期末实验实验报告

陈朴炎 2021211138

目录

1 给分点截图

1.1 local 模式部署安装

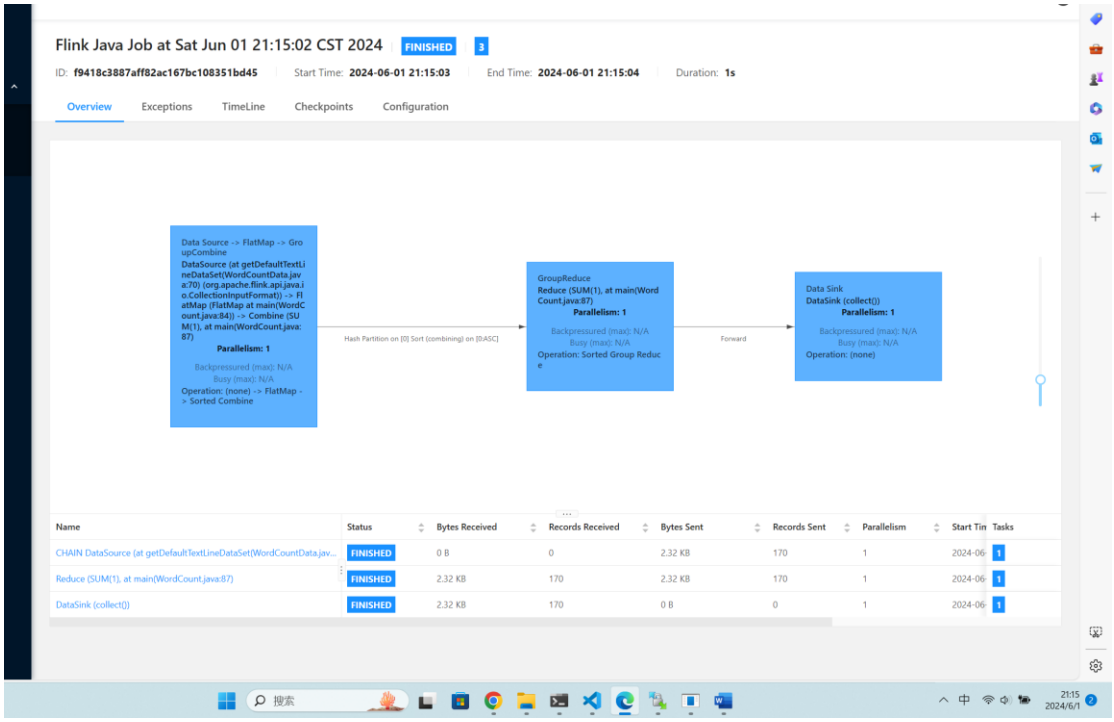
```
[root@cpy-2021211138 ~]# jps
2208 StandaloneSessionClusterEntrypoint
2529 Jps
2473 TaskManagerRunner
[root@cpy-2021211138 ~]# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.0.30  netmask 255.255.255.0  broadcast 192.168.0.255
        inet6 fe80::f816:3eff:fef5:4d6b  prefixlen 64  scopeid 0x20<link>
        ether fa:16:3e:f5:4d:6b  txqueuelen 1000  (Ethernet)
        RX packets 259332  bytes 387920380 (369.9 MiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 156329  bytes 10485200 (9.9 MiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
        inet 127.0.0.1  netmask 255.0.0.0
        inet6 ::1  prefixlen 128  scopeid 0x10<host>
        loop  txqueuelen 1000  (Local Loopback)
        RX packets 107  bytes 86181 (84.1 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 107  bytes 86181 (84.1 KiB)
```

主节点 jps 及 ifconfig 信息

终端单词输入截图，终端输出 out 文件查看单词统计结果

```
[root@cpy-2021211138 ~]# http://1.92.114.12:8081/#/overview
-bash: http://1.92.114.12:8081/#/overview: No such file or directory
[root@cpy-2021211138 ~]# nc -lk 9000
cpy cpy cpy pcy pcy pyc pyc abc 12456 3454564 123456 123456 123456 cpy ypc cyp pyypypypy 2021211138 2021211138 138 138 1
38
```

```
[root@cpy-2021211138 log]# tail -200f flink-root-taskexecutor-0-cpy-2021211138.out
cpy : 4
 : 1
138 : 3
2021211138 : 2
pyypypypy : 1
cyp : 1
ypc : 1
123456 : 3
3454564 : 1
12456 : 1
abc : 1
pyc : 2
pcy : 2
 : 3
```



1.2 standalone 模式部署安装

主节点的 jps：

```
[root@cpy-2021211138 modules]# jps
12123 Jps
12013 TaskManagerRunner
11713 StandaloneSessionClusterEntrypoint
[root@cpy-2021211138 modules]# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.0.30  netmask 255.255.255.0  broadcast 192.168.0.255
        inet6 fe80::f816:3eff:fef5:4d6b  prefixlen 64  scopeid 0x20<link>
        ether fa:16:3e:f5:4d:6b  txqueuelen 1000  (Ethernet)
        RX packets 69459  bytes 5297091 (5.0 MiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 59855  bytes 1077058398 (1.0 GiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

从节点的 jps：

node2：

```
[root@cpy-2021211138 ~]# jps
1968 Jps
1888 TaskManagerRunner
[root@cpy-2021211138 ~]# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.0.213  netmask 255.255.255.0  broadcast 192.168.0.255
        inet6 fe80::f816:3eff:fef5:4d22  prefixlen 64  scopeid 0x20<link>
        ether fa:16:3e:f5:4d:22  txqueuelen 1000  (Ethernet)
        RX packets 259244  bytes 372410061 (355.1 MiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 22928  bytes 1798836 (1.7 MiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

node3：

```
[root@cpy-2021211138 ~]# jps
1898 TaskManagerRunner
1978 Jps
[root@cpy-2021211138 ~]# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.0.161  netmask 255.255.255.0  broadcast 192.168.0.255
        inet6 fe80::f816:3eff:fef5:4dee  prefixlen 64  scopeid 0x20<link>
        ether fa:16:3e:f5:4d:ee  txqueuelen 1000  (Ethernet)
        RX packets 256712  bytes 372242342 (354.9 MiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 20954  bytes 1702853 (1.6 MiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

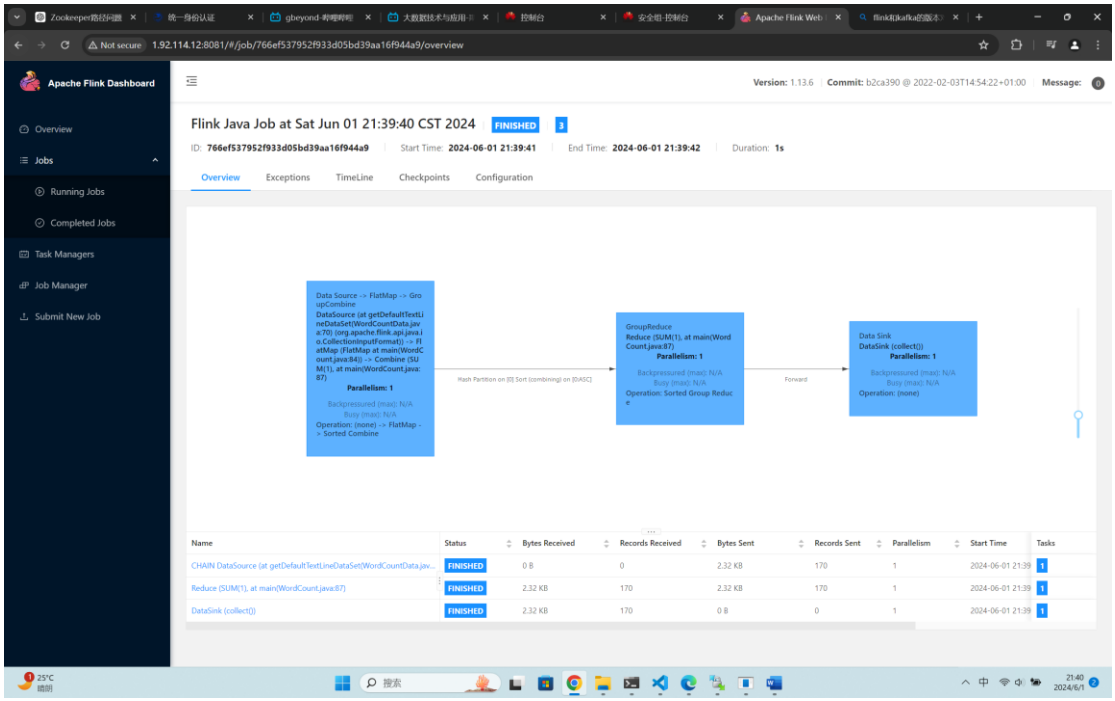
node4：

```
[root@cpy-2021211138 ~]# jps
1906 TaskManagerRunner
1986 Jps
[root@cpy-2021211138 ~]# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.0.135  netmask 255.255.255.0  broadcast 192.168.0.255
        inet6 fe80::f816:3eff:fef5:4dd4  prefixlen 64  scopeid 0x20<link>
        ether fa:16:3e:f5:4d:d4  txqueuelen 1000  (Ethernet)
        RX packets 255926  bytes 372193478 (354.9 MiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 20321  bytes 1668116 (1.5 MiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

运行测试用例：

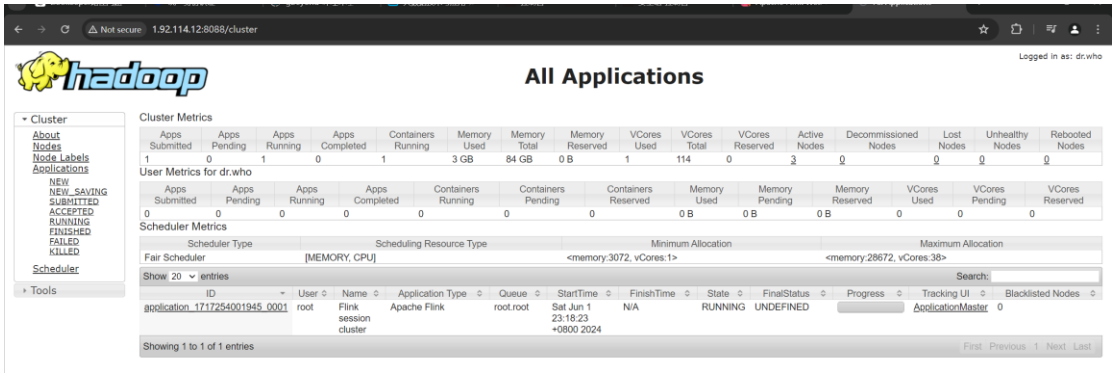
```
[root@cpy-2021211138 modules]# cd flink
[root@cpy-2021211138 flink]# bin/flink run examples/batch/WordCount.jar
Executing WordCount example with default input data set.
Use --input to specify file input.
Printing result to stdout. Use --output to specify output path.
Job has been submitted with JobID 766ef537952f933d05bd39aa16f944a9
Program execution finished
Job with JobID 766ef537952f933d05bd39aa16f944a9 has finished.
Job Runtime: 1072 ms
Accumulator Results:
- f96fa2439761acf45d1db4a102d1fe57 (java.util.ArrayList) [170 elements]


(a,5)
(action,1)
(after,1)
(against,1)
(all,2)
(and,12)
(arms,1)
(arrows,1)
(awry,1)
(ay,1)
```

查看 web 管理页面的 Completed job：



1.3 Flink on Yarn

给出 Yarn 的 Web 界面



这里我跟着实验指导手册，执行了 example 程序，如下所示：

```
[root@cpy-2021211138 flink]# bin/flink run examples/batch/WordCount.jar
2024-06-01 23:31:01,075 INFO  org.apache.flink.yarn.cli.FlinkYarnSessionCli                 [] - Found Yarn properties file under /tmp/.yarn-properties-root.
2024-06-01 23:31:01,075 INFO  org.apache.flink.yarn.cli.FlinkYarnSessionCli                 [] - Found Yarn properties file under /tmp/.yarn-properties-root.
Executing WordCount example with default input data set.
Use --input to specify file input.
Printing result to stdout. Use --output to specify output path.
2024-06-01 23:31:01,549 WARN  org.apache.flink.yarn.configuration.YarnLogConfigUtil        [] - The configuration directory ('/home/modules/flink/conf') already co
ntains a LOG4J config file.If you want to use logback, then please delete or rename the log configuration file.
2024-06-01 23:31:01,644 INFO  org.apache.hadoop.yarn.client.RMProxy                         [] - Connecting to ResourceManager at node1/192.168.0.30:8032
2024-06-01 23:31:01,795 INFO  org.apache.flink.yarn.YarnClusterDescriptor                   [] - No path for the flink jar passed. Using the location of class org.a
pache.flink.yarn.YarnClusterDescriptor to locate the jar
2024-06-01 23:31:01,864 INFO  org.apache.flink.yarn.YarnClusterDescriptor                   [] - Found Web Interface node2:44611 of application 'application_1717254
001945_0002'.
Job has been submitted with JobID 8049d44977529ffd49f5a05774aceecf
Program execution finished
Job with JobID 8049d44977529ffd49f5a05774aceecf has finished.
Job Runtime: 12869 ms
Accumulator Results:
- 70db26b72f0673f0edf77d3edaf7dbf9 (java.util.ArrayList) [170 elements]


(a,5)
(action,1)
(after,1)
(against,1)
(all,2)
(and,12)
(arms,1)
(arrows,1)
```

```
(turn,1)
(under,1)
(undiscover,1)
(unworthy,1)
(us,3)
(we,4)
(weary,1)
(what,1)
(when,2)
(whether,1)
(whips,1)
(who,2)
(whose,1)
(will,1)
(wish,1)
(with,3)
(would,2)
(wrong,1)
(you,1)
```

运行自己的输入：

```
[root@cpy-2021211138 flink]# bin/flink run examplex/batch/WordCount.jar -inputhdfs://node1:8020/flink_input -output hdfs://node1:8020
flink_coutput/wordcount-result.txt
2024-06-01 23:33:01,626 INF0 org.apache.flink.yarn.cli.FlinkYarnSessioncli                - Found Yarn properties file under /tmp/.
yarn-properties-root.
2024-06-01 23:33:01,626 INF0 org.apache.flink.yarn.cli.FlinkYarnSessioncli                - Found Yarn properties file under /tmp/.
yarn-properties-root.
2024-06-01 23:33:01,980 INFO org.apache.flink.yarn.cli.FlinkYarnSessioncli                - YARN properties set default parallelism
to 2
2024-06-01 23:33:01,980 INFO org.apache.flink.yarn.cli.FlinkYarnSessionCli                - YARN properties set default parallelism
to 2
YARN properties set default parallelism to 2
2024-06-01 23:33:01,059 INF0 org.apache.hadoop.yarn.client.RMProxy                        - Connecting to ResourceManager at node1/
192.168.0.30:8032
2024-06-01 23:33:02,162 INF0 org.apache.flink.yarn.cli.FlinkYarnSessioncli                - No path for the flink jar passed. Using
the location of class org.apache.flink.yarn.YarnclusterDescriptor to locate the jar
2024-06-01 23:33:02,162 INF0 org.apache.flink.yarn.cli.FlinkYarnSessioncli                - No path for the flink jar passed. Using
the location of class org.apache.flink.yarn.YarnclusterDescriptor to locate the jar
2024-06-01 23:33:02,165 WARN org.apache.flink.yarn.AbstractYarnclusterDescriptor          - Neither the HADOOP_CONF_DIR nor the
YARN_CONF_DIR environment variable is set. The Flink YARN Client needs one of these to be set to properly load the Hadoop
configuration for accessing YARN.
2024-06-01 23:33:02,217 INF0 org.apache.flink.yarn.AbstractYarnclusterDescriptor          - Found application Job Manager host name
'node2' and port '44739' from supplied application id 'application 1717254001945 0003'
Starting execution of program
Program execution finished
Job with JobID ea0b4f433a7dc8ec8c86deeb514b68e has finished.
Job Runtime: 13538 ms
```

查看 hdfs 文件 “

```
[root@cpy-2021211138 flink]# hdfs dfs -cat /flink_output/wordcount-result.txt
24/06/01 23:34:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
flink 1
cpy 1
hello 1
test 1
spark 1
world 1
```

1.4 Flink 消费 Kafka 数据

四个节点 jps 出现 Kafka 进程截图

这是主节点的 jps：

```
PS C:\Users\20531> ssh root@1.92.114.12
root@1.92.114.12's password:
Last login: Sun Jun  2 08:59:35 2024 from 117.129.58.146

       Welcome to Huawei Cloud Service

[root@cpy-2021211138 ~]# jps
18164 QuorumPeerMain
18585 Jps
18242 Kafka
[root@cpy-2021211138 ~]#
```

node2 的 jps：

```
PS C:\Users\20531> ssh root@120.46.149.118
root@120.46.149.118's password:
Last login: Sun Jun  2 09:37:37 2024 from 117.129.58.146

       Welcome to Huawei Cloud Service

[root@cpy-2021211138 ~]# jps
4742 Kafka
4633 QuorumPeerMain
5099 Jps
```

node3 的 jps：

```
PS C:\Users\20531> ssh root@1.92.86.3
root@1.92.86.3's password:
Last login: Sun Jun  2 08:59:04 2024 from 117.129.58.146

       Welcome to Huawei Cloud Service

[root@cpy-2021211138 ~]# jps
4577 QuorumPeerMain
4979 Jps
4641 Kafka
```

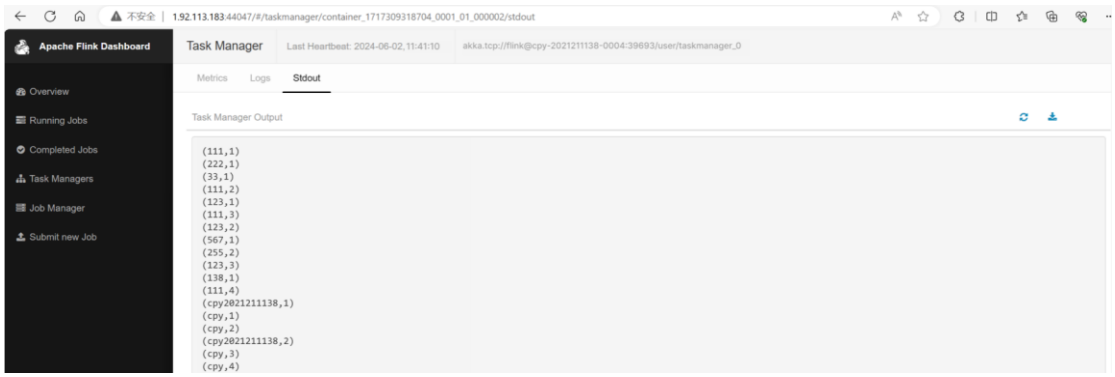
node4 的 jps：

```
PS C:\Users\20531> ssh root@120.46.87.42
root@120.46.87.42's password:
Last login: Sun Jun  2 08:59:16 2024 from 1.92.114.12

       Welcome to Huawei Cloud Service

[root@cpy-2021211138 ~]# jps
4454 QuorumPeerMain
4511 Kafka
4849 Jps
```

启动生产者、运行 jar 包、Web 界面输出结果截图。

```
[root@cpy-2021211138 kafka]# ./bin/kafka-topics.sh --create --zookeeper node1:2181 --replication-factor 1 --partitions 1 --topic wordsendertest
Created topic "wordsendertest".
[root@cpy-2021211138 kafka]# ./bin/kafka-topics.sh --create --zookeeper node1:2181 --replication-factor 1 --partitions 1 --topic test
Created topic "test".
[root@cpy-2021211138 kafka]# kafka-console-producer.sh --broker-list node1:9092 --topic test
>
>
```

```
[root@cpy-2021211138-0001 ~]# flink run -c Flink_Kafka WordCount.jar
2024-06-02 11:39:25,626 INFO org.apache.flink.yarn.cli.FlinkYarnSessioncli                - Found Yarn properties file under /tmp/.yarn-properties-root.
2024-06-02 11:39:25,626 INFO org.apache.flink.yarn.cli.FlinkYarnSessioncli                - Found Yarn properties file under /tmp/.yarn-properties-root.
2024-06-02 11:39:25,980 INFO org.apache.flink.yarn.cli.FlinkYarnSessioncli                - YARN properties set default parallelism to 2
2024-06-02 11:39:25,980 INFO org.apache.flink.yarn.cli.FlinkYarnSessionCli                - YARN properties set default parallelism to 2
YARN properties set default parallelism to 2
2024-06-02 11:39:25,059 INFO org.apache.hadoop.yarn.client.RMProxy                        - Connecting to ResourceManager at cpy-2021211138-0003/192.168.0.128:8032
2024-06-02 11:39:25,162 INFO org.apache.flink.yarn.cli.FlinkYarnSessioncli                - No path for the flink jar passed. Using the location of class org.apache.flink.yarn.YarnclusterDescriptor to locate the jar
2024-06-02 11:39:25,162 INFO org.apache.flink.yarn.cli.FlinkYarnSessioncli                - No path for the flink jar passed. Using the location of class org.apache.flink.yarn.YarnclusterDescriptor to locate the jar
2024-06-02 11:39:25,165 WARN org.apache.flink.yarn.AbstractYarnclusterDescriptor          - Neither the HADOOP_CONF_DIR nor the YARN_CONF_DIR environment variable is set. The Flink YARN Client needs one of these to be set to properly load the Hadoop configuration for accessing YARN.
2024-06-02 11:39:25,217 INFO org.apache.flink.yarn.AbstractYarnclusterDescriptor          - Found application Job Manager host name 'cpy-2021211138-0003' and port '44739' from supplied application id 'application 1717309318704 0003'
Starting execution of program
```



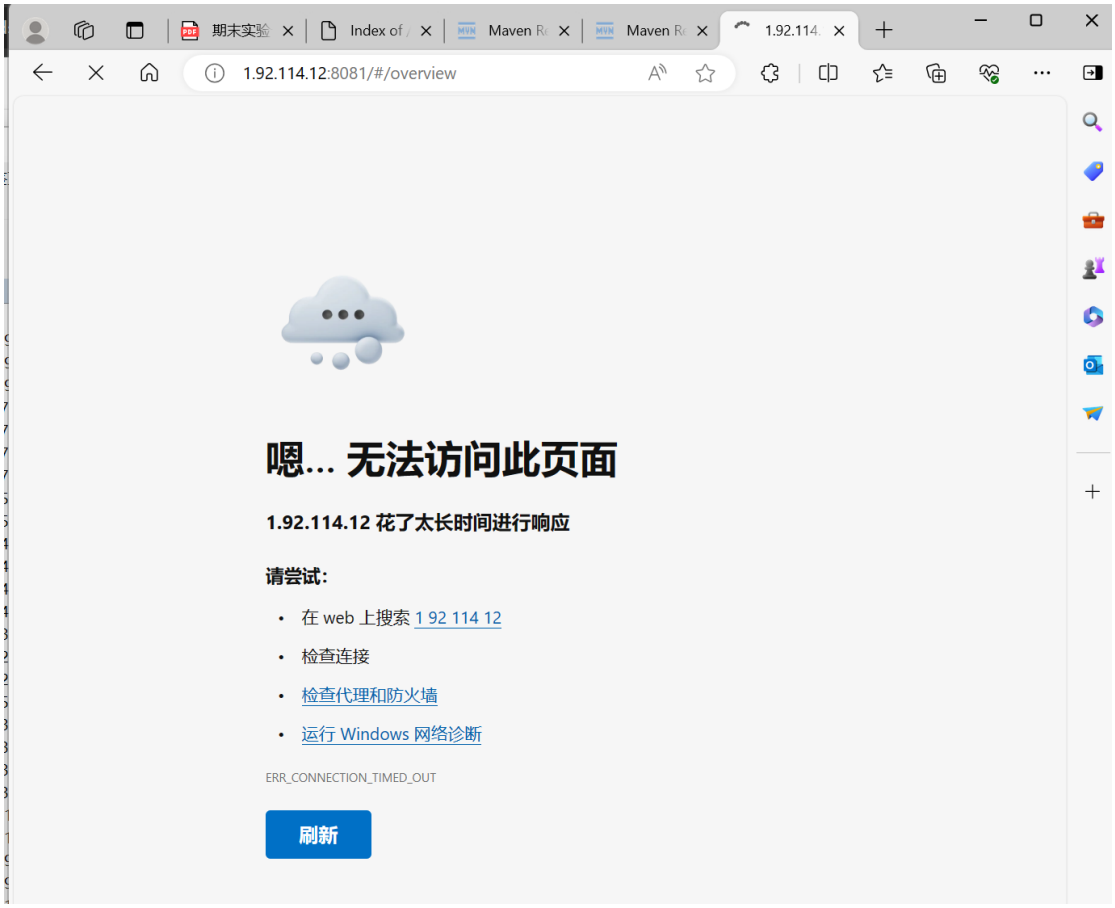
2 实验 bug

2.1 web 访问管理界面

Web 界面访问

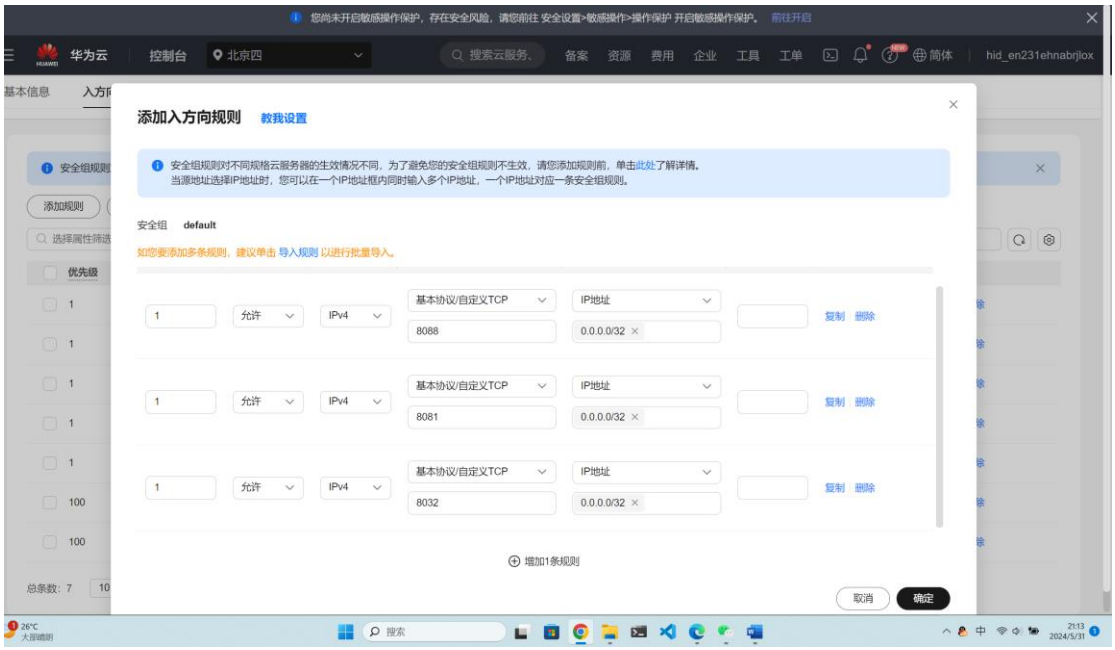
http://1.92.114.12:8081/#/overview

问题：



访问不了。

原因如下：



我的允许访问的 ip 地址网关设置为了/32，但如果是要让所有 ip 都能访问到这台服务器的 8081 端口的话，应该让 IP 地址和网关设置为 0.0.0.0/0，才可以。

修改如下：

之后就能正常访问了

2.2 运行测试用例报错

运行 Flink 自带的测试用例：

安装 nc 工具：

```
sudo yum -y install nc
```

遇到问题：

```
-bash: http://node1:8081/#/overview: No such file or directory
[root@cpy-2021211138 ~]# http://1.92.114.12:8081/#/overview
-bash: http://1.92.114.12:8081/#/overview: No such file or directory
[root@cpy-2021211138 ~]# sudo yum -y install nc
Loaded plugins: fastestmirror
```

```
Metadata Cache Created
[root@cpy-2021211138 ~]# sudo yum -y install nc
Loaded plugins: fastestmirror
[root@cpy-2021211138 ~]# sudo yum -y install nc
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
Resolving Dependencies
--> Running transaction check
---> Package nmap-ncat.aarch64 2:6.40-19.el7 will be installed
--> Finished Dependency Resolution

Dependencies Resolved
```

原因是，从 word 复制的命令 -y 前面这个‘-’是破折号，而命令行里需要用到的是连字符，所以要自己敲命令。

2.3 安全模式忘记去除

在实验三中，由于我忘记取出 hadoop 的安全模式，导致 flink 集群重启失败。

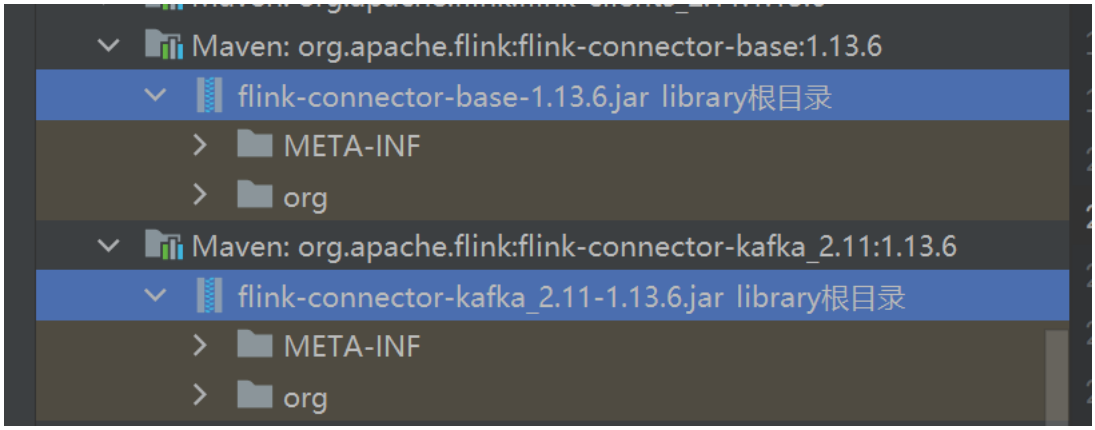
2.4 class no found

出现报错：

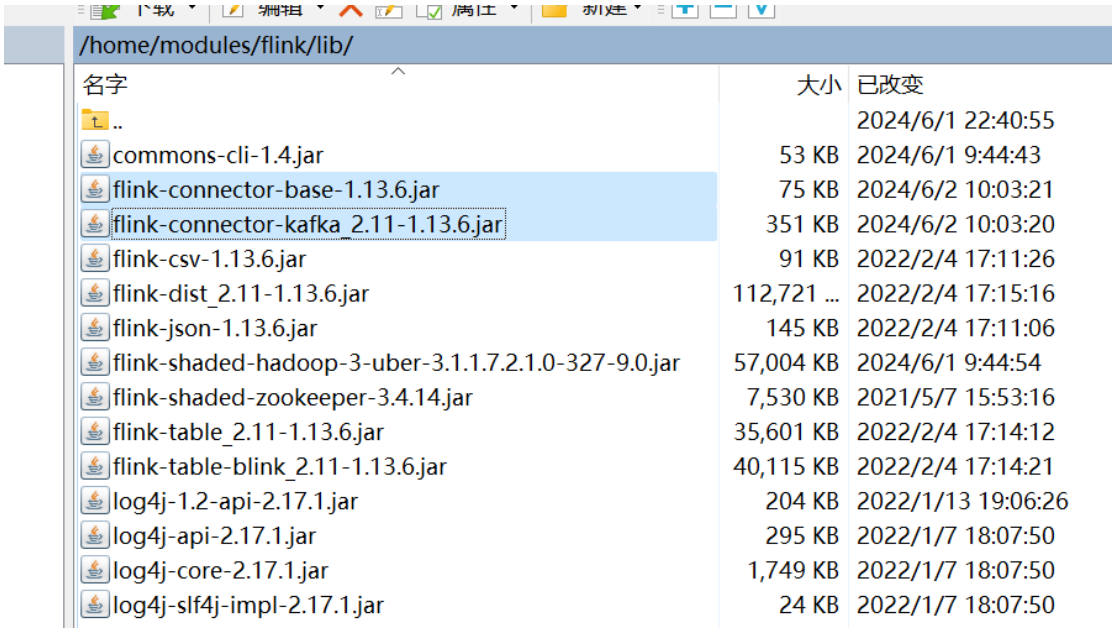
```
[root@cpy-2021211138 ~]# flink run -c Flink_Kafka WordCount.jar
2024-06-02 11:39:25,308 INFO  org.apache.flink.yarn.cli.FlinkYarnSessionCli                 [] - Found Yarn
properties file under /tmp/.yarn-properties-root.
2024-06-02 11:39:25,308 INFO  org.apache.flink.yarn.cli.FlinkYarnSessionCli                 [] - Found Yarn
properties file under /tmp/.yarn-properties-root.
java.lang.NoClassDefFoundError: org.apache.flink.streaming.connectors.kafka.FlinkKafkaConsumer
        at Flink_Kafka.main(Flink_Kafka.java:31)
        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.flink.client.program.PackagedProgram.callMainMethod(PackagedProgram.java:355)
        at org.apache.flink.client.program.PackagedProgram.invokeInteractiveModeForExecution(PackagedProgra
m.java:222)
        at org.apache.flink.client.ClientUtils.executeProgram(ClientUtils.java:114)
        at org.apache.flink.client.cli.CliFrontend.executeProgram(CliFrontend.java:812)
        at org.apache.flink.client.cli.CliFrontend.run(CliFrontend.java:246)
        at org.apache.flink.client.cli.CliFrontend.parseAndRun(CliFrontend.java:1054)
        at org.apache.flink.client.cli.CliFrontend.lambda$main$10(CliFrontend.java:1132)
```

解决：

在 maven 工程中的外部库中找到 FlinkKafkaConsumer 类和 FlinkKafkaConsumerBase 类，并将这两个 jar 包上传到服务器的 flink 下的 lib 中



上传：



3 实验总结

本次实验我花了非常久的时间才成功完成。我大概花了大半天时间将前面三个实验做完了。但是到第四个实验的时候，出现了很大的问题。一开始，我的 Flink 集群能够成功启动，但是每当运行我自己写的 jar 包时，就会有一些报错，告诉我什么什么类找不到了，不过这个类并不是指导书中出现的那个类。后面我把服务器重启，结果发现连 flink 都启动不了了。我在网上查了很久的资料，也尝试了很多办法，都没办法让 flink 重启。最后我把原来的四台服务器都扔了，重新建了四台服务器，重新配置 hadoop、zookeeper、flink、kafka，

重新做了实验四，这才成功。在这个实验中，我对原来的 hadoop、zookeeper、hdfs、flink、kafka 的配置都重新回顾了一遍，有了更深的认识。同时我也对那些配置文件有了更深刻的理解。