PODCAST TRANSCRIPT

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH
LELAND SPARKS
HOST: ALAN POOLE
GUEST: LELAND SPARKS

Alan Poole:

Welcome to another episode of *TAG Infrastructure Talks* podcast. I'm your host, Alan Poole, a corporate and infrastructure partner at Troutman Pepper. I have with me today Leland Sparks, principal at LNS Solutions and advisor at Emerald Operating Partners. We are finally getting back to data centers in this podcast series, and today we're going to be focusing on cooling. So, Leland, welcome and thanks for joining us.

Leland Sparks:

Thank you, Alan. It's my pleasure to be here.

Alan Poole:

We're really excited to get into the content here, but I think it's best if you could give our listeners a bit of your background and then we can jump into the background of data center cooling, I think, as our starting point.

Leland Sparks:

Sure. I spent six years in the Navy. And coming out of the Navy, this whole thing called networking had just started. So military personnel were good people to bring in because they knew how to deal with problems while they were busy setting up education for everybody. In 1999, I got introduced to my first data center build, and I have been doing data center design, data center build, data center consulting ever since. It is an interesting area to look at. When you think about 20 years ago what a data center was and what we think of data centers today, everything is exponential in the way that it grows.

One thing I learned early in my career, Gordon Moore, a former CEO of Intel, came up with Moore's Law, and that is in this technology space, everything doubles every five years. And I have seen that go through this cycle for about five cycles now, and the next cycles coming are going to be even more interesting when we think about the amount of power and thus the amount of cooling a data center requires. And we are not always conscious of the fact that our cell phone is nothing more than a handheld computer. And we watch videos on it. We email off of it. We text off of it. We do web searches off of it. So we have to have any data available to any device at any time in the data center world, which offers some unique challenges.

Alan Poole:

The issue at hand is cooling. So maybe could we start by talking about the history of cooling? And I guess that probably starts with some fans, right?

Leland Sparks:

Well, actually, if you go back about 30 years ago, most data centers were either mainframe or minis, and we used water to do cooling. When we moved to the PC-style server, it was felt that we didn't need to be that elaborate because the heat wasn't that bad. And in truth, that's true. The first data center I built would have no more than eight servers in it with a single or dual processor, and they were in large boxes. They didn't have a lot of connectivity. They weren't producing a lot of heat. They weren't drawing a lot of power.

But again, applying Moore's Law, every five years that doubles, Datacenter Dynamics came out with an article late in 2021 where they identified that the current cabinet in the data center average is running at 10 kW worth of power, producing 10 kW worth of heat. The next jump is in 2025 or a year and a half from now, where the average power draw at a cabinet will be 20 kW. And in 2030, it'll be 40 kW.

When we started out with this style of a data center using server-type processing, we've added more processors, larger cores to every server. Servers have gotten smaller. We have blade servers. You can produce a lot of heat by running all of those processors, and the whole reason for a data center is to run those processors. It's not about the size of the data center. It's not about the amount of power it uses. All of those are requirements to run those processors.

So when I first got into the data center arena, we just pumped air into the room. That was more than enough, as long as it was moving. And as that went on, we started having to create rows. So we were pumping air into the rows so that we were getting better air to all of the cabinets, to all of the servers. Then we had to start doing containment, and we had to start making pods so that we could either contain the cold air, and thus take that cold air and that pressure and get it closer to the server so that it's blowing into the server, cooling the CPU.

And now we're at a point where even with containment, hot aisle containment or cold aisle containment, whichever one you prefer, the temperatures are still going up. Once we get to 20 kW, a lot of data centers are going to have problems cooling their infrastructure, because when we first design and build a data center, everything's optimized and running at 100% efficiency. But if we have a 20-year life cycle on a data center, most of that equipment actually falls off of manufacturer support after 15 years. The efficiencies aren't there. Even with all of the maintenance that you do to keep it optimized, things wear out. Just think about, would you still be driving a car that you own 15 to 20 years after you bought it?

Alan Poole:

Not at this point in my life, but...

Leland Sparks:

But that's the same thing that's happening here, except that a data center runs 24 hours a day, seven days a week, 365 days a year. We don't shut them down. So this equipment has to be able to be taken offline for maintenance and then put back into cycle. There is no downtime. We don't drive our cars that much. We don't do anything that much. So there's a lot of strain on the system. So over time, just the longevity aspect of it, things do wear out, and thus our cooling efficiency goes down the older the data center gets.

And as we start having multiple CPUs and one server with multiple cores, you have to be able to remove that heat immediately. Heat sinks are there to capture the heat so that we can blow across the heat sink to then remove that heat. But it's not necessarily now about the temperature of the air, it's about the pressure.

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH LELAND SPARKS

Alan Poole:

Really?

Leland Sparks:

Am I putting enough pressure into the front of the server so that I'm actually forcing air across that heat sink to remove that heat? Which means I have to get more pressure, which means I have to get closer. So the distance that air has to be thrown in a data center, I start losing pressure. If I'm the 10th or 12th or 15th cabinet out, I don't get the same pressure as the first few cabinets. And as that heat builds up, it will actually start to cause problems for the CPU. So you have to turn over that heat, you have to remove it, and it's nonstop. So we are looking at the fact that we have to change the way we're going to do cooling in a data center.

In 1999, I did my first real data center job, and that was 100,000 square feet and it was a one-megawatt data center, and that was so impressive at that time. Now, there are so many data centers that have a million square feet of white space that are running 100 to 200 megawatts of power. That is unheard of. So what do you think happens with the heat there? You have to be able to move a lot of heat in that space. It's not just removing the heat from one server or one cabinet. It's all of those cabinets that are having the same issue. And as that heat builds up, it will start to impact the cabinets next to it.

Alan Poole:

Wow. So you have to get it out of the room.

Leland Sparks:

You have to get it out of the room. I was just going to say, with AI suddenly becoming the hot buzzword, we are talking about the processing power going up exponentially. We're talking about the heat problem becoming more severe over the next two or three years.

Alan Poole:

Yeah. The question that was on my mind that you were sort of getting at is AI. The processing power necessary for it is slated to cause a huge uptick in development. Are we going to be able to even hold true to Moore's Law with a doubling every five years because it's going to get faster?

Leland Sparks:

I think what it's going to do is the average is what we're looking at. So if I'm looking at high-end data centers that are doing cloud and they're doing hyperscale, they're the ones that are running a lot hotter than the average. But your typical enterprise data center is not running at 10 kW in a cabinet. They're probably running at four or five. But we are seeing more and more instances where 50 kW, 75 kW, even 100 kW in one cabinet.

Alan Poole:

Wow.

Leland Sparks:

You want to put that into perspective. One of those cabinets at 50 kW uses up the same power as about 1,700 homes.

Alan Poole:

That's quite a subdivision. That's much bigger than a subdivision.

Leland Sparks:

Now, what happens when you got 1,000 cabinets in a data center?

Alan Poole:

Sure. So what other types of cooling methods are being developed right now besides just moving cold air, HVAC?

Leland Sparks:

There are three approaches that are being looked at right now. One is actually building the data center and submerging it in the ocean.

Alan Poole:

Interesting.

Leland Sparks:

So if I can get power to it and I can get fiber to it, I can flood that data center and then lower it into the water, and the outside ambient temperature will remove the heat.

Alan Poole:

I can see some popularity problems with that issue.

Leland Sparks:

Well, it's something that was started a few years ago, and it is picking up speed, but it is not a major solution for most people. The problem is, if you have any maintenance requirements, you're not going to be able to go down there and fix them.

Alan Poole:

Right.

Leland Sparks:

You're going to have to bring the data center out of the ocean, bring it to the surface to work on it.

Alan Poole:

Gracious.

TAG Infrastructure Talks Podcast: The Data Center Cooling Conundrum With Leland Sparks

Leland Sparks:

Another approach that's being looked at right now, and there are samples working, is a system where we're bringing water out of a river or a lake. Bringing it in, straining that water, using it for cooling, and returning it to the source with less than a two-degree uptick in temperature of the water.

Alan Poole:

How do they get back up? Well, excuse me, back down? Or is it just, naturally, the mass of it only causes so much increase?

Leland Sparks:

It has to do with the volume.

Alan Poole:

Right. Volume. Right.

Leland Sparks:

And I'm not going to be able to build large data centers that way, because it's too much volume. So while it's a solution, not all data centers are close to a source of water, that they can take that much water, run it through, and return it back to that water source. And I think about a lot of the Southern states that have data centers. It's not an option for them.

Alan Poole:

No, not in our jurisdiction. Not at all.

Leland Sparks:

Not much.

Alan Poole:

Unless you get lucky like DataBank and find a spring under downtown Atlanta, but you can't plan around that.

Leland Sparks:

Yeah. And the problem with that is the more water you use, the sooner that spring can't supply enough water. I think about data centers in Phoenix using that cistern that they have and the impact that it has had on the water levels for Phoenix over the last 10 years.

Alan Poole:

So the data center itself has had a measurable impact on the water level in Phoenix?

Leland Sparks:

Oh, yeah.

Alan Poole:

Oh, dear.

Leland Sparks:

There were a lot of major data centers in Phoenix, and then the growth kind of stopped for a while and now it's starting up again, because a lot of them are turning away from water. So if I'm not using a water source, but air has a problem. Water can cool so much more effectively than air. It takes a lot of power to force air. And 30% of the power used in a data center, in most data centers, goes directly to cooling. So if you want to reduce your power and you want to work on your carbon neutrality for the future, you have to remove some of that power from cooling.

So the third option is liquid cooling, and this liquid cooling is right to the server. And there are three ways that we do that. The first is rear-door heat exchangers. There are multiple companies that have those. They have been around for a long time. You just attach those onto the back of the cabinet, provided the back of your cabinet is clear. You then bring in a liquid source that goes through a series of cold plates and it's removing the heat.

So 45-degree water ends going out at around 55 degrees out. Through that loop, I can tie multiple cabinets together. I have a liquid, and it could be water. It could be glycol. It could be a lot of different things. I'm using that to remove the heat and cycle that through. So I still have to take that output water or liquid and I'm going to have to cool that back down to 45.

Alan Poole:

And that's when the cold plate comes in.

Leland Sparks:

Well, that's where the rear-door heat exchanger comes in with its plate in the rear door. Now, there are some disadvantages to that. One is, I'm adding depth to the back of the cabinet. And if I have a hot aisle, I'm adding depth to the back of two cabinets. So how much space did I give myself? If I had three feet and these were six-inch doors, I now only have two feet. That is a safety issue. It's also an Americans with Disabilities Act (ADA) issue.

If I open the back door that has all of the cooling, as soon as I move that out, all that heat's coming out of the back. There's nothing there to absorb that heat. So I'm actually removing the cooling from my cabinet by opening the back door. Not that most data centers spend a lot of time in there opening those doors, but it doesn't take much time before all those CPUs start to overheat.

Alan Poole:

When would you have to open that door? For like a major maintenance issue?

Leland Sparks:

If I'm trying to work on a power strip that has a problem or I've got to add new cabling where I'm plugging it into the back. So there's a lot of activity going on in the back end of a cabinet. The next item there would be liquid immersion, and I like to look at liquid immersion as this. I take my cabinet, I push it over to its side, and then I fill it up with mineral oil.

Alan Poole:

Very cool. Yeah.

Leland Sparks:

There are different cooling solutions that go into it, but basically, I'm changing my data center by having a lot of pumps. I'm having a lot of liquid come in. I'm submerging the entire server into this liquid. Problems there is not all servers are intended to be submerged.

Alan Poole:

Sure.

Leland Sparks:

So you have to buy the right servers. Also, network switches are not submergible and storage is not submersible. So those continue to have to be outside. They draw less power than a server, but it does change how you do things. It also means that all of my cables going in, I cannot use the same cables because they're not meant to be submerged. So the liquid actually comes up through the center of the cable and goes out the other side.

If I have the kind of problems with the heat load for something that's that heavy of a processor, having to move from fiber optics back to copper cables means that I'm going to go from 100 gig or 200 gig back down to a maximum of 10 gig.

Alan Poole:

Oh my goodness. So fiber's not viable at all for that last connection?

Leland Sparks:

Well, we're going to have to come up with a different fiber connector that's protected, because there's an air gap between the fiber cable and what I'm plugging it into. This liquid, and think about the density of something like baby oil, if it gets into that gap, I can't pass light. I'm refracting that light. So now I've got big, heavy, thick, non-bendable cables and I have no cable management on this. And if you've ever seen the actual pictures of larger liquid immersion installs, they're a nightmare.

Alan Poole:

Oh.

Leland Sparks:

Also, if you look at the length of a server, they're between 39 and 42 inches. And if I'm going to completely submerge them, everything on the back end of the server, power and cable connectivity, now has to be moved to the front of the server because I can't get to the back. All of that has to be submerged. So my average liquid immersion tank is about 45 inches high. And if I have to reach across this at around 30 inches, how tall do your people have to be so that they're not getting their clothing into that liquid? And now I've got to have some kind of an overhead crane system, either manual or electronic, so that I can actually pull those servers out if I have to do any work on them. The liquid makes them slippery.

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH LELAND SPARKS

Alan Poole:

Oh, my.

Leland Sparks:

The weight of the server cannot just be...you cannot just reach across and pull that up. You're going to have to have a method of doing that. And even if all I want to do is add more memory to a server, I'm going to have to bring it up and I'm going to have to let this drip for about an hour before I can pull it out and work on it. And because there's no cable management, I have to disconnect just about everything in that liquid tank before I can pull one server out.

Operationally, unless I'm working on something where I'm never going to touch this and I'm not going to make changes, it does change the footprint. Cabinets being typically 24 to 30 inches wide, being six to eight feet tall, and being 40 to 48 inches in depth, when I lay those on their side, I can't get nearly the same number of cabinets into a data center than I do when they're vertical. So that changes how we lay things out and what we can do as well.

Alan Poole:

Do you have a sense of what percentage of the room is taken up by this strategy if you were to do every server like this?

Leland Sparks:

You would probably only be able to use about 20% of your space.

Alan Poole:

Wow.

Leland Sparks:

I still have to maintain a three-foot aisleway around to be able to work on anything. And I have lifts that help me lift servers up to push them into a vertical cabinet. When I lay this down, I now have to have something from overhead that I can lower this in, and then try and attach these and then hook everything up. So there's a lot more work to it. There's a lot more operational trade-offs.

Alan Poole:

Is there anything so attractive about this method that makes people actually want to use it right now?

Leland Sparks:

If I am running cryptocurrency farm, I have major heat issues, and liquid immersion is a ready, viable solution from a multitude of companies that I can use today.

Alan Poole:

Crypto. Is that one of the use cases where you don't need to go in and make too many alterations?

Leland Sparks:

No. You run that equipment till you burn it out.

Alan Poole:

You mentioned the third-

Leland Sparks:

Crypto data centers are a lot different than a typical data center. So it does have a very useful life for crypto. But for most data centers, it's...when I think about the size of the pumps and the piping that has to go into this to move this around, to do the cooling, typically, in a data center, security is an issue, and I don't want mechanical people coming into my data center that's digital, and the servers. Now, I've got to let piping people in. I've got to let maintenance in for mechanical systems, which are typically outside the white space and in the gray space.

Alan Poole:

Yeah. I can think of several customers of clients that would not be happy with that new procedure.

Leland Sparks:

Well, liquid immersion's been around for a long time. I mean, if I start looking at some of the companies out there, GRC (Green Revolution Cooling) has been in business for 22 years, and all they do is liquid immersion. So they have customers. They have agreements with major manufacturers, but they're still not growing at a rate that the market needs them to grow at to be successful.

Alan Poole:

Before we get to the technology you've been looking at, are there any other major types of liquid immersion worth summarizing?

Leland Sparks:

Well, I'm not looking at specific vendors. I'm looking at it as collectively liquid immersion itself. When I look at the trade-offs, I don't think... When you look at marketing, I have two favorite words for the industry, marchitecture and slideware. In marchitecture, I can say anything I want because it's marketing. And in slideware, I can prove it works in PowerPoint.

Alan Poole:

Sure.

Leland Sparks:

So you have to break that down. Anybody who's ever been to a data center show, you'll see all kinds of companies there with liquid immersion, and they might have one or two servers in that tank with one or two cables going to those units, so it looks clean. What you have do is you actually have to visit a site that has really deployed the liquid immersion and see what kind of problems that they're going through.

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH LELAND SPARKS

Alan Poole:

And it doesn't tell you anything about the building design or other infrastructure issues either.

Leland Sparks:

No. Well, I toured a facility last year and it's the first time I've ever seen paper towel dispensers hanging on equipment.

Alan Poole:

Oh, my. I would never have guessed that I ever would see that.

Leland Sparks:

Well, when I said it's like baby oil, because that's exactly what it feels like when your hand goes into this liquid. And then, how easy is it to get baby oil off your hands? The liquid is important because of its properties to remove heat, but it's not always people-friendly.

Alan Poole:

Not people-friendly and not digital infrastructure-friendly for sure, unless you're very careful.

Leland Sparks:

Yeah. Right. But like all technologies that have come along, we have to change how we do things. And as much as IT people in data centers like to think about change and new technology and learning new things, they're actually very averse to change.

Alan Poole:

A lot of unknowns. A lot of chaos. I get it.

Leland Sparks:

Well, yeah, but it's taken a long time to separate out the gray space and the white space and who has access and how we design things. And now, with this, and it's not that it's an unnecessary move, but it has major data center architectural and policy changes that you have to look at. You're going to be letting a lot more people into your white space that have no real background in the way servers work and what's going on. Doing a lot of mechanical maintenance and dealing with fluids in a data center that is production-ready can be a high risk level. I have not found one liquid immersion tank yet that has a way to drain the tank.

Alan Poole:

Wow, really?

Leland Sparks:

So what they tell you is you go out and you get a sump pump and you put it in there and you pump it out into a production data center into containers.

Alan Poole:

Uh-huh. Okay.

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH LELAND SPARKS

Leland Sparks:

The next thing is, is you can't use this liquid immersion anywhere. If I think about the size of the tank and I think about the piping and I think about the pumps that are in that, the tank itself can weigh 1,800 pounds.

Alan Poole:

Just one?

Leland Sparks:

Just one. Then I put 800 gallons of fluid in.

Alan Poole:

Oh, pre-fluid. Wow. I didn't realize.

Leland Sparks:

800 gallons of fluid before I put one server in. So I'm over 4,000 pounds before I put a server in. Are your floors rated for that kind of weight?

Alan Poole:

Not a great retrofit opportunity.

Leland Sparks:

No. No. It does fit new construction very well because I can plan for all of it. If I can handle all the operational differences and the configuration and building this out, liquid immersion has a very popular approach, this is an immediate approach, that will fit new construction better than retrofit.

Alan Poole:

That makes sense. I mean, you're talking about design that's completely nonstandard and the weight issues, but you can always figure that out for new builds. Let's start talking about cold plate technology. I understand you've been looking at this and are maybe trying to get to the white paper phase. Could you give us some background on that type of technology before we go into the implementation?

Leland Sparks:

That is the third liquid cooling that I can bring in. And the cold plate is, I have a metal plate that is the shape of the CPU. So all of these plates have to be unique to the CPU that you're running. I remove the heat sink. I put the cold plate on. I am pumping liquid in, and there are different types of liquid here. All the manufacturers have their own way of doing it. There's no standards around this yet.

But I'm pumping the liquid in, then it's flashing to vapor with the heat from the CPU, and then that vapor is coming out and it's going into a manifold and it's going to go into some form of heat exchanger, a CDU, HRU, depending on the manufacturer, what they call that. But I'm going to go to a heat exchanger, which is then going to remove that heat, bringing it back to a liquid, and that liquid cycles through. So it's a closed-loop system.

It does not need any outside cooling. It's all done within the cabinet. I can do it selectively from a server standpoint. So I'm doing one cabinet stand-alone and a group of 50 cabinets. I can add to it. I can, over designing, if I have a pod that has a lot of heat, I can set up an external source to bring cooling in, to help with that cooling. I can share the CDU amongst multiple cabinets by having it as a stand-alone CDU. So if I look at the Open Compute Project, they have put out a paper in November of 2021, putting some guidelines around these three different types of liquid cooling in a data center, and even how to put the mechanical systems in.

So most of the mechanical systems can still be outside the white space, and I'm just feeding it in. So I can do this on an individual basis. Now, I do have to have people come in and open up a server and take the heat sink out and put the cold plate in. So you're going to have to have some highly trained people to do this because of, they're not IT people, if you will.

The other aspect though is, companies like Dell are certifying these companies so that they can put the cold plates on and ship them out all ready to go. And when I install them, all I need is people to come in and put the manifolds in and hook everything up to a CDU and then hook these up directly into that manifolds. So I can either do a one-to-one manifold into a CPU out, or I can daisy-chain them inside the server depending on how much liquid that I'm putting in and what the return rate is for that.

So there's some engineering that has to go up front, but I can minimize the amount of people who are non-IT people in my white space by buying servers with selected manufacturer's product already installed. That has a lot of benefit because it means that Dell is extending the warranty on the server, because they did the exchange instead of a third party doing the exchange where Dell may not choose to provide that warranty anymore to a server that a third party has gone in and removed the heat sinks and put in cold plates.

Alan Poole:

Give me a quick description of the Open Compute Project (OCP). I'd like to talk a little bit about how these open groups participate in the pathway from conception to launch.

Leland Sparks:

Yeah. There are multiple work groups within Open Compute, some of them just focusing on servers, some on storage, some on the networking. Then you have those who are on the mechanical side. So I've been mostly working with the people on the mechanical side. A friend of mine is a co-chair of OCP for liquid cooling. So I talk to Don quite a bit. He looks at this from not the IT side as much as he looks at it from the mechanical side, "What kind of piping do I have to put in? How do I tie these together? How much heat can you actually handle?" And then that has to be translated to the manufacturers themselves.

To the best of my knowledge right now, there is no major manufacturer making a cold plate solution. There are multiple manufacturers who have partnered with major manufacturers. But if I suddenly had to go into a large rollout, I don't think most of the liquid or the chip manufacturers right now could handle the volume that's required. So we will see more partnership going here.

We will see more large names like Schneider and Vertiv getting heavily into this, partnering with somebody, buying somebody, or starting to build their own, because it is something that within the next couple of years, between Moore's Law, I'm going up to 20 kW per cabinet as an average, and AI coming out, it is a need that needs to be addressed right away. How do we do this? And we're not going to do it on a single cabinet basis for very long, because my problem's going to extend across multiple cabinets in a hurry.

Alan Poole:

How far are we from these... Well, let me ask it this way. What's the status of deployment of cold plate technology? Is it out there? Is it still in testing?

Leland Sparks:

No, it's being deployed. We actually thought that this time last year, it was going to take off. Unfortunately, we saw our economy take a hit. Now what we're seeing with all this talk around AI, it is picking up steam. There's a lot of things that are starting to happen. Vertiv has announced that they will be developing and supplying cold plate by next spring. So they're going to do it in-house. They already have most of the components. It's just the cold plate. And if Vertiv is doing that, it means that their customers are pushing them for it.

Alan Poole:

True.

Leland Sparks:

So when you look at somebody the size of Vertiv, or when Schneider finally makes their move, it'll gain a lot more popularity because of who those companies are and the number of data centers they're in now.

Alan Poole:

How long before we really need to be deploying cold plate to avoid slowdown and the massive ramp-up that everyone's predicting for data center development?

Leland Sparks:

We need to be doing it now. If I look at most data centers, they're running either a three-, four-, or five-year refresh rate, which means if it's a three-year, there are companies who are already refreshing now. They're not waiting till 2025 to get to 20 kWs an average cabinet. All of your high-end, your hyperscalers, your HPCs, your high-performance computing, all of those who are starting to develop and deploy AI, they need it now.

If you look at somebody like Amazon with all their data centers and with everything they're doing with cloud and who some of their customers are, they need it now. So they're working with companies right now trying to develop their solutions because they do a lot of development in-house, but they're going to have to take advantage of work that's already been done. They can't take the time to catch up.

Alan Poole:

Are you seeing the big players adopt this technology like you would hope?

Leland Sparks:

I'm seeing a lot of interest. I'm seeing a lot of conversations. Some of the speakers at Data Center World back in May were focusing on this. 7x24, they're starting to focus on it. OCP coming up in October, they're already looking for people to be speakers on this that have installed. Yeah. It's just a matter of how quickly it starts and how fast it ramps, but it's definitely taking on a more serious approach since AI has really started coming alive, if you will.

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH LELAND SPARKS

Artificial intelligence isn't new. In 2019, at the Gartner show in Las Vegas in December, they were talking about AI. Intel has been promoting AI products since 2020. Here's the gotcha. You got to determine what you need artificial intelligence to do for your company. It's not out of the box. You need the technology to handle it, but you have to first define what it is you're trying to do and what you need AI to do for you. Other people cannot define it for your business.

If I have AI-compatible equipment and a hardware standpoint, that's great, but it still has to be on the software development process, procedures, and the things where I need to develop that. Once I have that developed, then I can go out and get this hardware and I can see significant jumps in processing in companies almost overnight.

Alan Poole:

Oh, it seems like we're waiting for a big shoe to drop when there's enough momentum behind the actual implementation of AI in a commercially viable manner. Maybe we're not there yet, but we're going to be there and we may not even know when that is.

Leland Sparks:

I think it's starting to become a groundswell. I think it could have happened last year, except when we looked at inflation rates at 8.3 and higher. You're not going to necessarily run out and spend that money until you know what you're doing with it. And most companies are still going to do pilots, just because it's new technology. I have watched a lot of new technologies come into the data center space that have never gone anywhere. So if you want to do this, there's a term. An old CIO used to talk to me all the time, and she would say she wants to be leading edge but not bleeding edge.

Alan Poole:

That's very good.

Leland Sparks:

So it's a matter of, "I want to be one of the first, but I don't want to be first. I want somebody else to figure out the problems and I'll work on it once I know what they are and I'll jump into it, but I don't want to have all the problems and have that impact on my production data center." So I think that's where we're at with some of it. We're starting to see some of that bleeding edge taking place. So the leading edge is coming right behind it.

Alan Poole:

Well, that's what pilot programs are for, essentially. Right?

Leland Sparks:

That's exactly why you have pilot programs, because no matter what a manufacturer does, unless you're in a true production environment and you understand how they operate and what their priorities are, you do not have a data center-ready product day one. So you do need to get out and you need to find partners and you need to pilot, and it needs to be in a controlled environment that you can eventually roll into a production environment where you can control what's happening and you can make the adjustments that are necessary. But I think we're there. I think it's just going to be this groundswell, if you will. People are doing it now. We're just not hearing about it because they're still in the pilot stage.

TAG Infrastructure Talks Podcast: The Data Center Cooling Conundrum With Leland Sparks

Alan Poole:

Do you have any predictions about how prevalent different types of liquid immersion techniques will be in the next, let's say, five, ten years?

Leland Sparks:

I think because they've been out there longer and going to shows, we're going to see liquid immersion jump out of the gate and be the top provider. I think just due to the limitations of the rear-door heat exchanger, it's going to become a retrofit solution for a lot of people until they make the changes. I think cold plate, over the next two or three years, will take over the number one spot because just from the fact that I can have a lot of the work done at the server manufacturer and shipped in as opposed to having to do it on-site.

Alan Poole:

Yeah. That sounds huge.

Leland Sparks:

It is huge, because it means that the people coming in are only going to work in the back of the cabinet. They're going to put the heat exchanger in, they're going to put the manifold in, and they're going to hook things up. I don't have plumbers opening up my servers. And once people start to see that even though it's liquid and it isn't a contained system, if installed properly, there is no leakage or there's minimal aspect of leakage. I know it's one of those things that the white space folks do not want liquid in their data center, and yet, by law, they have pre-action sprinklers sitting over the top of their servers. So the liquid's already there.

Alan Poole:

Yeah. So where does the possible liquid leakage come in? From the vapor or something else?

Leland Sparks:

Well, you're putting liquid under pressure and you're going through fittings. If people don't get fittings into the manifold properly and tight enough, if those connection points aren't solid, and they need to be the kind of connections that cannot be knocked out accidentally. I do have hoses. So depending on how those hoses are set up to connectors, it can become a leak point.

So there's plenty of opportunity for leakage in a data center. The manufacturers themselves have to do a lot of quality work to make sure that building these and shipping them out to multiple entities, who are then going to install them into product and install them into your data center and into your cabinet and into your server, that you have the confidence that they're going to be the right quality and that the workmanship is good.

Alan Poole:

Sounds like you have a potential risk, which I guess is inherent to all new technology for critical needs that you have some newsworthy failure happen and then it just slows down for a while and it's behind five years where it should be.

Leland Sparks:

We've had that in so many different areas in data centers.

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH LELAND SPARKS

Alan Poole:

Sure.

Leland Sparks:

Yeah. The actual outage in the data center is not what costs you. The cost is to your reputation that you had an outage and you had to report it.

Alan Poole:

Sure.

Leland Sparks:

Yeah. This is what I would recommend anybody who wants to get into this and start seriously deciding which way to go. Plan a pilot first or find out from the manufacturer where they're conducting a pilot, and get them to allow you to come in and observe so that you have a better understanding of what it is and why the right company with the right quality program can do this for you.

And like I said, I think it's going to take some of the major players developing their own product, developing their own supply system and quality program, whether they're buying a company or they're investing into a company and they can lead where that's going, that they already have that reputation going into the data center. So I know that when they come in, they're going to do quality work.

Alan Poole:

We mentioned that there are certain types of use cases where, for example, liquid immersion maybe makes sense. Are there any particular types of data center buildings or maybe companies or use cases that are a good first shot for cold plate?

Leland Sparks:

I would say, if you're going into a colo (colocation data center) and you're moving cabinets in and you're going to have 50 kW cabinets, but you're only going to put 20 or 30 cabinets into a colo, work that out with the colo and do that there. That's a good use case, because one of the things with colo is we don't spend a lot of time on-site, so where I can monitor it remotely and I can put it out there, or any enterprise company that has an application that is driving hotter than other applications, where I can move them into a pod, anywhere from 10 to 16 cabinets in a pod, that I can tie those together and I can separate them out, because what normally happens when I have an application that's running hotter than everything else, I have to throw a lot more air at it and a lot more pressure.

But I don't always have the capability to do that without stealing from other areas. I have certain volume of air coming in with a certain pressure. So if I can build that out into its own pod and set that up from a cold plate standpoint, it'd be a lot easier to manage without impacting the rest of the data center.

Alan Poole:

Well, this has just been a fantastic conversation. I always like to close my podcast episodes by asking our experts to try and make a guess at the next big thing. So we've identified cooling as

the problem to solve. What do you see as the next big issue in this area after the types of liquid immersion or liquid cooling shake out?

Leland Sparks:

Yeah. Actually, it's happening in parallel, and that is micro data centers.

Alan Poole:

Ah. Is that the same thing as edge data centers or something else?

Leland Sparks:

Well, an edge data center is that I'm moving more the application closer to the user from a latency standpoint. I can put that in a colo. A micro data center is like an ISO container that I'm building a data center in and I'm dropping it wherever I need it. And the one area that's driving that is autonomous vehicles. If I look at Interstate 70 and I'm starting in California, I have to go up and over a mountain range, then I'm in the desert for a while, and then I got to go across the mountains in Colorado.

Well, if we think about latency being a factor of about 200 to 250 miles, when I'm talking about an autonomous vehicle with the thousands of commands that are happening every minute to be able to control that car, I don't have the same latency issues in the desert that I have in the mountains. So how often do I need to build a micro data center and put it out there so that I can hand off a car from one data center to the next so that you could drive across the country without ever steering the car?

That is the biggest problem with autonomous vehicles. It has to be able to talk to a data center. So I have to be able to build a lot of small data centers that I put out remotely, that I have to get power to and I have to get telecommunications to. So there are a lot of people building micro data centers, a lot of people that are looking at it from a mechanical standpoint. But I think from an operational standpoint, this is where we're going to see things like liquid immersion and cold plate becoming a prime factor so that I don't have to do as much maintenance because I have to have remote control with it.

And every auto manufacturer out there is doing their own thing when it comes to autonomous vehicle. So I cannot even share this between Ford and Tesla. Each of them are going to have to have their own. So it is a big challenge of, how do we move computing power out of the data center, maintaining data center rules, and putting that into remote areas? Anybody that's ever driven across Utah knows there's nothing exciting about that drive.

Alan Poole:

The power and telecom out there. That's tough.

Leland Sparks:

Yeah. So there's going to be a lot of rules around that. How do I maintain that? How do I set it up? How do I power it? How do I grow it? How do I update technology? So we will see a major shift in some companies in how they're going to address that. I know a lot of companies have jumped into this arena, but it's really more from the mechanical side of things. There's a lot of work that's going to have to be done on the computing side. If you think about cell towers, my cell phone works great until 4:00 in the afternoon when everybody is on the tower.

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH LELAND SPARKS

Alan Poole:

Yup.

Leland Sparks:

It's going to be the same thing. How do I handle rush hour?

Alan Poole:

Oh, boy.

Leland Sparks:

Bandwidth is enough. How much processing power do I need? How do I have enough telecommunications? How do I have multiple suppliers? Because not everybody is on AT&T or Verizon.

Alan Poole:

Yeah. Make it closed control here.

Leland Sparks:

So there are a lot of problems to work out there. And it's going to be interesting because there's going to be so many different use cases out there. Can you imagine somebody who uses their cell phone a lot driving across the top of Utah and they can't stream video? Yeah. As we see that, we were concentrating data centers for the longest time, making them bigger and larger. And even while doing that, we're having to tie multiple data centers together based on the type of application and replication and having data available. So if something goes down, the user's not down. Now, when we think about that from a car standpoint, an autonomous vehicle, that's an even bigger issue because of the life safety aspect.

Alan Poole:

Yeah. It can't go down, or you have to have some good contingency.

Leland Sparks:

You do. There are lots of other areas as well. I mean, there's a lot of push going on in the data centers about reducing power.

Alan Poole:

Right.

Leland Sparks:

There are a lot of data centers-

Alan Poole:

That's the climate perspective.

Leland Sparks:

Yeah. Well, everybody wants to go to clean, green power, except that the sun doesn't shine at midnight.

Alan Poole:

Right.

Leland Sparks:

So now I got to have some kind of battery storage system, and how big does that have to be? The wind doesn't blow all the time. I may not be close enough to hydroelectric power. So how do I do that? How do I generate my own power? How do I come up with non-carbon-based fuels to be able to run a data center? So that's a big one that people are working on as well.

Alan Poole:

Yeah. I think we'll be focusing on that at our next big TAG event on September 20th. The concept of green energy and data centers, it still seems like a twinkle in the eye, because those need some serious power.

Leland Sparks:

Well, it is a twinkle. But I mean, if I think about Amazon in Boardman, Oregon, they are going from 3.2 million square feet of white space to 6.4 million. They're already running over 325 megawatts of power, which means they're going to have to go up to closer to 650 megawatts of power. They're built right next to McNary Dam. So that dam is not shipping power off to Seattle or Portland anymore. Amazon's taking it all.

Alan Poole:

Goodness.

Leland Sparks:

And yet, they still have bought land to put in solar farms and, up the road, their wind farms, because they're trying to figure out how to bring all this green power in and have power for growth, because I'm not going to go from 2023 version of server, and in 2027, have one that uses the same amount of power. It's going to be bigger with more processors, more cores. So my power requirements are going to go up, not down.

Alan Poole:

Well, Leland, this has just been amazing. I really can't wait until we can share this with our listeners. And it ties in well with our... The fall is always big on data centers for us. So on behalf of TAG and my firm, I'd like to thank you for getting together with me and taking the time to talk with us about this.

Leland Sparks:

Well, I'm very happy to do so, Alan, and let me know if you need anything else from me.

TAG INFRASTRUCTURE TALKS PODCAST: THE DATA CENTER COOLING CONUNDRUM WITH LELAND SPARKS

Alan Poole:

Absolutely. And hey, best of luck out there. I hope we read more about cold plate technology soon and some of your efforts. So good luck out there.

Leland Sparks:

Thank you very much, Alan.

Alan Poole:

Absolutely. Have a good one.

Leland Sparks:

You too.

Alan Poole:

Everyone, this has been *TAG Infrastructure Talks*. If you like what you heard, please make sure to subscribe to this podcast on your listening channel of choice and follow TAG and Troutman Pepper on LinkedIn for more.