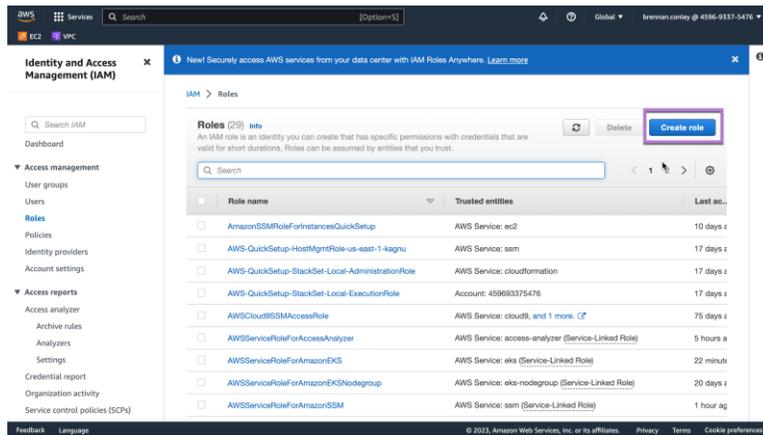


Implementing WEKA on Amazon EKS

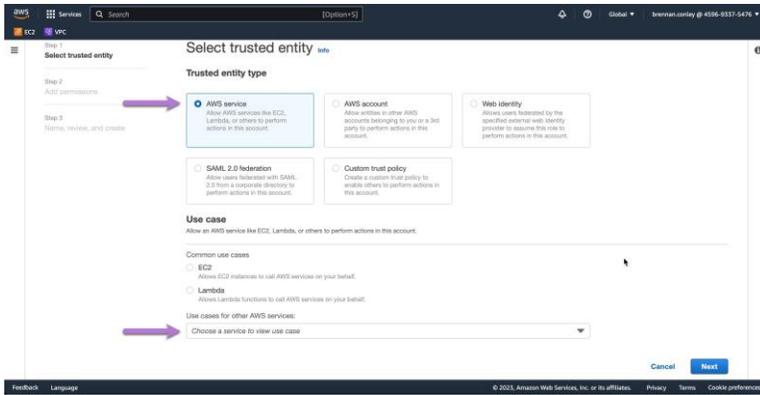
Establishing Pre-Requisites

1. Ensure an AWS account with adequate quota is being utilized.
2. Ensure AWS IAM account being used has sufficient roles and policies allocated to create additional IAM roles and policies for the purpose of EKS deployment.
3. Sign into the AWS console and visit <https://aws.amazon.com/iam/>.
4. Create a new IAM role allowing EC2 resource management and creation by EKS. This role can be referred to as the "Cluster role." The "Cluster role" will later be used as the EKS *****Cluster service role***** when creating the EKS cluster.

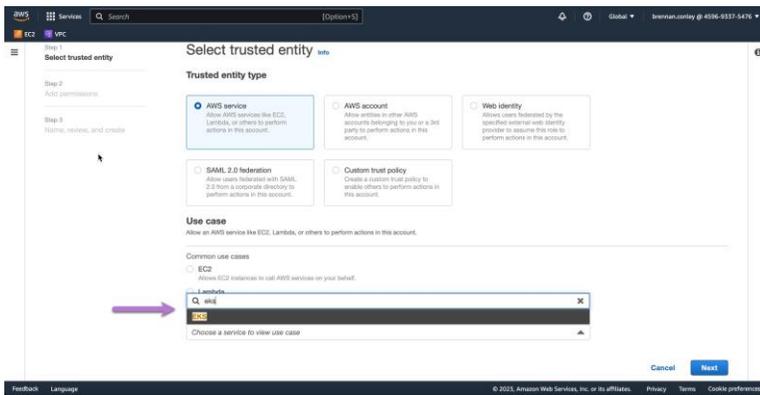
Click "Create role."



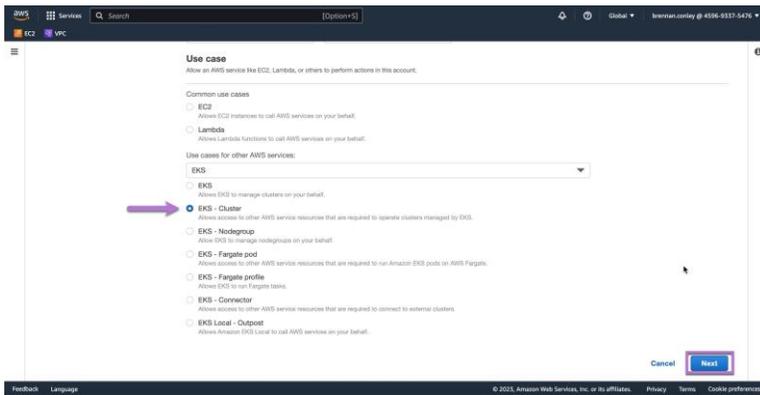
Select the "AWS service" trusted entity type. Then, utilize the indicated drop down.



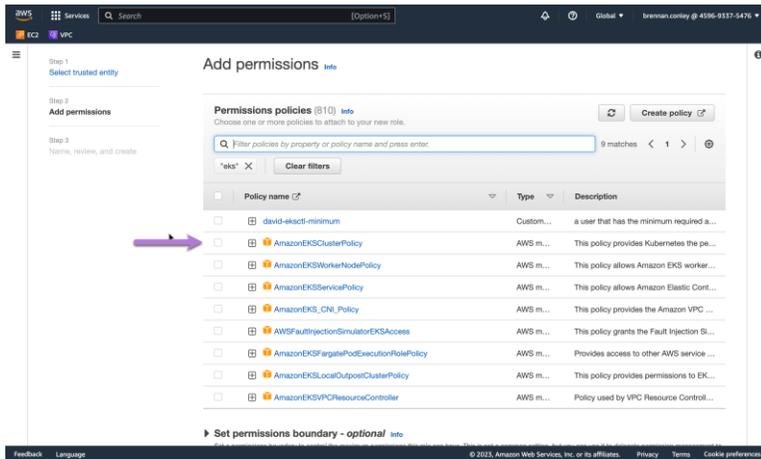
Search for and select "EKS." Click "Next."



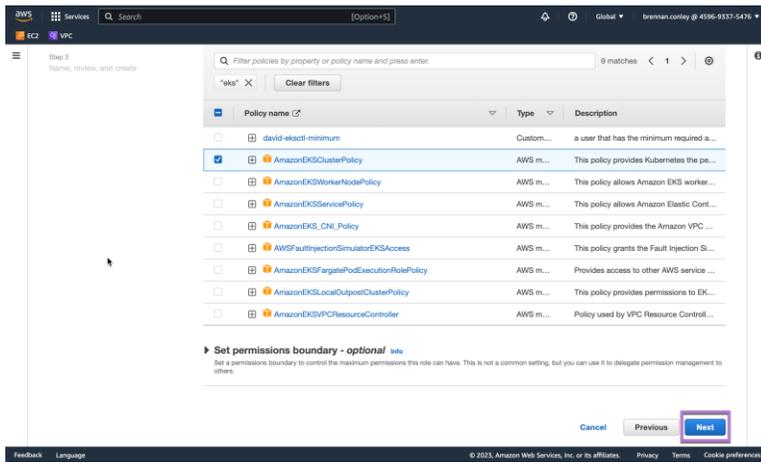
From the resulting EKS options, select "EKS - Cluster" and click "Next."



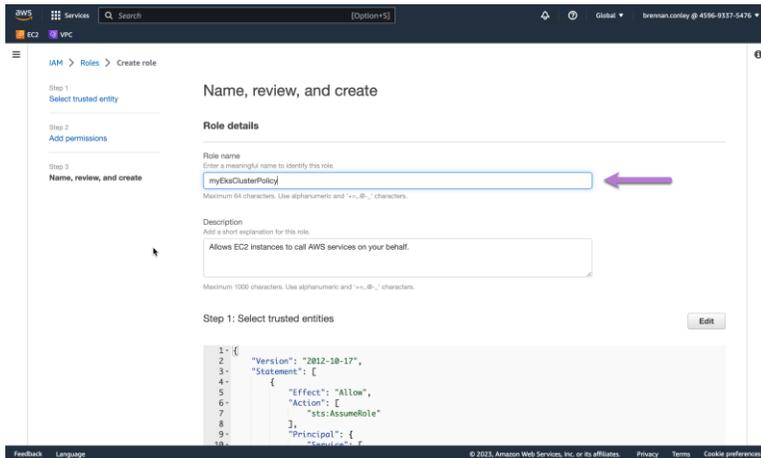
In the policies search bar, search for "eks." From the resulting policies, select "AmazonEKSClusterPolicy."



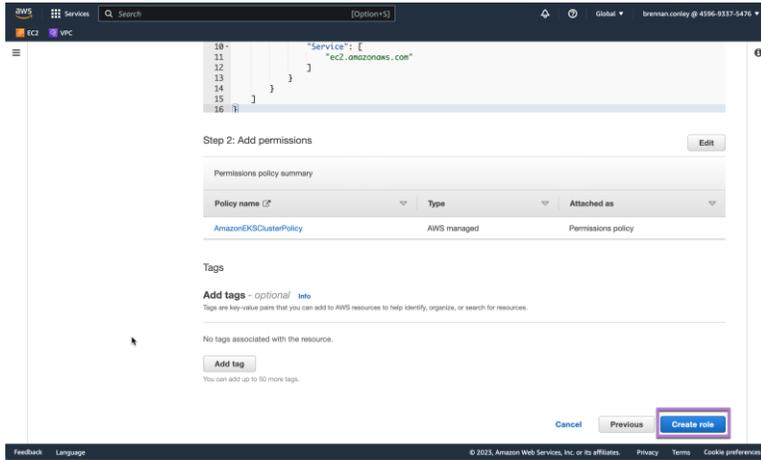
Once the policy is selected, click "Next."



Provide a name for the role



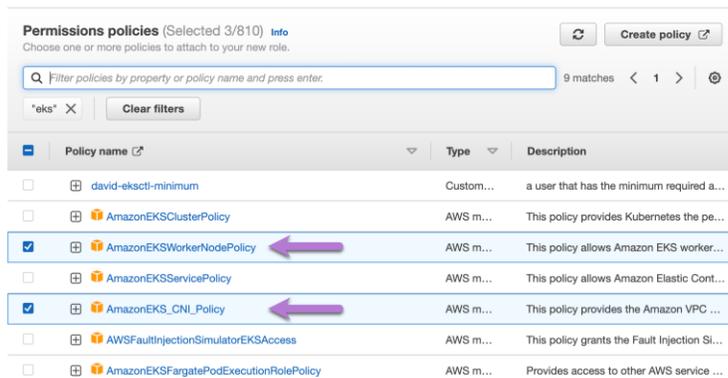
Click "Create role."



5. Following the same process as above, create another IAM role with *trusted entity type* as "AWS service" and the *use case* as "EC2." This role can be referred to as the "Worker node policy." This time, select the following policies:
 - a. "AmazonEKSWorkerNodePolicy"
 - b. "AmazonEC2ContainerRegistryReadOnly"
 - c. AmazonEKS_CNI_Policy

Create the role. This role will later be used as the EKS node group *Node IAM role*.

Add permissions [Info](#)



Add permissions [Info](#)

Permissions policies (Selected 3/810) [Info](#)
Choose one or more policies to attach to your new role.

Filter policies by property or policy name and press enter. 6 matches < 1 > [@](#)

containerregistry [X](#) [Clear filters](#)

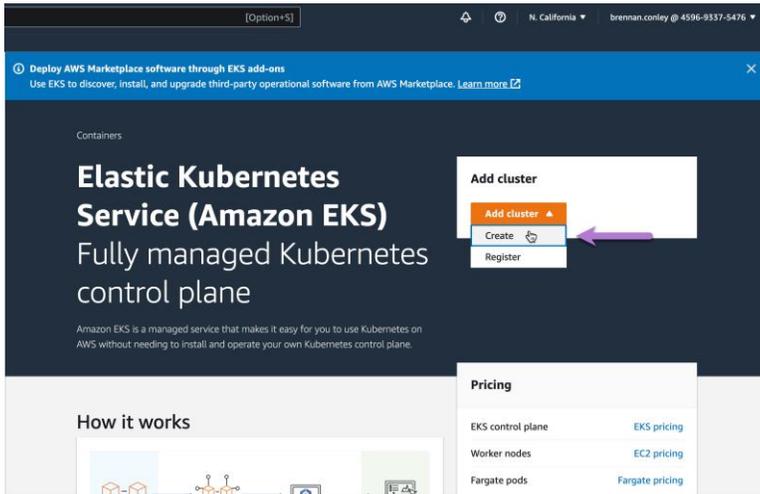
<input type="checkbox"/>	Policy name ↗	Type	Description
<input type="checkbox"/>	AmazonEC2ContainerRegistryFullAccess	AWS m...	Provides administrative access to Ama...
<input checked="" type="checkbox"/>	AmazonEC2ContainerRegistryReadOnly	AWS m...	Provides read-only access to Amazon ...
<input type="checkbox"/>	AmazonEC2ContainerRegistryPowerUser	AWS m...	Provides full access to Amazon EC2 C...
<input type="checkbox"/>	AmazonElasticContainerRegistryPublicReadOnly	AWS m...	Provides read-only access to Amazon ...
<input type="checkbox"/>	AmazonElasticContainerRegistryPublicFullAccess	AWS m...	Provides administrative access to Ama...
<input type="checkbox"/>	AmazonElasticContainerRegistryPublicPowerUser	AWS m...	Provides full access to Amazon ECR P...

These two roles will be used later in the EKS deployment process.

6. Ensure that a VPC has been created and is available for use by EKS.
7. Ensure at least two subnets have been created inside the VPC intended for use with EKS. These subnets should have public capabilities i.e. an internet gateway for public network access.
8. Ensure that a security group has been created for use with EKS.
9. Ensure that an AWS public / private key pair has been created, and that the private key has been downloaded and is accessible for use.
10. Either provision a linux jump box in the same VPC containing the Weka cluster and EKS worker nodes, or utilize a laptop configured for SSH communication with machines in the VPC for kubectl commands. Instructions for installing kubectl on various platforms can be found [here](#).
11. In advance, deploy a functional Weka cluster via start.weka.io. Make sure that the management interface is accessible. For the sake of simplicity, deploy the Weka cluster in the same VPC designated for EKS use. If deploying the Weka cluster on a subnet separate from EKS worker nodes, ensure proper subnet routing and security group configuration to allow access between the Weka cluster and EKS worker nodes.

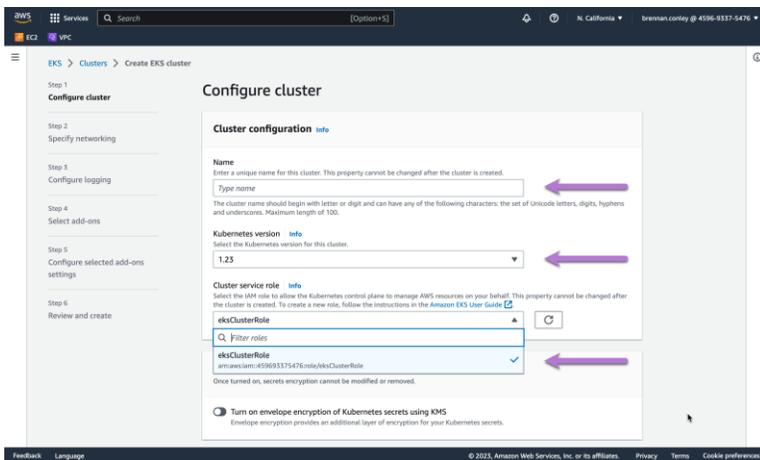
Deploying the EKS cluster

1. To begin, visit aws.amazon.com/eks. Login to the AWS console, and add a cluster.

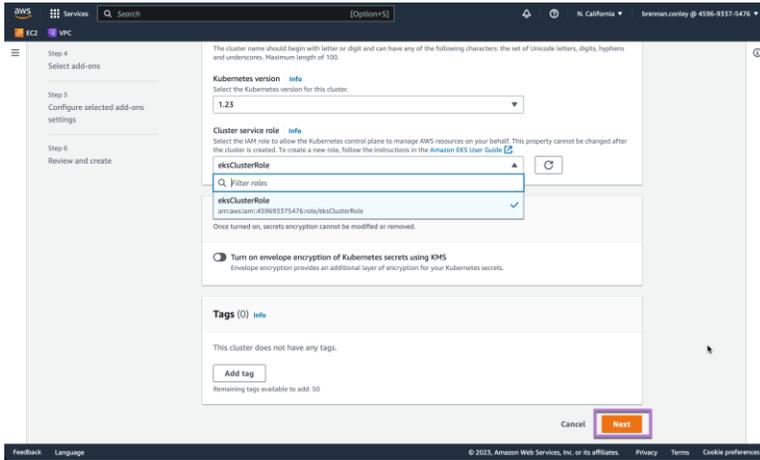


2. Configure the cluster.

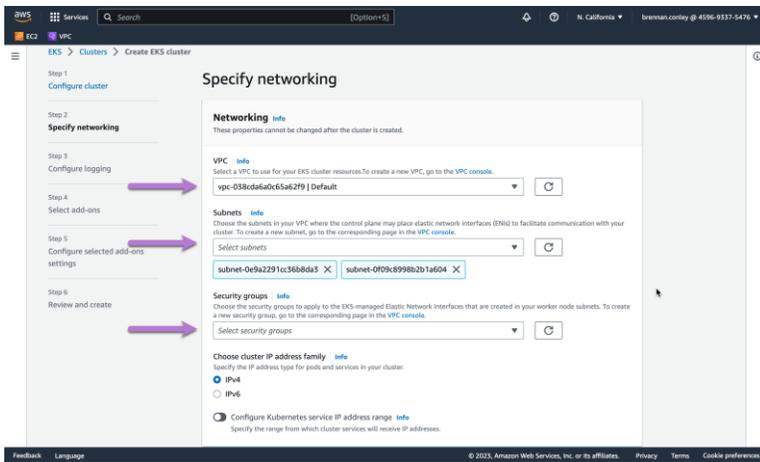
Enter an EKS cluster name. Select Kubernetes version 1.23. The default version of 1.24 currently (at the time of writing) has a bug that does not allow kubectl to properly communicate with the EKS cluster management interface. The actual issue is with kubectl, Kubernetes itself. For reference, the version of kubectl used must be -1, equal to, or +1 the version of the Kubernetes cluster. For *Cluster service role* select the "Cluster role" IAM role created in step 4 of the prerequisites section.



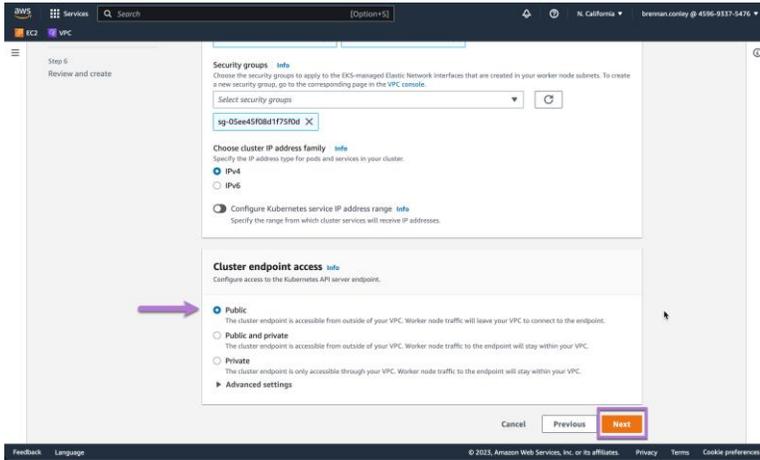
Leave all other options as default and click "Next."



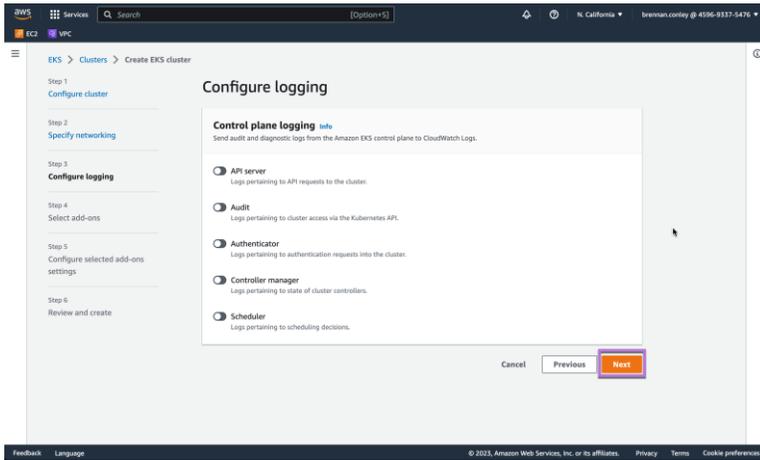
Continue cluster configuration by selecting the VPC, subnets, and security group created in the prerequisites section. EKS requires at least two subnets to be selected for load-balancing and redundancy purposes. Use the default of IPv4. Leave the Kubernetes service IP address range as default.



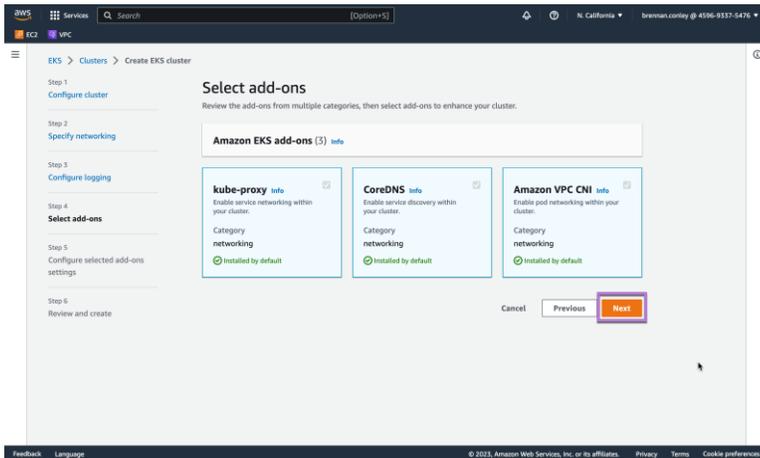
Under *Cluster endpoint access*, select the default of "Public." Click "Next."



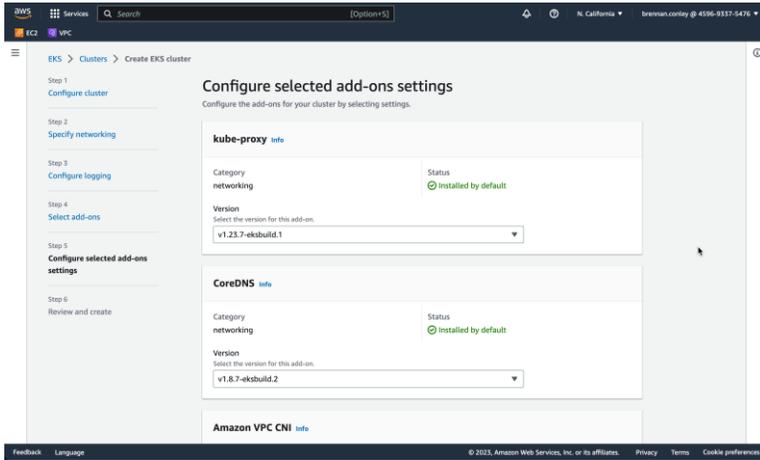
Configure logging as necessary, then click "Next."



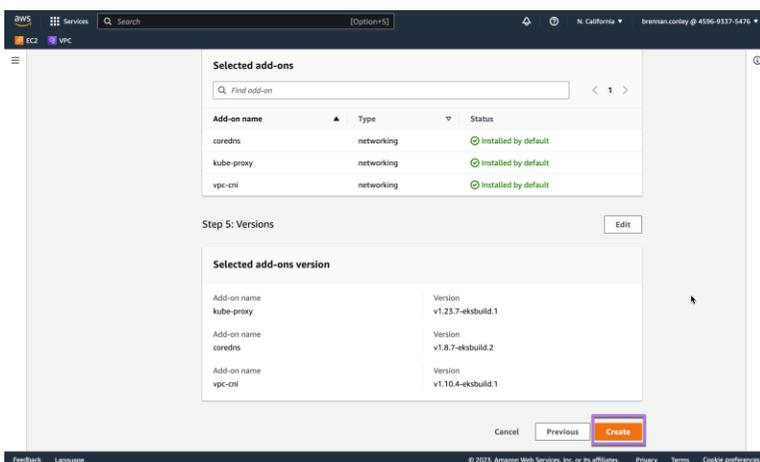
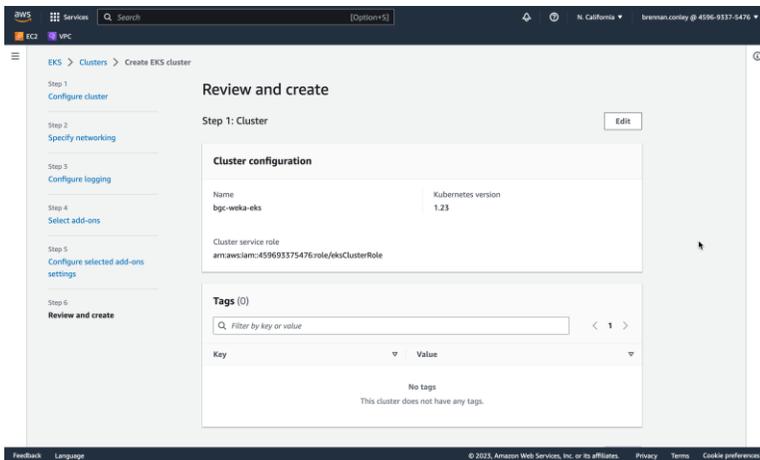
Under *Select add-ons* note the defaults and click "Next."



Under *Configure selected add-ons settings* note the defaults and click "Next."



Finally, under *Review and create* review all configured options and click "Create." The EKS cluster management framework will be created and necessary services started. No EKS worker nodes or node groups will be created at this time. Cluster creation will take approximately 10 minutes.

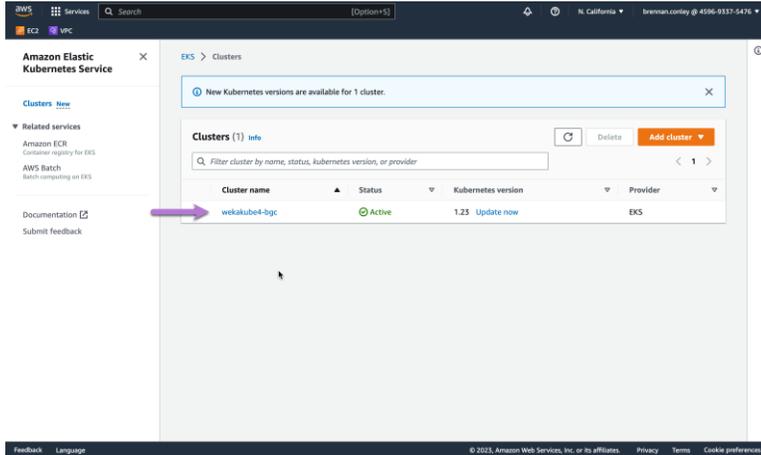


Creating node groups - non-autoscaling

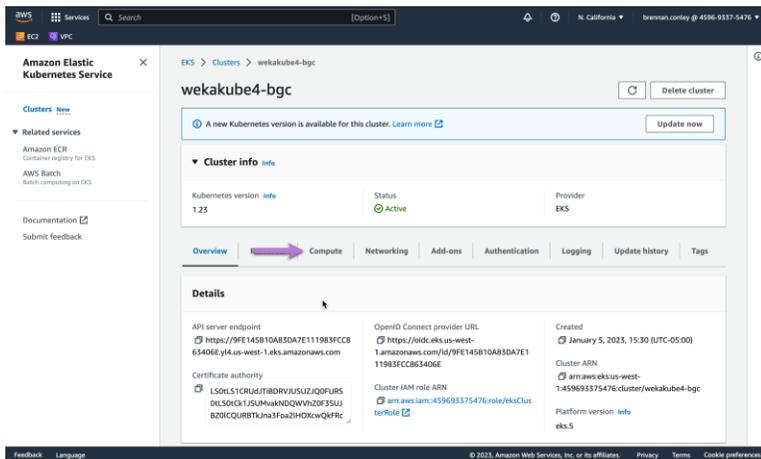
After the cluster has been created, a node group must be provisioned in order to deploy pods. Pods will contain worker nodes allocated from a worker node group. The following instructions will outline how to create a worker node group.

1. Navigate to EKS compute resources

Under *EKS, Clusters*, locate and select the previously created cluster.

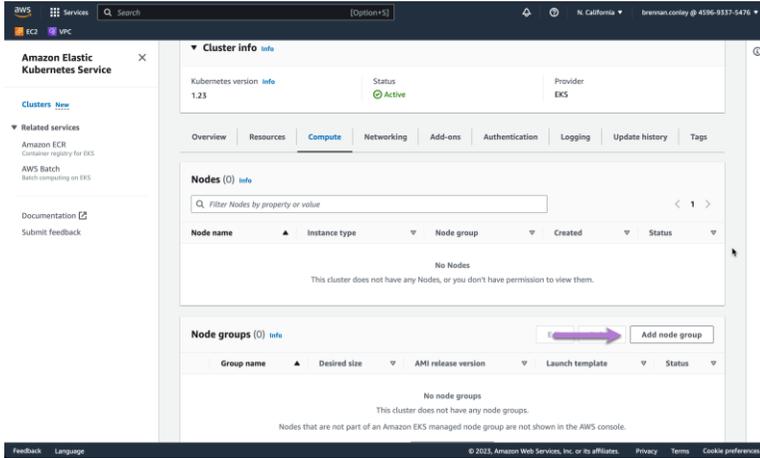


Select "Compute."

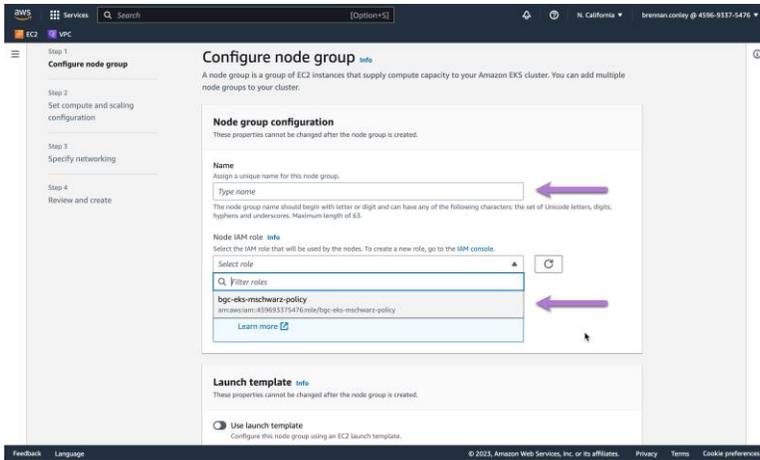


2. Create the node group

Scroll down and select "Add node group."



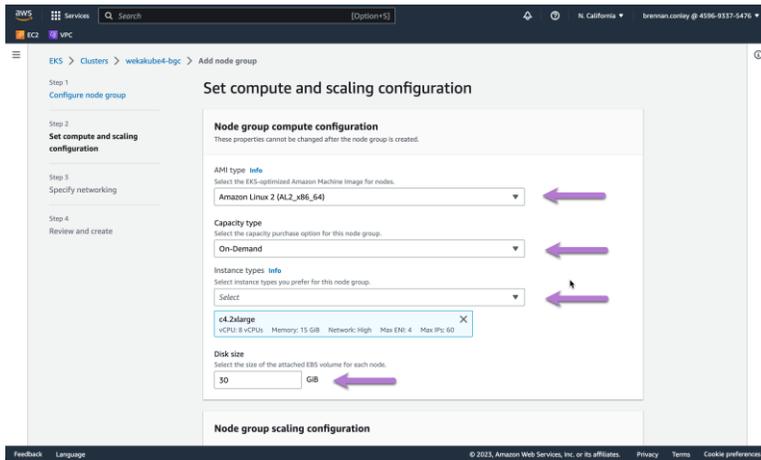
Under *Configure the node group* provide a name for the node group. For *Node IAM role* select the second IAM role created in the prerequisites section as the "Worker node policy." Leave all other options as default. Click "Next."



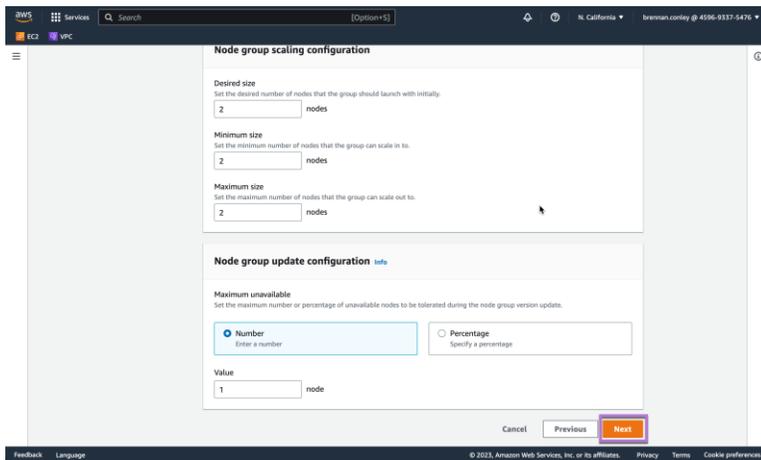
Under *Set compute and scaling information* configure desired compute and scaling parameters for the node group worker nodes.

For the purposes of this guide, select **Amazon Linux 2 (AL2_x86_64)** for the *AMI type*. For *capacity type* select *On-Demand*.

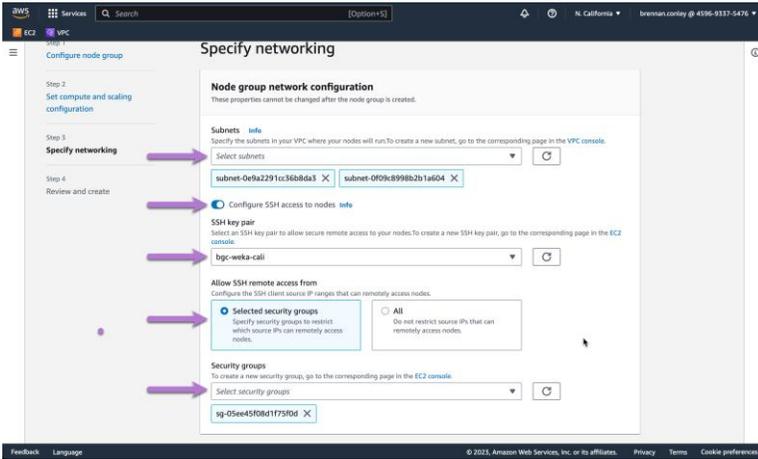
Select the **c4.2xlarge** (or otherwise desired instance size) under *Instance types*. Allocate **30GiB** for the *Disk size*.



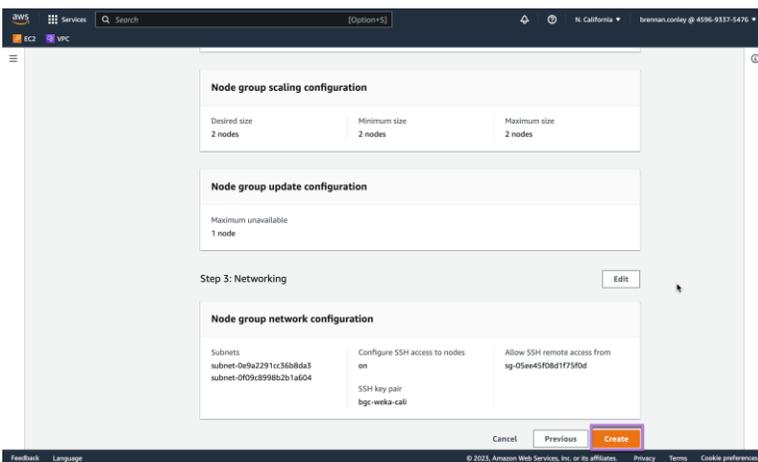
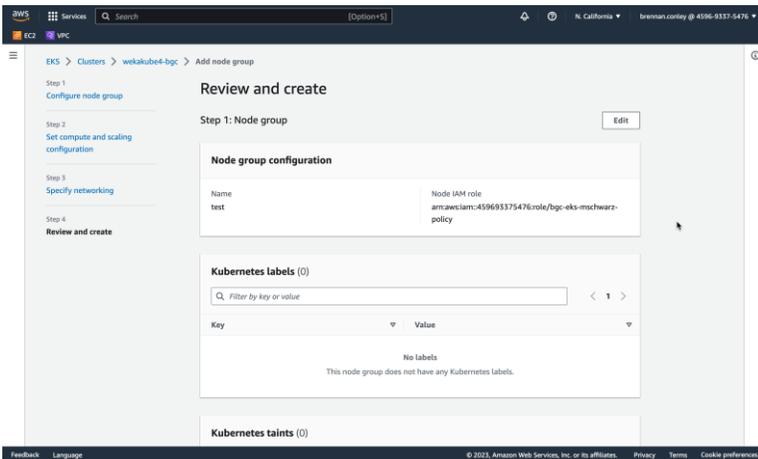
Select desired values for *Node group scaling configuration* and *Node group update configuration*. It's fine to leave the default values, unless a specific cluster size or scaling configuration is desired. Click "Next."



Specify networking. Under *Node group network configuration* select the subnets previously created for EKS in the prerequisites section. Toggle *Configure SSH access to nodes* to "on." Click "Next."



Review the configuration and parameter selection. If everything looks acceptable, click "Create."



Configuring autoscaling deployments and node groups

Pre-requisites

This section assumes the following:

- A Weka cluster has been deployed using start.weka.io and has the requisite amount of storage to support the workload being deployed via Kubernetes.
- The default Weka admin password has been changed and login credentials are available to be used by the CSI secrets YAML file.
- A small-scale file system (named something similar to "eksclient") has been created on the Weka cluster and will not be used for storing application data. This file system can be thin provisioned. A volume with minimum capacity of 5GB and a maximum capacity of 10GB is sufficient.
- YAML example files can be found on docs.weka.io in the CSI plugin section. These example files assume a filesystem named "default" and a nested directory named "shared" with the proper permissions for EKS worker node access. Executing `chmod 777` on the "shared" directory from a system with access to the Weka storage will provide the quickest route to granting access.
- An EKS cluster and node group have been created using the instructions earlier in this documentation.

Supporting resources

[Client installation script](#)

This script should be saved in an S3 bucket which is accessible to the EKS worker node EC2 instances or any other EC2 instances which require Weka client installation.

[Code snippet for launch template user data](#)

This code will be used in the launch template user data.

Obtaining Kubernetes User Data and Creating the Launch Template

Autoscaling groups in Amazon EKS allow for dynamic provisioning (or removal) of EKS worker nodes based upon user specified criteria or defined environmental parameters (CPU utilization, memory utilization, etc). The node groups can also be scaled manually by altering the number of nodes allocated in the related node group settings. In this example, the latter method will be used to increase / decrease node count within the node group.

Amazon AWS Launch Templates allow for various commands or scripts to be run upon instantiation of an EC2 instance. Launch Templates can be thought of as similar - but not identical - to first boot scripts. In the context of Kubernetes, launch templates are used to bootstrap Kubernetes. It is important to note that manually adding to the Launch Template of an existing worker nodes will not merge the new user data with the Kubernetes bootstrap scripts.

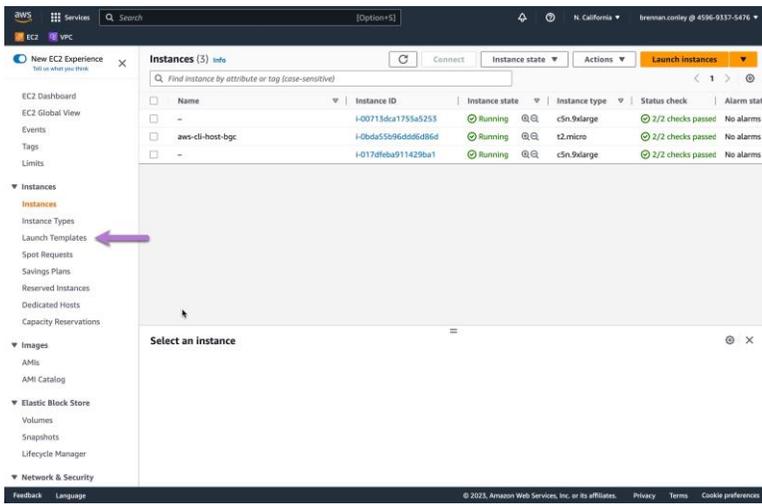
Important

When specifying an AMI, Amazon EKS doesn't merge any user data. Rather, you're responsible for supplying the required `bootstrap` commands for nodes to join the cluster. If your nodes fail to join the cluster, the Amazon EKS `CreateNodegroup` and `UpdateNodegroupVersion` actions also fail.

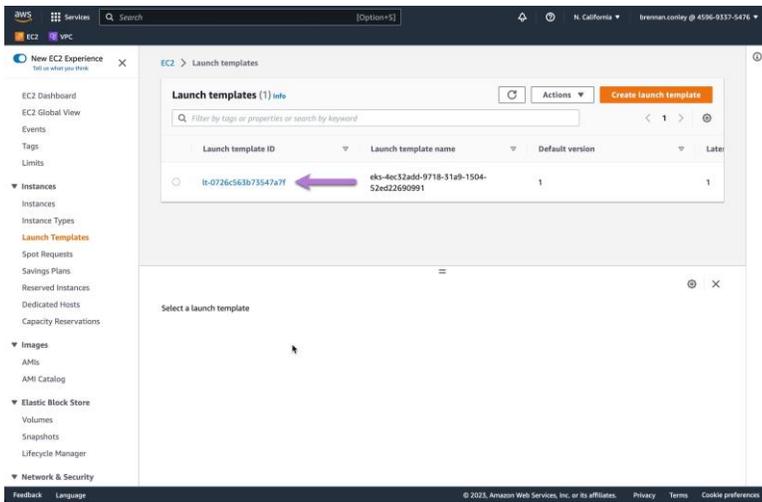
Due to this important nuance, we will need to examine the user data from an existing or temporarily created EKS node group. The instructions provided earlier in this documentation can be used to create the temporary node group, if necessary.

1. Obtain user data for the existing or newly created EKS worker nodes.

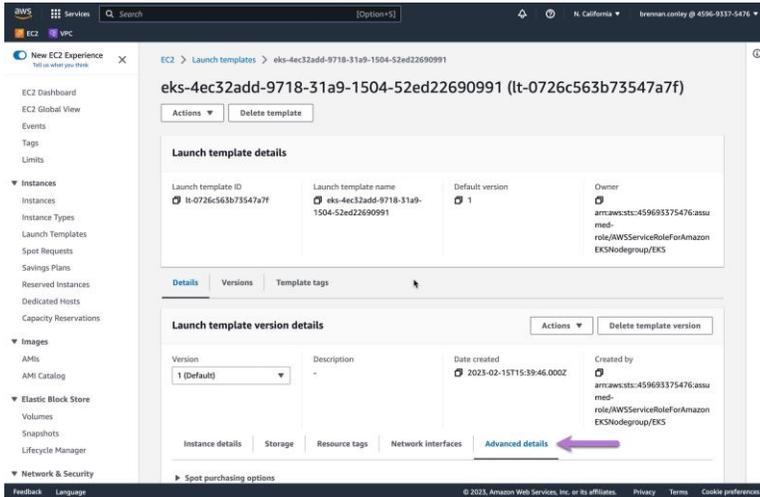
In the AWS console, navigate to the EC2 pane. Select "Launch Templates" from the left hand navigation pane.



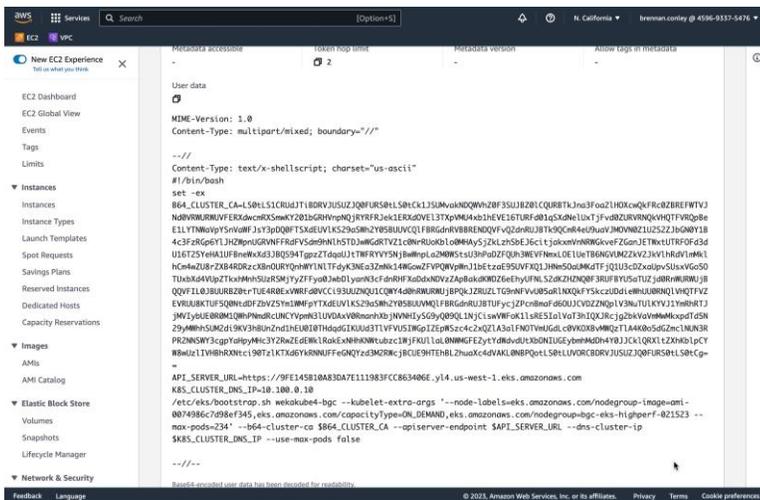
On the *Launch Templates* page, select the Launch Template associated with the existing or newly created EKS node group.



Once on the EKS *Launch Template* page, select "Advanced details."

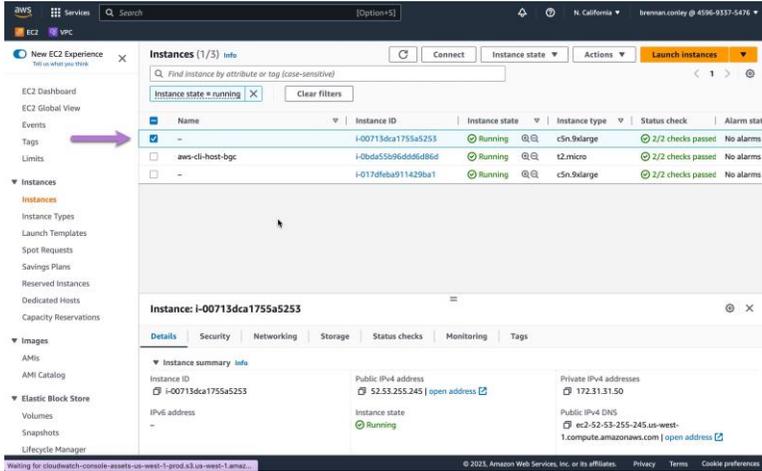


Under the *Advanced details* page, scroll down to view the applicable user data script which is used to bootstrap the EKS worker node EC2 instances into the EKS cluster. Please note, this user data is specific to your EKS cluster. This script will be used later when creating the Launch Template.

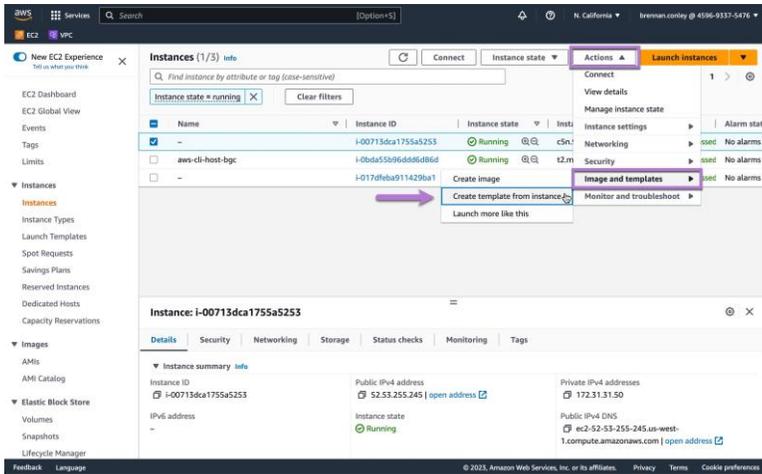


2. Create the new Launch Template.

Navigate to the EC2 instances page and select one of the EKS worker node EC2 instances.

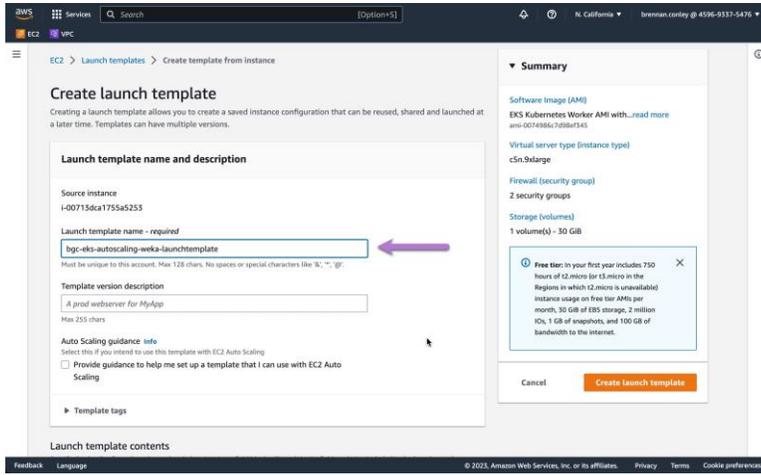


Next, from the "Actions" dropdown, select "Image and Templates," then select "Create template from instance."

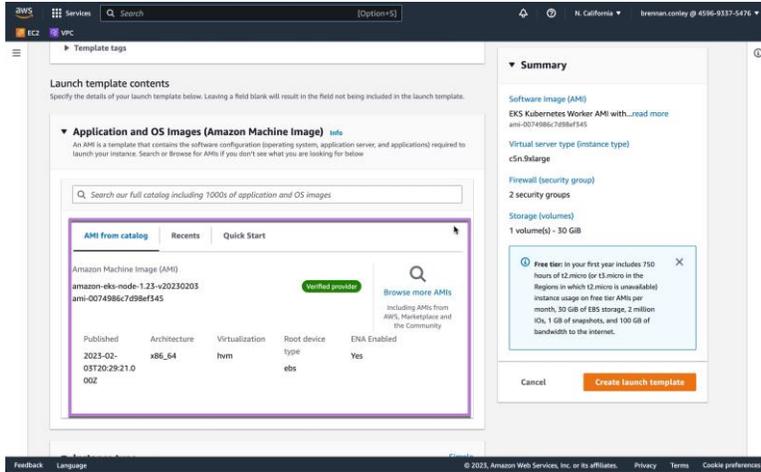


3. Modify the Launch Template parameters.

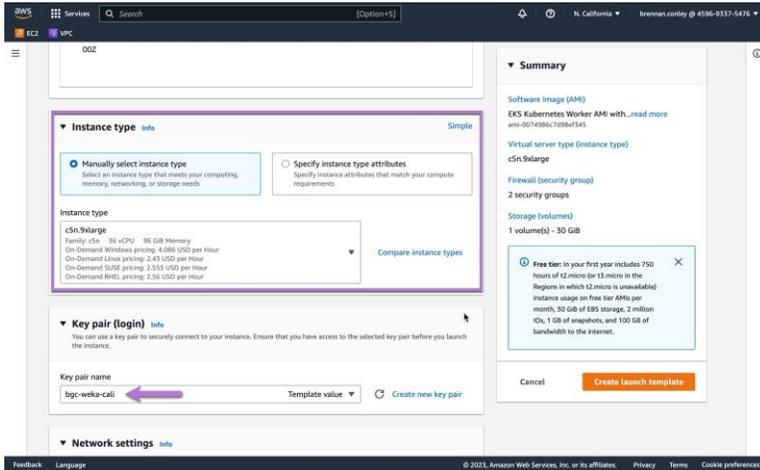
Enter a name for the Launch Template. The *“Template version description”* field is free text and can be used to describe the template’s purpose, though it is not required.



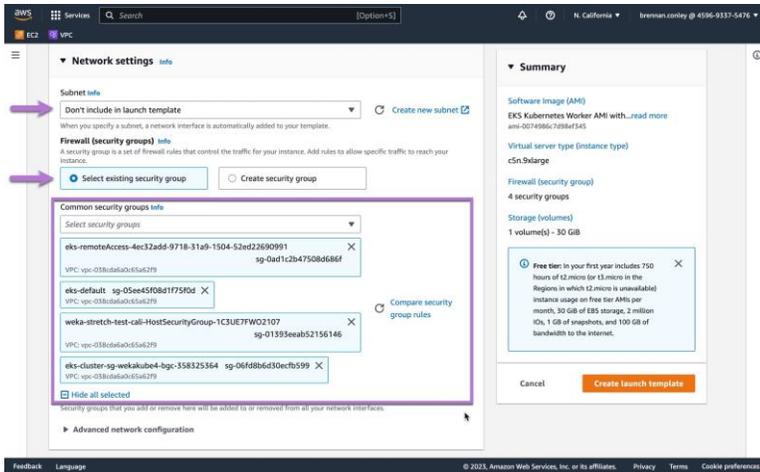
Under *“Application and OS Images”* retain the suggested AMI.



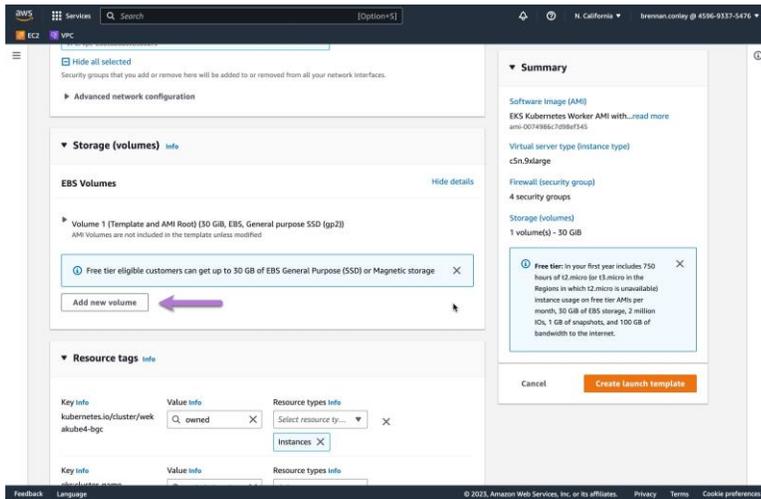
Select the desired instance type to be deployed when using the Launch Template. Additionally, specify the desired AWS SSH key pair. To simplify, the key pair specified should be the same as the key pair used to deploy the Weka cluster.



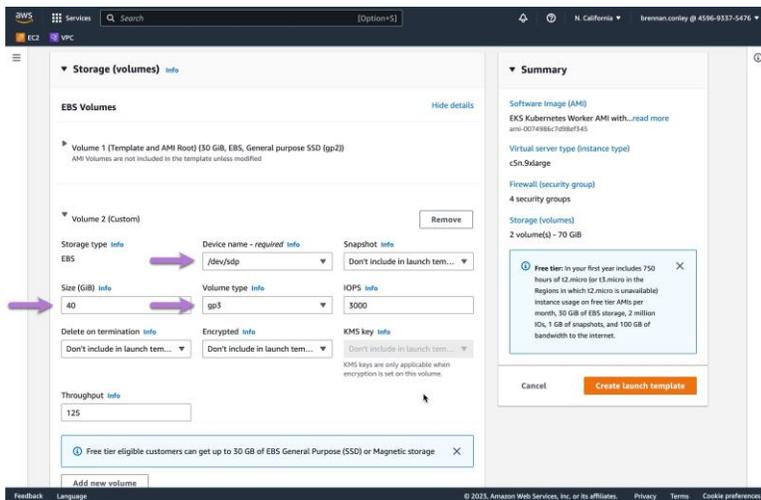
For "Network settings" configure the desired security groups. The Subnet should be set to "Don't include in launch template" as the subnet details will be specified as part of the Autoscaling group. Keep in mind that the selected subnet should have access to the Weka cluster. The security group selection should include the security group for the EKS cluster management interfaces (created upon creation of EKS cluster), the Weka cluster HostSecurityGroup, the EKS remote access security group (if remote access is desired for the node group) and any other relevant security groups for the environment (application specific). Advanced network configuration can remain as default.



Under "Storage (volumes)" EBS Volumes, select "Add new volume."

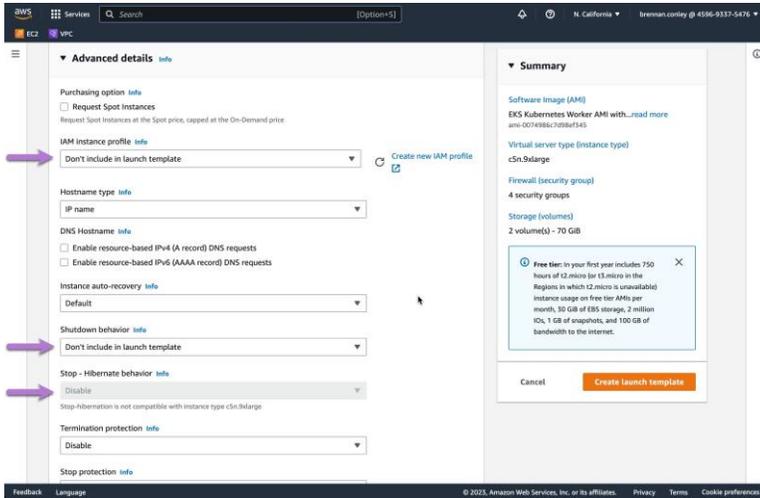


Create an additional small EBS volume (40GB will be sufficient) for the Weka client. The EBS volume will only hold the Weka software installation and debugging logs. Set the *device name* to `/dev/sdp` and the *volume type* to `gp3`. Do not modify the existing EBS volume (listed as Volume 1) which is configured by default.



Leave "Resource tags" as default and do not modify any values.

Expand "Advanced details" and configure the IAM instance profile to "Don't include in launch template." Additionally, configure both Shutdown behavior and Stop - Hibernate behavior to "Don't include in launch template." Some instances do not allow Stop - Hibernate behavior to be modified. In these cases, the field will be disabled and default should be accepted. All additional options can be left as default until User data - optional is reached.

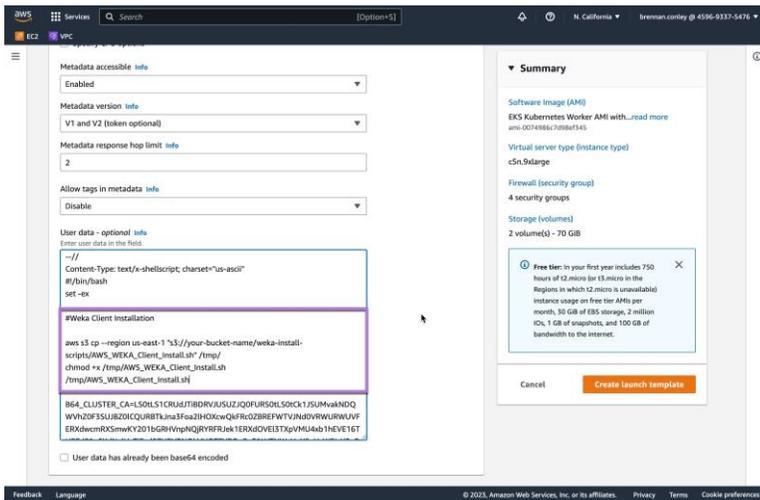


Finally, under *User data - optional* modify the script to include the code snippet below. Before copying and pasting, be sure to update the S3 URL to reflect your S3 bucket name and directory of the stored install script.

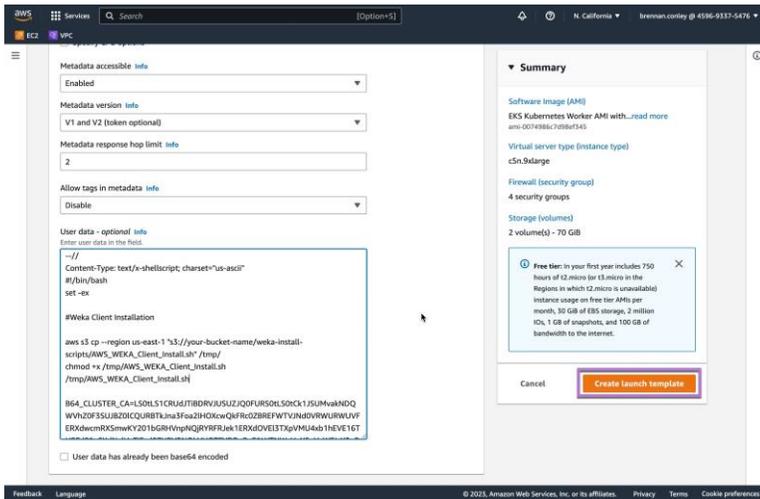
#Weka Client Installation

```
aws s3 cp --region us-east-1 "s3://your-bucket-name/weka-install-scripts/AWS_WEKA_Client_Install.sh" /tmp/
chmod +x /tmp/AWS_WEKA_Client_Install.sh
/tmp/AWS_WEKA_Client_Install.sh
```

The code should be pasted *after* `set -ex` and before `B64_CLUSTER_CA=` in the existing user data. An example is shown below.



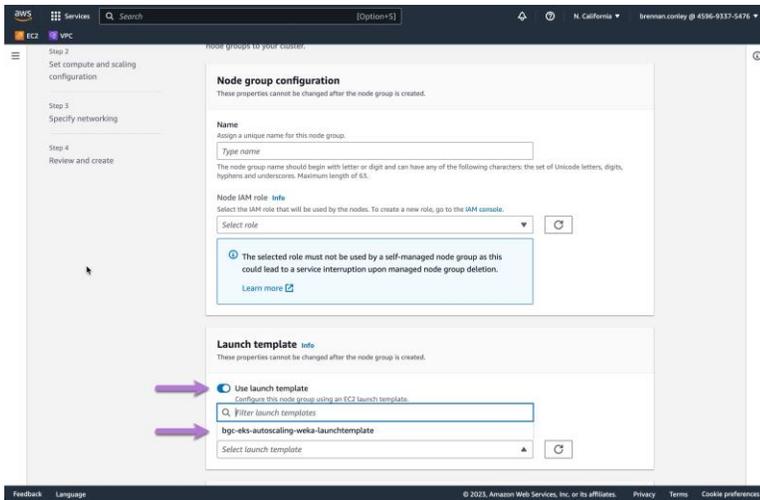
To complete the Launch Template configuration, select "Create launch template."



The Launch Template is now created.

4. Create an autoscaling node group by using the Launch Template

To create a node group with the Weka client pre-installed in preparation for autoscaling node groups, you can follow the instructions provided earlier in this documentation under "Creating node groups - non-autoscaling" with one small modification. When configuring the node group, under "Launch template" ensure "Use launch template" is toggled on as shown below. Select the newly created Launch Template from the "Launch template name" drop down as shown.



Proceed through the remainder of the node group configuration as shown before in the “Creating node groups - non-autoscaling” section. Once the node group has been configured and deployed, the Weka CSI driver can be installed and deployment of applications can begin.

Installing the WEKA CSI driver

Instructions for installing the Weka CSI driver can be found on docs.weka.io in the appendix, linked to [here](#).



weka.io

844.392.0665

