Designing & Implementing an Effective Backup Solution

EMC Proven Professional ^{TM} Knowledge Sharing 2008

Anuj Mediratta, Director
Neeraj Mediratta, CEO
Ace Data Devices Pvt. Ltd.

anuj@ace-data.com
neeraj@ace-data.com

:

Table of Contents

Disclaimer: The views, processes or methodologies published in this compilation are those of the authors. They do not necessarily reflect EMC Corporation's views, processes, or methodologies.

Introduction

The term backup has a very simple meaning. It is making a copy of your critical data on a different physical media to ensure that you are able to recover from a disaster. Backups are intended to be used at the time of a catastrophe where there is the opportunity for huge data loss. Fortunately, this rarely happens due to enhanced legacy technologies and emerging new technologies. Even a startup organization or a small workgroup with a few servers selects well protected RAID based DAS or NAS solution.

Data stored on disks on a standalone server is protected by various levels of RAID as suited to the application. A slightly larger enterprise enjoys the benefits of implementing mid-size to high end storage offering the additional benefit of using Hot Spare technology. This gives them one or even more spare disks to replace the failed disk in event of a disk failure. Enterprises keep the primary online storage so well protected that complete data loss rarely occurs.

Backups, therefore, are more frequently used to retrieve older versions of data and make copies of your original data. These copies can be used to simulate a test and development environment. Software development firms use copies of live data for testing. Recently, enterprises have also begun to create test environments for their mail servers. They prefer to test every new solution on the test environment before implementing it on the live setup.

One enterprise implemented EmailXtender® for their Lotus Domino servers; online journaling was on. One day, their mail server crashed. The Domino service provider thought that the newly implemented EmailXtender implementation was faulty. Now, we are doing everything on the test server. We have been rigorously testing and have found no problem in implementing and using any EmailXtender functionality in the simulated environment.

What needs to be backed up

Deciding what to back up is the most important challenge when implementing a backup solution. This would appear to be a very simple issue but enterprise decision-makers typically respond that they need to backup everything. This may not be always possible. It is important to know what is required to be backed up and what backup cycles are expected.

First, is the customer prepared to invest in a solution that backs up everything? Let us consider backing up a desktop. All desktops would have large amounts of non-business critical data ex. MP3, AVI files. Thanks to EMC Retrospect® for enabling the administrator to control and exclude certain file types from the user's desktops so that you can back up only required desktop data while keeping it transparent to the user.

Every executive may have multiple pst or nsf files. These would be the various mailing files he has been using since starting his career. Even in the current job, the organization would have a mail archiving process but the archived pst or nsf resides on the same desktop/laptop. Do we need to backup the same archived mail file daily? For most data types, an ideal implementation of EMC Retrospect makes it possible to avoid multiple backups of data files that have not changed since the last backup. It helps in smaller backup windows and occupies less space on the backup device.

Similarly, other than application servers, many servers have a large amount of unchanged data. We were working with an enterprise storing geographical data (maps etc.). The data kept growing and so did the backup window. The customer decided to upgrade to a Giga network for several reasons including larger backup windows. Eventually, a full backup would start on Friday evening and end on Tuesday evening. Most of the backup resources were busy backing up his server thereby hampering other backups. The customer was demanding a more robust backup solution so we audited the environment.

We discussed several issues with him and used our file manager utilities. At the end of the audit, we realized that 1.6 of the 2.8 Terabytes of data being backed up weekly had not been modified or used for more than 8 months. The customer was wasting resources in terms of backup window, network bandwidth and backup media by backing up the same data every week.

We immediately recommended and implemented DiskXtender® along with a low cost SATA disk shelf to automatically archive the unchanged data to a low cost SATA disk. This helped to utilize SAN storage more effectively and reduced the backup windows drastically. There was no need to upgrade the backup device to a higher speed, higher capacity device.

Expected frequency of recovery requests & Recovery Time Objective

What frequency of recovery requests does the enterprise expect? And how long do you think they can afford to wait for a recovery request to happen? Most people using EMC NetWorker® believe it is ideal to use the default pool and back up all data on tapes in the default pool. They still go ahead with multiple groups based on their desired backup schedules and levels. In reality, using a single Default pool only provides ease of management.

Backups should be divided over Media Pools & Groups for a more manageable backup system to run. It makes it easier to trace which set of media has your desired backups for recovery. This is particularly useful when you have to backup databases and you need the fastest possible recoveries.

You would put them all in the same pool with multiplexing ON. NetWorker® writes them on the same media block. While reading the media for recovery, the entire length of the tape needs to move across the head as it reads sequentially. If the tape backing up a critical database also has some other data, recovery would need that extra time to skip this unwanted length. NetWorker would need to skip many blocks from the media because the blocks won't have the data to be recovered. If you have dedicated media for a database backup, a large set of successive blocks would have the entire database. NetWorker would not be required to skip blocks thereby speeding up the recovery process.

You should always keep highly critical backups on different media especially if using a tape based backup device. This limitation can be avoided by using a fast disk based backup system where the read/write access on the device is done randomly.

How long to retain your backups

The other big question is how long does the customer want to retain their backups? Most want the backup tapes forever but this is not possible as tapes need to be recycled regularly. Older data can be stored on a set of tapes in a vault with critical old data archived, but then you need enough media and other resources to keep and store inventory.

Infrastructure, such as your network resources and the backup window that is utilized while backing up unchanged data, is also needed. While recovering an important file from the same tape media which has this data, the recovery window becomes large as the entire media needs to be scanned. The best solution is to backup this data on a longer life media and keep it in an archived state. Even if it resides on the same primary media, ignore it from the regular scheduled backups.

This is important when sizing the backup device and suggesting the backup policy while implementing a solution. For example, perform lesser wanted data backups once a month or quarter, and place a higher retention on them. As an organizational policy, if some one needs older data, he can obtain it from monthly or quarterly backups only. For recent and more wanted data, perform a daily backup. This helps to quickly answer a recovery request and has more versions available for recovery if a recent recovery has been requested.

Network issue

Network availability is a key factor for smooth backup operations. Backup performance depends on the network available for backups. We should ask questions like: What is the network bandwidth in use? How much load does the network already have? Many people would like to purchase high performance backup devices. We do understand that price benefits are more for new technology and no one wants to invest in an outgoing technology until there are monetary constraints, but expecting good backup performance just because you bought a high performance backup device is not the right expectation.

Despite the fact that the device is capable of writing up to 100 MB/s, it practically ends up writing @ 10-15 MB/s. This is because the data has to travel across the network before getting backed up and network performance is a key issue. If the network infrastructure is already being used heavily by other applications, backup data takes longer to travel.

Upgrading the network can be essential for a faster backup. In large enterprises where we have implemented NetWorker, we prefer to configure the backups across a separate dedicated backup network. This ensures that the backup setup is not impacting the existing customer infrastructure and that the backups and recoveries happen much more smoothly across the dedicated backup network.

In many implementations, we have seen backup fail due to network issues while the backup administrator and network administrator claim their network is working fine. These issues can be caused by improper network port configuration on the NIC or the switch. As a best practice, NetWorker should be configured with a NIC & switch port set to Full Duplex, 100 MBps OR 1000 MBps whichever is possible. Setting NIC or switch port to Full/Half Duplex, auto-negotiate does not give you a smooth network connectivity thereby failing backups intermittently.

Installing the backup device remotely

Installing the backup device remotely is one way to resolve the network bandwidth issue. You could either install the backup device directly attached to the large database server or you could install the backup device in the same network subnet. By doing so, you are directing the backup data to flow to the directly attached backup device or the backup device attached in your own subnet. The entire enterprise network is not affected in such cases. If the backup device is directly attached, you would achieve a faster throughput that would also relieve your application server more quickly.

It is also important to determine the sequence of servers to initiate backups. Large databases and file servers should be started first; smaller ones can be scheduled in between. The optimum performance of a backup device is largely dependent on the throughput at which the device receives the data to be written. The network plays a critical role here. Configuring the solution with the right approach can sometimes overcome network limitations. The challenge is to know the size of the data to be backed up.

We implemented a backup solution for a small enterprise with only 6 servers. As time progressed, the enterprise grew and is now hosting approximately 60 servers. We upgraded the backup device due to capacity requirements and also used the dedicated network. The enterprise continues to grow and currently hosts about 150 servers in a heterogeneous environment. Backing up all of them on a single backup device was becoming impossible. Even the dedicated backup network could not help us due to a steep increase in data size. All server groups wanted their own backup schedules and policies based on unique requirements.

We have further optimized the solution using the EMC NetWorker's storage node capabilities. We have grouped the servers based on their network subnets and applications groups and configured smaller storage nodes for their local setup. Backups are centrally controlled through NetWorker Management Console and the data flows smoothly to its respective storage node.

The critical applications servers are backed up through a dedicated network on the backup server. For faster backup performance, we are using an EMC Disk Library so the backups work very well. We integrated a tape library with the backup server to ensure that the customer can ship his tapes off-site. Data is cloned onto the physical tapes and shipped off-site during off-peak hours.

The application server's configuration and availability of resources work in conjunction with network bandwidth to ensure smooth backups. You need to know when the application server is least used. It would not be a good idea to perform backups while several users are accessing the online databases for their transaction processing.

Hardware & Application Resource Utilization

A backup application typically requires 25-30% memory and processor cycles from the server. Initiating a backup during peak or even standard working hours is likely to slow users' access to their applications and data.

In one case, a backup administrator initiated a backup during working hours near lunch time. He did not realize that the anti-virus application was also scheduled to run during lunch time. The moment the anti-virus application initiated, the application crashed. The server then had to be power recycled. Fortunately, there was no data loss but there was a 15 minute service loss during working hours.

We recently experienced a situation where when the backups were ON, compaction started and the application crashed. While both cases involved a mailing database, enterprises need to realize how important it is to allocate proper resources for smooth operations.

In the second example, we realized that the application server had 10 GB RAM. So where is the problem? We believe improper implementation & integration of applications is a key factor. The application might be able to cater to its own users but it needs much more to perform well. Perhaps, instead of configuring an automatic compaction when the database reaches a threshold, the administrator should have configured scheduled compaction once a week ensuring that it did not coincide with the backup schedule.

In another instance, we found that backups would start very well then slow down over a period of time. They would finally get aborted giving an application error. Again the same argument: The application is running fine and so are the users, so where has the application error come from? We then put the effective timeout parameters to use.

NetWorker configuration files for online backups give us the leisure of configuring the timeout settings while backing up an online database. Effective use of the Inactivity period and parallelism parameters ensures restricted data for backups being pulled out of the application. As a result, both the application and backups run smoothly. Slight performance degradation is possible but then there are no application or backup hang ups.

One of our customers complained that backups were very slow. The infrastructure was good: FC library, SAN based backups, NetWorker backing up on Dedicated Storage node. Though the library has LTO3 tape drives, the backup performance was not more than 15 MB/s. In fact, it averaged out to 8-210 MB/s. The customer wanted to test this on a LAN avoiding the FC path but we thought that it would not resolve the problem. We discussed this with his application administrator and found out that the application replicates the data online. Here was the answer.

Online replication places a heavy load on the application server especially if the data is getting updated very frequently. You can imagine the kind of load it would have when it is catering to 1000 mailbox users with around 400 GB of data with online replication ON. We requested them to stop replication and realized that the backup speed immediately increased and we achieved a speed of 100-120 MBps. For their confirmation, we started replication again and the backup performance slowed.

The memory and processor usage on a file server may not be huge but we still recommend backing them up during off-peak hours. Remember to consider the small overheads caused by Open File Manager when used for a file server.

The general practice of taking weekly full and daily incremental backups might also work well for a file server but does not help for application data or database files. Most of the time, the database would only have a full backup even if you configure it for an incremental/differential backup. This is because the incremental calculation is done on last modified date.

Many people have been asking us to configure their database backups incrementally using the database's transaction logging system. We have not seen a fully successful recovery of a database, especially a mailing database, using transaction logging. You need more disk space on the server to take care of the logs being generated and to ensure that they have been backed up. You truncate them regularly to recycle the disk space on the server.

Software compression is considered a good way to reduce the data flowing across the network, but is it really good? Software compression runs the compression algorithm on the client server, adding more load on CPU cycles. We need to be very careful before choosing our options. If the data to be backed up is not worth compression, like executables etc., software compression may actually increase the size by adding its own load to the non-compressible data files.

Basic Infrastructure Information: Critical

Basic infrastructure information is the most critical component of a successful backup solution's implementation & functioning. The IT administrator and implementation engineer have to be fully aware of their existing infrastructure with respect to data size, the applications installed, the applications and database utilized on each server, the availability of network resources etc.

Organizations have a critical need to implement a process to make the IT staff aware of how their infrastructure is being utilized. IT administrators often are not aware of many aspects such as applications installed on their servers. Often, they are reluctant to allow external service providers to spend time on their servers. You might wonder why this is so important for a backup application implementation.

Here is an example of one situation where everyone in the organization was sure nothing was wrong in their setup. It took approximately 15 days to conclude that their server had a non-required application which was causing the issue.

We implemented NetWorker 7.2 in a pure Windows setup. The customer wanted to utilize the advanced reporting tools available in the NetWorker Management Console (NMC). The backup administrator was also fascinated with the new Graphical User Interface. We installed the Management Console on the backup server.

We were never able to login at the server. Whenever we entered the default username & password, it rejected our credentials. What could go wrong with default credentials? Java implementation was perfect and we checked every possible cause in the administrator guide. NMC works using a Sybase database application at the backend and reserves port 2638 on the management server. We kept asking the customer to check for any application using Sybase as it would have blocked the port required.

The customer had not even heard about Sybase applications being used in his environment. It took us some time to convince him to allow us to run port management utilities to find out where port 2638 was being utilized. The port monitoring utilities suggested that the port was in use even if NMC services are down, but the entire IT staff was very sure that they never used any Sybase based application.

They were reluctant to provide information about the applications in use on that server as they wanted to safeguard their internal information. After repeated requests, we were allowed to scan the server. In the initial check we noticed another monitoring application with one of its components mentioning Sybase database. This tool was installed by one of the junior IT team members without mentioning it to others. He was doing research on a SAN management tool provided by their SAN vendor. He installed it on this server to help him learn more about the tool and forgot to uninstall it. Since this was done only for testing, senior management did not suspect the application was in their setup.

This was a small example where we could have installed our monitoring tool on any other server in the enterprise. Several such instances have suggested that if the right information is provided at the right time, the backup application implementation time could be reduced considerably. Most of the delayed implementations are due to incomplete information or something that no one cared for while planning the implementation.

Training the Implementation Engineer

Partners who are implementing solutions need to be rigorously trained in their data centers before entering an enterprise's production data center. It is important for the Implementation Engineer to have worked on the product for some time apart from reading product guides. We have often seen engineers work on a live setup and make errors that are costly to our customers and to our reputation.

Many enterprises are afraid to implement a solution because almost all of them have some past experience where the Implementation Engineer made an error. They are now reluctant to work with trained engineers.

Sometimes, the problem is in the customer's own setup. Because he knows that the engineer is not fully trained and confident, he may ignore the possibility of his own setup challenges and put it on the engineer to resolve. We need to be very careful when selecting the right person to implement a solution.

Refresher courses should be required for experienced engineers. There is no end to technology. We recently implemented a solution for NetWorker 7.4. Everything was smooth and then we realized that the customer wanted to purchase more client connections for his Virtual Machines. This is not correct, use the NetWorker VMware Consolidated Backup (VCB) feature and you just need a storage node license; you do not need to buy multiple client connections. It not only saves money in terms of license requirements but also does not put any load on the virtual machines for backups. We implemented this at the customer site and helped him save money.

Implementation Engineers need to regularly educate themselves on new enhancements. While we were integrating VCB with NetWorker 7.4, NetWorker 7.4 SP1 was released offering tighter integration with VCB backups. Therefore, it is very important to keep our knowledge current.

Training the Backup Person

The backup administrator needs training to support the configuration and working of backup technologies and applications. Remember the earlier example where we mentioned a server crash since online database, backup application & anti virus application were all running at the same time. His lack of knowledge caused application downtime and could have lead to data loss.

Backup administrators sometimes neglect the basics of technology and expect too much from it. Today, we were working on the Homebase solution at a customer site. We asked him to restore his data from his backup application, NetWorker, he was confused. He was expecting Homebase to take care of his backups as well as provide him with a BMR solution.

We have attended so many ‘critical calls’ where backups could not happen because of application failure. When analyzing the environment, you realize that it was a minor problem. The customer may have forgotten to label the right media for the media pool etc. The backup administrator needs to be trained on all the basic aspects of NetWorker and its behavior in various environments.

The enterprises’ biggest challenge is to retain their employees. This is perhaps one of the biggest challenges facing the IT industry is facing in our region. IT staff wants to learn, not to improve their productivity but to pursue a better job prospect. This makes it important for enterprises to rely on their service providers.

Lengthy training is difficult since backup administrators will spend more time learning and less time working. Capsule course should be designed for backup administrators to deliver foundations training in a couple of days; advanced courses should be left for administrators who show more interest in developing their expertise.

Documentation

A very strict documentation pattern has to be maintained. From planning the activity to the final handover, everything needs to be properly documented. We believe that if a proper plan and a proper document are not created, implementation will never be trouble free.

Get the right information from the customer during the sales cycle, check & re-check all compatibility issues, make a detailed and thorough plan for each client and module to be installed, think over it again: is this correct? If you can, test it in your environment.

The implementation documentation should have a clear procedure for installing new clients and application modules so that no IT administrator loses his valuable assets due to a lack of knowledge about his system. We document the NetWorker resources that have been created. This helps in quick implementation of backup policies and also provides faster support when the customer needs it.

Even if you have implemented NetWorker many times, you may still make mistakes. In fact, your experience should help you plan better and implement more smoothly. You should try to collect network information before stating the implementation and your plans for a customer should include his IP address and host names. Even a beginner should be able to execute a well drafted plan.

The above challenges arise with most enterprise environments. The right fit backup solution would also have these challenges.

Backup & Recovery Window

What is the right balance between the backup and recovery window? While most enterprises prefer to backup their application servers simultaneously, this may not always be wise. For example, every enterprise would prefer to backup all their SQL database servers at the same time and on the same media. As discussed earlier, this may not be a good idea for the ideal recovery window especially when tape media is the backup device used.

While reading the tape, the entire tape length needs to move across the head as it reads sequentially. If the tape backing up a critical database also has some other data, recovery would need that extra time to skip this unwanted length. You should always keep highly critical backups on different media especially if using a tape based backup device.

Sometimes pursuing higher backup windows might be good if they result in lower recovery times. After all, the backups are meant for recoveries.

Another good practice for an enterprise to successfully backup & recover is to regularly monitor its backup policies to see how their operations affect the policies. To ensure smooth recoveries and successful backup strategies, an enterprise should regularly test the restoration of data from the backup devices. Regular restoration drills should be considered by people using tape media for backups. The media has a life span. In many cases when you want to restore data, the media may not be good enough to help you.

Restoration drills are also important as they can be used to document the recovery process for each data type. It would be of great help to the IT administrator to quickly refer to the tested & documented restoration processes when there is a data loss rather then spending those extra minutes figuring out what to do.

Find the Right Fit

A centralized LAN based backup is a good fit for a simple network with multiple servers. Most of the backup applications support heterogeneous environments. It may be a good idea to use a remote device or a number of remote devices if there are a large number of client servers. This would reduce the load on the network bandwidth.

A NAS user could get improved backup performance by using the NDMP protocol to backup his NAS appliance. Some people would use NDMP protocol and backup to their device connected directly to the backup server. This sounds good for the NAS data but better NDMP performance can be achieved if the backup device is directly connected to the NAS box. NDMP also helps you reduce the total cost of ownership since you no longer need to invest in costly application online modules.

Many people purchase a Fibre based library and connect it to the HBA on the backup server. People believe that since it is a Fibre based library, they are taking a SAN based backup. This is not true as the backup data still travels on the LAN. Second, the performance expectation by just plugging in a Fibre based library is highly increased but you also need a fast network for performance. A Fibre library can perform near to its capabilities only if you have a 10/100/1000 network. Having latest tape drives and Fibre library may not always help.

A fibre based library is not SAN based backup. From the hardware perspective, a fibre library is required to be connected to the SAN switch and configured in the proper SAN zones. The backup application then must be configured so that the backup data flows from the SAN disk to the backup device as if the backup device was locally attached to each client server. The device can be dynamically shared among the client servers by configuration.

The choice of backup device is most important regardless of the backup technology– LAN-based, NDMP or SAN-based. Tape uses sequential access to reading and writing data; disk would be a better option for faster backups and recoveries. As a simple strategy, you can backup all your data on disk as the primary backup device, and move the older data on to tape for offsite storage. Media location can be maintained using the backup application. Disks also provide the added protection of RAID technologies so your backups become doubly secure. Some backup software vendors also support simultaneous backup & recovery from the same media while using the disk based backup device.

Simple Implementations that Worked Wonders

Here are a few examples of implementing NetWorker to its optimum use. These are very small enhancements that have enhanced productivity:

1. A customer has a corporate office and 45 branches across the country. The corporate office has a mid size implementation with NDMP backups on tape drives, and 8 Windows servers. The branches have only one server per branch with a standalone tape drive. The server in the branch is a file server and requires daily data backups. The customer does not want to deploy a backup administrator in each branch. However, someone needs to monitor backups daily.

   For the corporate office, there is a backup administrator who is busy logging into each branch to check the status of daily backups. We began simply. We configured automatic emails from each branch. Every morning the backup administrator received mail from each location with a different subject for a successful backup or a failed backup. To add value, we added our support email ID to the failed backup list which automatically sent us an alert. No more logging onto branches was required.

When they needed additional daily reports, we integrated a single NetWorker Management Server in the corporate office from where they could centrally monitor and generate reports for the desired location.

The next challenge was opening a large number of firewall ports. This was a NetWorker 7.1 implementation with a NetWorker Management Console add-on. Large port ranges were required to be opened for communications across the firewall. We worked with their security administrator to configure Windows IPsec so that only three open ports were required for each location.

As a result, the backup administrator was relieved from that responsibility. He was trained and promoted to a storage administrator maintaining NAS & SAN deployments along with backups. The organization saved one man’s salary apart from other resources every month.

2. SGI Irix based installation wanted to backup their online data. The backup device is a StorageTek Silo with 4 LTO-3 drives being used for Hierarchical Storage Manager application that stages data from Fibre disk to SATA disk and then to LTO-3 tapes and two sDLT tape drives to be used for data backup. The data is dmfdata and some archive bits are required to be set and data brought offline for backups. When backups were initiated with any other backup application, they would finish immediately without backing up any data. We implemented NetWorker 7.3.2 in the environment and configured the Silo with two drives. We also configured a dmfdata directive on the client.

   When backups are initiated, the backup server would request the ACSLS software to mount the required tapes for backup. The client side dmfdata directive takes care of switching the data offline-online and takes smooth backups.

Integrating Technologies

We would like to share a few of our robust NetWorker implementations here.

1. We have two NAS appliances, two SAP servers, one exchange server, one Oracle & a couple of file servers in one enterprise. The Exchange and Oracle servers host their data on SAN. The NAS appliance hosts their SAP database and file servers. The enterprise also has a SQL server with small SQL databases for payroll and HR. They already have an HP Fibre Channel (FC) based tape library.

We were asked to design a single solution. Since they already had a tape library, there was no option to suggest a new backup device. We used a combination of various options available with NetWorker and designed smooth backups. We configured the library on the Fibre Channel (FC) switch to be shared across the SAN servers and the backup server through the FC connectivity. The NAS appliance also has Fibre ports so we connected them to the FC switch. A combination of dynamic drive sharing and dedicated storage node has delivered smooth backups.

This is integrated with NDMP backups to take care of the backup of the NAS appliance. A drive is dynamically assigned when we initiate backup of a server on SAN. Using a dedicated storage node licenses only data from this server to get backed up on tape. Similarly when NDMP backup is initiated, a drive is dynamically assigned to the NAS appliance and its data gets backed up on tape. A NAS appliance does not need an additional storage node or dedicated storage node licenses. SQL server is backed up on LAN as it has very small databases.

We are using SAP online agents as the customer needs a tape based copy of backup for the last 28 days. This is a mandatory requirement. Offline backup of redo logs has also been scripted twice a week during off-peak hours.

2. We were shocked when we were unable to take NDMP backups smoothly. Initially, they worked well but they suddenly began failing. The NAS vendor politely suggested that this Library was not compatible for NDMP backups of the specific NAS appliance. They had supported the old library model but they had stopped supporting it with their new operating environment.

Once again, we were left to solve the NDMP backup issue. While we were thinking about how to support backing up the second filer, we learned that their NAS protocols do not support writing from the FC shelf to the SATA shelf within the same box.

The answer was simple. We took one shelf of one NAS appliance as SATA disks and configured it as an advanced file type device. We introduced a Gigabit Network Switch in the environment. Both NAS appliances have a dedicated Gigabit port that connects to the new Network Switch. The second network card of the backup server also connects to the same switch and the NAS backup happens the NDMP way on the dedicated Gigabit network. NDMP data travels across the Gigabit Network switch for backups and recoveries.

3. The environment has approximately 90 servers based on Windows, HP-UX, Solaris & Linux operating systems. The customer wants to back them up daily. We began with simple backups. Since the data size was small, we implemented a 1 TB disk device on the backup server. This was connected to the tape autoloader for regular staging of data to tapes.

The network becomes a challenge when you want to take a FULL backup of these servers. The customer did not agree to a dedicated backup network so we decided on another option. We first categorized the servers and save sets into more critical and less critical options. The more critical ones are backed up on a Weekly FULL basis and the less critical ones are backed up on a monthly FULL basis. Since the number of Weekly FULL backups were less, configuring two groups for Saturday FULL resolved our problems. Incremental backups occur daily.

Then, we configured the monthly FULL backups which were larger in number. We grouped them into four groups. Each group runs once on one of the Sundays of the month i.e. one group runs on first Sunday, the second on second Sunday and so on. Incremental backups also occur on a daily basis. So, with limited resources, we divided the backup groups so that all backups happen smoothly.

4. As it was destined, the above enterprise has grown in the number of servers and data size. The customer has installed EMC Symmetrix® storage for better operations. Most of the critical servers now have their data on EMC Symmetrix. So, why not better backups?

We configured disk based backups for 3 TB of SATA disks from Symmetrix. For smoother backups, each SAN based server was allocated space locally to backup on SAN. The disk backup option working with Dedicated Storage Node provided the right solution. When a backup is triggered, data is backed up locally on the storage node and indexes travel back to the backup server on the network.

It did not end here as we need the backups on tape. We configured a Fibre Channel tape library on the SAN. Dynamic drive sharing allocates the drive for each node, and backup tapes are created for each of them for off-site shipment. Tape utilization is not optimum but the backup & recovery performance is perfect. This is more important than media utilization.

5. We recently implemented a NetWorker solution for an enterprise with mixed requirements. The enterprise uses 5 windows servers and 200 desktops. They have a LTO-2 tape drive and back up their server locally using an entry level backup application that does not support online and NDMP backups. They wanted a backup solution from a single console to take care of their requirements. They also wanted to consolidate their Lotus & SQL databases on storage along with the file servers.

We implemented EMC unified storage NS 20. Lotus & SQL databases are hosted on the NAS portion of the storage backed up using NDMP protocol. Their existing LTO-2 tape drive has been used for NDMP backups. The file servers have been allocated space from the IP SAN form NS 20. One LTO-3 tape library was configured on the backup server. This takes care of their file servers through LAN. File servers have the Open File Agents configured on them. The backup is centrally monitored from the backup server which directs the file server data to the LTO-3 tape library attached to the backup server and NDMP backup to LTO-2 drive attached to the NS 20 storage. To make the solution more cost effective, we are using the NetWorker Business Edition for Windows as the new tape library has only 16 slots and the customer is not expected to add more servers for a few years.

For backing up desktops, we have configured EMC Retrospect on one of the servers with required disk space configured as the backup device. Both the backup applications are monitored through the same Management Console.

New Technologies Make it Possible

EMC Disk Library (EDL) is a good option for those who still want to use tape based backups with the protection and performance of disks. While configuring EDL, you have the choice to configure any number of drives with any tape technology. The disk system would be visible to the OS and the backup application as a tape system. So while the administrator backs up data on virtual tapes, data actually gets protected on disks. Regular copy of this data to the actual media tapes can be done easily for off-site shipment. NetWorker is smoothly integrated with EDL and its other competitive products.

While some people talk of disk failures because of the high input/output required during backups & recoveries, we feel it is still safer than the potential for tape media failure. Tape media failures are more frequent and can happen even when tape in not in use. We completed an implementation a couple of years ago. Several months after the implementation there were heavy rains in the city and everything stood still for a few weeks. When things settled down, our customer started his backups. They complained that every time they initiated a backup request, NetWorker came up with media related errors. The entire set of media had to be discarded since moisture had spoiled it. Disks always have the added advantage of RAID protection.

We were recently faced with a requirement for a backup solution for file servers where the files change a little but new files are regularly created. The customer needs a fast backup application as he has very small backup windows. They also have a few application servers on a different network subnet.

We have integrated a NetWorker 7.4.1 backup software with CLARiiON CX3-10 disk based system attached to the backup server. We are fully utilizing the Avamar integration into NetWorker using a disk based backup device. We created a separate group for clients that need data de-duplication features, the disk system has been configured as an advanced file type device and parallelism if set to 4. This helps smooth the file server backup. The remaining application servers have been grouped separately. We configured a storage node for them since they belong to a different subnet.

Snapshots and replication are another set of new technologies in the storage industry. Most people consider them backup. However, they cannot always substitute for traditional backups. Snapshots and replication always need exactly the same amount of disk space as the original data until you rely on pointer based snapshots. Pointer based snapshots are good only until the time you have the original data intact.

The destination would have the data in the same format as the source so no encryption. In most cases, neither of the two offers you reliable version-based restores. One big challenge of online replication is that if the source data gets corrupted, the destination gets corrupted before you realize it. Online replication also needs network bandwidth for the replication process.

While emerging technologies are arriving, you cannot replace the traditional ways of backing up your data on tapes. You need them for an off-site secure copy of your critical data. The disk backup option and the EMC Disk Library are good primary backup devices and the tapes for off-site shipment.

What to Know Before Designing the Solution

Before designing a backup solution, we must check on many parameters. These may not be relevant in drafting the Bill of Material or working out discounts to win the deal, but they are perhaps more important than selling a solution.

You should ask the following questions and find out the relevant information before proposing a solution:

1. *How many servers need to be backed up?*

You should know the number of servers that currently require backup and determine future plans for adding more servers.

2. *How many of these have online applications and which ones?*

This forms the backbone of the solution. Online applications and databases need to be well protected.

3. *Could the customer afford application downtime?*

This is perhaps the most critical question. Application downtime is something that no one can afford, but this question forms the basis of all commercial and technical evaluations of the solution. Thanks to the modular licensing of NetWorker, you can buy a module as and when required and do not need to invest all your money in the beginning.

4. *Which servers are more critical?*

This helps to prioritize the servers for backups. This also helps you to design the backup policy as to which servers need more attention and perhaps always FULL backup policy for faster recoveries.

5. *What is the actual data size from each of the above?*

You should know the current actual size of data being backed up. You should add approximately 10-15% to this value as it is very difficult for the customer to give you the exact GB size.

6. *What data size belongs to the more critical servers?*

This is important for the backup groups, media pools and schedules. How much critical data is generated every day? The critical server would start first. If you know the size and expected performance, you can schedule the other servers subsequently.

7. What is the expected growth in each of the above?

No backup solution is good if it cannot accommodate growth for a couple of years. The exception, of course, is if the enterprise grows exponentially.

8. How can we group them on the basis of their applications or network subnets?

As discussed earlier, if they belong to a different network subnet, give them a storage node for the subnet. If they are all same application or multiple databases from the same server, give them the same media pool.

9. What is the desired backup window?

Defining the backup window is important to decide the backup device. Of course, the most ideal would be a disk based primary device and a tape based secondary.

10. What is the desired recovery window?

Recovery window also helps decide the backup device. Based on the desired backup & recovery windows, you can decide the backup device and its connectivity like storage node, dedicated storage node etc.

11. Does the customer have a backup policy in mind?

The customer should know the policy for his backups. If his organization has not set one, you need to draft a policy with him.

Key factors to keep in mind while implementing a NetWorker backup solution:

1. Backup Device:
    a. If using disk based backups,
        i. Configure advanced file type device to get random and simultaneous read/write access.
        ii. Configure disks with RAID 3 or 5 and configure your devices one for each LUN.
    b. If using EMC Disk Library,
        i. Use optimum number of drives. A configuration of 4-6 drives should give good performance.
        ii. Configure slots starting with 5 GB capacity expandable to maximum value. It helps to optimize the usage of the disk space.
    c. If using tape drives,
        i. Create a separate media pool for databases. Prefer a separate pool for the most critical database or databases that need to be restored simultaneously.
        ii. Tape drives normally give good performance on Target session of 4-8. Test the performance in your setup before going beyond this.

2. Configure a separate media pool for Index & bootstrap backups. This helps in faster disaster recovery as you need to look for only one media to locate the latest bootstrap. For additional protection, clone these tapes regularly.

3. For faster disaster recovery, make a note of bootstrap Saveset ID on a daily basis. You don't need to spend time on scanner to locate this for you.

4. Configure your NICs for the backup network on Full Duplex or Half Duplex and 100 or 1000 MBps depending on the NIC used. Do not use Full/Half Duplex and 10/100/1000 Mbps type options. Many network issues are known to be resolved by this.

5. Make effective use of the comments field in each NetWorker resource.

6. Use the Server Network Interface attribute of the client resource for specifying the server's network address.

7. Keep similar nomenclature for a set of resources. Ex. If you want to name a Group as "SQL", name its media pool also as "SQL". It makes it easier to correlate while troubleshooting.

8. Use parameters like "Inactivity Timeout" effectively. For Lotus online backups, the configuration file has parameters for Connection timeout etc. If used effectively, they help streamline the backup performance.

9. Make a record of all ClientIDs for each client in the setup. They are more important than the client names themselves.

10. Keep optimum values for Client & save Group parallelism. Test them in your setup before using them.

11. Fine tune the Oracle RMAN backup channels for online Oracle backups. Normally 8 channels give a good writing speed. The speed goes down if more than the required number of channels is defined.

12. Divide save sets into multiple save sets. Ex: Instead of backing up D:\, backup: D:\1, D:\2 etc. A larger number of savesets produces more save streams for faster backups.

13. For security reasons, it is a good practice to type the correct server name in the client's \nsr\res\servers file rather than leaving it blank. Similarly, *.* should be used in the remote access field for a client resource carefully.

14. Configure online Lotus and Exchange backups as FULL backups. Use their transaction logs carefully if required. Sometimes, it is difficult to get a consistent database if transaction logs are not applied properly.

15. Regularly perform recovery drills to test the recovery of your critical data. Document the processes for others in the enterprise.

16. Effectively use technologies like cloning, staging, dedicated backup network, NDMP & SAN based backups, VCB backups, data de-duplication etc. These help to achieve a stable and successful backup setup.