

Data Recovery

- **Computer systems may crash. Files may be accidentally deleted. Disks may accidentally be reformatted. Computer viruses may corrupt files. Files may be accidentally overwritten.**
- **All of these can lead to the loss of your critical data**
- **The advanced tools should allow us to find the files and restore them for use.**
- **In those instances where the files have been irreparably damaged, your computer forensics expertise should allow you to recover even the smallest remaining fragments.**

- **Data recovery is, of course, of potential interest to anyone who has lost data to the ravages of time, malice, or carelessness.**
- **But in forensic computing or analysis, it takes on a new meaning—what other people have thrown away can become an important component in understanding what has happened in the past, as burglary tools, data files, correspondence, and other clues can be left behind by interlopers**

Data Recovery Defined

- **Data recovery is the process in which highly trained engineers evaluate and extract data from damaged media and return it in an intact format**

- **As computers are used in more important transactions and storage functions, and more important data is stored on them, the importance of qualified data recovery experts becomes evident.**
- Perhaps the information has been subjected to a virus attack, suffered damage from smoke or fire, or the drive has been immersed in water, the mainframe software has malfunctioned or file allocation tables are damaged—data recovery experts can help.

- **What would happen to the productivity of your organization in the event of a system-wide data center failure?**
- For most companies, the loss would be catastrophic.
 1. Hundreds, perhaps thousands, of employees would be rendered unproductive
 2. Sales transactions would be impossible to complete and customer service would suffer.
 3. The cost of replacing this data would be extraordinary—if it could be replaced at all

DATA BACKUP AND RECOVERY

- Companies that can provide reliable and rapid access to their information are now the fastest growing organizations in the world.
- **To remain competitive and succeed, they must protect their most valuable asset—data.**
- Fortunately, there are specialized hardware and software companies that manufacture products for the centralized backup and recovery of business critical data
- Hardware manufacturers offer automated tape libraries that can manage millions of megabytes of backed up information and eliminate the need for operators charged with mounting tape cartridges.
- **Software companies have created solutions that can back-up and recover dozens of disparate systems from a single console.**

- But Industry experts estimate that over 56% of the data in client/server networks is still not backed up on a regular basis?
- **It is often due to organizations' ever-shrinking back-up windows, inadequate network infrastructure, and a lack of system administration.**
- **Compounding the problem is an overall lack of experience in defining the proper features necessary for a successful; backup application.**
- **Finally, there is often a shortage of in-house expertise needed to implement sophisticated, enterprise-level backup applications.**

- The following are obstacles to backing up applications:
 - Backup window
 - Network bandwidth
 - System throughput
 - Lack of resources
 - Backup Window

Backup window

- **The backup window is the period of time when backups can be run.**
- **The backup window is generally timed to occur during nonproduction periods when network bandwidth and CPU utilization are low.**
- **However, many organizations now conduct operations 7 days a week, 24 hours a day—effectively eliminating traditional backup windows altogether**

Network Bandwidth

- **Many companies now have more data to protect than can be transported across existing local area networks (LANs) and wide area networks (WANs).**
- **If a network cannot handle the impact of transporting hundreds of gigabytes of data over a short period of time, the organization's centralized backup strategy is not viable.**

System Throughput

- **Three I/O bottlenecks are commonly found in traditional backup schemes. These are**
 - 1. The ability of the system being backed up to push data to the backup server**
 - 2. The ability of the backup server to accept data from multiple systems simultaneously**
 - 3. The available throughput of the tape device(s) onto which the data is moved**

Lack of Resources

- **Many companies fail to make appropriate investments in data protection until it is too late.**
- **Often, information technology (IT) managers choose not to allocate funding for centralized data protection because of competing demands resulting from emerging issues such as e-commerce, Internet and intranet applications, and other new technologies**

- **These are just a few of the impediments that make implementation of an enterprise backup and recovery solution a low priority for some organizations.**
- **Fortunately, there have been major advances in hardware and software technologies that overcome many or all of the traditional obstacles faced by IT professionals as they attempt to develop a comprehensive data-protection plan.**
- **In addition, companies such as StorNet provide specialized expertise in the deployment of complex, integrated storage solutions**

The Future of Data Backup

- Successful data backup and recovery is composed of four key elements:
- **The backup server, the network, the backup window, and the backup storage device (or devices).**
- **These components are highly dependent on one another, and the overall system can only operate as well as its weakest link.**
- To help define how data backup is changing to accommodate the issues described earlier, let's take a look at each element of a backup and recovery design and review the improvements being made

The Backup Server

- The backup server is responsible for managing the policies, schedules, media catalogs, and indexes associated with the systems it is configured to back up.
- The systems being backed up are called *clients*.
- Traditionally, all managed data that was being backed up had to be processed through the backup server.
- Conversely, all data that needed to be restored had to be accessed through the backup server as well.
- This meant that the overall performance of a backup or recovery was directly related to the ability of the backup server to handle the I/O load created by the backup process

- In the past, the only way to overcome a backup server bottleneck was to invest in larger, more powerful backup servers or data backup and recovery and divide the backup network into smaller, independent groups.
- Fortunately, backup-software developers have created methods to work around these bottlenecks.
- The most common workaround is to create *tape* servers that allow administrators to divide the backup tasks across multiple systems while maintaining scheduling and administrative processes on a primary or *backup* server

- This approach involves attaching multiple tape servers to a shared tape library, which reduces the overall cost of the system.
- Figure 5.1 is an example of a backup configuration such as this
- The newest backup architecture implements a server-less backup solution that allows data to be moved directly from disk to tape, bypassing the backup server altogether.
- This method of data backup removes the bottleneck of the backup server completely.
- However, the performance of server-less backup is then affected by another potential bottleneck—bandwidth.
- Figure 5.2 is an example of a server-less backup

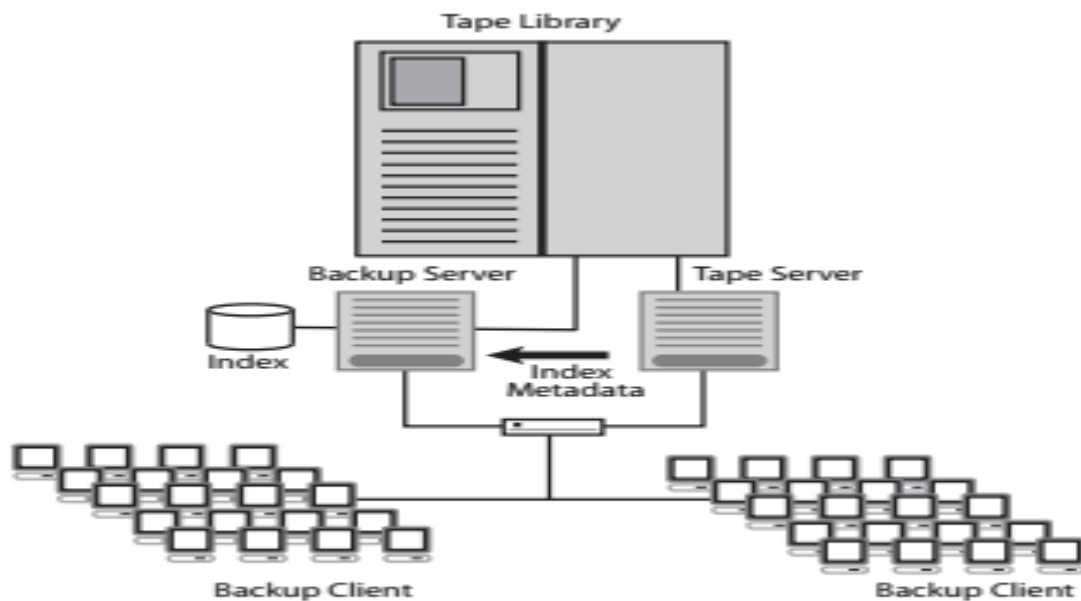


FIGURE 5.1 A backup using a shared tape library. (© Copyright 2002, StorNet. All rights reserved.)

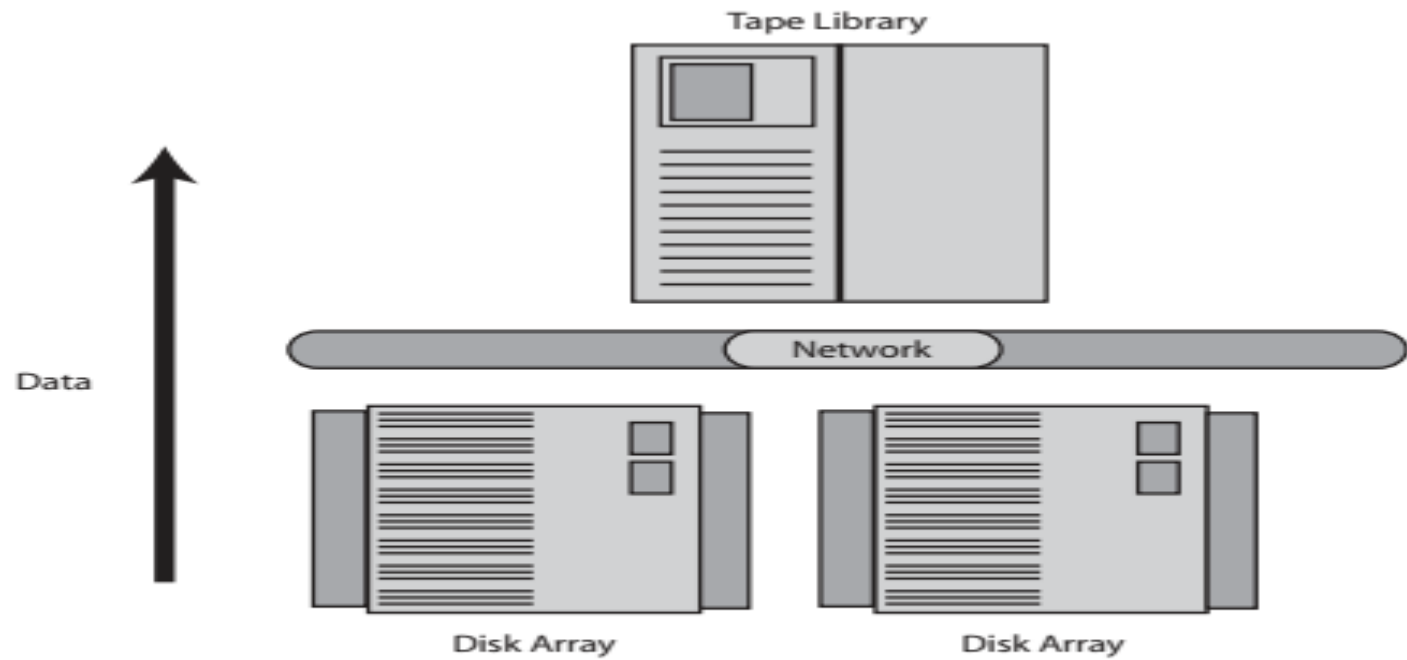


FIGURE 5.2 A serverless backup system. (© Copyright 2002, StorNet. All rights reserved.)

The Network Data Path

- **Centralization of a data-management process such as backup and recovery requires a robust and available network data path.**
- **The movement and management of hundreds or thousands of megabytes of data can put a strain on even the best-designed networks.**
- **Unfortunately, many companies are already struggling with simply managing the existing data traffic created by applications such as e-commerce, the Internet, email, and multimedia document management.**
- **Although technology such as gigabit Ethernet and asynchronous transfer mode (ATM) can provide relief, it is rarely enough to accommodate management of large amounts of data movement**

- **An enterprise class backup solution can distribute backup services directly to the data source, while at the same time centralizing the administration of these resources.**
- **For example, if there is a 600-gigabyte database server that needs to be backed up nightly, a tape backup device can be attached directly to that server.**
- **This effectively eliminates the need to move the 600-gigabyte database across the network to a centralized backup server.**
- **This approach is called a LAN-less backup, and it relies on a remote tape server capability.**
- **Figure 5.3 demonstrates how this approach is configured**

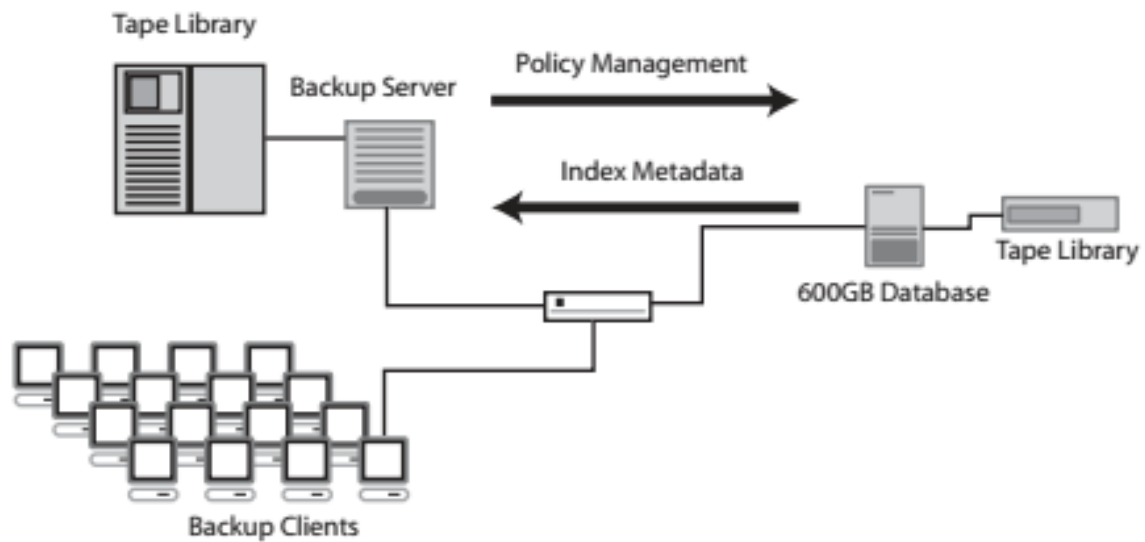


FIGURE 5.3 A LAN-less back-up using a remote tape server. (© Copyright 2002, StorNet. All rights reserved.)

- **Another option is the installation of a network path dedicated to the management and movement of data.**
- **This data path can be SCSI, Ethernet, ATM, fiber distributed data interface (FDDI), or fibre channel.**
- **Creating a dedicated data path is the beginning of a storage area network (SAN).**
- **SANs are quickly dominating the backup landscape, and applications such as server-less and LAN-less backup will continue to push this emerging technology forward.**
- **Figure 5.4 shows an example of a dedicated SAN topology**

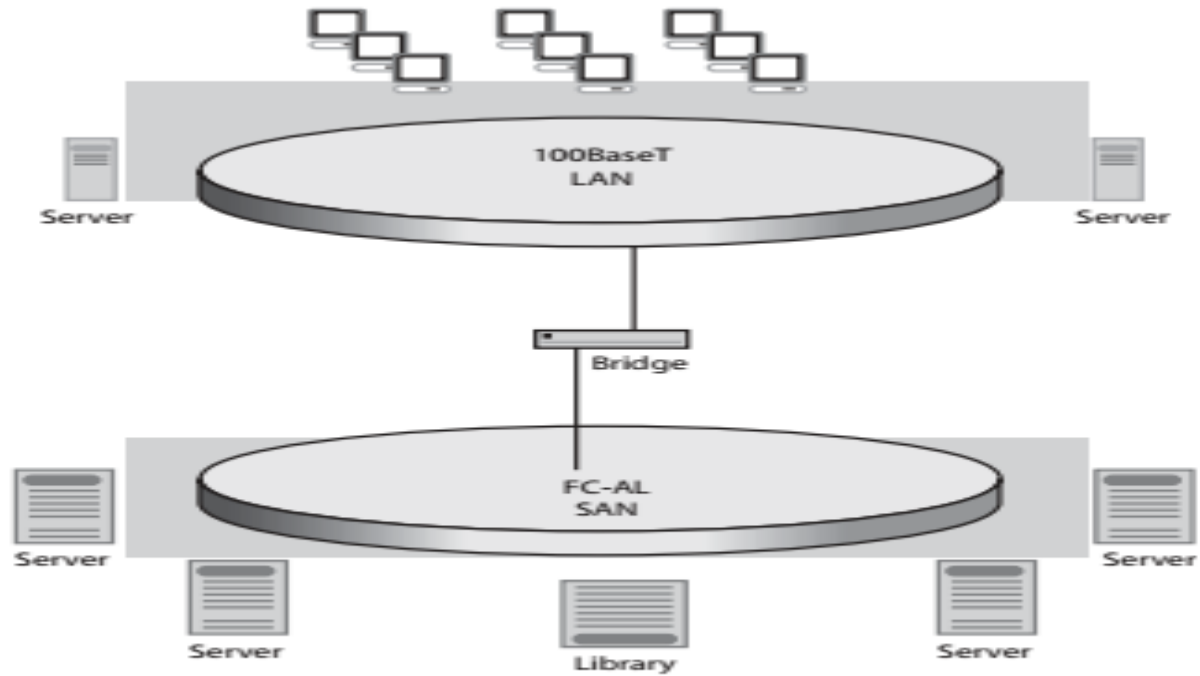


FIGURE 5.4 A storage area network using serverless backup. (© Copyright 2002, StorNet. All rights reserved.)

The Backup Window

- **Of all the parameters that drive the design of a backup application, one remains an absolute constant, and that is time.**
- **A backup window defines how much time is available to back up the network.**
- **Time plays an important role in choosing how much server, network, and resource support needs to be deployed.**
- **Today, most companies are managing too much data to complete backup during these ever shrinking backup windows**
- **In the past, companies pressured by inadequate backup windows were forced to add additional backup servers to the mix and divide the backup groups into smaller and smaller clusters of systems.**
- **However, the backup-software community has once again developed a way to overcome the element of time by using incremental backup, block-level backup, image backups, and data archiving**

Incremental Backup

- Incremental backups only transfer data that has changed since the last backup.
- On average, no more than 5% of data in a file server changes daily.
- That means an incremental backup may only require 5% of the time it takes to back up the entire file system.
- **Even then, a full backup had to be made regularly, or restoration of the data would take too long.**
- **Fortunately, there are now backup applications that combine incremental backups, thereby creating a *virtual* complete backup every night without necessitating a full backup during the limited backup window**

Block-Level Incremental Backup

- **Block-level incremental backups provide similar benefits as incremental backups, but with even more efficiency.**
- **Rather than backing up entire files that have been modified since the last backup, only the blocks that have changed since the last backup are marked for backup.**
- **This approach can reduce the amount of incremental data requiring backup nightly by orders of magnitude**
- **However, this benefit comes at a price.**
- **Often the file system of the client must be from the same vendor as the backup software.**
- **Also, there are databases such as Oracle that allow block-level backups, but the CPU requirements to do so may render the approach ineffective.**
- **Nevertheless, block-level backups may be the only viable option for meeting your backup window.**

Image Backups

- Image backups are quickly gaining favor among storage administrators.
- This type of backup creates copies, or *snapshots*, of a file system at a particular point in time.
- Image backups are much faster than incremental backups and provide the ability to easily perform a *bare bones* recovery of a server without loading the operating systems, applications, and the like.
- Image backups also provide specific point-in-time backups that can be done every hour rather than once a day

Data Archiving

- Removing infrequently accessed data from a disk drive can reduce the size of a scheduled backup by up to 80%.
- **By moving static, infrequently accessed data to tape, backup applications are able to focus on backing up and recovering only the most current and critical data.**
- **Static data that has been archived is easily recalled when needed but does not add to the daily data backup requirements of the enterprise.**
- This method also provides the additional benefit of freeing up existing disk space without adding required additional capacity

Backup Storage Devices

- **In many cases, the single most expensive item in a backup project is the backup storage device itself.**
- Therefore, it is important that the technical specifications of the storage device provide adequate capacity and performance to accommodate existing and planned data.
- **Determining the tape format, number of tape drives, and how many slots are required is predicated on many variables.**
- **Backup windows growth rates, retention policies, duplicate tape copies, and network and server throughputs all affect which backup storage device is best for your need**

- When designing a centralized data backup, take particular care selecting the right backup storage device.
 1. Make sure it can easily scale as your data rates increase.
 2. Verify that the shelf life of the media meets your long-term storage needs.
 3. Calculate the required throughput to meet your backup window and make sure you can support enough tape drives to meet this window

Recommended Backup Features

- Today's global economy means that applications such as email, relational databases, e-commerce, and enterprise resource planning (ERP) systems must be accessible and online 24 hours a day.
- Therefore, these applications cannot be shut down to perform administrative tasks such as backup.
- **A backup vendor should provide agents for the most common database and email applications that allow these databases to be backed up without shutting down applications.**

Data Interleaving

- To back up multiple systems concurrently, the backup application must be able to write data from multiple clients to tape in an interleaved manner.
- Otherwise, the clients must be backed up sequentially, which takes much longer

Remote Backup

- Many remote systems are exposed to unrecoverable data loss.
- Off-site locations are often not backed up at all because of the cost of deploying hardware and software remotely and the lack of administrative support in these remote locations.
- Laptop computers are especially vulnerable to data loss.
- **A backup application should have a method to back up systems across a WAN or over dial-up connections**

Global Monitoring

- **Companies are deploying applications that can be managed and monitored from any location in the enterprise.**
- Backup applications also need to be able to be accessed and administered from multiple locations.
- **A robust backup application should be able to support reporting and administration of any backup system, regardless of location.**

Performance

- **An enterprise backup application should be able to benchmark backup data rates exceeding one terabyte per hour.**
- These benchmarks show that backup performance is limited to the hardware and network and not to the application itself.

THE ROLE OF BACKUP IN DATA RECOVERY

Many factors affect back-up:

1. Storage costs are decreasing.
2. Systems have to be online continuously.
3. The role of backup has changed.

Storage Costs Are Decreasing

- **The cost per megabyte of primary (online) storage has fallen dramatically over the past several years and continues to do so as disk drive technologies advance.**
- **This has a huge impact on backup, as users become accustomed to having immediate access to more and more information online, the time required to restore data from secondary media is found to be unacceptable.**

Systems Have to Be Online Continuously

- (7 × 24) operations have become the norm in many of today's businesses.
- The amount of data that has to be kept online and available (operationally ready data) is very large and constantly increasing.
- **Higher and higher levels of fault tolerance for the primary data repository is a growing requirement.**
- **Because systems must be continuously online, the dilemma becomes that you can no longer take files offline long enough to perform backup**

The Role of Backup Has Changed

- It's no longer just about restoring data. Operationally, *ready* or *mirrored* data does not guard against data corruption and user error.
- The role of backup now includes the responsibility for recovering user errors and ensuring that *good* data has been saved and can quickly be restored

Conventional Tape Backup in Today's Market

- Current solutions offered by storage vendors and by backup vendors focus on network backup solutions.
- **To effectively accomplish backup in today's environment, tape management software is generally bundled with several other components to provide a *total* backup solution.**
- **A typical tape management system consists of a dedicated workstation with the front-end interfaced to the network and the backend controlling a repository of tape devices.**
- The media server runs tape management software. It can administer backup devices throughout an enterprise and can run continuous parallel backups and restores.

- An alternative to tape backup is to physically replicate or mirror all data and keep two copies online at all times. Because the cost of primary storage is falling, this is not as cost-prohibitive as it once was. The advantage is that the data does not have to be restored, so there are no issues with immediate data availability. There are, however, several drawbacks to all the backup and data availability solutions on the market today

Issues with Today's Backup

- **Network backup creates network performance problems. Using the production network to carry backup data, as well as for normal user data access, can severely overburden today's busy network resources.**
- **This problem can be minimized by installing a separate network exclusively for backups, but even dedicated backup networks may become performance bottlenecks**

- Offline backup affects data accessibility. Host processors must be quiescent during the backup.
- **Backup is not host-independent, nor is it non-disruptive to normal data access.**
- **Therefore, the time that the host is offline for data backup must be minimized.**
- **This requires extremely high-speed, continuous parallel backup of the raw image of the data.**
- Even in doing this, you have only deferred the real problem, which is the time needed to restore the information.
- Restoration of data needs to occur at the file level, not the full raw image, so the most critical information can be brought back into operation first

- **Live backups allow data access during the backup process but affect performance.**
- **Many database vendors offer *live* backup features. The downside to the live backup is that it puts a tremendous burden on the host.**
- **Redirection lives on the host, and journaling has to occur on the host. This requires consideration of local storage, host CPU cycles, and host operating system dependencies.**
- **Up to 50% of all host CPU cycles may be consumed during the backup process, severely impacting performance**

- Mirroring doesn't protect against user error and replication of *bad* data.
- **Fully replicated online data sounds great, albeit at twice the cost per megabyte of a single copy of online data.**
- **However, synchronizing, breaking, and re-synchronizing mirrors is not a trivial process and influences data access speeds while they are occurring.**
- Also, duplicating data after a user has deleted a critical file or making a mirrored copy of a file that has been corrupted by a host process doesn't help.
- Mirroring has its place in backup and recovery but cannot solve the problem by itself

New Architectures and Techniques Are Required

- Backup at extremely high speed, with host-processor independence of the underlying file structures supporting the data, is required.
- Recovery must be available at the file level. The time that systems are offline for back-up must be eliminated.
- Mirroring, or live data replication for *hot* recovery also has a role. For data that must be always available, highly fault-tolerant primary storage is not enough, nor is a time-consuming backup/restore

- Backup and mirroring are complementary, not competing technologies
- To achieve effective backup and recovery, the decoupling of data from its storage space is needed.
- Just as programs must be decoupled from the memory in which they're executed, the stored information itself must be made independent of the storage area it occupies.
- **It is necessary to develop techniques to journal modified pages, so that journaling can be invoked within the primary storage device, without host intervention.**
- **Two separate pipes for file access must be created: one pipe active and the other dynamic.**
- **The primary storage device must employ memory-mapping techniques that enable the tracking of all modified information.**
- Two copies of each change must be kept, with a thread composed of all old data stored in the journaled file.

- Part of the primary storage area must be set aside for data to be backed up. This area must be as large as the largest backup block (file, logical volume, etc.).
- The point-in-time *snapshot* of changed data will be used for backup, while the file itself remains in normal operation without impacting user access to data.
- To minimize this reserve storage area for backups, the storage device must support the reuse of this area by dynamically remapping.
- **Mechanisms must be put in place to allow for the backup of data to occur directly from the primary storage area to the backup area without host intervention**
- **Host CPU bottlenecks and network bottlenecks are then eliminated. The net result will be faster user response times during *live* backup, normal network performance levels throughout the process, and no backup downtime.**

- What about restore times? Fast, non-random restoration of critical data assumes that the user can select at the file level exactly which information comes back online first.
- Here again, the primary storage and its backup software must offload that burden from the host and take on the responsibility for understanding the underlying file structures of multiple heterogeneous hosts.
- Then, at the time of restore, the indices are accessible to allow selection at the file level for the recovery of the information from the backup device.

THE DATA-RECOVERY SOLUTION

- Availability once meant that an application would be available during the week, from 9 to 5, regardless of whether customers needed anything.
- Batch processing took over the evenings and nights, and most people didn't care because they were at home asleep or out having fun. But the world has changed.
- It's now common to offer extended service hours in which a customer can call for help with a bill, inquiry, or complaint. Even if a live human being isn't available to help, many enterprise applications are Web-enabled so that customers can access their accounts in the middle of the night while sitting at home

Shrinking Expertise, Growing Complexity

- Increased availability is good, except for one fact: many systems programmers, database administrators (DBAs), and other mainframe experts are maturing.
- It takes a lot of *care and feeding* to keep applications ready for work, and the people who have maintained these environments for so long have other things they want to do.

- The complex systems that have evolved over the past 30 years must be monitored, managed, controlled, and optimized.
- Batch windows are shrinking down to almost nothing. Backups often take place while an application is running.
- Application changes take place on the fly, under the watchful eye of the change-control police.
- If an outage occurs, the company stands to lose tens of thousands of dollars an hour.
- In today's gloomy economy, stockholders don't want to hear that their favorite investment is having system availability problems.

Failures

- Certainly, hardware failures were once more common than they are today. Disk storage is more reliable than ever, but failures are still possible.
- **Logic errors in programs or application of the wrong update at the wrong time can result in a system crash or, worse, an undetected error in the database—undetected, that is, until minutes, hours, or days later when a customer calls, a reconciliation fails, or some other checking mechanism points out the integrity exposure**
- **Finally, disasters do sometimes strike, and most often they occur without warning. Flooding doesn't always occur when it's convenient; tornadoes never do. Hurricanes and earthquakes are big-ticket events that ruin everyone's day.**
- **When they strike data center, wipe out your processing power, or even destroy your basement-level backup power supply, you have a lot of recovering to do**

Budgets and Downtime

- Does anyone need a reminder that budgets are tight? You have fewer resources (people, processing power, time, and money) to do more work than ever before, and you must keep your expenses under control.
- Shrinking expertise and growing complexity cry out for tools to make systems management more manageable, but the tools that can save resources (by making the most of the ones you have) also cost you resources to obtain, implement, and operate.
- **Businesses today simply cannot tolerate availability problems, no matter what the source of the problem. Systems must remain available to make money and serve customers.**
- **Downtime is much too expensive to be tolerated. You must balance your data management budget against the cost of downtime.**

Recovery: Think Before You Back Up

- One of the most critical data-management tasks involves recovering data in the event of a problem.
- **For this reason, installations around the world spend many hours each week preparing their environments for the possibility of having to recover.**
- **These preparations include backing up data, accumulating changes, and keeping track of all the needed resources.**
- **You must evaluate your preparations, make sure that all resources are available in usable condition, automate processes as much as possible, and make sure you have the right kind of resources.**

Evaluate Your Preparations

- Often the procedures that organizations use to prepare for recovery were designed many years ago.
- They may or may not have had *care and feeding* through the years to ensure that preparations are still sufficient to allow for recovery in the manner required today
- **Here is a simple example: say an organization has always taken weekly image copies on the weekend and has performed change accumulations at mid-week.**
- **Will this approach continue to satisfy their recovery requirements? Perhaps. If all of the resources (image copies, change accumulations, and logs) are available at recovery time, these preparations certainly allow for a standard recovery.**
- **However, if hundreds of logs must be applied, the time required for the recovery may be many hours—often unacceptable when the cost of downtime is taken into account.**

- This example illustrates the principle that, although your recovery strategy was certainly adequate when it was designed, it may be dangerously obsolete given today's requirements for increased availability.
- **What if a required resource is damaged or missing? How will you find out? When will you find out? Finding out at recovery time that some critical resource is missing can be disastrous!**

Don't Let Your Resources Fall Through the Cracks

- The previous example was unrealistically simplistic.
- Many organizations use combinations of batch and online image copies of various groups of databases, as well as change accumulations, all staggered throughout the week.
- **In a complex environment, how do you check to make sure that every database is being backed up? How do you find out whether you are taking image copies (either batch or online) as frequently as you planned? How do you determine whether your change accumulations are taken as often as you wanted? What if media errors occur? Identifying these types of conditions is critical to ensuring a successful recovery.**

Automated Recovery

- Having people with the required expertise to perform recoveries is a major consideration, particularly in disaster situations.
- For example, if the only person who understands your IBM Information Management System (IMS) systems (hierarchical database system) and can recover them moved far away, you're in trouble.
- However, if your recovery processes are planned and automated so that less-experienced personnel can aid in or manage the recovery process, then you're able to maximize all your resources and reduce the risk to your business.

- **Automation takes some of the human error factor and “think time” out of the recovery equation and makes the complexity of the environment less of a concern.**
- **Creating an automated and easy-to-use system requires the right tools and some planning for the inevitable, but compared to the possible loss of the entire business, it is worth the investment.**
- **With proper planning and automation, recovery is made possible, reliance on specific personnel is reduced, and the human-error factor is nearly eliminated**

- Creating the recovery job control language (JCL) for your IMS systems is not as simple as modifying existing JCL to change the appropriate names.
- In the event of a disaster, the IMS recovery control (RECON) data sets must be modified in preparation for the recovery.
- **RECON backups are usually taken while IMS is up, which leaves the RECONs in need of many clean-up activities before they can be used to perform a recovery: deleting online log data sets (OLDS), closing logistics supportability (LOGS), deleting subsystems (SUBSYS) records, and so on.**
- **This process often takes hours to perform manually, with the system down, equating to lost money.**
- Planning for RECON clean-up is an important but often-overlooked step of the preparation process; discovering a deficiency in this area at disaster recovery time is too late.

Make Recoveries Efficient

- Planning for efficient recoveries is also critical. Multithreading tasks shorten the recovery process.
- Recovering multiple databases with one pass through your log data certainly will save time.
- Taking image copies, rebuilding indexes, and validating pointers concurrently with the recovery process further reduce downtime.
- Where downtime is costly, time saved is money in the bank. Any measures you can take to perform recovery and related tasks more quickly and efficiently allow your business to resume faster and save money

Take Backups

- After you've thought about and planned for your recoveries, it's time to think about executing your plan.
- **Clearly the first step to a successful recovery is the backup of your data. Your goal in backing up data is to do so quickly, efficiently, and usually with minimal impact to your customers.**
- **If you have a large window where systems aren't available, standard image copies are your best option.**
- These clean copies are good recovery points and are easy to manage

- If, however, you need to take backups while systems are active, you may need some help.
- **You can take advantage of recent technological changes in various ways.**
- **You might need only very brief outages to take instant copies of your data, or you might have intelligent storage devices that allow you to take a snapshot of your data.**
- Both methods call for tools to assist in the management of resources.