AI | Expert Insights

Customer Prefers to Split Workloads Across Big Data Platforms According to Best Functionality

INTERVIEW DATE	EXPERT PERSPECTIVE	ANALYST PERSPECTIVE
10 Jan 2024	Customer	Investor-Led (Buy-Side)

CALL SUMMARY

1. Thinks Databricks has the best generative AI capabilities after its acquisition of MosaicML.
2. Believes TDC handles most of its AI/ML needs due to its scalability, multi-cloud, on-prem, and customer service.
3. In the expert's opinion, SNOW is a niche solution due to its vertical specialization and links with manufacturers and telecoms.

Expert Bio | Employment History

Expert is a Senior Director at Qualcomm, responsible for leading the AI product team and driving the mission statement. Expert reports to a Senior VP, ~2 levels from the C-suite

Show More

Governing Body Member - SFO CDO | Evanta, a Gartner Company | 6/2022 to Present (Public)

Member of The Board of Advisors | AI Forum | 1/2021 to Present (Private)

Industry Advisory Board - Big Data and Artificial Intelligence | San Francisco State University | 8/2020 to Present (Private)

Senior Director - Head of AI/ML Product Management | Qualcomm | 8/2020 to Present (Public)

Head of AI Product Management - Strategic Architecture | Intel Corporation | 9/2014 to 8/2020 (Public)

Director - Mobile Product | Lenovo | 11/2013 to 8/2014 (Public)

Senior Manager - Mobile Imaging Product Engineering | Aptina | 12/2003 to 11/2013 (Public)

Show More

Analyst
00:00:00

First, this call will be recorded so it can be transcribed. Second, you understand the definition of material non-public information and agree not to disclose any such information or any other confidential information during this interview. Third, you confirm that you do not have a non-disclosure agreement or any other type of agreement or arrangement that would prevent you from speaking about any of the companies that may be discussed in this call. If you can't answer any question I ask, please feel free to let me know and we'll move on. Finally, you agree not to discuss the details of your current employer. Do you agree with this?

Expert
00:00:24

Yes, I do.

Analyst
00:00:25

Okay, great. Just discussing your role, your current usage of data lake, data warehouse products.

Expert
00:00:35

Yeah. Currently, when we look at it from a platform perspective, we look at in terms of data construction, data construction could look at in-database analytics. When you try to bring models, how do you really train the models? Can we train the model at hyperscale? What preferred tools are necessary to optimize performance? How do you prepare the data once you train the models, and then how do you deploy it? If there's some kind of hybrid method of enablement, how do you scale from edge all the way to cloud, working across multiple cloud service providers? That's one option.

Expert
00:01:15

Another one is monitoring. Once you deploy the model in production, is it giving consistent performance? Is there a degradation from a drift standpoint? If there is, how do you monitor that, and what changes is necessary to really make sure that the drift can be made as minimal as possible? I would say pretty much in those lines.

Analyst
00:01:36

Got you. Okay, can you give me an example of just one of your big data applications that you work on in your current role? Just to have a sense for what you're doing with the data.

Expert
00:01:51

It could be in terms of chatbot deployment. Specifically, you have this available for enterprise vendors. Enterprise vendors would like to be talking to a lot of patients and having discussions on medical records, what transactions needs to be happened on, prescriptions needs to be updated or giving real-time diagnoses. For that, there's plenty of medical data and medical history per user, and then we really make the connection with the virtual assistant that's been created. It's customized for that user before it goes into a live person. The data associated with each single subscriber is quite large. That's what a lot of analytics was involved as an example.

Analyst
00:02:42

Okay, that's helpful. All right. What is your history with Teradata products?

Expert 00:02:52	Teradata products, when you look at the anchor point, to a large extent, I would say, you're starting off with LLMs, especially on generative AI is, how do you look at using the platform to, again, anchor on three things, preparing the data, training the models, and then deploying the models. There's a lot more focused attention specifically on the preparing portion and the training portion.
Expert 00:03:17	The intention is because these are large models, it takes a lot of time to do the complete training, and it's very important to really make sure that the training costs can be lowered as much as possible. It's important that you have the necessary infrastructure to do hyperparameter and hyperscale optimizations that can give you results as soon as possible, as fast as possible with better predictions. I would say that is pretty much one of the biggest anchor points.
Expert 00:03:44	Obviously, there is preparing data where we capture the data from a lot of warehouses, which means that if you happen to have a residency that's available that a lot of consumers are using, we need to be in a position to map on to warehouses when data is collected, so we can use it toward our data engineer. Those would be, I would say, key areas from a Teradata Vantage standpoint.
Analyst 00:04:07	**Got it. A more fundamental question, what is your history with Teradata? Did you start with an on-prem product and then go directly to lake on the cloud? Did they do go through enterprise in the interim? Did you try a bunch of different solutions? Did you do an RFP for what you wanted to do? What historically did you use it for versus what are you now doing with it?**
Expert 00:04:36	Yeah, I think it's a combination of a couple of options. Obviously, definitely started off with on-prem. On-prem is quite critical because if you happen to have sensitive data by definition, you cannot really send it off to any public or a private CSP to really manipulate it or get any control of it. That infra will always be there, by definition, it cannot scale, it's limited, so it's used for very specialized projects.
Expert 00:05:02	You have an opportunity there where you can scale, it's not sensitive data, and then here's the question all about scale. Can you use the on-prem-based services, or can you use a combination of on-prem plus cloud? If that is the case, it makes it much more easier. The most important thing is we happen to have the necessary platform infrastructure to be able to scale on-prem to cloud. This is also okay, but it puts a lot of investment on software to be in a position to share that resource with multiple CSPs. It's not easy.
Expert 00:05:39	The last option is going complete cloud, wherein we have the flexibility to doing a lot of things, wherein you don't have to go through the hybrid path of enablement. It's depending upon the use case, depending on the complexity, and depending upon, I would say, the stability of the system itself, you take one or the other.
Analyst 00:06:00	**Got it. Tell me what approach you're taking where you are. Are you going private versus public? You're doing hybrid? You're doing multi-CSP? What approach are you taking?**
Expert 00:06:19	I would say all of them. It really depends upon the segment. If it's anything financially related, or even government-related projects, you want to make sure it's very, I would say, edge-based. You're not showing that many resource. If you are getting more into, let's say, an automotive

environment wherein you're trying to collect data from an ADAS perspective and you're trying to do real-time analysis, to a large extent, it's the combination of hybrid wherein you have certain edge-based processing, and then certain aspect goes to the cloud, which happens behind the scenes. That is like an ADAS environment.

Expert
00:07:04

Complete cloud is wherein you're looking at advanced next-generation use cases like, as I mentioned, generative AI, a lot of LLMs you want to train from scratch or you want to really optimize or fine-tune the performance for a specific use case. You don't need to really do it on-prem, you can completely use the cloud services. That enterprise basically is all cloud-based.

Analyst
00:07:28

Got you. Are the decisions made at a very consolidated corporate level or does the different business units making their own decisions about what they're going to use?

Expert
00:07:41

I think there are two portions to it. One portion is the core technology. The core technology sets up certain rules and regulations on why a certain investment is important, why a certain infrastructure is important, and then they publish a KPI to understand, "Hey, this is the bounding boxes. This is how we're going to see success." Every business unit, depending upon the use case they would like to go for, they either latch on to it and say, "Okay, this makes sense. This meets my KPIs. I'll go for it."

Expert
00:08:12

There are going to be always specialized use cases wherein it does not meet the bounds, and then you have to have certain exceptions. Those exceptions are either back to the core technology team to help. If the business unit has their own, let's say, specialist services, then they try to go on their own. Going off in a rabbit trail, we try to keep it minimal because then integrating it back in the core becomes quite difficult, but that's how we try to maintain the balance.

Analyst
00:08:41

Okay, that's helpful. In some of the preview questions, talked about the various solutions that you use. Can you just go through them one by one, Teradata, but also the other solutions?

Expert
00:08:55

Yeah. I think the other one was, if I remember correct, was Snowflake and Databricks. Snowflake is definitely helpful. I think they really service the telecom industry quite well, and they also service the manufacturing segment quite well. The reason why we try to suggest that telecommunications is a very specialized field, there were a lot of work that happened on signal integrity, 5G, 6G, those kind of stuff wherein you need to have a close partnership to be able to invest in data from towers.

Expert
00:09:32

They have, I would say, a pretty good infrastructure to support telecom-based use cases. Also the same thing with manufacturing, especially when you're going into silicon packaging, detection, those kind of stuff. It makes it quite more helpful. I don't think Teradata really offers that point of easy integration, especially, as I mentioned, for telecom and you want to do fraud detection in telecom, it's not that easy.

Expert
00:10:06

Perhaps on Databricks, Databricks recently acquired a company, I forget the name of the company, for generative AI. Generative AI, they have something called retrieval augmentation, wherein the intention is if you want to do specialized queries which are based on documents. Especially for application developers, they want to be deploying a new application, they don't understand how a certain module works, and the company is dumping hundreds and hundreds of pages of documents.

Expert 00:10:46	Databricks offers the path wherein you put in a query, the query goes into a database, the database matches with specific context, that context gets added on to a query, and then goes into an LLM. Sometimes [Lang] implementation is very unique and available only on Databricks, and that makes it a big advantage for deployment, especially on productivity-based applications, which is heavily reliant on a lot of documents, and those kind of stuff.
Analyst 00:11:14	**Got it. I assume you've also tested out maybe some of the CSP-specific products as well as maybe MongoDB, things like that.**
Expert 00:11:24	Yeah. For example, you have the Vertex AI coming from Google and you have the Bedrock coming from Amazon. They're okay but they're very generic in nature, meaning it's very constrained to the fact that you have to deploy it on a cloud. There is no scalable option to run it on the edge. On top of it, you can't bring in third-party models. It is all, for example, if it's on Google, it's all Google-based models.
Expert 00:11:56	Amazon is opening up to also support third-party models where you can do optimizations. This is definitely a possibility, but it's very vanilla-based service. There's no specializations as a function of industry. That challenge hopefully changes, but that's where it stands today. Obviously, you have Lama and other stuff, that they're very open-source versions. They're very foundational models. They're the jack of all trades, master of none kind of thing. You can use it but it's very, again, very native, vanilla form.
Analyst 00:12:31	**Got you. It'd be helpful to understand when we're talking about your usage, and actually, I'd be curious on what the business units are requesting. It'd be interesting to understand if this is focused or if you could delineate when you're talking about maybe what you're seeing corporate-wide across the company and then what you're seeing in your specific application area.**
Analyst 00:13:02	**Interested in some of these topics that you're talking about. I want to understand how specific is it to AI and how specific is it to, let's just say, some finance function or some simple integration function somewhere, or data visualization. Anyway, just quick thing that you brought up, and then we'll get back on track, but edge work or the ability to run on the edge, does Snowflake do that well? Does Teradata products do that well?**
Expert 00:13:43	Yeah. They don't usually run it on the edge, but they give you an interface to map it on to your edge-based server, which means that at any point of time you want to partition the workload, it becomes much more easier.
Analyst 00:13:59	**Got it, interesting. What are your go-to products when you do want to run something on the edge? Is it Snowflake and Teradata? Are there other things that go in the mix there when you want to access big data on the edge?**
Expert 00:14:17	I think it's going to be a combination of all three. Especially if it is, let's say, generative AI with RAG-based options. RAG-based option is very enterprise and internally oriented, enterprise specifically, for vendors like banking where there's a lot of queries that come in or customer

requests come in. If request from a customer is also based on an older database, all that stuff gets mapped to a large extended database.

Expert
00:14:49

If you have use cases which are in the telecom and manufacturing segment, then it goes into Snowflake. Most of the other population bases internally, but specifically for finance, IT, marketing messaging, Vantage really helps out. I don't know the exact distribution within each of the domains. They would take equal, but there will always be product launches and marketing departments or IT departments scale up. Vantage possibly gets a higher portion of that in a given time, whereas if you're going towards next-generation research stuff, then Databricks takes over. That's...

Analyst
00:15:37

Got it, that's helpful. That's what I wanted to dig into a little bit with you is why are these products specialized in those areas? If I understood that right, Vantage is more for the marketing, even interactions with customers, it sounds like. Snowflake is used more on the manufacturing side, and then Databricks is used for anything that's new and upcoming. Is there a legacy reason for that? Are there technical reasons for that? Was there some CTO-type decision as to why that is? Is that historical accident?

Expert
00:16:24

I think to some extent, the infrastructure that has been set up, for example, Snowflake, has a very good connection with the Telecom Data Cloud, which helps the servers and the technology partners, which means that if you want to really improve any customer experience, maximize, let's say, operational efficiency real time, you already have the infra set, so you really don't have to work through it. You're good to go. They have a pretty good technology partnership with AT&T, with Verizon, those kind of stuff. It's a walk in the park.

Expert
00:17:02

That kind of setup doesn't exist much as an example. When you look at the Vantage point, they already have a fabric and a server VantageCloud setup for IT, which means that they have a lot of these customer inquiries that they've already worked through, maybe in big shops like Dell, HP, Lenovo kind of stuff, so they know exactly by model if it's a 15-inch laptop, a 13-inch laptop, what kind of questions they get. They have a very strong database, which means they can provide a faster response based on the database.

Expert
00:17:43

That becomes a much easier problem to handle because the infra already exists and you build on top of it. That's how I think it really depend upon specialty and exposure and experience. There might be a little bit of familiarity, I think, that people have, let's say, that they don't want to make changes, but they've started it but mostly incorporate it.

Analyst
00:18:07

Got it, okay. Digging a little bit on the VantageCloud Lake solution, were you part of the evaluation process? At some point somebody said, "We have on-prem, and now, let's go to the cloud version." Were you part of that? Are you familiar with how that decision was made?

Expert
00:18:38

Yes, we did, yeah.

Analyst
00:18:40

Got you. What's your take on VantageCloud Lake as a solution? Just very open-ended, what's good? What's bad? What's not working? What met expectations, what didn't?

Expert 00:18:58	There are two portions for that. One, let's look at purely the features that they offer in the solution kind of thing. Does it have the flexibility to integrate with multiple CSPs? The answer was absolutely, yes. You could do it with AWS, with the Microsoft, Google Cloud. There were no issue with that.
Expert 00:19:20	The second challenge was, can it support transactional-related activity? The transactional activity should be in a position to interface with various flavors of warehouses. This could be a data warehouse or even data lakes, if it happens to be in terms of collecting data. They really had that portion of infra was pretty good.
Expert 00:19:42	The next question becomes, as we start injecting data, as we start looking at doing a lot of manipulations on the data, do we have the necessary visualization tools that makes the job of a data scientist much more easier? This could be in terms of your IDE or analytic tools stuff. That was not great to begin with, but with time, it definitely improved. As it stands today, I think quite a few people are happy with it and they are all good to go.
Expert 00:20:15	Now, the question really becomes, "Okay, now I want to work on sensitive data. Can I be in a position to interface with edge-based servers and not just end up in a public-based space?" That connection of like [inaudible], which means we have a broken system wherein, "Oh my God, I have a gap. How do I close it up?" That's a problem. I think if you're looking at scalability, awesome with cloud, scalability with edge, an issue.
Analyst 00:20:50	**Help me understand that. I'm just being dense here, but why is it a problem when you have scalability with the edge?**
Expert 00:20:59	Because it's in front of nexus. They don't support doing partitioning of data between edge and cloud. Everything needs to be on the cloud. They need to bring it to their own server to really process information and then give out.
Analyst 00:21:13	**I see. Does anybody else do that?**
Expert 00:21:19	Databricks does it. Snowflake supports it as well.
Analyst 00:21:26	**Okay, got it. It's an inferior solution if you're dealing with versus Databricks or Snowflake. Is that accurate? If you're talking about stuff that you're trying to put on the edge partially?**
Expert 00:21:41	Exactly. It doesn't give you the variables. I don't think you're catering to the audience who prefer to have an edge option, especially working with sensitive data. Now, obviously, the pitch would be, "If you don't need edge, you come to me. You will have virtualization option completely available. Nobody is going to look at your data. Nobody is going to be accessing the data." It's often pitched, but it's always going to be a doubt that "Hey, we're going to miss something." That's where I think that support doesn't exist.

Analyst 00:22:09	**Got it. Despite those issues, do you use it for any edge applications right now? Do you exclusively use Snowflake and Databricks for that?**

Expert 00:22:22	Yes, exclusive. Teradata is not used for the edge at this time.

Analyst 00:22:27	**Okay, got it.**

Analyst 00:22:36	**Which are growing fastest? I don't know if there's any way to quantify how much usage is going on to each platform and/or which ones are growing faster, whatever sense you have.**

Expert 00:22:54	I think when you look at Teradata, the most important element, I would say, is the AI/ML scale at enterprise is what is scaling up specifically, and to support any transactional workloads mapped on to enterprise use cases. That's what a lot of traction I would suspect. Snowflake is, to a very large extent, specialized, which means mapping onto the telecom and the manufacturing line. They're very select. You're not going to see a lot spikes in that because it's very normalized.

Expert 00:23:36	Databricks is scaling up only because of what [inaudible] by the ecosystem and the demand for generative applications. Because they recently acquired a company, I forget the name, it's slipping me. That company is quite popular, having been quite popular for generative AI. Some people are trying to see if they can utilize that without complications.

Analyst 00:24:07	**Got it. MosaicML maybe.**

Expert 00:24:09	Yeah, MosaicML, exactly.

Analyst 00:24:12	**Okay, I just googled that. All right. I think you highlighted that AI/ML is being done on TDC, but generative AI is being done on Databricks. I just want to make sure I understand when you're talking about the ML/AI work on TDC, that excludes the generative AI stuff, but it includes, I take it, more analysis of existing data sets, rather than that interactive type data that generative AI works on. Is that the right way to think about that difference?**

Expert 00:24:55	Correct. Mostly productivity use cases or, I would say, what we call containerization data. It's all very traditional bread and butter use cases, not the [inaudible].

Analyst 00:25:19	**Maybe a better way to put it is TDC is more ML and DB is more GenAI. Is that a [inaudible]?**

Expert 00:25:28	Yeah.

Analyst 00:25:29	**Okay, got it. That makes sense. On VantageCloud Lake, do you stick unstructured data on it? Do you mostly use it as a data warehouse, and it's all relational and structured?**
Expert 00:25:45	We usually try to stick to structured data because as part of our first enablement, we want to make sure that the data is structured, it's vetted, it's labeled properly. It goes through an external vendor to make sure that it is consistent and clean. Now, there often going to be situations where it is unstructured to begin with. We try to avoid these scenarios because we don't know the source of it, no idea how good or bad it is, and there's been no filtering function on top of it. If possible, we try to avoid it, but mostly it's a structured option.
Analyst 00:26:18	**Got it. I'm curious, obviously, the advantage of the Lake product versus the VantageCloud Enterprise is the fact that it is more of a solution for unstructured data. Have you tried it for unstructured data? Do you find it, compared to other solutions, a competitive product or lacking or better? If you have use cases for that when you get to that point, would you be comfortable using it for that purpose?**
Expert 00:27:00	I think it is. We have done this as well, wherein we have a data warehouse and a data lake. Data warehouse, to a large extent, is being set up for a compliant environment, wherein you really want to make sure it's all clean setup. Data lake is obviously focusing on raw data. You want to select it and then get it and go. It's usually set up very nicely for staging, connection, but it's not processing.
Expert 00:27:31	Yes. Looked at it but to a large extent, we just don't operate on the raw original data. We try to really make sure that we can, in some shape or form, go through a warehouse, clean it up, get it up in a compliant manner, and then execute it. I would say, personally myself, I don't want to take raw data. I just feel uncomfortable.
Analyst 00:27:57	**Got it. Going back to one of the points you made before about trying to lower the cost of training, how much does that factor into your decision where to put the data? Maybe the original decision to use Teradata in the first place. Was that a key evaluation factor for you? Do you consider any of the product to be cheaper to scale, cheaper to run queries on? There's obviously various opinions out there. I'm curious where you fall.**
Expert 00:28:45	Absolutely. The intention is to reduce the training costs. The training costs are becoming quite expensive. If there's an opportunity wherein these private CSPs are providing an infrastructure to scale performance while reducing the training costs, we'll jump on it. That is absolutely important. All of them is purely going towards not licensed, but actually renting the infrastructure and the farm to train the data. If you can harness it for something else, fully appreciate that, and that's a great, important asset. A lot of folks are going to pay attention too.
Analyst 00:29:31	**Got you. I know I'm mixing up terms a little bit, but do you do training on your Teradata systems?**
Expert 00:29:43	Yes, we do. On Teradata, yes, for traditional ML use cases.
Analyst	**Yeah. Your ML use cases. Do you find a cost differential between using, I don't know if**

00:29:47	**you have a basis for comparison. It sounds like you use Teradata for most stuff for those ML cases. Maybe you don't have a point of comparison, but if you do, I'm curious if there's been any head-to-head versus Snowflake, or even Databricks for other things. Maybe another question would be, why don't you use Databricks for some of the stuff you're doing for Teradata? Why did you choose not to use Snowflake for the stuff that you've got on Teradata? These products are obviously very popular with some of your peers out there.**
Expert 00:30:37	Yes, I think it's a combination. I think when you look at most of these private CSPs, it's what kind of infra they really provide. The infras could be based off NVIDIA platforms or even Intel and AMD-based platform from a training standpoint. If you focus pretty much on NVIDIA platforms, you have three categories. You have the Ampere line, the Volta line, and the Hopper line, what we call the H100, V100, and A100.
Expert 00:31:05	A100, V100, and H100 are quite cheap in nature, which means it's been in generation for quite some time. Now, the question really becomes, is there going to be use cases, at least from the ML perspective, that you don't want to use the H100s, but you want to use the older generation architecture because it's cheaper, it takes more time, that it's fine to use it.
Expert 00:31:26	The question then becomes, if most of these private CSP offers that flexibility? Many of them don't. Even if you don't, then you're stuck with the most expensive solution and it's not very helpful. Databricks, for example, they'd only operate on the latest technology of NVIDIA architectures. They do have the previous generations, but they're completely packed most of the time because it's heavily utilized for many people. The question really comes up, do you have flexibility available at any given point of time? The answer is no, and that's the problematic thing.

Analyst
00:32:07

Got it. Let me ask that question a slightly different way. Why don't you do generative AI using Teradata or Snowflake? Just [trying to] dig a little bit.

Expert
00:32:27

Yeah. When you look at generative AI, there's training infrastructure and the software component that goes along with the training infrastructure. Teradata and Snowflake do have the latest configurations on platforms, but they don't have all the software portions. Remember I was talking about retrieval augmentation, fine-tuning services, low-rank adaptation techniques to really optimize performance on large language models. Those only exist on Databricks based on the acquisition from MosaicML. Teradata and Snowflake, a little behind on that. They're very vanilla-based offering on GenAI, because of which it's not really that good for anyone who compete in that space.

Analyst
00:33:13

Got you. I'm just learning about some of these concepts myself around what you were talking about with MosaicML. Are most of Databricks' advanced AI capabilities due to that acquisition? Did they have a better development platform to begin with?

Expert
00:33:37

They had a better development platform, but it was not complete. I think it was taking a little longer time for them to reach the commercial quality. My understanding is with this acquisition, they did a very hockey stick ramp, wherein they exposed a lot more features, much more faster into the market with the easy integration with MosaicML.

Analyst
00:33:57

Got it, that's helpful. Okay. Have you gone through any evaluation process firm-wide, I don't know, data platform type review in any recent period? Were there any conclusions as you were thinking about how things are developing, and if it needs to develop any differently?

Analyst
00:34:39

Were there any major outputs from that as to things that are missing, things that are not working well? Are you happy with the state of your various systems in relation to what you're hearing out in the industry that's possible?

Expert
00:34:59

Yeah. I think in the context of Teradata, there's definitely an opportunity for improving, especially in the analytics portion. Recollect I mentioned the command line interface with the visualization portion. The visualization portion really helps on model condition. It provides a lot of log information to data scientists on what optimizations can be done from their side. It's not the best, but I think they are improving with time on the analytics portion.

Expert
00:35:27

I think if there's an opportunity to improve, definitely do that. Obviously, the feature offering very weak on generative AI, at least on Teradata. If they can go beyond your traditional ML use cases and more generative AI, they can play catch up and definitely offer a flavor in comparison to Databricks. Again, the more the options, the pricing immediately drops, then the consumer has a lot more options. Right now, that is not the case. I would say that would be some of the pointers.

Expert
00:36:05

When did the data visualization and analytics start to improve? Do you know was there a turning point? I know they've been trying to make various improvements to their platform over the last

several years.

Expert
00:36:23

I think it would take about 12, 18 months wherein they offer the flexibility of bringing your own analytics, they have different languages and APIs, making it much more easy to ingest, those kind of stuff.

Analyst
00:36:37

Okay, got it. Given the timing of your moving on to VantageCloud Lake, it sounds like you might be one of the early users because they really only came out with that Q3, Q4 last year.

Expert
00:36:56

That's correct.

Analyst
00:36:57

Okay. Did you find any bugs with it? Is it working exactly the way, with high quality, high service levels, no major issues with it?

Expert
00:37:09

No. Whenever you try to get into these big contracts, there's always going to be an evaluation cycle. The evaluation cycle obviously starts, it's quite large in nature. When you try to do evaluations, not everything is perfect, but the expectation is do you happen to have a strong roadmap that really addresses it, especially if you want to bring your own model or you bring your own analytics, that capability and support for data science teams.

Expert
00:37:36

If it exists, confidence in the team can deliver, awesome, and then there's no complaint about it. The intention really becomes once you set up a certain vision, executing to that vision on a regular cadence, if that cadence slows down, then becomes a problem. That's what I think is the concern.

Analyst
00:37:56

Got you, okay. I just want to make sure I heard correctly, but it sounds like you're pretty happy with the base product. It does what it was meant to do. There's obviously the challenges of getting new systems up and running and making sure all those little kinks are worked out, but for what it was meant to do, it's good at. What it's missing are some capabilities, especially around data analytics. It doesn't really get any of the GenAI work because it doesn't have the capabilities that Databricks has.

Expert
00:38:32

That's correct.

Analyst
00:38:34

Okay, got you. How responsive are the various vendors for your needs? When you tell them, "We need this product feature or this thing is not working the way it should," are you able to get the help that you need? Do you work with any of their technology partners? Do you do it all yourself? Do you work directly with their engineers, consultants? Do you work with third parties, or again, do you do it all yourself?

Expert
00:39:17

No, I think there will be a lot of conversation with what you call customer enabling solutions team. They interact with the specialist, they are called, technology specialists, and things really roll that way for sure.

Analyst 00:39:37	**Okay. From that perspective, you don't see any issues, or let me know if you do, with any of the providers.**
Expert 00:39:47	Yeah, of course.
Analyst 00:39:50	**Okay, let's see. Yes, so I wanted to understand how aggressive are the vendors being with pricing. Are you seeing more aggressive offers? Is pricing pretty consistent with what it has been in the past? I don't know how you measure it, per query or per node. I've heard at different times. Are they taking...**
Expert 00:40:42	Yeah, I would say definitely becoming aggressive and it keeps changing on a regular basis. It really depends on how other vendors start to scale and start offering lower prices, and then also many private set-up companies are coming up like Lambda Labs. When you happen to have it drastically smaller, many of these private CSPs need to react to that situation and be in a position to supplement and support.
Expert 00:41:11	Yes, it happens on a much regular basis than anticipated. Obviously, if you have a contract that you're already finding, you don't have much of a choice to make adjustments, so you're going to have to live through it, and then once the contract ends, then you say, "This will be the updated version towards supporting that."
Analyst 00:41:29	**Yeah, that makes sense. I'd be curious, just a similar question where we talked about why you don't put GenAI on Teradata. Why don't you put more ML on Snowflake? Those guys are very aggressive and they've been capturing share in a lot of places. Are there any technology differences? Is it just where the data already is that makes you not do it? Do you see it just as a very niche solution? What's the reason for not doing more with Snowflake?**
Expert 00:42:11	No, I think it definitely can be done. It's one of those things we'll started off early on Teradata, and because of which, we let it go. Definitely, valuation is happening to see if we can transition some of the ML applications to Snowflake. Definitely they got quite impressive in terms of pricing and performance, so it is just a matter of time. I think if Snowflake continues to improve it, there will be a certain amount of transition for sure.
Analyst 00:42:40	**Yeah. The last time you evaluated Snowflake versus Teradata head to head, what were the technological differences?**
Expert 00:42:51	The technology differences, I would say it was specifically in terms of flexibility, available to multiple CSPs, and then does it have support for injecting structured and unstructured data. Data does not have to be only visual data, it could be textual data, audio data, speech data, or a combination of that.
Expert 00:43:18	Some of these warehouses, especially Snowflake, I don't think, if I remember correctly, had the opportunity to ingest audio data, whereas Teradata did. Again, the modalities make a difference. I would say one of those reasons why Teradata started off and obviously, many of these services

came in, I think, second half of last year. Great customer support. It's fine if that's the case, if not, just get started.

Analyst
00:43:49

Interesting. I don't want to put words in your mouth, but when you did the evaluation, which was the better product for your purposes?

Expert
00:44:03

I think it started off with Teradata. End of the day, it was just not the features, but also the customer support, integration, documentation, ease of use, scalability, all of those factors. It came up quite well compared to Snowflake. Now, Snowflake is a very big customer, multiple vendors. You're not guaranteed service. I think it's like a platinum, silver, gold, whatever that layered tiering is, you get one or the other.

Analyst
00:44:37

Got you. Okay, that's helpful. Maybe we're getting near the end of my question. As you look forward, are there any big trends that you see, whether related to these data vendors or just things that you are thinking about that will have a big impact on how you ingest data and then use the data?

Expert
00:45:13

This time, no. Through discussions on multimodality fusion, for example, if you happen to have a stream, it's a combination of vision plus text. How do you look at the data? Most of the data today, or even the infrastructure setup, is only built for unimodality ingestion, meaning it's only vision or only speech, but not a combination of both. This is not really sure, so that's one of the things I think we need to pay attention to. I would say that's probably will be one thing.

Analyst
00:45:47

Sure, got you. Do you work with any of the smaller non-public? Obviously, Databricks is also private, but do you work with any more up-and-coming technology providers that you think are pretty interesting?

Expert
00:46:03

Yeah. For example, Lambda Labs, one, then you have Together.ai, another one.

Analyst
00:46:15

Can you spell those for me? I think it didn't come through as clearly on the phone as I'd hoped.

Expert
00:46:21

Lambda, L-A-M-B-D-A, Lambda, Labs, L-A-B-S.

Analyst
00:46:27

Lambda Labs.

Expert
00:46:29

Yes. The other one is Together, T-O-G-E-T-H-E-R, Together.ai

Analyst
00:46:37

Okay, got it. That came through. That's pretty helpful. All right, time is about up. Any final thoughts just related to anything that we discussed or that you think are relevant?

Expert

No, this is good. Thank you so much for taking the time. I think we covered pretty much all of it.

Transcript produced in partnership with

MOSAIC
RESEARCH MANAGEMENT