



1. Spark and airflow image will be registered in ECR registry
2. EMR will use these images to create the cluster
3. Job will be orchestrated in Airflow which will run the main.py script (placed in S3)