EPISODE 763

[INTRODUCTION]

[00:00:00] JM: Real estate is an asset that is not straightforward to invest in. Real estate can generate excellent returns for investors, but can require much more time and expertise than stocks. Cadre is a company that allows users to invest in real estate more easily and intelligently. Cadre provides users with lots of data about potential investments and enables investments in those opportunities within the Cadre platform.

Leonid Movsesyan is the head of engineering at Cadre and joins the show to talk about the problems being solved by the company in terms of product development, infrastructure engineering, hiring and data science. To build a platform for evaluating real estate investments, Cadre ingests and merges lots of datasets, some of which are public and some of which are private. This gives investors a detailed picture of the value of these real estate investments, and the vision for cadre is much bigger. It's to be able to allow investors to invest in unusual assets, assets that are not simple to invest in today.

Before we get started, I want to mention that we have another podcast that we're working on, Fintech Daily. This is a new podcast from the Software Engineering Daily team. We're covering payments and cryptocurrencies and trading in the intersection of finance and technology, and we're looking for volunteer hosts for Fintech Daily. If you're interested in working with us to conduct interviews, send an email to host@fintechdaily.co. You can find the podcast on iTunes or Google or everywhere else, and if you're interested in hosting, don't hesitate to reach out. We are very much in the early days of Fintech Daily. We're flushing out the content strategy, and if you are interested and you're well-qualified and you want to be part of the burgeoning Fintech community, then send an email to host@fintechdaily.co.

[SPONSOR MESSAGE]

[00:02:09] JM: Kubernetes can be difficult. Container networking, storage, disaster recovery, these are issues that you would rather not have to figure out alone. Mesosphere's Kubernetes-as-a-service provides single click Kubernetes deployment with simple management, security

features and high availability to make your Kubernetes deployments easy. You can find out more about Mesosphere's Kubernetes-as-a-service by going to softwareengineeringdaily.com/mesosphere.

Mesosphere's Kubernetes-as-a-service heals itself when it detects a problem with the state of the cluster. So you don't have to worry about your cluster going down, and they make it easy to install monitoring and logging and other tooling alongside your Kubernetes cluster. With one click install, there's additional tooling like Prometheus, Linkerd, Jenkins and any of the services in the service catalog. Mesosphere is built to make multi-cloud, hybrid-cloud and edge computing easier.

To find out how Mesosphere's Kubernetes-as-a-service can help you easily deploy Kubernetes, you can check out softwareengineeringdaily.com/mesosphere, and it would support Software Engineering Daily as well.

One reason I am a big fan of Mesosphere is that one of the founders, Ben Hindman, is one of the first people I interviewed about software engineering back when I was a host on Software Engineering Radio, and he was so good and so generous with his explanations of various distributed systems concepts, and this was back four or five years ago when some of the applied distributed systems material was a little more scant in the marketplace. It was harder to find information about distributed systems in production, and he was one of the people that was evangelizing it and talking about it and obviously building it in Apache Mesos. So I'm really happy to have Mesosphere as a sponsor, and if you want to check out Mesosphere and support Software Engineering Daily, go to softwareengineeringdaily.com/mesosphere.

[INTERVIEW CONTINUED]

[00:04:27] JM: Leonid Movsesyan, you are the head of engineering at Cadre. Welcome to Software Engineering Daily.

[00:04:32] LM: Thank you.

[00:04:34] JM: Cadre is a company that enables investment in alternative assets, one example is real estate investment. That's what Cadre is focused on right now, and we will ease into a conversation about what Cadre does and then we can get into some engineering aspects. Real estate investment is a very large area. Describe the kinds of investments that Cadre is focused on.

[00:05:01] LM: Yeah, we are focusing primarily and at this point only on commercial real estate investment, which is basically an opportunity for individuals to invest and buy a share in a commercial real estate building, such as office building or a multifamily, or a hotel.

[00:05:19] JM: And this is an investment platforms. So what are the types of investors that Cadre is catering to?

[00:05:26] LM: Yeah. So we're focusing on individual investors as opposed to the traditional old-school private equity firms that are focusing primarily on institutions, such as pension funds, or big family offices. We focus on individual investors, the folks with a substantially lower net worth who would typically not be able to work with one of these big private equity firms.

Unfortunately, due to regulations that we need to be compliant with, right now we only can work with accredited investors and beyond, which means that as an individual you have to have – I think the rule is $200,000 of annual income for two years in a row, or more than $1 million in liquid investable assets to become a member of the platform.

[00:06:12] JM: One way to think about Cadre is as a two-sided marketplace between investors who want to put money into real estate and operators who manage real estate. I realized it's not a perfect description, but to some degree it is a platform and you get to operate this platform. Give an overview of what you're trying to build into this platform.

[00:06:37] LM: So first and foremost our goal, our sort of high-level mission is to improve people's financial futures. So the paramount of our platform is giving people access to really good investment opportunities. These investment opportunities are being vetted by our internal team, and right now the processes is semi-manual with a large emphasis on manual. Obviously, as a tech company, we want to remove the manual part as much as we can and make it as

automated as possible to enable the larger scale to lower the minimums and to invite more folks to the platform, give them opportunity to invest.

[00:07:18] JM: This is a common strategy for technology companies to look at something that is difficult to automate completely, but you start off with perhaps a good interface that makes it easier to get some transactions flowing through the system, and then you get some number of transactions and you're handling them manually under the surface to some degree. Over time you figure out what are the areas of a manual process that you can automate more and more and you gradually automate it more and more, and that allows you to build a moat. Can you give an overview for the strategy of what are the things that are manual today and what are you able to start automating?

[00:08:02] LM: Yeah, you're exactly right. So when we started about four years ago, there were no automation for any of these processes in place. So there was a lot of manual paperwork in terms of on boarding investors through the platform, doing all the necessary checks that regulators require us to do, performing all the transactions completely offline, sending people PDFs to print and sign and scan and send us back. Things like that, and slowly but surely we started automating different pieces. So like things like signing documents online, things like completing all the necessary checks online, and right now most of the user-related flow on our platform is automated. We have a fair amount of manual operations that still are being performed, but for majority of customers we were able to generalize and automate the flows.

Right now we're focusing on the other side of a marketplace which is substantially harder, I want to say, to automate, because we need to be able to get all the deal information about various deals that are coming into the platform and make a decision whether we think this deal is good or we think it's going to be not as great as a sponsor, an operator who sells this deal tells us it's going to be. So that part right now is mostly manual and we're looking for various ways to bring data that will help us to make this decision an automatic fashion.

[00:09:35] JM: If listeners are wanting to know more about this strategy of starting something that is difficult to automate, but presenting a shell around it that lets it be automated, I recommend – There's an episode we did about a company called Checkr, which is like the API for checking if people are – Or checking people's employment history and doing background

checks on them, and it's just like something that is really hard to build as an API because a background check has a lot of manual processes, but they did it and it builds an amazing business out of it. I think as developers, this is a bright future, because over time these will be APIs that will be able to leverage whether they're automated on the backend or not. They'll be things that we could build on top of.

[00:10:21] LM: Yeah, exactly.

[00:10:22] JM: Speaking about Cadre, you want to build this investment platform, and a lot of making smarter investments in alternative assets such as real estate is about getting information. How does information about real estate investments typically get aggregated and where are people finding it? Where are they accessing information about real estate?

[00:10:51] LM: There are three ways to gather this information for us as an investment platform. The first way is the most obvious. There is a lot of public sources of information that you can scrape in certain ways, which we do regularly. It's things like census publications and median income information that governments agencies are publishing. There are also things that we can scrape legally according to terms of service of various applications online, so that this is the first big bucket, the publicly available information.

Then the second big bucket is there is plenty of vendors in this market who work closely with a lot of operators and they aggregate information and then sell it, and we have partnership with some of them, partnerships with some of them. So we buy their information and incorporate it into our platform. The third way is to capture the data on our own assets that we already have on the platform and make informative decisions based on this information.

[00:11:56] JM: The world of real estate is not very well developed in terms of technology. There are some burgeoning technologies that are being used, but as far as I understand, the technology is not improving really rapidly. I remember when I was in high school, my mom was a real estate agent and she used something called MLS, which is the multiple listing service. It just felt like some dated technology.

Describe the technology that is used in the world of real estate today.

[00:12:29] LM: There is plenty of startups who are trying – You're exactly right, by the way. It's a pretty ancient field, and there is a lot of startups who are trying to disrupt this world and trying to build the software to solve these issues, because a lot of this old school, especially in commercial real estate field, obviously, a lot of the old school companies operate in Excel models that were created decades ago. A lot of email back and forth kind of thing, and obviously we are trying to be on the edge of technology, but at the same time you have to accept the fact that certain partners are just using the super old obsolete formats that we have to learn how to understand the exchange in. So that's kind of another part of the interesting engineering challenge for us, how to be very cutting-edge technology, but at the same time be able to work with this older formats of information.

[00:13:24] JM: Let's say I'm investor that wants to put money into real estate. So I make an account on Cadre and am I just looking at all of these different investments and I can just put money into them however I want to and Cadre gives me a bunch of data on these different investments? Can you describe the experience of the investor who is logging into Cadre and using it?

[00:13:46] LM: Yeah. So you basically can describe the two major ways how an individual can invest. The first way is to invest through what we call deal-by-deal engagement, where exactly as you described, you get access to the platform, you login and you see multiple investment opportunities, and you can click in every single one of them and get information about our model, about performance of this asset, about our hypotheses on the macroeconomics of the region, of the specific asset type in the region, and so on and so forth. We have a very informative video where you can see the asset filmed by our crew and also hear about all the assumptions that we have. You can see the schedule that we plan, like what sort of renovation we're planning to do with the asset and so on and so forth. So that's the one type of investment that you can do, and you can put your money in a specific deal, or alternatively, you can say, "I want to build a commercial real estate portfolio, and here's my commitment of a certain amount of dollars," and we will distribute your investment into multiple assets that are coming into our platform so you can have a diversification within commercial real estate alternative.

[00:15:03] JM: So that you got a few products. You got the product where you can find deals and look at those deals and dive deep into them and build your own little portfolio of real estate investments, or you've got something that's a little bit more like a Robo-advisor where you just put some money in and it's finding good real estate investments and maybe those investments fit some criteria that helps with your Robo advisory balancing, and then that requires maybe a little less effort. But then you'll get a little less specificity out of your real estate investment.

[00:15:35] LM: Exactly. The way we like to think about it is that if you are comfortable making your own investment decisions, the first product is probably better suited for you. But if you are say not that informed in terms of commercial real estate or alternative investment just in general and you just want to – Like say, you are heavily in tech stocks in your portfolio and you want to diversify, but you don't want to get a PhD in commercial real estate, then you can choose managed portfolio.

[00:16:05] JM: And then on the other side of the marketplace, you have these real estate owners or operators and they are putting up opportunities for the investors. Is that how it works?

[00:16:19] LM: Right now we're actually only putting the investment opportunities that Cadre underwrites our self. So will work with these operators who bring the deals to us and then we do underwriting ourselves and then decide whether or not we're going to put the asset to the platform.

[00:16:36] JM: Very cool. So that simplifies the equation a little bit since you're doing the underwriting and you don't have a bunch of back-and-forth with the other side of the marketplace. It's more like a one-sided marketplace. It could become a two-sided marketplace eventually.

[00:16:51] LM: Yes.

[00:16:52] JM: All right. Well, let's get into the technology. Give me an overview of the Cadre stack as it stands today.

[00:16:59] LM: So we are firm believers in engineering for the sake of the product and not for the sake of engineering. So I want to say we're relatively simple in terms of our stack. We are using a monolith Python application, because it works at our scale and our size. We are heavily on react-gs on the frontend and we're exclusively in AWS or we don't host any of our own infrastructure.

[00:17:25] JM: Can you tell me some of the AWS managed services that you're using?

[00:17:29] LM: Of course, yeah. We are obviously relying on their load balancer and our fleet is the Elastic servers, AWS in EC2, and we're also relying on their S3, and our database installation is also an RDs database.

On the other side of our AWS infrastructure, we are heavily using things like Redshift for our data efforts and ETL-ing. But yeah, that's pretty much what we have in place, yeah.

[00:17:58] JM: I like it. I like it because it's simple and pragmatic, and like earlier shows that I did, I would say maybe like two years ago, I was always excited when somebody was talking about a very complicated architecture. We've got containerized this and that and we've got these continuous delivery, microservices, complicated architecture, and over time you start to realize that there are some companies where it makes more sense to have a very simple architecture, particularly a company like yours where you've got a lot of complexity on the product design and the marketplace development side of things. It makes a lot more sense to keep things very, very simple in the way that the product stands today.

[00:18:45] LM: Yeah, exactly. For us as a product-oriented company, the biggest value that we need to extract from our infrastructure is ability to ship safe and fast. So we have to be able if we need to pivot technologically and implement certain features or say regulator comes to us and they want us to do certain things online, we have to move fast, we have to iterate fast. So velocity for us is essential. But at the same time, we try to keep infrastructure as simple as possible so the deployments are safe and we don't have to think about billion moving parts in our infrastructure.

[SPONSOR MESSAGE]

[00:19:33] JM: DigitalOcean is a reliable, easy to use cloud provider. I've used DigitalOcean for years whenever I want to get an application off the ground quickly, and I've always loved the focus on user experience, the great documentation and the simple user interface. More and more people are finding out about DigitalOcean and realizing that DigitalOcean is perfect for their application workloads.

This year, DigitalOcean is making that even easier with new node types. A $15 flexible droplet that can mix and match different configurations of CPU and RAM to get the perfect amount of resources for your application. There are also CPU optimized droplets, perfect for highly active frontend servers or CICD workloads, and running on the cloud can get expensive, which is why DigitalOcean makes it easy to choose the right size instance. The prices on standard instances have gone down too. You can check out all their new deals by going to do.co/sedaily, and as a bonus to our listeners, you will get $100 in credit to use over 60 days. That's a lot of money to experiment with. You can make a hundred dollars go pretty far on DigitalOcean. You can use the credit for hosting, or infrastructure, and that includes load balancers, object storage. DigitalOcean Spaces is a great new product that provides object storage, of course, computation.

Get your free $100 credit at do.co/sedaily, and thanks to DigitalOcean for being a sponsor. The cofounder of DigitalOcean, Moisey Uretsky, was one of the first people I interviewed, and his interview was really inspirational for me. So I've always thought of DigitalOcean as a pretty inspirational company. So thank you, DigitalOcean.

[INTERVIEW CONTINUED]

[00:21:40] JM: What about data? What's the volume of data that you're dealing with like? Is it so big that you have to have a complicated data platform?

[00:21:50] LM: I don't want to say it's big, but it's very different. So like the data sources that we have are extremely diverse. So it's not complicated in terms of the scale, but rather in terms of the ability to plug in and out different data sources. That's where we put a lot of our engineering thought on how to make it as pluggable as possible. Also the biggest problem that we have, a

challenge that we have is the fact that the data is not necessarily well-normalized or cleaned. So we have to build a lot of engineering and machinery around that.

[00:22:27] JM: And what are some of those datasets when you're talking about the diversity of the datasets and the range of different data formats and types? Give me a sample of some of the different data types.

[00:22:41] LM: So some of them could be just plain text file that one of the governmental agencies put on their website every year, and the other can be a binary format, like a proprietary binary format by one of our vendors that we need to get us back for and implement the parser on our side.

[00:23:01] JM: Wait, proprietary data format? So like there's real estate data companies that have like – I don't know, .realestate file?

[00:23:11] LM: Something like that. It would be almost like the Excel file, but slightly different.

[00:23:18] JM: Why is that? Is that like – Do you need some kind of license to be able to even open those kinds of files?

[00:23:24] LM: I think it's just things have been done this way for decades and there is no real – Like the companies that work in this field, they don't really see the value in unification. So I think that's the main reason why there is no unified formats for data.

[00:23:40] JM: Do you have a lot of overlap in the different datasets? Like do you have, for example, some TSV file that comes from one data company and it gives you the information about a particular property and then you get some .realestate file from some other proprietary dataset and it's about the same address and you have to merge those two datasets about the same piece of property and you have to keep them in sync? Is that an example of a kind of data problem you have?

[00:24:13] LM: Yes, and we also need to understand which of them is more trustworthy, because sometimes information would be vastly different. So that's where our certain data

science initiative comes in. We're trying to clean the data. We're trying to unify the data, normalize it and make sure that we can represent it in a very clean format to our investment team and we can start educating our own machinery on like decisions around the returns that the building can generate or cash flow of the building based on the macroeconomics of the region and the asset performance historically and things like that.

[00:24:50] JM: This is like when the data scientist has the diagram. It's like what people think I do versus what I actually do, and what people think I do is like developing artificial intelligence to take over the world, and what I'm actually doing is data cleaning and finding like which dataset is more trustworthy.

[00:25:09] LM: Yup.

[00:25:11] JM: Now, is that the majority of the challenges around building your "data platform" is doing – I hate the term data cleaning, because it makes it sound unglamorous or not fun, or not educational, but data cleaning is like a huge source of problems in software engineering.

[00:25:35] LM: Well, data cleaning is just the first step, because when you have to clean data, you need to start being able to make automated decisions on the investment, and that's where like the next phase and the next interesting challenge for data science comes in.

[00:25:50] JM: Right. Okay. So what's an example, once you have the data clean, let's say you have decided who is the more trustworthy source of any given conflict that could arise between two different datasets and you had this end result of properties with really good data around to them. How do you approach the process of building models and surfacing insights that classify those properties as being good investments?

[00:26:22] LM: The secret sauce, exact secret sauce of the platform is in this. So right now it's just our investment team that has combined decades of experience and have seen various cycles in economics and various types of assets and worked with various types of partners in terms of the operating the building and managing the building and so on and so forth.

So our goal as a tech platform is to actually codify this experience and codify this knowledge. So as long as the data is clean and trustworthy, we should be able to make the same level of decisions, but automatically.

[00:27:00] JM: So in the last five or six years there has been the devops movement, and the devops movement was all about breaking down the silos between developers and operations teams. There are other silos that have been broken down over time and silos that continue to exist today. You have the silos between software engineers and designers. You also have the silos between software engineers and data scientists and it seems like you would have a similar silo between people who are savvy real estate investors that maybe you've brought on to the Cadre team to help you develop your strategy and you have to figure out how do you extract their intelligence and put it into data science and machine learning algorithms. Have you developed any useful communication strategies that you can share about how you extract that knowledge from these investment professionals?

[00:27:58] LM: Yeah, it's actually a very good question, because that was my main concern before joining the company, because I was like, "Okay. So we have all these investment professionals who are highly motivated to keep their jobs," and you want to automate them out of their job. So it didn't make a lot of sense, right? But the good thing about Cadre is that all the members of our investment team are actually bought into this long-term strategy off automating the processes on both supply and demand side of our marketplace and they are actually more than willing to help to share their knowledge with our dated team and make our automation better. So it's just a matter – For us, it was a matter of finding rights people who are bought into the long vision and the mission of the company.

[00:28:43] JM: So once you find those people, what do you have them do? How do they actually convey that useful information to the data scientists?

[00:28:52] LM: So right now they do almost the same job that they would do in a traditional private equity firm but with more emphasis of capturing data and like building and sharing more information about the models that they built with our data team so we can start implementing the same models, but on the backend of our marketplace as opposed to an Excel spreadsheet, right? So that's the process. We have a weekly syncs between our investment team and the

data team specifically, because that's the biggest overlap. There is a lot of knowledge transfer going back and forth. Obviously, they have to step up a little bit in terms of their understanding of technology and capabilities of technology. So it's been a process on both sides, but we are getting to the place of balance where there is a very good continuous flow of information going either way.

[00:29:46] JM: That's actually really interesting. It's because you can have them make investment decisions for the Robo-advisor platform. You can have them basically do the full stack investing strategy, but then you're almost like you have an AWS kind of thing where you're exposing the same tooling that you would want to build for your in-house investors and giving that to the people who are making the à la carte investments.

[00:30:18] LM: Yup.

[00:30:19] JM: Very cool. Can you talk more about the product feedback loop between those two areas of the house and like what's your system for sort of transmuting product strategy from the Robo-advisor platform to the à la carte investment platform?

[00:30:39] LM: Well, our investment team works very closely with the customers with the investors on the platforms. So they continuously learn what Cadre investor wants from the investment opportunities. What kind of return profile people are looking for? What kind of risk people are willing to take and also what kind of pain points they have in terms of interaction with the platform.

So for example for managed portfolio, we want to automate as much as possible to the point where you just leave us with the number that you're willing to commit and information about your bank account and everything else is done automatically on our side, and we are continuously getting this feedback loop as well as the understanding of sort of what kind of assets our customers are looking for and what kind of portfolio composition is the best for current economics and things like that.

[00:31:35] JM: Then the investors that you have on the team, are they using any like off-the-shelf BI tools or off-the-shelf stuff and you're like, "Man, I wish we had something better to give

them, but unfortunately – Or maybe the BI tools are good enough." Maybe they're using Airtable or they're using Tableau, and that's just good enough, and maybe you can replicate that in your product. What kinds of software tools are they using?

[00:32:05] LM: Yeah. So there are BI tools, like off-the-shelf BI tools that we're using internally, but it's a steppingstone towards the better automation. So we are continuously learning of how their flows are changing our investment team in terms of make their lives continuously better. Then we implement that in our platform.

So like certain dashboards that they need, maybe it's faster to build it in the BI tool because you can do it within 20 minutes, but then it's going to take a couple of weeks to implement it. So if you want to experiment with something, it goes to BI tool. Then if you know exactly what you need or like you already experimented and you learn that this is very useful, we'd go ahead and implement that in platform.

[00:32:49] JM: Earlier you alluded to the emphasis on safety. Safety was something that's really important to your platform. I'd like to know more about why that's important. Is that because are you handling large financial transactions through the company's software or is there some third-party vendor who can like take care of the actual purchasing or the – I don't know, the transaction. Just imagine, these are very high-dollar transactions that are going somewhere, but how much of that is stuff that you actually have to write code for?

[00:33:23] LM: Yeah. So we have to write code for a lot of this processing, the money processing and money flows that are happening on the platform, but not only that. So as I've mentioned, we're regulated pretty heavily by FINRA and SEC and there is was a ward of the regulation that we need to follow. So there are certain security practices that we have to implement. They're not necessarily the most up-to-date. So we also want to be in spirit of the regulation, not only follow the ward of regulation. So we're actually going extra mile to make it even more secure and better. We work with external security consultants in terms of what kind of things that we need to implement on our platform. What kind of practices we need to have on our platform that go way beyond the typical expectations of the regulator.

But as you mentioned, we are moving around substantial amounts of money from people accounts and they trust us with their finance and we cannot compromise their trust, and that's why for us the most important thing is trustworthiness of the platform. People have to be 100% certain that their money is handled with the highest rigor of security.

[00:34:40] JM: What about fraud and risk of some kind of problematic transactions happening on the platform? Is that something you have to build checks and balances and do you have like a fraud and risk team or anything?

[00:34:56] LM: Yes. This is a big part of our know your customer process, because, again, as part of regulation, we have to run the checks on the people who are investors on the platform so we know exactly where their money comes from. It doesn't come from any illegal activities and things like that. So that's a big part of our process. Right now it is partially on our engineering team and partially in our legal team to run these checks. We're slowly sort of automating the processes to the point where for 99% of the customers it can be fully automated, but it is a big part of onboarding flow for us.

[00:35:37] JM: So the two products, the Cadre secondary market, which is the platform for buying and selling individual investments and the Cadre managed portfolio, the automatic commercial real estate investing tool. What's your strategy for dividing up resources between these two products? Because early companies, and I think Cadre is not super early. I mean, it's not like you just got started yesterday, but it's very hard to introduce the second product to a company. So what have you learned about developing a second product?

[00:36:17] LM: Secondary market place for us is a very exciting product just because it's something that has never been implemented online. There is not to my knowledge at least alternative investment platforms who are performing the secondary transactions online. So for us it was always in the back of her head that in order to have a marketplace, and this is the ultimate goal of ours to build this marketplace where individuals can invest in alternative assets. Se have to implement the secondary marketplace.

As it typically happens, you just allocate a certain part of your team to do this sort of moonshot kind of initiative within, and that was exactly the process was secondary marketplace, because

we had a very cross-functional team of individuals. It was all the way from like the business owner, all the way to the engineers and designers who are working on the experience for our customers to work solely on this project, and we allocated a certain amount of time of these individuals to make it happen. But it was a very big challenge for us to understand how much resources we need to put into it and how much of the process we need to automate versus keep manual. It's a very fine balance.

[00:37:31] JM: I'd like to go a little bit deeper into the data side of things. So we touched a little bit on the data cleaning process. Let's go deeper there, though. So I'd like to know more about your process for getting these datasets in the door, cleaning them. To what degree this is automated? If you have any scheduling systems, like if you use Airflow to the detail these data pipelines. Just tell me a little bit more about your data pipeline.

[00:38:00] LM: Yeah. So there are two types of data sources that we work with, is like basically if you think in terms of the ETL. It's pushed-based and pull-based. So some data comes to our system and then the others we need to scrape, and for these we're using the schedule-based systems where we're just scraping the data, we're cleaning it up, normalizing, storing in our data storage and – Data warehouse rather.

Then for the pushed-based, we just have an agreement with certain vendor and we are giving them continuous feedback on the quality, because our machinery when it accepts the data, it runs various checks and normalization so we can give our feedback to the vendor. Make sure that they are doing all the necessary checks on their side before they are giving us this data. Then we're trying to continuously evaluate how useful the data source is for our machinery, for our engineering, so we can invest our time and resources accordingly in the data sources that are actually helping us to make better, more informed decisions.

[00:39:09] JM: Are there any datasets that have surprised you in how valuable they've been?

[00:39:14] LM: There is a lot of interesting information that you can get from things like median income, but the most interesting sources are typically the ones that are scrapable online, information about the business activities around the building, information about the demographics within the certain neighborhood helps a lot making these informed decisions.

[00:39:38] JM: When you say scrape it online, do you mean like from satellite photos or Google Maps or something? What is it?

[00:39:47] LM: Basically you can work with various sources and some of them have, you know, hard rate limits. Some of them don't want to be scraped. So it's another challenge of getting this information.

[00:40:02] JM: So you got to set up some AWS lambda functions and kind of mask your IP address to overcome the rate limiting constraints.

[00:40:10] LM: I don't think it would be compliant with their terms of service.

[00:40:15] JM: I think that's right.

[00:40:17] LM: And we don't do that. Yeah.

[00:40:19] JM: Yeah. There's a company called SafeGraph and a guy I know named Auren Hoffman who runs that company, and I've had a number of conversations with him about this rise of the data company. I guess there's data companies that have been around for a long time, but relative to how many datasets are out there that will be very useful, there's not like a marketplace for datasets. There's not like at the AWS of datasets at least that I know of. Does that seem surprising to you? Why are there not more like data repositories?

[00:40:56] LM: I think part of the problem is that all these data vendors like to keep their data close to their hearts, because for them it's their main means of making money, and that's why you have the vast majority of – Vast variety rather of different formats and different types of data, and it's a very interesting question. It's a very hard problem I think to solve just based on everything I know now about the data, at least in commercial real estate.

[00:41:23] JM: What makes it so hard to solve?

[00:41:26] LM: Just the fact that all these companies are not really incentivized in opening up the information or bringing it to a format where you can just bootstrap this marketplace of data, not at least to my knowledge.

[00:41:41] JM: How do they build a proprietary edge in this data? How did these data aggregators, these proprietary data aggregators, get proprietary data? What's an example of some interesting data company you've heard about?

[00:41:58] LM: CBRE, for example, is a big company, and I think that they just captured a ton of their own proprietary information about the assets they have. There are some companies that –

[00:42:08] JM: Sorry. CBRE stands for what?

[00:42:10] LM: Oh! I actually don't know what they stand for, but RE is definitely real estate. But there are some companies that make partnerships with operators who operate buildings and then gather and capture their information. I guess that's how they bootstrap the data sources.

[00:42:30] JM: How do you figure out which data sources are more trustworthy?

[00:42:34] LM: Well, there is the back testing that we perform on our models and we'll try to understand what kind of weight a certain data source is made in the decision than understanding whether the information was correct or not. It's a variety of ways to do that. You cross compare the data sources to each other and like if two data sources are pointing at one direction and the third is pointing into a completely different direction, that is a good indicator, things like that.

[SPONSOR MESSAGE]

[00:43:11] JM: Triplebyte fast-tracks your path to a great new career. Take the Triplebyte quiz and interview and then skip straight to final interview opportunities with over 450 top tech companies, such as Dropbox, Asana and Reddit. After you're in the Triplebyte system, you stay there, saving you tons of time and energy.

We ran an experiment earlier this year and Software Engineering Daily listeners who have taken the test are three times more likely to be in their top bracket of quiz scores. So take the quiz yourself anytime even just for fun at triplebyte.com/sedaily. It's free for engineers, and as you make it through the process, Triplebyte will even cover the cost of your flights and hotels for final interviews at the hiring companies. That's pretty sweet.

Triplebyte helps engineers identify high-growth opportunities, get a foot in the door and negotiate multiple offers. I recommend checking out triplebyte.com/sedaily, because going through the hiring process is really painful and really time-consuming. So Triplebyte saves you a lot of time. I'm a big fan of what they're doing over there and they're also doing a lot of research. You can check out the Triplebyte blog. You can check out some of the episodes we've done with Triplebyte founders. It's just a fascinating company and I think they're doing something that's really useful to engineers. So check out Triplebyte. That's T-R-I-P-L-E-B-Y-T-E.com/sedaily. Triplebyte. Byte as in 8 bits.

Thanks to Triplebyte, and check it out.

[INTERVIEW CONTINUED]

[00:45:01] JM: How do you develop a back testing strategy? I mean, it's something that I've heard a lot about in like stock trading companies or securities trading companies. I know your background is like in product and like Dropbox and stuff. So how did you get a familiarity for back testing?

[00:45:18] LM: Well, actually I'm not the one who implement that. I've learned my lesson of a manager and I delegate to people who know what they're talking about. But basically, as I've mentioned, the buildings that we have on the platform, we actually control the information flow about this building. So we know exactly how they perform and we can compare it to our models and then back test it, but there are way more competent people at Cadre to answer this question than I am.

[00:45:48] JM: Yeah. So you did used to work at Dropbox in the past. How does the problem set of an investing platform like Cadre, how does that compare to your experience at Dropbox?

[00:45:59] LM: It is completely different. It's a very interesting field that I was not exposed to when I worked at Dropbox. I was always an infrastructure person at Dropbox. I was working on the reliability engineering and data storages. So it didn't really – Like it was really my area of expertise, but there is a lot of things that you take from working on the reliability engineering. For example, doing things manually versus automate, or just trustworthiness and safeguards on the platform and things like that. That was very easily transformable to the investment platform.

[00:46:38] JM: But generally speaking, I guess the problems are so much different, because Dropbox is this really large-scale, large number of transactions. Automation is super important. It's not like the Wizard of Oz automation later manual now kind of format. So it's just the core problems are so different.

[00:47:00] LM: Yeah, it is, and obviously at Cadre we don't have anywhere near the amount of scale and the amount of load that we had at Dropbox, but it's different type of problem. It's a problem of – Like from an engineering standpoint, it's a problem of how to implement things as fast as you can in an incredibly safe manner. From the business standpoint, it's a completely different problem, because Dropbox was operating in a very established market, I want to say. Obviously, they were one of the first storage solutions and one of the greatest in my opinion, but it was already an area where technology was prominent compared to cadre where we're coming into the market that is old-school and technology is not there yet.

[00:47:49] JM: Isn't it amazing when Dropbox came out and like the 5 years, 10 years after it came out, people were continually like, "This is going to get eaten by Google Drive. This is going to get eaten by Microsoft One Drive," and it hasn't really happened. Dropbox is just so much better at that core problem. Do you have any lessons about technology and startups from that phenomenon of Dropbox just continuing to crush that category?

[00:48:21] LM: Well, I think it's actually very applicable to Cadre as well, because in my opinion the reason why that happened was because the focus, because Dropbox knew that like Google has this infinite ATM machine that is called Google Search that can print cash and they don't really need to focus that much on the data sharing and the file sharing part of their infrastructure. When for Dropbox, it was the only way of doing things, and same for Cadre, we

are the platform for individuals to invest in alternative assets and that's the only thing we do and that's why I believe that we can win even though there are way bigger players in this market just because we were able and are able to focus on the task that we're solving.

[00:49:08] JM: Yeah, and the thing is like you can look at Google and say, "Well, Google has expanded into all these other products," and all these other products at Google are great. But when you think about the real cash cow of Google, it is kind of one core product that delivers the most cash. So it's like in terms of cash, it has been rewarding for Google to focus on its core competency. Even though if you zoom out, it looks like Google is doing a ton of different things. Focus in terms of the financial health of a business just seems to – That rule seems to carry over even to a company that seems as expansive as Google.

[00:49:46] LM: Yup.

[00:49:46] JM: And at Cadre, as you build technology for alternative investments – Alternative investments, that's a much broader category than real estate. So real estate is an example of an alternative investment. What are some other alternative investments that might make sense for Cadre to get into?

[00:50:05] LM: Well, the world is very fast in this field and obviously it could be infrastructure, things like that, or natural resources. I don't think we have clear knowledge of what is the next thing and the next big pillar for us right now, because we're so focused on making sure that our commercial real estate pillar is completely automated and scalable. But there is – Again, as I've mentioned, there is a variety of different alternatives that we can open up in the future.

[00:50:37] JM: There's a lot of real estate innovation today. Although, as we said before, the vast majority of real estate transactions remain kind of caught in the past, but you do have these companies like Opencoor, or Cadre, or Airbnb, or redfin, and they're making progress. How do you think real estate will look in 10 years?

[00:50:59] LM: Well, I think it's going to be very different from what we have right now. The analogy that I like to make is it's the same way as taxi evolved over the last 10 years from not being able to call a cap in a city that is not New York and to the ability to get your car within two

minutes and just go wherever you want. So I feel like we're going to achieve the same results were individuals, like on our side, individuals would be able to invest in commercial real estate almost automatically and have their investment in diversified portfolios without jumping through the hoops of the investment process.

[00:51:42] JM: And what are some downstream effects of this? Do you think we'll see more buildings getting built or just more real estate transactions period? More people invested in real estate? What's going to change because of these new technology companies?

[00:51:58] LM: Well, I think it's a little further than that, because you're basically – I think what we want to change is the way how people preserve their wealth, because if you think about that, we are automating things that were only accessible to ultra-rich before, and individual regular folks, they were not able to preserve their wealth the same way, and our goal is to allow them to do that. So like if you are an individual and say you are making $200,000 a year, you're a credit investor, your options to the date are public stock market bonds, maybe an investment account with your bank and like 401(k), but obviously no access to direct commercial real estate or any sort of alternative investments, and that's what we're trying to change and we're trying to give people this opportunity to have a stable financial future. So in that sense, I think that what can get changed is the way help people preserve their wealth.

[00:52:58] JM: Let's talk about your managerial role. So you're the head of engineering. What does that mean? How does that compare to roles like VP of engineering, or CTO, or engineering manager? What is head of engineering entail?

[00:53:14] LM: Well, I don't actually like know what the caveats of the role of CTO, versus VP, versus head of engineering is, because typically companies are just putting the different titles on the same role. But my day-to-day is basically being an advocate for our engineering team and making sure that everyone on the engineering team is happy and everything gets done. So we are delivering on the promises to the business part and the product part of our organization. We are innovating. We're keeping a very low-level of technical deb, and at the same time everyone on the engineering team actually feels fulfilled and they feel challenged and they feel that they belong in our company.

[00:53:56] JM: How do you maintain that? Because there must be work that nobody wants to do. How do you maintain happiness when you have to assign some tasks to people that nobody wants to do?

[00:54:07] LM: Well, again, I think it's a matter of motivating them to do this job, because if you package it as though it's just some busy work, obviously nobody would be excited about that. But if you're telling folks that this is the technical that will eventually make us less flexible, make us less agile, people generally tend to enjoy this kind of cleanups and this kind of process. If you found yourself in the position where majority of your engineers is working on something boring and something not challenging, it probably means that as engineering organization, you're doing something really, really wrong.

[00:54:49] JM: What about hiring? How much of your time do you spend on hiring?

[00:54:53] LM: I want to say 15% of my time. I typically spend on either reaching out to candidates or interviewing candidates or working on the processes around the recruiting in the company.

[00:55:06] JM: And do you have any – I mean, recruiting is one of those areas where everybody just feels like they're doing it wrong. Everybody feels so inadequate, I think, at least most of the people I talk to. They're just like, "We just cannot get enough engineers. We can't figure out how to do hiring right." Have you learned anything about recruiting that you can share?

[00:55:26] LM: Sure. I think that the most important is repeatability and ability to create a good experience for candidates, and then if you have like, for example, Cadre has an amazing team of sourcers who are very knowledgeable in various aspects of software engineering so they understand what kind of profiles we need to be reaching out for. As long as you have all the necessary pipes in place to make sure that candidates experience a seamless, their interview process is seamless and you have a very clear objective way of vetting candidates throughout the interview process, I think that hiring is not really a big issue.

Obviously, it's harder to hire engineers when you're a smaller startup, especially a company without a very flashy tech name. Obviously, I think that companies like Facebook or Google have way less trouble attracting people and making people to apply through their websites. But other than that, I mean we're compensating having an amazing sourcing team.

[00:56:33] JM: One thing I heard from an interview with a Stripe executive was that at Stripe, the interview process consists of interview questions that are designed to look like the job that they are hiring somebody for. So rather than these random whiteboard questions about how to reverse a string or how to make a hash map or something, they'll ask questions about fraud and risk or design a payments API, something like that.

Do you have any unique interview processes that you can share?

[00:57:13] LM: Yeah. So our interview process is somehow similar to that. We do have a phone screening interview where we're just making sure that a candidate is actually capable of writing software, understands the fundamentals of algorithms and data structures and things like that. But then once we bring candidates on site, will actually give them a laptop during the interview and we have two coding sessions where they're being faced with somehow life application issues, not necessarily the application that Cadre serves, but more applied to real-life problems. Then one of the sessions is for them to architect and design the system and it's based on their strength.

So if we're interviewing a frontend engineer, it would be like design this complex frontend. If we're interviewing an infrastructure engineer, we could be asking a question to design the big data storage. Another part of the interview is to talk to one of our product managers around their experience working in a very complicated project from the organizational standpoint so we can understand how engineers that we're interviewing interact with designers, how they interact with product managers, how they interact with business owners and things like that.

[00:58:31] JM: So we talked in detail about the simplicity of the software architecture as it stands today. I think there will be a point in your future where things are going to get more complicated. You're going to have some more complicated models you're going to build. You're going to probably have to use some multiple databases or get at least a more complex data

infrastructure. Are you starting to see on the horizon when that's going to happen or what you're going to need to do to expand the platform?

[00:59:05] LM: It's really hard to predict. We're trying to be ahead of this and there are some indicators that we can measure. So for example, if all of a sudden our deployment process becomes painful, it's a good indicator that something is not going right in our infrastructure, or if we start encountering more and more errors, or if we are compromising the durability and things like that and we are continuously measuring things like that, like how much time it takes us to do the full deployment. How many issues we encounter doing it and so on so forth?

So I cannot predict exactly when it's going to happen, but we're going to see the signs of that way, way before we actually need to make adjustments to the system. Also, there are two different types of complexity. One complexity that can come with the load and the increased number of customers and the other one that can come with the increased number of engineers. So if you all of a sudden needs to cater your infrastructure to 200 engineers who are working on the features at the same time, it's very different from doing the same work for 25 engineers. So we're trying to predict our headcount as well.

[01:00:21] JM: Okay. Well, Lenid, it's been really fun talking to you about Cadre in the present tense and in the future tense. I look forward to hearing more about the company.

[01:00:30] LM: Awesome. Thank you so much.

[END OF INTERVIEW]

[01:00:36] JM: HPE OneView is a foundation for building a software-defined data center. HPE OneView integrates compute, storage and networking resources across your data center and leverages a unified API to enable IT to manage infrastructure as code. Deploy infrastructure faster. Simplify lifecycle maintenance for your servers. Give IT the ability to deliver infrastructure to developers as a service, like the public cloud.

Go to softwareengineeringdaily.com/HPE to learn about how HPE OneView can improve your infrastructure operations. HPE OneView has easy integrations with Terraform, Kubernetes,

Docker and more than 30 other infrastructure management tools. HPE OneView was recently named as CRN's Enterprise Software Product of the Year. To learn more about how HPE OneView can help you simplify your hybrid operations, go to softwareengineeringdaily.com/HPE to learn more and support Software Engineering Daily.

Thanks to HPE for being a sponsor of Software Engineering Daily. We appreciate the support.

[END]