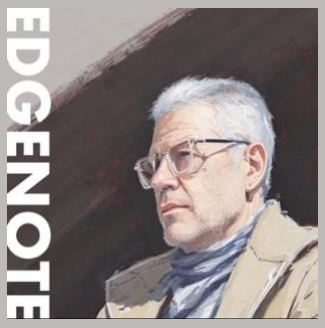


Across and Down: Orchestrating from Clouds to Edges

Pete Bernard – EDGECELSIOR

KEY THEME – PROLIFERATION OF SPECIALIZED CLOUDS WILL DRIVE WORKLOAD ORCHESTRATION AND MAKE WINNERS OUT OF EDGE PLATFORMS THAT LEVERAGE CLOUD NATIVE DEVOPS

In the beginning . . .

. . . it was all "cloud." Although back then, it was just a data center, usually in a basement of a non-descript building. I remember toiling down in one of these rooms at Boston University back in the 80s, humming with folks trying to get their jobs to run without errors.

Since I lived about a mile away, often in the cold Boston snow or sweltering heat, I quickly ginned up an "edge device" – a terminal connection from my dorm room over a slow modem (an IBM PC-1 with a fine monochrome monitor). Suffice to say, there was virtually NO processing happening on that edge device, no workload management and certainly no virtual machines or containers being spun up.

And since then, PCs phones and all kinds of edge devices have become way more sophisticated, and our hyper-scalar cloud build-out resembles a scene from The Creator. Our analog modems have been replaced with dark fiber, mmWave 5G and LoRaWAN.

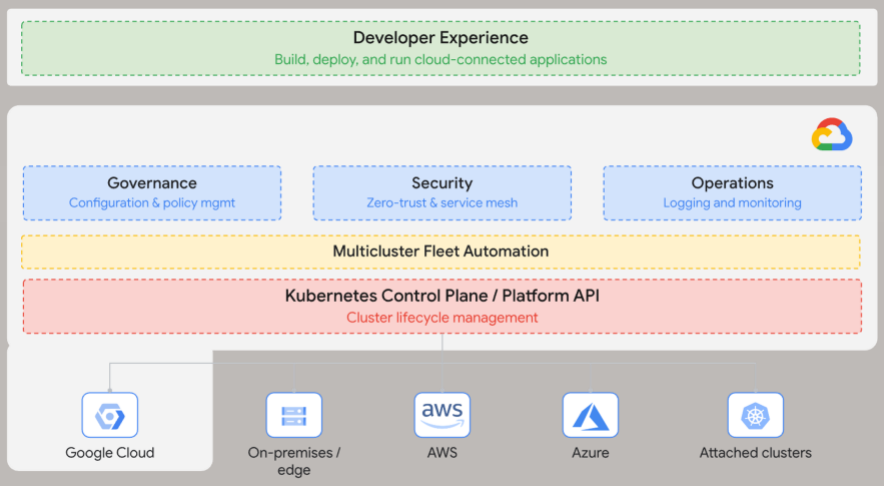
And yet, the architecture remains basically the same. You have a job that needs to run (without errors) – and that job, or workload, is running on the cloud, or on a PoP near the edge on a server box, or maybe a gateway on-prem, or maybe even on a sensor on the shelf at your local grocery. These workloads are running AI models, or simple compute, or updating software or firmware. How do you manage these workloads – when and where they run, what permissions they have, how does the data flow? This is the magic of orchestration, which means "the automated coordination and management of multiple systems, applications, and services."[1]

[1] redhat.com/en/topics/automation/what-is-orchestration

Orchestration is a nascent but healthy market. According to Insight Partners, the global cloud orchestration market size is expected to grow from USD 13.87 billion in 2021 to reach USD 53.40 billion by 2028; it is expected to grow at a CAGR of 21.5% from 2022 to 2028.[2]

Before we get into edge workload orchestration, we need to acknowledge that the advent of workload of orchestration into mainstream dev ops is bringing some really interesting "second order implications." If you can now manage workloads from cloud to edge, why not orchestrate them across clouds as well?

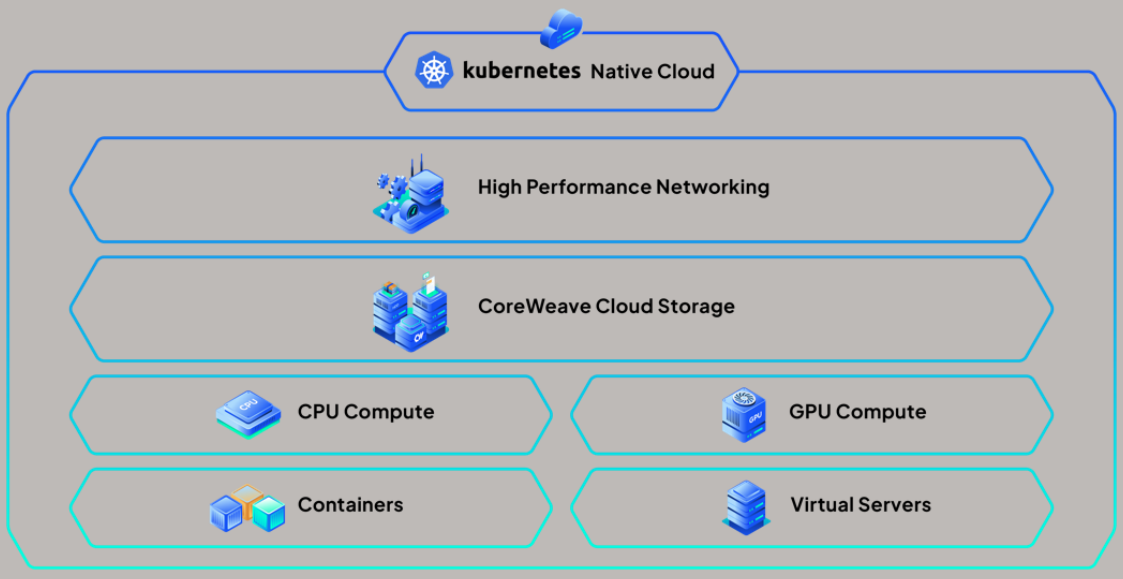
I have worked with many Fortune 50 companies on their digital transformations and virtually every one of them has a multi-cloud strategy. Yes, they may run many workloads on Azure but they may also be running on AWS, or GCP, or some workloads on Oracle – all for different reasons.



These companies may also have a legacy data center where they keep other workloads running, or want to shift workloads around to mitigate cost, availability, or run on "greener" platforms to benefit their carbon profile. Orchestrating these workloads require a new set of tools that have been coming on the market in the past couple of years, including Anthos from Google, Azure Arc from Microsoft, AWS CloudFormation and more "third party" orchestration tools that perhaps provide a more "neutral" approach, such as Terraform or Cloudify. It's too early to tell who will lead here and it's not necessarily a winner-take-all market.

Azure Arc[3], as an example, is a collection of capabilities specifically designed to "pull in" non-Azure data center workloads into a single pane of glass and deploy cloud like services in a cloud agnostic manner. The advantage of that platform is that you can keep a consistent dev environment with any GitOps compatible CI/CD framework, and still leverage your legacy data centers and alternative clouds.

[2] Cloud Orchestration Market Size to Hit $53.40 Billion, by (globenewswire.com)
[3] Azure Arc – Hybrid and Multi-Cloud Management and Solution (microsoft.com)



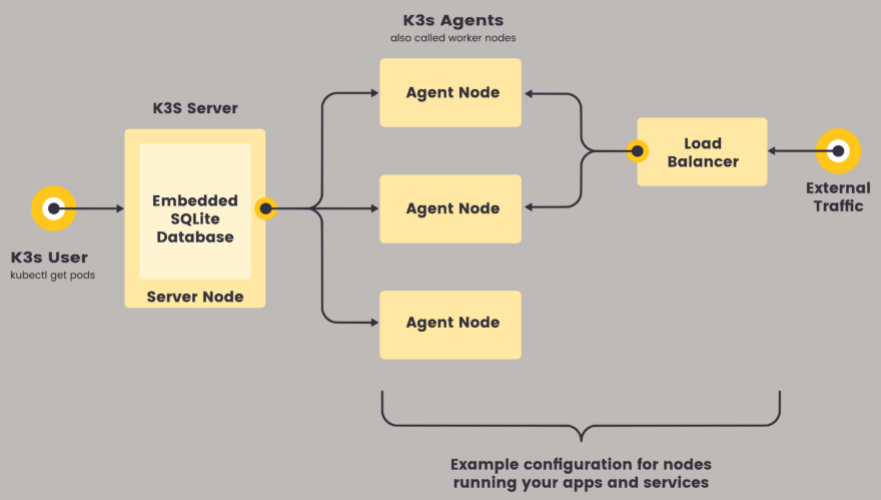
This multi-cloud capability is creating some really interesting offerings. Look at CoreWeave, who quickly established themselves with an AI focused cloud (with heavy investment from NVIDIA) and are looking at a $7b valuation[4]. We could see a proliferation of "industry vertical" clouds that are targeted at Healthcare, FinOps, sovereignty within geographies, and more. The hyper-scalar cloud universe could end up deconstructing itself into these more special purpose clouds, where performance, cost and capabilities will be fine-tuned for these specific workloads.

Now that we have kicked up so much dust here on managing cloud workloads, how do we then extend that capability outside the data center to the edge? This is easy to whiteboard but hard to implement and deploy.

The primary challenge is that creating a CNCF compliant container that acts like a cloud container requires device resources, especially memory, and edge device value propositions hinge on their judicious use of resources and cost. There is a paradox between the desire of consistent cloud-native devops on the edge and the gravitational pull to use only as many resources needed to get the job done. Fortunately, the CNCF[5] community has been working on taking the "K8" construct into resource-constrained edge devices with constructs like K3s. This provides an orchestration capability for workloads but with functional subsets in scheduling, self-healing, and other Kubernetes capabilities. There are other efforts for lighter-weight Kubernetes environments such as MicroK8s[6] (pronounced "microkates" and sold/serviced by Canonical for x86/Arm and Linux/Windows and Mac). Minikube[7] is another lighter-weight Kubernetes environment, although it too requires some edge resources – a minimum of 2GB of RAM and 20GB of storage.

[4] CoreWeave nearing stake sale that will value company at $7B - Bloomberg | Seeking Alpha
[5] Cloud Native Computing Foundation (cncf.io)
[6] MicroK8s - Zero-ops Kubernetes for developers, edge and IoT
[7] Welcome! | minikube (k8s.io)

A good example of an edge-based K8/K3 implementation is the Azure Kubernetes Services in Windows IoT (yes, a link for The EDGECELSIOR Show episode on this topic is here.) These types of "medium edge" environments play well with orchestration systems like Azure Arc, or Zededa[8], or Aarna Networks[9] for telco orchestration.



Beyond the MicroK8s, Minikubes and K3s – into the light and tiny edges - things are getting a bit murkier and creative. There is the notion of a K0 – not a CNCF project yet – and a project called AKRI (A Kubernetes Resources Initiative)[10], which is an actual CNCF project to run container-based orchestration on gateways that have multiple "leaf" sensors hanging off of it. There are also companies like MicroEJ[11] and Wind River with hypervisor based embedded containers[12].

Net net we should continue to see a great deal of innovation here as efforts begin to settle and deployments happen at scale. Cloud AND edge orchestration based on Kubernetes and container-based DevOps is here to stay, with functions including traditional compute but also fresh-trained AI workloads to even device provisioning and FOTA.

Companies that are betting their digital transformations on cloud to edge solutions will be insisting on seamless orchestration of workloads ACROSS clouds – from multiple hyper-scalars to "boutique" cloud providers, but also DOWN to their edge devices – from blade serves on prem to gateway, cameras, automotive IVI and CVC systems, and beyond to Cortex-M based sensors.

[8] ZEDEDA Home
[9] Simplify Edge Orchestration | Aarna Networks
[10] GitHub - project-akri/akri: A Kubernetes Resource Interface for the Edge
[11] MicroEJ - Software-Defined Everything - Design performant, secure, reliable and future-proof electronic products thanks to MicroEJ, the "tiny sibling of Android" for home appliances, industrial gateways, wearables, smart printers, medical devices, smart city and more.
[12] What Are Embedded Containers | Wind River | Wind River

Edge device builders would be wise to have container-based orchestration capabilities with service partners, orchestration companies should be enabling as broad of an ecosystem they can, and commercial customers should be insisting on a cloud devops approach to workload orchestration.

ACROSS and DOWN – Orchestrating a cloud of clouds all the way to the tiny edge.

We've come a long way.