* Spin up EMR cluster
* Include module in cluster
* Connect to head node
* Pull data from public bucket into Hadoop file system
* Analytics w/ spark
* Collect analytics to head node
* Write data to personal bucket
* Read data from S3 bucket to local machine