

# 实验报告

实验名称	实验一 Hadoop 本地运行模式		
实验教室	丹青 922	实验日期	2021 年 5 月 23 日
学 号	2019210173	姓 名	刘思远
专业班级	计算机科学与技术 04 班		
指导教师	卢洋		

东北林业大学  
信息与计算机科学技术实验中心

## 一、 实验目的

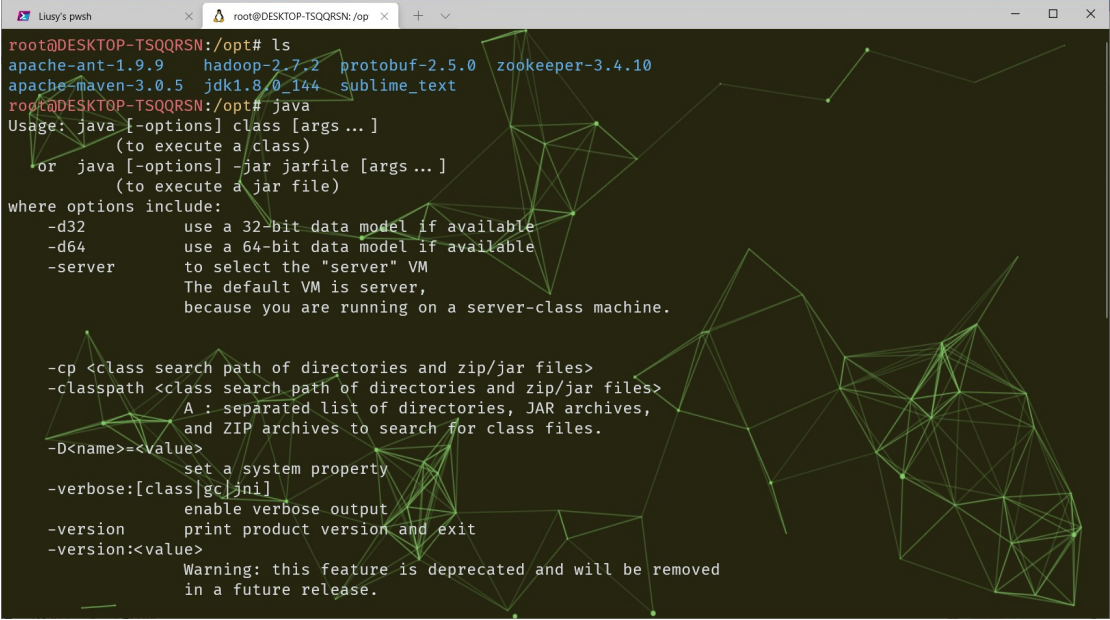
1. 掌握 JDK 的安装和配置；
2. 掌握 Hadoop 本地模式的安装和配置。

## 二、 实验环境

- (1) 计算机的硬件配置 PC 系列微机。
- (2) 计算机的软件配置 VMware 虚拟机软件及 Ubuntu 虚拟机。

## 三、 实验内容及结果

1. 下载安装超星在线课提供的 JDK ， 版本为 1.8.0\_144。



```
root@DESKTOP-TSQQRSN:/opt# ls
apache-ant-1.9.9  hadoop-2.7.2  protobuf-2.5.0  zookeeper-3.4.10
apache-maven-3.0.5  jdk1.8.0_144  sublime_text
root@DESKTOP-TSQQRSN:/opt# java
Usage: java [-options] class [args...]
           (to execute a class)
 or java [-options] -jar jarfile [args...]
           (to execute a jar file)
where options include:
  -d32          use a 32-bit data model if available
  -d64          use a 64-bit data model if available
  -server       to select the "server" VM
                 The default VM is server,
                 because you are running on a server-class machine.

  -cp <class search path of directories and zip/jar files>
  -classpath <class search path of directories and zip/jar files>
                 A : separated list of directories, JAR archives,
                 and ZIP archives to search for class files.
  -D<name>=<value>
                 set a system property
  -verbose:[class|gc|jni]
                 enable verbose output
  -version      print product version and exit
  -version:<value>
                 Warning: this feature is deprecated and will be removed
                 in a future release.
```

2. 验证所安装 Java 的版本；

```
root@DESKTOP-TSQQRN:/opt# java -version
java version "1.8.0_144"
Java(TM) SE Runtime Environment (build 1.8.0_144-b01)
Java HotSpot(TM) 64-Bit Server VM (build 25.144-b01, mixed mode)
root@DESKTOP-TSQQRN:/opt#
```

### 3. 下载安装超星在线课提供的 Hadoop ，版本为 2.7.2 ；

```
root@DESKTOP-TSQQRN:/opt# ls
apache-ant-1.9.9  hadoop-2.7.2  protobuf-2.5.0  zookeeper-3.4.10
apache-maven-3.0.5  jdk1.8.0_144  sublime_text
root@DESKTOP-TSQQRN:/opt# hadoop
Usage: hadoop [--config confdir] [COMMAND | CLASSNAME]
CLASSNAME
  run the class named CLASSNAME
or
where COMMAND is one of:
  fs
    run a generic filesystem user client
  version
    print the version
  jar <jar>
    run a jar file
    note: please use "yarn jar" to launch
    YARN applications, not this command.
  checknative [-a|-h]
    check native hadoop and compression libraries availability
  distcp <srcurl> <desturl>
    copy file or directories recursively
  archive -archiveName NAME -p <parent path> <src>* <dest>
    create a hadoop archive
  classpath
    prints the class path needed to get the
  credential
    interact with credential providers
  Hadoop jar and the required libraries
  daemonlog
    get/set the log level for each daemon
  trace
    view and modify Hadoop tracing settings

Most commands print help when invoked w/o parameters.
root@DESKTOP-TSQQRN:/opt#
```

### 4. 验证所安装 Hadoop 的版本；

```
root@DESKTOP-TSQQRN:/opt# hadoop version
Hadoop 2.7.2
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r b165c4fe8a74265c792ce23f546c64604acf0e41
Compiled by jenkins on 2016-01-26T00:08Z
Compiled with protoc 2.5.0
From source with checksum d0fda26633fa762bfff87ec759ebe689c
This command was run using /opt/hadoop-2.7.2/share/hadoop/common/hadoop-common-2.7.2.jar
```

### 5. 运行 Apache Hadoop 官方提供的 grep 案例；

```
root@DESKTOP-TSQQRN:/opt/hadoop-2.7.2# mkdir input
root@DESKTOP-TSQQRN:/opt/hadoop-2.7.2# cp etc/hadoop/*.xml input
root@DESKTOP-TSQQRN:/opt/hadoop-2.7.2# hadoop jar s
sbin/ share/
root@DESKTOP-TSQQRN:/opt/hadoop-2.7.2# hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar
grep input output 'dfs[a-z.]+'

```

```
Input split bytes=135
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=78
Reduce input records=3
Reduce output records=3
Spilled Records=6
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=88
CPU time spent (ms)=1120
Physical memory (bytes) snapshot=432201728
Virtual memory (bytes) snapshot=374525011488
Total committed heap usage (bytes)=321388544

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=176
File Output Format Counters
Bytes Written=48
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2# |
```

```
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2# ls
LICENSE.txt  README.txt  data  include  lib  logs  sbin  wcininput
NOTICE.txt  bin  etc  input  libexec  output  share
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2# cd output/
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2/output# ls
_SUCCESS  part-r-00000
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2/output# cat p*
1      dfsadmin
1      dfs.replication
1      dfs.http.address
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2/output# |
```

6. 运行 Apache Hadoop 官方提供的 WordCount 案例。

```
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2# ls
LICENSE.txt  README.txt  data  include  lib  logs  sbin  wcininput
NOTICE.txt  bin  etc  input  libexec  output  share
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2# cat wcininput/wc.input
hadoop yarn
hadoop lsy
lsy love lby 4ever
root
root

root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2# hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar
wordcount wcininput wcoutput
```



```
Input split bytes=119
Combine input records=10
Combine output records=7
Reduce input groups=7
Reduce shuffle bytes=84
Reduce input records=7
Reduce output records=7
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=74
CPU time spent (ms)=1180
Physical memory (bytes) snapshot=433127424
Virtual memory (bytes) snapshot=2506623086592
Total committed heap usage (bytes)=320339968

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=53
File Output Format Counters
Bytes Written=50
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2#
```

```
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2# ls
LICENSE.txt  README.txt  data  include  lib  logs  sbin  wcinput
NOTICE.txt  bin  etc  input  libexec  output  share  wcoutput
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2# cd wcoutput/
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2/wcoutput# ls
_SUCCESS  part-r-00000
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2/wcoutput# cat part-r-00000
4ever 1
hadoop 2
lby 1
love 1
lsy 2
root 2
yarn 1
root@DESKTOP-TSQQRSN:/opt/hadoop-2.7.2/wcoutput#
```

#### 四、 实验过程分析与讨论

主要的收获是学会了通过阅读官方文档解决问题，学习了如何安装 JDK 和 Hadoop，如何让配置 JAVA\_HOME 路径并写入 PATH。最后自己成功完成了官网所给的两个案例，Grep 和 WordCount。

#### 五、 指导教师意见

指导教师签字：卢洋