1 Spin up a new VM with Red hat Enterprise Linux 7 image

2 Download and Install Docker

<https://www.linux.com/learn/intro-to-linux/2017/11/how-install-and-use-docker-linux>

3 Create a docker container; I am using cloudera quickstart docker container

<https://medium.com/@SnazzyHam/how-to-get-up-and-running-with-clouderas-quickstart-docker-container-732c04ed0280>

Install Spark and all required APIs (PySpark, Pig, HBase, Hive)

4