
Amazon ECS

User Guide for AWS Fargate

API Version 2014-11-13



Amazon ECS: User Guide for AWS Fargate

Copyright © 2020 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

What is AWS Fargate?	1
Setting Up	3
Sign Up for AWS	3
Create an IAM User	3
Create a Virtual Private Cloud	5
Install the AWS CLI	5
Docker Basics	6
Installing Docker on Amazon Linux 2	6
Create a Docker Image	7
Push your image to Amazon Elastic Container Registry	8
Clean up	9
Getting Started with Amazon ECS on AWS Fargate	10
Prerequisites	10
Step 1: Create a Task Definition	10
Step 2: Configure the Service	11
Step 3: Configure the Cluster	12
Step 4: Review	12
Step 5: (Optional) View your Service	12
Step 6: Clean Up	12
Platform versions	14
Platform version considerations	14
Available platform versions	14
Migrating from platform version 1.3.0 to 1.4.0	16
Clusters	17
Cluster Concepts	17
Creating a Cluster	18
Cluster Capacity Providers	18
Cluster Capacity Provider Concepts	19
Cluster Capacity Provider Considerations	19
Using AWS Fargate Capacity Providers	20
Fargate Capacity Provider Considerations	20
Handling Fargate Spot Termination Notices	20
Creating a New Cluster That Uses Fargate Capacity Providers	21
Adding Fargate Capacity Providers To An Existing Cluster	22
Running Tasks Using a Fargate Capacity Provider	23
Updating Cluster Settings	23
Deleting a Cluster	24
Task Definitions	25
Task Definition Considerations	25
Network Mode	26
Task CPU and Memory	26
Logging	27
Amazon ECS Task Execution IAM Role	27
Example Task Definition	27
Task Storage	28
Application Architecture	29
Using the Fargate Launch Type	29
Creating a Task Definition	30
Task Definition Template	31
Task definition parameters	35
Family	36
Task execution role	36
Network mode	36
Container Definitions	36

Volumes	54
Launch types	56
Task size	57
Proxy configuration	58
Other task definition parameters	59
Launch Types	60
Fargate Launch Type	61
EC2 Launch Type	62
Using Data Volumes in Tasks	63
Fargate tasks using platform version 1.4.0 or later	28
Fargate tasks using platform version 1.3.0 or earlier	28
Example task definition	29
Amazon EFS Volumes	65
Fargate Task Networking	67
Fargate Task Networking Considerations	69
Using the awslogs Log Driver	69
Enabling the awslogs Log Driver for Your Containers	69
Creating a Log Group	70
Available awslogs Log Driver Options	70
Specifying a Log Configuration in your Task Definition	72
Viewing awslogs Container Logs in CloudWatch Logs	73
Custom Log Routing	75
Considerations	75
Required IAM Permissions	76
Using the AWS for Fluent Bit Image	77
Creating a Task Definition that Uses a FireLens Configuration	79
Using Fluent Logger Libraries	82
Filtering Logs Using Regular Expressions	82
Example Task Definitions	82
Private Registry Authentication for Tasks	85
Required IAM Permissions for Private Registry Authentication	86
Enabling Private Registry Authentication	86
Specifying Sensitive Data	87
Using Secrets Manager	88
Using Systems Manager Parameter Store	93
Example Task Definitions	97
Example: Webserver	98
Example: <code>splunk</code> Log Driver	98
Example: <code>fluentd</code> Log Driver	99
Example: <code>gelf</code> Log Driver	99
Example: Container Dependency	100
Updating a Task Definition	101
Deregistering Task Definition Revisions	101
Account Settings	103
Amazon Resource Names (ARNs) and IDs	104
Viewing Account Settings	105
Modifying Account Settings	105
Scheduling Tasks	108
Running Tasks	109
Running a Task Using the Fargate Launch Type	109
Scheduled Tasks (<code>cron</code>)	111
Task Retirement	113
Working with Tasks Scheduled for Retirement	114
Fargate Task Recycling	115
Services	116
Service scheduler concepts	116
Replica	117

Additional service concepts	117
Service Definition Parameters	117
Launch Type	117
Capacity Provider Strategy	118
Task Definition	119
Platform Version	119
Cluster	119
Service Name	119
Scheduling Strategy	120
Desired Count	120
Deployment Configuration	120
Deployment Controller	122
Task Placement	122
Tags	123
Network Configuration	124
Client Token	127
Service Definition Template	127
Creating a service	129
Step 1: Configuring Basic Service Parameters	129
Step 2: Configure a Network	131
Step 3: Configuring Your Service to Use a Load Balancer	132
Step 4: Configuring Your Service to Use Service Discovery	136
Step 5: Configuring Your Service to Use Service Auto Scaling	137
Step 6: Review and Create Your Service	139
Updating a Service	139
Deleting a Service	141
Deployment Types	142
Rolling Update	143
Blue/Green Deployment with CodeDeploy	143
External Deployment	147
Service Load Balancing	152
Service Load Balancing Considerations	153
Load Balancer Types	154
Creating a Load Balancer	156
Registering Multiple Target Groups with a Service	163
Service Auto Scaling	165
IAM Permissions Required for Service Auto Scaling	165
Target Tracking Scaling Policies	166
Step Scaling Policies	171
Service Discovery	173
Service Discovery Concepts	174
Service Discovery Considerations	174
Amazon ECS Console Experience	175
Service Discovery Pricing	175
Service Throttle Logic	176
Resources and Tags	177
Tagging Your Resources	177
Tag Basics	177
Tagging Your Resources	177
Tag Restrictions	178
Tagging Your Resources for Billing	179
Working with Tags Using the Console	179
Working with Tags Using the CLI or API	180
Usage Reports	181
Monitoring	183
Monitoring Tools	183
Automated Tools	183

Manual Tools	184
CloudWatch Metrics	184
Enabling CloudWatch Metrics	185
Available Metrics and Dimensions	185
Service Utilization	187
Service <code>RUNNING</code> Task Count	188
Viewing Amazon ECS Metrics	188
Events and EventBridge	189
Amazon ECS Events	190
Handling Events	196
CloudWatch Container Insights	197
Container Insights Considerations	198
Working with Container Insights-enabled clusters	198
Logging Amazon ECS API Calls with AWS CloudTrail	199
Amazon ECS Information in CloudTrail	200
Understanding Amazon ECS Log File Entries	200
Security	202
Identity and Access Management	202
Audience	203
Authenticating With Identities	203
Managing Access Using Policies	205
How Amazon Elastic Container Service Works with IAM	206
Identity-Based Policy Examples	210
Supported Resource-Level Permissions	219
Managed Policies and Trust Relationships	220
Service-Linked Role	227
Task Execution IAM Role	236
IAM Roles for Tasks	240
CodeDeploy IAM Role	243
CloudWatch Events IAM Role	246
Troubleshooting	249
Logging and Monitoring	251
Compliance Validation	252
Infrastructure Security	252
Interface VPC endpoints (AWS PrivateLink)	253
Using the Amazon ECS CLI	255
Installing the Amazon ECS CLI	255
Step 1: Download the Amazon ECS CLI	255
Step 2: Verify the Amazon ECS CLI	255
Step 3: Apply Execute Permissions to the Binary	260
Step 4: Complete the Installation	260
Configuring the Amazon ECS CLI	261
Profiles	261
Cluster Configurations	261
Order of Precedence	261
Migrating Configuration Files	262
Migrating Older Configuration Files to the v1.0.0+ Format	263
Tutorial: Creating a Cluster with a Fargate Task Using the Amazon ECS CLI	263
Prerequisites	263
Step 1: Create the Task Execution IAM Role	264
Step 2: Configure the Amazon ECS CLI	264
Step 3: Create a Cluster and Configure the Security Group	265
Step 4: Create a Compose File	265
Step 5: Deploy the Compose File to a Cluster	266
Step 6: View the Running Containers on a Cluster	266
Step 7: View the Container Logs	266
Step 8: Scale the Tasks on the Cluster	267

Step 9: View your Web Application	267
Step 10: Clean Up	267
Tutorial: Creating an Amazon ECS Service That Uses Service Discovery Using the Amazon ECS CLI	268
Prerequisites	268
Configure the Amazon ECS CLI	268
Create an Amazon ECS Service Configured to Use Service Discovery	268
Task metadata endpoint	271
Task metadata endpoint version 4	271
Enabling Task Metadata	271
Task Metadata Endpoint version 4 Paths	272
Task Metadata JSON Response	272
Examples	274
Task metadata endpoint version 3	279
Enabling Task Metadata	279
Task Metadata Endpoint Paths	279
Task Metadata JSON Response	279
Example Task Metadata Response	281
Service Quotas	283
Using Service Quotas	284
Savings Plans	285
Getting started with AWS App Mesh and Amazon ECS	286
Scenario	286
Prerequisites	286
Step 1: Create a mesh and virtual service	287
Step 2: Create a virtual node	287
Step 3: Create a virtual router and route	288
Step 4: Review and create	290
Step 5: Create additional resources	290
Step 6: Update services	294
Tutorials	305
Tutorial: Creating a VPC	305
Step 1: Create an Elastic IP Address for Your NAT Gateway	305
Step 2: Run the VPC Wizard	305
Step 3: Create Additional Subnets	306
Next Steps	307
Tutorial: Creating a Cluster with a Fargate Task Using the AWS CLI	307
Prerequisites	307
Step 1: Create a Cluster	308
Step 2: Register a Task Definition	308
Step 3: List Task Definitions	309
Step 4: Create a Service	309
Step 5: List Services	310
Step 6: Describe the Running Service	310
Step 7: Clean Up	312
Tutorial: Specifying Sensitive Data Using Secrets Manager Secrets	312
Prerequisites	312
Step 1: Create an Secrets Manager Secret	312
Step 2: Update Your Task Execution IAM Role	313
Step 3: Create an Amazon ECS Task Definition	314
Step 4: Create an Amazon ECS Cluster	315
Step 5: Run an Amazon ECS Task	315
Step 6: Verify	316
Step 7: Clean Up	317
Tutorial: Creating a Service Using Service Discovery	317
Prerequisites	318
Step 1: Create the Service Discovery Resources	318
Step 2: Create the Amazon ECS Resources	320

Step 3: Verify Service Discovery	323
Step 4: Clean Up	325
Tutorial: Creating a Service Using a Blue/Green Deployment	327
Prerequisites	327
Step 1: Create an Application Load Balancer	327
Step 2: Create an Amazon ECS Cluster	328
Step 3: Register a Task Definition	328
Step 4: Create an Amazon ECS Service	329
Step 5: Create the AWS CodeDeploy Resources	330
Step 6: Create and Monitor an CodeDeploy Deployment	332
Step 7: Clean Up	334
Tutorial: Listening for Amazon ECS CloudWatch Events	335
Prerequisite: Set Up a Test Cluster	336
Step 1: Create the Lambda Function	336
Step 2: Register Event Rule	336
Step 3: Test Your Rule	337
Tutorial: Sending Amazon Simple Notification Service Alerts for Task Stopped Events	337
Prerequisite: Set Up a Test Cluster	337
Step 1: Create and Subscribe to an Amazon SNS Topic	337
Step 2: Register Event Rule	338
Step 3: Test Your Rule	338
Troubleshooting	340
Checking stopped tasks for errors	340
Stopped tasks error codes	344
CannotPullContainer task errors	347
Service Event Messages	348
Service Event Messages	349
Invalid CPU or memory value specified	349
Troubleshooting service load balancers	350
Document History	352
AWS glossary	362

What is AWS Fargate?

AWS Fargate is a technology that you can use with Amazon ECS to run [containers](#) without having to manage servers or clusters of Amazon EC2 instances. With Fargate, you no longer have to provision, configure, or scale clusters of virtual machines to run containers. This removes the need to choose server types, decide when to scale your clusters, or optimize cluster packing.

When you run your Amazon ECS tasks and services with the Fargate launch type or a Fargate capacity provider, you package your application in containers, specify the CPU and memory requirements, define networking and IAM policies, and launch the application. Each Fargate task has its own isolation boundary and does not share the underlying kernel, CPU resources, memory resources, or elastic network interface with another task.

This topic describes the different components of Fargate tasks and services, and calls out special considerations for using Fargate with Amazon ECS.

Amazon ECS on AWS Fargate is currently only available in the following Regions:

Region Name	Region
US East (Ohio)	us-east-2
US East (N. Virginia)	us-east-1
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Africa (Cape Town)	af-south-1
Asia Pacific (Hong Kong)	ap-east-1
Asia Pacific (Mumbai)	ap-south-1 (aps1-az1 & aps1-az3 only)
Asia Pacific (Seoul)	ap-northeast-2
Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
Canada (Central)	ca-central-1
China (Beijing)	cn-north-1
China (Ningxia)	cn-northwest-1
Europe (Frankfurt)	eu-central-1
Europe (Ireland)	eu-west-1
Europe (London)	eu-west-2
Europe (Paris)	eu-west-3
Europe (Stockholm)	eu-north-1

Region Name	Region
South America (São Paulo)	sa-east-1
Middle East (Bahrain)	me-south-1
AWS GovCloud (US-East)	us-gov-east-1
AWS GovCloud (US-West)	us-gov-west-1

The following walkthroughs help you get started using Amazon ECS on Fargate.

- [Getting Started with Amazon ECS on AWS Fargate](#) (p. 10)
- the section called “Tutorial: Creating a Cluster with a Fargate Task Using the AWS CLI” (p. 307)
- the section called “Tutorial: Creating a Cluster with a Fargate Task Using the Amazon ECS CLI” (p. 263)

For more information about Amazon Elastic Container Service, see [What is Amazon ECS?](#).

Setting Up with Amazon ECS

If you've already signed up for Amazon Web Services (AWS) and have been using Amazon Elastic Compute Cloud (Amazon EC2), you are close to being able to use Amazon ECS. The set-up process for the two services is similar. The following guide prepares you for launching your first cluster using either the Amazon ECS first-run wizard or the Amazon ECS Command Line Interface (CLI).

Complete the following tasks to get set up for Amazon ECS. If you have already completed any of these steps, you may skip them and move on to installing the custom AWS CLI.

Sign Up for AWS

When you sign up for AWS, your AWS account is automatically signed up for all services, including Amazon EC2 and Amazon ECS. You are charged only for the services that you use.

If you have an AWS account already, skip to the next task. If you don't have an AWS account, use the following procedure to create one.

To create an AWS account

1. Open <https://portal.aws.amazon.com/billing/signup>.
2. Follow the online instructions.

Part of the sign-up procedure involves receiving a phone call and entering a verification code on the phone keypad.

Note your AWS account number, because you'll need it for the next task.

Create an IAM User

Services in AWS, such as Amazon EC2 and Amazon ECS, require that you provide credentials when you access them, so that the service can determine whether you have permission to access its resources. The console requires your password. You can create access keys for your AWS account to access the command line interface or API. However, we don't recommend that you access AWS using the credentials for your AWS account; we recommend that you use AWS Identity and Access Management (IAM) instead. Create an IAM user, and then add the user to an IAM group with administrative permissions or and grant this user administrative permissions. You can then access AWS using a special URL and the credentials for the IAM user.

If you signed up for AWS but have not created an IAM user for yourself, you can create one using the IAM console.

To create an administrator user for yourself and add the user to an administrators group (console)

1. Use your AWS account email address and password to sign in as the *AWS account root user* to the IAM console at <https://console.aws.amazon.com/iam/>.

Note

We strongly recommend that you adhere to the best practice of using the **Administrator** IAM user below and securely lock away the root user credentials. Sign in as the root user only to perform a few [account and service management tasks](#).

2. In the navigation pane, choose **Users** and then choose **Add user**.
3. For **User name**, enter **Administrator**.
4. Select the check box next to **AWS Management Console access**. Then select **Custom password**, and then enter your new password in the text box.
5. (Optional) By default, AWS requires the new user to create a new password when first signing in. You can clear the check box next to **User must create a new password at next sign-in** to allow the new user to reset their password after they sign in.
6. Choose **Next: Permissions**.
7. Under **Set permissions**, choose **Add user to group**.
8. Choose **Create group**.
9. In the **Create group** dialog box, for **Group name** enter **Administrators**.
10. Choose **Filter policies**, and then select **AWS managed -job function** to filter the table contents.
11. In the policy list, select the check box for **AdministratorAccess**. Then choose **Create group**.

Note

You must activate IAM user and role access to Billing before you can use the **AdministratorAccess** permissions to access the AWS Billing and Cost Management console. To do this, follow the instructions in [step 1 of the tutorial about delegating access to the billing console](#).

12. Back in the list of groups, select the check box for your new group. Choose **Refresh** if necessary to see the group in the list.
13. Choose **Next: Tags**.
14. (Optional) Add metadata to the user by attaching tags as key-value pairs. For more information about using tags in IAM, see [Tagging IAM Entities](#) in the *IAM User Guide*.
15. Choose **Next: Review** to see the list of group memberships to be added to the new user. When you are ready to proceed, choose **Create user**.

You can use this same process to create more groups and users and to give your users access to your AWS account resources. To learn about using policies that restrict user permissions to specific AWS resources, see [Access Management](#) and [Example Policies](#).

To sign in as this new IAM user, sign out of the AWS console, then use the following URL, where *your_aws_account_id* is your AWS account number without the hyphens (for example, if your AWS account number is 1234-5678-9012, your AWS account ID is 123456789012):

```
https://your_aws_account_id.signin.aws.amazon.com/console/
```

Enter the IAM user name and password that you just created. When you're signed in, the navigation bar displays "*your_user_name* @ *your_aws_account_id*".

If you don't want the URL for your sign-in page to contain your AWS account ID, you can create an account alias. From the IAM dashboard, choose **Create Account Alias** and enter an alias, such as your company name. To sign in after you create an account alias, use the following URL:

```
https://your_account_alias.signin.aws.amazon.com/console/
```

To verify the sign-in link for IAM users for your account, open the IAM console and check under **IAM users sign-in link** on the dashboard.

For more information about IAM, see the [AWS Identity and Access Management User Guide](#).

Create a Virtual Private Cloud

Amazon Virtual Private Cloud (Amazon VPC) enables you to launch AWS resources into a virtual network that you've defined.

Note

The Amazon ECS console first-run experience creates a VPC for your cluster, so if you intend to use the Amazon ECS console, you can skip to the next section.

If you have a default VPC, you also can skip this section and move to the next task, [Install the AWS CLI \(p. 5\)](#). To determine whether you have a default VPC, see [Supported Platforms in the Amazon EC2 Console](#) in the *Amazon EC2 User Guide for Linux Instances*. Otherwise, you can create a nondefault VPC in your account using the steps below.

Important

If your account supports Amazon EC2 Classic in a region, then you do not have a default VPC in that region.

To create a nondefault VPC

1. Open the Amazon VPC console at <https://console.aws.amazon.com/vpc/>.
2. From the navigation bar, select a region for the VPC. VPCs are specific to a region, so you should select the same region in which you created your key pair.
3. On the VPC dashboard, choose **Launch VPC Wizard**.
4. On the **Step 1: Select a VPC Configuration** page, ensure that **VPC with a Single Public Subnet** is selected, and choose **Select**.
5. On the **Step 2: VPC with a Single Public Subnet** page, enter a friendly name for your VPC in the **VPC name** field. Leave the other default configuration settings, and choose **Create VPC**. On the confirmation page, choose **OK**.

For more information about Amazon VPC, see [What is Amazon VPC?](#) in the *Amazon VPC User Guide*.

Install the AWS CLI

The AWS Management Console can be used to manage all operations manually with Amazon ECS. However, installing the AWS CLI on your local desktop or a developer box enables you to build scripts that can automate common management tasks in Amazon ECS.

To use the AWS CLI with Amazon ECS, install the latest AWS CLI, version. For information about installing the AWS CLI or upgrading it to the latest version, see [Installing the AWS Command Line Interface](#) in the *AWS Command Line Interface User Guide*.

Docker Basics for Amazon ECS

Docker is a technology that allows you to build, run, test, and deploy distributed applications that are based on Linux containers. Amazon ECS uses Docker images in task definitions to launch containers on Amazon EC2 instances in your clusters. For Amazon ECS product details, featured customer case studies, and FAQs, see the [Amazon Elastic Container Service product detail pages](#).

The documentation in this guide assumes that readers possess a basic understanding of what Docker is and how it works. For more information about Docker, see [What is Docker?](#) and the [Docker overview](#).

Installing Docker on Amazon Linux 2

Note

If you already have Docker installed, skip to [Create a Docker Image \(p. 7\)](#).

Docker is available on many different operating systems, including most modern Linux distributions, like Ubuntu, and even Mac OSX and Windows. For more information about how to install Docker on your particular operating system, go to the [Docker installation guide](#).

You don't even need a local development system to use Docker. If you are using Amazon EC2 already, you can launch an Amazon Linux 2 instance and install Docker to get started.

To install Docker on an Amazon EC2 instance

1. Launch an instance with the Amazon Linux 2 AMI. For more information, see [Launching an Instance](#) in the *Amazon EC2 User Guide for Linux Instances*.
2. Connect to your instance. For more information, see [Connect to Your Linux Instance](#) in the *Amazon EC2 User Guide for Linux Instances*.
3. Update the installed packages and package cache on your instance.

```
sudo yum update -y
```

4. Install the most recent Docker Community Edition package.

```
sudo amazon-linux-extras install docker
```

5. Start the Docker service.

```
sudo service docker start
```

6. Add the `ec2-user` to the `docker` group so you can execute Docker commands without using `sudo`.

```
sudo usermod -a -G docker ec2-user
```

7. Log out and log back in again to pick up the new `docker` group permissions. You can accomplish this by closing your current SSH terminal window and reconnecting to your instance in a new one. Your new SSH session will have the appropriate `docker` group permissions.
8. Verify that the `ec2-user` can run Docker commands without `sudo`.

```
docker info
```

Note

In some cases, you may need to reboot your instance to provide permissions for the `ec2-user` to access the Docker daemon. Try rebooting your instance if you see the following error:

```
Cannot connect to the Docker daemon. Is the docker daemon running on this host?
```

Create a Docker Image

Amazon ECS task definitions use Docker images to launch containers on the container instances in your clusters. In this section, you create a Docker image of a simple web application, and test it on your local system or Amazon EC2 instance, and then push the image to a container registry (such as Amazon ECR or Docker Hub) so you can use it in an Amazon ECS task definition.

To create a Docker image of a simple web application

1. Create a file called `Dockerfile`. A Dockerfile is a manifest that describes the base image to use for your Docker image and what you want installed and running on it. For more information about Dockerfiles, go to the [Dockerfile Reference](#).

```
touch Dockerfile
```

2. Edit the `Dockerfile` you just created and add the following content.

```
FROM ubuntu:18.04

# Install dependencies
RUN apt-get update && \
    apt-get -y install apache2

# Install apache and write hello world message
RUN echo 'Hello World!' > /var/www/html/index.html

# Configure apache
RUN echo '. /etc/apache2/envvars' > /root/run_apache.sh && \
    echo 'mkdir -p /var/run/apache2' >> /root/run_apache.sh && \
    echo 'mkdir -p /var/lock/apache2' >> /root/run_apache.sh && \
    echo '/usr/sbin/apache2 -D FOREGROUND' >> /root/run_apache.sh && \
    chmod 755 /root/run_apache.sh

EXPOSE 80

CMD /root/run_apache.sh
```

This Dockerfile uses the Ubuntu 18.04 image. The `RUN` instructions update the package caches, install some software packages for the web server, and then write the "Hello World!" content to the web server's document root. The `EXPOSE` instruction exposes port 80 on the container, and the `CMD` instruction starts the web server.

3. Build the Docker image from your Dockerfile.

Note

Some versions of Docker may require the full path to your Dockerfile in the following command, instead of the relative path shown below.

```
docker build -t hello-world .
```

4. Run **docker images** to verify that the image was created correctly.

```
docker images --filter reference=hello-world
```

Output:

REPOSITORY	TAG	IMAGE ID	CREATED	SIZE
hello-world	latest	e9ffedc8c286	4 minutes ago	241MB

5. Run the newly built image. The `-p 80:80` option maps the exposed port 80 on the container to port 80 on the host system. For more information about **docker run**, go to the [Docker run reference](#).

```
docker run -t -i -p 80:80 hello-world
```

Note

Output from the Apache web server is displayed in the terminal window. You can ignore the "Could not reliably determine the server's fully qualified domain name" message.

6. Open a browser and point to the server that is running Docker and hosting your container.
 - If you are using an EC2 instance, this is the **Public DNS** value for the server, which is the same address you use to connect to the instance with SSH. Make sure that the security group for your instance allows inbound traffic on port 80.
 - If you are running Docker locally, point your browser to <http://localhost/>.
 - If you are using **docker-machine** on a Windows or Mac computer, find the IP address of the VirtualBox VM that is hosting Docker with the **docker-machine ip** command, substituting *machine-name* with the name of the docker machine you are using.

```
docker-machine ip machine-name
```

You should see a web page with your "Hello World!" statement.

7. Stop the Docker container by typing **Ctrl + c**.

Push your image to Amazon Elastic Container Registry

Amazon ECR is a managed AWS Docker registry service. Customers can use the familiar Docker CLI to push, pull, and manage images. For Amazon ECR product details, featured customer case studies, and FAQs, see the [Amazon Elastic Container Registry product detail pages](#).

This section requires the following:

- You have the AWS CLI installed and configured. If you do not have the AWS CLI installed on your system, see [Installing the AWS Command Line Interface](#) in the *AWS Command Line Interface User Guide*.
- Your user has the required IAM permissions to access the Amazon ECR service. For more information, see [Amazon ECR Managed Policies](#).

To tag your image and push it to Amazon ECR

1. Create an Amazon ECR repository to store your hello-world image. Note the repositoryUri in the output.

```
aws ecr create-repository --repository-name hello-repository --region region
```

Output:

```
{
  "repository": {
    "registryId": "aws_account_id",
    "repositoryName": "hello-repository",
    "repositoryArn": "arn:aws:ecr:region:aws_account_id:repository/hello-
repository",
    "createdAt": 1505337806.0,
    "repositoryUri": "aws_account_id.dkr.ecr.region.amazonaws.com/hello-repository"
  }
}
```

2. Tag the hello-world image with the repositoryUri value from the previous step.

```
docker tag hello-world aws_account_id.dkr.ecr.region.amazonaws.com/hello-repository
```

3. Run the **aws ecr get-login-password** command. Specify the registry URI you want to authenticate to. For more information, see [Registry Authentication](#) in the *Amazon Elastic Container Registry User Guide*.

```
aws ecr get-login-password | docker login --username AWS --password-
stdin aws_account_id.dkr.ecr.region.amazonaws.com
```

Output:

```
Login Succeeded
```

Important

If you receive an error, install or upgrade to the latest version of the AWS CLI. For more information, see [Installing the AWS Command Line Interface](#) in the *AWS Command Line Interface User Guide*.

4. Push the image to Amazon ECR with the repositoryUri value from the earlier step.

```
docker push aws_account_id.dkr.ecr.region.amazonaws.com/hello-repository
```

Clean up

When you are done experimenting with your Amazon ECR image, you can delete the repository so you are not charged for image storage.

```
aws ecr delete-repository --repository-name hello-repository --region region --force
```

Getting Started with Amazon ECS on AWS Fargate

Amazon Elastic Container Service (Amazon ECS) is a highly scalable, fast, container management service that makes it easy to run, stop, and manage your containers. You can host your containers on a serverless infrastructure that is managed by Amazon ECS by launching your services or tasks on AWS Fargate. For a broad overview on AWS Fargate, see [What is AWS Fargate? \(p. 1\)](#).

Get started with Amazon ECS on AWS Fargate by using the Fargate launch type for your tasks. In the Regions where Amazon ECS supports AWS Fargate, the Amazon ECS first-run wizard guides you through the process of getting started with Amazon ECS using the Fargate launch type. The wizard gives you the option of creating a cluster and launching a sample web application. If you already have a Docker image to launch in Amazon ECS, you can create a task definition with that image and use that for your cluster instead.

Important

For more information about the Amazon ECS first-run wizard for EC2 tasks, see [Getting Started with Amazon ECS](#).

Complete the following steps to get started with Amazon ECS on AWS Fargate.

Prerequisites

Before you begin, be sure that you've completed the steps in [Setting Up with Amazon ECS \(p. 3\)](#) and that your AWS user has either the permissions specified in the AdministratorAccess or [Amazon ECS First Run Wizard Permissions \(p. 212\)](#) IAM policy example.

The first-run wizard attempts to automatically create the task execution IAM role, which is required for Fargate tasks. To ensure that the first-run experience is able to create this IAM role, one of the following must be true:

- Your user has administrator access. For more information, see [Setting Up with Amazon ECS \(p. 3\)](#).
- Your user has the IAM permissions to create a service role. For more information, see [Creating a Role to Delegate Permissions to an AWS Service](#).
- A user with administrator access has manually created the task execution role so that it is available on the account to be used. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

Step 1: Create a Task Definition

A task definition is like a blueprint for your application. Each time you launch a task in Amazon ECS, you specify a task definition. The service then knows which Docker image to use for containers, how many containers to use in the task, and the resource allocation for each container.

1. Open the Amazon ECS console first-run wizard at <https://console.aws.amazon.com/ecs/home#/firstRun>.
2. From the navigation bar, select the **US East (N. Virginia)** Region.

Note

You can complete this first-run wizard using these steps for any Region that supports Amazon ECS using Fargate. For more information, see [Fargate Launch Type \(p. 61\)](#).

3. Configure your container definition parameters.

For **Container definition**, the first-run wizard comes preloaded with the `sample-app`, `nginx`, and `tomcat-webserver` container definitions in the console. You can optionally rename the container or review and edit the resources used by the container (such as CPU units and memory limits) by choosing **Edit** and editing the values shown. For more information, see [Container Definitions \(p. 36\)](#).

Note

If you are using an Amazon ECR image in your container definition, be sure to use the full `registry/repository:tag` naming for your Amazon ECR images. For example, `aws_account_id.dkr.ecr.region.amazonaws.com/my-web-app:latest`.

4. For **Task definition**, the first-run wizard defines a task definition to use with the preloaded container definitions. You can optionally rename the task definition and edit the resources used by the task (such as the **Task memory** and **Task CPU** values) by choosing **Edit** and editing the values shown. For more information, see [Task definition parameters \(p. 35\)](#).

Task definitions created in the first-run wizard are limited to a single container for simplicity. You can create multi-container task definitions later in the Amazon ECS console.

5. Choose **Next**.

Step 2: Configure the Service

In this section of the wizard, select how to configure the Amazon ECS service that is created from your task definition. A service launches and maintains a specified number of copies of the task definition in your cluster. The **Amazon ECS sample** application is a web-based Hello World-style application that is meant to run indefinitely. By running it as a service, it restarts if the task becomes unhealthy or unexpectedly stops.

The first-run wizard comes preloaded with a service definition, and you can see the `sample-app-service` service defined in the console. You can optionally rename the service or review and edit the details by choosing **Edit** and doing the following:

1. In the **Service name** field, select a name for your service.
2. In the **Number of desired tasks** field, enter the number of tasks to launch with your specified task definition.
3. In the **Security group** field, specify a range of IPv4 addresses to allow inbound traffic from, in CIDR block notation. For example, `203.0.113.0/24`.
4. (Optional) You can choose to use an Application Load Balancer with your service. When a task is launched from a service that is configured to use a load balancer, the task is registered with the load balancer. Traffic from the load balancer is distributed across the instances in the load balancer. For more information, see [Introduction to Application Load Balancers](#).

Important

Application Load Balancers do incur cost while they exist in your AWS resources. For more information, see [Application Load Balancer Pricing](#).

Complete the following steps to use a load balancer with your service.

- In the **Container to load balance** section, choose the **Load balancer listener port**. The default value here is set up for the sample application, but you can configure different listener options for the load balancer. For more information, see [Service Load Balancing \(p. 152\)](#).

5. Review your service settings and click **Save, Next**.

Step 3: Configure the Cluster

In this section of the wizard, you name your cluster, and then Amazon ECS takes care of the networking and IAM configuration for you.

1. In the **Cluster name** field, choose a name for your cluster.
2. Click **Next** to proceed.

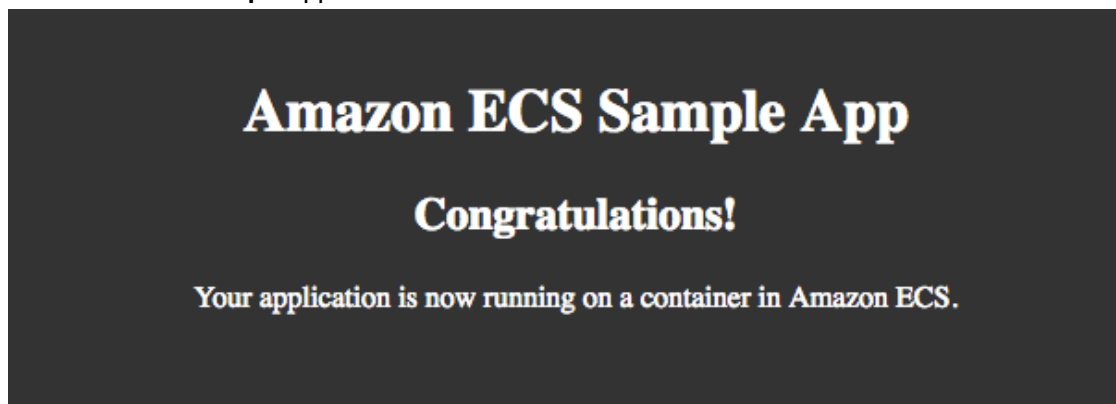
Step 4: Review

1. Review your task definition, task configuration, and cluster configuration and click **Create** to finish. You are directed to a **Launch Status** page that shows the status of your launch. It describes each step of the process (this can take a few minutes to complete while your Auto Scaling group is created and populated).
2. After the launch is complete, choose **View service**.

Step 5: (Optional) View your Service

If your service is a web-based application, such as the **Amazon ECS sample** application, you can view its containers with a web browser.

1. On the **Service: *service-name*** page, choose the **Tasks** tab.
2. Choose a task from the list of tasks in your service.
3. In the **Network** section, choose the **ENI Id** for your task. This takes you to the Amazon EC2 console where you can view the details of the network interface associated with your task, including the **IPv4 Public IP** address.
4. Enter the **IPv4 Public IP** address in your web browser and you should see a webpage that displays the **Amazon ECS sample** application.



Step 6: Clean Up

When you are finished using an Amazon ECS cluster, you should clean up the resources associated with it to avoid incurring charges for resources that you are not using.

Some Amazon ECS resources, such as tasks, services, clusters, and container instances, are cleaned up using the Amazon ECS console. Other resources, such as Amazon EC2 instances, Elastic Load Balancing load balancers, and Auto Scaling groups, must be cleaned up manually in the Amazon EC2 console or by deleting the AWS CloudFormation stack that created them.

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation pane, choose **Clusters**.
3. On the **Clusters** page, select the cluster to delete.
4. Choose **Delete Cluster**. At the confirmation prompt, enter **delete me** and then choose **Delete**. Deleting the cluster cleans up the associated resources that were created with the cluster, including Auto Scaling groups, VPCs, or load balancers.

AWS Fargate platform versions

AWS Fargate platform versions are used to refer to a specific runtime environment for Fargate task infrastructure. It is a combination of the kernel and container runtime versions.

New platform versions are released as the runtime environment evolves, for example, if there are kernel or operating system updates, new features, bug fixes, or security updates. Security updates and patches are deployed automatically for your Fargate tasks. If a security issue is found that affects a platform version, AWS patches the platform version. In some cases, you may be notified that your Fargate tasks have been scheduled for retirement. For more information, see [Task Retirement \(p. 113\)](#).

Topics

- [Platform version considerations \(p. 14\)](#)
- [Available AWS Fargate platform versions \(p. 14\)](#)
- [Migrating from platform version 1.3.0 to 1.4.0 \(p. 16\)](#)

Platform version considerations

The following should be considered when specifying a platform version:

- When specifying a platform version, you can use either a specific version number, for example 1.4.0, or `LATEST` (which uses the 1.3.0 platform version).
- To use a specific platform version, specify the version number when creating or updating your service. If you specify `LATEST`, your tasks use platform version 1.3.0.
- In the China (Beijing) and China (Ningxia) Regions, the only supported platform versions are 1.4.0 and 1.3.0. The AWS Management Console displays older platform versions but an error will be returned if they are chosen. The `LATEST` platform version is supported because it uses the 1.3.0 platform version.
- If you have a service with running tasks and want to update their platform version, you can update your service, specify a new platform version, and choose **Force new deployment**. Your tasks are redeployed with the latest platform version. For more information, see [Updating a Service \(p. 139\)](#).
- If your service is scaled up without updating the platform version, those tasks receive the platform version that was specified on the service's current deployment.

Available AWS Fargate platform versions

The following is a list of the platform versions currently available:

Fargate platform version-1.4.0

- Beginning on May 28, 2020, any new Fargate task that is launched using platform version 1.4.0 will have its ephemeral storage encrypted with an AES-256 encryption algorithm using an AWS Fargate-managed encryption key. For more information, see [Using Data Volumes in Tasks \(p. 63\)](#).
- Added support for using Amazon EFS file system volumes for persistent task storage. For more information, see [Amazon EFS Volumes \(p. 65\)](#).
- The ephemeral task storage has been increased to a minimum of 20 GB for each task. For more information, see [Using Data Volumes in Tasks \(p. 63\)](#).

- The network traffic behavior to and from tasks has been updated. Starting with platform version 1.4.0, all Fargate tasks receive a single elastic network interface (referred to as the task ENI) and all network traffic flows through that ENI within your VPC and will be visible to you through your VPC flow logs. For more information, see [Fargate Task Networking \(p. 67\)](#).
- Task ENIs add support for jumbo frames. Network interfaces are configured with a maximum transmission unit (MTU), which is the size of the largest payload that fits within a single frame. The larger the MTU, the more application payload can fit within a single frame, which reduces per-frame overhead and increases efficiency. Supporting jumbo frames will reduce overhead when the network path between your task and the destination supports jumbo frames, such as all traffic that remains within your VPC.
- CloudWatch Container Insights will include network performance metrics for Fargate tasks. For more information, see [Amazon ECS CloudWatch Container Insights \(p. 197\)](#).
- Added support for the task metadata endpoint version 4 which provides additional information for your Fargate tasks, including network stats for the task and which Availability Zone the task is running in. For more information, see [Task metadata endpoint version 4 \(p. 271\)](#).
- Added support for the `SYS_PTRACE` Linux parameter in container definitions. For more information, see [Linux Parameters \(p. 50\)](#).
- The Fargate container agent replaces the use of the Amazon ECS container agent for all Fargate tasks. This change should not have an effect on how your tasks run.
- The container runtime is now using Containerd instead of Docker. This change should not have an effect on how your tasks run. You will notice that some error messages that originate with the container runtime will change from mentioning Docker to more general errors. For more information, see [Stopped tasks error codes](#) in the *Amazon Elastic Container Service User Guide for AWS Fargate*.

Fargate platform version-1.3.0

- Beginning on Sept 30, 2019, any new Fargate task that is launched supports the `awsfirelens` log driver. FireLens for Amazon ECS enables you to use task definition parameters to route logs to an AWS service or AWS Partner Network (APN) destination for log storage and analytics. For more information, see [Custom Log Routing \(p. 75\)](#).
- Added task recycling for Fargate tasks, which is the process of refreshing tasks that are a part of an Amazon ECS service. For more information, see [Fargate Task Recycling \(p. 115\)](#).
- Beginning on March 27, 2019, any new Fargate task that is launched can use additional task definition parameters that enable you to define a proxy configuration, dependencies for container startup and shutdown as well as a per-container start and stop timeout value. For more information, see [Proxy configuration \(p. 58\)](#), [Container Dependency \(p. 51\)](#), and [Container Timeouts \(p. 52\)](#).
- Beginning on April 2, 2019, any new Fargate task that is launched supports injecting sensitive data into your containers by storing your sensitive data in either AWS Secrets Manager secrets or AWS Systems Manager Parameter Store parameters and then referencing them in your container definition. For more information, see [Specifying Sensitive Data \(p. 87\)](#).
- Beginning on May 1, 2019, any new Fargate task that is launched supports referencing sensitive data in the log configuration of a container using the `secretOptions` container definition parameter. For more information, see [Specifying Sensitive Data \(p. 87\)](#).
- Beginning on May 1, 2019, any new Fargate task that is launched supports the `spunk` log driver in addition to the `awslogs` log driver. For more information, see [Storage and Logging \(p. 45\)](#).
- Beginning on July 9, 2019, any new Fargate tasks that is launched supports CloudWatch Container Insights. For more information, see [Amazon ECS CloudWatch Container Insights \(p. 197\)](#).
- Beginning on December 3, 2019, the Fargate Spot capacity provider is supported. For more information, see [Using AWS Fargate Capacity Providers \(p. 20\)](#).

Fargate Platform Version-1.2.0

- Added support for private registry authentication using AWS Secrets Manager. For more information, see [Private Registry Authentication for Tasks \(p. 85\)](#).

Fargate Platform Version-1.1.0

- Added support for the Amazon ECS task metadata endpoint. For more information, see [Amazon ECS task metadata endpoint \(p. 271\)](#).
- Added support for Docker health checks in container definitions. For more information, see [Health Check \(p. 40\)](#).
- Added support for Amazon ECS service discovery. For more information, see [Service Discovery \(p. 173\)](#).

Fargate Platform Version-1.0.0

- Based on Amazon Linux 2017.09.
- Initial release.

Migrating from platform version 1.3.0 to 1.4.0

The following should be considered when migrating your Amazon ECS on Fargate tasks from platform version 1.3.0 to platform version 1.4.0.

- The network traffic behavior to and from tasks has been updated. Starting with platform version 1.4.0, all Amazon ECS on Fargate tasks receive a single elastic network interface (referred to as the task ENI) and all network traffic flows through that ENI within your VPC and will be visible to you through your VPC flow logs. For more information, see [Fargate Task Networking \(p. 67\)](#).
- If you are using interface VPC endpoints, the following should be considered.
 - When using container images hosted with Amazon ECR, both the **com.amazonaws.region.ecr.dkr** and **com.amazonaws.region.ecr.api** Amazon ECR VPC endpoints as well as the Amazon S3 gateway endpoint are required. For more information, see [Amazon ECR interface VPC endpoints \(AWS PrivateLink\)](#) in the *Amazon Elastic Container Registry User Guide*.
 - When using a task definition that references Secrets Manager secrets to retrieve sensitive data for your containers, you must create the interface VPC endpoints for Secrets Manager. For more information, see [Using Secrets Manager with VPC Endpoints](#) in the *AWS Secrets Manager User Guide*.
 - When using a task definition that references Systems Manager Parameter Store parameters to retrieve sensitive data for your containers, you must create the interface VPC endpoints for Systems Manager. For more information, see [Using Systems Manager with VPC endpoints](#) in the *AWS Systems Manager User Guide*.
- Ensure that the security group in the Elastic Network Interface (ENI) associated with your task has the security group rules created to allow traffic between the task and the VPC endpoints you are using.

Amazon ECS Clusters

An Amazon ECS cluster is a logical grouping of tasks or services. If you are using capacity providers, a cluster is also a logical grouping of capacity providers. When you first use Amazon ECS, a default cluster is created for you, but you can create multiple clusters in an account to keep your resources separate.

Topics

- [Cluster Concepts \(p. 17\)](#)
- [Creating a Cluster \(p. 18\)](#)
- [Amazon ECS Cluster Capacity Providers \(p. 18\)](#)
- [Using AWS Fargate Capacity Providers \(p. 20\)](#)
- [Updating Cluster Settings \(p. 23\)](#)
- [Deleting a Cluster \(p. 24\)](#)

Cluster Concepts

The following are general concepts about Amazon ECS clusters.

- Clusters are Region-specific.
- The following are the possible states that a cluster can be in.

ACTIVE

The cluster is ready to accept tasks and, if applicable, you can register container instances with the cluster.

PROVISIONING

The cluster has capacity providers associated with it and the resources needed for the capacity provider are being created.

DEPROVISIONING

The cluster has capacity providers associated with it and the resources needed for the capacity provider are being deleted.

FAILED

The cluster has capacity providers associated with it and the resources needed for the capacity provider have failed to create.

INACTIVE

The cluster has been deleted. Clusters with an `INACTIVE` status may remain discoverable in your account for a period of time. However, this behavior is subject to change in the future, so you should not rely on `INACTIVE` clusters persisting.

- A cluster may contain a mix of tasks using either the Fargate or EC2 launch types. For more information about launch types, see [Amazon ECS Launch Types \(p. 60\)](#).
- A cluster may contain a mix of both Auto Scaling group capacity providers and Fargate capacity providers, however when specifying a capacity provider strategy they may only contain one or the other but not both. For more information, see [Amazon ECS Cluster Capacity Providers \(p. 18\)](#).
- Custom IAM policies may be created to allow or restrict user access to specific clusters. For more information, see the [Cluster Examples \(p. 215\)](#) section in [Amazon Elastic Container Service Identity-Based Policy Examples \(p. 210\)](#).

- Clusters with Fargate tasks can be scaled using Service Auto Scaling. For more information, see [Service Auto Scaling](#) (p. 165).

Creating a Cluster

You can create an Amazon ECS cluster using the AWS Management Console, as described in this topic. Before you begin, be sure that you've completed the steps in [Setting Up with Amazon ECS](#) (p. 3).

The console cluster creation wizard provides a simple way to create the resources that are needed by an Amazon ECS cluster by creating a AWS CloudFormation stack. It also lets you customize several common cluster configuration options. However, the wizard does not allow you to customize every resource option. If your requirements extend beyond what is supported in this wizard, consider using our reference architecture at <https://github.com/aws-labs/ecs-refarch-cloudformation>.

If you add or modify the underlying cluster resources directly after they are created by the wizard you may receive an error when attempting to delete the cluster. AWS CloudFormation refers to this as *stack drift*. For more information on detecting drift on an existing AWS CloudFormation stack, see [Detect Drift on an Entire CloudFormation Stack](#) in the *AWS CloudFormation User Guide*.

To create a cluster

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. From the navigation bar, select the Region to use.
3. In the navigation pane, choose **Clusters**.
4. On the **Clusters** page, choose **Create Cluster**.
5. For **Select cluster compatibility**, choose **Networking only**, then choose **Next Step**.
Important
The `FARGATE` and `FARGATE_SPOT` capacity providers will be automatically associated with the cluster. For more information, see [Using AWS Fargate Capacity Providers](#) (p. 20).
6. On the **Configure cluster** page, enter a **Cluster name**. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
7. In the **Networking** section, configure the VPC for your cluster. You can keep the default settings, or you can modify these settings with the following steps.
 - a. (Optional) If you choose to create a new VPC, for **CIDR Block**, select a CIDR block for your VPC. For more information, see [Your VPC and Subnets](#) in the *Amazon VPC User Guide*.
 - b. For **Subnets**, select the subnets to use for your VPC. You can keep the default settings or you can modify them to meet your needs.
8. In the **CloudWatch Container Insights** section, choose whether to enable Container Insights for the cluster. For more information, see [Amazon ECS CloudWatch Container Insights](#) (p. 197).
9. Choose **Create**.

Using the EC2 Linux + Networking or EC2 Windows + Networking template

Amazon ECS Cluster Capacity Providers

Amazon ECS cluster capacity providers determine the infrastructure to use for your tasks. Each cluster has one or more capacity providers and an optional default capacity provider strategy. The capacity provider strategy determines how the tasks are spread across the capacity providers. When you run a

task or create a service, you may either use the cluster's default capacity provider strategy or specify a capacity provider strategy that overrides the cluster's default strategy.

Cluster Capacity Provider Concepts

Cluster capacity providers consist of the following components.

Capacity provider

A *capacity provider* is used in association with a cluster to determine the infrastructure that a task runs on.

For Amazon ECS on AWS Fargate users, the `FARGATE` and `FARGATE_SPOT` capacity providers are provided automatically. For more information, see [Using AWS Fargate Capacity Providers \(p. 20\)](#).

One or more capacity providers are specified in a capacity provider strategy, which is then associated with a cluster.

Capacity provider strategy

A *capacity provider strategy* gives you control over how your tasks use one or more capacity providers.

When you run a task or create a service, you specify a capacity provider strategy. A capacity provider strategy consists of one or more capacity providers with an optional *base* and *weight* specified for each provider.

The *base* value designates how many tasks, at a minimum, to run on the specified capacity provider. Only one capacity provider in a capacity provider strategy can have a base defined.

The *weight* value designates the relative percentage of the total number of launched tasks that should use the specified capacity provider. For example, if you have a strategy that contains two capacity providers, and both have a weight of 1, then when the base is satisfied, the tasks will be split evenly across the two capacity providers. Using that same logic, if you specify a weight of 1 for *capacityProviderA* and a weight of 4 for *capacityProviderB*, then for every one task that is run using *capacityProviderA*, four tasks would use *capacityProviderB*.

Default capacity provider strategy

A *default capacity provider strategy* is associated with each Amazon ECS cluster. This determines the capacity provider strategy the cluster will use if no other capacity provider strategy or launch type is specified when running a task or creating a service.

Cluster Capacity Provider Considerations

The following should be considered when using cluster capacity providers:

- When you specify a capacity provider strategy, the number of capacity providers that can be specified is limited to six.
- A cluster may contain a mix of both Auto Scaling group capacity providers and Fargate capacity providers, however when specifying a capacity provider strategy they may only contain one or the other but not both.
- A cluster may contain a mix of tasks and services using both capacity providers and launch types. A service may also be updated to use a capacity provider strategy rather than a launch type, however you must force a new deployment when doing so.
- When you specify a capacity provider strategy, the *base* value is only supported when running tasks. When creating a service, the capacity provider strategy *base* parameter is not supported.
- When using managed termination protection, managed scaling must also be used otherwise managed termination protection will not work.

- Using cluster capacity providers is not supported when using the blue/green deployment type for your services.
- Using cluster capacity providers is not supported when using Classic Load Balancers for your services.

Using AWS Fargate Capacity Providers

Amazon ECS cluster capacity providers enable you to use both Fargate and Fargate Spot capacity with your Amazon ECS tasks. For more information about cluster capacity providers, see [Amazon ECS Cluster Capacity Providers](#) (p. 18).

With Fargate Spot you can run interruption tolerant Amazon ECS tasks at a discounted rate compared to the Fargate price. Fargate Spot runs tasks on spare compute capacity. When AWS needs the capacity back, your tasks will be interrupted with a two-minute warning. This is described in further detail below.

Topics

- [Fargate Capacity Provider Considerations](#) (p. 20)
- [Handling Fargate Spot Termination Notices](#) (p. 20)
- [Creating a New Cluster That Uses Fargate Capacity Providers](#) (p. 21)
- [Adding Fargate Capacity Providers To An Existing Cluster](#) (p. 22)
- [Running Tasks Using a Fargate Capacity Provider](#) (p. 23)

Fargate Capacity Provider Considerations

The following should be considered when using Fargate capacity providers.

- The Fargate and Fargate Spot capacity providers do not need to be created. They are available to all accounts and only need to be associated with a cluster to be available for use.
- When a new cluster is created using the Amazon ECS console along with the **Networking only** cluster template, the `FARGATE` and `FARGATE_SPOT` capacity providers are associated with the new cluster automatically.
- To add the `FARGATE` and `FARGATE_SPOT` capacity providers to an existing cluster, you must use the AWS CLI or API. For more information, see [Adding Fargate Capacity Providers To An Existing Cluster](#) (p. 22).
- Using Fargate Spot requires that your task use platform version 1.3.0 or later. For more information, see [AWS Fargate platform versions](#) (p. 14).
- When tasks using the Fargate and Fargate Spot capacity providers are stopped, a task state change event is sent to Amazon EventBridge. The stopped reason describes the cause. For more information, see [Task State Change Events](#) (p. 191).
- A cluster may contain a mix of Fargate and Auto Scaling group capacity providers, however a capacity provider strategy may only contain either Fargate or Auto Scaling group capacity providers, but not both. For more information, see [Auto Scaling Group Capacity Providers](#) in the *Amazon Elastic Container Service Developer Guide*.

Handling Fargate Spot Termination Notices

When tasks using Fargate Spot capacity are stopped due to a Spot interruption, a two-minute warning is sent before a task is stopped. The warning is sent as a task state change event to Amazon EventBridge and a SIGTERM signal to the running task. When using Fargate Spot as part of a service, the service scheduler will receive the interruption signal and attempt to launch additional tasks on Fargate Spot if capacity is available.

To ensure that your containers exit gracefully before the task stops, the following can be configured:

- A `stopTimeout` value of 120 seconds or less can be specified in the container definition that the task is using. Specifying a `stopTimeout` value gives you time between the moment the task state change event is received and the point at which the container is forcefully stopped. For more information, see [Container Timeouts \(p. 52\)](#).
- The `SIGTERM` signal must be received from within the container to perform any cleanup actions.

The following is a snippet of a task state change event displaying the stopped reason and stop code for a Fargate Spot interruption.

```
{
  "version": "0",
  "id": "9bcdac79-b31f-4d3d-9410-fbd727c29fab",
  "detail-type": "ECS Task State Change",
  "source": "aws.ecs",
  "account": "111122223333",
  "resources": [
    "arn:aws:ecs:us-east-1:111122223333:task/b99d40b3-5176-4f71-9a52-9dbd6f1cebef"
  ],
  "detail": {
    "clusterArn": "arn:aws:ecs:us-east-1:111122223333:cluster/default",
    "createdAt": "2016-12-06T16:41:05.702Z",
    "desiredStatus": "STOPPED",
    "lastStatus": "RUNNING",
    "stoppedReason": "Your Spot Task was interrupted.",
    "stopCode": "TerminationNotice",
    "taskArn": "arn:aws:ecs:us-east-1:111122223333:task/
b99d40b3-5176-4f71-9a52-9dbd6fEXAMPLE",
    ...
  }
}
```

The following is an event pattern that is used to create an EventBridge rule for Amazon ECS task state change events. You can optionally specify a cluster in the `detail` field to receive task state change events for. For more information, see [Creating an EventBridge Rule](#) in the *Amazon EventBridge User Guide*.

```
{
  "source": [
    "aws.ecs"
  ],
  "detail-type": [
    "ECS Task State Change"
  ],
  "detail": {
    "clusterArn": [
      "arn:aws:ecs:us-west-2:111122223333:cluster/default"
    ]
  }
}
```

Creating a New Cluster That Uses Fargate Capacity Providers

When a new Amazon ECS cluster is created, you specify one or more capacity providers to associate with the cluster. The associated capacity providers determine the infrastructure to run your tasks on.

When using the AWS Management Console, the `FARGATE` and `FARGATE_SPOT` capacity providers are associated with the cluster automatically when using the **Networking only** cluster template. For more information, see [Creating a Cluster \(p. 18\)](#).

To create an Amazon ECS cluster using Fargate capacity providers (AWS CLI)

Use the following command to create a new cluster and associate both the Fargate and Fargate Spot capacity providers with it.

- [create-cluster](#) (AWS CLI)

```
aws ecs create-cluster \
  --cluster-name FargateCluster \
  --capacity-providers FARGATE FARGATE_SPOT \
  --region us-west-2
```

Adding Fargate Capacity Providers To An Existing Cluster

You can update the pool of available capacity providers for an existing Amazon ECS cluster by using the `PutClusterCapacityProviders` API.

Adding either the Fargate or Fargate Spot capacity providers to an existing cluster is not supported in the AWS Management Console. You must either create a new Fargate cluster in the console or add the Fargate or Fargate Spot capacity providers to the existing cluster using the Amazon ECS API or AWS CLI.

To add the Fargate capacity providers to an existing cluster (AWS CLI)

Use the following command to add the Fargate and Fargate Spot capacity providers to an existing cluster. If the specified cluster has existing capacity providers associated with it, you must specify all existing capacity providers in addition to any new ones you want to add. Any existing capacity providers associated with a cluster that are omitted from a `PutClusterCapacityProviders` API call will be disassociated from the cluster. You can only disassociate an existing capacity provider from a cluster if it's not being used by any existing tasks. These same rules apply to the cluster's default capacity provider strategy. If the cluster has an existing default capacity provider strategy defined, it must be included in the `PutClusterCapacityProviders` API call. Otherwise, it will be overwritten.

- [put-cluster-capacity-providers](#) (AWS CLI)

```
aws ecs put-cluster-capacity-providers \
  --cluster FargateCluster \
  --capacity-providers FARGATE
FARGATE_SPOT existing_capacity_provider1 existing_capacity_provider2 \
  --default-capacity-provider-strategy existing_default_capacity_provider_strategy \
  --region us-west-2
```

Running Tasks Using a Fargate Capacity Provider

You can run a task or create a service using either the Fargate or Fargate Spot capacity providers by specifying a capacity provider strategy. If no capacity provider strategy is provided, the cluster's default capacity provider strategy is used.

Running a task using the Fargate or Fargate Spot capacity providers is supported in the AWS Management Console. You must add the Fargate or Fargate Spot capacity providers to cluster's default capacity provider strategy if using the AWS Management Console. When using the Amazon ECS API or AWS CLI you can specify either a capacity provider strategy or use the cluster's default capacity provider strategy.

To run a task using a Fargate capacity provider (AWS CLI)

Use the following command to run a task using the Fargate and Fargate Spot capacity providers.

- `run-task` (AWS CLI)

```
aws ecs run-task \
  --capacity-provider-strategy capacityProvider=FARGATE,weight=1
  capacityProvider=FARGATE_SPOT,weight=1 \
  --cluster FargateCluster \
  --task-definition task-def-family:revision \
  --network-configuration
  "awsvpcConfiguration={subnets=[string,string],securityGroups=[string,string],assignPublicIp=string}" \
  --count integer \
  --region us-west-2
```

Create a service using a Fargate capacity provider (AWS CLI)

Use the following command to create a service using the Fargate and Fargate Spot capacity providers.

- `create-service` (AWS CLI)

```
aws ecs create-service \
  --capacity-provider-strategy capacityProvider=FARGATE,weight=1
  capacityProvider=FARGATE_SPOT,weight=1 \
  --cluster FargateCluster \
  --service-name FargateService \
  --task-definition task-def-family:revision \
  --network-configuration
  "awsvpcConfiguration={subnets=[string,string],securityGroups=[string,string],assignPublicIp=string}" \
  --desired-count integer \
  --region us-west-2
```

Updating Cluster Settings

Cluster settings enable you to configure parameters for your existing Amazon ECS clusters. You can update cluster settings using the Amazon ECS API, AWS CLI or SDKs. Currently, the only supported cluster setting is `containerInsights`, which allows you to enable or disable CloudWatch Container Insights for an existing cluster. To enable CloudWatch Container Insights for a new cluster, that can be done in the AWS Management Console during cluster creation. For more information, see [Creating a Cluster](#) (p. 18).

Important

Currently, if you delete an existing cluster that does not have Container Insights enabled and then create a new cluster with the same name with Container Insights enabled, Container Insights will not actually be enabled. If you want to preserve the same name for your existing cluster and enable Container Insights, you must wait 7 days before you can re-create it.

To update the settings for a cluster using the command line

Use one of the following commands to update the setting for a cluster.

- `update-cluster-settings` (AWS CLI)

```
aws ecs update-cluster-settings --cluster cluster_name_or_arn --settings  
name=containerInsights,value=enabled/disabled --region us-east-1
```

Deleting a Cluster

If you are finished using a cluster, you can delete it. Once deleted, the cluster will transition to the `INACTIVE` state. Clusters with an `INACTIVE` status may remain discoverable in your account for a period of time. However, this behavior is subject to change in the future, so you should not rely on `INACTIVE` clusters persisting.

When you delete a cluster in the Amazon ECS console, the associated resources that are deleted with it vary depending on how the cluster was created. [Step 5 \(p. 24\)](#) of the following procedure changes based on that condition.

If your cluster was created with the AWS Management Console then the AWS CloudFormation stack that was created for your cluster is also deleted when you delete your cluster. If you have added or modified the underlying cluster resources you may receive an error when attempting to delete the cluster. AWS CloudFormation refers to this as *stack drift*. For more information on detecting drift on an existing AWS CloudFormation stack, see [Detect Drift on an Entire CloudFormation Stack](#) in the *AWS CloudFormation User Guide*.

To delete a cluster

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. From the navigation bar, select the Region to use.
3. In the navigation pane, choose **Clusters**.
4. On the **Clusters** page, select the cluster to delete.
5. Choose **Delete Cluster**. You see a confirmation prompt.

Amazon ECS Task Definitions

A task definition is required to run Docker containers in Amazon ECS. Some of the parameters you can specify in a task definition include:

- The Docker image to use with each container in your task
- How much CPU and memory to use with each task
- The launch type to use, which determines the infrastructure on which your tasks are hosted
- The Docker networking mode to use for the containers in your task
- The logging configuration to use for your tasks
- Whether the task should continue to run if the container finishes or fails
- The command the container should run when it is started
- Any data volumes that should be used with the containers in the task
- The IAM role that your tasks should use

Your entire application stack does not need to exist on a single task definition, and in most cases it should not. Your application can span multiple task definitions by combining related containers into their own task definitions, each representing a single component. For more information, see [Application Architecture](#) (p. 29).

Topics

- [Task Definition Considerations](#) (p. 25)
- [Application Architecture](#) (p. 29)
- [Creating a Task Definition](#) (p. 30)
- [Task definition parameters](#) (p. 35)
- [Amazon ECS Launch Types](#) (p. 60)
- [Using Data Volumes in Tasks](#) (p. 63)
- [Fargate Task Networking](#) (p. 67)
- [Using the awslogs Log Driver](#) (p. 69)
- [Custom Log Routing](#) (p. 75)
- [Private Registry Authentication for Tasks](#) (p. 85)
- [Specifying Sensitive Data](#) (p. 87)
- [Example Task Definitions](#) (p. 97)
- [Updating a Task Definition](#) (p. 101)
- [Deregistering Task Definition Revisions](#) (p. 101)

Task Definition Considerations

Tasks that use the Fargate launch type do not support all of the Amazon ECS task definition parameters that are available. Some parameters are not supported at all, and others behave differently for Fargate tasks.

The following task definition parameters are not valid in Fargate tasks:

- `devices`
- `disableNetworking`
- `dnsSearchDomains`

- `dnsServers`
- `dockerSecurityOptions`
- `dockerVolumeConfiguration`
- `extraHosts`
- `host`
- `hostname`
- `links`
- `placementConstraints` — By default, Fargate tasks are spread across Availability Zones.
- `privileged`
- `sharedMemorySize`
- `tmpfs`

Important

When any task definition parameter is not supported, it is assumed that any subflags for that parameter are not supported either.

The following task definition parameters behave differently for Fargate tasks:

- When using `logConfiguration`, the supported log drivers for Fargate tasks are the `awslogs`, `splunk` and `awsfirelens` log drivers.
- When using `linuxParameters`, for capabilities the `drop` parameter can be used but the `add` parameter is not supported.
- The `healthCheck` parameter is only supported for Fargate tasks using platform version 1.1.0 or later.
- If you use the `portMappings` parameter, you should only specify the `containerPort`. The `hostPort` can either be left blank or be set to the same value as the `containerPort`.

To ensure that your task definition validates for use with the Fargate launch type, you can specify the following when you register the task definition:

- In the AWS Management Console, for the **Requires Compatibilities** field, specify **FARGATE**.
- In the AWS CLI, for the `--requires-compatibilities` option, specify `FARGATE`.
- In the API, specify the `requiresCompatibilities` flag.

Network Mode

Fargate task definitions require that the network mode is set to `awsvpc`. The `awsvpc` network mode provides each task with its own elastic network interface. A network configuration is also required when creating a service or manually running tasks. For more information, see [Fargate Task Networking](#) in the *Amazon Elastic Container Service User Guide for AWS Fargate*.

Task CPU and Memory

Fargate task definitions require that you specify CPU and memory at the task level. Although you can also specify CPU and memory at the container level for Fargate tasks, this is optional. Most use cases are satisfied by only specifying these resources at the task level. The table below shows the valid combinations of task-level CPU and memory.

CPU value	Memory value
256 (.25 vCPU)	0.5 GB, 1 GB, 2 GB

CPU value	Memory value
512 (.5 vCPU)	1 GB, 2 GB, 3 GB, 4 GB
1024 (1 vCPU)	2 GB, 3 GB, 4 GB, 5 GB, 6 GB, 7 GB, 8 GB
2048 (2 vCPU)	Between 4 GB and 16 GB in 1-GB increments
4096 (4 vCPU)	Between 8 GB and 30 GB in 1-GB increments

Logging

Fargate task definitions only support the `awslogs`, `splunk` and `awsfirelens` log drivers for the log configuration. The following shows a snippet of a task definition where the `awslogs` log driver is configured:

```
"logConfiguration": {
  "logDriver": "awslogs",
  "options": {
    "awslogs-group" : "/ecs/fargate-task-definition",
    "awslogs-region": "us-east-1",
    "awslogs-stream-prefix": "ecs"
  }
}
```

For more information about using the `awslogs` log driver in task definitions to send your container logs to CloudWatch Logs, see [Using the awslogs Log Driver \(p. 69\)](#).

For more information about using the `awsfirelens` log driver in a task definition, see [Custom Log Routing \(p. 75\)](#).

Amazon ECS Task Execution IAM Role

There is an optional task execution IAM role that you can specify with Fargate to allow your Fargate tasks to make API calls to Amazon ECR. The API calls pull container images as well as call CloudWatch to store container application logs. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

Example Task Definition

The following is an example task definition using the Fargate launch type that sets up a web server:

```
{
  "containerDefinitions": [
    {
      "command": [
        "/bin/sh -c \"echo '<html> <head> <title>Amazon ECS Sample App</title>
<style>body {margin-top: 40px; background-color: #333;} </style> </head><body> <div
style=color:white;text-align:center> <h1>Amazon ECS Sample App</h1> <h2>Congratulations!
</h2> <p>Your application is now running on a container in Amazon ECS.</p> </div></body></
html>' > /usr/local/apache2/htdocs/index.html && httpd-foreground\""
      ],
      "entryPoint": [
        "sh",
        "-c"
      ],
      "essential": true,
      "image": "httpd:2.4",
      "logConfiguration": {
        "logDriver": "awslogs",

```

```

        "options": {
            "awslogs-group" : "/ecs/fargate-task-definition",
            "awslogs-region": "us-east-1",
            "awslogs-stream-prefix": "ecs"
        },
        "name": "sample-fargate-app",
        "portMappings": [
            {
                "containerPort": 80,
                "hostPort": 80,
                "protocol": "tcp"
            }
        ]
    },
    "cpu": "256",
    "executionRoleArn": "arn:aws:iam::012345678910:role/ecsTaskExecutionRole",
    "family": "fargate-task-definition",
    "memory": "512",
    "networkMode": "awsvpc",
    "requiresCompatibilities": [
        "FARGATE"
    ]
}

```

Task Storage

For Fargate tasks, the following storage types are supported:

- Amazon EFS volumes for persistent storage. For more information, see [Amazon EFS Volumes \(p. 65\)](#).
- Ephemeral storage for nonpersistent storage.

When provisioned, each Amazon ECS task on Fargate receives the following ephemeral storage. The ephemeral storage configuration depends on which platform version the task is using. After a Fargate task stops, the ephemeral storage is deleted. For more information about Amazon ECS default service limits, see [Amazon ECS Service Quotas \(p. 283\)](#).

Fargate tasks using platform version 1.4.0 or later

All Amazon ECS on Fargate tasks using platform version 1.4.0 or later receive a minimum of 20 GB of ephemeral storage.

For tasks using platform version 1.4.0 or later that are launched on May 28, 2020 or later, the ephemeral storage is encrypted with an AES-256 encryption algorithm using an AWS Fargate-managed encryption key.

Fargate tasks using platform version 1.3.0 or earlier

For Amazon ECS on Fargate tasks using platform version 1.3.0 or earlier, each task receives the following ephemeral storage.

- 10 GB of Docker layer storage
- An additional 4 GB for volume mounts. This can be mounted and shared among containers using the `volumes`, `mountPoints` and `volumesFrom` parameters in the task definition.

Note

The `host` and `sourcePath` parameters are not supported for Fargate tasks.

Example task definition

In this example, you have two application containers that need to access the same scratch file storage location.

To provide nonpersistent empty storage for containers in a Fargate task

1. In the task definition `volumes` section, define a volume with the name `application_scratch`.

```
"volumes": [  
  {  
    "name": "application_scratch",  
    "host": {}  
  }  
]
```

2. In the `containerDefinitions` section, create the application container definitions so they mount the nonpersistent storage.

```
"containerDefinitions": [  
  {  
    "name": "application1",  
    "image": "my-repo/application",  
    "cpu": 100,  
    "memory": 100,  
    "essential": true,  
    "mountPoints": [  
      {  
        "sourceVolume": "application_scratch",  
        "containerPath": "/var/scratch"  
      }  
    ]  
  },  
  {  
    "name": "application2",  
    "image": "my-repo/application",  
    "cpu": 100,  
    "memory": 100,  
    "essential": true,  
    "mountPoints": [  
      {  
        "sourceVolume": "application_scratch",  
        "containerPath": "/var/scratch"  
      }  
    ]  
  }  
]
```

Application Architecture

How you architect your application on Amazon ECS depends on several factors, with the launch type you are using being a key differentiator. We give the following guidance which should assist in the process.

Using the Fargate Launch Type

When architecting your application using the Fargate launch type for your tasks, the main question is when should you put multiple containers into the same task definition versus deploying containers separately in multiple task definitions.

You should put multiple containers in the same task definition if:

- Containers share a common lifecycle (that is, they should be launched and terminated together).
- Containers are required to be run on the same underlying host (that is, one container references the other on a localhost port).
- You want your containers to share resources.
- Your containers share data volumes.

Otherwise, you should define your containers in separate tasks definitions so that you can scale, provision, and deprovision them separately.

Creating a Task Definition

Before you can run Docker containers on Amazon ECS, you must create a task definition. You can define multiple containers and data volumes in a task definition. For more information about the parameters available in a task definition, see [Task definition parameters \(p. 35\)](#).

To create a new task definition

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation pane, choose **Task Definitions**, **Create new Task Definition**.
3. On the **Select launch type compatibilities** page, choose **FARGATE**, **Next step**.

Note

The Fargate launch type is not compatible with Windows containers.

4. (Optional) If you have a JSON representation of your task definition, complete the following steps:
 - a. On the **Configure task and container definitions** page, scroll to the bottom of the page and choose **Configure via JSON**.
 - b. Paste your task definition JSON into the text area and choose **Save**.
 - c. Verify your information and choose **Create**.

Scroll to the bottom of the page and choose **Configure via JSON**.

5. For **Task Definition Name**, type a name for your task definition. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
6. For **Task execution IAM role**, either select your task execution role or choose **Create new role** so that the console can create one for you. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).
7. For **Task size**, choose a value for **Task memory (GB)** and **Task CPU (vCPU)**. The table below shows the valid combinations.

CPU value	Memory value
256 (.25 vCPU)	512 MB, 1 GB, 2 GB
512 (.5 vCPU)	1 GB, 2 GB, 3 GB, 4 GB
1024 (1 vCPU)	2 GB, 3 GB, 4 GB, 5 GB, 6 GB, 7 GB, 8 GB
2048 (2 vCPU)	Between 4 GB and 16 GB in 1 GB increments

CPU value	Memory value
4096 (4 vCPU)	Between 8 GB and 30 GB in 1 GB increments

8. For each container in your task definition, complete the following steps:
 - a. Choose **Add container**.
 - b. Fill out each required field and any optional fields to use in your container definitions. More container definition parameters are available in the **Advanced container configuration** menu. For more information, see [Task definition parameters \(p. 35\)](#).
 - c. Choose **Add** to add your container to the task definition.
9. (Optional) For **Service Integration**, to configure the parameters for App Mesh integration choose **Enable App Mesh integration** and then do the following:
 - a. For **Application container name**, choose the container name to use for the App Mesh application. This container must already be defined within the task definition.
 - b. For **Envoy image**, enter 840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod.
 - c. For **Mesh name**, choose the App Mesh service mesh to use. This must already be created in order for it to show up. For more information, see [Service Meshes](#) in the *AWS App Mesh User Guide*.
 - d. For **Virtual node name**, choose the App Mesh virtual node to use. This must already be created in order for it to show up. For more information, see [Virtual Nodes](#) in the *AWS App Mesh User Guide*.
 - e. For **Virtual node port**, this will be pre-populated with the listener port set on the virtual node.
 - f. Choose **Apply, Confirm**. This will create a new Envoy proxy container to the task definition, as well as the settings to support it. It will then pre-populate the App Mesh proxy configuration settings for the next step.
10. (Optional) For **Proxy Configuration**, verify all of the pre-populated values. For more information on these fields, see the JSON tab in [Update Services](#).
11. (Optional) For **Log Router Integration**, you can add a custom log routing configuration. Choose **Enable FireLens integration** and then do the following:
 - a. For **Type**, choose the log router type to use.
 - b. For **Image**, type the image URI for your log router container. If you chose the `fluentbit` log router type, the **Image** field prepopulates with the AWS for Fluent Bit image. For more information, see [Using the AWS for Fluent Bit Image \(p. 77\)](#).
 - c. Choose **Apply**. This creates a new log router container to the task definition named `log_router`, and applies the settings to support it. If you make changes to the log router integration fields, choose **Apply** again to update the FireLens container.
12. (Optional) To define data volumes for your task, choose **Add volume**. For more information, see [Using Data Volumes in Tasks \(p. 63\)](#).
 - For **Name**, type a name for your volume. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
13. In the **Tags** section, specify the key and value for each tag to associate with the task definition. For more information, see [Tagging Your Amazon ECS Resources](#).
14. Choose **Create**.

Task Definition Template

An empty task definition template is shown below. You can use this template to create your task definition, which can then be pasted into the console JSON input area or saved to a file and

used with the AWS CLI `--cli-input-json` option. For more information, see [Task definition parameters \(p. 35\)](#).

```
{
  "family": "",
  "taskRoleArn": "",
  "executionRoleArn": "",
  "networkMode": "awsvpc",
  "containerDefinitions": [
    {
      "name": "",
      "image": "",
      "repositoryCredentials": {
        "credentialsParameter": ""
      },
      "cpu": 0,
      "memory": 0,
      "memoryReservation": 0,
      "links": [
        ""
      ],
      "portMappings": [
        {
          "containerPort": 0,
          "hostPort": 0,
          "protocol": "tcp"
        }
      ],
      "essential": true,
      "entryPoint": [
        ""
      ],
      "command": [
        ""
      ],
      "environment": [
        {
          "name": "",
          "value": ""
        }
      ],
      "environmentFiles": [
        {
          "value": "",
          "type": "s3"
        }
      ],
      "mountPoints": [
        {
          "sourceVolume": "",
          "containerPath": "",
          "readOnly": true
        }
      ],
      "volumesFrom": [
        {
          "sourceContainer": "",
          "readOnly": true
        }
      ],
      "linuxParameters": {
        "capabilities": {
          "add": [
            ""
          ]
        }
      }
    }
  ]
}
```



```

        "drop": [
            ""
        ]
    },
    "devices": [
        {
            "hostPath": "",
            "containerPath": "",
            "permissions": [
                "read"
            ]
        }
    ],
    "initProcessEnabled": true,
    "sharedMemorySize": 0,
    "tmpfs": [
        {
            "containerPath": "",
            "size": 0,
            "mountOptions": [
                ""
            ]
        }
    ],
    "maxSwap": 0,
    "swappiness": 0
},
"secrets": [
    {
        "name": "",
        "valueFrom": ""
    }
],
"dependsOn": [
    {
        "containerName": "",
        "condition": "HEALTHY"
    }
],
"startTimeout": 0,
"stopTimeout": 0,
"hostname": "",
"user": "",
"workingDirectory": "",
"disableNetworking": true,
"privileged": true,
"readonlyRootFilesystem": true,
"dnsServers": [
    ""
],
"dnsSearchDomains": [
    ""
],
"extraHosts": [
    {
        "hostname": "",
        "ipAddress": ""
    }
],
"dockerSecurityOptions": [
    ""
],
"interactive": true,
"pseudoTerminal": true,
"dockerLabels": {
    "KeyName": ""
}

```

```

    },
    "ulimits": [
      {
        "name": "msgqueue",
        "softLimit": 0,
        "hardLimit": 0
      }
    ],
    "logConfiguration": {
      "logDriver": "awslogs",
      "options": {
        "KeyName": ""
      },
      "secretOptions": [
        {
          "name": "",
          "valueFrom": ""
        }
      ]
    },
    "healthCheck": {
      "command": [
        ""
      ],
      "interval": 0,
      "timeout": 0,
      "retries": 0,
      "startPeriod": 0
    },
    "systemControls": [
      {
        "namespace": "",
        "value": ""
      }
    ],
    "resourceRequirements": [
      {
        "value": "",
        "type": "GPU"
      }
    ],
    "firelensConfiguration": {
      "type": "fluentd",
      "options": {
        "KeyName": ""
      }
    }
  }
],
"volumes": [
  {
    "name": "",
    "host": {
      "sourcePath": ""
    },
    "dockerVolumeConfiguration": {
      "scope": "task",
      "autoprovision": true,
      "driver": "",
      "driverOpts": {
        "KeyName": ""
      },
      "labels": {
        "KeyName": ""
      }
    }
  },

```

```

        "efsVolumeConfiguration": {
            "fileSystemId": "",
            "rootDirectory": "",
            "transitEncryption": "ENABLED",
            "transitEncryptionPort": 0,
            "authorizationConfig": {
                "accessPointId": "",
                "iam": "ENABLED"
            }
        }
    },
    "placementConstraints": [
        {
            "type": "memberOf",
            "expression": ""
        }
    ],
    "requiresCompatibilities": [
        "EC2"
    ],
    "cpu": "",
    "memory": "",
    "tags": [
        {
            "key": "",
            "value": ""
        }
    ],
    "pidMode": "task",
    "ipcMode": "none",
    "proxyConfiguration": {
        "type": "APPMESH",
        "containerName": "",
        "properties": [
            {
                "name": "",
                "value": ""
            }
        ]
    },
    "inferenceAccelerators": [
        {
            "deviceName": "",
            "deviceType": ""
        }
    ]
}

```

You can generate this task definition template using the following AWS CLI command:

```
aws ecs register-task-definition --generate-cli-skeleton
```

Task definition parameters

Task definitions are split into separate parts: the task family, the IAM task role, the network mode, container definitions, volumes, task placement constraints, and launch types. The family and container definitions are required in a task definition, while task role, network mode, volumes, task placement constraints, and launch type are optional.

The following are more detailed descriptions for each task definition parameter.

Family

`family`

Type: string

Required: yes

When you register a task definition, you give it a family, which is similar to a name for multiple versions of the task definition, specified with a revision number. The first task definition that is registered into a particular family is given a revision of 1, and any task definitions registered after that are given a sequential revision number.

Task execution role

`executionRoleArn`

Type: string

Required: no

The Amazon Resource Name (ARN) of the task execution role that grants the Amazon ECS container agent permission to make AWS API calls on your behalf. The task execution IAM role is required depending on the requirements of your task. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

Network mode

`networkMode`

Type: string

Required: no

The Docker networking mode to use for the containers in the task. When using the Fargate launch type, the `awsvpc` network mode is required.

When the network mode is `awsvpc`, the task is allocated an elastic network interface, and you must specify a `NetworkConfiguration` when you create a service or run a task with the task definition. For more information, see [Fargate Task Networking](#) in the *Amazon Elastic Container Service User Guide for AWS Fargate*.

The `awsvpc` network mode offers the highest networking performance for containers because they use the Amazon EC2 network stack. Exposed container ports are mapped directly to the attached elastic network interface port, so you cannot take advantage of dynamic host port mappings.

Container Definitions

When you register a task definition, you must specify a list of container definitions that are passed to the Docker daemon on a container instance. The following parameters are allowed in a container definition.

Topics

- [Standard Container Definition Parameters \(p. 37\)](#)
- [Advanced Container Definition Parameters \(p. 40\)](#)

- [Other Container Definition Parameters \(p. 50\)](#)

Standard Container Definition Parameters

The following task definition parameters are either required or used in most container definitions.

Topics

- [Name \(p. 37\)](#)
- [Image \(p. 37\)](#)
- [Memory \(p. 38\)](#)
- [Port Mappings \(p. 38\)](#)

Name

name

Type: string

Required: yes

The name of a container. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed. If you are linking multiple containers together in a task definition, the `name` of one container can be entered in the `links` of another container to connect the containers.

Image

image

Type: string

Required: yes

The image used to start a container. This string is passed directly to the Docker daemon. Images in the Docker Hub registry are available by default. You can also specify other repositories with either `repository-url/image:tag` or `repository-url/image@digest`. Up to 255 letters (uppercase and lowercase), numbers, hyphens, underscores, colons, periods, forward slashes, and number signs are allowed. This parameter maps to `Image` in the [Create a container](#) section of the [Docker Remote API](#) and the `IMAGE` parameter of [docker run](#).

- When a new task starts, the Amazon ECS container agent pulls the latest version of the specified image and tag for the container to use. However, subsequent updates to a repository image are not propagated to already running tasks.
- Images in private registries are supported. For more information, see [Private Registry Authentication for Tasks \(p. 85\)](#).
- Images in Amazon ECR repositories can be specified by using either the full `registry/repository:tag` or `registry/repository@digest` naming convention. For example, `aws_account_id.dkr.ecr.region.amazonaws.com/my-web-app:latest` or `aws_account_id.dkr.ecr.region.amazonaws.com/my-web-app@sha256:94afdlf2e64d908bc90dbca0035a5b567EXAMPLE`.
- Images in official repositories on Docker Hub use a single name (for example, `ubuntu` or `mongo`).
- Images in other repositories on Docker Hub are qualified with an organization name (for example, `amazon/amazon-ecs-agent`).
- Images in other online repositories are qualified further by a domain name (for example, `quay.io/assemblyline/ubuntu`).

Memory

`memory`

Type: integer

Required: no

The amount (in MiB) of memory to present to the container. If your container attempts to exceed the memory specified here, the container is killed. The total amount of memory reserved for all containers within a task must be lower than the task `memory` value, if one is specified. This parameter maps to `Memory` in the [Create a container](#) section of the [Docker Remote API](#) and the `--memory` option to [docker run](#).

If using the Fargate launch type, this parameter is optional.

The Docker daemon reserves a minimum of 4 MiB of memory for a container, so you should not specify fewer than 4 MiB of memory for your containers.

`memoryReservation`

Type: integer

Required: no

The soft limit (in MiB) of memory to reserve for the container. When system memory is under contention, Docker attempts to keep the container memory to this soft limit; however, your container can consume more memory when needed, up to either the hard limit specified with the `memory` parameter (if applicable), or all of the available memory on the container instance, whichever comes first. This parameter maps to `MemoryReservation` in the [Create a container](#) section of the [Docker Remote API](#) and the `--memory-reservation` option to [docker run](#).

If a task-level memory value is not specified, you must specify a non-zero integer for one or both of `memory` or `memoryReservation` in a container definition. If you specify both, `memory` must be greater than `memoryReservation`. If you specify `memoryReservation`, then that value is subtracted from the available memory resources for the container instance on which the container is placed. Otherwise, the value of `memory` is used.

For example, if your container normally uses 128 MiB of memory, but occasionally bursts to 256 MiB of memory for short periods of time, you can set a `memoryReservation` of 128 MiB, and a `memory` hard limit of 300 MiB. This configuration would allow the container to only reserve 128 MiB of memory from the remaining resources on the container instance, but also allow the container to consume more memory resources when needed.

The Docker daemon reserves a minimum of 4 MiB of memory for a container, so you should not specify fewer than 4 MiB of memory for your containers.

Port Mappings

`portMappings`

Type: object array

Required: no

Port mappings allow containers to access ports on the host container instance to send or receive traffic.

For task definitions that use the `awsvpc` network mode, you should only specify the `containerPort`. The `hostPort` can be left blank or it must be the same value as the `containerPort`.

This parameter maps to `PortBindings` in the [Create a container](#) section of the [Docker Remote API](#) and the `--publish` option to [docker run](#). If the network mode of a task definition is set to `host`, then host ports must either be undefined or they must match the container port in the port mapping.

Note

After a task reaches the `RUNNING` status, manual and automatic host and container port assignments are visible in the following locations:

- Console: The **Network Bindings** section of a container description for a selected task.
- AWS CLI: The `networkBindings` section of the **describe-tasks** command output.
- API: The `DescribeTasks` response.

`containerPort`

Type: integer

Required: yes, when `portMappings` are used

The port number on the container that is bound to the user-specified or automatically assigned host port.

If using containers in a task with the Fargate launch type, exposed ports should be specified using `containerPort`.

If using containers in a task with the EC2 launch type and you specify a container port and not a host port, your container automatically receives a host port in the ephemeral port range. For more information, see `hostPort`. Port mappings that are automatically assigned in this way do not count toward the 100 reserved ports limit of a container instance.

`hostPort`

Type: integer

Required: no

The port number on the container instance to reserve for your container.

If using containers in a task with the Fargate launch type, the `hostPort` can either be left blank or be the same value as `containerPort`.

If using containers in a task with the EC2 launch type, you can specify a non-reserved host port for your container port mapping (this is referred to as *static* host port mapping), or you can omit the `hostPort` (or set it to 0) while specifying a `containerPort` and your container automatically receives a port (this is referred to as *dynamic* host port mapping) in the ephemeral port range for your container instance operating system and Docker version.

The default ephemeral port range Docker version 1.6.0 and later is listed on the instance under `/proc/sys/net/ipv4/ip_local_port_range`. If this kernel parameter is unavailable, the default ephemeral port range from 49153–65535 is used. Do not attempt to specify a host port in the ephemeral port range, as these are reserved for automatic assignment. In general, ports below 32768 are outside of the ephemeral port range.

The default reserved ports are 22 for SSH, the Docker ports 2375 and 2376, and the Amazon ECS container agent ports 51678–51680. Any host port that was previously user-specified for a running task is also reserved while the task is running (after a task stops, the host port is released). The current reserved ports are displayed in the `remainingResources` of **describe-container-instances** output, and a container instance may have up to 100 reserved ports at a time, including the default reserved ports. Automatically assigned ports do not count toward the 100 reserved ports limit.

`protocol`

Type: string

Required: no

The protocol used for the port mapping. Valid values are `tcp` and `udp`. The default is `tcp`.

If you are specifying a host port, use the following syntax:

```
"portMappings": [
  {
    "containerPort": integer,
    "hostPort": integer
  }
  ...
]
```

If you want an automatically assigned host port, use the following syntax:

```
"portMappings": [
  {
    "containerPort": integer
  }
  ...
]
```

Advanced Container Definition Parameters

The following advanced container definition parameters provide extended capabilities to the `docker run` command that is used to launch containers on your Amazon ECS container instances.

Topics

- [Health Check \(p. 40\)](#)
- [Environment \(p. 42\)](#)
- [Network Settings \(p. 45\)](#)
- [Storage and Logging \(p. 45\)](#)
- [Security \(p. 48\)](#)
- [Resource Limits \(p. 49\)](#)
- [Docker Labels \(p. 50\)](#)

Health Check

`healthCheck`

The container health check command and associated configuration parameters for the container. This parameter maps to `HealthCheck` in the [Create a container](#) section of the [Docker Remote API](#) and the `HEALTHCHECK` parameter of `docker run`.

Note

The Amazon ECS container agent only monitors and reports on the health checks specified in the task definition. Amazon ECS does not monitor Docker health checks that are embedded in a container image and not specified in the container definition. Health check parameters that are specified in a container definition override any Docker health checks that exist in the container image.

You can view the health status of both individual containers and a task with the `DescribeTasks` API operation or when viewing the task details in the console.

The following describes the possible `healthStatus` values for a container:

- **HEALTHY**—The container health check has passed successfully.
- **UNHEALTHY**—The container health check has failed.
- **UNKNOWN**—The container health check is being evaluated or there is no container health check defined.

The following describes the possible `healthStatus` values for a task. The container health check status of nonessential containers do not have an effect on the health status of a task.

- **HEALTHY**—All essential containers within the task have passed their health checks.
- **UNHEALTHY**—One or more essential containers have failed their health check.
- **UNKNOWN**—The essential containers within the task are still having their health checks evaluated or there are no container health checks defined.

If a task is run manually, and not as part of a service, the task will continue its lifecycle regardless of its health status. For tasks that are part of a service, if the task reports as unhealthy then the task will be stopped and the service scheduler will replace it.

The following are notes about container health check support:

- Container health checks are supported for Fargate tasks if you are using platform version 1.1.0 or later. For more information, see [AWS Fargate platform versions \(p. 14\)](#).
- Container health checks are not supported for tasks that are part of a service that is configured to use a Classic Load Balancer.

command

A string array representing the command that the container runs to determine if it is healthy. The string array can start with `CMD` to execute the command arguments directly, or `CMD-SHELL` to run the command with the container's default shell. If neither is specified, `CMD` is used by default.

When registering a task definition in the AWS Management Console, use a comma separated list of commands which will automatically converted to a string after the task definition is created. An example input for a health check could be:

```
CMD-SHELL, curl -f http://localhost/ || exit 1
```

When registering a task definition using the AWS Management Console JSON panel, the AWS CLI, or the APIs, you should enclose the list of commands in brackets. An example input for a health check could be:

```
[ "CMD-SHELL", "curl -f http://localhost/ || exit 1" ]
```

An exit code of 0 indicates success, and a non-zero exit code indicates failure. For more information, see `HealthCheck` in the [Create a container](#) section of the [Docker Remote API](#).

interval

The time period in seconds between each health check execution. You may specify between 5 and 300 seconds. The default value is 30 seconds.

timeout

The time period in seconds to wait for a health check to succeed before it is considered a failure. You may specify between 2 and 60 seconds. The default value is 5 seconds.

`retries`

The number of times to retry a failed health check before the container is considered unhealthy. You may specify between 1 and 10 retries. The default value is three retries.

`startPeriod`

The optional grace period within which to provide containers time to bootstrap before failed health checks count towards the maximum number of retries. You may specify between 0 and 300 seconds. The `startPeriod` is disabled by default.

Environment

`cpu`

Type: integer

Required: no

The number of `cpu` units the Amazon ECS container agent will reserve for the container. This parameter maps to `CpuShares` in the [Create a container](#) section of the [Docker Remote API](#) and the `--cpu-shares` option to [docker run](#).

This field is optional for tasks using the Fargate launch type, and the only requirement is that the total amount of CPU reserved for all containers within a task be lower than the task-level `cpu` value.

Note

You can determine the number of CPU units that are available per Amazon EC2 instance type by multiplying the number of vCPUs listed for that instance type on the [Amazon EC2 Instances](#) detail page by 1,024.

Linux containers share unallocated CPU units with other containers on the container instance with the same ratio as their allocated amount. For example, if you run a single-container task on a single-core instance type with 512 CPU units specified for that container, and that is the only task running on the container instance, that container could use the full 1,024 CPU unit share at any given time. However, if you launched another copy of the same task on that container instance, each task would be guaranteed a minimum of 512 CPU units when needed, and each container could float to higher CPU usage if the other container was not using it, but if both tasks were 100% active all of the time, they would be limited to 512 CPU units.

On Linux container instances, the Docker daemon on the container instance uses the CPU value to calculate the relative CPU share ratios for running containers. For more information, see [CPU share constraint](#) in the Docker documentation. The minimum valid CPU share value that the Linux kernel allows is 2. However, the CPU parameter is not required, and you can use CPU values below 2 in your container definitions. For CPU values below 2 (including null), the behavior varies based on your Amazon ECS container agent version:

- **Agent versions <= 1.1.0:** Null and zero CPU values are passed to Docker as 0, which Docker then converts to 1,024 CPU shares. CPU values of 1 are passed to Docker as 1, which the Linux kernel converts to two CPU shares.
- **Agent versions >= 1.2.0:** Null, zero, and CPU values of 1 are passed to Docker as two CPU shares.

On Windows container instances, the CPU limit is enforced as an absolute limit, or a quota. Windows containers only have access to the specified amount of CPU that is described in the task definition.

`essential`

Type: Boolean

Required: no

If the `essential` parameter of a container is marked as `true`, and that container fails or stops for any reason, all other containers that are part of the task are stopped. If the `essential` parameter of a container is marked as `false`, then its failure does not affect the rest of the containers in a task. If this parameter is omitted, a container is assumed to be essential.

All tasks must have at least one essential container. If you have an application that is composed of multiple containers, you should group containers that are used for a common purpose into components, and separate the different components into multiple task definitions. For more information, see [Application Architecture \(p. 29\)](#).

```
"essential": true|false
```

entryPoint

Important

Early versions of the Amazon ECS container agent do not properly handle `entryPoint` parameters. If you have problems using `entryPoint`, update your container agent or enter your commands and arguments as `command` array items instead.

Type: string array

Required: no

The entry point that is passed to the container. This parameter maps to `Entrypoint` in the [Create a container](#) section of the [Docker Remote API](#) and the `--entrypoint` option to **docker run**. For more information about the Docker `ENTRYPOINT` parameter, go to <https://docs.docker.com/engine/reference/builder/#entrypoint>.

```
"entryPoint": ["string", ...]
```

command

Type: string array

Required: no

The command that is passed to the container. This parameter maps to `Cmd` in the [Create a container](#) section of the [Docker Remote API](#) and the `COMMAND` parameter to **docker run**. For more information about the Docker `CMD` parameter, go to <https://docs.docker.com/engine/reference/builder/#cmd>. If there are multiple arguments, each argument should be a separated string in the array.

```
"command": ["string", ...]
```

workingDirectory

Type: string

Required: no

The working directory in which to run commands inside the container. This parameter maps to `WorkingDir` in the [Create a container](#) section of the [Docker Remote API](#) and the `--workdir` option to **docker run**.

```
"workingDirectory": "string"
```

environment

Type: object array

Required: no

The environment variables to pass to a container. This parameter maps to `Env` in the [Create a container](#) section of the [Docker Remote API](#) and the `--env` option to [docker run](#).

Important

We do not recommend using plaintext environment variables for sensitive information, such as credential data.

`name`

Type: String

Required: Yes, when `environment` is used

The name of the environment variable.

`value`

Type: String

Required: Yes, when `environment` is used

The value of the environment variable.

```
"environment" : [
  { "name" : "string", "value" : "string" },
  { "name" : "string", "value" : "string" }
]
```

`secrets`

Type: Object array

Required: No

An object representing the secret to expose to your container. For more information, see [Specifying Sensitive Data](#) (p. 87).

`name`

Type: String

Required: Yes

The value to set as the environment variable on the container.

`valueFrom`

Type: String

Required: Yes

The secret to expose to the container. The supported values are either the full ARN of the AWS Secrets Manager secret or the full ARN of the parameter in the AWS Systems Manager Parameter Store.

Note

If the Systems Manager Parameter Store parameter exists in the same Region as the task you are launching then you can use either the full ARN or name of the secret. If the parameter exists in a different Region then the full ARN must be specified.

```
"secrets": [
```

```
{
  "name": "environment_variable_name",
  "valueFrom": "arn:aws:ssm:region:aws_account_id:parameter/parameter_name"
}
```

Network Settings

dnsServers

Type: string array

Required: no

A list of DNS servers that are presented to the container. This parameter maps to `Dns` in the [Create a container](#) section of the [Docker Remote API](#) and the `--dns` option to [docker run](#).

Note

This parameter is not supported for Windows containers or tasks using the `awsvpc` network mode.

```
"dnsServers": ["string", ...]
```

Storage and Logging

readOnlyRootFilesystem

Type: Boolean

Required: no

When this parameter is true, the container is given read-only access to its root file system. This parameter maps to `ReadonlyRootfs` in the [Create a container](#) section of the [Docker Remote API](#) and the `--read-only` option to [docker run](#).

Note

This parameter is not supported for Windows containers.

```
"readOnlyRootFilesystem": true|false
```

mountPoints

Type: Object Array

Required: No

The mount points for data volumes in your container.

This parameter maps to `Volumes` in the [Create a container](#) section of the [Docker Remote API](#) and the `--volume` option to [docker run](#).

Windows containers can mount whole directories on the same drive as `$env:ProgramData`. Windows containers cannot mount directories on a different drive, and mount point cannot be across drives.

sourceVolume

Type: String

Required: Yes, when `mountPoints` are used

The name of the volume to mount.

`containerPath`

Type: String

Required: Yes, when `mountPoints` are used

The path on the container to mount the volume at.

`readOnly`

Type: Boolean

Required: No

If this value is `true`, the container has read-only access to the volume. If this value is `false`, then the container can write to the volume. The default value is `false`.

`volumesFrom`

Type: Object Array

Required: No

Data volumes to mount from another container. This parameter maps to `VolumesFrom` in the [Create a container](#) section of the [Docker Remote API](#) and the `--volumes-from` option to [docker run](#).

`sourceContainer`

Type: string

Required: yes, when `volumesFrom` is used

The name of the container to mount volumes from.

`readOnly`

Type: Boolean

Required: no

If this value is `true`, the container has read-only access to the volume. If this value is `false`, then the container can write to the volume. The default value is `false`.

```
"volumesFrom": [
  {
    "sourceContainer": "string",
    "readOnly": true|false
  }
]
```

`logConfiguration`

Type: [LogConfiguration](#) Object

Required: no

The log configuration specification for the container.

For example task definitions using a log configuration, see [Example Task Definitions \(p. 97\)](#).

This parameter maps to `LogConfig` in the [Create a container](#) section of the [Docker Remote API](#) and the `--log-driver` option to `docker run`. By default, containers use the same logging driver that the Docker daemon uses; however the container may use a different logging driver than the Docker daemon by specifying a log driver with this parameter in the container definition. To use a different logging driver for a container, the log system must be configured properly on the container instance (or on a different log server for remote logging options). For more information on the options for different supported log drivers, see [Configure logging drivers](#) in the Docker documentation.

The following should be noted when specifying a log configuration for your containers:

- Amazon ECS currently supports a subset of the logging drivers available to the Docker daemon (shown in the valid values below). Additional log drivers may be available in future releases of the Amazon ECS container agent.
- This parameter requires version 1.18 of the Docker Remote API or greater on your container instance.
- For tasks using the Fargate launch type, because you do not have access to the underlying infrastructure your tasks are hosted on, any additional software needed will have to be installed outside of the task. For example, the Fluentd output aggregators or a remote host running Logstash to send Gelf logs to.

```
"logConfiguration": {
  "logDriver": "awslogs", "fluentd", "gelf", "json-
file", "journald", "logentries", "splunk", "syslog", "awsfirelens",
  "options": { "string": "string"
    ... },
  "secretOptions": [{
    "name": "string",
    "valueFrom": "string"
  }]
}
```

logDriver

Type: string

Valid values: "awslogs", "fluentd", "gelf", "json-file", "journald", "logentries", "splunk", "syslog", "awsfirelens

Required: yes, when `logConfiguration` is used

The log driver to use for the container. The valid values listed earlier are log drivers that the Amazon ECS container agent can communicate with by default.

For tasks using the Fargate launch type, the supported log drivers are `awslogs`, `splunk`, and `awsfirelens`.

For more information on using the `awslogs` log driver in task definitions to send your container logs to CloudWatch Logs, see [Using the awslogs Log Driver \(p. 69\)](#).

For more information about using the `awsfirelens` log driver, see [Custom Log Routing](#).

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance.

options

Type: string to string map

Required: no

The configuration options to send to the log driver.

This parameter requires version 1.19 of the Docker Remote API or greater on your container instance.

secretOptions

Type: object array

Required: no

An object representing the secret to pass to the log configuration. For more information, see [Specifying Sensitive Data \(p. 87\)](#).

name

Type: String

Required: Yes

The value to set as the environment variable on the container.

valueFrom

Type: String

Required: Yes

The secret to expose to the log configuration of the container.

```
"logConfiguration": {
  "logDriver": "splunk",
  "options": {
    "splunk-url": "https://cloud.splunk.com:8080",
    "splunk-token": "...",
    "tag": "...",
    ...
  },
  "secretOptions": [{
    "name": "splunk-token",
    "valueFrom": "/ecs/logconfig/splunkcred"
  }]
}
```

Security

user

Type: string

Required: no

The user name to use inside the container. This parameter maps to `User` in the [Create a container](#) section of the [Docker Remote API](#) and the `--user` option to [docker run](#).

You can use the following formats. If specifying a UID or GID, you must specify it as a positive integer.

- `user`
- `user:group`
- `uid`
- `uid:gid`
- `user:gid`
- `uid:group`

Note

This parameter is not supported for Windows containers.

```
"user": "string"
```

Resource Limits

ulimits

Type: object array

Required: no

A list of `ulimits` to set in the container. This parameter maps to `Ulimits` in the [Create a container](#) section of the [Docker Remote API](#) and the `--ulimit` option to [docker run](#).

Fargate tasks use the default resource limit values with the exception of the `nofile` resource limit parameter which Fargate overrides. The `nofile` resource limit sets a restriction on the number of open files that a container can use. The default `nofile` soft limit is 1024 and hard limit is 4096 for Fargate tasks. These limits can be adjusted in a task definition if your tasks needs to handle a larger number of files.

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance.

Note

This parameter is not supported for Windows containers.

```
"ulimits": [
  {
    "name":
"core"|"cpu"|"data"|"fsize"|"locks"|"memlock"|"msgqueue"|"nice"|"nofile"|"nproc"|"rss"|"rtprio"|"r
    "softLimit": integer,
    "hardLimit": integer
  }
  ...
]
```

name

Type: string

Valid values: "core" | "cpu" | "data" | "fsize" | "locks" | "memlock" | "msgqueue" | "nice" | "nofile" | "nproc" | "rss" | "rtprio" | "rttime" | "sigpending" | "stack"

Required: yes, when `ulimits` are used

The type of the `ulimit`.

hardLimit

Type: integer

Required: yes, when `ulimits` are used

The hard limit for the `ulimit` type.

softLimit

Type: integer

Required: yes, when `ulimits` are used

The soft limit for the `ulimit` type.

Docker Labels

`dockerLabels`

Type: string to string map

Required: no

A key/value map of labels to add to the container. This parameter maps to `Labels` in the [Create a container](#) section of the [Docker Remote API](#) and the `--label` option to [docker run](#).

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance.

```
"dockerLabels": {"string": "string"
...}
```

Other Container Definition Parameters

The following container definition parameters are able to be used when registering task definitions in the Amazon ECS console by using the **Configure via JSON** option. For more information, see [Creating a Task Definition](#) (p. 30).

Topics

- [Linux Parameters](#) (p. 50)
- [Container Dependency](#) (p. 51)
- [Container Timeouts](#) (p. 52)
- [System Controls](#) (p. 53)
- [Interactive](#) (p. 54)
- [Pseudo Terminal](#) (p. 54)

Linux Parameters

`linuxParameters`

Type: [LinuxParameters](#) object

Required: no

Linux-specific options that are applied to the container, such as [KernelCapabilities](#).

Note

This parameter is not supported for Windows containers.

```
"linuxParameters": {
  "capabilities": {
    "add": ["string", ...],
    "drop": ["string", ...]
  }
}
```

capabilities

Type: [KernelCapabilities](#) object

Required: no

The Linux capabilities for the container that are dropped from the default configuration provided by Docker. For more information about the default capabilities and the non-default available capabilities, see [Runtime privilege and Linux capabilities](#) in the *Docker run reference*. For more detailed information about these Linux capabilities, see the [capabilities\(7\)](#) Linux manual page.

add

Type: string array

Valid values: "SYS_PTRACE"

Required: no

The Linux capabilities for the container to add to the default configuration provided by Docker. This parameter maps to CapAdd in the [Create a container](#) section of the [Docker Remote API](#) and the --cap-add option to [docker run](#).

Note

If you are using tasks that use the Fargate launch type, the add parameter is only supported if using platform version 1.4.0 or later.

drop

Type: string array

Valid values: "ALL" | "AUDIT_CONTROL" | "AUDIT_WRITE" | "BLOCK_SUSPEND" | "CHOWN" | "DAC_OVERRIDE" | "DAC_READ_SEARCH" | "FOWNER" | "FSETID" | "IPC_LOCK" | "IPC_OWNER" | "KILL" | "LEASE" | "LINUX_IMMUTABLE" | "MAC_ADMIN" | "MAC_OVERRIDE" | "MKNOD" | "NET_ADMIN" | "NET_BIND_SERVICE" | "NET_BROADCAST" | "NET_RAW" | "SETFCAP" | "SETGID" | "SETPCAP" | "SETUID" | "SYS_ADMIN" | "SYS_BOOT" | "SYS_CHROOT" | "SYS_MODULE" | "SYS_NICE" | "SYS_PACCT" | "SYS_PTRACE" | "SYS_RAWIO" | "SYS_RESOURCE" | "SYS_TIME" | "SYS_TTY_CONFIG" | "SYSLOG" | "WAKE_ALARM"

Required: no

The Linux capabilities for the container to remove from the default configuration provided by Docker. This parameter maps to CapDrop in the [Create a container](#) section of the [Docker Remote API](#) and the --cap-drop option to [docker run](#).

initProcessEnabled

Run an init process inside the container that forwards signals and reaps processes. This parameter maps to the --init option to [docker run](#).

This parameter requires version 1.25 of the Docker Remote API or greater on your container instance.

Container Dependency

dependsOn

Type: Array of [ContainerDependency](#) objects

Required: no

The dependencies defined for container startup and shutdown. A container can contain multiple dependencies. When a dependency is defined for container startup, for container shutdown it is reversed. For an example, see [Example: Container Dependency \(p. 100\)](#).

For tasks using the Fargate launch type, this parameter requires that the task or service uses platform version 1.3.0 or later.

```
"dependsOn": [
  {
    "containerName": "string",
    "condition": "string"
  }
]
```

containerName

Type: String

Required: Yes

The container name that must meet the specified condition.

condition

Type: String

Required: Yes

The dependency condition of the container. The following are the available conditions and their behavior:

- **START** – This condition emulates the behavior of links and volumes today. It validates that a dependent container is started before permitting other containers to start.
- **COMPLETE** – This condition validates that a dependent container runs to completion (exits) before permitting other containers to start. This can be useful for nonessential containers that run a script and then exit. This condition cannot be set on an essential container.
- **SUCCESS** – This condition is the same as **COMPLETE**, but it also requires that the container exits with a zero status. This condition cannot be set on an essential container.
- **HEALTHY** – This condition validates that the dependent container passes its Docker healthcheck before permitting other containers to start. This requires that the dependent container has health checks configured. This condition is confirmed only at task startup.

Container Timeouts

startTimeout

Type: Integer

Required: no

Example values: 120

Time duration (in seconds) to wait before giving up on resolving dependencies for a container. For example, you specify two containers in a task definition with containerA having a dependency on containerB reaching a **COMPLETE**, **SUCCESS**, or **HEALTHY** status. If a `startTimeout` value is specified for containerB and it does not reach the desired status within that time then containerA will give up and not start. This results in the task transitioning to a **STOPPED** state.

For tasks using the Fargate launch type, this parameter requires that the task or service uses platform version 1.3.0 or later. If this parameter is not specified, the default value of 3 minutes is used.

`stopTimeout`

Type: Integer

Required: no

Example values: 120

Time duration (in seconds) to wait before the container is forcefully killed if it doesn't exit normally on its own.

For tasks using the Fargate launch type, the task or service requires platform version 1.3.0 or later. The max stop timeout value is 120 seconds and if the parameter is not specified, the default value of 30 seconds is used.

System Controls

`systemControls`

Type: [SystemControl](#) object

Required: no

A list of namespaced kernel parameters to set in the container. This parameter maps to `Sysctl`s in the [Create a container](#) section of the [Docker Remote API](#) and the `--sysctl` option to [docker run](#).

It is not recommended that you specify network-related `systemControls` parameters for multiple containers in a single task that also uses either the `awsvpc` or `host` network mode for the following reasons:

- For tasks that use the `awsvpc` network mode, if you set `systemControls` for any container it will apply to all containers in the task. If you set different `systemControls` for multiple containers in a single task, the container that is started last will determine which `systemControls` take effect.
- For tasks that use the `host` network mode, the network namespace `systemControls` are not supported.

If you are setting an IPC resource namespace to use for the containers in the task, the following will apply to your system controls. For more information, see [IPC mode \(p. 60\)](#).

- For tasks that use the `host` IPC mode, IPC namespace `systemControls` are not supported.
- For tasks that use the `task` IPC mode, IPC namespace `systemControls` values will apply to all containers within a task.

Note

This parameter is not supported for Windows containers or tasks using the Fargate launch type.

```
"systemControls": [  
  {  
    "namespace": "string",  
    "value": "string"  
  }  
]
```

`namespace`

Type: String

Required: no

The namespaced kernel parameter to set a value for.

Valid IPC namespace values: "kernel.msgmax" | "kernel.msgmnb" | "kernel.msgmni" | "kernel.sem" | "kernel.shmall" | "kernel.shmmax" | "kernel.shmmni" | "kernel.shm_rmid_forced", as well as Sysctls beginning with "fs.mqueue.*"

Valid network namespace values: Sysctls beginning with "net.*"

value

Type: String

Required: no

The value for the namespaced kernel parameter specified in namespace.

Interactive

interactive

Type: Boolean

Required: no

When this parameter is `true`, this allows you to deploy containerized applications that require `stdin` or a `tty` to be allocated. This parameter maps to `OpenStdin` in the [Create a container](#) section of the [Docker Remote API](#) and the `--interactive` option to [docker run](#).

Pseudo Terminal

pseudoTerminal

Type: Boolean

Required: no

When this parameter is `true`, a TTY is allocated. This parameter maps to `Tty` in the [Create a container](#) section of the [Docker Remote API](#) and the `--tty` option to [docker run](#).

Volumes

When you register a task definition, you can optionally specify a list of volumes to be passed to the Docker daemon on a container instance, which then becomes available for access by other containers on the same container instance.

For more information, see [Using Data Volumes in Tasks \(p. 63\)](#).

The following parameters are allowed in a container definition:

name

Type: String

Required: No

The name of the volume. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed. This name is referenced in the `sourceVolume` parameter of container definition `mountPoints`.

host

Required: No

This parameter is specified when using bind mounts. To use Docker volumes, specify a `dockerVolumeConfiguration` instead. The contents of the `host` parameter determine whether your bind mount data volume persists on the host container instance and where it is stored. If the `host` parameter is empty, then the Docker daemon assigns a host path for your data volume, but the data is not guaranteed to persist after the containers associated with it stop running.

Bind mount host volumes are supported when using either the EC2 or Fargate launch types.

Windows containers can mount whole directories on the same drive as `$env:ProgramData`.

sourcePath

Type: String

Required: No

When the `host` parameter is used, specify a `sourcePath` to declare the path on the host container instance that is presented to the container. If this parameter is empty, then the Docker daemon has assigned a host path for you. If the `host` parameter contains a `sourcePath` file location, then the data volume persists at the specified location on the host container instance until you delete it manually. If the `sourcePath` value does not exist on the host container instance, the Docker daemon creates it. If the location does exist, the contents of the source path folder are exported.

efsVolumeConfiguration

Type: Object

Required: No

This parameter is specified when using Amazon EFS volumes.

fileSystemId

Type: String

Required: Yes

The Amazon EFS file system ID to use.

rootDirectory

Type: String

Required: No

The directory within the Amazon EFS file system to mount as the root directory inside the host. If this parameter is omitted, the root of the Amazon EFS volume will be used. Specifying `/` will have the same effect as omitting this parameter.

transitEncryption

Type: String

Valid values: `ENABLED` | `DISABLED`

Required: No

Whether or not to enable encryption for Amazon EFS data in transit between the Amazon ECS host and the Amazon EFS server. Transit encryption must be enabled if Amazon EFS IAM

authorization is used. If this parameter is omitted, the default value of `DISABLED` is used. For more information, see [Encrypting Data in Transit](#) in the *Amazon Elastic File System User Guide*.

`transitEncryptionPort`

Type: Integer

Required: No

The port to use when sending encrypted data between the Amazon ECS host and the Amazon EFS server. If you do not specify a transit encryption port, it will use the port selection strategy that the Amazon EFS mount helper uses. For more information, see [EFS Mount Helper](#) in the *Amazon Elastic File System User Guide*.

`authorizationConfig`

Type: Object

Required: No

The authorization configuration details for the Amazon EFS file system.

`accessPointId`

Type: String

Required: No

The access point ID to use. If an access point is specified, the root directory value will be relative to the directory set for the access point. If specified, transit encryption must be enabled in the `EFSVolumeConfiguration`. For more information, see [Working with Amazon EFS Access Points](#) in the *Amazon Elastic File System User Guide*.

`iam`

Type: String

Valid values: `ENABLED` | `DISABLED`

Required: No

Whether or not to use the Amazon ECS task IAM role defined in a task definition when mounting the Amazon EFS file system. If enabled, transit encryption must be enabled in the `EFSVolumeConfiguration`. If this parameter is omitted, the default value of `DISABLED` is used. For more information, see [IAM Roles for Tasks](#).

Launch types

When you register a task definition, you specify the launch type to use for your task. For more information, see [Amazon ECS Launch Types \(p. 60\)](#).

The following parameter is allowed in a task definition:

`requiresCompatibilities`

Type: string array

Required: no

Valid Values: `EC2` | `FARGATE`

The launch type the task is using. This enables a check to ensure that all of the parameters used in the task definition meet the requirements of the launch type.

Valid values are `FARGATE` and `EC2`. For more information about launch types, see [Amazon ECS Launch Types](#) (p. 60).

Task size

When you register a task definition, you can specify the total `cpu` and `memory` used for the task. This is separate from the `cpu` and `memory` values at the container definition level. If using the `EC2` launch type, these fields are optional. If using the `Fargate` launch type, these fields are required and there are specific values for both `cpu` and `memory` that are supported.

Note

Task-level CPU and memory parameters are ignored for Windows containers. We recommend specifying container-level resources for Windows containers.

The following parameter is allowed in a task definition:

`cpu`

Type: string

Required: no

Note

This parameter is not supported for Windows containers.

The hard limit of CPU units to present for the task. It can be expressed as an integer using CPU units, for example `1024`, or as a string using vCPUs, for example `1 vCPU` or `1 vcpu`, in a task definition. When the task definition is registered, a vCPU value is converted to an integer indicating the CPU units.

If using the `Fargate` launch type, this field is required and you must use one of the following values, which determines your range of supported values for the `memory` parameter:

CPU value	Memory value (MiB)
256 (.25 vCPU)	512 (0.5 GB), 1024 (1 GB), 2048 (2 GB)
512 (.5 vCPU)	1024 (1 GB), 2048 (2 GB), 3072 (3 GB), 4096 (4 GB)
1024 (1 vCPU)	2048 (2 GB), 3072 (3 GB), 4096 (4 GB), 5120 (5 GB), 6144 (6 GB), 7168 (7 GB), 8192 (8 GB)
2048 (2 vCPU)	Between 4096 (4 GB) and 16384 (16 GB) in increments of 1024 (1 GB)
4096 (4 vCPU)	Between 8192 (8 GB) and 30720 (30 GB) in increments of 1024 (1 GB)

`memory`

Type: string

Required: no

Note

This parameter is not supported for Windows containers.

The hard limit of memory (in MiB) to present to the task. It can be expressed as an integer using MiB, for example 1024, or as a string using GB, for example 1GB or 1 GB, in a task definition. When the task definition is registered, a GB value is converted to an integer indicating the MiB.

If using the Fargate launch type, this field is required and you must use one of the following values, which determines your range of supported values for the `cpu` parameter:

Memory value (MiB)	CPU value
512 (0.5 GB), 1024 (1 GB), 2048 (2 GB)	256 (.25 vCPU)
1024 (1 GB), 2048 (2 GB), 3072 (3 GB), 4096 (4 GB)	512 (.5 vCPU)
2048 (2 GB), 3072 (3 GB), 4096 (4GB), 5120 (5 GB), 6144 (6 GB), 7168 (7 GB), 8192 (8 GB)	1024 (1 vCPU)
Between 4096 (4 GB) and 16384 (16 GB) in increments of 1024 (1 GB)	2048 (2 vCPU)
Between 8192 (8 GB) and 30720 (30 GB) in increments of 1024 (1 GB)	4096 (4 vCPU)

Proxy configuration

`proxyConfiguration`

Type: [ProxyConfiguration](#) object

Required: no

The configuration details for the App Mesh proxy.

For tasks using the Fargate launch type, this feature requires that the task or service uses platform version 1.3.0 or later.

Note

This parameter is not supported for Windows containers.

```
"proxyConfiguration": {
  "type": "APPMESH",
  "containerName": "string",
  "properties": [
    {
      "name": "string",
      "value": "string"
    }
  ]
}
```

`type`

Type: String

Value values: APPMESH

Required: No

The proxy type. The only supported value is APPMESH.

`containerName`

Type: String

Required: Yes

The name of the container that will serve as the App Mesh proxy.

`properties`

Type: Array of [KeyValuePair](#) objects

Required: No

The set of network configuration parameters to provide the Container Network Interface (CNI) plugin, specified as key-value pairs.

- `IgnoredUID` – (Required) The user ID (UID) of the proxy container as defined by the `user` parameter in a container definition. This is used to ensure the proxy ignores its own traffic. If `IgnoredGID` is specified, this field can be empty.
- `IgnoredGID` – (Required) The group ID (GID) of the proxy container as defined by the `user` parameter in a container definition. This is used to ensure the proxy ignores its own traffic. If `IgnoredUID` is specified, this field can be empty.
- `AppPorts` – (Required) The list of ports that the application uses. Network traffic to these ports is forwarded to the `ProxyIngressPort` and `ProxyEgressPort`.
- `ProxyIngressPort` – (Required) Specifies the port that incoming traffic to the `AppPorts` is directed to.
- `ProxyEgressPort` – (Required) Specifies the port that outgoing traffic from the `AppPorts` is directed to.
- `EgressIgnoredPorts` – (Required) The egress traffic going to these specified ports is ignored and not redirected to the `ProxyEgressPort`. It can be an empty list.
- `EgressIgnoredIPs` – (Required) The egress traffic going to these specified IP addresses is ignored and not redirected to the `ProxyEgressPort`. It can be an empty list.

`name`

Type: String

Required: No

The name of the key-value pair.

`value`

Type: String

Required: No

The value of the key-value pair.

Other task definition parameters

The following task definition parameters are able to be used when registering task definitions in the Amazon ECS console by using the **Configure via JSON** option. For more information, see [Creating a Task Definition](#) (p. 30).

Topics

- [IPC mode](#) (p. 60)
- [PID mode](#) (p. 60)

IPC mode

`ipcMode`

Type: String

Required: No

The IPC resource namespace to use for the containers in the task. The valid values are `host`, `task`, or `none`. If `host` is specified, then all containers within the tasks that specified the `host` IPC mode on the same container instance share the same IPC resources with the host Amazon EC2 instance. If `task` is specified, all containers within the specified task share the same IPC resources. If `none` is specified, then IPC resources within the containers of a task are private and not shared with other containers in a task or on the container instance. If no value is specified, then the IPC resource namespace sharing depends on the Docker daemon setting on the container instance. For more information, see [IPC settings](#) in the *Docker run reference*.

If the `host` IPC mode is used, be aware that there is a heightened risk of undesired IPC namespace exposure. For more information, see [Docker security](#).

If you are setting namespaced kernel parameters using `systemControls` for the containers in the task, the following will apply to your IPC resource namespace. For more information, see [System Controls \(p. 53\)](#).

- For tasks that use the `host` IPC mode, IPC namespace related `systemControls` are not supported.
- For tasks that use the `task` IPC mode, IPC namespace related `systemControls` will apply to all containers within a task.

Note

This parameter is not supported for Windows containers or tasks using the Fargate launch type.

PID mode

`pidMode`

Type: String

Required: No

The process namespace to use for the containers in the task. The valid values are `host` or `task`. If `host` is specified, then all containers within the tasks that specified the `host` PID mode on the same container instance share the same process namespace with the host Amazon EC2 instance. If `task` is specified, all containers within the specified task share the same process namespace. If no value is specified, the default is a private namespace. For more information, see [PID settings](#) in the *Docker run reference*.

If the `host` PID mode is used, be aware that there is a heightened risk of undesired process namespace exposure. For more information, see [Docker security](#).

Note

This parameter is not supported for Windows containers or tasks using the Fargate launch type.

Amazon ECS Launch Types

An Amazon ECS launch type determines the type of infrastructure on which your tasks and services are hosted.

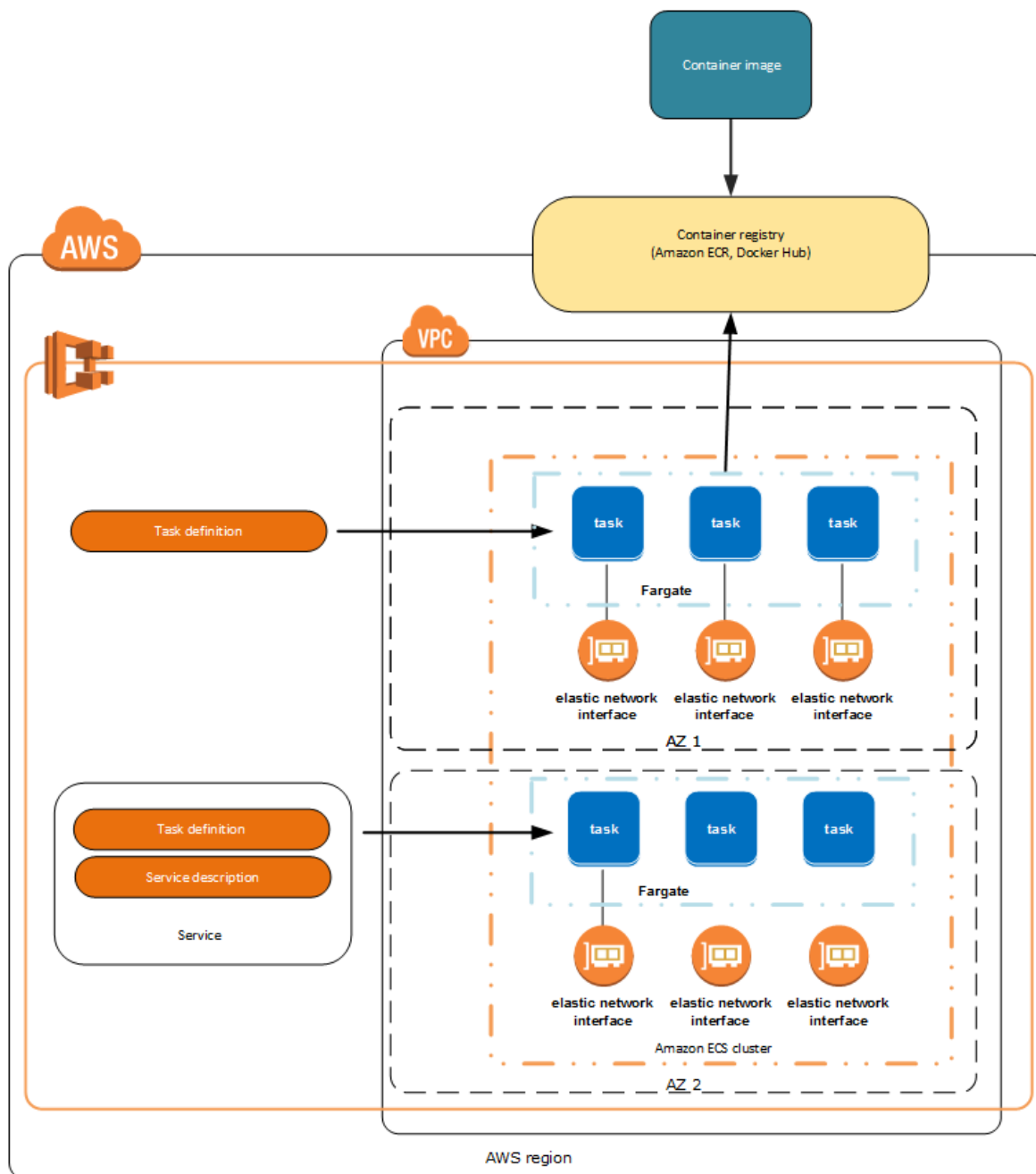
Fargate Launch Type

The Fargate launch type allows you to run your containerized applications without the need to provision and manage the backend infrastructure. Just register your task definition and Fargate launches the container for you.

The AWS Fargate launch type is currently available in the following Regions:

Region Name	Region
US East (Ohio)	us-east-2
US East (N. Virginia)	us-east-1
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Africa (Cape Town)	af-south-1
Asia Pacific (Hong Kong)	ap-east-1
Asia Pacific (Mumbai)	ap-south-1 (aps1-az1 & aps1-az3 only)
Asia Pacific (Seoul)	ap-northeast-2
Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
Canada (Central)	ca-central-1
China (Beijing)	cn-north-1
China (Ningxia)	cn-northwest-1
Europe (Frankfurt)	eu-central-1
Europe (Ireland)	eu-west-1
Europe (London)	eu-west-2
Europe (Paris)	eu-west-3
Europe (Stockholm)	eu-north-1
South America (São Paulo)	sa-east-1
Middle East (Bahrain)	me-south-1
AWS GovCloud (US-East)	us-gov-east-1
AWS GovCloud (US-West)	us-gov-west-1

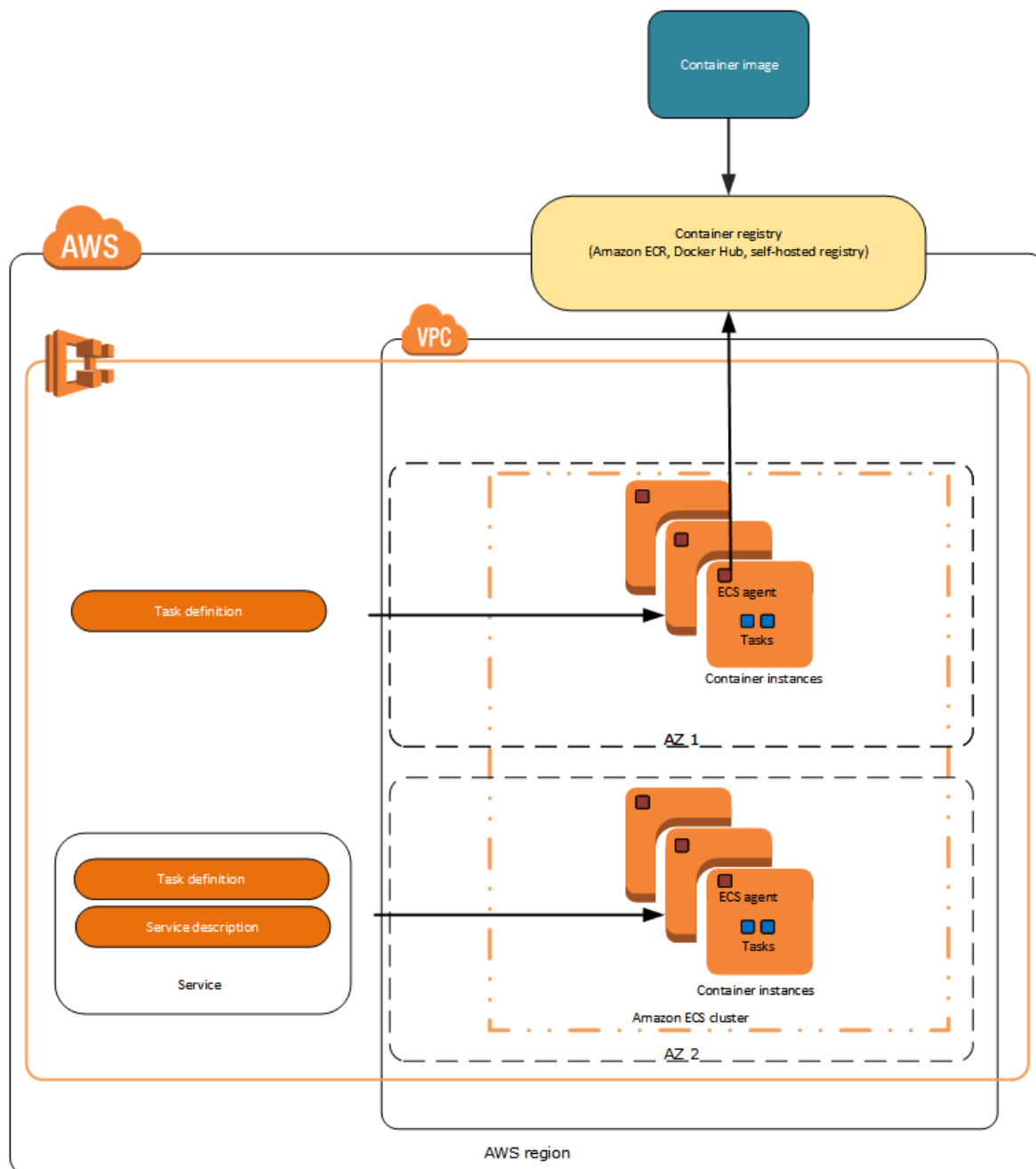
This diagram shows the general architecture:



EC2 Launch Type

The EC2 launch type allows you to run your containerized applications on a cluster of Amazon EC2 instances that you manage.

This diagram shows the general architecture:



Using Data Volumes in Tasks

For Fargate tasks, the following storage types are supported:

- Amazon EFS volumes for persistent storage. For more information, see [Amazon EFS Volumes \(p. 65\)](#).
- Ephemeral storage for nonpersistent storage.

When provisioned, each Amazon ECS task on Fargate receives the following ephemeral storage. The ephemeral storage configuration depends on which platform version the task is using. After a Fargate task stops, the ephemeral storage is deleted. For more information about Amazon ECS default service limits, see [Amazon ECS Service Quotas](#) (p. 283).

Fargate tasks using platform version 1.4.0 or later

All Amazon ECS on Fargate tasks using platform version 1.4.0 or later receive a minimum of 20 GB of ephemeral storage.

For tasks using platform version 1.4.0 or later that are launched on May 28, 2020 or later, the ephemeral storage is encrypted with an AES-256 encryption algorithm using an AWS Fargate-managed encryption key.

Fargate tasks using platform version 1.3.0 or earlier

For Amazon ECS on Fargate tasks using platform version 1.3.0 or earlier, each task receives the following ephemeral storage.

- 10 GB of Docker layer storage
- An additional 4 GB for volume mounts. This can be mounted and shared among containers using the `volumes`, `mountPoints` and `volumesFrom` parameters in the task definition.

Note

The `host` and `sourcePath` parameters are not supported for Fargate tasks.

Example task definition

In this example, you have two application containers that need to access the same scratch file storage location.

To provide nonpersistent empty storage for containers in a Fargate task

1. In the task definition `volumes` section, define a volume with the name `application_scratch`.

```
"volumes": [  
  {  
    "name": "application_scratch",  
    "host": {}  
  }  
]
```

2. In the `containerDefinitions` section, create the application container definitions so they mount the nonpersistent storage.

```
"containerDefinitions": [  
  {  
    "name": "application1",  
    "image": "my-repo/application",  
    "cpu": 100,  
    "memory": 100,  
    "essential": true,  
    "mountPoints": [  
      {  
        "sourceVolume": "application_scratch",  
        "containerPath": "/var/scratch"  
      }  
    ]  
  }  
]
```



```
    ],  
  },  
  {  
    "name": "application2",  
    "image": "my-repo/application",  
    "cpu": 100,  
    "memory": 100,  
    "essential": true,  
    "mountPoints": [  
      {  
        "sourceVolume": "application_scratch",  
        "containerPath": "/var/scratch"  
      }  
    ]  
  }  
]
```

Amazon EFS Volumes

Amazon Elastic File System (Amazon EFS) provides simple, scalable file storage for use with Amazon EC2 instances. With Amazon EFS, storage capacity is elastic, growing and shrinking automatically as you add and remove files. Your applications can have the storage they need, when they need it.

You can use Amazon EFS file systems with Amazon ECS to export file system data across your fleet of container instances. That way, your tasks have access to the same persistent storage, no matter the instance on which they land. However, you must configure your container instance AMI to mount the Amazon EFS file system before the Docker daemon starts. Also, your task definitions must reference volume mounts on the container instance to use the file system. The following sections help you get started using Amazon EFS with Amazon ECS.

Amazon EFS Volume Considerations

The following should be considered when using Amazon EFS volumes:

- For tasks using the Fargate launch type, Amazon EFS file system support was added when using platform version 1.4.0 or later. For more information, see [AWS Fargate platform versions \(p. 14\)](#).
- When specifying Amazon EFS volumes in tasks using the Fargate launch type, Fargate creates a supervisor container that is responsible for managing the Amazon EFS volume. The supervisor container uses a small amount of the task's memory. The supervisor container is visible when querying the task metadata version 4 endpoint, but is not visible in CloudWatch Container Insights. For more information, see [Task metadata endpoint version 4 \(p. 271\)](#).

Using Amazon EFS Access Points

Amazon EFS access points are application-specific entry points into an EFS file system that make it easier to manage application access to shared datasets. For more information on Amazon EFS access points and how to control access to them, see [Working with Amazon EFS Access Points](#) in the *Amazon Elastic File System User Guide*.

Access points can enforce a user identity, including the user's POSIX groups, for all file system requests that are made through the access point. Access points can also enforce a different root directory for the file system so that clients can only access data in the specified directory or its subdirectories.

You can use an Amazon ECS task IAM role to enforce that specific applications use a specific access point. By combining IAM policies with access points, you can easily provide secure access to specific datasets for your applications. For more information on using task IAM roles, see [IAM Roles for Tasks \(p. 240\)](#).

Specifying an Amazon EFS File System in your Task Definition

In order to use Amazon EFS file system volumes for your containers, you must specify the volume and mount point configurations in your task definition. The following task definition JSON snippet shows the syntax for the `volumes` and `mountPoints` objects for a container:

```
{
  "containerDefinitions": [
    {
      "name": "container-using-efs",
      "image": "amazonlinux:2",
      "entryPoint": [
        "sh",
        "-c"
      ],
      "command": [
        "ls -la /mount/efs"
      ],
      "mountPoints": [
        {
          "sourceVolume": "myEfsVolume",
          "containerPath": "/mount/efs",
          "readOnly": true
        }
      ]
    }
  ],
  "volumes": [
    {
      "name": "myEfsVolume",
      "efsVolumeConfiguration": {
        "fileSystemId": "fs-1234",
        "rootDirectory": "/path/to/my/data",
        "transitEncryption": "ENABLED",
        "transitEncryptionPort": integer,
        "authorizationConfig": {
          "accessPointId": "fsap-1234",
          "iam": "ENABLED"
        }
      }
    }
  ]
}
```

`efsVolumeConfiguration`

Type: Object

Required: No

This parameter is specified when using Amazon EFS volumes.

`fileSystemId`

Type: String

Required: Yes

The Amazon EFS file system ID to use.

`rootDirectory`

Type: String

Required: No

The directory within the Amazon EFS file system to mount as the root directory inside the host. If this parameter is omitted, the root of the Amazon EFS volume will be used. Specifying / will have the same effect as omitting this parameter.

`transitEncryption`

Type: String

Valid values: `ENABLED` | `DISABLED`

Required: No

Whether or not to enable encryption for Amazon EFS data in transit between the Amazon ECS host and the Amazon EFS server. Transit encryption must be enabled if Amazon EFS IAM authorization is used. If this parameter is omitted, the default value of `DISABLED` is used. For more information, see [Encrypting Data in Transit](#) in the *Amazon Elastic File System User Guide*.

`transitEncryptionPort`

Type: Integer

Required: No

The port to use when sending encrypted data between the Amazon ECS host and the Amazon EFS server. If you do not specify a transit encryption port, it will use the port selection strategy that the Amazon EFS mount helper uses. For more information, see [EFS Mount Helper](#) in the *Amazon Elastic File System User Guide*.

`authorizationConfig`

Type: Object

Required: No

The authorization configuration details for the Amazon EFS file system.

`accessPointId`

Type: String

Required: No

The access point ID to use. If an access point is specified, the root directory value will be relative to the directory set for the access point. If specified, transit encryption must be enabled in the `EFSVolumeConfiguration`. For more information, see [Working with Amazon EFS Access Points](#) in the *Amazon Elastic File System User Guide*.

`iam`

Type: String

Valid values: `ENABLED` | `DISABLED`

Required: No

Whether or not to use the Amazon ECS task IAM role defined in a task definition when mounting the Amazon EFS file system. If enabled, transit encryption must be enabled in the `EFSVolumeConfiguration`. If this parameter is omitted, the default value of `DISABLED` is used. For more information, see [IAM Roles for Tasks](#).

Fargate Task Networking

Amazon ECS tasks using Fargate require the `awsvpc` network mode, which provides each task with an elastic network interface (ENI) and a primary private IP address. When you run a task or create a service with this network mode, you must specify one or more subnets to attach the network interface to and

one or more security groups to apply to the network interface. Because each task gets its own ENI, you can also take advantage of other Amazon EC2 networking features like VPC Flow Logs so that you can monitor traffic to and from your tasks. Additionally, containers that belong to the same task can communicate over the `localhost` interface. A task can only have one ENI associated with it at a given time.

If you are using public subnets, decide whether to provide a public IP address for the network interface. For a Fargate task in a public subnet to pull container images, a public IP address needs to be assigned to the task's elastic network interface, with a route to the internet or a NAT gateway that can route requests to the internet. For a Fargate task in a private subnet to pull container images, the private subnet requires a NAT gateway be attached to route requests to the internet.

The following is an example of the `networkConfiguration` section for a Fargate task or service:

```
"networkConfiguration": {
  "awsvpcConfiguration": {
    "assignPublicIp": "ENABLED",
    "securityGroups": [ "sg-12345678" ],
    "subnets": [ "subnet-12345678" ]
  }
}
```

Services with tasks that use the Fargate launch type only support Application Load Balancers and Network Load Balancers. Classic Load Balancers are not supported. Also, when you create any target groups, you must choose `ip` as the target type, not `instance`. For more information, see [Service Load Balancing \(p. 152\)](#).

The network interfaces that are created are fully managed by AWS Fargate and there is an associated IAM policy that is used to grant permissions for Fargate. For tasks using Fargate platform version 1.4 or later, the task receives a single ENI (referred to as the task ENI) and all network traffic flows through that ENI within your VPC and will be visible to you through your VPC flow logs. For tasks that use Fargate platform version 1.3 and earlier, in addition to the task ENI, the task also receives a separate Fargate-owned ENI which is used for some network traffic which is not visible in the VPC flow logs. The following describes the network traffic behavior as well as the required IAM policy for each platform version.

Action	Traffic flow with platform version 1.3 and earlier	Traffic flow with platform version 1.4	IAM permission
Retrieving Amazon ECR login credentials	Fargate-owned ENI	Task ENI	Task execution IAM role
Image pull	Task ENI	Task ENI	Task execution IAM role
Sending logs through a log driver	Task ENI	Task ENI	Task execution IAM role
Sending logs through FireLens for Amazon ECS	Task ENI	Task ENI	Task IAM role
Retrieving secrets from Secrets Manager or Systems Manager	Fargate-owned ENI	Task ENI	Task execution IAM role
Amazon EFS file system traffic	Not available	Task ENI	Task IAM role
Application traffic	Task ENI	Task ENI	Task IAM role

Fargate Task Networking Considerations

There are several things to consider when using task networking.

- The Amazon ECS service-linked role is required to provide Amazon ECS with the permissions to make calls to other AWS services on your behalf. This role is created for you automatically when you create a cluster, or if you create or update a service in the AWS Management Console. For more information, see [Service-Linked Role for Amazon ECS \(p. 227\)](#). You can also create the service-linked role with the following AWS CLI command:

```
aws iam create-service-linked-role --aws-service-name ecs.amazonaws.com
```

- Amazon ECS populates the hostname of a task using task networking with an Amazon-provided (internal) DNS hostname when both the `enableDnsHostnames` and `enableDnsSupport` options are enabled on your VPC. If these options are not enabled, the DNS hostname of the task will be a random hostname. For more information on the DNS settings for a VPC, see [Using DNS with Your VPC](#) in the *Amazon VPC User Guide*.
- There is a limit of 16 subnets and 5 security groups that are able to be specified in the `awsVpcConfiguration`. For more information, see [AwsVpcConfiguration](#) in the *Amazon Elastic Container Service API Reference*.
- The ENIs that are created and attached by Fargate cannot be detached manually or modified by your account. This is to prevent the accidental deletion of an ENI that is associated with a running task. To release the ENIs for a task, stop the task.
- If a VPC is updated, for example to change the DHCP options set it uses, and you want tasks using the VPC to pick up the changes, those tasks must be stopped and new tasks started.
- For tasks using platform version 1.4 or later, the task ENIs support jumbo frames. Network interfaces are configured with a maximum transmission unit (MTU), which is the size of the largest payload that fits within a single frame. The larger the MTU, the more application payload can fit within a single frame, which reduces per-frame overhead and increases efficiency. Supporting jumbo frames will reduce overhead when the network path between your task and the destination supports jumbo frames, such as all traffic that remains within your VPC.

Using the awslogs Log Driver

You can configure the containers in your tasks to send log information to CloudWatch Logs. This allows you to view the logs from the containers in your Fargate tasks. This topic helps you get started using the `awslogs` log driver in your task definitions.

Note

The type of information that is logged by the containers in your task depends mostly on their `ENTRYPOINT` command. By default, the logs that are captured show the command output that you would normally see in an interactive terminal if you ran the container locally, which are the `STDOUT` and `STDERR` I/O streams. The `awslogs` log driver simply passes these logs from Docker to CloudWatch Logs. For more information on how Docker logs are processed, including alternative ways to capture different file data or streams, see [View logs for a container or service](#) in the Docker documentation.

Enabling the awslogs Log Driver for Your Containers

If you are using the Fargate launch type for your tasks, all you need to do to enable the `awslogs` log driver is add the required `logConfiguration` parameters to your task definition. For more information, see [Specifying a Log Configuration in your Task Definition \(p. 72\)](#).

Creating a Log Group

The `awslogs` log driver can send log streams to an existing log group in CloudWatch Logs or it can create a new log group on your behalf. The AWS Management Console provides an auto-configure option which creates a log group on your behalf using the task definition family name with `ecs` as the prefix. Alternatively, you can manually specify your log configuration options and specify the `awslogs-create-group` option with a value of `true` which will create the log groups on your behalf.

Note

To use the `awslogs-create-group` option to have your log group created, your IAM policy must include the `logs:CreateLogGroup` permission.

Using the Auto-configuration Feature to Create a Log Group

When registering a task definition in the Amazon ECS console, you have the option to allow Amazon ECS to auto-configure your CloudWatch logs. This option creates a log group on your behalf using the task definition family name with `ecs` as the prefix.

To use log group auto-configuration option in the Amazon ECS console

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the left navigation pane, choose **Task Definitions, Create new Task Definition**.
3. Select your compatibility option and choose **Next Step**.
4. Choose **Add container**.
5. In the **Storage and Logging** section, for **Log configuration**, choose **Auto-configure CloudWatch Logs**.
6. Enter your `awslogs` log driver options. For more information, see [Specifying a Log Configuration in your Task Definition](#) (p. 72).
7. Complete the rest of the task definition wizard.

Available awslogs Log Driver Options

The `awslogs` log driver supports the following options in Amazon ECS task definitions. For more information, see [CloudWatch Logs logging driver](#).

`awslogs-create-group`

Required: No

Specify whether you want the log group automatically created. If this option is not specified, it defaults to `false`.

Note

Your IAM policy must include the `logs:CreateLogGroup` permission before you attempt to use `awslogs-create-group`.

`awslogs-region`

Required: Yes

Specify the region to which the `awslogs` log driver should send your Docker logs. You can choose to send all of your logs from clusters in different regions to a single region in CloudWatch Logs so that they are all visible in one location, or you can separate them by region for more granularity. Be sure that the specified log group exists in the region that you specify with this option.

`awslogs-endpoint`

Required: No

By default, Docker uses either the `awslogs-region` log option or the detected Region to construct the remote CloudWatch Logs API endpoint. Use the `awslogs-endpoint` log option to override the default endpoint with the provided endpoint.

`awslogs-group`

Required: Yes

You must specify a log group to which the `awslogs` log driver sends its log streams. For more information, see [Creating a Log Group \(p. 70\)](#).

`awslogs-stream-prefix`

Required: Yes, when using the Fargate launch type.

The `awslogs-stream-prefix` option allows you to associate a log stream with the specified prefix, the container name, and the ID of the Amazon ECS task to which the container belongs. If you specify a prefix with this option, then the log stream takes the following format:

```
prefix-name/container-name/ecs-task-id
```

For Amazon ECS services, you could use the service name as the prefix, which would allow you to trace log streams to the service that the container belongs to, the name of the container that sent them, and the ID of the task to which the container belongs.

`awslogs-datetime-format`

Required: No

This option defines a multiline start pattern in Python `strftime` format. A log message consists of a line that matches the pattern and any following lines that don't match the pattern. Thus the matched line is the delimiter between log messages.

One example of a use case for using this format is for parsing output such as a stack dump, which might otherwise be logged in multiple entries. The correct pattern allows it to be captured in a single entry.

For more information, see [awslogs-datetime-format](#).

This option always takes precedence if both `awslogs-datetime-format` and `awslogs-multiline-pattern` are configured.

Note

Multiline logging performs regular expression parsing and matching of all log messages, which may have a negative impact on logging performance.

`awslogs-multiline-pattern`

Required: No

This option defines a multiline start pattern using a regular expression. A log message consists of a line that matches the pattern and any following lines that don't match the pattern. Thus the matched line is the delimiter between log messages.

For more information, see [awslogs-multiline-pattern](#).

This option is ignored if `awslogs-datetime-format` is also configured.

Note

Multiline logging performs regular expression parsing and matching of all log messages. This may have a negative impact on logging performance.

Specifying a Log Configuration in your Task Definition

Before your containers can send logs to CloudWatch, you must specify the `awslogs` log driver for containers in your task definition. This section describes the log configuration for a container to use the `awslogs` log driver. For more information, see [Creating a Task Definition \(p. 30\)](#).

The task definition JSON shown below has a `logConfiguration` object specified for each container; one for the WordPress container that sends logs to a log group called `awslogs-wordpress`, and one for a MySQL container that sends logs to a log group called `awslogs-mysql`. Both containers use the `awslogs-example` log stream prefix.

```
{
  "containerDefinitions": [
    {
      "name": "wordpress",
      "links": [
        "mysql"
      ],
      "image": "wordpress",
      "essential": true,
      "portMappings": [
        {
          "containerPort": 80,
          "hostPort": 80
        }
      ],
      "logConfiguration": {
        "logDriver": "awslogs",
        "options": {
          "awslogs-group": "awslogs-wordpress",
          "awslogs-region": "us-west-2",
          "awslogs-stream-prefix": "awslogs-example"
        }
      },
      "memory": 500,
      "cpu": 10
    },
    {
      "environment": [
        {
          "name": "MYSQL_ROOT_PASSWORD",
          "value": "password"
        }
      ],
      "name": "mysql",
      "image": "mysql",
      "cpu": 10,
      "memory": 500,
      "essential": true,
      "logConfiguration": {
        "logDriver": "awslogs",
        "options": {
          "awslogs-group": "awslogs-mysql",
          "awslogs-region": "us-west-2",
          "awslogs-stream-prefix": "awslogs-example"
        }
      }
    }
  ],
  "family": "awslogs-example"
}
```



```
}
```

In the Amazon ECS console, the log configuration for the `wordpress` container is specified as shown in the image below.

Log configuration ☐ Auto-configure CloudWatch Logs

Log driver awslogs ▾

Log options

Key

awslogs-group

awslogs-region

awslogs-stream-prefix

Add key

Viewing awslogs Container Logs in CloudWatch Logs

After your Fargate tasks that use the `awslogs` log driver have launched, your configured containers should be sending their log data to CloudWatch Logs. You can view and search these logs in the console.

To view your CloudWatch Logs data for a container from the Amazon ECS console

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. On the **Clusters** page, select the cluster that contains the task to view.
3. On the **Cluster: *cluster_name*** page, choose **Tasks** and select the task to view.
4. On the **Task: *task_id*** page, expand the container view by choosing the arrow to the left of the container name.
5. In the **Log Configuration** section, choose **View logs in CloudWatch**, which opens the associated log stream in the CloudWatch console.

Log Configuration	
Log driver: awslogs View logs in CloudWatch	
Key	Value
awslogs-group	awslogs-wordpress
awslogs-region	ap-northeast-1
awslogs-stream-prefix	awslogs-example

To view your CloudWatch Logs data in the CloudWatch console

1. Open the CloudWatch console at <https://console.aws.amazon.com/cloudwatch/>.
2. In the left navigation pane, choose **Logs**.
3. Select a log group to view. You should see the log groups that you created in [Creating a Log Group](#) (p. 70).

Create Metric Filter

Actions ▾

Filter:

Log Group Name Prefix

×

Log Groups

☐

awslogs-mysql

☐

awslogs-wordpress

4. Choose a log stream to view.

Filter events		
	Time (UTC -07:00)	Message
2016-09-09		
No older events found at the		
▶	12:56:47	WordPress not found in /var/www/html -
▶	12:56:47	Complete! WordPress has been success
▶	12:56:49	AH00558: apache2: Could not reliably d
▶	12:56:49	AH00558: apache2: Could not reliably d
▶	12:56:49	[Fri Sep 09 19:56:49.059245 2016] [mpm
▶	12:56:49	[Fri Sep 09 19:56:49.059273 2016] [core
▶	13:06:55	52.90.111.181 - - [09/Sep/2016:20:06:55
▶	13:06:56	52.90.111.181 - - [09/Sep/2016:20:06:55
▶	13:06:56	52.90.111.181 - - [09/Sep/2016:20:06:56
▶	13:06:57	54.210.246.190 - - [09/Sep/2016:20:06:5

Custom Log Routing

FireLens for Amazon ECS enables you to use task definition parameters to route logs to an AWS service or AWS Partner Network (APN) destination for log storage and analytics. FireLens works with [Fluentd](#) and [Fluent Bit](#). We provide the AWS for Fluent Bit image or you can use your own Fluentd or Fluent Bit image.

Creating Amazon ECS task definitions with a FireLens configuration is supported using the AWS SDKs, AWS CLI, and AWS Management Console.

Considerations

The following should be considered when using FireLens for Amazon ECS:

- FireLens for Amazon ECS is supported for tasks using both the Fargate and EC2 launch types.
- FireLens for Amazon ECS is supported in AWS CloudFormation templates. For more information, see [AWS::ECS::TaskDefinition FirelensConfiguration](#) in the *AWS CloudFormation User Guide*
- For tasks that use the `bridge` network mode, the container with the FireLens configuration must start before any application containers that rely on it start. To control the start order of your containers, use dependency conditions in your task definition. For more information, see [Container Dependency](#) (p. 51).

Note

If you use dependency condition parameters in container definitions with a FireLens configuration, ensure that each container has a `START` or `HEALTHY` condition requirement.

Required IAM Permissions

To use this feature, you must create an IAM role for your tasks that provides the permissions necessary to use any AWS services that the tasks require. For example, if a container is routing logs to Kinesis Data Firehose, then the task would require permission to call the `firehose:PutRecordBatch` API. For more information, see [Adding and Removing IAM Identity Permissions](#) in the *IAM User Guide*.

The following example IAM policy adds the required permissions for routing logs to Kinesis Data Firehose.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "firehose:PutRecordBatch"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

Your task may also require the Amazon ECS task execution role under the following conditions. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

- If your task uses the Fargate launch type and you are pulling container images from Amazon ECR or referencing sensitive data from AWS Secrets Manager in your log configuration, then you must include the task execution IAM role.
- If you are specifying a custom configuration file that is hosted in Amazon S3, your task execution IAM role must include the `s3:GetObject` permission for the configuration file and the `s3:GetBucketLocation` permission on the Amazon S3 bucket that the file is in. For more information, see [Specifying Permissions in a Policy](#) in the *Amazon Simple Storage Service Console User Guide*.

The following example IAM policy adds the required permissions for retrieving a file from Amazon S3. Specify the name of your Amazon S3 bucket and configuration file name.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject"
      ],
      "Resource": [
        "arn:aws:s3:::examplebucket/folder_name/config_file_name"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
```

```

        "s3:GetBucketLocation"
      ],
      "Resource": [
        "arn:aws:s3:::examplebucket"
      ]
    }
  ]
}

```

Using the AWS for Fluent Bit Image

AWS provides a Fluent Bit image with plugins for both CloudWatch Logs and Kinesis Data Firehose. We recommend using Fluent Bit as your log router because it has a lower resource utilization rate than Fluentd. For more information, see [CloudWatch Logs for Fluent Bit](#) and [Amazon Kinesis Firehose for Fluent Bit](#).

The **AWS for Fluent Bit** image is available on [Docker Hub](#). However, we recommend that you use the following images in Amazon ECR because they provide higher availability.

Region Name	Region	Image URI
US East (N. Virginia)	us-east-1	906394416424.dkr.ecr.us-east-1.amazonaws.com/aws-for-fluent-bit:latest
US East (Ohio)	us-east-2	906394416424.dkr.ecr.us-east-2.amazonaws.com/aws-for-fluent-bit:latest
US West (N. California)	us-west-1	906394416424.dkr.ecr.us-west-1.amazonaws.com/aws-for-fluent-bit:latest
US West (Oregon)	us-west-2	906394416424.dkr.ecr.us-west-2.amazonaws.com/aws-for-fluent-bit:latest
Asia Pacific (Hong Kong)	ap-east-1	449074385750.dkr.ecr.ap-east-1.amazonaws.com/aws-for-fluent-bit:latest
Asia Pacific (Mumbai)	ap-south-1	906394416424.dkr.ecr.ap-south-1.amazonaws.com/aws-for-fluent-bit:latest
Asia Pacific (Seoul)	ap-northeast-2	906394416424.dkr.ecr.ap-northeast-2.amazonaws.com/aws-for-fluent-bit:latest
Asia Pacific (Singapore)	ap-southeast-1	906394416424.dkr.ecr.ap-southeast-1.amazonaws.com/

Region Name	Region	Image URI
		aws-for-fluent-bit:latest
Asia Pacific (Sydney)	ap-southeast-2	906394416424.dkr.ecr.ap-southeast-2.amazonaws.com/aws-for-fluent-bit:latest
Asia Pacific (Tokyo)	ap-northeast-1	906394416424.dkr.ecr.ap-northeast-1.amazonaws.com/aws-for-fluent-bit:latest
Canada (Central)	ca-central-1	906394416424.dkr.ecr.ca-central-1.amazonaws.com/aws-for-fluent-bit:latest
Europe (Frankfurt)	eu-central-1	906394416424.dkr.ecr.eu-central-1.amazonaws.com/aws-for-fluent-bit:latest
Europe (Ireland)	eu-west-1	906394416424.dkr.ecr.eu-west-1.amazonaws.com/aws-for-fluent-bit:latest
Europe (London)	eu-west-2	906394416424.dkr.ecr.eu-west-2.amazonaws.com/aws-for-fluent-bit:latest
Europe (Paris)	eu-west-3	906394416424.dkr.ecr.eu-west-3.amazonaws.com/aws-for-fluent-bit:latest
Europe (Stockholm)	eu-north-1	906394416424.dkr.ecr.eu-north-1.amazonaws.com/aws-for-fluent-bit:latest
Middle East (Bahrain)	me-south-1	741863432321.dkr.ecr.me-south-1.amazonaws.com/aws-for-fluent-bit:latest
South America (São Paulo)	sa-east-1	906394416424.dkr.ecr.sa-east-1.amazonaws.com/aws-for-fluent-bit:latest

Region Name	Region	Image URI
AWS GovCloud (US-East)	us-gov-east-1	161423150738.dkr.ecr.us-gov-east-1.amazonaws.com/aws-for-fluent-bit:latest
AWS GovCloud (US-West)	us-gov-west-1	161423150738.dkr.ecr.us-gov-west-1.amazonaws.com/aws-for-fluent-bit:latest
China (Beijing)	cn-north-1	128054284489.dkr.ecr.cn-north-1.amazonaws.com.cn/aws-for-fluent-bit:latest
China (Ningxia)	cn-northwest-1	128054284489.dkr.ecr.cn-northwest-1.amazonaws.com.cn/aws-for-fluent-bit:latest

Creating a Task Definition that Uses a FireLens Configuration

To use custom log routing with FireLens you must specify the following in your task definition:

- A log router container containing a FireLens configuration. This container must be marked as essential.
- One or more application containers that contain a log configuration specifying the `awsfirelens` log driver.
- A task IAM role ARN containing the permissions needed for the task to route the logs.

When creating a new task definition using the AWS Management Console, there is a FireLens integration section that makes it easy to add a log router container. For more information, see [Creating a Task Definition](#) (p. 30).

Amazon ECS converts the log configuration and generates the Fluentd or Fluent Bit output configuration. The output configuration is mounted in the log routing container at `/fluent-bit/etc/fluent-bit.conf` for Fluent Bit and `/fluentd/etc/fluent.conf` for Fluentd.

To demonstrate how this works, the following is an example task definition example containing a log router container that uses Fluent Bit to route its logs to CloudWatch Logs and an application container that uses a log configuration to route logs to Amazon Kinesis Data Firehose.

```
{
  "family": "firelens-example-firehose",
  "taskRoleArn": "arn:aws:iam::123456789012:role/ecs_task_iam_role",
  "containerDefinitions": [
    {
      "essential": true,
      "image": "906394416424.dkr.ecr.us-west-2.amazonaws.com/aws-for-fluent-bit:latest",
      "name": "log_router",
      "firelensConfiguration": {
```

```
"type": "fluentbit"
},
"logConfiguration": {
  "logDriver": "awslogs",
  "options": {
    "awslogs-group": "firelens-container",
    "awslogs-region": "us-west-2",
    "awslogs-create-group": "true",
    "awslogs-stream-prefix": "firelens"
  }
},
"memoryReservation": 50
},
{
  "essential": true,
  "image": "httpd",
  "name": "app",
  "logConfiguration": {
    "logDriver": "awsfirelens",
    "options": {
      "Name": "firehose",
      "region": "us-west-2",
      "delivery_stream": "my-stream"
    }
  },
  "memoryReservation": 100
}
]
```

The key-value pairs specified as options in the `logConfiguration` object are used to generate the Fluentd or Fluent Bit output configuration. The following is a code example from a Fluent Bit output definition.

```
[OUTPUT]
Name      firehose
Match     app-firelens*
region    us-west-2
delivery_stream my-stream
```

Note

FireLens manages the match configuration. This configuration is not specified in your task definition.

Using ECS Metadata

When specifying a FireLens configuration in a task definition, you can optionally toggle the value for `enable-ecs-log-metadata`. By default, Amazon ECS adds additional fields in your log entries that help identify the source of the logs. You can disable this action by setting `enable-ecs-log-metadata` to `false`.

- `ecs_cluster` – The name of the cluster that the task is part of.
- `ecs_task_arn` – The full ARN of the task that the container is part of.
- `ecs_task_definition` – The task definition name and revision that the task is using.

The following shows the syntax required when specifying an Amazon ECS log metadata setting value:

```
{
  "containerDefinitions": [
    {
```



```
        "essential":true,
        "image":"906394416424.dkr.ecr.us-west-2.amazonaws.com/aws-for-fluent-bit:latest",
        "name":"log_router",
        "firelensConfiguration":{
            "type":"fluentbit",
            "options":{
                "enable-ecs-log-metadata":"true | false"
            }
        }
    }
}
]
```

Specifying a Custom Configuration File

In addition to the auto-generated configuration file that FireLens creates on your behalf, you can also specify a custom configuration file. The configuration file format is the native format for the log router you're using. For more information, see [Fluentd Config File Syntax](#) and [Fluent Bit Configuration Schema](#).

In your custom configuration file, for tasks using the `bridge` or `awsvpc` network mode, you should not set a Fluentd or Fluent Bit forward input over TCP because FireLens will add it to the input configuration.

Your FireLens configuration must contain the following options to specify a custom configuration file:

`config-file-type`

The source location of the custom configuration file. The available options are `s3` or `file`.

`config-file-value`

The source for the custom configuration file. If the `s3` config file type is used, the config file value is the full ARN of the Amazon S3 bucket and file. If the `file` config file type is used, the config file value is the full path of the configuration file that exists either in the container image or on a volume that is mounted in the container.

Important

When using a custom configuration file, you must specify a different path than the one FireLens uses. Amazon ECS reserves the `/fluent-bit/etc/fluent-bit.conf` filepath for Fluent Bit and `/fluentd/etc/fluent.conf` for Fluentd.

The following example shows the syntax required when specifying a custom configuration.

Important

To specify a custom configuration file that is hosted in Amazon S3, ensure you have created a task execution IAM role with the proper permissions. For more information, see [Required IAM Permissions \(p. 76\)](#).

The following shows the syntax required when specifying a custom configuration:

```
{
  "containerDefinitions":[
    {
      "essential":true,
      "image":"906394416424.dkr.ecr.us-west-2.amazonaws.com/aws-for-fluent-bit:latest",
      "name":"log_router",
      "firelensConfiguration":{
        "type":"fluentbit",
        "options":{
          "config-file-type":"s3 | file",
          "config-file-value":"arn:aws:s3:::mybucket/fluent.conf | filepath"
        }
      }
    }
  ]
}
```

```
    ]  
  }
```

Note

For tasks using the Fargate launch type, the only supported `config-file-type` value is `file`.

Using Fluent Logger Libraries

When the `awsfirelens` log driver is specified in a task definition, the ECS agent injects the following environment variables into the container:

`FLUENT_HOST`

The IP address assigned to the FireLens container.

`FLUENT_PORT`

The port that the Fluent Forward protocol is listening on.

The `FLUENT_HOST` and `FLUENT_PORT` environment variables enable you to log directly to the log router from code instead of going through `stdout`. For more information, see [fluent-logger-golang](#) on GitHub.

Filtering Logs Using Regular Expressions

Fluentd and Fluent Bit both support filtering of logs based on their content. FireLens provides a simple method for enabling this filtering. In the log configuration options in a container definition, you can specify the special keys `exclude-pattern` and `include-pattern` that take regular expressions as their values. The `exclude-pattern` key causes all logs that match its regular expression to be dropped. With `include-pattern`, only logs that match its regular expression are sent. These keys can be used together.

The following example demonstrates how to use this filter.

```
{  
  "containerDefinitions": [  
    {  
      "logConfiguration": {  
        "logDriver": "awsfirelens",  
        "options": {  
          "@type": "cloudwatch_logs",  
          "log_group_name": "firelens-testing",  
          "auto_create_stream": "true",  
          "use_tag_as_stream": "true",  
          "region": "us-west-2",  
          "exclude-pattern": "^([a-z][aeiou]).*$",  
          "include-pattern": "^.*[aeiou]$"   
        }  
      }  
    }  
  ]  
}
```

Example Task Definitions

The following are some example task definitions demonstrating common log routing options. For more examples, see [Amazon ECS FireLens Examples](#) on GitHub.

Topics

- [Forwarding Logs to CloudWatch Logs \(p. 83\)](#)

- [Forwarding Logs to an Amazon Kinesis Data Firehose Delivery Stream \(p. 83\)](#)
- [Forwarding to an External Fluentd or Fluent Bit \(p. 84\)](#)

Forwarding Logs to CloudWatch Logs

The following task definition example demonstrates how to specify a log configuration that forwards logs to a CloudWatch Logs log group. For more information, see [What Is Amazon CloudWatch Logs?](#) in the *Amazon CloudWatch Logs User Guide*.

In the log configuration options, specify the log group name and the Region it exists in. To have Fluent Bit create the log group on your behalf, specify `"auto_create_group": "true"`. You can also specify a log stream prefix, which assists in filtering. For more information, see [Fluent Bit Plugin for CloudWatch Logs](#).

```
{
  "family": "firelens-example-cloudwatch",
  "taskRoleArn": "arn:aws:iam::123456789012:role/ecs_task_iam_role",
  "containerDefinitions": [
    {
      "essential": true,
      "image": "906394416424.dkr.ecr.us-west-2.amazonaws.com/aws-for-fluent-bit:latest",
      "name": "log_router",
      "firelensConfiguration": {
        "type": "fluentbit"
      },
      "logConfiguration": {
        "logDriver": "awslogs",
        "options": {
          "awslogs-group": "firelens-container",
          "awslogs-region": "us-west-2",
          "awslogs-create-group": "true",
          "awslogs-stream-prefix": "firelens"
        }
      },
      "memoryReservation": 50
    },
    {
      "essential": true,
      "image": "nginx",
      "name": "app",
      "logConfiguration": {
        "logDriver": "awsfirelens",
        "options": {
          "Name": "cloudwatch",
          "region": "us-west-2",
          "log_group_name": "firelens-testing-fluent-bit",
          "auto_create_group": "true",
          "log_stream_prefix": "from-fluent-bit"
        }
      },
      "memoryReservation": 100
    }
  ]
}
```

Forwarding Logs to an Amazon Kinesis Data Firehose Delivery Stream

The following task definition example demonstrates how to specify a log configuration that forwards logs to an Amazon Kinesis Data Firehose delivery stream. The Kinesis Data Firehose delivery stream must

already exist. For more information, see [Creating an Amazon Kinesis Data Firehose Delivery Stream](#) in the *Amazon Kinesis Data Firehose Developer Guide*.

In the log configuration options, specify the delivery stream name and the Region it exists in. For more information, see [Fluent Bit Plugin for Amazon Kinesis Firehose](#).

```
{
  "family": "firelens-example-firehose",
  "taskRoleArn": "arn:aws:iam::123456789012:role/ecs_task_iam_role",
  "containerDefinitions": [
    {
      "essential": true,
      "image": "906394416424.dkr.ecr.us-west-2.amazonaws.com/aws-for-fluent-bit:latest",
      "name": "log_router",
      "firelensConfiguration": {
        "type": "fluentbit"
      },
      "logConfiguration": {
        "logDriver": "awslogs",
        "options": {
          "awslogs-group": "firelens-container",
          "awslogs-region": "us-west-2",
          "awslogs-create-group": "true",
          "awslogs-stream-prefix": "firelens"
        }
      },
      "memoryReservation": 50
    },
    {
      "essential": true,
      "image": "httpd",
      "name": "app",
      "logConfiguration": {
        "logDriver": "awsfirelens",
        "options": {
          "Name": "firehose",
          "region": "us-west-2",
          "delivery_stream": "my-stream"
        }
      },
      "memoryReservation": 100
    }
  ]
}
```

Forwarding to an External Fluentd or Fluent Bit

The following task definition example demonstrates how to specify a log configuration that forwards logs to an external Fluentd or Fluent Bit host. Specify the host and port for your environment.

```
{
  "family": "firelens-example-forward",
  "taskRoleArn": "arn:aws:iam::123456789012:role/ecs_task_iam_role",
  "containerDefinitions": [
    {
      "essential": true,
      "image": "906394416424.dkr.ecr.us-west-2.amazonaws.com/aws-for-fluent-bit:latest",
      "name": "log_router",
      "firelensConfiguration": {
        "type": "fluentbit"
      },
      "logConfiguration": {
        "logDriver": "awslogs",
```

```
"options": {
  "awslogs-group": "firelens-container",
  "awslogs-region": "us-west-2",
  "awslogs-create-group": "true",
  "awslogs-stream-prefix": "firelens"
},
"memoryReservation": 50
},
{
  "essential": true,
  "image": "httpd",
  "name": "app",
  "logConfiguration": {
    "logDriver": "awsfirelens",
    "options": {
      "Name": "forward",
      "Host": "fluentdhost",
      "Port": "24224"
    }
  },
  "memoryReservation": 100
}
]
```

Private Registry Authentication for Tasks

Private registry authentication for tasks using AWS Secrets Manager enables you to store your credentials securely and then reference them in your container definition. This allows your tasks to use images from private repositories. This feature is supported by tasks using both the Fargate or EC2 launch types.

Important

If your task definition references an image stored in Amazon ECR, this topic does not apply. For more information, see [Using Amazon ECR Images with Amazon ECS](#) in the *Amazon Elastic Container Registry User Guide*.

For tasks using the Fargate launch type, this feature requires platform version 1.2.0 or later. For information, see [AWS Fargate platform versions \(p. 14\)](#).

Within your container definition, specify `repositoryCredentials` with the full ARN of the secret that you created. The secret you reference can be from a different Region than the task using it, but must be from within the same account.

Note

When using the Amazon ECS API, AWS CLI, or AWS SDK, if the secret exists in the same Region as the task you are launching then you can use either the full ARN or name of the secret. When using the AWS Management Console, the full ARN of the secret must be specified.

The following is a snippet of a task definition showing the required parameters:

```
"containerDefinitions": [
  {
    "image": "private-repo/private-image",
    "repositoryCredentials": {
      "credentialsParameter":
        "arn:aws:secretsmanager:region:aws_account_id:secret:secret_name"
    }
  }
]
```

Required IAM Permissions for Private Registry Authentication

The Amazon ECS task execution role is required to use this feature. This allows the container agent to pull the container image. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

To provide access to the secrets that you create, manually add the following permissions as an inline policy to the task execution role. For more information, see [Adding and Removing IAM Policies](#).

- `secretsmanager:GetSecretValue`
- `kms:Decrypt`—Required only if your key uses a custom KMS key and not the default key. The ARN for your custom key should be added as a resource.

An example inline policy adding the permissions is shown below.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:Decrypt",
        "secretsmanager:GetSecretValue"
      ],
      "Resource": [
        "arn:aws:secretsmanager:<region>:<aws_account_id>:secret:secret_name",
        "arn:aws:kms:<region>:<aws_account_id>:key/key_id"
      ]
    }
  ]
}
```

Enabling Private Registry Authentication

To create a basic secret

Use AWS Secrets Manager to create a secret for your private registry credentials.

1. Open the AWS Secrets Manager console at <https://console.aws.amazon.com/secretsmanager/>.
2. Choose **Store a new secret**.
3. For **Select secret type**, choose **Other type of secrets**.
4. Select **Plaintext** and enter your private registry credentials using the following format:

```
{
  "username" : "privateRegistryUsername",
  "password" : "privateRegistryPassword"
}
```

5. Choose **Next**.
6. For **Secret name**, type an optional path and name, such as **production/MyAwesomeAppSecret** or **development/TestSecret**, and choose **Next**. You can optionally add a description to help you remember the purpose of this secret later.

The secret name must be ASCII letters, digits, or any of the following characters: `/_+=.@-`

7. (Optional) At this point, you can configure rotation for your secret. For this procedure, leave it at **Disable automatic rotation** and choose **Next**.

For information about how to configure rotation on new or existing secrets, see [Rotating Your AWS Secrets Manager Secrets](#).

8. Review your settings, and then choose **Store secret** to save everything you entered as a new secret in Secrets Manager.

To create a task definition that uses private registry authentication

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation pane, choose **Task Definitions**.
3. On the **Task Definitions** page, choose **Create new Task Definition**.
4. On the **Select launch type compatibility** page, choose the launch type for your tasks and then **Next step**.
5. For **Task Definition Name**, type a name for your task definition. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
6. For **Task execution role**, either select your existing task execution role or choose **Create new role** to have one created for you. This role authorizes Amazon ECS to pull private images for your task. For more information, see [Required IAM Permissions for Private Registry Authentication \(p. 86\)](#).

Important

If the **Task execution role** field does not appear, choose **Configure via JSON** and manually add the `executionRoleArn` field to specify your task execution role. The following shows the syntax:

```
"executionRoleArn": "arn:aws:iam::aws_account_id:role/ecsTaskExecutionRole"
```

7. For each container to create in your task definition, complete the following steps:
 - a. In the **Container Definitions** section, choose **Add container**.
 - b. For **Container name**, type a name for your container. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
 - c. For **Image**, type the image name or path to your private image. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
 - d. Select the **Private repository authentication** option.
 - e. For **Secrets manager ARN**, enter the full Amazon Resource Name (ARN) of the secret that you created earlier. The value must be between 20 and 2048 characters.
 - f. Fill out the remaining required fields and any optional fields to use in your container definitions. More container definition parameters are available in the **Advanced container configuration** menu. For more information, see [Task definition parameters \(p. 35\)](#).
 - g. Choose **Add**.
8. When your containers are added, choose **Create**.

Specifying Sensitive Data

Amazon ECS enables you to inject sensitive data into your containers by storing your sensitive data in either AWS Secrets Manager secrets or AWS Systems Manager Parameter Store parameters and then referencing them in your container definition.

Secrets can be exposed to a container in the following ways:

- To inject sensitive data into your containers as environment variables, use the `secrets` container definition parameter.

- To reference sensitive information in the log configuration of a container, use the `secretOptions` container definition parameter.

Topics

- [Specifying Sensitive Data Using Secrets Manager \(p. 88\)](#)
- [Specifying Sensitive Data Using Systems Manager Parameter Store \(p. 93\)](#)

Specifying Sensitive Data Using Secrets Manager

Amazon ECS enables you to inject sensitive data into your containers by storing your sensitive data in AWS Secrets Manager secrets and then referencing them in your container definition. Sensitive data stored in Secrets Manager secrets can be exposed to a container as environment variables or as part of the log configuration.

When you inject a secret as an environment variable, you can specify a JSON key or version of a secret to inject. This process helps you control the sensitive data exposed to your container. For more information about secret versioning, see [Key Terms and Concepts for AWS Secrets Manager](#) in the *AWS Secrets Manager User Guide*.

Topics

- [Considerations for Specifying Sensitive Data Using Secrets Manager \(p. 88\)](#)
- [Required IAM Permissions for Amazon ECS Secrets \(p. 88\)](#)
- [Injecting Sensitive Data as an Environment Variable \(p. 89\)](#)
- [Injecting Sensitive Data in a Log Configuration \(p. 92\)](#)
- [Creating a Task Definition that References Sensitive Data \(p. 92\)](#)

Considerations for Specifying Sensitive Data Using Secrets Manager

The following should be considered when using Secrets Manager to specify sensitive data for containers.

- For tasks that use the Fargate launch type, the following should be considered:
 - It is only supported to inject the full contents of a secret as an environment variable. Specifying a specific JSON key or version is not supported at this time.
 - To inject the full content of a secret as an environment variable or in a log configuration, you must use platform version 1.3.0 or later. For information, see [AWS Fargate platform versions \(p. 14\)](#).
- When using a task definition that references Secrets Manager secrets to retrieve sensitive data for your containers, if you are also using interface VPC endpoints, you must create the interface VPC endpoints for Secrets Manager. For more information, see [Using Secrets Manager with VPC Endpoints](#) in the *AWS Secrets Manager User Guide*.
- Sensitive data is injected into your container when the container is initially started. If the secret is subsequently updated or rotated, the container will not receive the updated value automatically. You must either launch a new task or if your task is part of a service you can update the service and use the **Force new deployment** option to force the service to launch a fresh task.

Required IAM Permissions for Amazon ECS Secrets

To use this feature, you must have the Amazon ECS task execution role and reference it in your task definition. This allows the container agent to pull the necessary Secrets Manager resources. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

To provide access to the Secrets Manager secrets that you create, manually add the following permissions as an inline policy to the task execution role. For more information, see [Adding and Removing IAM Policies](#).

- `secretsmanager:GetSecretValue`—Required if you are referencing a Secrets Manager secret.
- `kms:Decrypt`—Required only if your secret uses a custom KMS key and not the default key. The ARN for your custom key should be added as a resource.

The following example inline policy adds the required permissions.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "secretsmanager:GetSecretValue",
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:secretsmanager:<region>:<aws_account_id>:secret:<secret_name>",
        "arn:aws:kms:<region>:<aws_account_id>:key/<key_id>"
      ]
    }
  ]
}
```

Injecting Sensitive Data as an Environment Variable

Within your container definition, you can specify the following:

- The `secrets` object containing the name of the environment variable to set in the container
- The Amazon Resource Name (ARN) of the Secrets Manager secret
- Additional parameters that contain the sensitive data to present to the container

The following example shows the full syntax that must be specified for the Secrets Manager secret.

```
arn:aws:secretsmanager:<region>:<aws_account_id>:secret:<secret-name>:json-key:version-
stage:version-id
```

Important

If you are using AWS Fargate, it is only supported to specify the full ARN of the secret in your task definition. Specifying a specific JSON key or version is not supported at this time.

The following section describes the additional parameters. These parameters are optional, but if you do not use them, you must include the colons `:` to use the default values. Examples are provided below for more context.

`json-key`

Specifies the name of the key in a key-value pair with the value that you want to set as the environment variable value. Only values in JSON format are supported. If you do not specify a JSON key, then the full contents of the secret is used.

version-stage

Specifies the staging label of the version of a secret that you want to use. If a version staging label is specified, you cannot specify a version ID. If no version stage is specified, the default behavior is to retrieve the secret with the `AWSCURRENT` staging label.

Staging labels are used to keep track of different versions of a secret when they are either updated or rotated. Each version of a secret has one or more staging labels and an ID. For more information, see [Key Terms and Concepts for AWS Secrets Manager](#) in the *AWS Secrets Manager User Guide*.

version-id

Specifies the unique identifier of the version of a secret that you want to use. If a version ID is specified, you cannot specify a version staging label. If no version ID is specified, the default behavior is to retrieve the secret with the `AWSCURRENT` staging label.

Version IDs are used to keep track of different versions of a secret when they are either updated or rotated. Each version of a secret has an ID. For more information, see [Key Terms and Concepts for AWS Secrets Manager](#) in the *AWS Secrets Manager User Guide*.

For a full tutorial on creating a Secrets Manager secret and injecting it into a container as an environment variable, see [Tutorial: Specifying Sensitive Data Using Secrets Manager Secrets](#) (p. 312).

Example Container Definitions

The following examples show ways in which you can reference Secrets Manager secrets in your container definitions.

Example referencing a full secret

The following is a snippet of a task definition showing the format when referencing the full text of a Secrets Manager secret.

```
{
  "containerDefinitions": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:secret_name-AbCdEf"
    }]
  }]
}
```

Example referencing a specific key within a secret

The following shows an example output from a `get-secret-value` command that displays the contents of a secret along with the version staging label and version ID associated with it.

```
{
  "ARN": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-AbCdEf",
  "Name": "appauthexample",
  "VersionId": "871d9eca-18aa-46a9-8785-981dd39ab30c",
  "SecretString": "{\"username1\":\"password1\",\"username2\":\"password2\", \"username3\":\"password3\"}",
  "VersionStages": [
    "AWSCURRENT"
  ],
  "CreateDate": 1581968848.921
}
```

Reference a specific key from the previous output in a container definition by specifying the key name at the end of the ARN.

```
{
  "containerDefinitions": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf:username1::"
    }]
  }]
}
```

Example referencing a specific secret version

The following shows an example output from a [describe-secret](#) command that displays the unencrypted contents of a secret along with the metadata for all versions of the secret.

```
{
  "ARN": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-AbCdEf",
  "Name": "appauthexample",
  "Description": "Example of a secret containing application authorization data.",
  "RotationEnabled": false,
  "LastChangedDate": 1581968848.926,
  "LastAccessedDate": 1581897600.0,
  "Tags": [],
  "VersionIdsToStages": {
    "871d9eca-18aa-46a9-8785-981dd39ab30c": [
      "AWSCURRENT"
    ],
    "9d4cb84b-ad69-40c0-a0ab-cead36b967e8": [
      "AWSPREVIOUS"
    ]
  }
}
```

Reference a specific version staging label from the previous output in a container definition by specifying the key name at the end of the ARN.

```
{
  "containerDefinitions": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf::AWSPREVIOUS:"
    }]
  }]
}
```

Reference a specific version ID from the previous output in a container definition by specifying the key name at the end of the ARN.

```
{
  "containerDefinitions": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf::9d4cb84b-ad69-40c0-a0ab-cead36b967e8"
    }]
  }]
}
```

Example referencing a specific key and version staging label of a secret

The following shows how to reference both a specific key within a secret and a specific version staging label.

```
{
  "containerDefinitions": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf:username1:AWSPREVIOUS:"
    }]
  }]
}
```

To specify a specific key and version ID, use the following syntax.

```
{
  "containerDefinitions": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:appauthexample-
AbCdEf:username1::9d4cb84b-ad69-40c0-a0ab-cead36b967e8"
    }]
  }]
}
```

Injecting Sensitive Data in a Log Configuration

Within your container definition, when specifying a `logConfiguration` you can specify `secretOptions` with the name of the log driver option to set in the container and the full ARN of the Secrets Manager secret containing the sensitive data to present to the container.

The following is a snippet of a task definition showing the format when referencing an Secrets Manager secret.

```
{
  "containerDefinitions": [{
    "logConfiguration": [{
      "logDriver": "splunk",
      "options": {
        "splunk-url": "https://cloud.splunk.com:8080"
      },
      "secretOptions": [{
        "name": "splunk-token",
        "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:secret_name-
AbCdEf"
      }]
    }]
  }]
}
```

Creating a Task Definition that References Sensitive Data

You can use the Amazon ECS console to create a task definition that references an Secrets Manager secret.

To create a task definition that specifies a secret

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.

2. In the navigation pane, choose **Task Definitions, Create new Task Definition**.
3. On the **Select launch type compatibility** page, choose the launch type for your tasks and choose **Next step**.
4. For **Task Definition Name**, type a name for your task definition. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
5. For **Task execution role**, either select your existing task execution role or choose **Create new role** to have one created for you. This role authorizes Amazon ECS to pull private images for your task. For more information, see [Required IAM Permissions for Private Registry Authentication \(p. 86\)](#).

Important

If the **Task execution role** field does not appear, choose **Configure via JSON** and manually add the `executionRoleArn` field to specify your task execution role. The following code shows the syntax:

```
"executionRoleArn": "arn:aws:iam::aws\_account\_id:role/ec2TaskExecutionRole"
```

6. For each container to create in your task definition, complete the following steps:
 - a. Under **Container Definitions**, choose **Add container**.
 - b. For **Container name**, type a name for your container. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
 - c. For **Image**, type the image name or path to your private image. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
 - d. Expand **Advanced container configuration**.
 - e. For sensitive data to inject as environment variables, under **Environment**, for **Environment variables**, complete the following fields:
 - i. For **Key**, enter the name of the environment variable to set in the container. This corresponds to the `name` field in the `secrets` section of a container definition.
 - ii. For **Value**, choose **ValueFrom**. For **Add value**, enter the ARN of the Secrets Manager secret that contains the data to present to your container as an environment variable.
 - f. For sensitive data referenced in the log configuration for a container, under **Storage and Logging**, for **Log configuration**, complete the following fields:
 - i. Clear the **Auto-configure CloudWatch Logs** option.
 - ii. Under **Log options**, for **Key**, enter the name of the log configuration option to set.
 - iii. For **Value**, choose **ValueFrom**. For **Add value**, enter the full ARN of the Secrets Manager secret that contains the data to present to your log configuration as a log option.
 - g. Fill out the remaining required fields and any optional fields to use in your container definitions. More container definition parameters are available in the **Advanced container configuration** menu. For more information, see [Task definition parameters \(p. 35\)](#).
 - h. Choose **Add**.
7. When your containers are added, choose **Create**.

Specifying Sensitive Data Using Systems Manager Parameter Store

Amazon ECS enables you to inject sensitive data into your containers by storing your sensitive data in AWS Systems Manager Parameter Store parameters and then referencing them in your container definition.

Topics

- [Considerations for Specifying Sensitive Data Using Systems Manager Parameter Store \(p. 94\)](#)
- [Required IAM Permissions for Amazon ECS Secrets \(p. 94\)](#)
- [Injecting Sensitive Data as an Environment Variable \(p. 95\)](#)
- [Injecting Sensitive Data in a Log Configuration \(p. 95\)](#)
- [Creating an AWS Systems Manager Parameter Store Parameter \(p. 95\)](#)
- [Creating a Task Definition that References Sensitive Data \(p. 96\)](#)

Considerations for Specifying Sensitive Data Using Systems Manager Parameter Store

The following should be considered when specifying sensitive data for containers using Systems Manager Parameter Store parameters.

- For tasks that use the Fargate launch type, this feature requires that your task use platform version 1.3.0 or later. For information, see [AWS Fargate platform versions \(p. 14\)](#).
- Sensitive data is injected into your container when the container is initially started. If the secret or Parameter Store parameter is subsequently updated or rotated, the container will not receive the updated value automatically. You must either launch a new task or if your task is part of a service you can update the service and use the **Force new deployment** option to force the service to launch a fresh task.

Required IAM Permissions for Amazon ECS Secrets

To use this feature, you must have the Amazon ECS task execution role and reference it in your task definition. This allows the container agent to pull the necessary AWS Systems Manager resources. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

To provide access to the AWS Systems Manager Parameter Store parameters that you create, manually add the following permissions as an inline policy to the task execution role. For more information, see [Adding and Removing IAM Policies](#).

- `ssm:GetParameters`—Required if you are referencing a Systems Manager Parameter Store parameter in a task definition.
- `secretsmanager:GetSecretValue`—Required if you are referencing a Secrets Manager secret either directly or if your Systems Manager Parameter Store parameter is referencing a Secrets Manager secret in a task definition.
- `kms:Decrypt`—Required only if your secret uses a custom KMS key and not the default key. The ARN for your custom key should be added as a resource.

The following example inline policy adds the required permissions:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ssm:GetParameters",
        "secretsmanager:GetSecretValue",
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:ssm:<region>:<aws_account_id>:parameter/<parameter_name>",
        "arn:aws:secretsmanager:<region>:<aws_account_id>:secret:<secret_name>"
      ]
    }
  ]
}
```

```
        "arn:aws:kms:<region>:<aws_account_id>:key/<key_id>"
    ]
  }
}
```

Injecting Sensitive Data as an Environment Variable

Within your container definition, specify `secrets` with the name of the environment variable to set in the container and the full ARN of the Systems Manager Parameter Store parameter containing the sensitive data to present to the container.

The following is a snippet of a task definition showing the format when referencing an Systems Manager Parameter Store parameter. If the Systems Manager Parameter Store parameter exists in the same Region as the task you are launching, then you can use either the full ARN or name of the parameter. If the parameter exists in a different Region, then the full ARN must be specified.

```
{
  "containerDefinitions": [{
    "secrets": [{
      "name": "environment_variable_name",
      "valueFrom": "arn:aws:ssm:region:aws_account_id:parameter/parameter_name"
    }]
  }]
}
```

Injecting Sensitive Data in a Log Configuration

Within your container definition, when specifying a `logConfiguration` you can specify `secretOptions` with the name of the log driver option to set in the container and the full ARN of the Systems Manager Parameter Store parameter containing the sensitive data to present to the container.

Important

If the Systems Manager Parameter Store parameter exists in the same Region as the task you are launching, then you can use either the full ARN or name of the parameter. If the parameter exists in a different Region, then the full ARN must be specified.

The following is a snippet of a task definition showing the format when referencing an Systems Manager Parameter Store parameter.

```
{
  "containerDefinitions": [{
    "logConfiguration": [{
      "logDriver": "fluentd",
      "options": {
        "tag": "fluentd demo"
      },
      "secretOptions": [{
        "name": "fluentd-address",
        "valueFrom": "arn:aws:ssm:region:aws_account_id:parameter:parameter_name"
      }]
    }]
  }]
}
```

Creating an AWS Systems Manager Parameter Store Parameter

You can use the AWS Systems Manager console to create a Systems Manager Parameter Store parameter for your sensitive data. For more information, see [Walkthrough: Create and Use a Parameter in a Command \(Console\)](#) in the *AWS Systems Manager User Guide*.

To create a Parameter Store parameter

1. Open the AWS Systems Manager console at <https://console.aws.amazon.com/systems-manager/>.
2. In the navigation pane, choose **Parameter Store**, **Create parameter**.
3. For **Name**, type a hierarchy and a parameter name. For example, type `test/database_password`.
4. For **Description**, type an optional description.
5. For **Type**, choose **String**, **StringList**, or **SecureString**.

Note

- If you choose **SecureString**, the **KMS Key ID** field appears. If you don't provide a KMS CMK ID, a KMS CMK ARN, an alias name, or an alias ARN, then the system uses `alias/aws/ssm`, which is the default KMS CMK for Systems Manager. To avoid using this key, choose a custom key. For more information, see [Use Secure String Parameters](#) in the *AWS Systems Manager User Guide*.
- When you create a secure string parameter in the console by using the `key-id` parameter with either a custom KMS CMK alias name or an alias ARN, you must specify the prefix `alias/` before the alias. The following is an ARN example:

```
arn:aws:kms:us-east-2:123456789012:alias/MyAliasName
```

The following is an alias name example:

```
alias/MyAliasName
```

6. For **Value**, type a value. For example, `MyFirstParameter`. If you chose **SecureString**, the value is masked as you type.
7. Choose **Create parameter**.

Creating a Task Definition that References Sensitive Data

You can use the Amazon ECS console to create a task definition that references a Systems Manager Parameter Store parameter.

To create a task definition that specifies a secret

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation pane, choose **Task Definitions**, **Create new Task Definition**.
3. On the **Select launch type compatibility** page, choose the launch type for your tasks and choose **Next step**.
4. For **Task Definition Name**, type a name for your task definition. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
5. For **Task execution role**, either select your existing task execution role or choose **Create new role** to have one created for you. This role authorizes Amazon ECS to pull private images for your task. For more information, see [Required IAM Permissions for Private Registry Authentication](#) (p. 86).

Important

If the **Task execution role** field does not appear, choose **Configure via JSON** and manually add the `executionRoleArn` field to specify your task execution role. The following code shows the syntax:

```
"executionRoleArn": "arn:aws:iam::aws_account_id:role/ecsTaskExecutionRole"
```

6. For each container to create in your task definition, complete the following steps:

- a. Under **Container Definitions**, choose **Add container**.
- b. For **Container name**, type a name for your container. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
- c. For **Image**, type the image name or path to your private image. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
- d. Expand **Advanced container configuration**.
- e. For sensitive data to inject as environment variables, under **Environment**, for **Environment variables**, complete the following fields:
 - i. For **Key**, enter the name of the environment variable to set in the container. This corresponds to the `name` field in the `secrets` section of a container definition.
 - ii. For **Value**, choose **ValueFrom**. For **Add value**, enter the full ARN of the AWS Systems Manager Parameter Store parameter that contains the data to present to your container as an environment variable.

Note

If the Systems Manager Parameter Store parameter exists in the same Region as the task you are launching, then you can use either the full ARN or name of the secret. If the parameter exists in a different Region, then the full ARN must be specified.

- f. For secrets referenced in the log configuration for a container, under **Storage and Logging**, for **Log configuration**, complete the following fields:
 - i. Clear the **Auto-configure CloudWatch Logs** option.
 - ii. Under **Log options**, for **Key**, enter the name of the log configuration option to set.
 - iii. For **Value**, choose **ValueFrom**. For **Add value**, enter the name or full ARN of the AWS Systems Manager Parameter Store parameter that contains the data to present to your log configuration as a log option.

Note

If the Systems Manager Parameter Store parameter exists in the same Region as the task you are launching, then you can use either the full ARN or the name of the secret. If the parameter exists in a different Region, then the full ARN must be specified.

- g. Fill out the remaining required fields and any optional fields to use in your container definitions. More container definition parameters are available in the **Advanced container configuration** menu. For more information, see [Task definition parameters \(p. 35\)](#).
 - h. Choose **Add**.
7. When your containers are added, choose **Create**.

Example Task Definitions

This section provides some task definition examples that you can use to start creating your own task definitions. For more information, see [Task definition parameters \(p. 35\)](#) and [Creating a Task Definition \(p. 30\)](#).

For additional task definition examples, see [AWS Sample Task Definitions](#) on GitHub.

Topics

- [Example: Webserver \(p. 98\)](#)
- [Example: splunk Log Driver \(p. 98\)](#)
- [Example: fluentd Log Driver \(p. 99\)](#)

- [Example: gelf Log Driver \(p. 99\)](#)
- [Example: Container Dependency \(p. 100\)](#)

Example: Webserver

The following is an example task definition using the Fargate launch type that sets up a web server:

```
{
  "containerDefinitions": [
    {
      "command": [
        "/bin/sh -c \"echo '<html> <head> <title>Amazon ECS Sample App</title>
<style>body {margin-top: 40px; background-color: #333;} </style> </head><body> <div
style=color:white;text-align:center> <h1>Amazon ECS Sample App</h1> <h2>Congratulations!
</h2> <p>Your application is now running on a container in Amazon ECS.</p> </div></body></
html>' > /usr/local/apache2/htdocs/index.html && httpd-foreground\""
      ],
      "entryPoint": [
        "sh",
        "-c"
      ],
      "essential": true,
      "image": "httpd:2.4",
      "logConfiguration": {
        "logDriver": "awslogs",
        "options": {
          "awslogs-group" : "/ecs/fargate-task-definition",
          "awslogs-region": "us-east-1",
          "awslogs-stream-prefix": "ecs"
        }
      },
      "name": "sample-fargate-app",
      "portMappings": [
        {
          "containerPort": 80,
          "hostPort": 80,
          "protocol": "tcp"
        }
      ]
    }
  ],
  "cpu": "256",
  "executionRoleArn": "arn:aws:iam::012345678910:role/ecsTaskExecutionRole",
  "family": "fargate-task-definition",
  "memory": "512",
  "networkMode": "awsvpc",
  "requiresCompatibilities": [
    "FARGATE"
  ]
}
```

Example: splunk Log Driver

The following example demonstrates how to use the splunk log driver in a task definition that sends the logs to a remote service. The Splunk token parameter is specified as a secret option because it can be treated as sensitive data. For more information, see [Specifying Sensitive Data \(p. 87\)](#).

```
"containerDefinitions": [{
  "logConfiguration": {
    "logDriver": "splunk",
```

```
"options": {
  "splunk-url": "https://cloud.splunk.com:8080",
  "tag": "tag_name",
},
"secretOptions": [{
  "name": "splunk-token",
  "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:splunk-token-KnrBkD"
}],
```

Example: fluentd Log Driver

The following example demonstrates how to use the `fluentd` log driver in a task definition that sends the logs to a remote service. The `fluentd-address` value is specified as a secret option as it may be treated as sensitive data. For more information, see [Specifying Sensitive Data \(p. 87\)](#).

```
"containerDefinitions": [{
  "logConfiguration": {
    "logDriver": "fluentd",
    "options": {
      "tag": "fluentd demo"
    },
    "secretOptions": [{
      "name": "fluentd-address",
      "valueFrom": "arn:aws:secretsmanager:region:aws_account_id:secret:fluentd-address-KnrBkD"
    }]
  },
  "entryPoint": [],
  "portMappings": [{
    "hostPort": 80,
    "protocol": "tcp",
    "containerPort": 80
  },
  {
    "hostPort": 24224,
    "protocol": "tcp",
    "containerPort": 24224
  }]
}],
```

Example: gelf Log Driver

The following example demonstrates how to use the `gelf` log driver in a task definition that sends the logs to a remote host running Logstash that takes Gelf logs as an input. For more information, see [logConfiguration \(p. 46\)](#).

```
"containerDefinitions": [{
  "logConfiguration": {
    "logDriver": "gelf",
    "options": {
      "gelf-address": "udp://logstash-service-address:5000",
      "tag": "gelf task demo"
    }
  },
  "entryPoint": [],
  "portMappings": [{
    "hostPort": 5000,
    "protocol": "udp",
    "containerPort": 5000
  }]
}],
```

```
{
  "hostPort": 5000,
  "protocol": "tcp",
  "containerPort": 5000
}
],
}],
```

Example: Container Dependency

This example demonstrates the syntax for a task definition with multiple containers where container dependency is specified. In the following task definition, the `envoy` container must reach a healthy status, determined by the required container healthcheck parameters, before the `app` container will start. For more information, see [Container Dependency \(p. 51\)](#).

```
{
  "family": "appmesh-gateway",
  "proxyConfiguration": {
    "type": "APPMESH",
    "containerName": "envoy",
    "properties": [
      {
        "name": "IgnoredUID",
        "value": "1337"
      },
      {
        "name": "ProxyIngressPort",
        "value": "15000"
      },
      {
        "name": "ProxyEgressPort",
        "value": "15001"
      },
      {
        "name": "AppPorts",
        "value": "9080"
      },
      {
        "name": "EgressIgnoredIPs",
        "value": "169.254.170.2,169.254.169.254"
      }
    ]
  },
  "containerDefinitions": [
    {
      "name": "app",
      "image": "application_image",
      "portMappings": [
        {
          "containerPort": 9080,
          "hostPort": 9080,
          "protocol": "tcp"
        }
      ],
      "essential": true,
      "dependsOn": [
        {
          "containerName": "envoy",
          "condition": "HEALTHY"
        }
      ]
    }
  ],
  {
```

```
    "name": "envoy",
    "image": "840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-
prod",
    "essential": true,
    "environment": [
      {
        "name": "APPMESH_VIRTUAL_NODE_NAME",
        "value": "mesh/meshName/virtualNode/virtualNodeName"
      },
      {
        "name": "ENVOY_LOG_LEVEL",
        "value": "info"
      }
    ],
    "healthCheck": {
      "command": [
        "CMD-SHELL",
        "echo hello"
      ],
      "interval": 5,
      "timeout": 2,
      "retries": 3
    }
  },
  "executionRoleArn": "arn:aws:iam::123456789012:role/ecsTaskExecutionRole",
  "networkMode": "awsvpc"
}
```

Updating a Task Definition

To update a task definition, create a task definition revision. If the task definition is used in a service, you must update that service to use the updated task definition.

To create a task definition revision

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. From the navigation bar, choose the region that contains your task definition.
3. In the navigation pane, choose **Task Definitions**.
4. On the **Task Definitions** page, select the box to the left of the task definition to revise and choose **Create new revision**.
5. On the **Create new revision of Task Definition** page, make changes. For example, to change the existing container definitions (such as the container image, memory limits, or port mappings), select the container, make the changes, and then choose **Update**.
6. Verify the information and choose **Create**.
7. If your task definition is used in a service, update your service with the updated task definition. For more information, see [Updating a Service \(p. 139\)](#).

Deregistering Task Definition Revisions

If you decide that you no longer need a specific task definition revision in Amazon ECS, you can deregister the task definition revision so that it no longer displays in your `ListTaskDefinition` API calls or in the console when you want to run a task or update a service.

When you deregister a task definition revision, it is immediately marked as `INACTIVE`. Existing tasks and services that reference an `INACTIVE` task definition revision continue to run without disruption,

and existing services that reference an `INACTIVE` task definition revision can still scale up or down by modifying the service's desired count.

You cannot use an `INACTIVE` task definition revision to run new tasks or create new services, and you cannot update an existing service to reference an `INACTIVE` task definition revision (although there may be up to a 10-minute window following deregistration where these restrictions have not yet taken effect).

Note

At this time, `INACTIVE` task definition revisions remain discoverable in your account indefinitely; however, this behavior is subject to change in the future, so you should not rely on `INACTIVE` task definition revisions persisting beyond the lifecycle of any associated tasks and services.

Use the following procedure to deregister a task definition revision.

To deregister a task definition revision

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. From the navigation bar, choose the region that contains your task definition.
3. In the navigation pane, choose **Task Definitions**.
4. On the **Task Definitions** page, choose the task definition family that contains one or more revisions that you want to deregister.
5. On the **Task Definition Name** page, select the box to the left of each task definition revision you want to deregister.
6. Choose **Actions, Deregister**.
7. Verify the information in the **Deregister Task Definition** window, and choose **Deregister** to finish.

Account Settings

Amazon ECS provides the following account settings, which enable you to opt in or out of specific features.

Amazon Resource Names (ARNs) and IDS

Resource names: `serviceLongArnFormat`, `taskLongArnFormat`, and `containerInstanceLongArnFormat`

Amazon ECS is introducing a new format for Amazon Resource Names (ARNs) and resource IDs for Amazon ECS services, tasks, and container instances. You must opt in to the new format for each resource type to use features such as resource tagging for that resource type. For more information, see [Amazon Resource Names \(ARNs\) and IDs \(p. 104\)](#).

CloudWatch Container Insights

Resource name: `containerInsights`

CloudWatch Container Insights collects, aggregates, and summarizes metrics and logs from your containerized applications and microservices. The metrics include utilization for resources such as CPU, memory, disk, and network. Container Insights also provides diagnostic information, such as container restart failures, to help you isolate issues and resolve them quickly. You can also set CloudWatch alarms on metrics that Container Insights collects. For more information, see [Amazon ECS CloudWatch Container Insights \(p. 197\)](#).

When the `containerInsights` account setting is opted in, all new clusters created after opting in will have Container Insights enabled unless you disable it during cluster creation. Individual clusters can either be enabled or disabled during creation or by using the `UpdateClusterSettings` API.

For each Region, you can opt in to or opt out of each account setting at the account level or for a specific IAM user or role. The available account settings to opt in to or out of include the new ARN and resource ID format and the `awsvpc` trunking feature.

The following are supported scenarios:

- An IAM user or role can opt in or opt out for their individual user account.
- An IAM user or role can set the default opt in or opt out setting for all users on the account.
- The root user has the ability to opt in to or opt out of any specific IAM role or user on the account. If the account setting for the root user is changed, it sets the default for all the IAM users and roles for which no individual account setting has been selected.

The opt in and opt out option must be selected for each account setting separately. The ARN and resource ID format of a resource is defined by the opt-in status of the IAM user or role that created the resource.

Only resources launched after opting in receive the new ARN and resource ID format. All existing resources are not affected. In order for Amazon ECS services and tasks to transition to the new ARN and resource ID formats, the service or task must be re-created. To transition a container instance to the new ARN and resource ID format, the container instance must be drained and a new container instance registered to the cluster.

Note

Tasks launched by an Amazon ECS service can only receive the new ARN and resource ID format if the service was created on or after November 16, 2018, and the IAM user who created the service has opted in to the new format for tasks.

Topics

- [Amazon Resource Names \(ARNs\) and IDs \(p. 104\)](#)
- [Viewing Account Settings \(p. 105\)](#)
- [Modifying Account Settings \(p. 105\)](#)

Amazon Resource Names (ARNs) and IDs

When Amazon ECS resources are created, each resource is assigned a unique Amazon Resource Name (ARN) and resource identifier (ID). If you are using a command line tool or the Amazon ECS API to work with Amazon ECS, resource ARNs or IDs are required for certain commands. For example, if you are using the [stop-task](#) AWS CLI command to stop a task, you must specify the task ARN or ID in the command.

The ability to opt in to and opt out of the new Amazon Resource Name (ARN) and resource IDs is provided on a per-Region basis. Any new accounts created are opted out by default.

You can opt in or opt out of the new Amazon Resource Name (ARN) and resource ID format at any time. After you have opted in, any new resources that you create use the new format.

Note

A resource ID does not change after it's created. Therefore, opting in or out of the new format does not affect your existing resource IDs.

The following sections describe how ARN and resource ID formats are changing. For more information on the transition to the new formats, see [Amazon Elastic Container Service FAQ](#).

Amazon Resource Name (ARN) Format

Some resources have a friendly name, such as a service named `production`. In other cases, you must specify a resource using the Amazon Resource Name (ARN) format. The new ARN format for Amazon ECS tasks, services, and container instances includes the cluster name. For details about opting in to the new ARN format, see [Modifying Account Settings \(p. 105\)](#).

The following table shows both the current (old) format and the new format for each resource type.

Resource Type	ARN
Container instance	Old: <code>arn:aws:ecs:region:aws_account_id:container-instance/container-instance-id</code> New: <code>arn:aws:ecs:region:aws_account_id:container-instance/cluster-name/container-instance-id</code>
Amazon ECS service	Old: <code>arn:aws:ecs:region:aws_account_id:service/service-name</code> New: <code>arn:aws:ecs:region:aws_account_id:service/cluster-name/service-name</code>
Amazon ECS task	Old: <code>arn:aws:ecs:region:aws_account_id:task/task-id</code> New: <code>arn:aws:ecs:region:aws_account_id:task/cluster-name/task-id</code>

Resource ID Length

A resource ID takes the form of a unique combination of letters and numbers. New resource ID formats include shorter IDs for Amazon ECS tasks and container instances. The old resource ID format was 36 characters long. The new IDs are in a 32-character format that does not include any hyphens. For details about opting in to the new resource ID format, see [Modifying Account Settings \(p. 105\)](#).

Viewing Account Settings

You can use the AWS Management Console and AWS CLI tools to view the resource types that support the new ARN and ID formats.

To view your account settings using the console

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation bar at the top of the screen, select the Region for which to view your account settings.
3. From the dashboard, choose **Account Settings**.
4. On the **Amazon ECS ARN and resource ID settings** and **CloudWatch Container Insights** sections, you can view your opt-in status for each account setting for the authenticated IAM user and role.

To view your account settings using the command line

Use one of the following commands to view your account settings.

- [list-account-settings](#) (AWS CLI)

```
aws ecs list-account-settings --effective-settings --region us-east-1
```

- [Get-ECSAccountSetting](#) (AWS Tools for Windows PowerShell)

```
Get-ECSAccountSetting -EffectiveSetting true -Region us-east-1
```

To view the account settings for a specific IAM user or IAM role using the command line

Use one of the following commands and specify the ARN of an IAM user, IAM role, or root account user in the request to view their account settings.

- [list-account-settings](#) (AWS CLI)

```
aws ecs list-account-settings --principal-arn  
arn:aws:iam::aws_account_id:user/principalName --effective-settings --region us-east-1
```

- [Get-ECSAccountSetting](#) (AWS Tools for Windows PowerShell)

```
Get-ECSAccountSetting -PrincipalArn arn:aws:iam::aws_account_id:user/principalName -  
EffectiveSetting true -Region us-east-1
```

Modifying Account Settings

You can use the AWS Management Console and AWS CLI tools to modify the account settings.

To modify account settings using the console

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation bar at the top of the screen, select the Region for which to modify your account settings.
3. From the dashboard, choose **Account Settings**.
4. On the **Amazon ECS ARN and resource ID settings** and **CloudWatch Container Insights** sections, you can select or deselect the check boxes for each account setting for the authenticated IAM user and role. Choose **Save** once finished.

Important

IAM users and IAM roles need the `ecs:PutAccountSetting` permission to perform this action.

5. On the confirmation screen, choose **Confirm** to save the selection.

To modify the default account settings for all IAM users or roles on your account using the command line

Use one of the following commands to modify the default account setting for all IAM users or roles on your account. These changes apply to the entire AWS account unless an IAM user or role explicitly overrides these settings for themselves.

- [put-account-setting-default](#) (AWS CLI)

```
aws ecs put-account-setting-default --name serviceLongArnFormat --value enabled --  
region us-east-2
```

You can also use this command to modify the account settings for all tasks (`taskLongArnFormat`), container instances (`containerInstanceLongArnFormat`), and to opt in to the increased elastic network interface (ENI) limits for container instances (`awsvpcTrunking`). To do this, replace the name parameter with the corresponding resource type.

- [Write-ECSAccountSetting](#) (AWS Tools for Windows PowerShell)

```
Write-ECSAccountSettingDefault -Name serviceLongArnFormat -Value enabled -Region us-  
east-1 -Force
```

To modify the account settings for your IAM user account using the command line

Use one of the following commands to modify the account settings for your IAM user. If you're using these commands as the root user, changes apply to the entire AWS account unless an IAM user or role explicitly overrides these settings for themselves.

- [put-account-setting](#) (AWS CLI)

```
aws ecs put-account-setting --name serviceLongArnFormat --value enabled --region us-  
east-1
```

You can also use this command to modify the account settings for all tasks (`taskLongArnFormat`), container instances (`containerInstanceLongArnFormat`), and to opt in to the increased elastic network interface (ENI) limits for container instances (`awsvpcTrunking`). To do this, replace the name parameter with the corresponding resource type.

- [Write-ECSAccountSetting](#) (AWS Tools for Windows PowerShell)

```
Write-ECSAccountSetting -Name serviceLongArnFormat -Value enabled -Force
```

To modify the account settings for a specific IAM user or IAM role using the command line

Use one of the following commands and specify the ARN of an IAM user, IAM role, or root user in the request to modify the account settings for a specific IAM user or IAM role.

- [put-account-setting](#) (AWS CLI)

```
aws ecs put-account-setting --name serviceLongArnFormat --value enabled --principal-arn  
arn:aws:iam::aws_account_id:user/principalName --region us-east-1
```

You can also use this command to modify the account settings for all tasks (`taskLongArnFormat`), container instances (`containerInstanceLongArnFormat`), and to opt in to the increased elastic network interface (ENI) limits for container instances (`awsvpcTrunking`). To do this, replace the name parameter with the corresponding resource type.

- [Write-ECSAccountSetting](#) (AWS Tools for Windows PowerShell)

```
Write-ECSAccountSetting -Name serviceLongArnFormat -Value enabled -PrincipalArn  
arn:aws:iam::aws_account_id:user/principalName -Region us-east-1 -Force
```

Scheduling Amazon ECS Tasks

Amazon Elastic Container Service (Amazon ECS) is a shared state, optimistic concurrency system that provides flexible scheduling capabilities for your tasks and containers. The Amazon ECS schedulers leverage the same cluster state information provided by the Amazon ECS API to make appropriate placement decisions.

Each task that uses the Fargate launch type has its own isolation boundary and does not share the underlying kernel, CPU resources, memory resources, or elastic network interface with another task.

Amazon ECS provides a service scheduler (for long-running tasks and applications), the ability to run tasks manually (for batch jobs or single run tasks), with Amazon ECS placing tasks on your cluster for you. You can specify task placement strategies and constraints that allow you to run tasks in the configuration you choose, such as spread out across Availability Zones. It is also possible to integrate with custom or third-party schedulers.

Service Scheduler

The service scheduler is ideally suited for long running stateless services and applications. The service scheduler ensures that the scheduling strategy you specify is followed and reschedules tasks when a task fails (for example, if the underlying infrastructure fails for some reason).

There are two service scheduler strategies available:

- **REPLICA**—The replica scheduling strategy places and maintains the desired number of tasks across your cluster. By default, the service scheduler spreads tasks across Availability Zones. You can use task placement strategies and constraints to customize task placement decisions. For more information, see [Replica \(p. 117\)](#).
- **DAEMON**—The daemon scheduling strategy deploys exactly one task on each active container instance that meets all of the task placement constraints that you specify in your cluster. The service scheduler evaluates the task placement constraints for running tasks and will stop tasks that do not meet the placement constraints. When using this strategy, there is no need to specify a desired number of tasks, a task placement strategy, or use Service Auto Scaling policies. For more information, see [Daemon](#) in the *Amazon Elastic Container Service Developer Guide*.

Note

Fargate tasks do not support the **DAEMON** scheduling strategy.

The service scheduler optionally also makes sure that tasks are registered against an Elastic Load Balancing load balancer. You can update your services that are maintained by the service scheduler, such as deploying a new task definition, or changing the running number of desired tasks. By default, the service scheduler spreads tasks across Availability Zones, but you can use task placement strategies and constraints to customize task placement decisions. For more information, see [Amazon ECS services \(p. 116\)](#).

Manually Running Tasks

The `RunTask` action is ideally suited for processes such as batch jobs that perform work and then stop. For example, you could have a process call `RunTask` when work comes into a queue. The task pulls work from the queue, performs the work, and then exits. Using `RunTask`, you can allow the default task placement strategy to distribute tasks randomly across your cluster, which minimizes the chances that a single instance gets a disproportionate number of tasks. Alternatively, you can use `RunTask` to customize how the scheduler places tasks using task placement strategies and constraints. For more information, see [Running Tasks \(p. 109\)](#) and `RunTask` in the *Amazon Elastic Container Service API Reference*.

Running Tasks on a cron-like Schedule

If you have tasks to run at set intervals in your cluster, such as a backup operation or a log scan, you can use the Amazon ECS console to create a CloudWatch Events rule that runs one or more tasks in your cluster at specified times. Your scheduled event rule can be set to either a specific interval (run every **N** minutes, hours, or days), or for more complicated scheduling, you can use a `cron` expression. For more information, see [Scheduled Tasks \(cron\)](#) (p. 111).

Contents

- [Running Tasks](#) (p. 109)
- [Scheduled Tasks \(cron\)](#) (p. 111)
- [Task Retirement](#) (p. 113)
- [Fargate Task Recycling](#) (p. 115)

Running Tasks

Running tasks manually is ideal in certain situations. For example, suppose that you are developing a task but you are not ready to deploy this task with the service scheduler. Perhaps your task is a one-time or periodic batch job that does not make sense to keep running or restart when it finishes.

To keep a specified number of tasks running or to place your tasks behind a load balancer, use the Amazon ECS service scheduler instead. For more information, see [Amazon ECS services](#) (p. 116).

Contents

- [Running a Task Using the Fargate Launch Type](#) (p. 109)

Running a Task Using the Fargate Launch Type

To run a task using the Fargate launch type, do the following:

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation pane, choose **Task Definitions** and select the task definition to run.
 - To run the latest revision of a task definition shown here, select the box to the left of the task definition to run.
 - To run an earlier revision of a task definition shown here, select the task definition to view all active revisions, then select the revision to run.
3. Choose **Actions, Run Task**.
4. In the **Run Task** section, complete the following steps:
 - a. For **Launch type**, choose **FARGATE**. For more information about launch types, see [Amazon ECS Launch Types](#) (p. 60).
 - b. For **Platform version**, choose **LATEST**. For more information about platform versions, see [AWS Fargate platform versions](#) (p. 14).
 - c. For **Cluster**, choose the cluster to use.
 - d. For **Number of tasks**, type the number of tasks to launch with this task definition.
 - e. For **Task Group**, type the name of the task group.
5. In the **VPC and security groups** section, complete the following steps:
 - a. For **Cluster VPC**, choose the VPC for your tasks to use. Ensure that the VPC that you choose is not configured to require dedicated hardware tenancy, as that is not supported by Fargate tasks.

- b. For **Subnets**, choose the available subnets for your task.
 - c. For **Security groups**, a security group has been created for your task that allows HTTP traffic from the internet (0.0.0.0/0). To edit the name or the rules of this security group, or to choose an existing security group, choose **Edit** and then modify your security group settings.
 - d. For **Auto-assign public IP**, choose **ENABLED** if you want the elastic network interface attached to the Fargate task to be assigned a public IP address. This is required if your task needs outbound network access, for example to pull an image. If outbound network access is not required, then you can choose **DISABLED**.
6. In the **Advanced Options** section, complete the following steps:
- (Optional) To send command, environment variable, task IAM role, or task execution role overrides to one or more containers in your task definition, choose **Advanced Options** and complete the following steps:

Note

If you will be using the parameter values from your task definition there is no need to specify overrides. These fields are only used to override the values specified in the task definition.

- i. For **Task Role Override**, choose an IAM role for this task to override the task IAM role specified in the task definition. For more information, see [IAM Roles for Tasks \(p. 240\)](#).

Only roles with the `ecs-tasks.amazonaws.com` trust relationship are shown here. For more information about creating an IAM role for your tasks, see [Creating an IAM Role and Policy for your Tasks \(p. 241\)](#).
- ii. For **Task Execution Role Override**, choose a task execution role to override the task execution role specified in the task definition. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).
- iii. For **Container Overrides**, choose a container to which to send a command or environment variable override.
 - **For a command override:** For **Command override**, type the command override to send. If your container definition does not specify an `ENTRYPOINT`, the format should be a comma-separated list of non-quoted strings. For example:

```
/bin/sh, -c, echo, $DATE
```

If your container definition does specify an `ENTRYPOINT` (such as `sh, -c`), the format should be an unquoted string, which is surrounded with double quotes and passed as an argument to the `ENTRYPOINT` command. For example:

```
while true; do echo $DATE > /var/www/html/index.html; sleep 1; done
```

- **For environment variable overrides:** Choose **Add Environment Variable**. For **Key**, type the name of your environment variable. For **Value**, type a string value for your environment value (without surrounding quotes).

The screenshot shows a light blue dialog box for adding an environment variable. It has two input fields: 'Key' and 'Value'. The 'Key' field contains the text 'MY_ENV_VAR'. The 'Value' field contains the text 'This variable contains a string.' To the right of the 'Value' field is a small 'x' icon. Below the input fields is a blue button with a plus icon and the text 'Add Environment Variable'.

This environment variable override is sent to the container as:

```
MY_ENV_VAR="This variable contains a string."
```

7. In the **Task tagging configuration** section, complete the following steps:
 - a. Select **Enable ECS managed tags** if you want Amazon ECS to automatically tag each task with the Amazon ECS managed tags. For more information, see [Tagging Your Amazon ECS Resources](#).
 - b. For **Propagate tags from**, select one of the following:
 - **Do not propagate** – This option will not propagate any tags.
 - **Task Definitions** – This option will propagate the tags specified in the task definition to the task.

Note

If you specify a tag with the same key in the **Tags** section, it will override the tag propagated from the task definition.

8. In the **Tags** section, specify the key and value for each tag to associate with the task. For more information, see [Tagging Your Amazon ECS Resources](#).
9. Review your task information and choose **Run Task**.

Note

If your task moves from `PENDING` to `STOPPED`, or if it displays a `PENDING` status and then disappears from the listed tasks, your task may be stopping due to an error. For more information, see [Checking stopped tasks for errors \(p. 340\)](#) in the troubleshooting section.

Scheduled Tasks (cron)

Amazon ECS supports the ability to schedule tasks on either a `cron`-like schedule or in a response to CloudWatch Events. This is supported for Amazon ECS tasks using both the Fargate and EC2 launch types.

If you have tasks to run at set intervals in your cluster, such as a backup operation or a log scan, you can use the Amazon ECS console to create a CloudWatch Events rule that runs one or more tasks in your cluster at the specified times. Your scheduled event rule can be set to either a specific interval (run every `N` minutes, hours, or days), or for more complicated scheduling, you can use a `cron` expression. For more information, see [Schedule Expressions for Rules](#) in the *Amazon CloudWatch Events User Guide*.

You can also now set your Fargate tasks as a task target in CloudWatch Events, allowing you to launch tasks in response to changes that happen. Additionally, you can modify the network configuration when using the `awsvpc` network mode via the CloudWatch Events console and AWS CLI, giving Fargate tasks triggered by CloudWatch Events the same networking properties as Amazon EC2 instances. For more information, see [Tutorial: Run an Amazon ECS Task When a File is Uploaded to an Amazon S3 Bucket](#) in the *Amazon CloudWatch Events User Guide*.

Note

This feature is not yet available for Fargate tasks in the following Regions:

Region Name	Region
China (Beijing)	cn-north-1
China (Ningxia)	cn-northwest-1
Europe (Paris)	eu-west-3
Europe (Stockholm)	eu-north-1

Region Name	Region
South America (São Paulo)	sa-east-1
Middle East (Bahrain)	me-south-1

Creating a scheduled task

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. Choose the cluster in which to create your scheduled task. If you do not have any clusters, see [Creating a Cluster \(p. 18\)](#) for steps on creating a new cluster.
3. On the **Cluster: *cluster-name*** page, choose **Scheduled Tasks, Create**.
4. For **Schedule rule name**, enter a unique name for your schedule rule. Up to 64 letters, numbers, periods, hyphens, and underscores are allowed.
5. (Optional) For **Schedule rule description**, enter a description for your rule. Up to 512 characters are allowed.
6. For **Schedule rule type**, choose whether to use a fixed interval schedule or a cron expression for your schedule rule. For more information, see [Schedule Expressions for Rules](#) in the *Amazon CloudWatch Events User Guide*.
 - For **Run at fixed interval**, enter the interval and unit for your schedule.
 - For **Cron expression**, enter the cron expression for your task schedule. These expressions have six required fields, and fields are separated by white space. For more information, and examples of cron expressions, see [Cron Expressions](#) in the *Amazon CloudWatch Events User Guide*.
7. Create a target for your schedule rule.
 - a. For **Target id**, enter a unique identifier for your target. Up to 64 letters, numbers, periods, hyphens, and underscores are allowed.
 - b. For **Launch type**, choose the launch type for the tasks in your service. For more information, see [Amazon ECS Launch Types \(p. 60\)](#).
 - c. For **Task definition**, choose the family and revision (family:revision) of the task definition to run for this target.
 - d. For **Platform version**, choose the platform version to use for this target. For more information, see [AWS Fargate platform versions \(p. 14\)](#).

Note

Platform versions are only applicable to tasks that use the Fargate launch type.

- e. For **Number of tasks**, enter the number of instantiations of the specified task definition to run on your cluster when the rule executes.
- f. (Optional) For **Task role override**, choose the IAM role to use for the task in your target, instead of the task definition default. For more information, see [IAM Roles for Tasks \(p. 240\)](#). Only roles with the **Amazon EC2 Container Service Task Role** trust relationship are shown here. For more information about creating an IAM role for your tasks, see [Creating an IAM Role and Policy for your Tasks \(p. 241\)](#). You must add `iam:PassRole` permissions for any task role and task role overrides to the CloudWatch IAM role. For more information, see [Amazon ECS CloudWatch Events IAM Role \(p. 246\)](#).
- g. If your scheduled task's task definition uses the `awsvpc` network mode, you must configure a VPC, subnet, and security group settings for your scheduled task. For more information, see [Fargate Task Networking](#) in the *Amazon Elastic Container Service User Guide for AWS Fargate*.
 - i. For **Cluster VPC**, if you selected the EC2 launch type, choose the VPC in which your container instances reside. If you selected the Fargate launch type, select the VPC that the Fargate tasks should use. Ensure that the VPC you choose is not configured to require dedicated hardware tenancy as that is not supported by Fargate tasks.

- ii. For **Subnets**, choose the available subnets for your scheduled task placement.

Important

Only private subnets are supported for the `awsvpc` network mode. Because tasks do not receive public IP addresses, a NAT gateway is required for outbound internet access, and inbound internet traffic should be routed through a load balancer.

- iii. For **Security groups**, a security group has been created for your scheduled tasks, which allows HTTP traffic from the internet (`0.0.0.0/0`). To edit the name or the rules of this security group, or to choose an existing security group, choose **Edit** and then modify your security group settings.
 - iv. For **Auto-assign Public IP**, choose whether to have your tasks receive a public IP address. If you are using Fargate tasks, a public IP address must be assigned to the task's elastic network interface, with a route to the internet, or a NAT gateway that can route requests to the internet. This allows the task to pull container images.
 - h. For **CloudWatch Events IAM role for this target**, choose an existing CloudWatch Events service role (`ecsEventsRole`) that you may have already created. Or, choose **Create new role** to create the required IAM role that allows CloudWatch Events to make calls to Amazon ECS to run tasks on your behalf. For more information, see [Amazon ECS CloudWatch Events IAM Role \(p. 246\)](#).
- Important**
- If your scheduled tasks require the use of the task execution role, a task role, or if they use a task role override, then you must add `iam:PassRole` permissions for your task execution role, task role, or task role override to the CloudWatch IAM role. For more information, see [Amazon ECS CloudWatch Events IAM Role \(p. 246\)](#).
- i. (Optional) In the **Container overrides** section, you can expand individual containers and override the command and/or environment variables for that container that are defined in the task definition.
8. (Optional) To add additional targets (other tasks to run when this rule is executed), choose **Add targets** and repeat the previous substeps for each additional target.
 9. Choose **Create**.

To edit a scheduled task

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. Choose the cluster in which to edit your scheduled task.
3. On the **Cluster: *cluster-name*** page, choose **Scheduled Tasks**.
4. Select the box to the left of the schedule rule to edit, and choose **Edit**.
5. Edit the fields to update and choose **Update**.

Task Retirement

Amazon ECS task retirement affects tasks of both Fargate and EC2 launch types and you will be notified by email of the pending retirement.

A task can be scheduled for retirement in the following scenarios:

- AWS detects the irreparable failure of the underlying hardware hosting the task.
- Your task uses the Fargate launch type and is running on a platform version that has a security vulnerability that requires you to replace the tasks by launching new tasks using a patched platform version.

If your task is scheduled for retirement, you receive an email before the event with the task ID and retirement date. This email is sent to the address that's associated with your account, the same email address that you use to log in to the AWS Management Console. If you use an email account that you do not check regularly, then you can use the [AWS Personal Health Dashboard](#) to determine if any of your tasks are scheduled for retirement. To update the contact information for your account, go to the [Account Settings](#) page.

When a task reaches its scheduled retirement date, it is stopped or terminated by AWS. If the task is part of a service, then the task is automatically stopped and the service scheduler launches a new one to replace it. If you are using standalone tasks, then you receive notification of the task retirement and must launch new tasks to replace them.

Working with Tasks Scheduled for Retirement

If the task is part of a service, then the task is automatically stopped. The service scheduler starts a new one to replace it after it reaches its scheduled retirement date. If you would like to update your service tasks before the retirement date, you can use the following steps. For more information, see [Updating a Service](#) (p. 139).

To update a running service (AWS Management Console)

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. On the navigation bar, select the Region that your cluster is in.
3. In the navigation pane, choose **Clusters**.
4. On the **Clusters** page, select the name of the cluster in which your service resides.
5. On the **Cluster: *name*** page, choose **Services**.
6. Check the box to the left of the service to update and choose **Update**.
7. On the **Configure service** page, your service information is pre-populated. Select **Force new deployment** and choose **Next step**.

Note

For tasks using the Fargate launch type, forcing a new deployment launches new tasks using the patched platform version. Your tasks do not require you select a different platform version. For more information, see [AWS Fargate platform versions](#) (p. 14).

8. On the **Configure network** and **Set Auto Scaling (optional)** pages, choose **Next step**.
9. Choose **Update Service** to finish and update your service.

To update a running service (AWS CLI)

1. Obtain the ARN for the service.

```
aws ecs list-services --cluster cluster_name --region region
```

Output:

```
{
  "serviceArns": [
    "arn:aws:ecs:region:aws_account_id:service/MyService"
  ]
}
```

2. Update your service, forcing a new deployment that deploys new tasks.

```
aws ecs update-service --service serviceArn --force-new-deployment --  
cluster cluster_name --region region
```

If you are using standalone tasks, then you can start a new task to replace it. For more information, see [Running Tasks \(p. 109\)](#).

Fargate Task Recycling

Amazon ECS task recycling only affects tasks using the Fargate and no notification is sent prior to the recycling event.

A task can be recycled in the following scenarios:

- The task is using the Fargate launch type and using platform version 1.3.0 or later. For more information, see [AWS Fargate platform versions \(p. 14\)](#).

Note

Fargate tasks using platform versions prior to 1.3.0 are not affected.

- The task is part of an Amazon ECS service. Standalone tasks are not affected by task recycling, but may still be scheduled for retirement. For more information, see [Task Retirement \(p. 113\)](#).
- AWS determines there is cause for the task to be recycled, as described below.

When AWS determines that a security or infrastructure update is needed for a Fargate task, it will apply the necessary patches for the task. Most of these patches will be transparent and the task will not need to be stopped, but on occasion it is necessary for the task to be recycled. Starting with Fargate platform version 1.3.0, any Fargate tasks launched as part of a service may be stopped and a new one started by the Amazon ECS service scheduler in order to provide the best possible security and availability for the task. Task recycling begin after February 1, 2019 and will continue on a rolling basis. The service scheduler will ensure that the desired task count for your service will be maintained.

To prepare for this new process, we recommend testing your application behavior by simulating this scenario. You can do this by stopping an individual task in your service to test for resiliency.

Amazon ECS services

An Amazon ECS service enables you to run and maintain a specified number of instances of a task definition simultaneously in an Amazon ECS cluster. If any of your tasks should fail or stop for any reason, the Amazon ECS service scheduler launches another instance of your task definition to replace it in order to maintain the desired number of tasks in the service.

In addition to maintaining the desired number of tasks in your service, you can optionally run your service behind a load balancer. The load balancer distributes traffic across the tasks that are associated with the service.

Topics

- [Service scheduler concepts \(p. 116\)](#)
- [Additional service concepts \(p. 117\)](#)
- [Service Definition Parameters \(p. 117\)](#)
- [Creating a service \(p. 129\)](#)
- [Updating a Service \(p. 139\)](#)
- [Deleting a Service \(p. 141\)](#)
- [Amazon ECS Deployment Types \(p. 142\)](#)
- [Service Load Balancing \(p. 152\)](#)
- [Service Auto Scaling \(p. 165\)](#)
- [Service Discovery \(p. 173\)](#)
- [Service Throttle Logic \(p. 176\)](#)

Service scheduler concepts

If a task in a service stops, the task is killed and a new task is launched. This process continues until your service reaches the number of desired running tasks based on the scheduling strategy (also referred to as the *service type*) that you specified when creating the service.

The service scheduler includes logic that throttles how often tasks are restarted if they repeatedly fail to start. If a task is stopped without having entered a `RUNNING` state, determined by the task having a `startedAt` time stamp, the service scheduler starts to incrementally slow down the launch attempts and emits a service event message. This behavior prevents unnecessary resources from being used for failed tasks, giving you a chance to resolve the issue. After the service is updated, the service scheduler resumes normal behavior. For more information, see [Service Throttle Logic \(p. 176\)](#) and [Service Event Messages \(p. 348\)](#).

There are two service scheduler strategies available:

- **REPLICA**—The replica scheduling strategy places and maintains the desired number of tasks across your cluster. By default, the service scheduler spreads tasks across Availability Zones. You can use task placement strategies and constraints to customize task placement decisions. For more information, see [Replica \(p. 117\)](#).
- **DAEMON**—The daemon scheduling strategy deploys exactly one task on each active container instance that meets all of the task placement constraints that you specify in your cluster. The service scheduler evaluates the task placement constraints for running tasks and will stop tasks that do not meet the placement constraints. When using this strategy, there is no need to specify a desired number of tasks,

a task placement strategy, or use Service Auto Scaling policies. For more information, see [Daemon](#) in the *Amazon Elastic Container Service Developer Guide*.

Note

Fargate tasks do not support the `DAEMON` scheduling strategy.

Replica

The *replica* scheduling strategy places and maintains the desired number of tasks in your cluster.

When using the Fargate launch type with tasks, when the service scheduler launches new tasks or stops running tasks, it attempts to maintain balance across Availability Zones.

Additional service concepts

- You can optionally run your service behind a load balancer. For more information, see [Service Load Balancing](#) (p. 152).
- You can optionally specify a deployment configuration for your service. A deployment is triggered by updated the task definition or desired count of a service. During a deployment, the service scheduler uses the *minimum healthy percent* and *maximum percent* parameters to determine the deployment strategy. For more information, see [Service Definition Parameters](#) (p. 117).
- You can optionally configure your service to use Amazon ECS service discovery. Service discovery uses Amazon Route 53 auto naming APIs to manage DNS entries for your service's tasks, making them discoverable within your VPC. For more information, see [Service Discovery](#) (p. 173).
- When you delete a service, if there are still running tasks that require cleanup, the service status moves from `ACTIVE` to `DRAINING`, and the service is no longer visible in the console or in the `ListServices` API operation. After all tasks have transitioned to either `STOPPING` or `STOPPED` status, the service status moves from `DRAINING` to `INACTIVE`. Services in the `DRAINING` or `INACTIVE` status can still be viewed with the `DescribeServices` API operation. However, in the future, `INACTIVE` services may be cleaned up and purged from Amazon ECS record keeping, and `DescribeServices` calls on those services return a `ServiceNotFoundException` error.

Service Definition Parameters

A service definition defines how to run your Amazon ECS service. The following parameters can be specified in a service definition.

Launch Type

`launchType`

Type: String

Valid values: `EC2` | `FARGATE`

Required: No

The launch type on which to run your service. If a launch type is not specified, `EC2` is used by default. For more information, see [Amazon ECS Launch Types](#) (p. 60).

If a `launchType` is specified, the `capacityProviderStrategy` parameter must be omitted.

Capacity Provider Strategy

`capacityProviderStrategy`

Type: Array of objects

Required: No

The capacity provider strategy to use for the service.

A capacity provider strategy consists of one or more capacity providers along with the base and weight to assign to them. A capacity provider must be associated with the cluster to be used in a capacity provider strategy. The `PutClusterCapacityProviders` API is used to associate a capacity provider with a cluster. Only capacity providers with an `ACTIVE` or `UPDATING` status can be used.

If a `capacityProviderStrategy` is specified, the `launchType` parameter must be omitted. If no `capacityProviderStrategy` or `launchType` is specified, the `defaultCapacityProviderStrategy` for the cluster is used.

If specifying a capacity provider that uses an Auto Scaling group, the capacity provider must already be created. New capacity providers can be created with the `CreateCapacityProvider` API operation.

To use a AWS Fargate capacity provider, specify either the `FARGATE` or `FARGATE_SPOT` capacity providers. The AWS Fargate capacity providers are available to all accounts and only need to be associated with a cluster to be used.

The `PutClusterCapacityProviders` API operation is used to update the list of available capacity providers for a cluster after the cluster is created.

`capacityProvider`

Type: String

Required: Yes

The short name or full ARN of the capacity provider.

`weight`

Type: Integer

Valid range: Integers between 0 and 1,000.

Required: No

The weight value designates the relative percentage of the total number of tasks launched that should use the specified capacity provider.

For example, if you have a strategy that contains two capacity providers and both have a weight of 1, then when the base is satisfied, the tasks will be split evenly across the two capacity providers. Using that same logic, if you specify a weight of 1 for *capacityProviderA* and a weight of 4 for *capacityProviderB*, then for every one task that is run using *capacityProviderA*, four tasks would use *capacityProviderB*.

`base`

Type: Integer

Valid range: Integers between 0 and 100,000.

Required: No

The base value designates how many tasks, at a minimum, to run on the specified capacity provider. Only one capacity provider in a capacity provider strategy can have a base defined.

Task Definition

`taskDefinition`

Type: String

Required: No

The `family` and `revision` (`family:revision`) or full Amazon Resource Name (ARN) of the task definition to run in your service. If a `revision` is not specified, the latest `ACTIVE` revision of the specified family is used.

A task definition must be specified when using the rolling update (ECS) deployment controller.

Platform Version

`platformVersion`

Type: String

Required: No

The platform version on which your tasks in the service are running. A platform version is only specified for tasks using the Fargate launch type. If one is not specified, the latest version (`LATEST`) is used by default.

AWS Fargate platform versions are used to refer to a specific runtime environment for the Fargate task infrastructure. When specifying the `LATEST` platform version when running a task or creating a service, you get the most current platform version available for your tasks. When you scale up your service, those tasks receive the platform version that was specified on the service's current deployment. For more information, see [AWS Fargate platform versions \(p. 14\)](#).

Cluster

`cluster`

Type: String

Required: No

The short name or full Amazon Resource Name (ARN) of the cluster on which to run your service. If you do not specify a cluster, the `default` cluster is assumed.

Service Name

`serviceName`

Type: String

Required: Yes

The name of your service. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed. Service names must be unique within a cluster, but you can have similarly named services in multiple clusters within a Region or across multiple Regions.

Scheduling Strategy

schedulingStrategy

Type: String

Valid values: `REPLICA` | `DAEMON`

Required: No

The scheduling strategy to use. If no scheduling strategy is specified, the `REPLICA` strategy is used. For more information, see [Service scheduler concepts \(p. 116\)](#).

There are two service scheduler strategies available:

- **REPLICA**—The replica scheduling strategy places and maintains the desired number of tasks across your cluster. By default, the service scheduler spreads tasks across Availability Zones. You can use task placement strategies and constraints to customize task placement decisions. For more information, see [Replica \(p. 117\)](#).
- **DAEMON**—The daemon scheduling strategy deploys exactly one task on each active container instance that meets all of the task placement constraints that you specify in your cluster. The service scheduler evaluates the task placement constraints for running tasks and will stop tasks that do not meet the placement constraints. When using this strategy, there is no need to specify a desired number of tasks, a task placement strategy, or use Service Auto Scaling policies. For more information, see [Daemon](#) in the *Amazon Elastic Container Service Developer Guide*.

Note

Fargate tasks do not support the `DAEMON` scheduling strategy.

Desired Count

desiredCount

Type: Integer

Required: No

The number of instantiations of the specified task definition to place and keep running on your cluster.

This parameter is required if the `REPLICA` scheduling strategy is used. If the service uses the `DAEMON` scheduling strategy, this parameter is optional.

Deployment Configuration

deploymentConfiguration

Type: Object

Required: No

Optional deployment parameters that control how many tasks run during the deployment and the ordering of stopping and starting tasks.

maximumPercent

Type: Integer

Required: No

If a service is using the rolling update (ECS) deployment type, the `maximumPercent` parameter represents an upper limit on the number of your service's tasks that are allowed in the `RUNNING` or `PENDING` state during a deployment, as a percentage of the `desiredCount` (rounded down to the nearest integer). This parameter enables you to define the deployment batch size. For example, if your service is using the `REPLICA` service scheduler and has a `desiredCount` of four tasks and a `maximumPercent` value of 200%, the scheduler may start four new tasks before stopping the four older tasks (provided that the cluster resources required to do this are available). The default `maximumPercent` value for a service using the `REPLICA` service scheduler is 200%.

If your service is using the `DAEMON` service scheduler type, the `maximumPercent` should remain at 100%, which is the default value.

The maximum number of tasks during a deployment is the `desiredCount` multiplied by the `maximumPercent/100`, rounded down to the nearest integer value.

If a service is using either the blue/green (`CODE_DEPLOY`) or `EXTERNAL` deployment types and tasks that use the EC2 launch type, the **maximum percent** value is set to the default value and is used to define the upper limit on the number of the tasks in the service that remain in the `RUNNING` state while the container instances are in the `DRAINING` state. If the tasks in the service use the Fargate launch type, the maximum percent value is not used, although it is returned when describing your service.

`minimumHealthyPercent`

Type: Integer

Required: No

If a service is using the rolling update (ECS) deployment type, the `minimumHealthyPercent` represents a lower limit on the number of your service's tasks that must remain in the `RUNNING` state during a deployment, as a percentage of the `desiredCount` (rounded up to the nearest integer). This parameter enables you to deploy without using additional cluster capacity. For example, if your service has a `desiredCount` of four tasks and a `minimumHealthyPercent` of 50%, the service scheduler may stop two existing tasks to free up cluster capacity before starting two new tasks.

For services that *do not* use a load balancer, the following should be noted:

- A service is considered healthy if all essential containers within the tasks in the service pass their health checks.
- If a task has no essential containers with a health check defined, the service scheduler will wait for 40 seconds after a task reaches a `RUNNING` state before the task is counted towards the minimum healthy percent total.
- If a task has one or more essential containers with a health check defined, the service scheduler will wait for the task to reach a healthy status before counting it towards the minimum healthy percent total. A task is considered healthy when all essential containers within the task have passed their health checks. The amount of time the service scheduler can wait for is determined by the container health check settings. For more information, see [Health Check \(p. 40\)](#).

For services that *do* use a load balancer, the following should be noted:

- If a task has no essential containers with a health check defined, the service scheduler will wait for the load balancer target group health check to return a healthy status before counting the task towards the minimum healthy percent total.
- If a task has an essential container with a health check defined, the service scheduler will wait for both the task to reach a healthy status and the load balancer target group health check to return a healthy status before counting the task towards the minimum healthy percent total.

The default value for a replica service for `minimumHealthyPercent` is 100%. The default `minimumHealthyPercent` value for a service using the `DAEMON` service schedule is 0% for the AWS CLI, the AWS SDKs, and the APIs and 50% for the AWS Management Console.

The minimum number of healthy tasks during a deployment is the `desiredCount` multiplied by the `minimumHealthyPercent/100`, rounded up to the nearest integer value.

If a service is using either the blue/green (`CODE_DEPLOY`) or `EXTERNAL` deployment types and tasks that use the EC2 launch type, the **minimum healthy percent** value is set to the default value and is used to define the lower limit on the number of the tasks in the service that remain in the `RUNNING` state while the container instances are in the `DRAINING` state. If the tasks in the service use the Fargate launch type, the minimum healthy percent value is not used, although it is returned when describing your service.

Deployment Controller

`deploymentController`

Type: Object

Required: No

The deployment controller to use for the service. If no deployment controller is specified, the ECS controller is used. For more information, see [Amazon ECS Deployment Types \(p. 142\)](#).

`type`

Type: String

Valid values: `ECS` | `CODE_DEPLOY` | `EXTERNAL`

Required: yes

The deployment controller type to use. There are three deployment controller types available:

`ECS`

The rolling update (`ECS`) deployment type involves replacing the current running version of the container with the latest version. The number of containers Amazon ECS adds or removes from the service during a rolling update is controlled by adjusting the minimum and maximum number of healthy tasks allowed during a service deployment, as specified in the [deploymentConfiguration](#).

`CODE_DEPLOY`

The blue/green (`CODE_DEPLOY`) deployment type uses the blue/green deployment model powered by CodeDeploy, which allows you to verify a new deployment of a service before sending production traffic to it.

`EXTERNAL`

The external deployment type enables you to use any third party deployment controller for full control over the deployment process for an Amazon ECS service.

Task Placement

`placementStrategy`

Type: Array of objects

Required: No

The placement strategy objects to use for tasks in your service. You can specify a maximum of four strategy rules per service.

type

Type: String

Valid values: `random` | `spread` | `binpack`

Required: No

The type of placement strategy. The `random` placement strategy randomly places tasks on available candidates. The `spread` placement strategy spreads placement across available candidates evenly based on the `field` parameter. The `binpack` strategy places tasks on available candidates that have the least available amount of the resource that is specified with the `field` parameter. For example, if you `binpack` on memory, a task is placed on the instance with the least amount of remaining memory (but still enough to run the task).

field

Type: String

Required: No

The field to apply the placement strategy against. For the `spread` placement strategy, valid values are `instanceId` (or `host`, which has the same effect), or any platform or custom attribute that is applied to a container instance, such as `attribute:ecs.availability-zone`. For the `binpack` placement strategy, valid values are `cpu` and `memory`. For the `random` placement strategy, this field is not used.

Tags

tags

Type: Array of objects

Required: No

The metadata that you apply to the service to help you categorize and organize them. Each tag consists of a key and an optional value, both of which you define. When a service is deleted, the tags are deleted as well. A maximum of 50 tags can be applied to the service. For more information, see [Tagging Your Amazon ECS Resources \(p. 177\)](#).

key

Type: String

Length Constraints: Minimum length of 1. Maximum length of 128.

Required: No

One part of a key-value pair that make up a tag. A key is a general label that acts like a category for more specific tag values.

value

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

Required: No

The optional part of a key-value pair that make up a tag. A value acts as a descriptor within a tag category (key).

`enableECSTags`

Type: Boolean

Valid values: `true` | `false`

Required: No

Specifies whether to enable Amazon ECS managed tags for the tasks in the service. If no value is specified, the default value is `false`. For more information, see [Tagging Your Resources for Billing](#) (p. 179).

`propagateTags`

Type: String

Valid values: `TASK_DEFINITION` | `SERVICE`

Required: No

Specifies whether to copy the tags from the task definition or the service to the tasks in the service. If no value is specified, the tags are not copied. Tags can only be copied to the tasks within the service during service creation. To add tags to a task after service creation, use the `TagResource` API action.

Network Configuration

`networkConfiguration`

Type: Object

Required: No

The network configuration for the service. This parameter is required for task definitions that use the `awsvpc` network mode to receive their own Elastic Network Interface, and it is not supported for other network modes. If using the Fargate launch type, the `awsvpc` network mode is required. For more information, see [Fargate Task Networking](#) in the *Amazon Elastic Container Service User Guide for AWS Fargate*.

`awsvpcConfiguration`

Type: Object

Required: No

An object representing the subnets and security groups for a task or service.

`subnets`

Type: Array of strings

Required: Yes

The subnets associated with the task or service. There is a limit of 16 subnets that can be specified per `awsvpcConfiguration`.

`securityGroups`

Type: Array of strings

Required: No

The security groups associated with the task or service. If you do not specify a security group, the default security group for the VPC is used. There is a limit of 5 security groups that can be specified per `awsVpcConfiguration`.

`assignPublicIP`

Type: String

Valid values: `ENABLED` | `DISABLED`

Required: No

Whether the task's elastic network interface receives a public IP address. If no value is specified, the default value of `DISABLED` is used.

`healthCheckGracePeriodSeconds`

Type: Integer

Required: No

The period of time, in seconds, that the Amazon ECS service scheduler should ignore unhealthy Elastic Load Balancing target health checks, container health checks, and Route 53 health checks after a task enters a `RUNNING` state. This is only valid if your service is configured to use a load balancer. If your service has a load balancer defined and you do not specify a health check grace period value, the default value of 0 is used.

If your service's tasks take a while to start and respond to health checks, you can specify a health check grace period of up to 2,147,483,647 seconds during which the ECS service scheduler ignores the health check status. This grace period can prevent the ECS service scheduler from marking tasks as unhealthy and stopping them before they have time to come up.

`loadBalancers`

Type: Array of objects

Required: No

A load balancer object representing the load balancers to use with your service. For services that use an Application Load Balancer or Network Load Balancer, there is a limit of five target groups you can attach to a service.

After you create a service, the load balancer name or target group ARN, container name, and container port specified in the service definition are immutable.

For Classic Load Balancers, this object must contain the load balancer name, the container name (as it appears in a container definition), and the container port to access from the load balancer. When a task from this service is placed on a container instance, the container instance is registered with the load balancer specified here.

For Application Load Balancers and Network Load Balancers, this object must contain the load balancer target group ARN, the container name (as it appears in a container definition), and the container port to access from the load balancer. When a task from this service is placed on a container instance, the container instance and port combination is registered as a target in the target group specified here.

`targetGroupArn`

Type: String

Required: No

The full ARN of the Elastic Load Balancing target group associated with a service.

A target group ARN is only specified when using an Application Load Balancer or Network Load Balancer. If you are using a Classic Load Balancer the target group ARN should be omitted.

`loadBalancerName`

Type: String

Required: No

The name of the load balancer to associate with the service.

A load balancer name is only specified when using a Classic Load Balancer. If you are using an Application Load Balancer or a Network Load Balancer the load balancer name parameter should be omitted.

`containerName`

Type: String

Required: No

The name of the container (as it appears in a container definition) to associate with the load balancer.

`containerPort`

Type: Integer

Required: No

The port on the container to associate with the load balancer. This port must correspond to a `containerPort` in the task definition used by tasks in the service. For tasks that use the EC2 launch type, the container instance must allow ingress traffic on the `hostPort` of the port mapping.

`role`

Type: String

Required: No

The short name or full ARN of the IAM role that allows Amazon ECS to make calls to your load balancer on your behalf. This parameter is only permitted if you are using a load balancer with your service and your task definition does not use the `awsvpc` network mode. If you specify the `role` parameter, you must also specify a load balancer object with the `loadBalancers` parameter.

If your specified role has a path other than `/`, then you must either specify the full role ARN (this is recommended) or prefix the role name with the path. For example, if a role with the name `bar` has a path of `/foo/` then you would specify `/foo/bar` as the role name. For more information, see [Friendly Names and Paths](#) in the *IAM User Guide*.

Important

If your account has already created the Amazon ECS service-linked role, that role is used by default for your service unless you specify a role here. The service-linked role is required if your task definition uses the `awsvpc` network mode, in which case you should not specify a role here. For more information, see [Service-Linked Role for Amazon ECS](#) (p. 227).

`serviceRegistries`

Type: Array of objects

Required: No

The details of the service discovery configuration for your service. For more information, see [Service Discovery](#) (p. 173).

`registryArn`

Type: String

Required: No

The ARN of the service registry. The currently supported service registry is AWS Cloud Map. For more information, see [Working with Services](#) in the *AWS Cloud Map Developer Guide*.

`port`

Type: Integer

Required: No

The port value used if your service discovery service specified an SRV record. This field is required if both the `awsvpc` network mode and SRV records are used.

`containerName`

Type: String

Required: No

The container name value, already specified in the task definition, to be used for your service discovery service. If the task definition that your service task specifies uses the `bridge` or `host` network mode, you must specify a `containerName` and `containerPort` combination from the task definition. If the task definition that your service task specifies uses the `awsvpc` network mode and a type SRV DNS record is used, you must specify either a `containerName` and `containerPort` combination or a `port` value, but not both.

`containerPort`

Type: Integer

Required: No

The port value, already specified in the task definition, to be used for your service discovery service. If the task definition your service task specifies uses the `bridge` or `host` network mode, you must specify a `containerName` and `containerPort` combination from the task definition. If the task definition your service task specifies uses the `awsvpc` network mode and a type SRV DNS record is used, you must specify either a `containerName` and `containerPort` combination or a `port` value, but not both.

Client Token

`clientToken`

Type: String

Required: No

Unique, case-sensitive identifier you provide to ensure the idempotency of the request. Up to 32 ASCII characters are allowed.

Service Definition Template

The following shows the JSON representation of an Amazon ECS service definition.

```
{
```

```

"cluster": "",
"serviceName": "",
"taskDefinition": "",
"loadBalancers": [
  {
    "targetGroupArn": "",
    "loadBalancerName": "",
    "containerName": "",
    "containerPort": 0
  }
],
"serviceRegistries": [
  {
    "registryArn": "",
    "port": 0,
    "containerName": "",
    "containerPort": 0
  }
],
"desiredCount": 0,
"clientToken": "",
"launchType": "FARGATE",
"capacityProviderStrategy": [
  {
    "capacityProvider": "",
    "weight": 0,
    "base": 0
  }
],
"platformVersion": "",
"role": "",
"deploymentConfiguration": {
  "maximumPercent": 0,
  "minimumHealthyPercent": 0
},
"placementConstraints": [
  {
    "type": "distinctInstance",
    "expression": ""
  }
],
"placementStrategy": [
  {
    "type": "spread",
    "field": ""
  }
],
"networkConfiguration": {
  "awsvpcConfiguration": {
    "subnets": [
      ""
    ],
    "securityGroups": [
      ""
    ],
    "assignPublicIp": "ENABLED"
  }
},
"healthCheckGracePeriodSeconds": 0,
"schedulingStrategy": "REPLICA",
"deploymentController": {
  "type": "CODE_DEPLOY"
},
"tags": [
  {
    "key": "",

```



```
        "value": ""
      }
    ],
    "enableECSTags": true,
    "propagateTags": "SERVICE"
  }
}
```

You can create this service definition template using the following AWS CLI command.

```
aws ecs create-service --generate-cli-skeleton
```

Creating a service

When you create an Amazon ECS service, you specify the basic parameters that define what makes up your service and how it should behave. These parameters create a service definition.

You can optionally configure additional features, such as an Elastic Load Balancing load balancer to distribute traffic across the containers in your service. For more information, see [Service Load Balancing \(p. 152\)](#). You must verify that your container instances can receive traffic from your load balancers. You can allow traffic to all ports on your container instances from your load balancer's security group to ensure that traffic can reach any containers that use dynamically assigned ports.

The following documents take you through each step of the create service wizard in the AWS Management Console.

Topics

- [Step 1: Configuring Basic Service Parameters \(p. 129\)](#)
- [Step 2: Configure a Network \(p. 131\)](#)
- [Step 3: Configuring Your Service to Use a Load Balancer \(p. 132\)](#)
- [Step 4: Configuring Your Service to Use Service Discovery \(p. 136\)](#)
- [Step 5: Configuring Your Service to Use Service Auto Scaling \(p. 137\)](#)
- [Step 6: Review and Create Your Service \(p. 139\)](#)

Step 1: Configuring Basic Service Parameters

All services require some basic configuration parameters that define the service, such as the task definition to use, which cluster the service should run on, how many tasks should be placed for the service, and so on. This is called the service definition. For more information about the parameters defined in a service definition, see [Service Definition Parameters \(p. 117\)](#).

This procedure covers creating a service with the basic service definition parameters that are required. After you have configured these parameters, you can create your service or move on to the procedures for optional service definition configuration, such as configuring your service to use a load balancer.

Note

If your cluster is configured with a default capacity provider strategy, you will only be able to create a service using the default capacity provider strategy when using the console. Likewise, if no default capacity provider is defined, you will only be able to use a launch type when creating a service using the console. It is not currently possible to have a mixed strategy using both capacity providers and launch types in the console.

To configure the basic service definition parameters

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.

2. On the navigation bar, select the Region that your cluster is in.
3. In the navigation pane, choose **Task Definitions** and select the task definition from which to create your service.
4. On the **Task Definition name** page, select the revision of the task definition from which to create your service.
5. Review the task definition, and choose **Actions, Create Service**.
6. On the **Configure service** page, fill out the following parameters accordingly:
 - **Capacity provider strategy:** Choose whether your service should use the default capacity provider strategy defined for the cluster or a custom capacity provider strategy. A capacity provider must already be associated with the cluster in order to be used in a custom capacity provider strategy. For more information, see [Amazon ECS Cluster Capacity Providers \(p. 18\)](#).
 - **Launch type:** Choose whether your service should run tasks on Fargate infrastructure, or Amazon EC2 container instances that you maintain. This option is available if your cluster has no default capacity provider defined. For more information, see [Amazon ECS Launch Types \(p. 60\)](#).
 - **Platform version:** If you chose the Fargate launch type, then select the platform version to use.

Note
When the **LATEST** platform version is selected, the 1.3.0 platform version is used. To use platform version 1.4.0, you must select the **1.4.0** option.
 - **Cluster:** Select the cluster in which to create your service.
 - **Service name:** Type a unique name for your service.
 - **Service type:** Select a scheduling strategy for your service. For more information, see [Service scheduler concepts \(p. 116\)](#).
 - **Number of tasks:** If you chose the **REPLICA** service type, type the number of tasks to launch and maintain on your cluster.

Note
If your launch type is **EC2**, and your task definition uses static host port mappings on your container instances, then you need at least one container instance with the specified port available in your cluster for each task in your service. This restriction does not apply if your task definition uses dynamic host port mappings with the **bridge** network mode. For more information, see [portMappings \(p. 38\)](#).
7. On the **Deployments** page, fill out the following parameters accordingly:
 - For **Deployment type**, choose whether your service should use a rolling update deployment or a blue/green deployment using AWS CodeDeploy. For more information, see [Amazon ECS Deployment Types \(p. 142\)](#).
 - If you selected the blue/green deployment type, complete the following steps:
 - For **Deployment configuration** choose the deployment configuration to use for the service. This determines how traffic is shifted when your task set is updated. For more information, see [Blue/Green Deployment with CodeDeploy \(p. 143\)](#)
 - For **Service role for CodeDeploy** choose the IAM service role for AWS CodeDeploy. For more information, see [Amazon ECS CodeDeploy IAM Role \(p. 243\)](#)
8. In the **Task tagging configuration** section, complete the following steps:

- a. Select **Enable ECS managed tags** if you want Amazon ECS to automatically tag the tasks in the service with the Amazon ECS managed tags. For more information, see [Tagging Your Amazon ECS Resources](#).
- b. For **Propagate tags from**, select one of the following:
 - **Do not propagate** – This option will not propagate any tags to the tasks in the service.
 - **Service** – This option will propagate the tags specified on your service to each of the tasks in the service.
 - **Task Definitions** – This option will propagate the tags specified in the task definition of a task to the tasks in the service.

Note

If you specify a tag with the same key in the **Tags** section, it will override the tag propagated from either the service or the task definition.

9. In the **Tags** section, specify the key and value for each tag to associate with the task. For more information, see [Tagging Your Amazon ECS Resources](#).
10. Choose **Next step** and navigate to [Step 2: Configure a Network \(p. 131\)](#).

Step 2: Configure a Network

If your service's task definition uses the `awsvpc` network mode, you must configure a VPC, subnet, and security group for your service.

If your service's task definition does not use the `awsvpc` network mode, you can move on to the next step, [Step 3: Configuring Your Service to Use a Load Balancer \(p. 132\)](#).

The `awsvpc` network mode does not provide task ENIs with public IP addresses for tasks that use the EC2 launch type. To access the internet, tasks that use the EC2 launch type must be launched in a private subnet that is configured to use a NAT gateway. For more information, see [NAT Gateways](#) in the *Amazon VPC User Guide*. Inbound network access must be from within the VPC using the private IP address or DNS hostname, or routed through a load balancer from within the VPC. Tasks launched within public subnets do not have internet access.

Note

The above limitation does not apply to tasks that use the Fargate launch type. You can configure these tasks to receive public IP addresses.

To configure VPC and security group settings for your service

1. If you have not done so already, follow the basic service configuration procedures in [Step 1: Configuring Basic Service Parameters \(p. 129\)](#).
2. For **Cluster VPC**, if you selected the EC2 launch type, choose the VPC in which your container instances reside. If you selected the Fargate launch type, select the VPC that the Fargate tasks should use. Ensure that the VPC you choose is not configured to require dedicated hardware tenancy, as that is not supported by Fargate tasks.
3. For **Subnets**, choose the available subnets for your service task placement.
4. For **Security groups**, a security group has been created for your service's tasks, which allows HTTP traffic from the internet (0.0.0.0/0). To edit the name or the rules of this security group, or to choose an existing security group, choose **Edit** and then modify your security group settings.
5. For **Auto-assign Public IP**, choose whether to have your tasks receive a public IP address. If you are using Fargate tasks, in order for the task to pull the container image it must either use a public subnet and be assigned a public IP address or a private subnet that has a route to the internet or a NAT gateway that can route requests to the internet.

6. If you are configuring your service to use a load balancer or if you are using the green/blue deployment type, continue to [Step 3: Configuring Your Service to Use a Load Balancer \(p. 132\)](#). If you are not configuring your service to use a load balancer, you can choose **None** as the load balancer type and move on to the next section, [Step 5: Configuring Your Service to Use Service Auto Scaling \(p. 137\)](#).

Step 3: Configuring Your Service to Use a Load Balancer

Services can be configured to use a load balancer to distribute incoming traffic to the tasks in your service. If your service is using the rolling update deployment type, this is optional. If your service is using the blue/green deployment type, then it is required to use either an Application Load Balancer or Network Load Balancer.

If you are not configuring your service to use a load balancer, you can choose **None** as the load balancer type and move on to the next section, [Step 4: Configuring Your Service to Use Service Discovery \(p. 136\)](#).

If you have an available Elastic Load Balancing load balancer configured, you can attach it to your service with the following procedures, or you can configure a new load balancer. For more information, see [Creating a Load Balancer \(p. 156\)](#).

Important

Before following these procedures, you must create your Elastic Load Balancing load balancer resources.

Topics

- [Configuring a Load Balancer for the Rolling Update Deployment Type \(p. 132\)](#)
- [Configuring a Load Balancer for the Blue/Green Deployment Type \(p. 134\)](#)

Configuring a Load Balancer for the Rolling Update Deployment Type

If your service's tasks take a while to start and respond to Elastic Load Balancing health checks, you can specify a health check grace period of up to 2,147,483,647 seconds. During that time, the service scheduler ignores health check status. This grace period can prevent the service scheduler from marking tasks as unhealthy and stopping them before they have time to come up. This is only valid if your service is configured to use a load balancer.

To configure a health check grace period

1. If you have not done so already, follow the basic service configuration procedures in [Step 1: Configuring Basic Service Parameters \(p. 129\)](#).
2. For **Health check grace period**: Enter the period of time, in seconds, that the Amazon ECS service scheduler should ignore unhealthy Elastic Load Balancing target health checks after a task has first started.

To configure your service to use a load balancer, you must choose the load balancer type to use with your service.

To choose a load balancer type

1. If you have not done so already, follow the basic service creation procedures in [Step 1: Configuring Basic Service Parameters \(p. 129\)](#).

2. For **Load balancer type**, choose the load balancer type to use with your service:

Application Load Balancer

Allows containers to use dynamic host port mapping, which enables you to place multiple tasks using the same port on a single container instance. Multiple services can use the same listener port on a single load balancer with rule-based routing and paths.

Network Load Balancer

Allows containers to use dynamic host port mapping, which enables you to place multiple tasks using the same port on a single container instance. Multiple services can use the same listener port on a single load balancer with rule-based routing.

Classic Load Balancer

Requires static host port mappings (only one task allowed per container instance); rule-based routing and paths are not supported.

We recommend that you use Application Load Balancers for your Amazon ECS services so that you can take advantage of the advanced features available to them.

3. For **Select IAM role for service**, choose **Create new role** to create a new role for your service, or select an existing IAM role to use for your service (by default, this is `ecsServiceRole`).

Important

If you choose to use an existing `ecsServiceRole` IAM role, you must verify that the role has the proper permissions to use Application Load Balancers and Classic Load Balancers. For more information, see [Service Scheduler IAM Role \(p. 233\)](#).

4. For **ELB Name**, choose the name of the load balancer to use with your service. Only load balancers that correspond to the load balancer type you selected earlier are visible here.
5. The next step depends on the load balancer type for your service. If you've chosen an Application Load Balancer, follow the steps in [To configure an Application Load Balancer \(p. 133\)](#). If you've chosen a Network Load Balancer, follow the steps in [To configure a Network Load Balancer \(p. 134\)](#).

To configure an Application Load Balancer

1. For **Container to load balance**, choose the container and port combination from your task definition that your load balancer should distribute traffic to, and choose **Add to load balancer**.
2. For **Listener port**, choose the listener port and protocol of the listener that you created in [Creating an Application Load Balancer \(p. 157\)](#) (if applicable), or choose **create new** to create a new listener and then enter a port number and choose a port protocol for **Listener protocol**.
3. For **Target group name**, choose the target group that you created in [Creating an Application Load Balancer \(p. 157\)](#) (if applicable), or choose **create new** to create a new target group.

Important

If your service's task definition uses the `awsvpc` network mode (which is required for the Fargate launch type), your target group must use `ip` as the target type, not `instance`. This is because tasks that use the `awsvpc` network mode are associated with an elastic network interface, not an Amazon EC2 instance.

4. (Optional) If you chose to create a new target group, complete the following fields as follows:
 - For **Target group name**, a default name is provided for you.
 - For **Target group protocol**, enter the protocol to use for routing traffic to your tasks.
 - For **Path pattern**, if your listener does not have any existing rules, the default path pattern (`/`) is used. If your listener already has a default rule, then you must enter a path pattern that matches traffic that you want to have sent to your service's target group. For example, if your service is a

web application called `web-app`, and you want traffic that matches `http://my-elb-url/web-app` to route to your service, then you would enter `/web-app*` as your path pattern. For more information, see [ListenerRules](#) in the *User Guide for Application Load Balancers*.

- For **Health check path**, enter the path to which the load balancer should send health check pings.
5. When you are finished configuring your Application Load Balancer, choose **Next step**.

To configure a Network Load Balancer

1. For **Container to load balance**, choose the container and port combination from your task definition that your load balancer should distribute traffic to, and choose **Add to load balancer**.
2. For **Listener port**, choose the listener port and protocol of the listener that you created in [Creating a Network Load Balancer \(p. 161\)](#) (if applicable), or choose **create new** to create a new listener and then enter a port number and choose a port protocol for **Listener protocol**.
3. For **Target group name**, choose the target group that you created in [Creating a Network Load Balancer \(p. 161\)](#) (if applicable), or choose **create new** to create a new target group.

Important

If your service's task definition uses the `awsvpc` network mode (which is required for the Fargate launch type), your target group must use `ip` as the target type, not `instance`. This is because tasks that use the `awsvpc` network mode are associated with an elastic network interface, not an Amazon EC2 instance.

4. (Optional) If you chose to create a new target group, complete the following fields as follows:
 - For **Target group name**, a default name is provided for you.
 - For **Target group protocol**, enter the protocol to use for routing traffic to your tasks.
 - For **Health check path**, enter the path to which the load balancer should send health check pings.
5. When you are finished configuring your Network Load Balancer, choose **Next Step**.

Configuring a Load Balancer for the Blue/Green Deployment Type

To configure your service that uses the blue/green deployment type to use a load balancer, you must use either an Application Load Balancer or a Network Load Balancer.

To choose a load balancer type

1. If you have not done so already, follow the basic service creation procedures in [Step 1: Configuring Basic Service Parameters \(p. 129\)](#).
2. For **Load balancer type**, choose the load balancer type to use with your service:

Application Load Balancer

Allows containers to use dynamic host port mapping, which enables you to place multiple tasks using the same port on a single container instance. Multiple services can use the same listener port on a single load balancer with rule-based routing and paths.

Network Load Balancer

Allows containers to use dynamic host port mapping, which enables you to place multiple tasks using the same port on a single container instance. Multiple services can use the same listener port on a single load balancer with rule-based routing.

We recommend that you use Application Load Balancers for your Amazon ECS services so that you can take advantage of the advanced features available to them.

3. For **Load balancer name**, choose the name of the load balancer to use with your service. Only load balancers that correspond to the load balancer type you selected earlier are visible here.
4. The next step depends on the load balancer type for your service. If you've chosen an Application Load Balancer, follow the steps in [To configure an Application Load Balancer \(p. 133\)](#). If you've chosen a Network Load Balancer, follow the steps in [To configure a Network Load Balancer \(p. 134\)](#).

To configure an Application Load Balancer for the blue/green deployment type

1. For **Container to load balance**, choose the container and port combination from your task definition that your load balancer should distribute traffic to, and choose **Add to load balancer**.
2. For **Production listener port**, choose the listener port and protocol of the listener that you created in [Creating an Application Load Balancer \(p. 157\)](#) (if applicable), or choose **create new** to create a new listener and then enter a port number and choose a port protocol for **Production listener protocol**.
3. (Optional) Select **Test listener** if you want to configure a listener port and protocol on your load balancer to test updates to your service before routing traffic to your new taskset. Complete the following step:
 - For **Test listener port**, choose the listener port and protocol of the listener that you want to test traffic over, or choose **create new** to create a new test listener and then enter a port number and choose a port protocol in **Test listener protocol**.
4. For blue/green deployments, two target groups are required. Each target group binds to a separate taskset in the deployment. Complete the following steps:
 - a. For **Target group 1 name**, choose the target group that you created in [Creating an Application Load Balancer \(p. 157\)](#) (if applicable), or choose **create new** to create a new target group.

Important

If your service's task definition uses the `awsvpc` network mode (which is required for the Fargate launch type), your target group must use `ip` as the target type, not `instance`. This is because tasks that use the `awsvpc` network mode are associated with an elastic network interface, not an Amazon EC2 instance.

- b. (Optional) If you chose to create a new target group, complete the following fields as follows:
 - For **Target group name**, enter a name for your target group.
 - For **Target group protocol**, enter the protocol to use for routing traffic to your tasks.
 - For **Path pattern**, if your listener does not have any existing rules, the default path pattern (/) is used. If your listener already has a default rule, then you must enter a path pattern that matches traffic that you want to have sent to your service's target group. For example, if your service is a web application called `web-app`, and you want traffic that matches `http://my-elb-url/web-app` to route to your service, then you would enter `/web-app*` as your path pattern. For more information, see [ListenerRules](#) in the *User Guide for Application Load Balancers*.
 - For **Health check path**, enter the path to which the load balancer should send health check pings.
- c. Repeat the steps for target group 2.
- d. When you are finished configuring your Application Load Balancer, choose **Next step**. Navigate to [Step 4: Configuring Your Service to Use Service Discovery \(p. 136\)](#).

To configure a Network Load Balancer for the blue/green deployment type

1. For **Container to load balance**, choose the container and port combination from your task definition that your load balancer should distribute traffic to, and choose **Add to load balancer**.

2. For **Listener port**, choose the listener port and protocol of the listener that you created in [Creating an Application Load Balancer \(p. 157\)](#) (if applicable), or choose **create new** to create a new listener and then enter a port number and choose a port protocol for **Listener protocol**.
3. For **Target group name**, choose the target group that you created in [Creating an Application Load Balancer \(p. 157\)](#) (if applicable), or choose **create new** to create a new target group.

Important

If your service's task definition uses the `awsvpc` network mode (which is required for the Fargate launch type), your target group must use `ip` as the target type, not `instance`. This is because tasks that use the `awsvpc` network mode are associated with an elastic network interface, not an Amazon EC2 instance.

4. (Optional) If you chose to create a new target group, complete the following fields as follows:
 - For **Target group name**, enter a name for your target group.
 - For **Target group protocol**, enter the protocol to use for routing traffic to your tasks.
 - For **Health check path**, enter the path to which the load balancer should send health check pings.
5. When you are finished configuring your Network Load Balancer, choose **Next Step**. Navigate to [Step 4: Configuring Your Service to Use Service Discovery \(p. 136\)](#).

Step 4: Configuring Your Service to Use Service Discovery

Your Amazon ECS service can optionally enable service discovery integration, which allows your service to be discoverable via DNS. For more information, see [Service Discovery \(p. 173\)](#).

If you are not configuring your service to use a service discovery, you can move on to the next section, [Step 5: Configuring Your Service to Use Service Auto Scaling \(p. 137\)](#).

To configure service discovery

1. If you have not done so already, follow the basic service configuration procedures in [Step 1: Configuring Basic Service Parameters \(p. 129\)](#).
2. On the **Configure network** page, select **Enable service discovery integration**.
3. For **Namespace**, select an existing Amazon Route 53 namespace, if you have one, otherwise select **create new private namespace**.
4. If creating a new namespace, for **Namespace name** enter a descriptive name for your namespace. This is the name used for the Amazon Route 53 hosted zone.
5. For **Configure service discovery service**, select to either create a new service discovery service or select an existing one.
6. If creating a new service discovery service, for **Service discovery name** enter a descriptive name for your service discovery service. This is used as the prefix for the DNS records to be created.
7. Select **Enable ECS task health propagation** if you want health checks enabled for your service discovery service.
8. For **DNS record type**, select the DNS record type to create for your service. Amazon ECS service discovery only supports A and SRV records, depending on the network mode that your task definition specifies. For more information about these record types, see [Supported DNS Record Types](#) in the *Amazon Route 53 Developer Guide*.
 - If the task definition that your service task specifies uses the `bridge` or `host` network mode, only type SRV records are supported. Choose a container name and port combination to associate with the record.
 - If the task definition that your service task specifies uses the `awsvpc` network mode, select either the A or SRV record type. If the type A DNS record is selected, skip to the next step. If the type

SRV is selected, specify either the port that the service can be found on or a container name and port combination to associate with the record.

9. For **TTL**, enter the resource record cache time to live (TTL), in seconds. This value determines how long a record set is cached by DNS resolvers and by web browsers.
10. Choose **Next step** to proceed and navigate to [Step 5: Configuring Your Service to Use Service Auto Scaling](#) (p. 137).

Step 5: Configuring Your Service to Use Service Auto Scaling

Your Amazon ECS service can optionally be configured to use Auto Scaling to adjust its desired count of tasks in your Amazon ECS service up or down in response to CloudWatch alarms.

Amazon ECS Service Auto Scaling supports the following types of scaling policies:

- [Target Tracking Scaling Policies](#) (p. 166) (Recommended)—Increase or decrease the number of tasks that your service runs based on a target value for a specific metric. This is similar to the way that your thermostat maintains the temperature of your home. You select temperature and the thermostat does the rest.
- [Step Scaling Policies](#) (p. 171)—Increase or decrease the number of tasks that your service runs based on a set of scaling adjustments, known as step adjustments, which vary based on the size of the alarm breach.

For more information, see [Service Auto Scaling](#) (p. 165).

To configure basic Service Auto Scaling parameters

1. If you have not done so already, follow the basic service configuration procedures in [Step 1: Configuring Basic Service Parameters](#) (p. 129).
2. On the **Set Auto Scaling** page, select **Configure Service Auto Scaling to adjust your service's desired count**.
3. For **Minimum number of tasks**, enter the lower limit of the number of tasks for Service Auto Scaling to use. Your service's desired count is not automatically adjusted below this amount.
4. For **Desired number of tasks**, this field is pre-populated with the value that you entered earlier. You can change your service's desired count at this time, but this value must be between the minimum and maximum number of tasks specified on this page.
5. For **Maximum number of tasks**, enter the upper limit of the number of tasks for Service Auto Scaling to use. Your service's desired count is not automatically adjusted above this amount.
6. For **IAM role for Service Auto Scaling**, choose the `ecsAutoscaleRole`. If this role does not exist, choose **Create new role** to have the console create it for you.
7. The following procedures provide steps for creating either target tracking or step scaling policies for your service. Choose your desired scaling policy type.

These steps help you create target tracking scaling policies and CloudWatch alarms that can be used to trigger scaling activities for your service.

To configure target tracking scaling policies for your service

1. For **Scaling policy type**, choose **Target tracking**.
2. For **Policy name**, enter a descriptive name for your policy.
3. For **ECS service metric**, choose the metric to track. The following metrics are available:

- **ECSServiceAverageCPUUtilization**—Average CPU utilization of the service.
 - **ECSServiceAverageMemoryUtilization**—Average memory utilization of the service.
 - **ALBRequestCountPerTarget**—Number of requests completed per target in an Application Load Balancer target group.
4. For **Target value**, enter the metric value that the policy should maintain. For example, use a target value of 1000 for **ALBRequestCountPerTarget**, or a target value of 75(%) for **ECSServiceAverageCPUUtilization**.
 5. For **Scale-out cooldown period**, enter the amount of time, in seconds, after a scale-out activity completes before another scale-out activity can start. While the scale-out cooldown period is in effect, the capacity that has been added by the previous scale-out activity that initiated the cooldown is calculated as part of the desired capacity for the next scale out. The intention is to continuously (but not excessively) scale out.
 6. For **Scale-in cooldown period**, enter the amount of time, in seconds, after a scale-in activity completes before another scale-in activity can start. The scale-in cooldown period is used to block subsequent scale-in requests until it has expired. The intention is to scale in conservatively to protect your application's availability. However, if another alarm triggers a scale out activity during the cooldown period after a scale-in, Service Auto Scaling scales out your scalable target immediately.
 7. (Optional) To disable the scale-in actions for this policy, choose **Disable scale-in**. This allows you to create a separate scaling policy for scale-in later.
 8. Choose **Next step**.

These steps help you create step scaling policies and CloudWatch alarms that can be used to trigger scaling activities for your service. You can create a **Scale out** alarm to increase the desired count of your service, and a **Scale in** alarm to decrease the desired count of your service.

To configure step scaling policies for your service

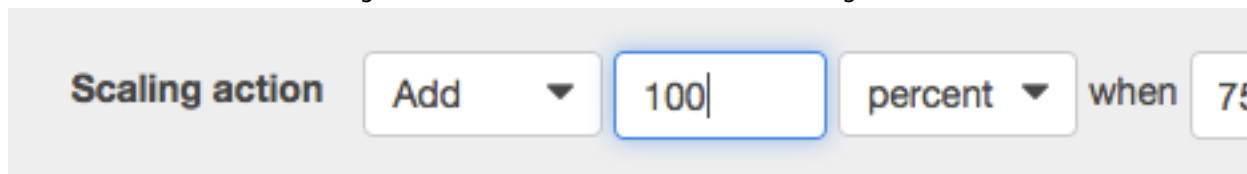
1. For **Scaling policy type**, choose **Step scaling**.
2. For **Policy name**, enter a descriptive name for your policy.
3. For **Execute policy when**, select the CloudWatch alarm to use to scale your service up or down.

You can use an existing CloudWatch alarm that you have previously created, or you can choose to create a new alarm. The **Create new alarm** workflow allows you to create CloudWatch alarms that are based on the **CPUUtilization** and **MemoryUtilization** of the service that you are creating. To use other metrics, you can create your alarm in the CloudWatch console and then return to this wizard to choose that alarm.

4. (Optional) If you've chosen to create a new alarm, complete the following steps.
 - a. For **Alarm name**, enter a descriptive name for your alarm. For example, if your alarm should trigger when your service CPU utilization exceeds 75%, you could call the alarm `service_name-cpu-gt-75`.
 - b. For **ECS service metric**, choose the service metric to use for your alarm. For more information, see [Service Auto Scaling \(p. 165\)](#).
 - c. For **Alarm threshold**, enter the following information to configure your alarm:
 - Choose the CloudWatch statistic for your alarm (the default value of **Average** works in many cases). For more information, see [Statistics](#) in the *Amazon CloudWatch User Guide*.
 - Choose the comparison operator for your alarm and enter the value that the comparison operator checks against (for example, > and 75).
 - Enter the number of consecutive periods before the alarm is triggered and the period length. For example, two consecutive periods of 5 minutes would take 10 minutes before the alarm

triggered. Because your Amazon ECS tasks can scale up and down quickly, consider using a low number of consecutive periods and a short period duration to react to alarms as soon as possible.

- d. Choose **Save**.
5. For **Scaling action**, enter the following information to configure how your service responds to the alarm:
 - Choose whether to add to, subtract from, or set a specific desired count for your service.
 - If you chose to add or subtract tasks, enter the number of tasks (or percent of existing tasks) to add or subtract when the scaling action is triggered. If you chose to set the desired count, enter the desired count that your service should be set to when the scaling action is triggered.
 - (Optional) If you chose to add or subtract tasks, choose whether the previous value is used as an integer or a percent value of the existing desired count.
 - Enter the lower boundary of your step scaling adjustment. By default, for your first scaling action, this value is the metric amount where your alarm is triggered. For example, the following scaling action adds 100% of the existing desired count when the CPU utilization is greater than 75%.



Scaling action Add 100 percent when 75

6. (Optional) You can repeat [Step 5 \(p. 139\)](#) to configure multiple scaling actions for a single alarm (for example, to add one task if CPU utilization is between 75-85%, and to add two tasks if CPU utilization is greater than 85%).
7. (Optional) If you chose to add or subtract a percentage of the existing desired count, enter a minimum increment value for **Add tasks in increments of *N* task(s)**.
8. For **Cooldown period**, enter the number of seconds between scaling actions.
9. Repeat [Step 1 \(p. 138\)](#) through [Step 8 \(p. 139\)](#) for the **Scale in** policy and choose **Save**.
10. Choose **Next step** to proceed and navigate to [Step 6: Review and Create Your Service \(p. 139\)](#).

Step 6: Review and Create Your Service

After you have configured your basic service definition parameters and optionally configured your service's networking, load balancer, service discovery, and automatic scaling, you can review your configuration. Then, choose **Create Service** to finish creating your service.

Note

After you create a service, the target group ARN or load balancer name, container name, and container port specified in the service definition are immutable. You cannot add, remove, or change the load balancer configuration of an existing service. If you update the task definition for the service, the container name and container port that were specified when the service was created must remain in the task definition.

Updating a Service

You can update an existing service to change some of the service configuration parameters, such as the number of tasks that are maintained by a service, which task definition is used by the tasks, or if your tasks are using the Fargate launch type, you can change the platform version your service uses. If you have an application that needs more capacity, you can scale up your service. If you have unused capacity to scale down, you can reduce the number of desired tasks in your service and free up resources.

If you want to use an updated container image for your tasks, you can create a new task definition revision with that image and deploy it to your service by using the **force new deployment** option in the console.

The service scheduler uses the minimum healthy percent and maximum percent parameters (in the deployment configuration for the service) to determine the deployment strategy.

If a service is using the rolling update (ECS) deployment type, the **minimum healthy percent** represents a lower limit on the number of tasks in a service that must remain in the `RUNNING` state during a deployment, as a percentage of the desired number of tasks (rounded up to the nearest integer). The parameter also applies while any container instances are in the `DRAINING` state if the service contains tasks using the EC2 launch type. This parameter enables you to deploy without using additional cluster capacity. For example, if your service has a desired number of four tasks and a minimum healthy percent of 50%, the scheduler may stop two existing tasks to free up cluster capacity before starting two new tasks. Tasks for services that do not use a load balancer are considered healthy if they are in the `RUNNING` state. Tasks for services that do use a load balancer are considered healthy if they are in the `RUNNING` state and they are reported as healthy by the load balancer. The default value for minimum healthy percent is 100%.

If a service is using the rolling update (ECS) deployment type, the **maximum percent** parameter represents an upper limit on the number of tasks in a service that are allowed in the `RUNNING` or `PENDING` state during a deployment, as a percentage of the desired number of tasks (rounded down to the nearest integer). The parameter also applies while any container instances are in the `DRAINING` state if the service contains tasks using the EC2 launch type. This parameter enables you to define the deployment batch size. For example, if your service has a desired number of four tasks and a maximum percent value of 200%, the scheduler may start four new tasks before stopping the four older tasks. That's provided that the cluster resources required to do this are available. The default value for the maximum percent is 200%.

If a service is using the blue/green (`CODE_DEPLOY`) deployment type and tasks that use the EC2 launch type, the **minimum healthy percent** and **maximum percent** values are set to the default values. They are only used to define the lower and upper limit on the number of the tasks in the service that remain in the `RUNNING` state while the container instances are in the `DRAINING` state. If the tasks in the service use the Fargate launch type, the minimum healthy percent and maximum percent values are not used. They are currently visible when describing your service.

When the service scheduler replaces a task during an update, the service first removes the task from the load balancer (if used) and waits for the connections to drain. Then, the equivalent of **docker stop** is issued to the containers running in the task. This results in a `SIGTERM` signal and a 30-second timeout, after which `SIGKILL` is sent and the containers are forcibly stopped. If the container handles the `SIGTERM` signal gracefully and exits within 30 seconds from receiving it, no `SIGKILL` signal is sent. The service scheduler starts and stops tasks as defined by your minimum healthy percent and maximum percent settings.

Important

If you are changing the ports used by containers in a task definition, you may need to update the security groups for the container instances to work with the updated ports.

If your service uses a load balancer, the load balancer configuration defined for your service when it was created cannot be changed. If you update the task definition for the service, the container name and container port that were specified when the service was created must remain in the task definition.

To change the load balancer name, the container name, or the container port associated with a service load balancer configuration, you must create a new service.

Amazon ECS does not automatically update the security groups associated with Elastic Load Balancing load balancers or Amazon ECS container instances.

To update a running service

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.

2. On the navigation bar, select the Region that your cluster is in.
 3. In the navigation pane, choose **Clusters**.
 4. On the **Clusters** page, select the name of the cluster in which your service resides.
 5. On the **Cluster: *name*** page, choose **Services**.
 6. Check the box to the left of the service to update and choose **Update**.
 7. On the **Configure service** page, your service information is pre-populated. Change the task definition, capacity provider strategy, platform version, deployment configuration, or number of desired tasks (or any combination of these). To have your service start a new deployment, which will stop and relaunch all tasks using the new configuration, select **Force new deployment**. Choose **Next step** when finished changing the service configuration.
 8. On the **Configure deployments** page, if your service is using the blue/green deployment type, the components of your service deployment is pre-populated. Confirm the following settings.
 - a. For **Application name**, choose the CodeDeploy application of which your service is a part.
 - b. For **Deployment group name**, choose the CodeDeploy deployment group of which your service is a part.
 - c. Select the deployment lifecycle event hooks and the associated Lambda functions to execute as part of the new revision of the service deployment. The available lifecycle hooks are:
 - **BeforeInstall** – Use this deployment lifecycle event hook to invoke a Lambda function before the replacement task set is created. The result of the Lambda function at this lifecycle event does not trigger a rollback.
 - **AfterInstall** – Use this deployment lifecycle event hook to invoke a Lambda function after the replacement task set is created. The result of the Lambda function at this lifecycle event can trigger a rollback.
 - **BeforeAllowTraffic** – Use this deployment lifecycle event hook to invoke a Lambda function before the production traffic has been rerouted to the replacement task set. The result of the Lambda function at this lifecycle event can trigger a rollback.
 - **AfterAllowTraffic** – Use this deployment lifecycle event hook to invoke a Lambda function after the production traffic has been rerouted to the replacement task set. The result of the Lambda function at this lifecycle event can trigger a rollback.
- For more information about lifecycle hooks, see [AppSpec 'hooks' Section](#) in the *AWS CodeDeploy User Guide*.
9. Choose **Next step**.
 10. On the **Configure network** page, your network information is pre-populated. In the **Load balancing** section, if your service is using the blue/green deployment type, select the listeners to associate with the target groups. Change the health check grace period (if desired) and choose **Next step**.
 11. (Optional) You can use Service Auto Scaling to scale your service up and down automatically in response to CloudWatch alarms.
 - a. Under **Optional configurations**, choose **Configure Service Auto Scaling**.
 - b. Proceed to [Step 5: Configuring Your Service to Use Service Auto Scaling \(p. 137\)](#).
 - c. Complete the steps in that section and then return.
 12. Choose **Update Service** to finish and update your service.

Deleting a Service

You can delete an Amazon ECS service using the console. Before deletion, the service is automatically scaled down to zero. If you have a load balancer or service discovery resources associated with the service, they are not affected by the service deletion. To delete your Elastic Load Balancing resources, see

one of the following topics, depending on your load balancer type: [Delete an Application Load Balancer](#) or [Delete a Network Load Balancer](#). To delete your service discovery resources, follow the procedure below.

To delete an Amazon ECS service

Use the following procedure to delete an Amazon ECS service.

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. On the navigation bar, select the Region that your cluster is in.
3. In the navigation pane, choose **Clusters** and select the name of the cluster in which your service resides.
4. On the **Cluster : *name*** page, choose **Services**.
5. Check the box to the left of the service to update and choose **Delete**.
6. Confirm the service deletion by entering the text phrase and choose **Delete**.

To delete the service discovery resources (AWS CLI)

To delete the remaining service discovery resources, you can use the AWS CLI to delete the service discovery service and service discovery namespace.

1. Ensure that the latest version of the AWS CLI is installed and configured. For more information about installing or upgrading your AWS CLI, see [Installing the AWS Command Line Interface](#).
2. Retrieve the ID of the service discovery service to delete.

```
aws servicediscovery list-services --region <region_name>
```

Note

If no service discovery service is returned, continue to step 4.

3. Using the service discovery service ID from the previous output, delete the service.

```
aws servicediscovery delete-service --id <service_discovery_service_id> --  
region <region_name>
```

4. Retrieve the ID of the service discovery namespace to delete.

```
aws servicediscovery list-namespaces --region <region_name>
```

5. Using the service discovery namespace ID from the previous output, delete the namespace.

```
aws servicediscovery delete-namespace --id <service_discovery_namespace_id> --  
region <region_name>
```

Amazon ECS Deployment Types

An Amazon ECS deployment type determines the deployment strategy that your service uses. There are three deployment types: rolling update, blue/green, and external.

Topics

- [Rolling Update \(p. 143\)](#)
- [Blue/Green Deployment with CodeDeploy \(p. 143\)](#)

- [External Deployment \(p. 147\)](#)

Rolling Update

The *rolling update* deployment type is controlled by Amazon ECS. This involves the service scheduler replacing the current running version of the container with the latest version. The number of tasks that Amazon ECS adds or removes from the service during a rolling update is controlled by the deployment configuration. A deployment configuration consists of the minimum and maximum number of tasks allowed during a service deployment.

To create a new Amazon ECS service that uses the rolling update deployment type, see [Creating a service \(p. 129\)](#).

Blue/Green Deployment with CodeDeploy

The *blue/green* deployment type uses the blue/green deployment model controlled by CodeDeploy. This deployment type enables you to verify a new deployment of a service before sending production traffic to it. For more information, see [What Is CodeDeploy?](#) in the *AWS CodeDeploy User Guide*.

There are three ways traffic can shift during a blue/green deployment:

- **Canary** — Traffic is shifted in two increments. You can choose from predefined canary options that specify the percentage of traffic shifted to your updated task set in the first increment and the interval, in minutes, before the remaining traffic is shifted in the second increment.
- **Linear** — Traffic is shifted in equal increments with an equal number of minutes between each increment. You can choose from predefined linear options that specify the percentage of traffic shifted in each increment and the number of minutes between each increment.
- **All-at-once** — All traffic is shifted from the original task set to the updated task set all at once.

The following are components of CodeDeploy that Amazon ECS uses when a service uses the blue/green deployment type:

CodeDeploy application

A collection of CodeDeploy resources. This consists of one or more deployment groups.

CodeDeploy deployment group

The deployment settings. This consists of the following:

- Amazon ECS cluster and service
- Load balancer target group and listener information
- Deployment roll back strategy
- Traffic rerouting settings
- Original revision termination settings
- Deployment configuration
- CloudWatch alarms configuration that can be set up to stop deployments
- SNS or CloudWatch Events settings for notifications

For more information, see [Working with Deployment Groups](#) in the *AWS CodeDeploy User Guide*.

CodeDeploy deployment configuration

Specifies how CodeDeploy routes production traffic to your replacement task set during a deployment. The following pre-defined linear and canary deployment configuration are available.

You can also create custom defined linear and canary deployments as well. For more information, see [Working with Deployment Configurations](#) in the *AWS CodeDeploy User Guide*.

Deployment configuration	Description
<code>CodeDeployDefault.ECSLinear10PercentEveryMinute</code>	Shifts 10 percent of traffic every minute until all traffic is shifted.
<code>CodeDeployDefault.ECSLinear10PercentEveryThreeMinutes</code>	Shifts 10 percent of traffic every three minutes until all traffic is shifted.
<code>CodeDeployDefault.ECSCanary10percent5Minutes</code>	Shifts 10 percent of traffic in the first increment. The remaining 90 percent is deployed five minutes later.
<code>CodeDeployDefault.ECSCanary10percent15Minutes</code>	Shifts 10 percent of traffic in the first increment. The remaining 90 percent is deployed 15 minutes later.
<code>CodeDeployDefault.ECSAllAtOnce</code>	Shifts all traffic to the updated Amazon ECS container at once.

Revision

A revision is the CodeDeploy application specification file (AppSpec file). In the AppSpec file, you specify the full ARN of the task definition and the container and port of your replacement task set where traffic is to be routed when a new deployment is created. The container name must be one of the container names referenced in your task definition. If the network configuration or platform version has been updated in the service definition, you must also specify those details in the AppSpec file. You can also specify the Lambda functions to run during the deployment lifecycle events. The Lambda functions allow you to run tests and return metrics during the deployment. For more information, see [AppSpec File Reference](#) in the *AWS CodeDeploy User Guide*.

Blue/Green Deployment Considerations

Consider the following when using the blue/green deployment type:

- When an Amazon ECS service using the blue/green deployment type is initially created, an Amazon ECS task set is created.
- You must configure the service to use either an Application Load Balancer or Network Load Balancer. Classic Load Balancers aren't supported. The following are the load balancer requirements:
 - You must add a production listener to the load balancer, which is used to route production traffic.
 - An optional test listener can be added to the load balancer, which is used to route test traffic. If you specify a test listener, CodeDeploy routes your test traffic to the replacement task set during a deployment.
 - Both the production and test listeners must belong to the same load balancer.
 - You must define a target group for the load balancer. The target group routes traffic to the original task set in a service through the production listener.
- Amazon ECS service auto scaling is not supported when using the blue/green deployment type. As a workaround, you can suspend scaling processes on the Amazon EC2 Auto Scaling groups created for your service before the service deployment, then resume the processes once the deployment has completed. For more information, see [Suspending and resuming scaling processes](#) in the *Amazon EC2 Auto Scaling User Guide*.
- Cluster capacity providers are not supported when using the blue/green deployment type.

- Tasks using the Fargate launch type or the `CODE_DEPLOY` deployment controller types don't support the `DAEMON` scheduling strategy.
- When you initially create a CodeDeploy application and deployment group, you must specify the following:
 - You must define two target groups for the load balancer. One target group should be the initial target group defined for the load balancer when the Amazon ECS service was created. The second target group's only requirement is that it can't be associated with a different load balancer than the one the service uses.
- When you create a CodeDeploy deployment for an Amazon ECS service, CodeDeploy creates a *replacement task set* (or *green task set*) in the deployment. If you added a test listener to the load balancer, CodeDeploy routes your test traffic to the replacement task set. This is when you can run any validation tests. Then CodeDeploy reroutes the production traffic from the original task set to the replacement task set according to the traffic rerouting settings for the deployment group.

Amazon ECS Console Experience

The service create and service update workflows in the Amazon ECS console supports blue/green deployments.

To create an Amazon ECS service that uses the blue/green deployment type, see [Creating a service](#) (p. 129).

To update an existing Amazon ECS service that is using the blue/green deployment type, see [Updating a Service](#) (p. 139).

When you use the Amazon ECS console to create an Amazon ECS service using the blue/green deployment type, an Amazon ECS task set and the following CodeDeploy resources are created automatically with the following default settings.

Resource	Default Setting
Application name	AppECS-< <i>cluster</i> [: 47]>-< <i>service</i> [: 47]>
Deployment group name	DgpECS-< <i>cluster</i> [: 47]>-< <i>service</i> [: 47]>
Deployment group load balancer info	The load balancer production listener, optional test listener, and target groups specified are added to the deployment group configuration.
Traffic rerouting settings	Traffic rerouting – The default setting is Reroute traffic immediately . You can change it on the CodeDeploy console or by updating the <code>TrafficRoutingConfig</code> . For more information, see CreateDeploymentConfig in the <i>AWS CodeDeploy API Reference</i> .
Original revision termination settings	The original revision termination settings are configured to wait 1 hour after traffic has been rerouted before terminating the blue task set.
Deployment configuration	The deployment configuration is set to <code>CodeDeployDefault.ECSAllAtOnce</code> by default, which routes all traffic at one time from the blue task set to the green task set. The deployment configuration can be changed using the AWS CodeDeploy console after the service is created.

Resource	Default Setting
Automatic rollback configuration	If a deployment fails, the automatic rollback settings are configured to roll it back.

To view details of an Amazon ECS service using the blue/green deployment type, use the **Deployments** tab on the Amazon ECS console.

To view the details of a CodeDeploy deployment group in the CodeDeploy console, see [View Deployment Group Details with CodeDeploy](#) in the *AWS CodeDeploy User Guide*.

To modify the settings for a CodeDeploy deployment group in the CodeDeploy console, see [Change Deployment Group Settings with CodeDeploy](#) in the *AWS CodeDeploy User Guide*.

Blue/Green Deployment Required IAM Permissions

Amazon ECS blue/green deployments are made possible by a combination of the Amazon ECS and CodeDeploy APIs. IAM users must have the appropriate permissions for these services before they can use Amazon ECS blue/green deployments in the AWS Management Console or with the AWS CLI or SDKs.

In addition to the standard IAM permissions for creating and updating services, Amazon ECS requires the following permissions. These permissions have been added to the `AmazonECS_FullAccess` IAM policy. For more information, see [AmazonECS_FullAccess](#) (p. 221).

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "codedeploy:CreateApplication",
        "codedeploy:CreateDeployment",
        "codedeploy:CreateDeploymentGroup",
        "codedeploy:GetApplication",
        "codedeploy:GetDeployment",
        "codedeploy:GetDeploymentGroup",
        "codedeploy:ListApplications",
        "codedeploy:ListDeploymentGroups",
        "codedeploy:ListDeployments",
        "codedeploy:StopDeployment",
        "codedeploy:GetDeploymentTarget",
        "codedeploy:ListDeploymentTargets",
        "codedeploy:GetDeploymentConfig",
        "codedeploy:GetApplicationRevision",
        "codedeploy:RegisterApplicationRevision",
        "codedeploy:BatchGetApplicationRevisions",
        "codedeploy:BatchGetDeploymentGroups",
        "codedeploy:BatchGetDeployments",
        "codedeploy:BatchGetApplications",
        "codedeploy:ListApplicationRevisions",
        "codedeploy:ListDeploymentConfigs",
        "codedeploy:ContinueDeployment",
        "sns:ListTopics",
        "cloudwatch:DescribeAlarms",
        "lambda:ListFunctions"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

```
}
```

Note

In addition to the standard Amazon ECS permissions required to run tasks and services, IAM users also require `iam:PassRole` permissions to use IAM roles for tasks.

CodeDeploy needs permissions to call Amazon ECS APIs, modify your Elastic Load Balancing, invoke Lambda functions, and describe CloudWatch alarms, as well as permissions to modify your service's desired count on your behalf. Before creating an Amazon ECS service that uses the blue/green deployment type, you must create an IAM role (`ecsCodeDeployRole`). For more information, see [Amazon ECS CodeDeploy IAM Role \(p. 243\)](#).

The [Create Service Example \(p. 218\)](#) and [Update Service Example \(p. 218\)](#) IAM policy examples show the permissions that are required for IAM users to use Amazon ECS blue/green deployments on the AWS Management Console.

External Deployment

The *external* deployment type enables you to use any third-party deployment controller for full control over the deployment process for an Amazon ECS service. The details for your service are managed by either the service management API actions (`CreateService`, `UpdateService`, and `DeleteService`) or the task set management API actions (`CreateTaskSet`, `UpdateTaskSet`, `UpdateServicePrimaryTaskSet`, and `DeleteTaskSet`). Each API action has a subset of the service definition parameters that it can manage.

The `UpdateService` API action updates the desired count and health check grace period parameters for a service. If the launch type, platform version, load balancer details, network configuration, or task definition need to be updated, you must create a new task set.

The `UpdateTaskSet` API action updates only the scale parameter for a task set.

The `UpdateServicePrimaryTaskSet` API action modifies which task set in a service is the primary task set. When you call the `DescribeServices` API action, it returns all fields specified for a primary task set. If the primary task set for a service is updated, any task set parameter values that exist on the new primary task set that differ from the old primary task set in a service are updated to the new value when a new primary task set is defined. If no primary task set is defined for a service, when describing the service, the task set fields are null.

External Deployment Considerations

Consider the following when using the external deployment type:

- Service auto scaling is not supported when using an external deployment controller.
- If using a load balancer for the task task, the supported load balancer types are either an Application Load Balancer or a Network Load Balancer.
- Tasks using the Fargate launch type or `EXTERNAL` deployment controller types don't support the `DAEMON` scheduling strategy.

External Deployment Workflow

The following is the basic workflow to managing an external deployment on Amazon ECS.

To manage an Amazon ECS service using an external deployment controller

1. Create an Amazon ECS service. The only required parameter is the service name. You can specify the following parameters when creating a service using an external deployment controller. All other service parameters are specified when creating a task set within the service.

`serviceName`

Type: String

Required: Yes

The name of your service. Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed. Service names must be unique within a cluster, but you can have similarly named services in multiple clusters within a Region or across multiple Regions.

`desiredCount`

The number of instantiations of the specified task set task definition to place and keep running within the service.

`deploymentConfiguration`

Optional deployment parameters that control how many tasks run during a deployment and the ordering of stopping and starting tasks. For more information, see [deploymentConfiguration](#).

`tags`

Type: Array of objects

Required: No

The metadata that you apply to the service to help you categorize and organize them. Each tag consists of a key and an optional value, both of which you define. When a service is deleted, the tags are deleted as well. A maximum of 50 tags can be applied to the service. For more information, see [Tagging Your Amazon ECS Resources \(p. 177\)](#).

`key`

Type: String

Length Constraints: Minimum length of 1. Maximum length of 128.

Required: No

One part of a key-value pair that make up a tag. A key is a general label that acts like a category for more specific tag values.

`value`

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

Required: No

The optional part of a key-value pair that make up a tag. A value acts as a descriptor within a tag category (key).

`enableECSTags`

Specifies whether to enable Amazon ECS managed tags for the tasks within the service. For more information, see [Tagging Your Resources for Billing \(p. 179\)](#).

`propagateTags`

Type: String

Valid values: `TASK_DEFINITION | SERVICE`

Required: No

Specifies whether to copy the tags from the task definition or the service to the tasks in the service. If no value is specified, the tags are not copied. Tags can only be copied to the tasks within the service during service creation. To add tags to a task after service creation, use the `TagResource` API action.

`healthCheckGracePeriodSeconds`

Type: Integer

Required: No

The period of time, in seconds, that the Amazon ECS service scheduler should ignore unhealthy Elastic Load Balancing target health checks, container health checks, and Route 53 health checks after a task enters a `RUNNING` state. This is only valid if your service is configured to use a load balancer. If your service has a load balancer defined and you do not specify a health check grace period value, the default value of 0 is used.

If your service's tasks take a while to start and respond to health checks, you can specify a health check grace period of up to 2,147,483,647 seconds during which the ECS service scheduler ignores the health check status. This grace period can prevent the ECS service scheduler from marking tasks as unhealthy and stopping them before they have time to come up.

`schedulingStrategy`

The scheduling strategy to use. Services using an external deployment controller support only the `REPLICA` scheduling strategy. For more information, see [Service scheduler concepts \(p. 116\)](#).

`placementConstraints`

An array of placement constraint objects to use for tasks in your service. You can specify a maximum of 10 constraints per task (this limit includes constraints in the task definition and those specified at run time). If you are using the Fargate launch type, task placement constraints aren't supported.

`placementStrategy`

The placement strategy objects to use for tasks in your service. You can specify a maximum of four strategy rules per service.

The following is an example service definition for creating a service using an external deployment controller.

```
{
  "cluster": "",
  "serviceName": "",
  "desiredCount": 0,
  "role": "",
  "deploymentConfiguration": {
    "maximumPercent": 0,
    "minimumHealthyPercent": 0
  },
  "placementConstraints": [
    {
      "type": "distinctInstance",
      "expression": ""
    }
  ],
  "placementStrategy": [
    {
      "type": "binpack",
```

```

        "field": ""
      }
    ],
    "healthCheckGracePeriodSeconds": 0,
    "schedulingStrategy": "REPLICA",
    "deploymentController": {
      "type": "EXTERNAL"
    },
    "tags": [
      {
        "key": "",
        "value": ""
      }
    ],
    "enableECSManagedTags": true,
    "propagateTags": "TASK_DEFINITION"
  }
}

```

2. Create an initial task set. The task set contains the following details about your service:

`taskDefinition`

The task definition for the tasks in the task set to use.

`launchType`

Type: String

Valid values: EC2 | FARGATE

Required: No

The launch type on which to run your service. If a launch type is not specified, EC2 is used by default. For more information, see [Amazon ECS Launch Types \(p. 60\)](#).

If a `launchType` is specified, the `capacityProviderStrategy` parameter must be omitted.

`platformVersion`

Type: String

Required: No

The platform version on which your tasks in the service are running. A platform version is only specified for tasks using the Fargate launch type. If one is not specified, the latest version (LATEST) is used by default.

AWS Fargate platform versions are used to refer to a specific runtime environment for the Fargate task infrastructure. When specifying the LATEST platform version when running a task or creating a service, you get the most current platform version available for your tasks. When you scale up your service, those tasks receive the platform version that was specified on the service's current deployment. For more information, see [AWS Fargate platform versions \(p. 14\)](#).

`loadBalancers`

A load balancer object representing the load balancer to use with your service. When using an external deployment controller, only Application Load Balancers and Network Load Balancers are supported. If you're using an Application Load Balancer, only one Application Load Balancer target group is allowed per task set.

The following snippet shows an example `loadBalancer` object to use.

```

"loadBalancers": [
  {

```

```
        "targetGroupArn": "",
        "containerName": "",
        "containerPort": 0
    }
]
```

Note

When specifying a `loadBalancer` object, you must specify a `targetGroupArn` and omit the `loadBalancerName` parameters.

`networkConfiguration`

The network configuration for the service. This parameter is required for task definitions that use the `awsvpc` network mode to receive their own elastic network interface, and it's not supported for other network modes. For more information, see [Fargate Task Networking](#) in the *Amazon Elastic Container Service User Guide for AWS Fargate*.

`serviceRegistries`

The details of the service discovery registries to assign to this service. For more information, see [Service Discovery](#) (p. 173).

`scale`

A floating-point percentage of the desired number of tasks to place and keep running in the task set. The value is specified as a percent total of a service's `desiredCount`. Accepted values are numbers between 0 and 100.

The following is a JSON example for creating a task set for an external deployment controller.

```
{
  "service": "",
  "cluster": "",
  "externalId": "",
  "taskDefinition": "",
  "networkConfiguration": {
    "awsvpcConfiguration": {
      "subnets": [
        ""
      ],
      "securityGroups": [
        ""
      ],
      "assignPublicIp": "DISABLED"
    }
  },
  "loadBalancers": [
    {
      "targetGroupArn": "",
      "containerName": "",
      "containerPort": 0
    }
  ],
  "serviceRegistries": [
    {
      "registryArn": "",
      "port": 0,
      "containerName": "",
      "containerPort": 0
    }
  ],
  "launchType": "EC2",
  "capacityProviderStrategy": [
```

```
{
  "capacityProvider": "",
  "weight": 0,
  "base": 0
},
"platformVersion": "",
"scale": {
  "value": null,
  "unit": "PERCENT"
},
"clientToken": ""
}
```

3. When service changes are needed, use the `UpdateService`, `UpdateTaskSet`, or `CreateTaskSet` API action depending on which parameters you're updating. If you created a task set, use the `scale` parameter for each task set in a service to determine how many tasks to keep running in the service. For example, if you have a service that contains `tasksetA` and you create a `tasksetB`, you might test the validity of `tasksetB` before wanting to transition production traffic to it. You could set the scale for both task sets to 100, and when you were ready to transition all production traffic to `tasksetB`, you could update the scale for `tasksetA` to 0 to scale it down.

Service Load Balancing

Your Amazon ECS service can optionally be configured to use Elastic Load Balancing to distribute traffic evenly across the tasks in your service.

Amazon ECS services support the Application Load Balancer, Network Load Balancer, and Classic Load Balancer load balancer types. Application Load Balancers are used to route HTTP/HTTPS (or Layer 7) traffic. Network Load Balancers and Classic Load Balancers are used to route TCP (or Layer 4) traffic. For more information, see [Load Balancer Types \(p. 154\)](#).

Application Load Balancers offer several features that make them attractive for use with Amazon ECS services:

- Each service can serve traffic from multiple load balancers and expose multiple load balanced ports by specifying multiple target groups.
- They are supported by tasks using both the Fargate and EC2 launch types.
- Application Load Balancers allow containers to use dynamic host port mapping (so that multiple tasks from the same service are allowed per container instance).
- Application Load Balancers support path-based routing and priority rules (so that multiple services can use the same listener port on a single Application Load Balancer).

We recommend that you use Application Load Balancers for your Amazon ECS services so that you can take advantage of these latest features, unless your service requires a feature that is only available with Network Load Balancers or Classic Load Balancers. For more information about Elastic Load Balancing and the differences between the load balancer types, see the [Elastic Load Balancing User Guide](#).

Topics

- [Service Load Balancing Considerations \(p. 153\)](#)
- [Load Balancer Types \(p. 154\)](#)
- [Creating a Load Balancer \(p. 156\)](#)
- [Registering Multiple Target Groups with a Service \(p. 163\)](#)

Service Load Balancing Considerations

Consider the following when you use service load balancing.

Application Load Balancer and Network Load Balancer Considerations

The following considerations are specific to Amazon ECS services using Application Load Balancers or Network Load Balancers:

- For services that use an Application Load Balancer or Network Load Balancer, you cannot attach more than five target groups to a service.
- For services with tasks using the `awsvpc` network mode, when you create a target group for your service, you must choose `ip` as the target type, not `instance`. This is because tasks that use the `awsvpc` network mode are associated with an elastic network interface, not an Amazon EC2 instance.
- If your service using an Application Load Balancer requires access to multiple load balanced ports, such as port 80 and port 443 for an HTTP/HTTPS service, you can configure two listeners. One listener is responsible for HTTPS that forwards the request to the service, and another listener that is responsible for redirecting HTTP requests to the appropriate HTTPS port. For more information, see [Create a Listener to Your Application Load Balancer](#) in the *User Guide for Application Load Balancers*.
- After you create a service, the target group ARN or load balancer name, container name, and container port specified in the service definition are immutable. You cannot add, remove, or change the load balancer configuration of an existing service. If you update the task definition for the service, the container name and container port that were specified when the service was created must remain in the task definition.
- If a service's task fails the load balancer health check criteria, the task is stopped and restarted. This process continues until your service reaches the number of desired running tasks.
- The Application Load Balancer slow start mode is supported. For more information, see [Application Load Balancer Slow Start Mode Considerations \(p. 153\)](#). about slow start mode, see [Target Groups for Your Application Load Balancers](#).
- When using Network Load Balancers configured with IP addresses as targets, requests are seen as coming from the Network Load Balancers private IP address. This means that services behind an Network Load Balancer are effectively open to the world as soon as you allow incoming requests and health checks in the target's security group.
- If you are experiencing problems with your load balancer-enabled services, see [Troubleshooting service load balancers \(p. 350\)](#).

Application Load Balancer Slow Start Mode Considerations

Application Load Balancers enabled for slow start mode are supported for Amazon ECS services. For more information about slow start mode, see [Target Groups for Your Application Load Balancers](#).

To ensure that the service scheduler ignores unhealthy container health checks until your tasks have warmed up and are ready to receive traffic, the following configurations are required:

- You must configure your container health check to return an `UNHEALTHY` status until the slow start period has ended.
- You must configure the health check grace period value for your Amazon ECS service for the same duration as the slow start mode duration.

Consider the following when you use different task network modes with Application Load Balancer slow start mode:

- When using `awsvpc` network mode, each task is assigned its own elastic network interface (ENI) and IP address which allows the Application Load Balancer to register each task as a target in the target group. This enables each newly registered target to have slow start mode enabled.
- When using `host` network mode, the task bypasses the Docker networking constructs and maps container ports directly to the Amazon EC2 instance's network interface or interfaces. You register the container instance as the Application Load Balancer target as opposed to the IP address of the task. This means you can only run one task per instance if you want slow start mode to work effectively. If you were to update an existing task or service, or restart the container instance, this does not re-register the container instance as an Application Load Balancer target, which would not cause the slow start duration to begin.
- When using `bridge` network mode, similarly to using `host` network mode, you register the container instance as the Application Load Balancer target as opposed to the Amazon ECS task so the same considerations described above apply.

Additionally, the following considerations are specific for using Application Load Balancer slow start mode and adding Amazon ECS tasks as targets:

- When you enable slow start for a target group, the targets already registered with the target group do not enter slow start mode.
- When you enable slow start for an empty target group and then register one or more targets using a single registration operation, these targets do not enter slow start mode. Newly registered targets enter slow start mode only when there is at least one registered target that is not in slow start mode.
- If you deregister a target in slow start mode, the target exits slow start mode. If you register the same target again, it enters slow start mode again.
- If a target in slow start mode becomes unhealthy and then healthy again before the duration period elapses, the target remains in slow start mode until the duration period elapses and then exits slow start mode. If a target that is not in slow start mode changes from unhealthy to healthy, it does not enter slow start mode.

Load Balancer Types

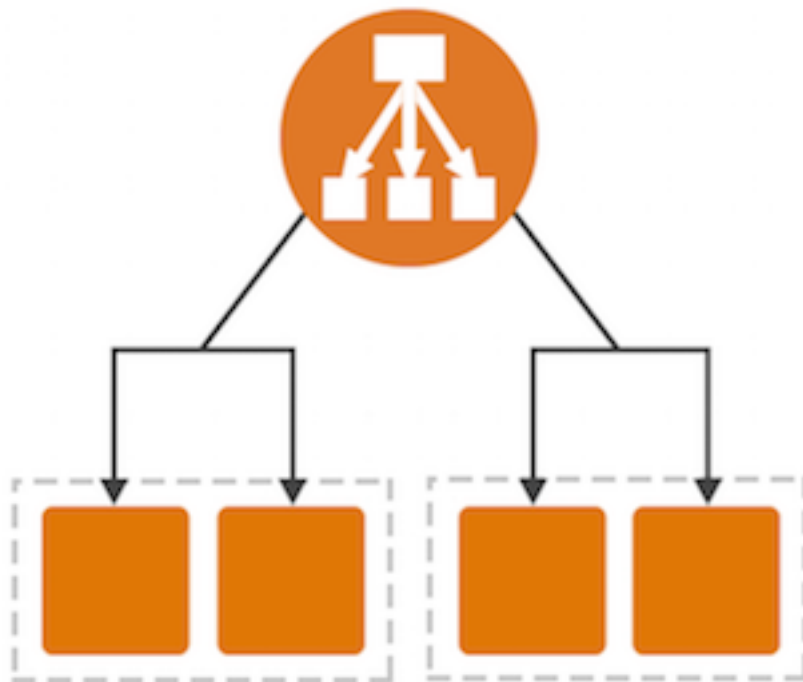
Elastic Load Balancing supports the following types of load balancers: Application Load Balancers, Network Load Balancers, and Classic Load Balancers. Amazon ECS services can use either type of load balancer. Application Load Balancers are used to route HTTP/HTTPS (or Layer 7) traffic. Network Load Balancers and Classic Load Balancers are used to route TCP (or Layer 4) traffic.

Topics

- [Application Load Balancer \(p. 154\)](#)
- [Network Load Balancer \(p. 155\)](#)

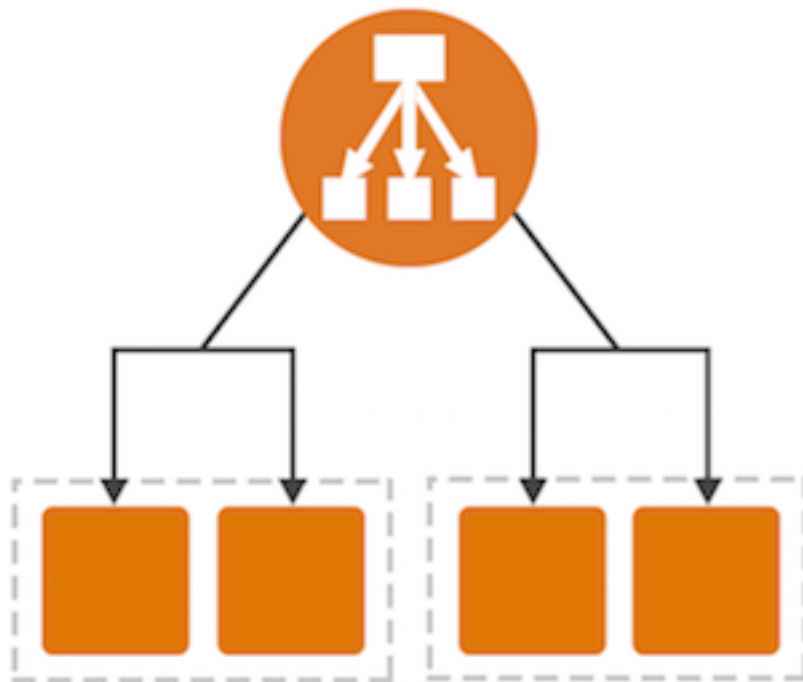
Application Load Balancer

An Application Load Balancer makes routing decisions at the application layer (HTTP/HTTPS), supports path-based routing, and can route requests to one or more ports on each container instance in your cluster. Application Load Balancers support dynamic host port mapping. For example, if your task's container definition specifies port 80 for an NGINX container port, and port 0 for the host port, then the host port is dynamically chosen from the ephemeral port range of the container instance (such as 32768 to 61000 on the latest Amazon ECS-optimized AMI). When the task is launched, the NGINX container is registered with the Application Load Balancer as an instance ID and port combination, and traffic is distributed to the instance ID and port corresponding to that container. This dynamic mapping allows you to have multiple tasks from a single service on the same container instance. For more information, see the [User Guide for Application Load Balancers](#).



Network Load Balancer

A Network Load Balancer makes routing decisions at the transport layer (TCP/SSL). It can handle millions of requests per second. After the load balancer receives a connection, it selects a target from the target group for the default rule using a flow hash routing algorithm. It attempts to open a TCP connection to the selected target on the port specified in the listener configuration. It forwards the request without modifying the headers. Network Load Balancers support dynamic host port mapping. For example, if your task's container definition specifies port 80 for an NGINX container port, and port 0 for the host port, then the host port is dynamically chosen from the ephemeral port range of the container instance (such as 32768 to 61000 on the latest Amazon ECS-optimized AMI). When the task is launched, the NGINX container is registered with the Network Load Balancer as an instance ID and port combination, and traffic is distributed to the instance ID and port corresponding to that container. This dynamic mapping allows you to have multiple tasks from a single service on the same container instance. For more information, see the [User Guide for Network Load Balancers](#).



Creating a Load Balancer

This section provides a hands-on introduction to using Elastic Load Balancing through the AWS Management Console to use with your Amazon ECS services. In this section, you create an external load balancer that receives public network traffic and routes it to your Amazon ECS container instances.

Elastic Load Balancing supports the following types of load balancers: Application Load Balancers, Network Load Balancers, and Classic Load Balancers, and Amazon ECS services can use either type of load balancer. Application Load Balancers are used to route HTTP/HTTPS traffic. Network Load Balancers and Classic Load Balancers are used to route TCP or Layer 4 traffic.

Application Load Balancers offer several features that make them attractive for use with Amazon ECS services:

- Application Load Balancers allow containers to use dynamic host port mapping (so that multiple tasks from the same service are allowed per container instance).
- Application Load Balancers support path-based routing and priority rules (so that multiple services can use the same listener port on a single Application Load Balancer).

We recommend that you use Application Load Balancers for your Amazon ECS services so that you can take advantage of these latest features. For more information about Elastic Load Balancing and the differences between the load balancer types, see the [Elastic Load Balancing User Guide](#).

Prior to using a load balancer with your Amazon ECS service, your account must already have the Amazon ECS service role created. For more information, see [Creating the Service Role for Your Account](#) (p. 157).

Topics

- [Creating the Service Role for Your Account \(p. 157\)](#)
- [Creating an Application Load Balancer \(p. 157\)](#)
- [Creating a Network Load Balancer \(p. 161\)](#)

Creating the Service Role for Your Account

Amazon ECS needs permissions to register and deregister container instances with your load balancer when tasks are created and stopped.

In most cases, the Amazon ECS service role is automatically created for you in the Amazon ECS console first run experience. You can use the following procedure to check and see if your account already has an Amazon ECS service role.

To check for the `ecsServiceRole` in the IAM console

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**.
3. Search the list of roles for `ecsServiceRole`. If the role does not exist, see [Service Scheduler IAM Role \(p. 233\)](#) to create the role. If the role does exist, select the role to view the attached policies.
4. Choose **Permissions**.
5. In the **Managed Policies** section, ensure that the **AmazonEC2ContainerServiceRole** managed policy is attached to the role. If the policy is attached, your Amazon ECS service role is properly configured. If not, follow the substeps below to attach the policy.
 - a. Choose **Attach Policy**.
 - b. For **Filter**, type **AmazonEC2ContainerServiceRole** to narrow the available policies to attach.
 - c. Select the box to the left of the **AmazonEC2ContainerServiceRole** policy and choose **Attach Policy**.
6. Choose **Trust Relationships, Edit Trust Relationship**.
7. Verify that the trust relationship contains the following policy. If the trust relationship matches the policy below, choose **Cancel**. If the trust relationship does not match, copy the policy into the **Policy Document** window and choose **Update Trust Policy**.

```
{
  "Version": "2008-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "ecs.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Creating an Application Load Balancer

This section walks you through the process of creating an Application Load Balancer in the AWS Management Console.

Define Your Load Balancer

First, provide some basic configuration information for your load balancer, such as a name, a network, and a listener.

A *listener* is a process that checks for connection requests. It is configured with a protocol and a port for the frontend (client to load balancer) connections, and protocol and a port for the backend (load balancer to backend instance) connections. In this example, you configure a listener that accepts HTTP requests on port 80 and sends them to the containers in your tasks on port 80 using HTTP.

To define your load balancer

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. From the navigation bar, select a Region for your load balancer. Be sure to select the same Region that you selected for your Amazon ECS container instances.
3. In the navigation pane, under **LOAD BALANCING**, choose **Load Balancers**.
4. Choose **Create Load Balancer**.
5. On the **Select load balancer type** page, choose **Application Load Balancer** and then choose **Continue**.
6. Complete the **Configure Load Balancer** page as follows:
 - a. For **Name**, type a name for your load balancer.
 - b. For **Scheme**, an internet-facing load balancer routes requests from clients over the internet to targets. An internal load balancer routes requests to targets using private IP addresses.
 - c. For **IP address type**, choose **ipv4** to support IPv4 addresses only or **dualstack** to support both IPv4 and IPv6 addresses.
 - d. For **Listeners**, the default is a listener that accepts HTTP traffic on port 80. You can keep the default listener settings, modify the protocol or port of the listener, or choose **Add** to add another listener.

Note

If you plan on routing traffic to more than one target group, see [ListenerRules](#) for details on how to add host or path-based rules.
 - e. For **VPC**, select the same VPC that you used for the container instances on which you intend to run your service.
 - f. For **Availability Zones**, select the check box for the Availability Zones to enable for your load balancer. If there is one subnet for that Availability Zone, it is selected. If there is more than one subnet for that Availability Zone, select one of the subnets. You can select only one subnet per Availability Zone. Your load balancer subnet configuration must include all Availability Zones that your container instances reside in.
 - g. Choose **Next: Configure Security Settings**.

Configure Security Settings

If you created a secure listener in the previous step, complete the **Configure Security Settings** page as follows; otherwise, choose **Next: Configure Security Groups**.

To configure security settings

1. If you have a certificate from AWS Certificate Manager, choose **Choose an existing certificate from AWS Certificate Manager (ACM)**, and then choose the certificate from **Certificate name**.
2. If you have already uploaded a certificate using IAM, choose **Choose an existing certificate from AWS Identity and Access Management (IAM)**, and then choose your certificate from **Certificate name**.

3. If you have a certificate ready to upload, choose **Upload a new SSL Certificate to AWS Identity and Access Management (IAM)**. For **Certificate name**, type a name for the certificate. For **Private Key**, copy and paste the contents of the private key file (PEM-encoded). In **Public Key Certificate**, copy and paste the contents of the public key certificate file (PEM-encoded). In **Certificate Chain**, copy and paste the contents of the certificate chain file (PEM-encoded), unless you are using a self-signed certificate and it's not important that browsers implicitly accept the certificate.
4. For **Select policy**, choose a predefined security policy. For details on the security policies, see [Security Policies](#) in the *User Guide for Application Load Balancers*.
5. Choose **Next: Configure Security Groups**.

Configure Security Groups

You must assign a security group to your load balancer that allows inbound traffic to the ports that you specified for your listeners. Amazon ECS does not automatically update the security groups associated with Elastic Load Balancing load balancers or Amazon ECS container instances.

To assign a security group to your load balancer

1. On the **Assign Security Groups** page, choose **Create a new security group**.
2. Enter a name and description for your security group, or leave the default name and description. This new security group contains a rule that allows traffic to the port that you configured your listener to use.

Note

Later in this topic, you create a security group rule for your container instances that allows traffic on all ports coming from the security group created here, so that the Application Load Balancer can route traffic to dynamically assigned host ports on your container instances.

Assign a security group: ☒ Create a **new** security group
☐ Select an **existing** security group

Security group name:

alb-example

Description:

Port 80 for HTTP ECS service

Type ⓘ	Protocol ⓘ	Port Range ⓘ	So
HTTP	TCP	80	A

Add Rule

3. Choose **Next: Configure Routing** to go to the next page in the wizard.

Configure Routing

In this section, you create a target group for your load balancer and the health check criteria for targets that are registered within that group.

To create a target group and configure health checks

1. For **Target group**, keep the default, **New target group**.
2. For **Name**, type a name for the new target group.
3. Set **Protocol** and **Port** as needed.
4. For **Target type**, choose whether to register your targets with an instance ID or an IP address.

Important

If your service's task definition uses the `awsvpc` network mode (which is required for the Fargate launch type), you must choose `ip` as the target type, not `instance`. This is because tasks that use the `awsvpc` network mode are associated with an elastic network interface, not an Amazon EC2 instance.

5. For **Health checks**, keep the default health check settings.
6. Choose **Next: Register Targets**.

Register Targets

Your load balancer distributes traffic between the targets that are registered to its target groups. When you associate a target group to an Amazon ECS service, Amazon ECS automatically registers and deregisters containers with your target group. Because Amazon ECS handles target registration, you do not add targets to your target group at this time.

To skip target registration

1. In the **Registered instances** section, ensure that no instances are selected for registration.
2. Choose **Next: Review** to go to the next page in the wizard.

Review and Create

Review your load balancer and target group configuration and choose **Create** to create your load balancer.

Create a Security Group Rule for Your Container Instances

After your Application Load Balancer has been created, you must add an inbound rule to your container instance security group that allows traffic from your load balancer to reach the containers.

To allow inbound traffic from your load balancer to your container instances

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. In the left navigation, choose **Security Groups**.
3. Choose the security group that your container instances use. If you created your container instances by using the Amazon ECS first run wizard, this security group may have the description, **ECS Allowed Ports**.
4. Choose the **Inbound** tab, and then choose **Edit**.
5. For **Type**, choose **All traffic**.
6. For **Source**, choose **Custom**, and then type the name of your Application Load Balancer security group that you created in [Configure Security Groups \(p. 159\)](#). This rule allows all traffic from your

Application Load Balancer to reach the containers in your tasks that are registered with your load balancer.

Type ⓘ	Protocol ⓘ
HTTP	TCP
All traffic	All

Add Rule

7. Choose **Save** to finish.

Create an Amazon ECS Service

After your load balancer and target group are created, you can specify the target group in a service definition when you create a service. When each task for your service is started, the container and port combination specified in the service definition is registered with your target group and traffic is routed from the load balancer to that container. For more information, see [Creating a service \(p. 129\)](#).

Creating a Network Load Balancer

This section walks you through the process of creating a Network Load Balancer in the AWS Management Console.

Define Your Load Balancer

First, provide some basic configuration information for your load balancer, such as a name, a network, and a listener.

A *listener* is a process that checks for connection requests. It is configured with a protocol and port for the frontend (client to load balancer) connections, and a protocol and port for the backend (load balancer to backend instance) connections. In this example, you configure an Internet-facing load balancer in the selected network with a listener that receives TCP traffic on port 80.

To define your load balancer

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. From the navigation bar, select a region for your load balancer. Be sure to select the same region that you selected for your Amazon ECS container instances.
3. In the navigation pane, under **LOAD BALANCING**, choose **Load Balancers**.

4. Choose **Create Load Balancer**.
5. On the **Select load balancer type** page, choose **Create** under **Network Load Balancer**.
6. Complete the **Configure Load Balancer** page as follows:
 - a. For **Name**, type a name for your load balancer.
 - b. For **Scheme**, choose either **internet-facing** or **internal**. An internet-facing load balancer routes requests from clients over the internet to targets. An internal load balancer routes requests to targets using private IP addresses.
 - c. For **Listeners**, the default is a listener that accepts TCP traffic on port 80. You can keep the default listener settings, modify the protocol or port of the listener, or choose **Add listener** to add another listener.
 - d. For **Availability Zones**, select the VPC that you used for your Amazon EC2 instances. For each Availability Zone that you used to launch your Amazon EC2 instances, select an Availability Zone and then select the public subnet for that Availability Zone. To associate an Elastic IP address with the subnet, select it from **Elastic IP**.
 - e. Choose **Next: Configure Routing**.

Configure Routing

You register targets, such as Amazon EC2 instances, with a target group. The target group that you configure in this step is used as the target group in the listener rule, which forwards requests to the target group. For more information, see [Target Groups for Your Network Load Balancers](#) in the *User Guide for Network Load Balancers*.

To configure your target group

1. For **Target group**, keep the default, **New target group**.
2. For **Name**, type a name for the target group.
3. Set **Protocol** and **Port** as needed.
4. For **Target type**, choose whether to register your targets with an instance ID or an IP address.

Important

If your service's task definition uses the `awsvpc` network mode (which is required for the Fargate launch type), you must choose `ip` as the target type, not `instance`. This is because tasks that use the `awsvpc` network mode are associated with an elastic network interface, not an Amazon EC2 instance.

You cannot register instances by instance ID if they have the following instance types: C1, CC1, CC2, CG1, CG2, CR1, G1, G2, H1, HS1, M1, M2, M3, and T1. You can register instances of these types by IP address.

5. For **Health checks**, keep the default health check settings.
6. Choose **Next: Register Targets**.

Register Targets with the Target Group

Your load balancer distributes traffic between the targets that are registered to its target groups. When you associate a target group to an Amazon ECS service, Amazon ECS automatically registers and deregisters containers with your target group. Because Amazon ECS handles target registration, you do not add targets to your target group at this time.

To skip target registration

1. In the **Registered instances** section, ensure that no instances are selected for registration.
2. Choose **Next: Review** to go to the next page in the wizard.

Review and Create

Review your load balancer and target group configuration and choose **Create** to create your load balancer.

Create an Amazon ECS Service

After your load balancer and target group are created, you can specify the target group in a service definition when you create a service. When each task for your service is started, the container and port combination specified in the service definition is registered with your target group and traffic is routed from the load balancer to that container. For more information, see [Creating a service \(p. 129\)](#).

Registering Multiple Target Groups with a Service

Your Amazon ECS service can serve traffic from multiple load balancers and expose multiple load balanced ports when you specify multiple target groups in a service definition.

To create a service specifying multiple target groups, you must create the service using the Amazon ECS API, SDK, AWS CLI, or an AWS CloudFormation template. After the service is created, you can view the service and the target groups registered to it with the AWS Management Console. It is not possible to update the load balancing configuration of an existing service.

Multiple target groups can be specified in a service definition using the following format. For the full syntax of a service definition, see [Service Definition Template \(p. 127\)](#).

```
"loadBalancers": [
  {
    "targetGroupArn": "arn:aws:elasticloadbalancing:region:123456789012:targetgroup/
target_group_name_1/1234567890123456",
    "containerName": "container_name",
    "containerPort": container_port
  },
  {
    "targetGroupArn": "arn:aws:elasticloadbalancing:region:123456789012:targetgroup/
target_group_name_2/6543210987654321",
    "containerName": "container_name",
    "containerPort": container_port
  }
]
```

Multiple Target Group Considerations

The following should be considered when you specify multiple target groups in a service definition:

- Multiple target groups are only supported when you use the Application Load Balancer or Network Load Balancer load balancer types.
- Multiple target groups are only supported when the service uses the rolling update (ECS) deployment controller type. If you are using the CodeDeploy or an external deployment controller, multiple target groups are not supported.
- Multiple target groups are supported for services containing tasks using both the Fargate and EC2 launch types.
- When creating a service that specifies multiple target groups, the Amazon ECS service-linked role must be created. The role is created by omitting the `role` parameter in API requests, or the `Role` property in AWS CloudFormation. For more information, see [Service-Linked Role for Amazon ECS \(p. 227\)](#).

Example Service Definitions

Following are a few example use cases for specifying multiple target groups in a service definition. For the full syntax of a service definition, see [Service Definition Template](#) (p. 127).

Example: Having separate load balancers for internal and external traffic

In the following use case, a service uses two separate load balancers, one for internal traffic and a second for internet-facing traffic, for the same container and port.

```
"loadBalancers":[
  //Internal ELB
  {
    "targetGroupArn":"arn:aws:elasticloadbalancing:region:123456789012:targetgroup/
    target_group_name_1/1234567890123456",
    "containerName":"nginx",
    "containerPort":8080
  },
  //Internet-facing ELB
  {
    "targetGroupArn":"arn:aws:elasticloadbalancing:region:123456789012:targetgroup/
    target_group_name_2/6543210987654321",
    "containerName":"nginx",
    "containerPort":8080
  }
]
```

Example: Exposing multiple ports from the same container

In the following use case, a service uses one load balancer but exposes multiple ports from the same container. For example, a Jenkins container might expose port 8080 for the Jenkins web interface and port 50000 for the API.

```
"loadBalancers":[
  {
    "targetGroupArn":"arn:aws:elasticloadbalancing:region:123456789012:targetgroup/
    target_group_name_1/1234567890123456",
    "containerName":"jenkins",
    "containerPort":8080
  },
  {
    "targetGroupArn":"arn:aws:elasticloadbalancing:region:123456789012:targetgroup/
    target_group_name_2/6543210987654321",
    "containerName":"jenkins",
    "containerPort":50000
  }
]
```

Example: Exposing ports from multiple containers

In the following use case, a service uses one load balancer and two target groups to expose ports from separate containers.

```
"loadBalancers":[
  {
```

```
"targetGroupArn": "arn:aws:elasticloadbalancing:region:123456789012:targetgroup/  
target_group_name_1/1234567890123456",  
  "containerName": "webserver",  
  "containerPort": 80  
},  
{  
  "targetGroupArn": "arn:aws:elasticloadbalancing:region:123456789012:targetgroup/  
target_group_name_2/6543210987654321",  
  "containerName": "database",  
  "containerPort": 3306  
}  
]
```

Service Auto Scaling

Automatic scaling is the ability to increase or decrease the desired count of tasks in your Amazon ECS service automatically. Amazon ECS leverages the Application Auto Scaling service to provide this functionality. For more information, see the [Application Auto Scaling User Guide](#).

Amazon ECS publishes CloudWatch metrics with your service's average CPU and memory usage. For more information, see [Service Utilization \(p. 187\)](#). You can use these and other CloudWatch metrics to scale out your service (add more tasks) to deal with high demand at peak times, and to scale in your service (run fewer tasks) to reduce costs during periods of low utilization.

Amazon ECS Service Auto Scaling supports the following types of automatic scaling:

- [Target Tracking Scaling Policies \(p. 166\)](#)—Increase or decrease the number of tasks that your service runs based on a target value for a specific metric. This is similar to the way that your thermostat maintains the temperature of your home. You select temperature and the thermostat does the rest.
- [Step Scaling Policies \(p. 171\)](#)—Increase or decrease the number of tasks that your service runs based on a set of scaling adjustments, known as step adjustments, that vary based on the size of the alarm breach.
- [Scheduled Scaling](#)—Increase or decrease the number of tasks that your service runs based on the date and time.

IAM Permissions Required for Service Auto Scaling

Service Auto Scaling is made possible by a combination of the Amazon ECS, CloudWatch, and Application Auto Scaling APIs. Services are created and updated with Amazon ECS, alarms are created with CloudWatch, and scaling policies are created with Application Auto Scaling.

In addition to the standard IAM permissions for creating and updating services, the IAM user that accesses Service Auto Scaling settings must have the appropriate permissions for the services that support dynamic scaling. IAM users must have permissions to use the actions shown in the following example policy.

```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Effect": "Allow",  
      "Action": [  
        "application-autoscaling:*",  
        "ecs:DescribeServices",  
        "ecs:UpdateService",  
      ]  
    }  
  ]  
}
```

```

        "cloudwatch:DescribeAlarms",
        "cloudwatch:PutMetricAlarm",
        "cloudwatch:DeleteAlarms",
        "cloudwatch:DescribeAlarmHistory",
        "cloudwatch:DescribeAlarms",
        "cloudwatch:DescribeAlarmsForMetric",
        "cloudwatch:GetMetricStatistics",
        "cloudwatch:ListMetrics",
        "cloudwatch:PutMetricAlarm",
        "cloudwatch:DisableAlarmActions",
        "cloudwatch:EnableAlarmActions",
        "iam:CreateServiceLinkedRole",
        "sns:CreateTopic",
        "sns:Subscribe",
        "sns:Get*",
        "sns:List*"
    ],
    "Resource": [
        "*"
    ]
}
]
}

```

The [Create Service Example \(p. 218\)](#) and [Update Service Example \(p. 218\)](#) IAM policy examples show the permissions that are required for IAM users to use Service Auto Scaling in the AWS Management Console.

The Application Auto Scaling service also needs permission to describe your Amazon ECS services and CloudWatch alarms, and permissions to modify your service's desired count on your behalf. If you enable automatic scaling for your ECS services, it creates a service-linked role named `AWSServiceRoleForApplicationAutoScaling_ECSService`. This service-linked role grants Application Auto Scaling permission to describe the alarms for your policies, to monitor the current running task count of the service, and to modify the desired count of the service. The original managed Amazon ECS role for Application Auto Scaling was `ecsAutoscaleRole`, but it is no longer required. The service-linked role is the default role for Application Auto Scaling. For more information, see [Service-Linked Roles](#) in the *Application Auto Scaling User Guide*.

Target Tracking Scaling Policies

With target tracking scaling policies, you select a metric and set a target value. Amazon ECS Service Auto Scaling creates and manages the CloudWatch alarms that trigger the scaling policy and calculates the scaling adjustment based on the metric and the target value. The scaling policy adds or removes service tasks as required to keep the metric at, or close to, the specified target value. In addition to keeping the metric close to the target value, a target tracking scaling policy also adjusts to the fluctuations in the metric due to a fluctuating load pattern and minimizes rapid fluctuations in the number of tasks running in your service.

Considerations

Keep the following considerations in mind:

- A target tracking scaling policy assumes that it should perform scale out when the specified metric is above the target value. You cannot use a target tracking scaling policy to scale out when the specified metric is below the target value.
- A target tracking scaling policy does not perform scaling when the specified metric has insufficient data. It does not perform scale in because it does not interpret insufficient data as low utilization.
- You may see gaps between the target value and the actual metric data points. This is because Service Auto Scaling always acts conservatively by rounding up or down when it determines how much

capacity to add or remove. This prevents it from adding insufficient capacity or removing too much capacity.

- To ensure application availability, the service scales out proportionally to the metric as fast as it can, but scales in more gradually.
- You can have multiple target tracking scaling policies for an Amazon ECS service, provided that each of them uses a different metric. The intention of Service Auto Scaling is to always prioritize availability, so its behavior differs depending on whether the target tracking policies are ready for scale out or scale in. It will scale out the service if any of the target tracking policies are ready for scale out, but will scale in only if all of the target tracking policies (with the scale-in portion enabled) are ready to scale in.
- Do not edit or delete the CloudWatch alarms that Service Auto Scaling manages for a target tracking scaling policy. Service Auto Scaling deletes the alarms automatically when you delete the scaling policy.

Tutorial: Service Auto Scaling with Target Tracking

The following procedures help you to create an Amazon ECS cluster and a service that uses target tracking to scale out (and in) automatically based on demand.

In this tutorial, you use the Amazon ECS first-run wizard to create a cluster and a service that runs behind an Elastic Load Balancing load balancer. Then you configure a target tracking scaling policy that scales your service automatically based on the current application load as measured by the service's CPU utilization (from the **ECS, ClusterName, ServiceName** category in CloudWatch).

When the average CPU utilization of your service rises above 75% (meaning that more than 75% of the CPU that is reserved for the service is being used), a scale-out alarm triggers Service Auto Scaling to add another task to your service to help out with the increased load. Conversely, when the average CPU utilization of your service drops below the target utilization for a sustained period of time, a scale-in alarm triggers a decrease in the service's desired count to free up those cluster resources for other tasks and services.

Prerequisites

This tutorial assumes that you are using administrator credentials, and that you have an Amazon EC2 key pair in the current region. If you do not have these resources, or you are not sure, you can create them by following the steps in [Setting Up with Amazon ECS \(p. 3\)](#).

Step 1: Create a Cluster and a Service

Start by creating a cluster and service using the Amazon ECS first-run wizard. The first-run wizard takes care of creating the necessary IAM roles for this tutorial, an Auto Scaling group for your container instances, and a service that runs behind a load balancer. The wizard also makes the clean-up process much easier, because you can delete the entire AWS CloudFormation stack in one step.

For this tutorial, you create a cluster called `service-autoscaling` and a service called `sample-webapp`.

To create your cluster and service

1. Open the Amazon ECS console first run wizard at <https://console.aws.amazon.com/ecs/home#/firstRun>.
2. From the navigation bar, choose the **US East (N. Virginia)** region.
3. On **Step 1: Container and Task**, for **Container definition**, select **sample-app**.
4. For **Task definition**, leave all of the default options and choose **Next**.
5. On **Step 2: Service**, for **Load balancer type**, choose **Application Load Balancer**, **Next**.

Important

Application Load Balancers do incur costs while they exist in your AWS resources. For more information, see [Elastic Load Balancing Pricing](#).

6. On **Step 3: Cluster**, for **Cluster name**, enter `service-autoscaling` and choose **Next**.
7. Review your choices and then choose **Create**.

You are directed to a **Launch Status** page that shows the status of your launch and describes each step of the process (this can take a few minutes to complete while your cluster resources are created and populated).

8. When your cluster and service are created, choose **View service**.

Step 2: Configure Service Auto Scaling

Now that you have launched a cluster and created a service in that cluster that is running behind a load balancer, you can enable Service Auto Scaling by creating a target tracking scaling policy.

To configure basic Service Auto Scaling parameters

1. On the **Service: sample-app-service** page, your service configuration should look similar to the image below, although the task definition revision and load balancer name are likely to be different. Choose **Update** to update your new service.

Service : sample-app-service

Cluster [service-autoscaling](#)

Status **ACTIVE**

Task definition [first-run-task-definition:5](#)

Launch type FARGATE

Platform version LATEST

Service role [aws-service-role/ecs.amazonaws.com/AWSServ](#)

Details

Tasks

Events

Auto Scaling

Deployments

Load Balancing

Target Group Name	Container Name	Com
EC2Co-Defau-13FL25TVMRZRO	sample-app	

2. On the **Update service** page, choose **Next step** until you get to **Step 3: Set Auto Scaling (optional)**.
3. For **Service Auto Scaling**, choose **Configure Service Auto Scaling to adjust your service's desired count**.
4. For **Minimum number of tasks**, enter 1 for the lower limit of the number of tasks for Service Auto Scaling to use. Your service's desired count is not automatically adjusted below this amount.
5. For **Desired number of tasks**, this field is pre-populated with the value that you entered earlier. This value must be between the minimum and maximum number of tasks specified on this page. Leave this value at 1.
6. For **Maximum number of tasks**, enter 2 for the upper limit of the number of tasks for Service Auto Scaling to use. Your service's desired count is not automatically adjusted above this amount.
7. For **IAM role for Service Auto Scaling**, choose the `ecsAutoscaleRole`. If this role does not exist, choose **Create new role** to have the console create it for you.

To configure a target tracking scaling policy for your service

1. Choose **Add scaling policy** to configure your scaling policy.

2. On the **Add policy** page, update the following fields:
 - a. For **Scaling policy type**, choose **Target tracking**.
 - b. For **Policy name**, enter `TargetTrackingPolicy`.
 - c. For **ECS service metric**, choose **ECSServiceAverageCPUUtilization**.
 - d. For **Target value**, enter 75.
 - e. For **Scale-out cooldown period**, enter 60 seconds. A scale-out activity increases the number of your service's tasks. While the scale-out cooldown period is in effect, the capacity that has been added by the previous scale-out activity that initiated the cooldown is calculated as part of the desired capacity for the next scale out. The intention is to continuously (but not excessively) scale out.
 - f. For **Scale-in cooldown period**, enter 60 seconds. A scale-in activity reduces the number of your service's tasks. The scale-in cooldown period is used to block subsequent scale-in requests until it has expired. The intention is to scale in conservatively to protect your application's availability. However, if another alarm triggers a scale out activity during the cooldown period after a scale-in, Service Auto Scaling scales out your scalable target immediately.
 - g. Choose **Save**.
3. Choose **Next step**.
4. Review all of your choices and then choose **Update Service**.
5. When your service status is finished updating, choose **View Service**.

Step 3: Trigger a Scaling Activity

After your service is configured with Service Auto Scaling, you can trigger a scaling activity by pushing your service's CPU utilization into the `ALARM` state. Because the example in this tutorial is a web application that is running behind a load balancer, you can send thousands of HTTP requests to your service (using the `ApacheBench` utility) to spike the service CPU utilization above the threshold amount. This spike should trigger the alarm, which in turn triggers a scaling activity to add one task to your service.

After the `ApacheBench` utility finishes the requests, the service CPU utilization should drop below your 75% threshold, triggering a scale-in activity that returns the service's desired count to 1.

To trigger a scaling activity for your service

1. From your service's main view page in the console, choose the load balancer name to view its details in the Amazon EC2 console. You need the load balancer's DNS name, which should look something like `EC2Contai-EcsElast-SMAKV74U23PH-96652279.us-east-1.elb.amazonaws.com`.
2. Use the `ApacheBench` (**ab**) utility to make thousands of HTTP requests to your load balancer in a short period of time.

Note

This command is installed by default on macOS, and it is available for many Linux distributions, as well. For example, you can install **ab** on Amazon Linux with the following command:

```
$ sudo yum install -y httpd24-tools
```

Run the following command, substituting your load balancer's DNS name.

```
$ ab -n 100000 -c 1000 http://EC2Contai-EcsElast-SMAKV74U23PH-96652279.us-east-1.elb.amazonaws.com/
```

3. Open the CloudWatch console at <https://console.aws.amazon.com/cloudwatch/>.

4. In the left navigation pane, choose **Alarms**.
5. Wait for your **ab** HTTP requests to trigger the scale-out alarm in the CloudWatch console. You should see your Amazon ECS service scale out and add one task to your service's desired count.
6. Shortly after your **ab** HTTP requests complete (between 1 and 2 minutes), your scale in alarm should trigger and the scale in policy reduces your service's desired count back to 1.

Step 4: Next Steps

Go to the next step if you would like to delete the basic infrastructure that you just created for this tutorial. Otherwise, you can use this infrastructure as your base and try one or more of the following:

- To view these scaling activities from the Amazon ECS console, choose the **Events** tab of the service. When scaling events occur, you see informational messages here. For example:

```
Message: Successfully set desired count to 1. Change successfully fulfilled by ecs.  
Cause: monitor alarm TargetTracking-service/service-autoscaling/sample-webapp-AlarmLow-  
fcd80aef-5161-4890-aeb4-35dde11ff42c in state ALARM triggered policy TargetTrackingPolicy
```

- If you have CloudWatch Container Insights set up and it's collecting Amazon ECS metrics, you can view metric data on the CloudWatch automatic dashboards. For more information, see [Introducing Amazon CloudWatch Container Insights for Amazon ECS](#) in the *AWS Compute Blog*.
- Learn how to set up CloudWatch Container Insights. Additional charges apply. For more information, see [Amazon ECS CloudWatch Container Insights \(p. 197\)](#) and [Updating Cluster Settings \(p. 23\)](#).

Step 5: Cleaning Up

When you have completed this tutorial, you may choose to keep your cluster, Auto Scaling group, load balancer, target tracking scaling policy, and CloudWatch alarms. However, if you are not actively using these resources, you should consider cleaning them up so that your account does not incur unnecessary charges.

To delete your cluster

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the left navigation pane, choose **Clusters**.
3. On the **Clusters** page, choose the **service-autoscaling** cluster.
4. Choose **Delete Cluster**, **Delete**. It may take a few minutes for the cluster AWS CloudFormation stack to finish cleaning up.

Step Scaling Policies

Although Amazon ECS Service Auto Scaling supports using Application Auto Scaling step scaling policies, we recommend using target tracking scaling policies instead. For example, if you want to scale your service when CPU utilization falls below or rises above a certain level, create a target tracking scaling policy based on the CPU utilization metric provided by Amazon ECS. For more information, see [Target Tracking Scaling Policies \(p. 166\)](#).

With step scaling policies, you create and manage the CloudWatch alarms that trigger the scaling process. If the target tracking alarms don't work for your use case, you can use step scaling. You can also use target tracking scaling with step scaling for an advanced scaling policy configuration. For example, you can configure a more aggressive response when utilization reaches a certain level.

Service Auto Scaling Considerations

- Metrics are not available until the clusters and services send the metrics to CloudWatch, and you cannot create CloudWatch alarms for metrics that do not exist yet.
- The scaling policies that you create for Amazon ECS services support a cooldown period. This is the number of seconds after a scaling activity completes where previous scaling policy-related scaling activities can influence future scaling activities.
 - For scale-out policies, while the cooldown period is in effect, the capacity that has been added by the previous scale-out activity that initiated the cooldown is calculated as part of the desired capacity for the next scale out. The intention is to continuously (but not excessively) scale out.
 - For scale-in policies, the cooldown period is used to block subsequent scale in requests until it has expired. The intention is to scale in conservatively to protect your application's availability. However, if another alarm triggers a scale-out policy during the cooldown period after a scale in, automatic scaling scales out your service immediately.
- The ECS service scheduler respects the desired count at all times, but as long as you have active scaling policies and alarms on a service, Service Auto Scaling could change a desired count that was manually set by you.
- If a service's desired count is set below its minimum capacity value, and an alarm triggers a scale-out activity, Service Auto Scaling scales the desired count up to the minimum capacity value and then continues to scale out as required, based on the scaling policy associated with the alarm. However, a scale-in activity does not adjust the desired count, because it is already below the minimum capacity value.
- If a service's desired count is set above its maximum capacity value, and an alarm triggers a scale in activity, Service Auto Scaling scales the desired count out to the maximum capacity value and then continues to scale in as required, based on the scaling policy associated with the alarm. However, a scale-out activity does not adjust the desired count, because it is already above the maximum capacity value.
- During scaling activities, the actual running task count in a service is the value that Service Auto Scaling uses as its starting point, as opposed to the desired count, which is what processing capacity is supposed to be. This prevents excessive (runaway) scaling that could not be satisfied, for example, if there are not enough container instance resources to place the additional tasks. If the container instance capacity is available later, the pending scaling activity may succeed, and then further scaling activities can continue after the cooldown period.

Amazon ECS Console Experience

Service Auto Scaling is disabled by default. You can enable it by configuring scaling policies from the **Auto Scaling** tab of your services in the AWS Management Console for Amazon ECS.

For step-by-step guidance for working with scaling policies from the console, see [Creating a service \(p. 129\)](#) and [Updating a Service \(p. 139\)](#). For more information about step scaling and a walkthrough, see [Automatic Scaling with Amazon ECS](#) in the *AWS Compute Blog*. For a target tracking walkthrough, see [Target Tracking Scaling Policies \(p. 166\)](#).

When you configure scaling policies for a service in the Amazon ECS console, your service is automatically registered as a scalable target with Application Auto Scaling, and your scaling policies are automatically in force as soon as they're successfully created.

AWS CLI and SDK Experience

Service Auto Scaling is made possible by a combination of the Amazon ECS, CloudWatch, and Application Auto Scaling APIs. Services are created and updated with Amazon ECS, alarms are created with CloudWatch, and scaling policies are created with Application Auto Scaling.

For more information about these specific API operations, see the [Amazon Elastic Container Service API Reference](#), the [Amazon CloudWatch API Reference](#), and the [Application Auto Scaling API Reference](#). For more information about the AWS CLI commands for these services, see the [ecs](#), [cloudwatch](#), and [application-autoscaling](#) sections of the [AWS CLI Command Reference](#).

To configure scaling policies for your ECS service using the AWS CLI

1. Register your ECS service as a scalable target using the [register-scalable-target](#) command.
2. Create a scaling policy using the [put-scaling-policy](#) command.
3. [Step scaling] Create an alarm that triggers the scaling policy using the [put-metric-alarm](#) command.

For more information about configuring scaling policies using the AWS CLI, see the [Application Auto Scaling User Guide](#).

Service Discovery

Your Amazon ECS service can optionally be configured to use Amazon ECS Service Discovery. Service discovery uses AWS Cloud Map API actions to manage HTTP and DNS namespaces for your Amazon ECS services. For more information, see [What Is AWS Cloud Map?](#) in the *AWS Cloud Map Developer Guide*.

Service discovery is available in the following AWS Regions:

Region Name	Region
US East (N. Virginia)	us-east-1
US East (Ohio)	us-east-2
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Asia Pacific (Hong Kong)	ap-east-1
Asia Pacific (Mumbai)	ap-south-1
Asia Pacific (Tokyo)	ap-northeast-1
Asia Pacific (Seoul)	ap-northeast-2
Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Canada (Central)	ca-central-1
Europe (Frankfurt)	eu-central-1
Europe (Ireland)	eu-west-1
Europe (London)	eu-west-2
Europe (Paris)	eu-west-3
Europe (Stockholm)	eu-north-1
Middle East (Bahrain)	me-south-1

Region Name	Region
South America (São Paulo)	sa-east-1

Service Discovery Concepts

Service discovery consists of the following components:

- **Service discovery namespace:** A logical group of service discovery services that share the same domain name, such as `example.com`.
- **Service discovery service:** Exists within the service discovery namespace and consists of the service name and DNS configuration for the namespace. It provides the following core component:
 - **Service registry:** Allows you to look up a service via DNS or AWS Cloud Map API actions and get back one or more available endpoints that can be used to connect to the service.
- **Service discovery instance:** Exists within the service discovery service and consists of the attributes associated with each Amazon ECS service in the service directory.
 - **Instance attributes:** The following metadata is added as custom attributes for each Amazon ECS service that is configured to use service discovery:
 - **AWS_INSTANCE_IPV4** – For an A record, the IPv4 address that Route 53 returns in response to DNS queries and AWS Cloud Map returns when discovering instance details, for example, `192.0.2.44`.
 - **AWS_INSTANCE_PORT** – The port value associated with the service discovery service.
 - **AVAILABILITY_ZONE** – The Availability Zone into which the task was launched. For tasks using the EC2 launch type, this is the Availability Zone in which the container instance exists. For tasks using the Fargate launch type, this is the Availability Zone in which the elastic network interface exists.
 - **REGION** – The Region in which the task exists.
 - **ECS_SERVICE_NAME** – The name of the Amazon ECS service to which the task belongs.
 - **ECS_CLUSTER_NAME** – The name of the Amazon ECS cluster to which the task belongs.
 - **EC2_INSTANCE_ID** – The ID of the container instance the task was placed on. This custom attribute is not added if the task is using the Fargate launch type.
 - **ECS_TASK_DEFINITION_FAMILY** – The task definition family that the task is using.
 - **ECS_TASK_SET_EXTERNAL_ID** – If a task set is created for an external deployment and is associated with a service discovery registry, then the `ECS_TASK_SET_EXTERNAL_ID` attribute will contain the external ID of the task set.
- **Amazon ECS health checks:** Amazon ECS performs periodic container-level health checks. If an endpoint does not pass the health check, it is removed from DNS routing and marked as unhealthy.

Service Discovery Considerations

The following should be considered when using service discovery:

- Service discovery is supported for tasks using the Fargate launch type if they are using platform version v1.1.0 or later. For more information, see [AWS Fargate platform versions \(p. 14\)](#).
- The Create Service workflow in the Amazon ECS console only supports registering services into private DNS namespaces. When a AWS Cloud Map private DNS namespace is created, a Route 53 private hosted zone will be created automatically.
- The DNS records created for a service discovery service always register with the private IP address for the task, rather than the public IP address, even when public namespaces are used.

- Service discovery requires that tasks specify either the `awsvpc`, `bridge`, or `host` network mode (`none` is not supported).
- If the task definition your service task specifies uses the `awsvpc` network mode, you can create any combination of A or SRV records for each service task. If you use SRV records, a port is required.
- If the task definition that your service task specifies uses the `bridge` or `host` network mode, an SRV record is the only supported DNS record type. Create an SRV record for each service task. The SRV record must specify a container name and container port combination from the task definition.
- DNS records for a service discovery service can be queried within your VPC. They use the following format: `<service discovery service name>.<service discovery namespace>`. For more information, see [Step 3: Verify Service Discovery \(p. 323\)](#).
- When doing a DNS query on the service name, A records return a set of IP addresses that correspond to your tasks. SRV records return a set of IP addresses and ports per task.
- If you have eight or fewer healthy records, Route 53 responds to all DNS queries with all of the healthy records.
- When all records are unhealthy, Route 53 responds to DNS queries with up to eight unhealthy records.
- You can configure service discovery for an ECS service that is behind a load balancer, but service discovery traffic is always routed to the task and not the load balancer.
- Service discovery does not support the use of Classic Load Balancers.
- It is recommended to use container-level health checks managed by Amazon ECS for your service discovery service.
 - **HealthCheckCustomConfig**—Amazon ECS manages health checks on your behalf. Amazon ECS uses information from container and health checks, as well as your task state, to update the health with AWS Cloud Map. This is specified using the `--health-check-custom-config` parameter when creating your service discovery service. For more information, see [HealthCheckCustomConfig](#) in the *AWS Cloud Map API Reference*.
- If you are using the Amazon ECS console, the workflow creates one service discovery service per ECS service. It maps all of the task IP addresses as A records, or task IP addresses and port as SRV records.
- Service discovery can only be configured when first creating a service. Updating existing services to configure service discovery for the first time or change the current configuration is not supported.
- The AWS Cloud Map resources created when service discovery is used must be cleaned up manually. For more information, see [Step 4: Clean Up \(p. 325\)](#) in the [Tutorial: Creating a Service Using Service Discovery \(p. 317\)](#) topic.

Amazon ECS Console Experience

The service creation workflow in the Amazon ECS console supports service discovery. Service discovery can only be configured when first creating a service. Updating existing services to configure service discovery for the first time or change the current configuration is not supported.

To create a new Amazon ECS service that uses service discovery, see [Creating a service \(p. 129\)](#).

Service Discovery Pricing

Customers using Amazon ECS service discovery are charged for Route 53 resources and AWS Cloud Map discovery API operations. This involves costs for creating the Route 53 hosted zones and queries to the service registry. For more information, see [AWS Cloud Map Pricing](#) in the *AWS Cloud Map Developer Guide*.

Amazon ECS performs container level health checks and exposes them to AWS Cloud Map custom health check API operations. This is currently made available to customers at no extra cost. If you configure additional network health checks for publicly exposed tasks, you are charged for those health checks.

Service Throttle Logic

The Amazon ECS service scheduler includes logic that throttles how often service tasks are launched if they repeatedly fail to start.

If tasks for an ECS service repeatedly fail to enter the `RUNNING` state (progressing directly from `PENDING` to `STOPPED`), then the time between subsequent restart attempts is incrementally increased up to a maximum of 15 minutes. This maximum period is subject to change in the future and should not be considered permanent. This behavior reduces the effect that unstartable tasks have on your Amazon ECS cluster resources or Fargate infrastructure costs. If your service triggers the throttle logic, you receive the following [service event message](#) (p. 349):

```
(service service-name) is unable to consistently start tasks successfully.
```

Amazon ECS does not ever stop a failing service from retrying, nor does it attempt to modify it in any way other than increasing the time between restarts. The service throttle logic does not provide any user-tunable parameters.

If you update your service to use a new task definition, your service returns to a normal, non-throttled state immediately. For more information, see [Updating a Service](#) (p. 139).

The following are some common causes that trigger this logic:

- The Amazon ECS container agent is unable to pull your task Docker image. This could be due to a bad container image name, image, or tag, or a lack of private registry authentication or permissions. In this case, you also see `CannotPullContainerError` in your [stopped task errors](#) (p. 340).

Important

Tasks that are stopped after they reach the `RUNNING` state do not trigger the throttle logic or the associated service event message. For example, if failed Elastic Load Balancing health checks for a service cause a task to be flagged as unhealthy, and Amazon ECS deregisters it and kills the task, this does not trigger the throttle. Even if a task's container command immediately exits with a non-zero exit code, the task has already moved to the `RUNNING` state. Tasks that fail immediately due to command errors do not trigger the throttle or the service event message.

Resources and Tags

Amazon ECS resources, including task definitions, clusters, tasks, services, and container instances, are assigned an Amazon Resource Name (ARN) and a unique resource identifier (ID). These resources can be tagged with values that you define, to help you organize and identify them.

The following topics describe resources and tags, and how you can work with them.

Contents

- [Tagging Your Amazon ECS Resources \(p. 177\)](#)
- [Amazon ECS Usage Reports \(p. 181\)](#)

Tagging Your Amazon ECS Resources

To help you manage your Amazon ECS tasks, services, task sets, task definitions, clusters, and container instances, you can optionally assign your own metadata to each resource in the form of *tags*. This topic describes tags and shows you how to create them.

Important

To use this feature, it requires that you opt-in to the new Amazon Resource Name (ARN) and resource identifier (ID) formats. For more information, see [Amazon Resource Names \(ARNs\) and IDs \(p. 104\)](#).

Tag Basics

A tag is a label that you assign to an AWS resource. Each tag consists of a *key* and an optional *value*, both of which you define.

Tags enable you to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. This is useful when you have many resources of the same type—you can quickly identify a specific resource based on the tags you've assigned to it. For example, you could define a set of tags for your account's Amazon ECS container instances that helps you track each container instance's owner and stack level.

We recommend that you devise a set of tag keys that meets your needs for each resource type. Using a consistent set of tag keys makes it easier for you to manage your resources. You can search and filter the resources based on the tags you add.

Tags don't have any semantic meaning to Amazon ECS and are interpreted strictly as a string of characters. Also, tags are not automatically assigned to your resources. You can edit tag keys and values, and you can remove tags from a resource at any time. You can set the value of a tag to an empty string, but you can't set the value of a tag to null. If you add a tag that has the same key as an existing tag on that resource, the new value overwrites the old value. If you delete a resource, any tags for the resource are also deleted.

You can work with tags using the AWS Management Console, the AWS CLI, and the Amazon ECS API.

If you're using AWS Identity and Access Management (IAM), you can control which users in your AWS account have permission to create, edit, or delete tags.

Tagging Your Resources

You can tag new or existing Amazon ECS tasks, services, task definitions, and clusters.

If you're using the Amazon ECS console, you can apply tags to new resources when they are created or existing resources by using the **Tags** tab on the relevant resource page at any time. The **Propagate tags from** option can be used when running a task to copy the tags from the task definition to the task or when creating a service to copy the tags from either the service or the task definition to the tasks in the service.

If you're using the Amazon ECS API, the AWS CLI, or an AWS SDK, you can apply tags to new resources using the `tags` parameter on the relevant API action or use the `TagResource` API action to apply tags to existing resources. For more information, see [TagResource](#). The `propagateTags` parameter can be used when running a task to copy the tags from the task definition to the task or when creating a service to copy the tags from either the service or the task definition to the tasks in the service. For more information, see [RunTask](#) and [CreateService](#).

Additionally, some resource-creating actions enable you to specify tags for a resource when the resource is created. If tags cannot be applied during resource creation, we roll back the resource creation process. This ensures that resources are either created with tags or not created at all, and that no resources are left untagged at any time. By tagging resources at the time of creation, you can eliminate the need to run custom tagging scripts after resource creation.

The following table describes the Amazon ECS resources that can be tagged, and the resources that can be tagged on creation.

Tagging Support for Amazon ECS Resources

Resource	Supports tags	Supports tag propagation	Supports tagging on creation (Amazon ECS API, AWS CLI, AWS SDK)
Amazon ECS tasks	Yes	Yes, from the task definition.	Yes
Amazon ECS services	Yes	Yes, from either the task definition or the service to the tasks in the service.	Yes
Amazon ECS task sets	Yes	No	Yes
Amazon ECS task definitions	Yes	No	Yes
Amazon ECS clusters	Yes	No	Yes

Tag Restrictions

The following basic restrictions apply to tags:

- Maximum number of tags per resource – 50
- For each resource, each tag key must be unique, and each tag key can have only one value.
- Maximum key length – 128 Unicode characters in UTF-8
- Maximum value length – 256 Unicode characters in UTF-8
- If your tagging schema is used across multiple services and resources, remember that other services may have restrictions on allowed characters. Generally allowed characters are: letters, numbers, and spaces representable in UTF-8, and the following characters: `+ - = . _ : / @`.
- Tag keys and values are case-sensitive.

- Don't use `aws :`, `AWS :`, or any upper or lowercase combination of such as a prefix for either keys or values as it is reserved for AWS use. You can't edit or delete tag keys or values with this prefix. Tags with this prefix do not count against your tags per resource limit.

Tagging Your Resources for Billing

When enabling Amazon ECS managed tags, Amazon ECS will automatically tag all newly launched tasks with the cluster name. For tasks that belong to a service, they will be tagged with the service name as well. These managed tags are helpful when reviewing cost allocation after enabling them in your Cost & Usage Report. For more information, see [Amazon ECS Usage Reports \(p. 181\)](#).

To see the cost of your combined resources, you can organize your billing information based on resources that have the same tag key values. For example, you can tag several resources with a specific application name, and then organize your billing information to see the total cost of that application across several services. For more information about setting up a cost allocation report with tags, see [The Monthly Cost Allocation Report](#) in the *AWS Billing and Cost Management User Guide*.

Important

To use this feature, it requires that you opt-in to the new Amazon Resource Name (ARN) and resource identifier (ID) formats. For more information, see [Amazon Resource Names \(ARNs\) and IDs \(p. 104\)](#).

Note

If you've just enabled reporting, data for the current month is available for viewing after 24 hours.

Working with Tags Using the Console

Using the Amazon ECS console, you can manage the tags associated with new or existing tasks, services, task definitions, clusters, or container instances.

When you select a resource-specific page in the Amazon ECS console, it displays a list of those resources. For example, if you select **Clusters** from the navigation pane, the console displays a list of Amazon ECS clusters. When you select a resource from one of these lists (for example, a specific cluster), if the resource supports tags, you can view and manage its tags on the **Tags** tab.

Contents

- [Adding Tags on an Individual Resource During Launch \(p. 179\)](#)
- [Adding and Deleting Tags on an Individual Resource \(p. 180\)](#)

Adding Tags on an Individual Resource During Launch

The following resources allow you to specify tags when you create the resource.

Task	Console
Run one or more tasks.	Running Tasks (p. 109)
Create a service.	Creating a service (p. 129)
Create a task set.	External Deployment (p. 147)
Register a task definition.	Creating a Task Definition (p. 30)
Create a cluster.	Creating a Cluster (p. 18)

Adding and Deleting Tags on an Individual Resource

Amazon ECS allows you to add or delete tags associated with your clusters, services, tasks, and task definitions directly from the resource's page.

To add a tag to an individual resource

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. From the navigation bar, select the region to use.
3. In the navigation pane, select a resource type (for example, **Clusters**).
4. Select the resource from the resource list and choose **Tags, Edit**.
5. In the **Edit Tags** dialog box, specify the key and value for each tag, and then choose **Save**.

To delete a tag from an individual resource

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. From the navigation bar, select the region to use.
3. In the navigation pane, choose a resource type (for example, **Clusters**).
4. Select the resource from the resource list and choose **Tags, Edit**.
5. On the **Edit Tags** page, select the **Delete** icon for each tag you want to delete, and choose **Save**.

Working with Tags Using the CLI or API

Use the following to add, update, list, and delete the tags for your resources. The corresponding documentation provides examples.

Tagging Support for Amazon ECS Resources

Task	AWS CLI	API Action
Add or overwrite one or more tags.	tag-resource	TagResource
Delete one or more tags.	untag-resource	UntagResource

The following examples show how to tag or untag resources using the AWS CLI.

Example 1: Tag an existing cluster

The following command tags an existing cluster.

```
aws ecs tag-resource --resource-arn resource_ARN --tags key=stack,value=dev
```

Example 2: Untag an existing cluster

The following command deletes a tag from an existing cluster.

```
aws ecs untag-resource --resource-arn resource_ARN --tag-keys tag_key
```

Example 3: List tags for a resource

The following command lists the tags associated with an existing resource.

```
aws ecs list-tags-for-resource --resource-arn resource_ARN
```

Some resource-creating actions enable you to specify tags when you create the resource. The following actions support tagging on creation.

Task	AWS CLI	AWS Tools for Windows PowerShell	API Action
Run one or more tasks.	run-task	Start-ECSTask	RunTask
Create a service.	create-service	New-ECSService	CreateService
Create a task set.	create-task-set	New-ECSTaskSet	CreateTaskSet
Register a task definition.	register-task-definition	Register-ECSTaskDefinition	RegisterTaskDefinition
Create a cluster.	create-cluster	New-ECSCluster	CreateCluster

The following examples demonstrate how to apply tags when you create resources.

Example 1: Create a cluster and apply a tag

The following command creates a cluster named `devcluster` and adds a tag with key `team` and value `devs`.

```
aws ecs create-cluster --cluster-name devcluster --tags key=team,value=devs
```

Example 2: Create a service and apply a tag

The following command creates a service named `application` and adds a tag with key `stack` and value `dev`.

```
aws ecs create-service --service-name application --task-definition task-def-app --tags key=stack,value=dev
```

Example 3: Create a service with tags and propagate the tags to the tasks in the service

The `--propagateTags` parameter can be used to copy the tags from either a task definition or a service to the tasks in a service. The following command creates a service with tags and propagates them to the tasks in that service.

```
aws ecs create-service --service-name application --task-definition task-def-app --tags key=stack,value=dev --propagateTags Service
```

Amazon ECS Usage Reports

AWS provides a free reporting tool called Cost Explorer that enables you to analyze the cost and usage of your Amazon ECS resources.

Cost Explorer is a free tool that you can use to view charts of your usage and costs. You can view data from the last 13 months, and forecast how much you are likely to spend for the next three months. You can use Cost Explorer to see patterns in how much you spend on AWS resources over time, identify areas

that need further inquiry, and see trends that you can use to understand your costs. You also can specify time ranges for the data, and view time data by day or by month.

The metering data in your Cost & Usage Report shows usage across all of your Amazon ECS tasks. The metering data includes CPU usage as `vCPU-Hours` and memory usage as `GB-Hours` for each task that was run. How that data is presented depends on the launch type of the task, as described below.

For tasks using the Fargate launch type, the `lineItem/Operation` column will show `FargateTask` and you will see the cost associated with each task.

You can also use the Amazon ECS managed tags to identify the service or cluster that each task belongs to. For more information, see [Tagging Your Resources for Billing \(p. 179\)](#).

Important

The metering data is only viewable for tasks launched on or after November 16, 2018. Tasks launched prior to this date will not show metering data.

Here's an example of some of the fields you can sort cost allocation data by when using Cost Explorer:

- Cluster name
- Service name
- Resource tags
- Launch type
- Region
- Usage type

For more information about creating an AWS Cost and Usage Report, see [AWS Cost and Usage Report](#) in the *AWS Billing and Cost Management User Guide*.

Monitoring Amazon ECS

You can monitor your Amazon ECS resources using Amazon CloudWatch, which collects and processes raw data from Amazon ECS into readable, near real-time metrics. These statistics are recorded for a period of two weeks, so that you can access historical information and gain a better perspective on how your clusters or services are performing. Amazon ECS metric data is automatically sent to CloudWatch in 1-minute periods. For more information about CloudWatch, see the [Amazon CloudWatch User Guide](#).

Monitoring is an important part of maintaining the reliability, availability, and performance of Amazon ECS and your AWS solutions. You should collect monitoring data from all of the parts of your AWS solution so that you can more easily debug a multi-point failure if one occurs. Before you start monitoring Amazon ECS; however, you should create a monitoring plan that includes answers to the following questions:

- What are your monitoring goals?
- What resources will you monitor?
- How often will you monitor these resources?
- What monitoring tools will you use?
- Who will perform the monitoring tasks?
- Who should be notified when something goes wrong?

When you are using the Fargate launch type, you get CPU and memory utilization metrics for each of your services to assist in the monitoring of your environment.

The next step is to establish a baseline for normal Amazon ECS performance in your environment, by measuring performance at various times and under different load conditions. As you monitor Amazon ECS, store historical monitoring data so that you can compare it with current performance data, identify normal performance patterns and performance anomalies, and devise methods to address issues.

Topics

- [Monitoring Tools \(p. 183\)](#)
- [Amazon ECS CloudWatch Metrics \(p. 184\)](#)
- [Amazon ECS Events and EventBridge \(p. 189\)](#)
- [Amazon ECS CloudWatch Container Insights \(p. 197\)](#)
- [Logging Amazon ECS API Calls with AWS CloudTrail \(p. 199\)](#)

Monitoring Tools

AWS provides various tools that you can use to monitor Amazon ECS. You can configure some of these tools to do the monitoring for you, while some of the tools require manual intervention. We recommend that you automate monitoring tasks as much as possible.

Automated Monitoring Tools

You can use the following automated monitoring tools to watch Amazon ECS and report when something is wrong:

- Amazon CloudWatch alarms – Watch a single metric over a time period that you specify, and perform one or more actions based on the value of the metric relative to a given threshold over a number of time periods. The action is a notification sent to an Amazon Simple Notification Service (Amazon SNS) topic or Amazon EC2 Auto Scaling policy. CloudWatch alarms do not invoke actions simply because they are in a particular state; the state must have changed and been maintained for a specified number of periods. For more information, see [Amazon ECS CloudWatch Metrics \(p. 184\)](#).

For services with tasks that use the Fargate launch type, you can use CloudWatch alarms to scale in and scale out the tasks in your service based on CloudWatch metrics, such as CPU and memory utilization. For more information, see [Service Auto Scaling \(p. 165\)](#).

- Amazon CloudWatch Logs – Monitor, store, and access the log files from the containers in your Amazon ECS tasks by specifying the `awslogs` log driver in your task definitions. This is the only supported method for accessing logs for tasks using the Fargate launch type. For more information, see [Using the awslogs Log Driver \(p. 69\)](#).
- Amazon CloudWatch Events – Match events and route them to one or more target functions or streams to make changes, capture state information, and take corrective action. For more information, see [Amazon ECS Events and EventBridge \(p. 189\)](#) in this guide and [What Is Amazon CloudWatch Events?](#) in the *Amazon CloudWatch Events User Guide*.
- AWS CloudTrail log monitoring – Share log files between accounts, monitor CloudTrail log files in real time by sending them to CloudWatch Logs, write log processing applications in Java, and validate that your log files have not changed after delivery by CloudTrail. For more information, see [Logging Amazon ECS API Calls with AWS CloudTrail \(p. 199\)](#) in this guide, and [Working with CloudTrail Log Files](#) in the *AWS CloudTrail User Guide*.

Manual Monitoring Tools

Another important part of monitoring Amazon ECS involves manually monitoring those items that the CloudWatch alarms don't cover. The CloudWatch, Trusted Advisor, and other AWS console dashboards provide an at-a-glance view of the state of your AWS environment. We recommend that you also check the log files on your container instances and the containers in your tasks.

- CloudWatch home page:
 - Current alarms and status
 - Graphs of alarms and resources
 - Service health status

In addition, you can use CloudWatch to do the following:

- Create [customized dashboards](#) to monitor the services you care about.
- Graph metric data to troubleshoot issues and discover trends.
- Search and browse all your AWS resource metrics.
- Create and edit alarms to be notified of problems.
- AWS Trusted Advisor can help you monitor your AWS resources to improve performance, reliability, security, and cost effectiveness. Four Trusted Advisor checks are available to all users; more than 50 checks are available to users with a Business or Enterprise support plan. For more information, see [AWS Trusted Advisor](#).

Amazon ECS CloudWatch Metrics

You can monitor your Amazon ECS resources using Amazon CloudWatch, which collects and processes raw data from Amazon ECS into readable, near real-time metrics. These statistics are recorded for a period of two weeks so that you can access historical information and gain a better perspective on how

your clusters or services are performing. Amazon ECS metric data is automatically sent to CloudWatch in 1-minute periods. For more information about CloudWatch, see the [Amazon CloudWatch User Guide](#).

Topics

- [Enabling CloudWatch Metrics \(p. 185\)](#)
- [Available Metrics and Dimensions \(p. 185\)](#)
- [Service Utilization \(p. 187\)](#)
- [Service RUNNING Task Count \(p. 188\)](#)
- [Viewing Amazon ECS Metrics \(p. 188\)](#)

Enabling CloudWatch Metrics

Any Amazon ECS service using the Fargate launch type is enabled for CloudWatch CPU and memory utilization metrics automatically, so you don't need to take any manual steps.

Available Metrics and Dimensions

The following sections list the metrics and dimensions that Amazon ECS sends to Amazon CloudWatch.

Amazon ECS Metrics

Amazon ECS provides metrics for you to monitor your resources. You can measure the CPU and memory reservation and utilization across your cluster as a whole, and the CPU and memory utilization on the services in your clusters. For your GPU workloads, you can measure your GPU reservation across your cluster.

The metrics made available will depend on the launch type of the tasks and services in your clusters. If you're using the Fargate launch type for your services, CPU and memory utilization metrics are provided to assist in the monitoring of your services. For the EC2 launch type, you will own and need to monitor the Amazon EC2 instances that make your underlying infrastructure. Accordingly, additional CPU, memory, and GPU reservation and CPU and memory utilization metrics are made available at the cluster, service, and task level.

Amazon ECS sends the following metrics to CloudWatch every minute. When Amazon ECS collects metrics, it collects multiple data points every minute. It then aggregates them to one data point before sending the data to CloudWatch. So in CloudWatch, one sample count is actually the aggregate of multiple data points during one minute.

The AWS/ECS namespace includes the following metrics.

CPUReservation

The percentage of CPU units that are reserved by running tasks in the cluster.

Cluster CPU reservation (this metric can only be filtered by `ClusterName`) is measured as the total CPU units that are reserved by Amazon ECS tasks on the cluster, divided by the total CPU units that were registered for all of the container instances in the cluster. Only container instances in `ACTIVE` or `DRAINING` status will affect CPU reservation metrics. This metric is only used for tasks using the EC2 launch type.

Valid dimensions: `ClusterName`.

Valid statistics: Average, Minimum, Maximum, Sum, Sample Count. The most useful statistic is Average.

Unit: Percent.

CPUUtilization

The percentage of CPU units that are used in the cluster or service.

Cluster CPU utilization (metrics that are filtered by `ClusterName` without `ServiceName`) is measured as the total CPU units in use by Amazon ECS tasks on the cluster, divided by the total CPU units that were registered for all of the container instances in the cluster. Only container instances in `ACTIVE` or `DRAINING` status will affect CPU utilization metrics. Cluster CPU utilization metrics are only used for tasks using the EC2 launch type.

Service CPU utilization (metrics that are filtered by `ClusterName` and `ServiceName`) is measured as the total CPU units in use by the tasks that belong to the service, divided by the total number of CPU units that are reserved for the tasks that belong to the service. Service CPU utilization metrics are used for tasks using both the Fargate and the EC2 launch type.

Valid dimensions: `ClusterName`, `ServiceName`.

Valid statistics: Average, Minimum, Maximum, Sum, Sample Count. The most useful statistic is Average.

Unit: Percent.

MemoryReservation

The percentage of memory that is reserved by running tasks in the cluster.

Cluster memory reservation (this metric can only be filtered by `ClusterName`) is measured as the total memory that is reserved by Amazon ECS tasks on the cluster, divided by the total amount of memory that was registered for all of the container instances in the cluster. Only container instances in `ACTIVE` or `DRAINING` status will affect memory reservation metrics. This metric is only used for tasks using the EC2 launch type.

Valid dimensions: `ClusterName`.

Valid statistics: Average, Minimum, Maximum, Sum, Sample Count. The most useful statistic is Average.

Unit: Percent.

MemoryUtilization

The percentage of memory that is used in the cluster or service.

Cluster memory utilization (metrics that are filtered by `ClusterName` without `ServiceName`) is measured as the total memory in use by Amazon ECS tasks on the cluster, divided by the total amount of memory that was registered for all of the container instances in the cluster. Only container instances in `ACTIVE` or `DRAINING` status will affect memory utilization metrics. Cluster memory utilization metrics are only used for tasks using the EC2 launch type.

Service memory utilization (metrics that are filtered by `ClusterName` and `ServiceName`) is measured as the total memory in use by the tasks that belong to the service, divided by the total memory that is reserved for the tasks that belong to the service. Service memory utilization metrics are used for tasks using both the Fargate and EC2 launch types.

Valid dimensions: `ClusterName`, `ServiceName`.

Valid statistics: Average, Minimum, Maximum, Sum, Sample Count. The most useful statistic is Average.

Unit: Percent.

GPUReservation

The percentage of total available GPUs that are reserved by running tasks in the cluster.

Cluster GPU reservation is measured as the number of GPUs reserved by Amazon ECS tasks on the cluster, divided by the total number of GPUs that was available on all of the GPU-enabled container instances in the cluster. Only container instances in `ACTIVE` or `DRAINING` status will affect GPU reservation metrics.

Valid dimensions: `ClusterName`.

Valid statistics: Average, Minimum, Maximum, Sum, Sample Count. The most useful statistic is Average.

Unit: Percent.

Note

If you're using tasks with the EC2 launch type and have Linux container instances, the Amazon ECS container agent relies on Docker `stats` metrics to gather CPU and memory data for each container running on the instance. For burstable performance instances (T3, T3a, and T2 instances), the CPU utilization metric may reflect different data compared to instance-level CPU metrics.

Dimensions for Amazon ECS Metrics

Amazon ECS metrics use the `AWS/ECS` namespace and provide metrics for the following dimensions. Metrics for a dimension only reflect the resources with running tasks during a period. For example, if you have a cluster with one service in it but that service has no tasks in a `RUNNING` state, there will be no metrics sent to CloudWatch. If you have two services and one of them has running tasks and the other doesn't, only the metrics for the service with running tasks would be sent.

ClusterName

This dimension filters the data that you request for all resources in a specified cluster. All Amazon ECS metrics are filtered by `ClusterName`.

ServiceName

This dimension filters the data that you request for all resources in a specified service within a specified cluster.

Service Utilization

Service utilization is measured as the percentage of CPU and memory that is used by the Amazon ECS tasks that belong to a service on a cluster when compared to the CPU and memory that is specified in the service's task definition. This metric is supported for services with tasks using the Fargate launch type.

$$\text{Service CPU utilization} = \frac{(\text{Total CPU units used by tasks in service}) \times 100}{(\text{Total CPU units specified in task definition}) \times (\text{number of tasks in service})}$$

$$\frac{(\text{Total MiB of memory used by tasks in service}) \times 100}{(\text{Total MiB of memory specified in task definition}) \times (\text{number of tasks in service})}$$

```
Service memory utilization =
-----
(Total MiB of memory specified in task definition) x (number
of tasks in service)
```

Each minute, the Amazon ECS container agent associated with each task calculates the number of CPU units and MiB of memory that are currently being used for each task owned by the service, and this information is reported back to Amazon ECS. The total amount of CPU and memory used for all tasks owned by the service that are running on the cluster is calculated, and those numbers are reported to CloudWatch as a percentage of the total resources that are specified for the service in the service's task definition. If you specify a soft limit (`memoryReservation`), it's used to calculate the amount of reserved memory. Otherwise, the hard limit (`memory`) is used. For more information about hard and soft limits, see [Task Definition Parameters](#).

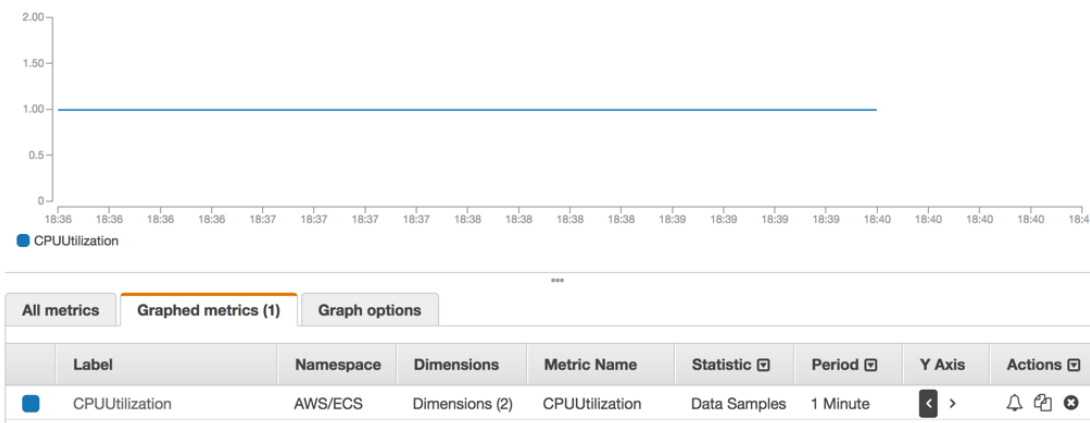
Service RUNNING Task Count

You can use CloudWatch metrics to view the number of tasks in your services that are in the `RUNNING` state. For example, you can set a CloudWatch alarm for this metric to alert you if the number of running tasks in your service falls below a specified value.

To view the number of running tasks in a service

1. Open the CloudWatch console at <https://console.aws.amazon.com/cloudwatch/>.
2. On the navigation pane, choose **Metrics**.
3. On the **All metrics** tab, choose **ECS**.
4. Choose **ClusterName**, **ServiceName** and then choose any metric (either `CPUUtilization` or `MemoryUtilization`) that corresponds to the service to view running tasks in.
5. On the **Graphed metrics** tab, change **Period** to **1 Minute** and **Statistic** to **Sample Count**.

The value displayed in the graph indicates the number of `RUNNING` tasks in the service.



Viewing Amazon ECS Metrics

After you have enabled CloudWatch metrics for Amazon ECS, you can view those metrics on the Amazon ECS and CloudWatch consoles. The Amazon ECS console provides a 24-hour maximum, minimum, and average view of your service metrics. The CloudWatch console provides a fine-grained and customizable display of your resources, as well as the number of running tasks in a service.

Topics

- [Viewing Service Metrics on the Amazon ECS Console \(p. 189\)](#)

- [Viewing Amazon ECS Metrics on the CloudWatch Console \(p. 189\)](#)

Viewing Service Metrics on the Amazon ECS Console

Amazon ECS service CPU and memory utilization metrics are available on the Amazon ECS console. The view provided for service metrics shows the average, minimum, and maximum values for the previous 24-hour period, with data points available in 5-minute intervals. For more information, see [Service Utilization \(p. 187\)](#).

To view service metrics in the console

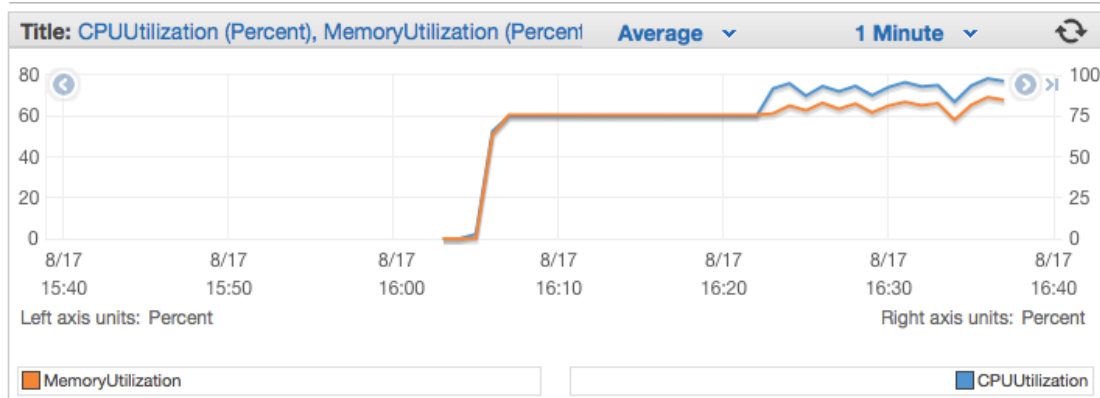
1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. Select the cluster that contains the service that you want to view metrics for.
3. On the **Cluster: *cluster-name*** page, choose **Services**.
4. Choose the service that you want to view metrics for.
5. On the **Service: *service-name*** page, choose **Metrics**.

Viewing Amazon ECS Metrics on the CloudWatch Console

Amazon ECS service metrics can also be viewed on the CloudWatch console. The console provides the most detailed view of Amazon ECS metrics, and you can tailor the views to suit your needs. You can view [Service Utilization \(p. 187\)](#) and the [Service RUNNING Task Count \(p. 188\)](#). For more information about CloudWatch, see the [Amazon CloudWatch User Guide](#).

To view metrics in the CloudWatch console

1. Open the CloudWatch console at <https://console.aws.amazon.com/cloudwatch/>.
2. In the **Metrics** section in the navigation pane, choose **ECS**.
3. Choose the metrics to view. Cluster metrics are scoped as **ECS > ClusterName** and service utilization metrics are scoped as **ECS > ClusterName, ServiceName**. The following example shows cluster CPU and memory utilization.



Amazon ECS Events and EventBridge

Amazon EventBridge enables you to automate your AWS services and respond automatically to system events such as application availability issues or resource changes. Events from AWS services are delivered to EventBridge in near real time. You can write simple rules to indicate which events are of interest

to you and what automated actions to take when an event matches a rule. The actions that can be automatically triggered include the following:

- Adding events to log groups in CloudWatch Logs
- Invoking an AWS Lambda function
- Invoking Amazon EC2 Run Command
- Relaying the event to Amazon Kinesis Data Streams
- Activating an AWS Step Functions state machine
- Notifying an Amazon SNS topic or an Amazon Simple Queue Service (Amazon SQS) queue

For more information, see [Getting Started with Amazon EventBridge](#) in the *Amazon EventBridge User Guide*.

You can use Amazon ECS events for EventBridge to receive near real-time notifications regarding the current state of your Amazon ECS clusters. When using the Fargate launch type, you can see the state of your tasks. For services, you can see events related to the health of your service.

Using EventBridge, you can build custom schedulers on top of Amazon ECS that are responsible for orchestrating tasks across clusters and monitoring the state of clusters in near real time. You can eliminate scheduling and monitoring code that continuously polls the Amazon ECS service for status changes and instead handle Amazon ECS state changes asynchronously using any EventBridge target. Targets might include AWS Lambda, Amazon Simple Queue Service, Amazon Simple Notification Service, or Amazon Kinesis Data Streams.

An Amazon ECS event stream ensures that every event is delivered at least one time. If duplicate events are sent, the event provides enough information to identify duplicates. For more information, see [Handling Events](#) (p. 196).

Events are relatively ordered, so that you can easily tell when an event occurred in relation to other events.

Topics

- [Amazon ECS Events](#) (p. 190)
- [Handling Events](#) (p. 196)

Amazon ECS Events

Amazon ECS tracks the state of each of your tasks and services. If the state of a task or service changes, an event is triggered and is sent to Amazon EventBridge. These events are classified as task state change events and service action events. These events and their possible causes are described in greater detail in the following sections.

Note

Amazon ECS may add other event types, sources, and details in the future. If you are programmatically deserializing event JSON data, make sure that your application is prepared to handle unknown properties to avoid issues if and when these additional properties are added.

Container state change and task state change events contain two `version` fields: one in the main body of the event, and one in the `detail` object of the event. The following describes the differences between these two fields:

- The `version` field in the main body of the event is set to 0 on all events. For more information about EventBridge parameters, see [Events and Event Patterns](#) in the *Amazon EventBridge User Guide*.
- The `version` field in the `detail` object of the event describes the version of the associated resource. Each time a resource changes state, this version is incremented. Because events can be sent multiple

times, this field allows you to identify duplicate events. Duplicate events have the same version in the `detail` object. If you are replicating your task state with EventBridge, you can compare the version of a resource reported by the Amazon ECS APIs with the version reported in EventBridge for the resource (inside the `detail` object) to verify that the version in your event stream is current.

Service action events only contain the `version` field in the main body.

Task State Change Events

The following scenarios trigger task state change events:

You call the `StartTask`, `RunTask`, or `StopTask` API operations, either directly or with the AWS Management Console, AWS CLI, or SDKs.

Starting or stopping tasks creates new task resources or modifies the state of existing task resources. The Amazon ECS service scheduler starts or stops a task.

Starting or stopping tasks creates new task resources or modifies the state of existing task resources. The Amazon ECS container agent calls the `SubmitTaskStateChange` API operation.

The Amazon ECS container agent monitors the state of your tasks and it reports any state changes. State changes might include changes from `PENDING` to `RUNNING` or from `RUNNING` to `STOPPED`.

A container in the task changes state.

The Amazon ECS container agent monitors the state of containers within tasks. For example, if a container that is running within a task stops, this container state change triggers an event.

A task using the Fargate Spot capacity provider receives a termination notice.

When a task is using the `FARGATE_SPOT` capacity provider and is stopped due to a Spot interruption, a task state change event is triggered.

Example Task State Change Event

Task state change events are delivered in the following format. The `detail` section below resembles the [Task](#) object that is returned from a [DescribeTasks](#) API operation in the *Amazon Elastic Container Service API Reference*. If your containers are using an image hosted with Amazon ECR, the `imageDigest` field is returned.

Note

The values for the `createdAt`, `connectivityAt`, `pullStartedAt`, `startedAt`, `pullStoppedAt`, and `updatedAt` fields are UNIX timestamps in the response of a `DescribeTasks` action whereas in the task state change event they are ISO string timestamps.

For more information about CloudWatch Events parameters, see [Events and Event Patterns](#) in the *Amazon EventBridge User Guide*.

```
{
  "version": "0",
  "id": "3317b2af-7005-947d-b652-f55e762e571a",
  "detail-type": "ECS Task State Change",
  "source": "aws.ecs",
  "account": "111122223333",
  "time": "2020-01-23T17:57:58Z",
  "region": "us-west-2",
  "resources": [
```

```

    "arn:aws:ecs:us-west-2:111122223333:task/FargateCluster/
c13b4cb40f1f4fe4a2971f76ae5a47ad"
  ],
  "detail": {
    "attachments": [
      {
        "id": "1789bcae-ddfb-4d10-8ebe-8ac87ddba5b8",
        "type": "eni",
        "status": "ATTACHED",
        "details": [
          {
            "name": "subnetId",
            "value": "subnet-abcd1234"
          },
          {
            "name": "networkInterfaceId",
            "value": "eni-abcd1234"
          },
          {
            "name": "macAddress",
            "value": "0a:98:eb:a7:29:ba"
          },
          {
            "name": "privateIPv4Address",
            "value": "10.0.0.139"
          }
        ]
      }
    ],
    "availabilityZone": "us-west-2c",
    "clusterArn": "arn:aws:ecs:us-west-2:111122223333:cluster/FargateCluster",
    "containers": [
      {
        "containerArn": "arn:aws:ecs:us-west-2:111122223333:container/
cf159fd6-3e3f-4a9e-84f9-66cbe726af01",
        "lastStatus": "RUNNING",
        "name": "FargateApp",
        "image": "111122223333.dkr.ecr.us-west-2.amazonaws.com/hello-
repository:latest",
        "imageDigest":
"sha256:74b2c688c700ec95a93e478cdb959737c148df3fbf5ea706abe0318726e885e6",
        "runtimeId":
"ad64cbc71c7fb31c55507ec24c9f77947132b03d48d9961115cf24f3b7307e1e",
        "taskArn": "arn:aws:ecs:us-west-2:111122223333:task/FargateCluster/
c13b4cb40f1f4fe4a2971f76ae5a47ad",
        "networkInterfaces": [
          {
            "attachmentId": "1789bcae-ddfb-4d10-8ebe-8ac87ddba5b8",
            "privateIpv4Address": "10.0.0.139"
          }
        ],
        "cpu": "0"
      }
    ],
    "createdAt": "2020-01-23T17:57:34.402Z",
    "launchType": "FARGATE",
    "cpu": "256",
    "memory": "512",
    "desiredStatus": "RUNNING",
    "group": "family:sample-fargate",
    "lastStatus": "RUNNING",
    "overrides": {
      "containerOverrides": [
        {
          "name": "FargateApp"
        }
      ]
    }
  }
}

```



```

    ],
    "connectivity": "CONNECTED",
    "connectivityAt": "2020-01-23T17:57:38.453Z",
    "pullStartedAt": "2020-01-23T17:57:52.103Z",
    "startedAt": "2020-01-23T17:57:58.103Z",
    "pullStoppedAt": "2020-01-23T17:57:55.103Z",
    "updatedAt": "2020-01-23T17:57:58.103Z",
    "taskArn": "arn:aws:ecs:us-west-2:111122223333:task/FargateCluster/c13b4cb40f1f4fe4a2971f76ae5a47ad",
    "taskDefinitionArn": "arn:aws:ecs:us-west-2:111122223333:task-definition/sample-fargate:1",
    "version": 4,
    "platformVersion": "1.3.0"
  }
}

```

For a tutorial walkthrough of setting up a simple AWS Lambda function that listens for Amazon ECS task events and writes them out to a CloudWatch Logs log stream, see [Tutorial: Listening for Amazon ECS CloudWatch Events \(p. 335\)](#).

For a tutorial walkthrough of creating an SNS topic to email you when a task state change event occurs, see [Tutorial: Sending Amazon Simple Notification Service Alerts for Task Stopped Events \(p. 337\)](#).

Service Action Events

Amazon ECS sends service action events with the detail type **ECS Service Action**. Unlike the container instance and task state change events, the service action events do not include a version number in the details response field. The following is an event pattern that is used to create an EventBridge rule for Amazon ECS service action events. For more information, see [Creating an EventBridge Rule](#) in the *Amazon EventBridge User Guide*.

```

{
  "source": [
    "aws.ecs"
  ],
  "detail-type": [
    "ECS Service Action"
  ]
}

```

Amazon ECS sends events with INFO, WARN, and ERROR event types. The following are the service action events.

Service Action Events with INFO Event Type

SERVICE_STEADY_STATE

The service is healthy and at the desired number of tasks, thus reaching a steady state.

TASKSET_STEADY_STATE

The task set is healthy and at the desired number of tasks, thus reaching a steady state.

CAPACITY_PROVIDER_STEADY_STATE

A capacity provider associated with a service reaches a steady state.

SERVICE_DESIRED_COUNT_UPDATED

When the service scheduler updates the computed desired count for a service or task set. This event is not sent when the desired count is manually updated by a user.

Service Action Events with **WARN** Event Type

`SERVICE_TASK_START_IMPAIRED`

The service is unable to consistently start tasks successfully.

`SERVICE_DISCOVERY_INSTANCE_UNHEALTHY`

A service using service discovery contains an unhealthy task. The service scheduler detects that a task within a service registry is unhealthy.

Service Action Events with **ERROR** Event Type

`SERVICE_DAEMON_PLACEMENT_CONSTRAINT_VIOLATED`

A task in a service using the `DAEMON` service scheduler strategy no longer meets the placement constraint strategy for the service.

`ECS_OPERATION_THROTTLED`

The service scheduler has been throttled due to the Amazon ECS API throttle limits.

`SERVICE_DISCOVERY_OPERATION_THROTTLED`

The service scheduler has been throttled due to the AWS Cloud Map API throttle limits. This can occur on services configured to use service discovery.

`SERVICE_TASK_PLACEMENT_FAILURE`

The service scheduler is unable to place a task. The cause will be described in the `reason` field.

A common cause for this service event being triggered is because of a lack of resources in the cluster to place the task. For example, not enough CPU or memory capacity on the available container instances or no container instances being available. Another common cause is when the Amazon ECS container agent is disconnected on the container instance, causing the scheduler to be unable to place the task.

`SERVICE_TASK_CONFIGURATION_FAILURE`

The service scheduler is unable to place a task due to a configuration error. The cause will be described in the `reason` field.

A common cause of this service event being triggered is because tags were being applied to the service but the user or role had not opted in to the new Amazon Resource Name (ARN) format in the Region. For more information, see [Amazon Resource Names \(ARNs\) and IDs \(p. 104\)](#). Another common cause is that Amazon ECS was unable to assume the task IAM role provided.

Example Service Steady State Event

Service steady state events are delivered in the following format. For more information about EventBridge parameters, see [Events and Event Patterns](#) in the *Amazon EventBridge User Guide*.

For a tutorial walkthrough of setting up a simple AWS Lambda function that listens for Amazon ECS service action events and writes them out to a CloudWatch Logs log stream, see [Tutorial: Listening for Amazon ECS CloudWatch Events \(p. 335\)](#).

For a tutorial walkthrough of creating an SNS topic to email you when a service event occurs, see [Tutorial: Sending Amazon Simple Notification Service Alerts for Task Stopped Events \(p. 337\)](#).

```
{
  "version": "0",
  "id": "af3c496d-f4a8-65d1-70f4-a69d52e9b584",
  "detail-type": "ECS Service Action",
```

```
{
  "source": "aws.ecs",
  "account": "111122223333",
  "time": "2019-11-19T19:27:22Z",
  "region": "us-west-2",
  "resources": [
    "arn:aws:ecs:us-west-2:111122223333:service/default/servicetest"
  ],
  "detail": {
    "eventType": "INFO",
    "eventName": "SERVICE_STEADY_STATE",
    "clusterArn": "arn:aws:ecs:us-west-2:111122223333:cluster/default",
    "createdAt": "2019-11-19T19:27:22.695Z"
  }
}
```

Example Capacity Provider Steady State Event

Capacity provider steady state events are delivered in the following format.

```
{
  "version": "0",
  "id": "b9baa007-2f33-0eb1-5760-0d02a572d81f",
  "detail-type": "ECS Service Action",
  "source": "aws.ecs",
  "account": "111122223333",
  "time": "2019-11-19T19:37:00Z",
  "region": "us-west-2",
  "resources": [
    "arn:aws:ecs:us-west-2:111122223333:service/default/servicetest"
  ],
  "detail": {
    "eventType": "INFO",
    "eventName": "CAPACITY_PROVIDER_STEADY_STATE",
    "clusterArn": "arn:aws:ecs:us-west-2:111122223333:cluster/default",
    "capacityProviderArns": [
      "arn:aws:ecs:us-west-2:111122223333:capacity-provider/ASG-tutorial-capacity-provider"
    ],
    "createdAt": "2019-11-19T19:37:00.807Z"
  }
}
```

Example Service Task Start Impaired Event

Service task start impaired events are delivered in the following format.

```
{
  "version": "0",
  "id": "57c9506e-9d21-294c-d2fe-e8738da7e67d",
  "detail-type": "ECS Service Action",
  "source": "aws.ecs",
  "account": "111122223333",
  "time": "2019-11-19T19:55:38Z",
  "region": "us-west-2",
  "resources": [
    "arn:aws:ecs:us-west-2:111122223333:service/default/servicetest"
  ],
  "detail": {
    "eventType": "WARN",
    "eventName": "SERVICE_TASK_START_IMPAIRED",
    "clusterArn": "arn:aws:ecs:us-west-2:111122223333:cluster/default",
    "createdAt": "2019-11-19T19:55:38.725Z"
  }
}
```

}

Example Service Task Placement Failure Event

Service task placement failure events are delivered in the following format. For more information about EventBridge parameters, see [Events and Event Patterns](#) in the *Amazon EventBridge User Guide*.

In the following example, the task was attempting to use the `FARGATE_SPOT` capacity provider but the service scheduler was unable to acquire any Fargate Spot capacity.

```
{
  "version": "0",
  "id": "ddca6449-b258-46c0-8653-e0e3a6d0468b",
  "detail-type": "ECS Service Action",
  "source": "aws.ecs",
  "account": "111122223333",
  "time": "2019-11-19T19:55:38Z",
  "region": "us-west-2",
  "resources": [
    "arn:aws:ecs:us-west-2:111122223333:service/default/servicetest"
  ],
  "detail": {
    "eventType": "ERROR",
    "eventName": "SERVICE_TASK_PLACEMENT_FAILURE",
    "clusterArn": "arn:aws:ecs:us-west-2:111122223333:cluster/default",
    "capacityProviderArns": [
      "arn:aws:ecs:us-west-2:111122223333:capacity-provider/FARGATE_SPOT"
    ],
    "reason": "RESOURCE:FARGATE",
    "createdAt": "2019-11-06T19:09:33.087Z"
  }
}
```

Handling Events

Amazon ECS sends events on an *at least once* basis. This means you may receive multiple copies of a given event. Additionally, events may not be delivered to your event listeners in the order in which the events occurred.

To enable proper ordering of events, the `detail` section of each event contains a `version` property. Each time a resource changes state, this `version` is incremented. Duplicate events have the same `version` in the `detail` object. If you are replicating your task state with EventBridge, you can compare the `version` of a resource reported by the Amazon ECS APIs with the `version` reported in EventBridge for the resource to verify that the `version` in your event stream is current. Events with a higher `version` property number should be treated as occurring later than events with lower `version` numbers.

Example: Handling Events in an AWS Lambda Function

The following example shows a Lambda function written in Python 2.7 that captures task state change events and saves them to the following Amazon DynamoDB table:

- `ECSTaskState` – Stores the latest state for a task. The table ID is the `taskArn` value of the task.

```
import json
import boto3

def lambda_handler(event, context):
    id_name = ""
    new_record = {}
```

```
# For debugging so you can see raw event format.
print('Here is the event:')
print(json.dumps(event))

if event["source"] != "aws.ecs":
    raise ValueError("Function only supports input from events with a source type of:
aws.ecs")

# Switch on task/container events.
table_name = ""
if event["detail-type"] == "ECS Task State Change":
    table_name = "ECSTaskState"
    id_name = "taskArn"
    event_id = event["detail"]["taskArn"]
else:
    raise ValueError("detail-type for event is not a supported type. Exiting without
saving event.")

new_record["cw_version"] = event["version"]
new_record.update(event["detail"])

# "status" is a reserved word in DDB, but it appears in containerPort
# state change messages.
if "status" in event:
    new_record["current_status"] = event["status"]
    new_record.pop("status")

# Look first to see if you have received a newer version of an event ID.
# If the version is OLDER than what you have on file, do not process it.
# Otherwise, update the associated record with this latest information.
print("Looking for recent event with same ID...")
dynamodb = boto3.resource("dynamodb", region_name="us-east-1")
table = dynamodb.Table(table_name)
saved_event = table.get_item(
    Key={
        id_name : event_id
    }
)
if "Item" in saved_event:
    # Compare events and reconcile.
    print("EXISTING EVENT DETECTED: Id " + event_id + " - reconciling")
    if saved_event["Item"]["version"] < event["detail"]["version"]:
        print("Received event is a more recent version than the stored event -
updating")
        table.put_item(
            Item=new_record
        )
    else:
        print("Received event is an older version than the stored event - ignoring")
else:
    print("Saving new event - ID " + event_id)

    table.put_item(
        Item=new_record
    )
```

Amazon ECS CloudWatch Container Insights

CloudWatch Container Insights collects, aggregates, and summarizes metrics and logs from your containerized applications and microservices. The metrics include utilization for resources such as CPU,

memory, disk, and network. The metrics are available in CloudWatch automatic dashboards. For a full list of Amazon ECS Container Insights metrics, see [Amazon ECS Container Insights Metrics](#) in the *Amazon CloudWatch User Guide*.

Operational data is collected as performance log events. These are entries that use a structured JSON schema that enables high-cardinality data to be ingested and stored at scale. From this data, CloudWatch creates higher-level aggregated metrics at the cluster and service level as CloudWatch metrics. For more information, see [Container Insights Structured Logs for Amazon ECS](#) in the *Amazon CloudWatch User Guide*.

Important

Metrics collected by CloudWatch Container Insights are charged as custom metrics. For more information about CloudWatch pricing, see [CloudWatch Pricing](#). Amazon ECS also provides monitoring metrics that are provided at no additional cost. For more information, see [Amazon ECS CloudWatch Metrics \(p. 184\)](#).

Container Insights Considerations

The following should be considered when using CloudWatch Container Insights.

- CloudWatch Container Insights metrics only reflect the resources with running tasks during the specified time range. For example, if you have a cluster with one service in it but that service has no tasks in a `RUNNING` state, there will be no metrics sent to CloudWatch. If you have two services and one of them has running tasks and the other doesn't, only the metrics for the service with running tasks will be sent.
- Network metrics are only available for tasks that either use the Fargate launch type or use the bridge network mode.

Working with Container Insights-enabled clusters

Container Insights can be enabled for all new clusters created by opting in to the `containerInsights` account setting, on individual clusters by enabling it using the cluster settings during cluster creation, or on existing clusters by using the `UpdateClusterSettings` API.

Opting in to the `containerInsights` account setting can be done with both the Amazon ECS console and the AWS CLI. You must be running version 1.16.200 or later of the AWS CLI to use this feature. For more information on creating Amazon ECS clusters, see [Creating a Cluster \(p. 18\)](#).

Important

To opt in all IAM users or roles on your account to Container Insights-enabled clusters using the console

1. As the root user of the account, open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation bar at the top of the screen, select the Region for which to opt in to Container Insights-enabled clusters.
3. From the dashboard, choose **Account Settings**.
4. For **IAM user or role**, ensure your root user or container instance IAM role is selected.
5. For **Container Insights**, select the check box. Choose **Save** once finished.

Important

IAM users and IAM roles need the `ecs:PutAccountSetting` permission to perform this action.

6. On the confirmation screen, choose **Confirm** to save the selection.

To opt in all IAM users or roles on your account to Container Insights-enabled clusters using the command line

Any user on an account can use one of the following commands to modify the default account setting for all IAM users or roles on your account. These changes apply to the entire AWS account unless an IAM user or role explicitly overrides these settings for themselves.

- [put-account-setting-default](#) (AWS CLI)

```
aws ecs put-account-setting-default --name containerInsights --value enabled --region us-east-1
```

- [Write-ECSAccountSettingDefault](#) (AWS Tools for Windows PowerShell)

```
Write-ECSAccountSettingDefault -Name containerInsights -Value enabled -Region us-east-1 -Force
```

To opt in an IAM user or container instance IAM role to Container Insights-enabled clusters as the root user using the command line

The root user on an account can use one of the following commands and specify the ARN of the principal IAM user or container instance IAM role in the request to modify the account settings.

- [put-account-setting](#) (AWS CLI)

The following example is for modifying the account setting of a specific IAM user:

```
aws ecs put-account-setting --name containerInsights --value enabled --principal-arn arn:aws:iam::aws_account_id:user/userName --region us-east-1
```

- [Write-ECSAccountSetting](#) (AWS Tools for Windows PowerShell)

The following example is for modifying the account setting of a specific IAM user:

```
Write-ECSAccountSetting -Name containerInsights -Value enabled -PrincipalArn arn:aws:iam::aws_account_id:user/userName -Region us-east-1 -Force
```

To update the settings for an existing cluster using the command line

Use one of the following commands to update the setting for a cluster.

- [update-cluster-settings](#) (AWS CLI)

```
aws ecs update-cluster-settings --cluster cluster_name_or_arn --settings name=containerInsights,value=enabled/disabled --region us-east-1
```

Logging Amazon ECS API Calls with AWS CloudTrail

Amazon ECS is integrated with AWS CloudTrail, a service that provides a record of actions taken by a user, role, or an AWS service in Amazon ECS. CloudTrail captures all API calls for Amazon ECS as events, including calls from the Amazon ECS console and from code calls to the Amazon ECS API operations.

If you create a trail, you can enable continuous delivery of CloudTrail events to an Amazon S3 bucket, including events for Amazon ECS. If you don't configure a trail, you can still view the most recent events in the CloudTrail console in **Event history**. Using the information collected by CloudTrail, you can determine the request that was made to Amazon ECS, the IP address from which the request was made, who made the request, when it was made, and additional details.

For more information, see the [AWS CloudTrail User Guide](#).

Amazon ECS Information in CloudTrail

CloudTrail is enabled on your AWS account when you create the account. When activity occurs in Amazon ECS, that activity is recorded in a CloudTrail event along with other AWS service events in **Event history**. You can view, search, and download recent events in your AWS account. For more information, see [Viewing Events with CloudTrail Event History](#).

For an ongoing record of events in your AWS account, including events for Amazon ECS, create a trail. A trail enables CloudTrail to deliver log files to an Amazon S3 bucket. By default, when you create a trail in the console, the trail applies to all regions. The trail logs events from all Regions in the AWS partition and delivers the log files to the Amazon S3 bucket that you specify. Additionally, you can configure other AWS services to further analyze and act upon the event data collected in CloudTrail logs. For more information, see:

- [Overview for Creating a Trail](#)
- [CloudTrail Supported Services and Integrations](#)
- [Configuring Amazon SNS Notifications for CloudTrail](#)
- [Receiving CloudTrail Log Files from Multiple Regions](#) and [Receiving CloudTrail Log Files from Multiple Accounts](#)

All Amazon ECS actions are logged by CloudTrail and are documented in the [Amazon Elastic Container Service API Reference](#). For example, calls to the `CreateService`, `RunTask` and `DeleteCluster` sections generate entries in the CloudTrail log files.

Every event or log entry contains information about who generated the request. The identity information helps you determine the following:

- Whether the request was made with root or IAM user credentials.
- Whether the request was made with temporary security credentials for a role or federated user.
- Whether the request was made by another AWS service.

For more information, see the [CloudTrail `userIdentity` Element](#).

Understanding Amazon ECS Log File Entries

A trail is a configuration that enables delivery of events as log files to an Amazon S3 bucket that you specify. CloudTrail log files contain one or more log entries. An event represents a single request from any source and includes information about the requested action, the date and time of the action, request parameters, and so on. CloudTrail log files are not an ordered stack trace of the public API calls, so they do not appear in any specific order.

Note

These examples have been formatted for improved readability. In a CloudTrail log file, all entries and events are concatenated into a single line. In addition, this example has been limited to a single Amazon ECS entry. In a real CloudTrail log file, you see entries and events from multiple AWS services.

The following example shows a CloudTrail log entry that demonstrates the `CreateCluster` action:

```
{
  "eventVersion": "1.04",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "AIDACKCEVSQ6C2EXAMPLE:account_name",
    "arn": "arn:aws:sts::123456789012:user/Mary_Major",
    "accountId": "123456789012",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
    "sessionContext": {
      "attributes": {
        "mfaAuthenticated": "false",
        "creationDate": "2018-06-20T18:32:25Z"
      },
      "sessionIssuer": {
        "type": "Role",
        "principalId": "AIDACKCEVSQ6C2EXAMPLE",
        "arn": "arn:aws:iam::123456789012:role/Admin",
        "accountId": "123456789012",
        "userName": "Mary_Major"
      }
    }
  },
  "eventTime": "2018-06-20T19:04:36Z",
  "eventSource": "ecs.amazonaws.com",
  "eventName": "CreateCluster",
  "awsRegion": "us-east-1",
  "sourceIPAddress": "203.0.113.12",
  "userAgent": "console.amazonaws.com",
  "requestParameters": {
    "clusterName": "default"
  },
  "responseElements": {
    "cluster": {
      "clusterArn": "arn:aws:ecs:us-east-1:123456789012:cluster/default",
      "pendingTasksCount": 0,
      "registeredContainerInstancesCount": 0,
      "status": "ACTIVE",
      "runningTasksCount": 0,
      "statistics": [],
      "clusterName": "default",
      "activeServicesCount": 0
    }
  },
  "requestID": "cb8c167e-EXAMPLE",
  "eventID": "e3c6f4ce-EXAMPLE",
  "eventType": "AwsApiCall",
  "recipientAccountId": "123456789012"
}
```

Security in Amazon Elastic Container Service

Cloud security at AWS is the highest priority. As an AWS customer, you benefit from a data center and network architecture that is built to meet the requirements of the most security-sensitive organizations.

Security is a shared responsibility between AWS and you. The [shared responsibility model](#) describes this as security *of* the cloud and security *in* the cloud:

- **Security of the cloud** – AWS is responsible for protecting the infrastructure that runs AWS services in the AWS Cloud. AWS also provides you with services that you can use securely. Third-party auditors regularly test and verify the effectiveness of our security as part of the [AWS compliance programs](#). To learn about the compliance programs that apply to Amazon Elastic Container Service, see [AWS Services in Scope by Compliance Program](#).
- **Security in the cloud** – Your responsibility is determined by the AWS service that you use. You are also responsible for other factors including the sensitivity of your data, your company's requirements, and applicable laws and regulations.

This documentation helps you understand how to apply the shared responsibility model when using Amazon ECS. The following topics show you how to configure Amazon ECS to meet your security and compliance objectives. You also learn how to use other AWS services that help you to monitor and secure your Amazon ECS resources.

Topics

- [Identity and Access Management for Amazon Elastic Container Service \(p. 202\)](#)
- [Logging and Monitoring in Amazon Elastic Container Service \(p. 251\)](#)
- [Compliance Validation for Amazon Elastic Container Service \(p. 252\)](#)
- [Infrastructure Security in Amazon Elastic Container Service \(p. 252\)](#)

Identity and Access Management for Amazon Elastic Container Service

AWS Identity and Access Management (IAM) is an AWS service that helps an administrator securely control access to AWS resources. IAM administrators control who can be *authenticated* (signed in) and *authorized* (have permissions) to use Amazon ECS resources. IAM is an AWS service that you can use with no additional charge.

Topics

- [Audience \(p. 203\)](#)
- [Authenticating With Identities \(p. 203\)](#)
- [Managing Access Using Policies \(p. 205\)](#)
- [How Amazon Elastic Container Service Works with IAM \(p. 206\)](#)
- [Amazon Elastic Container Service Identity-Based Policy Examples \(p. 210\)](#)
- [Supported Resource-Level Permissions for Amazon ECS API Actions \(p. 219\)](#)
- [Managed Policies and Trust Relationships \(p. 220\)](#)
- [Service-Linked Role for Amazon ECS \(p. 227\)](#)

- [Amazon ECS Task Execution IAM Role \(p. 236\)](#)
- [IAM Roles for Tasks \(p. 240\)](#)
- [Amazon ECS CodeDeploy IAM Role \(p. 243\)](#)
- [Amazon ECS CloudWatch Events IAM Role \(p. 246\)](#)
- [Troubleshooting Amazon Elastic Container Service Identity and Access \(p. 249\)](#)

Audience

How you use AWS Identity and Access Management (IAM) differs, depending on the work you do in Amazon ECS.

Service user – If you use the Amazon ECS service to do your job, then your administrator provides you with the credentials and permissions that you need. As you use more Amazon ECS features to do your work, you might need additional permissions. Understanding how access is managed can help you request the right permissions from your administrator. If you cannot access a feature in Amazon ECS, see [Troubleshooting Amazon Elastic Container Service Identity and Access \(p. 249\)](#).

Service administrator – If you're in charge of Amazon ECS resources at your company, you probably have full access to Amazon ECS. It's your job to determine which Amazon ECS features and resources your employees should access. You must then submit requests to your IAM administrator to change the permissions of your service users. Review the information on this page to understand the basic concepts of IAM. To learn more about how your company can use IAM with Amazon ECS, see [How Amazon Elastic Container Service Works with IAM \(p. 206\)](#).

IAM administrator – If you're an IAM administrator, you might want to learn details about how you can write policies to manage access to Amazon ECS. To view example Amazon ECS identity-based policies that you can use in IAM, see [Amazon Elastic Container Service Identity-Based Policy Examples \(p. 210\)](#).

Authenticating With Identities

Authentication is how you sign in to AWS using your identity credentials. For more information about signing in using the AWS Management Console, see [The IAM Console and Sign-in Page](#) in the *IAM User Guide*.

You must be *authenticated* (signed in to AWS) as the AWS account root user, an IAM user, or by assuming an IAM role. You can also use your company's single sign-on authentication, or even sign in using Google or Facebook. In these cases, your administrator previously set up identity federation using IAM roles. When you access AWS using credentials from another company, you are assuming a role indirectly.

To sign in directly to the [AWS Management Console](#), use your password with your root user email or your IAM user name. You can access AWS programmatically using your root user or IAM user access keys. AWS provides SDK and command line tools to cryptographically sign your request using your credentials. If you don't use AWS tools, you must sign the request yourself. Do this using *Signature Version 4*, a protocol for authenticating inbound API requests. For more information about authenticating requests, see [Signature Version 4 Signing Process](#) in the *AWS General Reference*.

Regardless of the authentication method that you use, you might also be required to provide additional security information. For example, AWS recommends that you use multi-factor authentication (MFA) to increase the security of your account. To learn more, see [Using Multi-Factor Authentication \(MFA\) in AWS](#) in the *IAM User Guide*.

AWS Account Root User

When you first create an AWS account, you begin with a single sign-in identity that has complete access to all AWS services and resources in the account. This identity is called the AWS account *root user* and

is accessed by signing in with the email address and password that you used to create the account. We strongly recommend that you do not use the root user for your everyday tasks, even the administrative ones. Instead, adhere to the [best practice of using the root user only to create your first IAM user](#). Then securely lock away the root user credentials and use them to perform only a few account and service management tasks.

IAM Users and Groups

An *IAM user* is an identity within your AWS account that has specific permissions for a single person or application. An IAM user can have long-term credentials such as a user name and password or a set of access keys. To learn how to generate access keys, see [Managing Access Keys for IAM Users](#) in the *IAM User Guide*. When you generate access keys for an IAM user, make sure you view and securely save the key pair. You cannot recover the secret access key in the future. Instead, you must generate a new access key pair.

An *IAM group* is an identity that specifies a collection of IAM users. You can't sign in as a group. You can use groups to specify permissions for multiple users at a time. Groups make permissions easier to manage for large sets of users. For example, you could have a group named *IAMAdmins* and give that group permissions to administer IAM resources.

Users are different from roles. A user is uniquely associated with one person or application, but a role is intended to be assumable by anyone who needs it. Users have permanent long-term credentials, but roles provide temporary credentials. To learn more, see [When to Create an IAM User \(Instead of a Role\)](#) in the *IAM User Guide*.

IAM Roles

An *IAM role* is an identity within your AWS account that has specific permissions. It is similar to an IAM user, but is not associated with a specific person. You can temporarily assume an IAM role in the AWS Management Console by [switching roles](#). You can assume a role by calling an AWS CLI or AWS API operation or by using a custom URL. For more information about methods for using roles, see [Using IAM Roles](#) in the *IAM User Guide*.

IAM roles with temporary credentials are useful in the following situations:

- **Temporary IAM user permissions** – An IAM user can assume an IAM role to temporarily take on different permissions for a specific task.
- **Federated user access** – Instead of creating an IAM user, you can use existing identities from AWS Directory Service, your enterprise user directory, or a web identity provider. These are known as *federated users*. AWS assigns a role to a federated user when access is requested through an [identity provider](#). For more information about federated users, see [Federated Users and Roles](#) in the *IAM User Guide*.
- **Cross-account access** – You can use an IAM role to allow someone (a trusted principal) in a different account to access resources in your account. Roles are the primary way to grant cross-account access. However, with some AWS services, you can attach a policy directly to a resource (instead of using a role as a proxy). To learn the difference between roles and resource-based policies for cross-account access, see [How IAM Roles Differ from Resource-based Policies](#) in the *IAM User Guide*.
- **AWS service access** – A service role is an IAM role that a service assumes to perform actions in your account on your behalf. When you set up some AWS service environments, you must define a role for the service to assume. This service role must include all the permissions that are required for the service to access the AWS resources that it needs. Service roles vary from service to service, but many allow you to choose your permissions as long as you meet the documented requirements for that service. Service roles provide access only within your account and cannot be used to grant access to services in other accounts. You can create, modify, and delete a service role from within IAM. For example, you can create a role that allows Amazon Redshift to access an Amazon S3 bucket on your behalf and then load data from that bucket into an Amazon Redshift cluster. For more information, see [Creating a Role to Delegate Permissions to an AWS Service](#) in the *IAM User Guide*.

- **Applications running on Amazon EC2** – You can use an IAM role to manage temporary credentials for applications that are running on an EC2 instance and making AWS CLI or AWS API requests. This is preferable to storing access keys within the EC2 instance. To assign an AWS role to an EC2 instance and make it available to all of its applications, you create an instance profile that is attached to the instance. An instance profile contains the role and enables programs that are running on the EC2 instance to get temporary credentials. For more information, see [Using an IAM Role to Grant Permissions to Applications Running on Amazon EC2 Instances](#) in the *IAM User Guide*.

To learn whether to use IAM roles, see [When to Create an IAM Role \(Instead of a User\)](#) in the *IAM User Guide*.

Managing Access Using Policies

You control access in AWS by creating policies and attaching them to IAM identities or AWS resources. A policy is an object in AWS that, when associated with an identity or resource, defines their permissions. AWS evaluates these policies when an entity (root user, IAM user, or IAM role) makes a request. Permissions in the policies determine whether the request is allowed or denied. Most policies are stored in AWS as JSON documents. For more information about the structure and contents of JSON policy documents, see [Overview of JSON Policies](#) in the *IAM User Guide*.

An IAM administrator can use policies to specify who has access to AWS resources, and what actions they can perform on those resources. Every IAM entity (user or role) starts with no permissions. In other words, by default, users can do nothing, not even change their own password. To give a user permission to do something, an administrator must attach a permissions policy to a user. Or the administrator can add the user to a group that has the intended permissions. When an administrator gives permissions to a group, all users in that group are granted those permissions.

IAM policies define permissions for an action regardless of the method that you use to perform the operation. For example, suppose that you have a policy that allows the `iam:GetRole` action. A user with that policy can get role information from the AWS Management Console, the AWS CLI, or the AWS API.

Identity-Based Policies

Identity-based policies are JSON permissions policy documents that you can attach to an identity, such as an IAM user, role, or group. These policies control what actions that identity can perform, on which resources, and under what conditions. To learn how to create an identity-based policy, see [Creating IAM Policies](#) in the *IAM User Guide*.

Identity-based policies can be further categorized as *inline policies* or *managed policies*. Inline policies are embedded directly into a single user, group, or role. Managed policies are standalone policies that you can attach to multiple users, groups, and roles in your AWS account. Managed policies include AWS managed policies and customer managed policies. To learn how to choose between a managed policy or an inline policy, see [Choosing Between Managed Policies and Inline Policies](#) in the *IAM User Guide*.

Resource-Based Policies

Resource-based policies are JSON policy documents that you attach to a resource such as an Amazon S3 bucket. Service administrators can use these policies to define what actions a specified principal (account member, user, or role) can perform on that resource and under what conditions. Resource-based policies are inline policies. There are no managed resource-based policies.

Access Control Lists (ACLs)

Access control lists (ACLs) are a type of policy that controls which principals (account members, users, or roles) have permissions to access a resource. ACLs are similar to resource-based policies, although they

do not use the JSON policy document format. Amazon S3, AWS WAF, and Amazon VPC are examples of services that support ACLs. To learn more about ACLs, see [Access Control List \(ACL\) Overview](#) in the *Amazon Simple Storage Service Developer Guide*.

Other Policy Types

AWS supports additional, less-common policy types. These policy types can set the maximum permissions granted to you by the more common policy types.

- **Permissions boundaries** – A permissions boundary is an advanced feature in which you set the maximum permissions that an identity-based policy can grant to an IAM entity (IAM user or role). You can set a permissions boundary for an entity. The resulting permissions are the intersection of entity's identity-based policies and its permissions boundaries. Resource-based policies that specify the user or role in the `Principal` field are not limited by the permissions boundary. An explicit deny in any of these policies overrides the allow. For more information about permissions boundaries, see [Permissions Boundaries for IAM Entities](#) in the *IAM User Guide*.
- **Service control policies (SCPs)** – SCPs are JSON policies that specify the maximum permissions for an organization or organizational unit (OU) in AWS Organizations. AWS Organizations is a service for grouping and centrally managing multiple AWS accounts that your business owns. If you enable all features in an organization, then you can apply service control policies (SCPs) to any or all of your accounts. The SCP limits permissions for entities in member accounts, including each AWS account root user. For more information about Organizations and SCPs, see [How SCPs Work](#) in the *AWS Organizations User Guide*.
- **Session policies** – Session policies are advanced policies that you pass as a parameter when you programmatically create a temporary session for a role or federated user. The resulting session's permissions are the intersection of the user or role's identity-based policies and the session policies. Permissions can also come from a resource-based policy. An explicit deny in any of these policies overrides the allow. For more information, see [Session Policies](#) in the *IAM User Guide*.

Multiple Policy Types

When multiple types of policies apply to a request, the resulting permissions are more complicated to understand. To learn how AWS determines whether to allow a request when multiple policy types are involved, see [Policy Evaluation Logic](#) in the *IAM User Guide*.

How Amazon Elastic Container Service Works with IAM

Before you use IAM to manage access to Amazon ECS, you should understand what IAM features are available to use with Amazon ECS. To get a high-level view of how Amazon ECS and other AWS services work with IAM, see [AWS Services That Work with IAM](#) in the *IAM User Guide*.

Topics

- [Amazon ECS Identity-Based Policies \(p. 206\)](#)
- [Amazon ECS Resource-Based Policies \(p. 210\)](#)
- [Authorization Based on Amazon ECS Tags \(p. 210\)](#)
- [Amazon ECS IAM Roles \(p. 210\)](#)

Amazon ECS Identity-Based Policies

With IAM identity-based policies, you can specify allowed or denied actions and resources as well as the conditions under which actions are allowed or denied. Amazon ECS supports specific actions, resources,

and condition keys. To learn about all of the elements that you use in a JSON policy, see [IAM JSON Policy Elements Reference](#) in the *IAM User Guide*.

Actions

The **Action** element of an IAM identity-based policy describes the specific action or actions that will be allowed or denied by the policy. Policy actions usually have the same name as the associated AWS API operation. The action is used in a policy to grant permissions to perform the associated operation.

Policy actions in Amazon ECS use the following prefix before the action: `ecs:`. For example, to grant someone permission to create an Amazon ECS cluster with the Amazon ECS `CreateCluster` API operation, you include the `ecs:CreateCluster` action in their policy. Policy statements must include either an **Action** or **NotAction** element. Amazon ECS defines its own set of actions that describe tasks that you can perform with this service.

To specify multiple actions in a single statement, separate them with commas as follows:

```
"Action": [
    "ecs:action1",
    "ecs:action2"
```

You can specify multiple actions using wildcards (*). For example, to specify all actions that begin with the word `Describe`, include the following action:

```
"Action": "ecs:Describe*"
```

To see a list of Amazon ECS actions, see [Actions, Resources, and Condition Keys for Amazon Elastic Container Service](#) in the *IAM User Guide*

Resources

The **Resource** element specifies the object or objects to which the action applies. Statements must include either a **Resource** or a **NotResource** element. You specify a resource using an ARN or using the wildcard (*) to indicate that the statement applies to all resources.

The Amazon ECS cluster resource has the following ARN:

```
arn:${Partition}:ecs:${Region}:${Account}:cluster/${clusterName}
```

For more information about the format of ARNs, see [Amazon Resource Names \(ARNs\) and AWS Service Namespaces](#).

For example, to specify the `my-cluster` cluster in your statement, use the following ARN:

```
"Resource": "arn:aws:ecs:us-east-1:123456789012:cluster/my-cluster"
```

To specify all clusters that belong to a specific account, use the wildcard (*):

```
"Resource": "arn:aws:ecs:us-east-1:123456789012:cluster/*"
```

Some Amazon ECS actions, such as those for creating resources, cannot be performed on a specific resource. In those cases, you must use the wildcard (*).

```
"Resource": "*" 
```

Some Amazon ECS API actions can be performed on multiple resources. For example, multiple clusters can be referenced when calling the `DescribeClusters` API action. To specify multiple resources in a single statement, separate the ARNs with commas.

```
"Resource": [  
    "resource1",  
    "resource2"
```

The following table describes the ARNs for each resource type used by the Amazon ECS API actions.

Important

The following table uses the new longer ARN format for Amazon ECS tasks, services, and container instances. If you have not opted in to the long ARN format, the ARNs will not include the cluster name. For more information, see [Amazon Resource Names \(ARNs\) and IDs \(p. 104\)](#).

Resource Type	ARN
All Amazon ECS resources	<code>arn:aws:ecs:*</code>
All Amazon ECS resources owned by the specified account in the specified region	<code>arn:aws:ecs:region:account:*</code>
Cluster	<code>arn:aws:ecs:region:account:cluster/cluster-name</code>
Container instance	<code>arn:aws:ecs:region:account:container-instance/cluster-name/container-instance-id</code>
Task definition	<code>arn:aws:ecs:region:account:task-definition/task-definition-family-name:task-definition-revision-number</code>
Service	<code>arn:aws:ecs:region:account:service/cluster-name/service-name</code>
Task	<code>arn:aws:ecs:region:account:task/cluster-name/task-id</code>
Container	<code>arn:aws:ecs:region:account:container/container-id</code>

To learn with which actions you can specify the ARN of each resource, see [Supported Resource-Level Permissions for Amazon ECS API Actions \(p. 219\)](#).

Condition Keys

The `Condition` element (or *Condition block*) lets you specify conditions in which a statement is in effect. The `Condition` element is optional. You can build conditional expressions that use [condition operators](#), such as equals or less than, to match the condition in the policy with values in the request.

If you specify multiple `Condition` elements in a statement, or multiple keys in a single `Condition` element, AWS evaluates them using a logical `AND` operation. If you specify multiple values for a single condition key, AWS evaluates the condition using a logical `OR` operation. All of the conditions must be met before the statement's permissions are granted.

You can also use placeholder variables when you specify conditions. For example, you can grant an IAM user permission to access a resource only if it is tagged with their IAM user name. For more information, see [IAM Policy Elements: Variables and Tags](#) in the *IAM User Guide*.

Amazon ECS defines its own set of condition keys and also supports using some global condition keys. To see all AWS global condition keys, see [AWS Global Condition Context Keys](#) in the *IAM User Guide*.

Amazon ECS implements the following service-specific condition keys.

Condition Key	Description	Evaluation Types
aws:RequestTag/ \${TagKey}	<p>The context key is formatted "aws:RequestTag/<i>tag-key</i>": "<i>tag-value</i>" where <i>tag-key</i> and <i>tag-value</i> are a tag key and value pair.</p> <p>Checks that the tag key-value pair is present in an AWS request. For example, you could check to see that the request includes the tag key "Dept" and that it has the value "Accounting".</p>	String
aws:ResourceTag/ \${TagKey}	<p>The context key is formatted "aws:ResourceTag/<i>tag-key</i>": "<i>tag-value</i>" where <i>tag-key</i> and <i>tag-value</i> are a tag key and value pair.</p> <p>Checks that the tag attached to the identity resource (user or role) matches the specified key name and value.</p>	String
aws:TagKeys	<p>This context key is formatted "aws:TagKeys": "<i>tag-key</i>" where <i>tag-key</i> is a list of tag keys without values (for example, ["Dept", "Cost-Center"]).</p> <p>Checks the tag keys that are present in an AWS request.</p>	String
ecs:ResourceTag/ \${TagKey}	<p>The context key is formatted "ecs:ResourceTag/<i>tag-key</i>": "<i>tag-value</i>" where <i>tag-key</i> and <i>tag-value</i> are a tag key and value pair.</p> <p>Checks that the tag attached to the identity resource (user or role) matches the specified key name and value.</p>	String
ecs:cluster	<p>The context key is formatted "ecs:cluster": "<i>cluster-arn</i>" where <i>cluster-arn</i> is the ARN for the Amazon ECS cluster.</p>	ARN, Null
ecs:container- instances	<p>The context key is formatted "ecs:container-instances": "<i>container-instance-arns</i>" where <i>container-instance-arns</i> is one or more container instance ARNs.</p>	ARN, Null
ecs:task-definition	<p>The context key is formatted "ecs:task-definition": "<i>task-definition-arn</i>" where <i>task-definition-arn</i> is the ARN for the Amazon ECS task definition.</p>	ARN, Null
ecs:service	<p>The context key is formatted "ecs:service": "<i>service-arn</i>" where <i>service-arn</i> is the ARN for the Amazon ECS service.</p>	ARN, Null

To learn with which actions and resources you can use a condition key, see [Supported Resource-Level Permissions for Amazon ECS API Actions](#) (p. 219).

Examples

To view examples of Amazon ECS identity-based policies, see [Amazon Elastic Container Service Identity-Based Policy Examples](#) (p. 210).

Amazon ECS Resource-Based Policies

Amazon ECS does not support resource-based policies.

Authorization Based on Amazon ECS Tags

You can attach tags to Amazon ECS resources or pass tags in a request to Amazon ECS. To control access based on tags, you provide tag information in the [condition element](#) of a policy using the `aws:RequestTag/key-name` or `aws:TagKeys` condition keys. For more information, see [Controlling Access Using Tags](#) in the *IAM User Guide*.

For more information about tagging Amazon ECS resources, see [Resources and Tags](#) (p. 177).

To view an example identity-based policy for limiting access to a resource based on the tags on that resource, see [Describing Amazon ECS Services Based on Tags](#) (p. 219).

Amazon ECS IAM Roles

An [IAM role](#) is an entity within your AWS account that has specific permissions.

Using Temporary Credentials with Amazon ECS

You can use temporary credentials to sign in with federation, assume an IAM role, or to assume a cross-account role. You obtain temporary security credentials by calling AWS STS API operations such as [AssumeRole](#) or [GetFederationToken](#).

Amazon ECS supports using temporary credentials.

Service-Linked Roles

[Service-linked roles](#) allow AWS services to access resources in other services to complete an action on your behalf. Service-linked roles appear in your IAM account and are owned by the service. An IAM administrator can view but not edit the permissions for service-linked roles.

Amazon ECS supports service-linked roles. For details about creating or managing Amazon ECS service-linked roles, see [Service-Linked Role for Amazon ECS](#) (p. 227).

Service Roles

This feature allows a service to assume a [service role](#) on your behalf. This role allows the service to access resources in other services to complete an action on your behalf. Service roles appear in your IAM account and are owned by the account. This means that an IAM administrator can change the permissions for this role. However, doing so might break the functionality of the service.

Amazon ECS supports service roles.

Amazon Elastic Container Service Identity-Based Policy Examples

By default, IAM users and roles don't have permission to create or modify Amazon ECS resources. They also can't perform tasks using the AWS Management Console, AWS CLI, or AWS API. An IAM administrator must create IAM policies that grant users and roles permission to perform specific API operations on the specified resources they need. The administrator must then attach those policies to the IAM users or groups that require those permissions.

To learn how to create an IAM identity-based policy using these example JSON policy documents, see [Creating Policies on the JSON Tab](#) in the *IAM User Guide*.

Topics

- [Policy Best Practices](#) (p. 211)
- [Allow Users to View Their Own Permissions](#) (p. 211)
- [Amazon ECS First Run Wizard Permissions](#) (p. 212)
- [Cluster Examples](#) (p. 215)
- [List and Describe Task Examples](#) (p. 217)
- [Create Service Example](#) (p. 218)
- [Update Service Example](#) (p. 218)
- [Describing Amazon ECS Services Based on Tags](#) (p. 219)

Policy Best Practices

Identity-based policies are very powerful. They determine whether someone can create, access, or delete Amazon ECS resources in your account. These actions can incur costs for your AWS account. When you create or edit identity-based policies, follow these guidelines and recommendations:

- **Get Started Using AWS Managed Policies** – To start using Amazon ECS quickly, use AWS managed policies to give your employees the permissions they need. These policies are already available in your account and are maintained and updated by AWS. For more information, see [Get Started Using Permissions With AWS Managed Policies](#) in the *IAM User Guide*.
- **Grant Least Privilege** – When you create custom policies, grant only the permissions required to perform a task. Start with a minimum set of permissions and grant additional permissions as necessary. Doing so is more secure than starting with permissions that are too lenient and then trying to tighten them later. For more information, see [Grant Least Privilege](#) in the *IAM User Guide*.
- **Enable MFA for Sensitive Operations** – For extra security, require IAM users to use multi-factor authentication (MFA) to access sensitive resources or API operations. For more information, see [Using Multi-Factor Authentication \(MFA\) in AWS](#) in the *IAM User Guide*.
- **Use Policy Conditions for Extra Security** – To the extent that it's practical, define the conditions under which your identity-based policies allow access to a resource. For example, you can write conditions to specify a range of allowable IP addresses that a request must come from. You can also write conditions to allow requests only within a specified date or time range, or to require the use of SSL or MFA. For more information, see [IAM JSON Policy Elements: Condition](#) in the *IAM User Guide*.

Allow Users to View Their Own Permissions

This example shows how you might create a policy that allows IAM users to view the inline and managed policies that are attached to their user identity. This policy includes permissions to complete this action on the console or programmatically using the AWS CLI or AWS API.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "ViewOwnUserInfo",
      "Effect": "Allow",
      "Action": [
        "iam:GetUserPolicy",
        "iam:ListGroupsWithUser",
        "iam:ListAttachedUserPolicies",
        "iam:ListUserPolicies",

```

```

        "iam:GetUser"
      ],
      "Resource": [ "arn:aws:iam::*:user/${aws:username}" ]
    },
    {
      "Sid": "NavigateInConsole",
      "Effect": "Allow",
      "Action": [
        "iam:GetGroupPolicy",
        "iam:GetPolicyVersion",
        "iam:GetPolicy",
        "iam:ListAttachedGroupPolicies",
        "iam:ListGroupPolicies",
        "iam:ListPolicyVersions",
        "iam:ListPolicies",
        "iam:ListUsers"
      ],
      "Resource": "*"
    }
  ]
}

```

Amazon ECS First Run Wizard Permissions

The Amazon ECS first-run wizard simplifies the process of creating a cluster and running your tasks and services. However, users require permissions to many API operations from multiple AWS services to complete the wizard. The [AmazonECS_FullAccess \(p. 221\)](#) managed policy below shows the required permissions to complete the Amazon ECS first-run wizard.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "application-autoscaling:DeleteScalingPolicy",
        "application-autoscaling:DeregisterScalableTarget",
        "application-autoscaling:DescribeScalableTargets",
        "application-autoscaling:DescribeScalingActivities",
        "application-autoscaling:DescribeScalingPolicies",
        "application-autoscaling:PutScalingPolicy",
        "application-autoscaling:RegisterScalableTarget",
        "appmesh:ListMeshes",
        "appmesh:ListVirtualNodes",
        "appmesh:DescribeVirtualNode",
        "autoscaling:UpdateAutoScalingGroup",
        "autoscaling:CreateAutoScalingGroup",
        "autoscaling:CreateLaunchConfiguration",
        "autoscaling:DeleteAutoScalingGroup",
        "autoscaling:DeleteLaunchConfiguration",
        "autoscaling:Describe*",
        "cloudformation:CreateStack",
        "cloudformation:DeleteStack",
        "cloudformation:DescribeStack*",
        "cloudformation:UpdateStack",
        "cloudwatch:DescribeAlarms",
        "cloudwatch:DeleteAlarms",
        "cloudwatch:GetMetricStatistics",
        "cloudwatch:PutMetricAlarm",
        "codedeploy:CreateApplication",
        "codedeploy:CreateDeployment",
        "codedeploy:CreateDeploymentGroup",
        "codedeploy:GetApplication",

```

```
"codedeploy:GetDeployment",
"codedeploy:GetDeploymentGroup",
"codedeploy:ListApplications",
"codedeploy:ListDeploymentGroups",
"codedeploy:ListDeployments",
"codedeploy:StopDeployment",
"codedeploy:GetDeploymentTarget",
"codedeploy:ListDeploymentTargets",
"codedeploy:GetDeploymentConfig",
"codedeploy:GetApplicationRevision",
"codedeploy:RegisterApplicationRevision",
"codedeploy:BatchGetApplicationRevisions",
"codedeploy:BatchGetDeploymentGroups",
"codedeploy:BatchGetDeployments",
"codedeploy:BatchGetApplications",
"codedeploy:ListApplicationRevisions",
"codedeploy:ListDeploymentConfigs",
"codedeploy:ContinueDeployment",
"sns:ListTopics",
"lambda:ListFunctions",
"ec2:AssociateRouteTable",
"ec2:AttachInternetGateway",
"ec2:AuthorizeSecurityGroupIngress",
"ec2:CancelSpotFleetRequests",
"ec2>CreateInternetGateway",
"ec2>CreateLaunchTemplate",
"ec2>CreateRoute",
"ec2>CreateRouteTable",
"ec2>CreateSecurityGroup",
"ec2>CreateSubnet",
"ec2>CreateVpc",
"ec2>DeleteLaunchTemplate",
"ec2>DeleteSubnet",
"ec2>DeleteVpc",
"ec2:Describe*",
"ec2:DetachInternetGateway",
"ec2:DisassociateRouteTable",
"ec2:ModifySubnetAttribute",
"ec2:ModifyVpcAttribute",
"ec2:RunInstances",
"ec2:RequestSpotFleet",
"elasticloadbalancing:CreateListener",
"elasticloadbalancing:CreateLoadBalancer",
"elasticloadbalancing:CreateRule",
"elasticloadbalancing:CreateTargetGroup",
"elasticloadbalancing>DeleteListener",
"elasticloadbalancing>DeleteLoadBalancer",
"elasticloadbalancing>DeleteRule",
"elasticloadbalancing>DeleteTargetGroup",
"elasticloadbalancing:DescribeListeners",
"elasticloadbalancing:DescribeLoadBalancers",
"elasticloadbalancing:DescribeRules",
"elasticloadbalancing:DescribeTargetGroups",
"ecs:*",
"events:DescribeRule",
"events>DeleteRule",
"events:ListRuleNamesByTarget",
"events:ListTargetsByRule",
"events:PutRule",
"events:PutTargets",
"events:RemoveTargets",
"iam:ListAttachedRolePolicies",
"iam:ListInstanceProfiles",
"iam:ListRoles",
"logs>CreateLogGroup",
"logs:DescribeLogGroups",
```

```

        "logs:FilterLogEvents",
        "route53:GetHostedZone",
        "route53:ListHostedZonesByName",
        "route53:CreateHostedZone",
        "route53>DeleteHostedZone",
        "route53:GetHealthCheck",
        "servicediscovery:CreatePrivateDnsNamespace",
        "servicediscovery:CreateService",
        "servicediscovery:GetNamespace",
        "servicediscovery:GetOperation",
        "servicediscovery:GetService",
        "servicediscovery:ListNamespaces",
        "servicediscovery:ListServices",
        "servicediscovery:UpdateService",
        "servicediscovery>DeleteService"
    ],
    "Resource": [
        "*"
    ]
},
{
    "Effect": "Allow",
    "Action": [
        "ssm:GetParametersByPath",
        "ssm:GetParameters",
        "ssm:GetParameter"
    ],
    "Resource": "arn:aws:ssm:*:*:parameter/aws/service/ecs*"
},
{
    "Effect": "Allow",
    "Action": [
        "ec2>DeleteInternetGateway",
        "ec2>DeleteRoute",
        "ec2>DeleteRouteTable",
        "ec2>DeleteSecurityGroup"
    ],
    "Resource": [
        "*"
    ],
    "Condition": {
        "StringLike": {
            "ec2:ResourceTag/aws:cloudformation:stack-name": "EC2ContainerService-
*"
        }
    }
},
{
    "Action": "iam:PassRole",
    "Effect": "Allow",
    "Resource": [
        "*"
    ],
    "Condition": {
        "StringLike": {
            "iam:PassedToService": "ecs-tasks.amazonaws.com"
        }
    }
},
{
    "Action": "iam:PassRole",
    "Effect": "Allow",
    "Resource": [
        "arn:aws:iam:*:*:role/ecsInstanceRole*"
    ],
    "Condition": {

```

```

        "StringLike": {
            "iam:PassedToService": [
                "ec2.amazonaws.com",
                "ec2.amazonaws.com.cn"
            ]
        }
    },
    {
        "Action": "iam:PassRole",
        "Effect": "Allow",
        "Resource": [
            "arn:aws:iam::*:role/ecsAutoscaleRole*"
        ],
        "Condition": {
            "StringLike": {
                "iam:PassedToService": [
                    "application-autoscaling.amazonaws.com",
                    "application-autoscaling.amazonaws.com.cn"
                ]
            }
        }
    },
    {
        "Effect": "Allow",
        "Action": "iam:CreateServiceLinkedRole",
        "Resource": "*",
        "Condition": {
            "StringLike": {
                "iam:AWSServiceName": [
                    "ecs.amazonaws.com",
                    "spot.amazonaws.com",
                    "spotfleet.amazonaws.com",
                    "ecs.application-autoscaling.amazonaws.com",
                    "autoscaling.amazonaws.com"
                ]
            }
        }
    }
]
}

```

The first run wizard also attempts to automatically create different IAM roles depending on the launch type of the tasks used. Examples are the Amazon ECS service role, container instance IAM role, and the task execution IAM role. To ensure that the first-run experience is able to create these IAM roles, one of the following must be true:

- Your user has administrator access. For more information, see [Setting Up with Amazon ECS \(p. 3\)](#).
- Your user has the IAM permissions to create a service role. For more information, see [Creating a Role to Delegate Permissions to an AWS Service](#).
- You have a user with administrator access manually create the required IAM role so it is available on the account to be used. For more information, see the following:
 - [Service Scheduler IAM Role \(p. 233\)](#)
 - [Amazon ECS Task Execution IAM Role \(p. 236\)](#)

Cluster Examples

The following IAM policy allows permission to create and list clusters. The `CreateCluster` and `ListClusters` actions do not accept any resources, so the resource definition is set to `*` for all resources.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:CreateCluster",
        "ecs:ListClusters"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

The following IAM policy allows permission to describe and delete a specific cluster. The `DescribeClusters` and `DeleteCluster` actions accept cluster ARNs as resources.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:DescribeClusters",
        "ecs>DeleteCluster"
      ],
      "Resource": [
        "arn:aws:ecs:us-east-1:<aws_account_id>:cluster/<cluster_name>"
      ]
    }
  ]
}
```

The following IAM policy can be attached to a user or group that would only allow that user or group to perform operations on a specific cluster.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "ecs:Describe*",
        "ecs:List*"
      ],
      "Effect": "Allow",
      "Resource": "*"
    },
    {
      "Action": [
        "ecs>DeleteCluster",
        "ecs:DeregisterContainerInstance",
        "ecs:ListContainerInstances",
        "ecs:RegisterContainerInstance",
        "ecs:SubmitContainerStateChange",
        "ecs:SubmitTaskStateChange"
      ],
      "Effect": "Allow",
      "Resource": "arn:aws:ecs:us-east-1:<aws_account_id>:cluster/default"
    }
  ]
}
```



```

    "Action": [
      "ecs:DescribeContainerInstances",
      "ecs:DescribeTasks",
      "ecs:ListTasks",
      "ecs:UpdateContainerAgent",
      "ecs:StartTask",
      "ecs:StopTask",
      "ecs:RunTask"
    ],
    "Effect": "Allow",
    "Resource": "*",
    "Condition": {
      "ArnEquals": {
        "ecs:cluster": "arn:aws:ecs:us-east-1:<aws_account_id>:cluster/default"
      }
    }
  }
]
}

```

List and Describe Task Examples

The following IAM policy allows a user to list tasks for a specified cluster:

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:ListTasks"
      ],
      "Condition": {
        "ArnEquals": {
          "ecs:cluster": "arn:aws:ecs:<region>:<aws_account_id>:cluster/<cluster_name>"
        }
      },
      "Resource": [
        "*"
      ]
    }
  ]
}

```

The following IAM policy allows a user to describe a specified task in a specified cluster:

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:DescribeTasks"
      ],
      "Condition": {
        "ArnEquals": {
          "ecs:cluster": "arn:aws:ecs:<region>:<aws_account_id>:cluster/<cluster_name>"
        }
      },
      "Resource": [
        "arn:aws:ecs:<region>:<aws_account_id>:task/<task_UUID>"
      ]
    }
  ]
}

```

```
]
}
```

Create Service Example

The following IAM policy allows a user to create Amazon ECS services in the AWS Management Console:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "application-autoscaling:Describe*",
        "application-autoscaling:PutScalingPolicy",
        "application-autoscaling:RegisterScalableTarget",
        "cloudwatch:DescribeAlarms",
        "cloudwatch:PutMetricAlarm",
        "ecs:List*",
        "ecs:Describe*",
        "ecs:CreateService",
        "elasticloadbalancing:Describe*",
        "iam:AttachRolePolicy",
        "iam:CreateRole",
        "iam:GetPolicy",
        "iam:GetPolicyVersion",
        "iam:GetRole",
        "iam:ListAttachedRolePolicies",
        "iam:ListRoles",
        "iam:ListGroups",
        "iam:ListUsers"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

Update Service Example

The following IAM policy allows a user to update Amazon ECS services in the AWS Management Console:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "application-autoscaling:Describe*",
        "application-autoscaling:PutScalingPolicy",
        "application-autoscaling>DeleteScalingPolicy",
        "application-autoscaling:RegisterScalableTarget",
        "cloudwatch:DescribeAlarms",
        "cloudwatch:PutMetricAlarm",
        "ecs:List*",
        "ecs:Describe*",
        "ecs:UpdateService",
        "iam:AttachRolePolicy",
        "iam:CreateRole",
        "iam:GetPolicy",
        "iam:GetPolicyVersion",

```

```

        "iam:GetRole",
        "iam:ListAttachedRolePolicies",
        "iam:ListRoles",
        "iam:ListGroups",
        "iam:ListUsers"
    ],
    "Resource": [
        "*"
    ]
}
]
}

```

Describing Amazon ECS Services Based on Tags

You can use conditions in your identity-based policy to control access to Amazon ECS resources based on tags. This example shows how you might create a policy that allows describing your services. However, permission is granted only if the service tag `Owner` has the value of that user's user name. This policy also grants the permissions necessary to complete this action on the console.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "DescribeServices",
      "Effect": "Allow",
      "Action": "ecs:DescribeServices",
      "Resource": "*"
    },
    {
      "Sid": "ViewServiceIfOwner",
      "Effect": "Allow",
      "Action": "ecs:DescribeServices",
      "Resource": "arn:aws:ecs:*:*:service/*",
      "Condition": {
        "StringEquals": {"ecs:ResourceTag/Owner": "${aws:username}"}
      }
    }
  ]
}

```

You can attach this policy to the IAM users in your account. If a user named `richard-roe` attempts to describe an Amazon ECS service, the service must be tagged `Owner=richard-roe` or `owner=richard-roe`. Otherwise he is denied access. The condition tag key `Owner` matches both `Owner` and `owner` because condition key names are not case-sensitive. For more information, see [IAM JSON Policy Elements: Condition](#) in the *IAM User Guide*.

Supported Resource-Level Permissions for Amazon ECS API Actions

The term *resource-level permissions* refers to the ability to specify which resources users are allowed to perform actions on. Amazon ECS has partial support for resource-level permissions. This means that for certain Amazon ECS actions, you can control when users are allowed to use those actions based on conditions that have to be fulfilled, or specific resources that users are allowed to use. For example, you can grant users permission to launch instances, but only of a specific type, and only using a specific AMI.

For more information about the resources that are created or modified by the Amazon ECS actions, and the ARNs and Amazon ECS condition keys that you can use in an IAM policy statement, see [Actions, Resources, and Condition Keys for Amazon Elastic Container Service](#) in the *IAM User Guide*.

Considerations for Resource-Level Permissions

When controlling access to Amazon ECS API actions by specifying the Amazon Resource Name (ARN) of a resource in an IAM policy, be mindful that ECS has introduced an account setting that affects the ARN format for container instances, services, and tasks. To use resource-level permissions, we recommend that you opt-in to the new, longer ARN format. For more information, see [Amazon Resource Names \(ARNs\) and IDs](#) (p. 104).

When an IAM policy is evaluated, the specified resources are evaluated based on their use of the new, longer ARN format. The following are examples of how access is controlled.

Specifying a Service with a Cluster Only with a Wildcard

Example: `arn:aws:ecs:region:aws_account_id:service/cluster_name*`

In this example, access will be controlled to the following services:

- All services using the new ARN format that are in the `cluster_name*` cluster.
- All services using the old ARN format that are in the `cluster_name*` cluster.

Important

This will **NOT** control access to services using the old ARN format that have a service name with the `cluster_name` prefix that are not in the `cluster_name*` cluster.

Specifying a Service with both a Cluster and Service Name with a Wildcard

Example: `arn:aws:ecs:region:aws_account_id:service/cluster_name/service_name*`

In this example, access will be controlled to the following services:

- All services using the new ARN format that are in the `cluster_name` cluster with the `service_name` prefix.
- All services using the old ARN format that are in the `cluster_name` cluster with the `service_name` prefix, even though the actual ARN of the service will still have the `arn:aws:ecs:region:aws_account_id:service/service_name*` ARN format.

Specifying a Service with a full ARN

Example: `arn:aws:ecs:region:aws_account_id:service/cluster_name/service_name`

In this example, access will be controlled to the following services:

- All services using the new ARN format that are in the `cluster_name` cluster with the `service_name` service name.
- All services using the old ARN format that are in the `cluster_name` cluster with the `service_name` service name, even though the actual ARN of the service will still have the `arn:aws:ecs:region:aws_account_id:service/service_name` ARN format.

Managed Policies and Trust Relationships

Amazon ECS and Amazon ECR provide several managed policies and trust relationships that you can attach to IAM users, EC2 instances, and Amazon ECS tasks that allow differing levels of control over resources and API operations. You can apply these policies directly, or you can use them as starting points for creating your own policies.

Topics

- [Amazon ECS Managed Policies and Trust Relationships \(p. 221\)](#)
- [Amazon ECR Managed Policies \(p. 226\)](#)

Amazon ECS Managed Policies and Trust Relationships

Amazon ECS provides several managed policies and trust relationships that you can attach to IAM users, EC2 instances, or Amazon ECS tasks that allow differing levels of control over Amazon ECS resources and API operations. You can apply these policies directly, or you can use them as starting points for creating your own policies. For more information about each API operation mentioned in these policies, see [Actions](#) in the *Amazon Elastic Container Service API Reference*.

Topics

- [AmazonECS_FullAccess \(p. 221\)](#)
- [AmazonEC2ContainerServiceFullAccess \(p. 224\)](#)
- [AmazonEC2ContainerServiceRole \(p. 225\)](#)
- [AmazonEC2ContainerServiceAutoscaleRole \(p. 225\)](#)

AmazonECS_FullAccess

This managed policy provides administrative access to Amazon ECS resources and enables ECS features through access to other AWS service resources, including VPCs, Auto Scaling groups, and AWS CloudFormation stacks.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "application-autoscaling:DeleteScalingPolicy",
        "application-autoscaling:DeregisterScalableTarget",
        "application-autoscaling:DescribeScalableTargets",
        "application-autoscaling:DescribeScalingActivities",
        "application-autoscaling:DescribeScalingPolicies",
        "application-autoscaling:PutScalingPolicy",
        "application-autoscaling:RegisterScalableTarget",
        "appmesh:ListMeshes",
        "appmesh:ListVirtualNodes",
        "appmesh:DescribeVirtualNode",
        "autoscaling:UpdateAutoScalingGroup",
        "autoscaling:CreateAutoScalingGroup",
        "autoscaling:CreateLaunchConfiguration",
        "autoscaling:DeleteAutoScalingGroup",
        "autoscaling:DeleteLaunchConfiguration",
        "autoscaling:Describe*",
        "cloudformation:CreateStack",
        "cloudformation:DeleteStack",
        "cloudformation:DescribeStack*",
        "cloudformation:UpdateStack",
        "cloudwatch:DescribeAlarms",
        "cloudwatch:DeleteAlarms",
        "cloudwatch:GetMetricStatistics",
        "cloudwatch:PutMetricAlarm",
        "codedeploy:CreateApplication",
        "codedeploy:CreateDeployment",
        "codedeploy:CreateDeploymentGroup",

```

```
"codedeploy:GetApplication",
"codedeploy:GetDeployment",
"codedeploy:GetDeploymentGroup",
"codedeploy:ListApplications",
"codedeploy:ListDeploymentGroups",
"codedeploy:ListDeployments",
"codedeploy:StopDeployment",
"codedeploy:GetDeploymentTarget",
"codedeploy:ListDeploymentTargets",
"codedeploy:GetDeploymentConfig",
"codedeploy:GetApplicationRevision",
"codedeploy:RegisterApplicationRevision",
"codedeploy:BatchGetApplicationRevisions",
"codedeploy:BatchGetDeploymentGroups",
"codedeploy:BatchGetDeployments",
"codedeploy:BatchGetApplications",
"codedeploy:ListApplicationRevisions",
"codedeploy:ListDeploymentConfigs",
"codedeploy:ContinueDeployment",
"sns:ListTopics",
"lambda:ListFunctions",
"ec2:AssociateRouteTable",
"ec2:AttachInternetGateway",
"ec2:AuthorizeSecurityGroupIngress",
"ec2:CancelSpotFleetRequests",
"ec2:CreateInternetGateway",
"ec2:CreateLaunchTemplate",
"ec2:CreateRoute",
"ec2:CreateRouteTable",
"ec2:CreateSecurityGroup",
"ec2:CreateSubnet",
"ec2:CreateVpc",
"ec2>DeleteLaunchTemplate",
"ec2>DeleteSubnet",
"ec2>DeleteVpc",
"ec2:Describe*",
"ec2:DetachInternetGateway",
"ec2:DisassociateRouteTable",
"ec2:ModifySubnetAttribute",
"ec2:ModifyVpcAttribute",
"ec2:RunInstances",
"ec2:RequestSpotFleet",
"elasticloadbalancing:CreateListener",
"elasticloadbalancing:CreateLoadBalancer",
"elasticloadbalancing:CreateRule",
"elasticloadbalancing:CreateTargetGroup",
"elasticloadbalancing>DeleteListener",
"elasticloadbalancing>DeleteLoadBalancer",
"elasticloadbalancing>DeleteRule",
"elasticloadbalancing>DeleteTargetGroup",
"elasticloadbalancing:DescribeListeners",
"elasticloadbalancing:DescribeLoadBalancers",
"elasticloadbalancing:DescribeRules",
"elasticloadbalancing:DescribeTargetGroups",
"ecs:*",
"events:DescribeRule",
"events>DeleteRule",
"events:ListRuleNamesByTarget",
"events:ListTargetsByRule",
"events:PutRule",
"events:PutTargets",
"events:RemoveTargets",
"iam:ListAttachedRolePolicies",
"iam:ListInstanceProfiles",
"iam:ListRoles",
"logs:CreateLogGroup",
```

```

        "logs:DescribeLogGroups",
        "logs:FilterLogEvents",
        "route53:GetHostedZone",
        "route53:ListHostedZonesByName",
        "route53:CreateHostedZone",
        "route53>DeleteHostedZone",
        "route53:GetHealthCheck",
        "servicediscovery:CreatePrivateDnsNamespace",
        "servicediscovery:CreateService",
        "servicediscovery:GetNamespace",
        "servicediscovery:GetOperation",
        "servicediscovery:GetService",
        "servicediscovery:ListNamespaces",
        "servicediscovery:ListServices",
        "servicediscovery:UpdateService",
        "servicediscovery>DeleteService"
    ],
    "Resource": [
        "*"
    ]
},
{
    "Effect": "Allow",
    "Action": [
        "ssm:GetParametersByPath",
        "ssm:GetParameters",
        "ssm:GetParameter"
    ],
    "Resource": "arn:aws:ssm:*:*:parameter/aws/service/ecs*"
},
{
    "Effect": "Allow",
    "Action": [
        "ec2>DeleteInternetGateway",
        "ec2>DeleteRoute",
        "ec2>DeleteRouteTable",
        "ec2>DeleteSecurityGroup"
    ],
    "Resource": [
        "*"
    ],
    "Condition": {
        "StringLike": {
            "ec2:ResourceTag/aws:cloudformation:stack-name": "EC2ContainerService-
*"
        }
    }
},
{
    "Action": "iam:PassRole",
    "Effect": "Allow",
    "Resource": [
        "*"
    ],
    "Condition": {
        "StringLike": {
            "iam:PassedToService": "ecs-tasks.amazonaws.com"
        }
    }
},
{
    "Action": "iam:PassRole",
    "Effect": "Allow",
    "Resource": [
        "arn:aws:iam:*:*:role/ecsInstanceRole*"
    ],

```

```

        "Condition": {
            "StringLike": {
                "iam:PassedToService": [
                    "ec2.amazonaws.com",
                    "ec2.amazonaws.com.cn"
                ]
            }
        },
        {
            "Action": "iam:PassRole",
            "Effect": "Allow",
            "Resource": [
                "arn:aws:iam::*:role/ecsAutoscaleRole*"
            ],
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": [
                        "application-autoscaling.amazonaws.com",
                        "application-autoscaling.amazonaws.com.cn"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
            "Resource": "*",
            "Condition": {
                "StringLike": {
                    "iam:AWSServiceName": [
                        "ecs.amazonaws.com",
                        "spot.amazonaws.com",
                        "spotfleet.amazonaws.com",
                        "ecs.application-autoscaling.amazonaws.com",
                        "autoscaling.amazonaws.com"
                    ]
                }
            }
        }
    ]
}

```

AmazonEC2ContainerServiceFullAccess

This managed policy allows full administrator access to Amazon ECS.

```

{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "autoscaling:Describe*",
                "autoscaling:UpdateAutoScalingGroup",
                "cloudformation:CreateStack",
                "cloudformation:DeleteStack",
                "cloudformation:DescribeStack*",
                "cloudformation:UpdateStack",
                "cloudwatch:GetMetricStatistics",
                "ec2:Describe*",
                "elasticloadbalancing:*",
                "ecs:*",
                "events:DescribeRule",
            ]
        }
    ]
}

```



```

        "events:DeleteRule",
        "events:ListRuleNamesByTarget",
        "events:ListTargetsByRule",
        "events:PutRule",
        "events:PutTargets",
        "events:RemoveTargets",
        "iam:ListInstanceProfiles",
        "iam:ListRoles",
        "iam:PassRole"
    ],
    "Resource": "*"
}
]
}

```

AmazonEC2ContainerServiceRole

This managed policy allows Elastic Load Balancing load balancers to register and deregister Amazon ECS container instances on your behalf. For more information, see [Service Scheduler IAM Role \(p. 233\)](#).

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:AuthorizeSecurityGroupIngress",
        "ec2:Describe*",
        "elasticloadbalancing:DeregisterInstancesFromLoadBalancer",
        "elasticloadbalancing:DeregisterTargets",
        "elasticloadbalancing:Describe*",
        "elasticloadbalancing:RegisterInstancesWithLoadBalancer",
        "elasticloadbalancing:RegisterTargets"
      ],
      "Resource": "*"
    }
  ]
}

```

AmazonEC2ContainerServiceAutoscaleRole

This managed policy allows Application Auto Scaling to scale your Amazon ECS service's desired count up and down in response to CloudWatch alarms on your behalf. For more information, see [Service Auto Scaling IAM Role \(p. 235\)](#).

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:DescribeServices",
        "ecs:UpdateService"
      ],
      "Resource": [
        "*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [

```

```
        "cloudwatch:DescribeAlarms",
        "cloudwatch:PutMetricAlarm"
    ],
    "Resource": [
        "*"
    ]
}
]
```

Amazon ECR Managed Policies

Amazon ECR provides several managed policies that you can attach to IAM users or EC2 instances that allow differing levels of control over Amazon ECR resources and API operations. You can apply these policies directly, or you can use them as starting points for creating your own policies. For more information about each API operation mentioned in these policies, see [Actions](#) in the *Amazon Elastic Container Registry API Reference*.

Topics

- [AmazonEC2ContainerRegistryFullAccess](#) (p. 226)
- [AmazonEC2ContainerRegistryPowerUser](#) (p. 226)
- [AmazonEC2ContainerRegistryReadOnly](#) (p. 227)

AmazonEC2ContainerRegistryFullAccess

This managed policy is a starting point for customers who are looking to provide an IAM user or role with full administrator access to manage their use of Amazon ECR. The [Amazon ECR Lifecycle Policies](#) feature enables customers to specify the lifecycle management of images in a repository. Lifecycle policy events are reported as CloudTrail events, and Amazon ECR is integrated with AWS CloudTrail to display a customer's lifecycle policy events directly in the Amazon ECR console. The `AmazonEC2ContainerRegistryFullAccess` managed IAM policy includes the `cloudtrail:LookupEvents` permission to facilitate this behavior.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecr:*",
        "cloudtrail:LookupEvents"
      ],
      "Resource": "*"
    }
  ]
}
```

AmazonEC2ContainerRegistryPowerUser

This managed policy allows power user access to Amazon ECR, which allows read and write access to repositories, but does not allow users to delete repositories or change the policy documents applied to them.

```
{
  "Version": "2012-10-17",
  "Statement": [
```

```
{
  "Effect": "Allow",
  "Action": [
    "ecr:GetAuthorizationToken",
    "ecr:BatchCheckLayerAvailability",
    "ecr:GetDownloadUrlForLayer",
    "ecr:GetRepositoryPolicy",
    "ecr:DescribeRepositories",
    "ecr:ListImages",
    "ecr:DescribeImages",
    "ecr:BatchGetImage",
    "ecr:GetLifecyclePolicy",
    "ecr:GetLifecyclePolicyPreview",
    "ecr:ListTagsForResource",
    "ecr:DescribeImageScanFindings",
    "ecr:InitiateLayerUpload",
    "ecr:UploadLayerPart",
    "ecr:CompleteLayerUpload",
    "ecr:PutImage"
  ],
  "Resource": "*"
}
```

AmazonEC2ContainerRegistryReadOnly

This managed policy allows read-only access to Amazon ECR, such as the ability to list repositories and the images within the repositories, and also to pull images from Amazon ECR with the Docker CLI.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecr:GetAuthorizationToken",
        "ecr:BatchCheckLayerAvailability",
        "ecr:GetDownloadUrlForLayer",
        "ecr:GetRepositoryPolicy",
        "ecr:DescribeRepositories",
        "ecr:ListImages",
        "ecr:DescribeImages",
        "ecr:BatchGetImage",
        "ecr:GetLifecyclePolicy",
        "ecr:GetLifecyclePolicyPreview",
        "ecr:ListTagsForResource",
        "ecr:DescribeImageScanFindings"
      ],
      "Resource": "*"
    }
  ]
}
```

Service-Linked Role for Amazon ECS

Amazon Elastic Container Service uses a service-linked role for the permissions it requires to call other AWS services on your behalf. For more information, see [Using Service-Linked Roles](#) in the *IAM User Guide*.

Prior to the introduction of a service-linked role for Amazon ECS, you were required to create an IAM role for your Amazon ECS services which granted Amazon ECS the permission it needed. This role is no longer

required, however it is available if needed. For more information, see [Legacy IAM Roles for Amazon ECS \(p. 233\)](#).

Permissions Granted by the Service-Linked Role

Amazon ECS uses the service-linked role named **AWSServiceRoleForECS** to enable Amazon ECS to call AWS APIs on your behalf.

The **AWSServiceRoleForECS** service-linked role trusts the `ecs.amazonaws.com` service principal to assume the role.

The role permissions policy allows Amazon ECS to complete the following actions on resources.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "ECSTaskManagement",
      "Effect": "Allow",
      "Action": [
        "ec2:AttachNetworkInterface",
        "ec2:CreateNetworkInterface",
        "ec2:CreateNetworkInterfacePermission",
        "ec2>DeleteNetworkInterface",
        "ec2>DeleteNetworkInterfacePermission",
        "ec2:Describe*",
        "ec2:DetachNetworkInterface",
        "elasticloadbalancing:DeregisterInstancesFromLoadBalancer",
        "elasticloadbalancing:DeregisterTargets",
        "elasticloadbalancing:Describe*",
        "elasticloadbalancing:RegisterInstancesWithLoadBalancer",
        "elasticloadbalancing:RegisterTargets",
        "route53:ChangeResourceRecordSets",
        "route53:CreateHealthCheck",
        "route53>DeleteHealthCheck",
        "route53:Get*",
        "route53:List*",
        "route53:UpdateHealthCheck",
        "servicediscovery:DeregisterInstance",
        "servicediscovery:Get*",
        "servicediscovery:List*",
        "servicediscovery:RegisterInstance",
        "servicediscovery:UpdateInstanceCustomHealthStatus"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AutoScaling",
      "Effect": "Allow",
      "Action": [
        "autoscaling:Describe*"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AutoScalingManagement",
      "Effect": "Allow",
      "Action": [
        "autoscaling>DeletePolicy",
        "autoscaling:PutScalingPolicy",
        "autoscaling:SetInstanceProtection",
        "autoscaling:UpdateAutoScalingGroup"
      ]
    }
  ]
}
```

```

        "Resource": "*",
        "Condition": {
            "Null": {
                "autoscaling:ResourceTag/AmazonECSManaged": "false"
            }
        }
    },
    {
        "Sid": "AutoScalingPlanManagement",
        "Effect": "Allow",
        "Action": [
            "autoscaling-plans:CreateScalingPlan",
            "autoscaling-plans>DeleteScalingPlan",
            "autoscaling-plans:DescribeScalingPlans"
        ],
        "Resource": "*"
    },
    {
        "Sid": "CWAlarmManagement",
        "Effect": "Allow",
        "Action": [
            "cloudwatch:DeleteAlarms",
            "cloudwatch:DescribeAlarms",
            "cloudwatch:PutMetricAlarm"
        ],
        "Resource": "arn:aws:cloudwatch:*:*:alarm:*"
    },
    {
        "Sid": "ECSTagging",
        "Effect": "Allow",
        "Action": [
            "ec2:CreateTags"
        ],
        "Resource": "arn:aws:ec2:*:*:network-interface/*"
    },
    {
        "Sid": "CWLogGroupManagement",
        "Effect": "Allow",
        "Action": [
            "logs:CreateLogGroup",
            "logs:DescribeLogGroups",
            "logs:PutRetentionPolicy"
        ],
        "Resource": "arn:aws:logs:*:*:log-group:/aws/ecs/*"
    },
    {
        "Sid": "CWLogStreamManagement",
        "Effect": "Allow",
        "Action": [
            "logs:CreateLogStream",
            "logs:DescribeLogStreams",
            "logs:PutLogEvents"
        ],
        "Resource": "arn:aws:logs:*:*:log-group:/aws/ecs/*:log-stream:*"
    }
]
}

```

Create the Service-Linked Role

Under most circumstances, you don't need to manually create the service-linked role. For example, when you create a new cluster (for example, with the Amazon ECS first-run experience, the cluster creation wizard, or the AWS CLI or SDKs), or create or update a service in the AWS Management Console, Amazon ECS creates the service-linked role for you, if it does not already exist.

You must configure permissions to allow an IAM entity (such as a user, group, or role) to create, edit, or delete a service-linked role.

To allow an IAM entity to create the `AWSServiceRoleForECS` service-linked role

Add the following statement to the permissions policy for the IAM entity that needs to create the service-linked role:

```
{
  "Effect": "Allow",
  "Action": [
    "iam:CreateServiceLinkedRole"
  ],
  "Resource": "arn:aws:iam::*:role/aws-service-role/ecs.amazonaws.com/
AWSServiceRoleForECS*",
  "Condition": {"StringLike": {"iam:AWSServiceName": "ecs.amazonaws.com"}}
}
```

Creating a Service-Linked Role in IAM (AWS CLI)

You can use IAM commands from the AWS Command Line Interface to create a service-linked role with the trust policy and inline policies that the service needs to assume the role.

To create a service-linked role (CLI)

Use the following command:

```
$ aws iam create-service-linked-role --aws-service-name ecs.amazonaws.com
```

Edit the Service-Linked Role

Amazon ECS does not allow you to edit the `AWSServiceRoleForECS` service-linked role. After you create a service-linked role, you cannot change the name of the role because various entities might reference the role. You can, however, edit the description of the role. For more information, see [Modifying a Role](#) in the *IAM User Guide*.

To allow an IAM entity to edit the description of the `AWSServiceRoleForECS` service-linked role

Add the following statement to the permissions policy for the IAM entity that needs to edit the description of a service-linked role:

```
{
  "Effect": "Allow",
  "Action": [
    "iam:UpdateRoleDescription"
  ],
  "Resource": "arn:aws:iam::*:role/aws-service-role/ecs.amazonaws.com/
AWSServiceRoleForECS*",
  "Condition": {"StringLike": {"iam:AWSServiceName": "ecs.amazonaws.com"}}
}
```

Delete the Service-Linked Role

If you no longer use Amazon ECS, we recommend that you delete the service-linked role. That way you don't have an unused entity that is not actively monitored or maintained. However, you must delete all Amazon ECS clusters in all regions before you can delete the service-linked role.

To allow an IAM entity to delete the `AWSServiceRoleForECS` service-linked role

Add the following statement to the permissions policy for the IAM entity that needs to delete a service-linked role:

```
{
  "Effect": "Allow",
  "Action": [
    "iam:DeleteServiceLinkedRole",
    "iam:GetServiceLinkedRoleDeletionStatus"
  ],
  "Resource": "arn:aws:iam::*:role/aws-service-role/ecs.amazonaws.com/AWSServiceRoleForECS*",
  "Condition": {"StringLike": {"iam:AWSServiceName": "ecs.amazonaws.com"}}
}
```

Cleaning up a Service-Linked Role

Before you can use IAM to delete a service-linked role, you must first confirm that the role has no active sessions and delete all Amazon ECS clusters in all AWS Regions.

To check whether the service-linked role has an active session

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles** and choose the **AWSServiceRoleForECS** name (not the check box).
3. On the **Summary** page, choose **Access Advisor** and review recent activity for the service-linked role.

Note

If you are unsure whether Amazon ECS is using the **AWSServiceRoleForECS** role, you can try to delete the role. If the service is using the role, then the deletion fails and you can view the regions where the role is being used. If the role is being used, then you must wait for the session to end before you can delete the role. You cannot revoke the session for a service-linked role.

To remove Amazon ECS resources used by the **AWSServiceRoleForECS** service-linked role

You must delete all Amazon ECS clusters in all AWS Regions before you can delete the **AWSServiceRoleForECS** role.

- Delete all Amazon ECS clusters in all regions. For more information, see [Deleting a Cluster \(p. 24\)](#).

Deleting a Service-Linked Role in IAM (Console)

You can use the IAM console to delete a service-linked role.

To delete a service-linked role (console)

1. Sign in to the AWS Management Console and open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane of the IAM console, choose **Roles**. Then select the check box next to **AWSServiceRoleForECS**, not the name or row itself.
3. Choose **Delete role**.
4. In the confirmation dialog box, review the service last accessed data, which shows when each of the selected roles last accessed an AWS service. This helps you to confirm whether the role is currently active. If you want to proceed, choose **Yes, Delete** to submit the service-linked role for deletion.
5. Watch the IAM console notifications to monitor the progress of the service-linked role deletion. Because the IAM service-linked role deletion is asynchronous, after you submit the role for deletion, the deletion task can succeed or fail.

- If the task succeeds, then the role is removed from the list and a notification of success appears at the top of the page.
- If the task fails, you can choose **View details** or **View Resources** from the notifications to learn why the deletion failed. If the deletion fails because the role is using the service's resources, then the notification includes a list of resources, if the service returns that information. You can then [clean up the resources](#) and submit the deletion again.

Note

You might have to repeat this process several times, depending on the information that the service returns. For example, your service-linked role might use six resources and your service might return information about five of them. If you clean up the five resources and submit the role for deletion again, the deletion fails and the service reports the one remaining resource. A service might return all of the resources, a few of them, or it might not report any resources.

- If the task fails and the notification does not include a list of resources, then the service might not return that information. To learn how to clean up the resources for that service, see [AWS Services That Work with IAM](#). Find your service in the table, and choose the **Yes** link to view the service-linked role documentation for that service.

Deleting a Service-Linked Role in IAM (AWS CLI)

You can use IAM commands from the AWS Command Line Interface to delete a service-linked role.

To delete a service-linked role (CLI)

1. Because a service-linked role cannot be deleted if it is being used or has associated resources, you must submit a deletion request. That request can be denied if these conditions are not met. You must capture the `deletion-task-id` from the response to check the status of the deletion task. Enter the following command to submit a service-linked role deletion request:

```
$ aws iam delete-service-linked-role --role-name AWSServiceRoleForECS+OPTIONAL-SUFFIX
```

2. Use the following command to check the status of the deletion task:

```
$ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
```

The status of the deletion task can be `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, or `FAILED`. If the deletion fails, the call returns the reason that it failed so that you can troubleshoot. If the deletion fails because the role is using the service's resources, then the notification includes a list of resources, if the service returns that information. You can then [clean up the resources](#) and submit the deletion again.

Note

You might have to repeat this process several times, depending on the information that the service returns. For example, your service-linked role might use six resources and your service might return information about five of them. If you clean up the five resources and submit the role for deletion again, the deletion fails and the service reports the one remaining resource. A service might return all of the resources, a few of them, or it might not report any resources. To learn how to clean up the resources for a service that does not report any resources, see [AWS Services That Work with IAM](#). Find your service in the table, and choose the **Yes** link to view the service-linked role documentation for that service.

Deleting a Service-Linked Role in IAM (AWSAPI)

You can use the IAM API to delete a service-linked role.

To delete a service-linked role (API)

1. To submit a deletion request for a service-linked role, call [DeleteServiceLinkedRole](#). In the request, specify the `AWSServiceRoleForECS` role name.

Because a service-linked role cannot be deleted if it is being used or has associated resources, you must submit a deletion request. That request can be denied if these conditions are not met. You must capture the `DeletionTaskId` from the response to check the status of the deletion task.

2. To check the status of the deletion, call [GetServiceLinkedRoleDeletionStatus](#). In the request, specify the `DeletionTaskId`.

The status of the deletion task can be `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, or `FAILED`. If the deletion fails, the call returns the reason that it failed so that you can troubleshoot. If the deletion fails because the role is using the service's resources, then the notification includes a list of resources, if the service returns that information. You can then [clean up the resources](#) and submit the deletion again.

Note

You might have to repeat this process several times, depending on the information that the service returns. For example, your service-linked role might use six resources and your service might return information about five of them. If you clean up the five resources and submit the role for deletion again, the deletion fails and the service reports the one remaining resource. A service might return all of the resources, a few of them, or it might not report any resources. To learn how to clean up the resources for a service that does not report any resources, see [AWS Services That Work with IAM](#). Find your service in the table, and choose the **Yes** link to view the service-linked role documentation for that service.

Legacy IAM Roles for Amazon ECS

Prior to the introduction of the `AWSServiceRoleForECS` IAM role, you were required to create separate IAM roles to enable Amazon ECS permissions to call Elastic Load Balancing and Application Auto Scaling APIs on your behalf.

The Amazon ECS service scheduler IAM role grants the Amazon ECS service scheduler permissions that it needs to register and deregister container instances with your load balancers. You can optionally create the service scheduler IAM role and specify it when creating a service, or preferably you can allow Amazon ECS to use the service-linked role.

The Amazon ECS Service Auto Scaling IAM role grants Amazon ECS permission to describe your CloudWatch alarms and registered services, as well as permission to update your Amazon ECS service's desired count on your behalf.

These legacy IAM roles are described in more detail below, but have effectively been replaced by the Amazon ECS service-linked role.

Service Scheduler IAM Role

Amazon ECS provides a managed IAM policy named `AmazonEC2ContainerServiceRole` to use for the service scheduler IAM role. The `AmazonEC2ContainerServiceRole` policy is shown below.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:AuthorizeSecurityGroupIngress",
        "ec2:Describe*"
      ]
    }
  ]
}
```

```
        "elasticloadbalancing:DeregisterInstancesFromLoadBalancer",
        "elasticloadbalancing:DeregisterTargets",
        "elasticloadbalancing:Describe*",
        "elasticloadbalancing:RegisterInstancesWithLoadBalancer",
        "elasticloadbalancing:RegisterTargets"
    ],
    "Resource": "*"
}
]
```

Note

The `ec2:AuthorizeSecurityGroupIngress` rule is reserved for future use. Amazon ECS does not automatically update the security groups associated with Elastic Load Balancing load balancers or Amazon ECS container instances.

To check for the `ecsServiceRole` in the IAM console

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**.
3. Search the list of roles for `ecsServiceRole`. If the role does not exist, use the procedure below to create the role. If the role does exist, select the role to view the attached policies.
4. Choose the **Permissions** tab.
5. In the **Managed Policies** section, ensure that the **AmazonEC2ContainerServiceRole** managed policy is attached to the role. If the policy is attached, your Amazon ECS service role is properly configured. If not, follow the substeps below to attach the policy.
 - a. Choose **Attach Policy**.
 - b. To narrow the available policies to attach, for **Filter**, type **AmazonEC2ContainerServiceRole**.
 - c. Check the box to the left of the **AmazonEC2ContainerServiceRole** policy and choose **Attach Policy**.
6. Choose **Trust Relationships, Edit Trust Relationship**.
7. Verify that the trust relationship contains the following policy. If the trust relationship matches the policy below, choose **Cancel**. If the trust relationship does not match, copy the policy into the **Policy Document** window and choose **Update Trust Policy**.

```
{
  "Version": "2008-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "ecs.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

To create an IAM role for your service scheduler load balancers

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles, Create role**.
3. For **Select type of trusted entity** section, choose **AWS service**.
4. For **Choose the service that will use this role**, choose **Elastic Container Service**.

5. For **Select your use case**, choose **Elastic Container Service** and choose **Next: Permissions**.
6. In the **Attached permissions policy** section, select the **AmazonEC2ContainerServiceRole** policy and choose **Next: Review**.
7. For **Role Name**, type `ecsServiceRole`, enter a **Role description** and then choose **Create role**.

Service Auto Scaling IAM Role

Amazon ECS provides a managed IAM policy named `AmazonEC2ContainerServiceAutoscaleRole` to use for the Service Auto Scaling IAM role. The `AmazonEC2ContainerServiceAutoscaleRole` policy is shown below.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:DescribeServices",
        "ecs:UpdateService"
      ],
      "Resource": [
        "*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "cloudwatch:DescribeAlarms",
        "cloudwatch:PutMetricAlarm"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

To check for the Service Auto Scaling role in the IAM console

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**.
3. Search the list of roles for `ecsAutoscaleRole`. If the role does not exist, use the procedure below to create the role. If the role does exist, select the role to view the attached policies.
4. Choose the **Permissions** tab.
5. In the **Permissions policies** section, ensure that the **AmazonEC2ContainerServiceAutoscaleRole** managed policy is attached to the role. If the policy is attached, your Amazon ECS service role is properly configured. If not, follow the substeps below to attach the policy.
 - a. Choose **Attach policies**.
 - b. To narrow the available policies to attach, for **Filter**, type `AmazonEC2ContainerServiceAutoscaleRole`.
 - c. Select the box to the left of the **AmazonEC2ContainerAutoscaleRole** policy and choose **Attach policy**.
6. Choose **Trust relationships, Edit trust relationship**.
7. Verify that the trust relationship contains the following policy. If the trust relationship matches the policy below, choose **Cancel**. If the trust relationship does not match, copy the policy into the **Policy Document** window and choose **Update Trust Policy**.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "application-autoscaling.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

To create an IAM role for Service Auto Scaling

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles** and then choose **Create role**.
3. In the **Choose the service that will use this role** section, choose **Elastic Container Service**.
4. In the **Select your use case** section, choose **Elastic Container Service Autoscale, Next: Permissions**.
5. For **Add tags (optional)**, enter any key value tags you wish to add to the IAM role. Choose **Next: Review** when finished.
6. In the **Role name** field, type `ecsAutoscaleRole` to name the role, and then choose **Create Role** to finish.

Amazon ECS Task Execution IAM Role

The Amazon ECS container agent, and the Fargate agent for your Fargate tasks, make calls to the Amazon ECS API on your behalf. The agent requires an IAM role for the service to know that the agent belongs to you. This IAM role is referred to as a task execution IAM role. You can have multiple task execution roles for different purposes associated with your account.

The following are common use cases for a task execution IAM role:

- Your task uses the Fargate launch type and...
 - is pulling a container image from Amazon ECR.
 - uses the `awslogs` or `awsfirelens` log driver. For more information, see [Using the awslogs Log Driver \(p. 69\)](#) and [Custom Log Routing \(p. 75\)](#).
- Your tasks uses either the Fargate or EC2 launch type and...
 - is using private registry authentication. For more information, see [Required IAM Permissions for Private Registry Authentication \(p. 238\)](#).
 - the task definition is referencing sensitive data using Secrets Manager secrets or AWS Systems Manager Parameter Store parameters. For more information, see [Required IAM Permissions for Amazon ECS Secrets \(p. 238\)](#).

Note

The task execution role is supported by Amazon ECS container agent version 1.16.0 and later.

Amazon ECS provides the managed policy named `AmazonECSTaskExecutionRolePolicy` which contains the permissions the common use cases described above require. It may be necessary to add inline policies to your task execution role for special use cases which are outlined below.

```
{
```

```

"Version": "2012-10-17",
"Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "ecr:GetAuthorizationToken",
      "ecr:BatchCheckLayerAvailability",
      "ecr:GetDownloadUrlForLayer",
      "ecr:BatchGetImage",
      "logs:CreateLogStream",
      "logs:PutLogEvents"
    ],
    "Resource": "*"
  }
]
}

```

An Amazon ECS task execution role is automatically created for you in the Amazon ECS console first-run experience; however, you should manually attach the managed IAM policy for tasks to allow Amazon ECS to add permissions for future features and enhancements as they are introduced. You can use the following procedure to check and see if your account already has the Amazon ECS task execution role and to attach the managed IAM policy if needed.

To check for the `ecsTaskExecutionRole` in the IAM console

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**.
3. Search the list of roles for `ecsTaskExecutionRole`. If the role does not exist, see [Creating the task execution IAM role \(p. 238\)](#). If the role does exist, select the role to view the attached policies.
4. On the **Permissions** tab, ensure that the **AmazonECSTaskExecutionRolePolicy** managed policy is attached to the role. If the policy is attached, your Amazon ECS task execution role is properly configured. If not, follow the substeps below to attach the policy.
 - a. Choose **Attach policies**.
 - b. To narrow the available policies to attach, for **Filter**, type **AmazonECSTaskExecutionRolePolicy**.
 - c. Check the box to the left of the **AmazonECSTaskExecutionRolePolicy** policy and choose **Attach policy**.
5. Choose **Trust relationships, Edit trust relationship**.
6. Verify that the trust relationship contains the following policy. If the trust relationship matches the policy below, choose **Cancel**. If the trust relationship does not match, copy the policy into the **Policy Document** window and choose **Update Trust Policy**.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "ecs-tasks.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}

```

Creating the task execution IAM role

If your account does not already have a task execution role, use the following steps to create the role.

To create the `ecsTaskExecutionRole` IAM role

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**, **Create role**.
3. In the **Select type of trusted entity** section, choose **Elastic Container Service**.
4. For **Select your use case**, choose **Elastic Container Service Task**, then choose **Next: Permissions**.
5. In the **Attach permissions policy** section, search for **AmazonECSTaskExecutionRolePolicy**, select the policy, and then choose **Next: Review**.
6. For **Role Name**, type `ecsTaskExecutionRole` and choose **Create role**.

Required IAM Permissions for Private Registry Authentication

The Amazon ECS task execution role is required to use the private registry authentication feature. This allows the container agent to pull the container image. For more information, see [Private Registry Authentication for Tasks](#) (p. 85).

To provide access to the secrets that you create, manually add the following permissions as an inline policy to the task execution role. For more information, see [Adding and Removing IAM Policies](#).

- `secretsmanager:GetSecretValue`
- `kms:Decrypt`—Required only if your key uses a custom KMS key and not the default key. The ARN for your custom key should be added as a resource.

An example inline policy adding the permissions is shown below.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:Decrypt",
        "secretsmanager:GetSecretValue"
      ],
      "Resource": [
        "arn:aws:secretsmanager:<region>:<aws_account_id>:secret:secret_name",
        "arn:aws:kms:<region>:<aws_account_id>:key/key_id"
      ]
    }
  ]
}
```

Required IAM Permissions for Amazon ECS Secrets

To use the Amazon ECS secrets feature, you must have the Amazon ECS task execution role and reference it in your task definition. This allows the container agent to pull the necessary AWS Systems Manager or Secrets Manager resources. For more information, see [Specifying Sensitive Data](#) (p. 87).

To provide access to the AWS Systems Manager Parameter Store parameters that you create, manually add the following permissions as an inline policy to the task execution role. For more information, see [Adding and Removing IAM Policies](#).

- `ssm:GetParameters`—Required if you are referencing a Systems Manager Parameter Store parameter in a task definition.
- `secretsmanager:GetSecretValue`—Required if you are referencing a Secrets Manager secret either directly or if your Systems Manager Parameter Store parameter is referencing a Secrets Manager secret in a task definition.
- `kms:Decrypt`—Required only if your secret uses a custom KMS key and not the default key. The ARN for your custom key should be added as a resource.

The following example inline policy adds the required permissions:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "secretsmanager:GetSecretValue",
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:secretsmanager:<region>:<aws_account_id>:secret:<secret_name>",
        "arn:aws:kms:<region>:<aws_account_id>:key/<key_id>"
      ]
    }
  ]
}
```

Optional IAM Permissions for Fargate Tasks Pulling Amazon ECR Images over Interface Endpoints

When launching tasks that use the Fargate launch type that pull images from Amazon ECR when Amazon ECR is configured to use an interface VPC endpoint, you can restrict the tasks access to a specific VPC or VPC endpoint. Do this by creating a task execution role for the tasks to use that use IAM condition keys.

Use the following IAM global condition keys to restrict access to a specific VPC or VPC endpoint. For more information, see [AWS Global Condition Context Keys](#).

- `aws:SourceVpc`—Restricts access to a specific VPC.
- `aws:SourceVpce`—Restricts access to a specific VPC endpoint.

The following task execution role policy provides an example for adding condition keys:

Important

The `ecr:GetAuthorizationToken` API action cannot have the `aws:sourceVpc` or `aws:sourceVpce` condition keys applied to it because the `GetAuthorizationToken` API call goes through the elastic network interface owned by AWS Fargate rather than the elastic network interface of the task.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecr:GetAuthorizationToken",
        "logs:CreateLogStream",

```

```

        "logs:PutLogEvents"
    ],
    "Resource": "*"
  },
  {
    "Effect": "Allow",
    "Action": [
      "ecr:BatchCheckLayerAvailability",
      "ecr:GetDownloadUrlForLayer",
      "ecr:BatchGetImage"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "aws:sourceVpce": "vpce-xxxxxx",
        "aws:sourceVpc": "vpc-xxxxxx"
      }
    }
  }
]
}

```

IAM Roles for Tasks

With IAM roles for Amazon ECS tasks, you can specify an IAM role that can be used by the containers in a task. Applications must sign their AWS API requests with AWS credentials, and this feature provides a strategy for managing credentials for your applications to use, similar to the way that Amazon EC2 instance profiles provide credentials to EC2 instances. Instead of creating and distributing your AWS credentials to the containers or using the EC2 instance's role, you can associate an IAM role with an ECS task definition or RunTask API operation. The applications in the task's containers can then use the AWS SDK or CLI to make API requests to authorized AWS services.

You define the IAM role to use in your task definitions, or you can use a `taskRoleArn` override when running a task manually with the RunTask API operation. The Amazon ECS agent receives a payload message for starting the task with additional fields that contain the role credentials. The Amazon ECS agent sets a unique task credential ID as an identification token and updates its internal credential cache so that the identification token for the task points to the role credentials that are received in the payload. The Amazon ECS agent populates the `AWS_CONTAINER_CREDENTIALS_RELATIVE_URI` environment variable in the `Env` object (available with the `docker inspect container_id` command) for all containers that belong to this task with the following relative URI: `/credential_provider_version/credentials?id=task_credential_id`.

Note

When you specify an IAM role for a task, the AWS CLI or other SDKs in the containers for that task use the AWS credentials provided by the task role exclusively and they no longer inherit any IAM permissions from the container instance.

From inside the container, you can query the credentials with the following command:

```
curl 169.254.170.2$AWS_CONTAINER_CREDENTIALS_RELATIVE_URI
```

Output:

```

{
  "AccessKeyId": "ACCESS_KEY_ID",
  "Expiration": "EXPIRATION_DATE",
  "RoleArn": "TASK_ROLE_ARN",
  "SecretAccessKey": "SECRET_ACCESS_KEY",

```



```
}  "Token": "SECURITY_TOKEN_STRING"
```

Note

The default expiration time for the generated IAM role credentials is 6 hours.

Topics

- [Benefits of Using IAM Roles for Tasks \(p. 241\)](#)
- [Creating an IAM Role and Policy for your Tasks \(p. 241\)](#)
- [Using a Supported AWS SDK \(p. 243\)](#)
- [Specifying an IAM Role for your Tasks \(p. 243\)](#)

Benefits of Using IAM Roles for Tasks

- **Credential Isolation:** A container can only retrieve credentials for the IAM role that is defined in the task definition to which it belongs; a container never has access to credentials that are intended for another container that belongs to another task.
- **Authorization:** Unauthorized containers cannot access IAM role credentials defined for other tasks.
- **Auditability:** Access and event logging is available through CloudTrail to ensure retrospective auditing. Task credentials have a context of `taskArn` that is attached to the session, so CloudTrail logs show which task is using which role.

Creating an IAM Role and Policy for your Tasks

You must create an IAM policy for your tasks to use that specifies the permissions that you would like the containers in your tasks to have. You have several ways to create a new IAM permission policy. You can copy a complete AWS managed policy that already does some of what you're looking for and then customize it to your specific requirements. For more information, see [Creating a New Policy](#) in the *IAM User Guide*.

You must also create a role for your tasks to use before you can specify it in your task definitions. You can create the role using the **Amazon Elastic Container Service Task Role** service role in the IAM console. Then you can attach your specific IAM policy to the role that gives the containers in your task the permissions you desire. The procedures below describe how to do this.

If you have multiple task definitions or services that require IAM permissions, you should consider creating a role for each specific task definition or service with the minimum required permissions for the tasks to operate so that you can minimize the access that you provide for each task.

The Amazon ECS Task Role trust relationship is shown below.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "ecs-tasks.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

To create an IAM policy for your tasks

In this example, we create a policy to allow read-only access to an Amazon S3 bucket. You could store database credentials or other secrets in this bucket, and the containers in your task can read the credentials from the bucket and load them into your application.

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Policies** and then choose **Create policy**.
3. Follow the steps under one of the following tabs, which shows you how to use the visual or JSON editors.

Using the visual editor

1. For **Service**, choose **S3**.
2. For **Actions**, expand the **Read** option and select **GetObject**.
3. For **Resources**, select **Add ARN** and enter the full Amazon Resource Name (ARN) of your Amazon S3 bucket, and then choose **Review policy**.
4. On the **Review policy** page, for **Name** type your own unique name, such as `AmazonECSTaskS3BucketPolicy`.
5. Choose **Create policy** to finish.

Using the JSON editor

1. In the **Policy Document** field, paste the policy to apply to your tasks. The example below allows permission to the `my-task-secrets-bucket` Amazon S3 bucket. You can modify the policy document to suit your specific needs.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject"
      ],
      "Resource": [
        "arn:aws:s3:::my-task-secrets-bucket/*"
      ]
    }
  ]
}
```

2. Choose **Create policy**.

To create an IAM role for your tasks

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**, **Create role**.
3. For **Select type of trusted entity** section, choose **AWS service**.
4. For **Choose the service that will use this role**, choose **Elastic Container Service**.
5. For **Select your use case**, choose **Elastic Container Service Task** and choose **Next: Permissions**.
6. For **Attach permissions policy**, select the policy to use for your tasks (in this example `AmazonECSTaskS3BucketPolicy`, and then choose **Next: Tags**.

7. For **Add tags (optional)**, enter any metadata tags you want to associate with the IAM role, and then choose **Next: Review**.
8. For **Role name**, enter a name for your role. For this example, type `AmazonECSTaskS3BucketRole` to name the role, and then choose **Create role** to finish.

Using a Supported AWS SDK

Support for IAM roles for tasks was added to the AWS SDKs on July 13th, 2016. The containers in your tasks must use an AWS SDK version that was created on or after that date. AWS SDKs that are included in Linux distribution package managers may not be new enough to support this feature.

To ensure that you are using a supported SDK, follow the installation instructions for your preferred SDK at [Tools for Amazon Web Services](#) when you are building your containers to get the latest version.

Specifying an IAM Role for your Tasks

After you have created a role and attached a policy to that role, you can run tasks that assume the role. You have several options to do this:

- Specify an IAM role for your tasks in the task definition. You can create a new task definition or a new revision of an existing task definition and specify the role you created previously. If you use the console to create your task definition, choose your IAM role in the **Task Role** field. If you use the AWS CLI or SDKs, specify your task role ARN using the `taskRoleArn` parameter. For more information, see [Creating a Task Definition \(p. 30\)](#).

Note

This option is required if you want to use IAM task roles in an Amazon ECS service.

- Specify an IAM task role override when running a task. You can specify an IAM task role override when running a task. If you use the console to run your task, choose **Advanced Options** and then choose your IAM role in the **Task Role** field. If you use the AWS CLI or SDKs, specify your task role ARN using the `taskRoleArn` parameter in the `overrides` JSON object. For more information, see [Running Tasks \(p. 109\)](#).

Note

In addition to the standard Amazon ECS permissions required to run tasks and services, IAM users also require `iam:PassRole` permissions to use IAM roles for tasks.

Amazon ECS CodeDeploy IAM Role

Before you can use the CodeDeploy blue/green deployment type with Amazon ECS, the CodeDeploy service needs permissions to update your Amazon ECS service on your behalf. These permissions are provided by the CodeDeploy IAM role (`ecsCodeDeployRole`).

Note

IAM users also require permissions to use CodeDeploy; these permissions are described in [Blue/Green Deployment Required IAM Permissions \(p. 146\)](#).

There are two managed policies provided. The `AWSCodeDeployRoleForECS` policy, shown below, gives CodeDeploy permission to update any resource using the associated action.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "ecs:DescribeServices",
```

```

        "ecs:CreateTaskSet",
        "ecs:UpdateServicePrimaryTaskSet",
        "ecs>DeleteTaskSet",
        "elasticloadbalancing:DescribeTargetGroups",
        "elasticloadbalancing:DescribeListeners",
        "elasticloadbalancing:ModifyListener",
        "elasticloadbalancing:DescribeRules",
        "elasticloadbalancing:ModifyRule",
        "lambda:InvokeFunction",
        "cloudwatch:DescribeAlarms",
        "sns:Publish",
        "s3:GetObject",
        "s3:GetObjectMetadata",
        "s3:GetObjectVersion"
    ],
    "Resource": "*",
    "Effect": "Allow"
}
]
}

```

The `AWSCodeDeployRoleForECSLimited` policy, shown below, gives CodeDeploy more limited permissions.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "ecs:DescribeServices",
        "ecs:CreateTaskSet",
        "ecs:UpdateServicePrimaryTaskSet",
        "ecs>DeleteTaskSet",
        "cloudwatch:DescribeAlarms"
      ],
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "sns:Publish"
      ],
      "Resource": "arn:aws:sns:*:*:CodeDeployTopic_*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "elasticloadbalancing:DescribeTargetGroups",
        "elasticloadbalancing:DescribeListeners",
        "elasticloadbalancing:ModifyListener",
        "elasticloadbalancing:DescribeRules",
        "elasticloadbalancing:ModifyRule"
      ],
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "lambda:InvokeFunction"
      ],
      "Resource": "arn:aws:lambda:*:*:function:CodeDeployHook_*",
      "Effect": "Allow"
    }
  ]
}

```

```
        "Action": [
            "s3:GetObject",
            "s3:GetObjectMetadata",
            "s3:GetObjectVersion"
        ],
        "Resource": "*",
        "Condition": {
            "StringEquals": {
                "s3:ExistingObjectTag/UseWithCodeDeploy": "true"
            }
        },
        "Effect": "Allow"
    }
}
```

To create an IAM role for CodeDeploy

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**, **Create role**.
3. For **Select type of trusted entity** section, choose **AWS service**.
4. For **Choose the service that will use this role**, choose **CodeDeploy**.
5. For **Select your use case**, choose **CodeDeploy - ECS**, **Next: Permissions**.
6. Choose **Next: Tags**.
7. For **Add tags (optional)**, you can add optional IAM tags to the role. Choose **Next: Review** when finished.
8. For **Role name**, type `ecsCodeDeployRole`, enter an optional description, and then choose **Create role**.

To add the required permissions to the Amazon ECS CodeDeploy IAM role

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. Search the list of roles for `ecsCodeDeployRole`. If the role does not exist, use the procedure above to create the role. If the role does exist, select the role to view the attached policies.
3. In the **Permissions policies** section, ensure that either the **AWSCodeDeployRoleForECS** or **AWSCodeDeployRoleForECSLimited** managed policy is attached to the role. If the policy is attached, your Amazon ECS CodeDeploy service role is properly configured. If not, follow the substeps below to attach the policy.
 - a. Choose **Attach policies**.
 - b. To narrow the available policies to attach, for **Filter**, type **AWSCodeDeployRoleForECS** or **AWSCodeDeployRoleForECSLimited**.
 - c. Check the box to the left of the AWS managed policy and choose **Attach policy**.
4. Choose **Trust Relationships**, **Edit trust relationship**.
5. Verify that the trust relationship contains the following policy. If the trust relationship matches the policy below, choose **Cancel**. If the trust relationship does not match, copy the policy into the **Policy Document** window and choose **Update Trust Policy**.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
```

```

    "Principal": {
      "Service": [
        "codedeploy.amazonaws.com"
      ]
    },
    "Action": "sts:AssumeRole"
  }
]
}

```

6. If the tasks in your Amazon ECS service using the blue/green deployment type require the use of the task execution role or a task role override, then you must add the `iam:PassRole` permission for each task execution role or task role override to the CodeDeploy IAM role as an inline policy. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#) and [IAM Roles for Tasks \(p. 240\)](#).

Follow the substeps below to create an inline policy.

- a. Open the IAM console at <https://console.aws.amazon.com/iam/>.
- b. Search the list of roles for `ecsCodeDeployRole`. If the role does not exist, use the procedure above to create the role. If the role does exist, select the role to view the attached policies.
- c. In the **Permissions policies** section, choose **Add inline policy**.
- d. Choose the **JSON** tab and add the following policy text.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:PassRole",
      "Resource": [
        "arn:aws:iam::<aws_account_id>:role/
<ecsTaskExecutionRole_or_TaskRole_name>"
      ]
    }
  ]
}

```

Note

Specify the full ARN of your task execution role or task role override.

- e. Choose **Review policy**
- f. For **Name**, type a name for the added policy and then choose **Create policy**.

Amazon ECS CloudWatch Events IAM Role

Before you can use Amazon ECS scheduled tasks with CloudWatch Events rules and targets, the CloudWatch Events service needs permissions to run Amazon ECS tasks on your behalf. These permissions are provided by the CloudWatch Events IAM role (`ecsEventsRole`).

The CloudWatch Events role is automatically created for you in the AWS Management Console when you configure a scheduled task. For more information, see [Scheduled Tasks \(cron\) \(p. 111\)](#).

The `AmazonEC2ContainerServiceEventsRole` policy is shown below.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {

```

```

        "Effect": "Allow",
        "Action": [
            "ecs:RunTask"
        ],
        "Resource": [
            "*"
        ]
    },
    {
        "Effect": "Allow",
        "Action": "iam:PassRole",
        "Resource": [
            "*"
        ],
        "Condition": {
            "StringLike": {
                "iam:PassedToService": "ecs-tasks.amazonaws.com"
            }
        }
    }
]
}

```

If your scheduled tasks require the use of the task execution role, a task role, or a task role override, then you must add `iam:PassRole` permissions for each task execution role, task role, or task role override to the CloudWatch Events IAM role. For more information about the task execution role, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

Note

Specify the full ARN of your task execution role or task role override.

```

{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": [
                "arn:aws:iam::<aws_account_id>:role/
<ecsTaskExecutionRole_or_TaskRole_name>"
            ]
        }
    ]
}

```

You can use the following procedure to check that your account already has the CloudWatch Events IAM role, and manually create it if needed.

To check for the CloudWatch Events IAM role in the IAM console

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**.
3. Search the list of roles for `ecsEventsRole`. If the role does not exist, use the next procedure to create the role. If the role does exist, select the role to view the attached policies.
4. Choose **Permissions**.
5. In the **Permissions policies** section, ensure that the **AmazonEC2ContainerServiceEventsRole** managed policy is attached to the role. If the policy is attached, your Amazon ECS service role is properly configured. If not, follow the substeps below to attach the policy.
 - a. Choose **Attach policies**.

- b. To narrow the available policies to attach, for **Filter**, type `AmazonEC2ContainerServiceEventsRole`.
 - c. Select the box to the left of the **AmazonEC2ContainerServiceEventsRole** policy and choose **Attach policy**.
6. Choose **Trust relationships**, **Edit trust relationship**.
7. Verify that the trust relationship contains the following policy. If the trust relationship matches the policy below, choose **Cancel**. If the trust relationship does not match, copy the policy into the **Policy Document** window and choose **Update Trust Policy**.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "events.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

To create an IAM role for CloudWatch Events

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles** and then choose **Create role**.
3. In the **Select type of trusted entity** section, choose **Elastic Container Service**. For **Select your use case** choose **Elastic Container Service Task**. Choose **Next: Permissions**.
4. In the **Attach permissions policy** section, select the **AmazonEC2ContainerServiceEventsRole** policy and choose **Next: Tags**.
5. In the **Add tags (optional)** section, enter any tags you would like to associate with the role and choose **Next: Review**.
6. For **Role name**, type `ecsEventsRole` to name the role, optionally enter a description, and then choose **Create role**.
7. Review your role information and choose **Create Role**.
8. Search the list of roles for `ecsEventsRole` and select the role you just created.
9. Choose **Trust relationships**, **Edit trust relationship**.
10. Replace the existing trust relationship with the following text in the **Policy Document** window and choose **Update Trust Policy**.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "events.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```


To add permissions for the task execution role to the CloudWatch Events IAM role

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Policies, Create policy**.
3. Choose **JSON**, paste the following policy, and then choose **Review policy**:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:PassRole",
      "Resource": [
        "arn:aws:iam::<aws_account_id>:role/
        <ecsTaskExecutionRole_or_TaskRole_name>"
      ]
    }
  ]
}
```

4. For **Name**, type `AmazonECSEventsTaskExecutionRole`, optionally enter a description, and then choose **Create policy**.
5. In the navigation pane, choose **Roles**.
6. Search the list of roles for `ecsEventsRole` and select the role to view the attached policies.
7. Choose **Attach policy**.
8. In the **Attach policy** section, select the **AmazonECSEventsTaskExecutionRole** policy and choose **Attach policy**.

Troubleshooting Amazon Elastic Container Service Identity and Access

Use the following information to help you diagnose and fix common issues that you might encounter when working with Amazon ECS and IAM.

Topics

- [I Am Not Authorized to Perform an Action in Amazon ECS \(p. 249\)](#)
- [I Am Not Authorized to Perform iam:PassRole \(p. 250\)](#)
- [I Want to View My Access Keys \(p. 250\)](#)
- [I'm an Administrator and Want to Allow Others to Access Amazon ECS \(p. 250\)](#)
- [I Want to Allow People Outside of My AWS Account to Access My Amazon ECS Resources \(p. 250\)](#)

I Am Not Authorized to Perform an Action in Amazon ECS

If the AWS Management Console tells you that you're not authorized to perform an action, then you must contact your administrator for assistance. Your administrator is the person that provided you with your user name and password.

The following example error occurs when the `mateojackson` IAM user tries to use the console to view details about a `widget` but does not have `ecs:GetWidget` permissions.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
ecs:GetWidget on resource: my-example-widget
```

In this case, Mateo asks his administrator to update his policies to allow him to access the `my-example-widget` resource using the `ecs:GetWidget` action.

I Am Not Authorized to Perform iam:PassRole

If you receive an error that you're not authorized to perform the `iam:PassRole` action, then you must contact your administrator for assistance. Your administrator is the person that provided you with your user name and password. Ask that person to update your policies to allow you to pass a role to Amazon ECS.

Some AWS services allow you to pass an existing role to that service, instead of creating a new service role or service-linked role. To do this, you must have permissions to pass the role to the service.

The following example error occurs when an IAM user named `marymajor` tries to use the console to perform an action in Amazon ECS. However, the action requires the service to have permissions granted by a service role. Mary does not have permissions to pass the role to the service.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform: iam:PassRole
```

In this case, Mary asks her administrator to update her policies to allow her to perform the `iam:PassRole` action.

I Want to View My Access Keys

After you create your IAM user access keys, you can view your access key ID at any time. However, you can't view your secret access key again. If you lose your secret key, you must create a new access key pair.

Access keys consist of two parts: an access key ID (for example, `AKIAIOSFODNN7EXAMPLE`) and a secret access key (for example, `wJalrXUtnFEMI/K7MDENG/bPxrFiCYEXAMPLEKEY`). Like a user name and password, you must use both the access key ID and secret access key together to authenticate your requests. Manage your access keys as securely as you do your user name and password.

Important

Do not provide your access keys to a third party, even to help [find your canonical user ID](#). By doing this, you might give someone permanent access to your account.

When you create an access key pair, you are prompted to save the access key ID and secret access key in a secure location. The secret access key is available only at the time you create it. If you lose your secret access key, you must add new access keys to your IAM user. You can have a maximum of two access keys. If you already have two, you must delete one key pair before creating a new one. To view instructions, see [Managing Access Keys](#) in the *IAM User Guide*.

I'm an Administrator and Want to Allow Others to Access Amazon ECS

To allow others to access Amazon ECS, you must create an IAM entity (user or role) for the person or application that needs access. They will use the credentials for that entity to access AWS. You must then attach a policy to the entity that grants them the correct permissions in Amazon ECS.

To get started right away, see [Creating Your First IAM Delegated User and Group](#) in the *IAM User Guide*.

I Want to Allow People Outside of My AWS Account to Access My Amazon ECS Resources

You can create a role that users in other accounts or people outside of your organization can use to access your resources. You can specify who is trusted to assume the role. For services that support

resource-based policies or access control lists (ACLs), you can use those policies to grant people access to your resources.

To learn more, consult the following:

- To learn whether Amazon ECS supports these features, see [How Amazon Elastic Container Service Works with IAM](#) (p. 206).
- To learn how to provide access to your resources across AWS accounts that you own, see [Providing Access to an IAM User in Another AWS Account That You Own](#) in the *IAM User Guide*.
- To learn how to provide access to your resources to third-party AWS accounts, see [Providing Access to AWS Accounts Owned by Third Parties](#) in the *IAM User Guide*.
- To learn how to provide access through identity federation, see [Providing Access to Externally Authenticated Users \(Identity Federation\)](#) in the *IAM User Guide*.
- To learn the difference between using roles and resource-based policies for cross-account access, see [How IAM Roles Differ from Resource-based Policies](#) in the *IAM User Guide*.

Logging and Monitoring in Amazon Elastic Container Service

Monitoring is an important part of maintaining the reliability, availability, and performance of Amazon Elastic Container Service and your AWS solutions. You should collect monitoring data from all of the parts of your AWS solution so that you can more easily debug a multi-point failure if one occurs. AWS provides several tools for monitoring your Amazon ECS resources and responding to potential incidents:

Amazon CloudWatch Alarms

Watch a single metric over a time period that you specify, and perform one or more actions based on the value of the metric relative to a given threshold over a number of time periods. The action is a notification sent to an Amazon Simple Notification Service (Amazon SNS) topic or Amazon EC2 Auto Scaling policy. CloudWatch alarms do not invoke actions simply because they are in a particular state; the state must have changed and been maintained for a specified number of periods. For more information, see [Amazon ECS CloudWatch Metrics](#) (p. 184).

For services with tasks that use the Fargate launch type, you can use CloudWatch alarms to scale in and scale out the tasks in your service based on CloudWatch metrics, such as CPU and memory utilization. For more information, see [Service Auto Scaling](#) (p. 165).

Amazon CloudWatch Logs

Monitor, store, and access the log files from the containers in your Amazon ECS tasks by specifying the `awslogs` log driver in your task definitions. This is the only supported method for accessing logs for tasks using the Fargate launch type. For more information, see [Using the awslogs Log Driver](#) (p. 69).

Amazon CloudWatch Events

Match events and route them to one or more target functions or streams to make changes, capture state information, and take corrective action. For more information, see [Amazon ECS Events and EventBridge](#) (p. 189) in this guide and [What Is Amazon CloudWatch Events?](#) in the *Amazon CloudWatch Events User Guide*.

AWS CloudTrail Logs

CloudTrail provides a record of actions taken by a user, role, or an AWS service in Amazon ECS. Using the information collected by CloudTrail, you can determine the request that was made to Amazon ECS, the IP address from which the request was made, who made the request, when it was

made, and additional details. For more information, see [Logging Amazon ECS API Calls with AWS CloudTrail \(p. 199\)](#).

AWS Trusted Advisor

Trusted Advisor draws upon best practices learned from serving hundreds of thousands of AWS customers. Trusted Advisor inspects your AWS environment and then makes recommendations when opportunities exist to save money, improve system availability and performance, or help close security gaps. All AWS customers have access to five Trusted Advisor checks. Customers with a Business or Enterprise support plan can view all Trusted Advisor checks.

For more information, see [AWS Trusted Advisor](#) in the *AWS Support User Guide*.

Another important part of monitoring Amazon ECS involves manually monitoring those items that the CloudWatch alarms don't cover. The CloudWatch, Trusted Advisor, and other AWS console dashboards provide an at-a-glance view of the state of your AWS environment. We recommend that you also check the log files on your container instances and the containers in your tasks.

Compliance Validation for Amazon Elastic Container Service

Third-party auditors assess the security and compliance of Amazon Elastic Container Service as part of multiple AWS compliance programs. These include SOC, PCI, FedRAMP, HIPAA, and others.

For a list of AWS services in scope of specific compliance programs, see [AWS Services in Scope by Compliance Program](#). For general information, see [AWS Compliance Programs](#).

You can download third-party audit reports using AWS Artifact. For more information, see [Downloading Reports in AWS Artifact](#).

Your compliance responsibility when using Amazon ECS is determined by the sensitivity of your data, your company's compliance objectives, and applicable laws and regulations. AWS provides the following resources to help with compliance:

- [Security and Compliance Quick Start Guides](#) – These deployment guides discuss architectural considerations and provide steps for deploying security- and compliance-focused baseline environments on AWS.
- [Architecting for HIPAA Security and Compliance Whitepaper](#) – This whitepaper describes how companies can use AWS to create HIPAA-compliant applications.
- [AWS Compliance Resources](#) – This collection of workbooks and guides might apply to your industry and location.
- [AWS Config](#) – This AWS service assesses how well your resource configurations comply with internal practices, industry guidelines, and regulations.
- [AWS Security Hub](#) – This AWS service provides a comprehensive view of your security state within AWS that helps you check your compliance with security industry standards and best practices.

Infrastructure Security in Amazon Elastic Container Service

As a managed service, Amazon Elastic Container Service is protected by the AWS global network security procedures that are described in the [Amazon Web Services: Overview of Security Processes](#) whitepaper.

You use AWS published API calls to access Amazon ECS through the network. Clients must support Transport Layer Security (TLS) 1.0 or later. We recommend TLS 1.2 or later. Clients must also support cipher suites with perfect forward secrecy (PFS) such as Ephemeral Diffie-Hellman (DHE) or Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). Most modern systems such as Java 7 and later support these modes.

Additionally, requests must be signed by using an access key ID and a secret access key that is associated with an IAM principal. Or you can use the [AWS Security Token Service](#) (AWS STS) to generate temporary security credentials to sign requests.

You can call these API operations from any network location, but Amazon ECS does support resource-based access policies, which can include restrictions based on the source IP address. You can also use Amazon ECS policies to control access from specific Amazon Virtual Private Cloud endpoints or specific VPCs. Effectively, this isolates network access to a given Amazon ECS resource from only the specific VPC within the AWS network. For more information, see [Amazon ECS interface VPC endpoints \(AWS PrivateLink\)](#) (p. 253).

Topics

- [Amazon ECS interface VPC endpoints \(AWS PrivateLink\)](#) (p. 253)

Amazon ECS interface VPC endpoints (AWS PrivateLink)

You can improve the security posture of your VPC by configuring Amazon ECS to use an interface VPC endpoint. Interface endpoints are powered by AWS PrivateLink, a technology that enables you to privately access Amazon ECS APIs by using private IP addresses. PrivateLink restricts all network traffic between your VPC and Amazon ECS to the Amazon network. You don't need an internet gateway, a NAT device, or a virtual private gateway.

For more information about AWS PrivateLink and VPC endpoints, see [VPC Endpoints](#) in the *Amazon VPC User Guide*.

Considerations for Amazon ECS VPC endpoints

Before you set up interface VPC endpoints for Amazon ECS, be aware of the following considerations:

- Tasks using the Fargate launch type don't require the interface VPC endpoints for Amazon ECS, but you might need interface VPC endpoints for Amazon ECR, Secrets Manager, or Amazon CloudWatch Logs described in the following points.
- To allow your tasks to pull private images from Amazon ECR, you must create the interface VPC endpoints for Amazon ECR. For more information, see [Interface VPC Endpoints \(AWS PrivateLink\)](#) in the *Amazon Elastic Container Registry User Guide*.

Important

If you configure Amazon ECR to use an interface VPC endpoint, you can create a task execution role that includes condition keys to restrict access to a specific VPC or VPC endpoint. For more information, see [Optional IAM Permissions for Fargate Tasks Pulling Amazon ECR Images over Interface Endpoints](#) (p. 239).

- To allow your tasks to pull sensitive data from Secrets Manager, you must create the interface VPC endpoints for Secrets Manager. For more information, see [Using Secrets Manager with VPC Endpoints](#) in the *AWS Secrets Manager User Guide*.
- If your VPC doesn't have an internet gateway and your tasks use the `awslogs` log driver to send log information to CloudWatch Logs, you must create an interface VPC endpoint for CloudWatch Logs. For more information, see [Using CloudWatch Logs with Interface VPC Endpoints](#) in the *Amazon CloudWatch Logs User Guide*.

- VPC endpoints currently don't support cross-Region requests. Ensure that you create your endpoint in the same Region where you plan to issue your API calls to Amazon ECS.
- VPC endpoints only support Amazon-provided DNS through Amazon Route 53. If you want to use your own DNS, you can use conditional DNS forwarding. For more information, see [DHCP Options Sets](#) in the *Amazon VPC User Guide*.
- The security group attached to the VPC endpoint must allow incoming connections on port 443 from the private subnet of the VPC.
- Controlling access to Amazon ECS by attaching an endpoint policy to the VPC endpoint isn't currently supported. By default, full access to the service will be allowed through the endpoint. For more information, see [Controlling Access to Services with VPC Endpoints](#) in the *Amazon VPC User Guide*.

Creating the VPC Endpoints for Amazon ECS

To create the VPC endpoint for the Amazon ECS service, use the [Creating an Interface Endpoint](#) procedure in the *Amazon VPC User Guide* to create the following endpoints. If you have existing container instances within your VPC, you should create the endpoints in the order that they're listed. If you plan on creating your container instances after your VPC endpoint is created, the order doesn't matter.

- `com.amazonaws.region.ecs-agent`
- `com.amazonaws.region.ecs-telemetry`
- `com.amazonaws.region.ecs`

Note

region represents the Region identifier for an AWS Region supported by Amazon ECS, such as `us-east-2` for the US East (Ohio) Region.

Create the Secrets Manager and Systems Manager endpoints

If you are referencing either Secrets Manager secrets or Systems Manager Parameter Store parameters in your task definitions to inject sensitive data into your containers, you need to create the interface VPC endpoints for Secrets Manager or Systems Manager so those tasks can reach those services. You only need to create the endpoints from the specific service your sensitive data is hosted in. For more information, see [Specifying Sensitive Data](#) (p. 87).

For more information about Secrets Manager VPC endpoints, see [Using Secrets Manager with VPC endpoints](#) in the *AWS Secrets Manager User Guide*.

For more information about Systems Manager VPC endpoints, see [Using Systems Manager with VPC endpoints](#) in the *AWS Systems Manager User Guide*.

Using the Amazon ECS Command Line Interface

The Amazon Elastic Container Service (Amazon ECS) command line interface (CLI) provides high-level commands to simplify creating, updating, and monitoring clusters and tasks from a local development environment. The Amazon ECS CLI supports Docker Compose files, a popular open-source specification for defining and running multi-container applications. Use the ECS CLI as part of your everyday development and testing cycle as an alternative to the AWS Management Console.

Important

At this time, the latest version of the Amazon ECS CLI only supports the major versions of [Docker Compose file syntax](#) versions 1, 2, and 3. The version specified in the compose file must be the string "1", "1.0", "2", "2.0", "3", or "3.0". Docker Compose minor versions are not supported.

The latest version of the Amazon ECS CLI is 1.17.0. For release notes, see [Changelog](#).

Note

The source code for the Amazon ECS CLI is [available on GitHub](#). We encourage you to submit pull requests for changes that you would like to have included. However, Amazon Web Services does not currently support running modified copies of this software.

Topics

- [Installing the Amazon ECS CLI \(p. 255\)](#)
- [Configuring the Amazon ECS CLI \(p. 261\)](#)
- [Migrating Configuration Files \(p. 262\)](#)
- [Tutorial: Creating a Cluster with a Fargate Task Using the Amazon ECS CLI \(p. 263\)](#)
- [Tutorial: Creating an Amazon ECS Service That Uses Service Discovery Using the Amazon ECS CLI \(p. 268\)](#)

Installing the Amazon ECS CLI

Follow these instructions to install the Amazon ECS CLI on your macOS, Linux, or Windows system.

Step 1: Download the Amazon ECS CLI

Download the Amazon ECS CLI binary.

- For macOS:

```
sudo curl -o /usr/local/bin/ecs-cli https://amazon-ecs-cli.s3.amazonaws.com/ecs-cli-darwin-amd64-latest
```

- For Linux systems:

```
sudo curl -o /usr/local/bin/ecs-cli https://amazon-ecs-cli.s3.amazonaws.com/ecs-cli-linux-amd64-latest
```

- For Windows systems:

Open Windows PowerShell and run the following commands:

```
PS C:\> New-Item -Path 'C:\Program Files\Amazon\ECSCLI' -ItemType  
Directory  
PS C:\> Invoke-WebRequest -OutFile 'C:\Program Files\Amazon\ECSCLI\ecs-cli.exe' https://  
amazon-ecs-cli.s3.amazonaws.com/ecs-cli-windows-amd64-latest.exe
```

Note

If you encounter permissions issues, ensure that you are running PowerShell as an administrator.

Step 2: Verify the Amazon ECS CLI

To verify the validity of the Amazon ECS CLI file, you can either use the provided MD5 sum or the PGP signatures. Both methods are described in the following sections.

Verify Using the MD5 Sum

Verify the downloaded binary with the MD5 sum provided.

- For macOS (compare the two output strings to verify that they match):

```
curl -s https://amazon-ecs-cli.s3.amazonaws.com/ecs-cli-darwin-amd64-latest.md5 && md5 -  
q /usr/local/bin/ecs-cli
```

- For Linux systems (look for an OK in the output string):

```
echo "$(curl -s https://amazon-ecs-cli.s3.amazonaws.com/ecs-cli-linux-amd64-latest.md5) /  
usr/local/bin/ecs-cli" | md5sum -c -
```

- For Windows systems:

Open Windows PowerShell and find the md5 hash of the executable that you downloaded:

```
PS C:\> Get-FileHash ecs-cli.exe -Algorithm MD5
```

Compare that with this md5 hash:

```
PS C:\> Invoke-WebRequest -OutFile md5.txt https://  
amazon-ecs-cli.s3.amazonaws.com/ecs-cli-windows-amd64-  
latest.md5  
PS C:\> Get-Content md5.txt
```

Verify Using the PGP Signature

The Amazon ECS CLI executables are cryptographically signed using PGP signatures. You can use the following steps to verify the signatures using the GnuPG tool.

- Download and install GnuPG. For more information, see the [GnuPG website](#).
 - For macOS, we recommend using Homebrew. Install Homebrew using the instructions from their website. For more information, see [Homebrew](#). After Homebrew is installed, use the following command from your macOS terminal:


```
brew install gnupg
```

- For Linux systems, install gpg using the package manager on your flavor of Linux.
 - For Windows systems, download and use the Windows simple installer from the GnuPG website. For more information, see [GnuPG Download](#).
2. Retrieve the Amazon ECS PGP public key. You can use a command to do this or manually create the key and then import it.
- a. Option 1: Retrieve the key with the following command.

```
gpg --keyserver hkp://keys.gnupg.net --recv BCE9D9A42D51784F
```

- b. Option 2: Create a file with the following contents of the Amazon ECS PGP public key and then import it:

```
-----BEGIN PGP PUBLIC KEY BLOCK-----
Version: GnuPG v2

mQINBfqlSasBEADliGcT1NVJ1ydfN8DqebYYe9ne3dt6jqKfMkOWLmm6LLGJe7HU
jGtqhCWRDkN+qPpHqdArRgDZAtn2pXY5fEipHgar4CP8QgRnRMO2f174lmavr4Vg
7K/KH8VHlq2uRw32/B94XLEgRbGTMdWfDKuxoPCttBQaMj3LGn6Pe+6xVWRkChQu
BoQAjhjBQ+bEmOkNy0LjNgjNlnL3UMAG56t8E3LANIgGgEnpNsB1UwFwluPoGZoTx
N+6pHBjRkIL/1v/ETU4FXpYw2zvhWNahxeNRnoYj3uycHkeliCrw4kj0+skizBgO
2K7oVX8Oc3j5+Zilhl/qDLXmUCb2az5cMM1mOoF8EKX5HaNuq1KfwJxqXE6NNiCO
lFTrT7QwD5fMNld3FanLgv/ZnIrsSaqJOL6zRSq8O4LN1OWBVbndExk2Kr+5kFxn
5lBPgfPgRj5hQ+KTHMa9Y8Z7yUc64BJiN6F9Nl7FJuSsfqbdkvRLsQRbcBG9qxX3
rJAEhieJzVMEUNl+EgeCkxj5xuSkNU7zw2c3hQZqEcrADLV+hvFJktOz9Gm6xxzbq
lTnWWCz4xrwIwteEBA2qE+MlDheVd78a3gIsEaStfQqOosYXaQbvlnSWOocly/5Zb
zizHTJIhLtUyIs9WisP2s0emeHZicVMfW61EgPrJAiupgc7kyZvFt4YwfwARAQAB
tCRBbWf6b24gRUNTIDx1Y3Mtc2VjdXJpdHlAYWlhem9uLmNvbT6JAhwEEAECAYF
AlrjL0YACgkQHivRXs0TaQrg1g/+JppwPqHn1VPmv7lessB8I5UqZeD6p6uVpHd7
Bs3pcPp8BV7BdRbs3sPLt5bV1+rkqOlw+0gZ4Q/ue/YbWtOAt4qY0OcEo0HgcnAX
lsB827QifZIVtGWMhuh94xzm/SJkvngml6KB3YJNWP61A9qJ37/VbVVLzvcmazA
McWB4HUMNrh0JgBCo0gIppCbpJEvUc02Bjn23eEJSs9kC7OUAHyQkVnx4d9UzXF
40oISF6hmQKIBoLnRrAlj5Qvs3GhvhQ0ThYq0Grk/KMJJX2CSqt7tWJ8gk1n3H3Y
SRERXJRnv7DsDDBwFgT6r5Q2HW1TBUvaoZy5hF6maD09nHcNnvBjQADzeT8Tr/Qu
bBCLZkNSYqqkpgtwv7seoD2P4n1giRvDAOEFmZpVkuR+C252IaH1HZFEZ+TvBVQM
Y8OWWxmIJW+J6evjo3N1e019UHv71jvoF8zljB14bsL2c+QTJmOv7nRqzDQgCWyp
Id/v2dUVVTk1j9omuLBBwNjzQCB+72LcIzJhYmaP1HC4LcKQG+/f4lexuItenatK
lEQuhYtyVXcBlh6Yn/wzNg2NWOwb3vqY/F7m6u9ixAwgtIMgPCDE4aJ86zrrXYFz
N2HqkTSQh77Z8KPKmyGopsmN/reMuilPdInb249nA0dzoN+nj+ttFOYCIaLaFyjs
Z0r1QA0JAjkEEwECACMFAlq1SasCGwMHcWkIBwMCAQYVCAIJGcsEFgIDAQIeAQIX
gAAKCRc86dmkLVF4T9iFEACEnkmlDNXsWUx34R3c0vamHrPxvfyI1fLEUen8D1h
uX9xy6jCEROHWEp0rjGK4QDPgM93sWJ+s1UAKg214QRVzft0y9/DdR+twApA0fzy
uavIthGd6+03jAAo6udYDE+cZC3P7XBbDiYEWk4XAF9I1JjB8htZUgVXBL046JhG
eM17+crgUyQeetkiOQemLbsbXQ40Bd9V7zf7XJraFd8VrwNUwNb+9KftgAsc9rk+
YIT/Pef+YOPysgcxI4sTWghTyCulVnuGoskgDv4v73PALU0ieUrvvQVqWMrvhVx1
0X90J7c1KOyhlEQQlaFTgmQjmXexVTwIBm8LvysFK6YXM41KjOrlz3+6xBIm/qe
bFyLUnf4WoIUplAaJhK9pRY+XENGNxdN4D26Kd0F+PLkm3Tr3Hy3b10k34FlGr
KVHUq1TZD7cvMnnKEELTUcKX+1mV3an16nmAg/my1JSUt6BNK2rJpY1s/kkSGSE
XQ4zuF2IGCpVBfHYAlt5Un5zwqkwQ3/n2kwAoDzonJcehDw/C/cGos5D0aIU7I
K2X2aTD3+pA7Mx3ImE2hqmYqRt9X42yF1PIEVRneBRJ3HDezAgJrNh0GQWRQkhIx
gz6/cTR+ekr5TptVsZS9few2GpI5bCgBKBisZIssT89aw7mAKWutOGcm4qM9/yK6
1bkCDQRatUmrARAAxNPvVwreJ2yAiFcUpdRlVhsuOgnxvs1QgsIw3H7+Pacr9Hpe
8uftYzQdC82KeSKhpHq7c8gMTMucIINTH25x9BCC73E33EjCL9Lqov1TL7+QkgHe
T+JihZwdD8Mx2K+LVVVu/aWkNrfMuNwyDUciSI4D5QHa8T+F8fgN40TpWYjirzel
5yoICMr9hVcbzDNv/ozKCxjx+XKgnFc3wrnDfJfntfDAT7ecwbUTL+viQKJ646s+
psiqXRYtVvYInEhLvrJ0aV6zHfOigE/Bils6/g7ru1Q6CEHqEw++APs5CcE8VzJu
WAGSVHZgun5Y9N4quR/M9Vm+IPMhTxxAg7rOvyRN9cAXfeSMf77I+XTifignNa8x
t/MDjXr1fjF4pThEi5u6WsuRdFwjY2azEv3vevodTi4HoJReH6dFRA6y8c+UDg1
2iHiOKIpQqLbHEfQmHcDd2fix+AaJKMnPGNku9qCFEMbgSRJpXz6BfwnY1QuKE+I
```

Amazon ECS User Guide for AWS Fargate
Step 2: Verify the Amazon ECS CLI

```
R6jA0frUnt2jhiGG/F8RceXzohaaC/Cx7LUCUFwc0n7z32C9/Dtj7I1PMOacdZzz
bJzRK0/ZDv+UN/c9dwAk1lZAYPMwGBkUaY68EBstnIliW34aWm6IiHhxioVPK5p
VJfyiXPO0EXqujtHLAeChfjcns3I12YshTldv2PafG53fp33ZdzeUgsBo+EAEQEA
AYkCHWQYAQIACUCWrvJqWtBDAKCRc86dmkLVF4T+ZdD/9x/8APzgNJF3o3STrF
jvnV1ycyhWYGAeBJiu7wjsNWwzMF0v15tLjB7AqeVxZn+WKDD/mIOQ450ZvnYZuy
X7DR0JszaH9wrYTxZLVruAu+t6UL0y/XQ4L1GZ9QR6+r+7t1Mvbfy7B1HbvX/gYt
Rwe/uwdibI0CagEzyX+2D3kT0lHO5XThbXaNf8AN8zha91Jt2Q2UR2X5T6JcwtMz
FBvZn13LSmZyE0EQehS2iUurU4uWOpGppuqVnb10jbCvCHKgDGrqZ0smKNAQng54
F365W3g8AfY48s8XQwzmcliowYX9bT8PziEi0J4QmQh0aXkpqZyFefuWeOL2R94S
XKzr+gRh3BAULoqF+qK+IUMxTip9KTPNVdPciC66yBiT6gFDji5Ca9pGpJXrC3xe
TXiKQ8DBWDhBPVPrRuLiaenTtZE0sPc4I85yt5U9RoPTStcOr34s3w5yEaJagt6S
Gc5r9ysjKfH6+6rbiluJxMgROSqtqr+RyB+V9A5/OgtNZc8l1K6u4UoOC3e8jUuW
vqWKvjJB/Kz3u4zaeNu2ZyyHaOqOuH+TETcW+jsY9IhbEzqN5yQYGi4pVmDkY5vu
lXbJnbqPKpRXgM9BecV9AMBpGbdQ/5LnHJJXg+G8YQOgp4lR/hC1TEFdIp5wM8AK
CWSENYt2o1rjgMXiZOMF8A5oBLkCDQRatUuSARAAr77kjj2QR2SZeOSlFBvV7oS
mFeSNnz9xZssqrsm6bTwSHM6YLDwc7Sdf2esDdyzONETwqrVCg+FxgL8hmo9hS4c
rR6tmrPomOmptr+XLsKcaP7ogIXsyZnrEAesvW8PnfayoiPCdc3cMCR/1tNHFGA
7EuR/XLBmi7Qg9tByVYQ5Yj5wB9V4B2yeCt3XtzPqeLKvaxl7PNelaHGJQY/xo+m
V0bndxf9IY+4oFJ4bLD32WqvyxES07vW6WBh7oqv3Zbm0yQrr8a6mDBpqLkvWwNI
3kpJR974tg5o5LfDu1BeeyHWP5Gm4U/G4JB+JIG1Ady+RmoWEt4BqTCZ/knnoGvw
D5sTCxbKdmuOmHgyTssog+30OcGYHV7pWYPHaxKHPm201xKCjH1RfzRULzGKjD+
ymTLT1I3AXFmLZJXikA0lvD3/wgMqCXschybcLjLD/bXlUfW03rzoeezXjgi/DJx
jKBAyBTY05nMcth109oaFd9d0HbsOUDkIMnsgGBE766Piro6MHo0T0rXl07T4pI
rwuSOsc6XzCzdImj0Wc6axS/HeUKRXWdXJwno5awTwXKRJMXGfhCvSvbcb2Wx+L
IKvmb7EB4K3fmjFFE67yolmiw2qRCUBfygtH3eL5XZU28MiCpue8Y8GKJoBAUyvf
KeMlrO8Jm3iRac5a/D0AEQEAAyKEPgQYAQIACUCWrvLkgIbAgIpCrc86dmkLVF4
T8FDIAQZAQIABgUCWrvLkgAKCRDePL1hra+LjtHYD/9MucxdFe6bX01dQR4tKhHq
POLRqy6z1BY9ILCLowNdGZdqorogUiUymgn3VhEhVtxTOoHcN7qOuM01PNsRnOeS
EYj7f8Xrb1clzkD6xULwmOclTb9bBxnBc/4PFvHABZW3QzusaZniNgkuxt6BTfLoS
Of4inq71kjmGK+TlzQ6mUMUQUG228NUQC+a84EPqYyAeY1sgvgB7hJBhYLOQAxhcW
6m20Rd8iEc6HyZJ3yCOCsKip/nRWAbf0OvfHfRBP0+m0ZwnJM8cPRFjOqqzFpKH9
HpDmTrC4wKP1+TL52LyEqNh4yZitXmZNV7giSRlkk0eDsko+bFy6VbMzKUMKJ3
D3eHFAMkujmbfJmSMTJOPGn5SB1HyjCZNX6bhIbQyEUB9gKCMUFaqXKwKpF6rj0
iQXAJxLR/shZ5Rk96VxzOphU17T90m/PnUEEPwq8KsBhnMRGxa0RFidDP+n9fgtv
HLmrOQX9zBCVXh0mdWYLrWvmzQFWzG7AoE55fkf8nAEPsalrCdtaNUBHRXA0OQXG
AHMOdJQqvBsmqMvuAdjkdWpFu5y0My5ddU+hiUzUyQLjL5Hhd5LOUDdewLZgIwlj
xrEAUzDKetnemM8GkHxDgg8koev5frmShJuce7vSJpCNg3EIJSGqMOPFJuLMTWz
vjHeDNbJy6uNL65ckJy6WhGjEADS2WAW1D6Tfekkc21SsIXk/LqEpLMR/og5OUif
wcEN1rS9IJXBWly8Me1N9qr5KcKQLmfdBNEyyceBhyVl0MDyHOK+7PoFmktGBg
13QierHv5GJ8LB3fclqHV8pwTt03Bc8z2g0TjmUYAN/ixETdReDoKavWJYSE9yom
aaJu279ioVTrwPEcse0XkiRyKtoTjwOb73CGkBZZpJyqux/rmCV/fp4ALdSW8zbz
FJvORaihvowwzjpfQKhwcU9LABXi2UvVm14v0AfeI7oiJPSU1zm4fEny4oiIBXLR
zhFNih1UjIu82X16mTm3BwbIga/s1fnQRGzyhquIMii+mWra23EwjChaxpvjjcUH
5ilLc5Zq781aCYRygYQw+hu5nFkOH1R+Z50Ubxjd/aqufngIAX7kPMD3LoF4K1dD
Q8ppQriUvXVo+4nPV6rPTy/PyqCLWDjkguHpJSEfSMkwaJrAz0QNSAU5CJ0G2Zu4
yxvYlumHCEl7nbFrm0vIia75Sa8KnywTdsyZsu3XcOcf3g+glxWtpjJqy2bYXlq
9uDOwTArWHOis6bq819RE6hxr1RBVXS6uqgQIZFBGyq66b0dIq4D2JdsUvqEMAHbc
e7tBfeB1CMBdA64e9Rq7bFR7Tvt8gasCZYlNr3lydh+dFHIeKH53HzQe6188HEic
+0jVnLkCDQRa55wJARAAYLya2Lx6gyoWoJN1a6740q3o8e9d4KggQOfGMTcflmeq
ivuzgN+3DZHN+9ty2KxXmtn0mhHBERZdbNjyMNT1gAgrhPNB4HtXBxum2wS57WK
DNmade914L7FWTPAWBG2Wn448OEHTqsClICXXWy9IICgclAEyIq0Yq5mAdTEgrJS
Z8t4GpwtDL9gNQyFXaWQmDmkAsCygQMvhAlmu9x0IzQG5CxsNzFk7zcuL60k14Z3
Cmt49k4T/7ZU8goWi8tt+rU78/IL3J/ff9+1civ1OwuUidgfPCsvOUW1JojsdCQA
L+RZJcoXq71fOFj/eNjeOSstCTDPfTCL+kThe6E5neDtbQHBYkEX1BRiTedSv4+M
ucgiTrdQFWKf89G72xdv8ut9AYYQ2BbeYU+JAYhUH8rYYui2dHKJIGjNvJscUWb
+QEqJIRleJRhrO+/CHgMs4fZakWF1VFhKBkcKmeJLn1f7EJJUW84ZhKXjO/AUPX
1CHSnjziRceUJCJYox1cwsom6jTE50GiNzcIxTn9xUC0UMKFeggNAFys1K+LWTm3
Bzo8H5ucjCUEmUm9lhkGwqTzgOlRX5eqPX+JBoSaObqhgqCa5IPinKRa6MgoFPHK
6sYKqroYwBgGZm6Js5chpNchvJMs/3WXNOEVg0J3z3vP0DMhxqWm+r+n9zLW8qsA
EQEAAyKEPgQYAQgACUCWuecCQIbAgIpCrc86dmkLVF4T8FDIAQZAQgABgUCWuec
CQAKCRBQ3szEcQ5hr+ykD/4tOLRHFHXuKucxgGaubUcVtsFrwBKmalcyJqaPms8u
6SkOwfGRI32G/GhOrp0Ts/MOKbObq6VLTh8N5Yc/53ME18zQFw9Y5AmRoW4PZXEU
uj5s7p4oR7xHMihMjCCBnlbvrR+34YPfgzTcgLiOEFHYT8UTxwnGmXOVnKMM7md
xD3CV5q6VAte8WKBo/220I13fcQlc9r/oWX4kXKkb0v9h0GwKbDJ1tzqTPrp/xFt
yohqnvImpnlz+Q9zXmbrWYL9/g8VCmW/NN2gju2G3Lu/TlFUWIT4v/50PK6TdeNb
VKJO4+S8bTayqSG9CML1S57KSGCo5HUHQWeSNHI+fpe5oX6FALPT9JLDce8OZz1i
```

```
cZzOMELP37mOOQun0AlmHm/hVzf0f311PtzbzcqWaE51tJvgUR/nZFO6Ta3O5Ezhs
3VlEJNQ1Ijff/6DH87SxvAoRIARCuZd0qxBCDK0avpFzUtbJd24lRA3WJpkEiMqKv
RDVZkE4b6TW61f0o+LaVfK6E8oLpixegS4fiqC16mFrOdyRk+RJJfIUyzOWTDVmt
g0U1CO1ezokMSqk7724pyjr2xf/r9/sC6aOJwB/lKgZkJfC6NqL7TlxVA31dUga
LEOVejTTE4gl+YtfsCDvALCtqL0jduSkUo+RXcBItmXhA+tShW0pbS2Rtx/ixua
KohVD/0R4QxiSwQmICNtm9mw9ydIllyjYXX5a9x4wMJracNY/LBybJPFnZnT4dYR
z4XjQysDwvVYZByaWoIe3QxjX84V6MlI2IdAT/xImu8gbacI8tmyfpIrLnPKiR9D
VFYfGBXuAX7+HgPPSFtrHQONCALxxzlbNpS+zxt9r0MiLgcLyspWxSdmoYGZ6nQP
RO5Nm/ZVS+u2imPCRzNUZEMa+dLE6kHxOrS0dPiuJ407NtPeYDKkoQtNagspsDvh
cK7CSqAiKMq06UBTxqlTSRkm62eOCtcs3p3OeHu5GRZF1uzTET0ZxYkaPgdrQknx
ozjp5mC7X+45lcCfmcVt94TFNL5HwEUVJpmOgmzILCI8yoDTWzloo+i+fPFsXX4f
kynhE83mSEcr5VHFYrTY3mQXGmNJ3bCLuc/jq7ysGq69xiKmTlUeXfm+aojCR05i
zyShIRJZ0GZfuzDYFDbMV9amA/YQGygLw//zP5ju5SW26dNxlf3MdFQE5J86rn9
MgZ4gcpazHEVUusbZsgkLizRp9imUiH8ymLqAXnfRGLU/LpNSefnvDFTtEIRcpOHc
bhayG0bk51Bd4mioOXnIsKy4j63nJXA27x5EVVHQ1sYRN8Ny4Fdr2tMamj20+X+J
qx2yy/UX5nSPU492e2CdZ1UhoU0SRFY3bxKHKb7SdbVeav+K5g==
=Gi5D
-----END PGP PUBLIC KEY BLOCK-----
```

The details of the Amazon ECS PGP public key for reference:

```
Key ID: BCE9D9A42D51784F
Type: RSA
Size: 4096/4096
Expires: Never
User ID: Amazon ECS
Key fingerprint: F34C 3DDA E729 26B0 79BE AEC6 BCE9 D9A4 2D51 784F
```

Import the Amazon ECS PGP public key with the following command.

```
gpg --import <public_key_filename>
```

3. Download the Amazon ECS CLI signatures. The signatures are ASCII detached PGP signatures stored in files with the extension `.asc`. The signatures file has the same name as its corresponding executable, with `.asc` appended.

- For macOS systems:

```
curl -o ecs-cli.asc https://amazon-ecs-cli.s3.amazonaws.com/ecs-cli-darwin-amd64-latest.asc
```

- For Linux systems:

```
curl -o ecs-cli.asc https://amazon-ecs-cli.s3.amazonaws.com/ecs-cli-linux-amd64-latest.asc
```

- For Windows systems:

```
PS C:\> Invoke-WebRequest -OutFile ecs-cli.asc https://amazon-ecs-cli.s3.amazonaws.com/ecs-cli-windows-amd64-latest.exe.asc
```

4. Verify the signature.

- For macOS and Linux systems:

```
gpg --verify ecs-cli.asc /usr/local/bin/ecs-cli
```

- For Windows systems:

```
PS C:\> gpg --verify ecs-cli.asc 'C:\Program Files\Amazon\ECSCLI\ecs-cli.exe'
```

Expected output:

```
gpg: Signature made Tue Apr  3 13:29:30 2018 PDT
gpg:                using RSA key DE3CBD61ADAF8B8E
gpg: Good signature from "Amazon ECS <ecs-security@amazon.com>" [unknown]
gpg: WARNING: This key is not certified with a trusted signature!
gpg:                There is no indication that the signature belongs to the owner.
Primary key fingerprint: F34C 3DDA E729 26B0 79BE  AEC6 BCE9 D9A4 2D51 784F
Subkey fingerprint: EB3D F841 E2C9 212A 2BD4  2232 DE3C BD61 ADAF 8B8E
```

Important

The warning in the output is expected and is not problematic. It occurs because there is not a chain of trust between your personal PGP key (if you have one) and the Amazon ECS PGP key. For more information, see [Web of trust](#).

Step 3: Apply Execute Permissions to the Binary

Apply execute permissions to the binary.

- For macOS and Linux systems:

```
sudo chmod +x /usr/local/bin/ecs-cli
```

- For Windows systems:

Edit the environment variables and add C:\Program Files\Amazon\ECSCLI to the PATH variable field, separated from existing entries by using a semicolon. For example:

```
PS C:\> C:\existing\path;C:\Program Files\Amazon\ECSCLI
```

Restart PowerShell (or the command prompt) so the changes go into effect.

Note

Once the PATH variable is set, the Amazon ECS CLI can be used from either Windows PowerShell or the command prompt.

Step 4: Complete the Installation

Verify that the CLI is working properly.

```
ecs-cli --version
```

Proceed to [Configuring the Amazon ECS CLI \(p. 261\)](#).

Important

You must configure the Amazon ECS CLI with your AWS credentials, an AWS region, and an Amazon ECS cluster name before you can use it.

Configuring the Amazon ECS CLI

The Amazon ECS CLI requires some basic configuration information before you can use it, such as your AWS credentials, the AWS Region in which to create your cluster, and the name of the Amazon ECS cluster to use. Configuration information is stored in the `~/.ecs` directory on macOS and Linux systems and in `C:\Users\<username>\AppData\local\ecs` on Windows systems.

To configure the Amazon ECS CLI

1. Set up a CLI profile with the following command, substituting *profile_name* with your desired profile name, *\$AWS_ACCESS_KEY_ID* and *\$AWS_SECRET_ACCESS_KEY* environment variables with your AWS credentials.

```
ecs-cli configure profile --profile-name profile_name --access-key $AWS_ACCESS_KEY_ID
--secret-key $AWS_SECRET_ACCESS_KEY
```

2. Complete the configuration with the following command, substituting *launch_type* with the task launch type you want to use by default, *region_name* with your desired AWS region, *cluster_name* with the name of an existing Amazon ECS cluster or a new cluster to use, and *configuration_name* for the name you'd like to give this configuration.

```
ecs-cli configure --cluster cluster_name --default-launch-type launch_type --
region region_name --config-name configuration_name
```

After you have installed and configured the CLI, you can try the [Tutorial: Creating a Cluster with a Fargate Task Using the Amazon ECS CLI \(p. 263\)](#). For more information, see the [Amazon ECS Command Line Reference](#) in the *Amazon Elastic Container Service Developer Guide*.

Profiles

The Amazon ECS CLI supports the configuring of multiple sets of AWS credentials as named *profiles* using the **ecs-cli configure profile** command. A default profile can be set by using the **ecs-cli configure profile default** command. These profiles can then be referenced when you run Amazon ECS CLI commands that require credentials using the `--ecs-profile` flag otherwise the default profile is used.

For more information, see the [Amazon ECS Command Line Reference](#) in the *Amazon Elastic Container Service Developer Guide*.

Cluster Configurations

A cluster configuration is a set of fields that describes an Amazon ECS cluster including the name of the cluster and the region. A default cluster configuration can be set by using the **ecs-cli configure default** command. The Amazon ECS CLI supports the configuring of multiple named cluster configurations using the `--config-name` option.

For more information, see the [Amazon ECS Command Line Reference](#) in the *Amazon Elastic Container Service Developer Guide*.

Order of Precedence

There are multiple methods for passing both the credentials and the region in an Amazon ECS CLI command. The following is the order of precedence for each of these.

The order of precedence for credentials is:

1. Amazon ECS CLI profile flags:
 - a. ECS profile (`--ecs-profile`)
 - b. AWS profile (`--aws-profile`)
2. Environment variables:
 - a. `ECS_PROFILE`
 - b. `AWS_PROFILE`
 - c. `AWS_ACCESS_KEY_ID`, `AWS_SECRET_ACCESS_KEY`, and `AWS_SESSION_TOKEN`
3. ECS config-attempts to fetch credentials from the default ECS profile.
4. Default AWS profile—Attempts to use credentials (`aws_access_key_id`, `aws_secret_access_key`) or `assume_role` (`role_arn`, `source_profile`) from the AWS profile name.
 - a. `AWS_DEFAULT_PROFILE` environment variable (defaults to default).
5. EC2 instance role

The order of precedence for Region is:

1. Amazon ECS CLI flags:
 - a. Region flag (`--region`)
 - b. Cluster config flag (`--cluster-config`)
2. ECS config-attempts to fetch the Region from the default ECS profile.
3. Environment variables—Attempts to fetch the region from the following environment variables:
 - a. `AWS_REGION`
 - b. `AWS_DEFAULT_REGION`
4. AWS profile-attempts to use the region from the AWS profile name:
 - a. `AWS_PROFILE` environment variable
 - b. `AWS_DEFAULT_PROFILE` environment variable (defaults to default)

Migrating Configuration Files

The process of configuring the Amazon ECS CLI has changed significantly in the latest version (v1.0.0) to allow the addition of new features. A migration command has been introduced that converts an older (v0.6.6 and older) configuration file to the current format. The old configuration files are deprecated, so we recommend converting your configuration to the newest format to take advantage of the new features. The configuration-related changes and new features introduced in v1.0.0 in the new YAML formatted configuration files include:

- Splitting up of credential and cluster-related configuration information into two separate files. Credential information is stored in `~/.ecs/credentials` and cluster configuration information is stored in `~/.ecs/config`.
- The configuration files are formatted in YAML.
- Support for storing multiple named configurations.
- Deprecation of the field `compose-service-name-prefix` (name used for creating a service `<compose_service_name_prefix> + <project_name>`). This field can still be configured. However, if it is not configured, there is no longer a default value assigned. For Amazon ECS CLI v0.6.6 and earlier, the default was `ecscompose-service-`.
- Removal of the field `compose-project-name-prefix` (name used for creating a task definition `<compose_project_name_prefix> + <project_name>`). Amazon ECS CLI v1.0.0 and later can still read old configuration files; so if this field is present then it is still read and used. However,

configuring this field is not supported in v1.0.0+ with the `ecs-cli configure` command, and if the field is manually added to a v1.0.0+ configuration file it causes the Amazon ECS CLI to throw an error.

- The field `cfn-stack-name-prefix` (name used for creating CFN stacks `<cfn_stack_name_prefix> + <cluster_name>`) has been changed to `cfn-stack-name`. Instead of specifying a prefix, the exact name of a CloudFormation template can be configured.
- Amazon ECS CLI v0.6.6 and earlier allowed configuring credentials using a named AWS profile from the `~/.aws/credentials` file on your system. This functionality has been removed. However, a new flag, `--aws-profile`, has been added which allows the referencing of an AWS profile inline in all commands that require credentials.

Note

The `--project-name` flag can be used to set the project name.

Migrating Older Configuration Files to the v1.0.0+ Format

While all versions of the Amazon ECS CLI support reading from the older configuration file format, upgrading to the new format is required to take advantage of some new features, for example using multiple named cluster profiles. Migrating your legacy configuration file to the new format is easy with the `ecs-cli configure migrate` command. The command takes the configuration information stored in the old format in `~/.ecs/config` and converts it to a pair of files in the new format, overwriting your old configuration file in the process.

When running the `ecs-cli configure migrate` command there is a warning message displayed with the old configuration file, and a preview of the new configuration files. User confirmation is required before the migration proceeds. If the `--force` flag is used, then the warning message is not displayed, and the migration proceeds without any confirmation. If `cfn-stack-name-prefix` is used in the legacy file, then `cfn-stack-name` is stored in the new file as `<cfn_stack_name_prefix> + <cluster_name>`.

For more information, see the [Amazon ECS Command Line Reference](#) in the *Amazon Elastic Container Service Developer Guide*.

Tutorial: Creating a Cluster with a Fargate Task Using the Amazon ECS CLI

This tutorial shows you how to set up a cluster and deploy a service with tasks using the Fargate launch type.

Prerequisites

Complete the following prerequisites:

- Set up an AWS account.
- Install the Amazon ECS CLI. For more information, see [Installing the Amazon ECS CLI \(p. 255\)](#).
- Install and configure the AWS CLI. For more information, see [AWS Command Line Interface](#).

Step 1: Create the Task Execution IAM Role

The Amazon ECS container agent makes calls to AWS APIs on your behalf, so it requires an IAM policy and role for the service to know that the agent belongs to you. This IAM role is referred to as a task execution IAM role. If you already have a task execution role created to use, you can skip this step. For more information, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

To create the task execution IAM role using the AWS CLI

1. Create a file named `task-execution-assume-role.json` with the following contents:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "",
      "Effect": "Allow",
      "Principal": {
        "Service": "ecs-tasks.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

2. Create the task execution role:

```
aws iam --region us-west-2 create-role --role-name ecsTaskExecutionRole --assume-role-policy-document file://task-execution-assume-role.json
```

3. Attach the task execution role policy:

```
aws iam --region us-west-2 attach-role-policy --role-name ecsTaskExecutionRole --policy-arn arn:aws:iam::aws:policy/service-role/AmazonECSTaskExecutionRolePolicy
```

Step 2: Configure the Amazon ECS CLI

The Amazon ECS CLI requires credentials in order to make API requests on your behalf. It can pull credentials from environment variables, an AWS profile, or an Amazon ECS profile. For more information, see [Configuring the Amazon ECS CLI \(p. 261\)](#).

To create an Amazon ECS CLI configuration

1. Create a cluster configuration, which defines the AWS region to use, resource creation prefixes, and the cluster name to use with the Amazon ECS CLI:

```
ecs-cli configure --cluster tutorial --default-launch-type FARGATE --config-name tutorial --region us-west-2
```

2. Create a CLI profile using your access key and secret key:

```
ecs-cli configure profile --access-key AWS_ACCESS_KEY_ID --secret-key AWS_SECRET_ACCESS_KEY --profile-name tutorial-profile
```


Step 3: Create a Cluster and Configure the Security Group

To create an ECS cluster and security group

1. Create an Amazon ECS cluster with the **ecs-cli up** command. Because you specified Fargate as your default launch type in the cluster configuration, this command creates an empty cluster and a VPC configured with two public subnets.

```
ecs-cli up --cluster-config tutorial --ecs-profile tutorial-profile
```

This command may take a few minutes to complete as your resources are created. The output of this command contains the VPC and subnet IDs that are created. Take note of these IDs as they are used later.

2. Using the AWS CLI, retrieve the default security group ID for the VPC. Use the VPC ID from the previous output:

```
aws ec2 describe-security-groups --filters Name=vpc-id,Values=VPC_ID --region us-west-2
```

The output of this command contains your security group ID, which is used in the next step.

3. Using AWS CLI, add a security group rule to allow inbound access on port 80:

```
aws ec2 authorize-security-group-ingress --group-id security_group_id --protocol tcp --port 80 --cidr 0.0.0.0/0 --region us-west-2
```

Step 4: Create a Compose File

For this step, create a simple Docker compose file that creates a simple PHP web application. At this time, the Amazon ECS CLI supports [Docker compose file syntax](#) versions 1, 2, and 3. This tutorial uses Docker compose v3.

Here is the compose file, which you can name `docker-compose.yml`. The web container exposes port 80 for inbound traffic to the web server. It also configures container logs to go to the CloudWatch log group created earlier. This is the recommended best practice for Fargate tasks.

```
version: '3'
services:
  web:
    image: amazon/amazon-ecs-sample
    ports:
      - "80:80"
    logging:
      driver: awslogs
      options:
        awslogs-group: tutorial
        awslogs-region: us-west-2
        awslogs-stream-prefix: web
```

Note

If your account already contains a CloudWatch Logs log group named `tutorial` in the `us-west-2` Region, choose a unique name so the ECS CLI creates a new log group for this tutorial.

In addition to the Docker compose information, there are some parameters specific to Amazon ECS that you must specify for the service. Using the VPC, subnet, and security group IDs from the previous step, create a file named `ecs-params.yml` with the following content:

```
version: 1
task_definition:
  task_execution_role: ecsTaskExecutionRole
  ecs_network_mode: awsvpc
  task_size:
    mem_limit: 0.5GB
    cpu_limit: 256
run_params:
  network_configuration:
    awsvpc_configuration:
      subnets:
        - "subnet ID 1"
        - "subnet ID 2"
      security_groups:
        - "security group ID"
  assign_public_ip: ENABLED
```

Step 5: Deploy the Compose File to a Cluster

After you create the compose file, you can deploy it to your cluster with **ecs-cli compose service up**. By default, the command looks for files called `docker-compose.yml` and `ecs-params.yml` in the current directory; you can specify a different docker compose file with the `--file` option, and a different ECS Params file with the `--ecs-params` option. By default, the resources created by this command have the current directory in their titles, but you can override that with the `--project-name` option. The `--create-log-groups` option creates the CloudWatch log groups for the container logs.

```
ecs-cli compose --project-name tutorial service up --create-log-groups --cluster-
config tutorial --ecs-profile tutorial-profile
```

Step 6: View the Running Containers on a Cluster

After you deploy the compose file, you can view the containers that are running in the service with **ecs-cli compose service ps**.

```
ecs-cli compose --project-name tutorial service ps --cluster-config tutorial --ecs-
profile tutorial-profile
```

Output:

Name	State	Ports
TaskDefinition Health		
tutorial/0c2862e6e39e4eff92ca3e4f843c5b9a/web	RUNNING	34.222.202.55:80->80/tcp
tutorial:1	UNKNOWN	

In the above example, you can see the web container from your compose file, and also the IP address and port of the web server. If you point your web browser at that address, you should see the PHP web application. Also in the output is the `task-id` value for the container. Copy the task ID as you use it in the next step.

Step 7: View the Container Logs

View the logs for the task:

```
ecs-cli logs --task-id 0c2862e6e39e4eff92ca3e4f843c5b9a --follow --cluster-config tutorial
--ecs-profile tutorial-profile
```

Note

The `--follow` option tells the Amazon ECS CLI to continuously poll for logs.

Step 8: Scale the Tasks on the Cluster

You can scale up your task count to increase the number of instances of your application with `ecs-cli compose service scale`. In this example, the running count of the application is increased to two.

```
ecs-cli compose --project-name tutorial service scale 2 --cluster-config tutorial --ecs-
profile tutorial-profile
```

Now you should see two more containers in your cluster:

```
ecs-cli compose --project-name tutorial service ps --cluster-config tutorial --ecs-
profile tutorial-profile
```

Output:

Name	TaskDefinition	Health	State	Ports
tutorial/0c2862e6e39e4eff92ca3e4f843c5b9a/web	tutorial:1	UNKNOWN	RUNNING	34.222.202.55:80->80/tcp
tutorial/d9fbbc931d2e47ae928fcf433041648f/web	tutorial:1	UNKNOWN	RUNNING	34.220.230.191:80->80/tcp

Step 9: View your Web Application

Enter the IP address for the task in your web browser and you should see a webpage that displays the **Simple PHP App** web application.

Simple PHP App

Congratulations

Your PHP application is now running on a container in Amazon ECS.

The container is running PHP version 5.3.10-1ubuntu3.15.

Step 10: Clean Up

When you are done with this tutorial, you should clean up your resources so they do not incur any more charges. First, delete the service so that it stops the existing containers and does not try to run any more tasks.

```
ecs-cli compose --project-name tutorial service down --cluster-config tutorial --ecs-
profile tutorial-profile
```

Now, take down your cluster, which cleans up the resources that you created earlier with `ecs-cli up`.

```
ecs-cli down --force --cluster-config tutorial --ecs-profile tutorial-profile
```

Tutorial: Creating an Amazon ECS Service That Uses Service Discovery Using the Amazon ECS CLI

This tutorial shows a simple walkthrough of creating an Amazon ECS service that is configured to use service discovery. Many of the service discovery configuration values can be specified with either the ECS parameters file or flags. When flags are used, they take precedence over the ECS parameters file if both are present. When using the Amazon ECS CLI, the compose project name is used as the name for your ECS service.

Prerequisites

It is expected that you have completed the following prerequisites before continuing on:

- Set up an AWS account.
- Install the Amazon ECS CLI. For more information, see [Installing the Amazon ECS CLI \(p. 255\)](#).

Configure the Amazon ECS CLI

Before you can start this tutorial, you must install and configure the Amazon ECS CLI. For more information, see [Installing the Amazon ECS CLI \(p. 255\)](#).

The Amazon ECS CLI requires credentials in order to make API requests on your behalf. It can pull credentials from environment variables, an AWS profile, or an Amazon ECS profile. For more information, see [Configuring the Amazon ECS CLI \(p. 261\)](#).

To create an Amazon ECS CLI configuration

1. Create a cluster configuration:

```
ecs-cli configure --cluster ec2-tutorial --region us-east-1 --default-launch-type EC2  
--config-name ec2-tutorial
```

2. Create a profile using your access key and secret key:

```
ecs-cli configure profile --access-key AWS_ACCESS_KEY_ID --secret-  
key AWS_SECRET_ACCESS_KEY --profile-name ec2-tutorial
```

Note

If this is the first time that you are configuring the Amazon ECS CLI, these configurations are marked as default. If this is not your first time configuring the Amazon ECS CLI, see the [Amazon ECS Command Line Reference](#) in the *Amazon Elastic Container Service Developer Guide* to set this as the default configuration and profile.

Create an Amazon ECS Service Configured to Use Service Discovery

Use the following steps to create an Amazon ECS service that is configured to use service discovery with the Amazon ECS CLI.

To create an Amazon ECS service configured to use service discovery

1. Create an Amazon ECS service named `backend` and create a private DNS namespace named `tutorial` within a VPC. In this example, the task is using the `awsvpc` network mode, so the `container_name` and `container_port` values are not required.

```
ecs-cli compose --project-name backend service up --private-dns-namespace tutorial --  
vpc vpc-04deee8176dce7d7d --enable-service-discovery
```

Output:

```
INFO[0001] Using ECS task definition                      TaskDefinition="backend:1"  
INFO[0002] Waiting for the private DNS namespace to be created...  
INFO[0002] Cloudformation stack status                  stackStatus=CREATE_IN_PROGRESS  
WARN[0033] Defaulting DNS Type to A because network mode was awsvpc  
INFO[0033] Waiting for the Service Discovery Service to be created...  
INFO[0034] Cloudformation stack status                  stackStatus=CREATE_IN_PROGRESS  
INFO[0065] Created an ECS service                      service=backend  
taskDefinition="backend:1"  
INFO[0066] Updated ECS service successfully             desiredCount=1  
serviceName=backend  
INFO[0081] (service backend) has started 1 tasks: (task 824b5a76-8f9c-4beb-  
a64b-6904e320630e). timestamp="2018-09-12 00:00:26 +0000 UTC"  
INFO[0157] Service status                             desiredCount=1 runningCount=1  
serviceName=backend  
INFO[0157] ECS Service has reached a stable state       desiredCount=1 runningCount=1  
serviceName=backend
```

2. Create another service named `frontend` in the same private DNS namespace. Because the namespace already exists, the Amazon ECS CLI uses it instead of creating a new one.

```
ecs-cli compose --project-name frontend service up --private-dns-namespace tutorial --  
vpc vpc-04deee8176dce7d7d --enable-service-discovery
```

Output:

```
INFO[0001] Using ECS task definition                      TaskDefinition="frontend:1"  
INFO[0002] Using existing namespace ns-kvhnzhhb5vxplfmls  
WARN[0033] Defaulting DNS Type to A because network mode was awsvpc  
INFO[0033] Waiting for the Service Discovery Service to be created...  
INFO[0034] Cloudformation stack status                  stackStatus=CREATE_IN_PROGRESS  
INFO[0065] Created an ECS service                      service=frontend  
taskDefinition="frontend:1"  
INFO[0066] Updated ECS service successfully             desiredCount=1  
serviceName=frontend  
INFO[0081] (service frontend) has started 1 tasks: (task 824b5a76-8f9c-4beb-  
a64b-6904e320630e). timestamp="2018-09-12 00:00:26 +0000 UTC"  
INFO[0157] Service status                             desiredCount=1 runningCount=1  
serviceName=frontend  
INFO[0157] ECS Service has reached a stable state       desiredCount=1 runningCount=1  
serviceName=frontend
```

3. Verify that the two services are able to discover each other within the VPC using DNS. The DNS hostname uses the following format:
`<service_discovery_service_name>.<service_discovery_namespace>`. For this example, the `frontend` service can be discovered at `frontend.tutorial` and the `backend` service can be discovered at `backend.tutorial`. Because these are private DNS namespaces, these DNS names only resolve when within the specified VPC.

4. To update the service discovery settings, update the settings for the `frontend` service. The values that can be updated are the DNS TTL and the value for the health check custom config failure threshold.

```
ecs-cli compose --project-name frontend service up --update-service-discovery --dns-type SRV --dns-ttl 120 --healthcheck-custom-config-failure-threshold 2
```

Output:

```
INFO[0001] Using ECS task definition                TaskDefinition="frontend:1"
INFO[0001] Updated ECS service successfully         desiredCount=1
  serviceName=frontend
INFO[0001] Service status                         desiredCount=1 runningCount=1
  serviceName=frontend
INFO[0001] ECS Service has reached a stable state   desiredCount=1 runningCount=1
  serviceName=frontend
INFO[0002] Waiting for your Service Discovery resources to be updated...
INFO[0002] Cloudformation stack status            stackStatus=UPDATE_IN_PROGRESS
```

5. To clean up, delete the Amazon ECS service and the service discovery resources. When the `frontend` service is deleted, the Amazon ECS CLI automatically removes the associated service discovery service.

```
ecs-cli compose --project-name frontend service rm
```

```
INFO[0000] Updated ECS service successfully         desiredCount=0
  serviceName=frontend
INFO[0001] Service status                         desiredCount=0 runningCount=1
  serviceName=frontend
INFO[0016] Service status                         desiredCount=0 runningCount=0
  serviceName=frontend
INFO[0016] (service frontend) has stopped 1 running tasks: (task 824b5a76-8f9c-4beba64b-6904e320630e). timestamp="2018-09-12 00:37:25 +0000 UTC"
INFO[0016] ECS Service has reached a stable state   desiredCount=0 runningCount=0
  serviceName=frontend
INFO[0016] Deleted ECS service                     service=frontend
INFO[0016] ECS Service has reached a stable state   desiredCount=0 runningCount=0
  serviceName=frontend
INFO[0027] Waiting for your Service Discovery Service resource to be deleted...
INFO[0027] Cloudformation stack status            stackStatus=DELETE_IN_PROGRESS
```

6. To complete the cleanup, delete the `backend` service along with the private DNS namespace that was created with it. The Amazon ECS CLI associates the AWS CloudFormation stack for the private DNS namespace with the Amazon ECS service for which it was created. When the service is deleted, the namespace is also deleted.

```
ecs-cli compose --project-name backend service rm --delete-namespace
```

Amazon ECS task metadata endpoint

Amazon ECS on Fargate provides a method to retrieve various metadata, network metrics, and [Docker stats](#) about your tasks and containers. This is referred to as the task metadata endpoint. The following task metadata endpoint versions are available for Fargate tasks:

- Task metadata endpoint version 4 – Available for tasks that use the Fargate launch type on platform version 1.4.0 or later.
- Task metadata endpoint version 3 – Available for tasks that use the Fargate launch type on platform version 1.1.0 or later.

All containers belonging to tasks that are launched with the `awsvpc` network mode receive a local IPv4 address within a predefined link-local address range. When a container queries the metadata endpoint, the container agent can determine which task the container belongs to based on its unique IP address, and metadata and stats for that task are returned.

Topics

- [Task metadata endpoint version 4](#) (p. 271)
- [Task metadata endpoint version 3](#) (p. 279)

Task metadata endpoint version 4

Beginning with Fargate platform version 1.4.0, an environment variable named `ECS_CONTAINER_METADATA_URI_V4` is injected into each container in a task. When you query the task metadata version 4 endpoint, various task metadata and [Docker stats](#) are available to tasks.

The task metadata version 4 endpoint functions like the version 3 endpoint but provides additional network metadata for your containers and tasks. Additional network metrics are available when querying the `/stats` endpoints as well.

Important

To avoid the need to create new task metadata endpoint versions in the future, additional metadata may be added to the version 4 output. We will not remove any existing metadata or change the metadata field names.

The following additional network metadata is included when querying the task metadata version 4 endpoint:

- `AttachmentIndex`
- `IPV4SubnetCIDRBlock`
- `MACAddress`
- `PrivateDNSName`
- `SubnetGatewayIPv4Address`
- `DomainNameServers`
- `DomainNameSearchList`

Enabling Task Metadata

The task metadata endpoint version 4 feature is enabled by default for tasks that use the Fargate launch type on platform version 1.4 or later.

Task Metadata Endpoint version 4 Paths

The following task metadata endpoints are available to containers:

`${ECS_CONTAINER_METADATA_URI_V4}`

This path returns metadata JSON for the container.

`${ECS_CONTAINER_METADATA_URI_V4}/task`

This path returns metadata JSON for the task, including a list of the container IDs and names for all of the containers associated with the task. For more information about the response for this endpoint, see [Task Metadata JSON Response \(p. 272\)](#).

`${ECS_CONTAINER_METADATA_URI_V4}/stats`

This path returns Docker stats JSON for the specific Docker container. For more information about each of the returned stats, see [ContainerStats](#) in the Docker API documentation.

`${ECS_CONTAINER_METADATA_URI_V4}/task/stats`

This path returns Docker stats JSON for all of the containers associated with the task. For more information about each of the returned stats, see [ContainerStats](#) in the Docker API documentation.

Task Metadata JSON Response

The following information is returned from the task metadata endpoint (`${ECS_CONTAINER_METADATA_URI_V4}/task`) JSON response.

Cluster

The full Amazon Resource Name (ARN) of the Amazon ECS cluster to which the task belongs.

TaskARN

The full Amazon Resource Name (ARN) of the task to which the container belongs.

Family

The family of the Amazon ECS task definition for the task.

Revision

The revision of the Amazon ECS task definition for the task.

DesiredStatus

The desired status for the task from Amazon ECS.

KnownStatus

The known status for the task from Amazon ECS.

Limits

The resource limits specified at the task level (such as CPU and memory). This parameter is omitted if no resource limits are defined.

PullStartedAt

The timestamp for when the first container image pull began.

PullStoppedAt

The timestamp for when the last container image pull finished.

AvailabilityZone

The Availability Zone the task is in.

Note

The Availability Zone metadata is only available for Fargate tasks using platform version 1.4 or later.

Containers

A list of container metadata for each container associated with the task.

DockerId

The Docker ID for the container.

Name

The name of the container as specified in the task definition.

DockerName

The name of the container supplied to Docker. The Amazon ECS container agent generates a unique name for the container to avoid name collisions when multiple copies of the same task definition are run on a single instance.

Image

The image for the container.

ImageID

The SHA-256 digest for the image.

Ports

Any ports exposed for the container. This parameter is omitted if there are no exposed ports.

Labels

Any labels applied to the container. This parameter is omitted if there are no labels applied.

DesiredStatus

The desired status for the container from Amazon ECS.

KnownStatus

The known status for the container from Amazon ECS.

ExitCode

The exit code for the container. This parameter is omitted if the container has not exited.

Limits

The resource limits specified at the container level (such as CPU and memory). This parameter is omitted if no resource limits are defined.

CreatedAt

The time stamp for when the container was created. This parameter is omitted if the container has not been created yet.

StartedAt

The time stamp for when the container started. This parameter is omitted if the container has not started yet.

FinishedAt

The time stamp for when the container stopped. This parameter is omitted if the container has not stopped yet.

Type

The type of the container. Containers that are specified in your task definition are of type `NORMAL`. You can ignore other container types, which are used for internal task resource provisioning by the Amazon ECS container agent.

Networks

The network information for the container, such as the network mode and IP address. This parameter is omitted if no network information is defined.

ExecutionStoppedAt

The time stamp for when the tasks `DesiredStatus` moved to `STOPPED`. This occurs when an essential container moves to `STOPPED`.

Examples

The following examples show sample outputs from the task metadata endpoints.

Example Container Metadata Response

When querying the `#{ECS_CONTAINER_METADATA_URI_V4}` endpoint you are returned only metadata about the container itself. The following is an example output.

```
{
  "DockerId": "c7a6b9b237934e9999f319ea3ccc9da4query-metadata",
  "Name": "query-metadata",
  "DockerName": "query-metadata",
  "Image": "mreferre/eksutils",
  "ImageID": "sha256:1b146e73f801617610dcb00441c6423e7c85a7583dd4a65ed1be03cb0e123311",
  "Labels": {
    "com.amazonaws.ecs.cluster": "arn:aws:ecs:us-west-2:&ExampleAWSAccountNo1;:cluster/default",
    "com.amazonaws.ecs.container-name": "query-metadata",
    "com.amazonaws.ecs.task-arn": "arn:aws:ecs:us-west-2:&ExampleAWSAccountNo1;:task/default/c7a6b9b237934e9999f319ea3ccc9da4",
    "com.amazonaws.ecs.task-definition-family": "query-metadata",
    "com.amazonaws.ecs.task-definition-version": "3"
  },
  "DesiredStatus": "RUNNING",
  "KnownStatus": "RUNNING",
  "Limits": {
    "CPU": 2
  },
  "CreatedAt": "2020-03-26T22:11:23.62831313Z",
  "StartedAt": "2020-03-26T22:11:23.62831313Z",
  "Type": "NORMAL",
  "Networks": [
    {
      "NetworkMode": "awsvpc",
      "IPv4Addresses": [
        "10.0.0.61"
      ],
      "AttachmentIndex": 0,
      "IPv4SubnetCIDRBlock": "10.0.0.0/24",
      "MACAddress": "0a:d2:d0:80:b6:b4",
      "DomainNameServers": [
        "10.0.0.2"
      ],
      "DomainNameSearchList": [
        "us-west-2.compute.internal"
      ]
    }
  ]
}
```

```

    ],
    "PrivateDNSName": "ip-10-0-0-61.us-west-2.compute.internal",
    "SubnetGatewayIpv4Address": ""
  }
}

```

Example Task Metadata Response

When querying the `#{ECS_CONTAINER_METADATA_URI_V4}/task` endpoint you are returned metadata about the task the container is part of. The following is an example output.

```

{
  "Cluster": "arn:aws:ecs:us-west-2:&ExampleAWSAccountNo1;:cluster/default",
  "TaskARN": "arn:aws:ecs:us-west-2:&ExampleAWSAccountNo1;:task/default/febee046097849aba589d4435207c04a",
  "Family": "query-metadata",
  "Revision": "7",
  "DesiredStatus": "RUNNING",
  "KnownStatus": "RUNNING",
  "Limits": {
    "CPU": 0.25,
    "Memory": 512
  },
  "PullStartedAt": "2020-03-26T22:25:40.420726088Z",
  "PullStoppedAt": "2020-03-26T22:26:22.235177616Z",
  "AvailabilityZone": "us-west-2c",
  "Containers": [
    {
      "DockerId": "febee046097849aba589d4435207c04aquery-metadata",
      "Name": "query-metadata",
      "DockerName": "query-metadata",
      "Image": "mreferre/eksutils",
      "ImageID": "sha256:1b146e73f801617610dcb00441c6423e7c85a7583dd4a65ed1be03cb0e123311",
      "Labels": {
        "com.amazonaws.ecs.cluster": "arn:aws:ecs:us-west-2:&ExampleAWSAccountNo1;:cluster/default",
        "com.amazonaws.ecs.container-name": "query-metadata",
        "com.amazonaws.ecs.task-arn": "arn:aws:ecs:us-west-2:&ExampleAWSAccountNo1;:task/default/febee046097849aba589d4435207c04a",
        "com.amazonaws.ecs.task-definition-family": "query-metadata",
        "com.amazonaws.ecs.task-definition-version": "7"
      },
      "DesiredStatus": "RUNNING",
      "KnownStatus": "RUNNING",
      "Limits": {
        "CPU": 2
      },
      "CreatedAt": "2020-03-26T22:26:24.534553758Z",
      "StartedAt": "2020-03-26T22:26:24.534553758Z",
      "Type": "NORMAL",
      "Networks": [
        {
          "NetworkMode": "awsvpc",
          "IPv4Addresses": [
            "10.0.0.108"
          ],
          "AttachmentIndex": 0,
          "IPv4SubnetCIDRBlock": "10.0.0.0/24",
          "MACAddress": "0a:62:17:7a:36:68",
          "DomainNameServers": [
            "10.0.0.2"
          ]
        }
      ]
    }
  ]
}

```

```

        "DomainNameSearchList": [
            "us-west-2.compute.internal"
        ],
        "PrivateDNSName": "ip-10-0-0-108.us-west-2.compute.internal",
        "SubnetGatewayIpv4Address": ""
    }
}
]
}
}

```

Example Task Stats Response

When querying the `${ECS_CONTAINER_METADATA_URI_V4}/task/stats` endpoint you are returned network metrics about the task the container is part of. The following is an example output.

```

{
  "1823e1f6-7248-43c3-bed6-eeaf7501a5query-metadata": {
    "read": "2020-04-06T16:12:01.090148907Z",
    "preread": "2020-04-06T16:11:56.083890951Z",
    "pids_stats": {
    },
    "blkio_stats": {
      "io_service_bytes_recursive": [
        {
          "major": 202,
          "minor": 26368,
          "op": "Read",
          "value": 3452928
        },
        {
          "major": 202,
          "minor": 26368,
          "op": "Write",
          "value": 0
        },
        {
          "major": 202,
          "minor": 26368,
          "op": "Sync",
          "value": 3452928
        },
        {
          "major": 202,
          "minor": 26368,
          "op": "Async",
          "value": 0
        },
        {
          "major": 202,
          "minor": 26368,
          "op": "Total",
          "value": 3452928
        }
      ],
      "io_serviced_recursive": [
        {
          "major": 202,
          "minor": 26368,
          "op": "Read",
          "value": 118
        },
        {

```

```

        "major": 202,
        "minor": 26368,
        "op": "Write",
        "value": 0
    },
    {
        "major": 202,
        "minor": 26368,
        "op": "Sync",
        "value": 118
    },
    {
        "major": 202,
        "minor": 26368,
        "op": "Async",
        "value": 0
    },
    {
        "major": 202,
        "minor": 26368,
        "op": "Total",
        "value": 118
    }
],
"io_queue_recursive": [

],
"io_service_time_recursive": [

],
"io_wait_time_recursive": [

],
"io_merged_recursive": [

],
"io_time_recursive": [

],
"sectors_recursive": [

]
},
"num_procs": 0,
"storage_stats": {

},
"cpu_stats": {
    "cpu_usage": {
        "total_usage": 410557100,
        "percpu_usage": [
            410557100,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0
        ]
    }
}

```

```

    ],
    "usage_in_kernelmode": 10000000,
    "usage_in_usermode": 250000000
  },
  "throttling_data": {
    "periods": 0,
    "throttled_periods": 0,
    "throttled_time": 0
  }
},
"precpu_stats": {
  "cpu_usage": {
    "total_usage": 0,
    "usage_in_kernelmode": 0,
    "usage_in_usermode": 0
  },
  "throttling_data": {
    "periods": 0,
    "throttled_periods": 0,
    "throttled_time": 0
  }
},
"memory_stats": {
  "usage": 4390912,
  "max_usage": 6488064,
  "stats": {
    "active_anon": 278528,
    "active_file": 344064,
    "cache": 3452928,
    "dirty": 0,
    "hierarchical_memory_limit": 536870912,
    "hierarchical_memsw_limit": 9223372036854772000,
    "inactive_anon": 0,
    "inactive_file": 3108864,
    "mapped_file": 2412544,
    "pgfault": 2800,
    "pgmajfault": 28,
    "pgpgin": 3144,
    "pgpgout": 2233,
    "rss": 278528,
    "rss_huge": 0,
    "total_active_anon": 278528,
    "total_active_file": 344064,
    "total_cache": 3452928,
    "total_dirty": 0,
    "total_inactive_anon": 0,
    "total_inactive_file": 3108864,
    "total_mapped_file": 2412544,
    "total_pgfault": 2800,
    "total_pgmajfault": 28,
    "total_pgpgin": 3144,
    "total_pgpgout": 2233,
    "total_rss": 278528,
    "total_rss_huge": 0,
    "total_unevictable": 0,
    "total_writeback": 0,
    "unevictable": 0,
    "writeback": 0
  },
  "limit": 9223372036854772000
},
"name": "query-metadata",
"id": "1823e1f6-7248-43c3-bed6-eealfa7501a5query-metadata",
"networks": {
  "eth1": {
    "rx_bytes": 564655295,

```

```
        "rx_packets": 384960,  
        "rx_errors": 0,  
        "rx_dropped": 0,  
        "tx_bytes": 3043269,  
        "tx_packets": 54355,  
        "tx_errors": 0,  
        "tx_dropped": 0  
      }  
    }  
  }  
}
```

Task metadata endpoint version 3

Beginning with Fargate platform version 1.1.0, an environment variable named `ECS_CONTAINER_METADATA_URI` is injected into each container in a task. When you query the task metadata version 3 endpoint, various task metadata and [Docker stats](#) are available to tasks.

Enabling Task Metadata

The task metadata endpoint feature is enabled by default for tasks using the Fargate launch type that use platform version v1.1.0 or later. For more information, see [AWS Fargate platform versions \(p. 14\)](#).

Task Metadata Endpoint Paths

The following API endpoints are available to containers:

`169.254.170.2/v2/metadata`

This endpoint returns metadata JSON for the task, including a list of the container IDs and names for all of the containers associated with the task. For more information about the response for this endpoint, see [Task Metadata JSON Response \(p. 279\)](#).

`169.254.170.2/v2/metadata/<container-id>`

This endpoint returns metadata JSON for the specified Docker container ID.

`169.254.170.2/v2/stats`

This endpoint returns Docker stats JSON for all of the containers associated with the task. For more information about each of the returned stats, see [ContainerStats](#) in the Docker API documentation.

`169.254.170.2/v2/stats/<container-id>`

This endpoint returns Docker stats JSON for the specified Docker container ID. For more information about each of the returned stats, see [ContainerStats](#) in the Docker API documentation.

Task Metadata JSON Response

The following information is returned from the task metadata endpoint (`169.254.170.2/v2/metadata`) JSON response.

Cluster

The full Amazon Resource Name (ARN) of the Amazon ECS cluster to which the task belongs.

TaskARN

The full Amazon Resource Name (ARN) of the task to which the container belongs.

Family

The family of the Amazon ECS task definition for the task.

Revision

The revision of the Amazon ECS task definition for the task.

DesiredStatus

The desired status for the task from Amazon ECS.

KnownStatus

The known status for the task from Amazon ECS.

Limits

The resource limits specified at the task level (such as CPU and memory). This parameter is omitted if no resource limits are defined.

PullStartedAt

The timestamp for when the first container image pull began.

PullStoppedAt

The timestamp for when the last container image pull finished.

AvailabilityZone

The Availability Zone the task is in.

Note

The Availability Zone metadata is only available for Fargate tasks using platform version 1.4 or later.

Containers

A list of container metadata for each container associated with the task.

DockerId

The Docker ID for the container.

Name

The name of the container as specified in the task definition.

DockerName

The name of the container supplied to Docker. The Amazon ECS container agent generates a unique name for the container to avoid name collisions when multiple copies of the same task definition are run on a single instance.

Image

The image for the container.

ImageID

The SHA-256 digest for the image.

Ports

Any ports exposed for the container. This parameter is omitted if there are no exposed ports.

Labels

Any labels applied to the container. This parameter is omitted if there are no labels applied.

DesiredStatus

The desired status for the container from Amazon ECS.

KnownStatus

The known status for the container from Amazon ECS.

ExitCode

The exit code for the container. This parameter is omitted if the container has not exited.

Limits

The resource limits specified at the container level (such as CPU and memory). This parameter is omitted if no resource limits are defined.

CreatedAt

The time stamp for when the container was created. This parameter is omitted if the container has not been created yet.

StartedAt

The time stamp for when the container started. This parameter is omitted if the container has not started yet.

FinishedAt

The time stamp for when the container stopped. This parameter is omitted if the container has not stopped yet.

Type

The type of the container. Containers that are specified in your task definition are of type **NORMAL**. You can ignore other container types, which are used for internal task resource provisioning by the Amazon ECS container agent.

Networks

The network information for the container, such as the network mode and IP address. This parameter is omitted if no network information is defined.

ExecutionStoppedAt

The time stamp for when the tasks **DesiredStatus** moved to **STOPPED**. This occurs when an essential container moves to **STOPPED**.

Example Task Metadata Response

The following JSON response is for a single-container task.

```
{
  "Cluster": "default",
  "TaskARN": "arn:aws:ecs:us-east-2:012345678910:task/9781c248-0edd-4cdb-9a93-f63cb662a5d3",
  "Family": "nginx",
  "Revision": "5",
  "DesiredStatus": "RUNNING",
  "KnownStatus": "RUNNING",
  "Containers": [
    {
      "DockerId": "731a0d6a3b4210e2448339bc7015aaa79bfe4fa256384f4102db86ef94cbbc4c",
      "Name": "~internal-ecs-pause",
      "DockerName": "ecs-nginx-5-internalecspause-acc699c0cbf2d6d11700",
      "Image": "amazon/amazon-ecs-pause:0.1.0",
```

```

    "ImageID": "",
    "Labels": {
      "com.amazonaws.ecs.cluster": "default",
      "com.amazonaws.ecs.container-name": "~internal-ecs-pause",
      "com.amazonaws.ecs.task-arn": "arn:aws:ecs:us-
east-2:012345678910:task/9781c248-0edd-4cdb-9a93-f63cb662a5d3",
      "com.amazonaws.ecs.task-definition-family": "nginx",
      "com.amazonaws.ecs.task-definition-version": "5"
    },
    "DesiredStatus": "RESOURCES_PROVISIONED",
    "KnownStatus": "RESOURCES_PROVISIONED",
    "Limits": {
      "CPU": 0,
      "Memory": 0
    },
    "CreatedAt": "2018-02-01T20:55:08.366329616Z",
    "StartedAt": "2018-02-01T20:55:09.058354915Z",
    "Type": "CNI_PAUSE",
    "Networks": [
      {
        "NetworkMode": "awsvpc",
        "IPv4Addresses": [
          "10.0.2.106"
        ]
      }
    ]
  },
  {
    "DockerId": "43481a6ce4842eec8fe72fc28500c6b52edcc0917f105b83379f88cac1ff3946",
    "Name": "nginx-curl",
    "DockerName": "ecs-nginx-5-nginx-curl-ccccb9f49db0dfe0d901",
    "Image": "nrdlngr/nginx-curl",
    "ImageID": "sha256:2e00ae64383cfc865ba0a2ba37f61b50a120d2d9378559dcd458dc0de47bc165",
    "Labels": {
      "com.amazonaws.ecs.cluster": "default",
      "com.amazonaws.ecs.container-name": "nginx-curl",
      "com.amazonaws.ecs.task-arn": "arn:aws:ecs:us-
east-2:012345678910:task/9781c248-0edd-4cdb-9a93-f63cb662a5d3",
      "com.amazonaws.ecs.task-definition-family": "nginx",
      "com.amazonaws.ecs.task-definition-version": "5"
    },
    "DesiredStatus": "RUNNING",
    "KnownStatus": "RUNNING",
    "Limits": {
      "CPU": 512,
      "Memory": 512
    },
    "CreatedAt": "2018-02-01T20:55:10.554941919Z",
    "StartedAt": "2018-02-01T20:55:11.064236631Z",
    "Type": "NORMAL",
    "Networks": [
      {
        "NetworkMode": "awsvpc",
        "IPv4Addresses": [
          "10.0.2.106"
        ]
      }
    ]
  }
],
"PullStartedAt": "2018-02-01T20:55:09.372495529Z",
"PullStoppedAt": "2018-02-01T20:55:10.552018345Z",
"AvailabilityZone": "us-east-2b"
}

```

Amazon ECS Service Quotas

The following table provides the default service quotas, also referred to as limits, for Amazon ECS for an AWS account which can be changed. For more information on the service limits for other AWS services that you can use with Amazon ECS, such as Elastic Load Balancing and Auto Scaling, see [AWS Service Limits](#) in the *Amazon Web Services General Reference*.

To request a quota increase, see [Requesting a Quota Increase](#) in the *Service Quotas User Guide*.

Service quota	Description	Default quota value
Clusters per account	The maximum number of clusters per Region, per account.	10,000
Services per cluster	The maximum number of services per cluster.	1,000
Tasks using the Fargate launch type or the FARGATE capacity provider, per Region, per account	The maximum number of tasks using the Fargate launch type or the FARGATE capacity provider, per Region. This limit applies to both standalone tasks and tasks launched as part of a service.	100
Fargate Spot tasks, per Region, per account	The maximum number of tasks using the FARGATE_SPOT capacity provider, per Region.	250
Public IP addresses for tasks using the Fargate launch type	The maximum number of public IP addresses used by tasks using the Fargate launch type, per Region.	100

The following table provides other limitations for Amazon ECS that cannot be changed.

Note

The quotas for Fargate task storage are dependent on the platform version used by the task. For more information, see [Using Data Volumes in Tasks \(p. 63\)](#).

Service quota	Description	Default quota value
Classic Load Balancers per service	The maximum number of Classic Load Balancers per service.	1
Tasks launched (count) per run-task	The maximum number of tasks that can be launched per RunTask API action.	10
Revisions per task definition family	The maximum number of revisions per task definition family. Deregistering a task definition revision does not exclude it from being included in this limit.	1,000,000

Service quota	Description	Default quota value
Task definition size limit	The maximum size, in KiB, of a task definition.	32
Task definition max containers	The maximum number of containers definitions within a task definition.	10
Subnets specified in an <code>awsvpcConfiguration</code>	The maximum number of subnets specified within an <code>awsvpcConfiguration</code> .	16
Security groups specified in an <code>awsvpcConfiguration</code>	The maximum number of security groups specified within an <code>awsvpcConfiguration</code> .	5
Tags per resource	The maximum number of tags per resource. This applies to tasks, services, task definitions, clusters, and container instances.	50

Using Service Quotas

Amazon Elastic Container Service (Amazon ECS) has integrated with Service Quotas, an AWS service that enables you to view and manage your quotas from a central location. Service quotas are also referred to as limits. For more information, see [What Is Service Quotas?](#) in the *Service Quotas User Guide*.

Service Quotas makes it easy to look up the value of all of the Amazon ECS service quotas.

To view Amazon ECS service quotas (AWS Management Console)

1. Open the Service Quotas console at <https://console.aws.amazon.com/servicequotas/>.
2. In the navigation pane, choose **AWS services**.
3. From the **AWS services** list, search for and select **Amazon Elastic Container Service (Amazon ECS)**.

In the **Service quotas** list, you can see the service quota name, applied value (if it is available), AWS default quota, and whether the quota value is adjustable.

4. To view additional information about a service quota, such as the description, choose the quota name.

To request a quota increase, see [Requesting a Quota Increase](#) in the *Service Quotas User Guide*.

Savings Plans and AWS Fargate

Savings Plans are a pricing model that offer significant savings on AWS usage. This pricing model offers lower prices on your AWS Fargate usage. You commit to a consistent amount of usage, in USD per hour, for a term of 1 or 3 years, and receive a lower price for that usage. For more information, see the [Savings Plans User Guide](#).

Each type of usage has an On-Demand rate and a Savings Plans price. For example, if you commit to \$10/hour of compute usage, you'll receive Savings Plans prices on all usage up to \$10. Any usage beyond the commitment is charged at On-Demand rates.

Savings Plans are available for 1 year or 3 year terms. You have the choice between **All Upfront**, **Partial upfront**, or **No upfront** payment options.

The following Savings Plans type is available for Fargate usage. For a full list of Savings Plans for both Amazon EC2 and Fargate usage, see [Savings Plans and Amazon ECS](#) in the *Amazon Elastic Container Service Developer Guide*.

- **Compute Savings Plans** provide the most flexibility and provide a discount of up to 66 percent. These plans automatically apply to your Fargate usage. You can move your Amazon ECS tasks from using Amazon EC2 to Fargate at any time and continue to receive the discounted rate provided by your Compute Savings Plan.

To get started, see [Get Started with Savings Plans](#) in the *Savings Plans User Guide*.

Getting started with AWS App Mesh and Amazon ECS

AWS App Mesh is a service mesh based on the [Envoy](#) proxy that helps you monitor and control services. App Mesh standardizes how your services communicate, giving you end-to-end visibility into and helping to ensure high-availability for your applications. App Mesh gives you consistent visibility and network traffic controls for every service in an application. For more information, see the [AWS App Mesh User Guide](#).

This topic helps you use AWS App Mesh with an actual service that is running on Amazon ECS. This tutorial covers basic features of App Mesh. To learn more about other features that you'll, but that aren't used when completing this tutorial, see the topics for [virtual nodes](#), [virtual services](#), [virtual routers](#), [routes](#), and the [Envoy proxy](#).

Scenario

To illustrate how to use App Mesh with Amazon ECS, assume that you have an application with the following characteristics:

- Includes two services named `serviceA` and `serviceB`.
- Both services are registered to a namespace named `apps.local`.
- `ServiceA` communicates with `serviceB` over HTTP/2, port 80.
- You've already deployed version 2 of `serviceB` and registered it with the name `serviceBv2` in the `apps.local` namespace.

You have the following requirements:

- You want to send 75 percent of the traffic from `serviceA` to `serviceB` and 25 percent of the traffic to `serviceBv2` to ensure that `serviceBv2` is bug free before you send 100 percent of the traffic from `serviceA` to it.
- You want to be able to easily adjust the traffic weighting so that 100 percent of the traffic goes to `serviceBv2` once it's proven to be reliable. Once all traffic is being sent to `serviceBv2`, you want to deprecate `serviceB`.
- You don't want to have to change any existing application code or service discovery registration for your actual services to meet the previous requirements.

To meet your requirements, you've decided to create an App Mesh service mesh with virtual services, virtual nodes, a virtual router, and a route. After implementing your mesh, you update the task definitions for your services to use the Envoy proxy. Once updated, your services communicate with each other through the Envoy proxy rather than directly with each other.

Prerequisites

App Mesh supports Linux services that are registered with DNS, AWS Cloud Map, or both. To use this getting started guide, we recommend that you have three existing services that are registered with DNS. You can create a service mesh and its resources even if the services don't exist, but you can't use the mesh until you have deployed actual services.

For more information about service discovery on Amazon ECS, see [Service Discovery](#). To create an Amazon ECS service with service discovery, see [Tutorial: Creating a Service Using Service Discovery](#).

The remaining steps assume that the actual services are named `serviceA`, `serviceB`, and `serviceBv2` and that all services are discoverable through a namespace named `apps.local`.

Step 1: Create a mesh and virtual service

A service mesh is a logical boundary for network traffic between the services that reside within it. For more information, see [Service Meshes](#) in the *AWS App Mesh User Guide*. A virtual service is an abstraction of an actual service. For more information, see [Virtual Services](#) in the *AWS App Mesh User Guide*.

Create the following resources:

- A mesh named `apps`, since all of the services in the scenario are registered to the `apps.local` namespace.
- A virtual service named `serviceb.apps.local`, since the virtual service represents a service that is discoverable with that name, and you don't want to change your code to reference another name. A virtual service named `servicea.apps.local` is added in a later step.

You can use the AWS Management Console or the AWS CLI version 1.18.71 or higher or 2.0.17 or higher to complete the following steps. If using the AWS CLI, use the `aws --version` command to check your installed AWS CLI version. If you don't have version 1.18.71 or higher or 2.0.17 or higher installed, then you must [install or update the AWS CLI](#). Select the tab for the tool that you want to use.

AWS Management Console

1. Open the App Mesh console first-run wizard at <https://console.aws.amazon.com/appmesh/get-started>.
2. For **Mesh name**, enter `apps`.
3. For **Virtual service name**, enter `serviceb.apps.local`.
4. To continue, choose **Next**.

AWS CLI

1. Create a mesh with the `create-mesh` command.

```
aws appmesh create-mesh --mesh-name apps
```

2. Create a virtual service with the `create-virtual-service` command.

```
aws appmesh create-virtual-service --mesh-name apps --virtual-service-name  
serviceb.apps.local --spec {}
```

Step 2: Create a virtual node

A virtual node acts as a logical pointer to an actual service. For more information, see [Virtual Nodes](#) in the *AWS App Mesh User Guide*.

Create a virtual node named `serviceB`, since one of the virtual nodes represents the actual service named `serviceB`. The actual service that the virtual node represents is discoverable through DNS with a hostname of `serviceb.apps.local`. Alternately, you can discover actual services using AWS Cloud

Map. The virtual node will listen for traffic using the HTTP/2 protocol on port 80. Other protocols are also supported, as are health checks. You will create virtual nodes for `serviceA` and `serviceBv2` in a later step.

AWS Management Console

1. For **Virtual node name**, enter **serviceB**.
2. For **Service discovery method**, choose **DNS** and enter **serviceb.apps.local** for **DNS hostname**.
3. Under **Listener configuration**, choose **http2** for **Protocol** and enter **80** for **Port**.
4. To continue, choose **Next**.

AWS CLI

1. Create a file named `create-virtual-node-serviceb.json` with the following contents:

```
{
  "meshName": "apps",
  "spec": {
    "listeners": [
      {
        "portMapping": {
          "port": 80,
          "protocol": "http2"
        }
      }
    ],
    "serviceDiscovery": {
      "dns": {
        "hostname": "serviceB.apps.local"
      }
    }
  },
  "virtualNodeName": "serviceB"
}
```

2. Create the virtual node with the `create-virtual-node` command using the JSON file as input.

```
aws appmesh create-virtual-node --cli-input-json file://create-virtual-node-
serviceb.json
```

Step 3: Create a virtual router and route

Virtual routers route traffic for one or more virtual services within your mesh. For more information, see [Virtual Routers](#) and [Routes](#) in the *AWS App Mesh User Guide*.

Create the following resources:

- A virtual router named `serviceB`, since the `serviceB.apps.local` virtual service doesn't initiate outbound communication with any other service. Remember that the virtual service that you created previously is an abstraction of your actual `serviceb.apps.local` service. The virtual service sends traffic to the virtual router. The virtual router will listen for traffic using the HTTP/2 protocol on port 80. Other protocols are also supported.
- A route named `serviceB`. It will route 100 percent of its traffic to the `serviceB` virtual node. You'll change the weight in a later step once you've added the `serviceBv2` virtual node. Though not covered in this guide, you can add additional filter criteria for the route and add a retry policy to cause

the Envoy proxy to make multiple attempts to send traffic to a virtual node when it experiences a communication problem.

AWS Management Console

1. For **Virtual router name**, enter **serviceB**.
2. Under **Listener configuration**, choose **http2** for **Protocol** and specify **80** for **Port**.
3. For **Route name**, enter **serviceB**.
4. For **Route type**, choose **http2**.
5. For **Virtual node name** under **Route configuration**, select **serviceB** and enter **100** for **Weight**.
6. To continue, choose **Next**.

AWS CLI

1. Create a virtual router.
 - a. Create a file named `create-virtual-router.json` with the following contents:

```
{
  "meshName": "apps",
  "spec": {
    "listeners": [
      {
        "portMapping": {
          "port": 80,
          "protocol": "http2"
        }
      }
    ]
  },
  "virtualRouterName": "serviceB"
}
```

- b. Create the virtual router with the `create-virtual-router` command using the JSON file as input.

```
aws appmesh create-virtual-router --cli-input-json file://create-virtual-router.json
```

2. Create a route.
 - a. Create a file named `create-route.json` with the following contents:

```
{
  "meshName" : "apps",
  "routeName" : "serviceB",
  "spec" : {
    "httpRoute" : {
      "action" : {
        "weightedTargets" : [
          {
            "virtualNode" : "serviceB",
            "weight" : 100
          }
        ]
      }
    },
    "match" : {
      "prefix" : "/"
    }
  }
}
```

```
    },  
    "virtualRouterName" : "serviceB"  
  }  
}
```

- b. Create the route with the [create-route](#) command using the JSON file as input.

```
aws appmesh create-route --cli-input-json file://create-route.json
```

Step 4: Review and create

Review the settings against the previous instructions.

AWS Management Console

Choose **Edit** if you need to make changes in any section. Once you're satisfied with the settings, choose **Create mesh**.

The **Status** screen shows you all of the mesh resources that were created. You can see the created resources in the console by selecting **View mesh**.

AWS CLI

Review the settings of the mesh you created with the [describe-mesh](#) command.

```
aws appmesh describe-mesh --mesh-name apps
```

Review the settings of the virtual service that you created with the [describe-virtual-service](#) command.

```
aws appmesh describe-virtual-service --mesh-name apps --virtual-service-name  
serviceb.apps.local
```

Review the settings of the virtual node that you created with the [describe-virtual-node](#) command.

```
aws appmesh describe-virtual-node --mesh-name apps --virtual-node-name serviceB
```

Review the settings of the virtual router that you created with the [describe-virtual-router](#) command.

```
aws appmesh describe-virtual-router --mesh-name apps --virtual-router-name serviceB
```

Review the settings of the route that you created with the [describe-route](#) command.

```
aws appmesh describe-route --mesh-name apps \  
--virtual-router-name serviceB --route-name serviceB
```

Step 5: Create additional resources

To complete the scenario, you need to:

- Create one virtual node named `serviceBv2` and another named `serviceA`. Both virtual nodes listen for requests over HTTP/2 port 80. For the `serviceA` virtual node, configure a backend of

`serviceb.apps.local`, since all outbound traffic from the `serviceA` virtual node is sent to the virtual service named `serviceb.apps.local`. Though not covered in this guide, you can also specify a file path to write access logs to for a virtual node.

- Create one additional virtual service named `servicea.apps.local`, which will send all traffic directly to the `serviceA` virtual node.
- Update the `serviceB` route that you created in a previous step to send 75 percent of its traffic to the `serviceB` virtual node and 25 percent of its traffic to the `serviceBv2` virtual node. Over time, you can continue to modify the weights until `serviceBv2` receives 100 percent of the traffic. Once all traffic is sent to `serviceBv2`, you can deprecate the `serviceB` virtual node and actual service. As you change weights, your code doesn't require any modification, because the `serviceb.apps.local` virtual and actual service names don't change. Recall that the `serviceb.apps.local` virtual service sends traffic to the virtual router, which routes the traffic to the virtual nodes. The service discovery names for the virtual nodes can be changed at any time.

AWS Management Console

1. In the left navigation pane, select **Meshes**.
2. Select the `apps` mesh that you created in a previous step.
3. In the left navigation pane, select **Virtual nodes**.
4. Choose **Create virtual node**.
5. For **Virtual node name**, enter `serviceBv2`, for **Service discovery method**, choose **DNS**, and for **DNS hostname**, enter `servicebv2.apps.local`.
6. For **Listener configuration**, select **http2** for **Protocol** and enter **80** for **Port**.
7. Choose **Create virtual node**.
8. Choose **Create virtual node** again. Enter `serviceA` for the **Virtual node name**. For **Service discovery method**, choose **DNS**, and for **DNS hostname**, enter `servicea.apps.local`.
9. For **Enter a virtual service name** under **New backend**, enter `servicea.apps.local`.
10. Under **Listener configuration**, choose **http2** for **Protocol**, enter **80** for **Port**, and then choose **Create virtual node**.
11. In the left navigation pane, select **Virtual routers** and then select the `serviceB` virtual router from the list.
12. Under **Routes**, select the route named `ServiceB` that you created in a previous step, and choose **Edit**.
13. Under **Targets**, **Virtual node name**, change the value of **Weight** for `serviceB` to **75**.
14. Choose **Add target**, choose `serviceBv2` from the drop-down list, and set the value of **Weight** to **25**.
15. Choose **Save**.
16. In the left navigation pane, select **Virtual services** and then choose **Create virtual service**.
17. Enter `servicea.apps.local` for **Virtual service name**, select **Virtual node** for **Provider**, select `serviceA` for **Virtual node**, and then choose **Create virtual service**.

AWS CLI

1. Create the `serviceBv2` virtual node.
 - a. Create a file named `create-virtual-node-servicebv2.json` with the following contents:

```
{
  "meshName": "apps",
  "spec": {
    "listeners": [
```

```
{
  "portMapping": {
    "port": 80,
    "protocol": "http2"
  }
},
"serviceDiscovery": {
  "dns": {
    "hostname": "serviceBv2.apps.local"
  }
},
"virtualNodeName": "serviceBv2"
}
```

- b. Create the virtual node.

```
aws appmesh create-virtual-node --cli-input-json file://create-virtual-node-
servicebv2.json
```

2. Create the serviceA virtual node.

- a. Create a file named `create-virtual-node-servicea.json` with the following contents:

```
{
  "meshName" : "apps",
  "spec" : {
    "backends" : [
      {
        "virtualService" : {
          "virtualServiceName" : "serviceb.apps.local"
        }
      }
    ],
    "listeners" : [
      {
        "portMapping" : {
          "port" : 80,
          "protocol" : "http2"
        }
      }
    ],
    "serviceDiscovery" : {
      "dns" : {
        "hostname" : "servicea.apps.local"
      }
    }
  },
  "virtualNodeName" : "serviceA"
}
```

- b. Create the virtual node.

```
aws appmesh create-virtual-node --cli-input-json file://create-virtual-node-
servicea.json
```

3. Update the `serviceb.apps.local` virtual service that you created in a previous step to send its traffic to the `serviceB` virtual router. When the virtual service was originally created, it didn't send traffic anywhere, since the `serviceB` virtual router hadn't been created yet.

- a. Create a file named `update-virtual-service.json` with the following contents:

```
{
  "meshName" : "apps",
  "spec" : {
    "provider" : {
      "virtualRouter" : {
        "virtualRouterName" : "serviceB"
      }
    }
  },
  "virtualServiceName" : "serviceb.apps.local"
}
```

- b. Update the virtual service with the [update-virtual-service](#) command.

```
aws appmesh update-virtual-service --cli-input-json file://update-virtual-
service.json
```

4. Update the serviceB route that you created in a previous step.

- a. Create a file named `update-route.json` with the following contents:

```
{
  "meshName" : "apps",
  "routeName" : "serviceB",
  "spec" : {
    "http2Route" : {
      "action" : {
        "weightedTargets" : [
          {
            "virtualNode" : "serviceB",
            "weight" : 75
          },
          {
            "virtualNode" : "serviceBv2",
            "weight" : 25
          }
        ]
      },
      "match" : {
        "prefix" : "/"
      }
    }
  },
  "virtualRouterName" : "serviceB"
}
```

- b. Update the route with the [update-route](#) command.

```
aws appmesh update-route --cli-input-json file://update-route.json
```

5. Create the serviceA virtual service.

- a. Create a file named `create-virtual-servicea.json` with the following contents:

```
{
  "meshName" : "apps",
  "spec" : {
    "provider" : {
      "virtualNode" : {
        "virtualNodeName" : "serviceA"
      }
    }
  }
}
```

```
    },  
    "virtualServiceName" : "servicea.apps.local"  
  }  
}
```

- b. Create the virtual service.

```
aws appmesh create-virtual-service --cli-input-json file://create-virtual-  
servicea.json
```

Mesh summary

Before you created the service mesh, you had three actual services named `servicea.apps.local`, `serviceb.apps.local`, and `servicebv2.apps.local`. In addition to the actual services, you now have a service mesh that contains the following resources that represent the actual services:

- Two virtual services. The proxy sends all traffic from the `servicea.apps.local` virtual service to the `serviceb.apps.local` virtual service through a virtual router.
- Three virtual nodes named `serviceA`, `serviceB`, and `serviceBv2`. The Envoy proxy uses the service discovery information configured for the virtual nodes to look up the IP addresses of the actual services.
- One virtual router with one route that instructs the Envoy proxy to route 75 percent of inbound traffic to the `serviceB` virtual node and 25 percent of the traffic to the `serviceBv2` virtual node.

Step 6: Update services

After creating your mesh, you need to complete the following tasks:

- Authorize the Envoy proxy that you deploy with each Amazon ECS task to read the configuration of one or more virtual nodes. For more information about how to authorize the proxy, see [Proxy authorization](#).
- Update each of your existing Amazon ECS task definitions to use the Envoy proxy.

Credentials

The Envoy container requires AWS Identity and Access Management credentials for signing requests that are sent to the App Mesh service. For Amazon ECS tasks deployed with the Amazon EC2 launch type, the credentials can come from the [instance role](#) or from a [task IAM role](#). Amazon ECS tasks deployed with the Fargate launch type do not have access to the Amazon EC2 metadata server that supplies instance IAM profile credentials. To supply the credentials, you must attach an IAM task role to any tasks deployed with the Fargate launch type. If a task is deployed with the Amazon EC2 launch type and access is blocked to the Amazon EC2 metadata server, as described in the *Important* annotation in [IAM Role for Tasks](#), then a task IAM role must also be attached to the task. The role that you assign to the instance or task must have an IAM policy attached to it as described in [Proxy authorization](#).

Update task definitions

You can update your task definitions by using the AWS Management Console or by modifying the JSON file for a task definition. The following steps only show updating the `taskB` task for the scenario. You also need to update the `taskBv2` and `taskA` tasks by changing the values appropriately. Select the method you prefer to use to update the task definition.

AWS Management Console

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.

2. From the navigation bar, choose the Region that contains your task definition.
3. In the navigation pane, choose **Task Definitions**.
4. On the **Task Definitions** page, select the box to the left of the task definition to revise. From the pre-requisites and previous steps, you might have task definitions named `taskA`, `taskB`, and `taskBv2`. Select `taskB` and choose **Create new revision**.
5. On the **Create new revision of Task Definition** page, make the following changes to enable App Mesh integration.
 - a. For **Service Integration**, to configure the parameters for App Mesh integration choose **Enable App Mesh integration** and then do the following:
 - i. For **Application container name**, choose the container name to use for the App Mesh application. This container must already be defined within the task definition.
 - ii. For **Envoy image**, enter one of the following images:
 - All [supported](#) Regions other than `me-south-1` and `ap-east-1`. You can replace `us-west-2` with any Region other than `me-south-1` and `ap-east-1`.

```
840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod
```
 - `me-south-1` Region:

```
772975370895.dkr.ecr.me-south-1.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod
```
 - `ap-east-1` Region:

```
856666278305.dkr.ecr.ap-east-1.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod
```
 - iii. For **Mesh name**, choose the App Mesh service mesh to use. In this topic, the name of the mesh that was created is `apps`.
 - iv. For **Virtual node name**, choose the App Mesh virtual node to use. For example, for the `taskB` task, you would choose the `serviceB` virtual node that you created in a previous step.
 - v. The value for **Virtual node port** is pre-populated with the listener port that you specified when you created the virtual node.
 - vi. Choose **Apply**, and then choose **Confirm**. A new Envoy proxy container is created and added to the task definition, and the settings to support the container are also created. The Envoy proxy container then pre-populates the App Mesh **Proxy Configuration** settings for the next step.
 - b. For **Proxy Configuration**, verify all of the pre-populated values.
 - c. For **Network Mode**, ensure that `awsvpc` is selected. To learn more about the `awsvpc` network mode, see [Task Networking with the awsvpc Network Mode](#).
6. Choose **Create**.
7. Update your service with the updated task definition. For more information, see [Updating a service](#).

JSON

Proxy configuration

To configure your Amazon ECS service to use App Mesh, your service's task definition must have the following proxy configuration section. Set the proxy configuration `type` to `APPMESH` and the `containerName` to `envoy`. Set the following property values accordingly.

IgnoredUID

The Envoy proxy doesn't route traffic from processes that use this user ID. You can choose any user ID that you want for this property value, but this ID must be the same as the `user` ID for the Envoy container in your task definition. This matching allows Envoy to ignore its own traffic without using the proxy. Our examples use `1337` for historical purposes.

ProxyIngressPort

This is the ingress port for the Envoy proxy container. Set this value to 15000.

ProxyEgressPort

This is the egress port for the Envoy proxy container. Set this value to 15001.

AppPorts

Specify any ingress ports that your application containers listen on. In this example, the application container listens on port `9080`. The port that you specify must match the port configured on the virtual node listener.

EgressIgnoredIPs

Envoy doesn't proxy traffic to these IP addresses. Set this value to `169.254.170.2,169.254.169.254`, which ignores the Amazon EC2 metadata server and the Amazon ECS task metadata endpoint. The metadata endpoint provides IAM roles for tasks credentials. You can add additional addresses.

EgressIgnoredPorts

You can add a comma separated list of ports. Envoy doesn't proxy traffic to these ports. Even if you list no ports, port 22 is ignored.

```
"proxyConfiguration": {
  "type": "APPMESH",
  "containerName": "envoy",
  "properties": [{
    "name": "IgnoredUID",
    "value": "1337"
  },
  {
    "name": "ProxyIngressPort",
    "value": "15000"
  },
  {
    "name": "ProxyEgressPort",
    "value": "15001"
  },
  {
    "name": "AppPorts",
    "value": "9080"
  },
  {
    "name": "EgressIgnoredIPs",
    "value": "169.254.170.2,169.254.169.254"
  },
  {
    "name": "EgressIgnoredPorts",
    "value": "22"
  }
  ]
}
```



```
]
}
```

Application container Envoy dependency

The application containers in your task definitions must wait for the Envoy proxy to bootstrap and start before they can start. To ensure that this happens, you set a `dependsOn` section in each application container definition to wait for the Envoy container to report as `HEALTHY`. The following code shows an application container definition example with this dependency. All of the properties in the following example are required. Some of the property values are also required, but some are *replaceable*.

```
{
  "name": "appName",
  "image": "appImage",
  "portMappings": [{
    "containerPort": 9080,
    "hostPort": 9080,
    "protocol": "tcp"
  }],
  "essential": true,
  "dependsOn": [{
    "containerName": "envoy",
    "condition": "HEALTHY"
  }]
}
```

Envoy container definition

Your Amazon ECS task definitions must contain one of the following [App Mesh Envoy container images](#):

- All [supported](#) Regions other than `me-south-1` and `ap-east-1`. You can replace `us-west-2` with any Region other than `me-south-1` and `ap-east-1`.

```
840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod
```

- `me-south-1` Region:

```
772975370895.dkr.ecr.me-south-1.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod
```

- `ap-east-1` Region:

```
856666278305.dkr.ecr.ap-east-1.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod
```

You must use the App Mesh Envoy container image until the Envoy project team merges changes that support App Mesh. For additional details, see the [GitHub roadmap issue](#).

All of the properties in the following example are required. Some of the property values are also required, but some are *replaceable*. The Envoy container definition must be marked as `essential`. The virtual node name for the Amazon ECS service must be set to the value of the `APPMESH_VIRTUAL_NODE_NAME` property. The value for the user setting must match the `IgnoredUID` value from the task definition proxy configuration. In this example, we use `1337`. The health check shown here waits for the Envoy container to bootstrap properly before reporting to Amazon ECS that the Envoy container is healthy and ready for the application containers to start.

The following code shows an Envoy container definition example.

```
{
  "name": "envoy",
  "image": "840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod",
  "essential": true,
  "environment": [{
    "name": "APPMESH_VIRTUAL_NODE_NAME",
    "value": "mesh/apps/virtualNode/serviceB"
  }],
  "healthCheck": {
    "command": [
      "CMD-SHELL",
      "curl -s http://localhost:9901/server_info | grep state | grep -q LIVE"
    ],
    "startPeriod": 10,
    "interval": 5,
    "timeout": 2,
    "retries": 3
  },
  "user": "1337"
}
```

Example task definitions

The following example Amazon ECS task definitions show how to merge the examples from above into a task definition for `taskB`. Examples are provided for creating tasks for both Amazon ECS launch types with or without using AWS X-Ray. Change the *replaceable* values, as appropriate, to create task definitions for the tasks named `taskBv2` and `taskA` from the scenario. Substitute your mesh name and virtual node name for the `APPMESH_VIRTUAL_NODE_NAME` value and a list of ports that your application listens on for the proxy configuration `AppPorts` value. All of the properties in the following examples are required. Some of the property values are also required, but some are *replaceable*.

If you're running an Amazon ECS task as described in the Credentials section, then you need to add an existing [task IAM role](#), to the examples.

Example JSON for Amazon ECS task definition - Fargate launch type

```
{
  "family" : "taskB",
  "memory" : "1024",
  "cpu" : "0.5 vCPU",
  "proxyConfiguration" : {
    "containerName" : "envoy",
    "properties" : [
      {
        "name" : "ProxyIngressPort",
        "value" : "15000"
      },
      {
        "name" : "AppPorts",
        "value" : "9080"
      },
      {
        "name" : "EgressIgnoredIPs",
        "value" : "169.254.170.2,169.254.169.254"
      },
      {
        "name": "EgressIgnoredPorts",
        "value": "22"
      }
    ]
  }
}
```

```

        {
            "name" : "IgnoredUID",
            "value" : "1337"
        },
        {
            "name" : "ProxyEgressPort",
            "value" : "15001"
        }
    ],
    "type" : "APPMESH"
},
"containerDefinitions" : [
    {
        "name" : "appName",
        "image" : "appImage",
        "portMappings" : [
            {
                "containerPort" : 9080,
                "protocol" : "tcp"
            }
        ],
        "essential" : true,
        "dependsOn" : [
            {
                "containerName" : "envoy",
                "condition" : "HEALTHY"
            }
        ]
    },
    {
        "name" : "envoy",
        "image" : "840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-
envoy:v1.12.3.0-prod",
        "essential" : true,
        "environment" : [
            {
                "name" : "APPMESH_VIRTUAL_NODE_NAME",
                "value" : "mesh/apps/virtualNode/serviceB"
            }
        ],
        "healthCheck" : {
            "command" : [
                "CMD-SHELL",
                "curl -s http://localhost:9901/server_info | grep state | grep -q LIVE"
            ],
            "interval" : 5,
            "retries" : 3,
            "startPeriod" : 10,
            "timeout" : 2
        },
        "memory" : "500",
        "user" : "1337"
    }
],
"requiresCompatibilities" : [ "FARGATE" ],
"taskRoleArn" : "arn:aws:iam::123456789012:role/ecsTaskRole",
"executionRoleArn" : "arn:aws:iam::123456789012:role/ecsTaskExecutionRole",
"networkMode" : "awsvpc"
}

```

Example JSON for Amazon ECS task definition with AWS X-Ray - Fargate launch type

X-Ray allows you to collect data about requests that an application serves and provides tools that you can use to visualize traffic flow. Using the X-Ray driver for Envoy enables Envoy to report tracing information to X-Ray. You can enable X-Ray tracing using the [Envoy configuration](#). Based on the

configuration, Envoy sends tracing data to the X-Ray daemon running as a [sidecar](#) container and the daemon forwards the traces to the X-Ray service. Once the traces are published to X-Ray, you can use the X-Ray console to visualize the service call graph and request trace details. The following JSON represents a task definition to enable X-Ray integration.

```
{
  "family" : "taskB",
  "memory" : "1024",
  "cpu" : "0.5 vCPU",
  "proxyConfiguration" : {
    "containerName" : "envoy",
    "properties" : [
      {
        "name" : "ProxyIngressPort",
        "value" : "15000"
      },
      {
        "name" : "AppPorts",
        "value" : "9080"
      },
      {
        "name" : "EgressIgnoredIPs",
        "value" : "169.254.170.2,169.254.169.254"
      },
      {
        "name": "EgressIgnoredPorts",
        "value": "22"
      },
      {
        "name" : "IgnoredUID",
        "value" : "1337"
      },
      {
        "name" : "ProxyEgressPort",
        "value" : "15001"
      }
    ],
    "type" : "APPMESH"
  },
  "containerDefinitions" : [
    {
      "name" : "appName",
      "image" : "appImage",
      "portMappings" : [
        {
          "containerPort" : 9080,
          "protocol" : "tcp"
        }
      ],
      "essential" : true,
      "dependsOn" : [
        {
          "containerName" : "envoy",
          "condition" : "HEALTHY"
        }
      ]
    }
  ],
  {
    "name" : "envoy",
    "image" : "840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-envoy:v1.12.3.0-prod",
    "essential" : true,
```

```
    "environment" : [
      {
        "name" : "APPMESH_VIRTUAL_NODE_NAME",
        "value" : "mesh/apps/virtualNode/serviceB"
      },
      {
        "name": "ENABLE_ENVOY_XRAY_TRACING",
        "value": "1"
      }
    ],
    "healthCheck" : {
      "command" : [
        "CMD-SHELL",
        "curl -s http://localhost:9901/server_info | grep state | grep -q LIVE"
      ],
      "interval" : 5,
      "retries" : 3,
      "startPeriod" : 10,
      "timeout" : 2
    },
    "memory" : "500",
    "user" : "1337"
  },
  {
    "name" : "xray-daemon",
    "image" : "amazon/aws-xray-daemon",
    "user" : "1337",
    "essential" : true,
    "cpu" : "32",
    "memoryReservation" : "256",
    "portMappings" : [
      {
        "containerPort" : 2000,
        "protocol" : "udp"
      }
    ]
  }
],
"requiresCompatibilities" : [ "FARGATE" ],
"taskRoleArn" : "arn:aws:iam::123456789012:role/ecsTaskRole",
"executionRoleArn" : "arn:aws:iam::123456789012:role/ecsTaskExecutionRole",
"networkMode" : "awsvpc"
}
```

Example JSON for Amazon ECS task definition - EC2 launch type

```
{
  "family": "taskB",
  "memory": "256",
  "proxyConfiguration": {
    "type": "APPMESH",
    "containerName": "envoy",
    "properties": [
      {
        "name": "IgnoredUID",
        "value": "1337"
      },
      {
        "name": "ProxyIngressPort",
        "value": "15000"
      },
      {
        "name": "ProxyEgressPort",
        "value": "15001"
      }
    ]
  }
}
```

```

    },
    {
      "name": "AppPorts",
      "value": "9080"
    },
    {
      "name": "EgressIgnoredIPs",
      "value": "169.254.170.2,169.254.169.254"
    },
    {
      "name": "EgressIgnoredPorts",
      "value": "22"
    }
  ]
},
"containerDefinitions": [
  {
    "name": "appName",
    "image": "appImage",
    "portMappings": [
      {
        "containerPort": 9080,
        "hostPort": 9080,
        "protocol": "tcp"
      }
    ],
    "essential": true,
    "dependsOn": [
      {
        "containerName": "envoy",
        "condition": "HEALTHY"
      }
    ]
  },
  {
    "name": "envoy",
    "image": "840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-
envoy:v1.12.3.0-prod",
    "essential": true,
    "environment": [
      {
        "name": "APPMESH_VIRTUAL_NODE_NAME",
        "value": "mesh/apps/virtualNode/serviceB"
      }
    ],
    "healthCheck": {
      "command": [
        "CMD-SHELL",
        "curl -s http://localhost:9901/server_info | grep state | grep -q LIVE"
      ],
      "startPeriod": 10,
      "interval": 5,
      "timeout": 2,
      "retries": 3
    },
    "user": "1337"
  }
],
"requiresCompatibilities" : [ "EC2" ],
"taskRoleArn" : "arn:aws:iam::123456789012:role/ecsTaskRole",
"executionRoleArn" : "arn:aws:iam::123456789012:role/ecsTaskExecutionRole",
"networkMode": "awsvpc"
}

```

Example JSON for Amazon ECS task definition with AWS X-Ray - EC2 launch type

X-Ray allows you to collect data about requests that an application serves and provides tools that you can use to visualize traffic flow. Using the X-Ray driver for Envoy enables Envoy to report tracing information to X-Ray. You can enable X-Ray tracing using the [Envoy configuration](#). Based on the configuration, Envoy sends tracing data to the X-Ray daemon running as a [sidecar](#) container and the daemon forwards the traces to the X-Ray service. Once the traces are published to X-Ray, you can use the X-Ray console to visualize the service call graph and request trace details. The following JSON represents a task definition to enable X-Ray integration.

```
{
  "family": "taskB",
  "memory": "256",
  "cpu": "1024",
  "proxyConfiguration": {
    "type": "APPMESH",
    "containerName": "envoy",
    "properties": [
      {
        "name": "IgnoredUID",
        "value": "1337"
      },
      {
        "name": "ProxyIngressPort",
        "value": "15000"
      },
      {
        "name": "ProxyEgressPort",
        "value": "15001"
      },
      {
        "name": "AppPorts",
        "value": "9080"
      },
      {
        "name": "EgressIgnoredIPs",
        "value": "169.254.170.2,169.254.169.254"
      },
      {
        "name": "EgressIgnoredPorts",
        "value": "22"
      }
    ]
  },
  "containerDefinitions": [
    {
      "name": "appName",
      "image": "appImage",
      "portMappings": [
        {
          "containerPort": 9080,
          "hostPort": 9080,
          "protocol": "tcp"
        }
      ],
      "essential": true,
      "dependsOn": [
        {
          "containerName": "envoy",
          "condition": "HEALTHY"
        }
      ]
    }
  ]
}
```

```
    "name": "envoy",
    "image": "840364872350.dkr.ecr.us-west-2.amazonaws.com/aws-appmesh-
envoy:v1.12.3.0-prod",
    "essential": true,
    "environment": [
      {
        "name": "APPMESH_VIRTUAL_NODE_NAME",
        "value": "mesh/apps/virtualNode/serviceB"
      },
      {
        "name": "ENABLE_ENVOY_XRAY_TRACING",
        "value": "1"
      }
    ],
    "healthCheck": {
      "command": [
        "CMD-SHELL",
        "curl -s http://localhost:9901/server_info | grep state | grep -q LIVE"
      ],
      "startPeriod": 10,
      "interval": 5,
      "timeout": 2,
      "retries": 3
    },
    "user": "1337"
  },
  {
    "name": "xray-daemon",
    "image": "amazon/aws-xray-daemon",
    "user": "1337",
    "essential": true,
    "cpu": 32,
    "memoryReservation": 256,
    "portMappings": [
      {
        "containerPort": 2000,
        "protocol": "udp"
      }
    ]
  }
],
"requiresCompatibilities" : [ "EC2" ],
"taskRoleArn" : "arn:aws:iam::123456789012:role/ecsTaskRole",
"executionRoleArn" : "arn:aws:iam::123456789012:role/ecsTaskExecutionRole",
"networkMode": "awsvpc"
}
```


Tutorials for Amazon ECS

The following tutorials show you how to perform common tasks when using Amazon ECS.

Topics

- [Tutorial: Creating a VPC with Public and Private Subnets for Your Clusters \(p. 305\)](#)
- [Tutorial: Creating a Cluster with a Fargate Task Using the AWS CLI \(p. 307\)](#)
- [Tutorial: Specifying Sensitive Data Using Secrets Manager Secrets \(p. 312\)](#)
- [Tutorial: Creating a Service Using Service Discovery \(p. 317\)](#)
- [Tutorial: Creating a Service Using a Blue/Green Deployment \(p. 327\)](#)
- [Tutorial: Listening for Amazon ECS CloudWatch Events \(p. 335\)](#)
- [Tutorial: Sending Amazon Simple Notification Service Alerts for Task Stopped Events \(p. 337\)](#)

Tutorial: Creating a VPC with Public and Private Subnets for Your Clusters

Container instances in your clusters need external network access to communicate with the Amazon ECS service endpoint. However, you might have tasks and services that you would like to run in private subnets. Creating a VPC with both public and private subnets provides you the flexibility to launch tasks and services in either a public or private subnet. Tasks and services in the private subnets can access the internet through a NAT gateway. Services in both the public and private subnets can be configured to use a load balancer so that they can still be reached from the public internet.

This tutorial guides you through creating a VPC with two public subnets and two private subnets, which are provided with internet access through a NAT gateway.

Step 1: Create an Elastic IP Address for Your NAT Gateway

A NAT gateway requires an Elastic IP address in your public subnet, but the VPC wizard does not create one for you. Create the Elastic IP address before running the VPC wizard.

To create an Elastic IP address

1. Open the Amazon VPC console at <https://console.aws.amazon.com/vpc/>.
2. In the left navigation pane, choose **Elastic IPs**.
3. Choose **Allocate new address**, **Allocate**, **Close**.
4. Note the **Allocation ID** for your newly created Elastic IP address; you enter this later in the VPC wizard.

Step 2: Run the VPC Wizard

The VPC wizard automatically creates and configures most of your VPC resources for you.

To run the VPC wizard

1. In the left navigation pane, choose **VPC Dashboard**.

2. Choose **Launch VPC Wizard, VPC with Public and Private Subnets, Select**.
3. For **VPC name**, give your VPC a unique name.
4. For **Elastic IP Allocation ID**, choose the ID of the Elastic IP address that you created earlier.
5. Choose **Create VPC**.
6. When the wizard is finished, choose **OK**. Note the Availability Zone in which your VPC subnets were created. Your additional subnets should be created in a different Availability Zone.

Non-default subnets, such as those created by the VPC wizard, are not auto-assigned public IPv4 addresses. Instances launched in the public subnet must be assigned a public IPv4 address to communicate with the Amazon ECS service endpoint.

To modify your public subnet's IPv4 addressing behavior

1. In the left navigation pane, choose **Subnets**.
2. Select the public subnet for your VPC. By default, the name created by the VPC wizard is **Public subnet**.
3. Choose **Actions, Modify auto-assign IP settings**.
4. Select the **Enable auto-assign public IPv4 address** check box, and then choose **Save**.

Step 3: Create Additional Subnets

The wizard creates a VPC with a single public and a single private subnet in a single Availability Zone. For greater availability, you should create at least one more of each subnet type in a different Availability Zone so that your VPC has both public and private subnets across two Availability Zones.

To create an additional private subnet

1. In the left navigation pane, choose **Subnets**.
2. Choose **Create Subnet**.
3. For **Name tag**, enter a name for your subnet, such as **Private subnet**.
4. For **VPC**, choose the VPC that you created earlier.
5. For **Availability Zone**, choose a different Availability Zone than your original subnets in the VPC.
6. For **IPv4 CIDR block**, enter a valid CIDR block. For example, the wizard creates CIDR blocks in 10.0.0.0/24 and 10.0.1.0/24 by default. You could use **10.0.3.0/24** for your second private subnet.
7. Choose **Yes, Create**.

To create an additional public subnet

1. In the left navigation pane, choose **Subnets** and then **Create Subnet**.
2. For **Name tag**, enter a name for your subnet, such as **Public subnet**.
3. For **VPC**, choose the VPC that you created earlier.
4. For **Availability Zone**, choose the same Availability Zone as the additional private subnet that you created in the previous procedure.
5. For **IPv4 CIDR block**, enter a valid CIDR block. For example, the wizard creates CIDR blocks in 10.0.0.0/24 and 10.0.1.0/24 by default. You could use **10.0.2.0/24** for your second public subnet.
6. Choose **Yes, Create**.
7. Select the public subnet that you just created and choose **Route Table, Edit**.
8. By default, the private route table is selected. Choose the other available route table so that the **0.0.0.0/0** destination is routed to the internet gateway (**igw-xxxxxxx**) and choose **Save**.

9. With your second public subnet still selected, choose **Subnet Actions, Modify auto-assign IP settings**.
10. Select **Enable auto-assign public IPv4 address** and choose **Save, Close**.

Next Steps

After you have created your VPC, you should consider the following next steps:

- Create security groups for your public and private resources if they require inbound network access. For more information, see [Working with Security Groups](#) in the *Amazon VPC User Guide*.
- Create Amazon ECS clusters in your private or public subnets. For more information, see [Creating a Cluster \(p. 18\)](#). If you use the cluster creation wizard in the Amazon ECS console, you can specify the VPC that you just created and the public or private subnets in which to launch your instances, depending on your use case.
 - To make your containers directly accessible from the internet, launch instances into your *public* subnets. Be sure to configure your container instance security groups appropriately.
 - To avoid making containers directly accessible from the internet, launch instances into your *private* subnets.
- Create a load balancer in your public subnets that can route traffic to services in your public or private subnets. For more information, see [Service Load Balancing \(p. 152\)](#).

Tutorial: Creating a Cluster with a Fargate Task Using the AWS CLI

The following steps help you set up a cluster, register a task definition, run a task, and perform other common scenarios in Amazon ECS with the AWS CLI. Ensure that you are using the latest version of the AWS CLI. For more information on how to upgrade to the latest version, see [Installing the AWS Command Line Interface](#).

Topics

- [Prerequisites \(p. 307\)](#)
- [Step 1: Create a Cluster \(p. 308\)](#)
- [Step 2: Register a Task Definition \(p. 308\)](#)
- [Step 3: List Task Definitions \(p. 309\)](#)
- [Step 4: Create a Service \(p. 309\)](#)
- [Step 5: List Services \(p. 310\)](#)
- [Step 6: Describe the Running Service \(p. 310\)](#)
- [Step 7: Clean Up \(p. 312\)](#)

Prerequisites

This tutorial assumes that the following prerequisites have been completed.

- The latest version of the AWS CLI is installed and configured. For more information about installing or upgrading your AWS CLI, see [Installing the AWS Command Line Interface](#).
- The steps in [Setting Up with Amazon ECS \(p. 3\)](#) have been completed.
- Your AWS user has the required permissions specified in the [Amazon ECS First Run Wizard Permissions \(p. 212\)](#) IAM policy example.

- You have a VPC and security group created to use. This tutorial uses a container image hosted on Docker Hub so your task must have internet access. To give your task a route to the internet, use one of the following options.
- Use a private subnet with a NAT gateway that has an elastic IP address.
- Use a public subnet and assign a public IP address to the task.

For more information, see [Tutorial: Creating a VPC with Public and Private Subnets for Your Clusters](#).

Step 1: Create a Cluster

By default, your account receives a default cluster.

Note

The benefit of using the default cluster that is provided for you is that you don't have to specify the `--cluster` *cluster_name* option in the subsequent commands. If you do create your own, non-default, cluster, you must specify `--cluster` *cluster_name* for each command that you intend to use with that cluster.

Create your own cluster with a unique name with the following command:

```
aws ecs create-cluster --cluster-name fargate-cluster
```

Output:

```
{
  "cluster": {
    "status": "ACTIVE",
    "statistics": [],
    "clusterName": "fargate-cluster",
    "registeredContainerInstancesCount": 0,
    "pendingTasksCount": 0,
    "runningTasksCount": 0,
    "activeServicesCount": 0,
    "clusterArn": "arn:aws:ecs:region:aws_account_id:cluster/fargate-cluster"
  }
}
```

Step 2: Register a Task Definition

Before you can run a task on your ECS cluster, you must register a task definition. Task definitions are lists of containers grouped together. The following example is a simple task definition that creates a PHP web app using the httpd container image hosted on Docker Hub. For more information about the available task definition parameters, see [Amazon ECS Task Definitions \(p. 25\)](#).

```
{
  "family": "sample-fargate",
  "networkMode": "awsvpc",
  "containerDefinitions": [
    {
      "name": "fargate-app",
      "image": "httpd:2.4",
      "portMappings": [
        {
          "containerPort": 80,
          "hostPort": 80,
          "protocol": "tcp"
        }
      ]
    }
  ]
}
```

```
        ],
        "essential": true,
        "entryPoint": [
            "sh",
            "-c"
        ],
        "command": [
            "/bin/sh -c \"echo '<html> <head> <title>Amazon ECS Sample App</title>
<style>body {margin-top: 40px; background-color: #333;} </style> </head><body> <div
style=color:white;text-align:center> <h1>Amazon ECS Sample App</h1> <h2>Congratulations!
</h2> <p>Your application is now running on a container in Amazon ECS.</p> </div></body></
html>' > /usr/local/apache2/htdocs/index.html && httpd-foreground\""
        ]
    },
    ],
    "requiresCompatibilities": [
        "FARGATE"
    ],
    "cpu": "256",
    "memory": "512"
}
```

The above example JSON can be passed to the AWS CLI in two ways: You can save the task definition JSON as a file and pass it with the `--cli-input-json file://path_to_file.json` option. Or, you can escape the quotation marks in the JSON and pass the JSON container definitions on the command line as in the below example. If you choose to pass the container definitions on the command line, your command additionally requires a `--family` parameter that is used to keep multiple versions of your task definition associated with each other.

To use a JSON file for container definitions:

```
aws ecs register-task-definition --cli-input-json file://$HOME/tasks/fargate-task.json
```

The **register-task-definition** command returns a description of the task definition after it completes its registration.

Step 3: List Task Definitions

You can list the task definitions for your account at any time with the **list-task-definitions** command. The output of this command shows the `family` and `revision` values that you can use together when calling **run-task** or **start-task**.

```
aws ecs list-task-definitions
```

Output:

```
{
  "taskDefinitionArns": [
    "arn:aws:ecs:region:aws_account_id:task-definition/sample-fargate:1"
  ]
}
```

Step 4: Create a Service

After you have registered a task for your account, you can create a service for the registered task in your cluster. For this example, you create a service with one instance of the `sample-fargate:1` task definition running in your cluster. The task requires a route to the internet, so there are two ways you can

achieve this. One way is to use a private subnet configured with a NAT gateway with an elastic IP address in a public subnet. Another way is to use a public subnet and assign a public IP address to your task. We provide both examples below.

Example using a private subnet.

```
aws ecs create-service --cluster fargate-cluster --service-name fargate-service --  
task-definition sample-fargate:1 --desired-count 1 --launch-type "FARGATE" --network-  
configuration "awsvpcConfiguration={subnets=[subnet-abcd1234],securityGroups=[sg-  
abcd1234]}"
```

Example using a public subnet.

```
aws ecs create-service --cluster fargate-cluster --service-name fargate-service --  
task-definition sample-fargate:1 --desired-count 1 --launch-type "FARGATE" --network-  
configuration "awsvpcConfiguration={subnets=[subnet-abcd1234],securityGroups=[sg-  
abcd1234],assignPublicIp=ENABLED}"
```

The **create-service** command returns a description of the task definition after it completes its registration.

Step 5: List Services

List the services for your cluster. You should see the service that you created in the previous section. You can take the service name or the full ARN that is returned from this command and use it to describe the service later.

```
aws ecs list-services --cluster fargate-cluster
```

Output:

```
{  
  "serviceArns": [  
    "arn:aws:ecs:region:aws_account_id:service/fargate-service"  
  ]  
}
```

Step 6: Describe the Running Service

Describe the service using the service name retrieved earlier to get more information about the task.

```
aws ecs describe-services --cluster fargate-cluster --services fargate-service
```

If successful, this will return a description of the service failures and services. For example, in services section, you will find information on deployments, such as the status of the tasks as running or pending. You may also find information on the task definition, the network configuration and time-stamped events. In the failures section, you will find information on failures, if any, associated with the call. For troubleshooting, see [Service Event Messages](#). For more information about the service description, see [Describe Services](#).

```
{  
  "services": [  
    {  
      "status": "ACTIVE",  
      "taskDefinition": "arn:aws:ecs:region:aws_account_id:task-definition/sample-  
fargate:1",  
    }  
  ]  
}
```

```

        "pendingCount": 2,
        "launchType": "FARGATE",
        "loadBalancers": [],
        "roleArn": "arn:aws:iam::aws_account_id:role/aws-service-role/
ecs.amazonaws.com/AWSServiceRoleForECS",
        "placementConstraints": [],
        "createdAt": 1510811361.128,
        "desiredCount": 2,
        "networkConfiguration": {
            "awsvpcConfiguration": {
                "subnets": [
                    "subnet-abcd1234"
                ],
                "securityGroups": [
                    "sg-abcd1234"
                ],
                "assignPublicIp": "DISABLED"
            }
        },
        "platformVersion": "LATEST",
        "serviceName": "fargate-service",
        "clusterArn": "arn:aws:ecs:region:aws_account_id:cluster/fargate-cluster",
        "serviceArn": "arn:aws:ecs:region:aws_account_id:service/fargate-service",
        "deploymentConfiguration": {
            "maximumPercent": 200,
            "minimumHealthyPercent": 100
        },
        "deployments": [
            {
                "status": "PRIMARY",
                "networkConfiguration": {
                    "awsvpcConfiguration": {
                        "subnets": [
                            "subnet-abcd1234"
                        ],
                        "securityGroups": [
                            "sg-abcd1234"
                        ],
                        "assignPublicIp": "DISABLED"
                    }
                },
                "pendingCount": 2,
                "launchType": "FARGATE",
                "createdAt": 1510811361.128,
                "desiredCount": 2,
                "taskDefinition": "arn:aws:ecs:region:aws_account_id:task-definition/
sample-fargate:1",
                "updatedAt": 1510811361.128,
                "platformVersion": "0.0.1",
                "id": "ecs-svc/9223370526043414679",
                "runningCount": 0
            }
        ],
        "events": [
            {
                "message": "(service fargate-service) has started 2 tasks: (task
53c0de40-ea3b-489f-a352-623bf1235f08) (task d0aec985-901b-488f-9fb4-61b991b332a3).",
                "id": "92b8443e-67fb-4886-880c-07e73383ea83",
                "createdAt": 1510811841.408
            },
            {
                "message": "(service fargate-service) has started 2 tasks: (task
b4911bee-7203-4113-99d4-e89ba457c626) (task cc5853e3-6e2d-4678-8312-74f8a7d76474).",
                "id": "d85c6ec6-a693-43b3-904a-a997e1fc844d",
                "createdAt": 1510811601.938
            }
        ],

```

```
{
  "message": "(service fargate-service) has started 2 tasks: (task cba86182-52bf-42d7-9df8-b744699e6cfc) (task f4c1ad74-a5c6-4620-90cf-2aff118df5fc).",
  "id": "095703e1-0ca3-4379-a7c8-c0f1b8b95ace",
  "createdAt": 1510811364.691
},
{
  "runningCount": 0,
  "placementStrategy": []
},
{
  "failures": []
}
```

Step 7: Clean Up

When you are finished with this tutorial, you should clean up the associated resources to avoid incurring charges for unused resources.

Delete the service.

```
aws ecs delete-service --cluster fargate-cluster --service fargate-service --force
```

Delete the cluster.

```
aws ecs delete-cluster --cluster fargate-cluster
```

Tutorial: Specifying Sensitive Data Using Secrets Manager Secrets

Amazon ECS enables you to inject sensitive data into your containers by storing your sensitive data in AWS Secrets Manager secrets and then referencing them in your container definition. For more information, see [Specifying Sensitive Data](#) (p. 87).

The following tutorial shows how to create an Secrets Manager secret, reference the secret in an Amazon ECS task definition, and then verify it worked by querying the environment variable inside a container showing the contents of the secret.

Prerequisites

This tutorial assumes that the following prerequisites have been completed:

- The steps in [Setting Up with Amazon ECS](#) (p. 3) have been completed.
- Your AWS user has the required IAM permissions to create the Secrets Manager and Amazon ECS resources described.

Step 1: Create an Secrets Manager Secret

You can use the Secrets Manager console to create a secret for your sensitive data. In this tutorial we will be creating a basic secret for storing a username and password to reference later in a container. For more information, see [Creating a Basic Secret](#) in the *AWS Secrets Manager User Guide*.

To create a basic secret

Use Secrets Manager to create a secret for your sensitive data.

1. Open the Secrets Manager console at <https://console.aws.amazon.com/secretsmanager/>.
2. Choose **Store a new secret**.
3. For **Select secret type**, choose **Other type of secrets**.
4. For **Specify the key/value pairs to be stored in this secret**, choose the **Plaintext** tab and replace the existing text with the following text. The text value you specify here will be the environment variable value in your container at the end of the tutorial.

```
password_value
```

5. Choose **Next**.
6. For **Secret name**, type `username_value` and choose **Next**. The secret name value you specify here will be the environment variable name in your container at the end of the tutorial.
7. For **Configure automatic rotation**, leave **Disable automatic rotation** selected and choose **Next**.
8. Review these settings, and then choose **Store** to save everything you entered as a new secret in Secrets Manager.
9. Select the secret you just created and save the **Secret ARN** to reference in your task execution IAM policy and task definition in later steps.

Step 2: Update Your Task Execution IAM Role

In order for Amazon ECS to retrieve the sensitive data from your Secrets Manager secret, you must have the Amazon ECS task execution role and reference it in your task definition. This allows the container agent to pull the necessary Secrets Manager resources. If you have not already created your task execution IAM role, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

The following steps assume you already have the task execution IAM role created and properly configured.

To update your task execution IAM role

Use the IAM console to update your task execution role with the required permissions.

1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane, choose **Roles**.
3. Search the list of roles for `ecsTaskExecutionRole` and select it.
4. Choose **Permissions, Add inline policy**.
5. Choose the **JSON** tab and specify the following JSON text, ensuring that you specify the full ARN of the Secrets Manager secret you created in step 1.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "secretsmanager:GetSecretValue"
      ],
      "Resource": [
        "arn:aws:secretsmanager:region:aws_account_id:secret:username_value-
u9bH6K"
      ]
    }
  ]
}
```

```
}  
]  
}
```

6. Choose **Review policy**. For **Name** specify `ECSSecretsTutorial`, then choose **Create policy**.

Step 3: Create an Amazon ECS Task Definition

You can use the Amazon ECS console to create a task definition that references a Secrets Manager secret.

To create a task definition that specifies a secret

Use the IAM console to update your task execution role with the required permissions.

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation pane, choose **Task Definitions**, **Create new Task Definition**.
3. On the **Select launch type compatibility** page, choose **EC2** and choose **Next step**.
4. Choose **Configure via JSON** and enter the following task definition JSON text, ensuring that you specify the full ARN of the Secrets Manager secret you created in step 1 and the task execution IAM role you updated in step 2. Choose **Save**.

Important

The value for the secret name in the task definition must match the name you specified for the secret name when the secret was created.

```
{  
  "executionRoleArn": "arn:aws:iam::aws_account_id:role/ecstaskExecutionRole",  
  "containerDefinitions": [  
    {  
      "entryPoint": [  
        "sh",  
        "-c"  
      ],  
      "portMappings": [  
        {  
          "hostPort": 80,  
          "protocol": "tcp",  
          "containerPort": 80  
        }  
      ],  
      "command": [  
        "/bin/sh -c \"echo '<html> <head> <title>Amazon ECS Sample  
App</title> <style>body {margin-top: 40px; background-color: #333;} </style> </  
head><body> <div style=color:white;text-align:center> <h1>Amazon ECS Sample App</h1>  
<h2>Congratulations!</h2> <p>Your application is now running on a container in Amazon  
ECS.</p> </div></body></html>' > /usr/local/apache2/htdocs/index.html && httpd-  
foreground\""  
      ],  
      "cpu": 10,  
      "secrets": [  
        {  
          "valueFrom":  
            "arn:aws:secretsmanager:region:aws_account_id:secret:username_value-u9bH6K",  
          "name": "username_value"  
        }  
      ],  
      "memory": 300,  
      "image": "httpd:2.4",  
      "essential": true,  
      "name": "ecs-secrets-container"  
    }  
  ]  
}
```

```
    ],  
    "family": "ecs-secrets-tutorial"  
  }  
}
```

5. Review the settings and then choose **Create**.

Step 4: Create an Amazon ECS Cluster

You can use the Amazon ECS console to create a cluster containing a container instance to run the task on. If you have an existing cluster with at least one container instance registered to it with the available resources to run one instance of the task definition created for this tutorial you can skip to the next step.

For this tutorial we will be creating a cluster with one `t2.micro` container instance using the Amazon ECS-optimized Amazon Linux 2 AMI.

To create a cluster

Use the Amazon ECS console to create a cluster and register one container instance to it.

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. From the navigation bar, select the Region that contains both the Secrets Manager secret and the Amazon ECS task definition you created.
3. In the navigation pane, choose **Clusters**.
4. On the **Clusters** page, choose **Create Cluster**.
5. For **Select cluster compatibility**, choose **EC2 Linux + Networking**, then choose **Next step**.
6. On the **Configure cluster** page, for **Cluster name** enter `ecs-secrets-tutorial`.
7. For **EC2 instance type**, choose **t2.micro**.
8. For **Key pair**, choose a key pair to add to the container instance.

Important

A key pair is required to complete the tutorial, so if you do not already have a key pair created follow the link to the EC2 console to create one.

9. In the **Networking** section, configure the VPC for your cluster. Select an existing VPC or you can choose **Create a new VPC** to use for the tutorial.
 - a. (Optional) If you choose to create a new VPC, for **CIDR Block**, select a CIDR block for your VPC. For more information, see [Your VPC and Subnets](#) in the *Amazon VPC User Guide*.
 - b. For **Subnets**, select the subnets to use for your VPC. You can keep the default settings or you can modify them to meet your needs.
10. For **Container instance IAM role**, choose your existing container instance IAM role or select **Create new role** to have one created for you.
11. Leave all other fields at their default values and choose **Create**.

Step 5: Run an Amazon ECS Task

You can use the Amazon ECS console to run a task using the task definition you created. For this tutorial we will be running a task using the EC2 launch type, using the cluster we created in the previous step.

To run a task

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. In the navigation pane, choose **Task Definitions** and select the **ecs-secrets-tutorial** task definition we created.

3. Select the latest revision of the task definition and then choose **Actions, Run Task**.
4. For **Launch Type**, choose **EC2**.
5. For **Cluster**, choose the **ecs-secrets-tutorial** cluster we created in the previous step.
6. For **Task tagging configuration**, deselect **Enable ECS managed tags**. They are unnecessary for the purposes of this tutorial.
7. Review your task information and choose **Run Task**.

Note

If your task moves from **PENDING** to **STOPPED**, or if it displays a **PENDING** status and then disappears from the listed tasks, your task may be stopping due to an error. For more information, see [Checking stopped tasks for errors \(p. 340\)](#) in the troubleshooting section.

Step 6: Verify

You can verify all of the steps were completed successfully and the environment variable was created properly in your container using the following steps.

To verify that the environment variable was created

1. Find the public IP or DNS address for your container instance.
 - a. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
 - b. Select the **ecs-secrets-tutorial** cluster that hosts your container instance.
 - c. On the **Cluster** page, choose **ECS Instances**.
 - d. On the **Container Instance** column, select the container instance to connect to.
 - e. On the **Container Instance** page, record the **Public IP** or **Public DNS** for your instance.
2. If you are using a macOS or Linux computer, connect to your instance with the following command, substituting the path to your private key and the public address for your instance:

```
$ ssh -i /path/to/my-key-pair.pem ec2-user@ec2-198-51-100-1.compute-1.amazonaws.com
```

For more information about using a Windows computer, see [Connecting to Your Linux Instance from Windows Using PuTTY](#) in the *Amazon EC2 User Guide for Linux Instances*.

Important

For more information about any issues while connecting to your instance, see [Troubleshooting Connecting to Your Instance](#) in the *Amazon EC2 User Guide for Linux Instances*.

3. List the containers running on the instance. Note the container ID for **ecs-secrets-tutorial** container.

```
docker ps
```

4. Connect to the **ecs-secrets-tutorial** container using the container ID from the output of the previous step.

```
docker exec -it container_ID /bin/bash
```

5. Use the **echo** command to print the value of the environment variable.

```
echo $username_value
```

If the tutorial was successful, you should see the following output:

```
password_value
```

Note

Alternatively, you can list all environment variables in your container using the `env` (or `printenv`) command.

Step 7: Clean Up

When you are finished with this tutorial, you should clean up the associated resources to avoid incurring charges for unused resources.

To clean up the resources

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. Select the **ecs-secrets-tutorial cluster** you created.
3. On the **Cluster** page, choose **Delete Cluster**.
4. Enter the delete cluster confirmation phrase and choose **Delete**. This will take several minutes but will clean up all of the Amazon ECS cluster resources.
5. Open the IAM console at <https://console.aws.amazon.com/iam/>.
6. In the navigation pane, choose **Roles**.
7. Search the list of roles for `ecsTaskExecutionRole` and select it.
8. Choose **Permissions**, then choose the **X** next to `ECSecretsTutorial`. Choose **Remove** to confirm the removal of the inline policy.
9. Open the Secrets Manager console at <https://console.aws.amazon.com/secretsmanager/>.
10. Select the `username_value` secret you created and choose **Actions, Delete secret**.

Tutorial: Creating a Service Using Service Discovery

Service discovery has been integrated into the Create Service wizard in the Amazon ECS console. For more information, see [Creating a service](#) (p. 129).

The following tutorial shows how to create an ECS service containing a Fargate task that uses service discovery with the AWS CLI.

For a list of Regions that support service discovery, see [Service Discovery](#) (p. 173).

Fargate tasks are only supported in the following Regions:

Region Name	Region
US East (Ohio)	us-east-2
US East (N. Virginia)	us-east-1
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Africa (Cape Town)	af-south-1

Region Name	Region
Asia Pacific (Hong Kong)	ap-east-1
Asia Pacific (Mumbai)	ap-south-1 (aps1-az1 & aps1-az3 only)
Asia Pacific (Seoul)	ap-northeast-2
Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
Canada (Central)	ca-central-1
China (Beijing)	cn-north-1
China (Ningxia)	cn-northwest-1
Europe (Frankfurt)	eu-central-1
Europe (Ireland)	eu-west-1
Europe (London)	eu-west-2
Europe (Paris)	eu-west-3
Europe (Stockholm)	eu-north-1
South America (São Paulo)	sa-east-1
Middle East (Bahrain)	me-south-1
AWS GovCloud (US-East)	us-gov-east-1
AWS GovCloud (US-West)	us-gov-west-1

Prerequisites

This tutorial assumes that the following prerequisites have been completed:

- The latest version of the AWS CLI is installed and configured. For more information, see [Installing the AWS Command Line Interface](#).
- The steps in [Setting Up with Amazon ECS \(p. 3\)](#) have been completed.
- Your AWS user has the required permissions specified in the [Amazon ECS First Run Wizard Permissions \(p. 212\)](#) IAM policy example.
- You have a VPC and security group created to use. For more information, see [Tutorial: Creating a VPC with Public and Private Subnets for Your Clusters](#).

Step 1: Create the Service Discovery Resources

Use the following steps to create your service discovery namespace and service discovery service.

To create the Service Discovery resources

1. Create a private service discovery namespace named `tutorial` within one of your existing VPCs:

```
aws servicediscovery create-private-dns-namespace --name tutorial --vpc vpc-abcd1234 --region us-east-1
```

Output:

```
{
  "OperationId": "h2qe3s6dxftvvt7riu6lfy2f6c3jlhf4-je6chs2e"
}
```

2. Using the OperationId from the previous output, verify that the private namespace was created successfully. Copy the namespace ID as it is used in subsequent commands.

```
aws servicediscovery get-operation --operation-id h2qe3s6dxftvvt7riu6lfy2f6c3jlhf4-je6chs2e
```

Output:

```
{
  "Operation": {
    "Id": "h2qe3s6dxftvvt7riu6lfy2f6c3jlhf4-je6chs2e",
    "Type": "CREATE_NAMESPACE",
    "Status": "SUCCESS",
    "CreateDate": 1519777852.502,
    "UpdateDate": 1519777856.086,
    "Targets": {
      "NAMESPACE": "ns-uejictsjen2i4eeg"
    }
  }
}
```

3. Using the NAMESPACE ID from the previous output, create a service discovery service named myapplication. Copy the service discovery service ID as it is used in subsequent commands:

```
aws servicediscovery create-service --name myapplication --dns-config 'NamespaceId="ns-uejictsjen2i4eeg",DnsRecords=[{Type="A",TTL="300"}]' --health-check-custom-config FailureThreshold=1 --region us-east-1
```

Output:

```
{
  "Service": {
    "Id": "srv-utcrh6wavdkggqtk",
    "Arn": "arn:aws:servicediscovery:region:aws_account_id:service/srv-utcrh6wavdkggqtk",
    "Name": "myapplication",
    "DnsConfig": {
      "NamespaceId": "ns-uejictsjen2i4eeg",
      "DnsRecords": [
        {
          "Type": "A",
          "TTL": 300
        }
      ]
    },
    "HealthCheckCustomConfig": {
      "FailureThreshold": 1
    },
    "CreatorRequestId": "e49a8797-b735-481b-a657-b74d1d6734eb"
  }
}
```

```
}
```

Step 2: Create the Amazon ECS Resources

Use the following steps to create your Amazon ECS cluster, task definition, and service.

To create Amazon ECS resources

1. Create an Amazon ECS cluster named `tutorial` to use:

```
aws ecs create-cluster --cluster-name tutorial --region us-east-1
```

Output:

```
{
  "cluster": {
    "clusterArn": "arn:aws:ecs:region:aws_account_id:cluster/tutorial",
    "clusterName": "tutorial",
    "status": "ACTIVE",
    "registeredContainerInstancesCount": 0,
    "runningTasksCount": 0,
    "pendingTasksCount": 0,
    "activeServicesCount": 0,
    "statistics": []
  }
}
```

2. Register a task definition that is compatible with Fargate. It requires the use of the `awsvpc` network mode. The following is the example task definition used for this tutorial.

First, create a file named `fargate-task.json` with the contents of the following task definition:

```
{
  "family": "tutorial-task-def",
  "networkMode": "awsvpc",
  "containerDefinitions": [
    {
      "name": "sample-app",
      "image": "httpd:2.4",
      "portMappings": [
        {
          "containerPort": 80,
          "hostPort": 80,
          "protocol": "tcp"
        }
      ],
      "essential": true,
      "entryPoint": [
        "sh",
        "-c"
      ],
      "command": [
        "/bin/sh -c \"echo '<html> <head> <title>Amazon ECS Sample App</title> <style>body {margin-top: 40px; background-color: #333;} </style> </head><body> <div style=color:white;text-align:center> <h1>Amazon ECS Sample App</h1> <h2>Congratulations!</h2> <p>Your application is now running on a container in Amazon ECS.</p> </div></body></html>' > /usr/local/apache2/htdocs/index.html && httpd-foreground\""
      ]
    }
  ]
}
```



```
    ],
    "requiresCompatibilities": [
        "FARGATE"
    ],
    "cpu": "256",
    "memory": "512"
}
```

Then, register the task definition using the `fargate-task.json` file that you created:

```
aws ecs register-task-definition --cli-input-json file://fargate-task.json --region us-east-1
```

3. Create a file named `ecs-service-discovery.json` with the contents of the ECS service that you are going to create. This example uses the task definition created in the previous step. An `awsvpcConfiguration` is required because the example task definition uses the `awsvpc` network mode.

```
{
  "cluster": "tutorial",
  "serviceName": "ecs-service-discovery",
  "taskDefinition": "tutorial-task-def",
  "serviceRegistries": [
    {
      "registryArn": "arn:aws:servicediscovery:region:aws_account_id:service/srv-utcrh6wavdkggqtk"
    }
  ],
  "launchType": "FARGATE",
  "platformVersion": "LATEST",
  "networkConfiguration": {
    "awsvpcConfiguration": {
      "assignPublicIp": "ENABLED",
      "securityGroups": [ "sg-abcd1234" ],
      "subnets": [ "subnet-abcd1234" ]
    }
  },
  "desiredCount": 1
}
```

Create your ECS service, specifying the Fargate launch type and the `LATEST` platform version, which supports service discovery:

```
aws ecs create-service --cli-input-json file://ecs-service-discovery.json --region us-east-1
```

Output:

```
{
  "service": {
    "serviceArn": "arn:aws:ecs:region:aws_account_id:service/ecs-service-discovery",
    "serviceName": "ecs-service-discovery",
    "clusterArn": "arn:aws:ecs:region:aws_account_id:cluster/tutorial",
    "loadBalancers": [],
    "serviceRegistries": [
      {
        "registryArn": "arn:aws:servicediscovery:region:aws_account_id:service/srv-utcrh6wavdkggqtk"
      }
    ]
  }
}
```

```
    ],
    "status": "ACTIVE",
    "desiredCount": 1,
    "runningCount": 0,
    "pendingCount": 0,
    "launchType": "FARGATE",
    "platformVersion": "LATEST",
    "taskDefinition": "arn:aws:ecs:region:aws_account_id:task-definition/tutorial-
task-def:1",
    "deploymentConfiguration": {
      "maximumPercent": 200,
      "minimumHealthyPercent": 100
    },
    "deployments": [
      {
        "id": "ecs-svc/9223370516993140842",
        "status": "PRIMARY",
        "taskDefinition": "arn:aws:ecs:region:aws_account_id:task-definition/
tutorial-task-def:1",
        "desiredCount": 1,
        "pendingCount": 0,
        "runningCount": 0,
        "createdAt": 1519861634.965,
        "updatedAt": 1519861634.965,
        "launchType": "FARGATE",
        "platformVersion": "1.1.0",
        "networkConfiguration": {
          "awsvpcConfiguration": {
            "subnets": [
              "subnet-abcd1234"
            ],
            "securityGroups": [
              "sg-abcd1234"
            ],
            "assignPublicIp": "ENABLED"
          }
        }
      }
    ]
  },
  "roleArn": "arn:aws:iam::aws_account_id:role/ECSServiceLinkedRole",
  "events": [],
  "createdAt": 1519861634.965,
  "placementConstraints": [],
  "placementStrategy": [],
  "networkConfiguration": {
    "awsvpcConfiguration": {
      "subnets": [
        "subnet-abcd1234"
      ],
      "securityGroups": [
        "sg-abcd1234"
      ],
      "assignPublicIp": "ENABLED"
    }
  }
}
```

Step 3: Verify Service Discovery

You can verify that everything has been created properly by querying your service discovery information. After service discovery is configured, you can query it using either the AWS Cloud Map API operations or by using `dig` from within your VPC, as described below.

To verify service discovery configuration

1. Using the service discovery service ID, list the service discovery instances:

```
aws servicediscovery list-instances --service-id srv-utcrh6wavdkggqtk --region us-east-1
```

Output:

```
{
  "Instances": [
    {
      "Id": "16becc26-8558-4af1-9fbd-f81be062a266",
      "Attributes": {
        "AWS_INSTANCE_IPV4": "172.31.87.2",
        "AWS_INSTANCE_PORT": "80",
        "AVAILABILITY_ZONE": "us-east-1a",
        "REGION": "us-east-1",
        "ECS_SERVICE_NAME": "ecs-service-discovery",
        "ECS_CLUSTER_NAME": "tutorial",
        "ECS_TASK_DEFINITION_FAMILY": "tutorial-task-def"
      }
    }
  ]
}
```

2. Using the service discovery namespace and service, use additional parameters to query the details about the service discovery instances:

```
aws servicediscovery discover-instances --namespace-name tutorial --service-name myapplication --query-parameters ECS_CLUSTER_NAME=tutorial --region us-east-1
```

Output:

```
{
  "Instances": [
    {
      "InstanceId": "16becc26-8558-4af1-9fbd-f81be062a266",
      "NamespaceName": "tutorial",
      "ServiceName": "ecs-service-discovery",
      "HealthStatus": "HEALTHY",
      "Attributes": {
        "AWS_INSTANCE_IPV4": "172.31.87.2",
        "AWS_INSTANCE_PORT": "80",
        "AVAILABILITY_ZONE": "us-east-1a",
        "REGION": "us-east-1",
        "ECS_SERVICE_NAME": "ecs-service-discovery",
        "ECS_CLUSTER_NAME": "tutorial",
        "ECS_TASK_DEFINITION_FAMILY": "tutorial-task-def"
      }
    }
  ]
}
```

3. The DNS records created in the Route 53 hosted zone for the service discovery service can be queried with the following AWS CLI commands.

Using the namespace ID, get information about the namespace, which includes the Route 53 hosted zone ID:

```
aws servicediscovery get-namespace --id ns-uejictsjen2i4eeg --region us-east-1
```

Output:

```
{
  "Namespace": {
    "Id": "ns-uejictsjen2i4eeg",
    "Arn": "arn:aws:servicediscovery:region:aws_account_id:namespace/ns-uejictsjen2i4eeg",
    "Name": "tutorial",
    "Type": "DNS_PRIVATE",
    "Properties": {
      "DnsProperties": {
        "HostedZoneId": "Z35JQ4ZFDRYPLV"
      }
    },
    "CreateDate": 1519777852.502,
    "CreatorRequestId": "9049a1d5-25e4-4115-8625-96dbda9a6093"
  }
}
```

4. Using the Route 53 hosted zone ID, get the resource record set for the hosted zone:

```
aws route53 list-resource-record-sets --hosted-zone-id Z35JQ4ZFDRYPLV --region us-east-1
```

Output:

```
{
  "ResourceRecordSets": [
    {
      "Name": "tutorial.",
      "Type": "NS",
      "TTL": 172800,
      "ResourceRecords": [
        {
          "Value": "ns-1536.awsdns-00.co.uk."
        },
        {
          "Value": "ns-0.awsdns-00.com."
        },
        {
          "Value": "ns-1024.awsdns-00.org."
        },
        {
          "Value": "ns-512.awsdns-00.net."
        }
      ]
    },
    {
      "Name": "tutorial.",
      "Type": "SOA",
      "TTL": 900,
      "ResourceRecords": [
        {

```

```
        "Value": "ns-1536.awsdns-00.co.uk. awsdns-hostmaster.amazon.com. 1
7200 900 1209600 86400"
      }
    ],
    },
    {
      "Name": "myapplication.tutorial.",
      "Type": "A",
      "SetIdentifier": "16becc26-8558-4af1-9fbd-f81be062a266",
      "MultiValueAnswer": true,
      "TTL": 300,
      "ResourceRecords": [
        {
          "Value": "172.31.87.2"
        }
      ]
    }
  ]
}
```

5. You can also query the DNS using `dig` from an instance within your VPC with the following command:

```
dig +short myapplication.tutorial
```

Output:

```
172.31.87.2
```

Step 4: Clean Up

When you are finished with this tutorial, you should clean up the associated resources to avoid incurring charges for unused resources.

To clean up the service discovery instances and Amazon ECS resources

1. Deregister the service discovery service instances:

```
aws servicediscovery deregister-instance --service-id srv-utcrh6wavdkggqtk --instance-
id 16becc26-8558-4af1-9fbd-f81be062a266 --region us-east-1
```

Output:

```
{
  "OperationId": "xhu73bsertlyffhm3faqi7kumsmx274n-jh0zimzv"
}
```

2. Using the `OperationId` from the previous output, verify that the service discovery service instances were deregistered successfully:

```
aws servicediscovery get-operation --operation-id xhu73bsertlyffhm3faqi7kumsmx274n-
jh0zimzv --region us-east-1
```

Output:

```
{
```

```
"Operation": {
  "Id": "xhu73bsertlyffhm3faqi7kumsmx274n-jh0zimzv",
  "Type": "DEREGISTER_INSTANCE",
  "Status": "SUCCESS",
  "CreateDate": 1525984073.707,
  "UpdateDate": 1525984076.426,
  "Targets": {
    "INSTANCE": "i-16becc26-8558-4af1-9fbd-f81be062a266",
    "ROUTE_53_CHANGE_ID": "C5NSRG1J4I1FH",
    "SERVICE": "srv-utcrh6wavdkggqtk"
  }
}
```

3. Delete the service discovery service:

```
aws servicediscovery delete-service --id srv-utcrh6wavdkggqtk --region us-east-1
```

4. Delete the service discovery namespace:

```
aws servicediscovery delete-namespace --id ns-uejictsjen2i4eeg --region us-east-1
```

Output:

```
{
  "OperationId": "c3ncqglftesw4ibgj5baz6ktaoh6cg4t-jh0ztysj"
}
```

5. Using the OperationId from the previous output, verify that the service discovery namespace was deleted successfully:

```
aws servicediscovery get-operation --operation-id c3ncqglftesw4ibgj5baz6ktaoh6cg4t-jh0ztysj --region us-east-1
```

Output:

```
{
  "Operation": {
    "Id": "c3ncqglftesw4ibgj5baz6ktaoh6cg4t-jh0ztysj",
    "Type": "DELETE_NAMESPACE",
    "Status": "SUCCESS",
    "CreateDate": 1525984602.211,
    "UpdateDate": 1525984602.558,
    "Targets": {
      "NAMESPACE": "ns-rymlehshst7hhukh",
      "ROUTE_53_CHANGE_ID": "CJP2A2M86XW3O"
    }
  }
}
```

6. Update the Amazon ECS service so that the desired count is 0, which allows you to delete it:

```
aws ecs update-service --cluster tutorial --service ecs-service-discovery --desired-count 0 --force-new-deployment --region us-east-1
```

7. Delete the Amazon ECS service:

```
aws ecs delete-service --cluster tutorial --service ecs-service-discovery --region us-east-1
```

8. Delete the Amazon ECS cluster:

```
aws ecs delete-cluster --cluster tutorial --region us-east-1
```

Tutorial: Creating a Service Using a Blue/Green Deployment

Amazon ECS has integrated blue/green deployments into the Create Service wizard on the Amazon ECS console. For more information, see [Creating a service \(p. 129\)](#).

The following tutorial shows how to create an Amazon ECS service containing a Fargate task that uses the blue/green deployment type with the AWS CLI.

Prerequisites

This tutorial assumes that you have completed the following prerequisites:

- The latest version of the AWS CLI is installed and configured. For more information about installing or upgrading the AWS CLI, see [Installing the AWS Command Line Interface](#).
- The steps in [Setting Up with Amazon ECS \(p. 3\)](#) have been completed.
- Your AWS user has the required permissions specified in the [Amazon ECS First Run Wizard Permissions \(p. 212\)](#) IAM policy example.
- You have a VPC and security group created to use. For more information, see [Tutorial: Creating a VPC with Public and Private Subnets for Your Clusters](#).
- The Amazon ECS CodeDeploy IAM role is created. For more information, see [Amazon ECS CodeDeploy IAM Role \(p. 243\)](#).

Step 1: Create an Application Load Balancer

Amazon ECS services using the blue/green deployment type require the use of either an Application Load Balancer or a Network Load Balancer. This tutorial uses an Application Load Balancer.

To create an Application Load Balancer

1. Use the [create-load-balancer](#) command to create an Application Load Balancer. Specify two subnets that aren't from the same Availability Zone as well as a security group.

```
aws elbv2 create-load-balancer \
  --name bluegreen-alb \
  --subnets subnet-abcd1234 subnet-abcd5678 \
  --security-groups sg-abcd1234 \
  --region us-east-1
```

The output includes the Amazon Resource Name (ARN) of the load balancer, with the following format:

```
arn:aws:elasticloadbalancing:region:aws_account_id:loadbalancer/app/bluegreen-alb/e5ba62739c16e642
```

2. Use the [create-target-group](#) command to create a target group. This target group will route traffic to the original task set in your service.

```
aws elbv2 create-target-group \  
  --name bluegreentarget1 \  
  --protocol HTTP \  
  --port 80 \  
  --target-type ip \  
  --vpc-id vpc-abcd1234 \  
  --region us-east-1
```

The output includes the ARN of the target group, with the following format:

```
arn:aws:elasticloadbalancing:region:aws_account_id:targetgroup/  
bluegreentarget1/209a844cd01825a4
```

3. Use the [create-listener](#) command to create a load balancer listener with a default rule that forwards requests to the target group.

```
aws elbv2 create-listener \  
  --load-balancer-arn  
  arn:aws:elasticloadbalancing:region:aws_account_id:loadbalancer/app/bluegreen-alb/  
e5ba62739c16e642 \  
  --protocol HTTP \  
  --port 80 \  
  --default-actions  
  Type=forward,TargetGroupArn=arn:aws:elasticloadbalancing:region:aws_account_id:targetgroup/  
bluegreentarget1/209a844cd01825a4 \  
  --region us-east-1
```

The output includes the ARN of the listener, with the following format:

```
arn:aws:elasticloadbalancing:region:aws_account_id:listener/app/bluegreen-alb/  
e5ba62739c16e642/665750bec1b03bd4
```

Step 2: Create an Amazon ECS Cluster

Use the [create-cluster](#) command to create a cluster named `tutorial-bluegreen-cluster` to use.

```
aws ecs create-cluster \  
  --cluster-name tutorial-bluegreen-cluster \  
  --region us-east-1
```

The output includes the ARN of the cluster, with the following format:

```
arn:aws:ecs:region:aws_account_id:cluster/tutorial-bluegreen-cluster
```

Step 3: Register a Task Definition

Use the [register-task-definition](#) command to register a task definition that is compatible with Fargate. It requires the use of the `awsvpc` network mode. The following is the example task definition used for this tutorial.

First, create a file named `fargate-task.json` with the following contents. Ensure that you use the ARN for your task execution role. For more information, see [Amazon ECS Task Execution IAM Role](#) (p. 236).


```
{
  "family": "tutorial-task-def",
  "networkMode": "awsvpc",
  "containerDefinitions": [
    {
      "name": "sample-app",
      "image": "httpd:2.4",
      "portMappings": [
        {
          "containerPort": 80,
          "hostPort": 80,
          "protocol": "tcp"
        }
      ],
      "essential": true,
      "entryPoint": [
        "sh",
        "-c"
      ],
      "command": [
        "/bin/sh -c \"echo '<html> <head> <title>Amazon ECS Sample App</title>
<style>body {margin-top: 40px; background-color: #333;} </style> </head><body> <div
style=color:white;text-align:center> <h1>Amazon ECS Sample App</h1> <h2>Congratulations!
</h2> <p>Your application is now running on a container in Amazon ECS.</p> </div></body></
html>' > /usr/local/apache2/htdocs/index.html && httpd-foreground\""
      ]
    }
  ],
  "requiresCompatibilities": [
    "FARGATE"
  ],
  "cpu": "256",
  "memory": "512",
  "executionRoleArn": "arn:aws:iam::aws_account_id:role/ec2TaskExecutionRole"
}
```

Then register the task definition using the `fargate-task.json` file that you created.

```
aws ecs register-task-definition \
  --cli-input-json file://fargate-task.json \
  --region us-east-1
```

Step 4: Create an Amazon ECS Service

Use the `create-service` command to create a service.

First, create a file named `service-bluegreen.json` with the following contents.

```
{
  "cluster": "tutorial-bluegreen-cluster",
  "serviceName": "service-bluegreen",
  "taskDefinition": "tutorial-task-def",
  "loadBalancers": [
    {
      "targetGroupArn":
        "arn:aws:elasticloadbalancing:region:aws_account_id:targetgroup/
        bluegreentarget1/209a844cd01825a4",
      "containerName": "sample-app",
      "containerPort": 80
    }
  ],
}
```

```
{
  "launchType": "FARGATE",
  "schedulingStrategy": "REPLICA",
  "deploymentController": {
    "type": "CODE_DEPLOY"
  },
  "platformVersion": "LATEST",
  "networkConfiguration": {
    "awsVpcConfiguration": {
      "assignPublicIp": "ENABLED",
      "securityGroups": [ "sg-abcd1234" ],
      "subnets": [ "subnet-abcd1234", "subnet-abcd5678" ]
    }
  },
  "desiredCount": 1
}
```

Then create your service using the `service-bluegreen.json` file that you created.

```
aws ecs create-service \
  --cli-input-json file://service-bluegreen.json \
  --region us-east-1
```

The output includes the ARN of the service, with the following format:

```
arn:aws:ecs:region:aws_account_id:service/service-bluegreen
```

Step 5: Create the AWS CodeDeploy Resources

Use the following steps to create your CodeDeploy application, the Application Load Balancer target group for the CodeDeploy deployment group, and the CodeDeploy deployment group.

To create CodeDeploy resources

1. Use the `create-application` command to create an CodeDeploy application. Specify the ECS compute platform.

```
aws deploy create-application \
  --application-name tutorial-bluegreen-app \
  --compute-platform ECS \
  --region us-east-1
```

The output includes the application ID, with the following format:

```
{
  "applicationId": "b8e9c1ef-3048-424e-9174-885d7dc9dc11"
}
```

2. Use the `create-target-group` command to create a second Application Load Balancer target group, which will be used when creating your CodeDeploy deployment group.

```
aws elbv2 create-target-group \
  --name bluegreentarget2 \
  --protocol HTTP \
  --port 80 \
  --target-type ip \
  --vpc-id "vpc-0b6dd82c67d8012a1" \
  --region us-east-1
```

The output includes the ARN for the target group, with the following format:

```
arn:aws:elasticloadbalancing:region:aws_account_id:targetgroup/  
bluegreentarget2/708d384187a3cfdc
```

3. Use the `create-deployment-group` command to create an CodeDeploy deployment group.

First, create a file named `tutorial-deployment-group.json` with the following contents. This example uses the resource that you created. For the `serviceRoleArn`, specify the ARN of your Amazon ECS CodeDeploy IAM role. For more information, see [Amazon ECS CodeDeploy IAM Role](#) (p. 243).

```
{  
  "applicationName": "tutorial-bluegreen-app",  
  "autoRollbackConfiguration": {  
    "enabled": true,  
    "events": [ "DEPLOYMENT_FAILURE" ]  
  },  
  "blueGreenDeploymentConfiguration": {  
    "deploymentReadyOption": {  
      "actionOnTimeout": "CONTINUE_DEPLOYMENT",  
      "waitTimeInMinutes": 0  
    },  
    "terminateBlueInstancesOnDeploymentSuccess": {  
      "action": "TERMINATE",  
      "terminationWaitTimeInMinutes": 5  
    }  
  },  
  "deploymentGroupName": "tutorial-bluegreen-dg",  
  "deploymentStyle": {  
    "deploymentOption": "WITH_TRAFFIC_CONTROL",  
    "deploymentType": "BLUE_GREEN"  
  },  
  "loadBalancerInfo": {  
    "targetGroupPairInfoList": [  
      {  
        "targetGroups": [  
          {  
            "name": "bluegreentarget1"  
          },  
          {  
            "name": "bluegreentarget2"  
          }  
        ]  
      },  
      {  
        "prodTrafficRoute": {  
          "listenerArns": [  
            "arn:aws:elasticloadbalancing:region:aws_account_id:listener/  
app/bluegreen-alb/e5ba62739c16e642/665750bec1b03bd4"  
          ]  
        }  
      }  
    ]  
  },  
  "serviceRoleArn": "arn:aws:iam::aws_account_id:role/ecsCodeDeployRole",  
  "ecsServices": [  
    {  
      "serviceName": "service-bluegreen",  
      "clusterName": "tutorial-bluegreen-cluster"  
    }  
  ]  
}
```

Then create the CodeDeploy deployment group.

```
aws deploy create-deployment-group \
  --cli-input-json file://tutorial-deployment-group.json \
  --region us-east-1
```

The output includes the deployment group ID, with the following format:

```
{
  "deploymentGroupId": "6fd9bdc6-dc51-4af5-ba5a-0a4a72431c88"
}
```

Step 6: Create and Monitor an CodeDeploy Deployment

Use the following steps to create and upload an application specification file (AppSpec file) and an CodeDeploy deployment.

To create and monitor an CodeDeploy deployment

1. Create and upload an AppSpec file using the following steps.
 - a. Create a file named `appspec.yaml` with the contents of the CodeDeploy deployment group. This example uses the resources that you created earlier in the tutorial.

```
version: 0.0
Resources:
  - TargetService:
      Type: AWS::ECS::Service
      Properties:
        TaskDefinition: "arn:aws:ecs:region:aws_account_id:task-definition/first-run-task-definition:7"
        LoadBalancerInfo:
          ContainerName: "sample-app"
          ContainerPort: 80
          PlatformVersion: "LATEST"
```

- b. Use the `s3 mb` command to create an Amazon S3 bucket for the AppSpec file.

```
aws s3 mb s3://tutorial-bluegreen-bucket
```

- c. Use the `s3 cp` command to upload the AppSpec file to the Amazon S3 bucket.

```
aws s3 cp ./appspec.yaml s3://tutorial-bluegreen-bucket/appspec.yaml
```

2. Create the CodeDeploy deployment using the following steps.
 - a. Create a file named `create-deployment.json` with the contents of the CodeDeploy deployment. This example uses the resources that you created earlier in the tutorial.

```
{
  "applicationName": "tutorial-bluegreen-app",
  "deploymentGroupName": "tutorial-bluegreen-dg",
  "revision": {
    "revisionType": "S3",
```

```
    "s3Location": {
      "bucket": "tutorial-bluegreen-bucket",
      "key": "appspec.yaml",
      "bundleType": "YAML"
    }
  }
}
```

- b. Use the `create-deployment` command to create the deployment.

```
aws deploy create-deployment \
  --cli-input-json file://create-deployment.json \
  --region us-east-1
```

The output includes the deployment ID, with the following format:

```
{
  "deploymentId": "d-RPCR1U3TW"
}
```

- c. Use the `get-deployment-target` command to get the details of the deployment, specifying the `deploymentId` from the previous output.

```
aws deploy get-deployment-target \
  --deployment-id "d-IMJU3A8TW" \
  --target-id tutorial-bluegreen-cluster:service-bluegreen \
  --region us-east-1
```

Continue to retrieve the deployment details until the status is `Succeeded`, as shown in the following output.

```
{
  "deploymentTarget": {
    "deploymentTargetType": "ECSTarget",
    "ecsTarget": {
      "deploymentId": "d-RPCR1U3TW",
      "targetId": "tutorial-bluegreen-cluster:service-bluegreen",
      "targetArn": "arn:aws:ecs:region:aws_account_id:service/service-bluegreen",
      "lastUpdatedAt": 1543431490.226,
      "lifecycleEvents": [
        {
          "lifecycleEventName": "BeforeInstall",
          "startTime": 1543431361.022,
          "endTime": 1543431361.433,
          "status": "Succeeded"
        },
        {
          "lifecycleEventName": "Install",
          "startTime": 1543431361.678,
          "endTime": 1543431485.275,
          "status": "Succeeded"
        },
        {
          "lifecycleEventName": "AfterInstall",
          "startTime": 1543431485.52,
          "endTime": 1543431486.033,
          "status": "Succeeded"
        },
        {
          "lifecycleEventName": "BeforeAllowTraffic",

```

```
        "startTime": 1543431486.838,
        "endTime": 1543431487.483,
        "status": "Succeeded"
    },
    {
        "lifecycleEventName": "AllowTraffic",
        "startTime": 1543431487.748,
        "endTime": 1543431488.488,
        "status": "Succeeded"
    },
    {
        "lifecycleEventName": "AfterAllowTraffic",
        "startTime": 1543431489.152,
        "endTime": 1543431489.885,
        "status": "Succeeded"
    }
],
"status": "Succeeded",
"taskSetsInfo": [
    {
        "identifer": "ecs-svc/9223370493425779968",
        "desiredCount": 1,
        "pendingCount": 0,
        "runningCount": 1,
        "status": "ACTIVE",
        "trafficWeight": 0.0,
        "targetGroup": {
            "name": "bluegreentarget1"
        }
    },
    {
        "identifer": "ecs-svc/9223370493423413672",
        "desiredCount": 1,
        "pendingCount": 0,
        "runningCount": 1,
        "status": "PRIMARY",
        "trafficWeight": 100.0,
        "targetGroup": {
            "name": "bluegreentarget2"
        }
    }
]
}
}
```

Step 7: Clean Up

When you have finished this tutorial, clean up the resources associated with it to avoid incurring charges for resources that you aren't using.

Cleaning up the tutorial resources

1. Use the `delete-deployment-group` command to delete the CodeDeploy deployment group.

```
aws deploy delete-deployment-group \
  --application-name tutorial-bluegreen-app \
  --deployment-group-name tutorial-bluegreen-dg \
  --region us-east-1
```

2. Use the `delete-application` command to delete the CodeDeploy application.

```
aws deploy delete-application \  
  --application-name tutorial-bluegreen-app \  
  --region us-east-1
```

3. Use the `delete-service` command to delete the Amazon ECS service. Using the `--force` flag allows you to delete a service even if it has not been scaled down to zero tasks.

```
aws ecs delete-service \  
  --service arn:aws:ecs:region:aws_account_id:service/service-bluegreen \  
  --force \  
  --region us-east-1
```

4. Use the `delete-cluster` command to delete the Amazon ECS cluster.

```
aws ecs delete-cluster \  
  --cluster tutorial-bluegreen-cluster \  
  --region us-east-1
```

5. Use the `s3 rm` command to delete the AppSpec file from the Amazon S3 bucket.

```
aws s3 rm s3://tutorial-bluegreen-bucket/appspec.yaml
```

6. Use the `s3 rb` command to delete the Amazon S3 bucket.

```
aws s3 rb s3://tutorial-bluegreen-bucket
```

7. Use the `delete-load-balancer` command to delete the Application Load Balancer.

```
aws elbv2 delete-load-balancer \  
  --load-balancer-arn  
  arn:aws:elasticloadbalancing:region:aws_account_id:loadbalancer/app/bluegreen-alb/  
e5ba62739c16e642 \  
  --region us-east-1
```

8. Use the `delete-target-group` command to delete the two Application Load Balancer target groups.

```
aws elbv2 delete-target-group \  
  --target-group-arn  
  arn:aws:elasticloadbalancing:region:aws_account_id:targetgroup/  
bluegreentarget1/209a844cd01825a4 \  
  --region us-east-1
```

```
aws elbv2 delete-target-group \  
  --target-group-arn  
  arn:aws:elasticloadbalancing:region:aws_account_id:targetgroup/  
bluegreentarget2/708d384187a3cfdc \  
  --region us-east-1
```

Tutorial: Listening for Amazon ECS CloudWatch Events

In this tutorial, you set up a simple AWS Lambda function that listens for Amazon ECS task events and writes them out to a CloudWatch Logs log stream.

Prerequisite: Set Up a Test Cluster

If you do not have a running cluster to capture events from, follow the steps in [Creating a Cluster \(p. 18\)](#) to create one. At the end of this tutorial, you run a task on this cluster to test that you have configured your Lambda function correctly.

Step 1: Create the Lambda Function

In this procedure, you create a simple Lambda function to serve as a target for Amazon ECS event stream messages.

1. Open the AWS Lambda console at <https://console.aws.amazon.com/lambda/>.
2. Choose **Create function**.
3. On the **Author from scratch** screen, do the following:
 - a. For **Name**, enter a value.
 - b. For **Runtime**, choose **Python 2.7**.
 - c. For **Role**, choose **Create a new role with basic Lambda permissions**.
4. Choose **Create function**.
5. In the **Function code** section, edit the sample code to match the following example:

```
import json

def lambda_handler(event, context):
    if event["source"] != "aws.ecs":
        raise ValueError("Function only supports input from events with a source type of: aws.ecs")

    print('Here is the event:')
    print(json.dumps(event))
```

This is a simple Python 2.7 function that prints the event sent by Amazon ECS. If everything is configured correctly, at the end of this tutorial, you see that the event details appear in the CloudWatch Logs log stream associated with this Lambda function.

6. Choose **Save**.

Step 2: Register Event Rule

Next, you create a CloudWatch Events event rule that captures task events coming from your Amazon ECS clusters. This rule captures all events coming from all clusters within the account where it is defined. The task messages themselves contain information about the event source, including the cluster on which it resides, that you can use to filter and sort events programmatically.

Note

When you use the AWS Management Console to create an event rule, the console automatically adds the IAM permissions necessary to grant CloudWatch Events permission to call your Lambda function. If you are creating an event rule using the AWS CLI, you need to grant this permission explicitly. For more information, see [Events and Event Patterns](#) in the *Amazon CloudWatch Events User Guide*.

To route events to your Lambda function

1. Open the CloudWatch console at <https://console.aws.amazon.com/cloudwatch/>.
2. On the navigation pane, choose **Events, Rules, Create rule**.

3. For **Event Source**, choose **ECS** as the event source. By default, the rule applies to all Amazon ECS events for all of your Amazon ECS groups. Alternatively, you can select specific events or a specific Amazon ECS group.
4. For **Targets**, choose **Add target**, for **Target type**, choose **Lambda function**, and then select your Lambda function.
5. Choose **Configure details**.
6. For **Rule definition**, type a name and description for your rule and choose **Create rule**.

Step 3: Test Your Rule

Finally, you create a CloudWatch Events event rule that captures task events coming from your Amazon ECS clusters. This rule captures all events coming from all clusters within the account where it is defined. The task messages themselves contain information about the event source, including the cluster on which it resides, that you can use to filter and sort events programmatically.

To test your rule

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. Choose **Clusters**, **default**.
3. On the **Cluster : default** screen, choose **Tasks**, **Run new Task**.
4. For **Task Definition**, select the latest version of **console-sample-app-static** and choose **Run Task**.
5. Open the CloudWatch console at <https://console.aws.amazon.com/cloudwatch/>.
6. On the navigation pane, choose **Logs** and select the log group for your Lambda function (for example, **/aws/lambda/my-function**).
7. Select a log stream to view the event data.

Tutorial: Sending Amazon Simple Notification Service Alerts for Task Stopped Events

In this tutorial, you configure a CloudWatch Events event rule that only captures task events where the task has stopped running because one of its essential containers has terminated. The event sends only task events with a specific `stoppedReason` property to the designated Amazon SNS topic.

Prerequisite: Set Up a Test Cluster

If you do not have a running cluster to capture events from, follow the steps in [Creating a Cluster \(p. 18\)](#) to create one. At the end of this tutorial, you run a task on this cluster to test that you have configured your Amazon SNS topic and CloudWatch Events event rule correctly.

Step 1: Create and Subscribe to an Amazon SNS Topic

For this tutorial, you configure an Amazon SNS topic to serve as an event target for your new event rule.

To create an Amazon SNS topic

1. Open the Amazon SNS console at <https://console.aws.amazon.com/sns/v3/home>.
2. Choose **Topics**, **Create topic**.
3. On the **Create topic** screen, for **Name**, enter **TaskStoppedAlert** and choose **Create topic**.
4. On the **TaskStoppedAlert** details screen, choose **Create subscription**.

5. On the **Create subscription** screen, for **Protocol**, choose **Email**. For **Endpoint**, enter an email address to which you currently have access and choose **Create subscription**.
6. Check your email account, and wait to receive a subscription confirmation email message. When you receive it, choose **Confirm subscription**.

Step 2: Register Event Rule

Next, you register an event rule that captures only task-stopped events for tasks with stopped containers.

To create an event rule

1. Open the CloudWatch console at <https://console.aws.amazon.com/cloudwatch/>.
2. On the navigation pane, choose **Events, Rules, Create rule**.
3. For Event Source, choose **Event Pattern**, select **Custom event pattern** and then replace the existing text with the following text:

```
{
  "source": [
    "aws.ecs"
  ],
  "detail-type": [
    "ECS Task State Change"
  ],
  "detail": {
    "lastStatus": [
      "STOPPED"
    ],
    "stoppedReason": [
      "Essential container in task exited"
    ]
  }
}
```

This code defines a CloudWatch Events event rule that matches any event where the `lastStatus` and `stoppedReason` fields match the indicated values. For more information about event patterns, see [Events and Event Patterns](#) in the *Amazon CloudWatch User Guide*.

4. For **Targets**, choose **Add target**. For **Target type**, choose **SNS topic**, and then choose **TaskStoppedAlert**.
5. Choose **Configure details**.
6. For **Rule definition**, type a name and description for your rule and then choose **Create rule**.

Step 3: Test Your Rule

Verify that the rule is working by running a task that exits shortly after it starts. If your event rule is configured correctly, you receive an email message within a few minutes with the event text. If you have an existing task definition that can satisfy the rule requirements, run a task using it. If you do not, the following steps will walk you through registering a Fargate task definition and running it that will.

To test the rule

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. Choose **Task Definitions, Create new Task Definition**.
3. For Select launch type compatibility, choose **FARGATE, Next step**.

4. Choose **Configure via JSON**, copy and paste the following task definition JSON into the field and choose **Save**.

```
{
  "containerDefinitions": [
    {
      "command": [
        "sh",
        "-c",
        "sleep 5"
      ],
      "essential": true,
      "image": "amazonlinux:2",
      "name": "test-sleep"
    }
  ],
  "cpu": "256",
  "executionRoleArn": "arn:aws:iam::012345678910:role/ecsTaskExecutionRole",
  "family": "fargate-task-definition",
  "memory": "512",
  "networkMode": "awsvpc",
  "requiresCompatibilities": [
    "FARGATE"
  ]
}
```

5. Choose **Create, View task definition**.
6. For **Actions**, choose **Run Task**.
7. For Launch type, choose **FARGATE**. For **VPC and security groups**, choose a VPC and Subnets for the task to use and then choose **Run Task**.
8. For **Container name**, type **Wordpress**, for **Image**, type **wordpress**, and for **Maximum memory (MB)**, type **128**.
9. On the **Tasks** tab for your cluster, periodically choose the refresh icon until you no longer see your task running. To verify that your task has stopped, for **Desired task status**, choose **Stopped**.
10. Check your email to confirm that you have received an email alert for the stopped notification.

Amazon ECS troubleshooting

You may need to troubleshoot issues with your load balancers, tasks, services, or container instances. This chapter helps you find diagnostic information from the Amazon ECS container agent, the Docker daemon on the container instance, and the service event log in the Amazon ECS console.

Topics

- [Checking stopped tasks for errors \(p. 340\)](#)
- [Stopped tasks error codes \(p. 344\)](#)
- [CannotPullContainer task errors \(p. 347\)](#)
- [Service Event Messages \(p. 348\)](#)
- [Invalid CPU or memory value specified \(p. 349\)](#)
- [Troubleshooting service load balancers \(p. 350\)](#)

Checking stopped tasks for errors

If you have trouble starting a task, your task might be stopping because of an error. For example, you run the task and the task displays a `PENDING` status and then disappears. You can view errors like this in the Amazon ECS console by displaying the stopped task and inspecting it for error messages.

To check stopped tasks for errors

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. On the **Clusters** page, select the cluster in which your stopped task resides.
3. On the **Cluster : *clustername*** page, choose **Tasks**.
4. In the **Desired task status** table header, choose **Stopped**, and then select the stopped task to inspect. The most recent stopped tasks are listed first.
5. In the **Details** section, inspect the **Stopped reason** field to see the reason that the task was stopped.

Details

Cluster	default
Container Instance	dd3599e9-2ca6-40f4-9da5-a0bb10408260
EC2 instance id	i-83c6ab47
Task Definition	curler:4
Last status	STOPPED
Desired status	STOPPED
Created at	2015-11-20 13:31:01 -0800
Stopped at	2015-11-20 13:31:03 -0800
Stopped reason	Essential container in task exited

6. If you have a container that has stopped, expand the container and inspect the **Status reason** row to see what caused the task state to change.

Containers

Name	Container Id	Status
▼ curler	3f871451-c9f1-4d6f-a...	STOPPED (CannotPullContainerError: Error: image tutum/bogus)
Details		
Status reason CannotPullContainerError: Error: image tutum/bogus:latest not found		
Command ["/usr/bin/watch","curl","-v","http://amazon-ecs-2004772631.us-west-2.elb.amazonaws.com/"]		

In the previous example, the container image name cannot be found. This can happen if you misspell the image name.

If this inspection does not provide enough information, see

The following are the possible error messages you may receive when your Fargate task is stopped unexpectedly. The error messages are returned by the container agent and the prefix is dependent on the platform version the task is using.

To check your stopped tasks for an error message using the AWS Management Console, see [Checking stopped tasks for errors](#) (p. 340).

DockerTimeoutError	ContainerRuntimeTimeoutError	This error occurs when a container was unable to transition to either a RUNNING or STOPPED state within the timeout period. The reason and timeout value is provided in the error message.	ContainerRuntimeTimeoutError: Could not transition to running; timed out after waiting 1m: <reason>
CannotStartContainerError	CannotStartContainerError	This error occurs when a container is unable to be started.	CannotStartContainerError: failed to get container status: <reason>
CannotStopContainerError	CannotStopContainerError	This error occurs when a container is unable to be stopped.	CannotStopContainerError: failed sending SIGTERM to container: <reason>
CannotInspectContainerError	CannotInspectContainerError	This error occurs when the container agent is unable to describe the container via the container runtime. When using platform version 1.3 or prior, the ECS agent will return the reason from Docker. When using platform version 1.4 or later, the Fargate agent will return the reason from containerd.	CannotInspectContainerError: <reason>

	ResourceInitializationError	This error occurs when the Fargate agent fails to create or bootstrap the resources required to start the container or the task is belongs to.	ResourceInitializationError: failed to initialize logging driver: <i><reason></i>
		This error only occurs if using platform version 1.4 or later.	
CannotPullContainerError	CannotPullContainerError	This error occurs when the agent is unable to pull the container image specified in the task definition. For more information, see CannotPullContainer task errors (p. 347).	CannotPullContainerError: <i><reason></i>
	CannotCreateVolumeError	This error occurs when the agent is unable to create the volume mount specified in the task definition. This error only occurs if using platform version 1.4 or later.	CannotCreateVolumeError: <i><reason></i>

	ContainerRuntimeError	This error occurs when the agent	ContainerRuntimeError: failed
		receives an unexpected error	to create
		from containerd	container IO set: <i><reason></i>
		for a runtime-specific operation.	
		This error is usually caused by an internal failure in the agent or the containerd runtime.	
		This error only occurs if using platform version 1.4 or later.	
OutOfMemoryError	OutOfMemoryError	This error occurs when a container exits due to processes in the container consuming more memory than was allocated in the task definition.	OutOfMemoryError: container killed due to memory usage
	InternalError	This error occurs when the agent encounters an unexpected, non-runtime related internal error.	InternalError: <i><reason></i>
		This error only occurs if using platform version 1.4 or later.	
Error message prefix in platform version 1.3 and prior	Error message prefix in platform version 1.4 and later	Details	Example

(p. 344) for more information.

Stopped tasks error codes

The following are the possible error messages you may receive when your Fargate task is stopped unexpectedly. The error messages are returned by the container agent and the prefix is dependent on the platform version the task is using.

To check your stopped tasks for an error message using the AWS Management Console, see [Checking stopped tasks for errors](#) (p. 340).

Error message prefix in platform version 1.3 and prior	Error message prefix in platform version 1.4 and later	Details	Example
DockerTimeoutError	ContainerRuntimeTimeoutError	This error occurs when a container was unable to transition to either a RUNNING or STOPPED state within the timeout period. The reason and timeout value is provided in the error message.	ContainerRuntimeTimeoutError: Could not transition to running; timed out after waiting 1m: <reason>
CannotStartContainerError	CannotStartContainerError	This error occurs when a container is unable to be started.	CannotStartContainerError: failed to get container status: <reason>
CannotStopContainerError	CannotStopContainerError	This error occurs when a container is unable to be stopped.	CannotStopContainerError: failed sending SIGTERM to container: <reason>
CannotInspectContainerError	CannotInspectContainerError	This error occurs when the container agent is unable to describe the container via the container runtime. When using platform version 1.3 or prior, the ECS agent will return the reason from Docker. When using platform version 1.4 or later, the Fargate agent will return the reason from containerd.	CannotInspectContainerError: <reason>
	ResourceInitializationError	This error occurs when the Fargate agent fails to create or bootstrap the resources required to start the container or the task is belongs to. This error only occurs if using platform version 1.4 or later.	ResourceInitializationError: failed to initialize logging driver: <reason>
CannotPullContainerError	CannotPullContainerError	This error occurs when the agent is unable	CannotPullContainerError: <reason>

Error message prefix in platform version 1.3 and prior	Error message prefix in platform version 1.4 and later	Details	Example
		to pull the container image specified in the task definition. For more information, see CannotPullContainer task errors (p. 347).	
	CannotCreateVolumeError	<p>This error occurs when the agent is unable to create the volume mount specified in the task definition.</p> <p>This error only occurs if using platform version 1.4 or later.</p>	CannotCreateVolumeError: <i><reason></i>
	ContainerRuntimeError	<p>This error occurs when the agent receives an unexpected error from containerd for a runtime-specific operation. This error is usually caused by an internal failure in the agent or the containerd runtime.</p> <p>This error only occurs if using platform version 1.4 or later.</p>	ContainerRuntimeError: failed to create container IO set: <i><reason></i>
OutOfMemoryError	OutOfMemoryError	This error occurs when a container exits due to processes in the container consuming more memory than was allocated in the task definition.	OutOfMemoryError: container killed due to memory usage
	InternalError	<p>This error occurs when the agent encounters an unexpected, non-runtime related internal error.</p> <p>This error only occurs if using platform version 1.4 or later.</p>	InternalError: <i><reason></i>

CannotPullContainer task errors

The following Docker errors indicate that when creating a task, the container image specified could not be retrieved.

Connection timed out

When a Fargate task is launched, its elastic network interface requires a route to the internet to pull container images. If you receive an error similar to the following when launching a task, it is because a route to the internet does not exist:

```
CannotPullContainerError: API error (500): Get https://111122223333.dkr.ecr.us-east-1.amazonaws.com/v2/: net/http: request canceled while waiting for connection"
```

To resolve this issue, you can:

Context canceled

The common cause for this error is because the VPC your task is using does not have a route to pull the container image from Amazon ECR.

Image not found

When you specify an Amazon ECR image in your container definition, you must use the full ARN or URI of your ECR repository along with the image name in that repository. If the repository or image cannot be found, you receive the following error:

```
CannotPullContainerError: API error (404): repository 111122223333.dkr.ecr.us-east-1.amazonaws.com/<repo>/<image> not found
```

To resolve this issue, verify the repository ARN or URI and the image name. Also ensure that you have set up the proper access using the task execution IAM role. For more information about the task execution role, see [Amazon ECS Task Execution IAM Role \(p. 236\)](#).

Insufficient disk space

If the root volume of your container instance has insufficient disk space when pulling the container image, you see an error similar to the following:

```
CannotPullContainerError: write /var/lib/docker/tmp/GetImageBlob11111111: no space left on device
```

To resolve this issue, free up disk space.

If you are using the Amazon ECS-optimized AMI, you can use the following command to retrieve the 20 largest files on your filesystem:

```
du -Sh / | sort -rh | head -20
```

Example output:

```
5.7G    /var/lib/docker/containers/50501b5f4cbf90b406e0ca60bf4e6d4ec8f773a6c1d2b451ed8e0195418ad0d2
1.2G    /var/log/ecs
594M    /var/lib/docker/devicemapper/mnt/c8e3010e36ce4c089bf286a623699f5233097ca126ebd5a700af023a5127633d/rootfs/data/logs
```

...

In some cases, like this example above, the root volume may be filled out by a running container. If the container is using the default `json-file` log driver without a `max-size` limit, it may be that the log file is responsible for most of that space used. You can use the `docker ps` command to verify which container is using the space by mapping the directory name from the output above to the container ID. For example:

CONTAINER ID	IMAGE	COMMAND	CREATED
STATUS	PORTS	NAMES	
50501b5f4cbf	amazon/amazon-ecs-agent:latest	"/agent"	4 days ago
Up 4 days		ecs-agent	

By default, when using the `json-file` log driver, Docker captures the standard output (and standard error) of all of your containers and writes them in files using the JSON format. You are able to set the `max-size` as a log driver option, which prevents the log file from taking up too much space. For more information, see [Configure logging drivers](#) in the Docker documentation.

The following is a container definition snippet showing how to use this option:

```
{
  "log-driver": "json-file",
  "log-opts": {
    "max-size": "256m"
  }
}
```

An alternative if your container logs are taking up too much disk space is to use the `awslogs` log driver. The `awslogs` log driver sends the logs to CloudWatch, which frees up the disk space that would otherwise be used for your container logs on the container instance. For more information, see [Using the awslogs Log Driver \(p. 69\)](#).

Service Event Messages

If you are troubleshooting a problem with a service, the first place you should check for diagnostic information is the service event log.

When viewing service event messages in the Amazon ECS console, duplicate service event messages are omitted until either the cause is resolved or six hours passes. If the cause is not resolved, you will receive another service event message after six hours has passed.

To check the service event log in the Amazon ECS console

1. Open the Amazon ECS console at <https://console.aws.amazon.com/ecs/>.
2. On the **Clusters** page, select the cluster in which your service resides.
3. On the **Cluster : *clustername*** page, select the service to inspect.
4. On the **Service : *servicename*** page, choose **Events**.

Tasks Events		
Last updated on April 24, 2015 6:33:30 AM (0m ago)		
Filter in this page		
Viewing 1-10 Events		
Event Id	Event Time	Message
22153606-5c...	2015-04-24 06:32:20 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
d863e60c-d3...	2015-04-24 06:30:47 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
dc59a716-b1...	2015-04-24 06:29:14 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
27c37c68-57...	2015-04-24 06:27:41 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
16a36873-8e...	2015-04-24 06:26:08 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
8cee3c0f-693...	2015-04-24 06:24:35 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
2137e914-18...	2015-04-24 06:23:02 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
4142d52d-62...	2015-04-24 06:21:29 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
d4f45e33-766...	2015-04-24 06:19:56 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.
9ad2546b-12...	2015-04-24 06:18:22 -0700	(service sample-webapp) was unable to place a task because the resources could not be found.

- Examine the **Message** column for errors or other helpful information.

Service Event Messages

The following are examples of service event messages you may see in the console:

- service (*service-name*) is unable to consistently start tasks successfully. (p. 349)

service (*service-name*) is unable to consistently start tasks successfully.

This service contains tasks that have failed to start after consecutive attempts. At this point, the service scheduler begins to incrementally increase the time between retries. You should troubleshoot why your tasks are failing to launch. For more information, see [Service Throttle Logic](#) (p. 176).

After the service is updated, for example with an updated task definition, the service scheduler resumes normal behavior.

Invalid CPU or memory value specified

When registering a task, if you specify an invalid cpu or memory value, you receive the following error:

```
An error occurred (ClientException) when calling the RegisterTaskDefinition operation:
Invalid 'cpu' setting for task. For more information, see the Troubleshooting section of
the Amazon ECS Developer Guide.
```

To resolve this issue, you must specify a supported value for the task CPU and memory in your task definition.

The cpu value can be expressed in CPU units or vCPUs in a task definition but is converted to an integer indicating the CPU units when the task definition is registered. If you are using the EC2 launch type, the

supported values are between 128 CPU units (0.125 vCPUs) and 10240 CPU units (10 vCPUs). If you are using the Fargate launch type, you must use one of the values in the following table, which determines your range of supported values for the `memory` parameter.

The `memory` value can be expressed in MiB or GB in a task definition but is converted to an integer indicating the MiB when the task definition is registered. If you are using the EC2 launch type, you must specify an integer. If you are using the Fargate launch type, you must use one of the values in the following table, which determines your range of supported values for the `cpu` parameter.

Supported task CPU and memory values for Fargate tasks are as follows.

CPU value	Memory value (MiB)
256 (.25 vCPU)	512 (0.5GB), 1024 (1GB), 2048 (2GB)
512 (.5 vCPU)	1024 (1GB), 2048 (2GB), 3072 (3GB), 4096 (4GB)
1024 (1 vCPU)	2048 (2GB), 3072 (3GB), 4096 (4GB), 5120 (5GB), 6144 (6GB), 7168 (7GB), 8192 (8GB)
2048 (2 vCPU)	Between 4096 (4GB) and 16384 (16GB) in increments of 1024 (1GB)
4096 (4 vCPU)	Between 8192 (8GB) and 30720 (30GB) in increments of 1024 (1GB)

Troubleshooting service load balancers

Amazon ECS services can register tasks with an Elastic Load Balancing load balancer. Load balancer configuration errors are common causes for stopped tasks. If your stopped tasks were started by services that use a load balancer, consider the following possible causes.

Important

Container health checks are not supported for tasks that are part of a service that is configured to use a Classic Load Balancer. The Amazon ECS service scheduler ignores tasks in an `UNHEALTHY` state that are behind a Classic Load Balancer.

Container instance security group

If your container is mapped to port 80 on your container instance, your container instance security group must allow inbound traffic on port 80 for the load balancer health checks to pass.

Elastic Load Balancing load balancer not configured for all Availability Zones

Your load balancer should be configured to use all of the Availability Zones in a region, or at least all of the Availability Zones in which your container instances reside. If a service uses a load balancer and starts a task on a container instance that resides in an Availability Zone that the load balancer is not configured to use, the task never passes the health check and it is killed.

Elastic Load Balancing load balancer health check misconfigured

The load balancer health check parameters can be overly restrictive or point to resources that do not exist. If a container instance is determined to be unhealthy, it is removed from the load balancer. Be sure to verify that the following parameters are configured correctly for your service load balancer.

Ping Port

The **Ping Port** value for a load balancer health check is the port on the container instances that the load balancer checks to determine if it is healthy. If this port is misconfigured, the load

balancer likely deregisters your container instance from itself. This port should be configured to use the `hostPort` value for the container in your service's task definition that you are using with the health check.

Ping Path

This value is often set to `index.html`, but if your service does not respond to that request, then the health check fails. If your container does not have an `index.html` file, you can set this to `/` to target the base URL for the container instance.

Response Timeout

This is the amount of time that your container has to return a response to the health check ping. If this value is lower than the amount of time required for a response, the health check fails.

Health Check Interval

This is the amount of time between health check pings. The shorter your health check intervals are, the faster your container instance can reach the **Unhealthy Threshold**.

Unhealthy Threshold

This is the number of times your health check can fail before your container instance is considered unhealthy. If you have an unhealthy threshold of 2, and a health check interval of 30 seconds, then your task has 60 seconds to respond to the health check ping before it is assumed unhealthy. You can raise the unhealthy threshold or the health check interval to give your tasks more time to respond.

Unable to update the service `servicename`: Load balancer container name or port changed in task definition

If your service uses a load balancer, the load balancer configuration defined for your service when it was created cannot be changed. If you update the task definition for the service, the container name and container port that were specified when the service was created must remain in the task definition.

To change the load balancer name, the container name, or the container port associated with a service load balancer configuration, you must create a new service.

Document History

The following table describes the major updates and new features for the *Amazon ECS User Guide for AWS Fargate*. We also update the documentation frequently to address the feedback that you send us.

Change	Description	Date
AWS Fargate platform version 1.4.0 update	Beginning on May 28, 2020, any new Fargate task that is launched using platform version 1.4.0 will have its 20 GB ephemeral storage encrypted with an AES-256 encryption algorithm using an AWS Fargate-managed encryption key. For more information, see Using Data Volumes in Tasks (p. 63) .	28 May 2020
AWS Fargate Region expansion	AWS Fargate with Amazon ECS has expanded to the Africa (Cape Town) Region.	11 May 2020
Service quota updated	The following service quota was updated: <ul style="list-style-type: none"> Clusters per account was raised from 2,000 to 10,000. For more information, see Amazon ECS Service Quotas (p. 283) .	17 April 2020
AWS Fargate platform version 1.4.0	AWS Fargate platform version 1.4.0 is released, which contains the following features: <ul style="list-style-type: none"> Added support for using Amazon EFS file system volumes for persistent task storage. For more information, see Amazon EFS Volumes (p. 65). The ephemeral task storage has been increased to 20 GB. For more information, see Using Data Volumes in Tasks (p. 63). The network traffic behavior to and from tasks has been updated. Starting with platform version 1.4.0, all Fargate tasks receive a single elastic network interface, referred to as the task ENI. All network traffic flows through that ENI within your VPC and will be visible to you through your VPC flow logs. For more information, see Fargate Task Networking (p. 67). Task ENIs add support for jumbo frames. Network interfaces are configured with a maximum transmission unit (MTU), which is the size of the largest payload that fits within a single frame. The larger the MTU, the larger the application payload that can fit within a single frame, which reduces per-frame overhead and increases efficiency. Supporting jumbo frames reduces overhead when the network path between your task and the destination supports jumbo frames, such as all traffic that remains within your VPC. CloudWatch Container Insights will include network performance metrics for Fargate tasks. For more 	8 April 2020

Change	Description	Date
	<p>information, see Amazon ECS CloudWatch Container Insights (p. 197).</p> <ul style="list-style-type: none"> Added support for the task metadata endpoint v4, which provides additional information for your Fargate tasks, including network stats for the task and which Availability Zone the task is running in. For more information, see Task metadata endpoint version 4 (p. 271). Added support for the <code>SYS_PTRACE</code> Linux parameter in container definitions. For more information, see Linux Parameters (p. 50). The Fargate container agent replaces the use of the Amazon ECS container agent for all Fargate tasks. This change should not affect how your tasks run. The container runtime is now using Containerd instead of Docker. This change should not affect how your tasks run. You may notice that some error messages that originate with the container runtime are more general and do not mention Docker. <p>For more information, see AWS Fargate platform versions (p. 14).</p>	
Amazon EFS file system support for task volumes	Amazon EFS file systems can be used as data volumes for your Fargate tasks. For more information, see Amazon EFS Volumes (p. 65).	8 April 2020
Amazon ECS Task Metadata Endpoint version 4	Beginning with Fargate platform version 1.4.0, an environment variable named <code>ECS_CONTAINER_METADATA_URI_V4</code> is injected into each container in a task. When you query the task metadata version 4 endpoint, various task metadata and Docker stats are available to tasks. For more information, see Task metadata endpoint version 4 (p. 271).	8 April 2020
Fargate Spot	Amazon ECS added support for running tasks using Fargate Spot. For more information, see Using AWS Fargate Capacity Providers (p. 20).	3 Dec 2019
Service Action Events	Amazon ECS now sends events to Amazon EventBridge when certain service actions occur. For more information, see Service Action Events (p. 193).	25 Nov 2019
Savings Plans	Savings Plans are a pricing model that offer significant savings on AWS usage. For more information, see Savings Plans and AWS Fargate (p. 285).	6 Nov 2019
FireLens for Amazon ECS	FireLens for Amazon ECS is in general availability. FireLens for Amazon ECS enables you to use task definition parameters to route logs to an AWS service or partner destination for log storage and analytics. For more information, see Custom Log Routing (p. 75).	30 Sept 2019

Change	Description	Date
AWS Fargate Region expansion	AWS Fargate with Amazon ECS has expanded to the Europe (Paris), Europe (Stockholm), and Middle East (Bahrain) Regions.	30 Sept 2019
FireLens for Amazon ECS	FireLens for Amazon ECS is in public preview. FireLens for Amazon ECS enables you to use task definition parameters to route logs to an AWS service or partner destination for log storage and analytics. For more information, see Custom Log Routing (p. 75) .	30 Aug 2019
CloudWatch Container Insights	CloudWatch Container Insights is now generally available. It enables you to collect, aggregate, and summarize metrics and logs from your containerized applications and microservices. For more information, see Amazon ECS CloudWatch Container Insights (p. 197) .	30 Aug 2019
AWS Fargate Region expansion	AWS Fargate with Amazon ECS has expanded to the Asia Pacific (Hong Kong) Region.	06 Aug 2019
Registering Multiple Target Groups with a Service	Added support for specifying multiple target groups in a service definition. For more information, see Registering Multiple Target Groups with a Service (p. 163) .	30 July 2019
CloudWatch Container Insights	Amazon ECS has added support for CloudWatch Container Insights. For more information, see Amazon ECS CloudWatch Container Insights (p. 197) .	9 July 2019
Resource-level permissions for Amazon ECS services and tasksets	Amazon ECS has expanded resource-level permissions support for Amazon ECS services and tasks. For more information, see How Amazon Elastic Container Service Works with IAM (p. 206) .	27 June 2019
AWS Fargate platform version 1.3.0 update	Beginning on May 1, 2019, any new Fargate task that is launched supports the <code>splunk</code> log driver in addition to the <code>awslogs</code> log driver. For more information, see Storage and Logging (p. 45) .	1 May 2019
AWS Fargate platform version 1.3.0 update	Beginning on May 1, 2019, any new Fargate task that is launched supports referencing sensitive data in the log configuration of a container using the <code>secretOptions</code> container definition parameter. For more information, see Specifying Sensitive Data (p. 87) .	1 May 2019
AWS Fargate platform version 1.3.0 update	Beginning on April 2, 2019, any new Fargate task that is launched supports injecting sensitive data into your containers by storing your sensitive data in either AWS Secrets Manager secrets or AWS Systems Manager Parameter Store parameters and then referencing them in your container definition. For more information, see Specifying Sensitive Data (p. 87) .	2 Apr 2019

Change	Description	Date
AWS Fargate platform version 1.3.0 update	Beginning on March 27, 2019, any new Fargate task can use additional task definition parameters that enable you to define a proxy configuration, dependencies for container startup and shutdown as well as a per-container start and stop timeout value. For more information, see Proxy configuration (p. 58) , Container Dependency (p. 51) , and Container Timeouts (p. 52) .	27 Mar 2019
Amazon ECS introduces the external deployment type	The <i>external</i> deployment type enables you to use any third-party deployment controller for full control over the deployment process for an Amazon ECS service. For more information, see External Deployment (p. 147) .	27 Mar 2019
Amazon ECS introduces the <code>PutAccountSettingDefault</code> API	Amazon ECS introduces the <code>PutAccountSettingDefault</code> API that allows a user to set the default ARN/ID format opt in status for all the IAM users and roles on the account. Previously, setting the account's default opt in status required the use of the root user. For more information, see Amazon Resource Names (ARNs) and IDs (p. 104) .	8 Feb 2019
Interface VPC Endpoints (AWS PrivateLink)	Added support for configuring interface VPC endpoints powered by AWS PrivateLink. This allows you to create a private connection between your VPC and Amazon ECS without requiring access over the Internet, through a NAT instance, a VPN connection, or AWS Direct Connect. For more information, see Interface VPC Endpoints (AWS PrivateLink) . 26 Dec 2018	
AWS Fargate platform version 1.3.0	New AWS Fargate platform version released, which contains: <ul style="list-style-type: none">Added support for using AWS Systems Manager Parameter Store parameters to inject sensitive data into your containers. For more information, see Specifying Sensitive Data (p. 87).Added task recycling for Fargate tasks, which is the process of refreshing tasks that are a part of an Amazon ECS service. For more information, see Fargate Task Recycling (p. 115). For more information, see AWS Fargate platform versions (p. 14) .	17 Dec 2018
AWS Fargate Region expansion	AWS Fargate with Amazon ECS has expanded to the Asia Pacific (Mumbai) and Canada (Central) Regions.	07 Dec 2018

Change	Description	Date
Amazon ECS blue/green deployments	<p>Amazon ECS added support for blue/green deployments using CodeDeploy. This deployment type allows you to verify a new deployment of a service before sending production traffic to it.</p> <p>For more information, see Blue/Green Deployment with CodeDeploy (p. 143).</p>	27 Nov 2018
Resource tagging	<p>Amazon ECS added support for adding metadata tags to your services, task definitions, tasks, clusters, and container instances.</p> <p>For more information, see Resources and Tags (p. 177).</p>	15 Nov 2018
AWS Fargate Region expansion	<p>AWS Fargate with Amazon ECS has expanded to the US West (N. California) and Asia Pacific (Seoul) Regions.</p> <p>For more information, see AWS Fargate platform versions (p. 14).</p>	07 Nov 2018
Service limits updated	<p>The following service limits were updated:</p> <ul style="list-style-type: none"> • Number of tasks using the Fargate launch type, per Region, per account was raised from 20 to 50. • Number of public IP addresses for tasks using the Fargate launch type was raised from 20 to 50. <p>For more information, see Amazon ECS Service Quotas (p. 283).</p>	31 Oct 2018
AWS Fargate Region expansion	<p>AWS Fargate with Amazon ECS has expanded to the Europe (London) Region.</p> <p>For more information, see AWS Fargate platform versions (p. 14).</p>	26 Oct 2018
Private registry authentication support for Amazon ECS using AWS Fargate tasks	<p>Amazon ECS introduced support for Fargate tasks using private registry authentication using AWS Secrets Manager. This feature enables you to store your credentials securely and then reference them in your container definition, which allows your tasks to use private images.</p> <p>For more information, see Private Registry Authentication for Tasks (p. 85).</p>	10 Sept 2018

Change	Description	Date
Amazon ECS CLI v1.8.0	<p>New version of the Amazon ECS CLI released, which added the following functionality:</p> <ul style="list-style-type: none"> • Added support for Docker volumes in Docker compose files. • Added support for task placement constraints and strategies in Docker compose files. • Added support for private registry authentication in Docker compose files. • Added support for <code>--force-update</code> on <code>compose up</code> to force relaunching of tasks. <p>For more information, see the Amazon ECS Command Line Reference in the <i>Amazon Elastic Container Service Developer Guide</i>.</p>	7 Sept 2018
Amazon ECS service discovery Region expansion	<p>Amazon ECS service discovery has expanded support to the Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Tokyo), EU (Frankfurt), and Europe (London) Regions.</p> <p>For more information, see Service Discovery (p. 173).</p>	30 August 2018
Scheduled tasks with Fargate tasks support	<p>Amazon ECS introduced support for scheduled tasks for the Fargate launch type.</p> <p>For more information, see Scheduled Tasks (cron) (p. 111).</p>	28 August 2018
AWS Fargate Region expansion	<p>AWS Fargate with Amazon ECS has expanded to the Europe (Frankfurt), Asia Pacific (Singapore), and Asia Pacific (Sydney) Regions.</p> <p>For more information, see AWS Fargate platform versions (p. 14).</p>	19 July 2018
Amazon ECS CLI v1.7.0	<p>New version of the Amazon ECS CLI released, which added the following functionality:</p> <ul style="list-style-type: none"> • Added support for container healthcheck and devices in Docker compose files. For more information, see the Amazon ECS Command Line Reference in the <i>Amazon Elastic Container Service Developer Guide</i>. 	18 July 2018

Change	Description	Date
Amazon ECS service scheduler strategies added	<p>Amazon ECS introduced the concept of service scheduler strategies.</p> <p>There are two service scheduler strategies available:</p> <ul style="list-style-type: none"> • REPLICA—The replica scheduling strategy places and maintains the desired number of tasks across your cluster. By default, the service scheduler spreads tasks across Availability Zones. You can use task placement strategies and constraints to customize task placement decisions. For more information, see Replica (p. 117). • DAEMON—The daemon scheduling strategy deploys exactly one task on each active container instance that meets all of the task placement constraints that you specify in your cluster. The service scheduler evaluates the task placement constraints for running tasks and will stop tasks that do not meet the placement constraints. When using this strategy, there is no need to specify a desired number of tasks, a task placement strategy, or use Service Auto Scaling policies. For more information, see Daemon in the <i>Amazon Elastic Container Service Developer Guide</i>. <p>Note Fargate tasks do not support the DAEMON scheduling strategy.</p> <p>For more information, see Service scheduler concepts (p. 116).</p>	12 June 2018
Amazon ECS CLI v1.6.0	<p>New version of the Amazon ECS CLI released, which added the following functionality:</p> <ul style="list-style-type: none"> • Added support for Docker compose file syntax version 3. For more information, see the Amazon ECS Command Line Reference in the <i>Amazon Elastic Container Service Developer Guide</i>. 	5 June 2018
AWS Fargate Region expansion	<p>AWS Fargate with Amazon ECS has expanded to the US East (Ohio), US West (Oregon), and EU West (Ireland) Regions.</p> <p>For more information, see AWS Fargate platform versions (p. 14).</p>	26 April 2018

Change	Description	Date
Amazon ECS CLI v1.5.0	<p>New version of the Amazon ECS CLI released, which added the following functionality:</p> <ul style="list-style-type: none"> Added support for the ECS CLI to automatically retrieve the latest stable Amazon ECS-optimized AMI by querying the Systems Manager Parameter Store API during the cluster resource creation process. This requires the user account that you are using to have the required Systems Manager permissions. Added support for the <code>shm_size</code> and <code>tmpfs</code> parameters in compose files. <p>For more information, see the Amazon ECS Command Line Reference in the <i>Amazon Elastic Container Service Developer Guide</i>.</p>	19 April 2018
Amazon ECS CLI download verification	<p>Added new PGP signature method for verifying the Amazon ECS CLI installation file. For more information, see Installing the Amazon ECS CLI (p. 255).</p>	5 April 2018
AWS Fargate Platform Version	<p>New AWS Fargate platform version released, which contains:</p> <ul style="list-style-type: none"> Added support for Amazon ECS task metadata endpoint (p. 271). Added support for Health Check (p. 40). Added support for Service Discovery (p. 173) <p>For more information, see AWS Fargate platform versions (p. 14).</p>	26 March 2018
Amazon ECS Service Discovery	<p>Added integration with Route 53 to support Amazon ECS service discovery. For more information, see Service Discovery (p. 173).</p>	22 March 2018
Amazon ECS CLI v1.4.2	<p>New version of the Amazon ECS CLI released, which added the following functionality:</p> <ul style="list-style-type: none"> Updated the AMI to <code>amzn-ami-2017.09.k-amazon-ecs-optimized</code>. <p>For more information, see the Amazon ECS Command Line Reference in the <i>Amazon Elastic Container Service Developer Guide</i>.</p>	20 March 2018

Change	Description	Date
Amazon ECS CLI v1.4.0	<p>New version of the Amazon ECS CLI released, which added the following functionality:</p> <ul style="list-style-type: none"> Added support for the us-gov-west-1 Region. Added <code>--force-deployment</code> flag for the <code>compose service</code> command. Added support for <code>aws_session_token</code> in ECS profiles. Updated the AMI to <code>amzn-ami-2017.09.j-amazon-ecs-optimized</code>. <p>For more information, see the Amazon ECS Command Line Reference in the <i>Amazon Elastic Container Service Developer Guide</i>.</p>	09 March 2018
Container Health Checks	<p>Added support for Docker health checks in container definitions. For more information, see Health Check (p. 40).</p>	08 March 2018
Amazon ECS Task Metadata Endpoint	<p>Beginning with version 1.17.0 of the Amazon ECS container agent, various task metadata and Docker stats are available to tasks that use the <code>awsvpc</code> network mode at an HTTP endpoint that is provided by the Amazon ECS container agent. For more information, see Amazon ECS task metadata endpoint (p. 271).</p>	8 February 2018
Amazon ECS Service Auto Scaling using target tracking policies	<p>Added support for ECS Service Auto Scaling using target tracking policies in the Amazon ECS console. For more information, see Target Tracking Scaling Policies (p. 166).</p> <p>Removed the previous tutorial for step scaling in the ECS first run wizard. This was replaced with the new tutorial for target tracking.</p>	8 February 2018
Amazon ECS CLI v1.3.0	<p>New version of the Amazon ECS CLI released, which added the following functionality:</p> <ul style="list-style-type: none"> Ability to create empty clusters with the <code>up</code> command. Added <code>--health-check-grace-period</code> flag for the <code>compose service up</code> command. Updated the AMI to <code>amzn-ami-2017.09.g-amazon-ecs-optimized</code>. <p>For more information, see the Amazon ECS Command Line Reference in the <i>Amazon Elastic Container Service Developer Guide</i>.</p>	19 January 2018
New service scheduler behavior	<p>Updated information about the behavior for service tasks that fail to launch. Documented new service event message that triggers when a service task has consecutive failures. For more information about this updated behavior, see Additional service concepts (p. 117).</p>	11 January 2018

Change	Description	Date
Task-level CPU and memory	Added support for specifying CPU and memory at the task-level in task definitions. For more information, see TaskDefinition .	12 December 2017
Amazon ECS console CodePipeline integration	Added Amazon ECS integration with CodePipeline. CodePipeline supports Amazon ECS as a deployment option to help set up deployment pipelines. For more information, see ??? .	12 December 2017
Task execution role	<p>The Amazon ECS container agent makes calls to the Amazon ECS API actions on your behalf, so it requires an IAM policy and role for the service to know that the agent belongs to you. The following actions are covered by the task execution role:</p> <ul style="list-style-type: none"> • Calls to Amazon ECR to pull the container image • Calls to CloudWatch to store container application logs <p>For more information, see Amazon ECS Task Execution IAM Role (p. 236).</p>	7 December 2017
Amazon ECS CLI v1.1.0 with Fargate support	<p>New version of the Amazon ECS CLI released, which added the following features:</p> <ul style="list-style-type: none"> • Support for task networking • Support for AWS Fargate • Support for viewing CloudWatch Logs data from a task <p>For more information, see ECS CLI changelog.</p>	29 November 2017
AWS Fargate GA	Added support for launching Amazon ECS services using the Fargate launch type. For more information, see Amazon ECS Launch Types (p. 60) .	29 November 2017

AWS glossary

For the latest AWS terminology, see the [AWS glossary](#) in the *AWS General Reference*.