

大数据分析-作业6：基于 MRS 的 Kafka 实验报告

Kafka 简介

Apache Kafka起源于LinkedIn，后来于2011年成为开源Apache项目，然后于2012年成为First-class Apache项目。Kafka是用Scala和Java编写的。Apache Kafka是基于发布订阅的容错消息系统。它是快速，可扩展和设计分布。

关键步骤

1. 集群购买

在华为云的大数据：MapReduce 服务中购买集群，购买过程和配置参考华为云的手册。

集群部署成功的结果。

现有集群

评价 使用指南 购买集群

所有状态

请输入集群名称或ID

标签搜索

名称/ID	集群版...	集群类...	节...	状态	计费类...	创建时间	可用区	操作
mrs_demo 58812746-9348-47d3-81e9-a7b...	MRS 3.1.0	流式集...	5	运行中	按需计费	2022/10/18 19:14:3...	可用区2	转包周期 删除

← mrs_demo

使用指南 评价 下载认证凭证 运维

概览 节点管理 组件管理 告警管理 文件管理 作业管理 租户管理 引导操作 弹性伸缩 标签管理

基本信息

集群名称

mrs_demo

集群状态

运行中

集群管理页面

前往 Manager

付费类型

按需计费

集群版本

MRS 3.1.0

集群类型

流式集群

IAM用户同步

未同步

集群ID

58812746-9348-47d3-81e9-a7b625670fa3

创建时间

2022/10/18 19:14:38 GMT+08:00

可用区

可用区2

虚拟私有云

vpc-1927

默认生效子网

subnet-1951 切换子网

OBS权限控制

单击管理

数据连接

单击管理

委托

管理委托

弹性公网IP

123.60.37.127

Kerberos认证

关闭

日志记录

安全组

mrs_mrs_demo_nOYF

通信安全授权

组件版本

Kafka 版本

2.11-2.4.0

Flume 版本

1.9.0

ZooKeeper 版本

3.5.6

2. Kafka 客户端安装

前往 Manager, 以管理员身份进入, 选择 Kafka



下载安装包之后 scp 到服务器，解压之后安装 kafka 客户端

```
/opt/hadoopclient/FusionInsight_Cluster_1_Kafka_ClientConfig/Kafka
[22-10-18 20:44:41]: Copy Kafka config file to "/opt/client/Kafka/./kafka/config"
[22-10-18 20:44:41]: Create Kafka env file "/opt/client/Kafka/component_env".
[22-10-18 20:44:41]: Kafka installation is complete.
[22-10-18 20:44:41]: The component client is installed successfully
[22-10-18 20:44:41]: Persisting client information to the database...
[2022-10-18 20:44:41]: Platform=x86_64
[2022-10-18 20:44:41]: ClientIp=192.168.0.204
[2022-10-18 20:44:41]: InstallUser=root
[2022-10-18 20:44:41]: ClientPath=/opt/client
[2022-10-18 20:44:42]: Do Save client information to database Successfully.
[root@node-master2YFqX FusionInsight_Cluster_1_Kafka_ClientConfig]#
```

klist 检查，安装成功

```
[root@node-master2YFqX client]# kinit admin
Password for admin@HADOOP.COM:
[root@node-master2YFqX client]# klist
Ticket cache: FILE:/tmp//krb5cc_0
Default principal: admin@HADOOP.COM

Valid starting    Expires          Service principal
10/18/22 20:46:40 10/19/22 20:46:34 krbtgt/HADOOP.COM@HADOOP.COM
[root@node-master2YFqX client]#
```

3. Kafka 消息传递

创建 topic

在安装目录下输入命令 `source bigdata_env`

然后再输入 `kafka-topics.sh --create --zookeeper 192.168.0.73:2181/kafka --partitions 2 --replication-factor 2 --topic hw6` 来创建 topic hw6

```
root@node-master2YEvZ hadoopclient]# ls
autoRefreshConfig.sh  clientregister.keytab  Install.in1  Kafka  refreshConfig.sh  uninstall.sh
bigdata_env          client-registry-1.0.0.jar  JDK          log    register-client.sh
client.properties    conf.py               jython-standalone-2.7.2.jar  readme  switchuser.py
root@node-master2YEvZ hadoopclient]# source bigdata_env
root@node-master2YEvZ hadoopclient]# kafka-topics.sh --create --zookeeper 192.168.0.73:2181/kafka --partitions 2 --replication-factor 2 --topic hw6
Created topic hw6.
```

产生消息

输入命令 `kafka-console-producer.sh --broker-list 192.168.0.11:9092 --topic hw6 --producer.config /opt/hadoopclient/kafka/kafka/config/producer.properties`

```
[root@node-master2yEvZ config]# kafka-console-producer.sh --broker-list 192.168.0.11:9092 --topic hw6 --producer.config /opt/hadoopclient/Kafka/kafka/config/producer.properties
[2022-10-18 23:04:07,334] WARN The configuration 'producer.type' was supplied but isn't a known config. (org.apache.kafka.clients.producer.ProducerConfig)
[2022-10-18 23:04:07,334] WARN The configuration 'serializer.class' was supplied but isn't a known config. (org.apache.kafka.clients.producer.ProducerConfig)
>sunlifan
>201250181
```

消费消息

输入命令 `kafka-consoconsumer.sh --topic hw6 --bootstrap-server 192.168.0.11:9092 --consumer.config /opt/hadoopclient/Kafka/kafka/config/consumer.properties`

```
>^C[root@node-master2yEvZ config]# kafka-consoconsumer.sh --topic hw6 --bootstrap-server 192.168.0.11:9092 --consumer.config /opt/hadoopclient/Kafka/kafka/config/consumer.properties
201250181
sunlifan
```

4. Python 使用 Kafka

1. 准备环境：安装 python 版的 kafka 客户端

使用 `pip install kafka-python==2.0.1` 来安装

```
Python 2.7.5 (default, Nov 25 2020, 18:33:06)
[GCC 4.8.5 20150623 (EulerOS 4.8.5-28)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()
[root@node-master2yEvZ config]# pip install kafka-python==2.0.1
Collecting kafka-python==2.0.1
  Downloading https://files.pythonhosted.org/packages/aa/34/12f219f7f9e68e79a54874d26f974db1ab4efac4e6dae665b421df48f9/kafka_python-2.0.1-py2.py3-none-any.whl (232kB)
    100% |#####| 235kB 42kB/s
Installing collected packages: kafka-python
Successfully installed kafka-python-2.0.1
```

2. 导入 producer.py 和 consumer.py,这里我是从本地通过 scp 上传到云服务器上

```
Successfully installed pip-20.3.4
You are using pip version 20.3.4, however version 22.3 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
root@node-master2yEvZ config]# ls ~
consumer.py  env_file  producer.py
root@node-master2yEvZ config]# cd ~
root@node-master2yEvZ ~]#
```

具体的 producer.py 和 consumer.py 代码如下

```
# producer.py
#!/usr/bin/python3

from kafka import KafkaProducer

conf = {
    'bootstrap_servers': ['192.168.0.11:9092'],
    'topic_name': 'hw6',
}

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

data = bytes('sunlifan201250181').encode('utf-8')
producer.send(conf['topic_name'], data)
```

```
producer.close()
print('end producer')
```

```
# consumer.py
#!/usr/bin/python3

from kafka import KafkaConsumer

conf = {
    'bootstrap_servers': ['192.168.0.11:9092'],
    'topic_name': 'hw6',
    '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')
```

3. 执行 producer.py 和 consumer.py

```
[root@node-master2yEvZ ~]# python producer.py
start producer
end producer
[root@node-master2yEvZ ~]# python consumer.py
start consumer
hw6: value=sunlifan201250181
```

参考资料

实验手册

华为云官方文档

Kafka 官网