

2023-06-14 Fabric CAT Office Hours - AMA

0:02

So thanks for joining everyone. I know that I did change the topic of today's office hours very late, but the speaker, yeah, they loud on us. So that's how it is.

0:15

Before we get started with all your great questions, I will let you know that we are planning on having an official kickoff of yeah, what we are now calling Fabric Cat Office Hours and that will be in late August and I will send an invite so that hopefully by the end of this, this week. So please get your questions ready for today, write them in the chat or raise your hands. I can see that we already have a few people.

0:46

And so Edward, let's get started with yield. I just got off a call with another CSA and they said somebody in GBB said data bricks can't connect to one link yet. Is that true? Can it connect to the ABFS send point or not? Anybody know? So what cannot connect to?

1:15

They they wanna connect to data bricks. They want data bricks to read from one leg. Is that possible. I thought that was possible they were they were being told it wasn't. Yes it is. It is. There is a video from the I'm not sure if it's a that's in the cube with just Kaplan and he shows that excellent of OK awesome data bricks reading from one leg not one leg no I'm sure coming out to.

1:45

No, actually they they are writing the the. Sorry. I think it's Dr. Rick's writing to one link, but okay. Now I understand that your question is the other way around. Yeah. Yep. Take a look. There's a the links in the chat there as well from the doc page on how to do that. Thank you very much. Yeah, I literally got the call and didn't have a chance to check and I saw nobody else had a quote and hand. So I figured I'd jump in and ask. Thank you.

2:18

OK, so check to clarify, 1 Lake is fully ADLS compliant, so you can use the same APIs, so you can rewrite, et cetera. That's what I was saying. But they were saying somebody told them they couldn't yet and I'm like.

2:41

Maybe the question relates to Unity and the Managed Identity access to 1 Lake. Yeah so I suppose the question could be asked differently is can you use a managed an Azure Managed Identity to connect to 1 lake and can you?

3:12

In most technical right, it's it's it's the default authentication on Azure. Ohh, sorry, change identity. Sorry, I got it wrong. Scratch that. Can you can you restate that question? Or was it using a managed identity to connect from data works to one leg? Yeah, it went on Azure, yeah. So can you connect to one leg using a managed identity?

3:43

Because all the paths right now are pass through authentication. Yeah. Did you want me to pick that one up? Of course. So right now, no in the very. This is internal, isn't it? Yes. Yeah. Yeah. OK, good. So.

4:12

On I think it's on, yeah, on MSIT you can use service principles to connect to 1 Lake, but other than that on the customers can't use that yet it's not being deployed out and then eventually we will have, you know other authentication methods. But right now the answer is no, Soon service principles and soon other things.

4:41

Thank you. Great, Mark. Thanks a lot. So I think we can move on to Youton. Next one line. Thanks, Lars. So I have a question on git integration and deployment pipelines. So I've had a couple customers ask me since we announced git integration that why should I use deployment pipelines?

5:06

As well, and I've seen kind of our recommendation is to use git integration for the development piece of it to get to your development workspace and then deployment pipelines from your you know, to your development to your test to your prod. But I guess I'm having a hard time articulating the real benefit of using deployment pipelines for those later 2 stages versus just using three different branches in in a git repo pointing to those 3 workspaces instead, because then you get version control.

5:35

I mean you still would have kind of the same, you'd be able to have the same separation of duties in terms of workspace permissions and and permissions in your DevOps branches. So I'm just curious if anybody else has more, I guess information on why it or or how it works better to use get integration and deployment pipelines together versus, you know, just get integration.

6:05

Yeah. So I think that the, the deployment pipelines and now we're talking about Power BI specific, right? Yeah, Yep, Power BI specific for for these particular questions. Yeah, the deployment pipelines are meant for having your three environments like development says production and.

6:27

And the git is obviously brand new. So I think most customers will still who are using deployment pipelines, will still use deployment pipelines and probably use that in their development workspace and then deploy and and git to production using deployment pipelines. Yeah, I mean that makes sense especially since git integration still in preview.

6:53

But I know with DevOps you can do a pull request from one branch to another, so we could seemingly accomplish roughly the same thing in DevOps and then use that get integration to sync. But yeah, I think I think that's just something that I'm I'm, I'm wondering if more people are hearing or if there's more kind of something I'm missing there in terms of a benefit of using deployment pipelines. I I don't think you miss another benefit.

7:20

That that that's the same scenario that we have today in in Synapse actually and and the best practice is to connect only development to to the repository. There will be more problems down the line than

the benefits of just having to copy to push files between branches. And there's some articles, public ones written by some of us, I can share them.

7:50

Awesome. Yeah, that'd be great. And deployment pipelines have a built in notion of environment configuration that you would have to figure out some alternative solution for with git. Yeah, there's no logical organization in git of this is my dev, test, production, branch type thing.

8:13

Yeah, and which branches and each ended having parameters that are that differ by environment like. Not that it's not doable like but somebody just has to figure it out and see if that's a better pattern. Yep, that makes sense. So I have like a similar question on that. So if you have like you know dev, QA and prod branch in dev, OPS and.

8:41

We have three different workspace like you know Dell QA and production workspace. Can we like link Dell to Dell branch and QA to QA branch and Prod to Prod branch or main branch something like that. Is it kind of recommended practice there or do you want me to share this screen But and then just asked and the answer is we don't really know. We what we know works is you know.

9:10

Do get integration with your dev workspace and and do everything else with the deployment pipeline because that's a well worn path for Power BI deployment. There may be other good patterns that rely mostly or entirely on DevOps integration, but we haven't figured them out yet. Okay.

9:30

And in the development pipeline, it supports 3 and three environment there, right? Right now they have Q and PROD, but what if customer needs like UAT is 1/4 environment between the PROD and QA? Is it customizable the pipeline? Yeah, I was told by Lee Benjamin who's a program manager I think is running deployment pipelines that that's coming this summer.

9:59

Is the ability to get up to 10 deployment pipeline stages. You can use the APIs and do as many as you like, but yeah, do we before it's the same? Is it since the same I thought it was I asked the API would support more as well, but that was I think several months ago. So it could have changed. It was the same. OK, but think about deployment pipelines as the low code, no code way of pushing stuff.

10:28

In having environments for local users, so they don't need to know about any of the DevOps stuff, they don't need to care. They just, you know, it's just a nice way of them to have multiple environments so that you know people aren't trashing the production environment. They can test it and then if you've got something a bit more complicated or maybe you're in a larger organization, then you might have actually you know, proper pipelines that will actually do it and do it on a nightly build or do it on a on a schedule or do it on a check in or do it in a different way.

10:59

Because they don't want users, you know, just pushing stuff into production. So they're they're complementary. There's not really one and then you would only ever use one and then and not the other. That makes sense. That's the kind of another thing that we talked about with one of my

customers is that deployment pipelines are suited well when you have limited developers, whereas the more developers you have working on the same artifacts, that's where.

11:28

Git integration becomes more well not no no git integration, so can so just consider git integration as another feature. Deployment pipelines have got nothing to do with git integration. So when ohh OK maybe is it because we're talking at cross purposes so when people are talking about deployment pipelines they're talking about these in here So click on there.

11:58

I don't think I've got one. Ohh. OK OK great. I don't have one and I'll go something create development and that has nothing to do with my git integration. It's just gonna. So these are ones we don't support yet and so so git. So git integration is recommended only if they wanna do version control right for their source. I mean or artifact. Well yeah.

12:27

Yes, if you want to do version control, it has nothing to do with this. This is just pushing it through the deployment pipelines. You can you can set up each workspace separately and have them with a different git repo so I can manually push through here and this is me doing it as a load code user. Not knowing what I'm doing, I'll go back to my workspace. You know the the git integration that I've got.

12:57

Yeah, right. So with the same scenario here, right. So if we integrate this three workspace with the three branches in a Git, right, Dev, QA and prod. So if we move code from DEV to QA and we come back here and synchronize that workspace like you know, QA workspace that brings the code from git, that's another approach, a possible approach I think. So this workspace is connected to Git.

13:23

So it's the it's really easy to do you just connect it up and I've actually got two things that are outstanding. Yeah if I choose deployment pipeline, the next the this this would have to be checked into git. This this could be checked into completely different git repo, completely different settings. This is just a workspace. This has nothing to do with git. If you wanted to have that version control between different environments then set up git, you know we can have we can have git set up on here on this workspace.

13:53

OK, yeah, the the two features. But I guess the similarity is the ability, the continuous deployment part, the the, the, the fact that you can do it with both and you can do it problematically with both. So I guess that's where the similarity is. Yes, yes, changes across different environments. If I'm doing proper CI CD, you know I'll do it my my dev workspace.

14:15

I can have the others connected up to other workspaces and when they get checked in there, I could do my deployment process, my pipeline. But this is a manual process and it has nothing to do with it's just taking stuff from one environment and pushing right you can you can automate it through the APIs. So yeah yeah you could automate through the APIs yeah and and I think previously people also used the deployment pipeline APIs together with git so. So there is some kind of symbol uses there it.

14:45

If you would, yes. But obviously, remember that the the Git stuff for Power BI reports essentially is very, very new. It's only just coming out now. Yeah, so never really did that before. But now you can. So we got about 20 words, please. You can see stuff is, yeah, see these, right. Let's, let's, let's call it on this topic. I feel like we've spent way too much time here. We have a lot of hands in the queues. Let's be respectful of others that have shown up. Let's move on to Dan. Whenever you're ready, please feel free to come off mute.

15:14

Hey, everybody. My question is around security. A lot of my customers are really interested in trying Fabric out, but it first has to get through the security team, right. And you know right now there's a little bit up on the the documentation and in our material on Gear Up, we have kind of like 1 slide that talks about one security at the very high level, but not really much else and any conversation with the security team is going to be a little bit more in depth.

15:44

And I'm just curious if we don't have anything in the works in terms of like security focused deck or something to take with us to have security conversations because you know right now I don't really feel comfortable doing that. Great comment there. And then I'll call on bank and Sergei if you want to expand upon the birth you're doing.

16:10

Yeah. Is anyone taking the question or I can take it. Oh, you, you, I saw you're off mute. I know you wanted to jump on this one. Please feel free. Oh, no, I was just waiting to see if any folks want to chime in. But so, yeah, thanks. That's a great question. And so we are actively tracking this request. So the generic guidance as of now is to we are in when we are in public preview. Not all the security features and elements have been lighted up.

16:37

So our general guidance in this area is try out fabrics functionality with the non sensitive anonymized data. Also make sure that you evaluate fabric only with a limited set of people, not everybody in your organization needs to check it out. So this is the generic guidance. However, we are actively working on all the security aspects including updating our Power BI security white paper.

17:06

All the securities and compliance work is going on and we are also building all the security features. So by the time when we are near our GA, we will have many more to come in this area. But for now our guidance is to just this the functionality and the security pieces coming down the line. Does that answer your question? Yeah, I mean I I have no doubt you guys are working extremely hard on you know rolling out.

17:35

Security features I I guess my question was more along the lines of would it be possible to share the road map for security with at least internally. So I have something to you know talk to the security teams about like you know, hey this is the plan. I can't give you timelines, but you know this is how it's gonna look. Valid point. We'll take this feedback and we'll circle back with the PM teams on this. Sure appreciate it.

18:04

Yeah. And forever, it's obviously going to be a hot topic for everybody. It is not a good idea to share

any of this information because things change. And then you guys on this call probably familiar with Vnet Gateway saga, how many times did VNET Gateway GA get delayed? It's been over a year and a half.

18:28

So Dan, if you do get your hands on this information for high level discussion, the number of disclaimers you're going to have to file is going to be enormous and even by GA like one security for example, this is Microsoft confidential, it's not locked down yet and it's fundamental like how do you sell Lake House without selling and explaining what one security means. So for now the message to customers should be pretty cold.

18:55

If if those customers are security conscious and you have to say use non sensitive data in your exploration and more information is gonna come and it is Microsoft, you can point to how good security is in Synapse and how good security story is in Power BI with OLS and Ros. Yep, makes sense. Appreciate it.

19:21

Thanks, Venga 1st and Sergey. Yeah. And I just want to highlight again that we are in public preview now and it's not ready for production fabric unless you're talking about Power BI. So so that should also be your, your measure to customer, just to emphasize that. Next up is Andrew.

19:42

Yeah, hi, good morning everyone. So I have a question, I have two question here. So my first question is like in Synapse, right. So we when we write a notebook, so for that we have to create a a spark pool, right. So when we run the notebook, so it will takes, it's took some time to spin up the cluster, the computer engine. So I'm I'm just trying to understand when, when, whenever we are running in the fabric 1 so it's.

20:11

It's running very fast. So what is the engine we are using in the fabric because in whenever we are running the notebook so it's not it's not taking any any time it's just throwing giving us the data. So what is the engine we are using at the back end because in the synapse whenever we run the any notebook it's take it's take two minutes or three minutes or 5 minutes to run to on the computer engine or sparkle. So what what what we are doing in the fabric side.

20:39

I I can help with that one. This is Tammy. So one of the biggest changes we made is we keep warm spark pools running for you so you don't have to, within your sessions, sit and hang out and wait for one to start. It was a huge ask in Synapse cuz like you said, it took forever to start them. Even if you're just developing you hit run and go get coffee. So now what we have waiting in your background while you're while you're there are clusters ready to be used.

21:06

Or it will cost, it will cost you, the customer, right. The cost will be very high, no only when they're running, only when you're using them, when when you're using them in the sense when we, let's say if I I wrote a notebook, right, so it will start immediately or just how how it how it works in the back end, it's while you are executing the notebooks.

21:26

And The thing is, is we haven't gotten all the costing numbers all worked out yet. So I mean, you'll need to compare, but one of the biggest asks was these start too slow. So we spin up some stuff in the background for you so it doesn't start so slow anymore. It should be, you know, as you probably already noticed, it's kinda nice.

21:44

Yeah, yeah, that's what I because I was running few of the sample notebook. So it it was running very far, it started running immediately. So I thought yeah, so it's like suspiciously fast, isn't it? You go what's wrong? It's working. Yeah, yeah, yeah. Okay. Thanks for thanks for that. So my yeah, this is Usher from Cat team. If I can add add to what Tammy said, right. So that that the concept that we are kind of referring to is called a starter tool, right. So just to kind of reiterate.

22:14

The cost of keeping those those pool live all the time for you to use when whenever you're coming and start executing your code will be born by Microsoft. Unlike competition where they charge you to keep it live all the time, right. So that's that's one thing Tammy has already clarified. So that's a starter pool and for now if you see the configuration of that starter pool is it goes all the way from one node to kind of 10 node, right. So based on your workload if you have a long running job.

22:44

It scales out, it scales out to up to 9 executor in majority of the cases. Based on our experience what we have seen is like 80 to 85% of the jobs are short lived job and they should be all good with running it with the starter pool, right. So that's that's that's that covers majority of the scenario. There are other certain scenario where customer need more control the way they have more control in in synapse.

23:12

With respect to the number of executors, say for example if customer is processing hundreds of terabytes of data and job is running for a couple of hours and then what they want more than 10 nodes, what comes with a starter pool right? So we have another concept called custom pool where you can define your own node type, a number of node, auto scaling, dynamic allocation and everything the the way you define with.

23:40

Synapse today, right? So the startup pool will cover majority of the scenario, but forest, so the other scenario, you have an ability to go and define a custom pool with your own control, with your own configuration. That's something will work the same way as as the way it's working in Synapse. So if you are running your job against custom pool, that's going to take 2 to 3 minutes to bring your session up because it's going to work in the same way.

24:09

However, this is only going to be for some of the jobs which take longer, say couple of hours. So with for for a job which takes couple of hours, 2 minutes, 3 minutes. Time would not vary and much or or matter and much. So that's the one thing. The third thing is say for example even for that type of job you want to keep those like underlying cluster up and running all the time so that you do not have to wait for.

24:39

Those those cluster to come the way we have to for for two to three minutes. So we are coming up with another concept that's the third type of pool you can say that's called light pool of custom pool. So basically you can define the light pool configuration of your custom pool and in that specific case customer will have to pay for keeping it light. So these are three different options. So most of the cases the startup will take care of your your use cases.

25:07

In other cases, you can go with the custom pool and in the third case you can have custom pool with light pool ability. So this these are the options customer will have. Yeah. Yeah. Got it. Yeah. Thanks for the for the detail. Yeah. And OK. My second question is like, so we have one leg concept here, right? So one leg is nothing, so one copy. So when I was doing it, yeah, let's say I'm writing some of the data. So all the everything is coming on the one leg.

25:36

So it's creating A1 copy of that one. It's similar like one drive. So what I can see here, let's say I have generated a Parker file or CSC file. So that Parker file or CSC file I can I can move to my C drive or any personal folder. So customer can ask me right how save this data. This is not is this is not a safe right? I I can I can take the customer data and put in my local folder. How? How much customer can ask how? What is the security here?

26:10

Sorry, what is that? Let's say let's say whenever we are creating a warehouse or any data data flow pipeline or something. So we have we are generating a pocket file or we are creating a CSV file or any any target location where we are in in the one lake where we can see the data warehouse and every all the folders are getting created. Let me share you my skin I can show you.

26:36

No, no I get your point. Whenever you create a file it is creating one link yeah yeah this so it's a one link. So if you see here so I have created this this all these files here I can move these files to my local drive right? I can move this my download or documents picture anywhere right. So how much save this customer can ask right this is not saved my data is you can anyone can move in the local drive anywhere.

27:02

You have to be a work member. Yeah. The user has to belong to the workspace though, right? Sorry. So that's the yeah. Short answer ananj. You get access to the data what you can see and work with regardless of any engines that you use. So this applies to even today. You can use any engine of your choice and you can go and get access to the data that you want.

27:32

And download it to your local computer, it's our same as that how different is it from that but that but that is not secure, right because anyone let's say and and if there is a development number of developers right they they can take that their data and they can receive their data right. OK. So that that will be handled. We are what we you maybe you're familiar with the concept of MIP labels in Power BI today. Yes.

28:01

Yeah, we will have a similar capability across the one leg. And again, these are not security controls. These are for you to understand the sensitivity of the data so that you are consciously sharing what

needs to be shared in which platform. But again, as I said, even today if you have access to a piece of information with the given compute, you can download it to your local machine.

28:30

It's no different than that. So is there any way where we can where we can restrict that the person cannot download to from the one link to the their local drive. Is there any something security we can we can set it there don't give them access to the one way Yeah yeah it's not possible and I don't see a reason why it is. It might be required but but hey this is.

29:00

The way we are going to handle this is by using probably labels and controls around that governance around that would be the approach to handle these kind of conversations. If others Sergei or somebody please come in and if you have other inputs, yeah, this is.

29:27

Understanding is with one one security you can define the security and based on the definition, say for example there are like 10 tables created and I give you access on five of them only. So you'll be able to see the data for the five tables only. So it because you have access on the data anyway, you'll be you'll be able to copy it right? So that's that's how it's going to work.

29:52

For rest of the other five tables for which you have not been given access, you'll not be able to see the files and the data when when the one security feature is available. OK, OK OK yeah so with related to that we have some Power BI admin portal settings or like you know export to excel, export to CSV. Are those settings applicable to?

30:15

Content of like one leg or deliver house or artifact as well or it just still limited to Power BI artifacts. Is it something? Let me put a screenshot here. Yeah, I I think those tenant settings are specific to power Gas today. I don't think you can export to Excel anywhere else. But but again, I mean a lot of these questions are related to security like we talked about.

30:50

We are in public preview and not everything is ready and we have a lot of people lined up in the queue here for questions. So I think we should move on. So let's take Brian, you're up next.

31:09

OK. There we go. All right. Yeah. So I've got a couple of questions. Just do we know or what is the state of the Dmv's in fabric in particular to the data warehouse? I've got a customer that collects telemetry now in in Synapse that tracks their queries and things. And I'm trying to figure out which direction we're tracking with Dmv's, if we're going more to the SQL Server route or if we're going more the Synapse dedicated pool route or what?

31:39

And there seems to just be there's 742 DMV showing now, of which like 500 or so return errors saying that they're not supported by this version. I didn't know if there's any way or if that was gonna get cleaned up, you know, closer to the GA Excellent question. Oh, go ahead. No, go ahead Tamuel. I was gonna say anything else. There are essentially 4.

32:06

That are going to be available, not as many as were before if you were used to the ones available in the dedicated pool. I bet Brian's in the back coming up with the link as fast as he can type because he's way better at that than I am and so not as many will be available because it's. There's less knobs to tune in the 1st place. More of the information will be available through the monitoring front ends and and.

32:35

Applications and reporting and less tedious detail, but we do have plans to be you know, you can put things in query story, you could put things in log analytics, you could download them yourself. There was like too many ways to do it. We are in the process of kind of doing a roundup to have it all load to one place so that you can write your own queries against it.

32:58

Yeah, So as long as we've got history, I mean I know that the you can go through the the portal or the studio or whatever you wanna call it and and look at it there. But we're using it programmatically in some cases to validate and and you know flag where things had started running longer and stuff. So we've just been collecting what flows through the Dmvs and Synapse because we those don't stay around very long.

33:25

In a highly active environment. So we've had. Yep. Land them into a table. Yes. The wonderful world of ring buffers. Yes. Yes. Yes. Yeah. So, yeah. There's. No. The link was just put in the chat. Thank you, Mark of right now there are sort of four, I think that that are supported. A lot of the others don't it, since it doesn't work the same way, you can't actually collect some of the same data you used to be able to. Right. So yeah. And that's what I've seen the sessions. Yeah. These I've already looked at and they don't.

33:53

We actually can't collect what the query was that ran. I mean, you can grab it while it's running, but you can't really get it until it gets into the history. But then it's rolled up across every time that query is run. So if I got one that's, you know, spiked up and the time, you know, took like I've got an export process that took 15 hours, it usually takes 9. And we're trying to figure out why that all of a sudden. This one export of basically the same size of data now is what happened. Yeah. Yeah, exactly.

34:23

Why is it six hours longer and costing us, you know, \$2000 more to export data? What would be great since you're doing this programmatically, if you if you want, I would love to sit down and talk to you about what it is you're collecting now, exactly what you're looking at while you're doing it so that we can line up what you guys need with what is getting built. OK, yeah, that's what we should be e-mail. Let's talk about it. OK.

34:47

Second question is CTE support. So I've got Ctas's that use CTE and everyone notice that they are not supported in CTAS. So is that something that's near term? Long term I gotta find another way to rewrite all this stuff. That one I don't know the timing on. Sounds like rewrite all this stuff. OK, yeah Mark, do you happen to know if Ctes are on the?

35:21

By GA roadmap or is that something you need to follow up on separately? Yeah, we'd have to follow up on that separately. I can't remember on that feature. OK, OK, fair enough. And then the last one, which I'm sure everybody's pitching to hear about, is the support for being able to shortcut Adls that are behind firewalls and in private Vnets. Is that near term, Long term, Because at this point my customer can't use Fabric.

35:50

You know in the state that it's in, sorry, what can you repeat the question. So if we want to shortcut, so if we want to shortcut into a data like storage that's already existing that everything that the customer's deployed is, is in private Vnets and behind firewalls, OK. So there's no concept of a public facing data set anywhere in the in the enterprise. Got, got it, got it, got it. OK, I'll answer that.

36:23

So again to reiterate, we are in public preview right now, so the asks would be to evaluate the functionality of shortcuts to your existing Adls. Yes, it might not be possible or even not recommended for them to open up all their storage accounts to their existing ADLS and start using Fabric at this point of time, but is that is it blocking them from reviewing the functionality of shortcuts?

36:53

In a separate ADLS storage that they can open up with anonymized the data if there is a blocker please reach out in our so the channel they've got these policies in place they can't even deploy it an ADLS storage that's open to the Internet. Fine, we please bring it to our teams and tag us with the customer names that this they have a policy that's blocking.

37:20

From previewing the functionality, we will take it. OK, Thank you. Thank you. Ryan, I have a follow up question for you. When you're asking about CE Test, obviously custom can use the factory, right? Did a factory extract the data out? Is that an option for the customer? And if not, why on the export process, you mean that? Yeah, was talking about. Yeah, yeah, we are using it. OK. I've got actually eight of them running in parallel to try to get this export speed up to something reasonable.

37:51

And running it, DW 15,000 seems to be the most cost optimized export process. But yeah, we just, we take over the entire service and run eight of them in, in parallel. Yeah. OK Yeah. Yeah. In Fabric or even within Gen. 2, you try to use them partitioning, either dynamic or partitioning. Yeah, we've got that. So yeah, we've got random.

38:18

Key in the in the data that we can split the table into the how many of our chunks we need basically. And the eight is the one that seems to be the most reliable. I went above 8 parallel processes then we started running out of writers or Jvms, started having problems. So OK, OK, Thanks Mark. Great. Thanks for answering that question all of you. So Next up is Phil.

38:48

Yeah, so I put a question in at the seven after the hour and Brad Sketch gave a pretty good answer here, but I just wanted to make sure we're all on the same page. So I'm having a hard time

understanding when to use a warehouse as my source for Power BI compared to a lake house as my source to Power BI. So.

39:07

If I could lay out the facts like normally if I'm just a Power BI developer, I went to my team and I said I need a star schema for Power BI. They'd make it SQL Server or Synapse. Now we're saying we're landing the data in one lake and it's going to be in parquet delta. And I could go down one path to say go make a lake house out of this. And I'm probably using my data scientists to do that. They're doing some Python jobs. Or I could say go ahead and make a warehouse and I'm probably going to someone with sequel skills to go ahead and make that.

39:36

I was wondering which one was better for Power BI to consume, the lake house side or the warehouse side since it is the same data and Brad's response said that by GA time it's most likely gonna be the same speed. So it sounds like the answer is depending on the skill set of my person creating the data side, if they're good at Python, go one route, if they're good at SQL, go another route, but I just wanted confirmation on that.

40:05

But you're not mentioning anything about Direct link here. It's not not an option. Is that an option here for you? Possibly that could be the be the third one. I thought Direct Lake was basically just like when I see my drop downs in like my Power Bi desktop today, I'm just seeing you know, I think the Lake house or the, yeah, the warehouse, but today if you wanna create a a Direct lake.

40:33

Connectivity mode, you can only do that in in the fabric in the the browser, but I think that will yeah. Eventually it will be either the the connectivity type that you prefer to use. But whether you should use one of the two options I I don't know. If someone else will comment on that, I don't. I don't know which one is best.

41:08

So, yeah, so I guess that you're and say sorry Brad, I was going to say Brad's rights. Yeah, I think the, the general, the general view from the RBI side should ultimately be that it doesn't matter if you're using Direct Lake over the Lake house or the warehouse then.

41:31

The data is being accessed the same way we're we've got the data in one lake, it's in Delta and Power BI is working on top of it. So it's really a matter of who's building it and which tool are they building it in like calcer warehouse. From there Power BI shouldn't really care any cost factor difference. One other one other extra thing to say about that directly doesn't work on warehouses yet but it will do which is I think what you're referring to about the by GA.

42:00

Right. Sorry, Correct. Yeah. Thank you for clarifying that, Chris. Sorry, I'll just go over something. Thank you. Who who was gonna ask about costs or something? Yes, I did ask, is there any cost difference between both by picking choosing which one to use should be from the Power BI points of view for everything else we will find out in August, I guess.

42:36

You okay. Thank you. Then the Next up here is Sachin. Yeah. Hi. I'm sorry, I my hand must have been up for quite some time, customer call, so I had to go on hold. Couple of questions I had, I don't know if it was covered already. So one question was?

43:02

Those customers that are, again, I'm talking about Microsoft federal customers, so GCC here. So those customers that are currently GCC but also Azure Gov, is there going to be a complexity related to that whenever Fabric finally comes into GCC? Of course it's not in GCC yet right now.

43:29

But when it comes, will there be an issue that a customer is GCC but Azure Gov as opposed to Azure commercial? Because that's that's really the usual combination that happens, GCC with Azure commercial and GCC hive with Azure Gov But there are a couple of customers we have that are GCC plus Azure Gov.

43:58

So I don't know timelines or what what deployments look like, but I'll just kind of say globally anytime your question includes Azure Gov or GCC, the answer is yes. There's going to be added complexity, there's extra, there's extra complexity there that goes into the way that we do identity and the way that environment segregation works between Azure Gov and Azure commercial.

44:27

And that's historically caused a lot of challenges. If you want to to shoot me an IM, maybe that's something we can chat about separately cuz I've gotta, I'll have to follow up with some folks to see what some of the, the actual deployment and integration's gonna look like between those environments. But I can follow up on that and get back to you. Yeah, yeah, I mean exactly. I mean that the, the.

44:51

Because the because of the identities, because these customers have two identities, one the GCC tenant for GCC and GCC high tenant for Azure Gov And of course once Fabric comes in to consume Fabric and Fabric, but you buy capacity in Azure and that's where I was worried about. So yeah, I'll definitely send you an IM and we can talk about that. And then the second question that I had was around sharing of data. So currently Azure.

45:22

Or rather Microsoft Purview provides data share but some customers are saying that we cannot use that if the the party that we are sharing with they don't have an Azure subscription because there is no way for them to consume the data, then will there be a better way to share data with one lake? Maybe Which will which will probably serve this use case?

45:59

There are vague plans for a data marketplace. Don't ask me any details about it, but there are vague plans for a data marketplace and I think they're going to start thinking about it in the next semester. OK. So data market, no idea whether that's going to apply to U.S. government stuff though. OK, but data marketplace in in fabric, you meant Chris.

46:28

Yes, OK. And Sachin, when you send me a message, we can chat about that as well. Just as Chris said,

there's stuff going on on that front that's getting started and I'm trying to collect some feedback from from some folks on that. So let's chat about that as well. OK. All right. Thanks a lot, Brad. I'm going to reach out to you. Thanks, Brian. Chris, Sasha, you're up next, everyone.

46:58

So this is an open question, so hopefully it's OK. It might be a bit of a change with gear, but just kind of curious to know from you guys if in this very short period of time you're starting to see any sort of emerging patterns with what customers are doing when it comes to trying out stuff in fabric. Yeah, that is a very open question.

47:29

I don't have any experiences in detail yet. I don't know if anyone else can chime in. I would say it's too early as well. People like playing around with it, but I I don't think it's we're we're way off getting best practices or anything apart from the fact that there's, you know lots of missing functionality and we don't know how much things are really gonna cost.

48:00

Yeah, they, you know, people are evaluating, but we have no idea what's gonna work better than. Yeah, which, which pattern is gonna work better than anything else. Yeah. What? And I was going to say Ohh, sorry, Lars, go ahead. Ohh. I was gonna say that that I hear from a lot of customers that what this is gonna cost. And yeah, that that's the, the, yeah, the only pattern that I've seen so far. But it's not related to. So any test practices of usage, many things.

48:28

Go ahead, Sammy. Yeah. I was just gonna say most of the customers I have, I have discussed with and talked with. Everyone is very much at the, I have a lot of questions stage, which I think is what you're experiencing too. As we go through this AMA, I have a lot of questions and I'm not sure if I can try it yet. A lot of them around some of the security that we know is on its way. So there's a lot of guidance supplying on. So how do we, how do we test it out if for instance all of our data is behind a firewall?

48:58

And and so a lot of help to get them started on a trial, even if it's the sort of sample data, just to let them push some buttons and and get a look at it and see what it will be like for them. And of course that generates more questions which means they're trying it. But we aren't really at, like Lars said, a best practices stage yet. A lot of people are just tipping their sticking their toe in, I think.

49:25

Something that I've been saying to my customers is apart from the fact, yeah, customers love to come to Microsoft and think we're omnipotent and we've worked out all the best practices in advance, which of course we haven't. Yeah, I always like to kind of turn it back over to them and say, well, it's a public preview. You're testing it. We would love to see how you get on. Please let us know what works and doesn't work for you so we can deal your ideas and share them with everybody else.

49:53

And, you know, if you find something that doesn't work well, you think you want to work, you should work better. Well, again, let us know. It's a public preview. And if you were saying something and hundreds of other customers are saying the same thing, we'll prioritize fixing it. Yeah. OK, We only have 8 minutes left and still your hands raised. So, Anson, you're up next.

50:24

So quick question is I have a I have a customer right now they they do not want their end users to run trials on, on, on, on Fabric. So they've disabled the the option in their production environment.

50:42

They're asking is there a way to do it across the tenant or is that is, is that one environment at a time, sorry, by environment you mean capacity. So the way that they showed, showed it to me yesterday was the fact that they were they went in and did a they they signed up for the trial.

51:10

But when they signed up for the trial, the trial showed up in their their their production environment and and and I it wasn't power *** but it was in their their production environment under the power platform admin admin center and it still want other users to have the option to try to to activate the trial.

51:38

Let's do this. Why don't you capture details of this scenario and post them on teams channel and we will tag PM that's responsible for trial experience. Perfect. I'll do a I'll do a video of it and just post it there. We're going to meet tomorrow. Yeah yeah, video screenshots or details scenario description will surface. This there are so you guys know right? There's an approach to.

52:07

Reduce friction in how customers can try out users and the tent can try out fabric. So we do collect this feedback and what I was going at is since you touched on the subject is everyone needs to be aware of the fact that if tenant admin disables fabric workloads is a global setting then an admin is not capable of disabling.

52:37

And removing capability from capacity admins to enable Fabric workloads that primarily applies to Psqs. Your case is different relates to FSQS and we will tag appropriate PMS and get an answer. But for everyone since we have this big audience, be aware of the fact that disabling trial kicking tires on Fabric capacities is not as easy as it might seem. Thank you.

53:08

Thanks. So great. And we move on to Brett. So just a question. Thanks for the session. A question about the default data set that gets created with the Lake House. So you can modify the contents of what goes into that default data set from the Lake House experience. And you can use the model tab, but if you browse to the default data set, you can't view the model.

53:34

Is there a reason for that? Or is that you don't intend for people to do that and you just want them to browse the table structures on the right hand side? And while you're conjuring that question a second one when you're running Spark workloads in a capacity?

54:09

You we you know the understanding I I it's a little bit hazy about how you will burst and then use burn down the rest of that you know the capacity for the 24 hour period but what happens if you run out of

capacity is in a. So let's say I'm running a spark job and then I've got a capacity dedicated just to doing that because it's a very big job.

54:37

And then it runs out of MOJO because it's now, you know, we've used all of the available capacity plus the flexing and all that kind of stuff. Is there an auto scaling kind of capability that's in the wings somewhere like full Power BI? I mean if auto scaling is there for Spark, yes, No, I'm talking about the, so let's say I, I, I buy F64, yeah.

55:14

And then I use, I Max out all the capacity and it's it it it tries to burst the workload and so on. But now I've used all of the all of the CPU that Microsoft has allocated to the capacity for the time period that we talked about. What happens? Does the job just say because there there's in the docs that says that things will wait in a queue if you're enabling that.

55:43

Because you'll either get an error when there's no more capacity left at a particular point in time, or you'll wait in a queue, but the capacity is a block of compute for that period of time. If you exhaust that, what is the what happens to the customer, right Currently in Power BI? In a situation like that, you'll get throttled, it won't just fail. So I think I saw something similar on one of the build videos, but not sure.

56:23

Okay. Thank you. Let's move on. If there's no immediate thoughts. Thank you. Thanks. So, Diana, you're up next. But what I wanted to bring up to the top, I posted this issue in the cat.

56:44

Team chat is I'm helping my customer trying to find the data source user activity and this is not the same thing as the person who configured the data source. We are able to get that via the API and with the PBI monitoring tool. We are using the Room Romano solution but the customer is using. Basically this is configured to go through the gateway and.

57:11

We are trying to find this specific activity. We don't find it through the Romano logs. Somebody posted to use the audit logs to get this activity, but I believe we're getting the audit logs through the Romano solutions. So does anyone know if this information can be found in the gateway logs? Because that's the direction that I'm going now with my customer. I want to know if I'm heading in the right direction or if this is something that we cannot get.

57:49

I put some screenshots in the in the the post that I use and I'm going to bring it up to putting it in the chat window here of this conversation to see if anybody can help me. Thank you so much. Maybe we don't have time to discuss it, but if you can post it in here that'd be great.

58:10

Excuse me, we're at the top of the hour now and I I can set it a bit longer. We have one more hand raise and then I think that's the last one for today to see if you can answer it. Darwin your, yeah, yeah I just said a question. So, so I've already on my own M365 tenant I've essentially enabled you know

Fabric and I'm going to be doing a training with a group of say 10 people and I wanted to on my M365 tenant I'll create like user one through user 10.

58:39

And so my question is this is, if I take all those users and I make them part of the trial, I've heard that the trial is like an F64, you get that? If I does each of my 10 users get their own F64 or or or should I just go build a big F SKU and put everybody in this in the same workspace for my workshop? So the question is because I I would like the 10 people that are using this to have a good experience and if I.

59:08

Enable the tenant for each of those 10 users. Is each of those 10 users getting some capacity each separate F64 or or should I go with a bigger capacity because it on the lower capacities like that F4 is F8. So I found them really unstable as far as as anyway anyway that's my question. So to to answer your question, there will be one trial capacity that you create and then if each you should have their own trial capacity they will need to create that on their own.

59:39

Or you will have to create them for them, right. So when I create, when I go to user 2 or whatever, I create a new workspace and I get my own capacity for me as user 2. Is that the case? Or or is or am I sharing it with the person that created the first one? Well, it's. Yeah, so. So each user, I think you can create up to 10 trial capacities in a tenant today. And so you could assign.

1:00:09

Each user, each user of their own workspace and their own trial capacity assigned to that workspace. Okay, That's. Thank you. That's all I need to know. Thanks very much. Thanks everyone for all your great questions. We are two minutes after the hour so I will stop the recording. We will be back next week with a presentation.