

Hadoop-Spark-Hive 环境

使用 dockerhub 上的镜像: <https://hub.docker.com/r/suhothayan/hadoop-spark-pig-hive>。

请按以下步骤获取:

- 安装 docker, 修改镜像源:
 - Windows : <https://www.runoob.com/docker/windows-docker-install.html>
 - MacOS : <https://www.runoob.com/docker/macos-docker-install.html>
 - Ubuntu : <https://www.runoob.com/docker/macos-docker-install.html>
- 获取镜像: `docker pull suhothayan/hadoop-spark-pig-hive:2.9.2`

Wordcount Demo

- 运行镜像: `docker run -it suhothayan/hadoop-spark-pig-hive:2.9.2 # -i`: 以交互式运行容器, 通常与 `-t` 同时使用; `-t`: 为容器重新分配一个伪输入终端, 通常与 `-i` 同时使用; `-v` 可以挂在一个本地的卷到镜像环境中 详细使用请自查
- 创建一个文件做为输入, 随便写入内容: `vim input.txt`, 写入内容并保存
- 将文件倒入 hdfs 指定路径下: `hadoop fs -put input.txt input/ #hdfs` 常用命令上网查
- 运行 wordcount 演示程序: `hadoop jar /usr/local/hadoop-2.9.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.9.2.jar wordcount /user/root/input/input.txt /user/root/output/out`
- 从 hdfs 获取结果到 linux 文件系统: `hadoop fs -getmerge /user/root/output/out result.txt #也可直接查看` 命令自查