November 2010

## **Overview**

Between November 17 and December 2 2010, a survey was run on the NetWorker blog (**http://nsrd.info/blog**) to gather a high level overview of the NetWorker usage of as many respondents as possible.

The **NetWorker** Bloa

nentary from a long-term NetWorker consultant and b

This survey aimed to review:

- Operating systems;
- NetWorker versions;
- · Licensed modules;
- Cloning;
- Deduplication take-up;
- Backup to disk;
- Open source databases.

# **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	4
How many datazones are you running?	5
Comments and Conclusions	5
What is the version of your NetWorker server(s)?	6
Comments and Conclusions	6
Number of clients being backed up	7
Comments and Conclusions	7
NetWorker Server Operating Systems	8
Comments and Conclusions	8
Aside – Oracle, Sun and Solaris	8
Client Operating Systems	9
Comments and Conclusions	9
Omission	9
Sites using Deduplication	10
Comments and Conclusions	10
Modules in Use	11
Comments and Conclusions	11
Open Source Database Usage	12
Raw responses	13
Responses only from users of OSDBs	13
Factoring "would not use" non-OSDB deployments out	14
Comments and Conclusions	14
Backup to Disk	15
Comments and Conclusions	15
Cloning policies	16
Worker Usage Survey, November 2010	2

Comments and Conclusions	16
In Conclusion	17

# **Survey Introduction**

The primary goal of this survey was to act as a follow-up from the NetWorker Usage Survey conducted in March 2010, with particular focus on NetWorker versions as well as getting clarification on options not previously polled, such as backup to disk and deduplication. Special consideration was made for polling respondents about backup of open source databases.

There were 152 responses to the survey. Since the survey needed to cover multiple data zones, and most questions allowed multiple selections, there are several questions where the total number of selected answers exceeds the number of individual survey responses.

Much gratitude is owed to all respondents.

#### How many datazones are you running?

The numbers for the responses were as follows:



# Comments and Conclusions

While the clear majority of sites have a single NetWorker server deployed, what is telling is that the majority is (just) less than 50%. Despite FUD to the contrary, NetWorker clearly remains a strong enterprise solution, with some sites deploying a large number of datazones to suit their requirements.

Anecdotal and personal evidence would suggest that sites with a large number of datazones typically fall into one of two categories:

- National or multinational companies where disparate datazones exist across multiple geographic regions;
- Companies that offer backup services, and isolate customers or clusters of customers by datazone.

# What is the version of your NetWorker server(s)?

With 231 answers on this (out of 152 respondents), it's clear that the environments with multiple datazones don't keep them all running at the same version.



# Comments and Conclusions

It's reassuring to see the clear majority of NetWorker datazones running a supported version of the NetWorker server (v7.5.x and higher as of this report publication date). Particularly telling is the high take up rate for NetWorker 7.6 SPI. In the previous report I'd theorised that NetWorker adoption followed a similar bell curve to the classic product adoption model, but that model is certainly not supported by the high percentage of v7.6 SPI users.

It would seem that the high number of useful new features in 7.6 SP1, including scheduled cloning, enhancements to ADV\_FILE devices, Data Domain support, etc., has been sufficient to encourage many NetWorker users to jump to a leading edge release of the product.

A telling fact remains that a number of datazones (almost 10%) remain on 7.2.x releases. While undoubtedly some of these will be due to no longer being on maintenance, others would be representative of a 'fear' in some sites of the changes that were introduced 'nsrjobd' – the new job management system. While these fears were justified in v7.3.x and early versions of v7.4, any site reluctant to upgrade due to those factors should now be planning an update to a 7.5.x or 7.6.1 release.

The strong numbers for v7.6 SP1, versus the relatively low numbers for v7.6, would suggest that a considerable percentage of customers who made the jump to v7.6 updated to SP1 very quickly on its release.

# Number of clients being backed up

It should be noted that there were only 151 responses to this question; one of the survey results did not include an answer on this.



#### Comments and Conclusions

This result clearly demonstrates one thing: NetWorker is not really a workgroup backup product. To be sure, it can be deployed and maintained into a workgroup environment, but most sites that deploy it will have more than 100 clients. Indeed over 20% of the deployments surveyed have over 1000 clients being backed up. Note that question asked for a total number of clients across *all* datazones deployed. However, a NetWorker datazone can (when properly designed) easily handle thousands of clients.

In fact, based on the deployment model, more than 50% of sites using NetWorker are backing up more than 250 clients, positioning it very clearly in the datacentre for many organisations. Clearly NetWorker scales well.

# **NetWorker Server Operating Systems**

With 195 answers, clearly sites that run multiple NetWorker servers do not run them all on the same operating system type.



#### Comments and Conclusions

In the March 2010 survey results, we had server operating systems of 43% Solaris, 29% Windows and 22% Linux – this survey sees Solaris and Windows numbers remain the same, with Linux shrinking by 3% – quite likely a small enough difference to be a blip based on survey responses.

Clearly these three operating systems maintain the lions share of operating system support – and proof positive that many companies don't want to be forced to use a particular server platform as their backup server. For example, products such as CommVault, which stubbornly only support a single server operating system attempt to force a design standard on users, rather than working with real world requirements.

#### Aside – Oracle, Sun and Solaris

Since the last survey, the Oracle takeover of Sun has been completed, and a common message occurring in the marketplace now is that Sun systems, as they come off maintenance, are being replaced by cheaper Linux or Windows systems due to the changes in Oracle's pricing model – particularly in the educational market. (For author opinion on why this will create significant long-term challenges for Oracle, refer to http://nsrd.info/blog/2010/04/22/rip-solaris/)

It will be particularly interesting revisiting the Solaris market share for NetWorker servers over the coming 1-2 years to see what impact Oracle's changes have made in the market.

# **Client Operating Systems**

With over 500 responses, NetWorker clearly continues to be very popular in heterogenous environments.



## Comments and Conclusions

Like the previous survey, the highest number of respondents were backing up Windows, Linux and Solaris. Reflecting its position in the mid-range server market, Windows remains the most backed up operating system, and this is likely to continue for some time, given the continued growth of virtualisation use within datacentres. Yet reflecting its heritage, NetWorker remains quite popular for backing up Unix based operating systems.

#### Omission

Embarrassingly, the author neglected to poll for Mac OS X client usage. In the previous survey Mac OS X protection had reached almost 5%; the next survey will again poll Mac OS X client usage and report on growth between the two (longer) periods.

## **Sites using Deduplication**

Deduplication is not a binary activity; sites may choose not to use it, or use it at the target level, or the source level, or a mix of the two depending on the circumstances at hand.



# Comments and Conclusions

While it is a strong growth market, deduplication has not yet hit commodity status within the backup environment. Part of this would be attributed to the high cost, regardless of product in use, compared to traditional backup. While there are compelling arguments to say, deploy 10TB of target deduplication storage instead of 50TB of raw target storage, the cost of said deployments mean that deduplication will typically be integrated into an environment as part of a major refresh cycle, rather than an ad-hoc update.

Without a prior poll to provide hard data, we can only draw on anecdotal evidence and personal experience; however, from these factors data deduplication is a growth market in storage at this point.

However, data deduplication products are currently at the premium end of the pricing range; while this is somewhat justified given the raw space savings they present, it is still not in the "commodity" storage price range, and this will see a slower growth/uptake than say, the reasonably fast adoption of backup to disk technologies within the data protection industry.

# Modules in Use

With 340 responses, this was aimed at not determining actual databases/mail servers/etc in use, but to find out what the layout of actual *module* use for those products were.

The numbers for the responses were as follows:



## Comments and Conclusions

The continued dominance of particular modules in use (Oracle, Exchange and Microsoft SQL Server) reflect both anecdotal evidence from social networking environments and this author's personal experience that these are the most commonly deployed/used modules within NetWorker datazones of any size.

One expects that the release of NMM 2.3 will be the catalyst for sites to move from dedicated Microsoft Exchange/SQL modules across to the consolidated Microsoft applications module.

# **Open Source Database Usage**

In this survey, two questions were asked regarding Open Source Database (OSDB) usage; the first was to get an understanding of respondent use; the second was to find out whether the community feels there is viability of a module.



Clearly MySQL retains the lions share of usage, though as a long-term PostgreSQL user, it is heartening to see that platform retaining such a strong share.

The supplemental question to this was: if there *was* an OSDB Module, how much would sites be prepared to pay, in US dollars, for the use of the module on a per database-server basis?

This turned out to require a bit more analysis than intended. The available answers to this question were:

- Would not use
- *≩* <\$500
- Between \$500 and \$1000
- Between \$1000 and \$2000
- Over \$2000

What wasn't anticipated was the prevalence of responses from sites *not* currently using open source databases. These fell into two obvious categories – those who answered "would not use", and those who answered with a price they'd be prepared for a module.

This leads to three takes on the data provided by respondents:

- Raw responses Unadjusted responses
- Only polling users of OSDBs Filtering out the responses from anyone who answered the pricing question but were not currently using OSDBs
- Adjusted "would not use" Filtering out only the "would not use" responses from people who answered they were not currently using OSDBs

#### Raw responses

The raw responses seemed to present a fairly negative view of the notion of an OSDB Module, with the breakdown as follows:



However, the "Would not use" category was somewhat influenced by sites not actually using OSDBs themselves, and this needs to be factored in order to see a clearer picture.

# Responses only from users of OSDBs

When we factor in OSDB usage, and remove from the results any answers from sites without OSDBs deployed, the break-down of answers changes considerably:



At this point, counting only "immediate" potential customers, the numbers change considerably on this; rather than "Would not use" being the dominant answer, a pricing model starts to present itself.

#### Factoring "would not use" non-OSDB deployments out

An alternate interpretation is that sites currently *not* using OSDBs but who answered with a pricing option (rather than "would not use") indicate potential customers – there is significant potential that these are sites who *would* use an OSDB *if* there were a reliable way of integrating it into the backup process. (The author's direct experience with customers reflects this.)

Evaluated on this basis, the numbers become:

- Would not use 19
- 🎽 **<\$500** 42
- Between \$500 and \$1000 23
- Between \$1000 and \$2000 3
- 🎐 Over \$2000 🗎





#### Comments and Conclusions

Regardless of how you look at the numbers (raw, fully or partially filtered), there's a significant number of sites out there currently using OSDBs who would be willing to pay for a module in order to integrate the backups of those databases with NetWorker. It's also worthwhile to remember that a cousin to MySQL, MaxDB, is increasingly deployed with SAP systems, and is begging for an integrated backup solution. (While there is a "published guide" to integrating MaxDB backups with NetWorker, it's so inept and dangerous that any site which attempts to trust production backups to this method needs to go back and re-learn data protection fundamentals.)

Clearly the most popular option was "sub \$500". Is this a viable pricing option? Unlikely – it's a somewhat unrealistic pricing option, but it's at least a door opening. There is, fortunately, hard pricing data already out in the marketplace. Zmanda, the company providing an "enterprise AMANDA" solution also provide a ZRM module, designed specifically to provide backups for MySQL. However, rather than paying a one-off license cost with yearly maintenance, Zmanda charge a yearly subscription for the use of the module. Assuming price fluctuations and an average yearly price of \$500 US, a 5 year ZRM investment for a single database server taken at individual years will cost \$2500. (While Zmanda offer discounts for longer-term purchases, I've suggested 5 x 1 year purchases to approximate yearly maintenance cycles.)

So, a one-off license cost of \$500 or less does seem unrealistic, even if an OSDB Module is simpler than proprietary database modules. A higher initial price but with lower yearly maintenance is very likely to deliver a cheaper long-term price than the Zmanda offering. It's also worthwhile remembering there's often a gap between what people would *like* to pay, and what they're *willing* to pay for a product.

Hopefully this will serve as food for thought for EMC product management. The market clearly wants – and is willing to pay for – an OSDB Module. Pricing users are prepared to accept appear balanced towards the lower end of the spectrum, but one suspects given the alternative subscription model offered by Zmanda, an adjustment to that spectrum spread would be acceptable to many sites.

## **Backup to Disk**

There were 176 responses to this question, reflecting the option of choosing multiple backup to disk options.



This shows beyond a doubt that backup to disk well and truly has become a mainstream configuration option. With only 16% of sites not using backup to disk, the majority of environments have moved through the process of optimising their backup environments to incorporate nearline storage.

It's not particularly surprising that ADV\_FILE usage is significantly higher than VTL usage within NetWorker.ADV\_FILE devices allow sites to bring their own storage (be it spare, or "cheaply" acquired) and just add a license or two.VTLs typically appear to be a larger capital investment in comparison; done properly they'll be specialised arrays running a custom operating system (or a "black box" environment).

Having closely followed the development of LinuxVTL (an open source VTL product aimed solely at lab and training environments), it's clear that the market has significant interest in VTLs, but price tends to shoe-horn VTL deployments into major hardware refresh activities.

That being said, as the global financial crisis eases, it will be interesting to gauge the growth of VTLs within NetWorker deployments in the coming 1-2 years.

# **Cloning policies**

The final question of the survey was around whether NetWorker sites are actively cloning in order to protect their data. All 151 respondents answered this question.

The answers were:



# Comments and Conclusions

First, it should be noted that as an oversight, no provision was made for sites that do not clone because of alternate replication processes (e.g., replicated VTLs with no tape-out); one suspects though based on personal and anecdotal experience that the number of sites that have actually cut tape completely out of their environment in such a way remain relatively low. (Depressingly, anecdotal evidence suggests that it's becoming increasingly common for mid-size sites to cut out tape, but not replicate their VTLs, backup grids, etc.)

As a data protection specialist, one heartening message that can be taken from this is that the majority of sites are guarding against their backup environments being a single point of failure. Indeed, 74% of respondents are at least cloning *something*. One would certainly hope that those sites cloning "very selectively" are indeed keeping a very careful eye on what *isn't* being cloned to avoid exposure to backup failure.

## In Conclusion

Without a doubt, NetWorker has a strong user following, and a strong usage base in the enterprise environment. With deployments featuring multiple datazones, thousands of clients with a plethora of operating systems and databases/applications being protected, many organisations, regardless of size, rely on NetWorker daily to ensure successful continued operations.

The work EMC product management and development have put into more recent releases (particularly v7.6 / v7.6 SPI) has clearly paid off: users are upgrading to take advantage of the new features and remain well within support ranges. This should be added incentive for meeting deadlines and feature requirement lists in planned releases – the market will clearly respond favourably to these.

The growth of open source databases within the *enterprise* show one key factor in particular in this survey – it's clearly time for a module capable of backing up *at least* MySQL; however, more generally one would suggest that it's time that all vendors (not just EMC) start to beef up their support for these products. The market is clearly saying that they want to use these products, and they're willing to pay for enterprise support.