REPORT REPRINT

Cloud-native ecosystems will answer the call for edge computing support

JUNE 4 2020

By Brian Partridge, Jay Lyman

Edge computing will see hybrid cloud resources exploding out of centralized datacenter locations to users and devices, while cloud-native technologies can deliver a spectrum of abstractions of underlying multicloud IT infrastructure. These two technology trends reinforce each other, but elements of the cloud-native stack must be optimized to work well for the constraints of edge computing systems.

451 Research®
Now a Part of
S&P Global Market Intelligence

Introduction

Cloud-native technologies can deliver a spectrum of abstractions of underlying hybrid, multicloud IT infrastructure, while edge computing physically arranges hybrid IT infrastructure closer to users and devices. These two technology trends reinforce each other, but elements of the cloud-native stack must be optimized to work well for the constraints of edge computing systems.

451 TAKE

Several trends are converging and maturing in unison – edge computing, cloud-native technologies, and emerging workloads such as IoT and machine learning conducted on edge machines and in the cloud. The requirement to develop software that is easily portable and can be run on any hybrid, multicloud infrastructure including edge locations will require cloud-native tool chains and strong DevOps practices. We see these trends as overlapping and reinforcing. We contend that cloud-native technologies are necessary to ensure that edge computing can achieve business outcomes with the right speed and security, and at scale.

Containers, VMs and Kubernetes are becoming critical to enterprise DevOps teams for speed, efficiency and flexibility, but in some cases edge deployments will require specific optimizations to deal with the constraints that come with edge such as limited compute and storage resources and intermittent network availability. We also see edge deployments emerging as yet another venue for cloud-native implementations that have already driven hybrid and multicloud architectures that span on-premises, private clouds and multiple public clouds. Support for these different environments has enabled many vendors to expand their hybrid IT story to include edge.

Why cloud-native for edge?

Edge computing has become an incredibly hot topic, so much so that it often feels like the entire IT ecosystem is staking a claim in one form or another. Edge computing setups offer a mix of performance benefits (such as lower latency, autonomous operations if a network link goes down), cost savings (by reducing data backhaul to centralized computing locations), and security and data sovereignty advantages.

With so much edge infrastructures coming online, a DevOps environment that can support any number of execution venues will be ideal in a small datacenter, wiring closet, gateway or extending all the way down to relatively low-end devices.

Containers are well suited to edge applications where consistent performance is required across a wide variety of infrastructure, including resource-constrained, edge locations – think of ruggedized servers, but also cars, satellites or ATM machines. Cloud-native software darling Kubernetes is also well-positioned to handle the management and orchestration of container environments, as well as distributed infrastructure and applications.

Using Kubernetes, containers are able to run stand-alone (nodes) or as part of clustered environments, with the ability to easily upgrade from a centralized control plane (master). The lightweight, scalable and ephemeral nature of cloud-native software is aligning with advances in hardware and use cases, such as Raspberry Pi – a low-cost/low-footprint computer, sized like a credit card.

As more cloud-native providers focus on edge, additional advantages of containers are emerging, such as faster rollbacks so edge deployments that break or have bugs and can be rapidly returned to a working state. We are also seeing more granular, layered container support whereby updates are portioned into smaller chunks or targeted at limited environments, and thus don't require an entire container image update.

Edge challenges

The challenges of deploying and managing workloads on edge infrastructure are well suited to the capabilities of cloud-native technologies, with some important modifications. Edge computing environments will bring a mix of underlying infrastructure capacity, and will be physically distributed at a massive scale. Managing application performance of edge workloads will require a significant amount of automation to avoid the need to physically visit infrastructure.

While containers are lightweight enough to run in edge environments, Kubernetes wasn't designed with edge computing in mind, since most Kubernetes distributions don't support the entire portfolio of ARM reference designs (such as Raspberry Pi, popular in IoT), and can require 1-4GB of available RAM to properly function. So limited edge node resources must be dealt with. Luckily, there are several options to deal with limited compute capacity. The other issue is that edge nodes must be able to run with intermittent or limited network capacity, so edge architectures must be designed to operate in the absence of a consistent network.

Edge may require lightweight K8s

Several options are available for edge application developers who wish use Kubernetes container orchestration techniques but require lighter-weight implementations due to the resource constraints that, by definition, characterize edge computing setups.

We describe two of the most popular below – MicroK8s and K3s – but there are others including Minikube, KubeEdge and Kind. Each of these options has its pros and cons, and choosing the right one typically comes down to workload requirements, operating system in use, infrastructure and chipsets in use, and organizational preference.

MicroK8s – This is a fully certified distribution developed by the Kubernetes team at Canonical, so optimized for Ubuntu Linux environments. The container delivery system allows for 24-hour updates, and advertises essentially no downtime. Canonical describes this as Zero Ops.

K3s – This fully certified distribution from Rancher is designed for both development and production environments. K3s is a 'stripped down' version of the full Kubernetes stack (about 25% of normal size). This allows for a 40MB binary application download that can operate with just 512KB of RAM.

Recommendations

If you are considering an edge deployment, we recommend initially beginning with application vision, requirements and expected outcomes. After targeting edge use cases that can support a desired business objective such as lower operations costs, higher or diversified revenue, and greater customer intimacy, you can turn attention to the system required.

The goal is to design a system that is secure, well integrated with existing systems and processes, and designed for scale and performance. The performance requirements will narrow the underlying technology choices, including edge node compute capacity and network setups, and which distribution of Kubernetes will make sense. This is where cloud-native technologies such as containers and optimized Kubernetes have an opportunity to shine.