CloudTweaks Podcast 28 - Taming Kubernetes Complexity - Automation, Cost, and Control.docx

Steve Intro

Cloud infrastructure has become more powerful, more complex, and more expensive than ever. At the center of it all sits Kubernetes, quietly acting as the operating system for modern data, cloud, and AI workloads. But while Kubernetes enables massive scale and flexibility, it also introduces new operational, financial, and strategic challenges that many organizations underestimate. In this episode, I'm joined by Maxim Melamedov, CEO and founder of Zesty, to unpack why Kubernetes remains so critical in today's data-driven world, where automation can help or hurt cost, performance, and stability, and how companies can strike the right balance between control and efficiency as their workloads continue to evolve. Maxim, welcome to the CloudTweaks podcast.

Maxim

Hey, Steve, thank you for having me here.

Steve

Obviously, cloud optimization is a crowded space, so what makes Zesty's approach to automating Resource Management unique or different from others?

Maxim

Thank you. That's a great question, as you mentioned, that is that is a space that it has a lot of players. What differentiates us is our ability and approach. We are looking at the optimization from two lenses, both the operational part, where we're operating on the cluster level, and both on the financial part, when we're leveraging financial tooling such as Spot Instances and saving plans. So having one algorithm that can perfectly size your infrastructure and then make sure that it leverages all the financial tooling to make it as cost efficient as possible, that what makes us unique in the market.

Steve

So you basically are customizing to each client's requirements is that, and you're doing that? I mean, it's interesting to find out that you're doing this, but I'm surprised that it is unique. So you're going sort of above and beyond most other organizations in terms of tailoring the products to their needs. Is that correct?

Maxim

You have to think about your consumption. Your consumption has the CPU RAM layer of consumption that is within a cluster, and there is the financial elements of that consumption that impacts your cost and SLA. So by putting all of this together under one umbrella, we allow customers to set better strategies and make sure that their infrastructure meets their business goals, not only cost savings or not only operational efficiency. It's both.

Steve

When we talk about cloud infrastructure, people often think about Kubernetes. Why is this? Why has Kubernetes become so central to how companies run workloads today?

Maxim

I would say Kubernetes is the operating system of cloud infrastructure. There were a lot of container orchestration solutions out there. The reason Kubernetes won, I would assume that it's because it was open source by Google, and as an open source tool, it requires configuration management and setup, but it also allows you to customize it for your needs, so you have the managed version of Kubernetes in every

public cloud you have, the GKE, AKs, ECS, and it also allows you to run your clusters or orchestrate your containerized clusters on prem. So it became a standard, and once you have something that is standardized, that allows you to simply configure, simply manage and simply scale, that how Kubernetes became so popular,

Steve

So with all the benefits of cloud and containerization, you're speaking about the good things that you can provide. What are some of the hidden dangers that organizations might overlook, whether it's financial or operational or security related?

Maxim

Security you want as less tools as possible, right? That's rule number one, I would say, more tools, more abilities to have a problem. From an operational standpoint, there is a fine balance between stability, performance and cost. That's kind of the goal. The first thing that any person who is responsible for cloud infrastructure wants to risk is stability or performance. He doesn't want to think things to break, because when things break company, lose money. Business is losing credibility. Customers are getting disappointed, and then there's a whole spiral effect, and the balance with performance and stability is cost. Because, in a way, I can take extremely large infrastructure and to have it available for me, 24/7, the probability of us getting an outage is significantly low. So you know, it's easy, but in a in a vital business where cloud computing is not always available, especially on the GPU side, you have to figure out how to share those resources in a more efficient way so you are running efficiently, and in that realm, you are factoring in different containers behavior, different applications behavior, different scaling parameters and different performances of different cloud vendors. So to figure out how you are able to set the right strategy per cluster, per workload, per container, requires understanding, and you have to also adjust your setup as time passes, because you have consumption. That is not stable or steady. That's the whole interesting part, that your consumption is fluctuating, and you want to be sure that you are as efficient as possible, otherwise you are either wasteful or crashing.

Steve

How does automation, when done right, help reduce risks like resource overprovisioning or outages or compliance gaps?

Maxim

When automation is done right, it means that you, as the person responsible for operating infrastructure, would put the right guardrails around that workload. And once you put the right guardrails around that workload, the automation will operate within those guardrails, you can afford to be more aggressive when you're talking about your QA and Dev environments. In some cases, you want to save more money. You want to be more cautious when you are identifying workloads that are critical for your production. With that being said, you want to be dynamic as well, because static is just expensive and risky. So by setting the right guardrails around different workloads, around different applications, and allowing automation to operate according to the strategy you have set, that's the simplest way. That's the most efficient way.

Steve

How would you answer to people who are worried that automation is taking over too much control? I mean, how do you balance, let's say, AI driven decision making, with human oversight in a system as complicated as Kubernetes?

Maxim

I would say quite simple. You put you are the one that needs to set up the policy so you are in control of how much the automation or the machine learning algorithm will act. You're putting the structure around

it. So I don't expect people say, hey, you know, I trust you do whatever you want, because there is a lot of missing context and things can break having a person in the middle when you are setting something up, I think, is crucial, especially today, where you don't really know what will be the next action and there is no full context. So this is a way to simplify it and allow people to develop and build rather than maintain and reinforce.

Steve

Very nicely put, is there anything else that you would like to have our listeners be aware of with regards to Zesty specifically, or about the industry itself?

Maxim

So regarding Zesty in the industry, yes, there is an interesting shift that is happening with the recent announcement of AWS around saving plans for databases as an FYI if you are using reserved instances on your EC two machines, note that they are no longer available if you want to launch a new machine type. So start figuring out how you transition from reserved instances to saving plans, and if you need help automating that and how you build a strategy, set up strategy and monitor and act on your strategy, we can help.

Steve

So let's learn about Zesty, where you're based, what parts of the world you primarily provide your services, or everywhere. Some interesting context about the company itself.

Maxim

We have been around since 2018 we operate out of three centralized locations. One is Israel, the second one is UK, and the third one is us, specifically the Bay Area. And we're really excited about the future of how the cloud infrastructure, and especially Kubernetes, is being the driver behind all the AI workloads today. So if you have to look at all the Neo clouds that are offering GPU related value, they are all driven by Kubernetes. So it's kind of the next shift. And if you have to think about Kubernetes, it's currently the version of 1.35 that's the that's the latest one. So we're, we're still in the diaper phase.

Steve

Thank you so much for joining me today on the cloud tweaks podcast.

Maxim

Thank you for having me, Steve.

Steve

You can find out more about Zesty by going to their website, that is Zesty.co.

And as for us, remember, if you like what you hear, be sure to subscribe and tell a friend or a colleague or a client or all of them about the CloudTweaks podcast. We will be back again soon with another episode and another discussion on the topics that you need to know about to successfully carry on in the business of information security. Until then, I'm Steve Prentice. Thanks for listening.