

WeData Data Development Platform

FAQs



Copyright Notice

©2013–2025 Tencent Cloud. All rights reserved.

The complete copyright of this document, including all text, data, images, and other content, is solely and exclusively owned by Tencent Cloud Computing (Beijing) Co., Ltd. ("Tencent Cloud"); Without prior explicit written permission from Tencent Cloud, no entity shall reproduce, modify, use, plagiarize, or disseminate the entire or partial content of this document in any form. Such actions constitute an infringement of Tencent Cloud's copyright, and Tencent Cloud will take legal measures to pursue liability under the applicable laws.

Trademark Notice

 Tencent Cloud

This trademark and its related service trademarks are owned by Tencent Cloud Computing (Beijing) Co., Ltd. and its affiliated companies("Tencent Cloud"). The trademarks of third parties mentioned in this document are the property of their respective owners under the applicable laws. Without the written permission of Tencent Cloud and the relevant trademark rights owners, no entity shall use, reproduce, modify, disseminate, or copy the trademarks as mentioned above in any way. Any such actions will constitute an infringement of Tencent Cloud's and the relevant owners' trademark rights, and Tencent Cloud will take legal measures to pursue liability under the applicable laws.

Service Notice

This document provides an overview of the as-is details of Tencent Cloud's products and services in their entirety or part. The descriptions of certain products and services may be subject to adjustments from time to time.

The commercial contract concluded by you and Tencent Cloud will provide the specific types of Tencent Cloud products and services you purchase and the service standards. Unless otherwise agreed upon by both parties, Tencent Cloud does not make any explicit or implied commitments or warranties regarding the content of this document.

Contact Us

We are committed to providing personalized pre-sales consultation and technical after-sale support. Don't hesitate to contact us at 4009100100 or 95716 for any inquiries or concerns.

Contents

FAQs

Basics

Usage

Data Integration

Data Development

Data Operations

FAQs

Basics

Last updated: 2024-08-23 17:35:27

What is Tencent Cloud WeData?

Tencent Cloud WeData is a one-stop data collaboration development governance platform created by Tencent Cloud.

What are the common scenarios of WeData?

Common scenarios of WeData include enterprise data warehouse construction, collaborative data visualization development, data asset management, and data governance. For more details, see [Scenarios](#).

What product capabilities does WeData provide?

WeData offers full-link data processing capabilities including ad-hoc analysis, workflow collaborative orchestration, data asset management, and data governance.

WeData Product Architecture and Advantages?

As a big data development and governance platform, WeData aims to provide open source, easy to use, and secure and stable product services. The overall architecture consists of an operational management system and development and operation tools. For details, please refer to [Product Architecture](#).

What types of data sources does WeData support?

Supports various on-cloud and self-built data sources, such as:

- Relational Database: MySQL, PostgreSQL, Oracle, SQL Server, IBM DB2, Dameng DM, SAP HANA, etc.
- Big Data Storage: Hive, HDFS, HBase, Kudu, Clickhouse, DLC, Impala, Gbase, Tbase, etc.
- Semi-structured: FTP, SFTP, COS.
- NoSQL: Redis, Elasticsearch.
- Message Queue: Kafka, etc.

Does WeData support private deployment?

WeData supports private deployment. You can [Submit a ticket](#) to contact us.

Usage

Last updated: 2024-08-24 18:08:06

Which sub-accounts can authorize CAM roles?

Policy name	Policy Description	Function
AdministratorAccess	This policy allows you to manage all users and their permissions, financial-related information, and cloud service assets within the account	<ul style="list-style-type: none"> Can manage sub-account permissions Can log in to WeData Console Can create WeData Project
QCloudResourceFullAccess	This policy allows you to manage all cloud service assets within the account	<ul style="list-style-type: none"> Can log in to WeData Console Can create WeData Project
ReadOnlyAccess	This policy allows you read-only access to all cloud service assets within the account that support interface-level or resource-level authentication	<ul style="list-style-type: none"> Can log in to WeData Console Cannot create WeData Project

How to obtain/view EMR VPC?

Go to Tencent Cloud [EMR Console](#), select the region and ID corresponding to the cluster. Enter the **Cluster Detail Page > Basic Information > Instance information** page to view the cluster network information.

The screenshot shows the '实例信息' (Instance Information) page in the EMR console. Under the '基础配置' (Basic Configuration) section, the '网络信息' (Network Information) field is highlighted with a red box. The value is 'wedat---oc-weda', where 'wedat' is the ID and 'oc-weda' is the display name, separated by '---'. Other fields include '实例ID' (Instance ID), '地域信息' (Region Information), '节点高可用' (Node High Availability), '安全组' (Security Group), '云硬盘加密' (Cloud Disk Encryption), '创建时间' (Creation Time), 'Master公网IP' (Master Public IP), '计费模式' (Billing Mode), and '对象存储' (Object Storage).

Note

The network information of EMR by default uses "---" to separate the ID and display name. You can copy the information before "---" (inside the red box) for search.

How to obtain/view the data source VPC?

Taking MySQL as an example, go to the corresponding Tencent Cloud MySQL Console, select the region and ID of the corresponding instance. Enter the **Instance Detail Page > Instance Details** page, and view the instance network information.



When synchronizing the whole MySQL database, click "Pause" immediately after running, the task status always shows "In operation"?

Cause Analysis: After the whole database synchronization task starts running, clicking **Pause** immediately will cause the task status to remain "In operation". Backend data is continuously synchronizing, far exceeding the actual size of the source table.

Solution: When the MySQL whole database synchronization task is in the snapshot phase, the "Pause" operation is not supported.

After clicking "Continue" on a "Paused" task, some data is duplicated?

Cause Analysis: In the current version, resuming a real-time task after it was paused will start from the last checkpoint, so partial data duplication is a normal occurrence.

During real-time synchronization, after deleting the data source or changing the data source structure, the task shows "Running" but is not actually synchronizing?

Cause Analysis: During real-time task operation, if the following scenarios occur, the task still shows as running but is not actually synchronizing any data.

- Deleting the data source in the MySQL source database. For example, during synchronization from MySQL to Hive, deleting the MySQL or Hive table in the MySQL source database.
- Changing the data source structure in the MySQL source database. For example, during synchronization from MySQL to Hive, deleting or modifying certain fields in the MySQL or Hive table in the MySQL source database.

Both operations of deleting the data source and changing data structures are considered abnormal issues. WeData temporarily cannot obtain the abnormal status of the task.

Solution: After obtaining the abnormal logs, the solutions for the above two situations are as follows:

1. Delete the data source. Users need to reselect a new data source on the WeData real-time synchronization interface, configure the mapping relationship, and then click **submit**, choosing "Keep job status data, continue running".
2. Changing the data source structure. Users need to click **Refresh** in the WeData Real-time Synchronization Interface to obtain the latest data source structure, configure the mapping relationship, then click **Submit**, and choose "Retain Job Status Data, Continue Running".

When synchronizing MySQL sharded tables with different schemas, the task shows "Running" but the task is not successfully synchronized?

Cause Analysis: When synchronizing MySQL sharded tables, if the schemas of the configured tables are different and the fields establishing the mapping relationship do not exist in certain tables, the task status will show "Running" but will not actually execute successfully. Additionally, since Oceanus currently uses a strategy of continuous restarting when a task encounters an error, the task remains in a "Running" state.

Solution: For the unsuccessful scenario mentioned above, exclude non-compliant tables in the WeData Real-time Synchronization Interface, configure the mapping relationship, then click **Submit**, and choose "Retain Job Status Data, Continue Running".

When performing MySQL whole database synchronization, the "Oceanus job config length exceeds 100k limit" warning appears?

Problem Description: When performing MySQL whole database synchronization, the "Oceanus job config length exceeds 100k limit" warning appears.

Cause Analysis: To ensure the performance of the whole database migration task, the total length of the task configuration cannot exceed 100k in the current version. If too many data tables or the total number of table fields are too large, the system prompts a limit warning.

Solution: It is recommended that a single whole database migration task does not configure more than 20 tables. It can be divided into multiple whole database migrations or sharded tasks.

Does the Definition field configuration support Complex Types?

The current version does not support Complex Types such as Array, Map, and Struct, only Primitive Types.

An error occurs when submitting the task: "ache.calcite.sql.validate.SqlValidatorException: Cast function cannot convert value of type ..."?

Problem Description: When submitting a real-time synchronization task, the system reports the error "[TencentCloudSDKError] Code=FailedOperation, Message=ache.calcite.sql.validate.SqlValidatorException: Cast function cannot convert value of type ...".

Cause Analysis: The types between the source and target fields of real-time synchronization cannot be converted.

Solution:

1. Please check whether the field types configured between the source and target ends match. For specific type conversions between data sources and target ends, refer to [Synchronized Data Source](#).
2. If the task is a MySQL-Ckafka single table synchronization task, check whether the mapping relationship for pkNames_wedata_di and sqlType_wedata_di with the target end is established. Since the current version does not support Complex Types such as Array, Map, and Struct in Ckafka fields, the pkNames_wedata_di and sqlType_wedata_di parameters cannot be written to the Ckafka target end.

Sub-account prompts "No PassRole" when redirected to Tencent Cloud Oceanus Console through operation and maintenance?



Solution: Please refer to the [Oceanus Permission Management](#) document.

Data Integration

Last updated: 2024-11-25 09:31:41

1. How to handle ClickHouse to DLC sync reporting network timeout

Problem description:

When users sync ClickHouse data to DLC, it succeeds with small datasets, but reports network timeout with larger tables.

```
at com.***.***.plugin.rdbms.util.DBUtil.query(DBUtil.java:470)
at com.***.***.plugin.rdbms.util.DBUtil.query(DBUtil.java:431)
at com.***.***.plugin.rdbms.reader.CommonRdbmsReader$Task.startRead(CommonRdbmsReader.java:196)
at com.***.***.plugin.reader.clickhousereader.ClickhouseReader$Task.startRead(ClickhouseReader.java:72)
at com.***.***.core.taskgroup.runner.ReaderRunner.run(ReaderRunner.java:59)
at java.lang.Thread.run(Thread.java:748)
[2022-04-14 21:16:08]-[INFO] Total 0 records, 0 bytes | Speed 0B/s, 0 records/s | Error 0 records, 0 bytes
[2022-04-14 21:16:08]-[INFO]
```

该任务最可能的错误原因是：

socketTimeoutException: Read timed out

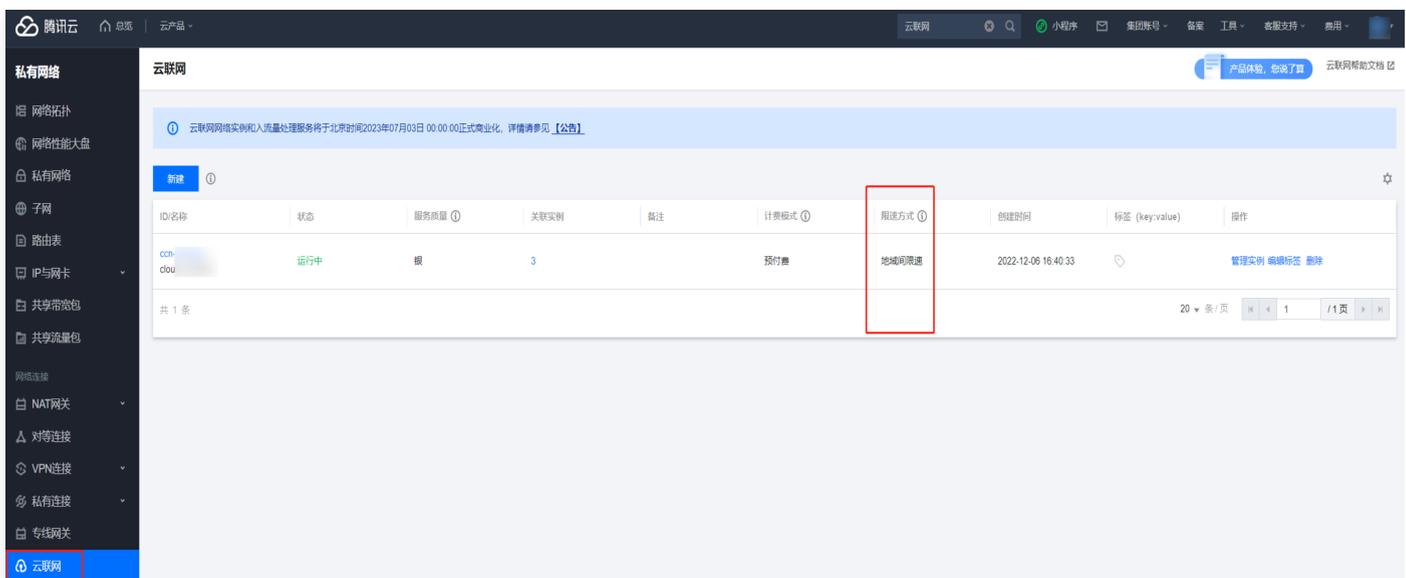
```
[2022-04-14 21:16:08]-[ERROR] Task execution failed
```

```
[2022-04-14 21:16:08]-[INFO] commitTask success.
```

```
[2022-04-14 21:16:08]-[INFO] taskId: <20[REDACTED]2>
```

Locating the Issue:

1. Logging into the execution environment POD, Telnet to the target ClickHouse datasource connection shows connectivity, and logging in is also successful, thus ruling out network issues.
2. It is suspected that large ClickHouse data volume or high concurrency leads to read timeout. Further reduce concurrency in [CCN](#).



- If the same error persists after limiting concurrency and increasing the data read operation timeout in ClickHouse and still encounter errors, further investigation is required to ensure that the customer's execution environment resource group and ClickHouse VPC are consistent. If not, the network must be connected via **CCN** (Cloud Connect Network).
- Confirm the required public network bandwidth for cross-regional network connectivity with the CCN product, and verify whether the user has connected the two VPCs across regions via CCN but without enabling public network bandwidth.



- For testing purposes, CCN defaults to 10Kb of traffic, explaining why small data syncs and network tests are normal, but large data volumes report network issues.



6. Cause:

If the execution resource group and datasource are not in the same VPC and cross-region VPC connection via CCN is without enabling public network bandwidth, default CCN supports less than

10Kb of traffic, exceeding which causes a disconnection, making small data volumes feasible but large data volumes fail.

7. Solution:

Solve the issue by adding public network bandwidth via CCN.

2. How to handle issues when requiring public network access but network is disconnected

Problem description:

```

        "connection": {
          "jdbcUrl": "jdbc:mysql://gz-cdb-bwejl37.sql.tencentcdb.com:57198/?rewriteBatchedStatements=true&tinyInt1is0it=false&serverTimezone=Asia/Shanghai&characterEncoding=utf8",
          "table": "sor_siteinfo",
          "sor_siteinfo": {
            "writeMode": "replace",
            "batchSize": 1024,
            "username": "root",
            "password": "123456",
            "preSql": "truncate table sor_siteinfo"
          }
        },
        "name": "mysqlwriter"
      },
      "setting": {
        "errorLimit": {
          "record": 0
        },
        "speed": {
          "byte": -1,
          "channel": 1
        }
      }
    }
  }
}
}

2022-06-05 16:05:15] [WARN] priority set to 0, because NumberFormatException, the value is: null
2022-06-05 16:05:15] [INFO] PerfTrace traceId=job_0, isEnabled=false, priority=0
2022-06-05 16:05:15] [INFO] JobContainer starts job.
2022-06-05 16:05:15] [INFO] Driver registered
2022-06-05 16:05:15] [INFO] register random thin driver
2022-06-05 16:05:15] [INFO] Supplied authorities: 172.17.147.237:7001,172.17.147.136:7001
2022-06-05 16:05:15] [INFO] Resolved authority: 172.17.147.237:7001
2022-06-05 16:05:15] [INFO] Will try to open Client transport with JDBC Uri: jdbc:hive2://172.17.147.237:7001;serverTimezone=Asia/Shanghai;tinyInt1is0it=false
2022-06-05 16:05:16] [INFO] Available jdbcUrl[jdbc:hive2://172.17.147.237:7001;serverTimezone=Asia/Shanghai;tinyInt1is0it=false]
2022-06-05 16:05:16] [ERROR] Exception when calling callable, 异常Msg:[Code: MySQLErrCode-02], Description: [数据库服务的IP地址或者Port错误, 请检查填写的IP地址和Port或者联系DBA确认IP地址和Port是否正确。如果是同步中心用户请联系DBA确认idb上录入的IP和PORT信息和数据库的当前实际信息是一致的]。 - 具体错误信息为: com.mysql.jdbc.exceptions.CommunicationsException: Communications link failure

he last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.
2022-06-05 16:06:17] [ERROR] Exception when calling callable, 即将尝试执行第1次重试,本次重试计划等待 [1000]ms,实际等待 [1000]ms, 异常Msg:[Code: MySQLErrCode-02], Description: [数据库服务的IP地址或者Port错误, 请检查填写的IP地址和Port或者联系DBA确认IP地址和Port是否正确。如果是同步中心用户请联系DBA确认idb上录入的IP和PORT信息和数据库的当前实际信息是一致的]。 - 具体错误信息为: com.mysql.jdbc.exceptions.CommunicationsException: Communications link failure

he last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.
2022-06-05 16:07:19] [ERROR] Exception when calling callable, 即将尝试执行第2次重试,本次重试计划等待 [2000]ms,实际等待 [2001]ms, 异常Msg:[Code: MySQLErrCode-02], Description: [数据库服务的IP地址或者Port错误, 请检查填写的IP地址和Port或者联系DBA确认IP地址和Port是否正确。如果是同步中心用户请联系DBA确认idb上录入的IP和PORT信息和数据库的当前实际信息是一致的]。 - 具体错误信息为: com.mysql.jdbc.exceptions.CommunicationsException: Communications link failure

he last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.
2022-06-05 16:08:23] [ERROR] Exception when calling callable, 即将尝试执行第3次重试,本次重试计划等待 [4000]ms,实际等待 [4000]ms, 异常Msg:[Code: MySQLErrCode-02], Description: [数据库服务的IP地址或者Port错误, 请检查填写的IP地址和Port或者联系DBA确认IP地址和Port是否正确。如果是同步中心用户请联系DBA确认idb上录入的IP和PORT信息和数据库的当前实际信息是一致的]。 - 具体错误信息为: com.mysql.jdbc.exceptions.CommunicationsException: Communications link failure

he last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.
2022-06-05 16:09:21] [ERROR] Exception when calling callable, 即将尝试执行第4次重试,本次重试计划等待 [8000]ms,实际等待 [8000]ms, 异常Msg:[Code: MySQLErrCode-02], Description: [数据库服务的IP地址或者Port错误, 请检查填写的IP地址和Port或者联系DBA确认IP地址和Port是否正确。如果是同步中心用户请联系DBA确认idb上录入的IP和PORT信息和数据库的当前实际信息是一致的]。 - 具体错误信息为: com.mysql.jdbc.exceptions.CommunicationsException: Communications link failure

he last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.

```

Locating the Issue:

Locate through logs.

Cause:

The data source uses a public network, and the resource group's network does not default to public.

Solution:

Add the subnet of the resource group to the NAT Gateway. For details, refer to [NAT Gateway Configuration](#) and [Access the Internet through NAT Gateway](#) .

Note

If it is the main account's resource group, you can replace the resource group's subnet [routing policy](#) with EKS.

3. How to handle data synchronization failure when it prompts that the data source is not accessible

Problem description:

Offline synchronization prompts that the database cannot be connected, but the database is actually accessible.

```

运行日志
r.init(JobContainer.java:306)
r.start(JobContainer.java:114)
gine.java:232)
gine.java:122)
s.wedata.wedataHelper.startwedata(wedataHelper.java:126)
s.WeDataDiRunner.execute(WeDataDiRunner.java:161)
api.AbstractTaskRunner.executeWithCommonLog(AbstractTaskRunner.java:509)
api.AbstractTaskRunner.startWork(AbstractTaskRunner.java:547)
s.WeDataDiRunner.main(WeDataDiRunner.java:75)
of [jdbc:hive2://[redacted]/ola_feather;serverTimezone=Asia/Shanghai;tinyInt1isBit=false] failed, for Code: [DBUtilErrorCode-10], Description: [连接数
calling callable, 异常Msg:无法连接对应的数据库, 可能原因是: 1) 配置的ip/port/database/idbc错误 无法连接. 2) 配置的username/pass**rd*****
calling callable, 即将尝试执行第1次重试.本次重试计划等待[1000]ms,实际等待[1000]ms, 异常Msg:[无法连接对应的数据库, 可能原因是: 1) 配置的ip/port/database/jdbc错误, 无法连接.]
ities: [redacted]
ity: [redacted]

```

Locating the Issue:

The network of the WeData DataInLong resource group and CDB are not under the same VPC and are not interconnected.

Cause:

The DataInLong execution resource group and data source are under the user's VPC network. They need to be in the same VPC network, otherwise, the network will not be accessible, and the execution resource group cannot synchronize data correctly. If they are in different VPC networks.

Solution:

Use [CCN](#) or [Peering Connection](#) to enable cross-VPC network intercommunication. If the data source is a public network instance, configure the [NAT Gateway](#).

4. How to handle Hive On COS table permission error during data write

Problem description:

Offline synchronization, hive data source, table data stored in COS, error during data write:

java.lang.Exception: Retrieve the file metadata file failure.

```

e', isPartition='false', isBucket='false'], Column[name='app_version', datatype='string', nullable='true', isPartition='false', isBucket='false'], Column[name='remark', dataTy
pe='string', nullable='true', isPartition='false', isBucket='false']]
[2022-08-26 10:37:58]-[ERROR] 判断文件路径 [message:filePath =cosn://dev-data-lake-1306106726/user/hive/warehouse/ods_tsp_login_history_di_tmp1661481478536]是否存在时发生网络IO
异常,请检查您的网络是否正常!
[2022-08-26 10:37:58]-[ERROR] Exception when job run: java.io.IOException: cosn://dev-data-lake-1306106726/user/hive/warehouse/ods_tsp_login_history_di_tmp1661481478536 : java
.lang.Exception: Retrieve the file metadata file failure. cos key: /user/hive/warehouse/ods_tsp_login_history_di_tmp1661481478536, exception: com.qcloud.cos.exception.CosServi
ceException: Forbidden (Status Code: 403; Error Code: 403 Forbidden; Request ID: NjMwODMyMDZlMzFkMjc2MmVfYmYyNF8zNDA0MTdm); Trace ID: OGVmYzZiMmQzYjA2OWNhODk0NTRkMTBjOWVmdAxO
Dc00WRkZjk0ZDM1NmIIM2E2MTRlY2MzZDhmNmI5MmI1OTBjYzE2MjAxN2M1MzJiOTdkZjMxMDVlYTZjN2FiNmI0ZTRkY2MzNGMxNzE2YTQxZjhkNTAxM2ZmMGwIwYzVjZDM=
[wrapped] com.tencent.wedata.common.exception.wedataException: Code: [HiveWriter-06], Description: [与HDFS建立连接时出现IO异常.]. - java.io.IOException: cosn://dev-data-lake-13
06106726/user/hive/warehouse/ods_tsp_login_history_di_tmp1661481478536 : java.lang.Exception: Retrieve the file metadata file failure. cos key: /user/hive/warehouse/ods_tsp_lo
gin_history_di_tmp1661481478536, exception: com.qcloud.cos.exception.CosServiceException: Forbidden (Status Code: 403; Error Code: 403 Forbidden; Request ID: NjMwODMyMDZlMzFkMjc2MmVfYmYyNF8zNDA0MTdm); Trace ID: OGVmYzZiMmQzYjA2OWNhODk0NTRkMTBjOWVmdAxO
Dc00WRkZjk0ZDM1NmIIM2E2MTRlY2MzZDhmNmI5MmI1OTBjYzE2MjAxN2M1MzJiOTdkZjMxMDVlYTZjN2FiNmI0ZTRkY2MzNGMxNzE2YTQxZjhkNTAxM2ZmMGwIwYzVjZDM=
at org.apache.hadoop.fs.CosNativeFileSystemStore.handleException(CosNativeFileSystemStore.java:1056)
at org.apache.hadoop.fs.CosNativeFileSystemStore.queryObjectMetadata(CosNativeFileSystemStore.java:524)
at org.apache.hadoop.fs.CosNativeFileSystemStore.retrieveMetadata(CosNativeFileSystemStore.java:544)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:422)
at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeMethod(RetryInvocationHandler.java:165)
at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invoke(RetryInvocationHandler.java:157)
at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeOnce(RetryInvocationHandler.java:95)
at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:359)

```

Locating the Issue:

Check if the account is associated with the WeData_QCSRole role and if this role has the COS CAM policy configured.

Cause:

When integrating tasks for hive on cos, write data through the WeData_QCSRole role to obtain temporary credentials from CAM. If the COS-related CAM policy is not added under the WeData_QCSRole role, the obtained temporary credentials will not have permission to read/write the COS bucket.

Solution:

On the role page of CAM, search for "wedata", find WeData_QCSRole, and check if it includes COS in the associated policies. If not, add QcloudCOSFullAccess.

访问管理

- ☰ 概览
- 👤 用户
- 👤 用户组
- 📄 策略
- 👤 角色
- 👤 身份提供商
- 👤 联合账号
- 🔑 访问密钥

角色信息

角色名称 WeData_Q

RoleArn qcs::cam::uin-[redacted]:WeData_QCSRole

角色ID 461-[redacted]

角色描述 当前角色为 WeData 数据开发平台 服务角色，该角色将在已关联策略的权限范围内访问您的其他

创建时间 2021-11-22 22:37:03

标签 暂无标签 ✎

权限 角色载体 (1) 撤销会话 服务

▼ 权限策略

关联策略以获取策略包含的操作权限。解除策略将失去策略包含的操作权限。

关联策略
批量解除策略

COS
✕ 🔍

<input type="checkbox"/> 策略名	描述
<input type="checkbox"/> QcloudCOSBucketConfigWrite	对象存储 (COS) 修改存储桶配置的操作权限
<input type="checkbox"/> QcloudCOSBucketConfigRead	对象存储 (COS) 读取存储桶配置的操作权限
<input type="checkbox"/> QcloudCOSListOnly	对象存储 (COS) 获取存储桶、对象列表的操作权限
<input type="checkbox"/> QcloudCOSDataReadOnly	对象存储 (COS) 数据只读的访问权限
<input type="checkbox"/> QcloudCOSDataWriteOnly	对象存储 (COS) 数据只写的访问权限
<input type="checkbox"/> QcloudCOSDataFullControl	对象存储 (COS) 数据读、写、删除、

Data Development

Last updated: 2025-04-09 14:28:37

Unable to Select COS Bucket for Resource Upload

1. issue phenomenon

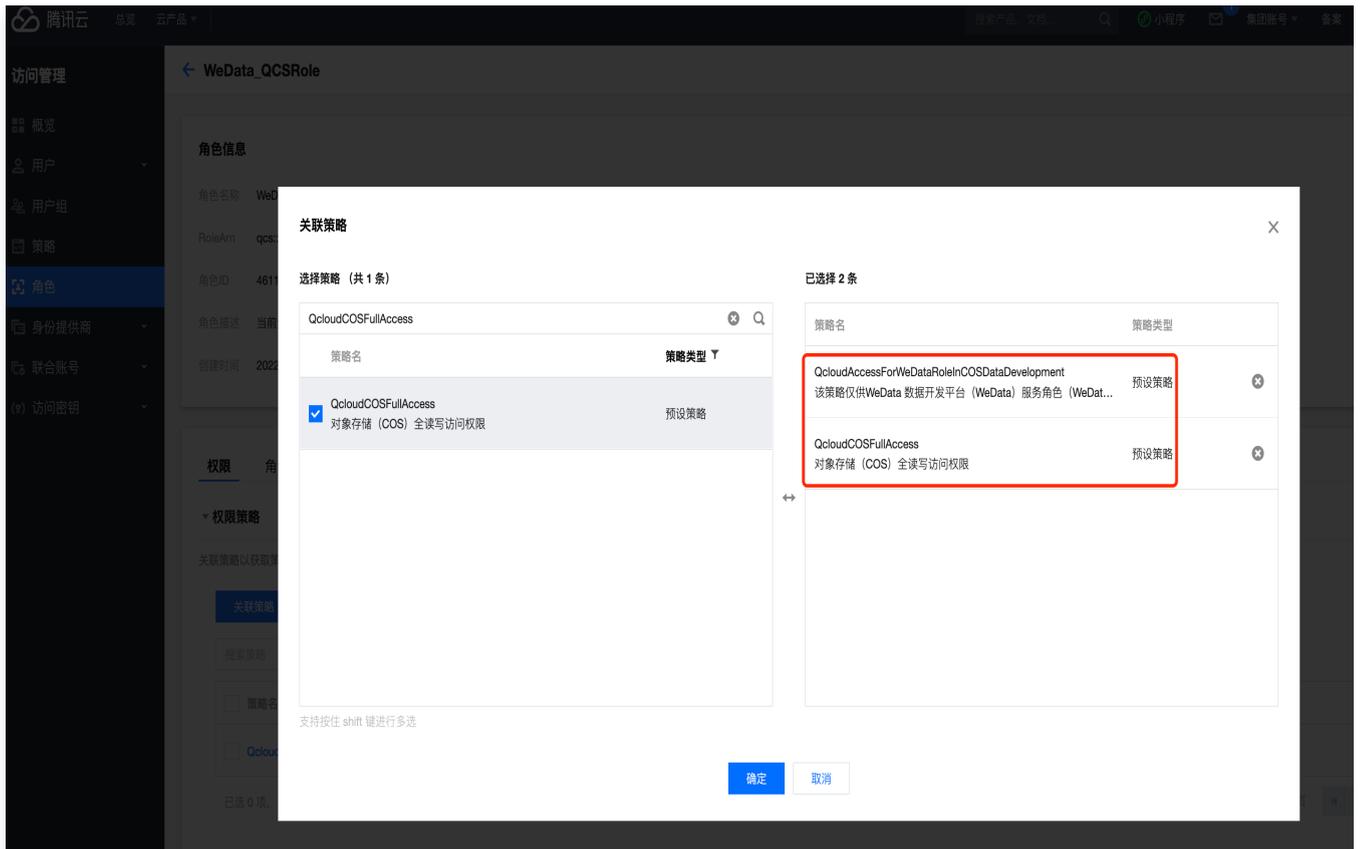


2. Solution

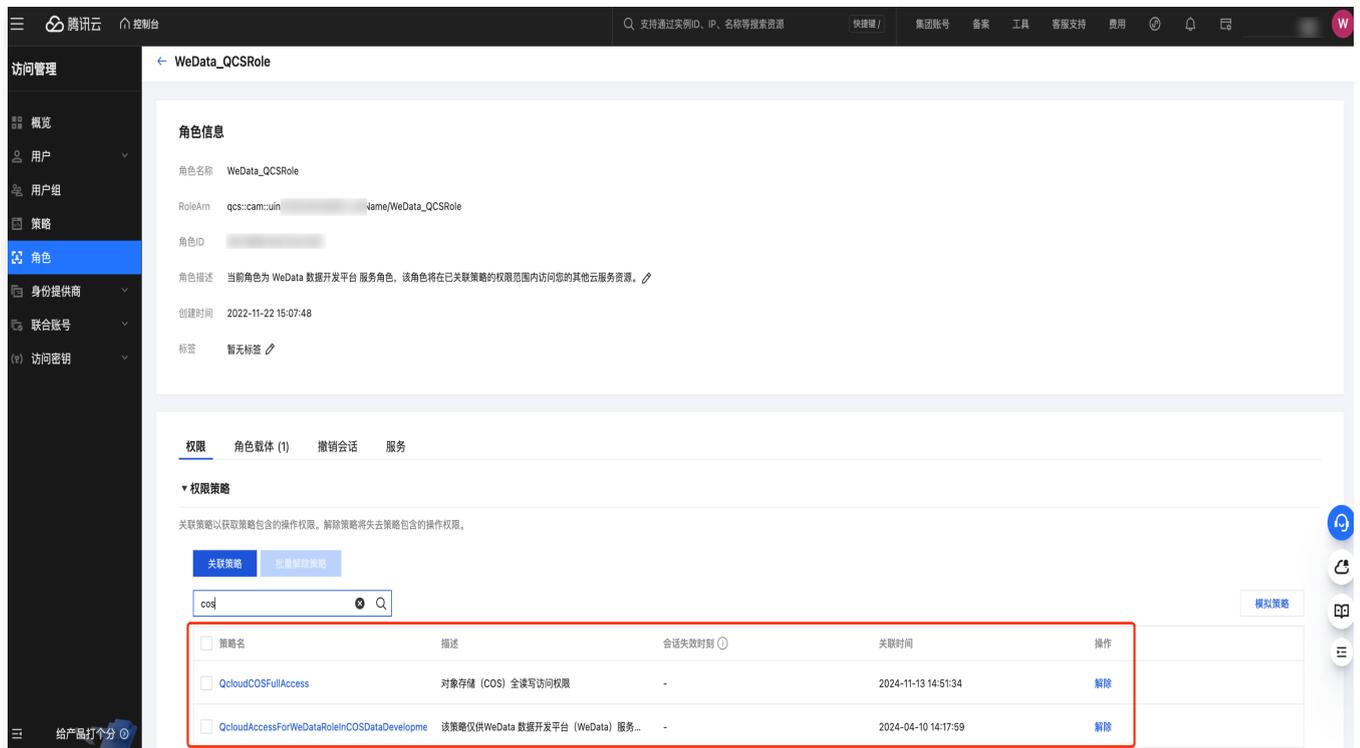
2.1 Enter Tencent Cloud [Cloud Access Management](#), select the **role** menu, and search for **WeData_QCSRole**.



2.2 1. Click the role name **WeData_QCSRole**, select the associated policy, search for policies "QcloudAccessForWeDataRoleInCOSDataDevelopment" and "QcloudCOSFullAccess", and confirm the association.



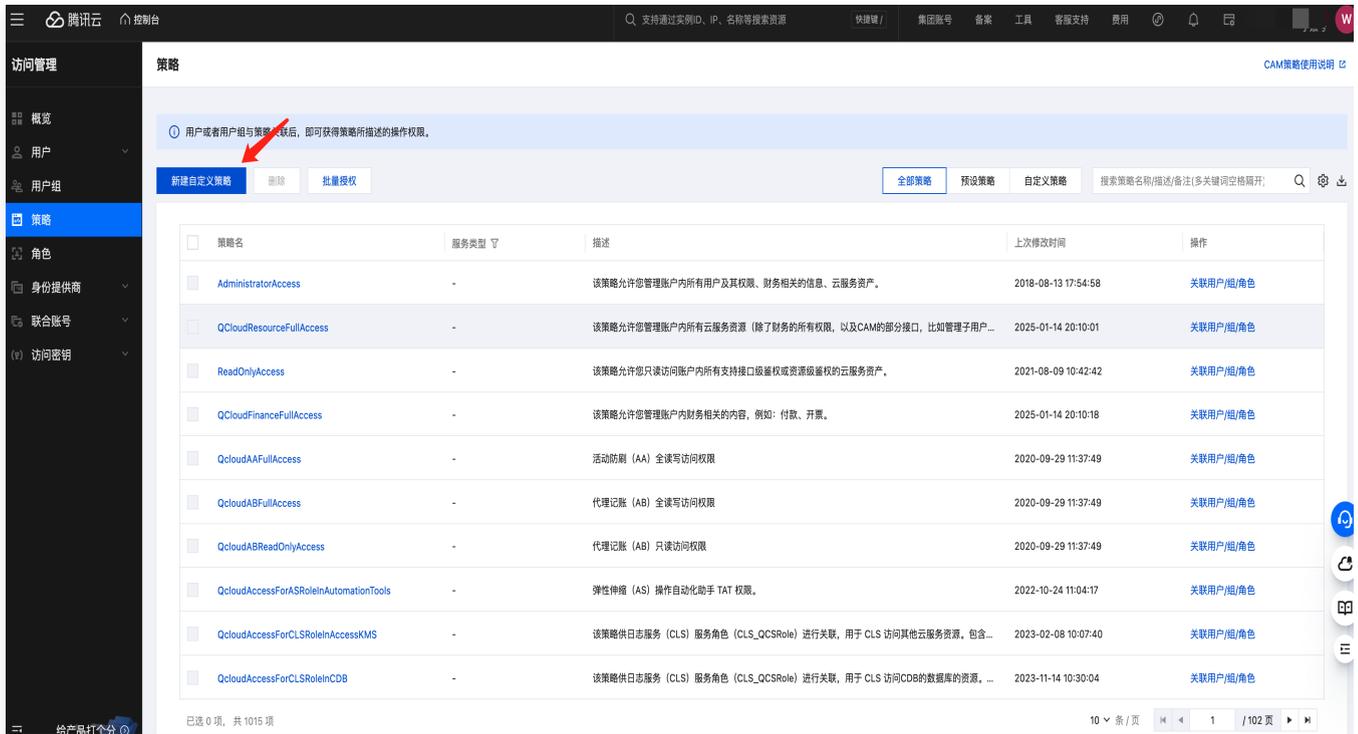
After successful association, this policy can be seen under the permission policy. Try uploading resources to the COS bucket again and it will succeed.



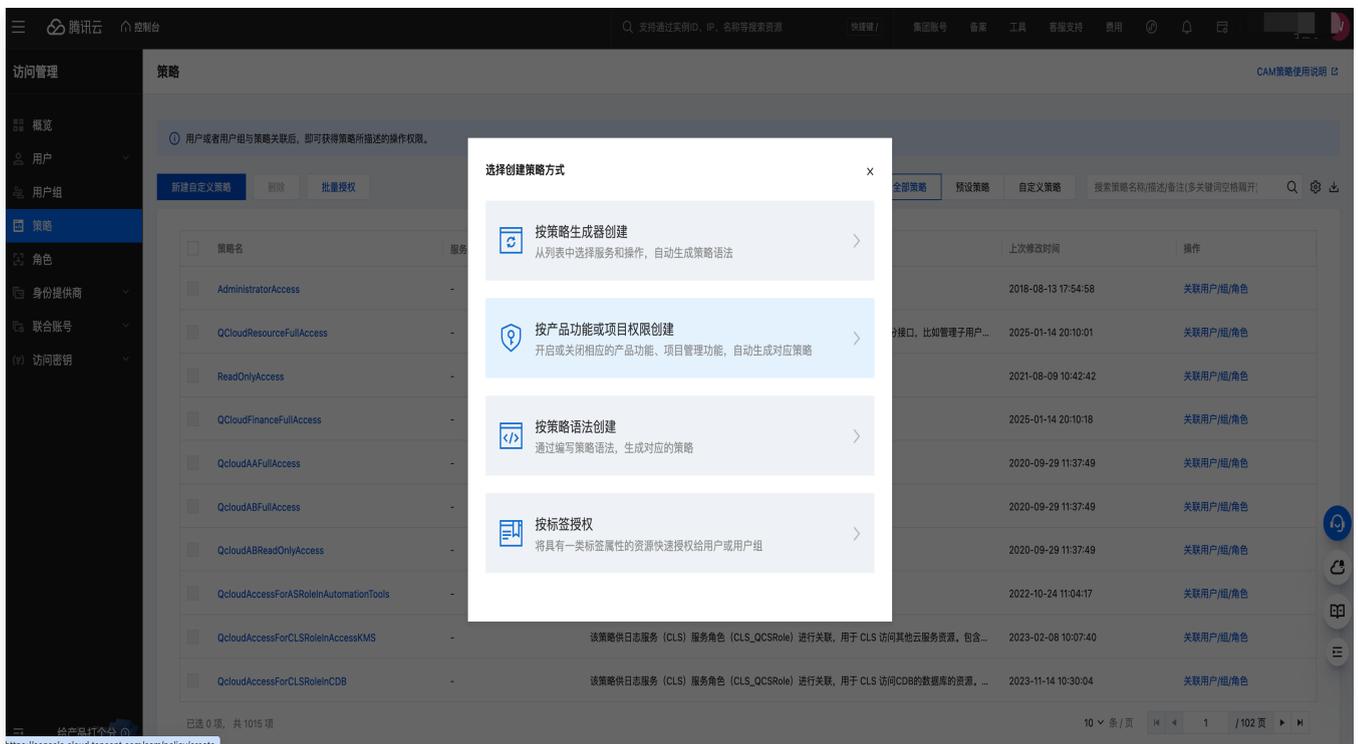
3. Supplementary solution

If you are concerned that the excessive permissions of COSFullAccess granted in the previous step, you can see the following operations to grant WeData only read access to a certain bucket specifically used for development.

3.1 Create a custom policy



Select create by policy generator.



3.2 Generate a COS Policy

The focus is on limiting the specified cos resource, simultaneously expanding the permission of the drop-down list, the location of specific resources, and filling in the COS bucket provided to

WeData Development Engineer. Fill in other content by referring to the screenshot.

按策略生成器创建

1 编辑策略 > 2 关联用户/用户组/角色

可视化策略生成器 JSON

对象存储 (全部操作)

效果 (Effect) 允许 拒绝

服务 (Service) 对象存储 (cos)

操作 (Action) 全部操作 (*)

资源 (Resource) 全部资源 特定资源

不拆分资源级和操作级接口

[添加自定义资源六段式 来限制访问](#)

条件 (Condition) 来源 IP 添加其他条件

+ 添加权限

下一步 字符数: 160 (最多6144)

添加资源六段式

资源六段式 用于唯一描述腾讯云的资源对象

qcs::cos::uid/1315051789/

服务: cos

地域: 所有地域

账户: uid/1315051789

资源前缀:

资源:

确定

The completed Summary Strategy Map is as follows:

访问管理

- 概览
- 用户
- 用户组
- 策略**
- 角色
- 身份提供商
- 联合账号
- 访问密钥

← 按策略生成器创建

1 编辑策略 > 2 关联用户/用户组/角色

可视化策略生成器 JSON

▼ 对象存储 (全部操作)

效果 (Effect)	<input checked="" type="radio"/> 允许 <input type="radio"/> 拒绝
服务 (Service)	对象存储 (cos)
操作 (Action)	全部操作 (*)
资源 (Resource)	qcs::cos::uid/1258756906:wedatabeijing-1258756906/*
条件 (Condition)	<input type="checkbox"/> 来源 IP ⓘ 添加其他条件

▼ 对象存储 (1 个操作)

效果 (Effect)	<input checked="" type="radio"/> 允许 <input type="radio"/> 拒绝
服务 (Service)	对象存储 (cos)
操作 (Action)	读操作 编辑
	GetService 拉取存储桶列表
资源 (Resource)	<input checked="" type="radio"/> 全部资源 <input type="radio"/> 特定资源 收起
条件 (Condition)	<input type="checkbox"/> 来源 IP ⓘ 添加其他条件

[+ 添加权限](#)

[下一步](#) 字符数: 338 (最多6144)

3.3 Save policy, associated with the role of WeData.

访问管理

- 概览
- 用户
- 用户组
- 策略**
- 角色
- 身份提供商
- 联合账号
- 访问密钥

← 按策略生成器创建

1 编辑策略 > 2 关联用户/用户组/角色

基本信息

策略名称

描述

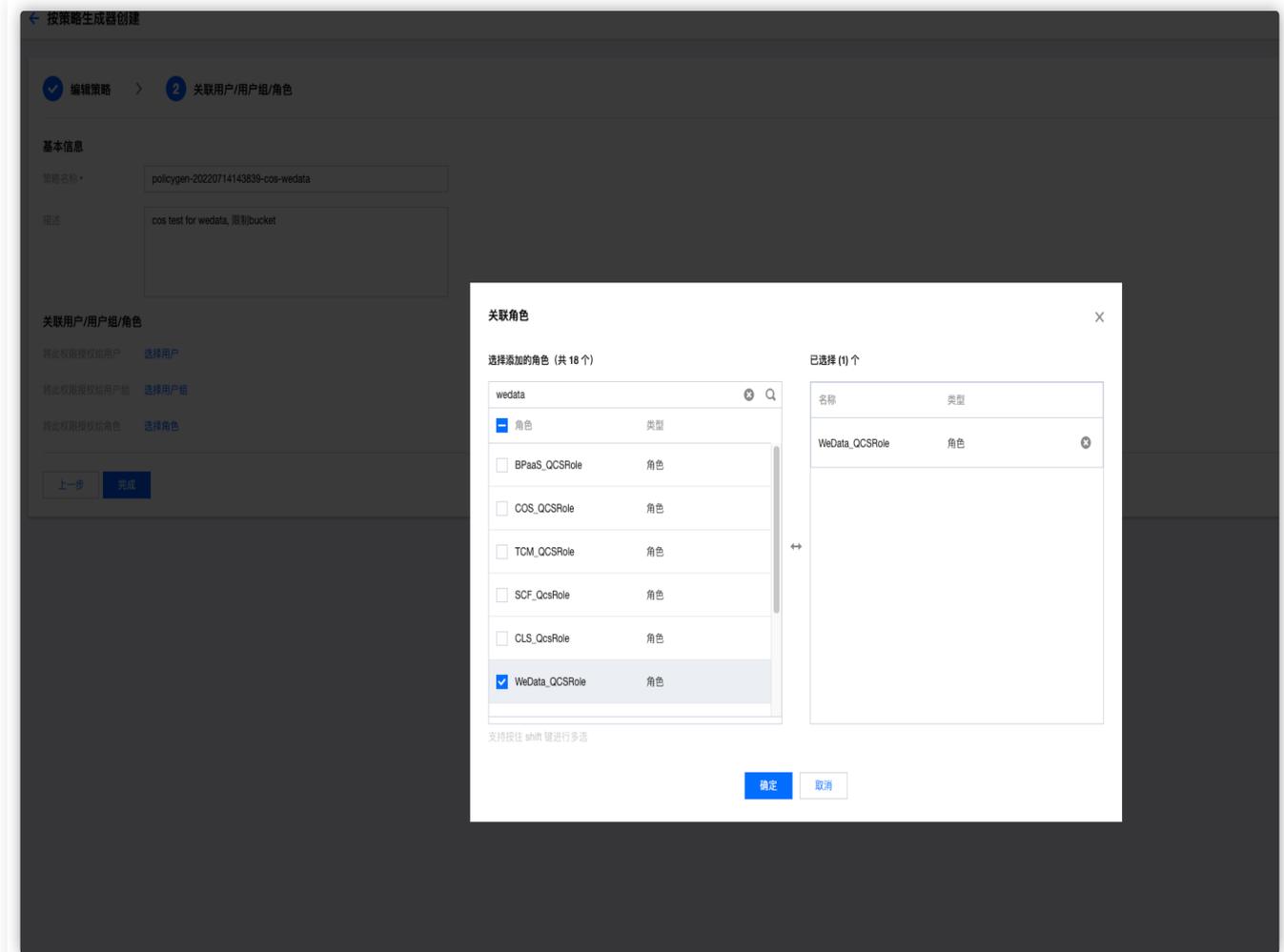
关联用户/用户组/角色

将此权限授权给用户 [选择用户](#)

将此权限授权给用户组 [选择用户组](#)

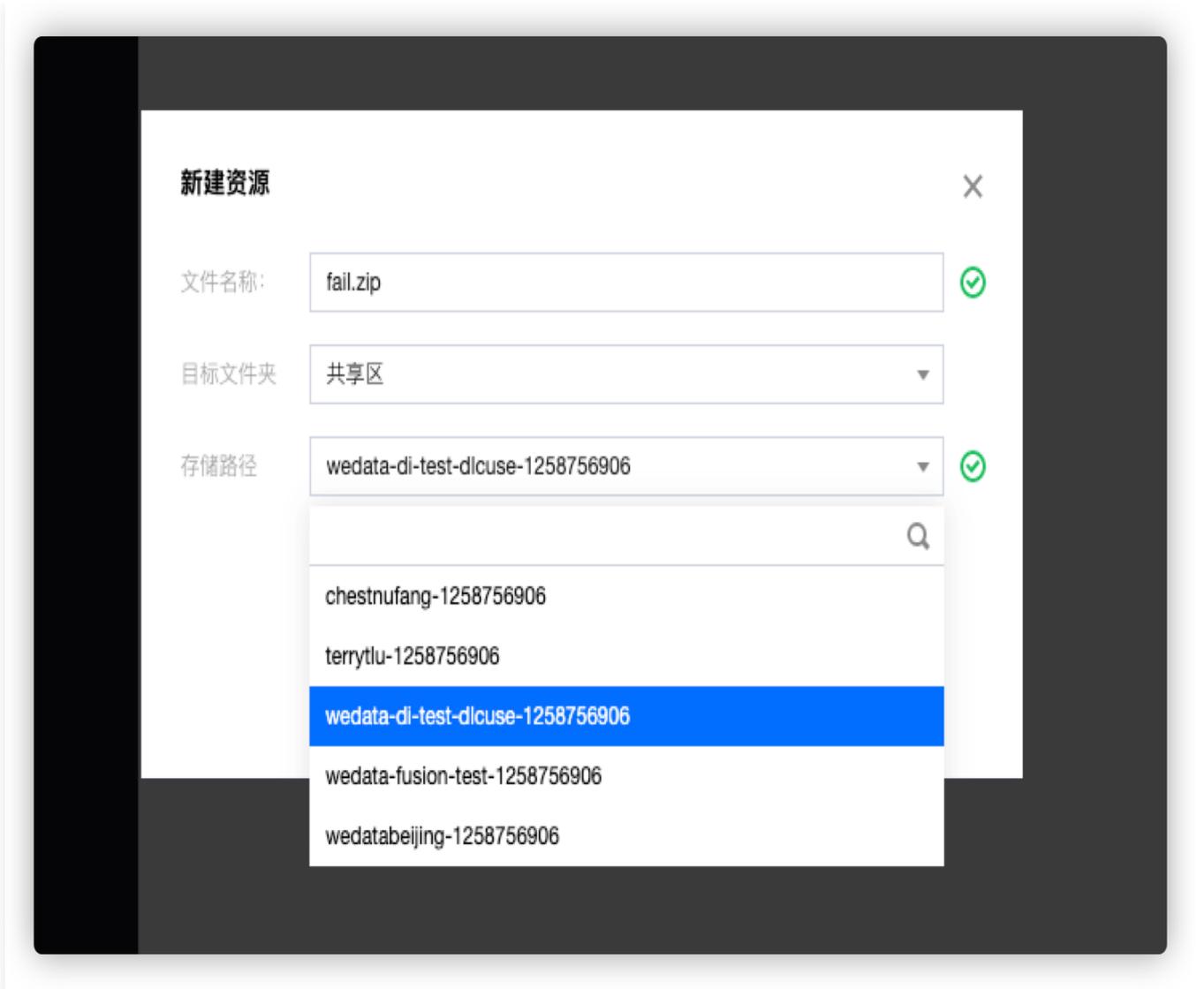
将此权限授权给角色 [WeData_QCSRole](#)
[重新选择角色](#)

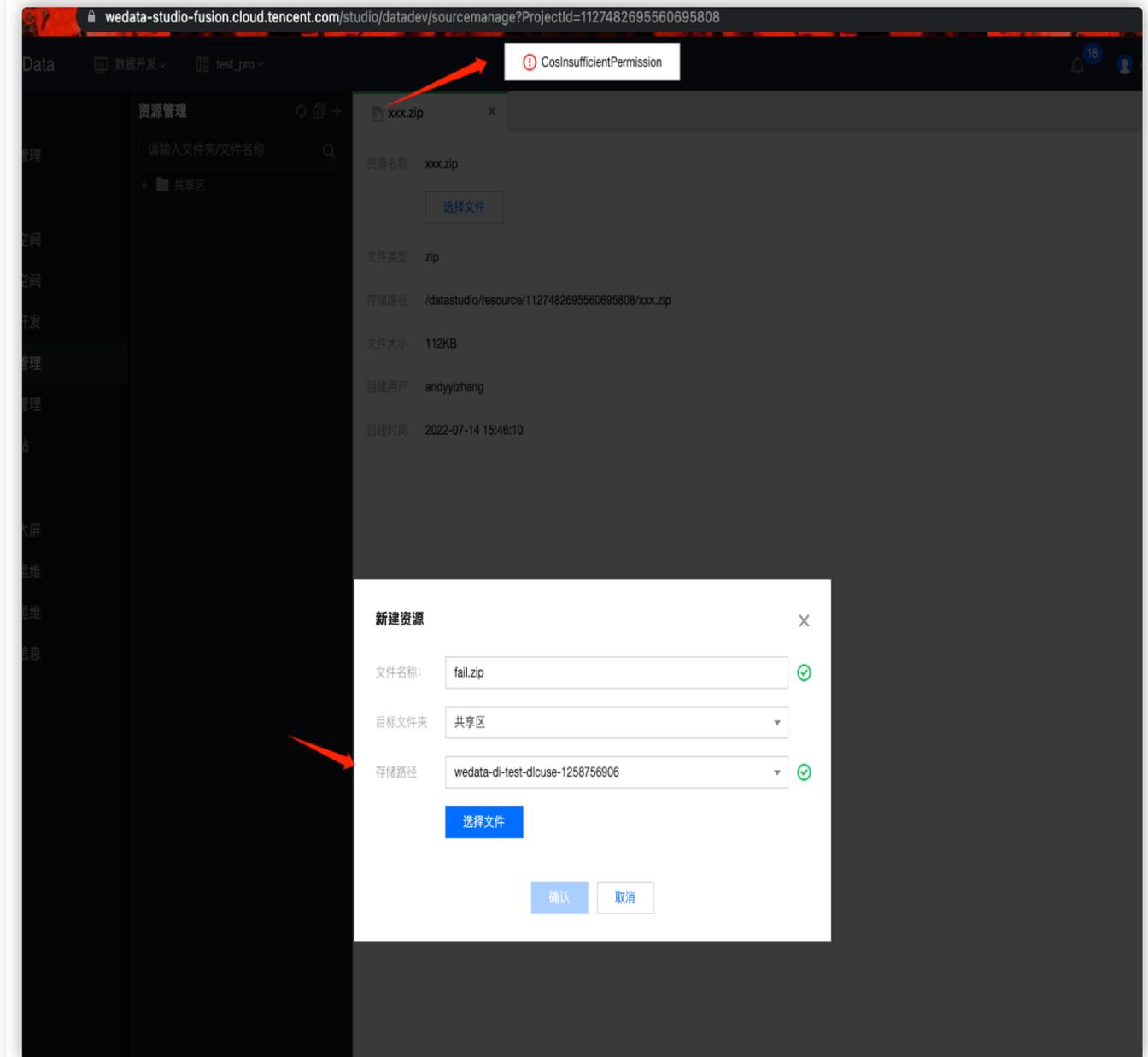
[上一步](#) [完成](#)



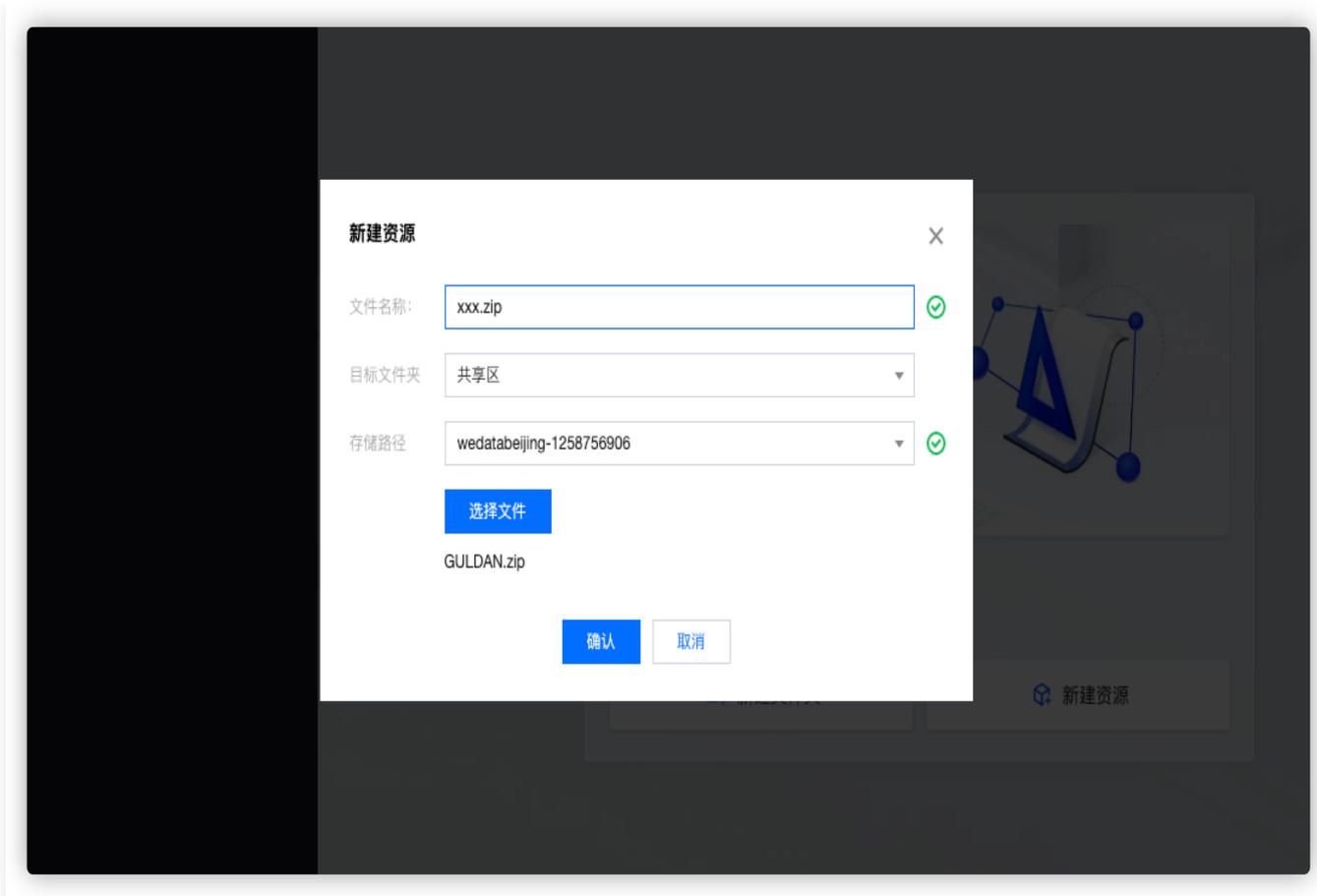
3.4 Test WeData

The bucket dropdown list still shows ALL buckets. However, if you select a bucket without authorization, an error message indicating no permission will appear.





If you select an authorized bucket, you can operate.



3.5 Final Generated Strategy Text for reference

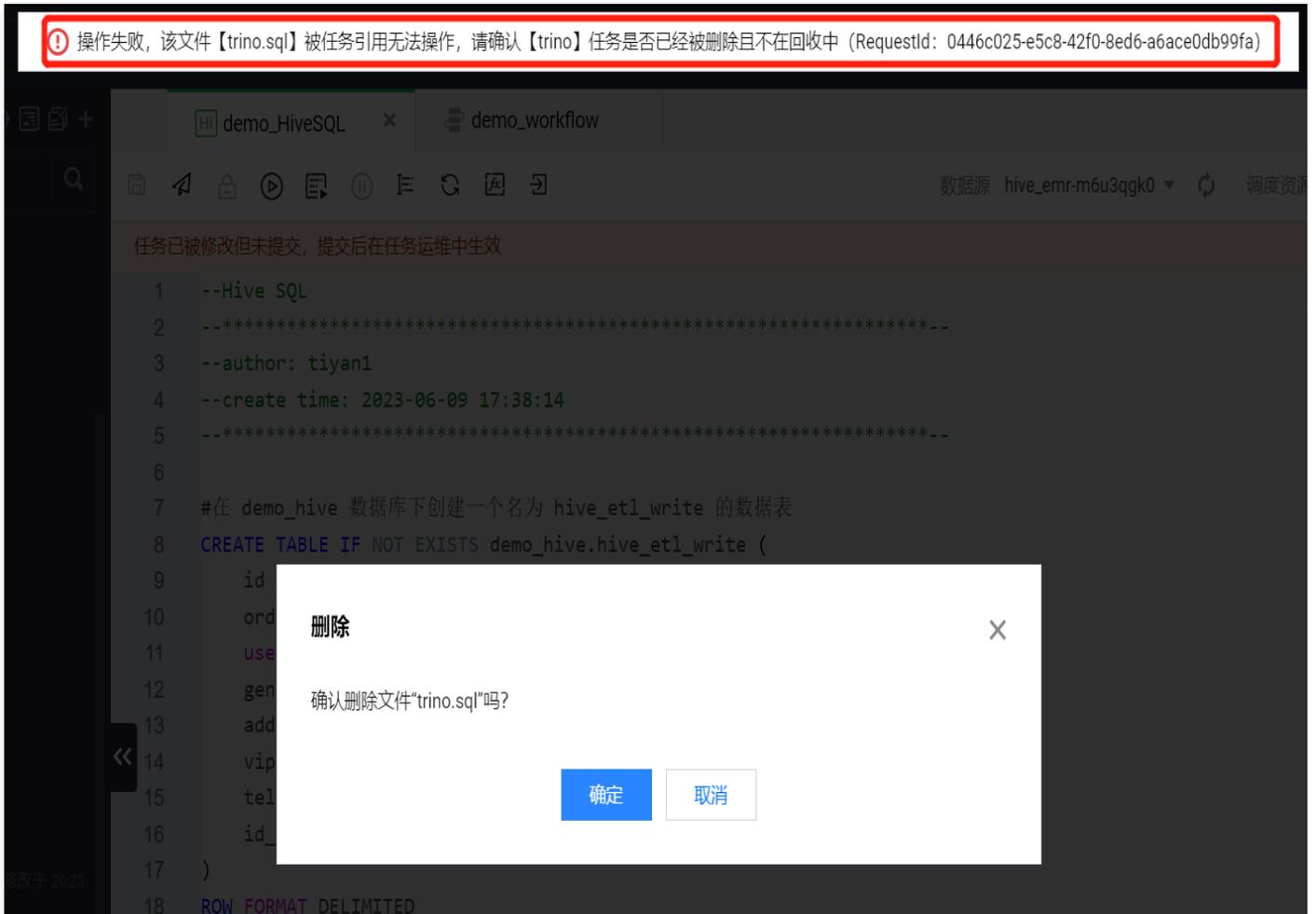
Directly create policies by filling in text. Without page operation, note that parameters in the resource part need to be modified.

```
{
  "version": "2.0",
  "statement": [
    {
      "effect": "allow",
      "action": [
        "cos:*"
      ],
      "resource": [
        "qcs::cos::uid/1258756906:wedatabeijing-1258756906/*"
      ]
    },
    {
      "effect": "allow",
      "action": [
        "cos:*"
      ],
      "resource": [
        "*"
      ]
    }
  ]
}
```

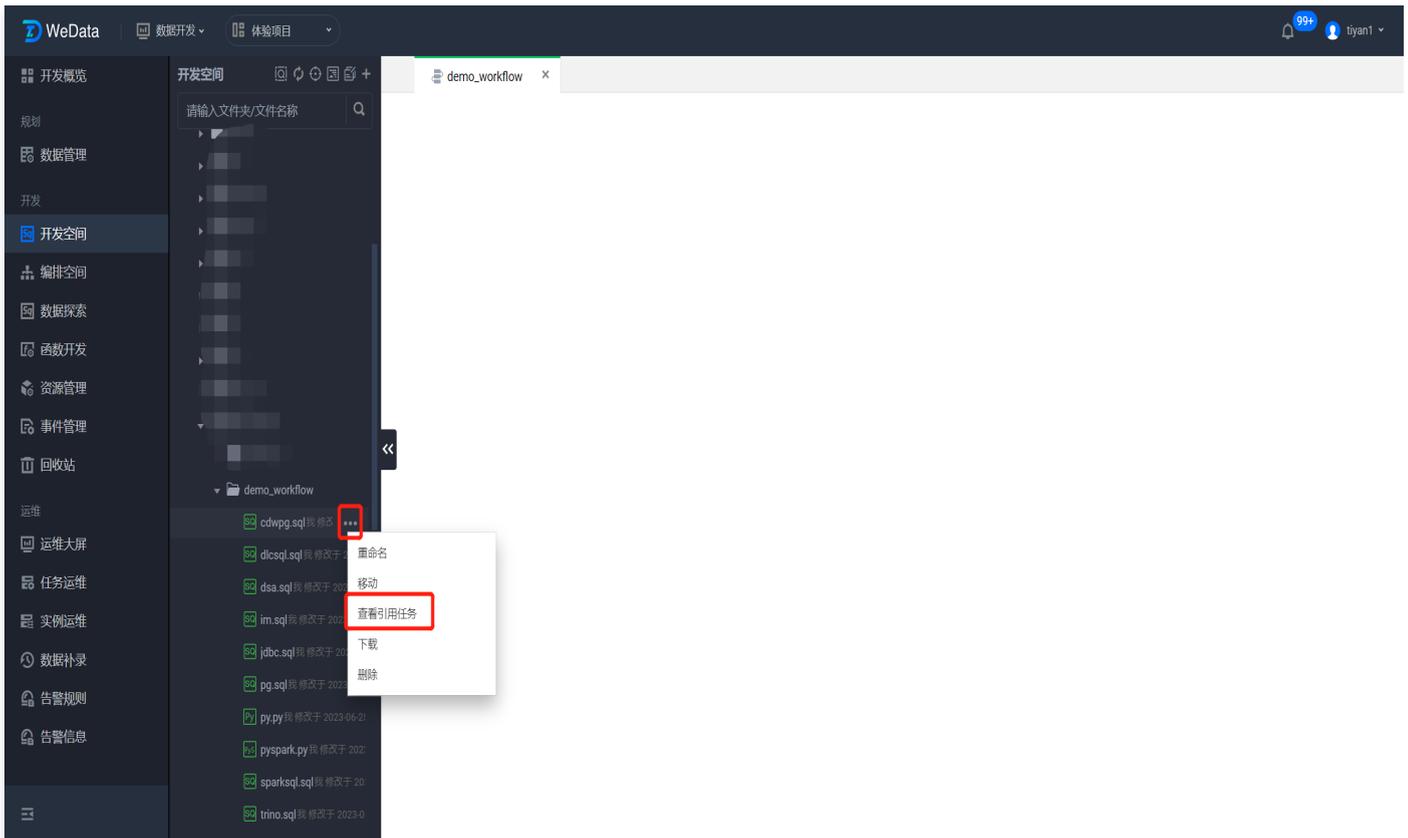


Clicking on the Develop Space Task Cannot Be Deleted, It Is Referenced

1. issue phenomenon



- Solution 1:** Select the undeletable development task in the Develop Space directory, open the operation menu of the task, and click **View Referenced Tasks**. You can see the orchestration workflow information that refers to this task. Find the corresponding workflow according to the information. After corresponding deletion in the workflow, it can be normally deleted in the Develop Space.



3. Solution 2: According to the task name in the pop-up window, search and delete in the orchestration space and recycle bin.

3.1 If the task is in the orchestration space, you can delete the task in the orchestration space at this point and check the delete script option.

3.2 If queried in the recycle bin, delete the task in the recycle bin and then the development space script can be deleted.

3.3 If it cannot be queried in the recycle bin either, you need to check whether the responsible person of this task is yourself. Currently, permission control has been implemented for task display in the recycle bin. You can only see tasks whose responsible person is the current logged-in account.

Task Creation Failure in the Orchestration Space, with a Notification That the Task Has the Same Name, but the Task Cannot Be Queried in the Orchestration Space

Solution: Query the script files in the development space and check whether there are existing files with the same name under the directory.

Development Task Debugging and Execution: Notification That No Available Nodes

1. issue phenomenon

When users are performing debugging and running of development scripts and task nodes, they see from the logs the task diagnostic information notification: No available nodes. At this point, it is

necessary to consider whether the scheduling resource group has reached saturation. Currently, the development task is unable to apply for scheduling resources, leading to a running failure.

任务已被修改但未提交, 提交后在任务运维中生效

```
1  --Hive SQL
2  -----
3  --author: tiyan1
4  --create time: 2023-06-09 17:38:14
5  -----
6
7  #在 demo_hive 数据库下创建一个名为 hive_etl_write 的数据表
8  CREATE TABLE IF NOT EXISTS demo_hive.hive_etl_write (
9      id STRING,          #序号
10     dd_id STRING,       #订单编号
11     user_id STRING,     #用户编号
12     gender STRING,     #性别
13     address STRING,    #地址
14     vip STRING,        #用户等级标识
15     telephone STRING,  #电话
16     id_card STRING     #用户身份证
```

运行日志

```
正在等待提交测试...任务[recordId=4610]创建成功, 正等待提交统一调度...
任务[recordId=4610]创建成功, 正等待提交统一调度...
[2023-06-25 01:54:29] [ERROR] 任务诊断信息: 没有可用节点, resource group 20230413174849011308, reason:
loader key: 1314991481_vpc-6npcniro_30.0.0.8, status: RESOURCE_NOT_ENOUGH, issue limit: 23, issue count: 19, cpu usage: 0.00%, cpu cores: 8, total mem: 163788
```

2. Solution

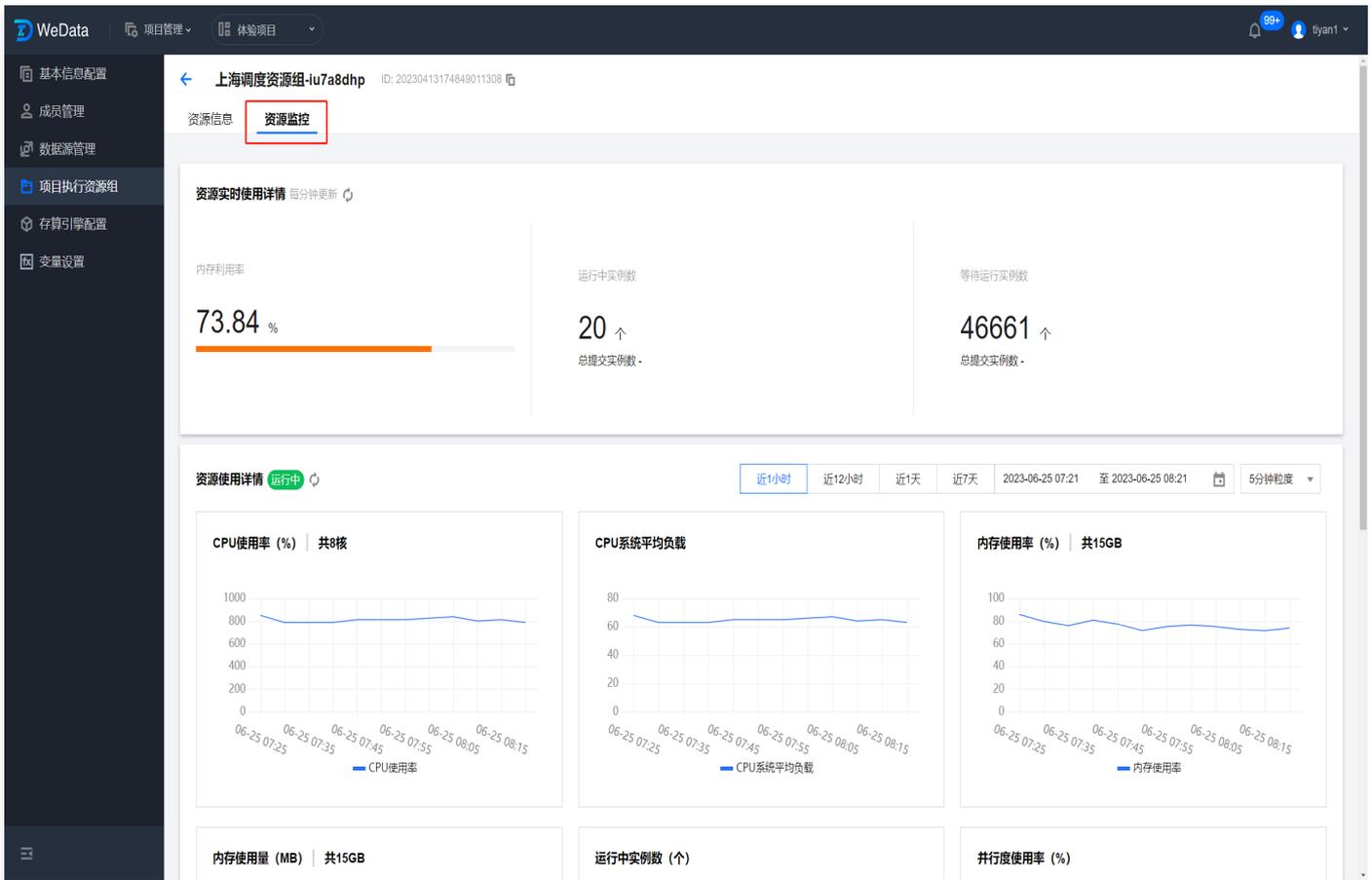
Enter **Project Management > Project Execution Resource Group**, find the selected scheduling resource group when the task is running, and click the resource group name to view the resource group usage status.

The screenshot displays the WeData Data Development Platform interface. The left sidebar contains navigation options: '基本信息配置', '成员管理', '数据源管理', '项目执行资源组', '存算引擎配置', and '变量设置'. The main content area is titled '项目执行资源组' and has four tabs: '调度资源' (highlighted with a red box), '集成资源', '数据服务资源', and '流计算资源组'. Under the '调度资源' tab, there are two sub-tabs: '标准调度资源' and '自定义调度资源'. A blue information box contains text about standard execution resources. Below this is a table with the following data:

资源组名称	状态	资源包规格/数量
上海调度资源组-iu7a8dhp	运行中	8C 16G / 1

共 1 条

Through resource monitoring, you can see the current load information of the resource group. Based on monitoring contents such as the frequency of use of the resource group and high-concurrency time points, reasonably plan the scheduling cycle of development tasks, and you can reduce the occurrence frequency of node occupancy issues.



Additionally, you can also add new scheduling resource groups to alleviate load pressure. Enter console → go to the Execution Resource Group page, and click Create under the Scheduling Resource Group tab to add new scheduling resource groups. For configuration methods, see [Scheduling Resource Group Configuration](#).

腾讯云 数据开发治理 WeData 执行资源组

调度资源组 | 集成资源组 | 数据服务资源组 | 流计算资源组

创建

调度资源组: 适用于专业版、企业版WeData, 提供不同规格标准化托管调度执行资源用于任务调度, 详情请参考计费说明

调度资源组名称/ID	地域	网络	绑定项目名称/ID	状态	资源包规格/数量	到期时间	操作
[模糊]	上海	[模糊]	[模糊]	运行中	8C 16G / 1	2023-07-12 21:57:50	关联项目 解除关联 调整配置 续费 销毁
[模糊]	上海	[模糊]	[模糊]	运行中	8C 16G / 1	2023-07-13 17:57:36	关联项目 解除关联 调整配置 续费 销毁
[模糊]	上海	[模糊]	[模糊]	运行中	16C 32G / 1	2023-06-29 16:07:53	关联项目 解除关联 调整配置 续费 销毁
[模糊]	上海	[模糊]	[模糊]	运行中	16C 32G / 1	2023-07-16 11:29:40	关联项目 解除关联 调整配置 续费 销毁
[模糊]	上海	[模糊]	[模糊]	运行中	16C 32G / 1	2023-07-13 11:16:48	关联项目 解除关联 调整配置 续费 销毁

共 5 条

10 条 / 页 1 / 1 页

Notably, during the configuration process of creating a scheduling resource group, it should be ensured that the region, network, associated project are consistent with the project where the

development task is located.

地域 广州 **上海** 北京 成都 南京 新加坡 上海金融

调度资源组所在地域。处于不同地域的云产品间网络不互通，创建成功后不可切换地域。请您谨慎选择，建议与产品版本选择同一地域，可降低服务访问时延。

网络 vpc-8npcniro | 上海VPC subnet-hy7vloxx | WeData_sh_... 共253个子网IP，剩余可用250个

调度资源所选VPC需具备访问公网能力。如现有的网络不合适，您可以去控制台[新建私有网络](#)或[新建子网](#)。

规格 **4C8G** 8C16G 16C32G

规格与性能说明详见[规格与性能说明](#)

数量 - 1 + 个

资源组名称 上海调度资源组-n7j5u4ca

描述 请输入调度资源组描述

计费配置

计费模式 **包年包月**

购买时长 **1个月** 2个月 3个月 4个月 5个月 6个月 7个月 8个月 9个月 1年 2年 3年 4年

续订 自动续订
账户余额充足时，设备到期后按月自动续费。

关联项目空间

关联项目 **立即关联** 暂不关联

关联项目空间后，仅此项目内任务可使用本资源组运行，后续可在产品控制台中修改。

项目空间 上海 体验项目

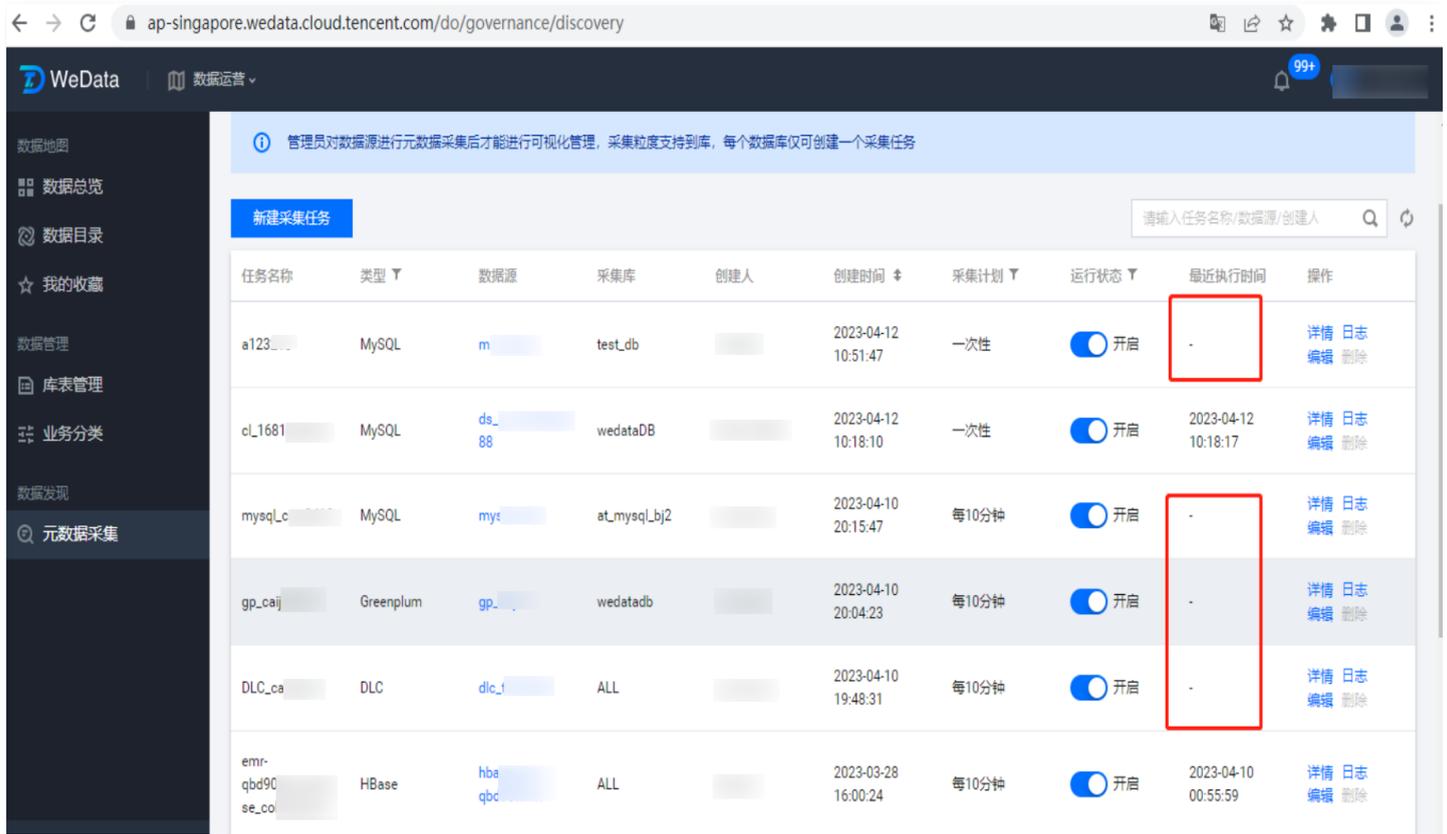
您已购买 上海 地域的产品服务，当前调度资源组仅可关联本地域下的项目空间。

Data Operations

Last updated: 2024-08-23 17:51:41

The newly created collection task did not execute when scheduled

Symptom: The scheduled collection task did not execute.



The screenshot shows the WeData Data Development Platform interface. The main content area displays a table of collection tasks. The table has the following columns: 任务名称 (Task Name), 类型 (Type), 数据源 (Data Source), 采集库 (Collection Library), 创建人 (Creator), 创建时间 (Creation Time), 采集计划 (Collection Plan), 运行状态 (Running Status), 最近执行时间 (Last Execution Time), and 操作 (Operations). The '运行状态' column shows '开启' (On) for all tasks. The '最近执行时间' column shows '-' for several tasks, indicating they did not execute as scheduled. These empty cells are highlighted with red boxes.

任务名称	类型	数据源	采集库	创建人	创建时间	采集计划	运行状态	最近执行时间	操作
a123	MySQL	m	test_db		2023-04-12 10:51:47	一次性	开启	-	详情 日志 编辑 删除
cl_1681	MySQL	ds_88	wedataDB		2023-04-12 10:18:10	一次性	开启	2023-04-12 10:18:17	详情 日志 编辑 删除
mysqlc	MySQL	mys	at_mysql_bj2		2023-04-10 20:15:47	每10分钟	开启	-	详情 日志 编辑 删除
gp_caij	Greenplum	gp_	wedatadb		2023-04-10 20:04:23	每10分钟	开启	-	详情 日志 编辑 删除
DLC_ca	DLC	dlc_f	ALL		2023-04-10 19:48:31	每10分钟	开启	-	详情 日志 编辑 删除
emr-qbd90-se-co	HBase	hba-qbc	ALL		2023-03-28 16:00:24	每10分钟	开启	2023-04-10 00:55:59	详情 日志 编辑 删除

Cause: Execution Resource Group configuration error.

Solution: Change the Execution Resource Group and try again.

The Business Information Tag cannot be deleted properly

Symptom: The Tag does not have a delete key and cannot be deleted.

The screenshot shows the '智慧能源生态平台' (Smart Energy Ecosystem Platform) interface. The main content area is titled '信息配置' (Information Configuration) and '标签管理' (Tag Management). A table lists several tags with columns for '标签名称' (Tag Name), '描述' (Description), '创建时间' (Creation Time), '更新时间' (Update Time), '创建人' (Creator), and '操作' (Action). The tag 'limye' is highlighted with a red box. The table also includes a search bar and pagination controls.

标签名称	描述	创建时间	更新时间	创建人	操作
121	-	2023-01-09 22:50:23	2023-01-09 22:50:23	egonsun	编辑 删除
biaoqian1	-	2022-12-22 22:55:32	2022-12-22 22:55:32	limye	编辑 删除
xxx	1111	2023-01-03 16:17:00	2023-01-03 16:17:00	egonsun	编辑 删除
光伏云	-	2023-03-07 10:36:08	2023-03-07 10:36:08	limye	编辑 删除
测试	-	2023-03-29 10:24:57	2023-03-29 10:24:57	tangweichen	编辑 删除
测试数据	-	2023-01-04 16:45:27	2023-01-04 16:45:27	v_vwchtang	编辑 删除

Cause: Business logic issue.

Solution:

- When User A accesses the table, a box Tag is added to the table.
- When User B accesses the table again, they cannot delete the box Tag added by User A. They can only choose +1 (up arrow) or do nothing.
- After User B chooses +1, they can then choose -1 (down arrow).

Misinterpretation of information display in Data Panorama



The meaning of the number "0" in the circle node indicates that there are no more sub-directories under this directory. It is not related to the number of tables present in that directory.