Leaders in the storage industry ponder upcoming technologies and trends.

BY MACHE CREEGER, MODERATOR

CTO Storage Roundtable, Part Two

THE FOLLOWING CONVERSATION is the second installment of a CTO roundtable forum featuring seven world-class experts on storage technologies. This series of CTO forums focuses on the near-term challenges and opportunities facing the commercial computing community. Overseen by the ACM Professions Board, the goal of the series is to provide IT managers access to expert advice to help inform

their decisions when investing in new architectures and technologies.

Once again we'd like to thank Ellie Young, Executive Director of USENIX, who graciously invited us to hold our panel during the USENIX Conference on File and Storage Technologies (FAST '08) in San Jose, CA, Feb. 27, 2008. Ellie and her staff were extremely helpful in supporting us during the conference and all of us at ACM greatly appreciate their efforts. —*Stephen Bourne*

Participants

Steve Kleiman—Senior Vice President and Chief Scientist, Network Appliances.

Eric Brewer—Professor, Computer Science Division, University of California, Berkeley, Inktomi co-founder (acquired by Yahoo).

Erik Riedel—Head, Interfaces &

Architecture Department, Seagate Research, Seagate Technology.

Margo Seltzer—Herchel Smith Professor of Computer Science, Professor in the Division of Engineering and Applied Sciences, Harvard University, Sleepycat Software founder (acquired by Oracle Corporation), architect at Oracle Corporation.

Greg Ganger—Professor of Electrical and Computer Engineering, School of Computer Science, Director, Parallel Data Lab, Carnegie Mellon University.

Mary Baker—Research Scientist, HP Labs, Hewlett-Packard.

Kirk McKusick—Past president, Usenix Association, BSD and FreeBSD architect.

Moderator

Mache Creeger—Principal, Emergent Technology Associates.

IBM introduced the System/360 Model 30 with the tape drive system in 1965.

MACHE CREEGER: What can people who have to manage storage for a living take from this conversation? What recommendations can we make? What technologies do you see on the horizon that would help them?

STEVE KLEIMAN: Storage administrators today have tremendous problems that are not adequately solved by any tools. They have home directories, databases, LUNs. It's not just one set of bits on one set of drives; they're all over the place. They've got replicas and perhaps have to manage mirroring relationships between them. They have to manage a disaster recovery scenario and the server infrastructure on the other site if the whole thing fails. They have all these mechanisms for all these data sets that they must process day-in and day-out and they have to monitor the whole thing to see if it's working correctly. Just being able to manage that mess, the thousands of data sets they have to deal with, is a big problem that isn't solved yet.

MACHE CREEGER: Nobody's in the business of providing enterprise-level storage infrastructure management?

STEVE KLEIMAN: The people who have

solved it best in the past have been the backup people. They actually give you a data transfer mechanism that manages everything in the background and they give you a GUI that allows you to say, "I want to look for this particular data set, I want to see how many copies of it I have, and I want to restore that particular thing." Or "I want to know that these many copies have been made across this much time."

Of course, the problem is that it's all getting blown up. So now, it's not just, "What copies do I have on tape? What copies do I have in various locations spread around the world? What mirroring relationships do I have?" The trouble is that today it's all managed in someone's head. I call it "death by mirroring." It's hard. We'll sort it all out eventually.

KIRK MCKUSICK: What do you see as a possible solution?

STEVE KLEIMAN: Currently people are building outrageous ad hoc system scripts—Perl scripts and other types. My company is working on this as are lots of other people in the storage industry, but it's more than a single box problem. It's managing across boxes,

even managing heterogeneously. We have to understand that we're solving the convergence of QoS, replication, disaster recovery, archive, and backup. What we need is a unified UI for handling all these functions, each of which used to be handled for different reasons by different mechanisms.

ERIC BREWER: That is a core issue. How many copies do you have and why do you have them? Every copy is serving some purpose, whether as a backup, or a replication for read throughput, or a cache copy in flash. Because they're automatically distributed you can't keep track of all these things. I think you actually can manage the file system—broadly speaking, storage system, whereby you proactively assign how many copies you have of something.

MARGO SELTZER: Users make copies outside the scope of the storage administrator all the time.

ERIK RIEDEL: Because the amount of data and what it's used for both increase constantly, you have to get the machines to help the users tag content with metadata—to help them know what the data is, what the copy is for, where it came from, why they have it,

and what it actually represents.

MARGO SELTZER: With the data provenance you can identify copies, whether they were made intentionally or unintentionally. That's a start. However, answering the other semantic questions, such as "Why was the copy made?" will still require user intervention, which historically has been very difficult to get.

STEVE KLEIMAN: Each set of data—a database, a user's home directory—has certain properties associated with it. With a database you want to make sure it has a certain quality of service, a disaster recovery strategy, and a certain number of archival copies so that they can go back a number of years. They may also want to have a certain number of backup checkpoints to go back to in case of corruption.

Those are all properties of the data set that can be predefined. Once set, the system can do the right thing, including making as many copies as is relevant. It's not that people are making copies for the sake of making copies; they're trying to accomplish this higher-level goal and not telling the system what that goal is.

MARGO SELTZER: You're saying that you need provenance and you need the tools to add the provenance, so that when Photoshop makes a copy there's a record that says, "Okay, this is now a Photoshop document, but it came from this other document and then it was transformed by Photoshop."

ERIC BREWER: I completely agree with provenance, but I thought you said that it was inherently not going to work because users could always make copies that are not under anyone's control. I think that's the breach and not the observance. Most copies are made by software.

MARGO SELTZER: I agree, but I think that those copies have a way of leaking outside of the domain where things like de-duplication can't do anything about them. What typically happens is I go through the firewall, open up something on the corporate server, and then, as I am about to go on my trip, I save a file to my laptop and take my laptop away. Steve's de-duplication software is never going to see my laptop again.

ERIC BREWER: Yes, and that was my earlier point about managing the data. If you were to go to any system administrator with that scenario they'd get

STEVE KLEIMAN

Over the next decade enterprise-level data is going to migrate to a central archive function that is compressed and de-duplicated, potentially with compliance and whatever other disaster recovery features that you might want.

these big eyes and be really afraid. It should be a lot harder to do exactly what you just stated. That particular problem is perceived as a huge problem by lawyers and system administrators everywhere. The leakage of that data is a big issue.

STEVE KLEIMAN: Companies that actually own the end-user applications will have to set architectures and policies around this area. They'll certainly sign and possibly encrypt the document. Over time, they will also take responsibility for the things that we have been talking about: encryption, controlling usage, and external copies. Part of this problem is solved in the application universe and there are only a few companies that are practical owners of that space.

MARGO SELTZER: There are times when you want that kind of provenance and there are times when you really don't.

MACHE CREEGER: There's going to be a hazy line between the two. Defining what is an extraneous copy or derivation of a data object will be intimately tied up with the original object's semantics. Storage systems are going to be called on to have a more semantic understanding of the objects they store, and deciding that information is redundant and delete-able will be a much more complex decision.

STEVE KLEIMAN: The good news is the trend for end-user application companies, such as Microsoft, is to be relatively open about their protocols. Having those protocols open and accessible will allow people to leverage a common model across the entire system. So, yes, if you kept encrypting blindly you'd defeat any de-duplication because everything is Klingon poetry at that point. I should be able to determine whether two documents that are copied and separately encrypted are the same or not. I'm hoping that will be possible.

MACHE CREEGER: What recommendations are we going to be able to make? If IT managers are going to be making investments in archival types of solutions, disaster recovery, de-duplication, and so on, what should they be thinking about in terms of how they design their architectures today and in the next 18 months?

STEVE KLEIMAN: Over the next decade enterprise-level data is going to mi-

grate to a central archive function that is compressed and de-duplicated, potentially with compliance and whatever other disaster recovery features that you might want. Once data is in this archive and has certain known properties, the enterprise storage manager can control how it is accessed. They may have copies out on the edges of the network for performance reasons—maybe it's flash, maybe its high-performance disks, maybe it's something else—but for all that data there's a central access and control point.

MACHE CREEGER: So people should be looking at building a central archival store that has known properties. Then, once a centralized archive is in place, people can take advantage of other features, such as virtualization or de-duplication, and not sweat the peripheral/edge storage stuff as much.

STEVE KLEIMAN: I do that today at home, where I use a service that backs up all the data on my home servers to the Internet. When I tell them to back up all my Microsoft files, the Microsoft files don't go over the network. The service knows that they don't have to

copy Word.exe.

MARY BAKER: I'm going to disagree a little bit. One of the things I've been doing the last few years is looking at how people and organizations lose data. There's an amazing richness of ways in which you can lose stuff and a lot of the disaster stories were due to, even in a virtual sense, a centralized archive.

There's a lot to be said for having those edge copies under other administrative domains. The effectiveness of securing data in this way depends on how seriously you want to keep the data, for how long, and what kind of threat environment you have. The convenience and economics of a centralized archive are very compelling, but it depends on what kinds of risks you want to take with your data over how long a period of time.

MARGO SELTZER: What happens if Steve's Internet archive service goes out of business?

STEVE KLEIMAN: In my case, I still have a copy. I didn't mean to imply that the archive is in one location and that there's only one copy of that data in the archive. It's a distributed archive, which has bet-

ter replication properties because you want that higher long-term reliability. From the user's point, it's a cloud that you can pull documents out of.

ERIK RIEDEL: The general trend for the last several years is for more distribution, not less. People use a lot of high-capacity, portable devices of all sorts, such as BlackBerrys, portable USB devices, and laptops. For a system administrator, the ability to capture data is much more threatening today. Five or 10 years ago all you had to worry about were tightly controlled desktops. Today things are a great deal more complicated.

I was at a meeting where someone predicted that within two or three years, corporations were going to allow you to buy your own equipment. You'd buy your own laptop, bring it to work, and they'd add a little bit of software to it. But even in the age in which corporate IT departments control your laptop and desktop, certainly the train has left the station on BlackBerrys, USBs, and iPods. So for a significant segment of what the administrator is responsible for, pulling data back into a central store is not going to work.

MACHE CREEGER: That flies in the face of Steve's original argument.

STEVE KLEIMAN: I don't think so. I do think that there will be a lot of distributed data that will be on the laptops. There will be some control of that data, perhaps with DRM mechanisms. Remember, in an enterprise the family jewels are really two things: the bits on the disks and the brain cells in the people. Both are incredibly important and for the stuff that the enterprise owns, that it pays its employees to produce, it's going to want to make sure those bits exist in a secure place and not just on somebody's laptop. There may be a copy encrypted on somebody's laptop and the enterprise may have the key, but in order for the company to assert intellectual property rights on those bits, you are going to have to centrally manage and secure them in some way, shape, or form.

ERIC BREWER: I agree that's what corporations want, but the practice may be quite different.

STEVE KLEIMAN: That's the part I disagree with because part of the employee contract is that when they generate bits that are important to the company, the

The DEC Hard Drive Assembly, circa 1970.

company has to have a copy of them.

GREG GANGER: Let's be careful. There are two interrelated things going on here: Does the company have a copy of the information, and can a company control who else gets a copy? What Erik just brought up is an example of the latter. What Steve has been talking about is more of the former.

STEVE KLEIMAN: Margo has been saying that companies may not have a copy. I fundamentally disagree with that. That's what it pays the employees to generate. The question is, can the company control the copy? My working assumption is that this is beyond the scope of any storage system. DRM systems are going to have to come into play and then it's key management on top of that.

MARGO SELTZER: I'm not sure I buy this. Yes, companies care that employees do their jobs, but very few companies tell their employees how to do their job. If my job is to produce some information and data, I may be traveling for a week and it may take some time for that to happen. In the meantime, I may be producing valuable corporate data on my laptop that is not yet on any corporate server. Whether it gets there or not is a process issue and process issues don't always get resolved in the way we intend.

MACHE CREEGER: You're both right. Margo wants to create value for her company in whatever way she is comfortable—on a laptop while she's traveling, at home—whichever way works that produces the highest value for her employment contract. If the company values Margo's work, they will be willing to live, within reason, with Margo's work style.

On the other hand, from Steve's perspective, sooner or later, Margo will have to take what is a free-form edge document and check it into a central protected repository and live with controls. She can then go on to the next production phase, which might be a Rev. 2 derivative of that original work, or perhaps something completely different.

ERIK RIEDEL: You certainly have to be careful. You're moving against the trend here. The trend is toward decentralization. Corporations are encouraging people to work on the beach and at home.

ERIK RIEDEL

The general trend for the last several years is for more distribution, not less. People use a lot of high-capacity, portable devices of all sorts, such as BlackBerrys, portable USB devices, and laptops. For a system administrator, the ability to capture data is much more threatening today.

STEVE KLEIMAN: Nothing I've said is in conflict with that. Essentially, the distilled intellectual property has to come back to the corporation at some point.

MARGO SELTZER: Sometimes it's the process that's absolutely critical. Did I steal the code or write it myself? That information is only encapsulated on my laptop. Regardless of whether I check it into Steve's repository, when Mary's company comes and sues me because I stole her software, what you really care about is the creation process that did or did not happen on my laptop.

ERIC BREWER: I don't think that's the day-to-day problem of a storage administrator. What we're talking about is whether the first goal is to know which of the copies you don't want to lose, which is a different problem than copies leaking out to others.

STEVE KLEIMAN: I do think that the legal system still counts. Technology can't make that obsolete. You still have a legal obligation to a company. You still have an obligation not to break the law. Any technology that we can come up with, someone will probably find a way of circumventing it, and that will require the legal system to fill in the gaps. That's absolutely true with all the stuff on laptops that we don't know how to control right now.

MARGO SELTZER: I also think it's more than just copies that we need to be concerned with; it's also derivative works, to use the copyright term. It's "Oh, look: File A was an input to File B which was an input to File C and now I have File D, and that might actually be tainted because I can see the full path of how it got there."

MACHE CREEGER: Maybe what we're seeing here is that we need to intuit more semantics about the bits we are storing. Files are not just a bunch of bits; they have a history and fit in a context, and to solve these kinds of problems, companies are going to have to put processes and procedures in place to define the context of the storage objects they want to retain.

MARY BAKER: You can clamp down to some extent, but it's the hidden channel problem, even through processes that are not malicious. Say I'm on the beach and the only thing I've got is a non-company PDA and I have some ideas or I talk to somebody and I record something. It can be very hard to

bring all these different sources into a comprehensive storage management policy. Storage has gotten so cheap; it's in everything around us. It's very easy to store bits in lots of places that may be hard to incorporate as part of an integrated system.

STEVE KLEIMAN: There's not just one answer to these problems. Look at what happens in the virus scanning world. It's very much a belt and suspenders approach. They do it on laptops, on storage system, in networks, and on gateways. It's a hard problem, no doubt about it.

There are a variety of technologies for outsourcing markets, such as China and India, where people who are working on a particular piece of source code for a particular company are restricted from copying that source code in any way, shape, or form. The software disables that.

Similar things are possible for the information proliferation issues we have been talking about. All these types of solutions have pros and cons and depend on what cost you are willing to pay. This is not just a technological issue or a storage issue; it's a policy issue that also includes management and legal issues.

ERIC BREWER: In some ways it's a triumph of the storage industry that we have moved from the main concern being how to store stuff to trying to manage the semantics of what we're storing.

MACHE CREEGER: Again, from the storage manager's standpoint, what is he to do? What should he be doing in the next 18 to 24 months?

STEVE KLEIMAN: Today people are saving a lot of time, money, and energy doing server virtualization and storage virtualization. Those two combined are very powerful and I think that's the next two, three, or four years right there.

GREG GANGER: And the products are available now. Multiple people over the course of time have talked about snapshots. If you're running a decent-sized IT operation, you should make sure that your servers have the capability of doing snapshots.

ERIC BREWER: On the security side, encryption. Sometimes there are limited areas where you can do the right kind of key management and hierarchies, but encryption is an established way

in the storage realm to begin to protect the data in a comprehensive way.

MARGO SELTZER: Backup, archival, and disaster recovery are all vital functions, but they're different functions and you should actually think carefully about what you're doing and make sure that you're doing all three.

GREG GANGER: Your choice for what you're doing for any one of the three might be to do nothing, but it should be an explicit choice, not an implicit one.

ERIK RIEDEL: And the other way around. When we're talking about energy efficiency, being efficient about copies, and not allowing things to leak, then you want to think explicitly about why you are making another copy.

ERIC BREWER: And which copies do you really not want to lose? I differentiate between master copies, which are the ones that are going to survive, and cache copies, which are ones that are intentionally transient.

GREG GANGER: For example, if you're running an organization that does software development, the repository, CVS, SVN—whatever it is that you're using—is much more important than the individual copies checked out to each of the developers.

ERIC BREWER: It's the master copy. You've got to treat it differently. No one can weaken your master copy.

MACHE CREEGER: I know that the first CAD systems were developed for and by computer people. They did them for IC chip and printed circuit board design and then branched out to lots of other application areas.

Is the CVS main development tree approach going to be applicable to lots of different businesses and areas for storage problems or do you think the paradigm will be substantially different?

GREG GANGER: It will absolutely be relevant to lots of areas.

ERIC BREWER: I think most systems have cache copies and master copies.

GREG GANGER: In fact, all of these portable devices are fundamentally instances of taking cached copies of stuff.

ERIC BREWER: Any device you could lose ought to contain only cache copies.

MARGO SELTZER: Right, but the reality of the situation is that there are a lot of

portable devices you can lose that are the real copy. We've all known people who've lost their cell phone and with it, every bit of contact information in their lives.

GREG GANGER: They have learned an important lesson and it never happens to them again.

MARGO SELTZER: No, they do it over and over again, because then they send mail out to their Facebook networks that says "send me your contact information."

MACHE CREEGER: They rebuild from the periphery.

ERIC BREWER: The periphery is the master copy; that's exactly right.

MACHE CREEGER: We've talked about security and storage infrastructure. We've touched on copyright, archival, and talked a lot about energy. We've talked about various architectures and argued passionately back and forth between repositories and the free cloud spirit.

Storage managers have a huge challenge. They don't have the luxury of taking the long view of seeing all these tectonic forces moving. They have to make a stand today. They've got a fire hose of information coming at them and they have to somehow structure it to justify their job. They have to do all of this, with no thanks or gratitude from management, because storage is supposedly a utility. Like the lights and the plumbing, it should just work.

STEVE KLEIMAN: They have a political problem as well. The SAN group will not talk to the networking group. The backup group is scared that their jobs are going to go away. Looking at the convergence of technologies, even for something simple like FCoE (Fibre Channel over Ethernet), the SAN Fibre Channel people are circling the wagons.

MACHE CREEGER: Or iSCSI over 10 gigabit Ethernet.

STEVE KLEIMAN: Absolutely. There are a lot of technical issues in there, but there are very serious people and political issues as well.

Mache Creeger (mache@creeger.com) is a longtime technology industry veteran based in Silicon Valley. Along with being a columnist for *ACM Queue*, he is the principal of Emergent Technology Associates, marketing and business development consultants to technology companies worldwide.