



Seamless Kubernetes on premises and in the cloud



Executive summary

In this paper, we address current challenges with modernizing on-premises systems with Kubernetes. Although Amazon Web Services (AWS) public cloud services offer several advantages, customers still need to run workloads on premises for a number of reasons, including low latency and data locality, which we cover in-depth in this guide. To help address these challenges, we present a set of solutions powered by Amazon Elastic Kubernetes Service (Amazon EKS). Specifically, we focus on Amazon EKS Anywhere, a solution to run Kubernetes on premises, allowing for both perpetual on-premises container and infrastructure management, as well as a vehicle for migrating to the cloud.

This paper is intended to accelerate company decision-makers' evaluation process determine a solution that best fits your business needs.

Preface

In his book *The Paradox of Choice: Why More Is Less*, psychologist Barry Schwartz argues that the freedom of choice has not given us freedom but rather paralyzed us. A quick glance at 2,000+ open-source projects in the Cloud Native Computing Foundation (CNCF) can seem daunting. The explosion of hosting options and management tools means that cloud architects need to make more decisions in their infrastructure management than ever before. We have seen customers paralyzed with choice at nearly every architectural decision point, such as deciding on logging providers, storage interfaces, container runtimes, and more.

To reduce decision fatigue with running containers on premises, AWS launched Amazon EKS Anywhere, a deployment option for Amazon Elastic Kubernetes Service (EKS), allowing customers to create and operate Kubernetes clusters running on their infrastructure virtually anywhere. Amazon EKS Anywhere is bundled with opinionated architectural patterns, tooling, and open-source projects to accelerate your Kubernetes adoption with AWS-recommended approaches and to help alleviate the paralysis of choice.

Introduction

AWS has made cloud computing a broadly accessible and a reliable form of IT infrastructure. You no longer need to commit to significant upfront investments to build and operate data centers before developing applications. Lowering the risk of experimentation with minimal upfront costs means novel ideas are able to go from prototypes to running applications sooner. Cloud computing ushered in an era of innovation, the effects of which we are still seeing today.

But those models are no longer unique to the cloud. On-premises environments benefit from recent innovations in infrastructure management, application packaging, and open-source tooling, minimizing risk and maximizing existing infrastructure investments. In this paper, we take the perspective that innovations driven by public cloud providers are accessible to everyone, whether your workloads are running in AWS Regions or in your own data centers.

Data center modernization

One of the primary benefits of the cloud is allowing someone else to manage the “undifferentiated heavy lifting,” a term coined by Jeff Bezos in 2006 in reference to the value that Amazon Elastic Compute Cloud (EC2) provides to its customers. Undifferentiated heavy lifting refers to work that is not unique to your business—it’s simply the price of admission. For example, managing application infrastructure is critical to your company’s success, but it is by no means unique or differentiated work.

Another benefit of AWS is having a partner who will work with you to accelerate your innovation efforts and support your application needs. Infrastructure has changed a lot since 2006—from physical servers and virtual machines to containers and serverless. Amazon has been at the forefront of many of these innovations, helping customers adopt the appropriate technologies and operate them at scale.

Kubernetes has standardized container and infrastructure management through a common API interface and collection of open-source software. Amazon EKS Anywhere brings the combined Kubernetes and AWS partnership to your data center. Modernization isn’t simply deploying specific software; it’s also leveraging partner expertise with AWS to accelerate your innovations.

Despite the amount of undifferentiated work associated with managing data centers, customers still have several reasons to run workloads in on-premises or edge environments. These include:

- **Capitalizing on existing investments**
- **Data locality requirements**
- **Latency-sensitive compute**

Let’s cover these in-depth.

Capitalizing on existing investments

Your data center is a long-term investment. There is significant capacity planning, training, and process creation that require specialties for your company and industry. You have invested in the hardware and space required to build a data center, and you've equally invested time to train people to operate the hardware you've purchased.

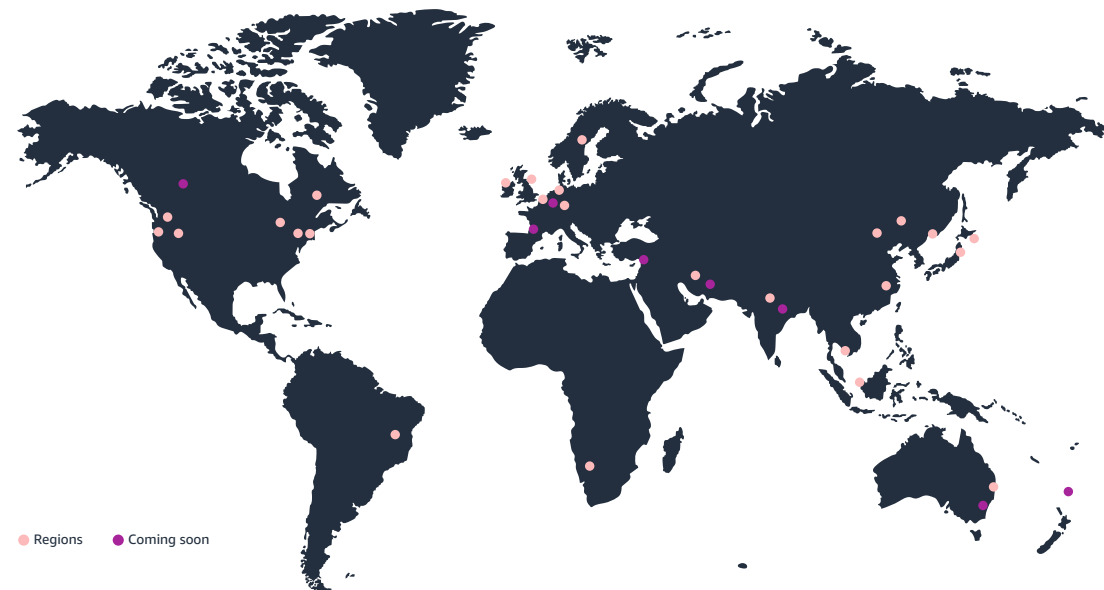
This is not something that can be easily migrated away from before receiving your return on investment (ROI). The amount of time to see returns varies for every company; it is often calculated using hardware refresh cycles, project funding cycles, or general industry trends for your company.

Once you have realized your ROI, you can choose to keep the existing processes and expertise to reinvest in your data center, or you can choose to partner with an external service provider with new investments. No matter what decision you make, you can still modernize with Kubernetes on premises to help gain efficiencies, deployment speed, and stability. In addition, if you choose to migrate to the cloud at a later date, running Kubernetes on premises can greatly streamline the migration process.

Running a data center doesn't exclude you from taking advantage of cloud services. For example, leveraging databases in the cloud while running compute on premises could augment your development team's skill set with the expertise of AWS to manage provisioning, maintenance, and reliability. This allows you to focus on training and hiring the best employees for your business.

Figure 1: AWS Global Infrastructure Map

The AWS Cloud spans 84 Availability Zones within 26 geographic regions around the world, with announced plans for 24 more Availability Zones and eight more AWS Regions in Australia, India, Israel, New Zealand, Spain, Switzerland, and United Arab Emirates (UAE).



Latency-sensitive compute

Some environments don't have regulatory restrictions, but they may have physical limitations. Communication is only as fast as the speed of light. Every mile of wire between your CPU and your data adds 8.2 microseconds of latency. It doesn't matter if that mile is from another data center, cloud provider, or wires within your data center. If latency matters, location matters.

For some applications, latency to your customers is important. For applications that need real-time processing like augmented reality, autonomous vehicles, game streaming, and real-time trading—lower latency can improve the customer experience and create a competitive advantage. Moving your data processing closer to your customers is a great option for lower latency.

This is especially true for applications that serve functions like routing for telecommunications. You may need your business logic to be near the routing layer to make fast decisions for customer calls or to serve data from the closest Point of Presence (POP).

Industries that process large amounts of data, such as the film and entertainment industry, develop petabytes of information requiring rendering or processing. The data is often created in a studio where artists create digital assets or filmmakers record directly to network storage. Moving this data takes time. Adding compute close to the storage is architecturally the simplest and most performant solution.



Why Kubernetes for on premises and edge?

Since its release in 2014, Kubernetes has exponentially grown in popularity because of its ability to adapt to a multitude of different applications and environments. Today, it is the key to unlocking cloud-native practices in your own data centers and applications. To learn why Kubernetes is ideal for seamless management of workloads across on premises and the cloud, we first need to understand the rise of containers: the technology that paved the way for cloud-native applications as we know them.

Applications start as simple solutions to a problem. As time passes, with more features, bug fixes, and other changes, they can sprawl the same way that infrastructure did. Applications grow in size and complexity and may need to be separated into multiple components. Today, instead of creating large, feature-complex applications, many companies separate them into domain-specific services—an architectural pattern known as microservices.

Containers help provide consistent application packaging and deployment and are ideal for running microservices. They give the flexibility to package practically any application—internally developed or by a third party—with all of the library dependencies required to run it. However, operations teams are faced with a new set of challenges associated with managing containers: reliable deployment, service discovery, scheduling, and dynamic scaling. Collectively, these challenges can be solved with container orchestrators.

Kubernetes was created to address these problems and more. Now is a great time to get started with Kubernetes, but the ecosystem has grown so rapidly that it is hard to know where to start. With a seemingly endless number of architectural decisions for managing Kubernetes on premises, customers are faced with a paradox of choice.

Amazon EKS for on premises, the cloud, and edge

Today, tens of thousands of customers choose to run their Kubernetes workloads with Amazon EKS because of its security, reliability, scalability, and ease of management. Through Amazon EKS, AWS enables customers to run Kubernetes wherever their applications require. Customers can run the same AWS-managed Amazon EKS offering in **AWS Regions**, **AWS Outposts**, **AWS Local Zones**, and **AWS Wavelength Zones** for a consistent Kubernetes experience across environments.

Customers who want to leverage existing investments have the flexibility to run Amazon EKS Anywhere on their hardware. With this breadth of deployment options, Amazon EKS meets customers where they are so they can modernize at their own pace with a consistent set of solutions, whether they are migrating to the cloud or modernizing their on-premises and edge environments.

Figure 3



At the root of all Amazon EKS deployment options is **Amazon EKS Distro**, an open-source Kubernetes distribution built and maintained by AWS. This is the same Kubernetes distribution used by Amazon EKS in the cloud and on premises in **Amazon EKS Anywhere**. AWS offers a number of ways and locations to deploy Amazon EKS clusters:

- **AWS Regions** – in a single AWS Region across one or more availability zones, powered by AWS-managed data centers.
- **AWS Local Zones** – in industry centers and close to large populations, this type of infrastructure deployment includes compute, storage, database, and other select AWS services.
- **AWS Wavelength Zones** – within 5G networks, AWS Wavelength can embed AWS compute and storage services, providing mobile edge computing infrastructure for developing, deploying, and scaling ultra-low-latency applications.
- **AWS Outposts** – in virtually any on-premises data center or colocation space, AWS Outposts creates a truly consistent hybrid experience. This fully managed service offers the same AWS infrastructure, AWS services, APIs, and tools and runs on AWS-provided hardware in your data center.
- **Amazon EKS Anywhere** – on your own infrastructure with the support of AWS, customers can create and operate Kubernetes clusters. Today, customers can run Amazon EKS Anywhere on **bare metal** servers or using VMware vSphere.

Benefits of Amazon EKS Anywhere

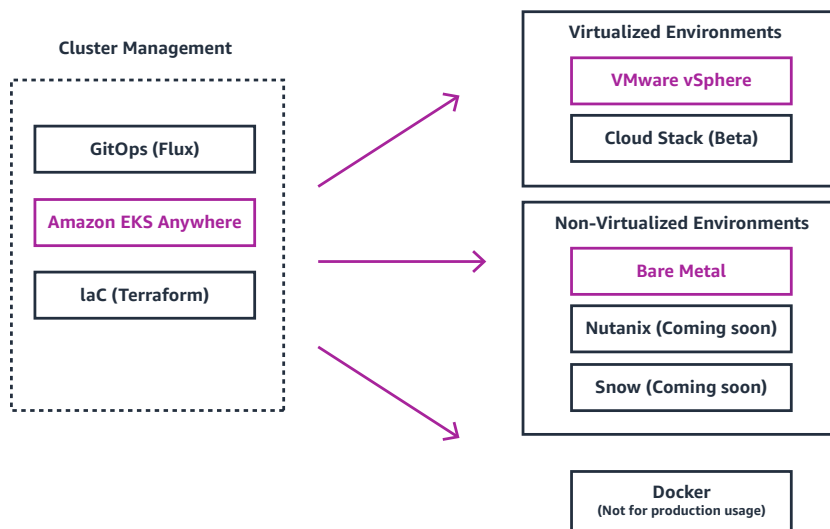
Amazon EKS Anywhere is open source: you don't need to be an AWS customer to use it or have existing contracts to get started. Instead, AWS offers enterprise support through a subscription. When compared to the other deployment options on AWS, Amazon EKS Anywhere is distinct, as it solves a combination of the three problems we outlined earlier:

- Capitalizing on existing investments
- Data locality requirements
- Latency-sensitive compute

Deployment options—Amazon EKS Anywhere gives you the freedom to bring your own compute.

There are a number of options for deploying Amazon EKS Anywhere, whether it is in virtualized environments using VMware vSphere or non-virtualized environments like bare metal. We are actively working to support even more environments, such as Nutanix and the AWS Snow Family of devices. You can even run Amazon EKS Anywhere on your local machine for non-production usage, giving teams an easy way to have a consistent experience when developing and testing applications.

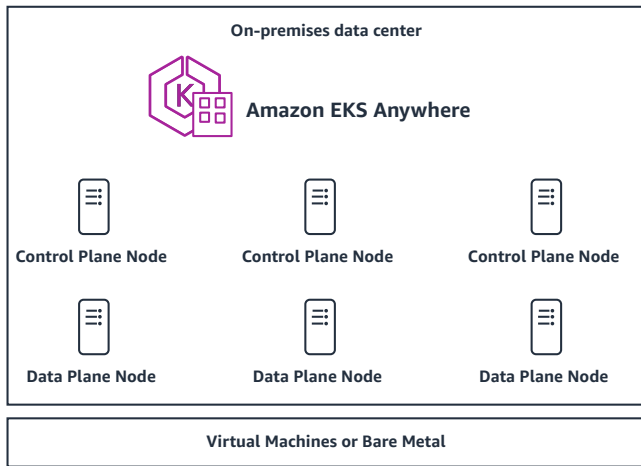
Figure 4



Flexible connectivity—Amazon EKS Anywhere allows you to connect back to Amazon EKS on AWS or run in disconnected environments.

Notably, Amazon EKS Anywhere builds on Amazon EKS Distro, allowing you to leverage your own infrastructure running on premises. You’ll have the ability to administer the control plane nodes and the data plane nodes where your workloads are running, giving you the flexibility in operating the complete Kubernetes cluster. A notable advantage of this is the ability to run workloads in a “disconnected” mode without needing to communicate back to AWS data centers.

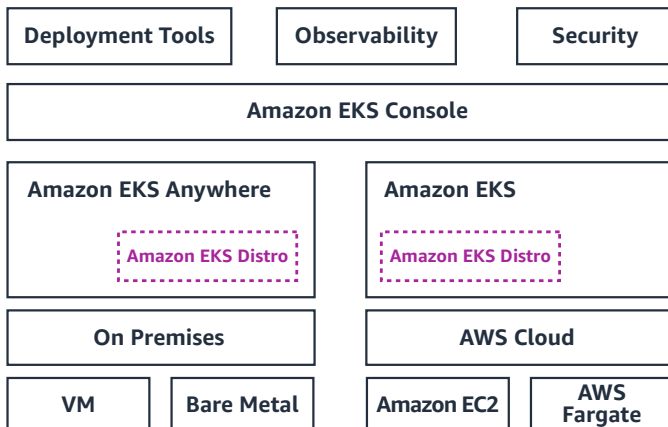
Figure 5



Consistent tooling and opinionated packaging—Avoid the paradox of choice with AWS-supported tooling.

With consistent tooling and opinionated packaging, you get the same operational experience across on-premises and cloud environments. This means you can use the same command line interface (CLI) tool, eksctl, to work with both Amazon EKS and Amazon EKS Anywhere clusters or choose to use other open-source solutions like Terraform and [cdk8s](#). In addition, you can use the [Amazon EKS Connector](#) to see your clusters visually represented in the Amazon EKS console.

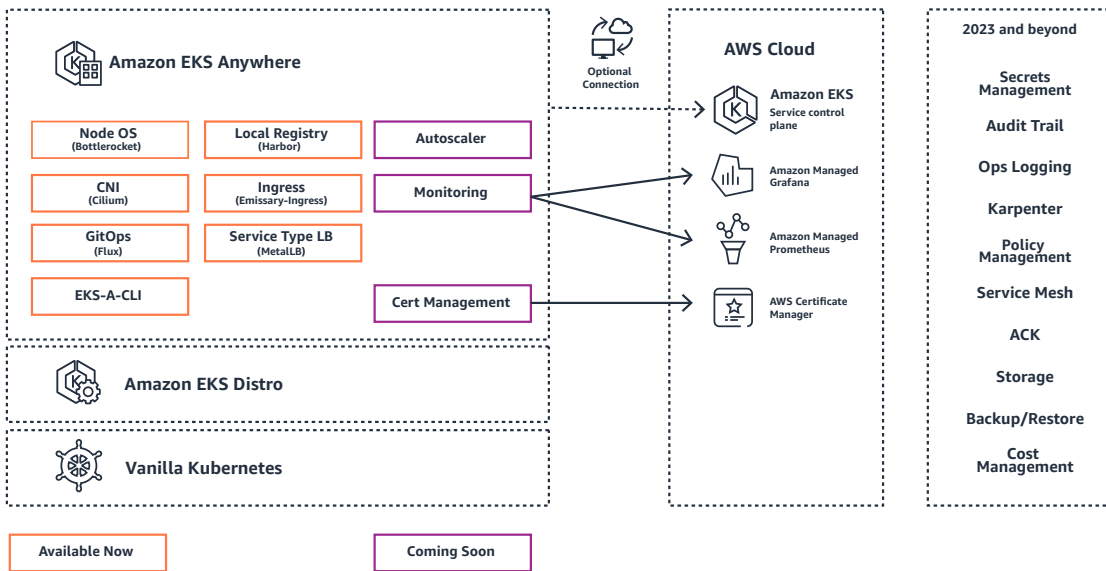
Figure 6



Curated Packages—Extend Kubernetes with capabilities from third-party vendors and open-source projects.

To help customers overcome decision fatigue when deciding on tooling (including open-source projects), Amazon EKS Anywhere introduced Curated Packages. These are Amazon-curated software packages that extend the core functionalities of Kubernetes on your Amazon EKS Anywhere clusters. Like Amazon EKS add-ons, Curated Packages are versioned and upgraded with your clusters for compatibility. Better yet, these are supported directly by AWS, leading us to the final advantage.

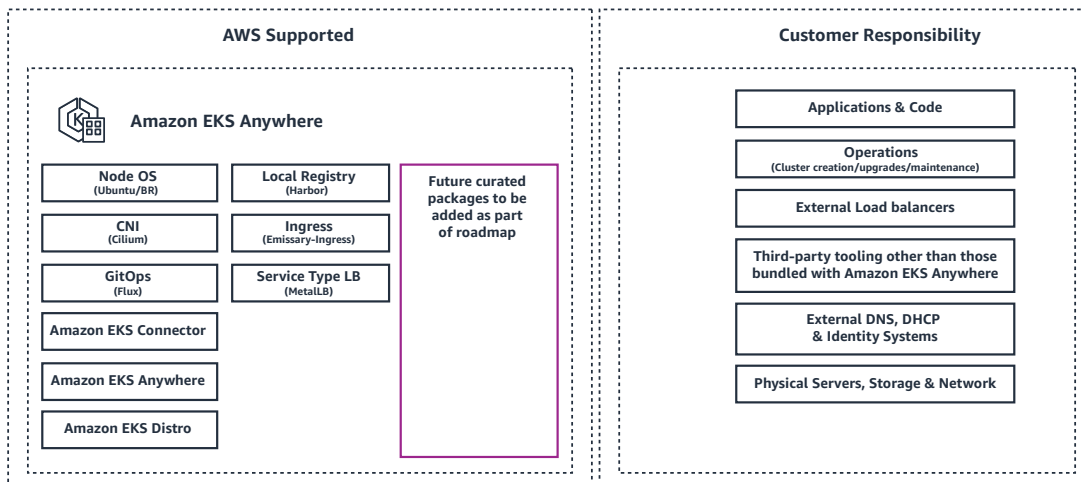
Figure 7



Shared Support—One stop for all your support needs.

For customers, a single point of contact for support is critical. With AWS supporting the complete stack from Amazon EKS Distro, tooling such as the CLI, and a number of curated packages, customers can rest easy knowing that they have a one-stop shop for getting the support their teams need.

Figure 8



CONCLUSION

Ready to get started?

As Kubernetes gains adoption, more and more applications will be built with containers and cloud-native technologies. Wherever you are in your cloud modernization journey, AWS can provide the appropriate services and support—and help you make the right decision for your business. Amazon EKS Anywhere is free, with **Enterprise Support** available through a subscription.

Learn more about Amazon EKS Anywhere or get started today:

[Get in touch with an EKS Anywhere specialist ›](#)

[Visit our documentation and get started for free ›](#)

Interested in learning more? Watch Amazon EKS, Amazon EKS Anywhere, and other container-focused content on **[Containers from the Couch](#)**.