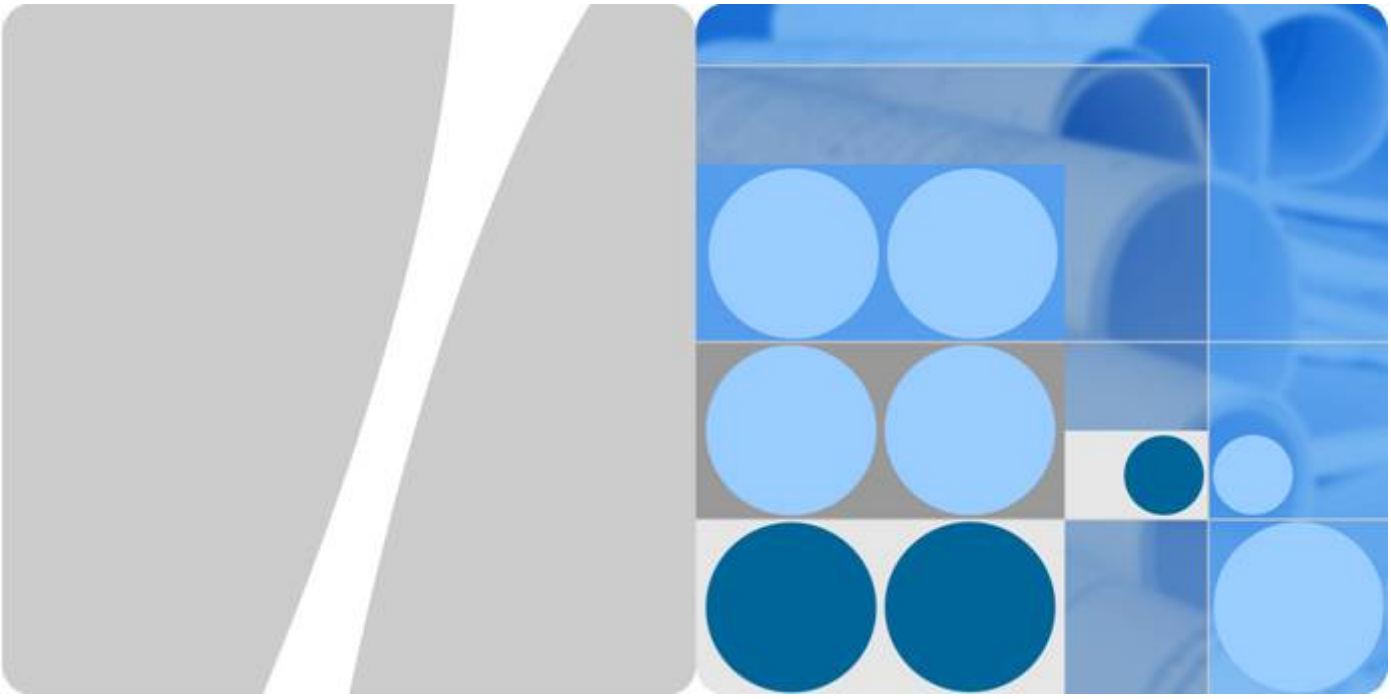


Cloud Stream Service

FAQ

Issue 18

Date 2020-09-29

HUAWEI TECHNOLOGIES CO., LTD.

HUAWEI

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.

Contents

 Copyright © Huawei Technologies Co., Ltd.

 Copyright © Huawei Technologies Co., Ltd.

1 General FAQs

1.1 What Is Real-Time Stream Computing?

"Real-time" is a description of events that are happening. "Stream", like the water flow, has the one-by-one feature. "Computing" involves arithmetic operations, mathematical analysis, and execution of algorithm models.

Real-time stream computing implements online analysis or algorithm operations one by one based on events that are happening.

1.2 What Is CS?

Cloud Stream Service (CS) is a real-time big data stream analysis service running on the public cloud. Computing clusters are fully managed by CS, allowing you to focus on StreamSQL services. CS is fully compatible with Apache Flink 1.5.3 and Apache Spark 2.2.1 APIs.

Recommended in the IT field, CS delivers distributed real-time stream computing with low latency (millisecond-level latency), high throughput, and high reliability. Powered on Flink, CS integrates enhanced features and security, and supports both stream processing and batch processing methods. It provides Stream SQL features required for data processing, and will add algorithms of machine learning and graph computing to Stream SQL in the future.

1.3 What Are the Advantages of CS?

CS has the following advantages:

- Distributed real-time stream computing

  CS supports large-scale distributed clusters and cluster auto-scaling. You can adjust your cluster capacity based on the resources required by your jobs, minimizing the costs.

- Easy to use

  You can use the online SQL editor to compile Stream SQL statements to define the source stream, sink stream, and data processing logic to rapidly

 Copyright © Huawei Technologies Co., Ltd.

implement business logic. With CS, you can analyze streaming data without managing clusters and learning more programming skills.

- Support of exclusive clusters

  CS supports auto scaling and is fully managed, which frees you from managing clusters, big data frameworks, and resource scheduling frameworks. It also visualizes the running status of your submitted jobs. You can run your jobs in a shared cluster or exclusive cluster. Exclusive clusters are physically isolated from shared clusters and other tenants' clusters. You can also manage the quota of exclusive clusters.

- Secure isolation

  Security protection mechanisms for tenants ensure secure job running. Tenants' computing clusters are physically isolated from each other and protected by independent security configurations.

- Pay-per-use billing mode

  You are only charged for the stream processing unit (SPU) resources you use by usage duration (seconds). An SPU contains one core and 4 GB memory.

- High throughput and low latency

  CS uses the Dataflow model of Apache Flink, a real-time computing framework. High-performance computing resources are used to consume data from your created Kafka, DMS Kafka, and MRS Kafka clusters. A single SPU processes about 10,000 messages per second.

1.4 What Are the Features of CS?

CS provides the following functions:

- Abundant StreamSQL online analysis capabilities

  Aggregation functions, such as Window and Join, geographical functions, and CEP functions are supported. SQL is used to express business logic, facilitating service implementation. For details, see the **Cloud Stream Service SQL Syntax Reference**.

- StreamingML

  CS Provides multiple streaming machine learning methods to analyze and predict data in real time. You only need to call related functions through SQL statements to perform data statistics, anomaly detection, real-time clustering, and time series analysis. For details, see **StreamingML**.

- Geographical location analysis

  CS offers geographical location analysis functions to analyze geospatial data in real time. You can fulfill yaw detection and geo-fencing through SQL statements. For details, see **Geographical Functions**.

- CEP SQL

  CS provides pattern matching and detection based on Match Recognize to assist business personnel in anomaly detection based on complex event rules through SQL. You can apply this function in various scenarios, such as fraud detection, abnormal vehicle behavior detection, and abnormal running status detection of industrial devices. For details, see **Pattern Matching**.

- Data visualization

 Copyright © Huawei Technologies Co., Ltd.

CS works with charts of various types to display job output data in real time. You can directly use API Gateway (APIG) to access job data as well as customize data to be visualized. For details, see **Data Visualization**.

- Visual SQL editor

  CS provides a visual editor for users who are not familiar with SQL development. The visual editor encapsulates upstream and downstream services (such as DIS and CloudTable) and internal logic operators (such as filter and window) that need to be interconnected with CS into drag-and-drop components. It allows you to easily create a job topology by dragging required elements into the canvas and then connecting them. By clicking each element in the canvas, you can set related parameters. For details, see **Visual Editor**.

- Exclusive cluster creation and resource quota allocation for jobs

  Tenants can create exclusive clusters, which are physically isolated from shared clusters and other tenants' clusters and are not subject to other jobs. Tenants can also configure the maximum SPU quota for their exclusive clusters and allocate available clusters and SPU quota for their sub-users. For details, see **Cluster Management**.

- Online SQL job testing

  Job debugging helps you check whether a SQL statement's logic is correct. After sample data is input manually or using OBS, the correct SQL statement logic will export results as expected. For details, see **Debugging a Job**.

- Support for user-defined Flink and Spark jobs

  You can submit user-defined Flink and Spark jobs in exclusive clusters.

- Support for Spark streaming and structured streaming

  You can submit user-defined Spark streaming jobs in exclusive clusters.

- Interconnection with multiple cloud services to form a rich stream ecosystem

  The CS ecosystem consists of cloud service ecosystems and open-source ecosystems.

  - Cloud service ecosystems: CS can be interconnected with other cloud services by using Stream SQL. You can directly use SQL statements to read and write data from various cloud services, such as Data Ingestion Service (DIS), Object Storage Service (OBS), CloudTable Service (CloudTable), Distributed Cache Service (DCS), MapReduce Service (MRS), Relational Database Service (RDS), and Simple Message Notification (SMN).
  - Open-source ecosystems: After connections to other VPCs are established through VPC peering connections, you can access all data sources and output targets (such as HBase, Elasticsearch, and Kafka) supported by Flink and Spark in your exclusive CS cluster.

  For details, see the **Cloud Stream Service Stream Ecosystem Development Guide**.

- Interconnection with Intelligent EdgeFabric (IEF)

  IEF works together with CS to provide on-cloud management, stream processing on edge devices, and real-time stream processing.

 Copyright © Huawei Technologies Co., Ltd.

1.5 What Are the Application Scenarios of CS?

CS focuses on Internet and IoT service scenarios that require real-time processing and high throughput. Basically, CS provides IoV services, online log analysis, online machine learning, online graph computing, and online recommendation algorithm application for multiple industries, such as small- and medium-sized enterprises in the Internet industry, IoT, IoV, and financial anti-fraud.

Real-Time Stream Analysis

CS provides real-time stream analysis with ease of use, low latency, and high throughput. You can use Stream SQL or customize jobs to perform stream analysis.

Advantages

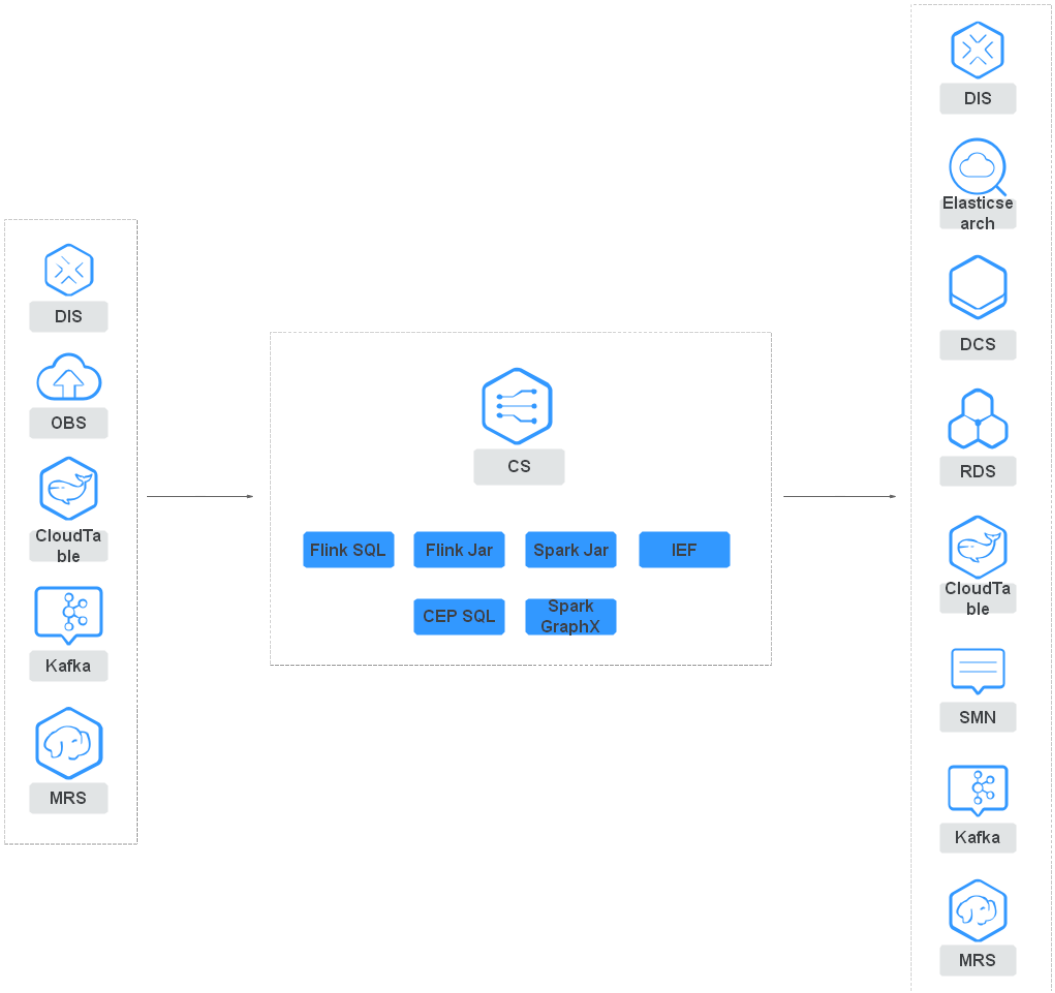
- Easy to use: Stream SQL is edited online. Rich SQL functions meet the requirements of complex services.
- Fully-managed: You can focus on stream analysis without managing computing clusters.
- Pay-per-use billing mode: You are billed for the usage duration (precise to seconds) of SPUs used by your jobs.

Features: Complex stream analysis methods, such as Window, CEP, and Join, can be performed on streaming data with millisecond-level latency.

Application scenarios: real-time log analysis, network traffic monitoring, real-time risk control, real-time data statistics, and real-time data ETL

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-1 Real-time stream analysis



IoT

IoT or edge devices upload data to DIS or other storage services. CS reads data from DIS, analyzes data in real time (involving anomaly detection, data cleansing, statistical analysis, and metric warning), and makes the stream analysis result persistent or reports alarms in real time.

Advantages

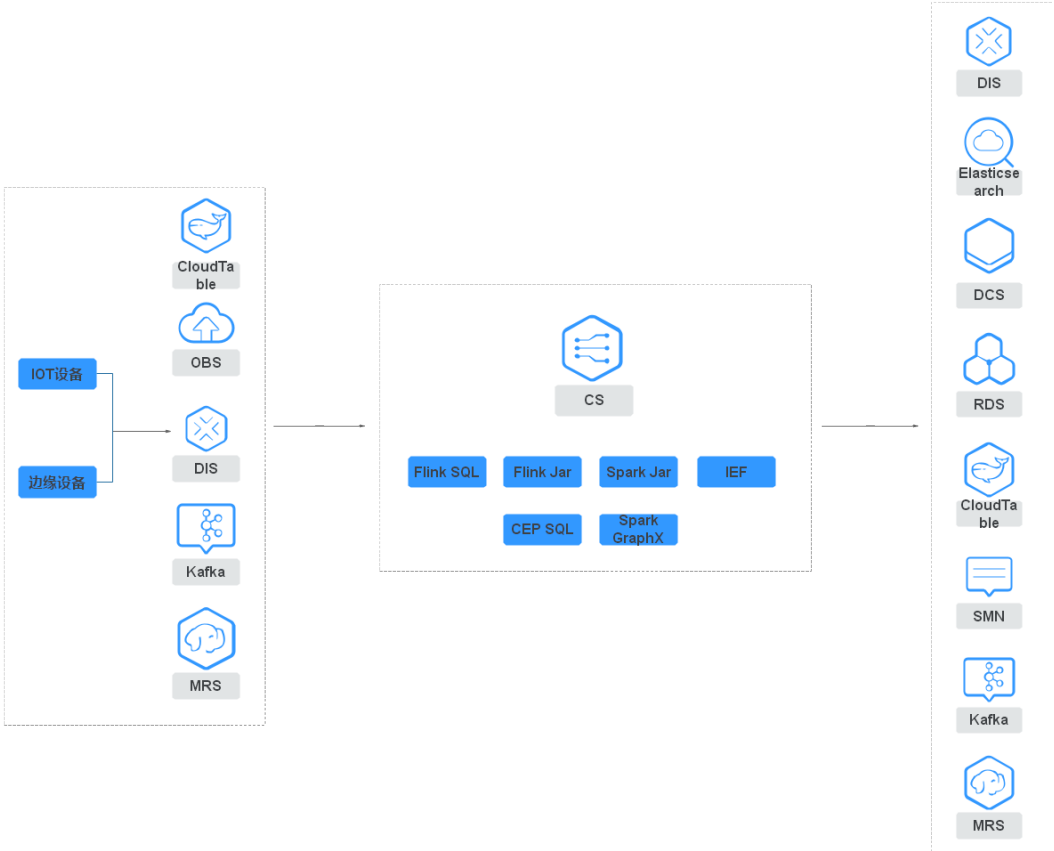
- Various IoT SQL functions: include region detection function, yaw detection function, and common IoT functions used for relative position determination.
- High throughput and low latency: CS adopts the Apache Flink engine to deliver a real-time computing framework.
- Secure isolation: Tenants are isolated from each other to ensure data security.

Feature: IoT services call the APIs of CS. CS then reads sensor data in real time and executes users' analysis logic. Analysis results are sent to services, such as DIS and RDS, for data persistency, alarm or report display, or visual display of results.

Application scenarios: elevator IoT, industrial IoT, shared bicycles, IoV, and smart home

 Copyright © Huawei Technologies Co., Ltd.

Figure 1-2 IoT



1.6 What Skills Do I Need to Use CS?

You need to be able to write SQL statements. To develop user-defined jobs, you also need to be familiar with Flink APIs.

1.7 What Is an SPU?

SPU is short for stream processing unit. An SPU is the charging unit used for CS. In a standard configuration, a single SPU provides you with 1 vCPU and 4 GB of memory. Multiple SPUs can be configured for the same job.

1.8 What Should I Do to Get Started with CS?

The procedure for getting started with CS is as follows:

1. Prepare data sources and data output channels.
2. Create an OBS bucket for saving output data.
3. Apply for CS.
4. Create and then submit a job.
5. Send data to the data source.
6. View the job information and the job execution result.

 Copyright © Huawei Technologies Co., Ltd.

For details, see **Getting Started**.

1.9 What Is User Quota?

In the public , quotas are used to limit the number of resources available to users. If the existing resource quota cannot meet your service requirements, you can submit a work order to increase your quota. Once your application is approved, we will update your resource quota accordingly and send you a notification. For details about operations on quotas, see **Quotas**.

1.10 How Do I Obtain the AK/SK Pair?

The access key ID (AK) and secret access key (SK) are a pair of access keys used together to authenticate users who wish to make API requests. The AK/AS pair provides functions similar as a password. When users make API requests to manage cloud resources (for example, creating a cluster), the AK/SK pair is required to sign the requests. This mechanism ensures the confidentiality and integrity of the requests as well as the correctness of the identities of both parties. Access keys can be generated and managed on the **My Credentials** page. To obtain the AK/SK pair, perform the following steps:

1. Register and log in to the **public cloud management console**.
2. Move the cursor over your username in the upper right corner of the management console and click **My Credentials** from the drop-down list.
3. Click the **Access Keys** tab.
4. Click **Add Access Key** to switch to the **Add Access Key** dialog box.
5. Enter your login password.
6. Enter the verification code sent to your mail or mobile phone.

   NOTE

   For users created in IAM, if no email address or mobile phone was specified during user creation, the login password is enough. No verification code will be required.

7. Click **OK**.

   NOTE

   Keep the AK/SK file somewhere safe to prevent information leakage.

1.11 How Do I Obtain the Project ID?

A project ID is the ID of the region where a system resides. When you access the public cloud system through APIs to perform operations on cloud resources (for example, creating a cluster), you must provide a project ID.

To view the project ID, perform the following steps:

1. Log in to the **public cloud management console**.
2. Move the cursor over your username in the upper right corner of the management console and click **My Credentials** from the drop-down list.

 Copyright © Huawei Technologies Co., Ltd.

On the displayed **My Credentials** page, view the project ID on the **Projects** page, for example, **5a3314075bfa49b9ae360f4ecd333695**.

1.12 Where Can I Get Help When Encountering Troubles?

To better help and serve users, we provide multiple methods to provide support. For details, see **https://www.huaweicloud.com/en-us/about/contact.html**.

1.13 What Should I Do If the System Displays a Message Indicating that No SMN Topics Exist?

Go to the IAM console, select the user group that your member account belongs to, and add an SMN policy for the corresponding region.

1.14 Regions and AZs

What Are Regions and AZs?

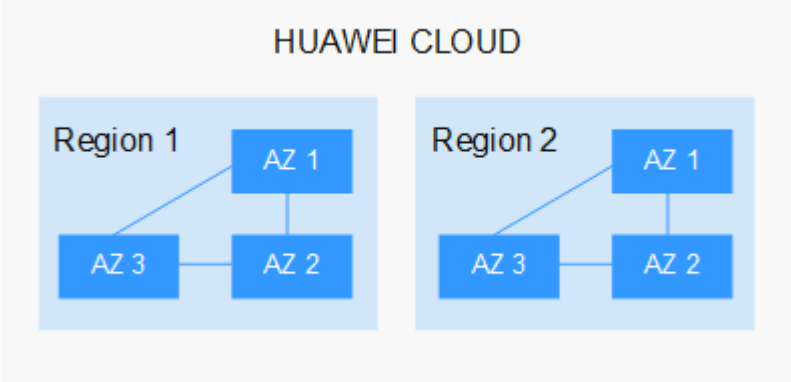
A region and availability zone (AZ) identify the location of a data center. You can create resources in a specific region and AZ.

- Regions are classified based on geographical location and network latency. Public services, such as Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Object Storage Service (OBS), Virtual Private Cloud (VPC), Elastic IP (EIP), and Image Management Service (IMS), are shared within the same region. Regions are classified as universal regions and dedicated regions. A universal region provides universal cloud services for common tenants. A dedicated region provides services of the same type or only provides services for specific tenants.
- An AZ contains one or more physical data centers. Each AZ has independent cooling, fire extinguishing, moisture-proofing, and electricity facilities. Within an AZ, computing, network, storage, and other resources are logically divided into multiple clusters. AZs within a region are interconnected using high-speed optical fibers to allow you to build cross-AZ high-availability systems.

Figure 1-3 shows the relationship between regions and AZs.

Copyright © Huawei Technologies Co., Ltd.

Figure 1-3 Regions and AZs



HUAWEI CLOUD provides services in many regions around the world. You can select a region and AZ as needed. For more information, see **HUAWEI CLOUD Global Regions**.

How to Select a Region?

When selecting a region, consider the following factors:

- Location

  You are advised to select a region close to you or your target users. This reduces network latency and improves access speed. However, Chinese mainland regions provide the same infrastructure, BGP network quality, as well as resource operations and configurations. Therefore, if you or your target users are in the Chinese mainland, you do not need to consider the network latency differences when selecting a region.

  The countries and regions outside the Chinese mainland, such as Bangkok and Hong Kong, provide services for users outside the Chinese mainland. If you or your target users are in the Chinese mainland, these regions are not recommended due to high access latency.

  - If you or your target users are in the Asia Pacific area (excluding the Chinese mainland), select the **AP-Hong Kong**, **AP-Bangkok**, or **AP-Singapore** region.
  - If you or your target users are in Africa, select the **AF-Johannesburg** region.
  - If you or your target users are in Europe, select the **EU-Paris** region.

- Resource price

  Resource prices may vary in different regions. For details, see **Product Pricing Details**.

How to Select an AZ?

When determining whether to deploy resources in the same AZ, consider your applications' requirements for disaster recovery (DR) and network latency.

- For high DR capability, deploy resources in different AZs in the same region.
- For low network latency, deploy resources in the same AZ.

 Copyright © Huawei Technologies Co., Ltd.

Regions and Endpoints

When using an API to use resources, you must specify its region and endpoint. For details about HUAWEI CLOUD regions and endpoints, see **Regions and Endpoints**.

 Copyright © Huawei Technologies Co., Ltd.

2 Job Related FAQs

2.1 Which Data Sources Does CS Support?

CS can process data from the following:

- DIS
- OBS
- MRS
- Kafka cluster
- CloudTable

2.2 Where Can CS Job Results Be Exported?

The CS job result data can be exported to the following:

- DIS
- OBS
- RDS
- DWS
- DDS
- SMN
- Kafka cluster
- CloudTable
- Cloud Search Service (CSS)
- DCS Redis instance
- MRS
- APIG

For details about the syntax used to create the preceding sink streams, see the **Cloud Stream Service Stream Ecosystem Development Guide**.

 Copyright © Huawei Technologies Co., Ltd.

2.3 Which Data Formats Does CS Support?

- CS can read data in CSV, JSON, or EMAIL format from DIS.
- CS can write data in CSV or JSON format to DIS.
- CS can read data in CSV format from OBS.
- CS can write data in CSV, JSON, or ORC format to OBS.
- CS can send data in text format to SMN.
- CS can read and write data in CSV or JSON format from and to Kafka.
- CS can read data from CloudTable and store data in tables in CloudTable.
- CS can send JSON data to CSS.
- CS can send data in key-value format to DCS Redis.
- CS can write data in CSV or JSON format to APIG.

For details, see **Data Formats** in the *Cloud Stream Service Stream Ecosystem Development Guide*.

2.4 What Kind of Code-based Jobs Does CS Support?

CS supports jobs developed using SQL statements and user-defined jobs using JAR files.

2.5 How Is Job Concurrency Implemented?

In CS, job concurrency indicates to start multiple concurrent tasks in a job. A SQL-statement-based job does not support concurrency of a single operator.

2.6 How Can I Check Job Results?

- CS can output job results to DIS. You can view the results in DIS. For details, see **Retrieving Data from DIS** in the **Data Ingestion Service User Guide**.
- CS can output job results to RDS. You can view the results in RDS. For details, see **Relational Database Service Getting Started Guide**.
- CS can output job results to SMN, and SMN sends the results to the user's terminal. For details, see **Simple Message Notification Getting Started Guide**.
- CS can output job results to Kafka. You can view the results in Kafka clusters. For details, visit the **Kafka official website**.
- CS can output job results to CloudTable. You can view the results in CloudTable. For details, see **Getting Started** in the **CloudTable Service User Guide**.
- CS can output job results to IEF. You can view the results in IEF. For details, see the *Intelligent EdgeFabric User Guide*.
- CS can export job results to CSS. You can view the results in CSS. For details, see **Cloud Search Service Getting Started Guide**.

 Copyright © Huawei Technologies Co., Ltd.

- CS can export job results to DCS. You can view the results in DCS. For details, see **Distributed Cache Service Getting Started Guide**.

2.7 What Should I Do If the OBS Bucket Selected for a Job Is Not Authorized?

If the OBS bucket selected for a job is not authorized, perform the following steps:

Step 1 On the CS management console, click **Job Management**.

Step 2 In the row where the target job resides, click **Edit** under **Operation** to switch to the **Edit** page.

Step 3 Configure parameters under **Running Parameter** on the **Edit** page.

1. Select **Enable Checkpoint** or **Save Job Log**.
2. Specify **OBS Bucket**.
3. Select **Authorize OBS**.

----End

2.8 How Do I Confirm that My Exclusive Cluster Is Connected to a Specified Address?

After an exclusive cluster is created, you can test the connectivity between the cluster and your specified address.

The procedure is as follows:

Step 1 Log in to the CS console.

Step 2 In the navigation tree on the left pane of the CS management console, click **Cluster Management** to switch to the **Cluster Management** page.

Step 3 In the row where the cluster you want to view is located, click the cluster name in the **Name** column to switch to the **Cluster Details** page.

Step 4 Click **Test Address Connectivity**.

Step 5 In the displayed dialog box, enter the address to be tested and click **Test**.

The system starts to test whether the cluster and the specified address are connected. The address can be a domain name, an IP address. You can specify the port for the IP address. If a port is specified, use the ip:port or hostname:port format.

NOTE

If the specified IP address is not enabled with ICMP, you are advised to use the specified port to test the connectivity.

If the network between the exclusive cluster and the peer end is disconnected, create a VPC peering connection.

 Copyright © Huawei Technologies Co., Ltd.

For details about how to set up the VPC peering connection, see **VPC Peering Connection** in *Cloud Stream Service User Guide*.

----End

2.9 What Should I Do If My CS Job Fails to Be Submitted Due to Connection Timeout?

Cause:

The network connection between the exclusive cluster where your CS job runs and the target address has been lost.

Solution:

Set up a VPC peering connection to ensure network connectivity between the exclusive cluster where your CS job runs and the target address.

For details about how to set up the VPC peering connection, see **VPC Peering Connection** in *Cloud Stream Service User Guide*.

After a VPC peering connection is created, test the address connectivity by referring to **How Do I Confirm that My Exclusive Cluster Is Connected to a Specified Address?**.

If the network is still disconnected, check related configurations by referring to **What Can I Do If VPCs in a VPC Peering Connection Cannot Communicate with Each Other?**.

CS jobs must run on exclusive clusters in any of the following scenarios:

- CS outputs the job result to RDS.
- CS outputs the job result to DWS in JDBC mode.
- CS outputs the job result to the CSS cluster.
- CS outputs the job result to the DCS Redis instance.
- The source stream ingests data from Kafka as input data for CS jobs.
- CS outputs the job result to Kafka.
- CS outputs the job result to MRS HBase.

Before running a job in any of the preceding scenarios, ensure that a VPC peering connection has been established between the CS exclusive cluster where the job resides and the address to be connected.

2.10 What Should I Do If the DIS Stream Does Not Exist During the Semantic Check?

1. View the job SQL statements to check whether the DIS stream was created. If it was not created, create a DIS stream by referring to **Creating a DIS Stream** in the **Data Ingestion Service User Guide**.
2. Verify that the DIS stream and CS job are located in the same region.

 Copyright © Huawei Technologies Co., Ltd.

2.11 How Does CS Interconnect with DMS Using SASL_SSL?

A Kafka premium instance of DMS can be the source and sink streams of CS.

If you enable SASL_SSL when creating a DMS instance, data will be encrypted before transmission for enhanced security.

CS connects to a DMS instance in command line mode. The procedure is as follows:

Step 1 Obtain the **client.truststore.jks** certificate.

Click **here** to download a package and decompress the package to obtain the client certificate file **client.truststore.jks**.

Step 2 Upload the **client.truststore.jks** file to a directory on the CS cluster, for example, **/opt/cs/user_files**.

Step 3 Add the following parameters to the SQL code: **kafka_properties_delimiter** and **kafka_properties**.

The sample code is as follows:

```
CREATE SINK STREAM kafka_sink (
    id STRING
)
WITH (
  type="kafka",
  kafka_bootstrap_servers = "192.168.0.143:9093,192.168.0.113:9093,192.168.0.180:9093",
  kafka_topic = "csdms",
  encode = "csv",
  kafka_properties_delimiter = ",",
  kafka_properties = "sasl.jaas.config=org.apache.kafka.common.security.plain.PlainLoginModule required
username=\"test\" password=\"*******
\";,sasl.mechanism=PLAIN,security.protocol=SASL_SSL,ssl.truststore.location=/opt/cs/user_files/
client.truststore.jks,ssl.truststore.password=dms@kafka"
);
```

username and **password** are the username and password configured when enabling SASL_SSL for DMS.

After the configurations are complete, CS interconnects with DMS.

----End

For details about how to connect to a Kafka premium instance using SASL, see **Distributed Message Service User Guide**.

2.12 How Do I Save Checkpoints to OBS for a Custom Flink Job?

CS can save checkpoints of a custom Flink job to OBS. The code example is as follows:

```
import org.apache.flink.streaming.api.environment.StreamExecutionEnvironment;
import org.apache.flink.runtime.state.filesystem.FsStateBackend;
import org.apache.flink.contrib.streaming.state.RocksDBStateBackend;
```

 Copyright © Huawei Technologies Co., Ltd.

```
StreamExecutionEnvironment env = StreamExecutionEnvironment.getExecutionEnvironment();
RocksDBStateBackend rocksDbBackend = new RocksDBStateBackend(new FsStateBackend("OBS path"),
true);
env.setStateBackend(rocksDbBackend);
```

The OBS path is **obs://ak:sk@OBS endpoint/bucket name/bucket directory**. **ak** and **sk** are the AK/SK of the account owning the OBS bucket to be uploaded.

2.13 Why I Can't Select an Exclusive Cluster for an Existing Fink SQL Job?

You need to bind the cluster with your account.

1. On the left of the CS management console, click **Cluster Management**.
2. On the **Cluster Management** page, click **User Quota Management**.
3. Bind a specified cluster with the user who creates the job.

Copyright © Huawei Technologies Co., Ltd.

3 Billing Related FAQs

3.1 How Am I Billed by CS?

CS charging is simple and controllable. You can choose either pay-per-use (hourly) or a yearly/monthly package, which is more economic. For details about the service tariff, see the **Cloud Stream Service Price Details**.

3.2 How Am I Billed for Other Services?

You are billed for other cloud services based on the billing standards of other services. For details about the billing standards, visit the **Product Pricing Details page**.

3.3 Do I Need to Pay If CS Does Not Ingests Data?

Yes. Once a job gets running, you will be billed no matter whether CS ingests data.

 Copyright © Huawei Technologies Co., Ltd.

4 SQL Related FAQs

4.1 Does CS Support JOIN Between Stream Data and Table Data?

Currently, the JOIN operation between stream data and DCS Redis table data is supported. For details, see the **[Ecosystem]DIS-CS=DCS-DIS_SAMPLE_TEMPLATE** in **Job Template** on the CS management console. For details about the syntax of JOIN operation between stream data and table data, see **JOIN Between Stream Data and Table Data** in the *Cloud Stream Service SQL Syntax Reference*.

 Copyright © Huawei Technologies Co., Ltd.