

MPLab1 - 实验报告

曾许墨秋 181240004

18国院数理 曾许墨秋

系统安装情况

系统采用WSL (Windows Subsystem for Linux)，之前装了Linux，发现Linux和华硕的电脑适配不好，而且WSL也便于开发，所以重装电脑后就没有再装Linux了，直接用的WSL。

本次实验采用新建群组hadoop-user，并在群组下添加用户hadoop，然后由root新建文件夹/home/hadoop分配给hadoop用户作为根目录，使用chmod 777赋予Hadoop对该文件夹的权限并加入sudo组。因为已经有较好的环境隔离效果，所以就没有使用docker。

java尝试了jdk13和jdk8，jdk13会有warning，查询了hadoop2.7.1使用的Java版本后采用了jdk8+hadoop2.7.1的配置。

实验数据说明

测试的wordcount文件的话，因为伪单机，也没必要太多太大的文件，测试一下就行，所以选取了hadoop的[Wikipedia](#)和我从我的[php](#)上随便选的两个网页。


实验遇到的问题

1. hadoop-env.sh内本来有 `export JAVA_HOME=${JAVA_HOME}`，应该只要用户的环境变量配置好就行（使用apt安装 `JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64`），但实际上需要重新再写一次，否则虽然Hadoop可以运行，但启动进程后会报错找不到JAVA_HOME，应该是localhost找不到，不清楚为什么。
2. 每次启动WSL后都要 `service ssh restart` 才能连上localhost
3. 一开始没有配置SecondaryNode，配置后jps结果如图1

```
LAPTOP-59V240RG% jps
994  DataNode
1635 NodeManager
4631 Jps
1208 SecondaryNameNode
824  NameNode
1373 ResourceManager
```

图1 java进程 (jps)

实验结果



Application application_1618562610547_0004

Logged in as: dr.who

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Kill Application

Application Overview

User: hadoop

Name: word count

Application Type: MAPREDUCE

Application Tags:

YarnApplicationState: FINISHED

FinalStatus Reported by AM: SUCCEEDED

Started: Fri Apr 16 17:28:55 +0800 2021

Elapsed: 12sec

Tracking URL: History

Diagnostics:

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>

Total Number of Non-AM Containers Preempted: 0

Total Number of AM Containers Preempted: 0

Resource Preempted from Current Attempt: <memory:0, vCores:0>

Number of Non-AM Containers Preempted from Current Attempt: 0

Aggregate Resource Allocation: 51503 MB-seconds, 31 vcore-seconds

Show 20 entries

Search:

Attempt ID	Started	Node	Logs
appattempt_1618562610547_0004_000001	Fri Apr 16 17:28:55 +0800 2021	http://LAPTOP-59V240RG.localdomain:8042	Logs

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

图2 Hadoop Application

Browse Directory

/test_out

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	hadoop	supergroup	0 B	2021/4/16下午5:29:06	1	128 MB	_SUCCESS
-rw-r--r--	hadoop	supergroup	214.84 KB	2021/4/16下午5:29:06	1	128 MB	part-r-00000

Hadoop, 2015.

图3 Word Count 输出文件

```
bllovetx@LAPTOP-59V240RG: /mnt/c/Users/ASUS
!important;      3
!important;border:0      1
!important;height:1px    1
!important;padding:0     1
!important;width:1px     1
"4      1
"<a      1
"All      1
"Benchmarking      1
"Characterization      1
"Distributed      1
"MapReduce:      1
"compatible"      1
"ext.visualEditor.desktopArticleTarget.init","ext.visualEditor.targetloader","ext.eventlogging","ext.wikimediaEvents","ext.navigationTiming","ext.uls.compactlinks","ext.uls.interface","ext.cx.eventlogging.campaigns","ext.centralNotice.geoIP","ext.centralNotice.startUp"]</script>      1
"php"      1
"wgULSCurrentAutonym":"English","wgNoticeProject":"wikipedia","wgCentralAuthMobileDomain":!1,"wgEditSubmitButtonLabelPublish":!0,"wgULSPosition":"interlanguage","wgWikibaseItemId":"Q29120";RLSTATE={"ext.globalCssJs.user.styles":"ready","site.styles":"ready","noscript":"ready","user.styles":"ready","ext.globalCssJs.user":"ready","user":"ready","user.options":"loading","ext.cite.styles":"ready","skins.vector.styles.legacy":"ready","jquery.makeCollapsible.styles":"ready","ext.visualEditor.desktopArticleTarget.noscript":"ready","ext.uls.interlanguage":"ready","ext.wikimediaBadges":"ready","wikibase.client.init":"ready"};RLPAGEMODULES=["ext.cite.ux-enhancements","site","mediawiki.page.ready","jquery.makeCollapsible","mediawiki.toc","skins.vector.legacy.js","ext.gadget.ReferenceTooltips","ext.gadget.charinsert","ext.gadget.extra-toolbar-buttons","ext.gadget.refToolbar","ext.gadget.switcher","ext.centralauth.centralautologin","mmv.head","mmv.bootstrap.autostart","ext.popups","ext.centralNotice.geoIP","ext.centralNotice.startUp"]</script>
"}.mw-parser-output      1
#3366cc,inset      2
#3366cc}.mw-ui-button.mw-ui-progressive:hover{background-color:#447ff5;border-color:#447ff5}.mw-ui-button.mw-ui-progressive:focus{box-shadow:inset      1
#36c;border-radius:0.875em;box-shadow:0 1
#a2a9b1;border-radius:2px;cursor:pointer;vertical-align:middle;font-family:inherit;font-size:1em;font-weight:bold;line-height:1.28571429em;text-align:center;-webkit-appearance:none}.mw-ui-button:visited{color:#202122}.mw-ui-button:hover{background-color:#ffffff;color:#404244;border-color:#a2a9b1}.mw-ui-button:focus{background-color:#ffffff;color:#202122;border-color:#3366cc;box-shadow:inset      1
#a2a9b1;bottom:-8px;left:10px}.mwe-popups.flipped-y:after{content:'';position:absolute;border:11px      1
#a2a9b1;bottom:-9px;left:293px;z-index:111}.mwe-popups.flipped-x-y:after{content:'';position:absolute;border:12px      1
#a2a9b1;bottom:-9px;left:420px}.mwe-popups.flipped-x-y.mwe-popups-is-tall:after{content:'';position:absolute;border:12px      1
#a2a9b1;box-shadow:0      1
#a2a9b1;top:-9px;left:293px}.mwe-popups.mwe-popups-image-pointer.flipped-x:after{content:'';position:absolute;border:12px      1
#a2a9b1;top:-9px;left:420px;z-index:111}.mwe-popups.flipped-x.mwe-popups-is-tall      1
#a2a9b1;top:-9px;left:9px;z-index:111}.mwe-popups.mwe-popups-image-pointer:after{content:'';position:absolute;border:12px      1
#aaa      1
#c8ccd1;      4
#c8ccd1;border-radius:3px;box-shadow:0 1
#c8ccd1;box-shadow:0      1
#c8ccd1;opacity:0.8}.uls-icon-back:hover{opacity:1;cursor:pointer}.uls-menu      1
1,1      Top
```

图4 Word Count 结果

Summary

Security is off.
Safemode is off.
28 files and directories, 13 blocks = 41 total filesystem object(s).
Heap Memory used 138.94 MB of 351 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 56.86 MB of 57.88 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	250.98 GB
DFS Used:	1.07 MB (0%)
Non DFS Used:	15.42 GB
DFS Remaining:	235.56 GB (93.85%)
Block Pool Used:	1.07 MB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0

图5 Hadoop 进程 Summary

实验体会

1. Hadoop 功能非常完备，包括用于备份与特殊情况处理的SecondaryNode、一系列只要配置好，启动Hadoop就会开始监听的端口及相应的网页，甚至还有可视化的文件系统和状态监控界面
2. 也许是应为功能比较多，所有功能集中在localhost上仅仅只是启动就很占用电脑资源，不清楚是否有WSL的因素在内。