EPISODE 1340

[INTRODUCTION]

[00:00:00] GM: Modern companies leverage dozens or even hundreds of software solutions to solve specific needs of their businesses. Organizations need to collect all of these disparate data sources into a data warehouse in order to add value. The raw data typically needs transformation before it can be analyzed. In many cases, companies develop homegrown solutions. Thus, reinventing the wheel and possibly planting deep rooted seeds of technical debt.

Mozart Data helps you collect all of your data sources in under an hour. They provide managed data pipelines, data warehousing and transformation automation. In this episode, I interview cofounders CEO Peter Fishman, and CTO Dan Silberman about the modern data stack.

[INTERVIEW]

[00:00:47] GM: Dan and Peter, welcome to Software Engineering Daily.

[00:00:53] PF: Great to be here.

[00:00:55] GM: Well, to kick things off, tell me a little bit about how the two of you guys got connected. How did you first meet?

[00:01:01] PF: Well, Dan and I met over 20 years ago when we were both in college. Dan's best friend in high school was my college roommate. So on a trip down to Duke, Dan and I ended up meeting each other, and then we lived in Boston together, and then alter we lived in Brooklyn together. We've known each other for over 20 years. For the last kind of 15, we've both been bouncing around different technology companies across Silicon Valley. And then last year, at the start of the pandemic, we decided to found Mozart Data together.

Dan and I also cofounded a hot source company together that we're really proud of. About 12 years ago, we started Bacon Hot Sauce, the world's first and greatest bacon-flavored hot sauce.

So for 10 years, we were hot sauce entrepreneurs, and now we've sort of pivoted to data space full time.

[00:01:51] GM: Very cool. Well, I don't generally see a lot of overlap between the two industries. Tell me a little bit about some of each of your backgrounds in the data world.

[00:01:59] DS: Sure. I'm more of the data engineer. Pete is more of a data analyst, data scientist. I've worked basically all over the Bay Area, generally startups that are a little bit smaller. So I founded a few companies, or I've been an early engineer at a few companies. I generally prefer working organizations 20 or fewer people. I've done some data science myself. But more often, I'm sort of the engineer alongside folks who are doing analysis. And I really enjoy building tools to help them do their job better.

[00:02:33] PF: And I've been leading and running data and analytics teams at mostly late stage startups. And in the past, only late stage startups could really take advantage of all the data that they were generating. So I've generally built teams and data stacks and data tooling at late stage startups like Playdom, and Yammer, and Zenefits, and Opendoor, and most recently, Ease.

[00:03:00] GM: What have you guys seen changing in the data landscape over the time you've been in the industry?

[00:03:06] PF: So I think the data landscape has massively changed. I hate to do the back in my day, but I'll start there. Back in my day, to get sort of your data pipelines going, it would require hiring a few data engineers spending a lot of money on a fancy database, like essentially buying a large contract that really locks you in.

And today, I think like the biggest change is that to get started with data, instead of building everything, a lot of times you can get away with buying, or buying for a long amount of time. And the other part is that you need many fewer data engineering resources than before. I think the biggest change is that the buy solutions have just gotten so much better, so much more standard. And now, basically, people can start a data team with basically a credit card. And that's, to me, a giant change in data tooling. And then, of course, the obvious things like there

are more data professionals. Similarly, there's a lot more data being generated, collected, but also it's become table stakes.

When Dan and I were starting our careers, the types of companies that would hire data people were like Google. Today, you talk to most seed stage startups, they're already thinking about hiring not just like, say, someone at a college to analyze some of the data that they're creating, but like actually a full-fledged data professional at a pretty early stage, because data has become table stakes for competing in the standard sort of categories of in b2b SaaS, in DTC. Really, even just the reporting of it has become table stakes for raising series seed, series A and beyond.

[00:05:01] DS: I'd add to that, that there's a lot more people that are comfortable in things like reading SQL and answering their own questions. So people that would never call themselves a data engineer or a data analyst, but they've picked up SQL over the years just to be better at their job. And they're able to now sort of take advantage of a lot of these advances as well.

[00:05:27] GM: So what is Mozart Data? And where does it fit into the landscape?

[00:05:32] PF: So Mozart Data is the easiest way to spin up a modern data stack. What that means is that we manage EL for you. We provide you a managed Snowflake warehouse. And we provide a layer for doing transforms, so that in under an hour you can get up and running with a world-class stack that late stage companies would be implementing or using. But you can really do it with no data engineering team.

So the way that we fit into the landscape is we try to bring together the best in class pieces and the necessary pieces, and the types of pieces that real practitioners end up using. And we try to put that all under one roof or one throat to choke.

[00:06:18] GM: So I imagine, I think of the modern technology company, or even really any company, they probably have a long list of SaaS services they're using. Something for accounting, maybe something for HR. Add onto that probably Stripe and a few other things. And we now have all these siloed data sources. How does Mozart Data help me with those?

[00:06:40] PF: That's exactly it. There's this sort of incredible explosion in SaaS tools. But a lot of the sort of winners of these spaces are becoming quite evident, and writing the extract and load from a tool like Stripe to database is a solved problem. And I think the idea is that data has power in these silos. So you can go to Stripe, or you can go to Shopify, and you can look at a count of your customers.

But what you really want to do is actually bring all this data together. So data becomes really valuable for making like insight or doing an ad hoc analysis when you're able to combine often data from multiple different places typically in a central warehouse. So that's really – You hit on really the punch line for why of Mozart's existence, which is the right way. And you see like most sophisticated companies that use their data do it the following way, which is they're doing their analysis out of a powerful columnar.

[00:07:45] GM: When data engineers maybe aren't aware of a tool like Mozart, I think it's very easy to think, “Well, let's just spin this up. I can go to the vendors website.” Maybe they've got very developer friendly documentation. I've kind of got my Hello World going in a few minutes even. That happens quite often with a lot of modern tooling.

So for a developer with a not build here sort of attitude, what's the pitch for Mozart Data?

[00:08:10] DS: There's so much functionality that data engineers all the world over are just building over and over again. I've personally in my career built similar data stacks, similar tools, five, six times in some cases. So I mean, if you really only want to build things again, like I have that syndrome, that's one of the reasons why I have done that. But the some of the underlying tools have gotten so good that there's really no reason to focus on building the mechanics of pulling data from one API and putting it into a database, or snapshotting tables, or scheduling transformations. The more that you can rely on tried and true tested other people's code that will do that for you, the more you can focus on the specifics of your data, the questions that can actually help your company, and really focusing on the problems that are unique to your data.

[00:09:04] GM: Well, you have a lot of, I think, 120 plus connectors I read about on the site. Could you give a sampling of what types of systems you're currently integrated with?

[00:09:14] DS: Yeah. So our most popular are Shopify, Stripe, Salesforce. Everybody's kind of running half their business off of Google Sheets. So being able to pull that into your central database is really valuable. Most of our customers have some sort of application database, MySQL or Postgres, that ideally you're not running your analytics queries on. Much better to replicate that into a columnar database and then do your analysis off of that.

And then like you mentioned before, accounting tools, CRM. Marketing is a big one. I think marketing attribution is a really good way to kind of understand the power of this where you can look at how your ads in Google ads are doing. You can look at how your ads in Facebook or Instagram are doing. But if you really want to know what is our average customer acquisition cost, a simple thing like that is going to require pulling all that data into one place, mixing it up with however you're being paid from your customers. And so a tool like this, getting a central database and then pulling all of your data into it, that's when you can really unlock like some other questions that are going to be very valuable to you.

[00:10:25] GM: Can you describe a little bit about the experience once all the data has landed there? Let's say I, at my organization, have connected five or six of those tools you mentioned, or any commerce company with the Shopify store and Stripe and a few other things, and I'm fully wired up with Mozart Data. Now I have a question, “Okay, great, I can write some SQL queries. How do I know the schema and what kind of opportunities do I have to cut between or join between those data sets?”

[00:10:55] PF: So a lot of the inspiration for Mozart comes from a quote of a former boss of mine. And there's this line that I've always loved, which he says, is that data scientists spend 95% of their time cleaning data and 5% of their time complaining about cleaning data. And they basically never spent any time doing data science. And obviously, the punch line of the joke is you're expecting them to say 5% of the time doing work, but it's actually not even that.

So a lot of what a good transform layer is, is basically the ability, like you said, to explore the data, to understand essentially the columns and the tables and what the definition of – Like what key business definitions actually mean in terms of SQL.

What are what are transform layer is, is essentially one that's easy to do those things, things that practitioners do, to explore your tables and also to then clean up your data by writing those transforms and scheduling them really easily.

[00:12:05] GM: Let's unpack a little bit more about that transform process. Let's say that I'm a data engineer and I'm at the helm of Mozart Data and I've got some very custom need. Something as the data flows through, I want to do some aggregation. How do I get in there and get the system to help me with that?

[00:12:23] PF: I mean, basically, a transform, anything you can do in SQL. If you can write a SQL query in your BI tool, and that result is going to be useful for more than five minutes. In our tool, it's very easy to cut and paste that. You can run it and then you can schedule it however often you want it to run. And we'll do things like we'll automatically parse your sequel to show you the data lineage so that you can see if some transformation breaks, say, you'll get a notification, but then you can also see what else depended on this. What might have broken because this broke? Or generally, more importantly, you can look back and see, if nobody changed the code in this, it shouldn't have broken, unless something upstream of it broke. So kind of being able to see the lineage of how your data is flowing from the time it lands from your SaaS tools until it's in what would you might think of as your productionized tables that are ready for easier analysis. You can see very easily how the data flows through if something breaks everything downstream of it that's impacted by that and things like that.

[00:13:35] GM: Gotcha. Each of those integrations, I don't want to use the word brittle. It's not that they're brittle. But something could happen. An IT professional could decide to rotate a key and suddenly you're no longer able to sync data until that update has to happen. So to some degree, maybe that's the most minor case. You have to keep an eye or maintain these connections. To what degree do users of Mozart Data have to think about those connections or monitor them on a day to day basis?

[00:14:04] DS: Basically none. That's kind of one of the big points of us doing that for you. If Salesforce changes their API's a bit, a million people can go and change their EL process, or the few companies that kind of support that sort of EL, they can go and change that code. Mozart will go and update our Salesforce connector so that you don't have to worry about it. The credential rotation, that'll be something that our customers would have to go and deal with, but they'll get a notification that it failed, and then they could see why. And then they can go and create a new token or whatever.

[00:14:42] GM: Very neat. Well, I'd love to talk through maybe a hypothetical integration. Let's say some ecommerce company, we've got a Shopify store, and I would like to update my HubSpot CRM records when the sales are made. What's the process like to get that going?

[00:15:00] DS: To clarify, do you mean like a reverse ETL? You want you want data pushed from Shopify into HubSpot?

[00:15:07] GM: Or maybe we should revise the question, just looking to explore like a concrete user story.

[00:15:13] DS: Got it. So we do have some customers that are using exactly Shopify and HubSpot. So we don't – I'll kind of go back to my previous statement. We don't actually do reverse ETL much. We'll let you sync things to Google Sheets. But if you want to go like from Shopify and HubSpot into your data warehouse, we cover that. And then if you want to do some analysis or combine your Shopify user info with your HubSpot support info perhaps and then push that back into Shopify or HubSpot, we don't currently cover that part of the data stack. We partner with a couple of companies. I really like Hightouch for that, for example.

[00:15:55] GM: Gotcha. So it's really about empowering the analyst then, I guess. Pulling all data together in a common SQL interface?

[00:16:01] PF: Yeah. I think like Dan highlighted that. We sort of stop where the data becomes specific to you. We are big believers in having humans that have sort of an intimate knowledge of the tables and the business. And the business logic and definitions really drive the insights and the combining. Very often, the types of tools that sort of out of the box can do this for you, I

think fall short, or miss the key insight that you're trying to drive out of the data, just because there's – I mean, I think sort of, we can call it messy data, but it's really just sort of misleading data. Because, essentially, typically more is better. And then like someone who trained that knows the data can refine it down.

[00:16:50] DS: Yeah. I think the centralization here ends up paying off in dividends that you don't really expect to see with one person's work empowering somebody else's work. So like the cleaning and organizing that Pete was talking about a few minutes ago, that could be done in a BI tool. But if one analyst does that in their BI tool, that's not helping anybody else do their job better. But if you centralize that and it's effectively a shared repository of code and you've got these production tables, then maybe one team can be empowered to answer their questions in tableau now. The analysts have also have cleaned up data to do whatever machine learning, or just make some charts that they want to do. And then maybe, like we're saying, some other team wants that data piped back into HubSpot so that their customer support folks can have some extra information when they're dealing with people that doesn't actually live in HubSpot, unless it goes through a process like this. It gets cleaned up, and then put back into the tool.

[00:17:52] PF: And this is I think a lot of what has inspired the change of the acronym. So in general, people are now saying ELT or ETLT. It's really highlighting the importance of that last sort of tea stage before the BI, so the last transform stage before the BI. So it has really become a common best practice to be cleaning and transforming and sort of having a uniform set of definitions one layer above your BI tool.

I think, back in my day, the practice was always do messy copy and paste of SQL or have sort of drag and drop definitions in your BI tool. Today, it's been well surfaced that a very, very, very common problem is that it's one question, several different “correct answers”. One of the – Not cure alls, but one of the best ways of addressing this very common problem is to have common definitions written essentially before the BI layer.

[00:18:59] GM: Interesting. So in that regard, I would describe Mozart Data is like foundational. That it's going to have maybe a few key services that you'd offer in that transform layer. I'm thinking like data enrichment could be one. A little bit of cleanup. What are some of the appropriate things that you see people using it for before hitting their BI layer?

[00:19:21] PF: Mostly joining. Like unioning and joining. Again, I think this is where the power of – This is where one plus one definitely doesn't equal two. It is certainly the case. And Dan gave an example where you can very easily sort of assess like, “How are my Google Ads doing? Or how are my Facebook ads doing?” But what you really actually want to know is often further down the funnel that is easy to measure. When you're sort of assessing things like CAC and LTV, you're often making assumptions once users get to a certain stage if you're only able to analyze the data from, essentially, your marketing sources.

You also have a slightly tough time comparing marketing sources, because you might see that each is providing you the same cost to get leads to the same place. But those leads might perform incredibly differently. And then last, if there is such a thing as attribution, you can't do it well when you're doing an individual source. Because most often, if you were to ask, say, Google, how they're doing in terms of what credit they deserve on a lead, they would say all of it. And that might not actually be true. You might have learned about in another place, and the last touch might have been actually through search.

But I think the typical answer is that it is really important to join data together or union data. And that's not necessarily specific to Mozart. It certainly is the case that there are many tools in both the transfer space and the BI space that enable you to do it. But it is certainly a best practice to get, basically, as much as you're sort of going to be making decisions off of data together in your data warehouse.

[00:21:17] GM: Makes sense. And I'm curious if you have any thoughts on the evolution of a team growing in a startup or a midsize company. With the availability of tools like yours, it seems like maybe a data engineer could be delayed a bit because some of the fundamental stuff can be done in an automated fashion. I don't believe we'll ever eliminate the data engineer position. I think the tooling just scales the professional. But I'm curious to get your thoughts on how these things might be evolving in the modern tech group.

[00:21:43] PF: Sure, I agree with you that the data engineer’s job is not going away. In fact, the demand for data engineers is massively outpacing the supply. So I wouldn't be worried about your employment prospects if you were a data engineer not because of the existence of Mozart

or similar tooling, just because the demand for data and people very capable of moving and manipulating and summarizing data is just honestly increasing. It feels like by the minute.

So the point is not to totally obsolete the data engineer. That said, it's not to do that. And I actually attended a talk by George Fraser of Fivetran. And Fivetran is an organization of many hundreds of people that doesn't have a data engineer. Now their business is to obsolete the data engineer. But I would say that you can see very large, very successful businesses that would typically rely on not a single data engineer, but a team of data engineers to do certain work.

What you don't want is people that are becoming increasingly more scarce and increasingly more valuable, and their output is becoming increasingly more leveraged doing highly rote work that, like as Dan had mentioned, in some sense is a solved problem where people are solving it at scale. So I think I'm religiously of the camp of the tooling, including Mozart, is now at a stage that's good enough that can certainly push your first data higher not to necessarily be an engineer, or a data engineer, or a hybrid data engineer and data analyst. Instead, you should be looking for essentially the types of skills that you should leverage a tool like Mozart and then hire for the types of skills that are important for teasing out insights from the data.

[00:23:34] GM: I'm curious if you have any thoughts or seeing patterns as the product grows around the right point of adoption. Is this something early stage companies are picking up because they see the need and they don't want to build out all these connectors? Or is it something a later stage company is adopting?

[00:23:50] PF: So the answer is both. I am surprised at just how early companies are essentially getting on board with data. Again, most of the companies that I joined, I joined as employee number 100. And that was the time where they were making big investments in data teams. Today, we obviously have sold to companies in our YC batch that were just a few people even before they started generating a lot of revenue or certainly a lot of data.

Again, like I mentioned in an earlier answer, part of that is that you can just get started so much more inexpensively, right? So aside from the people cost, you can buy a trivial amount of compute. Basically, there's a variety of reasons why this sort of use of data is happening at an

earlier and earlier stage. But it's mostly because it's demanded of these companies. So it can be demanded by the market to like behave optimally, right? So to figure out what's working and what's not, and double down into those that's working. And there are so many tools that are much better at sending you signals about what's working. You can hook up a product analytics tool and instantly see what users are doing and how they're interacting with your website.

Similarly, it's the case that it might be demanded by the capital market. So in order to get venture funding, people like to know, these are your metrics. And the ability to gather like the standard set of metrics very quickly and update them with a single click is a really powerful tool for raising money.

Really, I would say that the reason that the movement has gotten data to show up earlier in earlier stages, and now it's really not uncommon for pre-seed companies, especially technically savvy ones, to be using data and even a lot of data is because, one, again, it's sort of like supply and demand. There's an increased demand for it. And again, it's also like a lot easier. So there's an increased supply of tooling.

[00:26:02] GM: Avoiding the reinvention of the wheel is one of the most appealing things to me for a tool like Mozart Data. I've got plenty of stuff I need my engineers to build. Why redo yet another Stripe integration or something like that? But then, of course, maybe an added benefit of that that a lot of people miss is that I also don't have to maintain it. That's your problem. Are there any challenges you've seen, especially over time maintaining such a large number of connections in light of possible breaking changes from vendor API's and things like that?

[00:26:34] DS: I mean, this stuff scales pretty well. When we fix a connector, we fix it for all of our customers. And if we didn't exist, and those people weren't using a similar tool, that's 10x the work, 100x the work. So I mean, this, to me seems like the efficient way to do it. Basically pull it in in a very generic way. And then each person can transform it in whatever way is unique to their needs. And then we just have to maintain like that initial, "Let's just it all out." We're going to naively put it in your database, and then you can take it from there.

[00:27:11] GM: Right, right. And I like the appeal of that, that effectively, any challenges on the API end become your problem. I was just curious, and maybe this hasn't happened. Maybe it's

been fortunate. But it wouldn't surprise me to know that, let's say, one of your connections, and we don't have to name them or anything, would do a botched release. And suddenly you're getting all these rate limit errors. And now that's your issue to solve for 100 people using that connection. Do you face challenges like that? Or is the ecosystem of API's pretty stable?

[00:27:40] PF: We definitely do face that. I mean, the answer is just hire really good engineers and have them build systems to monitor and staging environments so we don't push broken things, kind of standard engineering practices or standard good engineering practices. We do use our 120 connectors. We use a mix of powered by Fivetran. So Fivetran maintains a lot of those connectors. We use Singer taps, many open source public Singer taps. So the community is maintaining those. And then we write our own Singer taps. So it does get complicated. But my advice to a listener is use a tool that's doing that maintaining for you. If you're going to maintain it, then build a good engineering team with good engineering practices.

[00:28:27] GM: What is the getting started story look like today if there's maybe a listener at a startup thinking they need a tool like this? I know you guys are doing demos and things like that. Are you accepting new customers? And what's an onboarding process look like?

[00:28:41] PF: So we're definitely accepting new customers. And we will continue to be doing that for a while. The onboarding process, we try to make it painless, right? So that's the brag, right? Like so what we think of as our value prop is the ability to – Is sort of like the easy button for onboarding a world-class data pipeline.

What it typically involves is, one, you can you can sign up for it on our website, and you don't have to talk to me or anyone like me to get started. The flip of it is, is I find it to be a lot like superhuman, where like a 30 minute sort of consultation that's just simply a push in the back. When I signed up for Superhuman, I didn't really need to know how to use an email client. It was just really nice to have somebody sort of walk me through kind of using the platform.

And when you do that, I think like the biggest challenge with most things data is just getting started. And I think, again, it's an intuitive enough platform that you should be off to the races in under an hour. But it's always great to essentially have those best practices or a lot of that experience with the data cleaning really at your fingertips or the standard sort of outputs or

tables that are going to be really critical to most businesses when they fall into the categories of b2b, DTC, etc.

[00:30:16] GM: What are some of the unique challenges in industries like that? Direct to consumer is kind of an interesting one, in particular, if you have time to unpack it?

[00:30:23] PF: Well. I mean, that's not a particular – Most businesses have, especially at the early stage, are really trying to solve a mystery, which is what is my LTV and what is my CAC? And if you can sort of get your – And some DTC businesses sell a one-off product. I think at the start of this, I mentioned that Dan and I sold hot sauce. So we had a Shopify site, baconhotsauce.com, where we sold hot sauce. And a lifetime value as a hot sauce seller tends to be, honestly, a lot of one-off if you're doing sort of online hot sauce sales. So there's not really much of a lifetime value calculation. For other companies, there's sort of repeat purchases, there's increase in basket size, there's sort of churn, there's all of these things that people are trying to measure.

If you're a DTC company, you're really trying to measure two things. One, like I said, is lifetime value. The problem is you don't have a lifetime to let that live out. So you're trying to make inference about how often a customer comes back based on, typically, cohort analysis. And shorter time windows then, sort of it's not all that useful to make the prediction. If you do a one year LTV, or a two year LTV, or a 10 year LTV, it's certainly not useful to make that prediction 10 years later.

So the other side of it is CAC. So you want to just measure your marketing spend and you want to measure its efficacy. And also you want to measure that on the margin. And then you want to essentially divide one by the other. And hopefully if you're a good business, that ratio is good. And then you sort of dial in how much you want to grow rapidly versus sort of make that ratio look good. As you as you amp it up, that ratio gets worse. So that's really a business decision.

But, really, in the DTC space, those are the main metrics. And then the question is how do you get at them? Especially when you're doing multiple channels for advertising, especially when there's real like ambiguity about kind of customer lifetimes. So there, there's a little bit more art than science. But at the end of the day, those are the questions. And the best way to be

answering that question is generally to, again, get the data to a central place and then be able to join the tables and then and then start analyzing it.

[00:32:51] GM: If I'm in a situation, I happen to know a lot of those direct to consumer products. They'll make pretty swift decisions early on with some early indicators and things like that. So recency is critical. What's the story for recency in Mozart Data?

[00:33:06] PF: Yeah. So let me say that we're not like a company that believes in analysis paralysis. We have a company core value. Like, most often, in doing our own analyses, we like to do the 80/20 principle. I'm somebody that did my PhD, and I spent probably of the six years I was in grad school, I probably spent like five of them getting the standard errors of my point estimate in my dissertation to like four decimals. That's not kind of how one makes decisions. But it's very much the case that cohorts change, right? And they change rapidly when the source changes, or they change rapidly over time.

So, most often, when you're looking at a cohort analysis, you are looking at like a time-based cohort analysis or channel-based cohort analysis. And I think often where companies get in trouble is they have great success in a specific channel, and they think that it generalizes. So the ability to look at your data with a short lag is a really powerful solution to a world where things sort of shift under your feet by the day, week, month, whatever it is.

[00:34:18] GM: Gotcha. No, I meant in terms of recency that you're, in some sense, at the behest of whatever the API does. If it's, let's say, an ad tech server, I'm going to pull in some campaign data. If they only updated midnight, well, that's as quickly as I can get new data in. Do you face any engineering challenges. Or I guess as a user, I might like to know what I should expect in terms of how in sync my systems are. What's the typical experience like?

[00:34:47] DS: I would say these days, most API's are pretty good about their data that they'll provide is, in general, almost real time. It might go through some internal ETL process. But that's going to be running very, very constantly, and will let you sync as often as every five minutes for a connector so you can at least get it into your database. And then it really depends on how much work is going to happen from the time it lands in your data warehouse till it's at the tables that are then feeding into your other systems or your reporting. And that's kind of just on

a case by case basis. Depends on how many hours, if you have hours of compute, if it's seconds. You can get pretty real time.

I think the important thing is that like you're often comparing this. Definitely, some of our smaller customers, they would say I like to lead often if I'm talking to a new potential customer with trying to tease out what are some questions that you wish you knew about your business and you know are out there in some combination of your data, but you can't put it together. And oftentimes they'll say, “Here's our main question.” And we do have the answer. It's just that it takes three days to pull data from all these different sources. Then we put it into Excel. And then we do a bunch of manual joins. And then we have this monthly report. So in a lot of cases, we're automating, and then making data. Have one hour of lag rather than what used to be sometimes a month of lag and much more manual labor to produce.

[00:36:22] GM: Gotcha. Do you face any sort of cold start problem? A new customer could sign up tomorrow with a 20-year history in their Salesforce account.

[00:36:32] DS: I'll say a tiny bit. We had one potential customer that we lost because we were syncing their HubSpot data. And three weeks later, it wasn't done. But that's a real outlier. Most of these connectors, you can connect it and tomorrow your data will be there. In the worst case, in some cases, it's almost instantaneous. It really depends on how much data we're talking. And some of the details of the connector, HubSpot happens to have pretty strict rate limiting. So if you have millions of users in your HubSpot account, there's just no way around it. It's going to take a while to get that data in.

[00:37:10] GM: Makes sense. Yeah. I'm curious about who are the people that find you guys and bring you into their organization. Is there a commonality to the title of the person who first engages?

[00:37:21] PF: So the answer is, actually, it's quite varied. We typically interact with a lot of data folks. So data analysts, and heads of analytics, and data scientists. But actually, our best customers tend to actually have Ops related titles, SalesOps, MarketingOps, BizOps, RevOps. Usually, it's somebody that's been kind of saddled with the challenge of bringing data into the organization or doing reporting at the executive level. But they might not have a really deep

background in in data or data pipelines. And they certainly don't have a budget for hiring one, two, or even more data engineers. Those are really our ideal customers.

As Dan mentioned, there's this real great growing population of folks that have been at companies that have very successfully leveraged data. And in doing so they've really empowered themselves by learning SQL and learning how to manipulate and think about data. And now kind of they find themselves at companies that don't have that existing infrastructure, and they sort of they want it. They miss that ability to answer kind of questions in a really automated fashion and the ability to sort of dive into the tables and not be sort of living out of spreadsheets exclusively.

So there's this growing population, and these folks used to have the title like data analyst or business analyst. But now they're really anyone. There're folks in product orgs, in marketing works, a lot of folks in sales and sales operations orgs as well that all are becoming really data savvy. And that's a great trend that we really hope to take advantage of.

[00:39:06] GM: I see a number of directions you could go, but I'm curious about the ways in which Mozart data is going to grow in the future. What are the plans for where the product will head?

[00:39:16] DS: I mean, the way that we think about product direction, we work really closely with a lot of our customers. And our goal is to make your job easier as an analyst. So we work closely with our customers. And we're pretty aware of what is currently hard for them with their current tools when that includes Mozart. I mean, some of the specifics kind of as a data team, or as an organization matures in their use of data, the stuff that comes up that needs a little bit more work are things like data cataloging and observability. So understanding how the data is flowing through your system and then being able to take advantage one person’s work, being able to be taken advantage of by another person, that requires some cataloguing and so that you can basically share the information so that others can take advantage of that.

And also a lot of stuff around permissioning. As a company grows, and maybe the initial users of a system like this are effectively admins of the whole data. But then they start empowering other people in the organization to answer questions. Maybe they hook up possibly even like a no-

code BI tool, or a low-code BI tool. And you start empowering other people in the org to answer questions. Then you have issues around permissioning. And you might want the sales team is now going to use some of the outputs of your data pipeline, but you don't want them to have access to the HR data that's also in the database for a different team to use. And those are tricky DBA problems to manage that. So we have plans on our roadmap to try to make that easier to manage.

[00:40:58] GM: When Mozart data gets deployed to one of your customers, is it a set it and forget it kind of thing? Or does someone in my org end up operating that in some way.

[00:41:09] DS: So we have some of both. I mean, you can connect something, maybe transform something, and then set up a dashboard, and that dashboard will work forever. And some of our customers do that. They come to us. They have some problem. They know they can solve it with the combination of data across their tools. And we help them get there.

But most companies at that point will then also see, now they understand kind of the power that they have. And they'll start building more pipelines. They'll start answering other questions. And one thing we definitely see, the core Mozart users at a company, they'll often – They are then more efficient at answering questions. And so they'll start getting more and more questions coming into them. And the use of data tends to grow when you see it being successful and easier to do.

[00:42:01] GM: And do you envision – To make a comparison to the Elk stack, where you have Elasticsearch, LogStash and Kibana, do you see Mozart being the M of some other stack?

[00:42:12] DS: Yeah. So we're trying to kind of be – We try to cover everything in a data stack. Like right now, you can build a data stack using 10 different services, right? There's a ton of companies that are like all we do is data cataloging. All we do is data lineage. All we do is transformations, or EL. We're trying to be the turnkey all in one. But in many cases, you're going to want to go deep on one dimension of that. So you might get most of what you need out of Mozart. But then you want to hook up one dedicated cataloguing tool, or you want to use a machine learning tool, things like that. And so that's a pretty common use case.

Like I was mentioning before with like, reverse ETL. That's something that we dabble in a little bit. But if you want to go deep on that, then there's whole big companies of people that are dedicated just to moving data out of a data warehouse and back into SaaS tools. So those are pretty common partners for us.

[00:43:17] GM: What's your common use case for people getting started? And can you tell anyone who's listening and fits that model where they can get going?

[00:43:24] PF: Yeah. When companies are really ready to make that investment in data, whether that's a person, whether that's data infrastructure, we want to be the easiest, fastest, most cost effective way of getting you successful at the start of your data journey. Our goal is as a company is to basically empower business users to take advantage of their data to write definitions in languages and tooling they're comfortable in, like SQL. And we think of ourselves as the easiest way really to get started with a world class modern data stack.

[00:44:06] DS: And I'll throw in there, you can go to www.mozartdata.com and either sign up for a demo if you want to chat, or sign up for a 14-day free trial.

[00:44:16] GM: Well, Dan and Peter, thank you both so much for taking the time to come on Software Engineering Daily.

[00:44:21] PF: Thanks, it was a pleasure.

[END]