数据接入服务

最佳实践

文档版本	01
发布日期	2023-06-20

华为云计算技术有限公司

商标声明


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。

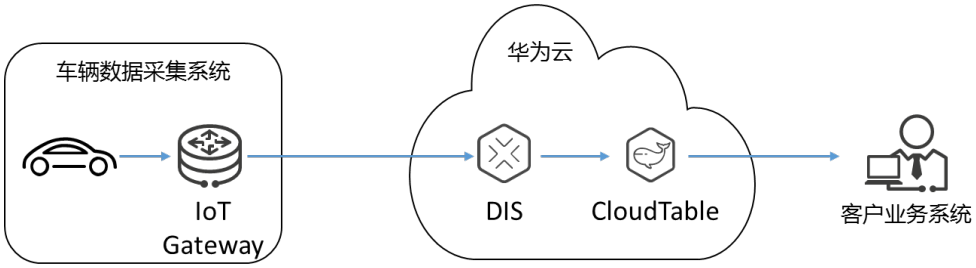
目 录

1 使用 DIS 实时分析车辆位置

场景介绍

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。

图 1-1 业务流程图



本次实践基本流程如下所示：

1. **申请CloudTable集群**
2. **在CloudTable中创建数据表**
3. **申请DIS通道**
4. **添加转储任务**
5. **获取认证信息**
6. **准备DIS应用开发环境**
7. **编写发送数据到DIS的应用程序**
8. **启动数据上传程序**
9. **在CloudTable中查看上传数据**
10. **CloudTable查询指定车辆位置**

申请 CloudTable 集群

创建一个CloudTable集群用于存放DIS转储的数据。

在 CloudTable 中创建数据表

用户创建DIS通道，选择将数据转储到CloudTable中，需要创建CloudTable数据表。

采集获得数据是JSON格式，样例如下：

```
{"DeviceID":"4d3a27c13dc21ae056044b818a03dwen002","Mileage":"55378500","DataTime":"2017-10-23 12:19:35.000","Latitude":"34.585639","IsACCOpen":"true","Longitude":"119.193524","Velocity":0,"Direction":"null","Carnum":"WL66666","BaiDuLatitude":"34.607069","BaiDuLongitude":"119.190093","BaiDuAdress":"江苏省连云港市新浦区通灌北路78号","ReceiveTime":"2017-10-23 12:19:34.846","Altitude":"null"}
```

本实践中，通过使用HBase shell客户端完成建表操作。

步骤1 准备Linux弹性云服务器。假设该弹性云服务器名称为“ecs-385d”。

步骤2 安装客户端并启动Shell访问CloudTable集群。

步骤3 在HBase shell客户端执行**create 'tbl1',{NAME => 'i'}**命令，创建数据表。界面显示如下表示创建成功。

----结束

申请 DIS 通道

请参见**开通DIS通道**创建通道。

添加转储任务

步骤1 使用注册帐户登录DIS控制台。

步骤2 在左侧列表栏中选择“通道管理”。

步骤3 单击**申请DIS通道**中创建的通道名称，进入所选通道的管理页面，选择“转储管理”页签。

步骤4 单击“添加转储任务”按钮，在弹出的“添加转储任务”页面配置转储相关配置项。

说明

- 每个通道最多可创建5个转储任务。
- 源数据类型为FILE的通道，不允许添加转储任务。

步骤5 单击“立即创建”。

表 1-1 转储任务参数说明

参数	参数解释	配置值
转储服务类型	选择CloudTable，通道里的流式数据存储在DIS中，并实时导入表格存储服务Cloudtable集群的HBase表和OpenTSDB。	CloudTable

参数	参数解释	配置值
任务名称	用户创建转储任务时，需要指定转储任务名称，同一通道的转储任务名称不可重复。任务名称由英文字母、数字、中划线和下划线组成。长度为1～64个字符。	-
偏移量	• 最新：最大偏移量，即获取最新的数据。 • 最早：最小偏移量，即读取最早的数据。	最新
CloudTable集群	单击“选择”，在“选择CloudTable集群”窗口选择一个集群名称。 此配置项不可配置为空。仅支持选择，不可手动输入。	cloudtable-demo
CloudTable表类型	HBase和openTSDB两种。	HBase
CloudTable数据表	CloudTable数据表：单击“选择”，在“选择CloudTable数据表”窗口选择一个数据表。 此处路径仅支持选择，不可手动输入。 **说明** 配置此项必须已配置“CloudTable集群”并创建了HBase表。	tbl1
备份开关	用户数据转储CloudTable服务失败时，是否将转储失败的数据备份至OBS服务。 • 开启：是，转储失败的数据备份至OBS服务。 • 关闭：否，转储失败的数据不备份至OBS服务。 开关默认关闭。 **说明** 关闭开关，转储失败的数据会存储在DIS中，并在“生命周期”配置的时间到达时将数据清除。	关闭

参数	参数解释	配置值
Row Key	• Json属性名，取值范围为英文字母、数字、下划线和小数点，最大取值为32个字符，不可为空，不可以小数点开头，不可包含连续的小数点且不可以小数点结尾。最多可添加64个属性。 • 数据类型，从下拉框选择。 – Bigint – Double – Boolean – Timestamp – String – Decimal	-
Row Key 分隔符	支持“.”、“,”、“|”、“;”、“-”、“_”、和“~”七种字符取值，也可配置为NULL。 最大长度为一个字符。	-
Schema 列	• 列名，取值范围为英文字母、数字和下划线，最大取值为32个字符，不可为空。最多可添加4096个列。 • 数据类型，从下拉框选择。 – Bigint – Double – Boolean – Timestamp – String – Decimal • Json属性名，取值范围为英文字母、数字、下划线和小数点，最大取值为32个字符，不可为空，不可以小数点开头，不可包含连续的小数点且不可以小数点结尾。 • 所属列族，从下拉框选择，不可为空。配置此项必须已配置“CloudTable 集群”、“CloudTable 数据表”且CloudTable表类型为HBase。	参见表1-2表2Schema 列填写。

表 1-2 Schema 列填写

列名	数据类型	JSON属性名	列族
DeviceID	String	DeviceID	i
Mileage	Bigint	Mileage	i
Latitude	Decimal	Latitude	i
IsACCOpen	Boolean	IsACCOpen	i
Longitude	Decimal	Longitude	i
Velocity	Bigint	Velocity	i
Direction	String	Direction	i
BaiDuLatitude	Decimal	BaiDuLatitude	i
BaiDuLongitude	Decimal	BaiDuLongitude	i
BaiDuAdress	String	BaiDuAdress	i
ReceiveTime	Timestamp	ReceiveTime	i
Altitude	String	Altitude	i

----结束

获取认证信息

- 获取AK/SK

  您可以通过如下方式获取访问密钥。

  a. 登录控制台，在用户名下拉列表中选择“我的凭证”。

  b. 进入“我的凭证”页面，选择“访问密钥 > 新增访问密钥”，如图1-2所示。

  图 1-2 单击新增访问密钥

  c. 单击“确定”，根据浏览器提示，保存密钥文件。密钥文件会直接保存到浏览器默认的下载文件夹中。打开名称为“credentials.csv”的文件，即可查看访问密钥（Access Key Id和Secret Access Key）。

  说明

  - 每个用户仅允许新增两个访问密钥。
  - 为保证访问密钥的安全，访问密钥仅在初次生成时自动下载，后续不可再次通过管理控制台界面获取。请在生成后妥善保管。

- 获取项目ID和帐号ID

项目ID表示租户的资源，帐号ID对应当前帐号。用户可在对应页面下查看不同Region对应的项目ID和帐号ID。

a. 注册并登录管理控制台。

b. 在用户名的下拉列表中单击“我的凭证”。

c. 在“API凭证”页面，查看帐号名和帐号ID，在项目列表中查看项目ID。

- 获取endpoint

  终端节点（Endpoint）即调用API的**请求地址**，不同服务不同区域的终端节点不同。本服务的Endpoint可从**终端节点Endpoint**获取。

准备 DIS 应用开发环境

具体操作请参见**准备DIS应用开发环境**。

编写发送数据到 DIS 的应用程序

步骤1 准备数据样例。

步骤2 修改样例代码。

样例工程为**准备DIS应用开发环境**中下载的“huaweicloud-sdk-dis-java-.zip ”压缩包“\dis-sdk-demo\src\main\java\com\bigdata\dis\sdk\demo”路径下的“ProducerDemo.java”文件。

根据实际情况更改“AK”、“SK”和“ProjectId”的值。

```
private static void runProduceDemo()
  {
      // 创建DIS客户端实例
      DIS dic = DISClientBuilder.standard()
              .withEndpoint("https://dis.cn-north-1.myhuaweicloud.com:20004")
              .withAk("${your_AK}")
              .withSk("${your_SK}")
              .withProjectId("${your_projectId}")
              .withRegion("cn-north-1")
              .build();

      // 配置流名称
      String streamName = "dis-demo";

      // 配置上传的数据
      PutRecordsRequest putRecordsRequest = new PutRecordsRequest();
      putRecordsRequest.setStreamName(streamName);

      List<PutRecordsRequestEntry> putRecordsRequestEntryList = new
ArrayList<PutRecordsRequestEntry>();
       String[] messages = {     此处填写上一步准备的数据样例    };

      for (int i = 0; i < messages.length; i++)
      {
         ByteBuffer buffer = ByteBuffer.wrap(messages[i].getBytes());
          PutRecordsRequestEntry putRecordsRequestEntry = new PutRecordsRequestEntry();
          putRecordsRequestEntry.setData(buffer);

putRecordsRequestEntry.setPartitionKey(String.valueOf(ThreadLocalRandom.current().nextInt(1000000)));
          putRecordsRequestEntryList.add(putRecordsRequestEntry);
      }
      putRecordsRequest.setRecords(putRecordsRequestEntryList);

      log.info("========== BEGIN PUT ============");

      PutRecordsResult putRecordsResult = null;
```

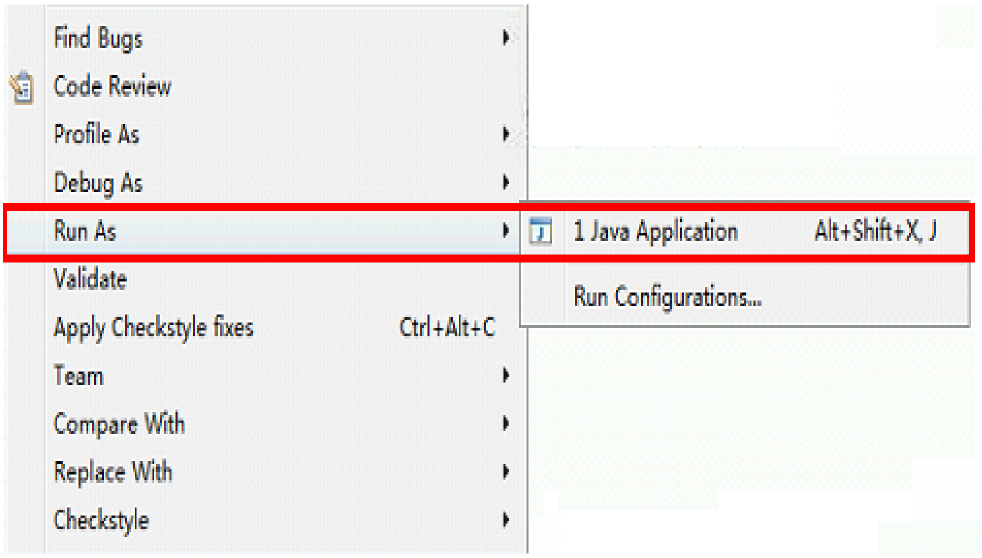
```
    try
    {
        putRecordsResult = dic.putRecords(putRecordsRequest);
    }
    catch (DISClientException e)
    {
        log.error("Failed to get a normal response, please check params and retry. Error message [{}]",
            e.getMessage(),
            e);
    }
    catch (ResourceAccessException e)
    {
        log.error("Failed to access endpoint. Error message [{}]", e.getMessage(), e);
    }
    catch (Exception e)
    {
        log.error(e.getMessage(), e);
    }

    if (putRecordsResult != null)
    {
        log.info("Put {} records[{} successful / {} failed].",
            putRecordsResult.getRecords().size(),
            putRecordsResult.getRecords().size() - putRecordsResult.getFailedRecordCount().get(),
            putRecordsResult.getFailedRecordCount());

        for (int j = 0; j < putRecordsResult.getRecords().size(); j++)
        {
            PutRecordsResultEntry putRecordsRequestEntry = putRecordsResult.getRecords().get(j);
            if (putRecordsRequestEntry.getErrorCode() != null)
            {
                // 上传失败
                log.error("[{}] put failed, errorCode [{}], errorMessage [{}]",
                    new String(putRecordsRequestEntryList.get(j).getData().array()),
                    putRecordsRequestEntry.getErrorCode(),
                    putRecordsRequestEntry.getErrorMessage());
            }
            else
            {
                // 上传成功
                log.info("[{}] put success, partitionId [{}], partitionKey [{}], sequenceNumber [{}]",
                    new String(putRecordsRequestEntryList.get(j).getData().array()),
                    putRecordsRequestEntry.getPartitionId(),
                    putRecordsRequestEntryList.get(j).getPartitionKey(),
                    putRecordsRequestEntry.getSequenceNumber());
            }
        }
    }
    log.info("========== END PUT ============");
}
```

----结束

启动数据上传程序

程序开发完成后，右键选择“Run As > 1 Java Application”运行程序，如图1-3所示。

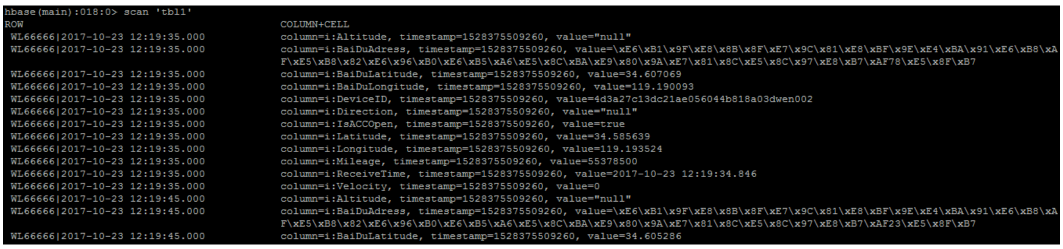
图 1-3 运行上传数据程序



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。

在 CloudTable 中查看上传数据

在HBase shell客户端执行**scan 'tbl1'**命令，显示如下表示数据上传成功。



CloudTable 查询指定车辆位置

以查询车辆“**WL66666**”在**2017-10-23 12:22:00**时间以后的位置为例。

步骤1 登录HBase shell客户端。

步骤2 执行**scan 'tbl1',{COLUMNS => ['i:Latitude','i:Longitude'], FILTER=>"RowFilter(>=,'binary:WL66666|2017-10-23 12:22:00')"}**命令，查询车辆位置。

HBase客户端查询结果如下所示。

```
hbase(main):024:0> scan 'tbl1',{COLUMNS => ['i:Latitude','i:Longitude'], FILTER=>"RowFilter(>=,'binary:WL66666|2017-10-23 12:22:00')"}
ROW                                   COLUMN+CELL
 WL66666|2017-10-23 12:22:25.000      column=i:Latitude, timestamp=1528375509260, value=34.587286
 WL66666|2017-10-23 12:22:25.000      column=i:Longitude, timestamp=1528375509260, value=119.190901
 WL66666|2017-10-23 12:22:35.000      column=i:Latitude, timestamp=1528375509260, value=34.587913
 WL66666|2017-10-23 12:22:35.000      column=i:Longitude, timestamp=1528375509260, value=119.190339
 WL66666|2017-10-23 12:22:45.000      column=i:Latitude, timestamp=1528375509260, value=34.588418
 WL66666|2017-10-23 12:22:45.000      column=i:Longitude, timestamp=1528375509260, value=119.189996
 WL66666|2017-10-23 12:22:55.000      column=i:Latitude, timestamp=1528375509260, value=34.588466
 WL66666|2017-10-23 12:22:55.000      column=i:Longitude, timestamp=1528375509260, value=119.189966
 WL66666|2017-10-23 12:23:05.000      column=i:Latitude, timestamp=1528375509260, value=34.588468
 WL66666|2017-10-23 12:23:05.000      column=i:Longitude, timestamp=1528375509260, value=119.189966
 WL66666|2017-10-23 12:23:15.000      column=i:Latitude, timestamp=1528375509260, value=34.588574
 WL66666|2017-10-23 12:23:15.000      column=i:Longitude, timestamp=1528375509260, value=119.189836
 WL66666|2017-10-23 12:23:25.000      column=i:Latitude, timestamp=1528375509260, value=34.589244
 WL66666|2017-10-23 12:23:25.000      column=i:Longitude, timestamp=1528375509260, value=119.189254
 WL66666|2017-10-23 12:23:35.000      column=i:Latitude, timestamp=1528375509260, value=34.589901
 WL66666|2017-10-23 12:23:35.000      column=i:Longitude, timestamp=1528375509260, value=119.188708
 WL66666|2017-10-23 12:23:45.000      column=i:Latitude, timestamp=1528375509260, value=34.590371
 WL66666|2017-10-23 12:23:45.000      column=i:Longitude, timestamp=1528375509260, value=119.188324
 WL66666|2017-10-23 12:23:55.000      column=i:Latitude, timestamp=1528375509260, value=34.590391
 WL66666|2017-10-23 12:23:55.000      column=i:Longitude, timestamp=1528375509260, value=119.188309
10 row(s) in 0.0610 seconds
```

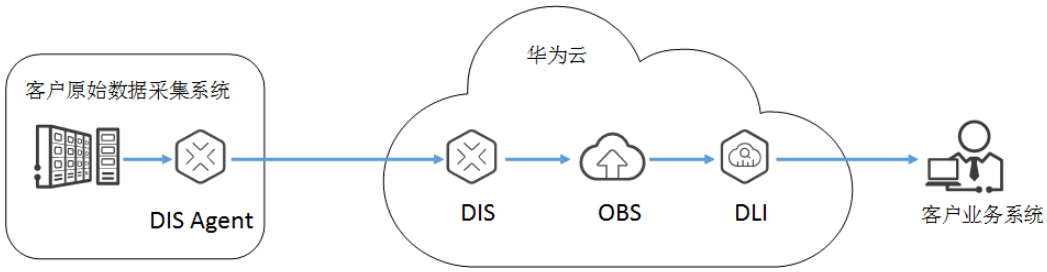
----结束

2 使用 DIS 采集增量驾驶行为日志数据

场景简介

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。

图 2-1 业务流程图



本次实践基本流程如下所示：

1. **申请OBS桶**
2. **申请DIS通道**
3. **添加转储任务**
4. **获取认证信息**
5. **安装Agent**
6. **准备数据样例**
7. **配置DIS Agent**
8. **启动DIS Agent**
9. **在OBS查看上传文件**
10. **创建数据库**
11. **创建OBS表**
12. **查询数据样例**
13. **结果查询**

申请 OBS 桶

创建一个OBS桶用来存放DIS转储的数据，请参见**创建桶**。

申请 DIS 通道

请参见**开通DIS通道**创建通道。

添加转储任务

步骤1 使用注册帐户登录DIS控制台。

步骤2 在左侧列表栏中选择“通道管理”。

步骤3 单击**申请DIS通道**中创建的通道名称，进入所选通道的管理页面，选择“转储管理”页签。

步骤4 单击“添加转储任务”按钮，在弹出的“添加转储任务”页面配置转储相关配置项。

说明

- 每个通道最多可创建5个转储任务。
- 源数据类型为FILE的通道，不允许添加转储任务。

步骤5 单击“立即创建”。

表 2-1 转储任务参数说明

参数	参数解释	配置值
转储服务类型	选择OBS。 通道里的流式数据存储在DIS中，并周期性导入对象存储服务（Object Storage Service，简称OBS）。 通道里的实时文件数据传输完成后，导入OBS。	OBS
任务名称	用户创建转储任务时，需要指定转储任务名称，同一通道的转储任务名称不可重复。任务名称由英文字母、数字、中划线和下划线组成。长度为1～64个字符。	-
转储文件格式	• text • csv • parquet • carbon	根据需要选择。
数据转储地址	存储该通道数据的OBS桶名称。桶名称在“对象存储服务”中“创建桶”时创建。	**申请DIS通道**创建的桶名称。

参数	参数解释	配置值
转储文件目录	在OBS中存储通道文件的自定义目录，多级目录可用“/”进行分隔，不能以“/”开头。 取值范围：0~50个字符。 默认配置为空。	-
时间目录格式	数据将存储在OBS桶中转储文件目录下，按时间格式作为层级的目录中。 当选择的时间目录格式精确到日时，存储目录为“桶名称/转储文件目录/年/月/日”。 取值范围： ● N/A：置空，不使用日期时间目录。 ● yyyy：年 ● yyyy/MM：年/月 ● yyyy/MM/dd：年/月/日 ● yyyy/MM/dd/HH：年/月/日/时 ● yyyy/MM/dd/HH/mm：年/月/日/时/分 此配置项仅支持选择，不可手动输入。	-
记录分隔符	进行OBS周期转储时，分隔不同转储记录的分隔符。 取值范围： ● 逗号 "," ● 分号 ";" ● 竖线 "|" ● 换行符 "\n" ● NULL 此配置项仅支持选择，不可手动输入。	-
偏移量	● 最新：最大偏移量，即获取最新的数据。 ● 最早：最小偏移量，即读取最早的数据。	最新

参数	参数解释	配置值
数据转储周期	根据用户配置的时间，周期性的将数据导入OBS，若某个时间段内无数据，则此时间段不会生成打包文件。 取值范围：30～900。 单位：秒。 默认配置为300秒。	-

----结束

获取认证信息

- 获取AK/SK

  您可以通过如下方式获取访问密钥。

  a. 登录控制台，在用户名下拉列表中选择“我的凭证”。

  b. 进入“我的凭证”页面，选择“访问密钥 > 新增访问密钥”，如图2-2所示。

  图 2-2 单击新增访问密钥

  c. 单击“确定”，根据浏览器提示，保存密钥文件。密钥文件会直接保存到浏览器默认的下载文件夹中。打开名称为“credentials.csv”的文件，即可查看访问密钥（Access Key Id和Secret Access Key）。

  说明

  - 每个用户仅允许新增两个访问密钥。
  - 为保证访问密钥的安全，访问密钥仅在初次生成时自动下载，后续不可再次通过管理控制台界面获取。请在生成后妥善保管。

- 获取项目ID和帐号ID

  项目ID表示租户的资源，帐号ID对应当前帐号。用户可在对应页面下查看不同Region对应的项目ID和帐号ID。

  a. 注册并登录管理控制台。

  b. 在用户名的下拉列表中单击“我的凭证”。

  c. 在“API凭证”页面，查看帐号名和帐号ID，在项目列表中查看项目ID。

- 获取endpoint

  终端节点（Endpoint）即调用API的**请求地址**，不同服务不同区域的终端节点不同。本服务的Endpoint可从**终端节点Endpoint**获取。

安装 Agent

步骤1 获取Agent安装包。

步骤2 解压“dis-agent-*X.X.X*.zip”压缩包到当前文件夹。

----结束

准备数据样例

步骤1 获取数据样例压缩包。

步骤2 解压压缩包到当前文件夹。

----结束

配置 DIS Agent

步骤1 使用文件管理器进入DIS Agent程序的conf目录，例如“C:\dis-agent-X.X.X|*conf*”。

步骤2 使用编辑器打开“agent.yml”文件，根据实际情况修改各配置项的值并保存。

说明

- 各配置项与值之间必须以英文格式的“冒号+空格”形式分隔。
- “agent.yml”文件为linux格式，建议使用“Sublime Text”工具编辑文件。

表 2-2 agent.yml 配置文件说明

配置项	是否必填	说明	默认值
region	是	DIS服务所在区域。 获取方式请参见**获取认证信息**。	-
ak	是	用户的Access Key。 获取方式请参见**获取认证信息**。	请根据实际情况配置
sk	是	用户的Secret Key。 获取方式请参见**获取认证信息**。	请根据实际情况配置
projectId	是	用户所属区域的项目ID。 获取方式请参见**获取认证信息**。	请根据实际情况配置
endpoint	是	DIS数据网关地址。格式：https://DIS终端节点。 获取方式请参见**获取认证信息**。	-

配置项	是否必填	说明	默认值
body.serialize.type	否	DIS数据包上传格式。（非原始数据格式） • json：DIS数据包封装为json格式，满足普通使用。 • protobuf：DIS数据包封装为二进制格式，可以减少体积约1/3，在数据量较大的情况下推荐使用此格式。	json
body.compress.enabled	否	是否开启传输数据压缩。	false
body.compress.type	否	开启压缩时选择的数据压缩格式，目前支持的压缩格式如下： lz4：综合来看效率最高的压缩算法,更加侧重压缩解压速度,压缩比并不是第一。 zstd：一种新的无损压缩算法，旨在提供快速压缩，并实现高压缩比。	lz4
PROXY_HOST	否	配置代理IP，请求走代理服务器的需要配置。	请根据实际情况配置
PROXY_PORT	否	配置代理端口。	80
PROXY_PROTOCOL	否	配置代理协议。支持http和https。	http
PROXY_USERNAME	否	配置代理用户名。	请根据实际情况配置
PROXY_PASSWORD	否	配置代理密码。	请根据实际情况配置
[flows] 监控的文件信息，可同时配置多个监控文件信息。 当前支持如下模式上传： DISStream:持续监控文本文件，实时收集增量数据按分隔符解析并上传到DIS通道(通道源数据类型为BLOB/JSON/CSV)，配置项说明请参见**表2-3**。 具体配置格式可以参见版本包中的“agent.yml”的样例。			

表 2-3 DISStream 配置项说明

配置项	是否必填	说明	默认值
DISStream	是	DIS 通道名称。 将“filePattern”所匹配到的文件内容按分隔符解析并上传到此通道。	请根据实际情况配置
filePattern	是	文件监控路径，只能监控一个目录下的文件，无法递归目录监控。 如果要监控多个目录，可以在flows下面配置多个“DISStream”，文件名可使用“*”进行匹配。 ● “/tmp/*.log”表示匹配“/tmp”目录下所有以“.log”结尾的文件。 ● “/tmp/access-*.log”表示匹配“/tmp”目录下所有以“access-”开头，以“.log”结尾的文件。 ● Windows上路径范例为“D:\logs*.log”。	请根据实际情况配置
directoryRecursionEnabled	否	是否查找子目录 ● false：不递归查找子目录，只匹配根目录下的文件 ● true： 递归查找所有子目录。如filePattern配置为/tmp/*.log，此时可以匹配到/tmp/one.log，/tmp/child/two.log，/tmp/child/child/three.log	false
initialPosition	否	监控起始位置。 ● END_OF_FILE：开始启动时不解析当前匹配的文件，而是从新增文件或新增的内容开始按分隔符解析并上传。 ● START_OF_FILE：将“filePattern”配置的所有匹配文件按照修改时间，从旧到新按分隔符解析并上传到DIS服务。	START_OF_FILE
maxBufferAgeMillis	否	最长上传等待时间。 单位：毫秒 ● 记录队列满则立即上传。 ● 记录队列未满，等待此配置项配置的时间后上传到DIS服务。	5000
maxBufferSizeRecords	否	记录队列缓存的最大记录数，如果队列达到此值则立刻上传这批数据。	500

配置项	是否必填	说明	默认值
partitionKeyOption	否	每条记录会携带一个PartitionKey，相同PartitionKey的记录会分配到同一个分区。此配置项可设置每条记录的PartitionKey值，取值如下： ● RANDOM_INT：PartitionKey的值为随机数字的字符串，记录均匀分布在每个分区。 ● FILE_NAME：PartitionKey的值为文件名称字符串，记录分布在特定的一个分区中。 ● FILE_NAME,RANDOM_INT：PartitionKey的值为文件名称字符串与随机数字字符串的组合体，以英文逗号分隔，记录携带所属的文件名并均匀分布在所有分区。	RANDOM_INT
recordDelimiter	否	每条记录之间的分隔符。 取值范围：任意一个字符，且包含在双引号内。 取值不可为空，即该配置项不可配置为“ ”。 **说明** 如果取值为特殊字符，使用反斜杠（\）转义，如分隔符为引号（"），可配置为"\""，如果为反斜杠（\），可配置为"\\"。 如果为控制字符如STX（正文开始），可配置为"\u0002"。	"\n"
isRemainRecordDelimiter	否	上传记录时，是否携带分隔符。 ● true：携带分隔符。 ● false：不携带分隔符。	false
isFileAppendable	否	文件是否有追加内容的可能。 ● true：文件可能会追加内容。Agent持续监控文件，若文件追加了内容则根据recordDelimiter解析后上传记录。此时要保证文件以recordDelimiter结尾，否则Agent会认为文件追加未完成，继续等待recordDelimiter写入。 ● false：文件不会追加内容。文件最后一行不以recordDelimiter结尾，Agent仍会当做最后一条记录上传，上传完成后根据“deletePolicy”和“fileSuffix”的配置执行文件删除或重命名操作。	true

配置项	是否必填	说明	默认值
maxFileCheckingMillis	否	最长文件变动检查时间，如果文件在此时间内“大小”、“修改时间”和“文件ID”都没有变化，则认为文件已经完成并开始上传。 请根据实际文件变动的频率配置此值，避免文件未完成已开始上传的情况。 若文件上传后有变动，则会重新全量上传。 单位：毫秒 **说明** “isFileAppendable”配置为“false”时该配置项生效。	5000
deletePolicy	否	文件内容上传完成之后的删除策略。 ● never：文件内容上传完毕后不删除文件。 ● immediate：文件内容上传完毕后删除文件 。 **说明** “isFileAppendable”配置为“false”时该配置项生效。	never
fileSuffix	否	文件内容上传完成之后添加的文件名后缀。 例如：原文件名为“x.txt”，“fileSuffix”配置为“.COMPLETED”，则文件上传后的命名为“x.txt.COMPLETED”。 **说明** “isFileAppendable”配置为“false”，同时“deletePolicy”配置为“never”，该配置项生效。	.COMPLETED
sendingThreadSize	否	发送线程数。默认单线程发送。 **须知** 使用多线程会导致如下问题： ● 数据发送不保证顺序。 ● 程序异常停止并重新启动时会丢失部分数据。	1
fileEncoding	否	文件编码格式，支持UTF8, GBK, GB2312, ISO-8859-1等	UTF8

配置项	是否必填	说明	默认值
resultLogLevel	否	每次调用DIS数据发送接口后的结果日志级别。 ● OFF：日志中不输出每次接口调用的结果。 ● INFO：每次接口调用的结果以INFO级别输出到日志。 ● WARN：每次接口调用的结果以WARN级别输出到日志。 ● ERROR：每次接口调用的结果以ERROR级别输出到日志。	INFO

步骤3 （可选）使用Windows自带的记事本修改“agent.yml”文件，需要将文件另存为选“UTF-8”编码。

1. 选择“文件 > 另存为”。
2. 在弹出的“另存为”窗口中选择“编码”为“UTF-8”。
3. 单击“保存”，弹出“确认另存为”对话框。
4. 单击“是”。

----结束

启动 DIS Agent

步骤1 使用文件管理器进入DIS Agent程序的bin目录，例如“C:\dis-agent-X.X.X|*bin*”。

步骤2 双击“start-dis-agent.bat”文件，在弹出的控制台窗口显示如下内容表示启动成功。

```
[INFO ] (main) com.bigdata.dis.agent.Agent Agent: Startup completed in XXX ms.
```

DIS Agent启动后会立即上传文件，并持续打印日志。如果没有ERROR日志表示上传正常。

当日志输出不频繁(每30s打印一次)，且有如下类似信息，表示已经上传完成。

```
Agent: Progress: [0 records (0 bytes) / 10 files (32573229 bytes)] parsed, and [0  records / 10 files] sent successfully to destinations. Uptime: 30146ms
```

----结束

在 OBS 查看上传文件

步骤1 登录对象存储服务管理控制台。

步骤2 在左侧导航栏选择“桶列表”页签。

步骤3 在右侧表格中，“桶名称”列单击对应的桶名称，即**申请DIS通道**中配置的“桶名称”。

步骤4 在弹出的桶页面中单击左侧导航栏“对象”页签，查看已上传的文件。

----结束

创建数据库

步骤1 在Console页面上方菜单栏中单击“产品”，单击“大数据”分类中的“数据湖探索DLI”。

步骤2 创建demo数据库，在DLI控制台总览页面，选择“SQL作业”，单击“创建作业”，进入SQL作业编辑器。

步骤3 在SQL作业编辑器左侧，选择 “数据库”，单击创建数据库。

说明

“default”为内置数据库，不能创建名为“default”的数据库。

----结束

创建 OBS 表

步骤1 选择demo数据库，在编辑框中输入以下SQL语句：

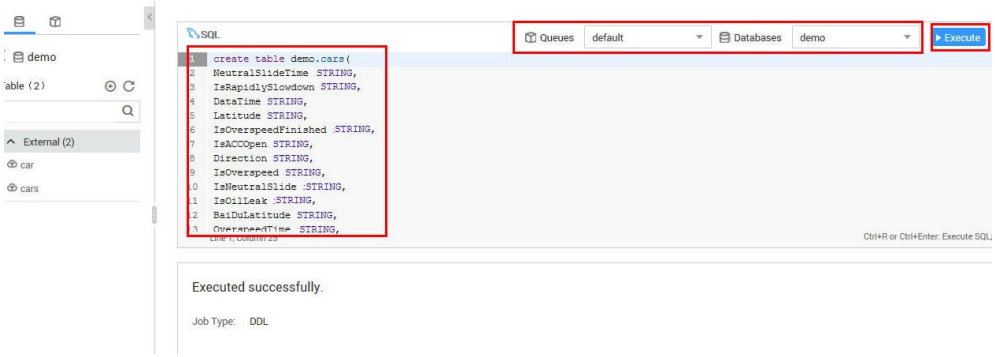
```
create table demo.cars(
  NeutralSlideTime STRING,
  IsRapidlySlowdown STRING,
  DataTime STRING,
  Latitude STRING,
  IsOverspeedFinished STRING,
  IsACCOpen STRING,
  Direction STRING,
  IsOverspeed STRING,
  IsNeutralSlide STRING,
  IsOilLeak STRING,
  BaiDuLatitude STRING,
  OverspeedTime STRING,
  IsRapidlySpeedup STRING,
  DeviceID STRING,
  Mileage STRING,
  Longitude STRING,
  Velocity STRING,
  IsNeutralSlideFinished STRING,
  IsFatgueDriving STRING,
  Carnum STRING,
  BaiDuLongitude STRING,
  BaiDuAdress STRING,
  IsHthrottleStop STRING,
  ReceiveTime STRING,
  Altitude STRING
) USING csv OPTIONS (path "obs://.....")
```

说明

请注意，将SQL语句中的“csv”修改为转储到OBS的文件格式，OBS路径修改为实际存放数据的OBS路径。

步骤2 单击“执行”，创建表，如图2-3所示。

图 2-3 创建表



表中的各字段含义请参见表2-4。

表 2-4 表字段含义

列名称(en)	数据类型	说明
DeviceID	string	设备ID
DataTime	string	数据时间
ReceiveTime	string	接收时间
IsACCOpen	string	ACC是否打开
Longitude	string	经度
Latitude	string	纬度
Velocity	string	速度
Direction	string	方向
Altitude	string	高度
Mileage	string	里程数
BaiDuLongitude	string	百度地图经度
BaiDuLatitude	string	百度地图纬度
BaiDuAdress	string	百度地图地址
Carnum	string	车牌号
IsRapidlySpeedup	string	急加速
IsRapidlySlowdown	string	急减速
IsNeutralSlide	string	空挡滑行
IsNeutralSlideFinished	string	空挡滑行结束
NeutralSlideTime	string	空挡滑行时长(s)

列名称(en)	数据类型	说明
IsOverspeed	string	超速
IsOverspeedFinished	string	超速结束
OverspeedTime	string	超速时长(s)
IsFatgueDriving	string	疲劳驾驶
IsHthrottleStop	string	停车轰油门

----结束

查询数据样例

- 数据查询统计

```
SELECT * FROM demo.cars
```

结果查询

查询语句执行结果如图2-4所示。

图 2-4 统计结果