

Run Hadoop job on AWS

1. In order to run any hadoop job, you first need to create a S3 buckets and put your code and input there.
2. From your AWS Console, navigate to S3
3. Click create bucket and enter your desired bucket name (no space or special characters). Then click Create



Create a Bucket - Select a Bucket Name and Region Cancel X

A bucket is a container for objects stored in Amazon S3. When creating a bucket, you can choose a Region to optimize for latency, minimize costs, or address regulatory requirements. For more information regarding bucket naming conventions, please visit the [Amazon S3 documentation](#).

Bucket Name:

Region: ▼


4. Click on your bucket name to browse its content (currently empty). Upload your Jar file and input (if any) using **Upload** button
5. All done. Next step will be configure AWS Elastic MapReduce to run your job
6. From your AWS Console, navigate to **Elastic MapReduce**
7. Click Create cluster. You only need to change the following configuration, **keep everything else default**.
8. Cluster Configuration: Enter your desired name and s3 bucket name for logs

Cluster Configuration

Cluster name

Termination protection Yes
 No

Logging Enabled

Log folder S3 location
 
s3://<bucket-name>/<folder>/

Debugging Enabled

4. Software Configuration: Select AMI version 2.4.8 (hadoop 1.0.3)

Software Configuration

Hadoop distribution Amazon

AMI version

MapR

5. Steps: Select Custom Jar from Add step selection. Change Auto-terminate to **No** (otherwise cluster will automatically turn off and you need to create another cluster)

Steps

i A step is a unit of work you submit to the cluster. A step might contain one or more Hadoop tasks that run on the cluster to configure an application. You can submit up to 256 steps to a cluster. [Learn more](#)


Name	Action on failure	JAR location
Add step Custom JAR 		
Configure and add		
Auto-terminate	<input type="radio"/> Yes	
	<input checked="" type="radio"/> No	

6. Click **Configure and add** button and fill in the location of your Jar, location of your input and output in S3 as **Arguments**. Click **Add** when finish


Add Step

Step type Custom JAR

Name* Custom JAR

JAR location* s3://tuananhtest/wordcount.jar 

Arguments s3://tuananhtest/input s3://tuananhtest/output

Action on failure Continue 

7. Click Create cluster. Your cluster will be up and running shortly.
8. When everything is finished, navigate to Amazon S3 and check your output.
9. If you want to run another job, navigate to Elastic MapReduce and click on your cluster, then click **Add step**.
10. **Important: When you finished running all jobs, navigate to Elastic MapReduce and Terminate your cluster. Otherwise the cluster you created will keep running and your credit will run out shortly.**