1. 安装jdk

1、创建目录jdk mkdir

tar -zxvf jdk-8u121-linux-x64.tar.gz

查看下bin目录

设置环境变量

sudo vi /etc/profile

添加 ：

export JAVA\_HOME=/home/hadoop/jdk/jdk1.8.0\_121/

export PATH=$PATH:$JAVA\_HOME/bin

1. 安装hadoop

1、解压安装hadoop

tar -zxvf hadoop-2.10.0.tar.gz

2、hadoop目录结构简介

bin：一些可执行脚本

etc：hadoop系统配置文件

include：本地包

lib：本地包

sbin：hadoop启动的一些脚本

share：java编程包

3、更改etc/hadoop下的配置文件

hadoop-env.sh：添加java环境变量

core-site.xml

hdfs-site.xml

mapred-site.xml.template

yarn-site.xml

Hadoop启动时要读取(读取一些参数)上述四个文件内容

vi hadoop-env.sh

修改export JAVA\_HOME=/home/hadoop/jdk/jdk1.8.0\_121/

vi core-site.xml

<property>

<name>fs.defaultFS</name>

<value>hdfs://user1:9000/</value>

</property> Hadoop本地机器使用的文件系统

<property>

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

<value>/home/hadoop/data/</value>

</property> Hadoop文件数据存储的目录，一般namenode和datanode分开，这里配在同一目录下

vi hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>1</value>

</property> hadoop文件块的份数

mv mapred-site.xml.template mapred-site.xml 更改mapreduce配置文件的名字

vi mapred-site.xml

<property>

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

<value>yarn</value>

</property> Hadoop mapreduce程序yarn会把程序分配到多个节点中运行，否则只在本地机器运行

vi yarn-site.xml

<property>

<name>yarn.resourcemanager.hostname</name>

<value>user1</value>

</property> Hadoop yarn的主节点地址

<property>

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

<value>mapreduce\_shuffle</value>

</property> reduce程序获取数据的方式

1. hadoop文件系统格式化（使用新硬盘需要格式化文件系统先）

export JAVA\_HOME=/home/hadoop/jdk/jdk1.8.0\_121/

export HADOOP\_HOME=/home/hadoop/hadoop/hadoop-2.10.0

export PATH=$PATH:$JAVA\_HOME/bin:$HADOOP\_HOME/bin

source /etc/profile 使环境变量立即生效

hadoop namenode -format(先查看data目录下)

1. 启动hadoop

查看sbin目录下：

export JAVA\_HOME=/home/hadoop/jdk/jdk1.8.0\_121/

export HADOOP\_HOME=/home/hadoop/hadoop/hadoop-2.10.0

export PATH=$PATH:$JAVA\_HOME/bin:$HADOOP\_HOME/bin: $HADOOP\_HOME/sbin

source /etc/profile

start-dfs.sh 启动dfs

启动了几个进程，datanode如何找到，secondnamenode是什么，0.0.0.0是什么，每次登录都需要设置密码。

jps查看启动的java进程