

初识数据接入服务(DIS)

1 任务介绍

数据接入服务（Data Ingestion Service，DIS）是华为云提供的实时数据接入服务。它提供了灵活数据采集、高效数据传输、实时数据分发能力，让用户可以轻松构建基于实时数据的分析和应用。

本课程内容的主要是DIS服务快速入门操作指导。课程参与者在华为云页面创建DIS通道，使用DIS Agent采集数据并上传数据至DIS。

2 任务执行

2.1 前提准备

2.1.1 计算机环境准备

本教程适用于已经安装JAVA开发环境的windows系统计算机，请预先准备计算机环境。

2.1.2 拥有华为云账号

本教程基于华为云服务，前提是拥有华为云账户。若尚未注册成为华为云账户，请登录华为云官网注册。

2.1.3 获取账户相关信息

使用DIS服务需要通过用户账户相关信息访问指定资源，用户需要记录以下账户信息：

- 访问密钥（AK/SK）

DIS通过用户帐户中的AK和SK进行签名验证，确保通过授权的帐户才能访问指定的DIS资源。创建访问密钥步骤如下：

1. 登录DIS控制台。
2. 单击页面右上角的用户名，选择“我的凭证”。
3. “我的凭证”页面，单击“管理访问密钥”区域下方的“新增访问密钥”

4. 根据界面提示输入相关信息，会下载一个“credentials.csv”文件，即为访问密钥，中间一列为AK信息，最后一列为SK信息。

说明

每个用户最多可创建两个有效的访问密钥。为防止访问密钥泄露，建议您将其保存到安全的位置。

项目列表 管理访问密钥	
新增访问密钥 您还可以添加0个访问密钥。	
访问密钥ID	创建时间
PM1XKXTX5	2017/09/28 16:42:26 GMT+08:00
OD0387IHK	2017/10/14 16:43:09 GMT+08:00

- 项目ID

项目ID表示租户的资源。用户可登录DIS控制台，单击页面右上角的用户名，选择“我的凭证”。进入“我的凭证”页面，在“项目列表”栏下查看不同Region对应的项目ID。示例如下图：

项目列表 管理访问密钥	
所属区域	项目
华北-北京一	cn-north-1
	497431229f6c451ae

- Region

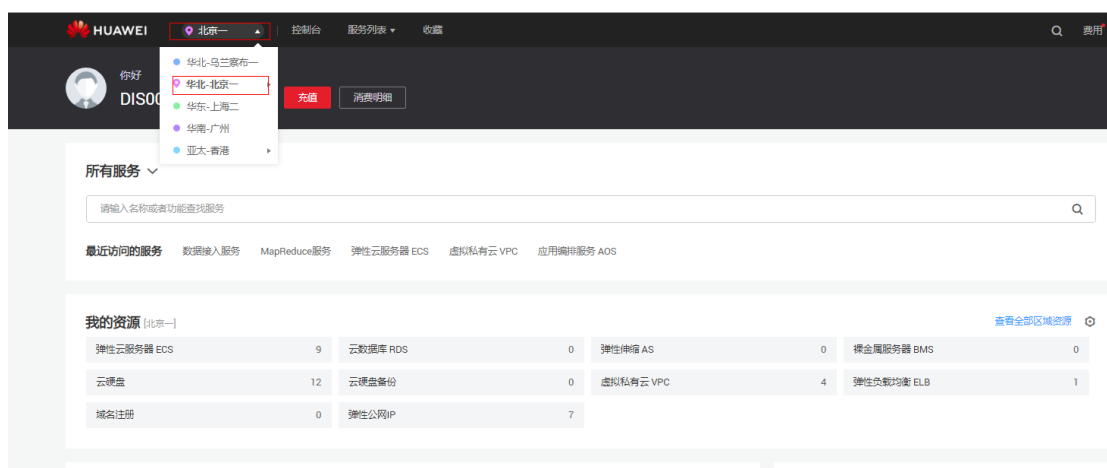
DIS通道所在区域。本教程默认在华北区进行操作，即region信息为cn-north-1。

- Endpoint

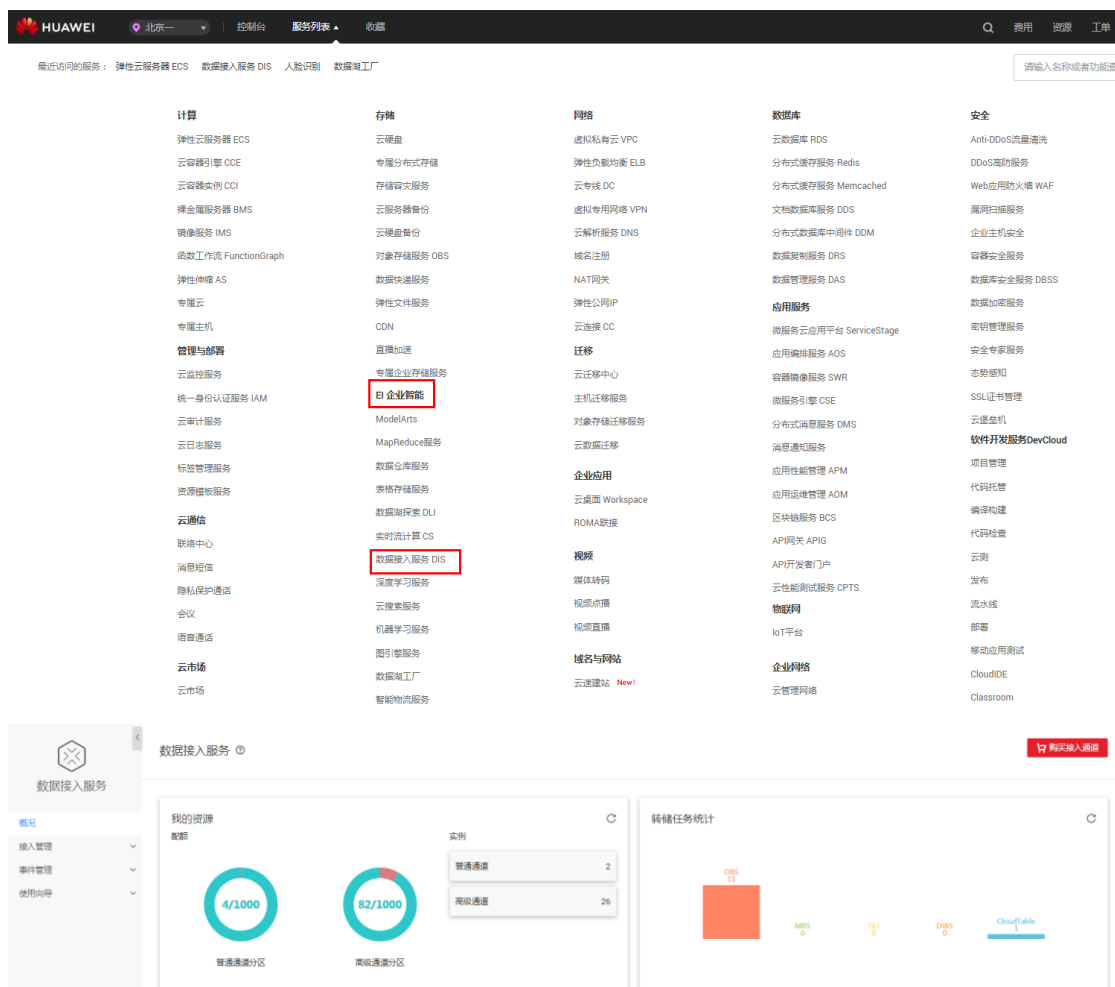
DIS服务数据网关地址。本教程默认在华北区进行操作，即endpoint信息为<https://dis.cn-north-1.myhuaweicloud.com:20004>

2.2 开通 DIS 通道

登录华为云控制台<https://console.huaweicloud.com/console/>，输入用户名密码后，切换到华北-北京一区域。



在“服务列表”下选择“EI 企业智能-数据接入服务”，进入数据接入服务界面，点击右上角“购买接入通道”按钮，创建DIS通道。



新建DIS通道相关参数配置如下：

- 计费模式：按需付费

- 当前区域：华北-北京一
- 通道名称：**dis-demo**（可根据命名规则自行命名，通道名称将用于本教程打卡任务，请使用包含用户名的通道名称）
- 通道类型：高级（普通通道：最高发送速度可达1MB/秒及1000条记录/秒，最高读取速度可达2MB/秒。高级通道：最高发送速度可达5MB/秒及2000条记录/秒，最高读取速度可达10MB/秒。目前高级通道开放免费试用。）
- 分区数量：1
- 生命周期（天）：1
- 源数据类型：BLOB

购买接入通道 ?

返回通道列表

★ 计费模式

按需付费

★ 当前区域

华北-北京一

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。

★ 通道名称

dis-demo

可使用自动生成的由前缀“dis-”加4位随机字符或数字组成的名称，例如：dis-HvB1或者dis_HvB1，也可自定义。

★ 通道类型

普通

高级 ?

★ 分区数量 ?

-

1

+

分区计算

您还有823个可用分区。申请更多分区配额请单击[申请扩大配额](#)。

选择的规格为：高级通道 | 1 个分区 | 通道理论容量：5 MB/秒（接入）或 2000 记录数/秒（接入）；10 MB/秒（读取）

★ 生命周期（天）

-

1

+

?

★ 源数据类型

BLOB

JSON

CSV

FILE

单击“立即购买”按钮，弹出“规格确认”页面，单击“提交”按钮，完成DIS通道购买。

创建DIS通道后，进入DIS控制台“通道管理”页面查看已创建的通道。

数据接入服务

通道管理

您还可以使用 996 个普通通道分区，917 个高级通道分区。

名称/ID

状态

通道类型

分区数量

源数据类型

生命周期（天）

创建时间

计费模式

操作

dis-demo
U1Z1ocMm81H2Bjw1Q

运行中

高级

1

BLOB

1

2018/11/29 15:54:51 GMT+08:00

按需计费

删除 更多

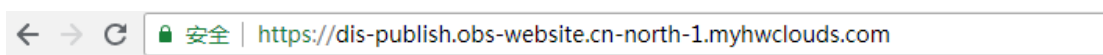
2.3 使用 DIS Agent 采集并上传数据

2.3.1 安装 DIS Agent

使用浏览器登录<https://dis-publish.obs-website.cn-north-1.myhwclouds.com/>，下载

“dis-agent-1.0.4.zip”压缩包，在工作目录解压压缩包。本教程使用目录：

D:\work\dis-agent-1.0.4。



DIS SDK

[dis-sdk-1.3.0.zip](#) ([sha256sum](#))

[dis-sdk-1.2.3.zip](#) ([sha256sum](#))

[dis-sdk-1.2.2.zip](#) ([sha256sum](#))

[dis-sdk-1.2.1.zip](#) ([sha256sum](#))

[dis-sdk-1.2.0.zip](#) ([sha256sum](#))

[dis-kafka-adapter-1.0.4.zip](#) ([sha256sum](#))

[dis-kafka-adapter-1.0.3.zip](#) ([sha256sum](#))

[dis-kafka-adapter-1.0.2.zip](#) ([sha256sum](#))

[dis-kafka-adapter-1.0.1.zip](#) ([sha256sum](#))

[dis-kafka-adapter-1.0.0.zip](#) ([sha256sum](#))

[dis-agent-1.0.4.zip](#) ([sha256sum](#))

[dis-agent-1.0.3.zip](#) ([sha256sum](#))

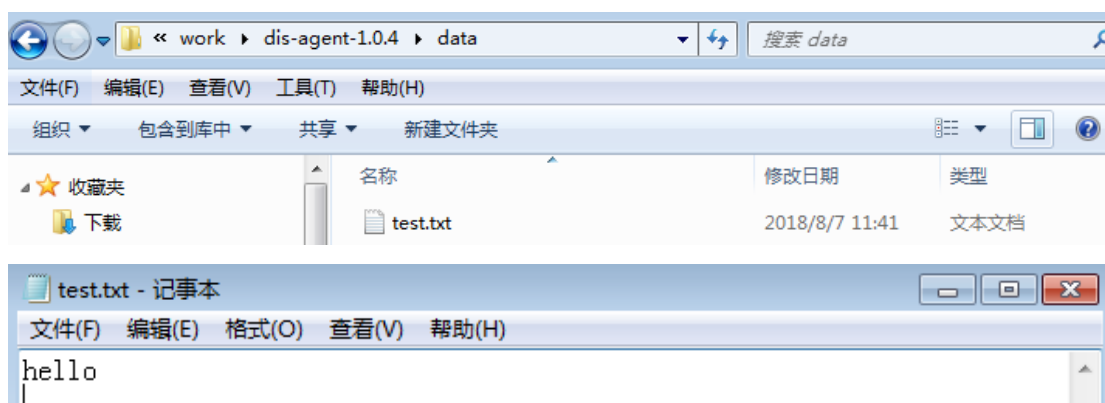
[dis-agent-1.0.2.zip](#) ([sha256sum](#))

[dis-agent-1.0.1.zip](#) ([sha256sum](#))

[dis-agent-1.0.0.zip](#) ([sha256sum](#))

2.3.2 构造数据文件

在工作目录创建要监控的数据文件，当该文件的内容增长时，增长的数据会通过DIS Agent上传至云上DIS。本教程使用的数据文件示例为D:\work\dis-agent-1.0.4\data\test.txt，即在“dis-agent-1.0.4”目录下新建“data”子目录，并在“data”目录下新建一份“test.txt”文件。



2.3.3 配置 DIS Agent

在“\dis-agent-1.0.4\conf”目录中打开“agent.yml”修改DIS Agent的配置。需要自行修改的部分如下图红框所示：

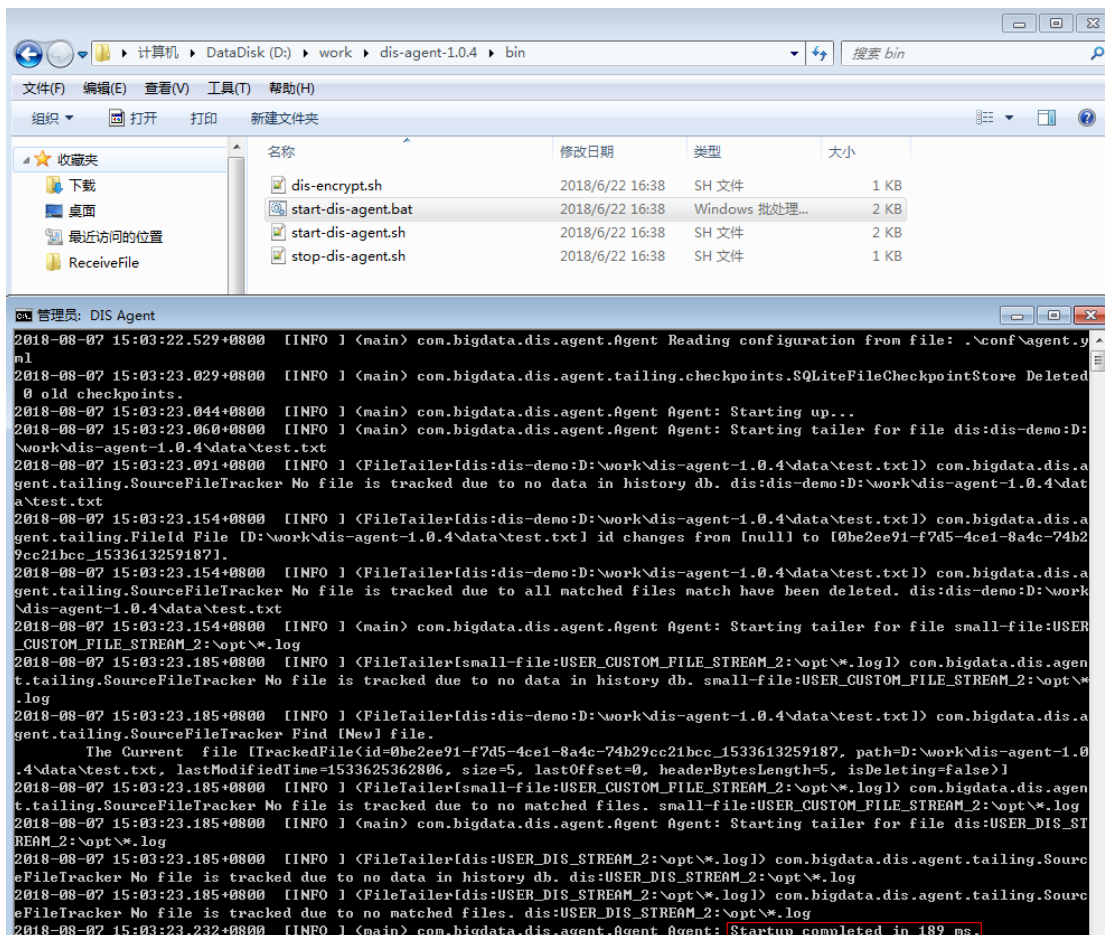
```
1  ---
2  # cloud region id
3  region: cn-north-1
4  # user ak (get from 'My Credential')
5  ak: USER_AK
6  # user sk (get from 'My Credential')
7  sk: USER_SK
8  # user project id (get from 'My Credential')
9  projectId: USER_PROJECTID
10 # the dis gw endpoint
11 endpoint: https://dis.cn-north-1.myhuaweiclouds.com:20004
12 # config each flow to monitor file.
13 flows:
14   # DIS stream
15   - DISStream: USER_DIS_STREAM_1
16     # only support specified directory, filename can use * to match some files. eg. * means match all file, test*.log means match test1.log or test-12.log and so on.
17     filePattern: /tmp/*.log
18     # from where to start: 'START_OF_FILE' or 'END_OF_FILE'
19     initialPosition: START_OF_FILE
20     # upload max interval(ms)
21     maxBufferAgeMillis: 5000
22
23   # Custom File Stream (file maximum size is 128MB)
24   - CustomFileStream: USER_CUSTOM_FILE_STREAM_2
25     # only support specified directory, filename can use * to match some files. eg. * means match all file, test*.log means match test1.log or test-12.log and so on.
26     filePattern: /opt/*.log
27     # from where to start: 'START_OF_FILE' or 'END_OF_FILE'
28     initialPosition: START_OF_FILE
29     # when the file size and lastModifyTime does not change within 'maxFileCheckingMillis', start to upload the file
30     maxFileCheckingMillis: 5000
31
32   # If there are other monitor files, continue to follow the above configuration
33   # another dis stream monitor config
34   - DISStream: USER_DIS_STREAM_2
35     filePattern: /opt/*.log
36     initialPosition: START_OF_FILE
37     maxBufferAgeMillis: 5000
```

配置说明：

- ak：账户的访问密钥信息。参考章节2.1.2获取。
- sk：账户的访问密钥信息。参考章节2.1.2获取。
- projectId：账户的项目ID信息。参考章节2.1.2获取。
- DISStream：使用到的DIS通道信息，即章节2.2中创建的通道名称。
- filePattern：DIS Agent数据采集来源，即使用DIS Agent把该文件中的数据上传至DIS。本教程以章节2.3.2中创建的文件为例，文件路径为D:\work\dis-agent-1.0.4\data\test.txt。

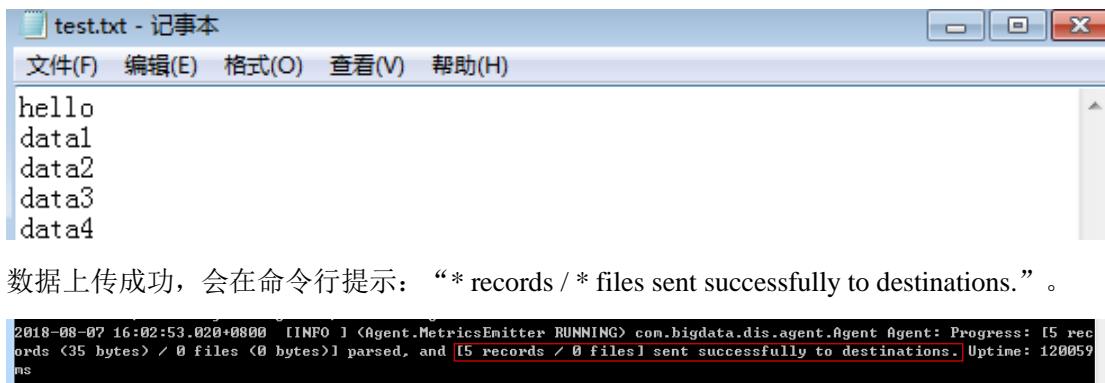
2.3.4 启动 DIS Agent

进入“dis-agent-1.0.4\bin”目录，双击“start-dis-agent.bat”脚本。命令行出现“Startup completed in **ms”表示DIS Agent启动成功。



2.3.5 追加数据并上传

在“data\test.txt”中追加数据并保存，DIS Agent会将新增的数据上传。



数据上传成功，会在命令行提示：“* records / * files sent successfully to destinations.”。

2.3.6 查看数据上传情况

进入华为云DIS控制台，点击左侧“接入管理->通道管理”，进入“通道管理”界面，点击章节2.2创建的DIS通道。

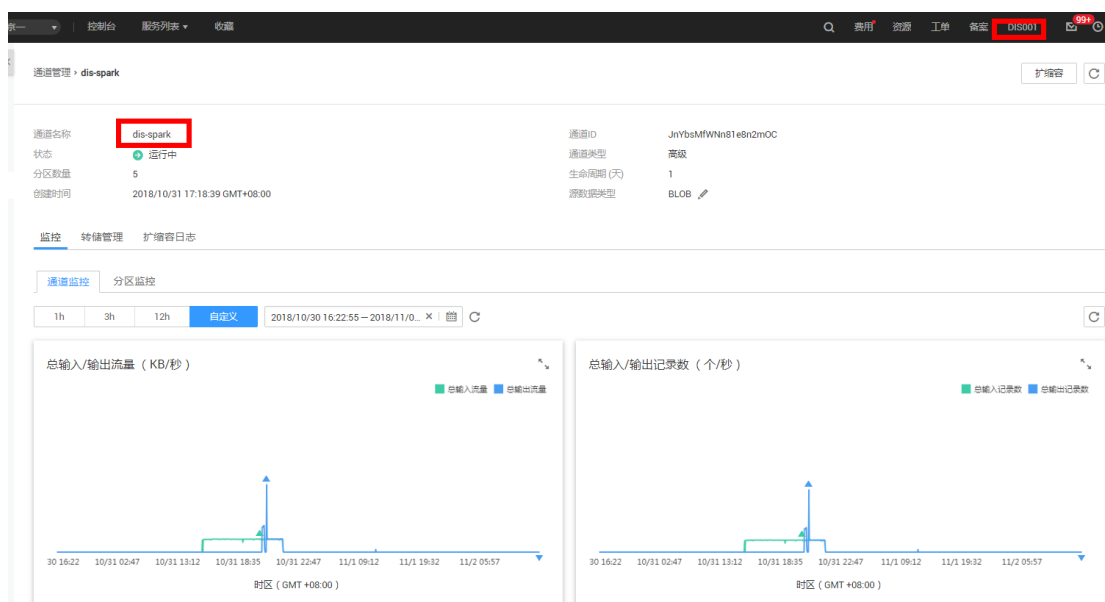


在“监控 > 通道监控”界面查看数据是否从用户侧上传至云上DIS服务。



3 打卡任务

完成上述操作，参照章节2.3.6中截取“监控 > 通道监控”界面完成打卡，注意截图信息需要包含用户名信息（仅用于记录打卡）。



4 温馨提示

明日课程需要用户自行申请cloudtable集群，点击该网址

<https://console.huaweicloud.com/cloudtable/?region=cn-north-1#/cloudtable/management/dashboa>

rd进入cloudtable控制台，点击右上角按钮“免费试用”即可，立即购买（使用默认参数）

免费试用申请时间：12月4日 10:00-12月5日 24:00