The **NetWorker** Blog

Commentary from a long-term NetWorker consultant and backup theorist

## Overview

Between March 11 and March 20 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 three broad categories sampled were NetWorker versions, operating systems types and licensed features.

## 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, 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

The initial goal of the NetWorker Usage Survey was to help the author understand topics that should be considered for the NetWorker Blog moving forward. However, it quickly turned into something much more interesting: a good high level overview of how NetWorker is deployed. Additionally, this information should provide some lessons for EMC in relation to NetWorker.

There were 211 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.)
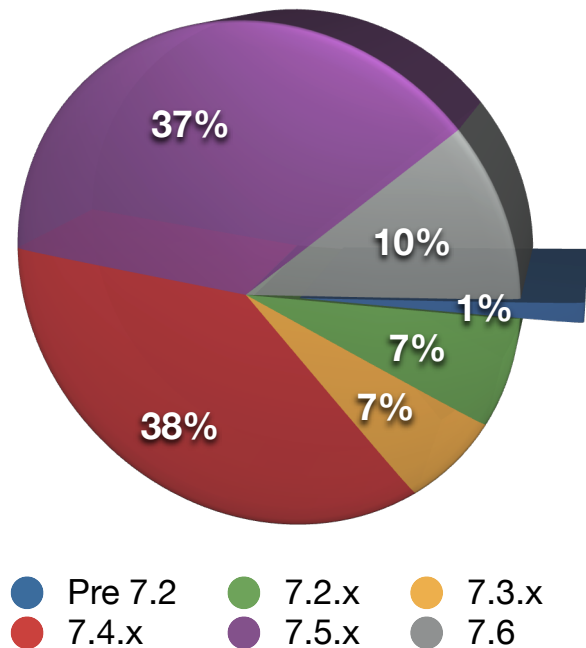
Much gratitude is owed to all respondents, but particularly to those who filled out additional details in the free-form comment section of the survey, regardless of whether those comments were further clarification of their environment(s), or to express thanks for the NetWorker Blog.

## What version of NetWorker are your server(s) running?

Anticipating multiple data zones, this question allowed multiple answers. A total of 258 responses were received to this question.

The numbers for the responses were as follows:

- **Pre 7.2** – 3
- **7.2.x** – 18
- **7.3.x** – 17
- **7.4.x** – 98
- **7.5.x** – 96
- **7.6** – 26



Pie chart showing: 37%, 10%, 1%, 7%, 7%, 38%

Legend:
- Pre 7.2
- 7.2.x
- 7.3.x
- 7.4.x
- 7.5.x
- 7.6

*Comments and Conclusions*

Based on the low "Pre 7.2" responses, it would appear that the use of significantly older versions of NetWorker is quite low – or, users of legacy products are in a "maintenance only" cycle that reduces the need for them to be involved in online NetWorker communities.

It is actually somewhat lucky that the survey could be conducted at a point where a new release of NetWorker (v7.6) has been out for a while, but there has not yet been a service pack release. The advantage of this scenario is that it allows us to draw fairly accurate conclusions of adoption rates of new releases.

Therefore, it is very interesting to note that the percentage of NetWorker 7.6 users is still eclipsed by the percentage of NetWorker users running with unsupported software (versions 7.3 and lower). It should be noted that in the free-form comments section, 5 respondents indicated that they planned to upgrade from various versions of NetWorker to v7.6 "in the next few weeks".

From the survey results as well as anecdotal evidence from online forums such as the NetWorker Mailing List and feedback on the NetWorker Blog, a few conclusions can be drawn:

i.   Overall, NetWorker administrators are cautious about moving to newer versions of the software once existing versions are "running smoothly".

ii.  EMC product management could perhaps speed up the transition to newer versions of the product by aiming to release a service pack within short order of the initial release. The majority of administrators tend to talk about upgrading "after the first service pack". It would seem sensible then that any "service pack 1" for a new release of NetWorker should focus on bug fixes only, with an aggressive release schedule (e.g., within 3 months of initial product release). It is likely that with this strategy, the number of early adopters would increase substantially.

iii. 7.3.x was a reasonably unpopular version. Bearing in mind this was the first version of the product that introduced *nsrjobd*, many users were initially reluctant to upgrade to it at all, and more remain on

7.2.x/earlier than do on 7.3.x. While one might claim that users would have been upgrading to 7.4.x or 7.5.x, this is not born out by the comparable number of users remaining on 7.2.x.

iv. Some EMC account managers have had a somewhat dubious reputation for histrionically pushing customers towards the "latest and greatest" releases for NetWorker as soon as they're released. It would seem the lure of new features attracts only a small percentage (which is fairly typical of the "new adopters" segment of a product adoption curve) to immediately upgrade.

One might claim that the low adoption rate for new NetWorker versions points to EMC having poor quality controls, etc. (This would be a fast-FUD claim made by a competitor reading this report, for instance.) However, if one considers the percentages of users against the classic product adoption curve, a very different perspective comes out. The product adoption curve resembles the following:



| Innovators | Early Adopters | Early Majority | Late Majority | Laggards | Extreme Laggard |
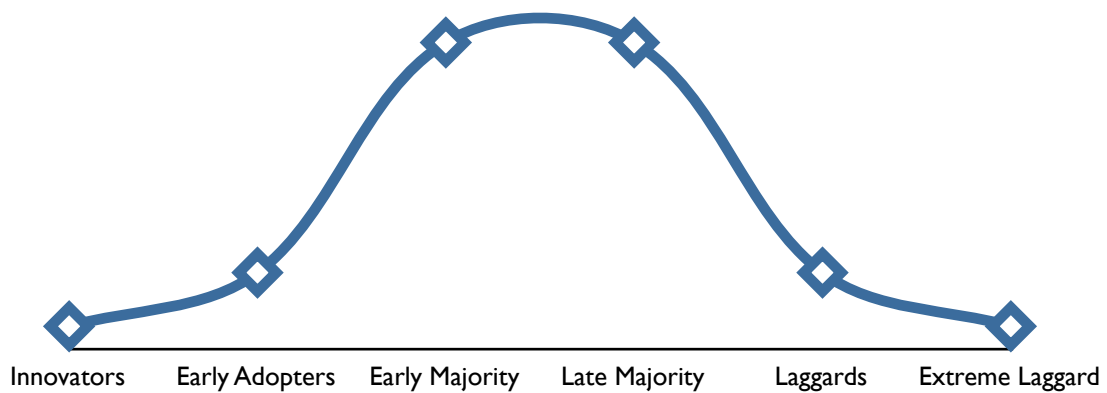
If we consider this scale against the NetWorker usage plotted as a curve as well, we see:



| 7.6 | 7.5.x | 7.4.x | 7.3.x | 7.2.x | Pre 7.2 |
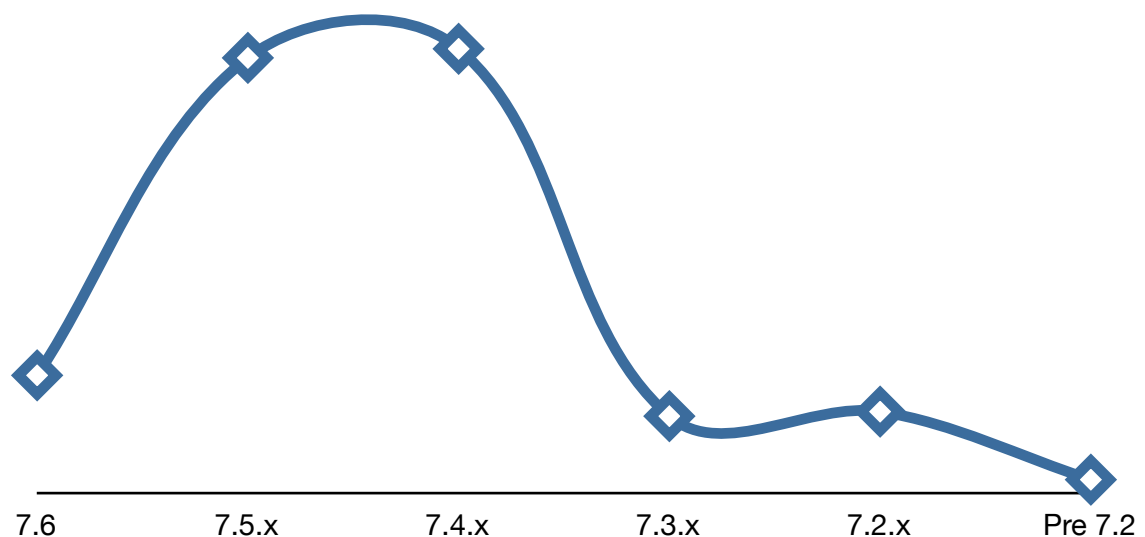
While not a perfect fit, the similarities between the product adoption curve and NetWorker's primary version usage cannot be denied. This adds further credence that EMC would be well advised ensure that Service Pack 1 for any new release is (a) focused on bug fixes and stability enhancements, and (b) released within or around 3 months of the initial version release. Doing so would encourage more users to make that jump from innovators/early adopters to the early or late majority sooner.

## What operating system is used for your NetWorker server?

There were 239 responses to this, and they reflected that a vast majority of NetWorker users stick to three primary platforms for running the actual server from:

The numbers for the responses were as follows:

- **Linux** – 52
- **Solaris** – 103
- **Windows** – 69
- **AIX** – 9
- **HP/UX** – 5
- **Tru64** – 1

29%

4%

2% 0%

43%

22%

● Linux ● Solaris ● Windows
● AIX ● HP/UX ● Tru64
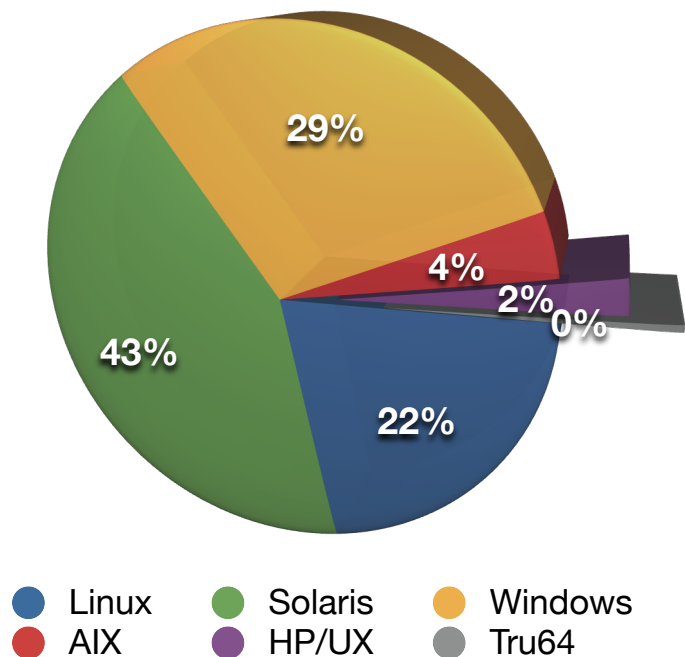
*Comments and Conclusions*

While Solaris remains the dominant platform for running a NetWorker server from, both Linux and Windows have the lions' share of the remaining amount – and together eclipse the Solaris usage. We can draw the following conclusions:

i.   There is a strong shift *away* from legacy Unix platforms to those operating systems that are traditionally seen as either cheaper (Linux) or easier (Windows) to manage. (The merits of these general attitudes towards Linux and Windows respectively are beyond the scope of this report.)

ii.  While it's unlikely in the short to medium term that lesser used Unix platforms such as AIX and HP/ UX will see a lessening of support, we can draw the conclusion that the entrance of any new operating system at server levels will be quite difficult due to the current high use of Windows, Linux and Solaris.

iii. From a personal experience, these results suggest that EMC need to concentrate more on quality control for Linux. It has frequently seemed the case that the Linux platform has played poor cousin to traditional Unix and Windows in terms of testing and fixes – this should be addressed, and quickly.

iv.  EMC remains a strong heterogenous backup system by offering a broad selection of operating systems for the server platform. Other backup products that are limited in the server operating system/architecture effectively self-limit themselves to much smaller market segments.

## Storage Node Versions

With 197 responses to this, it's clear there are a fair number of NetWorker sites that are actively running storage nodes. The distributed, 3-tier approach to backups offered by various enterprise architectures, including NetWorker, is considerably more advanced than 2-tier "server/client only" model seen in many workgroup products.

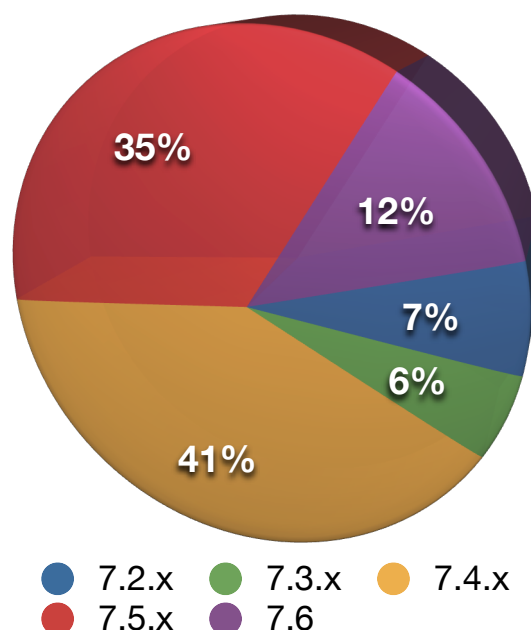The numbers for the responses were as follows:

- **Pre 7.2** – None
- **7.2.x** – 13
- **7.3.x** – 11
- **7.4.x** – 80
- **7.5.x** – 69
- **7.6** – 24



*Comments and Conclusions*

The survey perhaps should have included a question to ask whether storage node versions were different to the NetWorker server versions, which makes drawing any firm conclusions from this data challenging.
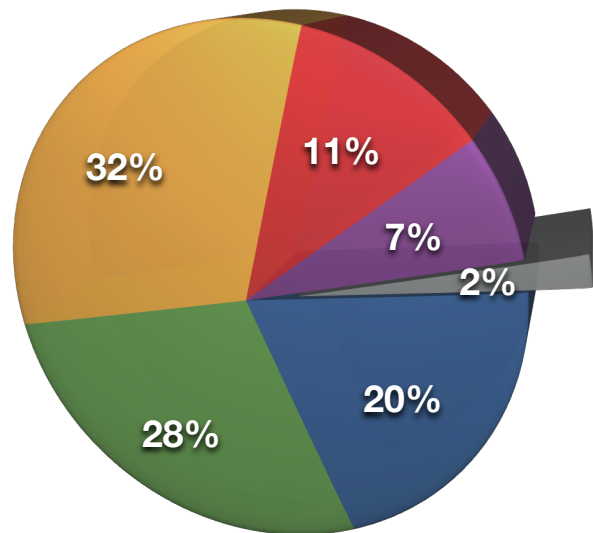
However, it would appear that there is a reasonably good correlation (at the percentage level) between storage node and server versions, suggesting that most users are likely to be running reasonably similar if not the versions of NetWorker on their storage nodes as they do on their servers. I.e., current recommended best practices for NetWorker datazones do *appear* to be mostly followed in relation to storage node and server versions.

## Storage Node Operating Systems

With 274 total responses, we see a broader distribution in storage node operating system types than we do in backup server operating system types.

The numbers for the responses were as follows:

- **Linux** – 55
- **Solaris** – 78
- **Windows** – 87
- **AIX** – 30
- **HP/UX** – 19
- **Tru64** – 5



*Comments and Conclusions*

While there is a majority 3-way share of operating systems for NetWorker servers, the distribution of operating systems for storage nodes is considerably more broad.

A future survey will aim to discover the correlation between storage node operating systems and whether the storage nodes are full nodes or dedicated storage nodes, as it would be interesting to see how many storage nodes in this broader OS selection are provisioned just for self-backup.

While Linux, Solaris and Windows still retain the lions share of operating system use in the storage node component, HP/UX, AIX and even Tru64 share a broader use as storage nodes.

With a broader distribution of operating systems at the storage node level, one would hope that EMC are at least prepared to consider to add new operating systems/platforms into the supported list for this tier in the backup environment when sufficient justification is present. (Using Mac OS X as an example, many sites that use the Mac OS X client have a considerable amount of data on that platform and would prefer to use the same platform as their storage node. Currently the number of sites using the Mac OS X client however would not seem to – as yet – justify calls for building a storage node.)

Regardless, it's clear that EMC's continued support for a wide range of operating systems at the storage node level is a product strength; by allowing sites to use preferred or most-suitable operating systems for their storage nodes, datazone configurations remain quite flexible.

## Number of protected clients

This was a single-choice question, which means there were exactly 211 responses, as per the number of survey responses. This reflected the total number of clients backed up across all datazones within the environment.

The numbers for the responses were as follows:

- **Clients < 10** – 9
- **10 >= Clients < 50** – 21
- **50 > Clients <= 100** – 31
- **100 > Clients <= 500** – 84
- **> 500 Clients** – 66

31%

4%

10%

15%

40%
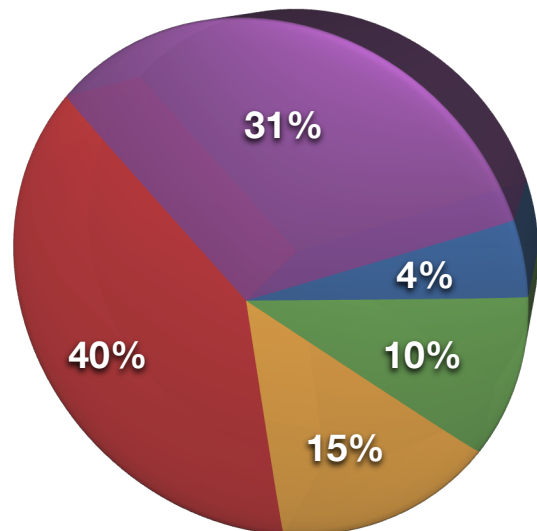
● <10          ● >=10 and <50
● >50 and <=100   ● >100 and <=500
● >500

*Comments and Conclusions*

Several respondents used comment/feedback fields to indicate that they backup well over 1000 or 1500 clients within their environments, indicating that a follow-up survey should provide more options in the larger number ranges. (One respondent indicated a total of 90 datazones protecting over 10,000 clients and 1.5PB of data weekly.)

The clearest conclusion to draw from this is that despite FUD spread by competitors, EMC NetWorker is *not* a small-environment backup product. If we look at sites that protect 100 or more hosts, that represents 71% of installs. With 31% of those installs protecting more than 500 clients, it sends a clear message that in the enterprise backup space NetWorker more than holds its own in terms of numbers of hosts protected.

The low number of sites where less than 10 clients are protected could be equally attributed to one of two reasons (or indeed, both). First, NetWorker's price is not really geared towards workgroup protection. Secondly, it may equally be argued that sites with smaller numbers of clients have less involvement in social networking activities associated with NetWorker due to minimum complexity within their environments.

While EMC may argue that Retrospect offers a better priced workgroup backup model than NetWorker, it is the author's gut feeling is that it will not offer true competition in the workgroup space until a clear upgrade path from Retrospect to NetWorker can be offered. In particular, NetWorker's ongoing inability to recover Retrospect data (or direct a Retrospect server as a "recover node" as was setup for EDM) must be acting as a limiting factor in Retrospect take up.

With such a large percentage of sites backing up 100 or more clients, it also suggests that EMC need to treat the performance tuning guide (finally updated in NetWorker 7.6) as a living document, keeping it as up to date as possible to provide fresh and new advice as new technologies come into play. This will help companies with larger numbers of clients stay appraised of best practices approaches to keeping their datazones running at peak efficiency.

## Client operating systems

With 751 total responses to this question, it's clear that there are a wide variety of client operating systems being protected by NetWorker.

The numbers for the responses were as follows:

- **Linux** – 174
- **Solaris** – 155
- **Windows** – 209
- **AIX** – 76
- **HP/UX** – 50
- **Tru64** – 17
- **NetWare** – 23
- **OpenVMS** – 19
- **Mac OS X** – 29
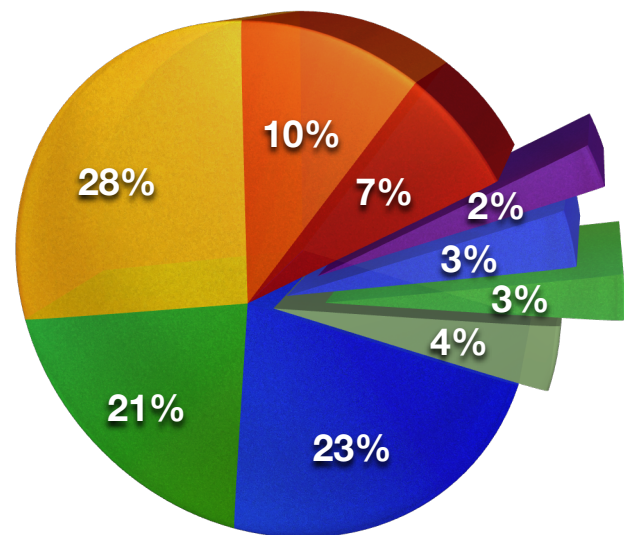


- Linux
- Solaris
- Windows
- AIX
- HP/UX
- Tru64
- Netware
- OpenVMS
- Mac OS X

*Comments and Conclusions*

The most pertinent comment to make on this front is: if you build it, they will come. Overall most enterprises are keen to deploy as few backup products/solutions as possible – while at the IT staffing perspective centralisation vs decentralisation remains an eternal pendulum, at the backup perspective the battle was won years ago.

One of the things we can draw from anecdotal evidence when combined with the survey results is that with the high percentage of Linux use within NetWorker, EMC needs to expand the support for different distributions – and in particular, offer support for the Debian style distributions that use a different package management system than the RPM-based platform currently supported. From personal as well as anecdotal experience, many sites that use NetWorker and Linux will have a mix of Linux based package management platforms, with the majority shared between the RedHat/SuSE ".rpm" approach and the Debian ".deb" approach. These latter companies have to choose between manually extracting and installing package contents, running the NetWorker client on "unsupported" systems, and picking alternate backup methodologies for those platforms. Neither choice is palatable.

One might also add that given the relative similarities between Mac OS X and BSD, it seems odd that EMC do not at least officially support BSD on the x86/x86_64 platform. Again, the broad distribution of clients suggests that if client versions are made available for new operating systems, customers will be found for them. (Indeed, the creation of a Linux/PPC client several years ago would seem to back this statement up.)

## Modules in Use

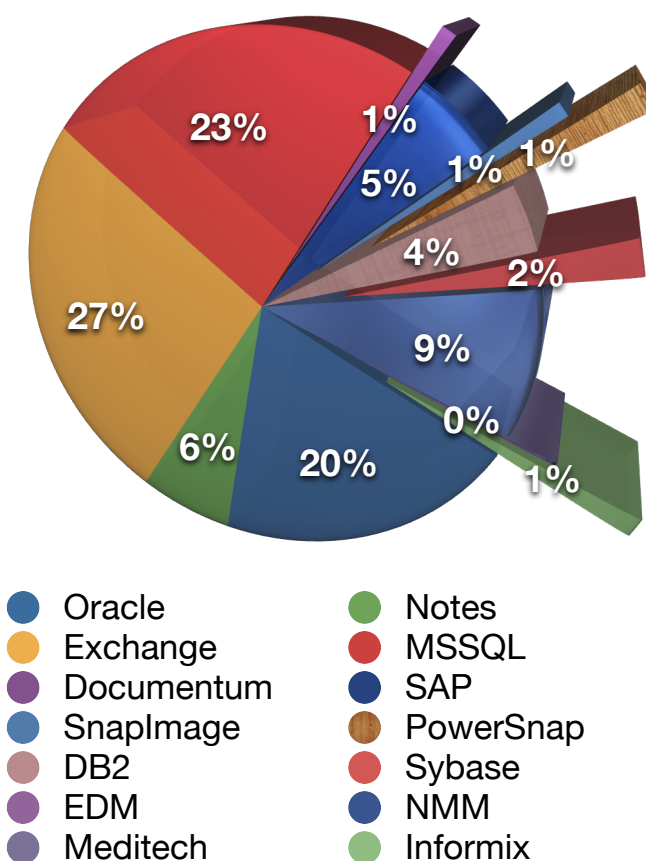With 481 responses, this was aimed at not determining actual databases/mail servers/etc in use, but to find out what the layout of actual *module* use for those products were.

The numbers for the responses were as follows:

- **Oracle** – 95
- **Notes** – 30
- **Exchange** – 128
- **MSSQL** – 111
- **Documentum** – 4
- **SAP on Oracle** – 23
- **SnapImage** – 4
- **PowerSnap** – 7
- **DB2** – 18
- **Sybase** – 10
- **EDM** – 0
- **NMM** – 44
- **Meditech** – 1
- **Informix** – 6



Legend:
- Oracle
- Notes
- Exchange
- MSSQL
- Documentum
- SAP
- SnapImage
- PowerSnap
- DB2
- Sybase
- EDM
- NMM
- Meditech
- Informix

*Comments and Conclusions*

The 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. (The comparative low use of the NetWorker Module for Lotus Notes reflects both the decreasing popularity of that platform *as well as* the ongoing misconception that Notes can be safely backed up without a module.)

In certain cases, limited use of modules will more reflect limited capabilities of those modules rather than limited need. For instance, the SnapImage module should be frequently used to counter issues with dense filesystems, yet remains largely stagnant in terms of development, operating system support and capabilities.

Over time one expects to see the use of dedicated Exchange and MSSQL modules to reduce, with comparable growth in their replacement, the NetWorker Module for Microsoft Applications (NMM). It's still too early to factor in use of the NetWorker Module for Databases and Applications.

## Open Source Database Usage

Several times in the last few years the author has discussed the continuing lack of support for open source databases within NetWorker with both EMC product and account management. One hopes the 163 responses received in this survey gives EMC some renewed incentive on this front.
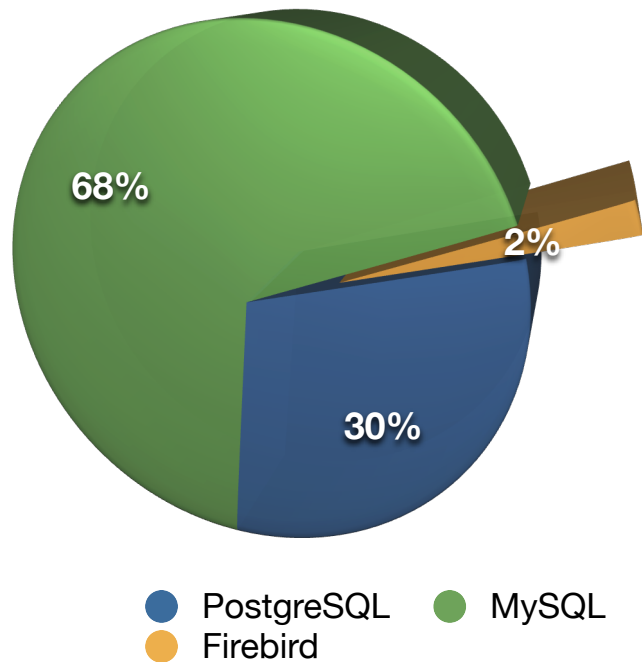
The numbers for the responses were as follows:

- **PostgreSQL** – 49
- **MySQL** – 111
- **Firebird** – 3

(While room was provided for other open source databases to be specified, none were.)



*Comments and Conclusions*

When we consider the 163 declared users of open source databases versus 211 survey responses (i.e., 77% of respondents), there are a few inevitable conclusions that must be drawn:

i.   EMC are continuing to put themselves at a competitive disadvantage by not offering modules for MySQL and PostgreSQL.

ii.  While there are other open source databases (e.g,. Firebird), there is currently insufficient usage patterns within the surveyed NetWorker user community to consider modules outside of MySQL and PostgreSQL.

iii. While MySQL usage outnumbers PostgreSQL at slightly more than 2:1, the percentage of PostgreSQL use still justifies consideration of a PostgreSQL module.

iv.  While it is unlikely Open Source database users will be willing to pay comparable prices for a backup module for their products as say, Oracle or Microsoft SQL users, an appropriate price structure could be developed by EMC.

v.   An oft-used complaint by many of the enterprise backup companies for why Open Source databases are not supported is a lack of commercial support/assistance available. Given there are companies that offer commercialised versions of both packages, this seems a dubious argument.

Anecdotally, companies are quite keen to see integrated open source database backup solutions within their enterprise backup strategy. Companies such as Zmanda, who offer a MySQL backup utility, attempt to fill this void, but struggle to offer truly a holistic user experience. It is time for larger data protection companies to step up to the plate and recognise this emerging market.

## Miscellaneous Options

A "catch all" question for miscellaneous options within NetWorker received 317 answers – indication that NetWorker is frequently used to extend beyond standard backup options. It is pertinent to consider this data set in respect to the number of survey responses – i.e., 211 filled in surveys. (Obviously more than one selection could be made by the respondents.)

Examined from this perspective, common options for NetWorker are as follows:
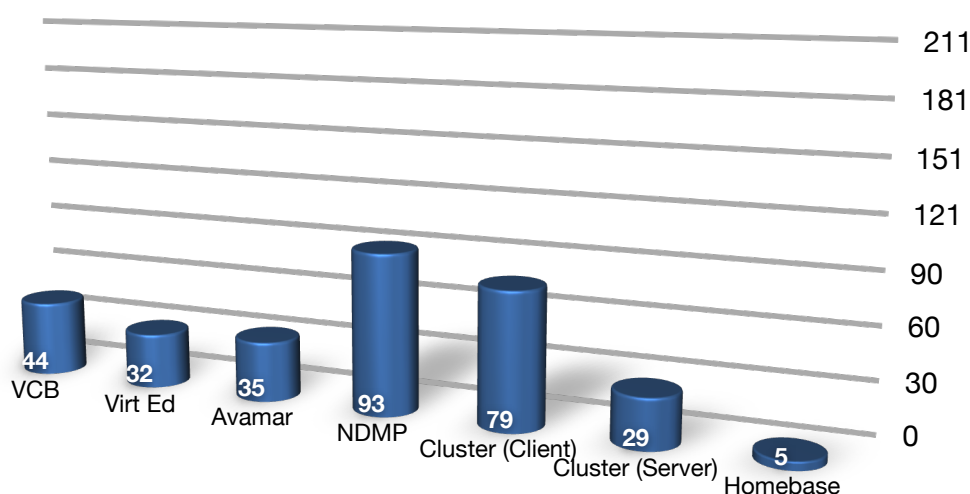


Expressed as a percentage:

- **VCB Usage** – 21% of respondents

- **Virtual Edition Client** – 15% of respondents

- **Avamar** – 17% of respondents

- **NDMP** – 44% of respondents

- **Cluster (Client)** – 37% of respondents

- **Cluster (Server)** – 14% of respondents

- **Homebase** – 2% of respondents

*Comments and Conclusions*

The first comment is regarding survey clarity. It seems odd that such a relative high number of respondents would be clustering their backup server, which was what the "Cluster (Server)" option was meant to cite. There is some probability that due to a lack of explanation, some survey users may have been confused on these terms. With this in mind, no distinction will be made between the two clustering options for the conclusions.

Based on the survey results, we can draw the following conclusions:

i.   The NDMP option by far and away is the most popular one surveyed in this section. Given the tiny usage patterns for the SnapImage module, we can conclude that in all likelihood this option is being used for backup of NAS systems. This indicates that NAS remains popular (or has been growing in popularity) to assist the management of larger amounts of non-direct storage.

ii.  The relatively low VCB usage is not surprising; it is likely this would represent a percentage of VMware environments deployed rather than all VMware environments deployed. The extremely limited support in VMware's VCB functionality for non-Windows guests has not only been disappointing, but has limited the intention of many sites from deploying VCB generally – i.e., why deploy a technology that only works half of the time? (It will be interesting to revisit VMware backup options around 3-6 months after NetWorker includes support for the vSphere API.)

iii.  The Avamar adoption rate indicates that source based deduplication is getting some attention within the NetWorker datazone. However, its penetration is not as high as some pundits would suggest. This perhaps as much as anything derives from companies being (by and large) reluctant to introduce *any* change into the backup processes for individual clients. As EMC continues the integration of Data Domain within their product line, source based deduplication will likely suffer a continued slow adoption rate due to the competition from alternatives that do not require end-client changes.

iv.  Homebase adoption has been poor at best. While anecdotally many users have clamored for BMR support for years, it seems that business *desire* for support exceeds business *intention* to adopt available solutions. It might also be concluded that by the time that homebase support was added, most existing NetWorker deployments had either found alternate solutions, or decided that BMR was not a driving factor within their environments.

## In Conclusion

NetWorker's ability to operate within a heterogenous environment remains one of its core strengths. The ability to support a plethora of operating systems, databases and applications allows it to slot into the requirements of a large number of businesses.

When we evaluate the overall survey responses along with previous benchmarks set by NetWorker for backup and recovery performance, as well as key data protection design strategies, it seems clear that there is no credibility to periodic competitive statements which accuse the product of not being a "real" enterprise data protection product. Indeed, another core strength for NetWorker is that ability to readily scale from the smallest to the largest of environments.

From a historical perspective, the long-term decision by EMC to acquire Legato has provided the company with an almost-complete vertical solution for storage, virtualisation and data protection. (The notable exception of course is an owned tape solution, but given EMC's history and overall outsider perceived opinion of tape, this seems unlikely to change.)