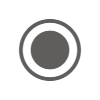
**AnnexCloud-Kanini Connect-20240313\_183204-Meeting Recording**

March 13, 2024, 1:05PM

40m 9s

 **Speaker 1** 1:15  
So I think it's going on outside, OK.

 **Prashantkumar sharma** started transcription

 **Speaker 1** 1:19  
So if you see this one, this is the graph.  
This is the comparison from 27 March.  
Right.  
You see this via have taken Tuesdays because we made changes on March 7th.

 **Tuhin** 1:27  
Yep.

 **Speaker 1** 1:30  
So I wanted to show you exact comparison day-to-day because as I said earlier, right, there's a bell curve.

 **Tuhin** 1:30  
Yep.

 **Speaker 1** 1:35  
So the cost was 4025 the last week, Feb 27 to March 4th and this after this March 5th to March 12th.

 **Tuhin** 1:36  
Yeah.  
Yep.

 **Speaker 1** 1:44  
It went down to 2955.  
And this is 2 and just for putting the legend just I know it's smaller.  
I'll change this screenshot for you, but before I share it with you, so if you see this blue one that you see is the resource group that we create for the data bricks, right for users to go in and you know use workspace or run a job.

 **Tuhin** 1:57  
Umm.

 **Speaker 1** 2:06  
So it will vary based on how much usage we have correct and the loop and that you see this is the default created by databricks.

 **Tuhin** 2:09  
Yep.

 **Speaker 1** 2:14  
The VMS that Karthik has been talking about the cores that was being used right?  
This blue one represents those directly used by, so this blue one is based on DU, right?

 **Tuhin** 2:24  
OK.

 **Speaker 1** 2:25  
And this dark blue one is the one where data breaks.  
You know, use the storage or the create the VM machine.  
And I have put a detailed this thing me and uh you know Rupali with them.  
So this is the previous week, just to give you an overview, there's a comparison Data Azure data bricks.  
If you see the resource group that is given that charges you as your data bricks, that was 2574, right in that week and this is DB use.

 **Tuhin** 2:50  
Umm.

 **Speaker 1** 2:52  
This is party view that you are charging does not show on Azure.  
What kind?  
What you're doing, but this is based on how much compute?  
How much storage?  
Everything you're using right and this all is also in this group.

 **Tuhin** 3:01  
Yep.

 **Speaker 1** 3:04  
I'm assuming DB and VPN gateway.  
This is something on the design side.  
We did so that is just I've put because it was a major amount, right storage on the that one is being it is using at $8.00 directly on Azure on databricks peoc this is default DBFS stuff is being used.  
The VM machines that were being created by every job, every cluster coming up.  
This was the cost right?  
864 and 269 I put a total for you the last week.

 **Tuhin** 3:29  
Yep.

 **Speaker 1** 3:31  
It was 4024 and you can look at the day-to-day costings also for detail if you want to go right.

 **Tuhin** 3:33  
Yep.

 **Speaker 1** 3:37  
If you want me to go, I can go over them.

 **Tuhin** 3:37  
Umm, no, no, no, that's OK.

 **Speaker 1** 3:39  
Otherwise, this is for your information, right?

 **Tuhin** 3:40  
Let's yeah, yes, yeah.

 **Speaker 1** 3:42  
You can look at that and see and this is what Carl you're requested and this we were making.  
This is what were the changes made exactly before and after right there we made and this is just configurations right?

 **Carl Droste** 3:53  
You know, OK.

 **Speaker 1** 3:55  
So we make sure so we reduce the number of workers, the minimum number of workers also, and maximum number on because it must be OCP.  
And we looked at the metrics, right?  
And if you see the cores we maximum core, then we change the kind of instances also that they were using and same for you know and this was so the three that you see these were running on the job cluster and this is the job cluster changes on which the job each job can be configured right.  
So we configured on what we saw the maximum usage for for last month and this was the POC cluster the normal clusters.

 **Tuhin** 4:24  
OK.

 **Speaker 1** 4:29  
So we changed them just for like me Rupali we are using a work instance and we are coding something that is being used.  
So we reduced the configuration alongside with the Vemula and with uh Karthik and this was something doesn't the major change right?  
One more change we made is on as we have changed only schedules and higher environment and lower environment.  
Also, we saw that there was no sequence of and everything was running continuously for UAT.  
So we changed and Tarang and thought of agreed that this is actually what should be doing and we put it again on schedule.  
So these are the only changes that we did, very minor ones.  
If you see, this is the after change past sheet for March 5th to March 11th because we wanted to give you exact number because there's a forecast changes and all those things the cost came down to 2954.

 **Tuhin** 5:09  
OK.  
Yep.

 **Speaker 1** 5:19  
If you see the VM machine, it's come down half.  
This has come down through right?  
It's not as much slower as the I think we can go if we use ADLS has your data bricks.

 **Carl Droste** 5:26  
Yeah.

 **Speaker 1** 5:30  
Also, it's not much 800, but I was expecting not much on this, just for the course because users are there right?  
We are doing development, there's UAT, running and then you have, you know, users running to report using that directly.  
Thus we can evaluate 3, so 294 and this you'll see the changes, right?

 **Tuhin** 5:45  
Yep.

 **Speaker 1** 5:48  
So this is the cost different sheet from the first one minus the second one.

 **Tuhin** 5:49  
Yep.

 **Speaker 1** 5:54  
Thursday, we made the changes.  
After that, the cost started going on 163 till yesterday.  
The difference?  
This is the savings that you had 304.  
So on an average you can see 242-3300.  
That is, this is in last 12345 days.  
This is the money.

 **Tuhin** 6:11  
OK.

 **Speaker 1** 6:13  
That was it.  
By these small changes.

 **Tuhin** 6:14  
But.  
OK.

 **Speaker 1** 6:19  
OK, I'm not.  
Not.  
Not much, but I'm hoping that is what you were expecting.

 **Tuhin** 6:21  
So no, no, no.  
So OK.

 **Carl Droste** 6:23  
Uh.

 **Tuhin** 6:23  
So no, this is good.  
This is?  
I asked for 100, so we got to 16, so that's that's a win.  
But so, uh, when I present or talk to internal people, Priyanka, can you guide me in terms of?

 **Speaker 1** 6:35  
Umm.  
Hmm. Hmm.

 **Tuhin** 6:41  
I know it depends on a lot of things, but what's your best estimation in terms of?

 **Speaker 1** 6:43  
Umm.

 **Tuhin** 6:46  
Is it a linear thing that we're gonna see between POC and production is a could it be nonlinear, you know?

 **Speaker 1** 6:52  
Hey.

 **Tuhin** 6:56  
Either way, right?  
Maybe it's not gonna be like, you know what?

 **Speaker 1** 6:58  
So so I'm thinking based on because we have 3 environments, three of them have different loans, so every one of them right, the usage is different everywhere and the data is also different.

 **Tuhin** 7:00  
What are you thinking right?  
Yeah.

 **Speaker 1** 7:11  
So right now I cannot commit on it, but there will be change.  
Definitely it will come down.

 **Tuhin** 7:16  
Yeah.

 **Speaker 1** 7:16  
I don't think it will be linear for all this thing, but but there will be a similar kind of change you'll see and that is one thing why I wanted to show you this one.

 **Tuhin** 7:19  
OK.

 **Speaker 1** 7:25  
If you see the prod that you know we just reschedule.  
I did not put that.  
I'll put it for you.  
You'll find it for that USA.

 **Tuhin** 7:32  
Yep.

 **Speaker 1** 7:33  
If you see the prod database right, US E the costings 5 starting one that was there right?  
And then Thursday, we changed it.

 **Tuhin** 7:38  
Umm.

 **Speaker 1** 7:39  
It went down to 498 on an average, then it becomes standard on the weekends when you are not using it.  
Went down to 268, if you if you want me to compare it was double the cost on weekends.  
So there's a if you see there was difference on Saturdays and Sundays earlier, now they have become standard, right?

 **Tuhin** 7:51  
OK.

 **Speaker 1** 7:56  
Like they are similar Saturday, Sunday when the workload is down and this is because your jobs are running there you the API you can see Abby, the cost of gone down drastically, right?

 **Tuhin** 7:57  
Umm.

 **Speaker 1** 8:07  
I know it's not comparison.  
It was one 43185, so this is just schedulings right of the jobs that we did and your job failure stopped.

 **Tuhin** 8:14  
OK.

 **Speaker 1** 8:16  
And that is what VM machines, right?  
So you will see this kind of so production if you want to show I can put that the 143, one 85168 on an average above 140 right even on a weekend it was 108.

 **Tuhin** 8:29  
OK.

 **Speaker 1** 8:31  
On this weekend, it was, yeah, $50.

 **Tuhin** 8:35  
Yep.

 **Speaker 1** 8:36  
Both the nothing more, nothing less, right.  
And on weekdays, 5150, so that is 140 to 50, but as.

 **Tuhin** 8:39  
OK.  
OK.  
Makes sense.  
Yeah, one question right.  
Uh, with the job schedule failing, right, as we are attributing some of that to the cost savings, is it fair to assume on the job fails there is like even if it takes one second to fail over the job, right, that were failing, is there, what what do we incurring like a minimum cost because sometimes DBU or whatever, right?

 **Speaker 1** 8:48  
Umm.  
Hey.  
Yes, umm.

 **Tuhin** 9:07  
It doesn't.

 **Speaker 1** 9:07  
Yes.

 **Tuhin** 9:08  
It doesn't charge you for it.  
One second it there is a minimum threshold.  
Maybe it's one minute.

 **Speaker 1** 9:12  
So on.

 **Tuhin** 9:13  
I don't know. Yeah.

 **Speaker 1** 9:14  
Yeah.  
So that's why if you notice Tuhin you were not because the same kind of job failures were happening on POC also, right?

 **Tuhin** 9:19  
Umm.

 **Speaker 1** 9:21  
If you notice the cost for POC the the which you are charged by dbus right the that resource group and I'm showing them separately because of this reason only because I know you wanted that clarification right?

 **Tuhin** 9:22  
Yeah.  
Umm Yep.  
Yeah.

 **Speaker 1** 9:32  
How how we have to do that so that device went down right, but they did not go so much down.

 **Tuhin** 9:36  
OK, sure. Yep.

 **Speaker 1** 9:37  
They went from 2800 to 2000, right, but the data breaks one that was running right.  
We did not do anything except code changes and scheduling, right?

 **Tuhin** 9:44  
Umm.

 **Speaker 1** 9:47  
And it the if you see everywhere, that is what made the change right in production, we did not do anything apart from scheduling because you see these virtual machines, they are divided.

 **Tuhin** 9:47  
OK.

 **Speaker 1** 9:58  
This is what is coming up for databricks.  
For example, here uh reason, right?

 **Tuhin** 10:02  
Umm.

 **Speaker 1** 10:02  
So even if your job fails, these VMS are brought up and they go down after certain time only.

 **Tuhin** 10:05  
Umm.

 **Speaker 1** 10:08  
It's first the job will the machine will be created right on the Azure side?  
You're paying for that machine because it's a virtual machine, so it might be just.  
There is a differences in costs.  
You can see it might be only $1.00 per machine coming up half a dollar you were paying, but that accounted $200.00 because it was failing and it was running every minute 24 into 60 right for the whole day.

 **Tuhin** 10:29  
But.

 **Speaker 1** 10:32  
That many VM machines were coming up costing you .36 point 4-5 and then going down.  
Only the ones that was continuously running was costing you mood, but there was minute costs, but that was adding to your cost.  
That is, I can show you exactly if you want and that will help you take it forward.

 **Tuhin** 10:43  
OK.  
Yeah.

 **Speaker 1** 10:48  
This sheet is, I think the best I can, so I'll divide it by the service, right?  
So, but the VM machines out of mostly for data breaks and why I restricted it to to those two because I know Karthik and team is using, they are using the VM machines for different ones they have restricted.

 **Tuhin** 10:54  
Umm.

 **Speaker 1** 11:07  
They know the costing, but here virtual machines was last time.  
This is for December.  
No, I want to just go for the last seven days.  
I'll just go back there and then March that'll be.

 **Tuhin** 11:20  
Umm.

 **Speaker 1** 11:22  
So 16,004 Feb, right?  
Just virtual machines and these virtual machines.

 **Tuhin** 11:25  
Umm.

 **Speaker 1** 11:30  
If you see all three sorry, prod Annex, cloud data bricks prod, this is all databricks that we created right POC and this so these are VM machine.

 **Tuhin** 11:37  
OK.

 **Speaker 1** 11:39  
Ohh sorry, I actually my mistake.  
I actually did the service name.  
I put this thing I wanted to put on a VM machine, so this is what one thing I was showing earlier, right?  
That if I put a resource group that is being created by data bricks.

 **Tuhin** 11:51  
Umm.

 **Speaker 1** 12:00  
Yeah.  
So this one do these are everywhere right of your machine cost?

 **Tuhin** 12:03  
Umm.

 **Speaker 1** 12:05  
I'll show you one production that will be better actually rather than so.  
And I'll show you the custom date only so that we can you can take it in the same way, right when you're taking forward.  
So the previous week this you see the that group only has virtual machine 616 and 157, right?  
It's not doing anything, it's just running the jobs.

 **Tuhin** 12:27  
Umm.

 **Speaker 1** 12:29  
It's it's in your AP region, right?

 **Tuhin** 12:30  
Yeah.

 **Speaker 1** 12:32  
If I keep everything same and all the only thing in production we did is rescheduling right?

 **Tuhin** 12:32  
Umm.  
Umm.

 **Speaker 1** 12:38  
So let's change it to I'll I'll put it seventh.  
Well, 5th to 11th for now, right?

 **Tuhin** 12:45  
Umm.

 **Speaker 1** 12:47  
And I'll show you 2 two days as well because I still warehouse went down.  
This is $100 cost.  
That was on this small 1, right?

 **Tuhin** 12:55  
OK.

 **Speaker 1** 12:56  
So the $100 cost difference right on storage also because it's creating copies?  
That is the difference you get.

 **Tuhin** 13:03  
Makes sense?  
OK.

 **Speaker 1** 13:06  
So that will be changing, right?  
It'll make a change and this we haven't even changed the course of this.

 **Tuhin** 13:08  
Umm.

 **Speaker 1** 13:11  
It's still has the CPUs the same, so I'm hoping we can keep it down.

 **Tuhin** 13:11  
Yeah.  
Yep.

 **Speaker 1** 13:16  
I hope that answers your question.

 **Tuhin** 13:17  
Yeah.  
Yep, makes sense.  
Makes sense.  
Yeah, I thought, you know.

 **Speaker 1** 13:19  
Umm.

 **Tuhin** 13:21  
Yeah.  
You know, there's, you know.

 **Speaker 1** 13:22  
Yeah, I cannot say linear or nonlinear because there are other factors there, right?

 **Tuhin** 13:23  
Yeah.  
Like he's actually.  
I I knew it.

 **Speaker 1** 13:27  
We have no.

 **Tuhin** 13:27  
I knew it, but I just want to double check because sometimes you know, I don't know from.

 **Speaker 1** 13:30  
Yes.

 **Tuhin** 13:31  
OK.  
OK, that's good.  
I have two other points that I wanna talk about and I do have a hard stop so want to be cognizant of other topic that anyone might have.

 **Speaker 1** 13:37  
Hmm.  
Umm.

 **Tuhin** 13:43  
I do wanna talk about the change request and the approach with data anonymization and things of that nature.

 **Speaker 1** 13:49  
OK.

 **Tuhin** 13:50  
That's one of the topic that I want to talk through.

 **Speaker 1** 13:52  
Hmm.

 **Tuhin** 13:55  
Then I just wanted to quickly talk through next steps.  
So there's so unless someone else had or you had some other items to cover today.

 **Speaker 1** 14:07  
Yeah.  
I think for my side it's the same topics I wanted to bring up.

 **Tuhin** 14:10  
OK, Kathy.  
Carl, is that?  
Is that OK with you guys too?  
Do you guys have anything Vishal?  
OK, so Priyanka, you know, so you must have seen the email exchange between me and Anand, right?

 **Speaker 1** 14:28  
Yes.

 **Tuhin** 14:28  
Umm.  
Umm, so what's my assumption?  
Kind of correct that the data anonymization, sanitization of PII and things of that nature that was essentially the motivation behind it was to replicate the production data in POC to run a for like a better word apples to apples comparison was that the thought process or or and OK.

 **Speaker 1** 14:56  
Yes, that is the thought process that we are going to have the symbol of volume similar kind of data in the lower environment.

 **Tuhin** 15:00  
And.

 **Speaker 1** 15:04  
So that the testing before going and we can you know this will be like we learn the performance and we then scrape the data and we can again get it right because there be a code that you just run script get scrubbed the data and get it in the lower environment.  
But the the cost and performance issue, three things that you're facing that are not because there is no testing in the lower environment for cost.

 **Tuhin** 15:17  
Mm-hmm.

 **Speaker 1** 15:25  
First of all, approximate at least right the second is the performance and 3rd is data quality.

 **Tuhin** 15:25  
Hmm.  
Umm.  
Umm.

 **Speaker 1** 15:32  
Also, like right now we are not.  
That's why I cannot give you a comment on POC and you know, compare the POC environment to production environment.

 **Tuhin** 15:39  
Yeah.

 **Speaker 1** 15:39  
If I had that I could give you better view right?  
Even if I run them slowly, I can give you a better view that OK for 10 minutes it cost me this much.  
I can stop that and I love that.

 **Tuhin** 15:48  
Yeah.

 **Speaker 1** 15:49  
OK, this will go for this much at least one time.  
Run.

 **Tuhin** 15:51  
Yeah.

 **Speaker 1** 15:51  
I can cost it right?  
Same for data quality.  
Right now, the team is not able.  
We understood where the issues might be in the code like where all where the deletes for example where the duplicates are coming because they're not putting the records are duplicating because the update it's taking with the update date, the time stamp and these kind of things are, but we cannot exactly replicate in the lower environment because the data is not similar kind data quality issues.

 **Tuhin** 15:58  
Umm.  
Mm-hmm.

 **Speaker 1** 16:15  
The main concern that otherwise we can always have a faker data out applicate data, but that is not going to resolve the data quality, cost and performance that can be done right without taking production.

 **Tuhin** 16:16  
OK.

 **Speaker 1** 16:25  
But we'll still have to put a volume of data there right to test.

 **Tuhin** 16:28  
OK.  
OK. OK.  
So two other questions, right?  
So the first thing in POC right and in terms of the anonymization and the PII things. Right?  
So the first question is, uh, from a time point of view, because the reason I mentioned this Priyanka and Karthik can correct me if I'm wrong right?

 **Speaker 1** 16:53  
Who?

 **Tuhin** 16:55  
If you ask me and you, you ask me, you can ask, partake, or you ask Tarang, or you ask anyone in development.  
Right.  
You won't get it like you know, there is like a because of a.  
But that you say it lack of documentation or data modeling or whatever that exists right now, right?

 **Speaker 1** 17:13  
Umm.

 **Tuhin** 17:14  
I can't tell you that this table or this column has Pi right, and it's going to be very difficult even to do that discovery.  
So keeping that constraint in mind like how are you planning on scrubbing PII?

 **Speaker 1** 17:23  
Umm.

 **Tuhin** 17:30  
Or, like anonymizing the data, right?  
No, or was was.  
What are you gonna use the tool?  
Did we have to write some code and what I'm getting towards not Priyanka is that how long is that exercise gonna take?

 **Speaker 1** 17:43  
So that is actually the.  
This is the one discussion that we Prashant Rupali were having in the day before.

 **Tuhin** 17:50  
Yeah.

 **Speaker 1** 17:50  
Like in the earlier in the day like because I had that call with Karthik yesterday, we all met and I understood what you're saying exactly.

 **Tuhin** 17:52  
Yeah.

 **Speaker 1** 17:58  
That was the impression I got.  
He conveyed it very well and I understood that my that estimation part right.  
The third part that is there, the scrubbing we have done it so many times.

 **Tuhin** 18:04  
Mm-hmm.  
Yep.

 **Speaker 1** 18:07  
I can say in Prashant keep me honest here 3213 days to one week we'll be done with that coding on Lambda will put a function for you.

 **Tuhin** 18:14  
Umm.

 **Speaker 1** 18:15  
It's a scrub down the major part and I think the first assumption we took when putting that is that OK, the identification and everything that was already done right, because this is something an SME, we need a live, we can do it with you, but it'll take more time, right?

 **Tuhin** 18:23  
Umm yeah.

 **Speaker 1** 18:30  
And we would need an SME to help and Prashant, this is something I think we have done that that is not, but it has been our time consuming exist.

 **Prashantkumar sharma** 18:36  
They didn't. Yep.

 **Tuhin** 18:37  
No, no, no, I'm not.  
I'm not.  
I'm not.  
I'm not.  
I'm not questioning the sorry.  
Yeah, this is gonna sound bad.

 **Speaker 1** 18:41  
To.

 **Tuhin** 18:43  
I don't mean it that way, right?

 **Speaker 1** 18:44  
Uh.

 **Tuhin** 18:45  
I'm not.  
I'm not questioning capabilities that whether you can do it right, I'm I'm questioning whether uh.

 **Speaker 1** 18:50  
Hmm hmm.

 **Tuhin** 18:57  
Whether we have the time or like, you know, and to do this and the second question is frankly, if we don't do that anonymization and we say like, OK, we're gonna do that dumping POC, right?

 **Speaker 1** 19:00  
Hmm.  
The.  
Hmm.

 **Tuhin** 19:08  
I So what are some of the challenges that you guys foresee apart from any regulatory things that we might add here need to adhere to from an annex cloud point of view?

 **Speaker 1** 19:20  
So the the main problem is.  
See, we have no issues, but it's a regulatory thing.  
And this is in US specially PII laws everywhere, right?

 **Tuhin** 19:25  
OK.

 **Speaker 1** 19:29  
So your company registered in California, right?

 **Tuhin** 19:34  
No, it's in Virginia, Virginia.

 **Speaker 1** 19:34  
Where is it, Virginia.  
OK.  
So, Virginia, all the privacy laws are there.  
California's more stricter than Virginia, and one thing that I saw in is like you have money constraints, right?

 **Tuhin** 19:42  
Mm-hmm.

 **Speaker 1** 19:48  
You have loyalty points.  
What is the cost and all those things?

 **Tuhin** 19:51  
Umm.

 **Speaker 1** 19:51  
That is one concern that is coming down.  
So we were thinking like we were discussing about it, that we have to shuffle that so that it cannot.  
1st is PII data you have to do that that is a cross right?  
You can put it if it doesn't get caught in this.  
OK.  
But it is a responsibility that Annex Cloud will have to take.

 **Tuhin** 20:07  
Yeah, I mean, I mean, I mean I, I I understand that you know from a compliance point of view, right, you you you need to you need to be you know you need to be comfortable right and you know and again right this is not my decision but I'm just trying to understand my option.

 **Speaker 1** 20:11  
Umm.  
That is the big one.  
Umm.

 **Tuhin** 20:25  
Priyanka, frankly, that because Matt essentially asked me, like you know, essentially like you know, do we need to talk out to our in-house console to figure out if that's if that's if that's we are OK with not anonymizing.

 **Speaker 1** 20:26  
Hmm hmm.  
Hmm.  
So.  
Umm.

 **Tuhin** 20:41  
But I I I understand that you guys might your your legal counsel in-house might might want something from us right even if we said that was true, right?

 **Speaker 1** 20:49  
Yeah. Yes.

 **Tuhin** 20:51  
So I I I I I I I understand all of that right and I'm not giving it decision.  
I'm just.  
I'm just talking through my options yet because essentially that trade off right now, if I if I put it on a very high level that I can present right, is that bring the Pi and things of that nature we think it's gonna take anytime between 3 to 7, three to three days to a week, right.

 **Speaker 1** 20:56  
Umm.  
Umm.  
Umm.

 **Tuhin** 21:16  
And these are the things, yeah.

 **Speaker 1** 21:19  
Yes, but the first step that is necessary for that, if we don't know which columns to end on device, it's apps to use.

 **Tuhin** 21:22  
Yeah.  
So you need any of that?  
Yeah, you need.  
Yes, I yeah, I understand that.

 **Speaker 1** 21:27  
You understand that.

 **Tuhin** 21:29  
That's why that's why the first thing I said that right, you know, so the end result might end up being the same.  
But you know, if we get if we get any regulatory concern, we can we can we can show intent and it doesn't become gross negligence.  
So I I I get that right.  
So from that point of view, I think I think that is that is good.  
OK.  
Anyhow, let me let me talk to Matt and I will get back to you on the change request.  
OK, I understand. Yeah.

 **Speaker 1** 21:58  
So Prashant, we want to add to anything to this.

 **Prashantkumar sharma** 22:01  
I guess you're covered.  
What we can do, Priyanka is like pretty much because we we we have done all those three steps multiple times what we can go and provide to Tuhin and because now you Rupali have very good idea of the data as well, right we'll give you obviously not the accurate but but some tentative timeline OK this is the time it is gonna go ahead and take and obviously how much help we need in in terms of classification right let's OK it's a seeing the data it's might be like one hour of effort from one of the SMEs.

 **Tuhin** 22:09  
Umm.  
Umm.

 **Prashantkumar sharma** 22:30  
Or two hours or effort on half a day of effort.  
So we will provide that data and then we can decide like let's as long as like if you're OK doing like with in terms of law and all that we can go ahead and use the data as it is nothing like that.

 **Tuhin** 22:33  
So.  
Yep.

 **Prashantkumar sharma** 22:41  
So we'll provide you both the options maybe today and then we can you can decide.

 **Tuhin** 22:42  
OK.  
Umm.  
Yeah, absolutely.  
And you know, it's A and again, right, the decisions could not gonna be up to me.  
I think we'll have to run it through our in-house console to get their sign off too, but you know, it's essentially it's.  
It's more than the money part, right?  
More than the effort part, I think they it's the time constraint more so, right, you know, because I don't know.

 **Prashantkumar sharma** 23:04  
Yep, Yep.

 **Tuhin** 23:06  
I don't know if Matt so linked to spend week plus one 1 1/2 weeks to two weeks to do this just from a timing point of view.

 **Prashantkumar sharma** 23:12  
Umm.

 **Tuhin** 23:14  
So if not, then you know, let's talk through like, you know, what are the illegal steps?  
So you'd have to make and if there is an exception, right not my call.  
I just have to present the options, yeah.

 **Prashantkumar sharma** 23:24  
Sure.  
Makes sense.  
Yeah, we can do that.

 **Tuhin** 23:28  
OK, sounds good.  
Priyanka.  
Then what are we focusing on rest of the week then?

 **Speaker 1** 23:33  
So one thing is we are looking at storage costs.  
That is a smaller part, right?  
Other thing is the team is testing out two approaches for concurrency.

 **Tuhin** 23:38  
OK.

 **Speaker 1** 23:42  
We have an approach, but it's like we cannot test it out if we don't have all the like.  
We have not developed right how much it will definitely we know how to do that and what has to be done.  
We'll keep it like half a page.  
We can give you the approach right?

 **Tuhin** 23:55  
Yep.

 **Speaker 1** 23:55  
That is already just have to be documented.  
Second thing is that we are looking at the approach if we can do what you said right, like having a delta in the bronze layer.  
Although I'm.

 **Tuhin** 24:06  
Umm.

 **Speaker 1** 24:06  
I'm I wanted to check how because now I I and me and you know Rupali we have the access to production we wanted to see how much data get an approximate on the Azure price calculator, how much it will cost to you know dump the data like making bronze layer actually the raw layer and then maybe archiving everything else or deleting everything else that is coming having the current seven days or one day you know runs so that you always have a copy is it so the only comparison I'm trying to do now is that is.

 **Tuhin** 24:19  
Yeah.

 **Speaker 1** 24:36  
It better for you cost and performance wise that we get all the data from AWS RDS into the bronze layer and then deal with it in this layer.  
Or is it better that you actually, you know, do all those things and then get it in the bronze layer?  
Then the purpose of silver layer.  
Actually, we'll have to discuss this is one, you know, brainstorming we were having and we were talking about it.

 **Tuhin** 24:54  
Yeah.

 **Speaker 1** 24:58  
Whether silver layer is useful or not, if we do the delta there itself, do you have any other data quality concerns?

 **Tuhin** 25:01  
OK.

 **Speaker 1** 25:03  
Because there are other things that obviously I see that you have mentioned about data 30 missing data and you know accuracy of data, those things that will be built in silver layer.  
But then I'm feeling like that is time consuming because you have to read through the entire thing or you have to make changes to your RDS for, you know creating new table for each table where the deleted data go.

 **Tuhin** 25:24  
Can you?

 **Speaker 1** 25:24  
And after we pick it up, then it's truncated and it's then it comes.  
So I'm just walking.  
Talking you and walking you through the approach you can see, right?

 **Tuhin** 25:31  
Yeah.

 **Speaker 1** 25:34  
That is a lot of scanning to the RDS from Azure to AWS, but I can give you exact numbers when I see and I think this is what I told Karthik.

 **Tuhin** 25:38  
OK.

 **Speaker 1** 25:43  
I'll come back to him with that approach and see what is the costing right now for your fetch of that data and based on that we can see how much data is coming and if we take the total data, will it make a lot of difference?  
Will it be like it's quicker for performance wise that kind of thing?  
Maybe that is something we can run on production environment.

 **Tuhin** 25:58  
OK.

 **Speaker 1** 26:00  
Maybe one of our team can test it out.  
The entire bulk one time and delete it from there.

 **Tuhin** 26:04  
Umm.

 **Speaker 1** 26:06  
That can give you a better idea.  
What do you think?  
Like you were cluster.

 **Tuhin** 26:10  
Yep, sounds good.

 **Speaker 1** 26:10  
That is right that so that it's not in the lower environment and it's a higher environment we get the bulk and we'll have the exact you know comparison of the current one.  
It's not going to change much in the cost and performance wise if we do the other approach and how much difference we are doing when we bulk dump you, take a snapshot and put it just there and then deal with it.

 **Tuhin** 26:30  
OK.

 **Speaker 1** 26:30  
Everything in the silver layer, and so I was actually call that out that we are going to do that, but because we were concentrating on keeping the cost down and we're looking at duplicating, we know certain points where what are the issues, we'll document them at least what we have known till now.

 **Tuhin** 26:35  
Good enough, yeah.  
Umm.

 **Speaker 1** 26:47  
But as I said, I'll disclaimer is that we have not been able to replicate it properly.  
This is we are just doing sample data and what we found out in the code, right?

 **Tuhin** 26:52  
Umm.

 **Speaker 1** 26:56  
So these two things and the comparison of the approach will spill over to next week also, right, because it's already Wednesday in uh, Friday afternoon.

 **Tuhin** 26:56  
Yeah.

 **Speaker 1** 27:05  
I'll give you something at least, but exact things I'll be putting them across in next week.  
Is that alright?

 **Tuhin** 27:13  
Yeah.  
Yeah, that's fine.  
Yeah, that sounds good, but I just want to nothing as a call out that doesn't make sense.  
So appreciate that transparency. Yep.

 **Speaker 1** 27:22  
Sure.

 **Tuhin** 27:23  
Yep.  
OK, sounds good.  
That's all from me.  
I gotta run.  
But thank you everyone, unless there is any question for me, yeah.

 **Speaker 1** 27:31  
But that's all call you have anything.

 **Tuhin** 27:32  
OK.

 **Carl Droste** 27:34  
No, no, not specifically.  
I guess is there any other challenges that you need help with?  
Uh getting traction on things.  
I know, I see that the the the space is helping a little bit go through that.

 **Natarajan Sundararajan** 27:48  
Yeah.

 **Carl Droste** 27:50  
Uh, last night.  
So drink, I guess there's any other challenges that you guys need, let me know and let me see what I can do to help streamline that on this side.

 **Speaker 1** 27:52  
Umm.

 **Tuhin** 27:54  
Yeah.

 **Speaker 1** 28:02  
I think we are good.  
Otherwise, as I said, I was against putting any of my developers developing in the production environment.  
But to give you those numbers, we might have to run that, you know, experiment in the production, because we I I'm just clarifying you we don't want that production data to come through into the lower environment and sit in the delta, right.

 **Carl Droste** 28:12  
Umm.

 **Speaker 1** 28:22  
So we can have one of the team members run that on the production and then work the access.

 **Carl Droste** 28:22  
Right.

 **Tuhin** 28:23  
Yeah.

 **Speaker 1** 28:26  
But we have everything set up and that is one big reason to him.  
If you see you see this is the reason we I think that that exercise all of we are OK with production data if you decide whether we can run with it right.  
But that is long term.  
It will help to have that set up.

 **Tuhin** 28:41  
Yep, yeah.

 **Carl Droste** 28:43  
Yes, Doug, document everything that we're doing as we do it.  
No, yes.

 **Speaker 1** 28:48  
Yes, yes.

 **Tuhin** 28:49  
Make sense?

 **Speaker 1** 28:49  
And it's coming of the data at least then and call this is 1 suggestion from my side I'll give maybe for now you might go with the production approach because you have to resolve it, but that exercise should also start because this is on your business side as well, right.  
But you will have to do those things and identifying cataloging even on the side.

 **Tuhin** 29:04  
Umm.  
Yeah.

 **Speaker 1** 29:08  
Right.  
So that will be a good practice to have for later on so that these kind of issues don't occur again.

 **Carl Droste** 29:14  
Umm.

 **Speaker 1** 29:14  
Then it'll be just developing some features rather than dealing with these issues in the upper environment, we can do everything in the lower one.  
And no, I'm over there was, you know, stating this.  
But just like serious that that has always worked out better rather than, you know, long term, not planning for that.

 **Carl Droste** 29:29  
Alright.  
When you put it into a process that we can repeat going forward so that we don't get lost like we have before.

 **Tuhin** 29:32  
Yep.

 **Speaker 1** 29:35  
Yes.

 **Tuhin** 29:40  
Yep.

 **Speaker 1** 29:40  
Exactly.

 **Carl Droste** 29:41  
Yep.

 **Tuhin** 29:44  
Sounds good guys.  
Thank you.  
We'll talk soon.

 **Carl Droste** 29:45  
All right.

 **Tuhin** 29:46  
Yeah.  
Thank you.

 **Carl Droste** 29:47  
Thank you all.

 **Tuhin** 29:47  
Bye bye.

 **Speaker 1** 29:47  
OK.  
Thank you.

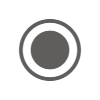
 **Natarajan Sundararajan** 29:49  
Yeah.

 **Carl Droste** 29:49  
Have a good rest of your Thursday, Wednesday bye but.

 **Prashantkumar sharma** 29:50  
Thank you enough.

 **Rupali Ghosh** 29:50  
Thank you.

 **Speaker 1** 29:51  
You do.  
Have a nice day.  
Bye bye.

 **Prashantkumar sharma** stopped transcription