TRANSCRIPT OF CONVERSATION WITH BRIAN PAWLOWSKY / DAVE HITZ (NetApp) / MARGO SELTZER

FACILITATOR: RIK FARROW

BRIAN: How shall we begin?

DAVE: She's going to ask us questions. I guess we should introduce ourselves or something.

MARGO: Except we know who everybody is. So here I am with Dave Hitz and Brian Pawlowsky of NetApp and our agenda for today is to talk about network storage and wherever that may take us. So I guess my first question is that people use this term "network storage," but I think different people use it to mean different things. So in order to lay some context, I'd like you guys to tell me what you think network storage is all about.

DAVE: I think we should start with the technical answer.

BRIAN: Storage that's on a network?

DAVE: Ta-da!!

MARGO: Now if you tell us that, do you have to kill us?

BRIAN: I'll let Dave handle this. Maybe it's more complicated than I give it credit for.

DAVE: The place that I've heard this get complicated is, there's a whole bunch of different dimensions when you look at network storage. Brian gave the answer, it's storage over a network. Yes, obviously, but does Fibre Channel network count as a network or does network storage only include Ethernet? Sometimes people say network-attached storage—which almost always they mean Ethernet by that—but is that only file-based protocols or would iSCSI be a form of network-attached storage? And so you can get into really funny kind of technical semantic arguments about whether a particular type of storage like iSCSI is a form of network-attached storage or not, so I'm not that interested in the vocabulary of it but I think that there's two dimensions that matter. The first dimension is, are you using Ethernet or are you using some other form of networking like Fibre Channel? That's important dimension number 1; and then the other interesting dimension is, is it block-based storage like Fibre Channel or iSCSI—basically read a block, write a block, talk directly to the disk drive—or is it file-based storage like NFS or CIFS?

MARGO: So let's look at each of those dimensions. Why does it matter whether you're talking over an Ethernet or something else?

DAVE: From a technical perspective, technical people tend to look at the difference between Fibre Channel and Ethernet and they say you know it's not that big of a difference. What really matters is where I plug into the operating system, and plugging in at the block-device layer is an important distinction as opposed to plugging into the file-system layer. Very, very different ways that the traffic moves over.

MARGO: So that goes back to your other dimension; I guess the question is, are those dimensions really separable then?

DAVE: The block file really is where you plug into the OS, and technical people almost always argue that that's the much more important distinction. Business people tend to focus on Fibre Channel versus Ethernet, and the reason business people tend to focus on that is because they worry about things like capital expenditure and group organizations; if they've spent millions of dollars on a Fibre Channel infrastructure and they're about to buy more storage, they care a whole lot about whether that new storage is going to plug into the millions of dollars' worth of Fibre Channel infrastructure they already bought; whether they're going to plug it into their corporate Ethernet infrastructure, in which case they may need to beef that up; and when it gets to a management ownership perspective because business people often care about who owns what and who do I delegate a problem to. If I'm plugging into my Ethernet, does the storage group buy their own Ethernets or do they plug into the corporate Ethernet? And if they do, does the corporate Ethernet guarantee the service level that will meet the—from an organizational perspective and a capital perspective . . . Fibre Channel versus Ethernet suddenly becomes a very interesting discussion, even though, from a technical perspective, technical people would tend to say, hey, it's a network either way, who cares?

BRIAN: So there is a historical artifact here though that, I think, was interpreted as a technical truism; that essentially the evolution of block storage went into the Fibre Channel network and Fibre Channel SANS, which were much better than using run-of-the-mill Ethernet and TCP networking, which was used for low-grade file sharing along the lines of NFS or things that you see in the Microsoft Windows network. And there was this line between the two that was more an artifact of the evolution of the 2 technologies than a technical requirement. Where we are today is just a total blur of first with iSCSI going over TCP/IP; Fibre Channel protocols being put over Ethernet; block protocols being tunneled through Fibre Channel networks; and Infini-band just playing merrily between the 2 camps.

MARGO: Maybe it's worth actually stepping back a minute and giving a little context for all those different technologies because we talk about them as if they're [unintelligible] but perhaps some of our readership would like a little bit more of an explanation about how—

DAVE: If you go back 15 years it is pretty interesting because originally everybody just attached disk drives directly, and the 2 emerging technologies for attaching storage over some kind of a network were either if you wanted to attach over the Ethernet you got to use NAS, NFS or CIFS—NFS for UNIX, CIFS for Windows—and so the file type of storage attachment, the NAS style, became closely associated with TCP because of that history. Meanwhile, companies like EMC and Hitachi and IBM were promoting Fibre Channel for block-based storage and so in the block world as people moved into the network; so you asked originally, can you really decouple these things? For almost a decade Ethernet meant file-based storage and Fibre Channel meant block-based storage and you really couldn't decouple them. And then iSCSI came along and suddenly you had block-based storage going over Ethernet, and that's when everybody said, well, wait a minute then; is iSCSI a form of SAN or a form of NAS? That's kind of where the confusion I think really started.

You look in Infini-band—relatively small percentage of the market, not very many people running storage over Infini-band. That's really more of a server interconnect technology, so that's more something that futurists care about as opposed to people who deploy stuff for the most part. Man, we got pretty deep and technical pretty quick!

MARGO: That's good. And I think that historical background is useful. So as NetApp, how do you deal with—seems, as you describe it, there is this confusion in some sense that TCP/IP means file-based and Fibre Channel means block-based, so as a company how do you deal with this in helping people figure out what's the right solution for them?

DAVE: This is tricky; I don't want to turn this thing into an ad for NetApp. The specific approach—

MARGO: As technologists.

DAVE: But the specific approach that NetApp has taken with respect to this issue is we build storage systems that will speak either protocol, so you buy a storage system, you put a bunch of disk drives into it; if you want to talk to it over Ethernet that's great, we support iSCSI, we support NFS and CIFS. You want to talk—and that really reduces the burden of having to decide ahead of time which one you want.

MARGO: I guess my question though is, do customers come to you knowing that they want one or the other? I'm sort of guessing that people come looking for a storage solution and because you guys have looked at the market, you provide whatever language they want, but in some sense they're coming to you asking for a storage solution. And so my question I guess boils down to how should a customer think about making that decision?

BRIAN: Let me dive in. I think there's a couple things that are pre-conditions that may cause a customer to go one way or the other. Many large enterprise customers have established data centers with existing equipment, existing procedures and processes around a certain form of interconnect from the history of how they developed their deployments of storage, and they may walk in with a basic structure that you have to fit into. So that would be one clear place where they have an idea of what they need to do. I think a more compelling driver for decision from a customer perspective is application support; the types of data they want to deploy and the partnerships they have and how they—what version of storage deployments they support, what architectures they support. Those two things can factor into a customer having a preconceived notion of how they want to deploy their storage. I think more and more as the technologies are spanning the different underlying physical

interconnects at the storage protocol layer, that it's becoming less and less clear that there's any one answer which some people might try to give.

A long time ago, even though I'm deeply involved in NFS, I backed off the "there's one true solution to all problems." I think what we are trying to spend more of our time moving forward is partnering with a customer to determine the best way and the most cost-effective way to deploy their applications so that they can run their business. And that's, I think, how we're trying to change our mindset in terms of working with customers, giving them a map through all of the possible choices they can do in storage deployments. Because from my perspective, I think it's important that NetApp maintain a rather non-dogmatic stance around the final decisions the customer makes regarding storage deployments. We just have to support their architectures as they want to run their applications.

DAVE: This has been something that evolved over time. Ten years ago it was pretty clear where Fibre Channel would make sense and where Ethernet would make sense. If you were looking at heavy-duty database business kinds of apps you definitely wanted a Fibre Channel. If you were looking at more distributed, users' home directories, you definitely wanted NAS, and it was pretty distinct.

MARGO: Why? Is it again just —

DAVE: Because Ethernet reliability was not as strong; because the applications had not yet been modified to support NAS. If you went and talked to Oracle, they would give you a list of reasons why NFS was not a good solution for running your databases. So 10 or 15 years ago there really was a strong distinction.

BRIAN: And even if it ran over NFS they would say they wouldn't support it, which was a deal breaker for a lot of customers even if—but they said, we just ran the application over NFS and it works fine. I'm sorry, that's a non-supported configuration. And in one sense, I mean they have a business to run and they basically make decisions as to what [unintelligible]

DAVE: They hadn't chosen to train up their problem-solving people on those technologies and so they couldn't really help you. So it was very clear what happened is that Ethernet got to be much, much faster and better. , driven a lot by the requirements of the internet. People expected their web pages to be up 24/7, and so Cisco really put a lot of the same kinds of higher quality data center quality features in the Ethernet that had been in Fibre Channel earlier. And a lot of the out-vendors matured. Oracle said, you know this NFS stuff can save people money. So if you look at it like a Venn diagram of what are all the problems you could solve with NFS and what are all the problems you could solve with SAN, 10 years ago they were disjoint sets. ; there was not really any overlap. Today the vast majority of things you might consider using storage for, you could use either one. It's gone to a Venn diagram with 90% overlap. So the bulk of the problems—you say I want to run an SAP environment—realistically you can do that on NFS, iSCSI or Fibre Channel, whichever one you want. And so in that environment you tend to look at things from 2 different perspectives: there's the business and the technical. Business ones tend to dominate—if you already invested a lot of money in Fibre Channel, you've trained your people up on Fibre Channel and you have undepreciated capital expense of Fibre Channel, no smart technical guy will talk the business guy out of that.

However, if you're talking about we're going to do something new: we're just doing a major upgrade or, look, we're going to get a bunch of VMware, what should we do there? Then in that situation you really can dig in on the technical aspects. So from a business perspective what I tend to believe is whatever you're already doing is probably the cheapest thing to keep doing. From a technical perspective if you've got the opportunity to come in and say, look we're gonna redesign a bunch of stuff from scratch—not always but 80, 90% of the time Ethernet storage—either NAS or iSCSI—is almost always going to be easier to manage and lead to the lower over cost.

MARGO: So let me see if I can summarize that. So if I'm working in the context of an existing deployment, then my solution is stay the course for the most part. Whatever technology I'm using, that's going to be my simplest solution to move forward.

DAVE: There will certainly be exceptions but in general that's what we see.

MARGO: First approximation—and if I'm starting with a clean slate I'm about to build a new data center. Then I get to think this over from scratch; I reasonably can choose either solution, and I think what I'm hearing you say is most of the time people seem to be going with more of the Ethernet-based solutions.

DAVE: No, I wouldn't yet say most of the time people do because there's still a lot of people who are familiar with Fibre Channel and more comfortable with it. What I would say is when people really roll up their sleeves and do a technical analysis—when Oracle decided to build their outsourcing data center and they said, what technologies are we going to use to make this thing as low cost to operate as we can, they answered, we're going to use Linux because that's going to really drive down the operating cost; we're going to use it on Intel as compared to more . . . the commodity approach there. And they chose to go with NFS, so when someone really rolls up their sleeves and says we're really going to dig in on the technical merits and decide what to do technically to drive costs down, they will tend to land on an Ethernet answer.

MARGO: So—

BRIAN: I do have one comment here. Two comments. I think in full disclosure, you are talking to the company that invented the NAS industry and I think that's really critical in this conversation because I think more so—there were a couple other companies playing around, most notably Auspex, in terms of providing a storage solution over Ethernet and NFS. But it was—NetApp blew the doors off the NFS industry and then our introduction of the Windows and NFS file-sharing server—that basically was our first step in multi-protocol, and we ended up inventing consciously with a lot of Dave's direction and drive the NAS industry, the NAS segment; took a long time to get people to recognize that that existed and it was a proof by deployment and numbers that convinced the analysts to start tracking it because there were hundreds of millions of dollars now being invested in the NAS segment versus the SAN segment, which was the Fibre Channel industry. So there was a—

DAVE: Your point is we might be biased?

BRIAN: No, no! That we were a major player in the beginning in the NAS segment of the industry. As we grew we basically entered the other storage markets, and I think the second point that's sort of important here is that on a lot of levels this discussion of Fibre Channel SANS, Ethernet NAS I find boring beyond tears, right? Because no one—the customer base in the industry has matured tremendously since 1993 when you guys started shipping, when I wasn't here; I started in '94. And NetApp started shipping out those file servers. They came out with an NFS file server that had a backup utility on it, and you could back up your storage using dump and restore from Berkeley with an NFS connection to it and some disks. The expectation today about storage deployments is that the interconnect and how you choose to read and write your data and over what wire you choose to write it is the irrelevant part of the equation.

The more important stuff is what you do with the data, the data protection, the data integrity guarantees you provide, disaster recovery capabilities, and the value-add that you layer on top of that data to allow people to reuse it in different forms and leverage that data in different contexts. That's why people are interested in the differences between storage vendors today, not the protocols over the wire.

MARGO: That's perfect because my next question was going to say—so what we can take away from this is that in some sense these decisions are no longer important. So you lead into exactly the next area which is, okay, you're a storage vendor; it used to be that when I wanted storage I went and I bought the best price-performing disk I could. And it's no longer a pure price-performance choice in storage, so what are those other characteristics that you started alluding to and what are the value-adds that storage manufacturers are really gonna have to compete on?

DAVE: Let me start top down. It's humbling as a storage vendor to recognize that CIOs do not give a shit about storage. CIOs have some list of business problems that they want to solve and in general each of those business problems links to a particular application, like all the employees need to be able to send each other email; okay, we've chosen Exchange, and so the CIO's top-level concern then is, how do I run Exchange or are we going to balance the books, Oracle financials, how do I run Oracle's financials? The more that a storage vendor can talk to the CIO about how their storage makes some kind of difference for running that application, the better off you are as a storage vendor. So—your face is all scrunched up.

MARGO: So the question is then, that makes it sound like your value-adds are all application-specific and I'm going to claim that there's gotta be a set of common value add-ons that you can argue will help your Exchange server and will help your financial apps and will help something else, and that you can't possibly run a business having to argue each individual application independently.

DAVE: There are common technologies that can help a lot of different apps, but I'll tell you when you get into actually working with someone doing an Exchange deployment versus an Oracle deployment, they care about fundamentally different things; but the underlying technologies that you link in—let me use Exchange as just a really specific example. One of the things that people have noticed in Exchange deployments is the Exchange database tends to get corrupted, just something that tends to happen and we're not going to make any comments about any partners we happen to have; Oracle people don't complain about that, we've noticed. So in an Exchange world, something that Exchange administrators care a lot about is, how do I get back to the earlier version of the stuff I had that used to be good? And snapshots are a beautiful tool for doing that, so if you can get back to that earlier version . . . and the more automated the better, right? If you could take the Exchange tools and say, oh, what about this snapshot, is that good yet? No, still corrupted; or that snapshot from an hour before that, is that good? The more you can automate and link into that so that the storage-centric idea which is snapshots becomes relevant to that Exchange person who—think about the challenge in the real data center—you got an Exchange administrator who typically doesn't own his own server, so there's a server administrator who typically doesn't own the network to the storage, so there's like a Fibre Channel or an Ethernet administrator, and then down somewhere further on there's a storage administrator, and often each of those people reports to a different director, sometimes a different VP. And the poor Exchange guy is just trying to get his database back the way it used to be, right? If you can somehow work with that Exchange guy and say, look, here's a tool that lets you do all this stuff and now your Exchange environment is back up and running again without having to have even talked to the storage guy—that's a whole different model.

In Oracle, on the other hand, one of the big challenges is people are always running test and development environments. They're not so worried that the database is corrupt but they say, I've got this giant production database that I'm not allowed to touch, but I'm doing some little tweak in the customization that I have for SAP, say, or Oracle financials—I wish I had a playpen I could work in. Snapshots, writeable snapshots or clones, are a great tool for that. But again does the Oracle administrator know how to go get the server guy to mount and the storage guy to create new LANS? So your point, can you really win in this industry by optimizing one app at a time? No, you do underlying technologies like snapshots and clones and thin provisioning and data replication that are useful in a wide variety of areas, but then you really do have to look—there's a bazillion apps but you look at the combination of the major apps—the Microsoft Suite, the Oracle, the SAP, VMware is an emerging one that has common characteristics—test and development environment, sort of the typical UNIX, home directory—you look at that set of apps, optimize for them based on a common set of underlying capabilities. But you look at the underlying capabilities, they tend to be around how do you virtualize your storage more? How do you create snapshots, how do you do thin provisioning, how do you do clones? You've got lots of data here, how do you get it to there? De-duplication—those are the kinds of building blocks.

BRIAN: I want to make a comment because Dave just kind of glossed over a large part of our history, and I just want to bring a history into our current kind of strategy around leverage. So snapshots—it wasn't that there weren't a lot of NFS servers out there; one of the key differentiators was basically instantaneous point-in-time copy of an entire file system at essentially zero cost except for—because it was basically in place and copy on write. And that shattered the Exchange deployment model from what other people were providing around Exchange when it came out.

DAVE: Shattered good or bad?

BRIAN: Shattered the preconceptions about the time required for backup and the number of recovery points you could have in your Exchange environment; when it divoted on you, what you hope to recover and how fast the recovery was. Snapshot just blew away the traditional methods of doing backup to tape or any other—backup to some mirror disk, and the cost of making a mirror disk that was probably corrupted also because it wasn't—at that point in time you weren't breaking the mirrors often enough. Fast forwarding, that experience from the late 90s when we started seeing vast incursions into Exchange deployments for our product, a lot of times our

customers were coming to us kicking and screaming about many different applications before we were giving them the tools around it.

I think there was a recognition that it's not the primary copy of data that's what is most important and of most concern to people in an organization. It's the secondary copies of data: the recovery points, the archives, the ability to leverage and reuse data that has to be managed because of the cost of making those copies for different purposes, but also their usefulness in terms of business continuity. The primary copy—everybody was designing around and everything was optimizing for it but what came, I think, circling back to everybody was the cost and value of the secondary copies that our fundamental technology enables interesting processes and techniques around, regardless of how you access the data. How we do data management with snapshots, how we do disaster recoveries and secondary copy management applies to Fibre Channel SANS and to NAS and file access.

MARGO: So what I can take away is that snapshots were a truly fundamental value-add that helped you differentiate early and that continued to be leveraged to solve a bunch of different business problems.

BRIAN: Yes.

MARGO: Can you give me another example?

BRIAN: Of how snapshots are leveraged or—

MARGO: No, so snapshot—I was a big fan of snapshots from day 1, I still am—

DAVE: What have you done for me lately?

MARGO: Yes, what have you done for me lately? So snapshots were a great idea but it's 2008 and what's the next piece of core technology?

DAVE: There's a handful of different ones that we can work our way through. One that I think is interesting and people don't understand the ramifications of it as much as they might is RAID 6 or RAID DP, which is the ability to allow any 2 disks to fail instead of just 1. And the main reason that's important—as you look at disks getting bigger and bigger the question becomes how large of a RAID group can you build? When I say RAID group, one parity drive with how many different disk drives; and as the number of disks gets bigger and each disk drive itself gets bigger—so say you put 10, then your overhead's 10%; you put 20, your overhead's 5%. The more disks you put in there, the more data you have to read to reconstruct a bad one. In fact, if you look at the math, disks are getting so big these days that just looking at the bit failure rate—current sized disks you build a standard RAID group of 7 disks—that's how many we fit in our shelves—we have 14 shelves so we put in 2—you would expect to see failures 1% of the time on your RAID reconstructions just as a result of the raw bit failure rate of the underlying disk drives.

So imagine that doubles again because, remember, if one drive fails you have to read all of the other drives. So it was already getting bad for regular disk drives; the real challenge is what about those cheap ATA drives? Wouldn't it be nice to be able to use SATA drives, which are both bigger and slower so it's gonna take longer; and less reliable.

MARGO: The whole triple play!

DAVE: Yes, the trifecta, to quote the president, that's at the top of many people's lists. So RAID DP is obviously good because instead of going with mirroring, which doubles the number of disks you have to have, you can have, say, 14 disks and 2 for parity; so that's cool; but what it really enables is it enables you to go into ATA disk drives and use them in production environments instead of just secondary environments. And ATA is way more space-efficient and as a result of—an ATA drive uses about as much power as a Fibre Channel drive, but it has a lot more storage, so if you're in an environment which is power-cooling floor-space-constrained, ATA is a win; obviously, they're cheaper.

MARGO: So I'm hearing remnants of the conversation I heard in the early 80s, which is you replace big expensive drives with redundant arrays of inexpensive disks; and I think what I'm hearing from you is—

DAVE: Same idea—

MARGO: It's the next generation; now we're going to replace arrays of moderately inexpensive disks with arrays of really super cheap disks that are unreliable and so therefore we're going to have to go to even bigger parity.

DAVE: Absolutely. If you look to EMC, the way that EMC enabled to transition from the DASD style of drives to the cheaper emerging, more commoditized drives of that era was through the invention RAID 4, and I do think that ATA or SATA drives enable the next generation of this transition.

BRIAN: I want to connect the 2 topics of snapshots and RAID DP in a way because it goes to, I think, NetApp strengths and when we're at our best in terms of product development. So it's not as if there weren't solutions with [unintelligible] mirrors for business continuation and recovery, stuff like that. What was the strength of snapshots was basically the commoditization and making snapshots available for everyone at no cost, essentially, compared to all other solutions. The clever parts about RAID DP—having double disk protection—was not like we invented double disk protection; that wasn't the clever part. RAID 6 was certainly [unintelligible] just that no one could ever enable it because they would regret that decision forever more because of the performance applications. The clever part about RAID DP is that it was enabled with no more than a 1 to 3% performance drop versus single parity disk protection on all our storage systems to the point where we ship it out by default in all our system from our low-end SMB product, the 300 S500, up to our high-end systems—

DAVE: And all our published benchmarks have it enabled.

BRIAN: All our published benchmarks have RAID DP enabled. So the clever part in terms of products versus technology is being able to actually give this technology to customers and allow them to use it in their daily operations.

MARGO: Which reasonable performance?

BRIAN: With better—

DAVE: In order to use it in your daily operations it has to have reasonable performance.

BRIAN: Otherwise it's an interesting little project, and I think in a lot of ways that's what—at NetApp we try to surf the technology curves in terms of delivering products to customers that take technology that other people may have touched on or may have—other people are working on—by making it available in commodity storage systems. And when we make those, connect those dots together and deliver the product to customers, I think that's where we shine.

DAVE: Another zone of technology is that they're built on top of snapshots, but it's really something different; all of the data replication technology that NetApp has—it turns out that snapshots are a beautiful starting point for replicating data to a remote location, and the reason for that—or even making a local replication but especially remote—the reason for that is one of the biggest challenges of replicating data is, if the data that you're copying is changing underneath you and you're moving the blocks, depending on what order the data changes and depending on what order the blocks move in, you may get corruption on the copy that's of a form that even FSCK can't fix up. You've gotta be really careful, so in a lot of situations when you do bulk copies to a remote location, you may even have to quiesce the system. If you have snapshots as a foundation you can just copy all of the blocks in a snapshot, and you know that a snapshot is static and those blocks are locked down so that changes don't go back on top of those blocks. The result is that starting with the foundation of snapshots, NetApp has built a whole collection of technologies that will do remote replication; the simple one is just mirroring a system so that you've got a local system, and the remote one looks just the same; and that's obviously cool for DR—

MARGO: You said "just the same," but for using snapshots there's a time delay.

DAVE: Sorry, just the same as it looked at some point in the past; just the same but with a delay.

BRIAN: And with a well-defined consistency point.

DAVE: Yes, with a well-defined consistency model. What a lot of customers started looking at as they moved away from tapes for backup was they said, I like that model but I wish on the remote machine it's probably made with cheap—even if my primary storage is still Fibre Channel the remote machine, let's say, is really cheap, boatloads of ATA, and probably a lot more drives per head, so the performance wouldn't be there necessarily for

running an app but just kind of for reference. They want to keep the snapshots for a lot longer; so enabling the capability to say, yeah, in your primary systems maybe you only keep snapshots for a day or two or maybe a week; on your remote system what if you could keep snapshots for literally a year or multiple years? We started getting banks looking at this and saying tape just isn't scaling; disks are getting bigger and faster, faster than tapes are getting bigger and faster; especially the faster part. And so a lot of banks have a regulatory requirement to keep data around for 7 years, and so they started saying can we have 7 years' worth of snapshots, please?

And as opposed to the snap mirror, which is make another machine just the same as the one you've got, how about a snap vault? My visual image of a vault is it's dusty and shit lives there forever, right? So that whole domain of how much of your data would you like to replicate and how easy is it to replicate? That's a place that we've looked at a lot.

MARGO: And it seems that when you get into that model of okay, I need my snapshots for 7 years and they're going to be spinning; then I also have a disk lifetime problem and the disks don't necessarily last 7 years, and so I also have a problem of refreshing my disk farm as it's running with these [unintelligible]

DAVE: Sure, I mean typically keep things for—the capital lifetime of this equipment for most customers is 3 or 4 years. Some people keep it for an extra year, but really 3 to 5 years is typically the replacement cycle. So the point of keeping a snapshot for 7 years, that snapshot may not be living on the same system or the same disks, but that snapshot—it's the same bytes in the same file system organized structure. The snapshot may live for much, much longer than any of the physical components live.

MARGO: So as a customer when I'm refreshing my vault, I assume I want to keep my vault spinning, so am I doing sort of a real-time online migration to my new vault and then sort of replacing incrementally, or am I really doing—okay time to copy the vault!

BRIAN: I had a comment from one customer; I asked him about in the movie industry you do all these movies; what do you do with all the results of the stuff and I guess it's a migrate forward and prune operation. For the regulation compliance purposes where you want to keep things for 7 years, then that's some category of data, right? And I think part of managing your expenses and your business is determining what you have to apply those policies to and how much data you have to keep around, and you reduce your costs as long as you don't inadvertently delete important data that you need—you reduce your costs by carefully maintaining the data longevity and how long you keep data for compliance purposes and for business purposes.

So from data that's not constrained by compliance regulations I think the policies people generally apply are migrate it forward to new equipment and prune on the way. Dave may have a comment but I asked this question of people who are managing increasingly large amounts of digital storage in the movie industry.

DAVE: It's funny, it depends on the industry how much they spend on storage, but here's something interesting: the disk drive industry keeps giving bigger and bigger disk drives for the same price and so, roughly speaking, every 1.5 years the capacity in a disk drive or the capacity at a given dollar amount doubles. So even if all you do is every year you keep spending the same amount of storage; whatever your storage budget is, every 18 months you got twice as much storage. So if you think about storage that you saved that you're still keeping 6 years later, it's a good question to ask how much effort would it be worth to go pruning it, because 18 months later everything that you owned forever going back will fit on half of the storage you bought this year; 3 years later it'll fit on a quarter of the storage this year; 4, 5, 6 years later—just stick it in the teeniest little corner of the new—now that's assuming that your budget's continuing to grow and you keep buying new stuff, but just like—

MARGO: But there's another parameter in addition to budget, which is how long does it take me to copy the disk drive and is that constant or is that becoming—is that getting bigger or smaller? If it takes me an hour to copy one disk drive today; 18 months from now, how long does it take me to copy one disk drive?

DAVE: It's getting worse; the answer is it's getting worse, but I haven't seen that be the main limiting factor as opposed to money and the cost of—

BRIAN: I think the way people deploy—people use our storage—it's fairly transparent, so you're making a migration forward; you don't have a gun to your head you have to have this done in an hour; but the important

thing is you do it transparently and non-disruptively. And that's where we kind of focus our efforts. I'm sure the speed is important but—

MARGO: But my question is, is there a wall out there—

DAVE: Is there a wall is an interesting question. Let me tell ya, we had a customer cancel and one of our customers said—because we were talking about these issues of do you delete it, do you keep it, how long do you keep it; and he said, look, realistically all data falls into 2 categories—data that I decide to delete immediately or say within the first week of creating it and data I choose to keep forever. And he goes, once I haven't decided to delete it within the first week, I just never can afford to delete it. And then one of the other customers says, wait a minute—friendly amendment—he said, there's a third category which is data that within a week of creating it you decide exactly how long you'll keep it.

So, for instance, for banks, customer records—right away when you create that data you decide, I'm legally required to keep it for 7 years and the policy at our bank is we will keep it for exactly 7 years or maybe we'll keep it for 14 but whatever it is, at the instant the data is created, either you tag it with a deletion time, which might be right away—it's a temp directory—or it might be 7 years or—in the medical, the retention requirement is lifetime of the patient plus 7 years. So if you imagine a pediatrics department, I mean they need a plan for how do I guarantee I can get the data back in like 107 years or something like that.

MARGO: And read it, thank you.

DAVE: Yes, in a format that will still have some relevance. But in general they—nobody's figured out a cost model—I mean people are looking at automating this. There's companies like Kazion that are looking at how do we go through automatically scanned data; look for searches, look for usage patterns and see if we can't figure out—you know what? This is just data that doesn't look like data anybody's gonna care about.

BRIAN: So I mentioned there was a little event going at the university; 8 professors coming in talking to our advanced development group; and I mentioned that at the beginning because I just used the movie studio as a kind of migrate and prune operation; so I'm suspecting that that's the ideal and that happens less than it should and that the trends in the industry about the disk drives increasing in size make it not as important; just throw more disk space at the problem, why make this decision? And the possibility of inadvertently deleting something like a minute and a half of the last Oscar-winning animation would be a very bad thing to happen, right? Especially with Blu-ray on the horizon; you want to remaster that.

So adding on to this pile; you talk about technologies—is de-duplication. So NetApp has a de-duplication technology in which we—I think we best call it dense volumes and dense data format. And basically we're taking data, compressing out the—not compressing, eliminating the duplicate data by a couple different algorithms we have playing around; and not only are you getting larger disks but you're basically squeezing out the redundant bits of the data to get a lot more space. And one would think—this is the comment I made to the professors—one would argue that, well, now we're gonna have a lot of free space; we can start pulling these disks out of our data center and then we can basically reduce our footprint in the data center and power consumption, and realistically that's not gonna happen.

And this has nothing to do with technologies; everything to do with human nature and this is your garage at home. Your garage gets full at home, you think the solution is I gotta buy a bigger house with a garage twice the size and I will never have this problem; no, you will have that problem again; in the future your garage will be just as full as it was before, eventually. So what we are having problems with, I think, is all these technologies—and Dave just touched on it, search and indexing stuff—is that what we're facing is a massive pile of data, and I think one of the moving-forward challenges for NetApp and other companies and researchers is wading through piles of data because you basically have been paralyzed deleting stuff that you might need in the future and you think—there's something there that you want and you have to find it through this gigantic amount of storage you deployed; NetApp will provide this storage that'll keep it around forever, right? I mean data availability, integrity and durability of the data; so you'll be able to get access if you can only for the life of you find it and I think that's a lot of where we're looking at—adding value in terms of another layer of data management.

DAVE: There's a handful of exceptions to the never throw anything away or never delete stuff; the only places I see that really working are when you get—

BRIAN: [unintelligible] you really, really want.

DAVE: Then you delete the stuff—no, very large amounts of data that are very structured so you know when a whole big chunk of data is . . . so someone is designing a car and you take all of the schematics and all that stuff for a car, and 10 years later you may say, you probably don't want to delete it but you really can decide we really will move that to a different place. A movie's another example; during the production period of a film it is in very, very active use at the point that the movie has shipped . . . now you may still want to keep it around in some form, but that enormous chunk of data suddenly becomes just not accessed frequently. Another example is oil fields, the guys that do seismic processing for fields; there's a period where they're going to go in and drill a whole bunch of stuff and all of that's very active and then they say, you know what? We're retiring this field. Those are the kinds of examples where you get enormous amounts of data that goes from actively engaged to retired just kind of at a snap. Very little data is of that nature.

BRIAN: Can I push back on the oil field thing? Oddly, about the oil field exploration when I'm sitting with these guys is once you do undersea exploration the initial exploration is a very expensive proposition where they layer the bottom of the sea floor with explosives and measurement devices to develop a 3D map of what it looks like underneath the sea floor where oil pockets might be. What they concentrate on is getting as much data as possible; I mean as much data as possible. And 3 to 5 years out they change algorithms or ways of viewing the data and reprocess the data again to find, because it's become—because the price of oil has gone up and those oil fields that may be there, they might be a little more expensive to get to now—become quite lucrative. But they didn't have the mapping technology and the processing technology and the algorithms that they're continually developing just like in animation. In one sense it's a parallel here; increasingly realistic—though not too realistic, please—animation and better viewer experience.

DAVE: To go back and reactivate the data—you gave the Blu-ray example—a 10-year-old movie that they cut and shipped, that whole thing's dead until Blu-ray comes out and then suddenly they do want to go back and get the (increasingly, it's going to be) digital masters and re-cut the thing.

MARGO: And there's still cases where we can't keep all the data and so we just want to keep as much of it as we can; the Large Hadron Collider that's supposed to go online this year or next year produces data at such an alarming rate that they actually threw out something like 98% of it, and the data they keep is algorithmically determined, and the real bummer is 5 years from now when they discover there was a bug in that algorithm.

DAVE: Here's an interesting thing about that, though; that problem applies to an increasingly small part of the industry. The rate that disk drives are growing; it used to be that individual people—I remember I used to literally worry about how much stuff could I fit on my disk and should I throw away that old PowerPoint presentation; it really was an issue, and you look today—the amount of space I've got on my laptop, and if I ever run out of it I'll plug a USB drive in; that's just not an issue and that same effect is moving further and further up-market; simple math—suppose you're a company and you've got a million customers and you want to keep 1000 bytes on all of them—that's only a gig!

MARGO: And pictures of them too.

DAVE: Let's say I want to keep a megabyte on each of them with pictures and stuff; that's only a terabyte.

BRIAN: So, Dave, you once said large amounts of data arose from 3 general ways; I don't know if you still—

DAVE: Yes, I have these rules of large data; it used to be you could generate large data by typing and now you can type for the rest of your life and you'll never fill a disk drive; in fact, you can talk for your life and with decent compression you can never fill a disk drive and I think we're approaching that for—you can look for the rest of your life and you won't fill a disk drive; we're not there yet. So my conclusion was there's only 3 ways to create enormous amounts of data. The first way is generated by computer. So when you look at things like object files or computer-generated graphics or—

MARGO: Large Hadron Colliders.

DAVE: Well, no, that's actually a different category; so the first is generated by computer; the second is one person typing can't fill a disk drive but a million people typing at once, it turns out, can still fill a disk drive.

BRIAN: Facebook.

DAVE: Facebook, Yahoo email, Google email—

MARGO: Fourteen-year-old girls can fill a drive!

DAVE: And the third category is sample the real world, which is the category that I put the Hadron collider in. So sample the real world would include videos, photos—and these things are multiplicative. I just argue it's hard for me personally to fill a disk drive with photos if I don't use cameras like—if I use my Instamatic. But you get a million people using photos all at once and suddenly you can; so you combine sample the real world, a million people at once, and computer-generated, and those are going to be the sources for big data.

BRIAN: So I kind of rephrase—Dave's the one that told me this but I rephrase it to be the law of large numbers: lots of people, sample the real world, and that would be physics and the Hadron collider is an example, and then there's simulate the real world and that's animation and special effects—

DAVE: No, I don't have a fourth category. Lots of people have suggested categories which I generally put into the same ones, but Eisner's category was make another copy of the same old data, which I have to say it is amazing—when you go in and talk to customers who use data, the percentage of the time that they go, my database, yeah, it's big but it's not that big. The thing I just did—million customers times a million bytes for each one—but do you know how many copies I have of that exact same data just for test and development? You'll get numbers like factor of 20—

MARGO: It's an arms race, right? You've got customers who insist on copying data for a variety of reasons and you've got de-duplication technology that is trying to remove those duplicates, and those are just competing in an arms race.

DAVE: And you need to figure out—the key thing here is—and this is where you get all back to what should be virtual and what should be real. Sometimes you copy data because you really want a copy of it somewhere else. Like the last thing you want to do is de-dupe the copies between your primary data center and your backup data center. That's completely not the point, so those are copies typically made to respond to physical problems; I'm worried the disk drive will break or worried my building will burn down, something like that. And then there's copies you make for virtual reasons, like I've got my production environment and want to do an experiment and I only want to change one byte but that's a virtual copy. So when you're creating copies for those type of reasons, then you don't really want to physically copy them; you would just like to do a virtual copy.

MARGO: Yes and I guess the question is, do you really trust users to make those distinctions for you?

DAVE: That gets back to what we were talking about very early on; when you look at who makes those kinds of copies and why, they tend to be different people in the data center; the person worried about the disaster recovery infrastructure typically is somebody involved in servers and storage, networks, buildings, that stuff; the person making a test and dev copy is typically the Oracle admin and so that's why you would like tools that let the Oracle admin just say, I would like to run an experiment. Dear tool, may I run an experiment? Yes. Here's a new instance of Oracle and we have a clone test; in that context nobody is expecting that a real copy happened anywhere; the fact that it happens to be fast and cheap—that's cool because now you can let people do it a lot.

But they weren't designing a DR environment and they weren't expecting to have any impact on it. You could imagine some confusion and admin could think, oh, gee, I just created this, now it's safe, but in a corporate environment—typically it's somebody else's responsibility to have made it safe already.

MARGO: Yes, I was just thinking of my own personal experience where the [unintelligible] got that wrong; so make a mailbox on your Macintosh and move some things over there; whether it's a virtual copy or real copy is not always clear.

DAVE: Yes, drag and drop interface; as it turns out they were all sim links, so guess what? Sure was fast though.

MARGO: So what keeps you awake at night? What do you worry about?

DAVE: Those are 2 different questions. My wife and I have a 9-month baby. To say what keeps us awake at night?

BRIAN: My wife wants a baby, that keeps me awake. I think when you talk about technology there's one thing that we're investing a tremendous amount in and that's certainly keeping me awake in terms of how are we going to do this; what are the different things pulling at it. So it's certainly the case in storage systems today, essentially virtualized disks. Nobody really in any serious storage array, no one is actually talking to a disk from the application server perspective or the network perspective; that's just so boring and useless just because protection is required, integrity, everything else. So, essentially, storage arrays take disks and horizontally scale them and provide capabilities that no single disk can give to an application; in the same way that we virtualize disks with horizontal scaling across drives and getting capabilities that weren't there from the base unit, we're basically looking at horizontally scaling storage clusters. And there's 2 things driving this, or 3 things. One, there is regardless of all the performance improvements, either through clock rate and now multi-cores on processors and interconnects, you eventually hit a single-box skin limit as to what you can cost-effectively put together in any way that results in a viable product. And you eventually have to connect these things together and have to have more than 1 box.

So NetApp customers have—most customers have way more than 10 boxes and often far more than 100 within their sites, and what we are looking at is cluster technology to make that experience of horizontal scaling of performance capacity a seamless experience, just as you add disk drives for additional capacity in a single box, adding controllers with their associated disks and horizontal scaling to provide a kind of seamless expansion in the face of these growing storage demands.

The 2 things that—I think playing with it just because of the fundamental limit of a box and forcing you to consider horizontal scaling is that there's also sweet spots at any given point in technology for what's a price performance solution that makes interesting building blocks for a horizontal scale-out solution. Horizontal scale-out allows you to do mid-range competitive products that provide a capacity and performance scaling that is more cost-effective than putting gigantic bricks together from smaller configurations. I think the third thing that's coming out is emergence of standards, and this goes to the ITF work and the NFS work; that NFS is this old protocol that was invented at Sun in 1984 and here we are in 2008 and we're coming up with an NFS version 4.1. I guess that would make it—I don't know how you count, because there was never an NFS version 1; it was version 2, 3, 4.0 and now 4.1.

Where one of the significant features in that protocol is basically providing horizontal scaling from the application server perspective in what I'm gonna call a commodity cluster file system called pNFS that can exploit a horizontally scaling storage solution at the back end, and the 2 together is a match of emerging technologies and capabilities from the storage side of the applications.

DAVE: A list of things to worry about—Flash has the potential to really change stuff. Flash memory has some very interesting characteristics: it's roughly 10 times as expensive as disk and people say, hey but it's getting cheap faster. Well, disks are cheap pretty fast as well so I'm not counting on Flash crossing over. So it's 10 times as expensive as disk. In random access situations it's about 100 times faster. So that's—you do the math, that comes back to be 10 times as cheaper per op of performance. So the concerning question there is what do you do with it? It's too expensive to realistically expect it simply to replace disk drives as people want more and more data. I mean it certainly will in some applications as it gets cheaper; when it gets to be a small enough percentage of an iPod, you just say, hey let's just use Flash instead, which we've done. When it gets to be a small enough percentage of a laptop's cost, clearly that's what is coming next. So for low-end systems it'll work its way up. In larger storage systems I think the real question is, how do you design them differently to take advantage of it? And the answer won't simply be build enormous storage systems and put all Flash-based—I mean there may be a teeny market of super high performance but more realistically I think figuring out how—it's the first technology in a long time that I think realistically can be an additional tier in the caching hierarchy and figuring out the best kind of architecture to take advantage of that.

Where there's some easy things to do with it, there's some more difficult things to do with it; that's something Steve Kleiman's worrying a lot about these days. He's our chief science officer?

BRIAN: Chief scientist.

DAVE: But when you just talk about what kind of stuff you're losing sleep over, that's another one. It's interesting, I used to have a much more technical role. These days I am hanging out much more with the sales guys and the marketing guys—that whole side of the business. So I worry about things like, overall, in the storage attached over some kind of network market—NAS, SAN, iSCSI, Fibre Channel, that whole market—we've got about 10% market share. We've got about 10% unaided recognition, which is, roughly speaking, all of our customers know who we are and if you're not our customer you probably don't. I lose sleep over that.

How can you sell stuff to people who have never heard of you? It's a whole different domain of challenge and it leaves you more to answers like, gee, maybe you should figure out how to tell more people and how you go about doing that.

MARGO: It's the same question that every startup faces, and perhaps it's encouraging to every startup to hear that big companies face it too, just in a very different scale.

DAVE: Absolutely. When you're a startup you just assume that nobody would have heard of you and so you just go selling based on the assumption that everybody you encounter hasn't heard of you and your first thing you explain to them is yes, you haven't heard of us but here's what we do. You get a certain size and you would hope that that starts to change, but in fact the only reason it does is because you actually spend money and put billboards up in airports or whatever thing—these days Google ads are actually more effective.

BRIAN: Six or 7 papers at the FAST '08 conference, depending how you count it.

RIK: It got slashdotted last year. How about open source competition? Free BSD also came up with a way of doing copy and write, snapshots, and now we have ZFS from Sun which is an open source.

DAVE: Sun is open source but absolutely you can grab it and get hold of it.

RIK: You can grab a free copy and of course a lot of the people who will be reading this are going to be thinking, gee, couldn't I do this myself for less? The do–it-yourself approach. We had a guy write about something he called the data monster a couple years ago. He was out at University of Colorado atmospheric research (UCAR) and he needed a certain thing; he had a limited budget and so he had to figure out how to roll his own essentially SAN for some giant project.

DAVE: I see that as part of a larger category that is—I call it goodness over supply after Clayton Christensen's book The Innovator's Dilemma: basically he argues that in a lot of industries you get to the point where what people are providing is just good enough for most of the customers; in the context of laptops, for instance, if you're not doing development, if your laptop experience is like 99% of everybody—web, Word, email, maybe PowerPoint—Intel keeps wanting to sell you a faster chip and at some point you go, you know what? I don't want a faster chip. How about one that the battery lasts longer or how about one that's lighter or how about it wasn't so hot so if I kept it on my lap I didn't burn my leg. You start to get to other criteria, so one of the things that I do worry about is you look at the amount of storage you can hold on disk drives in storage environments; there's categories of customers for whom that continues to be critical: the Hadron collider guys. And then there's a lot of categories of customers that just say, I buy a PD with a handful of disk drives and I can store more than I can imagine saving for a long time as long as I don't let my kids near the downloads.

MARGO: And one of the interesting metrics we've always tracked is the Costco price of a terabyte of storage, which is not reliable and not secure and doesn't have many of the qualities, but I can go to Costco and buy a terabyte for about $300, and I think that hits your "good enough" point for many, many things.

DAVE: So when you look at it from that perspective, the heavy-duty storage companies—NetApp and EMC and Hitachi, folks in that business—they're already, we are already in the business of finding the people with the worst problems and—our vice chairman Tom Mendoza used to be the head of worldwide sales, he always puts things in simple pragmatic terms. As a company your job is to find people with problems so painful they're willing to pay money to help get them solved. That's kind of the essence; if you find someone with a problem that they're actually not willing to pay money to solve, then you don't have a business. If they don't have a problem at all.

MARGO: That's the zero billion a year market.

DAVE: That's what you're out looking for, so the question you always have to be asking is what's the next hard problem and what parts of it does a particular technology solve? DFS has a pretty cool file system. We've got, obviously, a list of reasons why we think it's better and our cloning and blahblah, but if you look and you go really today why would a customer care about one thing—what would they care about in comparison—they would care a lot more about the overall support relationship; when something breaks who do I call? How does Oracle feel? Over time, yeah, those will keep changing. So our job needs to be to keep finding the next harder problem. That open source isn't that different from Microsoft commoditizing things. It commoditizes stuff in a lower price that you get kind of for free. Microsoft doesn't commoditize it for free but once they bundle something with Windows, it's tough to charge money for Netscape when IE comes for free in Windows. It doesn't from a commoditization perspective matter that much whether it's coming up at you from a low-end Windows environment or coming up at you from open source. Either way you better keep looking for the next hard problem somebody wants to pay money to solve.

Here's the good news. As long as disk drives keep doubling every 18 months, the amount of data everybody has to manage—to store, replicate, de-dupe, to do all this stuff—that's a pretty steep innovation curve to track. And I think the steeper the innovation curve, the better an opportunity—when you're a company charging money for stuff, in the end I look at that—in essence what the customer is paying for over and above the price of the metal that you sell them; they're paying for you to hire lots of smart people to innovate to solve their problem or to go visit their site and solve their problem.

MARGO: And give them some [unintelligible]

DAVE: The problem of [unintelligible]

MARGO: Absolutely right.

DAVE: And open source can do some of that stuff too; it doesn't solve the problem of someone to come on your site and help you with stuff, which is why the open source business model tends to be, gosh, give this part away for free; charge money for this other part.

BRIAN: Let me give my personal experience with the open source business model. So in 1999 several of us were noticing that a lot of our customers were using a lot of Linux in their environments, and why did we know this? Because they were basically calling up NetApp and asking us what was wrong. And we were sitting there going why are you asking us? It seems there's a problem on your Linux systems. So they walk into their data center and they look around and they go—okay, who do we give money to? And that was NetApp; they were basically seeing Linux as a free operating system running whatever version of NFS. So basically—open source doesn't eliminate the need for the relationship and the responsibility for providing products and solutions that work, and customers kind of have a tendency to basically value those relationships over time. Our reaction by the way to the open source movement and to our customers' movements was to actually ramp up a Linux NFS development team internally and take control of our destinies and contribute to the community to reduce our overall support cost in our center—answering Linux calls that we could fix and get fixes out to the community.

We use open source, we contribute to open source; it's this trend and strategy in the industry. I think it doesn't turn over business models; I mean, I don't know if those books are still on Amazon about how you can make money on free software and all the things came out about 5 years ago.

MARGO: I'll get on my little soapbox which is that I think the term "open source business model" is an oxymoron. Open source is a licensing technology and you can build various different businesses around products that happen to have an open source licensing technology, but open source is absolutely not a business model and so many companies have taken the approach you guys have which is that, okay, we have customers who are using open source operating systems and they don't have anyone to yell at and so we'll give them someone to yell at because they will pay for the privilege of yelling at us.

DAVE: I want to circle back to it just to kind of make a comment on disk technology and the limits of technology; in one sense open source provides an avenue to get technology out, and I made a comment earlier on in the discussion that NetApp's strength is not just getting technology out the door; it's actually getting a product out the door that works and solves a problem. And makes the customer's life better than it was before. So it's not just simply throwing technology over the wall, that's not the point; having double disk protection

wasn't the trick, it was having something that—double disk protection had to be enabled by default and actually does what it's supposed to do. I think the thing, the terabyte disk thing from Costco and stuff, what's the difference between a terabyte disk and NetApp storage is—if I were you I would put your baby's digital photos on the NetApp storage and not the terabyte disk if you want to see them later on. I mean I have—my wife has a picture of me when I was—like an old English bar song—was washing me in a sink surrounded by dishes. And that picture endures because the media endures and it's fairly solid and it's proven technology.

Disks and digital data are ephemeral under many circumstances that I don't think people have come to appreciate yet, but the more and more we put in digital format I think the more and more personal—people are [unintelligible] feel losses of things that were once considered very permanent like photographs and things. I think the differentiation between NetApp and what the people pay for is the difference between the unreliable disk that we know lies repeatedly; and the thing I like about Flash, when I first heard about Flash was, great, a new technology; let's find new ways to lie, right? And that's going to be the challenge of—my interest about Flash, is this peculiar kind of interest in what is different about Flash is that it can fail compared to how disks fail? We've already seen new ways of failure in Flash. And NetApp's job is to basically hide all that from you; hide all that from our customers. And that bar will always keep going up; commoditization will always keep occurring; the level of sophistication regarding the solutions put out there changes; and NetApp's business revolves around continually adding value over things like the competition from open source, which are just standard run-of-the-mill competition.

RIK: I want to circle around for one minute; as you were talking, actually, I was thinking, well, there's something else you actually offer. Before we came out here, remember the guy who created this thing called data monster; his name is Mark Uris and he worked at UCAR.

MARGO: I haven't seen this, it sounds interesting.

RIK: It was an interesting—he needed a SAN, he didn't have enough money, it's partly government, part university and they had a certain budget and they had these huge requirements and so on, so he got something that works. So I said—I ought to contact him because he can tell us how it's still working today 3 years later. I sent him an email and he's no longer there, so I'm wondering how the solution actually works when the bright guy who figured it all out and made it all work disappears? Or he's bored with it now and he wants to walk away and nobody knows how to use the command line.

DAVE: Let me tell you a story. I think it was some students from Berkeley, and I can't remember who it was but this was maybe 10 years ago at NetApp, and these guys came down from Berkeley, a team of like 3 grad students, and they had done a project to build an absolutely enormous storage system and I can't remember how many drives it had in it—1000 or 10,000, something that they thought was enormous, and they were trying to solve the problem of how would one build such a thing and what would be the issues and how do you connect it up and how would you get to all of the storage. And I remember, I was looking at this and it all sounded pretty cool and they were doing this—like wow, grad students, it's great. Then I asked them, so I'm really interested in the question of failure modes; like how do you protect against this kind of failure and that kind of failure. That is the heart of heavy-duty storage.

And they said, you know, we really view that as the zone of research that we were going to dig into and so the answer—punted, 100% punted! They didn't even have RAID! And I was just looking at that; that's not to say that no open source can work; I mean that is not where I was going but when you get one person does one of these things; I mean these are big hard problems. Were you at SOSP?

RIK: No, I wasn't.

DAVE: I went to SOSP in the summer, I slum with the technical people still occasionally and the most hilarious talk—

MARGO: I wasn't there.

DAVE: God, the most hilarious talk was the one done by Amazon about their Dynamo system. So SOSP, ACM preeminent operating system conference; a lot of research going on there, a lot of historically fundamental papers have changed the way people view computing, and this Amazon guy is up there describing his key

redistribution algorithm for their basically distributed storage that they use for their online store. And somebody stands up in the Q&A and says, so I read your paper and I'm looking at your algorithm; can you prove to me that this algorithm actually converges as you add nodes to the storage. And the guy just kind of blinks and says it's converged so far. Everybody's laughing our asses off. And the hilarious part was he was—I mean people took it as his word but what they had produced was a database underlayment for one of the largest online retailers—the largest online retailer? I don't know where Amazon stands—in the world and he basically said here, let me demo the system. He pops open a web browser, goes in one, clicks a book. He goes demonstration finished. It wasn't the theory that was interesting there—but the theory was interesting that the Dynamo system is very clever. It was the actually delivery and productization of a scaled working system that was phenomenal. And I think one of the things that—it's that leaping across from the theory and from the research area to the productization and delivery of a solution; that is what companies and businesses are about and it might separate—some of this discussion we have about technology versus product. But it was the most hilarious talk I sat through in a long time. Especially at SOSP, not a lot of yucks going on there.

RIK: I had a few notes so Margo brought up the wall of copy time and I think just wanted to touch on that for 1 more minute. I guess—at USENIX we have a guy called . . . a past president of USENIX.

MARGO: Kirk?

RIK: Not Kirk. The big Australian.

MARGO: Andrew Hume.

RIK: Andrew Hume, who likes to rant on about disk failures because he works for AT&T and has enormous databases, and he says he checks almost everything he does because he finds when he just copies from one array of disks to another he'll find bit errors.

BRIAN: Absolutely.

DAVE: That was the bit error rate that I was talking about earlier. The one thing I would say about copy times, they have already gotten so long that you really can't assume you can do them in any type of a window in any convenient slot; like you said, what if I've got this big system and I need to copy the stuff to another place? What do I do while that's happening? You have to have a model of copying where while the copy's happening you can still use it.

MARGO: Right, but the question is what happens when my data collection is so big; and how big does it need to be that it takes more than 3 years to copy it?

DAVE: Right, so—

RIK: Because we have exponential growth here and we don't have exponential growth of the disk transfer rate. So we're going to hit that wall at some point.

DAVE: Yes. Currently the limit that's the bigger one now. I mean, a lot of people are still doing tape-based backup which is creating a copy and the kinds of windows that they're doing that is they're maybe not doing full nightlies because that won't fit but they're still doing full weeklies and they're doing incremental nightlies. So what that tells us is we can do a full copy in the weekend slot from all of our disks to tape. So that's the kind of zone that it is and that's breaking down because you'd like nightlies full and what's the recovery time, and that's the underlying technological thing that's driving people from tape-based backup to disk-to-disk backup. So I completely agree that your wall is coming, but all I'm saying is that's not currently the wall that's the big issue; the wall that you would like to get, how do you get this second copy created and then if you're doing—it's not that it's gonna take more than 3 years, it's that it's gonna take more than a day. And so how do you use it while you're doing that? The way that you do it is you take a snapshot, you start copying everything; then when you finish that first copy you look and say, well how much change since the snapshot?

You take another snapshot and you copy everything that changed since the snapshot and then you move that; then you look and say, well but how much change since that? Well, it's sort of like Zeno's paradox. You keep getting closer and closer and eventually you have to do some magic where you say access will shut down for the period to get the last one over and do the transfer; and either you can get that cut over period short enough that you can pretend to still be up and running because you just hold off ops for a little while or else you have to

have a window that's a few minutes or an hour, but for now that's more the nature of the wall that you're hitting as opposed to you physically can't do it.

I guess exponentially that's coming but it's at least enough years off that it's not the front of my list.

MARGO: I mean that problem is the exact same problem that you've got in database replication solutions and you've gotten the virtual machine snapshots; I mean everybody's doing exactly that copying—

DAVE: You have to keep running while you're copying, so in some form or another you have to start with it here in your running and you get it to another place and you're still running and all the—and you can do proxying. I mean one of the things people do is you copy the stuff but still start here and then eventually flip over, but you proxy back for the stuff that's not fresh. I mean those are the kinds of solutions you do—

BRIAN: I think what has changed a lot—let's take the migration to a new system off the table for a second—what has changed a lot is that primarily what people used to do was they had backup windows to tape and that physical operation was critical for their recovery in terms of any disaster or data loss. What any storage solution worth its salt today kind of thing is basically providing mechanisms for giving you reliability, availability, and integrity of your data over larger windows for doing things like offsite tape generation, and providing enough guarantees around that space to lengthen that window more and more to allow you to manage that back end of your process.

I mean that's just been fundamentally grown up around all the storage technologies. The migration thing, I think the saving grace probably today is that there are overlapping ways of storage; nobody comes in and says I'm gonna replace my 1000 petabytes of data tonight and then do it tomorrow; there's overlapping waves and [unintelligible] of cycles running out, and periods where different applications and applications data are moving around. And the strategy now is the transparency of it and increasing that transparency, but continuous migration of the data and then doing a cutover in non-disruptive ways, like I said before. I never sat down and went through the numbers of when this becomes completely impossible.

There's a lot of research going on—going back to our university-based stuff—basically when disks become so cheap and so large and so randomly available that massive redundancy of basically continuous copies with RAID-sure technology; the step beyond RAID; and having redundant copies all over the place, in one sense—except for some consistency issues like Google's storage system—is one of massive distribution. They don't provide a lot of guarantees of that data and then that's not required for the applications they're serving off it, right? I still use traditional storage within their business. But somewhere in there may be solutions and that's part of, I think, some of the reason we had investments within NetApp for very speculative work in our advanced development group of looking at research areas and emerging technologies that are much more highly distributed environments. Something that we are kind of interested in and I know other people are interested in this area too.

There's a potential there. I think the issue practically speaking for customers today with these solutions is one of practicality and manageability of highly distributed diffuse kind of storage approaches and what consistency guarantees are they providing; and then describing which of those—what copy out there is actually the valid one given the high distribution of the data. Which when you're on Google and you go "I'm feeling lucky," what does Google mean, that you're feeling lucky is actually the top hit or was the top hit from before because they just lost a section of their cloud, right?

DAVE: This is an old story but it's right in the zone you're talking about; I was talking ages ago to Alta Vista, which kind of dates it, but they were doing a large data center migration and they were going to shut down—they were going to move all of their service and storage from one location to another location, and I asked the guy how are you going to keep your environment online? You going to replicate all the data and move it; what are you going to do? And what he said is, no we're just going to shut half of it down, put it in trucks, move it over, bring up the service on a half way shut down and then we'll bring over the other half. This was the underlying storage migration method: bring over the other half and then bring it all back online. And I was like, what do you mean you're going to shut down half of people's searches? You just won't get the stuff you were expecting, and he said, oh no, the top 5 million hits we have 20 or 30 copies of; the only time you will notice is if you're searching for something so obscure, it'll be the 100th entry or the 500th entry; and so he said,

realistically if that data is missing for a day, who is likely even to notice? You can try and construct the search that—what would be the search that would get something that's not in the top million or 5 million hits into the top page?

I'm sure if you knew the exact website or something; but I thought that was a really interesting observation about search; but Brian's point is it's going to become true of storage as a whole.

RIK: You mentioned Amazon earlier; I mean they have S3 and the reason he made me think of that—

BRIAN: That's the Dynamo stuff I believe. The Dynamo paper that describes S3.

RIK: Okay, because essentially they say, well you want to store stuff; you can store it, we're going to charge you this much per month per gigabyte.

DAVE: They use it internally but now they're selling it as a service.

RIK: But the deal with it is there's no guarantee that your data will survive.

BRIAN: Because there would be legalities around providing that guarantee and stuff, but people still want to get involved with it. And there's no service level, never mind integrity, but service level about the data probably. Which, by the way, doesn't make it not useful.

RIK: They certainly use S3. I know people have used it because they looked at buying the disks themselves, doing this project, and when they were done with the project they didn't need the disk space anymore so it's like, oh, well rather than having to sink cost we'll just pay to transfer it up there; we'll actually run our apps on the hosted Amazon service as well and the whole thing is going to be cheaper than buying the hardware. So just another model, but this is sort of like storage as a cloud.

BRIAN: So let's be realistic here; we're not breaking any new intellectual ground here. We had monks copying books in little scriptoriums years ago, and the only reason we have the books we have today was because they basically applied the multiple redundancy thing and having enough copies around such that disasters didn't take them all out and the ones that we have are the ones that were copied enough. Sorry—I like monks and scriptoriums by the way.

DAVE: A redundant array of independent scribes.

MARGO: I thought it was redundant arrays of independent monks!

RIK: Probably working in China today.

DAVE: The interesting thing about the storage industry; I made the comment earlier that CIOs don't actually care about storage; they care about some app they have to run that solves some problem they have; increasingly what's cool for us is they're finding they don't want to care about storage but to do what they want to do they have to. Because it's become . . . .the storage that people used to have—I mean you look 20 years ago it almost certainly was financial records about something boring; I mean Wall Street cares at the end of the month you have to close the books; at the end of the year you have to do it because they didn't do quarterly earnings reports yet. But the nature of it was more like that.

Today if you look at—the storage is growing but the stuff we're keeping on it, like I go to the hospital and whether or not they put penicillin in me, which I might be allergic to, depends on the storage. Suddenly I care a lot more about that storage; the personal relevance to me, the immediacy of it matters a lot more, and so suddenly governments are getting involved and regulating how long you have to keep it. Nobody had a regulation that you had to keep any storage 100 years—the life of the patient plus 7 years—it's becoming much more part of the fabric of our lives—the stuff we're storing. Brian mentioned baby photos; so you look . . . .you may not want to care about storage exactly but suddenly you're finding you have to. I mean CIOs but just everybody; it's becoming much more relevant because of what we're keeping there and because it's legally relevant and personally relevant. So it's kind of a cool, fun business.

RIK: I had one more really basic question. You answered it in one way but I want to try to answer it in a slightly different way, and thinking again our audience doesn't have a lot of CIOs, unfortunately; we're not that magazine, not that organization. We have—

DAVE: They're your boss's boss.

RIK: Yes. Essentially the people who read or listen to it are going to be looking for those pearls of wisdom and also how they can convince somebody higher up in the food chain to say, yes, we really want this type of solution because sure, I can build this great [unintelligible] solution for you, but what if I leave? Who is going to run it? The thing I still have questions about is block versus file system; so we have block storage—you essentially have something like, say, Oracle who says fine, stick it all out on iSCSI; oh, by the way, I'll manage it for you; but oh, you can also install Oracle on a file system. From a sys admin's perspective down in the trenches, how do they make these decisions? Is there a performance benefit? I'll do iSCSI because I think it's faster, or I'm going to do a file system because it looks more familiar to me and I can just say, oh here—I could mount this as a file system, say, here are the re-do logs; this is what I need to copy over to this other machine to recreate the transactions. So I can save my ass by doing this, whereas if it were a block device I wouldn't have a clue.

BRIAN: I thought you were starting asking about how do people know NetApp by the way but then you went to technologies; I'll make a comment. I sat down and looked at different product offerings from our competitors and I think one of the clearest strengths that NetApp brings to the table is what we called a unified storage strategy; literally all of our investments around data management and the technologies around de-duplication—snapshots, mirroring capabilities, asset recovery—apply regardless of how you access the storage, and our platform is common regardless of how you decide to access it; whether it be Fibre Channel, SAN, iSCSI, NFS or CIFS. The important part being, should you change your mind, or worse yet, should your application vendor change their mind, then you have a protection in terms of your capital investment and your data management investment which involves processes you built around your data center and your applications for how you provide protection and continuity. Those processes are a common underlayment regardless of how you access data, and I think that's one of our strongest stories.

So in terms of allowing people to sleep at night that are looking at this vast array of tools they could bring out and approaches to deploying storage, I think NetApp has this value proposition that—don't sweat it—if you have to change your mind, which may be out of your hands, you can reconfigure the storage for use in those applications. I think if you get to that point then I think you can sit back and you follow a checklist; that's gonna become increasingly the case I believe—with virtualization—that the application is going to drive underlying choices for deployment and storage. I did a couple of talks in India on advanced development and research and what does NetApp think; and there were professors and students in the audience and I'm saying, by the way, in case no one told you, operating systems research is pretty much dead; no one cares anymore. They look at me, what are you talking about?

It's basically virtualization has allowed people to essentially wrap an application with run-time libraries and operating system on a master; and they manage that unit, that virtual machine instance that they can replicate and move around. And in one sense you really don't want to care too much about how that instance requires the storage and storage infrastructure as long as you can be flexible in how you deploy it. There are actually some fine points about manageability of a TCP/IP Ethernet-based network, and the ability to scale beyond what I think Fibre Channel is capable of physically today that I think virtualization of the data center is going to be forcing some hard questions that heretofore have been pushed off regarding the technology choices of Fibre Channel versus Ethernet, SAN versus file protocols. And I think as you get virtualization, commoditization, this disruption is going to make people step back. And I think they will be looking at more going towards standard solutions around networking and stuff. But I think it's going to proceed from the applications and the applications vendors first; rather than having to sit there and go, what's my fastest solution? because that's in the end for a serious deployment that you're running your business on—I mean performance is one aspect of the whole equation.

DAVE: Fast, cheap, reliable—there's a whole bunch of stuff. So you're asking a pretty pragmatic question; you're a systems admin, say, and you're going to deploy a new app and you're trying to figure out for this app should I use Fibre Channel, iSCSI or NFS let's say. CIFS is seldom used for . . . .

RIK: Anything but Exchange.

DAVE: No, not even Exchange. CIFS is really optimized for home directory. In fact in UNIX if you're interested in using Ethernet you might as well use NFS because it tends to work pretty well; you could use iSCSI but it's not as mature there. In Windows for running apps, CIFS does not have the consistency modeling, so iSCSI tends more to be Windows-centric. But you're trying to figure out, so which one do I want to run? The most common approach is just to do whatever everyone else does. So why bother doing something different because everybody else is doing it? You probably won't get fired that way but it might be less fun; but I mean realistically it's the quick way to do it and it's the safe way to do it. So most people just do what everyone else is doing. Assuming that you'd like to do something on the new side, there's 2 approaches: one is to test the shit out of an approach and the other is the penguin's approach. So the test the shit out of it approach is that you're not sure whether NFS would be good enough; you think that it might be and so you're just really going to test the shit out of it. Is it fast enough? You better test reliability consistence—is it really gonna be cheaper when I leave and other . . . .so you test the shit out of it.

The other approach is the penguin's approach. Here's what I mean by that. Apparently—I saw the documentary—penguins when they're wanting to go into the water they're all standing at the edge of the ice and they're all looking into the water and none of them wants to go in because there could be a sea lion there that's going to eat them. So they're all just kind of shuffling around looking around and eventually as they all shuffle, one of them gets knocked in as they crowd and that penguin—either the surface of the water turns bloody red or the penguin pops his head back up; and once the penguin pops his head back up the rest jump in. And so assuming that you're not the guy who gets to test the shit out of it because you might not have those resources, but you're interested in trying something new that—how do I get my boss to sign up for that. You want to look and see if there's some penguins in the water ahead of you.

So for instance if you're interested in doing an Oracle deployment and you want to know if it's NFS, it would be interesting to know, so who is running Oracle on NFS? Turns out the largest Oracle data center in the world they chose to run on NFS; that might be a fact to share with your boss, right? But then your boss will ask questions, well, about in your industry, and Oracle is Oracle but we're a bank. Okay, are any banks doing it? And then your boss will say yes but they're a bank in Maine and we're a bank in New York; New York is special. I mean just realistically—you're asking kind of a pragmatic question. I know how bosses can be.

RIK: No, that's good.

DAVE: And we've got customer reference programs to try and help people find that kind of stuff because people ask those questions all the time. God, my personal instinct is I'd rather use the iSCSI, but how do I find enough references to prove that it could work in my industry? But typically those are the 2 approaches; either you have a reasonably big budget and you're willing to be the first penguin; either a big budget or a big appetite for risk; like they'll fire me but I'm good, I'll get another job. You either choose to be the penguin or else you look around for one.

MARGO: You either choose to be the front penguin or one of the back penguins.

RIK: That was my list of questions.