**基于Docker部署hadoop集群：**

初次学习，没有自己去制作镜像，选择阿里云上现成的hadoop镜像。

1.下载镜像，运行容器

docker pull registry.cn-hangzhou.aliyuncs.com/kaibb/hadoop

下载完成后，运行三个容器,Master,Slave1和Slave2

docker run -i -t --name **Master** -h Master registry.cnhangzhou.aliyuncs.com/kaibb/hadoop /bin/bash

docker run -i -t --name **Slave1** -h Slave1 registry.cn-hangzhou.aliyuncs.com/kaibb/hadoop /bin/bash

docker run -i -t --name **Slave2** -h Slave1 registry.cn-hangzhou.aliyuncs.com/kaibb/hadoop /bin/bash

（--name指代容器名，-h指代主机名）

1. 配置无秘ssh:



大体操作是,生成秘钥，获取秘钥，在各个节点中保存秘钥。

在三个节点上重复此操作，将生成的秘钥存入各自的authorized\_keys文件中(路径~/.ssh)。

最后三个节点的authorized\_keys文件均如图：



1. 添加各节点ip：

使用 **ip addr**命令获取各节点ip



修改 **/etc/hosts** , 将三个节点的ip , 主机名添加进去。



之后测试直接使用 **ssh+主机名**(如ssh Slave1)，顺利登入则证明成功（第一次会让输入yes或no，输入yes即可).

1. 配置hadoop

在Master节点上,

所有配置文件的路径在 **/opt/tools/hadoop-2.7.2/etc/hadoop**



我们需要配置的是**hadoop-env.sh , core-site.xml , hdfs-site.xml , mapred-site.xml , yarn-site.xml**及文件 **slaves**

**hadoop-env.sh**:

配置jdk路径：找到对应位置补充jdk路径



**Core-site.xml**:



指定**namenode**的地址hdfs://主机名:9000 端口为9000

hadoop.tmp.dir 在value路径下创建tmp文件夹

**hdfs-site.xml**:



dfs.replication指备份的数量

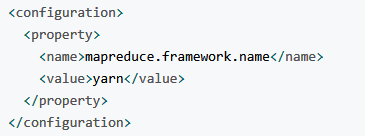
dfs.datanode.data.dir 在对应节点对应value下创建data文件夹

dfs.namenode.name.dir在对应节点对应value下创建name文件夹

**yarn-site.xml**: (不懂意思)



**mapred-site.xml**:



slaves：（指定作为slave节点的主机，将主机名添加进文件）



至此便完成了hadoop的配置

接着将这些文件通过 scp发送到各个Slave节点上覆盖原来的文件：

**scp core-site.xml hadoop-env.sh hdfs-site.xml mapred-site.xml yarn-site.xml Slave1:/opt/tools/hadoop/etc/hadoop/**

**scp core-site.xml hadoop-env.sh hdfs-site.xml mapred-site.xml yarn-site.xml Slave2:/opt/tools/hadoop/etc/hadoop/**

1. **运行hadoop:**

进行格式化操作：hadoop namenode -format （格式化时会自动创建上面配置文件里所写的目录）

(**！: 多次进行格式化命令操作会使得datanode无法打开，每次格式化之前把上面配置文件里提到的tmp,name.data文件夹均删除**）

到启动脚本路径下启动hadoop

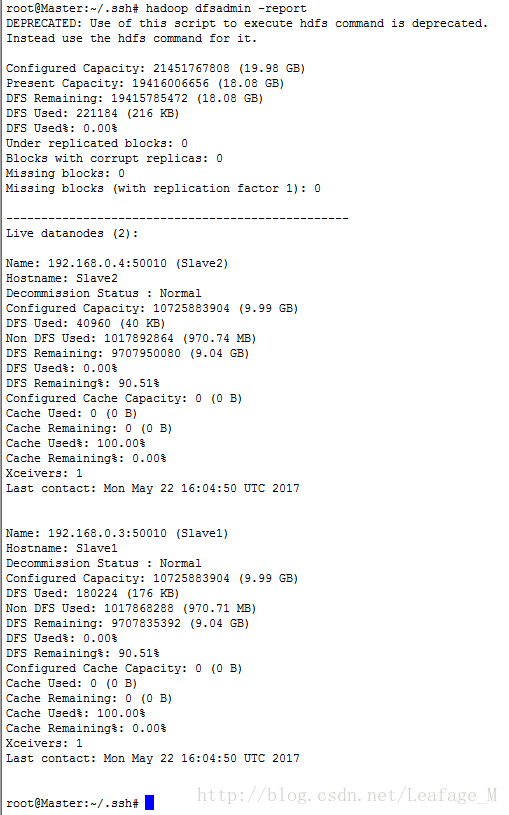


在各个节点上通过**jps**命令查看相关进程是否已经启动





通过hadoop dfsadmin -report 可以查看各节点信息



1. 运行world count实例：

在HDFS上创建一个文件夹：**hadoop fs -mkdir /input**

查看该文件夹： **hadoop fs -ls /**

上传文件到HDFS上，这里直接将README.txt来进行测试：

**hadoop fs -put README.txt /input/**

****

运行word count实例(mapreduce任务实际是一个jar包，直接运行即可）：

在**/opt/tools/hadoop/share/hadoop/mapreduce**目录中执行：

**hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordcount /input /output**

****

最后在**output目录**下会有文件 **part-r-00000** 查看该文件即可获取实例结果：

**hadoop fs -cat /output/part-r-00000**

****

1. 检测集群状况**：**

若是本地搭建可通过Master节点的50070端口监测集群节点运行状态，由于是在容器里运行，没有图形界面，我们可以采用端口映射的方案：

****

之后就可以通过本地浏览器访问50070端口查看集群情况。