一：hadoop window部署

1 软件下载

下载 <https://archive.apache.org/dist/hadoop/common/> 中有  hadoop 所有版本

链接：<https://pan.baidu.com/s/1hyBKUg4qXqN_wBWuiK3ysA>

http://192.168.3.78/tutorial/hadoop-%E7%AC%AC%E4%B8%80%E8%AF%BE%EF%BC%9Awin%E5%92%8Clinux%E5%8D%95%E6%9C%BA%E9%83%A8%E7%BD%B2%E8%AF%95%E7%94%A8/

hadoop-2.7.3.tar.gz

安装 jdk-8u11-windows-x64.exe

设置 JAVA\_HOME  和 CLASSPATH

JAVA\_HOME=C:\Program Files\Java\jdk1.8.0\_11

PATH=.....;%JAVA\_HOME%\bin;

CLASSPATH=%JAVA\_HOME%\lib/dt.jar;%JAVA\_HOME%\lib\tools.jar

解压  hadoop-2.7.3.tar.gz

解压 [winutils-master.zip](http://192.168.3.78/tutorial/hadoop-%e7%ac%ac%e4%b8%80%e8%af%be%ef%bc%9awin%e5%92%8clinux%e5%8d%95%e6%9c%ba%e9%83%a8%e7%bd%b2%e8%af%95%e7%94%a8/winutils-master.zip)

copy winutils/hadoop2.7/bin 里所有文件到 /hadoop\_home/bin 下

2 hadoop 配置

配置hadoop文件

配置文件示例下载

http://192.168.3.78/tutorial/hadoop-%E7%AC%AC%E4%B8%80%E8%AF%BE%EF%BC%9Awin%E5%92%8Clinux%E5%8D%95%E6%9C%BA%E9%83%A8%E7%BD%B2%E8%AF%95%E7%94%A8/hadoop-%E9%85%8D%E7%BD%AE/

hadoop-env.cmd

hdfs-site.xml

core-site.xml

mapred-site.xml

yarn-site.xml

hadoop-env.cmd 里set JAVA\_HOME=C:\Progra~1\Java\jdk1.8.0\_11

core-site.xml

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://172.16.0.10:8020</value>

</property>

</configuration>

hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:///hadoop-2.7.3/dfs/nn</value>

</property>

<property>

<name>dfs.namenode.checkpoint.dir</name>

<value>file:///hadoop-2.7.3/dfs/sn</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:///hadoop-2.7.3/dfs/dn</value>

</property>

<property>

<name>dfs.namenode.secondary.http-address</name><value>localhost:50090</value>

</property>

<property>

<name>dfs.webhdfs.enabled</name><value>true</value>

</property>

</configuration>

mapred-site.xml

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

yarn-site.xml

<configuration>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name><value>172.16.0.10:8031</value>

</property>

<property>

<name>yarn.resourcemanager.scheduler.address</name><value>172.16.0.10:8030</value>

</property>

<property>

<name>yarn.resourcemanager.admin.address</name><value>172.16.0.10:8033</value>

</property>

<property>

<name>yarn.resourcemanager.webapp.address</name><value>172.16.0.10:8088</value></property>

<property>

<name>yarn.resourcemanager.address</name><value>172.16.0.10:8032</value>

</property>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

</configuration>

路径问题

[file:///dfs/nn](file:///\\dfs\nn)

反应的是当下的磁盘 如果命令执行在 D盘，指的是 D:/dfs/nn

3 hfs格式化

始化，格式化hdfs磁盘

hadoop namenode -format

4 启动hadoop 和试用

hadoop\_home/sbin/start-dfs.cmd 启动 hdfs

hadoop\_home/sbin/start-yarn.cmd 启动 yarn

hadoop\_home/sbin/start-all.cmd 启动 yarn 和 dfs

启动问题

yarn.cmd里有编码问题

最终解决，使用Notepad++, 如下操作 Edit -> EOL Conversion -> Windows Format，保存并在hadoop的sbin目录下运行start-all.cmd，可以看到resourcemanager成功运行。

将\r 替换成 \r\n

在sublime text3 下，对 yarn文件  ，alt+r 替换\n  为 \r\n

试用

192.168.3.172:50070/ hdfs web界面

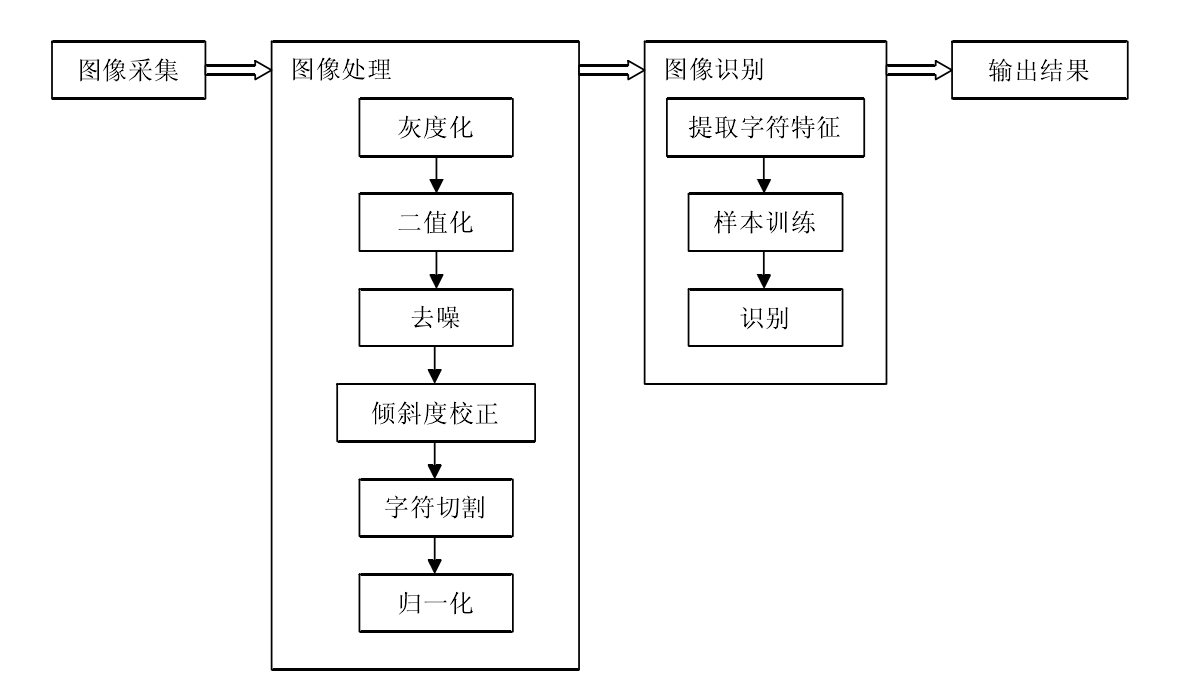
192.168.3.172:8088 yarn web 界面

hadoop  fs -ls /

hadoop fs -put file1 /usr

hadoop jar D:\tutorial\ap1\hadoop-2.7.3\share\hadoop\mapreduce\hadoop-mapreduce-examples-2.7.3.jar pi 10 10

二、python 验证码识别

过程

地址 http://192.168.3.78/tutorial/

代码 及样本