EPISODE 532

[INTRODUCTION]

[0:00:00.3] JM: Farmers have lots of data. A corn farmer needs to monitor the chemical composition of soil. A soybean farmer needs to track crop yield. A chicken farmer needs to count the number of eggs produced. If this data is captured, it can be acted upon. For example, a dry farm can automatically turn up its irrigation system if the data is recorded and the system can respond appropriately, or the data can simply be gathered and studied. If you work in a pure software business, you might take for granted how easy it is to track your metrics. On the farm, you need to use sensors and drones to gather that data.

Mike Prorock is the CTO of mesur.io, M-E-S-U-R.io, a company that makes sensors and software infrastructure for agriculture. He joins the show to describe the use cases for agricultural technology and the architecture behind it. Today's episode is a great complement to our recent episodes on streaming data.

Mesur.io offers a case study in how streaming systems can be put into practice. Mike will also be speaking at the upcoming Strata Data Conference in San Jose, which I'll be attending as well, and I'll also be attending meet ups for Software Engineering Daily, which you can find at softwareengineeringdaily.com/meetup, if you want to register for an upcoming meet up. In March I will be at Datadog in New York, and then I'll be at HubSpot in Boston, and in April I will be at TeleSign in LA. We're going to have some great meet ups with some speakers and people and stickers, probably some food. I hope to see you there.

[SPONSOR MESSAGE]

[0:01:47.1] JM: I listen to a lot of podcasts about technical content, and the Google Cloud Platform Podcast is one of my favorites. The GCP Podcast covers the technologies that Google cloud is building through interviews with the people building them, and these are often unique Google Cloud Services, like BigQuery, AutoML and Firebase.

I'm a big Firebase user, so I try to learn about how it works under the hood, and I want to hear about new features that they're releasing. I also listen to the GCP Podcast to prepare for episodes of Software Engineering Daily, because when I do shows about Google cloud technologies, I'm doing research around them, and I find that the GCP podcast covers topics before I do. So if you want to stay on the leading edge of what is being released at Google and how these technologies are built, check out gcppodcast.com.

I've been a listener for a few years now and the content is consistently good. A few of my favorite recent episodes are the interview with Vint Cerf, who is one of the creators of TCP/IP. He's one of the fathers of the internet, and also the show about BigQuery was super useful. You can find those episodes and more by going to gcppodcast.com, and follow GCP podcast on Twitter.

Thanks to Google Cloud Platform, the podcast, for being a sponsor of Software Engineering Daily. Much appreciated.

[INTERVIEW]

[0:03:20.9] JM: Mike Prorock is the CTO it mesur.io. Mike, welcome Software Engineering Daily.

[0:03:25.4] MP: Thanks so much, man.

[0:03:27.3] JM: You live on a farm and you are a technologist. Tell me what is at the intersection
of farming and modern technology.

[0:03:36.4] MP: Yeah. No, I think it's a very good question. I think this is something that's actually becoming more and more important, especially when you think globally, right? A lot of this ultimately has to do with resource availability, and we haven't — You'll see things in the news or have to cut your irrigation down your lawns and stuff like that, but we in the US are the walking from a water availability standpoint, right? So unless you live in an area where it happened to used to have been desert, by and large in the US you're fine. But as you get out

broader in the world, places like, say, Southeast Asia, you get into these scenarios where 92% of the available water there that's potable and usable for something is actually going to agriculture.

Just that kind of resource availability alone says, “Hey, we’ve got to get a little bit smarter about how do we make decisions around what resources do we use, when do we use them,” and then down the line, “what’s the impact of that.” Everything from runoff of fertilizer or pesticides out into the main water supply to unintended consequences, bolstering — Killing bees off or bolstering up immunity of certain pesticides don't work anymore. So we have to get a little bit smarter at a large-scale with agriculture about how we deal with these things. Really, that's what technology is therefore, is to go through and help us make better decisions.

[0:04:59.1] JM: And to what degree do we need Big Bang new technologies? Do we need an artificial meat patty, or do we need sophisticated desalination techniques, or can we get by with things like more intelligent deployment of sensors, better aggregation of data, better water allocation techniques? Is this simply a matter of blocking and tackling and basic technological implementations, or do we need some Big Bang developments in order to make a measurable impact on, say, water allocation in the Southeast Asia?

[0:05:38.6] MP: Yeah, I think it's a little bit twofold. There is certain Big Bang stuff that we've got to address in desalination or mobile water purification. That is actually Big Bang thing that has to happen, and a good example of that would be like when a hurricane rolls through and things like that. You still have to keep people drinking water and you still have to get things back on track. So we have to do — We do have to make some major innovations on those kinds of fronts.

Artificial meat patties, probably a lot less, and it's not so much about getting better necessarily, though there is always room for movement all on kind of the analysis of the data and the collecting the data, putting it all together in the right place. A lot of it has to do with the democracy of that data and the accessibility to it.

Part of this is especially when you're dealing with technical audiences or like when we’re talking to investors and stuff like that, when you say farm, certain things come to mind. For most people in the US, you think big cornfield in Iowa. Well, the reality is that's not what agriculture is, and in

most cases that it doesn't even resemble that. 80% of farms in America have an average growing size of 6 acres with an average revenue of less than 100 K. So that's a totally different picture than what most people think of when they think of farm.

So the solutions that are out there that say at the top tier or the big major players in agriculture today, what they are doing at that commodity row crop level of corn and soy and everything else, that's good. It's useful to those guys, but it's really tailored to those guys and that solution doesn't scale down to the average farm size in the US, much less the average farm size in the world, and this is where I see the big, both the opportunity and kind of the problem, is the right kind of data to make the right decisions may exist in someone's head. It may exist across a bunch of different books. It may exist across a bunch of different sensors, but people aren't really putting that all in one place. That's really, I think, the root of the problem.

That's kind of why we moved back to the country and got our farm going, that's really what lead me down the path of saying, "Hey, let's build a company that actually does do this," that glues all that disparate data together including a lot of that knowledge wealth from the previous couple of generations that starting to die out and go away. I'm constantly picking the brain of some of the older guys that are running either cattle farms or whatever else around the area by us. Just to kind of get, "Hey, how do you know when to go plant something?" "Well, I do it based off of this — The first thing after Easter, and if the weather is still holding, then I go do this," and it's like, "Well, there's probably some science behind that."

So then I go chase down the science behind that and build a model around that, and if I have the right information feeding into that model, I can actually make a localized recommendation out, and it turns out there's a whole lot of that kind of stuff that we tend to forget about, or is literally dying out, or is going away or being ignored. Part of it is, "Okay, how do we combine this quantitative hard-core real data, like all these stuff we poll from USDA around plaint characteristics and optimal germination temps and soil moisture rates. How do you take that and translate that back into a way that a non-technical audience can get value out of it without ignoring and while going through discovering kind of that qualitative and institutional knowledge that people tend to forget about when they think about it from a pure technical standpoint.

[0:09:04.9] JM: Let's break down what your company does and get to some of the use cases. So mesur.io, which you're the CTO of, you have a sensor that you've built and people on farms can deploy the sensor throughout their soil, and the sensor will gather and aggregate information and shuttle that information back to a central repository where you can create machine learning models, you can basically assess your resource allocation, for example, how often are you watering your plants, or how often are you fertilizing your plants, and correlate that or compare it to other strategies that may present better crop yield or less resource usage. I am describing your company correctly?

[0:09:57.7] MP: Yeah, I think so. Some other things like alternates is a big thing for us. So if you think about just avoiding certain pests, sometimes it can be as simple as, "Hey, did you know if you did X," and I'll give you a good example of like one of our — Like what we call our little kitchen garden, the stuff we grow just for us on our farm. We run marigolds around the outside.

The reason we do that is twofold. That goes through and attracts in all sorts of good pollinators for the rest of the stuff we're actually growing, but it also keeps the rabbits out. So there's little simple things like that that a lot of people tend not to think about that actually make a big difference, because then all of a sudden you're cutting down on your fencing cost and it has all these additional side benefits that you wouldn't otherwise think of, and that's a very simple example, but there's lots of things like that, like from a companion planting standpoint, where if you plan X-plant with something else, they actually help dry and something that benefits the other, whether that's a pollinator or they repel a certain kind of pest that would normally attack, say, tomatoes or peppers. There're lots of things like that that come into play.

Then the other things we're kind of doing, we actually had a really good foothold in golf of all places, and it was kind of unexpected, but we had a bunch of agronomists kept saying, "Hey, have you talked to any golf course superintendents?" It turns out those guys are all agronomists. So they're highly interested in this kind of stuff, and they've got huge one water utilization needs, and therefore knows that they're not having necessarily the most optimal impact back on the environment, and they actually care about that right and they're trying to do something about that. So we make heavy amounts of recommendations around irrigation and irrigation schedule there.

I think an important difference that we do is though, yeah, of course, you could use our platform to automate — Do the full kind of like farm automation kind of stuff or whatever you want to call it, you could do that, but then you take a little bit of the respect for the end user there, because if I'm talking to a guy who's had 30 years in the field and has two or three plus college degrees on the stuff, he actually does something, so I should respect that gave him the ability to do that better and take his input put back in an factor that into our models. So if we make a recommendation about X, I want to find out why he didn't follow that, like, “Hey, I recommend you cut you watering down by 3 inches for these three days,” and he goes through and says, “Well, I didn't do that. I actually kept it at four, and here's why.” That's then feedback we can factor in to adjust the model and like that model learn not just from the quantitative data coming in on the sensor side or the remote-sensing side from either drones, UAVs, satellites, etc., but it's actually also kind of tapping that actual knowledge of something we would never be exposed to otherwise, and we can learn from that user feedback as well.

So it’s really trying to kind of combing all these techniques around dynamic modeling and bringing kind of that optimal self-learning system that everyone's always after. It’s kind of combining all of those techniques and applying it to a very discreet field, which is how do you grow stuff.

[0:12:59.4] JM: The classic; you manage what you measure phraseology, and you think about what we have in software, I mean, many startups do not have good metrics instrumented throughout their servers and their key performance indicators, their churn rate, their infrastructure usage. So many startups are wasting much of money on infrastructure that they don't —

[0:13:27.3] MP: Oh, yeah. They have no idea what they're doing.

[0:13:29.1] JM: They have no idea.

[0:13:30.3] MP: Yeah. It’s a really common — Well, and honestly though, I think — And this is amongst other things, but when you look at what makes a startup succeed, it's a couple of things. One, you got to have the right idea at the right time when the markets are ready, but beyond that, you do have to be 100% dedicated, whether you do it from an OKR perspective or

something else, but you have to be 100% in lockstep on the engineering side with the actual business goals and the end business goals, because if you're not doing that, you're going to burn out. You're going to burn resources where you shouldn't, and it's entirely metrics-driven. It has to be.

[0:14:03.1] JM: And the same is true for farms.

[0:14:05.2] MP: Yes.

[0:14:05.8] JM: But unfortunately it's pretty hard historically for these farms to draw convincing correlations between things that they're doing and outcomes. I mean, they —

[0:14:20.8] MP: They got it, and I'll give you a simple case, and it's really funny. A lot of these actually just has to do with how we as engineers approach problems versus what our users tend to be. I think there's that classic comic that shows what the user requested, what the designer came up with, what the programmer actually built and all that stuff, and it comes together and you wind up with something that's useless. This is actually a large part of the problem, I think, when dealing with agricultural software. There are some great companies out there that have tried to attack this, and have done an okay job for the really big guys, but if you're really big guy, you can hire someone to go run an SAP, ERP and tune it into your farm right, but most guys aren't that, and we have a couple of customers who are doing that.

The vast majority do not have time to go plug in a spreadsheet. They're working around the clock as much or more than we think we work, and they're always on. So the interfaces has to make it, one; as be a simple and noninvasive as possible. So some of that is on the sensor side for sure, some of it is human. Like when my wife, for instance, is going out and — We do a ton of pasture raised eggs and stuff like that. So when she's going out at the end of the day and logging, she had been just logging it in a book, like, "A, today, here is the date. Here's how many eggs I collected." She's not writing down all the weather conditions for that day. She's not writing down any of the other notes that were going that day. So that's useful in formation. You can see what your overall trend was, but she can't see why and what else was going on. So that's one of the features where we actually just cut on that's actually been really well-responded to, is just going through and looking at your logbook of, "Okay. Whether it was eggs, whether it was

something else, what was going on in the weather historically while that was happening?" both on a point basis from our sensors and then combining in other sources to fill in the gaps. So that's turned out to be really insightful to say, "Oh yeah, we know there's an impact of why, but we're not thinking about humidity impacts, for instance, on egg production," and that stuff all actually comes into play.

[SPONSOR MESSAGE]

[0:16:31.4] JM: Your enterprise produces lots of data, but you aren't capturing as much as you would like. You aren't storing in the right place and you don't have the proper tools to run complex queries against your data.

MapR is a converged data platform that runs across any cloud. MapR provides storage, analytics and machine learning engines. Use the MapR operational database and event streams to capture your data. Use the MapR analytics and machine learning engines to analyze your data in batch or interactively across any cloud, on-premise or at the edge.

MapR's technology is trusted by major industries like Audi, which uses MapR to accelerate deep learning in autonomous driving applications. MapR also powers Aadhaar, the world's largest biometric database, which adds 20 million biometrics per day.

To learn more about how MapR can solve problems for your enterprise, go to softwareengineeringdaily.com/mapr to find whitepapers, videos and ebooks. MapR can leverage the high volumes of data produced within your company, whether you're an oil company like Anadarko or a major fin-tech provider, like Cabbage, who uses MapR to automate loan risk and has done $3 billion of automated loans to date.

Go to softwareengineeringdaily.com/mapr to find out how MapR can help your business take full advantage of its data.

Thanks to MapR for being a sponsor of Software Engineering Daily.

[INTERVIEW CONTINUED]

[0:18:13.9] JM: Let's talk about the engineering side of things. So we did some shows recently about edge computing, and this seems like a perfect example of edge computing in action. You take a farm, where you've got a bunch of sensors strewn throughout the farm, it's an open question as to whether there is connectivity throughout that area of the farm. I mean, maybe you've got cell towers nearby, maybe not. Maybe you can do something interesting with mesh networking, where only one of the sensors or a couple of the sensors need to be close to a Wi-Fi signal so that the other sensors are just close enough to —

[0:18:53.6] MP: Yeah. Actually, to tap into that, I think this is one of those things where we, as nerds, tend to try to over analyze things all the time just by default. So if I think about statements, like premature optimization is the root of all evils, there're some real truth to that, and we tend to overcomplicate these things.

In face I can think of a number of people also playing in this space, whether it's, "Oh, we just do soil and moisture probes are, or we just do weather stations," who are going, "Oh, well, we're going to have to roll out this Wi-Fi mesh or this BTLE or [inaudible 0:19:26.0] or some other — Like take your pick of hot network of the day, and that's going to solve the problem.

The reality is if these guys don't have cell phone coverage, at least at the GSM level, then they're probably not functioning too much anyways, and the vast majority, even globally, do you have that coverage. Like you can get into these areas in Sub-Saharan Africa and still have the GSM signal. Now that might mean that you have to think about different ways of sending your data back. It may mean that you're sending in alert back, not through a mobile app, but through a text message, but you have to think about it in that kind of a away rather than saying, "Hey, let's go," because I don't know a single farmer who's going to be like, "Hey, let me go roll out a whole bunch of gateways and do a bunch of edge processing." You would lose them at the first moment that you said, "Oh, yeah! Now, here's these 10 extra pieces of hardware. I know you only made $45,000 last year, but go roll out all these Wi-Fi networks." You would just get laughed out the door. Whereas if you say kind of like we do, "Hey, you take this single device, you put it out there and we'll let you know if it's working or it will let you know if there's a problem and it's not working, but you just go put it out there and walk away."

[0:20:30.6] JM: So are you saying that these sensors that you use at Mesur, they have sim cards and they're just sending text messages to servers?

[0:20:37.7] MP: I mean, not exactly, but we are using e-sims and it's all cellular, and especially now that CATM and some of the new low-power LTE stuff that's coming out, we're leveraging that, but we started out just doing 2G just to get good global coverage and trickling data back. So you have to think about it from like, "How little data can I get back in a reliable way and on what schedule do I need it?" Rather than thinking about, "Hey! Let me guarantee that I've got 100% delivery uptime and all of these and everything else."

That gives you a little bit, like in our case, that gave us a great degree of focus, because we said, "Hey, we may go in —" For instance, we're working with Comcast on a lot of the LoRaWAN initiative stuff with machineQ, that it's an awesome thing. It's got a lot of promise. Netherlands already has LoRaWAN rolled out across the nation. So there's a lot of really cool things you can do with that, but that doesn't mean it's going to be everywhere right off bat. Yeah, 5, 10 years from now as you get things like [inaudible 0:21:36.1] we're updating and stuff like that on the LoRaWAN side, yeah, that might get kind of interesting. But for right now you can do that stuff on cellular already, so take advantage of it.

[0:21:44.8] JM: Fascinating. Okay. So we've solved the connectivity problem. We've got data that's just coming in on a regular basis from our sensors, and so who are you text if all these sensors throughout my cornfield are texting something and that data is getting —

[0:22:02.9] MP: Yeah, we're trickling data back over just regular GSM data or whatever, right? When a message goes out, we're typically actually hitting either the farmer or the operator themselves, or like if you take a golf course case, we may actually send a first pat and then we do this in a couple of cases. We may send a first batch of alerts to say, "Hey, we see the dry down and a burnout area of the 11th stairway.

[0:22:22.9] JM: There must be a server that is at least aggregating that information.

[0:22:27.2] MP: Exactly. So this is where we do —

[0:22:29.0] JM: So you're texting Twilio, or AWS, or what exactly?

[0:22:33.1] MP: Yeah. I mean, we run all the data, and I wouldn't think of it as texting, right? Text come outbound for us. Whereas we trickle data in over just regular data packets and UDP and things like that, but that all makes its way back into AWS and Google Cloud Platform and things like that. We're entirely Kubernetes now. I guess we still have a little bit of Docker Swarm stuff that hasn't been fully merged into that new infrastructure yet.

Yeah. So we're leveraging the right cloud platforms for the right things. Starting things are way cheaper on storage. Use those things. When you think about imagery, especially coming off of drones, you want to cheap storage. Yeah, we bring all that stuff together, and so we have some intelligence about how we collect that data and how we route that data in. It may come in. Typically, we're routing in, or pulling stuff in or listening for stuff in a variety of different ways, and then dumping that into a bit old Kafka stream. It's just getting into Kafka.

Once we get it there we could start to do stuff with that. You have different things that you need to do with that data. Some of it needs to be point in time decision, "Hey, this was a standard deviation above. Therefore I'm going to send an alert on that," or hit some rule that the user said, "If I see a drop in pressure, that means I've got a leak on this well line. So freaking send me alert even if it's 3 in the morning. Wake me up so that I don't lose a bunch of money." Those are the kind of things that once we get it into our infrastructure, that's really where the real work gets done.

[0:23:58.2] JM: Okay. So Kafka. You've got all your events getting logged into Kafka on these topics. Maybe you've got a topic for each different farm or something like that, and then what do you have doing the streaming workloads of the data enrichment? Are you using Spark Streaming, or Kafka Streams, or Flink?

[0:24:15.4] MP: Yeah, we got a little bit of Spark streaming going. We're actually doing some interesting stuff with the rules as well just because it lets us define business rules in a fairly easy way for certain kinds of stuff. Most of it were actually just doing little micro-service consumers in Go though. So most of our codebase now, like if we're to think about our codebase and what we're writing in, yeah, we're doing machine learning stuff.

[0:24:36.9] JM: Because the data volume — I guess the data volume for a given farm is not overwhelming, right? How much do you really need to sample soil? If you could sample it every 15 minutes, and that probably is enough and that's not an overwhelming amount of data.

[0:24:54.4] MP: You got it. Obviously, this kind of scales up. There's a lot of farmers out there. There're a lot of golf courses out there. We actually have some pretty large amounts of data that we pull from public sources as well, so like USGS Sampling Stations and stuff like that. The volumes do start to stack up overtime, but it's really by just taking a very clean, like, "Hey, we're going to publish something out to a topic," and then a couple of different consumers are going to pick that up and decide if they need to do anything with it or not. That lets us very discretely kind of break out these functions and build pretty quickly in the event that, "Hey, we got something wrong. Let's go swap this out. It's not the end of the world."

[0:25:30.0] JM: What about Google Cloud versus AWS? So the things that I have heard about Google Cloud that are super appealing are BigQuery, for one, and the APIs, for example, image recognition or speech-to-text, things like that.

[0:25:48.5] MP: Yeah. We're not really using any of that on our —

[0:25:50.1] JM: You're not using any of that.

[0:25:51.3] MP: Yeah. We're using them because everything is fairly outside of some bare metal hadoop or "bare metal", if there is bare metal today in most startups. It's really just, "Hey, we're virtualizing some whole machines," but we're containerized outside of our Hadoop stuff. Even there, a lot of our Kafka stuff is — Like our producers and consumers are all containerized.

We've taken a pretty heavy container first style approach for a lot of different reasons. Once it kind of got stabled, we held off on that for up until a couple of months ago and then starting cutting stuff. We had it ready to go. We're developing that way, but we held off on rolling that out in production.

We were kind of ready to go, started rolling it out on production. Where Google Cloud really wins out on my side is just from a management standpoint is really slick. Obviously if you're doing stuff that's defense related or whatever, you don't necessarily want to be doing certain kinds of workloads up there, because AWS has that space pretty well —

[0:26:46.1] JM: So are you using UI, or are you using GKS, or the Google Kubernetes, the hosted Kubernetes —

[0:26:53.8] MP: Kind of all the above there. In particular storage, that's a big thing for us. There are certain kinds of workloads we do, like raster image processing which I probably can't go into too many details on how we're attacking some of that stuff. But at a top level, think about if you get a satellite image in or a UAV or a drone image in, that's multispectral. You could tell a whole lot from the different wavelengths that you're looking at, like a weed is showing up in a certain kind of area. So if you could process that kind of imagery very quickly and determine where a problem is, you can actually then extract out a recommendation for someone to actually go out and deal with that problem in the real-world. That's the kind of stuff we're typically work loading over in Google, is that kind of raster processing, the image storage, things like that.

The rest for us right now is all in Amazon, and they've been great. They're hugely supportive of us and we've had just great experience rolling that that kind of stuff out there. But certain things like S3, from a storage perspective, they're good. They're very well-acknowledged. We use S3 for exchanging policies, some of the bigger guys. Really, it's not optimal compared to the way Google approach a storage at a lower level.

[0:28:10.8] JM: That's because of price, or because of the UI, or what exactly?

[0:28:14.4] MP: Mostly pricing is the biggest thing. I mean, if you think about like just a flyover of like 100 acres, you could easily be quickly looking at 11 gigs, and if you're doing that a couple of times a week, that starts getting up there quick. Then when you think about that's your starting input and then you generate a whole bunch of other assets out of that, like, "Hey, let me go generate out some kind of NDVI type output that looks at the actual vegetative indexes across a given area," or maybe looks at sub-spectrum out of infrared. Those are the kinds of things like, "Oh, yeah. I got this one image pack in. Now all of a sudden I'm busting it out into all

these other things that are six to 12 gigs," then I'm doing that twice a day in some cases or even only doing that a couple of times a week across a bunch of customers. That starts to stack up pretty quick.

[0:29:02.7] JM: Drone image workloads, it sounds like are a really big percentage of your workloads.

[0:29:08.7] MP: Yeah. For actual compute side, that's actually probably the largest amount of compute that we're doing, followed by kind of ongoing modeling and learning from the data and comparing results we got or weather forecast that we received in from take your pick, a different service, to see what actually happened so that we can start to score and wait and prefer whether forecast for individual locations. It's kind of like the — You have the one workload that's very heavy raster-wise, and that's just the volume of the data. If a pixel if representing 10 centimeters or less, and you may have 12 different pixels according — Or five different pixel accounts for that same area, you could stack up just a lot of volume very quick without meaning to.

[0:29:51.1] JM: Have you had to become a drone operator, or do you partner Airware or something?

[0:29:55.5] MP: No. We have a guy actually who is a great drone operator who kind of came out of that space and knows GIS very well, but then we are partnering in with other kinds of providers on that side. We take in a lot of data from PrecisionHawk, DreamHammer, folks in that space are typically —

[0:30:13.2] JM: What do those companies do?

[0:30:14.1] MP: They basically operate fleets of drones and UAVs and then crunch the raw initial outputs out of that. Think of it as you can at least get a single file that shows what that flyover is, rather than tons and tons of images, and then we kind of pick it up once you got that single file, which is usually in GeoTIFF format, and we kind of pick that up from there and then do all sorts of stuff to it.

[0:30:37.2] JM: What's the state of regulation in those kinds of drones? Do the drones have to be relatively positioned close to the farms, or can you have like a single drone that just like handles a bunch of different farms in a close area?

[0:30:49.8] MP: Yeah. We're seeing a variety of different things. Actually a lot of the crop consultants who are going out hands-on anyways are starting to get into this, are partnering with guys who will go out and say, "Yeah, I've got a drone. You've got six farms you're working with in this side of the county. So we're going to go out all day and just fly over all of those places."

So you'll see that kind of thing, but you have a lot of line of sight that comes into play, and so this is kind of the limiting scaling factor from a regulation standpoint, is by and large, and don't take this as a heart statement. By and large, the assumption you should make is that if you're flying a drone of any kind of commercial weight or size or for a commercial purpose, that you've got to have someone who's licensed and that knows what they're doing, because otherwise you're just going to burn a whole bunch of money up, and that they're going to be more or less within line of sight of that device at all times. That really limits a lot of kind of that crossing the horizon thing, where really that's the limitation.

[0:31:43.9] JM: There are not autonomous drones.

[0:31:45.4] MP: Well, there's some. I would view them as semi-autonomous. So they will fly on their own if you let them. But the way the regulatory bodies are right now, it's like you may go in and program in your flight plan and let it make adjustments based on weather and crosswind conditions and stuff like that, but at that point you still have an operator sitting there and watching it ready to make corrections if there's an issue.

[0:32:06.4] JM: Are people using this data to set up automatic responsive systems, like where they have some data feed and there's data coming in and it's like, "Oh, my crops are not doing well. They're dehydrated," and I have this automated system where I can take in data and the system recognizes that things are under-hydrated and then automatically trigger, or my sense is that we're quite far from that kind of automation and that there's still a whole lot of manual intervention required to turn on the faucet and make more water come out.

[0:32:43.9] MP: Yeah, that's actually not — We're a lot closer to that than you might think.

[0:32:47.3] JM: Really? That's great.

[0:32:49.0] MP: Yeah. We've done some really good demos of that, and it works quite well. But what you find is that humans are kind of corky animals, and we do like an element of control or at least the thought that we're in control. So most often what we're seeing instead of, say, full automation, is that someone will go through and schedule their irrigation to come on at a certain time, but then we, from an automation side, will interrupt that. We'll either prevent it from coming on because it doesn't need it, or we'll shut it off early when the right amount of water has been applied.

So there's still that automation — Totally automated. Someone went in at the beginning of the week and plugged in a schedule and walked away, and then we control shutting it off. But there's that perception of control, that feeling of control by the actual operator, whether that's a superintendent, weather that's a farm manager, whoever. Same thing with feed application and stuff like that.

That's really more the type of automation that we're seeing, or at least that we're seeing our customers respond well to, because initially me being me, and I think with this audience, most people sympathize with this. it's like, "Oh, man! We could just go in and automate all of these and then I can go work on something else instead of having to go flip a hose on or whatever."

I also tend to philosophically, from a design perspective, take the assumption that I really don't know anything even if I think I do. So rather than coming to a customer with a product and saying, "Hey, here's what we built, and it's awesome because I'm a great program. Therefore you should use it this way." I usually tend to approach them and say, "Okay. Let me talk to the five new guys we're thinking of onboarding, and let me see what they're pain points are first. Then we show them with real data what's going on and what's some of our capabilities are, but post it in, "Well, what would you like changed about this? What would make your life easier?" Whether that's an automation process question. Whether that's, "Are we not showing you a data point you want to see." Like we're flipping on a growing degree days right now, which is a huge thing from a pest emergence standpoint, planning out of certain kinds of like growth

management type chemicals, and that's something that we found out just by user querying, like, "Okay, great. We've got some good users on board. We're making some money. Let's go back. Let's survey them. Let's fill out what our needs are from our top guys and from the guys that aren't real happy with us and see what the pain points are and see what's making their lives difficult and then tailor around that."

By taking that assumption that we don't actually have the final answer already, that's been hugely valuable to us from an adoption standpoint, and we've seen a lot less resistance coming into markets that aren't necessarily the most technical markets. When you got your 65-year-old neighbor who is still running a tractor that was his dad's from 1935, and I'm using an actual real case here. This is not the kind of guy who's necessarily running around with the latest iPhone, right? So can you actually make something that is accessible to that user into the next generation coming in and the next generation coming in?

What's interesting when we think about these age and persona profiles, especially kind of our target market is sustainable and organic ad at that midsize, because that's who really needs this to start and it can go from there.

[0:36:00.2] JM: So these are the people who are like selling to farmers market and stuff.

[0:36:02.7] MP: Got it. It's the farmers market. They're selling to the farmer table restaurants, that kind of thing. Those guys, what's interesting there about the profile is it's actually a much higher amount of women using our products there. They're the actual runs running the farm day-to-day and a lot of times it's the granddaughters.

[0:36:22.4] JM: The millennials.

[0:36:23.5] MP: You got it, and so it's really interesting, because they actually are used to this notion of, "Hey, I'm just going to go in, and, oh great, you got a little web app I can just add to my home screen and we're actually —" Don't quote me on this, but in April we should be popping out in the app store for a more broad — Various app stores for more broad adoption, but we're managing that closely. That's another lesson when we think about software, just because something is working or it's there. Don't always think, "Oh hey, what's that end state?

Let's control it," just like you do back to your investors, just like you do to your internal teams, your management side on the business side. Always control that message and control and adoption so that you don't overwhelm yourself right off the bat without meaning to. Make sure you're making money, obviously, right?

[SPONSOR MESSAGE]

[0:37:12.9] JM: Apps today are built on a wide range of backends, from traditional databases like PostgreS, to MongoDB and Elasticsearch, to file systems, like S3. When it comes to analytics, the diversity and scale of these formats makes delivering data science and BI workloads very challenging. Building data pipelines seems like a never ending job as each new analytical tool requires designing from scratch.

There's a new open-source project called Dremio that is designed to simplify analytics on all these sources. It's also designed to handle some of the hard work, like scaling performance of analytical jobs. Dremio is the team behind Apache Arrow, a new standard for in-memory columnar data analytics. Arrow has been adopted across dozens of projects, like Pandas, to improve the performance of analytical workloads on CPUs and GPUs. It's free and open-source. It's designed for everyone, from your laptop, to clusters of over 1,000 nodes.

Check out Dremio today at dremio.com/sedaily. Dremio solved hard engineering problems to build their platform, and you can hear about how it works under the hood by checking out our interviews with Dremio CTO, Jacques Nadeau as well, as the CEO, Tomer Shiran. And at dremio.com/sedaily, you can find all the necessary resources to get started with Dremio for free.

I'm really excited about Dremio. The shows we did about it were really technical and really interesting. If you like those episodes or you like Dremio itself, be sure to tweet @dremiohq and let them know you heard about it from Software Engineering Daily.
Thanks again to Dremio, and check it out at dremio.com/sedaily to learn more.

[INTERVIEW CONTINUED]

[0:39:14.6] JM: Let me ask you this, because you're closely involved with some of these farms and you operate a farm yourself. When I go to Whole Foods or I go to Buy Right or I go to some place where I am buying my organic produce or my organic meat, what do those labels mean? Do they actually mean anything and how safe is that supply chain? Are there interlopers in the supply chain who are diluting my organic produce such that there is really no purpose to me paying that 40% premium on organic produce?

[0:39:51.0] MP: Yeah. There're a couple of things here. Some of it depends on what it is. I will tell you that for sure. A classic example is milk. If you're buying milk in the US and you're not buying the store brand, you're wasting your money. It's that simple, because it can only travel so far. It's processed same way. Yeah, you might have some better grade over it. Yeah, this was all grass fed and that genuinely does make a difference. Whether or not it's certified organic or not, it actually doesn't have quite the impact.

My favorite example of this that just kind of drives me nuts and blows my mind at the same time is when you see things like free range on poultry of any kind, whether that's chickens or any kind of poultry byproducts.

[0:40:31.1] JM: Or vegetarian diet.

[0:40:32.8] MP: Yeah, exactly. What free range means in a chicken context for actual chicken farms if you go out to a real poultry house is that there's a door to the outside somewhere.

[0:40:43.9] JM: Right.

[0:40:44.7] MP: I'm not kidding, and there's a little run outside. There's an FDA and USDA specified amount of space that needs to be available. Now if you've ever been in a chicken house you will know that those chickens have never been in the outdoors. They are terrified to go in the outdoors and, yeah, one might get kicked out by accident, but they will never be outside and that's just because of the way we have to grow a certain amount of food and raise a certain amount of food to sustain our population. That's just the raw reality of it. So this is why I tend not to come down on that kind of stuff as hard as probably some other people. So even though I may have my ideal case, like in our case, we actually pasture everything. The chickens

are in the coupe at night and that's it. Everything else they're out in the field, eating bugs, and having a good like, and we like that.

[0:41:33.5] JM: Do the eggs taste better?

[0:41:34.3] MP: We have people who buy from — Oh, it absolutely does. You could tell. Especially with eggs, you'll get these bright RNGOs just from the extra beta carotene that's in the diet, and you see a different throughout the year seasonally as well. So it's a really — It's something you can actually tell the difference on. This is why I have trouble eating catfish out at a restaurant too just over the simple fact that it tastes like corn. Once you realize that almost everything is being corn fed, everything kind of tastes like corn. So just like me, personally, I have my preferences. I have my ideals. We have our people that, you know, in RCSA and stuff that we work with closely on that side, but obviously that can only scale so far. Now if we were to make major changes to the way we distribute and manage food and everything else in this country, sure, that might be a different story, but I would be skeptical of our ability to do that without some major technological breakthroughs.

[0:42:27.5] JM: What I'll say is on the occasion that I go to the farmer's market, and to a lesser extent when I go to Buy Right, which is a smaller grocery chain in California, in San Francisco at least, I do notice the difference in taste. What I wonder about is Whole Foods, because Whole Foods has organic food and some of the listeners, by the way, are going to kill me by now, because this is so off topic of Software Engineering, but I just have to ask you because you're in the know.

[0:42:52.0] MP: It's totally fine. I mean, I'm actually loving this conversation, because this is something I'm really passionate about, because this is something that I think we as a society have to think about a little more, and I had a great conversation with, let's just say a senior person over at USDA, and one of the things — This is a technical guy over there, and we had a great conversation about this exact topic. One of the things that people don't —

[0:43:12.2] JM: To be clear. Is the Whole Foods supply chain management.

[0:43:16.3] MP: Whole Foods supply chain management and what organic as a label means. There are certain things like, say, tomatoes, that it's probably safer for you from a food poisoning standpoint to not buy organic. So the actual health impacts to you are probably better to just buy whatever old fashion meat, just regular stuff, and a lot of that has to do with cross contamination and handling and what's used as pesticides or as fertilizer. So that's something that a lot of people tend not to think of, but the "organic industry" has created this perception that organic is somehow automatically better, and what's interesting to me is like we have a group by in North and South Carolina called Carolina Farm Stewards Association, which is very well-focused on what they call sustainable agriculture, and you have this out in California too and kind of all across the US.

The focus there is on doing things in a sustainable way whether or not your label organic doesn't really matter, because at this point, because we had to open the regulations up so much on the what it means to be a organic, that there's a very low bar to do that. So there is definitely a difference when you go out and buy grass-raised fed and finished beef. But if you know how to go find that without talking to your butcher who don't knows his supplier and you're trying to buy that at Whole Foods, good luck on you, because you're not going to find it, or they'll be mislabeled or whatever else.

That's why I think we have this kind of perception versus reality disparity where we think because of marketing, literally, that organic is somehow this magic thing and that if I buy this same thing at Whole Foods versus in rural North Carolina, it's Food Line or Piggly Wiggly are the two grocery stores that are everywhere, you literally will go to a Whole Foods and you're buying the exact tomato from the exact same farm that you're buying somewhere else or collard greens of whatever, and I know this because I sell into — Like I don't sell to the grocery stores, but I know the guys who do, and we typically sell farmers markets, restaurants and caterers, but like I know the guy up the road who grows the collard greens for most of that area and they might use a slightly different set of fertilizer on one field just to match the organic regulations and go through these inspections on that field, and that's why he's got to charge more for it, and that might be what he sells off to Whole Foods, but chances are once you've gone through that process, you're just going to go do that everywhere so you're selling the exact same thing in both places is the reality.

[0:45:50.2] JM: All right, Mike, we're close to the end of our time, because I wish I could keep talking to you, but I got to get going in a bit. I want to ask you about a few futuristic areas. I guess the one I really want to focus on that I'm most interested by is vertical farming. Tell me what you think of vertical farming. Maybe you can define the term and just summarize why vertical farming is important.

[0:46:11.1] MP: Yeah. I'm not going to necessarily summarize why it's important, but I'll discuss that topic. So vertical farming for those of us that just sit behind a computer all day is this approach to say, "Hey," whether it's indoor or outdoor or whatever, "let's go through and start to stack stuff up and grow stuff in an up and down way as opposed to a side to side way."

Now the first real good example of this that I'll bring up, and this might be a difficult topic for those of us that actually bond with their livestock or whatever, but if you ever go to a hog farm, there is a company in North Carolina that produces most of the hogs in the US and they have locations everywhere, but they largely invented this notion of vertical farming where you could stack animals on top of other animals with the right kind of confinement and containment and get them to grow to the right size to go to market, because you don't have enough space literally to do that otherwise. This actually, in a buyproduct way, ended up having a positive impact, because all of a sudden you had a whole lot of waste that wasn't just spread across all sorts of fields and running into runoff. So there are both positives and negatives to this. Where most people now, from like, "Hey, this is cool now," tend to think of it is in terms of folks that are stacking up like veggie garden or spring green gardens up and down specially in inner cities, and that's where I think the important and cool side of this is, is when we think about our food web or our supply chain around egg and food, we have a big problem here in this country, which is most of our stuff is coming from like California, or if it's pork it comes from North Carolina or Iowa. We don't live in those places. We live across the country. So what happens when that supply chain get interrupted? Well, of a sudden grocery stores that only have three days of food on the hand for their area are starting to run out of stuff.

So this is actually food security concern, and this also as a byproduct because of extra shipping has all sorts of other impacts on the environment at their own cost and on — You name it, the rest of the things around supply chain. So if we could start to bring some of that farming back in, whether it's on an inner city perspective and do that in some really cool ways that we haven't

thought about otherwise, whether that's vertical farming or aquaponics of whatever else, there are some really neat things as a potential from that. One, you could start to potentially solve some of the food dessert kind of problems where, "Hey, all that's available is McDonalds," you could start to provide additional jobs without meaning to just as a byproduct. You could do things like community gardens in a lot easier way, areas that wouldn't otherwise have that. So I think that's something like I personally I'm highly interested in. I've had some great conversations with folks over at USDA who are highly interested in that idea of how do we start to localize our food sources in a little better way.

So that's something that I'm very passionate about watching how that grows. That's also something, because it tends to be confined that you have to do in a semi-automated way. You have to do it with a certain amount of data knowledge, because you're controlling inputs, you're controlling the amount of light if you're doing it indoors or controlling what goes into the water if you're doing it in a hydroponic way. That's where our technology starts to really have some fun, is because then you could start to apply it in an even broader way and kind of get towards that full automation thing we're talking about earlier. Then that automation ends up becoming a benefit as supposed to a hindrance to you doing your job. So it is definitely a really interesting topic and I could probably go back and forth with you for a while on it, but we'll run out of time if we do that.

[0:49:45.3] JM: Absolutely. Mike, it's been really great talking to you. The time flew by. I look forward to seeing you at Strata, I believe. You're going to Strata, right?

[0:49:53.5] MP: Yeah. I'm going to be out at Strata San Jose, and it's funny because I got double booked also talking at Gartner at the same time. So I'm sending my CEO down with a technical handler.

[0:50:02.8] JM: There you go.

[0:50:03.7] MP: I'm coming out to Strata. So it will just be a replay. If you are at my Singapore, you can come out and ask different questions, But the Strata Singapore, it will be a replay of that. Then — I mean, I'm trying to think if I had any other interesting ones coming up —

[0:50:16.0] JM: I've never been to a Gartner Conference. Are those fun?

[0:50:19.0] MP: No.

[0:50:20.3] JM: Okay.

[0:50:22.3] MP: They're valuable.

[0:50:22.9] JM: Valuable. Okay. All right.

[0:50:24.5] MP: Depending what you're doing. Fun isn't necessarily the right way to think about it. They actually can be — When you think from a practical business and cost impact on how do you make decisions about build versus buy with software and tie all that stuff together, that's where that stuff becomes really, really awesome, because you'll see things like, "Hey, let's understand what this new technology is, and then let's not do that ourselves. Let's make sure we avoid that danger."

[0:50:52.0] JM: Someone was telling me they pay $90,000 a year for a Gartner subscription, and I was like, "Oh, okay. That's how Gartner makes a lot of money, because you get great information from it."

[0:51:03.3] MP: When I was working more on the consulting side with the big Fortune 50 guys, let's just put it that way, that stuff is taken very, very seriously and for a good reason. There are certain things that they cover very, very well. Like Cindi Howson over at Gartner who does the data analytics and data discovery magic quadrant. She's phenomenal She's got a huge amount of insight into that space in particular. I love sitting down and talking with her, because when you're talking about analytics and understanding what's going on, there's stuff going on worldwide that you just otherwise would have no idea about.

So it's a great way to say, "Hey, let me break out of my box." This is the same argument for like pro-diversity in the workplace and in engineering, because as soon as you start to break your core mindset, that's when you start to get really successful and come up with cool new things that you otherwise wouldn't, and it feeds into that same thing.

[0:51:51.9] JM: All right. Mike Prorock, thanks for coming on the show. It's been great talking to you.

[0:51:55.4] MP: Yeah. Thanks so much for having me, Matt.

[END OF INTERVIEW]

[0:52:00.0] JM: If you are building a product for software engineers or you are hiring software engineers, Software Engineering Daily is accepting sponsorships for 2018. Send me an email, jeff@softwareengineeringdaily.com if you're interested.

With 23,000 people listening Monday through Friday and the content being fairly selective for a technical listener, Software Engineering Daily is a great way to reach top engineers. I know that the listeners of Software Engineering Daily are great engineers because I talked to them all the time. I hear from CTOs, CEOs, directors of engineering who listen to the show regularly. I also hear about many newer hungry software engineers who are looking to level up quickly and prove themselves, and to find out more about sponsoring the show, you can send me an email or tell your marketing director to send me an email, jeff@softwareengineering.com.

If you're listening to the show, thank you so much for supporting it through your audienceship. That is quite enough, but if you're interested in taking your support of the show to the next level, then look at sponsoring the show through your company. So send me an email at jeff@softwarengineeringdaily.com.

Thank you.

[END]