

# The State of Resilience and Optimization on IBM Power Systems

Research Findings Based on Surveys of IBM i and AIX Users

POWER SYSTEMS 20TH ANNIVERSARY



## Contents

Executive Summary	1
Foreword: Power Systems, 20 Years On	2
Report Focus	3
Research Environment and Methodology	4
The State of Disaster Recovery Expectations	5
Service Level Agreements	5
Recovery Time Objectives	6
Recovery Point Objectives	7
DR Plan Completeness	8
The State of Disaster Recovery Technologies	8
Data Protection Methods	9
Data Replication Technologies	11
Continuous Data Protection	13
The State of Systems Optimization	14
System Management Capabilities	15
Automation of Optimization	16
Optimization Tools and Methods	16
Optimization Time Requirements	17
Optimization of Large Files and Objects	18
Summary: The State of Resilience and Optimization	19

## Executive Summary

The AS/400 and RS/6000 platforms have evolved greatly over 20 years, and now IBM has brought them together to create the Power Systems server line. The technologies and methods for data protection and system optimization for both platforms have seen great changes as well. Even as transaction and data storage volumes have increased, ensuring that the right data protection and management practices are in place has moved up from a line-level responsibility to a core boardroom mandate.

### Central Issues for Executives

- Ensuring that the technologies and processes currently in place will achieve the recovery time objectives (RTOs) and recovery point objectives (RPOs) of the organization is imperative. A large percentage of firms surveyed do not have appropriate technologies in place to achieve their stated goals.
- Data protection and High Availability are related but still distinctly different objectives. In general, companies surveyed show increasing adoption of solutions to address both areas, but keeping data and applications continuously available has not been addressed as fully or consistently as protecting against catastrophic data loss.
- The older focus on Disaster Recovery (recovering from fire, flood, earthquake, or other natural catastrophe) is yielding to a broader understanding of protection against all causes of business interruption. Technologies for recovering from and actively avoiding downtime caused by wide-area power outages, loss of telecomm services, and even accidental damage or deletion of individual files are maturing and gaining support.
- Data quality—in terms of its currency, accuracy, and relevance to current business, as well as data access efficiency and productivity—is another issue that is moving front and center for executives and operational managers in all areas of the company. As data stores grow, so do the costs of the labor and downtime resulting from data maintenance activities. This area is ripe for automation, and a new range of technologies is emerging to fill the need.

## Next Steps

- Evaluate the costs associated with downtime of any duration from any cause.
- Budget for and implement technologies and processes to address the issues of data quality and availability based not just on the avoidance of potential downtime costs but also on the measureable savings that accrue from operational efficiencies and flexibility that automatically result from those solutions.
- Where technologies are in place, ensure that the company's actual ability to recover from downtime or data loss events is in fact sufficient to meet stated Service-Level Agreements (SLAs) and internal expectations. Live testing of systems and procedures is necessary, for both the training and readiness of personnel and the evaluation of current capabilities versus present and future needs.

## Foreword: The AS/400, 20 Years On

“The value of data will be treated as an asset on the balance sheet and reported by the Chief Financial Officer while the quality of data will become a technical reporting metric and key IT performance indicator.”

IBM Data Governance Council, 7 July 2008

There is little doubt that data is increasingly viewed as perhaps the single most valuable asset in any business. Quantifying that value monetarily, as suggested by IBM’s Data Governance Council, may indeed be the next step in international accounting standards.

But even without putting an exact number on it, working business leaders understand intuitively the value of their data and know that they must daily address the practical, tangible issues of protecting their business information from loss or corruption and putting it to work effectively, efficiently, and profitably.

Since its inception, IBM® has, of course, been at the forefront of responding to such core business requirements. And while its mainframe computers are legendary, it is clear that IBM’s introduction of personal computers and midrange servers in the 1980s had the greatest impact for the majority of businesses, the so-called mid-market or “small and medium businesses” (SMBs). The introduction of the AS/400® server in 1988 opened up a new realm of possibilities for many such companies. Powerful yet affordable and much less complex to program and manage than a mainframe system, the AS/400 enabled SMBs to compete with larger companies and to operate far more efficiently, forever raising the bar for what it took to compete in any marketplace.

In the twenty remarkable years since, the sophistication of the platform and the range of business functions depending upon it have grown exponentially. Through continuous development, the AS/400 evolved into the iSeries® and then the System i™ and became capable of running not only its native OS, but also IBM’s AIX® and even Linux®, becoming over time a core computing platform for hundreds of thousands of businesses globally.

Today, IBM has merged the System i server line with its AIX-based System p™ counterpart, creating the Power™ Systems server. Nevertheless, the soul of the AS/400 lives on in the IBM i operating system. And the core mission of the combined IBM i and AIX platform remains the same, with businesses of all sizes around the world depending on their IBM Power Systems servers and the data they process and store.

## Report Focus

As part of its core mission, the Information Availability Institute (IAI) regularly surveys the IBM i and AIX customer base, seeking to understand current and developing information availability and data protection needs. And clearly, the range and sophistication of the tools and technologies available to business of all sizes to address these needs have grown right along with the market.

So, at this twentieth anniversary of the introduction of the AS/400, considering our focus on information availability, we felt it would be appropriate to undertake a review of the data protection and recovery technologies and strategies that IBM i and AIX customers have adopted. We also looked into their related database and systems optimization practices.

## Research Environment and Methodology

The data set used for this report includes the results of over a dozen surveys, conducted between July 2007 and July 2008. In total, responses were received from over 2,000 technical professionals and executives involved in the management of IBM i and AIX environments.

Throughout this report, we will identify responses as pertaining to users of the IBM i and AIX operating systems. As a practical matter, under these “umbrella” identifiers, we include all prior variants of the IBM i (i5/OS, OS/400) and do not differentiate between, for example, AIX running on a legacy System i or System p server.

## The State of Disaster Recovery Expectations

To begin our study, we asked four questions about our survey participants' perceptions of their organizations' current readiness to handle business interruptions (planned and unplanned downtime events).

### 1. Are Service-Level Agreements (SLAs) important to your organization?

For this question, the available responses were: "Yes, we have SLAs in place today" and "No, we don't use SLAs at this time."

Sixty-nine percent of respondents indicated that SLAs were indeed a part of the environment (and performance pressures) with which they must contend. Though business contract terms specifying performance criteria were not uncommon twenty years ago, Service-Level Agreements were not a common IT issue. In today's on-demand, right-now economy, even smaller businesses are heavily dependent upon e-mail and Internet services to conduct business. Access to IT services has become an imperative. There is no reason to expect the prevalence of SLAs to do anything but grow.



Figure 1

## 2. What is your organization's recovery time objective (RTO) after a disaster or complete server or application failure?

With this question, we expected and in fact observed a significant bias toward shorter time frames, which is certainly in line with ever-increasing SLA pressures. Overall, 45 percent said recovery within 6 hours was the mark, while 78 percent indicated that, at a minimum, a 24-hour (intra-day) recovery capability was required. (Figure 2)

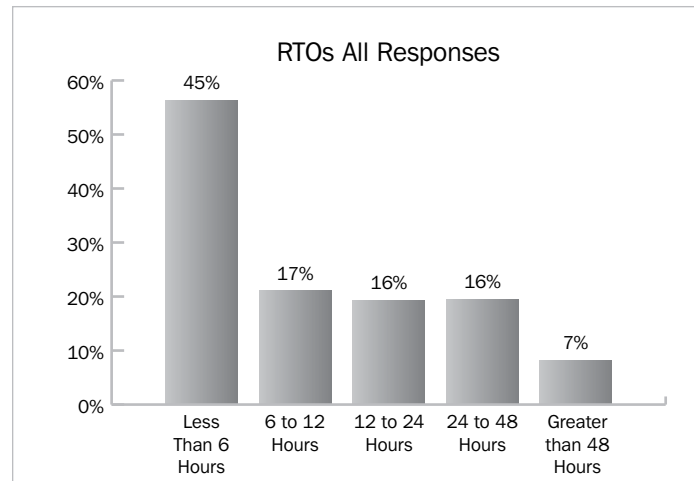


Figure 2

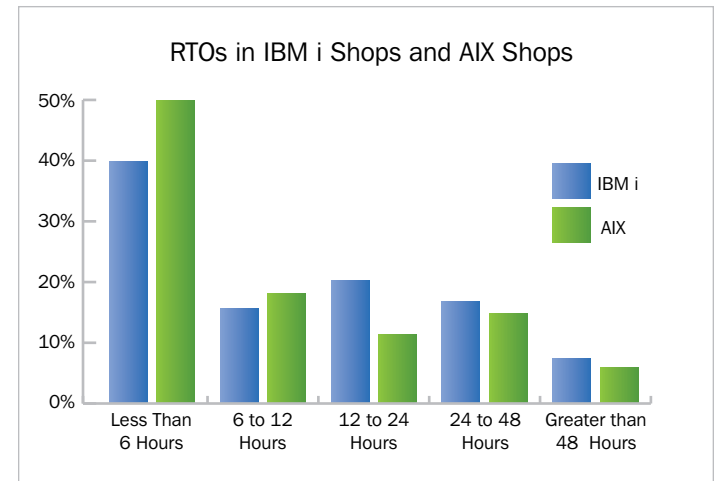


Figure 3

Within these results, comparing the responses from the IBM i shops to those from the AIX shops, there was a notably stronger bias in the AIX responses toward shorter, more aggressive recovery time requirements. (Figure 3) For example, 68 percent of AIX respondents cited less than 12 hours as the objective, compared to 56 percent in IBM i shops.

**3. What is your organization’s recovery point objective (RPO), expressed in time? That is, what is the maximum volume of transactions you are willing to lose as a result of a disaster or a complete server or application failure?**

Compared to the results for the recovery time question, the responses for this question were somewhat more evenly spread across the provided range. But data loss still appears to be of greater concern than is the time to resume operations. Most notably, while 55 percent of respondents indicated that they could accept business downtime of six hours or more after a server or application failure, a full 59 percent indicated that they were unwilling to lose more than a few minutes worth of transactions or operational data.

But there remains a noticeable contingent of respondents whose organizations are apparently not especially worried about losing many hours or even a day’s worth of online data in the event of a catastrophic server failure. This may reflect a subset of businesses that are still comfortable with relying upon manual re-entry of transactions from paper records.

For this question, the responses from IBM i users were roughly similar to those of AIX users, with differences of 3 percent or less in most categories.

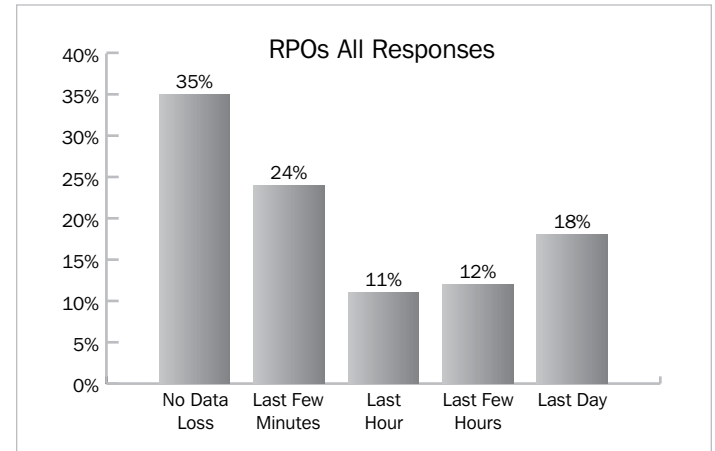


Figure 4

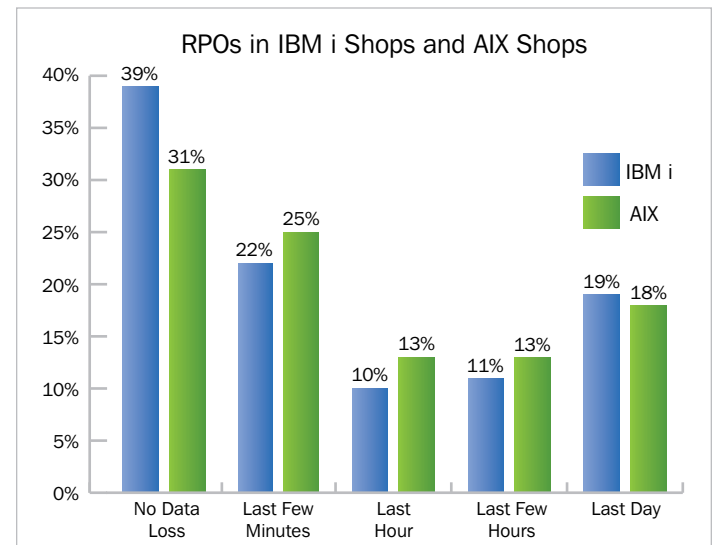


Figure 5

#### 4. How do you rate the completeness of your organization's disaster recovery plan for IT systems, including testing of the DR plan?

This question was asked specifically to gauge the respondents' self-confidence in their DR readiness.

For this question, we have responses from companies that use IBM i or AIX, as well as companies that use both operating systems.

Only about 10–15 percent of the respondents had full confidence (100 percent) that their DR plan was complete, tested, and ready to go. About 50 percent of all respondents expressed at least some doubt or concern, as they rated their confidence at 75 percent to 90 percent. But overall, the distribution of the responses between AIX and IBM i shops was very similar.

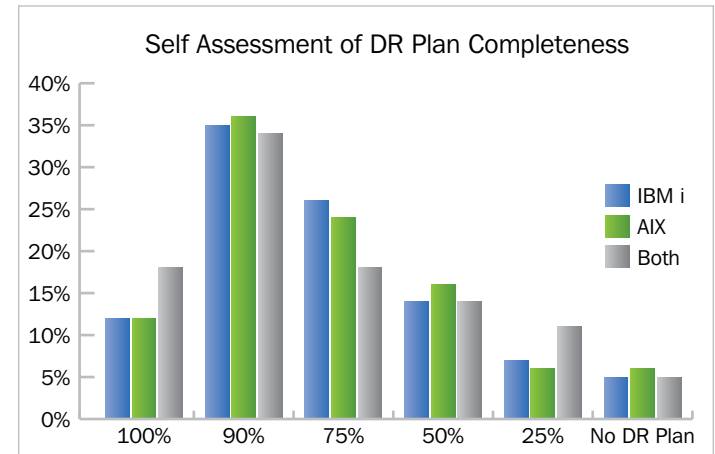


Figure 6

## The State of Disaster Recovery Technologies

When we discuss High Availability (HA) and Disaster Recovery (DR), we are really addressing a wide range of technologies and processes, all of which are focused on protecting data from loss or corruption. For this report, we bring together results from questions that attempt to define and measure which of these technologies and processes are more or less prevalent in businesses today. Clearly, when the AS/400 was first introduced in 1988, and likewise the RS/6000 some two years later, the options for data protection were very limited. Backup to tape was clearly the predominant method. Compared to today's standards, disk was very expensive, and disk parity (RAID) and disk-to-disk replication were in their infancy.

Since then, of course, many technologies have been developed, generally offering more efficient and granular management of data using journaling, logical replication, clustering, disk mirroring, continuous data protection (CDP), and more.

## 5. What data protection methods do you currently employ?

For this question, a range of options was presented, and the participants were invited to “check all that apply.” Thus the percentages reported add up to more than 100 percent.

This is a combined result for IBM i and AIX shops. Several features of the results are very clear. First, tape backup is still very much in use. In some shops, it has become more of an adjunct technology: clearly, the shops using disk-to-disk replication or mirroring must have contributed to the 79 percent ranking of tape stored offsite. In such a case, it seems reasonable to state that the primary recovery plan depends upon the secondary disk, with recovery from tape being reserved for “worst-case” scenarios.

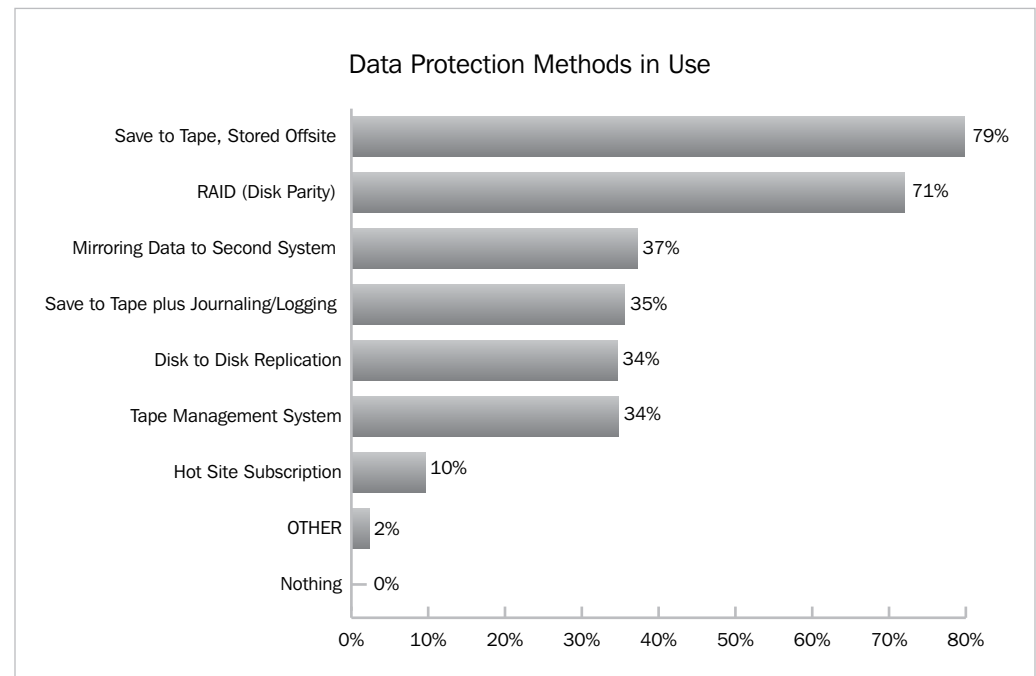


Figure 7

However, based on these numbers, a significant percentage of organizations are still relying solely or almost entirely on tape-based methods for DR, whether tape stored offsite, a tape management system, or tape in conjunction with

disk-based journaling or logging. It is interesting to note that alternative tape-based options, specifically tape backup automation and tape plus journaling/logging, are reported in similar numbers as the far more sophisticated and powerful replication options.

RAID disk technology is also nearly universal. It is a mature technology that requires minimal intervention. By its strong prevalence, it coexists in many of the reporting companies, alongside mirroring and replication technologies, as a fundamental element of a multi-tiered approach to data protection.

But perhaps the most interesting result comes when this set of responses is compared with the earlier questions regarding RTO and RPO, as well as the estimation of DR plan completeness. Recalling, some 56 percent to 68 percent of respondents declared recovery time objectives (RTOs) being under 12 hours, and some 70 percent identified recovery point objectives (RPOs) in the range of one hour or less of data loss. In addition, self-reported DR plan completeness was very high, with roughly 65 percent saying their DR plan was 75 percent or more complete.

### **Key Finding**

From the present chart, it appears that a large percentage of companies have a disconnect between their goals and perception of DR readiness and the actual state of affairs. Again, a large percentage of shops reported relying solely or predominantly on tape-based methods. Achieving intra-day recovery time with these methods is generally unlikely, and to expect tape to support RPOs of one hour or less is just not realistic.

To borrow a phrase from Alan Greenspan, former Chief of the U.S. Federal Reserve, the likely reality is that the reported RPOs and RTOs, as well as the numbers for confidence in DR plan completeness, suffer from a bit of “irrational exuberance,” unless more advanced technologies are implemented to support the more aggressive objectives.

## 6. Does your company use any of the following data replication technologies?

Setting aside tape backup and RAID, in this question we focused more closely on replication technologies in use. Once again, respondents were invited to report all technologies they have in use, so the percentages add up to over 100 percent.

For this question, we again have responses from companies that use both IBM i and AIX. This group provided us with an especially interesting result.

First, looking only at those shops that identified with one OS, the unambiguous answer “None” was strong, with that response exceeding 45 percent in IBM i shops. (Figure 8)

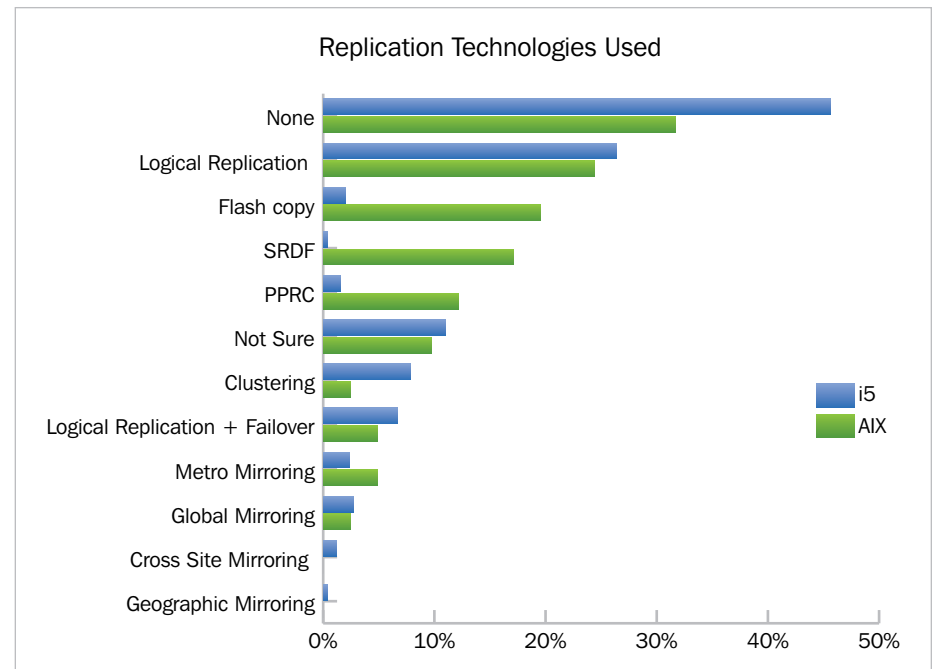


Figure 8

### Key Finding

Software-based logical replication was in use in about 25 percent of all businesses reporting, while only 5 percent indicated that they use logical replication in conjunction with automated failover (High Availability, or HA). This result is consistent with all previous reviews of shops using logical replication, based on formal surveys and informal conversations with end users and software providers. Despite the continuing development by replication software vendors of easier and more effective failover capabilities, the vast majority of end users who are using logical replication have not yet implemented the failover capability.

Adding to the wonder of this result is the fact that, within the IBM i market, IBM has made major efforts recently to make software-based server failover more popular and affordable through the introduction and strong promotion of low-cost Capacity Backup Unit (CBU) servers. Despite all this, the untapped potential for High Availability failover solutions remains large.

Looking at the results for other replication technologies, a technological divide between IBM i users and their AIX counterparts is apparent. Flash Copy, EMC's SRDF (Symmetrix™ Remote Data Facility) for storage network mirroring, and IBM's PPRC (now referred to as "Metro Mirror") all play strongly in AIX shops, but not within the IBM i community. (Note that the lag in recognition of the Metro Mirror naming is apparent in the results: while the PPRC and Metro Mirror results could logically have been combined, we have kept them separate for this report.)

This difference is not unexpected. In AIX environments, clustering does not necessarily involve replication. Instead, server and application availability is maintained through "switchable" storage—that is, two servers that can attach to one set of storage (though, due to OS design restrictions, not at the same time). As a result, replication of data to a secondary site tends to be storage-system-centric, with Flash Copy, SRDF, and Metro Mirroring all fitting into that category.

In contrast, within IBM i architectures, separate storage is nearly always associated with each deployed server. In this architecture, clustering involves replication between separate servers with clustering software managing failover when needed. So while clustering is a topological concept for AIX pros, it is a distinct set of server availability management software in the context of the IBM i architecture.

The prevalence of storage hardware-based mirroring over long distances, or "Global Mirroring," is lower than that of "Local Mirroring." In general, Global Mirroring continues to be an option taken only by a smaller number of large enterprises.

At the beginning of this section, we mentioned an interesting result. Within the community of companies reporting the use of both AIX and IBM i, the use of logical replication jumps, along with Flash Copy and Metro Mirroring, while the reported use of the other technologies is generally similar to the single-OS shops. From other data reviewed, we also know that these same firms more often report using other UNIX- and Linux-based applications in their IT architecture as well.

From our own research and from numbers provided by IBM, we know that the percentage of shops within the IBM customer base utilizing both operating systems, though growing, has historically been in the 5–10 percent range. Taken together, the data suggests that there is a small but growing group of IBM midrange customers that tend to be more aggressive in the adoption of multiple technologies to help achieve their HA and DR goals.

**7. How useful would it be if you had the ability to quickly recover accidentally deleted or otherwise corrupted records or objects back to any specific point in time?**

This question probes for a general reaction to the desirability of continuous data protection (CDP). A relatively new category of data protection technology, CDP is, in essence, a replication-based

DR technology that is focused on recovering from “micro disasters,” those all-too-common events in which a single file or object is accidentally deleted or corrupted. By keeping a time-stamped record of every change made to the backup copy during real-time replication, CDP software makes it possible to recover, on demand, a copy of that file as it was just before the damage occurred or at any prior point.

In addition to the advantage of a more granular recovery (that is, recovery of individual objects at any chosen point in time), the most current and advanced CDP solutions feature user interfaces that speed and simplify the recovery

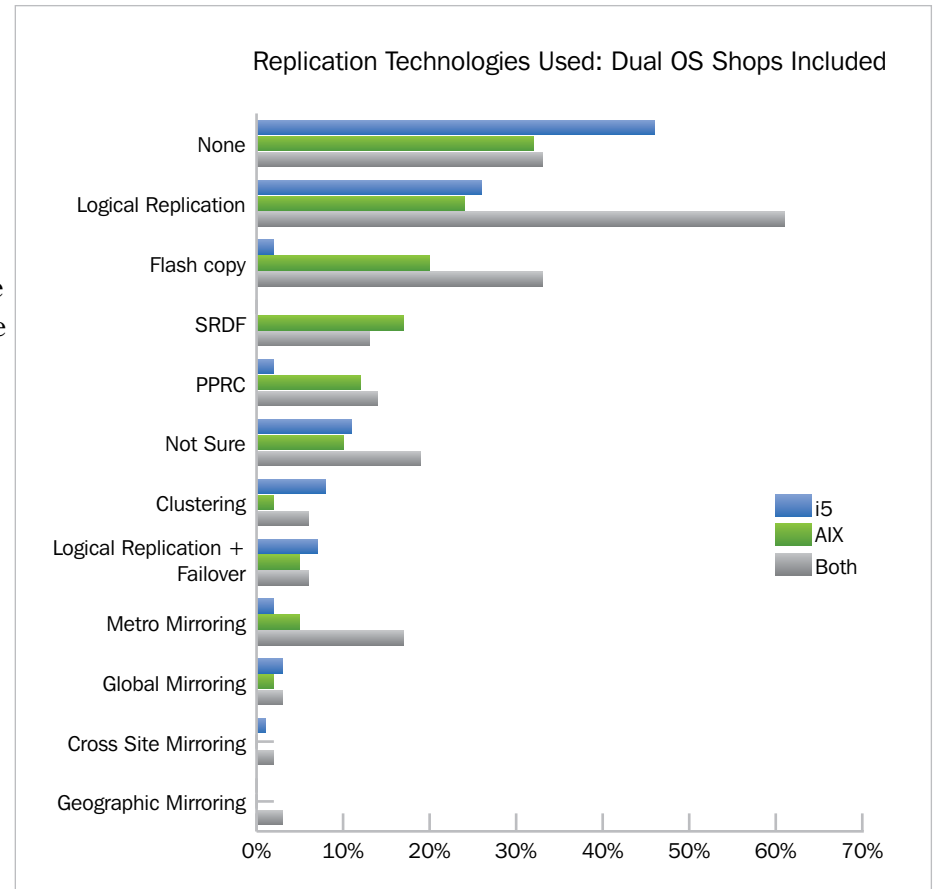


Figure 9

process, so much so that it is little more difficult than scrolling back through a DVD to replay a favorite scene in a movie.

Again, a preponderance of responses was positive, with 64 percent indicating that this would be very useful or extremely so. Keep in mind that this was the response from IT professionals who are accustomed to the world-class reliability and stability of their Power Systems servers. But even the best application, OS, and hardware designs cannot fully protect against human error. It is clear that easier file- and object-level recovery technology, such as CDP, is desirable and will without a doubt be increasingly adopted by IBM server customers.

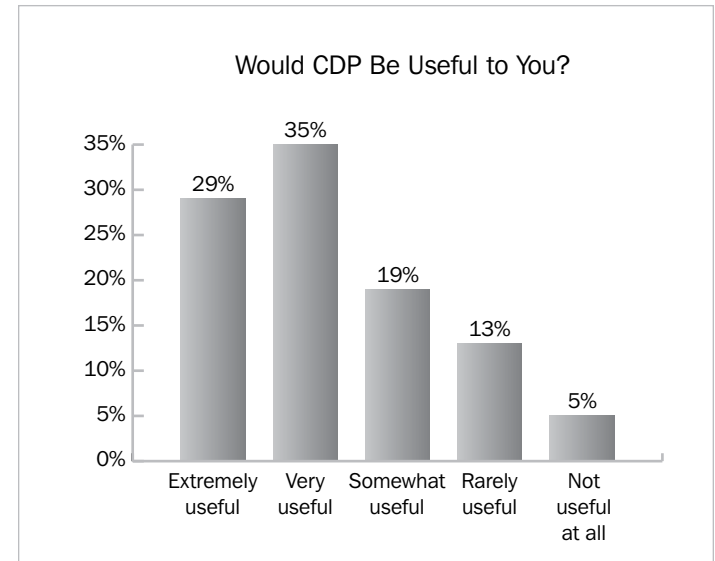


Figure 10

## The State of Systems Optimization

Data stores have always required maintenance and quality assurance work. That much has not changed in the twenty years since the AS/400 was introduced. But what has changed is the sheer volume of data being added annually to the average corporation's database and the level to which those companies' livelihoods depend upon the availability and accuracy of that data. All of the data protection technologies previously reviewed focus on ensuring the existence of data, with some amount of built-in error detection and correction being applied to the copying process.

But the measures of data quality go beyond simple accuracy. For example, a million-record database, in which each record is individually accurate, can still be of low quality if the records are not ordered or keyed properly or if half of the records are in fact marked as deleted but still must be manipulated for reporting. Data stores that cause poor interactive response or batch processing time are poor quality data, no matter how accurate the information is. Thus, rather than referring to data quality, per se, we will use the term "optimization" in order to address the wider range of data management needs.

With so much data flowing daily, the work of keeping the databases and systems operating efficiently is a constant challenge. Just as in the realm of data protection, data and systems management tools and technologies have been developing apace with the growing challenges since the birth of the AS/400 twenty years ago. So this is another area that we wanted to explore to review how well current systems management and optimization practices are supporting the greater goals of information availability and overall business resilience.

A note about the survey results we are about to discuss: for this subject, we conducted surveys only among the IBM i community. As you will see, the results are rather interesting, and we do intend to expand future work on this subject to include the AIX community.

## 8. Which of the following system management capabilities would be beneficial to you?

We'll begin by examining survey questions that probed for perceived needs. The full wording of the options—from which respondents could choose all that applied to them—was as follows:

- Ongoing optimization of my IBM i environment
- Reclaiming DASD—we're growing close to capacity now
- Improving response times for our interactive application users
- Improving response times for batch processing

Essentially, the three primary benefits of optimizing files and objects were identified as priorities by the respondents in fairly equal proportion. But the desirability of “ongoing” optimization, which implies automation and real-time processing, was much more universally acknowledged. Clearly, we identified an operational pain point with this question.

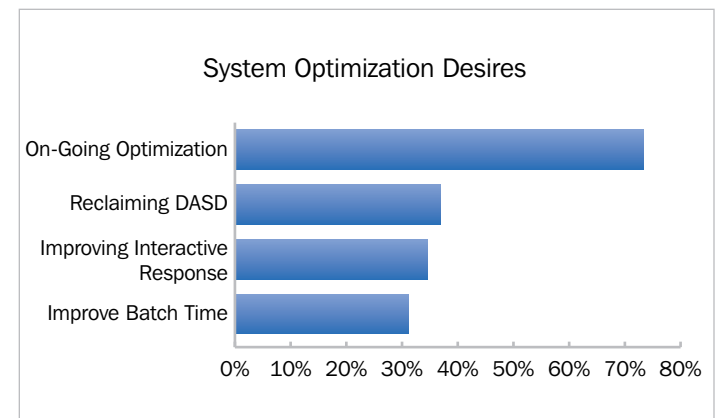


Figure 11

The response we received to a related though somewhat simpler question confirms automation as an important need:

**9. Would you like to be able to automatically perform optimization and maintenance on physical and logical files, programs, queues, and spool files?**

On a basic up or down vote, the Yes vote for this question was 76 percent. From the strength of the vote, we surmise that automating optimization is a new idea—or at least a newer option that these IT pros have not likely already implemented.

Based upon this, our line of questioning moved on to determining, in the absence of automation, what is currently being done to optimize data in these companies.

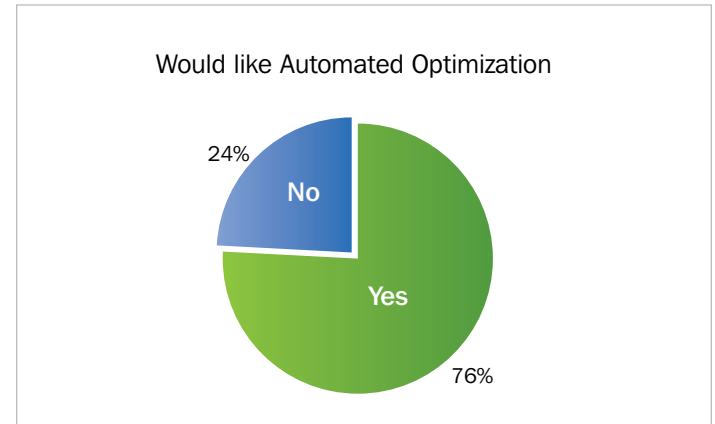


Figure 12

**10. How do you handle ongoing maintenance and optimization for your IBM i systems and applications?**

Respondents were allowed to identify whether more than one method or tool was in use. Given its OEM nature, we expected to see IBM’s Systems Director as a common response, so it was included by name in the list of possible responses.

The use of command line invocation for optimization tasks, whether manually or as part of scripted Control Language (CL) programs, was quite prevalent. Given the audience for this question, we expect that in-house developed programs that utilize other than IBM OS CL scripting are included in the results but do not predominate.

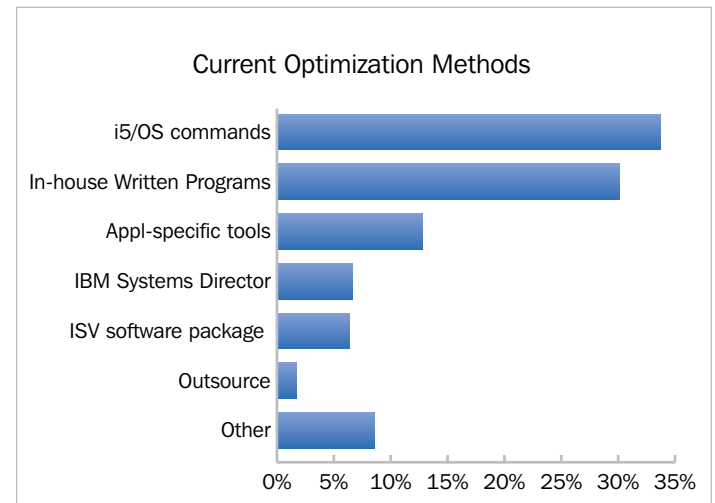


Figure 13

For both the Other and ISV categories, respondents were invited to name their alternate method or product. Director™ from Vision Solutions was identified in more than 1 percent of the responses while 40 other products, including just a few services, made up the remaining responses for these write-in votes, indicating a fragmented marketplace that has not yet matured or chosen a reigning champion.

Despite the fact that we know automation is desired, Do It Yourself is still the most common modality by far. So if manual optimization is in fact an operational pain point, what are the root causes of that pain? Why is Do It Yourself not loved? The next question addresses this.

**11. How often do you ensure that your IBM i environment is clean and optimized? Include time spent performing tasks such as reorganizing DASD, tracking jobs by CPU or resource usage, and optimizing and maintaining applications and physical and logical files.**

Given the relatively new territory in which we are working (for the participants as well as ourselves), we opted for the detailed question to ensure the goal of the question was understood. Responses were limited to four time-frame frequencies.

Note the bias to the extremes. Not to say that two completely differing camps exist, but there appears to be a tendency either to make optimization a daily routine or to defer it as occasional “project-level” work.

To help explain this, we now correlate the prior two questions regarding methods and frequency. In other words, does how you do it affect how often you do it? Indeed, it does. Breaking out the responses into three groups, effectively by level of automation in their methods, we plotted responses against their declared frequency.

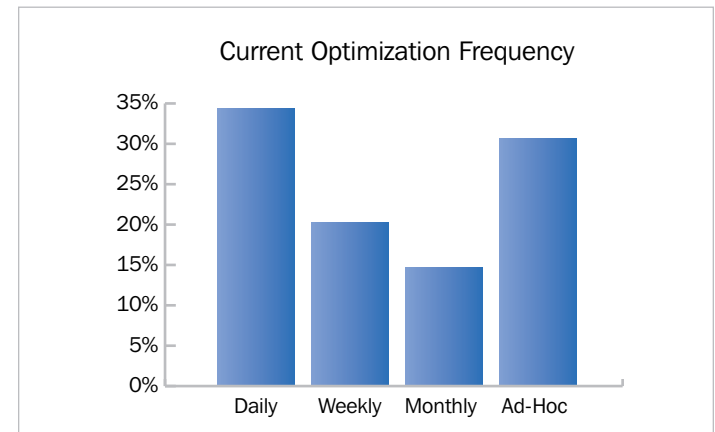


Figure 14

**Key Finding**

The results are clear. Automation enables and encourages daily application of the optimization discipline. Lack thereof influences a deferral pattern of behavior. In a similar exercise, we asked in the same survey about the impact of large files or objects on the optimization process.

**12. Do you have any large files that you have not optimized because you believe that the time to reorganize them might exceed your planned downtime window?**

This self-classification question was limited to a Yes or No response. The result was a near tie: 48 percent said Yes, they avoid optimizing large files due to the downtime impact.

Perhaps ambiguous on its own, this result gains more meaning when, as with the prior question, we overlay these Yes and No responses on their corresponding responses to the optimization frequency question.

In Figure 16, we see an amplification of the extreme-biased result of the frequency question. Clearly, the avoidance of optimizing large files enhanced the likelihood that optimization, though incomplete, would be done daily. In other words, “keep it quick, and I’ll do it more often.” Alternately, dedication to inclusion of large files tended to influence a decision to defer optimization altogether until large, project-level windows of opportunity were available.

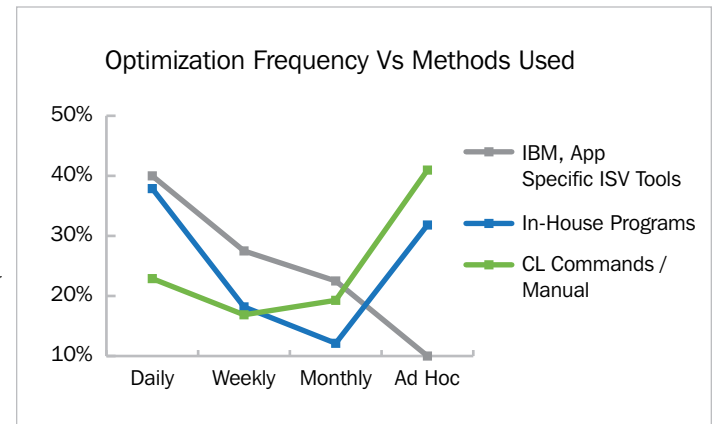


Figure 15

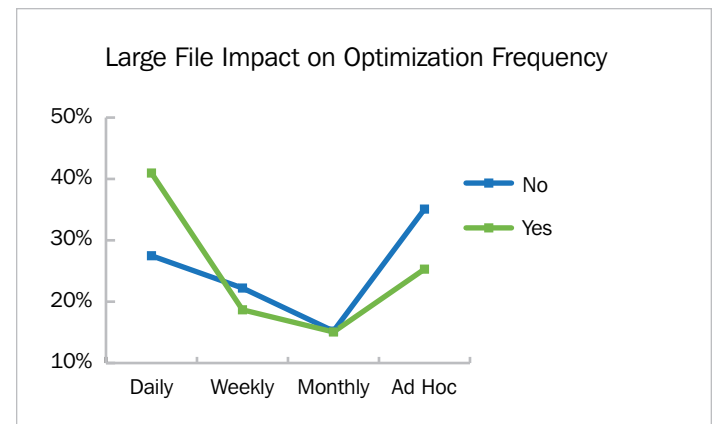


Figure 16

So, while optimization is clearly understood to be necessary by most practitioners, it is certainly not viewed as painless, simple, or quick. It is, in short, a chore. And any IT task that is a chore is ripe for automation.

## Summary: The State of Resilience and Optimization

Overall, we find that Power Systems customers continue to reflect the same values they always have. They are at once very pragmatic and yet open to change when new technologies prove themselves valuable and truly ready for deployment. The global proliferation of the AS/400 occurred not in spite of, but rather primarily because of, twenty years of constant innovation and change. And because of these same advances, business today moves and changes at a pace that makes 1988 seem like “the good old days.”

Up against demands for greater operational efficiencies, faster processes, and increasingly, 24 x7 operations, businesses cannot afford to take chances with unproven or immature technologies. But these very demands for efficient, uninterrupted operations also make these same businesses ripe for adoption of newer resiliency and optimization options that better protect their data and more effectively meet internal and external availability SLAs.

And so we see in our data evidence of a dynamic tension, a balance under pressure. Our survey respondents are well aware of the resiliency and optimization that modern business requires and acknowledge increasingly stringent RTOs and RPOs. They also exhibit a pattern of careful and deliberate choices among the increasing number and range of data protection, availability, and optimization technologies that have become available over the past twenty years.

Some technologies, such as tape backup and RAID, have become ubiquitous. Others, such as replication with failover (HA), show focused adoption within particular groups of businesses. But in other areas, legacy methods, such as relying upon CL scripts for optimizing data stores, still persist while their practitioners, clearly hungering for an easier and better way, carefully gauge the readiness of more automated options.

Thus, we may confidently expect to see the adoption of all of the discussed data protection and optimization technologies to accelerate within the Power Systems market, with those that are the most pragmatic and simplest to use gaining the most new mind and market share. And undoubtedly, the resiliency of businesses using Power Systems servers, and their successors, will continue to improve over the next twenty years and beyond.

## About the Information Availability Institute

The Information Availability Institute provides research and education that helps business professionals of all disciplines to understand, evaluate and apply information availability technologies.

Drawing upon the deep resources of Vision Solutions, a leading provider of IT resiliency, recovery, and optimization technologies, as well as its technology partners and independent industry experts, the IAI is committed to identifying and communicating improvements in technologies that increase Information Availability and overall business resilience across the entire enterprise.

