



Critical Concepts of Data Protection



Abstract

The protection of data has always been an important part of the IT world. Recently, it's become a critical function that leaves no room for error. Ineffective backup or restore processes can slow or halt daily operations, resulting in costs running into the thousands or even millions of dollars, depending on the degree of automation in an organization. GST's Research & Engineering group has identified a number of critical concepts that must be understood by anyone wanting to contribute to the safeguarding of critical data and programs for the enterprise.

Some background on the theory and trends in data protection is provided, along with specific guidelines for safeguarding data and applications for the enterprise.

A dedicated manufacturer of tape storage solutions, GST, Inc. produced this GST Research Report as part of its commitment to provide information leading to improved safeguarding of mission-critical applications and data within the IT industry.



Table of Contents

Executive Summary	3
Introduction	4
1. Source and target	4
2. Types of backups	5
3. File lockout vs. fully available	5
4. The shrinking backup window	6
5. Growth of faster backup solutions	6
6. Growth of larger capacity backup solutions	6
7. Automation of the backup process	7
8. Elimination of the backup window	7
9. Get data off your premises quickly	8
10. Give the person backing up the system proper authority	8
11. Expanding data volumes	8
12. The backup dilemma	9
13. The danger of computer downtime	9
Summary	9
About the author	10
Feedback	10
About GST, Inc.	10
Trademarks	10



GST Research Report

Data Protection

Executive Summary

Today's IT systems permit the interconnection of a wide variety of servers and other devices in more complex networks; these often require complex backup strategies to protect them. Such data protection scenarios require a thorough understanding of backup options and which approaches are best under a variety of circumstances

The importance of understanding the concepts presented here is to enable the development of data protection strategies that contain the necessary depth to withstand a wide range of threats to the enterprise's critical information resource. The need for such a capability is well-understood by Risk Managers today, and only ignored at an organization's peril.

The enterprise's data is now considered one of its most critical assets. Even the loss of critical data for as little as a few days can result in the eventual failure of the business, as documented in the aftermath of the first World Trade Center bombing on February 26, 1993.

Recent Federal, state and local legislation in the U.S. and Europe has focused on the mandated protection of data of specified types in regulated industries. Penalties for non-compliance to these mandates include high fines and prison terms for senior officers. This has spurred a renewed interest in data protection at the boardroom level.



Critical Concepts of Data Protection

The guidance provided here is meant to assist organizations in developing better data protection strategies. This results in lower risks within the organization and better backup and restore functionality at the operational level. Often, improvements in strategy or technique can be implemented with no increase in cost to the organization.

Introduction

Top management's understanding of basic concepts that underlie data protection today is essential. It ensures that management can properly exercise its function to assure that the necessary data protection measures are implemented within the enterprise. A basic knowledge of important concepts is especially necessary where top management does not possess the technical skills to assess the individual backup, recovery, archiving and disaster recovery practices in effect within the organization.

This paper provides some basic fundamentals that top management must understand to discharge their role as the provider of policy-level guidance in the areas of data and application protection. These areas are of great concern to management, all the more so because of legislated mandates by oversight bodies that carry stiff fines and prison sentences if not followed.

Top management must be committed to the accurate and efficient performance of all data protection strategies, including backup operations, recovery, archiving and disaster recovery measures. Moreover, anyone who wants to effectively participate in the assessment of data protection strategies must understand today's critical backup/recovery issues and the environment they in which they operate.

1. Source and target

All backups involve at least one source device, which contains the data being backed up, and one target device, where the backup data is written for safekeeping. We read from the source device and write to one or more target devices. The data on the source device is usually stored on disk media. The media on target backup devices is usually removable so that it can be stored in a remote and safe location. The vast majority of backup media used today is magnetic



GST Research Report

Data Protection

tape; CD-ROM disks are also used. Sometimes there are dual targets that permit making multiple copies of the same backup data.

2. Types of backups

One way that backup operations are characterized is by the way that data is grouped in the operation.

Physical backup: This type of backup copies a byte-for-byte mirror image of the data from disk files to the target media.

Logical backup: This type of backup copies the logical structures from the source file to the target device.

Physical backups are generally faster and can be less prone to error than logical backups. Being faster, they can reduce the time for a backup over the time required for a logical backup operation when entire volumes/files must be backed up.

However, in many cases only a few records might change between backup operations, and since logical backups only backup data that has been changed, use of logical backup operations can result in backing up much less data where few changes to the file(s) being backed up have occurred.

3. File lockout vs. fully available

When a file is being backed up, there are two possibilities regarding its availability to users:

Read-only dumps: the file being backed up is frozen in time so that it cannot be altered during the backup operation. This is called a *file lockout* or *off-line* backup and users may be able to read but definitely not update the file.

Full availability dumps: the file being backed up is also available to be accessed and updated. This is called an *on-line* backup. To accomplish this process, various schemes exist including the mirroring of the database onto a second mirrored file and then backing up from that mirrored set of source files.

Read-only or static backups are much simpler and place less of a burden on the backup operation, but they make the source files unavailable during the backup operation. These read-only or off-line backups generally perform better than on-line backups and are used until users can no longer be locked out of their files during backup operations.



4. Shrinking backup window.

As production jobs on the computer expand, this leaves less time for backup operations.....the backup window begins to shrink. At the same time, as we noted above, the dilemma is that data to be backed up is growing. So the backup window is shrinking while the work that must be done in that window is growing. This has put great pressure on IT organizations to find approaches to backup that are faster and more resilient. Backup solution manufacturers have worked to produce more robust media and drives to meet this need; the result being a wide variety of backup solutions to meet almost any backup need today.

5. Growth of faster backup solutions.

In the past few years, tape manufacturers in particular have formed alliances and consortia to bring more resources together for the development of faster tape backup solutions that can keep up with backup requirements within the context of a shrinking backup window. The result was a consortium of IBM, HP and Seagate (now Certance) that produced the Linear Tape-Open (LTO) tape technology with speeds in the Gigabyte per hour range. Sony followed with their Advanced Intelligent Tape (AIT) drives....then Quantum developed their Super DLT drives....all performing at the hundreds of GB an hour speeds. Each of these new tape technologies has its own unique strengths. Collectively, these faster tape drives have meant that the smaller and mid-sized computers can now match the giant mainframe computers in backup speeds, and the backup window can be further shrunk with the faster backup operations possible.

6. Growth of larger capacity backup solutions.

Speed is important, as we have just seen. It's one of the three legs of the backup stool; the other two being *capacity* and *reliability*...all three need to be addressed in any backup strategy. Capacity is important for several reasons. First, many backup operations used a single tape drive to backup a server; if the tape in the drive fills up during the backup operation, then an operator is need to change the tape... a costly and error-prone process. The ideal, is to perform a backup with no operator intervention...not just to cut down on the cost of hiring an operator to switch tapes, but because tape handling can introduce costly operator errors. The goal: backup with no operator intervention. This can be done in three different type of backup devices:

Single or Dual Tape Drives: the unattended backup capacity here is the capacity of one or two tape cartridges (approximately a half TB in the higher capacity drives).

Autoloaders: these devices have one tape drive and hold between 8 and 12 tape cartridges for much larger unattended backup capacities in the multiple TB range.

Tape Libraries: these devices have multiple tape drives and hold hundreds of cartridges with unattended capacities in to the hundreds of TB.



GST Research Report

Data Protection

Since many enterprises have no more than a few hundred GB of data to backup each day, a single or dual drive high-capacity tape drive can meet their needs. As organizations grow, they tend to convert to tape autoloaders for larger backup needs, and eventually to a tape library for very large backups and more sophisticated backups that take advantage of multiple drives all backing up different servers at the same time.

7. Automation of the backup process.

An important part of any backup process or strategy is how the overall backup policies and underlying procedures are carried out. In the past, manual execution of backup procedures was the norm. In the past few years, these procedures and the policies they enforce have been automated to a considerable extent. In fact, many organizations run a “lights out” operations environment, at least at nights and on weekends and holidays, where all operations including backup are performed completely automatically. New tape management and backup software initiates backup jobs at the scheduled times, positions the correct tape cartridges for backup, executes the backups and returns the tape cartridges to their destination for pickup and delivery to the arranged Disaster Recovery Center.

This automation process has many advantages. Procedures are verifiable and even auditable. They occur more efficiently and with no operator assistance or chance for operator error. This ensures both the integrity of the backup process, improves speed and reduces the likelihood of destroying or losing important backup tapes.

8. Elimination of the backup window.

A logical extension of automating the backup process is the elimination of any backup window. This allows all computer time to be devoted to production time, except when downtime must be scheduled to upgrade or repair the hardware. Such a condition permits a 24x7 operation (now frequently referred to as 24x365 to include operation over holidays).

Computer operations that require continuous availability of the computer for mission-critical applications often occurs in the following circumstances:

- Worldwide offices in the organization mean that at anytime there is an office open somewhere.
- Internet applications, like Internet Commerce for taking orders.
- A 24 hour everyday online Order Department.
- Supply chain applications that require constant availability to suppliers and vendors.
- Customer support applications.
- 24 hour teller or other banking/financial applications.
- 24 hour medical or police support applications.



GST Research Report

Data Protection

To eliminate the backup window to achieve continuous availability of computer operations is no simple task. It generally takes at least two servers, with sophisticated software that mirrors all data and applications from the source server to the target server, enabling backup operations to be performed from the target server while production jobs are continuously executed on the source server.

9. Get data off your premises quickly.

Hand in hand with backup and recovery of lost data, is the off-site protection of data. The best destination for backup data is to get it to a safe and remote location as quickly as possible. This can be a secure storage room/vault, or a Disaster Recovery Site where computers are available to replace your computer operation should your computer go down for an extensive time or be destroyed in a disaster.

A key element is transferring backup data to a DR site is to have it removed quickly from the computer site, since as long as it sits at the site it is not safe from a disaster at the site. Another key element, is practicing what to do with the data at the DR site, in case a DR operation is ever needed; these disaster rehearsals are critical to a working data protection strategy.

10. Give the person backing up the system proper authority.

Whoever is in charge of backups and restore operations within your organization, holds the institution's lifeblood in their hands. Make sure they have enough authority to do the job fully. There should be a clear channel of communication between the person responsible for data protection and the CIO and Risk Manager within the organization. This is now doubly important with new compliance regulations like the Sarbanes-Oxley Act that mandates the way data must be backed up and stored, with stiff fines and prison sentences mandated when regulations are not followed correctly.

11. Expanding data volumes.

In recent years, the volume of data that organizations store and routinely sift through has gone from hundreds of Megabytes to Gigabytes (billions of bytes) to Terabytes (trillions of bytes) as organizations expand the complexity of their mission-critical computer applications. Larger databases became data warehouses where firms look for trends that can mean millions in additional sales revenue or savings if detected early (at least earlier than their competitors). Larger data volumes often necessitate different backup and recovery strategies to accommodate the larger amounts of data that must be protected. When very large data volumes are involved, backup software should be tested to be sure it can scale up to handle these larger data volumes.



GST Research Report

Data Protection

12. The backup dilemma.

The explosion in the size and number of data files that need to be backed up has created a backup dilemma, since the greater the amount of data involved, the more crucial that data becomes to the enterprise. This necessitates improved protection strategies. Also, more data means longer times to complete backup operations and this robs servers of valuable productive time to support on-line applications. These two countervailing trends: (1) expanding databases with resulting longer backups and (2) the need to spend as little time doing backups so more production time is available, is what is referred to as *the backup dilemma*. This is one of the biggest challenges that computer operations professionals face today.

13. The danger of computer downtime.

The failure of a backup operation, can quickly reach catastrophic proportions. This occurs when in the same day that a backup fails there is a computer failure and the system goes down. The previous backup data must be accessed to initiate a Restore Operation to restore the computer to the state it was in when the most recent backup data was collected. Each hour that the system remains down (called downtime) can cost millions of dollars for organizations where critical applications are handled on-line in a real-time environment. For every organization, there is a point, beyond which it cannot survive if computer operations are not restored.

Summary

Top management's understanding of basic concepts that underlie data protection today is essential. It ensures that management can properly exercise its function to assure that the necessary data protection measures are implemented within the enterprise. A basic knowledge of data protection concepts is especially important where top management does not possess the technical skills to assess the individual backup, recovery, archiving and disaster recovery practices in effect within the organization.

As commerce and institutions begin to move into the 21st Century and with its attendant uncertainties, awareness of the critical role of data protection strategies permeates not only the Boardroom but executive, middle and supervisory management of every organization concerned about its long-term survival.

These basic principles are part of the foundation of understanding needed to properly manage and guide the data protection function.

###



GST Research Report

Data Protection

About the author

This GST Research Report was prepared by GST's Research & Engineering group under the leadership of David Breisacher, CEO/Chairman at GST. David is the founder of several successful companies, including GST and BCC Technologies, a manufacturer of eServer disk, tape and memory storage devices. A visionary for the storage industry since the early 90's, David lends his market insight and predictions for the IBM midrange storage marketplace to the research conducted at GST. His experience in sensing shifts in technology and industry directions has made it possible for him to organize and structure successful companies to rapidly meet the evolving needs of storage users.

Feedback

We value your feedback on this GST Research Report. Please send your comments, suggestions and questions to: research@gstinc.com.

About GST, Inc.

GST, Inc. (<http://www.gstinc.com>) engineers, manufactures, markets and sells a line of innovative storage products to meet the need for high-performance, continuous reliability and cost-effective data storage. These products include tape solutions available today, and will include storage-related services, software and disk subsystems in the future. A comprehensive array of tape solutions range from single and dual tape subsystems, autoloaders, midrange tape libraries, to modular enterprise-wide tape libraries, with focus on improved backup and disaster recovery solutions. Modular design enables field upgrades, scalability, investment protection for existing GST tape solutions, and lower life-cycle costs. GST's product development is guided by several advisory boards to closely track market needs and fully utilize the latest engineering developments in product design. Complete information about products, support and company background can be found at the company's Website.

Trademarks

GST, InternalDR, EntryDR, SafeDR, AutoDR, GrowthDR, ScalableDR, Commander, BridgeLink, SanMatrix and StorMount are trademarks of GST, Inc. in the United States and other countries. AS/400, iSeries, IBM, UNIX, Linux and Windows are the property of their respective owners.