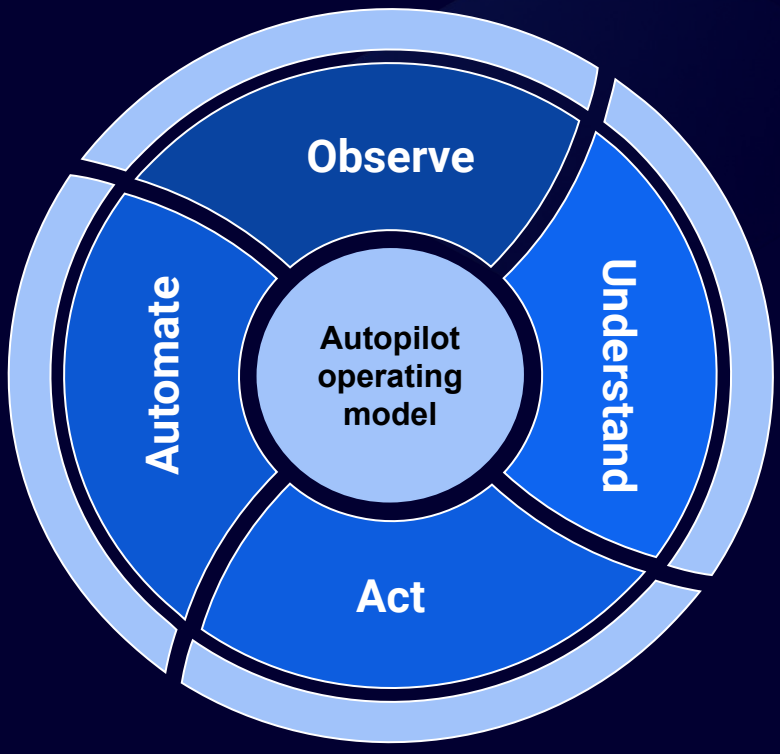
CloudNatix

Autopilot operating model

Observe, understand, act and automate operations across your global IT infrastructure fleet spanning multiple VMs, Containers and Kubernetes Clusters

Author: Somik Behera, Head of Products



© 2021 CloudNatix Inc. All rights reserved.

Contents

© 2021 CloudNatix Inc. All rights reserved.

Executive Summary

In this white paper, we look at the need for the Autopilot operating model (AOM), the implications of the Autopilot operating model, and describe its benefits for minimizing cloud waste and accelerating DevOps, FinOps, SecOps, NetOps and Developer productivity.

Enterprise organizations are now moving to cloud and multi-cloud infrastructure en-masse in service of their digital transformation projects [1,2]. One of the key application architectures that drives the adoption of cloud and increasingly multi-cloud remains building cloud-native Apps using Kubernetes[3] in service of enterprise digital transformation projects. Some leading organizations have also started exploring and building multi-cloud-native applications as well [4]. As cloud-native application architecture matures, the number of Kubernetes clusters within every enterprise is rapidly increasing [5]. With the rise in multiple Kubernetes clusters, comes the challenge of operating efficiently in this complex environment [6], and this is impacting DevOps and Developer productivity.

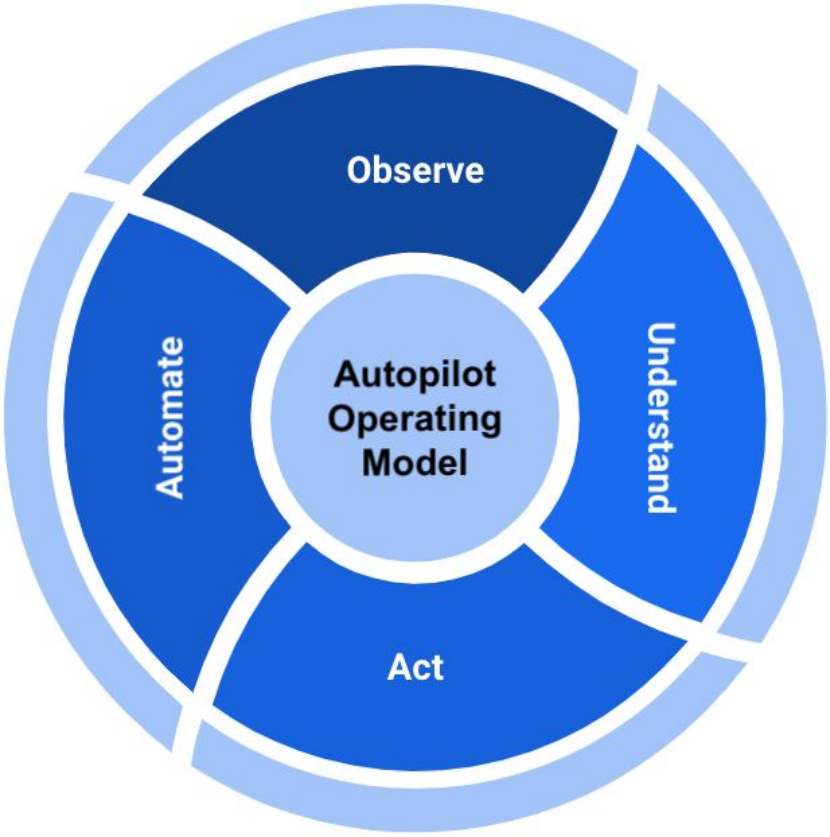
A recent industry survey by FinOps foundation, covering 804 responses and $30.9B in enterprise cloud spend, found that 53% of enterprises have 2 or more clouds and enterprise infrastructure is becoming increasingly multi-cloud [7]. FinOps survey for Kubernetes further shows that 58% respondent are struggling with optimizing rapidly increasing spend on Kubernetes. Experts have cited anywhere from 30%-40% of many organizations' cloud spend is wasted. This is a big problem as cloud efficiency and cost are top of mind for every enterprise that is increasingly becoming a cloud-first enterprise adopting cloud-native application architecture.

The biggest challenge in operating an efficient global cloud environment remains in "getting engineers to take action", second only to "accurate forecasting" and "automating cloud waste reduction" [8]

This white paper will explore the causes of cloud waste and solutions to minimize it. But first we will need to understand the Autopilot operating model and the existing cloud architecture that have led to these inefficiencies. We will then explain how we can help FinOps, DevOps, SecOps, and NetOps leverage the Autopilot operating model that has helped hyperscalers achieve extreme efficiency [9] and increased productivity.

© 2021 CloudNatix Inc. All rights reserved.

Autopilot operating model overview



Autopilot operating model

To thrive in an era of multi-cloud infrastructure where enterprises are increasingly building multi-cloud-native apps leveraging Kubernetes, driven by digital transformation, enterprise IT must evolve from gatekeeping based on Information Technology Infrastructure Library (ITIL) best practices to enabling shared self-service, automated processes for DevOps excellence leveraging the Autopilot operating model.

For most enterprises, the goals of digital transformation focus on delivering new business and customer value more quickly and at a very large scale, akin to how the web-scale pioneers such as Uber and Twitter have conditioned the modern consumer expectations. The cloud is an inevitable part of this transformation as it presents the opportunity to rapidly deploy on-demand services with limitless scale and unparalleled compute capabilities to ultimately deliver next-generation experiences to customers.

In the cloud, however, enterprises have the challenge of maintaining their existing resources, private clouds, and datacenters while simultaneously developing new applications and services that leverage the public cloud's benefits, which includes modern cloud-native application architecture using Kubernetes among other changes. Many enterprises soon discover each cloud provider operates differently and they must choose among a vast array of cloud-based services and also be attuned to leveraging the dynamic, programmable nature of cloud to their advantage and not become a victim to cloud waste.

© 2021 CloudNatix Inc. All rights reserved.

Autopilot operating model overview (contd.)

The dynamic nature of cloud infrastructure and cloud-native Application architecture requires adopting the Autopilot operating model that continuously optimizes the enterprise fleet for efficiency and availability instead of using the static provisioning, updating and operating model of the pre-cloud era.

The Autopilot operating model consists of the following steps:

1) **Observe** your unified cost & operational Intelligence across your multi-cloud infrastructure consisting of virtual machines and cloud-native applications built on containers and Kubernetes.
2) **Understand** what you can optimize with machine generated insights and recommendations
3) **Act** using an Operations Manager that spans multiple clouds as well as Kubernetes clusters to optimize your K8s and VM workloads for availability & cost
4) **Automate** recommendation implementation using operations manager and realize efficiencies and more importantly put your infrastructure and applications on Autopilot, when you have become confident in the Autopilot operating model.

The Autopilot operating model needs a planet-scale cluster manager

A planet-scale cluster manager allows enterprises to provide for unified access management, while controlling multiple clusters, automating resource management and more, all across multiple regions and clusters of capacity, aggregating vs disaggregating cluster resources, using logical services that span clouds versus namespaces that span a single cluster.

In essence, a planet-scale cluster manager manifests to Autopilot operating model by providing enterprises a control plane that lets multiple teams, be it FinOps, DevOps, SecOps or NetOps, to *observe* cost and operations, *understand* opportunities with actionable insights, *act* across environments and finally *automate* using autopilot techniques that were pioneered by the hypescalers [9].

© 2021 CloudNatix Inc. All rights reserved.

Implication of Autopilot operating model

FinOps on Autopilot

Today's Status Quo	Autopilot Model using a Planet-Scale Cluster Manager
• Fragmented or missing cost visibility • Extensive time and effort required to understand allocation and perform forecasting • Inability to discover top offenders and drive engineering action • No automated and unified reporting capabilities across multiple accounts, clouds, and K8s clusters	Planet-scale cluster manager becomes the single source of truth across FinOps, DevOps, and Development teams.

DevOps on Autopilot

Today's Status Quo	Autopilot Model using a Planet-Scale Cluster Manager
• Fragmented or missing operational visibility across clusters • Extensive time and effort required to access VMs or Pods across clusters - multiple IAM roles, RBAC roles, differing credentials • Multiple tools required to find pods, historical metrics, discover logs, and remediate with actions in an outage situation. • Multi-week effort for right-sizing of instances and K8s pods and many manual adjustments, resulting in missed opportunities, cloud waste and availability degradation	• Observe global cost and operations. • Understand with ML generated insights • Act with human-in-the loop automation • Autopilot to ensure continuous cost and operational optimization.

© 2021 CloudNatix Inc. All rights reserved.

Implication of Autopilot operating model (contd.)

SecOps on Autopilot

Today's Status Quo	Autopilot Model using a Planet-Scale Cluster Manager
• Fragmented or missing access logs when the enterprise environment spans multiple K8s clusters or Cloud environments. • Extensive time and effort required to understand the level access given to particular identity across multiple clusters or clouds, requiring multiple teams and tools such as VPNs and Network Firewall rules that are not identity centric. • Brittle access revocation mechanisms for a particular identity across multiple clusters or clouds. • Multi-week effort for right-sizing of access policies and ensuring compliance.	• While development teams "shift-left", planet-scale cluster manager helps you "secure-right" with an unified access plane across clusters.

NetOps on Autopilot

Today's Status Quo	Autopilot Model using a Planet-Scale Cluster Manager
• Fragmented or missing operational visibility across clusters around application centric network ingress and egress points for K8s based services • Extensive time and effort required to access VMs or Pods across clusters - VPNs, Firewall holes, multiple IAM roles, RBAC roles. • Multiple tools required to find pods, historical metrics, discover logs, and remediate connectivity issues in an outage situation.	• Observe global workload connectivity patterns - egress & ingress across clusters, VMs and K8s • 1 click automated, orchestrated, secure, zero trust access across pods in any cloud or cluster

© 2021 CloudNatix Inc. All rights reserved.

Reduce waste and increase productivity

Cloud waste and multi-K8s complexity problem

The functionality of the cloud is simple enough: on-demand, pay only for what you use, and limitless capacity. But as cloud use grows, the cost of cloud can dig into the gross margins of the cloud providers' customers and thus requires more than refinement and optimization. It's an issue Sarah Wang and Martin Casado pursued in their research published in the Trillion-Dollar Paradox.

Public cloud spend was up 41% from the previous year, according to Gartner. But, it is also estimated that upto $62B of cloud spend is wasted [10]. With that much money on the line, organizations need to think about how to create and enforce policies to ensure operational consistency to manage costs in the cloud as well as adopt techniques to optimize cloud capacity and therefore cloud costs.

The problem isn't new. In 2017, Gartner noted that "through 2020, 45% of organizations that perform lift-and-shift to cloud IaaS without optimization will be overprovisioned by as much as 55%, and will overspend by 70% during the first 18 months."Unnecessary cloud costs come in many forms: Paying for unutilized cloud infrastructure Many organizations moving to the cloud suffer from a lack of oversight, visibility, and tracking around what developers can provision, which can lead to idle resources, over provisioning, and orphaned resources. These unnecessary costs can quickly inflate your cloud bill.

This should be no surprise, given that Google, a hyperscale pioneer, also noticed average utilization in their private cloud environment in early days of around 10-50%, with Google websearch servers spending 30% of their capacity as Idle.

More recently, Microsoft research published research around how the "Idle" server capacity in Azure datacenters is significant and Azure could potentially leverage emerging techniques around Idle Capacity Harvesting, to "monetize" this Idle capacity.

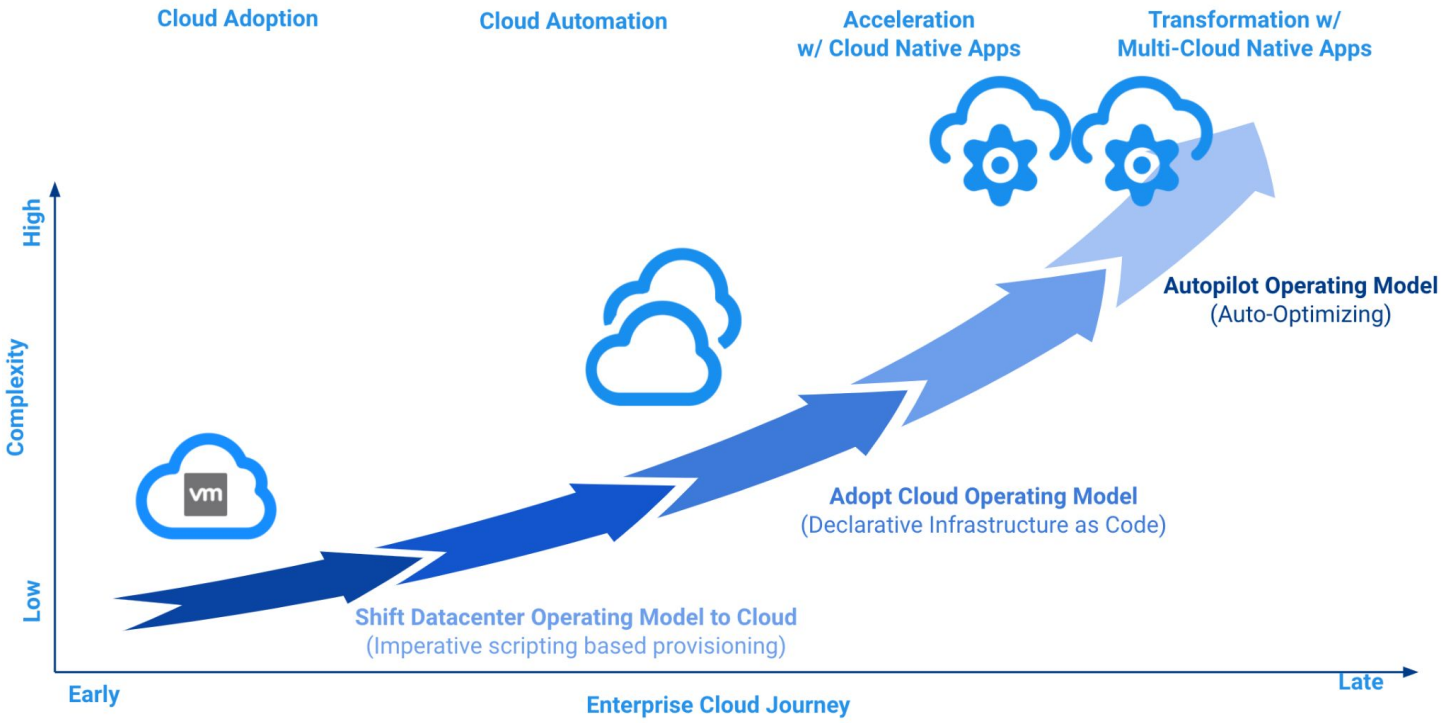
As you can see, the hyperscalers were and are acutely aware of the challenges and opportunities created by cloud waste and cloud underutilization.

For enterprise IT teams, these shifts in approach are compounded by the realities of running on hybrid- and multi-cloud infrastructures and the varying tools required to work with each technology and vendor. To create effective multi-cloud teams that maximize productivity and minimize waste, they need to apply their skills consistently regardless of the target environment.

© 2021 CloudNatix Inc. All rights reserved.

Reduce waste and increase productivity (contd.)

The autopilot solution to cloud waste & devops productivity



While at Google, the invention of Linux containers by Rohit Seth, back in 2006 was a key piece of the puzzle, Google invested heavily in running a cost, capacity and availability optimized infrastructure that spans the planet.

We can see evidence of Google leveraging Artificial Intelligence (AI) and Machine Learning (ML) techniques dubbed "Autopilot" to autonomously run planet-scale applications with high availability and cost/capacity efficiency.

Autopilot continuously profiles and understands application needs as well its observed performance via service level objectives (SLOs) and then tries to "recommend" corrective actions to human operators. Once, approved, the "Autopilot" engine automatically and continuously "right-sizes" containerized applications that span multiple servers, then finally "bin-packing" multiple workloads into a single server to maximize the server's utilization while minimizing the application down time as measured by the application's SLOs.

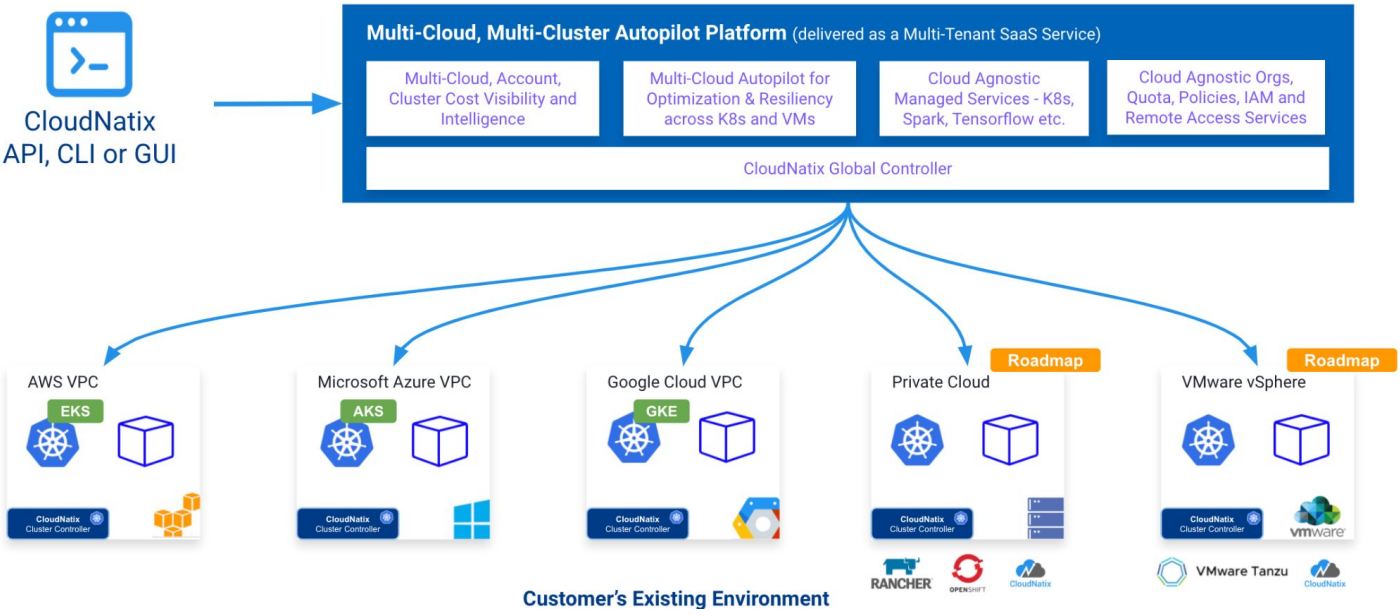
The "Autopilot operating model" addresses the key challenge, as identified by the FinOps Foundation - Encourage engineering action for cost and outage avoidance. The Autopilot model enables engineering action by leveraging ML techniques to generate insights and recommendations and simplifies

© 2021 CloudNatix Inc. All rights reserved.

CloudNatix puts cloud operations on autopilot

Unique planet-scale cluster management approach

Planet-scale cluster management approach simplifies management and dramatically boosts cost optimization across fleets of diverse infrastructure operations—VMs or containers, legacy or greenfield, on-prem or cloud-based—all delivered as a service.



© 2021 CloudNatix Inc. All rights reserved.

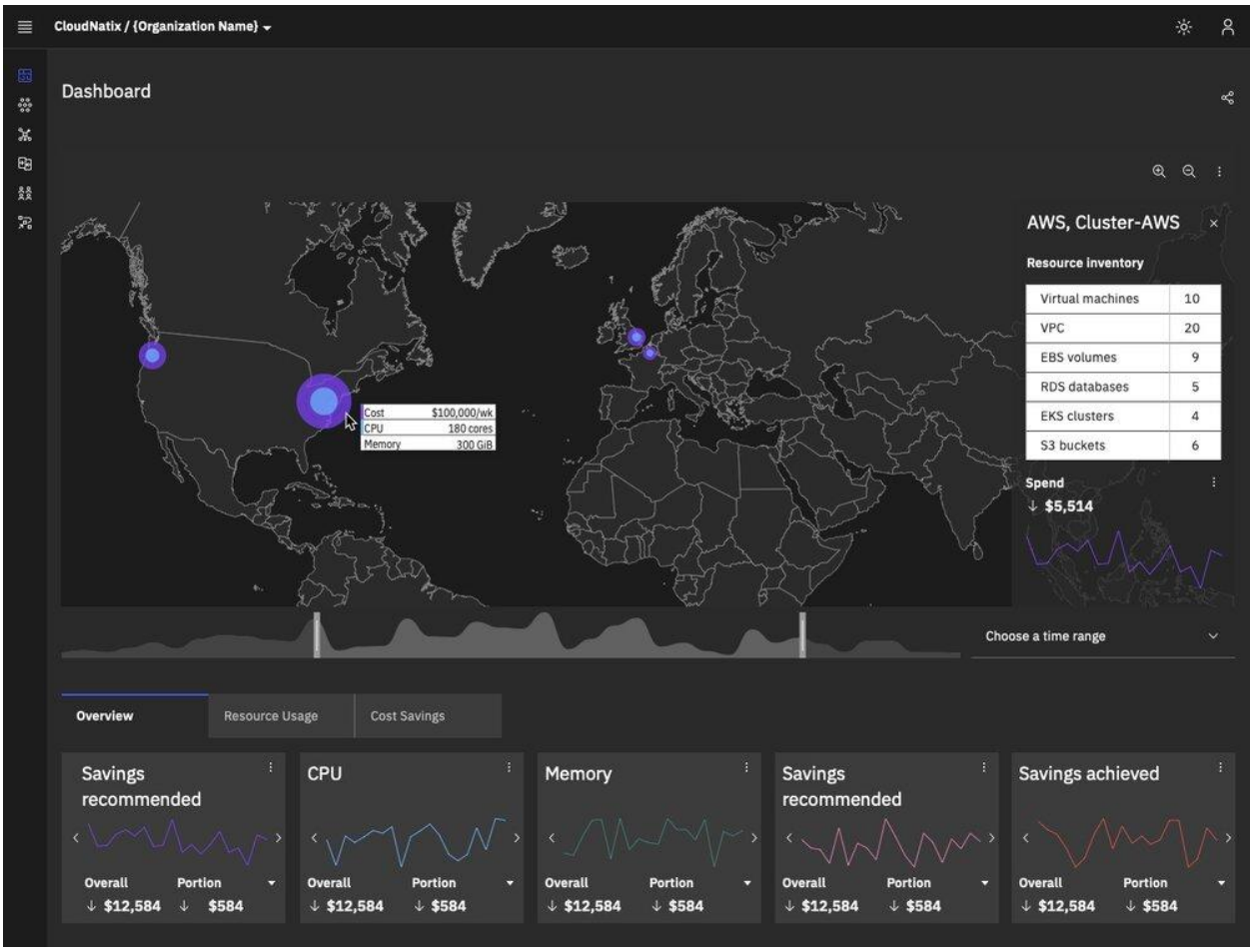
CloudNatix puts cloud operations on autopilot

Works where you work

CloudNatix can connect to any infrastructure, anywhere, from cloud to datacenter to edge, across VM, Kubernetes and Managed Kubernetes clusters.

Observe with CloudNatix Dashboard

Global Cost & Operational Intelligence across your multiple Public/Private Cloud & K8s clusters.

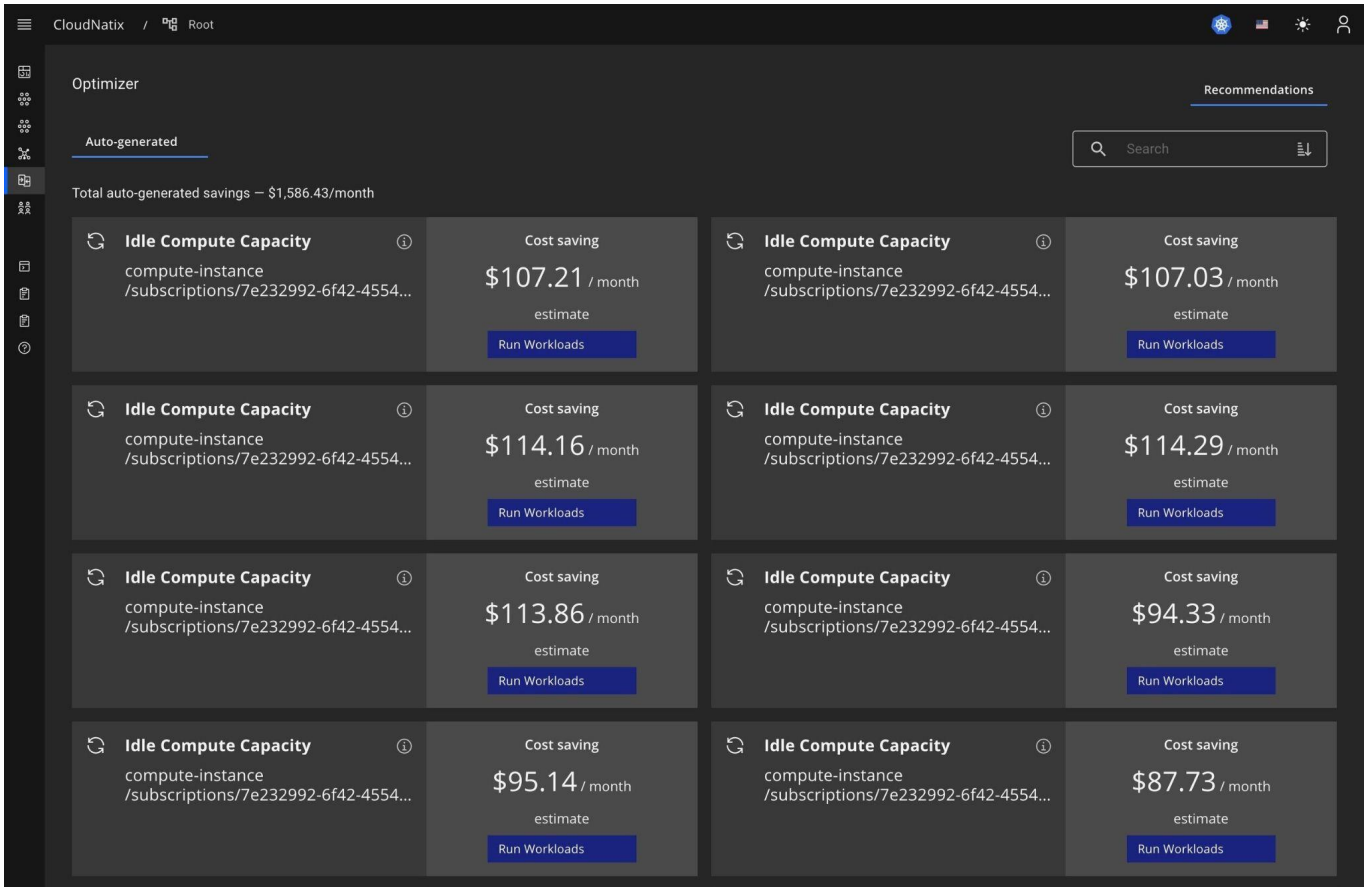


© 2021 CloudNatix Inc. All rights reserved.

CloudNatix puts cloud operations on autopilot (contd.)

Understand with CloudNatix Insights

Understand your opportunities to optimize for cost and automation using CloudNatix recommendations engine.

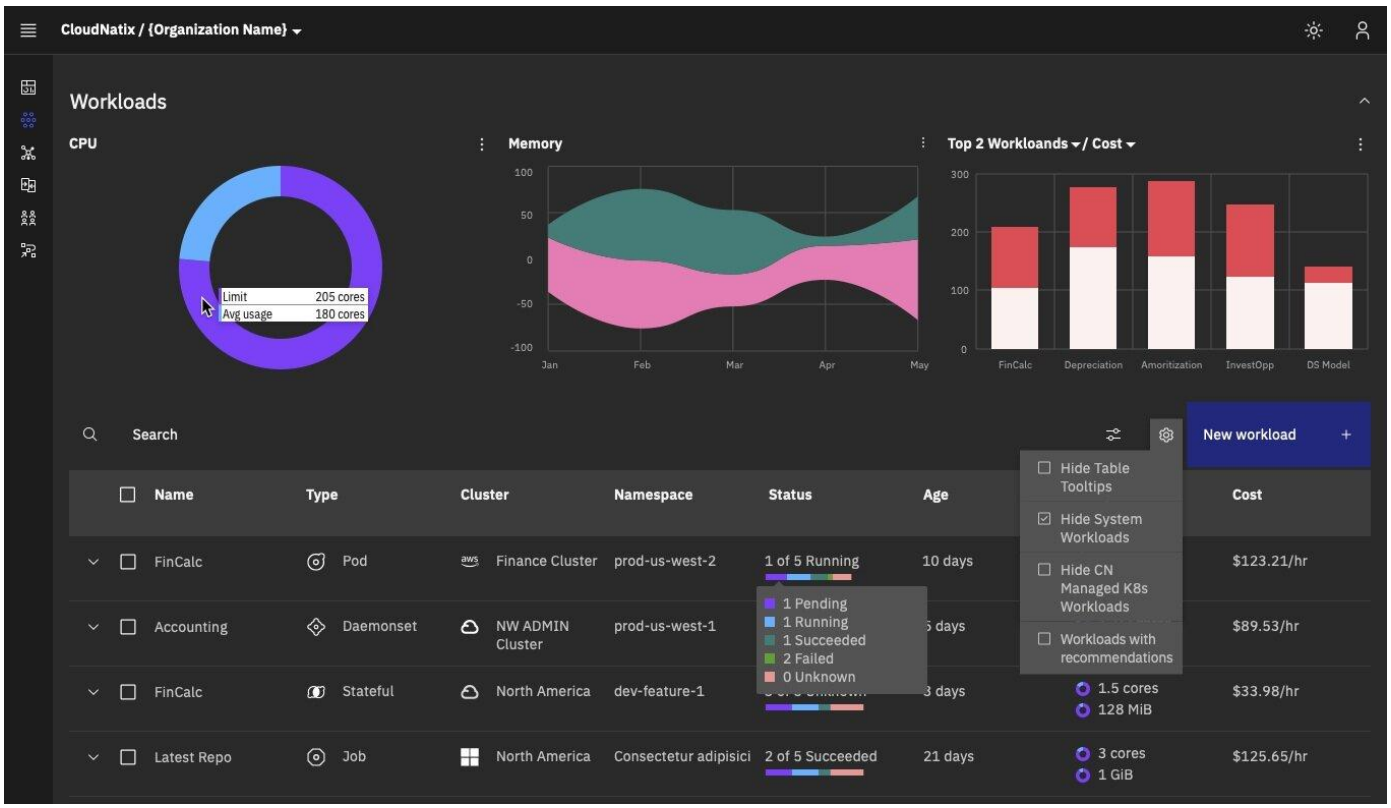


© 2021 CloudNatix Inc. All rights reserved.

CloudNatix puts cloud operations on autopilot (contd.)

Act with CloudNatix Operations manager

A federated global control plane to operate and optimize multiple K8s and VM clusters with a unified command and control center.

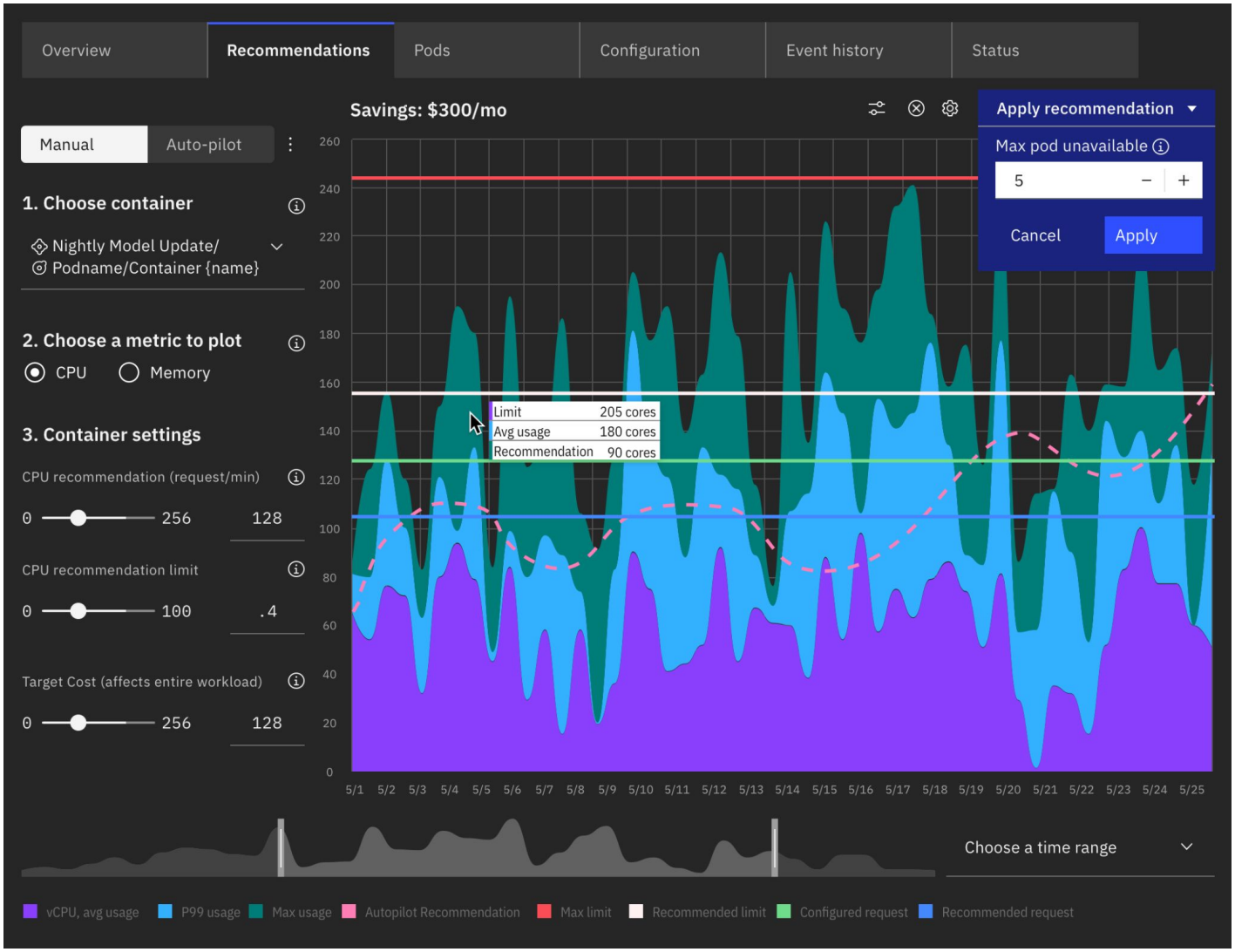


© 2021 CloudNatix Inc. All rights reserved.

CloudNatix puts cloud operations on autopilot (contd.)

Automate with CloudNatix Autopilot engine

Perform human-in-the-loop impact analysis of CloudNatix Insights, and when satisfied put your VM and K8s clusters on Autopilot to ensure your environment is always cost optimized while meeting SLOs.



© 2021 CloudNatix Inc. All rights reserved.

Conclusion

The bottom line is that the architecture that web-scale companies are using is the same one enterprises will need to adopt tomorrow to remain competitive. That architecture relies on hyper-automation and a set of techniques dubbed as “Autopilot” in a paper published by Google [11].

Alternatively, these enterprises will cede market share and relevance to the next webscale innovator to disrupt an industry. In adopting this Autopilot model, enterprises will have to deal with unified access management, controlling multiple clusters, resource management and more, all across multiple regions and clusters of capacity, aggregating vs disaggregating cluster resources, using logical services that span clouds versus namespaces that span a single cluster.

They will need a planet scale cluster manager to do this, something that can observe costs, operations, performance, and Service Level Indicators(SLIs) across clusters, across public and private clouds, spanning VMS and Kubernetes on a single pane of glass so they can operate above the clouds with visibility, control and hyper-automation. Bandage solutions like running multiple spreadsheet models to figure out how to optimize are not scalable. What’s needed are tools like machine learning to automate and optimize at scale, bin packing technology to intelligently and automatically optimize across a planet scale infrastructure fleet, and high levels of automation either with humans in the loop or out. That’s how hyperscalers and webscalers are already doing it.

It’s not a fiction: it’s already happening across web-scale companies, not just hyperscalers. Autopilot operating model is the path that every enterprise is going to need to follow if they want to transform themselves to compete with the next web-scale upstart. The good news is, the pioneers have left us a map to follow and CloudNatix can be your partner in helping put your enterprise cloud and cloud-native infrastructure on autopilot.

Infrastructure utilization	Cost Saving	Devops to Server (VM)	Developer Velocity
CloudNatix	*CloudNatix*	*CloudNatix*	*CloudNatix*
50-60%	**25-60%** (Automatic)	**1 to 1000**	**High**
w/o CloudNatix	*w/o CloudNatix*	*w/o CloudNatix*	*w/o CloudNatix*
20-25%	**5-8%** (Manual scripting)	**1 to 50**	**Low**

Results from adopting autopilot operating model

© 2021 CloudNatix Inc. All rights reserved.

Appendix

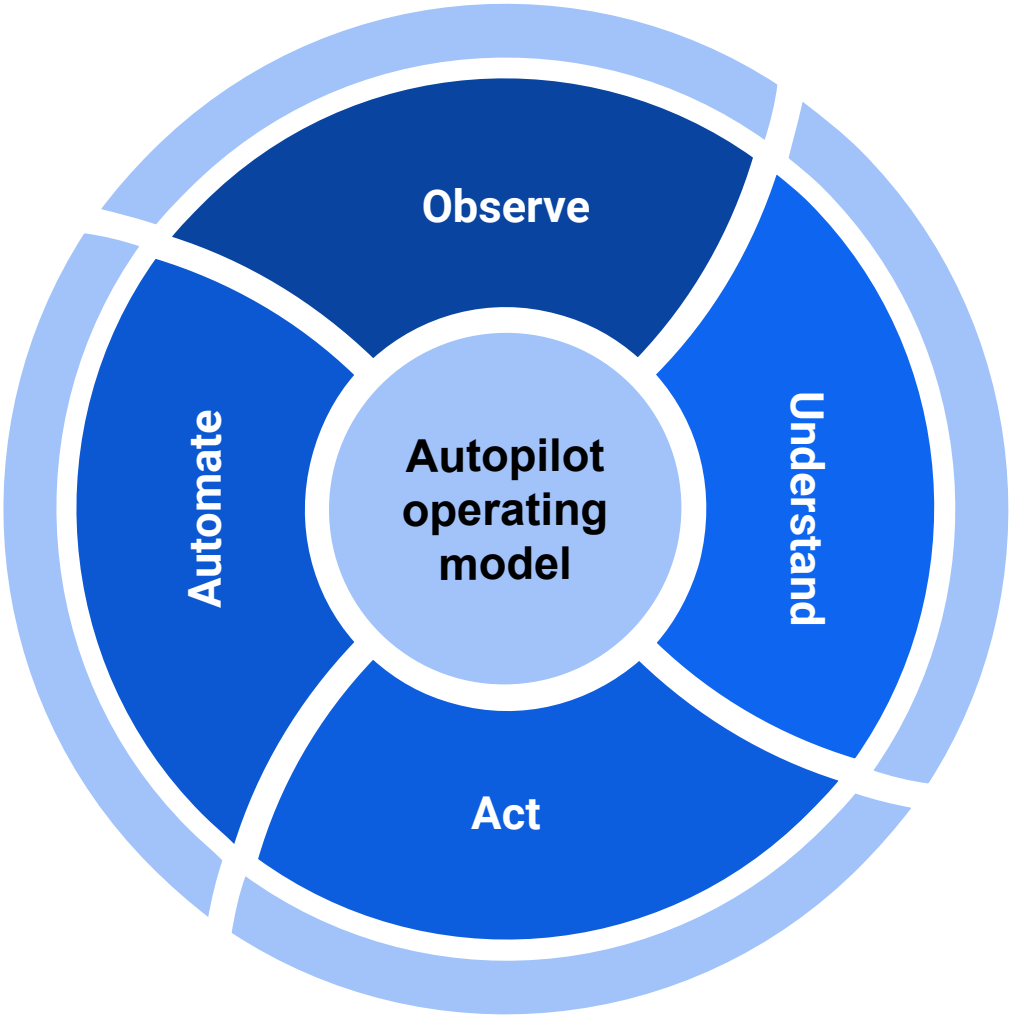
[1] https://www.hashicorp.com/blog/hashicorp-state-of-cloud-strategy-survey-welcome-to-the-multi-cloud-era
[2] https://www.cncf.io/wp-content/uploads/2020/12/CNCF_Survey_Report_2020.pdf
[3] https://www.cncf.io/wp-content/uploads/2020/12/CNCF_Survey_Report_2020.pdf [Page 4]
[4] https://www.cncf.io/wp-content/uploads/2020/12/CNCF_Survey_Report_2020.pdf [Page 4]
[5] https://www.cncf.io/wp-content/uploads/2020/12/CNCF_Survey_Report_2020.pdf [Page 9]
[6] https://www.cncf.io/wp-content/uploads/2020/12/CNCF_Survey_Report_2020.pdf [Page 8]
[7] https://data.finops.org/
[8] https://data.finops.org/ [challenges section]
[9] https://www.eurosys2020.org/wp-content/uploads/2020/04/slides/149_rzadca_slides.pdf
[10] https://www.businessinsider.com/companies-waste-62-billion-on-the-cloud-by-paying-for-storage-they-dont-need-according-to-a-report-2017-11
[11] https://dl.acm.org/doi/pdf/10.1145/3342195.3387524

© 2021 CloudNatix Inc. All rights reserved.

To learn more about about how CloudNatix can be your partner in the journey to Cloud, cloud-native and beyond, go to www.cloudnatix.com.

contact@cloudnatix.com | www.cloudnatix.com | @cloudnatix

© 2021 CloudNatix Inc. All rights reserved.


Observe
Understand
Act
Automate
Autopilot operating model