Disaster Recovery in GIS

Introduction

Today enterprise geographic information systems (GIS) are mission-critical for the data center. As more and more organizations implement spatially enabled applications it is increasingly important that these applications remain up, running and available under all circumstances. Spatially enabled applications are often the applications most needed in an emergency. Consequently, GIS being down for any reason is not an option.

The typical data center today has evolved into a complex, inflexible environment that can be costly to manage and difficult to recover in the event of a disaster. "Disaster" is a relative term and can mean anything from loss of an important file to loss of an entire site. A disaster can come in many forms, including but not limited to operator error, malicious virus, hardware failure, or act of nature; and can also result from system maintenance and planned downtime. The fact is, for one reason or another systems go down, and for many organizations this means lost revenue or the inability to serve customers. When GIS is involved that may also mean the inability to save lives.

Consequently, it is not only important to plan for disasters, but to also have a methodology in place for fast recovery. Downtime must be reduced if not eliminated. This may mean a disaster recovery (DR) strategy that includes offsite data replication and the ability to run operations remotely or from an alternate site.

Today's Complex Data Center

Data and system availability is the number one priority in today's IT and GIS departments. Unfortunately this is difficult to achieve with traditional technology and even more difficult to restore when something goes wrong. The complexity of today's IT and GIS environments makes sustaining availability or recovering data and systems a daunting task, and a challenge that many managers find difficult if not impossible to meet.

How can we gain control of our data and recover quickly? A growing number of organizations turn to virtualization as the answer. Storage is at the heart of the data center and enterprise GIS; therefore storage is the place to start if you want to gain control of the data. Virtualization provides cost-savings and efficiencies, and when deployed correctly, enables organizations to be more flexible and able to respond to disasters and recover operations more quickly.

Step One: Storage Virtualization

When you hear "virtualization" technology most people immediately think "VMWare™" or similar virtual server technologies. Virtualization is an industry buzzword and products like VMWare™ are becoming increasingly popular. But to create a truly virtual data center managers must take a holistic approach and seek solutions that leverage all computing resources. Processing power is only one resource. Virtualization should really begin with your backend storage environment. Starting there gives you tremendous power over your data.

Virtualization is available with today's more advanced storage technologies. Using virtualization, storage managers can provision storage for multiple servers, automatically classify and migrate data to the appropriate tiers of storage, and create unlimited numbers of recovery points using snapshot technology. These features make it possible to quickly respond and completely recovery data and systems.

If you are free from the limitations of traditional storage you can truly virtualize your data and in effect separate the hardware from software. Storage virtualization frees you from the restraints associated with managing RAID sets and traditional backup and gives you the power to recover data and systems quickly and completely.

Some storage solutions today offer "snapshot" technology. Snapshots create point and time copies of data and can be an excellent alternative to traditional backup. Snapshots offer protection against data loss and corruption. A virtual storage solution includes snapshot capability with the ability to take an unlimited number of snapshots in a space-efficient manner, and keep them as long as necessary. With unlimited snapshot capability managers can create a complete matrix of protection around the data.

Storage virtualization also brings the ability to "boot from SAN". With boot from SAN the operating system, applications and data all reside on the SAN. Servers are merely the processing "engines". Booting from SAN reduces complexity, lowers costs and accelerates server recovery. With boot from SAN you are no longer tied to the physical limitations of server hardware. When a server goes down you do not have to stop operations while you figure out what went wrong; you can simply redirect processing to a spare server and be up and running again in boot-cycle time. For GIS this can be critical.

In organizations with more than a few servers, deploying and updating server software can be a laborious task. No one knows this better than the GIS manager. Applying software patches and upgrades can bring unexpected results, but with boot from SAN you can easily create Test Server instances, apply your patch to the test environment and test it thoroughly before bringing it into production. Then when you are ready to roll out the new version you can be confident the rollout will go smoothly. If it doesn't, you can go back to the server instance prior to that patch and bring things back up while you figure out what went wrong. Downtime for system upgrades becomes a thing of the past. Boot from SAN offers managers numerous advantages and should be a key component of any DR strategy.

Step Two: Server Virtualization

If you walk into most traditional data centers you will see server after server underutilized and running at a fraction of its capacity. When server hardware is underutilized it makes sense to virtualize and run applications across fewer servers. Server virtualization may not be appropriate for every application and every server, but when used correctly it can bring efficiencies and cost-savings.

Server virtualization can make DR easier, provided the storage architecture is also virtualized. Without storage virtualization virtual servers may create more complexity and make DR even more difficult. Managing virtual servers can be complicated and traditional storage technology often lacks the flexibility needed to optimize a virtual server environment. This may explain why many managers in a virtual server environment still lack a comprehensive data recovery strategy. Such problems can be eliminated when virtualization is applied to both servers and storage to create a truly virtual data center.

Replication

Backup and recovery when a data center is intact is one thing, but what if you lose access to the data center, or the data center itself? Real DR must provide a method for recovery in the event of site loss. Some assume the answer is storing a copy of data offsite using backup to tape or similar media. But how current is that data and how quickly can it be recovered? What about the systems and the applications necessary to access the data? How long will we be down while you put everything back together? Remote replication may be the answer.

Replication technology provides protection against site loss, yet remote replication is one of the most frequently required but least implemented technologies today. With many operations running 24x7 managers understand more than ever before the importance of DR in the data center. But when it comes to implementing a solution, the cost and complexity of traditional remote replication offerings have prevented widespread use. In planning for DR most organizations want the ability to replicate data to a remote site but find the available options expensive and complex to implement and manage. It is also very difficult to test a DR plan that involves remote replication.

The key is finding a remote replication solution that leverages the power of virtualization to speed recovery. Such solutions are available today and able reduce the cost and complexity associated with traditional replication. With the right replication technology snapshots can be replicated continuously between sites, synchronously or asynchronously, to ensure data is current. The right solution should be reasonably simple to deploy, manage and test.

Testing Your DR Strategy is Key

Choose a replication technology that can be tested while all users are connected and the system is up and running. This allows managers to easily verify that their volumes are being replicated successfully. The ability to test your DR strategy on an ongoing basis will ensure you success in the event you have to put your plan into action.

Conclusion

With today's more advanced storage, virtualization and replication technologies, it is possible to recovery the data center quickly and efficiently. Managers are able to eliminate narrow backup windows on production systems by taking advantage of features that allow physical and virtual servers to boot from SAN, by using snapshot capability to recover servers within seconds, and by using replication technology to protect against site loss.

Sustaining data and system availability in today's complex data center is a real challenge for most managers. Virtualization may be the key to simplifying the data center and recovering data and systems in the event of a disaster. The first step is storage virtualization. By virtualizing both storage and processing power managers can build resilient data centers and recover from data hazards quickly.

The result is an affordable disaster recovery solution, one that leverages the full capabilities of today's virtualization technologies. IT and GIS managers can create a complete data center DR solution by combining server virtualization with a feature-rich storage solution that includes storage virtualization capabilities.

Server and storage virtualization used together will provide real DR for today's IT & GIS.