WHITE PAPER

Taking it to the Edge

Where edge computing is going and how enterprises can successfully scale and deploy to it

Taking it to the Edge

Where edge computing is going, and how enterprises can successfully scale and deploy to it

Edge deployments—specifically, the concept of “scaling the edge”—is one of the most challenging directives an organization can face in today’s economic and technology landscape.

It’s also a difficult concept to summarize, since the edge can mean different things to different organizations.

Enterprises that have mostly transitioned to cloud-native operations or were “born in the cloud,” such as smaller organizations and startups, often see the edge as an extension of their public cloud. Or, they see it simply as a resource belonging to a cloud vendor that can be leveraged on demand.

But for mid-sized to large organizations with hybrid workloads, both on premises and in the public cloud, the edge can be seen as an extension of their network. It may not be the best performing part of it, but it is vital to their business and deserves more attention.

This white paper is focused on this second group of organizations and how they can successfully leverage edge computing technologies to deliver better products, services, and experiences to customers.

Specifically, we’ll be covering:

- Trends at the edge
- Edge deployment considerations
- How to create an edge deployment strategy
- Running Kubernetes at the edge

PART 1:

Trends at the edge

In many ways, the rise of the edge has coincided with the rise of the internet of things (IoT).

The proliferation of connected devices that can communicate through the internet or through other devices has created a massive amount of data that companies across industries can utilize—if they are able to process all of it efficiently.

Given that IoT devices are generally constrained and limited, at least from a compute perspective, the edge compute component is increasingly valuable.

Sending an ocean of data between IoT devices in the wild and datacenters or the cloud is not just time-consuming, it can quickly become prohibitively expensive. And while edge computing doesn't entirely solve these problems, being able to process at least some of the data locally can help dramatically.

For a sense of where edge computing is going, let's look at two industries: manufacturing and retail.

In manufacturing, data is increasingly being mined for things like safety measures and robotic learning systems. These are being powered by sensors that collect an almost unimaginable amount of data. When a hazardous situation arises, or a segment on a line suddenly fails, the ability to correct the issue immediately is critical.

With compute of that data located at the edge, manufacturers are able to identify and remedy issues much more quickly than if the data had to travel outside of the facility. They're also able to provide employees with the ability to react in real-time.

Similarly, in retail, companies are able to lean on edge computing to quickly flag items that show a spike in popularity in a given region and promptly organize their products in stores to capture the attention of buyers while the iron is still hot.

Edge computing is also informing everything from store security and theft protection, to alleviating the need to have static point-of-sales locations within stores—or physical brick and mortar stores altogether.

Then there is the inevitable issue of a networked point-of-sale device going down across a retailer's footprint, which in the past would lead to a rather dramatic loss in sales until the system is back up in running.

But with edge computing, these outages are much less of a concern since critical devices at stores can still function properly even as their communication to the company's headquarters is compromised.

PART 2:

Edge deployment considerations

When deployed successfully, edge solutions can open up entire new avenues for organizations to create a competitive advantage. Among those avenues are:

Localized real-time data analysis

Processing data at edge sites rather than a centralized hub can potentially lead to faster data-driven decisions.

Reduced operational costs and storage needs

Processing data locally can also potentially cut down on transmission and storage costs, as well as bandwidth and throughput concerns.

Enhanced security

Ensuring security at the edge requires site key management, data encryption at rest and in flight, and other considerations.

In order to open these and other avenues, organizations need to take into account three considerations.

1 Infrastructure and configuration management

Deploying and scaling at the edge is not just about devices and sensors. It's also about managing an assortment of hardware components and platforms with varying capabilities and configurations in both the hardware and software stacks.

Configuration drift is a considerable challenge, and managing that drift—from various edge sites to evacuating edge workloads due to planned or unplanned downtime—can be complex.

2 Edge visibility

Due to the nature of edge environments being very heterogeneous and, in some cases, isolated with degraded or under-provisioned networks, visibility becomes a much bigger challenge at the edge compared to centralized cloud and datacenter ecosystems.

Real-time traffic visibility is not enough because when something goes wrong at the edge, your response, mitigation, and resolution timeframes are dramatically extended. This greatly increases the need for predictive and proactive analytics to create insights into your edge network.

3 Last mile connectivity

Co-opting a term from the telco world, last mile connectivity is more important than ever now that we are on the cusp of global 5G transformation.

At the edge, any extension of a centralized and distributed network pushed to the outer limits will encounter typical issues inherent to all edge environments—latency, jitter, bandwidth, quality of service, and throughput.

While the above recommendations focus on pre-deployment considerations, there are still additional layers to building an effective and holistics strategy. Which brings us to...

PART 3:

How to create an edge deployment strategy

If your organization is at a point where it is ready to create or build an edge deployment strategy, there are four critical factors to consider. These are:

Carrier independence

For greater flexibility in service and pricing, your strategy should be designed for multiple carriers with proper coverage, bandwidth, and reliability.

Additionally, design considerations can involve private long-term evolution (LTE), multi-access edge computing (MEC) layers, and network edge computing (NEC) deployments across different carriers where applicable.

The goal is to avoid being locked into any one carrier in order to maximize potential cost savings and achieve flexibility in networking.

Compute horizon

The concept of "compute horizon" (a term coined by Redapt) is the act of spreading out edge computing power from the device edge to the centralized network—and everywhere in between.

The goal is to distribute compute resources in such a way that if network degradation were to occur at any point, an organization could still carry out decisions closer to the device edge.

Having this flexibility to process data at various points near your edge sites allows you to optimize costs, speed up decision making, and reduce data transfer and processing overhead.

The various mesh- and "fog" -based topologies are designed to increase the concept of "velocity to insight," but they are not without caveats. There is a CAPEX/OPEX conversation to be had within your organization when evaluating cost/benefit analysis to distributed decision making.

5G adoption and mobile edge

With so many organizations looking to capitalize on 5G data transfer speeds, it's easy to lose sight of the fact that 5G adoption and infrastructure are still sporadic across the globe.

Conversations around how to leverage the potential of (MEC) is crucial for organizations looking to take advantage of 5G at the edge. MEC infrastructure is very complex, but the premise is very simple: a network architecture bridging cloud compute with the radio-access network (RAN) edge of a mobile network.

The key takeaway here is that MEC is no longer reserved for telecommunications operators. It is a continually evolving technology that organizations will increasingly be able to take advantage of to grow and impact their businesses.

Security

When it comes to ensuring security at the edge, it is essential to focus not just on cloud environments and software but also on the underlying platforms.

While security practices like public key infrastructure (PKI), authentication and compliance, modern KMS, and other on-premises and cloud-native security services are critical, a key confidential computing technology to explore is a security platform with the ability to leverage hardened security features based on the concept of a trust execution environment (TEE) within chipsets.

For example, Intel's flagship chip security Software Guard Extensions (SGX). By leveraging the SGX ecosystem, you can secure on-premises, hybrid, multi, and edge cloud environments by tapping into secure memory enclaves at the hardware level independent of the OS or any hypervisor.

PART 4:

Running Kubernetes at the edge

Edge Capable Architecture is an edge site able to run entire Kubernetes clusters. There is also a lightweight distribution of Kubernetes that can reduce the edge resource requirements even further, such as those devices sporting lower power ARM-based chipsets.

A use case for this would be legacy M2M remote environments heavily focused on local data processing or situations where network SLAs are either not reliable or not important.

While this is an evolving topic, for a baseline reference let's look at some of the most common approaches for deploying Kubernetes at the edge:

KubeEdge Architecture involves placing the control plan in a public IaaS or within a datacenter while the worker nodes reside at the edge. Because this is a completely decentralized deployment model, the edge nodes can devote a lot of resources to performance rather than capacity given the management plane separation.

A particular type of artificial intelligence/machine learning AI/ML edge model where an organization needs to refresh and push out newly trained models constantly would do well to adopt this approach. Similarly, organizations looking to leverage edge serverless development or deal with edge bursting, cloud backup, and regulatory compliance scenarios would also do well to adopt this approach.

Virtual Kubelet is an open-sourced kubelet implementation that is essentially an abstraction-based model approach whereby the Virtual Kubelet puts forth all the components that make up a traditional Kubernetes kubelet, effectively allowing the Virtual Kubelet to act as an interface between outside APIs and Kubernetes APIs.

While there may be some interesting cloud streaming IoT use cases where a Virtual Kubelet approach makes sense, the ecosystem services leveraging this technology are still fairly new.

SaaS Optimized Management Planes are in a similar boat as Virtual Kubelet, as there are not too many distributions on the market offering this approach. Still, it's possible that if the pioneering vendors working on this architecture are successful, it could potentially change the way we consume and deploy edge workloads on Kubernetes.

One interesting note about the SaaS Optimized Management Planes approach is its potential for centralized management across all edge sites while still providing a local control plane to edge Kubernetes clusters—which could solve for degraded network and connectivity problems.

Conclusion

Deploying a scalable edge solution that meets the needs of your organization can allow you to deliver better products and services to your customers.

While there is considerable effort involved in deploying scalable edge solutions properly, the investment and risk are usually worth it.

redapt

Is your company interested in deploying scalable edge solutions?

Contact us today at redapt.com/contact