**SPE-GCS Podcast – Gustavo Sanchez**
**10-05-21**

**Marty**:  Hi everyone, and welcome. I'm Marty Stetzer, president of EKT Interactive in Houston. We're proud to be the podcast sponsor with the Society of Petroleum Engineers, Gulf Coast Section. The SPE Section was founded in 1935 and now has over 11,000 members. It's a volunteer organization that provides member forums to upgrade and maintain professional competency. This podcast is one of a series and another learning resource available to the members. Numerous on demand webinars can be accessed at www.spegcs.org.

Today, our topic is the importance of a solid digital infrastructure before you try to optimize any system. And I'll be speaking with Mr. Gustavo Sanchez, the co-founder of Panda Tech in Houston, with over 10 years in the oil and gas industry. We're really happy to have his input on this topic at this time of unprecedented challenges in our industry; especially with numerous upstream and downstream companies looking at the advantages of digital tools like machine learning and AI to help optimize their business.

Gustavo, thank you so much for taking the time today.

**Gustavo**: Thanks, Marty. It's good to be here.

Marty:  Gustavo, to get started, can you share with our audience how you got interested in this topic in the first place?

**Gustavo**: Yeah, Marty, and I like to go back to even before my energy experience. When I graduated college, my first job was in back office financial operation, doing data work. So what I mean by back office, we were managing very complicated type of accounts. And, for example, if you owned the account, Marty, and you wanted to make a trade or something, there was an army of people in a warehouse doing all these data processes. And this is when I first realized the importance of infrastructure, as far as digital, because we were doing all kinds of crazy data work, some of it in the workflow, some of it in the policies, management and process.

And this is when I first saw the data quality problem that I'm keen to. Okay? So let's say, Marty, you wanted to do that trade, and someone in the back had, let's say, fat fingers. And through all those data processes, there would be a missed input. And that mistake, depending on the input, it was really just depending on the input, would be a dollar or millions of dollars, right? And the way we would solve those issues is by adding more workflows, but at the end of the day, they were band aids to solutions instead of holistic ways to solve the problem.

Now, moving forward, this is about a little bit over 10 years ago, moved to Houston, back in the oil and gas business. And I was doing market analysis and business development for mid- to small-size service companies… then wanted to do business in Ecuador, Colombia, and Peru. I was representing them and doing some of that market analysis. And there was something else that I thought data would be great, as far as to help me with my job.

And let's say I would sell some sort of pump or skid system: Let's use Echo Patrol as an example. And then the Echo Patrol engineer would call me and say, "Hey, Gustavo, the pump, the system, it's having issues. It's not working." And obviously, the first thing they blame is a bad system. You could get blackballed, big issues. And the supplier has spare parts in the warehouse in Louisiana… and it's down for a lot of time. So the first iteration, the first idea of the company that we created was to give the

salespeople for small and mid-sized service companies notifications on their phones that allow them to provide aftermarket service. Now, that's not what we do. And I'll tell you the pivoting story later, I'm sure.

But then, move now, and what we do is data quality, right, speaking about digital infrastructure. And what that means is that we help companies, energy companies and federal organizations that manage thousands of sensors, collecting millions of data points, reduce the time and the cost it takes to clean and validate this time series data. And it's important because 80% of a data professional's time is spent cleaning and validating the data. And it doesn't matter what people say. Their AI algorithms or any algorithm outside of their sandboxes will be unreliable unless you have robust validation quality pipelines along other types of infrastructures.

**Marty**:  Gustavo, you mentioned the importance of a digital infrastructure. And in your example on trading was awesome. I'm very familiar with that side of the business and there are thousands of transactions every day that need to be carefully matched. What do you see as the key components of this concept?

**Gustavo**: Yeah, sure. And there's many, many things, but there's I would say five things that I like to think about and lose some sleep over sometimes. In a little bit of a priority order – I would say we have data quality, we have scalability of products, we have cyber security, data gathering and pipelines, and the overall usability of both systems.

Marty:  Gustavo, can we now go through each one, with you explaining why it's important and any client example… starting with data quality?

Gustavo: Sure, of course, Marty. So I always like to give this example, and I'm not going to name any company names or anything… but in 2016, 17-ish, it was one of the biggest companies in the world had launched a giant digital platform. And in our industry, in the oil and gas industry, this company had to sell all their oil and gas assets. What they did is go to the technology before going to the infrastructure, right? So let's optimize maintenance, let's do X, Y and Z without thinking of the infrastructure. So there's many reasons, but that's one of them.

So as we're pitching our main idea that I talked about before, before we pivoted to data quality, first, we realized that, okay, the small to mid-size, we may be early for that. They're not ready yet. Tough sales, especially in 2016 – absolute crisis. So we pivoted to more of a general predictive maintenance thing, solution. Again, very competitive. We were about a year late to that party and the big company got it wrong and a lot of other smaller companies also got it wrong.

So we get, eventually, to an offshore drilling contractor, one of the biggest ones in the world. And they have built a wonderful, wonderful digital twin to predict the maintenance and optimize the operation of their blowout prevention equipment. - a wonderful algorithm from some of the most brilliant minds in mechanical engineering that I've met. And they say - you know what? We have this own digital twin that we've developed internally. We can't give you that predictive maintenance aspect, because two or three companies already failed for us.

Instead, we have this problem, this algorithm. We know it's good, but it's not reliable. And we have a feeling it has to do with the data in theri PI system. We think it's trash. So if you can solve that

and identify what data is trash and why – aka validate that that data is useful for the purpose of that algorithm – then we'll work with you.

So that's when we actually pivoted to data quality, right? So just like in anything else, if you have bad data in, you're going to have bad data out - garbage in, garbage out. And that was true for them, and it's true for the entire industry, right? So instead of going directly to the maintenance, this company decided to focus on some of the infrastructure to make those production algorithms reliable.

**Marty**:  Thanks, Gustavo. That was terrific. Now, what about scalability? Again, why is it important and is there a client example that you can share?

**Gustavo**: Of course. So scalability is not so much infrastructure, but it's a combination of different things that allow systems to be scalable, right? So I'll give you… going back to that same example of technology before infrastructure and process. I was talking to a person that worked in rigs for the major, major oil producers. And the same big company had a big pitch event for the company. And essentially, what the users at the rigs realized is for them to deploy all these maintenance algorithms, they had to deploy small sandboxes in every piece of equipment. Now that's not scalable, right?

Every time we are going through a sales cycle with one of them, with an oil operator or any sort of energy operator, we eventually have to sit down with a dev ops department. And if it's going to cost them a lot of money to deploy and maintain, then there's ROI in the, let's say, the maintenance side of it, but there's no ROI, or the ROI cancels out, because of the cost of developing, deploying, and maintaining. So you have to keep that cost in mind.

So the way we like to think about this is everything we build at the lower level – the smallest building block, the lowest Lego block – is the API. Meaning, we build little Lego pieces for absolutely everything we do. On top of this API, we build applications. So putting back to the Lego thing, we put the Lego together and we build the house, or whatever. And these applications… we have to make sure that from a central location of the user's choosing, it can be deployed on cloud, on the edge, and managed and deployed all from a central location of the user's choosing. So if that's not considered, then you don't have a place to put your AI algorithm.

**Marty**:  It sure makes sense to me that cyber security is a big part of this whole initiative, as you're managing thousands of bits of data across a network.

**Gustavo**: Absolutely. Cybersecurity is huge. And we define cyber in two ways. So think about the word, cyber security, which by the way, it's one word or two words, both is fine, how ever you want to spell it. But the *cyber* is digital, *security* means risk reduction, right? So that's pretty broad.

So there's two types of risks that we need to approach, and I'm going to put this into context, Marty, of artificial intelligence systems. If we need to reduce the risk of artificial intelligence systems, we need to look at two things. Number one  is external attacks. So think of the pipeline that was hacked not too long ago, Marty. So that's a hacker from the outside, a bad actor. We have to be able to protect from that. That's one thing.

Number two – and in the context of AI systems, I would argue even more important – is reducing intrinsic risk. What do I mean by this? It's good-intentioned engineers, configuration and OSI

PIs, sensors that are mis-calibrated… All these things and processes can cause bad data inputs to the algorithms. That means your AI system is going to be untrustworthy. Attackers may not only shut it down, they may want to poison the data as well. A mis-calibrated sensor, in fact, poisons the data. When you're training, when you have your training data set, a well-intentioned engineer may not realize that they don't have representation of states in the data. That's a risk that we have to look into, because it's a digital risk for cyber, for AI systems.

So a way to address all these risks as one, yes, the cyber security firewalls, and so on and so forth. But validating and checking data quality throughout the flow of data, the pipelines of data of these systems, will help us reduce those cyber security risks as well, Marty.

**Marty**:  Gustavo, that's an interesting insight. I, like most of the world, think of cyber security as only being an external problem. That's not it at all, is it?

**Gustavo**: No, not at all, Marty. Again, I always like to tell people that are skeptical, where in the definition of cyber security does it say it's external? Its digital and its risk.

**Marty**:  When you discussed security, you talked about data gathering and pipelines. Where do those two concepts fit in the list of important components?

**Gustavo**: Oh, they're huge. They're the first building block, I would say. So companies, as they go through the digital transformation, this is one thing they did right, which is okay, we're going to go spend millions of dollars. We're going to put sensors, we're going to grab edge computing, and we're going to have data pipelines to put that on the enterprise. So then, they get OSI PI and so on, and boom, they got all this data. It's important, right? Think about the midstream section. If you produce gas, oil, whatever it is, on upstream without the midstream, the pipeline, then the refinery, what's the point? It's not going to do anything, right? So that's the heart of the operation.

The same thing with the data pipelines... If you have an artificial intelligence algorithm, but you don't have a way to get at data, then what's the point? So it's a necessary infrastructure. Obviously, this data pipeline has to be protected from the cyber security standpoint of both intrinsic and external. Okay? So I'll give you an example of a couple of these things.

If you have sensors in edge computing connected to your controllers, and you're collecting all this data for all your, let's say, 40,000 sensors in an offshore drilling rig. It's cost prohibitive to actually store all of that data on your enterprise, right? So you need to have a data pipeline that can validate, check the quality, of what data is good and is worth storing. So then you can decide to throw away or just keep it in a different storage locally, the data that you don't need to send to the enterprise, right? So that is an importance of validation in these data pipelines.

Now, sensors can be mis-calibrated and when they send it to the control, that's part of the data pipeline. We need to validate the data there as well. Telemetry has issues – it drops, it's limited. We need to validate that that's not happening as well. So even though it's an incredible, the number one building block, there's a lot of checks around it that need to happen, because data travels long distances and it can be an issue.

**Marty**:  The final concept that you mentioned was usability. Tell us a little bit more about that, Gustavo.

**Gustavo**: Absolutely, Marty. So the one thing about usability, which is something we all learn the hard way, is that misconnect between people on the field and people with tucked-in polo and collared shirts at the office, right? We need to make sure that all the products that we create are used, or want to be used, by the people on the field. Right? So we need to think about user experience, user interface – because anyone that's been on a field knows that if they don't know what it is, they're not going to use it. If they don't like it and it breaks, they're not going to call anybody about it, because it's not useful to them. So then, it's just an expense for the company.

In oil and gas… and it's changing very rapidly, so this is a different conversation than we would have had five years ago… we had engineers building UI/UXs. We did not do human-centered design. We are now, and we have front-end engineers that do UI and UX, doing these things to make sure that people know how to use them. It's a completely different skill set that we need to continue to attract to our industry. So that's the front-end.

On the back-end, so the people developing – because at the end of the day, this is an engineering industry – some people may just want to have access to your APIs and code. You need to make sure that that code is usable and robust, that it handles errors with proper documentation, that you protect your IP, that people can actually use it, because without it, again, no one's going to use it. And then it's just an expense with no ROI.

**Marty**:  That's a great checklist for our listeners of the five key items needed for a digital infrastructure: data quality, scalability, cyber security, data-gathering pipelines, and usability.

There's one more thing I'd like to talk over with you, Gustavo. In our planning session, you mentioned your work on a geothermal project. In this era of energy transition, geothermal has been written up in the October 2020 issue of JPT. Can you elaborate on your project and where your digital infrastructure concepts helped the client?

**Gustavo**: Of course, Marty. And I think this is a great place to talk about that. So we worked a lot with the federal government. Federal government includes Department of Defense, all different branches. This is an example of a logistics base of the United States Marine Corps in Albany, Georgia. This base, even though small because it's just a logistics base, is one of the hotbeds of innovation. The Marine Corps want to have net zero emissions as a branch of the military in the next 15… 10 to 15 years or so. Don't quote me on the number, but it's around there. And a lot of this experimentation, these pilots, this technology is happening at logistics base in Albany, Georgia.

Okay. With that in mind, there's two ways that this base is piloting these technologies: doing power generation from waste and using geothermal systems for utilities. Okay? So it's all renewable. And keep in mind, Marty, bases are, a lot of the times, disconnected from the city's grids and power systems and they're run like cities themselves, just small cities. In fact, some bases are going to be bigger than some small cities.

Back to the geothermal plan. So what they've done in the base, they've drilled… and it's actually kind of similar to the oil and gas process. They've drilled a bore field, which is a series of water wells

where they store this water. This water is to be distributed to facilities for utilities. I'm going to give you one example of a, let's use the air conditioning and heating example. These wells are produced by a pump and that water is produced and is sent to a multi-phasic heat pump. Okay? This multi-phasic feed pump, again, I'm sure everyone listening knows about this, uses compression to heat water in different stages. So depending on the temperature, it needs to be, it can go to one of two places. It can go to a system that helps cool the facilities, and if it needs to be hot, then it goes to a system that helps heat the facilities.

Remember, we got pumps, valves, and everything, right? When the water's being pumped back to the bore field, it has to be a certain temperature. If it's too hot, it's going to alter the thermodynamics of the bore field. So the wells are very, it's bad if that happens for the wells. So there's dry coolers. The dry coolers cool the water, and then it gets injected back. So this is a cyclical process that, again, is green. There's treatment of the water at the wells. And in the bore fields alone, there's hundreds of sensors. And the whole system alone, pumps, pipes, valves, and so on, there's probably about 50 sensors per one of these facilities. They got two at that plant, at that base right now. So it's pretty special.

So with all these sensors… with all this data that's being collected at, some data's on change, some data's sampled every minute, so on, so forth, very similar to what we deal with in oil and gas. You want to be able to do predictive maintenance, which that's the goal – but one of the things they realized after they had an issue with the bore field, which is the heart of this, right, that's the hardest thing to fix – is they didn't get the alarms they were supposed to get from their condition-based monitoring predictive maintenance algorithm. They didn't happen. It's because they had issues with the sensors that were installed in the bore field. So no one was checking that data quality.

So as they keep building these pilots and processes, they make sure that they have the appropriate digital infrastructure to reach the goal – which is being net zero in an efficient way, being able to predict the maintenance of these systems and optimizing the processes. And you've mentioned their planning session a couple of times, Marty, and something you said really, really stood up to me, which is people think technology before process. When in reality, you got to do process before technology – so infrastructure before predictive maintenance. If not, it won't work.

**Marty**:  Gustavo, thanks so much for your insights, especially your out-of-the-box Marine Corps example. They will definitely be valuable to the SPEGCS audience and our own community of 10,000 EKT Interactive listeners. Do you have any recent articles, papers, or webinars that you recommend for our communities to get more background on this important topic?

**Gustavo**: Yes, Marty, and thanks for asking. In fact, if people go to [pandatatech.com](pandatatech.com), and I'll say that slowly, [pandatatech.com](pandatatech.com), you'll be able to find the links of papers we've published at the officers' technology conference. Our LinkedIn page has a lot more information about data quality, some use cases that are important to us in our journey of evangelizing digital infrastructure, to both the federal government and the energy industry.

**Marty**:  That was perfect. Thanks again, Gustavo. If you want to learn more about the SPE Gulf Coast Section, go to [www.spegcs.org](www.spegcs.org). You can access recorded webinars in the on-demand library, or support our scholarship program by contributing to the scholarship endowment fund. If you're not an SPE member and would like to join, please visit [www.spe.org/join](www.spe.org/join) to enjoy all the SPE membership benefits.

Thanks for listening. Our company name, EKT Interactive, stands for energy knowledge transfer, digitally capturing the innovative perspectives of industry experts like Gustavo Sanchez. If you're new to the oil and gas industry and want to quickly learn how this industry works, check us out at [www.EKTinteractive.com](www.EKTinteractive.com).

Thanks again.