

The NetWorker Blog

Commentary from a long-term NetWorker consultant and backup theorist

Overview

Between September 16 2012 and October 1, a survey was run on the NetWorker blog (<http://nsrd.info/blog>) to gather a snapshot of data recovery volumes, frequency and initiators.

This survey aimed to review:

- Common age of data being recovered;
- Frequency of data being recovered;
- By whom the recoveries are being completed;
- Samples of how much data are being recovered.

The survey was open to users of all backup products.

About the Author

Preston de Guise has been specialising in data protection services since 1996, and has provided consulting services to a diverse selection of companies ranging from small sites to companies in the Global Fortune 500.

Preston is the author of “Enterprise Systems Backup and Recovery: A corporate insurance policy” (ISBN-10 1420076396, ISBN-13 978-1420076394). Written for both technical and management users, “Enterprise Systems Backup and Recovery: A corporate insurance policy” provides insight into best practice approaches to designing policies and procedures for ensuring that data protection solutions installed form a cohesive and reliable system within an enterprise. Details of the book can be found at <http://www.enterprisesystemsbackup.com>.

Preston de Guise currently works for IDATA Resolutions, an Australian/New Zealand company that specialises in storage, archiving, data protection, virtualisation and high availability solutions. IDATA provides a wide range of services including installation and configuration, training, remote support, remote audits, on-site support, operational assistance and managed services. IDATA Resolutions can be found on the net at <http://www.idataresolutions.com>.

Table of Contents

Survey Introduction	3
How frequently do you recover data?	4
<i>Comments and Conclusions</i>	4
Age of backup data being recovered	5
<i>Comments and Conclusions</i>	5
Percentage of Data Recovered that is Production	6
<i>Comments and Conclusions</i>	6
Percentage of recoveries performed by backup administrators	7
<i>Comments and Conclusions</i>	7
Percentage of Recoveries performed by Operators	8
<i>Comments and Conclusions</i>	8
Percentage of recoveries performed by system/application administrators	9
<i>Comments and Conclusions</i>	9
Percentage of recoveries performed by end users	10
<i>Comments and Conclusions</i>	10
GB per week recovered	11
<i>Comments and Conclusions</i>	11
In Conclusion	12

Survey Introduction

Everyone is well aware that we backup in order to recover.

Yet, surprisingly, even rough details of what is recovered, how frequently it is recovered, and by whom it is recovered, as well as some other basic details, are often difficult to determine.

This minimum information sometimes makes the life of a backup team more challenging – it fosters a belief that backup systems are rarely, if ever, used for recovery, and therefore, a waste of money. (Or at least, by those who don't understand the concept of electronic insurance.)

Due to the relative brevity of the survey, 62 responses were received; while this is a lower number than other NetWorker Information Hub surveys, the fairly broad distribution of results suggests it may be largely representative of the more broader backup and recovery community.

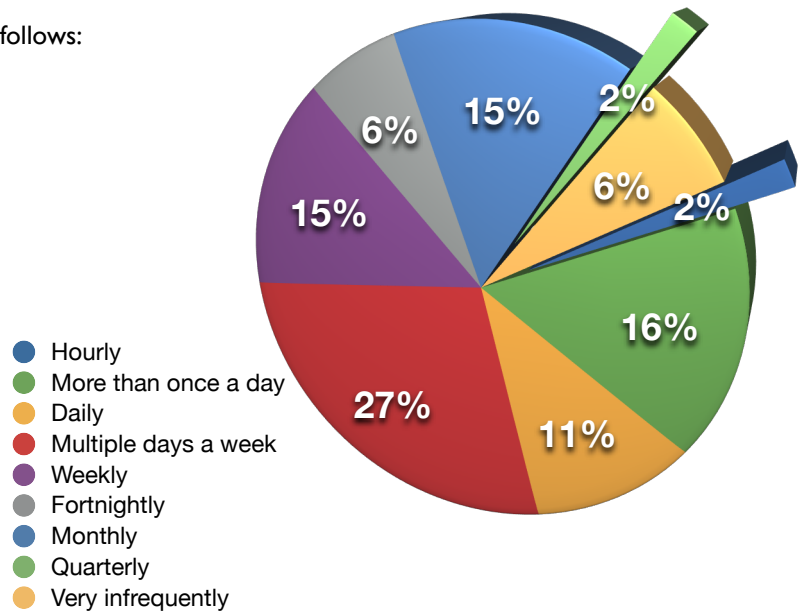
Note all respondents responded to all questions; where fewer responses were received, percentages have been calculated off the number of responses to that question, rather than the total number of respondents.

Much gratitude is owed to all respondents.

How frequently do you recover data?

The numbers for the responses were as follows:

- 👤 **Hourly** - 1
- 👤 **More than once a day** - 10
- 👤 **Daily** - 7
- 👤 **Multiple days a week** - 17
- 👤 **Weekly** - 9
- 👤 **Fortnightly** - 4
- 👤 **Monthly** - 9
- 👤 **Quarterly** - 1
- 👤 **Very infrequently** - 4



Comments and Conclusions

Anecdotally a general consensus amongst many in the backup community is that recoveries are only run rarely. Yet, evidence from backup administrators completing the survey indicate quite a contrary response. Just on 70% of respondents indicated a recovery frequency of at least weekly, with over 50% indicating multiple recoveries a week.

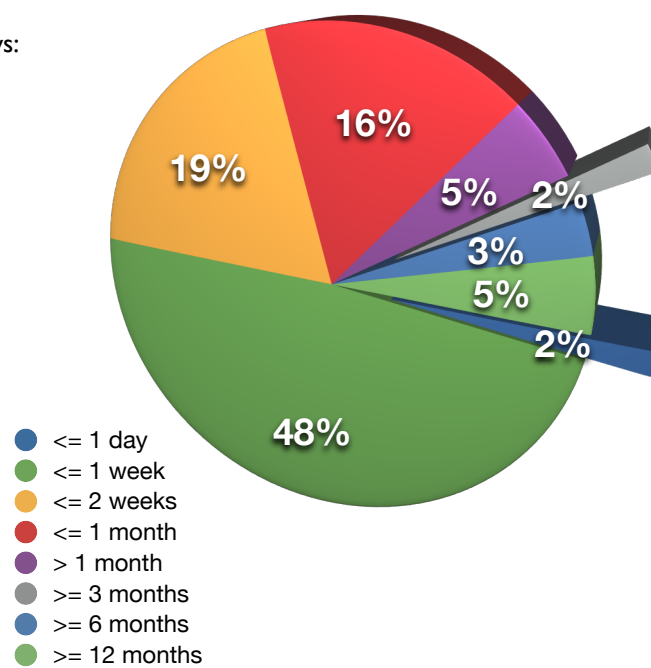
Conversely, sites that recover data on only a monthly basis or even longer are clearly in the minority – just under 23% of respondents indicated this was their recovery frequency profile.

This emphasises a core design consideration – backup systems must be oriented to facilitate recoveries as quickly and painlessly as possible, with an absolute minimum number of constraints on factors that may interfere with recoveries.

Age of backup data being recovered

The numbers for the responses were as follows:

- 👤 ≤ 1 day – 1
- 👤 ≤ 1 week – 30
- 👤 ≤ 2 weeks – 12
- 👤 ≤ 1 month – 10
- 👤 > 1 month – 3
- 👤 ≥ 3 months – 1
- 👤 ≥ 6 months – 2
- 👤 ≥ 12 months – 3



Comments and Conclusions

This held little in the way of surprises – the vast majority of data recovery requests occur for data that has only recently been backed up. In this case, 50% of recovery requests are deemed to be against data that has been backed up within a period of 7 days from the recovery date.

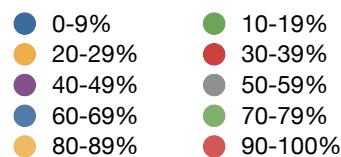
As would also be expected based on typical retention cycles, the amount of data backed up more than a month ago being recovered tends to be quite low – based on the survey results, it sits at a reasonably meager 15% (rounded).

Taking into account our previous question, this demonstrates that not only does a backup system have to be designed to facilitate recoveries with minimum fuss, but it must also be most optimised to facilitate recoveries of the most recently backed up data. Further, where the number of recovery requests for more frequently backed up data is high, or where end users may need to recover that data directly, it can also indicate a requirement to work backup into a larger *information lifecycle protection* strategy, instead of trying to handle all data protection just within the backup product. (This is covered in greater detail on the blog, “ILP Policies vs Backup Policies”, <http://nsrd.info/blog/2012/05/15/ilp-policies-vs-backup-policies/>)

Percentage of Data Recovered that is Production

The numbers for the responses were as follows:

- 0-9% – 4
- 10-19% – 3
- 20-29% – 5
- 30-39% – 1
- 40-49% – 3
- 50-59% – 5
- 60-69% – 1
- 70-79% – 12
- 80-89% – 13
- 90-100% – 15



Comments and Conclusions

The survey confirmed that the vast majority of data being recovered is categorised as production data – more than 64 per cent of respondents indicated that production data accounted for 70% or higher of the data recovered within their environments. Almost 25 per cent indicated that it accounted for 90 per cent or higher of the data recovered.

This too would indicate a design strategy – assume almost two thirds of data being recovered is production data. While non-production data may typically be recovered as a non-urgent activity, this is rarely the case with production data.

Percentage of recoveries performed by backup administrators

The numbers for the responses were as follows:

0-19% – 20

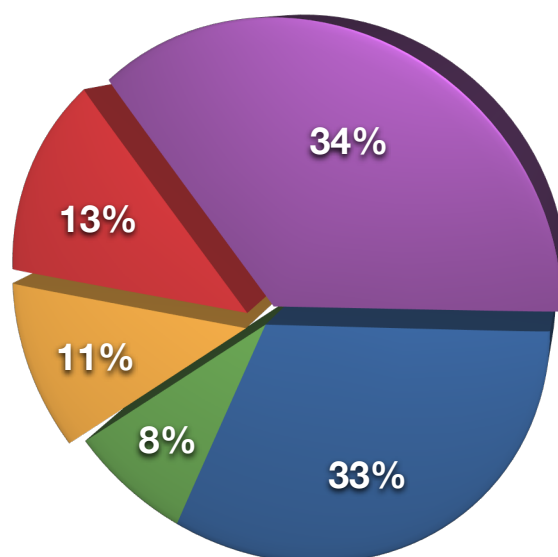
20-39% – 5

40-59% – 7

60-79% – 8

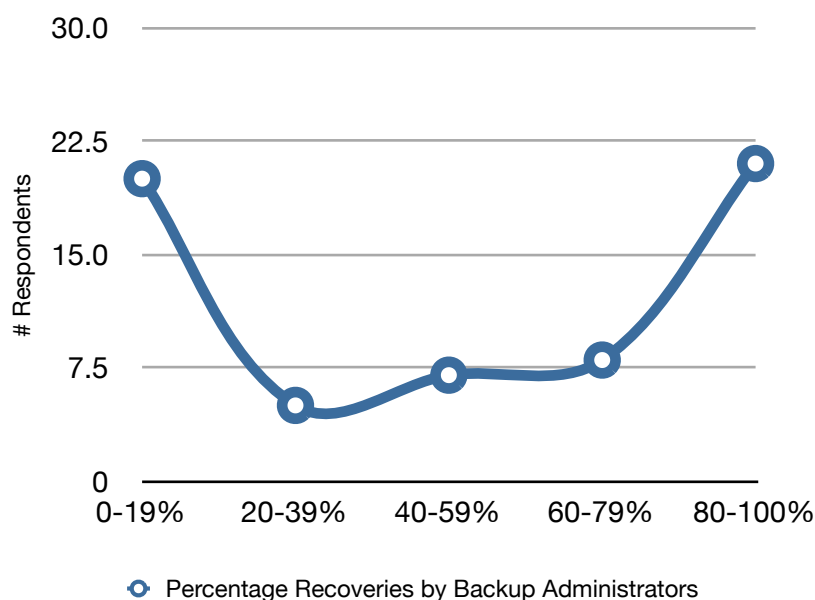
80-100% – 21

0-19%
20-39%
40-59%
60-79%
80-100%



Comments and Conclusions

Interestingly, this profile suggests a fairly polarised approach to recovery facilitation, and is best demonstrated through a line graph:



While there are some sites where the recovery workload is shared around (as indicated by the middle of the graph), it would appear that the majority of sites fall into two different categories – those where backup administrators perform the majority of the recoveries, and those where backup administrators are rarely involved in recoveries.

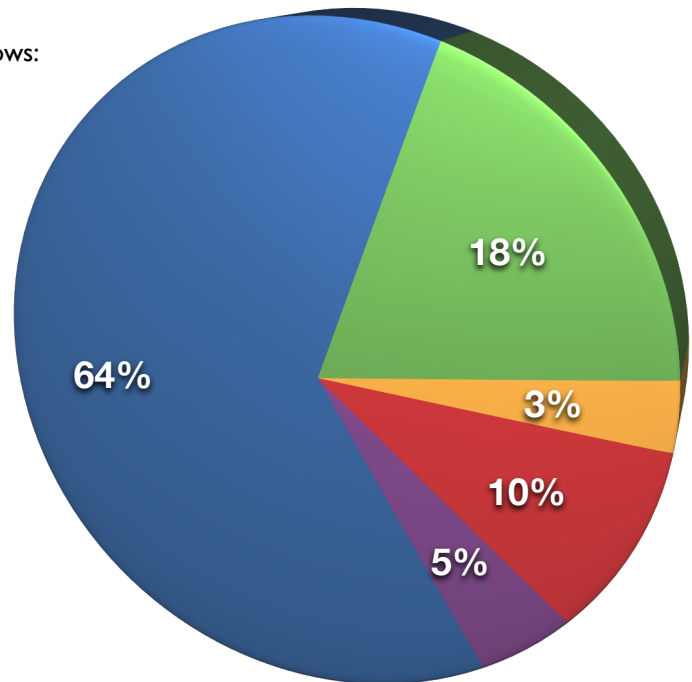
Sites should be carefully aware of which profile they fall into, so that training can be directed and oriented appropriately.

Percentage of Recoveries performed by Operators

The numbers for the responses were as follows:

- 👤 **0-19%** – 39
- 👤 **20-39%** – 11
- 👤 **40-59%** – 2
- 👤 **60-79%** – 6
- 👤 **80-100%** – 3

- 0-19%
- 20-39%
- 40-59%
- 60-79%
- 80-100%








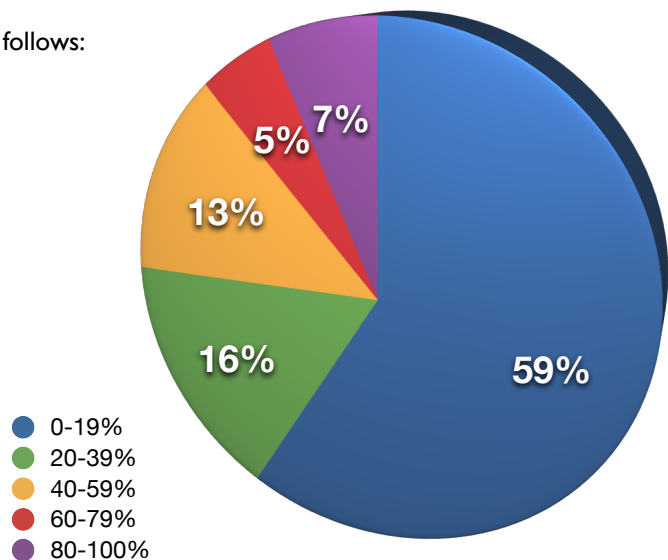
Comments and Conclusions

While I was still a system and backup administrator in the 1990s, before entering full time consulting, operators were the most likely group to be performing recoveries within an environment. This was starting to decline in the early 2000s, and the responses indicate a continued shift away from operator-led recovery.

Percentage of recoveries performed by system/application administrators

The numbers for the responses were as follows:

-  **0-19%** – 36
-  **20-39%** – 10
-  **40-59%** – 8
-  **60-79%** – 3
-  **80-100%** – 4



Comments and Conclusions

Only a small percentage of sites leaves recovery operations in the hands of the system and application administrators. Observationally this would be for a simple reason – the majority of recovery requests are typically for end-user data, be it email or file, and both types of recoveries are typically simple enough so as to not require a specialist to be involved.

Percentage of recoveries performed by end users

Unlike the other questions focusing on roles involved in recoveries, this question explicitly included an option for 0.

The numbers for the responses were as follows:

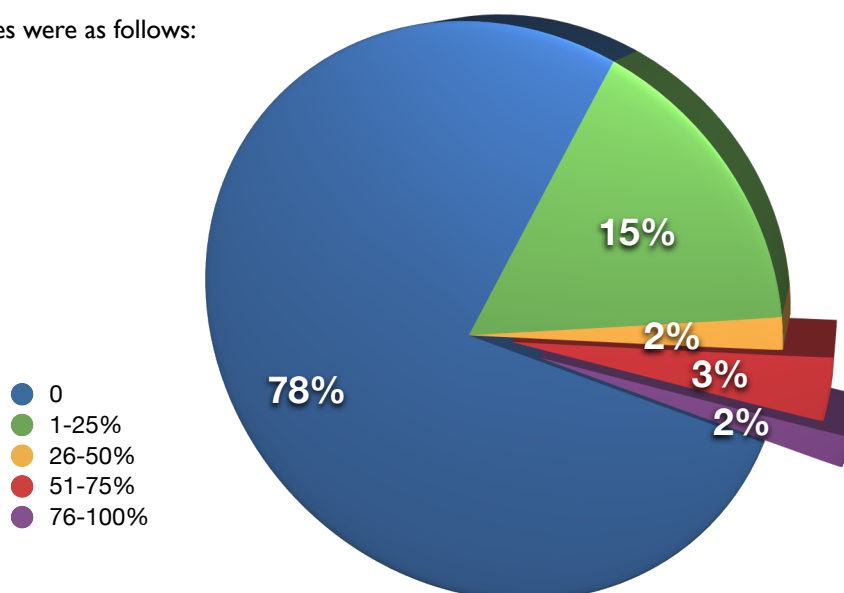
👤 0% – 47

👤 1-25% – 9

👤 26-50% – 1

👤 51-75% – 2

👤 76-100% – 1



Comments and Conclusions

Few sites facilitate end-user initiated recoveries. Historically, while most products have supplied an interface that allows for end-user recovery, a key dissuading point for end-user recoveries has been the overall architecture. (In physical tape based environments, allowing a large number of users access to the recovery interface typically leads to chaos.)

One might have assumed the ongoing changes in backup architecture – the move to disk backup for instance – might have lead to more recoveries being facilitated by end-users themselves. The survey result shows this isn't the case (or at least, not yet), and there are a couple of likely reasons for this:

- **Resistance to change** – Backup and recovery functions tend to be largely conservative in nature; modifying an old quote, one might say “there are old backup administrators and there are bold backup administrators, but there are no old, bold backup administrators”. Since the entire purpose of backup and recovery systems is to provide dependable data restoration services, once a process is introduced and found to work, modifying that process becomes non-trivial. Thus, since historically recoveries have been facilitated without the end-users being involved, many sites see no real reason to change.
- **Appropriateness of function to role** – Another key reason why end-users may not be involved in the recovery process is whether such an activity is actually appropriate to their role within the company. While many technical people might argue that recovery of ones own data should be integral to someone's job, many other professions would see otherwise. Why should an accountant, or a financial controller, or an academic or secretary – or indeed, any other profession – need to train in activities that can be more appropriately performed by other roles within the organisation? After all, any IT worker with a basic understanding of spreadsheets would likely be able to perform all manner of financial functions (e.g., expense reimbursements, mileage reimbursements, etc.) yet we recognise that only certain roles can appropriately perform these functions.

Ultimately the ‘dream’ of end-users running the majority of data recovery processes within an organisation may be a bit like the paperless office – a nice idea, but not likely to take hold any time soon.

GB per week recovered

How much data per week is recovered within an environment? This question was always aimed at being a guestimate only – yet it's an answer many backup administrators who properly manage their environment would have a fair idea of.

This was a free-form question; rather than provide scales, respondents could enter any amount in GB they wished. The lowest numbers entered were in the order of 0.1 GB per week; yet, at the opposite end of the scale, 1500 GB, 2000 GB, 4000 GB and 9000 GB were all provided as answers, too.

The answers provided were broken into several ranges as follows:

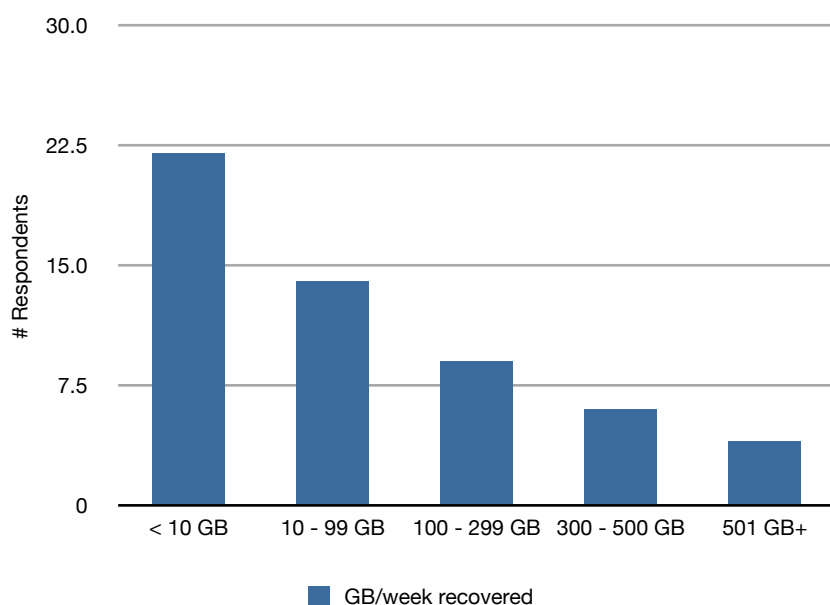
👤 < 10 GB – 22

👤 10 - 99 GB – 14

👤 100 - 299 GB – 9

👤 300 - 500 GB – 6

👤 501+ GB – 4



Comments and Conclusions

While some sites use their backup product to recover large amounts of data on a regular basis, the majority of sites are only recovering small, targeted amounts of data – 65% of sites recover less than 100GB of data per week.

In Conclusion

The recovery survey results allow us to draw a few conclusions about how backup environments should (generically) be designed in order to maximise recoverability:

- Assume the majority of recoveries will occur for data backed up within the last 7 days;
- Assume two thirds or more of recoveries will be of production data;
- Assume a reasonably high recovery frequency – at least weekly, likely daily.

Of course, there will always be exceptions to the above – yet for many sites, they'll work well as a good rule of thumb in the absence of other data.

The survey will be re-run next year, with some additional questions targeted at refining and drawing additional conclusions from the data.