

HIGH AVAILABILITY AND CLUSTERING SOLUTIONS FOR IBM eSERVER iSERIES AND IBM AS/400



DataMirror[®]
THE EXPERIENCE OF NOW.[™]

High Availability and Clustering Solutions For IBM eServer iSeries and AS/400

Executive Summary

In the new economy, data drives all business. Data is precious and making sure it's available to your customers, decision makers, employees and associates 7 days a week, 24 hours per day is critical. Why are some businesses able to rapidly react to unplanned system downtime while others are forced to close their doors? Prepared companies have acquired technology and taken steps to ensure that their valuable data and mission-critical applications are thoroughly backed up and continuously available anywhere in the world in real-time. These companies have implemented high availability solutions for their IBM AS/400 and iSeries environments.

In the Internet era, organizations simply can not afford to lose either valuable data or the use of that data. Disasters strike suddenly and only prepared companies remain open for business as if nothing happened or with minimal inconvenience. No one can predict natural disaster, system failure, user error or illegal intrusion. Unplanned outages aside, every system must be taken off-line periodically for routine maintenance such as hardware and software upgrades.

What is high availability? By definition, the goal of high availability is near continuous access to the critical data and applications that drive the business. But how does it work and who needs to worry about it? What are the implementation options? What should organizations look for in a high availability solution?

These are just some of the considerations that must be taken into account when developing a high availability strategy. Perhaps the easiest to answer is: who should worry about high availability? Every organization that relies on data to effectively and efficiently conduct business must be concerned about ensuring high availability. In some industries, such as the financial services and insurance industries, regulations and laws mandate companies to implement high availability systems. In these cases, companies may be held accountable if system outage affects their ability to service consumers. In the retail sector, and especially in e-Business environments where companies are selling goods and services via the Internet, high availability is a common sense imperative. The remaining questions are discussed in detail in the pages that follow.

This paper discusses high systems availability for companies with IBM eServer iSeries and AS/400 computing environments. It provides a high level overview of factors driving the increasing need for high availability solutions and a brief overview of high availability terms and concepts including logical partitioning and clustering. It also provides information regarding options for high availability, what to look for in an AS/400 availability solution, and an expanded section on the features and benefits that DataMirror High Availability Suite and iCluster software offer companies.

Why back up iSeries and AS/400 systems?

Although availability of iSeries and AS/400 systems is reportedly as high as 99.8 percent, things can still go wrong. With today's global business hours, and the growth of 24/7 e-Business, many companies must be on-line all the time. This means that AS/400 organizations must have a contingency plan in place for ensuring continuous business operations.

Downtime is at best an inconvenience to users. At worst, system failure can result in lost business, lost opportunities and diminished customer loyalty. If system outages prevent users from completing tasks and conducting business via corporate systems, the long-term damage may be expensive and may even be beyond repair.

Even if systems are used primarily for increased collaboration within the organization, the adverse affects of downtime can be devastating. For example, an outage can make it difficult or even impossible to share order, inventory and other information with suppliers, partners or customers. If users are unable to access business data stored on servers, decision making, business monitoring, and other critical functions become impossible and the promise of unparalleled collaboration created by computing systems remains unfulfilled.

Cost of Downtime

Cost of downtime is the key factor in deciding on a high availability strategy. Calculating the exact cost of downtime is difficult given the range of variables and complex network of dependencies that IS infrastructures generate. However, most organizations have the information required to make a reasonable estimate of cost per hour of system outage.

The elements required to make a reasonable estimate on cost per hour of downtime involve losses in both labor and revenue, and any other costs incurred due to the organization's inability to access data. For example, lost access to an orders or inventory database is much more costly during peak business hours than at 2:00 am. Lost customer goodwill and loyalty must also be factored into the equation. Exactly how much is a loyal customer worth?

What is 0.1% of downtime actually worth? There are approximately 8,760 hours in one year. One percent of that is 90 hours and 0.1% amounts to 9 hours of lost access to the data and applications that drive your business. When you consider how critical these data and applications are to the success of your business, it is clear why you should seriously consider purchasing a proven AS/400 high availability solution.

Labor Costs

Part of the formula for calculating cost of downtime includes totaling the cost of normal operations for all staff affected by system outage. This involves estimating the extent to which employees are affected by downtime. Are some employees able to carry out other tasks during outage? What percentage of their normal assignments are they able to complete? Order processing departments, for example, may be totally reliant on continuous access to data about inventory, shipping, and cataloging information. Many companies find that even when they calculate a very conservative cost of labor during downtime, the figures are still astounding.

Lost Revenue

As with labor costs, arriving at an accurate figure for revenue loss due to downtime is challenging. However, an assessment of percentage of revenue loss directly attributable to downtime must still be made. Lost revenue can vary greatly from organization to organization depending on industry, business model and other factors. An on-line business dependent on Internet order entry applications would be seriously impacted by downtime (essentially 100% lost revenue for those days). Similarly, a call center that involves staff entering orders directly to production systems would be seriously affected.

Other Costs

There are often losses incurred due to downtime that are less tangible. For example, a recent weather disaster left several insurance company branches unable to process claims in a timely fashion due to data loss incurred during the outage. The organization not only lost revenue and existing customers, but had to run a public relations and media outreach campaign to try and reestablish its reputation in the area. The sample case study below provides a guide and calculations that organizations can use to estimate the approximate cost of downtime.

Sample case study

A manufacturing company's production AS/400 Server goes down for 2 hours:

Annual gross revenue of the company	50,000,000
Number of employees	200 employees
Average value of an employee working hour (including salary, benefits and administrative cost)	\$ 55.00

Cost of data loss

Employee hours required to restore lost data	50 hours
Value of lost data (hours needed to restore data) x (value of employee working hour)	\$ 2,750

Cost of employee downtime

Hours the system is down	2 hours
Cost of employee downtime (number of employees) x (hrs downtime) x (cost of employee working hour)	\$ 22,000

Cost of lost sales opportunities

Number of sales per year	10,000
Number of customers	1,500
Estimated number of sales lost due to system failure	2 lost sales
Cost of sales lost due to downtime (gross revenue) by (number of sales) = (cost per sale) x (number of lost sales)	\$ 10,000

Cost of lost customer goodwill

Estimated lost customers	2 lost customers
Cost of lost customers	\$ 67,000

(number of sales / year) by (number of customers) =
(number of sales / customer / year) x (cost / sale)
x (number of lost customers) = cost of lost customers

Estimated total cost of downtime

Cost of lost data	\$ 2,750
Cost of employee downtime	\$ 22,000
Cost of sales lost due to downtime	\$ 10,000
Cost of lost customers	\$ 67,000

Total cost	\$ 101,750
-------------------	-------------------

The Bottom Line

As the sample case study provided illustrates, the costs of downtime can be staggering. Given the potential for loss, having a reliable high availability solution in place becomes a business necessity. Tape backups prove a degree of availability, but the incremental nature of these backup procedures may require considerable time and effort to restore data lost since the last backup. Paper trails may be difficult to trace and re-input into the system. In addition, tape backups must be scheduled for off peak hours and use up processing time that could be utilized for other tasks.

What are the Options?

With a good fiscal incentive for backing up systems established, an examination of high availability options including benefits and drawbacks for each approach is useful. There are several high availability alternatives. Depending on the organization's need for data and system access, each is viable. Built-in AS/400 backup and recovery strategies may very well be sufficient depending on the organization's individual requirements.

Inherent Methods

Some applications require more availability than others. The scale of a company's operations, the budget available for high availability projects, and the feasibility of alternatives must be carefully considered before commencing a high availability project. For example, in the case of a web-enabled stock and pricing database for remote sales representatives to check daily or weekly updates, straight media backup, AS/400 journal backup or system-to-system backup may suffice for high availability. The remote sales team may be able to function well enough without access to the database for several hours or even several days without seriously impacting business operations.

This is seldom the case with e-Business applications. In the case of an e-commerce application that provides on-line customers with real-time access to an inventory database, the need for continuous access to inventory data, as well as customer profiles and other data rich content, is critical. Additionally, the IBM AS/400 has several inherent methods that offer various levels of system availability.

Media Backups

Media backups involve copying critical files directly to tape or optical disk for recovery purposes. This means that the informational database is manually restored once the system comes back on-line after a failure. Media backups do have a serious downside. They often require significant time and resources to restore the data, and critical AS/400 objects may be lost. They become impractical when the amount of data required for backup becomes large. The higher the volume of data that is backed up, the longer it will take to restore. Essentially, more downtime is required as a result of lengthy restoration processes.

Journal Backups

A similar alternative for backing up applications involves copying AS/400 system journals to media. Rather than backing up entire database files, companies can choose to make copies of journal files that, when necessary, can be used to restore databases in certain situations. Journals provide a snapshot of a database at a given point in time whether it is weekly, nightly or hourly. When the time comes, organizations can restore to the last full database backup. While an effective solution for certain applications, journal backups do not provide protection for complete system failure or natural disasters resulting in system loss. They are useful for temporary outages, like power loss or recoverable system failure.

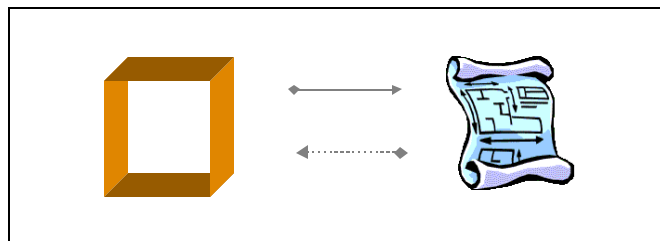


Figure 1: Backup AS/400 journal and restore from journal.

Remote Journals

The remote journal function is part of the base OS/400 beginning with V4R2. Implementing remote journals enables the association of journals and journal receivers between source and target AS/400 systems. By creating an association of a local journal and its journal receivers to a remote journal and its associated receivers, the remote journal function enables journal receiver replication from a source system to a target system. When used in coordination with communications interfaces, such as OptiConnect for OS/400, SNA or TCP/IP, customers can create clustered environments that provide 24

x 365 availability and full system redundancy. Remote journals help off-load CPU consumption from the local system so it can achieve more throughput. Remote journals replace customer programming with a more efficient system programming interface to capture and transmit journal images between source and target systems.

RAID 5

RAID 5 (redundant arrays of independent disks) offers organizations the ability to maintain shadow copies of information across separate disks. It allows missing data to be recalculated from one disk to another in the event of a disk failure. While offering increased availability, RAID technology does not protect against multiple disk failures. Another availability solution involves DASD, or disk, mirroring.

Disk Mirroring

Disk mirroring involves duplicating data writes to a twin set of disks. In other words, if an organization's iSeries has five disks, then it maintains a twin set of another five disks for backup and recovery. This protects against other failures such as a power outage within one disk set, multiple disk failures, and physical damage to disks. While a more expensive alternative to RAID 5, DASD mirroring does not require another iSeries system to be purchased and remains an affordable solution for many organizations.

File Transfer Technologies

A final option for applications not requiring continuous availability involves copying files between primary systems and one or more secondary machines. This means copying entire database files to a backup machine via a communication line and, when required, copying them back to the primary server.

While useful for smaller data stores, file transfer to secondary machines becomes time consuming as the database becomes larger. Selecting only critical data is a variation on this method, but requires careful administration to ensure that absolutely all the data required to drive the business is backed up. Additionally, if transmission fails due to communication problems, a complete retransmission of all files is necessary.

Advanced Availability

The trouble with media storage, journal copies, and server-to-secondary system copies is that they all mandate some degree of system downtime. It takes time to copy critical files—time that could be better spent serving users. Additionally, the server also has to be taken off-line when it comes time to restore the database. As pointed out, this process is relatively straightforward but can take a significant amount of time to carry out depending on data volumes.

When inherent iSeries backup and recovery processes fail to meet the requirements of organizations, they must develop a more advanced strategy for ensuring access to corporate applications and the data that powers their business.

While it may be perfectly acceptable to schedule periods of downtime for backups to an intra-organization solution, this is definitely not the case for many applications. In commercial applications such as e-Business or direct customer access applications like call centers, access to real-time information 24 hours a day is mandatory. As reliance on computer systems to maintain operations as well as competitive advantages grows, continuous availability becomes a necessity. The ability to replicate data, the applications that support them, and other iSeries object types in real-time is critical.

What is iSeries and AS/400 High Availability?

There are four major components involved in iSeries and AS/400 high availability systems: hardware configuration, mirroring software, role switching facilities, and a process monitoring interface. There are a number of different hardware configurations that support varying degrees of high availability.

Fully Redundant Systems

Companies with a great need for secondary or recovery systems may elect to install another system to serve as an identical twin to the primary production servers. This involves setting up another AS/400 with identical applications and data as well as CPU and DASD. In the event of outage, users have an exact replica of their production server at their disposal. In effect, any outage will be transparent to users.

Balanced Servers

Another option is to off-load batch jobs, such as query and reporting activity, as well as critical data and other objects to a secondary system. The secondary system does not need to have identical functionality as the production server. Organizations can make full-time use of their secondary machine, putting it to work as a batch job server as well as a recovery machine in the event of outage, while saving some cost on hardware investment. However, if there is an outage, performance will be sacrificed until such time as normal primary server operations can be restored.

Mixed-Mode Servers

The specialized iSeries mixed-mode server is designed for running applications in environments that demand both interactive as well as batch processes. By purchasing a mixed-mode iSeries, organizations can satisfy the requirements of iSeries system redundancy as well as implement new applications such as a data warehouse on the same system. This is because the mixed-mode server is optimized to handle both interactive operations such as user entered transaction updates, and batch processing such as querying and reporting.

Switching

Switching mechanisms and high availability software are also critical components of any high availability solution. Switches handle operational role switching between primary and recovery machines. Automatic sensing and switching devices are available from several vendors but full hardware switching capabilities may not necessarily be a critical requirement for all companies. In today's sophisticated network environments, a virtual switchover to a new machine may be as simple as logging on to a second machine. Users and IS staff can be alerted to an outage at which time network traffic can be re-routed by IS or users can click a different icon to establish a connection to the recovery machine.

Switching mechanisms are a convenient way to automate user and role switching. Some switching software includes decision control facilities for establishing events and reactions. Switching software can either notify administrators of an outage or initiate other events after a user-defined delay. These types of switching facilities are valuable to organizations wanting to make system outages as seamless to users as possible.

Mirroring Software

The most important element in a solid iSeries availability solution is the mirroring software that will synchronize the recovery system in real-time from the production machine. This software component must be fast, transaction-consistent, and guarantee data delivery at the secondary site even in the event

of a network or system outage. The software must also be two-way to allow both the production and the secondary machines to act as a source and target to each other. Full two-way functionality gives the secondary system real-time capabilities to re-synchronize the production machine, while enabling users to continue working on the secondary system.

For most companies, production applications are static and change infrequently. Manually synchronizing these to the secondary machine is fairly straightforward but doing it automatically may be desired. Organizations wanting an easy-to-use solution that requires minimal administration should look for a high availability package that includes both data and object mirroring capabilities. User profiles, data queues, spool files and other critical iSeries objects must be replicated to recovery systems for full redundancy.

Ultimately, organizations look for a situation that involves switching users to the secondary site while recovery takes place on the production system. During operation at the backup site, database changes are being captured on that system. Once the production system is recovered and restored to the last backup, the mirroring engine can be turned on to enable all the database changes captured on the secondary machine to be replicated back to the production system. In time, the production machine will be exactly synchronized with the secondary system and users can be switched back to the production system.

Process Monitoring Facilities

The final component of the high availability solution is an interface for tracking the entire mirroring process. The ability to monitor historical and current activities is important as administrators need to know the performance of their solution. What is the latency in moving data and objects between primary and secondary systems? What CPU cycles are being used to support the availability solution? Many solutions do not incorporate an intuitive and reliable monitoring tool.

What to look for in an iSeries and AS/400 high availability solution

Implementing a high availability solution for iSeries and AS/400 computing environments involves moving data and objects in real-time between primary systems and recovery machines in order to secure operations from business interruption. The software used to enable such a solution should meet certain standards. It should be capable of refreshing the primary systems while active. It should also be capable of moving data and objects in a two-way configuration. That is, the primary server should be able to act as both the source or the target of data and object mirroring while it is active. If the software doesn't include this functionality, near continuous availability is impossible.

Fault Tolerance

Ironically, some products require downtime in order to re-synchronize the primary system before it is brought back on-line. Because many data and object mirroring solutions are not truly two-way enabled, administrators must shut down access in order to refresh the primary server after an outage. More sophisticated products allow organizations to refresh primary iSeries systems while allowing users to continue working on the recovery server. This ensures a smooth transition back to normal operations.

Transaction Consistency

The software should be transaction-consistent. This ensures that transactions are performed on the secondary machine in precisely the same order in which they occurred on the primary AS/400 server.

Technologies use the iSeries and AS/400 audit journal to track the order of transactions. Transaction consistency ensures data and object integrity.

Fault Tolerance

The software should include fault tolerance or guaranteed data delivery at the secondary server. If there is a network, system or communication outage during replication, the secondary server should reflect the last complete transaction. This eliminates the possibility of duplicate entries or partial transactions when it comes time to switch operations back to the primary server.

Ease of Use and Administration

The application should be easy to install and administer. The last thing companies want to worry about when setting up a high availability solution is custom programming and other time-consuming, costly and resource-draining activities. Object and data replication software should install quickly and be ready for operation with minimal effort. Additionally, users should be able to administer both the object and data mirroring elements from a unified interface.

Integrated Product Suite

The component integration of iSeries and AS/400 high availability solutions is another key consideration. For companies holding low maintenance as a prime concern, product suites that offer mirroring software, monitoring, and switching as a single solution are desirable. Integrated suites that incorporate single-install procedures and feature complete and central monitoring facilities for the entire availability process are ideal for organizations looking for total peace of mind.

Incremental Change Mirroring

The software should send data on a net change basis. For most companies, program files do not change frequently. These objects need only be mirrored to the secondary server on an as-needed basis, such as when a new version is released. However, data and several other objects such as user and security profiles, and data areas are subject to rapid changes.

Mirroring software should be capable of sending only changed data to the secondary system. This avoids the transmission of redundant data and the depletion of system resources caused by updating the target machine with entire transaction records instead of only those elements that actually changed since the last update. This technology is not only applicable for high availability solutions, but for any production system environment that demands access to critical data and the applications.

In summary, companies requiring an iSeries and AS/400 high availability solution should ensure that they choose a package that integrates as many of the critical components as possible. Piecemeal solutions -- separate components for switching, mirroring, and monitoring -- can involve separate installation, configuration and administration of each element. High Availability Suite from DataMirror Corporation offers organizations a fully integrated solution for iSeries high availability.

About DataMirror High Availability Suite

DataMirror High Availability Suite is a single integrated product for continuous business operations in iSeries and AS/400 environments. It enables real-time mirroring using sophisticated technology that guarantees that events are properly synchronized between primary and secondary systems. High Availability Suite has a single installation, as well as integrated processes for configuration, mirroring

and monitoring both data and objects. High Availability Suite also provides operational switching with the flexibility to initiate tasks based on conditional events. Both protection and simplicity are maximized to deliver the most robust and secure technology for iSeries and AS/400 high availability. The advanced match/merge architecture of High Availability Suite enables real-time auto-registration, making the product easy to administer while maximizing data integrity.

Key Features of High Availability Suite

Please note that all features and functionality of DataMirror High Availability Suite are also included in DataMirror iCluster, the iSeries and AS/400 cluster management solution, discussed later in this document.

High performance data and object mirroring

High Availability Suite captures database transactions from primary systems and mirrors them in real-time to one or more secondary iSeries servers. Mirroring can be performed on either a relative record number basis or by unique key. Real-time iSeries object mirroring from primary AS/400 to one or multiple backup systems. High Availability Suite supports continuous mirroring or on-demand full refresh of specific object groups.

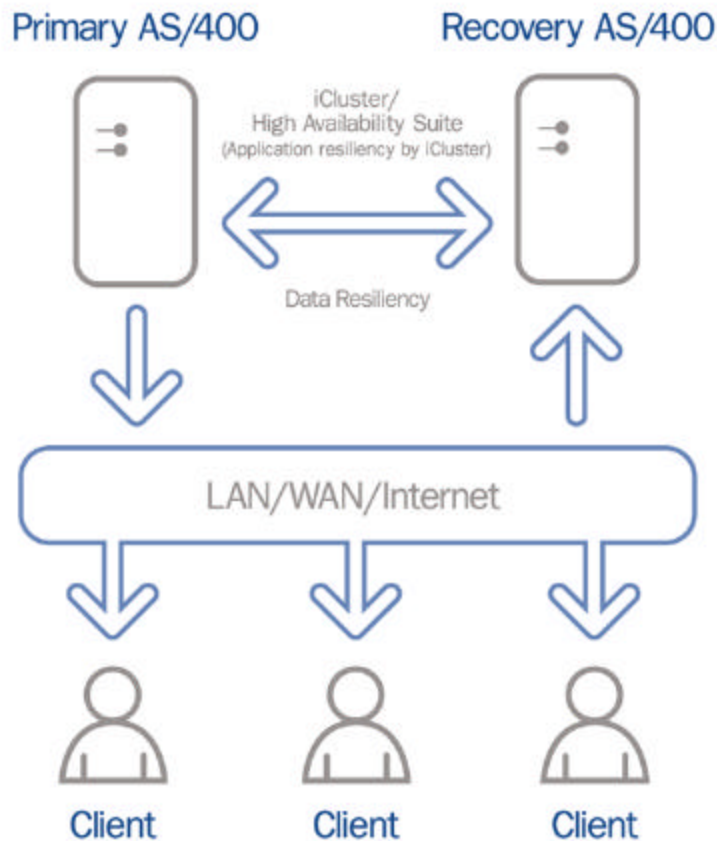


Figure 2: High Availability Suite mirrors data and objects between primary and recovery iSeries and AS/400 servers. Cluster management and mirroring is provided by iCluster software.

Ease of Use

High Availability Suite installs without the need for programming changes to existing applications. Intuitive menu-driven screens simplify the setup process, allowing users to map selected data and objects to recovery systems without scripting or coding. High Availability Suite is a single-setup product with no need to run separate installs for each component. There is no need to run separate tasks from different applications. Data and object mirroring, switching and monitoring processes are all centrally administered. High Availability Suite also features additional enhancements to simplify initial synchronization between primary and secondary systems.

Match/Merge Architecture

The iSeries uses both audit journals and database journals to record transactions. Both are critical to high availability. Many mirroring solutions simply dump transactions to the secondary machine regardless of when they arrive at the secondary machine. This means that if the audit journal posts a transaction on September 23 at 2:43 pm and the database journal records an event at 2:44 pm, it is possible for the latter to be transmitted and acknowledged by the secondary system before the prior. High Availability Suite utilizes synchronous event processing to ensure that all operational events on the primary system are synchronized with the events on the secondary system regardless of when they arrive at the secondary machine.

High Availability Suite maximizes data integrity by performing a match and merge of all records transmitted to the target secondary machine to ensure that events are recorded in the exact sequence they occurred on the primary system. Records are matched and merged in real-time to enable real-time auto-registration and eliminate the need to continually perform sync checks to ensure synchronization integrity.

High Performance Data Draining

Another unique feature of High Availability Suite involves the draining of data to the secondary system. As data and other objects are captured from primary systems, many products send them one at a time for integrity reasons. High Availability Suite sends the maximum amount of data at any given time, limited only by communication channels. With this draining feature, larger transactions occupy only one communication channel, freeing up other channels for the mirroring of smaller transactions. While this means that the single, larger transaction may take somewhat longer to reach the target system, the many smaller transactions, which may have been lost in the event of outage, are mirrored on the secondary machine. Draining maximizes the amount of data successfully mirrored to secondary systems to minimize data loss resulting from system failure.

Sync Check

Sync check allows you to determine whether the objects on the target are the same as the objects on the source. All objects for the specified target/group combination are sync checked. You can perform two types of sync checks. The Object List sync check compares the number, object name and types of objects on the source system with the number, object name and types of objects on the target system. It does not compare the attributes of the objects. The Object Attribute sync check looks for the existence of replicated *FILE objects on the target system. It also compares the attributes of the source and target *FILE objects. How often you perform a sync check depends on how often you want to check that the source and target objects are the same. The sync check obtains the file attributes for the selected source objects and compares them with the file attributes of the target objects. You can indicate the attributes that you want to include in the sync check.

Commitment Control

High Availability Suite commitment control stages database transactions so that they are assembled before being applied, or additionally, opened in a commitment control environment to ensure that all updates are received. It also makes sure that the changes are applied in the correct sequence. Commitment control staging is performed on the target system.

High Availability Suite's commitment control staging has the added benefit of holding journal entries while the target apply process has been suspended so that you can perform backups and other maintenance operations that require exclusive access to target files and objects. Mirroring is allowed to continue as well as source scrape processes and the transfer of journal information to the target system. Consequently, it helps minimize potential data loss in the event of a primary system failure and allows for a high degree of synchronization.

High Speed Operational Switching

High Availability Suite provides all the necessary tools to implement operational switching for high availability. When an outage occurs, High Availability Suite can be used to switch physical operations to a backup system so that normal activities can resume as soon as possible. This provides high availability in environments that demand continuous operations. High Availability Suite automatically detects when communications to the primary system have failed and invokes the corrective measures that are specified by the user. You can specify how frequently the communications should be polled to verify that the primary system is working properly. A switch condition can be defined by specifying the maximum waiting time to receive a response from a polled communication link and how many consecutive communication failures are tolerated. It also provides the necessary support for user exits and messaging.

Metadata is refreshed to the backup system to support an operational switching without requiring any user intervention. Metadata refers to internal files that define the replication behaviour specified by the user and to identify which objects and physical files are replicated, and the targets that will receive these objects and files. Metadata is maintained on both the production and backup systems. This allows the backup system to replicate objects and physical files after an operational switch has been performed

Status Monitor

High Availability Suite allows you to monitor replication on both the source and target systems. This is especially useful when the source and target systems are located in different geographical regions. Regardless of where you happen to be situated (in either the source or target environment), replication can be monitored from either site. Interactive, real-time inquiry screens snapshot current activities while historical inquiry screens track past replication statistics. Monitoring is an important component of a high availability solution when combined with operational switching.

Multiple Views

High Availability Suite provides a number of different ways to view the relationships between object specifiers, groups, and target systems. For example, in certain situations, you may want to view all the object specifiers that are currently selected in a single group. In other situations, you may want to identify the groups that have a particular object specifier currently selected. This flexibility allows you to work with High Availability Suite in a way that meets your requirements. Depending on your preference, you can work with High Availability Suite by navigating through menus and screens or by invoking commands from a command line.

Exclusive XtremeCache™ Technology

Systems that accommodate high volume transactions, or ones that undergo end-of-day heavy batch processes typically cause congestion in the mirroring process called latency. These systems require premium real-time mirroring performance. High Availability Suite addresses these issues with XtremeCache technology, which optimizes the flow of data into and out of the primary and secondary systems. XtremeCache can be configured in multiple high availability instances to achieve extreme performance scalability. This breakthrough performance enhancement is designed to help customers attain near zero latency for iSeries high availability environments, resulting in the highest level of availability of any real-time data/object mirroring system.

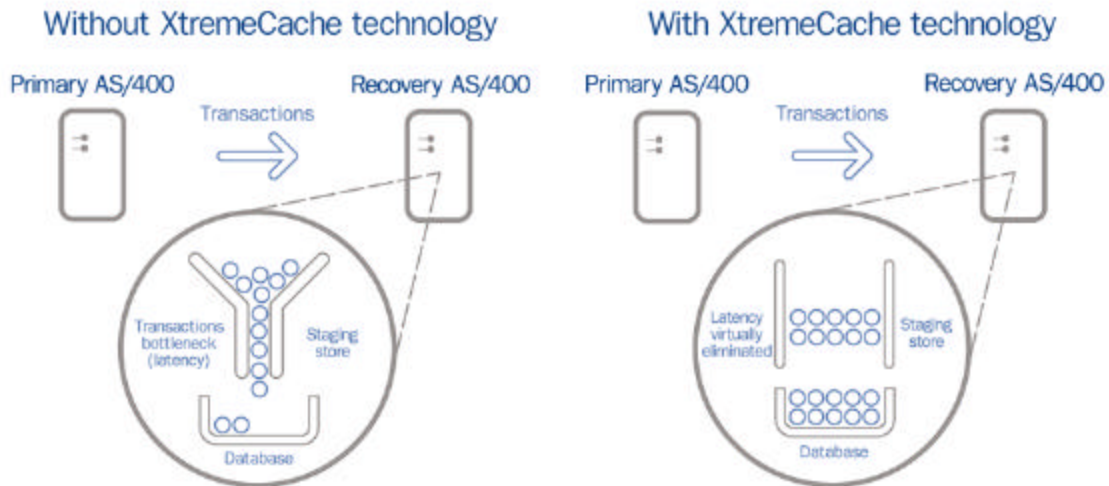


Figure 3: XtremeCache technology exclusive to DataMirror High Availability Suite and iCluster.

Real-time IFS Mirroring

High Availability Suite utilizes state-of-the-art byte stream file recording introduced into the OS/400 operating system by IBM to manage the underlying Integrated File System (IFS) requirements of e-Business applications and Enterprise Resource Planning (ERP) software. IFS support is especially beneficial for e-Business implementations because many types of web page objects are typically stored under the Integrated File System. Users can now mirror their large IFS files in real-time to ensure that AS/400 web servers and e-Business sites are thoroughly backed up and resilient.

Checkpoint User Exits

Automated synchronization exit processing simplifies the process of completing days-end batch processing and updating tape backups on recovery iSeries and AS/400 systems while allowing the production systems to remain live. Automated checkpoint processing allows users to launch processes at selected quiesce or synchronization points to maximize data integrity.

Full Range of Services

DataMirror in-house professionals are available for any implementation assistance that organizations may require. Seasoned high availability and cluster experts can help outline requirements, install software, test the implementation, and help with any other project needs. DataMirror also offers global 24/7 support via telephone, e-mail and the web.

What added value does High Availability Suite deliver?

With DataMirror High Availability Suite, companies can not only gain peace of mind knowing that they have a reliable, high performance solution that ensures access to their AS/400 data and applications during both planned and unplanned outage, but they can also simultaneously gain improved overall system performance.

Through a workload balancing strategy, organizations can off-load user query activity to the secondary system to improve overall efficiency of the AS/400 environment. Because High Availability Suite mirrors data and other objects in real-time, companies can actually create a shadow database on the secondary machine and use it for batch jobs, queries, reports, and other resource intensive tasks. Data integrity and consistency is ensured and organizations can extend their technology investment by maximizing system resources.

DataMirror's complete high availability solution also provides enhanced support for recent IBM high availability initiatives including logical partitioning and iSeries and AS/400 clustering.

Logical Partitioning (LPAR)

Logical partitioning (LPAR) is an advance in V4R4 that makes the iSeries much more highly available. LPAR enables customers to run multiple independent servers or partitions—each with their own processors, memory and disks—within a single iSeries server. Logical Partitioning is ideal for server consolidation, running mixed production and test environments, and supporting multiple languages and time zones in a single server.

Each partition may be configured independently to accommodate different system names, user profiles, languages and time zones. Partitions may also be customized to address individual business requirements for example, a single AS/400 may be optimized for both interactive and batch workloads to support both transaction processing (OLTP) and e-Business or business intelligence applications.

Enhanced levels of high availability

Logical partitioning has several advantages for customers implementing a high availability solution. One problem that many companies face in the Internet era is that production systems are required to be online 7/24/365. There is simply no good time to lock files on the production system in order to perform backups to tape. Logical partitioning enables these customers to run their mission-critical systems continuously in one partition, while high availability software mirrors all critical data and objects in real-time to an adjoining partition. Backups to tape can then be easily made on the mirrored system as required.

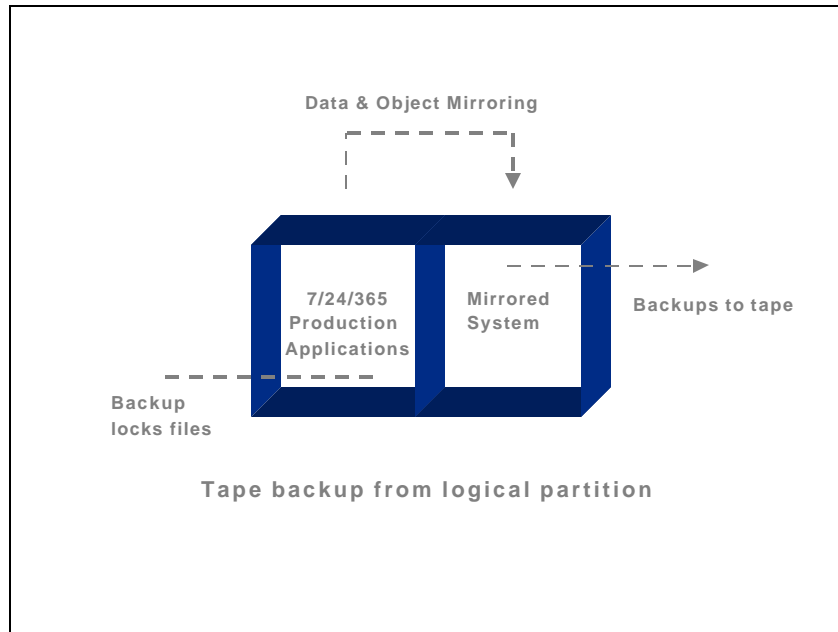


Figure 4: Backup systems with LPAR without locking files

LPAR also solves a user profile issue that some customers have experienced in attempting to replicate data and objects from multiple servers into a single non-partitioned AS/400 allocated for backup and recovery. Suppose a company is mirroring data from two AS/400s to one failover server. Server One has the user profile Bob and the designation Programmer. However, Server Two also has the user profile Bob, but this Bob is designated a User. The discrepancy in user profiles creates a conflict on the backup system. Similar problems can occur at the library and individual object level as well.

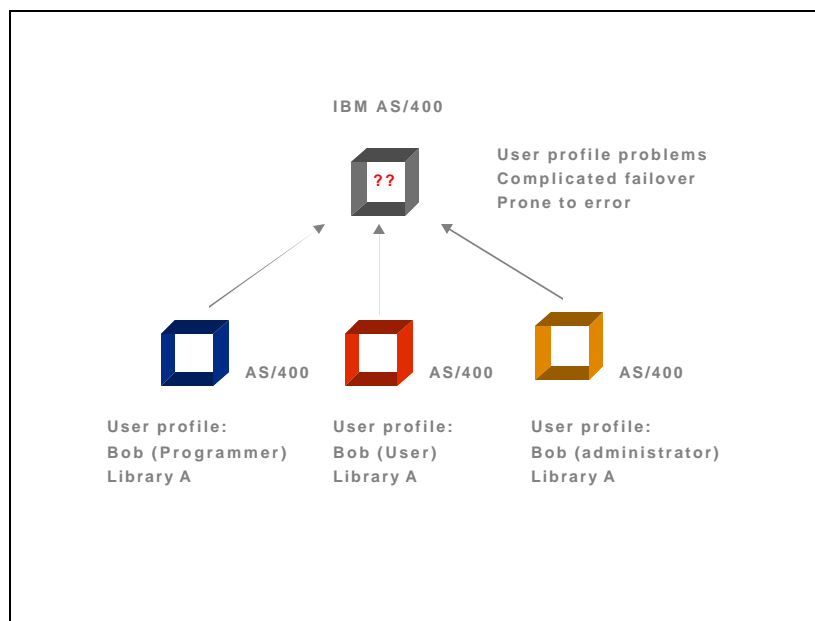


Figure 5: User profiles prone to error during failover

User profile problems mean that the high availability solution could be prone to error, and can result in a complicated and lengthy failover procedure to restore production systems after a planned or unplanned outage. Consulting is often required to resolve these conflicts and automate the failover procedure. With logical partitioning this problem is all but eliminated. Data and objects from Server One are replicated to partition one (SYS1), while data and objects from Server Two are mirrored to partition two (SYS2) and so on. Each partition is assigned a separate system name and configured with its own user profiles and libraries.

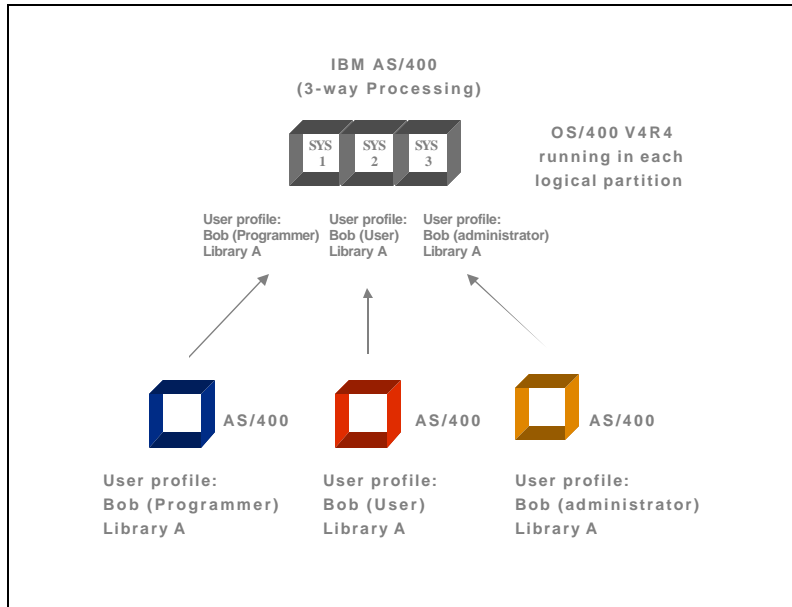


Figure 6: Easily manage user profiles with LPAR.

LPAR also enables customers to run production and test environments in a single partitioned iSeries. Companies can install a new operating system in a partition and thoroughly test the system and its compatibility with applications before gradually migrating applications to the partition and initiating switchover from the production environment to the test partition.

Given that logical partitioning can allow one 12-way multi-processing AS/400 to be partitioned into multiple separate systems, each running their own applications and data stores, the advantages of this technology are clear. Fewer servers means more efficient management of human, financial and information resources and greater availability of critical systems. With support for logical partitioning, as well as integrated cluster management in V4R4, the AS/400 is well positioned as a platform on which to consolidate a company's vast information resources.

iSeries and AS/400 Clustering

While server clusters are common in NT and UNIX environments, clustering is new technology for the iSeries and AS/400. Clustering makes it possible to share computing load over multiple systems without users necessarily needing to know that more than one system is involved. If a single node in a clustered environment is taken off-line or goes down, users may notice a decrease in performance, but they will not lose access to either applications or critical business data.

The minimum requirements for an iSeries and AS/400 cluster are two servers connected by a network and cluster management software available from a High Availability Business Partner like DataMirror. Cluster management software provides clustering services such as failure recovery, operational switching between primary and secondary nodes in a cluster, data resiliency, and the ability to manage all nodes or systems in the cluster from a single workstation.

Clustering provides a way to increase the overall availability of business critical applications and the data that drives these applications. Clustering, together with third-party provided ClusterProven[™] applications and Cluster Management Utilities represents a highly efficient and reliable method to establish, monitor, and manage data and application resiliency across an organization.

Cluster Architecture

Cluster Resource Services (CRS) is the first step in IBM's multistage clustering plan. CRS consists of an open set of APIs incorporated into the base OS/400 operating system in V4R4 that enable cluster facilities. These APIs are available to iSeries and AS/400 application providers, Independent Solution Vendors (ISVs) and customers to enhance their application resiliency.

A Cluster Management Utility (CMU), such as that provided by DataMirror and other HABPs, allows users to create, configure and administer the cluster. Communication interface addresses are defined to form the cluster's node-to-node links. Resilient resources--objects replicated to one or more nodes--are associated with a Cluster Resource Group (CRG) so they can be managed as a single unit. Since a group usually contains all of the elements needed to run a specific application and for client systems to connect to the service, full data resiliency is enabled over multiple backups. Groups allow an administrator to combine resources into larger logical units and more easily manage them as a single unit.

Two types of CRGs are supported: one for data resiliency and one for application resiliency. Data resiliency provides the capability to switch the point of access for a set of data to another node that maintains an exact replica of that data. Application resiliency provides the capability to replicate all critical data required by a ClusterProven[™] application. A ClusterProven[™] application is a highly available application that can switch to a backup node when the primary node becomes unavailable. These applications are aware that they are operating in a clustered environment and have intelligent exit programs that are executed in the event of primary node failure. ClusterProven[™] applications coordinate application restarts with the data elements of CRGs when the time comes to switch operations back to the primary node.

Lastly, application resiliency enables the specification of a takeover IP address that will allow for near seamless switchover of a cluster aware application in the event of failure. As part of the information conveyed to a ClusterProven[™] application, the takeover IP address is fed to the CMU for near transparent switchover of users to the secondary, tertiary, or other iSeries node.

As organizations grow more dependent on data for decision support, order processing, inventory control, accounting and other business-critical operations, they may require a more advanced solution like clustering to ensure uninterrupted access to business critical data and the systems that house it.

About DataMirror iCluster

In today's 24/7 world, there is simply no time for either unplanned or planned system outages. The business challenges of globalization, consolidation, and e-Business have created a growing industry demand for continuous availability. Cluster technology is implemented on the IBM iSeries and AS/400 to provide customers with continuous availability during both planned and unplanned outages. iCluster is for customers who require even greater levels of availability than can normally be attained through a non-clustered high availability environment. The goal of iSeries clustering is to achieve unprecedented system availability approaching 99.999% and beyond.

A cluster consists of networks of iSeries and AS/400 systems (or nodes) that work together to provide seamless operations. If the node in the cluster that provides a service can no longer perform this role, operations are automatically switched to a designated recovery node. To achieve this objective, objects and data must be replicated from the primary system to the backup node so that operations can be easily moved to this node.

While the iSeries provides the basic cluster infrastructure, a total AS/400 clustering solution requires the involvement of cluster middleware partners like DataMirror, as well as the involvement of iSeries Solution Developers (SDs). DataMirror iCluster software provides data resiliency, while the Solution Developer provides application resiliency. Together, these parts provide the whole solution to continuous availability. DataMirror iCluster software further enhances the iSeries and AS/400 cluster solution by providing the easy to use cluster management and robust data resiliency required in today's 24x365 e-Business world.

iCluster includes an intuitive Java graphical user interface (GUI), as well as optional menu-driven green screens, to define nodes in a cluster, determine what data and other critical objects are mirrored to specific nodes, and monitor the entire cluster from a single station. It also handles the data and object mirroring requirements of clustered environments. Users can define cluster groups, initiate cluster operations, and set cluster security levels.

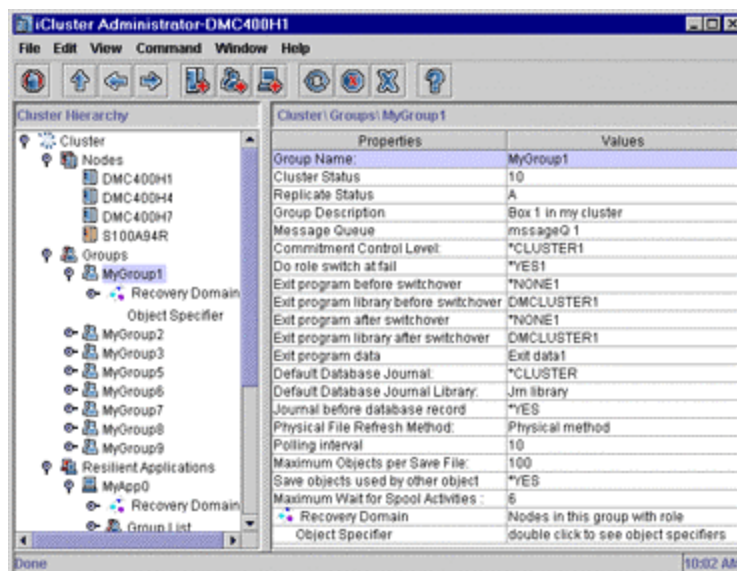


Figure7: Intuitive iCluster GUI

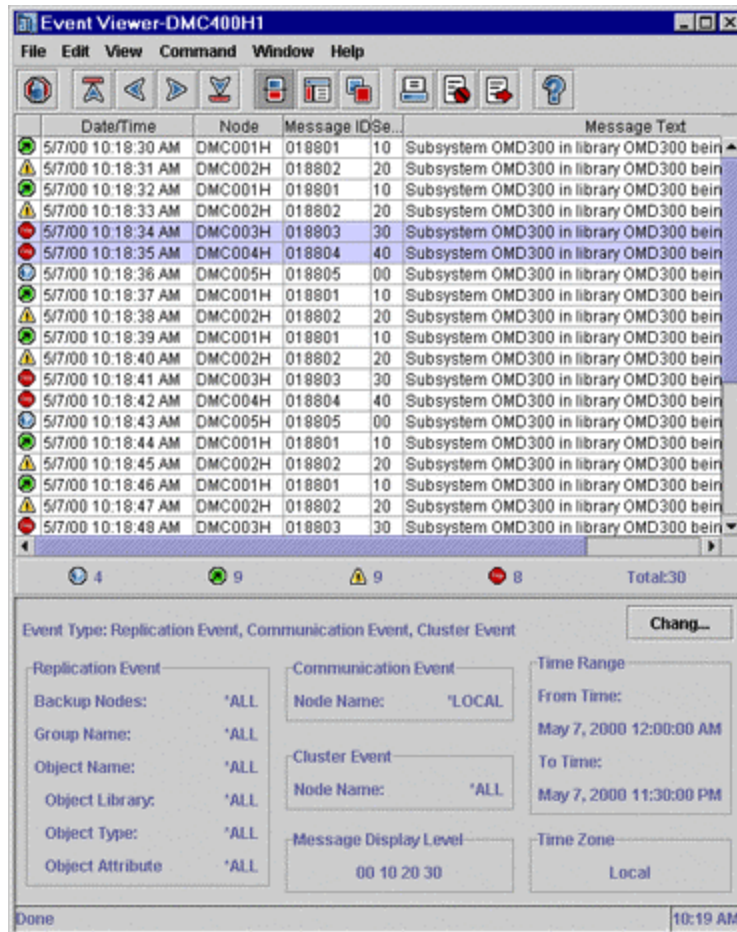


Figure 8: Another view of iCluster Java GUI.

DataMirror iCluster is an easy to implement and high performance solution for ensuring the continuous availability of business-critical applications such as e-Business, ERP and customer-facing applications. iCluster can be installed on distributed systems to maintain round-the-clock system availability. iCluster also features exclusive XtremeCache technology that optimizes the flow of data into and out of a high speed software cache. This breakthrough performance enhancement leads the industry in its ability to help customers attain near zero latency for clustered high availability environments.

IBM ClusterProven™ Program for Applications Resiliency

Solution Developers can build enhanced continuous availability applications using a common interface architecture to cluster management and data resiliency products like iCluster. Properly designed, these applications can be configured and activated from cluster management utilities like iCluster and made resilient to planned and unplanned outages. Applications meeting the criteria may receive the right to be recognized as ClusterProven™ for iSeries and AS/400. When an application is IBM ClusterProven, it is aware of the status of other cluster nodes. In the event of failure, a ClusterProven application is able to sense failure and initiate automatic fail-over of the primary node, to the secondary node.

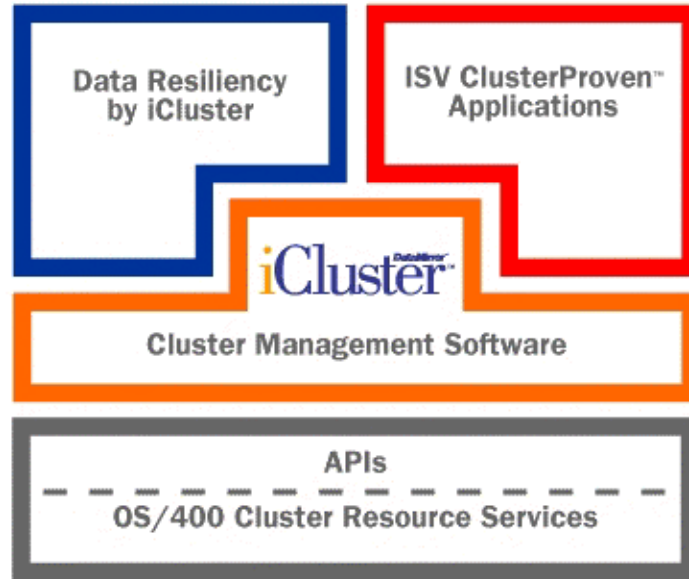


Figure 9: Role of iCluster in facilitating AS/400 resiliency

iCluster supports IBM's ClusterProven™ standard for applications resiliency by providing auto-configuration and activation of ClusterProven for iSeries applications. iCluster can further simplify the implementation of data and application resiliency projects and handle role switching for new ClusterProven™ applications.

Summary

In closing, it is important to remember that as an organization grows more dependant on data for e-commerce, decision support, order processing, inventory control, accounting and other operations, the need for high availability becomes imperative. Moreover, while traditional tape backups or FTP file transfer may be suitable for current business activities, future business growth may necessitate a more advanced high availability solution. For example, a traditional bricks-and-mortar company that is entering the e-commerce space will need to ensure 24/7 operations for its web servers, e-commerce applications, and dynamic databases that make the e-commerce experience. When that time comes, it is imperative that companies have technology in place to ensure uninterrupted access to organizational data and the systems that house it. While there are several software packages for IBM eServer iSeries and AS/400 high systems availability, none offer the integrated design, high performance architecture, and ease of use of DataMirror High Availability Suite and iCluster.

Copyright © 2003 DataMirror Corporation. All rights reserved. DataMirror, iCluster, High Availability Suite and XtremeCache are trademarks or registered trademarks of DataMirror Corporation. All other brand or product names are trademarks or registered trademarks of their respective companies.

ABOUT DATAMIRROR

DataMirror (Nasdaq: DMCX; TSX: DMC) delivers LiveBusiness™ software solutions that give companies the power to manage, monitor and protect their corporate data in real-time. DataMirror's comprehensive family of solutions enables customers to easily and cost-effectively capture, transform and flow data throughout the enterprise. DataMirror unlocks *the experience of now™* by providing the instant data access, integration and availability companies require today across all computers in their business. Over 1,700 companies have gone live with DataMirror software. DataMirror is headquartered in Markham, Canada, and has offices around the globe.

DataMirror®
THE EXPERIENCE OF NOW.™

HOW TO DO BUSINESS WITH DATAMIRROR

North America	1 800 362 5955
UK	+ 44 (0)20 7633 5200
France	+ 33 (0) 1 55 70 30 18
Germany	+ 49 6151 8275 0
Hong Kong	+ 852 2251 8226

FOR MORE INFORMATION VISIT
WWW.DATAMIRROR.COM