RunAs Radio

http://www.runasradio.com

RunAs Radio is a weekly Internet Audio Talk Show for IT Professionals working with Microsoft products. The full range of IT topics is covered from a Microsoft-centric viewpoint.

Text Transcript of Show #0107
(Transcription services provided by PWOP Productions)

Robert Smith Diagnoses Our Storage Performance Problems!
April 29, 2009

[Music]

Brandon Wenn: From runasradio.com, you're listening to RunAs Radio, the Internet audio talk show for IT professionals with Richard Campbell and Greg Hughes. This is Brandon Wenn, announcing show #107, with guest Robert Smith, recorded Thursday, April 16, 2009. RunAs Radio is produced each week by PWOP Productions, providing professional media and podcasting services online at pwop.com. You can follow the boys on Twitter at twitter.com/runasradio.

Richard Campbell: This is Richard Campbell and you're listening to RunAs Radio. With me as always, my co-host Greg Hughes.

Greg Hughes: That's me. Hey, Richard.

Richard Campbell: How are you, sir?

Greg Hughes: Good. How about yourself?

Richard Campbell: Oh, you know, going like crazy and getting ready for TechEd. Things are exciting as usual and we still have the sweepstakes going on over at the dotnetrocks.com.

Greg Hughes: Right.

Richard Campbell: The .NET Rocks folks, and I'm one of them, we're giving away a free pass to TechEd: airfare, hotel, and a pass for one lucky winner. All you got to do is enter the sweepstakes.

Greg Hughes: Pretty cool deal.

Richard Campbell: Yup. Swing over there and have a little look at what we're doing. But let's start right into this because you know I always know we're having a great show when the conversation base will start even before recording. We've got Robert Smith with us, and Robert, you're one of the premiere field engineers and I don't think enough folks know about PFEs.

Robert Smith: Okay, then do you want me to kind of talk about that a little bit?

Richard Campbell: Absolutely. Just dive right in. We've talked a clean half minute with a couple of other premiere engineers so just tell us about your role and what is it you do.

Robert Smith: Okay, I want to start by saying thank you for having me on the program. It's very informative and I've listen to several episodes.

Richard Campbell: Thanks.

Robert Smith: Anyway, PFE, Premiere Field Engineers, we are a group and we do support consulting for Microsoft. We fall under the customer service and support umbrella which encompasses -- it used to be called PFF and now it's called CFS Premiere but we do the same type of thing. We work closely with customers often geographically and we try to build a relationship and often go to the customer site everyday and spend time and build this relationship and so we know the customer, we know what they're trying to do, what they would like to do in the future, and we do everything possible to facilitate success in whatever their venture is.

Greg Hughes: So which customers do you work with, or what's the relationship of the type of customer, and how they usually engage you?

Robert Smith: I work with Lockheed Martin mostly. I've been also working with a relatively new customer called United Launch Alliance and they actually are in rocket business and they're a new company that was formed from the rocket business from Boeing and Lockheed Martin and they're now a standalone entity.

Greg Hughes: All right, yeah.

Robert Smith: They have a big facility in Denver and I go and consult with them, but also I go and work with Lockheed. Sometimes I have to travel because Lockheed is spread all over the place.

Greg Hughes: Sure, scattered away.

Robert Smith: East coast, west coast, whatever. But generally, you know, we have a pretty good presence out here in Denver.

Richard Campbell: So it's generally big companies like Lockheed Martin that has PFE relationships?

Robert Smith: Generally yes. The contracts are a certain size before they're signed. We don't just take the smaller contracts. It's not that it's some great expenses, just that we have a threshold because we allocate head counts based on the hours that we sell so we try to sign a certain size contract and say okay, this is your person. I'm dedicated to Lockheed Martin and there are others that are dedicated to Lockheed Martin. So that's my job, it's just to go in there and build the relationship and sometimes it's reactive, sometimes I'll walk in there and they're having a cribbage set or something but...

Greg Hughes: Sure.

Robert Smith: That's kind of why I got pointed in the direction of RunAs Radio, it's because I spent a

lot of time troubleshooting storage performance and we thought it would be a good topic to bring up.

Richard Campbell: Well, absolutely and I think a lot of folks are surprised and frustrated with their storage systems. The vendors tell one story and people's experiences tend to be somewhat different and there are lots of reasons why, it just turns out more complicated than people really think.

Robert Smith: Yeah, it turns out that, I don't know, storage and managing storage technology is pretty intensive and it takes a lot of, you know, you have to keep up with evolving technologies and a lot of it is not really quite automated as it should be. A lot of the things, maybe they don't work the way they do at the trade show simply because the environment is not the same. It's not that the product doesn't do what it says it's going to do, it's that you get into the environment and things are different and you have different requirements, and so to me I think we're still evolving and we've made a lot of progress, there's no doubt about that and I've seen some great things. Solid State, SSD, that's going to be great because then hopefully at some point we can start doing away with this term called spindles, this to me is kind of almost antique.

Greg Hughes: Sure.

Richard Campbell: Yeah, it is a little okay. It's like vinyl records.

Robert Smith: Vinyl records are great but they have their place and you just have all these complex problems that arise because of, you know, allocate spindles and it's just very complicated.

Richard Campbell: Now I've been paying close attention to the SSD's situation and this maybe a bit of a digression from our topic but it's obviously something that's near and dear to your heart too. We're not seeing a lot of traction on SSDs on the enterprise level. Seems like it's a desktop product so far.

Robert Smith: SSD is expensive. The last I heard it was, what is it, like a hundred dollars a gigabyte or so.

Greg Hughes: Right.

Robert Smith: So for the mainstream enterprise customers, it's just not cost-effective and at some point I think they were talking about hybrid storage where it's partially SSD and partially spindles, you know, spinning platters...

Richard Campbell: Right.

Robert Smith: The spinning platters have evolved so much and they are so high-performing now that they're just still cost effective to run those things and that's where we are at right now. Our storage performance problems, a lot of them arise because of the, and I want to mention this, it seems like everything we do in storage is a trade-off of some kind or another.

Richard Campbell: Sure.

Robert Smith: We have a disk drive that the vendors want you to get a 300-gig or 600-gig but if your cost model is based on gigabytes, then you're not going to get the performance because you have one spindle that does 300 gigs and maybe your budget says, well, you guys just need a terabyte so you just needs three drives. Really we need 6, or 8, or 10 to meet the performance requirements. So there's always a trade off.

Greg Hughes: You know, if you think back long enough, I can remember one that was a hundred dollars a megabyte, it was a big deal, and the idea of a gigabyte was so far off in the future and it seemed so unrealistic to so many of us but that's where we are now. Gigabytes are super, super cheap. The world has changed. The question that I have too about the SSD stuff is just like Meantime to Failure, how many times are we going to be able to read and write to the stuff before it doesn't work anymore?

Robert Smith: That's one of the big problems that everybody is working on because it's true that the cycle, the duty cycle is not the same. There's even APIs in Windows that are being developed and of course I'm not revealing any big secrets or anything, but there are APIs that are being developed to change the way that we write to the disks so that we don't sit there and write anymore than we have to just to help improve the life, the duty cycle of SSD.

Richard Campbell: Yeah and part of the challenge of the SSDs, we've seen this in IBM X.25 research, there's been a big baffle around this that they know these cells only last so long so they're constantly rewriting things to make sure that the writes are even across all the cells and that's becoming part of the problem of longevity as well. There's just so much chatter on these drives to try and keep them balanced.

Robert Smith: This is true. One of the things, because I go to the trade shows once in a while and one of the things that I saw that was pretty neat is you're running, I don't know how many terabytes, 16, 32 terabytes and you're running it using like 80 watts of power.

Richard Campbell: Right.

Greg Hughes: Yeah.

Robert Smith: You can zero every cell in the entire array in five seconds. So there are some amazing potential out there, and then IOP, I don't want to get to start using lingo without saying what an IOP is, it's simply an input/output operation per second, a disk transfer operation. So these SSD devices, some of them are doing, you know what, a hundred thousand IOPs. It could be any kind, read, write, random, sequential, it doesn't matter and so the future of those is really good, but today, that's kind of what I'm talking today about storage performance, it's we're still bound to these problems that we have based on spindle, spinning platters.

Richard Campbell: The number of platters and the number of heads, I mean all of these things matter. I've definitely just gone through this with a customer where the reality came down to too many apps competing for the same spindle at the same time and the disk queue just blew out the window. That's what's killing them and they'd never looked at their application that way before.

Robert Smith: Right and you see that's a whole problem that you get into especially on Storage Area Network or SAN...

Greg Hughes: Right.

Robert Smith: It's that because of disk cooling or disk virtualization or conglomeration, whatever the term is, where you group together disks, because of the resulting amount of gigabytes, you might have to share those.

Richard Campbell: Yes.

Robert Smith: You might have to curve up those, they call them LAN but it's really just a logical disk device and you pass those out to the different people, and the servers, you know one server might be busy doing something really disk intensive and it makes another one softer because they're sharing the same spindle in the backend and that's really, really hard to troubleshoot.

Richard Campbell: Well, I think part of problem is that SANs make that process more opaque.

Robert Smith: Right. That's the nirvana of SAN, it's that, well, now you can more efficiently utilize your storage and I'll tell you that captive storage, which is storage that's tied to servers, the utilization is low 10% to 15% so you have a lot of wasted storage. So in SAN we can make up for that, we can cool and present and reallocate and do lots of things but there's still, I don't know, there are still problems out there that have to be addressed. Some people try to addressed those by using the old method of we're okay, I'll just manually create my disk group and I'm going to give it to a server so I'll know I'll have better knowledge on what's on the backend.

Greg Hughes: Right.

Richard Campbell: So how do folks go about actually instrumenting their storage infrastructure knowing that just finding out that this is the problem, I know my apps is running too slow, I've got the complaints, how do I get started in saying this is where the problem lies?

Robert Smith: From a Windows perspective, we really don't have too much visibility into this. Let's say it's a SAN, okay. We could have direct attached, but right now we'll say it's a SAN.

Richard Campbell: Right.

Robert Smith: So we don't know what's going on. SAN is just a transport so I'd say we start with Performance Monitor, PerfMon, and we'll look at things like, you know, we'll start with disk transfer per second. How many IOP am I doing, what is my response time, and how many IOs am I building up in my queue. A queue is simply an IO that's waiting to be serviced. There's a trade-off, like I said everything is a trade-off. We have to try to keep the queue if at all possible without making IO wait excessively. So we go back to the spindle and you say, well, a fiber channel spindle is capable of doing about, on average, about three IOPs or IO operation per second. Other drives maybe, like say the ASTA drive, it might be two. So therefore, if you have a disk group that's comprised of five disks, you might say, well, the queue link should be around 15. Now nothing is static in the computer world, things are going to spike and they're going to go up and down so these ephemeral or the short-lived spikes, we don't worry too much about those as long as the storage can recover. We look at sort of averages overtime and make sure that we're not making IO wait excessively. PerfMon is a great tool but there are other tools. There is a tool called Server Performance Advisor, it's a free download. It's really meant for Server 2003. I think you can use it on Windows 2000, but possibly we're getting away from Windows 2000 by now. It used a combination of Event Tracing for Windows and PerfMon counters to provide you just almost a report-based, here's what your load is and here's your response.

Richard Campbell: It all gets back to PerfMon, right. Everything we ever wanted to know about our system is in PerfMon, we just can't find it.

Robert Smith: Yeah, it takes a while to get accustomed to know what you're looking for in PerfMon, and the thing about storage, if you're troubleshooting storage you want to make sure that your sample interval is sufficient that you're not losing a lot of what's happening. In other words, if you say you're interval at five minutes or even 30 seconds...

Richard Campbell: Right.

Robert Smith: That's too long. When I do storage troubleshooting, I'll say, okay, let's capture physical disk and logical disk at a one second interoperable and nothing else.

Greg Hughes: Yeah.

Richard Campbell: Doesn't testing now become part of the dataset?

Robert Smith: You mean the overhead of sampling?

Richard Campbell: Yeah.

Robert Smith: It's mostly in memory. I mean, you know, you're ticking the drive every so often but it's not really disk intensive, there's very little overhead, it's not enough if you queued the results too much.

Richard Campbell: Okay. I mean, it's interesting to know. Where would you be writing this out to? I mean you just keep it in memory or are you actually trying to log into disk or log it into SQL Server or something like that?

Robert Smith: Well, you can log it to disk and you can capture from another machine over the network interface.

Richard Campbell: Yeah. I was thinking that the thing to do, especially when you're analyzing storage performance, is to ship it off to another machine so that your writes to the disk aren't impacting your record keeping.

Robert Smith: Yeah, that's a very good point. You're absolutely right and that's why we say let's do it over the network so it's pretty much mostly handled in memory and then through the network interface, not too much goes to disks in that scenario. So you capture from another machine, you do your analysis, you can automate the whole thing, what log mean, what's the tool that's built into Windows. I do want to mention another tool though. PerfMon is and has been probably the greatest tool but there's a new one. It's been around for a little while. It's based on Event Tracing for Windows.

Richard Campbell: Oh yeah.

Robert Smith: It's called the Windows Performance Toolkit. It's another free download and actually I believe the latest version is in the Software Development Kit, the SDK, which that can be a big download but there's a lot of good information and tools in there. So you can download that, now you got the Windows Performance Toolkit. Now you have to install it on Vista or Server 2008, however, you can copy. There are two files that you can copy to a Server 2003 machine, it's Xperf.exe and perfctrl.dll, those two files and this is all documented on the blog, it's the Windows Performance Toolkit blog, you don't have to remember all these right now, but anyway, you copy those two binaries to Server 2003 and now you can collect traces from a Server 2003 machine and then you have to copy the trace over to the Vista or Server 2008. It's just amazing, the information that you can get from this. You can see a visual of where on the volume the IO is taking place, the overview of things you didn't know were happening. What if you have concurrent IO operations like two intense data streams at the same time. Well, a lot of your delay might be simply latency from the seeking that's going back and forth and you see that visually in the Xperf tool and you also see every single IO so here you wouldn't capture for a long time, you might capture for five minutes because files can get really big.

Richard Campbell: Right.

Robert Smith: The great thing about Xperf though is it's all in memory in kernel mode and you can start and stop a trace without any kind of reboot, there's no disruption to the server.

Richard Campbell: What's interesting just looking over performance analyzer and so forth in the WPT here is this is aimed to performance tuning so it feels like its PerfMon data, they just stripped away the stuff that isn't related to performance.

Robert Smith: Yeah and Event Tracing for Windows, I mean that's a whole topic, you can go on for a long time and because you can trace CPU activity, power states, it's just amazing. But as far as storage performance, to me it reveals things like, for example, a cluster. I can see all the cluster IO that you wouldn't see that in PerfMon because it's more of a SCSI command than it is a read or a write, and I can see Flash IO which would come from the Volume Shadow Copy Service that I may not see in PerfMon.

Richard Campbell: Right and these all come down to just trying to get a picture of where the performance problem lies that you are hammering a driver that these particular spindles are too busy. I guess my question then is what does this failure mode looks like? How do you tell that that's the problem?

Robert Smith: When at storage?

Richard Campbell: Yeah.

Robert Smith: Generally, you start with response time. Okay, you start saying, well, my response times are bad, what could it be? So let's start with the storage and we will, you know, I like to go write first into the Windows Performance Toolkit. I wish I could show you, but one of the things that you'll see is you get almost like an Excel spreadsheet-type of view of every single IO. Two of the columns, one of them is the time that the IO spent from beginning to end. In other words, from the time it was built until the time it was marked complete. There is also another queuing counter that measures the time that that IO left disk, that disk which is the very last component in Windows until the time it was completed. The difference between those two counters will show you how long that IO spent queued up waiting to be serviced. That will give you an idea of whether you're latency looks like it's coming from the storage or if it's in Windows itself.

Richard Campbell: Right.

Robert Smith: I've seen a lot of cases where SQL Server, which is really tough on storage sometimes, can really batched-up the IO. I mean, it's just like getting a bucket of water and just pour it all at once.

Richard Campbell: So literally the LAN doesn't have the ability to service the number of requests it's getting all at once and disk queue just pile up.

Robert Smith: Exactly, that's exactly right even though in most cases or in many cases we're talking to cache.

Richard Campbell: Right.

Robert Smith: Now we have this big, huge cache in front of spindle that helps optimize, and they do to some degree, but depending on the workload you get pretty random, you can actually use up that cache pretty quick.

Richard Campbell: And you get these sort of states of cache meltdown where now we're actually out of memory again with SQL Server shocker and there's not enough stuff to stay in the cache to give good overall performance on certain tasks so you got such desperate tasks competing with that cache that each thing has half the stuff needing to be cache and for me it's always been I know I have a storage problem when I go and look and I got a lot of disk queuing. Disk queuing is the sort of my red flag. Now I know I should pay more attention to disk. Is there any reason that a long disk queue is good?

Robert Smith: If you're disk queue is too high overtime, then that generally means that the disk can't keep up with the load that you've presented them.

Richard Campbell: Right.

Robert Smith: In other words, I've kind of alluded earlier the rule of two or three IO per spindle as it were. Here's another whole another problem. You may not know how many spindles are back there.

Richard Campbell: Right.

Robert Smith: What if you have a storage device that likes to cool storage and then just allocate LAN. So you get to the point where maybe you do go to the Windows Performance Toolkit to look at the queuing and see whether it's the IO. Where is it waiting? Does the disk service time look good? In other words, by the time I let go of that IRP, the IO Request Packet, from disk.sys driver until the time it was completed. In other words, that's a calculated measurement of how well storage is doing.

Greg Hughes: Right.

Robert Smith: So that's one indication right there if it's a storage problem. Interestingly enough, there's another place you can go look if you happen to have, I don't know if I should mention names, vendor names, but Emulex and their utility, I'm sure QLogic has it somewhere, but there is a counter that you can look at, it's called FBSY and that means Fabric Busy, there's another one called PBSY, Port Busy, and if those counters are non-zero or the number in there is non-zero then you're likely stalling somewhere out on the SAN so you're actually overrunning a port somewhere on the SAN, so that's just another tool to help you troubleshoot where is this slowdown happening.

Greg Hughes: This is query in the HBA or who is it that you're talking to here when you're getting this information?

Robert Smith: PBSY or FBSY is a response from the fabric, from a fabric port somewhere.

Greg Hughes: Ah, okay.

Robert Smith: It could be in the switch if you have fiber channel switch, or if you have the fabric port on the storage device itself, somewhere there, some devices returning this response. So that's a fiber channel thing.

Greg Hughes: Okay.

Robert Smith: SCSI or iSCSI technology may return busy, you know, iSCSI response. So that's just the places you can look for more information.

Richard Campbell: Robert, isn't the correct answer to this problem every time more spindles?

Robert Smith: No. That's the first thing people like I wish it had more spindles.

Richard Campbell: Yeah.

Robert Smith: It's not always. Sometimes it can be a matter of changing your cache settings on your controller, maybe because SQL tries to cache a lot of reading in memory on the server itself, maybe we can allocate some cache more towards writing. There are some other things we could do when we look at things like partition alignment. How is my partition aligned on my array disk? What that means is the clusters are lined up with the elements of the RAID storage. So we call them chunks or we call them elements or stripe units and if we cause a boundary of a stripe unit with the partition allocation then we incur two times performances. So we have delays by having misaligned partition. Very simple to fix, the problem is you have to do it when you create the partition, you can't do it after the fact. So either use Server 2008 which aligns every time or use disk parts and you do it manually, align=1024.

Richard Campbell: What does an aligned partition looks like? Are we talking about like the 64k block size that SQL Server likes so much?

Robert Smith: Yeah, it's the starting offset, the partition starting offset is lined up either -- it used to be called cylinder align I believe, so it started off it like Sector 63. So if you happen to look in whatever tool you're looking at like MSN FL32, you see that your partition starts on Sector 63. That doesn't divide evenly by any RAID element there is.

Greg Hughes: Okay.

Robert Smith: So you know right off the bat that's misaligned. A lot of vendors will tell you, you know 128, that's okay because that lines up and we say, you know, what does that number mean, 128? So that's the starting offset for that partition, offset from let's say zero, Sector Zero on a disk, so we go out 128 blocks and we'll start the partition and that happens to line up with the 64 kilobytes that are pretty common in enterprise storage.

Greg Hughes: Sure. So by doing this and by optimizing it at the storage level that way, we're really offloading the work that would otherwise have to be done by the operating system and the subsystems to compensate for the fact that it's not set-up that way.

Robert Smith: What happens if you have a misalign partition is you might seek to a certain block on the partition and to fulfill that IO might incur two accesses to the disk where otherwise it would only have taken one because we cross a RAID element boundary, and so we've seen, we've measured, there's been people at Microsoft and other places that have measured up to 30% performance right there so this is, you know, when you're building your partition, align it every time. That's just a very simple thing you can do. You can script it if you want and you'll save yourself a little headache by not having to worry, you can worry about something else later.

Greg Hughes: Sure but you do have to do it ahead of time.

Robert Smith: Uh-hum, you have to do it ahead of time.

Richard Campbell: And like you said, SQL 2008 does this by default now.

Robert Smith: Windows 2008 does it by default, yes.

Richard Campbell: Right, sorry, it is a Windows 2008 thing, not SQL.

Robert Smith: Windows Server 2008, yes. For most disks, it will start the partition at 2,048 sector offset. There's a KB article after that explains this pretty well and it's pretty easy to find in your favorite search engine and if not I can follow up because I can't remember the number, it's just off the top.

Richard Campbell: I understand and it's just one of those very odd things. I wonder how many people who are listening are like I knew nothing about this.

Greg Hughes: Right.

Richard Campbell: You're telling me literally, cutting my performance in half if I get this wrong and I've never looked at it.

Robert Smith: This is true and part of that is because of backwards compatibility with the old, it's called cylinder head and sector...

Richard Campbell: Right.

Greg Hughes: Right.

Robert Smith: Geometry of older -- because disks don't use that anymore, they just don't. Everything is abstracted. It's all about sectors now.

The hardware on the disk handles all that so there's no such thing really anymore as cylinder head and sectors. Windows is, for the longest time, trying to be backwards compatible when you create a partition in the GUI, in the UI, it would misalign right off the bat so this one of those free things you can do and pick up some performance right there if you're not doing it already.

Greg Hughes: So that's one really great proactive thing. I mean, that sounds like that's something that could impact a lot of people potentially. Are there one or two others that you run up against all the time that if people were to be proactive, plan ahead, or do something that they would solve a lot of their problems that you typically have to deal with?

Robert Smith: Keep up with your drivers. In Windows, if you're using, let's say you're using direct attached, you know I'll start with the simple things, still keep up with your drivers and firmware, Scsiport.sys and Storport.Sys, those are Microsoft's technologies for the Storage Port Driver, but anyway, Microsoft, the thing about the way that we do business, we don't generally proactively notify people when there's a hot fix. We'll say, well, wait for the next service pack. Sometimes you can pick up a little performance by updating Storport or SCSI Port and then at the same time look at your mass storage drivers, go ahead and update those, check the firmware. Now when you start talking about SAN, and when I say SAN I want to include iSCSI and fiber channel in the same context even though a lot of people think of them differently. With fiber channel, you can look at your two-depth settings on your HBAs. Now this is another one of these areas that takes a lot of -- you get yourself in trouble because you think, well, fine, increase queue depth then I can, you know, it's just like opening the spigot on the faucet and that's true but if you overrun your SAN then we're back where we started with the FBSY and the PBSY. In other words, the SAN will return, instead of serving your IO or passing your IO it will return an error and say you have to wait. It's just like slamming the door shut on your IO.

Greg Hughes: Sure.

Robert Smith: So you can tune a little bit, just make sure that you're not overrunning or causing a slowdown because you're just setting too many IO. On the other hand, you want to keep this figure open far enough that you're keeping enough IO in flight, that there's always one or two or three waiting to be serviced. So those are some of the things that you can do without, you know, that's the short of it, having to add new disks to your array or whatever.

Greg Hughes: In theory, I mean you could be in a situation where you go and, like we said, add this to your array just assuming that that's where your problem is but then you could actually have a problem where you're not properly tuned in terms of how much water you're shoving down the throat of your storage system.

Robert Smith: That's exactly right. You know, you think, well, I have some more spindles and that should solve my problem. Well, maybe that's not the problem in the first place. Sure, maybe it's not a bad thing to have more spindles but we have to make sure that we're keeping the pipe full because if we're not -- or make sure we're not overrunning the pipe and make IO wait at the gate as it were.

Richard Campbell: Like you said, it's okay to have a couple of items in the disk queue just to keep those spindles busy and you get the sense of what that right number is, but adding more spindles is not necessarily going to drive that down if the line is strangled.

Robert Smith: This is true, yeah, and there's actually more that you can do but it actually gets pretty tough sometimes unless you have access to the storage device yourself because the Performance Monitor tool, PerfMon, our old friend, there's usually a counterpart to live in the storage device.

Richard Campbell: Right.

Robert Smith: Even in a fiber channel switch or even iSCSI switches, you might have performance counters there too so you can always look there if you have access but a lot of times in a lot of cases that I run into you have the server group, you have the SAN group, you may even have the storage group and a lot of times those groups don't go along so well, they don't trust each other.

Greg Hughes: Yeah.

Richard Campbell: Yeah.

Greg Hughes: Unfortunately.

Richard Campbell: It gets easy to get into the finger-pointing game that's why we need good measurement data so that we don't point at each other, we point at the data.

Robert Smith: That's exactly right. You're right.

Greg Hughes: This is not a good reason for plan ahead and plan together, don't throw something over the wall until the next set of people when you're chain, bring them in for the design phase and you can solve a lot of those problems early on probably.

Robert Smith: Yeah. There are a couple of other tools. There's a Fiber Channel Information tool, it is another free download from microsoft.com and you can actually look at it, there's some steps that are kept in a certain place in the HBA for fiber channel and once in a while I've actually solved the performance problems because I saw that there was some CRC errors building up and we were able to point to a specific, you know, it turned out I forgot, where was it, a cable or something, but there was a piece of hardware that was causing the problem. So that's just another little tool that you can throw out there. Some of those stats are in the Fiber Channel Management Tool, HB anywhere, QLogic has an answer for, and there are other tools out there.

Richard Campbell: All right, Robert, we're getting down towards the end of the show here. I don't think any conversation about storage can be complete, especially while we're still living on spindles, maybe some mentioned a defragmentation. Where do you stand on defrag?

Robert Smith: Defrag is certainly an issue and if your data gets spread out sufficiently, then there's unnecessary seek time. There are certain storage devices that will intentionally scatter like one uses this technology called WAFL, Write Anywhere File Layout.

Richard Campbell: Right.

Robert Smith: But there is still some benefit to defragmenting. The way I look at it though is you got to make sure that you're not inducing more overhead to try to solve that problem. So if there's a tool that you're using, I mean Microsoft, you know Windows has built-in some defrag tools. If you can run those during maintenance hours or off hours or something like that, that's great, I'd say go ahead and do it. Some of the tools that I've seen, they run services and things. Just make sure that the tools don't induce more overhead than the benefit that we're getting.

Greg Hughes: Right.

Robert Smith: Because the high-end storage tries to write the blocks that you throw at it. It tries to write them contiguous in the first place so there are some optimizations that happen on the backend, I'd say just use it, use it judiciously like anything else.

Richard Campbell: Absolutely. All right, I think we're about at the end of the show. Robert, any final shout outs, things people should be looking at?

Robert Smith: Just keep up, look at SSD, I think it's going to be the future. Holler at your storage vendors, holler at Microsoft if there's something you want fixed. Please speak up and let people know because if you don't, then we won't know what you want fixed and the more voices the better.

Richard Campbell: I'll make sure to throw a link up on the website for the Windows Performance Toolkit. It looks like a great tool for digging into these sorts of problems.

Robert Smith: I love it. It was silent for a while, it wasn't release to the public and then finally it was released and then a lot of people love it too.

Richard Campbell: Excellent. Robert Smith, thanks so much for coming on the show.

Greg Hughes: Thanks very much.

Robert Smith: Thank you.

Richard Campbell: And we'll talk to you next week on RunAs Radio.