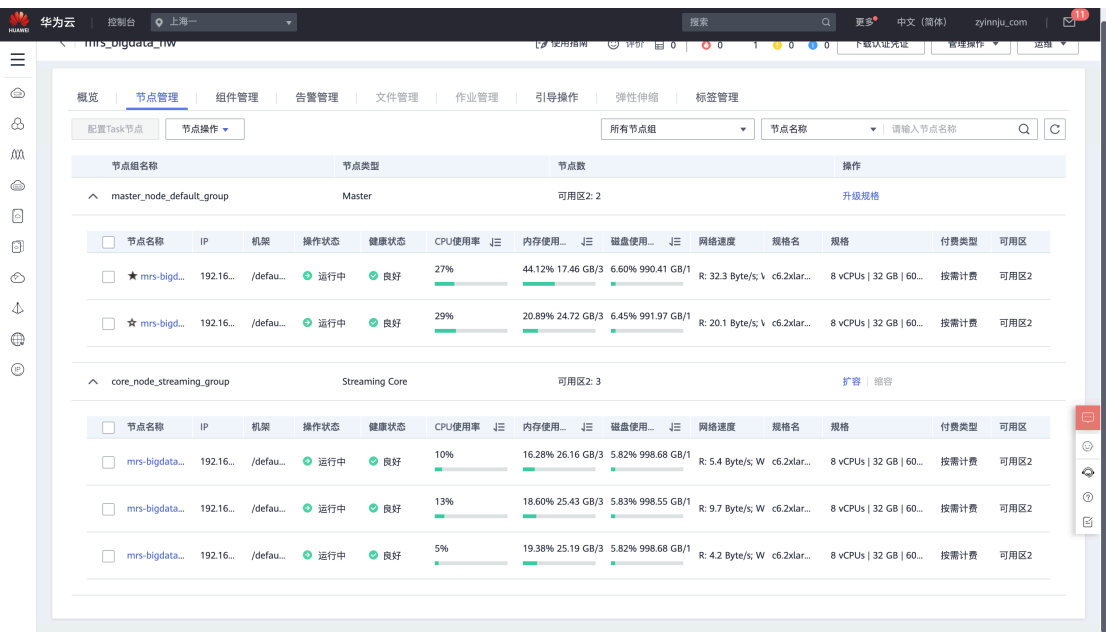
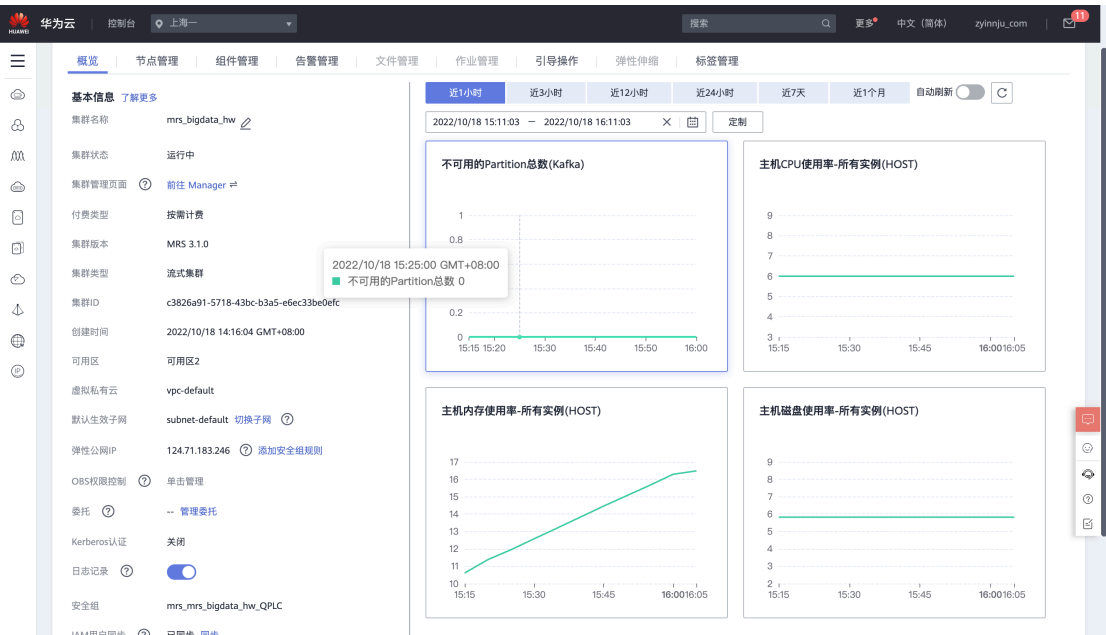


Big Data Analysis -- Homework 6

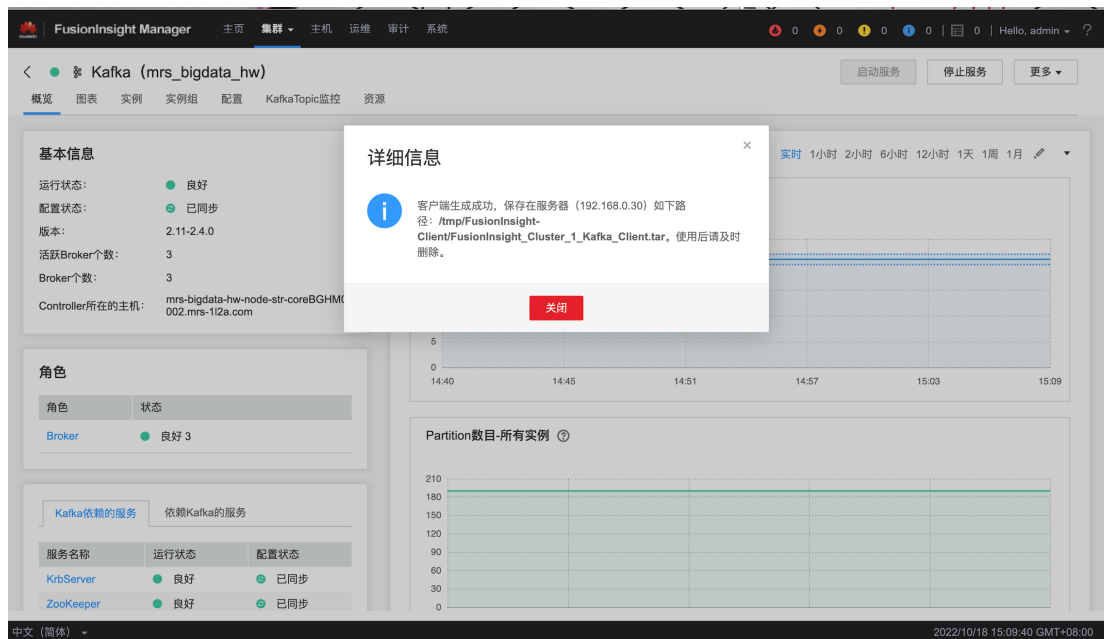
201250182 郑义

一、创建 Kafka 集群

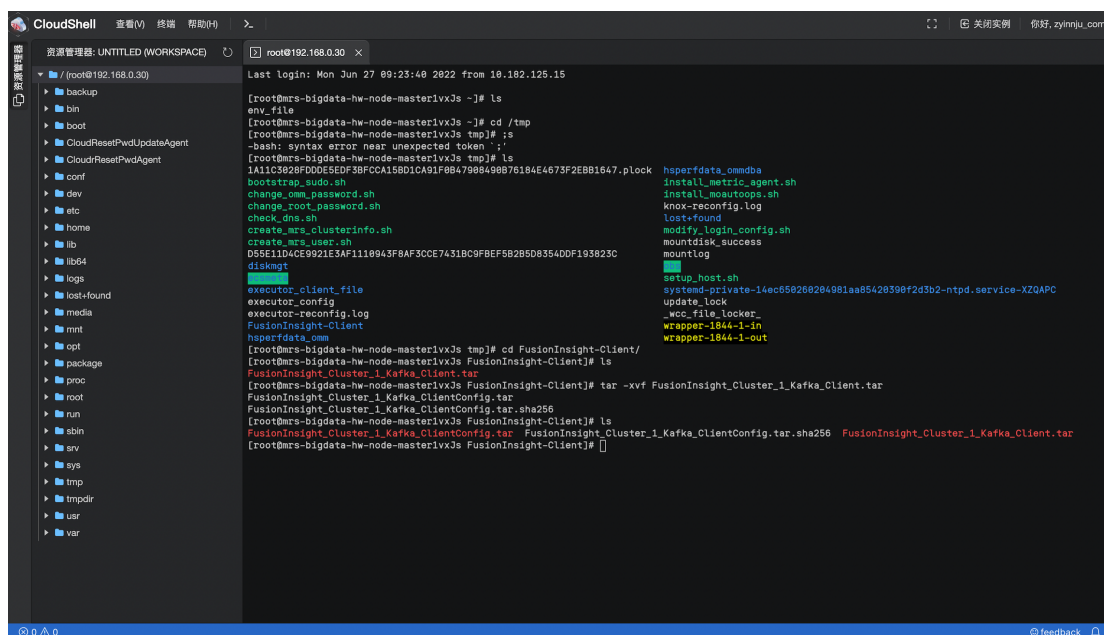
创建集群，两台 master 节点，三台 node 节点，配置为 8 核 32 GB。这里创建步骤和教程相同，不过多赘述，区别为：给 master 的一个节点多了一张弹性网卡，可以在本机通过 ssh 访问。



二、下载 Kafka 客户端



通过 manager 下载 kafka 客户端，然后进行安装。



解压后并进行 sha256 之后运行 ./install.sh /opt/client 进行安装。之后可以在 /opt/client 目录下执行 source bigdata_env 然后执行 kafka 相关命令。

三、创建 topic、生成消息、消费消息

在 /opt/client 目录下进行操作，首先 source bigdata_env

```
[root@mrs-bigdata-hw-node-master1vx3s ~]# cd /opt/client
[root@mrs-bigdata-hw-node-master1vx3s client]# ls
autoRefreshConfig.sh  client.properties  client-registry-1.0.0.jar  install.ini  jython-standalone-2.7.2.jar  log  refreshConfig.sh  switchuser.py
bigdata_env          clientregister.keytab  conf.py           JDK          Kafka          readme  register-client.sh  uninstall.sh
[root@mrs-bigdata-hw-node-master1vx3s client]#
```

然后创建 topic, 这里的 zookeeper 节点选择的是 192.168.0.120。

```
[root@mrs-bigdata-hw-node-master1vx3s client]# kafka-topics.sh --create --zookeeper 192.168.0.120:2181/kafka --partitions 2 --replication-factor 2 --topic bigdata-hw-topic
Created topic bigdata-hw-topic.
```

然后使用 consumer 和 producer 来生成和消费消息, kafka 实例选择的都是 192.168.0.120

```
[root@mrs-bigdata-hw-node-master1vx3s client]# kafka-console-producer.sh --broker-list 192.168.0.120:9092 --topic bigdata-hw-topic --producer.config /opt/client/Kafka/kafka/config/producer.properties
(2022-10-18 16:21:55,750) WARN The configuration 'producer.type' was supplied but isn't a known config. (org.apache.kafka.clients.producer.ProducerConfig)
(2022-10-18 16:21:55,751) WARN The configuration 'serializer.class' was supplied but isn't a known config. (org.apache.kafka.clients.producer.ProducerConfig)
>201250182郑义
>
```

可以看到消息被正确生成和消费

四、 Python 版本的 Kafka 连接
修改后的 producer.py 如下:

```
[root@mrs-bigdata-node-master1ocTe hw]# cat producer.py
#!/usr/bin/python3

from kafka import KafkaProducer

conf = {
    'bootstrap_servers': ["192.168.0.61:9092"],
    'topic_name': 'bigdata-hw-topic',
}

print('start producer')
producer = KafkaProducer(bootstrap_servers=conf['bootstrap_servers'])

data = bytes("201250182郑义", encoding="utf-8")
producer.send(conf['topic_name'], data)
producer.close()
print('end producer')
```

修改后的 consumer.py 如下:

```
[root@mrs-bigdata-node-master1ocTe hw]# cat consumer.py
#!/usr/bin/python3

from kafka import KafkaConsumer

conf = {
    'bootstrap_servers': ["192.168.0.61:9092"],
    'topic_name': 'bigdata-hw-topic',
    'consumer_id': 'consumer-id'
}

print('start consumer')
consumer = KafkaConsumer(conf['topic_name'],
                          bootstrap_servers=conf['bootstrap_servers'],
                          group_id=conf['consumer_id'])

for message in consumer:
    print("%s: value=%s" % (message.topic, message.value))

print('end consumer')
```

运行结果:

```
[root@mrs-bigdata-node-master1ocTe hw]#
[root@mrs-bigdata-node-master1ocTe hw]# python3 producer.py
start producer
end producer
[root@mrs-bigdata-node-master1ocTe hw]#

[root@mrs-bigdata-node-master1ocTe hw]#
[root@mrs-bigdata-node-master1ocTe hw]# python3 consumer.py
start consumer
bigdata-hw-topic: value=b'201250182\x09\x83\x91\xe4\xb9\xa9'
```

可以看到正确收到消息, 其中后面的编码是 “郑义” 的编码。