

EPISODE

04

Your data isn't perfect. Start anyway.

Guest: John Young, Global Managing Director, Databricks



JOHN YOUNG: But we hear, you know, my data's in terrible shape. I can't really begin this journey of AI, which we want to stand up and say pretty loudly, "Not true." And in fact, AI is really well suited to support that and help in transitioning. You don't have to have perfect data in order to create valuable solutions that customers can benefit from.

VOICEOVER: Outwitting complexity starts here. Welcome to The GuideLine.

STUART BROWN: John, welcome to Guideline. I'm glad to have you here. Really excited about this conversation. I think it's very relevant with our clients. You bring a really deep perspective on tech, data management, AI innovation. You've been around the business, like I have, for a while. So, before we get into all the questions, let me start with you. Can you walk through a little bit of your background and your current role at Databricks and sort of where you see Databricks fitting in there?

JOHN YOUNG: Yeah, thanks Stuart. So, happy to be on GuideLine. Happy to be here with you. I'm just under five years at Databricks. I lead a team in field engineering globally that works with all of our system integration partners. And those could be the very large, that you might all hear and be familiar with: Accenture, Deloitte, all the way to the smallest one- or two-person shops who just love Databricks and are working to, you know, build data and AI solutions with customers. Prior to joining Databricks, I ran the data science and machine learning practice for Hitachi. So have had large teams building data science and machine learning solutions over the past decade or so focused on Databricks.

STUART BROWN: Great. I'm going to pull those threads later. Cause that history of data science and to where we are now with generative is a fun one because I think it's been an interesting conflagration of the two worlds coming together. But let's start with Databricks a little bit. So, some people may not be aware of Databricks, what the breadth and depth is. Obviously, everybody knows the big hyperscalers. They've maybe heard of Databricks, they've heard of Snowflake and others, right? But can you give a little bit of overview of Databricks as a company, as a platform, and sort of how you talk to clients and partners about 'why Databricks.'

JOHN YOUNG: Yeah, so I've been with Data... so Databricks started about fifteen years ago out of a lab in Berkeley. A team of founders were working on a research project around high volume, high velocity data processing, and they created Spark, which is a very high throughput data processing engine. They open sourced that and then they built a platform around that. And that was kind of the genesis of Databricks, thinking, hey, you know, there's got to be a better way to, in this world of machine learning and AI, to process data more cost effectively. And that kind of initiated or instantiated Databricks.

A few years later, we recognized that data warehouses were good for some things, and data lakes were what most data science and machine learning used and leveraged. So, but the two of those lived in sort of different worlds. And we innovated a lot to try to bring those together into this category of the "lakehouse," which back in 2019-2020 was sort of this, you know, no one thought this was a legit thing. Turns out it is and now everyone, hyperscalers, other data platforms, are really understanding the value of lakehouse, which brings the best of data warehousing with the best of data lakes and allows for all of those AI and machine learning as well as the more traditional data warehousing functions to happen.

So, we're now kind of at the, you know, sort of foundation of AI. We've been an innovator in data and we believe, you know, with our kind of open platform that we're well positioned to support generative AI and the advances that come with the AI world.

STUART BROWN: Yeah, I definitely, it's a very interesting and forward-thinking journey that Databricks has been on and getting ahead of the game back in the, as you said, in the lakehouse days. I think back in those days people were like, well, I have a data warehouse, why do we need a lakehouse? Why do we need a data lake? And we started, I remember saying to people, the data warehouses are going to go away and they're, 'No, they can't, I've spent millions and millions of dollars on these things.' Yeah, but they're basically like putting things in the attic. You don't know where it is, you don't really go at it, you don't use it. And if you really think about how the business operates, if you've got trapped value in your organization you can't access, then is the data very, is it really valuable? And as we've seen, the unlock from, I think, generative and now I think the push towards AI, what we're seeing with data is, the data is the fuel of the organization. And I think companies are starting to realize that. So, I want to pull that thread a little bit here and talk about, what you've seen from the innovative companies and what the challenges are for those that have been resisting.

But in that, I think it's, as we think about the challenges of the data, one of the things you talked about was open, right? So, open source and open data. I think where I see things going, especially with the rise of things like OpenClaw and Hermes and other stuff, really sort of this resident intelligence inside my organization, which, you know Microsoft's now adopted with Scout and others. Data residency, the situ of that data matters.

So, how does, when you think about the data advice that you're giving clients and we go together to have those conversations, what is the recommendation out there if you're a CIO, CTO, great, but even more so a CEO, a CMO, CFO around, look, understanding how to think about this data differently and understanding where the world is going from a from a data sovereignty perspective. I'd love to get your and Databricks' thought on that.

JOHN YOUNG: Yeah, I think, look, all companies have always talked about the importance of securing data and governing data and controlling data. Very few companies do it well. I think, as a data scientist, there were, there was like this, sort of, foundation of every good solution comes, it required three things.

Data, an algorithm, sort of like the creativity of an algorithm, and feature, feature engineering. So, not just the raw data, you're always doing something with that data. And those three things, like, data within the enterprise is super, super valuable. And most companies don't really recognize that. They have ambitions about monetizing data, their data, but the way they combine it, how they package it and how they share it, choose to share it, or choose to protect it, is really what differentiates what we see as you know, companies who are taking, who are benefiting from this age of AI, this transition into AI. It's those who are recognizing the value of their data and really getting thoughtful about how they curate and control and then provide that choice to their customers.

STUART BROWN: I think that's really insightful. I think when, cause when I think of data and the interesting thing about it, like the problem is there's so much data in an organization. And when I think about the functions the businesses do, there's almost, there's not many businesses that what they do is one of one, right? So, whether it's meter-to-cash in utilities or, you know, hire-to-retain, whatever it is, right? The processes are generally the same at the algorithm, to your point.

So, there's sort of two things I want to ask here. Number one is I'd like to understand a little bit, sort of, the health of the data, you know, because everybody talks about, well, my data is not clean enough or whatever. And I don't think anybody's going to have perfectly clean data. So, I'd love to understand a little bit, when you're talking to clients about the validity and health of their data, how we have that conversation. And the last piece is how do you segment or tier that data? Because the reality is, you know, I worked a long time in the utilities practice, right? So, the meter-to-cash is meter-to-cash. Everybody reads a meter, everybody sends a bill, everybody gets cash in.

How you apply the processes on top of that is the unique sauce. So, how do you then tier that logic to say, look, the data that we really need to retain and replicate and put across the intelligence and give to tokens is not everything below the water line, it's the stuff above the water line. So, is that a conversation you have with clients, especially as data scientists? And then the validity of that data and the health of that data, how do you, how do you guide a client to say, you can't wait till it's perfect? But how does Databricks and how do you look at that and helping them take that journey?

JOHN YOUNG: Good question. So, one of the things that we hear a lot is, my data's, you know, yes, data's in silos. That is not unique.

STUART BROWN: Shocking.

JOHN YOUNG: Yeah, right. But we hear you know, my data's in terrible shape. I can't really begin this journey of AI, which we want to stand up and say pretty loudly, "Not true." And in fact, AI is really well suited to support that and help in transitioning. You don't have to have perfect data in order to create valuable solutions that customers can benefit from.

That's just a fact. And AI can support filling in the gaps, the quality gaps, identifying the ones that are important to identify and segregate or segment. Decide what you need to do and prioritize those systems that need to get either migrated away from because they're not supporting the long term for you. Or you resolve, you know, some of the quality issues and find out where that source is. And also move forward and build something valuable.

I think, as it relates to, how do you tier and or how our customers are tiering, there's a few things that are emerging, and it has to do with how to find and identify the productivity gains or the productivity opportunities from AI. And that is, right, we got all these build tools. You know, you can build better PowerPoints, you can build cooler images to put in those PowerPoints, you can summarize a lot faster.

There's a lot of activity and personal productivity, which is cool and it's great, but that's not going to drive enterprise value, generally speaking. What we do see is kind of along three lanes. We see, the end-to-end revenue, the pre, if you will, pre-sales, get us to sale. So, revenue per employee. And then what is it that we're doing from lifetime, you know, value or customer lifecycle management. Once we have the customer, how do we manage that customer for its lifetime?

And we start to introduce along the, sort of, horizontal across both of those, this notion of operational simulation, which in, you mentioned, you know, energy, certainly in manufacturing, we see this, the notion of digital twins. "Operational twins." It is very easy for these tools to create simulations that approximate, right, either side of that, you know, revenue. Get to customer, keep the customer.

And so creating operational simulations really allows the companies to focus where they should apply and go after data. And it also informs what, you know, what data sources are viable within your enterprise.

STUART BROWN: I think that's great. I mean, the digital twin example is a great one. So, I did a lot in manufacturing in the early days of digital twins and it was such a boon to them. There were always challenges with plant managers and operators and, you know, we can't shut things down. But getting to that point where you could say, can we get five percent, two percent? And then the efficiency gains were small, but actually big for, especially on the maintenance side of the equipment. And there's no reason you can't put, to your point, the same analogy to an enterprise function and to the goal of the enterprise as a whole and breaking out the silos.

And it's interesting, one of the lines I used to use with boards when I'd be talking about AI, this was back in 2022, '23, I'd be in front of boards of directors and they'd say, well, my CIO tells me our data's bad, so we can't take advantage of this. I'm like, great, so you're making decisions with bad data today and you don't know it. The bad data's still there, and it's being, it's in a silo, so where it's being applied may not have the context of the data being bad.

So, the whole idea in my mind of putting in a platform and an approach like Databricks, and what you guys are talking about is, it gets to the people who understand the context of how that data is being applied and used and goes, 'Something's not right. And I think we need to go look at it,' and allows them to focus very quickly, almost like a smoke tower in a forestry setting, to say this is where we need to go. Because if not, to your point, it is this massive jumble of, you know, especially these big organizations, legacy tech, legacy databases, some of which may be in mainframes, some of which may be in CLI-based systems, you've got all these different, where do I look, where did it fall apart on the journey?

And a lot of the times in the past, we would have spent a ton of time and money and effort to, you know, refactor that data and whatever. But now I think with AI, and love to get your thoughts on this, we can surround it. We can surround that legacy data and have it be part of the intelligence layer without really messing with the source system if we don't want to.

JOHN YOUNG: I totally agree with that. So a couple of things that you've said there have triggered a thought, right? One thing that we see, within Databricks, we've been consuming these tools, our own tools as well as some of the, you know, the frontier tools and helping us build, right? So, we have, you know, initiatives internally.

We call it Databricks on Databricks. It's like, how do you drive AI transformation using Databricks as the platform within Databricks?

STUART BROWN: Yeah.

JOHN YOUNG: And I think what, you know, one thing that we observe is, these tools allow you to create net new very quickly. So, if we thought there was technical debt before in an organization, the amount of technical debt that can be accumulated virtually, you know, literally overnight is kind of astounding. And so that's one thing that we really drove a lot in our internal transformations using Databricks, using our governance tools.

We get to see all of the build that's happening. And we also get insight into how the teams are functioning. Like, are you just building, you know, something over the weekend that you want to post in, you know, on YouTube and say, hey, look what I hacked together?

Or is this really driving, you know, some kind of enterprise value and efficiency? And so Databricks on Databricks has, informed us in some of the cautions of building new technical debt and also kind of cemented the importance of governing and controlling all of those data and AI assets. So, it's a whirlwind, we're still learning a lot every day.

STUART BROWN: Well, I think everybody is. Everybody's building stuff, right? Everybody's, because it's like when we first got PowerPoint and we got fonts. Everybody used every font, right? And I'm that old. So, you know. But you know, everyone used everything they had, right? And it's sort of the same thing.

I mean, I develop stuff all the time with my bots and I'm like, did I really have to develop that? I mean, I probably could have found something that would have been simpler, but the idea is, like, it's so much fun to do. But I like the idea of looking at the measurement of personal productivity or individual or even one-off versus, is there enterprise value in what we've created? And if there is, how do we tap that into the broader construct to get across?

Because I think we're going to see more and more personal agents. I've always said, like, we're going to have, there's going to be a Stu bot that I take with me when I go to other companies. Like, there's any personality that I have that knows how I work and operate and how I like to interact. And then there's the role that I step into. I should step into a role with a level of intelligence that that company deems appropriate for that role. So, if I'm, you know head of technology, if I come in to be the CTO of Databricks, I should walk into an intelligence layer that when I get emails, pings, whatever, I'm getting context with that from how the company has decided this role operates in the past. Now I get to shape it a little bit and what I bring to the table, but having that organizational intelligence staying inside of an, in that algorithm, like you talked about, is really important.

And this is, I think, the promise of where data science wanted to be fifteen, twenty years ago, but frankly it was a really heavy lift.

JOHN YOUNG: I think your concept of personalization, right? Agent personalization is super important. It's what we start to get, we immediately recognize when we're

interacting with our data and, through an agent, is whether they are in tune with us or whether they are sort of, like, whoop, wait, what's happening here? You just got lazy or you went off the rails.

We sort of have a sense, which is why it's super important as you talk about any of these business processes, why human-in-the-loop is so critical to the transformation. You can't just say agents communicate with other agents all the way through. We're not at that place yet. We need humans in the loop, both in design and observation. And that, I think design, and it's just a quick aside, I think design is really the next sort of anchor of what are these systems like?

How are they good, how are they bad, you know, design of the systems, design of workflows is a critical skill. And the concept of jobs going away, I think they're shifting for sure, and shifting to need more and better design.

STUART BROWN: Yeah, I couldn't agree more.

JOHN YOUNG: Because the scale is staggering, right? It's kind of hard to get your head around how fast and how many things can be, you know, operating at the same time. And, just to be, more thoughtful in design and control of that is super critical.

STUART BROWN: I love that and I could not agree more. I think the shifting of jobs, and I think we've started, everyone's starting to realize it's not a net reduction, but it is a shifting. I think that's the right way to put it. And I think the more context and content you have in your field, the more valuable you're going to become because it's going to require the, I like the idea of the design. I mean, we've been thinking about this as we go to clients and we talk, sort of, about our construct of how you look at regulated industries and what we should be looking at and, I'm trying to go in and talk one, three, five years out, even though we're designing for today. What I don't want to have happen, I mean, you and I, we've been through the cloud journeys, we've been through the data warehouse journeys, we've been through the, back in the day, the middleware journeys and all the other stuff.

And I want, like, we know where this is headed. This is going to head to the idea that the data is going to dictate the software, and the software generally is going to be built by us. It doesn't mean we're not going to have software companies that we leverage, but odds are, personal opinion, we're going to be leveraging their data models more than anything else.

And those data models are very valuable to them, and they need to monetize. And I think some of them are and some of them are not. And those that don't, I think will have a very hard, hard road ahead. But I think the idea of, okay, I want to build what my CRM looks like. I want to build what my ERP looks like. And it's going to be based on this process, not on this system. I don't need to buy all the other junk. And so, I think the data becomes very critical. And then how you structure that data.

And I use the term “fungible” all the time. I want my architectures to be as fungible as possible for all the innovation that’s going to be coming over the next 12 to 36 months. And I will be vehemently annoying to a client if I find them going down a path that is going to be restricting them. And even if they’re like, religious argument, I’ll be like, look, I’m going to keep warning you, you get to make the final call, you’re the client. But what you’re about to do is going to lock you into something that you don’t want to be locked into.

And we’ve gotten away from that mostly. I mean we got away from the hardware lock-ins and everything else. Now, having said that, I may flip back and say, hey, I believe this data needs to be on premise, right? As opposed to, I would have said all the data needs to be on cloud in the past. I’m going to say, look, there needs to be some stuff here and some stuff there, especially in my industries. But, you know, you may use this hyperscaler today. The next one, it might be it might be an xAI, it might be on, we don’t know who the next ones are going to be.

JOHN YOUNG: Exactly. Right.

STUART BROWN: And you you’ve got to be able to take advantage of that quickly to stay ahead of your competition, no matter what that competition is, even if you’re a regulated agency, I want to make sure that I’m getting the best intelligence I can. So, I agree with you. I think the idea of the role of the individual in the organization becomes more valuable, but it does require change adoption in an organization that is tough. I did, I’d love to get your thought on this, I was in front of a lot of CEOs and I said, look, the biggest challenge of AI is not going to be the tech. It’s going to be the change in your organization.

And I said it’s the biggest change to hit a company since Y2K. It’s bigger than COVID because I’m not, I mean, we’re about to hit a COVID moment where those that are, you know, in my mind the COVID moment was, if you weren’t remote work enabled, you got smacked in the face really hard and had, like, weeks to catch up, which was very obviously very profitable for the SIs, right? I think we’re in the same thing in AI. If you’ve been dragging your feet and you don’t even have a data lake, you’re still sitting on a data warehouse, you can’t take advantage of all of this, and you’re going to have to do it either quickly or you’re going to be in deep trouble.

I think that same thing exists. So, if this, if there are executives out there listening to this or regulators, agencies, if you don’t have a data lake, if you’re not looking at a data lakehouse, if you’re not thinking about this right now, if you don’t have a strategy at least, you’re behind. Right? You’re behind. And it’s going to cost you more money to catch up.

JOHN YOUNG: Yeah, I think, I would say, I would add one piece to that. I would say, look, you’re behind, and you are advantaged because you can skip the line, right?

STUART BROWN: Correct. Agreed. Agreed.

JOHN YOUNG: You can, please do not just keep waiting because it's not going to settle anytime soon, and you could literally be out of business before you, and you won't see it coming. Like, there are things that are happening so fast that, you know, one day you think you're cruising and the next you're in deep trouble because you just got ate by software.

The good news is you can skip the line. To your point, skip the line to open, to not be locked in infrastructure, you know, infrastructure plus software, which is really your previous choices. It's like, pick this cloud, pick that cloud, pick that cloud, and you're stuck into the whole stack.

That is not the case anymore. Databricks is a perfect example. We run on all clouds. We have the software and we can consume all models, like, all through this. And we're not charging customers for that access, for that control. We're saying, look, the control should sit with you for your data, for your AI. So, grab it. Come with us. Skip the line, jump in, grab it.

Take control, start building things of value that drive, you know, customer engagement, that drive top line revenue and operational efficiency with your own data. Because 'with your own data' is the unique value prop that you come to the table with. So, I think you're spot on with that, Stuart.

STUART BROWN: I love the point around you can skip the line. I had this when I was running a big cloud practice in another place, right? I mean, it was especially in the oil and gas / energy space, right? They were the last to go. And it was interesting, the very last to go in the cloud space were healthcare and financial services. Like they were kicking and screaming. We can't do it because regulatory, PII, HIPAA, whatever. First to go to AI.

Right. And I think because they were so regulated in process, they were able to take advantage early on because they were so process oriented. But it does get to that point of, you can skip the line and you should be doing it. And so I say that a lot, especially when, you can imagine, I'm talking to government agencies. I'm like, hey, it's okay that you haven't, that you don't have all this other crap we got to unwind that you haven't already capitalized out. But in order to get this done, this is going to require organizational leadership, we're doing this.

And so, like, in the federal government, I'm seeing that now. There's a big shift in obviously the technology organizations where there's a big push in a short amount of time to really, frankly, make up twenty years of non-innovation, which is great. State and local, I haven't seen that at the same level yet that I'm seeing in the federal level, but I think the idea of saying, do you have the change agent in your organization and do you have the decision authority to make this happen because if you don't, even in a commercial organization, it's challenging.

It's even more difficult in the governmental organizations. But I agree that you can make leaps and bounds and not have to unwind, look the leading companies, they're fine unwinding spend. Banks, I don't care, these are companies that innovate on the edge all the time. So, their idea of unwinding something put in two, three years ago is not a big deal. It is a big deal when you're talking about taxpayer dollars, or you're talking about, you know, regulated, you know, regulated monopolistic industries that are regulated by the government. You can't spend willy-nilly on that, right?

JOHN YOUNG: Yeah, you know, I love your point about change agent, right? And being that leader. Like one of the things that I look for when we are meeting with partners and customers is those inspirational people, right? Like, I get it. Companies, people, it comes down to individuals in their heart. They've invested a lot in a company.

They've put their careers on the line. They've made decisions that their careers rest on. Hey, we just sold this big thing to the board. We haven't reaped the value of it yet. If I come in and say, hey, we got to do something different, they're going, I'm running a lot of risk. And what I would say to that is, the next generation is looking for inspiration here. It's looking for inspiring leaders. And I think whether you've made a big commitment and submitted, you know, a lot of personal chips on the table with your company, or you are looking for that opportunity, like be inspirational, inspirational to the teams that are under, you know, that that that are working with you and for you, and really embrace what's out there, I think, because the change agent is super critical to all of this.

We can come in and say, hey, whether it's a forward deployed engineer or, whatever the current term of your great builder is, we can build stuff. But change, like getting your org to change is what's critical. And we need inspirational leaders as much as we need powerful technology.

STUART BROWN: Yeah, I agree. And I do the same thing. When I go into a room, my first thing is, I'm like, are the people advocating for this, how long have they been enrolled? And if they've been there fifteen, twenty years and, is this, do they finally feel like they've got the, like I basically read them out. Are they here because finally someone listened to them and they get to do what they want to do and the others didn't? And is there a leader above them that says we're going the same direction? Or are they just checking the box?

If they're just checking the box, I'm very much like, look, I hear what you want to do, but unless you have this kind of authority, like unless you really want to get it done, you can't do these kind of things by committee, especially with the leapfrog like we talked about.

Well, I think we're at the end, John. So, I'm going to ask you the one question I ask everybody, which is, you know, what's your guideline? So, one final insight or piece of advice that you could leave our listeners with today around sort of what your guideline is and what they should be listening, what they should be thinking.

JOHN YOUNG: I sort of, so maybe I'll bring forward a couple of thoughts that we shared already, right? One is, when you can, skip the line. So, I think that's an important thing. Don't feel like you are mired or locked into something. Explore some of these leading technologies like Databricks and the innovations that come with Databricks or our platform or others, like, jump into this.

Jump ahead and don't feel you're locked in, and I think anchor on, for sure, anchor on the ability to control data and AI assets as you are building. Do that in a place that allows you this openness to pick infrastructure, to pick software, because that is going to allow you to go fast.

The open technologies move faster than closed. The open enterprises move faster than closed. It's proven out. And that'd be my advice. Just don't be afraid to go fast.

STUART BROWN: I love it. Skip the line and be open to open. I think it's the right way to look at it. Yeah, I love it. Well, John, first of all, thank you. I know you're on PTO day, so thank you for joining us. I really appreciate it. I think it's a great conversation. I think people will get a lot out of this. I think, you know, data is, we used to talk about it as being sort of the brain. It's the brain and the heart of organizations now. And it's more and more important that we get this right.

John, thanks for your time. I appreciate it. Thanks for joining GuideLine. I appreciate it.

JOHN YOUNG: Thanks so much, Stuart. I really enjoyed this.

The GuideLine is a podcast brought to you by Guidehouse. [Subscribe](#) wherever you listen to podcasts so you never miss an episode. To see more from Guidehouse, head to our website, [guidehouse.com](https://www.guidehouse.com).

About Guidehouse

Guidehouse is a global professional services firm delivering advisory, technology, and managed services to the commercial and government sectors. With an integrated business technology approach, Guidehouse drives efficiency and resilience in the healthcare, financial services, energy, infrastructure, and national security markets. [guidehouse.com](https://www.guidehouse.com)

© 2026 Guidehouse Inc. All rights reserved. This content is for general information purposes only and should not be used as a substitute for consultation with professional advisors.