

## 第01讲 Windows10 中安装Hadoop

- 1、安装JDK
- 2、安装Hadoop
- 3、安装Hadoop在windows专用二进制文件和工具类依赖库winutils和Hadoop.dll文件
- 4、启动Hadoop平台
- 5、打开web控制台GUI
- 6、hadoop、hdfs命令使用
- 7、QA常见问题

# 第01讲 Windows10 中安装Hadoop

## 1、安装JDK

- 1、到Oracle官网，选择jdk-8u202-windows-x64，下载jdk-8u202-windows-x64.exe
- 2、配置JDK的环境变量

WIN+R找开运行框，输入cmd 回车， 输入java -version, 显示如下页面，JDK安装成功。



```
C:\Windows\system32\cmd.exe
Microsoft Windows [版本 10.0.18362.356]
(c) 2019 Microsoft Corporation。保留所有权利。

C:\Users\hadoop>java -version
java version "1.8.0_202"
Java(TM) SE Runtime Environment (build 1.8.0_202-b08)
Java HotSpot(TM) 64-Bit Server VM (build 25.202-b08, mixed mode)

C:\Users\hadoop>javac -version
javac 1.8.0_202
C:\Users\hadoop>_
```

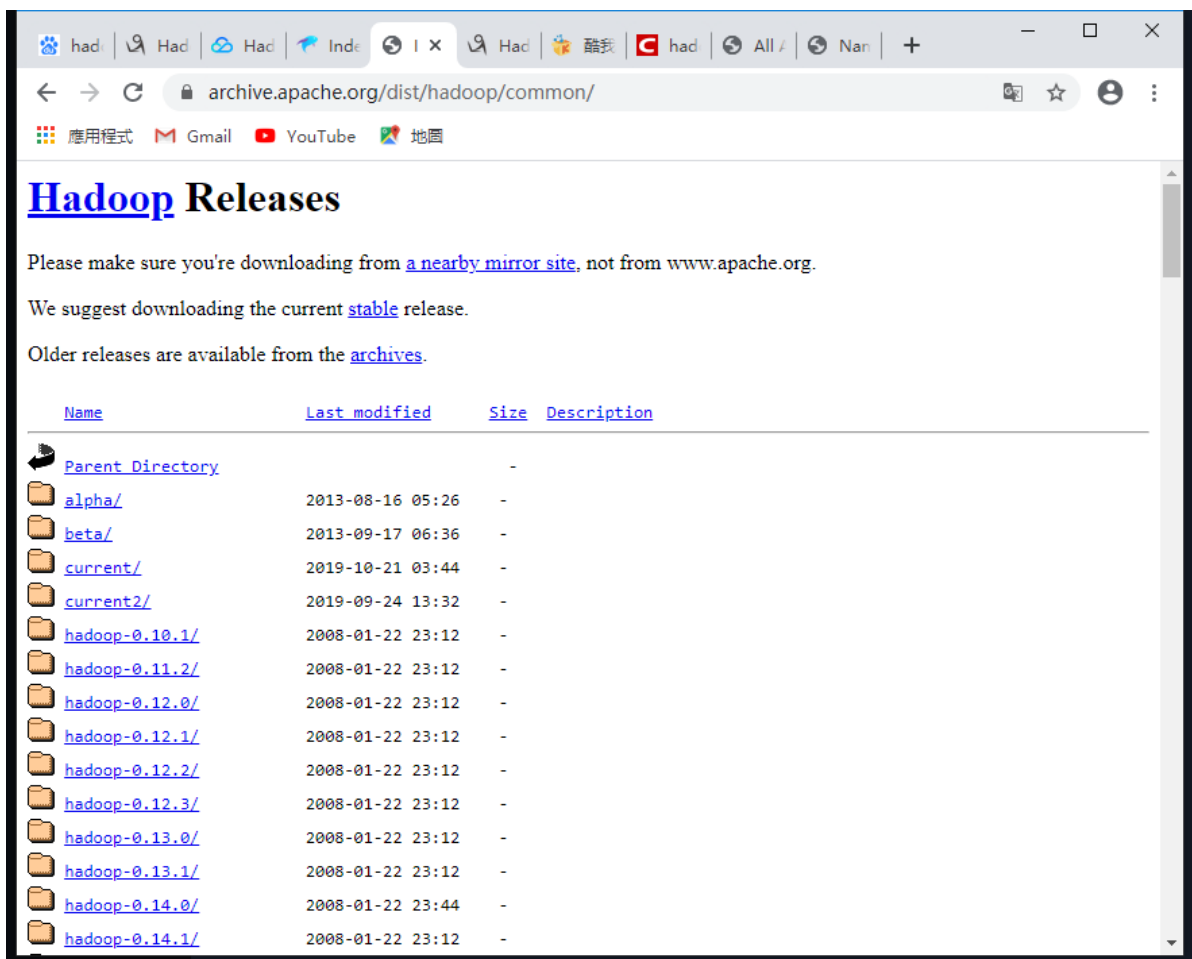
Java

Javac编译器

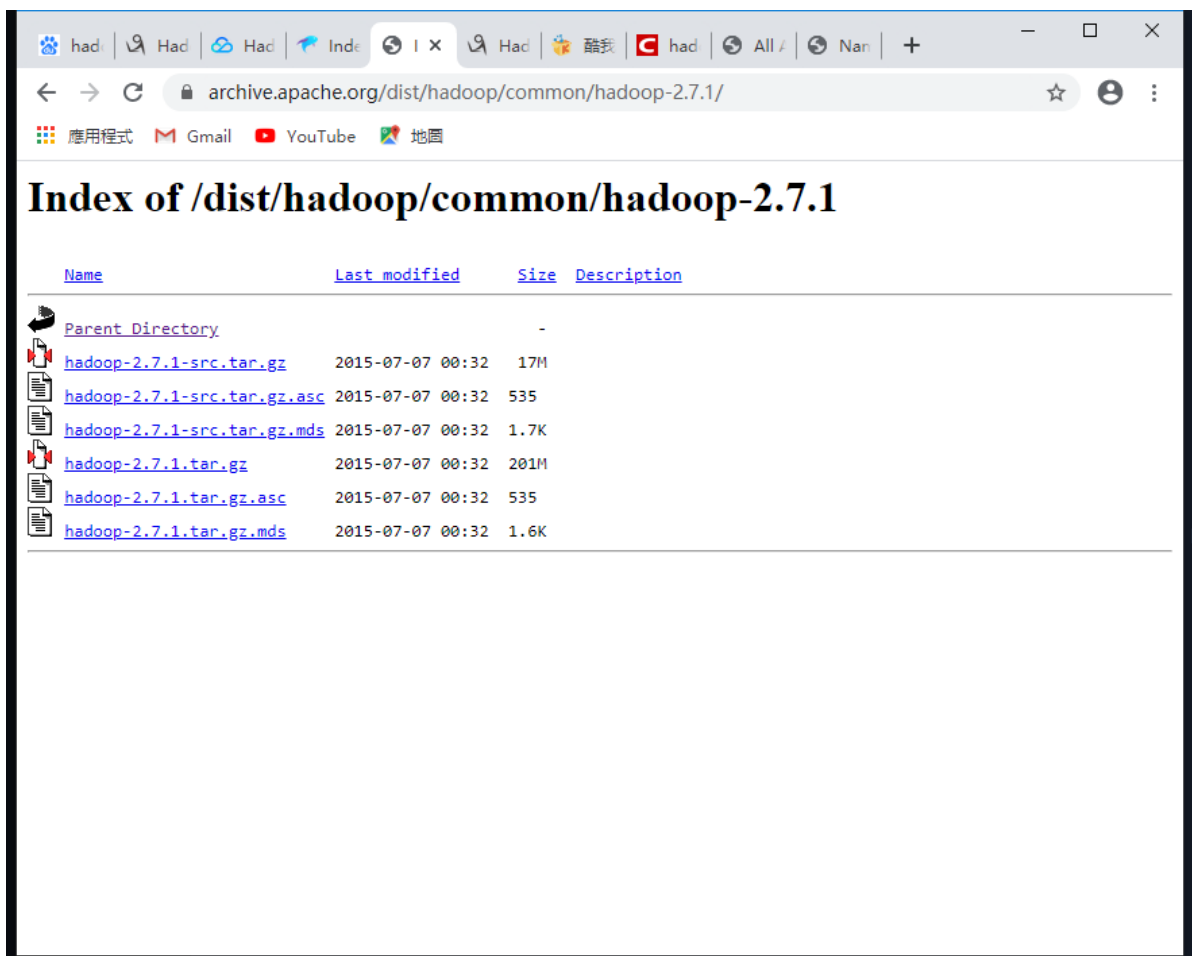
## 2、安装Hadoop

- 1、下载Hadoop镜像安装文件

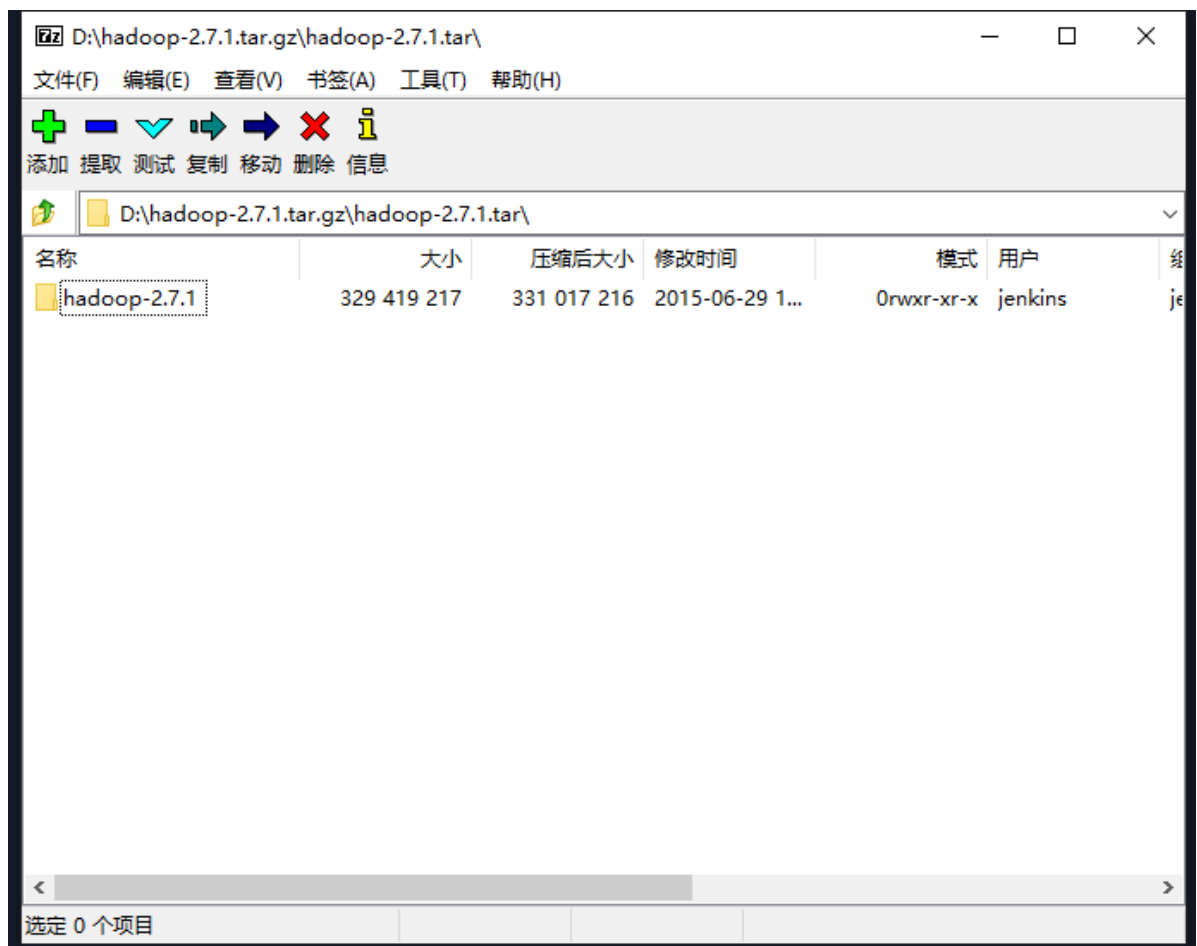
到`<https://archive.apache.org/dist/hadoop/common/>网址下载Hadoop安装包



## 2、下载Hadoop-2.7.1

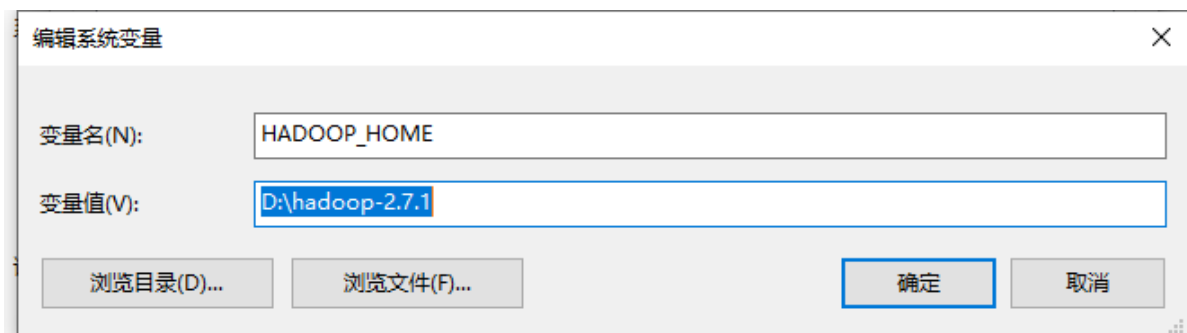


## 3、通过7-zip File Manager的管理员权限打开7-zip工具，将hadoop-2.7.1.tar.gz解压



#### 4、配置Hadoop的环境变量

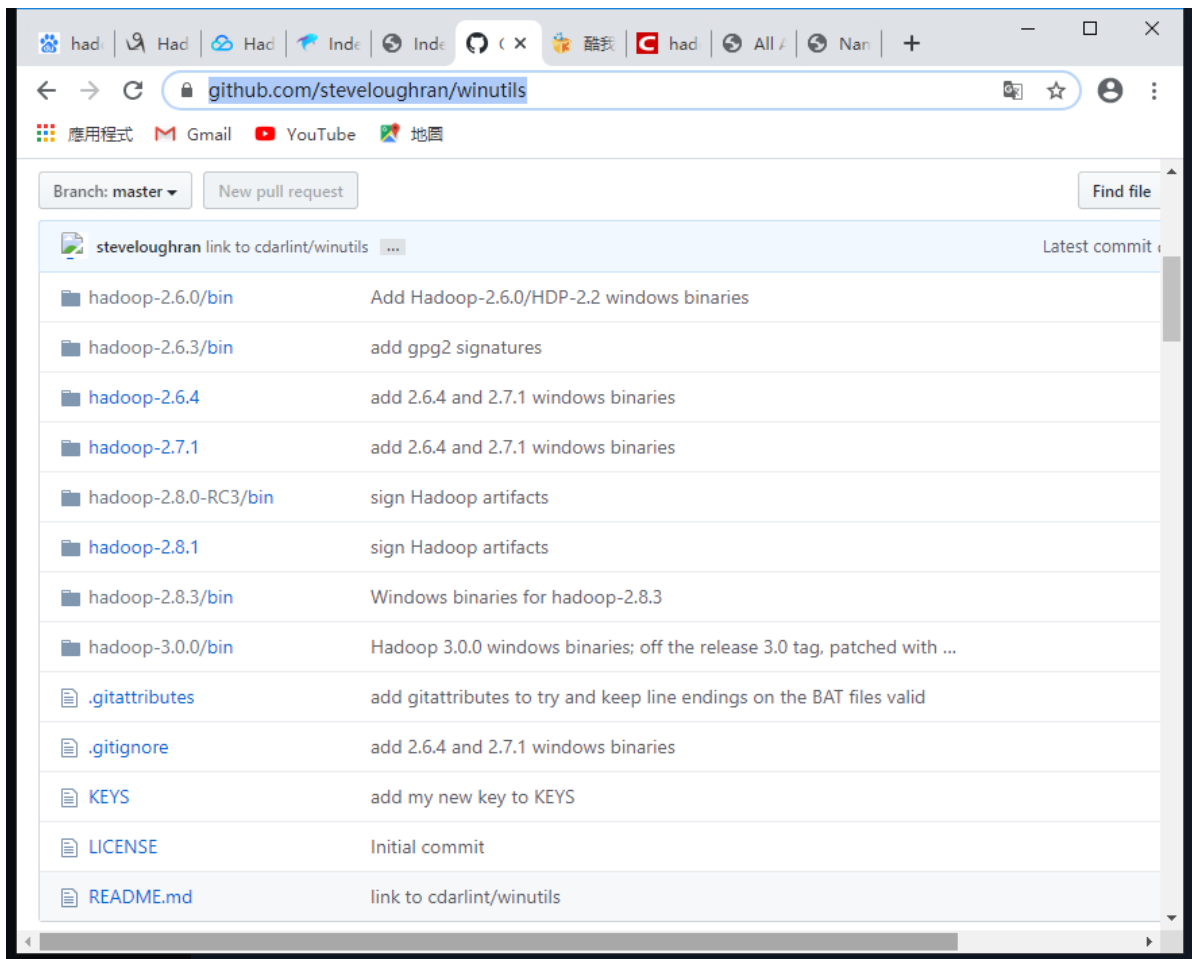
添加HADOOP\_HOME环境变量



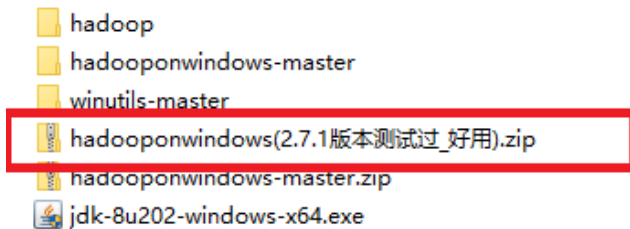
5、将%HADOOP\_HOME%\bin和%HADOOP\_HOME%\sbin添加到PATH环境变量中。

### 3、安装Hadoop在windows专用二进制文件和工具类依赖库winutils和Hadoop.dll文件

1、到<https://github.com/steveloughran/winutils>下载winutils库



2、将winutils文件夹中的hadoop-2.7.1压缩包解压，将解压后的bin目录和etc目录覆盖D:\hadoop-2.7.1\bin文件夹和D:\hadoop-2.7.1\etc文件夹。



名称	修改日期	类型	大小
bin	2018/5/29 15:34	文件夹	
etc	2018/5/29 15:34	文件夹	
README.md	2018/5/29 15:34	Markdown File	1 KB

3、解压，将目录和目录文件替换至<hadoop-2.7.1\bin>目录和<hadoop-2.7.1\etc>目录，修改<etc/hadoop/hdfs-site.xml>文件中datanode和datanode文件路径。

```
<configuration>
  <!-- 这个参数设置为1，因为是单机版hadoop -->
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>/D:/hadoop/workplace/data/namenode</value>
  </property>
```

```
<property>
  <name>dfs.datanode.data.dir</name>
  <value>/D:/hadoop/workplace/data/datanode</value>
</property>
</configuration>
```

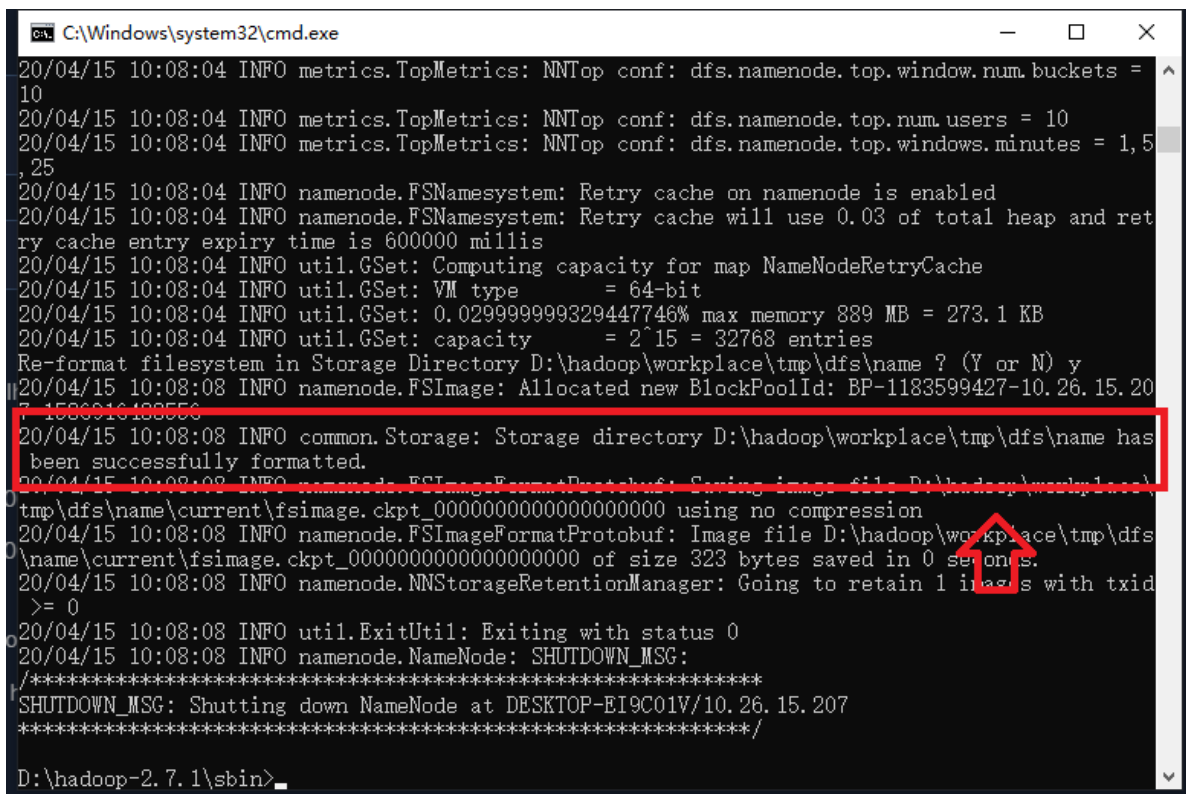
4、修改D:\hadoop-2.7.1\etc\hadoop目录下的hadoop-env.cmd文件，将JAVA\_HOME用@rem注释掉，编辑为JAVA\_HOME的路径，然后保存。

```
@rem The java implementation to use. Required.
set JAVA_HOME=%JAVA_HOME%
```

## 4、启动Hadoop平台

1、运行win+R，启动cmd窗口。

```
hdfs namenode -format
```

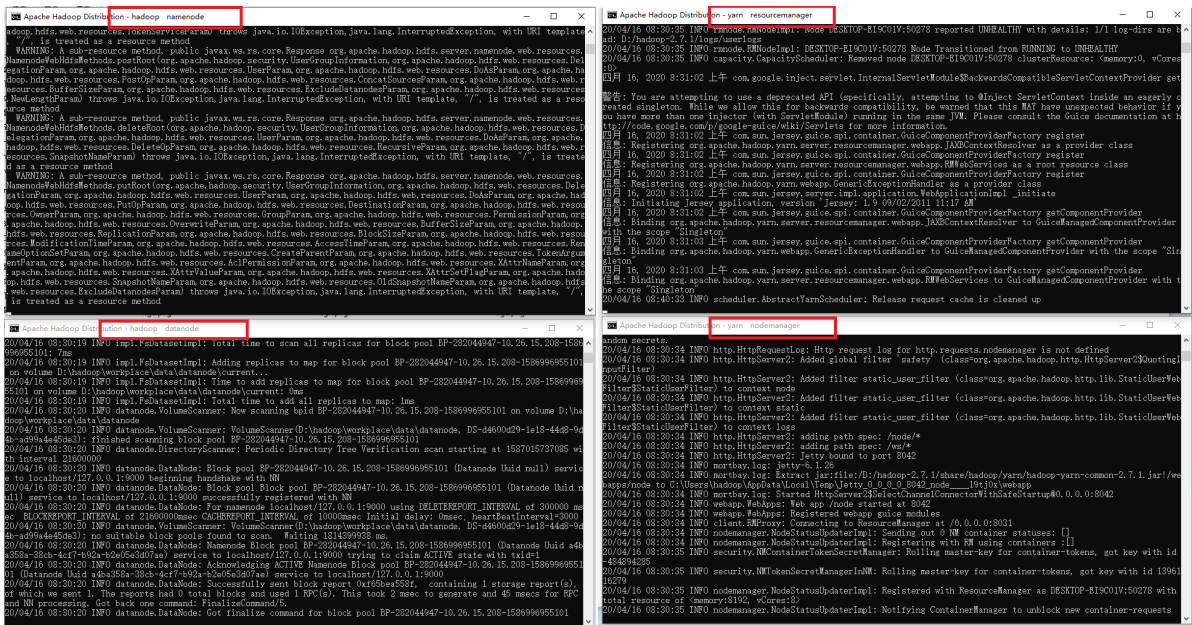


```
C:\Windows\system32\cmd.exe
20/04/15 10:08:04 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
20/04/15 10:08:04 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
20/04/15 10:08:04 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
20/04/15 10:08:04 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
20/04/15 10:08:04 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
20/04/15 10:08:04 INFO util.GSet: Computing capacity for map NameNodeRetryCache
20/04/15 10:08:04 INFO util.GSet: VM type = 64-bit
20/04/15 10:08:04 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
20/04/15 10:08:04 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format filesystem in Storage Directory D:\hadoop\workplace\tmp\dfs\name ? (Y or N) y
20/04/15 10:08:08 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1183599427-10.26.15.207-1586916188556
20/04/15 10:08:08 INFO common.Storage: Storage directory D:\hadoop\workplace\tmp\dfs\name has been successfully formatted.
20/04/15 10:08:08 INFO namenode.FSImageFormatProtobuf: Saving image file D:\hadoop\workplace\tmp\dfs\name\current\fsimage.ckpt_000000000000000000 using no compression
20/04/15 10:08:08 INFO namenode.FSImageFormatProtobuf: Image file D:\hadoop\workplace\tmp\dfs\name\current\fsimage.ckpt_000000000000000000 of size 323 bytes saved in 0 seconds.
20/04/15 10:08:08 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
20/04/15 10:08:08 INFO util.ExitUtil: Exiting with status 0
20/04/15 10:08:08 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at DESKTOP-EI9C01V/10.26.15.207
*****/
D:\hadoop-2.7.1\sbin>
```

格式化命令执行结果成功。

2、切换到hadoop的sbin目录，启动dfs文件系统，启动yarn系统。

```
cd d:\hadoop-2.7.1\sbin
start-dfs.cmd
start-yarn.cmd
```



启动四个Apache Hadoop Distribution窗口。

## 5、打开web控制台GUI

1、打开资源管理web页面 <http://localhost:8088/>

← → ↺

localhost:8088/cluster

应用程式

Gmail

YouTube

地图

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used
0	0	0	0	0	0 B

Scheduler Metrics

Scheduler Type		Scheduling Resource		
Capacity Scheduler		[MEMORY]		
Show 20 entries				
ID	User	Name	Application Type	Queue
Showing 0 to 0 of 0 entries				

2、打开节点管理web页面 <http://localhost:50070/>

← → ↺

localhost:50070/dfshealth.html#tab-overview

☆ 👤 ⋮

應用程式 Gmail YouTube 地圖

Hadoop

Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

# Overview 'localhost:9000' (active)

Started:	Wed Apr 15 10:13:45 CST 2020
Version:	2.7.7, rc1aad84bd27cd79c3d1a7dd58202a8c3ee1ed3ac
Compiled:	2018-07-18T22:47Z by stevel from branch-2.7.7
Cluster ID:	CID-6626b80c-1b63-4077-8f09-8f7a4cf55553
Block Pool ID:	BP-1183599427-10.26.15.207-1586916488556

## Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 63.7 MB of 224.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 38.2 MB of 39.44 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	0 B
DFS Used:	0 B (100%)

## 6、hadoop、hdfs命令使用

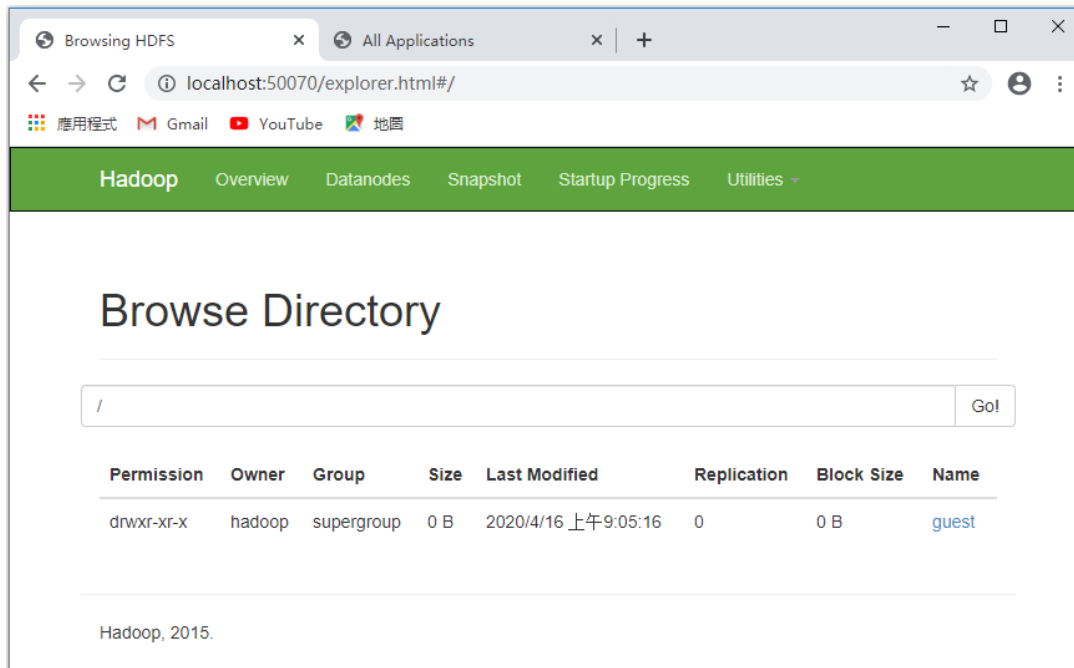
### 1、创建输入目录

```
hadoop fs -mkdir hdfs://localhost:9000/guest/  
hadoop fs -mkdir hdfs://localhost:9000/guest/input
```

### 2、上传数据到Hadoop目录

```
hadoop fs -put D:\Data\data.txt hdfs://localhost:9000/guest/input
```

### 3、通过web页面查看创建的目录和上传的数据



## 7、QA常见问题

```
WARN hdfs.DFSClient: DataStreamer Exception:
org.apache.hadoop.ipc.RemoteException: java.io.IOException: File
/user/root/input/configuration.xml could only be replicated to 0 nodes, instead
of 1
```

- 1、删除 [D:\hadoop\workplace](#) 目录下的data文件夹和tmp文件夹