Paper #2065

The Case for Networked, Centralized Storage

Carmen J. Marchionni

Open Systems Solutions, Inc.(OSSI)

1 Lakeview Drive Newtown, PA 18940

(215) 579-8111

<u>Summary</u>

Remember back in the late 70's when Hewlett-Packard was touting the benefits of Distributed Data Processing and Decentralization of MIS? The message was clear…move the computing power down near those who were using it…*empower* the User!

Of course back then when you talked about decentralization of MIS you talked about moving the sophisticated and expensive computing resources away from *corporate* MIS to *departmental* MIS…but in each case data processing *professionals* were managing those resources with standardized practices and procedures for data storage and data management. The size, cost and complexity of the computing equipment dictated that only specially trained computer experts could be entrusted with it's 'care and feeding'.

As times changed and computing equipment became less costly and more friendly, the User became more computer literate, and application software for almost any application became 'shrink-wrapped' and available at your local computer 'supermarket', dumb terminals connected to the host were replaced by personal workstations connected to the network…the day of real User empowerment has arrived!

If information is power and decisions grounded in data are better ones then surely today's business professionals who are truly 'connected' and 'on-line' are making quicker and better decisions for their company's.

So what's wrong with that?!

One small glitch…these financial, engineering, marketing and other non-data processing professionals were never trained on proper back-up procedures and software/hardware compatibility issues and never thought to consider how their one (or two) small purchases could impact the total capital budget for computing equipment for the company.

So now those same companies who have benefited (or think they did) from the empowerment of their user community are faced with the real fear that the very data so vital to their company's success and profitability may be at jeopardy from theft because it's not being controlled in any way or by loss due to hardware failure or lack of proper back-up procedures. In addition, the sum of all those small purchases by each User is adding up to a number that nobody thinks makes sense.

The pendulum may have swung too far the other way!
<u>Distributed Data Processing</u>

What was the impetus for changing the ground rules by which the company's computing resources were managed…and controlled? Was it the arrogance of Corporate MIS in knowing that they had something you needed and only they could deliver it because it was such complex, intimidating equipment? Surely if you are over 40 and a User you can remember 'begging' for urgent help only to find yourself at the mercy of someone from Corporate MIS who has never read In Search of Excellence and doesn't know how to spell Customer Satisfaction. You had no choice. You couldn't do it yourself. Computer literacy was binary…you either were literate and a DP professional *or* were not literate and a User.

Was it the corporate "tax" (cost allocation) that each department had to pay without any opportunity for negotiation or challenge? Marketing and other departments were forced to cancel important programs due to lack of budget because Corporate MIS reported to the CFO, or directly to the President, and there was no arguing about it. Department management was being measured on the profitability of their department but were not totally in control of the spending of their budget. General managers usually did not take kindly to this.

So, it was no wonder that when vendors like HP started marketing their Distributed Data Processing systems, they spent more time with the User community as they did with MIS. In fact, in many companies HP was shunned by MIS because they were siding with 'the enemy' and putting evil thoughts into there minds. Thoughts like freedom from dictatorship and control of their own destiny. Thoughts which appealed very strongly to General Mangers who felt hand-cuffed with the current centralized MIS function.

Distributed Data Processing took off and hardware and software developers designed products to accommodate this new market trend. Major applications like Materials Management and Production Control were now able to be run and managed by the user department who owned the materials and who produced the products. Department MIS staffs grew as Corporate MIS staffs dwindled. The Users were empowered and in control and Corporate MIS was relegated to a 'liaison' role for these applications. The arrogance had shifted along with the power and control.

Distributed Data Processing II aka Networked Client/Server Computing

As dumb terminals became intelligent terminals and then PC's and now workstations and networks became flexible and pervasive, the ability for almost everyone in the company to be 'connected' to everyone else and most importantly to all the data became a reality…and not just for large corporations. Small and medium-sized companies also saw the value and the light…and the 'Net' and got on the 'on-line' bandwagon.

Networks were installed everywhere and Corporate MIS, with arrogance in check, became an important player in helping to establish some intelligence to the selection and implementation of the network architecture. Network Administrator specialists came into being and provided the much needed focus there as the 'network became the computer'.

Networked Computing replaced the outdated Distributed Data Processing terminology and Client/Server further defined Networked Computing with implications of specialized resources located on the network serving all those who were signed up to get access to it.

"Open" systems standards were well developed and UNIX, even with it's many proprietary variants dominated Client/Server for awhile, only to see a challenge from Windows/NT as that O/S becomes more comprehensive with features like clustering and the equipment that birthed it becomes more and more powerful. Netware sites are making conversion plans to NT at a brisk pace and UNIX shops are preparing for a peaceful coexistence of a mixed UNIX/NT environment.

As this huge increase in data, and sometimes information, helps companies to make better and quicker decisions, the question of how and where to store, manage and protect this data becomes a key strategic issue.

The Effect on Storage and Storage Management

Before the Distributed Data Processing era data storage was simple. It resided on the mainframe and was managed by Corporate MIS in a very controlled fashion. You told MIS what you needed access to and if you were authorized to get it you did…sometime well after you requested it.

As functional and departmental DP equipment moved out of the Corporate MIS glass room and into the department MIS facility, the data was much more readily available but still centrally controlled at the department level. Uploads to Corporate MIS were routinely performed and the company's data was tightly managed.

As Client/Server networked computing became pervasive and PC's were placed on everyone's desk to give them more data for more informed decision making, it was easy for individuals to make computing resource purchases for their workstations that they felt would improve that resource. Adding a disk drive or upgrading one was not only cheap…it was pretty easy to do…so a lot of people did it…and with little regard to any guidelines on compatibility, serviceability, performance, etc. Disk was a commodity and you buy commodities on price…only!

The joy of Division GM's seeing their employees at the leading edge of technology, making proactive decisions ahead of the competition quickly turned to sorrow when one of the lowest price commodity disk drives, which held months worth of key information, crashed sent the whole Division back to GO. Why not simply restore the data from tape you say? Good idea if it had been backed-up at all or backed-up properly. Those details about the equipment were not paid much attention to.

This story and variants of it are the downside of the networked, decentralized computing resource explosion. The benefits of having the data are immense but the chaos it causes in figuring out how to manage and protect it requires a new paradigm.

The Data about the Data

Let's find out more about data storage. First let's define Decentralized and Centralized in their new contemporary terms:

Classification	Location of Servers	Administrative Control	Purchasing Authority
De-Centralized	Distributed	Distributed	Facility
Centralized (new definition)	Distributed	Central	Site
Centralized (old definition)	Centralized	Central	Central

Now for the data:

* The annual cost of managing storage is $7.00/MB in a De-Centralized environment and $3.50/MB in a Centralized environment.

* Disk & File Management represent 45% of the cost of managing storage and Back-up & Restore represent 22%.

* Storage failures represent 55% of all Server failures.

* The cost of downtime for a "medium" site, with >100 but <1000 workstations, is $39K per hour.

* IBM estimates that distributed storage costs 7 times the cost of the storage component for storage management.

"We're looking for …really robust operations that serve decentralized systems with decentralized management." ------Steve Davis, Rohm & Haas Co.

In Search of Solutions

Now that we know the problem how do we solve it?

Well, like all tough problems the solution is not always easy nor straight-forward. One approach is to follow a paradigm shift from distributed, host-attached storage to "pooled", network-attached storage.

FROM	TO
Distributed, isolated pockets of storage	Pooled, centralized storage
Administering Servers	Administering network resources
Storage attached to Sever & Server attached to network	Network storage resources available directly to clients
Storage performance limited by Server bandwidth	Storage performance limited only by the network

General Purpose File Servers will not cut the mustard here. Dedicated, specialized Storage Servers (Symbios Logic/MetaStor, Auspex) must be used. Among the many benefits of dedicated Storage Servers over File Servers are:

* Storage Performance

* Back-up / Recovery Performance

* Increased data Availability to the network

* Simple, Centralized Administration

* Expandability

* Connectivity

Centralized management of distributed Storage Servers is key from a performance and data avilability aspect but combining this with Hierarchical Storage Management (HSM) is the ultimate in Centralized storage management. HSM implemented as part of each Storage Server provides the most cost-effective means of storing the data and provides a means for Disaster Recovery in the event of a major disaster such as flood, earthquake, sabotage, etc.

HSM integrates an archiving function to make multiple copies of the data to various storage media and migrates the least used data from expensive hard disk media to less expensive optical and tape yet ensuring quick access to the older data.

Conclusion

Storage Management is a key, strategic issue which must get the attention it deserves. If ever the adage " you get what you pay for" ever held true it holds true with storage management. Time and energy spent up-front in thinking through a cogent storage management strategy could be the competitive

differentiation that you need. You already believe in empowerment of the User and that data is power. Why not put manage that precious resource more carefully?