8] Check with Gravic at www.gravic.com for the future availability of this product.

[9] B. D. Holenstein, P. J. Holenstein, W. H. Highleyman, Asynchronous coordinated commit replication and dual write with replication transmission and locking of target database on updates only, *U.S. Patent 7,177,866*; February 13, 2007.

Shadowbase ZDL

Shadowbase ZDL is a future technology that is a modification of ***Shadowbase Plus SR.*** It also joins the transaction at the source system and uses the Shadowbase asynchronous replication engine to replicate data. However, rather than directly applying the changes to the target database, it safe-stores them in a target-side persistent queue file and then applies them to the target database in the background. At commit time, ***Shadowbase ZDL*** will vote to commit if it has successfully safe-stored all changes in its queue file, whether or not they have been applied to the target database.[10]

Consequently, application latency is reduced since it is much faster to queue changes than to apply them to the database. However, since changes are applied asynchronously, applications at the target system can modify the same data items before the replicated changes are applied, thus leading to data collisions.

Shadowbase ZDL can also be used with the Shadowbase asynchronous replication engine to provide a safe-store on the target system to queue changes should the target database be taken offline.

Summary

The technology exists today to achieve arbitrarily fast recovery times following a system failure with little if any loss of data. The key to this technology is data replication.

Data replication comes in several forms – asynchronous or synchronous, unidirectional or bidirectional. Each combination supports different ranges of recovery times (RTOs) and data loss (RPOs). By understanding the costs of downtime and data loss for each application and the costs of achieving various levels of high availability and continuous availability, IT management can make informed decisions concerning the availability approach that is right for each application.

The Shadowbase suite of replication products provides the full range of replication technologies to satisfy the most demanding IT availability requirements.

[10] Check with Gravic at www.gravic.com for the future availability of this product.