Roadmap to Availability on the AS/400e: *A White Paper*

Worldwide AS/400 Availability Team, July 2000

Second Edition (July 2000)

This edition applies to OS/400™ Version 4 Release 4 and later.

Comments may be addressed to:
Eric Hess, 01-03-50
IBM Corporation
7 Campus Circle
Roanoke, Texas
76262

- or -

texas@us.ibm.com

When you send information to IBM, you grant IBM a non-exclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Table of Contents

The members of Worldwide IBM AS/400 Availability Team that authored this paper are:

Sue Baker
Nick Harris
Paul Kirkdale
Valerie Smith
Steve Finnes
Eric Hess
Sue Nee

Thanks to the following for their invaluable contributions:

Janice Dunstan
Mike Snyder
Glenn Van Benschoten (LakeView Technology)
Joy Mosley
Brenda Thompson

Roadmap to Availability on the AS/400*e*: *A White Paper*

Roadmap to Availability on the AS/400e

AS/400e 24x365…no time for down time.

e-business, globalization, industry consolidation, merger mania, and server consolidation are the forces setting new standards for system availability. The AS/400 has always been known for legendary availability, and its success in the commercial computing arena has made it the central database or enterprise server for a significant share of the industries around the world. Name your favorite company, there is a very good chance that the AS/400 is playing the central role for the commercial On-Line Transaction Processing. The so called "back office".

The corporate clients using the AS/400 as the enterprise server, the mainframe if you will, are demanding continuous availability. No time for planned outages and certainly no time for unplanned outages is fast becoming the norm. The luxury of having a window of time to perform system backups has vanished, along with the idea that one can risk the possibility of an unplanned outage. In the increasingly seamless world of e-business, CRM, and SCM, where financial transactions are dependent on a chain of events, no link in the availability chain can be broken and certainly, the central engine of commerce, the AS/400 can not be down.

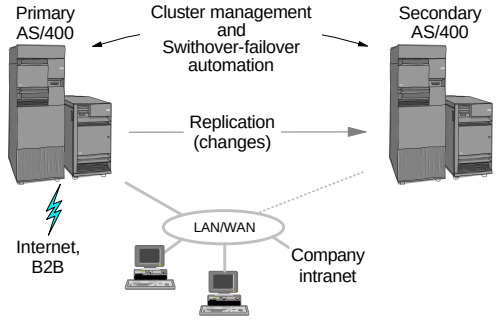
Availability is a discipline, it is a technology, it is a corporate culture. It involves systems management, application design, even organizational structure and above all, a commitment from the IT organization and the business to make it a focus. There is no free lunch. In this paper, we are going to help you understand how you can achieve world class results. Our goal is to arm you with knowledge so you are the informed decision maker in this critical area of IT management.

Meeting your business' availability needs with AS/400 Clusters

A well managed single AS/400 can provide the highest levels of availability of any non-clustered sever in the industry. However, we have now moved deep into an era where Information Technology and information do more than simply support a business…it provides businesses a competitive edge in the marketplace. Small outages, tolerated just a few years ago, can now mean significant losses of revenue and future opportunities for your business today.

Clusters are the best solution for continuous availability requirements on an AS/400, providing rapid recovery for the widest range of outages possible, with minimal cost.

HABP Cluster Solution:



A basic AS/400 Cluster.

What is an AS/400 Cluster? An AS/400 cluster is a combination of redundant AS/400 hardware and software, and clustering middleware solutions from an IBM High Availability Business Partner (*a.k.a. HABP*), providing the capability to minimize the impact from both *planned and unplanned outages*. A clustering solution can provide both data and application resiliency (prior to AS/400 Version 4 Release 4, clustering solutions provide only data resiliency). In addition, a modern clustering solution will provide a single point of control and GUI interfaces to make the job of managing a cluster much easier. Like all server cluster solutions, the level of availability provided by an AS/400 cluster is directly related to the level of commitment from the business' leaders within an organization, the focus and effort put into the initial planning of the

cluster, the level of recoverability built into applications, how well the systems are managed, and the investment in the education of the operations staff.

Estimating the Value of Availability

As higher levels of availability increase the investments required, it is important to have a good understanding of the dollar value IT systems provide to the business, as well as the costs to the business if these systems are not available. Analysis in this area is often delayed because the exercise is time consuming and difficult, considering the number of variables that exist within the company.

Once the Value of Availability for your IT services has been determined, you have an invaluable tool that can be referenced to establish availability requirements, to cost justify appropriate investments in availability management solutions, and to measure all returns on that investment. Consider the following:

Analyze by major application or service provided - The major cost of an outage is the cumulative total of not having the applications available to continue business.

What is value of the system being up? - It is not easy to determine the cost of outages. The inaccessibility of each application or program could have widely varying effects on the productivity of its users. You must start with a reasonable estimation of what each critical application is worth to the business. Some applications are critical throughout major portions of the day, while others can be run any time or on demand.

Direct vs. indirect costs - *Direct costs* are the time and revenue lost directly because the system is down. *Indirect costs* are those incurred by another department or function as a result of an outage. For example, a marketing department may absorb the cost of a manufacturing line being shut down because the system is unavailable. This is an indirect cost of the outage, but nonetheless a real cost to the company.

Tangible vs. intangible costs - *Tangible costs* are direct and indirect costs that can be measured in dollars and cents. *Intangible costs* are those for which cash never changes hands, such as lost opportunity, good will, market share, etc.

Fixed vs. variable costs - *Fixed costs* are direct, indirect, tangible or intangible that would result from the failure regardless of the length of the outage. *Variable costs* are those that vary with the duration of the down time, but they are not necessarily directly proportional.

Primary considerations:

1. **Loss of revenue**
   *Loss per hour = gross annual revenue / annual business hrs * % of the business impacted * # of outage hrs*

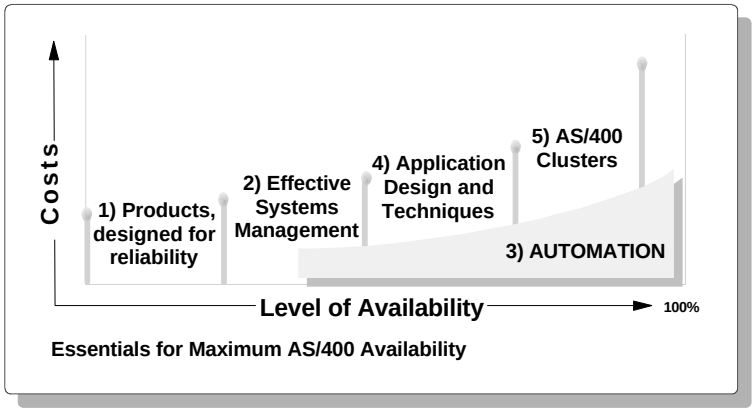
2. **Cost of labor**
   *Cost of labor = # of people * % affected * average $ / hr*

3. **Lack of service to your customers, lost opportunity, and damage to business image**

(1) For more detailed calculations and methodology refer to: "So you want to estimate the value of availability" (IBM publication GG22-9318)

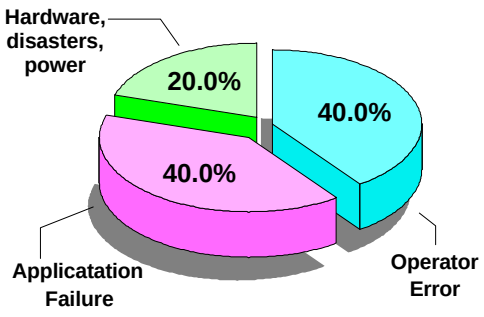
Essentials for Maximum AS/400 Availability

Achieving the highest level of availability possible cannot be accomplished with clustering software and hardware solutions alone. When working towards maximum availability, it becomes very important that planning be performed not only for the operating system, hardware, and database, but also for applications and operational processes. Regardless of sever platform, an unreliable system, poor systems management, security exposures, lack of automation, and applications that do not provide transaction recovery and restart capabilities can weaken availability solutions. The biggest value from your availability investment is achieved when system reliability, effective systems management, automation, and application design are addressed just as vigorously.



Essentials for Maximum AS/400 Availability

1. Utilize products designed for availability - A single AS/400 system delivers an average of 99.9+% availability. According to data collected by IBM over the last two years, AS/400 owners have experienced on average less than 9 hours of unplanned down time per year. IBM is able to deliver a very reliable AS/400 due to the fact that the IBM Development team designs, creates, builds, tests, and services the AS/400 as a single entity. The Rochester Development Laboratory has the responsibility for the hardware, operating system, database, and the other middleware facilities that make up a complete system.



Unplanned outage factors

2. Employ effective system management process - Lack of attention to systems management disciplines and process can effect availability to a very measurable degree. Availability solutions such as clusters are undermined when systems management processes are lacking or nonexistent. An effective system management strategy ties heavily into automation, such as automatic archival of data, continuous system auditing, responding to security exposures, and monitoring of error logs, backup and restores, and so on. A quote by Gartner group puts this in perspective: *By year 2003, "100% availability will remain elusive as user-controlled disciplines have an ever-greater relative impact on achieving availability"* (Gartner Group, June Conference, Dallas, Tx., 1997).

It is easy to see why an investment in systems management disciplines, automation, and application recovery is necessary. Only a few of additional hours of yearly downtime will reduce availability from 99.99% availability to 99.9%.

Availability Percentage		Total Outage Per Year
99.9999	=	32 seconds
99.999	=	5 minutes
99.99	=	53 minutes
99.9	=	8.8 hours
99	=	87 hours (3.6 days)
90	=	876 hours (36 days)

How long is an outage really?

3. Increase automation -- Increased availability means the need for reducing the possibility of errors and recovery delays and increased consistency. In many cases, human errors create more down time than hardware failures or power outages. More effective automation through the use of automation software and tools can help offset an overburdened staff and allows them to attend to more unique and critical decisions and tasks. As availability requirements increase, investments in automation must increase as well.

4. Exploit availability techniques, and applications design for availability - Decrease unplanned outages or their effects by utilizing availability options such as disk protection, access path protection, file journaling, and user ASPs (Auxiliary Storage Pools).

Target a phased approach at increasing application resiliency (recoverability) and availability. As a general rule, if an application on a non-clustered system is difficult to recover, it will just as difficult to recover in a clustered environment, and may reduce the value of your cluster investment. Said another way, a cluster solutions cannot overcome a poor application design.

Decrease application recovery times and database integrity issues (incomplete transactions) by using applications that incorporate commitment control or other methods of transaction, and application recovery. In fact, any environment that is not using journaling techniques cannot achieve high availability. Journaling is the programmatic movement of data from memory to disk and is the only way to insure completed transactions. Therefore, at a minimum, all systems must deploy journaling while the more sophisticated and highly available applications also use commitment control. These two techniques define the building blocks to all highly available environments.

5. Implement special solutions to meet your availability goals - To reach your availability goals, special solutions such as AS/400 clusters with monitoring, automatic switch over and recovery automation are implemented to control both planned and unplanned outages.
If you side step issues 1 through 4 (described above), even sophisticated options like AS/400 Clusters may not provide the highest levels of availability possible, since small outages (such as recovering or reentering transactions) add up.

Prerequisites to implementing AS/400 Clusters

To accomplish data replication, the clustering middleware from IBM's HABPs make use of OS/400 journaling (called a log on other servers). A journal's receiver contains the entries that the system adds when events occur that are journaled, such as changes to database files, changes to other journaled objects, or security-relevant events. The HABP clustering software coordinates and replicates these changes between a primary AS/400 and a backup AS/400…providing an extremely efficient method of continuous replication.

Prior to implementing clustering, you will want to test the impact journaling may add to your system, if it is not currently activated. In most cases, the impact is minimal. However, some older applications migrated from older platforms may see more of an impact. An IBM Integrated Technology Services representative or the High Availability Business Partner's services organization can help you plan journaling implementation, as well as provide techniques to increase journaling performance and minimize its impact to the system.

OS/400 journaling is also a prerequisite to Commitment Control. Commitment control is an extension of journal management which allows applications to keep database files synchronized and ensures transaction integrity

Commitment control helps ensure that:

- All changes within a transaction are completed for all files that are affected.

- If processing is interrupted before the transaction is completed, all changes within a transaction are removed.
- Changes that are made during a transaction can be removed when the user program determines that it is necessary to do so. This is called a rollback operation.

Without commitment control, recovering data for a complicated program requires detailed application and database knowledge. Although adding Commitment Control to your applications is optional, as the volume *and* value of business transactions increase, transaction integrity will become more critical. Commitment Control should be a priority in your next application upgrade.

What level of availability do you need an AS/400 cluster to provide?

Single system AS/400 availability options (such as backup power, OS/400 System Managed Access Path Protection, OS/400 Disk Mirroring or RAID-5 disk protection, redundant I/O technologies, redundant cooling and power supplies, double bit memory error correction, etc.) provide very high levels of availability. The implementation and operation of these features are virtually transparent to your business' applications, and have very little impact on your daily operations. Likewise, most application development departments or organizations find it unnecessary to be concerned with the implementation or day-to-day operations of these features (after all, OS/400, AS/400 hardware or the operations department will always be on top if it). We need to change that way of thinking with clusters.

Like all cluster technologies in use today, AS/400 Clusters should not be thought of as a shrink wrap or a plug-and-play solution. Although the AS/400's superior architecture, integrated nature, and ease-of-use provide fast solution deployments, clustering will be new to your business and staff. Your business will be planning, learning, implementing, and operating an AS/400 technology that may be very new to your staff. The most efficient AS/400 cluster installations in operation today were accomplished with thorough business and project planning, continuous involvement from business's IT operations and application development organizations

In the event of an outage, speed is of the essence. How you design and implement your AS/400 cluster will affect the time it takes to get back to business. The speed in which the backup AS/400 can take over (in the event of an unplanned or planned outage) is controlled by a number of factors such as:

- **How well the data replication function has been planned and tuned** - Of utmost importance is insuring that data is quickly replicated from the primary AS/400 to the backup. Since the HABP data replication software provides the ability to multithread data replication, you have the capability to tune data replication to your environment. Initial design of your cluster's data replication, followed by fine tuning after implementation, will help assure that the backup AS/400 stays up to date with the primary AS/400. Focus on HABP solutions that take advantage of AS/400 technologies like remote journaling. Remote journaling moves the task of transporting the contents of the journal receiver to the back up system into OS/400. Remote journaling is a critical advancement in the high availability foundation of the AS/400 and represents new investments by IBM in AS/400 clustering technology. Your choice of a high availability solution provider should be based in part on their ability to bring AS/400 core technology to market in a timely fashion.
- **Communications performance between the AS/400s** - For larger more complex environments, local clusters can easily take advantage of additional, low cost, high speed network cards, or OptiConnect/400 when additional bandwidth is needed.
- **The level of training for the operations staff, and IT operation policies** - Clustering requires that the operations staff be well trained on the daily use of an AS/400 cluster and in the event of an outage, efficient in performing a role swap between AS/400s. The operations staff must also be trained to monitor and address any data replication problems

that may occur. Equally important is the decision on when to switch. It must be crisply defined and documented. Many minutes (perhaps hours) are lost when a debate among the IT staff erupts on whether to switch over (e.g., perform a role swap) to the backup AS/400, or attempt repairs on the primary.

- **The ability for the application to recover from an outage quickly, automatically** - In the event of an outage, once you have switched over to the backup AS/400, are your applications and databases ready for business? Do you have incomplete transactions that you must manually investigate, then delete, or reenter? Must batch jobs be restarted from the beginning? These types of application recovery issues can delay business recovery on the backup AS/400. As your business grows, so do transaction sizes, volumes, and complexities, requiring applications that provide automatic recovery.

AS/400 clusters provide significant flexibility and growth to meet your availability need. Your business can start with a *base cluster implementation* (sometimes referred to as AS/400 Dual Systems), in which data replication alone is the primary vehicle for availability. In the event of an outage, some amount of time will be required by your staff to verify the state of applications and databases on the backup AS/400. Although switch over times for an unplanned outage may be slightly longer, there is very little investment in application changes.

To reduce the time required to verify the status of a database on the backup AS/400 in the event of an outage, you business can invest in a *cluster with increased application recoverability*. In this case, incomplete transactions are eliminated by integrating OS/400 Commitment Control in critical applications.

Ultimately, when an outage occurs and your business needs to switch over to a backup AS/400 in as little time as possible, a *cluster utilizing continuously available application*s is required. In addition to OS/400 Commitment Control, such applications are designed to automatically checkpoint themselves (to facilitate restart on the backup AS/400), and make use of new OS/400 Version 4 Release 4 clustering APIs (Application Program Interfaces). This requires a bigger investment in application changes, however, it also reaps the biggest reward in increased availability. AS/400 Solution Providers that offer applications which are ClusterProven™ meet this criteria.

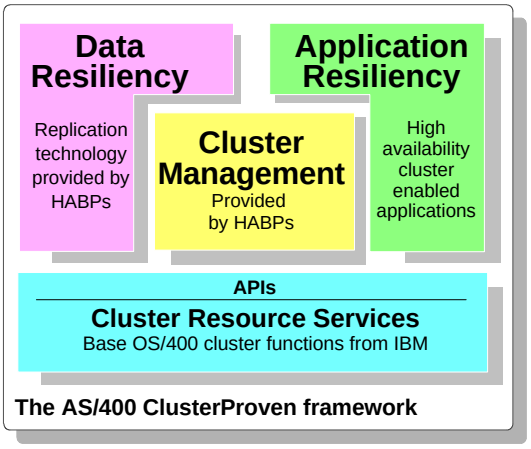
AS/400 Clusters for Version 4 Release 4

V4R4 Introduces a New Level of AS/400 Clustering Support - This support provides a common architected interface for application developers, AS/400 Software providers, and HABPs to use in building high availability solutions for the AS/400. The architecture is built around a framework, and is the foundation to build continuous availability solutions for the AS/400.

Cluster Resource Services, a component of OS/400, provides:

- Tools to create and manage clusters, the ability to detect a failure within a cluster, and switch over and fail over mechanisms to move work between cluster nodes for planned or unplanned outages
- A common method for setting up object replication for nodes within a cluster. This includes the data and program objects necessary to run applications that are cluster enabled
- Mechanisms to automatically switch applications and users from a primary to a backup node within a cluster for planned or unplanned outages.

This clustering framework is built around a set of system APIs, system services, and exit programs. This clustering architecture requires teamwork between IBM and business partners to provide the total solution. Data replication services and the cluster management interface is provided by IBM's HABPs. AS/400 software solution providers deliver ClusterProven applications that utilize AS/400 cluster support, or you can implement AS/400 cluster support within your own applications.



The AS/400 ClusterProven framework

The business benefits of this new level of clustering support for the AS/400 are:

- Easier to plan, implement, and manage clusters.
- A common method for setting up replication services, making replication set up faster and allowing customers more flexibility in choosing an HABP vendor.
- Coordinated switch over of data and applications.
- A more complete solution to maintain high availability of critical applications by reducing the business impact of scheduled or unscheduled server outages.
- A consistent view of application and object replication and availability within the cluster.

How AS/400 Clustering Works - With OS/400 V4R4 a cluster is created and activated by the HABP's cluster middleware via a system API. Activation of the cluster starts OS/400 Cluster Resource Services (CRS). CRS provides the foundation for clustering on AS/400. CRS includes new system objects called Cluster Resource Groups (CRGs), cluster event action codes, and takeover IP address support for transparently switching clients from a primary to a backup node for planned or unplanned outages.

There are two types of CRGs - data CRGs and application CRGs. These CRGs are used to define and manage the resilient objects used in a clustering environment. CRGs are views of information kept consistent across all nodes in a cluster - things like the programs, data areas, data queues, files and directories. The data itself is replicated within the cluster via HABP

replication services. These CRGs can be thought of as objects that provide information about replicated objects. Each CRG has an exit program associated with it. Exit programs govern the actions to be taken when a cluster event occurs and control TCP/IP address takeover. Data and application CRGs are created by the application provider and processed by the High Availability Business Partner clustering middleware.

What Needs to be Done to Make an Application Continuously Available? - To make an application continuously available, you need to:

1. Set up an AS/400 cluster
2. Create and activate data and application CRGs
3. Provide an exit program to process cluster events
4. Provide application restart functions on the backup node

Steps one and two are performed with the HABP clustering middleware products.

The exit program in step three above can be written in any ILE language and may be used by several applications. It defines what action to take when a cluster event occurs (e.g., start clustering, switch over, add a node or other possible cluster events). The exit program also uses a list of which programs and data objects that must exist on nodes in a cluster for switch over purposes. In most cases, this exit program is supplied by the application provider you choose, or by your company's application development department.

Application restart options - step four above - will be different depending on whether the application is host centric or client centric.

Some client centric applications keep much of the application state information in the client itself. If the primary server fails a client may receive a "server unavailable" message and after retrying, the user can re- access the application without even being aware that they have been switched to a back up server. An example would be an application serving static web pages. Multi-tier client centric applications using a database server for transaction processing need to have the database transaction portion of the application kept in sync when the user is switched between nodes.

The user's communications session is automatically switched to the back up host server with AS/400's Cluster Resource Services TCP/IP address takeover.

Host centric applications are more difficult to reposition near the point of failure than client centric applications. There are three possible options for restart on a backup node after failure for host centric applications:

1. The user signs on to the backup node and restarts the application.
2. The user signs on to the backup node and is returned to the last menu option selected.
3. The user signs on to the backup node and is returned to the last completed transaction.

Because there is no application logic available on host terminals, the work station user must perform the sign on to the back up server node after switch over. To use option 2 - returning the user to the last menu option selected - the application must be modified to record users' menu selections and this information must be replicated to the backup nodes. A system administrator or the end user still needs to determine the last replicated transaction when the application is restarted on the backup system but it is still less disruptive than restarting the application.

IBM can help you integrate V4R4 clustering support into your applications. The IBM Technology Solutions Center in Rochester, MN. offers education on V4R4 cluster support, and application design and consulting.

In order to benefit from the AS/400 V4R4 clustering technology, you must select an HABP solution which is built upon the V4R4 infrastructure and is generally available in the marketplace. Since the technology is relatively new, the solutions tend to be in beta installations at this time. The ultimate solution will result from the combination of a ClusterProven application and the high availability business partner software featuring the AS/400 V4R4 clustering technology

AS/400 ClusterProven™ Applications

ClusterProven is an IBM Server-wide program that recognizes applications which continue to be available in the event of a failure. It recognizes true availability cannot be achieved without application involvement. The ClusterProven program is designed to increase the high availability application portfolio and raise the availability bar to move towards continuous operations. Each IBM server has its own definitions for ClusterProven that relate to the specific platform.

To be ClusterProven, an application must reposition the user to the last main menu option selected prior to failover or switchover. In order for an AS/400 application to achieve ClusterProven status, it must:

1. Identify data and objects needed to run the application.
2. Write an exit program to handle cluster events (guidelines are provided in the AS/400 ClusterProven specifications at web site listed below).
3. Provide application restart.

Planning for a successful AS/400 Cluster

The number one success factor - is understanding that deploying a replicated data environment or cluster is not the same as deploying other products. It is the start of a new continuous process for your organization. There MUST be a commitment from both business management and IT to insure a success. This commitment should also include a commitment to training and a knowledge transfer during the implementation to ensure you understand all of the technical and business impacts of the project.

The second success factor - is deploying a project team that covers your entire IT infrastructure. While deploying the cluster is primarily the responsibility of the operations department, it will not be successful without the involvement of application development, networking, database and end users.

The third success factor - is the project plan. A successful project plan will include:

- A definition of the scope of the deployment
  - Identify the critical business processes in your enterprise, and classify the applications within those processes as to their impact to the business process. Develop a policy that guides developers and operations on availability requirements for each application classification.
  - Business units and IT should jointly take steps to develop a high availability and business continuity strategy for critical business processes. A key step in that process is understanding downtime costs for critical business processes.

- For each application suite that will be replicated:
  - What database networks, programs, and other objects are required for operation?
  - What subsystems, devices, program starts, etc., are required for the application to be operational?
  - Is the application well behaved, highly tuned, and scalable?
    - Has the application been thoroughly analyzed for its performance characteristics?
    - Has the application been bench marked utilizing the Rochester Customer Benchmark Center?
    - The following techniques will impact the deployment of the HA solution. Does the application or the database employ them?
      - ✓ check constraints, referential integrity, and/or triggers
      - ✓ spool file commitment control, data queues, and/or data areas
      - ✓ user spaces, and/or message queues
      - ✓ temporary objects or work files
    - If the database network for the application is not journaled, plan the implementation of journaling, utilize appropriate techniques to minimize the impact of journaling.
      - ✓ Understand that the best performing journal environments deploy the journal receivers in a user auxiliary storage pool (ASP) that is protected with OS/400 disk mirror and contains a minimum of 3 usable disk arms (6 mirrored disks) and maximum of 15 usable disk arms (30 mirrored disks). Note: OS/400 Version 4 Release 5 may increase this maximum even further.

- ✓ When multiple journal/journal receiver pairs are required, the receivers are deployed to different user ASPs based on analyzed performance requirements.
- ✓ The attribute for system managed is set on each journal.
- ✓ The journal object is not in the same library as the files that are journaled. (This enables library rename on the fly.)

- Education plans are put in place and executed for
  - Operations management personnel
  - Lead applications personnel
  - System operators (especially night and weekend operators)
  - All members of the implementation team
- Alter change management practices to incorporate new replicated data environment (e.g., programs are moved into production on BOTH systems or HABP software is used to replicate changed programs from production to backup).
- For extremely large, complex cluster proposals, consider bench marking the solution at the Rochester Customer Benchmark center.

Regular, successful tests - Because the AS/400 is such a reliable system and failures occur seldom, your staff will not always have a chance to gain experience switching from the primary AS/400 to the backup. Regular successful switch tests are the final element of a successful implementation. Simulated failure testing or planned maintenance switches performed on a regular basis enable you to be confident that your HA solution is working for you.

If you've chosen to set up your cluster environment such that the backup system is utilized as a standby secondary, you could consider the following general plan for regular switch overs:

1. Apply PTFs and/or release upgrade to the backup system, allow these changes to burn in and be fully tested.
2. Perform a planned switch from the primary to the backup. Allow the users to run in this environment for a week.
3. Perform same changes on the primary.
4. Either allow the systems to stay with their current roles or perform a switch back to the original roles.

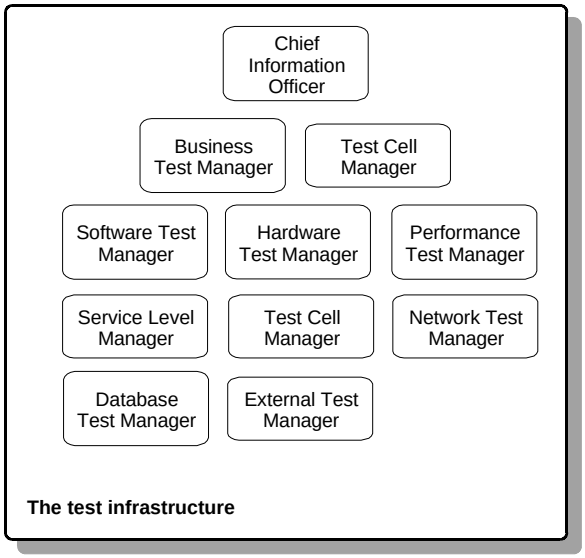
Testing a Clustering Solution

By employing AS/400 clusters, your business has made a serious commitment to exceptional levels of availability. These levels are not only reliant on hardware and software, but can only be achieved with stringent problem and change management. Hence, identified problems must be replicated. If a change is put in place, this change must tested before further risk to production.

Defining and implementing an HA test environment - The highly available business in its truest sense should be an end-to-end entity. Therefore, when planning the test environment a holistic approach should be taken. This means that not only are the obvious things like hardware, network and applications tested, but also the less obvious components. Examples of these components are the business processes associated with the computer systems, the facilities the business uses, data and applications from external providers, and accurately documented job responsibilities.

One must also consider whether cluster environment is being planned or already exists. If a company was just starting to plan their HA solution, there is considerable implementation testing to be performed. If the HA environment exists it may be a matter of verifying that all recovery conditions have been documented and a test environment created. If this is a new solution, then much of the implementation testing can be carried forward to the problem, change management processes of the production systems.

Test Infrastructure - One of the first tasks to do is create the test infrastructure. This consists of small group of business and IT personnel. One member of the group should have overall responsibility for the test facility, and another very important member is the business test manager. The business test manager should be from a business unit, and have sufficient status to demand the attention of board members. Once these members have been appointed other members can be selected. For example, development test, hardware test, network test. External organizations should not be ignored. Make sure anyone who provides data, service or applications to your business or receives data or service from your business is involved in testing. These organizations may not be involved in all tests, or be required to attend meetings, but they must be considered. Define all roles, responsibilities and escalation paths within the test infrastructure.



The test infrastructure

Environment Inventory - Once the infrastructure group is formed then tasks can be assigned. The group can then inventory all the components within the E2E environment. This documentation will also form the basis of a good disaster/recovery plan. Once all the components are known, then create a test cell. This should be an integrated process between the business and IT. An example of this documentation that is often overlooked, is an accurately documented job responsibility and role of each component and member of the business. While these components can be well documented internally, some firms have used an outside employment agency to interview staff as if they were trying fill a position. This information is then held as a DR asset, and could be used in the event of a major disaster with loss of life.

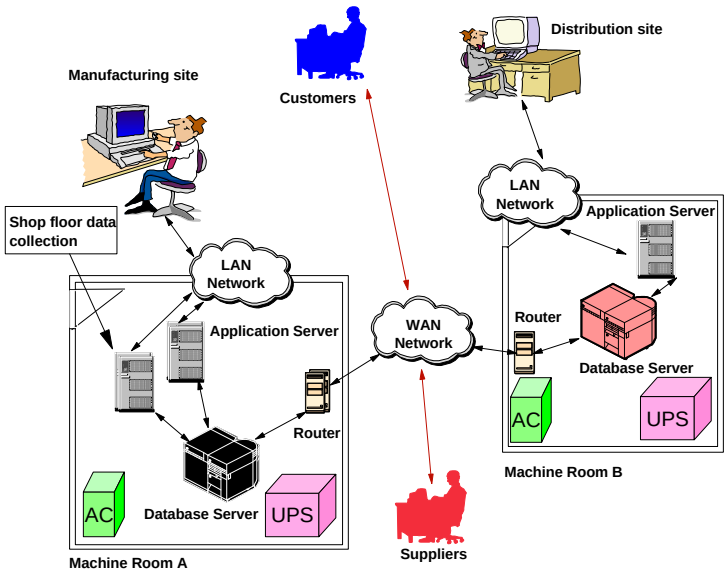
Test Cell - A typical test cell should include replicants of the systems, and components that exist in production. These replicants do not necessarily need to be equivalent size, but should be able to perform the same functions. Volume testing is a separate topic that will be discussed later.

Start with the simple stuff. Make sure there is a true representation of all applications on the systems. A copy of Gold code (e.g., final code version) must exist, too often changes are made to the application and these changes are not reflected in the gold code. This creates the sort of integrity problem that causes recovery to fail. A strict development cycle, with control over changes, movement into production, gold code update must be implemented.

With hardware, make sure that all critical and active components are replicated and if possible form part of the test cell. As new hardware is added, it too much be replicated into the test cell. In some cases if the component is very expense, a loan or rental may be available for testing.

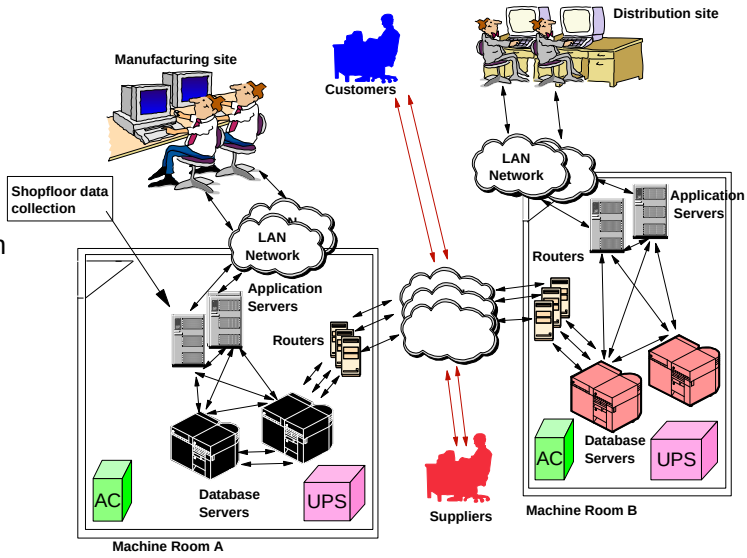
Volume testing - Volume testing is another key element of pre-production testing. Making changes, such as adding a new application or some new hardware can create a potential for loss in availability through degraded performance. If the value of the availability is high, you might consider a dedicated performance test facility like the Rochester Benchmark Center. There, applications can be tested on the latest hardware models, or after significant application changes or additions have occurred, verifying that the changes meets your agreed service levels.

Moving to an HA environment - This illustration depicts a business without an HA implementation. There are two sites with separate but dependent processing environments. Site A is a manufacturing site with a three-tier model, clients, application servers, and database servers. Site B is a distribution site. There is shop floor data collection input to the application. Customers and suppliers both have EDI links to both sites through the WAN, and information flows in both directions. As can be seen, this is already a fairly complex processing environment, and one that desperately needs to be highly available. Looking at this scenario, suppose there is a change to the application, and that change creates an unexpected problem for the end-user. The user is unable to perform the business function, and in effect that particular part of the business becomes unavailable, along with any associated business functions. Had the application been fully tested, the unexpected problem would have been detected and rectified before the change was implemented in production.



A non-clustered environment

The next illustration shows the same scenario, but made highly available. Now we can see the true level of complexity of a very highly available environment as it applies to the users and hardware. It does not show the business process level. In this scenario the database servers have been given a switch-over/fail-over backup. Application servers and database servers are a four-node mutual takeover cluster. From a network perspective, three network providers connect the sites, suppliers and customers. The user population has been split across multiple LANs providing access for some users in the event that a LAN fails. This scenario contains many different components and multiple sites.



A fully clustered environment

End-to-End testing in production - In this environment testing is both important and complicated. If a change or problem occurs on any component, it must be tested before being brought into production. For example, if the IT group decides to upgrade a router, they should be load tested to ensure that they meet the performance required to support the service level agreements. Once these tests have been carried out the new routers are installed in pairs. As there are three routers in the network, resiliency is only partially degraded. These routers will then be monitored for a couple of weeks to finally check that there are no unexpected problems. Once the IT group has accepted these new components, the other pairs of routers are swopped out.

The database and application server is another critical resource that needs careful handling when making changes. Let's look at a database server, and a change needs to be made to the database. There is a definite requirement to test the change before implementing in production.

Testing Examples - The following are examples of testing activities that could be carried out in a highly available business.

Business process testing

- Annual DR testing There should be a full DR test at least once a year. It very important that all parts of the business understand what needs happen in the event of a major failure. The only way to achieve this is by familiarity.
- Telephone systems - after upgrade or relocation, testing the phone system is a very critical part of testing. Can customers still use 1-800 numbers connect to the remote disaster site. If there is a failure on the local PABX, are there direct lines available in key departments to allow outgoing calls?
- Staff access (physical access to local and remote locations). If there is a security system in place, does it allow access during a system failure. When the remote location is many miles away, what is the method for moving staff to the remote location and accommodating them.
- Customer/supplier access (network and telephony). Alternative numbers and system access point should be published to customers and suppliers.

- Emergency power - many systems will have UPS support to the machine rooms. However this is no good if there is no power to office space. Client UPS or limited generated UPS to some areas should be considered.

Integration testing

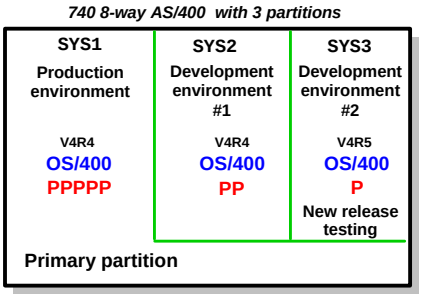
- Hardware components (new and upgrade) - as new hardware or upgrades are added. These must be tested. Some functions will be able to be tested in the test cell, others may require testing during the upgrade process.
- Operating systems upgrade (servers and clients) - most of these tests can be carried out in the test cell.
- Application upgrade - after an application upgrade, new functions should be tested within the business structure. The performance under load conditions should be tested. These load conditions should be representative of peak business periods and for systems that are expected to have the ability to accept workload from other systems during switch/fail over.
- Database integrity (disaster, upgrade, switchover) - if there is a switch/fail over the business must be able to test the integrity of the database. In planning for these tests, establish key business indicators that can be reviewed on restart. Develop a plan for managing the business if the integrity can not be established in a very short period of time.
- Interactive job (restart and performance) - interactive jobs lose much of their individual environment during a failure. Test the recovery of these jobs after switch/fail over. After a change to the application, hardware, network ensure the performance of interactive work meets expectations.
- Batch (restart and performance) - batch is normally more complicated than interactive. Jobs can be single or multithread, short or very long running. To test these jobs the test cell will be vital. Allow the jobs to run to completion and check their results. Test performance on similar hardware to the proposed solution.
- Backup subsystems (recovery and performance) - backup management during change and testing is critical. It can be very easy to get out of sync. Normal cyclic backups may have to be supplemented to retain similar levels of backup redundancy during major changes. This is particularly pertinent to staged or phased upgrades and where the tape backup window is very small.
- Access to systems internally and externally - there should be sufficient redundancy to allow changes to the network where old and new components can coexist in product for a short period after the change.
- Application/database switch/fail over - in a clustered solution the clustering functions should be tested along with the changes made to other parts of the business or systems.
- Replication testing (performance, apply process function/start/stop) - similar to the cluster testing, replication performance and function must be tested after changes to other components.

Complementary Availability Solutions for the AS/400

An AS/400 cluster not only offers flexibility in application availability but, in the configuration of the systems used in the cluster as well. Other AS/400 hardware and software options that complement AS/400 availability:

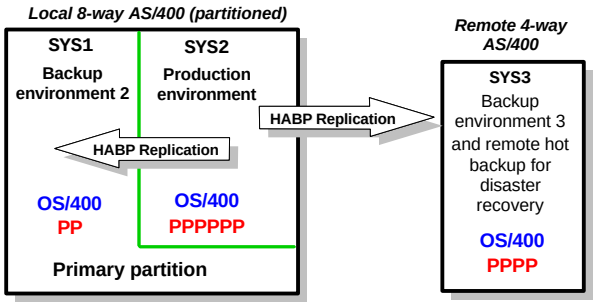
AS/400 Logical Partitioning (LPAR) - Available beginning with V4R4, AS/400 logical partitions let you run multiple independent OS/400 instances or partitions (each with its own processors, memory, and disks) in an n-way symmetric multiprocessing AS/400, model 6xx, Sxx, 7xx and 8xx. With logical partitioning you can address multiple system requirements in a single machine to achieve server consolidation, business unit consolidation, and mixed production/test environments.

By itself, LPAR does not provide a significant availability increase. It can however, be used to complement other availability strategies.



A basic LPAR configuration

Since each partition is treated as a separate AS/400, you can run a cluster environment on an single system image. Up to twelve cluster nodes (one per CPU) can exist within one LPAR system. This can provide for a more cost efficient clustering solution, however, an LPAR clustered environment does increase single points of failure. For example, if the AS/400's primary partition becomes unavailable, all secondary partitions will be unavailable as well (though the opposite is not true).
The highest level of availability will be obtained with two separate AS/400s. In some environments, LPAR lends itself well to situations where both a local and remote backup AS/400 is desired. A good example is when a business wants to provide it's own disaster recovery capability. In this LPAR example, a single 8-way AS/400 is partitioned into two severs (a 2-way and a 6-way). A second AS/400 (4-way) in a remote location is put in place to provide disaster recovery capabilities. With clustering active, data is efficiently and simultaneously replicated to both the local backup AS/400 and the remote AS/400. The local backup AS/400 can now be used for local nightly saves, hardware and software upgrades and software maintenance. In the event of a disaster (or the need for the entire local hardware to be powered off), the remote backup AS/400 is available. In some cases, this more cost efficient (including floor space), than separate servers.



LPAR and a local and remote AS/400 cluster

Integrated AS/400 availability options - The AS/400 offers many integrated availability technologies that complement an AS/400 clusters solution. In most cases, it is recommended that these solutions be used with an AS/400 cluster to further mask or reduce downtime, and to increase a cluster's efficiency.

- **AS/400 Disk protection** includes AS/400 *Device Parity Protection* (RAID-5) and OS400™ *Disk Mirroring*. In the vent of a disk failure, these disk protection methods prevent data from being lost, and allow concurrent maintenance of disks on your system.
- **Auxiliary storage pools** (ASP) are a group of units that are defined from all the disk units that make up auxiliary storage. It is a software definition of how the disk units are arranged. You can use auxiliary storage pools to separate your disk units into logical subsets. User

ASPs allow you to isolate objects on one or more specific disk units. User ASPs are used to manage system performance and availability requirements:

- Provide dedicated resources for frequently used availability objects, such as journal receivers.
- Allow on-line, unattended saves.
- Place infrequently used objects, such as large history files, on disk units with slower performance.

- **Access path protection** - An access path (view) describes the order in which records in a database file are processed. A file can have multiple access paths, if different programs need to see the records in different sequences. If your system ends abnormally when access paths are in use, the system may have to rebuild the access paths before you can use the files again. This can be a time-consuming process, since an IPL on a large, busy AS/400 that had ended abnormally may take many hours. Two methods of access-path protection are available:

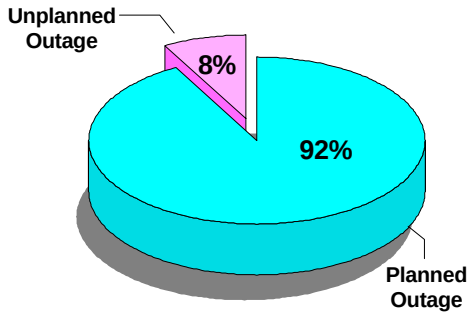
  - OS/400 System Managed Access Path Protection (SMAPP)
  - Explicit journaling of access paths

- **Concurrent maintenance** - Depending on AS/400 model, concurrent maintenance allows disks, I/O Processors, adapters, power supplies, fans, CD-ROM, and tape to replaced without powering down the AS/400.

In all cases, it is highly recommended that these integrated availability options be used in an AS/400 clustered environment, as well as on a stand-alone AS/400.

Competitive Availability Solutions

Not too far in the past, the primary concern with server availability has been with unplanned outages - unpredictable failures that caused a significant outage of varying duration for the business. For non-AS/400 servers, recent studies suggest that 20% of all outages are unplanned. For the AS/400, unplanned outages probably account for no more than 6-8% or less of server downtime. The remaining 92-94% of downtime are the results of planned outages for backups, maintenance, upgrades, etc. Other server platforms must rely on clusters to even approach the level of availability that a single AS/400 can provide.

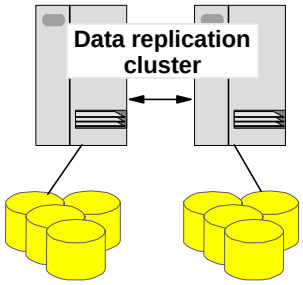


Typical AS/400 Planned outage frequency

The number of unplanned outages will continue to shrink for the entire industry as hardware and software technology become more resilient. Although unplanned outages must still be eliminated, in the desire to achieve 24x365, planned downtime has now become a primary focus point. Clusters provide the ability to mask planned outages as well as unplanned.

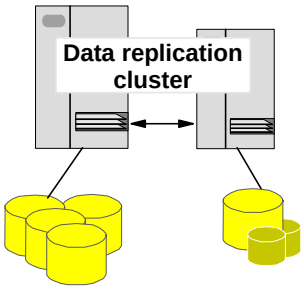
Comparing the solutions: *AS/400e Servers*

Data replication cluster - A cluster that utilizes data replication is the best solution for approaching continuous availability requirements, providing rapid recovery for the *widest range of outages possible*, with minimal cost and maximum application flexibility. Sometimes referred to as a non-shared or shared-nothing cluster, data replication provides the ability for a cluster to eliminate or mask both *unplanned outages* (hardware and software failures, human errors, disasters, etc.), and *planned outages* (application maintenance, software and hardware upgrades, backups, etc.).



Typical data replication cluster

AS/400 clusters accomplish redundancy by utilizing Data Replication. Each AS/400 in the cluster has it's own set of independent disks. Data is replicated from the primary AS/400 to a backup AS/400, and the cluster is managed using one of IBM's High Availability Business Partners middleware products (*DataMirror, Lakeview Technology, and Vision Solutions*). This allows applications and data to be available for use on the backup AS/400, should the primary system fail or need to be taken off-line for planned maintenance (e.g., backups, application changes, etc.). This type of cluster provides control over all types of outages, and the ability to provide near continuous availability. There are additional benefits to a cluster that utilizes data replication. Since each system can be an independent configuration the backup (or secondary) AS/400 may be a different size, with different amounts or types of disk units, and even a different release of software. In many cases this allows you to protect your investments by better utilizing existing hardware, reducing costs when implementing a cluster. Equally important is that a data replication cluster is the only availability solution that provides realistic remote disaster/recovery capabilities. In

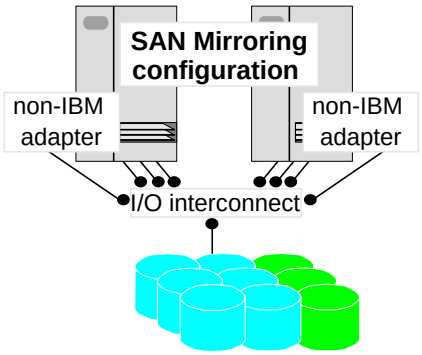


A data replication cluster allows different configurations

many cases, standard communications will allow you to place a backup AS/400 on another floor and/or power grid, or in an even safer location…across the complex, in another city, or even around the world. Other availability solutions are simply limited in the distances they can span, reducing their feasibility as a disaster/recovery solution.

SAN mirroring configurations (external disk) - There are features that SAN (Storage Area Networks) vendors offer which are targeted at system availability enhancements. One is aimed at minimizing the impact of system backups on availability, and another offers a disaster recovery option. Carefully understand what your availability requirements are and what the SAN environment can offer in this arena. Are your primary concerns planned outages? For example, to perform a save operation using this SAN feature, your system will need to be brought to a restricted state (an outage) prior to detaching the objects from the host environment. This is necessary to insure that you have a "clean" consistent copy, a snapshot of your system data at the time of the operation. And depending on the SAN configuration, during this period the disk units may be left in an unprotected state, creating an exposure for a very long, unplanned outage should a single disk unit fail.

A SAN disaster/recovery configuration has the benefit of being somewhat simpler than a cluster but, there are recovery and performance issues with this solution. The first issue to consider is this solution must implement synchronous, remote disk mirroring. This is required for several reasons, the primary being to assure that data is written in the same order on the backup system as it was on the primary system. The greater the distance between the primary and backup servers, the more the response time on the primary system is impacted. An AS/400 cluster utilizing data replication offers greater efficiency and does not encounter this problem (since only the changes to critical business data are replicated, there is less of a performance penalty, and the systems can be separated by greater distances, such as another city, state or country)



Typical SAN mirroring configuration

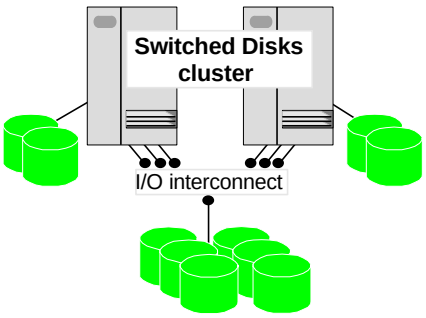
The second issue is the potential for toxic data. In his book entitled "In Search of Clusters", Gregory F. Pfister defines toxic data best, by describing two servers, Alice - the backup server, and Bozo - the primary server: "*Alice (the target) now has physical access to the data. Data that was last written by Bozo. Dead Bozo. Bozo who may well have been deranged before kicking the bucket. How do we know he didn't scramble the data in its death throes?*"[(2).] For an AS/400, toxic data can represent objects on your system that become damaged due to a system failure, or it can result in incomplete transactions that leave your databases partially updated. The AS/400 has features that help to reduce the possibility of damaged objects, but these capabilities can be circumvented in some cases in a SAN mirroring configuration so additional recovery on the backup AS/400 may be required. To eliminate incomplete transactions, OS/400 Journaling along with commitment control must be implemented.

Finally is the issue of price/performance. When attempting to utilize a SAN mirroring configuration for disaster/recovery you effectively end up with twice the processing capacity installed, but are effectively utilizing only half of it. The backup copy of data that is there for disaster protection, can not be accessed for real time processing such as off-line batch or queries as can be done with an AS/400 data replication cluster.

(2) "In Search of Clusters", Gregory F. Pfister, 1998 Prentice Hall PTR, ISBN 0-13-899709-8

Comparing the solutions: *non-AS/400 Servers*

Switched disks cluster solutions - Other (non-AS/400) server platforms increase availability by utilizing cluster technologies such as server controlled *switch disk*s. Switch disks use some form of external storage (e.g., SCSI or fiber connected) with a mechanism that allows the disks to be switched between servers in the cluster, in the event of an outage. Cluster management middleware software manages the switchover, and for high availability purposes insures that no two system access the disks (e.g., data) at the same time.



A properly designed switch disk cluster can offer some advantages over a data replication cluster. Because a switch disk cluster does not use data replication, there is slightly less overhead on the systems and more available for processing of transactions. In the event of an unplanned outage (such as the loss of memory or a processor), fail over to the backup system may be slightly faster in some cases since it is not necessary to wait for data replication to complete on the backup sever. A switched disk cluster can be simpler to operate.

	Replication		Switched Disks
+	Easier to add to existing single system		Harder to add; must alter system (disk cables, fiber, adapters, etc.)
+	Easier to configure; systems are independent		Harder to configure; disk configurations must be synchronized
+	Nodes can be long distances apart, providing disaster recovery		Nodes must be physically close, disk in one place - single point of failure
+	Can use any combination of I/O controllers (e.g., disk adapters)		Requires disk and adapters that actually implement their specifications
+	Provides more options for planned outages - less planned downtime		Single copy of objects reduces planned outage capabilities
+	Complete application maintenance independence and granularity		Limited maintenance, controlled by disk configuration granularity
+	Backup system can be productive and process other workloads		Backup system must typically remain idle
+	No single point of failure, (requires additional disk storage)	+	Shares a single copy of disk storage, (single point of failure)
	Uses CPU and I/O capacity	+	Little overhead in normal operation
	Tight synchronization can cause performance loss; Loose synchronization can lose data	+	Tight synchronization, without performance loss
	Fail back requires resynchronization	+	Fail back issues are the same as fail over

Data Replication Cluster vs. Switch Disk Cluster

However, there are limitations in a switched disk cluster. Because switched disk solutions rely on a single copy of the data, they face problems masking planned outages and outages for planned maintenance are still incurred. The ability to provide application maintenance without causing an outage is reduced or not possible with a switched disk, since data and objects are in

use by the primary system. In the case of a release upgrade, many application solution providers do not offer n-x release support, so a switched disk solution will require extended downtime during the application upgrade. Many hardware upgrades (including changes to disk configuration) will still require downtime of the entire cluster. Similarly, concurrent backups from the secondary server may not be possible and will usually require that either the application or entire cluster be stopped.

Summary

In many cases, an AS/400 cluster solution is coined as too expensive since it requires file journaling which can add overhead to the system. The fact is that all solutions, whether another platform's switched disks cluster, or a SAN mirroring configuration requires the use of journaling (or as it is called on other server, logging) in order to insure data integrity.

Once the value of data replication is understood, AS/400 clusters become a logical choice for maximizing system availability, since they provide superior control over both planned and unplanned outages.

The Hidden Benefits of AS/400 Clusters

Clusters can actually be an efficient solution to not only availability issues, but in some cases, server capacity and performance issues as well. Depending on your business availability requirements, the backup AS/400 in a cluster can be utilized in one of two modes:

1. **Idle backup AS/400** - The only active workload on the backup AS/400 is the HABP cluster middleware (providing data replication and cluster management). In the event of an outage, this server stands ready for a switch over and to assume primary AS/400 production processing responsibilities. Typically this provides for the fastest recovery in the event of an outage on the primary AS/400, since no time is required to manage or end workload on the backup AS/400.
2. **Active backup AS/400** - In addition to the HABP cluster middleware, other work takes place on the backup AS/400 allowing it to be productive throughout the day (ideally this is 'read' only type of activity, in order to avoid interference with the data replication processes). By utilizing the replicated databases on the backup AS/400, these 'read only' type of workloads can be relocated to the backup AS/400, making use of available CPU and system resources. Functions that impact the availability, capacity and performance of the primary AS/400, may be candidates for relocation to the backup AS/400. While providing additional processing capabilities, this method will usually increase the amount of time, and the complexity, to switch over to the backup AS/400.

For example:

- **Provide query, reporting, and inquiry capabilities any time of the day without impacting the primary AS/400** - Adhoc queries, reporting, web-enabled customer inquiry, and Business Intelligence (BI) can have an impact on the primary (production) AS/400 performance. By utilizing the replicated databases on the backup AS/400, 'read only' types of workloads can be easily relocated to the backup AS/400, making use of available CPU and system resources. In the case of BI, the heavy read only workloads such as BI's data extraction, data transformation, and Data Mart builds can be relocated to the backup AS/400. In some BI environments, this may also help your networks by eliminating large data movements, since BI can now take advantage of the replication databases already on the backup AS/400.
- **Perform system maintenance without impacting production:**

  1. *Nightly saves with reduced impact to production work* **-** Even with fast, multiple tape drives, and OS/400 save options such as Save-While-Active, backup windows that require quiescing production work are becoming a problem. By making use of the data replicated on the backup AS/400, most backup activity can be relocated to the backup AS/400, reducing planned downtime on the primary AS/400.
  2. *Perform PTF and application maintenance while reducing the impact to production work* **-** The backup AS/400 is available to assume production work, while the primary AS/400 has regular PTF (program fixes) applied, or your business applications have maintenance performed. Once completed, production work can be returned to the primary AS/400, all with minimal interruption to business.
  3. *Perform release updates and software or upgrades and system upgrades while reducing the impact to production work* **-** The backup AS/400 is available to assume production work, while the primary AS/400 has its software (such as OS/400 and related programs) upgraded to a new release, or when a new hardware upgrade needs to be performed. A benefit of the data replication method of clustering is that is supports a wide variation of versions and releases between systems, virtually eliminating the possibility that both systems must be upgraded at the same time. Once completed, production work can be returned to the primary AS/400, all with minimal interruption to business.

- **Eliminate development and testing from the production AS/400** - Periodically, an outage may occur due to application development and testing on the primary (production) AS/400. Whether it is a planned or unplanned outage (e.g., accidental alteration or deletion of production data), it is probably an outage your business cannot tolerate. A safer place for application development and testing is on the backup AS/400, where an outage does not immediate impact availability.

The ability to utilize a backup AS/400 in your cluster for other workloads (such as those mentioned above), will depend on your availability requirements, configurations, spare capacity and performance of the backup AS/400, and the training of your operations staff on the HABP clustering solution. During the design phase of your cluster, you should work closely with an HABP consultant or an IBM Integrated Technology Services (ITS) consultant to ensure that failover and switchover capabilities meet your objectives.

.

Selecting an AS/400 Clustering Solution Provider

The good news is that for your clustering solution you have several choices. The bad news is that you have choice to make! IBM has three High Availability Business Partners (HABPs): DataMirror, Lakeview Technology, and Vision Solutions, and each offer a very competitive, robust cluster solution.

1. The AS/400 clustering solution: Each HABP has their own suite of products that make up their high availability solution. It is accepted that each one is packaged differently and it is not possible to draw a direct comparison between them, however when reviewing them, each solution proposed should offer the following functions:

- *Do the products support the latest technologies ?* - Each AS/400 product cycle will bring enhancements designed to make the high availability /clustering environment more robust and efficient. As technology continues to evolve it is crucial that the HABP solutions adopt and support these emerging technologies in a timely manner. For example, remote journaling, MQ Series support and integration, and the Integrated File System.
- *Robustness* - There are various ways for a clustered environment to fail. Objects on the secondary and primary systems being out of sync is an example. What tools does the HABP provide to prevent and/or detect problems that may prevent you from using a backup node for an unplanned outage.
- *Project discipline* - Is the HABP able to demonstrate a sophisticated level of project management? Does the HABP understand the total solution including the application environment? Are there services offered to analyze the entire system environment including the application? Even when the consulting services are provided by IBM's Integrated Technology Services, project discipline must be a part of the design discussion and a step in the process to achieve the end objectives.
- *Fast replication* - Changes that are made to the production machine need to be reflected on the backup machine in a timely manner. The solution should be capable of handling the calculated workload that is going to be generated, and this should also allow for different processing environments such as the online daytime interactive and overnight batch processing. You should also be provided with the ability to prioritize replication workloads on the system.
- *Transaction sequencing* - Changes that are made on the production machine need to be reflected on the backup machine in the same sequence as they are generated. The solution should ensure this occurs by providing a built in integrity checking and if any problems are encountered, operations are notified quickly (e.g., Message, pager, etc.).
- *Role swap* - It should be possible to change the direction of the replication using tools supplied as part of the solution. For example, should the production machine require a software upgrade, the backup machine should be able to take over the workload of the production machine and store any changes generated so the production machine can be synchronized later when it comes back online.
- *Complete environment replication* - The AS/400 has many different object types. The solution needs to be able to replicate any changes made to any object in the production environment to the backup environment. For example, a user signs on to the production machine, changes his password and then performs some database changes. The database changes are replicated to the backup machine but the password change isn't. In the event that the production machine fails the user will not be able to sign on to the backup machine.
- *Progress tracking and reporting tools* - The solution should be able to track and report the status of the production and backup AS/400. It should be able to report how many changes have been generated on the production machine, how many have been received by the

backup machine and how many have been applied on the backup machine. This information should also be available historically.

- *Error handling* - The solution should be able to ensure delivery from the production to the backup machine. If data is sent from the production machine to the backup machine, but in the process the communications link fails, then the solution should be able to re-send the data that has not been received on the backup machine.

2. What references are there ? - References should be available for environments with similar volume requirements. If third party applications are being used, where possible, the references should be for the same applications.

3. What implementation assistance is going to be given ? - What assistance is going to be given to implement the solution? Implementing an HABP solution is not just about installing and configuring software. There are many more steps involved which, depending on your staff's skill and experience levels, require different amounts of assistance.

4. What product education is available ? - Detailed training should be available for the proposed solution. This should cover the setup of the environment, changing the configuration of the environment (switching), problem determination procedures and the recovery processes required in the event that a controlled switch is made.

Ideally, this training should be tailored to your business and operational environment.

5. What is the ongoing support structure ? - Once the solution has been installed an ongoing support function needs to be provided. In the event that you have an urgent question or encounter a problem with the solution, you do not want to have to wait until the start of the next working day before you can ask someone about it or get the problem resolved.

- What ongoing support is available?
- If the support is provided via a local agent, does the agent have a support channel to the HABPs
- Is there a 24 x 7 help desk that you can call?
- Is on site assistance available in a timely manner, if required?

6. Are the products easy to use ?

- Is it intuitive?
- Is it easy to navigate around the display formats?
- Is it possible to imbed functions in batch jobs or are they command line interface only?
- Does it offer an interface that is similar that your staff is comfortable with ?

7. What is the total cost ? - Obviously, one factor in the selection process is the cost of the solution. However, the initial cost alone should not be a primary deciding factor. Consideration should also be given to ongoing product maintenance costs, upgrade costs for future releases and support costs.

Appendix

IBM Web Sites:

IBM ClusterProven™ program	http://www.ibm.com/servers/clusters
AS/400 main home page	http://www.as400.ibm.com
AS/400 Availability home page	http://www.as400.ibm.com/ha

IBM High Availability Business Partner Web Sites:

3100 Steeles Ave E, suite 700
Markham, Ontario, Canada L3R 8T3
800-362-5955 or 905-415-0310
web page: http://www.datamirror.com
e-mail: info@datamirror.com

2301 W 22 St, suite 206
Oakbrook, IL 60523
630-537-0440
web page: http://www.lakeviewtech.com
e-mail: info@lakeviewtech.com

17911 Von Carman Ave, 5th floor
Irvine, CA 92614
800-683-4667 or 949-253-6500
web page: http://www.visionsolutions.com
e-mail: info@visionsolutions.com

Publications References:

1) “So you want to estimate the value of availability” (IBM publication GG22-9318)

2) “In Search of Clusters”, Gregory F. Pfister, 1998 Prentice Hall PTR, ISBN 0-13-899709-8