

分布式消息服务 Kafka 版

快速入门

文档版本 01
发布日期 2024-07-05



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目录

1 快速连接 Kafka 并生产消费消息.....	1
2 入门实践.....	13

1 快速连接 Kafka 并生产消费消息

本章节将为您介绍分布式消息服务Kafka版入门的使用流程，以在“华东-上海二”区域创建一个开启密文接入、安全协议为SASL_SSL的Kafka 2.7实例，客户端使用内网通过同一个VPC连接Kafka实例生产消费消息为例，帮助您快速上手Kafka。

图 1-1 Kafka 使用流程



1. 准备工作

Kafka实例运行于虚拟私有云（Virtual Private Cloud，简称VPC）中，在创建实例前需要确保有可用的虚拟私有云。

Kafka实例创建后，您需要在弹性云服务器中下载和安装Kafka开源客户端，然后才能进行生产消息和消费消息。

2. 创建Kafka实例

在创建实例时，您可以根据需求选择需要的实例规格和数量，并开启密文接入，安全协议设置为“SASL_SSL”。

连接安全协议为“SASL_SSL”的Kafka实例时，使用SASL认证，数据通过SSL证书进行加密传输，安全性更高。

3. 创建Topic

Topic用于存储消息，供生产者生产消息以及消费者订阅消息。

本文以在控制台创建Topic为例介绍。

4. 连接实例

在连接安全协议为“SASL_SSL”的Kafka实例时，需要下载证书，并在客户端配置文件中设置连接信息。

步骤一：准备工作

步骤1 注册华为账号并实名认证。

在创建Kafka实例前，请先注册华为账号并实名认证，具体步骤请参考[注册华为账号并开通华为云](#)和[实名认证介绍](#)。

如果您已有一个华为账号并实名认证，请跳过此步骤。

步骤2 为账户充值。

在创建Kafka实例前，确保账户有足够金额。账户充值的具体步骤，请参考[账户充值](#)。

步骤3 为用户添加Kafka实例的操作权限。

如果您需要对云上的资源进行精细管理，请使用统一身份认证服务（Identity and Access Management，简称IAM）创建IAM用户及用户组，并授权，以使得IAM用户获得Kafka实例的操作权限。具体操作请参考[创建用户并授权使用DMS for Kafka](#)。

步骤4 创建VPC和子网。

在创建Kafka实例前，确保已存在可用的VPC和子网。创建VPC和子网的具体步骤，请参考[创建虚拟私有云和子网](#)。

须知

创建的VPC与Kafka实例必须在相同的区域。

步骤5 创建安全组并添加安全组规则。

在创建Kafka实例前，确保已存在可用的安全组。创建安全组的具体步骤，请参考[创建安全组](#)。

连接Kafka实例前，请添加[表1-1](#)所示安全组规则，其他规则请根据实际需要添加。

表 1-1 安全组规则


方向	协议	端口	源地址	说明
入方向	TCP	9093	0.0.0.0/0	使用内网通过同一个VPC访问Kafka实例（密文接入）。

说明

安全组创建后，系统默认添加入方向“允许安全组内的弹性云服务器彼此通信”规则和出方向“放通全部流量”规则。此时使用内网通过同一个VPC访问Kafka实例，无需添加[表1-1](#)的规则。

步骤6 构建生产消费客户端。

本文以Linux系统的弹性云服务器（Elastic Cloud Server，简称ECS）作为生产消费客户端。在创建Kafka实例前，请先创建开启弹性公网IP的ECS、安装JDK、配置环境变量以及下载Kafka开源客户端。

1. 登录管理控制台，在左上角单击，选择“计算 > 弹性云服务器”，创建一个ECS实例。
创建ECS的具体步骤，请参考[购买弹性云服务器](#)。如果您已有可用的ECS，可重复使用，不需要再次创建。
2. 使用root用户登录ECS。
3. 安装Java JDK，并配置JAVA_HOME与PATH环境变量。

- a. 下载Java JDK。

说明

ECS默认自带的JDK可能不符合要求，例如OpenJDK，需要配置为Oracle的JDK，可至Oracle官方下载页面[下载Java Development Kit 1.8.111及以上版本](#)。

- b. 解压Java JDK。

```
tar -zxvf jdk-8u321-linux-x64.tar.gz
```

“jdk-8u321-linux-x64.tar.gz”为JDK的版本，请根据实际情况修改。

- c. 打开“.bash_profile”文件。

```
vim ~/.bash_profile
```

- d. 添加如下内容。

```
export JAVA_HOME=/opt/java/jdk1.8.0_321
export PATH=$JAVA_HOME/bin:$PATH
```

“/opt/java/jdk1.8.0_321”为JDK的安装路径，请根据实际情况修改。

- e. 按“Esc”，然后输入以下命令，按“Enter”，保存并退出“.bash_profile”文件。

```
:wq
```

- f. 执行如下命令使修改生效。

```
source .bash_profile
```

- g. 查看Java JDK是否安装成功。

```
java -version
```

回显信息中包含如下信息，表示Java JDK安装成功。

```
java version "1.8.0_321"
```

4. 下载开源的Kafka客户端。

```
wget https://archive.apache.org/dist/kafka/2.7.2/kafka_2.12-2.7.2.tgz
```

5. 解压Kafka客户端文件。

```
tar -zxf kafka_2.12-2.7.2.tgz
```

----结束

步骤二：创建 Kafka 实例

步骤1 进入[购买Kafka实例页面](#)。

步骤2 设置实例基础信息，配置详情请参考[表1-2](#)。

表 1-2 设置实例基础信息

参数	说明
计费模式	选择“按需计费”，即先使用再付费，按照Kafka实例实际使用时长计费，秒级计费，按小时结算。
区域	不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。 选择“华东-上海二”。
项目	每个区域默认对应一个项目，这个项目由系统预置，用来隔离物理区域间的资源（计算资源、存储资源和网络资源）。 选择“华东-上海二(默认)”。
可用区	可用区指在同一区域下，电力、网络隔离的物理区域，可用区之间内网互通，不同可用区之间物理隔离。 选择“可用区1、可用区2、可用区3”。
实例名称	实例名称支持自定义，但需要符合命名规则：长度为4~64个字符，由英文字母开头，只能由英文字母、数字、中划线、下划线组成。 输入“kafka-test”。
企业项目	该参数针对企业用户使用。企业项目是对企业不同项目间资源的分组和管理，属于逻辑隔离。 选择“default”。
规格选择模式	选择“集群版”，创建一个集群Kafka实例。
版本	Kafka的版本号。Kafka实例创建后，版本号不支持修改。 选择“2.7”。
代理规格	根据业务需求选择相应的代理规格。 选择“kafka.2u4g.cluster”。
代理数量	根据业务需求选择相应的代理数量。 选择“3”。
单个代理存储空间	根据实际需要选择存储Kafka数据的磁盘类型和磁盘大小。 实例总存储空间 = 单个代理的存储空间 * 代理数量，Kafka实例创建后，磁盘类型不支持修改。 选择“超高I/O 100GB”。
容量阈值策略	选择“自动删除”，即磁盘达到95%的容量阈值后，可以正常生产和消费消息，但是会删除最早的10%的消息，以保证磁盘容量充足。该场景优先保障业务不中断，数据存在丢失的风险。

图 1-2 实例基础信息

计费模式

包年/包月

按需计费

区域

华东-上海二

不同区域的资源之间内网不互通。请选择靠近您客户的区域，可以降低网络时延、提高访问速度。

项目

华东-上海二(默认)

可用区

可用区1

可用区2

可用区3

可用区4

温馨提示：不支持选2个可用区，请选择1个或者3个及以上可用区。
单个可用区无法保证可靠性和服务SLA，建议选择多个可用区。

实例名称

kafka-test

企业项目

default

[新建企业项目](#)

规格选择模式

集群版

规格测算

单机版

自定义Kafka实例的规格配置。

版本

3.x

2.7

1.1.0

实例安装的软件版本。版本支持公告 [版本发布说明](#)

实例创建后，不支持变更版本。建议服务端版本和客户端版本保持一致。

规格名称	单个代理TPS	单个代理最大分区数	单个代理消费组数	单个代理流量规格 (MB/s)
<input type="radio"/> kafka.2u4g.cluster.small	20,000	100	15	40
<input checked="" type="radio"/> kafka.2u4g.cluster	30,000	250	20	100
<input type="radio"/> kafka.4u8g.cluster	100,000	500	100	200
<input type="radio"/> kafka.8u16g.cluster	150,000	1,000	150	375
<input type="radio"/> kafka.12u24g.cluster	200,000	1,500	200	625
<input type="radio"/> kafka.16u32g.cluster	250,000	2,000	200	750

当前选择规格 kafka.2u4g.cluster | 单个代理TPS 30,000 | 单个代理最大分区数 250 | 单个代理消费组数 20 | 单个代理流量规格 (MB/s) 100

代理数量

-
3
+

实例创建后，代理数量不支持扩容。

当前实例网络总流量规格：300 MB/s (当前实例网络总流量规格 = 单个代理流量规格 * 代理数量)

当前实例业务流量规格：

- * 建议读流量：50 MB/s
- * 建议写流量：50 MB/s

实例网络总流量规格和业务流量规格参考说明请参考 [实例规格和网络带宽说明](#)。

由于一般业务系统流量存在波动，建议您预留30%冗余流量，防止出现流量波动影响您的业务。

单个代理存储空间

超高性能

-
100
+

GB

实例总存储空间 300 GB

磁盘类型创建完成后不可修改，存储空间不支持扩容，所选磁盘性能和并发请参考 [如何选择磁盘类型](#)，并根据实际业务IO要求进行选择。

当前实例磁盘总带宽规格：510 MB/s (当前实例磁盘总带宽规格 = 单个代理磁盘规格 * 代理数量)

单个代理磁盘带宽：170 MB/s

容量刷值策略

自动删除

生产受限

步骤3 设置实例网络环境信息，配置详情请参考表1-3。

表 1-3 设置实例网络环境信息

参数	说明
虚拟私有云	虚拟私有云和子网在Kafka实例创建完成后，不支持修改。 选择 准备工作 中设置好的虚拟私有云和子网。
安全组	选择 准备工作 中设置好的安全组。

图 1-3 实例网络环境信息



步骤4 设置实例的访问方式，配置详情请参考表1-4。

表 1-4 设置实例的访问方式

参数	子参数	说明
内网访问	明文接入	选择“不开启”。
	密文接入	<p>密文接入表示客户端连接Kafka实例时，需要进行SASL认证。</p> <ol style="list-style-type: none"> 开启“密文接入”。 “kafka安全协议”选择“SASL_SSL”，设置用户名和密码。用户名输入“test”。密文接入成功开启后，用户名不支持修改。 用户名需要符合以下命名规则：由英文字母开头，且只能由英文字母、数字、中划线、下划线组成，长度为4~64个字符。 密码需要符合以下命名规则： <ul style="list-style-type: none"> 长度为8~32个字符。 至少包含以下字符中的3种：大写字母、小写字母、数字、特殊字符`~!@#\$\$%^&*()-_+=\ []{};:","<.>?`和空格，并且不能以-开头。 不能与用户名或倒序的用户名相同。 开启“SASL PLAIN机制”。
	内网IP地址	选择“自动分配”，即系统自动分配子网中可用的IP地址。
公网访问	-	选择“不开启”。

图 1-4 实例的访问方式

内网访问

- 明文接入:
- 密文接入:
- kafka安全协议: **SASL_SSL** | SASL_PLAINTEXT
SASL_SSL加密传输, 安全性较好。
- 跨VPC访问协议: SASL_SSL 实例创建成功后, 跨VPC访问安全协议无法修改。
- 内网IP地址: **自动分配** | 手动选择

公网访问:

用户名:
SSL用户名用于实例访问认证, 实例创建后将不可修改。

密码:

确认密码:

SASL PLAIN 机制:
默认支持SCRAM-SHA-512机制, 开启后同时支持SCRAM-SHA-512和PLAIN机制。

步骤5 单击“更多配置”，设置更多相关信息，配置详情请参考表1-5。

表 1-5 更多配置

参数	说明
Smart Connect	选择“不开启”。
Kafka自动创建Topic	选择“不开启”。
标签	不设置。
描述	不设置。

图 1-5 更多配置

更多配置 ^

Smart Connect Smart Connect 用于异构系统间的数据同步。

Kafka自动创建Topic 开启后表示生产或消费一个未创建的Topic时, 系统会自动创建此Topic, 此Topic的默认参数值如下: 分区数为3, 副本数为3, 老化时间为72小时, 不开启同步复制和同步消息。建议仅在迁移场景或测试场景使用。

标签 如果您需要使用同一标签标识多种云资源, 即所有服务均可在标签输入框下拉选择同一标签, 建议在TMS中创建预定义标签, [查看预定义标签](#)

您还可以添加20个标签。

描述

0/1,024

步骤6 单击“立即购买”，进入规格确认页面。

步骤7 确认实例信息无误后，提交请求。

步骤8 单击“返回Kafka专享版列表”，查看Kafka实例是否创建成功。

创建实例大约需要3到15分钟，此时实例的“状态”为“创建中”。

- 当实例的“状态”变为“运行中”时，说明实例创建成功。
- 当实例的“状态”变为“创建失败”，请删除创建失败的实例，然后重新创建。如果重新创建仍然失败，请联系客服。


说明

创建失败的实例，不会占用其他资源。

步骤9 实例创建成功后，单击实例名称，进入实例详情页。

步骤10 在“连接信息”区域，查看并记录连接地址。

图 1-6 使用内网通过同一个 VPC 访问 Kafka 实例的连接地址

内网密文连接地址	192.0.0.238:9093,192.0.0.32:9093,192.0.0.236:9093 
----------	-------------------------------------------------------------------------------------------------------------------------------------

----结束

步骤三：创建 Topic

步骤1 在“Kafka专享版”页面，单击Kafka实例的名称，进入实例详情页面。

步骤2 在左侧导航栏单击“Topic管理”，进入Topic列表页。

步骤3 单击“创建Topic”，弹出“创建Topic”对话框。

步骤4 填写Topic名称和配置信息，配置详情请参考表1-6，单击“确定”，完成创建Topic。

表 1-6 Topic 参数说明

参数	说明
Topic名称	<p>名称支持自定义，但需要符合命名规则：以英文字母、数字、下划线开头，且只能由英文字母、数字、句点、中划线、下划线组成，长度为3~200个字符。</p> <p>名称不能为以下内置Topic：</p> <ul style="list-style-type: none"> • <code>__consumer_offsets</code> • <code>__transaction_state</code> • <code>__trace</code> • <code>__connect-status</code> • <code>__connect-configs</code> • <code>__connect-offsets</code> <p>创建Topic后不能修改名称。</p> <p>输入“topic-01”。</p>
分区数	<p>如果分区数与消费者数一致，分区数越大消费的并发度越大。</p> <p>输入“3”。</p>
副本数	<p>Kafka会自动在每个副本上备份数据，当其中一个Broker故障时数据依然是可用的，副本数越大可靠性越高。</p> <p>输入“3”。</p>
老化时间（小时）	<p>消息的最长保留时间，消费者必须在此时间结束前消费消息，否则消息将被删除。删除的消息，无法被消费。</p> <p>输入“72”。</p>
同步复制	<p>选择“不开启”，即Leader副本接收到消息并成功写入到本地日志后，就马上向客户端发送写入成功的消息，不需要等待所有Follower副本复制完成。</p>
同步落盘	<p>选择“不开启”，即生产的消息存在内存中，不会立即写入磁盘。</p>
消息时间戳类型	<p>选择“CreateTime”，即生产者创建消息的时间。</p>
批处理消息最大值（字节）	<p>Kafka允许的最大批处理大小，如果在生产客户端配置文件或代码中启用消息压缩，则表示压缩后的最大批处理大小。</p> <p>输入“10485760”。</p>
描述	<p>不设置。</p>

图 1-7 创建 Topic

×

创建 Topic

Topic 名称

分区数 ? 取值范围: 1-200
Topic创建后, 分区数不支持扩容。

副本数 ? 取值范围: 1-3 , 建议取3副本
副本数需要小于等于代理个数。

老化时间 (小时) ? 取值范围: 1-720
Topic中消息的过期时间, 超过时间的消息将被删除, 无法被消费。

同步复制 ?

同步落盘 ?

消息时间戳类型 ?

批处理消息最大值 (字节) ?

描述

0/200 ↗

---结束

步骤四：连接实例生产消费消息

步骤1 配置生产消费配置文件。

1. 登录Linux系统的ECS。
2. 下载client.jks证书并上传到ECS的“/root”目录下。

获取证书的方法如下：在Kafka控制台单击Kafka实例名称，进入实例详情页面，在“连接信息 > SSL证书”所在行，单击“下载”。解压压缩包，获取压缩包中的客户端证书文件：client.jks。

📖 说明

“/root”为证书存放路径，请根据实际情况修改。

3. 进入Kafka客户端文件的“/config”目录下。

```
cd kafka_2.12-2.7.2/config
```

4. 在“consumer.properties”和“producer.properties”文件中分别增加如下行（示例以PLAIN机制为例介绍）。

```
sasl.jaas.config=org.apache.kafka.common.security.plain.PlainLoginModule required \
username="*****" \
password="*****";
sasl.mechanism=PLAIN

security.protocol=SASL_SSL
ssl.truststore.location={ssl_truststore_path}
ssl.truststore.password=dms@kafka
ssl.endpoint.identification.algorithm=
```

参数说明：

- username和password为创建Kafka实例过程中开启密文接入时填入的用户名和密码。
- ssl.truststore.location配置为[步骤1.2](#)证书的存放路径。
- **ssl.truststore.password为服务器证书密码，不可更改，需要保持为dms@kafka。**
- ssl.endpoint.identification.algorithm为证书域名校验开关，为空则表示关闭。这里需要保持关闭状态，必须设置为空。

- 步骤2** 进入Kafka客户端文件的“/bin”目录下。

```
cd ../bin
```

- 步骤3** 生产消息。

```
./kafka-console-producer.sh --broker-list ${连接地址} --topic ${Topic名称} --producer.config ../config/producer.properties
```

参数说明如下：

- 连接地址：从[创建实例](#)中获取的连接地址。
- Topic名称：从[创建Topic](#)中获取的Topic名称。

示例如下，“192.xxx.xxx.xxx:9093,192.xxx.xxx.xxx:9093,192.xxx.xxx.xxx:9093”为Kafka实例连接地址。

执行完命令后，输入需要生产的消息内容，按“Enter”发送消息到Kafka实例，输入的每一行内容都将作为一条消息发送到Kafka实例。

```
[root@ecs-kafka bin]#./kafka-console-producer.sh --broker-list
192.xxx.xxx.xxx:9093,192.xxx.xxx.xxx:9093,192.xxx.xxx.xxx:9093 --topic topic-01 --producer.config ../config/producer.properties
>Hello
>DMS
>Kafka!
>^C[root@ecs-kafka bin]#
```

如需停止生产使用Ctrl+C命令退出。

- 步骤4** 消费消息。

```
./kafka-console-consumer.sh --bootstrap-server ${连接地址} --topic ${Topic名称} --from-beginning --consumer.config ../config/consumer.properties
```

参数说明如下：

- 连接地址：从[创建实例](#)中获取的连接地址。
- Topic名称：从[创建Topic](#)中获取的Topic名称。

示例如下：

```
[root@ecs-kafka bin]# ./kafka-console-consumer.sh --bootstrap-server
192.xxx.xxx.xxx:9093,192.xxx.xxx.xxx:9093,192.xxx.xxx.xxx:9093 --topic topic-01 --from-beginning --
consumer.config ../config/consumer.properties
Hello
Kafka!
DMS
^CProcessed a total of 3 messages
[root@ecs-kafka bin]#
```

如需停止消费使用**Ctrl+C**命令退出。

----结束

相关信息

- 了解Kafka实例的相关概念，请参考[Kafka基本概念](#)。
- 了解Kafka实例的价格，请参考[分布式消息服务Kafka版价格详情](#)。
- 在控制台查看消息，请参考[查看Kafka消息](#)。
- 查看消费进度，请参考[查看Kafka消费进度](#)。
- 查看Kafka实例的监控指标，请参考[查看Kafka监控数据](#)。

2 入门实践

当您购买了Kafka实例后，可以根据自身的业务需求使用分布式消息服务Kafka版提供的一系列常用实践。

表 2-1 常用最佳实践

实践	描述
使用MirrorMaker跨集群同步数据	使用MirrorMaker可以实现将源集群中的数据镜像复制到目标集群中。
配置消息堆积数监控	介绍如何设置消息组的消息堆积数超过阈值时，消息通知服务及时发送告警短信/邮件通知用户。