AICPA CITP Credential Examination Series

Topic: Data Backup & Recovery

Presenters: Chris Fraser and Kevin Martin

Chris Fraser: Hello, and welcome to the AICPA CITP Credential Examination series. This podcast will assist you in preparing for the examination specific to Data Backup & Recovery topics. By way of introductions, my name is Chris Fraser. I am an AICPA CITP credential holder based out of St. Petersburg, Florida, where I'm employed by Optum360 UnitedHealth Group. I am joined today by my peer and colleague, Kevin Martin.

Kevin Martin: Thank you, Chris. I'm excited to be here today. I also am a CITP holder and I'm based out of Cincinnati Ohio. I'm currently employed with Martin and Associates, a firm that I'm a 50% business partner in, and delivering IT services for the last 28 years. We are pleased for this opportunity to share insight in the surrounding areas of data backup and recovery.

Chris: Great, let's get started. Kevin, I thought it would be interesting to talk about how we got to data backup today, because you and I both started back when technology was a little bit different. I remember going through processes where we had to schedule tape backups.

We went through a grandfather-father tape cycle process, because the cost and transportation issues-- You'd actually have delivery services who are moving tapes around. There was not the same level of redundancy and or recovery that we have today, so I'm excited by the new technologies and the new advantages that we have.

Kevin: Great. A thing that I remember a lot is on the recovery side alone. In today's environment, what you're able to do on a recovery mode, is you can get back to something immediately. Sometimes within a minute, sometimes within a half hour, depending on the size of what you're bringing back. In the days of tape, if you wanted something from a month ago, you might have to wait a day for the tape to be delivered. And then have a couple hour process to read off of that tape to hope that the files that you are looking for were there.

Chris: I always looked at the perspective when I talk to the companies would be, how long can you afford to be down? Part of that was, how long does it take you to get back up? We've made some good progress with new technologies. I'm excited where that has reduced that issue significantly.

Kevin: The one point that you brought up there is a key policy that our end users and customers have to understand, is how long can you be down? That really is one of the things that dictates the most what your costs are going to be, which will be the first threshold that you're going to look at to say, "Is there a return on investment for this cost versus how long you're going to be down."

Chris: Good point on that. If you had an infinite budget you could obviously have really great backup systems, but most companies are bound by real constraints, and so they

need to make the individual decision on what kind of a recovery process is worth it to them.

Kevin: Correct.

Chris: Kevin, why don't we talk a little bit about today's backup policies and procedures?

Kevin: The thing that is most important for me that I keep stressing to all of my customers is, once you get those policies and procedures in place, you really need to make sure that you have a procedure in place to make sure that those policies really get enforced every day.

There's so many times that people say, "Yes, we understand. We need to back this up once a day or every hour depending on what the nature of the backup is," but then they just don't put in place any type of procedure to say what happens or what type of alert do they get when that process is not done so that they can react to it.

Chris: Good point, so what happens if the procedures are not followed? Is there a good process in place to alert you of that? Another factor I find is, since data has gotten so cheap, the storage is immense.

A lot of times you have to make sure that is it worth your while to backup everything, or do you want just certain important data, and then is the data being stored in the right location? I still see in today's world that people might use their local hard drive when the backups are only being done on a server or a shared drive.

Kevin: I have a client that part of their backup problem that they were solving over the last six months was, part of their data sizes were so large on a couple of their image-based files that fed into their inventory that whenever they had to roll back to go get the data, or push the files up to a vendor, it simply took too long in the standpoint that it might take three or four hours to get the data pushed up, that there was no efficient way for them to then do a live test to do that. They had to revisit what type of bandwidth they were using to get that to work.

Chris: A good hands-on story. You also make me think of two important considerations I see. One is, you have a process, what I call the data gardening. As that data grows, and grows, and grows, I know people that have ancient data that didn't even know why they have it anymore.

If nobody is actively trying to cull through that and keep it current or relevant, it grows at an enormous rate and can become a real nightmare to worry about. And then, are they testing the recovery to make sure they are getting what they expect back.

Kevin: Correct, and while they're doing that recovery, and they are doing that test to the recovery, that is a wonderful time to look at the data that you do have archived so that you can start purging.

As you said, pruning that data out of your archives so that you don't end up with 15 copies of the same thing. And then not sure of which one's the master copy or how

much time you need to look at those other 14 to make sure you're now not going to delete everything.

It is a lot easier to keep things cleaner every time you touch it versus doing a quick touch not looking at the dirty part, and then a month later or five years later, you would just realize how much dirty data that you have.

Chris: That's actually one of the projects I have right now. There's these terabytes, and terabytes of data with folders and nobody can explain what's what. It's obviously easier to take care of that before it gets bad instead of after.

Kevin: Yes, very much so. On the other part of the frequency part of the data backups, the one thing to keep to mind as you're putting together your backup policy and your backup plan, is you need to settle into a frequency that you understand not just what time of the day things are happening, but you also understand what impact that time of the day means if you need to roll anything back.

We have certain clients that run nearly a 24/7 environment. If they're doing backup just once a day at the tail end of that 24 hours, but then the first eight hours of the next day someone has a major issue, they really have to evaluate how they fix that in relation to their backups. The frequency also comes into play with what people need to do in relation to knowing where their data is at a certain point in the day.

Chris: That's a good point, Kevin. Along the lines of that, I'm wondering now, are we able to backup all the right data? Certain things are now live in making sure you can update those, I'll say, ever changing or constant databases that are open at the time. My understanding of technology has gotten better about handling that, but I still see that as an issue.

Kevin: Technology does handle that, that you can backup all that. The issue that you have is really two things. One is going to be the cost of that might be relatively exorbitant compared to what you need.

Two, understanding that how you do bring something back, then, three minutes after that time period, or 30 minutes after that time period, and how that affects the rest of the work in between there and roll back.

SQL does a good job with its transactional logs within the SQL database set. You can bring back logs and then you can roll backwards the log which basically acts like rolling back the clock, minute by minute by minute, transaction by transaction by transaction. You're literally rewinding back to a point that you want to get to exactly.

Chris: I can see that being real important with an accounting system, because you don't need to miss a transaction or duplicate a transaction.

Kevin: Correct. When you say accounting, that always means inventory, sales order, manufacturing, project management, service calls. It's not just the debits and credits of what's going in the general ledger or the accounts payable check that you cut. It's everything related to your enterprise resource or ERP or CRM system that has all those data points.

Chris: Right, agreed. One other thing I found interesting, and I'm happy about the technology, is that we have this almost shadow copy and real time replication going on. That way you're almost getting constant backups being made. Have you seen that?

Kevin: Yes. We just had a case yesterday in our office that one of our staff had deleted one of the virtuals that they were trying to get some production code organized in. No one realized that the virtual got deleted for about three weeks. The good news was, is because nothing was being done on that for the last three weeks, we had to shadow backup from a few days before that three week time period.

We're able to pull that shadow backup and pull down the project and the source code, and to be able to move on. We might have lost about 40, 45 minutes finding what we needed, and the stress related to it, but overall, we ended up saving about 40 hours of not having to redo that project.

Chris: Always good to hear a success story. You also remind me of one other thing that I'm seeing as an issue with the frequencies. The ransomware has become so malicious, that a lot of times it's encrypting and not telling you with the design that you're going to actually make backups of these encrypted ones. By the time you realize you've been infected, you don't have a good backup to go to.

Kevin: Yes, that's a big problem now. A lot of companies now are starting to face is those time bomb ransomwares that go off three days later after they've had a chance to propagate through the master files and then be backed up a couple nights. You really need to make sure that within your backup policies and the type of backup that you're doing, that all the appropriate safeguards from malware and ransomware are all strictly enforced.

We had something this week or last week with a client that we were trying to do part of an update on their system. Their system wouldn't let part of an update go on because there was a file that was already infected inside where we were going to install to.

The backup software, backup mechanism, that was going to shadow that said, "No, you can't install that because we won't let that current image come up to our remote site for being backed up." You'll be at risk. That was an interesting warn you before you replicate bad data up to your good backup. The system was taking care of that. That was nice.

Chris: I also like to do what I call a "Milestone or Archive Checkpoint." Even though you're doing a series of constantly updating backups, make your snapshot at year end, make a snapshot at a quarter end. You'll always have those to fall back on also if something catastrophic happens.

Kevin: With that point, the one thing that the small businesses have the biggest problems with, is they just don't have the resource and the technical skills to do that correctly on a timely basis so they know that they're covered. Some of the smaller clients are just running a little bit more blind of making sure that they have a backup, but not really doing the right thing in testing a full restore, and such like that.

Chris: Now Kevin, you bring up a good subject. That leads into what I would think of as the types of data storage. Especially with costs coming down, I'm a big favor of multiple options. Having your Cloud backup, and whether it's the Dropbox type stuff, Carbonite, OneDrive, all those kind of items.

Add in your own hard drive backup you take off-site or even some of the things-- say for my personal use, I'm still backing up to a DVD, because I know ransomware can't overwrite that once I've written it once.

Kevin: You just locked in to a very, very important part there. When you're writing your backup, if you can write it to a device like a DVD or some other device that you can put a lock on, that nothing else can write on top of, that gives you a great safeguard. The other point you brought up of doing things, in essence, in two mediums, is something that we adhere to, both individually within our firm.

I backup part of my data to a hard drive. Part of my data, or the same data's backed up in the Cloud, but I have access to that hard drive not to be on the internet. If I ever need to get back to anything, I have that safe. Our office does, basically, the same thing. We're backing up to a Carbonite-type storage class device out in the Cloud, but then we have hard media in-house that we rotate through on a weekly basis.

Chris: Just a couple other things that I'm obviously using, I've got my own NAS. You can rotate drives out of that if you need to. Obviously, the days of tape to me have come and gone. That the amount of data you can store on the hard disk has grown dramatically, and the costs have come down enough. That's great for these large amounts of data.

Kevin: Overall, in the last three or four years, the only people buying tapes are people that had existing tape backup sets they still have life on and they can get big enough tapes for them, but everything new is not in that media.

When you opened up and talked about how things used to be back in the day, I was talking to a client the other week that actually remembers the good old days of accounting software coming on floppy disk. That might have been 26 floppy disks.

You're going through an install and the 16th disk has a bad sector in it. You got to wait to get a 16th disk. Or you're backing up, and you know that's going to take 26 floppy disks , but it locked up on the 13th disk and you can't finish your backup. Some of those older media, it is nice to see go away and tape is heading in that direction.

Chris: One of things we've also talked about was using off-site or Cloud type backups. One component that I don't know is always well-explained is, a lot of times you need a seed device, just because of the amount of bandwidth that you get for the internet.

Now, a lot times when we're setting up an initial one, we will do a dump to a portable hard disk, send it up to the Cloud location where they get the starting point. From there, they can do the differentials for the backups online.

Kevin: With that seeding, the biggest thing you're saving is that initial load of not having to push it through your internet connection. As more and more businesses get to gigabyte fiber type connectivity or 300 mbps, if that's what they're at, if they're not

on a full gig, part of that seeding is a little bit easier now that you can seed maybe over a couple nights. Pushing the stuff over the internet versus pushing it to a disk or some type of USB device and sending it to it. They're all very good options, but the seed thing definitely is needed in today's large amount of data that so many of our customers have.

Chris: Right, and then the other component of that is if you do have a complete failure where your system goes down and you need to restore from the Cloud, bandwidth, again, becomes an issue. And sometimes you need to look at doing a portable seed device that goes back the other way.

Kevin: Yes, very much so. We do that a lot of times at clients when we need to test major upgrades for a customer. It's just easier to seed the backup on-site and then walk back to your office, and then have everything you need on that seed versus trying to download off of that.

Chris: Kevin, one of the things that I'm seeing a lot of these days is being able to use virtual systems, virtual servers. You can now just make a backup of that whole virtual disk. I'm interested to get your thoughts on what you see going on with that.

Kevin: A lot of clients will use that as a second type of backup. If they're a heavy SQL environment with their accounting software, and it's running on a virtual server, the virtual server's running its traditional backups, from a data standpoint that needs to be done. But then having that virtual being screenshot or saved once a night, and archived off in weekly rotations, is very nice to have.

The other nice thing about a virtual system is, if you need to do any type of testing or moving something from one server to another because the server is going down or being replaced it's a lot of times now easier to move virtuals versus just trying to move data and programs, then having to do any type of reinstall or configuration.

Chris: I agree. It gives you some nice options, and I sometimes find you may need to-- If you need that granularity returned, restoring a single file, certain backups work better for that versus we have to restore the entire server as soon as possible. Just restoring that virtual hard disk image gives you the advantage for that. They're different in their trade offs.

Kevin: The other thing too with the trade off of the virtuals as well is, you really need to pay attention to what operating system your virtual is on so that you don't accidentally have it become incompatible to move to the next generation of a virtual server.

You might have had a virtual running in the 2003 environment that would work on a 2008 environment, but it won't work in a 2012 server environment. You need to make sure that you don't let your virtual operating systems lag too far behind in Microsoft's road map of what they support.

Chris: That's a good point. I've also sometimes thought about people going back to archives to restore things and then realizing they don't have the application that it goes with anymore.

Kevin: It is never any fun when you have data, but you have nothing that you can run it on.

Chris: Right. Then I guess some people have talked about having their recovery stuff from old tapes, and they even have old tape drives to restore things from. Hopefully, in today's world we don't need to go back that far for data, but you never know with the environment.

Kevin: The environment and the data backups, or how far to go back on some things with data, I always remind clients that they really should save things for seven years, and then find the right way to electronically shred data that's older than seven years that they don't consider to be a permanent type of document that needs to be saved. Like a lease or a loan covenant and or some other type of thing that is permanent in nature that you want to keep after seven years. Transactional documents, like accounts payable checks, electronically paperless office type of things, you've got to find a way to make sure you electronically shred those by deleting those.

Chris: Excellent point. That also goes back to our managing the data size.

Kevin: We're well past the days of the simple 10 Meg IBM XT, that's for sure.

Chris: Right. Hopefully, we don't have to worry about restoring an old Word Perfect file. I don't think I have the application to do that anymore either.

Kevin: I do have the Word Perfect manual on my desk as one of my mementos of a historical nature.

Chris: I did find a Windows 95 startup disk, if you need one of those.

Kevin: Yes, okay. I'll keep that in mind.

Chris: Kevin, I wanted to talk a little bit about the idea of the RAID, the redundant arrays.

Kevin: The thing to keep in mind with RAID is, you really are doing one of two things. You're either not doing RAID, where you have one disk and if that disk fails, you have to go back to some type of backup device. Or, you're running some type of RAID numbered one through six that is using multiple drives that will give you the flexibility that if a drive fails, because it's in a RAID of other drives, your data still goes, works, is accessible and then you have X number of days to swap out that bad drive for a new good drive and then your RAID's at full features again. That's something that you need to make sure that your server has the right level of RAID in the drives to give you the redundancy you want.

But then on your workstations, most people never bother putting RAID on their workstation, because if their workstation fails, they're going to get a new workstation. It's only affecting them, not an entire office of five users or a hundred users or 500 users.

Chris: Good point on that Kevin. The things I'm seeing RAIDs on, obviously agreed, is on the server and then also on the NAS, or the network attached storage. I'm also

a big fan of-- The hard disks have always been one of those things that fail more so than other items because of the moving part.

If you've got a mirrored drive, or in the case of a RAID five, you have a series of drives that you're striped with a parity drive, you can have a failure and it does not let you lose your data. You simply pop out the bad drive, plug in the new one, it goes back, updates the RAID, and you're able to continue without any downtime.

Kevin: That's the power of RAID. I can't tell you how many times I've gotten a note from a customer saying we had a drive fail but we're okay because RAID's in effect. Can you send somebody out here to get ready to swap out the drive? Or they tell us we're swapping out a drive because RAID kicked in and showed us that we had a failed drive and--

Chris: Hopefully, they know to not ignore the beeping noise. I've had one client that's like, "I have this annoying noise come from my server." Finally, when I looked at it, I said, "You've got a bad drive in your RAID. If you lose another one, you're going to be out of luck."

One other point on RAID real quick, is when you buy disks, a lot of times they're made from the same manufacturing run. A lot of times their mean time between failure is actually about the same time. It always makes me get worried when one fails, you might have a whole series of failures coming up after that.

Kevin: Correct. Today's environment as well as we see more and more solid state drives, we have more comfort knowing that, because it's not a mechanical moving part, there is less to break. Hopefully, that drive will last much longer than a normal old-fashioned spindle drive.

Chris: Agreed. The space capacity on these SSDs is going up dramatically which is really nice to see. All right. Why don't we cover a little bit of the recovery testing? You want to cover what you think are some of the best practices?

Kevin: Yes. When I see recovery testing, what immediately comes back to mind is where we started: data backup policies and procedures. The recovery testing is so important to make sure that somebody is going to make sure that what is being backed up, really is being backed up. When I'm working with outside IT firms, I'm always making sure that I tell them as we're implementing different items on servers or applications, "You need to test the recovery by doing a restore." I don't want to hear someone say, "Yes, it works." I want someone to say, "I tested it and it worked."

Chris: I completely agree with that. I've had way too many times when somebody says, "Yes, I'm confident it works." Until you actually do that recovery, you don't have the validation that you really need.

Kevin: Correct, and there are a lot of IT people in the workforce that I'll say can get a little flippant sometime where they just say, "Yes, it'll work." They think just because they said it will work, that it actually does for certain, work. That's what recovery testing is all about.

Chris: It's one of the things you want to find out in a safe test environment, not when you actually have a critical loss, that you're able to retrieve what you need. There's a funny cartoon I've kept on my bulletin board for a long time of a guy who's wheeling up this big giant crate that says, "Warning, wild rhino," and they're going to open up the server room and say, "Let's test how well the recovery system really works."

Kevin: [laughs] I like that. That's cute.

Chris: A couple of thoughts I have on recovery testing, is again, it's a different level. Do you test that individual file level? Can you restore folders? Can you restore servers? Or even sometimes network configs have their own issue of Active Directory and other things like that.

Kevin: That's a great point. Each of those four items that you just mentioned are all individual subtests within that recovery testing that you need to make certain that you're comfortable works. Especially, as you start getting to, well, if I lose an entire server, can I truly restore it back to its state that it needs to be?

Chris: Again, that gets back to how much are you willing to spend to be down or not be down? All right, how about a little bit of the subject area relating to controls over backup data? Some of the things that I look at when I'm doing some of the reviews is how much protection do you put on the backups themselves?

Do you keep them encrypted? What's the chain of custody for them? How do you make sure that if you have sensitive data, that it's properly kept in the right controls, but the same point not so restrictive that you can't do what you need to do from the backup standpoint?

Kevin: Correct. If you layer into that whole scenario, the fact that are you HIPAA compliant or whatever other regulatory agency is putting down certain edicts of how your data needs to be backed up and saved, and where it's saved and what format it is, that's also important.

Almost every one of us now in a business environment come through and touch some type of data that's either fiscally, or from a credit card standpoint, needs to be encrypted. Or at some type of health care or HIPAA type documents or information that needs to be saved and encrypted correctly, we just can't say, "I have a backup." We also have to make sure that the controls are all there based on whatever agency is overseeing us.

Chris: Great point. Especially, how is your Cloud transfer? Is that encrypted properly? Is the facility that you're using to receive that backup, is that meeting the requirements that you need? Are you looking at their SOC reports to make sure that they're doing things that they say they're going to do if you're using a third party?

Kevin: Correct. Getting into those SOC reports is so much more important today than it was years ago. Those SOC reports are giving you, as a reader, a third party's opinion on if your vendor is doing all the things that need to be done day in day out, year in year out. Not just what's on the marketing on their website, but in real life.

Chris: Right. Especially, since we're using so many third parties for backups and services as we mentioned, the Carbonite and the OneDrive, those are all third-party. Looking at what they're doing is important because you're placing reliance for your own data on them.

Kevin Correct. I'm amazed how many people years ago would lock all their paper filing cabinets, et cetera, when they left the office at night. You look at the amount of data that they're leaving on their PCs with their PCs turned on or what they have backed up on a USB Drive sitting next to their desk that is last night's backup from the Cloud, and all that data is so accessible versus years ago, the data was not. You really need to make sure you're putting in the controls that your management team's expecting.

Chris: That's a good point. It's always the weakest link in the chain, that can be the problem. Now, that things are propagated so much, you have to really expand your scope when you're looking at things. Well, Kevin, any other good backup stories you can think of? I always find stories are interesting to share.

Kevin: I'm old enough and have had enough fun that I've had tapes that have broken or come off the spindles that we had to hand wind back on to another cassette so we could restore items. The hard drives that have crashed over the years that you had to send out and pay thousands of dollars for some type of data recovery, or sector recovery of different items. All those stories are fun to have.

I think the stories of the future are going to be more of what's going to happen to all this Cloud storage that's out there of all of our data if people don't do a good job of purging data out of the systems that they're backing up on to. We're all generating just so much electronic noise or exhaust that's just data. That's really not important after so many days or so many months, but it still exists and no one's getting rid of it.

Chris: That's probably a data pollution. It could be our next big issue, but I agree with that. I looked back at how many versions of something are being stored and which one is the correct version of what you need. As this stuff grows, it gets harder and harder to maintain. It's going to become a real issue.

I appreciate everybody's time and on behalf of the AICPA Information Management & Technology Assurance Division, we would like to thank you for tuning in to this CITP exam series podcast on Data Backup and Recovery.

This is one in a series of podcasts that the AICPA's IMTA division is pleased to offer around a variety of topics of importance for the CITP exam. Be sure to check out other podcasts in this series, on topics that include Data Analysis & Reporting Infrastructures, COSO Framework, Information Lifecycle Management, Service Organization Controls (SOC), Internal Audit, PCI Compliance, and HIPAA Compliance. Thanks for listening.

AICPA CITP Credential Examination Series

Disclaimer

This podcast is designed to provide illustrative information with respect to the subject matter covered, and does not represent an official opinion or position of the AICPA or AICPA.Org. It is provided with the understanding that the AICPA and AICPA.Org are not engaged in offering legal, accounting or other professional service. If such advice or expert assistance is required, the services of a competent, professional person should be sought. The AICPA and AICPA.Org make no representations, warranties or guarantees as to, and assume no responsibility for, the content or application of the material contained herein, and especially disclaim all liability for any damages arising out of the use of, reference to, or reliance on such material.