Windows10下hadoop单机环境的安装、配置和部署

1、下载hadoop

https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-2.7.7/hadoop-2.7.7.tar.gz

2、解压hadoop至本地，设置环境变量

D:\Hadoop\hadoop-2.7.7

系统变量：变量名：HADOOP\_HOME，变量值：D:\Hadoop\hadoop-2.7.7

Path：%HADOOP\_HOME%\bin

至此hadoop已经成功安装上了，可以在cmd中输入hadoop -version查看版本

3、在D:\Hadoop\hadoop-2.7.7下新增文件夹Workspace，进入Workspace，新增data、name、tmp这3个子文件夹

4、进入D:\Hadoop\hadoop-2.7.7\etc\hadoop，修改hadoop的4个主要配置文件

（1）修改core-site.xml

<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/D:/Hadoop/hadoop-2.7.7/workspace/tmp</value>

<description>namenode上本地的hadoop临时文件夹</description>

</property>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

<description>HDFS的URI，文件系统://namenode标识:端口号</description>

</property>

</configuration>

（2）修改hdfs-site.xml

<configuration>

<!-- 这个参数设置为1，因为是单机版hadoop -->

<property>

<name>dfs.replication</name>

<value>1</value>

<description>副本个数，配置默认是3,应小于datanode机器数量</description>

</property>

<property>

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

<value>/D:/Hadoop/hadoop-2.7.7/workspace/data</value>

<description>datanode上数据块的物理存储位置</description>

</property>

<property>

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

<value>/D:/Hadoop/hadoop-2.7.7/workspace/name</value>

<description>namenode上存储hdfs名字空间元数据</description>

</property>

</configuration>

（3）修改mapred-site.xml（如果不存在就先copy mapred-site.xml.template，再修改文件名为mapred-site.xml）

<configuration>

<property>

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

<value>yarn</value>

</property>

<property>

<name>mapred.job.tracker</name>

<value>hdfs://localhost:9001</value>

</property>

</configuration>

（4）修改yarn-site.xml

<configuration>

<!-- Site specific YARN configuration properties -->

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

</configuration>

5、配置jdk路径

打开hadoop-env.cmd，修改jdk安装路径，如果环境变量配置有%JAVA\_HOME%，则不需要修改

@echo off

@rem Licensed to the Apache Software Foundation (ASF) under one or more

@rem contributor license agreements. See the NOTICE file distributed with

@rem this work for additional information regarding copyright ownership.

@rem The ASF licenses this file to You under the Apache License, Version 2.0

@rem (the "License"); you may not use this file except in compliance with

@rem the License. You may obtain a copy of the License at

@rem

@rem http://www.apache.org/licenses/LICENSE-2.0

@rem

@rem Unless required by applicable law or agreed to in writing, software

@rem distributed under the License is distributed on an "AS IS" BASIS,

@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

@rem See the License for the specific language governing permissions and

@rem limitations under the License.

@rem Set Hadoop-specific environment variables here.

@rem The only required environment variable is JAVA\_HOME. All others are

@rem optional. When running a distributed configuration it is best to

@rem set JAVA\_HOME in this file, so that it is correctly defined on

@rem remote nodes.

@rem The java implementation to use. Required.

set JAVA\_HOME=%JAVA\_HOME%

@rem The jsvc implementation to use. Jsvc is required to run secure datanodes.

@rem set JSVC\_HOME=%JSVC\_HOME%

@rem set HADOOP\_CONF\_DIR=

@rem Extra Java CLASSPATH elements. Automatically insert capacity-scheduler.

if exist %HADOOP\_HOME%\contrib\capacity-scheduler (

if not defined HADOOP\_CLASSPATH (

set HADOOP\_CLASSPATH=%HADOOP\_HOME%\contrib\capacity-scheduler\\*.jar

) else (

set HADOOP\_CLASSPATH=%HADOOP\_CLASSPATH%;%HADOOP\_HOME%\contrib\capacity-scheduler\\*.jar

)

)

@rem The maximum amount of heap to use, in MB. Default is 1000.

@rem set HADOOP\_HEAPSIZE=

@rem set HADOOP\_NAMENODE\_INIT\_HEAPSIZE=""

@rem Extra Java runtime options. Empty by default.

@rem set HADOOP\_OPTS=%HADOOP\_OPTS% -Djava.net.preferIPv4Stack=true

@rem Command specific options appended to HADOOP\_OPTS when specified

if not defined HADOOP\_SECURITY\_LOGGER (

set HADOOP\_SECURITY\_LOGGER=INFO,RFAS

)

if not defined HDFS\_AUDIT\_LOGGER (

set HDFS\_AUDIT\_LOGGER=INFO,NullAppender

)

set HADOOP\_NAMENODE\_OPTS=-Dhadoop.security.logger=%HADOOP\_SECURITY\_LOGGER% -Dhdfs.audit.logger=%HDFS\_AUDIT\_LOGGER% %HADOOP\_NAMENODE\_OPTS%

set HADOOP\_DATANODE\_OPTS=-Dhadoop.security.logger=ERROR,RFAS %HADOOP\_DATANODE\_OPTS%

set HADOOP\_SECONDARYNAMENODE\_OPTS=-Dhadoop.security.logger=%HADOOP\_SECURITY\_LOGGER% -Dhdfs.audit.logger=%HDFS\_AUDIT\_LOGGER% %HADOOP\_SECONDARYNAMENODE\_OPTS%

@rem The following applies to multiple commands (fs, dfs, fsck, distcp etc)

set HADOOP\_CLIENT\_OPTS=-Xmx512m %HADOOP\_CLIENT\_OPTS%

@rem set HADOOP\_JAVA\_PLATFORM\_OPTS="-XX:-UsePerfData %HADOOP\_JAVA\_PLATFORM\_OPTS%"

@rem On secure datanodes, user to run the datanode as after dropping privileges

set HADOOP\_SECURE\_DN\_USER=%HADOOP\_SECURE\_DN\_USER%

@rem Where log files are stored. %HADOOP\_HOME%/logs by default.

@rem set HADOOP\_LOG\_DIR=%HADOOP\_LOG\_DIR%\%USERNAME%

@rem Where log files are stored in the secure data environment.

set HADOOP\_SECURE\_DN\_LOG\_DIR=%HADOOP\_LOG\_DIR%\%HADOOP\_HDFS\_USER%

@rem The directory where pid files are stored. /tmp by default.

@rem NOTE: this should be set to a directory that can only be written to by

@rem the user that will run the hadoop daemons. Otherwise there is the

@rem potential for a symlink attack.

set HADOOP\_PID\_DIR=%HADOOP\_PID\_DIR%

set HADOOP\_SECURE\_DN\_PID\_DIR=%HADOOP\_PID\_DIR%

@rem A string representing this instance of hadoop. %USERNAME% by default.

set HADOOP\_IDENT\_STRING=%USERNAME%

6、格式化HDFS文件系统

进入D:\Hadoop\hadoop-2.7.7\bin，执行hdfs namenode -format命令，会报错

C:\Windows\system32>hdfs namenode -format

21/03/19 20:14:17 ERROR util.Shell: Failed to locate the winutils binary in the hadoop binary path

java.io.IOException: Could not locate executable D:\Hadoop\hadoop-2.7.7\bin\winutils.exe in the Hadoop binaries.

at org.apache.hadoop.util.Shell.getQualifiedBinPath(Shell.java:382)

at org.apache.hadoop.util.Shell.getWinUtilsPath(Shell.java:397)

at org.apache.hadoop.util.Shell.<clinit>(Shell.java:390)

at org.apache.hadoop.util.StringUtils.<clinit>(StringUtils.java:80)

at org.apache.hadoop.hdfs.server.common.HdfsServerConstants$RollingUpgradeStartupOption.getAllOptionString(HdfsServerConstants.java:80)

at org.apache.hadoop.hdfs.server.namenode.NameNode.<clinit>(NameNode.java:250)

21/03/19 20:14:17 INFO namenode.NameNode: STARTUP\_MSG:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

STARTUP\_MSG: Starting NameNode

STARTUP\_MSG: host = LAPTOP-JAQME9KQ/192.168.56.1

STARTUP\_MSG: args = [-format]

STARTUP\_MSG: version = 2.7.7

STARTUP\_MSG: classpath = D:\Hadoop\hadoop-2.7.7\etc\hadoop;D:\Hadoop\hadoop-2.7.7\share\hadoop\common\lib\activation-1.1.jar;D:\Hadoop\hadoop-2.7.7\share\hadoop\common\lib\apacheds-i18n-2.0.0-M15.jar;D:\Hadoop\hadoop-2.7.7\share\hadoop\common\lib\apacheds-kerberos-codec-2.0.0-M15.jar;D:\Hadoop\hadoop-2.7.7\share\hadoop\common\lib\api-asn1-api-

意思是在hadoop/bin目录下缺少了winutils.exe和hadoop.dll，winutils.exe是在Windows系统上需要的hadoop调试环境工具，里面包含一些在Windows系统下调试hadoop、spark所需要的基本的工具类，另外在使用eclipse调试hadoop程序时，也需要winutils.exe，需要配置上面的环境变量。下载winutils，注意需要与hadoop的版本相对应：

https://github.com/4ttty/winutils

https://codechina.csdn.net/mirrors/4ttty/winutils?utm\_source=csdn\_github\_accelerator

解压，将winutils.exe、hadoop.dll放入D:\Hadoop\hadoop-2.7.7\bin

（我的hadoop是2.7.7版本，github上也没有对应的winutils.exe，直接使用2.8.1版本的winutils.exe和hadoop.dll）

7、进入D:\Hadoop\hadoop-2.7.7\bin，再次执行hdfs格式化：hdfs namenode -format命令，执行成功

不用管报错：SHUTDOWN\_MSG: Shutting down NameNode at LAPTOP-JAQME9KQ/192.168.56.1

8、进入D:\Hadoop\hadoop-2.7.7\sbin

执行start-all.cmd，会弹出4个窗口，分表代表name，data，yarn，resourcemanager四个进程，也可以使用jps命令查看是否开启这4个进程

说明全部启动成功，网页访问http://localhost:50070，即可看到hdfs文件系统，网页访问http://localhost:8088，即可进入管理界面

可以使用stop-all.cmd停止运行

C:\Windows\system32>d:

D:\>cd D:\Hadoop\hadoop-2.7.7\sbin

D:\Hadoop\hadoop-2.7.7\sbin>cd D:\Hadoop\hadoop-2.7.7\sbin

D:\Hadoop\hadoop-2.7.7\sbin>start-all.cmd

This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd

starting yarn daemons

D:\Hadoop\hadoop-2.7.7\sbin>jps

24980 ResourceManager

20024 DataNode

36856 NodeManager

37784

38172 Jps

8460 NameNode