

数据开发治理平台 WeData

常见问题



腾讯云

【 版权声明 】

©2013–2025 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【 商标声明 】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

文档目录

常见问题

产品基础相关

购买相关

使用相关

公共模块

数据集成

数据开发

数据运营

常见问题

产品基础相关

最近更新时间：2024-09-10 17:13:41

什么是腾讯云数据开发治理平台 WeData？

腾讯云 WeData 是腾讯云打造的一站式数据协作开发治理平台。

数据开发治理平台 WeData 常见的应用场景有哪些？

数据开发治理平台 WeData 常见的应用场景，包括企业级数据仓库构建、数据可视化协同开发、数据资产管理和数据治理等，详情请参见 [应用场景](#)。

数据开发治理平台 WeData 提供哪些产品能力？

数据开发治理平台 WeData 提供了即席分析、工作流协同编排、数据资产管理和数据治理等全链路数据加工能力。

WeData 产品架构与优势？

WeData 作为一款大数据开发治理平台，旨在提供开源、简单易用、安全稳定的产品服务，整体架构由运营管理体系和开发运营工具构成。详情请参见 [产品架构](#)。

数据开发治理平台 WeData 支持哪些类型数据源？

支持多种云上及自建数据源，如：

- 关系型数据库：Mysql、PostgreSQL、Oracle、SQL Server、IBM DB2、达梦 DM、SAP HANA 等。
- 大数据存储：Hive、HDFS、HBase、Kudu、Clickhouse、DLC、Impala、Gbase、Tbase 等。
- 半结构化：FTP、SFTP、COS。
- NoSQL：Redis、Elasticsearch。
- 消息队列：Kafka 等。

数据开发治理平台 WeData 支持私有化部署吗？

WeData 支持私有化部署，您可 [提交工单](#) 联系我们。

购买相关

最近更新时间：2024-10-14 14:48:21

WeData 如何计费？

WeData 目前已开启商业化，整体计费项目包括产品版本、和执行资源组等。详情请参见 [购买指南](#)。

WeData 产品版本地域与资源组地域有什么区别？

- 产品版本地域是指 WeData 服务实际部署地域，目前 WeData 在腾讯云上已部署支持北京、广州、硅谷等国内站地域，以及新加坡等国际站地域。WeData 支持一个主账号同时开通多个地域 WeData 服务，不同地域 WeData 服务之间功能版本、项目空间、业务数据等互相隔离不可互通，且不同地域项目空间仅可在该地域控制台查看。
- 资源组地域是 WeData 所提供的全托管执行机器的物理机实际部署地域，建议您根据实际业务数据实际所在地域、以及网络规划等购买对应区域的执行资源组。

如果开通了多个地域的产品版本，这两个地域产品版本以及项目等是否可以共享？

不同地域版本不可通用，且同一地域内同一时段仅可购买一个产品版本。在某一个地域下创建项目空间，仅支持在当前地域查看，您可以通过切换控制台查看不同地域下项目空间。

WeData 版本之间有什么区别？

WeData 采用包年包月的方式对产品版本进行计费，目前已提供专业版、企业版等多个产品版本。版本间功能差别请参见 [产品版本购买说明](#)。

WeData 执行资源组如何计费？

针对由 WeData 提供的全托管执行资源组支持包年包月的方式购买。执行资源组规格及性能请参见 [执行资源购买说明](#)。

调度资源组与集成资源组有什么区别？

- 调度资源组主要将周期任务提交至调度系统按计划周期运行，如 Hive SQL、Spark SQL、shell、数据质量等。
- 集成资源组主要用于实际运行数据同步任务，如离线同步、实时同步等。

WeData 提供的执行资源组与自定义资源组有什么区别？

- 执行资源组是 WeData 提供的全托管执行资源组，整体资源运行环境、物理机部署维护都由 WeData 服务提供，用户可根据自身业务需要购买相应资源规格，同时也支持进行资源组扩缩容、续费等操作。目前，调度资源组和集成资源组均支持全托管形式。
- 自定义执行资源组仅在企业版及以上的产品版本中支持，用户可将腾讯云上 CVM 机器关联到项目空间。CVM 费用、以及环境部署安装与维护等都由用户自行承担。

是否可以不购买调度资源组，使用 CVM 来构建自定义调度资源组？

购买企业版及以上的产品版本的用户，可以使用自定义调度资源组。自定义调度资源组请参见 [项目执行资源组](#)。

使用相关

最近更新时间：2024-07-03 11:14:31

哪些子账号可进行 CAM 角色授权？

策略名称	策略说明	作用
AdministratorAccess	该策略允许您管理账户内所有用户及其权限、财务相关的信息、云服务资产	- 可管理子账号权限 - 可登录 WeData 控制台 - 可创建 WeData 项目
QCloudResourceFullAccess	该策略允许您管理账户内所有云服务资产	- 可登录 WeData 控制台 - 可创建 WeData 项目
ReadOnlyAccess	该策略允许您只读访问账户内所有支持接口级鉴权或资源级鉴权的云服务资产	- 可登录 WeData 控制台 - 不可创建 WeData 项目

如何获取/查看 EMR VPC？

进入腾讯云 [EMR 控制台](#)，选择集群对应所在的地域及 ID。进入[集群详情页](#) > [基本信息](#) > [实例信息](#)页面，查看集群网络信息。

← emr-
ZL-110-EMR

基本信息

实例信息

引导操作

集群服务

集群资源

集群监控

实例信息

基础配置

实例ID

emr-

地域信息

北京(北京七区)

节点高可用

已开启

网络信息

wedat

安全组

sg-

配置安全组

云硬盘加密

未开启

创建时间

2021-11-23 20:37:33

Master公网IP

82

检查更新

计费模式

按量计费

对象存储

已授权

自定义服务角色

未设置

设置

说明

EMR 的网络信息默认使用"--"隔开 ID 和显示名称，可复制"--"前部分（红框内）信息进行搜索。

如何获取/查看数据源 VPC?

以 MySQL 为例，进入对应腾讯云 MySQL 控制台，选择对应实例所在的地域及 ID。进入**实例详情页 > 实例详情**页面，查看实例网络信息。



MySQL 整库同步时，运行后立刻单击“暂停”，任务状态一直显示"操作中"?

原因分析：整库同步任务开始运行后，立刻单击暂停，任务状态始终处于"操作中"。后端数据在持续同步，已经远远超过源表实际大小

解决方案：当 MySQL 整库同步任务问题处于 snapshot 阶段时，不支持“暂停”操作。

任务“暂停”后点击“继续”操作出现部分数据重复?

原因分析：当前版本，实时任务暂停之后的恢复将从上个 checkpoint 开始，因此存在部分数据重复属于正常现象。

实时同步过程中，删除数据源、更改数据源结构等操作后，任务显示“运行中”，但实际上并没有同步?

原因分析：在实时任务运行过程中，出现下面情况，任务仍然显示运行中，但是实际上没有数据同步。

- 在 MySQL 源数据库中删除数据源。例如，MySQL 到 Hive 的同步过程中，在 MySQL 源数据库中将 MySQL 或 Hive 的表删除。
- 在 MySQL 源数据库中更改数据源结构。例如，MySQL 到 Hive 的同步过程中，在 MySQL 源数据库中将 MySQL 或 Hive 的表删除或修改部分字段。

删除数据源和更改数据结构操作都属于异常问题，WeData 暂时无法获取任务的异常状态。

解决方案：在获取到异常日志后，针对上述两种情况的解决方法如下：

1. 删除数据源。用户需要在 WeData 实时同步界面重新选择新的数据源，配置好映射关系，然后单击**提交**，选择“保留作业状态数据，继续运行”。
2. 更改数据源结构。用户需要在 WeData 实时同步界面，单击**刷新**，获取最新数据源结构，配置好映射关系，然后单击**提交**，选择“保留作业状态数据，继续运行”。

MySQL 分库分表同步表的 schema 不同，任务显示“运行中”却未同步任务成功？

原因分析：由于 MySQL 分库分表同步时若配置表之间的 schema 不同，并且建立映射关系的字段不存在于某张表中，因此虽然任务状态显示“运行中”但实际未成功执行。同时，由于 Oceanus 目前采用任务出错时一直重启的策略，因此任务持续处于“运行中”的状态。

解决方案：针对上面第2种不成功的情况，需要在 WeData 实时同步界面，将不符合要求的表排除，配置好映射关系，然后单击**提交**，选择“保留作业状态数据，继续运行”。

MySQL 整库同步时，出现“Oceanus job config length exceeds 100k limit”提醒？

问题描述：MySQL 整库同步时，出现“Oceanus job config length exceeds 100k limit”提醒。

原因分析：为保证整库迁移任务的性能，当前版本中任务配置总体长度不可超过100k。若包含过多的数据表或者表字段总数过大，系统则提示限制。

解决方案：建议单个整库迁移任务中不要配置超过20张表，可分割为多个整库迁移或分库分表任务。

自定义字段配置是否支持复杂类型？

当前版本不支持 Array、Map、Struct 等复杂类型，仅支持基本类型。

提交任务时候报错“ache.calcite.sql.validate.SqlValidatorException: Cast function cannot convert value of type ...”？

问题描述：提交实时同步任务时候，系统报错 “[TencentCloudSDKError] Code=FailedOperation, Message=ache.calcite.sql.validate.SqlValidatorException: Cast function cannot convert value of type ...”。

原因分析：实时同步的来源和目标字段之间的类型无法转换。

解决方案：

1. 请检查配置的来源和目标端之间的字段类型是否匹配，具体数据来源和目标端间类型转换请参见 [同步数据源](#)。
2. 若任务为 MySQL-Ckafka 单表同步任务中，请检查是否建立了管理 pkNames_wedata_di、sqlType_wedata_di 与目标端映射关系。由于当前版本 Ckafka 字段不支持 Array、Map、Struct 等复杂

类型，因此无法将 pkNames_wedata_di 、 sqlType_wedata_di 参数写入 Ckafka 目标端。

子账号通过运维跳转到腾讯云 Oceanus 控制台，提示无 “PassRole” ？



解决方案：请参考 [Oceanus 权限管理](#) 文档。

公共模块

最近更新时间：2024-08-23 18:13:31

商业化购买

关于自动续费

- **自动续费时间：**自动续费是在订单到期当天至到期6天内系统进行扫描自动续费，具体续费时间点不固定。
- **是否开启自动续费修改：**需要在续费管理页根据资源 ID（UUID）找到对应资源修改。

买了某地域的资源，在另一个地域看到

在购买页面上，您需要选择两个地域。上面的地域代表资源地域，下面的项目地域将决定您购买的是哪个地域的 WeData。通常，如果出现了这种情况，很可能是由于项目地域的选择有误。您可以在订单中确认所选择的地域是否正确。

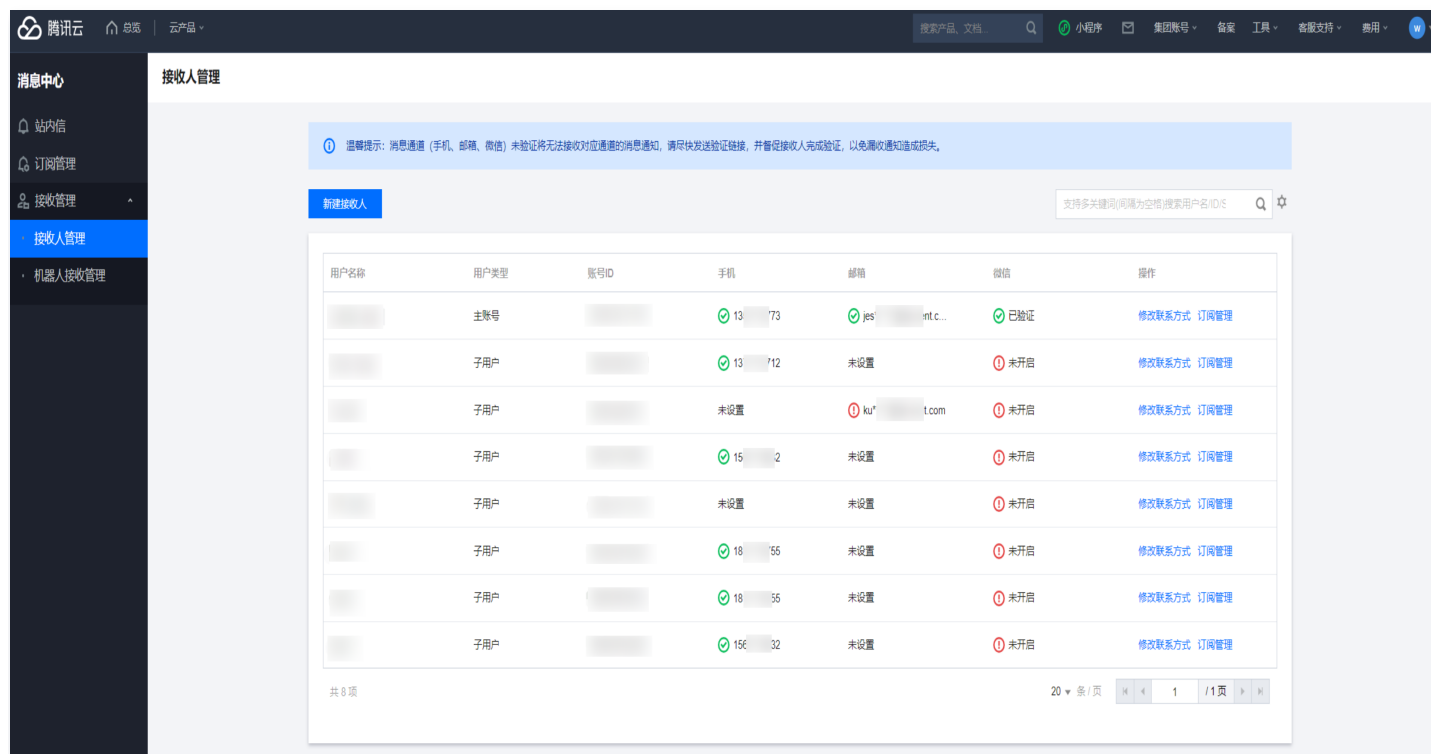
订单一直处于处理中

需要确认订单中购买的 WeData 调度资源/集成资源都处于运行状态，WeData 可以在够在购买的地域登录就是 WeData 处于运行中。调度资源/集成资源需要在资源界面，确定资源是否处于正常状态，新购/续费订单需要资源均处于运行中。确认均处于运行中后（表明资源已经可用了），等待云计费下一次回调将会更新订单状态。

消息服务

消息发送接收不到

进入 [消息中心](#) 查看各种联系方式是否正常配置。如果显示配置正常仍无法接收到告警，请联系 WeData 值班人员或 [售后支持](#) 进行处理。



集群同步

EMR 的配置信息与 WeData 执行机的配置信息不一致

1. WeData 配置信息没有及时更新

用户可以通过手动操作或自动同步的方式刷新，使两边的配置信息保持一致。

1.1 手动同步配置信息：WeData 集群管理会从 EMR 中同步配置文件到数据库。需要主动单击下图中的刷新按钮，WeData 才会去拉取配置文件。

1.2 自动同步配置信息：执行机定时从集群管理同步配置文件，每小时执行一次。

2. EMR 配置文件位置错误

WeData 集群管理同步 EMR 配置文件时，同步的是 EMR 控制台上的配置文件，而不是 EMR 机器上的原始配置文件。遇到这种问题，用户需要将 EMR 机器上的配置文件在 EMR 控制台页面上再配置一遍，这个操作不会影响 EMR 机器上配置文件的正确性，EMR 在遇到相同的配置的时候会进行覆盖操作。

数据开发在执行任务时，报错缺少 jar 包

首先，需要请 EMR 产品人员提供相应的 jar 包。将其放到执行机对应的目录下，然后执行任务看是否报错。如果没有报错，说明确实是 COS 中缺少 jar 包。接着，需要确认是哪个版本的 EMR，联系 WeData 值班人员或 [售后支持](#) 进行处理。

数据集成

最近更新时间：2024-11-25 09:31:22

1、如何处理 ClickHouse 到 DLC 同步报网络超时

问题现象：

用户同步 ClickHouse 数据到 DLC，数据量很小的时候可以成功，数据表较大的情况会报出网络超时。

```
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

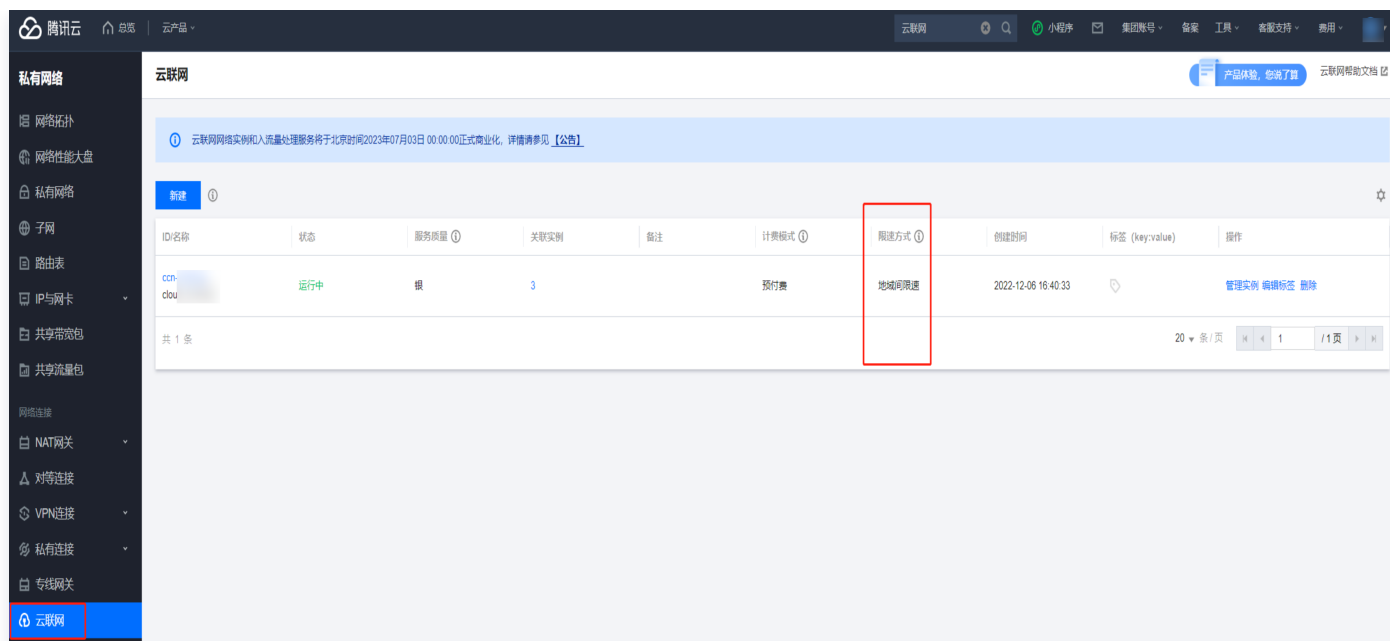
```
[2022-04-14 21:16:08]-[ERROR] Task execution failed
```

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

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

问题定位：

1. 通过登录执行环境 POD，Telnet 目标数据源（ClickHouse）连接信息可以连通，同时登录也是成功的，则排除网络故障。
2. 推测是 ClickHouse 数据量大或者高并发导致的读取超时，需要在 [云联网](#) 进一步降低并发。



3. 并发限制后，如果发现还是一样的报错，执行 ClickHouse 数据读取 SQL 误报错异常，同时也提高了读取任务超时时间的情况下，问题依然存在。则需要进一步了解客户执行环境资源组与 ClickHouse 的 VPC 是否一致，如果不一致，两边网络需要通过通过 [云联网](#)（CCN）打通。
4. 确认 CCN 产品跨区域网络打通需要的公网带宽，用户是否通过云联网打通了跨区域的两个 VPC，但是没有开通公网带宽。



5. 这时 CCN 为了方便测试，默认开通10Kb的流量，这解释了为什么小数据量同步和网络测试都是正常的，但是数据量大的情况下会报出网络问题。



6. 问题原因:

执行资源组与数据源不在相同 VPC，通过 CCN 打通跨区域的两个 VPC 后没有开通公网带宽，默认 CCN 支持小于10Kb的流量，超过后断流，导致小数据量可行，大数据量不行。

7. 解决方案:

通过 CCN 添加公网带宽后解决问题。

2、如何处理需要访问公网但网络不通的问题

问题现象:

```

2022-06-05 16:05:15] [WARN] priority set to 0, because NumberFormatException, the use is: null
2022-06-05 16:05:15] [INFO] PerfTrace traceJob=0, isEnable=false, priority=0
2022-06-05 16:05:15] [INFO] jdbcContainer starts job.
2022-06-05 16:05:15] [INFO] Driver registered.
2022-06-05 16:05:15] [INFO] register lindom this 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/ods;serverTimezone=Asia/Shanghai;tinyIntTrisBit=false
2022-06-05 16:05:16] [INFO] Available [jdbcUrl=jdbc:hive2://172.17.147.237:7001, 172.17.147.136:7001/ods;serverTimezone=Asia/Shanghai;tinyIntTrisBit=false]
2022-06-05 16:05:16] [ERROR] Exception when calling callable, 异常码: 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id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

The last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.
2022-06-05 16:05:17] [ERROR] Exception when calling callable, 异常码: 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id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

The 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, 异常码: 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id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

The 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, 异常码: 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id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

The last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.
2022-06-05 16:09:31] [ERROR] Exception when calling callable, 异常码: 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id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

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

```

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/jdbc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. 2
ities: [redacted]
ity: [redacted]
```

问题定位:

WeData 数据集成资源组的网络和 CDB 不在同一个 VPC 下, 网络不互通。

问题原因:

数据集成执行资源组和数据源都在用户的 VPC 网络下, 需要确保是同一个 VPC 网络, 否则网络无法连通, 执行资源组无法正常同步数据。如果是不同的 VPC 网络。

解决方案:

采用 [云联网](#) 或 [对等连接](#) 打通跨 VPC 的网络互通, 如果数据源是公网实例, 则需要配置 [NAT 网关](#)。

4、如何处理Hive On COS表写入报权限错误

问题现象:

离线同步, hive数据源, 表数据存储在cos中, 数据写入报错: 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: NjmWODMyMDZfMzFkMjc2MmVfYmYyNF8zNDh0MTdm); Trace ID: OGVmYzZiMmQzYjA2OWNhODk0NTRkMTBiOWVwMDAxO
Dc0OWRkZj k0ZDM1NmI1M2E2MTRlY2MzZDhmNmI5MmI1OTBjYzE2MjAxN2M1MzJiOTdkZjMxMDVlYTZjN2FjMmI0ZTRkY2MzNGMxNzE2YTQxZjhkNTAxM2ZmMGiWYzVjZDM=
[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: NjmWODMyMDZfMzFkM
jc2MmVfYmYyNF8zNDh0MTdm); Trace ID: OGVmYzZiMmQzYjA2OWNhODk0NTRkMTBiOWVwMDAxODc0OWRkZj k0ZDM1NmI1M2E2MTRlY2MzZDhmNmI5MmI1OTBjYzE2MjAxN2M1MzJiOTdkZjMxMDVlYTZjN2FjMmI0ZTRkY2MzNGM
xNzE2YTQxZj hkNTAxM2ZmMGiWYzVjZDM=
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)
```

问题定位：

查看账户是否关联WeData_QCUserRole角色，以及WeData_QCUserRole角色是否设置 COS 的 CAM 策略。

问题原因：

集成任务hive on cos写入数据，通过WeData_QCUserRole角色在CAM获取临时密钥，WeData_QCUserRole 角色下面没有添加 COS 相关的 CAM 策略，导致获取的临时密钥没有权限去读写 COS 桶。

解决方案：

在 [CAM](#) 的角色页面，搜索"wedata"，找到WeData_QCUserRole，看关联策略中是否包含了COS。如果没有，添加上 QcloudCOSFullAccess 即可。

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_QCUserRole

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

✕

🔍

☐

策略名

描述

☐

QcloudCOSBucketConfigWrite

对象存储（COS）修改存储桶配置的操作权限

☐

QcloudCOSBucketConfigRead

对象存储（COS）读取存储桶配置的操作权限

☐

QcloudCOSListOnly

对象存储（COS）获取存储桶、对象列表的操作权限

☐

QcloudCOSDataReadOnly

对象存储（COS）数据只读的访问权限

☐

QcloudCOSDataWriteOnly

对象存储（COS）数据只写的访问权限

☐

QcloudCOSDataFullControl

对象存储（COS）数据读、写、删除、

数据开发

最近更新时间：2025-02-06 17:20:02

资源上传无法选择 COS 桶

1. 问题现象

新建资源

文件名称：

请输入任务名称

目标文件夹：

共享区

存储路径：

请选择

无法选择路径

请选择存储路径

选择文件

确认取消

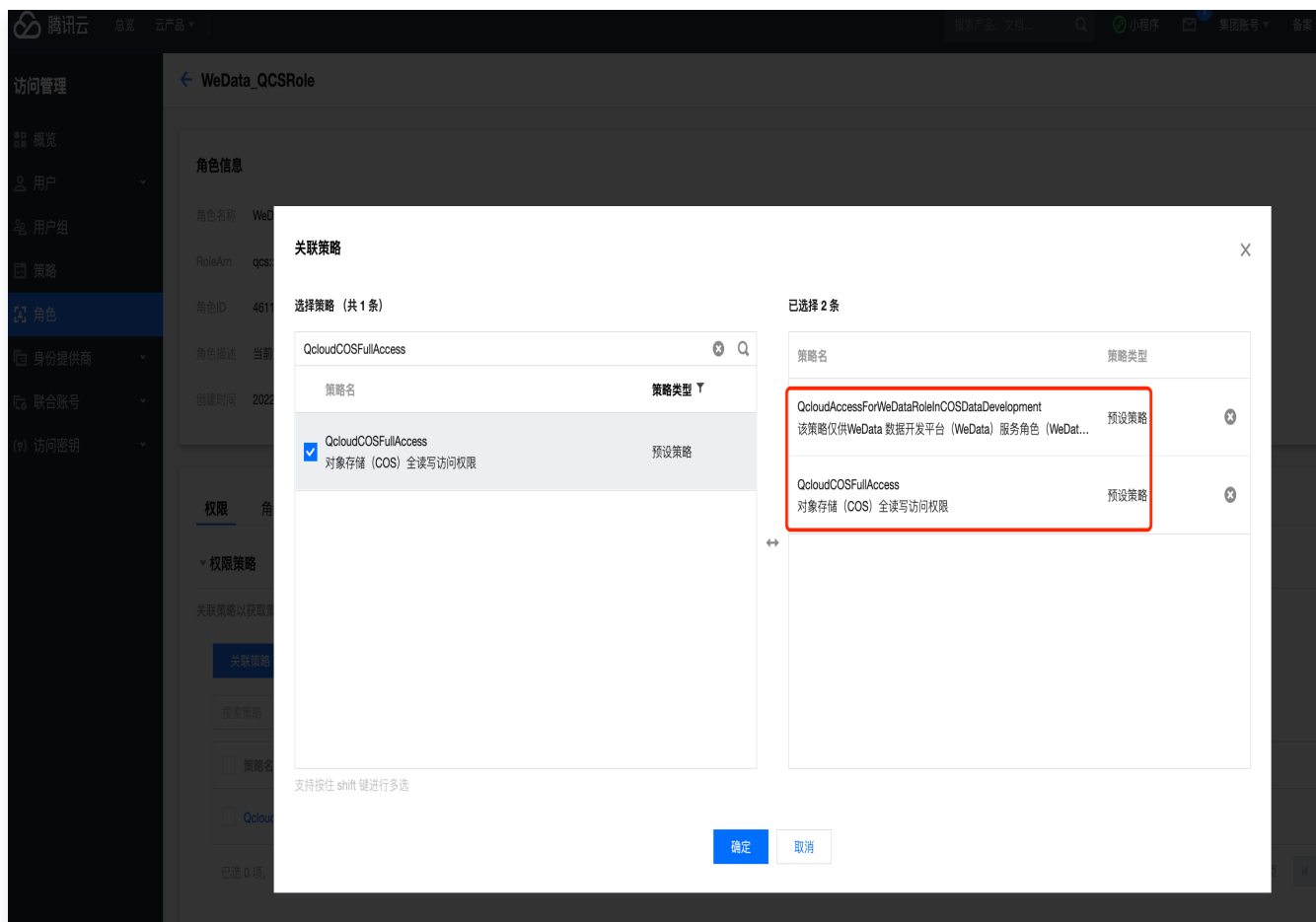
2. 解决方案

2.1 进入腾讯云 [访问管理](#)，选择角色菜单，搜索 WeData_QCSRole。

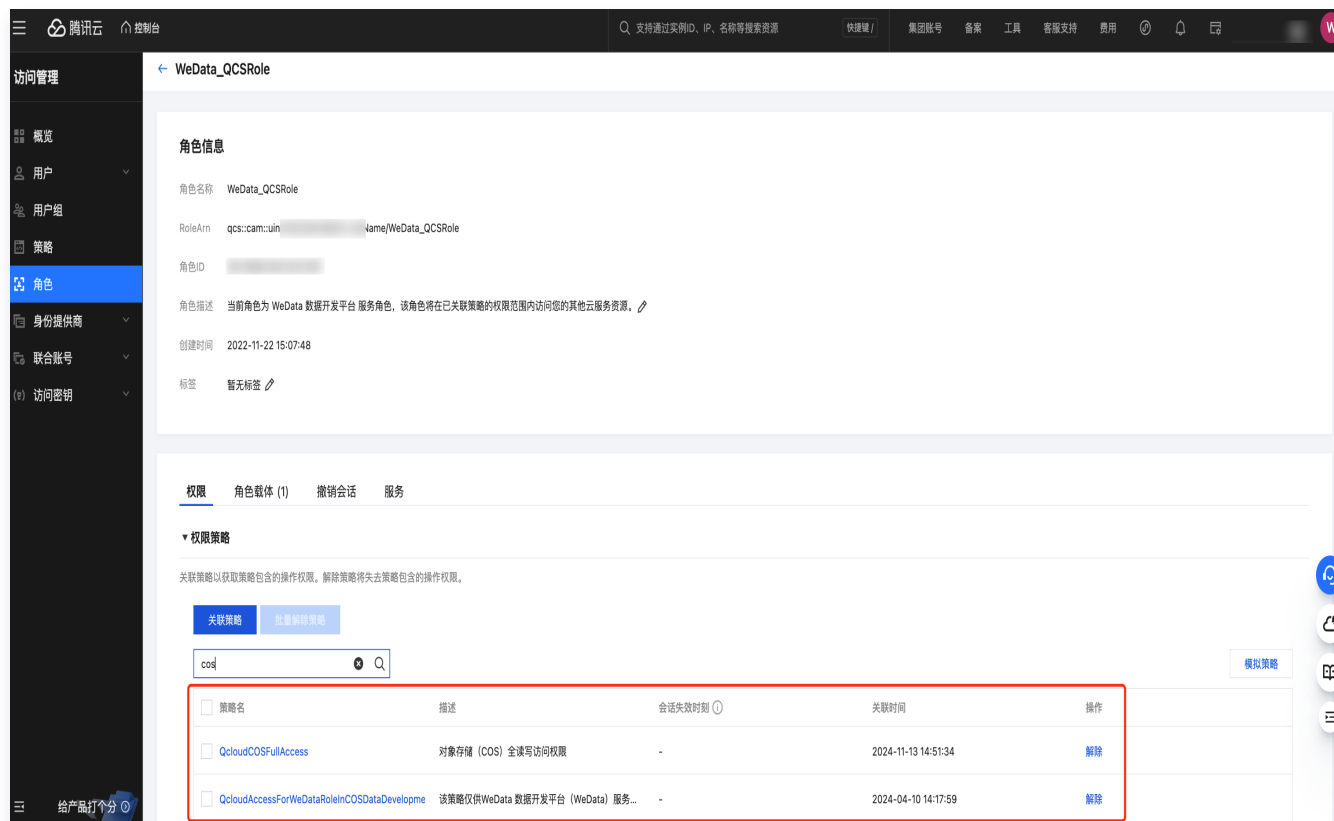


2.2 单击 WeData_QCSRole 角色名称，选择关联策略，搜索策

略“QcloudAccessForWeDataRoleInCOSDataDevelopment”和“QcloudCOSFullAccess”，并关联确认。



关联成功后在权限策略下可以看到该策略。再次尝试资源上传 COS 桶即可成功。



3. 方案补充

如果担心上一步授予 COSFullAccess 权限过大，可以参考下面操作，授予 WeData 只读取某个具体用来开发的 bucket 桶。

3.1 创建自定义策略

策略

① 用户或者用户组与策略关联后，即可获得策略所描述的操作权限。

新建自定义策略 删除 批量授权

全部策略 预设策略 自定义策略

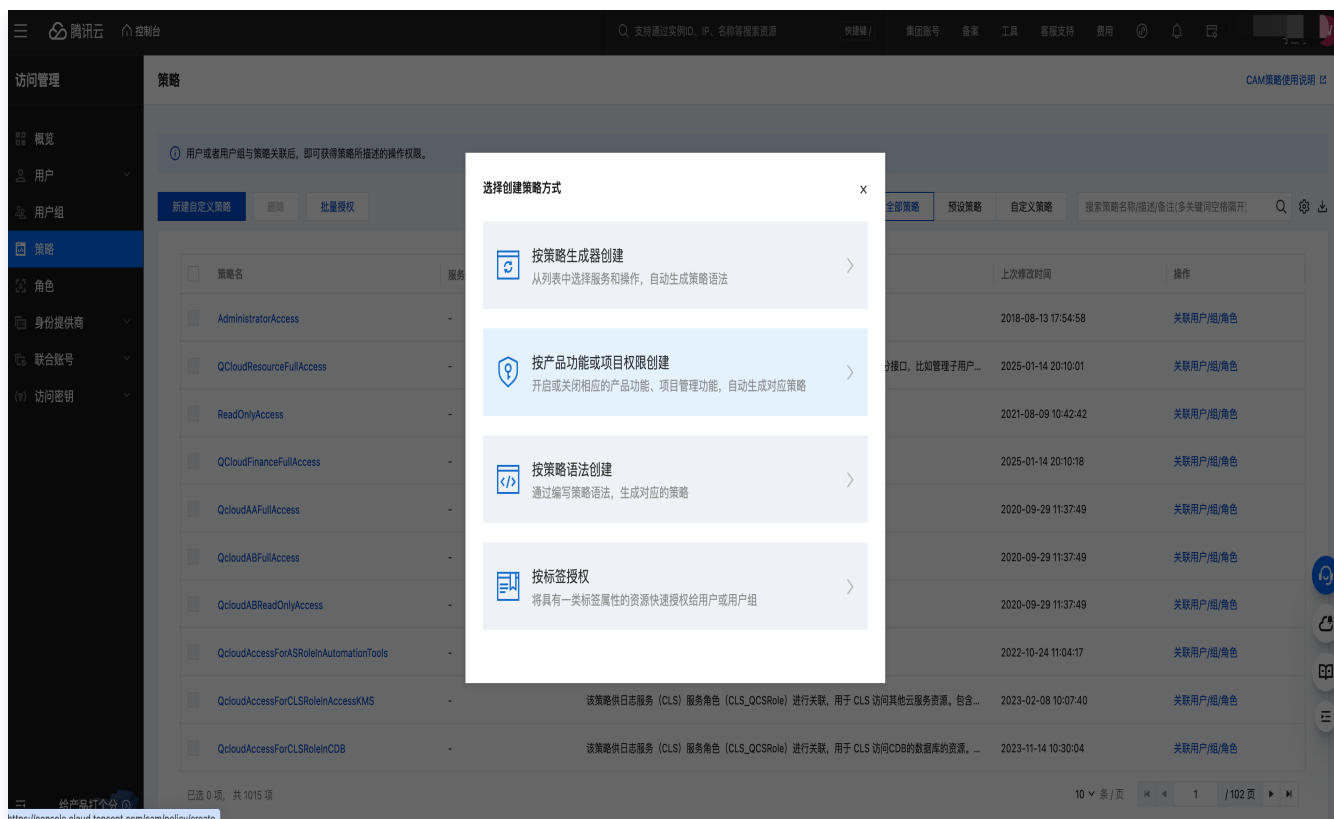
搜索策略名称/描述/备注(多关键词空格隔开)

策略名	服务类型	描述	上次修改时间	操作
AdministratorAccess	-	该策略允许您管理账户内所有用户及其权限、财务相关的信息、云服务资产。	2018-08-13 17:54:58	关联用户/组/角色
QCloudResourceFullAccess	-	该策略允许您管理账户内所有云服务资源（除了财务的所有权限，以及CAM的部分接口，比如管理子用户...）	2025-01-14 20:10:01	关联用户/组/角色
ReadOnlyAccess	-	该策略允许您只读访问账户内所有支持接口级鉴权或资源级鉴权的云服务资产。	2021-08-09 10:42:42	关联用户/组/角色
QCloudFinanceFullAccess	-	该策略允许您管理账户内财务相关的内容，例如：付款、开票。	2025-01-14 20:10:18	关联用户/组/角色
QcloudAAFullAccess	-	活动防刷（AA）全读写访问权限	2020-09-29 11:37:49	关联用户/组/角色
QcloudABFullAccess	-	代理记账（AB）全读写访问权限	2020-09-29 11:37:49	关联用户/组/角色
QcloudABReadOnlyAccess	-	代理记账（AB）只读访问权限	2020-09-29 11:37:49	关联用户/组/角色
QcloudAccessForASRoleInAutomationTools	-	弹性伸缩（AS）操作自动化助手 TAT 权限。	2022-10-24 11:04:17	关联用户/组/角色
QcloudAccessForCLSRoleInAccessKMS	-	该策略供日志服务（CLS）服务角色（CLS_QCSRole）进行关联，用于 CLS 访问其他云服务资源。包含...	2023-02-08 10:07:40	关联用户/组/角色
QcloudAccessForCLSRoleInCDB	-	该策略供日志服务（CLS）服务角色（CLS_QCSRole）进行关联，用于 CLS 访问CDB的数据库的资源。...	2023-11-14 10:30:04	关联用户/组/角色

已选 0 项，共 1015 项

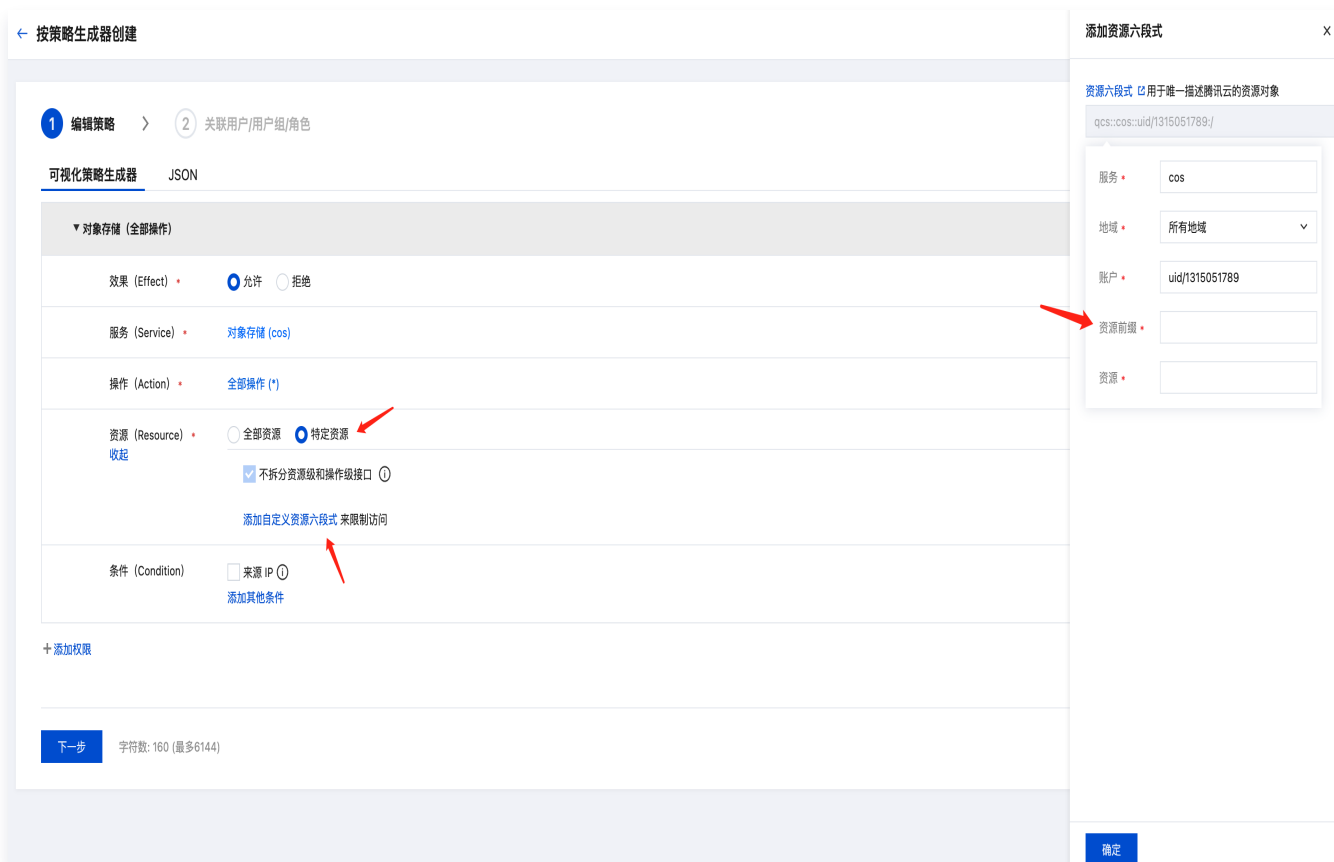
10 条 / 页

选择按策略生成器创建。



3.2 生成 COS 策略

重点在于限制制定的 COS 资源，同时增加下拉列表的权限，特定资源的位置，填写提供给 WeData 开发工程师的 COS bucket。其他内容参考截图填写。



填写完的汇总策略图如下：

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) ☒ 允许 ☐ 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) ☐ 来源 IP ⓘ[添加其他条件](#)

▼ 对象存储 (1 个操作)

效果 (Effect) ☒ 允许 ☐ 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) ☒ 全部资源 ☐ 特定资源
[收起](#)条件 (Condition) ☐ 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 保存策略, 关联到 WeData 的角色。

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

✓ 编辑策略

2 关联用户/用户组/角色

基本信息

策略名称 *

policygen-20220714143839-cos-wedata

描述

cos test for wedata, 限制bucket

关联用户/用户组/角色

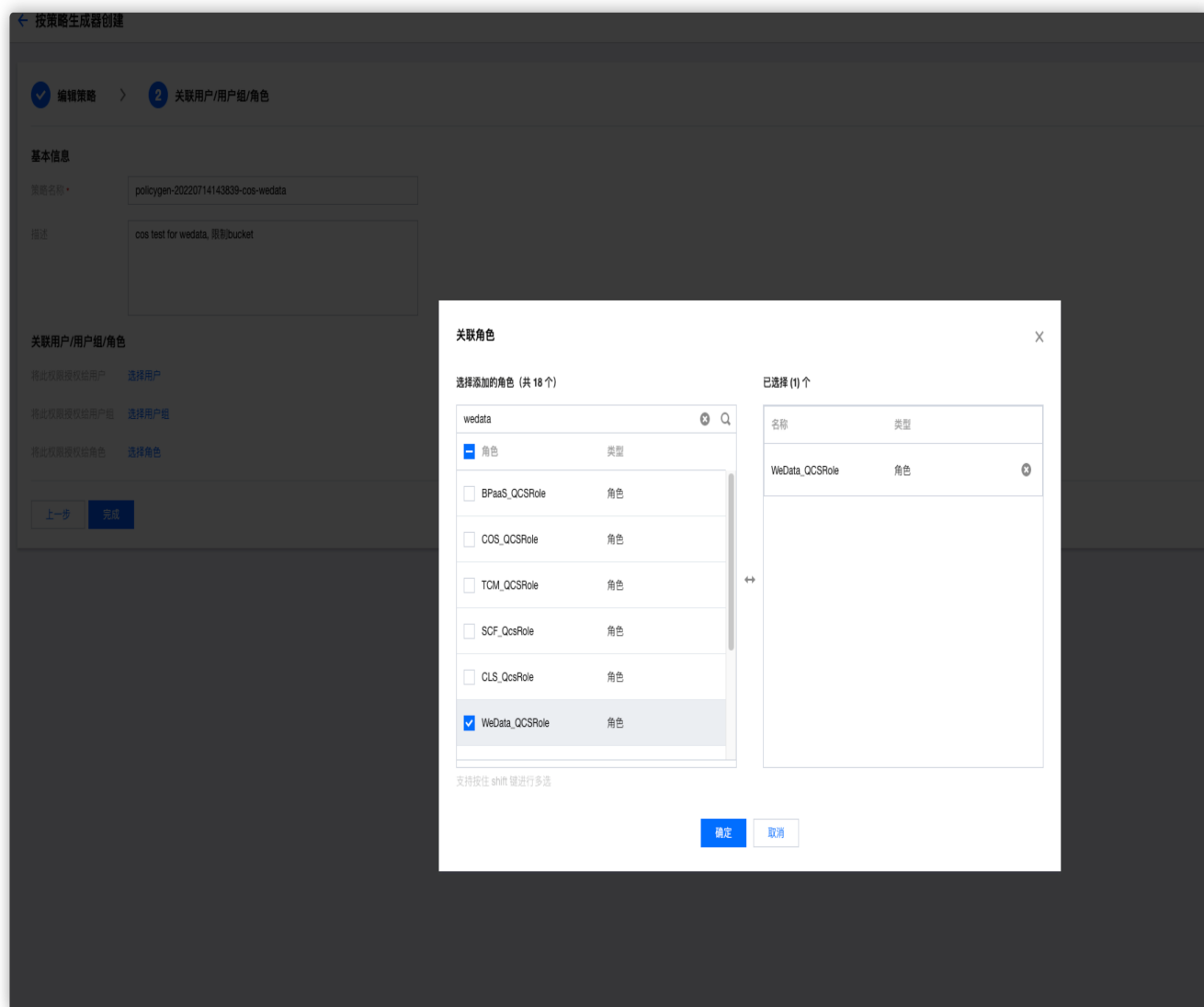
将此权限授权给用户 [选择用户](#)

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](#)

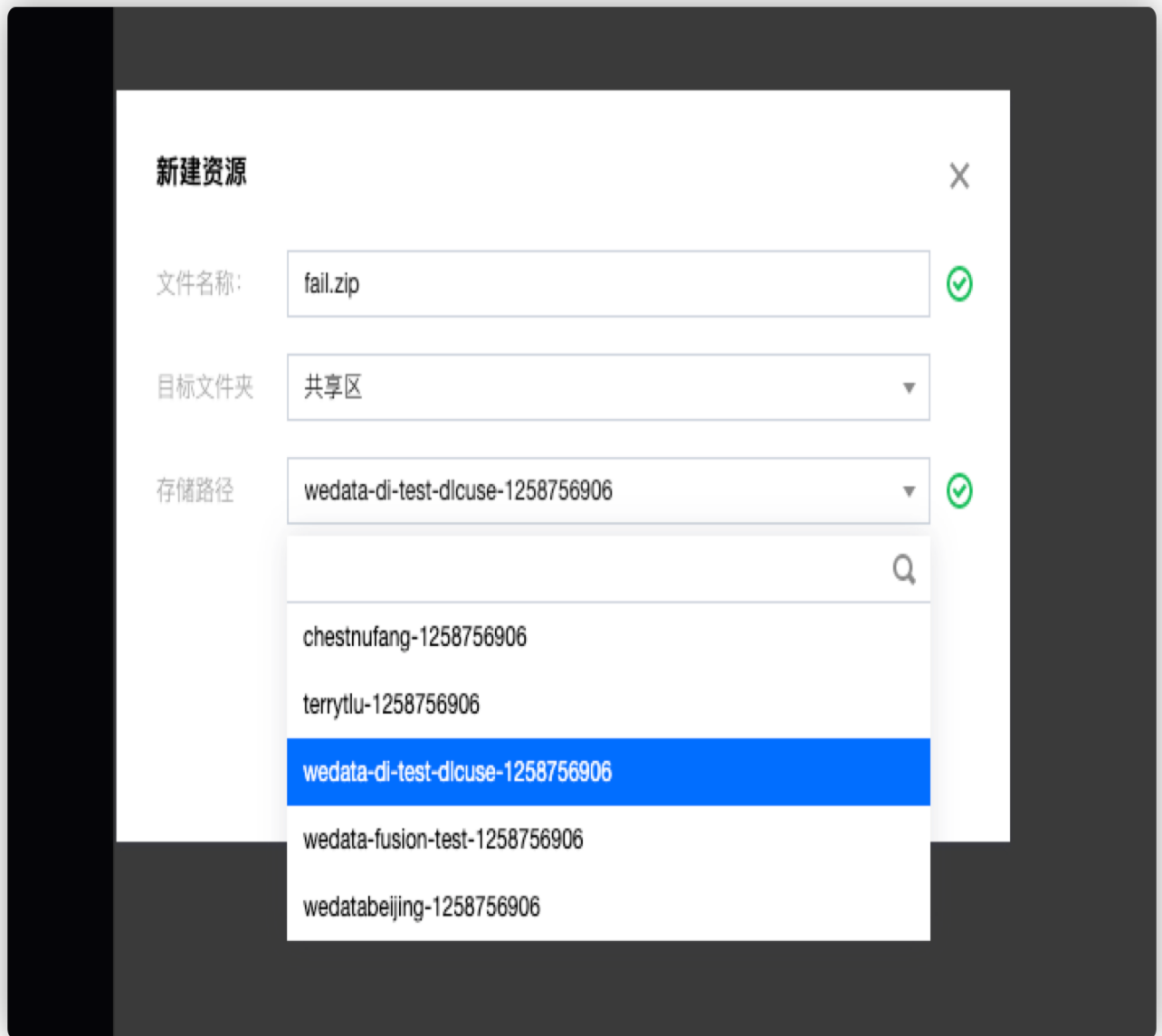
上一步

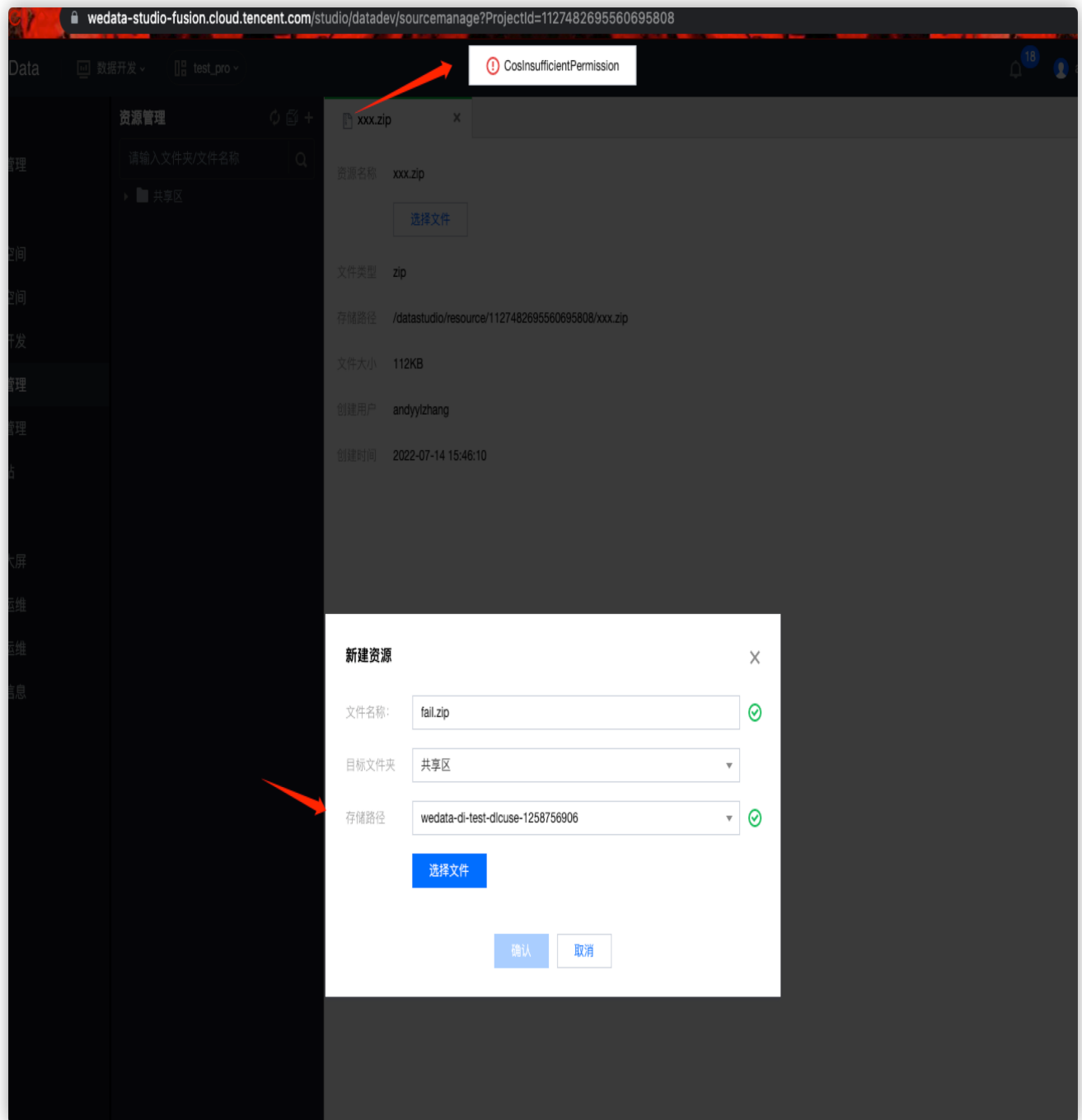
完成



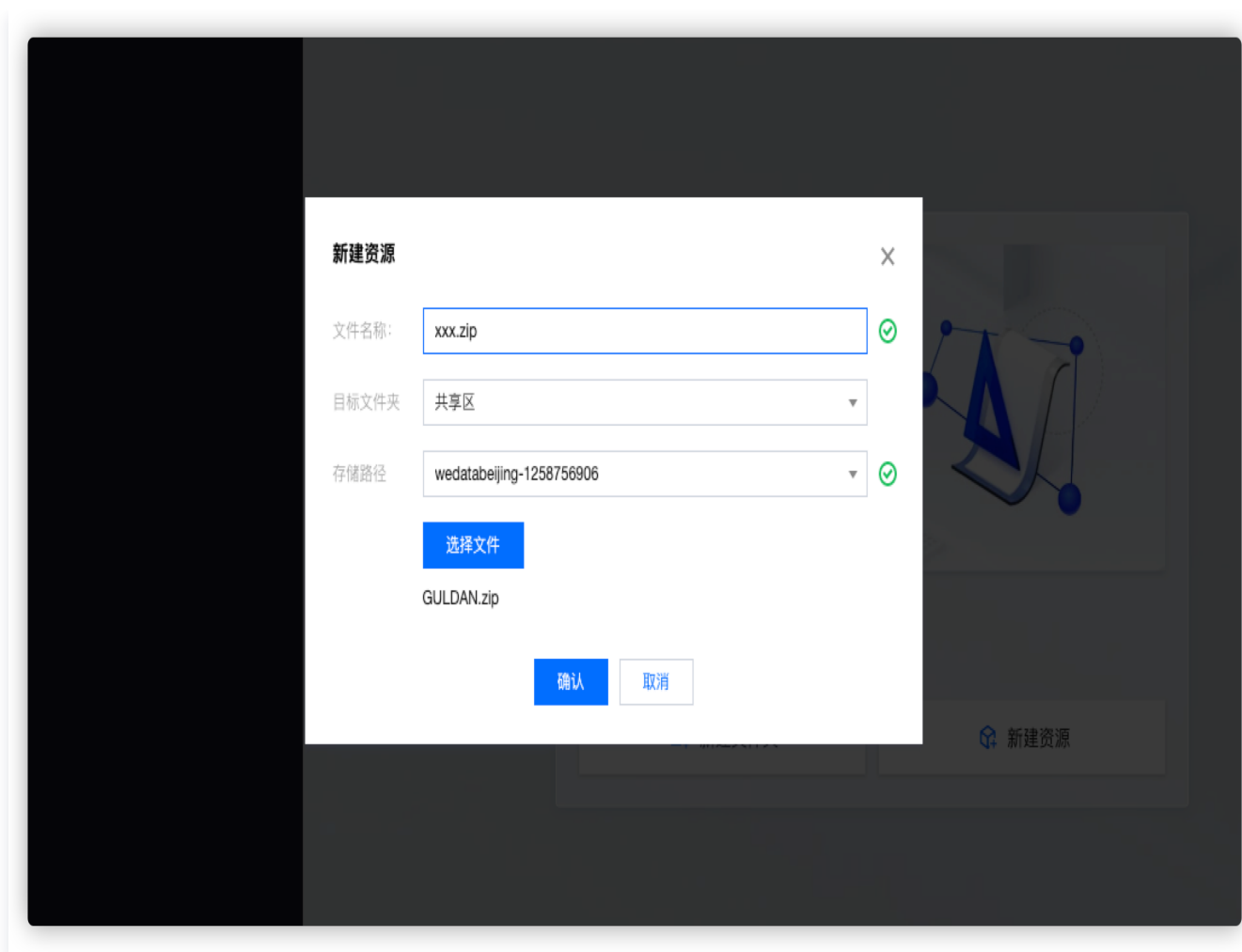
3.4 测试 WeData

bucket 下拉列表依然能看到所有 bucket，但是选择没有授权的 bucket 的话，会报错没有权限。





选择已被授权的 bucket，则可以操作。



3.5 最终生成的策略文本参考

可以直接按照文本填写生成策略，无需页面操作，注意需要修改 resource 部分的参数。

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

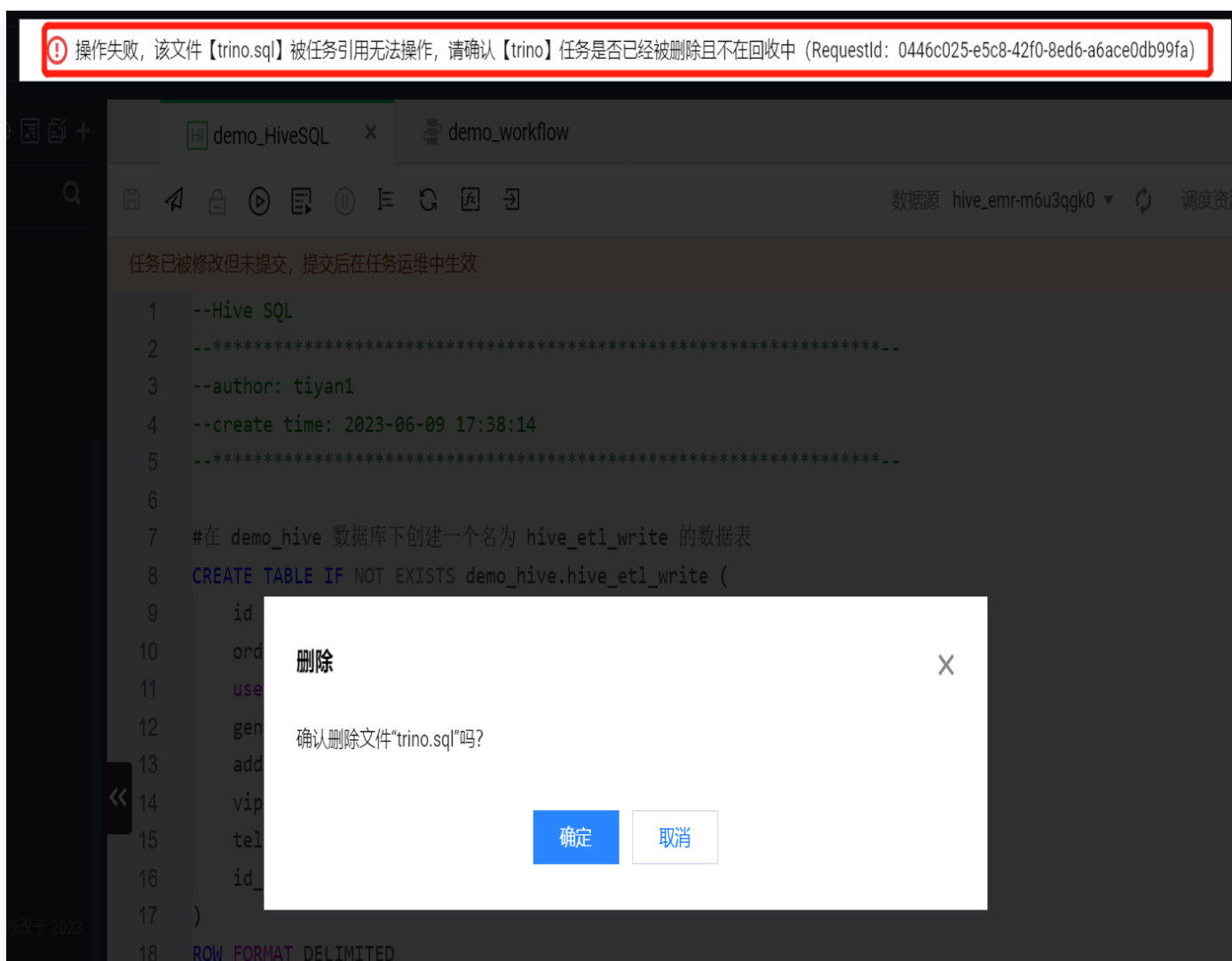
```

        "cos:*"
      ],
      "resource": [
        "*"
      ]
    }
  ]
}

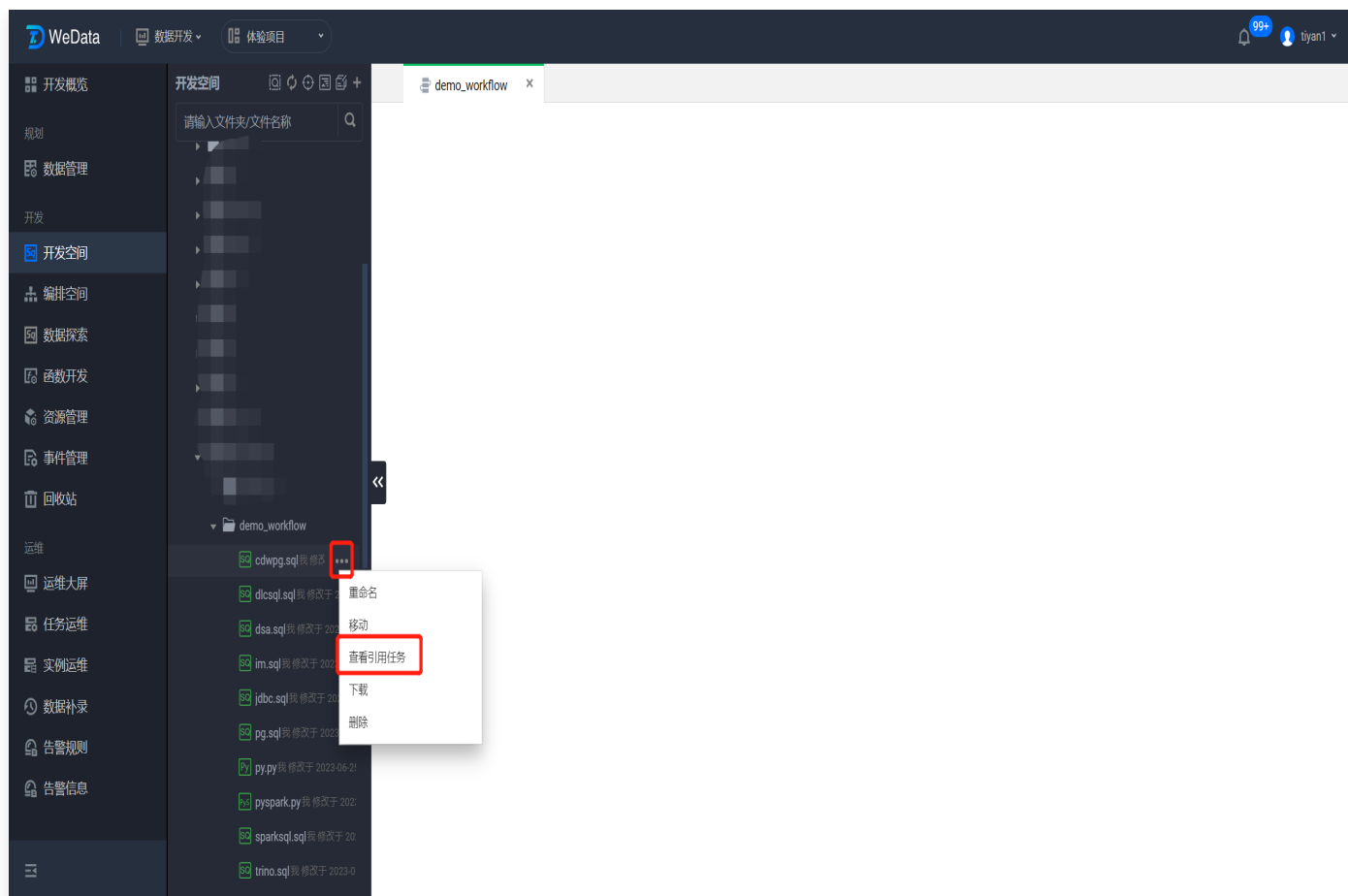
```

单击开发空间任务无法删除，显示被引用

1. 问题现象



2. 解决方案1：在开发空间目录中选中无法删除的开发任务，点开该任务的操作菜单，单击查看引用任务。可以看到引用该任务的编排工作流信息，根据信息找到对应工作流，在工作流中对应删除后，即可在开发空间中正常删除。



3. 解决方案2：根据弹窗中的任务名在编排空间和回收站中搜索进行删除。

3.1 如果该任务在编排空间，此时可以在编排空间中删除任务并勾选删除脚本选项。

3.2 如果在回收站查询到，在回收站中删除任务后即可删除开发空间脚本。

3.3 如果在回收站中也没有查询到，需要查看这个任务的责任人是否是自己，目前回收站的任务展示做了权限控制，只能看到责任人为当前登录账号的任务。

编排空间任务创建失败，提示任务已重名，但是在编排空间查询不到

解决方案：查询开发空间的脚本文件，查询同目录下是否有同名文件。

开发任务调试与执行时提示没有可用节点

1. 问题现象

用户在进行开发脚本、任务节点调试运行时，通过日志看到任务诊断信息提示：没有可用节点。这时需要考虑调度资源组是否已经饱和，当前开发任务无法申请到调度资源，从而导致运行失败。

demo_HiveSQL x demo_workflow

数据源 hive_emr-m6u3qgk0 调度资源组 上海调度资源组-uu7a8dhp 资源队列 default

任务已被修改但未提交, 提交后在任务运维中生效

任务属性

调度设置

版本

元

```
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. 解决方案

进入项目管理 > 项目执行资源组, 找到任务运行时所选的调度资源组, 单击资源组名称查看资源组使用状况。

WeData

项目管理

体验项目

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

变量设置

项目执行资源组

调度资源

集成资源

数据服务资源

流计算资源组

标准调度资源

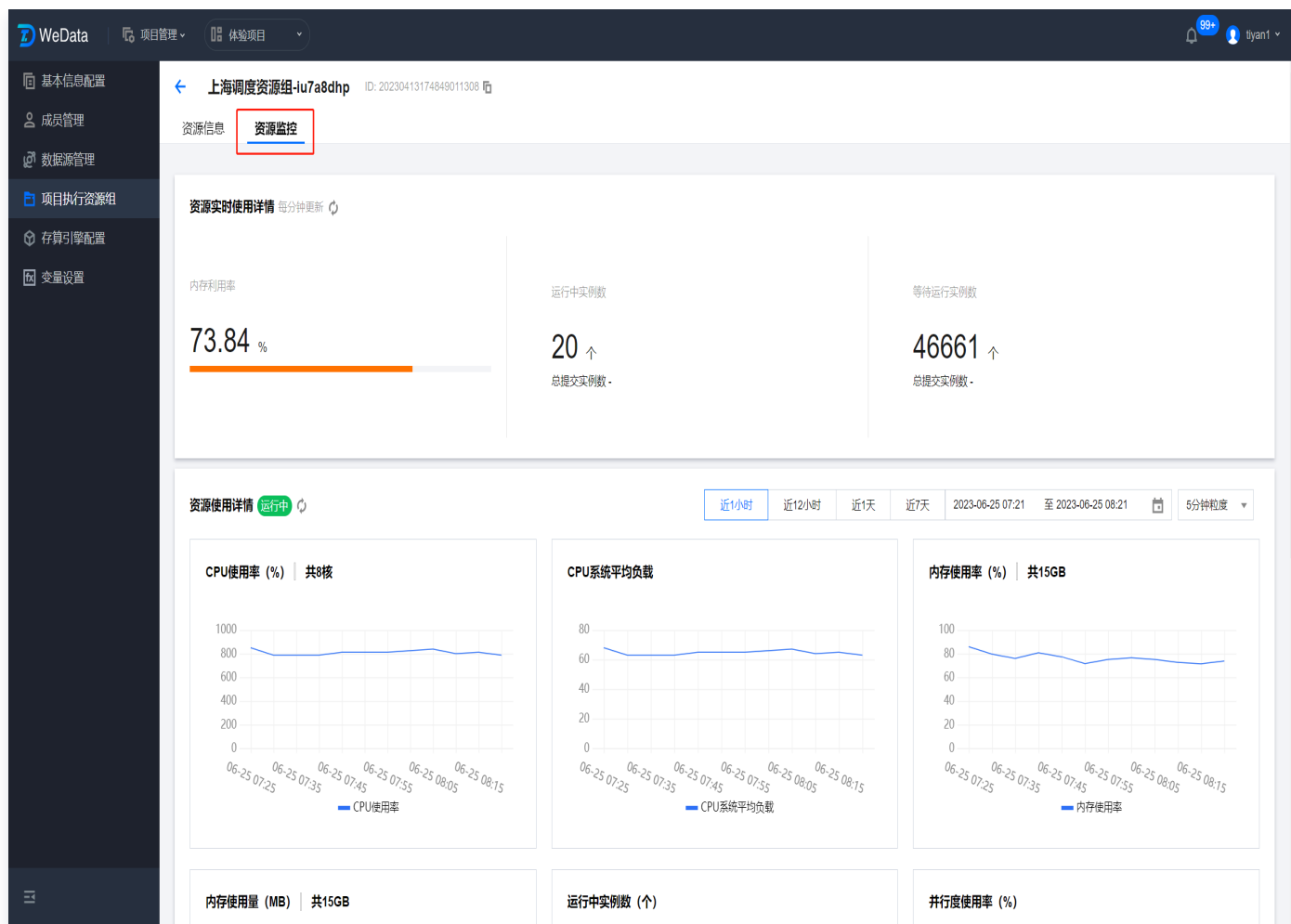
自定义调度资源

标准执行资源：专业版、企业版WeData提供不同规格标准化全托管执行资源用于任务调度，详情请参考[计费说明](#)。标准资源组变配、销毁等请前往[控制台](#)操作。

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

通过资源监控可以看到当前资源组的负载信息，根据资源组使用频率与高并发时间点等监控内容，合理规划开发任务的调度周期，即可降低节点占用问题的发生频率。



另外，也可以增加新的调度资源组来分担负载压力。进入控制台->执行资源组页面，在调度资源组页签下单击创建来增加新的调度资源组，配置方法详见[调度资源组配置](#)。

执行资源组 上海

调度资源组 集成资源组 数据服务资源组 流计算资源组

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

创建

请输入资源组名称

调度资源组名称/ID	地域	网络	绑定项目名称/ID	状态	资源规格/数量	到期时间	操作
	上海			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
	上海			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
	上海			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
	上海			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
	上海			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

需要注意的是，在创建调度资源组的配置过程中，需要保证地域、网络、关联项目与开发任务所在项目保持一致。

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-6npcnro | 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。

数据运营

最近更新时间：2024-08-22 20:48:11

新建的采集任务，下发调度没有执行

现象：采集的任务，下发调度没有执行。

ap-singapore.wedata.cloud.tencent.com/do/governance/discovery

WeData 数据运营

① 管理员对数据源进行元数据采集后才能进行可视化管理，采集粒度支持到库，每个数据库仅可创建一个采集任务

新建采集任务

请输入任务名称/数据源/创建人

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_Lbj2		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-qbd90se_co	HBase	hbaqbc	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

原因：执行资源组配置错误。

解决方案：更换执行资源组，重试。

业务信息标签不能正常删除

现象：标签没有删除键，不能删除。

智慧能源生态平台

数据运营

数据总览

数据字典

我的收藏

信息配置

数据发现

港华能源项目

总览 能源物联 应用集成 能源数仓 项目空间

数据简报 数仓管理 应用管理 平台管理 叶东林

信息配置

数据类目 标签管理

创建

请输入标签名称

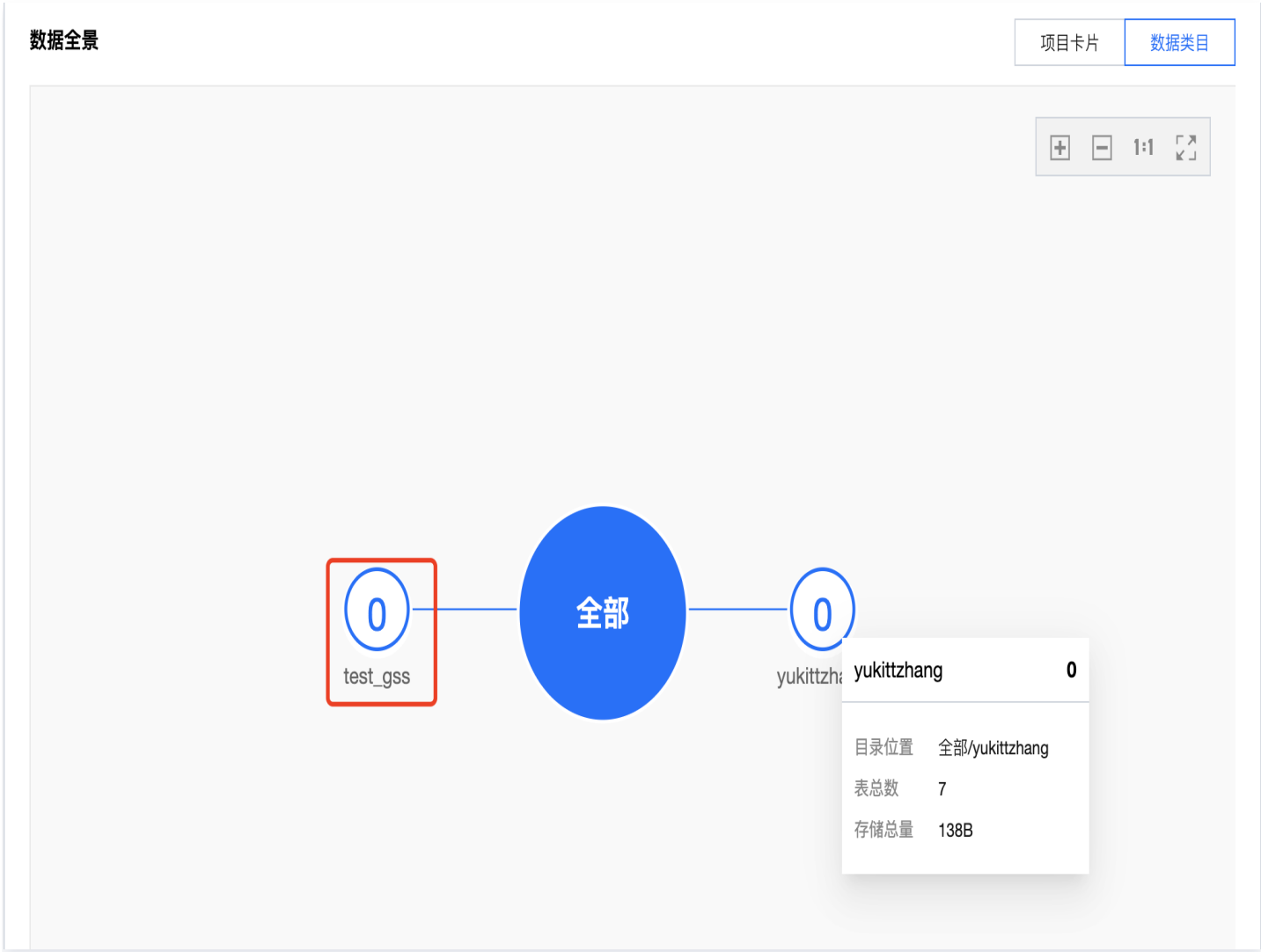
标签名称	描述	创建时间	更新时间	创建人	操作
121	-	2023-01-09 22:50:23	2023-01-09 22:50:23	egonsun	编辑 删除
biaoqian1	-	2022-12-22 22:55:32	2022-12-22 22:55:32	limye	编辑 删除
xxx	1111	2023-01-03 16:17:00	2023-01-03 16:17:00	egonsun	编辑 删除
光伏云	-	2023-03-07 10:36:08	2023-03-07 10:36:08	limye	编辑 删除
测试	-	2023-03-29 10:24:57	2023-03-29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

共 6 条

10 条 / 页 1 / 1 页

- 原因：业务逻辑问题。
- 解决方案：
- 当 A 用户访问 table 表的时候，给 table 增加了一个 box 标签。
 - 当 B 用户再次访问 table 表的时候，是无法删除 A 用户增加的 box 标签的，只能选择+1（即向上的箭头），或者啥都不处理。
 - 当 B 用户+1后，可以选择-1（即向下的箭头）。

数据全景中信息展示误解



圆圈节点中数字“0”的含义指的是该目录下已无其他目录了，与该目录下存在的表数量无关。