



“Achieving Century Uptimes” An Informational Series on Enterprise Computing

**As Seen in *The Connection*, An ITUG Publication
December 2006 – Present**

About the Authors:

Dr. Bill Highleyman, Paul J. Holenstein, and Dr. Bruce Holenstein, have a combined experience of over 90 years in the implementation of fault-tolerant, highly available computing systems. This experience ranges from the early days of custom redundant systems to today’s fault-tolerant offerings from HP (NonStop) and Stratus.

Gravic, Inc.
Shadowbase Products Group
301 Lindenwood Drive, Suite 100
Malvern, PA 19355
610-647-6250
<http://www.gravic.com/shadowbase>

Achieving Century Uptimes
Part 2: What Will Active/Active Cost Me?
January/February 2007

Dr. Bill Highleyman
Dr. Bruce Holenstein
Paul J. Holenstein

Before embarking on a migration to an active/active configuration, management is certain to ask the obvious question, “How much is this going to cost the company? Are the advantages which it brings worth the cost?”

This is not a simple question. The cost of an active/active system is complicated by the very fact of its flexibility – there are simply many ways in which such a system may be implemented. Choices include how many nodes should there be, what kind of systems should be used for these nodes, how many database copies across the network are required, where should the nodes and database systems be located, can “lights-out” operations be used at some nodes, what will be the communication infrastructure, and so on. There are no simple answers to these questions.

Objectives

Before any of these questions can be answered, some fundamental parameters describing what the active/active system is intended to do must be established.

Recovery Time Objective (RTO)

RTO is, in effect, the target mean time to repair (MTR) the system. Unlike monolithic systems, the recovery time of a failed node is immaterial. This is true if the time that it takes to switch users from a failed node to a surviving node is so short that the users don’t see the node failure as an outage.

However, this wonderful statement is not quite true. Though recovery from a node failure is very fast, there is the possibility that multiple node failures could take down an active/active system. Although this might happen with a frequency measured in centuries, it still can happen. The time to get at least one node returned to service so that the system is once again up and running is an important parameter and becomes the RTO for an active/active system.

Mean Time Before Failure (MTBF)

The MTBF of the nodes is not very meaningful since node failures are virtually transparent. The MTBF for the system is more significant. However, if we are measuring MTBF in centuries, it may be meaningless to try to specify system MTBF.

Availability

A more meaningful measure of system reliability than its MTR and MTBF is its availability. This is the proportion of time that it is expected to be up. Availability is important because in most active/active applications, downtime has very serious consequences, financial costs being only one. Here we speak of five 9s (an average of five minutes downtime per year), six nines (thirty seconds downtime per year), seven 9s (three seconds downtime per year), and so on.

Recovery Point Objective (RPO)

RPO is the amount of data loss that is tolerable following a node failure. If asynchronous replication is being used, some data may be lost in the replication pipeline following a failure. Is 100 msec. of data loss acceptable? 10 seconds of data loss? Or must no data be lost?

Facility Separation

One of the main reasons for using any distributed system architecture, whether it be active/active or active/backup, is to achieve a degree of disaster tolerance. What is the minimum distance that the nodes must be separated? Can they be collocated? On a campus? Within a metropolitan area? Or must they be thousands of miles apart? In addition, the number of nodes that are desired must be established. This may be determined by the needs for data locality relative to user communities.

Configuration

Now that the important parameters have been established, some configuration decisions that affect the cost of the system can be made.

Number of Nodes

The number of nodes that will be required in the active/active system can be established. It is at least as many as determined in the Facilities Separation step above. However, there are tradeoffs between the amount of total capacity that must be purchased and the total number of nodes. The greater the number of nodes, the less the total purchased capacity need be. For instance, assuming that full capacity is needed in the event of a node failure, each system must be able to support the full load in a two-node system. Thus, 200% of the required capacity must be purchased. However, in a five-node system, each node needs only to carry 25% of the load. Therefore, only 125% of the required capacity need be purchased.

The number of nodes and the required availability objective determine the number of spare nodes that are required. Furthermore, it can be shown that the use of fault-tolerant

nodes (such as NonStop servers) will generally require fewer nodes than the use of high-availability nodes (such as UNIX servers).¹

Number of Database Copies

The number of database copies that are required in the application network can also be determined. This is a matter of the system availability objective and also of the distribution of the nodes if data locality to user communities is important. If users are to be served by local nodes, partitioning the database so that each node only contains that portion of the data needed by that user community (backed up by at least one other node) can reduce the total amount of disk storage needed in the network.

Network Redundancy

The system availability objective and the geographic separation of nodes as well as transaction activity will determine the choices that are available for the communication network. Does the availability requirement dictate redundant WAN networks? Redundant LAN networks? Can the Internet be used as an inexpensive WAN backup? Can dialed lines be used as backup?

Database Synchronization

The required RPO and the geographic distribution of the nodes dictate the type of data synchronization technique that can be used.

- If the RPO is zero, synchronous replication must be used.²
- If the RPO is zero and the nodes are close enough together, network transactions can be used.
- If the RPO is zero and the nodes are far apart, synchronous replication must be used.³
- If the RPO is not zero, asynchronous replication may be used. The RPO value dictates the replication latency of the replication engine. If the RPO specifies that one second of data may be lost, the asynchronous replication engine must have a replication latency less than one second.

Note that very small but non-zero RPO values may dictate synchronous replication if there are no available asynchronous replication engines with replication latencies less than the RPO specification.

¹ See Chapter 1, [Achieving Century Uptimes](#), in *Breaking the Availability Barrier: Achieving Century Uptimes with Active/Active Systems*, by Paul J. Holenstein, Dr. Bill Highleyman, and Dr. Bruce Holenstein, referred to as Volume 2 hereafter.

² Though not synchronous replication, split mirrors for audit is also a technique to avoid data loss.

³ See Chapter 4, [Synchronous Replication](#), in *Breaking the Availability Barrier: Survivable Systems for Enterprise Computing*, by Paul J. Holenstein, Dr. Bill Highleyman, and Dr. Bruce Holenstein, referred to as Volume 1 hereafter.

Active/Active Costs

The costs associated with going active/active can now be established. They include both initial costs and recurring costs. Many of these costs are similar to what would be incurred for a monolithic system, but others are unique to active/active systems.

Initial Costs

Initial costs include processor, storage, network, facility, and software costs.

Processors

Under consideration may be several configurations which vary the number of nodes and the type of system used for each node. Each of these can be priced out by getting quotes from the system vendors.

One note of caution is voiced by The Standish Group: “In order to calculate a meaningful predicted Total Cost of Ownership, one must first properly size the system.” Their point is that system vendors will be quick to undersize their systems for proposal purposes while inflating the systems of their competitors.

Data Storage

Storage also can now be priced. Again, there may be several configurations to consider, including direct-attached or network-attached storage and the use of single disks, mirrored disks, or RAID.

Note that every node need not have a full complement of storage depending upon the partitioning strategy chosen. Also, not every node may need storage depending upon data access routing strategies.

Network

The objectives have determined whether networks need to be redundant or not. This may not have much of an impact on WAN networks, but redundant LANs will probably mean more than twice the amount of hardware than single LANs.

Facilities

The objectives have determined the number of facilities required and probably their location. Initial facilities costs include construction, raised flooring, air conditioning and heating, desks, telephone systems, and so on. If it has been determined that some facilities will be “lights-out,” facility costs will be less for those sites than for manned facilities.

Software

There are several startup software costs to consider:

- *Migration*: If the applications are not active/active-ready, they may have to be modified to be able to run in parallel with other copies of the application.⁴
- *Licenses*: Initial software licenses must be purchased. In addition to the normal licenses that would be purchased for a monolithic system, licenses for the selected data replication engine, if any, must be obtained.
- *Network Management Subsystem*: Though this is just another software license, it is a major one. The network management system is a key active/active system component that allows the application network to be configured, managed, and monitored.

Recurring Costs

Recurring costs include hardware maintenance, software maintenance and licensing, communication networks, personnel, facilities, and insurance.

Hardware Maintenance

In an active/active system, it is possible that a more relaxed hardware maintenance policy can be incorporated. This is because in the event of a node failure, the system is still operational. Satisfactory service may be achieved by contracting for next-day service rather than same-day service, for instance.

Relaxed maintenance service means that a node is down for a longer period of time, thus reducing its availability. Furthermore, its extended downtime raises the probability of a second node failure before the first is returned to service. This, in turn, reduces the system availability. Also, in the event of a total system failure, it may take longer to restore the system to service – a parameter governed by the RTO determined as one of the objectives. These must all be considered when deciding to what extent hardware maintenance responsiveness can be relaxed.

Software Maintenance and Licensing

Software maintenance from the system vendor and third party vendors is typically included in the recurring software license.

An important and costly element of this category is the licensing cost for the network management system.

⁴ See Chapter 8, Eliminating Planned Downtime with Zero Downtime Migrations, in Volume 2, referenced earlier.

Communication, Personnel, and Facilities

Given the type and redundancy of the communication channels, the number of personnel needed at each facility, and the location of the facilities, the cost for these items can be determined. In addition, there will be internal staff required to maintain the applications and to keep the system properly configured.

Insurance

The cost of business risk insurance may well be reduced since an active/active system is less susceptible to loss.

The Cost of Downtime

A major difference between the cost of a monolithic system and an active/active system is the cost of downtime. Studies have shown that for many companies, the cost of downtime can range from hundreds of thousands to millions of dollars per hour. Bad publicity may occur. Even worse, property and life may be put in danger.

Often, the savings in downtime cost can justify an active/active system decision, whatever its cost. Sometimes, this decision is also affected by regulatory requirements.

Calculating the Cost of an Active/Active System

The elements of active/active system cost calculations is a veritable alphabet soup, including TCO, ROI, NPV, IRR, and others.

Total Cost of Ownership (TCO)

TCO is the total cost of the system over a period of years, including all initial and recurring costs, the cost of downtime, and the cost of money. This is the bottom line that management may be seeking.

Return on Investment (ROI)

If there is a savings to be achieved relative to the current operations (which there almost certainly will be if the cost of downtime is significant), ROI is a measure of the time in which the investment in the new active/active system will pay for itself.

The Standish Group (www.standishgroup.com) has a powerful tool, Virtual Advisor, to help in calculating TCO and ROI.

Net Present Value (NPV)

NPV is used to factor in the cost of money in the calculation of the total cost of the system.

Internal Rate of Return (IRR)

IRR is a method that can be used to determine the breakeven cost of money (the interest rate) at which the cost of two approaches is equal.⁵

Summary

The comparison of the costs of an active/active system versus a monolithic system is a complex task. On the one hand, the active/active system incurs the cost of redundancy and network management. On the other hand, it can dramatically reduce the cost of downtime and the related insurance costs. Only by going through a rigorous analysis can this comparison be determined.

⁵ See Chapter 9, Total Cost of Ownership, of Volume 2, previously referenced, for a detailed description of NPV and IRR.