

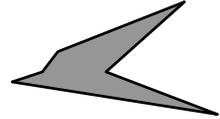


Infotech SA Inc. White Paper



Data Vaulting

May 2002



DATA VAULTING

Review of Data Protection Technologies and Options

1.0 Introduction

In light of recent events, the protection of data in Federal Computer Installations is a particularly important and sensitive issue. Suppose that a large portion of Agency's computer resources and data reside at a single Computer Center. *In the event of a catastrophic disaster at this Center, the Agency is at very serious risk of irrecoverable data loss. It is therefore of the utmost importance to consider strategies for providing more complete protection for the Agency's data.*

Let us consider a situation where an Agency, A, has two data centers, several miles apart, and roughly equal computing resources are housed at each center. An ATM link is available between the two centers. The existing Enterprise Backup System (EBS) backs up data at both locations and stores it into silo libraries at the respective location.

Several of the data protection strategies discussed here involve upgrades to the existing Enterprise Backup System and the inter-site network ATM link. This paper describes these higher data protection levels, their advantages and disadvantages and their likely costs.

Data protection concepts, which are based on local and remote copies of backup data, are discussed first. Their advantages, disadvantages and limitations are considered. A description of the various data protection levels and the degree of security offered by each is given next. The Appendix describes fibre channel inter-connect technology.

2.0 Data Protection Levels

Four data protection levels are defined, with each succeeding level offering more protection. The Enterprise Backup System provides the first level of protection today. Two additional levels (Levels 2 and 4) can be provided at a later date when the network is upgraded to Fibre and one of the levels (Level 3) could be provided with additional tapes, trucking services and personnel.

Level 1 – This is a single tape backup copy of data kept at the same location as the server (i.e. our current level of data protection).

Level 2 – This is a single tape backup copy of data, but it is kept at a remote location. For Suitland, backups would be stored in Bowie; for Bowie, backups would be stored in Suitland. Remote interconnection technologies described later would have to be in place for Level 2 protection.

Level 3 – There are two tape copies of the backup data. The first copy is kept at the same location as the server. The second tape copy is manually moved to a remote location. Suitland copies are moved to the Building 4 vault. Bowie copies are trucked to Suitland to the Building 4 vault. This could begin now if we had additional tapes, a trucking company and personnel. This method also needs more processes to insure against human error.

Level 4 – There are two tape copies of the backup data. The first copy is a tape backup copy kept at the same location as the server. The second copy is a tape backup copy kept at a remote location. The second copy is made electronically over an inter-site link. Remote interconnection technologies described later would have to be in place for Level 4 protection

In each of the above protection levels, a "data verify" step can be added optionally. Verification ensures that the copies of the data are readable.

Level	Protection Type	Verifi- cation	Method	Available
1	1 Tape Copy kept locally	YES / NO	Automated	Now
2	1 Tape Copy kept remotely (stored at remote site using new inter-site link)	YES / NO	Automated	Pending new inter-site link
3	1 Tape Copy kept locally + 1 Tape Copy kept remotely (moved daily to remote site)	YES / NO	Automated / Manual	Could provide now, but need more resources
4	1 Tape Copy	YES /	Automated	Pending

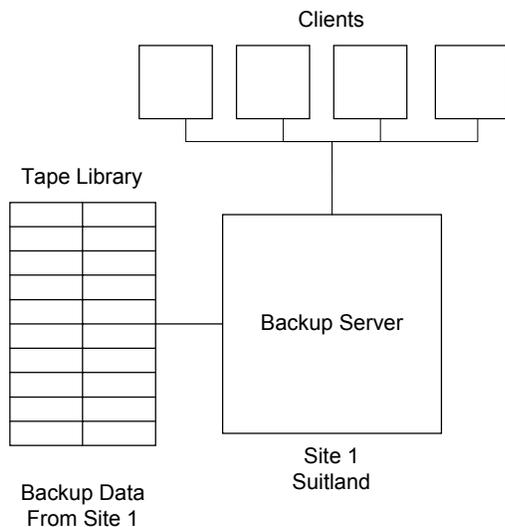
	kept locally + 1 Tape Copy remotely (stored at remote site using new inter-site link)	NO		new inter-site link
--	---	----	--	---------------------

offered this level of protection. Level 2 Data Protection for all backup clients requires a new high-speed link between Suitland and Bowie.

Table 1 – Data Protection Levels

3.0 Approaches to Implementing Data Protection Levels

This Section discusses the methods via which the BOC’s Enterprise Backup System (EBS) may implement the data protection levels described in the preceding Section. These methods, their costs, advantages and disadvantages are described



below.

3.1 Level 1 Data Protection (Backup Copy of data)

Level 1, a local backup copy of the data, is already done by the Enterprise Backup System and does not need further discussion in this paper.

Level 1 Data Protection

Similarly for Bowie.

3.2 Level 2 Data Protection (Remote Backup Copy)

In Level 2 Data Protection, a single backup copy of the data is stored remotely. At the present time, because of the limited network bandwidth between Suitland and Bowie, very small amounts of data are

Level 2 Data Protection – Remote Copies

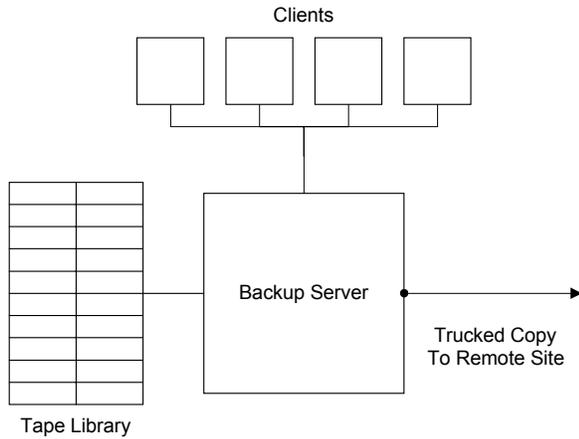
The main advantages of this method are that the backup copy is kept at a remote location from the server and the additional costs associated with keeping 2 tape copies of the data are avoided (i.e. tape drives, silos, and media). This would require rearranging our current tape silos and drives. More silos and drives would be needed in Suitland to handle the Bowie workload. The main disadvantage is there is only 1 backup copy of the data. If this copy is destroyed, all that remains is the current data on the server. This method would also require extensive testing once the fibre link is established.

3.3 Level 3 Data Protection (Local Tape Copy + Remote Tape Copy stored manually)

In Level 3 Data Protection, there is one local tape copy and one remote tape copy. The remote copy is stored manually (i.e. by trucking).

Media, trucking and personnel costs are the extra expenditures required for Level 3 data protection. It is possible that at a later date, depending on growth in workload, more tape drives would be required.

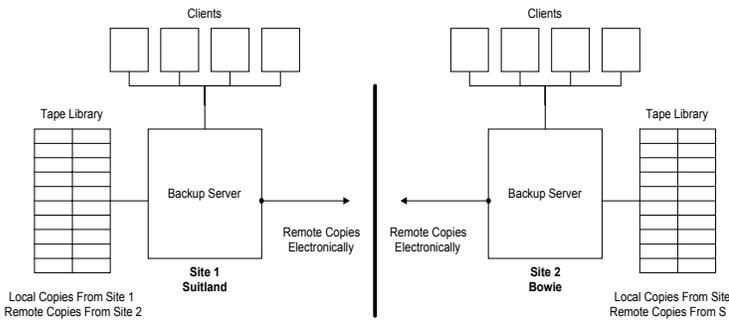
The main disadvantage of this method concerns security and scalability. The manual tape transportation and administrative problems implicit in this method leave much scope for error. These problems are likely to become overwhelming, especially as the volume of backup data at the BOC continues to grow. The security requirements for some of the backup data may also make this method unacceptable. This method would need new verification procedures to be sure no human error had occurred.



Level 3 Data Protection

3.4 Level 4 Data Protection (Local Tape Copy + Remote Tape Copy stored electronically)

In Level 4 Data Protection there is a local backup tape copy and there is also a second, remote backup tape copy. The remote backup tape copy is stored electronically. Level 4 Data Protection for all clients requires a new Suitland-Bowie link. A diagram for Level 4 Data Protection is shown below. It would also require additional silos in Suitland.



Level 4 Data Protection – Local and Remote Copies

The main advantages are two backup copies of the data exist and one is electronically stored at a remote location. Errors associated with manual handling of the remote second copy are eliminated. The main disadvantage of this (highest) level of protection is its initial cost.

4.0 Summary

The following tables summarize the main advantages, disadvantages and costs for each of the Data Protection levels described above. Note that the costs stated are costs for the extra connectivity hardware and include the implicit costs for additional media and tape libraries.

Data Protection Level	Vaulting Approach	Main Advantages	Main Disadvantages	Comments
1	Local Backup Copy	N/A	N/A	Covered in the current system except for media
2	Remote Backup Copy (electronic)	Provides security in case site housing the client machines is destroyed	Only provides 1 copy of the data. Requires high-speed inter-site link to be operationally viable. Will need testing and re-arranging silos and drives. More capacity needed in Suitland.	Currently implemented on a very limited basis due to bandwidth constraints
3	Local Tape Backup Copy + Remote Tape Backup Copy stored manually	Does not require high-speed inter-site link. Provides security above and beyond that of Level 2 in case site housing the client machine is destroyed.	Security concerns because of manual handling. Trucking and personnel costs are needed. Human error introduced when handling large amounts of tapes daily.	Currently implemented on a limited basis. Would require additional procedures to protect against human error
4	Local Tape Backup Copy + Remote Tape Backup Copy stored electronically	Provides security above and beyond that of Level 2 in case site housing the client machines is destroyed. Non-manual, much more reliable than other levels	Requires high-speed inter-site link to be operationally viable. Will need testing.	Currently not implemented

Table 2 - Summary of Data Protection Levels

Protection Level	Method	Cost Considerations - excluding media, future library capacity and drives)	Cost
1	Local Copy made in library	N/A	N/A
2	Remote Backup Copy	DWDM Boxes and interfaces (\$300K), moving silos (\$100K)	\$400K
3	Local Copy + Manual Remote Copy (Truck)	Media (\$325K), 8 Vault Drives (\$200K), Library (\$75K), Trucking Costs, and additional personnel costs	\$600K + Trucking Costs (TBD) + additional personnel costs
4	Remote Copy and Local Copy	Media (\$325K), 8 Vault drives (\$200K), Library (\$75K), DWDM Boxes and interfaces (\$300K),	\$900K

Table 3 - Summary of Costs for Data Protection Levels

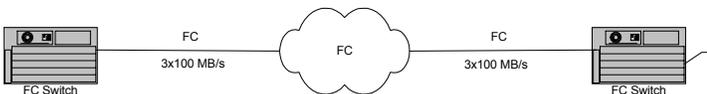
Note: In Level 3, the Trucking Cost is an annual recurring expense.

Appendix

This Appendix describes fibre channel technology, which is a prerequisite for both Level 2 and Level 4 Data Protection as defined in this paper.

A.1 Dark Fibre

In this option, data is sent to tape drives at the other site using Fibre Channel (FC).



The most important advantage of this method is performance. Since tape drives at the remote sites are attached directly to the backup servers via FC, we expect to easily achieve 70 MB/s.

The disadvantage of this method is that today, there is no dark fibre between the two Census sites! However, three such lines, offering a combined

throughput of 300 MB/sec are due to become available by the beginning of March 2002.

If one of these three lines could be devoted to the Enterprise Backup System, this would provide a bandwidth of 100 MB/sec for backup.

The actual throughput should be 2/3 of the maximum capacity. Given a 10-hour window, 2.2 TB of data could be moved between the two sites. This is not enough to satisfy our current workload of between 3 – 4 TB a night and any future needs.

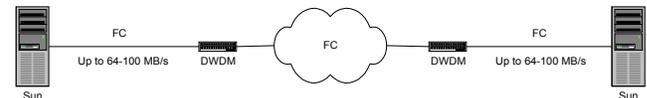
The Backup Project could lease its own dark fibre line at a cost of about \$3,500/month to provide additional capacity. However, it would need to be determined if the additional lines are available.

A.2 Multiplexed Dark Fibre

In this option, data is sent to tape drives at the other site using a multiplexed fibre line. A single fibre line would be multiplexed thereby giving it a capacity of N fibre lines all on the same single fibre line.

Multiplexed Dark Fibre

The most important advantage of this method is high performance. Since tape drives at the remote sites are attached directly to the backup servers via



FC, there is available the equivalent of several individual 100 MB/sec connections. A very high bandwidth, up to 6.4 GB/s (depending on multiplexor configuration) should be available using a single dark fibre line. The only additional hardware over the previous option is the multiplexor box at each site. Even under the 2/3 assumption, we would get well over 4 GB/s

No dark fibre line between Suitland and Bowie exists today, but it is thought that one of the planned three lines could be available to the Enterprise Backup System. The only other disadvantage is the initial cost of the multiplexor.

Assuming conservatively, that the equivalent of 10 FC lines becomes available through such multiplexing, and given a 10-hr backup period, a total of 24 TB per backup period could be moved. This could satisfy the current workload plus it allows room for growth.