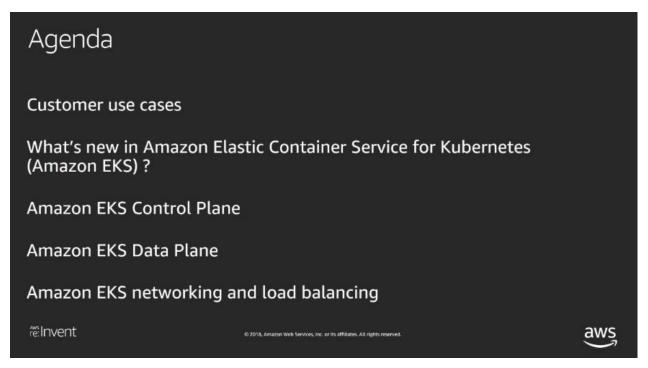


In this talk, we cover the configuration and networking details needed to run a production-ready Kubernetes cluster with **Amazon Elastic Container Service for Kubernetes (Amazon EKS)**. We walk through the new features and updates for Amazon EKS in 2018.



The EKS Control Plane deals with every that AWS runs for you while the EKS Data Plane deals with everything that you run in your account.

How are customers using Amazon EKS?

Microservices

PaaS

Enterprise app migrations

Machine learning

ML with TF based workloads using EKS

Which customers are using Amazon EKS?



"We built the next generation of our PaaS using EKS for large enterprise workloads. We manage thousands of applications and have hundreds of DevOps teams."

Which customers are using Amazon EKS?



"Snapchat serves millions of people around the world every day, and we're thrilled to now leverage Amazon EKS as a core compute service that can meet our needs now, as well as upcoming plans to host several critical workloads in the coming months."

Which customers are using Amazon EKS?



"Kubernetes is fast becoming the preferred solution for container orchestration. Its biggest downside is that it is not simple to set up and operate. EKS gives us all the benefits of Kubernetes, but takes care of managing the hard stuff. We can dedicate less resources to deployment and operations as result."

Which customers are using Amazon EKS?



"The performance from Amazon EKS makes it feasible to effectively manage large-scale databases delivering over a million queries per second. EKS also helps with our cluster management and scalability challenges."

What's new?

April: Amazon EKS achieved K8s conformance

June: Amazon EKS is HIPAA-eligible

July: Amazon EKS AMI build scripts available in GitHub

August: New EKS-optimized AMI and updated AWS CloudFormation template for provisioning

worker nodes

August: Amazon EKS supports GPU-enabled Amazon Elastic Compute Cloud (Amazon EC2)

instances

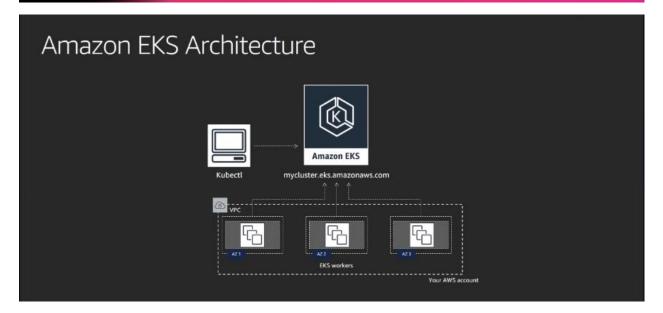
August: Amazon EKS platform version 2 launched August: Amazon EKS supports HPA with custom metrics September: Amazon EKS launches in Dublin, Ireland

September: Amazon EKS simplifies cluster setup with update-kubeconfig CLI command

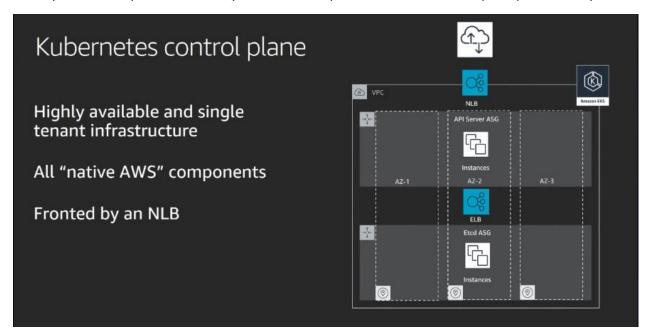
October: Amazon EKS adds support for Dynamic Admission Controllers (Istio)

November: Amazon EKS launches in Ohio

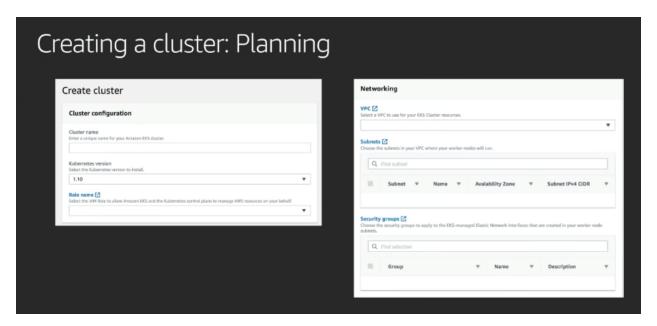
Amazon EKS Control Plane

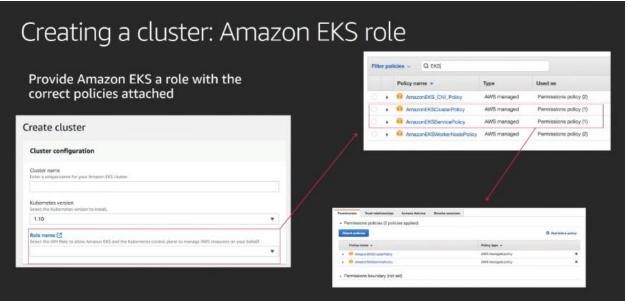


We handle all the control plane functions like etcd functions, losing nodes, connectivity issues between AZs, etc and provide you with a cluster endpoint dedicated to your cluster via Route53. You need to focus on your worker nodes that run in your VPCs. If you use Kubectl, you will need to point it to the cluster endpoint provided to you.



This is the infrastructure of the K8s control plane that AWS runs for you.

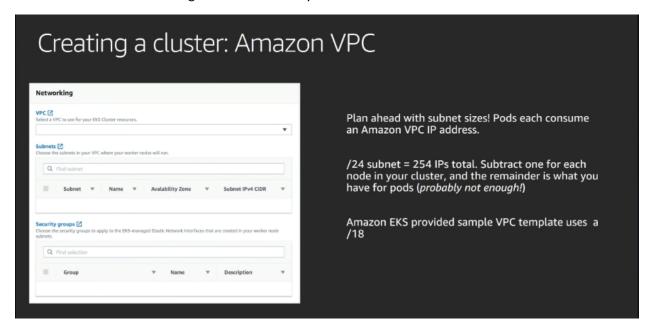




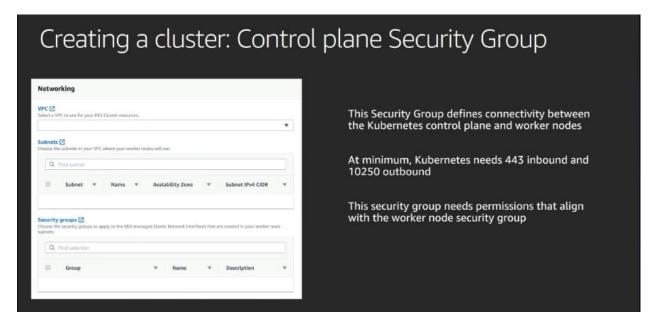
You need to provide a role so that AWS can manage resources on your behalf, the role will have the needed policies to allow AWS manage the resources for you. In IAM, there are 4 different policies available and 2 of the policies are applicable to the EKS control plane. You need to takes these 2 policies **AmazonEKSClusterPolicy** and **AmazonEKSServicePolicy** and associate them with a role, then give EKS the role so that it can create the control plane for your cluster. The **AmazonEKSServicePolicy** is going to be used for managing network infrastructure in your account for things like the network interfaces that will be created. The **AmazonEKSClusterPolicy** contains all the permissions that K8s needs to do things in your account, like creating LBs, EBS volumes, modifying tags, etc.



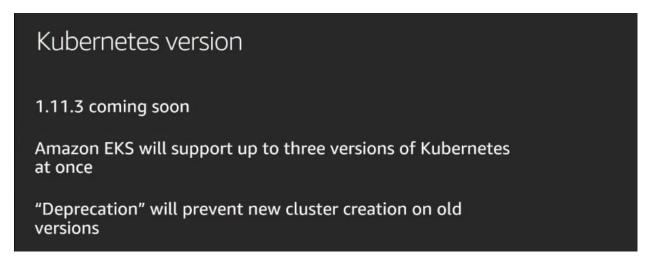
There are also VPC requirements, this is for the VPC that your data plane is going to be running inside and it includes the LBs and worker nodes. You need to give EKS all the private and/or public subnets that you are going to use within the VPC and then you might then decide to run your worker nodes in the private subnets and run the ELBs in the public subnets. This allows k8s to tag those subnets for you as above.



You have to plan your subnet sizes ahead of time and use a slightly bigger subnet size, this is because EKS allocates a lot of pods for you and it normally allocates an IP per pod created. The /24 subnet is a little too small with 254 possible pods, so we use a default /18 subnet that gives you more possible IPs with a larger CIDR block size.



K8s needs the port 443 inbound to the control plane and the control plane needs the port 10250 outbound, the 10250 is the port that the kubelet listens on and it is what the control plane uses to send instructions to the kubelet running in your worker nodes. The 443 outbound on your worker nodes so that you can reach the HTTP endpoint that provides the k8s API. There are 2 security groups that you need to note, the control plane SG and the worker node SG, both of these SGs have to have permissions that are aligned with no exclusive denies.



You also need to provide the k8s version you want to run in your cluster, EKS will run the last 3 K8s versions and deprecate older ones.

Amazon EKS platform version

Platform version revisions represent API server configuration changes or Kubernetes patches

Platform versions increment within a Kubernetes version only

K8s 1.10 K8s 1.11 K8s 1.12 eks.1 eks.1 eks.1

eks.2 eks.2

eks.3

API server configuration

Kubernetes version	Kubernetes patch version	Amazon EKS platform version	Enabled admission controllers	Release notes
1.10	1.10.3	eks.2	Initializers, NamespaceLifecycle, LimitRanger, ServiceAccount, Def aultStorageClass, ResourceQuot a, DefaultTolerationSeconds, Nod eRestriction, MutatingAdmission Webhook, ValidatingAdmissionWebhook	*Added support for Kubernetes <u>Horizontal Pod</u> <u>Autoscaler(HPA).</u>

You can look at the documentation to see what changes are available in the new EKS K8s version changes

EKS Kubernetes version updates – coming soon

New UpdateClusterVersion API – supports in-place updates of Kubernetes version

Introduces an update EKS API object

ListUpdates and DescribeUpdate APIs to provide visibility into the status of a given update

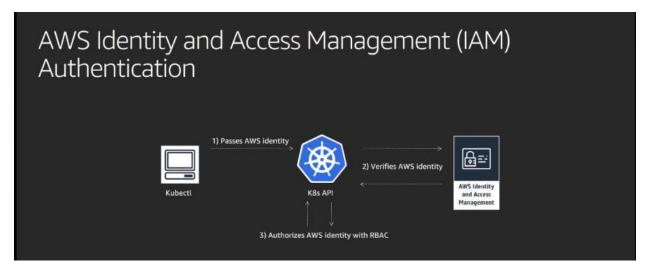
There are some APIs to help you listed above.



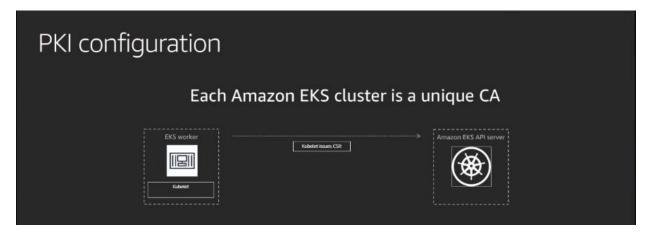
We have integrated IAM with the K8s RBAC, we split the responsibilities of authentication and authorization.



In Kubectl, we have an external authentication binary called the **AWS Authenticator** that allows kubectl to pass your AWS identity to the K8s API Server. This means that kubectl can look in the exact credential chain as the AWS CLI like looking at your session token, your environment variables, your secret access keys or active roles. It can package this information up and send it to the k8s API server.



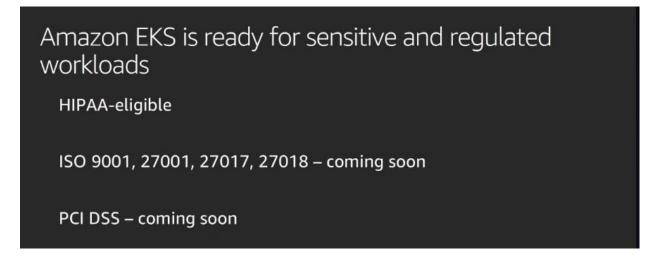
We then have a webhook on the K8s API server that verifies/authenticates the identity as a valid user and has the authorized permissions to make the specific kubectl call, it will then allow or deny the action.



Each EKS cluster is also a unique CA, and AWS manages this for you since the PKIs in K8s is complicated to set up by you. We use the built-in K8s CA system and generate a set of certificates for the control plane.



The kubelet can also get a certificate by issuing a CRS to the EKS API Server that will generate both public and private keys. Then the kubelet knows how to install the generated certificates. The kubelet's are also set up to do certificate rotation automatically.

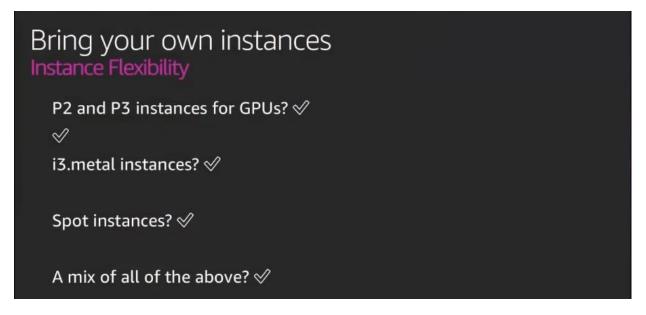


Wrapping it all up Easily export control plane config \$aws eks update-kubeconfig --name my-cluster CLI Helper command to create kubeconfig file Creates a new context for each cluster in the config

This helper command will dump all the config information you care about from the control plane like the certificate data, cluster name into a kubeconfig file that you can start to use.

Amazon EKS Data plane

This is the stuff that runs on your account.



You have some things to stand up on your end but we give you an AMI and a CloudFormation (CF) template that you can use. You are not limited on the instance types that you can use and launch for your worker nodes group.

Bring your own OS

EKS AMI build scripts

https://github.com/awslabs/amazon-eks-ami

Source of truth for Amazon EKS Optimized AMI

Easily build your own Amazon EKS AMI

Build assets for Amazon EKS AMI for each supported Kubernetes version



AI/ML with Amazon EKS

Amazon EKS-optimized AMI with GPU support

Easily run Tensorflow on Amazon EKS



Includes NVIDIA packages to support Amazon P2 and P3 instances

Available on AWS Marketplace

Worker node setup – Bootstrapping

/etc/eks/bootstrap.sh [options] <cluster-name>

Supports arbitrary kubelet bootstrap args and defining `-- max-pods` for kubelet

You have a couple of options for bootstrapping your worker nodes, AWS provides a CF template to use. You can also use the **bootstrap.sh** script in the GitHub repo. This allows you to tell the kubelet what is the cluster name that it should be joining, that instance will then join that cluster as a worker node. It also allows you to pass flags to kubelet to modify its behavior at boot time, this could be for defining max-pods, etc.

Worker node setup — Bootstrapping Amazon-eks-nodegroup.yaml AWS Cloudformation template supports `BootstrapArguments` parameter Passes parameter directly to bootstrap.sh: /etc/eks/bootstrap.sh \${clusterName} \${BootstrapArguments} Result? Modify Kubelet config directly through CFN parameter: --kubelet-extra-args --node-labels=mykey=myvalue, nodegroup=NodeGroup1

This is also extensible to our CF template using the parameter called **BootstrapArguments** that allow you to feed these arguments directly to the **bootstrap.sh** script. This allows you to modify the cluster configuration at scale across all your instances using the flags via CF.

```
Worker node setup - Authentication

apiversion: v1
kind: ConfigMap
metadata:
    name: aws-auth
    namespace: kube-system

data:

mapRoles: |
    - rolearn: <ARN of instance role>
    username: system:node:{{Ec2PrivateDNSName}}
    groups:
    - system:bootstrappers
    - system:nodes

**Elnvent**

0.2018, Arrazoo Web Service, Inc. or its uffiliates Al Tophn reserved.

aws.
```

An important part of this is the Authentication component, all of the bootstrapping of the instance that you do requires that you tell the cluster to let the particular IAM policy join the cluster as a node. This works by using a ConfigMap on the AWS cluster and we have a group for the node, you basically tell it the ARN for the node that should be allowed to join that cluster.

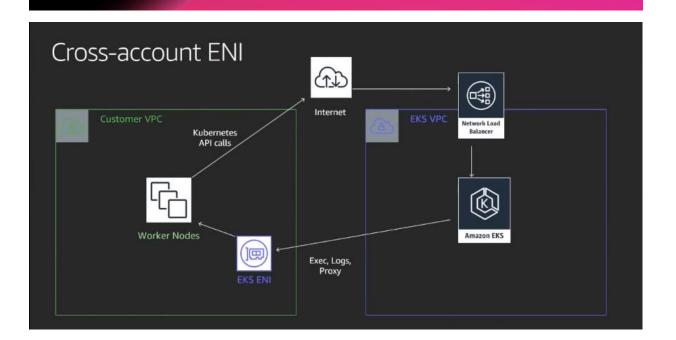
Updating worker nodes

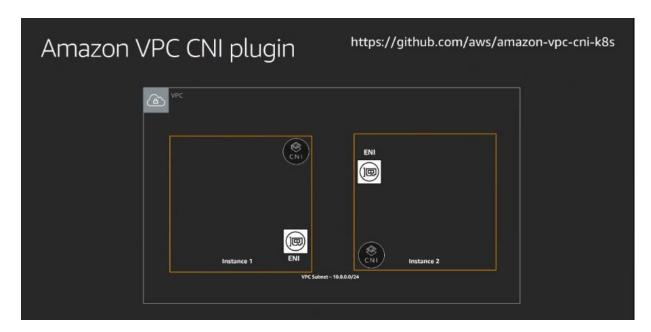
Two options:

- 1) Create new node group with latest Amazon EKS AMI; drain old nodes; terminate old CFN template
- 2) Simply update AMI in CFN template; "rolling" replacement policy terminates nodes

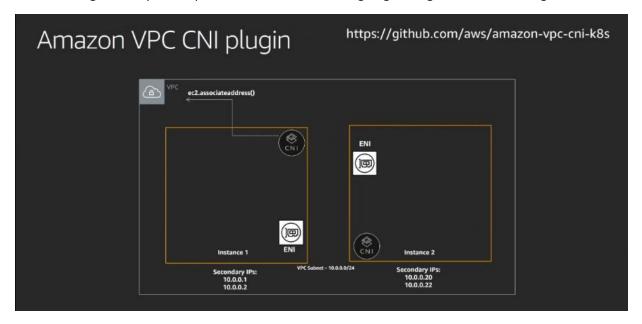
(Downsides: un-graceful termination of applications)

Amazon EKS networking

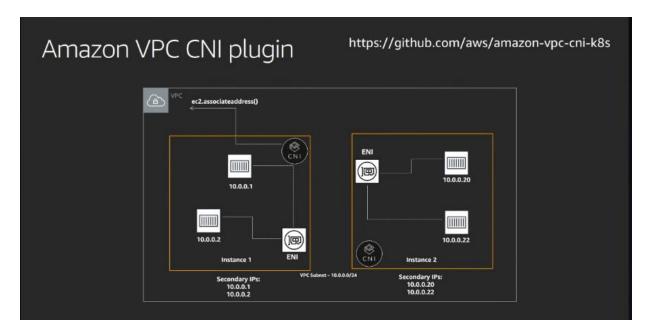




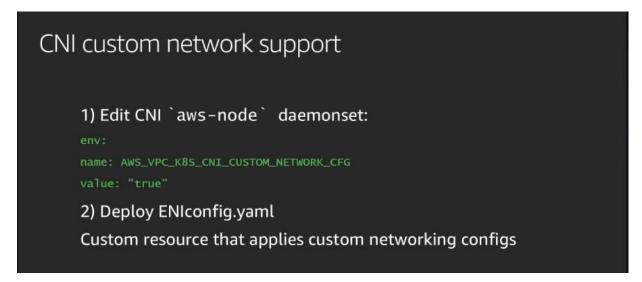
This is what runs in your account and VPC K8s pods. We also have and use a CNI plugin which is an OSS project for allowing your applications communicate with each other and to external resources. This means that every pod in your K8s cluster gets a native VPC IP address. Using the CNI allows you to go directly to your pods for debugging purposes, LBs can also go directly to the pod's IP address instead of going through a NodePort using an additional hop, etc.



The CNI plugin attaches an ENI on-demand to the worker nodes where the pods run, it then uses secondary IP addresses by calling **ec2.associateaddress()** to the instances. This approach allows us to get more pod density per worker node instance. Different EC2 instance types can support varying number of ENIs per instance, bigger ENIs can support more secondary addresses. The T2 instance has 2 ENIs and 6-8 secondary IP addresses.



As pods are scheduled to your worker nodes, the CNI wires up the network interfaces for you. The secondary IP addresses are going to attaches to an ENI and CNI plugin knows that when a pod is scheduled, it should take one of those IP addresses and wire it into the pod. You need to size your VPC appropriately.



You can also expand the IP address ranges to include secondary ranges to be used by pods in your worker nodes.

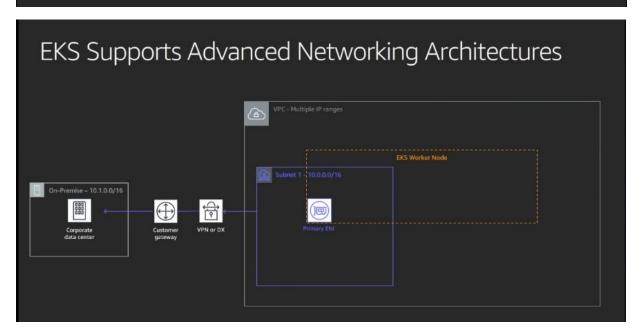


CNI custom network support

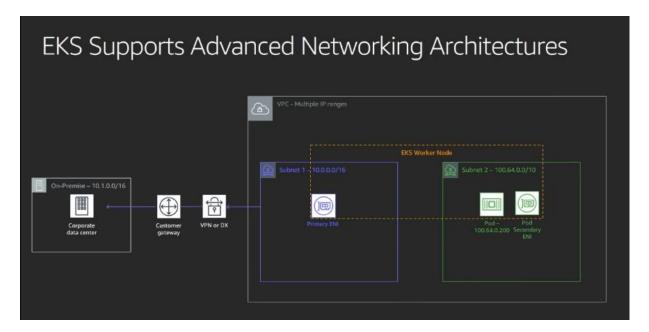
4) Annotate nodes to use custom config:

`kubectl annotate node <nodename>.<region>.compute.internal
k8s.amazonaws.com/eniConfig=group1-pod-netconfig`

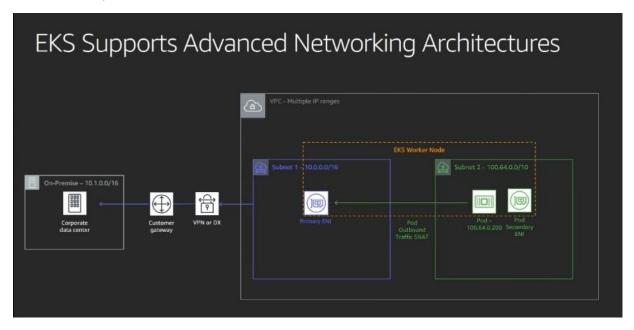
EKS Supports Advanced Networking Architectures VPC - Multiple IP ranges Corporate Curstomer VPN or DX Corporate Curstomer VPN or DX



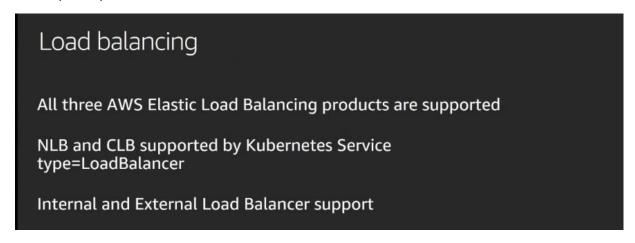
The primary ENI always remains on the primary subnet in your VPC, this ensures that pods running in your worker nodes have communication access back to your on-premises workloads.



When you have a secondary subnet in your VPC, the worker node will see this and add a secondary ENI to that subnet and also add a pod to it.



We then use source-NATting to allow pods on the secondary subnet to communicate with the on-premises workloads via the primary ENI.

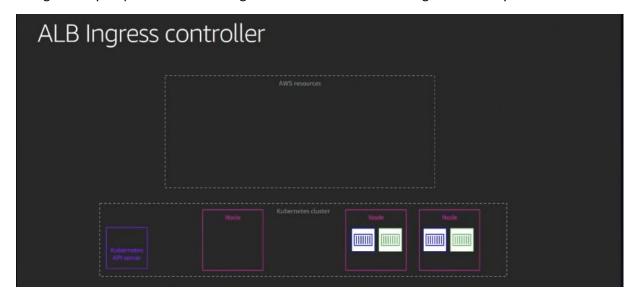


Load balancing Want to use an Internal Load Balancer? Use annotation: service.beta.kubernetes.io/aws-load-balancerinternal: 0.0.0.0/0 Want to use an NLB? Use annotation: service.beta.kubernetes.io/aws-load-balancertype: nlb

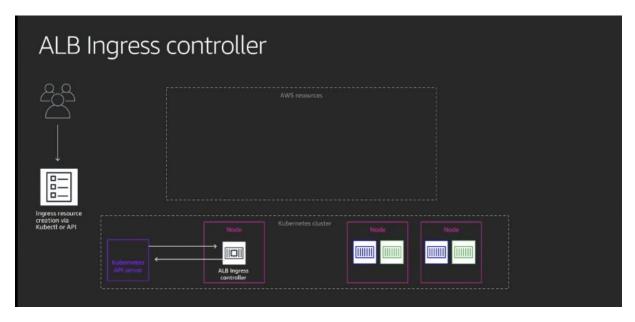
You need to have a private subnet for the internal LB when you tell K8s to create the LB using the command above. Do the same for creating the NLB as above.



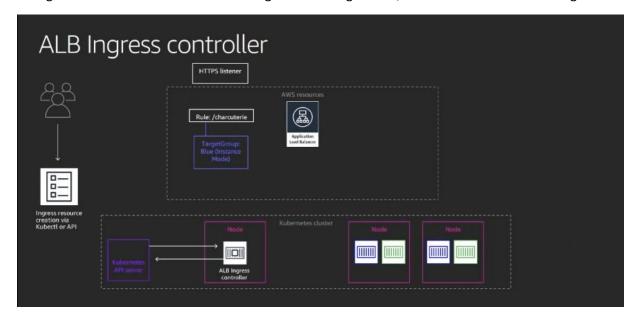
This gracefully maps the ALB to the ingress resource or API in K8s. Ingress is the layer-7 abstraction for K8s.



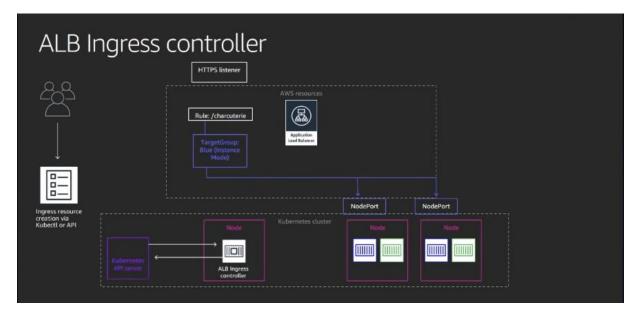
The ALB Ingress controller enables you to expose applications running inside your K8s cluster as an external HTTP or HTTPS service, it relies on the ALB to provide all the context-based routing.



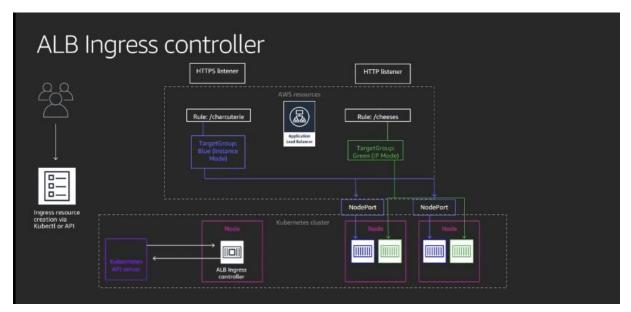
To use the ALB Ingress controller, you would first deploy it as a deployment on one of your worker nodes. This is just a simple app that watches the API Server and looks for notifications from the API Server that says 'someone has created an ingress resource'. This will include things like wanting an ALB, what routes should be configured in the ALB, what SGs



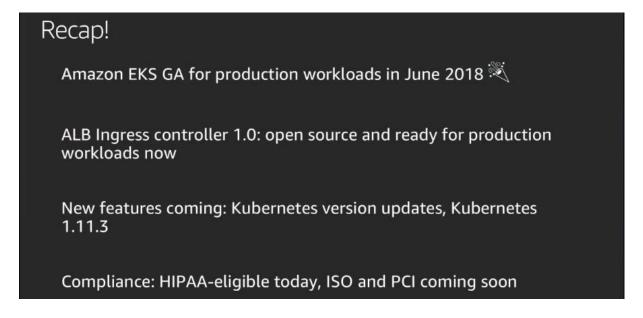
The ingress controller then starts to create the ALB for you and all other things like the HTTPS Listener and a Rule to what TargetGroup.



Traffic is actually routed to the NodePorts on your worker nodes that allows the traffic to get to your pods



You can also have another HTTP Listener, it uses IP Mode that points directly to the running pods IP addresses



Global availability

Launch Regions: Virginia, Oregon

Recent Regions: Ireland, Ohio

Coming soon: Singapore, Sydney, Seoul, Tokyo, Frankfurt

Demo: EKS updates

We have a couple of pods running on a pineapple cluster

