Daniel Newman: Hey everyone, welcome back to The Six Five Summit. It's day one, this is the track you have all been waiting for. It's the AI track. If you haven't been paying attention to all the talk that's been going on this year about generative AI, about AI as a whole, about enterprise AI, then you probably haven't turned on your TV or read anything recently, because it is all the rage. And this is a track that so many people are excited for, we got a great day ahead, and our track opener is a returning guest to the Six Five Summit, Rob Bearden, CEO of Cloudera. Rob, welcome back. So happy to have you here at the Six five Summit.

Rob Bearden: Hey, good morning, Dan. It's great to be here. We appreciate you having us and congratulations on your recent announcement on Futurum AI. I think it's really cool.

Daniel Newman: Yeah, it couldn't have been more timely, Rob. We are all having to reflect on our businesses right now, and the impact that AI is going to have. It's probably something you and your team is going through. It's certainly something our team here at Futurum, we are all going through. This is going to change the world. So I'm really excited to have you here. Big topic. I'm really glad you're coming back and you're leading this track and this conversation, because so much of what AI is able to do, it's all about building on data. And so that's something that Cloudera is built around, it's been doing for a very long time. So let's start here, because if you've read my commentary, I've been following Cloudera closely for many years. You guys are doing a lot of important things for data management, but not everyone knows who Cloudera is. So start with that, Rob, talk a little bit about the business, where you're at, and why it's so important to the enterprise AI ecosystem?

Rob Bearden: Yeah, sure. Absolutely. So when you think about Cloudera, what we focus on is enabling our customers to manage and unlock the value of their data. And we're a hybrid data company, and we solve data management challenges across both the public and private clouds. In fact, we've been the big data management leader for the past decade and we have actually 25 exabytes of data under management across our customer base. And for context that's on par with every one of the hyperscaler's data under management. But in addition to that, we've got another 25 exabytes of data that's under management on our open source platforms. And from a commercial standpoint, we have over a billion dollars of revenue and that puts us in the top 1% of all software companies in the world. And we have very healthy operating margins, generating over 450 million of adjusted EBITDA this year.

And if you take the top 10 largest global companies for any given industry, on average, nine of those are Cloudera customers. And the important point is that we've been the market maker for big data and analytics for the past decade. And when you look at how we've evolved as a company over the last 10 or 12 years, you're going to see why we're so well positioned for this new world of enterprise AI. And so when Cloudera really began, we think about that as Cloudera 1.0, that was all about building an opensource based enterprise data platform, with the highest levels of security and governance, that allowed our customers to finally get rid of those data silos and manage petabyte scale data. And what this did is it finally gave them the ability to drive insights across all their data. And then we transitioned to Cloudera 2.0, and that's where Hortonworks and Cloudera emerged.

And that allowed us to rationalize all our product portfolios, capture a lot of synergy, and really accelerate our path to the hybrid cloud. And then we began our journey on Cloudera 3.0, and

that's been all about making the hybrid, multi-cloud data market function, and function at scale. And we've created the first true hybrid, multi-cloud data platform, and that's allowed 90% of our legacy Hadoop customers to migrate to our new platform, which is CDP, both public and private cloud. And now by leveraging the CDP open data lakehouse, they can do enterprise AI across all their enterprise data. And so the powerful part about this is it positions our customers to leverage CDP open data lakehouse for the foundational models and LLMs, and drive generative AI based applications in a secure, trusted, and responsible manner, better than any other platform in the world.

Daniel Newman: So Rob, first of all, I just want to double click on something you said. You mentioned 25 exabytes you have under management, and then another 25 exabytes out in the opensource ecosystem running on Cloudera. As far as I know, that has to be something like 100 plus times the amount of data on a cloud platform like a Snowflake. I mean it's a huge amount of data, and it's obviously so critical for companies to be able to unlock the value through what we're talking about in generative AI, through LLMs. So you mentioned that, talk a little bit about how the open data lakehouse and the LLMs that it can consequentially deploy in their businesses can unlock that enterprise value, Rob?

Rob Bearden: Yeah, sure. Well actually, Cloudera has been helping enterprises gain value from AI and machine learning models for years now. Our machine learning service has become really the best in class, and it covers the full ML life cycle, from exploratory data scientists to model training, and actually the overall deployment of those models. We even have a library of the end-to-end applied machine learning prototypes, where we call them AMPs, that help our customers get started quickly developing their own applications, and see value in a rapid way. And when we talk about generative AI and LLMs, what we've learned is they're only as good as the data they've been trained on. And the feedback that we've had from our customers is they want to create their own versions of GPT, and build their own foundation models, rather than trying to train data on public API plugins.

And for these models and the AI to be successful, it needs to be trusted, and trusting the AI starts with being able to trust your data. So what we've done is enabled our CDP open data lakehouse to provide the foundation data, with security and governance and the enterprise context, that our customers can use to deploy their foundation models they build, in the data center or across the cloud, and we deliver this in an enterprise ready fashion. And what this really means is that with all our CDP open data lakehouse, all our customers' data is enterprise AI ready. And look, this is a perfect time to announce that we have an immediate availability of the first CDP LLM AMP. And what this does is it allows our customers to build their own AI applications that are powered by an open source LLM with their enterprise data. And it's available both for CDP public cloud and private cloud right now, it's free, and it's really easy to get started with. So Dan, we want to see you leverage this with the Futurum AI stack, and I think you'll really like it.

Daniel Newman: Yeah, I assure you of this, Rob, we're going to play with it. Because A, it's going to give us the opportunity to provide the market with some real analysis of just how flexible and fast. You said just a few minutes, so I'll be texting you when I'm done, and say, "Hey, I just built an app in just a few minutes." Now, actually, Rob, what you said, first of all, I was very excited. Second of all, I

love announcements at the Six Five Summit, so very excited to hear that you're launching the availability of the CDP LLM AMP. Having done this, I can tell you this is achievable.

And I know a lot of people out there are probably in varying states of how much to embrace this AI, how quickly to deploy it, how accessible is it, is there a democratization, can small companies really play, or is this going to be a large enterprise's world? And I think this maybe democratizes and makes the playing field more level than we've ever seen it before. This is really an opportunity. And so everything you just said, it's really thought provoking. And one of the things that I'm always really interested in hearing is how are organizations doing this? So as the CEO of the company, you're out there not only building product and building the business, but you're out talking to customers and partners. Talk a little bit about how your customers are already taking advantage of what you built.

Rob Bearden: Yeah, it's mind numbing how quickly you can transform an overall business model by leveraging LLMs. And a great example of this is one of our large banking customers in Singapore, which is OCVC. They implemented an open source LLM on CDP private cloud, and their goal was to automatically generate code that was using models from Hugging Face. And they were actually able to do that in a matter of just a few days, by leveraging the CDP machine learning capability. And they saw immediate value, and it's something that would've taken them many months to do without those tools, and without those LLMs. Another great example is the IRS's Research Applied Analytics and Statistics Group. And what they're doing is they're leveraging CDP to train and run their fraud detection models. And what they've been able to accomplish is improve their performance by three or 4X for their data science and data engineering pipelines. So it really drives a TCO very, very significantly.

Daniel Newman: And Rob, you mentioned Hugging Face, and Hugging Face, I've seen on a few stages now, definitely becoming a bit of a beacon of opensource, large language model partnerships. So I'm really glad to see you've gone down that path. But I've also just noticed as a whole, this AI boom has created a ton of ecosystem development. What is your approach, and who are some of the strategic partners that you're focused on for AI at Cloudera?

Rob Bearden: Well, our strategy is to be able to allow our customers to bring their own data, their own LLM models, transformer libraries, and really run those across all the CDP enterprise data. And that's going to significantly accelerate the development of open source AI applications for the entire community. And a core enabler of this is going to be leveraging IBM's new Watson X Studio in order to apply the foundation data from the CDP open data lakehouse actually to the Watson X Foundation models. And obviously the AI workloads are very GPU intent, very compute intensive, and our partnership with NVIDIA and Intel are really key to accelerating the innovation for our customers. A great example is the CDB NVIDIA configurations that are increasing the performance up to seven, eight, nine times, at less than half the cost relative to what modern CPU only alternatives are. We also have very similar configurations with Intel on their Xeon processor family, and are getting very, very similar performance results as well.

Daniel Newman: So one of the things that, as I've looked more deeply under the hood of Cloudera, we talked a little bit about the volume, Rob, of data under management, and that is incredibly impressive. And I think anybody that's tracking what's going on with AI knows that there's this inflection

where you need large models to create smaller models, and then smaller models to create tinier models. And then in the end, we need to create efficiency, Rob, we need to... Less power, we need to be able to do more with less data. That's where we're all going. But in the beginning it's all about having that right initial data set. That's why these large language models are so big.

I think as I reviewed your stack, Rob, I said, "You've really been making customers AI ready for a long time." This isn't a new thing. Customers, clients in running Cloudera have been putting their data, organizing it in a way, structured, unstructured, which is ready for this day. So I'm guessing that's probably why you feel it, besides you being the CEO, that Cloudera is the best open data lakehouse. But why do you think customers are choosing, why do you think you have nine of the 10 or that stat you mentioned, why are they choosing to build on CDP open data lakehouse when obviously there are many different approaches they could take these days?

Rob Bearden: Yeah, well what our customers tell us they want in an open data lake house is first they want an integrated platform to unify the capabilities of both a data warehouse and a data lake, and that powers business intelligence, machine learning, and AI solutions across one single platform. They also want to be able to leverage opensource innovations that come from the community, such as Iceberg and Airflow and Unicorn. And most importantly, they want to have the flexibility of a hybrid multi-cloud environment. And they want to be able to deploy across both the public and private cloud of their choice, and be able to have interoperability and freedom to store and drive storage and compute wherever it's most architecturally and financially pragmatic to do so. And Cloudera is the only company on the planet that offers a true hybrid open data lakehouse, with an identical experience for both public and private cloud, at scale, with the best TCO in the industry. And we're able to do that because of the scale and capabilities of our data fabric.

Daniel Newman: So let's double click on that one. We talked about the AI angle of it, but in terms of the open data lakehouse, this is more foundational. What are your customers saying there?

Rob Bearden: Well, as you can imagine, I meet with many customers every week, and the feedback's been extremely positive. A good example is Santander. They had multiple business units, including the risk group, finance, capital markets, and each one of those groups were building their own data warehouses using very different sources of inconsistent data. And they all had different ways of calculating the same metrics, leading to multiple versions of the truth. And now, by leveraging CDP open data lakehouse and our hybrid data fabric, they have a system of record now across each line of business, and they're finally able to have a single version of the truth, which is generated from over 40 million customer records spanning 10 years of historical data. And this has delivered, as you can imagine, tremendous customer benefits. And what we've been able to do is really drive the best TCO for them across any of their data management platforms.

Daniel Newman: Yeah, so one thing we haven't talked about is more about the infrastructure and fabric for really building for the future. I've been on the record with you a number of times, I've been on the record with other companies, industries, and media, talking about hybrid being the model that we believe is the future. And it's improving, in fact, because every prem-based focused company has moved towards the public cloud, and every public cloud provider has moved back towards prem. And so I think we've come to the consensus that hybrid is the answer, but you have a very specific approach at Cloudera to hybrid. What does it mean to you and the company?

Rob Bearden: Well, everyone seems to have their own definition of hybrid, but everybody wants to get there it seems. But what our customers tell us is they want a platform that has the ability to move workloads, and data, and the metadata, and the applications, bidirectionally across both public and private clouds. And we've delivered on this hybrid vision through a single control plane, which manages a common security and governance framework across the platform in all of our data services. And what this has done is it's allowed our customers to deploy the CDP hybrid capabilities without any application refactoring, and they now have an identical experience on both public and private cloud. And so what hybrid really means to us is that all your data is enterprise AI ready now.

Daniel Newman: Yeah, I think the hybrid adoption curve, though, is maybe the most challenging thing, Rob, for companies. It's that they all know they need to get there, and it's, like I said, it's been a bit of a sliding scale. You had an era of born on cloud companies that started out, and the economics just didn't work that well. That's largely what we found out is when you hit a certain scale, there was laws of diminishing returns. You saw Andreessen Horowitz published something really interesting about that. And then on the other end of things, you saw companies that were built entirely on-prem, and they're realizing that they needed more of that flexibility to be able to be on-prem and in the cloud concurrently, and again, created challenges to move that way. So what are you seeing in terms of that adoption curve for Cloudera companies being able to make that move? Because they're going one way or the other, in the end, to land in some hybrid mix of cloud and prime?

Rob Bearden: Yeah. Well, we've seen really strong adoption trends, and they're actually accelerating right now. And what we've seen is this happen in two phases. The first phase has been our customers who were migrating from our legacy platforms to CDP, and actually 87% of our ARR is now on CDP, and we're trending that by the end of the year, to have 94% of our ARR actually tied to customers on CDP. So that's been a rapid adoption rate. And that cohort of customers is also seeing a very, very strong net expansion. And so obviously that means that once they start using CDP, they rapidly expand their use cases and applications. And this is now leading us to the second phase of the adoption cycle of the CDP data services. And in our last fiscal year, we've really seen the adoption of our data services explode, and overall, data services ARR is roughly doubling every three quarters.

And this continues to accelerate as enterprise AI is going mainstream now. But I also share a few other customer success stories that might help realize really the scale we're actually talking about here. I'll start with one of the largest banks in the US. They began migrating petabyte scale workloads from one of our competitors' platforms, and they now have 250 petabytes of data on our private cloud data services, and they estimate they're now saving as much as $2 billion a year in infrastructure cloud cost. We also have a global pharmaceutical company that's using CDP public cloud and our data services to manage 25 petabytes of R&D data that's used by more than 12,000 users. And this has led them to be more than two times as productive, and massively accelerated their time to market through their R&D pipelines.

Daniel Newman: So Rob, it seems that you're hitting on all cylinders. We've talked a lot about hybrid and AI, and this is the AI track at the Six Five Summit, so that's what people do want to hear about. But foundationally, and beneath the AI, there's a lot of data management requirements. Talk a little

bit just about some of the other innovations in data management that you've been working towards.

Rob Bearden: Oh, wow. Yeah, this is the exciting part. Of course we're all in on AI, and we'll continue to invest very heavily into the open data lakehouse that powers the enterprise AI applications. But in our next CDP open data lakehouse release, we'll integrate the platform and the data services with Iceberg, across both public and private cloud, and this is exactly what our customers have been asking for. And also stay tuned here because there's going to be some more detail on what we're doing. And we've also recently launched Ozone, which is our next generation storage technology. It allows customers to manage billions of objects which are supported by Amazon S3 APIs.

And the adoption curve on this has been really impressive. We already have one customer that's got 30 petabytes of data in production on Ozone. On the streaming side, we now offer a flow-based, low-code environment called Cloudera Data Flow. We have over 450 connectors that help our customers connect almost any target source including Salesforce, and GitHub, and Snowflake, and many, many others. And hopefully you've seen today's announcement about our Cloudera Observability product, which gives our customers unprecedented visibility into workload and resource utilization, allowing them to better control and manage budget overruns. So obviously you can see our continued commitment to innovate across all our entire product suite, so that our customers can get insights from all their data for the enterprise AI services.

Daniel Newman: Well first of all, Rob, congratulations on the launch. Observability is a big topic. I've talked about it for a long time. People could follow a lot of my commentary, but you need to have that visibility into your data ecosystem in order to maximize value in your business. So makes a lot of sense with all the data you guys have under management there. It's also really great that you have this broad-based customer ecosystem running on both opensource and natively on Cloudera, that are able to be your first customers out of the gate when you're launching these new services, because it's so important to build these proof points. As we've gone through, Rob, all these, you've been able to really say, "Here's a large global institution in finance, or in manufacturing," or whichever industry, that's able to really build on this. And I think it's important for those that are making mind decisions to hear from the community.

Say, "Hey, who else is taking this risk?" Or, "Who else is making this commitment?" So I love all that. Now, as you know, the event was all about dealing with rough waters. We called the event Navigating Rough Waters because after two or three years of really booming technology, we've seen a bit of a pullback in the market, and I don't think any company has been able to completely run away from it. So just curious, Rob, your take. We've had a bit of a downturn, we've seen it pick back up a bit, and it's been ebbing and flowing, but we saw some really tough waters in '22 into early '23. There's still a lot of calls for recession that may still be ahead of us. Interest rates are skyrocketing, we've got inflation that's been stubborn and hard to break. What is your view on how this is impacting the market in Cloudera?

Rob Bearden: Yeah, there's obviously a lot of headwinds and turbulence, but as a quick refresher, we went private in October of '21, and our goal behind that was to transform the company deeper and faster than we could as a public company. And we focused on transitioning our customers to

CDP, and enabling their hybrid journey. And we did this with a deep focus on getting to best in class unit economics across all parts of the business. And as a result, we're likely to end the year with 94% of our ARR tied to CDP. And our fiscal year plan is to become a Rule of 50 company, and deliver more than $450 million of adjusted EBITDA. And this discipline has made us a much better company, with higher quality products, a faster rate of innovation, and really accelerating value for our customers through our hybrid and enterprise AI solutions.

Daniel Newman: So first of all, it's always nice when a company that doesn't have to disclose its financial does, especially when you're at a billion-dollar size scale, because now that... You guys were public for a long time, and now as private, it's sometimes hard to say, "How are we tracking?" And it sounds like you're tracking really well, and delivering close to 45% or so of adjusted EBITDA against revenue is a very impressive outcome for any business. So congratulations Rob, you're doing something right. So just in case anybody was wondering, Rob's doing really well out there. No, I'm kidding. But in all serious, let's summate this, because you've covered a lot of ground. We talked about LLM, we talked about AI, we talked about a couple of launches and observability in large language model, opensource that you're making available. We talked about the market. What are a couple of the key takeaways that you want our audience here at the Six Five Summit to remember from our conversation?

Rob Bearden: So Dan, there's no doubt there's a lot of turbulence and headwinds that are going right now, but what we know is that data moats are real, and data's the key to enabling enterprise AI. The AI market is changing rapidly, and you have all this happening with new foundation models, new LLMs, new model parameters, and it's changing by the day. But the reality is, data will always be the constant to AI's success. But for enterprise AI to succeed, it needs to be trusted, and trusting AI starts with trusting the data.

And enterprises need to be able to trust the data that is used to train the models. They need to have secure access only for the authorized users, and they need to govern and secure that lineage that tracks that data through the entire transformation process as it's happening along the way. And Cloudera's open data lakehouse and data fabric delivers enterprise AI workloads in the most secure and governed fashion for all the enterprise data. There's no other platform that controls more data types across more clouds with open data innovation at scale, and that's who Cloudera is. We're the data company that enables you to trust your data so that you can actually trust your AI.

Daniel Newman: Rob, this has been really timely, and it's a great way to kick off the rest of our AI track here at The Six Five Summit. It's great to have you back. Thanks for leading the way, thanks for talking about both the foundational, the picks and shovels, as well as the vision and innovation that's going to drive the future of AI. Hope to have you back for the next Six Five Summit Rob, great job, good luck, and congratulations on all the success of Cloudera.

Rob Bearden: Thank you very much, and it's been great being here, and look forward to the rest of the summit. The agenda looks great.

Daniel Newman: And there you have it. You heard it here first from Rob Bearden, CEO of Cloudera. Stay with us. It's day one, this is the AI track at The Six Five Summit. We appreciate you tuning in.