Spark AWS EMR Guide

1. Login to AWS Console
2. Select EMR
3. Create cluster (software configuration: Spark; choose Large image)
4. Upload input files to S3
5. Get public DNS of the master instance
6. Login to master node via ssh to master node’s public DNS
7. pyspark (now similar to XSEDE)