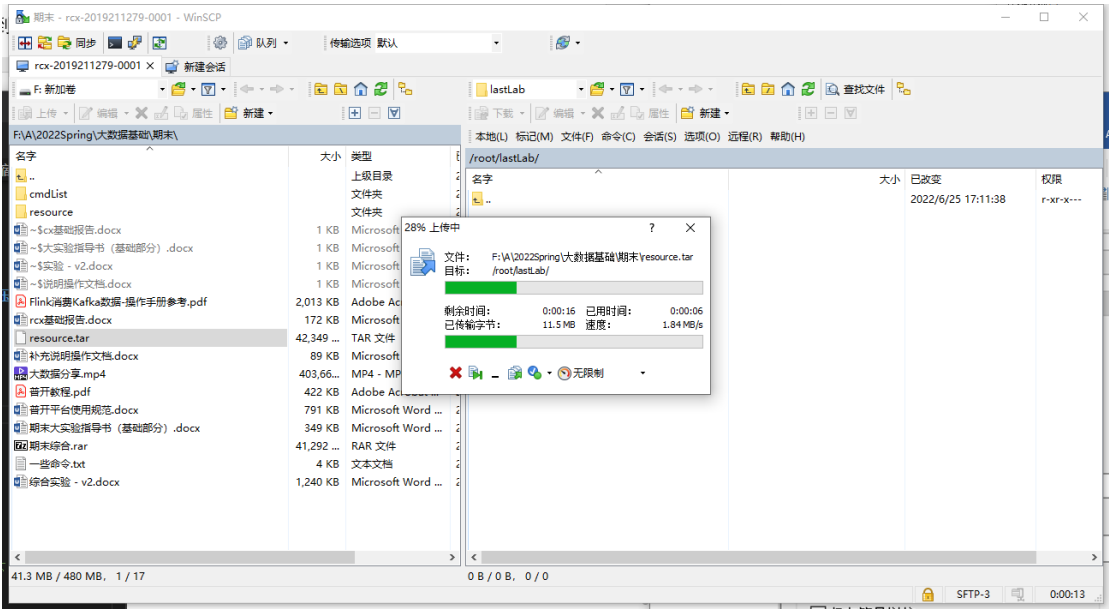
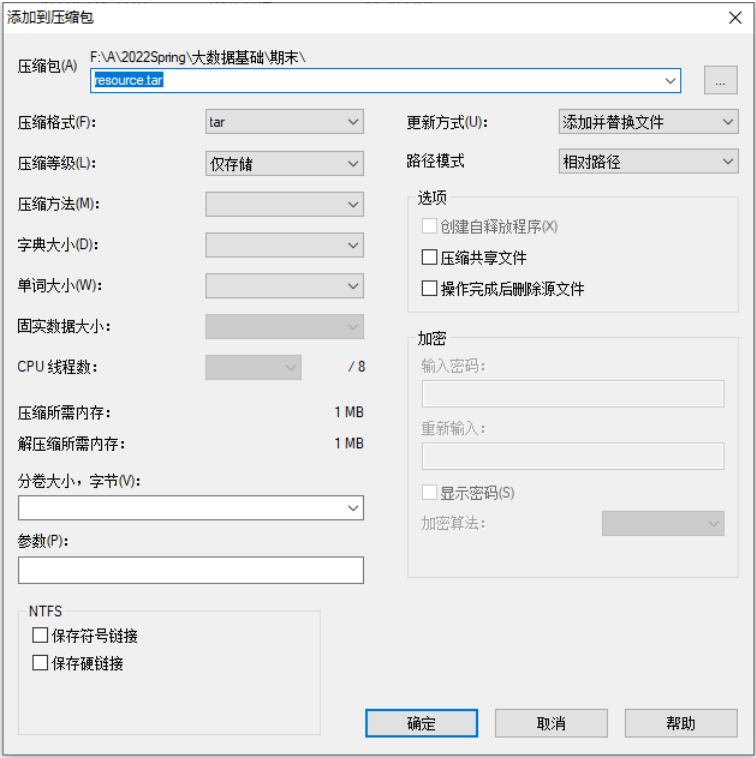


大数据综合实验-基础部分

一 安装 Kafka

1.1 将本地相关资源归档上传服务器并解压



```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER
resource/code/server-client/
resource/code/server-client/recommend_client.py
resource/code/server-client/recommend_server.py
resource/code/spark-sparkstreaming-recommend/
resource/code/spark-sparkstreaming-recommend/hbase2spark.scala
resource/code/spark-sparkstreaming-recommend/kafkaStreaming.scala
resource/code/spark-sparkstreaming-recommend/pom.xml
resource/code/spark-sparkstreaming-recommend/recommend.scala
resource/data/
resource/data/json_test_ratings.json
resource/data/json_train_ratings.json
resource/data/movies.csv
resource/kafka_2.11-0.10.2.2.tgz
resource/redis-6.0.6.tar.gz
[root@rcx-2019211279-0001 lastLab]# ls
resource resource.tar
[root@rcx-2019211279-0001 lastLab]# cd resource
[root@rcx-2019211279-0001 resource]# ls
code data kafka_2.11-0.10.2.2.tgz redis-6.0.6.tar.gz
[root@rcx-2019211279-0001 resource]#
```

1.2 解压 kafka

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER
[root@rcx-2019211279-0001 resource]# ls
code data kafka_2.11-0.10.2.2.tgz redis-6.0.6.tar.gz
[root@rcx-2019211279-0001 resource]# tar -zxvf kafka_2.11-0.10.2.2.tgz
kafka_2.11-0.10.2.2/
kafka_2.11-0.10.2.2/LICENSE
kafka_2.11-0.10.2.2/NOTICE
kafka_2.11-0.10.2.2/bin/
kafka_2.11-0.10.2.2/bin/kafka-preferred-replica-election.sh
kafka_2.11-0.10.2.2/bin/kafka-console-consumer.sh
kafka_2.11-0.10.2.2/bin/kafka-consumer-perf-test.sh
kafka_2.11-0.10.2.2/bin/zookeeper-server-stop.sh
kafka_2.11-0.10.2.2/bin/kafka-verifiable-consumer.sh
kafka_2.11-0.10.2.2/bin/kafka-acls.sh
kafka_2.11-0.10.2.2/bin/kafka-consumer-offset-checker.sh
kafka_2.11-0.10.2.2/bin/zookeeper-server-start.sh
kafka_2.11-0.10.2.2/bin/kafka-server-stop.sh
kafka_2.11-0.10.2.2/bin/kafka-simple-consumer-shell.sh
kafka_2.11-0.10.2.2/bin/kafka-configs.sh
kafka_2.11-0.10.2.2/bin/kafka-reassign-partitions.sh
kafka_2.11-0.10.2.2/bin/kafka-producer-perf-test.sh
kafka_2.11-0.10.2.2/bin/kafka-replay-log-producer.sh
kafka_2.11-0.10.2.2/bin/kafka-topics.sh
kafka_2.11-0.10.2.2/bin/connect-standalone.sh
kafka_2.11-0.10.2.2/bin/kafka-broker-api-versions.sh
[root@rcx-2019211279-0001 resource]# ls
code data kafka_2.11-0.10.2.2 kafka_2.11-0.10.2.2.tgz redis-6.0.6.tar.gz
[root@rcx-2019211279-0001 resource]#
```

1.3 编辑 config/server.properties 文件，修改 delete.topic.enable 和 zookeeper.connect

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]# ls
2019211279-rcx-input.txt lastLab
aarch64 linux
apache-hive-2.1.1-bin metastore_db
apache-hive-2.1.1-bin.tar.gz MyHbase.jar
derby.log mysql-5.7.30.tar.gz
hadoop-2.7.7 mysql-connector-java-5.1.28.jar
javasharedresources myWordCount.jar
lab4 OpenJDK8U-jdk_aarch64_linux_openj9_8u292b10_openj9-0.26.0.tar
[root@rcx-2019211279-0001 ~]# cd lastLab/
[root@rcx-2019211279-0001 lastLab]# ls
resource resource.tar
[root@rcx-2019211279-0001 lastLab]# cd resource
[root@rcx-2019211279-0001 resource]# ls
code data kafka_2.11-0.10.2.2 kafka_2.11-0.10.2.2.tgz redis-6.0.6.tar.gz
[root@rcx-2019211279-0001 resource]# mv kafka_2.11-0.10.2.2 /home/modules/
[root@rcx-2019211279-0001 resource]# ls
code data kafka_2.11-0.10.2.2 kafka_2.11-0.10.2.2.tgz redis-6.0.6.tar.gz
[root@rcx-2019211279-0001 resource]# cd /home/modules/
[root@rcx-2019211279-0001 modules]# ls
hadoop-2.7.7 hadoop-2.8.3 kafka_2.11-0.10.2.2
[root@rcx-2019211279-0001 modules]#
```

```
/home/modules/kafka_2.11-0.10.2.2/config/server.properties - rcx-2019211279-0001 - 编辑器 - WinSCP

#
# 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.

# see kafka.server.KafkaConfig for additional details and defaults

##### Server Basics #####

# The id of the broker. This must be set to a unique integer for each broker.
broker.id=0

# Switch to enable topic deletion or not, default value is false
delete.topic.enable=true

##### Socket Server Settings #####

# The address the socket server listens on. It will get the value returned from
# java.net.InetAddress.getCanonicalHostName() if not configured.
# FORMAT:
# listeners = listener_name://host_name:port
# EXAMPLE:
# listeners = PLAINTEXT://your.host.name:9092
#listeners=PLAINTEXT://:9092

# Hostname and port the broker will advertise to producers and consumers. If not set,
# it uses the value for "listeners" if configured. Otherwise, it will use the value
# returned from java.net.InetAddress.getCanonicalHostName().
#advertised.listeners=PLAINTEXT://your.host.name:9092

行: 1/124      列: 1      字符: 35 (0x23)      编码: 936 (ANSI/OEM)
```

```
/home/modules/kafka_2.11-0.10.2.2/config/server.properties - rcx-2019211279-0001 - 编辑器 - WinSCP

# The following configurations control the disposal of log segments. The policy can
# be set to delete segments after a period of time, or after a given size has accumulated.
# A segment will be deleted whenever *either* of these criteria are met. Deletion always happens
# from the end of the log.

# The minimum age of a log file to be eligible for deletion due to age
log.retention.hours=168

# A size-based retention policy for logs. Segments are pruned from the log as long as the remaining
# segments don't drop below log.retention.bytes. Functions independently of log.retention.hours.
#log.retention.bytes=1073741824

# The maximum size of a log segment file. When this size is reached a new log segment will be created.
log.segment.bytes=1073741824

# The interval at which log segments are checked to see if they can be deleted according
# to the retention policies
log.retention.check.interval.ms=300000

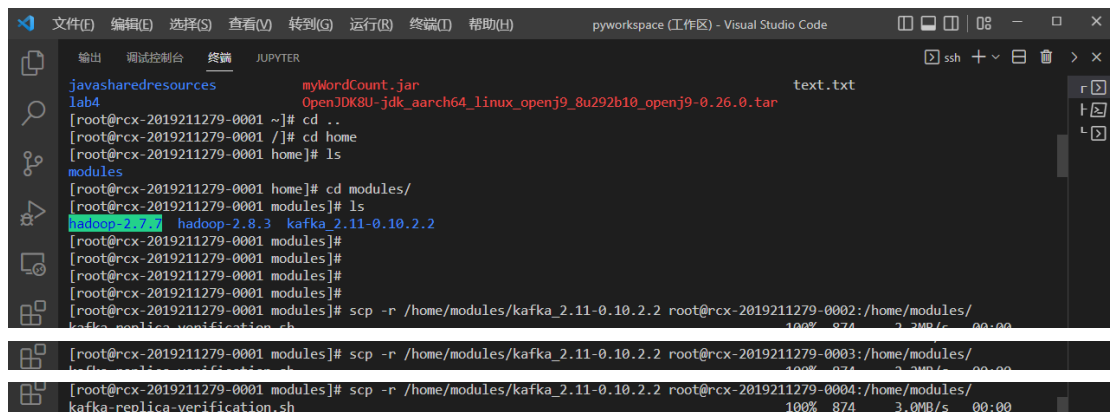
##### Zookeeper #####

# Zookeeper connection string (see zookeeper docs for details).
# This is a comma separated host:port pairs, each corresponding to a zk
# server. e.g. "127.0.0.1:3000,127.0.0.1:3001,127.0.0.1:3002".
# You can also append an optional chroot string to the urls to specify the
# root directory for all kafka znodes.
zookeeper.connect=rcx-2019211279-0001,rcx-2019211279-0002,rcx-2019211279-0003,rcx-2019211279-0004

# Timeout in ms for connecting to zookeeper
zookeeper.connection.timeout.ms=6000

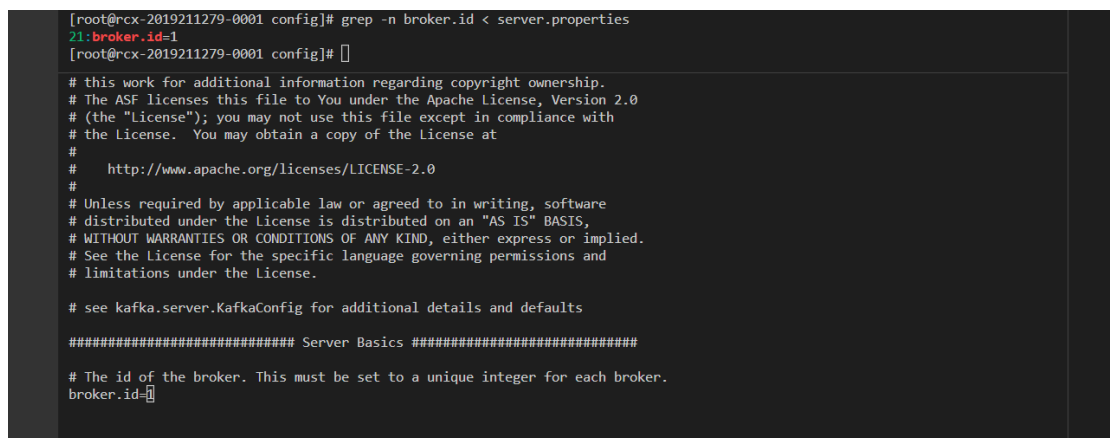
行: 1/124      列: 1      字符: 35 (0x23)      编码: 936 (ANSI/OEM)
```

1.4 将 kafka 文件夹通过 scp 发送到其余结点对应目录下



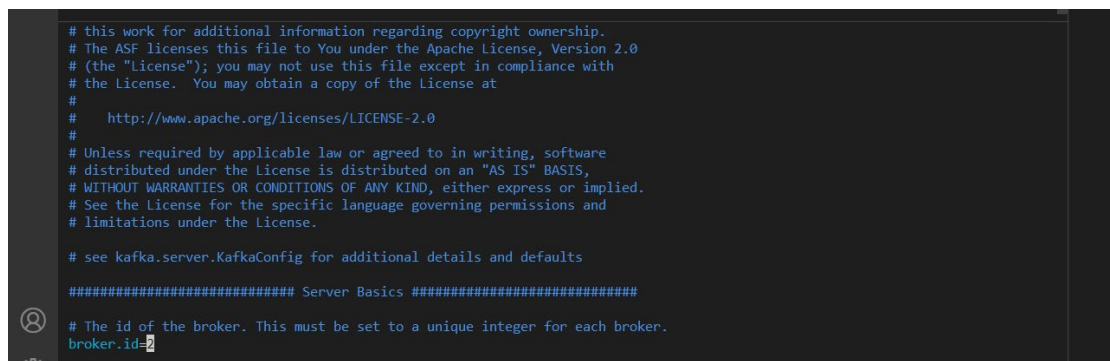
```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER
myWordCount.jar text.txt
lab4 OpenJDK8U-jdk_aarch64_linux_openj9_8u292b10_openj9-0.26.0.tar
[root@rcx-2019211279-0001 ~]# cd ..
[root@rcx-2019211279-0001 /]# cd home
[root@rcx-2019211279-0001 home]# ls
modules
[root@rcx-2019211279-0001 home]# cd modules/
[root@rcx-2019211279-0001 modules]# ls
hadoop-2.7.7 hadoop-2.8.3 kafka_2.11-0.10.2.2
[root@rcx-2019211279-0001 modules]#
[root@rcx-2019211279-0001 modules]#
[root@rcx-2019211279-0001 modules]#
[root@rcx-2019211279-0001 modules]#
[root@rcx-2019211279-0001 modules]# scp -r /home/modules/kafka_2.11-0.10.2.2 root@rcx-2019211279-0002:/home/modules/
kafka-replica-verification.sh 100% 874 3.2MB/s 00:00
[root@rcx-2019211279-0001 modules]# scp -r /home/modules/kafka_2.11-0.10.2.2 root@rcx-2019211279-0003:/home/modules/
kafka-replica-verification.sh 100% 874 3.2MB/s 00:00
[root@rcx-2019211279-0001 modules]# scp -r /home/modules/kafka_2.11-0.10.2.2 root@rcx-2019211279-0004:/home/modules/
kafka-replica-verification.sh 100% 874 3.0MB/s 00:00
```

1.5 编辑 config/server.properties 文件, 修改 broker.id 分别为 1、2、3、4

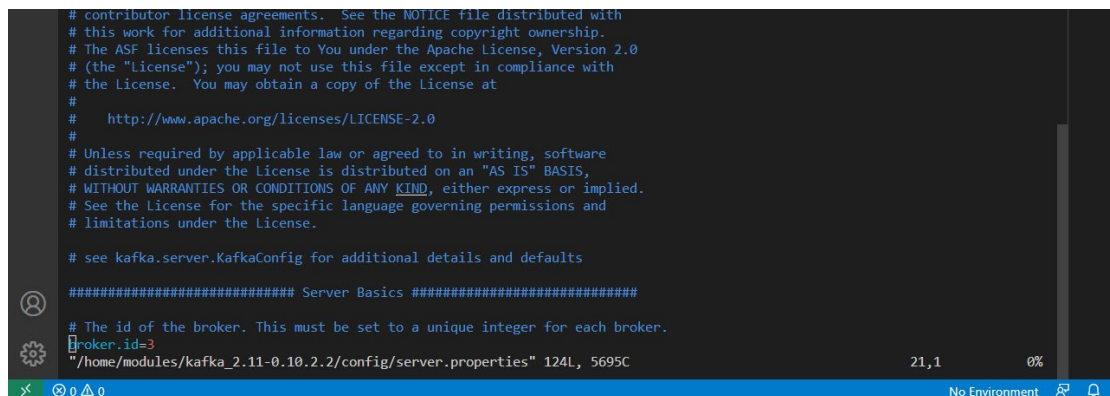


```
[root@rcx-2019211279-0001 config]# grep -n broker.id < server.properties
21:broker.id=1
[root@rcx-2019211279-0001 config]#
```

```
# 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.
#
# see kafka.server.KafkaConfig for additional details and defaults
##### Server Basics #####
# The id of the broker. This must be set to a unique integer for each broker.
broker.id=1
```



```
# 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.
#
# see kafka.server.KafkaConfig for additional details and defaults
##### Server Basics #####
# The id of the broker. This must be set to a unique integer for each broker.
broker.id=2
```



```
# 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.
#
# see kafka.server.KafkaConfig for additional details and defaults
##### Server Basics #####
# The id of the broker. This must be set to a unique integer for each broker.
broker.id=3
"/home/modules/kafka_2.11-0.10.2.2/config/server.properties" 124L, 5695C 21,1 0%
```

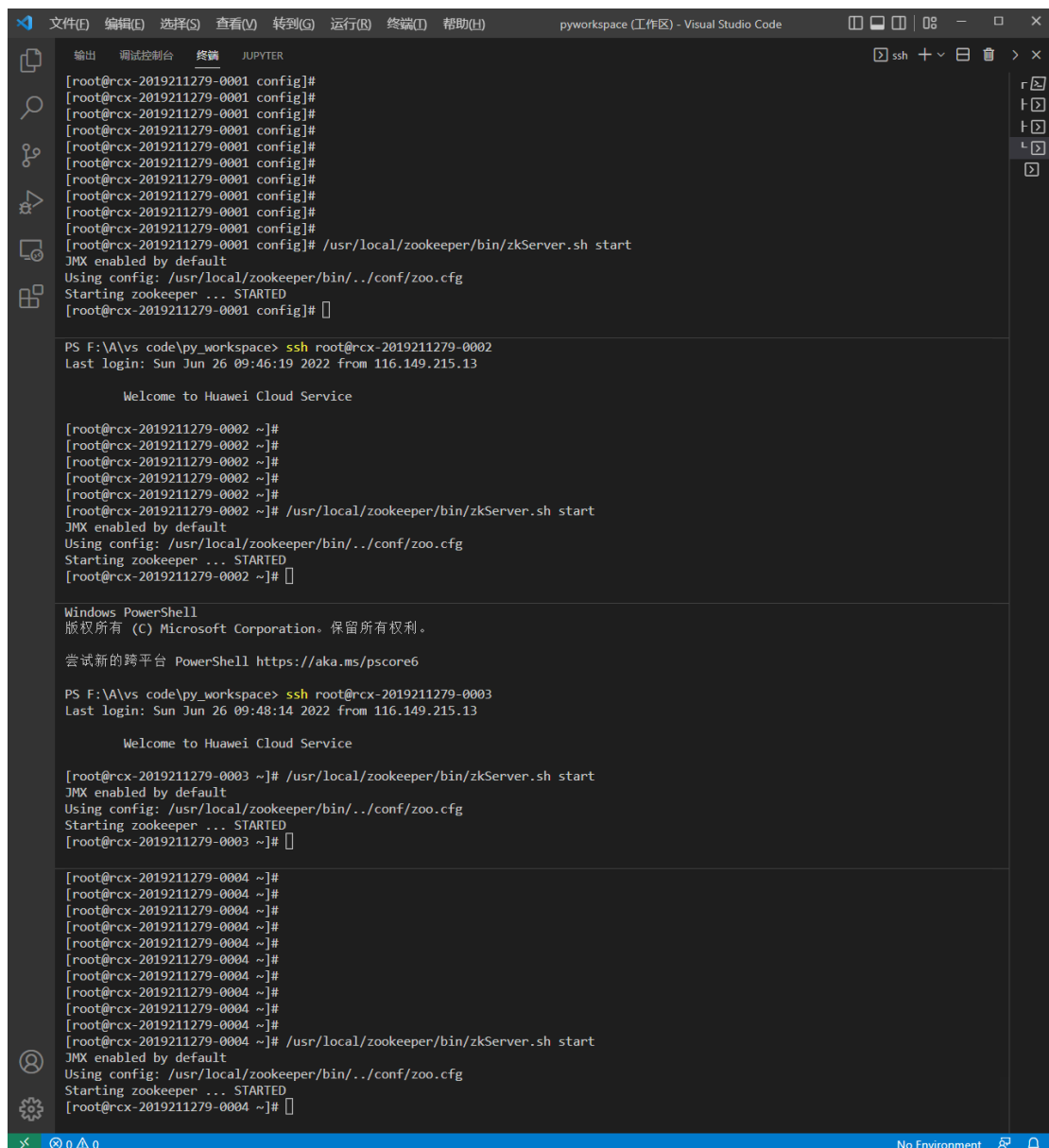
```
# this work for additional information regarding copyright ownership.
# The ASF licenses this file to You under the Apache License, Version 2.0
# (the "License"); you may not use this file except in compliance with
# the license. You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# 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.

# see kafka.server.KafkaConfig for additional details and defaults

##### Server Basics #####

# The id of the broker. This must be set to a unique integer for each broker.
broker.id=4
```

1.6 各节点启动 zookeeper



The screenshot shows the Visual Studio Code interface with a terminal window titled "JUPYTER". The terminal displays four separate sessions for starting ZooKeeper on different nodes. Each session begins with an SSH connection to a node (e.g., root@rcx-2019211279-0001) and the execution of the command `/usr/local/zookeeper/bin/zkServer.sh start`. The output for each session shows that ZooKeeper is successfully started, displaying messages like "JMX enabled by default", "Using config: /usr/local/zookeeper/bin/./conf/zoo.cfg", and "Starting zookeeper ... STARTED".

```
输出 调试控制台 终端 JUPYTER
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]#
[root@rcx-2019211279-0001 config]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/./conf/zoo.cfg
Starting zookeeper ... STARTED
[root@rcx-2019211279-0001 config]#

PS F:\A\vs code\py_workspace> ssh root@rcx-2019211279-0002
Last login: Sun Jun 26 09:46:19 2022 from 116.149.215.13

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/./conf/zoo.cfg
Starting zookeeper ... STARTED
[root@rcx-2019211279-0002 ~]#

Windows PowerShell
版权所有 (C) Microsoft Corporation。保留所有权利。

尝试新的跨平台 PowerShell https://aka.ms/pscore6

PS F:\A\vs code\py_workspace> ssh root@rcx-2019211279-0003
Last login: Sun Jun 26 09:48:14 2022 from 116.149.215.13

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0003 ~]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/./conf/zoo.cfg
Starting zookeeper ... STARTED
[root@rcx-2019211279-0003 ~]#

[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/./conf/zoo.cfg
Starting zookeeper ... STARTED
[root@rcx-2019211279-0004 ~]#
```

1.7 各节点启动 kafka

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER ssh + - 日 垃圾桶 > x
inator.GroupMetadataManager)
[2022-06-26 09:54:21,372] INFO Will not load MX4J, mx4j-tools.jar is not in the classpath (kafka.utils.Mx4JLoader$)
[2022-06-26 09:54:21,413] INFO Creating /brokers/ids/1 (is it secure? false) (kafka.utils.ZKCheckedEphemeral)
[2022-06-26 09:54:21,422] INFO Result of znode creation is: OK (kafka.utils.ZKCheckedEphemeral)
[2022-06-26 09:54:21,425] INFO Registered broker 1 at path /brokers/ids/1 with addresses: EndPoint(rcx-2019211279-0001,9092,LicenseName(PLAINTEXT),PLAINTEXT) (kafka.utils.ZkUtils)
[2022-06-26 09:54:21,426] WARN No meta.properties file under dir /tmp/kafka-logs/meta.properties (kafka.server.BrokerMetadataCheckpoint)
[2022-06-26 09:54:21,467] INFO New leader is 1 (kafka.server.ZookeeperLeaderElector$LeaderChangeListener)
[2022-06-26 09:54:21,469] INFO Kafka version : 0.10.2.2 (org.apache.kafka.common.utils.AppInfoParser)
[2022-06-26 09:54:21,469] INFO Kafka commitId : cd80bc412b9b9701 (org.apache.kafka.common.utils.AppInfoParser)
[2022-06-26 09:54:21,470] INFO [Kafka Server 1], started (kafka.server.KafkaServer)

[2022-06-26 09:55:28,016] INFO [GroupCoordinator 2]: Starting up. (kafka.coordinator.GroupCoordinator)
[2022-06-26 09:55:28,017] INFO [GroupCoordinator 2]: Startup complete. (kafka.coordinator.GroupCoordinator)
[2022-06-26 09:55:28,018] INFO [Group Metadata Manager on Broker 2]: Removed 0 expired offsets in 1 milliseconds. (kafka.coordinator.GroupMetadataManager)
[2022-06-26 09:55:28,036] INFO Will not load MX4J, mx4j-tools.jar is not in the classpath (kafka.utils.Mx4JLoader$)
[2022-06-26 09:55:28,063] INFO Creating /brokers/ids/2 (is it secure? false) (kafka.utils.ZKCheckedEphemeral)
[2022-06-26 09:55:28,074] INFO Result of znode creation is: OK (kafka.utils.ZKCheckedEphemeral)
[2022-06-26 09:55:28,075] INFO Registered broker 2 at path /brokers/ids/2 with addresses: EndPoint(rcx-2019211279-0002,9092,LicenseName(PLAINTEXT),PLAINTEXT) (kafka.utils.ZkUtils)
[2022-06-26 09:55:28,076] WARN No meta.properties file under dir /tmp/kafka-logs/meta.properties (kafka.server.BrokerMetadataCheckpoint)
[2022-06-26 09:55:28,096] INFO Kafka version : 0.10.2.2 (org.apache.kafka.common.utils.AppInfoParser)
[2022-06-26 09:55:28,097] INFO Kafka commitId : cd80bc412b9b9701 (org.apache.kafka.common.utils.AppInfoParser)
[2022-06-26 09:55:28,098] INFO [Kafka Server 2], started (kafka.server.KafkaServer)

[2022-06-26 09:55:32,356] INFO [GroupCoordinator 3]: Starting up. (kafka.coordinator.GroupCoordinator)
[2022-06-26 09:55:32,358] INFO [GroupCoordinator 3]: Startup complete. (kafka.coordinator.GroupCoordinator)
[2022-06-26 09:55:32,359] INFO [Group Metadata Manager on Broker 3]: Removed 0 expired offsets in 1 milliseconds. (kafka.coordinator.GroupMetadataManager)
[2022-06-26 09:55:32,377] INFO Will not load MX4J, mx4j-tools.jar is not in the classpath (kafka.utils.Mx4JLoader$)
[2022-06-26 09:55:32,402] INFO Creating /brokers/ids/3 (is it secure? false) (kafka.utils.ZKCheckedEphemeral)
[2022-06-26 09:55:32,415] INFO Result of znode creation is: OK (kafka.utils.ZKCheckedEphemeral)
[2022-06-26 09:55:32,416] INFO Registered broker 3 at path /brokers/ids/3 with addresses: EndPoint(rcx-2019211279-0003,9092,LicenseName(PLAINTEXT),PLAINTEXT) (kafka.utils.ZkUtils)
[2022-06-26 09:55:32,417] WARN No meta.properties file under dir /tmp/kafka-logs/meta.properties (kafka.server.BrokerMetadataCheckpoint)
[2022-06-26 09:55:32,439] INFO Kafka version : 0.10.2.2 (org.apache.kafka.common.utils.AppInfoParser)
[2022-06-26 09:55:32,439] INFO Kafka commitId : cd80bc412b9b9701 (org.apache.kafka.common.utils.AppInfoParser)
[2022-06-26 09:55:32,440] INFO [Kafka Server 3], started (kafka.server.KafkaServer)

[2022-06-26 09:55:35,100] INFO [ExpirationReaper-4], Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2022-06-26 09:55:35,101] INFO [ExpirationReaper-4], Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2022-06-26 09:55:35,101] INFO [ExpirationReaper-4], Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2022-06-26 09:55:35,115] INFO [GroupCoordinator 4]: Starting up. (kafka.coordinator.GroupCoordinator)
[2022-06-26 09:55:35,116] INFO [GroupCoordinator 4]: Startup complete. (kafka.coordinator.GroupCoordinator)
[2022-06-26 09:55:35,137] INFO Will not load MX4J, mx4j-tools.jar is not in the classpath (kafka.utils.Mx4JLoader$)
[2022-06-26 09:55:35,137] INFO [Group Metadata Manager on Broker 4]: Removed 0 expired offsets in 2 milliseconds. (kafka.coordinator.GroupMetadataManager)
[2022-06-26 09:55:35,163] INFO Creating /brokers/ids/4 (is it secure? false) (kafka.utils.ZKCheckedEphemeral)
[2022-06-26 09:55:35,171] INFO Result of znode creation is: OK (kafka.utils.ZKCheckedEphemeral)
[2022-06-26 09:55:35,172] INFO Registered broker 4 at path /brokers/ids/4 with addresses: EndPoint(rcx-2019211279-0004,9092,LicenseName(PLAINTEXT),PLAINTEXT) (kafka.utils.ZkUtils)
[2022-06-26 09:55:35,173] WARN No meta.properties file under dir /tmp/kafka-logs/meta.properties (kafka.server.BrokerMetadataCheckpoint)
[2022-06-26 09:55:35,194] INFO Kafka version : 0.10.2.2 (org.apache.kafka.common.utils.AppInfoParser)
```

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER ssh + - 日 垃圾桶 > x
Windows PowerShell
版权所有 (C) Microsoft Corporation。保留所有权利。

尝试新的跨平台 PowerShell https://aka.ms/pscore6

PS F:\> cd \\py_workspace> ssh root@rcx-2019211279-0001
Last login: Sun Jun 26 09:41:53 2022

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0001 ~]# jps
2027 Jps
1717 Kafka
1678 QuorumPeerMain
[root@rcx-2019211279-0001 ~]#
```


二、安装 Redis（单机部署）

2.1 升级 gcc



```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
ssh + - 日 窗 > x
输出 调试控制台 终端 JUPYTER
Last login: Sun Jun 26 09:57:02 2022 from 116.149.215.13

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]# yum -y install centos-release-scl
Loaded plugins: fastestmirror
Determining fastest mirrors
base                                     | 3.6 kB  00:00:00
epel                                   | 5.4 kB  00:00:00
extras                                | 2.9 kB  00:00:00
updates                               | 2.9 kB  00:00:00
(1/2): extras/7/aarch64/primary_db     | 250 kB  00:00:00
(2/2): updates/7/aarch64/primary_db    | 3.2 MB  00:00:00
Resolving Dependencies
--> Running transaction check
--> Package centos-release-scl.noarch 0:2-3.el7.centos will be installed
--> Processing Dependency: centos-release-scl-rh for package: centos-release-scl-2-3.el7.centos.noarch
--> Running transaction check
--> Package centos-release-scl-rh.noarch 0:2-3.el7.centos will be installed
--> Finished Dependency Resolution

Dependencies Resolved

=====
Package                                Arch              Version            Repository          Size
=====
Installing:
centos-release-scl                     noarch            2-3.el7.centos     extras              12 k
Installing for dependencies:
centos-release-scl-rh                  noarch            2-3.el7.centos     extras              12 k
=====

Transaction Summary
=====
Install 1 Package (+1 Dependent package)

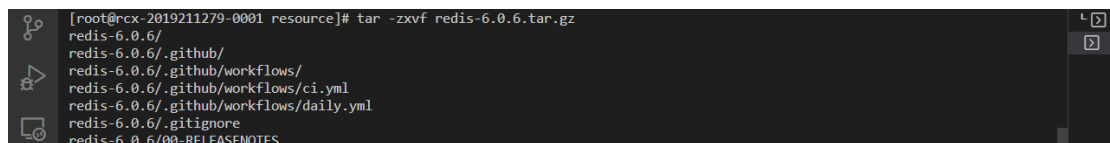
Total download size: 24 k
Installed size: 39 k
Downloading packages:
(1/2): centos-release-scl-rh-2-3.el7.centos.noarch.rpm | 12 kB  00:00:00
(2/2): centos-release-scl-2-3.el7.centos.noarch.rpm    | 12 kB  00:00:00
-----
Total                                                    86 kB/s | 24 kB  00:00:00
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Installing : centos-release-scl-rh-2-3.el7.centos.noarch 1/2
  Installing : centos-release-scl-2-3.el7.centos.noarch   2/2
  Verifying   : centos-release-scl-2-3.el7.centos.noarch 1/2
  Verifying   : centos-release-scl-rh-2-3.el7.centos.noarch 2/2

Installed:
centos-release-scl.noarch 0:2-3.el7.centos

Dependency Installed:
centos-release-scl-rh.noarch 0:2-3.el7.centos

Complete!
[root@rcx-2019211279-0001 ~]#
```

2.2 解压 redis



```
[root@rcx-2019211279-0001 resource]# tar -zxvf redis-6.0.6.tar.gz
redis-6.0.6/
redis-6.0.6/.github/
redis-6.0.6/.github/workflows/
redis-6.0.6/.github/workflows/ci.yml
redis-6.0.6/.github/workflows/daily.yml
redis-6.0.6/.gitignore
redis-6.0.6/00-RELEASENOTES
```

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER ssh + v 日 垃圾桶 X
* centos-sclo-sclo: mirrors.tuna.tsinghua.edu.cn
Package devtoolset-9-gcc-9.3.1-2.2.el7.aarch64 already installed and latest version
Package devtoolset-9-binutils-2.32-16.el7.aarch64 already installed and latest version
Resolving Dependencies
--> Running transaction check
--> Package devtoolset-9-gcc-c++.aarch64 0:9.3.1-2.2.el7 will be installed
--> Processing Dependency: devtoolset-9-libstdc++-devel = 9.3.1-2.2.el7 for package: devtoolset-9-gcc-c++-9.3.1-2.2.el7.aarch64
--> Package devtoolset-9-libatomic-devel.aarch64 0:9.3.1-2.2.el7 will be installed
--> Processing Dependency: libatomic(aarch-64) >= 4.8.0 for package: devtoolset-9-libatomic-devel-9.3.1-2.2.el7.aarch64
--> Running transaction check
--> Package devtoolset-9-libstdc++-devel.aarch64 0:9.3.1-2.2.el7 will be installed
--> Package libatomic.aarch64 0:8.3.1-2.1.1.el7 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

=====
Package Arch Version Repository Size
=====
Installing:
devtoolset-9-gcc-c++ aarch64 9.3.1-2.2.el7 centos-sclo-rh 10 M
devtoolset-9-libatomic-devel aarch64 9.3.1-2.2.el7 centos-sclo-rh 19 k
Installing for dependencies:
devtoolset-9-libstdc++-devel aarch64 9.3.1-2.2.el7 centos-sclo-rh 3.1 M
libatomic aarch64 8.3.1-2.1.1.el7 base 34 k
=====

Transaction Summary
=====
Install 2 Packages (+2 Dependent packages)

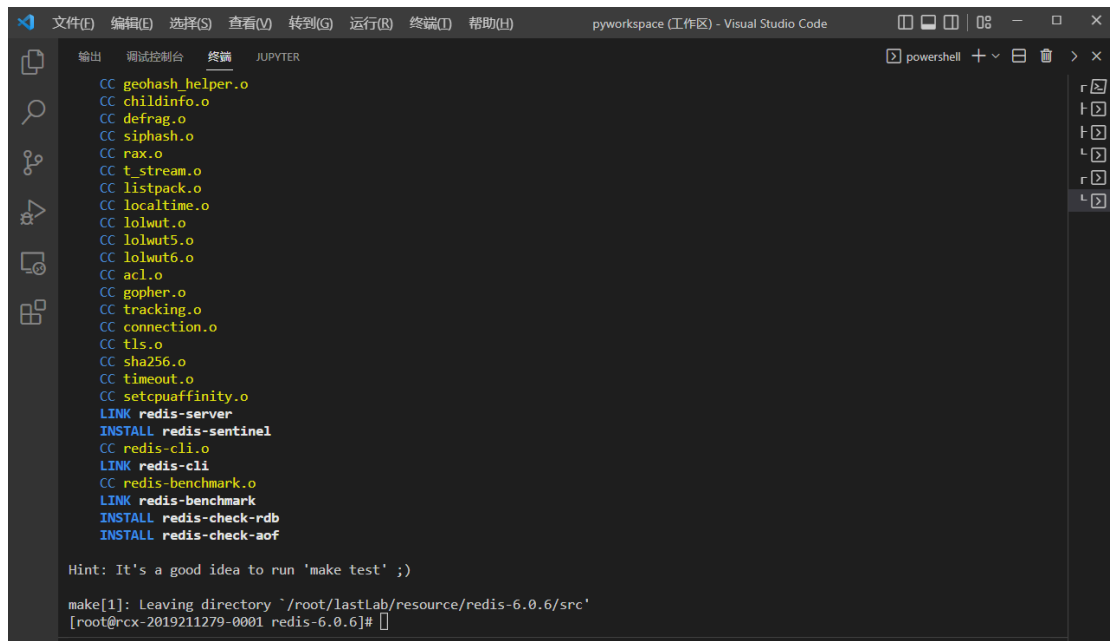
Total download size: 14 M
Installed size: 43 M
Downloading packages:
(1/4): devtoolset-9-libatomic-devel-9.3.1-2.2.el7.aarch64.rpm | 19 kB 00:00:00
(2/4): libatomic-8.3.1-2.1.1.el7.aarch64.rpm | 34 kB 00:00:00
(3/4): devtoolset-9-libstdc++-devel-9.3.1-2.2.el7.aarch64.rpm | 3.1 MB 00:00:02
(4/4): devtoolset-9-gcc-c++-9.3.1-2.2.el7.aarch64.rpm | 10 MB 00:00:08
-----
Total 1.7 MB/s | 14 MB 00:00:08
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
Installing : libatomic-8.3.1-2.1.1.el7.aarch64 1/4
Installing : devtoolset-9-libstdc++-devel-9.3.1-2.2.el7.aarch64 2/4
Installing : devtoolset-9-gcc-c++-9.3.1-2.2.el7.aarch64 3/4
Installing : devtoolset-9-libatomic-devel-9.3.1-2.2.el7.aarch64 4/4
Verifying : devtoolset-9-libstdc++-devel-9.3.1-2.2.el7.aarch64 1/4
Verifying : devtoolset-9-libatomic-devel-9.3.1-2.2.el7.aarch64 2/4
Verifying : devtoolset-9-gcc-c++-9.3.1-2.2.el7.aarch64 3/4
Verifying : libatomic-8.3.1-2.1.1.el7.aarch64 4/4

Installed:
devtoolset-9-gcc-c++.aarch64 0:9.3.1-2.2.el7 devtoolset-9-libatomic-devel.aarch64 0:9.3.1-2.2.el7

Dependency Installed:
devtoolset-9-libstdc++-devel.aarch64 0:9.3.1-2.2.el7 libatomic.aarch64 0:8.3.1-2.1.1.el7

Complete!
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]# scl enable devtoolset-9 bash
[root@rcx-2019211279-0001 ~]#
```

2.3 进入解压得到的文件夹，编译，安装

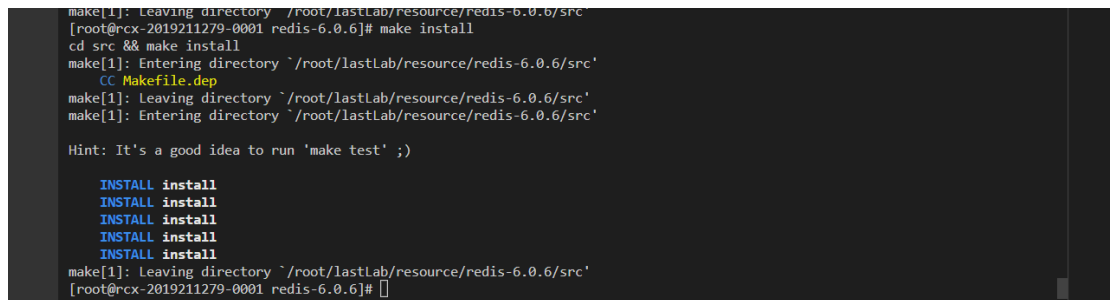


The screenshot shows the Visual Studio Code interface with a terminal window. The terminal displays the compilation progress of Redis 6.0.6. It lists various object files being compiled (e.g., geohash_helper.o, childinfo.o) and then shows the linking of redis-server, redis-sentinel, redis-cli, and redis-benchmark. The process is currently in the 'INSTALL' phase for redis-check-rdb and redis-check-aof. A hint suggests running 'make test'.

```
CC geohash_helper.o
CC childinfo.o
CC defrag.o
CC siphash.o
CC rax.o
CC t_stream.o
CC listpack.o
CC localtime.o
CC lolwut.o
CC lolwut5.o
CC lolwut6.o
CC acl.o
CC gopher.o
CC tracking.o
CC connection.o
CC tls.o
CC sha256.o
CC timeout.o
CC setcpuaffinity.o
LINK redis-server
INSTALL redis-sentinel
CC redis-cli.o
LINK redis-cli
CC redis-benchmark.o
LINK redis-benchmark
INSTALL redis-check-rdb
INSTALL redis-check-aof

Hint: It's a good idea to run 'make test' ;)
```

make[1]: Leaving directory `/root/lastLab/resource/redis-6.0.6/src'
[root@rcx-2019211279-0001 redis-6.0.6]#



The screenshot shows the continuation of the Redis 6.0.6 installation process. It shows the execution of 'make install' and the subsequent installation of various components (e.g., Makefile.dep, install). The process is currently in the 'INSTALL' phase for install. A hint suggests running 'make test'.

```
make[1]: Leaving directory `/root/lastLab/resource/redis-6.0.6/src'
[root@rcx-2019211279-0001 redis-6.0.6]# make install
cd src && make install
make[1]: Entering directory `/root/lastLab/resource/redis-6.0.6/src'
CC Makefile.dep
make[1]: Leaving directory `/root/lastLab/resource/redis-6.0.6/src'
make[1]: Entering directory `/root/lastLab/resource/redis-6.0.6/src'

Hint: It's a good idea to run 'make test' ;)
```

INSTALL install
INSTALL install
INSTALL install
INSTALL install
INSTALL install

make[1]: Leaving directory `/root/lastLab/resource/redis-6.0.6/src'
[root@rcx-2019211279-0001 redis-6.0.6]#

2.4 修改 redis.conf, 设置 redis 为守护进程, 并允许远程连接, 关闭保护模式

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]# cd /root/lastlab/resource/
[root@rcx-2019211279-0001 resource]# ls
code kafka 2.11-0.10.2.2.tgz redis-6.0.6 redis-6.0.6.tar.gz
[root@rcx-2019211279-0001 resource]# cd redis-6.0.6
[root@rcx-2019211279-0001 redis-6.0.6]# grep -n daemon < redis.conf
220:# By default Redis does not run as a daemon. Use 'yes' if you need it.
221:# Note that Redis will write a pid file in /var/run/redis.pid when daemonized.
222:daemonize no
238:# When the server runs non daemonized, no pid file is created if none is
239:# specified in the configuration. When the server is daemonized, the pid file
256:# output for logging but daemonize, logs will be sent to /dev/null
[root@rcx-2019211279-0001 redis-6.0.6]# grep -n protect-mode < redis.conf
[root@rcx-2019211279-0001 redis-6.0.6]# grep -n protected-mode < redis.conf
88:protected-mode yes
[root@rcx-2019211279-0001 redis-6.0.6]# grep -n requirepass < redis.conf
387:# If the master is password protected (using the "requirepass" configuration
780:# IMPORTANT NOTE: starting with Redis 6 "requirepass" is just a compatibility
786:# requirepass foobared

214 # seconds.
215 #
216 # tls-session-cache-timeout 60
217
218 ##### GENERAL #####
219
220 # By default Redis does not run as a daemon. Use 'yes' if you need it.
221 # Note that Redis will write a pid file in /var/run/redis.pid when daemonized.
222 daemonize yes
223
224 # If you run Redis from upstart or systemd, Redis can interact with your
225 # supervision tree. Options:
226 # supervised no - no supervision interaction
227 # supervised upstart - signal upstart by putting Redis into SIGSTOP mode
228 # supervised systemd - signal systemd by writing READY=1 to $NOTIFY_SOCKET
229 # supervised auto - detect upstart or systemd method based on
230 # UPSTART_JOB or NOTIFY_SOCKET environment variables
231 # Note: these supervision methods only signal "process is ready."
232 # They do not enable continuous liveness pings back to your supervisor.

223,0-1 11%
```

```
80 # The server only accepts connections from clients connecting from the
81 # IPv4 and IPv6 loopback addresses 127.0.0.1 and ::1, and from Unix domain
82 # sockets.
83 #
84 # By default protected mode is enabled. You should disable it only if
85 # you are sure you want clients from other hosts to connect to Redis
86 # even if no authentication is configured, nor a specific set of interfaces
87 # are explicitly listed using the "bind" directive.
88 protected-mode no
89
90 # Accept connections on the specified port, default is 6379 (IANA #815344).
91 # If port 0 is specified Redis will not listen on a TCP socket.
92 port 6379
93
94 # TCP listen() backlog.
95 #
96 # In high requests-per-second environments you need an high backlog in order
97 # to avoid slow clients connections issues. Note that the Linux kernel
98 # will silently truncate it to the value of /proc/sys/net/core/somaxconn so

89,0-1 4%
```

```
786:# requirepass foobared
778 # aclfile /etc/redis/users.acl
779
780 # IMPORTANT NOTE: starting with Redis 6 "requirepass" is just a compatibility
781 # layer on top of the new ACL system. The option effect will be just setting
782 # the password for the default user. Clients will still authenticate using
783 # AUTH <password> as usually, or more explicitly with AUTH default <password>
784 # if they follow the new protocol: both will work.
785 #
786 # requirepass foobared
787 # requirepass rcx@200012291918
788 # Command renaming (DEPRECATED).
789 #
790 # -----
791 # WARNING: avoid using this option if possible. Instead use ACLs to remove
792 # commands from the default user, and put them only in some admin user you
793 # create for administrative purposes.
794 # -----
795 #
796 # It is possible to change the name of dangerous commands in a shared

787,1 42%
```

```
[root@rcx-2019211279-0001 redis-6.0.6]#
[root@rcx-2019211279-0001 redis-6.0.6]#
[root@rcx-2019211279-0001 redis-6.0.6]#
[root@rcx-2019211279-0001 redis-6.0.6]#
[root@rcx-2019211279-0001 redis-6.0.6]# grep -n daemon < redis.conf
220:# By default Redis does not run as a daemon. Use 'yes' if you need it.
221:# Note that Redis will write a pid file in /var/run/redis.pid when daemonized.
222:daemonize yes
238:# When the server runs non daemonized, no pid file is created if none is
239:# specified in the configuration. When the server is daemonized, the pid file
256:# output for logging but daemonize, logs will be sent to /dev/null
[root@rcx-2019211279-0001 redis-6.0.6]# grep -n protected-mode < redis.conf
88:protected-mode no
[root@rcx-2019211279-0001 redis-6.0.6]# grep -n requirepass < redis.conf
387:# If the master is password protected (using the "requirepass" configuration
780:# IMPORTANT NOTE: starting with Redis 6 "requirepass" is just a compatibility
786:# requirepass foobared
787:requirepass rcx@200012291918
1490:# So use the 'requirepass' option to protect your instance.
[root@rcx-2019211279-0001 redis-6.0.6]#
```

2.5、启动 redis

```
[root@rcx-2019211279-0001 resource]# redis-server redis-6.0.6/redis.conf
7051:C 26 Jun 2022 11:51:55.423 # oO0oO00oO00o Redis is starting oO0oO00oO00o
7051:C 26 Jun 2022 11:51:55.423 # Redis version=6.0.6, bits=64, commit=00000000, modified=0, pid=7051, just started
7051:C 26 Jun 2022 11:51:55.423 # Configuration loaded
[root@rcx-2019211279-0001 resource]#
```

2.6 ps 确认 redis 运行成功

```
7051:C 26 Jun 2022 11:51:55.423 # Configuration loaded
[root@rcx-2019211279-0001 resource]# ps -ef | grep redis
root      7052      1  0 11:51 ?        00:00:00 redis-server 127.0.0.1:6379
root      7058    7016  0 11:53 pts/5    00:00:00 grep --color=auto redis
```

三、安装 python3

```
Loaded plugins: fastestmirror
[root@rcx-2019211279-0001 resource]# yum install python3
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
 * centos-scl-rh: mirrors.tuna.tsinghua.edu.cn
 * centos-scl-scl: mirrors.tuna.tsinghua.edu.cn
Resolving Dependencies
--> Running transaction check
--> Package python3.aarch64 0:3.6.8-18.el7 will be installed
--> Processing Dependency: python3-libs(aarch-64) = 3.6.8-18.el7 for package: python3-3.6.8-18.el7.aarch64
--> Processing Dependency: python3-setuptools for package: python3-3.6.8-18.el7.aarch64
--> Processing Dependency: python3-pip for package: python3-3.6.8-18.el7.aarch64
--> Processing Dependency: libpython3.6m.so.1.0()(64bit) for package: python3-3.6.8-18.el7.aarch64
--> Running transaction check
--> Package python3-libs.aarch64 0:3.6.8-18.el7 will be installed
--> Processing Dependency: libtirpc.so.1()(64bit) for package: python3-libs-3.6.8-18.el7.aarch64
--> Package python3-pip.noarch 0:9.0.3-8.el7 will be installed
--> Package python3-setuptools.noarch 0:39.2.0-10.el7 will be installed
--> Running transaction check
--> Package libtirpc.aarch64 0:0.2.4-0.16.el7 will be installed
```

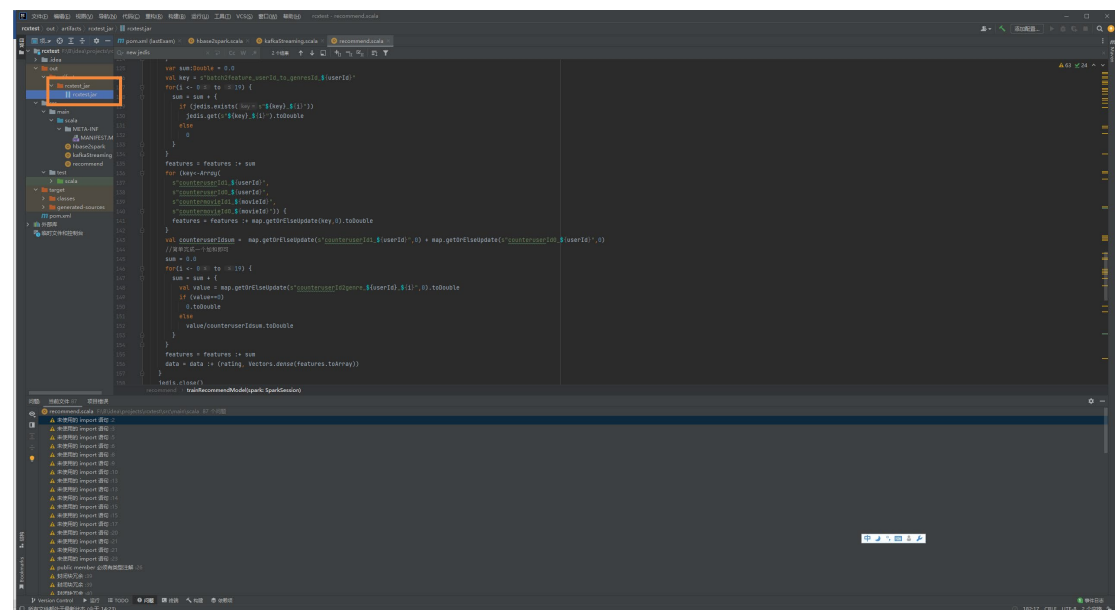
```
dwz.aarch64 0:0.11-3.el7          perl-srpm-macros.noarch 0:1-8.el7          python-rpm-macros.noarch 0:3-34.el7
python-srpm-macros.noarch 0:3-34.el7  python3-rpm-generators.noarch 0:6-2.el7  python3-rpm-macros.noarch 0:3-34.el7
redhat-rpm-config.noarch 0:9.1.0-88.el7.centos
```

```
Complete!
[root@rcx-2019211279-0001 resource]# pip3 install happybase
WARNING: Running pip install with root privileges is generally not a good idea. Try 'pip3 install --user' instead.
Collecting happybase
  Using cached https://files.pythonhosted.org/packages/d1/9c/f5f7bdb5439cda2b7da4e20ac24ec0e2455fd68aade8397f211d2994c39d/happ
ybase-1.2.0.tar.gz
Requirement already satisfied: six in /usr/local/lib/python3.6/site-packages (from happybase)
Collecting thrift2>=0.4 (from happybase)
  Using cached https://files.pythonhosted.org/packages/1d/d1/6b041449bd04b953294f3a070fc96bd8ce23ff81e96cc4c2920f7d555fe0/thri
ft2-0.4.14.tar.gz
Requirement already satisfied: pyc4.0,>=3.4 in /usr/local/lib/python3.6/site-packages (from thrift2>=0.4->happybase)
Installing collected packages: thrift2, happybase
  Running setup.py install for thrift2 ... done
  Running setup.py install for happybase ... done
Successfully installed happybase-1.2.0 thrift2-0.4.14
[root@rcx-2019211279-0001 resource]#
```

修改 python 代码（加入密码）

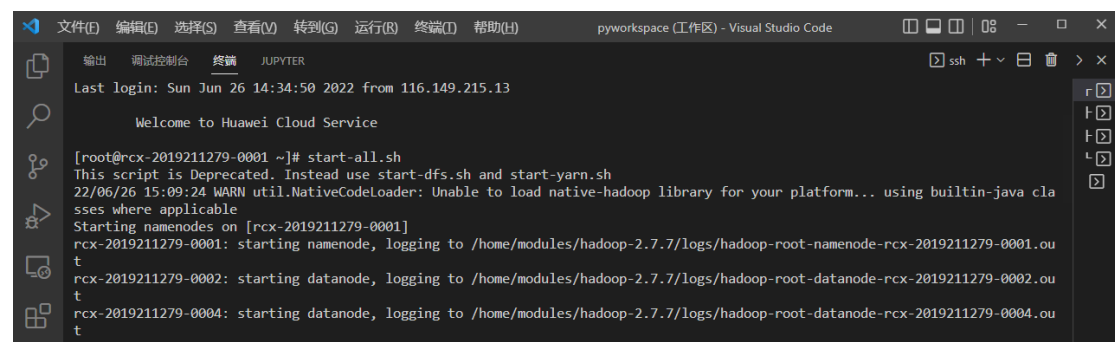
```
[root@rcx-2019211279-0001 load]#  
[root@rcx-2019211279-0001 load]# grep -n redis.Con <generatorRecord.py  
[root@rcx-2019211279-0001 load]# grep -n redis.Con <load_movie_redis.py  
16: pool = redis.ConnectionPool(host=host,port=port,decode_responses=True)  
[root@rcx-2019211279-0001 load]# grep -n redis.Con <load_train_ratings_hbase.py  
  
[root@rcx-2019211279-0001 server-client]# grep -n redis.Con <recommend_client.py  
[root@rcx-2019211279-0001 server-client]# grep -n redis.Con <recommend_server.py  
18: pool = redis.ConnectionPool(host=host,port=port,decode_responses=True)  
[root@rcx-2019211279-0001 server-client]#
```

四、打 jar 包

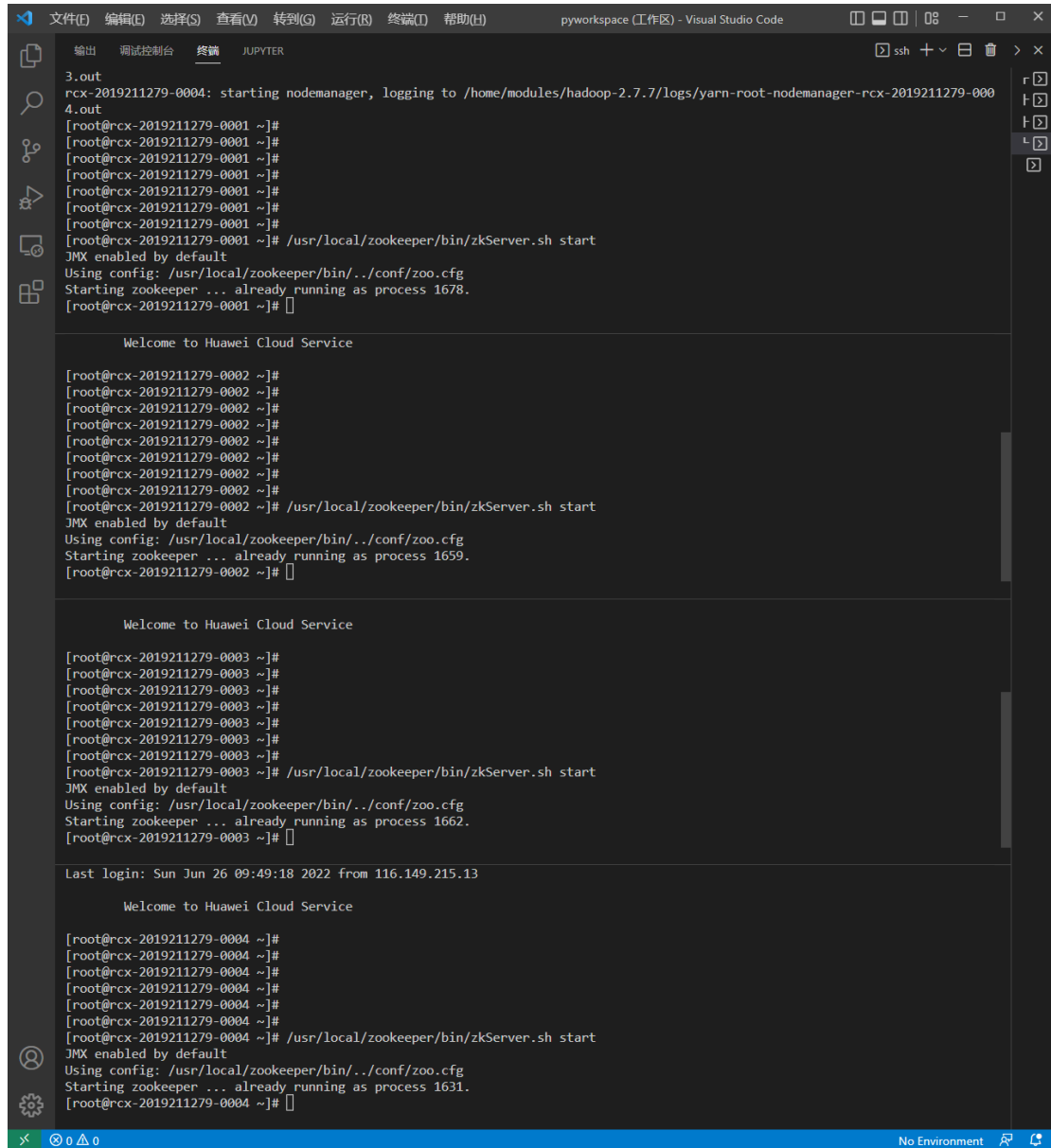


五、运行简易推荐系统

5.1 启动 HDFS



5.2 启动 zookeeper (所有节点)



```
3.out
rcx-2019211279-0004: starting nodemanager, logging to /home/modules/hadoop-2.7.7/logs/yarn-root-nodemanager-rcx-2019211279-000
4.out
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/../conf/zoo.cfg
Starting zookeeper ... already running as process 1678.
[root@rcx-2019211279-0001 ~]#

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/../conf/zoo.cfg
Starting zookeeper ... already running as process 1659.
[root@rcx-2019211279-0002 ~]#

Welcome to Huawei Cloud Service

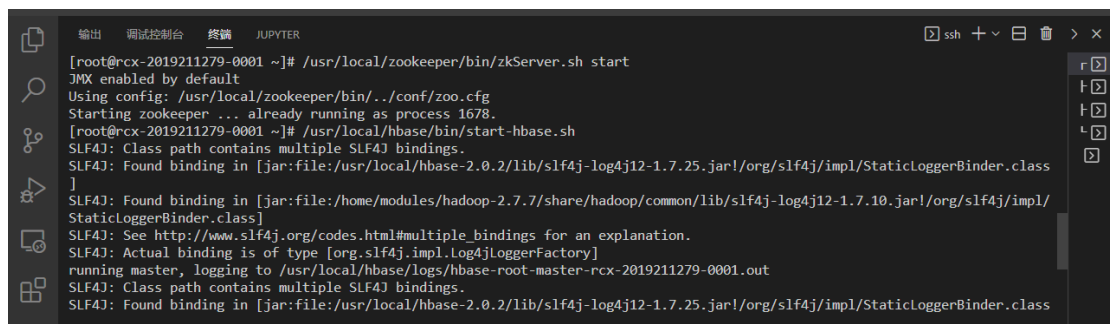
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/../conf/zoo.cfg
Starting zookeeper ... already running as process 1662.
[root@rcx-2019211279-0003 ~]#

Last login: Sun Jun 26 09:49:18 2022 from 116.149.215.13

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/../conf/zoo.cfg
Starting zookeeper ... already running as process 1631.
[root@rcx-2019211279-0004 ~]#
```

5.3 启动 HBase



```
[root@rcx-2019211279-0001 ~]# /usr/local/zookeeper/bin/zkServer.sh start
JMX enabled by default
Using config: /usr/local/zookeeper/bin/../conf/zoo.cfg
Starting zookeeper ... already running as process 1678.
[root@rcx-2019211279-0001 ~]# /usr/local/hbase/bin/start-hbase.sh
SLF4J: Found binding in [jar:file:/usr/local/hbase-2.0.2/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/modules/hadoop-2.7.7/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
running master, logging to /usr/local/hbase/logs/hbase-root-master-rcx-2019211279-0001.out
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase-2.0.2/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
```

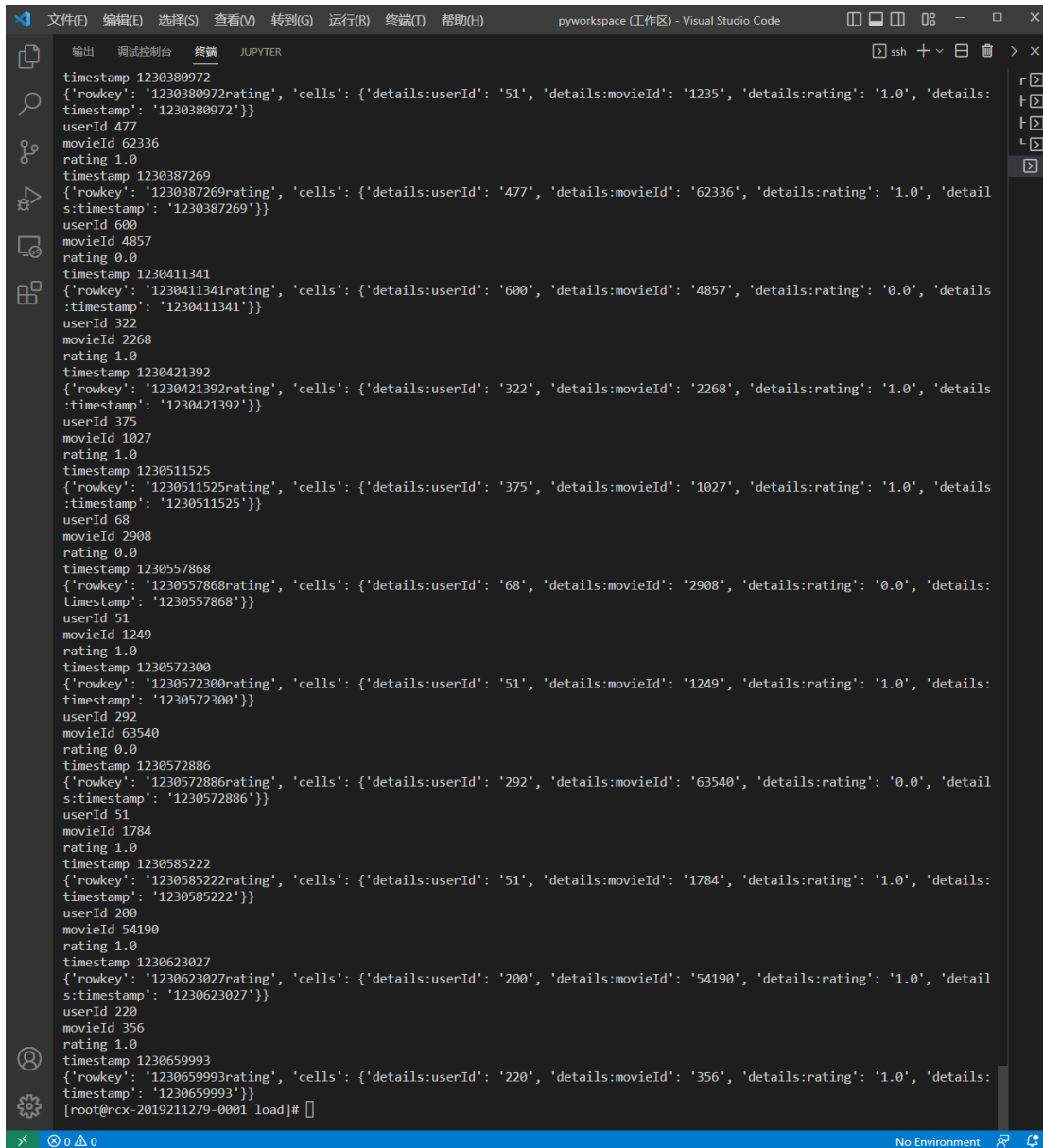
5.4 配置 HBase Thrift 连接，以便 python 中的 happybase 库能够连接 Hbase

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER ssh + - [ ] > x
.out
rcx-2019211279-0004: running regionserver, logging to /usr/local/hbase/bin/../logs/hbase-root-regionserver-rcx-2019211279-0004
.out
rcx-2019211279-0003: running regionserver, logging to /usr/local/hbase/bin/../logs/hbase-root-regionserver-rcx-2019211279-0003
.out
[root@rcx-2019211279-0001 ~]# /usr/local/hbase/bin/hbase-daemon.sh start thrift
running thrift, logging to /usr/local/hbase/logs/hbase-root-thrift-rcx-2019211279-0001.out
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase-2.0.2/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class
]
SLF4J: Found binding in [jar:file:/home/modules/hadoop-2.7.7/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
[root@rcx-2019211279-0001 ~]#
```

5.5 在 HBase 中创建对应的表

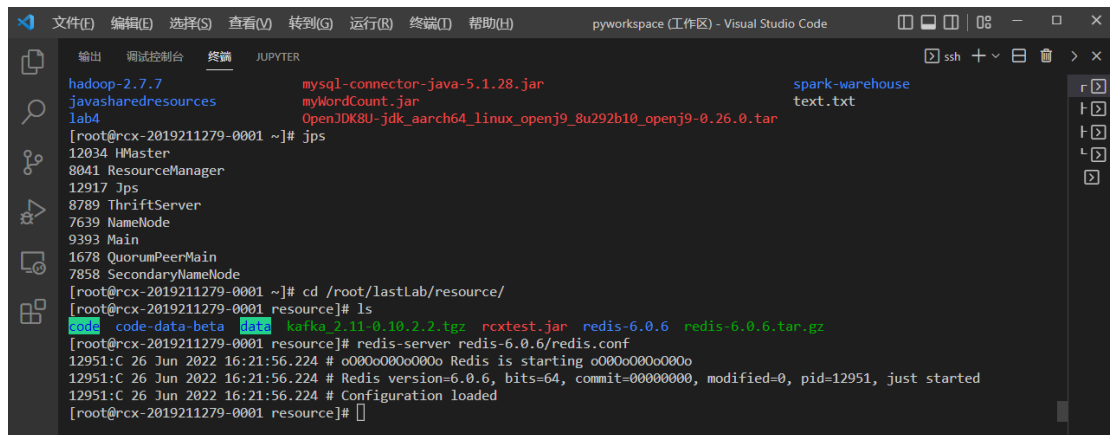
```
2022-06-26 16:03:36,683 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bu
iltin-java classes where applicable
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 2.0.2, r1cfab033e779df840d5612a85277f42a6a4e8172, Tue Aug 28 20:50:40 PDT 2018
Took 0.0084 seconds
hbase(main):001:0> create 'movie_records','details'
Created table movie_records
Took 2.0322 seconds
=> Hbase::Table - movie_records
hbase(main):002:0> scan 'movie_records'
ROW COLUMN+CELL
0 row(s)
Took 0.1525 seconds
hbase(main):003:0> list
TABLE
movie_records
1 row(s)
Took 0.0388 seconds
=> ["movie_records"]
hbase(main):004:0> exit
[root@rcx-2019211279-0001 ~]#
```


5.6. 启动 load_train_ratings_hbase.py (需要运行完)



```
timestamp 1230380972
{'rowkey': '1230380972rating', 'cells': {'details:userId': '51', 'details:movieId': '1235', 'details:rating': '1.0', 'details:timestamp': '1230380972'}}
userId 477
movieId 62336
rating 1.0
timestamp 1230387269
{'rowkey': '1230387269rating', 'cells': {'details:userId': '477', 'details:movieId': '62336', 'details:rating': '1.0', 'details:timestamp': '1230387269'}}
userId 600
movieId 4857
rating 0.0
timestamp 1230411341
{'rowkey': '1230411341rating', 'cells': {'details:userId': '600', 'details:movieId': '4857', 'details:rating': '0.0', 'details:timestamp': '1230411341'}}
userId 322
movieId 2268
rating 1.0
timestamp 1230421392
{'rowkey': '1230421392rating', 'cells': {'details:userId': '322', 'details:movieId': '2268', 'details:rating': '1.0', 'details:timestamp': '1230421392'}}
userId 375
movieId 1027
rating 1.0
timestamp 1230511525
{'rowkey': '1230511525rating', 'cells': {'details:userId': '375', 'details:movieId': '1027', 'details:rating': '1.0', 'details:timestamp': '1230511525'}}
userId 68
movieId 2908
rating 0.0
timestamp 1230557868
{'rowkey': '1230557868rating', 'cells': {'details:userId': '68', 'details:movieId': '2908', 'details:rating': '0.0', 'details:timestamp': '1230557868'}}
userId 51
movieId 1249
rating 1.0
timestamp 1230572300
{'rowkey': '1230572300rating', 'cells': {'details:userId': '51', 'details:movieId': '1249', 'details:rating': '1.0', 'details:timestamp': '1230572300'}}
userId 292
movieId 63540
rating 0.0
timestamp 1230572886
{'rowkey': '1230572886rating', 'cells': {'details:userId': '292', 'details:movieId': '63540', 'details:rating': '0.0', 'details:timestamp': '1230572886'}}
userId 51
movieId 1784
rating 1.0
timestamp 1230585222
{'rowkey': '1230585222rating', 'cells': {'details:userId': '51', 'details:movieId': '1784', 'details:rating': '1.0', 'details:timestamp': '1230585222'}}
userId 200
movieId 54190
rating 1.0
timestamp 1230623027
{'rowkey': '1230623027rating', 'cells': {'details:userId': '200', 'details:movieId': '54190', 'details:rating': '1.0', 'details:timestamp': '1230623027'}}
userId 220
movieId 356
rating 1.0
timestamp 1230659993
{'rowkey': '1230659993rating', 'cells': {'details:userId': '220', 'details:movieId': '356', 'details:rating': '1.0', 'details:timestamp': '1230659993'}}
[root@rcx-2019211279-0001 load]#
```

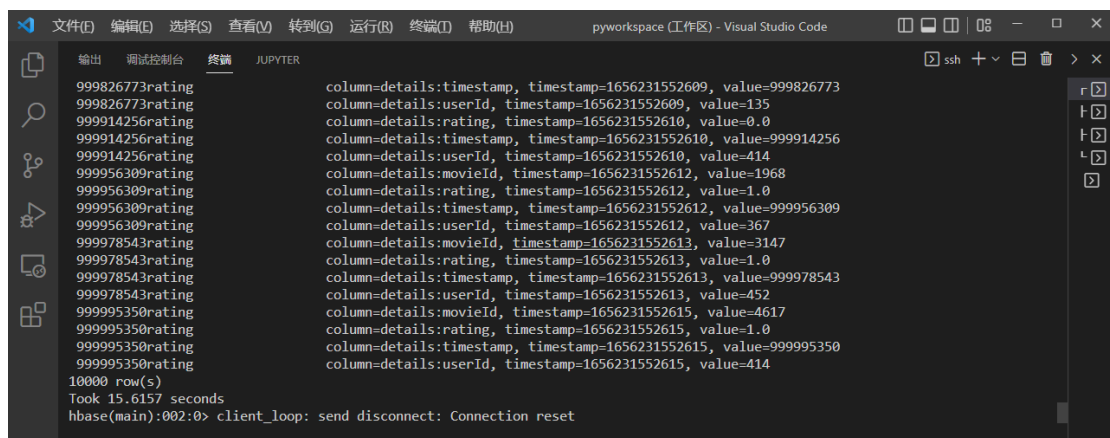
5.7 启动 redis



```
hadoop-2.7.7      mysql-connector-java-5.1.28.jar      spark-warehouse
javasharedresources  myWordCount.jar      text.txt
lab4      OpenJDK8U-jdk_aarch64_linux_openj9_8u292b10_openj9-0.26.0.tar

[root@rcx-2019211279-0001 ~]# jps
12034 HMaster
8041 ResourceManager
12917 Jps
8789 ThriftServer
7639 NameNode
9393 Main
1678 QuorumPeerMain
7858 SecondaryNameNode
[root@rcx-2019211279-0001 ~]# cd /root/lastLab/resource/
[root@rcx-2019211279-0001 resource]# ls
code  code-data-beta  data  kafka_2.11-0.10.2.2.tgz  rcxtest.jar  redis-6.0.6  redis-6.0.6.tar.gz
[root@rcx-2019211279-0001 resource]# redis-server redis-6.0.6/redis.conf
12951:C 26 Jun 2022 16:21:56.224 # oO0oO00oO00o Redis is starting oO0oO00oO00o
12951:C 26 Jun 2022 16:21:56.224 # Redis version=6.0.6, bits=64, commit=00000000, modified=0, pid=12951, just started
12951:C 26 Jun 2022 16:21:56.224 # Configuration loaded
[root@rcx-2019211279-0001 resource]#
```

5.8 启动 load_movie_redis.py（需要运行完）



```
999826773rating      column=details:timestamp, timestamp=1656231552609, value=999826773
999826773rating      column=details:userId, timestamp=1656231552609, value=135
999914256rating      column=details:rating, timestamp=1656231552610, value=0.0
999914256rating      column=details:timestamp, timestamp=1656231552610, value=999914256
999914256rating      column=details:userId, timestamp=1656231552610, value=414
999956309rating      column=details:movieId, timestamp=1656231552612, value=1968
999956309rating      column=details:rating, timestamp=1656231552612, value=1.0
999956309rating      column=details:timestamp, timestamp=1656231552612, value=999956309
999956309rating      column=details:userId, timestamp=1656231552612, value=367
999978543rating      column=details:movieId, timestamp=1656231552613, value=3147
999978543rating      column=details:rating, timestamp=1656231552613, value=1.0
999978543rating      column=details:timestamp, timestamp=1656231552613, value=999978543
999978543rating      column=details:userId, timestamp=1656231552613, value=452
999995350rating      column=details:movieId, timestamp=1656231552615, value=4617
999995350rating      column=details:rating, timestamp=1656231552615, value=1.0
999995350rating      column=details:timestamp, timestamp=1656231552615, value=999995350
999995350rating      column=details:userId, timestamp=1656231552615, value=414
10000 row(s)
Took 15.6157 seconds
hbase(main):002:0> client_loop: send disconnect: Connection reset
```

5.9 启动 Kafka 并创建 Kafka Topic

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H)    pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER
7639 NameNode
14500 Kafka
1678 QuorumPeerMain
7858 SecondaryNameNode
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]#
[root@rcx-2019211279-0001 ~]# /home/modules/kafka_2.11-0.10.2.2/bin/kafka-topics.sh --zookeeper rcx-2019211279-0001:2181 --cre
ate --topic movie_rating_records --partitions 1 --replication-factor 1
WARNING: Due to limitations in metric names, topics with a period('.') or underscore('_') could collide. To avoid issues it
is best to use either, but not both.
Created topic "movie_rating_records".
[root@rcx-2019211279-0001 ~]#

PS F:\vs code\py_workspace> ssh root@rcx-2019211279-0002
Last login: Sun Jun 26 18:50:57 2022 from 116.149.183.138

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0002 ~]# jps
3884 Kafka
4203 Jps
2308 NodeManager
2188 DataNode
3105 HRegionServer
1659 QuorumPeerMain
[root@rcx-2019211279-0002 ~]#
[root@rcx-2019211279-0002 ~]#

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]# jps
2322 NodeManager
3836 Kafka
3114 HRegionServer
4148 Jps
2202 DataNode
1662 QuorumPeerMain
[root@rcx-2019211279-0003 ~]#
[root@rcx-2019211279-0003 ~]#

Last login: Sun Jun 26 18:54:19 2022 from 116.149.183.138

Welcome to Huawei Cloud Service

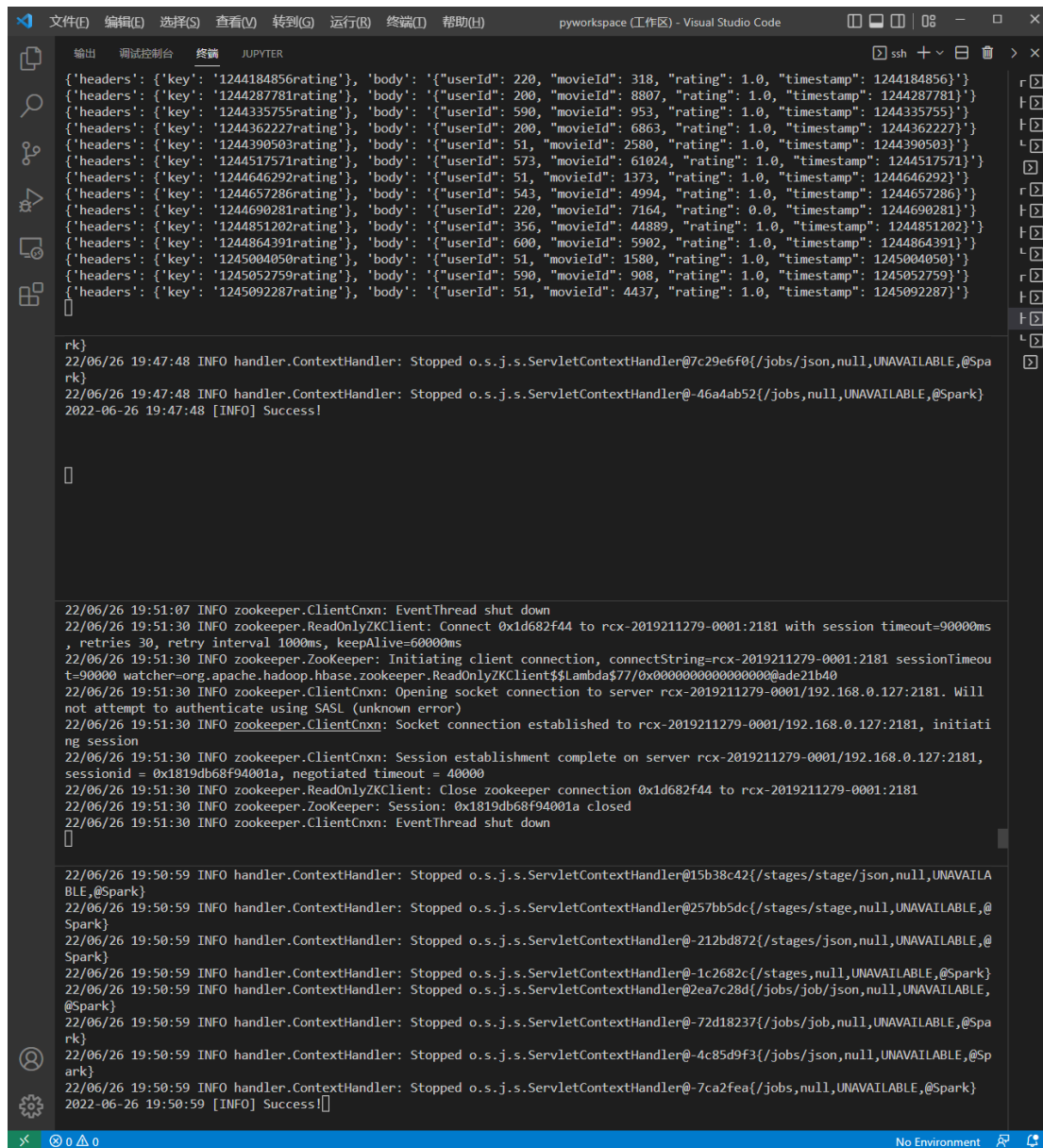
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]# jps
3139 HRegionServer
4276 Jps
2307 NodeManager
2187 DataNode
1631 QuorumPeerMain
3962 Kafka
[root@rcx-2019211279-0004 ~]#
[root@rcx-2019211279-0004 ~]#

193587 Bungo Stray Dogs: Dead Apple (2018)
193609 Andrew Dice Clay: Dice Rules (1991)
Load .././data/movies.csv success!
[root@rcx-2019211279-0001 load]#
```

5.10 启动 generatorRecord.py

```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER
load server-client spark-sparkstreaming-recommend
[root@rcx-2019211279-0001 code]# cd load/
[root@rcx-2019211279-0001 load]# ls
generatorRecord.py load_movie_redis.py load_train_ratings hbase.py
[root@rcx-2019211279-0001 load]# python3 generatorRecord.py -h rcx-2019211279-0001:9092 -f ../../data/json_test_ratings.json
Traceback (most recent call last):
  File "generatorRecord.py", line 11, in <module>
    from kafka import KafkaProducer
ModuleNotFoundError: No module named 'kafka'
[root@rcx-2019211279-0001 load]# pip3 install kafka
Collecting kafka
  Downloading kafka-1.3.5-py2.py3-none-any.whl (207 kB)
    | 207 kB 24 kB/s
Installing collected packages: kafka
Successfully installed kafka-1.3.5
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv
[root@rcx-2019211279-0001 load]#
[root@rcx-2019211279-0001 load]#
[root@rcx-2019211279-0001 load]# python3 generatorRecord.py -h rcx-2019211279-0001:9092 -f ../../data/json_test_ratings.json
rcx-2019211279-0001:9092 ../../data/json_test_ratings.json
rcx-2019211279-0001:9092
{'headers': {'key': '1230673202rating'}, 'body': '{"userId": 51, "movieId": 59615, "rating": 1.0, "timestamp": 1230673202}'},
{'headers': {'key': '1230843298rating'}, 'body': '{"userId": 200, "movieId": 32587, "rating": 1.0, "timestamp": 1230843298}'},
{'headers': {'key': '1230866243rating'}, 'body': '{"userId": 51, "movieId": 8874, "rating": 1.0, "timestamp": 1230866243}'},
{'headers': {'key': '1230934983rating'}, 'body': '{"userId": 53, "movieId": 1441, "rating": 1.0, "timestamp": 1230934983}'},
{'headers': {'key': '1230963272rating'}, 'body': '{"userId": 390, "movieId": 1217, "rating": 1.0, "timestamp": 1230963272}'},
{'headers': {'key': '1231040959rating'}, 'body': '{"userId": 600, "movieId": 39444, "rating": 0.0, "timestamp": 1231040959}'},
{'headers': {'key': '1231063111rating'}, 'body': '{"userId": 375, "movieId": 48385, "rating": 1.0, "timestamp": 1231063111}'},
{'headers': {'key': '1231113698rating'}, 'body': '{"userId": 600, "movieId": 2840, "rating": 0.0, "timestamp": 1231113698}'},
{'headers': {'key': '1231135164rating'}, 'body': '{"userId": 239, "movieId": 260, "rating": 1.0, "timestamp": 1231135164}'},
{'headers': {'key': '1231214050rating'}, 'body': '{"userId": 390, "movieId": 6377, "rating": 1.0, "timestamp": 1231214050}'},
{'headers': {'key': '1231220011rating'}, 'body': '{"userId": 53, "movieId": 2686, "rating": 1.0, "timestamp": 1231220011}'},
{'headers': {'key': '1231265524rating'}, 'body': '{"userId": 219, "movieId": 2628, "rating": 0.0, "timestamp": 1231265524}'},
{'headers': {'key': '1231438908rating'}, 'body': '{"userId": 51, "movieId": 4014, "rating": 1.0, "timestamp": 1231438908}'},
{'headers': {'key': '1231441227rating'}, 'body': '{"userId": 223, "movieId": 1342, "rating": 0.0, "timestamp": 1231441227}'},
{'headers': {'key': '1231573897rating'}, 'body': '{"userId": 600, "movieId": 1199, "rating": 1.0, "timestamp": 1231573897}'},
{'headers': {'key': '1231600566rating'}, 'body': '{"userId": 356, "movieId": 35836, "rating": 1.0, "timestamp": 1231600566}'},
{'headers': {'key': '1231651501rating'}, 'body': '{"userId": 573, "movieId": 2918, "rating": 1.0, "timestamp": 1231651501}'},
{'headers': {'key': '1231679240rating'}, 'body': '{"userId": 600, "movieId": 3248, "rating": 0.0, "timestamp": 1231679240}'},
{'headers': {'key': '1231688474rating'}, 'body': '{"userId": 220, "movieId": 260, "rating": 1.0, "timestamp": 1231688474}'},
{'headers': {'key': '1231781595rating'}, 'body': '{"userId": 200, "movieId": 1380, "rating": 1.0, "timestamp": 1231781595}'},
{'headers': {'key': '1231862532rating'}, 'body': '{"userId": 553, "movieId": 5513, "rating": 1.0, "timestamp": 1231862532}'},
{'headers': {'key': '1231862566rating'}, 'body': '{"userId": 51, "movieId": 2183, "rating": 1.0, "timestamp": 1231862566}'},
{'headers': {'key': '1231907123rating'}, 'body': '{"userId": 239, "movieId": 1704, "rating": 1.0, "timestamp": 1231907123}'},
{'headers': {'key': '1231955833rating'}, 'body': '{"userId": 219, "movieId": 63082, "rating": 1.0, "timestamp": 1231955833}'},
{'headers': {'key': '1232005148rating'}, 'body': '{"userId": 318, "movieId": 1884, "rating": 1.0, "timestamp": 1232005148}'},
{'headers': {'key': '1232025085rating'}, 'body': '{"userId": 51, "movieId": 1243, "rating": 1.0, "timestamp": 1232025085}'},
{'headers': {'key': '1232048579rating'}, 'body': '{"userId": 220, "movieId": 2716, "rating": 1.0, "timestamp": 1232048579}'},
{'headers': {'key': '1232128184rating'}, 'body': '{"userId": 356, "movieId": 2918, "rating": 1.0, "timestamp": 1232128184}'},
{'headers': {'key': '1232138195rating'}, 'body': '{"userId": 356, "movieId": 58559, "rating": 1.0, "timestamp": 1232138195}'},
{'headers': {'key': '1232156555rating'}, 'body': '{"userId": 239, "movieId": 475, "rating": 1.0, "timestamp": 1232156555}'},
{'headers': {'key': '1232261718rating'}, 'body': '{"userId": 600, "movieId": 4642, "rating": 1.0, "timestamp": 1232261718}'},
{'headers': {'key': '1232290680rating'}, 'body': '{"userId": 220, "movieId": 7147, "rating": 1.0, "timestamp": 1232290680}'},
{'headers': {'key': '1232332121rating'}, 'body': '{"userId": 600, "movieId": 25947, "rating": 1.0, "timestamp": 1232332121}'},
{'headers': {'key': '1232346833rating'}, 'body': '{"userId": 600, "movieId": 2193, "rating": 1.0, "timestamp": 1232346833}'},
{'headers': {'key': '1232509252rating'}, 'body': '{"userId": 51, "movieId": 5816, "rating": 1.0, "timestamp": 1232509252}'},
{'headers': {'key': '1232639166rating'}, 'body': '{"userId": 600, "movieId": 371, "rating": 0.0, "timestamp": 1232639166}'},
{'headers': {'key': '1232711497rating'}, 'body': '{"userId": 600, "movieId": 41566, "rating": 0.0, "timestamp": 1232711497}'},
{'headers': {'key': '1232729275rating'}, 'body': '{"userId": 51, "movieId": 1967, "rating": 1.0, "timestamp": 1232729275}'},
{'headers': {'key': '1232733235rating'}, 'body': '{"userId": 239, "movieId": 6502, "rating": 1.0, "timestamp": 1232733235}'}
```

5.11 启动 hbase2spark、kafkaStreaming、recommend



```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H) pyworkspace (工作区) - Visual Studio Code

输出 调试控制台 终端 JUPYTER

[{"headers": {"key": "1244184856rating"}, "body": {"userId": 220, "movieId": 318, "rating": 1.0, "timestamp": 1244184856}}, {"headers": {"key": "1244287781rating"}, "body": {"userId": 200, "movieId": 8807, "rating": 1.0, "timestamp": 1244287781}}, {"headers": {"key": "1244335755rating"}, "body": {"userId": 590, "movieId": 953, "rating": 1.0, "timestamp": 1244335755}}, {"headers": {"key": "1244362227rating"}, "body": {"userId": 200, "movieId": 6863, "rating": 1.0, "timestamp": 1244362227}}, {"headers": {"key": "1244390503rating"}, "body": {"userId": 51, "movieId": 2580, "rating": 1.0, "timestamp": 1244390503}}, {"headers": {"key": "1244517571rating"}, "body": {"userId": 573, "movieId": 61024, "rating": 1.0, "timestamp": 1244517571}}, {"headers": {"key": "1244646292rating"}, "body": {"userId": 51, "movieId": 1373, "rating": 1.0, "timestamp": 1244646292}}, {"headers": {"key": "1244657286rating"}, "body": {"userId": 543, "movieId": 4994, "rating": 1.0, "timestamp": 1244657286}}, {"headers": {"key": "1244690281rating"}, "body": {"userId": 220, "movieId": 7164, "rating": 0.0, "timestamp": 1244690281}}, {"headers": {"key": "1244851202rating"}, "body": {"userId": 356, "movieId": 44889, "rating": 1.0, "timestamp": 1244851202}}, {"headers": {"key": "1244864391rating"}, "body": {"userId": 600, "movieId": 5902, "rating": 1.0, "timestamp": 1244864391}}, {"headers": {"key": "1245004050rating"}, "body": {"userId": 51, "movieId": 1580, "rating": 1.0, "timestamp": 1245004050}}, {"headers": {"key": "1245052759rating"}, "body": {"userId": 590, "movieId": 908, "rating": 1.0, "timestamp": 1245052759}}, {"headers": {"key": "1245092287rating"}, "body": {"userId": 51, "movieId": 4437, "rating": 1.0, "timestamp": 1245092287}}]

rk}
22/06/26 19:47:48 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@7c29e6f0[/jobs/json,null,UNAVAILABLE,@Spark]
22/06/26 19:47:48 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@-46a4ab52[/jobs,null,UNAVAILABLE,@Spark]
2022-06-26 19:47:48 [INFO] Success!

[]

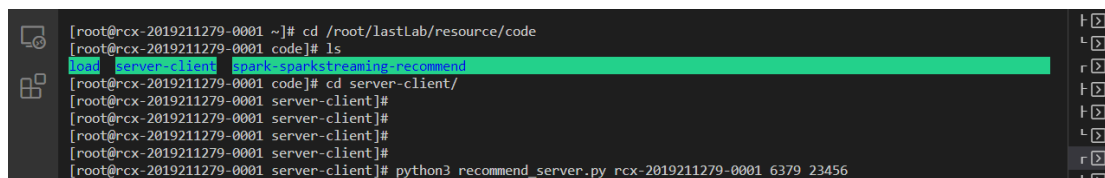
22/06/26 19:51:07 INFO zookeeper.ClientCnxn: EventThread shut down
22/06/26 19:51:30 INFO zookeeper.ReadOnlyZKClient: Connect 0x1d682f44 to rcx-2019211279-0001:2181 with session timeout=90000ms, retries 30, retry interval 1000ms, keepAlive=60000ms
22/06/26 19:51:30 INFO zookeeper.ZooKeeper: Initiating client connection, connectString=rcx-2019211279-0001:2181 sessionTimeout=90000 watcher=org.apache.hadoop.hbase.zookeeper.ReadOnlyZKClient$$Lambda$77/0x0000000000000000@ade21b40
22/06/26 19:51:30 INFO zookeeper.ClientCnxn: Opening socket connection to server rcx-2019211279-0001/192.168.0.127:2181. Will not attempt to authenticate using SASL (unknown error)
22/06/26 19:51:30 INFO zookeeper.ClientCnxn: Socket connection established to rcx-2019211279-0001/192.168.0.127:2181, initiating session
22/06/26 19:51:30 INFO zookeeper.ClientCnxn: Session establishment complete on server rcx-2019211279-0001/192.168.0.127:2181, sessionId = 0x1819db68f94001a, negotiated timeout = 40000
22/06/26 19:51:30 INFO zookeeper.ReadOnlyZKClient: Close zookeeper connection 0x1d682f44 to rcx-2019211279-0001:2181
22/06/26 19:51:30 INFO zookeeper.ZooKeeper: Session: 0x1819db68f94001a closed
22/06/26 19:51:30 INFO zookeeper.ClientCnxn: EventThread shut down

[]

22/06/26 19:50:59 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@15b38c42[/stages/stage/json,null,UNAVAILABLE,@Spark]
22/06/26 19:50:59 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@257bb5dc[/stages/stage,null,UNAVAILABLE,@Spark]
22/06/26 19:50:59 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@-212bd872[/stages/json,null,UNAVAILABLE,@Spark]
22/06/26 19:50:59 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@-1c2682c[/stages,null,UNAVAILABLE,@Spark]
22/06/26 19:50:59 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@2ea7c28d[/jobs/job/json,null,UNAVAILABLE,@Spark]
22/06/26 19:50:59 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@-72d18237[/jobs/job,null,UNAVAILABLE,@Spark]
22/06/26 19:50:59 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@-4c85d9f3[/jobs/json,null,UNAVAILABLE,@Spark]
22/06/26 19:50:59 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@-7ca2fea[/jobs,null,UNAVAILABLE,@Spark]
2022-06-26 19:50:59 [INFO] Success![]

No Environment
```

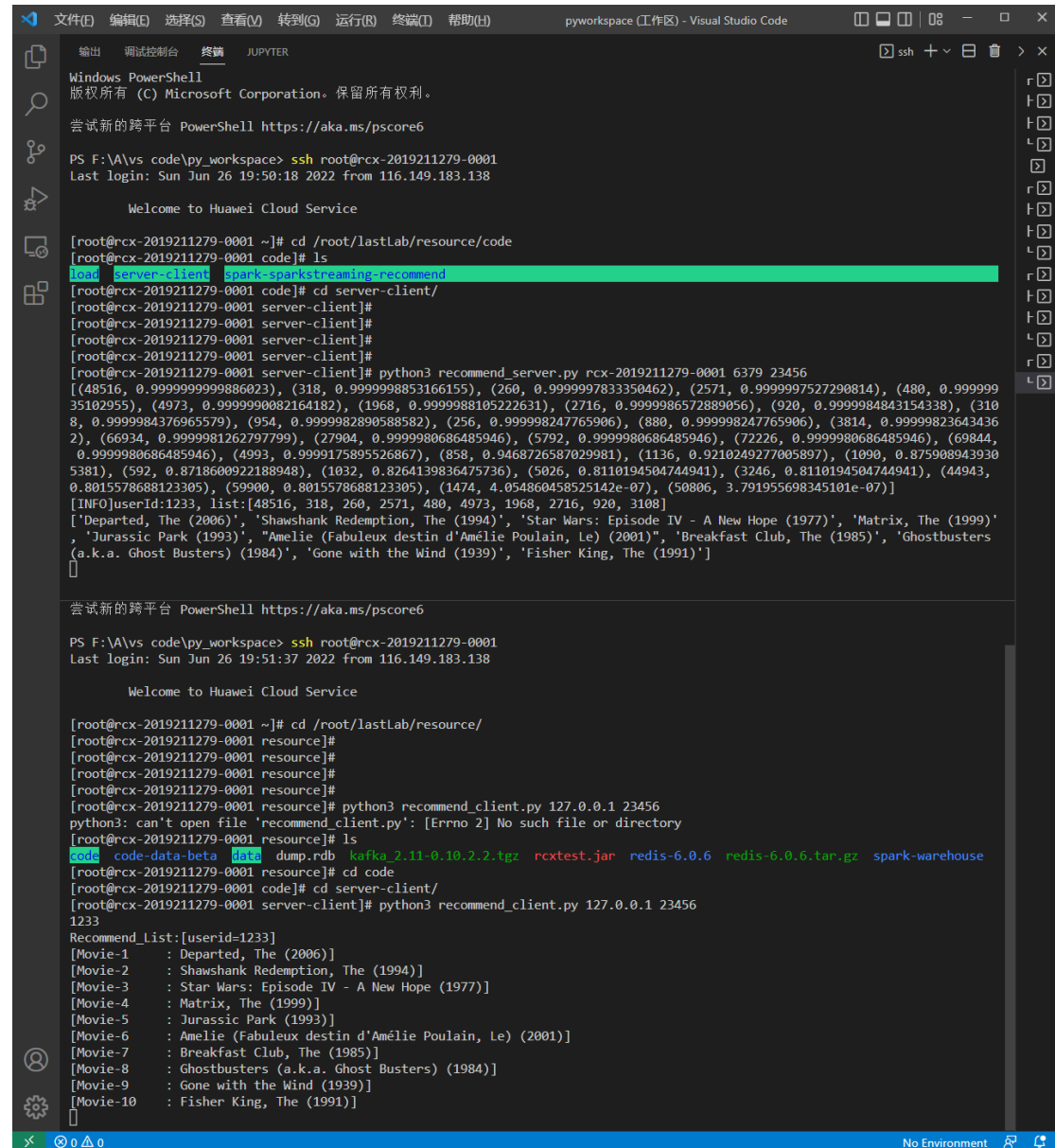
5.12 启动 recommend_server.py



```
[root@rcx-2019211279-0001 ~]# cd /root/lastLab/resource/code
[root@rcx-2019211279-0001 code]# ls
load server-client spark-sparkstreaming-recommend
[root@rcx-2019211279-0001 code]# cd server-client/
[root@rcx-2019211279-0001 server-client]#
[root@rcx-2019211279-0001 server-client]#
[root@rcx-2019211279-0001 server-client]#
[root@rcx-2019211279-0001 server-client]# python3 recommend_server.py rcx-2019211279-0001 6379 23456
```

5.13 启动 recommend_client.py

```
[root@rcx-2019211279-0001 ~]# cd /root/lastLab/resource/
[root@rcx-2019211279-0001 resource]#
[root@rcx-2019211279-0001 resource]#
[root@rcx-2019211279-0001 resource]#
[root@rcx-2019211279-0001 resource]#
[root@rcx-2019211279-0001 resource]# python3 recommend_client.py 127.0.0.1 23456
python3: can't open file 'recommend_client.py': [Errno 2] No such file or directory
[root@rcx-2019211279-0001 resource]# ls
code  code-data-beta  data  dump.rdb  kafka_2.11-0.10.2.2.tgz  rcxtest.jar  redis-6.0.6  redis-6.0.6.tar.gz  spark-warehouse
[root@rcx-2019211279-0001 resource]# cd code
[root@rcx-2019211279-0001 code]# cd server-client/
[root@rcx-2019211279-0001 server-client]# python3 recommend_client.py 127.0.0.1 23456
```



```
文件(F) 编辑(E) 选择(S) 查看(V) 转到(G) 运行(R) 终端(T) 帮助(H)  pyworkspace (工作区) - Visual Studio Code
输出 调试控制台 终端 JUPYTER
Windows PowerShell
版权所有 (C) Microsoft Corporation。保留所有权利。

尝试新的跨平台 PowerShell https://aka.ms/pscore6

PS F:\A\vs code\py_workspace> ssh root@rcx-2019211279-0001
Last login: Sun Jun 26 19:50:18 2022 from 116.149.183.138

Welcome to Huawei Cloud Service

[root@rcx-2019211279-0001 ~]# cd /root/lastLab/resource/code
[root@rcx-2019211279-0001 code]# ls
load  server-client  spark-sparkstreaming-recommend
[root@rcx-2019211279-0001 code]# cd server-client/
[root@rcx-2019211279-0001 server-client]#
[root@rcx-2019211279-0001 server-client]#
[root@rcx-2019211279-0001 server-client]#
[root@rcx-2019211279-0001 server-client]#
[root@rcx-2019211279-0001 server-client]# python3 recommend_server.py rcx-2019211279-0001 6379 23456
[[{"id": 48516, "score": 0.999999999886023}, {"id": 318, "score": 0.999998853166155}, {"id": 260, "score": 0.999997833350462}, {"id": 2571, "score": 0.999997527290814}, {"id": 480, "score": 0.9999935102955}, {"id": 4973, "score": 0.999990082164182}, {"id": 1968, "score": 0.999988105222631}, {"id": 2716, "score": 0.999986572889056}, {"id": 920, "score": 0.999984843154338}, {"id": 3108, "score": 0.999984376965579}, {"id": 954, "score": 0.999982890588582}, {"id": 256, "score": 0.99998247765906}, {"id": 880, "score": 0.99998247765906}, {"id": 3814, "score": 0.99998236434362}, {"id": 66934, "score": 0.999981262797799}, {"id": 27904, "score": 0.999980686485946}, {"id": 5792, "score": 0.999980686485946}, {"id": 72226, "score": 0.999980686485946}, {"id": 69844, "score": 0.999980686485946}, {"id": 4993, "score": 0.9999175895526867}, {"id": 858, "score": 0.9468726587029981}, {"id": 1136, "score": 0.9210249277005897}, {"id": 1090, "score": 0.8759089439305381}, {"id": 592, "score": 0.871860922188948}, {"id": 1032, "score": 0.8264139836475736}, {"id": 5026, "score": 0.8110194504744941}, {"id": 3246, "score": 0.8110194504744941}, {"id": 44943, "score": 0.8015578688123305}, {"id": 59900, "score": 0.8015578688123305}, {"id": 1474, "score": 4.054860458525142e-07}, {"id": 50806, "score": 3.791955698345101e-07}]]
[INFO]userId:1233, list:[{"id": 48516, "score": 0.999999999886023}, {"id": 318, "score": 0.999998853166155}, {"id": 260, "score": 0.999997833350462}, {"id": 2571, "score": 0.999997527290814}, {"id": 480, "score": 0.9999935102955}, {"id": 4973, "score": 0.999990082164182}, {"id": 1968, "score": 0.999988105222631}, {"id": 2716, "score": 0.999986572889056}, {"id": 920, "score": 0.999984843154338}, {"id": 3108, "score": 0.999984376965579}, {"id": 954, "score": 0.999982890588582}, {"id": 256, "score": 0.99998247765906}, {"id": 880, "score": 0.99998247765906}, {"id": 3814, "score": 0.99998236434362}, {"id": 66934, "score": 0.999981262797799}, {"id": 27904, "score": 0.999980686485946}, {"id": 5792, "score": 0.999980686485946}, {"id": 72226, "score": 0.999980686485946}, {"id": 69844, "score": 0.999980686485946}, {"id": 4993, "score": 0.9999175895526867}, {"id": 858, "score": 0.9468726587029981}, {"id": 1136, "score": 0.9210249277005897}, {"id": 1090, "score": 0.8759089439305381}, {"id": 592, "score": 0.871860922188948}, {"id": 1032, "score": 0.8264139836475736}, {"id": 5026, "score": 0.8110194504744941}, {"id": 3246, "score": 0.8110194504744941}, {"id": 44943, "score": 0.8015578688123305}, {"id": 59900, "score": 0.8015578688123305}, {"id": 1474, "score": 4.054860458525142e-07}, {"id": 50806, "score": 3.791955698345101e-07}]]
["Departed, The (2006)", "Shawshank Redemption, The (1994)", "Star Wars: Episode IV - A New Hope (1977)", "Matrix, The (1999)", "Jurassic Park (1993)", "Amelie (Fabuleux destin d'Am\u00e9lie Poulain, Le) (2001)", "Breakfast Club, The (1985)", "Ghostbusters (a.k.a. Ghost Busters) (1984)", "Gone with the Wind (1939)", "Fisher King, The (1991)"]

尝试新的跨平台 PowerShell https://aka.ms/pscore6

PS F:\A\vs code\py_workspace> ssh root@rcx-2019211279-0001
Last login: Sun Jun 26 19:51:37 2022 from 116.149.183.138

Welcome to Huawei Cloud Service

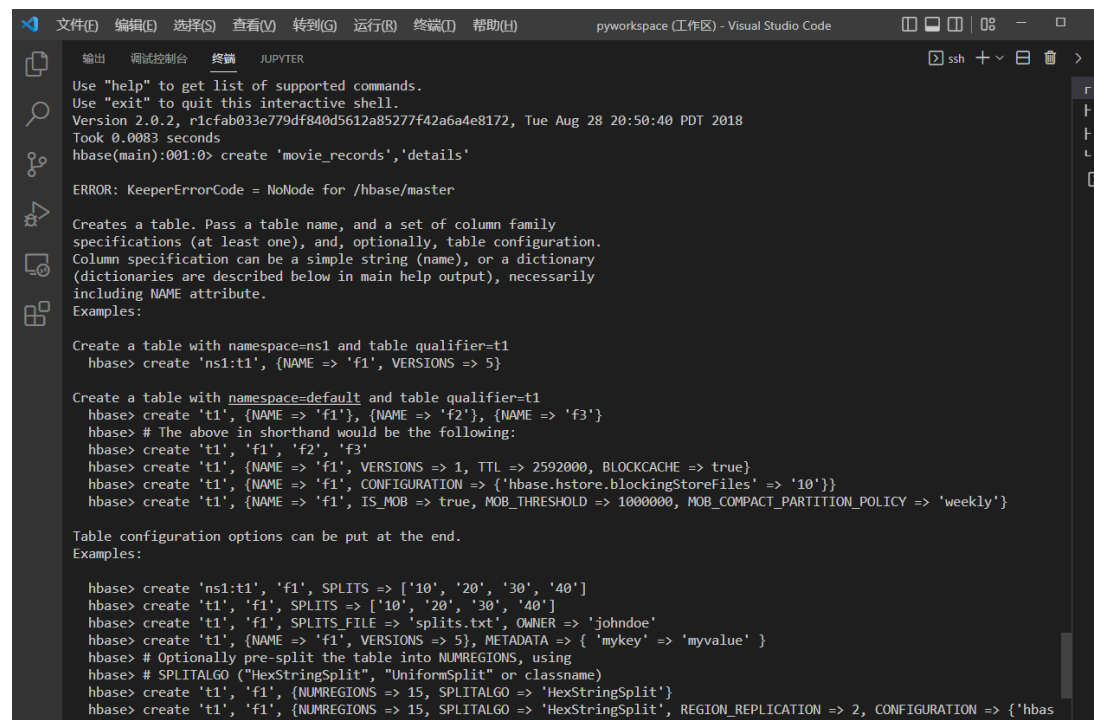
[root@rcx-2019211279-0001 ~]# cd /root/lastLab/resource/
[root@rcx-2019211279-0001 resource]#
[root@rcx-2019211279-0001 resource]#
[root@rcx-2019211279-0001 resource]#
[root@rcx-2019211279-0001 resource]#
[root@rcx-2019211279-0001 resource]# python3 recommend_client.py 127.0.0.1 23456
python3: can't open file 'recommend_client.py': [Errno 2] No such file or directory
[root@rcx-2019211279-0001 resource]# ls
code  code-data-beta  data  dump.rdb  kafka_2.11-0.10.2.2.tgz  rcxtest.jar  redis-6.0.6  redis-6.0.6.tar.gz  spark-warehouse
[root@rcx-2019211279-0001 resource]# cd code
[root@rcx-2019211279-0001 code]# cd server-client/
[root@rcx-2019211279-0001 server-client]# python3 recommend_client.py 127.0.0.1 23456
1233
Recommend_List:[{"id": 48516, "score": 0.999999999886023}, {"id": 318, "score": 0.999998853166155}, {"id": 260, "score": 0.999997833350462}, {"id": 2571, "score": 0.999997527290814}, {"id": 480, "score": 0.9999935102955}, {"id": 4973, "score": 0.999990082164182}, {"id": 1968, "score": 0.999988105222631}, {"id": 2716, "score": 0.999986572889056}, {"id": 920, "score": 0.999984843154338}, {"id": 3108, "score": 0.999984376965579}, {"id": 954, "score": 0.999982890588582}, {"id": 256, "score": 0.99998247765906}, {"id": 880, "score": 0.99998247765906}, {"id": 3814, "score": 0.99998236434362}, {"id": 66934, "score": 0.999981262797799}, {"id": 27904, "score": 0.999980686485946}, {"id": 5792, "score": 0.999980686485946}, {"id": 72226, "score": 0.999980686485946}, {"id": 69844, "score": 0.999980686485946}, {"id": 4993, "score": 0.9999175895526867}, {"id": 858, "score": 0.9468726587029981}, {"id": 1136, "score": 0.9210249277005897}, {"id": 1090, "score": 0.8759089439305381}, {"id": 592, "score": 0.871860922188948}, {"id": 1032, "score": 0.8264139836475736}, {"id": 5026, "score": 0.8110194504744941}, {"id": 3246, "score": 0.8110194504744941}, {"id": 44943, "score": 0.8015578688123305}, {"id": 59900, "score": 0.8015578688123305}, {"id": 1474, "score": 4.054860458525142e-07}, {"id": 50806, "score": 3.791955698345101e-07}]]
["Movie-1 : Departed, The (2006)"]
["Movie-2 : Shawshank Redemption, The (1994)"]
["Movie-3 : Star Wars: Episode IV - A New Hope (1977)"]
["Movie-4 : Matrix, The (1999)"]
["Movie-5 : Jurassic Park (1993)"]
["Movie-6 : Amelie (Fabuleux destin d'Am\u00e9lie Poulain, Le) (2001)"]
["Movie-7 : Breakfast Club, The (1985)"]
["Movie-8 : Ghostbusters (a.k.a. Ghost Busters) (1984)"]
["Movie-9 : Gone with the Wind (1939)"]
["Movie-10 : Fisher King, The (1991)"]
```


实验结果：可以得到以下推荐电影

```
Recommend_List:[userid=1233]
[Movie-1      : Departed, The (2006)]
[Movie-2      : Shawshank Redemption, The (1994)]
[Movie-3      : Star Wars: Episode IV - A New Hope (1977)]
[Movie-4      : Matrix, The (1999)]
[Movie-5      : Jurassic Park (1993)]
[Movie-6      : Amelie (Fabuleux destin d'Amélie Poulain, Le) (2001)]
[Movie-7      : Breakfast Club, The (1985)]
[Movie-8      : Ghostbusters (a.k.a. Ghost Busters) (1984)]
[Movie-9      : Gone with the Wind (1939)]
[Movie-10     : Fisher King, The (1991)]
```

实验总结&心得体会：

由于时间关系，我只完成了大数据综合实验的基础部分。实验过程中遇到了一个问题，在 hbase shell 中无法建表，hbase shell 其他操作也有问题，如图所示：



```
pyworkspace (工作区) - Visual Studio Code
ssh + v 日 0% - x
输出 调试控制台 终端 JUPYTER
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 2.0.2, r1cfab033e779df840d5612a85277f42a6a4e8172, Tue Aug 28 20:50:40 PDT 2018
Took 0.0083 seconds
hbase(main):001:0> create 'movie_records','details'

ERROR: KeeperErrorCode = NoNode for /hbase/master

Creates a table. Pass a table name, and a set of column family
specifications (at least one), and, optionally, table configuration.
Column specification can be a simple string (name), or a dictionary
(dictionaries are described below in main help output), necessarily
including NAME attribute.
Examples:

Create a table with namespace=ns1 and table qualifier=t1
hbase> create 'ns1:t1', {NAME => 'f1', VERSIONS => 5}

Create a table with namespace=default and table qualifier=t1
hbase> create 't1', {NAME => 'f1'}, {NAME => 'f2'}, {NAME => 'f3'}
hbase> # The above in shorthand would be the following:
hbase> create 't1', 'f1', 'f2', 'f3'
hbase> create 't1', {NAME => 'f1', VERSIONS => 1, TTL => 2592000, BLOCKCACHE => true}
hbase> create 't1', {NAME => 'f1', CONFIGURATION => {'hbase.hstore.blockingStorefiles' => '10'}}
hbase> create 't1', {NAME => 'f1', IS_MOB => true, MOB_THRESHOLD => 1000000, MOB_COMPACT_PARTITION_POLICY => 'weekly'}

Table configuration options can be put at the end.
Examples:

hbase> create 'ns1:t1', 'f1', SPLITS => ['10', '20', '30', '40']
hbase> create 't1', 'f1', SPLITS => ['10', '20', '30', '40']
hbase> create 't1', 'f1', SPLITS_FILE => 'splits.txt', OWNER => 'johndoe'
hbase> create 't1', {NAME => 'f1', VERSIONS => 5, METADATA => { 'mykey' => 'myvalue' }}
hbase> # Optionally pre-split the table into NUMREGIONS, using
hbase> # SPLITALGO ("HexStringSplit", "UniformSplit" or classname)
hbase> create 't1', 'f1', {NUMREGIONS => 15, SPLITALGO => 'HexStringSplit'}
hbase> create 't1', 'f1', {NUMREGIONS => 15, SPLITALGO => 'HexStringSplit', REGION_REPLICATION => 2, CONFIGURATION => {'hbas
```

在网上查询相关资料，发现这个可能是 hbase 和 zookeeper 数据一致性的问题，StackOverflow 上有这一问题的解决方案：[apache zookeeper - hbase shell cannot use: ERROR: KeeperErrorCode = NoNode for /hbase/master - Stack Overflow](https://stackoverflow.com/questions/45442424/apache-zookeeper-hbase-shell-cannot-use-error-keepererrorcode-nodenode-for-hbase-master)

我的同学也给出了最佳的修改方案：首先修改 hbase-env.sh，再输入 zkCli.sh -server localhost:2181，然后 rmr /hbase

1: I have changed below property in hbase-env.sh, because I want to use separate ZK service instead of embedded one in HBase -

```
# Tell HBase whether it should manage it's own instance of ZooKeeper or not.
export HBASE_MANAGES_ZK=false
```

```
2022-06-26 16:01:13,244 [myid:] - INFO [main:Environment@100] - Client environment:java.compiler=j9jit29
2022-06-26 16:01:13,244 [myid:] - INFO [main:Environment@100] - Client environment:os.name=Linux
2022-06-26 16:01:13,244 [myid:] - INFO [main:Environment@100] - Client environment:os.arch=aarch64
2022-06-26 16:01:13,244 [myid:] - INFO [main:Environment@100] - Client environment:os.version=4.18.0-80.7.2.el7.aarch64
2022-06-26 16:01:13,244 [myid:] - INFO [main:Environment@100] - Client environment:user.name=root
2022-06-26 16:01:13,244 [myid:] - INFO [main:Environment@100] - Client environment:user.home=/root
2022-06-26 16:01:13,245 [myid:] - INFO [main:Environment@100] - Client environment:user.dir=/root
2022-06-26 16:01:13,245 [myid:] - INFO [main:ZooKeeper@438] - Initiating client connection, connectString=localhost:2181 sessionTimeout=30000 watcher=org.apache.zookeeper.ZooKeeperMain$MyWatcher@4f6a933d
Welcome to ZooKeeper!
2022-06-26 16:01:13,263 [myid:] - INFO [main-SendThread(localhost:2181):ClientCnxn$SendThread@975] - Opening socket connection to server localhost/0:0:0:0:0:0:0:1:2181. Will not attempt to authenticate using SASL (unknown error)
JLine support is enabled
2022-06-26 16:01:13,283 [myid:] - INFO [main-SendThread(localhost:2181):ClientCnxn$SendThread@852] - Socket connection established to localhost/0:0:0:0:0:0:0:1:2181, initiating session
2022-06-26 16:01:13,294 [myid:] - INFO [main-SendThread(localhost:2181):ClientCnxn$SendThread@1235] - Session establishment complete on server localhost/0:0:0:0:0:0:0:1:2181, sessionId = 0x1819db68f940001, negotiated timeout = 30000

WATCHER::

WatchedEvent state:SyncConnected type:None path:null
[zookeeper:localhost:2181(CONNECTED) 0] rmr /hbase
[zookeeper:localhost:2181(CONNECTED) 1]
```

接着操作就可以得到正确结果

```
hbase(main):001:0> create 'movie_records','details'
Created table movie_records
Took 2.0322 seconds
=> Hbase::Table - movie_records
hbase(main):002:0>
```

大数据综合实验-调研报告

大数据软件架构及应用

一 主流框架类型

目前比较常见的大数据软件架构有仅批处理框架、仅流处理框架、混合框架这几类。

二 仅批处理框架

主流的仅批处理框架有我们很熟悉的 Hadoop 框架

批处理是大数据处理当中的普遍需求，批处理主要操作大容量静态数据集，并在计算过程完成后返回结果。鉴于这样的处理模式，批处理有个明显的缺陷，就是面对大规模的数据，在计算处理的效率上，不尽如人意。

目前来说，批处理在应对大量持久数据方面的表现极为出色，因此经常被用于对历史数据进行分析。

Hadoop 擅长批处理、吞吐量大、做全量数据的离线分析，Hadoop 擅长于分析统计大批量的数据，在 hadoop 权威指南中，指出：MapReduce 的设计目标是服务于那些只需要数分钟或者数小时既可以完成的作业。例如有这样的案例，美国国家气候中心使用 hadoop 分析统计 1901 年到 2001 年的气象数据，统计出气温最高的年份。实际上这里已经体现出了 hadoop 处理数据的一个特点：对于已经存在的大量数据进行统计分析。这种特性注定了 hadoop 是高延迟的，即使我们在几秒中内可以算出结果，但是还是不够实时。

三 仅流处理框架

常见的仅流处理框架有 Apache Storm 和 Apache Samza

批处理之后出现的另一种普遍需求，就是流处理，针对实时进入系统的数据进行计算操作，处理结果立刻可用，并会随着新数据的抵达继续更新。

在实时性上，流处理表现优异，但是流处理同一时间只能处理一条（真正的流处理）或很少量（微批处理，Micro-batch Processing）数据，不同记录间只维持最少量的状态，对硬件的要求也要更高。

Storm 是 Twitter 开源的分布式实时大数据处理框架，最早开源于 github，从 0.9.1 版本之后，归于 Apache 社区，被业界称为实时版 Hadoop。随着越来越多的场景对 Hadoop 的 MapReduce 高延迟无法容忍，比如网站统计、推荐系统、预警系统、金融系统(高频交易、股票)等等，大数据实时处理解决方案（流计算）的应用日趋广泛，目前已是分布式技术领域最新爆发点，而 Storm 更是流计算技术中的佼佼者和主流。与 hadoop 不同的是，storm 不是批量处理已经存在的大量数据，而是实时计算每一条数据。例如，一个促销活动，假设首页上有 100 种商品，同时有几亿用户访问。运营者就需要实时统计每种商品的点击率，如果在一段时间内，某种商品的访问量太低，就应该使用其他商品替换这个商品。因为在促销中，首页上的商品位置资源是比较稀缺的，如果一个商品长时间没多少人访问，应该让更有价值

的商品来放在这个位置上。而因为用户量太大，商品太多，可能几分钟内就能产生几亿甚至几十亿的点击数据。

Apache Samza 是一种与 Apache Kafka 消息系统紧密绑定的流处理框架。虽然 Kafka 可用于很多流处理系统，但按照设计，Samza 可以更好地发挥 Kafka 独特的架构优势和保障。该技术可通过 Kafka 提供容错、缓冲，以及状态存储。Samza 可使用 YARN 作为资源管理器。这意味着默认情况下需要具备 Hadoop 集群（至少具备 HDFS 和 YARN），但同时也意味着 Samza 可以直接使用 YARN 丰富的内建功能。

四 混合处理框架

主流的混合处理框架主要为 Apache Spark 和 Apache Flink。实现这样的功能重点在于两种不同处理模式如何进行统一，以及要对固定和不固定数据集之间的关系进行何种假设在实际的应用当中，批处理和流处理同时存在的场景也很多，混合处理框架就旨在解决这类问题。提供一种数据处理的通用解决方案，不仅可以提供处理数据所需的方法，同时提供自己的集成。

Apache Spark 是一个围绕速度、易用性和复杂分析构建的大数据处理框架，最初在 2009 年由加州大学伯克利分校的 AMPLab 开发，并于 2010 年成为 Apache 的开源项目之一项、库、工具，可满足图形分析、机器学习、交互式查询等多种场景。Spark 及其 RDD 是在 2012 年开发的，以应对 MapReduce 集群计算范式的局限性，该范式在分布式程序上强制使用特定的线性数据流结构：MapReduce 程序从磁盘读取输入数据，在数据上映射函数，减少映射结果，并在磁盘上存储减少结果。Spark 的 RDD 作为分布式程序的工作集，提供限制形式的分布式共享内存。与 Hadoop 和 Storm 等其他大数据和 MapReduce 技术相比，Spark 有如下优势：Spark 提供了一个全面、统一的框架用于管理各种有着不同性质（文本数据、图表数据等）的数据集和数据源（批量数据或实时的流数据）的大数据处理的需求官方资料介绍 Spark 可以将 Hadoop 集群中的应用在内存中的运行速度提升 100 倍，甚至能够将应用在磁盘上的运行速度提升 10 倍。

Apache Flink 是由 Apache 软件基金会开发的开源，统一的流处理和批处理框架。Apache Flink 的核心是用 Java 和 Scala 编写的分布式流数据流引擎。Flink 是一个针对流数据和批数据的分布式处理引擎。对 Flink 而言，其所要处理的主要场景就是流数据，批数据只是流数据的一个极限特例而已。在 Flink 框架当中，所有的任务当成流来处理，因此实现了更低延迟的实时流处理。

参考文献：

- [1] [大数据常用处理框架 - 知乎 \(zhihu.com\)](#)
- [2] [Apache Hadoop - 维基百科, 自由的百科全书 \(wikipedia.org\)](#)
- [3] [Apache Storm - Wikipedia](#)
- [4] [Apache Samza - Wikipedia](#)
- [5] [Spark: 基本架构及原理 - 知乎 \(zhihu.com\)](#)
- [6] [Apache Spark - 维基百科, 自由的百科全书 \(wikipedia.org\)](#)
- [7] [Apache Flink - Wikipedia](#)

期末基础实验

基本命令

```
# 登录
ssh root@rcx-2019211279-0001
ssh root@rcx-2019211279-0002
ssh root@rcx-2019211279-0003
ssh root@rcx-2019211279-0004

cd /root/lastLab/resource/
cd /root/lastLab/resource/data
cd /root/lastLab/resource/code
```

环境配置部分

一 安装Kafka

1.1将本地相关资源压缩上传服务器并解压

```
# 本地压缩
# 上传服务器 使用winSCP
# 解压文件
tar -xvf resource.tar
```

1.2 解压Kafka

```
# 解压kafka
tar -zxvf kafka_2.11-0.10.2.2.tgz
# 将解压得到的文件夹移到/home/modules目录下
mv kafka_2.11-0.10.2.2 /home/modules/
```

1.3 编辑config/server.properties文件，修改delete.topic.enable和zookeeper.connect

1.4 将kafka文件夹通过scp发送到其余结点对应目录下

```
scp -r /home/modules/kafka_2.11-0.10.2.2 root@rcx-2019211279-0002:/home/modules/  
scp -r /home/modules/kafka_2.11-0.10.2.2 root@rcx-2019211279-0003:/home/modules/  
scp -r /home/modules/kafka_2.11-0.10.2.2 root@rcx-2019211279-0004:/home/modules/
```

1.5 编辑config/server.properties文件，修改broker.id分别为1、2、3、4

```
vim /home/modules/kafka_2.11-0.10.2.2/config/server.properties
```

1.6 各节点启动zookeeper

```
/usr/local/zookeeper/bin/zkServer.sh start
```

1.7 各节点启动kafka

```
/home/modules/kafka_2.11-0.10.2.2/bin/kafka-server-start.sh -daemon /home/modules/kafka_2.11-0.1
```

1.8 jps确认kafka启动成功

```
jps
```

二、安装Redis（单机部署）

2.1 升级gcc

```
yum -y install centos-release-scl
```

```
yum -y install devtoolset-9-gcc devtoolset-9-gcc-c++ devtoolset-9-binutils devtoolset-9-libatomic
```

```
scl enable devtoolset-9 bash
```

2.2 解压redis

```
tar -zxvf redis-6.0.6.tar.gz
```

2.3进入解压得到的文件夹，编译，安装

```
cd redis-6.0.6
```

```
make
```

```
make install
```


2.4 修改redis.conf，设置redis为守护进程，并允许远程连接，关闭保护模式

```
cd redis-6.0.6
grep -n daemon < redis.conf
grep -n protected-mode < redis.conf
grep -n requirepass < redis.conf
grep -n bind < redis.conf

vim /root/redis-6.0.6/redis.conf
vim redis.conf
```

2.5启动redis

```
redis-server redis-6.0.6/redis.conf
```

2.6ps确认redis运行成功

```
ps -ef | grep redis
```

三、安装python3

```
yum install python3
# 安装需要的库 happybase pandas redis kafka
# happybase
pip3 install happybase
# python-dev
yum search python | grep python-devel
yum install python3-devel
# pandas
python3 -m pip install --upgrade --force pip
pip3 install setuptools==33.1.1
pip3 install pandas
# redis
pip3 install redis
# kafka
pip3 install kafka
```

修改python代码

加密码

rcx@200012291918

```
# 字符集问题
# 注明utf-8
# coding=utf-8

# 在redis_connect () 中
grep -n redis.Con < generatorRecord.py
grep -n redis.Con < load_movie_redis.py
grep -n redis.Con < load_train_ratings_hbase.py
vim generatorRecord.py

pool(...password = 'rcx@200012291918'...)

grep -n redis.Con < recommend_client.py
grep -n redis.Con < recommend_server.py
vim recommend_server.py
vim recommend_server.py
```

运行

1 启动HDFS

```
start-all.sh
```

2 启动zookeeper(所有节点)

```
zkServer.sh start
```

```
/usr/local/zookeeper/bin/zkServer.sh start
```

3 启动HBase

```
start-hbase.sh
```

```
/usr/local/hbase/bin/start-hbase.sh
```

4配置HBase Thrift连接，以便python中的happybase库能够连接Hbase

```
/usr/local/hbase/bin/hbase-daemon.sh start thrift
```

5 在HBase中创建对应的表

需要进入到hbase.bash

```
hbase shell
```

```
create 'movie_records','details'
```

```
scan 'movie_records'
```

6.启动load_train_ratings_hbase.py（需要运行完）

```
python3 load_train_ratings_hbase.py rcx-2019211279-0001 9090 "movie_records" "../..data/json_tr
```

7 启动redis

```
redis-server redis-6.0.6/redis.conf
```

```
redis-cli -h 127.0.0.1 -p 6379
```

8 启动load_movie_redis.py（需要运行完）

```
python3 load_movie_redis.py rcx-2019211279-0001 6379 "../..data/movies.csv"
```

10 启动Kafka 并创建 Kafka Topic

```
/home/modules/kafka_2.11-0.10.2.2/bin/kafka-server-start.sh /home/modules/kafka_2.11-0.10.2.2/cc
```

```
-daemon
```

```
/home/modules/kafka_2.11-0.10.2.2/bin/kafka-server-start.sh -daemon /home/modules/kafka_2.11-0.1
```

```
# 创建topic
```

```
/home/modules/kafka_2.11-0.10.2.2/bin/kafka-topics.sh --zookeeper rcx-2019211279-0001:2181 --cre
```

```
#删除kafka topic
/home/modules/kafka_2.11-0.10.2.2/bin/kafka-topics.sh --zookeeper localhost:2181 --delete --topi

#查看kafka topic
/home/modules/kafka_2.11-0.10.2.2/bin/kafka-topics.sh --list --zookeeper localhost:2181

#查看kafka topic消息数量
/home/modules/kafka_2.11-0.10.2.2/bin/kafka-run-class.sh kafka.tools.GetOffsetShell --broker-li

# 关闭kafka
cd /home/modules/kafka_2.11-0.10.2.2/bin
. kafka-server-stop.sh
```

10 启动 generatorRecord.py

```
python3 generatorRecord.py -h rcx-2019211279-0001:9092 -f "../data/json_test_ratings.json"
```

11 启动hbase2spark、 kafkaStreaming、 recommend

```
#spark提交hbase2spark任务
/root/spark-2.1.1-bin-hadoop2.7/bin/spark-submit --class hbase2spark --master yarn --num-executc
#spark提交kafkaStreaming任务
/root/spark-2.1.1-bin-hadoop2.7/bin/spark-submit --class kafkaStreaming --master yarn --num-exec
#spark提交recommend任务
/root/spark-2.1.1-bin-hadoop2.7/bin/spark-submit --class recommend --master yarn --num-executors
```

12启动recommend_server.py

```
#启动推荐系统server (可以在本地Windows/macOS运行)
python3 recommend_server.py rcx-2019211279-0001 6379 23456
```

13 启动recommend_client.py

```
#启动推荐系统server (可以在本地Windows/macOS运行)
python3 recommend_client.py 127.0.0.1 23456
```

退出

```
# 关闭C&S
在recommend_client和recommend_server终端中直接ctrl+C
# 关闭步骤11的三个任务终端
直接ctrl+C
# 关闭generatorRecord
该终端中直接直接ctrl+C
# 关闭kafka (4台)
cd /home/modules/kafka_2.11-0.10.2.2/bin
. kafka-server-stop.sh
# 关闭redis
ps -ef | grep redis
kill -9 pid
# 关闭thrift
/usr/local/hbase/bin/hbase-daemon.sh stop thrift
# 关闭hbase
/usr/local/hbase/bin/stop-hbase.sh
# 关闭zookeeper (4台)
/usr/local/zookeeper/bin/zkServer.sh stop
# 关闭hdfs
/home/modules/hadoop-2.7.7/sbin/stop-all.sh
```