Map Reduce

Deployment Guidelines

Prerequisites:

Install Hadoop related to your Operating system.

how to run:

create a text file 'names.txt', add data (space separated) and put it into HDFS.

hadoop fs -put names.txt /user/cloudera/names.txt

run the jar file

hadoop jar wordcount.jar /user/cloudera/names.txt /user/cloudera/wordcountoutput

check the output in HDFS

hadoop fs -cat /user/cloudera/wordcountoutput

Running in a docker

1. install docker
2. pull the image

    docker pull sequenceiq/hadoop-docker:2.7.0

3. Start a container

    docker run -it sequenceiq/hadoop-docker:2.7.0 /etc/bootstrap.sh -bash

4. Run MapReduce job.

    cd $HADOOP_PREFIX

       bin/hadoop dfs -put names.txt /user/root/input/

       bin/hadoop jar wordcount.jar /user/root/input/names.txt /user/root/output

       bin/hadoop dfs -cat /user/root/output/*