

BACKUP TRANSFORMATION FROM THE INSIDE

EMC's Chief Oracle Architect Speaks Out



"If your backup infrastructure can't keep pace with the explosion of data, or match the mobility of cloud computing, your database is going to suffer... So, it's absolutely important to make sure your backup infrastructure is much more agile and much more dynamically scalable than it is today."—Darryl Smith

What is backup transformation and why is it so critical to an organization's health and well-being? In this Q&A, Darryl Smith, EMC chief oracle architect and veteran of the backup wars, provides an insider's view of the need for and impact of backup transformation within EMC IT and on our business.

Smith, who is responsible for all of EMC's databases, primarily Oracle but also SQL Server, MySQL, Postgres, and Greenplum, describes the people and process changes EMC IT has gone through and the game-changing effects they have had. He also takes a look at what lies ahead and gives his view of backup nirvana.

For more information and resources, be sure to check out www.emc.com/BackupLeader.

AS AN ORACLE ARCHITECT, WHY IS BACKUP IMPORTANT TO YOUR PEACE OF MIND?

I'm responsible for making sure EMC databases are healthy, up and running, and able to service our business users to get them the information they need. If I, or my team, don't show up for work, everything stops. If the database supporting your CRM application, or your customer service application, or whatever application you're running, isn't functioning, your application's not working.

And for me to do my job, backup is critical. One of our main struggles as we began our journey to the cloud was that our backup infrastructure wasn't keeping pace. We were relying on legacy-based infrastructures and on point-to-point backup solutions. That was a problem for my peace of mind—and it's a problem for a lot of organizations today.

SO, BACKUP TRANSFORMATION IS A BIG DEAL?

Yes. It's really something that we have to do. You can't continue to back up the way that you've always done in the past because the old way doesn't scale, and it doesn't work in virtualized cloud infrastructures. The reason is that all applications, from small conference room scheduling systems to your most mission-critical ERP or customer service-type applications, really revolve around data. And so it is critical that that data is protected, and without backup transformation this isn't possible.

AND THIS CAN AFFECT CLOUD AND VIRTUALIZATION EFFORTS?

Right. If your backup infrastructure can't keep pace with the explosion of data, or match the mobility of cloud computing, your database is going to suffer. And this

could mean anything from reduced application performance to data loss or even complete application failure, which obviously can be critical for a business. So, it's absolutely important to make sure your backup infrastructure is much more agile, and much more dynamically scalable, than it is today.

WHAT HAVE YOU DONE OVER THE LAST COUPLE YEARS TO TRANSFORM BACKUP?

Today, from a server perspective, we are 94% virtualized today. From a database perspective, we're around 84% virtualized, and at this point we are virtualizing as quickly as possible. And this has big backup implications.

When we started, we used traditional backup methodologies. We would back up using traditional RMAN to our backup scheduler, EMC NetWorker, and back it up to either tape devices or disk—virtual tape devices backed by disk. Tapes, obviously, are a very old technology, so we replaced those with virtual tape libraries. But even those didn't scale out, and we were constantly measuring, quantifying and restricting the use of these resources.

That was really the core of our problem: We had a limited number of resources, a growing number of databases, and a growing amount of data that needed to get backed up. To keep pace with our virtualization, we've basically gotten off all of our legacy platforms. Even our NetWorker storage nodes have been virtualized.

WHERE DOES YOUR BACKUP TRANSFORMATION STAND TODAY?

So now, 24 months in, we're moving toward scaling out our infrastructure. Our tape devices are gone, and we're replacing our virtual tape devices with Data Domain appliances, which have much more capacity and much more throughput, as well as the ability to store data much, much longer due to the deduplication. And that first step has really helped in reducing the restrictions on our capacity. We now know that if we need to get a backup restored, we can get it back fairly quickly and not have to rely on tapes and off-site storage. But we're still managing backup like we had some of those old restrictions, and that's really where we need to start taking this to the next level.

Also, we've done some "database-as-a-service" work, where we were able to fully automate backup and recovery so it's basically hands-off. Back when we didn't have this big data sprawl, when we didn't have mobile computing, we were able to set up our backups and kind of forget about them, and that was nice. Now, life is a whole lot more complicated but by being able to automate our backups, and fully script it, we no longer worry about whether the backup takes place, because we've removed the variables.

We're doing the same thing with some of our very big databases. We've worked with the backup teams and engineered an offloaded backup. By that, I mean we take a storage clone, mount it on a proxy server or an NetWorker storage node, and back up the database up from there. In that case, the backup database or the proxy server is connected directly to the Data Domain, and so we don't have resource constraints. And so those backups now work flawlessly, and they work at very high speed as well.

So, that's two examples of where we've been working with the backup team in order to provide a backup that is reliable and scalable, that we can then basically configure it and forget about it.

SOUNDS LIKE YOU'RE WORKING DIFFERENTLY WITH YOUR BACKUP TEAM. HOW HAVE PROCESSES CHANGED?

Replacing our infrastructure is only a starting point in transforming how our backup infrastructure works. We really need to work more on process and procedures. It's good to have, for instance, an enterprise-wide scheduler. But trying to manage individual components and resources as a manual effort gets in the way of being able to scale out dynamically, as well as give you the agility that cloud computing really requires.

Really, the key to this whole transformation is melding of all of these individual groups and members and skills into one greater skill set. We talk about this all the time when we talk about IT transformation—that all the skills need to come together. And that doesn't necessarily mean that one person has to know it all, but people with different skill sets need to work closely together to really become part of that core architecture team and at the same time, that knowledge gets shared.

HOW HAVE RELATIONSHIPS WITH THE BACKUP TEAM HAVE CHANGED?

Well, the relationship was becoming strained. We were having many difficulties with backups, and the backup teams were constantly responding to issues. The strain really was on both sides—the backup team, as well as the DBAs, was very frustrated. They were dealing with day-to-day issues and constantly responding to problems. It was a constant problem, and it not only strained relationships but also our ability to provide a service to our business users.

The backup team has typically been more of a backroom team. You don't hear about them; they're not really part of the bigger IT story; they're not part of the decision-making process. They just do the backups. But that kind of thinking or culture doesn't really work anymore. The backup team really needed to become more a part of the bigger IT team and help with the overall IT transformation. If they don't, you're going to have problems. Your databases aren't going to run the way they should—they're going to be held up trying to write off backups. In the event of a failure, you're not going to be able to recover your systems or your data. Backup really has to be more integrated into the mainstream than it ever has been before.

WHERE ARE YOU HEADED? WHAT'S BACKUP NIRVANA FOR YOU?

Well, we're not 100% at backup-as-a-service today, but we are using it in a limited form. For instance, our other IT-as-a-service offerings, infrastructure-as-a-service, as well as our database-as-a-service offerings all make use of backups, and these backups are fully automated and provided as a service.

So, the endgame, or the goal of our transformation of our backup infrastructure, is that the backup administrators will be responsible for the infrastructure, for the capacity planning, to make sure that there's enough capacity for us. As for me, as a consumer, as the DBA who's consuming these backup services, I need to be able to

configure my own backups, run them at will, and be able to do restores as necessary without having to rely on another group to make sure that something's working. So my nirvana is basically that I am running a console where I can schedule all my database backups and not worry about whether or not I have capacity.

It's kind of like power. When I turn on the lights, I expect the lights to come on, because I have enough power coming to them. I would like my backups to be just as easy.

WHAT EMC TECHNOLOGIES ARE YOU'RE USING? HOW HAVE THEY CHANGED THE GAME?

One of the first steps we took was buying EMC Data Domain systems, even before EMC bought the company. We chose them because of deduplication. Of course, we had Avamar already, and that does some pretty good deduplication. It's great for backing up operating systems and desktops, and we had a large investment in Avamar to do that. But for databases, we needed something with a much higher capacity, and that's why we went with Data Domain for our databases.

We've always had NetWorker, which makes backups a lot easier. DBAs are used to using RMAN to do backups, but RMAN has to write to a tape device. Operating systems don't typically have built-in tape devices. So NetWorker acts as that virtual tape device for the RMAN backup, and so that's fully integrated for us, and the DBA can just kick off a backup, and NetWorker, through the storage nodes, will back it up to, in our case, the Data Domain device. That integration is very important.

One of the newer technologies that we're looking at is DD Boost from Data Domain. That also integrates with RMAN. And so the DBA can use their traditional RMAN backups, get the benefits of a product like DD Boost, without really having to know anything about the product, but it puts full control in the DBA's hands.

Another reason, we're working on aggressively rolling out DD Boost, is it reduces the resources required to do our backups, because a lot of that backup deduplication gets offloaded on the database server itself, so I'm pushing a lot less data across my network. This is going to allow me to back up many more things at once, whether that is databases or operating systems. But more important, it's going to mean that my backups can travel a bit farther. So now I can have infrastructure that's going to fit the mobility of cloud computing.

The other thing we're doing is using Data Domain appliances to back up to directly. So we can use products like VMware Data Director, configure some of our data stores to be Data Domain NFS mounts, and just take storage clones and write them straight to the Data Domain appliance. And we can do the same thing with tools like Export or Data Pump from Oracle, or manual backups. In SQL Server, sometimes we just do manual file system backups that go straight to Data Domain.

EMC², EMC, and the EMC logo are registered trademarks or trademarks of EMC Corporation in the United States and other countries. VMware are registered trademarks or trademarks of VMware, Inc., in the United States and other jurisdictions. © Copyright 2012 EMC Corporation. All rights reserved. Published in the USA. 11/12 Handout H11297.

EMC believes the information in this document is accurate as of its publication date. The information is subject to change without notice.