<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 General FAQs</strong></td>
</tr>
<tr>
<td>1.1 What Is DLI?</td>
</tr>
<tr>
<td>1.2 What Are the Application Scenarios of DLI?</td>
</tr>
<tr>
<td>1.3 Which Data Formats Does DLI Support?</td>
</tr>
<tr>
<td>1.4 Where Can DLI Data be Stored?</td>
</tr>
<tr>
<td>1.5 How Can I Use DLI If Data Is Not Uploaded to OBS?</td>
</tr>
<tr>
<td>1.6 Regions and AZs</td>
</tr>
<tr>
<td><strong>2 Billing</strong></td>
</tr>
<tr>
<td>2.1 What Is the Billing Mode of DLI?</td>
</tr>
<tr>
<td>2.2 How Do I Check the Billing?</td>
</tr>
<tr>
<td>2.3 What Is the Difference Between the Following Two Payment Modes: One Is to Purchase 4,000-CU Resources for Three Months at a Time, the Other Is to Purchase 4,000-CU Resources for One Month for Three Times?</td>
</tr>
<tr>
<td><strong>3 Problems Related to Quota</strong></td>
</tr>
<tr>
<td>3.1 What Is User Quota?</td>
</tr>
<tr>
<td>3.2 How Do I View My Quotas?</td>
</tr>
<tr>
<td>3.3 How Do I Increase a Quota?</td>
</tr>
<tr>
<td><strong>4 Problems Related to Permissions</strong></td>
</tr>
<tr>
<td>4.1 How Do I Create a Sub-user?</td>
</tr>
<tr>
<td>4.2 How Do I Modify a User Policy?</td>
</tr>
<tr>
<td>4.3 What Is Column Permission Granting of a DLI Partition Table?</td>
</tr>
<tr>
<td><strong>5 Problems Related to Management Console</strong></td>
</tr>
<tr>
<td>5.1 What Should I Do If a Message Indicating that the SMN Topic Does Not Exist Is Displayed When I Use the SMN Topic in DLI?</td>
</tr>
<tr>
<td><strong>6 Problems Related to SQL Jobs</strong></td>
</tr>
<tr>
<td>6.1 How Can I Avoid Garbled Characters Caused by Inconsistent Character Codes?</td>
</tr>
<tr>
<td>6.2 The Compression Ratio of OBS Tables Is Too High</td>
</tr>
<tr>
<td>6.3 What Should I Do If Error “path obs://xxx already exists” Is Reported When Data Is Exported From DLI to OBS?</td>
</tr>
<tr>
<td>6.4 A Schema Parsing Error Is Reported When You Create a Hive Table Using CTAS</td>
</tr>
<tr>
<td><strong>7 Problems Related to Spark Jobs</strong></td>
</tr>
<tr>
<td>7.1 What Can I Do When Timeout Occurs During Creation of a Spark Session?</td>
</tr>
</tbody>
</table>
7.2 How Do I Set the AK/SK for a Spark Queue to Operate an OBS Table? ........................................................... 16
7.3 What Can I Do When Receiving java.lang.AbstractMethodError in the Spark Job? ........................................ 17

8 Problems Related to Datasource Connections ................................................................................................. 18
8.1 How Can I Perform Query on Data Stored on Services Rather Than DLI?...................................................... 18
8.2 How Do I Set the Auto-increment Primary Key or Other Fields That Are Automatically Filled in the RDS Table When Creating a DLI and Associating It with the RDS Table? ................................................................. 18
8.3 Why Is the Error Message "communication link failure" Displayed When I Use a Newly Activated Datasource Connection? ............................................................................................................................................................ 18

9 Problems Related to Flink Jobs ......................................................................................................................... 22
9.1 How Do I Save Checkpoints to OBS for a Custom Flink Job? ........................................................................... 22
9.2 What Should I Do If the DIS Stream Does Not Exist During the Semantic Check? ............................................... 22
9.3 How Can I Check Job Results? ........................................................................................................................... 22
9.4 What Should I Do If the OBS Bucket Selected for a Job Is Not Authorized? ....................................................... 23
9.5 How Do I Write Data to Different Elasticsearch Clusters in a Flink Job? ............................................................ 23
9.6 Does Data in the Temporary Stream of Flink SQL Need to Be Cleared Periodically? How Do I Clear the Data?................................................................................................................................................................................................ 23
9.7 How Does a DLI Flink Job Interact with Data Sources on the Public Network? .................................................. 24
9.9 What Should I Do If Logs Are Not Written to the OBS Bucket After a DLI Flink Job Fails to Be Submitted for Running? ............................................................................................................................................................. 24

10 Problems Related to APIs ............................................................................................................................... 26
10.1 How Do I Obtain the AK/SK Pair? ........................................................................................................................ 26
10.2 How Do I Obtain the Project ID? ........................................................................................................................ 26
1.1 What Is DLI?

Data Lake Insight (DLI) is a Serverless big data compute and analysis service that is fully compatible with Apache Spark and Apache Flink ecosystems and supports batch streaming. With multi-model engines supported by DLI, enterprises can use SQL statements or programs to easily complete batch processing, stream processing, in-memory computing, and machine learning of heterogeneous data sources.

1.2 What Are the Application Scenarios of DLI?

DLI is applicable to large-scale log analysis, federated analysis of heterogeneous data sources, and big data ETL processing.

**Large-scale Log Analysis**

- Game operation data analysis
  
  Different departments of a game company analyze daily new logs via the game data analysis platform to obtain required metrics and make decision according to the obtained metric data. For example, the operation department obtains required metric data, such as new players, active players, retention rate, churn rate, and payment rate, through the platform to learn the current game status and determine follow-up actions. The placement department obtains the channel sources of new players and active players through the platform to determine the platforms for placement in the next cycle.

- Advantages
  
  - Efficient Spark programming model: DLI uses Spark Streaming to directly ingest data from DIS and perform preprocessing such as data cleaning. You only need to edit the processing logic, without the need to pay attention to the multi-thread model.
  
  - Easy to use: You can use standard SQL statements to compile metric analysis logic without paying attention to the complex distributed computing platform.
Pay-per-use: Log analysis is scheduled periodically based on the time requirements. There is a long idle period between each two scheduling operations. DLI adopts the pay-per-use billing mode, which saves the cost by more than 50% compared with the exclusive queue mode.

- It is recommended that you use the following related services:
  OBS, DIS, DWS, RDS

Federated Analysis of Heterogeneous Data Sources

- Digital service transformation of car companies
  In the face of new competition pressures and changes in travel services, car companies build the IoV cloud platform and IVI OS to streamline Internet applications and vehicle use scenarios, completing digital service transformation for car companies. This delivers better travel experience for vehicle owners, increases the competitiveness of car companies, and promotes sales growth. For example, DLI can be used to collect and analyze daily vehicle metric data (such as batteries, engines, tire pressure, and airbags), and give feedback on maintenance suggestions to vehicle owners in time.

- Advantages
  - No need for migration in multi-source data analysis: RDS stores the basic information about vehicles and vehicle owners, CloudTable stores real-time vehicle location and health status information, and DWS stores periodic metric statistics. DLI allows federated analysis on data from multiple sources without data migration.
  - Tiered data storage: Car companies need to retain all historical data to support auditing and other services that requiring infrequent data access. Warm and cold data is stored in OBS and frequently accessed data is stored in CloudTable and DWS, reducing the overall storage cost.
  - Rapid and agile alarm triggering: There are no special requirements for the CPU, memory, hard disk space, and bandwidth.

- It is recommended that you use the following related services:
  DIS, CDM, OBS, DWS, RDS, and CloudTable

Big Data ETL Processing

- Carrier big data analysis
  Carriers typically require petabytes, or even exabytes of data storage, for both structured (base station details) and unstructured (messages and communications) data. They need to be able to access the data with extremely low data latency. Extracting value from this data efficiently is a major challenge. DLI provides multi-mode engines such as batch processing and stream processing to break down data silos and perform unified data analysis.

- Advantages
  - Big Data ETL: You can enjoy TB to EB-level data governance capabilities to quickly perform ETL processing on massive carrier data. Distributed datasets are provided for batch processing.
  - High Throughput, Low Latency: DLI uses the Dataflow model of Apache Flink, a real-time computing framework. High-performance computing resources are provided to consume data from your created Kafka, DMS...
Kafka, and MRS Kafka clusters. A single CU processes 1,000 to 20,000 messages per second.

- Fine-grained Permissions Management: Your company may have numerous departments, where data needs to be shared and isolated. Using DLI, you can apply for resource queues by tenant to isolate computing resources (CPUs and memory), ensuring job SLA. DLI supports table- or column-level data permission control, allowing for secure access for different departments.

- It is recommended that you use the following related services: OBS, DIS, and DAYU

**Geographic Big Data Analysis**

- Geographic big data analysis

Geographic big data has big data characteristics. It features large data volume (for example, PB-scale global satellite remote sensing image data is generated) and numerous data varieties (for example, structured remote sensing image raster data, vector data, unstructured spatial location data, and 3D modeling data). Users focus on how to use efficient mining tools or mining methods to get insights from the large volume of geographic big data.

- Advantages
  - Spatial Data Analysis Operators: With full-stack Spark capabilities and rich Spark spatial data analysis Spatial Data Analysis Operators With full-stack Spark capabilities and rich Spark spatial data analysis algorithm operators, DLI delivers comprehensive support for real-time processing of dynamic streaming data with location attributes and offline batch processing. DLI can handle massive data, including structured remote sensing image data, unstructured 3D modeling, and laser point cloud data.
  - CEP SQL: DLI delivers geographical location analysis functions to analyze geospatial data in real time. You can fulfill yaw detection and geo-fencing through SQL statements.
  - Big Data Processing: DLI allows you to quickly migrate remote sensing image data at the TB to EB scale to the cloud and perform image data slicing to offer resilient distributed datasets (RDDs) for distributed batch computing.

- It is recommended that you use the following related services: DIS, CDM, DES, OBS, RDS, and CloudTable

### 1.3 Which Data Formats Does DLI Support?

DLI supports the following data formats:

- Parquet
- CSV
- ORC
- Json
- Carbon
1.4 Where Can DLI Data be Stored?

DLI data can be stored in either of the following:

- OBS: Data used by SQL jobs, Spark jobs, and Flink jobs can be stored in OBS, reducing storage costs.
- DLI: The column-based Parquet format is used in DLI. That is, the data is stored in the Parquet format. The storage cost is relatively high.
- Datasource connection jobs can be stored in the connected services. Currently, CloudTable, CSS, DCS, DDS, DWS, MRS, and RDS are supported.

1.5 How Can I Use DLI If Data Is Not Uploaded to OBS?

Currently, DLI supports analysis only on the data uploaded to the cloud. In scenarios where regular (for example, on a per day basis) one-off analysis on incremental data is conducted for business, you can do as follows: Anonymize data to be analyzed and store anonymized data on OBS temporarily. After analysis is complete, export the analysis report and delete the data temporarily stored on OBS.

1.6 Regions and AZs

Concept

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 divided from the dimensions of 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 only or for specific tenants.
- An AZ contains one or more physical data centers. Each AZ has independent cooling, fire extinguishing, moisture-proof, 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 1-1 shows the relationship between 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.

Region Selection

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 rate. However, Chinese mainland regions provide basically the same infrastructure, BGP network quality, as well as operations and configurations on resources. 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 Asia Pacific excepting 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.

AZ Selection

When determining whether to deploy resources in the same AZ, consider your applications’ requirements on 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.
Regions and Endpoints

Before using an API to call resources, specify its region and endpoint. For details about HUAWEI CLOUD regions and endpoints, see Regions and Endpoints.
2.1 What Is the Billing Mode of DLI?

DLI queues are billed in pay-per-use mode. You are billed by calendar hour only after you purchase a queue.

Note that the computing resources of a pay-per-use queue are released after the queue is idle for one hour. When the queue is used again, the computing resources need to be reallocated, which may take 5 to 10 minutes.

If you need dedicated resources for a short period of time, you can select the dedicated resource mode when purchasing a pay-per-use queue. In this mode, resources are dedicated and used on demand (the billing period is equal to the life cycle of the queue).

For details about the billing modes, see Data Lake Insight Service Overview > Billing Description.

For details about the prices, see the Data Lake Insight Pricing Details.

2.2 How Do I Check the Billing?

If you feel that the billing is incorrect when you use DLI, perform the following steps:

- **NOTE**

  The following operations are performed on the DLI management console. For details, see Job Management in the Data Lake Insight User Guide.

- SQL Job
  a. Log in to the DLI management console.
  b. Choose Job Management > SQL Jobs.
  c. View the job details to be confirmed and check whether the following operations are performed within the fee deduction period:
     i. Use the self-created queue.
ii. Executed SQL jobs.

- If you used queues created by yourself to execute jobs, you are not billed incorrectly. The default billing mode is by CUH. For details, see the Data Lake Insight Pricing Details.
- If you used the default queue to execute jobs, you are billed based on the amount of data scanned. For details, see the Data Lake Insight Pricing Details.
- If no SQL job is executed, continue to check the Spark jobs.

- Spark Job
  b. Check whether the job is completed. If not, the job is billed by CUH. For the DLI pricing details, see the Data Lake Insight Pricing Details.
     If no Spark job is executed, continue to check the Flink jobs.

- Flink Job
  a. Choose Job Management > Flink Jobs.
  b. Check whether the job is completed. If not, the job is billed by CUH. For the DLI pricing details, see the Data Lake Insight Pricing Details.
     If no job is executed within the fee deduction period, check whether the billing is caused by the storage of a large amount of data.

- Data Storage
  a. Choose Data Management > Databases and Tables.
  b. Check whether the created database contains data stored in DLI.
     i. Click the name of the database to be viewed. The Table Management page is displayed.
     ii. Check whether the Data Location of the corresponding table is DLI.
        If you have stored data in the DLI, you are billed based on the amount of data stored. For details, see the Data Lake Insight Pricing Details.

If no problem is found after checking the preceding items, you can submit a service ticket. Service support personnel will help you check the problem.

2.3 What Is the Difference Between the Following Two Payment Modes: One Is to Purchase 4,000-CU Resources for Three Months at a Time, the Other Is to Purchase 4,000-CU Resources for One Month for Three Times?

If you purchase a queue of 4,000 CUs for three months at a time, the validity period of the queue is three months.
If you purchase queues for three times and each time purchase a queue with 4000 CUs for one month, three queues with 4000 CUs are purchased. When the one-month validity period expires, the three queues expire.
3 Problems Related to Quota

3.1 What Is User Quota?

HUAWEI CLOUD restricts the number and capacity of user resources. If the existing resource quota cannot meet your service requirements, you can submit a service ticket to increase your quota. Once your application is approved, we will update your resource quota accordingly and send you a notification. For details about the quotas, see Quotas.

3.2 How Do I View My Quotas?

1. Used to log in to the management console.
2. Click 🗺️ in the upper left corner of the management console to select a region.
3. In the upper right corner of the page, choose Resources > My Quotas. The Service Quota page is displayed.

Figure 3-1 My Quotas

![My Quotas](image)

4. View the used and total quota of each type of DLI resources on the displayed page.
   If a quota cannot meet service requirements, click Increase Quota to adjust it.
3.3 How Do I Increase a Quota?

1. Log in to the management console.
2. Click in the upper left corner and select the desired region.
3. In the upper right corner of the page, choose Resources > My Quota. The Service Quota page is displayed.

![Figure 3-2 My Quota](image)

4. Click Increase Quota.
5. On the Create Service Ticket page, configure parameters as required.
   In the Problem Description area, fill in the content and reason for adjustment.
6. After all mandatory parameters are configured, select I have read and agree to the Tenant Authorization Letter and click Submit.
4 Problems Related to Permissions

4.1 How Do I Create a Sub-user?

IAM users are useful if you have purchased multiple resources, such as ECSs, Elastic Volume Services (EVSs), and Bare Metal Servers (BMSs), on HUAWEI CLOUD as an administrator. To allocate the resources to different employees or applications in your enterprise, you can create IAM users for the employees or applications and grant them permissions required to complete tasks. The IAM users have their own usernames and passwords to log in to HUAWEI CLOUD.

For details, see Creating an IAM User and Assigning Permissions to an IAM User in the Identity and Access Management User Guide.

4.2 How Do I Modify a User Policy?

You can modify user policies by creating or modifying custom policies on the IAM console.

For details, see Creating a Custom Policy and Modifying and Deleting a Custom Policy in the Identity and Access Management User Guide.

4.3 What Is Column Permission Granting of a DLI Partition Table?

You cannot perform permission-related operations on the partition column of a partition table. However, when you grant the permission of any non-partition column in a partition table to another user, the user gets the permission of the partition column by default. When the user views the permission of the partition table, the permission of the partition column will not be displayed.
5.1 What Should I Do If a Message Indicating that the SMN Topic Does Not Exist Is Displayed When I Use the SMN Topic in DLI?

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

6.1 How Can I Avoid Garbled Characters Caused by Inconsistent Character Codes?

DLI supports only UTF-8-encoded texts. Ensure that data is encoded using UTF-8 during table creation and import.

6.2 The Compression Ratio of OBS Tables Is Too High

A high compression ratio of OBS tables in the Parquet or ORC format (for example, a compression ratio of 5 or higher compared with text compression) will lead to large data volumes to be processed by a single task. In this case, you are advised to set `dli.sql.files.maxPartitionBytes` to `33554432` (default: `134217728`) in the `conf` field in the `submit-job` request body to reduce the data to be processed per task.

6.3 What Should I Do If Error "path obs://xxx already exists" Is Reported When Data Is Exported From DLI to OBS?

Create an OBS directory that does not exist. Alternatively, you can manually delete the existing OBS directory and submit the job again. However, exercise caution when deleting the existing OBS directory because the operation will delete all data in the directory.

6.4 A Schema Parsing Error Is Reported When You Create a Hive Table Using CTAS

Currently, DLI supports the Hive syntax for creating tables of the TEXTFILE, SEQUENCEFILE, RCFILE, ORC, AVRO, PARQUET, and CARBON file types. If the file format specified for creating a table in the CTAS is AVRO and digits are directly
used as the input of the query statement (SELECT), for example, if the query is
CREATE TABLE tb_avro STORED AS AVRO AS SELECT 1, a schema parsing
exception is reported.

Description: If the column name is not specified, the content after SELECT is used
as both the column name and inserted value. However, the column name of the
AVRO table cannot be a digit. As a result, an error is reported, indicating that the
schema fails to be parsed.

Solution: You can use CREATE TABLE tb_avro STORED AS AVRO AS SELECT 1 AS
colName to specify the column name or set the storage format to a format other
than AVRO.
7 Problems Related to Spark Jobs

7.1 What Can I Do When Timeout Occurs During Creation of a Spark Session?

Increase the value of the `heartbeatTimeoutInSecond` parameter to prolong the session timeout period.

If you do not submit statements in a session or no request is sent to the session within the time interval specified by the `heartbeatTimeoutInSecond` parameter, the session automatically exits. For details, see “Table 2 Request parameters” in Creating a Session in the Data Lake Insight API Reference.

7.2 How Do I Set the AK/SK for a Spark Queue to Operate an OBS Table?

- When creating the `SparkContext`, do as follows:
  ```scala
  val sc: SparkContext = new SparkContext()
  sc.hadoopConfiguration.set("fs.obs.access.key", ak)
  sc.hadoopConfiguration.set("fs.obs.secret.key", sk)
  ```
- When creating the `SparkSession`, do as follows:
  ```scala
  val sparkSession: SparkSession = SparkSession
    .builder()
    .config("spark.hadoop.fs.obs.access.key", ak)
    .config("spark.hadoop.fs.obs.secret.key", sk)
    .enableHiveSupport()
    .getOrCreate()
  ```

The temporary AK/SK is recommended. For details, see Obtaining a Temporary Access Key and Security Token in the Identity and Access Management API Reference.

**NOTE**

For security purposes, you are advised not to include the AK and SK information in the OBS path. In addition, if a table is created in the OBS directory, the OBS path specified by the Path field cannot contain the AK and SK information.
7.3 What Can I Do When Receiving java.lang.AbstractMethodError in the Spark Job?

The Spark 2.3 has changed the behavior of the internal interface **Logging**. If the user code directly inherits the **Logging** and the earlier version Spark is used during compilation, the `java.lang.AbstractMethodError` is reported when the application runs in the Spark 2.3 environment.

Solutions are as follows:

- You can recompile the application based on Spark 2.3.
- You can use the **slf4j+log4j** to implement the log function instead of inheriting the internal interface **Logging** of the Spark. Details are described as follows:

```xml
<dependency>
  <groupId>org.slf4j</groupId>
  <artifactId>slf4j-api</artifactId>
  <version>1.7.16</version>
</dependency>
<dependency>
  <groupId>org.slf4j</groupId>
  <artifactId>slf4j-log4j12</artifactId>
  <version>1.7.16</version>
</dependency>
<dependency>
  <groupId>log4j</groupId>
  <artifactId>log4j</artifactId>
  <version>1.2.17</version>
</dependency>
```

```scala```
private val logger = LoggerFactory.getLogger(this.getClass)
logger.info("print log with slf4j+log4j")
```
8 Problems Related to Datasource Connections

8.1 How Can I Perform Query on Data Stored on Services Rather Than DLI?

To perform query on data stored on services rather than DLI, perform the following steps:

1. Assume that the data to be queried is stored on multiple services (for example, OBS) on HUAWEI CLOUD.
2. On the DLI management console, create a table and set the Path parameter to the save path of the target data, for example, path of an OBS bucket. The data is actually stored on OBS and data migration is not required.
3. On the DLI management console, edit SQL statements to query and analyze the data.

8.2 How Do I Set the Auto-increment Primary Key or Other Fields That Are Automatically Filled in the RDS Table When Creating a DLI and Associating It with the RDS Table?

When data is inserted into DLI, set the ID field to NULL.

8.3 Why Is the Error Message "communication link failure" Displayed When I Use a Newly Activated Datasource Connection?

- Cause
The network connectivity is abnormal. You are advised to check whether the security group is correctly selected and whether the VPC is correctly configured.

- **Solution:**
  
  Example: When you create an RDS datasource connection, the system displays the error message *Communication link failure*.

  a. Delete the original datasource connection and create a new one. When you create a new connection, ensure that the selected **Security Group**, **VPC**, **Subnet**, and **Destination Address** are the same as those in RDS.

    **NOTE**

    Select a correct **Service Type**. In this example, select **RDS**.

    **Figure 8-1 Creating a basic datasource connection-RDS**

    ![Creating a basic datasource connection-RDS](image)

  b. Check the configurations of VPC.

    If the error message is still displayed after creating a new datasource connection according to **Step 1**, check the VPC configuration.

    - **Classic datasource connections:**
      
      - **Inbound rule:** Check whether the inbound CIDR block and port in the security group have been enabled. If not, create the CIDR block and port you need.

    Check whether the CIDR block and port are configured.

    **Figure 8-2 Checking whether the CIDR block and port are configured**

    ![Checking whether the CIDR block and port are configured](image)
If no, add it.

**Figure 8-3 Adding an inbound rule**

- **Outbound rules**: Check whether the CIDR block and port of the outbound rule are enabled. (It is recommended that all CIDR blocks be enabled.)

  Check whether the CIDR block and port are configured.

  **Figure 8-4 Checking whether the CIDR block and port are configured**

  If no, add it.

  **Figure 8-5 Adding an outbound rule**

- **Enhanced datasource connections**:
  Check whether the CIDR block corresponding to the DLI queue is open. If no, add an outbound CIDR block in the VPC.

  1) Find the CIDR block corresponding to the queue bound to the datasource connection in DLI.
2) In the VPC security group, check whether the CIDR block corresponding to the DLI queue has been configured.

If no, add it.

If the fault persists, contact Huawei technical support.
9 Problems Related to Flink Jobs

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

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

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

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

To rectify this fault, perform the following steps:

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 Flink job are located in the same region.

9.3 How Can I Check Job Results?

- DLI can output stream 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.
- DLI can output stream job results to RDS. You can view the results in RDS. For detailed operations, see Relational Database Service Getting Started Guide.
• DLI can output stream job results to SMN, and SMN sends the results to the user's terminal. For details, see Simple Message Notification Getting Started Guide.

• DLI can output stream job results to Kafka. You can view the results in Kafka clusters. For details, visit the Kafka official website.

• DLI can output stream job results to CloudTable. You can view the results in CloudTable. For detailed operations, see Getting Started in the CloudTable Service User Guide.

• DLI can export stream job results to CSS. You can view the results in CSS. For details, see Getting Started in the Cloud Search Service User Guide.

• DLI can export stream job results to DCS. You can view the results in DCS. For details, see Distributed Cache Service Getting Started Guide.

9.4 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 DLI management console, click Job Management < Flink Jobs.

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 Parameters on the Edit page.

1. Select Enable Checkpointing or Save Job Log.

2. Specify OBS Bucket.

3. Select Authorize OBS.

----End

9.5 How Do I Write Data to Different Elasticsearch Clusters in a Flink Job?

Add the following SQL statements to the Flink job:

create source stream ssource(xx);
create sink stream es1(xx) with (xx);
cREATE sink stream es2(xx) with (xx);
insert into es1 select * from ssource;
insert into es2 select * from ssource;

9.6 Does Data in the Temporary Stream of Flink SQL Need to Be Cleared Periodically? How Do I Clear the Data?

The temp stream in Flink SQL is similar to a subquery. It is a logical stream used to simplify the SQL logic and does not generate data storage. Therefore, there is no need to clean data.
9.7 How Does a DLI Flink Job Interact with Data Sources on the Public Network?

The running environment of DLI queues cannot directly communicate with the public network environment. Therefore, DLI queues cannot directly interact with data sources on the public network. To connect to a public network or an offline data source, you can use an elastic IP address (EIP) or use Direct Connect to connect to a VPC instance on HUAWEI CLOUD, and then use the enhanced datasource function of DLI to enable the network link between the VPC of the DLI queue and the VPC instance.

9.8 Does a User-Defined Flink Job Support Configuration File Upload? How Do I Upload a Configuration File?

Configuration files can be uploaded for user-defined jobs (JAR).

1. Upload the configuration file to DLI through Package Management.
2. In the Other Dependencies area of the Flink JAR job, select the created DLI package.
3. Load the file through `ClassName.class.getClassLoader().getResource("userData/fileName")` in the code. In the file name, `fileName` indicates the name of the file to be accessed, and `ClassName` indicates the name of the class that needs to access the file.

9.9 What Should I Do If Logs Are Not Written to the OBS Bucket After a DLI Flink Job Fails to Be Submitted for Running?

Mode for storing generated job logs when a DLI Flink job fails to be submitted or executed. The options are as follows:

- If the submission fails, a submission log is generated only in the `submit-client` directory.
- You can view the logs generated within 1 minute when the job fails to be executed on the management console. Choose Job Management > Flink Jobs, click the target job name to go to the job details page, and click Run Log to view real-time logs.
- If the running fails and exceeds 1 minute (the log dump period is 1 minute), run logs are generated in the `application_xx` directory.

Flink dependencies have been built in the DLI server and security hardening has been performed based on the open-source community version. To avoid dependency package compatibility issues or log output and dump issues, be careful to exclude the following files when packaging:
- Built-in dependencies (or set the package dependency scope to "provided")
- Log configuration files (example, log4j.properties/logback.xml)
- JAR file for log output implementation (example, log4j).

On this basis, the `taskmanager.log` file rolls as the log file size and time change.
10 Problems Related to APIs

10.1 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 to 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 with and log in to the HUAWEI 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 Create Access Key. The Create Access Key dialog box is displayed.
5. On the Create Access Key page, enter the 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.

10.2 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. Register with and log in to the HUAWEI 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.
   On the displayed **My Credentials** page, view the project ID on the **Projects** page, View project IDs. For example, 5a3314075bfa49b9ae360f4ecd333695.