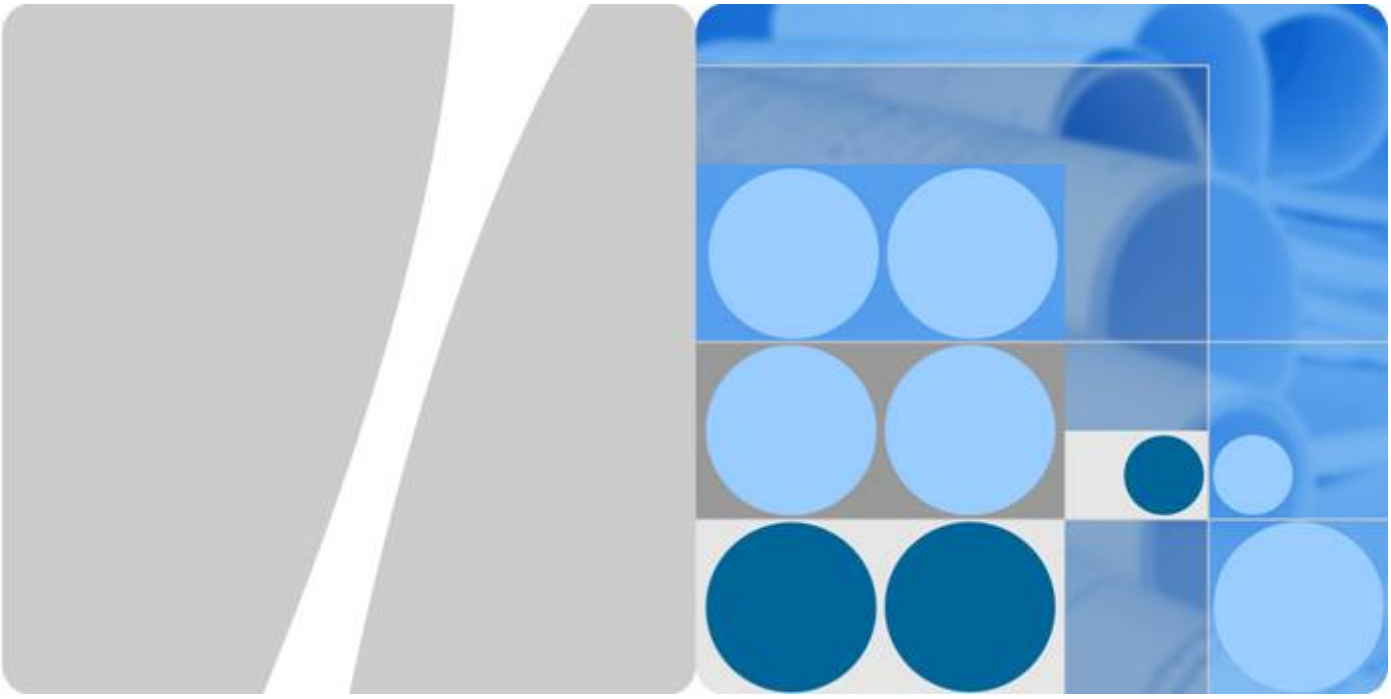


Cloud Stream Service

Quick Start

Issue 19

Date 2018-11-30

HUAWEI TECHNOLOGIES CO., LTD.

HUAWEI

Copyright © Huawei Technologies Co., Ltd. 2019. 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.

Contents

1 Getting Started.....1

 Copyright © Huawei Technologies Co., Ltd.

1 Getting Started

This section describes how to use CS. The general procedure for using CS is as follows:

Step 1: Prepare Data Sources and Data Output Channels

Step 2: Create an OBS Bucket for Saving Outputs

Step 3: Apply for CS

Step 4: Create a Job and Submit It

Step 5: Send Data to DIS

Step 6: View Job Information and Job Execution Result

Step 7: View Related Documents

Step 8: Delete the Job

This document uses a Flink streaming SQL job as an example to help you understand how to use CS. For details about Flink streaming SQL edge jobs, user-defined Flink jobs, and user-defined Spark jobs, see **Job Management** in the **Cloud Stream Service User Guide**.

In the example, vehicle data are recorded in real time and the data for Audi vehicles whose prices are lower than ¥300,000 are exported.

You need to create a job that has one source stream and one sink stream. The source stream is used to record vehicle information in real time. The sink stream is used to export information about the Audi vehicles whose prices are lower than ¥300,000.

Prerequisites

You have registered a CS account with the public cloud.

Step 1: Prepare Data Sources and Data Output Channels

CS supports other services as data sources and data output channels. For details, see **Preparing Data** in the *Cloud Stream Service User Guide*.

In this example, DIS serves as the data source and data output channel. You need to deploy DIS for **JobSample**.

For example, create the following DIS streams for **JobSample**. For details, see **Creating a DIS Stream** in the **Data Ingestion Service User Guide**.

 Copyright © Huawei Technologies Co., Ltd.

- DIS stream as the source stream

  Stream Name: **csinput**

  Stream Type: **Common**

  Partitions: **1**

  Source Data Type: **BLOB**

  Data Retention Period (days): **1**

  Data Dumping: **Off**

- DIS stream as the sink stream

  Stream Name: **csoutput**

  Stream Type: **Common**

  Partitions: **1**

  Source Data Type: **BLOB**

  Data Retention Period (days): **1**

  Data Dumping: **Off**

NOTE

Retain the default values for **Stream Type**, **Source Data Type**, and **Data Retention Period (days)**.

Step 2: Create an OBS Bucket for Saving Outputs

In this example, you need to enable OBS for **JobSample** to provide CS with the functions of checkpoint, saving job logs, and debugging data.

For example, create the following OBS bucket for **JobSample**. For details, see **Creating a Bucket** in the **Object Storage Service Console Operation Guide**.

Information about the created OBS bucket is as follows:

Region: **CN North-Beijing1**

Bucket Name: **smoke-test**

Storage Class: **Standard**

Bucket Policy: **Private**

NOTE

Retain the default settings for other parameters.

Step 3: Apply for CS

You can log in to the CS management console through a browser and apply for CS.

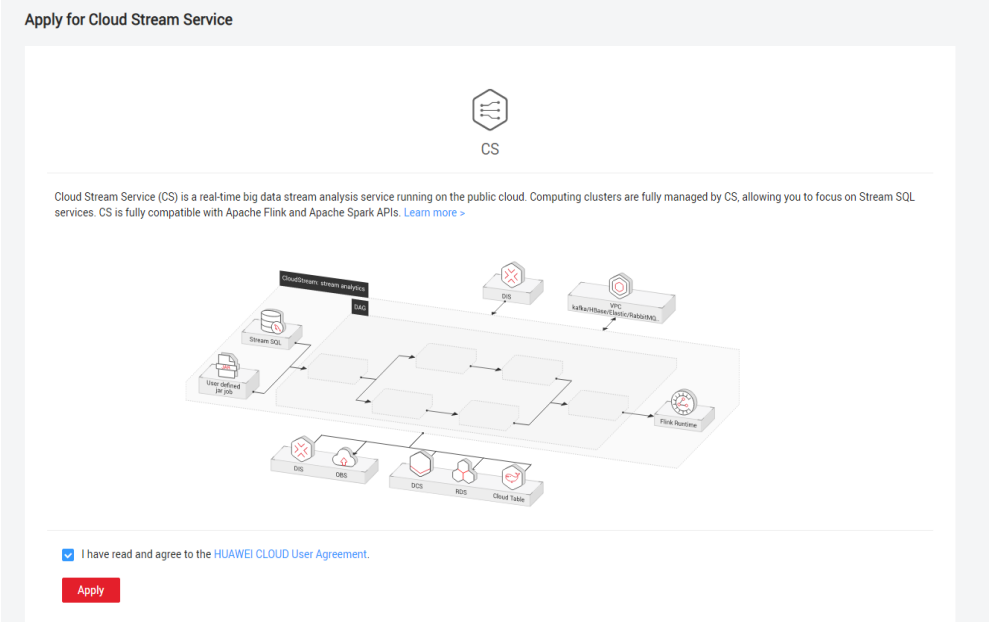
Step 1 Log in to the **CS management console**.

If you have not registered with the public cloud, click **Free Registration** to register an account with the public cloud as prompted.

Step 2 The **Apply for Cloud Stream Service** page is displayed.

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-1 Applying for CS

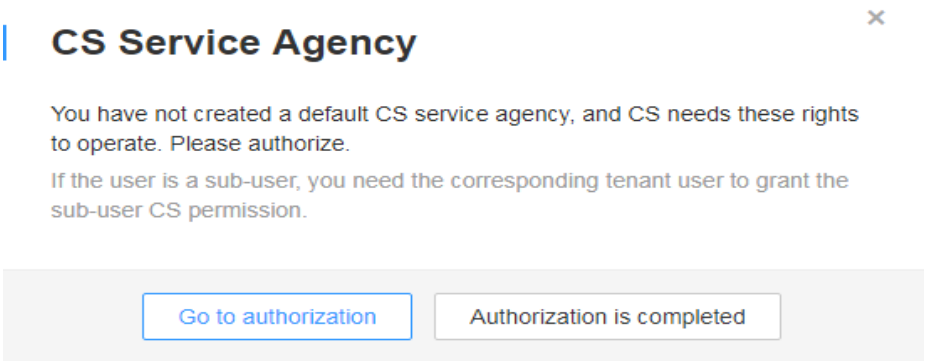


Step 3 Select **I have read and agree to the HUAWEI CLOUD User Agreement** and click **Apply**.

Step 4 After the application completes, the system redirects you the **Overview** page.

Step 5 In the **CS Service Agency** window that is automatically displayed, click **Go to authorization**.

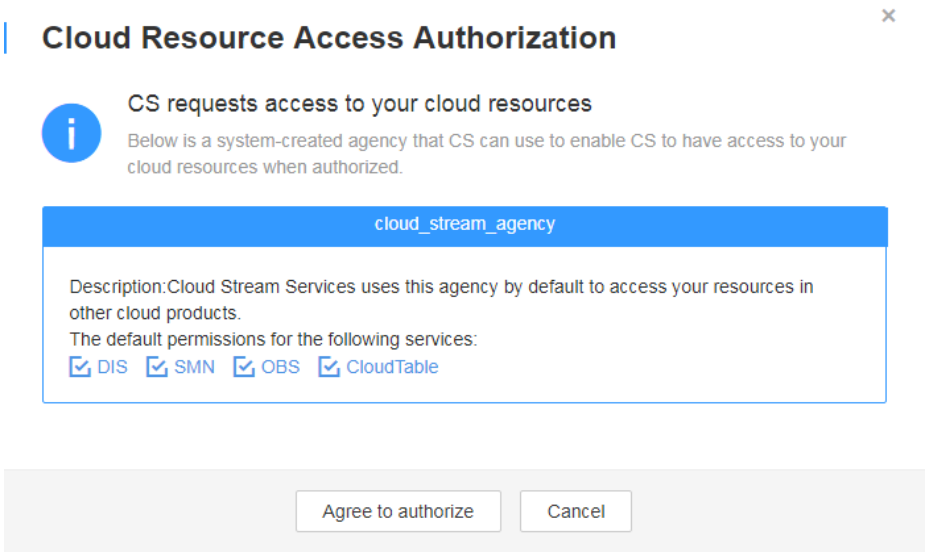
Figure 1-2 Creating an agency



Step 6 On the **Cloud Resource Access Authorization** page that is displayed, click **Agree to authorize**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-3 Cloud resource access authorization



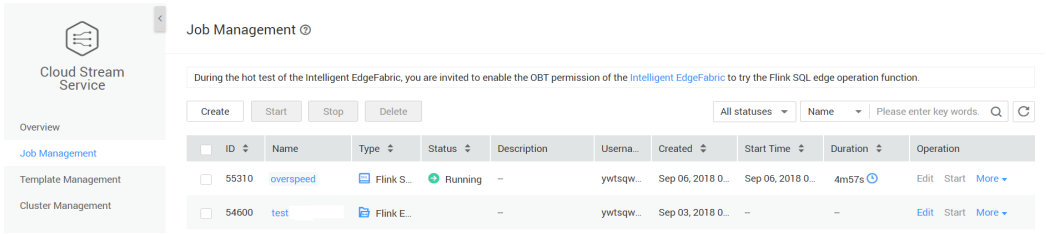
----End

Step 4: Create a Job and Submit It

To use CS, you need to create a job first, for example, **JobSample**.

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

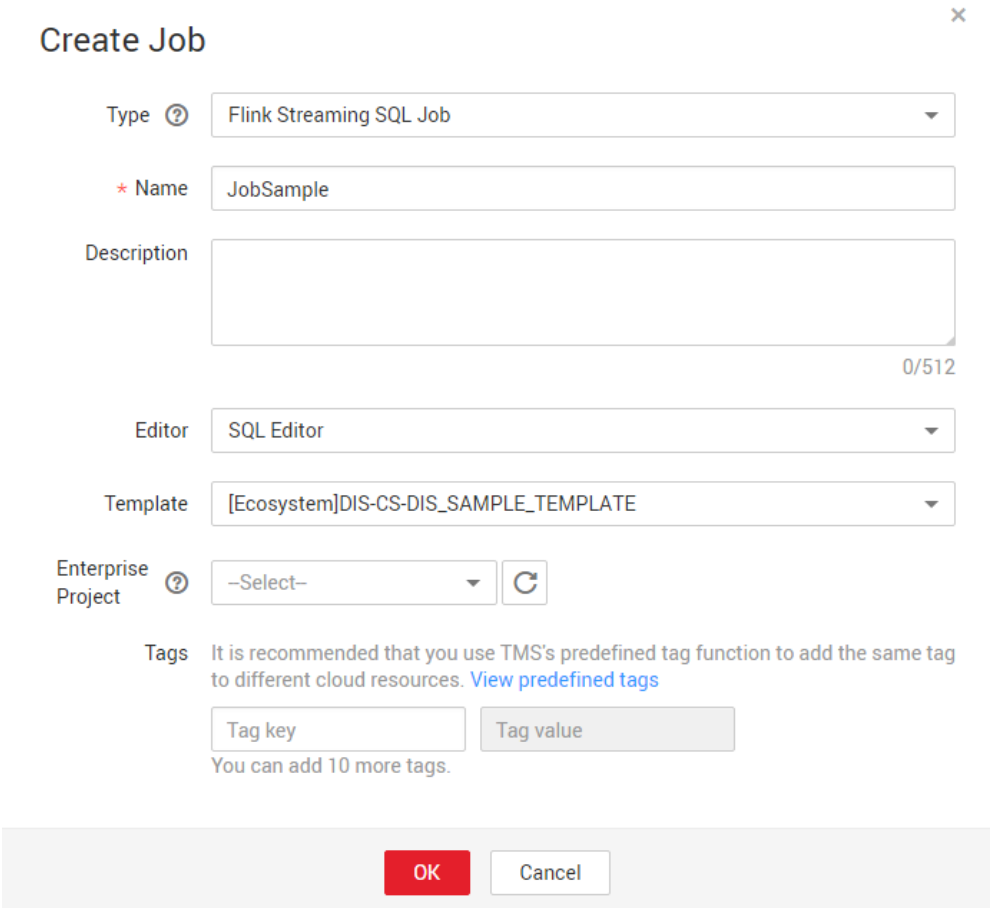
Figure 1-4 Job Management



Step 2 On the **Job Management** page, click **Create Job**. In the displayed **Create Job** dialog box, specify job details.

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-5 Creating a job



Step 3 Click **OK** to switch to the **Edit** page, as shown in the following figure. In the SQL statement editing area, SQL statements in the template selected in **Step 2** are displayed.

Figure 1-6 Editing a job



Step 4 Edit SQL statements as required.

This example uses the SQL statements from the template selected in **Step 2**. The SQL statement details are as follows:

```
/**
* This example uses a general stream analysis template. DIS serves as both the source stream and sink stream. You need to enable DIS and create source and sink streams.
  * >>>>>>>>> Ensure that you have created desired DIS streams with your account.
```

 Copyright © Huawei Technologies Co., Ltd.

```
<<<<<<<<<<
  *
  * >>>>>Sample input<<<<<
  *  Stream name: car_infos(car_id,car_owner,car_brand,car_price):
  *  1,lilei,bmw320i,28
  *  2,hanmeimei,audia4,27
  * >>>>>Sample output<<<<<
  *  Stream name: audi_cheaper_than_30w(car_id,car_owner,car_brand,car_price):
  *  2,hanmeimei,audia4,27
  **/

/** Obtain data from DIS stream csinput and create a source stream.
  *
  * Reconfigure the following options according to actual conditions:
  * channel: Indicates the name of the stream where data is located.
  * partition_count: Indicates the number of partitions of the stream.
  * encode: Indicates the data encoding format, which can be csv or json.
  * field_delimiter: Indicates the separator used to separate every two
attributes when the CSV encoding format is adopted.
  **/
CREATE SOURCE STREAM car_infos (
  car_id STRING,
  car_owner STRING,
  car_brand STRING,
  car_price INT
)
WITH (
  type = "dis",
  region = "cn-north-1" ,
  channel = "csinput",
  partition_count = "1",
  encode = "csv",
  field_delimiter = ","
);

/** Create a sink stream and export the data to DIS stream csoutput.
  *
  * Reconfigure the following options according to actual conditions:
  * channel: Indicates the name of the stream where data is located.
  * partition_key: Indicates the primary key used for data provisioning when
there are multiple partitions in the stream.
  * encode: Indicates the result encoding format, which can be csv or json.
  * field_delimiter: Indicates the separator used to separate every two
attributes when the CSV encoding format is adopted.
  **/
CREATE SINK STREAM audi_cheaper_than_30w (
  car_id STRING,
  car_owner STRING,
  car_brand STRING,
  car_price INT
)
WITH (
  type = "dis",
  channel = "csoutput",
  region = "cn-north-1" ,
  partition_key = "car_owner",
  encode = "csv",
  field_delimiter = ","
);

/** Output information about the Audi vehicles whose prices are lower than
¥300,000. **/
INSERT INTO audi_cheaper_than_30w
SELECT *
FROM car_infos
WHERE car_brand like "audi%" and car_price < 30;

/************************Insert test data into source stream
csinput************************/
```

 Copyright © Huawei Technologies Co., Ltd.

```
CREATE SINK STREAM car_info_data (
  car_id STRING,
  car_owner STRING,
  car_brand STRING,
  car_price INT
)
WITH (
  type ="dis",
  region = "cn-north-1" ,
  channel = "csinput",
  partition_key = "car_owner",
  encode = "csv",
  field_delimiter = ",\n"
);

INSERT INTO car_info_data
SELECT "1", "lilei", "bmw320i", 28;
INSERT INTO car_info_data
SELECT "2", "hanmeimei", "audia4", 27;
/********************** Stop the process of inserting test data to the source
stream. ******************************/
```

NOTE

The SQL statement consists of three parts:

- Creating the source stream: Ensure that fields defined for the stream are in the same format as the data source of the sink stream. Parameters in the **WITH** field defined for the stream must be the same as those of the source DIS stream.
- Creating a sink stream: Ensure that parameters in the **WITH** field defined for the stream are same as those of the sink DIS stream.
- Business logic: Compile SQL statements according to the scenario and insert the result data into the output stream.

Step 5 Click **Check Semantics**.

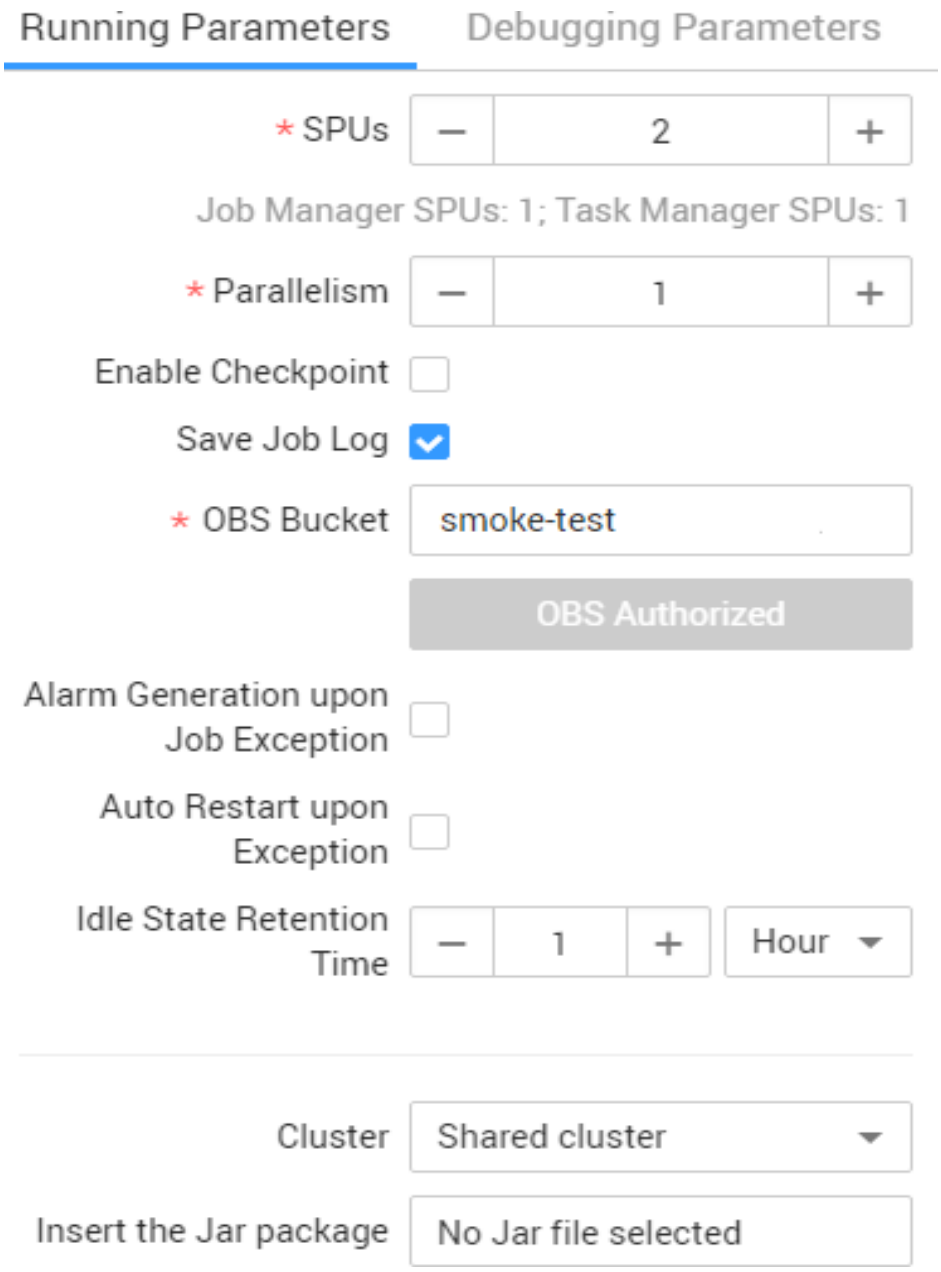
- You can only debug, submit, or start a job after semantic verification succeeds.
- If verification is successful, the message "The SQL semantic verification is complete. No error." will be displayed.
- If verification fails, a red "X" mark will be displayed next to each SQL statement with errors. You can move the cursor to the "X" mark to view error details and change the SQL statement as prompted.

Step 6 On the **Running Parameters** page in the right pane of the **Edit** page, set the parameters as follows:

- Retain the default settings for **SPUs** and **Parallelism**.
- Select **Save Job Log** and set **OBS Bucket** to **smoke-test**. If the selected OBS bucket is not authorized, click **Authorize OBS**.
- Retain the default setting **Shared cluster** for **Cluster**. Alternatively, you can select a user-defined exclusive cluster. For details about how to create a user-defined exclusive cluster, see **Creating a Cluster** in the *Cloud Stream Service User Guide*.

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-7 Running parameters



NOTE

CS provides a debugging function to help you verify the business logic of jobs by using test data. For details, see **Debugging a Job** in the *Cloud Stream Service User Guide*.

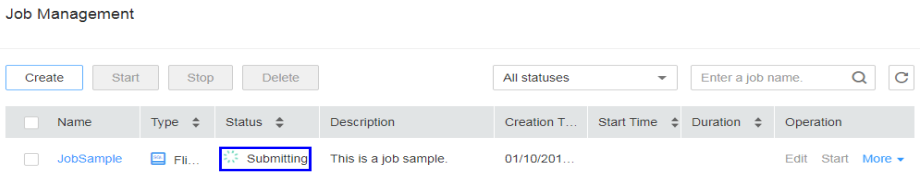
Step 7 Click **Save**.

Step 8 Click **Submit**. On the displayed **Job Bill** page, click **OK** to submit and start the job.

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**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-8 Job status



If **Status** of a job is **Submission failed** or **Running exception**, the job submission failed or the job did not execute successfully. In this case, click the job name **JobSample** in the job list and click **Running Logs** to query the job's run log. Rectify faults based on the log and resubmit the job.

----End

Step 5: Send Data to DIS

Use the obtained DIS stream as the data source. After the job is submitted, you can continuously upload data to DIS over the DIS stream to provide real-time stream data sources to CS jobs.

In this example, local data is uploaded to DIS through the csinput stream. For details, see **Sending Data to DIS** in the **Data Ingestion Service User Guide**.

In the following example, each data record includes four fields: the license plate number, owner name, brand, and price of the vehicle.

```
1,lilei,bmw320i,28
2,hanmeimei,audia4,27
```

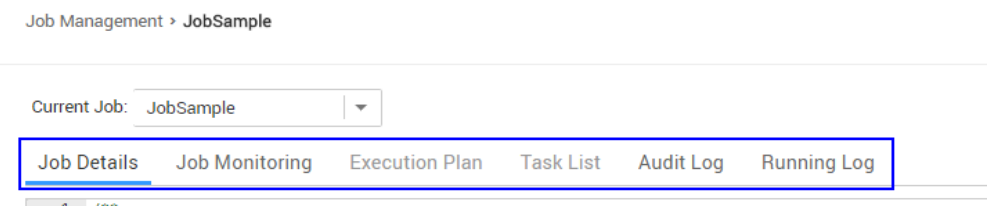
Step 6: View Job Information and Job Execution Result

After a job is started, you can view the job status by performing the following steps:

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 job list, click job name **JobSample** to view job details.

Figure 1-9 Job monitoring



For details, see **Monitoring a Job** in the *Cloud Stream Service User Guide*.

To view the results of a job, log in to the DIS management console and view data on the DIS stream that serves as the sink stream. For details, see **Retrieving Data from DIS** in the **Data Ingestion Service User Guide**.

----End

 Copyright © Huawei Technologies Co., Ltd.

Step 7: View Related Documents

After performing the preceding steps, you can refer to the following to learn more about CS:

- **Cloud Stream Service User Guide**: This document explains concepts of a job, template, and cluster, as well as details about related operations.
- **Cloud Stream Service API Reference**: This document explains how to use CS APIs to perform job operations.
- **Cloud Stream Service Stream Ecosystem Development Guide**: This document discusses the cloud service and open-source ecosystems related to Cloud Stream Service and provides guidance on how to use stream ecosystems for development.
- **Cloud Stream Service SDK Reference**: This document describes how to install and configure the development environment, and how to perform secondary development by invoking API functions provided by CS SDK.
- **Cloud Stream Service SQL Syntax Reference**: This document details common SQL syntaxes used for CS.

Step 8: Delete the Job

You can delete unwanted example jobs to avoid resource waste or quota occupation.

NOTE

Deleted jobs cannot be restored. Exercise caution when deleting jobs.

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 row where job **JobSample** is located on the **Job Management** page, choose **More > Delete**.

Step 3 In the displayed dialog box, click **OK**.

----End

 Copyright © Huawei Technologies Co., Ltd.