



<http://www.runasradio.com>



Richard
Campbell

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.



Greg
Hughes

Text Transcript of Show #001

(Transcription services provided by [PWOP Productions](#))



**Pat Hynds on Storage Technology
April 11, 2007**

Carl Franklin: From runasradio.com, you're listening to RunAs Radio. The weekly Internet audio talk show for IT Professionals, with Richard Campbell and Greg Hughes. This is Carl Franklin, announcing show number 1 with guest Patrick Hynds, recorded April 6, 2007. RunAs Radio is produced each week by PWOP Productions, offering professional media and podcasting services online at pwop.com.

Hello, this is Carl Franklin from .NET Rocks! and I am here as a special guest for this introductory edition of RunAs Radio. Richard Campbell, how are you sir?

Richard Campbell: I am well; and thanks for doing this for us Carl.

Carl Franklin: It's my pleasure. .NET Rocks! has been a wildly successful Microsoft .NET developer show. 220 some odd shows in the can so far; and we decided to do a Microsoft IT focused show, in which Richard Campbell is going to play the host.

Richard Campbell: Yeah, because I am kind of a mutant creature having a foot in both camps. I not only wrote the software for the machine, but I built the machine from scratch. So, it was not a logical fit for me; I'm usually more focused on issues around how gear works and so on. In my career, I've spent a lot of time building out data centers, and dealing with making infrastructure work, as well as how the software worked.

Carl Franklin: Needless to say, you're qualified for the job. And Greg Hughes, how are you sir?

Greg Hughes: I am doing just fine, and how are you?

Carl Franklin: I am fine, I am well. You were a natural choice for the co-host, not only because of your association with Scott Hanselman in -- which we're going to hear about in a second, but you are also a popular IT blogger who is Microsoft-centric.

Greg Hughes: Well, I do have a blog, it's over at gregghughes.net, and I've been writing on that for a few years. My background really, I also share Richard's interest and experience in doing a lot of infrastructure. Over the last few years, I've really migrated quite a bit toward the security side of things and security operations, both on the IT and the process side.

Carl Franklin: That's great. I can't wait to hear what you guys are going to do with this show, and I am sure a lot of people are too - a lot of listeners. There's quite a lot of untapped potential

here, because there aren't a lot of Microsoft centric IT shows out there; are there?

Richard Campbell: I am surprised at the diversity of topics. We have a good time on .NET Rocks!, with all the things we talk about around software development, but when you start thinking about what IT professionals need to be concerned with, it's vast. I am really glad Greg is with us, because he is a security guy, and that's always high on the list, but data storage, mail services, networking in general, the issues around regulatory compliance - I think these are all things that IT guys worry about, and we're going to go out and round up the best and brightest, and try and get some answers.

Greg Hughes: Yeah, I think that's absolutely right, and we have some real exciting topics on the plate, and I think it's going to be a lot of fun. This will be somewhat unique, I think, having the Microsoft focus, but also -- I am just really excited about the fact that we're going to be putting together some really interesting conversations.

Carl Franklin: This show is typically going to be about a half-an-hour, but we ran a little bit long this time, because of this introduction, and of course the talk with Pat Hynds which is coming up, but we typically want to run this at about a half-an-hour. Is that right guys?

Richard Campbell: My real vision is that we have a single topic each show -- and that will tend to run shorter. I mean, experience with .NET Rocks! we found that, typically we talk about a few different things along the way, whichever way the conversation goes, but I really want to try and focus on an individual thing. We'll see what happens; if it works to go longer, I'll go longer. I hope the listeners will let us know what they like.

Carl Franklin: Okay, and of course you can send us your feedback and get all sorts of information about the show at www.runasradio.com. Guys, this is where I step off. You have a good show.

Richard Campbell: Thanks very much Carl.

(00:05:06)

(Music)

Richard Campbell: All right Greg, let's introduce our guest. Patrick Hynds is the President of CriticalSites and a Microsoft Regional Director. Named by Microsoft as Regional Director for Boston, he has been recognized as a leader in the technology fields. An expert on Microsoft technology and experienced with other technologies as well (Websphere, Sybase Storage Management, Java, UNIX, Netware, and

Network Security), Patrick has taught freelance software development and Network Architecture, and has been a successful consultant who has enjoyed mastering difficult troubleshooting assignments. A graduate of West Point, and a Gulf War veteran, Patrick brings an uncommon level of dedication to his leadership role at CriticalSites. Experienced in addressing business challenges with a special emphasis on security issues involving leading-edge database storage, web and hardware systems, Patrick often speaks at major technical conferences, including TechEd US, TechEd Hong Kong, and the Middle East Developers Conference in Cairo, Egypt, the Microsoft Security Summit in New York and Boston, DevDays, Code Camp, and many more.

In spite of the demands of his management role at CriticalSites, Patrick stays technical and in the trenches, acting as Project Manager and/or developer/engineer on selected projects throughout the year. Welcome Patrick.

Patrick Hynds: Thank you.

Richard Campbell: So, what is it you don't do?

Patrick Hynds: Windows.

Richard Campbell: Ah, there you go.

Patrick Hynds: Actually that's not true; I do Windows.

Richard Campbell: So, we've talked many times. I mean, Pat, you and I have been at conferences and so forth together.

Patrick Hynds: Yeah, just recently.

Richard Campbell: Today though, I really wanted to talk about storage technologies, something you and I have chatted about over dinner once or twice. So, what area -- well, maybe we should sort of back off and say, what do we mean by storage technologies?

Patrick Hynds: Well, so when you say storage technologies, it evokes to me, where am I going to put my bits? Where am I throwing the data that I care about? And one of the things that I think we should start with is, how do you think about that; how do you think about storage? Is it just a disk? Is it just a product that you buy? Or is it a service that you have to provide to whatever community - developer community, user community, knowledge worker or customer community?

Richard Campbell: Does it really get into something that ultimately I am selling to my customer?

Patrick Hynds: It depends on who you are; I mean, if you're running -- if you're a storage manager at Fortune 100 Company or Global 1000 Company, you're trying to just stay ahead of demand. Right now, we're seeing -- we are still on the cusp of people seeing a 100% storage growth per year. We were talking to a large financial organization over in London earlier this year, who said that they started off this year at 700 Terabytes, and they will be over 1.5 Petabytes by the end of the year.

Richard Campbell: Those are big numbers.

Patrick Hynds: Yes, they are very big numbers - and they're scary numbers, because how do you backup to tape a petabyte? You don't, is really the answer. So, there's a lot of things that - a lot of paradigms are breaking and shattering now under the -- just sheer weight of the growth.

Richard Campbell: I wonder how many times we've actually been successful at using tape backup in a Disaster Recovery scenario ever.

Patrick Hynds: I've seen many, many more stories, where it has failed and where it has succeeded - and most companies are realizing that. The vast majority of companies I talk to are going to disk to disk, and they're doing Replication between sites, and things of that nature - but that increases the amount of disk you have, because now you have double the disk or triple sometimes.

Greg Hughes: Not to mention the network connectivity and all the different infrastructures that has to be in place to be able to support that.

Patrick Hynds: Exactly; and bandwidth has always been the most expensive resource after manual labor. So, if I need to double the disk on a project -- let's say I have an office in Memphis, and I come to you and say, "Look, the project is going, I need to double the disk," most people will just say, "No problem." It doesn't matter what the disk is, if it's 50 Terabytes or 100 Terabytes, double the disk is not an interesting number; the maintenance of it is, but most people don't think in those terms. But, if I said I had to double the bandwidth to that office, it's probably got to go all the way up to the CTO level and the CIO level to get approval, because bandwidth is so expensive.

Richard Campbell: You couple bandwidth into storage technologies -- let me rephrase that. You couple backup into storage technologies.

Patrick Hynds: Well, you have to; I mean, I can give you a free puppy; does that mean the puppy is free?

Richard Campbell: Tell that to a Linux guy.

Patrick Hynds: So, the cost of having something -- that's where the whole definition of whether it's a product or a service comes in. If it's a product, then the cost of it is the cost of it. If I buy a pen, the cost of the pen is what it is, I don't have to maintain the pen, I don't have to keep the pen online - when the pen is done, I throw it away. I have never seen that happen with an online storage solution. Whatever data is on that storage is got to be support; put a Terabyte up, let the users use it for a year, try to take it offline, God help you.

(00:10:09)

Richard Campbell: As soon as you provide the service, it's a one-way trip.

Patrick Hynds: If it's not the same terabyte, you'll be replacing it and migrating that data for the rest of eternity.

Richard Campbell: All right. So, let's talk about technologies then. What are the options?

Patrick Hynds: Well, there's still stone tablets. The big three are SAN, Regular Server Storage -- which is really falling out of favor just because of limitations in size and flexibility -- and NAS. Those are the buckets; you either have a big server with a big disk, you have a number of servers with a SAN, which they're front ending for, or you've got NAS devices that act as storage points throughout the enterprise.

Greg Hughes: I think for the purpose of people that maybe are familiar with the terms, you want to explain the Storage Area Network and what the different terms mean?

Patrick Hynds: Sure. So we refer to the first one, it's just a server with storage as direct attached storage. This is where the storage is attached to the bus of the box - I think any RAID whatever you've ever considered. That's what everybody is used to, because that's what you buy. But SAN is a much more expensive and much more flexible mechanism where you have a Big Honking Disk - that's a technical term. Big Honking Disk, a big monster fast disk that sits out there and allows LUNs, Logical Unit Numbers, to be split off, and shared and allocated. So, let's say I have a SAN that's a petabyte - or make it easy, a 100 terabyte SAN. I may have server one that gets 15 terabytes, server two that gets 25, server three that gets 50, and I hold some more in reserve.

So, each of those chunks, or those Logical Unit Numbers or LUNs, can be assigned to a server and act as if it's a local drive to the server, but the speeds can be much faster, because they are typically Fibre Channel connected historically, although TCP/IP, iSCSI has become more popular. So, this is where -- you still connect to the server, you still connect to a share in the server, but the disk is actually somewhere else - it's abstracted. Typically SAN is more expensive; and it was the first off-the-box solution. SAN was the thing that came out, and big companies started doing it - large banks, think of the top five banks, they were probably the first ones to go down this road. Then NAS, Network Attached Storage is where you get a better model as far as cost goes. With NAS, the biggest competitors would be Windows based NASs - EMC is pretty big in that space. You can buy an EMC NAS through Dell, but the two big competitors would be EMC, with their Solera, and Network Appliance with their Filer, and also IBM, which just came out with their own Filer.

So, what those are, is those are boxes with disk and their own operating system and their own ability to share; and so it's like a network printer. The printer doesn't have to be connected to a server to do its thing and for you to connect to it - the disk is the same way, it's out on the network, it has its own address, it has its own resolution, it has its own tools. NAS is typically much cheaper byte for byte than SAN; and it's much more flexible in performance than Server Based Storage. So those are the three choices, and most of the big customers I talk to have all three.

Richard Campbell: Does iSCSI sort of blur the line? I mean I've always seen iSCSI as a SAN technology, yet it seems to act very NAS like.

Patrick Hynds: Everybody is trying to figure out how to use it; everybody is advertising it. It's really about how is your infrastructure? Do you want Fibre Channel, which is difficult, and what if I want to throw Fibre Channel between two buildings? Now I am talking about very specialized technology -- or can I just use a gigabit Ethernet or 10 gigabit Ethernet to bridge those two gaps?

Richard Campbell: So, in your mind, all iSCSI really is this direct connect CAT5E gigabit Ethernet between the devices?

Patrick Hynds: Well, so it's got some -- it's evolved beyond that. It's got some other features and some other capabilities that make it more competitive with Fibre Channel, but easy to implement, cheaper to implement and eventually higher bandwidth than Fibre Channel. Problem

with Fibre Channel is, its wicked high bandwidth, but it's not evolving nearly as fast as the iSCSI is.

Richard Campbell: It seems like Fibre in general has hit a wall here in terms of utilization.

(00:14:59)

Patrick Hynds: Yeah, because again it's glass, everybody -- it's very difficult to work with, it's brittle. It just got too many -- it's a first generation of a technology.

Richard Campbell: It's never really evolved.

Patrick Hynds: Yeah. The odds that they'd get it right on the first try are zero.

Greg Hughes: Having implemented NAS, SAN -- a few different SANs and an awful lot of local attached storage, or server direct attached, I can echo the fact that the cost -- the cost between these is pretty exorbitantly different.

Patrick Hynds: Yeah.

Greg Hughes: Maybe you could talk a little bit about how extreme those cost differences really are.

Patrick Hynds: They are very different; it's like a Ferrari versus a good GM Car I guess -- or maybe that's an oxymoron depending on who you are.

Richard Campbell: Oh, goodness!

Patrick Hynds: I know people don't like it, but to be honest with you, I can get all the functionality I need out of a Jeep; I don't have to buy a Land Rover, but a Land Rover is going to set me back a lot more, it's going to give me some flexibility, maybe it's better for specifics. I kind of always thought about the SAN versus NAS as, depends on what you're using it for, and how you want to use it. For instance, we have a SAN here; we also have several NASs. The SAN is used for the primary storage for our servers -- our exchange server, or our file servers have a large -- our backup servers. We have all the disks centralized; we have access to true management utilities. The users can map their home directories in the same way. It wasn't that big of a difference in price, because we went with the EMC SAN -- and through Dell that's a much cheaper option. So, the price difference is narrowing, but in the early days, it was the difference between buying IBM versus buying Novell. If you remember those days, you bought IBM Token Ring, because that's what everybody bought, but you could buy a Novell network, and

you could do it a lot cheaper and it even had some advantages.

Richard Campbell: When you had -- and networking worked. So, I mean your point is, NAS works, and has always worked; SAN was sort of the Deluxe solution. I used to see that SAN was the 'larger' solution; when you were talking petabytes, weren't you always talking SAN?

Patrick Hynds: To some extent you still are; although that's changing now. The thing about SAN is, it's a more mature technology; it's actually been around longer. NAS was a reaction of, "Wow, these SANs are a great idea, how can we make this cheaper?"

Richard Campbell: Makes sense.

Patrick Hynds: Network Appliance is probably the big innovator here. EMC is a force, but they've still got such a wide product line, whereas NetApp is actually more focused just on the NAS version. So, they don't have all the other software, all the other sub companies. EMC bought Veritas, so they have a bunch of solutions from them as well. So, I watched them quite a bit to see where the innovation is coming from. But again, EMC is a big company and they're doing a lot too.

Greg Hughes: One of the things that crosses my mind, the questions that comes up for me, I'd be interested to hear your thought on this. We're seeing a lot more -- we've talked about backup versus storage and really how that's very much a blurred line now. We're seeing a lot of like, home backup systems, a lot of services that are out there. Microsoft has in beta their Home Server, and there's a lot of ASP or service type of offerings out there. What's your thought about that, and maybe do you have any thoughts about the future?

Patrick Hynds: Yeah, I think it's a great phenomenon. I like the idea that people are looking to backup their stuff, because up till now the answer has almost always been, "Oh, your stuff has gone? It's too bad it's gone." Almost every time somebody came to me with a burned out computer or something, there was no backups. A lot of these backups are the one-touch backups. So, you plug the 300 gigabyte device in, you hit a button, and it just goes and grabs everything it can from the server that it's connected to, and just copies it. I haven't played with them, because I've got a more mature backup solution at home, disk to disk, one remote building to the other, but I think it's a good thing. I think it's a logical place for that market to go to,

and I like the fact that it skipped tape - it just went right to disk to disk.

Greg Hughes: Sure, I think, having worked with all these technologies, and sort of being in the online banking industry, I see SAN as being -- you mentioned, the more 'mature' model. It's been around longer, it also has some feature functionality that can be really very robust if you need to be able to, for example, synchronize LUNs and drive arrays, or if you need to be able to do splits so that you can take one off -- make a copy and take it offline - then you can do your backup or do some other form of protection.

(00:20:10)

The NAS world hasn't quite caught up there yet; what's happening technology wise in the NAS world, to bring it really on par with SAN, not just from a bandwidth perspective, but from a feature functionality perspective?

Patrick Hynds: So, the thing is, you are right in all points, and you might not be paying as much attention to it as I am, because I do actually go and speak at some of the source conferences is, the NAS people have a Napoleonic complex against SAN. They really want to beat them, and what they are doing is they're adding in functionality much faster than most people realize. For instance, let's say I have a 30-50 filer.

Richard Campbell: This is a NetApp?

Patrick Hynds: That's a NetApp filer, it's a cluster, it's two heads and humongous number of drives, and the way you normally set it up, is you take half the drives and allocate them to one head, and the other half of the drives allocate them to the other head, and what they have done is, if the drives don't fail, because the drives typically don't fail, and if they do they are all raided on the shelf, and therefore there is a swap out, but if the drives don't fail, but one of the heads fails, they are set up to assume the identity of each other. So, this is a really nice fault-tolerant feature. So, I am on the network, I have got my two NAS heads on the same switch, if one of the brains cooks off and dies, the other one detects that the heartbeat is gone, and it assumes the identity and manages the disk that it used to manage and answers any calls to it's dead brother.

Richard Campbell: Sounds an awful lot like Active/Active clustering in something like SQL Server.

Patrick Hynds: Right. So, SQL is actually something that's a big deal now on NAS, as is

Exchange, as is Oracle. So, it's becoming an application platform in a way that SAN has traditionally been the application platform.

Greg Hughes: Right, so this is similar to the redundant HBA's or Host Bus Adaptors on a SAN?

Patrick Hynds: Yes, but at a lower cost because I am not just paying for the redundancy, I actually get to use the object and use it as a backup. And we are starting to talk about much higher, as I said earlier, numbers of terabytes. We are in the petabyte range now with some of the NAS's, which just wasn't heard of before.

Richard Campbell: That used to be SAN country only.

Patrick Hynds: Exactly.

Greg Hughes: Well, one of the topics of discussion that, sort of, tracked over time has to do with density of storage, right? I think a lot of us know what a standard Data Center rack looks like. Quite often you will see SANs, for example taking up a single rack. What's density looking like now and what are they doing in the NAS space, and how much storage can you fit into some given area?

Patrick Hynds: That's an interesting question. I haven't really thought it out recently. If I had to swag it, I would say you could probably fit a petabyte in two or three racks.

Richard Campbell: Now, are we still talking about SCSI only for the drives in any of these density storage solutions? Is it still we are stuck at a 147 gigabyte SCSI drives?

Patrick Hynds: Well, we are getting beyond that, we're seeing 340s. I don't what the exact multiples are, but we are seeing bigger drives, which is why we can see a petabyte in less than a Data Center.

Greg Hughes: In all cases, it's not always SCSI anymore either, is that right? We are seeing other technologies that are maybe a little more complacent in what you might think of as being -- I guess the term would be commodity type of drive.

Patrick Hynds: Right, because they want to lower the price of the disks. Because, if the mean time between failures is 300,000 hours, that's a long time especially when you can afford to lose a disk and still work.

Richard Campbell: Because you are running in some kind of redundant array that tolerates this.

Patrick Hynds: Right, and now there is Dual pairing, so you can lose two disks.

Richard Campbell: Right.

Patrick Hynds: So, we are getting to the point where I would actually want my Cardiac Care Unit running on these storage devices as opposed to being horrified by the idea that a PC might control my heart rate.

Greg Hughes: One of the unfortunate facts of the world even today is that there is still an awful lot of places that are running mission critical applications on a machine that has been sitting under somebody's desk for a few years. The beauty of centralized storage and being able to connect to that as you become independent of that machine, but also you can become much more reliable.

(00:25:00)

Patrick Hynds: True, and again what we talked about alluded to before is, the disaster recovery scenarios after 9/11 have become much more sophisticated. It's disk-to-disk over long distances is a much more interesting thing now, because people see you actually have to have recover.

Richard Campbell: So, two different backup scenarios. Is the local backup scenario essentially dead, the disk-array dies, I need another one here to switch over to, just because that array is redundant enough now it's just not going to fail?

Patrick Hynds: No, it depends on your data. So, I was actually working on a whitepaper and I haven't written it yet, about data classification, for exactly what you are talking about. So, let's take an example of my organization. We do software development; we have some commercial products that we are about to release in the security space, and we have got a sister company that's historically been an ISV. So, when I look at backups for our organization, all things are not created equal. Think about encryption, if I want to encrypt the secrets to the nuclear arsenal, I want that encryption to stand the test of time. I don't want you to hack it, you can hack it for three years, and I don't want you to break it.

Richard Campbell: Right.

Patrick Hynds: So, if I am firing artillery and I am going to transmit the coordinates that I am going to hit in 40 seconds, I can use much weaker encryption.

Richard Campbell: Yes, because the data is not meaningful a minute later.

Patrick Hynds: Right, anybody who would care about it is gone by the time they could break the code. If we do it right.

Richard Campbell: Spoken like a west pointer.

Patrick Hynds: Yes, I got to get those military analogies in. But, the same is true with data, all data is not created equal, and classification is the missing component. There are a bunch of people out there claiming to do classification, almost no one does. So, let's take a look at my organization. I got a Visual SourceSafe repository.

Richard Campbell: Right, source code place.

Patrick Hynds: Critical, I have got Exchange, critical. I have got a financial package, invoicing PO's, all that critical. Customer database, CRM, critical. Home directories, important but I am not going to die if they don't come back. Websites, I would be annoyed if I had to recreate them, but technically they should be in my SourceSafe, shouldn't they?

Richard Campbell: Right.

Patrick Hynds: How about copies of my CDs that we posted into a directory, so we can install from them. That's not even worth backing up. So, there are categories of stuff that I can't live without, stuff that I would be annoyed or impaired if it was gone, but the world would go on. There is stuff that I would probably actually be better off if it went away. Have you ever had a PC die, lose all the data and be relieved?

Richard Campbell: Oh absolutely.

Patrick Hynds: Because, you knew that there was just so much muck on it that the little good stuff you lost is worth the sacrifice.

Richard Campbell: Ultimately the cleanup cost of that machine hadn't died is higher than the value we would have gotten back from it.

Patrick Hynds: Exactly.

Richard Campbell: How cynical is that?

Patrick Hynds: It's realistic though. So, I think the local backup is still viable. For instance, we have several sites, we have offices overseas as well as here in the US, and I am in New Hampshire right now, we are opening an office in Massachusetts, and we are going to be replicating our data, our mission-critical data

between those two sites. So, we do a backup here, and then we copy that backup over a VPN to that other location. So, that way if a tactical nuke goes off here, we have got the fallback plan.

Richard Campbell: Let's go with just a site catastrophic event.

Patrick Hynds: So, we are right next to an airport, so we keep thinking about the plane that's going to hit the building. So, does that qualify?

Richard Campbell: Yeah, but it could also be a -
- I mean I like a site catastrophic event, because it's just like the site is going to go down, I have dealt with places in the hurricane space, think about the plan you want to have when you are in New Orleans and Katrina is coming. So, now we are talking about an offsite backup solution.

Patrick Hynds: For the things that you care about?

Richard Campbell: Right.

Patrick Hynds: Again there are things even if you had extra time and money, you wouldn't spend it on getting that stuff over, because it just doesn't matter. I mean maybe I install Office from a share on the network, it's data on the network. Do I need to back it up? No, I have got it on DVD.

Greg Hughes: One of the interesting things that you mentioned, when you were using your analogy related to 45-seconds, the coordinates that I am sending out for the mortar round doesn't apply anymore. Another kind of an interesting time based data backup or movement of data, and where do you store and when do you backup question comes into play, and I think that we are seeing more and more of this now related to what information is it that I touch on a regular basis on my network versus what information is it that is just sitting there, but nobody has touched in quite a period of time.

(00:30:01)

Patrick Hynds: Right.

Greg Hughes: What's the value of the data based on that and how available does it need to be? Like a lot of the backup question now isn't just, do I have a copy of it, but what form of availability do I have? Where is technology going or what's the current story in relation to that?

Patrick Hynds: So, again it comes down to the classification, because most organizations don't have as clear a handle on what data matters to

them as an organization a size of ours. I know crystal clear what data matters, and what data doesn't within a certain extent and so it becomes challenging and hopefully I am not going off topic, but have you heard about what happened to Merrill about last year? They were given a court order to produce some data that they had, that was on their network, but they couldn't find it.

Greg Hughes: Absolutely.

Patrick Hynds: They paid a \$10 million fine for not producing data they actually had.

Richard Campbell: Wow! Nice cost of operations there.

Patrick Hynds: Right, so the problem is if you talk to Gartner, I don't know what their exact numbers are, but I think it's 20-40% or 45% of everything on the average network, this isn't like the worst run networks in the world, the average network is junk, the stuff that just shouldn't be there, it's either no longer valid or it was never valid. The big problem that people are fighting beyond recovery is, are we recovering the junk along with the good stuff?

Richard Campbell: And consequently, are we backing up the junk along with the good stuff?

Patrick Hynds: Here is a scary statistic. I will just ask you the question, you may guess correctly but, most people don't. What's the size of a one hour uncompressed HD Video, roughly?

Richard Campbell: An uncompressed HD Video, I don't know. I mean even MPEG 2, that's about 2 Gigabytes.

Patrick Hynds: Yeah this is half a terabyte.

Richard Campbell: Half a terabyte.

Patrick Hynds: Yeah, so I tape one hour of March Madness in HD, to my one terabyte USB drive that I got for Christmas, and I don't think about it and I copy it down to the drive overnight, and I come into the office and I plug it in and I decide to throw it up on the server to share with Bill, and I copy it while I am out during the day and I don't even pay attention to the size, I could be a terabyte or more. Two hour football game is going to be a full terabyte.

Richard Campbell: Right.

Greg Hughes: In an unintelligent backup system, you are backing up half a terabyte worth of HD video.

Patrick Hynds: Right and it's probably worse than that, because I might even crash the server by consuming all the disk space.

Richard Campbell: Yeah, you have found a pretty extreme example, but it's true that these things happen. This is how we end up with getting two megabyte photographs from our friends, because they just don't know about JPEG.

Patrick Hynds: I have seen the results of an employee at a client, who was terminated for a cause, and then they find out that their entire home directory is a terabyte of YouTube videos.

Richard Campbell: Right.

Patrick Hynds: Because they have no controls. I mean policies are really the key here. If you know what goes into the disks in the right locations and you can control that, then the users can't get off the reservation if you will. For instance, who was it we were working with? Merrill, it was Merrill Lynch, they had a problem where they had PST files.

Greg Hughes: Outlook data files.

Patrick Hynds: In home directories, now you wouldn't think that's a problem, but they had incremental backups. Well, one thing about PST's is if Outlook connects to a PST, even if you never read anything out of it, and you never write anything to it, it modifies the date and the archive date of the PST, which means all the PST's in the organization were getting picked up on incremental backup, and the incremental backup were terabytes in size.

Richard Campbell: Of course they were, and yeah you don't actually get an incremental backup of a PST, you get a whole backup of PST.

Patrick Hynds: Exactly. So, they were hoping to just capture the PST's that had actually been modified, but they were getting the ones that were just connected to and most of them were connected to automatically. So, they used our sister company's product, NTP Software QFS to set a policy that said, you can't put PST's in your home directory, they go in your PST directory. So, they completely eliminated the problem, and did some other policies about MP3's can't go on this disk, and you can't put any video content here and that's the next generation. The next generation is to actually control what you can do with the storage, not whether you have access to it or not.

Richard Campbell: Provisioning is just not enough. If I limited you to 100 gigabytes of personal storage on the network, you wouldn't have written your March Madness file out to me. But, that's not necessarily a good enough solution?

(00:35:02)

Patrick Hynds: Well, because I would have written out a hundred -- what did you say, 100 gigabyte?

Richard Campbell: Yeah.

Patrick Hynds: You would have given me 100 gigabytes of family photos, and wedding invitations, and MP3 files, and bootleg copies of Shrek and everything else.

Greg Hughes: Typically followed by a request for additional storage because I have run out of space.

Richard Campbell: Right.

Patrick Hynds: Exactly. Merck is another one. So, I should clarify this, NTP Software is a sister company for CriticalSites, they are our biggest customer, and I do work with them quite a bit. So, I am intimately involved with their products. NTP Software does a lot of software related to this specific problem, policy based management, storage, firewall basically. And Merck had a situation where they wanted a lights-out strategy. They wanted to set limits on users, but they didn't want to have the IT group managing increases to those limits, because users use up the space they're allocated, and if they reasonably can use more, and there's a business use, you want their manager to approve it, not go through the storage guy. So, we implemented what's called Self Managing Storage, where when they hit a limit, they can go through a website, and they can request for more storage, but part of the request is an analysis of what they're using the storage for.

The manager gets an email that says, this user wants an increase in storage, this is how much they're using, this is what they're using it for, and the manager can look and say, they have a gigabyte of MP3 files. Send a message to the user - get rid of the MP3 files, denied. So, getting the user out of the business of competing with storage is really the next big push. Trying to get the user to be part of the solution as opposed to part of the problem is the real big push. I mean I kind of dragged us here from backup, but backups have traditionally been the problem, because we're backing up too much junk. If we can increase the quality, get the signal-to-noise

ratio to a logical place then that's the big win in storage management.

Richard Campbell: Alright. We've danced around offsite solutions. What methods have you used, Pat, to make offsite work for you?

Patrick Hynds: So, I am a low-tech guy sometimes. I know that doesn't show, but I kind of like the idea of doing the backups local, and then piping them based on available bandwidth or when you have the time offsite. There are solutions where a bit's written here, and a bit's transmitted over the WAN and written in Chicago at the same instant. Those are organizations that are spending \$100 Million plus on storage management a year.

Greg Hughes: Usually you hear the term replication associated with that.

Patrick Hynds: Yeah. They have more bandwidth than most small suburbs, between their sites. I've weaved out with those companies. In fact, there are a couple of companies that we helped recover from 9/11 in New York, that, that was their strategy, and that's one of the reasons that they were back up in hours. We had one financial institution we worked with, and I can't name names here, that we had them back up through a lot of help of their own in four hours and another one in eight, and the difference really was just about process and procedures, but they were using most of the same technologies.

Richard Campbell: Anything under a day, and you had a pretty good plan.

Patrick Hynds: Yeah.

Richard Campbell: That's redundant gear, a separate copy of everything; you don't have time to configure much of anything to get back up in a day.

Greg Hughes: One of the interesting things about where the world is now and where technology is moving right now is if you're over a day, there is a good chance that you're flying tape somewhere, and actually bringing something up. You start to get into the hours, and you're talking about occasional replication. We even have companies now that are literally live in more than one location, doing truly distributed applications and storage, and you're talking a matter of seconds to a matter minutes to be able to be fully synchronized and fully online, and the loss of one site is really something that is almost a non-event from a production standpoint.

Patrick Hynds: Yeah, but when we do security audits, there's a definite break in the character of the conversation. We do a lot of security audits here at CriticalSites, and I had a conversation with typically senior management, sometimes the owners, depending on how big the company is, and the conversation is always kind of touchy, and that is, "Okay, you lose this site, it's completely wiped out. You've got a Disaster Recovery plan, this is what you're going to do, we've already talked about this, we reviewed this. What if all the people are gone too, what if 95% of the staff doesn't survive, or is not available anymore? Do you continue on?" In a company with under a thousand people, the answer is typically no.

(00:40:07)

Richard Campbell: Interesting truth. I am just thinking about this idea of Active/Active Redundancy of Data Centers; you have two different Data Centers that are doing similar work in different areas and are the failover site for each other.

Patrick Hynds: Kind of like what NetApp does.

Richard Campbell: Yeah, it's the same scenario or what you do in SQL Server and so forth. I mean if being able to recover quickly is the critical feature, then the real answer is two data centers that you're using routinely.

Patrick Hynds: Yeah, it's a tough sell though, because the Fortune 1000 and even some of them not so much are hard to convince, because this is for something that might never happen.

Richard Campbell: Right. Now, I think it only makes sense if you can actually make that other Data Center do valuable work on its own, and actually split the load between the Data Centers. I've certainly run into the issue with my customers where we just couldn't get more space in this Data Centre, so we went somewhere else, because we needed the space, and ended up with this advantage of having a fail-over solution.

Patrick Hynds: It's funny you mentioned space, because the other factor that I'm seeing more and more is companies saying, "I'm in Central London, I've got a Data Center, and I can't put one more watt or get one more bit of AC into this Data Center, unless I get something out."

Greg Hughes: Because the power is just not available, it's not always the floor space quite often, it's the power.

Richard Campbell: It is electricity; it is surprising how much that comes up.

Patrick Hynds: And the HVAC to get rid of it.

Richard Campbell: I was doing the numbers while you were talking about how much space a Petabyte takes up, and if you're using 500-gig drives at high density, 13-drives per three U's, on a 48-U, you're still talking about a dozen full racks.

Patrick Hynds: Really?

Richard Campbell: Yeah, maybe I'm doing my math wrong.

Patrick Hynds: No, you might be right. A dozen full racks.

Greg Hughes: Well, if you think about what it would've taken it to rack a Petabyte five years ago, a dozen racks is nothing.

Patrick Hynds: No, I guess I must have been thinking about a 100-gig, a 100 Terabytes then.

Richard Campbell: A Petabyte is just an awful lot of place, and irrespective of the fact of the cost of all those racks. Then you get immediately into, how much electricity, how much 'H' fact for that storage?

Patrick Hynds: What I really like is the metric of, how big is a Petabyte, well 5-Petabytes, I am told, depends again how you do your estimates, but if you took every word ever spoken by every human who has ever lived, including these and turn them into plain text, it would fit on 5-Petabytes.

Richard Campbell: Yeah I'd buy that.

Greg Hughes: Or every item ever sold on eBay.

Patrick Hynds: Whichever comes first.

Richard Campbell: We're well into this idea now of what is valuably stored, like you can gather a tremendous amount of data about your systems and about your customers, and how much of it is truly valuable.

Greg Hughes: I think also, there is the question of what is valuably stored where. Kind of going earlier, if there are files that I maybe need to be able to have access to, but I don't touch them, nobody has touched them in the last two years, then what level of access to those files do I need, and what's the cost of the storage where I'm putting those?

Patrick Hynds: Yeah, Hierarchical Storage Management is another of those, "Oh, I wish we

could do that," I can't say no one is doing it, but the problem is, how do you do it, and how do you classify? If you can't classify, it's tough to know what and what not to move. Ideally, you have your primary storage, which is expensive. You have your secondary storage, which you only move the stuff that you will need. So, let's say half of the stuff never makes it to secondary storage, and then you have tertiary storage, which is the archive, the record books, and hopefully that's less than 20% of the size of your original storage. But, then it accumulates.

Richard Campbell: So, ultimately it ought to be bigger.

Patrick Hynds: Eventually yeah, but again you have to understand your data and classify it correctly, and it's something no one's got a handle on.

Richard Campbell: I see a theme here Pat. The real challenge of good storage management is effective classification of data.

Patrick Hynds: Yeah, well, you do it in your home life everyday, you know the difference between good furniture and junk, you know what goes in the garage versus what goes in the attic, versus what goes in the yard for the burn pile.

Richard Campbell: Right, yeah the cutlery in the drawer, the cutlery in the china cabinet, and the cutlery in the attic.

Patrick Hynds: And in the safe deposit box.

Greg Hughes: But if you take all the cutlery and throw it all in one drawer, and just keep tossing stuff in there, then ultimately you're going to understand that you may have an idea of what's out on the surface, but what lies underneath is something that is no longer very apparent.

Patrick Hynds: Which is how everybody lives right now.

Richard Campbell: Yeah, Data wise.

(0:45:00)

Patrick Hynds: Yeah, I mean we've seen the sitcoms where you walk into the apartment, typically Friends I think, and the place is just, it looks like a bomb went off. You couldn't find anything in that room if you tried. Well, that's how we live in our storage world everyday. Most corporations couldn't lay their hands on anything that they haven't used recently to save their life.

Richard Campbell: Literally to the tune of \$10 Million.

Patrick Hynds: Exactly.

Greg Hughes: Yeah, there's an old adage note, if only we knew what we know, then everything would be fine. So, that corpus of knowledge at a corporation or a company or an organization, and the insight into that is still, and it has been adage that's been used for the last 10 years, but even still today to really be able to have deep insight into that aggregated knowledge, which is of course taking up storage, it's something that's still quite a challenge.

Richard Campbell: Pat, thanks so much, really enjoyed talking to you.

Patrick Hynds: Thanks for having me.

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

[Music]