



EMR-LAB-Setupcreate s3 bucket



Status

Not Started

step1- creat

Amazon S3 > Buckets > Create bucket

Create bucket [Info](#)

Buckets are containers for data stored in S3.

General configuration

AWS Region

Asia Pacific (Singapore) ap-southeast-1

Bucket type [Info](#)

☒ **General purpose**
Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.

☐ **Directory**
Recommended for specialized low-latency use cases supported by AWS Availability Zones or data residency use cases supported by AWS Local Zones.

Bucket name [Info](#)

clahan-EMR-bucket-1

Bucket names must be 3 to 63 characters and unique within the global namespace. Bucket names must also begin and end with a letter or number. Valid characters are a-z, 0-9, periods (.), and hyphens (-). [Learn more](#)

Copy settings from existing bucket - optional

next bucket version

Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. [Learn more](#)

Bucket Versioning

☐ Disable

☒ Enable

next go to default encryption

Default encryption [Info](#)

Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type [Info](#)

Secure your objects with two separate layers of encryption. For details on pricing, see DSSE-KMS pricing on the Storage tab of the [Amazon S3 pricing page](#).

☒ Server-side encryption with Amazon S3 managed keys (SSE-S3)

☐ Server-side encryption with AWS Key Management Service keys (SSE-KMS)

☐ Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)

Bucket Key

Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. [Learn more](#)

☐ Disable

☒ Enable

inside the bucket we need to create the folders

EMR-LAB-Setupcreate s3 bucket

1

Amazon S3 > Buckets > clahan-emr-bucket-1 > Create folder

[configuration](#) to upload an empty folder and specify the appropriate settings.

Folder

Folder name

monthly_build /

Folder names can't contain "/". [See rules for naming](#)

Server-side encryption [Info](#)

Server-side encryption protects data at rest.

The following encryption settings apply only to the folder object and not to sub-folder objects.

Server-side encryption

☐ Don't specify an encryption key
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.

☒ Specify an encryption key
The specified encryption key is used to encrypt the folder object before storing it in Amazon S3.

Encryption settings [Info](#)

☐ Use bucket settings for default encryption

☒ Override bucket settings for default encryption

inside this bucket also we need to create fodlers

Amazon S3 > Buckets > clahan-emr-bucket-1 > monthly_build/ > 2025-11/ > Create folder

Folder

Folder name

input /

Folder names can't contain "/". [See rules for naming](#)

Server-side encryption [Info](#)

Server-side encryption protects data at rest.

The following encryption settings apply only to the folder object and not to sub-folder objects.

Server-side encryption

☐ Don't specify an encryption key
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.

☒ Specify an encryption key
The specified encryption key is used to encrypt the folder object before storing it in Amazon S3.

Encryption settings [Info](#)

☒ Use bucket settings for default encryption

☐ Override bucket settings for default encryption

VPC > Your VPCs > Create VPC

VPC settings

Resources to create [Info](#)
Create only the VPC resource or the VPC and other networking resources.

☐ VPC only ☒ VPC and more

Name tag auto-generation [Info](#)
Enter a value for the Name tag. This value will be used to auto-generate Name tags for all resources in the VPC.

☒ Auto-generate
clahan-emr-vpc

IPv4 CIDR block [Info](#)
Determine the starting IP and the size of your VPC using CIDR notation.

10.0.0.0/16 65,536 IPs
CIDR block size must be between /16 and /28.

IPv6 CIDR block [Info](#)

Preview

[Show details](#)
AWS virtual network

ian-emr-vpc-vpc

Subnets (2)
Subnets within this VPC

ap-southeast-1a
clahan-emr-vpc-subnet-public1-ap-

ap-southeast-1b
clahan-emr-vpc-subnet-public2-ap-

Route tables (1)
Route network traffic to resourc

clahan-emr-vpc-rtb-public

VPC > Your VPCs > Create VPC

Default

► **Encryption settings - optional**

Number of Availability Zones (AZs) [Info](#)
Choose the number of AZs in which to provision subnets. We recommend at least two AZs for high availability.

1 2 3

► **Customize AZs**

Number of public subnets [Info](#)
The number of public subnets to add to your VPC. Use public subnets for web applications that need to be publicly accessible over the internet.

0 2

Number of private subnets [Info](#)
The number of private subnets to add to your VPC. Use private subnets to secure backend resources that don't need public access.

0 2 4

Preview

[Show details](#)
AWS virtual network

ian-emr-vpc-vpc

Subnets (2)
Subnets within this VPC

ap-southeast-1a
clahan-emr-vpc-subnet-public1-ap-

ap-southeast-1b
clahan-emr-vpc-subnet-public2-ap-

Route tables (1)
Route network traffic to resou

clahan-emr-vpc-rtb-publ

Press the "Print Scrn" on your keyboard to

VPC > Your VPCs > Create VPC > Create VPC resources

Create VPC workflow

✓ Success

▼ Details

- ✓ Create VPC: [vpc-0ec8eb9eba0f4f1c6](#)
- ✓ Enable DNS hostnames
- ✓ Enable DNS resolution
- ✓ Verifying VPC creation: [vpc-0ec8eb9eba0f4f1c6](#)
- ✓ Create S3 endpoint: [vpce-06e9270d82ac40c20](#)
- ✓ Create subnet: [subnet-075f16772f818f4c9](#)
- ✓ Create subnet: [subnet-0f0d7774cd62b526a](#)
- ✓ Create internet gateway: [igw-07ccd70269edb2559](#)
- ✓ Attach internet gateway to the VPC
- ✓ Create route table: [rtb-03eb0ca724c29155a](#)
- ✓ Create route
- ✓ Associate route table
- ✓ Associate route table
- ✓ Verifying route table creation

Step3:- Create EMR cluster

Create cluster [Info](#)

▼ Name and applications - *required* [Info](#)








Name your cluster and choose the applications that you want to install to your cluster.

Name

Amazon EMR release [Info](#)

A release contains a set of applications which can be installed on your cluster.

Application bundle

Spark Interactive 	Core Hadoop 	Flink 	HBase 	Presto 	Trino 	Custom 
---	---	--	--	---	--	---

▼ Cluster configuration - *required* [Info](#)

Choose a configuration method for the primary, core, and task node groups for your cluster.

☒ Uniform instance groups

Choose the same EC2 instance type and purchasing option (On-Demand or Spot) for all nodes in your node group. [Learn more](#)

☐ Flexible instance fleets

Choose from the widest variety of provisioning options for the EC2 instances in your cluster. Diversify instance types and purchasing options, and use an allocation strategy. [Learn more](#)

Uniform instance groups

Primary

Choose EC2 instance type

4 vCore 16 GiB memory

EBS only storage

On-Demand price: \$0.240 per instance/...

Lowest Spot price: \$0.057 (ap-southeast-1c)

Actions ▼

☐ Use high availability

Launch highly available, more resilient cluster with three primary nodes on On-Demand Instances. This configuration applies for the lifetime of your cluster. [Learn more](#)

▼ Cluster termination and node replacement [Info](#)

Choose termination settings and protect your cluster from accidental shutdown.

Termination option

- ☐ Manually terminate cluster
- ☐ Automatically terminate cluster after last step ends
- ☒ Automatically terminate cluster after idle time (Recommended)

Idle time

Enter the time until your cluster terminates.

0 days ▼

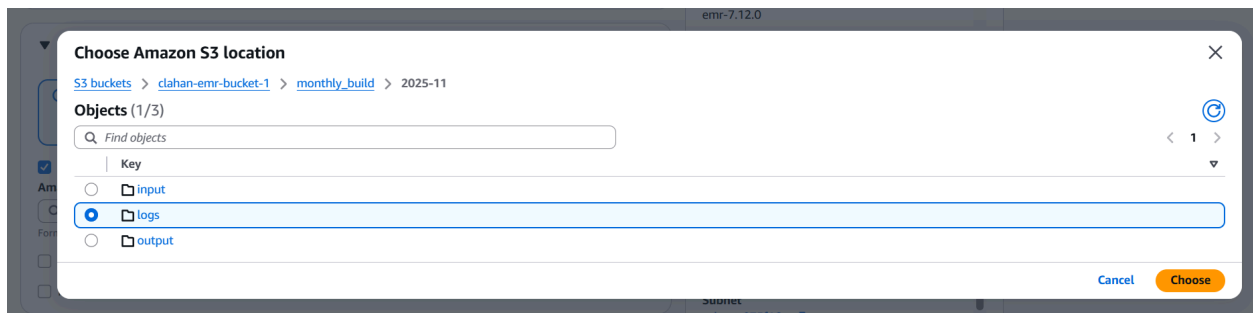
03:00:00

Choose a time that is greater than 1 minute (00:01:00) and less than 7 days. The time is in hh:mm:ss (24-hour) format.

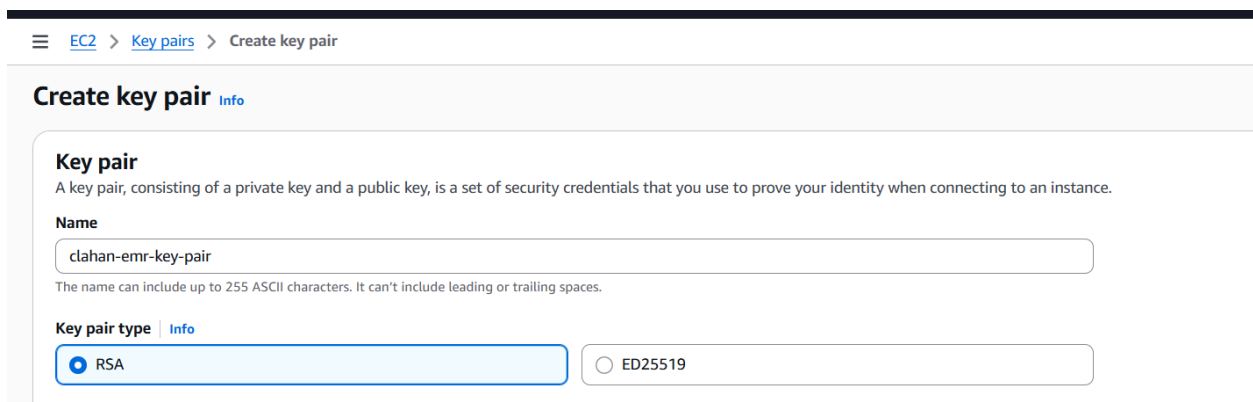
☒ Use termination protection

Protects your cluster from accidental termination. If on, you must first turn off protection to terminate the cluster. We recommend turning on termination protection for your long running clusters.

cluster logs



for security we need to create key pair



▼ Security configuration and EC2 key pair [Info](#)

Choose a security configuration or create a new one that you can reuse with other clusters.

Security configuration

Select your cluster encryption, authentication, authorization, and instance metadata service settings.






[Browse](#) 

[Create security configuration](#) 

Amazon EC2 key pair for SSH to the cluster [Info](#)





[Browse](#)

[Create key pair](#) 

create identity and access Management (IAM) roles

▼ Identity and Access Management (IAM) roles - *required* [Info](#)

Choose or create a service role and instance profile for the EC2 instances in your cluster.

Amazon EMR service role [Info](#)

The service role is an IAM role that Amazon EMR assumes to provision resources and perform service-level actions with other AWS services.

☐ **Choose an existing service role**
Select a default service role or a custom role with IAM policies attached so that your cluster can interact with other AWS services.


☒ **Create a service role**
Let Amazon EMR create a new service role so that you can grant and restrict access to resources in other AWS services.

Networking resources

We've already added the resources that you configured in the [Networking](#) section. Choose the VPC, subnet, and security groups that the service role can access.


Virtual Private Cloud (VPC)





Subnet





EC2 instance profile for Amazon EMR

The instance profile assigns a role to every EC2 instance in a cluster. The instance profile must specify a role that can access the resources for your steps and bootstrap actions.

☐ Choose an existing instance profile

Select a default role or a custom instance profile with IAM policies attached so that your cluster can interact with your resources in Amazon S3.

☒ Create an instance profile

Let Amazon EMR create a new instance profile so that you can specify a custom set of resources for it to access in Amazon S3.

S3 bucket access [Info](#)

☐ Specific S3 buckets or prefixes in your account [Info](#)

Choose the buckets or prefixes that you want this instance profile to access.

☒ All S3 buckets in this account with read and write access

Grant the instance profile access to all buckets that have read and write access enabled in your account.

THE RED CIRCLE RECOMENDED FOR PRACTICE