International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 11 Issue 4 –NOVEMBER 2014.

DATA RECOVERY: CHOOSING THE RIGHT TECHNOLOGIES

Dr. Y.Narasimha Rao

Director, KBN College, PG Centre

Abstract— In today's information driven organizations, the costs to generate, keep available, and recover data are staggering. With the widespread increased reliance upon data, the costs of interrupted access to data or data losses could financially compromise the organization, and in some cases, leave it no room to recover. Given that, decisions pertaining to storing and recovering the organization's data assets can no longer be the exclusive responsibility of IT experts, but must involve discussions with technologically informed C-level managers as well. With that as a backdrop, C- level managers along with IT experts must consider their organizations' business needs, state and federal regulatory requirements, and the potential threats that system downtime and potential data losses pose to the success of their organizations.

After thoroughly analyzing the business needs and potential threats to the data asset, the next activity is to apply the appropriate technologies to ensure that efficient backup schemes and adequate recovery is in place for the organization's data asset. Once a solution is in place, the organization must rigorously adhere to best practices that support the solution, and regularly test the system's ability to recover data.

I. INTRODUCTION

The concept of data recovery once prevailed as a relatively straightforward practice. Ever since the commercial introduction of magnetic-data storage devices, in 1947, users of this technology became aware of the vulnerability of the data stored on these devices. Therefore, the introduction of data recovery was quickly introduced as a means to protect the investment made in data stored on this medium. Originally, the goal was simply to create a second copy of a primary data set, storing it on a less expensive medium than the primary medium. This second copy was retained and generally only accessed in the event of a primary disk failure, where it would be copied onto a new or repaired disk.

Current need for data recovery solutions

In recent years, several trends have caused IT professionals to become more diligent in designing and maintaining data recovery systems. Some major factors to consider with regard to justifying and designing an appropriate data recovery solution include:

Growth in stored data
Cost of downtime and data unavailability
Cost of data management
Data as a core business asset

Data storage has grown rapidly in recent years. Conservative estimates from the International Data Corporation (IDC) show data expanding at 50% to 80% per year approximately, while other industry analysts place the growth rate closer to 100% per year, even in the current struggling economy. There are many data types driving this growth and at the enterprise level, some notables are:

Databases
Email
Multimedia
Enterprise Resource Planning (ERP)
Supply Chain Management
E-procurement
Content Management
Data Mining
Customer Relationship Management (CRM)
Electronic Document Management (EDM)

This statistic "55%" is meaningful from a data backup and recovery perspective because database-resident data can be challenging to successfully backup and recovery in production environments.

Email:

In addition, in 1998 the Securities Exchange Act and the Internal Revenue Service (IRS) codes were amended, requiring organizations with ties to the securities industry to maintain all electronic correspondence between employees and clients. Combined, email growth and retention factors led email communications to drive many 100s of terabytes of storage in the United States alone in the coming years.

Multimedia: Applications used for audio, video, and graphics creation and manipulation generate large files that consume tremendous amounts of storage. Marketing communications groups increasingly leverage these technologies as part of their overall corporate branding efforts.

Data Recovery Assessments

A basic data recovery assessment should include a review of all pertinent data types, along with the availability needs for

each type of data. An assessment of the data types should answer the following questions:

What would be the cost to the organization to recreate each type of data? The higher the cost, the more comprehensive the data recovery strategy should be. For data that can be recreated without extensive costs or delays, lower levels of protection should be adequate.

What is the cost to the organization for each hour that a data source is unavailable? Considerations include lost productivity, lost revenue, and the possible loss of customers if the period of unavailability reaches a certain threshold. This information is used to determine the organization's Recovery Time Objective (RTO).

What is the projected business impact if data is lost permanently? This information is critical when assessing the requirement for data redundancy and off-site data vaulting.

In the event of a system's failure, how many transactions could the organization afford to lose? In other words, how frequently must backups be performed? This information is used to determine the organization's Recovery Point Objective (RPO).

How much regularly planned system downtime or slowdown time can the production environment tolerate for backup activity? In other words, what is the backup window?

How much administrator time is currently allocated to backup activities per terabyte of data? Can this ratio be sustained at your current data growth rate? Consider technologies that would automate and centralize backups to improve storage administration efficiency. This should be done with a clear understanding of the potential Return on Investment (ROI) for each technology. The key is to assess those that would yield the most benefit to the organization.

What is the retention period for each data type? Retention period requirements affect many infrastructure design decisions, such as the storage medium, rotation policies, library size, vaulting service selection, and backup data redundancy.

Is there a requirement to store a copy of the data off-site? Off-site archival of backup media can influence several decisions, including backup software, vaulting service, and storage media.

Is this data currently mirrored or replicated within the environment or at a disaster recovery site? Duplicate copies of the produced data enable the consideration of such activities as off host backup, using a third copy of data, or off-site backup performed at a replica site.

Data types and uses

How much data of each type is currently stored within the organization?

What is the rate of growth for each data type?

What is the configuration of the server(s) that hosts this data, if applicable? Include make/model, number of processors, operating system and level, and a list of applications hosted by each system. This platform information can affect the data recovery infrastructure.

Is workstation-resident data currently protected by backup software? If so, how much administrative time is dedicated to workstation backup?

How long would it take to rebuild a workstation in the event of a failure, or if it were stolen?

What would be the cost to the organization if this data were not recoverable?

Are there legal or regulatory requirements for data retention?

Mapping Technologies to Business Needs

Needs-based application of technology

Upon completion of a comprehensive recovery assessment, the business needs can be mapped to available technologies that would ultimately render an optimized solution for the organization.

File-level and block-level backups

The host processor, during a file-level backup, makes numerous I/O and system calls to locate and open the header of each file for backup.

Server slowdown due to high processor utilization

Poor tape performance, resulting from frequent rewinds to establish tape repositioning (known as "shoe shining" due to the back and forth motion of the tape on the tape heads)

Accelerated wear and tear on tape heads and tape resulting from shoe shining.

LAN-free backup

LAN-free backups transfer data across the network in order to centralize tape administration and leverage network resources. See Figure 1. This is a good method, if the backup activity can be contained within an acceptable backup window, and the LAN is not used by other applications during backups to prevent network saturation

Server less (server-free) backups

The concept of server less backup has been touted as an up-and-coming technology by vendors for several years, but in reality it has trickled into the marketplace rather than taking it by storm.

The allure of server less backup is that it allows backups to occur at any time, without taxing the production server with extensive file system thrashing or I/O processing.

File-system data snapshots

The data that exists at the time of the snapshot is protected from being overwritten on the physical disk so that it may be referenced from the snapshots.

This enables consistent static access to files at an identified point in time, which offers tremendous benefit to both backup and recovery processes.

Backup to disk

By utilizing disks as part of a data recovery infrastructure, an IT operation can improve its ability to meet its backup windows and RTO.

Backup to disk: Traditional disk storage, often ATA based, can be configured as the target for backup data. While this approach can yield backup performance benefits in some environments, its primary benefits are backup reliability and recovery performance. When backing up to disk, it is generally recommended that a copy of the backup data be written to tape or replicated off site, for disaster recovery and archive purposes.

 Tape emulation: Another approach to introducing disk into the backup infrastructure is to utilize emulation technology as a front end to the disk system, thereby presenting disk to the backup application in a way that makes it appear as tape. Tape emulation can yield a couple of benefits over traditional backup to disk:

It enables seamless integration of disk into the backup operations, versus traditional disk which, generally requires some reengineering of the backup workflow.

It results in measurable performance gains without file a backup client to the backup server. It may also be advisable to encrypt the data as it is written to tape, so that as the tape is placed on a shelf or transferred off-site it does not fall prey to malicious tampering.

Bare matel restore

As applications and servers evolve IT personnel often tweak and manipulate them in subtle ways to improve functionality, performance, and stability. Unfortunately, these changes are not always documented fully or clearly in a change control process, in spite of good intentions on the part of the organization. Virtually every IT group has a horror story about the recovery process of a server that failed, where the server had undergone many undocumented modifications over the course of its useful life.

Consequently, before the system's data could be restored, the operating system must be manually recovered, patched, and configured. Then the file system and storage volumes must be arranged exactly as they were before the crash occurred. The slightest error or misalignment in this process can, and often does, result in a server being less than fully functional, due to subtle nuances in the mapping of data to applications, operating systems and file systems.

In order to protect workstations using a traditional backup product, the workstations need to be running in a stable state during the backup period, requiring both a disciplined user and a robust workstation operating system.

Regular full backups as part of a typical backup procedure require substantial bandwidth, which may be challenging in large organizations with limited backup windows, and relatively impossible for remote users with dial up connections.

Systems rebuild time is prohibitively long for workstations, even with a traditional backup process in place and given system installation and configuration times.

In most organizations, it has been determined that it was too resource intensive to include workstations in the enterprise backup operations.

As a workaround, users were generally instructed to copy their backup-worthy files to a network drive, where they would be backed up regularly. In theory, this sounds acceptable, but in practice most organizations find that users lack the discipline to adhere strictly to this standard, leaving countless valuable files unprotected.

II. Conclusion

In today's competitive marketplace, business data is expanding at rates that outpace the technologies used to manage this growth. This puts tremendous pressure on IT staffs to develop timely, effective, and cost-efficient solutions to meet the data recovery needs for their evolving and expanding data. Given the need for timeliness and the cost factors associated with data recovery in data rich organizations, it may be wise to consider a partnership with an independent storage architect such as Data link Corporation. Data link's expertise is delivering custom storage infrastructures that meet business needs and maximize the value of data.

III. References

[1] http://en.wikipedia.org/wiki/Magic_number(programming).
[2] Carrier, Brian. File System Forensic Analysis. Addison Wesley Professional, 2005.
[3] Forensiks Wiki. Forensics Wiki. AFF. [Online] [Cited: Mar 29, 2011.] http://www.forensicswiki.org/wiki/AFF.
[4] Bunting, Steve and Wei, William. The Official EnCE Enchase Certified Examiner Study Guide. Indianapolis In: Wiley Publishing, Inc., 2006.
[5] United States Computer Emergency Readiness Team. USCERT Vulnerability Note VU#836068. US-CERT: United States Computer Emergency Readiness Team. [Online] [Cited: March 5, 2011.] http://www.kb.cert.org/vuls/id/836068.
[6] Xiaoyun Wang, Yiqun Lisa Yin, Hongbo Yu. Collision Search Attacks on SHA1. 2005.