

一、实验运行步骤与截图

1. 单机 hadoop 系统安装 (Linux Ubuntu16.04)

步骤:

- 1) 安装 ssh
- 2) 安装 java8
- 3) 下载 hadoop 安装包, 解压安装 Hadoop3.3.1 (我安装在 /home/jiayaoyao/Hadoop/Hadoop-3.3.1 中)
- 4) 在/etc/profile 中配置环境变量
- 5) 运行 java -version 以及 hadoop version 检查是否成功配置
- 6) 运行 grep 程序测试

截图:

运行过程:

```
jiayaoyao@jiayaoyao-virtual-machine:~/hadoop/hadoop-3.3.1$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar
grep input output 'dfs[a-z.]+'
2021-10-07 01:51:38,409 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2021-10-07 01:51:38,678 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2021-10-07 01:51:38,678 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2021-10-07 01:51:39,209 INFO input.FileInputFormat: Total input files to process : 10
2021-10-07 01:51:39,255 INFO mapreduce.JobSubmitter: number of splits:10
2021-10-07 01:51:39,735 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local834201022_0001
2021-10-07 01:51:39,735 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-07 01:51:39,951 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2021-10-07 01:51:39,951 INFO mapreduce.Job: Running job: job_local834201022_0001
2021-10-07 01:51:39,974 INFO mapred.LocalJobRunner: OutputCommitter set in Config null
2021-10-07 01:51:39,988 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-10-07 01:51:39,988 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2021-10-07 01:51:39,989 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2021-10-07 01:51:40,037 INFO mapred.LocalJobRunner: Waiting for map tasks
2021-10-07 01:51:40,038 INFO mapred.LocalJobRunner: Starting task: attempt_local834201022_0001_m_000000_0
2021-10-07 01:51:40,068 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-10-07 01:51:40,068 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2021-10-07 01:51:40,133 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2021-10-07 01:51:40,142 INFO mapred.MapTask: Processing split: file:/home/jiayaoyao/hadoop/hadoop-3.3.1/input/hadoop-policy.xml:0+11765
2021-10-07 01:51:40,295 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2021-10-07 01:51:40,295 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2021-10-07 01:51:40,295 INFO mapred.MapTask: soft limit at 83886080
2021-10-07 01:51:40,295 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2021-10-07 01:51:40,295 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2021-10-07 01:51:40,303 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2021-10-07 01:51:40,333 INFO mapred.LocalJobRunner:
2021-10-07 01:51:40,333 INFO mapred.MapTask: Starting flush of map output
2021-10-07 01:51:40,333 INFO mapred.MapTask: Spilling map output
2021-10-07 01:51:40,333 INFO mapred.MapTask: bufstart = 0; bufend = 17; bufvoid = 104857600
2021-10-07 01:51:40,333 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214396(104857584); length = 1/6553600
2021-10-07 01:51:40,346 INFO mapred.MapTask: Finished spill 0
2021-10-07 01:51:40,362 INFO mapred.Task: Task:attempt_local834201022_0001_m_000000_0 is done. And is in the process of committing
2021-10-07 01:51:40,371 INFO mapred.LocalJobRunner: map
2021-10-07 01:51:40,371 INFO mapred.Task: Task 'attempt_local834201022_0001_m_000000_0' done.
2021-10-07 01:51:40,402 INFO mapred.Task: Final Counters for attempt_local834201022_0001_m_000000_0: Counters: 18
File System Counters
FILE: Number of bytes read=294239
```

运行结果:

```
jiayaoyao@jiayaoyao-virtual-machine:~/hadoop/hadoop-3.3.1$ cat output/*
1
dfsadmin
jiayaoyao@jiayaoyao-virtual-machine:~/hadoop/hadoop-3.3.1$
```

可以正确输出, 表明 hadoop 单机系统搭建成功。

2. 单机伪分布式 Hadoop 系统安装和运行

2.1 不含 yarn 的单机伪分布

步骤:

- 1) 配置 etc/hadoop/core-site.xml 与 etc/hadoop/hdfs-site.xml
- 2) 免密码 SSH 访问配置

3) 格式化 NameNode

4) 启动 namenode 并使用 jps 查看集群状态

```

root@jiayaoyao-virtual-machine: /home/jiayaoyao/hadoop/hadoop-3.3.1
0 of size 399 bytes saved in 0 seconds .
2021-10-10 20:15:46,538 INFO namenode.NNStorageRetentionManager: Going to retain
1 images with txid >= 0
2021-10-10 20:15:46,557 INFO namenode.FSNamesystem: Stopping services started fo
r active state
2021-10-10 20:15:46,558 INFO namenode.FSNamesystem: Stopping services started fo
r standby state
2021-10-10 20:15:46,564 INFO namenode.FSImage: FSImageSaver clean checkpoint: tx
id=0 when meet shutdown.
2021-10-10 20:15:46,565 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at jiayaoyao-virtual-machine/127.0.1.1
*****/
root@jiayaoyao-virtual-machine: /home/jiayaoyao/hadoop/hadoop-3.3.1# sbin/start-d
fs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [jiayaoyao-virtual-machine]
root@jiayaoyao-virtual-machine: /home/jiayaoyao/hadoop/hadoop-3.3.1# jps
61670 SecondaryNameNode
61816 Jps
61416 DataNode
61247 NameNode
root@jiayaoyao-virtual-machine: /home/jiayaoyao/hadoop/hadoop-3.3.1#

```

5) 浏览 NameNode 的 Web 接口

The screenshot shows the Hadoop NameNode Web UI in a browser. The address bar shows 'localhost:9870/dfshealth.html#tab-overview'. The page has a green navigation bar with tabs: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The 'Overview' tab is selected, showing 'localhost:9000' (✓active).

Started:	Sun Oct 10 20:16:21 +0800 2021
Version:	3.3.1, ra3b9c37a397ad4188041dd80621bdeefc46885f2
Compiled:	Tue Jun 15 13:13:00 +0800 2021 by ubuntu from (HEAD detached at release-3.3.1-RC3)
Cluster ID:	CID-222f6c79-1dbc-48eb-b1ad-901ffc91b6
Block Pool ID:	BP-1763845669-127.0.1.1-1633868146078

Summary

Security is off.
 Safemode is off.
 1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
 Heap Memory used 106 MB of 242 MB Heap Memory. Max Heap Memory is 873 MB.
 Non Heap Memory used 49.11 MB of 51.09 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	18.62 GB
----------------------	----------

Configured Capacity:	18.62 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	9.69 GB
DFS Remaining:	7.96 GB (42.76%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Sun Oct 10 20:16:21 +0800 2021
Last Checkpoint Time	Sun Oct 10 20:15:46 +0800 2021
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 3

Journal Manager	State
FileJournalManager(root=/home/jiayao/haooop/hadoop-3.3.1/tmp/dfs/name)	EditLogFileOutputStream(/home/jiayao/haooop/hadoop-3.3.1/tmp/dfs/name/current/edits_inprogress_0000000000000000003)

NameNode Storage

Storage Directory	Type	State
/home/jiayao/haooop/hadoop-3.3.1/tmp/dfs/name	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	18.62 GB	24 KB (0%)	7.96 GB (42.76%)	24 KB	1

- 6) 创建执行 MapReduce 作业的 Hdfs 目录并拷贝文件至分布式文件系统
- 7) 运行 grep 程序
- 8) 停止 daemon 运行

运行结果截图：


```

root@jiayaoyao-virtual-machine: /home/jiayaoyao/hadoop/hadoop-3.3.1
snapshotDiff      diff two snapshots of a directory or diff the current
                  directory contents with a snapshot
version           print the version

  Daemon Commands:

balancer          run a cluster balancing utility
datanode          run a DFS datanode
dfsrouter         run the DFS router
diskbalancer      Distributes data evenly among disks on a given node
httpfs           run HttpFS server, the HDFS HTTP Gateway
journalnode       run the DFS journalnode
mover            run a utility to move block replicas across storage types
namenode         run the DFS namenode
nfs3             run an NFS version 3 gateway
portmap          run a portmap service
secondarynamenode run the DFS secondary namenode
sps              run external storagepolicysatisfier
zkfc             run the ZK Failover Controller daemon

SUBCOMMAND may print help when invoked w/o parameters or with -h.
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# cat output/*
1      dfsadmin
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1#

```

2.2 基于 yarn 的单机伪分布

步骤:

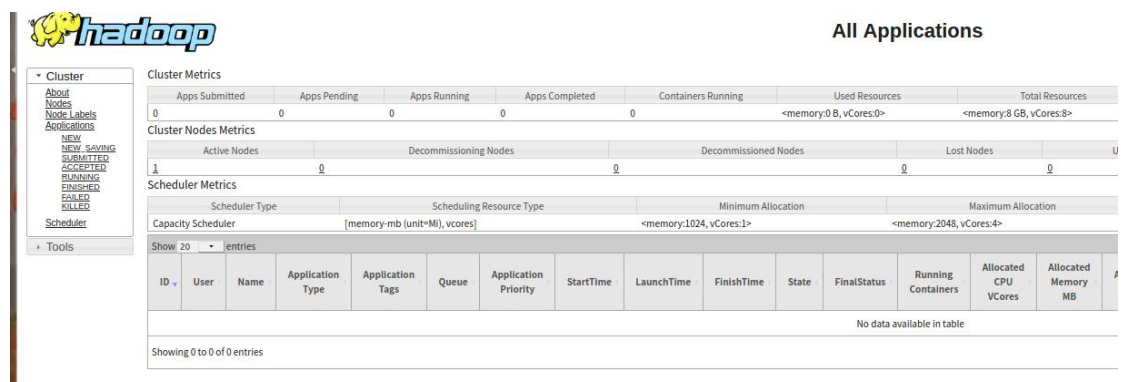
- 1) 配置 etc/hadoop/mapred-site.xml 与 etc/hadoop/yarn-site.xml
- 2) 启动 NameNode 与 ResourceManage , 使用 jps 查看

```

root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# sbin/start-y
arn.sh
Starting resourcemanager
Starting nodemanagers
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# jps
80144 Jps
78819 DataNode
78651 NameNode
79755 NodeManager
79034 SecondaryNameNode
79594 ResourceManager
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1#

```

- 3) 浏览 NameNode 的 Web 接口与 ResourceManager 的 Web 接口



hadoop

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources
0	0	0	0		<memory:0 B, vCores:0>	<memory:8 GB, vCores:8>

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>	<memory:2048, vCores:4>

Showing 0 to 0 of 0 entries

- 4) 运行 grep 程序
- 5) 停止 daemon 运行

运行过程以及结果截图：

运行过程：

```

/job_1634200597130_0006
2021-10-14 20:09:10,707 INFO input.FileInputFormat: Total input files to process : 31
2021-10-14 20:09:10,791 INFO mapreduce.JobSubmitter: number of splits:31
2021-10-14 20:09:10,929 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1634200597130_0006
2021-10-14 20:09:10,929 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-14 20:09:11,452 INFO conf.Configuration: resource-types.xml not found
2021-10-14 20:09:11,453 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-10-14 20:09:11,758 INFO impl.VarnClientImpl: Submitted application application_1634200597130_0006
2021-10-14 20:09:11,868 INFO mapreduce.Job: The url to track the job: http://jiayaoyao-virtual-machine:8088/proxy/application_1634200597130_0006/
2021-10-14 20:09:11,869 INFO mapreduce.Job: Running job: job_1634200597130_0006
2021-10-14 20:09:23,338 INFO mapreduce.Job: Job job_1634200597130_0006 running in uber mode : false
2021-10-14 20:09:23,339 INFO mapreduce.Job: map 0% reduce 0%
2021-10-14 20:10:25,648 INFO mapreduce.Job: map 3% reduce 0%
2021-10-14 20:10:29,863 INFO mapreduce.Job: map 6% reduce 0%
2021-10-14 20:10:30,891 INFO mapreduce.Job: map 19% reduce 0%
2021-10-14 20:10:45,805 INFO mapreduce.Job: map 23% reduce 0%
2021-10-14 20:10:48,951 INFO mapreduce.Job: map 35% reduce 0%
2021-10-14 20:10:49,956 INFO mapreduce.Job: map 39% reduce 0%
2021-10-14 20:11:01,629 INFO mapreduce.Job: map 45% reduce 0%
2021-10-14 20:11:02,778 INFO mapreduce.Job: map 55% reduce 0%
2021-10-14 20:11:08,965 INFO mapreduce.Job: map 55% reduce 18%
2021-10-14 20:11:12,086 INFO mapreduce.Job: map 58% reduce 18%
2021-10-14 20:11:14,628 INFO mapreduce.Job: map 61% reduce 18%
2021-10-14 20:11:15,675 INFO mapreduce.Job: map 65% reduce 20%
2021-10-14 20:11:16,684 INFO mapreduce.Job: map 71% reduce 20%
2021-10-14 20:11:21,702 INFO mapreduce.Job: map 71% reduce 24%
2021-10-14 20:11:23,802 INFO mapreduce.Job: map 74% reduce 24%
2021-10-14 20:11:25,883 INFO mapreduce.Job: map 81% reduce 24%
2021-10-14 20:11:26,894 INFO mapreduce.Job: map 87% reduce 24%
2021-10-14 20:11:27,902 INFO mapreduce.Job: map 87% reduce 27%
2021-10-14 20:11:34,314 INFO mapreduce.Job: map 87% reduce 29%
2021-10-14 20:11:35,320 INFO mapreduce.Job: map 94% reduce 29%
2021-10-14 20:11:36,333 INFO mapreduce.Job: map 100% reduce 29%
2021-10-14 20:11:37,385 INFO mapreduce.Job: map 100% reduce 100%
2021-10-14 20:11:38,397 INFO mapreduce.Job: Job job_1634200597130_0006 completed successfully
2021-10-14 20:11:38,506 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=467
  FILE: Number of bytes written=8741336
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=111762
  HDFS: Number of bytes written=577

```

结果：

```

root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# bin/hdfs dfs -get ou
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# cat output/*
cat: output/output: 是一个目录
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# cd output
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1/output# cat output/*
5 dfs.audit.logger
3 dfs.logger
3 dfs.server.namenode.
2 dfs.audit.log.maxfilesize
2 dfs.audit.log.maxbackupindex
2 dfs.sh
1 dfsadmin
1 dfs.replication
1 dfs.namenode.servicerpc
1 dfs.namenode.rpc
1 dfs.namenode.name.dir
1 dfs.log
1 dfs.datanode.data.dir
1 dfs.namenode.ec.policies.max.cellsize
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1/output#

```

3. 集群 Hadoop 系统安装与运行（基于 Hadoop）

步骤：

- 1) 使用 `sudo docker network create --driver=bridge hadoop` 创建虚拟网络
- 2) 拉取 Ubuntu16.04 镜像文件
- 3) 进入容器进行基础文件配置（首先安装 vim、sudo、ssh、net-tools、java, Scala, 并进行免密登录配置）

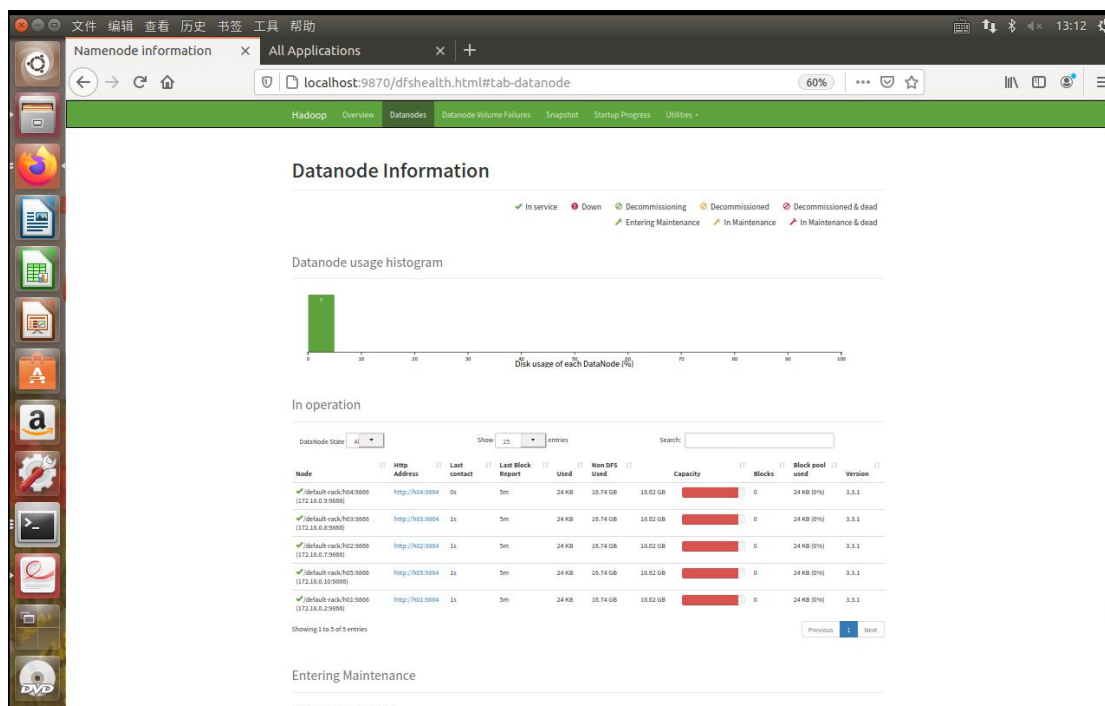
- 4) 安装 hadoop 并对环境变量进行配置。
- 5) 在 docker 中启动集群
- 6) 格式化 NameNode, 使用 `./hadoop dfsadmin -report` 查看分布式文件系统的状态, 使用 web 查看节点状态。

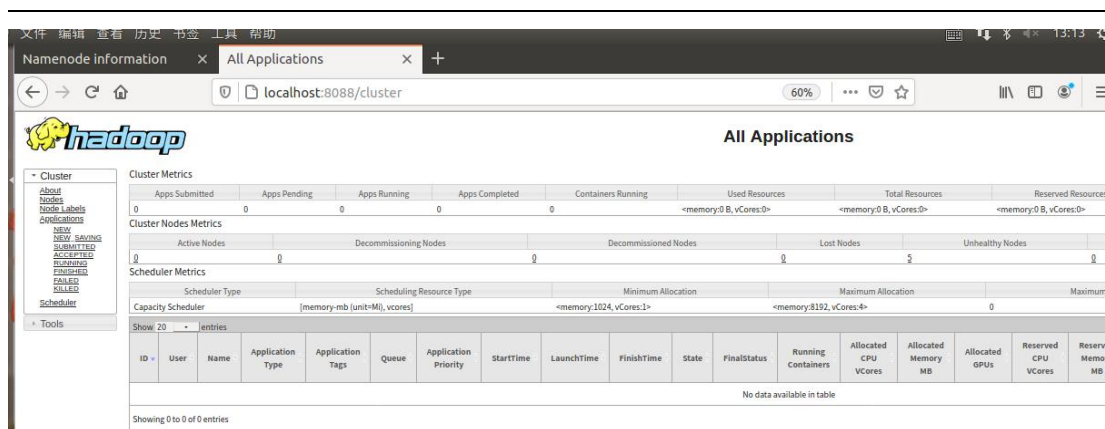
```
root@h01: /usr/local/hadoop# cd bin
root@h01: /usr/local/hadoop/bin# ./hadoop dfsadmin -report
WARNING: Use of this script to execute dfsadmin is deprecated.
WARNING: Attempting to execute replacement "hdfs dfsadmin" instead.

Configured Capacity: 99970560000 (93.10 GB)
Present Capacity: 4886835200 (4.55 GB)
DFS Remaining: 4886691840 (4.55 GB)
DFS Used: 143360 (140 KB)
DFS Used%: 0.00%
Replicated Blocks:
  Under replicated blocks: 0
  Blocks with corrupt replicas: 0
  Missing blocks: 0
  Missing blocks (with replication factor 1): 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0
Erasure Coded Block Groups:
  Low redundancy block groups: 0
  Block groups with corrupt internal blocks: 0
  Missing block groups: 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0

-----
Live datanodes (5):

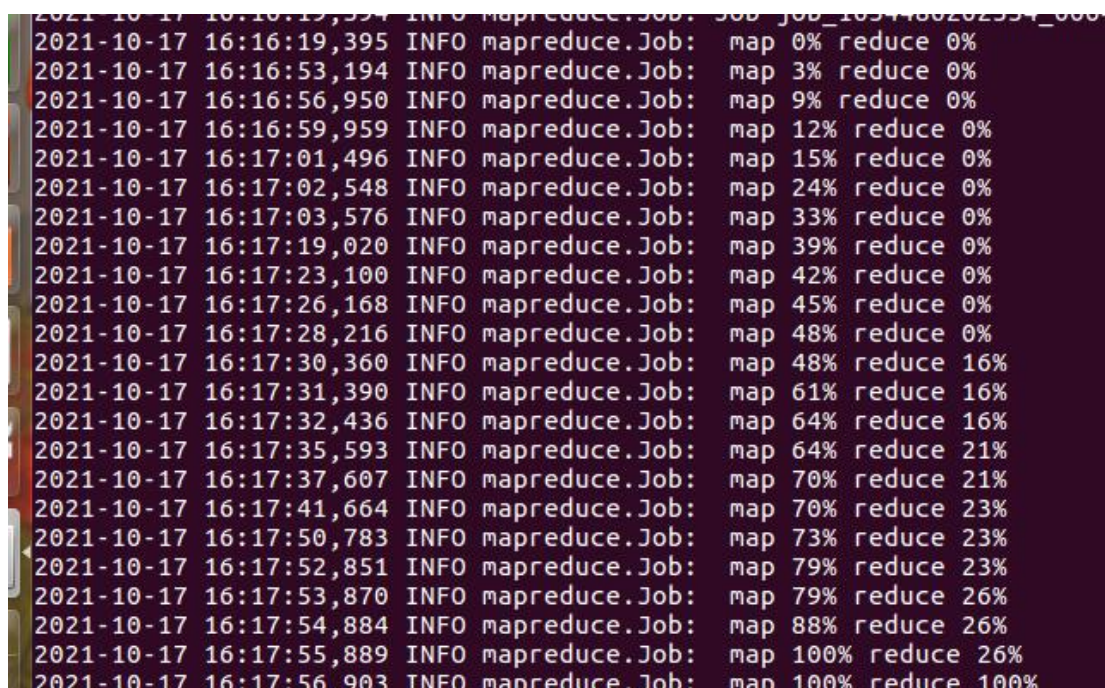
Name: 172.18.0.10:9866 (h05.hadoop)
Hostname: h05
Decommission Status : Normal
Configured Capacity: 19994112000 (18.62 GB)
DFS Used: 28672 (28 KB)
Non DFS Used: 17977503744 (16.74 GB)
DFS Remaining: 977338368 (932.06 MB)
DFS Used%: 0.00%
DFS Remaining%: 4.89%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Sun Oct 17 05:36:52 UTC 2021
Last Block Report: Sun Oct 17 04:53:22 UTC 2021
Num of Blocks: 0
```



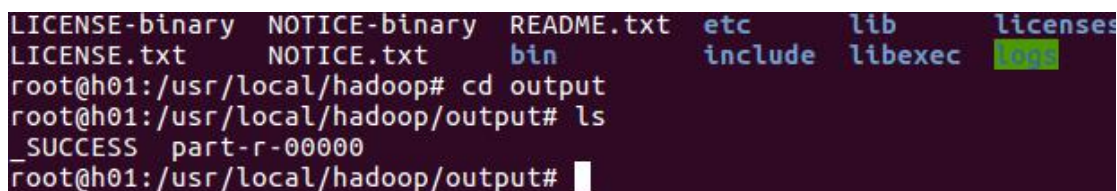


- 7) 创建执行 MapReduce 作业的 Hdfs 目录并拷贝文件至分布式文件系统
- 8) 运行 grep 程序
- 9) 停止 deamon 运行

实验结果截图：



结果：



二、实验过程中遇到的问题：

1. 单机 hadoop 系统安装:

1) 安装 Hadoop 后, 在 /etc/environment 配置环境变量后, 输入 whereis hadoop, 显示为空。

```
jiayaoyao@jiayaoyao-virtual-machine:~$ whereis java
java: /usr/local/java /usr/share/java /usr/local/java/jdk1.8.0_301/bin/java
/local/java/jdk1.8.0_301/jre/bin/java
jiayaoyao@jiayaoyao-virtual-machine:~$ whereis hadoop
hadoop:
jiayaoyao@jiayaoyao-virtual-machine:~$ bin/hadoop
bash: bin/hadoop: 没有那个文件或目录
```

通过查询, 又在 /etc/profile.d 创建了自己的环境变量.sh 文件。

输入以下内容:

```
jiayaoyao@jiayaoyao-virtual-machine:/
export JAVA_HOME=/usr/local/java/jdk1.8.0_301
export PATH=${JAVA_HOME}/bin:$PATH
export HADOOP_HOME=/home/jiayaoyao/hadoop/hadoop-3.3.1
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

再运行 source /etc/profile 和 hadoop version, 可以正常显示版本了。

```
jiayaoyao@jiayaoyao-virtual-machine:/$ hadoop version
Hadoop 3.3.1
Source code repository https://github.com/apache/hadoop.git -r a3b9c37a397ad4188
041dd80621bdeefc46885f2
Compiled by ubuntu on 2021-06-15T05:13Z
Compiled with protoc 3.7.1
From source with checksum 88a4ddb2299aca054416d6b7f81ca55
This command was run using /home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/comm
on/hadoop-common-3.3.1.jar
```

2. 伪分布式 Hadoop 系统安装和运行

2.1 无 yarn 伪分布式 Hadoop 系统安装和运行

配置好公钥后, 连接本地——ssh localhost 没有遇到问题。


```

jiayaoyao@jiayaoyao-virtual-machine: ~
jiayaoyao@jiayaoyao-virtual-machine:~/.ssh$ ssh localhost
Welcome to Ubuntu 16.04.7 LTS (GNU/Linux 4.15.0-142-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

UA Infra: Extended Security Maintenance (ESM) is not enabled.

0 updates can be applied immediately.

84 additional security updates can be applied with UA Infra: ESM
Learn more about enabling UA Infra: ESM service for Ubuntu 16.04 at
https://ubuntu.com/16-04

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by

```

打开 daemon:

```

bash: cd: /home/hadoop/hadoop-3.3.1: 没有那个文件或目录
jiayaoyao@jiayaoyao-virtual-machine:~$ cd /home/jiayaoyao/hadoop/hadoop-3.3.1
jiayaoyao@jiayaoyao-virtual-machine:~/hadoop/hadoop-3.3.1$ sbin/start-dfs.sh
/home/jiayaoyao/hadoop/hadoop-3.3.1/etc/hadoop/hadoop-env.sh: 行 1: #: 未找到命令
Starting namenodes on [localhost]
localhost: /home/jiayaoyao/hadoop/hadoop-3.3.1/etc/hadoop/hadoop-env.sh: 行 1: #
未找到命令
Starting datanodes
localhost: /home/jiayaoyao/hadoop/hadoop-3.3.1/etc/hadoop/hadoop-env.sh: 行 1: #
未找到命令
localhost: ERROR: Cannot set priority of datanode process 44201
Starting secondary namenodes [jiayaoyao-virtual-machine]

```

根据提示将首行的“#”删除，又显示‘(’ undefined。

最后将 hadoop-env.sh 删除了重写，在里面重新添加了路径以及：

```
export HDFS_NAMENODE_USER=root
```

```
export HDFS_DATANODE_USER=root
```

```
export HDFS_SECONDARYNAMENODE_USER=root
```

没有出现类似的报错了。

继续运行，出现：

```

jiayaoyao@jiayaoyao-virtual-machine:~/hadoop/hadoop-3.3.1$ sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
localhost: ERROR: Cannot set priority of datanode process 48980
Starting secondary namenodes [jiayaoyao-virtual-machine]
jiayaoyao@jiayaoyao-virtual-machine:~/hadoop/hadoop-3.3.1$

```

```

jiayaoyao@jiayaoyao-virtual-machine:~/hadoop/hadoop-3.3.1$ sbin/start-dfs.sh
Starting namenodes on [localhost]
ERROR: namenode can only be executed by root.
Starting datanodes
ERROR: datanode can only be executed by root.
Starting secondary namenodes [jiayaoyao-virtual-machine]
ERROR: secondarynamenode can only be executed by root.

```

需要权限登录，然后以 root 用户的形式登录。

root 用户出现了无法连接的问题，如下：

```

jiayaoyao@jiayaoyao-virtual-machine:~/hadoop/hadoop-3.3.1$ su
密码：
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# sbin/start-dfs.sh
Starting namenodes on [localhost]
localhost: Permission denied (publickey,password).
Starting datanodes
localhost: Permission denied (publickey,password).
Starting secondary namenodes [jiayaoyao-virtual-machine]
jiayaoyao-virtual-machine: Permission denied (publickey,password).

```

运行 ssh localhost，不成功（普通用户 ssh localhost 可以成功）

重新修改一些文件的权限：

```

jiayaoyao-virtual-machine: Permission denied (publickey,password).
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# chmod 700 ~/.ssh
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# chmod 644 ~/.ssh/authorized_keys

```

在 root 用户下运行 ssh localhost，可以成功：

```

root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# ssh localhost
Welcome to Ubuntu 16.04.7 LTS (GNU/Linux 4.15.0-142-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

UA Infra: Extended Security Maintenance (ESM) is not enabled.

```

再打开 daemon 没有出现 warning 和报错：

```

root@jiayaoyao-virtual-machine:~# cd /home/jiayaoyao/hadoop/hadoop-3.3.1
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [jiayaoyao-virtual-machine]
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1#

```

查看状态却是不正常的（没有 namenode）：

```

root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# jps
54641 Jps
54170 DataNode
54394 SecondaryNameNode
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1#

```

查看日志：


```

org.apache.hadoop.hdfs.server.common.InconsistentFSStateException: Directory /home/jiayayao/hadoop/hadoop-3.3.1/tmp/dfs/name is in
an inconsistent state: storage directory does not exist or is not accessible.
    at org.apache.hadoop.hdfs.server.namenode.FSImage.recoverStorageDirs(FSImage.java:392)
    at org.apache.hadoop.hdfs.server.namenode.FSImage.recoverTransitionRead(FSImage.java:243)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFSImage(FSNamesystem.java:1197)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFromDisk(FSNamesystem.java:779)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.loadNamesystem(NameNode.java:677)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.initialize(NameNode.java:764)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:1018)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:991)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1767)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1832)
2021-10-10 19:55:13,772 INFO org.apache.hadoop.util.ExitUtil: Exiting with status 1:
org.apache.hadoop.hdfs.server.common.InconsistentFSStateException: Directory /home/jiayayao/hadoop/hadoop-3.3.1/tmp/dfs/name is in
an inconsistent state: storage directory does not exist or is not accessible.
2021-10-10 19:55:13,774 INFO org.apache.hadoop.hdfs.server.namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at jiayayao-virtual-machine/127.0.1.1
*****/

```

查询网上各种解决办法，都没有成功。

于是我重新格式化了 namenode，这次使用 jps 可以看到 namenode 了。

对 namenode 进行 web 查看，也是可以正确显示的。

上传文件不成功：

```

oot@jiayayao-virtual-machine:/home/jiayayao/hadoop/hadoop-3.3.1# bin/hdfs dfs
-mkdir /user/jiayayao
oot@jiayayao-virtual-machine:/home/jiayayao/hadoop/hadoop-3.3.1# bin/hdfs dfs
-put etc/hadoop input
ut: 'input': No such file or directory: 'hdfs://localhost:9000/user/root/input'
oot@jiayayao-virtual-machine:/home/jiayayao/hadoop/hadoop-3.3.1# bin/hdfs dfs
-mkdir /user/root

```

开始的时候直接用自己的名字命名，结果报错没有该目录，也没有办法上传文件。经过查询，由于我以 root 用户的身份登录，所以建立的用户名字应该是 root。

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/user Go! [Icons]

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	root	supergroup	0 B	Oct 10 20:53	0	0 B	hadoop
drwxr-xr-x	root	supergroup	0 B	Oct 10 21:07	0	0 B	jiayayao
drwxr-xr-x	root	supergroup	0 B	Oct 10 21:13	0	0 B	root

Showing 1 to 3 of 3 entries Previous 1 Next

Hadoop, 2021.

重新建立 root 文件夹后，进行代码的执行，最终成功。

2.2 基于 yarn 的单机伪分布式运行

运行 sbin/start-dfs.sh 以及 sbin/start-yarn.sh 后使用 jps 查看，显示正常。

查看 web 端，显示正常

但运行代码，出现错误：


```

root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar grep input output1 'dfs[a-z.]+'
2021-10-11 00:36:46,065 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2021-10-11 00:36:58,058 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1633881097685_0011
2021-10-11 00:36:59,204 INFO input.FileInputFormat: Total input files to process : 32
2021-10-11 00:37:01,950 INFO mapreduce.JobSubmitter: number of splits:32
2021-10-11 00:37:02,759 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1633881097685_0011
2021-10-11 00:37:02,759 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-11 00:37:04,874 INFO conf.Configuration: resource-types.xml not found
2021-10-11 00:37:04,874 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-10-11 00:37:05,020 INFO impl.YarnClientImpl: Submitted application application_1633881097685_0011
2021-10-11 00:37:05,098 INFO mapreduce.Job: The url to track the job: http://jiayaoyao-virtual-machine:8088/proxy/application_1633881097685_0011/
2021-10-11 00:37:05,098 INFO mapreduce.Job: Running job: job_1633881097685_0011
2021-10-11 00:37:15,420 INFO mapreduce.Job: Job job_1633881097685_0011 running in uber mode : false
2021-10-11 00:37:15,428 INFO mapreduce.Job: map 0% reduce 0%
2021-10-11 00:37:15,444 INFO mapreduce.Job: Job job_1633881097685_0011 failed with state FAILED due to: Application application_1633881097685_0011 failed 2 times due to AM Container for appattempt_1633881097685_0011_000002 exited with exitCode: 1
Failing this attempt.Diagnostics: [2021-10-11 00:37:14.942]Exception from container-launch.
Container id: container_1633881097685_0011_02_000001
Exit code: 1

[2021-10-11 00:37:14.945]Container exited with a non-zero exit code 1. Error file: prelaunch.err.
Last 4096 bytes of prelaunch.err :
Last 4096 bytes of stderr :
错误: 找不到或无法加载主类 org.apache.hadoop.mapreduce.v2.app.MRAppMaster

[2021-10-11 00:37:14.946]Container exited with a non-zero exit code 1. Error file: prelaunch.err.
Last 4096 bytes of prelaunch.err :
Last 4096 bytes of stderr :
错误: 找不到或无法加载主类 org.apache.hadoop.mapreduce.v2.app.MRAppMaster

For more detailed output, check the application tracking page: http://jiayaoyao-virtual-machine:8088/cluster/app/application_1633881097685_0011 Then click on links to logs of each attempt.
. Failing the application.
2021-10-11 00:37:15,582 INFO mapreduce.Job: Counters: 0
2021-10-11 00:37:15,628 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032

```

经查找，需要补充 classpath 到文件中。

```

root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# hadoop classpath
/home/jiayaoyao/hadoop/hadoop-3.3.1/etc/hadoop:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/common/lib/*:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/common/*:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/hdfs:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/hdfs/lib/*:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/hdfs/*:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/mapreduce/*:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/yarn:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/yarn/lib/*:/home/jiayaoyao/hadoop/hadoop-3.3.1/share/hadoop/yarn/*
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1#

```

将这些值添加到 yarn-site.xml 文件中的 yarn.application.classpath 下面。

重启 yarn，显示如下：

```

79834 SecondaryNameNode
83213 MRAppMaster
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar grep input output 'dfs[a-z.]+'
2021-10-11 01:21:15,542 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2021-10-11 01:21:17,250 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1633886456395_0001
2021-10-11 01:21:18,497 INFO input.FileInputFormat: Total input files to process : 32
2021-10-11 01:21:18,804 INFO mapreduce.JobSubmitter: number of splits:32
2021-10-11 01:21:19,397 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1633886456395_0001
2021-10-11 01:21:19,397 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-11 01:21:19,732 INFO conf.Configuration: resource-types.xml not found
2021-10-11 01:21:19,732 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-10-11 01:21:20,224 INFO impl.YarnClientImpl: Submitted application application_1633886456395_0001
2021-10-11 01:21:20,272 INFO mapreduce.Job: The url to track the job: http://jiayaoyao-virtual-machine:8088/proxy/application_1633886456395_0001/
2021-10-11 01:21:20,273 INFO mapreduce.Job: Running job: job_1633886456395_0001
2021-10-11 01:22:17,782 INFO mapreduce.Job: Job job_1633886456395_0001 running in uber mode : false
2021-10-11 01:22:17,888 INFO mapreduce.Job: map 0% reduce 0%

```

等待一小时多毫无反应。

```

2021-10-11 01:22:17,888 INFO mapreduce.Job: map 0% reduce 0%

ACACAC
2021-10-11 01:37:00,492 INFO ipc.Client: Retrying connect to server: jiayaoyao-virtual-machine/127.0.1.1:44495. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1#

```

最后的报错是这样的：


```

终端 文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
Stopping secondary namenodes [jiayaoyao-virtual-machine]
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [jiayaoyao-virtual-machine]
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# jps
29041 SecondaryNameNode
29843 Jps
29476 NodeManager
29316 ResourceManager
28790 DataNode
28620 NameNode
26590 MRAppMaster
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3
.3.1.jar grep input output 'dfs[a-z.]+'
2021-10-13 18:46:35,115 INFO client.DefaultNoHARMFaloverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2021-10-13 18:46:35,902 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging
/job_1634121982208_0001
2021-10-13 18:46:36,885 INFO input.FileInputFormat: Total input files to process : 33
2021-10-13 18:46:37,442 INFO mapreduce.JobSubmitter: number of splits:33
2021-10-13 18:46:37,737 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1634121982208_0001
2021-10-13 18:46:37,737 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-13 18:46:38,138 INFO conf.Configuration: resource-types.xml not found
2021-10-13 18:46:38,138 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-10-13 18:46:38,707 INFO impl.YarnClientImpl: Submitted application application_1634121982208_0001
2021-10-13 18:46:38,755 INFO mapreduce.Job: The url to track the job: http://jiayaoyao-virtual-machine:8088/proxy/application_1634121
982208_0001/
2021-10-13 18:46:38,755 INFO mapreduce.Job: Running job: job_1634121982208_0001
2021-10-13 18:46:52,001 INFO mapreduce.Job: Job job_1634121982208_0001 running in uber mode : false
2021-10-13 18:46:52,022 INFO mapreduce.Job: map 0% reduce 0%
2021-10-13 18:50:25,669 INFO ipc.Client: Retrying connect to server: jiayaoyao-virtual-machine/127.0.1.1:43417. Already tried 0 time(
s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2021-10-13 18:50:26,672 INFO ipc.Client: Retrying connect to server: jiayaoyao-virtual-machine/127.0.1.1:43417. Already tried 1 time(
s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2021-10-13 18:50:27,677 INFO ipc.Client: Retrying connect to server: jiayaoyao-virtual-machine/127.0.1.1:43417. Already tried 2 time(
s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2021-10-13 19:25:27,063 INFO ipc.Client: Retrying connect to server: 0.0.0.0/0.0.0.0:8032. Already tried 0 time(s); retry policy is R
etryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2021-10-13 19:25:37,145 INFO ipc.Client: Retrying connect to server: 0.0.0.0/0.0.0.0:8032. Already tried 1 time(s); retry policy is R
etryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2021-10-13 19:25:38,615 INFO ipc.Client: Retrying connect to server: 0.0.0.0/0.0.0.0:8032. Already tried 2 time(s); retry policy is R
etryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)

```

我尝试删除 tmp 后重新格式化 namenode，并不成功。

于是我删除了 hdfs 文件再进行格式化，仍然不可行。

最后只能删了 hadoop 再重装。

配置好环境后，首先我运行了 pi 程序作为先验测试，报错如下：

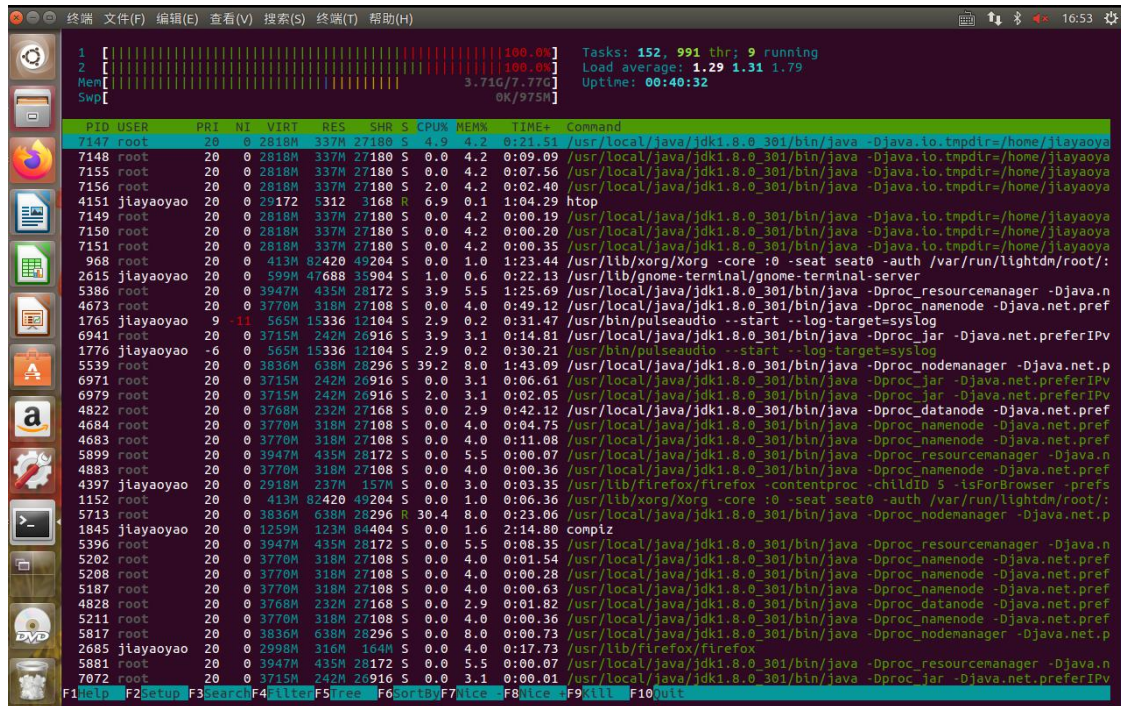
```

终端 文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
Wrote input for Map #15
Wrote input for Map #16
Wrote input for Map #17
Wrote input for Map #18
Wrote input for Map #19
Starting Job
2021-10-14 14:28:42,682 INFO client.DefaultNoHARMFaloverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2021-10-14 14:28:50,876 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging
/job_1634192886207_0001
2021-10-14 14:28:51,609 INFO input.FileInputFormat: Total input files to process : 20
2021-10-14 14:28:51,694 INFO mapreduce.JobSubmitter: number of splits:20
2021-10-14 14:28:52,452 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1634192886207_0001
2021-10-14 14:28:52,452 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-14 14:28:53,289 INFO conf.Configuration: resource-types.xml not found
2021-10-14 14:28:53,291 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-10-14 14:28:58,258 INFO impl.YarnClientImpl: Submitted application application_1634192886207_0001
2021-10-14 14:28:58,417 INFO mapreduce.Job: The url to track the job: http://jiayaoyao-virtual-machine:8088/proxy/application_1634192
886207_0001/
2021-10-14 14:28:58,418 INFO mapreduce.Job: Running job: job_1634192886207_0001
2021-10-14 14:30:29,226 INFO mapreduce.Job: Job job_1634192886207_0001 running in uber mode : false
2021-10-14 14:30:29,229 INFO mapreduce.Job: map 0% reduce 0%
2021-10-14 14:30:29,247 INFO mapreduce.Job: Job job_1634192886207_0001 failed with state FAILED due to: Application application_16341
92886207_0001 failed 2 times due to AM Container for appattempt1634192886207_0001_000002 exited with exitCode: 1
Failing this attempt.Diagnostics: [2021-10-14 14:30:28.323]Exception from container-launch.
Container id: container_1634192886207_0001_02_000001
Exit code: 1
[2021-10-14 14:30:28.332]Container exited with a non-zero exit code 1. Error file: prelaunch.err.
Last 4096 bytes of prelaunch.err :
Last 4096 bytes of stderr :
错误: 找不到或无法加载主类 org.apache.hadoop.mapreduce.v2.app.MRAppMaster
[2021-10-14 14:30:28.332]Container exited with a non-zero exit code 1. Error file: prelaunch.err.
Last 4096 bytes of prelaunch.err :
Last 4096 bytes of stderr :
错误: 找不到或无法加载主类 org.apache.hadoop.mapreduce.v2.app.MRAppMaster
For more detailed output, check the application tracking page: http://jiayaoyao-virtual-machine:8088/cluster/app/application_16341928
86207_0001 Then click on links to logs of each attempt.
Failing the application.
2021-10-14 14:30:29,282 INFO mapreduce.Job: Counters: 0
Job job_1634192886207_0001 failed!
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1#

```

经查询，在 mapred 中增添了 MapReduce 的 classpath 可以运行，但一直停留在 map0reduce0 的界面，等待无果。

重启了电脑和虚拟机，后了解到 htop 可查看整个虚拟机的进程以及运行情况，于是下载安装了一下然后进行查看，发现 cpu 的使用率达到了 100%。



重新运行 pi 程序并且将参数 20 10 改为 6 3，等待了大约 5min，pi 程序终于出现了结果。

```

root@jiayao-yao-virtual-machine: /home/jiayao-yao/hadoop/hadoop-3.3.1
Merged Map outputs=6
GC time elapsed (ms)=53834
CPU time spent (ms)=87780
Physical memory (bytes) snapshot=2272698368
Virtual memory (bytes) snapshot=18458378240
Total committed heap usage (bytes)=1917321216
Peak Map Physical memory (bytes)=346701824
Peak Map Virtual memory (bytes)=2638434304
Peak Reduce Physical memory (bytes)=208429056
Peak Reduce Virtual memory (bytes)=2642685952
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=708
File Output Format Counters
Bytes Written=97
Job Finished in 265.285 seconds
Estimated value of Pi is 3.7777777777777777
root@jiayao-yao-virtual-machine: /home/jiayao-yao/hadoop/hadoop-3.3.1#

```

这次运行 grep 程序终于不会卡在 map0%reduce0%了。


```

HDFS: Number of bytes read=0
HDFS: Number of bytes written=0
HDFS: Number of read operations=5
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched reduce tasks=1
  Total time spent by all maps in occupied slots (ms)=0
  Total time spent by all reduces in occupied slots (ms)=1930
  Total time spent by all reduce tasks (ms)=1930
  Total vcore-milliseconds taken by all reduce tasks=1930
  Total megabyte-milliseconds taken by all reduce tasks=1976320
Map-Reduce Framework
  Combine input records=0
  Combine output records=0
  Reduce input groups=0
  Reduce shuffle bytes=0
  Reduce input records=0
  Reduce output records=0
  Spilled Records=0
  Shuffled Maps =0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=41
  CPU time spent (ms)=330
  Physical memory (bytes) snapshot=233574400
  Virtual memory (bytes) snapshot=2639888384
  Total committed heap usage (bytes)=191365120
  Peak Reduce Physical memory (bytes)=233574400
  Peak Reduce Virtual memory (bytes)=2639888384
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Output Format Counters
  Bytes Written=0
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# bin/hdfs dfs -get output output

```

查看结果:

```

root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# bin/hdfs dfs -get output output
get: 'output/output/_SUCCESS': File exists
get: 'output/output/part-r-00000': File exists
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# bin/hdfs dfs -get output output
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# cat output/*
part-r-00000 _SUCCESS
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1#

```

翻日志看到了一些问题:

```

2021-10-14 17:03:22,645 INFO mapreduce.Job: Running job: job_1634200597130_0004
2021-10-14 17:03:41,756 INFO mapreduce.Job: Job job_1634200597130_0004 running in uber mode : false
2021-10-14 17:03:41,834 INFO mapreduce.Job: map 0% reduce 0%
2021-10-14 17:04:05,422 INFO mapreduce.Job: map 0% reduce 0%
2021-10-14 17:04:06,429 INFO mapreduce.Job: map 16% reduce 0%
2021-10-14 17:04:07,433 INFO mapreduce.Job: map 19% reduce 0%
2021-10-14 17:04:15,488 INFO mapreduce.Job: map 22% reduce 0%
2021-10-14 17:04:18,586 INFO mapreduce.Job: map 28% reduce 0%
2021-10-14 17:04:19,611 INFO mapreduce.Job: map 38% reduce 0%
2021-10-14 17:04:22,676 INFO mapreduce.Job: map 47% reduce 0%
2021-10-14 17:04:30,700 INFO mapreduce.Job: map 53% reduce 0%
2021-10-14 17:04:35,730 INFO mapreduce.Job: map 53% reduce 18%
2021-10-14 17:04:37,746 INFO mapreduce.Job: map 56% reduce 18%
2021-10-14 17:04:40,770 INFO mapreduce.Job: map 69% reduce 18%
2021-10-14 17:04:41,778 INFO mapreduce.Job: map 69% reduce 22%
2021-10-14 17:04:47,812 INFO mapreduce.Job: map 72% reduce 23%
2021-10-14 17:04:49,827 INFO mapreduce.Job: map 75% reduce 23%
2021-10-14 17:04:50,832 INFO mapreduce.Job: map 78% reduce 23%
2021-10-14 17:04:51,835 INFO mapreduce.Job: map 84% reduce 23%
2021-10-14 17:04:53,844 INFO mapreduce.Job: map 84% reduce 28%
2021-10-14 17:04:56,864 INFO mapreduce.Job: map 88% reduce 28%
2021-10-14 17:04:59,885 INFO mapreduce.Job: map 91% reduce 29%
2021-10-14 17:04:59,887 INFO mapreduce.Job: Task Id : attempt_1634200597130_0004_m_000031_0, Status : FAILED
Error: java.io.FileNotFoundException: Path is not a file: /user/root/input/shellprofile.d
    at org.apache.hadoop.hdfs.server.namenode.INodeFile.valueOf(INodeFile.java:90)
    at org.apache.hadoop.hdfs.server.namenode.INodeFile.valueOf(INodeFile.java:76)
    at org.apache.hadoop.hdfs.server.namenode.FSDirStatAndListingOp.getBlockLocations(FSDirStatAndListingOp.java:156)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getBlockLocations(FSNamesystem.java:2071)
    at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.getBlockLocations(NameNodeRpcServer.java:773)
    at org.apache.hadoop.hdfs.protocolPB.ClientNameNodeProtocolServerSideTranslatorPB.getBlockLocations(ClientNameNodeProtocolServerSideTranslatorPB.java:458)
    at org.apache.hadoop.hdfs.protocol.proto.ClientNameNodeProtocolProtos$ClientNameNodeProtocol$2.callBlockingMethod(ClientNameNodeProtocolProtos.java)
    at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:600)
    at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:568)
    at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:552)
    at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1093)
    at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:1035)
    at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:963)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)
    at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2966)
    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)

```

这是因为上传 input 文件中还套有一个文件夹，把文件夹后删掉重新上传并运行。

```
/job_1634200597130_0006
2021-10-14 20:09:10,707 INFO input.FileInputFormat: Total input files to process : 31
2021-10-14 20:09:10,791 INFO mapreduce.JobSubmitter: number of splits:31
2021-10-14 20:09:10,929 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1634200597130_0006
2021-10-14 20:09:10,929 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-14 20:09:11,452 INFO conf.Configuration: resource-types.xml not found
2021-10-14 20:09:11,453 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-10-14 20:09:11,758 INFO impl.YarnClientImpl: Submitted application application_1634200597130_0006
2021-10-14 20:09:11,868 INFO mapreduce.Job: The url to track the job: http://jiayaoyao-virtual-machine:8088/proxy/application_1634200597130_0006/
2021-10-14 20:09:11,869 INFO mapreduce.Job: Running job: job_1634200597130_0006
2021-10-14 20:09:23,338 INFO mapreduce.Job: Job job_1634200597130_0006 running in uber mode : false
2021-10-14 20:09:23,339 INFO mapreduce.Job: map 0% reduce 0%
2021-10-14 20:10:25,648 INFO mapreduce.Job: map 3% reduce 0%
2021-10-14 20:10:29,863 INFO mapreduce.Job: map 6% reduce 0%
2021-10-14 20:10:30,891 INFO mapreduce.Job: map 19% reduce 0%
2021-10-14 20:10:45,805 INFO mapreduce.Job: map 23% reduce 0%
2021-10-14 20:10:48,951 INFO mapreduce.Job: map 35% reduce 0%
2021-10-14 20:10:49,956 INFO mapreduce.Job: map 39% reduce 0%
2021-10-14 20:11:01,629 INFO mapreduce.Job: map 45% reduce 0%
2021-10-14 20:11:02,778 INFO mapreduce.Job: map 55% reduce 0%
2021-10-14 20:11:08,965 INFO mapreduce.Job: map 55% reduce 18%
2021-10-14 20:11:12,086 INFO mapreduce.Job: map 58% reduce 18%
2021-10-14 20:11:14,628 INFO mapreduce.Job: map 61% reduce 18%
2021-10-14 20:11:15,675 INFO mapreduce.Job: map 65% reduce 20%
2021-10-14 20:11:16,684 INFO mapreduce.Job: map 71% reduce 20%
2021-10-14 20:11:21,702 INFO mapreduce.Job: map 71% reduce 24%
2021-10-14 20:11:23,802 INFO mapreduce.Job: map 74% reduce 24%
2021-10-14 20:11:25,883 INFO mapreduce.Job: map 81% reduce 24%
2021-10-14 20:11:26,894 INFO mapreduce.Job: map 87% reduce 24%
2021-10-14 20:11:27,902 INFO mapreduce.Job: map 87% reduce 27%
2021-10-14 20:11:34,314 INFO mapreduce.Job: map 87% reduce 29%
2021-10-14 20:11:35,320 INFO mapreduce.Job: map 94% reduce 29%
2021-10-14 20:11:36,333 INFO mapreduce.Job: map 100% reduce 29%
2021-10-14 20:11:37,385 INFO mapreduce.Job: map 100% reduce 100%
2021-10-14 20:11:38,397 INFO mapreduce.Job: Job job_1634200597130_0006 completed successfully
2021-10-14 20:11:38,506 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=467
  FILE: Number of bytes written=8741336
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=111762
  HDFS: Number of bytes written=577
```

这次终于运行成功。

```
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# bin/hdfs dfs -get ou
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# cat output/*
cat: output/output: 是一个目录
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1# cd output
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1/output# cat output/*
5      dfs.audit.logger
3      dfs.logger
3      dfs.server.namenode.
2      dfs.audit.log.maxfilesize
2      dfs.audit.log.maxbackupindex
2      dfs.sh
1      dfsadmin
1      dfs.replication
1      dfs.namenode.servicerpc
1      dfs.namenode.rpc
1      dfs.namenode.name.dir
1      dfs.log
1      dfs.datanode.data.dir
1      dfs.namenode.ec.policies.max.cellsize
root@jiayaoyao-virtual-machine:/home/jiayaoyao/hadoop/hadoop-3.3.1/output#
```

3. 利用 docker 构建 hadoop 集群：

完全按照知乎的教程配置，但是运行 grep 程序，卡住：


```

at org.apache.hadoop.mapreduce.JobSubmitter.writeSplits(JobSubmitter.java:327)
at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:200)
at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1571)
at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1568)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.mapreduce.Job.submit(Job.java:1568)
at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1589)
at org.apache.hadoop.examples.Grep.run(Grep.java:78)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
at org.apache.hadoop.examples.Grep.main(Grep.java:103)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:71)
at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
Caused by: java.io.IOException: Input path does not exist: hdfs://h01:9000/user/root/input
at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.singleThreadedListStatus(FileInputFormat.java:313)
... 28 more
root@h01:/usr/local/hadoop# bin/hdfs dfs -mkdir /user
root@h01:/usr/local/hadoop# bin/hdfs dfs -mkdir /user/root
root@h01:/usr/local/hadoop# bin/hdfs dfs -put etc/hadoop input
root@h01:/usr/local/hadoop# bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar grep input output 'dfs[a-z.]+'
2021-10-17 06:11:07,318 INFO client.DefaultHadoopMFailoverProxyProvider: Connecting to ResourceManager at h01/172.18.0.2:8032
2021-10-17 06:11:08,126 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging
/job_1634446552423_0002
2021-10-17 06:11:08,827 INFO input.FileInputFormat: Total input files to process : 34
2021-10-17 06:11:09,875 INFO mapreduce.JobSubmitter: number of splits:34
2021-10-17 06:11:12,251 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1634446552423_0002
2021-10-17 06:11:12,251 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-17 06:11:17,286 INFO conf.Configuration: resource-types.xml not found
2021-10-17 06:11:17,287 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-10-17 06:11:22,354 INFO impl.YarnClientImpl: Submitted application application_1634446552423_0002
2021-10-17 06:11:22,623 INFO mapreduce.Job: The url to track the job: http://h01:8088/proxy/application_1634446552423_0002/
2021-10-17 06:11:22,624 INFO mapreduce.Job: Running job: job_1634446552423_0002

```

查看 log，没有发现报错，按照网上的好多教程，也没有正确运行，把 h01 等 5 个节点删除并重新进入容器中进行排查。

在运行 hadoop version 时发生了报错：

```

root@ceeef339f07b:/usr/local/hadoop/etc/hadoop# hadoop version
WARNING: log4j.properties is not found. HADOOP_CONF_DIR may be incomplete.
Hadoop 3.3.1
Source code repository https://github.com/apache/hadoop.git -r a3b9c37a397ad4188041dd80621bdeefc46885f2
Compiled by ubuntu on 2021-06-15T05:13Z
Compiled with protoc 3.7.1
From source with checksum 88a4ddb2299aca054416d6b7f81ca55
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-3.3.1.jar

```

查询该报错，是 conf_dir 没有配置好。进入 etc/profile 查看，知乎提供的 conf_dir 是错误的。

更改配置后，可以使用 hadoop version 进行查询。

重新配置节点。

进行运行，仍然不可以。

查看 node，发现：

/default-rack	UNHEALTHY	h01:37211	<u>h01:8042</u>	Sun Oct 17 15:04:13 +0000 2021	1/1 local-dirs usable space is below configured utilization percentage/no more usable space [/usr/local /hadoop /tmp/nm-local-dir: used space above threshold of 90.0%]; 1/1 log-dirs usable space is below configured utilization percentage/no more usable	0	0B	0B	0	0
---------------	-----------	-----------	-----------------	--------------------------------	--	---	----	----	---	---

似乎是内存空间不够，对虚拟机进行了扩容。

使用 `gparted` 进行查看，果然空间只剩下不到 1G。



The screenshot shows the GParted interface for disk /dev/sda (60.00 GiB). The main view displays four partitions: /dev/sda1 (19.04 GiB, ext4, /), /dev/sda2 (975.00 MiB, extended), /dev/sda5 (975.00 MiB, linux-swap), and a large unallocated space of 40.00 GiB. A table below provides a detailed summary of these partitions.

分区	文件系统	挂载点	大小	已用	未用	标识
/dev/sda1	ext4	/	19.04 GiB	17.07 GiB	1.97 GiB	boot
/dev/sda2	extended		975.00 MiB	—	—	
/dev/sda5	linux-swap		975.00 MiB	0.00 字节	975.00 MiB	
未分配	未分配		40.00 GiB	—	—	

扩充内存，再次运行 `grep` 程序，终于运行成功了。