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](image)

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

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)
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.

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.