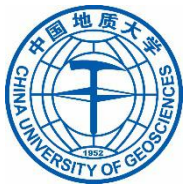


中国地质大学（北京）

《数据库小学期》

Storm 实验报告

学　　院：信息工程学院

专　　业：计算机科学与技术

班　　级：10041811、10041812

指导老师：孙大为

日　　期：2021 年 7 月 6 日

成　　员：如　下

学　号：1005183121　　姓　名：周子杰

学　号：1004181221　　姓　名：汪　航

学　号：1010183115　　姓　名：冯锦元

学　号：1004181222　　姓　名：李　可

学　号：1004181223　　姓　名：孙郡忆

一、 实验内容

Storm 的配置

- JDK 的安装与配置
- ZooKeeper 的安装与配置
- Storm 的安装与配置

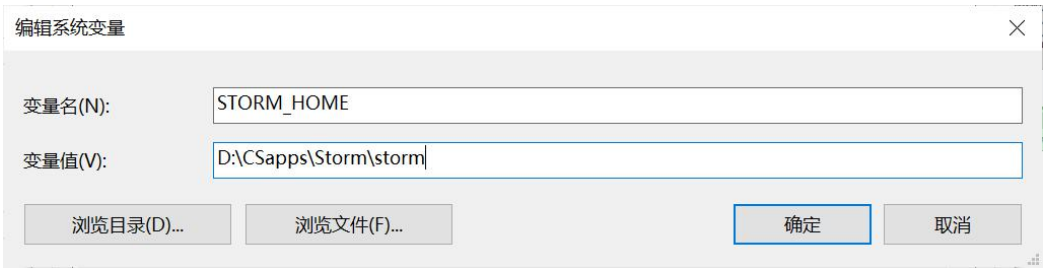
二、 实现方法

JDK 的安装与配置，这个在很早以前就已经配置过了，由于最新版本的 java 直接一键安装即可，故这个部分非常简单。

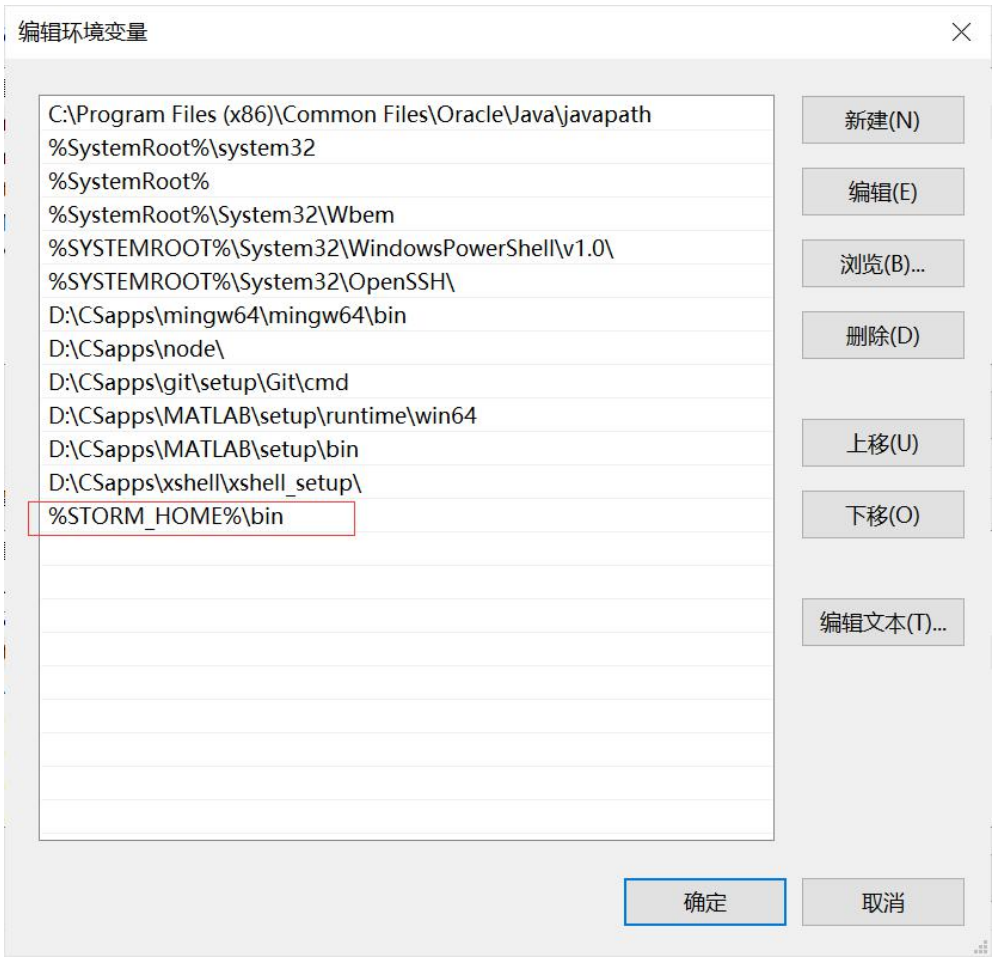
```
C:\Users\24763>java -version
java version "1.8.0_192"
Java(TM) SE Runtime Environment (build 1.8.0_192-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.192-b12, mixed mode)
```

设置环境变量：

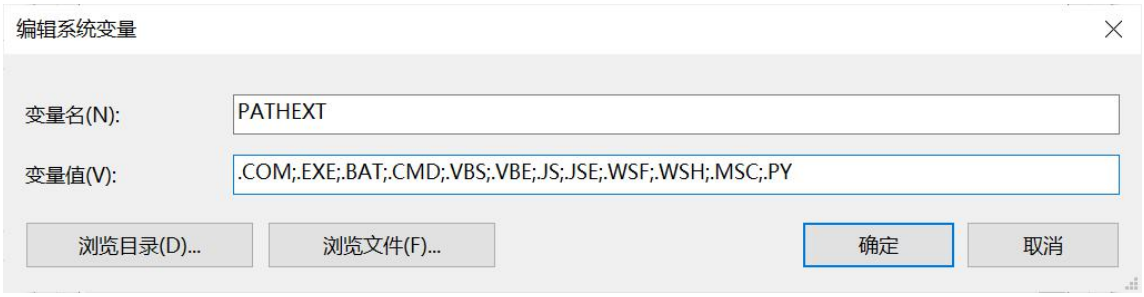
- 新增环境变量 STORM_HOME 并设置为 D:\apache-storm-1.2.2



- 在 Path 里新增%STORM_HOME%\bin



➢ 在 PATHEXT 路径中加入.PY



➢ 在 storm 下找到 E\conf storm.yaml 将其内容复制为：

```
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements.  See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership.  The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License.  You may obtain a copy of the License at
```

```
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

########### These MUST be filled in for a storm configuration
# storm.zookeeper.servers:
#         - "server1"
#         - "server2"
storm.zookeeper.servers:
      - "127.0.0.1"
#
# nimbus.seeds: ["host1", "host2", "host3"]
nimbus.seeds: ["127.0.0.1"]
storm.local.dir: "D:\\storm-local\\data3"
supervisor.slots.ports:
      - 6700
      - 6701
      - 6702
      - 6703
#
#
# ##### These may optionally be filled in:
#
## List of custom serializations
# topology.kryo.register:
#         - org.mycompany.MyType
#         - org.mycompany.MyType2: org.mycompany.MyType2Serializer
#
## List of custom kryo decorators
# topology.kryo.decorators:
#         - org.mycompany.MyDecorator
#
## Locations of the drpc servers
# drpc.servers:
#         - "server1"
#         - "server2"

## Metrics Consumers
## max.retain.metric.tuples
```

```
## - task queue will be unbounded when max.retain.metric.tuples is equal or less than 0.
## whitelist / blacklist
## - when none of configuration for metric filter are specified, it'll be treated as 'pass all'.
## - you need to specify either whitelist or blacklist, or none of them. You can't specify both
of them.
## - you can specify multiple whitelist / blacklist with regular expression
## expandMapType: expand metric with map type as value to multiple metrics
## - set to true when you would like to apply filter to expanded metrics
## - default value is false which is backward compatible value
## metricNameSeparator: separator between origin metric name and key of entry from map
## - only effective when expandMapType is set to true
# topology.metrics.consumer.register:
#     - class: "org.apache.storm.metric.LoggingMetricsConsumer"
#       max.retain.metric.tuples: 100
#       parallelism.hint: 1
#     - class: "org.mycompany.MyMetricsConsumer"
#       max.retain.metric.tuples: 100
#       whitelist:
#           - "execute.*"
#           - "^__complete-latency$"
#       parallelism.hint: 1
#       argument:
#           - endpoint: "metrics-collector.mycompany.org"
#       expandMapType: true
#       metricNameSeparator: "."

## Cluster Metrics Consumers
# storm.cluster.metrics.consumer.register:
#     - class: "org.apache.storm.metric.LoggingClusterMetricsConsumer"
#     - class: "org.mycompany.MyMetricsConsumer"
#       argument:
#           - endpoint: "metrics-collector.mycompany.org"
#
# storm.cluster.metrics.consumer.publish.interval.secs: 60

# Event Logger
# topology.event.logger.register:
#     - class: "org.apache.storm.metric.FileBasedEventLogger"
#     - class: "org.mycompany.MyEventLogger"
#       arguments:
#           endpoint: "event-logger.mycompany.org"

# Metrics v2 configuration (optional)
#storm.metrics.reporters:
```

```
#   # Graphite Reporter
#   - class: "org.apache.storm.metrics2.reporters.GraphiteStormReporter"
#     daemons:
#           - "supervisor"
#           - "nimbus"
#           - "worker"
#     report.period: 60
#     report.period.units: "SECONDS"
#     graphite.host: "localhost"
#     graphite.port: 2003
#
#   # Console Reporter
#   - class: "org.apache.storm.metrics2.reporters.ConsoleStormReporter"
#     daemons:
#           - "worker"
#     report.period: 10
#     report.period.units: "SECONDS"
#     filter:
#           class: "org.apache.storm.metrics2.filters.RegexFilter"
#           expression: ".*my_component.*emitted.*"
```

最后就是运行部分：首先，先确保 zookeeper 是打开的状态（运行 bin 目录下的 zkSever.cmd），之后进入到 Strom 的 bin 目录下，分别运行 storm.py nimbus、storm.py supervisor 和 storm.py ui 指令。

- zookeeper 打开

```
D:\CSapps\zookeeper\zookeeper\bin>call "D:\CSapps\Java\JDK\javajdk8_setup"\bin\java "-Dzookeeper.log.dir=D:\CSapps\zooke
eper\zookeeper\bin\.." "-Dzookeeper.root.logger=INFO,CONSOLE" -cp "D:\CSapps\zookeeper\zookeeper\bin\..\build\classes;D:
\CSapps\zookeeper\zookeeper\bin\..\build\lib\*;D:\CSapps\zookeeper\zookeeper\bin\..\*;D:\CSapps\zookeeper\zookeeper\bin\
..\lib\*;D:\CSapps\zookeeper\zookeeper\bin\..\conf" org.apache.zookeeper.server.quorum.QuorumPeerMain "D:\CSapps\zookeep
er\zookeeper\bin\..\conf\zoo.cfg"
2021-07-05 13:20:28,238 [myid:] - INFO  [main:QuorumPeerConfig@136] - Reading configuration from: D:\CSapps\zookeeper\zo
okeeper\bin\..\conf\zoo.cfg
2021-07-05 13:20:28,241 [myid:] - INFO  [main:DatadirCleanupManager@78] - autopurge.snapRetainCount set to 3
2021-07-05 13:20:28,242 [myid:] - INFO  [main:DatadirCleanupManager@79] - autopurge.purgeInterval set to 0
2021-07-05 13:20:28,243 [myid:] - INFO  [main:DatadirCleanupManager@101] - Purge task is not scheduled.
2021-07-05 13:20:28,244 [myid:] - WARN  [main:QuorumPeerMain@116] - Either no config or no quorum defined in config, run
ning  in standalone mode
2021-07-05 13:20:28,281 [myid:] - INFO  [main:QuorumPeerConfig@136] - Reading configuration from: D:\CSapps\zookeeper\zo
okeeper\bin\..\conf\zoo.cfg
2021-07-05 13:20:28,281 [myid:] - INFO  [main:ZooKeeperServerMain@98] - Starting server
2021-07-05 13:20:28,293 [myid:] - INFO  [main:Environment@100] - Server environment:zookeeper.version=3.4.12-e5259e43754
0f349646870ea94dc2658c4e44b3b, built on 03/27/2018 03:55 GMT
2021-07-05 13:20:28,293 [myid:] - INFO  [main:Environment@100] - Server environment:host.name=LAPTOP-U34NI66Q
2021-07-05 13:20:28,293 [myid:] - INFO  [main:Environment@100] - Server environment:java.version=1.8.0_192
2021-07-05 13:20:28,293 [myid:] - INFO  [main:Environment@100] - Server environment:java.vendor=Oracle Corporation
2021-07-05 13:20:28,294 [myid:] - INFO  [main:Environment@100] - Server environment:java.home=D:\CSapps\Java\JDK\javajdk
8_setup\jre
2021-07-05 13:20:28,294 [myid:] - INFO  [main:Environment@100] - Server environment:java.class.path=D:\CSapps\zookeeper\
```

- 运行 storm.py nimbus

```
D:\CSapps\Storm\storm\bin>python storm.py nimbus
Running: D:\CSapps\Java\JDK\javajdk8_setup\bin\java.exe -server -Ddaemon.name=nimbus -Dstorm.options= -Dstorm.home=D:\CS
apps\Storm\storm -Dstorm.log.dir=D:\CSapps\Storm\storm\logs -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -
Dstorm.conf.file= -cp D:\CSapps\Storm\storm\*;D:\CSapps\Storm\storm/lib\*;D:\CSapps\Storm\storm/extlib\*;D:\CSapps\Storm
\storm/extlib-daemon\*;D:\CSapps\Storm\storm\conf -Xmx1024m -Dlogfile.name=nimbus.log -DLog4jContextSelector=org.apache.
logging.log4j.core.async.AsyncLoggerContextSelector -Dlog4j.configurationFile=D:\CSapps\Storm\storm\log4j2\cluster.xml o
rg.apache.storm.daemon.nimbus
b''
```

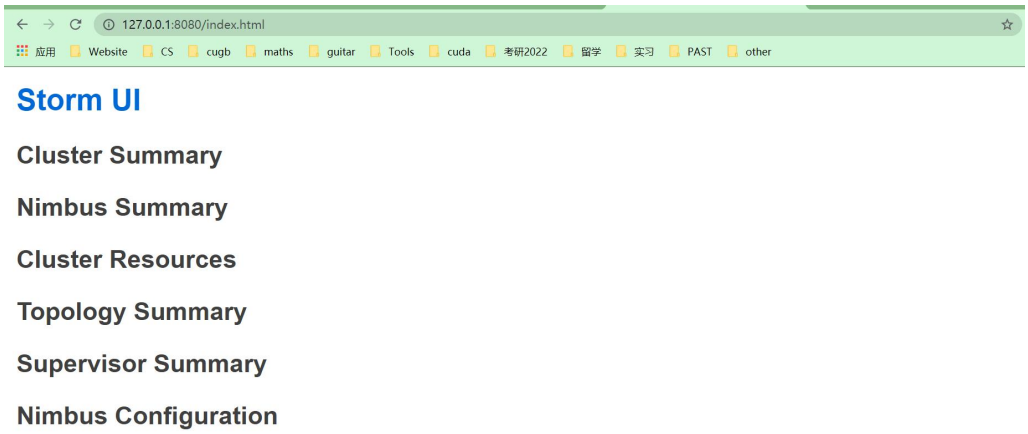
- Storm.py supervisor

```
D:\CSapps\Storm\storm\bin>python storm.py supervisor
Running: D:\CSapps\Java\JDK\javajdk8_setup\bin\java.exe -server -Ddaemon.name=supervisor -Dstorm.options= -Dstorm.h
ome=D:\CSapps\Storm\storm -Dstorm.log.dir=D:\CSapps\Storm\storm\logs -Djava.library.path=/usr/local/lib:/opt/local/
lib:/usr/lib -Dstorm.conf.file= -cp D:\CSapps\Storm\storm\*;D:\CSapps\Storm\storm/lib\*;D:\CSapps\Storm\storm/extli
b\*;D:\CSapps\Storm\storm/extlib-daemon\*;D:\CSapps\Storm\storm\conf -Xmx256m -Dlogfile.name=supervisor.log -Dlog4j
.configurationFile=D:\CSapps\Storm\storm\log4j2\cluster.xml org.apache.storm.daemon.supervisor.Supervisor
b''
```

- Storm.py ui

```
D:\CSapps\Storm\storm\bin>python storm.py ui
Running: D:\CSapps\Java\JDK\javajdk8_setup\bin\java.exe -server -Ddaemon.name=ui -Dstorm.options= -Dstorm.home=D:\CSapps
\Storm\storm -Dstorm.log.dir=D:\CSapps\Storm\storm\logs -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dsto
rm.conf.file= -cp D:\CSapps\Storm\storm\*;D:\CSapps\Storm\storm/lib\*;D:\CSapps\Storm\storm/extlib\*;D:\CSapps\Storm\sto
rm/extlib-daemon\*;D:\CSapps\Storm\storm;D:\CSapps\Storm\storm\conf -Xmx768m -Dlogfile.name=ui.log -DLog4jContextSelecto
r=org.apache.logging.log4j.core.async.AsyncLoggerContextSelector -Dlog4j.configurationFile=D:\CSapps\Storm\storm\log4j2\
cluster.xml org.apache.storm.ui.core
```

可以看到，成功运行，配置完成。



三、 结果分析

在打开 zookeeper、nimbus、supervisor、ui 可以成功得检验 storm 的安装。在 windows 系统下的安装较为简单，但是对于初学的我们仍然带来了很大的挑战。

四、 结论与展望

Storm 是一个分布式的、容错的实时计算系统。Storm 为分布式实时计算提供了一组通用原语，可被用于“流处理”之中，实时处理消息并更新数据。Storm 可以方便地在一个计算机集群中编写与扩展复杂的实时计算，Storm 用于实时处理。综上所述，Storm 具有广阔的应用前景和丰厚的应用价值，学习好 Storm 对同学们今后的学习和工作都有十分巨大的帮助。