Cloud Stream Service Best Practices

Cloud Stream Service Best Practices

Issue 01

Date 2020-05-21

HUAWEI TECHNOLOGIES CO., LTD.

Copyright © Huawei Technologies Co., Ltd. 2020. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: https://www.huawei.com

Email: support@huawei.com

Contents

 Copyright © Huawei Technologies Co., Ltd.

Cloud Stream Service Best Practices

1 Using Spark to Collect Statistics on Vehicle Startup Duration in Real Time

Scenario Overview

This practice helps you understand the basic functions of CS. In this practice, use the user-defined Spark job of CS to collect vehicle driving data that is ingested in real time and output information about each vehicle per startup, including the vehicle startup time, vehicle stop time, vehicle startup duration, and much more.

The original data is a simplified simulation of the real-time status information about vehicle access to the Internet of Vehicles (IoV). Generally, the bus data of a vehicle accessing to IoV includes information uploaded during driving, such as the longitude and latitude information, speed, direction, acc status, throttle position, and brake position. In this practice, the simulation data format is CSV, and only the vehicle ID, acc status, and data reporting time are used.

In this practice, the status management capability of the Spark Streaming is used to collect statistics on the duration of each vehicle startup based on the real-time access data of vehicles, and output the result. Compared with the solution of writing original data to disks and then conducting offline analysis, the solution adopting stream computing delivers more real-time performance and reduces data writing operations, thereby reducing costs.

In this practice, two DIS streams are used to separately serve as source and sink streams to simplify the process. In this practice, two DIS streams are used to separately serve as source and sink streams to simplify the process. CS supports source and sink streams of various types. For details, see the **Cloud Stream Service Stream Ecosystem Development Guide**.

The procedure is as follows:

1. **Prepare a Spark Sample Program**
2. **Prepare Spark Sample Data**
3. **Create a Cluster**
4. **Create a Job**
5. **View the Job Execution Results**

 Copyright © Huawei Technologies Co., Ltd.

Cloud Stream Service Best Practices

Prepare a Spark Sample Program

Step 1 Create an OBS bucket for storing the CS sample programs, logs, and output results.

1. Log in to the public cloud management console.
2. Click **Service List** and choose **Storage** > **Object Storage Service**.
3. Click **Create Bucket** to create a bucket named **cs-test**. Retain the default values for **Storage Class** and **Bucket Policy**.
4. Click **Create Now**.

Step 2 Download the sample program **SparkStreamingCountOnTime-assembly-1.0.jar** to your local PC.

1. Contact technical support to obtain the **cs-sparkstreaming-CarAccTime.zip** package.
2. Decompress the package in your directory.

Step 3 Enter the **cs-test** bucket, click **Upload File**. In the displayed dialog box, select the **SparkStreamingCountOnTime-assembly-1.0.jar** file in the **cs-sparkstreaming-CarAccTime** directory.

----End

Prepare Spark Sample Data

Step 1 Create a DIS stream for receiving vehicle data (that is, Spark sample data) in real time.

1. Log in to the public cloud management console.
2. Choose **Service List** > **EI Enterprise Intelligence** > **Data Ingestion Service**.
3. On the DIS console, click **Buy Stream**.
4. On the displayed **Buy Stream** page, configure the stream as follows:

   Region: **CN North-Beijing1**

   Stream Name: **sparkstreaming-input**

   Stream Type: **Common**

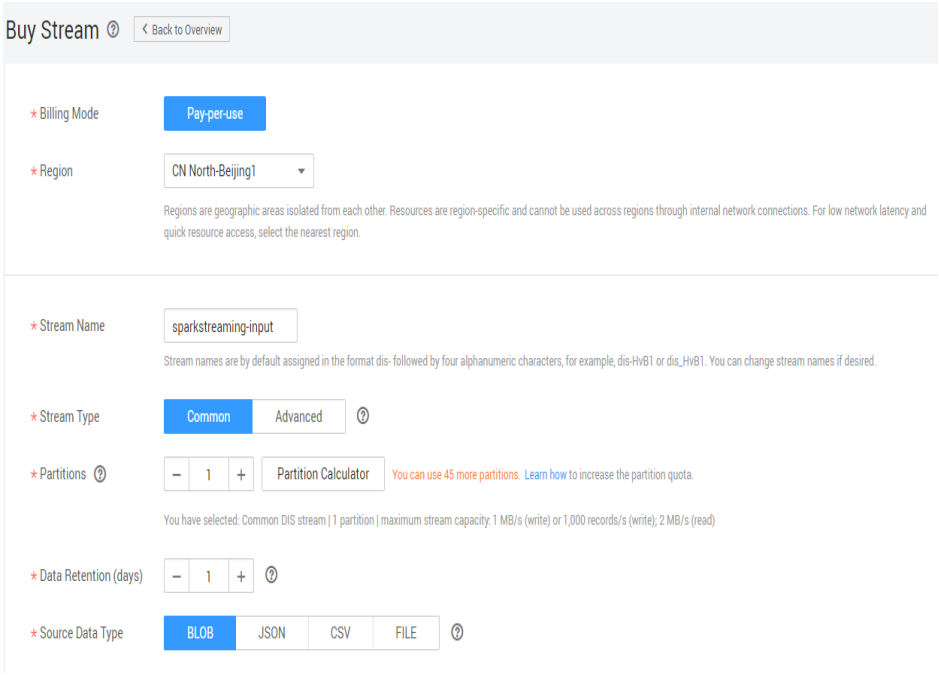
   Partitions: **1**

   Data Retention (days): **1**

   Source Data Type: **BLOB**

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-1 Creating a DIS stream for receiving real-time vehicle data



5. Click **Buy Now**.

Step 2 Repeat **Step 1** to create a DIS stream named **sparkstreaming-output** for receiving the CS job output data.

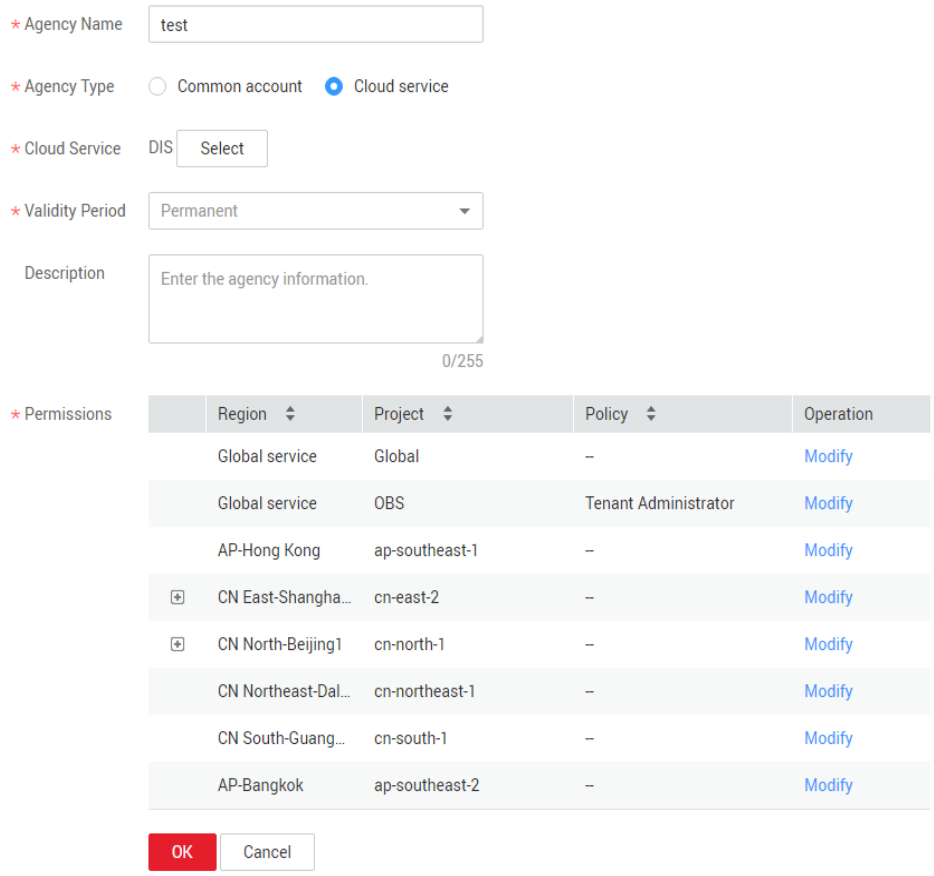
Step 3 Create a DIS agency on the IAM management console.

When you create a DIS stream and dump data to OBS, you need to create an IAM agency to authorize DIS to access your OBS resources.

1. Log in to the public cloud management console.
2. Choose **Service List** > **Management & Deployment** > **Identity and Access Management**.
3. Click **Agencies.** On the displayed page, click **Create Agency**.
4. Configure a DIS agency named **test** with the configurations shown in **Figure 1-2**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-2 Configuring DIS agency information



5. Click **OK**.

Step 4 Configure a data dump task for the DIS stream.

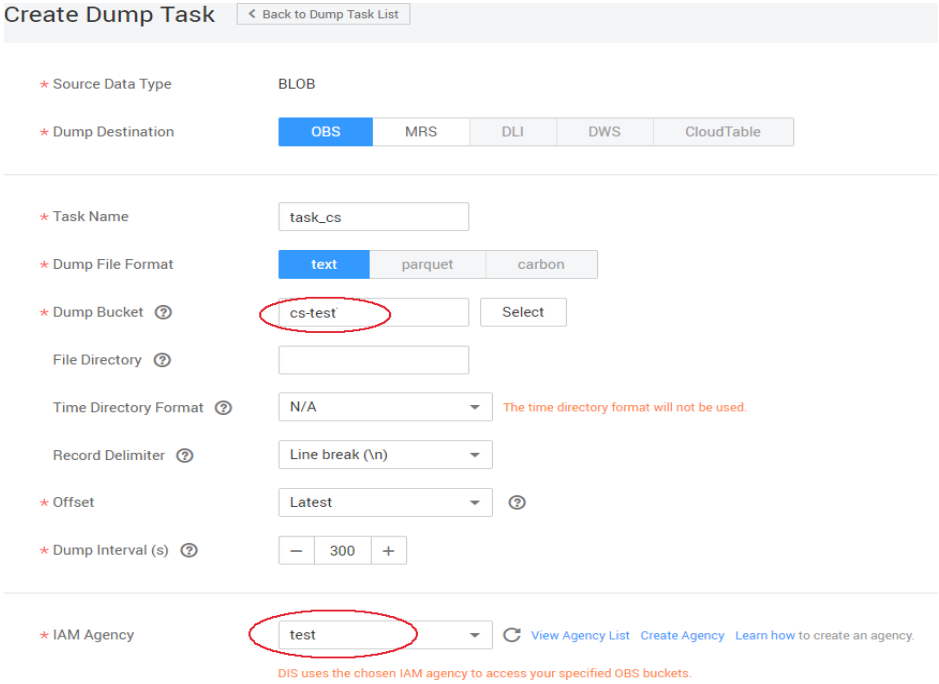
1. Log in to the public cloud management console.
2. Choose **Service List** > **EI Enterprise Intelligence** > **Data Ingestion Service**.
3. In the left navigation pane, click **Stream Management**. On the displayed page, click **sparkstreaming-output** in the **Name/ID** column of the stream list.
4. On the displayed page, click **Dump Management**, and then click **Add Dump Task**.
5. Configure the dump task by referring to the configurations shown in **Figure 1-3** and then click **Create Now**.

NOTE

- **Dump Destination**: Select the **cs-test** bucket created in **Step 1**.
- **IAM Agency**: Select the **test** agency created in **Step 3**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-3 Configuring the dump task



Step 5 Download the sample program **SparkStreamingCountOnTime-assembly-1.0.jar** to your local PC.

1. Contact technical support to obtain the **cs-sparkstreaming-CarAccTime.zip** package.
2. Decompress the package in your directory.

Step 6 Run the JAR file in the application package to generate the simulation data, and send the data to the DIS stream.

1. Obtain the **SparkStreamingCountOnTime-assembly-1.0.jar** file from the **cs-sparkstreaming-CarAccTime** directory.
2. Run the following command in the directory where the **SparkStreamingCountOnTime-assembly-1.0.jar** file is stored:

   java -cp SparkStreamingCountOnTime-assembly-1.0.jar util.GenerateData CarNumber ReportPeriod DisEndpoint AK SK ProjectId Region DisChannel

   The following table describes the parameters involved in the preceding command.

Table 1-1 Parameter configurations involved in the preceding command

Parameter	Configuration Method
CarNumber	Set this parameter to the total vehicle quantity contained in the simulation data. The value is an integer. The recommended value is **1000**.
ReportPeriod	Set this parameter to the interval (unit: s) for reporting data by each vehicle in the simulation data. The value is an integer. The recommended value is **10**.

 Copyright © Huawei Technologies Co., Ltd.

Cloud Stream Service Best Practices

Parameter	Configuration Method
DisEndpoint	Set this parameter to the endpoint address of DIS. You can obtain the parameter value from **Regions and Endpoints**. For example, **https://dis.cn-north-1.myhuaweicloud.com:443**.
AK/SK	Perform the following steps to obtain the parameter value: 1. Log in to the public cloud management console. 2. Move the cursor to the username in the upper right corner and select **My Credentials** from the drop-down list. 3. On the displayed **My Credentials** page, click **Access Keys**. 4. Click **Add Access Key**, enter the password and verification code as prompted, and click **OK**. 5. In the **Download Access Key** dialog box that is displayed, click **OK** to save the access keys to your browser's default download path. 6. Open the downloaded **credentials.csv** file to obtain the access keys (AK and SK).
ProjectId	Perform the following steps to obtain the parameter value: 1. Log in to the public cloud management console. 2. Move the cursor to the username in the upper right corner and select **My Credentials** from the drop-down list. 3. On the **Project List** page, obtain the project ID corresponding to the **CN North-Beijing1** region.
Region	Set this parameter to **cn-north-1**.
DisChannel	Set this parameter to **sparkstreaming-input**, which is the DIS stream used for receiving data.

3. If "*** data sent" is displayed on the CLI console, the data is successfully sent.

----End

Create a Cluster

Step 1 Log in to the public cloud management console.

Step 2 Click **Service List** and choose **EI Enterprise Intelligence** > **Cloud Stream Service** to access the CS console.

Copyright © Huawei Technologies Co., Ltd.

Cloud Stream Service Best Practices

Step 3 In the left navigation pane, click **Cluster Management** to switch to the **Cluster Management** page.

Step 4 Click **Create Cluster**. On the displayed **Create Cluster** page, configure basic cluster information by referring to **Table 1-2**.

Table 1-2 Parameters related to cluster configuration

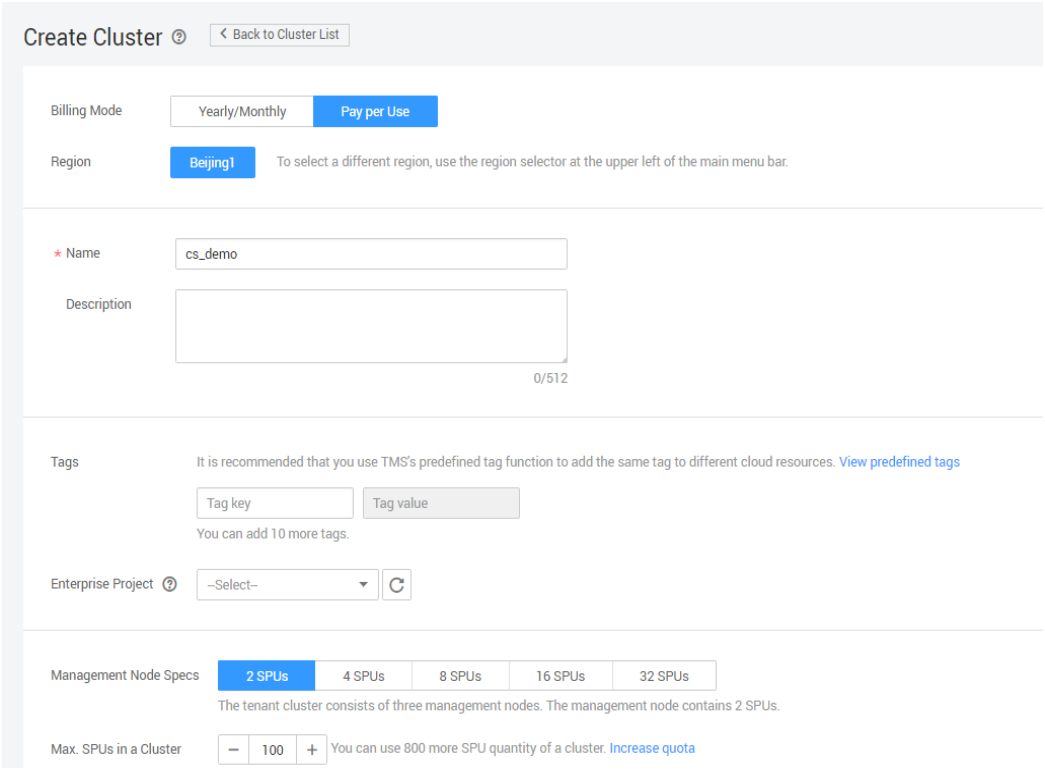
Parameter	Configuration Method
Billing Mode	The pay-per-use billing mode is used.
Region	Click  in the upper left corner of the management console and choose CN North-Beijing1.
Name	**cs_demo**
Description	Description of a cluster. This parameter can be left unspecified.
Tag	If you want to use the same tag to identify multiple cloud resources, you are advised to create predefined tags in the TMS. In this way, the same tag can be selected for all services. This parameter can be left unspecified.
Enterprise Project	For **Enterprise Project**, select the enterprise project that you have created on the Enterprise Management console. For details about how to create an enterprise project on the Enterprise Management console, see **Creating an Enterprise Project** in the *Enterprise Management User Guide*. If you do not select an enterprise project for the cluster, use the **default** project will be used.
Management Node Specs	Specifications of management nodes used by an exclusive cluster. The parameter value is positively correlated with the number of jobs running in the cluster. Select **2 SPUs**.
Max. SPUs in a Cluster	Maximum number of SPUs in a cluster for the purpose of dynamic capacity expansion. The default value is **100**. The parameter value can be changed after the cluster is created. Retain the default value.
Advanced Configuration	You can configure and adjust the VPC and subnet to which the cluster belongs based on the network plan. Retain the default values for related parameters.

The following figure shows an example of the basic cluster information.

Copyright © Huawei Technologies Co., Ltd.

Cloud Stream Service Best Practices

Figure 1-4 Creating a cluster



Step 5 Click **OK**. The system automatically switches to the **Cluster Management** page, where **Status** of the created cluster is **Requesting resources**.

It takes about 1 to 3 minutes to create a cluster. If the value of **Status** changes to **Running**, the cluster is successfully created.

----End

Create a Job

Step 1 In the navigation tree on the left pane of the CS management console, choose **Job Management** to switch to the **Job Management** page.

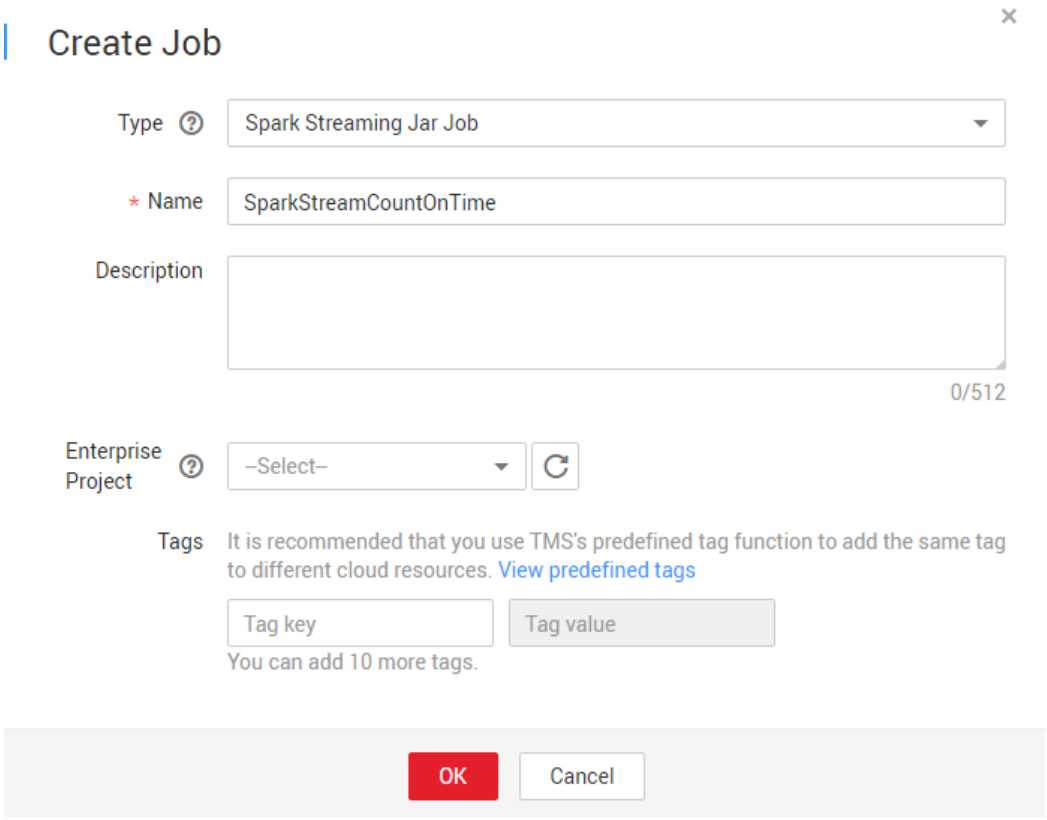
Step 2 Click **Create Job** to switch to the **Create Job** dialog box.

Step 3 Specify job parameters as required.

- Select **Spark Streaming Jar Job** For **Type**.
- Set **Name** to **SparkStreamingCountOnTime**.

Copyright © Huawei Technologies Co., Ltd.

Figure 1-5 Creating a job



Step 4 Click **OK** to enter the **Edit** page.

Step 5 Upload the JAR file.

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-6 Uploading the JAR file

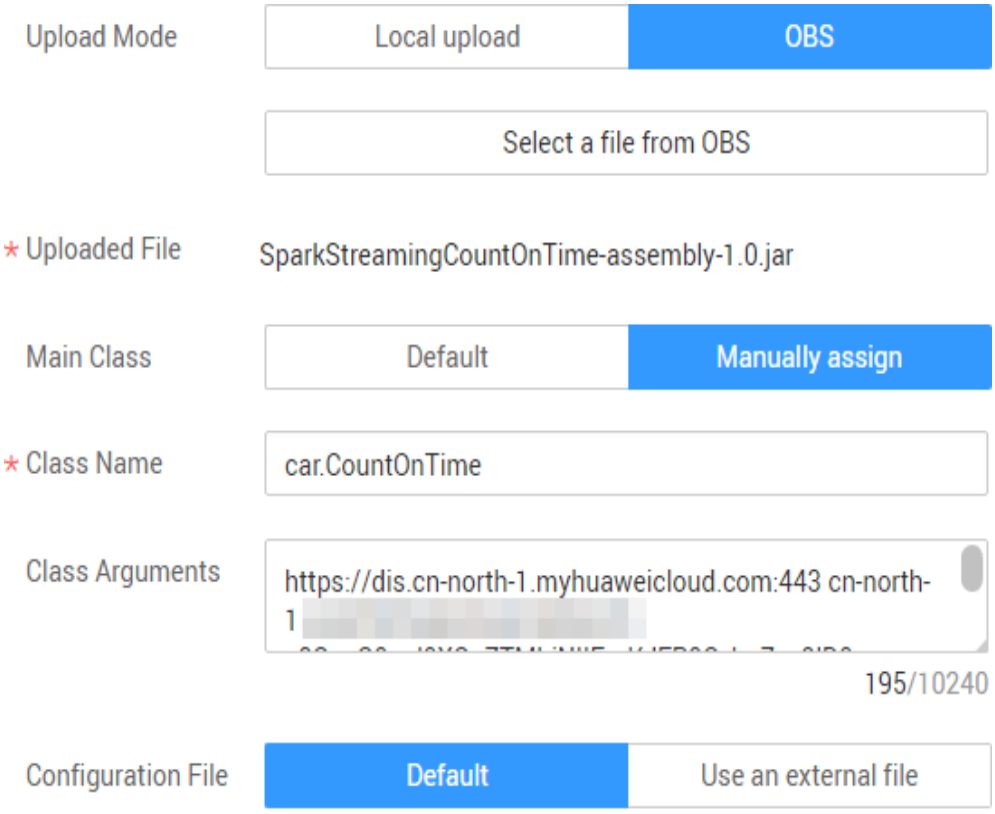


Table 1-3 Parameters for uploading a JAR file

Parameter	Configuration Method
Upload Mode	1. Select **OBS**. 2. Click **Select a File from OBS** and select the **SparkStreamingCountOnTime-assembly-1.0.jar** file that is uploaded to the **cs-test** bucket in **Prepare a Spark Sample Program**. 3. Click **OK**.

 Copyright © Huawei Technologies Co., Ltd.

Cloud Stream Service Best Practices

Parameter	Configuration Method
Main Class	1. Select **Manually assign**. 2. For **Class Name**, enter **car.CountOnTime**. 3. Enter the following parameters in the text box next to **Class Arguments**: *DisEndpoint Region AK SK ProjectId InputStream Duration OutputStream* **NOTE** – For details about how to set **DisEndpoint**, **Region**, **AK**, **SK**, and **ProjectId**, refer to **Table 1-1**. – **InputStream** refers to the input channel of the CS job. In this practice, set this parameter to **sparkstreaming-input**. – **OutputStream** refers to the output channel of the CS job. In this practice, set this parameter to **sparkstreaming-output**. – **Duration** refers to the batch interval (unit: second) of the Spark Streaming application. The value is an integer. In this practice, set this parameter to **1**.
Configuration File	You do not need to set this parameter because there are no available user-defined configuration files.
Cluster	Select the **cs_demo** cluster created in **Create a Cluster**.

Step 6 Click **Configure Parameters** on the left to configure job parameters.

Copyright © Huawei Technologies Co., Ltd.

Figure 1-7 Configuring job parameters

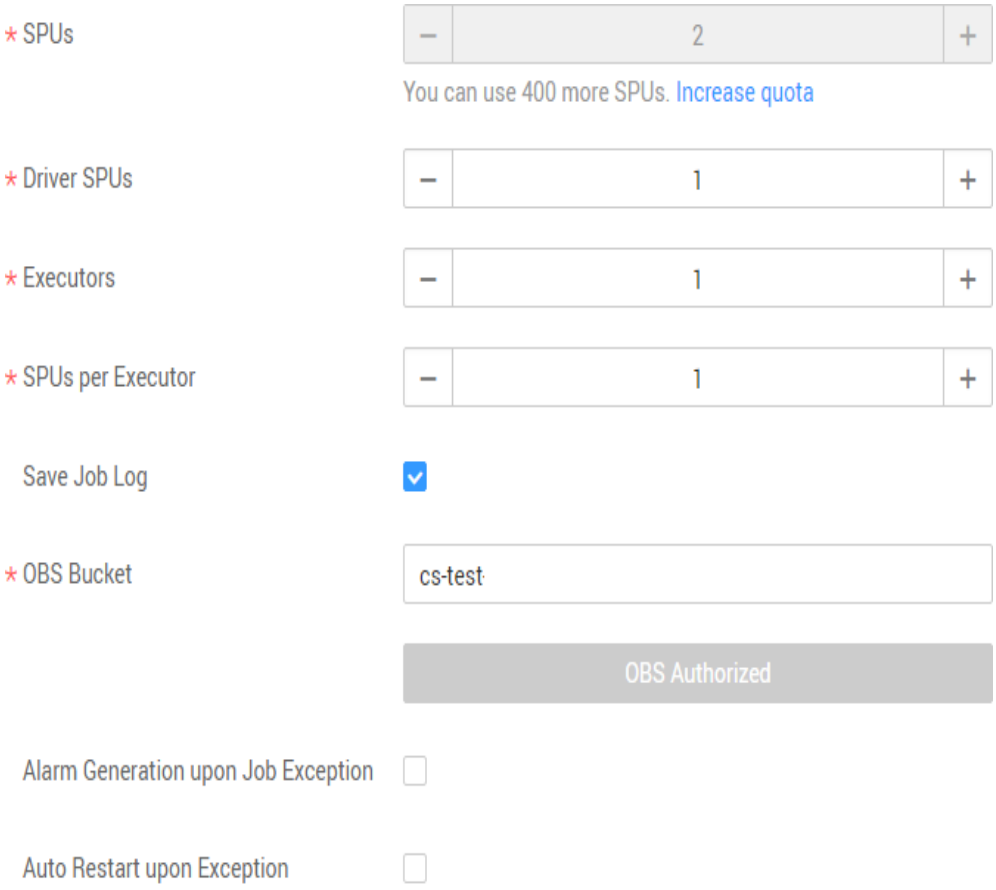


Table 1-4 Configuring job parameters

Parameter	Configuration Method
SPUs	This parameter indicates the total number of SPUs used by a job. The parameter value is automatically generated based on the values of parameters **Driver SPUs**, **Executors**, and **SPUs per Executor**.
Driver SPUs	This parameter indicates the number of SPUs used by the driver node. Use the default value **1**.
Executors	This parameter indicates the number of executor nodes. Use the default value **1**.
SPUs per Executor	This parameter indicates the number of SPUs per executor node. Use the default value **1**.

 Copyright © Huawei Technologies Co., Ltd.

Cloud Stream Service Best Practices

Parameter	Configuration Method
Save Job Log	This parameter indicates whether to save job logs. 1. Select this option to enable the job log saving function. 2. Click the text box next to **OBS Bucket**. In the dialog box that is displayed, select the **cs-test** bucket created in **Step 1**. 3. Click **Authorize OBS**.
Alarm Generation upon Job Exception	After you enable this function, CS sends related alarm information over SMSs or emails if a job fails or arrears occur. In this practice, do not select this option.
Auto Restart upon Exception	If you enable this function, CS automatically restarts and restores abnormal jobs upon job exceptions. In this practice, do not select this option.

Step 7 Click **Select the Target Cluster**. From the **Cluster** drop-down menu, select the **cs_demo** cluster created in **Create a Cluster**.

Figure 1-8 Selecting the cluster

Step 8 Click **Submit**. On the page that is displayed, click **OK**.

After the job is submitted, the system automatically switches to the **Job Management** page, and the created job is displayed in the job list. You can view the **Status** column to query the job status. After a job is successfully submitted, **Status** of the job will change from **Submitting** to **Running**.

----End

View the Job Execution Results

Step 1 In the navigation tree on the left pane of the CS management console, choose **Job Management** to switch to the **Job Management** page.

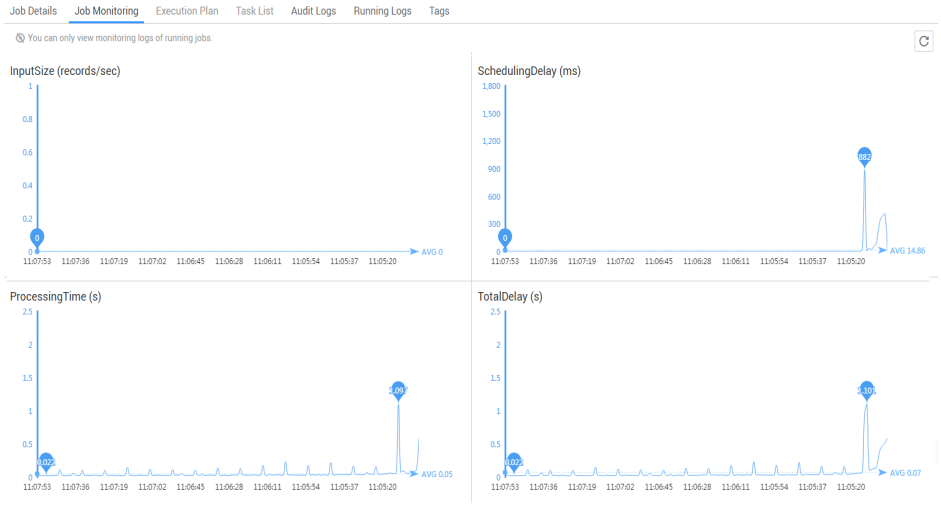
Step 2 In the **Name** column, click **SparkStreamingCountOnTime** to switch to the **Job Details** page.

Step 3 Click **Job Monitoring**. You can view information about the following four metrics: **InputSize**, **SchedulingDelay**, **ProcessingTime**, and **TotalDelay**.

You can also log in to the ECS connected to the cluster to view more monitoring information.

Copyright © Huawei Technologies Co., Ltd.

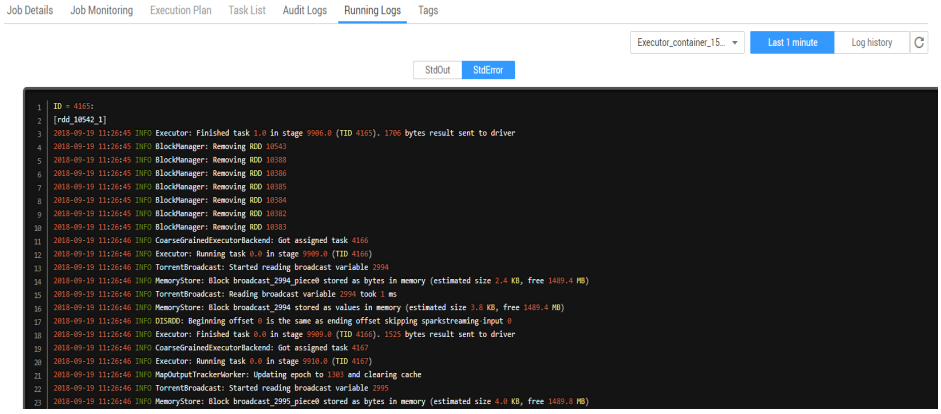
Figure 1-9 Checking the job monitoring information



Step 4 Click **Running Logs** to view job logs.

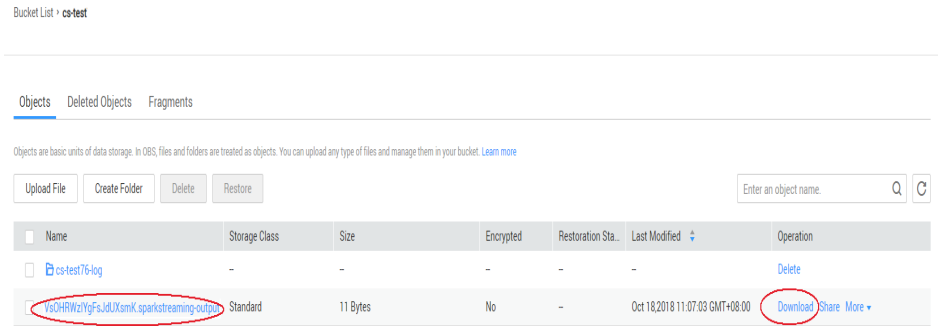
You can view the driver or executor logs by selecting the options in the drop-down list box in the upper right corner or clicking **StdOut** or **StdError**.

Figure 1-10 Viewing job logs



Step 5 After a dump interval elapses (the dump interval is set in **Step 4** and the default value is 300 seconds), log in to the OBS management console. Click **cs-test** in the **Bucket Name** column. Click **Objects** from the left navigation pane. On the displayed page, locate the row where the dump file for the **sparkstreaming-output** stream resides and click **Download**.

Figure 1-11 Exporting the dump file



 Copyright © Huawei Technologies Co., Ltd.

Cloud Stream Service Best Practices

Step 6 After the file is downloaded, open it. You can view the statistics about vehicle startup durations collected by CS jobs.

The file is in the CSV format. As shown in the following figure, the fields, from left to right, indicate the vehicle ID, startup time, stop time, and startup duration separately.

Figure 1-12 Vehicle statistics about vehicle startup durations

```
848,1537327813015,1537327823020,10005
879,1537327813516,1537327823520,10004
311,1537327808015,1537327818017,10002
522,1537327810014,1537327820018,10004
72,1537327815516,1537327825521,10005
727,1537327822019,1537327832022,10003
654,1537327811515,1537327831522,20007
698,1537327811515,1537327831523,20008
897,1537327823520,1537327833523,10003
151,1537327826522,1537327836522,10000
646,1537327811014,1537327841023,30009
710,1537327812015,1537327842024,30009
609,1537327831022,1537327841023,10001
806,1537327833022,1537327843024,10002
893,1537327833523,1537327843525,10002
427,1537327809014,1537327849026,40012
488,1537327839524,1537327849527,10003
```

----End

Copyright © Huawei Technologies Co., Ltd.

2 Using Flink SQL to Analyze Logs in Real Time

Scenario Overview

In this practice, data is read from DIS. Create a Flink SQL job on the CS console to analyze logs in real time, and then output the result to OBS.

The procedure is as follows:

1. **Creating a DIS Stream and an OBS Bucket**
2. **Creating a Flink SQL Job**
3. **Sending DIS Data and Viewing the Result**

Creating a DIS Stream and an OBS Bucket

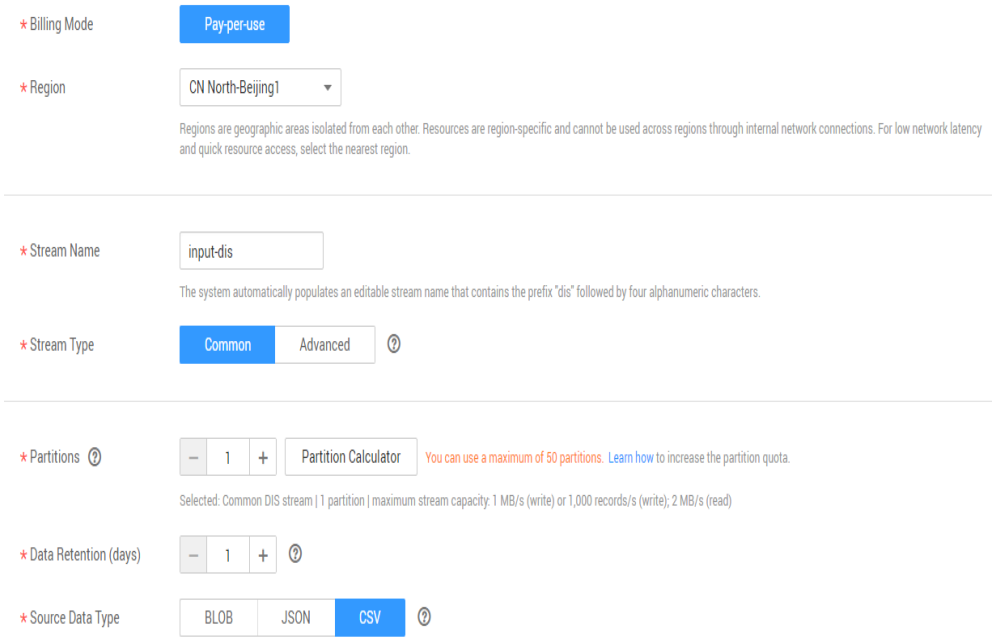
In this practice, data is read from DIS and then written to OBS after stream processing. Therefore, you need to create a DIS stream and OBS bucket first.

Step 1 Create a DIS stream.

1. Log in to the management console.
2. Click **Service List** and choose **EI Enterprise Intelligence** > **Data Ingestion Service** to access the DIS console.
3. On the DIS console, click **Buy Stream**.
   - **Region**: Set the same region for DIS, OBS, and CS. **CN North-Beijing 1** is used as an example.
   - **Stream Name**: **input-dis**
   - **Source Data Type**: **CSV**
   - Retain the default values for other parameters.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-1 Creating a DIS stream

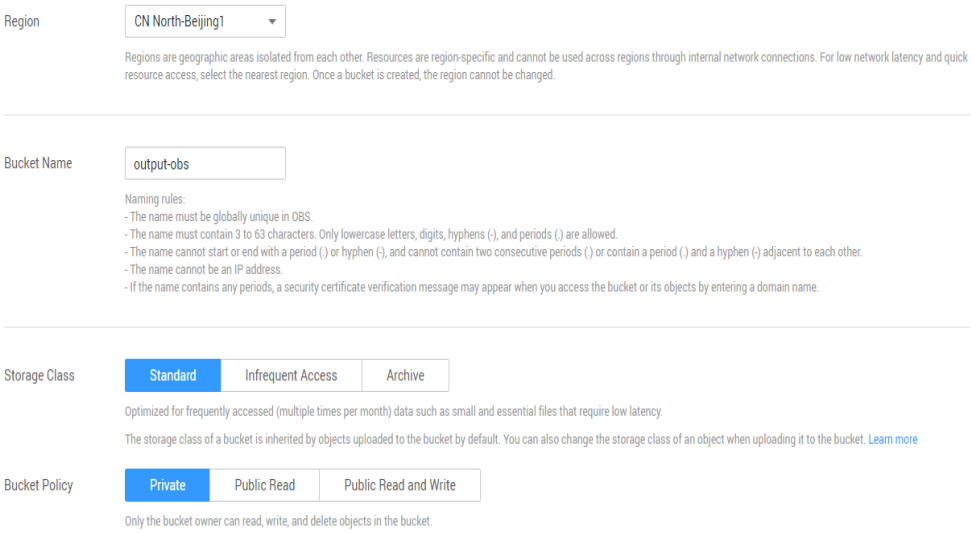


4. Click **Next**, confirm stream specifications, and click **Submit**.
5. Click **Back to Stream Management**. The created stream **input-dis** is displayed. This stream is used as the source stream of CS.

Step 2 Create an OBS bucket.

1. Click **Service List** and choose **Storage** > **Object Storage Service**.
2. Click **Create Bucket**. On the displayed page, configure related parameters.
   - **Region**: Select **CN North-Beijing1**.
   - **Bucket Name**: **output-obs**
   - Retain the default values for other parameters.

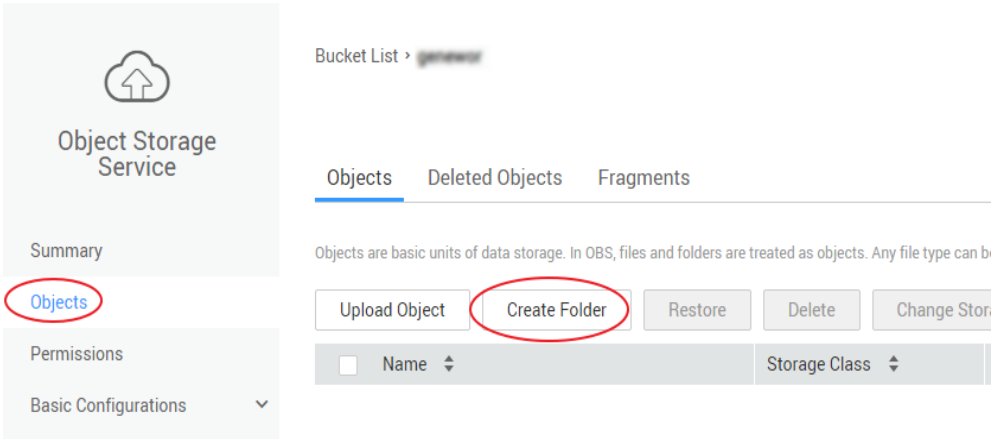
Figure 2-2 Creating an OBS bucket



 Copyright © Huawei Technologies Co., Ltd.

3. Click **Create Now**. The new bucket is displayed in the bucket list.
4. Click the bucket name **output-obs**. In the navigation pane, choose **Objects**. Click **Create Folder** and enter **logInfos** in the **Folder Name** field. The **logInfos** folder is used to store the output.

Figure 2-3 Creating a folder



----End

Creating a Flink SQL Job

Step 1 Click **Service List** and choose **EI Enterprise Intelligence** > **Cloud Stream Service** to access the CS console.

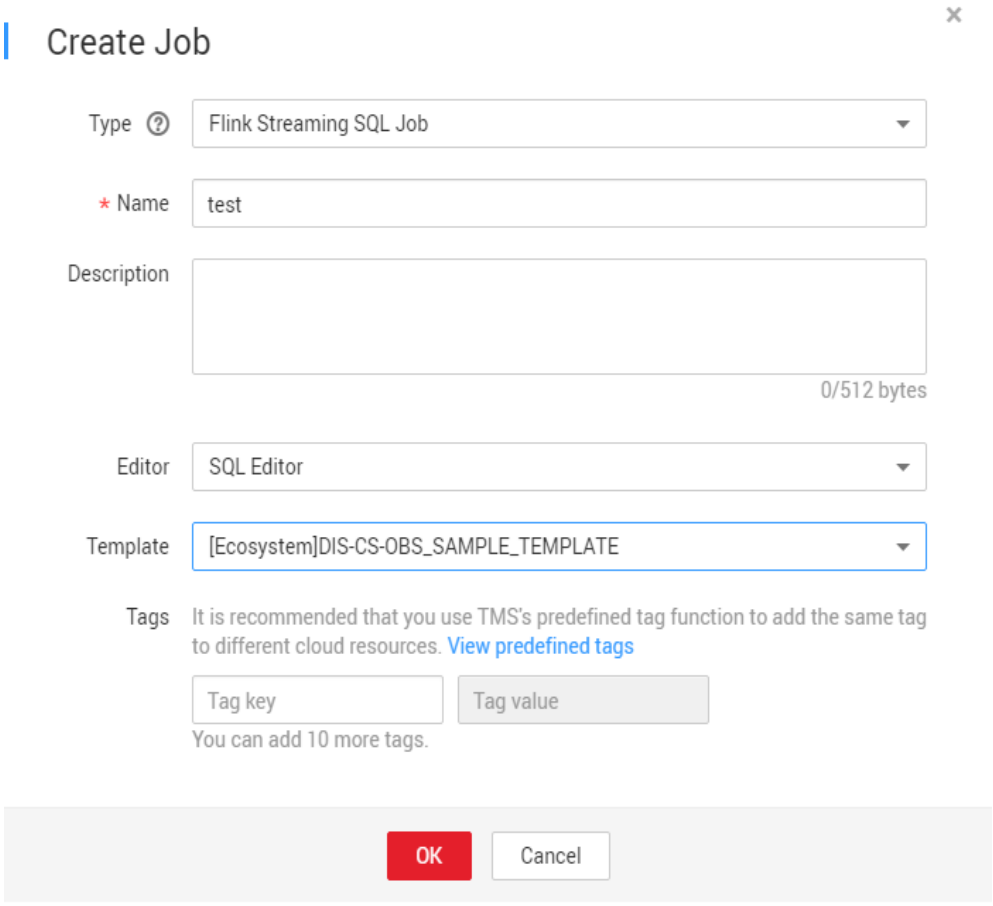
If you log in to the CS console for the first time, apply for CS and authorize CS as prompted.

Step 2 Click **Create Job**. In the displayed **Create Job** dialog box, specify job information.

- **Type**: Select **Flink Streaming SQL Job**.
- **Name**: **test**
- **Template**: Select **[Ecosystem]DIS-CS-OBS_SAMPLE_TEMPLATE**.
- Retain the default values for other parameters.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-4 Creating a Flink SQL job



Step 3 Click **OK**. On the displayed page, edit the job.

[Ecosystem]DIS-CS-OBS_SAMPLE_TEMPLATE is used as an example.

The SQL editor contains the following parts:

- Source stream: Configure it in the **with** statement to interconnect with the **input-dis** stream of DIS so that CS can obtain data from the **input-dis** stream in real time. Set the following parameters:
  - type = "dis"
  - region = "cn-north-1". **cn-north-1** indicates **North China-Beijing 1**.
  - channel = "input-dis" (DIS stream name)
  - partition_count = "1" (number of partitions of the DIS stream)
  - encode = "csv" (data encoding format)
  - field_delimiter = "\\|\\|" (delimiter between attributes if the encoding format is CSV)
  - quote = "\u005c\u0022" (quoted symbol in a data format). The attribute delimiters between two quoted symbols are treated as common characters. If double quotation marks are used as the quoted symbol, set this parameter to **\u005c\u0022** for character conversion.
  - offset = "0". This indicates that CS starts to process data from data record 0 in DIS.

 Copyright © Huawei Technologies Co., Ltd.

For details, see **DIS Source Stream**.

- Sink stream: Configure it in the **with** statement to interconnect with the OBS bucket so that CS can output the result to the OBS bucket. Set the following parameters:
  - type = "obs"
  - region = "cn-north-1". **cn-north-1** indicates **North China-Beijing 1**.
  - encode = "csv" (data encoding format)
  - field_delimiter = "\\|\\|" (delimiter between attributes if the encoding format is CSV)
  - row_delimiter = "\n" (row separator)
  - obs_dir = "output-obs/logInfos" (file storage directory). The format is **{bucket name}/{folder name}**.
  - file_prefix = "log_out" (prefix of the exported file name). The default value is **temp**.
  - rolling_size = "100m" (maximum size of a file)

  For details, see **OBS Sink Stream**.
- The following are SQL querying examples:

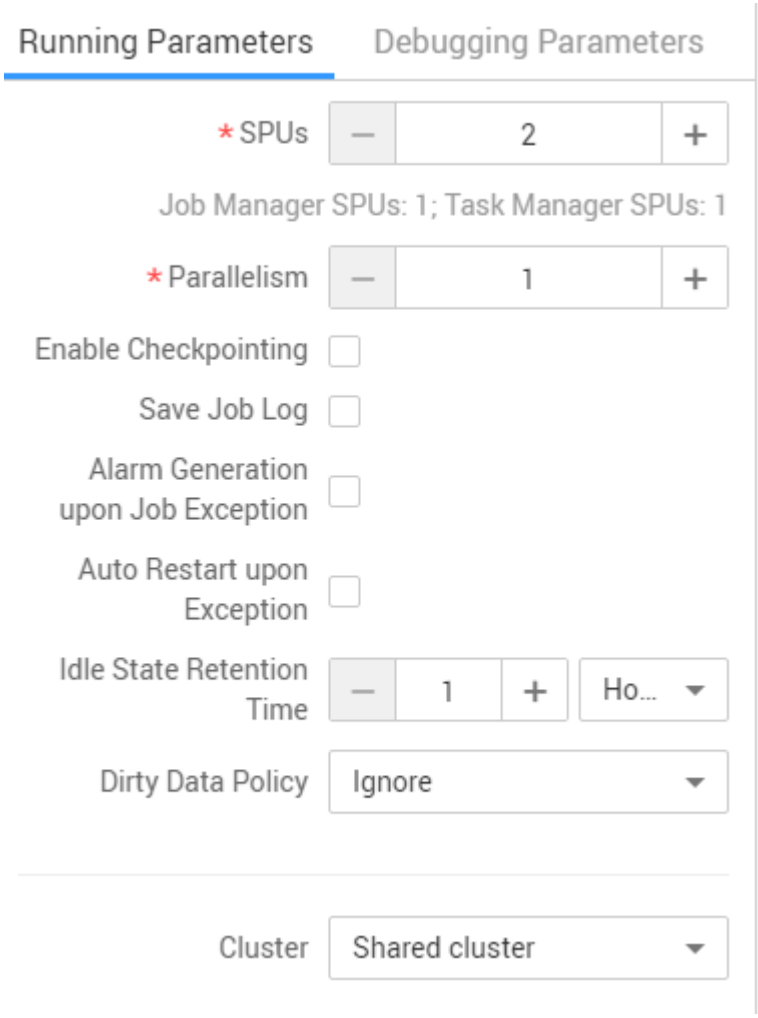
```
INSERT INTO log_out
SELECT http_host,forward_ip,cast(cast(msec * 1000 as bigint) + 28800000  as
timestamp),status,request_length, bytes_sent,string_to_array(request, '\\ ')[1],string_to_array(request,
'\\ ')[2],http_referer,http_user_agent,
upstream_cache_status,upstream_status,request_time,cookie_DedeUserID_cookie_sid_sent_http_logdat
a,upstream_response_time,
upstream_addr,
case IP_TO_PROVINCE(forward_ip) when "Guangxi" then "Guangxi Zhuang Autonomous Region"
when "Ningxia" then "Ningxia Hui Autonomous Region"
when "Taiwan" then "Taiwan Province"
when "Macao" then "Macau"
else IP_TO_PROVINCE(forward_ip) end,
case when http_user_agent like "%Chrome%" then "Chrome"
when http_user_agent like "%Firefox%" then "Firefox"
when http_user_agent like "%Safari%" then "Safari"
else "Others" end
FROM log_infos;
```

Step 4 Set job running parameters.

- **SPUs**: One Stream Processing Unit (SPU) is a unit of stream processing capacity comprised of 1 vCPU compute and 4 GB memory. One SPU costs ¥0.5 per hour. At least two SPUs are required.
- **Parallelism**: number of concurrent operators in a Flink job. The default value is **1**.
- **Enable Checkpointing**: whether to enable the Flink snapshot
- **Save Job Log**: whether to save job logs to the OBS bucket
- **Alarm Generation upon Job Exception**: whether to send an email and SMS message after a job exception occurs

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-5 Setting job running parameters



Step 5 Click **Check Semantics**. You can perform **Debug**, **Submit**, and **Start** operations on a job only after semantics check succeeds.

Step 6 Click **Submit**. Review job configurations and then click **OK**.

The system automatically switches to the **Job Management** page, and the created job is displayed in the job list. The **Status** column displays the job status. If a job is successfully submitted, the job status will change to **Running**.

----End

Sending DIS Data and Viewing the Result

The DIS Agent is used to upload CSV data to the DIS stream. Similar to Flume, the DIS Agent is a local agent that monitors local file changes. Once new data is added to a file, the data is immediately uploaded to the DIS stream.

For details about how to use DIS Agent, see **Uploading Data by Using Agent**.

Step 1 Start DIS Agent.

1. Download DIS Agent from **https://dis-publish.obs-website.cn-north-1.myhuaweicloud.com/dis-agent-1.1.0.zip**.

 Copyright © Huawei Technologies Co., Ltd.

2. Decompress the downloaded DIS Agent package.
3. Modify the **conf/agent.yml** file.

```
---
# Keep unchanged.
region: cn-north-1
# user ak (get from 'My Credentials')
ak: Enter your AK.
# user sk (get from 'My Credentials')
sk: Enter your SK.
ak/sk: Log in to the management console, hover the cursor on the username in the upper right corner, and choose My Credentials > Access Keys > Create Access Key.
# user project id (get from 'My Credentials')
projectId: Log in to the management console, hover the cursor on the username in the upper right corner, choose My Credentials > Projects, locate the row containing cn-north-1, and use the value of Project ID.
# Keep unchanged.
endpoint: https://dis.cn-north-1.myhuaweicloud.com:20004
# config each flow to monitor file.
flows:
# Enter the name of the DIS stream you created.
- DISStream: input-dis
# Enter the directory for storing data files.
filePattern: D:/disagent-cw/dis-agent-1.1.0/data/*.log
# Keep unchanged.
```

4. Start DIS Agent.
   - Linux operating system: **bin/start-dis-agent.sh**
   - Windows operating system: **bin/start-dis-agent.bat**

Step 2 Send DIS data.

Place your data file in the file path configured in **agent.yml**. If you choose to add data to the file by writing a program, move the program to the file path configured in **agent.yml**.

```
import time

for idx in range(10000):
    with open("test.log", mode = "a+") as f:
        f.write("api.huaweicloud.com||45.249.212.44||15421010072.675||200||651||228||POST /x/report/heartbeat HTTP/1.1||-||Mozilla/5.0 (Windows NT 6.0; rv:34.0) Gecko/20100101 Firefox/34.0||-||200||0.033||-.918nw0fj-||0.033||140.206.227.10:80" + "\n" + "api.huaweicloud.com||45.249.212.52||15421010072.875||200||651||228||POST /details/jobs HTTP/1.1||-||Mozilla/5.0 (Windows NT 6.0; rv:34.0) Gecko/20100101 Firefox/34.0||-||200||0.033||-.918nw0fj-||0.033||140.206.227.10:80" + "\n")
    time.sleep(60)
```

Step 3 Log in to the OBS console, go to the **logInfos** folder of the **output-obs** bucket, click **Download**, and view the output.

----End