SDS PODCAST EPISODE 979:

AGENTIC DATA MANAGEMENT AND THE FUTURE OF ENTERPRISE AI, WITH ROHIT CHOUDHARY

Jon Krohn:	00:00:00	For years, I've been quoting the stat that the world's data are roughly doubling every year. My guest today says that's way too conservative. He's seeing enterprise data soon growing at close to 10X per year, and most organizations are nowhere near ready for what that means.
	00:00:16	Welcome to episode number 979 of the SuperDataScience Podcast. I'm your host, Jon Krohn. Today, I've got the super zen, super knowledgeable Rohit Choudhary as my guest on the show. Rohit is founder and CEO of Acceldata, a Bay Area startup that has raised nearly a hundred million dollars in venture capital to make the increasingly vast amounts of data that enterprises collect, self-aware, self-optimizing, and AI ready. This is a really cool episode. Enjoy.
	00:00:43	This episode of SuperDataScience is made possible by Anthropic, Cisco, Acceldata, and the Open Data Science Conference.
	00:00:52	Rohit, welcome to the SuperDataScience Podcast. It's a treat to have you on the show. How you doing? Where are you calling in from?
Rohit C.:	00:00:59	Thanks, Jon. I'm calling in from San Francisco. Great to be on your show.
Jon Krohn:	00:01:03	Nice. Yes. I see a familiar background because a couple of months ago we had Ashwin Rajeeva, who is your co-founder and CTO on the show, and he had a very similar background to you, and he was in the Bay Area as well. His episode is sensational, episode number 957 for folks who want to check that out. That one was a great dive into the technical things that you're doing at Acceldata. And of course, we'll get into technical stuff as

well in your episode. But I have lots of questions for you around how all this agentic stuff is going to transform the way that operations work and strategy, all these kinds of big questions that we have.

Rohit C.:   00:01:42   For sure, Jon. Looking forward to diving in. And yeah, Ashwin and my background are the same, not just the background that you see, but also the background.

Jon Krohn: 00:01:51   Tell us a bit about that before we get started.

Rohit C.:   00:01:53   Well, we've been now working for about 12 to 14 years together. This is our third company together, and we figured out that we had common interests, common beliefs about technology, and we wanted to build products that a lot of users could use in serious environments. And we felt that we could build technologies that did not exist. And so it's been a very exciting journey.

Jon Krohn: 00:02:17   That's so nice. And so I imagine you guys are also friends as well. You must be very close friends doing all that together.

Rohit C.:   00:02:22   Of course. I mean, the founders go back a long way. I keep telling everyone that this is a team that has been now together for at least 13 to 14 years, if not longer. And there are people in the company who I have personally worked with for over 20 years now.

Jon Krohn: 00:02:36   Right. Investors must love that kind of consistency.

Rohit C.:   00:02:40   I guess. I think it's pretty predictable for everyone because everybody knows where everybody stands and what are their viewpoints. So there's not much discovery that's left to do.

Jon Krohn: 00:02:50   Exactly. That's really nice. It's important to obviously co-found companies with people that you trust and really enjoy spending time with. And so your role, so Ashwin, he

was co-founder and CTO, your co-founder and CEO of Acceldata, which is the industry's first agentic data management platform, now revolutionizing governance, optimization and AI-driven operations, but Acceldata has actually been a leader in data observability for some time. In fact, you can correct me if I'm wrong on this, but from our research, it looks like you coined the term data observability in 2018.

Rohit C.:   00:03:26   That is right. I think we were the pioneers of this industry. What we figured out when we were starting the company and we started thinking about building this kind of company in 2017, we came to the conclusion that, and going back and drawing from my own background, which was earlier 50% application engineer in the next 50% data engineer. I figured out that data engineering tools were way behind in terms of operational support, production support, real-time visibility into what was going on inside data pipelines vis-a-vis where the application stack and observability had already arrived and reached. And so therefore, it seemed like here was an opportunity of going and building a really big company. And then that was just the beginning of the rabbit hole. And so to your point, obviously we've come a long way from building observability for infrastructure, data pipelines, data quality, data monitoring, to now this whole new vision of Agentic data management platform, which I'm really excited about.

Jon Krohn: 00:04:23   Yeah. It's really cool. And it's amazing to hear that you coined the term data observability. The way that you answered that question, because I asked, did you coin it? And you kind of answered yes, we, but I mean, you individually. It was an interview in Chief Data Officer Magazine in 2018, which is pretty cool because it's a term that is ubiquitous now. You kind of assume everybody who's in our field knows what it means.

Rohit C.:	00:04:47	That was kind of the hypothesis as well for starting the company that this is the missing layer as far as infrastructure is concerned. Because you had databases, you had data warehouses, data lakes, lots of management applications, but there was nothing that was giving real-time visibility to users, to system administrators, to data engineers, and obviously to executives. So we just felt that this is going to be a big thing. And now obviously Gartner talks about it and says that it's a really important component that enterprise leaders should think about, et cetera. So obviously we've come a long way from where we started.
Jon Krohn:	00:05:23	Yeah. And we're going to talk about that journey a bit later in the episode, as my regular listeners will know, I always want to dive right into the exciting, cool things that you're doing right now as early as we can into the episode. And so talking about an Agentic data management platform that Acceldata has evolved into these days, are you able to provide a couple of key customer use cases or user journeys so that my listeners get a great sense of what it means to use an Agentic data management platform and how that improves my work as say a data scientist or another technical practitioner?
Rohit C.:	00:06:04	So let's back up a little bit. Let's go back to ChatGPT when it arrived. I think it took us all in and we figured out that this is going to be a completely new evolution. If you, for example, today want to do research on public market equity and you're planning to build an investment portfolio, you're actually not necessarily going to your wealth advisor. Well, you should, or that's a discretion that you apply, but there's so much information that you can glean from any of these LLMs. Now, the specialty of these LLMs has been that it actually does two things. One, it allows you to do broad-based research, but also allows you to do narrow research, which is in your area of expertise/interest/curiosity. What we found out was that there is so much data lying around in the enterprise, in

the semi-structured and the structured world, and we had been toying with the metadata.

00:07:00   And as we just talked about, we were synthesizing signals across the infrastructure layer, the user layer, and the application layer as far as data is concerned. And the ChatGPT moment essentially explained to us that if you put all of those synthesis of signals together and you allowed people to do deep research on their data, they will find some really, really interesting insights. Now, does that mean that you will be building BI reports with the data that you have access through the ADM platform? The answer is quite really maybe, and I'll explain in a moment because there are several visualization tools you can build reports elsewhere, and therefore that's not the primary use case. But where we are situated in the data pipeline is at a very interesting place. And the opportunity for data leaders and data stewards and data engineers to go curate the data sets that are most useful to manage and operate the data sets that are going to produce the most amount of operational benefits to the business has become easy.

00:08:05   So if you thought of a couple of concrete use cases, think of a CMO who's trying to improve the quality of their targeting in the US around zip code data. Now, the problem with zip code data is that it always comes from variety of different sources, including internal and external data sources. And the challenge with that is that that data is never accurate. Now, please note, if you are a CMO, you're dealing with possibly petabytes of data, and if you're not dealing with petabytes of data, you're most likely dealing with terabytes and high gigabytes of data. And it's just not possible anymore to go find your critical data elements, apply data quality rules, and improve and fix the data pipelines that are producing this 40 data. And so if you went back to the ADM premise, what you can do, and I'm tying these two things together, which is that you can actually use the capabilities on the Acceldata

platform with your favorite LLM, and we will help you identify, sift, and sort through the metadata that exists in the Acceldata platform and take you to the most important data assets, tell you what's wrong with it, and tell you how to fix it.

00:09:20   And what it does as an outcome is that you as a CMO are allocating your budget much, much better than in a non-Agentic world. Another way of thinking about it, and if I would just summarize, you've had this backlog of bad data for at least 20 years, and what have you been able to do with it? Not much, because the problem is that manual effort on this is extremely tedious. And so the promise of AI, which is extreme automation and extreme convenience is actually now coming to the data management space. We're super excited about that.

Jon Krohn: 00:09:53   That's a great example with the zip codes. It makes it very easy to understand these kinds of capabilities. And you use the term, I hadn't used it yet, the short form of agentic data management ADM, and so you'll probably hear us say that term ADM a number of times through this episode. It's clear. So when a CMO is doing this, is it possible, does Acceldata allow kind of a CMO who doesn't write code to be able to use the agentic data management platform themself?

Rohit C.:   00:10:19   Well, the beauty of this whole thing, Jon, at the moment, is that code is no longer a static artifact anymore, and it's not an assumption that code can no longer be written. So the ADM platform has two capabilities. One is that obviously it allows you to go prompt and write these complex workflows, which we completely support. But the second, which is the more advanced users, and obviously as you're going towards fingers on keyboard, what you'll find out is that there's always this thing that, can I just build it as opposed to using a drag and drop interface? And this is something that Ashwin personally is extremely passionate about, and we as a company believe that

Fingers and Keyboard will win. So if you have a really complex scenario, that complex scenario can actually be now generated and that code can be deployed within the ADM platform.

00:11:08   You can actually go and make sure that the use case that you had in mind can be not only expressed as workflow or through a drag and drop interface or through promise, but also through code. So you have the extremes, which is if you're a business user, you only like plain English, you're very welcome, but if you'd like to see some generated code which is doing work for you, absolutely will support that as well. And I think going forward, this is going to be the paradigm, not just for generation of code, but also maintenance of production systems and of a wide variety of use cases within the data management sphere, which today are represented through multiple products.

Jon Krohn: 00:11:48   Nice. And a key part, you mentioning Ashwin there reminded me that a key part of the magic that you offer at Acceldata is allowing this kind of agentic capability over huge amounts of data, petabytes of data very efficiently, very cost effectively. And so a big part of the episode that we had with him, episode 957 talked about that technically. You have made a strong claim in the past, Rohit, that the cost of fixing data when it's ready for consumption is almost a thousand times more expensive than if you can fix the data as it enters your system. And you said that Acceldata is the only product that monitors the entire supply chain of data inside the enterprise. So that way you're getting things when the data are being created or entering your system, which is way more inexpensive in your view. So explain to us why it's so much more inexpensive to fix your data as it's coming in and how the ADM platform does that.

Rohit C.:   00:12:46   It's a very simple way of thinking about it. Let's say that I'm driving on 101, God forbid, I go and crash. The cost of

the crash is extremely high. I've got to call the insurance guy and then I've got to call the cops and then I've got to pay for whoever I crashed into. And so the cost of that is extremely expensive. When you are producing bad data, you're not just creating inconvenience for yourself. You're actually inviting regulators, you're supplying bad quality of decision making to your customers and your stakeholders. And on top of that, you're actually paralyzed because the next set of innovation is completely contingent on the availability of your time. And it's very hard to find time slots where you can go and innovate. Instead, you're spending most of the times in dealing with that kind of crash. And therefore the cost of breaking, which is taking cognizance of the cars that are ahead of anyone driving half a mile or a mile out is much easier and much lower than the cost of crash.

00:13:44   And that's very similar in data. Data essentially keeps getting transformed every step of the way along the data pipeline as it heads from the point of consumption, or sorry, the other way around, the point or the source where it is generated to the point of consumption. And that's a complex journey. And that journey touches data in many different ways through transformations, through copying, through just routine delivery across streaming, batch, and many other different methods of data transformation and movement. But the problem is that every step has changed or modified something. And if you're unable to fix it as it is getting transformed as it is moving, then the challenge is you've left the problem too late and you discover that only when a customer's complaining about the data that you're delivering to him or her and that she's not happy with the quality of data and therefore she's not going to be happy with the quality of the outcomes.

00:14:38   Now, if you were to sort of take an organizational perspective on that, 70 to 75% of data engineering effort today is actually not in modernization. It is not on moving

to the next best thing that exists in the data world. It's actually about just going and solving the old problems.

Jon Krohn: 00:14:55   Right. Yes. And speaking of solving real world problems, a big part of your mission, according to a number of quotes that we found from you, is that your mission is to make enterprise data AI ready so that it's ready for use. How do you know when your data are AI ready and how does Acceldata help you get there?

Rohit C.:   00:15:15   So a few things that breaks down into multiple dimensions, and let's talk about that. So what are the different kinds of data sources? So today you're getting bombarded by data which is coming out of system of records, which is employee database, HR databases, finance databases, customer interactions such as ATM transactions if you're a bank, credit card transactions, if you provide credit cards, let's assume that you're getting all of that data. In addition to that, you're getting credit worthiness reports of consumers, and you're putting all of that together in a package. You're harmonizing that and you're saying, "Okay, all right, I'm going to make some assumptions and inferences based on the data sets that we have so that we may be able to offer better loan guarantees schemes, et cetera, to our customer base." Now, assuming that Jon's credit report had a 690 as opposed to a 760, but he was offered the loans in the terms that applied to 760.

                  00:16:14   I don't think, Jon, you're a 690, but let's assume for a moment. I imagine that that-

Jon Krohn: 00:16:19   You've seen the data, you know too much around. I

Rohit C.:   00:16:22   Haven't. I haven't. But just imagine that you just got a bad deal and most likely you won't accept it, and that is the real challenge. Now, as you're making these decisions, as agents in the future will make these decisions of underwriting, off loan processing, off loan providing, they

will have to have access to highly accurate data as opposed to the differences between the real credit and the assumed credit score that is in the database. So how it breaks down is that there are a few parameters and a few vectors along which you have to test for data readiness. Number one, it has to be technically accurate. The data that is stored should comply with what was expected, which is numbers must be numbers, strings must be strings, addresses must be addresses, and that's simple. I think in this day and age, a lot of companies are getting to a place where the technical accuracy of that data is more or less guaranteed.

00:17:20 And I think we've seen a lot of progress in the last five years. Let's talk about the next stuff, which is the business context in which that data can be considered ready or not. Now, the reality is that the context in an enterprise is stretched across structured and unstructured data, and also in the data that lies inside the semi-structured and the structured data. Now, what it means is that the business context is hidden in the documents and your policies and how you operate as a business, the regulators that you report to. But the real facts are still placed in the semi-structured and the structured databases which power all your decisions, which effectively means that you have to marry the context of the business and ensure that the facts are compliant with those policies, business procedures, and everything else. And when you put these two things together, obviously with the verifiability of the data as it moves through these complex data processes, pipelines, processing steps, then you're fine, but it is a little bit of an effort right now for enterprises, which is where ADM kicks in and it says, "Okay, I'm going to take you to the most important data sets.

00:18:30 I'm going to recommend the rules for you. I'm going to apply them. I'm going to execute that on this data plane and I'm going to tell you what the remediation steps are. "

And so it's super interesting as to how this entire process is getting automated. Do we see a world in which the availability of AI-ready datasets is going to get accelerated even further? I think what's going to happen is that the backlog of data, which could not be used for machine learning analytics in the past will be available for machine learning, analytics, and AI in the next couple of years.

Jon Krohn: 00:19:07   Right. So what you're saying is historically, most enterprise data is unstructured. It's just kind of raw documents, PDF files, Excel spreadsheets without any kind of clear structure, and that makes it difficult for machine learning algorithms to be trained on those data, for AI systems to operate over those data. And so by having agents and workflows that are more intelligent, you can be structuring all those unstructured data into something structured that then we can train machine learning models on and so on.

Rohit C.:   00:19:41   Absolutely. So you vectorize the business context and then you use your data lakes and data warehouses to provide the facts. And so once you have the context of the business in a vector store, and then you have the data which is available in vast quantities in your data warehouse data lakes and system of records, you can actually go and identify the entire business. And that's an incredibly fascinating place to be.

Jon Krohn: 00:20:02   Yes, fascinating indeed. And I'm sure a lot of people are getting excited about what you're describing here where they can finally make use of these vast amounts of data that they have in their data lakes just sitting there when they know there's some business value in there. When we think about the amount of data that we have available to us increasing so much, how do you suggest that your clients rethink the role of human judgment in data operations when now it's impossible with the amounts of

data that we have for a human to be in the loop on every data flow?

Rohit C.: 00:20:39 So I think it's a very exciting time in the world in general and in AI in specific. I think what has happened with the build time of all software processes is incredible. I think everybody knows the success of Codex, Anthropic Cloud Code, all of that. I think the most important distinction I would make is that runtime is different than execution time, and I'll explain how. When you are in the process of building, you have the opportunity of being able to go rectify, correct, alter mistakes that you make or change specifications as you please. But once you're in runtime, you have two kinds of users, and I'm talking about the future, I'm not talking about just now. You have human users who are actually coming and doing things as they are trained to be using regular UX, but the new mode of interaction is also going to be agentic.

00:21:35 And when you put both of those together, the systems that we are now talking about will have to cater to both of these kinds of users. So in that world where humans and agents will use access data for both things, for getting into the agentic flows, making interactions with each other, and finally making judgment calls, I think human judgment will continue to be important in the runtime. Unlike the build time, the build time, the developer is already there on the machine. He's verifying everything that's going on, and in some cases, agents are verifying that. So there is a world in the future where I foresee that there will be agents that humans will manage who will go and verify the work that other agents have gone and produced. It's a very interesting world. It's a dream within a dream. It's inception at the moment, but I think what is going to happen is that these agents will end up speeding up a lot of mundane work.

00:22:33 And what will end up coming to the humans in the loop is the ability and the opportunity of making high quality

decisions. And I can give you a couple of examples. You don't have to go and identify the alerts that you should be paying attention to and what is alert noise. And when you get the alert, there are a lot of existing systems which go and do the alert sifting and they tell you which alerts to pay attention to and what not to. But I think the world that we are going to go look for, and I'm looking forward to that, is that which of the alerts that I should be fixing right away and which of the alerts that I have time? But not just that, give me the fix which I can then go and deploy in production and sort a lot of issues.

00:23:17 So the speed and the velocity of execution will increase. Now, one of the things I've been talking about for the last few months now is the cognitive overload of these systems, producing so much information at such rapid pace is something that we'll have to get used to. We as humans are not used to this mode of communication. And I think while there will be agents which will keep assisting us, I think we will also have to develop mechanisms through software, which will take us to the most important things upfront as opposed to asking us to sift through all the details.

Jon Krohn: 00:23:53 Right. And that makes a lot of sense. We've picked up on other interviews where you've given where the whole thinking shifts around how you need to build your data systems, that the governance needs to be completely different, that you can't just have a one-time governance initiative that's fixed and is going to be lasting for years. Instead, you've said that we need active governance to be watching the system and figuring out where we can be fully automated versus where we need to still have judgment made by humans.

Rohit C.: 00:24:29 Absolutely. I mean, applying regulation and governance post data is already ready for consumption is not the best idea in this world because agents are going to apply themselves and extract information from along the

lifecycle, from the point of origin to the point of consumption, agents are going to make queries and access data for their own purposes and benefits across this pipeline. And so if you think of governance only being limited to the end, to the very end, which is data which is ready for consumption, I think it's already too late for factors that we talked earlier, but also because agents are no longer limited to the consumption data, which was known through the medallion architecture traditionally for human consumption, that's changing. And so therefore the governance which is required is actually more operational in nature as opposed to a process. It's no longer a regulation/compliance need.

00:25:21   It's actually a company operability operations need and therefore governance has to change. It is very, very different from the governance that was applied for many, many years ago. And I'll just break it down for you. If you entered a data governance officer today, maybe two years ago, maybe the view is shifting pretty dramatically right now, but two years ago, every single governance unit or team in any significant enterprise would go with the following stack. Let's go get a data catalog. Let's find ways of doing master data management. Let's go then find out a data quality tool. The problem is that each of these companies, they were operating off a substrate of metadata that was disconnected and that disconnected metadata was fixed by this whole new paradigm of data observability. So data observability provided us the capability to have metadata, which was not just from the consumption layer, but all the way from where data was getting produced to all the way, how it was being shifted, moved, translated, and finally how it was ready for consumption.

00:26:27   Now, on top of that substrate of metadata and access to high quality of data, now it is possible to do governance in a completely different way, which is real time, which is model centric, which is inference centric because

inferences are not waiting for two months. Inferences will happen when the agents are able to glean information across your pipeline as opposed to just the end. And so I think there's going to be a massive shift in the way people think about data governance. It's going to become operational first. The second thing, which is an even bigger trend, is many of these tools that I talked about earlier, they were owned by different groups, some with technology, some with business. There's going to be a big convergence of personas because as the tools converge to the substrate of high quality of metadata and the workflows on top, what you will end up finding is that different governance groups which are using different products and tools today will rationalize those tools and say, "All right, we're going to build these agentic flows on this platform, on this upstate of metadata." And I think that's a big shift, which is coming in the next five years, particularly when humans and agents have to get access to the same data.

Jon Krohn: 00:27:38 Speaking of organizations trying to do things the old way when it doesn't work anymore, you've also stated previously that enterprises are trying to retrofit the previous generation of application monitoring platforms to monitor their data platforms and that that paradigm doesn't work at all.

Rohit C.: 00:27:56 Yeah, that paradigm doesn't work at all. And I'll just break it down. It's fairly simple to understand. Applications were built for human interaction to monitor clicks, to monitor browser strokes, click strokes, experiences, and the interaction between microservices and the information that it was extracting through crop databases. Now, given that there are some hyperscaling or hyperscaler like companies like Meta, Google, Facebook, they have completely different requirement off their relational databases. But if you go to a typical enterprise, they probably have hundreds of thousands of users, in some cases, high million users, and that is not

actually creating the problem. The real problem is getting created after the fact when petabytes of data flow into data lakes every day. The customer conversations that we have are as follows. We have thousand plus applications, IT applications in general, which span different domains, different countries, different lines of businesses.

00:28:58   All of that data needs to be harmonized inside data lakes and data warehouses. And when you think of that world, the application monitoring paradigm was built for the system of records and the thousand IT applications that get deployed, including microservices, monolithic services, et cetera. It has nothing to do with how data is moving, migrating, translating inside data lakes and data warehouses. The consumption unit and the users who consume this data, the kinds of processing that happens today across CPUs and GPUs, they have nothing to do with a primary CPU-based architecture, which essentially is meant to store and provide back the information when it is queried. These are completely different paradigms, and therefore the APM world just doesn't apply to this modern new data world.

Jon Krohn: 00:29:46   And speaking of things moving quickly and past paradigms no longer being relevant, while reading your blog posts and watching your presentations and interviews as we were doing research for this episode, it was apparent that your vision for Acceldata transcends classic terms like IT, data platforms, infrastructure and pipelines as passive tools. It's more about the enterprises, teams and individuals who will be active participants in an intelligent operating environment where data, AI, human judgment converge to drive decisions and outcomes. In fact, we've got a blog post that I'll put in the show notes. It's called Convergence of Personas: How AI is reshaping data management functions. And in that, you wrote that AI is eroding clear boundaries, collapsing personas into a more fluid dynamic model where functions blend and expertise shifts. How should

organizations rethink career paths and incentives to ensure depth of expertise isn't lost, but we have that kind of flexibility anyway?

Rohit C.:   00:30:49   I think there's a lot of value in what AI is bringing, and there's a lot of value that humans will bring to these AI systems. There's a period of transition right now. And the transition is from going from completely human-centric systems, designs that were built only for humans, to a world in which this is going to be a collaborative world where agents and humans will have to work together. It's a reality. I think people will have to wake up to that. I think OpenClaw has been a fascinating moment for all of the AI world, and I'm actually pretty excited about seeing where this world goes from here on, because there's so much of work that these agents are autonomously going to do and accomplish for these teams. Now, if you think of what that world means, it effectively means that if you are a person who has the capability of doing critical thinking and structured language, you can accomplish a lot.

00:31:42   So the previous paradigm had certain restrictions that you had to be a great programmer, and then on top of it, you would need the business domain and the expertise. Today, what has happened is that the level at which you can interact with these business systems, the GPUs, and the agents which can do work for you, depends upon the clarity of your thinking and your own curiosity. And if you are a person who has a lot of determination of being very clear in what you expect the output to be, along with the expertise that you know your industry better than the others, you're more likely and most likely to produce outcomes and systems which are way better than some of your competitors peers will produce. In that world, I think one will have to prioritize the presence of these skills within individuals who possess both of those things, creative thinking and a lot of structure, clarity of thought.

00:32:39   And I think that is the only way that organizations will progress further. In terms of the value of domain, like I mentioned, there's this period of interim transition between human-centric to agentic plus human-centric world. And in that world, I think a lot of domain expertise is going to get used to train internal agents to fine-tune LLMs for LLMs or SLMs to suit the purposes of your own organization, and those skills will then be required in the future.

Jon Krohn: 00:33:10   Nice. You used a term there, SLM. I'm sure all of our listeners are familiar with LLM, large language model. SLM, probably most listeners are familiar with that as well, but small language model. And this is a really exciting area where you don't necessarily need to have these big … You think about Claude Opus. You don't need to have a Claude Opus size model running for every kind of task. I've had a lot of success in doing consulting for enterprises, or in startups that I've been a part of, fine-tuning very small, say, open-source LAMA models that just have a few billion parameters. They can become very, very good at a narrow task when they have a high quality training data. You wouldn't ask them to do anything, but for that one task or this relatively narrow set of tasks that they're fine-tuned for, they can excel.

Rohit C.:   00:33:58   100%. I think SLMs and LLMs will be part of the enterprise stack. I think as will CPUs and GPUs and ASIC, it's going to be an XPU architecture with SLMs and LLMs. And I think a lot of IP will start getting embedded in the SLMs as well, obviously because of privacy concerns and everything else. And if you were to just think and just abstract it out, I think AI is breaking the centralization model. When I think about this whole world, for the last 10, 12 years, the whole thing was about centrality of let's get all your data together into a data lake, into a warehouse, into a location, into a cloud on- premise environment. Whatever's your preference? I think there isn't enough time for such large-scale

migrations to take place. I think AI models will have to get closer to the data where it resides and operate on that to provide the enterprise outcome that people need.

00:34:55   I don't think the speed of AI matches the physical realities of large-scale migration and centrality. So it is going to be a decentralized world in the future. And if you just extend or extrapolate that argument, it's hard for me to see that only LLMs will win. I think LLMs plus SLMs will win.

Jon Krohn: 00:35:15   I agree a hundred percent. And the frontier LLMs from proprietary providers, they allow us a very easy playground to experiment with what's possible, but it's not always going to be the best production decision.

Rohit C.:   00:35:28   That's right. It also depends upon what's the complexity of your use case. Are you looking for a full-time PhD researcher or are you just looking for ... And what's the best way to put it? An intern who's willing to work on a few projects for you? It just depends upon what the complexity of your scenario is.

Jon Krohn: 00:35:44   Right. Exactly. For navigating these kinds of complex scenarios, a particular role that a lot of enterprises have latched onto that have made important in their organizations is a CAIO, a chief AI officer. And last year you wrote an article, a chief AI officer won't fix your AI problems, where you criticized organizations that appoint a CAIO without spreading the responsibility across the organization. You argued that AI isn't a standalone initiative. It's a capability that should be woven into every facet of the business, and that over the next few years, CAIOs will likely begin to disappear from organizational charts. And I've actually, I've had guests on in the past who are experts at bringing AI into enterprises, and they've said that the chief AI officer, when one is hired, they often barely last more than a year because they're kind of expected to change everything, but then they're

not given the responsibility or the power to make a lot of change.

Rohit C.:   00:36:48   So that's definitely one part of it. I think there's the other part of it, which is what is the real problem of AI? I think it's a three-part problem. One, absolutely is the availability of the right set of data, which most of the enterprises are still struggling with. I think it's also identification of the most critical use cases and being very clear about what are deterministic outcomes and what are probabilistic outcomes at the moment. Because what ends up happening is that many of the POCs that are currently running, they never make it to production. And if that continues to happen for over a period of time, the overall enterprise disappointment with these AI initiatives just keeps growing. And so while it is extremely easy to see the ChatGPT-like excitement and seeing how much content and tokens can it generate and what answers can it give, the answers that the enterprise needs are not generic in nature.

00:37:49   They're actually more specific in nature. And so the more specific that the AI officers can get the use cases and define and have narrow boundaries for AI projects, I think the faster that they can actually accomplish a lot of change. What I also find, which is the third pillar, which is it is also a question of adoption because AI is a big hit, a roaring hit in consumer, it's a roaring hit in prosumer for companies like CursorCloud, you name it, but where it is kind of not there is in the enterprise right now. And I think I'm going to make a prediction that in the next 12 to 18 months, we'll see massive use cases coming out of the enterprise. And it is because what it requires is that enterprise AI is not a solution to all your problems, but it is a great solution to a lot of your complex problems.

00:38:44   And I think across the enterprise, when you go and look at and ask the leaders, what are you doing today? They are already doing a lot of automation on the front office.

They've done a lot of automation in the middle office, and I think the back office is just there. We saw this with cloud adoption as well, and so they're ringing parallels on this. Initially, it seemed like cloud adoption was great, and then it slowed down, and then it picked up like it never picked up before. So what ends up happening is that certain enterprise harness capabilities that you require from a security, privacy, governance point of view, along with a reference architecture, it evolves, and after that, it's go, go. And I think we are at the cusp of that moment.

Jon Krohn: 00:39:26   Really great perspective there and so easy to understand. Another piece of this puzzle, in addition to having the right people or giving them the right responsibilities, say a chief AI officer, another big issue that enterprise face is having data or not having data, their own proprietary data. And this seems to have a big impact on their competitive positioning. So given that both high quality data and AI will drive large parts of the economy in the future, you've previously made a sharp distinction between companies that are data haves and data have nots, as well as companies that are AI haves and AI have nots. As agentic systems like those that you have in your ADM, in your Agentic data management platform, as those Agentic systems become more powerful, do you see this divide between haves and have nots becoming a long-term structural advantage in most industries if you do have the data and if you do have AI?

Rohit C.:   00:40:26   I think so. I think public markets are a great place to look at. If you look at what has happened in the SaaS economy in the last few weeks, I think everybody's aware of what is going on. One of the perspectives that's evolving, and obviously these change from time to time, but the important thing that I found out from this recent collapse in the SaaS valuations is because the market is actually not willing to price a future AI. It's actually looking at what are the capabilities of AI that you have today and

how soon are you going to be able to generate revenue out of that? And I think that distinction is going to grow wider and wider. And the comparatives are shifting quite dramatically because an anthropic is going from strength to strength, destroying value accrued in different industries and domains by just the announcement of agents, which are expected to do the bulwark of what these industries were supposed to do.

00:41:23   So I think it is a moment in which the AI halves will have the preparedness and the willingness to have AI first revenue versus the AI have nots will unfortunately be on the other side of the equation because either they don't have the comprehensive AI capabilities in their platform or those AI platform capabilities are not being used by their customer base. And I think both of those would be disastrous.

Jon Krohn: 00:41:51   Right, particularly in an environment where the amount of data that these haves have is increasing so rapidly, something that surprised me because I very frequently state that the amount of data ... So when I'm giving a lecture, giving a keynote or whatever, a stat that often I bring up is how the amount of data that we have on the planet that we're storing is about doubling every year, something like every year, year and a half. But at a conference, you said you expected enterprise data in particular, not to grow twofold year over year, but you're seeing now more like four to five X and that in a couple years from now, you expect it'll be seven to 8x. So basically, let's call it 10X. It's kind of like it's exponentially increasing.

Rohit C.:   00:42:38   It is happening. I mean, think of it this way. If there were a million users using your product and they were creating logs, interactions and all of that data for you, imagine what'll happen when you'll have 10 million agents running a mock on your database, on your customer base, on your data. And so the amount of activity that is

going on right now in the enterprise and is foreseeable in the near future, then you're actually looking at data which is going to go through the roof and are we prepared for that? And that is the real question and the real answer that everybody's trying to figure out. Think of it the other way around, and I'll give you a different perspective now, and I'm not talking about activity monitoring. I'm talking about, imagine the number of applications that are expected to go live in the next two years.

00:43:24   You and I today, Jon, can sit here and while we are speaking, we can build applications and deploy it on our favorite platform and it'll take us less than 20 minutes, and it will be an extremely meaningful workflow that our customers might like. And so what ends up happening is that a lot more applications are therefore going to be live, and therefore they're going to capture a lot of data, and therefore that's going to hit your databases, data lakes, and data warehouses. It is going to hit your network. It is going to hit your storage. It is going to hit the speed at which you are able to store and retrieve data. So this is a completely different world, and data therefore is going to go explode, expand. And the question that I keep thinking about is that in this decentralized architecture, do people have a unified control plane through which they can manage their data?

00:44:16   And that is what our mission all along has been that we're going to enable the management of this massive scale data which is coming to all the enterprises today.

Jon Krohn: 00:44:25   Yeah. You're definitely building the right product for today, as well as for the future over there at Acceldata. No question. Let's talk a little bit about how you came to be in this market leading position, market defining category, the leader in agentic data management. From founding a mobile meeting solution and leading engineering at Hortonworks to creating this new product category with

Acceldata, your journey has been driven by a product first customer anchored philosophy. You've in the past described startup building as not a marathon, but like a Ironman triathlon followed by an ultra marathon right after it. And yeah, that it takes six or seven years before a product becomes mainstream. And that whole time you're educating analysts, buyers, users about what it's like. So tell us about that journey from when you founded Acceldata almost eight years ago now to being able to now have created this new product category of agentech data management.

Rohit C.:　　　00:45:33　　Yeah. I think obviously my leanings have always been in terms of products because I think that's the best form of expression that I know. And I appreciate the problems that enterprises face at a very, very deep level having been an operator myself. I understand how difficult and complex it gets to run engineering teams, production support, and what taxes engineers the most. And so I had deep appreciation and empathy for that, which is why we chose this industry because it was closest to the personas that not just I, but the rest of my co-founders also shared. I think the second thing which you touched upon, which is I think equally interesting is what actually goes into building companies. And I think it's super important to understand that obviously products are products, but you also need people and processes along with that. And you have to have the ability to keep your year to the ground and keep listening to your customers.

　　　　　　　　00:46:33　　I had attended a great talk from somebody at BusinessObjects a couple of years ago and on one of his slides, he had this beautiful quote that said, dance with who brung yeah. And I said, okay, that's great because if your customers brought you along and they gave you the first few opportunities, then you might as well just listen to them because their problems are not solved. Our share of wallet for the problems that they have is not yet done. And so therefore you've got to keep listening to what are

they telling you, where the head is at, whether they headed into what is the direction that they're going to take. And the more you can align yourself with the vision of the customers and keep abstracting it for the industries that you intend to operate in, I think the easier it becomes to build products.

00:47:19   I think at this moment in time, the entire thing is about who can build better products, who can serve your customers better, who can maintain relationships with all the customers and your customer base in the best fashion. Can you serve your customers in the way that they expect you to be served? And I think those are key things that go along in building a strong company. Obviously, none of this is possible without people. It's questionable as to how many people do you need going forward, but I think you need the best people with the most amount of clarity and the most amount of curiosity. I think those things go about building a big company. Obviously we are still starting out. The final point I want to make was that yes, it takes a lot of time. I reflect on the journeys that we had at Hortonworks, which was already a built technology we had already tested.

00:48:09   Most of the technology in the open source community had already stamped it as accurate. The same story happened at Kafka and Confluent and the same story we saw with Spark and Databricks. For many, many years, you're almost just building without any real market traction, but once the product comes together, once you've identified, understood, and delivered to the customer expectation, use cases, critical challenges, then it's just a GTM problem. I'm not dismissing that as not being a problem, but that GTM problem does not rise till the time that you didn't get the right product.

Jon Krohn: 00:48:43   I like how you say that you're just getting started when you've been doing this for eight years and already raised over a hundred million dollars in venture capital or about

a hundred million dollars in venture capital. That's pretty wild. And yeah, it shows your big vision for what you're doing.

Rohit C.:   00:48:58   Jon, 100 million these days are being spread on seed rounds, just to let you know.

Jon Krohn: 00:49:04   Yeah, but not very often. I mean, that does happen, and that also, that really only happens in the Bay Area. I think you're around seeing some of that. It's a hundred million dollar seed round anywhere else in the world, I think it's still unheard of. I know. But yeah, you guys are doing a phenomenal job over there getting everything going. Speaking of having the right people, it sounded like a big part of how you are able to have so much success at Acceldata. In a tweet, you wrote that products with global appeal are not short on capital. They can get a hundred million dollars to round. And eyes are on those teams, founders, products, where it's the imagination of different, better, and when articulated well enough, the money, the teams, the customers, they recruit themselves into the mission. Tell us about that, because it sounds like there's a lot of useful insights there for any of our listeners who are trying to build a product or build a company.

Rohit C.:   00:50:12   When I go back to my In Mobi days where I was a founding engineer, what I found out that the team that assembled together at Inmobi, and this is way back in 2008, they had the need of expressing themselves through high quality engineering work, which was then reflected in the products that we built. And so if you think that you want to build the best team, then you've got to get a mission that aligns some of the people who care about that mission. And it's therefore at that moment self-selecting. I mean, the amount of ours that we work in startups, it's insane. And a lot of people have that as the most important aspect of their life. Obviously, everything else is important, but for them, the ability to go and

create something out of nothing matters more than anything else. And so there are people who actually come and join startups at different phases.

00:51:02   There are people who are experts at getting zero to one, and then one to 10, and 10 to 100, and 100 and beyond. And these are different kinds of people. But if you are a company which is continuously doing zero to ones, one to 10s, 10 to 100s, then obviously you have a lot of space for people who can join you at any part of that journey, because if you're an innovative company, you're going to keep building newer products, which gives the ability for the creatives to come and express themselves. And then you give the operators the ability to go and manage the products that are coming through the pipeline every year, every couple of years. I think the second most important thing is that what is your personality there? Do you believe that you have the need to build something or do you have the need to manage something?

00:51:45   And that both of those options are great, but those are two different options. And therefore it depends upon what you want to select and what outcomes do you want from it from a life perspective. I think the third thing which we have emphasized on is that we have values. Now, we have global offices and therefore culture in all these offices are local to those centers, but the values are global. And the values for us are that we are builders at heart. We're going to be extremely genuine, respectful, and humble about our interactions and simple about our interactions with our customers, with our internal employees, with our investors, with any other stakeholder that we find either internally or externally. And if you align with those values, then you're a great fit. Obviously, we also pay a lot of attention on your craftsmanship. Do you care about the work that you're putting out?

00:52:38   Would you actually take it very personally if the work that you produced had some critique and comment, or really

just look at it and say, "Okay, here's an opportunity that I'm going to just iterate on it and make it 10 times better." And if you look for these kind of mission-driven people who care a lot about the quality of the work that they put out and the kind of teams that they would like to work in, you end up finding the right people.

Jon Krohn: 00:53:01   Nice. Beautifully said. And yeah, it sounds like going back to a conversation we had right at the beginning of this episode, that if these are people that you already have worked with before and you trust, there can be fewer bumps in the road potentially as you go and build a product company together. Really cool. Rohit, we're starting to wrap up the episode. There's something that I was supposed to tell you before we started recording. So usually I tell my guests in advance that at the end of the episode, I always ask for a book recommendation, but I forgot to tell you that. So I don't know how well prepared you are on that question, but do you happen to have a book recommendation for us, Rohit?

Rohit C.:   00:53:46   Yeah, sure. I mean, one of the books that I really liked was, there are several. Can I make three?

Jon Krohn: 00:53:52   You can make three Rohit.

Rohit C.:   00:53:54   Zero to one continues to be a great book. It's a great read for no matter what stage of your life you are at, it's an awesome book. And the second would be the beginning of Infinity, which I have come to appreciate a lot. It's a fascinating book, and if you're interested in things in abundance, in the concept of abundance, I think it's a great book.

Jon Krohn: 00:54:15   And so when you say that, so Zero to One is a book that I'm familiar with. I do like it a lot as well. It's got great insights on, yes, life in general, but also building a startup, building an AI product, especially now today. It was remarkably forward looking given that it must be

about a decade old, that book, Zero to One. This book, Beginning of Infinity, that you just started describing. So I thought maybe it was kind of about astronomy, but it sounds like it's about having abundance on our planet thanks to AI and automation and abundant energy and all these kinds of things.

Rohit C.:   00:54:53   No, I think it talks about abundance. It's also about the creation of knowledge, and it also argues that all kinds of progress, whether it's scientific, societal, technological, it stems from a relentless, rational pursuit of reasonable explanations, which are consistent over long periods of time. And David Deusch, he proposes that this idea of knowledge creation is always at the beginning. It's never at the end, which means that if physics was at a certain place in the 16th century, it is very, very different in the 21st century, and it's got to be very different in five centuries from now, but it's just the beginning of the process. And I think that's a fascinating concept and it helps me to have a beginner's mindset all the time that you can just forget about the work that you did until yesterday. You can build on that, but the reality is today.

                            00:55:52   So anybody, and this especially applies in the age of AI, all of the things that we learned in the last 50 years, I think you're just saying, "Okay, that was a good foundation, but what's next?" And I think that's super interesting.

Jon Krohn: 00:56:04   Sorry to quickly interrupt you before you give us your third book recommendation, but you talked about beginner's mind there. Are you a meditator, Rohit?

Rohit C.:   00:56:10   I am, yeah. I

Jon Krohn: 00:56:12   Am.

Rohit C.:   00:56:13   I've been practicing yoga for the last 10 years and it's a great practice, highly recommended. It just cleans your

mind, takes the rest of you of your day-to-day friction that you encounter in your life, whether it's traffic, active work with customers. It's just a great leg there.

Jon Krohn: 00:56:29   Yeah, yeah. Well, actually, it shows in the way that you conduct yourself in this interview from even before we started recording, listeners and viewers couldn't possibly know this because it happened before the episode formally started, but we took a little break before we got going. And even just the way you handled yourself in that, there was this sense of calm and ease that came across, and I just knew that this was going to be such a great episode because of that. And so I came back, sat down, was so excited to start recording. So it's working.

Rohit C.:   00:57:04   No, I appreciate it. Thank you. I highly recommend it.

Jon Krohn: 00:57:07   Nice. All right. And yeah, your third recommendation, Rohit, your third book, Rec?

Rohit C.:   00:57:10   There are a few, but I'm going to go with the gospel of Ramakrishna. It's a deeply spiritual book. It actually moved some of my thinking, and I know it's not a very popular opinion at this point in time to talk about spirituality, but I think it's the most popular opinion in terms of figuring out what gives you satisfaction and what gives you contentment in life.

Jon Krohn: 00:57:34   Nice. I like that a lot. Yeah. I've read a number of translations of the Baghdad Gita, and those have been pretty influential on me as an adult. And yeah, so this kind of thing I am into. And so Gospel of Ramakrishna sounds interesting. Maybe I'll check that out as well. Well, Rohit, thank you so much for the sensational episode. For people who want to be getting more of your ideas and thoughts after this podcast episode, how should they follow you in Acceldata?

Rohit C.:	00:58:09	Well, Acceldata is on LinkedIn and on Twitter, and I'm RC online, both on LinkedIn and on Twitter. Most of my thoughts are posted on LinkedIn these days. That's where most of my professional circle is, and I care a lot about that network, so best to connect with me on LinkedIn.
Jon Krohn:	00:58:27	Nice. Yeah, that is where most of our guests are converging over the past couple of years, for sure. It's funny how LinkedIn strength is that it doesn't change. It's just so familiar. It's so funny. Social media platforms, it's like they're all … I guess like most product companies, they're constantly innovating, constantly changing, but it seems like that means a lot of the platforms that were popular in the past have kind of faded. Meanwhile, LinkedIn hasn't really changed, and that's turned out to be a strength.
Rohit C.:	00:58:59	Yeah. I mean, I think I don't know what the next generation of LinkedIn will be. Will there be a social network for agents who are doing work that we are doing today? Will they require … The whole moldbook thing was so fascinating and obviously now OpenClaw, but I don't know. Maybe LinkedIn is due to change. Maybe there's a disruption coming.
Jon Krohn:	00:59:19	I don't know. If we have any listeners making product decisions at LinkedIn never change, just stay the way you are. It's perfect. Okay. Yeah. Great. Great, great, great. Great episode, Rohit. Thank you so much for being on the show. You have such a busy schedule, so to make the time for us, it really means a lot. We appreciate it.
Rohit C.:	00:59:36	Of course, Jon. Thanks for having me today.
Jon Krohn:	00:59:39	What a wonderful episode with Rohit Choudhary. In it, he covered how he coined the term data observability in 2018, how fixing bad data at the point of consumption can be roughly a thousand times more expensive than catching and fixing it as it flows through the pipeline, how

for your enterprise data to be AI ready, they need to satisfy multiple dimensions, technical accuracy where a number or numbers and strings or strings, business context compliance where the facts in your structured data actually align with your policies, your procedures, and regulatory requirements. He talked about how he predicts enterprise data will grow four to 5X year over year right now, accelerating to nearly 10X soon, driven largely by the explosion of AI agents generating queries and activity at a scale that dwarfs human users. And he talked about how in the age of AI, the most valuable professionals won't necessarily be the best programmers.

01:00:31   They'll be the ones with the clearest thinking, the deepest domain expertise, and the curiosity to articulate precisely what outcomes they need. As always, you can get all the show notes, including the transcript for this episode, the video recording, any materials mentioned on the show, the URLs for Rohit's social media profiles, as well as my own at superdatascience.com/979. Thanks, of course, to everyone. On the SuperDataScience Podcast team, our podcast manager, Sonja Brajovic, media editor, Mario Pombo, our partnerships team Natalie Ziajski, our researcher, Serg Masís writer, Dr. Zara Karschay, and our founder Kirill Eremenko. Thanks to all of them for producing this super episode for us today, for enabling that super team to create this free podcast for you. We are so grateful to our sponsors. You can support the show listener by checking out our sponsor's links, which are in the show notes. And if you'd ever like to sponsor an episode, you can get the details on how by making your way to jonkrohn.com/podcast.

01:01:25   Otherwise, share this episode with someone that would like to hear it. Review it on your favorite podcasting app. If you write an Apple podcast review, I will read that on air. We really appreciate you doing those. Helps get the word out about the show. Subscribe if you're not already a subscriber, but most importantly, just keep on tuning

in. I'm so grateful to have you listening, and I hope I can continue to make episodes you love for years and years to come. Till next time, keep on rocking it out there, and I'm looking forward to enjoying another round of the SuperDataScience Podcast with you very soon.