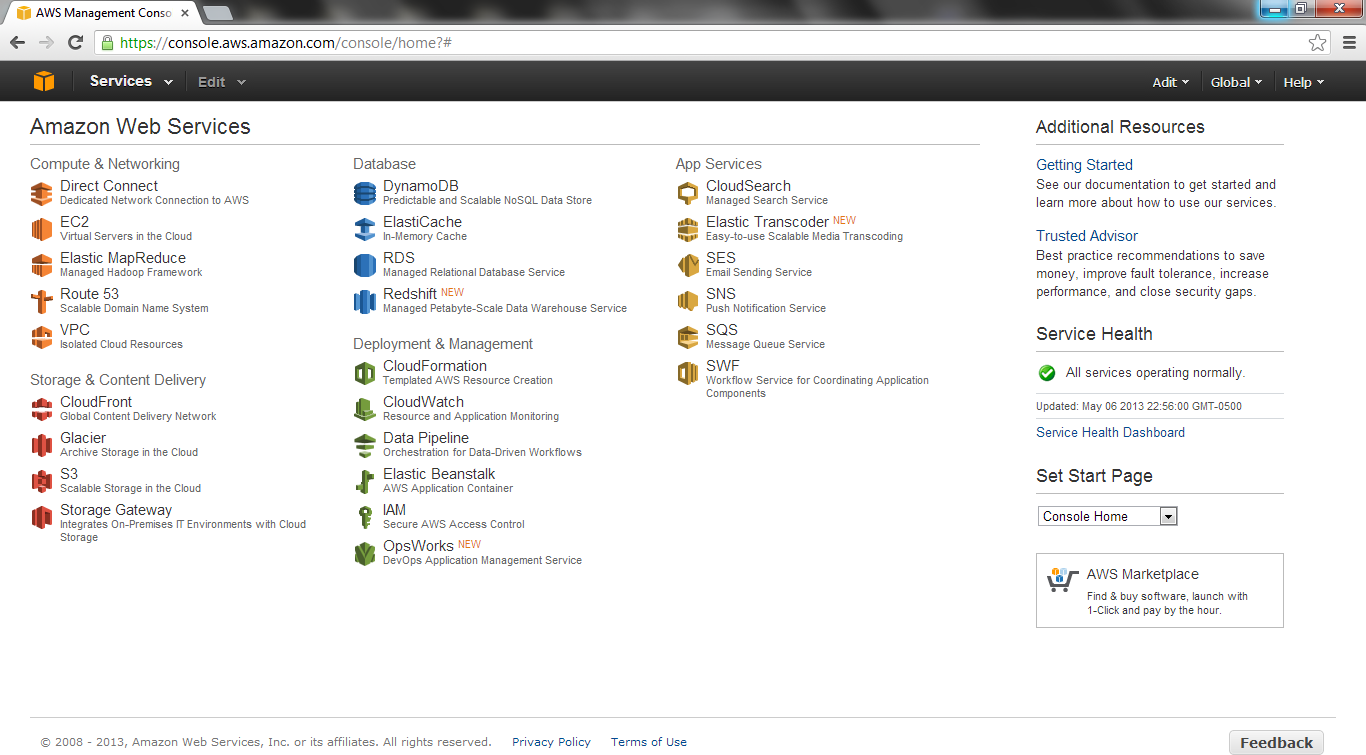
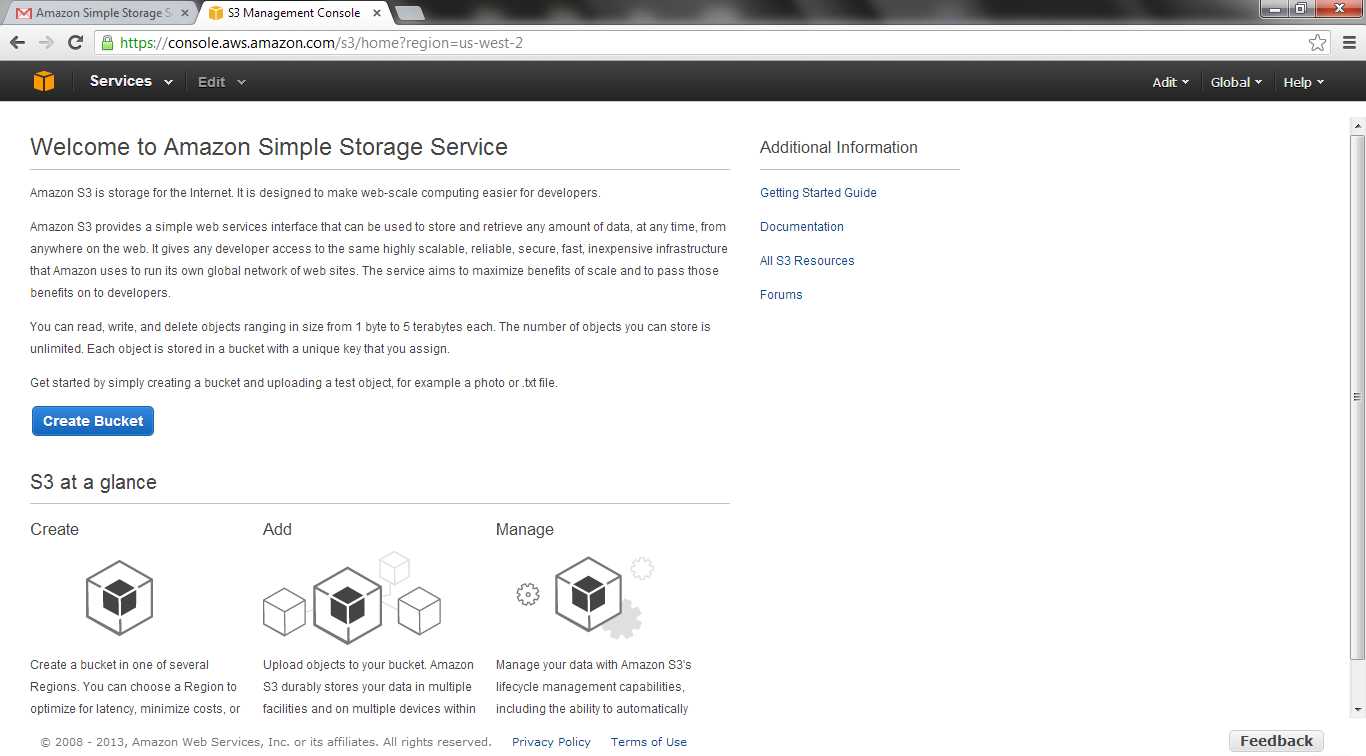
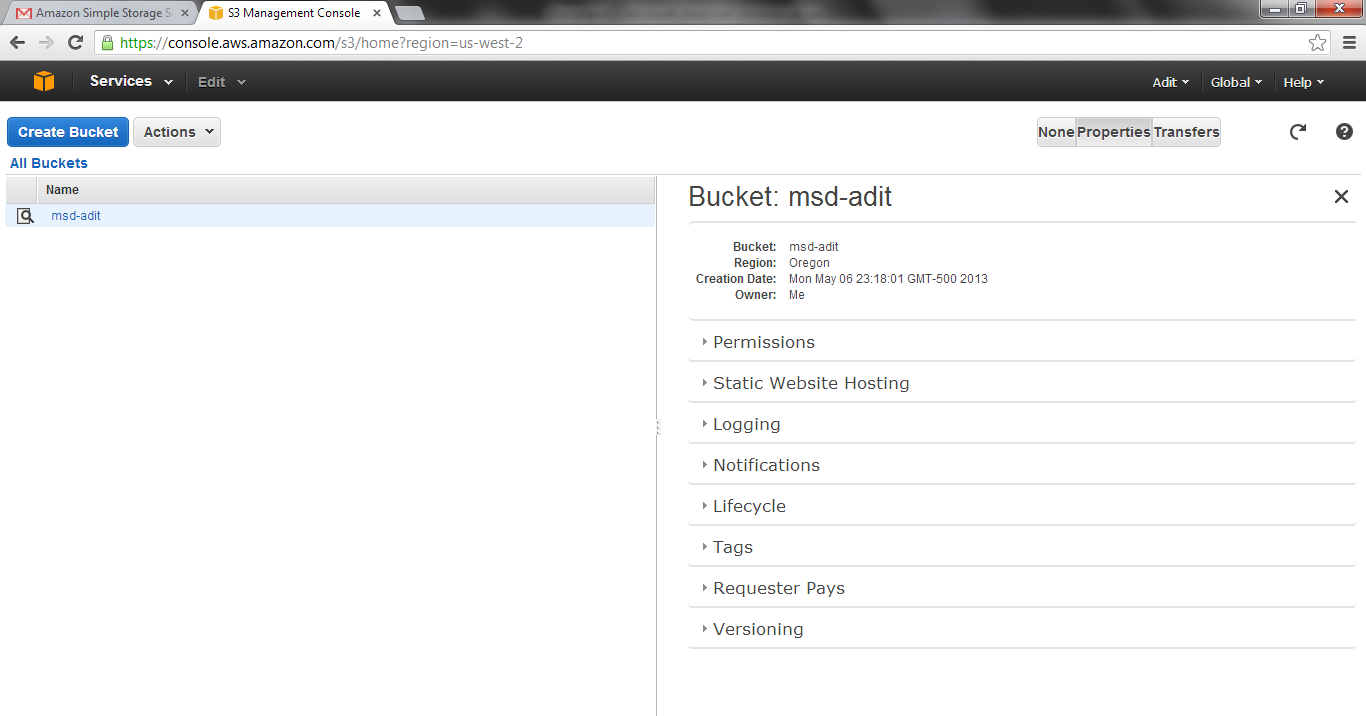
1. Sign in to the AWS Management console

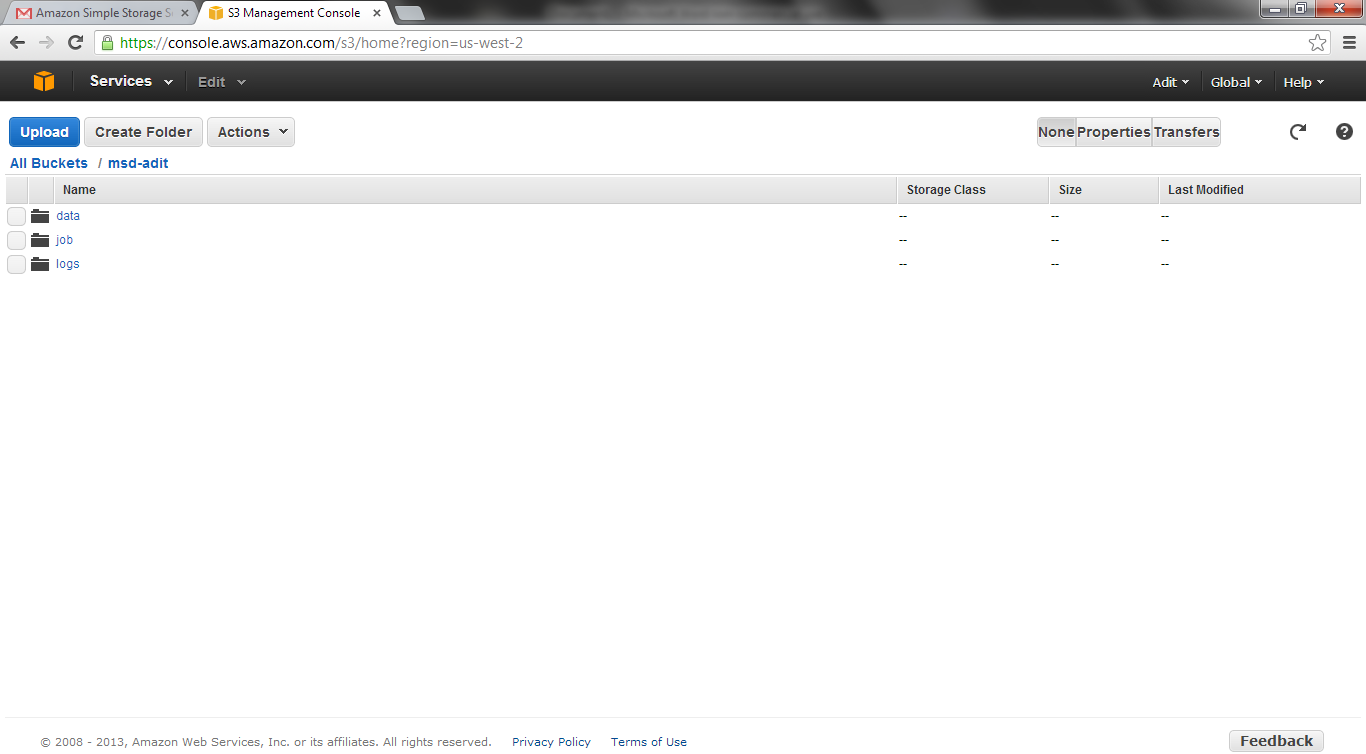


1. Create Amazon S3 bucket

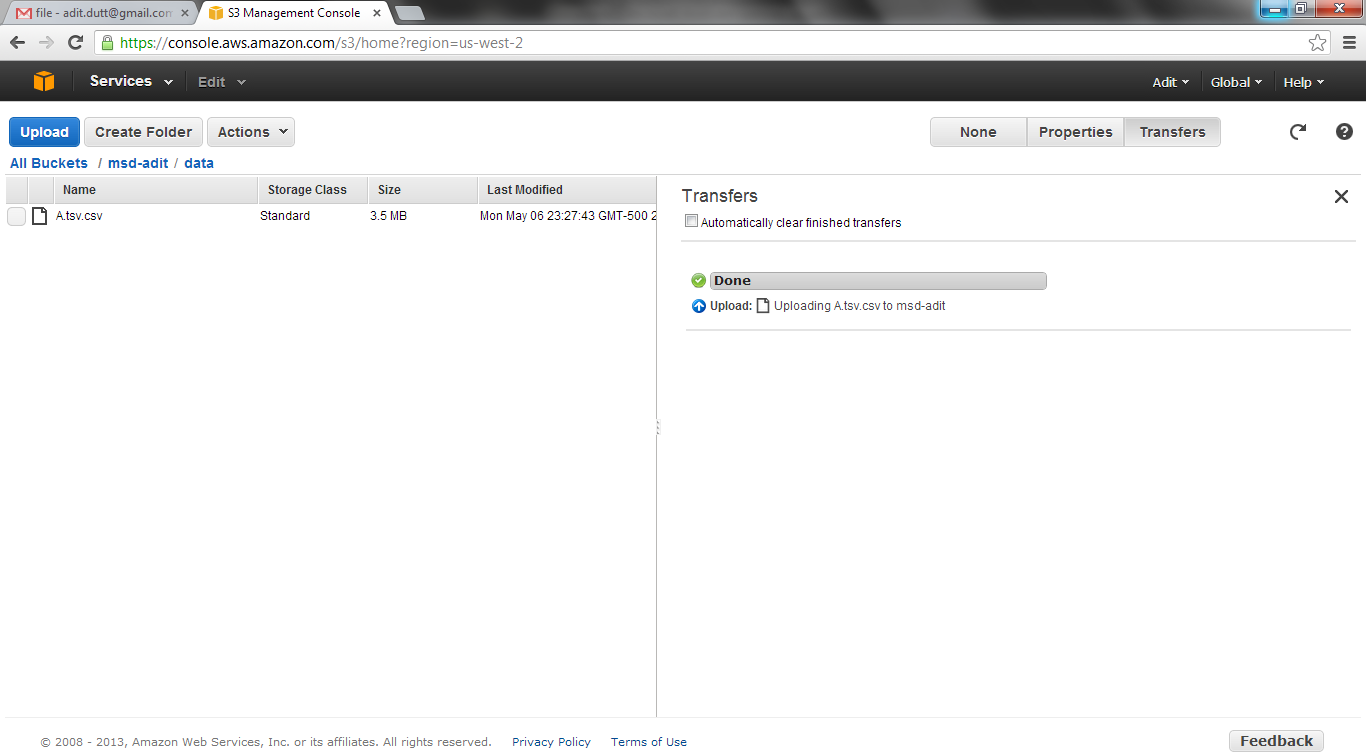




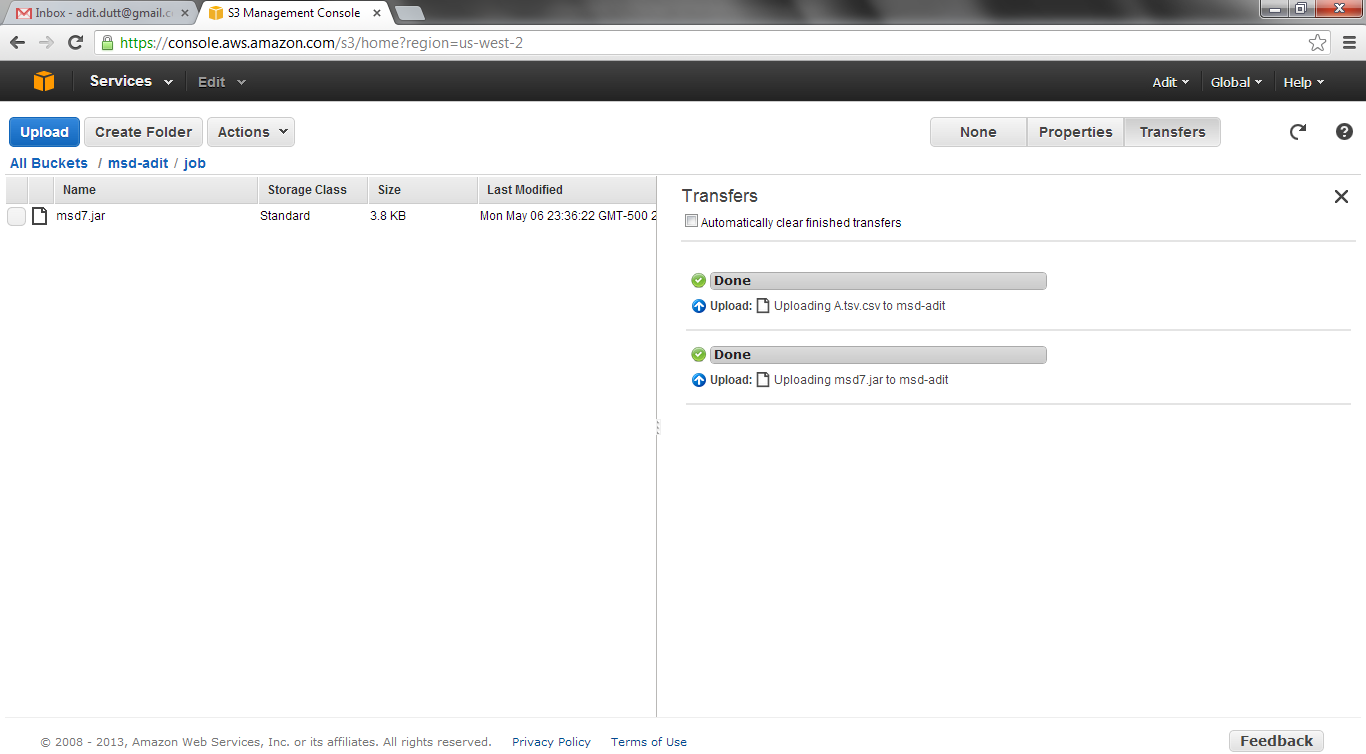
1. Create folders in the bucket



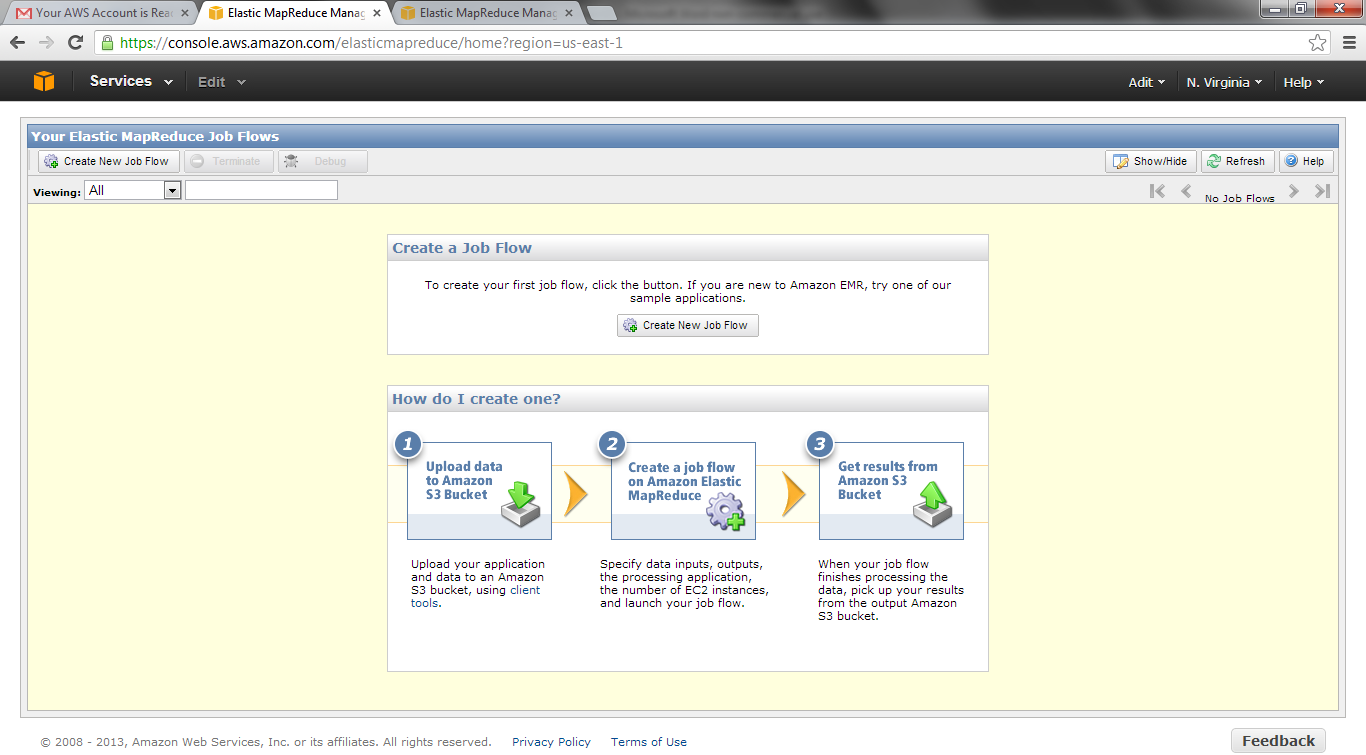
1. Upload sample input (test) data

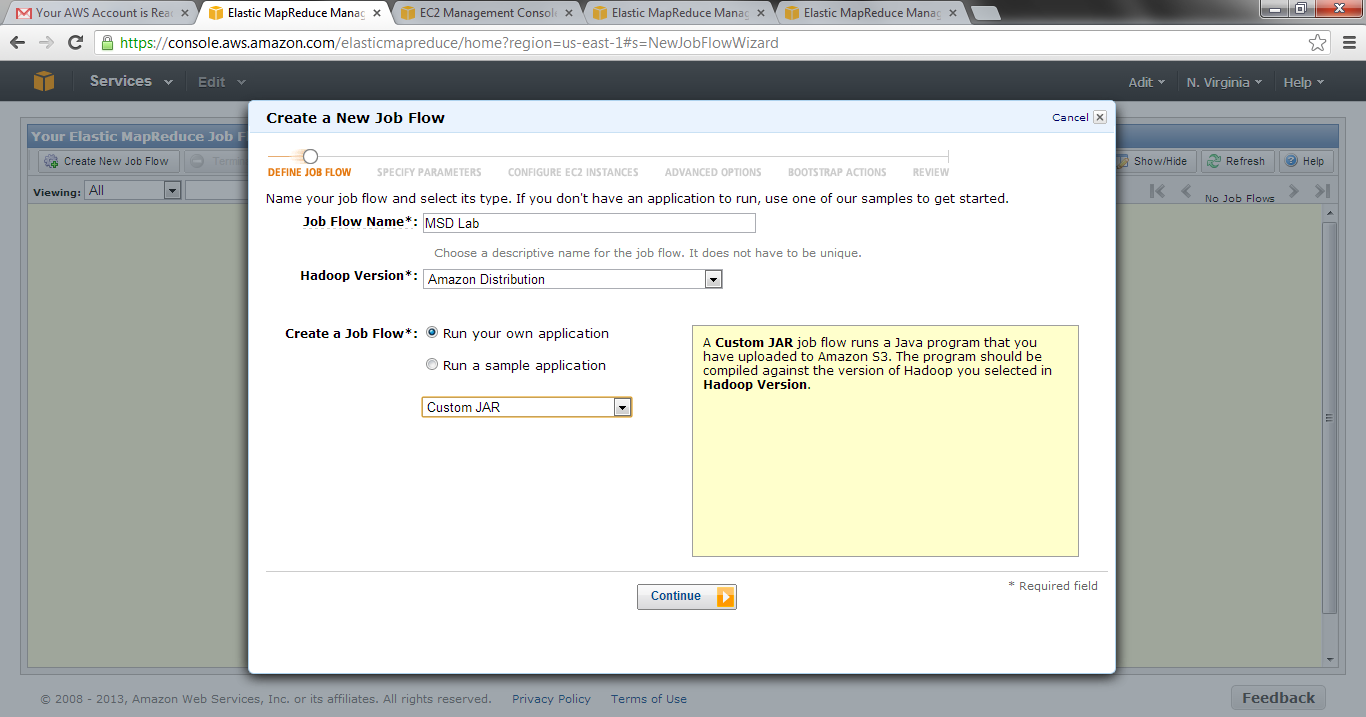


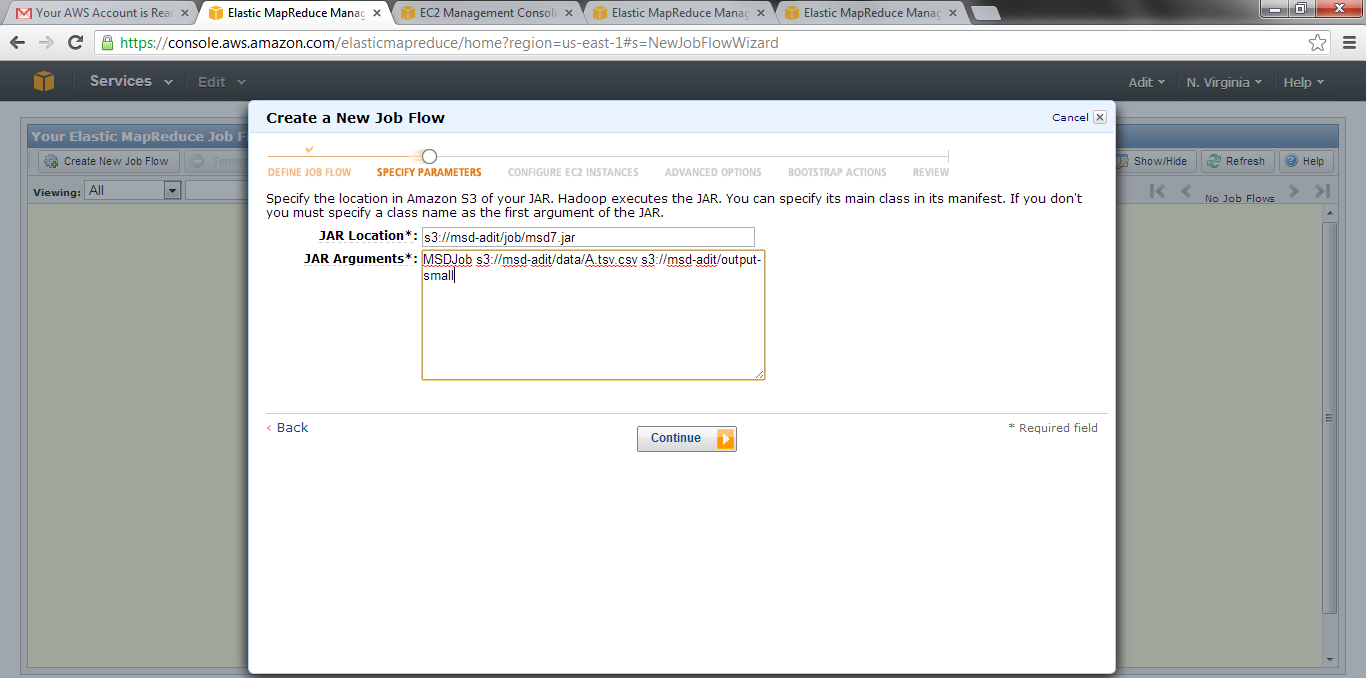
1. Upload the custom jar file for the MSD MapReduce Job:

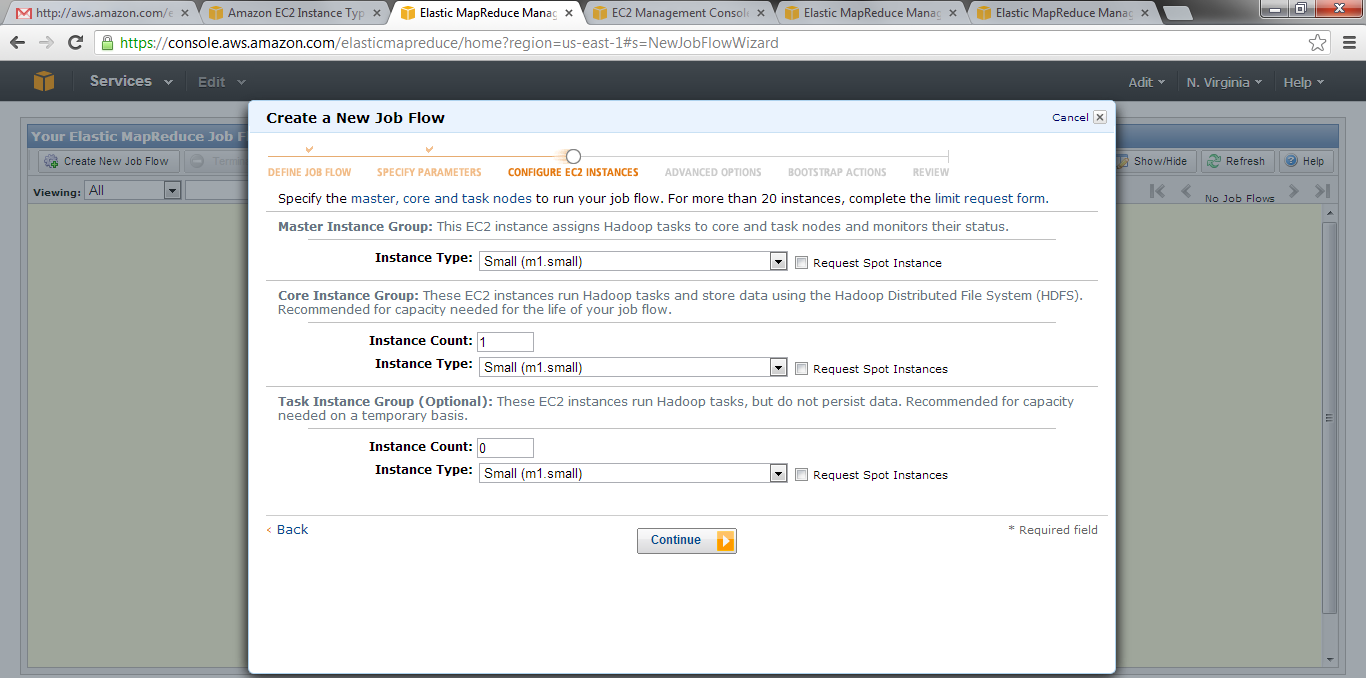


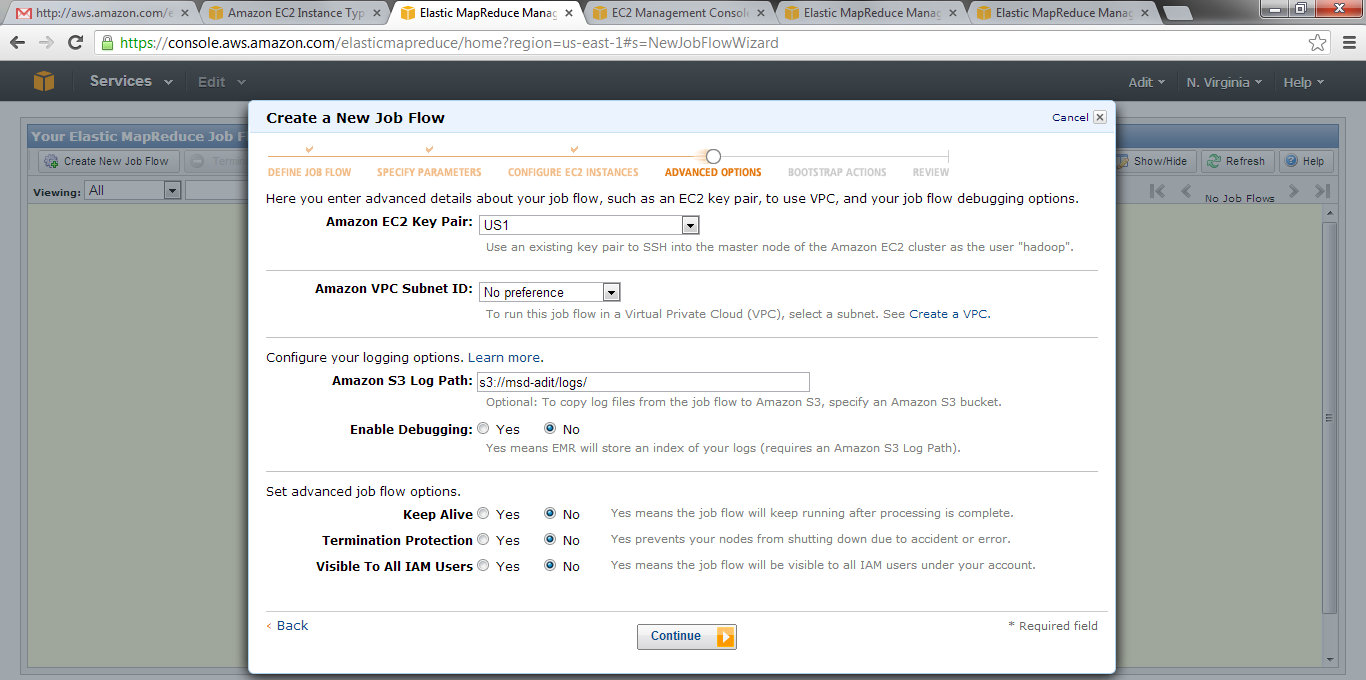
1. Create new job flow and launch the cluster

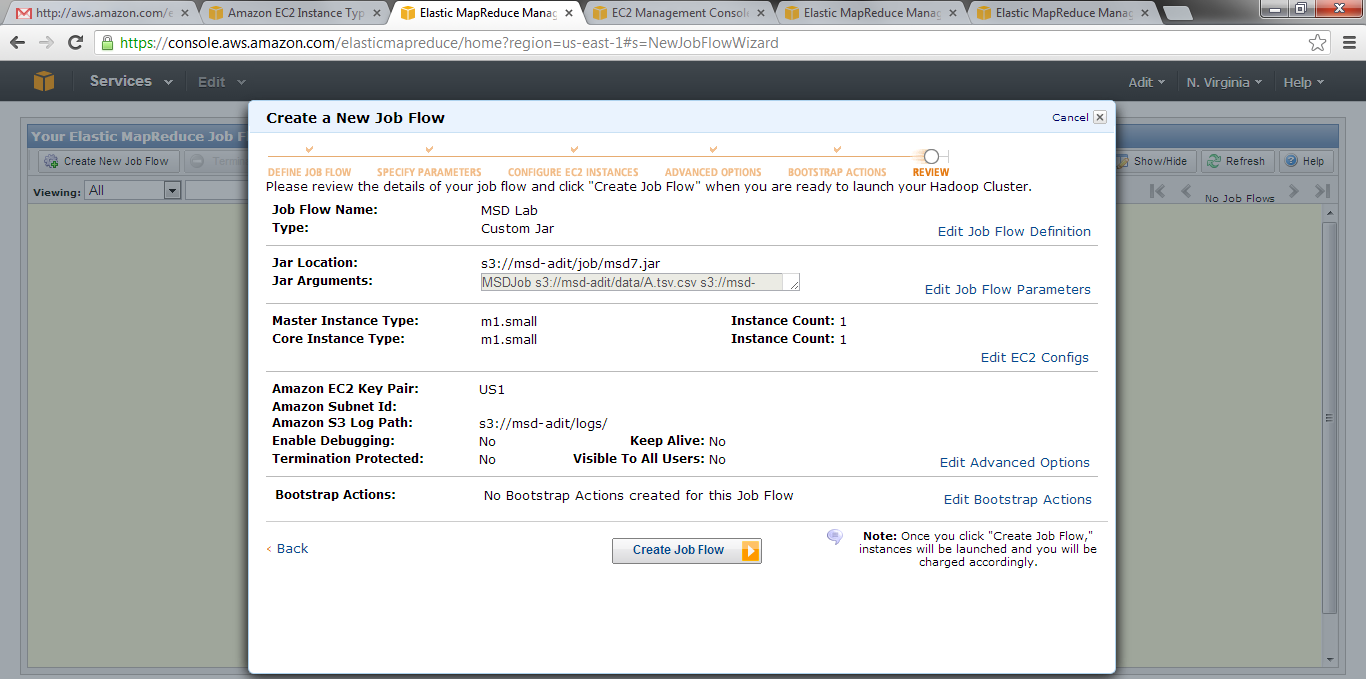


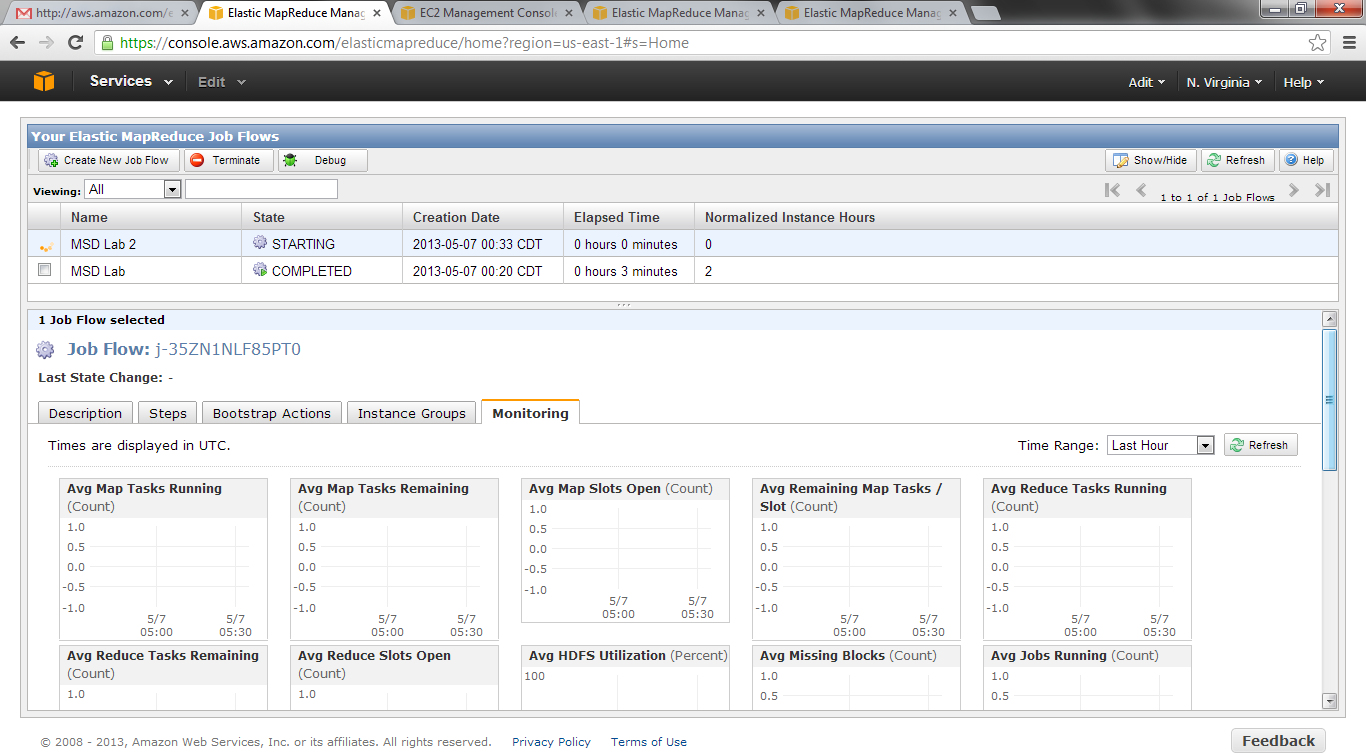


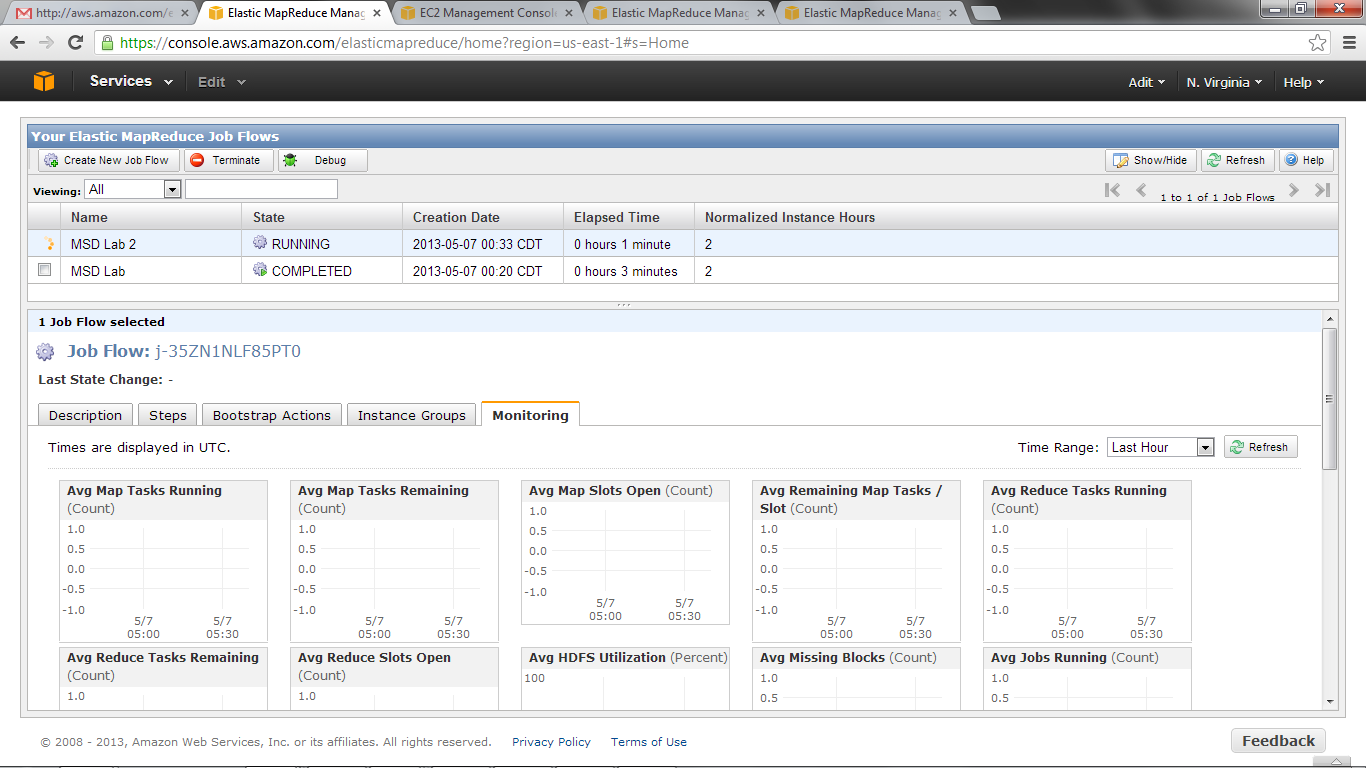


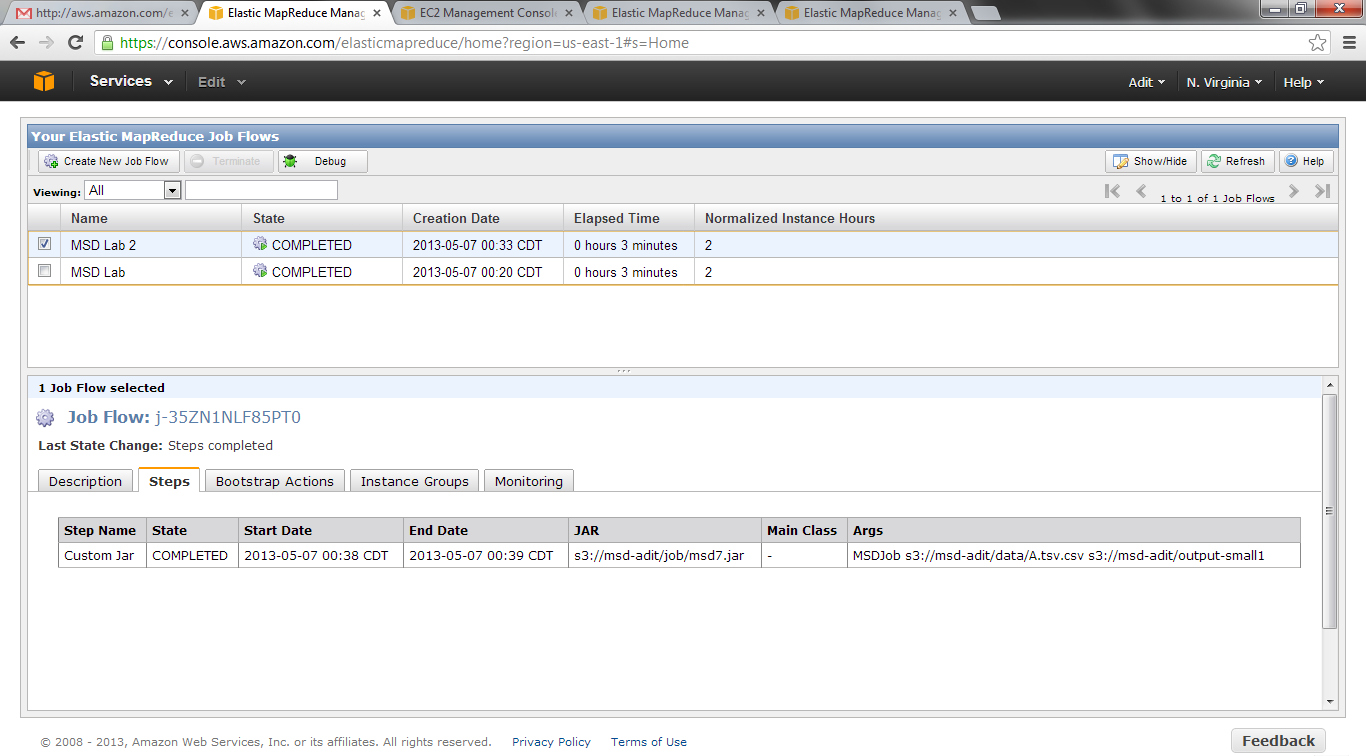




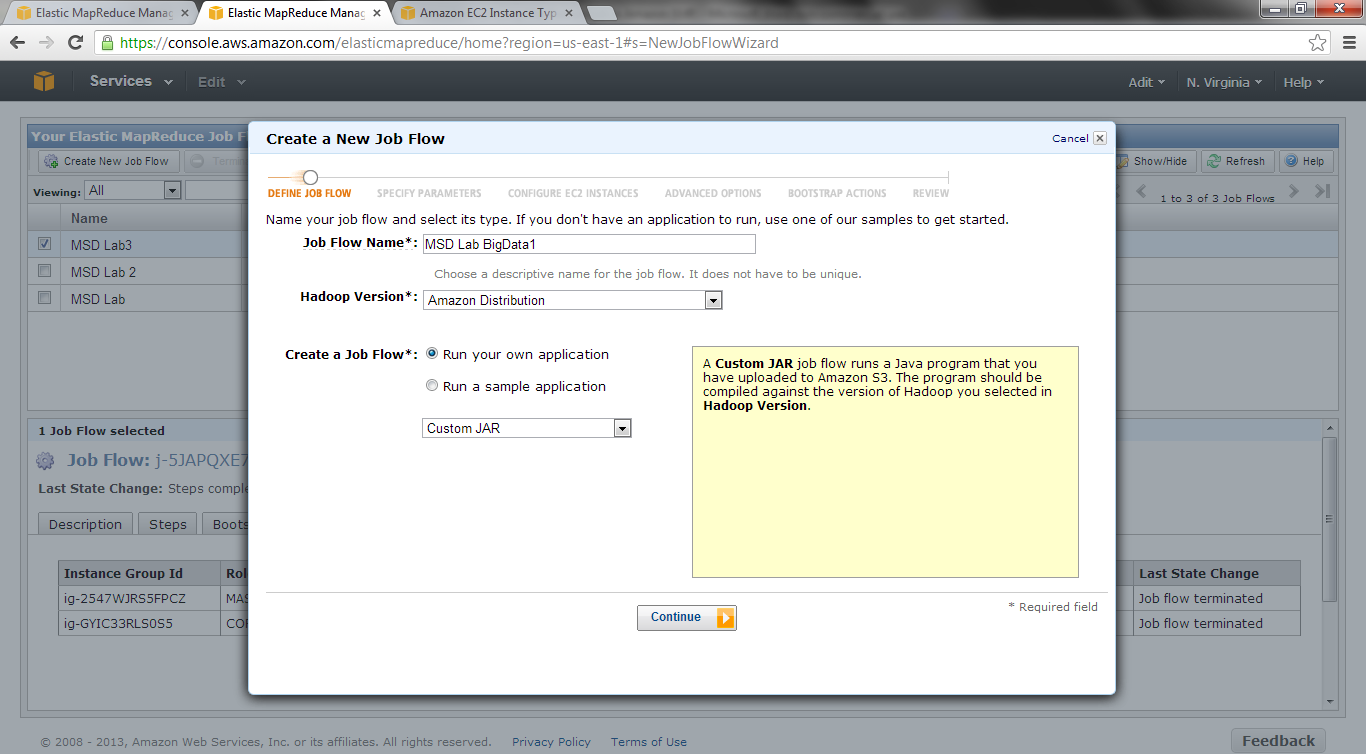


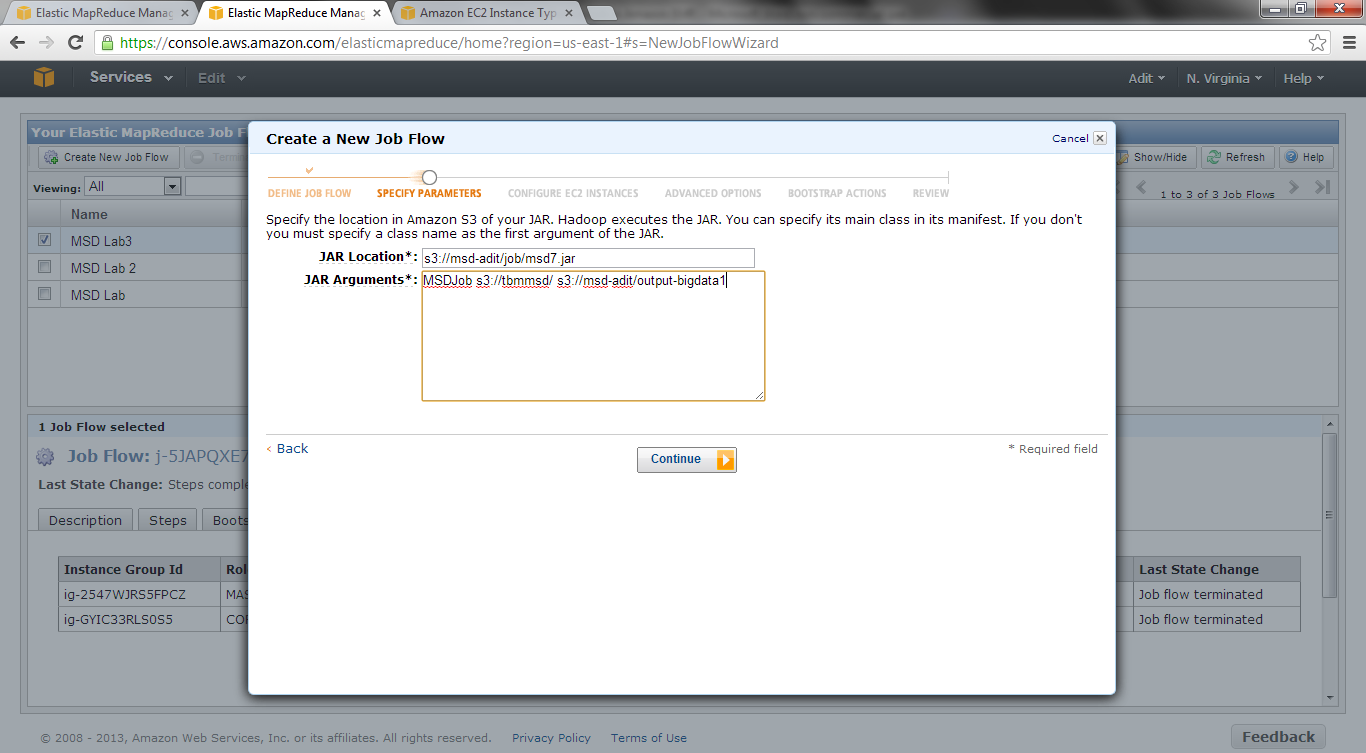




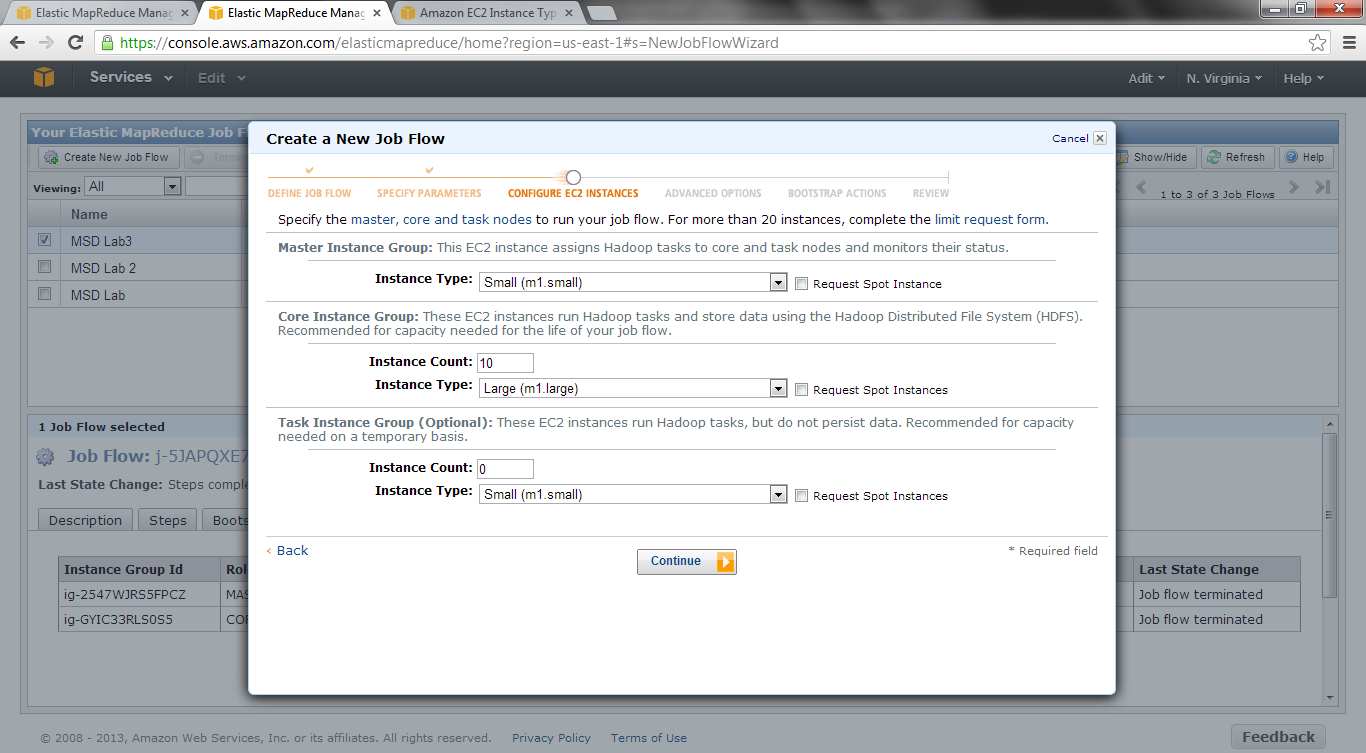


Big Data:

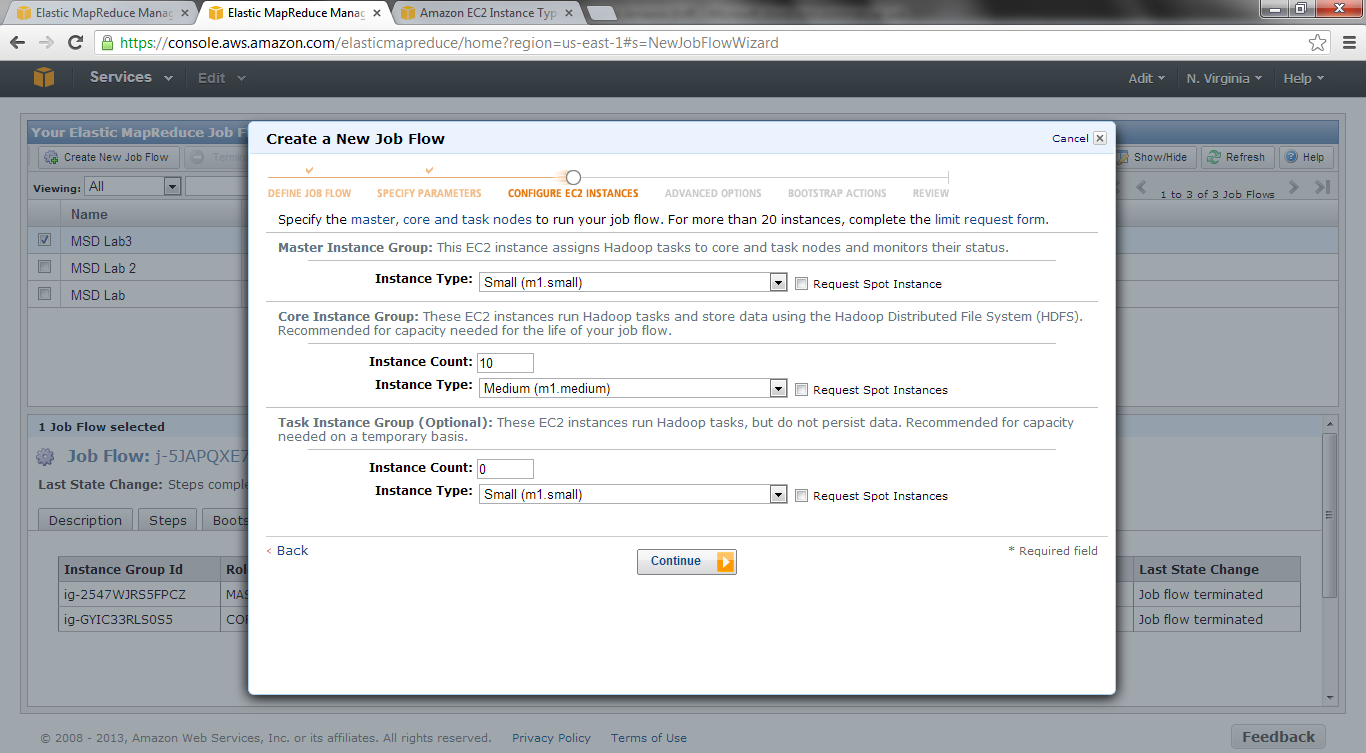


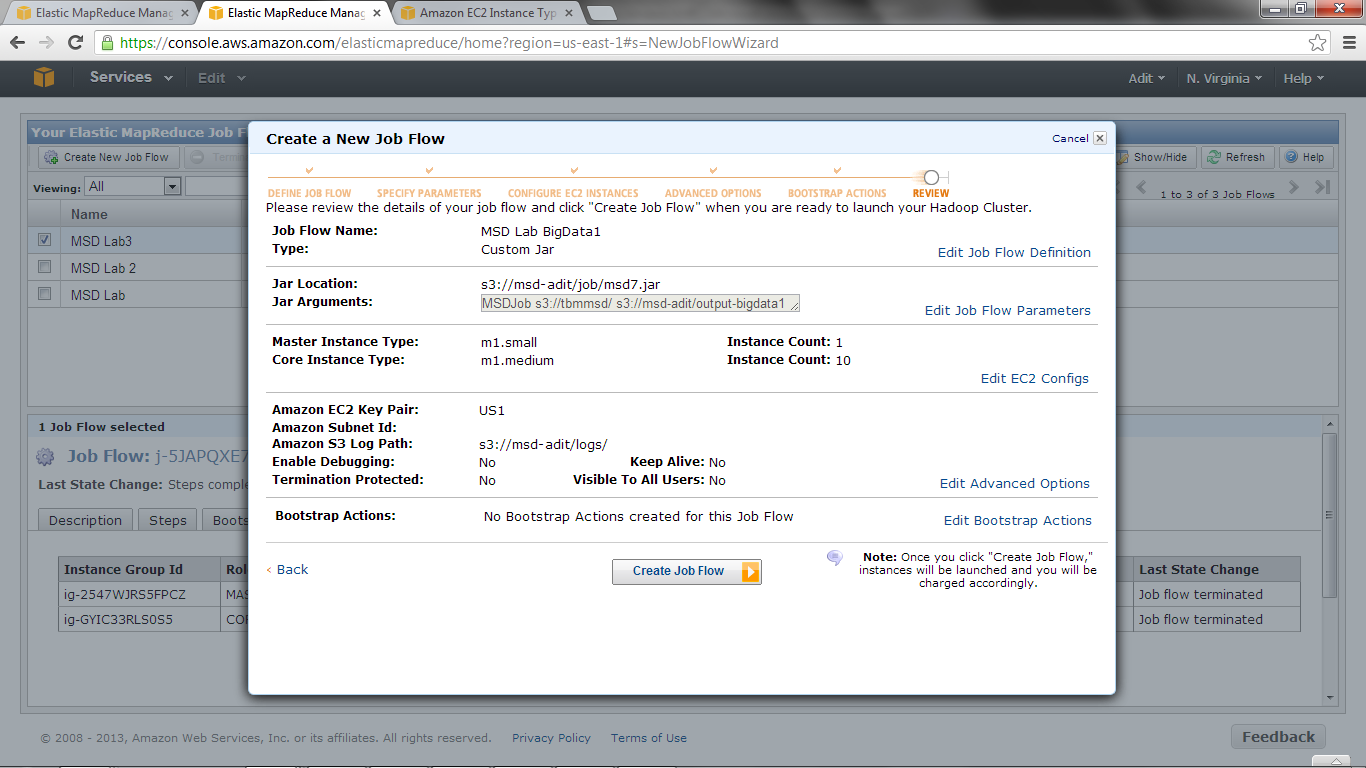


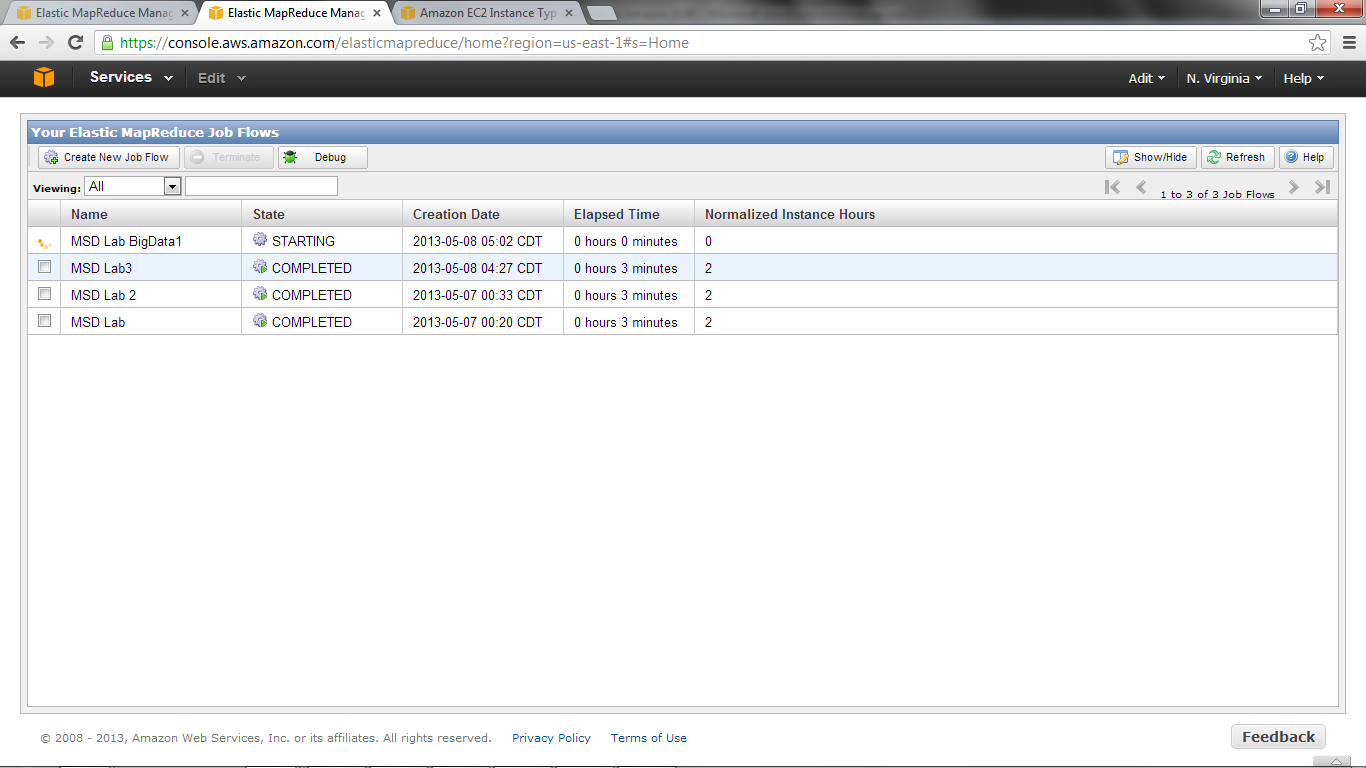
If we consider large:



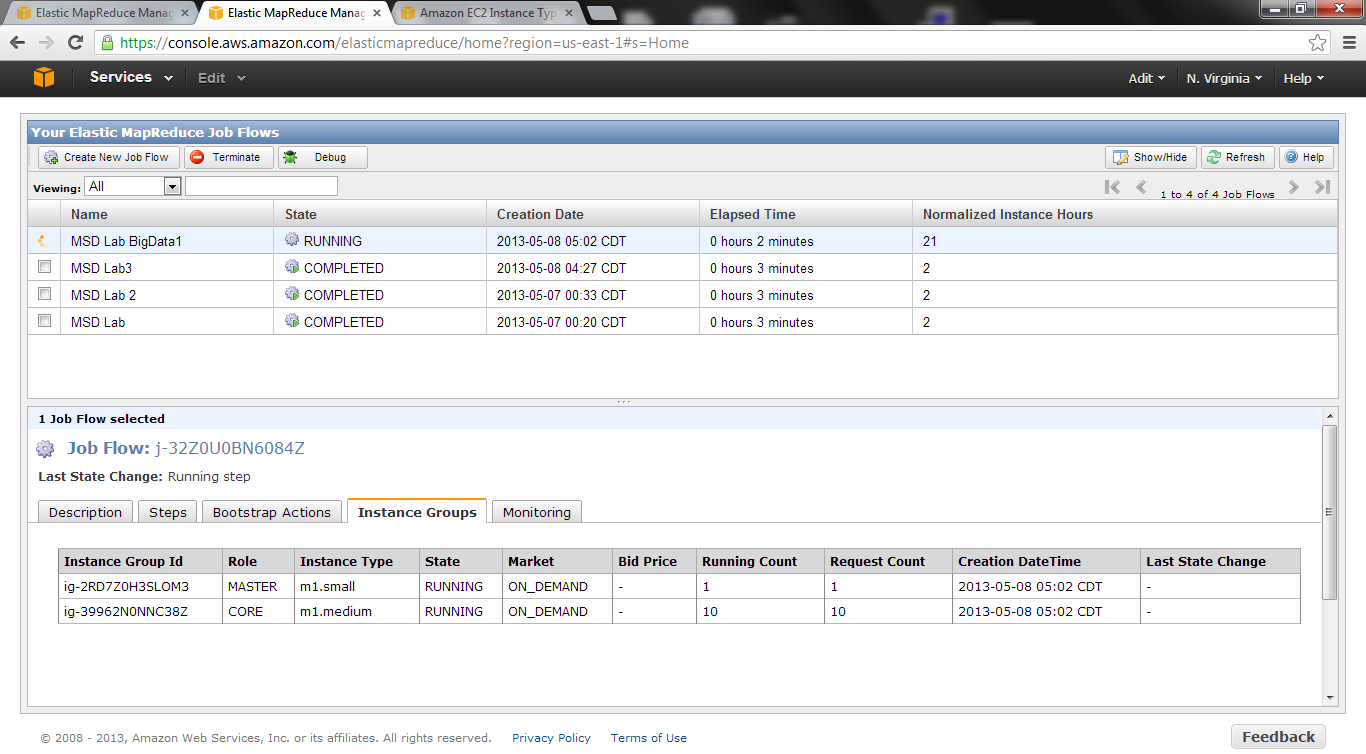
Medium:



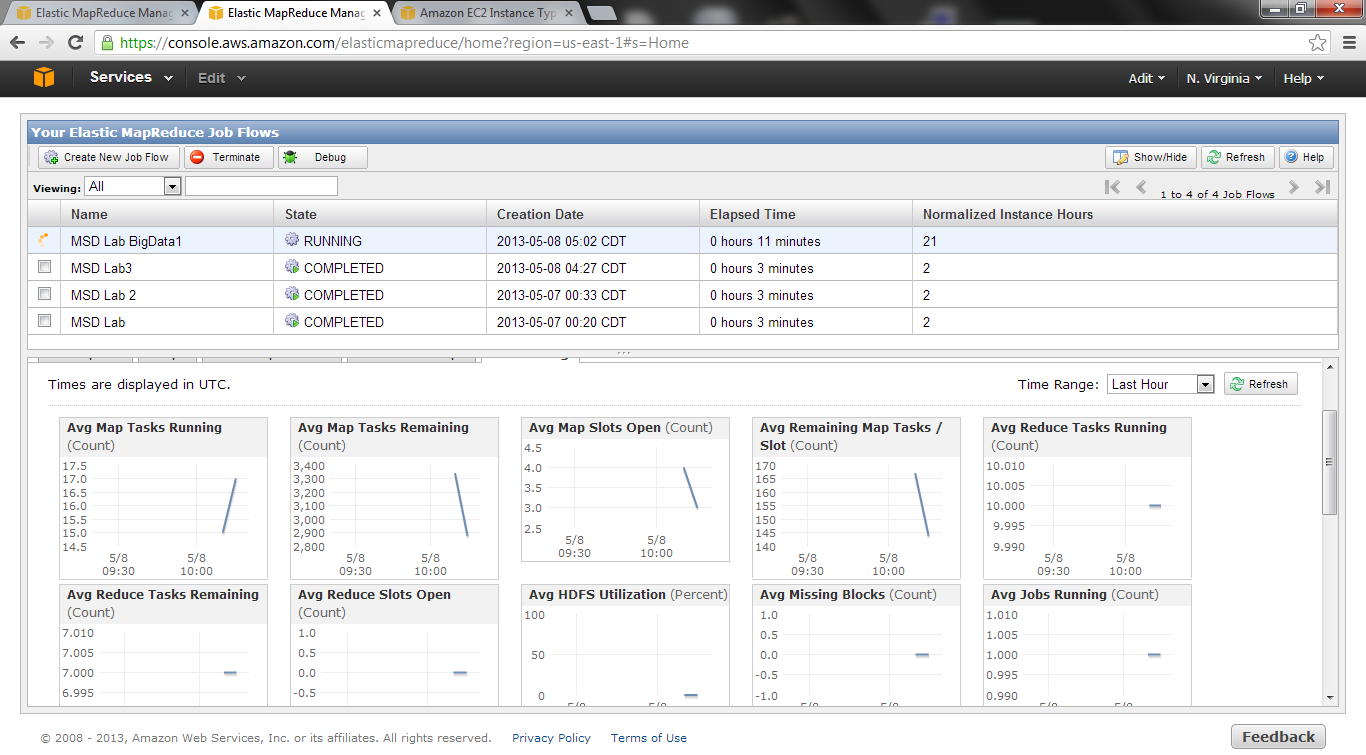




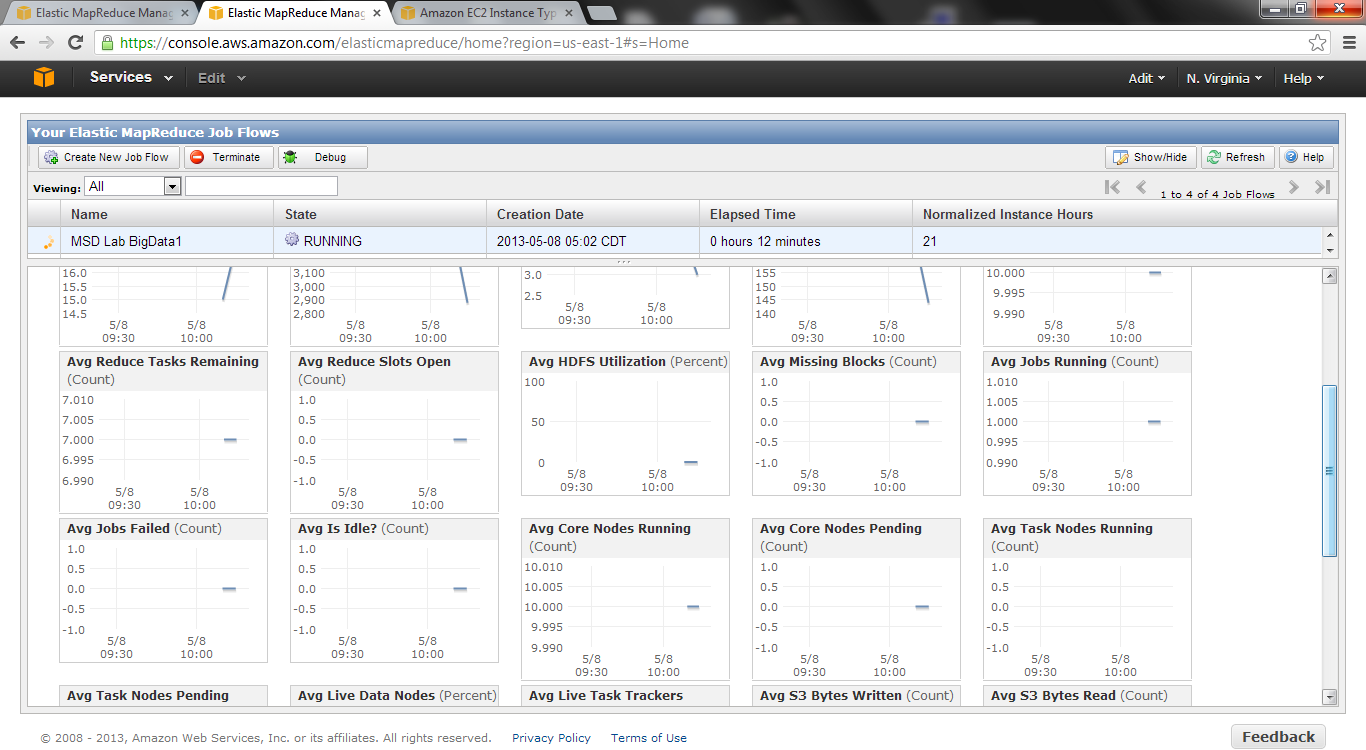
EMR Snapshot 1:



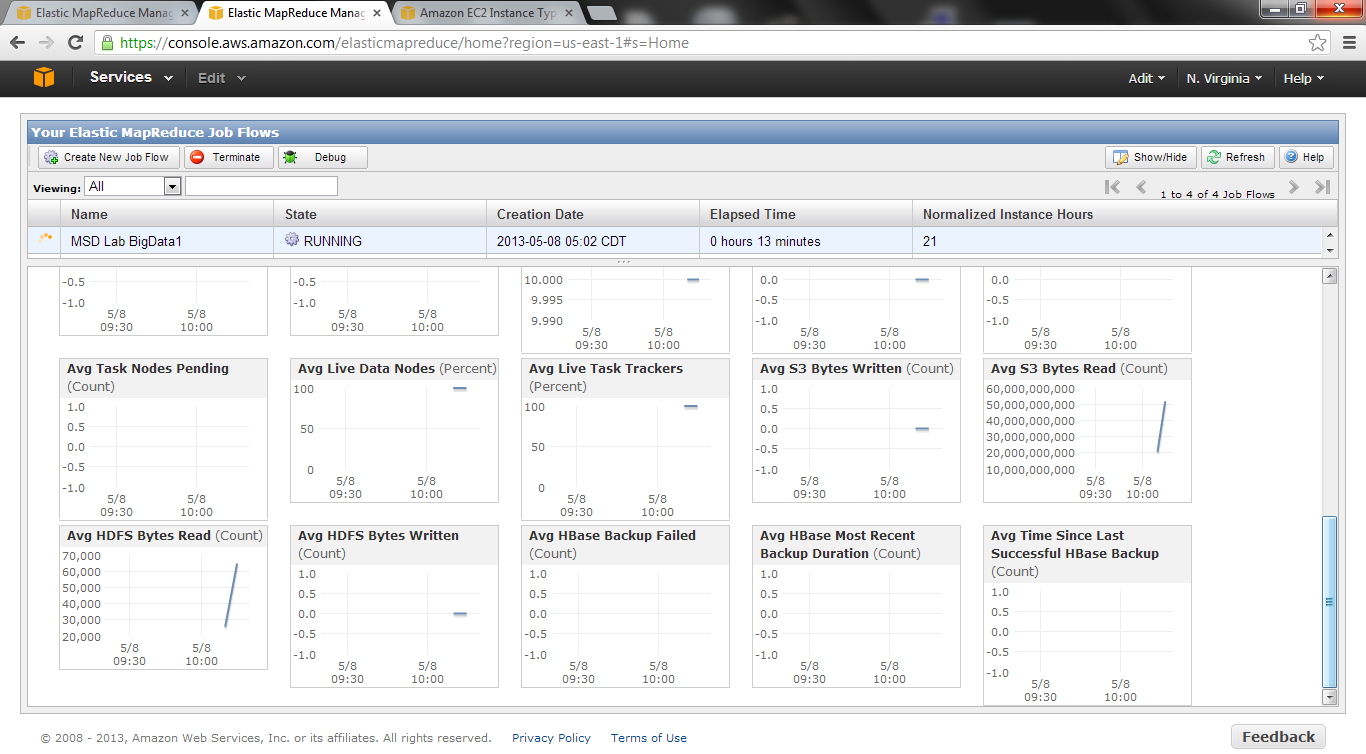
EMR Snapshot 2:



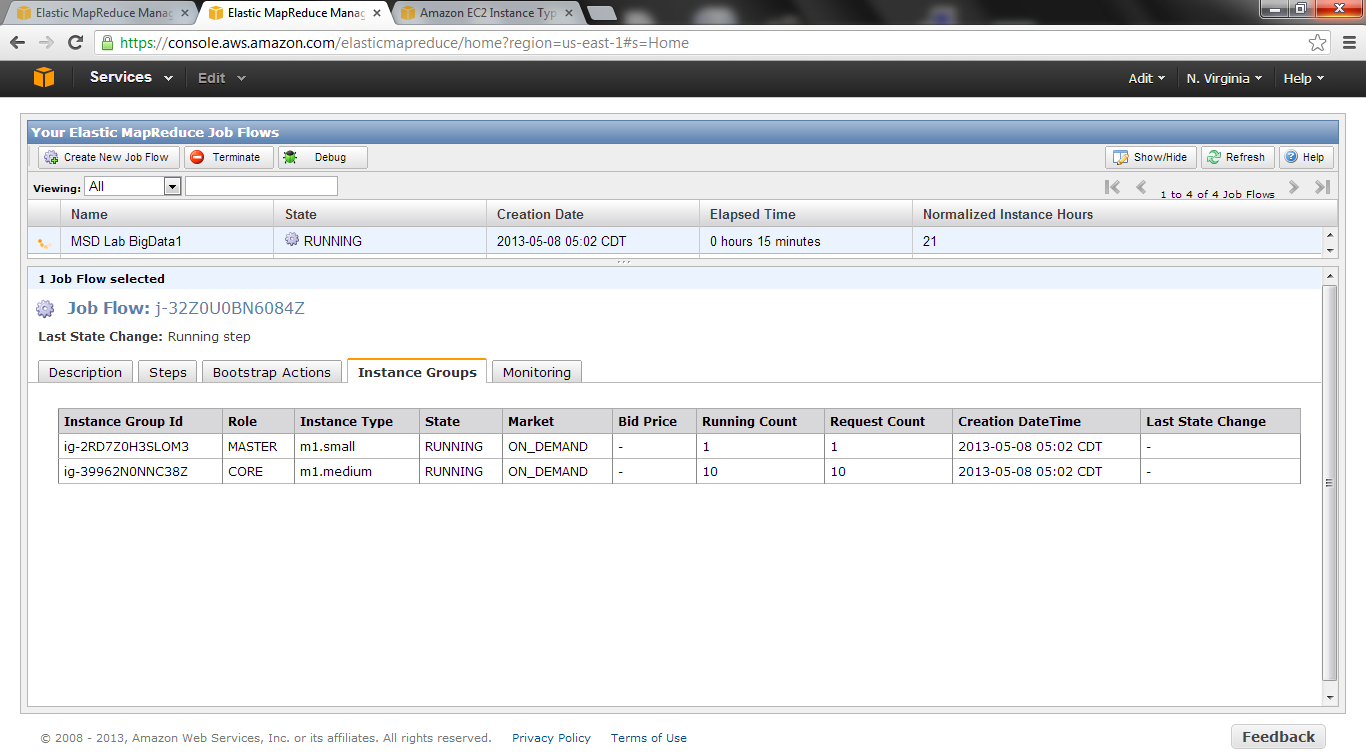
EMR Snapshot 3:



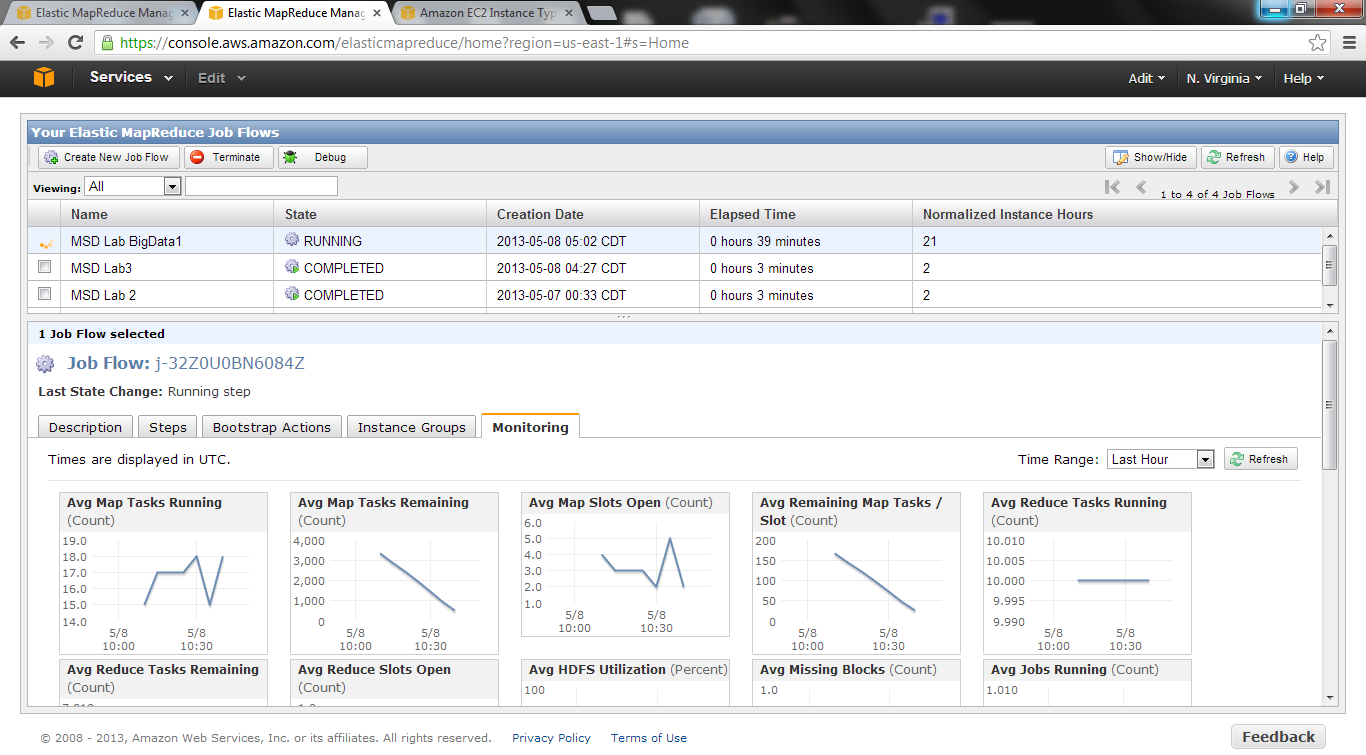
EMR Snapshot 4:



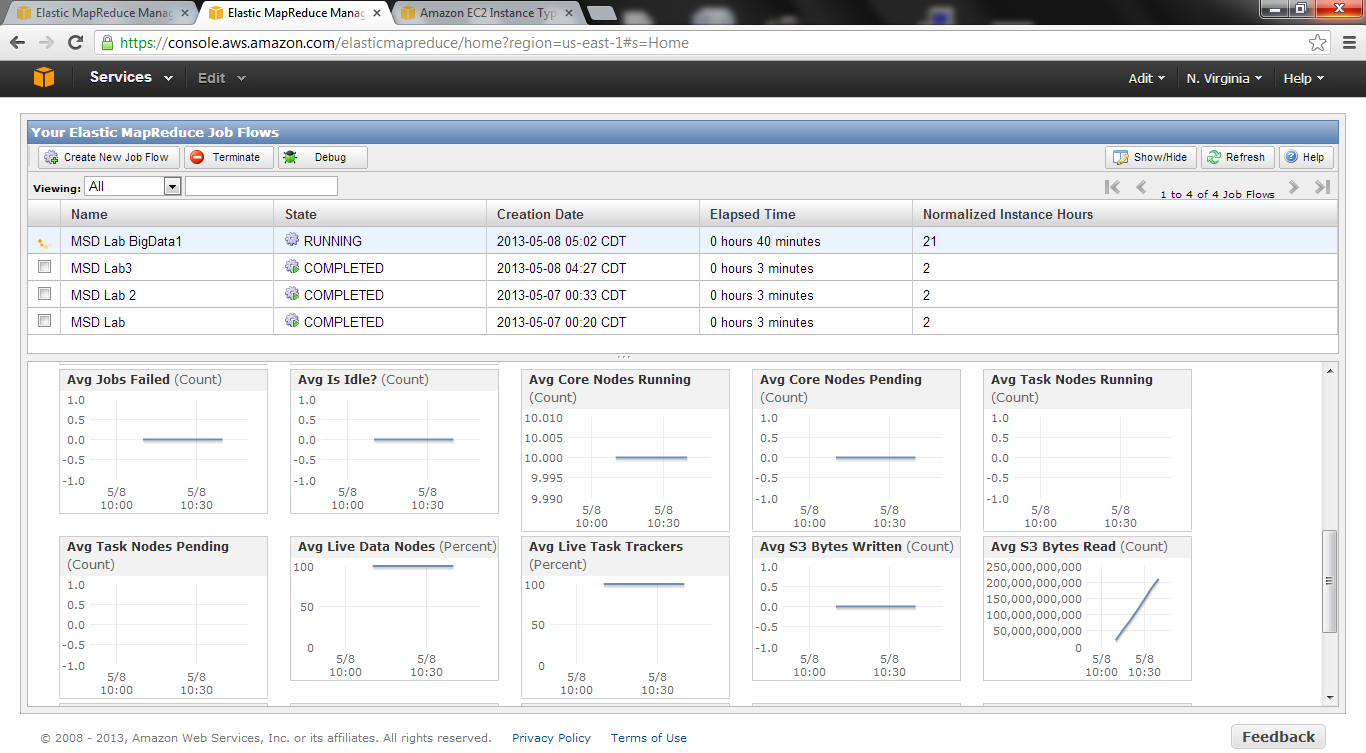
EMR Snapshot 5:



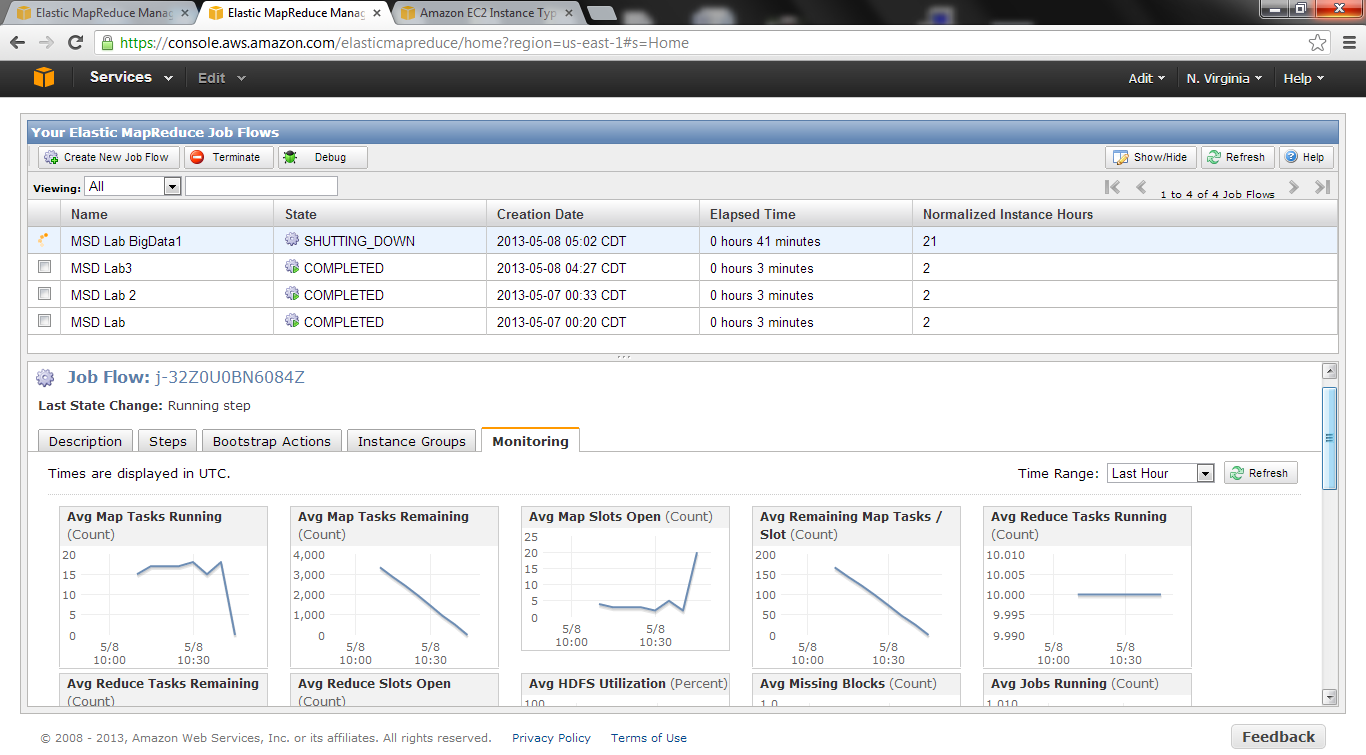
EMR Snapshot 6:



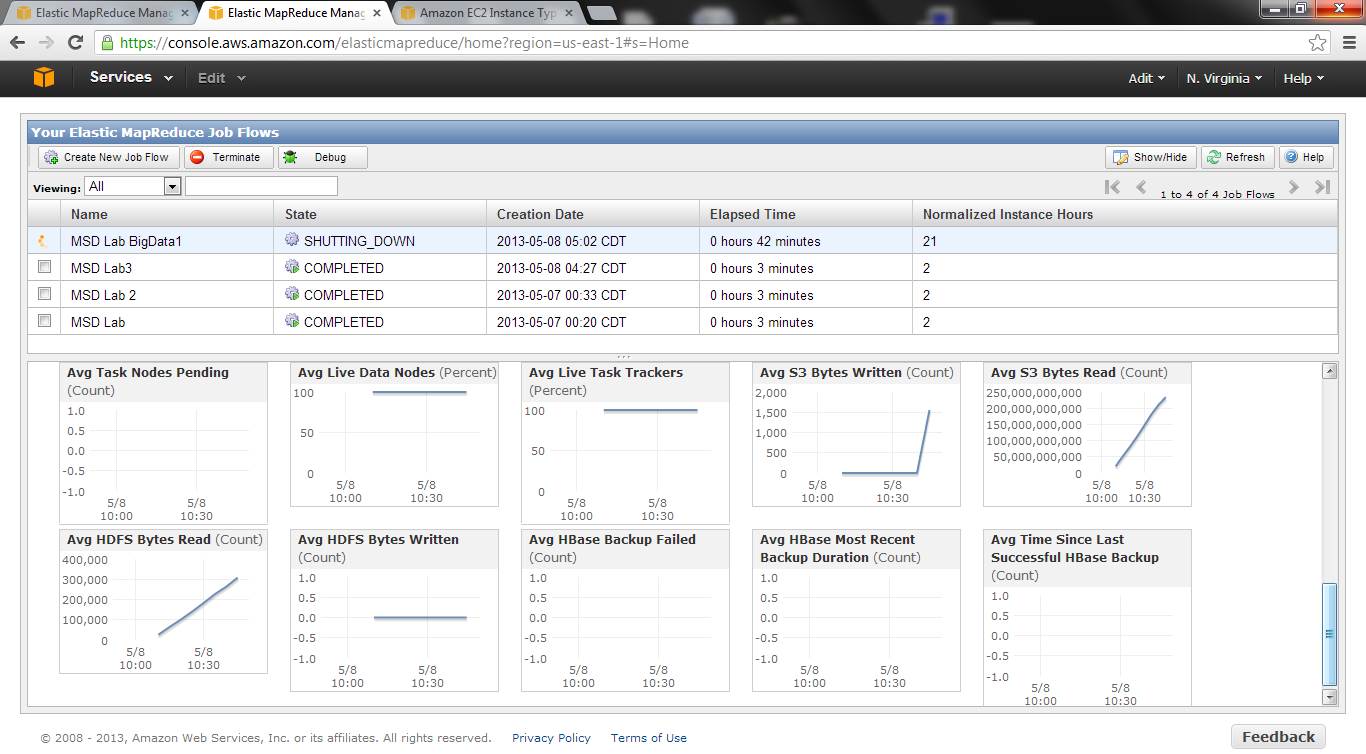
EMR Snapshot 7:



EMR Snapshot 8:



EMR Snapshot 9:



EMR Snapshot 10:

