

数据接入服务

# 快速入门

文档版本 02  
发布日期 2023-07-12



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 DIS 使用流程简介.....	1
2 步骤 1: 开通 DIS 通道.....	2
3 步骤 2: 准备 DIS 应用开发环境.....	6
4 步骤 3: 发送数据到 DIS.....	11
5 步骤 4: 从 DIS 获取数据.....	13
6 获取认证信息.....	14
7 连接 OBS.....	15
8 自动创建委托.....	16
9 入门实践.....	17

# 1 DIS 使用流程简介

---

DIS的使用流程如下：

## **步骤1：开通DIS通道**

用户使用DIS前需要先开通DIS通道。

## **步骤2：准备DIS应用开发环境**

用户开发DIS应用程序前，首先需要安装应用开发工具。然后获取SDK和样例工程，并导入到用户的开发环境中。

## **步骤3：发送数据到DIS**

基于数据上传业务开发应用程序，并运行程序，实现数据上传功能。数据上传过程中可在Console控制台查看数据上传通道相关信息。

## **步骤4：从DIS获取数据**

基于数据下载业务开发应用程序，并运行程序，实现数据下载功能。

# 2 步骤 1：开通 DIS 通道

用户可以基于云管理平台Web界面开通DIS通道。

## 前提条件

使用DIS前需要注册公有云账户，再开通DIS。

## 操作步骤












- 步骤1** 使用注册账户登录[DIS控制台](#)。
- 步骤2** 单击管理控制台左上角的 ，选择区域和项目。
- 步骤3** 单击“购买接入通道”配置相关参数。

表 2-1 接入通道参数说明

参数	参数解释	参数示例
计费模式	按需计费	按需计费
区域	指的是云服务所在的物理位置。您可以在下拉框中选择并切换区域。	华北-北京1
<b>基本信息</b>		
通道名称	用户发送或者接收数据时，需要指定通道名称，通道名称不可重复。通道名称由英文字母、数字、中划线和下划线组成。长度为1~64个字符。	dis-Tido

参数	参数解释	参数示例
通道类型	<ul style="list-style-type: none"> <li>普通通道单分区容量：最高发送速度可达 1MB/秒或1000条记录/秒（达到任意一种速度上限才会被限流），最高提取速度可达 2MB/秒，单次请求的记录总大小不能超过1MB（不包含partitionKey数据大小）。</li> <li>高级通道单分区容量：最高发送速度可达 5MB/秒或2000条记录/秒（达到任意一种速度上限才会被限流），最高提取速度可达 10MB/秒，单次请求的记录总大小不能超过5MB（不包含partitionKey数据大小）</li> </ul>	-
分区数量	分区是DIS数据通道的基本吞吐量单位。	5
分区计算	<p>用户可以根据实际需求通过系统计算得到一个建议的分区数量值。</p> <ol style="list-style-type: none"> <li>单击“分区计算”，弹出“计算所需分区数量”对话框。</li> <li>根据实际需求填写“平均记录大小”、“最大写入记录数”和“消费程序数量”，“预估所需分区数量”选项框中将显示所需的分区数量，此值不可修改。</li> </ol> <p><b>说明</b> 所需分区计算公式：</p> <ul style="list-style-type: none"> <li>按流量计算所需写分区数：（所得数值需向上取整后作为分区数） 普通通道：平均记录大小*（1+分区预留比例20%）*最大写入记录数/（1*1024KB） 高级通道：平均记录大小*（1+分区预留比例20%）*最大写入记录数/（5*1024KB）</li> <li>按消费程序数量计算读分区数：（消费程序数量/2后的数值需要保留两位小数，然后乘以“按流量计算所需写分区数”，最终取值需向上取整） （消费程序数量/2）*按流量计算所需的写分区数</li> </ul> <p>获取“按流量计算所需写分区数”、“按消费程序数量计算读分区数”中的最大值作为预估所需分区数量。</p> <ol style="list-style-type: none"> <li>单击“使用计算值”将系统计算出的建议值应用于“分区数量”。</li> </ol>	-
生命周期（小时）	<p>存储在DIS中的数据保留的最长时间，超过此时长数据将被清除。</p> <p>取值范围：24~72的整数。</p>	24

参数	参数解释	参数示例
源数据类型	<ul style="list-style-type: none"> <li>• BLOB: 存储在数据库管理系统中的一组二进制数据。“源数据类型”选择“BLOB”，则支持的“转储服务类型”为“OBS”、“MRS”。</li> <li>• JSON: 一种开放的文件格式，以易读的文字为基础，用来传输由属性值或者序列性的值组成的数据对象。“源数据类型”选择“JSON”，则支持的“转储服务类型”为“OBS”、“MRS”、“DLI”和“DWS”。</li> <li>• CSV: 纯文本形式存储的表格数据，分隔符默认采用逗号。“源数据类型”选择“CSV”，则支持的“转储服务类型”为“OBS”、“MRS”、“DLI”、“DWS”。</li> </ul>	JSON
自动扩缩容	<p>创建通道的同时是否开启自动扩缩容功能。</p> <p>通过单击  或  来关闭或开启自动扩缩容开关。</p>	<p><b>说明</b></p> <p>用户可在创建通道时定义是否自动扩缩容，也可对已创建的通道修改自动扩缩容属性。</p>
自动缩容最小分区数	设置自动缩容的分区下限，自动缩容的目标分区数不小于下限值。	-
自动扩容最大分区数	设置自动扩容的分区上限，自动扩容的目标分区数不超过上限值。	-
源数据分隔符	源数据为CSV格式时的数据分隔符。	-
Schema开关	<p>创建通道的同时是否为其创建数据Schema。源数据类型为JSON或CSV时可配置该参数。</p> <p>通过单击  或  来关闭或开启Schema配置开关。</p> <p><b>说明</b></p> <p>若创建通道时，没有同时创建数据Schema,可待通道创建成功后。到通道的管理页面创建数据Schema，详情请参见<a href="#">管理源数据Schema</a>。</p>	“源数据类型”为“JSON”和“CSV”时，可选择创建数据Schema。

参数	参数解释	参数示例
源数据 Schema	<p>支持输入和导入源数据样例，源数据样例格式为 JSON 或者 CSV，详细操作请参见<a href="#">管理源数据 Schema</a>。</p> <p>1. 在左侧文本框中输入 JSON 或者 CSV 格式的源数据样例，也可单击  导入源数据样例。</p> <p>2. 在左侧文本框中单击 ，可删除左侧文本框中已输入或导入的源数据样例。</p> <p>3. 在左侧文本框中单击 ，可在右侧文本框中根据源数据样例生成 Avro schema。</p> <p>4. 在右侧文本框中单击 ，可删除已生成的 Avro schema。</p> <p>5. 在右侧文本框中单击 ，可修改已生成的 Avro schema。</p>	<p>仅当“Schema 配置开关”配置为“开启”：  时需要配置此参数。</p>
企业项目	<p>配置通道所属的企业项目。已开通企业项目管理服务的用户才可以配置该参数。默认值为 default。</p> <p>企业项目是一种云资源管理方式，企业项目管理服务提供统一的云资源按项目管理，以及项目内的资源管理、成员管理。</p> <p>您可以选择默认的企业项目“default”或其他已有的企业项目。如果要创建新的企业项目，请登录企业管理控制台进行创建，详细操作请参考《企业管理用户指南》。</p>	-
现在配置	<p>单击“现在配置”，呈现添加标签。</p> <p>添加标签具体请参考<a href="#">管理通道标签</a>。</p>	-
暂不配置	暂不配置任何信息。	-
标签	<p>标签是通道的标识。为通道添加标签，可以方便用户识别和管理拥有的通道资源。</p>	-

**步骤4** 单击“立即购买”，弹出“规格确认”页面。

**步骤5** 单击“提交”，完成通道接入。

----结束



# 3 步骤 2: 准备 DIS 应用开发环境

用户开发DIS应用程序前，首先需要安装和配置应用开发环境。获取SDK和样例工程，并导入到用户的开发环境中。

## 前提条件

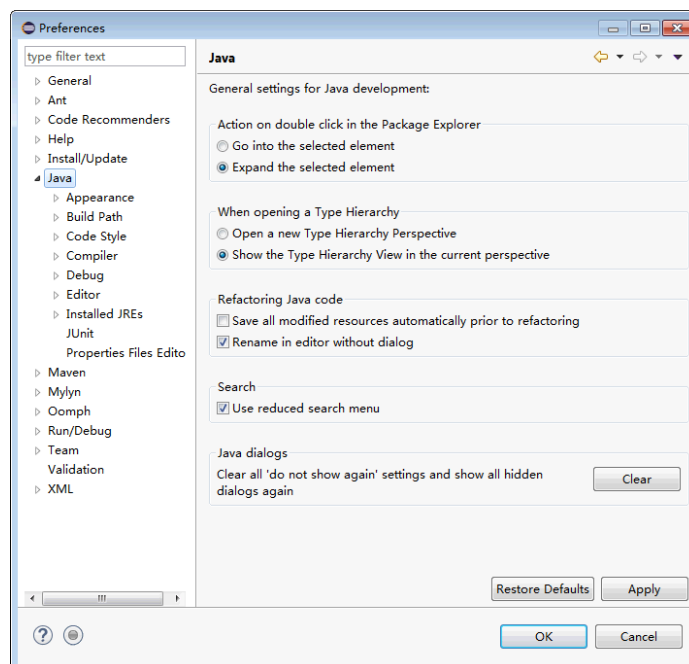
- JDK(1.8版本或以上版本)工具已安装成功。
- Eclipse工具已安装成功。

## 操作步骤

### 步骤1 Eclipse中配置JDK。

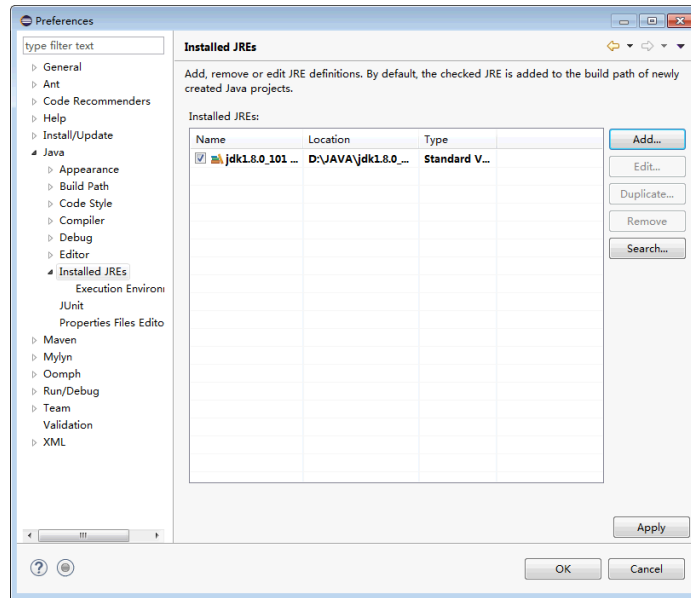
1. 打开Eclipse工具，选择“Window > Preferences”，弹出“Preferences”窗口。
2. 在左侧菜单栏单击“Java”，显示如图3-1所示内容，选择相关配置，单击“OK”。

图 3-1 Preferences



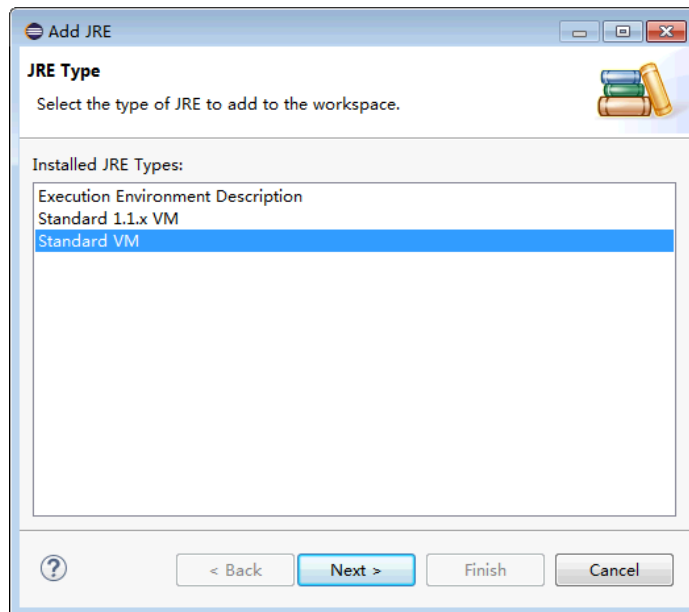
3. 在左侧菜单栏选择“Java > Installed JREs”配置JDK环境变量，显示如图3-2所示。
  - 右侧窗口中显示已配置好的JDK变量，执行步骤1.3.a完成JDK变量配置。
  - 如需配置多个不同的变量对应不同版本的JDK，请执行步骤1.3.b ~ 步骤1.3.d。

图 3-2 Installed JREs



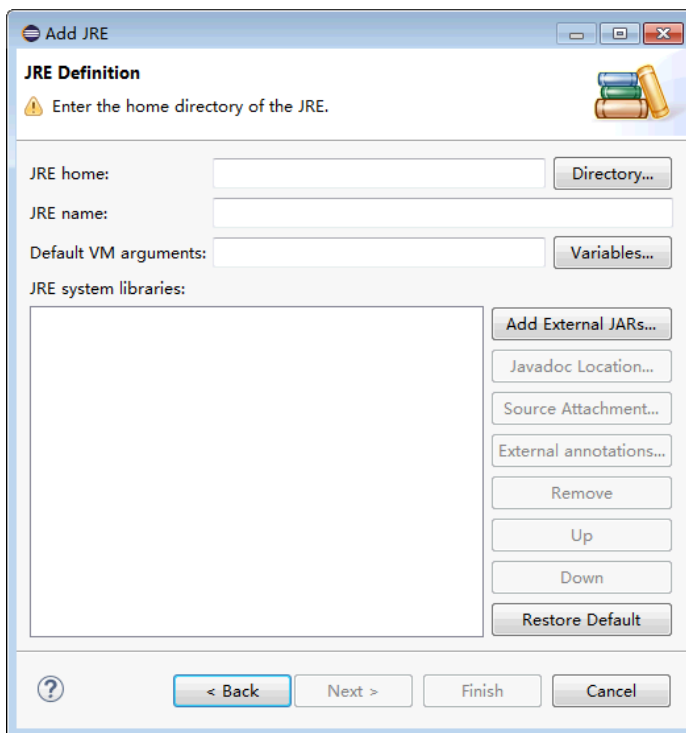
- a. 勾选已安装的JDK，单击“OK”。
- b. 单击“Add”按钮，弹出“Add JRE”窗口，如图3-3所示。

图 3-3 JRE Type



- c. 选择一个JRE类型，单击“Next”，弹出如图3-4所示窗口。

图 3-4 JRE Definition



- d. 配置JDK基本信息，单击“Finish”完成配置。
- JRE home: JDK安装路径。
  - Default VM arguments: JDK运行参数。

**步骤2** 下载资源包。

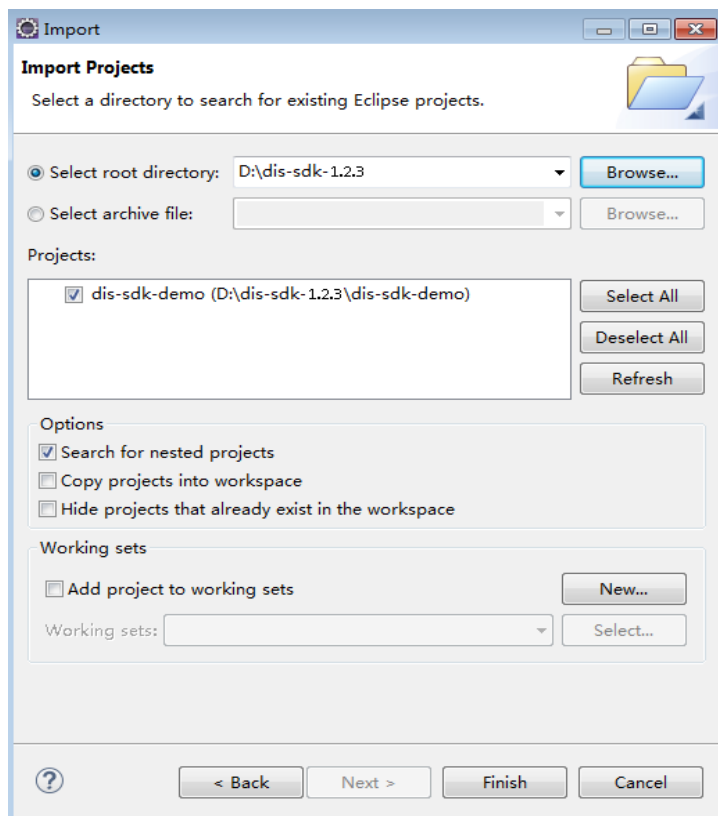
在<https://github.com/huaweicloud/huaweicloud-sdk-java-dis>中下载DIS的Java SDK压缩包。

在DIS SDK桶中获取“huaweicloud-sdk-dis-java-X.X.X.zip”压缩包，其中包含了示例工程demo包。

**步骤3** 导入Eclipse项目。

1. 打开Eclipse。选择“File > Import”弹出“Import”窗口。
2. 选择“Maven > Existing Maven Projects”，单击“Next”，进入“Import Maven Projects”页面。
3. 单击“Browse”按钮，根据实际情况选择“dis-sdk-demo”样例工程的存储位置，勾选样例工程，如图3-5所示。

图 3-5 Import Maven Projects



4. 单击“Finish”完成项目导入。

#### 步骤4 配置Demo工程。

1. 配置项目编码为“UTF-8”。
  - a. 在左侧导航栏“Project Explorer”中右键单击所需工程，选择“Properties”，进入“Properties for dis-sdk-demo”页面。
  - b. 左侧页签栏选择“Resource”，右侧对话框显示“Resource”页面。
  - c. 在“Text file encoding”栏中选择“Other”，单击下拉框选择“UTF-8”。
  - d. 单击“Apply and Close”完成编码配置。
2. 添加JDK。
  - a. 在左侧导航栏“Project Explorer”中右键单击所需工程，选择“Properties”，进入“Properties for dis-sdk-demo”页面。
  - b. 左侧页签栏选择“Java Build Path”，右侧对话框显示“Java Build Path”页面。
  - c. 在“Java Build Path”页面选择“Libraries”页签，单击“Add Library”，弹出“Add Library”对话框。
  - d. 选择“JRE System Library”，单击“Next”确认“Workspace default JRE”为jdk1.8及以上版本。
  - e. 单击“Finish”退出“Add Library”对话框。
  - f. 单击“Apply and Close”完成JDK添加。

**步骤5** 初始化DIS客户端实例。其中，“endpoint”，“ak”，“sk”，“region”，“projectId”信息请参见[获取认证信息](#)。

----结束

# 4 步骤 3：发送数据到 DIS

## 功能简介

将用户本地数据通过DIS通道不断上传至DIS服务。

### 📖 说明

目前数据支持存储至DIS和对象存储服务（Object Storage Service，简称OBS）MapReduce服务（MapReduce Service，简称MRS）、数据仓库服务（Data Warehouse Service，简称DWS）、数据湖探索（Data Lake Insight，简称DLI），具体存储位置在[新增转储任务](#)的“数据转储”中配置。

DIS为临时存储器，存储在DIS中的数据最长保留时间为[步骤3](#)中配置的“生命周期”的值。

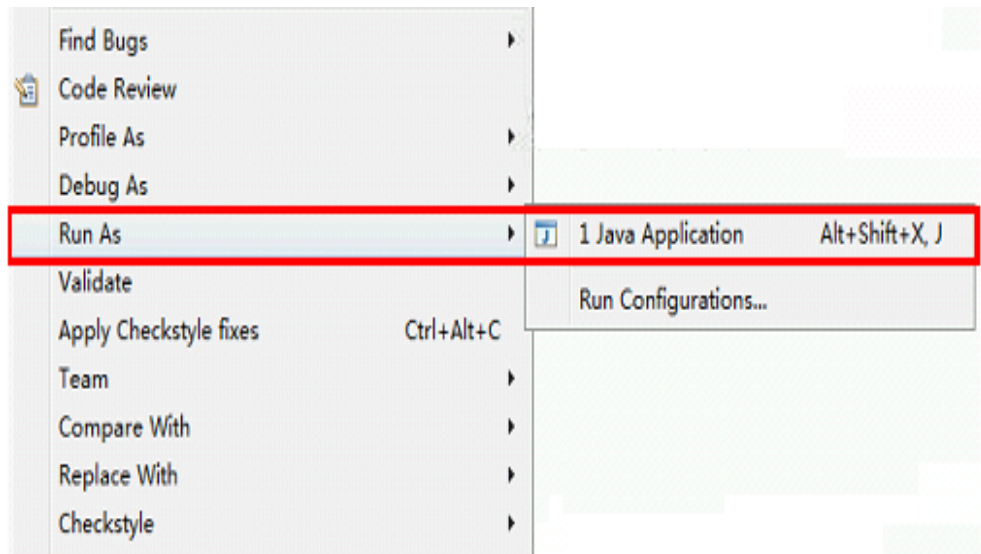
## 样例代码

样例工程为[DIS SDK桶](#)中下载的“huaweicloud-sdk-dis-java-X.X.X.zip”压缩包中“\dis-sdk-demo\src\main\java\com\bigdata\dis\sdk\demo”路径下“ProducerDemo.java”文件。

## 运行程序

程序开发完成后，右键选择“Run As > 1 Java Application”运行程序，如[图4-1](#)所示。

图 4-1 运行上传数据程序



数据上传过程中可在Console控制台查看数据上传通道量信息。出现类似信息表示数据上传成功。

```
14:40:20.090 [main] INFOcom.bigdata.dis.sdk.DISConfig - get from classLoader
14:40:20.093 [main] INFODEMOT - ===== BEGIN PUT =====
14:40:21.186 [main] INFOcom.bigdata.dis.sdk.util.config.ConfigurationUtils - get from classLoader
14:40:21.187 [main] INFOcom.bigdata.dis.sdk.util.config.ConfigurationUtils - propertyMapFromFile size : 2
14:40:22.092 [main] INFOcom.bigdata.dis.sdk.demo.ProducerDemo - Put 3 records[3 successful / 0 failed].
14:40:22.092 [main] INFOcom.bigdata.dis.sdk.demo.ProducerDemo - [hello world.] put success, partitionId
[shardId-0000000000], partitionKey [964885], sequenceNumber [0]
14:40:22.092 [main] INFOcom.bigdata.dis.sdk.demo.ProducerDemo - [hello world.] put success, partitionId
[shardId-0000000000], partitionKey [910960], sequenceNumber [1]
14:40:22.092 [main] INFOcom.bigdata.dis.sdk.demo.ProducerDemo - [hello world.] put success, partitionId
[shardId-0000000000], partitionKey [528377], sequenceNumber [2]
14:40:22.092 [main] INFOcom.bigdata.dis.sdk.demo.ProducerDemo - ===== PUT OVER =====
```

# 5 步骤 4: 从 DIS 获取数据

## 功能简介

从DIS服务中下载数据。

## 样例代码

样例工程为**DIS SDK桶**中下载的“huaweicloud-sdk-dis-java-*X.X.X*.zip”压缩包中“\dis-sdk-demo\src\main\java\com\bigdata\dis\sdk\demo”路径下“ConsumerDemo.java”文件。

## 运行程序

出现类似信息表示下载数据成功：

```
14:55:42.954 [main] INFOcom.bigdata.dis.sdk.DISConfig - get from classLoader
14:55:44.103 [main] INFOcom.bigdata.dis.sdk.util.config.ConfigurationUtils - get from classLoader
14:55:44.105 [main] INFOcom.bigdata.dis.sdk.util.config.ConfigurationUtils - propertyMapFromFile size : 2
14:55:45.235 [main] INFOcom.bigdata.dis.sdk.demo.ConsumerDemo - Get stream
streamName[partitionId=0] cursor success :
eyJnZXRJdGVyYXRvclBhcmFtIjpw7InN0cmVhbS1uYW1lIjoizGlzLTEzbW9uZXkiLCJwYXJ0aXRpb24taWQqOiIiwliwiY
3Vyc29yLXR5cGUiOiIjBVF9TRVfVRU5DRV9OVU1CRViiLCJzdGFydGluZy1zZXF1ZW5jZS1udW1iZXliOilxMDY4O
TcyIn0slmdlbnVvYXRlVGlZNXN0YW1wIjoxNTEzNjY2NjMxMTYxfQ
14:55:45.305 [main] INFOcom.bigdata.dis.sdk.demo.ConsumerDemo - Get Record [hello world.],
partitionKey [964885], sequenceNumber [0].
14:55:45.305 [main] INFOcom.bigdata.dis.sdk.demo.ConsumerDemo - Get Record [hello world.],
partitionKey [910960], sequenceNumber [1].
14:55:46.359 [main] INFOcom.bigdata.dis.sdk.demo.ConsumerDemo - Get Record [hello world.],
partitionKey [528377], sequenceNumber [2].
```



# 6 获取认证信息

---

## 获取 AK/SK

AK/SK (Access Key ID/Secret Access Key) 是用户调用接口的访问密钥。由用户在IAM中创建，可在“我的凭证 > 访问密钥”页面下载生成。

## 获取项目 ID

项目ID表示租户的资源。用户可在“我的凭证 > > API凭证”页面下查看项目列表中不同Region对应的项目ID。

## 获取 region 和 endpoint

请参见[终端节点及区域说明](#)。

# 7 连接 OBS

---

## 介绍

DIS可以向对象存储服务（Object Storage Service，简称OBS）上传数据。

## 前提条件

已参考[自动创建委托](#)创建IAM委托，授权DIS服务去访问用户的OBS。

## 数据转储

用户在[新增转储任务](#)时可设置“数据转储地址”。当“数据转储”设置为“OBS”时，DIS会将通道数据周期性导入OBS。

# 8 自动创建委托

---

用户创建DIS通道，选择将数据转储到对象存储服务（Object Storage Service，简称OBS）、MapReduce服务（MRS）集群、数据仓库服务（Data Warehouse Service，简称DWS）或数据湖探索（Data Lake Insight，简称DLI）中，需要通过创建IAM委托授权DIS服务去访问用户的OBS、MRS、DWS或DLI资源。

1. 使用账号首次进入界面添加转储任务时，系统会自动弹出创建委托界面。
2. 单击“同意授权”则平台会自动创建委托。
3. 委托授权成功后，隶属于该账号下的IAM子用户也可添加转储任务。

# 9 入门实践

当您参考[开通DIS](#)、[从DIS获取数据](#)等一系列操作后，可以根据自身的业务需求使用DIS提供的一系列常用实践。

表 9-1 常用最佳实践

实践		描述
案例	<a href="#">使用DIS实时分析车辆位置</a>	数据接入服务（Data Ingestion Service，简称DIS）实时采集车辆位置数据并上传到华为云的表格存储服务（CloudTable Service，简称CloudTable）中，用户可以使用CloudTable查询指定车辆在指定时间段的车辆位置。
	<a href="#">使用DIS采集增量驾驶行为日志数据</a>	数据接入服务（Data Ingestion Service，简称DIS）采集增量驾驶行为日志数据并上传到华为云对象存储服务（Object Storage Service，简称OBS），通过数据湖探索（Data Lake Insight，简称DLI）分析上传的日志数据，获取驾驶员的驾驶行为，以支持车企提供驾驶习惯优化等增值服务。