

CAS WEBINARS
EQUINOX TRACK 2: SYSADMIN INTEREST GROUP

JUNE 11, 2020

CAPTIONING PROVIDED BY:

CAPTIONACCESS

contact@captionaccess.com

www.captionaccess.com

>> Good afternoon, everyone. We are going to get started with this program. This track is sponsored by NC Cardinal with captioning made possible by Equinox Open Library Initiative. And we would like to thank our captioner. We would also like to thank the other conference sponsored for making this event possible. Mobius, Bibliomation and Evergreen Indiana. The event is being recorded and will be available on YouTube following the conclusion of the conference. We would like to encourage everyone to use the chat window to post questions. The facilitators will be collecting questions along the way and causing them to the presenter at the end of the session when she asks for questions.

We would like to encourage everyone to chat freely, and changing the text a little. We would like to introduce Chris Sharp with Georgia Pines.

>> Thank you very much, April. If you have not participated in this before, this group traditionally has happened at each conference since 2010, maybe. And it has traditionally been a fairly informal attendee driven discussion about topics of interest involving Evergreen system administration. It has actually broken up into side groups. That is not possible under this particular Zoom situation like it would be if we were together.

And we typically have a note taker who has done this in a shared Google document. That could be me or it could be someone else who volunteers. But basically, giving us some notes. Not just the recorded record here but an actual document that you can do. So we have a question, are we in presentation mode, and I don't know exactly what that is.

>> I think we are in meeting mode.

>> This is meeting board. Which is why you're able to have video and audio.

>> That question might be coming because with the desktop share we can only see four people at a time.

>> Maybe April, if you stopped sharing the image.

>> I can see.

>> It depends on the screen size.

>> We have gallery view versus speaker view on the right. That will give you more faces. Now, we are in Brady Bunch mode.

So just depending on the room, you can switch between screens to see everyone.

So anyway, as I mentioned, this has been kind of a meeting without an agenda. But I have some ideas of things we can do if no one has a particular option, like a particular topic that they want to delve into or questions that they have always wondered. Even though there are lots of experience in the room, we do encourage newer people to ask questions

that they have concerns about so that we can assist. Because a lot of us have been doing this for many years. I started at GPLS in 2008 have and have been doing this work since then. Some of us have had similar career arcs.

So we are open to questions. We can do that, but if there are no questions, I had the idea that we might just sort of do a guided tour, not of the install process, which some of you might repeat, but maybe a run in Evergreen system so we know what the different moving parts are and how you might take a look at those parts and know what to administer, what it actually means to be a system administrator.

And that idea came to mind as something that people can benefit from and do.

If there are questions about how we are administering, hardware, virtual environments, containers, cloud hosting, that kind of stuff, that is all stuff we have talked about in the past. So I will open the floor and you can use the chat feature or you can speak up with your microphone and we will see what we can do.

>> I will start. Has anyone done any tinkering at all with the -- as it relates to Evergreen?

>> Blake?

>> I'm actually planning this your very soon putting all the open I'll don't open IOL listen to their own containers and introducing more containers for each service.

>> Excellent. Because I have seen -- your documentation. I've used that in the past. S thank are putting that there. It was helpful.

>> I have gotten beyond novice but not much further than that. And Chris and I and the GPLS Pines people, continue conversations about the playbook pickets all under the same umbrella so I would be curious to see it -- if you show that, what it looks like.

>> Thanks.

>> I'm excited about the idea of using more -- resident container for VM. Right now you take the branch I got out there to look at the repo and tell it to go, it will create a container that runs Evergreen. But that's it. It does a lot of stuff, no doubt about it, but it is all in one container. So the next step that I'm really hoping to get to is to divide each little piece up and deal with open ILS routing. (laughing) what's up, Bill?

>> I'm upside down. fact it matters your background.

>> So I don't know if open ILS -- anyway, I'm not sure, I don't have to edit sorry, OpenSRF is what I was talking about. I don't think I have two edit OpenSRF. I think the OpenSRF XML can refer to all the pieces, and then

the router would look it up. And maybe there's a special DNS. I haven't fleshed out all that stuff. But it's something I have been thinking about.

>> And for people that don't know -- it might be like us, this is for me. -- is a software application that will allow you to herd all of your things in a correlated way. And that sort of the high level estimation of what that is. And it sounds like Adam is curious and Blake has actually done it. So that's where we are. I did my people to get lost too much in the weeds about what that is.

>> I will add that we actually are using Kubernetes on the Google cloud. And it does work with the container. Evergreen is automatically the end load balanced and healed and health checked by the Cooper nays engine. So we are doing that.

>> Following up on that, Blake, does that make it more reliable, do you think? Have you really seen a benefit over running on -- machines in the past?

>> Yes. It's more reliable. We didn't use it to begin with, and I remember those days and troubleshooting which app servers, and all that stuff. If you have your health checked soporific, it will find the bad one, kill it, and make anyone from your container by resolving whatever before anybody knows, before we even know half the time. So it is self-healing. And that increases reliability for sure.

>> Getting in from the chat, Jane said she's been using to compose with other projects.

>> I'm liking a little behind on updating our image. This discussion reminds me I haven't published the version that you might want to be wanting to test with, so you get around the Mobius office, and it will just magically bring a server into existence. It's been around a couple of years. It seems like we have around 8000 downloads. So people are using it.

>> To avoid the shame of that, Blake, I'm going to just put that in the notes here.

>> Have you actually found yourself needing something?

>> Not me personally. That the kind of thing we have always had somebody who has taken on -- there was VMware and virtual box to kept up-to-date for a while and then started lagging behind. It's just a mouth to feed and we don't always get to it. We actually were providing the packages as well. So that project, we use -- we are no longer sort of putting them out there. Because we didn't see a lot of interest or use from the community. After.

>> That's pretty much the same here, too. We had a bunch of scripts, we can build a bunch of machines. As Chris says, that -- is another mouth to feed and it isn't helping anywhere else.

>> Yeah.

>> What I've actually been doing lately is virtual machines, I billed three models, basically. We use bricks. So each brick will be virtual machine host, three or four virtual machines on it. And I make three models, and deny condom when I need them and then the IP addresses and that sort of thing. So that's what I've been doing.

>> Reacquainting myself with these containers. 3, 4, 1.

>> I want to hear from people that haven't spoken up yet. Are there things that you're curious about, don't understand quite yet? Need some guidance on?

>> Bill is having a lot of fun with scraping stuff while ignoring the usual means of turning them away.

>> You mean like the TFC file?

>> Robots.txt and so on.

>> I use IP tables and drop. That's one approach, is kill their IP if they don't respect robots.txt. back I've always suspected that to be an honor system.

>> Given the annoying bots --

>> Or a legit company that behaves maliciously by ignoring them.

>> Or it's Google?

>> No, it's a turn it in bot.

>> On the subject of robots .txt, but on that subject, there are several examples on community servers, you can go to the web root and look at it. In some cases, their elaborate setups where they allow certain directories but not other directories.

>> Hours are pretty much no robots. And then things like directives in nginx, if the string matches a certain -- I forget the name --

>> User agent.

>> User agent. To turn them away sometimes that doesn't do the trick, either. So okay, IP tables, you are done. It would be nice to automate that a little better.

>> We have had things on the Evergreen server, but on web application service. Jason were you going to say something?

>> You could do it by user agent. And I think some of them are using an extremely pedantic, very specific way to parse robots .txt. Her people may or may not bother to write, so they think that they are not being bad actors, just extremely picky.

>> Ignoring star is pretty impolite.

>> It space, where they got should have a tab. That sort of thing.

>> And SQL injection attempts seem to be rather rampant and most ignorable. But occasionally they clog stuff down.

>> Anybody else out there? We have a bunch of quiet people who have already switched to happy hour mode or are asleep at their desks because they have been looking at Zoom for days on end.

>> To follow up on Jeremiah's comment, what I see is often stuff for Microsoft SQL server, and it's like loads of parentheses, and I get a search box, which in they put in a password search to just run forever. I think there's a word about that. It's improved over the years. I remember seeing that quite a bit.

>> We had this one this morning that just caused headaches until I got them there and. I missed have the presentations I wanted to actually pay attention to, stuck in that, trapped basically.

>> All right.

>> I got a question. How do you go -- end up identifying while running queries for the database? I've had a lot of auditor queries long-running that end up actually crashing the database because it runs out of memory. I've tried to set some limits to it, but sometimes they just end up catching.

>> Those are just sitting there. They are idle sessions, and the last thing that was done was select auditor. With that what you mean?

>> Yeah.

>> They are holding the connection open. We have seen the same thing. I don't know why they don't sort of release the connection, whether that's intentional to have OpenSRF be efficient. So Bill, do you have any ideas about that?

>> Like a lot of us, I'm doing too many things at once. So the question was on queries?

>> Chris was mentioning things are sitting there idle, and the last thing they have done is an auditor query like -- I forgot what the terminology is.

>> Shut them down. Well, one thing we do is we volunteer for idle transactions and if they are out of for a certain amount of time will go in there and kill the transactions if we have to.

I'm not quite sure how to prevent those. I don't what the cause of those are.

>> I just want to clarify. You're talking about any transaction, not just idle?

>> Yeah, I was. back some are idle and some are --

>> Okay. I've not really seen too big of a -- idle. We get long-running queries that are active. And we use -- we have a script, I will have to get out, that basically, every time, it says give me all the running queries and checks along they have been running. If it's been running over so many hours, I get an email. And I go in and see what it's doing. And most of the time you can cancel the back end or whatever it is. On the back end process. Because

there end up being searches that for some reason I just never ending. It's not as much as it used to be, though.

>> I actually got a SQL from Jason Boyer, I think, that attended to pull out the IDs, Helen the queries have been running, like the first 100 or so, there is a SQL that they are running. I can put that up later if anyone is interested. It may already be on my -- been no writing about it.

>> I think everyone is doing what I'm posting a chat. It's exactly the same issue where someone wants something -- I don't know if it's an optimization issue or something in the way that we are formulating the queries that just means it's never going to come back. I think all of us see that. And Chris Burton, in our case, this is the thing that makes our database start slowing and eventually stop receiving new connections is if we don't have something like this running to kill off the bad -- that did searches, they do pile up. And becomes an issue.

>> Yeah. Sometimes I get some of those bad bib searches that sit there. Okay.

>> Our favorite is the staff person that goes in and make the search and basically put something in that ends up being -- user. That one is never going to finish either.

>> That could be one of those autogenerated interfaces where it's trying to render a select drop-down of all the users. I don't know if there is a way from the staff interface to force a -- and everyone. Pretty long.

>> To be fair, too, based on what I see in the database, on the queries thing, not actually looking to the logs to see if I can find the actual deck if you month ago I think I found one. It was someone at this library that basically did something.

>> That has happened before. Where an autogenerated interface was mentoring a table and I had a link on the table for users and you have a list of all the user so you can pick the user? So we just say give me the whole table and at least in one instance that's been fixed. And there other bugs opened the have to do with preventing that generally for tables that are assumed to be larger. But that's still floating out there.

>> There is one the carousel admin interface where will start links to something -- I've forgotten, every user setting or something, it was like digging deep and just returning thousands and thousands. I think that's not been fixed yet.

>> The other thing that goes along with this, because if it's a case of something that the staff person is doing, and I'm not blaming staff here, it's a bug in the software. But they will try it, and it will not come back and they will try it again. And very often I will go in and see three or four of these running.

More or less simultaneously because the first one hasn't come back yet. So will see that sometimes, too. There are some things in decode for that, but I think it only works if one of them has returned. And Bill, I don't think Mike is around, but I think Mike wrote. So correct me if I'm wrong -- I'm wrong, but I think that's the way it works. If the person comes back, the other one stop run. There's something going on there.

>> Do you think it was specific to search maybe, the case where enter was getting tapped a bunch of times and the exact same search would come through? I don't know if that applies to other things or not.

>> I have a query in chat, running the same search. I was trying to get to that finding somebody who held down the enter key. And generated like a billion queries on the database.

That little snippet seems to work, and I would monitor that output and I would get in along, saying someone's -- sent an alarm saying someone holding down the key. Something like that.

>> We have our clients stripped down our whole system because I bring a book down and right screen or wrong screen, put the book down on the keyboard. You just hammered the database.

>> That has happened to us.

>> That's hilarious. Brings down Georgia Pines.

>> And it was a great day. So this happened on that great day and some kid or somebody in the children's section -- we were thankful for the way they named their web station script we were able to figure out which one it was. But just sitting there, it was hitting F5, and it was just reload, reload, reload. And anyway --

>> I just put the book on the table and the whole server crashed?

>> I think that is 72 -- or somebody had they debug down and shut the whole system down.

>> You can't break it.

>> Don't tell the hackers.

>> You can do anything, just not put your book on the keyboard.

>> Protect the keyboard at all times.

>> That's effectively what I am talking about.

>> Nothing happens to Evergreen except for loaded down going what is this?

>> There should be ways to adjust some of this in nginx, too. One of the big things I came to a years ago was some of these things can be resolved with a proxy server. But if you are using a load balancer, it's in front of everything, then woodlot the proxy, you would want to load-balancing proxy.

Filter out some of those two picket requests. I have seen examples of those for nginx online. It's something definitely worth looking into.

>> The last session of the day after a venture of using Perl for database and developer update, and a Perl tour and just the 4:00 witching hour, I thicketts okay that we are low energy. But I don't want anyone who hasn't spoken yet to feel free to bring up any questions they have or issues. Because we are here to learn from each other. And I'm not trying to badger anyone, I just want to encourage participation.

>> I have a question about search speed. Does everybody feel like there's is slow?

>> Yes.

>> Yes.

>> Come on, Bill. This is where we do a big old high-five. Bill really backed me up. Don't tell anybody, but me, too.

>> I didn't say that. (laughing).

>> I have been whispering about it for years. Jeff has but was it?

>> Its elastic search. And Bill made a branch expanding on that. Integrated into a ranting Evergreen, and as far as I know, breading production?

>> No. In fact, we did a testing cluster if you weeks ago. Because I only integrated it with the staff, I can't get it going until we have worked out all the issues with that. So that -- we are just now starting to work with that.

>> I still plan on putting that into the public side. We will see. I have a couple of searches running, and it's really looking promising. It's looking like a tiny little bit at a time, it could be very granular on the SysAdmin side to enable elastic search for certain kind of searches. I was thinking along the lines of if it was a keyword search for consortium, for over and use elastic search as a starting place, maybe opening it up for more and more types of searches as we go on.

>> Yeah. You might get slightly different behavior. I mean, for example, not every feature is going to be -- so the search might return different results to checkboxes and toggles are activated. And as a heads up on that, I am calling my branch over to github, and it works great, but make it so the indexing was more segregated from Evergreen so that I could re-ingest pretty quickly. I have a little over two or 3 million bits that have reindexed. An actual full index would take long to than that.

>> Github somehow makes that faster?

>> No, I made a ton of changes. And I will point you to it. And part of the reason I didn't leave that on the mainland, is it creates a much more significant departure from the way it is working. And it's creating a separate indexing process. It's a little bit harder of a cell, but I will keep my eyes on it.

>> Something entirely separate from the full text search?

>> Yes. The indexing assault happens in elastic search. And if it gets to pushed to Elasticsearch, it's extracted separately from the configuration that Evergreen has for extracting search indexes. So it's a different thing that is doing. It uses XLT. It's basically just one style sheet. So it's just one style sheet, record, record, record. It make that fast.

>> Does the data update, does that happen at regular intervals or does it sort of update live, what?

>> The indexing script has a mode that it can run that says only update records that have been modified since X time. I'm going to start running every 5 minutes, and it will look for call number, copy, changes. That's what is looking at right now.

>> That's cool. I have thought about him lamenting something like this, but I have just never gotten to it. And I always have had the excuse that I didn't want another mouth to feed. But that is a consistent complaint from our users that that is usually a lower score that we get on our patient satisfaction surveys and things like that, that we put out. I think actually, that issue has made the difference in at least one Evergreen evaluation. Should be moved to Pines evaluation, that we have seen in libraries in Georgia because it's concerning. Is it something that can be mitigated or fixed? But they don't go any further than that. They just seek what it looks like right now.

>> We look at that and say this isn't good enough. We don't have other options.

>> Chris, just a personal comment there, you're in the cab.

>> Affects back going -- it's not a faster than Pines. I've used Dynex all the time.

>> In Georgia they have reason they would not want to go on to Pines, nothing but I just want to do it. This might be like finding reasons not to. But was free remove that obstacle, that excuse can be used. I'd love to take a look at what you've done, Bill.

>> And I don't know when we are going to be able to do production, because there's so much going on, but regardless I will do a session on it whenever I go through the initial rollout.

>> Not to take away from elastic, but I have been expanding took us remitting with more versions of -- I have a server with 9, 10, 11, and 12, and I see huge speed increases with 11 and 12. 12 is actually sometimes soft much better than 9.6, I think it's actually going to work. Something that took several minutes on 9.6, finishes in a matter of seconds on 12.

We have an issue with updating employee records, particularly when you have someone -- one example I'm working on is there are hundred 50 records associated with an authority. We have to update all those records, and many what going on and what is in cache, that can take 6 minutes, which is way too long.

So on 12, that finish in less than a minute. I don't member the numbers off the top of my head, but it was a lot faster. It was so fast I had to go in a double check that everything actually happened. I thought maybe it blew up. But it finished.

>> That's a nice problem to have.

>> How does 12 work with Evergreen?

>> I think is at compatible is 10. So if you fix the handful of functions that need to be fixed and maybe one or two of the storage -- then basically if you install the latest version, should be good to go.

>> We are looking at -- I was updating to 12. I had some things in the meantime, because I have never been able to set the yet, but that's one of the biggest things that we have.

>> You're saying, I'm 3.4.2, I think. So you're saying, I'm ready to upgrade to 12, and I do everybody will love me?

>> I would caution.

>> I would say I wouldn't do that in production right away.

>> I can't guarantee that everything works. I've looked at a few other things, that weren't broken with 10, and everything I have looked at so far has worked. But I haven't really stress tested a system on 12. But from what I can see, 11 is -- so 10 is a little bit faster than 9.6. 11 is a lot faster, and -- then 9.6, and 12 is a little bit faster than 11, from what I've been able to see.

>> I was going to say that other significant factors and just database feed in general are the hardware is running on. And whether or not it's configured correctly, optimized for the hardware it's running on. So figuration is a come, nation of art and science. I'm sort of paraphrasing something someone said to me at once, and you have to learn what those are. There is documentation out there pick sometimes it takes a little bit of research other than just reading the documents to figure out what each section does, and what each setting does, and it's very much -- you have to experiment, but that's kind of the only place because that's what you're trying to fix. So you might tweak things like memory allowed, and what sort of storage it's running on and how fast the processor is, all that is going to be a huge factor.

>> I want to add that in my case, this is all the same hardware because I have four versions running simultaneously on the same machine. I use the

same -- I basically get the production data, so thought production data. Same hardware. 10, 11, 12 I am running on default configuration. In the 9.6 has been optimized, but not to use the full hardware pickets and optimized partially. Anti-hardware itself is not equal to our production server, it doesn't have enough RAM to hold the whole database. But it has tons of disk space. But the RAM is not the best. I do see improvements going from 9.6 to 12.

>> What's the highest version that we have rented the Evergreen system?

>> We are on 9.6. We actually upgraded us one night. This was in advance of December. It was right before 3.4. And because I did not -- it wasn't clear to me that there were photos that were breaking, I just saw tons and tons of errors. So I spent a couple more hours upgrading again to 9.6. After I think I went to -- which is a fun thing to be able to say. That was it. I went to 11.

>> You regaled that experience, I believe.

>> Yeah.

>> So to anybody's knowledge, and .6 is the highest in production?

>> We have already touched on the community server. It is not exactly high stress.

>> I'm running 10 on my personal server. It's a very low provision box.

>> Also, are these fixes to this function, some part of the upgrade happens automatically?

>> If you upgrade test version for the habit -- we made sure that the fixes still worked on 9.6 and didn't cause performance degradation on 9.6, either.

>> I was on 3.2, that's right. If I had upgraded Evergreen and denigrated the database, I would not have seen problems.

>> So upgraded Evergreen to 3.4 or newer -- or newer first.

>> Or use the patches. It's just changing a couple of function calls.

>> And everybody is not showing up. Nobody wants to go to 10 first?

>> Actually I'm planning over they were date weekend. My plan is to move to anywhere version. I originally thought 10, but I'm thinking maybe another.

>> We are talking about going to 12. But that depends on having it tested.

>> Chris, did you say you want to go to 10 or 9.8?

>> That was Jason.

>> No, it was you.

>> Sorry, I was saying I wanted to go I will probably go to something newer than 10. I guess I was being somewhat conservative when I chose 10, and I was back in December. I don't mind being somewhat pioneering. Probably 11 or 12. This will be in September or the end of August.

>> And I'm definitely excited about this subject. Has enemy got any stress testing, type script or anything so we could do it on a dead server and then

hit it really hard? I have one that hits the server and looks up a bunch of words from an open API online that feeds those words into a search engine and parallels 10 at a time, and measures how long it took each one to come back, and that's interfacing with the database directly and with the OPAC so it can post stress queries and is going to the allocation stack. Anybody else got something like that?

>> The only thing we do is that we haven't done that lately, because I've been using the upgrade feature. But when we were doing PG restore, if you do that at three quarters of your processor core number, that's a pretty dang good test. So you've got a box with 24 cores, getting up to 18 or something like that at a time, if it doesn't fall over, that's probably a good test of can it work.

But I guess Bill had that constrictor think years ago, but I don't think we've seen -- that I've seen, anyway, scripts that pound a server that I am aware of it.

>> It's pretty easy if you are really searching -- worried about searching, just to do gateway calls is pretty straightforward. And there is also things like PC bench, which I have never personal use, but I know it's out there.

>> You could just get a cat to sit on the enter key.

(Laughter)

>> The way I determined that PG 12 and 11 and 10 were faster than 90.6, as I use ethical PL profiler. You have to install it from source, but you basically install it. You build the abstention for the database, you create the extension the database, and you can run a query, and will go through until he was part of it are slow. So in the case of say, update authority records, the mentor and said oh, it's spinning all this time doing metabib. So that's how I knew where the slowness was. And then where the speedups were.

>> That's awesome.

>> I will post my thing, but it's a couple of years old now.

>> I have often wanted to write something to stress test the server. And I haven't really gotten around to it because there are so many other things to do.

I just wanted to add, too, that also upgrade the operating system. In doing that made a difference, too.

>> Going from 16 -- unless you install the hardware -- HWE, whatever that stands for. When you're going from 4.4 to 5.4 -- go from 16 to 18 to 20.

>> Anything estimate faster, Jason?

>> They still don't have much love for team users.

>> Why don't we hold a conference on teams?

>> Let's switch over now.

>> Let's do that.

(Multiple speakers)

>> I'm going to call the session to a close unless there are some final lingering questions or comments.

>> Any particular reasons for going with the platform -- one part from than others?

>> I will answer that one. Originally, I think it was beyond generation to coerce nothing crazy. But Bill can correct me if I'm wrong, but several levels in the community around 2013 are so started moving into platforms and that's -- that was our preferred platform for other things going on. So we decided to standardize everything and move it over to -- and that's where it lives. On that note, I actually have a couple of branches that I have been working on over the last few years to increase support for -- and Red Hat. I got things running on both, but I'm still fighting with the dependency problems and C libraries for whatever reason. So I would love for that not to be a requirement. There are lots of shops where we have just say, no. If it had to run something that is not used to running. That's a great question.

>> We're going to be pretty much across the board.

>> Okay. Well, thank you so much for your time today and I guess this is the last session of the conference. I hope you all take care and we will see you on ISE and email lists. So just keep in touch and good luck and you can always pass questions to us if you need them.

>> Thanks, everybody.