

HIGHLIGHTS

- » **NetBackup/NetApp Disk Integration**
- » **6 Tips for Archive/Compliance**
- » **Simplified Sizing for Databases**
- » **FAS6000 Virtual Tour**
- » **Win Free BMW ORACLE Gear**

Winning by a Fraction of a Second: BMW ORACLE Racing

Ian Burns, Design Coordinator, BMW ORACLE Racing

BMW ORACLE gains a competitive edge with the ultimate database storage machine.



[More »](#)

BLOGGING WITH DAVE

Dave Hitz, NetApp Founder and EVP

"Last quarter was a milestone for NetApp. We passed EMC in **TBs** shipped to become the largest *networked storage* vendor."

» [Dave's Blog](#)

DRILL DOWN

» **Virtual Tour: FAS6000 Series**

Rotate the controller, zoom in for detail, and look under the covers with this 3D model.



» **Multiprotocol Use Guide**

How native support bridges UNIX and Windows file security styles.

» **Voice Your Opinion**

Design your ideal professional services.



ADMIN RESOURCES

» **Free Tool: Angry IP Scanner**

In seconds identify online devices, open ports, host names, rouge entries and more.

TIPS FROM THE TRENCHES

Six Tips for Archive/Compliance

Mike Riley, Global Systems Engineer, NetApp

What would you do if someone slid a 5¼" disk across the table and asked you to open up the WordStar document on it? Check out six ways to ensure the decisions you make today don't limit your future options.

[More »](#)

Sizing Storage for Oracle Databases: Size Matters (But I/O Matters Too!)

Aaron Sims, Sr. Professional Services Consultant, NetApp

Advice on how to work with NetApp to efficiently size your storage system including sample queries, tips for using Oracle STATSPACK, and more.

[More »](#)

ENGINEERING TALK

Whiteboarding Session: Veritas® NetBackup™/NetApp Disk Integration

Stephen Manley, Technical Director, NetApp

Bill Roth, Technical Product Manager, Symantec

An in-depth look at the integration points between NetBackup and NetApp disk and how functionality is evolving in new releases.

Plus, see how data reduction using single instance storage works to eliminate redundant block data.



[More »](#)



IAN "FRESH" BURNS

Design coordinator, BMW ORACLE Racing

In an abiding quest to capture the coveted America's Cup sailing title and trophy for 2007, native Australian Ian Burns has grown accustomed to moving himself and his family to the latest international hub of yacht racing design and testing. Tech OnTap caught up with Ian in Valencia, Spain where long days of intensive sailing practice are punctuated by equally long and intensive scrutiny of computer models, real-time statistics, and streaming video footage. In a world in which fractions of a second mean the difference between winning and losing, technology plays an increasingly vital role in the team's efforts to achieve the "ultimate win."

Winning by a Fraction of a Second:

BMW ORACLE Gains the Competitive Edge

Q&A with Ian Burns

Q: Congratulations on winning the first two races of the 2006/2007 World Match Racing Tour, the Portugal Match Cup and the German Sailing Grand Prix. What's it like to win a match by no more than half a boat length, as the team did in Portugal?

Ian: This is what the America's Cup is all about. It's a game played in meters, a game of tiny fractions of a percent of advantage. All of the racing teams already operate at an incredibly high standard of performance, so to win you have to get performance from 99.5% to 99.6% in some critical area.

In the corporate world, if people were to take their company from 100th place in their industry to 10th place, some might say they were doing great. In our event, second place is an abject failure. There's only one winner, so every decision we make is based on whether or not it will give us an advantage over our competition that we can exploit.

Q: Much of your team's R&D work involves generating and analyzing huge amounts of information. How has this aspect of the race changed since the team's first run for the America's Cup in 1995?

Ian: The contrast is really quite amazing. We spend a lot of our time doing research that results in gigabytes and gigabytes of data that all attempt to answer one basic question: "Does Boat A beat Boat B, or vice versa?"

In our last run for the America's Cup, we could run only about one to two computer models per month. Sometimes it might take several months to run a particularly complex model.

Today, we've achieved a lot more with the scale of technology generally available, not to mention the use of Oracle® databases and NetApp storage systems. Whereas before we could only run a couple computer models a month, we now run much more sophisticated models every two days, or even every day. Virtually any team member can start with raw data and process it for immediate results without needing the help of a massive team of analysts.

We've stepped up to maybe 10 to 15 times the data production speed we were at before, plus the actual size of the files and the models has doubled or tripled. So you're talking about 100 to even 1,000 times faster data production than we had during the last America's Cup.



The Power of One Team
(2:24 min video)

RELATED INFORMATION

- [The Power of One Team: BMW ORACLE Racing, Oracle, and NetApp \(2:24 min video\)](#)
- [Improving Agility for Database Development, Test, and QA](#)
- [Win a FREE Racing Team Jacket](#)

Divining between Art and Science in International Yacht Racing

The BMW ORACLE Racing Team logs over 250 variables each second during typical tests on the water. The team on shore then collects the data in real-time and begins the task of analyzing it for clues to help boost sailing performance by even a fraction.

The testing payload on the boat consists of measuring equipment for any number of variables, including:

- Water pressure
- Hydraulic pressure
- Strain on the mast
- Load on the sails

Hear Ian explain how the team uses data to [design a better yacht](#). (3 minutes)



Q: How do you manage the sheer volume of data generated as the team produces models and analyzes data as much as 1,000 times faster than in previous races?

Ian: We always struggled with the scale of our data. When we try something new, the data doesn't double, it goes up by the square, 10 times, or even 100 times. We were forever archiving things and could never fit everything on disk storage.

NetApp lets us scale easily. We just buy another shelf of disk when we need to, which makes it a lot better than running into hardware walls. When you couple that with the scalability of the Oracle system with clustered technology and a grid that protects you when something goes down, it's a great combination that has let us pull up huge files from five years ago and compare them, virtually instantaneously, to what we did a few seconds ago.

Q: Are there areas beyond R&D in which the team has leveraged technology to improve its overall performance?

Ian: Absolutely. In the last Cup, we handed out computer printouts at meetings to discuss our findings. Today we have access to simultaneous videos, data traces, different views of the sails, and more, all integrated into our performance analysis meeting. We can fast-forward or zoom through the data in a huge video or data file very quickly.

This puts a major load on the Oracle database, servers, and storage. We have had to invest in our infrastructure so that it could handle this load. We have a Fibre Channel connection to our database servers so that we can get the video files ripped off the NetApp systems as fast as possible. The bandwidth is breathtaking in terms of how fast the database can pull the video files up and serve them to multiple users at one time, even when we double, triple, or quadruple the load.

We've also benefited from the integrated nature of NetApp and Oracle solutions. Oracle natively accesses NetApp in the fastest possible method you can rip data off a disk. That level of integration is very powerful, since we have none of the bottlenecks or layers of operating systems on top of file systems you might otherwise expect.

We went from working almost entirely on paper before to now having huge archives of video, data, and still photography, all time-synched and displayed in real time. The change has been dramatic.

Q: The America's Cup is a global event with design and development happening across multiple sites and races around the world. How does this factor in?

Ian: In our previous run for the Cup, we had geographically dispersed offices much like we do now. We would find ourselves running into problems when we tried to share and access files between the sites. Sometimes we had a kind of server you had to get to from a remote site. And with the vagaries of the Internet in Spain, sometimes this could be okay and sometimes there would not be enough bandwidth to grab a decent-sized file and pull it down the line.

Now we use NetApp SnapMirror® software to replicate files between the sites. When a file is changed in our Seattle or Valencia location, it begins spreading itself all around to our different sites, on our different filers around the world. We take a backup of those systems locally with NetApp Snapshot™ and we actually have the same version replicated nightly to the other sites via SnapMirror. This gives us the opportunity to roll back in multiple places around the world. We now have backups of the exact same file structure and file setup in three to four structures around the world.

This is an added benefit of the system, and it makes us confident of having the data protected and easily recovered in the event something unforeseen happens at one of the sites.

Q: Your architecture has changed significantly in the past five years. What kind of impact has this had on the team from a support perspective?



NetApp vs. EMC and HP: IT Decision-Maker Perspectives

After interviewing storage managers and DBAs at more than 20 European and North American enterprises, Mercer Management Consulting identified key differentiators between database environments using NetApp, EMC, and HP storage.

Findings include:

- In a typical FC SAN environment, NetApp solutions cost less than EMC CLARiiON (32%), HP EVA (42%), and EMC Symmetrix (62%).
- NetApp environments require less primary disk capacity (52%) and data protection disk capacity (20%) than EMC and HP environments.
- Database storage administrators using NetApp storage can recover twice as fast from a database crash or corruption.

Get the details. [Read the full report.](#)

Sailing Lore and the Early Quest for the America's Cup

The America's Cup challenge dates back to the first race between England's Team Aurora vs. Team America in 1851. The trophy was called, at the time, the Royal Yacht Squadron's "100 Guinea Cup."

Shortly after Schooner "America" won the 100 Guinea Cup, the original team owners donated the winning cup to the New York Yacht Club under a "Deed of Gift" that indicated the trophy should be used as "a perpetual challenge cup for friendly competition between nations."

Following a steep set of rules developed by the Yacht Club representing the current defender of the trophy, all challengers are required to test their sailing prowess in a series of pre-qualifying matches prior to the ultimate race. At the end, only one challenger alone is deemed fit enough to meet the defender on the water for the final, decisive sailing match to determine the next winner.

[Visit the official America's Cup site.](#)

Ian: For a group like ours, which doesn't invest heavily in IT staff, we rely on systems that are based on great technology but also easy to use and maintain. With a NetApp system, you have a full understanding of how it works after just reading a one-page instruction set. These things are so straightforward I could probably train my kids to set them up and they'd have them up and running in just a few hours.

The simplicity of the system with a small understanding allows us to deal with it ourselves on a daily basis, in just 5 or 10 minutes a week. This fits our model: a relatively low spend on actual technician cost and a maximum spend on the absolute best technology we can buy.

We've also had good support from NetApp when we need it. Valencia is not the technical center of the world, but we've had technicians on call once or twice and they've been here in almost no time.



Q: Are there areas in which your close ties with Oracle and NetApp have provided an advantage over your competitors?

Ian: Our competitors use similar ideas and technologies, but without the depth of technical ability or application we have by our association with Oracle and NetApp. Our expertise with this type of technology allows us to "turn the crank" and do more iterations of design. We can test more, try more, and come up with a better solution than our competitors.

I also believe our association with tech suppliers like Oracle and NetApp has led us to go to a level higher than what we would have thought was adequate if we'd been sitting in a sales office buying the cheapest product or what we thought was good at the time. These types of relationships are incredibly useful to us when we need every fraction of a percentage gain. Quite often small gains in individual areas quickly add up to large gains in the total product.





MIKE RILEY

Global Systems Engineer, Network Appliance

For the past seven years Mike has worked in various capacities as a NetApp systems engineer. He currently works out of the NetApp Global Engineers team focused on data protection and retention solutions. In this role, Mike works with customers from Des Moines, Idaho, to Tel Aviv, Israel, to help protect their critical data. Mike also received a recent promotion: he is now defensive coordinator for

the St. Raphael Football Giants, where his son lines up as the right defensive end.

Six Tips for Archive and Compliance Planning

By Mike Riley

What would you do if someone slid a 5¼" disk across the table and asked you to open up the WordStar document on it? I realize I'm addressing the Tech OnTap audience, but let's face it: those of us who even *remember* WordStar are an ever-shrinking minority. The irony is that WordStar only breathed its last gasps as a commercial technology in 1998 or so—barely eight years ago.

The point is that there's a lot more to long-term data storage than just squirreling away data. You've got to think about how the data is stored, what format the data is stored in, and what it will take to read it back. Time works against you when you're talking about storing data for three, five, seven, 30, or more years.

I've spoken with my fair share of NetApp customers about archive and compliance over the past seven years, and the main advice I give them is to stay nimble. "Be nimble" ranks right up there with "don't worry; be happy" when it comes to technical advice, but the point is that you don't want to make decisions now that lock you into proprietary technologies or limit your future options.

Here are six things to keep in mind as you design an infrastructure to cope with long-term data storage for archive and compliance.

Tip 1: Avoid Storage Silos

Long-term archival is largely driven by rapidly growing volumes of e-mail and unstructured files, while compliance must satisfy external regulatory requirements and enable litigation support and e-discovery. Whether you're dealing with archive or compliance planning, many of the storage infrastructure requirements are the same.

In fact, from a storage perspective there are really only two critical differentiators:

- Archive is driven by a need for mailbox management. This typically involves leveraging tiered storage to move e-mail from the primary e-mail server storage to lower-cost secondary storage.
- Compliance adds an absolute requirement for data permanence that can cast every misstep, every wrong decision, and every virus in stone for years to come.

Because the underlying storage requirements are similar, wherever possible you should try to meet long-term storage requirements with the same basic infrastructure. Deploying separate silos of storage for archive and compliance distinct from your enterprise storage adds management complexity and limits your ability to adapt flexibly to future needs.

Don't make decisions now that limit your future options.

RELATED INFORMATION

- [VeriTest Performance Comparison: NetApp FAS3050, R200, and EMC Centera G4 \(pdf\)](#)
- [NetApp Archive and Compliance Solutions](#)

Head-to-Head Performance Comparison: NetApp vs. EMC

VeriTest is the leading third-party independent source for competitive analysis and performance benchmarks. The company provides benchmark results for companies like Cisco, IBM, and Microsoft and is the exclusive testing provider for the EMC Centera Certification program.

Recently VeriTest studied archival performance using Symantec® Enterprise Vault on three platforms:

- NetApp NearStore® R200
- NetApp FAS3050
- 16-node EMC Centera G4

The results of a multithreaded test showed that as the archival load was increased, the FAS3050 scaled 42% higher than the Centera and delivered 38% better performance. The FAS3050 also showed little or no performance impact during single-drive failure, a double-drive failure, and when running a discovery request.

Get the details. [Read the VeriTest report.](#)

Tip 2: Avoid Proprietary Data Formats and Technologies

Think about that WordStar document again for a minute. It's bad enough that it was written by an application that hasn't been active for at least eight years, but what about that 5¼" floppy? What format is that in? Where can you find a drive that can read it? How can you get the file onto some medium where you can even begin to work with it?



I'm not suggesting that the 5¼" floppy was ever considered a good medium for archival storage, but it's far from the only storage device that ever passed quickly from the mainstream into obsolescence. When you evaluate storage solutions for archive and compliance you need to focus on open, standards-based solutions.

Here are a few critical questions to ask:

1. **What format is used to store the data?** You want your file data to be stored in its native format as much as possible. The more software layers that separate you from your data, the more trouble you could have in the future.
2. **How difficult will it be to migrate to new storage?** Let's face it; if you're going to store data for years, you're eventually going to have to migrate to new storage. If your data is stored in a proprietary format, it may make future migration more complicated.
3. **What access protocols are supported?** If you can only access your data using one protocol, you're locked in. You want the option of using multiple standard protocols.
4. **What archive and compliance applications are supported, and how are they integrated?** A good long-term solution must support the major applications for e-mail archival, compliance, and so on. Beyond that, how are the applications integrated? Does the storage system have proprietary APIs that the application must follow, or does it use standard access protocols? Proprietary APIs make it harder for application vendors to support a particular platform.

In an enterprise setting where by law you must lock down data for years, the best way to preserve flexibility is to keep the data in its original format as much as possible and make it accessible using standard protocols. As much as security architectures such as Active Directory have evolved over the past five years, I also suggest that the compliance solution slipstream natively into that security model. This gives you file-level control at the individual user level. The more rigid your solution, the more time becomes your enemy.

The more rigid your solution, the more time becomes your enemy.

At this point you may be thinking, "Of course NetApp would say this is the best solution because it's what NetApp offers." In any evaluation, you want to pull the vendor tag off of the solution and look at the data. We're not reading tea leaves, nor are we splitting the atom. Eliminating the vendor religion from the discussion, the idea in archive and compliance planning is to meet (or exceed) regulatory/corporate governance demands and yet prepare yourself for the fact that laws can change, software evolves (or becomes extinct), and sometimes you change your mind. You want to be able to adapt without radically changing your architecture. If you work from there, the solutions start to fall into place.

Tip 3: Don't Consciously Limit Your Options

There's one thing I can tell you about any technology forecast: it's wrong. How wrong it is will determine the price you pay for following it. When you're dealing with data that must be kept for years, limiting your options from the start is the wrong strategy. Based on your environment and requirements, you might decide not to implement certain features, but that's different than closing doors or leaving off features entirely.

You know what you expect of your storage solutions today—how to back them up; how to replicate them; how to secure and protect them; how they should perform in different business critical situations. How about next year? What about 10 years from now?

Support Your Compliance Technology with the Right Processes

The latest and greatest technology (from any vendor) is not enough to ensure you are in compliance. There are important processes that must be part of any compliance solution. Overlook them, and the auditors may never bother to take a look at your cool technology.

Here are key processes that seem to be getting the most attention right now:

- All physical access and maintenance of the media or system must be audited and recorded and must have preapproval by a control source.
- You should have a security log recording any configuration changes made to your compliance solution.
- You need to be able to provide an auditable topology and workflow report of how data gets to your compliant storage.
- There must be an audit trail for outage remediation. In other words, should you experience an outage, what steps would be taken in order to bring the system back into compliance, and how did you know when you had reached that point?
- There must be a method to verify and audit the quality and accuracy of the data over time.

Another consideration involves the resiliency options you use today and what you might need in the future. If a solution includes ATA disks (which are considered less reliable than Fibre Channel disks), will standard RAID be enough, or is double-parity RAID a necessity? If you need double-parity RAID, can the solution you choose deliver the performance you need now and in the future?

It's easy to rationalize options away, but it's extremely difficult to bolt on features after the fact. Search tools and SRM reporting can be added on quickly with minimal pain, but you may not be so lucky with other features. Don't be too ready to give up the features you've gotten used to in your existing storage environment. A solution that can't be backed up using your current enterprise backup strategy may not make much sense. The same goes for ACLs and directory services.

It's easy to rationalize options away, but it's extremely difficult to bolt on features after the fact.

Tip 4: Plan for Performance

I think it's fair to say that no one in the storage industry thought that performance would be a big factor in disk-based archive and compliance storage. IT departments were mainly replacing tape and optical solutions, so any disk-based solution would be a marked improvement, right? The main consideration was providing inexpensive but reliable storage.

Then the market changed. There was a geometric increase in the number of audit and discovery requests (facilitated by all this data being on a random access device), and the scope expanded beyond e-mail to include home directories, instant messages, digital images, phone records, database records—you name it. Ingestion rates and queries skyrocketed. In effect, we had changed our tape and optical replacement strategy into a data warehouse. Who knew?!

For e-mail archival you have to guarantee performance for a number of simultaneous activities. End users need adequate performance when browsing archived e-mail even when a discovery search is under way. Discovery searches or other activities can't get in the way of ongoing archive processes, and when it comes time to expire and delete messages, the storage system has to support rapid deletion (potentially hundreds of thousands of messages a day) along with everything else.

You can put this under the "I'd rather be lucky than good" category, but having the ability to mix drive types and move up to more powerful controllers for greater horsepower worked in the favor of NetApp customers. Today, we have a number of customers turning to us as an alternative simply because the market changed, performance became an issue, and they needed a new strategy.

Tip 5: Utilize Storage Virtualization

A technology that has gained a lot of interest recently is storage virtualization. The appeal is clear—what's not to like about the idea of having a storage pool of infinite bigness?

When it comes to archive and compliance, look for virtualization technologies that offer you the best of both worlds: a pool of infinite bigness *and* the granular control of file systems.

Having a large storage pool certainly simplifies an archival architecture. This type of storage virtualization makes it easy to add additional storage where it's needed and adjust the size of volumes as necessary to adapt to changing policies. For instance, you may decide that you need to retain archived e-mail for a longer period and thus increase the size of volumes dedicated for that purpose.

Archival is worthless; discovery is priceless.

But—as the saying goes—"archival is worthless; discovery is priceless." Having granular access and control of the data is critical because the meter starts running when a discovery request is issued. If external auditors and lawyers want to see results by the end of the week, your lawyers will want to see that information tomorrow. Would you rather run your initial search through 100TB or 10PB?

Having a granular look at the data can also provide flexibility if you need to streamline backup operations or balance replication schedules. Rather than attempting to cope with one massive "junk drawer" of information, you can focus on levels of granularity that make sense.

Tip 6: Think Carefully about Encryption

Recent compromises of credit card and other personal data reported in the news media have made companies keenly aware of the need to protect data from prying eyes. FBI surveys suggest that 50–80% of attacks come from inside the firewall, making encryption the best option to secure valuable trade secrets, financial data, and customer records. Even where compliance is not mandated, companies are discovering the importance of encryption. Most industry experts agree it's likely to become a long-term trend for e-mail and e-mail archives because of the intellectual property and confidential information contained within.

FBI surveys suggest that 50–80% of attacks come from inside the firewall.

Because security is particularly important for compliance data, many customers want to leverage encryption there. One potential problem with encryption is that, depending on the encryption used, identical blocks of input data (plaintext) may encrypt to different blocks of output data (ciphertext). If you have two files that are exactly the same going in, they'll be different coming out. This can significantly impact single instance storage strategies and potentially increase storage requirements.

Also, if you're going to encrypt data, it's better to do it before it enters your system. Encryption is typically run on a third-party server or appliance such as a Decru® system, so you'd think it would be pretty easy to add after the fact, but you have to think twice before you encrypt data that has already entered into compliance.

Here's the issue: encryption changes the data. If you have unencrypted data already in compliant storage, you can't change it. If you encrypt it, you will need twice the storage *and* a certified audit to confirm that the data itself did not change during the encryption process (also known as a chain of custody). Some archival and compliance software packages do include an encryption option, but the tradeoff tends to be an impact on the archival system performance or increased server requirements.

In a Nutshell: Be Nimble

The only thing certain in archive and compliance right now is that things are going to change further. Although you can't predict every possibility or circumstance, if you utilize open data formats and architectures and avoid limiting your options unnecessarily, you'll be ahead of the game when it comes to deploying a nimble compliance solution that can adapt readily to change.

Comment on this article



AARON SIMS

Senior Professional Services consultant, Network Appliance

With over six years of experience in NetApp Professional Services, Aaron specializes in designing storage architectures for Oracle® databases. He works closely with NetApp customers to ensure that they get the best performance possible from new and existing Oracle applications. Prior to joining NetApp, Aaron spent several years designing solutions using NetApp storage as a systems engineer for BellSouth.net. Aaron is also an avid private pilot, and when he's not thinking about storage he's working on building a larger plane so his kids can visit their grandmother more often.

Sizing Storage for Oracle Databases:

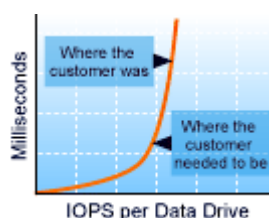
Size Matters (But I/O Matters Too!)

By Aaron Sims

There's more to sizing storage for an Oracle database than simply buying enough disk capacity to accommodate the data. While data warehouse applications sometimes require more storage than I/O, the reverse is typically true of OLTP applications. Supplying enough disk spindles to deliver good performance is essential, and sizing based on capacity alone can lead to unexpected problems.

Several years ago I worked with a customer who learned this lesson the hard way. The customer had deployed an Oracle database across only nine disk drives, which provided more than enough capacity for his small (in size) Oracle application. The application was business critical and the customer was surprised when performance didn't meet expectations. Why? Because each disk drive can only do so many I/O operations per second (IOPS), and this application—although it didn't require a lot of capacity—was pushing those nine disk drives beyond their limits.

No matter what type of storage system they are connected to, even the best disk drives can only deliver so many IOPS with an acceptable response time. While Oracle best practices suggest that response times should not exceed 10 to 15 milliseconds, this particular application generated so much I/O that it was pushing average disk response times up to 50 milliseconds. Ouch! End users definitely noticed. In the end, the customer needed 50 disk drives to bring response times down to an acceptable level and satisfy the I/O needs of the application.



This article describes the critical elements of storage sizing and explains why each is important. There are five key pieces of information that NetApp needs to help you size your storage system correctly:

- Database size
- Read rate at peak (MB/sec)
- Write rate at peak (MB/sec)
- Transaction log rate
- Block size of database

Gathering the Database Configuration

The first step in sizing storage for your database is to figure out how big the various components of the database are, including data files, temp files, online redo logs, and archived redo logs. The following queries can be executed to determine the size of the first three:

RELATED INFORMATION

- [Database Layout with Data ONTAP® 7G and Flexible Volumes \(pdf\)](#)
- [Improving Agility for Database Development, Test, and QA](#)
- [FlexVol™ and FlexClone™ for Database Environments \(Demo\)](#)

Technology-Driven Storage Sizing

Rather than rely on rules of thumb and back-of-the-envelope calculations, NetApp has developed a unique tool to produce highly accurate results when sizing a NetApp storage system for a database. Typical vendor tools deal only with raw I/O requirements, but the NetApp sizing tool offers application-specific "flavors" and converts application requirements to I/O requirements.

The NetApp database sizer is continuously updated and takes into account a variety of critical factors. For instance, it includes the capacity and performance profile of each available disk and the performance capabilities of each NetApp model.

The varying amounts of NVRAM and cache memory in the different NetApp models can affect disk I/O requirements, and thus the final configuration depends on the storage system model as well as the type of disk.

The primary inputs to this application can be gathered from your existing database. The database sizer also takes into account a variety of other factors including:

Component

Query

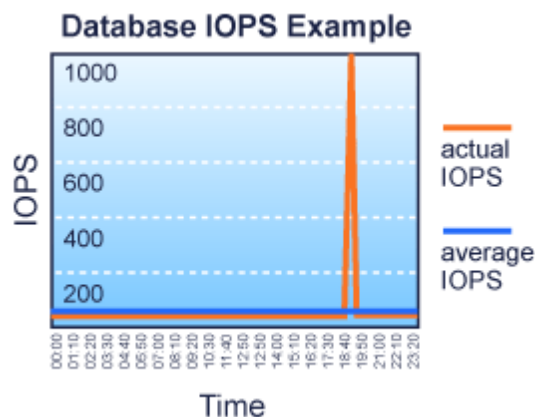
Data Files	<code>SELECT SUM(bytes)/1024/1024/1024 "Datafile Size (GB)" FROM v\$datafile;</code>
Temp Files	<code>SELECT SUM(bytes)/1024/1024/1024 "Temp Size (GB)" FROM v\$tempfile;</code>
Redo Logs	<code>SELECT SUM(bytes)/1024/1024/1024 "Redo Size (GB)" FROM v\$log;</code>

There is no query that can determine how much space you need to store your archived log files since it depends on how long you retain the files. If you're not sure about this, you can simply go to your archive directory to see how many log files you are keeping and the size of each file.

Measuring the I/O Load at Peak Operation

The next step in sizing your storage needs is to gather information about the I/O load created by the database. This step is critical to a successful outcome and a little trickier. My favorite tool for gathering this information is Oracle STATSPACK, which was introduced in Oracle 8i™. The output from STATSPACK gives you the remaining four of the five key items: read rate, write rate, transaction log rate, and database block size. STATSPACK has varying levels of detail. Use level 5 to capture the right level of detail for this purpose. The key to getting good data with STATSPACK is to run it for only 10 to 15 minutes during the period in which your database is at peak load. If you run it for much longer than that, you risk hiding the peak I/O. It's very common for database administrators to send in STATSPACK data collected over a 24-hour period, but unless your database runs at a constant rate for 24 hours (unlikely), this type of data is of very limited use for sizing.

Imagine you have a database that averages 100 IOPS for 23 hours 50 minutes each day, but does 1,000 IOPS for 10 minutes. A 24-hour average as represented by STATSPACK would indicate just over 106 IOPS. If you size your storage for 106 IOPS, you're going to be very surprised when that critical 10-minute interval comes along!



Most database administrators have a pretty good idea when the peak I/O load occurs during the day. If you aren't certain, you can run STATSPACK for 15-minute intervals for 24 hours (96 total intervals) to find the peak period. For some applications the peak may occur on a particular day of the week, at month end, quarter end, during the holiday season, etc. If this is the case with your database, it is obviously best to gather data during that peak period if at all possible.

If you are using Oracle 10g™, you also have the option of using an alternative tool called AWR, which is installed by default and produces output very similar to STATSPACK. A background job in Oracle 10g samples AWR counters every hour (by default) and saves information for 10 days. You can tune these defaults to gather data every 10 to 15 minutes and retain the data for an appropriate period to easily find your peak I/O.

Factoring in Growth Expectations

Once you've gathered the numbers, the next step is to estimate your growth expectations in terms of both capacity and I/O so you can size to accommodate future needs. If this is a database you've had for a while, you probably already know how quickly your capacity requirement is growing, but estimating how quickly your I/O requirements will grow is another question.

The impact of business changes to your application's I/O requirements is highly dependent on the application. For instance, adding 1,000 new customers to a customer service application will certainly increase the size of the database, but if the number of

- Use of NetApp Snapshot™ and other technologies
- Data access protocol (FC, iSCSI, or NFS)
- Oracle RAC

Optimizing Database Layout with NetApp FlexVol

Once you've sized your storage system, the next step is to lay out your database to take full advantage of the underlying capabilities of Data ONTAP 7G, including aggregates and flexible volumes.

The most common questions that database administrators ask about layout are:

- How many aggregates?
- How many flexible volumes?
- How many flexible volumes per database?
- Is it necessary to separate transaction logs from data files?
- What are the performance gains and risks of doing so?
- Can OLTP and DSS workloads be combined on a single aggregate?

These topics and more are discussed in the technical report [Database Layout with Data ONTAP 7G](#). ([pdf](#))

customer service representatives stays the same, your I/O needs may not grow. On the other hand, if you add a new call center and increase the number of representatives by 10%, your I/O needs will obviously increase.

You should think carefully about the likely future changes to your business—growth, expansion into new areas, etc.—and how that will impact your application. If you add a new activity to your database, what impact will it have? Will it make the database busier, require more storage capacity, or both?

Putting It All Together

Once you've gathered your data and thought about your growth requirements, it's time to sit down with your NetApp or channel partner account team to size a storage system that meets your needs. In addition to the hard numbers you've gathered, your NetApp sales professionals may ask you a few additional questions tailored to your specific environment and requirements.

If you're deploying a new database, your account team will use information gathered from a detailed requirements questionnaire to estimate the above data and work with you to come up with a good starting configuration for your new application.

After they've got a good understanding of your needs, your sales team will feed your data into a special database sizing tool developed at NetApp (see sidebar). This application will determine whether your database is capacity bound or I/O bound and produce a range of storage system and disk combinations that will meet both your capacity and your I/O requirements with headroom to meet your future growth. With the best possible options in hand, you can then choose the configuration best suited to your needs and budget.



[Comment on this article](#)



STEPHEN MANLEY, Technical Director, Network Appliance

Since 1997, Stephen has helped define and develop NetApp tape and disk-to-disk backup, mirroring, and compliance technologies. He is the NetApp technical lead driving NetApp/NetBackup integration.

BILL ROTH, NetBackup Technical Product Manager, Symantec

With more than 20 years of industry experience, Bill's background includes technical positions in marketing, support, and manufacturing at VERITAS, NetApp, Auspex Systems, Amdahl, and IBM.

Technical Whiteboarding Session: Veritas® NetBackup™ Integration with NetApp Disk

Over two years ago engineers from VERITAS (now Symantec) and NetApp first started brainstorming ways to improve disk-to-disk backup by integrating the robust management capability of NetBackup, the efficiency of NetApp SnapVault® technology, and the reliability of NetApp NearStore® systems. Today, the fully integrated solution is in use in production environments worldwide.

Tech OnTap recently caught up with two of the technical leads helping to drive integration efforts for an in-depth look at the project and how functionality is evolving in recent releases.

Topics include:

- Customer demands driving integration
- Technical integration points
- The benefits of the joint solution versus tape and other disk-to-disk backup solutions
- How data reduction using single instance storage works



Watch the Whiteboard Session

[Comment on this article](#)

RELATED INFORMATION

- [SnapVault for NetBackup TechBrief](#)
- [Disk-Based Data Protection with Symantec and NetApp \(pdf\)](#)
- [Integrated Solution Demo](#)
- [Information-Optimized Backups](#)

Information-Optimized Backups

It used to be that a larger library, faster tape drive, upgraded network, or new backup server kept backups running.

Over the past five years, however, disk capacities have grown 20x, with a scant four-fold performance improvement for backup. Today, many companies are turning to smarter backup solutions that:

- Reduce backup storage consumption by eliminating redundant files within a backup
- Shrink backup windows by copying only incremental data changes
- Ensure reliable, recoverable backups by validating backups as they are created

Read [When Brute Force Isn't Enough: The Value of Information for Backups](#).

TECH ONTAP ARCHIVE



Prassana V.

Senior Engineer, Sling Media Private Limited

Prasanna is the IT Guy at Sling Media R&D in Bangalore, where he is in charge of system and network infrastructure. He has four years of experience in the IT field, and plays a significant part in enhancing and improving customer experience with technical support. In his free time he maintains a [technical blog](#) with tips and tricks related to system administration and writes reviews of various gadgets and tools. He is currently writing a review of the upcoming Windows® Vista and Office 11 for his friends and colleagues. Prasanna is also a big fan of Open Source software.

Tool of the Month: Angry IP Scanner

Every month Tech OnTap showcases a free tool that just might make your life a little easier. [Recommend a tool](#) and get a FREE NetApp jacket.

Author: Angry IP Scanner was developed on SourceForge.net.

What it is: Angry IP Scanner is a very fast IP scanner and port scanner. It can scan IP addresses in any range as well as any range of ports at each address.

How it works: Angry IP Scanner simply pings each IP address to check whether it's alive, then optionally resolves the hostname, determines the MAC address, scans ports, etc. Plug-ins are available that can gather additional specific information from each host. Plug-ins include FTP Detect, Windows Shares, MAC address via ARP, DNS Aliases, and others. Anyone who is able to write code can develop new plug-ins to extend the functionality of Angry IP Scanner.

Why Prasanna thinks it's cool: Our network consists of resources such as Wireless Access Points, routers, network printers, IP phones, PDAs, and of course PCs. A part of the network has static IP addresses, while other parts have dynamically assigned IP addresses. Using Angry IP Scanner I can quickly and reliably scan the network for devices online, IP addresses assigned to these network devices, open ports, and the host names and usernames on PCs.

How Prasanna uses Angry IP Scanner: Our network consists of resources such as Wireless Access Points, routers, network printers, IP phones, PDAs, and of course PCs. A part of the network has static IP addresses, while other parts have dynamically assigned IP addresses. Using Angry IP Scanner I can quickly and reliably scan the network for devices online, IP addresses assigned to these network devices, open ports, and the host names and usernames on PCs.

I have installed additional plug-ins so that I can extract NetBIOS information such as computer name, currently logged-in user, network shares, and MAC addresses. A scan is initiated by entering the IP address range. It takes a few seconds to complete the scan, and once done all the results are displayed graphically, which makes it easier to analyze.

I also use it to check if there are any rogue entries in the network by comparing it with a previously saved list, and it helps me determine if any systems are running unwanted services such as FTP, by using the plug-in FTP Detect. The tool is very flexible, with options to customize the port range, time delays, number of threads, ping count, ARP MAC for determining MAC addresses independent of NETBIOS information, and favorite IP range. It can be used as a command-line utility or in a batch file. Results can be exported to a text, CSV, HTML, XML, or IP-port list file for reference at a later date.

Recently, one of the security processes in our office required the MAC addresses of all network-enabled devices. All that I did was run the Angry IP Scanner to gather the required information without having to ask each user.

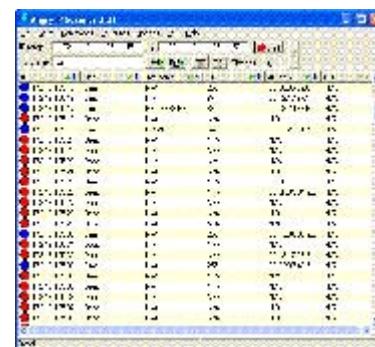
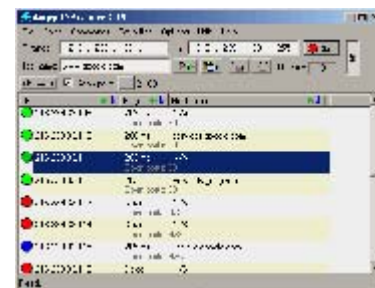
RELATED INFORMATION

■ [Angry IP Scanner Page](#)

■ **Previously Featured Tools:**

- [Firefox](#)
- [Cacti](#)
- [MRTG for Filers](#)
- [Nagios](#)
- [Simulate ONTAP](#)
- [SIO](#)
- [ToasterView](#)

Angry IP Scanner Screens:



Angry IP Scanner is one of the tools that will simplify the life of an IT Engineer.

Caveats: The scan is slightly slowed down if lot of plug-ins are loaded. And of course, a tool that gathers this kind of information can also be used by hackers or others wishing to penetrate security, so you have to be careful about who has the ability to run it.

Want to recommend a tool? If we feature it, you get a free FREE NetApp jacket.

✦ **Nominate a tool**

