



PieDataCS 云原生平 台用户手册 (draft)

V2.20.0 EE

2024-12

目录

声明	ix
法律说明	ix
商标声明	ix
免责声明	ix
注意	ix
前言	xi
文档概述	xi
适用读者	xi
通用约定	xi
1 关于 PieDataCS 云原生平台	1
系统架构概览	1
云原生虚拟数仓相关概念	2
PieCloudDB 实例	2
数仓	2
虚拟数仓	3
组织	3
用户	3
角色	4
版本差异性说明	4
2 权限管理体系概览	5
权限分类	5
角色的继承与赋予	6
角色的赋权流程	6
3 快速上手	9
3.1 云上云版免费试用指导	9
适用场景	9
试用说明	9
申请方式	9
3.2 SQL 查询基础操作	12
登录组织下的计算空间	13
创建虚拟数仓	14
管理虚拟数仓	16
创建数据库和数据库对象	17
执行 SQL 查询并展示查询结果	19
4 控制台管理指南	21
4.1 控制台功能概览	21
4.2 管理组织	23

创建组织.....	23
添加组织用户.....	24
配置组织第三方身份认证.....	25
4.3 管理计算空间.....	27
访问计算空间.....	28
新建计算空间.....	28
修改计算空间名称.....	29
管理计算空间用户.....	29
审核计算空间新用户.....	31
4.4 查看组织用户.....	32
4.5 管理组织角色.....	33
关于控制台的系统预设角色.....	33
查看角色权限信息.....	35
添加自定义角色.....	37
继承和收回角色.....	38
授予角色给用户和收回授权.....	42
4.6 管理计费与订阅.....	43
查看计费模式与计费标准.....	44
订阅包年包月服务.....	44
4.7 费用中心.....	45
消费记录概览.....	46
充值购买 π 点.....	47
查看消费记录.....	48
申请开具发票.....	49
管理支付订单.....	51
4.8 我的工单.....	51
创建和提交工单.....	52
查看工单详情.....	52
删除工单.....	53
5 计算空间使用指南.....	55
5.1 计算空间功能概览.....	55
5.2 用户信息栏.....	57
个人中心.....	57
邀请注册.....	57
控制台.....	58
退出登录.....	58
5.3 虚拟数仓.....	59
创建虚拟数仓.....	59
删除虚拟数仓.....	61
启停虚拟数仓.....	61

修改虚拟数仓配置.....	62
虚拟数仓的扩缩容.....	63
查看虚拟数仓详情.....	65
配置外部接入.....	70
访问虚拟数仓.....	71
5.4 数据洞察.....	75
数据库对象概览.....	75
创建 SQL 文件.....	76
创建 SQL 文件夹.....	78
执行 SQL 查询并展示查询结果.....	79
查看执行计划.....	81
5.5 数据库.....	83
管理数据库对象权限.....	83
查看数据库对象信息.....	86
授予和收回数据库对象权限.....	89
全量授权或收回 Schema 下所有表和视图的权限.....	92
批量授权或收回 Schema 下数据库对象的权限.....	94
5.6 快速导入.....	96
导入数据.....	96
查看导入历史.....	102
5.7 查询历史.....	102
查询历史列表.....	102
查询历史详情.....	104
5.8 用量统计.....	105
5.9 管理计算空间用户.....	106
新建用户.....	107
移除用户.....	108
查看用户详情.....	108
授予/收回用户的角色.....	109
5.10 管理计算空间角色.....	110
关于系统预设角色.....	111
查看角色的权限信息.....	113
添加自定义角色.....	115
管理系统权限.....	116
继承和收回角色.....	118
授予角色给用户和收回授权.....	122
6 计算空间高级功能.....	125
6.1 管理定时任务.....	125
查看任务运行记录.....	125
创建定时任务.....	125

6.2 审计日志.....	127
6.3 数据脱敏.....	129
使用说明.....	129
启用数据脱敏.....	129
切换脱敏 Schema.....	131
添加表的脱敏规则.....	132
添加脱敏用户.....	133
卸载脱敏扩展.....	135
6.4 外部数据包装器（FDW）扩展.....	136
6.4.1 外部数据包装器（FDW）扩展使用说明.....	136
6.4.2 postgres_fdw.....	137
6.4.3 mysql_fdw.....	157
6.4.4 oracle_fdw.....	162
6.5 数据集成.....	172
6.5.1 数据集成简介.....	172
6.5.2 授予用户数据集成功能权限.....	175
6.5.3 信息总览.....	176
6.5.4 配置调度系统.....	179
6.5.5 创建数据源.....	181
6.5.6 配置表的映射关系.....	184
6.5.7 创建和执行 DDL 任务.....	188
6.5.8 创建离线同步任务.....	190
6.5.9 创建实时同步任务.....	202
6.5.10 管理元数据.....	215
6.6 数据共享.....	219
6.6.1 数据共享概述.....	219
6.6.2 关于数据共享权限.....	219
6.6.3 创建与管理数据共享.....	223
6.6.4 订阅数据共享与取消订阅.....	226
7 附录.....	231
7.1 PieProxy 外部接入示例.....	231
通过 Postgres Client PSQL 命令连接虚拟数仓.....	231
通过 Python psycopg2 连接虚拟数仓.....	231
通过 DBeaver 连接虚拟数仓.....	231
通过 PGAdmin4 连接虚拟数仓.....	234
通过 DataGrip 连接虚拟数仓.....	236
通过 Navicat 连接虚拟数仓.....	240
通过 JDBC URL 连接虚拟数仓.....	242
7.2 角色的系统权限清单.....	244
计算空间的系统权限.....	244

控制台的系统权限.....	245
---------------	-----

声明

法律说明

版权所有© 2024 杭州拓数派科技发展有限公司。

本手册仅为所涉及的软件产品的信息性介绍和帮助手册。本手册不可成为全面判断该软件产品具备或不具备某些功能和/或技术参数的依据，亦不可用于判定该软件产品是否符合某些要求、技术任务和/或参数以及其他第三方规范文件。

本手册中的所有资料为杭州拓数派科技发展有限公司（以下简称“OpenPie”或“拓数派”）所有并只能为产品购买者使用。事先未经 OpenPie 书面许可，本手册的任何部分均不得复制、篡改、转载于网络资源或利用通讯渠道或在大众媒体加以传播。

本手册描述的产品中，除非获得权利人的许可，否则，任何人不能以任何形式对前述软件进行复制、分发、修改、摘录、反编译、反汇编、解密、反向工程、出租、转让、分许可等侵犯软件版权的行为。

商标声明

OpenPie、PieCloudDB、πCloudDB、PieDataCS、πDataCS 及其他 OpenPie 相关的商标均为拓数派在中国大陆地区的注册商标。本文档中提及的其他注册商标、徽标和公司名称为其所有人所有。

免责声明

任何情况下 OpenPie 对本文档可能的任何错误和/或遗漏以及因其导致的产品购买方损失（直接或间接损失，包括未得利益）均不承担任何责任。

注意

本手册描述的产品及其附件的某些特性和功能，取决于当地网络的设计和性能，以及您安装的环境；本手册中产品经过详细的测试，但不能保证与所有的软硬件系统完全兼容，不能保证本产品完全没有错误。

本手册仅供参考，不构成任何形式的承诺，产品相关内容请以实际使用时的呈现为准。OpenPie 保留随时修改本手册中任何信息的权利，无需提前通知且不承担任何责任。

前言

文档概述

本文档旨在帮助数据库管理员和其他 IT 专业人员了解 PieCloudDB 数据库技术、管理方法和最佳实践。云原生数据库管理员需要掌握数据库设计、配置、监视、备份和恢复、性能优化和故障排除等方面的技能和工具，以确保数据安全、可用性和性能。

该文档主要介绍 PieDataCS 云原生平台的界面操作，为数据库管理员提供知识储备。

适用读者

本文档旨在为 PieDataCS 大模型数据计算系统的管理者（数据库管理员、数据库运维人员）提供参考信息，帮助其更好地管理与运维大模型数据计算系统。

本文档假设读者已掌握 Linux/Unix 系统管理、数据管理系统、数据库管理和 SQL 语言等方面的知识。

通用约定

约定	含义
危险	指示会出现严重的故障或问题，可能导致系统完全崩溃或无法正常运行。
警告	指示会出现重要的问题或异常情况，可能会对系统的功能或性能产生明显的影响，但系统仍然可以正常运行。
注意	指示可能会出现问题或异常情况，但不会对系统的正常运行产生严重影响。此类告警需要用户加以关注。
提示	提供附加信息、指导或补充说明。
>	多级菜单路径。
粗体	UI 页面的按键或菜单等元素。
等宽字体	命令、URL、示例中的代码、屏幕输出的文本或输入的文本等。
斜体	参数或者变量等。

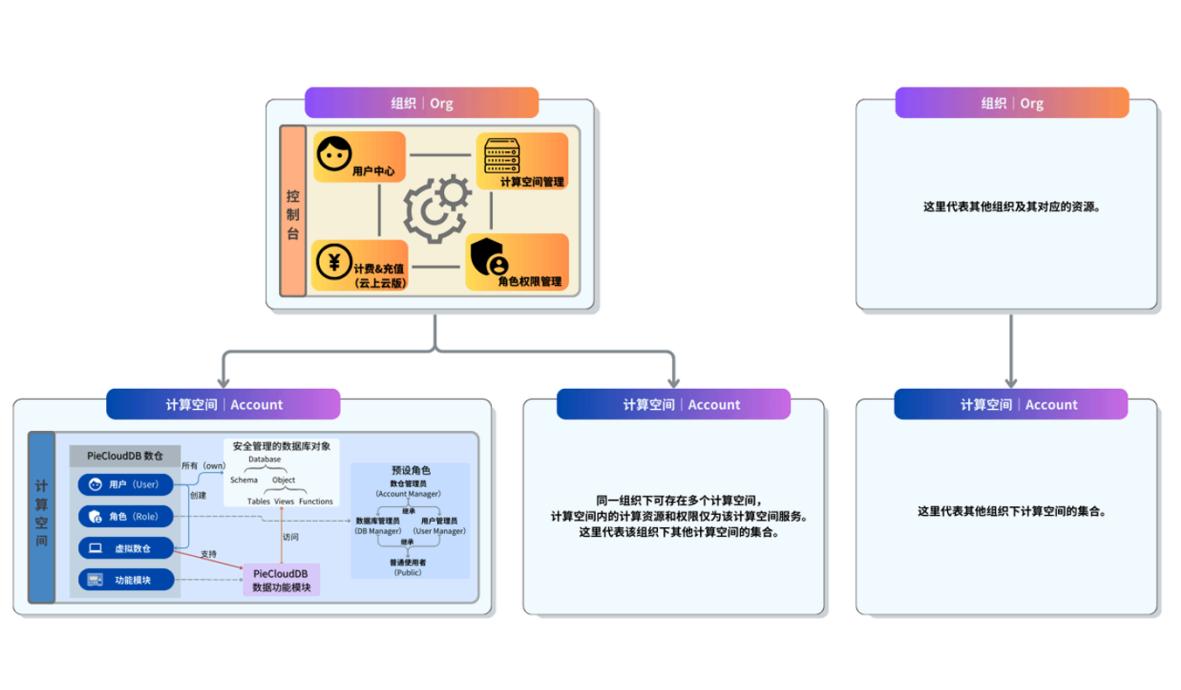
1 关于 PieDataCS 云原生平台

PieDataCS 云原生平台是一款以 PieCloudDB 为核心的可视化数据库管理平台，它用于管理部署实例上所拥有的实体物理资源，例如 CPU、内存、存储等。这些部署实例可以运行在物理硬件、容器或者虚拟机上。

在 PieDataCS 云原生平台注册账号成功后，系统会自动为初始用户创建一个组织以及该组织下的数仓。作为组织管理员，初始用户可以在「控制台」页面查看数仓列表信息，并支持在数仓下创建新用户，这些新用户将获得相应数仓的访问权限。用户在进入指定的“数仓”页面后，可以根据需求在「虚拟数仓」页面创建多个虚拟数仓。在数仓的「数据洞察」页面指定虚拟数仓和数据库之后，用户就可以运行 SQL 查询或执行 DML 操作来进行数据交互。

系统架构概览

PieDataCS 云原生平台的整体系统架构如下图所示。



PieDataCS 云原生平台将部署实例上所拥有的实体物理资源（如 CPU、内存、存储等）集合称为计算空间（Account）。对于 PieCloudDB，一个计算空间实质为一个物理数据仓库（简称数仓）。计算空间隶属于组织，并通过控制台进行统一管理。

控制台既是用户中心，负责管理组织下的所有用户，也具备计算空间的管理功能，同时还包括计费、充值等需要在组织层级控制的其他功能。

一个组织可以对应一个实体（个人、公司或其他组织），并且可以包含多个数仓。每个数仓都会同步建立对应的计算空间。组织下的用户可以加入不同的计算空间并进入数仓，从而在数仓中创建和使用各自的计算资源，即虚拟数仓。

每个数仓可以创建和管理多个虚拟数仓、数据库、用户和角色等。同一组织下各个数仓产生的所有存储和计算费用，都会以组织为单位进行结算。

云原生虚拟数仓相关概念

PieCloudDB 实例

数据库实例（Database Instance）是由操作系统进程和相关资源组成的独立且运行中的数据库环境。它可以被视为一个完整的、具有独立身份和特定功能的数据库系统，其中包括数据文件、日志文件、缓存区、网络端口等多个组件。

PieCloudDB 实例指的是已经安装完成的任意版本的 PieCloudDB 产品，它可以在物理硬件、容器或虚拟机上运行。无论其底层载体是什么，都可以被视为一个实例。例如，用户可以在自己的 K8s 集群上部署一套 PieCloudDB 实例，也可以在阿里云上部署一套 PieCloudDB 实例。用户通过访问 PieCloudDB 实例来获取数据库服务。

数仓

数据仓库，即数仓，是一种专门化的数据库系统，用于存储和管理企业级数据。它是一个集成数据的中央存储库，汇集了来自一个或多个不同源的数据。数据仓库将当前和历史数据存储在一起，用于为整个企业的员工创建分析报告。与传统的事务处理型数据库不同，数据仓库着重于数据的分析、报告和智能决策。数仓的主要特点包括：

- 面向主题：数据仓库以企业或组织为中心，按照业务领域或主题对数据进行组织和存储。这样的组织结构不仅更符合业务需求，也便于数据的查询、分析和统计。
- 集成性强：数据仓库通常汇集来自多个不同来源和系统的数据，通过对数据进行提取、转换和加载，实现数据的集成和统一管理。
- 时间性：数据仓库通常包含历史数据和当前数据，可以追溯数据的历史变化和趋势，支持时间序列分析和预测模型的建立与应用。
- 面向决策：数据仓库中的数据经过清洗、整理和加工，转化为更适合决策分析和报表统计的形式。可以通过 OLAP（Online Analytical Processing）等技术进行复杂查询和多维分析，为企业决策提供有力支持。

虚拟数仓

虚拟数据仓库（Virtual Data Warehouse），即虚拟数仓，是一种基于数据虚拟化技术的数据管理解决方案。它通过抽象化数据源和数据存储之间的逻辑关系，实现对分散、异构数据的整合和管理。虚拟数仓提供了统一的数据视图和访问接口，满足企业级应用和分析需求。虚拟数仓代表纯计算资源，是动态的，包括由后台云上的虚拟机提供的CPU、内存和临时存储的计算资源集群。

虚拟数仓中不存储真实数据，仅提供SQL执行服务。一个虚拟数仓至少包括一个协调器和一个执行器，每个节点都是一组CPU、内存和磁盘资源的虚拟资源集合。用户无法直接获取这些节点的信息，但对有权限的用户来说，虚拟数仓的启动、停止、扩展和缩减过程是可见的。

协调器负责响应客户端的SQL请求，生成查询计划，分发执行信息和元数据，并执行最终结果的汇总（如LIMIT、ORDER BY等）。每个集群有一个且只有一个协调器，通常只有协调器对用户可见，而执行器对用户不可见。但在某些特殊场景下，客户端也可能直接访问执行器。

执行器用于执行协调器分发的任务。每个集群可以有一个或多个执行器，它们可以并行执行同一个SQL查询任务。

组织

一个组织可以对应一个实体，无论是个人、公司还是其他组织形式，并且可以包含多个数仓。每个数仓都会同步建立其对应的计算空间。在一个组织下，可以有多个用户，这些用户可以加入不同的计算空间并进入数仓，并在数仓中创建和使用他们各自的计算资源，即虚拟数仓。

每个数仓能够创建和管理多个虚拟数仓、数据库、用户和角色等。同一组织下各个数仓产生的所有存储和计算费用，都将以组织为单位进行统一结算。

用户

用户（User）是指隶属于组织的用户，他们可以加入不同的计算空间。在同一个计算空间下，所有用户共享该计算空间的资源。

PieDataCS云原生平台的控制台和计算空间（数仓）各自拥有一套独立的权限管理体系和用户功能。用户可以先加入组织，然后再申请加入计算空间，或者直接在计算空间下创建新用户。用户可使用的系统功能和PieCloudDB数据库权限取决于他们各自角色的权限设置。计算空间所对应数仓下的用户功能允许查看当前数仓下所有用户的激活和审核状态。拥有用户管理权限的用户可以执行新增、删除、修改用户等操作。用户通过邮箱或手机号码进行身份区分，一个邮箱地址或手机号码仅对应一个用户。

角色

角色（Role）是系统和数仓（计算空间）中用于分配数据库权限的一种机制，代表了一个权限的集合。权限通过角色这个载体下放给用户，一个角色可以对应一个用户，也可以对应一类用户。用户与角色之间的关系不是一一对应的，也就是说，一个用户可以被授予多个角色，而一个角色也可以被授予给多个用户。每个用户至少会被赋予一种角色，而每一种角色都有其对应的权限集。

版本差异性说明

PieDataCS 云原生平台为 PieCloudDB 数据库内核提供了数据洞察、元数据浏览、用户管理、权限管理、SQL 查询历史、ETL 管理等功能的可视化界面，使用户能够通过直观的操作与 PieCloudDB 内核进行交互。

PieDataCS 云原生平台支持企业版和云上云版，两者在功能上存在差异，包括但不限于：

- 用户登录方式不同。

PieDataCS 云原生平台企业版的登录信息可由部署之后显示的打印信息获取，而对于云上云版，用户登录 OpenPie 官网 (<https://www.openpie.com/product-version#coc>) 即可进入产品界面。

- 虚拟数仓的外部接入方式不同。

云上云版的虚拟数仓仅支持 PieProxy 作为外部接入方式，而企业版则提供了 Node Port 和 PieProxy 两种选项。在企业版中，默认的外部接入方式为 Node Port。

- 云上云版独有功能如下：

- 计费模式
- 费用中心
- 我的工单
- 免费试用
- 用量统计

- 企业版独有功能如下：

- 定时任务
- 审计日志
- 数据脱敏
- FDW
- 数据集成

2 权限管理体系概览

PieDataCS 云原生平台的权限管理体系以角色为基本管理单位，角色是用户权限概念的具体化体现。

权限分类

当用户访问 PieCloudDB 中的数据时，PieDataCS 云原生平台会从网络连接、数据加密、数据库对象权限三个维度对数据访问权限进行约束和限制。权限分配所涉及的相关概念包括：

- 数据库对象的访问权限

数据库对象是数据库中用于管理数据的实体，包括表、视图、Schema 和序列（Sequence）等。

- 角色

角色是一组权限的集合，具备继承和授予的功能，可以将权限授予给其他用户或角色。

- 用户

用户通常指拥有登录凭据和访问权限的个人或应用程序。

- 系统权限

系统权限指的是云原生平台的功能使用权限。

需要注意的是，PieDataCS 云原生平台的权限系统会因版本不同而略有差异，并且控制台界面和计算空间操作界面各自拥有一套独立的权限体系。

控制台界面的权限仅涉及系统权限，详细信息请参见 [管理组织角色](#)。

而计算空间操作界面的权限分为两个部分：PieDataCS 系统权限和 PieCloudDB 数据库对象权限。通过在计算空间下建立的角色，既可以控制数据库对象权限，也可以控制计算空间的系统权限。数据库对象权限源自数据库中的角色权限，所有者可以将数据库对象权限下放给特定角色。对于在计算空间中创建的角色，其在数据库中的权限与计算空间下的数据库对象权限保持一致。有关系统权限的详细信息，请参见 [管理系统权限](#)。

计算空间操作界面中的功能因其特性不同，对权限的要求也各不相同，通常可以分为以下三类：

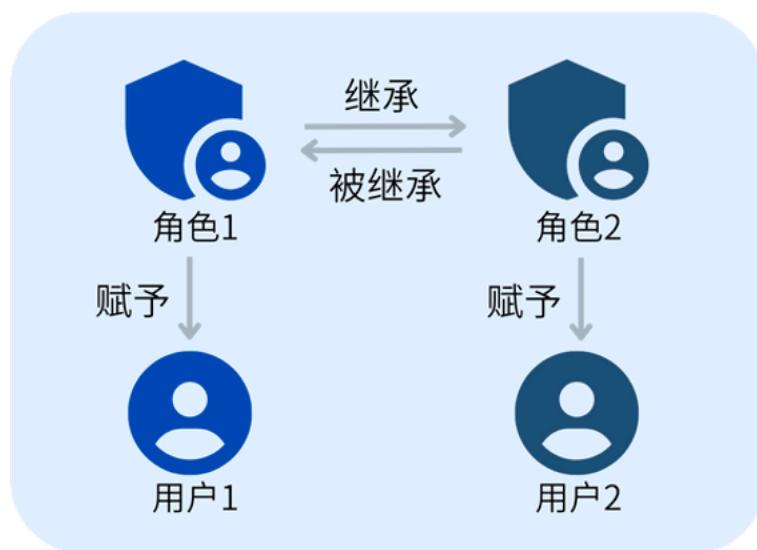
- 功能仅涉及计算空间的系统权限。
- 功能仅涉及数据库对象权限。

- 功能同时涉及系统权限和数据库对象权限。

在计算空间操作界面，有关数据库对象权限的详细信息，请参见 [管理数据库对象权限](#)。

角色的继承与赋予

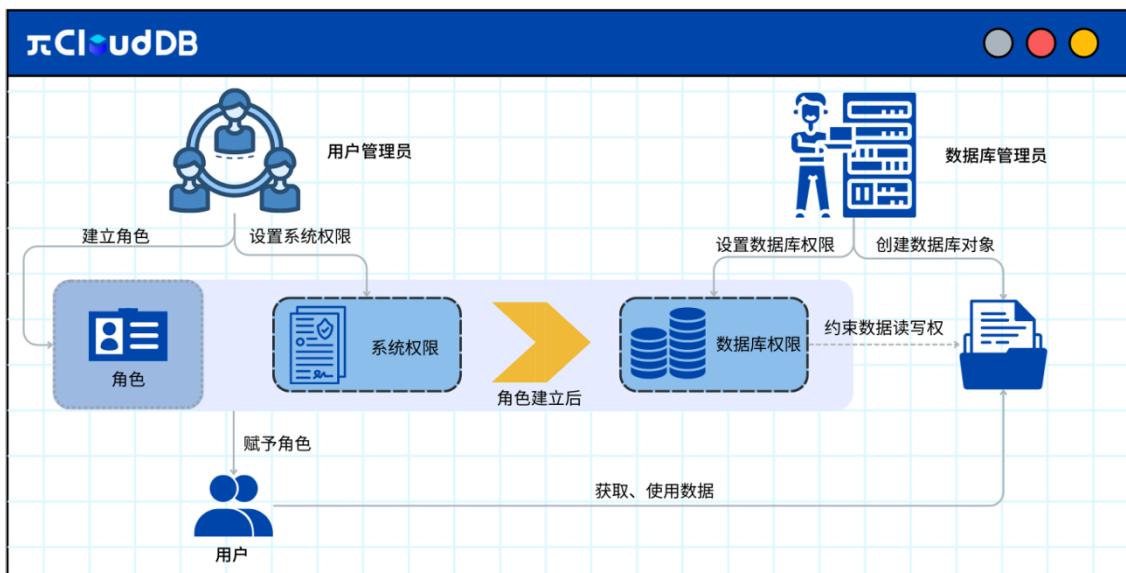
继承是指将权限以角色的形式传递给另一个角色的过程。在层级结构中，继承的角色应高于被继承的角色，且继承角色的权限与被继承角色的权限保持一致。这意味着，即使收回了继承角色所继承的权限，该权限状态也会与被继承角色保持一致，而不会被实际收回。一个角色可以继承另一个角色的权限，且继承的权限即使被改动，其状态也会与被继承角色相应的权限保持一致。这种继承关系在各角色间形成了一个层次结构。



赋予是指将权限以角色的形式下放给用户的行为。拥有角色管理权限的用户或角色，可以将特定角色及其关联的权限授予给其他用户。

角色的赋权流程

PieDataCS 云原生平台支持创建自定义角色并设置权限，一般流程如下图所示。



假设数仓管理员已将用户管理员及数据库管理员赋予了相应的用户，用户管理员可为新用户组建立自定义角色，并对该角色的系统权限进行设置。设置完毕后，用户管理员再将该角色赋予用户组。

在自定义角色设立后，数据库管理员创建该角色所对应的数据库对象，并将相应的数据库对象权限赋予该角色，通过数据库对象权限来约束相应数据的读写权限。角色权限管控下的用户组，可通过其对应的角色，在适当的权限条件下探索数据。

在数仓中的用户（除数仓管理员外）未被赋予任何角色的情况下，数仓管理员需要先设置好用户管理员及数据库管理员，再由这两个角色对新角色权限进行设置。

本文将通过具体示例详细介绍 PieDataCS 云原生平台的赋权流程。

示例：假设某公司需要对某几类线上产品的销售额进行存储和分析。该公司的数据管理团队决定由 David、Betty 和 Charlie 三位负责管理这部分数据及其附属的权限，并支持业务部门的 Adam 进行销售数据分析工作。

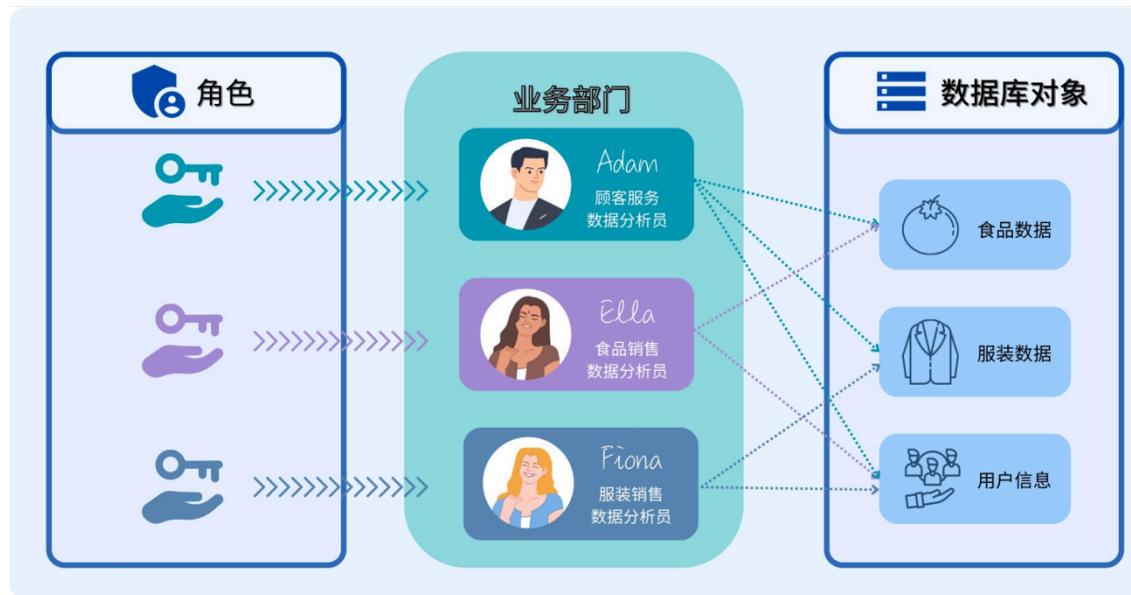
针对本示例的赋权步骤如下：

1. 设置用户管理员及数据库管理员：David作为数仓管理员，将用户管理员角色赋予 Betty。Betty 被赋予用户管理员权限后，开始承担用户权限的管理职责，并将数据库管理员角色赋予给 Charlie。
2. 用户管理员创建自定义角色并赋予系统权限：各管理员角色就位后，Betty 为 Adam 创建其对应的角色，并设置系统权限，然后将该角色赋予给 Adam。
3. 数据库管理员创建虚拟数仓并赋予数据库对象权限：根据 Adam 的需求，Charlie 为其角色创建对应的数据库对象，并将这些数据库对象的权限分配给 Adam 的角色。

通过类似的步骤，可以根据每组用户的特性（例如数据类型、功能需求等）来创建对应的角色。

在上述示例中，用户管理员 Betty 和数据库管理员 Charlie 可以使用同样的步骤为来自其他业务部门的 Ella 和 Fiona 建立相应的角色，并根据她们的需求赋予系统和数据库对象权限。

下图展示了 Adam、Ella 和 Fiona 在数据库中的读取权限。



Adam、Ella 和 Fiona 都来自业务部门，他们分别属于不同的团队，并且对数据权限和开发需求各有不同。PieCloudDB 的权限系统允许这些用户通过分配的角色，根据各自的角色权限安全地进行数据的读取和写入操作。这使得他们能够按照自己的需求进行数据清理、业务分析和开发工作。

3 快速上手

3.1 云上云版免费试用指导

为满足使用者的试用需求，PieDataCS 云原生平台的云上云版本提供免费试用 30 天的服务。

适用场景

适用于个人学习、产品测试等数据体量偏小的场景，不建议用于涉及企业核心业务的高可用场景。

试用说明

- 创建虚拟数仓的数量上限为 2 个，执行节点的数量上限为 4 个。
- 上传数据的体量上限为 1G。
- 下载数据的体量上限为 100M。

申请方式

1. 登录 OpenPie 官网 (<https://www.openpie.com/product-version#coc>) 并进入产品界面。
2. 在产品界面，点击云上云版的 **免费试用** 即可进入 PieDataCS 云原生平台的登录页面。



3. 点击下方的 **没有账号?立即注册** 以跳转到注册账号页面。



4. 默认使用邮箱注册，根据系统提示填写用户名、邮箱、邮箱验证码和密码。



如果需要切换为使用手机号码注册，点击右上方的 **切换至手机注册** 即可，并填写用户名、手机号、手机验证码和密码。



5. 阅读并同意《隐私政策》和《服务条款》后，点击 **完成注册** 即可自动跳转到系统为其分配的默认计算空间操作界面。

初始用户在 PieDataCS 云原生平台注册账号成功后，即会自动创建一个组织和一个组织下的计算空间。初始用户作为组织管理员可以在控制台的「**计算空间**」页面查看计算空间列表信息，并支持在计算空间下创建新用户，该用户即可拥有该计算空间的访问权限。

进入指定计算空间页面后，初始用户可以使用数据洞察、数据库和查询历史的功能。在「**数据洞察**」页面选择目标虚拟数仓和数据库，就可以执行 SQL 查询等进行数据交互。如果用户被赋予了相应的管理权限，还可以根据需求在「**虚拟数仓**」页面创建虚拟数仓。有关 SQL 查询的基础操作步骤，请参见 [SQL 查询基础操作](#)。

3.2 SQL 查询基础操作

本文通过具体示例来介绍已注册用户如何快速开始使用 PieDataCS 云原生平台的计算空间来执行 SQL 查询，一般操作流程如下：

1. 登录组织下的计算空间
2. 创建虚拟数仓
3. 管理虚拟数仓
4. 创建数据库和数据库对象
5. 执行 SQL 查询并展示查询结果

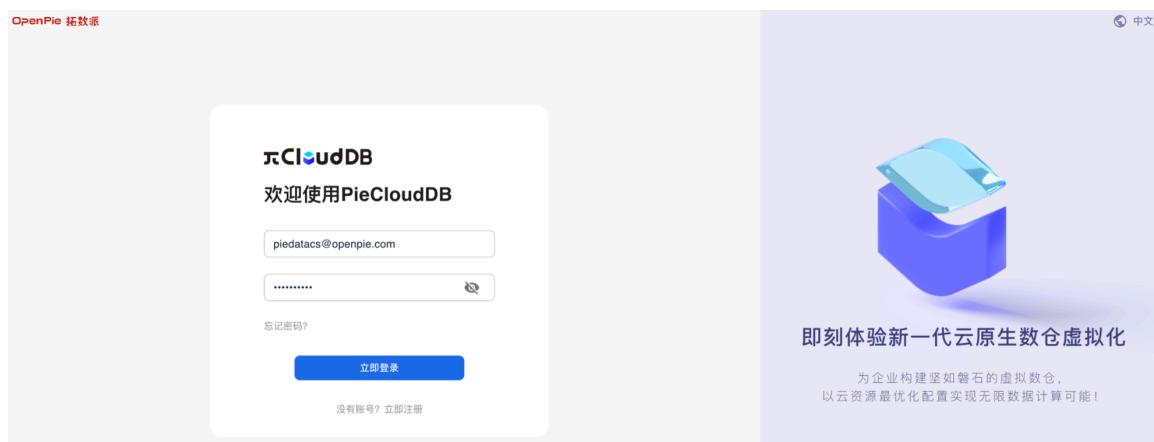
注意：

创建虚拟数仓和管理虚拟数仓操作需要相应的系统权限，一般由数仓管理员（accountmanager）、数据库管理员（dbmanager）或具有相应管理权限的用户来操作。

登录组织下的计算空间

1. 进入平台登录页面，输入登录账号（邮箱或者手机号）和密码，点击 **立即登录**，即可进入“选择计算空间登录”的页面。

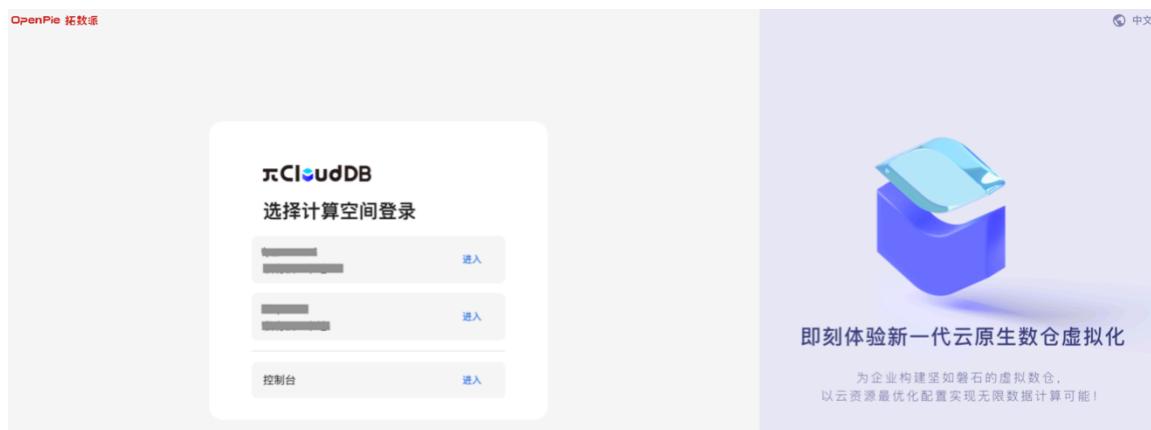
对于 PieDataCS 云原生平台的云上云版本，用户在浏览器输入网址：<https://app.pieclouddb.com>，即可进入 PieDataCS 云原生平台的登录页面；而对于 PieDataCS 云原生平台企业版，需要使用登录 URL 信息，该信息可由部署之后显示的打印信息获取，即“Console”信息（例如 <http://xx.xx.xx.xx:xxxxx>）。



提示：

- 用户命名规则：用户名必须以英文字母开头，可以包含英文字母（区分大小写）、数字和下划线，最长为 63 个字符。
- 密码规则：密码为 8 位以上，必须包含大小写字母、数字和特殊符号 (!@#~\$%^&*()|_)。
- 验证码：如果没有配置验证码服务，则无需输入验证码。

2. 在“选择计算空间登录”页面，选择所需的计算空间并点击 **进入**，即可登录计算空间的功能页面。

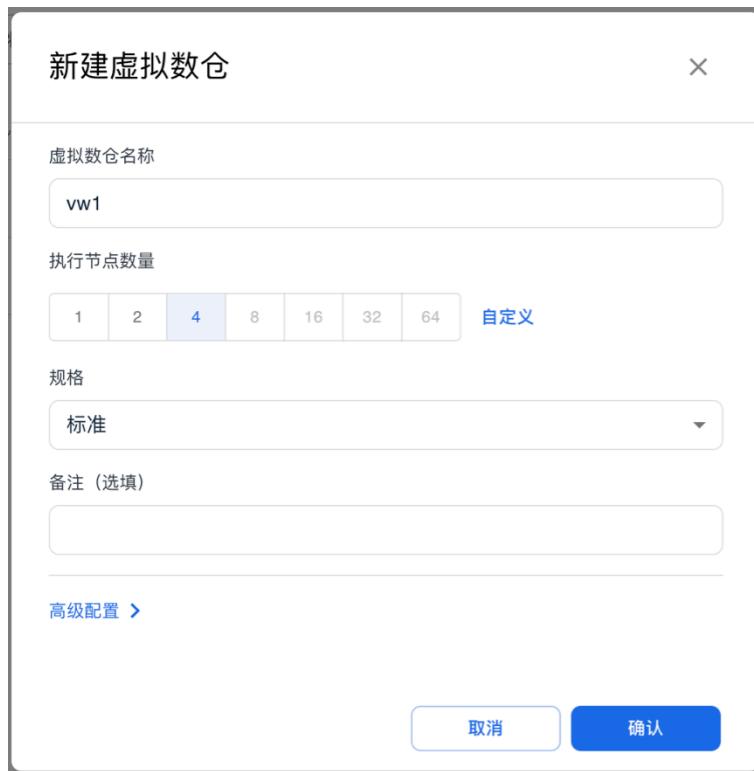


对于初始用户，如果没有加入其他计算空间，则系统会在其注册并立即登录后自动跳转至注册时为其分配的默认计算空间的功能页面。

创建虚拟数仓

在「**虚拟数仓**」页面，数仓管理员（accountmanager）、数据库管理员（dbmanager）或具有相应管理权限的用户可以创建虚拟数仓，具体操作步骤如下：

1. 在「**虚拟数仓**」页面点击 **新建虚拟数仓**，即可进入虚拟数仓的创建页面。
2. 输入自定义的虚拟数仓名称，并选择执行节点数量和节点规格。



3. 展开“高级配置”选项，设置虚拟数仓的自动启停功能。

- 如果打开“自动启动”功能，则当虚拟数仓处于关闭状态而用户尝试运行查询命令、数据导入、外部接入等功能时，虚拟数仓会自动启动并运行相关任务。
- 如果打开“自动关闭”功能，当虚拟数仓在没有运行任务时，会等待至所设置的“自动关闭等待时长”后自动关闭。系统默认的自动关闭等待时长为 10 分钟。



4. 点击确认。一个新的虚拟数仓便可创建完成并自动同步到列表中。

注意：

PieDataCS 云原生平台的云上云版的试用账号可以最多创建两个虚拟数仓，且默认开启自动启停功能。

管理虚拟数仓

在「虚拟数仓」页面，数仓管理员（accountmanager）、数据库管理员（dbmanager）或具有相应管理权限的用户可以在虚拟数仓列表中，选择目标虚拟数仓并点击隐藏菜单「...」，并根据对应选项执行虚拟数仓的启停、扩缩容等管理操作。

启停虚拟数仓

- 当虚拟数仓处于“运行中”状态时，点击**关闭**选项，即可关闭运行中的虚拟数仓。

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	运行中	4	标准	不开启	未开启	按量计费	10分钟前	查看详情	查看日志
vw2	运行中	4	标准	不开启	未开启	按量计费	36分钟前	查看详情	查看日志

每页 10 条 | < | < | 1 | > | > | 前往 | [删除](#)

- 当虚拟数仓处于“已关闭”状态时，点击**启动**选项，即可启动已关闭的虚拟数仓。

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	已关闭	4	标准	不开启	未开启	按量计费	9分钟前	查看详情	查看日志
vw2	运行中	4	标准	不开启	未开启	按量计费	35分钟前	查看详情	查看日志

每页 10 条 | < | < | 1 | > | > | 前往 | [删除](#)

- 当虚拟数仓处于“失败”、“关闭中”或者“启动中”状态时，点击**强制关闭**选项，即可强制关闭该虚拟数仓。

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	启动中	4	标准	不开启	未开启	按量计费	12分钟前	查看详情	查看日志
vw2	运行中	4	标准	不开启	未开启	按量计费	38分钟前	查看详情	查看日志

每页 10 条 | < | < | 1 | > | > | 前往 | [页](#)

虚拟数仓扩缩容

虚拟数仓可以通过修改执行节点数量来实现扩缩容。当虚拟数仓处于“已关闭”或“运行中”状态时，点击**修改**选项，即可进入目标虚拟数仓的配置界面。

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	运行中	4	标准	不开启	未开启	按量计费	33分钟前	查看详情 查看日志	修改
vw2	运行中	4	标准	不开启	未开启	按量计费	1小时前	查看详情 查看日志	修改
共 2 条									
每页	10	条	<	<	1	>	>	前往	

在虚拟数仓的配置页面，用户可以通过减少执行节点的数量来执行缩容操作。例如，将节点数量从“4”调整为“2”。一旦确认变更，设置会立即生效，从而将虚拟数仓的执行器数量从4个缩容到2个。



同理，用户可以通过增加虚拟数仓的执行节点数量来执行扩容操作。

创建数据库和数据库对象

在「数据洞察-文件」页面，用户可以根据如下步骤创建数据库和数据库对象：

1. 点击 **新建>新建 SQL 文件**，即可新建一个的 SQL 文件。
2. 在所打开的 SQL 文件的编辑区中，输入 SQL 命令 `CREATE DATABASE` 来创建一个新的数据库（例如 testdb）。

```
CREATE DATABASE testdb;
```

3. 选择用于执行该 SQL 语句的数据库（一般使用默认数据库“openpie”）和一个已创建的虚拟数仓。

4. 在当前 SQL 文件的编辑区，选中创建数据库的 SQL 语句后，点击 **执行选中** 以运行该命令。



The screenshot shows the πDataCS interface. In the top navigation bar, there are dropdown menus for 'openpie' and 'vw1 (运行中)', a '保持会话' (Keep Session) toggle, and several buttons: '执行选中' (Execute Selected), '全部执行' (Execute All), and '执行计划' (Execute Plan). The '执行选中' button is highlighted with a red box. Below the toolbar, the SQL editor contains the command: '1 CREATE DATABASE testdb;'. Underneath the editor, the results pane shows a green success icon, the request ID 'qu01jfcb87wjgj297wp7n62j0ch6#0', the execution time '666ms', and the number of affected records '0'. The message 'ok' is also displayed.

执行结果显示，示例数据库已成功创建。

5. 在“数据库”下拉列表中选择并切换到目标数据库（例如 testdb）。虚拟数仓可以保持不变。
 6. 在当前 SQL 文件的编辑区，输入 SQL 命令 `CREATE TABLE` 来创建目标数据库下的表，并使用 `INSERT INTO` 插入一些数据。

如下示例为，创建数据库 testdb 下的表 test_table，并插入一些数据。

```
CREATE TABLE test_table
(id CHAR(10),
name CHAR(50),
length INT,
date CHAR(10),
type CHAR(20)
);

INSERT INTO test_table VALUES
('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'),
('HG120', 'The Dinner Game', 140, '1985-02-10', 'Comedy');
```

7. 在当前 SQL 文件的编辑区，选中上述 SQL 语句后，点击 **执行选中** 以运行该命令。



The screenshot shows the πDataCS interface with the database set to 'testdb'. The 'Execute Selected' button is highlighted with a red box. The SQL editor contains the following code:

```
1 CREATE DATABASE testdb;
2
3 v CREATE TABLE test_table
4   (id CHAR(10),
5    name CHAR(50),
6    length INT,
7    date CHAR(10),
8    type CHAR(20)
9  );
10
11 v INSERT INTO test_table VALUES
12   ('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'),
13   ('HG120', 'The Dinner Game', 140, '1985-02-10', 'Comedy');
```

Below the editor, the results pane shows two successful operations: '1 CREATE TABLE...' and '2 INSERT INTO ...'. Each operation has a green success icon, the request ID, the execution time (279ms), and the number of affected records (2). The message 'ok' is also displayed.

执行结果显示，示例表已成功创建，并且示例数据已成功插入。

执行 SQL 查询并展示查询结果

在「数据洞察-文件」页面中打开的 SQL 文件的编辑区，用户首先需要选择要查询的数据库（例如 testdb）和已创建的虚拟数仓（例如 vm1）。然后，输入 `SELECT` 语句（例如 `SELECT * FROM test_table;`）。最后，选中该语句并点击 **执行选中** 来执行 SQL 查询。执行 SQL 查询语句成功后，系统会自动显示查询结果。



The screenshot shows the OpenPie interface with the following details:

- Top navigation bar: Includes dropdowns for "testdb" and "vm1 (运行中)" (Running), a "保持会话" (Keep Session) toggle, and search/filter icons.
- Tool buttons: "执行选中" (Execute Selected), "全部执行" (Execute All), and "执行计划" (Execution Plan).
- SQL Editor area:
 - Code content:

```
1 v CREATE TABLE test_table
2 (
3     id CHAR(10),
4     name CHAR(50),
5     length INT,
6     date CHAR(10),
7     type CHAR(20)
8 );
9
10 v INSERT INTO test_table VALUES
11     ('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'),
12     ('HG120', 'The Dinner Game', 140, '1985-02-10', 'Comedy');
13
14 SELECT * FROM test_table;
```
 - Execution status message: "● 执行成功 请求ID: qu01jfs1bwp6nsg9rjy1tvwhrd3f#0 扫描时间: 14ms 执行时间: 476ms 查询记录数: 2"
 - Result summary: "结果集大小: 221 Bytes"
 - Result table:

	id	name	length	date	type
1	HG120	The Dinner Game	140	1985-02-10	Comedy
2	B6717	Tampopo	110	1985-02-10	Comedy

SQL 查询结果包括该 SQL 语句的执行状态、请求 ID、扫描时间、执行时间以及受影响的记录数等详细信息。

如果在执行 SQL 语句的过程中发生错误，用户可以根据提示信息进行问题排查并修订，随后再次执行查询。

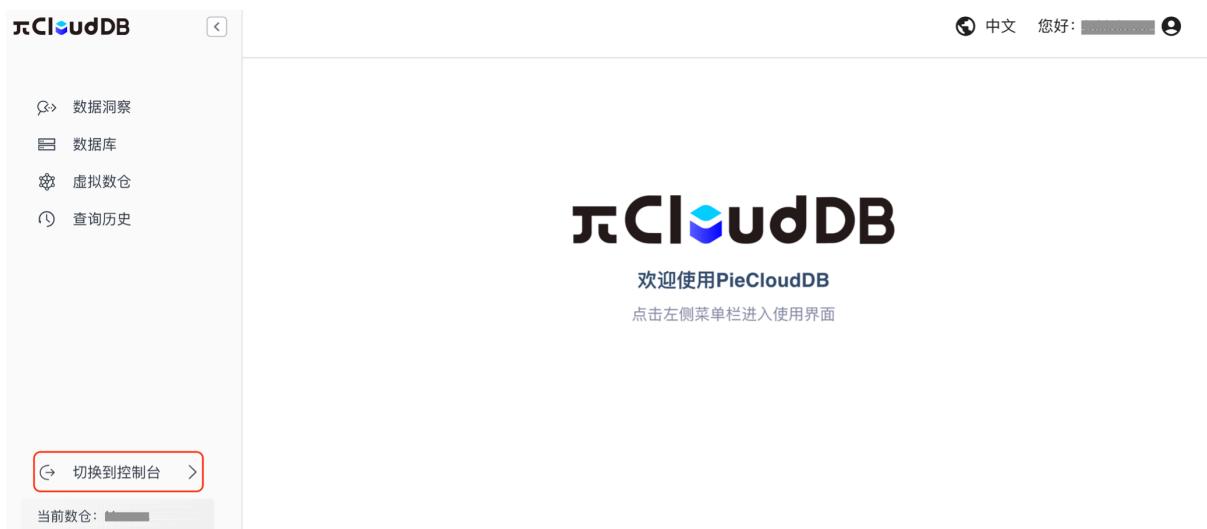
4 控制台管理指南

4.1 控制台功能概览

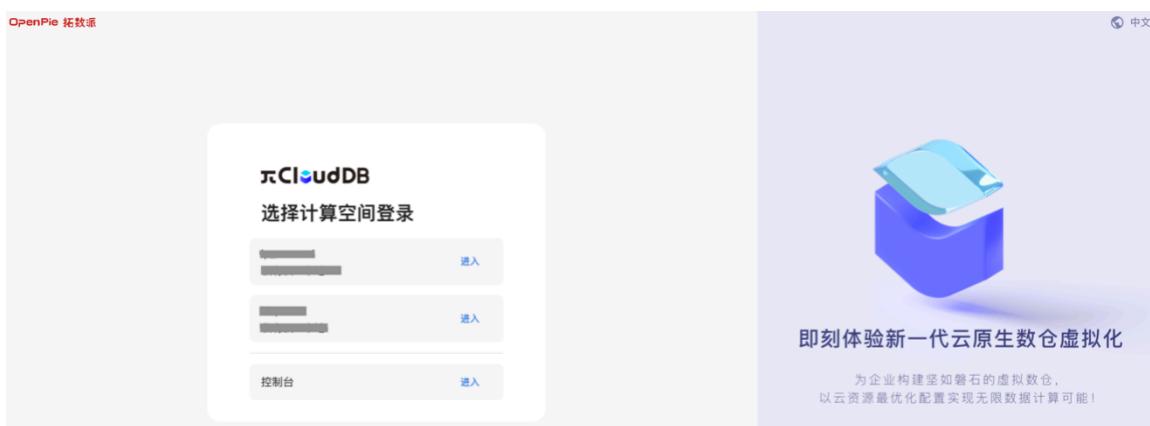
控制台具有组织、计算空间、用户、费用、角色权限等多方位管理功能。

在 PieDataCS 云原生平台的登录页面，云上云版用户输入登录账号（邮箱或者手机号）和密码后，点击 **立即登录**，之后选择“控制台”并点击 **进入** 即可进入「控制台」页面。

对于初始用户，如果没有加入其他计算空间，则系统会在其注册并立即登录后自动跳转至注册时为其分配的默认计算空间页面，此种情况下，初始用户在该计算空间页面点击**切换到控制台**，即可快速切换到对应的控制台页面。



在 PieDataCS 云原生平台的登录页面，企业版用户输入登录账号（邮箱或者手机号）和密码，点击 **立即登录**，之后选择“控制台”并点击 **进入** 即可进入「控制台」页面。



而对于云上云版用户，输入登录账号（邮箱或者手机号）和密码后，点击 **立即登录** 之后，系统只提供“计算空间”选项，选择目标计算空间并点击 **进入** 即可进入「**计算空间**」页面。在该计算空间页面点击 **切换到控制台**，即可快速切换到对应的控制台页面。

「**控制台**」页面主要提供如下功能：

- **管理组织**

三方身份认证配置。

- **管理空间**

管理在组织下所创建的计算空间。一个计算空间实质为一个数仓，包含用户数据、元数据、虚拟数仓等。

- **管理用户**

查看当前组织下所有用户的信息。

- **管理角色**

查询控制台中各角色的关系，以及查看和管理角色的各项权限。

- **管理计费与订阅**

支持按量计费和订阅包年包月服务。

提示：

仅 PieDataCS 云原生平台的云上云版本支持计费与订阅管理的功能。

- **费用中心**

展示资源用量及其对应的花销。

提示：

仅 PieDataCS 云原生平台的云上云版本支持费用中心的功能。

- **我的工单**

创建和管理工单，向 PieDataCS 云原生平台的技术支持团队进行提问和反馈。

提示：

仅 PieDataCS 云原生平台的云上云版本支持我的工单功能。

4.2 管理组织

组织（Organization）在用户层面对应一个注册实体（例如个人用户、企业用户等），在架构层面对应计算空间和用户的管理单位。

创建组织

初始用户在 PieDataCS 云原生平台登录页面注册账号成功后，即会自动创建一个组织和一个组织下的计算空间。如果组织下的新增用户想要登录计算空间使用数据计算服务，需要申请加入对应的计算空间，或由管理员手动将其加入该组织下的计算空间中。

初始用户可以根据如下步骤创建组织：

1. 在登录页面，点击下方的 **没有账号？立即注册**，即可跳转到注册账户页面。



2. 填写用户名、登录凭据信息（支持邮箱或手机号）、密码和确认密码。
 - 用户命名规则：用户名必须以英文字母开头，可以包含英文字母（区分大小写）、数字和下划线，最长为 63 个字符。
 - 密码规则：密码为 8 位以上，必须包含大小写字母、数字和特殊符号（!@#~\$%^&*()+=_）。
 - 验证码：如果没有配置验证码服务，则无需输入验证码。
3. 填写好信息后点击 **完成注册**，如果显示注册成功的信息，页面会自动跳转进入计算空间并显示正在初始化。

4. 点击右上角用户信息栏的 **控制台** 选项，如果查看到组织信息和空间信息，这说明初步创建组织和空间成功。



添加组织用户

在「**控制台**」界面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户点击菜单栏中的**用户管理**即可进入对应的功能页面并查看该组织下的用户信息。

组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以使用控制台的“**邀请注册**”功能，将用户加入到当前组织。具体操作步骤如下：

1. 在「**控制台**」界面展开右上角的菜单，点击**邀请注册**。



2. 复制弹窗中的邀请链接，将链接地址发送给需要被添加到组织的用户。



- 新用户点开链接后会进入到邀请注册页面。凭借收到的邀请码，输入相关信息和验证码完成注册。

注册成功后，该用户会自动加入组织，同时组织管理员可以在控制台的「**用户管理**」菜单下的用户列表中查看到该用户的基本信息。

- 新用户申请加入计算空间。

注意：

新增用户加入组织后，如果需要登录计算空间并使用数据计算服务，则要申请加入计算空间，或由组织管理员手动将其加入组织下的计算空间中。

组织管理员或该组织下具有相应用户管理权限的用户审核通过后，新用户即可登录控制台的「**计算空间**」页面，在计算空间列表中找到已申请加入的目标计算空间，在“操作”字段下点击**进入空间**即可。

配置组织第三方身份认证

PieDataCS 云原生平台支持配置 LDAP（轻量级目录访问协议）功能，该功能兼容 LDAP 和 LDAPS 协议，并通过身份验证和授权实现统一的访问控制和用户管理。

在「**控制台**」界面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户点击菜单栏中的**组织管理**即可进入对应的功能页面并配置三方身份认证。

组织管理

中文 您好: [用户名]

组织信息

组织名称

组织ID

or01j5mr204ndv1s2gacw62eze7n

组织类型

普通

用户身份认证

添加身份认证

身份认证类型

默认计算空间ID

需要管理员审核

启用状态

登录入口

操作

LDAP

否

 已停用

添加身份认证的具体操作步骤如下：

- 在控制台的「组织管理」页面，点击 **添加身份认证**，即可进入 LDAP 配置页面。
- 在 LDAP 配置页面，选择用户通过 LDAP 功能登录默认被分配的计算空间，如果不选则默认用户只加入组织。
- 选择“首次登录是否需要管理员审核”，默认为不需要审核。
- 填写 LDAP 或 LDAPS 地址、Bind DN（管理员 DN）、密码（管理员密码）、Base DN、搜索过滤器等信息。
- 点击 **测试连接**，以验证 LDAP 配置是否可以连通。

添加 LDAP 配置

用户默认的计算空间（选填）

首次登录需要管理员审核

LDAP / LDAPS 地址
ldap://[REDACTED]:389

LDAP 使用明文传输，建议使用加密传输的LDAPS，提高安全性。

Bind DN

密码

Base DN

搜索过滤器
(sAMAccountName=%s)

LDAP 连接测试

连通性测试成功

取消 添加

6. 连通性验证成功后，点击 **添加**，即可将 LDAP 配置加入身份认证配置列表，且默认为停用状态。
7. 在用户身份认证配置列表中，用户点击目标身份认证配置所在行的 **启用**，即可启用 LDAP 服务。

用户身份认证					
身份认证类型	默认计算空间ID	需要管理员审核	启用状态	登录入口	操作
LDAP		否	<input checked="" type="radio"/> 已停用		删除 启用 修改配置

启用后，目标身份认证配置所在行的登录入口即可显示登录信息。该组织下的用户可以直接复制登录入口提示的 URL 信息进行登录。

用户身份认证					
身份认证类型	默认计算空间ID	需要管理员审核	启用状态	登录入口	操作
LDAP		否	<input checked="" type="radio"/> 已启用		删除 复制

根据管理员配置的规则，用户需要输入用户名和密码。通常，用户名是邮箱地址，或者是用户名一般为邮箱地址或去掉邮箱域名后缀的前部分对应的用户名。

用户名具有如下固定的规则：

- 用户名必须以英文字母开头，可以包含英文字母（区分大小写）、数字和下划线，最长为 63 个字符。
- 组织下的用户名可能包含特殊字符等不符合规则的情况，此时需要用户修改用户名以在组织内使用，修改符合规则后方可登录成功。
- 首次登录成功后，当再次登录时，用户依然使用的是原来企业内部的用户名和密码。但在 PieDataCS 云原生平台内，用户的用户名会显示成适配数据库规则后的用户名。

4.3 管理计算空间

计算空间包括在部署实例上所拥有的实体物理资源，包括 CPU、内存、存储等。

在「控制台」界面，用户点击菜单栏中的 **计算空间** 即可进入对应的功能页面。该页面支持如下操作：

- 访问计算空间
- 新建计算空间
- 修改计算空间名称
- 管理计算空间用户

- 管理计费与订阅

提示：

仅 PieDataCS 云原生平台的云上云版本支持管理计费与订阅功能。

访问计算空间

在组织下新建的计算空间具有完全独立的存储资源（用户数据）和计算资源（虚拟数仓）。初始用户注册成功后自动创建一个组织和一个计算空间，初始用户作为组织管理员可以在控制台的「**计算空间**」页面查看所有计算空间的信息。

计算空间名称	计算空间ID	空间类型	状态	创建时间	空间用户	操作
Uranus	ac01j5mr204s0b3pg7y8c6za8021	PieCloudDB 虚拟数仓	正常	2024-08-19 15:19:07	查看详情	进入空间

在控制台的「**计算空间**」页面，用户可以通过以下两种方式进入计算空间：

- 在计算空间列表中找到目标计算空间，点击“操作”字段下的**进入空间**，即可进入指定的计算空间。
- 在计算空间列表中找到目标计算空间，鼠标放置计算空间名称处会出现“URL链接”标识，点击即可复制登录 URL，通过 URL 登录云原生平台并直接跳转进入目标计算空间。

新建计算空间

在控制台的「**计算空间**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以创建 PieCloudDB 虚拟数仓的计算空间，具体操作步骤如下：

- 点击**新建空间**并在下拉列表中选择目标计算空间类型（本文以“PieCloudDB”为例），即可进入计算空间创建页面。

- 针对 PieCloudDB 虚拟数仓，在创建页面需要选择云服务厂商和地域，输入计算空间名称，并从下拉列表选择该组织下的指定用户作为初始管理员。

提示：

如果所选择的计算空间初始管理员是组织下的其他用户，则该计算空间的管理权限归该用户所有。

- 点击 **完成**，计算空间创建成功后会自动跳转至计算空间详情页面。

如果步骤 2 中选择的初始管理员为当前计算空间的操作用户，则可以在空间详情页面点击 **进入空间**，直接进入该计算空间对应的一个新的、独立的数仓。

修改计算空间名称

在初始化创建组织和计算空间时，系统会随机分配一个计算空间名称。新建计算空间时，用户也需要输入一个新的计算空间名称。此外，已有的计算空间名称也支持修改，以便管理员能够更有效地管理组织下的计算空间。

在控制台的「**计算空间**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以修改计算空间的名称，具体操作步骤如下：

- 在计算空间列表中，点击目标计算空间所在行的“操作”栏下的**修改名称**，即可进入修改页面。



计算空间							操作
空间类型	计算空间ID	空间类型	状态	创建时间	空间用户	操作	
PieCloudDB 虚拟数仓	ac01j5mr204s0b3pg7y8c6za8021	共享存储	正常	2024-08-19 15:19:07	查看详情	进入空间	⋮
Uranus	ac01j5z27qcmg49pdb0jn0f0gmt	PieCloudDB 虚拟数仓	正常	2024-08-23 15:29:25	查看详情	进入空间	修改名称
demo1							查看详情

- 输入新的计算空间名称。
- 点击**确定**，即可完成修改。

管理计算空间用户

组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以将组织下已创建的用户添加到计算空间中或将该用户从计算空间中移除；也支持在计算空间下创建新用户，在计算空间下创建的新用户也可以被添加到其他计算空间中。

在控制台的「**计算空间**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户点击目标计算空间所在行的“空间用户”栏下的**查看详情**，即可进入该计算空间的详情页面。

The screenshot shows the 'Space Details' page for a space named 'Uranus'. At the top, there's a header with '计算空间 / Uranus', a language switch ('中文'), and a greeting ('您好: [redacted]'). Below the header, there's a 'Space Details' section with the following information:

名称	Uranus	ID	ac01j5mr204s0b3pg7y8c6za8021
类型	PieCloudDB 虚拟数仓	云服务商	私有云
地域	本地	状态	正常
创建时间	2024-08-19 15:19:07		

Below the details, there's a 'Space Users' section with a table:

用户ID	用户名	邮箱	手机号	验证状态	验证方式	审核状态	创建时间	操作
us01j8ps2h4k81dp3kd1h72sr9zc	testuser02	[redacted]		已验证	自动跳过	审核通过	3 个月前	移出用户

At the bottom right of the 'Space Users' section, there are two buttons: 'Create New User' and 'Add User'.

该页面支持如下操作：

- 创建计算空间新用户
- 添加用户到计算空间
- 移出计算空间用户

创建计算空间新用户

在控制台的「**计算空间**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以新建组织下的计算空间用户，具体操作步骤如下：

1. 选择指定计算空间，点击“空间用户”栏下的**查看详情**，即可进入该计算空间的详情页面。
2. 在“空间用户”区域，点击**创建新用户**，即可弹出创建新用户窗口。
3. 在创建新用户页面，输入新用户的相关注册信息，包括用户名、登录方式信息和密码等。
4. 在窗口点击**创建新用户**，即可将一个新用户加入组织并分配到目标计算空间中。

需要注意的是，如果该用户想登录其他计算空间并使用数据计算服务，则需要申请加入目标计算空间，或由管理员手动将其加入组织下的目标计算空间中。

计算空间									操作	
空间类型	全部	计算空间名称	计算空间ID	空间类型	云服务商	地域	状态	创建时间	空间用户	操作
计算空间名称	Pluto	ac01jf1rsyn6kk2xwbpzfc450hjz	共享存储	PieCloudDB 虚拟数仓	私有云	本地	正常	2024-12-23 11:46:11	查看详情	进入空间
DataSpace01	ac01jf1fxdtk81gggv68v8rqm2rgz			PieCloudDB 虚拟数仓	私有云	本地	正常	2024-12-24 16:44:37	查看详情	申请加入

添加用户到计算空间

在控制台的「**计算空间**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以将组织中已创建且已验证的用户添加至计算空间中，具体操作步骤如下：

1. 选择目标计算空间，点击“空间用户”栏下的**查看详情**，即可进入该计算空间的详情页面。
2. 在“空间用户”区域，点击**添加用户**，即可弹出添加用户窗口。
3. 在下拉列表中选择要添加的目标用户。
4. 在窗口点击**添加**，即可将该用户加入目标计算空间。

移出计算空间用户

在控制台的「**计算空间**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以移出计算空间中的用户，具体操作步骤如下：

1. 选择指定计算空间，点击“空间用户”栏下的**查看详情**，即可进入该计算空间的详情页面。
2. 在“空间用户”区域的指定计算空间的“操作”栏下，依次点击**移出用户 > 确定**，该用户就会被移出计算空间。

审核计算空间新用户

新用户申请加入计算空间后，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以选择如下两种途径审核新用户：

- 在控制台的「**计算空间**」页面审核

在控制台的「**计算空间**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以在计算空间列表中的“空间用户”字段下点击**查看详情**，查看该计算空间下的用户列表。

在用户列表找到申请加入该计算空间的用户，在“操作”字段下点击**审核**，选择“通过”（或“拒绝”），通过审核后的用户即可获得登录计算空间的权限；如果被拒

绝，用户侧的“已申请加入”状态会变成“申请加入未通过”，只有当管理员将该用户移出用户列表，该用户才能再次发送申请。

用户名	邮箱	验证状态	验证方式	审核状态	创建时间	操作
us01jfvyjm46n5nmry2xte107mm6	██████████	已验证	自动跳过	未审核	34分钟前	移出用户 审核
us01jfvxh4dnbs3esr38d1wxp4ne	██████████	已验证	自动跳过	审核通过	1小时前	移出用户

- 在计算空间的「**用户**」页面审核

在计算空间页面，具有对应管理权限的用户（被授予数仓管理员或者用户管理员的用户），可进入“**用户**”菜单查看用户列表。

在用户列表找到申请加入该计算空间的用户，点击“操作”字段下的**审核**，选择审核“通过”（或“拒绝”），通过审核后的用户即可获得登录计算空间的权限；如果被拒绝，用户侧的“已申请加入”状态会变成“申请加入未通过”，只有当管理员将该用户移出用户列表，该用户才能再次发送申请。

用户名	激活状态	激活方式	审核状态	操作
██████████	● 已激活	自动激活	未审核	审核
██████████	● 已激活	自动激活	审核通过	⋮

4.4 查看组织用户

在控制台的「**用户管理**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以查看当前组织下的所有用户信息。

用户					
用户名	用户ID	激活状态	激活方式	邮箱	手机号
██████████	us01h65ry5yxejpqjkwwgxyq4ez	已验证	自动跳过	██████████	██████████
██████████	us01h5940wsfnawfwljsh9p6pvbm	已验证	组织身份认证	██████████	██████████
██████████	us01h4mhgs7j2rw0627a5q1cc0d4	已验证	自动跳过	██████████	██████████

新用户加入组织可以使用控制台下的“邀请注册”功能，相关信息请参见[管理组织-添加组织用户](#)。

控制台的邀请注册功能和计算空间操作界面的邀请注册的方式类似，两者的区别是计算空间操作界面的邀请注册需要在计算空间下的用户列表完成新用户审核，而控制台的邀请注册则需要在控制台的用户列表完成新用户审核。相关信息请参见[管理计算空间-审核计算空间新用户](#)。

新用户注册成功后，会自动加入该组织，管理员可以在用户列表中查看用户的基本信息。如果新用户需要登录计算空间并使用数据计算服务，则需要申请加入该组织下对应的计算空间，或由管理员手动将其加入组织下的计算空间中。详细信息请参见[管理计算空间-添加用户到计算空间](#)。

4.5 管理组织角色

在控制台的「**角色管理**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以查询当前组织下各角色之间的关系，以及管理角色所具备的各项系统权限。

PieDataCS 云原生平台的角色分为两类，一类是控制台组织下的角色，另一类是计算空间的角色。与计算空间的角色相比，控制台组织下的角色不具有数据库对象的相关权限，也具有不同的系统权限。

关于控制台的系统预设角色

PieDataCS 云原生平台设置了控制台系统预设角色，预设角色不可被删除，管理员类角色建议使用系统预设角色。控制台预设角色与相关权限信息如下：

- 普通使用者（public）

普通使用者是控制台预设角色中最基础的角色。每一位新用户在创建后都会自动被授予普通使用者角色，具有查看组织信息和组织下计算空间的权限。

- 组织管理员 (orgmanager)

组织管理员在控制台权限管理系统中的等级最高，具有控制台中所有的系统权限，可以进行组织层级的参数管理、计算空间管理、用户管理和控制台下的权限管理。

- 财务管理员 (finmanager)

财务管理员主要管理组织下的充值账户和账单。

提示：

仅 PieDataCS 云原生平台的云上云版本支持财务管理员 (finmanager) 角色。

上述系统预设角色的初始化权限如下表所示。

角色名称	初始化权限
public	<ul style="list-style-type: none">公共参数数仓下虚拟数仓列表组织信息产品规格和定价信息数仓列表
orgmanager	<ul style="list-style-type: none">组织管理用户信息赋予/取消赋予角色给用户赋予//取消赋予权限给角色角色管理数仓下用户列表用户审核数仓信息继承//取消（被）继承角色给角色用户管理组织下用户列表数仓管理角色信息
finmanager	<ul style="list-style-type: none">发票申请组织和数仓用量信息订单列表充值购买订阅、续费和自动续费管理

有关各项权限的详细描述，请参见 [角色的系统权限清单-控制台的系统权限](#)。

控制台与计算空间的角色功能基本相同，有关角色的更多相关信息，请参见 [管理计算空间角色](#)。

查看角色权限信息

控制台的「**角色管理**」页面通过关系图和列表两种形式呈现角色的权限信息。

关系图直观展示了各角色的层次结构，也是默认的角色呈现方式。



在关系图中，组织管理员（orgmanager）或该组织下具有相应管理权限的用户点击任意角色的卡片，即可打开该角色的权限详情页面，并展示如下角色信息：

- 系统权限：系统权限页面展示当前角色已具有的系统权限，包括继承所得权限和初始权限。在该页面，点击 **查看所有权限描述** 即可查阅全量系统权限的详细说明。

角色详情(orgmanager)

Organization Manager角色作为组织管理员，管理组织和组织下的数仓

[系统权限](#)

[角色继承](#)

[用户授予](#)

此角色(orgmanager)已拥有权限



[查看所有权限描述](#)

- 角色继承：角色继承页面展示当前角色所继承的角色，或被哪些角色所继承。该页面同时支持**继承角色**和**收回角色**的操作。

角色详情(orgmanager)

Organization Manager角色作为组织管理员，管理组织和组织下的数仓

[系统权限](#)

[角色继承](#)

[用户授予](#)

角色(orgmanager)被以下角色继承

orgmanager无法被任何角色继承，只能被授予给用户

角色(orgmanager)继承了以下角色

finmanager, public

+ 继承角色 - 收回角色

- 用户授予：用户授予页面展示当前角色已授予用户的信息。该页面同时支持**将角色授予用户**和**取消对用户的角色授权**的操作。

角色详情(orgmanager)

Organization Manager角色作为组织管理员，管理组织和组织下的数仓

系统权限	角色继承	用户授予
将角色 (orgmanager) 授予用户		
<input type="text" value="请输入用户名"/> <button>授予</button>		
<div style="border: 2px dashed red; padding: 5px;">已授予用户 (1)</div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;"> testUser01 (ID:us01hc4re1e609v1ax8ytcphk3bm) </div>		

与关系图相比，**列表**是以表格的方式呈现角色信息。在列表中，用户点击目标角色名称即可展开「**角色详情**」页面，并展示各角色的系统权限、所继承的角色和被继承的角色，以及授予用户的信息。组织管理员 (orgmanager) 或该组织下具有相应管理权限的用户点击列表中的目标角色名称即可展开**角色详情**页面，并展示各角色的系统权限、所继承的角色和被继承的角色，以及角色已授予用户的信息。

The screenshot shows the 'Role Management' page. On the left, there's a navigation bar with '关系图' (Relationship Diagram) and '列表' (List), with '列表' highlighted. Below it is a search bar with placeholder '输入角色名查询'. To the right are three tabs: '系统权限' (System Permissions), '角色继承' (Role Inheritance), and '用户授予' (Granted to User). The '用户授予' tab is active, showing a table with three rows: finmanager, orgmanager, and public. The 'orgmanager' row is selected, and its details are shown on the right side of the screen. This includes a summary section for 'orgmanager' with a note that it has been granted permissions, and two large sections for 'Inherited Permissions' (10 items) and 'Initial Permissions' (13 items), each listing various system functions like '订单列表' (Order List), '发票申请' (Invoice Application), etc.

添加自定义角色

在控制台的「**角色管理**」页面，组织管理员 (orgmanager) 或该组织下具有相应管理权限的用户可以添加自定义角色，具体操作步骤如下：

1. 在**关系图**或者**列表**页面点击**添加角色**，即可弹出**创建角色**窗口。
2. 输入要创建的角色名称。

3. (可选) 从下拉列表中选择要继承其权限的角色。如果不选，会默认继承“public”角色及其权限。
4. (可选) 输入对所创建角色的描述。该描述信息会展示在**角色详情**页面。
5. 点击**创建角色**，自定义角色创建成功后会同步更新在角色关系图和列表中。

与系统预设角色不同的是，自定义角色支持修改和删除操作，**关系图**页面在角色名称卡片上提供了“编辑角色”和“删除角色”的快捷键图标。**列表**页面中自定义角色的“操作”栏下的隐藏菜单同样支持快捷“编辑角色”和“删除角色”的操作。

继承和收回角色

继承角色是将权限以角色的形式传递给另一个角色，但继承角色在层级上需要高于被继承的角色，且继承角色的权限与被继承角色的权限保持一致。

继承角色操作包括“当前角色被其他角色继承”和“当前角色继承其他角色”。如果“当前角色被其他角色继承”，则当前角色还可以执行“收回角色”的操作。

注意：

public 作为系统角色无法继承任何角色，但默认被所有角色继承。

当前角色被其他角色继承

在控制台的**「角色管理」**页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以执行新增角色并继承当前角色的操作：

1. 打开目标角色的**角色详情**页面并切换到**角色继承**面板。
2. 在**角色被以下角色继承**区域，点击**继承角色**。

角色详情(demo-role)

系统权限 角色继承 用户授予

角色(demo-role)被以下角色继承

暂无任何角色继承此角色

+ 继承角色 - 收回角色

角色(demo-role)继承了以下角色

public

+ 继承角色 - 收回角色

- 从下拉列表中选择一个选择一个角色来继承当前角色。

角色详情(demo-role)

系统权限 角色继承 用户授予

新增一个角色继承当前角色(demo-role)

已继承角色 (0)

选择一个角色，继承当前角色(demo-role)

orgmanager

- 点击 **保存修改**，则所选角色即可继承当前角色。

如果需要取消其他角色对目标角色的继承，在控制台的「**计算空间**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户在指定角色的**角色详情**页面的**角色被以下角色继承**区域，点击**收回角色**并从下拉列表中选择一个要收回继承权限的角色。一旦确认变更，该角色将不再继承当前角色。

角色详情(demo-role)

The screenshot shows the 'Role Details' page for 'demo-role'. At the top, there are three tabs: 'System Permissions', 'Role Inheritance' (which is selected and underlined), and 'User Assignment'. Below the tabs, a red dashed box highlights the section titled 'The role (demo-role) is inherited by the following roles'. Inside this box, the 'orgmanager' role is listed in a dropdown menu. At the bottom of this section are two buttons: '+ Inherit Role' (highlighted with a red border) and '- Revoke Role'.

角色(demo-role)被以下角色继承

orgmanager

+ 继承角色 - 收回角色

当前角色继承其他角色

在控制台的「**角色管理**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以执行当前角色继承其他角色的操作：

1. 打开目标角色的**角色详情**页面并切换到**角色继承**面板。
2. 在**角色继承了以下角色**区域，点击**继承角色**。

The screenshot shows the 'Role Details' page for 'demo-role'. The 'Role Inheritance' tab is selected. A red dashed box highlights the section 'The role (demo-role) inherits the following roles'. Inside this box, the 'orgmanager' role is listed in a dropdown menu. At the bottom of this section are two buttons: '+ Inherit Role' (highlighted with a red border) and '- Revoke Role'.

角色(demo-role)继承了以下角色

orgmanager

+ 继承角色 - 收回角色

3. 从下拉列表中选择一个要继承的角色。

角色详情(demo-role)

系统权限 角色继承 用户授予

使当前角色(demo-role)继承一个新角色

被继承角色 (1)

public

选择一个要继承的角色

admin

4. 点击 **保存修改**，当前角色就继承了所选的角色。需要注意的是，如果要继承角色的权限低于当前角色，会出现报错。

如果需要取消目标角色所继承的角色，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以在目标角色的**角色详情**页面的**角色继承了以下角色**区域，点击**收回角色**并从下拉列表中选择一个要取消继承的角色。一旦确认变更，当前角色将不再继承该角色。

角色详情(demo-role)

系统权限 角色继承 用户授予

角色(demo-role)被以下角色继承

orgmanager

+ 继承角色 - 收回角色

角色(demo-role)继承了以下角色

public

admin

+ 继承角色 - 收回角色

授予角色给用户和收回授权

授予角色给用户

授予角色给用户是将权限以角色的形式下发给用户。组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以将角色及其权限赋予给其他用户。

注意：

public 角色不支持此操作，因为 public 已被默认授予给所有用户。

在控制台的「**角色管理**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以授予角色给其他用户，具体操作步骤如下：

1. 打开目标角色的**角色详情**页面并切换到**用户授予**面板。
2. 在**将角色授予用户**区域，从下拉列表中选择一个合适的用户。

角色详情(finmanager)

Financial Manager角色作为财务管理员，管理组织下的充值账户和账单

系统权限	角色继承	用户授予
将角色 (finmanager) 授予用户		
<input type="text" value="请输入用户名"/>		授予
已授予用户 (1)		
testUser01 (ID:us01hc4re1e609v1ax8ytcpfhk3bm)		x

3. 点击**授予**，即可将角色授予新用户并将更改信息同步更新到已授予用户列表中。

角色详情(demo-role)

系统权限	角色继承	用户授予
将角色 (demo-role) 授予用户		
<input type="text" value="请输入用户名"/>		授予
已授予用户 (1)		
testUser01 (ID:us01hc4re1e609v1ax8ytcpfhk3bm)		x

收回角色授权

在控制台的「**角色管理**」页面，组织管理员（orgmanager）或该组织下具有相应管理权限的用户可以收回对其他用户的角色授权，具体操作步骤如下：

1. 打开指定角色的**角色详情**页面并切换到**用户授予**面板。
2. 在已授予用户列表中，选中目标用户并点击**X**。

角色详情(demo-role)

系统权限	角色继承	用户授予
------	------	-------------

将角色 (demo-role) 授予用户

请输入用户名

授予

已授予用户 (1)

testUser01 (ID:us01hc4re1e609v1ax8ytcphk3bm) X

3. 在弹出的窗口中点击**确认收回**，即可取消对用户的角色授权并同步更新已授予用户的列表。



提示:

用户被取消某一角色授权后可以再次被授予相同的角色。

4.6 管理计费与订阅

PieDataCS 云原生平台的云上云版本支持多种计费模式，包括按量计费和包年包月。用户可以选择一次性支付费用订阅包年包月服务。此外，平台还支持自动续费功能，确保用户能够享受连续且稳定的服务。

提示：

仅 PieDataCS 云原生平台的云上云版本支持此功能。

查看计费模式与计费标准

在控制台的「计算空间」页面的计算空间列表中，财务管理员（finmanager）或具有相应管理权限的用户点击目标计算空间的“计费模式”栏下的**管理**，即可进入该计算空间下的**虚拟数仓**面板，该页面显示了当前计算空间所使用的虚拟数仓及其计费模式。

用户列表	虚拟数仓	计费模式			
ID	名称	状态	规格	执行节点	计费模式
██████████	vm1	● 运行中	标准	2	按量计费
██████████	vw2	● 已关闭	标准	4	按量计费

财务管理员（finmanager）或具有相应管理权限的用户在该页面切换到**计费模式**面板，点击**按量计费**即可查看当前的按量计费方式的收费标准，如下表所示。

计费方式	收费标准
虚拟数仓计算(标准)	1.5101元/节点/小时
虚拟数仓计算(中等)	3.0202元/节点/小时
数据库存储(metadata)	0.0042元/GB/小时
数据库存储(s3)	0.000377元/GB/小时
数据请求临时缓存(s3)	0.000377元/GB/小时
数据导入临时缓存(s3)	0.000377元/GB/小时
PieProxy 出口流量(pieproxy)	0.72元/GB

订阅包年包月服务

在控制台的「计算空间」页面，财务管理员（finmanager）或具有相应管理权限的用户可以订阅包年包月服务，具体操作步骤如下：

1. 在控制台的「计算空间」页面的计算空间列表中，点击指定计算空间的“计费模式”栏下的**管理**，即可进入当前计算空间的计费模式页面。
2. 切换到**计费模式**面板，点击**去订阅**，即可进入订阅页面并按需以包年包月的方式订购计算空间。

用户列表 虚拟数仓 **计费模式**

流量与存储为所有虚拟数仓共用，超出套餐的流量与存储的部分，将根据按量计费进行收费。

计费项	名称	规格	数量	计费模式	有效期
虚拟数仓	vw2	标准	协调节点: 1 执行节点: 4	按量计费	-
虚拟数仓	vm1	标准	协调节点: 1 执行节点: 2	按量计费	-

[去订阅](#)

3. 根据虚拟数仓规格分别指定新虚拟数仓的执行节点数量（或者可以直接添加现有的虚拟数仓），以及存储容量和流量大小。
4. 选择订阅时长（支持 1 个月和 12 个月）
5. 选择是否开通自动续费、到期后是否转为按量付费。此时，系统会在页面自动计算出总金额。

① 请知悉：订阅套餐外新建的虚拟数仓，以及存储和流量超过套餐规格上限的部分，会自动转为按量计费。

计费项	规格	数量	名称	单价
虚拟数仓	标准	2	执行节点	300元/月(协调节点) + 300元/月(执行节点)
存储	存储-500G	1		130元/月
流量	出口流量-50G (每月)	1		35元/月

[+ 添加新的虚拟数仓](#) [+ 添加现有的虚拟数仓](#)

订阅时长

1 个月 **12 个月**

开通自动续费
 到期后转为按量付费
 阅读并同意[《订阅条款》](#)

总金额: **¥ 1065** [重新计算](#)

[返回](#) [提交订单](#)

6. 阅读《订阅条款》并勾选后点击**提交订单**。
7. 确认订阅信息无误后，点击**提交**。如果需要修改，则点击**取消**并返回页面修改后重新提交。
8. 订单提交后会跳转至控制台的「**费用中心-订单管理**」页面，用户在该页面选择支付方式并支付，即可完成订阅流程。

4.7 费用中心

费用中心用于展示资源用量及其对应的费用。财务管理员 (finmanager) 或具有相应管理权限的用户可以查询对应账户下的资源用量，并进行付费等操作。

提示：

仅 PieDataCS 云原生平台的云上云版本支持此功能。

在「**控制台**」界面，点击菜单栏中的**费用中心**即可进入功能页面。该页面支持如下操作：

- 充值购买 π 点
- 查看消费记录
- 申请开具发票
- 管理支付订单

消费记录概览

在控制台的「**费用中心-费用总览**」页面，财务管理员（finmanager）或具有相应管理权限的用户可以查看当前账户的消费记录的概览。

PieDataCS 云原生平台使用 π 点（Pie Point）作为衡量不同计费项的计价标准。π 点是拓数派（OpenPie）推出的虚拟账户计量单位，适用于拓数派旗下产品 PieDataCS 及相關服務。PieDataCS 云原生平台以小时为单位实行计费，以日为单位实行扣费。

费用总览		
本月已用π点	今日已用π点 / 可用余额	日均使用量（本月已用 / 天数）
22.14	0 / 977.86 ?	0.96(22.14 / 23)

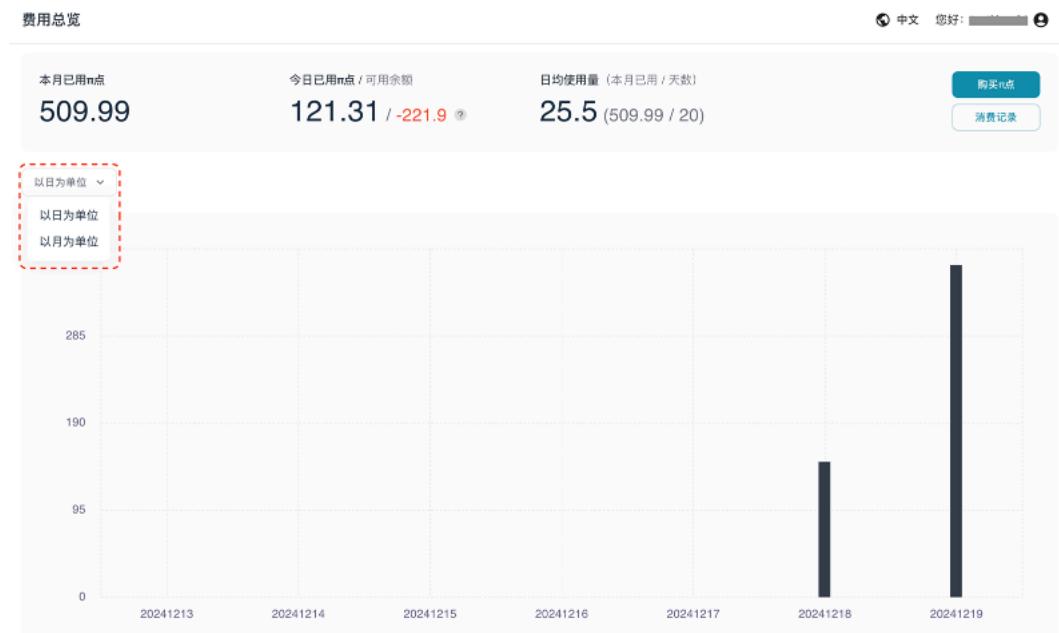
- “本月已用 π 点” 表示当月已使用 π 点的总和。
- “今日已用 π 点” 表示当日已记录但尚未从 π 点余额中扣除的消费。该数据每小时更新一次。每日的已用 π 点总和将在北京时间每日 0 点后自动从余额中扣除。

“可用余额”表示当前账户中以 π 点形式表示的可用资源总和，其更新频率同样为每小时一次。需要注意的是，由于更新频率的限制，如果一个小时内使用的资源量超过了当前的可用 π 点，“可用余额”可能会显示为负值。在这种情况下，除非用户及时充值，否则将无法继续使用服务。

PieDataCS 云原生平台在检测到欠费后，会为试用平台的组织提供 24 小时的缓冲期以完成充值；如果账户在 24 小时内未能充值并回到非欠费状态，系统将冻结该组织。届时，如果账户希望恢复使用服务，必须联系 OpenPie 技术支持团队。

- “日均使用量”表示当月每天平均使用的π点，由公式“本月已用π点/本月使用天数”计算得出。

此外，控制台的「**费用中心-费用总览**」页面还支持以柱形图的形式展示消费趋势，用户可以选择日或月作为展示的时间单位。当前仅支持显示近7天及近6个月所消费的历史π点消费记录。



充值购买π点

财务管理员（finmanager）或具有相应管理权限的用户可以通过在线充值或者对公转账的方式购买π点。

在线充值购买π点的步骤如下：

- 进入控制台的「**费用中心-充值购买**」页面，或者在控制台的「**费用中心-费用总览**」页面点击**购买π点**，即可跳转至在线充值页面。

The screenshot shows the 'Expense Center - Purchase Points' page. It features tabs for 'Online Recharge' (highlighted with a red box) and 'Corporate Transfer'. A 'Recharge History' button is also present. Below is a list of instructions:

1. Deducting π points will prioritize deducting赠送的π points.
2. If you need to understand specific consumption details, please go to 'Expense Overview' - 'Expense History'.
3. Users can issue invoices based on account statements and actual consumption amounts (taxable amount), and apply for invoices through the 'Invoice Application' menu.

Below the instructions are sections for 'Recharge Amount' (1 RMB = 1 π point) and payment methods. It includes a self-defined amount input field, a checkbox for accepting the Openpie recharge agreement, and a 'Pay Now' button.

2. 选择充值金额（换算公式为“1元人民币=1π点”），或者自定义充值金额。
3. 勾选“阅读并同意 Openpie 充值协议”。
4. 点击**立即支付**，即可通过扫码的方式来付款充值。

财务管理员（finmanager）或具有相应管理权限的用户在控制台的「**费用中心-充值购买**」页面切换至**对公转账**面板，可以查看拓数派的对公账号信息并进行线下汇款。



此外，财务管理员（finmanager）或具有相应管理权限的用户在该页面点击**充值记录**，即可跳转至「**费用中心-费用总览-消费记录**」页面并查看当前组织的全部充值记录。

查看消费记录

在控制台的「**费用中心-费用总览**」页面，财务管理员（finmanager）或具有相应管理权限的用户点击**消费记录**，可以查看当前账号的消费记录，其中包括：

- **月度账户变动** 查看当前组织下所有计算空间的月度交易明细。用户可以根据计算空间的名称、收支类型（包括收入和支出）、交易类型（包括消费、充值、退款、系统调整和注册赠送）以及交易时间（以月为单位）等字段进行筛选，以便查看相关信息。
- **账户变动明细** 查看当前组织下所有计算空间的以天为单位的交易明细及对应的计费项。用户可以根据计算空间的名称、收支类型（包括收入和支出）、交易类型（包括消费、充值、退款、系统调整和注册赠送）以及交易时间等字段进行筛选，以便查看相关信息。
- **月度 π 点消费** 查看当前组织下所有计算空间的 π 点（Pie Point）的月度消耗明细。用户可以根据计算空间的名称和交易时间（以月为单位）进行筛选，以便查看相关信息。

展开子列表可以显示数据库存储、虚拟数仓计算和数据请求临时缓存分别消耗的 π 点和对应使用量。

- **π 点消费明细** 查看当前组织下所有计算空间的每小时消费 π 点（Pie Point）明细及对应的计费项。用户可以根据计算空间的名称、计费项（包括虚拟数仓计算、数据库存

储、数据请求临时缓存、数据导入临时缓存和 PieProxy 出口流量）、消费类型（包括预付费和后付费）和交易时间等字段进行筛选，以便查看相关信息。

- 充值记录 查看当前组织的全部充值记录。如果用户充值失败或并未进行支付，也会被记录。用户可以根据交易类型（包括支付、关闭和退款）字段进行筛选，以便查看相关信息。

申请开具发票

在控制台的「费用中心-申请开票」页面，财务管理员（finmanager）或具有相应管理权限的用户可以申请开票和查看开票记录。

申请开具发票的步骤如下：

- 选择开票账单记录。选定开票账单后，点击 **下一步**，即可进入选择抬头页面。

The screenshot shows the 'Invoice Management' section of the 'Fee Center'. It has tabs for 'Apply for Invoice' (selected) and 'Invoice Record'. Below is a note about invoice application rules:

1. Invoice application is based on monthly consumption statements. Single consumption statements do not support splitting into multiple invoices. Multiple months' consumption statements can be merged into one invoice.
2. When the organization passes the special tax audit, it can apply for value-added tax electronic invoices.
3. The invoice item name is: "Information Technology Services" - Cloud Service. The tax rate is 6%. The invoice item and tax rate are based on the tax law's regulations, and they cannot be modified.
4. In response to the Zhejiang Provincial Tax Bureau's work on fully digitalizing electronic invoices, our company only supports issuing electronic invoices. Please fill in the invoice header and recipient email address correctly.
5. After you submit an invoice application, the system will issue and send an invoice to your email within one hour. Please wait patiently.

Below the note is a three-step progress bar:

- 1. Select Invoice Statement Record**
- 2. Select Header**
- 3. Confirmation Information**

The main area shows a table of monthly statement records:

Statement Period	Actual Consumption Points	Invoice Amount
202309	2949.01	1899.76
202402	3847.89	1357.41
202403	10180	10180
202404	1594.56	362.71

At the bottom left, it says 'Selected Invoice Amount: 1899.76'. At the bottom right is a blue button labeled 'Next Step'.

在该页面点击 **查看历史月账单记录**，可以跳转至控制台的「费用中心-费用总览-消费记录」页面，查看月度π点消费记录。

- 选择抬头。在选择抬头页面，从列表中选择已有的发票抬头信息，并输入接收发票的邮箱等信息。点击 **下一步**，进入确认信息页面。

发票管理

申请开票 开票记录

1. 发票申请基于月度消费账单。单月消费账单不支持拆分为多张发票，多个月的消费账单可支持合并，开具一张发票。
 2. 仅当组织名通过专票申请审核时，可申请开具增值税电子专用发票。[专票申请审核](#)
 3. 开票类目名称为：“信息技术服务”云服务，税率为6%，发票类目和税率基于税法的规定，不支持修改。
 4. 为响应浙江省税务局关于开展全面数字化的电子发票试点工作，当前我司仅支持开具电子发票，请您务必正确填写发票抬头和收件邮箱地址。
 5. 在您提交开票申请后，系统将在一个小时为您开具并发送到您的邮箱，请您耐心等待。

1. 选择开票账单记录 2. 选择抬头 3. 确认信息

选择发票抬头

发票类型	发票抬头	抬头类型	纳税人识别号	抬头状态	操作
<input checked="" type="radio"/> 增值税普通发票	██████████	企业	██████████	就绪	修改 删除
<input type="radio"/> 增值税电子专用发票	██████████	企业	██████████	就绪	

电子发票收件邮箱地址

选中的开票金额：1899.76 元

[上一步](#) [下一步](#)

或者，点击 **添加抬头**，即可添加新的抬头信息。在目标抬头所在行，点击 **修改** 即可编辑普通发票的抬头信息，但专用发票的抬头不支持直接修改。同时，用户也可以选择直接删除抬头信息。

3. 确认信息。在确认信息页面，点击 **确认申请** 即可提交开票信息。

发票管理

申请开票 开票记录

1. 发票申请基于月度消费账单。单月消费账单不支持拆分为多张发票，多个月的消费账单可支持合并，开具一张发票。
 2. 仅当组织名通过专票申请审核时，可申请开具增值税电子专用发票。[专票申请审核](#)
 3. 开票类目名称为：“信息技术服务”云服务，税率为6%，发票类目和税率基于税法的规定，不支持修改。
 4. 为响应浙江省税务局关于开展全面数字化的电子发票试点工作，当前我司仅支持开具电子发票，请您务必正确填写发票抬头和收件邮箱地址。
 5. 在您提交开票申请后，系统将在一个小时为您开具并发送到您的邮箱，请您耐心等待。

1. 选择开票账单记录 2. 选择抬头 3. 确认信息

发票抬头

发票类型	发票抬头	抬头类型	纳税人识别号
增值税普通发票	██████████	企业	██████████

电子发票收件邮箱地址

发票备注

选中的开票金额：1899.76 元

[上一步](#) [确认申请](#)

在控制台的「**费用中心**」页面切换至**开票记录**面板，财务管理员（finmanager）或具有相应管理权限的用户可以查看开票记录，包括开票金额、发票抬头、发票类型、申请人、申请时间和发票状态等信息。

The screenshot shows the 'Invoicing Record' section of the Control Console. At the top, there are tabs for 'Apply for Invoice' and 'Invoicing Record', with 'Invoicing Record' being the active tab. Below the tabs, there's a button labeled 'All Records'. The main area displays a table with the following data:

Invoice Amount	Invoice Header	Invoice Type	Applicant	Application Time	Invoice Status
432.18元	[Redacted]	VAT General Ordinary Invoice	[Redacted]	2023.09.11	Invoiced successfully

管理支付订单

控制台的「**费用中心-订单管理**」页面以列表的形式展示当前组织下所有计算空间按量计费或者按包年包月计费的订单信息。

财务管理员（finmanager）或具有相应管理权限的用户可以根据计算空间的名称、订单类型（包括订阅和续费）、订单状态（包括未支付、已支付、已完成和已取消）以及订单号等字段进行筛选和查找，以便查看相关信息。

该页面支持如下操作：

- **查看订单详情**

在“操作”栏下点击**详情**即可跳转至订单详情页面。该页面展示了计费项、规格、数量、单价和总金额等信息。用户可以点击**去支付**跳转至订单支付页面进行付款操作。同时，用户也可以点击**取消订单**，一旦订单被取消，则该订单会关闭且在订单列表不会再出现支付的操作项。

- **支付未付款的订单**

如果列表中存在订单的订单状态为“未支付”，在“操作”栏下点击**去支付**即可跳转至订单支付页面，点击**确定支付**后扫码即可付款充值。

4.8 我的工单

PieDataCS 云原生平台的云上云版本支持创建和管理工单，对 OpenPie 技术支持团队进行提问和反馈。

提示：

仅 PieDataCS 云原生平台的云上云版本支持此功能。

在「**控制台**」界面，点击菜单栏中的**我的工单**即可进入功能页面。该页面支持如下操作：

- 创建和提交工单
- 查看工单详情
- 删除工单

创建和提交工单

在「**我的工单**」页面，用户可以根据如下步骤创建和提交工单：

1. 点击**创建工作单**，弹出新建工单窗口。
2. 选择所属问题的分类，例如数据洞察、数据库、虚拟数仓、导入数据、查询历史、用户、角色等。
3. 填写详细的问题描述，说明所遇到的问题。
4. 点击**提交**，即可创建一个新的工单。



新工单创建成功后会同步在工单列表中，而且订单状态为“等待回复”。

查看工单详情

在工单列表中，用户点击指定的工单信息列即可进入工单详情页面。

工单列表 / 工单详情

工单#tk01j7n4zlj6r5sc26a5hztkere 15分钟前 导入数据 一分钟后

testUser01 15分钟前
test: 导入csv文件的数据时发生报错。

返回列表 已解决 回复

工单详情页面支持如下操作：

- 回复工单。用户可以点击 **回复 > 提交** 来回复反馈或者继续补充问题描述。
- 关闭工单。如果工单已解决，用户可以点击 **已解决 > 确定** 来结束提问并关闭工单，之后工单列表中的订单状态会变为“已关闭”。工单关闭后会在页面出现 **重新开启** 按钮，点击该按钮后还可以继续使用工单提问。
- 返回工单列表。用户点击 **返回列表** 即可返回工单列表主页面。

删除工单

在工单列表中的“操作”栏下点击 **删除**，用户可以撤回并删除已提交的工单。

工单列表

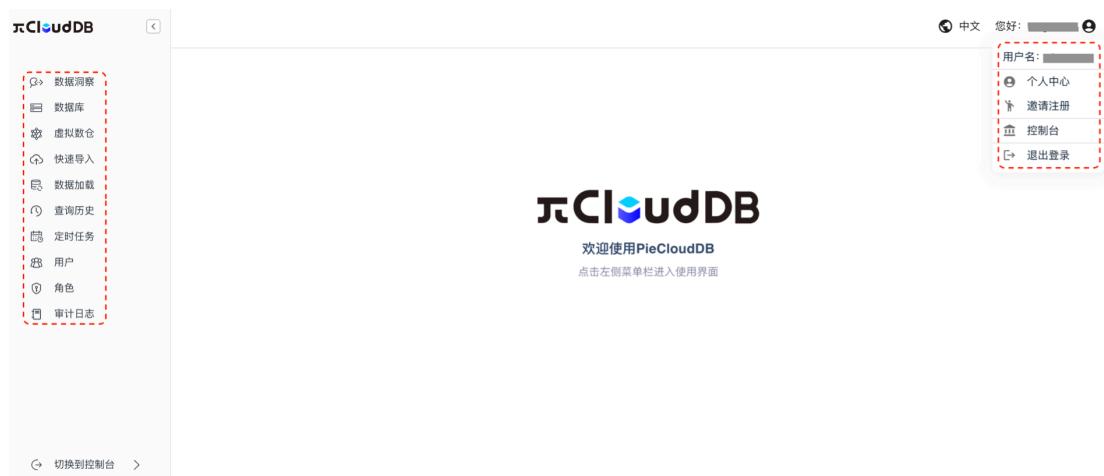
创建工单

类型	工单描述	状态	创建时间	操作
导入数据	test: 导入csv文件的数据时发生报错。	已关闭	3个月前	删除
数据洞察	无法执行 SQL 查询。	已回复	3个月前	删除

5 计算空间使用指南

5.1 计算空间功能概览

PieDataCS 云原生平台的计算空间操作界面分为**功能菜单栏**和**用户信息栏**两部分。



提示:

PieDataCS 云原生平台的计算空间操作界面菜单栏中的功能会因版本不同而略有差异。

有关用户信息栏的详细信息，请参见 [用户信息栏](#)。

计算空间功能菜单栏包括以下通用功能：

- [数据洞察](#)

创建 SQL 文件、执行查询任务以及查看执行结果。

- [数据库](#)

查询当前数据库的元数据及数据库对象的权限信息。

- [虚拟数仓](#)

创建计算资源来执行所提交的 SQL 任务。

- [快速导入](#)

提供 ETL 功能，支持上传本地的数据文件到平台。

- [查询历史](#)

查询请求历史的相关信息。

- [用户](#)

查看当前计算空间下所有用户的激活和审核状态；拥有相关权限的用户可进行多项用户相关的操作。

- [角色](#)

查询当前数仓下各角色的关系，以及查看和管理角色的各项权限。

- [新手指引](#)

集合了指引、视频、产品使用手册及样本数据集，提供新手指引及相关参考内容。

提示：

仅 PieDataCS 云原生平台的云上云版本支持该功能。

- [帮助与支持](#)

提供 OpenPie 联系邮箱、客服电话以及技术交流群二维码。

提示：

仅 PieDataCS 云原生平台的云上云版本支持该功能。

综合来看，在一套完整的 SQL 执行流程中，依次涉及到的平台功能包括虚拟数仓、数据洞察、快速导入、数据库以及查询历史等。

此外，PieDataCS 云原生平台企业版的独有功能如下：

- [定时任务](#)

创建并定期执行后台 SQL 任务。

- [审计日志](#)

查看虚拟数仓的用户行为审计日志。

- [数据脱敏扩展](#)

在脱敏 Schema 中创建脱敏的表并设置字段的脱敏规则。

- [外部数据包装器（FDW）扩展](#)

实现了访问外部数据源数据的功能。

- [数据集成](#)

从不同的数据库抽取数据并加载到同构或者异构的目标数据库。

5.2 用户信息栏

在计算空间操作界面，用户点击登录所使用的“用户名”，即可在下拉列表中选择个人中心、邀请注册、控制台和退出登录等多项与用户相关的功能。

个人中心

用户在个人中心可以完善基本信息，例如绑定邮箱、手机号或更改密码等。

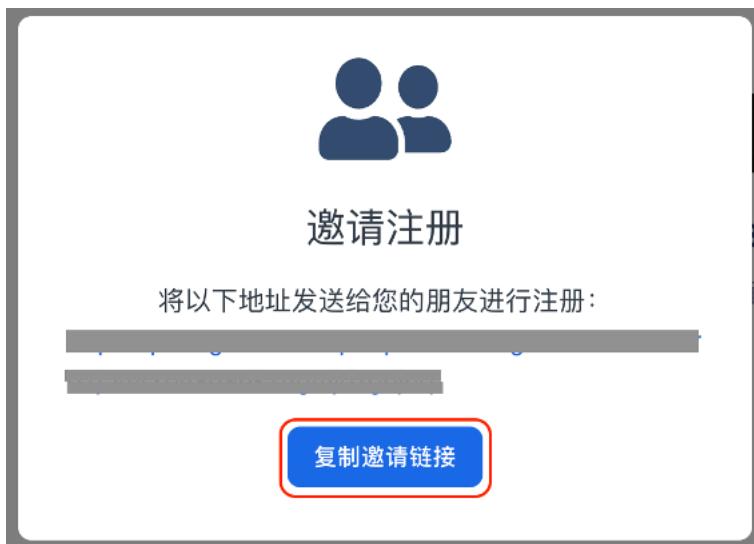
在“用户信息栏”的下拉列表中选择**个人中心**即可进入功能页面，用户可以执行如下操作：



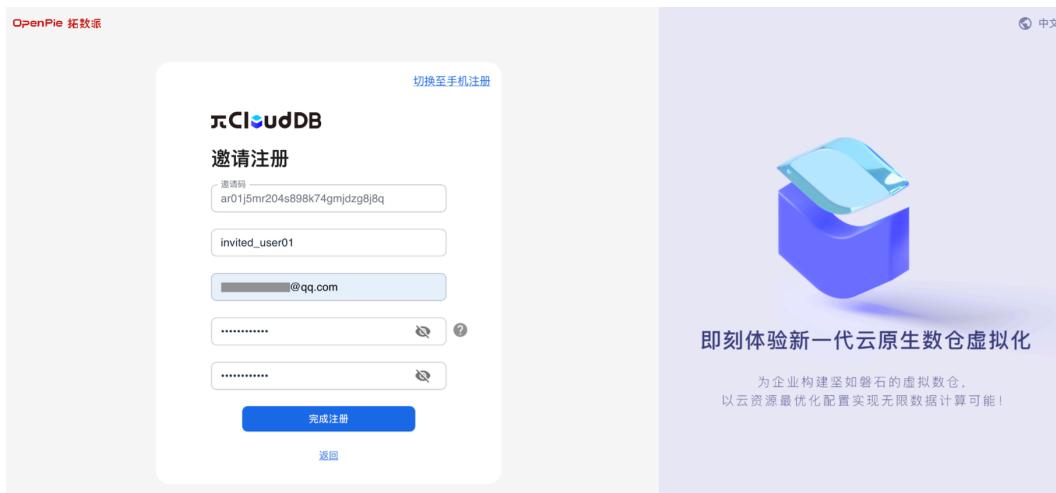
- **更改绑定方式**：如果用户初始使用邮箱（或手机号）的方式注册，则点击**绑定手机号**（或**绑定邮箱**），可以切换为使用新的绑定方式用于登录使用。
- **修改密码**：点击**修改密码**可以修改用户密码。

邀请注册

在当前计算空间下，用户可以使用邀请注册功能添加新用户。在“用户信息栏”的下拉列表中选择**邀请注册**，之后在弹出的窗口中点击**复制邀请信息**并将链接发送给新用户即可完成邀请。



新用户点开链接后会进入到邀请注册页面。凭借邀请码，输入相关注册信息和验证码完成注册。新用户注册步骤完成后，将自动加入邀请人所在数仓。



控制台

控制台主要包括组织管理、计算空间管理、用户管理、角色管理、费用中心（仅云上云版支持）和我的工单（仅云上云版支持）等功能。有关各功能的详细信息，请参见 [控制台功能概览](#)。

退出登录

退出 PieDataCS 云原生平台。

5.3 虚拟数仓

虚拟数仓是计算空间（数仓）内可按需创建的极具弹性的计算资源，它用于提供数据计算服务。在使用 PieCloudDB 运行 SQL 查询或 DML 操作之前，用户需要先指定一个虚拟数仓和数据库。之后，相应的 SQL 任务将通过指定的虚拟数仓与指定数据库中的数据进行交互。

由于 PieCloudDB 采用存算分离的架构，即使不同的虚拟数仓访问的是同一份数据，每个虚拟数仓的计算资源也是相互隔离的。因此，任何一个虚拟数仓与存储层的交互、启动、停止或扩缩容操作都不会影响其他虚拟数仓的计算性能。用户可以根据自己的计算需求，创建、启动、停止或调整虚拟数仓的资源规格。

在计算空间操作界面，具有相应管理权限的用户（例如数仓管理员和数据库管理员）点击菜单栏中的 **虚拟数仓** 即可进入功能页面。该页面主要支持如下操作：

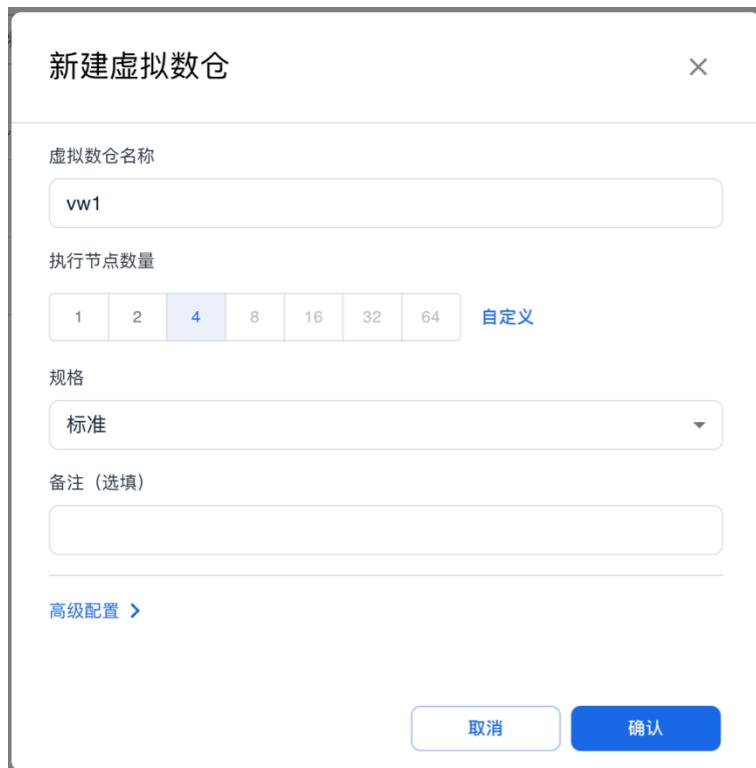
- 创建虚拟数仓
- 删除虚拟数仓
- 启停虚拟数仓
- 修改虚拟数仓配置
- 虚拟数仓的扩缩容
- 配置外部接入
- 查看虚拟数仓详情

创建虚拟数仓

一个 PieCloudDB 数仓下可根据需求创建多个虚拟数仓。创建虚拟数仓的具体操作步骤如下：

1. 在「虚拟数仓」页面点击 **新建虚拟数仓**，进入虚拟数仓创建页面。
2. 输入虚拟数仓的名称，并选择执行节点数量和节点规格。

如下示例为，创建一个名称为“vw1”，且拥有“4”个“标准”规格（即 2C4G）执行器的虚拟数仓。



提示:

当前 PieCloudDB 默认只支持标准的节点规格，但用户可根据业务场景联系 OpenPie 定制虚拟数仓的规格。

3. (可选) 设置虚拟数仓的自动启停功能并按需设置自动关闭等待时长。

“高级配置”选项用于设置虚拟数仓的自动启停功能，该功能需要在部署时手动开启。

- 如果打开“自动启动”功能，则当虚拟数仓处于关闭状态而用户尝试运行查询命令、数据导入、外部接入等功能时，虚拟数仓会自动启动并运行相关任务。
- 如果打开“自动关闭”功能，当虚拟数仓在没有运行任务时，会等待至所设置的“自动关闭等待时长”后自动关闭。系统默认的自动关闭等待时长为 10 分钟。



提示：

PieDataCS 云原生平台的云上云版本默认开启自动启停功能。

4. 点击确认，一个新的虚拟数仓即创建完成并同步到虚拟数仓列表中。

删除虚拟数仓

虚拟数仓支持删除操作。如果需要删除一个虚拟数仓，用户可以在「虚拟数仓」页面的虚拟数仓列表中选中目标虚拟数仓，然后点击隐藏菜单栏「...」并选择 **删除** 选项，即可删除该虚拟数仓。但需要注意，一旦虚拟数仓被删除，它将无法被恢复。

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	运行中	4	标准	PieProxy	接入详情 高级配置	未开启	按量计费	2 天前	查看详情 查看日志
vw2	运行中	4	标准	PieProxy	接入详情 高级配置	未开启	按量计费	2 天前	查看详情 查看日志

共 2 条

每页 10 条 < < 1 > > 前往

启停虚拟数仓

在 PieCloudDB 数仓下，用户可以根据需求开启多个虚拟数仓。每个虚拟数仓都支持手动启动和关闭。此外，如果虚拟数仓出现状态异常，系统还提供了强制关闭操作的选项。

在「虚拟数仓」页面的虚拟数仓列表中选中目标虚拟数仓，然后点击隐藏菜单栏「...」并选择相应的启停操作。

- 当虚拟数仓处于“运行中”状态时，点击 **关闭** 选项，即可关闭运行中的虚拟数仓。

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	● 运行中	4	标准	不开启	未开启	按量计费	10 分钟前	查看详情	查看日志
vw2	● 运行中	4	标准	不开启	未开启	按量计费	36 分钟前	查看详情	查看日志

共 2 条 每页 10 条 | < | < | 1 | > | > | 前往 | [关闭](#) | [修改](#) | [删除](#)

- 当虚拟数仓处于“已关闭”状态时，点击**启动**选项，即可启动已关闭的虚拟数仓。

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	● 已关闭	4	标准	不开启	未开启	按量计费	9 分钟前	查看详情	查看日志
vw2	● 运行中	4	标准	不开启	未开启	按量计费	35 分钟前	查看详情	查看日志

共 2 条 每页 10 条 | < | < | 1 | > | > | 前往 | [启动](#) | [修改](#) | [删除](#)

- 当虚拟数仓处于“失败”、“关闭中”或者“启动中”状态时，点击**强制关闭**选项，即可强制关闭该虚拟数仓。

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	● 启动中	4	标准	不开启	未开启	按量计费	12 分钟前	查看详情	查看日志
vw2	● 运行中	4	标准	不开启	未开启	按量计费	38 分钟前	查看详情	查看日志

共 2 条 每页 10 条 | < | < | 1 | > | > | 前往 | [强制关闭](#) | [修改](#) | [删除](#)

修改虚拟数仓配置

在「**虚拟数仓**」页面的虚拟数仓列表中，用户选择目标虚拟数仓并点击隐藏菜单「...」，然后选择**修改**选项，即可进入目标虚拟数仓的配置界面。

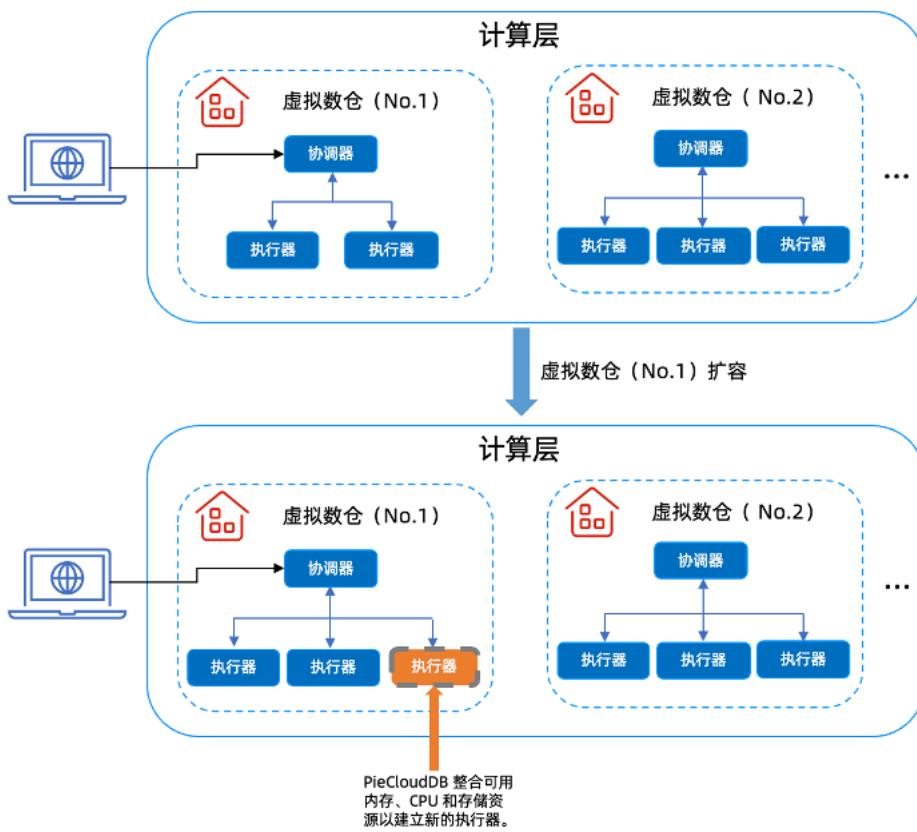
虚拟数仓名称		状态	执行节点数量	规格	外部接入	审计开关	计费模式	创建时间	备注	查看更多
vw1	● 运行中	4	标准	不开启	未开启	按量计费	33分钟前		查看详情 查看日志	⋮
vw2	● 运行中	4	标准	不开启	未开启	按量计费	1小时前		查看详情 查看日志	修改
共 2 条										
每页	10	条	<	1	>	>1	前往			关闭 删除

虚拟数仓支持修改基本配置和高级配置。当虚拟数仓处于“运行中”或“已关闭”状态时，包括虚拟数仓名称、执行器数量、规格、备注、自动启停项等在内的配置项均支持修改。

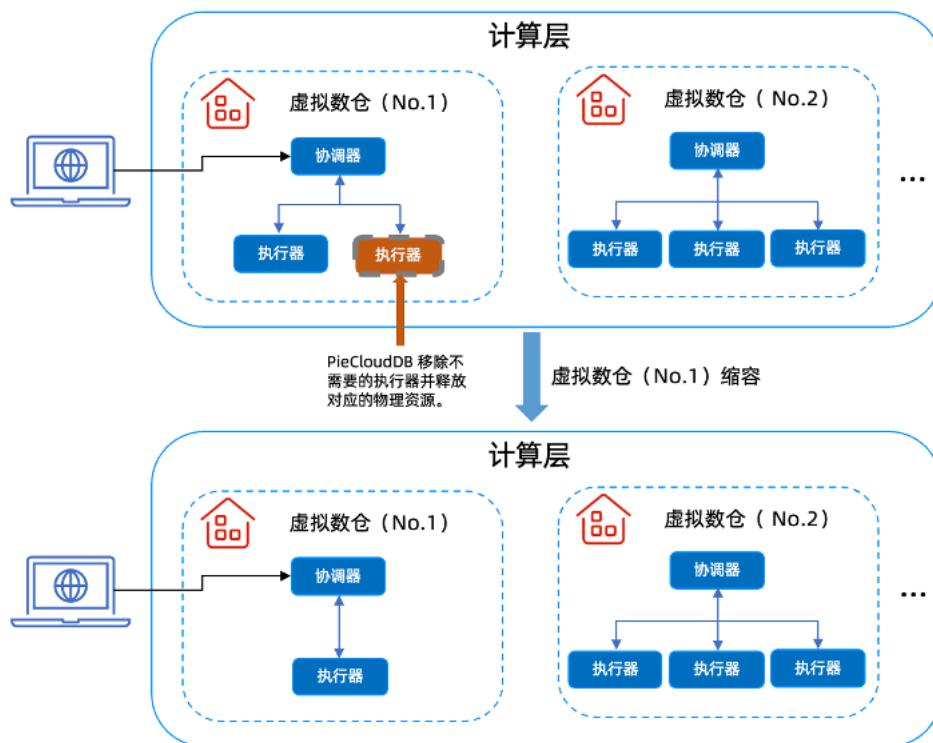
虚拟数仓的扩缩容

虚拟数仓可以根据用户需求，通过增加或减少执行器的数量来实现扩容（Scale Out）或缩容（Scale In）。通过这种方式，虚拟数仓中的协调器可以保持不变，即用户在完成虚拟数仓的扩缩容操作后，执行查询任务时与执行器相交互的协调器与扩缩容操作前是相同的。

下图展示了虚拟数仓通过增加执行器数量进行扩容的过程。



下图展示了虚拟数仓通过减少执行器数量进行缩容的过程。



当虚拟数仓处于“已关闭”或“运行中”状态时，可以通过修改执行器（节点）数量来实现扩缩容。

在「虚拟数仓」页面的虚拟数仓列表中，选择目标虚拟数仓并点击隐藏菜单「...」，然后选择 **修改** 选项，即可进入目标虚拟数仓的配置界面。

在虚拟数仓的配置页面，用户可以通过减少执行节点的数量来执行缩容操作。例如，将节点数量从“4”调整为“2”。一旦确认变更，设置会立即生效，从而将虚拟数仓的执行器数量从 4 个缩容到 2 个。



同理，用户可以通过增加虚拟数仓的执行节点数量来执行扩容操作。

此外，在新建虚拟数仓或者修改虚拟数仓配置时，用户还可以根据业务需求，通过配置不同的虚拟数仓规格来实现扩缩容（Scale Out/In）。不同的数仓规格对应不同的协调器和执行器资源配置，包括标准和中等两个等级，这对应不同的 CPU 和内存配置。

查看虚拟数仓详情

在「虚拟数仓」页面的虚拟数仓列表中，用户选择目标虚拟数仓并点击 **查看详情** 即可进入指定虚拟数仓的详情页面。

虚拟数仓 / 虚拟数仓详情

vw1

名称	vw1	ID	cl01jfc7q2d1ty8czv27txba6y4d	外部接入方式	PieProxy	审计开关	未开启	执行节点数量	4
状态	运行中	规格	标准	自动启动	已开启	自动关闭	已开启	自动关闭等待时长	10分钟
创建时间	2 天前	备注							

启停记录

名称	ID	执行节点数量	规格	启动时间	停止时间	停止原因	运行时间
vw1	cl01jfc7q2d1ty8czv27txba6y4d	4	标准	2024年12月18日星期三 14:48	2024年12月20日星期五 08:02	数仓冻结	1d 17h
vw1	cl01jfc7q2d1ty8czv27txba6y4d	4	标准	2024年12月18日星期三 14:48	2024年12月18日星期三 14:48	用户关闭	16s
vw1	cl01jfc7q2d1ty8czv27txba6y4d	4	标准	2024年12月18日星期三 14:47	2024年12月18日星期三 14:47	用户关闭	20.2s
vw1	cl01jfc7q2d1ty8czv27txba6y4d	4	标准	2024年12月18日星期三 14:46	2024年12月18日星期三 14:47	用户关闭	1m 7s
vw1	cl01jfc7q2d1ty8czv27txba6y4d	4	标准	2024年12月18日星期三 14:36	2024年12月18日星期三 14:45	用户关闭	8m 31s

该页面主要展示指定虚拟数仓的如下信息：

- 启停记录
- 错误日志
- SQL 审计日志
- 用户的查询信息

提示：

仅 PieDataCS 云原生平台的企业版支持此功能。

- 用户的会话信息

提示：

仅 PieDataCS 云原生平台的企业版支持此功能。

启停记录

启停记录功能以列表形式展示虚拟数仓的启动和关闭历史，方便用户监控和审查虚拟数仓的运行状态。

启停记录列表通常包含以下字段信息：

- 名称：虚拟数仓的名称。
- ID：虚拟数仓的 ID。
- 执行节点数量：虚拟数仓在关闭前所配置的执行节点数量。
- 规格：虚拟数仓的规格类型。

- 启动时间：虚拟数仓的启动时间。
- 关闭时间：虚拟数仓的关闭时间。
- 停止原因，即虚拟数仓关闭的原因，可能为如下之一：
 - 用户关闭：用户手动关闭虚拟数仓。
 - 自动关闭：虚拟数仓开启“自动关闭”开关后，在没有任务运行时，等待至“自动关闭等待时长”后将自动关闭。
 - 数仓冻结：虚拟数仓所属的数仓被冻结，无法使用。
 - 用户修改：用户手动修改了虚拟数仓的配置而触发关闭，待修改生效后，虚拟数仓会自动启动。
- 运行时间：虚拟数仓由启动到关闭的时间间隔。

错误日志

注意：

安装 Monitor 模块后，即可开通 PieDataCS 云原生平台的虚拟数仓的错误日志功能。

错误日志功能以列表形式展示虚拟数仓的错误日志信息，帮助用户更有效地定位和解决相关问题。

错误日志信息列表通常包含以下字段信息：

- 日志级别：日志级别从轻到重分别为 ERROR、FATAL 和 PANIC，默认显示全部级别的日志信息。
- 节点名称：节点类型，包括协调节点（coordinator）和执行节点（executor）。
- 日志内容：日志的详细信息。
- 时间：日志的生成时间。

错误日志信息列表提供如下快捷操作：

- 根据日志级别筛选日志：使用**日志级别**控件可以筛选出指定的一个或多个日志级别的日志信息。默认情况下，展示所有日志级别的历史信息。
- 根据节点类型筛选日志：通过**节点类型**下拉列表中选择特定的节点类型，可以筛选出相应节点类型的日志信息。默认情况下，显示所有节点类型的错误日志。
- 指定时间范围筛选日志：使用**开始时间-结束时间**控件可以指定只显示某个时间范围内的日志信息。
- 根据日志级别排序：使用**排序方式**控件可以根据日志级别对日志信息进行正序或者倒序排列。

SQL 审计

注意:

安装 Monitor 模块后，即可开通 PieDataCS 云原生平台的虚拟数仓的 SQL 审计功能。

SQL 审计功能以列表形式记录并监控在虚拟数仓中执行的所有 SQL 操作。要使用此 SQL 审计功能，用户需要在虚拟数仓列表中启用审计开关。

SQL 审计列表通常包含以下字段信息：

- 执行用户：使用虚拟数仓执行 SQL 的用户名。
- 虚拟数仓 ID：虚拟数仓的 ID。
- 时间：执行 SQL 的时间。
- 数据库：执行 SQL 的数据库。
- SQL 类别：SQL 语句所属的类别，可能为如下之一：
 - READ：数据源存在关系或查询的命令，包括 SELECT 和 COPY。
 - WRITE：目标数据存在关系的命令，包括 INSERT、UPDATE、DELETE、TRUNCATE 和 COPY。
 - FUNCTION：函数调用和 DO 块。
 - ROLE：与角色和权限相关的语句，包括 GRANT、REVOKE、CREATE/ALTER/DROP ROLE。
 - DDL：ROLE 中未包含的所有 DDL。
 - MISC：杂项命令，例如 DISCARD、FETCH、CHECKPOINT、VACUUM、SET。
 - MISC_SET：杂项 SET 命令，例如 SET ROLE。
 - ALL：包括以上所有内容。
 - none：默认项。
- 命令类型：SQL 命令的类型。
- 会话 ID：执行 SQL 的会话（客户端连接）ID。
- 事务 ID：执行 SQL 的事务 ID。
- SQL 参数：执行的 SQL 里所包含的参数。如果设置了 pgaudit.log_parameter，则此字段将包含引用的 CSV 格式的语句参数，如果没有参数则显示 *<none>*。否则，该字段显示 *<not logged>*。
- SQL 文本：所执行的 SQL 语句的完整内容。

SQL 审计列表提供如下快捷操作：

- 根据 SQL 文本搜索信息：搜索框支持通过 SQL 文本来检索相关的审计信息。

- 指定时间范围筛选信息：使用**开始时间-结束时间**控件可以指定只显示某个时间范围内的 SQL 审计信息。
- 根据执行时间排序：使用**排序方式**控件可以根据执行时间的长短将 SQL 审计信息进行正序或者倒序排列。

终止查询

查询功能允许用户查看并在需要时终止自己在该虚拟数仓中执行的 SQL 查询。

查询列表默认显示全部用户在当前虚拟数仓正在执行的 SQL 查询，用户选中自己在该虚拟数仓所执行的 SQL 查询后，点击**终止查询 > 确定** 即可结束该查询任务。



启停记录						
错误日志		SQL 审计		查询		
用户名 全部				会话		
<input type="checkbox"/>	pid	状态	查询语句	用户	数据库	查询开始时间
<input type="checkbox"/>	3542	active	select pg_sleep(1100);	testUser	openpie	2024-08-02 16:18:14
<input checked="" type="checkbox"/>	3705	active	select pg_sleep(20000000);	demoUser	openpie	2024-08-02 16:23:36

共 2 个查询

终止会话

会话功能支持用户查看和管理自己在该虚拟数仓中发起的会话（客户端连接）。

会话列表默认展示所有用户在当前虚拟数仓中发起的会话。用户可以选中自己发起的特定会话，然后点击**终止会话 > 确定** 即可结束该会话。一旦会话被终止，其中的所有查询也将随之终止。



启停记录						
错误日志		SQL 审计		查询		
会话状态 全部		用户名 全部		会话		
<input type="checkbox"/>	pid	会话 ID	用户	数据库	状态	开始时间 应用程序
<input type="checkbox"/>	4178	320	demoUser	openpie	active	2024-08-02 16:38:17
<input checked="" type="checkbox"/>	5017	386	testUser	openpie	idle	2024-08-02 17:09:07

共 2 个会话

会话列表提供如下快捷操作：

- 根据会话状态筛选会话：通过在**会话状态**下拉列表中选择特定的会话状态，可以筛选出相应状态的会话。默认情况下，显示所有状态的会话。会话状态可能为如下之一：
 - active：会话正在工作中。
 - idle：会话处于空闲状态。
 - idle in transaction：事务正在运行但不活跃，可能正在等待用户输入。
 - idle in transaction (aborted)：与“idle in transaction”类似，两者的区别是该状态指示事务中的某条语句导致了事务中止。

- fastpath function call: 会话正在执行一个高效的函数调用。
- disabled: 会话已被关闭。
- 根据用户筛选会话: 通过在**用户名**下拉列表中选择特定用户名, 可以筛选出相应用户的会话。默认情况下, 显示所有用户的会话。
- 按照进程 ID 排序: 点击 “pid” 字段旁边的排序快捷键, 可以根据进程 ID 的大小对列表信息进行排序。
- 按照会话 ID 排序: 点击 “会话 ID” 字段旁边的排序快捷键, 可以根据会话 ID 的大小对列表信息进行排序。
- 按照开始时间排序: 点击 “开始时间” 字段旁边的排序快捷键, 可以根据会话开始时间的由近到远或者由远到近的顺序对列表信息进行排序。

配置外部接入

虚拟数仓支持外部接入功能, 但需要在部署时手动启用。新建的虚拟数仓默认不开启外部接入的功能。一旦启用该功能, 用户在创建虚拟数仓后, 可以通过打开外部接入开关来进行连接操作。

虚拟数仓

The screenshot shows a table with columns: 虚拟数仓名称 (Virtual Warehouse Name), 状态 (Status), 执行节点数量 (Number of Execution Nodes), 规格 (Specification), 外部接入 (External Access), 审计开关 (Audit Switch), and 计费模式 (Billing Mode). A row for 'vw1' is selected, showing '运行中' (Running) status, 4 nodes, Standard specification, and '不开启' (Disabled) external access. The 'External Access' column has a dropdown menu with three options: '不开启' (Disabled), 'Node Port', and 'PieProxy'. The 'Node Port' and 'PieProxy' options are highlighted with a red dashed box.

虚拟数仓名称	状态	执行节点数量	规格	外部接入	审计开关	计费模式
vw1	运行中	4	标准	不开启	未开启	按量计费

共 1 条

需要注意的是, PieDataCS 云原生平台云上云版的虚拟数仓仅支持 PieProxy 作为外部接入方式, 而企业版则提供了 Node Port (默认方式) 和 PieProxy 两种选项。说明如下:

- Node Port: 通过每个 Node 上的固定端口号直接访问。
- PieProxy: 用户可以通过虚拟数仓使用 JDBC、ODBC 或 Postgres 驱动从外部连接数据库来获取数据。

企业版用户可以根据需要更改外部接入的方式。在目标虚拟数仓的“外部接入”操作栏下, 选择相应的外部接入选项即可。但只有虚拟数仓处于“已关闭”或“运行中”状态时, 才能进行外部接入方式的更改操作。

访问虚拟数仓

用户可以通过以下两种方式访问和连接虚拟数仓：

- 通过 WebSQL 的方式连接指定的虚拟数仓来执行 SQL。
- 启用外部接入功能，通过客户端连接虚拟数仓。

通过 WebSQL 方式连接虚拟数仓

用户通过数据洞察功能执行 SQL 查询以获取所需结果。在执行 SQL 查询的过程中，还需要连接到指定的数据库和虚拟数仓。

虚拟数仓具备为查询任务分配计算资源的能力。通过指定数据库，可以确保查询操作仅涉及与当前业务相关的数据；同时，指定虚拟数仓可以根据查询需求动态分配计算资源，以确保查询操作的执行不会影响其他任务的性能。

在「数据洞察」页面的“文件”面板，点击 **新建 SQL 文件** 快捷按钮，或者点击 **新建 > 新建 SQL 文件**，又或者在左侧的文件管理栏中选择一个已创建的 SQL 文件。打开 SQL 文件后，用户可以通过选择数据库和虚拟数仓来执行 SQL 查询。



其中，可选的虚拟数仓状态类型如下：

- 运行中的虚拟数仓。
- 非运行中、非关闭中和非失败状态，但配置了自动启动的虚拟数仓。

上述两种虚拟数仓状态的主要区别在于：当选择的虚拟数仓处于“运行中”状态时，点击 **执行** 会立即执行 SQL 查询；如果选择的虚拟数仓未处于“运行中”状态，但已配置了“自动启动”功能，那么点击 **执行** 会先启动虚拟数仓，随后再执行 SQL 查询。除以上情况外，其他状态的虚拟数仓均不可选。

通过客户端连接虚拟数仓

在通过客户端连接虚拟数仓之前，用户首先需要启用目标虚拟数仓的外部接入功能。在成功获取连接信息后，再进行客户端的连接配置。本文以通用的外部接入方式 PieProxy 为例，连接配置的步骤如下：

第一步：开启 PieProxy 外部接入

在「虚拟数仓」页面的虚拟数仓列表中，找到目标虚拟数仓，并在其外部接入栏下选择“PieProxy”以启动外部接入。成功开启外部接入功能之后，目标虚拟数仓所在列的外部接入栏下就会出现“高级配置”选项。

The screenshot shows a table of virtual warehouses. The columns are: 虚拟数仓名称 (Virtual Warehouse Name), 状态 (Status), 执行节点数量 (Number of Execution Nodes), 规格 (Specification), and 外部接入 (External Access). A row for 'vw1' is selected, showing '运行中' (Running) status, 4 nodes, Standard specification, and 'PieProxy' under External Access. To the right of 'PieProxy', there are two links: '接入详情' (Access Details) and '高级配置' (Advanced Configuration). The '高级配置' link is highlighted with a red dashed box.

虚拟数仓名称	状态	执行节点数量	规格	外部接入
vw1	● 运行中	4	标准	PieProxy ▼ 接入详情 高级配置

目标虚拟数仓的“PieProxy 高级配置”可以用于设置 CIDR 白名单和用户白名单。CIDR 白名单用于限定允许进行外部接入的客户端 IP 地址范围，而用户白名单则用于限定允许进行外部接入的用户范围。如果相应的白名单未设置任何条目，则表示对所有 IP 地址或用户开放访问权限；相反，只有当 IP 地址或用户被列入白名单时，才能访问目标虚拟数仓。



第二步：配置客户端连接

在「虚拟数仓」页面的虚拟数仓列表中，目标虚拟数仓开启 PieProxy 外部接入后，在其“外部接入”栏下点击 **接入详情**，即可查看虚拟数仓的连接信息和连接示例。

提示：

目前支持的客户端工具包括 psql、pgadmin、DBeaver、Navicat（在兼容模式下）等；支持的驱动或封装库包括 pgx、gorm、MyBatis、HikariCP、psycopg2 等。在正常情况下，对于兼容 PostgreSQL V11 及以上版本的客户端工具，外部连接可以同步支持。对于其他版本的客户端工具，可以在外部连接的兼容模式下获得支持。

配置客户端的外部连接时一般涉及如下信息：

- 用户连接信息，包括用户名（即登录平台的用户名称）和虚拟数仓 Token 密码。
- 虚拟数仓信息，包括虚拟数仓服务器的接入 IP 地址、端口号和虚拟数仓 ID。
- 数据库信息，一般只需数据库名称即可。



其中，虚拟数仓的 Token 密码需要通过重置密码的方式获取：首先点击 **忘记了 Token 密码**，然后点击 **重置 Token 密码**。出于安全考虑，PieCloudDB 不会存储此密码，因此用户需要在实时重置后立即记录并保存新密码，或者在每次连接时都进行密码重置操作。

外部接入页面提供了多种连接方式，包括常用的 psql 连接方法、带环境变量的 psql 连接方法，以及一些常用驱动和客户端的连接方法。此外，用户还可以在兼容模式下通过客户端连接到虚拟数仓。这涉及到忽略 PGOPTIONS 配置，并在数据库名称后使用 “#” 号来附加对应的虚拟数仓 ID。例如，如果需要连接的数据库名为 “db1”，虚拟数仓 ID 为 “vw1”，则在兼容模式下，连接的数据库名称应为 “db1#vw1”。在这种情况下，其他连接参数应与 PostgreSQL 的标准连接方式保持一致。

注意：

如果需要使用 SSL/TSL 连接，用户可以选择自行提供相应的证书，或者申请证书以完成对接。目前，此功能尚未支持界面化操作。如需使用，请联系 OpenPie 获取支持。

客户端的默认连接方式分为 Session 模式和 Transaction 模式。其中，Session 模式指的是在客户端连接过程中，会持续占用同一条数据库连接，这便于用户执行连续性或有状态的操作。而 Transaction 模式指的是在客户端连接过程中，多次语句执行可能会使用多条不同的数据库连接，这适合用户执行无状态的操作以及大量并发操作。默认情况下，客户端使用 Session 模式。如果用户希望切换到 Transaction 模式，可以打开“Transaction 模式”按钮，以获取专为 Transaction 模式配置的连接信息和示例。



关于更多客户端的外部接入示例，请参见 [外部接入示例](#)。

连接故障排查

- 问题：客户端通过外部接入连接虚拟数仓，执行 SQL 时报错 “prepared statement "stmtcache_XX" does not exist” 。

排查思路：查看使用的连接模式是否是 Transaction 模式，且是否使用到了 Prepared Statement 的语法。或是 Transaction 模式不支持 Prepared Statement 以及 Session 级别的配置设置，建议切换到 Session 模式后再进行尝试。

- 问题：连接过程中客户端卡住无响应，连接一段时间后发生报错。

排查思路如下：

- 查看对应虚拟数仓是否开启或者自动启动的配置是否开启。
- 查看客户端的并发连接数是否达到上限。默认同用户且同虚拟数仓且同数据库的并发上限为 20，可根据需要进行调整。

- 问题：客户端配置完成后，在连接虚拟数仓时，发生报错 “FATAL: the address from client request is not in the whitelist” 。

排查思路：查看外部接入的客户端 IP 地址是否加入目标虚拟数仓的“PieProxy 高级配置”中的 CIDR 白名单中，只有当客户端 IP 地址被列入白名单时，才能访问目标虚拟数仓。

5.4 数据洞察

在将数据上传至平台并执行 SQL 查询之前，请确保所在计算空间至少有一个可用的虚拟数仓。此外，还需要使用数据洞察功能来创建查询所需的数据库、Schema 以及表等数据库对象。

在计算空间操作界面，用户点击菜单栏中的 **数据洞察** 即可进入功能页面。该页面支持如下操作：

- 数据库对象概览
- 创建 SQL 文件
- 创建 SQL 文件夹
- 执行查询并展示查询结果
- 查看执行计划

数据库对象概览

在「**数据洞察-数据库**」页面，用户可以查看所有数据库对象的列表及其相应的属性信息。用户只需将鼠标悬停在数据库对象名称之上，即可显示该对象的属性信息，并调出隐藏菜单及其选项。

The screenshot shows the 'Database Object Overview' page. On the left, there's a sidebar with tabs for 'Database' and 'File'. Under 'Database', it shows '常用数据库对象' (Common Database Objects) with '暂无常用数据库对象' (No common database objects yet) and '所有数据库对象' (All Database Objects) which lists 'openpie' and 'testdb'. A context menu is open over 'testdb', with options '刷新' (Refresh), '复制对象名到SQL' (Copy object name to SQL), and '添加字段名到SQL' (Add field name to SQL) highlighted with a red box. On the right, there's a large panel titled '数据洞察' (Data Insight) showing the properties of 'testdb'. The properties listed include: oid (17384), owner (stageuser), tablespace (pg_default), encoding (UTF8), collate (en_HK.UTF-8), ctype (en_HK.UTF-8), istemplate (false), allowconn (true), and connlimit (-1). At the bottom right of this panel are two buttons: '新建文件夹' (Create New Folder) and '本地导入 SQL 创建文件' (Import Local SQL File to Create File).

通常，数据库对象右侧的隐藏菜单支持的操作包括：**刷新**、**复制对象名到SQL** 以及 **添加字段名到SQL**（仅适用于表和视图）。

在目标数据对象的属性信息窗口中，用户可以点击窗口右上角的跳转按钮，即可切换到「**数据库**」页面，以便进一步查看该数据库对象的详细信息。更多详细信息，请参见 [数据库](#)。

创建 SQL 文件

在「**数据洞察-文件**」页面，用户可以创建新的 SQL 文件或从本地导入现有的 SQL 文件。基于这些文件，用户能够编写并执行 SQL 语句。为了便于操作，查询区域的初始页面还提供了快捷按钮：**新建 SQL 文件** 和 **本地导入 SQL 创建文件**。此外，用户还可以创建文件夹来组织和管理 SQL 文件。



用户可以通过如下两种方式创建 SQL 文件：

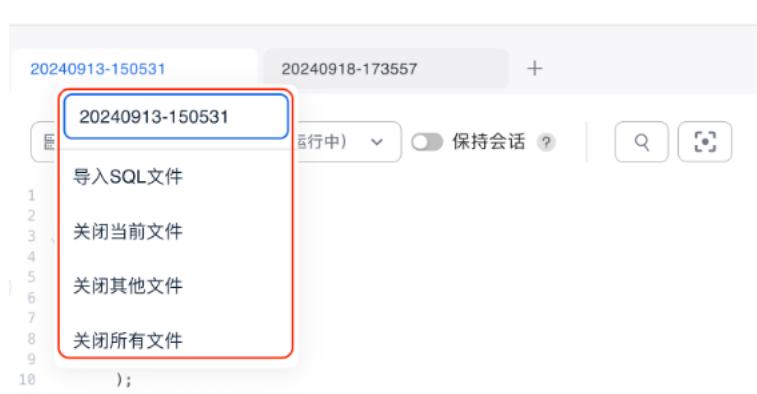
- 新建 SQL 文件

在「**数据洞察-文件**」页面，用户可以点击 **新建>添加 SQL 文件** 来新建一个 SQL 文件。创建 SQL 文件成功后会自动在查询区域打开，用户即可在右侧数据洞察的编辑区域内编写查询语句并执行 SQL 查询。

- 本地导入 SQL 文件

在「**数据洞察-文件**」页面，用户可以点击 **导入**，在随后弹出的窗口中选择存储在本地计算机上的 SQL 文件。选择文件并点击 **打开** 后，SQL 文件中的内容会立即同步显示在数据洞察的查询区域内。通过这种方式导入的本地 SQL 文件与在页面上直接创建的 SQL 文件具有相同的操作功能和特性。

数据洞察



在查询区域，对于已打开的 SQL 文件，点击其名称的下拉列表，可以执行如下操作：

- 重命名：在文件名称编辑框内，删除原有的文件名称，之后输入新的文件名称，点击回车即可修改完成。
- 导入 SQL 文件：将本地 SQL 文件中的内容导入到当前的 SQL 文件窗口中。此操作不会覆盖之前的 SQL。
- 关闭当前文件：在查询区域关闭当前打开的 SQL 文件。
- 关闭其他文件：如果查询区域中存在其他打开的 SQL 文件，则会关闭其他打开的 SQL 文件。
- 关闭所有文件：如果查询区域中存在其他打开的 SQL 文件，则会关闭所有打开的 SQL 文件。



在 SQL 文件列表区域，如果将鼠标悬浮于 SQL 文件名称上方会出现隐藏菜单，可以执行如下操作：

- 重命名：更改目标 SQL 文件的名称。
- 移动到文件夹：如果存在文件夹，可以将目标 SQL 文件移动至指定文件夹内。
- 删除文件：删除目标 SQL 文件。
- 导出文件：自动下载 SQL 文件到本地。

创建 SQL 文件夹

根据实际的使用需求，用户还可以创建文件夹来组织和管理 SQL 文件。

在「数据洞察-文件」页面，用户可以点击 **新建>新建文件夹** 并弹出创建文件夹窗口。在该窗口中输入文件夹的名称，然后点击 **确认**，新创建的文件夹将立即显示在文件夹列表中。此外，查询区域的初始页面还提供了一个快捷按钮 **新建文件夹**，以便用户快速创建文件夹。



在 SQL 文件夹列表区域，如果将鼠标悬浮于文件夹名称上方会出现隐藏菜单，可以执行如下操作：

- 重命名：更改目标 SQL 文件夹的名称。
- 删除文件夹：删除目标 SQL 文件夹及其所包含的文件。需要注意的是，文件夹删除后不可恢复。
- 在此文件夹中新建 SQL 文件：在目标文件夹中新建一个 SQL 文件并自动在查询区域打开。
- 本地导入 SQL 创建文件：选择存储在本地计算机上的 SQL 文件并放入目标文件夹。

在 SQL 文件夹创建完成后，用户可以点击目标 SQL 文件名称上方的隐藏菜单并选择 **移动到文件夹** 选项，之后从弹出的文件夹列表中选中文件夹并单击，当前 SQL 文件即被移动到该文件夹进行归档。



如果用户需要将文件夹中的 SQL 文件移出，点击目标 SQL 文件名称上方的隐藏菜单并选择 **移动到>从文件夹中移出** 选项，即可将目标 SQL 文件移出文件夹。被移出的 SQL 文件会重新出现在 SQL 文件列表中。



执行 SQL 查询并展示查询结果

「数据洞察」页面的查询区域用于编写的 SQL 语句并执行这些查询语句，以及展示查询结果。

用户获得相应权限并进入计算空间后，使用**数据洞察**功能执行查询的步骤一般如下：

1. 创建运行 SQL 查询所需的虚拟数仓。

关于创建虚拟数仓的详细信息，请参见 [虚拟数仓-创建虚拟数仓](#)。

2. 通过**创建 SQL 文件**或者**导入 SQL 文件**的方式来新建一个 SQL 文件，根据查询需求编写相应的 SQL 语句。

本文所使用的示例语句如下：

```
CREATE DATABASE testdb;

CREATE TABLE test_table
(
    id CHAR(10),
    name CHAR(50),
    length INT,
    date CHAR(10),
    type CHAR(20)
);

INSERT INTO test_table VALUES
('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'),
('HG120', 'The Dinner Game', 140, '1985-02-10', 'Comedy');

SELECT * FROM test_table;
```

3. 在数据库下拉列表中选择要查询的数据库（默认数据库为“openpie”）。本文使用 `CREATE DATABASE` 创建了一个数据库 testdb。

```
20240913-150531 +  
testdb vvw1 (运行中) 保持会话  
选择数据库  
openpie  
testdb  
CREATE TABLE test_table  
7   CHAR(10),  
8   length INT,  
9   date CHAR(10),  
10  type CHAR(20)  
);  
11  
12  INSERT INTO test_table VALUES  
13    ('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'),  
14    ('HG120', 'The Dinner Game', 140, '1985-02-10', 'Comedy');  
15  
16  SELECT * FROM test_table;  
17
```

4. 选择一个虚拟数仓运行 SQL 查询。

5. (可选) 打开“保持会话”功能。

“保持会话”功能默认是关闭的。如果开启“保持会话”，则用户在同一次登录会话中执行 SQL 任务时会复用同一个数据库连接。如果需要切换到不同的数据库，请先关闭“保持会话”功能，之后再进行数据库切换操作。

6. 选中需要运行的 SQL 语句后，点击 **执行选中** 运行指定的 SQL 语句。

如果需要执行当前文件中的所有 SQL 语句，用户可以直接点击 **全部执行**，系统将依次运行文件内的所有 SQL 语句。

7. 查看执行结果。

查询结果区域展示了运行该 SQL 语句的执行状态、请求 ID、扫描时间、执行时间以及执行影响记录数等信息。

```

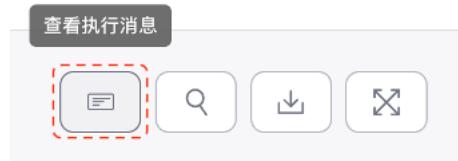
1 v CREATE TABLE test_table
2 (
3     id CHAR(10),
4     name CHAR(50),
5     length INT,
6     date CHAR(10),
7     type CHAR(20)
8 );
9
10 v INSERT INTO test_table VALUES
11     ('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'),
12     ('HG120', 'The Dinner Game', 140, '1985-02-10', 'Comedy');
13
14 SELECT * FROM test_table;

```

执行成功 请求ID: qu01jfs2cf0f7ne4j89rxfzwawwq#3 扫描时间: 16ms
执行时间: 247ms 查询记录数: 2 结果集大小: 221 Bytes

	id	name	length	date	type
1	HG120	The Dinner Game	140	1985-02-10	Comedy
2	B6717	Tampopo	110	1985-02-10	Comedy

在展示查询结果区域，用户点击 **查看执行消息** 快捷按钮，可以在执行查询请求时返回执行信息。“查看执行消息”功能支持在执行过程中显示数据库的提示信息，用户可以通过 RAISE NOTICE 语法将期望输出的信息展示在执行消息中。

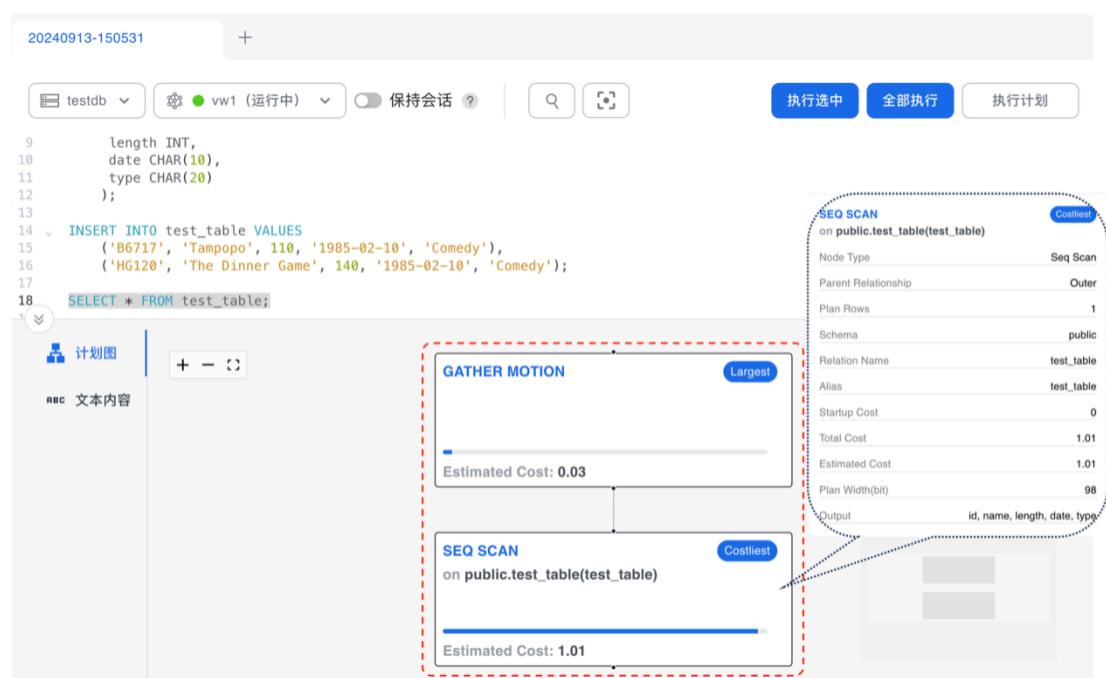


此外，用户还可以单击请求 ID 并直接跳转到「**查询历史**」页面，以便查看该查询的详细信息。详细信息请参见 [查询历史](#)。

查看执行计划

在 SQL 查询执行完成后，对于包含 SELECT、INSERT、UPDATE 和 DELETE 等操作的 SQL 语句，用户可以选择相应的语句，然后点击 **执行计划**。随后，查询结果区域将以图形化和文本形式展示对应 SQL 语句的详细执行计划信息，从而帮助用户理解查询的执行过程和性能特性。

如下为计划图示例，每个节点代表执行计划中的一个步骤，窗口底部的蓝色进度条表示对应节点的执行计划的开销，通常用来快速识别哪些步骤可能成为性能瓶颈。



如果单击某个计划图的步骤，对应窗口会展开并显示详细的执行信息，可能包含如下字段：

- Slice: 所在 Slice 的编号。
- Start Time: 执行节点的初始化时刻。
- Finish Time: 获取最后一条数据的时刻。
- Execution Time: 获取全部数据所需的时间。
- Startup Time: 获取第一条数据所需的时间。
- Actual Rows: 查询在当前步骤生成的行数。
- Node Type: 节点类型，例如 Gather Motion。
- Parent Relationship: 父节点关系。
- Plan Rows: 查询的估计结果行数。
- Startup Cost: 执行计划的初始开销。
- Total Cost: 执行计划的总开销。
- Estimated Cost: 执行计划的估计开销。
- Plan Width: 计划节点的每行宽度。
- Output: 输出结果。
- 与当前运算相关的信息，例如哈希键、合并键、连接条件或过滤条件。

执行计划的文本内容通常由 EXPLAIN 命令生成，该命令输出的结果详细描述了 SQL 语句的执行步骤。在文本执行计划中，每个步骤用箭头 (->) 来标记，表示查询计划的顺序。同时，使用缩进来表示查询计划树中的层级结构，使得整个执行流程的逻辑关系更加清晰。如下为一个计划文本的示例。

```

4
5   CREATE TABLE test_table
6   (
7     id CHAR(10),
8     name CHAR(50),
9     length INT,
10    date CHAR(10),
11    type CHAR(20)
12  );
13
14  INSERT INTO test_table VALUES
15    ('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'),
16    ('HG120', 'The Dinner Game', 140, '1985-02-10', 'Comedy');
17
18  SELECT * FROM test_table;
19

```

计划图
文本内容

Gather Motion 2:1 (slice1; segments: 2) (cost=0.00..1.04 rows=2 width=98)
 Output: id, name, length, date, type
 -> Seq Scan on public.test_table (cost=0.00..1.01 rows=1 width=98)
 Output: id, name, length, date, type
 Optimizer: Postgres query optimizer

5.5 数据库

在「数据库」页面，用户可以查看已创建的数据库及其对象的详细信息，包括属性和权限列表。这些信息有助于用户管理和维护数据库对象。

在计算空间操作界面，用户点击菜单栏中的 **数据库** 即可进入功能页面。该页面还支持如下操作：

- 授予和收回数据库对象的权限
- 全量授权或收回 Schema 下所有表和视图的权限
- 批量授权或收回 Schema 下同一类数据库对象的权限

管理数据库对象权限

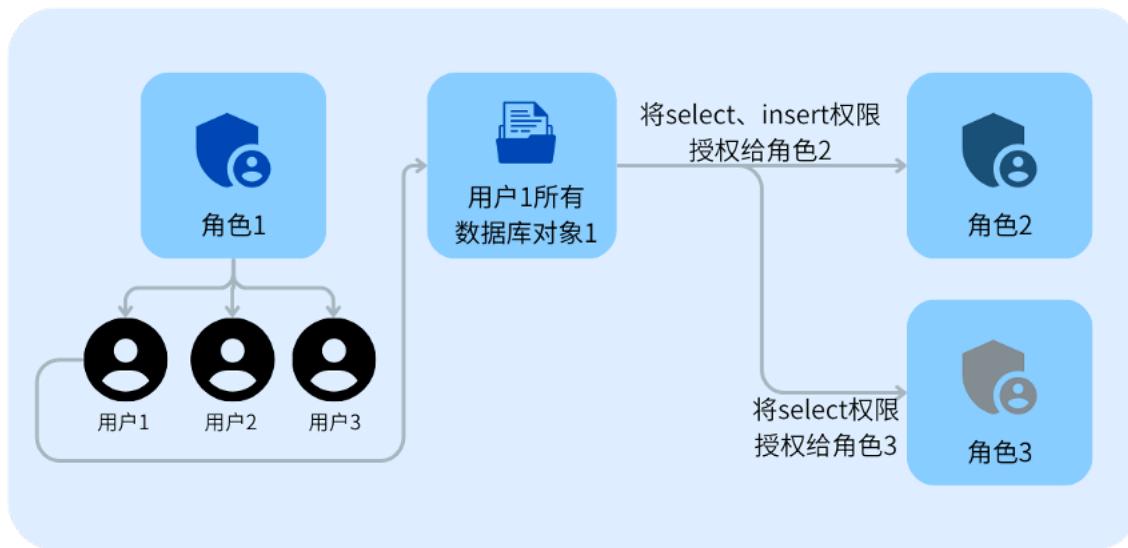
数据库对象的权限分为三个层级：数据库层、Schema 层，以及数据库或 Schema 下的其他数据库对象层。如果某个角色需要访问特定数据库对象的权限，所有者应从数据库层级开始，自上而下地授予相应的权限。通常建议由数据库管理员统一创建和管理数据库对象的权限。

PieCloudDB 的数据库对象可授予的权限参见下表。

数据库对象	可授予权限
数据库	<ul style="list-style-type: none"> CONNECT：允许连接到数据库。 TEMPORARY：允许在使用数据库时创建临时表。 CREATE：允许在数据库中创建新 Schema 和 Publication，并允许在数据库中安装受信任的扩展。

数据库对象	可授予权限
Schema	<ul style="list-style-type: none">USAGE: 允许访问 Schema 中的对象。CREATE: 允许在 Schema 内创建新对象。
表	<ul style="list-style-type: none">SELECT: 允许从表、视图等对象的任意列或特定列执行 SELECT。INSERT: 允许将新行插入到表、视图等。UPDATE: 允许更新表、视图等的所有列或特定列。DELETE: 允许从表、视图等中删除行 (DELETE 命令通常也需要 SELECT 权限来确定要删除哪些行)。TRUNCATE: 允许对表执行 TRUNCATE。REFERENCES: 允许创建表或表中特定列的外键约束。TRIGGER: 允许创建触发器。
外部表	<ul style="list-style-type: none">SELECT: 允许从表、视图等对象的任意列或特定列执行 SELECT。INSERT: 允许将新行插入到表、视图等。UPDATE: 允许更新表、视图等的所有列或特定列。DELETE: 允许从表、视图等中删除行 (DELETE 命令通常也需要 SELECT 权限来确定要删除哪些行)。TRUNCATE: 允许对表执行 TRUNCATE。REFERENCES: 允许创建表或表中特定列的外键约束。TRIGGER: 允许创建触发器。
视图和物化视图	<ul style="list-style-type: none">SELECT: 允许从表、视图等对象的任意列或特定列执行 SELECT。INSERT: 允许将新行插入到表、视图等。UPDATE: 允许更新表、视图等的所有列或特定列。DELETE: 允许从表、视图等中删除行 (DELETE 命令通常也需要 SELECT 权限来确定要删除哪些行)。TRUNCATE: 允许对表执行 TRUNCATE。REFERENCES: 允许创建表或表中特定列的外键约束。TRIGGER: 允许创建触发器。
函数	默认将 EXECUTE 权限授予 PUBLIC。

PieDataCS 云原生平台以角色 (Role) 为基础来管理数据库对象权限。下图展示了常见的数据库对象权限的授权流程。



例如，Adam 需要销售数据的只读权限，作为数仓管理员的 Charlie 需要为 Adam 赋予数据库对象的权限，赋权流程如下：

1. 在「**角色**」页面，管理员 Charlie 创建一个自定义角色“顾客数据分析”。详细操作信息请参见 [角色管理-添加自定义角色](#)。
2. 在「**数据洞察**」页面，管理员 Charlie 创建所需的数据库对象，例如一张存放销售数据的表，其名称为“服装产品数据”。关于创建数据库对象的详细操作信息，请参见 [数据洞察](#)。
3. 在「**数据库**」页面，数仓管理员 Charlie 将所创建的表“服装产品数据”的权限授权给所对应的角色“顾客数据分析”。赋权流程一般如下：
 1. 数据库层级赋权：将表“服装产品数据”所在数据库的 CONNECT 权限授予“顾客数据分析”角色。

提示：

CONNECT 权限是用户对数据库的访问权限，该权限默认已分配给了所有角色。

2. Schema 层级赋权：将表“服装产品数据”的 Schema 所对应的权限授予“顾客数据分析”角色（例如 USAGE 等权限）。
3. 数据库对象层级赋权：将表“服装产品数据”的只读权限，即 SELECT 权限授予“顾客数据分析”角色。

有关授予数据库对象权限的详细操作信息，请参见本章节的 [授予和收回数据库对象的权限](#)。

4. 将角色“顾客数据分析”授予用户 Adam，则 Adam 便具有了服装产品数据的只读权限。详细信息参见 [角色管理-授予角色给用户和收回授权](#)。

查看数据库对象信息

用户在「**数据洞察**」页面创建的数据库对象会同步显示在「**数据库**」页面的数据库列表中。此外，用户也可以点击**刷新**图标，手动刷新数据库信息以确保列表是最新的。

名称	所有者
openpie	openpie
testdb	stageuser
db3	stageuser

在数据库列表中，用户可以通过点击目标数据库对象的名称来展开或收回相应的层级。每个层级都展示了其特有的数据库对象相关信息，具体如下：

- **数据库**

数据库层级展示数据库的详细信息，包括对象类型、所有者、数据库名称、所在的表空间名称、字符编码以及其权限授予情况。此外，该层级还会显示该数据库下的 Schema 和 Extension 列表。

角色名称	权限名称	操作
accountmanager	CONNECT,TEMPORARY	收回
dbmanager	CONNECT,TEMPORARY	收回
usermanager	CONNECT,TEMPORARY	收回

- **Schema**

Schema 层级展示 Schema 的详细信息，包括对象类型、所有者以及其权限授予情况。此外，该层级还会显示该 Schema 下的各类数据库对象，包括表、外表、视图、物化视图和函数，用户可以点击**Tables**、**Views**、**Foreign Tables**、**Views**、**Materialized Views** 或者**Functions** 选项，直接切换到相应用对象类型的页面进行查看和管理。

数据库 / testdb / schemas / tsch

中文 您好: []

Schema 详情	Tables	Foreign Tables	Views	Materialized Views	Functions									
对象类型 Schema	所有者 stageuser													
权限 表 / 视图权限 ▾ 授权 <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>角色名称</th> <th>权限名称</th> <th>操作</th> </tr> </thead> <tbody> <tr> <td>accountmanager</td> <td>USAGE</td> <td>收回</td> </tr> <tr> <td>dbmanager</td> <td>USAGE</td> <td>收回</td> </tr> </tbody> </table>						角色名称	权限名称	操作	accountmanager	USAGE	收回	dbmanager	USAGE	收回
角色名称	权限名称	操作												
accountmanager	USAGE	收回												
dbmanager	USAGE	收回												

- 表

表层级展示表的详细信息，包括对象类型、所有者以及其权限授予情况。此外，该层级还会显示该表的字段信息列表。

数据库 / testdb / tsch / tables / test_table

中文 您好: []

Table 详情	字段信息	数据预览									
对象类型 Table	所有者 stageuser										
权限 授权 <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>角色名称</th> <th>权限名称</th> <th>操作</th> </tr> </thead> <tbody> <tr> <td>accountmanager</td> <td>INSERT,SELECT,UPDATE</td> <td>收回</td> </tr> <tr> <td>dbmanager</td> <td>INSERT,SELECT,UPDATE</td> <td>收回</td> </tr> </tbody> </table>			角色名称	权限名称	操作	accountmanager	INSERT,SELECT,UPDATE	收回	dbmanager	INSERT,SELECT,UPDATE	收回
角色名称	权限名称	操作									
accountmanager	INSERT,SELECT,UPDATE	收回									
dbmanager	INSERT,SELECT,UPDATE	收回									

表层级还提供了一个便捷的“数据预览”功能。用户如果拥有 SELECT 权限，则可以查看目标表的内容，系统默认会展示前 100 条数据。

数据库 / testdb / tsch / tables / test_table

中文 您好: []

Table 详情	字段信息	数据预览																		
		⌚ ● vw2 (运行中) ▾ <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th id="id">id</th> <th>name</th> <th>length</th> <th>date</th> <th>type</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>DC101</td> <td>The Dinner Game</td> <td>140</td> <td>1985-02-10</td> <td>Comedy</td> </tr> <tr> <td>2</td> <td>B6717</td> <td>Tampopo</td> <td>110</td> <td>1985-02-10</td> <td>Comedy</td> </tr> </tbody> </table>		id	name	length	date	type	1	DC101	The Dinner Game	140	1985-02-10	Comedy	2	B6717	Tampopo	110	1985-02-10	Comedy
	id	name	length	date	type															
1	DC101	The Dinner Game	140	1985-02-10	Comedy															
2	B6717	Tampopo	110	1985-02-10	Comedy															

在数据库列表中，与表同一层级的数据库对象还包括：

- 外部表：展示外部表的详细信息，包括表的名称、所有者、FDW 名称与 Oid、Foreign Server 名称与 Oid、相关配置信息以及其权限授予情况。此外，该层级还会显示该外部表的字段信息列表。

The screenshot shows the 'Foreign Table 详情' tab selected for a table named 't3'. It displays basic information: Name: t3, Owner: stageuser, FDW Name: postgres_fdw, FDW Oid: 17439, Foreign Server Name: fdw_server_a2db3, Foreign Server Oid: 17440, Configuration: schema_name=sch3,table_name=t3, and Description: -. Below this, there's a 'Permissions' section showing grants for 'accountmanager' and 'usermanager' roles with 'SELECT' privileges, each with a 'Revoke' link. A blue '授权' (Grant) button is located at the top right.

角色名称	权限名称	操作
accountmanager	SELECT	收回
usermanager	SELECT	收回

外部表页面也支持便捷的“数据预览”功能。用户如果拥有 SELECT 权限，则可以查看目标外部表的内容，系统默认会展示前 100 条数据。

- 视图/物化视图：展示视图的详细信息，包括对象类型、所有者以及其权限授予情况。此外，该层级还会显示该视图的源码和字段信息列表。

The screenshot shows the 'View 详情' tab selected for a view named 'v1'. It displays basic information: Object Type: View, Owner: stageuser. Below this, there's a code editor showing the SQL source code:

```
SELECT tbl1.id  
FROM tbl1;
```

 with 'Fullscreen View' and 'Copy Code' buttons. A 'Permissions' section shows grants for 'accountmanager' and 'usermanager' roles with 'SELECT' privileges, each with a 'Revoke' link. A blue '授权' (Grant) button is located at the top right.

角色名称	权限名称	操作
accountmanager	SELECT	收回
usermanager	SELECT	收回

视图页面也支持便捷的“数据预览”功能。用户如果拥有 SELECT 权限，则可以查看目标视图的内容，系统默认会展示前 100 条数据。

- 函数：展示函数的详细信息，包括对象类型、所有者以及其权限授予情况。此外，该层级还会显示该视图的源码信息。

数据库 / testdb / public / functions / add_two_numbers

中文 您好: 

Function 详情

对象类型 Function 所有者 stageuser

名称 add_two_numbers 语言 plpgsql 参数类型 int1 integer, int2 integer

返回值类型 integer

```
源码 CREATE OR REPLACE FUNCTION public.add_two_numbers(int1 integer, int2 integer)
RETURNS integer
LANGUAGE plpgsql
AS $function$
BEGIN
    RETURN int1 + int2;
END;
$function$
```

权限

角色名称	权限名称
accountmanager	EXECUTE
dbmanager	EXECUTE
usermanager	EXECUTE

授予和收回数据库对象权限

在数据库列表中，用户可以点击目标数据库对象的名称来查看其权限信息。对于当前用户所拥有的数据库对象，用户还可以执行授予和收回权限的相关操作，但不适用于函数。默认情况下，所有用户都被赋予了函数的使用权限。

授予数据库对象权限的操作步骤如下：

1. 在「数据库」页面选中目标数据库对象后，在权限区域点击 **授权** 以弹出授予权限窗口。

数据库 / testdb / sch3 / tables / t3

中文 您好: 

Table 详情 字段信息 数据预览

对象类型 Table 所有者 stageuser

权限

授权

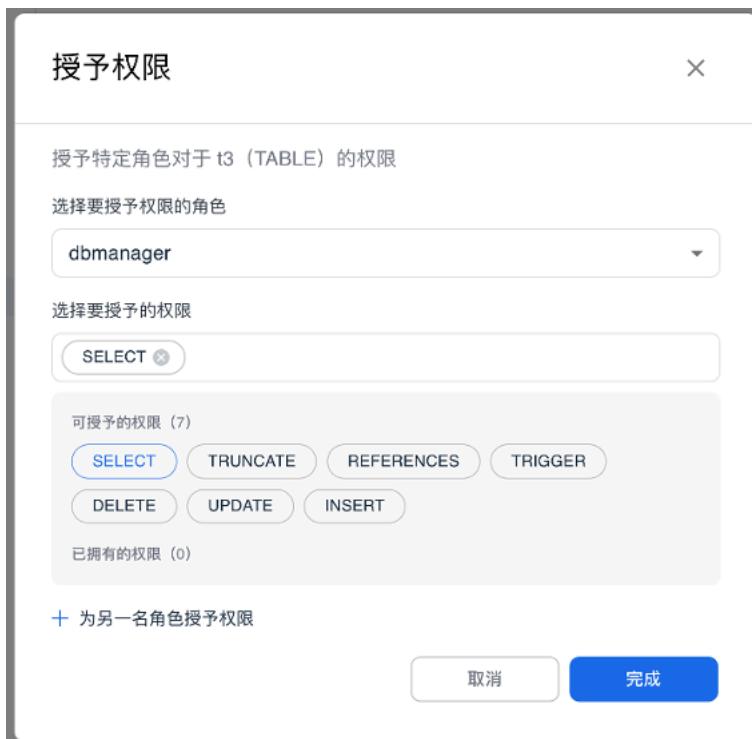
角色名称 权限名称 操作

暂无被授权的角色

- 在下拉列表中选择要授予权限的角色（支持同时为多个角色授予不同的目标数据库对象权限），页面将同步显示可授予该角色的数据库对象权限和该角色已拥有的数据库对象权限。

如果所选择的角色具有较低的权限级别，则高于该级别的角色也会自动继承该数据库对象的权限。例如，如果将目标数据库对象的权限授予给 dbmanager，则 accountmanager 也会自动继承该数据库对象的权限。关于角色的继承与权限赋予的更多详细信息，请参见 **管理角色**。

- 选择要授予该角色的目标数据库对象权限。



- 点击 **完成**，操作成功后，被授予权限的角色名称及其对应的权限信息将被同步更新至权限列表中。

Table 详情			字段信息	数据预览	操作									
对象类型: Table	所有者: stageuser													
权限 <table border="1"> <thead> <tr> <th>角色名称</th> <th>权限名称</th> <th>操作</th> </tr> </thead> <tbody> <tr> <td>accountmanager</td> <td>SELECT</td> <td>收回</td> </tr> <tr> <td>dbmanager</td> <td>SELECT</td> <td>收回</td> </tr> </tbody> </table>						角色名称	权限名称	操作	accountmanager	SELECT	收回	dbmanager	SELECT	收回
角色名称	权限名称	操作												
accountmanager	SELECT	收回												
dbmanager	SELECT	收回												

如果需要收回授予某角色的数据库对象的权限，参考步骤如下：

- 在「数据库」页面选中目标数据库对象后，在权限区域点击目标角色的“操作”菜单栏下的**收回**以弹出收回权限窗口。

The screenshot shows the 'Table 详情' tab selected in the top navigation bar. Under the 'Permissions' section, there is a table listing roles and their permissions:

角色名称	权限名称	操作
accountmanager	SELECT	收回
dbmanager	SELECT	收回

- 在下拉列表中选择选择要收回权限的角色（支持同时收回多个角色的不同的目标数据库对象权限），页面将同步显示可收回的数据库对象权限。

当收回权限时，如果目标角色的权限级别较低，则所有高于该级别的角色也会自动失去相应的权限。此外，如果尝试仅收回继承的角色的数据库对象权限，即使系统显示成功也是不生效的。关于角色的继承与权限赋予的更多详细信息，请参见**管理角色**。

- 选择要收回的目标数据库对象权限。



- 点击**完成**，操作成功后，数据库对象的权限列表会同步更新为权限更改后的信息。

全量授权或收回 Schema 下所有表和视图的权限

对于当前用户所拥有的 Schema，支持全量授权或收回对应 Schema 下属于该用户的所
有表和视图的权限。

全量授权目标 Schema 下的表和视图权限的操作步骤如下：

1. 在「数据库」页面选中目标 Schema 后，点击权限列表的表/视图权限下拉列表中的全量授权以弹出全量授予权限的窗口。



2. 在下拉列表中选择要授予权限的角色和相应的数据库对象权限。



3. 点击 完成，操作成功后，被授予权限的角色名称及其对应的权限信息将被同步更新至
权限列表中。

提示：

如果对某个角色进行表和视图的全量授权，那么该角色也将自动获得目标 Schema 的 USAGE 权限。

如果需要全量收回目标 Schema 下的表和视图的权限，参考步骤如下：

1. 在「数据库」页面选中目标 Schema 后，点击权限列表的**表/视图权限**下拉列表中的**全量收回**以弹出全量收回权限窗口。

The screenshot shows the 'testdb / schemas / sch3' schema permissions page. The 'Tables' tab is selected. In the 'Permissions' section, there is a dropdown menu labeled 'Table / View Permissions' with two options: 'Grant' and 'Revoke'. The 'Revoke' option is highlighted with a red rectangle. Below the dropdown, there is a table listing roles and their permissions:

Role Name	Permission Name	Action
accountmanager	USAGE	Revoke
usermanager	USAGE	Revoke

2. 在下拉列表中选择要收回权限的角色和相应的数据库对象权限。

The screenshot shows the 'Full Table / View Revoke Permission' dialog box. It has the following fields:

- Role selection: 'usermanager' (selected)
- Permission selection: 'SELECT' (selected)
- Reversible permission list: 'SELECT', 'INSERT', 'UPDATE', 'DELETE', 'TRUNCATE', 'TRIGGER', 'REFERENCES' (with 'SELECT' highlighted)
- Buttons: 'Cancel' and 'Finish' (the 'Finish' button is highlighted with a blue rectangle)

3. 点击**完成**，操作成功后，数据库对象的权限列表会同步更新为权限更改后的信息。

批量授权或收回 Schema 下数据库对象的权限

对于当前用户所拥有的 Schema 下的某一类数据库对象执行批量授权或收回数据库对象权限，包括表、外部表、视图和物化视图，但该功能不适用于函数。默认情况下，所有用户都被赋予了函数的使用权限。

全量授权目标 Schema 下的某一类数据库对象权限的操作步骤如下：

- 在「数据库」页面进入目标 Schema 层级后，切换到目标数据库对象的面板，点击**批量授权**下拉列表中的**批量授权**以弹出批量授予权限的窗口。

Database / db3 / schemas / sch3

中文 您好: [REDACTED]

Schema 详情 Tables Foreign Tables Views Materialized Views Functions

Table 列表

名称	所有者
t3	stageuser
tbl1	stageuser

- 在下拉列表中选择要授予权限的角色，并支持将权限批量授予多个角色。

批量授予权限

授予特定角色对于 sch3 中的 (TABLE) 的权限

选择要授予权限的角色 (支持批量选择)

usermanager, dbmanager

第1组TABLE

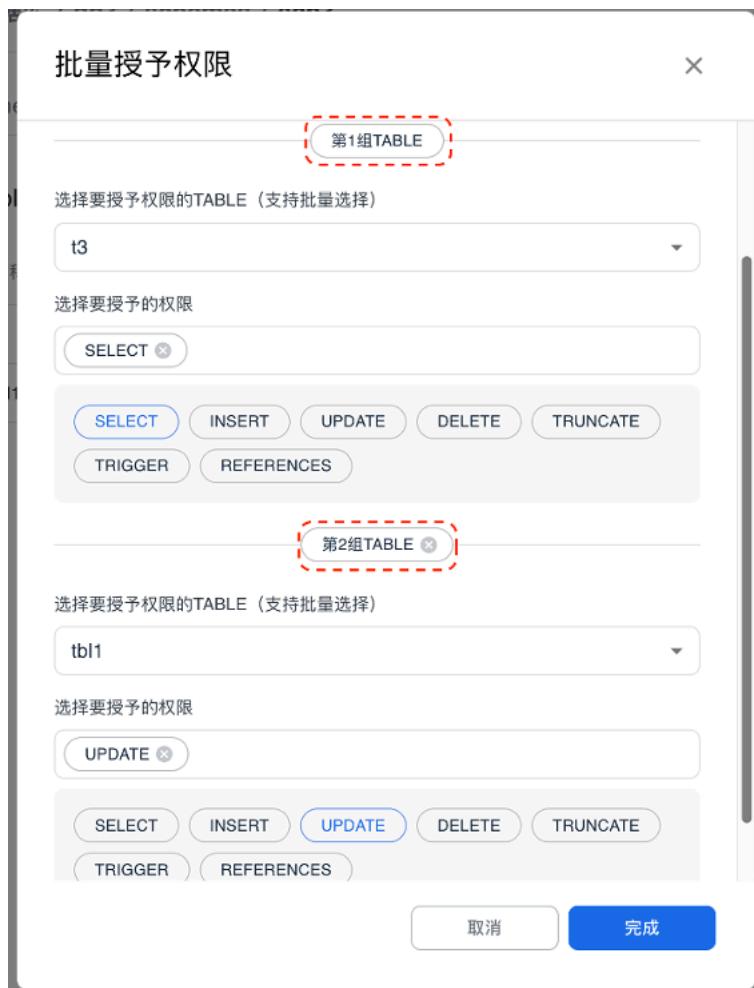
选择要授予权限的TABLE (支持批量选择)

选择要授予的权限

SELECT, INSERT, UPDATE, DELETE, TRUNCATE, TRIGGER, REFERENCES

+ 为另一组TABLE授予权限

- 选择至少一组数据库对象并选择要授予的权限。



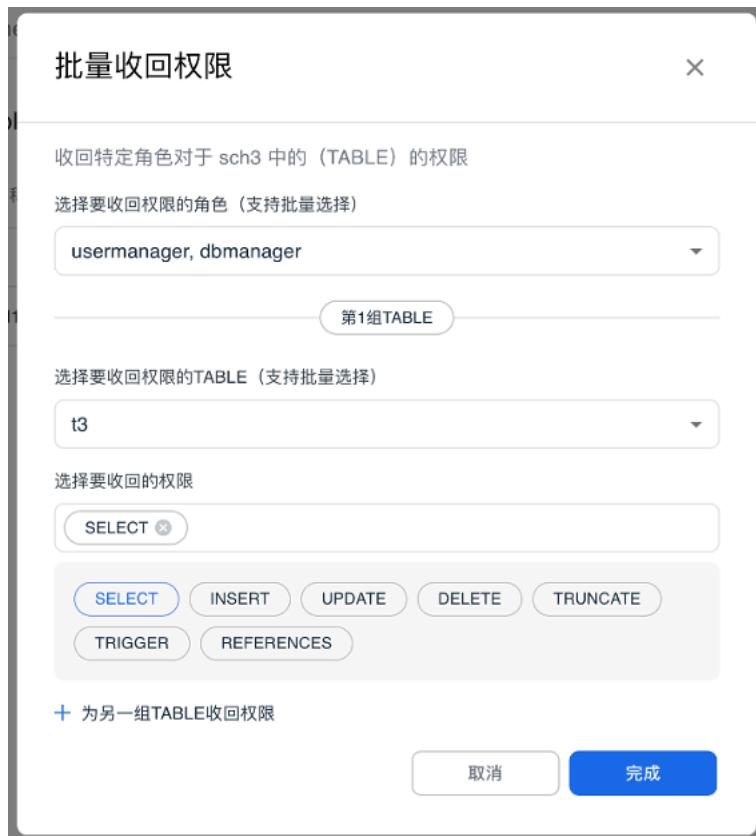
4. 点击 **完成**，操作成功后，被授予权限的角色名称及其对应的权限信息将被同步更新至所选数据库对象的权限列表中。

如果需要全量收回目标 Schema 下的某一类数据库对象的权限，参考步骤如下：

1. 在「数据库」页面进入目标 Schema 层级后，切换到目标数据库对象的面板，点击**批量收回**下拉列表中的**批量收回**以弹出批量收回权限的窗口。

名称	所有者
t3	stageuser
tbl1	stageuser

2. 在下拉列表中选择要收回权限的角色，并支持批量收回多个角色的权限。
3. 选择至少一组数据库对象并选择要收回的权限。



4. 点击 **完成**，操作成功后，数据库对象的权限列表会同步更新为权限更改后的信息。

5.6 快速导入

在计算空间操作界面，具有相应管理权限的用户（例如数仓管理员和数据库管理员）点击菜单栏中的 **快速导入** 即可进入功能页面。该页面支持如下操作：

- 导入数据
- 查看导入历史

导入数据

PieDataCS 云原生平台支持从本地上传 csv 或 text 文件中的数据到目标表并对数据进行分析。

本文以上传一份虚拟的电商数据为例，介绍导入数据的具体操作步骤。

1. 在「**数据洞察**」页面，创建所需的数据库对象，包括数据库、Schema 和表。

1. 选择默认数据库“openpie”和已创建的虚拟数仓，运行以下 SQL 语句以创建所需的数据库。本文所使用的数据库名称为“线上销售数据”。

```
CREATE DATABASE 线上销售数据;
```

2. 切换到数据库“线上销售数据”，然后运行以下 SQL 语句以创建三个 Schema：“服装销售数据”、“食品销售数据”和“顾客数据”。

```
CREATE SCHEMA 服装销售数据;
```

```
CREATE SCHEMA 食品销售数据;
```

```
CREATE SCHEMA 顾客数据;
```

3. 运行以下 SQL 语句以在新创建的 Schema 上创建所需的表。

```
-- 创建存放食品产品数据的表

CREATE TABLE 食品销售数据.食品产品数据 (
    产品编号 VARCHAR(10) NOT NULL,
    原料 VARCHAR(5),
    类型 VARCHAR(5),
    价格 FLOAT,
    库存 INT,
    产品图片 TEXT
);
-- 创建存放 2020 年至 2023 年食品交易数据的表

CREATE TABLE "食品销售数据".交易数据_2020_2023 (
    交易编号 VARCHAR(10) NOT NULL,
    顾客序号 VARCHAR(10) NOT NULL,
    产品编号 VARCHAR(10) NOT NULL,
    交易日期 VARCHAR(10),
    交易时间 TIME,
    件数 INT,
    平台 VARCHAR(5)
);

-- 创建存放服装产品数据的表

CREATE TABLE 服装销售数据.服装产品数据 (
    产品编号 VARCHAR(10) NOT NULL,
    颜色 VARCHAR(5),
    类型 VARCHAR(5),
    价格 FLOAT,
    库存 INT,
    产品图片 TEXT
);
-- 创建存放 2020 年至 2023 年服装交易数据的表

CREATE TABLE "服装销售数据".交易数据_2020_2023 (
    交易编号 VARCHAR(10) NOT NULL,
    顾客序号 VARCHAR(10) NOT NULL,
    产品编号 VARCHAR(10) NOT NULL,
    交易日期 VARCHAR(10),
    交易时间 TIME,
    件数 INT,
    平台 VARCHAR(5)
);

-- 创建存放 2020 年至 2023 年顾客数据的表

CREATE TABLE 顾客数据.顾客数据_2020_2023 (
    顾客序号 VARCHAR(10) NOT NULL,
    顾客姓名 VARCHAR(5),
    生日 DATE,
    注册日期 DATE,
    手机号 VARCHAR(11),
    省份 VARCHAR(10),
    城市 VARCHAR(10),
    地区 VARCHAR(10),
    地址 VARCHAR(100)
);
```

2. 在「快速导入」页面，点击 **导入数据** 进入导入数据的操作页面，根据如下步骤上传单份数据文件至目标表。



1. 选择可用的虚拟数仓、数据库和要上传数据的目标表。

在上传数据前，用户可以通过页面上的“查看目标表字段”功能来查看目标表的字段信息，如下图所示。

This screenshot shows the same 'Quick Import' interface as above, but with the 'View Target Table Fields' button highlighted with a red rectangle. Below it, a table displays the fields for the '顾客数据_2020_2023' table:

列名	类型	是否非空
顾客序号	character varying(10)	true
顾客姓名	character varying(5)	false
生日	date	false
注册日期	date	false
手机号	character varying(11)	false
省份	character varying(10)	false
城市	character varying(10)	false

2. 点击 **选择文件** 或拖拽本地数据文件上传至平台。
3. (可选) 点击预览图标可以查看所上传文件的内容。

4. 点击“高级选项”快捷按钮可以调整单个文件的上传设置。修改完成后，点击**确认**保存设置即可。



“高级选项”中的设置项包括如下：

- 分隔符：指定文件中的字段分隔符。默认值为逗号（，）。
- 表头：如果文件包含表头，则需要开启此选项按钮。默认为不开启。
- 空字符串：指定字符串来表示空值。默认使用不带引号的空字符。
- 换行符：指定换行字符。
- 转义字符：指定转义字符。默认使用双引号（"）。
- 异常记录日志：是否开启记录异常日志。默认为不开启。
- 错误数量限制：设置文件中允许的错误数量。如果文件中的错误数量小于这个设定值，文件将会被成功上传。请注意，该选项的取值应为 0 或者大于 1 的整数。
- 文件完整性校验：比较导入数据前后上传文件的校验和。启用该功能后，如果校验和不匹配，则不导入任何数据。默认为不开启。

注意：

该步骤中的上传设置只针对单个文件，而非全局文件的设置。在上传多个文件时，可以通过全局的“高级选项”调整全部的文件上传设置。如果文件未单独设置上传选项，则默认使用全局设置；如果需要上传的数据都包含表头，在开始上传前，需要开启高级选项中的表头功能。

5. 点击**开始**上传文件。数据导入成功后，系统会提示“上传成功”。

The screenshot shows the πDataCS interface with the following details:

- Top navigation bar: OpenPie | πDataCS
- Left sidebar: vm1 (运行中) | 线上销售数据 | 顾客数据 / 顾客数据_2020_2023
- Main area:
 - Info message: 所有文件将被上传到 线上销售数据 数据库的 顾客数据_2020_2023 表, [查看目标表字段](#)
 - Upload area: 拖拽文件到这里, 或者点击选择文件 (with a cloud icon and an upward arrow)
 - File list:
 - customer.csv (导入格式: csv 916.65 KB) - status: 上传成功 ✓
 - Message: 导入数据成功, 共导入 8000 条数据
- Bottom buttons: 高级选项, 清空, 全部开始

快速导入功能也支持将多份数据上传至数据库的同一张表。

The screenshot shows the πDataCS interface with the following details:

- Top navigation bar: OpenPie | πDataCS
- Left sidebar: vm1 (运行中) | 线上销售数据 | 服装销售数据 / 交易数据_2020_2023
- Main area:
 - Info message: 所有文件将被上传到 线上销售数据 数据库的 交易数据_2020_2023 表, [查看目标表字段](#)
 - Upload area: 拖拽文件到这里, 或者点击选择文件 (with a cloud icon and an upward arrow)
 - File list:
 - clothing_transaction_p0.csv (导入格式: csv 25.48 MB) - status: 点击开始 ↑
 - clothing_transaction_p1.csv (导入格式: csv 21.08 MB) - status: 点击开始 ↑
 - clothing_transaction_p2.csv (导入格式: csv 19.63 MB) - status: 点击开始 ↑
 - clothing_transaction_p3.csv (导入格式: csv 13.73 MB) - status: 点击开始 ↑
 - clothing_transaction_p4.csv (导入格式: csv 9.46 MB) - status: 点击开始 ↑
- Bottom buttons: 高级选项 (circled), 清空, 全部开始

例如，用户的“服装交易数据”由多个数据文件组成，选中这些文件，然后点击**全部开始**来依次上传多个文件。在上传多个文件前，用户可以使用全局的“高级选项”来统一调整所有文件的上传设置。一旦这些文件的数据被加载到相应的表中，用户就可以执行联合查询，以进行深入的数据分析。

查看导入历史

用户每次导入数据的信息（包括分隔符、表头等设置）都会被记录在导入历史中。

在「**快速导入**」页面，选中目标历史任务并展开子任务列表，点击展开按钮(>)，即可展示详细的导入任务信息。这些信息包括了导入数据时所应用的高级选项设置，如下图所示。

The screenshot shows the 'Import History' page with a list of completed imports. One specific import is expanded to show its details:

任务ID	上传人	虚拟数仓名称	原始文件名	文件大小	已导入数据行数	导入状态	创建时间	结束时间
519cb46b91ac29ac	stageuser	vw1	customer.csv	916.65 KB	8000	导入成功	21小时前	21小时前

Below the table, a 'Task Details' section is highlighted with a red dashed box, showing configuration parameters:

- database: 线上销售数据
- 限制错误数量: 2
- 转义字符: "
- schema: 账客数据
- 分隔符: ,
- 文件是否包含表头: false
- table: 顾客数据_2020_2023
- 将此字符串视为null: null
- 是否记录异常日志: false
- 换行符: LF

5.7 查询历史

在「**查询历史**」页面，用户可以查看在其数据洞察所提交的所有查询信息，包括SQL文本、查询状态、请求ID、执行时间、开始时间以及结果集信息等。

在计算空间操作界面，用户点击菜单栏中的**查询历史**即可进入功能页面。该页面主要展示两部分内容：

- **查询历史列表**
- **查询历史详情**

查询历史列表

查询历史列表通常包含以下字段信息：

- SQL 文本：查询所执行的 SQL 语句。
- 状态，即查询的执行状态，可能为如下状态之一：
 - 已创建：查询已创建请求。
 - 运行中：查询正在执行中。
 - 已完成：查询已完成。
 - 已取消：查询被取消。
 - 无权限：没有查询权限。
 - 失败：查询在执行时发生报错。
- 请求 ID：标识查询请求的字符串。
- 执行时间：查询产生结果所需的时间。如果查询包含子任务，则只展示子任务的执行时间。
- 开始时间：查询开始执行的时间。
- 结果集大小：查询产生的结果集所占存储的大小。

当执行包含多条 SQL 语句的查询时，用户可以通过点击 **展开子任务列表** 按钮 (>) 来展示每条查询语句的详细历史信息。如下图所示。

The screenshot shows the 'Query History' interface. At the top, there's a search bar labeled '输入SQL关键词查询' and a dropdown menu for '状态' (Status) set to '全部' (All). Below the header is a table with columns: SQL文本 (SQL Statement), 状态 (Status), 请求ID (Request ID), 执行时间 (Execution Time), 开始时间 (Start Time), 结果集大小 (Result Set Size), and 操作 (Operations).

SQL文本	状态	请求ID	执行时间	开始时间	结果集大小	操作																												
> -- 创建存放食品产品数据的表 CREATE TA...	● 已完成	qu01j8hmvgv3gnjwhey9xfmkddbmy	1 天前			删除																												
▼ CREATE SCHEMA 服装销售数据; CREAT...	● 已完成	qu01j8hmg00r6stxyh0p596scz2a	1 天前			删除																												
<table border="1"> <thead> <tr> <th>子任务</th> <th>状态</th> <th>子任务ID</th> <th>结果记录数</th> <th>执行时间</th> <th>开始时间</th> <th>结果集大小</th> </tr> </thead> <tbody> <tr> <td>CREATE SCHEMA 服装销售数据;</td> <td>● 已完成</td> <td>1</td> <td>0</td> <td>329ms</td> <td>1 天前</td> <td></td> </tr> <tr> <td>CREATE SCHEMA 食品销售数据;</td> <td>● 已完成</td> <td>2</td> <td>0</td> <td>123ms</td> <td>1 天前</td> <td></td> </tr> <tr> <td>CREATE SCHEMA 顾客数据;</td> <td>● 已完成</td> <td>3</td> <td>0</td> <td>127ms</td> <td>1 天前</td> <td></td> </tr> </tbody> </table>							子任务	状态	子任务ID	结果记录数	执行时间	开始时间	结果集大小	CREATE SCHEMA 服装销售数据;	● 已完成	1	0	329ms	1 天前		CREATE SCHEMA 食品销售数据;	● 已完成	2	0	123ms	1 天前		CREATE SCHEMA 顾客数据;	● 已完成	3	0	127ms	1 天前	
子任务	状态	子任务ID	结果记录数	执行时间	开始时间	结果集大小																												
CREATE SCHEMA 服装销售数据;	● 已完成	1	0	329ms	1 天前																													
CREATE SCHEMA 食品销售数据;	● 已完成	2	0	123ms	1 天前																													
CREATE SCHEMA 顾客数据;	● 已完成	3	0	127ms	1 天前																													
CREATE DATABASE 线上销售数据;	● 已完成	qu01j8hmf05rkdq1qc5y3fed53gn	1.2s	1 天前		删除																												
> CREATE VIEW sch3.v1 AS SELECT * FR...	● 已完成	qu01j8gwqf6yvar69addqayzs4am	1 天前			删除																												
> CREATE TABLE sch3.tbl1 (id INTEGER); ...	● 失败	qu01j8gwq99cj7kexk0kc1xqe0ny	1 天前			删除																												

「**查询历史**」页面中的查询历史列表提供如下快捷操作：

- 根据 SQL 关键词搜索：使用搜索框，用户可以通过 SQL 关键词来搜索相关的 SQL 文本信息。
- 根据查询的执行状态筛选信息：在 **状态** 控件的下拉列表中，用户可以选择特定的执行状态来筛选历史查询信息。默认情况下，展示所有状态的历史查询信息。

- 指定时间范围筛选信息：使用**时间范围控件**，用户可以指定一个时间范围（可按分钟、小时或天为单位）来筛选历史查询信息。
- 查看历史查询详情：将鼠标悬浮于查询历史列表中高亮显示的 SQL 文本上，直接点击文本或在弹出的浮窗中点击**查看详情**，即可进入「**查询历史-查询详情**」页面。
- 按照开始时间排序：点击“开始时间”字段旁边的排序快捷键，用户可以根据开始时间的先后顺序来排序列表信息。
- 下载结果集（csv 文件格式）到本地：在“结果集大小”操作栏下，点击**下载**快捷键，用户可以将结果集导出为 CSV 文件并下载到本地。
- 删除查询信息：在“结果集大小”操作栏下，点击**删除**并确认，即可删除当前行的查询历史记录。

查询历史详情

SQL文本	状态	请求ID	执行时间	开始时间	结果集大小
<code>SELECT * FROM sch3.t3;</code>	已完成	qu01j8pgwbfty1dyzrnclj5222w8	595ms	1 分钟前	42 Bytes
<code>SELECT * FROM sch3.t3;</code>	× 查看详情	J01j8hmvg3gnjwhey9xfmkddbmy		2 天前	
<code>UNATCH UNATCH 放弃消息队列; UNATCH...</code>	...	qu01j8hmgo0r6stxyh0p596scz2a		2 天前	

在「**查询历史**」页面，用户可以通过将鼠标悬浮于查询历史列表中高亮的 SQL 文本，直接点击文本或在浮窗点击**查看详情** 即可查看特定查询的详细信息。

状态	已完成	执行时间	595ms	扫描时间	23ms
请求耗时	595ms	数据库	db3	用户名	stageuser
开始时间	4 分钟前	结束时间	4 分钟前	虚拟数仓	vw1
执行节点数量	2	查询记录数	3	结果集大小	42 Bytes
所属会话	se01j8pgvaeft756vvvf6dc026e7				

SQL文本

```
SELECT * FROM sch3.t3;
```

查询详情 页面主要展示如下信息：

- 状态，即查询的执行状态，可能为如下状态之一：
 - 已创建：查询已创建请求。
 - 运行中：查询正在执行中。

- 已完成：查询已完成。
- 已取消：查询被取消。
- 无权限：没有查询权限。
- 失败：查询在执行时发生报错。
- 执行时间：查询产生结果所需的时间。
- 扫描时间：执行查询时，数据库遍历找到满足查询条件的数据所需要的时间。请注意，只有 SELECT 查询会显示此信息。
- 请求耗时：发送查询请求到接收到响应所经过的时间。
- 数据库：被查询的数据库的名称。
- 用户名：提交查询的用户。
- 开始时间：查询开始执行的时间。
- 结束时间：查询终止执行的时间。
- 虚拟数仓：执行此查询的虚拟数仓名称。
- 执行节点数量：执行此查询的虚拟数仓的执行节点的数量。
- 执行影响记录数：查询执行数据库操作（例如插入、更新、删除等）后，所影响的数据库表中的数据数量。
- 结果记录数：执行查询操作后，从数据库返回的结果集中包含的记录数量。请注意，只有 SELECT 查询会显示此信息。
- 所属会话：查询所属会话的信息。请注意，如果没有在「**数据洞察**」开启“保持会话”功能，则不会出现此信息。
- 错误信息：如果查询执行失败，则会显示对应的报错信息。

此外，在 SQL 文本区域会展示查询所执行的 SQL 语句。点击 **在数据洞察中打开** 可以跳转到「**数据洞察**」编辑区域并自动创建一个包含该 SQL 的文件窗口，用户可以直接执行 SQL 文本对应的查询任务。详细信息请参见 [数据洞察](#)。

5.8 用量统计

在「**用量统计**」页面，用户可以查看用户数据和元数据的存储总量、虚拟数仓的使用量和所使用的总流量。

提示：

仅 PieDataCS 云原生平台的云上云版本支持此功能。

在计算空间操作界面，具有相应管理权限的用户（例如数仓管理员和数据库管理员）点击菜单栏中的 **用量统计** 即可进入功能页面。



该页面以柱形图的形式展示如下指标信息：

- 存储量：分别统计元数据、用户数据以及数据请求临时缓存所消耗的存储量和它们各自的百分比，并计算这三者的存储量总和。
存储量包括元数据存储、数据库使用的对象存储、数据洞察查询缓存使用的对象存储以及数据导入缓存使用的对象存储。
- 虚拟数仓使用量：根据计算公式“节点数*小时”，统计各虚拟数仓的使用量，并计算它们占总使用量的百分比。
- 流量：统计通过 PieProxy 外部接入方式连接至 PieCloudDB 虚拟数仓时，执行查询等操作所消耗的流量及其所占的百分比。

用户将鼠标悬停在任意柱形图上方，系统即会显示各指标对应时间点的值。

5.9 管理计算空间用户

在「**用户**」页面，用户可以查看当前数仓下所有用户的激活和审核状态。

在计算空间操作界面，具有相应管理权限的用户（例如用户管理员、数仓管理员和数据库管理员）点击菜单栏中的 **用户** 即可进入功能页面。该页面支持如下操作：

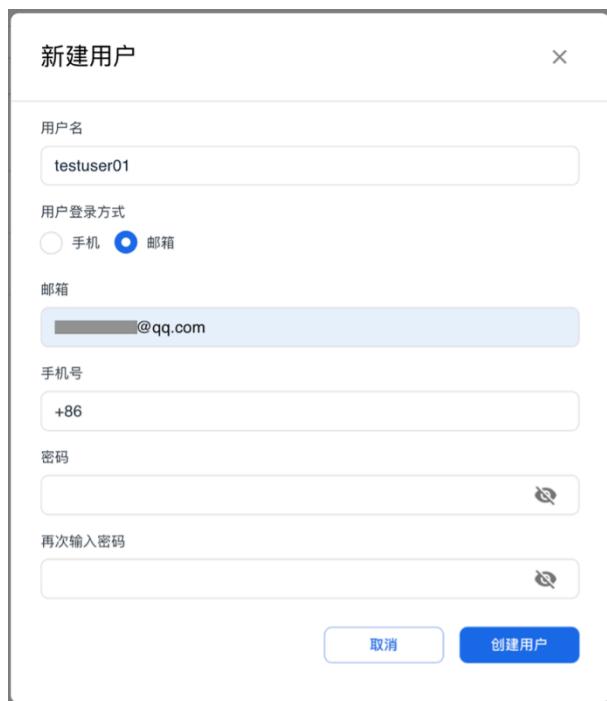
- 新建用户
- 移除用户
- 查看用户详情

- 授予/收回用户的角色

新建用户

在「**用户**」页面，新建用户的具体操作步骤如下：

- 点击**新建用户**以弹出新建用户窗口。
- 输入用户名。
- 选择新用户的注册方式并输入手机号码或者邮箱地址。当前支持使用手机或者邮箱的注册方式。



- 设置用户的登录密码，并再次输入密码以确认。
- 点击**创建用户**，新创建的用户会同步更新在用户列表中。

用户名	激活状态	激活方式	审核状态	操作
demouser01	● 已激活	自动激活	审核通过	⋮
testuser01	● 已激活	自动激活	审核通过	⋮
stageuser	● 已激活	自动激活	审核通过	⋮

共 3 条 每页 10 条 < < 1 > > 前往 页

注意：

新创建的用户会被系统自动审核通过，但仅具有 public 权限和部分功能的使用权限，可以访问的页面包括控制台的计算空间以及计算空间操作页面的数据洞察、数据库、虚拟数仓和查询历史。

此外，PieDataCS 云原生平台还支持邀请注册的方式，邀请新用户注册。详细信息请参见 [用户信息栏-邀请注册](#)。

移除用户

如果需要删除某一个用户，组织管理员可以在「**用户**」页面的用户列表中选中目标用户，然后点击隐藏菜单栏「...」并选择 **移除** 选项并确认，即可删除该用户。请注意，如果需要恢复该用户，组织管理员必须手动将其重新添加到当前计算空间。

The screenshot shows the 'User Management' page. At the top, there is a search bar labeled '输入用户名查询' and a blue button labeled '新建用户'. Below the search bar is a table with columns: '用户名' (Username), '激活状态' (Activation Status), '激活方式' (Activation Method), '审核状态' (Review Status), and '操作' (Operations). The table contains three rows of data:

用户名	激活状态	激活方式	审核状态	操作
demouser01	● 已激活	自动激活	审核通过	...
testuser01	● 已激活	自动激活	审核通过	移除
stageuser	● 已激活	自动激活	审核通过	...

At the bottom of the table, it says '共 3 条' (3 results) and has a pagination section with '每页 10 条' (10 items per page) and navigation buttons.

查看用户详情

在「**用户**」页面的用户列表中选中目标用户并点击该用户所在列，即可进入用户详情页面查看用户的详细信息，包括用户注册信息和被授予的角色信息。

The screenshot shows the 'User Detail' page for the user 'testuser01'. At the top, there is a breadcrumb '用户管理 / 用户详情' and a status bar with '中文 您好: [redacted]'. The main content area shows the user's basic information:

用户名	testuser01	用户ID	us01j8pqhcm6t9xafa9hhfppm7nw
手机号	[redacted]	邮箱	[redacted]@qq.com

Below this, a message says 'testuser01被授予了1个角色' (testuser01 has been granted 1 role). There is a search bar labeled '查询角色' and a blue button labeled '授予角色'.

The table of assigned roles is as follows:

角色名称	操作
public	收回

注意：

只有被赋予用户管理员（usermanager）角色的用户才能在「用户」页面查看用户详情。

授予/收回用户的角色

在「**用户**」页面的用户列表中选中目标用户并点击该用户所在列，即可进入用户详情页面查看用户被授予的角色信息。

注意：

只有被赋予用户管理员（usermanager）角色的用户才能进入用户详情页面并执行授予/收回用户的角色的相关操作。

如果需要将某一角色授予目标用户，参考步骤如下：

1. 在用户详情页面点击 **授予角色** 以弹出将角色授予用户的窗口。
2. 在授予的角色的下拉列表中选择合适的角色。



3. 点击 **授予**，操作成功后，所授予的角色信息会同步更新在角色列表中。

用户名	testuser01	用户ID	us01jcd0xjbxnxnytzasetcv3zvn
手机号		邮箱	████████@qq.com

testuser01被授予了2个角色

角色名称	操作
public	收回
usermanager	收回

如果需要收回授予目标用户的某一角色，参考步骤如下：

1. 在「**用户**」页面的用户列表中选中目标用户并点击该用户所在列，即可进入用户详情页面。
2. 在用户详情页面的角色列表中，找到指定的角色，在“操作”栏下点击**收回**以弹出收回授予给用户的角色的窗口。

The screenshot shows the 'User Management / User Details' page for a user named 'testuser01'. The user's details are listed: Username: testuser01, User ID: us01j8pqhcm6t9xafa9hhfppm7nw, and Email: [REDACTED]@qq.com. Below this, it says 'testuser01被授予了2个角色' (testuser01 has been granted 2 roles). A search bar labeled '查询角色' (Search Role) is present. The role list shows two entries: 'public' with an '收回' (Revoke) button and 'dbmanager' with another '收回' button, which is highlighted with a red rectangle. A blue button labeled '授予角色' (Grant Role) is located at the top right of the role list area.

3. 点击**确定**。收回角色的操作成功后，角色列表会同步更新为更改后的信息。



5.10 管理计算空间角色

在「**角色**」页面，用户可以查询系统各角色的关系和管理角色的各项权限。

在计算空间操作界面，具有相应管理权限的用户（例如用户管理员、数仓管理员和数据库管理员）点击菜单栏中的**角色**即可进入功能页面。该页面主要支持如下操作：

- 查看角色的权限信息
- 添加自定义角色
- 管理系统权限
- 继承和收回角色

- 授予角色给用户和收回授权

关于系统预设角色

PieDataCS 云原生平台的计算空间支持如下系统预设角色，系统预设角色不支持修改和删除操作。

- 普通使用者 (public)

普通使用者是最基础的角色。所有新用户在创建账户后，系统将自动赋予他们普通使用者角色，以及访问数据库和使用数据洞察等基础功能的权限。

注意:

普通使用者 (public) 作为系统角色无法继承任何角色，但默认被所有角色继承。

- 用户管理员 (usermanager)

用户管理员角色继承了普通使用者 (public) 的权限，主要负责创建和管理用户及角色，并管理系统权限。

- 数据库管理员 (dbmanager)

数据库管理员继承了普通使用者 (public) 的权限，主要负责管理数据库对象权限和计算资源。他们有权使用数据导入功能，并可以创建、更改、删除和监控虚拟数仓。

- 数仓管理员 (accountmanager)

在权限管理系统中，数仓管理员拥有最高等级的权限，不仅继承了用户管理员 (usermanager) 和数据库管理员 (dbmanager) 的权限，还负责管理包括数据库对象权限在内的所有功能。

注意:

除数仓管理员 (accountmanager) 和数据库管理员 (dbmanager) 之外，其他预设角色仅有权限查询和使用已有的虚拟数仓。

上述系统预设角色的初始化权限如下表所示。

角色名称	初始化权限
public	<ul style="list-style-type: none">· 查看虚拟数仓基本信息· 查询历史· 数据洞察· 数仓信息· 赋予/取消赋予数据库权限给角色· 数据库对象权限信息· 数据库· 公共参数
usermanager	<ul style="list-style-type: none">· 用户管理· 角色详情· 用户详情· 用户信息· 赋予/取消赋予角色给用户· 赋予/取消赋予系统权限给角色· 用户审核· 继承/取消（被）继承角色给角色· 角色信息· 角色管理
dbmanager	<ul style="list-style-type: none">· 角色信息· 虚拟数仓错误日志· 虚拟数仓巡检· 虚拟数仓外部接入能力· 查询和会话管理虚拟数仓详情· 数据导入历史数据· 虚拟数仓 SQL 审计· 数据导入功能操作· 角色详情· 虚拟数仓基础管理能力· 用户信息
accountmanager	<ul style="list-style-type: none">· 数仓管理· 用户行为审计日志

有关各项权限的详细描述，请参见 [角色的系统权限清单-计算空间的系统权限](#)。

注意：

上述表格中的系统权限同时适用于 PieDataCS 云原生平台的企业版和云上云版。然而，还有一些额外的权限仅适用于企业版，这些权限目前尚未在表格中列出。相关信息请参见 [角色的系统权限清单](#)。

查看角色的权限信息

「角色」页面以关系图和列表两种形式呈现角色的权限信息。关系图直观展示了各角色的层次结构，也是默认的角色呈现方式。越靠近角色继承的源头，就越靠近关系图底部。

在关系图中点击任意角色的卡片即可打开角色详情页面，并展示如下角色信息：

- 系统权限：系统权限页面展示当前角色已具有的系统权限，包括继承所得权限和初始权限。

在该页面，点击 **查看所有权限描述** 即可查阅全量系统权限的详细说明。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限	数据库对象权限	角色继承	用户授予
此角色(usermanager)已拥有权限			
<div style="border: 1px dashed red; padding: 5px;">继承所得权限 (8) 查询历史 虚拟数仓基本信息 数仓信息 赋予/取消赋予数据库权限给角色 数据库 数据洞察 公共参数 数据库对象权限信息</div>			
<div style="border: 1px dashed red; padding: 5px;">初始权限 (10) 用户管理 角色详情 用户详情 用户信息 赋予/取消赋予角色给用户 赋予/取消赋予系统权限给角色 用户审核 继承/取消（被）继承角色给角色 角色信息 角色管理</div>			
+ 修改/添加自有权限		查看所有权限描述	

- 数据库对象权限：数据库对象权限页面展示当前角色所具有的数据库对象权限。

在该页面，选中目标数据库对象名称并点击 **详情** 即可跳转至「数据库」页面来进一步查看数据库对象的属性信息、权限列表和该数据库下的 Schema 列表等。有关数据库页面功能的详细信息，请参见 **数据库**。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限	数据库对象权限	角色继承	用户授予
	> openpie CONNECT TEMPORARY > testdb CONNECT TEMPORARY > db3 CONNECT TEMPORARY > 线上销售数据 CONNECT TEMPORARY		

[详情]

- 角色继承：角色继承页面展示当前角色所继承的角色，或被哪些角色所继承。该页面同时支持**继承角色**和**收回角色**的操作。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限	数据库对象权限	角色继承	用户授予
		角色(usermanager)被以下角色继承	
		<div style="border: 1px solid #ccc; padding: 5px; border-radius: 10px; display: inline-block;">accountmanager</div>	
		+ 继承角色 - 收回角色	
		角色(usermanager)继承了以下角色	
		<div style="border: 1px solid #ccc; padding: 5px; border-radius: 10px; display: inline-block;">public</div>	
		+ 继承角色 - 收回角色	

- 用户授予：用户授予页面展示当前角色已授予用户的信息。该页面同时支持**将角色授予用户**和**取消对用户的角色授权**的操作。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

The screenshot shows a user interface for managing roles. At the top, there are tabs: 'System Permissions', 'Database Object Permissions', 'Role Inheritance', and 'Grant'. The 'Grant' tab is selected. Below the tabs, a section titled 'Grant the role (usermanager) to users' contains a search input field 'Enter user name' and a blue 'Grant' button. A red dashed box highlights a list titled 'Granted to users (2)' which includes 'stageuser (ID:us01j5mr204neeq2eec7pagzc0vz)' and 'demouser01 (ID:us01j8psbwsx1exsszrgk2100r5q)'. Each item in the list has a small 'X' icon to its right.

与关系图相比，**列表**是以表格的方式呈现角色信息。在列表中，用户点击目标角色名称即可展开「**角色详情**」页面，并展示各角色的系统权限、数据库对象权限，所继承的角色和被继承的角色，以及授予用户的信息。

添加自定义角色

系统不仅支持预设角色，还支持创建自定义角色来实现系统权限和数据库权限的分配。

在「**角色**」页面添加自定义角色的步骤如下：

1. 点击 **添加角色** 即可弹出创建角色窗口。
2. 输入角色名称。
3. (可选) 选择需要继承其权限的角色，可以选择继承系统预设角色或者自定义角色的权限。
4. (可选) 输入对角色的描述。
5. 点击 **创建角色**。创建成功后，自定义的角色会同步更新在角色关系图和列表中。

与预设角色不同的是，自定义角色支持修改和删除操作，关系图的角色名称卡片上提供了“编辑角色”和“删除角色”的快捷键。

同时，角色列表中自定义角色的“操作”栏下的隐藏菜单提供了“编辑角色”和“删除角色”的选项。

角色管理

关系图 列表

输入角色名查询 添加角色

角色名称	创建时间	更新时间	类型	描述	操作
public	6个月前	6个月前	系统角色		
usermanager	6个月前	6个月前	系统角色		
dbmanager	6个月前	6个月前	系统角色		
accountmanager	6个月前	6个月前	系统角色		
demorole	2小时前	2小时前	自定义角色		

共 5 条 每页 10 条 < 1 > >> 编辑角色 删除角色

注意：

对于删除角色操作，如果要删除的角色被额外授予了数据库权限，则删除角色前需要清空相关权限。

管理系统权限

在数仓操作界面的「**角色**」页面，拥有相关权限的用户（例如用户管理员）可以管理系统权限。参考步骤如下：

1. 在「**角色**」页面，选中目标角色并打开角色详情的**系统权限**页面。
2. 点击 **修改/添加自有权限**。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限 数据库对象权限 角色继承 用户授予

此角色(usermanager)已拥有权限

继承所得权限 (8)

查询历史 虚拟数仓基本信息 数仓信息 赋予/取消赋予数据库权限给角色
 数据库 数据洞察 公共参数 数据库对象权限信息

初始权限 (10)

用户管理 角色详情 用户详情 用户信息 赋予/取消赋予角色给用户
 赋予/取消赋予系统权限给角色 用户审核 继承/取消（被）继承角色给角色
 角色信息 角色管理

+ 修改/添加自有权限

查看所有权限描述

3. 在 **可授予权限区域**，选择要添加或者删除的权限。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限	数据库对象权限	角色继承	用户授予
选择要添加或者删除的权限 系统初始化赋予的权限和继承自其他角色的权限无法修改			
<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> 数据导入功能操作 × </div> <div style="border: 2px dashed red; padding: 5px; background-color: #f0f0f0;"> 可授予权限(17) <div style="display: flex; justify-content: space-around; gap: 10px;"> 虚拟数仓详情虚拟数仓基础管理能力虚拟数仓外部接入能力虚拟数仓监控 虚拟数仓巡检数据导入功能操作数据导入历史数据虚拟数仓错误日志 虚拟数仓SQL审计数仓管理定时任务-查看任务运行记录定时任务-任务管理 用户行为审计日志数据源/流/作业功能只读数据源/流/作业功能管理 脱敏信息脱敏设置管理 </div> </div> <div style="text-align: right; margin-top: 10px;"> 查看所有权限描述 </div>			
取消 保存修改			

4. 点击 **保存修改** 即可完成系统权限变更。角色详情的 **自有权限** 区域会展示该角色变更后的系统权限。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限	数据库对象权限	角色继承	用户授予
此角色(usermanager)已拥有权限			
<div style="border: 2px dashed red; padding: 5px; background-color: #f0f0f0;"> 自有权限 (1) <div style="display: flex; justify-content: space-around; gap: 10px;"> 数据导入功能操作 </div> </div> <div style="margin-top: 10px;"> 继承所得权限 (8) <div style="display: flex; justify-content: space-around; gap: 10px;"> 数仓信息赋予/取消赋予数据库权限给角色数据库数据洞察查询历史 公共参数数据库对象权限信息虚拟数仓基本信息 </div> </div> <div style="margin-top: 10px;"> 初始权限 (10) <div style="display: flex; justify-content: space-around; gap: 10px;"> 角色管理角色详情角色信息用户详情用户信息 赋予/取消赋予角色给用户赋予/取消赋予系统权限给角色用户审核 继承/取消 (被) 继承角色给角色用户管理 </div> </div>			
+ 修改/添加自有权限 查看所有权限描述			

在系统权限设置完成后，用户管理员可以授予相关用户以相应的角色。

继承和收回角色

继承角色是指将权限以角色的形式传递给另一个角色的过程。在层级结构中，继承的角色的层级需要位于被继承角色之上。

继承角色的操作包括两个方面：“当前角色被其他角色继承”以及“当前角色继承其他角色”。

注意：

public 作为系统角色无法继承任何角色，但默认被所有角色继承。而 accountmanager 角色无法被其他任何角色继承，但默认情况下会继承所有其他角色的权限。

新增角色继承当前角色

如果需要执行新增角色继承当前角色的操作，参考步骤如下：

1. 在「**角色**」页面，选中目标角色并打开角色详情的**角色继承**页面。
2. 在**角色被以下角色继承**区域中，点击**继承角色**。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限

数据库对象权限

角色继承

用户授予

角色(usermanager)被以下角色继承

accountmanager

+ 继承角色 - 收回角色

角色(usermanager)继承了以下角色

public

+ 继承角色 - 收回角色

3. 从下拉列表中选择一个可用角色以继承当前角色。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限 数据库对象权限 角色继承 用户授予

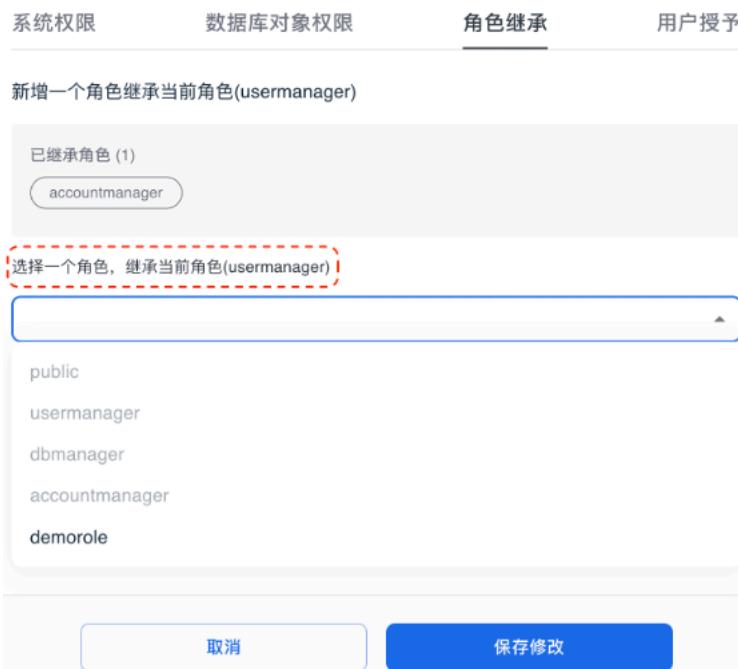
新增一个角色继承当前角色(usermanager)

已继承角色 (1)
accountmanager

选择一个角色，继承当前角色(usermanager)

public
usermanager
dbmanager
accountmanager
demorole

取消 保存修改



4. 点击 **保存修改**，操作成功后，即可新增一个角色来继承当前角色。

收回角色

如果需要执行收回角色的操作，参考步骤如下：

1. 在「**角色**」页面，选中目标角色并打开角色详情的**角色继承**页面。
2. 在**角色被以下角色继承**区域中，点击**收回角色**。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限 数据库对象权限 角色继承 用户授予

角色(usermanager)被以下角色继承

accountmanager demorole

+ 继承角色 - 收回角色

角色(usermanager)继承了以下角色

public

+ 继承角色 - 收回角色



3. 从下拉列表中选择一个可用角色以不再继承当前角色。

角色详情(usermanager)
User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限 数据库对象权限 角色继承 用户授予

使当前角色(usermanager)不再继承某个角色

被继承角色 (2)
public demorole

选择一个要取消继承的角色

public
demorole

取消 保存修改

4. 点击 **保存修改**，操作成功后，该角色将不再继承该角色。

当前角色继承其他角色

如果需要执行当前角色继承其他角色的操作，参考步骤如下：

1. 在「**角色**」页面，选中目标角色并打开角色详情的**角色继承**页面。

角色详情(usermanager)
User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限 数据库对象权限 角色继承 用户授予

角色(usermanager)被以下角色继承

accountmanager

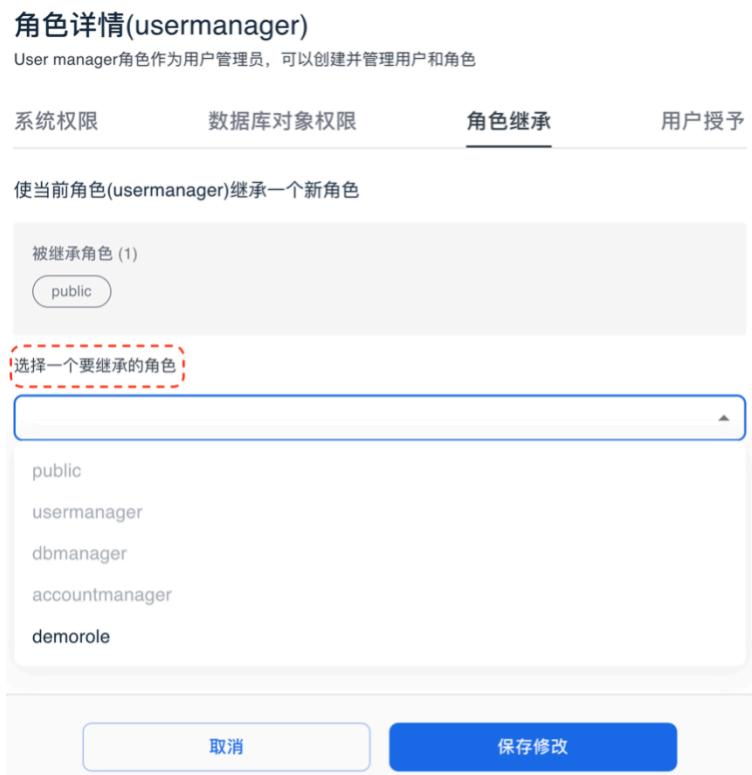
+ 继承角色 - 收回角色

角色(usermanager)继承了以下角色

public

+ 继承角色 - 收回角色

2. 在**角色继承了以下角色区域**, 点击**继承角色**。
3. 从下拉列表中选择一个要继承的角色。



4. 点击**保存修改**, 操作成功后, 当前角色就继承了所选的角色。

取消继承的角色

如果需要取消当前角色被其他角色所继承, 参考步骤如下:

1. 在**「角色」**页面, 选中目标角色并打开**角色详情**的**角色继承**页面。
2. 在**角色继承了以下角色区域**, 点击**收回角色**。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限	数据库对象权限	角色继承	用户授予
角色(usermanager)被以下角色继承			
<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> accountmanager </div>			
+ 继承角色 - 收回角色			
角色(usermanager)继承了以下角色			
<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> public demorole </div>			
+ 继承角色 - 收回角色			

- 从下拉列表中选择一个要取消继承的角色。

角色详情(usermanager)

User manager角色作为用户管理员，可以创建并管理用户和角色

系统权限	数据库对象权限	角色继承	用户授予
使当前角色(usermanager)不再继承某个角色			
<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> 被继承角色 (2) </div>			
<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> public demorole </div>			
<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> 选择一个要取消继承的角色 </div>			
<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> <div style="border: 1px solid #ccc; width: 100%; height: 30px; margin-bottom: 5px;"></div> <div style="display: flex; justify-content: space-around;"> public demorole </div> </div>			
<div style="text-align: right; margin-top: 20px;"> 取消 保存修改 </div>			

- 点击 **保存修改**，操作成功后，当前角色将不再继承所选的角色。

授予角色给用户和收回授权

授予角色给用户是将权限以角色的形式下放给用户。拥有角色管理权限的角色可以将角色及其权限赋予给指定用户。

注意：

public 角色不支持此操作，因为 public 已被默认授予给所有用户。

授予角色给用户

如果需要执行授予角色给用户的操作，参考步骤如下：

1. 在「**角色**」页面，选中目标角色并打开角色详情的**用户授予**页面。
2. 在**将角色授予用户的**下拉列表中选择一个可用的用户。



3. 点击**授予**，操作成功后，当前角色即可授予给所选用户并同步更新到已授予用户的列表中。

取消对用户的角色授权

如果需要取消对用户的角色授权，参考步骤如下：

1. 在「**角色**」页面，选中目标角色并打开角色详情的**用户授予**页面。
2. 在已授予用户列表中，选中目标用户并点击**X**。



3. 在弹出的窗口中，点击 **确认收回** 即可取消对目标用户的角色授权。



提示：

用户被取消某一角色授权后可以再次被授予相同的角色。

6 计算空间高级功能

6.1 管理定时任务

PieDataCS 云原生平台支持在预定的时间点或时间间隔内自动执行 SQL 任务，适用于定期执行后台任务、生成报告、清理垃圾数据、更新数据和执行自动化操作等运维或者开发场景。

提示：

仅 PieDataCS 云原生平台的企业版支持管理定时任务的功能。

查看任务运行记录

在计算空间操作界面，具有相应管理权限的用户（例如数据库管理员）点击菜单栏中的**定时任务**就会默认进入任务运行记录页面。

该页面以列表形式显示所有定时任务的信息。用户可以输入“所属任务 ID”或者选择任务状态（包括运行中、成功、失败、等待调度、取消调度和停止）来筛选指定的任务信息。

记录ID	所属任务ID	所属任务名称	状态	结果	触发时间	结束时间	操作
ji01j14n1bt9h4ftvhpz067swtlyw	jb01j14jhq0abhx708512as8km80	自动刷新视图	● 等待调度				取消
ji01j14m27q755vt2y1b9jqhf3ry	jb01j14jhq0abhx708512as8km80	自动刷新视图	● 取消调度		2024年6月24日星期一 15:59		
ji01j14m0d4827aa0w4bhvatpnz2	jb01j14jhq0abhx708512as8km80	自动刷新视图	● 成功	ok	2024年6月24日星期一 15:58	2024年6月24日星期一 15:58	
ji01j14kyjhre26pdbr3spfhpwp6e	jb01j14jhq0abhx708512as8km80	自动刷新视图	● 成功	ok	2024年6月24日星期一 15:57	2024年6月24日星期一 15:57	

创建定时任务

在「**定时任务**」页面，创建定时任务的步骤如下：

1. 点击**任务管理**以切换到任务管理页面。
2. 点击**创建任务**，进入创建任务页面。

任务ID	任务名称	描述	下次触发时间	状态	创建时间	操作
暂无任务						

3. 分别输入如下定时任务信息：

- 任务名称：自定义的任务名称。
 - （选填）任务执行超时时间：如果超过该阈值，则任务会终止，并等待下一个周期再重新执行。
 - 最大同时运行任务数：当前为固定值“1”。
 - 任务执行周期：输入 cron 表达式，指定任务的执行周期。例如“*/1 * * * *”表示每分钟一次。
 - （选填）任务描述：对该定时任务的自定义描述。
4. 选择创建后是否立即启用该任务。如果没有勾选，则需要在创建定时任务完成后，在任务管理列表手动启用。
5. 选择运行该任务的虚拟数仓和数据库。
6. 输入执行该定时任务的 SQL 语句。

定时任务 / 创建任务

任务名称

自动刷新视图

任务执行超时时间 (选填)

10

秒

最大同时运行任务数

1

任务执行周期

*/1 * * * *



任务描述 (选填)

自动刷新视图, 每分钟一次

 创建后立即启用

虚拟数仓

vh1 (运行中)

数据库

openpie

定时任务 SQL

1 REFRESH MATERIALIZED VIEW mymatview;

7. 点击 **完成**, 定时任务创建成功后会自动同步到任务管理列表中。

任务运行记录	任务管理														
	<div style="display: flex; justify-content: space-between;">C创建任务</div> <table border="1"><thead><tr><th>任务ID</th><th>任务名称</th><th>描述</th><th>下次触发时间</th><th>状态</th><th>创建时间</th><th>操作</th></tr></thead><tbody><tr><td>jb01j4jhq0abhx708512as8km80</td><td>自动刷新视图</td><td>自动刷新视图, 每分钟一次</td><td>2024年6月24日星期一 16:22</td><td>已启用</td><td>2024年6月24日星期一 15:32</td><td>修改 禁用 删除</td></tr></tbody></table>	任务ID	任务名称	描述	下次触发时间	状态	创建时间	操作	jb01j4jhq0abhx708512as8km80	自动刷新视图	自动刷新视图, 每分钟一次	2024年6月24日星期一 16:22	已启用	2024年6月24日星期一 15:32	修改 禁用 删除
任务ID	任务名称	描述	下次触发时间	状态	创建时间	操作									
jb01j4jhq0abhx708512as8km80	自动刷新视图	自动刷新视图, 每分钟一次	2024年6月24日星期一 16:22	已启用	2024年6月24日星期一 15:32	修改 禁用 删除									

在任务管理列表中的操作栏下提供 **修改**、**禁用** 或者 **删除** 选项, 以方便用户管理目标定时任务。如果定时任务处于未启用或者禁用状态, 用户可以点击 **启用** 运行目标定时任务。

6.2 审计日志

PieDataCS 云原生平台支持对数仓用户的行为进行审计, 并统计相关的 API 调用信息, 以帮助用户排查和解决问题。

提示:

仅 PieDataCS 云原生平台的企业版支持审计日志的功能。

在数仓操作界面，具有相应管理权限的用户（例如数仓管理员）点击菜单栏中的 **审计日志** 即可进入用户行为审计页面。该页面以列表形式展示当前所有用户的审计日志。

审计日志							
开始时间		结束时间		请输入方法名	请输入 API 地址	请输入请求参数搜索关键词	C
API 地址	方法名	请求参数	状态码	响应消息	请求时间	客户端 IP	用户代理
/svc/auth/login	login	{ "accType": "2", "loginName": "1449723..."}	200		2024-06-24 14:59:51	██████████	Mozilla/5.0 (Macintosh; ...)
/svc/auth/login	login	{ "accType": "2", "loginName": "1449723..."}	200		2024-06-21 10:23:07	██████████	Mozilla/5.0 (Macintosh; ...)
/svc/auth/login	login	{ "accType": "2", "loginName": "1449723..."}	200		2024-06-20 17:07:45	██████████	Mozilla/5.0 (Macintosh; ...)
/svc/auth/login	login	{ "accType": "2", "loginName": "1449723..."}	200		2024-06-20 17:07:03	██████████	Mozilla/5.0 (Macintosh; ...)

审计日志的列表主要展示如下字段信息：

- **API 地址**: 所调用的 API 地址的资源路径。
- **方法名**: 所调用的函数名称。
- **请求参数**: 请求所涉及的配置或指令。
- **状态码**: 该请求响应的状态码。
- **响应消息**: 该请求对应的响应消息。
- **请求时间**: 用户登录数仓的时间。
- **客户端 IP**: 登录用户所使用的云原生平台的 IP 地址。
- **用户代理**: 用户代理自身的信息，例如操作系统类型、浏览器版本、语言偏好等信息。

审计日志的列表提供如下快捷操作：

- **指定时间范围筛选信息**: 使用**开始时间-结束时间**控件可以筛选出某个时间范围（单位为天）内的日志信息。
- **根据方法名称搜索信息**: 使用**请输入方法名**搜索框可以通过所调用的函数名称来搜索相关信息。
- **根据 API 地址搜索信息**: 使用**请输入 API 地址**搜索框可以通过所调用的 API 地址来搜索相关信息。
- **根据请求参数搜索信息**: 使用**请输入请求参数搜索关键词**搜索框可以通过请求所涉及的配置或指令来搜索相关信息。
- **立即刷新**: 点击立即刷新图标可以即刻同步当前的列表信息为最新。

6.3 数据脱敏

PieCloudDB 支持一键启用数据脱敏扩展功能。该功能通过在数据库层级创建脱敏 Schema，并根据脱敏规则对脱敏 Schema 下指定表中的敏感数据进行脱敏处理，可以降低脱敏用户在执行数据查询时发生数据和隐私泄露的风险。

提示：

仅 PieDataCS 云原生平台的企业版支持数据脱敏的功能。

在数仓操作界面，用户点击菜单栏的「**数据库**」后，在数据库列表中选择目标数据库并点击目标数据库名称就会进入数据库详情页面，之后切换到**Extension**页面，即可使用数据脱敏的相关功能。

数据脱敏扩展功能的一般使用流程如下：

1. 指定数据库下开启数据脱敏
2. 切换脱敏 Schema
3. 添加表的脱敏规则
4. 添加脱敏用户
5. 卸载脱敏扩展

使用说明

在目标数据库启用脱敏扩展后，该数据库下会自动创建两个 Schema：anon 和 mask。其中，anon 用于存储脱敏规则函数和假数据；mask 中的视图用来存储经过匿名化处理或掩码处理后的脱敏数据，实现与原始数据分开存储和管理。

同时，用户在该数据库下的 `search_path` 会指向 mask Schema，在该数据库下指定脱敏 Schema（默认为 public）内创建的表，也会映射到 mask Schema 下。例如，设置脱敏 Schema（例如 s1）后，用户连接该数据库并对指定的脱敏表（例如 mt1）进行查询，如果 SQL 语句中不指定 Schema（例如 `SELECT * FROM mt1;`），则查询结果会自动指向 mask Schema 下的表 mt1；如果指定脱敏 Schema（例如 `SELECT * FROM s1.mt1;`），则会报无权限的错误。

启用数据脱敏

启用数据脱敏扩展功能的操作步骤如下：

- 在数仓操作界面点击菜单栏「数据库」即可进入功能页面，在数据库列表中选择目标数据库。

提示：

在「数据洞察」页面可以创建数据库、Schema 和表，以及准备原始数据和执行查询。在「快速导入」页面可以批量导入数据。

- 点击数据库名称展示数据库详情后，切换到**Extension**页面。
- 在**脱敏扩展**区域，点击**启用>确定**即可启用数据脱敏功能，同时脱敏扩展的**当前状态**由“未启用”变为“已启用”。

扩展名	命名空间	版本号	所有者	是否可以重新定位
plpgsql	pg_catalog	1.0	openpie	false
gp_exttable_fdw	pg_catalog	1.0	openpie	true

数据脱敏功能启用后，该数据库会自动创建两个 Schema：anon 和 mask。

- 点击**查看详情**即可查看脱敏扩展详情。

脱敏扩展详情的相关字段含义如下：

- 名称：扩展的名称。
- 所有者：扩展的所有者。

- 所属命名空间：包含扩展导出对象的 Schema。
- 版本号：扩展的版本号。
- 是否可以重新定位：如果扩展可以重新定位到另一个 Schema，则为 True。
- 配置参数列表：扩展配置的参数列表，如果没有则为空（NULL）。
- 条件表达式：扩展配置表的 WHERE 子句过滤条件，如果没有则为空（NULL）。

切换脱敏 Schema

启用脱敏扩展后，系统会默认设置 public 作为脱敏 Schema。但由于 public Schema 的权限控制比较特殊，所以建议用户在启用脱敏扩展后立即切换脱敏 Schema，不要使用 public。

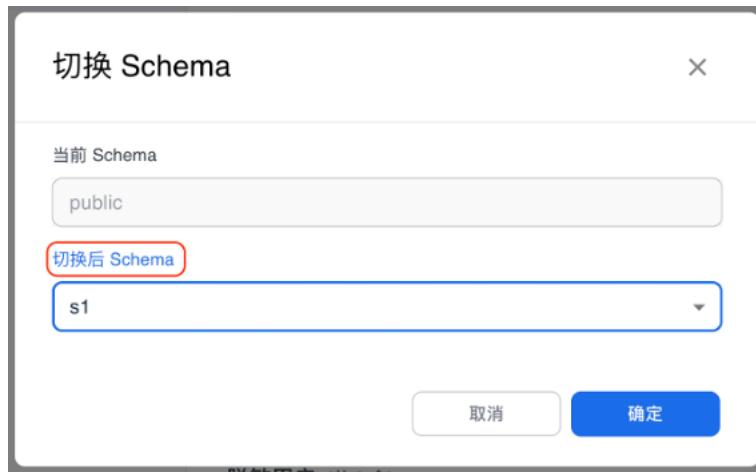
1. 启用数据脱敏扩展后，点击 **查看详情**，在扩展详情页面的**脱敏信息**区域会显示当前指定的脱敏 Schema 信息。

脱敏信息 ② (当前脱敏 Schema: public)

表名 脱敏字段 操作

暂无数据

2. 点击 **切换 Schema** 即可弹出切换 Schema 的窗口。
3. 在下拉列表中选择要切换的 Schema。



注意：

切换脱敏 Schema 的操作不会清空原 Schema 下表的所有已设置的脱敏规则。

4. 点击 **确定** 即可切换到新的 Schema。

添加表的脱敏规则

设置脱敏 Schema 后，用户可以在该 Schema 下设置表的脱敏规则。

1. 在扩展详情页面的**脱敏信息**区域，点击**新建脱敏**即可弹出添加脱敏的窗口。
2. 在下拉列表中选择需要脱敏的表，并添加一个或者分别添加多个脱敏字段和脱敏规则，之后点击**添加**即可。



注意:

当前版本只支持为 TEXT 数据类型的字段添加脱敏规则。

当前版本预置了 TEXT 数据类型的脱敏规则，包括如下：

- 全置*: 所有文本都置*。
- 保留m+n字符，即保留前m位、末n位字符，中间所有字符都置*。当前支持的组合包括6+4、3+4和6+3。
- 邮箱地址遮盖：遮盖邮箱名称和地址的关键字符。

提示:

如果用户需要配置其他脱敏规则可以联系 OpenPie。

3. 点击**确定并提交**即可将所添加的脱敏信息同步到列表中。

表名	脱敏字段	操作
mt1	c1	管理脱敏字段 脱敏预览 删除

在脱敏信息列表中，对已添加的脱敏信息支持如下操作：

- 管理脱敏字段：点击 **管理脱敏字段**，可以再次添加脱敏字段并设置脱敏规则，或者删除已添加的脱敏字段及其脱敏规则。

管理mt1的脱敏字段

×

添加脱敏字段

请选择脱敏字段
请选择脱敏规则
添加

已脱敏字段(共1个)

c1 全置* (将字段全置为*符号) — 删除

- 脱敏数据预览：点击 **脱敏预览**，选择可用的虚拟数仓后即可预览脱敏后的数据，用于查验是否符合预期。

mt1脱敏预览

●
vm1 (运行中)
▼

	c1	
1	*****	
2	*****	

- 删除脱敏规则：点击 **删除>确定**，即可删除指定的脱敏字段及其脱敏规则。

添加脱敏用户

每个数据库下的脱敏 Schema 都可以指定自己的脱敏用户。

1. 在扩展详情页面的**脱敏用户**区域，点击**管理用户**即可弹出管理脱敏用户的窗口。

2. 在下拉列表中选择需要脱敏的一个或者多个用户，点击 **添加** 即可。

管理脱敏用户

×
添加脱敏用户

×
添加

已添加的脱敏用户(共0个)

3. 点击 **X** 可以返回扩展详情页面，新添加的用户信息也会同步显示在**脱敏用户**区域。
 4. 使用脱敏用户重新登录 PieDataCS 云原生平台，在指定数据库的脱敏 Schema 下执行 SQL 查询，系统就会根据所设置的脱敏规则返回查询结果。下图为脱敏后的得到的查询结果的示例。

数据洞察
中文 您好:

20240829-104505
+

保持会话

```
1   select * from mt1;
```

● 执行成功 请求ID: [qu01j6e0bp0nc8g1crp56vtd6z2s#0](#) 扫描时间: 20ms
 执行时间: 266ms 查询记录数: 2 结果集大小: 21 Bytes

		c1
1	*****	*****
2	*****	*****

脱敏用户的相关注意事项如下：

- 当用户被指定为脱敏用户后，系统会自动撤销其在指定脱敏 Schema 下的 CREATE 和 USAGE 权限。这意味着这些用户将无法在原 Schema 中执行创建表、查询数据、更新数据等操作。因此，管理员在设置脱敏用户名单时应谨慎行事，建议不要将 Schema 的 Owner 设置为脱敏用户。对于未被指定为脱敏用户的用户，他们查询数据库和 Schema 的结果不会受到影响。

- 对于在指定数据库下设置的脱敏用户，他们对映射的 mask Schema 下的所有视图都具有查看权限，但查看的结果将是脱敏后的数据。如果脱敏 Schema 下的表未设置脱敏规则，脱敏用户将能看到完整的表信息。因此，管理员需要谨慎设计脱敏 Schema 及其包含的表，并合理配置脱敏用户。
- 当用户被授予某个角色，且该角色对脱敏 Schema 有相关的数据库对象权限时，用户将继承角色的权限，从而对脱敏 Schema 具有相应的权限。因此，在设置脱敏用户时，如果该用户被授予了某个角色，且该角色具有脱敏 Schema 及其下所有表的权限，建议管理员在操作前先撤销用户所继承的角色，或直接撤销角色的数据库对象权限。

卸载脱敏扩展

在指定数据库的**Extension**页面的**脱敏扩展**区域，点击**卸载>确定**即可卸载数据脱敏功能。

扩展名	命名空间	版本号	所有者	是否可以重新定位
plpgsql	pg_catalog	1.0	openpie	false
gp_exttable_fdw	pg_catalog	1.0	openpie	true

卸载完成后，脱敏扩展的**当前状态**由“已启用”变为“未启用”。需要注意的是，该数据库在启用脱敏扩展时自动创建的 Schema anon 和 mask 也会被自动删除，之前创建的脱敏规则会保留但不再生效。

6.4 外部数据包装器（FDW）扩展

6.4.1 外部数据包装器（FDW）扩展使用说明

PieCloudDB 支持使用外部数据包装器（FDW）扩展 postgres_fdw、mysql_fdw 和 oracle_fdw 来实现访问外部数据源数据的功能。

提示：

仅 PieDataCS 云原生平台的企业版支持外部数据包装器（FDW）扩展的功能。

在数仓操作界面，用户点击菜单栏的「**数据库**」后，在数据库列表中选择目标数据库并点击目标数据库名称就会进入数据库详情页面，之后切换到**Extension**页面，即可使用 FDW 扩展的相关功能。

FDW 扩展功能具有严格的使用权限限制，基于用户权限的管控层面，FDW 扩展的使用流程大致如下：

- 启用和配置 FDW 扩展
- 访问外表数据并执行相关操作
- 卸载 FDW 扩展

启用和配置 FDW 扩展功能

针对具有较高权限的用户，通常是管理员，请优先执行如下配置操作：

1. 启用 FDW 扩展

在「**数据库**」页面的目标数据库下的 Extension 列表中，点击 **启用扩展**，并在下拉列表中选择所需要的扩展，点击 **完成** 即可添加扩展。

2. 创建外部服务器

外部服务器定义了连接到外部数据源的方式。它是访问外部数据的入口，使得本地数据源能够通过 SQL 查询的方式访问远程数据。

在「**数据库**」页面的目标数据库下的 Extension 列表中，点击目标扩展名称以进入扩展详情页面，在点击 **创建 Server** 即可进入外部服务器的配置页面。

3. 配置用户映射

配置远程用户映射到本地用户，配置的远程用户和本地用户是一一对应的。

在「**数据库**」页面的目标数据库下的 Extension 列表中，点击目标扩展名称以进入扩展详情页面。在关联 Server 列表中，选择目标外部服务器并点击隐藏菜单栏「...」下的 **用户映射** 选项即可进入用户映射的配置页面。

访问外表数据并执行相关操作

完成启用 FDW、创建外部服务器和配置用户映射的操作后，用户就可以执行外部表相关操作了，例如 *SELECT*、*UPDATE*、*DELETE*、*INSERT* 等。

注意：

具有执行外部表关联查询等操作权限的用户是由所配置的用户映射中的“映射本地用户”指定的，而且该用户需要具有外部表的相关操作权限。

本文以执行外部表的关联查询为例，介绍如何在数仓操作界面访问外部表的数据。

1. 在「**数据洞察**」页面打开一个 SQL 文件，连接目标数据库（即需要执行关联查询的数据库）和虚拟数仓。
2. 创建外部表。在「**数据洞察**」页面的 SQL 文件中，使用 *CREATE FOREIGN TABLE* 命令创建一个外部表，用于访问远程的表。

提示：

建议外部表的列与被引用的远程表相匹配。

3. 访问外部表的数据。在「**数据洞察**」页面的 SQL 文件中，使用 *SELECT* 命令访问远程表中的数据或者执行其他相关 DML 操作。

卸载 FDW 扩展

具有相应权限的用户，通常是管理员，在「**数据库**」页面的目标数据库下的“Extension”列表中找到需要卸载的扩展。在该扩展旁边的隐藏菜单栏「...」下点击 **卸载**，即可进行扩展的卸载操作。需要注意的是，在卸载扩展之前，请先确保清除了所有相关依赖，否则可能会遇到错误，导致卸载失败。

6.4.2 postgres_fdw

PieCloudDB 支持使用 postgres_fdw 扩展来直接访问存储在外部数据库服务器中的数据，包括 PieCloudDB、PieCloudDB-TP 和 PostgreSQL 数据源。

postgres_fdw 主要支持如下外部表相关的使用场景：

- 直接访问 PieCloudDB 数据源的数据并执行数据查询（*SELECT*）

- 直接访问 PieCloudDB-TP 数据源的数据并执行数据查询 (SELECT)
- 直接访问 PostgreSQL 数据源的数据并执行增删改查操作 (SELECT/UPDATE/DELETE/INSERT)

注意事项

针对外部数据源 PieCloudDB 和 PieCloudDB-TP, `postgres_fdw` 仅支持 SELECT 操作。

配置项说明

外部服务器

- Server 名称：外部服务器的名称。

外部服务器的名称应遵循数据库的标准命名方式，并保持一定的规律性，以便于用户进行管理和识别。建议命名只包含字母、数字和下划线，并且不要以数字开头。如果命名中包含空格和特殊字符，或者需要区分大小写，建议在创建外部表时，使用双引号将所关联的外部服务器名称括起来。

- 主机：外部数据源服务器的主机 IP 地址。
- 端口：外部数据源服务器的端口号。
- (可选) Options 参数：关联外部数据源服务器的 Options 参数。
- 连接数据库名称：关联外部数据源服务器的数据库名称。
- (可选) 将 Server 的 Usage 权限授予给指定用户：将外部数据源服务器的使用权限授予指定用户。默认情况下，该权限会授予执行该操作的当前用户。

用户映射

- 远程用户身份信息：远程用户名和密码。如果所连接的远程数据源为 PieCloudDB 虚拟数仓，则密码为该虚拟数仓的 token 密码。
- 映射本地用户：指定本地执行外部表关联查询等操作的用户。需要注意的是，当指定所映射的本地用户为 PUBLIC 时，会创建一个公共映射，这个映射适用于所有没有指定远程用户映射的用户。

外部表

对于 `postgres_fdw`, 使用 `CREATE FOREIGN TABLE` 命令创建外部表时必须指定外部数据源表所在的 Schema 名称和表名。

```
CREATE FOREIGN TABLE foreign_table_name (
    column_name1 data_type1,
    column_name2 data_type2,
    ...
) SERVER server_name OPTIONS (schema_name 'remote_schema', table_name
    'remote_table');
```

使用 postgres_fdw 访问 PieCloudDB 数据的示例

本文通过具体示例介绍在如下场景中使用 postgres_fdw 访问 PieCloudDB 数据：

- 跨 PieCloudDB 数仓执行关联查询
- 同一个 PieCloudDB 数仓下执行跨数据库的关联查询

示例 1：跨数仓执行关联查询

如下示例为，数仓 account1 具有数据库 db1 和虚拟数仓 vw1，数仓 account2 具有数据库 db2 和虚拟数仓 vw2，将数仓 account1 下 db1 的表 t1 分享到数仓 account2 下的 db2 中。

此外，请确保数仓 account1 下已创建示例数据库 db1 和用于分享数据的表 t1，并置于 db1 的 sch1 Schema 下。相关 SQL 语句示例如下：

```
CREATE DATABASE db1;
-- 连接数据库 db1 和虚拟数仓 vw1
CREATE SCHEMA sch1;
CREATE TABLE sch1.t1 (id INT, name VARCHAR(16), age INT);
INSERT INTO sch1.t1 (id, name, age) VALUES (1, 'John', 25),(2, 'Jane', 30),(3, 'Mike', 35);
```

1. 登录数仓 account2 下的数仓操作界面点击菜单栏「**数据库**」，在数据库 db2 下启用和配置 FDW 扩展功能。

1. 启用 FDW 扩展。

在数据库 db2 下的 Extension 列表中，点击 **启用扩展** 并在下拉列表中选择“postgres_fdw”，再点击 **完成** 即可添加 postgres_fdw 扩展到 Extension 列表中。

扩展名	命名空间	版本号	所有者	是否可以重新定位
plpgsql	pg_catalog	1.0	openpie	false
gp_exttable_fdw	pg_catalog	1.0	openpie	true
postgres_fdw	public	1.0	openpie	true
⋮				

此时，系统也会提示“启用成功”。

2. 创建外部服务器对象。

在 Extension 列表中，点击扩展名称 postgres_fdw 以进入扩展详情页面，再点击 **创建 Server** 即可进入外部服务器的配置页面。

数据库 / db2 / extensions / postgres_fdw

Extension 详情

名称 postgres_fdw	所有者 openpie	所属命名空间 public
版本号 1.0	是否可以重新定位 true	配置参数列表
条件表达式		

关联 Server

创建 Server

Server 名称	授权可用的用户	配置信息
-----------	---------	------

输入如下外部服务器配置信息：

- Server 名称：外部服务器的名称。本示例使用 fdw_server_a1db1。
- 主机：数仓 account1 中虚拟数仓 vw1 的 Host IP。在虚拟数仓外部接入信息页面，点击 **接入详情** 获取 Host IP。
- 端口：数仓 account1 中虚拟数仓 vw1 的端口号。在虚拟数仓外部接入信息点击 **接入详情** 获取端口号。
- （可选）Options 参数：如果连接的外部数据源为 PieCloudDB 虚拟数仓且使用 PieProxy 外部接入方式，则须填入外部数据源的虚拟数仓 ID 信息，格式为 “-c warehouse=<虚拟数仓ID>”；如果使用 Node Port 接入方式，则不需要填写虚拟数仓 ID 信息。
- 连接数据库名称：要访问的数仓 account1 下的数据库名称。该示例为 db1。
- （可选）将 Server 的 Usage 权限授予给指定用户：选择并指定其他需要外部数据源服务器使用权限的用户。该示例为 testUser。

3. 配置用户映射。

在扩展详情页面的关联 Server 列表中，选择目标外部服务器并点击隐藏菜单栏「...」下的 **用户映射** 选项即可进入用户映射的配置页面。

Extension 详情

名称 postgres_fdw

所有者 openpie

所属命名空间 public

版本号 1.0

是否可以重新定位 true

配置参数列表

条件表达式

关联 Server



创建 Server

Server 名称

授权可用的用户

配置信息

fdw_server_a1db1

{testUser=U/openpie}

{host=[REDACTED] port=[REDACTED] dbname=db1}

用户映射

修改

删除

选项配置说明如下：

- 远程用户名：可以连接外部数据源的用户。
- 密码：要执行关联查询的虚拟数仓的 token 密码。在虚拟数仓外部接入信息页面，依次点击 **接入详情>忘记了Token密码>重置Token密码** 获取 token 密码，如下图所示。



- 映射本地用户：指定本地执行外部表关联查询等操作的用户。管理员还可以将所映射的本地用户指定为PUBLIC，这样可以创建一个公共映射。这个映射适用于所有未指定远程用户映射的用户。



2. 以用户映射所指定的本地用户身份登录数仓 account2 下的「数据洞察」页面并打开一个 SQL 文件，连接数据库 db2 和虚拟数仓 vw2，以执行外部表关联查询。

1. 使用 `CREATE FOREIGN TABLE` 创建一个外部表，并按需创建 Schema 用以区分外部表的存放位置。

```
CREATE SCHEMA foreign_sch;
CREATE FOREIGN TABLE foreign_sch.t1 (id INT, name VARCHAR(16), age INT)
    SERVER fdw_server_a1db1 options (schema_name 'sch1', table_name 't1');
```

选项配置说明如下：

- `server_name`: 使用所创建的外部服务器对象的名称。该示例为 `fdw_server_a1db1`。
- `schema_name`: 要访问的外部表所属 Schema 名称。该示例为 `sch1`。
- `table_name`: 要访问的外部表的实际表名。该示例为 `t1`。

2. 创建表 t2 用于关联查询。

```
CREATE TABLE t2 (id INT, name VARCHAR(16));
INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
```

3. 访问外表数据并执行关联查询。

```
SELECT t1.id, t1.name, t1.age FROM foreign_sch.t1
    AS t1 INNER JOIN t2 ON t1.name=t2.name;
```

查询结果如下图所示。

The screenshot shows the πDataCS Data Observatory interface. At the top, it says "数据洞察" and "跨数仓执行关联查询". Below that, there are dropdown menus for "db2" and "vw2 (运行中)", a "保持会话" checkbox, and three execution buttons: "执行选中", "全部执行", and "执行计划". The SQL code is as follows:

```

1 CREATE SCHEMA foreign_sch;
2
3 v CREATE FOREIGN TABLE foreign_sch.t1 (id INT, name VARCHAR(16), age INT)
4           SERVER fdw_server_a1db1 options (schema_name 'sch1', table_name 't1');
5
6 DROP TABLE t2;
7 CREATE TABLE t2 (id INT, name VARCHAR(16));
8 INSERT INTO t2 (id, name) VALUES (1, 'John'), (2, 'James'), (3, 'Mike');
9
10 v SELECT t1.id, t1.name, t1.age FROM foreign_sch.t1
11      AS t1 INNER JOIN t2 ON t1.name=t2.name;

```

Execution results summary: 执行成功, 请求ID: qu01j6h36ckn59h2e8s6fmxebh4d#0, 扫描时间: 20ms, 执行时间: 926ms, 查询记录数: 2, 结果集大小: 32 Bytes.

	id	name	age	
1	1	John	25	
2	3	Mike	35	

示例 2: 同一个数仓下执行跨数据库的关联查询

如下示例为，数仓 account2 具有数据库 db2 和 db3，以及虚拟数仓 vw2，将数仓 account2 下 db3 的表 t3 分享到同一数仓下的 db2 中。

此外，请确保数仓 account2 下已创建示例数据库 db3 和用于分享数据的表 t3，并置于 db3 的 sch3 Schema 下。相关 SQL 语句示例如下：

```

CREATE DATABASE db3;
-- 连接数据库 db3 和虚拟数仓 vw2
CREATE SCHEMA sch3;
CREATE TABLE sch3.t3 (id INT, name VARCHAR(16), age INT);
INSERT INTO sch3.t3 (id, name, age) VALUES (1, 'John', 25),(2, 'Jane', 30),(3, 'Mike', 35);

```

1. 登录数仓 account2 下的数仓操作界面点击菜单栏「数据库」，在数据库 db2 下启用和配置 FDW 扩展功能。
1. 启用 FDW 扩展。

在数据库 db2 下的 Extension 列表中，点击 **启用扩展** 并在下拉列表中选择 “postgres_fdw”，再点击 **完成** 即可添加 postgres_fdw 扩展到 Extension 列表中。

The screenshot shows the 'Extensions' tab selected in the top navigation bar. Below it is a table titled 'Extension 列表' (Extension List) with columns: 扩展名 (Extension Name), 命名空间 (Namespace), 版本号 (Version), 所有者 (Owner), and 是否可以重新定位 (Can be repositioned). The 'postgres_fdw' row is highlighted with a red dashed box. A blue button labeled '启用扩展' (Enable Extension) is visible at the top right of the table area.

扩展名	命名空间	版本号	所有者	是否可以重新定位
plpgsql	pg_catalog	1.0	openpie	false
gp_exttable_fdw	pg_catalog	1.0	openpie	true
postgres_fdw	public	1.0	openpie	true
				:

此时，系统也会提示“启用成功”。

2. 创建外部服务器对象。

在 Extension 列表中，点击扩展名称 postgres_fdw 以进入扩展详情页面，再点击 **创建 Server** 即可进入外部服务器的配置页面。

The screenshot shows the 'Extension 详情' (Extension Details) page for the 'postgres_fdw' extension. It displays basic information: 名称 (Name) - postgres_fdw, 所有者 (Owner) - openpie, 所属命名空间 (Namespace) - public, 版本号 (Version) - 1.0, 是否可以重新定位 (Can be repositioned) - true, and 条件表达式 (Conditional Expression). Below this, there's a section titled '关联 Server' (Associate Server) with a blue button labeled '创建 Server' (Create Server) which is highlighted with a red box.

名称 postgres_fdw	所有者 openpie	所属命名空间 public
版本号 1.0	是否可以重新定位 true	配置参数列表
条件表达式		

输入如下外部服务器配置信息：

- Server 名称：外部服务器的名称。本示例使用 fdw_server_a2db3。
- 主机：数仓 account2 中虚拟数仓 vw2 的 Host IP。在虚拟数仓外部接入信息页面，点击 **接入详情** 可以获取 Host IP。

- 端口：数仓 account2 中虚拟数仓 vw2 的端口号。在虚拟数仓外部接入信息页面，点击 **接入详情** 获取端口号。
- （可选）Options 参数：如果连接的外部数据源为 PieCloudDB 虚拟数仓且使用 PieProxy 外部接入方式，则须填入外部数据源的虚拟数仓 ID 信息，格式为 “-c warehouse=<虚拟数仓ID>”；如果使用 Node Port 接入方式，则不需要填写虚拟数仓 ID 信息。
- 连接数据库名称：要访问的数仓 account2 下的数据库名称。该示例为 db3。
- （可选）将 Server 的 Usage 权限授予给指定用户：选择并指定其他需要外部数据源服务器使用权限的用户。该示例为 testUser。

3. 配置用户映射。

在扩展详情页面的关联 Server 列表中，选择目标外部服务器并点击隐藏菜单栏「...」下的 **用户映射** 选项即可进入用户映射的配置页面。

The screenshot shows the 'Extension Detail' page for the 'postgres_fdw' extension. At the top, there are tabs for 'Database / db2 / extensions / postgres_fdw'. On the right, there are language and user settings. Below the tabs, there's a section for 'Extension Detail' with fields: 'Name' (postgres_fdw), 'Owner' (openpie), 'Namespace' (public), 'Version' (1.0), 'Can be repositioned' (true), and 'Condition expression'. Under 'Associated Servers', there's a table with one row: 'fdw_server_a2db3' (Server Name), '{testUser=U/openpie}' (Granted users), and a configuration string '{host=[REDACTED],port=[REDACTED],dbname=db3}' (Configuration information). To the right of this table is a context menu with options: 'User Mapping' (highlighted with a red box), 'Modify', and 'Delete'.

选项配置说明如下：

- 远程用户名：可以连接外部数据源的用户。
- 密码：要执行关联查询的虚拟数仓的 token 密码。在虚拟数仓外部接入信息页面，依次点击 **接入详情>忘记了Token密码>重置Token密码** 获取 token 密码，如下图所示。



- 映射本地用户：指定本地执行外部表关联查询等操作的用户。管理员还可以将所映射的本地用户指定为PUBLIC，这样可以创建一个公共映射。这个映射适用于所有未指定远程用户映射的用户。



- 以用户映射所指定的本地用户身份登录数仓 account2 下的「数据洞察」页面并打开一个 SQL 文件，连接数据库 db2 和虚拟数仓 vw2，以执行外部表关联查询。

1. 使用 `CREATE FOREIGN TABLE` 创建一个外部表，并按需创建 Schema 用以区分外部表的存放位置。

```
CREATE SCHEMA foreign_sch;  
CREATE FOREIGN TABLE foreign_sch.t3 (id INT, name VARCHAR(16), age INT)  
    SERVER fdw_server_a2db3 options (schema_name 'sch3', table_name 't3');
```

选项配置说明如下：

- `server_name`: 使用所创建的外部服务器对象的名称。该示例为 `fdw_server_a2db3`。
- `schema_name`: 要访问的外部表所属 Schema 名称。该示例为 `sch3`。
- `table_name`: 要访问的外部表的实际表名。该示例为 `t3`。

2. 创建表 t2 用于关联查询。

```
CREATE TABLE t2 (id INT, name VARCHAR(16));  
INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
```

3. 访问外表数据并执行关联查询。

```
SELECT t3.id, t3.name, t3.age FROM foreign_sch.t3  
    AS t3 INNER JOIN t2 ON t3.name=t2.name;
```

查询结果如下图所示。

数据洞察

同一个数仓下执行跨数据库的关联...

db2 vw2 (运行中) 保持会话 ?

执行选中 全部执行 执行计划

```

1 CREATE SCHEMA foreign_sch;
2
3 CREATE FOREIGN TABLE foreign_sch.t3 (id INT, name VARCHAR(16), age INT)
4 SERVER fdw_server_a2db3 options (schema_name 'sch3', table_name 't3');
5
6 CREATE TABLE t2 (id INT, name VARCHAR(16));
7
8 INSERT INTO t2 (id, name) VALUES (1, 'John'), (2, 'James'), (3, 'Mike');
9
10 SELECT t3.id, t3.name, t3.age FROM foreign_sch.t3
11 AS t3 INNER JOIN t2 ON t3.name=t2.name;

```

● 执行成功 请求ID: qu01j6h6ngzrw31q0he0hwmqakgy#0 扫描时间: 22ms
执行时间: 893ms 查询记录数: 2 结果集大小: 32 Bytes

	id	name	age
1	1	John	25
2	3	Mike	35

使用 postgres_fdw 访问 PieCloudDB-TP 数据的示例

如下示例为，将 PieCloudDB-TP 数据源的数据库 db1 中表 tt1 的数据分享给 PieCloudDB 数仓 account2 下的数据库 db2，数仓 account2 具有虚拟数仓 vw2。

此外，请确保 PieCloudDB-TP 数据源实例已创建示例数据库 db1 和用于分享数据的表 tt1，并置于 db1 的 tps1 Schema 下。相关 SQL 语句示例如下：

```

CREATE DATABASE db1;
-- 连接数据库 db1
CREATE SCHEMA tps1;
CREATE TABLE tps1.tt1 (id INT, name VARCHAR(16), age INT);
INSERT INTO tps1.tt1 (id, name, age) VALUES (1, 'John', 25), (2, 'Jane', 30), (3, 'Mike', 35);

```

1. 登录数仓 account2 下的数仓操作界面点击菜单栏「数据库」，在数据库 db2 下启用和配置 FDW 扩展功能。
 1. 启用 FDW 扩展。

在数据库 db2 下的 Extension 列表中，点击 **启用扩展** 并在下拉列表中选择 “postgres_fdw”，再点击 **完成** 即可添加 postgres_fdw 扩展到 Extension 列表中。

The screenshot shows the 'Extensions' tab selected in the database configuration interface. The 'Extension 列表' section displays several extensions, with 'postgres_fdw' highlighted by a red dashed box. The table columns are: 扩展名 (Extension Name), 命名空间 (Namespace), 版本号 (Version), 所有者 (Owner), and 是否可以重新定位 (Can be repositioned). The 'postgres_fdw' row has values: public, 1.0, openpie, and true.

扩展名	命名空间	版本号	所有者	是否可以重新定位
plpgsql	pg_catalog	1.0	openpie	false
gp_exttable_fdw	pg_catalog	1.0	openpie	true
postgres_fdw	public	1.0	openpie	true

此时，系统也会提示“启用成功”。

2. 创建外部服务器对象。

在数据库 db2 下的 Extension 列表中，点击扩展名称 postgres_fdw 以进入扩展详情页面，再点击 **创建 Server** 即可进入外部服务器的配置页面。

The screenshot shows the 'Extension 详情' (Details) tab for the 'postgres_fdw' extension. It displays basic information: 名称 (Name) - postgres_fdw, 所有者 (Owner) - openpie, 所属命名空间 (Namespace) - public, 版本号 (Version) - 1.0, 是否可以重新定位 (Can be repositioned) - true, and 条件表达式 (Conditional Expression). Below this, the '关联 Server' (Associate Server) section shows fields for Server 名称 (Name), 授权可用的用户 (Authorized User), and 配置信息 (Configuration Information). A blue button labeled '创建 Server' (Create Server) is highlighted with a red box.

输入如下外部服务器配置信息：

- Server 名称：外部服务器的名称。本示例使用 fdw_server_tp1db1。
- 主机：PieCloudDB-TP 数据源实例的 Host IP。在实例管理的外部接入信息页面，点击 **接入详情** 获取 Host IP。

- 端口：PieCloudDB-TP 数据源实例的端口号。在实例管理的外部接入信息页面，点击 **接入详情** 获取端口。
- （可选）Options 参数：如果 PieCloudDB-TP 数据源实例使用 PieProxy 外部接入方式，则需要填写实例 ID 信息，格式为 “-c warehouse=<实例ID>”；如果 PieCloudDB-TP 数据源实例使用 Node Port 接入方式，则不需要填写该信息。
- 连接数据库名称：要访问的 PieCloudDB-TP 数据源实例的数据库名称。该示例为 db1。
- （可选）将 Server 的 Usage 权限授予给指定用户：选择并指定其他需要外部数据源服务器使用权限的用户。该示例为 testUser。

3. 配置用户映射。

在扩展详情页面的关联 Server 列表中，选择目标外部服务器并点击隐藏菜单栏「...」下的 **用户映射** 选项即可进入用户映射的配置页面。

名称	postgres_fdw	所有者	openpie	所属命名空间	public
版本号	1.0	是否可以重新定位	true	配置参数列表	
条件表达式					

关联 Server			C	创建 Server
Server 名称	授权可用的用户	配置信息		
fdw_server_tp1db1	{testUser=U/openpie}	{host=1.0.0.0, port=5432, dbname=db1}	... <div style="margin-top: 5px;"> 用户映射 修改 删除 </div>	
-	-	-	...	
-	-	-	...	

选项配置说明如下：

- 远程用户名：可以连接外部数据源的用户。
- 密码：要执行关联查询的 PieCloudDB-TP 数据源实例的 token 密码。在实例管理的外部接入信息页面，依次点击 **接入详情>忘记了Token密码>重置Token密码** 获取 token 密码，如下图所示。



- 映射本地用户：指定本地执行外部表关联查询等操作的用户。管理员还可以将所映射的本地用户指定为 PUBLIC，这样可以创建一个公共映射。这个映射适用于所有未指定远程用户映射的用户。



- 以用户映射所指定的本地用户身份登录数仓 account2 下的「数据洞察」页面并打开一个 SQL 文件，连接数据库 db2 和虚拟数仓 vw2，以执行外部表关联查询。

1. 使用 `CREATE FOREIGN TABLE` 创建一个外部表，并按需创建 Schema 用以区分外部表的存放位置。

```
CREATE SCHEMA foreign_sch;  
CREATE FOREIGN TABLE foreign_sch.tt1 (id INT, name VARCHAR(16), age INT)  
    SERVER fdw_server_tp1db1 options (schema_name 'tps1', table_name 'tt1');
```

选项配置说明如下：

- `server_name`: 使用所创建的外部服务器对象的名称。该示例为 `fdw_server_tp1db1`。
- `schema_name`: 要访问的外部表所属 Schema 名称。该示例为 `tps1`。
- `table_name`: 要访问的外部表的实际表名。该示例为 `tt1`。

2. 创建表 t2 用于关联查询。

```
CREATE TABLE t2 (id INT, name VARCHAR(16));  
INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
```

3. 访问外表数据并执行关联查询。

```
SELECT t1.id, t1.name, t1.age FROM foreign_sch.tt1  
    AS t1 INNER JOIN t2 ON t1.name=t2.name;
```

查询结果如下图所示。

数据洞察

中文 您好: testUser

The screenshot shows the PieCloudDB Data Observatory interface. At the top, there's a toolbar with a dropdown for 'PieCloudDB-TP', a '+' button, and several other icons. Below the toolbar is a search bar with dropdowns for 'db2' and 'vw2 (运行中)' (running), and a '保持会话' (keep session) checkbox. To the right of the search bar are three buttons: '执行选中' (execute selected), '全部执行' (execute all), and '执行计划' (execute plan). The main area contains a block of SQL code:

```

1 CREATE SCHEMA foreign_sch;
2
3 v CREATE FOREIGN TABLE foreign_sch.tt1 (id INT, name VARCHAR(16), age INT)
4     SERVER fdw_server_tp1db1 options (schema_name 'tps1', table_name 'tt1');
5
6 CREATE TABLE t2 (id INT, name VARCHAR(16));
7
8 INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
9
10 v SELECT t1.id, t1.name, t1.age FROM foreign_sch.tt1
11      AS t1 INNER JOIN t2 ON t1.name=t2.name;

```

Below the SQL code, a message indicates the execution was successful: '● 执行成功 请求ID: qu01j6rt70tzj4aw63wk3jvmnkxn#0 扫描时间: 18ms' and '执行时间: 394ms 查询记录数: 2 结果集大小: 32 Bytes'. To the right of this message are three small icons: a magnifying glass, a downward arrow, and a cross. At the bottom, there's a table with two rows of data:

	id	name	age
1	1	John	25
2	3	Mike	35

使用 postgres_fdw 访问 PostgreSQL 数据的示例

如下示例为，将 PostgreSQL 数据库中表 pgt1 的数据分享给 PieCloudDB 数仓 account2 下的数据库 db2，数仓 account2 具有虚拟数仓 vw2。

此外，请确保 PostgreSQL 数据库下已创建用于分享数据的示例表 pgt1，并置于 public Schema 下。相关 SQL 语句示例如下：

```

CREATE TABLE pgt1 (id INT, name VARCHAR(16), age INT);
INSERT INTO pgt1 (id, name, age) VALUES (1, 'John', 25),(2, 'Jane', 30),(3, 'Mike', 35);

```

1. 登录数仓 account2 下的数仓操作界面点击菜单栏「数据库」，在数据库 db2 下启用和配置 FDW 扩展功能。

1. 启用 FDW 扩展。

在数据库 db2 下的 Extension 列表中，点击 **启用扩展** 并在下拉列表中选择 “postgres_fdw”，再点击 **完成** 即可添加 postgres_fdw 扩展到 Extension 列表中。

数据库 / db2

中文 您好: [REDACTED]

数据库详情 Schemas Extensions

Extension 列表



启用扩展

扩展名	命名空间	版本号	所有者	是否可以重新定位
plpgsql	pg_catalog	1.0	openpie	false
gp_exttable_fdw	pg_catalog	1.0	openpie	true
postgres_fdw	public	1.0	openpie	true

此时，系统也会提示“启用成功”。

2. 创建外部服务器对象。

在数据库 db2 下的 Extension 列表中，点击扩展名称 postgres_fdw 以进入扩展详情页面，再点击 **创建 Server** 即可进入外部服务器的配置页面。

数据库 / db2 / extensions / postgres_fdw

中文 您好: [REDACTED]

Extension 详情

名称 postgres_fdw

所有者 openpie

所属命名空间 public

版本号 1.0

是否可以重新定位 true

配置参数列表

条件表达式

关联 Server



创建 Server

Server 名称

授权可用的用户

配置信息

输入如下外部服务器配置信息：

- Server 名称：Postgres 服务器的名称。本示例使用 fdw_server_pg1。
- 主机：Postgres 服务器的 Host IP。
- 端口：Postgres 服务器的端口号。
- （可选）Options 参数：关联 Postgres 服务器可能需要的 Options 参数。
- 连接数据库名称：关联 Postgres 服务器下的数据库名称。该示例为 testdb。

- （可选）将 Server 的 Usage 权限授予给指定用户：选择并指定其他需要外部数据源服务器使用权限的用户。该示例为 testUser。

数据库 / db2 / extensions / postgres_fdw

Extension 详情

名称 postgres_fdw	所有者 openpie	所属命名空间 public
版本号 1.0	是否可以重新定位 true	配置参数列表
条件表达式		

关联 Server

Server 名称	授权可用的用户	配置信息
fdw_server_pg1	{testUser=U/openpie}	{host=... port=... dbname=testdb...}

C 创建 Server

3. 配置用户映射。

在扩展详情页面的关联 Server 列表中，选择目标外部服务器并点击隐藏菜单栏「...」下的 **用户映射** 选项即可进入用户映射配置页面。

数据库 / db2 / extensions / postgres_fdw

Extension 详情

名称 postgres_fdw	所有者 openpie	所属命名空间 public
版本号 1.0	是否可以重新定位 true	配置参数列表
条件表达式		

关联 Server

Server 名称	授权可用的用户	配置信息
fdw_server_pg1	{testUser=U/openpie}	{host=... port=... dbname=testdb...}

C 创建 Server

用户映射

修改

删除

选项配置说明如下：

- 远程用户名：可以连接外部数据源的用户。
- 密码：该用户的密码。

- 映射本地用户：指定本地执行外部表关联查询等操作的用户。管理员还可以将所映射的本地用户指定为 PUBLIC，这样可以创建一个公共映射。这个映射适用于所有未指定远程用户映射的用户。



2. 以用户映射所指定的本地用户身份登录数仓 account2 下的「数据洞察」页面并打开一个 SQL 文件，连接数据库 db2 和虚拟数仓 vw2，以执行外部表关联查询。

1. 使用 `CREATE FOREIGN TABLE` 创建一个外部表，并按需创建 Schema 用以区分外部表的存放位置。

```
CREATE SCHEMA foreign_sch;
CREATE FOREIGN TABLE foreign_sch.pgt1 (id INT, name VARCHAR(16), age INT)
  SERVER fdw_server_pg1 options (schema_name 'public', table_name 'pgt1');
```

选项配置说明如下：

- `server_name`: 使用所创建的外部服务器对象的名称。该示例为 `fdw_server_pg1`。
- `schema_name`: 要访问的外部表所属 Schema 名称。该示例为 `public`。
- `table_name`: 要访问的外部表的实际表名。该示例为 `pgt1`。

2. 创建表 t2 用于关联查询。

```
CREATE TABLE t2 (id INT, name VARCHAR(16));
INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
```

3. 访问外表数据并执行关联查询。

```
SELECT t1.id, t1.name, t1.age FROM foreign_sch.pgt1
  AS t1 INNER JOIN t2 ON t1.name=t2.name;
```

查询结果如下图所示。

The screenshot shows the PieCloudDB Data Explorer interface. At the top, there are tabs for 'postgres' and '+'. Below the tabs are dropdown menus for 'db2' and 'vw2 (运行中)', a '保持会话' switch, and search/refresh buttons. To the right are buttons for '执行选中' (Execute Selected), '全部执行' (Execute All), and '执行计划' (Execution Plan). The main area contains the following SQL code:

```
1 v CREATE FOREIGN TABLE foreign_sch.pgt1 (id INT, name VARCHAR(16), age INT)
2           SERVER fdw_server_pg1 options (schema_name 'public', table_name 'pgt1');
3
4 CREATE TABLE t2 (id INT, name VARCHAR(16));
5
6 INSERT INTO t2 (id, name) VALUES (1, 'John'), (2, 'James'), (3, 'Mike');
7
8 v SELECT t1.id, t1.name, t1.age FROM foreign_sch.pgt1
9           AS t1 INNER JOIN t2 ON t1.name=t2.name;
```

Below the code, a message indicates the execution was successful: '● 执行成功 请求ID: qu01j6hc3c7fmrhm8zxctm1j782z#0 扫描时间: 15ms' and '执行时间: 187ms 查询记录数: 2 结果集大小: 32 Bytes'. To the right are search, refresh, and close buttons. The results are displayed in a table:

	id	name	age
1	1	John	25
2	3	Mike	35

6.4.3 mysql_fdw

PieCloudDB 支持使用 mysql_fdw 扩展来直接访问 MySQL 数据源，允许用户对 MySQL 服务器上的数据执行增删改查操作（SELECT/UPDATE/DELETE/INSERT）。

注意事项

当使用 mysql_fdw 扩展在 PieCloudDB 中执行 DML 操作，例如更新（UPDATE）、删除（DELETE）或插入（INSERT）时，必须确保 MySQL 数据源的表结构中首列是单一主键或者作为复合主键的一部分。这是因为 mysql_fdw 需要使用主键来唯一确定记录，从而准确地执行这些操作。然而，对于 SELECT 操作，并没有主键约束的限制。

请注意，为了确保 mysql_fdw 能够正确地执行 DML 操作，用户应该在创建外部表时指定主键列，或者确保外部表的定义与 MySQL 数据源的表结构保持一致。

配置项说明

- **外部服务器配置项**

- Server 名称：外部服务器的名称。

外部服务器的名称应遵循数据库的标准命名方式，并保持一定的规律性，以便于用户进行管理和识别。建议命名只包含字母、数字和下划线，并且不要以数字开头。如果命名中包含空格和特殊字符，或者需要区分大小写，建议在创建外部表时，使用双引号将所关联的外部服务器名称括起来。

- 主机：外部数据源服务器的主机 IP 地址。
- 端口：外部数据源服务器的端口号。
- （可选）将 Server 的 Usage 权限授予给指定用户：将外部数据源服务器的使用权限授予指定用户。默认情况下，该权限会授予执行该操作的当前用户。

• 用户映射

- 远程用户身份信息：远程用户名和密码。
- 映射本地用户：指定本地执行外部表关联查询等操作的用户。需要注意的是，当指定所映射的本地用户为 PUBLIC 时，会创建一个公共映射，这个映射适用于所有没有指定远程用户映射的用户。

• 外部表

对于 mysql_fdw，使用 *CREATE FOREIGN TABLE* 命令创建外部表时必须指定外部数据源表所在的数据库名称和表名。

```
CREATE FOREIGN TABLE foreign_table_name (
    column_name1 data_type1,
    column_name2 data_type2,
    ...
) SERVER server_name OPTIONS (dbname 'db_name', table_name 'tb_name');
```

使用 mysql_fdw 访问 MySQL 数据

本文通过具体示例介绍如何使用 mysql_fdw 访问 MySQL 数据。

如下示例为，将 MySQL 数据库中表 myt1 的数据分享给 PieCloudDB 数仓 account2 下的数据库 db2，数仓 account2 具有虚拟数仓 vw2。

此外，请确保 MySQL 数据库已创建示例数据库 df_demo 和用于分享的表 myt1。相关 SQL 语句示例如下：

```
CREATE DATABASE df_demo;
CREATE TABLE myt1 (id INT, name VARCHAR(16), age INT);
INSERT INTO myt1 (id, name, age) VALUES (1, 'John', 25),(2, 'Jane', 30),(3, 'Mike', 35);
```

1. 登录数仓 account2 下的数仓操作界面点击菜单栏「数据库」，在数据库 db2 下启用和配置 FDW 扩展功能。
 1. 启用 FDW 扩展。

在数据库 db2 下的 Extension 列表中，点击 **启用扩展** 并在下拉列表中选择 “mysql_fdw”，再点击 **完成** 即可添加 mysql_fdw 扩展到 Extension 列表中。

扩展名	命名空间	版本号	所有者	是否可以重新定位
plpgsql	pg_catalog	1.0	openpie	false
gp_exttable_fdw	pg_catalog	1.0	openpie	true
postgres_fdw	public	1.0	openpie	true
mysql_fdw	public	1.1	openpie	true

此时，系统也会提示“启用成功”。

2. 创建外部服务器对象。

在数据库 db2 下的 Extension 列表中，点击扩展名称 mysql_fdw 以进入扩展详情页面，再点击 **创建 Server** 即可进入外部服务器的配置页面。

名称	mysql_fdw	所有者	openpie	所属命名空间	public
版本号	1.1	是否可以重新定位	true	配置参数列表	
条件表达式					

输入如下外部服务器配置信息：

- Server 名称：MySQL 服务器的名称。本示例使用 fdw_server_my1。
- 主机：MySQL 服务器的 Host IP。

- 端口：MySQL 服务器的端口号。
- (可选) 将 Server 的 Usage 权限授予给指定用户：选择并指定其他需要外部数据源服务器使用权限的用户。该示例为 testUser。

The screenshot shows the 'Extension 详情' (Extension Details) page for the 'mysql_fdw' extension. It includes basic information like name, owner, and namespace, and a section for '关联 Server' (Associated Servers) with one entry for 'fdw_server_my1'.

名称	所有者	所属命名空间
mysql_fdw	openpie	public

版本号	是否可以重新定位	配置参数列表
1.1	true	

条件表达式

关联 Server

Server 名称	授权可用的用户	配置信息
fdw_server_my1	{testUser=U/openpie}	{host=[REDACTED], port=[REDACTED]}

创建 Server

3. 配置用户映射。

在扩展详情页面的关联 Server 列表中，选择目标外部服务器并点击隐藏菜单栏「...」下的 **用户映射** 选项即可进入用户映射的配置页面。

The screenshot shows the same 'Extension 详情' page for 'mysql_fdw'. A red box highlights the 'User Mapping' option in the context menu for the 'fdw_server_my1' server entry.

用户映射

修改

删除

选项配置说明如下：

- 远程用户名：可以连接外部数据源的用户。
- 密码：该用户的密码。
- 映射本地用户：指定本地执行外部表关联查询等操作的用户。管理员还可以将所映射的本地用户指定为PUBLIC，这样可以创建一个公共映射。这个映射适用于所有未指定远程用户映射的用户。



2. 以用户映射所指定的本地用户身份登录数仓 account2 下的「**数据洞察**」页面并打开一个 SQL 文件，连接数据库 db2 和虚拟数仓 vw2，以执行外部表关联查询。

1. 使用 `CREATE FOREIGN TABLE` 创建一个外部表。

```
CREATE FOREIGN TABLE foreign_sch.myt1 (id INT, name VARCHAR(16), age INT)
  SERVER fdw_server_my1 options (dbname 'df_demo', table_name 'myt1');
```

选项配置说明如下：

- `server_name`: 使用所创建的外部服务器对象的名称。该示例为 `foreign_sch.myt1`。
- `dbname`: 要访问的外部表所属数据库的名称。该示例为 `df_demo`。
- `table_name`: 要访问的外部表的实际表名。该示例为 `myt1`。

2. 创建表 t2 用于关联查询。

```
CREATE TABLE t2 (id INT, name VARCHAR(16));
INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
```

3. 访问外表数据并执行关联查询。

```
SELECT t1.id, t1.name, t1.age FROM foreign_sch.myt1
  AS t1 INNER JOIN t2 ON t1.name=t2.name;
```

查询结果如下图所示。

The screenshot shows the Data Observatory interface with the following details:

- Session:** mysql (selected), db2 (available), vw2 (running)
- Buttons:** 保持会话 (Keep Session), 执行选中 (Execute Selected), 全部执行 (Execute All), 执行计划 (Execution Plan)
- SQL Script:**

```

1 v CREATE FOREIGN TABLE foreign_sch.myt1 (id INT, name VARCHAR(16), age INT)
2      SERVER fdw_server_my1 options (dbname 'df_demo', table_name 'myt1');
3
4 CREATE TABLE t2 (id INT, name VARCHAR(16));
5
6 INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
7
8 v SELECT t1.id, t1.name, t1.age FROM foreign_sch.myt1
9      AS t1 INNER JOIN t2 ON t1.name=t2.name;

```
- Execution Result Summary:**
 - Execution successful (执行成功)
 - Request ID: qu01j6rjvyrw7h7age230z7r1583#0
 - Scan time: 24ms
 - Execution time: 293ms
 - Number of query records: 2
 - Result set size: 32 Bytes
- Table Data:**

	id	name	age
1	1	John	25
2	3	Mike	35

6.4.4 oracle_fdw

PieCloudDB 通过 oracle_fdw 扩展支持直接访问 Oracle 数据源的数据，允许用户对 Oracle 服务器上的数据执行增删改查操作（SELECT/UPDATE/DELETE/INSERT）。

注意事项

使用 `CREATE FOREIGN TABLE` 命令创建外部表时，其表结构必须与 Oracle 数据库中的相应表结构保持一致。

配置项说明

• 外部服务器配置项

- Server 名称：Oracle 服务器的名称。
- 连接串：Oracle 数据库的连接串，格式为 “`//host:port/dbname`”。
- （可选）事务隔离级别：在下拉列表中选择 Oracle 的事务隔离级别以控制并发事务之间的相互影响。

当前版本支持序列化（SERIALIZABLE）、读已提交（READ COMMITTED）和只读（READ ONLY）。默认的事物隔离级别是序列化（SERIALIZABLE）。

- （可选）是否在 Oracle 端开启字符转换：开启后，系统将自动转换 Oracle 数据库中的字符集，以确保数据的完整性和准确性。默认情况下，该配置项是关闭的。
- （可选）是否让会话时间和 pdb 时区保持一致：开启后，系统将会话时区设置为 pdb 时区，但这一设置仅影响单个用户的会话。默认情况下，该配置项是关闭的。如果查询中涉及日期和时间的相关计算，建议开启该配置项。
- （可选）将 Server 的 Usage 权限授予给指定用户：将外部数据源服务器的使用权限赋予给指定用户。默认授予给当前执行操作的用户。

- **用户映射**

- 远程用户身份信息：当前会话所连接的 Oracle 数据库用户的用户名和密码。
- 映射本地用户：指定本地执行外部表关联查询等操作的用户。需要注意的是，当指定所映射的本地用户为 PUBLIC 时，会创建一个公共映射，这个映射适用于所有没有指定远程用户映射的用户。

- **外部表**

- table: Oracle 数据库中被访问的表名，一般要求大写。

注意：

在 Oracle 数据库中，标识符（包括表名、列名、索引名等）的默认行为是大小写不敏感的，当用户创建一个标识符时，Oracle 会自动将其转换为大写字母存储。创建外表时应和 Oracle 数据库的标识符大小写一致。

- （可选）dblink: 访问表所用的 Oracle 数据库链接。这个名称必须与 Oracle 系统目录中的完全一致。
- （可选）schema: Oracle 数据库中被访问的表所属用户的名称，要求是大写的形式。如果要访问的表不属于当前所连接的 Oracle 用户（即用户映射中的用户）的表，则需要指定该选项。
- （可选）max_long: 设置 Oracle 的表中 LONG、LONG RAW、XMLTYPE 类型列的最大长度限制。有效的取值范围为 1 到 1073741823 之间的整数，默认值为 32767。
- （可选）readonly: 设置 readonly 参数为 YES、ON 或 TRUE 时，外部表将被限制为仅允许执行只读操作（如 SELECT）。默认情况下，该参数设置为 FALSE，即允许执行 INSERT、UPDATE 和 DELETE 操作。

- (可选) sample_percent: 该选项只影响 ANALYZE 处理，设置用于计算因统计信息而随机选择的 Oracle 表中数据的比例。有效的取值范围为 1 到 100 之间的整数，默认值为 100。
- (可选) prefetch: 设置扫描外表时，与 Oracle 数据库之间进行单次传输的数据的行数。有效的取值范围为 1 到 1000 之间的整数，默认值为 50。
- (可选) lob_prefetch: 为 BLOB、CLOB 和 BFILE 值设置预取的字节数，默认值为 1048576。

外部表的列

- (可选) key: 如果选项设置为 YES、ON 或 TRUE，则对应的 Oracle 外部表上的列将被视为主键列。如果需要对 Oracle 数据库中的映射表执行 UPDATE 和 DELETE 操作，则必须在使用 `CREATE FOREIGN TABLE` 所创建的外表的主键的列上设置该选项为 YES、ON 或 TRUE。
- (可选) strip_zeros: 如果选项设置为 YES、ON 或 TRUE，则从 Oracle 数据库的映射表中导入数据时，会去除目标字段中的所有 ASCII 0 字符。这个选项只对 character, character varying 和 text 类型的列有意义。

使用说明

- Oracle 权限: Oracle 用户需要具有 CREATE SESSION 权限以及从相关表或视图中选择数据的权限。对于 EXPLAIN VERBOSE，用户还需要对 `V$SQL` 和 `V$SQL_PLAN` 有 SELECT 权限。
- 连接: 由于为每个单独的查询创建一个 Oracle 会话的成本很高，所以 `oracle_fdw` 会缓存 Oracle 连接。当 PieCloudDB 会话结束时，所有 Oracle 连接将自动关闭。
使用函数 `oracle_close_connections()` 也可以关闭所有缓存的 Oracle 连接。这对于不总是访问外部表的长时间运行的会话很有用，但不能在修改 Oracle 数据的事务中调用此函数。
- 列: 当您定义一个外部表时，Oracle 表的列将按照它们的定义顺序映射到所创建的外表的列。`oracle_fdw` 只会包含 Oracle 查询中实际需要的列，如果需要执行 UPDATE 或 DELETE，请确保在所有属于表的主键的列上设置了 key 选项。

有关更多 `oracle_fdw` 的使用信息，请参见 [oracle_fdw 官方文档](#)。

使用 `oracle_fdw` 访问 Oracle 数据的示例

本文通过具体示例介绍如何使用 `oracle_fdw` 访问 Oracle 数据。

如下示例为，将 Oracle 数据库中表 `otab` 的数据分享给 PieCloudDB 数仓 `account2` 下的数据库 `db2`，数仓 `account2` 具有虚拟数仓 `vw2`。

此外，请确保 Oracle 数据库已创建用户 ORAS1 和用于分享数据的表 otab。相关 SQL 语句示例如下：

```
CREATE USER ORAS1 IDENTIFIED BY 123456 DEFAULT TABLESPACE users TEMPORARY
TABLESPACE temp QUOTA 100M ON users;

GRANT CREATE SESSION TO ORAS1;

CREATE TABLE ORAS1.otab (id INT PRIMARY KEY, name VARCHAR(16), age INT);

INSERT INTO ORAS1.otab (id, name, age) VALUES (1, 'John', 25);
INSERT INTO ORAS1.otab (id, name, age) VALUES (2, 'Jane', 30);
INSERT INTO ORAS1.otab (id, name, age) VALUES (3, 'Mike', 35);
```

1. 管理员登录数仓 account2 下的数仓操作界面点击菜单栏「数据库」，在数据库 db2 下启用和配置 FDW 扩展功能。

1. 启用 FDW 扩展。

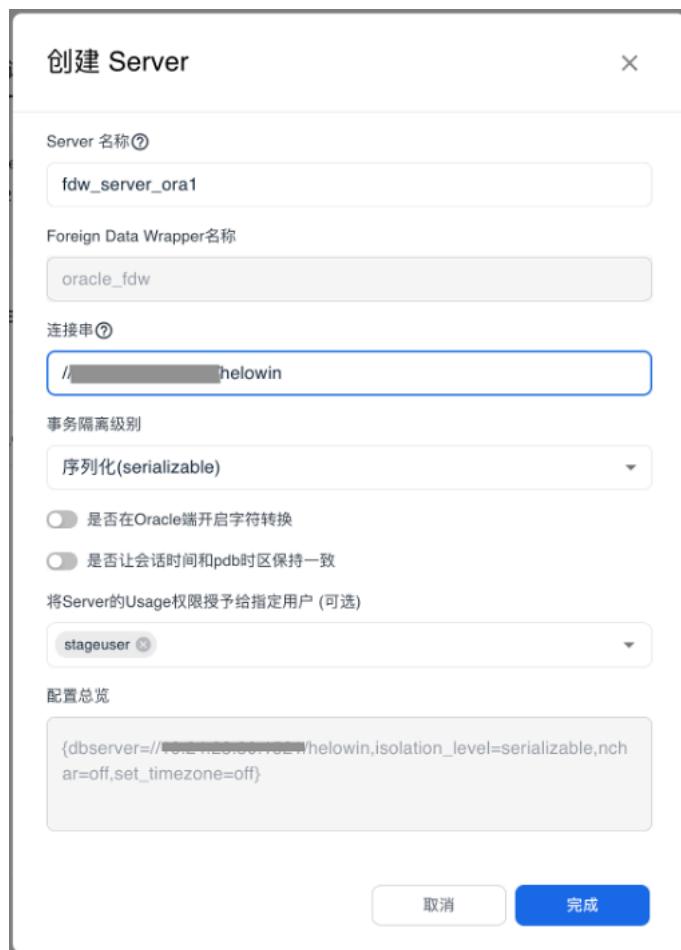
在数据库 db2 下的 Extension 列表中，点击 **启用扩展** 并在下拉列表中选择 “oracle_fdw”，再点击 **完成** 即可添加 oracle_fdw 扩展到 Extension 列表中。

扩展名	命名空间	版本号	所有者	是否可以重新定位
plpgsql	pg_catalog	1.0	openpie	false
gp_exttable_fdw	pg_catalog	1.0	openpie	true
oracle_fdw	public	1.2	openpie	true

此时，系统也会提示“启用成功”。

2. 创建外部服务器对象。

在数据库 db2 下的 Extension 列表中，点击目标扩展 oracle_fdw 以进入相应的扩展详情页面，再点击 **创建 Server** 即可进入外部服务器的配置页面。



设置如下外部服务器的配置信息：

- Server 名称：Oracle 服务器的名称。本示例使用 fdw_server_ora1。
- 连接串：Oracle 数据库的连接串，格式为 “//host:port/dbname”。
- （可选）事务隔离级别：在下拉列表中选择 Oracle 的事务隔离级别以控制并发事务之间的相互影响。默认的事物隔离级别是序列化（SERIALIZABLE）。
- （可选）是否在 Oracle 端开启字符转换：开启后，系统将自动转换 Oracle 数据库中的字符集，以确保数据的完整性和准确性。默认情况下，该配置项是关闭的。
- （可选）是否让会话时间和 pdb 时区保持一致：开启后，系统将会话时区设置为 pdb 时区，但这一设置仅影响单个用户的会话。默认情况下，该配置项是关闭的。如果查询中涉及日期和时间的相关计算，建议开启该配置项。
- （可选）将 Server 的 Usage 权限授予给指定用户：将外部数据源服务器的使用权限赋予给指定用户。默认授予给当前执行操作的用户。

3. 配置用户映射。

在扩展详情页面的关联 Server 列表中，点击目标服务器的隐藏菜单栏「...」并选择 **用户映射** 选项即可进入用户映射的配置页面。

Extension 详情

名称 oracle_fdw 所有者 openpie 所属命名空间 public
版本号 1.2 是否可以重新定位 true 配置参数列表
条件表达式

关联 Server

Server 名称	授权可用的用户	配置信息
fdw_server_ora1	{stageuser=U/openpie}	{dbserver=//,isolation_...}

...
用户映射
修改
删除

选项配置说明如下：

- 远程用户名：当前会话所连接的 Oracle 数据库用户的用户名。

注意：

请确保该 Oracle 数据库用户具有当前会话的连接权限。授予用户会话连接权限的参考命令为 GRANT CREATE SESSION TO oracle_user。

- 密码：当前会话所连接的 Oracle 数据库用户的密码。
- 映射本地用户：指定执行外部表关联查询等操作的用户。管理员还可以将所映射的本地用户指定为 PUBLIC，这样可以创建一个公共映射。这个映射适用于所有未指定远程用户映射的用户。



2. 以用户映射所指定的本地用户身份登录数仓 account2 下的「数据洞察」页面并打开一个 SQL 文件，连接数据库 db2 和虚拟数仓 vw2，以执行外部表相关操作。

1. 使用 `CREATE FOREIGN TABLE` 命令创建一个外部表。

```
CREATE Schema foreign_sch;
CREATE FOREIGN TABLE foreign_sch.otab1 (
    id INT OPTIONS (key 'true') NOT NULL,
    name VARCHAR(16),
    age INT
) SERVER fdw_server_ora1 OPTIONS (
    table 'OTAB',
    schema 'ORAS1',
    max_long '32767',
    readonly 'false',
    sample_percent '100',
    prefetch '200'
);
```

配置项说明如下：

- **table**: Oracle 数据库中被访问的表名，一般是大写。但在 Oracle 数据库中，用户也可以使用双引号强制转换标识符为小写，所以需要注意，创建外表时应和 Oracle 内的标识符大小写一致。
- **schema**: 如果要访问的表不属于当前所连接的 Oracle 用户（即用户映射中的用户）的表，则需要指定该选项。由于本示例中的 Oracle 连接用户为 SYSTEM，所以需要指定 schema 为 “ORAS1”。
- （可选）**max_long**: 设置 Oracle 的表中 LONG、LONG RAW、XMLTYPE 类型列的最大长度限制。
- （可选）**readonly**: 该参数设置为 FALSE，即允许执行 INSERT、UPDATE 和 DELETE 操作。
- （可选）**sample_percent**: 设置用于计算因统计信息而随机选择的 Oracle 表中数据的比例，默认值为 100。

- （可选）prefetch：设置扫描外表时，与 Oracle 数据库之间进行单次传输的数据的行数。

2. 创建表 t2 用于关联查询。

```
CREATE TABLE t2 (id INT, name VARCHAR(16));
INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
```

3. 访问外表数据并执行增删改查操作。

- 访问外表数据并执行关联查询。

```
SELECT t1.id, t1.name, t1.age FROM foreign_sch.otab1
AS t1 INNER JOIN t2 ON t1.name=t2.name;
```

查询结果如下图所示。

The screenshot shows the πDataCS Data Observatory interface. The top navigation bar includes '数据洞察' (Data Observatory), '中文' (Chinese), and a user profile. The main window has a tab labeled 'oracle_fdw'. Below the tabs are buttons for 'db2' (selected), 'vw2 (运行中)' (running), '保持会话' (keep session), and execution options: '执行选中' (execute selected), '全部执行' (execute all), and '执行计划' (execution plan). The SQL editor displays the following code:

```

2 CREATE Schema foreign_sch;
3 DROP FOREIGN TABLE foreign_sch.otab1;
4 v CREATE FOREIGN TABLE foreign_sch.otab1 (
5     id INT OPTIONS(key 'true') NOT NULL,
6     name VARCHAR(16),
7     age INT
8 ) SERVER fdw_server_ora1 OPTIONS (
9     table 'OTAB',
10    schema 'ORAS1',
11    max_long '32767',
12    readonly 'false',
13    sample_percent '100',
14    prefetch '200'
15 );
16 CREATE TABLE t2 (id INT, name VARCHAR(16));
17
18 INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
19
20 v SELECT t1.id, t1.name, t1.age FROM foreign_sch.otab1
21 AS t1 INNER JOIN t2 ON t1.name=t2.name;

```

Execution status: 执行成功 (Execution successful). Request ID: qu01j9zv1mmynh4nxtkykwpa6mr4#0. Scan time: 15ms. Execution time: 1.6s. Result count: 2. Result size: 32 Bytes.

The results table shows the following data:

	id	name	age
1	1	John	25
2	3	Mike	35

- 访问外表数据并执行 INSERT 操作。

```
INSERT INTO foreign_sch.otab1 (id, name, age) VALUES (4, 'Mary', 40);
SELECT * FROM foreign_sch.otab1;
```

查询结果如下图所示。

数据洞察

oracle_fdw +

db2 vw2 (运行中) 保持会话 ? 执行选中 全部执行 执行计划

```
11 max_long '32767',
12 readonly 'false',
13 sample_percent '100',
14 prefetch '200'
15 );
16 CREATE TABLE t2 (id INT, name VARCHAR(16));
17
18 INSERT INTO t2 (id, name) VALUES (1, 'John'), (2, 'James'), (3, 'Mike');
19
20 v SELECT t1.id, t1.name, t1.age FROM foreign_sch.otab1
21      AS t1 INNER JOIN t2 ON t1.name=t2.name;
22
23 INSERT INTO foreign_sch.otab1 (id, name, age) VALUES (4, 'Mary', 40);
24 SELECT * FROM foreign_sch.otab1;
```

● 1 INSERT INTO ...
● 2 SELECT * FR...

● 执行成功 请求ID: qu01j9zvp9c7w3sgb8x416njcvzz#2 扫描时间: 14ms
执行时间: 40ms 查询记录数: 4 结果集大小: 52 Bytes

	id	name	age
1	4	Mary	40
2	1	John	25
3	2	Jane	30
4	3	Mike	35

- 访问外表数据并执行 UPDATE 操作。

```
UPDATE foreign_sch.otab1 SET age=20 WHERE id=4;
SELECT * FROM foreign_sch.otab1;
```

查询结果如下图所示。

数据洞察

oracle_fdw

中文 您好: []

db2 vvw2 (运行中) 保持会话 ? 执行选中 全部执行 执行计划

```
14      );
15  CREATE TABLE t2 (id INT, name VARCHAR(16));
16  );
17
18 ) INSERT INTO t2 (id, name) VALUES (1,'John'),(2,'James'),(3,'Mike');
19
20 v  SELECT t1.id, t1.name, t1.age FROM foreign_sch.otab1
21      AS t1 INNER JOIN t2 ON t1.name=t2.name;
22
23 INSERT INTO foreign_sch.otab1 (id, name, age) VALUES (4, 'Mary', 40);
24
25 UPDATE foreign_sch.otab1 SET age=20 WHERE id=4;
26 SELECT * FROM foreign_sch.otab1;
27
```

1 UPDATE forei...
2 SELECT * FR...

● 执行成功 请求ID: qu01j9zvsr4jqq7hbrnb9ghbxq57#2 扫描时间: 20ms
执行时间: 115ms 查询记录数: 4 结果集大小: 52 Bytes

	id	name	age
1	4	Mary	20
2	1	John	25
3	2	Jane	30
4	3	Mike	35

- 访问外表数据并执行 DELETE 操作。

```
DELETE FROM foreign_sch.otab1 WHERE id=4;  
SELECT * FROM foreign_sch.otab1;
```

查询结果如下图所示。

数据洞察

oracle_fdw +

db2 vw2 (运行中) 保持会话 ? 执行选中 全部执行 执行计划

```

16 CREATE TABLE t2 (id INT, name VARCHAR(16));
17
18 INSERT INTO t2 (id, name) VALUES (1, 'John'),(2, 'James'),(3, 'Mike');
19
20 SELECT t1.id, t1.name, t1.age FROM foreign_sch.otab1
21      AS t1 INNER JOIN t2 ON t1.name=t2.name;
22
23 INSERT INTO foreign_sch.otab1 (id, name, age) VALUES (4, 'Mary', 40);
24
25 UPDATE foreign_sch.otab1 SET age=20 WHERE id=4;
26
27 DELETE FROM foreign_sch.otab1 WHERE id=4;
28 SELECT * FROM foreign_sch.otab1;
29

```

● 执行成功 请求ID: qu01j9zvxbn1pp5kmsww6ny5q09y#2 扫描时间: 13ms
执行时间: 72ms 查询记录数: 3 结果集大小: 42 Bytes

	id	name	age
1	1	John	25
2	2	Jane	30
3	3	Mike	35

6.5 数据集成

6.5.1 数据集成简介

PieCloudDB 的数据集成 (Dataflow) 功能支持从不同的数据库抽取数据，并将其加载到同构或异构的目标数据库中。这一数据迁移任务管理服务依托于 PieCloudDB 管控平台实现。用户只需在平台上配置数据源、数据流和作业调度，即可实现数据全量迁移与备份的可视化和精细化管理。

数据集成功能主要适用于将多源数据库的历史数据进行全量或者增量迁移至中央历史数仓，并进行汇聚处理的场景。

在当前版本中，数据集成支持两种数据迁移任务类型：离线同步和实时同步。对于非周期性作业，系统会创建一个单次任务，这可以是离线同步任务或实时同步任务；对于周期性作业，系统会创建一个定期执行的任务并按照预设周期运行。但请注意，周期性作业目前仅适用于表到表的离线同步任务。

注意:

数据集成功能属于 PieCloudDB 企业版的定制化功能，需要进行独立部署。有关部署方式，请联系 OpenPie 技术支持团队。

核心特性

数据集成的核心特性包括但不限于：

- 支持多种数据源

能够从多种数据源抽取数据并加载到目标数据库，实现了对异构数据源（例如 MySQL、PostgreSQL 等）的自动化配置。同时，提供目标数据库 Schema 创建语法的自动化支持，并内置及自定义字段改写规则。

- 支持多种内置校验算法

在数据迁移过程中，提供数据一致性校验，包括全量目标数据表校验和单次作业范围校验，确保数据的全面和高效准确性。

- 任务调度精细化管理

用户可以按需选择单表、多表或全库的方式进行历史数据迁移，并支持灵活的单次或周期性迁移作业策略。

- 多线程高效运行

数据集成过程采用多线程技术，高效处理数据的导出、压缩、导入以及数据校验等任务，显著提升数据处理效率。

- 可视化监控

基于 PieCloudDB 管控平台，实现数据集成流程的可视化监控。例如，实时展示数据导出和导入的进度条；通过数据集成总览面板全局查看数据集成任务的状态。

- 国产化部署

支持部署在 x86 和 ARM CPU 架构的服务器，并适配主流国产操作系统，例如 Kylin OS 等。

前期准备

在执行数据集成之前，请确保完成以下准备工作：

- 源数据库准备工作

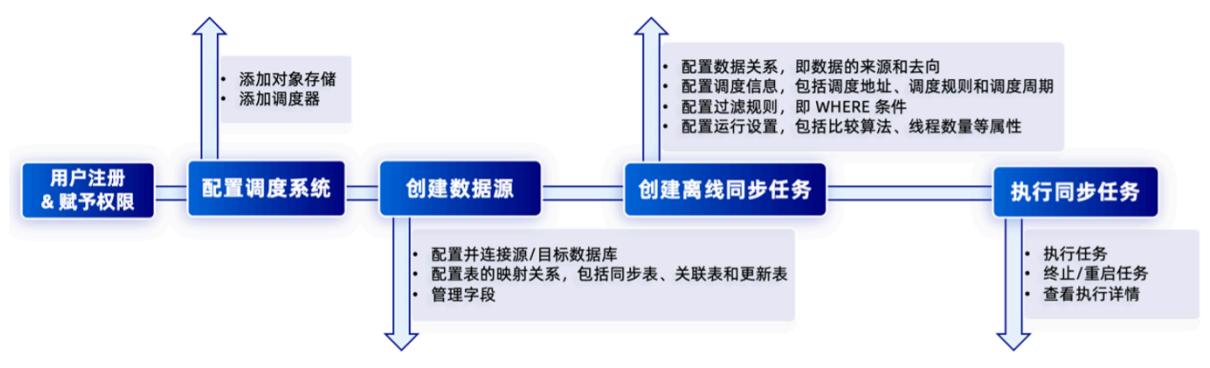
确保源数据库提供了一个具有所需迁移库表读取权限的用户。此外，该用户需要能够从运行数据集成的服务器或容器访问源数据库。

- 目标数据库准备工作

预先在目标数据库中创建好需要导入数据的表，并确保有一个用户具备对这些表的写入权限。同样，该用户需要能够从运行数据集成的服务器或容器访问目标数据库。

离线同步任务使用流程

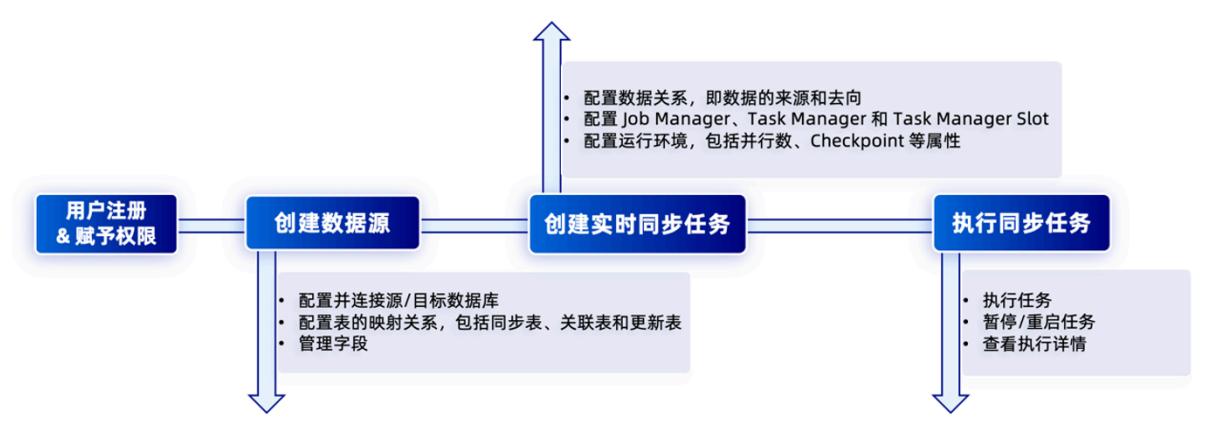
数据集成功能支持创建离线同步任务，适用于批处理数据的场景。对于离线同步任务，建议使用流程如下：



1. 授予用户数据集成功能权限
2. 配置调度系统
3. 创建源数据源并配置连接
4. 配置源数据源的表的映射关系
5. 创建和执行 DDL 任务为目标数据源创建新的表
6. 配置目标数据源的表的映射关系
7. 创建离线同步任务

实时同步任务使用流程

数据集成功能支持创建实时同步任务，适用于流处理数据的场景。对于实时同步任务，建议使用流程如下：



1. 授予用户数据集成功能权限
2. 创建源数据源并配置连接
3. 配置源数据源的表的映射关系

4. 创建和执行 DDL 任务为目标数据源创建新的表
5. 配置目标数据源的表的映射关系
6. 创建实时同步任务

6.5.2 授予用户数据集成功能权限

数据集成功能具有两类权限：只读和管理，属于系统权限范畴。具有“只读”权限的用户仅能在总览、数据源、数据流和作业调度页面执行一些查询操作；除了包含只读权限所开放的功能外，具有“管理”权限的用户还可以对各个页面上的功能执行创建、删除和修改等操作。

数仓预设角色中，accountmanager 角色作为数仓管理员，默认同时具有数据集成功能的管理权限（权限名称为“数据集成功能管理”）和只读权限（权限名称为“数据集成功能只读”）。

本文通过具体示例来介绍如何在 PieDataCS 云原生平台对一位新的数仓用户授予数据集成功能的管理权限。

1. 以管理员身份登录云原生平台并创建自定义角色（例如 dataflow-manager），并将数据集成的管理权限“数据集成功能管理”授予该角色。



如果当前其他角色已具有“数据集成功能管理”的权限，在创建自定义角色时也可以通过直接继承角色的方式获得该权限。

2. 以管理员身份在「用户」页面创建一个数仓用户，例如 dataflow-user。
3. 在「角色」页面，以管理员身份为新数仓用户 dataflow-user 赋予步骤 1 中所创建的角色 dataflow-manager，则该用户会继承该角色的数据集成功能管理权限。

4. 以步骤 3 所创建的数仓用户 dataflow-user 登录 PieDataCS 云原生平台，即可开始使用数据集成的功能。

6.5.3 信息总览

数据集成的「总览」页面展示指定时间内，离线同步的数据导入和导出的历史统计信息，包括执行次数、导入记录数、平均执行时间和总执行时间。

The screenshot shows the 'Overview' page for data integration. At the top, there are tabs for 'Overview', 'Source Dimension', and 'Table Dimension'. Below the tabs, there is a date range selector from '2024/11/18' to '2024/11/24', and buttons for 'This Week' and 'This Month'. On the right, there is a 'System Configuration' button.

Data Import Statistics:

Execution Count	Import Record Count	Avg Execution Time (s)	Total Execution Time (s)
89	326,553,094	38.40	3,417.68

Data Export Statistics:

Execution Count	Export Record Count	Avg Execution Time (s)	Total Execution Time (s)
103	326,553,171	129.21	13,308.80

该页面上方提供时间筛选控件，以方便用户根据时间筛选信息：

- 使用**时间范围** 控件可以筛选指定起止时间范围内的数据导入和导出的信息，单位为天。
- 使用**看本周** 控件可以筛选显示本周内的数据导入和导出的相关信息。
- 使用**看本月** 控件可以筛选显示本月内的数据导入和导出的相关信息。

Source 维度

在「总览」页面，**Source 维度** 基于数据源的维度展示数据集成的统计信息。

The screenshot shows the 'Source Dimension' page. At the top, there are tabs for 'Overview', 'Source Dimension', and 'Table Dimension'. Below the tabs, there is a dropdown for 'Operation Type' set to 'All', a date range selector from '2024/11/15' to '2024/11/22', a search bar for 'Data Source Name Search', and a clear button.

Data Source Statistics:

Data Source	Operation Type	Execution Count	Record Count	Avg Execution Time (s)	Total Execution Time (s)	Action
pdb_pg_nodeport	Data Import	1	13	1.76	1.76	View Table Statistics
pdb_mysql	Data Import	1	119,994,608	895.04	895.04	View Table Statistics
pdb_pg	Data Import	20	84	1.89	37.77	View Table Statistics
pg12	Data Export	23	112	0.39	9.07	View Table Statistics

Source 维度的信息列表显示如下信息：

- 数据源：数据集成所使用的数据源的名称。
- 操作类型：数据集成操作类型，包括“数据导入”和“数据导出”。
- 执行次数：数据集成操作被执行的次数。
- 记录数：数据集成操作所输出数据的行数。
- 平均执行时间(秒)：该数据源的数据集成操作的平均执行时间，该值的计算公式为“总执行时间/执行次数”。单位为秒。
- 总执行时间(秒)：该数据源的数据集成操作的总执行时间。单位为秒。
- 操作：相关扩展操作。当前版本支持“查看table统计”，点击该按钮即可跳转至当前数据源所对应的表维度页面，详细信息请参见本章节 **Table 维度** 的内容。

表名	所属数据源	操作类型	↑ 执行次数	记录数	平均执行时间(秒)	总执行时间(秒)
tb	pdb_pg_nodeport	数据导入	1	13	1.76	1.76

Source 维度的信息列表提供如下快捷操作：

- 根据操作类型筛选信息：使用**操作类型**控件可以筛选指定操作类型的数据集成信息。默认展示全部操作类型的信息。
- 指定时间范围筛选信息：使用**开始时间-结束时间**控件可以筛选显示某个时间范围（单位为天）内的信息。
- 根据数据源名称搜索信息：使用**输入数据源名称搜索**的搜索框可以通过数据源的名称搜索相关信息。
- 按照执行次数排序：点击表头的“执行次数”字段，可以按照该操作的执行次数的大小顺序显示列表信息。
- 按照记录数排序：点击表头的“记录数”字段，可以按照该操作所执行的记录数的大小顺序显示列表信息。
- 按照平均执行时间排序：点击表头的“平均执行时间”字段，可以按照该操作的平均执行时间的大小顺序显示列表信息。
- 按照总执行时间排序：点击表头的“总执行时间”字段，可以按照该操作的总执行时间的大小顺序显示列表信息。
- 立即刷新：点击立即刷新图标可以即刻同步当前的列表信息为最新。

Table 维度

在「总览」页面，**Table 维度** 基于数据源中表的维度展示数据集成的统计信息。



The screenshot shows a user interface for managing database dimensions. At the top, there are tabs: '总览' (Overview), 'Source 维度' (Source Dimension), and 'Table 维度' (Table Dimension), with 'Table 维度' being the active tab and highlighted with a red border. Below the tabs is a toolbar with filters for '操作类型' (Operation Type) set to '全部' (All), date range from '2024/11/15' to '2024/11/22', and search fields for '输入表名搜索' (Search by Table Name) and '输入数据源名称搜索' (Search by Data Source Name). On the far right of the toolbar is a blue button labeled '系统配置' (System Configuration). The main area is a table listing four data integration operations:

表名	所属数据源	操作类型	↑ 执行次数	记录数	平均执行时间(秒)	总执行时间(秒)
part	pdb_oracle	数据导入	1	2,000,000	67.73	67.73
supplier	pdb_oracle	数据导入	1	100,000	5.10	5.10
customer	pdb_oracle	数据导入	1	1,500,000	68.58	68.58
orders	pdb_oracle	数据导入	1	15,000,000	229.77	229.77

Table 维度的信息列表显示如下字段：

- 表名：表的名称。
- 所属数据源：该表所属的数据源。
- 操作类型：数据集成的操作类型，包括“数据导入”和“数据导出”。
- 执行次数：该表被执行的次数。
- 记录数：数据集成操作所流转的记录数量。
- 平均执行时间(秒)：该表的数据集成操作的平均执行时间，该值的计算公式为“总执行时间/执行次数”。单位为秒。
- 总执行时间(秒)：该表的数据集成操作的总执行时间。单位为秒。

Table 维度的信息列表提供如下快捷操作：

- 根据操作类型筛选信息：使用**操作类型**控件可以筛选指定操作类型的数据集成信息。默认展示全部操作类型的信息。
- 指定时间范围筛选信息：使用**开始时间-结束时间**控件可以筛选显示某个时间范围（单位为天）内的信息。
- 根据表名称搜索信息：使用**输入表名搜索**的搜索框可以通过表的名称搜索相关信息。
- 根据数据源名称搜索信息：使用**输入数据源名称搜索**的搜索框可以通过数据源的名称搜索相关信息。
- 按照执行次数排序：点击表头的“执行次数”字段，可以按照该表被执行次数的大小顺序显示列表信息。
- 按照记录数排序：点击表头的“记录数”字段，可以按照该表所执行的记录数的大小顺序显示列表信息。
- 按照平均执行时间排序：点击表头的“平均执行时间”字段，可以按照该表的平均执行时间的大小顺序显示列表信息。
- 按照总执行时间排序：点击表头的“总执行时间”字段，可以按照该表的总执行时间的大小顺序显示列表信息。
- 立即刷新：点击立即刷新图标可以即刻同步当前的列表信息为最新。

6.5.4 配置调度系统

系统配置涵盖 S3 配置和调度器设置。在数据的离线同步过程中，调度器（dataflow-worker）负责管理数据的导出、导入和校验任务。数据源中导出的数据将被保存到 S3，作为备份存储。

提示：

S3 配置和调度器配置仅适用于批处理数据的场景，该配置信息会应用于创建离线同步任务。

数据集成的「总览」页面提供了一个进入系统配置的入口。用户在该页面点击 **系统配置** 即可进入系统配置页面。

添加 S3 配置

S3 对象存储作为数据流转的中间站，负责存储从不同的数据源抽取的数据，然后加载数据到目标数据库（例如 PieCloudDB）。

添加 S3 配置的步骤如下：

1. 在系统配置的「**S3 配置**」页面，点击 **添加 S3 配置** 即可进入创建 S3 配置的页面。
2. 在创建 S3 配置页面，输入如下 S3 配置信息：
 - 名称：S3 配置的自定义名称。
 - Endpoint：与 S3 服务通信的终端 URL 地址。
 - Region：S3 服务所在的区域，每个区域都有一个唯一的标识符。
 - Bucket Name：存放备份文件的 S3 Bucket 名称，需要提前创建。
 - S3 Access Key：访问密钥。
 - S3 Secret Key：私有访问密钥。
3. 点击 **完成**。操作成功后，新添加的 S3 配置会自动同步到列表中。

系统配置的「**S3 配置**」页面以列表形式展示当前已有的 S3 配置信息。该页面还支持添加 S3 配置，以及对 S3 配置的修改和删除操作。

ID	名称	Endpoint	Region	Bucket Name	S3 Access Key	S3 Secret Key	设为默认配置	更多操作
gxtan_S3	http://[REDACTED]			gxtan-type-demo	minioadmin	*	<input checked="" type="checkbox"/>	⋮

针对于添加后的 S3 配置，用户可以执行如下操作：

- 修改/删除：指定某条 S3 配置，用户点击隐藏的列表并选择对应选项来“修改”或者“删除”该 S3 配置。
- 设为默认配置：如果系统配置了多个 S3 对象存储，第一个 S3 配置会自动成为默认配置。用户可以滑动开关指定默认的 S3 配置。

该设置会影响离线同步任务的运行配置中所涉及的 S3_config_id 字段值，如果在创建离线同步任务时未手动配置 S3_config_id，系统就会使用默认的 S3 配置。有关离线同步任务的详细信息，请参见 [创建离线同步任务](#)。

添加调度器配置

在数据集成过程中，调度器负责发起数据的导出、导入和校验任务。调度系统会根据执行周期、调度环境、运行配置等因素，对数据的离线同步任务进行协调和管理，这称为作业调度。

添加调度器配置的步骤如下：

- 在系统配置的「调度器配置」页面，点击 **添加调度器** 即可进入创建调度器的页面。
- 在创建调度器配置页面，输入如下调度器信息：
 - 名称：调度器的自定义名称。
 - 主机：Piescheduler 所在宿主机的 IP 地址。
 - 端口：Piescheduler 的 nodePort 暴露出的端口号。
 - Workflow Names：工作流的名称。
 - Task Queues：任务队列的名称。

提示：

基于客户真实环境中所部署的 LoadBalancer，主机和端口号可以为 LoadBalancer 的 IP 和其所暴露的端口号。

- 点击 **完成**。操作成功后，新添加的调度器配置会自动同步到列表中。

系统配置的「调度器配置」页面以列表形式展示当前已有的调度器配置信息。在目标调度器配置所在行，启用默认配置开关，可以将目标调度器设置为默认调度器。

S3 配置	调度器配置	操作
		添加调度器
端口	Workflow Names	Task Queues

以下展示了调度器配置列表：

端口	Workflow Names	Task Queues	设为默认配置	创建时间	更新时间	更多
PyFileBatchWorkflow	PyFileBatchWorkflow	PYTHON_FILE_BATCH_QUEUE	<input checked="" type="checkbox"/>	2024-05-21 09:36:28	2025-01-16 10:34:02	⋮
PyFileBatchWorkflow,FlinkFileBatchWorkflow		PYTHON_FILE_BATCH_QUEUE,FLINK_FILE_BATCH_QUEUE	<input type="checkbox"/>	2024-05-07 10:33:03	2024-05-07 10:33:03	⋮

在列表中，用户点击目标调度器配置所在行的隐藏的下拉列表并选择对应选项来 **修改** 或者 **删除** 该调度器配置。

6.5.5 创建数据源

数据集功能支持与多种不同数据源的连接，当前版本支持连接到 PostgreSQL、PieCloudDB TP、PieCloudDB XP、PieCloudDB、MySQL 以及 Oracle 等数据源类型。用户通过配置相应的连接参数并进行连通性测试后，即可成功连接到数据源，为数据集成做好准备。

创建数据源并配置连接

用户必须首先创建数据源，然后配置相应的连接参数，之后才能执行数据集成操作。

在数据集成的「**数据源**」页面，创建数据源并配置连接的步骤如下：

1. 点击 **创建数据源** 即可进入创建数据源的页面。
2. 在“**类型**”下拉列表中选择一种数据源类型。当前版本支持 Oracle、MySQL、PostgreSQL、PieCloudDB TP、PieCloudDB XP 和 PieCloudDB。
3. 输入数据源名称和对该数据源的描述（选填）。
4. 点击 **完成**。操作成功后，新创建的数据源会自动同步到列表中。
5. 在数据源列表中的新创建数据源的“**操作**”栏下，点击 **连接配置** 即可进入配置页面。
6. 在数据源连接配置页面，分别输入所选数据源的连接信息（各数据源的连接信息参考下表），并点击 **完成** 以确认更改。

字段	含义	适用的数据源类型
host (必填)	要连接数据源的主机 IP 地址	<ul style="list-style-type: none">• Oracle• MySQL• PostgreSQL• PieCloudDB TP• PieCloudDB XP• PieCloudDB
port (必填)	要连接数据源的主机端口	<ul style="list-style-type: none">• Oracle• MySQL• PostgreSQL• PieCloudDB TP• PieCloudDB XP• PieCloudDB

字段	含义	适用的数据源类型
database (必填)	要连接数据源的目标数据库的名称	<ul style="list-style-type: none"> • MySQL • PostgreSQL • PieCloudDB TP • PieCloudDB XP • PieCloudDB
user (必填)	要连接数据源的数据库的用户名	<ul style="list-style-type: none"> • Oracle • MySQL • PostgreSQL • PieCloudDB TP • PieCloudDB XP • PieCloudDB
password (必填)	要连接数据源的数据库的密码	<ul style="list-style-type: none"> • Oracle • MySQL • PostgreSQL • PieCloudDB TP • PieCloudDB XP • PieCloudDB
warehouse (选填)	如果连接的外部数据源为 PieCloudDB 虚拟数仓且使用 PieProxy 外部接入方式，则须填入外部数据源的虚拟数仓 ID 信息	PieCloudDB
schema (选填)	要连接数据源的目标 Schema 名称	<ul style="list-style-type: none"> • PostgreSQL • PieCloudDB TP • PieCloudDB XP • PieCloudDB
servertimezone (选填)	要连接数据源的服务器所在的时区	MySQL
service (必填)	要连接数据源的目标数据库服务	Oracle
sid (必填)	要连接数据源的目标数据库实例的系统标识符	Oracle
connparam (选填)	连接参数	Oracle
extparam (选填)	外部参数	<ul style="list-style-type: none"> • PostgreSQL • PieCloudDB TP • PieCloudDB XP • PieCloudDB

7. 点击 **连通性测试**，如果显示“连通性测试成功”的信息，则说明已与所配置的数据源建立了连接。下图以 PieCloudDB 数据源为例。



8. 点击页面的图标“X”，即可退出配置页面并返回到数据源主页面。

提示：

如果需要修改一个数据源的连接配置，请重复执行上述步骤 5~8。

创建数据源之后，还需要配置表的映射关系，详细信息请参见 [配置表的映射关系](#)。

管理数据源

在数据源列表中，用户可以对目标数据源进行修改（仅限于修改数据源的名称和描述，数据源类型无法更改）和删除操作。这些功能都可以通过数据源列表每行旁边的隐藏菜单「...」来快速访问。

用户也可以点击目标数据源名称以进入数据源详情页面，该页面支持执行修改数据源的基本信息和配置信息等操作。

数据集成 / 数据源

中文 您好: 

状态 正常	输入名称查询	输入描述查询	 DDL 任务	创建数据源
pdb_mysql	pieclouddb	2024年11月20日星期三 14:11	连接配置 关联的表	
pdb_pg_nodeport	pieclouddb	2024年11月19日星期二 15:11	连接配置 关联的表	
pg_test	postgres	2024年11月15日星期五 11:15	连接配置 关联的表	
pdb_up_low	pieclouddb	2024年11月14日星期四 19:48	连接配置 关联的表	
mysql_up_low	mysql	2024年11月14日星期四 19:46	连接配置 关联的表	
pdb_oracle	pieclouddb	2024年11月14日星期四 17:07	连接配置 关联的表	
oracle_stage	oracle	2024年11月14日星期四 17:03	连接配置 关联的表	

数据源信息列表提供如下快捷操作：

- 根据状态筛选信息：使用**状态**控件可以筛选指定状态的数据源信息，包括“正常”和“回收站”两种状态。默认展示“正常”状态的数据源信息。当数据源被删除后，它将被移至回收站。

如果数据源被删除，则会被标记为“回收站”状态，点击**恢复**快捷键即可将其返回到“正常”状态。

数据集成 / 数据源

中文 您好: 

状态 回收站	输入名称查询	输入描述查询	 DDL 任务	创建数据源
demo	oracle	2024年11月21日星期四 15:27	连接配置 关联的表	
pdb_mysql2	mysql	2024年11月20日星期三 14:10	连接配置 关联的表	

- 根据数据源名称搜索信息：使用**输入名称查询**的搜索框可以基于数据源的名称搜索相关信息。
- 根据描述搜索信息：使用**输入描述查询**的搜索框可以基于数据源的描述信息搜索相关信息。

6.5.6 配置表的映射关系

数据源在初始创建状态下不包含任何表，因此需要配置表的映射关系。

在数据集成的「**数据源**」页面，用户点击目标数据源所在行的“操作”栏下的**关联的表**选项，即可进入目标数据源的详情页面。

数据集成 / 数据源

状态 正常

输入名称查询

输入描述查询

C DDL 任务 创建数据源

数据源名称	数据源类型	描述	创建时间	操作
pdntp	pieclouddbtp	local	2024年12月16日星期一 11:15	连接配置 关联的表

在数据源详情页面中，“关联的表”区域会展示与当前数据源相关联的所有通用表（如果存在），并提供配置表映射关系的相关操作功能。

关联的表 [查看全部数据表](#)

输入 schema 搜索

输入 表名搜索

表名	schema	标题	表名是否大小写敏感	操作
交易数据_2020_2023	服装销售数据	服装销售数据.交易数据_2020_2023	否	解除关联 查看字段
顾客数据_2020_2023	顾客数据	顾客数据.顾客数据_2020_2023	否	解除关联 查看字段

提示：

在数据集成的「数据源」页面，用户也可以点击目标数据源名称来进入数据源详情页面。

同步表：从真实数据源获取表信息并与之关联

如果需要从当前实际的数据源中获取表信息并建立关联，可以使用同步表功能，该功能支持整个数据库的同步以及部分表的同步。从当前数据源获取的全量或者增量的表信息会被自动维护到通用表模型中，以方便其他数据源与之关联。

同步表的操作步骤如下：

- 在数据源详情页面中的“关联的表”区域，点击**同步表**即可弹出关联窗口。
- 选择同步数据源中表的方式。
 - 如果需要与当前数据源中全量表相关联，则勾选**整库同步**。
 - 如果需要与当前数据源中指定表相关联，则勾选**选择部分表进行同步**，之后在“要同步的表”下拉列表中选择需要同步的表（至少选择一个表）。
- 点击**完成**。如果关联成功，相关同步信息会显示在同步结果的窗口中。



4. 点击 X，即可关闭同步表的窗口。

上述操作完成后，目标数据源即可与从当前数据源所选的表相关联并将其同步到列表中。

更新表：与真实数据源中数据表的结构做同步更新

从真实数据源获取表信息并同步表的操作仅涉及增量同步表的信息。如果数据源的表结构发生变化，则需要使用更新表的功能来与真实数据源的数据表结构进行同步更新，这一操作支持全量更新和部分更新。

与真实数据源的指定表的结构做同步更新后，最新的表信息会被自动维护到通用表模型，用于构建数据同步任务。

更新表的操作步骤如下：

1. 在数据源详情页面中的“关联的表”区域，点击 **更新表** 即可弹出更新窗口。
2. 选择同步更新数据源中表的方式。
 - 如果需要与当前数据源中全量表的结构做同步更新，则勾选 **更新所有表**。
 - 如果需要与当前数据源中指定表的结构做同步更新，则勾选 **选择部分表进行更新**，之后在“要更新的表”下拉列表中选择要同步结构的表（至少选择一个表）。
3. 点击 **更新**。如果操作成功，相关更新信息会显示在更新结果的窗口中。



4. 点击 X，即可关闭更新表的窗口。

关联表：将通用表模型关联到当前数据源

数据源在初始创建状态下，通常不与任何通用表模型关联。如果数据集成模块中存在可用的通用表模型，用户可以点击 **关联表**，并从下拉框中选择要关联的表。



然后点击 **完成**，即可将选定的通用表与目标数据源关联，并将表信息同步到列表中。

The screenshot shows a table listing associated tables. The 'CUSTOMER' and 'ORDERS' rows are highlighted with a dashed red border. The table has columns: 表名 (Table Name), schema, 标题 (Title), 表名是否大小写敏感 (Case-sensitive), and 操作 (Operations).

表名	schema	标题	表名是否大小写敏感	操作
CUSTOMER		CUSTOMER	否	解除关联 查看字段
ORDERS		ORDERS	否	解除关联 查看字段
交易数据_2020_2023	服装销售数据	服装销售数据.交易数据_2020_2023	否	解除关联 查看字段
顾客数据_2020_2023	顾客数据	顾客数据.顾客数据_2020_2023	否	解除关联 查看字段

解除关联：将当前数据源与指定表解除关联

如果需要将当前数据源与指定表解除关联，在目标表所在行的“操作”列中，点击**解除关联**，确认后即可该表与当前数据源解除关联，同时该表也会从关联列表中移除。



关联的表 查看全部数据表				
表名	schema	标题	表名是否大小写敏感	操作
CUSTOMER		CUSTOMER	否	解除关联 查看字段
ORDERS		ORDERS	否	解除关联 查看字段
交易数据_2020_2023	服装销售数据	服装销售数据.交易数据_2020_2023	否	解除关联 查看字段
顾客数据_2020_2023	顾客数据	顾客数据.顾客数据_2020_2023	否	解除关联 查看字段

查看字段信息

在目标数据源详情页面的“关联的表”列表中，在目标表所在行的“操作”列中，点击**查看字段**，即可查看该表的所有字段信息。



关联的表 查看全部数据表				
表名	schema	标题	表名是否大小写敏感	操作
CUSTOMER		CUSTOMER	否	解除关联 查看字段
ORDERS		ORDERS	否	解除关联 查看字段
交易数据_2020_2023	服装销售数据	服装销售数据.交易数据_2020_2023	否	解除关联 查看字段
顾客数据_2020_2023	顾客数据	顾客数据.顾客数据_2020_2023	否	解除关联 查看字段

6.5.7 创建和执行 DDL 任务

在数据仓库或数据库迁移过程中，DDL 任务可以将源数据源的表结构迁移到目标数据源下并创建新的表。

提示：

请确保在创建和执行 DDL 任务之前，源数据源的连接配置和表的映射关系均已正确设置。

在数据集成的「**数据源**」页面，点击**DDL 任务**即可进入相应的功能页面。

在「**DDL 任务**」页面，创建 DDL 任务的步骤如下：

1. 点击**创建任务**即可进入创建任务页面。
2. 在“源数据源”下拉列表中选择源数据源，并根据实际需要填写 Schema。如果没有指定 Schema，则系统会迁移源数据源的默认 Schema 下的表。

注意：

对于源数据源，在多个 Schema 的场景下建议指定 Schema。

- 在“目标数据源”下拉列表中选择目标数据源，并根据实际需要填写 Schema。如果没有指定 Schema，则会保存在目标数据源的默认 Schema 下。

注意：

对于目标数据源，在多个 Schema 的场景下需要指定 Schema。

- 点击**添加**，分别添加源数据源中需要迁移的表和需要排除的表。
- (可选) 设置执行该任务的并行数。默认情况下，使用系统预定义的并行数。
- (可选) 根据实际使用需求开启相关任务配置。在默认情况下，如下功能选项是关闭的。
 - 在提交后立刻执行：如果没有开启该选项，则 DDL 任务在创建后为“初始”状态并不会执行。
 - 若表不存在，则创建表：如果在目标数据源中找不到与源数据源同名的表，则系统将在指定的 Schema 下自动创建一个具有相同结构的表。
 - 若表已存在，则删除表重新创建：如果在目标数据源中存在与源数据源同名的表，则系统将删除该表并重新创建一个具有相同结构的表。
 - 是否同步索引：在执行 DDL 任务时同步源数据源中表的索引。
 - 是否同步主键：在执行 DDL 任务时同步源数据源中表的主键。
- 点击**完成**。DDL 任务创建成功后，其信息会同步到列表中。

如果在步骤 6 中开启“在提交后立刻执行”，则该 DDL 任务会在创建成功后就开始执行，否则，需要用户点击操作栏的隐藏菜单「...」下的**执行**选项来执行该任务。

源数据源	源数据源Schema	目标数据源	目标数据源Schema	状态	执行信息	执行时间	表的同步情况	操作
pdb-pdb	pdb-np			● 初始		开始时间： 结束时间：	成功： 0 失败： 0	查看
oracle_stage	pdb_oracle			● 已完成	finished	开始时间：2024-11-21 19:05:56 结束时间：2024-11-21 19:06:04	成功： 19 失败： 0	查看
mysql	pdb_mysql			● 已完成	finished	开始时间：2024-11-20 15:33:06 结束时间：2024-11-20 15:33:11	成功： 13 失败： 0	查看

在 DDL 任务执行成功后，用户点击目标 DDL 任务所在行的“表的同步情况”栏下的**查看**，即可查看已成功同步的表和执行 SQL 明细等信息，以及同步失败的表信息。

表的同步情况				
表名	状态	执行信息	SQL 明细	
TB	● 成功	success		查看
lower_tb	● 成功	success		查看
UP_TB	● 失败	pq: relation "up_tb" already exists		无
TEST_NUMERIC	● 失败	pq: relation "test_numeric" already exists		无
PT_LIST_TEST	● 成功	success		查看
PART_HASH_T1	● 成功	success		查看
MID_tb	● 成功	success		查看
TRANS_SMT_ATP_T_REPORT	● 失败	pq: relation "trans_smt_atp_t_report" already exists		无
TEST_NUMERIC_2	● 失败	pq: relation "test_numeric_2" already exists		无
PT_RANGE_TEST1	● 成功	success		查看

DDL 任务列表显示已创建的 DDL 任务信息，包括源数据源、目标数据源、表的同步情况等。

数据集成 / 数据源 / DDL 任务							
源数据源	源数据源Schema	目标数据源	目标数据源Schema	状态	执行信息	执行时间	表的同步情况
oracle_stage		pdb_oracle		● 已完成	finished	开始时间: 2024-11-21 19:05:56 结束时间: 2024-11-21 19:06:04	成功: 19 查看 失败: 0 查看
mysql		pdb_mysql		● 已完成	finished	开始时间: 2024-11-20 15:33:06 结束时间: 2024-11-20 15:33:11	成功: 13 查看 失败: 0 查看
mysql		pdb_mysql		● 已完成	finished	开始时间: 2024-11-20 14:06:38 结束时间: 2024-11-20 14:06:42	成功: 12 查看 失败: 0 查看
pg12		pdb_pg		● 已完成	finished	开始时间: 2024-11-20 11:45:43 结束时间: 2024-11-20 11:45:48	成功: 11 查看 失败: 0 查看

用户也可以点击目标 DDL 任务的操作栏的隐藏菜单「...」下的 **修改** 或者 **删除** 选项来执行相应的操作。

需要注意的是，在 DDL 任务执行完成后，目标数据源也需要配置表的映射关系，详细操作信息请参见 [配置表的映射关系-同步表](#)。

6.5.8 创建离线同步任务

离线同步适用于批处理场景，支持以下类型的数据同步：

- 单表的全量同步：针对单个表进行完整的数据同步。
- 整库批量表全量同步：对整个数据库中的多个表进行批量的全量数据同步。
- 数据库指定表列表的全量同步：根据指定的表列表，对选定的表进行全量数据同步。

创建表到表的离线同步任务

在数据集成的「**数据同步**」页面，用户可以根据业务需求分别设置数据关系、调度配置、过滤规则和运行配置以创建表到表的离线同步任务。参考步骤如下：

1. 切换到「**离线同步**」页面，点击 **新建离线同步** 即可进入创建页面。
2. 输入该离线同步任务的名称。
3. 从“类型”下拉列表中选择离线同步的类型为“表到表”。
- 4.（可选）输入对该任务的描述。
5. 点击 **下一步** 以进入离线同步任务的配置详情页面。



6. 设置“数据关系”以定义数据流向。

在「**数据关系**」页面，首先点击 **编辑**，然后分别选择源数据表和目标数据表。完成选择后，点击 **保存** 以确认更改。

7. 添加调度器配置。

切换到「**调度配置**」页面，点击 **编辑** 并执行如下操作：

1. 在“调度地址”下拉列表中选择运行该数据流的调度器。

有关配置调度器的详细信息，请参见 [配置调度系统-添加调度器配置](#)。

提示：

通常情况下，如果没有特殊环境需求，设置一个调度地址即可。

2. 在“工作流名称”下拉框中选择“PyFileBatchWorkflow”。工作流的作用是将作业任务按照指定顺序组织起来。
3. 在“任务队列”下拉框中选择“PYTHON_FILE_BATCH_QUEUE”。任务队列是轻量级的、动态分配的队列，用于轮询任务。

注意：

系统初始化时应该已经配置了调度器信息。在当前版本的系统中，调度器的工作流名称和任务队列都是固定配置。在选择调度地址后，系统会自动协助用户完成选择。如果在使用过程中发现缺少这类信息，请联系 OpenPie 技术支持团队获取帮助。

4. 设置任务执行周期为“立即执行”或者“周期”。

“立即执行”表示只会创建一个任务并执行一次；“周期”表示会按照周期表达式（Cron 表达式）的设置来创建周期任务并定时执行。

对于周期性的任务，可以通过输入 cron 表达式来定义执行的周期。点击“？”可以显示 cron 表达式的格式说明。例如，“*/60 20 1-30 1-12 6”表示“对于 1 月到 12 月，每月的 1 号到 30 号，每个星期六晚上八点，每隔 60 分钟执行一次作业”。

5. 点击保存**以确认更改。****8. (可选) 添加过滤规则。**

切换到「**过滤规则**」页面，首先点击**编辑**，之后输入 where 条件用于过滤源数据表和目标数据表中的数据。

过滤规则分为如下三类：

- 添加普通条件：需要输入字段名和字段值，并指定逻辑关系。例如“l_quantity > 100”。
- 添加时间范围条件：需要输入字段名和满足时间范围（年/月）的字段值。例如，字段“l_shipdate”满足时间范围过去“12 月”。
- 自定义过滤条件：点击**切换为自定义**，分别输入源表和目标表的自定义 where 表达式（请注意，不要在表达式末尾加分号“;”）。例如，“where update_time > '2024-01-01'"。

在所有过滤条件添加完成后，点击**保存**以确认更改。

9. 修改运行配置。

“运行配置”展示数据同步任务的系统信息。由于离线同步任务需要做数据校验和文件导出的场景，由此运行配置包括比较算法、差异行数量限制、S3 桶信息、导出文件格式等配置项。

切换到「**运行配置**」页面，点击**修改**，并执行如下操作：

- 在运行配置的编辑页面修改 `s3_config_id` 字段以设置 S3 连接配置信息，之后点击 **完成**。对于其他配置项，可以保持它们的默认设置不变。

提示：

对于批处理类型的导出文件，当前版本默认将其存储路径设置为 S3。用户可以通过数据集成的「总览」页面来添加 S3 配置信息。一旦配置完成，用户就可以在运行配置中通过 `s3_config_id` 字段选择对应的 S3 配置 ID。如果需要指定除默认配置之外的存储桶，用户还可以通过配置 `s3_config_bucket` 字段来指定 S3 存储桶，从而覆盖默认的存储桶信息配置。

- 有关运行配置项的详细信息，请参见 [离线同步任务运行配置项说明](#)。
- 运行配置修改完成后，点击 **X** 即可返回运行配置页面。
 - 点击 **测试 S3 连接**，如果配置信息无误，系统会显示“连接测试成功”的信息。
 - 在当前离线同步任务的配置详情页面，点击 **运行** 即可开始运行表到表的离线同步任务。



如果在点击 **运行** 后就发生配置相关报错，用户可以点击 **返回配置** 并根据报错信息修改相应的配置，之后可以重新运行该任务。在任务运行期间，如果需要强制停止运行该任务，点击 **终止** 即可，相应的调度任务也会被同步删除。

提示：

如果终止了一个正在运行的任务，它可能会在数据导出导入过程中的任意点停止。如果需要在终止后重新启动任务，建议先检查目标表，确认数据是否需要被清空，然后重新创建并启动任务。

在表到表的离线同步任务的详情页面，运行结果区域会动态展示离线同步任务在 `ExportTable`、`ImportTable` 和 `TableDataDiff` 各个阶段的执行结果、响应消息、执行时间，以及是否执行成功。



在离线同步任务执行完成后，用户可以点击 **查看调度详情** 来查看执行该任务时详细的调度情况。

数据集成 / 同步详情 / 调度详情 / 1c8773ac-1434-4738-9613-0c4edf94e495

1c8773ac-1434-4738-9613-0c4edf94e495 - Running

时间	事件类型	事件描述	更多信息
47 2024-11-25 17:20:18.619	ActivityTaskScheduled	Activity Type export_table_task	更多信息
46 2024-11-25 17:20:18.619	WorkflowTaskCompleted	Scheduled Event ID 44	更多信息
45 2024-11-25 17:20:18.608	WorkflowTaskStarted	Scheduled Event ID 44	更多信息
44 2024-11-25 17:20:18.602	WorkflowTaskScheduled	Task Queue Name 8@dataflow-pyworker-857c99c44f-tlj4p-91c55b04... 更多信息	更多信息
43 2024-11-25 17:20:18.602	ActivityTaskCompleted	Result {"heartbeat_timeout":60}	更多信息

对于周期性的表到表的离线同步任务，要查看当前的调度信息，只需切换到「**调度配置**」页面，并点击图标 **>>**，即可显示详细的调度信息。

调度配置

工作流名称: PyFileBatchWorkflow
任务队列: PYTHON_FILE_BATCH_QUEUE
调度地址: gx_scheduler(temporal-frontend.openpie-infra:7233)
周期设置: 周期
任务执行周期: */60 20 1-30 1-12 6

当前调度

调度状态	创建时间	最近执行时间	下次执行时间
运行中 暂停调度	2024-11-25 16:43:01		2024-11-30 20:00:00 2024-12-07 20:00:00 2024-12-14 20:00:00 2024-12-21 20:00:00 2024-12-28 20:00:00 2025-01-04 20:00:00 2025-01-11 20:00:00 2025-01-18 20:00:00 2025-01-25 20:00:00 2025-02-01 20:00:00

对于周期性的表到表的离线同步任务，如果需要暂停未来的调度周期，只需点击 **暂停调度**，之后可以随时重启暂停的调度。请注意，暂停操作并不会影响当前正在运行的任务。如果决定取消未来的所有调度周期，可以点击 **删除**。这一操作同样不会删除任何正在执行的任务的调度信息。

创建库到库的离线同步任务

在数据集成的「**数据同步**」页面，用户可以根据业务需求分别设置数据关系、调度配置、过滤规则和运行配置以创建库到库的离线同步任务。参考步骤如下：

1. 切换到「**离线同步**」页面，点击 **新建离线同步** 即可进入创建页面。
2. 输入该离线同步任务的名称。
3. 从“类型”下拉列表中选择离线同步的类型为“库到库”。
- 4.（可选）输入对该任务的描述。
5. 点击 **下一步** 以进入离线同步任务的配置详情页面。



6. 设置“数据关系”以定义数据流向。

在「**数据关系**」页面，点击 **编辑** 并执行如下操作：

1. 分别选择源数据源和目标数据源。
2. 选择数据复制方式为“整库复制”或者“多表复制”。

提示:

整库复制方式仅适用于源数据库和目标数据库中表名完全相同的情况。如果表名不一致，请选择多表复制方式。

如果选择“多表复制”，点击 **添加表** 并根据需要选择源表和对应的目标表。

3. 点击 **保存** 以确认更改。
7. 添加调度器配置。

切换到「**调度配置**」页面，点击 **编辑** 并执行如下操作：

1. 在“调度地址”下拉列表中选择运行该数据流的调度器。

有关配置调度器的详细信息，请参见 [配置调度系统-添加调度器配置](#)。

提示:

通常情况下，如果没有特殊环境需求，设置一个调度地址即可。

2. 在“工作流名称”下拉框中选择“PyFileBatchWorkflow”。工作流的作用是将作业任务按照指定顺序组织起来。
3. 在“任务队列”下拉框中选择“PYTHON_FILE_BATCH_QUEUE”。任务队列是轻量级的、动态分配的队列，用于轮询任务。

注意：

系统初始化时应该已经配置了调度器信息。在当前版本的系统中，调度器的工作流名称和任务队列都是固定配置。在选择调度地址后，系统会自动协助用户完成选择。如果在使用过程中发现缺少这类信息，请联系 OpenPie 技术支持团队获取帮助。

4. 设置任务执行周期。库到库的离线同步仅支持“立即执行”，即只会创建一个任务并执行一次。
5. 点击**保存**以确认更改。
8. (可选) 添加过滤规则。

切换到「**过滤规则**」页面，首先点击**编辑**，之后输入 where 条件以用于过滤源数据表和目标数据表中的数据。

过滤规则分为如下三类：

- 添加普通条件：需要输入字段名和字段值，并指定逻辑关系。例如“l_quantity > 100”。
- 添加时间范围条件：需要输入字段名和满足时间范围（年/月）的字段值。例如，字段“l_shipdate”满足时间范围过去“12 月”。
- 自定义过滤条件：点击**切换为自定义**，分别输入源表和目标表的自定义 where 表达式（请注意，不要在表达式末尾加分号“;”）。例如，“where update_time > '2024-01-01'"。

在所有过滤条件都添加完成后，点击**保存**以确认更改。

9. 修改运行配置。

“运行配置”展示数据同步任务的系统信息。由于离线同步任务需要做数据校验和文件导出的场景，由此运行配置包括比较算法、差异行数量限制、S3 桶信息、导出文件格式等配置项。

切换到「**运行配置**」页面，点击**修改**，并执行如下操作：

1. 在运行配置的编辑页面修改 `s3_config_id` 字段以设置 S3 连接配置信息，之后点击**完成**。对于其他配置项，可以保持它们的默认设置不变。

提示：

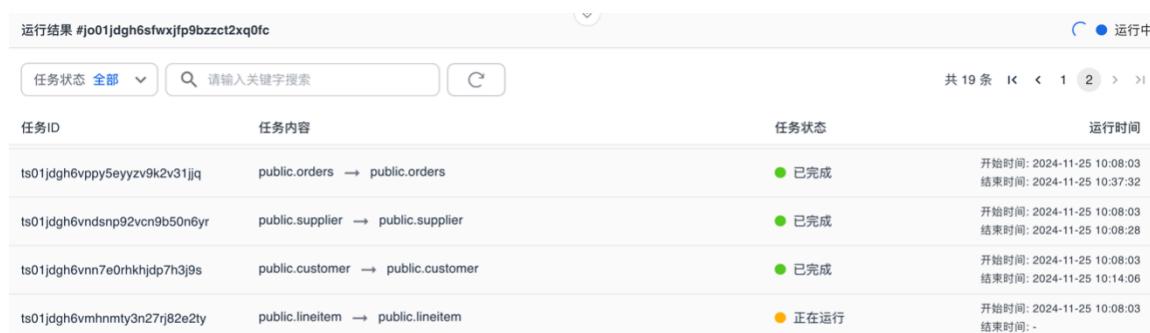
对于批处理类型的导出文件，当前版本默认将其存储路径设置为 S3。用户可以通过数据集成的「总览」页面来添加 S3 配置信息。一旦配置完成，用户就可以在运行配置中通过 `s3_config_id` 字段选择对应的 S3 配置 ID。如果需要指定除默认配置之外的存储桶，用户还可以通过配置 `s3_config_bucket` 字段来指定 S3 存储桶，从而覆盖默认的存储桶信息配置。

有关运行配置项的详细信息，请参见 [离线同步任务运行配置项说明](#)。

2. 运行配置修改完成后，点击 **X** 即可返回运行配置页面。
3. 点击 **测试 S3 连接**，如果配置信息无误，系统会显示“连接测试成功”的信息。
10. 在当前离线同步任务的配置详情页面，点击 **运行** 即可开始运行库到库的离线同步任务。

与表到表的离线同步任务不同，库到库的离线同步任务一般会以表为单位拆分成多个子任务，子任务会自动同步到任务列表中并实时显示子任务的执行状态，可能为如下之一：

- 已创建
- 正在运行
- 已完成
- 正在终止
- 正在重启
- 已终止
- 失败
- 数据不一致

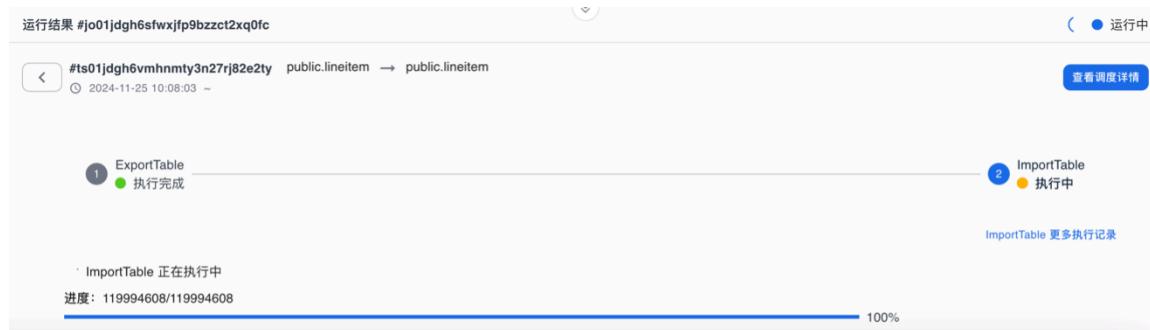


The screenshot shows a table of tasks with the following data:

任务ID	任务内容	任务状态	运行时间
ts01jdgh6vppp5eyyzv9k2v31jjq	public.orders → public.orders	● 已完成	开始时间: 2024-11-25 10:08:03 结束时间: 2024-11-25 10:37:32
ts01jdgh6vndsn92vcn9b50n6yr	public.supplier → public.supplier	● 已完成	开始时间: 2024-11-25 10:08:03 结束时间: 2024-11-25 10:08:28
ts01jdgh6vnn7e0rhkhjp7h3j9s	public.customer → public.customer	● 已完成	开始时间: 2024-11-25 10:08:03 结束时间: 2024-11-25 10:14:06
ts01jdgh6vmhnmt3n27rj82e2ty	public.lineitem → public.lineitem	● 正在运行	开始时间: 2024-11-25 10:08:03 结束时间: -

用户选中库到库的离线同步任务的子任务并单击，即可进入目标子任务的执行详情页面。

对于正在运行中的离线同步子任务，运行结果区域会动态展示离线同步任务在 ExportTable、ImportTable 和 TableDataDiff 各个阶段的执行进度。



与表到表的离线同步任务相同，各个子任务的运行结果区域会展示该任务在 ExportTable、ImportTable 和 TableDataDiff 各个阶段的执行结果、响应消息、执行时间，以及是否执行成功。

在离线同步任务执行完成后，用户可以点击 **查看调度详情** 来查看执行该任务时详细的调度情况。

离线同步任务运行配置项说明

配置项	说明	默认值
key_columns	表的主键	无
extra_columns	用于比较的非主键列	无
algorithm	比较算法。取值范围: <ul style="list-style-type: none"> md5diff: 直接计算 MD5 值，且只能确定文件或数据之间是否存在差异。 hashdiff: 递归计算 md5 值，并找出具体不同的行。 countdiff: 通过计算记录数量判断是否有差异。 	md5diff
limit	最多能返回的有差异的行的数量。取值范围 [1,1000]。	1000
threads	执行比较的线程数量	1
bisection_threshold	在 diff 操作期间每个段中的行数阈值。如果段内的行数低于此阈值，则将在内存中直接进行比较。	无
bisection_factor	在每次比较迭代中数据集被划分为校验和片段的数量。该数值应小于 bisection_threshold。	无
verbose	校验时是否输出详细日志	false
s3_config_id	s3 连接配置信息	无
s3_config_bucket	s3 桶的名称。如果为空，则使用默认值。	空

配置项	说明	默认值
mssql_server_name	mssql 服务的名称，如果用 Parquet 就是必填项。	无
export_batch_rows	导出文件被分割成多个文件，每个文件的行数。	100000
batch_parallel_size	批作业并行运行子作业的个数	5
batch_file_format	批作业的文件格式，支持 csv 和 parquet 两种。	csv
heartbeat_timeout	activity 的心跳超时时间。单位：秒。	60
export_compress_parallel	压缩上传任务最大并行数量	5
enable_pk_duplicate_detect	是否开启检测主键重复。开启后，在导入数据前会根据主键进行查重。	True
source_timezone	源库时区。连接源数据库和目标数据库时都使用该时区。	Asia/Shanghai
import_order_by	对于大表，将该字段设置为常用于查询过滤条件的字段（例如日期），可以显著提高数据库查询对数据文件的过滤效率，从而提升查询性能。语法和 SQL 的 ORDER BY 一致。	无

离线同步任务失败问题排查

在执行同步任务过程中，可能遇到的执行失败的情况，在任务详情页面的运行结果区域，用户可以通过查看任务的状态和执行详情来了解失败的原因。

通常情况下，库到库的离线同步任务很少需要再次执行，除非失败原因是由于资源、环境等不可控因素引起的，在这种情况下，任务可以被重新执行。

如果任务执行失败是由于系统稳定性、数据库状态等服务质量问题引起的，用户可以在任务详情页面的运行结果区域，找到当前执行的任务记录以及执行失败的具体步骤，并点击该步骤下方的 **重试** 来重新执行该步骤。

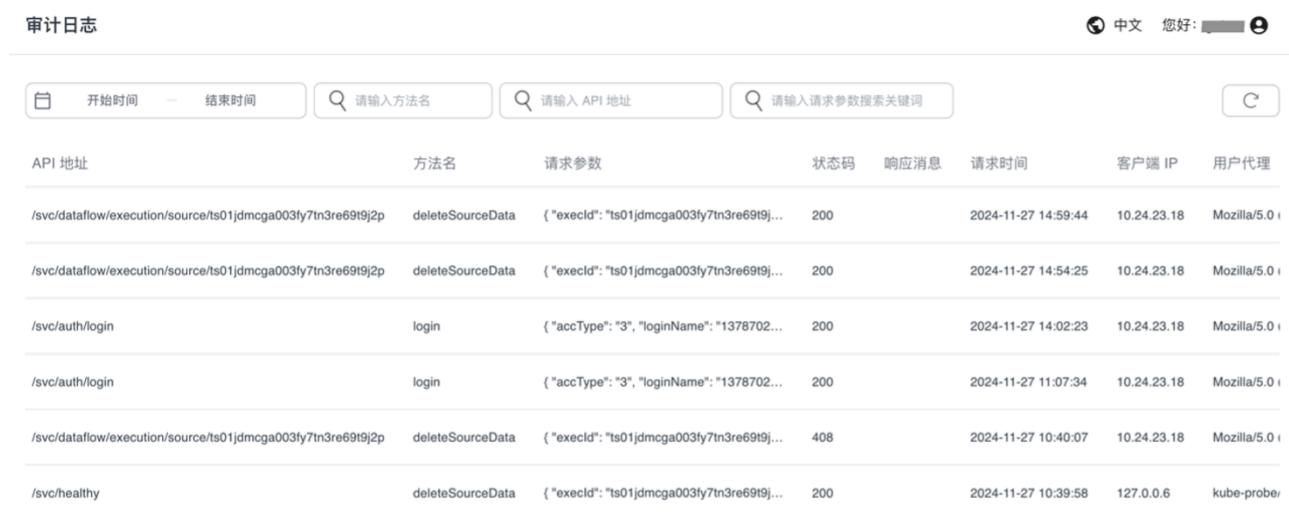


如果任务失败是由于配置问题引起的，那么需要重新修改任务配置。用户可以点击 **返回配置** 并根据报错信息修改相应的配置，之后可以重新运行该任务。

在执行同步任务时，通常不会出现数据不一致的情况。目前，数据不一致的最主要原因是重复导入数据，而次要原因可能是数据导出、导入或数据校验流程中出现了问题。

如果任务状态显示“数据不一致”，建议首先清理目标数据源的数据，然后确认环境配置等的正确性，最后重新执行任务。如果执行上述操作后数据仍然不一致，请联系 OpenPie 技术支持团队。

此外，用户可以通过审计日志来了解任务的方法调用情况。在 PieDataCS 云原生平台的数仓操作界面，点击菜单栏「审计日志」即可进入该功能页面。



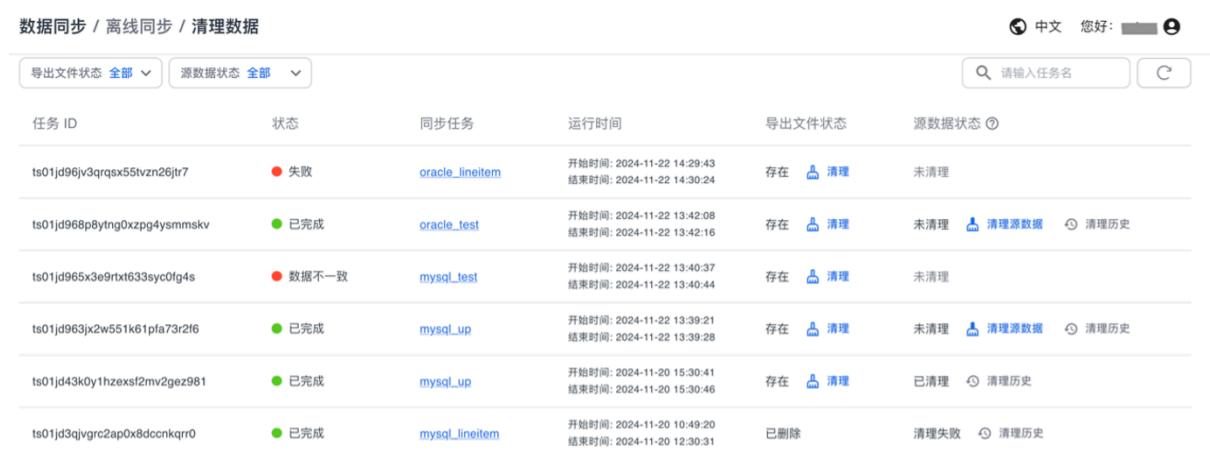
The screenshot shows a table of audit log entries. The columns are: API 地址 (API Address), 方法名 (Method Name), 请求参数 (Request Parameters), 状态码 (Status Code), 响应消息 (Response Message), 请求时间 (Request Time), 客户端 IP (Client IP), and 用户代理 (User Agent). The data includes various API calls like /svc/auth/login and /svc/dataflow/execution/source, with status codes ranging from 200 to 408.

API 地址	方法名	请求参数	状态码	响应消息	请求时间	客户端 IP	用户代理
/svc/dataflow/execution/source/ts01jdmcg003fy7tn3re69t9j2p	deleteSourceData	{ "execId": "ts01jdmcg003fy7tn3re69t9j..." }	200		2024-11-27 14:59:44	10.24.23.18	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/115.0.5790.102 Safari/537.36
/svc/dataflow/execution/source/ts01jdmcg003fy7tn3re69t9j2p	deleteSourceData	{ "execId": "ts01jdmcg003fy7tn3re69t9j..." }	200		2024-11-27 14:54:25	10.24.23.18	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/115.0.5790.102 Safari/537.36
/svc/auth/login	login	{ "accType": "3", "loginName": "1378702..." }	200		2024-11-27 14:02:23	10.24.23.18	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/115.0.5790.102 Safari/537.36
/svc/auth/login	login	{ "accType": "3", "loginName": "1378702..." }	200		2024-11-27 11:07:34	10.24.23.18	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/115.0.5790.102 Safari/537.36
/svc/dataflow/execution/source/ts01jdmcg003fy7tn3re69t9j2p	deleteSourceData	{ "execId": "ts01jdmcg003fy7tn3re69t9j..." }	408		2024-11-27 10:40:07	10.24.23.18	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/115.0.5790.102 Safari/537.36
/svc/healthy	deleteSourceData	{ "execId": "ts01jdmcg003fy7tn3re69t9j..." }	200		2024-11-27 10:39:58	127.0.0.6	kube-probe

清理源数据

对于离线同步任务，用户可以根据设定的周期和 where 条件来实现定时定量的数据迁移。数据迁移完成后，如果用户希望清理源数据库的表中的特定数据，可以通过执行清理源数据的操作来实现。

在数据集成的「数据同步」页面，用户点击 **数据清理** 即可进入清理数据的功能页面。该页面以列表形式展示了当前所有的离线同步任务。



The screenshot shows a table of offline sync tasks. The columns are: 任务 ID (Task ID), 状态 (Status), 同步任务 (Sync Task), 运行时间 (Run Time), 导出文件状态 (Export File Status), and 源数据状态 (Source Data Status). The data includes tasks like oracle_lineitem, oracle_test, mysql_test, mysql_up, and mysql_lineitem, each with its own run details and source data status.

任务 ID	状态	同步任务	运行时间	导出文件状态	源数据状态
ts01jd96jv3qrqxsx55tvzn26jlr7	失败	oracle_lineitem	开始时间: 2024-11-22 14:29:43 结束时间: 2024-11-22 14:30:24	存在 清理	未清理
ts01jd96p8ytng0xpg4ysmmskv	已完成	oracle_test	开始时间: 2024-11-22 13:42:08 结束时间: 2024-11-22 13:42:16	存在 清理	未清理 清理源数据 清理历史
ts01jd965x3e9rtxt633sy0fg4s	数据不一致	mysql_test	开始时间: 2024-11-22 13:40:37 结束时间: 2024-11-22 13:40:44	存在 清理	未清理
ts01jd963jx2w551k61pfa73r2f6	已完成	mysql_up	开始时间: 2024-11-22 13:39:21 结束时间: 2024-11-22 13:39:28	存在 清理	未清理 清理源数据 清理历史
ts01jd43k0y1hzexsf2mv2gez981	已完成	mysql_up	开始时间: 2024-11-20 15:30:41 结束时间: 2024-11-20 15:30:46	存在 清理	已清理 清理历史
ts01jd3qvgrc2ap0x8dccnkqrr0	已完成	mysql_lineitem	开始时间: 2024-11-20 10:49:20 结束时间: 2024-11-20 12:30:31	已删除	清理失败 清理历史

在「清理数据」页下的离线任务列表中的目标任务的“源数据状态”栏下，用户可以点击 **清理源数据**，并基于 where 条件来按需清理目标任务源数据，如下图所示。



对于数据量较大的场景，为了保证源数据库的稳定性，可以实现源数据的分批多轮自动清理。默认情况下，每轮删除的数据量为 10000 条，但用户也可以输入自定义的整数值来设定每轮删除的数据量。

如果未设置 where 条件，源数据清理操作将会清空整张表的所有数据。

源数据清理完成后，系统会显示“已清理”。如果用户执行了清理数据源的操作，可以点击 **查看清理历史** 来查看历史清理信息，其中包括调度器的 IP 地址、执行的 SQL 语句、当前用户名、清理时间以及本次清理所删除的记录条数等详细信息。

源数据清理历史						
ip	状态	sql	用户	清理时间	删除记录条数	更多信息
【遮挡】	● 已清理	DELETE FROM apitest_mysql_table_1	【遮挡】	2024-06-04 18:18:51	1	更多

清理导出文件

对于离线同步任务，数据迁移完成后或在迁移过程中，如果通过数据校验失败等方式发现了导出数据的问题，用户可以清理这些有问题的导出文件。

在数据集成的「**数据同步**」页面，用户点击 **数据清理** 即可进入清理数据的功能页面。该页面以列表形式展示了当前所有的离线同步任务。

数据同步 / 离线同步 / 清理数据

任务 ID	状态	同步任务	运行时间	导出文件状态	源数据状态
ts01jd96jv3qrqxsx55tvzn26jtr7	失败	oracle_lineitem	开始时间: 2024-11-22 14:29:43 结束时间: 2024-11-22 14:30:24	存在 清理	未清理
ts01jd968p8ytng0xzpg4ysmmskv	已完成	oracle_test	开始时间: 2024-11-22 13:42:08 结束时间: 2024-11-22 13:42:16	存在 清理	未清理 清理源数据 清理历史
ts01jd965x3e9rtxt633sycofg4s	数据不一致	mysql_test	开始时间: 2024-11-22 13:40:37 结束时间: 2024-11-22 13:40:44	存在 清理	未清理
ts01jd963jx2w551k61pfa73r2f6	已完成	mysql_up	开始时间: 2024-11-22 13:39:21 结束时间: 2024-11-22 13:39:28	存在 清理	未清理 清理源数据 清理历史
ts01jd43k0y1hzexsf2mv2gez981	已完成	mysql_up	开始时间: 2024-11-20 15:30:41 结束时间: 2024-11-20 15:30:46	存在 清理	已清理 清理历史
ts01jd3qjvgrc2ap0x8dcnknqr0	已完成	mysql_lineitem	开始时间: 2024-11-20 10:49:20 结束时间: 2024-11-20 12:30:31	已删除	清理失败 清理历史

在「清理数据」页下的离线任务列表中的目标任务的“导出文件状态”栏下，用户可以点击清理，确认后即可清理目标导出文件。

任务 ID	状态	同步任务	运行时间	导出文件状态	源数据状态
ts01jd96jv3qrqxsx55tvzn26jtr7	失败	oracle_lineitem	开始时间: 2024-11-22 14:29:43 结束时间: 2024-11-22 14:30:24	存在 清理	未清理
ts01jd968p8ytng0xzpg4ysmmskv	已完成	oracle_test	开始时间: 2024-11-22 13:42:08 结束时间: 2024-11-22 13:42:16	未清理 清理源数据 清理历史	
ts01jd965x3e9rtxt633sycofg4s	数据不一致	mysql_test	开始时间: 2024-11-22 13:40:37 结束时间: 2024-11-22 13:40:44	未清理	

文件清理成功后，系统会显示“已删除”。如果导出文件状态栏显示为“不存在”，则表明该任务没有生成导出文件。

6.5.9 创建实时同步任务

实时同步任务实现了利用 PieCloudDB 的 Flink 动态执行器中间件与 Flink 实时流计算引擎进行对接，进而将计算结果数据写入 PieCloudDB 的虚拟数据仓库中。

实时同步适用于流处理场景，支持以下数据同步类型：

- 单表全量与 CDC (Change Data Capture) 增量同步
- 整库全量与 CDC 增量同步
- 整库指定表的全量与 CDC 增量同步

创建表到表的实时同步任务

在数据集成的「数据同步」页面，用户可以根据业务需求分别设置数据关系、机器配置和运行配置以创建表到表的实时同步任务。参考步骤如下：

- 切换到「实时同步」页面，点击 新建实时同步 即可进入创建页面。
- 输入该实时同步任务的名称。
- 从“类型”下拉列表中选择同步类型为“表到表”。

4. (可选) 输入对该任务的描述。
5. 点击 **下一步** 以进入该实时同步任务的配置详情页面。



6. 设置“数据关系”以定义数据流向。

由于实时同步任务是用于处理实时数据的，为了避免在实时处理过程中因源端和目标端的表结构变化而导致的问题，用户需要手动配置字段之间的映射关系。

注意:

- 当前版本的实时表同步功能仅支持单个源表到单个目标表的同步，尚不支持涉及多个源表或目标表的同步。关于此类定制化需求，请联系 OpenPie 技术支持团队。
- 为保证数据一致性和完整性，在进行实时同步时，源表必须包含主键。主键能够唯一标识每条记录，防止因故障导致数据丢失或重复。如果进行部分字段关联而未将源表的主键与目标表的对应字段关联，系统将自动寻找与源表主键字段名相同的目标字段以进行关联和同步。如果未找到相应的目标字段，同步将失败。

在「**数据关系**」页面，点击 **编辑** 并执行如下操作：

1. 点击 **添加源表** 选择源数据源类型、源数据源和源表，点击 **确定** 后，系统将自动将其添加到编辑区，用于配置映射关系。

注意:

对于源数据源，当前仅支持 MySQL、PostgreSQL 和 PieCloudDB TP。

用户也可以输入源表的别名，该选项同 SQL 中的 `AS` 语法，例如 “`table1 AS t1`”，则表别名可输入 `t1`。

2. 点击 **添加目标表** 选择目标数据源类型、目标数据源和目标表，点击 **确定** 后，系统将自动将其添加到编辑区，用于配置映射关系。

注意:

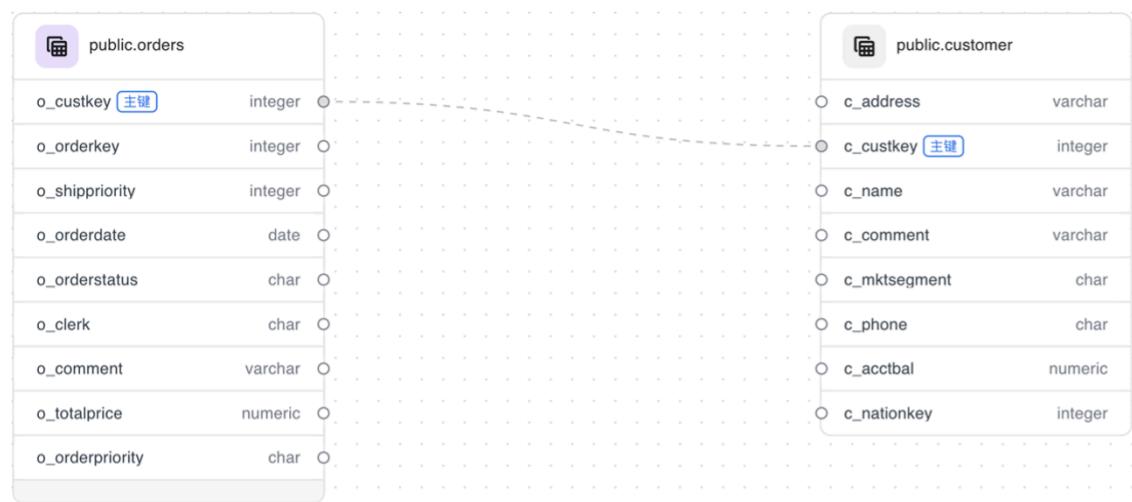
对于目标数据源的类型，当前仅支持 PieCloudDB。而且目标表不支持多任务共享同步。

用户也可以输入目标表的别名，该选项同 SQL 中的 `AS` 语法，例如“`table1 AS t1`”，则表别名可输入 `t1`。

3. 在编辑区域配置源表和目标表之间的字段关联。**提示:**

在配置数据关系时，如果进行的是整表实时同步，并且源表与目标表的表结构（即字段）完全一致，那么就无需手动进行字段关联。反之，如果源表和目标表的表结构不完全一致，或者只同步部分字段，那么就需要手动连接相关字段以建立关联。

选择指定源表（或目标表）的字段从原点进行拖拽到目标表（或源表）的对应字段生成关系线，多个字段映射可以重复关联操作，直到源表和目标表及其字段之间的所有关系都配置完成。如下图所示。

**7. 修改机器配置。**

切换到「**机器配置**」页面，点击**编辑**以修改机器配置。如下机器配置设置完成后，点击**保存**以确认更改。

机器配置项说明如下：

- Job Manager CPU 核数：Job Manager 所能使用的 CPU 总数。默认值：1 Core。

- Job Manager CPU 内存：Job Manager 的总进程内存大小。默认值：1024 Mi。
- Task Manager CPU 核数：Task Manager 所能使用的 CPU 总数。默认值：1 Core。
- Task Manager CPU 内存：Task Manager 的总进程内存大小。默认值：1024 Mi。
- Task Manager Slot：Task Manager 的任务槽数量。默认值：16。

其中，Job Manager 是实时同步作业的管理中心，负责接收用户提交的作业，并进行作业的调度和管理；Task Manager 是实际负责执行计算的 Worker，在其上执行实时同步作业的一组 Task；Task Manager Slot 是 Task Manager 中的最小资源调度单位，用于分配和托管任务的内存空间。

注意：

为避免资源浪费，当 Task Manager Slot 数量超过作业执行的默认并行度（parallelism.default）时，系统会自动将 Slot 数量调整为与并行度相等。也就是说，仅当并行度大于 Slot 数量时，系统会按照用户实际设定的 Slot 数量执行作业；反之，如果并行度小于 Slot 数量，系统会按照并行度执行作业。

8. 修改运行配置。

“运行配置”展示数据同步任务的系统信息。对于实时同步任务，运行配置包括 flink checkpoint 触发间隔、超时时长、重启策略、作业并行度等配置项。

切换到「**运行配置**」页面，首先点击 **修改** 以根据实际需求修改相应的配置信息，之后点击 **完成**。

有关运行配置项的详细信息，请参见 **实时同步任务运行配置项说明**。

点击 **X** 即可退出并返回配置页面。

9. 在实时同步任务的详情页面面，点击 **运行** 即可开始运行表到表的实时同步任务。

The screenshot shows the configuration interface for a task named 'tp-pdb #3'. At the top, there are buttons for '表到表' (Table to Table), '数据关系' (Data Relationship), '机器配置' (Machine Configuration), and '运行配置' (Run Configuration). The status bar indicates the task is '运行中' (Running). On the left, a table from 'pieclouddbt' is shown with columns: 'c_acctbal' (numeric), 'c_nationkey' (integer), and 'c_address' (varchar). On the right, a table from 'npdb' is shown with columns: 'c_address' (varchar), 'c_custkey' (integer), and 'c_name' (varchar). Below the tables, the text '运行结果 #jo01jf953z0t38gbn5nt6zk11agd' is displayed. In the bottom right corner, there is a '任务列表' (Task List) button.

如果在点击 **运行** 后就发生配置相关报错，用户可以点击 **返回配置** 并根据报错信息修改相应的配置，之后可以重新运行该任务。

在任务运行期间，如果需要强制停止运行该任务，点击 **终止** 即可。如需重新运行该任务，用户可以点击 **重启作业** 或者 **重启任务**，重启任务将从上一个 Savepoint 开始执行。

The screenshot shows the configuration interface for a task named 'tp-pdb #2'. The status bar indicates the task is '已关闭' (Closed). There are buttons for '重启作业' (Restart Job), '返回配置' (Return Configuration), and a refresh icon. The data relationship section shows two tables: 'public.customer' from both 'pieclouddbt' and 'npdb' databases. Below the tables, the text '运行结果 #jo01jf99fy24mjk3m3echfxgx064' is displayed. In the bottom right corner, there is a '任务列表' (Task List) button. The results section shows the message 'execution terminated gracefully'.

创建库到库的实时同步任务

在数据集成的「**数据同步**」页面，用户可以根据业务需求分别设置数据关系、机器配置和运行配置以创建库到库的实时同步任务。参考步骤如下：

1. 切换到「**实时同步**」页面，点击 **新建实时同步** 即可进入创建页面。

2. 输入该实时同步任务的名称。
3. 从“类型”下拉列表中选择同步类型为“库到库”。
4. (可选) 输入对该任务的描述。
5. 点击**下一步**以进入该实时同步任务的配置详情页面。



6. 设置“数据关系”以定义数据流向。

在「数据关系」页面，点击**编辑**并执行如下操作：

1. 分别选择源数据源和目标数据源。
2. 选择数据复制方式为“整库复制”或者“多表复制”。

提示:

整库复制方式仅适用于源数据库和目标数据库中表名完全相同的情况。如果表名不一致，请选择多表复制方式。

如果选择“多表复制”，则点击**添加表**，并根据需要选择源表和对应的目标表。在库到库的多表映射中，系统提供了同名表的自动匹配和填充功能。



3. 点击 **保存** 以确认更改。
7. 修改机器配置。

切换到「**机器配置**」页面，点击 **编辑** 以修改机器配置。如下机器配置设置完成后，点击 **保存** 以确认更改。

机器配置项说明如下：

- Job Manager CPU 核数：Job Manager 所能使用的 CPU 总数。默认值：1 Core。
- Job Manager CPU 内存：Job Manager 的总进程内存大小。默认值：1024 Mi。
- Task Manager CPU 核数：Task Manager 所能使用的 CPU 总数。默认值：1 Core。
- Task Manager CPU 内存：Task Manager 的总进程内存大小。默认值：1024 Mi。
- Task Manager Slot：Task Manager 的任务槽数量。默认值：16。

其中，Job Manager 是实时同步作业的管理中心，负责接收用户提交的作业，并进行作业的调度和管理；Task Manager 是实际负责执行计算的 Worker，在其上执行实时同步作业的一组 Task；Task Manager Slot 是 Task Manager 中的最小资源调度单位，用于分配和托管任务的内存空间。

注意：

为避免资源浪费，当 Task Manager Slot 数量超过作业执行的默认并行度 (parallelism.default) 时，系统会自动将 Slot 数量调整为与并行度相等。也就是说，仅当并行度大于 Slot 数量时，系统会按照用户实际设定的 Slot 数量执行作业；反之，如果并行度小于 Slot 数量，系统会按照并行度执行作业。

8. 修改运行配置。

“运行配置”展示数据同步任务的系统信息。对于实时同步任务，运行配置包括 flink checkpoint 触发间隔、超时时长、重启策略、作业并行度等配置项。

切换到「**运行配置**」页面，点击 **修改** 以根据实际需求修改相应的配置信息，之后点击 **完成**。

有关运行配置项的详细信息，请参见 **实时同步任务运行配置项说明**。

点击 **X** 即可退出并返回配置页面。

9. 在实时同步任务的详情页面，点击 **运行** 即可开始运行库到库的实时同步任务。

如果在点击 **运行** 后就发生配置相关报错，用户可以点击 **返回配置** 并根据报错信息修改相应的配置，之后可以重新运行该任务。

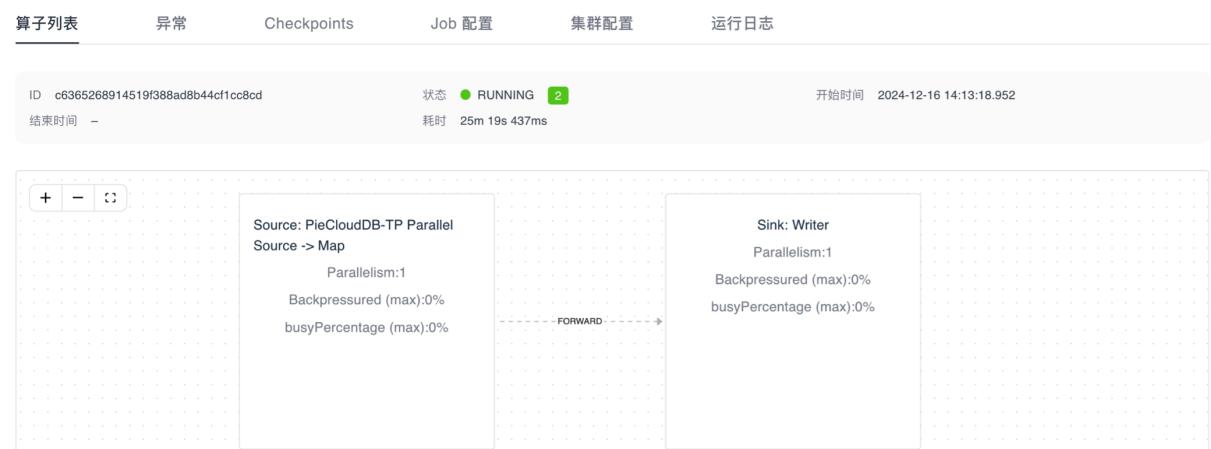
与表到表的实时同步任务相同，在任务运行期间，如果需要强制停止运行该任务，点击 **终止** 即可。如需重新运行该任务，用户可以点击 **重启作业** 或者 **重启任务**，重启任务将从上一个 Savepoint 开始执行。

查看实时同步任务详情

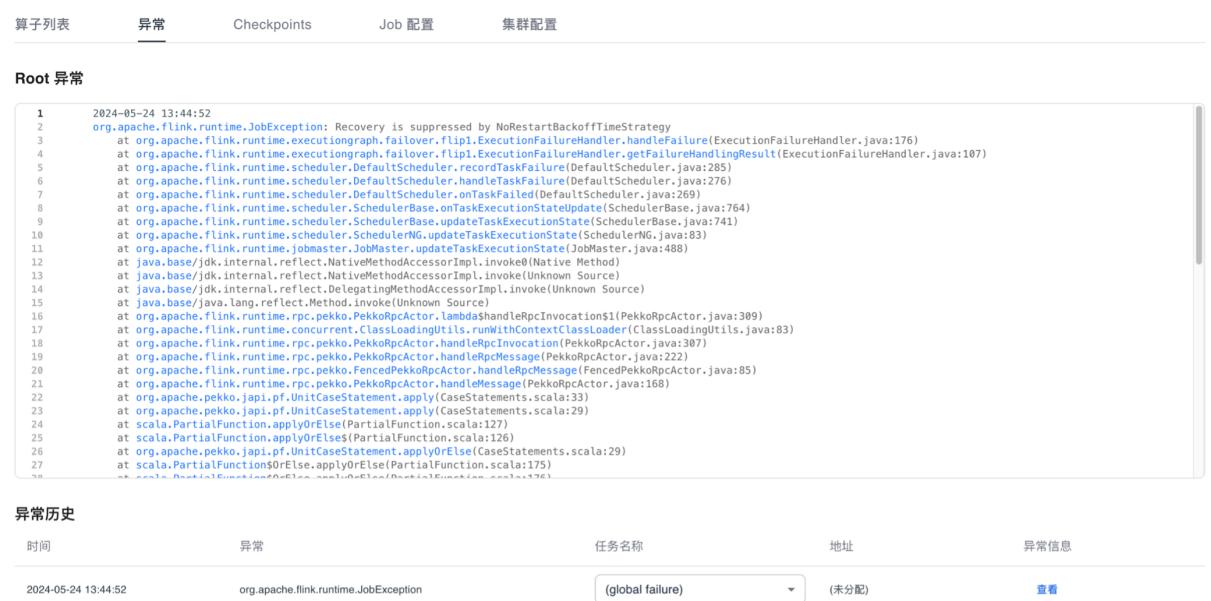
对于执行成功的实时同步任务，用户在运行结果区域点击 **查看详情** 即可查看 flink 算子信息和运行日志等信息。

• **算子列表：**算子列表页面展示运行中的 flink 作业的概览信息和算子详情。概览信息包括作业的运行时长、状态、开始时间等；算子列表以图形化方式展示了作业的逻辑。

执行图，可以查看作业中每个算子的名称、并行度、反压等信息。有关各个算子的更多信息展示在下方列表中。



- 异常：异常页面展示 flink 作业执行与重试过程中发生的所有异常堆栈，以协助排查问题。



- Checkpoints: Checkpoints 页面展示 flink 作业执行过程中所有的 checkpoint 详情、统计信息和配置信息。

算子列表 异常 Checkpoints Job 配置 集群配置 运行日志

总览 历史 统计 配置

Checkpoint 数量 Triggered:1104 In Progress:0 Completed:1091 Failed:13 Restored:0

最新完成的 Checkpoint ID:1091 | Completion Time:2024-12-16 14:31:50.992 | End to End Duration:20073d 6h 31m 50s | Checkpointed Data Size:3.32 KB | Full Checkpoint Data Size:3.32 KB

Checkpoint 详情: Path:/file/flink/its01/j71jev1p3emew2kerz598px/checkpoints/c6365268914519f388ad8b4cf1cc8cd/chk-1091 Discarded:- Checkpoint Type:aligned checkpoint

算子:

Name	Acknowledged	Latest Acknowledgment	End to End Duration	Checkpointed Data Size	Full Checkpoint Data Size	Processed (per)
> Source: PieCloudDB-TP Parallel Source -> Map	1 / 1 (100 %)	2024-12-16 14:31:50.992	14ms	3.09 KB	3.09 KB	
> Sink: Writer	1 / 1 (100 %)	2024-12-16 14:31:50.992	14ms	234 B	234 B	

最新失败的 Checkpoint 无

最新保存的 Checkpoint 无

最新保存的 Checkpoint 无

- Job 配置: Job 配置页面展示 flink 作业的执行配置。

算子列表 异常 Checkpoints Job 配置 集群配置 运行日志

执行配置

执行模式 PIPELINED	最大重试次数 Restart deactivated.	任务并行数 1
对象重用模式 false		

用户配置

暂无数据

- 集群配置: 集群配置页面展示 flink 作业所属集群的配置详情、JVM 启动参数和 classpath。

算子列表	异常	Checkpoints	Job 配置	集群配置	运行日志
Job Manager 配置					
\$internal.application.program-args					--jobId:jo01jf71jetqea7n82h3gm8gpew9;--serverAddr:pie-dataflowserver-grpc.openpie-connect.svc.cluster.local:8099
\$internal.flink.version					v1_18
\$internal.pipeline.job-id					c6365268914519l388ad8b44cf1cc8cd
blob.server.port					6124
execution.checkpointing.externalized-checkpoint-retention					RETAIN_ON_CANCELLATION
execution.checkpointing.interval					5 min
execution.shutdown-on-application-finish					false
execution.submit-failed-job-on-application-error					true
JVM					
version					OpenJDK 64-Bit Server VM - Eclipse Adoptium - 11/11.0.22+7
arch					amd64
options					-Xmx1530082096 -Xms1530082096 -XX:MaxMetaspaceSize=268435456 -XX:+IgnoreUnrecognizedVMOptions -Dlog.file=/opt/flink/log/flink-kubernetes-application-0-ts01jf71jev1p3emew2kerz598px-676db68df9-jq748.log

- 运行日志：运行日志页面展示 flink 作业所属集群的 Job manager 和 Task manager 的运行日志。

算子列表	异常	Checkpoints	Job 配置	集群配置	运行日志
Job Manager ▾					
Job Manager					
Task Manager					
					最后更新时间 文件大小
					flink-kubernetes-application-0-ts01jf71jev1p3emew2kerz598px-676db68df9-jq748.log 2024-12-16 14:37:24 1.03 MB

实时同步任务运行配置项说明

配置项	说明	默认值
execution.checkpointing.interval	触发检查点（Checkpoint）的时间间隔。它指定了在连续两个检查点之间应有的最长时间（毫秒）。如果设置为大于 0 的值，系统将在这个时间间隔后自动触发检查点。	1000
execution.checkpointing.min-pause	这是在连续两个检查点之间的最小暂停时间（以毫秒为单位）。如果任务处理速度落后于数据生成速度，系统会暂停发送数据以等待任务赶上。该参数定义了暂停操作的最小时长。	500
execution.checkpointing.timeout	检查点操作的超时时间（以毫秒为单位）。如果在指定的超时时间内检查点未能完成，系统将取消当前检查点，并尝试执行下一次检查点。	600000
parallelism.default	作业执行的默认并行度。它决定了作业中所有算子的默认并行实例数。	1

配置项	说明	默认值
state.checkpoints.num-retained	保留的检查点（Checkpoint）数量。当启用了基于时间的触发检查点时，这个参数决定了在磁盘上保留多少个最近的检查点。	1
restart-strategy.type	作业失败时的重启策略类型。可能的值包括： <ul style="list-style-type: none">• none：无重启策略。• fixed-delay：固定延迟启动策略。• failure-rate：故障率重启策略。	none
restart-strategy.fixed-delay.delay	当重启策略类型为 fixed-delay 时，该参数设置了两次连续重启尝试之间的延迟。单位：秒（s）。	1
restart-strategy.fixed-delay.attempts	当重启策略类型为 fixed-delay 时，该参数设置了将作业声明为失败之前重试执行的次数。-1 表示次数无上限。	-1
restart-strategy.failure-rate.delay	当重启策略类型为 failure-rate 时，该参数设置了两次连续重启尝试之间的延迟。单位：秒（s）。	1
restart-strategy.failure-rate.failure-rate-interval	当重启策略类型为 failure-rate 时，该参数设置了测量重新启动策略故障率的时间间隔。单位：秒（s）。	60
restart-strategy.failure-rate.max-failures-per-interval	当重启策略类型为 failure-rate 时，该参数设置了在 failure-rate.delay 指定的时间窗口内允许的作业失败之前的最大重启次数。	1
state.backend.dir	状态后端存储的目录路径用于保存实时同步作业的状态信息，包括检查点和保存点。	默认使用 flink 集群默认配置

实时同步常见报错与解决方案

在执行同步任务过程中，可能遇到的执行失败的情况，在任务详情页面的运行结果区域，用户可以通过查看任务的状态和执行详情来了解失败的原因。

如果任务失败是由于配置问题引起的，那么需要重新修改任务配置。用户可以点击 **返回配置** 并根据报错信息修改相应的配置，之后可以重新运行该任务。如果实时同步任务执行失败，用户可以在运行结果区域点击 **查看详情** 来查看 flink 算子信息和运行日志等信息，以辅助排查问题。

运行结果 #jo01jf953z0t38gbn5nt6zk11agd

#ts01jf953z1mt7f6yhc26p4frv5
⌚ 2024-12-17 09:53:28 ~ 2024-12-17 09:54:27 ● 失败

结果

错误信息：

```
java.io.IOException: Could not perform checkpoint 5 for operator Sink: Writer (1/1)#0.
    java.lang.RuntimeException: org.postgresql.util.PSQLException: ERROR: tuple concurrently deleted (mstoream.c:5739)
        org.postgresql.util.PSQLException: ERROR: tuple concurrently deleted (mstoream.c:5739)
```

在执行同步任务时，如果发生如下报错，可以参考相应的解决方案。

- 报错信息:** `java.lang.IllegalArgumentException: source table [database_name.table_name] do not contain any primary key.`

原因: 源数据表未包含主键。

解决方案: 为源数据表指定一个主键。同时，如果源数据库中实际的表没有主键，而用户仅在关联表模型的字段管理中修改“主键”的信息，实时同步任务在校验时同样会报错。

- 报错信息:** `java.lang.IllegalArgumentException: source[database_name.table_name] wal level is replica, should be set to logical.`

原因: 日志级别未设置为“replica”。

解决方案: 对于源数据库（通常是 PostgreSQL 系列），将其日志级别应设置为“replica”。为了启用逻辑复制（logical replication）功能，这是实时同步功能的底层依赖，需要将数据库的日志级别设置为 `wal_level=logical`。对于 PostgreSQL 系列数据库，可以通过修改配置文件并重启数据库实例来实现这一设置；而对于 PieCloudDB TP 数据库，除了直接修改配置文件外，还可以通过在计算空间平台上设置实例的 GUC 模板来进行配置。

- 报错信息:** `java.lang.IllegalArgumentException: table [table_name] replication identity is [d], should be set to full.`

原因：源表的 *REPLICA IDENTITY* 属性未设置为 *FULL*。

解决方案：修改源表的 *REPLICA IDENTITY* 属性为 *FULL*，参考 SQL 命令如下：

```
ALTER TABLE table_name REPLICA IDENTITY FULL;
```

在 PostgreSQL 中，*REPLICA IDENTITY* 属性用于确定在逻辑复制过程中，如何识别被更新或删除的行。这个属性对于确保变更数据流（Change Data Capture, CDC）的准确性至关重要。为了在实时同步中准确地识别更新（UPDATE）和删除（DELETE）操作的前后状态，需要将 *REPLICA IDENTITY* 设置为 *FULL*。这样，PostgreSQL 会在 WAL（Write-Ahead Logging）中记录足够的信息来支持逻辑复制。

6.5.10 管理元数据

在数据集成模块中，元数据构成了一个统一的、可配置的通用表模型，包含表信息、字段信息和字段排序等属性。

通用表模型可以关联多个数据源，使不同的数据源能够共享该模型，从而减少元数据冗余，增加系统的灵活性和可扩展性。

关于通用表模型

在数据集成的「数据源」页面的数据源列表中，用户点击目标数据源名称即可进入数据源详情页面。

进入数据源详情页面后，在数据源“关联的表”区域，用户点击 **查看全部数据表** 即可进入数据集成的「全部数据表」页面。该页面以列表形式显示所有的通用表模型，并支持创建通用表模型和从数据源导入通用表模型。目标通用表模型所在行的隐藏菜单「...」还提供修改和删除通用表模型以及字段管理的选项。

表名	schema	表名是否大小写敏感	标题	数据源	数据源类型	描述	创建时间
tb	public	否	public.tb	adb_pg	pieclouddb		2024年11月25日星期一 10:07
region	public	否	public.region	adb_pg	pieclouddb		2024年11月25日星期一 10:07
tb3	public	否	public.tb3	adb_pg	pieclouddb		2024年11月25日星期一 10:07

通用表模型的列表包含如下字段信息：

- 表名：通用表模型的表名称。
- Schema（选填）：Schema 名称。如果未指定，则默认为“public”。

- 表名是否大小写敏感：表名是否开启了大小写敏感。
- 标题（选填）：自定义表的标题。
- 数据源：如果通用表是“从数据源导入”的并绑定到了源数据源，则会显示该数据源的名称。
- 数据源类型：如果通用表是“从数据源导入”的并绑定到了源数据源，则会显示该数据源的类型。
- 描述（选填）：对通用表的描述。
- 创建时间：创建通用表的时间。

通用表模型的列表提供如下快捷操作：

- 根据数据源类型筛选信息：使用**数据源类型**控件可以通过数据源的类型筛选相关的通用表信息。
- 根据 Schema 名称搜索信息：使用**输入 Schema 查询**的搜索框可以通过 Schema 的名称搜索相关的通用表信息。
- 根据表名搜索信息：使用**输入表名查询**的搜索框可以通过表的名称搜索相关的通用表信息。
- 立即刷新：点击立即刷新图标可以即刻同步当前的列表信息为最新。
- 修改通用表：在通用表模型列表的行隐藏菜单下提供了修改的快捷键，可以用于修改指定的表信息。
- 删除通用表：在通用表模型列表的行隐藏菜单下提供了删除的快捷键，可以用于删除指定的表。

创建一个新的通用表模型

在数据集成的「**全部数据表**」页面，如果需要创建一个新的通用表模型，操作步骤如下：

1. 点击**创建**即可进入创建表的页面。
2. 输入该通用表的表名。
3. （可选）指定该通用表所属的 Schema。
4. 指定是否启用表名的大小写敏感。默认设置为不启用，勾选该选项后即可启用。
5. （可选）分别输入该通用表的标题和描述信息。
6. 点击**完成**。通用表模型创建成功后，其信息会自动同步到列表中。

通过上述步骤创建的通用表模型，只是创建了一张表并无字段信息。在新创建的通用表模型所在行的隐藏菜单「...」下，用户选择**字段管理**选项即可进入目标通用表的「**字段管理**」页面，并根据实际需求进行新增字段、修改字段属性和删除字段等管理操作，详细信息请参见本章节的**管理通用表模型的字段**。

从数据源导入一个通用表模型

在数据集成的「**全部数据表**」页面，如果需要从数据源导入一个通用表模型，操作步骤如下：

1. 点击**从数据源导入**即可进入从数据源导入表的页面。
2. 在“选择数据源”的下拉列表中选择一个目标数据源，之后页面会显示“选择原始表”字段。
3. 在“选择原始表”下拉列表中选择目标数据源的一张表。如果需要查看该原始表的字段信息，点击**查看字段**即可显示所选原始表的字段名、字段类型、字段精度、主键等信息。

指定原始表后，页面会随即展开需要配置的字段信息。

The screenshot shows the 'Import from Data Source' dialog box. At the top, it says 'Import from Data Source' and has a close button 'X'. Below that is a 'Select Data Source' section with a dropdown menu showing 'pdb-pdb'. A dashed red box highlights the 'Select Original Table' section, which contains another dropdown menu showing 'Customer Data.Customer Data_2020_2023'. Below these are several configuration options:

- A checked checkbox 'Bind the table to the data source pdb-pdb'.
- An optional field 'Rename Table Name (Optional)' with an empty input field.
- A checkbox 'Table name case sensitivity' which is unchecked.
- An optional field 'Title (Optional)' with an empty input field.
- An optional field 'Description (Optional)' with an empty input field.

At the bottom right are two buttons: 'Cancel' and 'Finish'.

4. 指定所导入的原始表是否绑定到数据源。该原始表会默认被绑定到所选的目标数据源，取消勾选即可解除绑定。

提示：

解除绑定后，该表在通用表模型列表的“数据源”和“数据源类型”字段不会显示相应的信息。

5. (可选) 如果需要重新命名要导入的表，可以输入一个新的表名。
6. 指定是否启用表名的大小写敏感。默认设置为不启用，勾选该选项后即可启用。
7. (可选) 分别输入该通用表的标题和描述信息。
8. 点击**完成**。通用表模型导入成功后，其信息会自动同步到列表中。

在所导入的通用表模型所在行的隐藏菜单「...」下，用户选择**字段管理**选项即可进入目标通用表的「**字段管理**」页面，并进行新增字段、修改字段属性和删除字段等管理操作，详细信息请参见本章节的**管理通用表模型的字段**。

管理通用表模型的字段

注意：

建议用户谨慎使用字段管理功能。

用户可以根据不同的数据源，灵活地定义表结构以适应不同的业务场景，包括新增字段、自定义已有字段的名称和字段数据类型等属性信息。

在通用表模型列表的行隐藏菜单下，用户点击**字段管理**选项即可进入指定通用表的「**字段管理**」页面。该页面支持如下操作：

- 新增字段

在指定通用表的字段列表下方，点击**新增**来添加新的自定义字段。分别输入字段的属性信息，包括字段名、字段类型、字段长度（如有）、字段精度（如有）、是否主键和注释等。

完成每个字段的操作后，点击**完成**按钮，该字段就会被添加到字段列表中。

- 修改字段属性

在字段列表的“操作”栏下，点击**修改**来修改指定字段的属性信息，包括字段名、字段类型、字段长度（如有）、字段精度（如有）、是否主键和注释等。完成每个字段的操作后，点击**完成**即可保存对目标字段的更改。

- 删除字段

在字段列表的隐藏菜单下，点击**删除**，确认后即可删除指定字段并将其从列表中移除。

- 字段排序

字段排序管理功能使用户能够根据个人喜好或业务需求定义字段的显示顺序。用户可以通过拖拽或手动调整字段顺序，来合理地安排字段的排序。

6.6 数据共享

6.6.1 数据共享概述

数据共享（Data Sharing）功能支持同一组织下的计算空间（PieCloudDB 虚拟数仓类型）之间的跨计算空间数据共享。在当前版本中，数据共享以表（Table）为单位进行，一个共享可以包含来自多个 Schema 和多个数据库的多张表。

数据共享功能主要适用于内部跨团队协作、数据整合与分发、敏捷数据分析等场景。其核心特性如下：

- 跨计算空间共享

支持同一组织内的不同计算空间（PieCloudDB 虚拟数仓类型）之间直接共享数据，减少中间层数据复制，打破数据孤岛，并加速分析流程。

- 以表为共享粒度

共享的最小单位是表（Table），可灵活选择需要共享的特定表，并支持从多个数据库（Database）和多个 Schema 中抽取表以组合成共享内容。

- 多层级数据整合

允许将分散在不同计算空间、数据库或 Schema 中的表整合到一个共享中，简化跨业务域的数据调用，而无需数据迁移。

- 权限与安全控制

基于组织内计算空间的权限体系，通过共享机制仅开放脱敏后的表给到指定计算空间，避免未授权访问，以满足合规要求。

提示：

数据共享属于 PieCloudDB 企业版的独有功能，需要在部署时开启功能开关。有关部署的相关信息，请联系 OpenPie 技术支持团队。

6.6.2 关于数据共享权限

数据共享功能的目标用户通常分为两类：数据分享者和数据订阅者。数据分享者可以将数据加入共享，并指定共享的计算空间；数据订阅者可以订阅公开的共享数据，同时管

理那些被指定共享至当前计算空间的数据。基于用户视角，数据共享功能的权限设置涵盖了系统权限以及数据库对象权限。

权限说明

- 系统权限

数据共享功能支持如下系统预设权限：

- 数据共享菜单：控制数据共享功能的显示与隐藏。如果用户未被授予相应权限，则无法访问数据共享功能。
- 共享数据分享信息：允许用户查看指定计算空间下共享的数据信息。
- 共享数据分享管理：允许用户在指定计算空间下创建、删除和编辑共享。
- 共享数据接收信息：允许用户查看共享给所在计算空间的数据信息。
- 共享数据接收管理：允许用户订阅或取消订阅公共数据共享，并将共享表映射到本地。

其中，“共享数据分享信息”为 PUBLIC 权限。除此权限外，数据分享者还需要同时具有“数据共享菜单”与“共享数据分享管理”的权限；而数据订阅者还需要同时具有“数据共享菜单”、“共享数据接收信息”与“共享数据接收管理”的权限。

提示：

Account Manager 角色的初始权限默认包括上述数据共享功能的系统预设权限。

- 数据库对象权限

数据共享的数据库对象权限与表的 *SELECT* 权限相对应。如果用户对某张表具有 *SELECT* 权限，并且同时具备“数据共享菜单”与“共享数据分享管理”的系统权限，那么该用户就可以创建共享，并且能够将自己拥有 *SELECT* 权限的表加入到共享中，进而分享给其他计算空间。

如何授予用户相关权限

本文通过具体示例来介绍如何在 PieDataCS 云原生平台对一位新用户授予数据共享功能的接收和管理权限。

1. 创建自定义角色并授予其数据订阅者的相应权限。

以管理员身份进入目标计算空间并创建自定义角色（例如 data-subscriber）。

将“数据共享菜单”、“共享数据接收信息”与“共享数据接收管理”权限授予该角色之后，保存并应用所做的修改。

角色详情(data-subscriber)

数据订阅者

系统权限

数据库对象权限

角色继承

用户授予

选择要添加或者删除的权限

系统初始化赋予的权限和继承自其他角色的权限无法修改

共享数据接收信息 × 数据共享菜单 × 共享数据接收管理 ×

可授予权限(30)

虚拟数仓详情 虚拟数仓基础管理能力 虚拟数仓外部接入能力 虚拟数仓监控
虚拟数仓巡检 数据导入功能操作 数据导入历史数据 用户信息
用户管理 用户审核 用户详情 角色信息 角色详情 角色管理
赋予/取消赋予系统权限给角色 继承/取消(被)继承角色给角色
赋予/取消赋予角色给用户 用量统计 定时任务-查看任务运行记录
定时任务-任务管理 数仓管理 虚拟数仓错误日志 虚拟数仓SQL审计
用户行为审计日志 脱敏信息 脱敏设置管理
获取虚拟数仓中的Query和Session信息 数据集功能管理 数据集功能只读
共享数据分享管理

取消

保存修改

提示:

如果当前其他角色已具有相关权限，在创建自定义角色时也可以通过直接继承角色的方式获得该权限。

有关创建自定义角色的详细信息，请参见 [管理计算空间角色-添加自定义角色](#)。

2. 创建新用户作为数据订阅者。

以管理员身份在计算空间的「**用户**」页面创建一个新用户（例如 `data_receiver`），并填写相应的注册信息。

用户名
data_receiver

用户登录方式
 手机 邮箱

邮箱
[Redacted]

手机号
+86

密码
.....

再次输入密码
.....

取消 创建用户

有关新建用户的详细信息，请参见 [管理计算空间用户-新建用户](#)。

3. 授予新用户以角色并继承相应的数据订阅者的权限。

新用户通过审核后，在计算空间的「**用户**」页面，管理员可以进入该用户的详情页面，并为其赋予 **步骤 1** 中创建的角色。完成这一步骤后，该用户将继承相应的权限。

用户名: data_receiver 用户ID: [Redacted]
手机号: [Redacted] 邮箱: [Redacted]

data_receiver被授予了2个角色

角色名称	操作
public	收回
data-subscriber	收回

有关授予用户以角色的详细信息，请参见 [管理计算空间用户-授予/收回用户的角色](#)。

在 **步骤 3**所创建的用户被赋予相关权限后，即可登录 PieDataCS 云原生平台，使用订阅数据共享功能，请参见 [订阅数据共享并映射到本地](#)。

如果需要为用户赋予数据分享者的相关权限（“数据共享菜单”和“共享数据分享管理”），可以参考上述步骤进行操作。需要注意的是，数据分享者和数据订阅者需要处于不同的计算空间中。

6.6.3 创建与管理数据共享

数据共享的数据为公共数据和私有数据。公共数据是指在目标组织内公开可见的数据，但该组织以外的计算空间无法查看这些数据，因此也无法订阅这些数据。而私有数据不在组织下公开，只能共享给指定的计算空间。

创建数据共享

在「**数据共享**」页面，数据分享者在创建数据共享时，需依次添加共享表和共享对象，具体步骤如下：

1. 切换到“我共享的”面板，该面板会显示当前计算空间下创建的所有共享。
2. 在页面点击**新建共享**，在弹出的窗口中，参考如下步骤创建共享：
 1. 输入所创建共享的名称和描述。
 2. 选择是否在组织下公开，即组织下所有计算空间可用。选项“是”表明该共享的数据是公共数据，选项“否”则表明该共享的数据是私有数据。



3. 点击**确认**，所创建的共享信息会同步到“我共享的”面板的列表中。

数据共享

共享给我的		我共享的				
是否公开	全部	输入名称查询				
名称	描述	是否公开	共享对象	更新时间	创建时间	操作
data-for-share	共享数据	是	查看	2025-02-25 17:56:56	2025-02-25 17:56:56	管理共享数据 ...
private-data	私有数据	否	查看	2025-02-25 17:00:07	2025-02-25 17:00:07	管理共享数据 ...

3. 在“我共享的”面板的列表中的目标共享所在行的“操作”栏下，点击**管理共享数据**，在弹出的窗口中，参考如下步骤添加共享表。

1. 在对应的下拉列表中，依次选择数据库和数据库中指定 Schema 下的目标表。



2. 点击**添加表**即可将目标表添加到共享数据列表（支持添加多个共享表）。



注意:

如果要添加的指定数据库下的目标表已存在于其他共享中，会导致添加失败，并报错“记录已存在”。

3. 共享表添加完成后，点击 **X** 即可退出窗口，并返回到“我共享的”面板。
4. 在“我共享的”面板的列表中的目标共享所在行的“共享对象”栏下，点击 **查看**，在弹出的窗口中，参考如下步骤添加共享对象（即计算空间）。
1. 输入指定的计算空间 ID，点击 **添加** 即可将该计算空间添加为共享对象（支持添加多个共享对象）。

**提示:**

如果共享数据为公共数据，也可以省略此步骤，因为本计算空间以外的数据订阅者在订阅侧也能获取此公共共享信息并执行订阅操作；而对于私有数据，则必须通过此步骤添加共享对象。

如需移除目标共享对象，在目标共享对象所在行的“操作”栏下，点击 **移除** 即可。

2. 共享对象添加完成后，点击 **X** 即可退出窗口，并返回到“我共享的”面板。

管理数据共享

数据共享者在管理共享数据时，需注意以下事项：

- 如果公共共享数据的分享者希望将公共数据转为私有数据，只需将共享数据的“是否公共”属性设置为“否”。此时，如果已有计算空间订阅了该数据，订阅关系不会失效；但如果分享者将已订阅的计算空间从共享中移除，那么映射到接收者本地的数据访问将会失效。

在“我共享的”面板的目标共享所在行的隐藏菜单栏下，选择**编辑**选项即可修改共享数据的属性。

- 如果公共共享数据的分享者想要删除该共享，必须先移除所有共享对象，之后才能成功删除共享。

在“我共享的”面板的目标共享所在行的隐藏菜单栏下，选择**删除**选项即可删除共享。

6.6.4 订阅数据共享与取消订阅

对于公共数据，任何计算空间的数据订阅者都能够主动进行订阅；而对于私有数据，仅限于特定计算空间的数据订阅者可以订阅并查看。

订阅公开数据共享并映射到本地

在「**数据共享**」页面，数据订阅者需先订阅（公开）共享，再将共享表映射到本地数据库之后，才能访问共享数据。具体步骤如下：

- 切换到“共享给我的”面板，该面板会显示当前用户所在计算空间内所订阅的所有共享。

The screenshot shows a user interface for managing shared data. At the top, there are two tabs: 'Shared with Me' (selected) and 'My Shared'. Below the tabs is a search bar with placeholder text '输入名称查询' (Enter name query). To the right of the search bar is a button labeled '订阅公开共享' (Subscribe to public sharing), which is highlighted with a red border. Underneath the search bar are four columns: '共享来源' (Sharing source), '共享名称' (Shared name), '更新时间' (Update time), and '创建时间' (Create time). A horizontal line labeled '操作' (Operations) is positioned below these columns. At the bottom of the list area, it says '暂无数据' (No data).

提示：

对于私有数据，数据订阅者可以在“共享给我的”面板中的列表中直接查看到指定给该计算空间的私有共享。

- 点击**订阅公开共享**即可进入公开共享页面，查看所有可以订阅的公开共享。

The screenshot shows a user interface for managing public sharing. At the top, it says '数据共享' (Data sharing) and has a language switch to '中文' (Chinese) and a greeting '您好:' (Hello:). Below this is a header '共享给我的 / 公开共享' (Shared with me / Public sharing) with a search bar '输入名称查询' (Enter name query) and a refresh button. The main content area has columns: '共享来源' (Sharing source), '名称' (Name), '描述' (Description), '更新时间' (Update time), '创建时间' (Create time), '订阅状态' (Subscription status), and '操作' (Operations). There is one entry: 'Neptune (01jm...)' with 'data-for-share' as the name, '共享数据' (Shared data) as the description, update and creation times as '2025-02-25 17:56:56', and a status of '未订阅' (Not subscribed). To the right of the status are buttons for '添加订阅' (Add subscription) and '查看共享数据' (View shared data).

- 在所选的目标共享所在行的“操作”栏下，点击**添加订阅**并确认共享来源，之后点击**确定**即可添加目标共享到“共享给我的”面板。此时，订阅状态会显示为“已订阅”。

共享给我的 / 公开共享

共享来源	名称	描述	更新时间	创建时间	订阅状态	操作
Neptune (01jmy@khbkpy5...)	data-for-share	共享数据	2025-02-25 17:56:56	2025-02-25 17:56:56	已订阅	取消订阅 查看共享数据

如需查看共享数据，在当前公开共享页面的目标共享所在行的“操作”栏下，点击**查看共享数据**即可显示所订阅共享的数据库下的表和其字段信息。

4. 返回“共享给我的”面板，在所订阅的目标共享所在行的“操作”栏下，点击**查看详情**即可进入该共享的详情页面。

数据共享

共享来源	共享名称	更新时间	创建时间	操作
Neptune (ac01jh5548h0q2fyqyrsr0r6y35)	private-data	2025-02-28 15:39:02	2025-02-28 15:39:02	查看详情 ...
Neptune (ac01jh5548h0q2fyqyrsr0r6y35)	data-for-share	2025-02-27 17:57:07	2025-02-27 17:57:07	查看详情 ...

5. 在目标共享的详情页面可以执行映射共享数据到本地的操作。参考步骤如下：

1. 切换到共享详情页面的“共享表”面板，即可显示所订阅的共享数据中表的相关信息。

数据共享

所属数据库	表	字段	操作
openpie_sample_data	食品销售数据.交易数据_2020_2023	共 7 个字段	映射表

2. 在目标表所在行的“操作”栏下，点击**映射表**即可进入映射表配置界面。
3. 在下拉列表中选择要引入目标共享表的本地数据库和 Schema。



4. 点击 **创建**，操作成功后，相关信息将同步显示在共享详情页面的“映射表”面板的列表中。

数据共享

共享给我的 / 共享详情

共享来源	更新时间	共享名称	创建时间
dbshare	2025-02-27 17:57:07	data-for-share	2025-02-27 17:57:07

映射表 **共享表** **C**

所属数据库	表	共享表	描述	创建人	操作
dbshare	public.交易数据_2020_2023	交易数据_2020_2023		data_receiver	移除

共 1 条 每页 10 条 1 前往 页

注意：

当共享数据映射到本地后，如果数据分享者移除了共享表，然后又重新添加了同一张表，这也会导致映射失效。此时，数据订阅者需要移除并重新添加映射。

完成上述操作步骤后，数据订阅者就可以在当前计算空间的「**数据库**」页面，查看本地映射的共享表的详细信息，相关操作信息请参见 [数据库](#)。

数据订阅者也可以在当前计算空间的「**数据洞察**」页面，通过映射表对所订阅的共享数据执行相关查询，相关操作信息请参见 [数据洞察](#)。

取消订阅数据共享

数据订阅者可以通过如下两种方式取消所订阅的数据共享：

注意：

如果已经对共享数据完成映射操作，则需要在移除映射表之后再取消订阅，否则会报错。

- 在「**共享给我的**」页面的目标共享所在行的“操作”栏下，点击隐藏菜单并选择**取消订阅**即可。

共享给我的	我共享的
<input type="text"/> 输入名称查询	<input type="button"/> C <input type="button"/> 订阅公开共享
共享来源	共享名称
Neptune [REDACTED]	data-for-share

更新时间 创建时间 操作
2025-03-04 15:05:34 2025-03-04 15:05:34 [查看详情](#) ...
2025-03-04 15:04:54 2025-03-04 15:04:54 [查看详情](#)

- 在「**共享给我的**」页面，点击**订阅公开共享**以进入公开共享页面。之后在目标共享所在行的“操作”栏下，点击**取消订阅**即可。

共享给我的 / 公开共享						
共享来源	名称	描述	更新时间	创建时间	订阅状态	操作
Neptune (01 jmy@khbkpy6...)	data-for-share	共享数据	2025-02-25 17:56:56	2025-02-25 17:56:56	已订阅	<input type="button" value="取消订阅"/> 查看共享数据

7 附录

7.1 PieProxy 外部接入示例

通过 Postgres Client PSQL 命令连接虚拟数仓

通过 Postgres Client PSQL 命令连接虚拟数仓的步骤如下：

1. 设置环境变量。命令格式如下：

```
export PGOPTIONS='-c warehouse= 虚拟数仓ID'
```

示例如下：

```
export PGOPTIONS='-c warehouse=cl*****'
```

2. 运行 psql 命令。命令格式如下：

```
psql -h 接入IP地址 -U 用户名 -p 端口号 -d 数据库名称
```

示例如下：

```
psql -h 192.x.x.x -U openpie -p 5432 -d openpie
```

通过 Python psycopg2 连接虚拟数仓

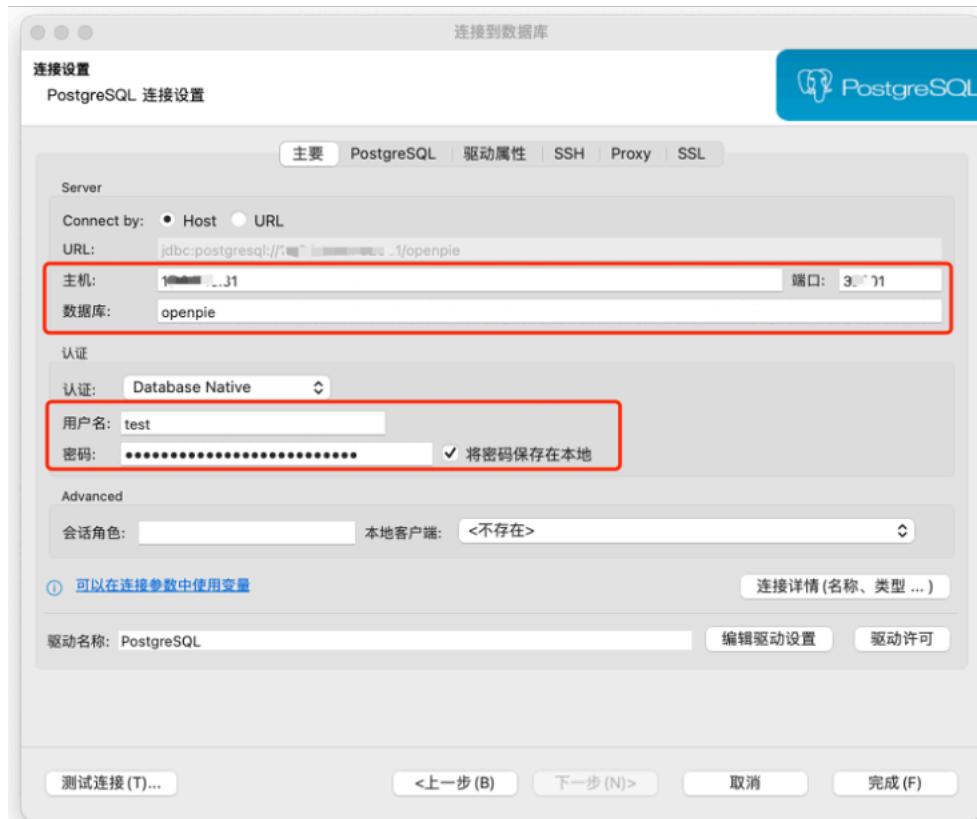
通过 Python psycopg2 连接的前提条件是当前 Python 环境已安装 psycopg2。连接示例如下：

```
import psycopg2
conn=psycopg2.connect(host="1**.*.*.*",
                      port="5432",
                      dbname="openpie",
                      user="david",
                      password="****",
                      sslmode="disable",
                      options="-c warehouse=cl*****")
cursor=conn.cursor()
cursor.execute("select * from pg_roles limit 5")
result=cursor.fetchall()
```

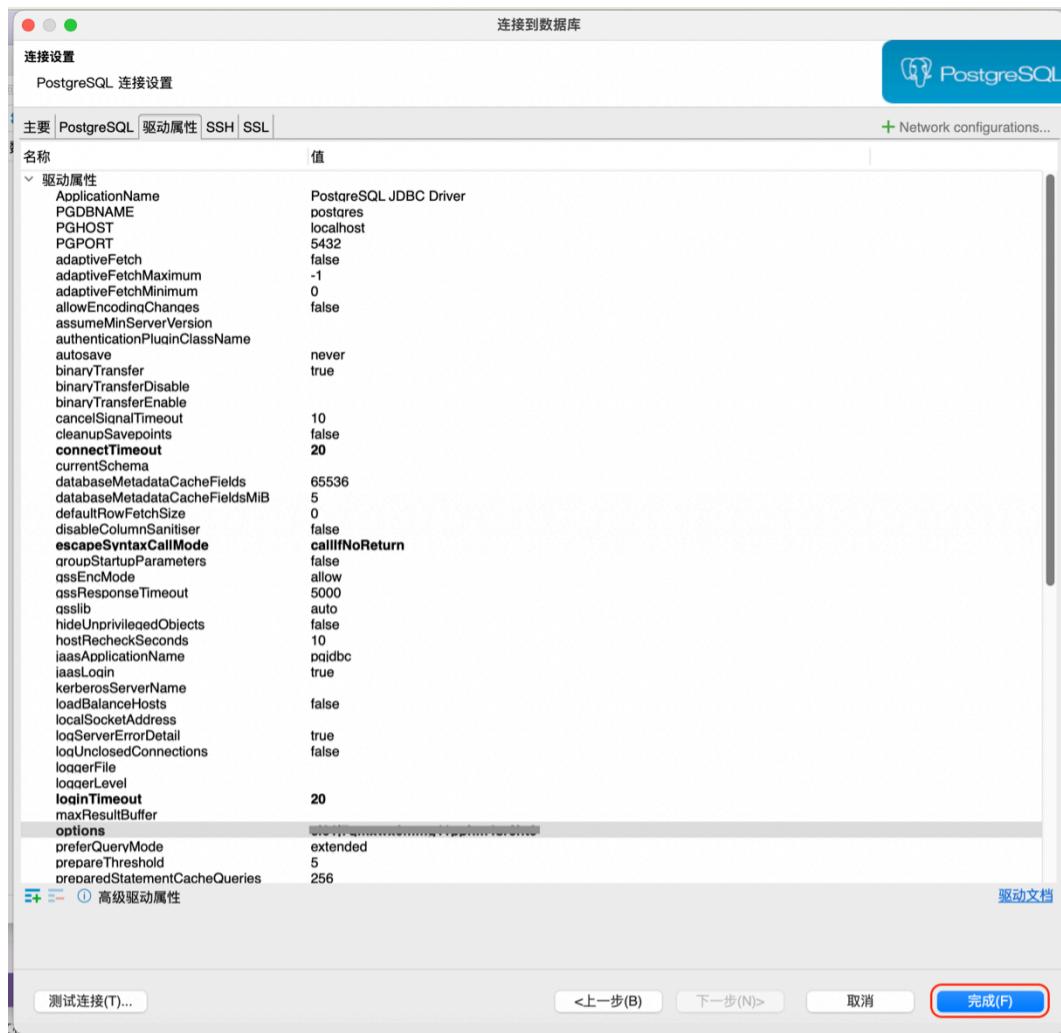
通过 DBeaver 连接虚拟数仓

通过 DBeaver 连接虚拟数仓的步骤如下：

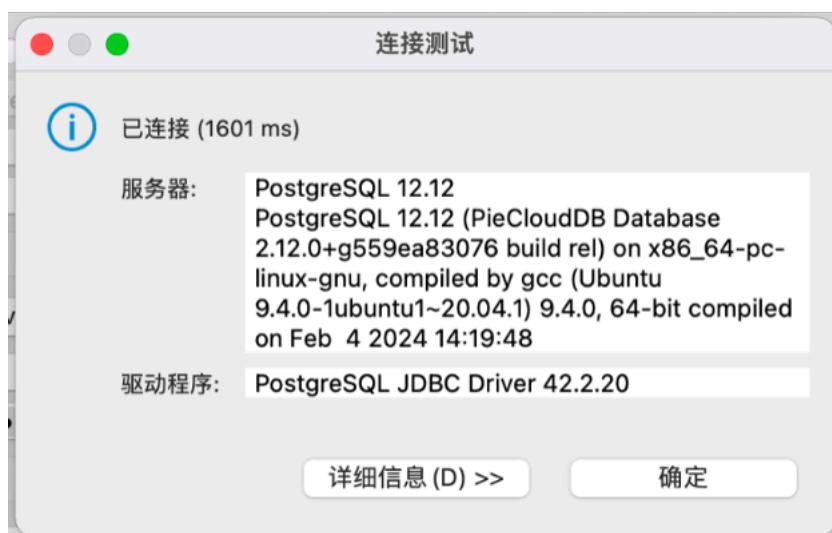
- 进入 DBeaver 软件操作界面后，点击 **数据库 > 新建数据库连接**，选择 PostgreSQL，并设置主机 Host、数据库 Database、用户名 Username 和密码 Password。



- 切换驱动属性界面，在“options”选项处填入虚拟数仓 ID（例如，`-C warehouse=cl*****`），并点击 **完成**。

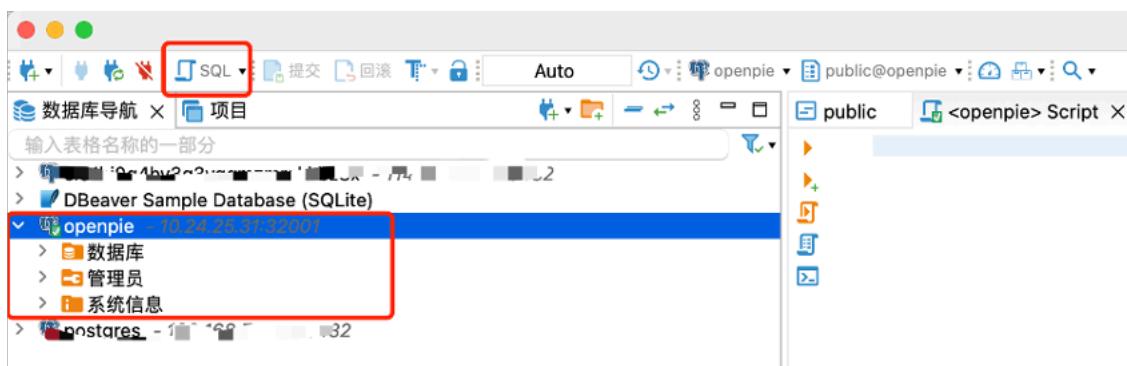


3. 点击 **测试连接**，验证是否成功连接数据库。下图为连接成功示例。



4. 点击 **完成**，完成连接设置。

5. 选择所设置的连接，点击 **SQL**，添加一个 SQL 文件。



6. 在所连接的“openpie”数据库中新建一张表 test_1，插入数据并查询验证。

```
CREATE TABLE test_1(id int);
INSERT INTO test_1 VALUES (1),(2),(3);

SELECT * FROM test_1;
```

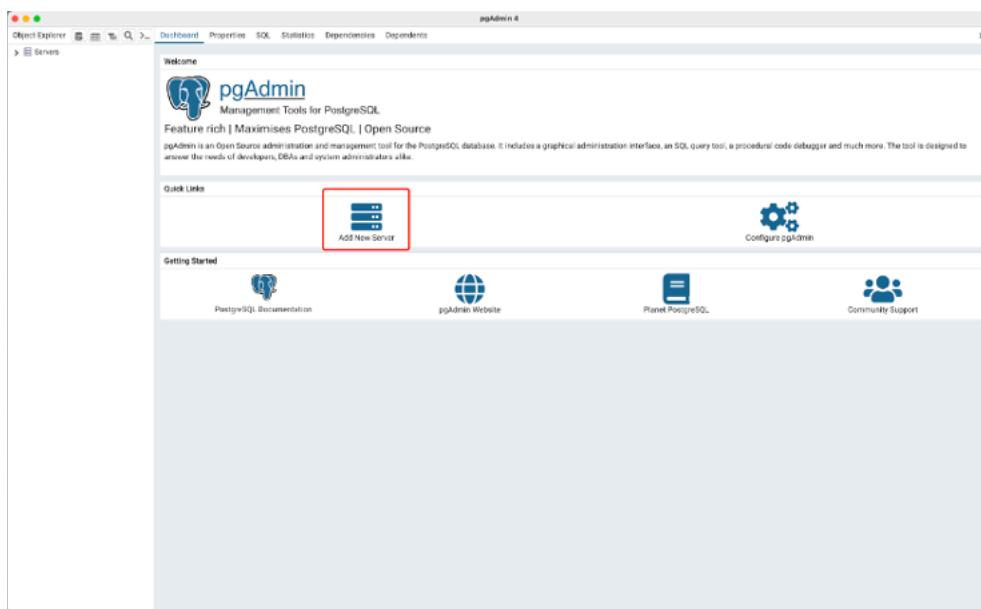
运行结果如下图所示。

test_11 X	
CREATE TABLE test_1(id int); INSERT INTO test_1 VALUES (1),(2),(3);	
网格	123 id
1	1
2	2
3	3
文本	
SQL	
二进制	
树状	
视图	
帮助	

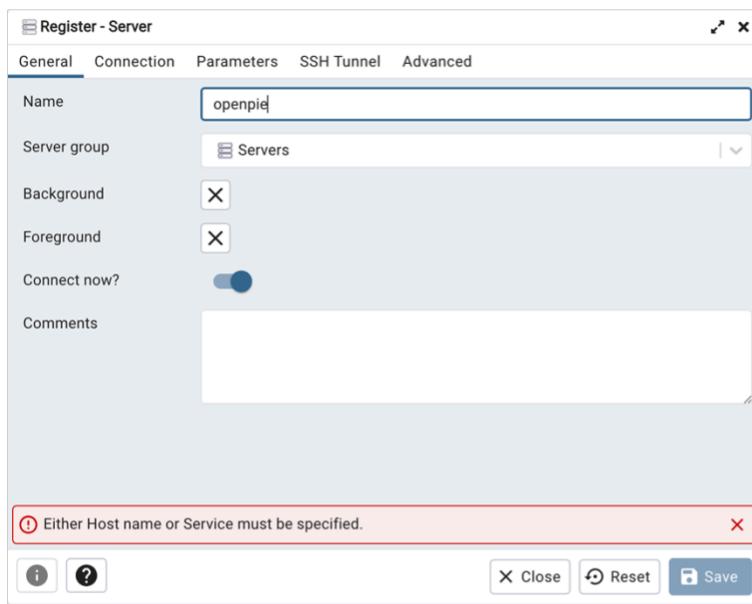
通过 PGAdmin4 连接虚拟数仓

通过 PGAdmin4 连接虚拟数仓的步骤如下：

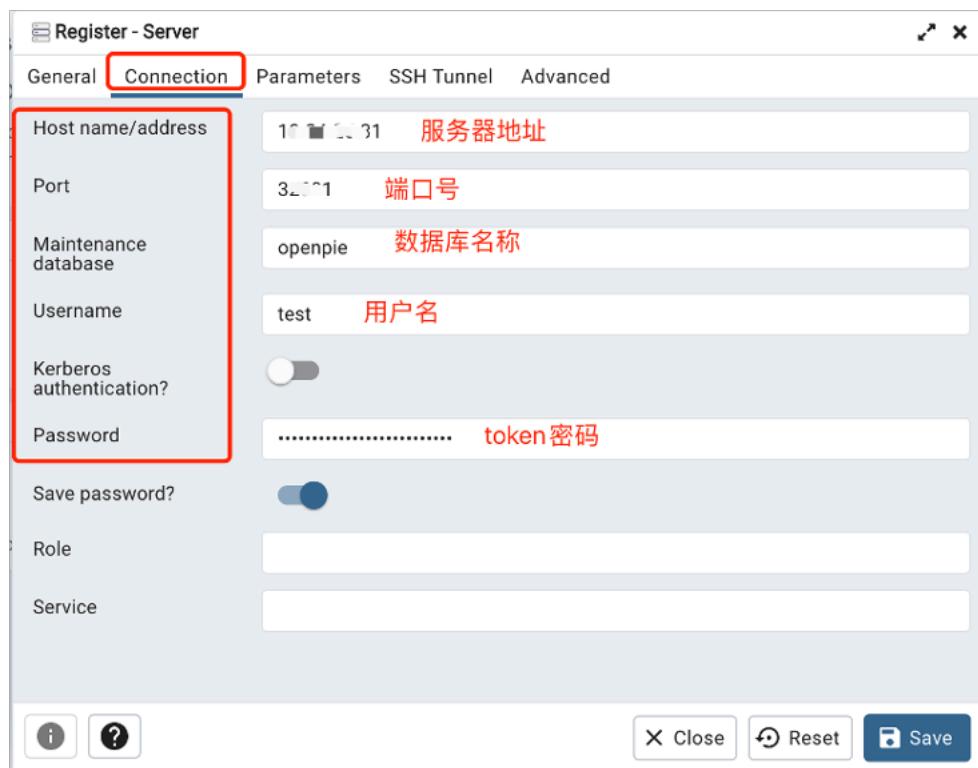
1. 进入PGAdmin4软件操作界面后，点击 **Add New Server**，新建连接。



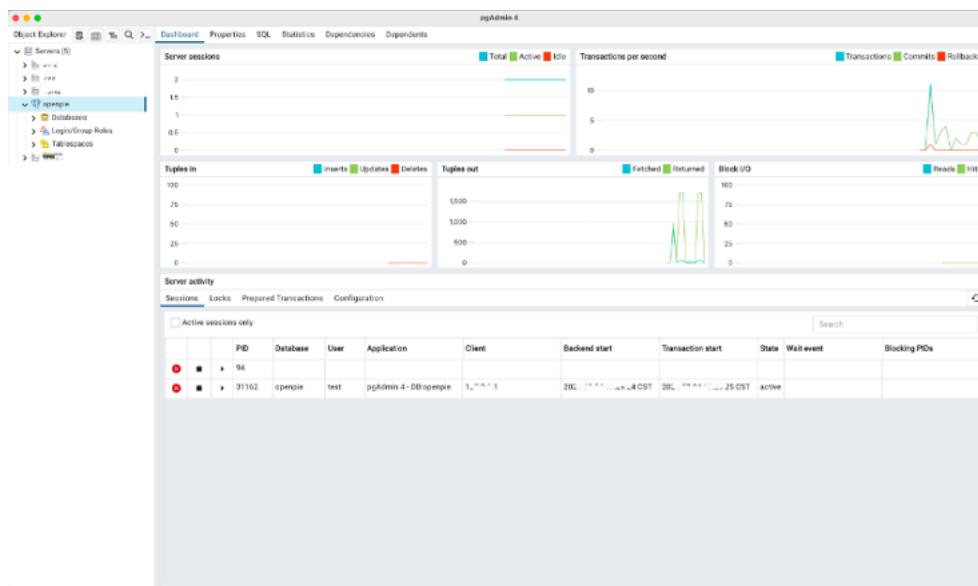
2. 在 General 配置页面输入连接名称并保存。



3. 在 Connection 配置页面输入通过平台获取的服务器、数据库和用户信息，并保存。



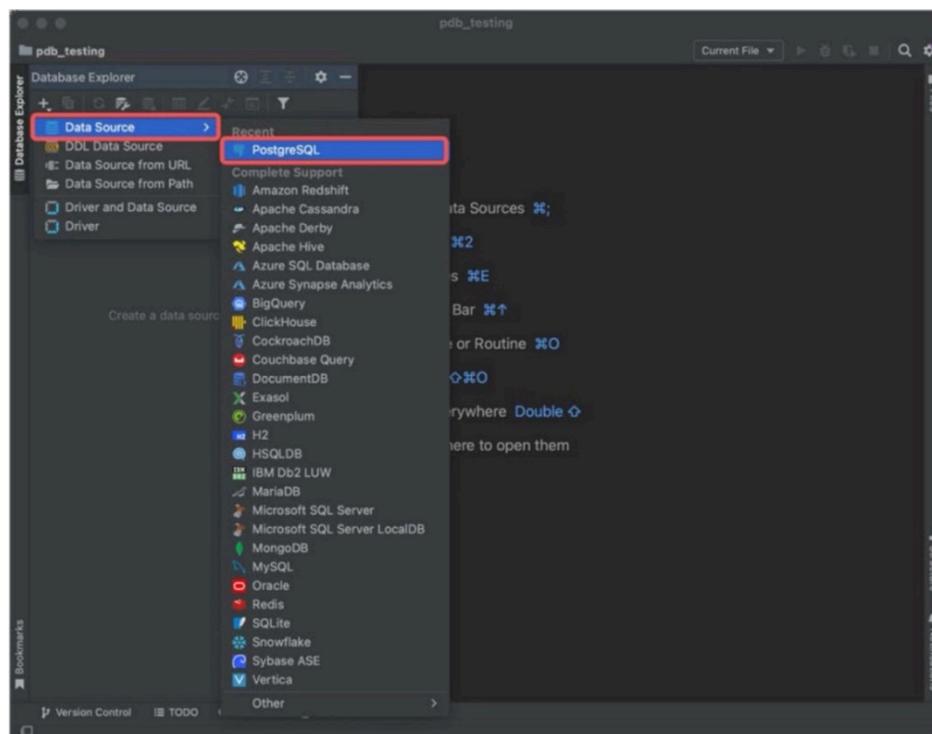
PGAdmin 会自动进行连接测试，下图为连接成功的示例。



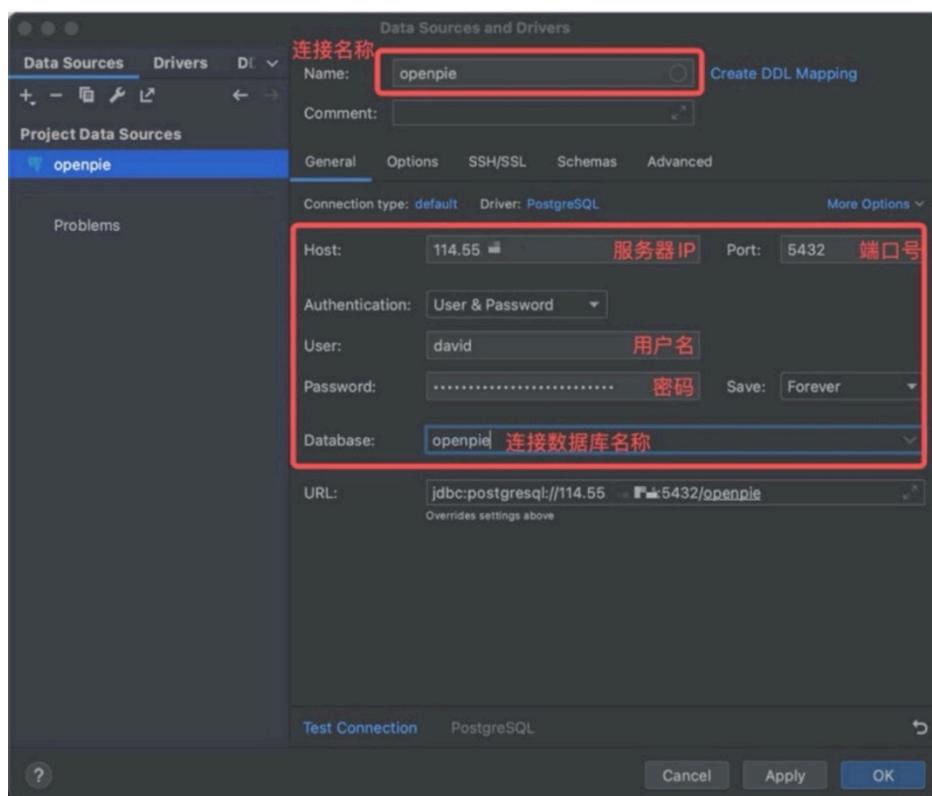
通过 DataGrip 连接虚拟数仓

通过 DataGrip 连接虚拟数仓的步骤如下：

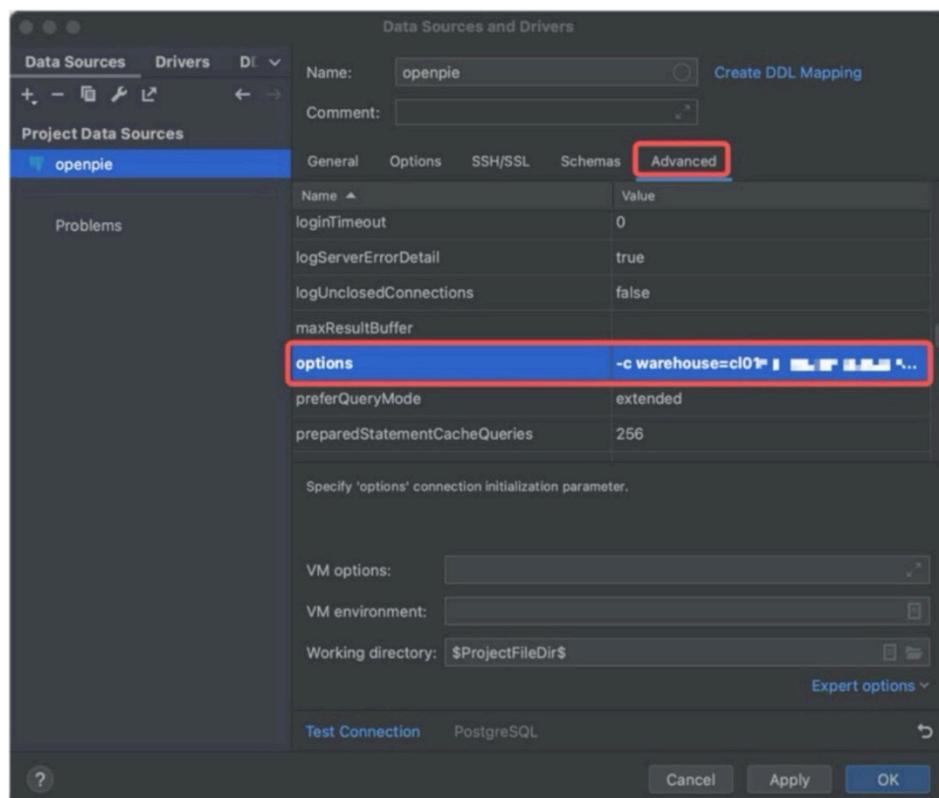
1. 进入 DataGrip 软件操作界面后，点击按钮 +，选择 **Data Source->PostgreSQL** 新建连接。



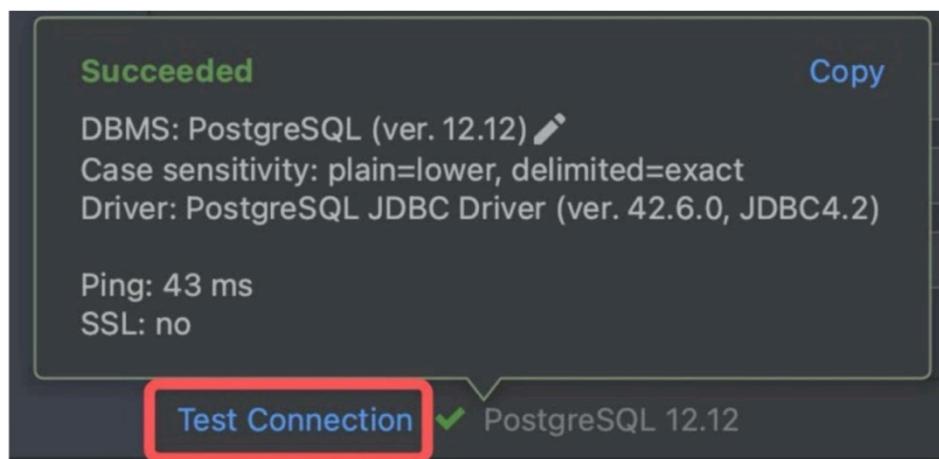
2. 在连接配置页面，输入连接名称和数据库相关信息。



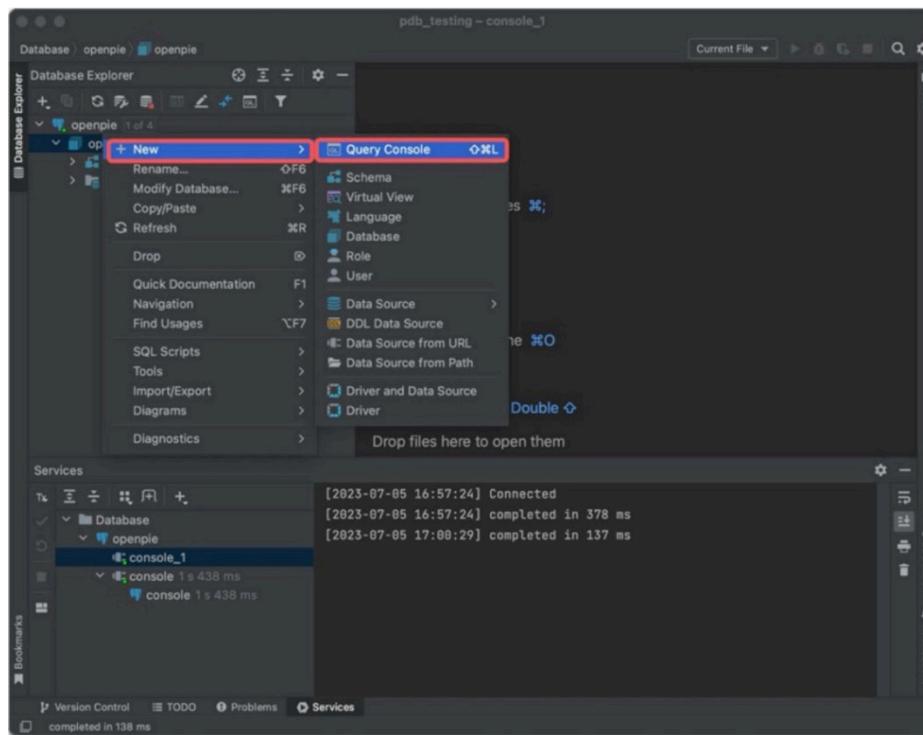
3. 在 Advanced 页面设置连接选项。在“options”一栏填入要连接的虚拟数仓 ID，例如，`-c warehouse=c1*****`。



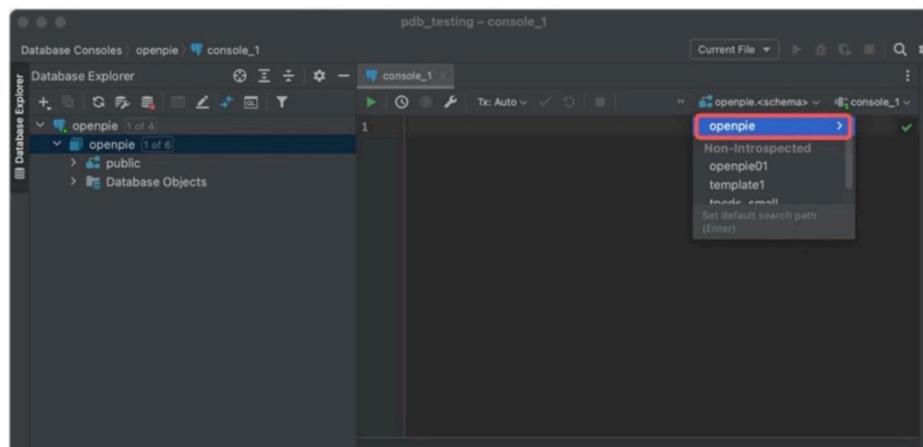
4. 点击 **Test Connection**, 验证是否连接数据库成功。下图为连接成功示例。



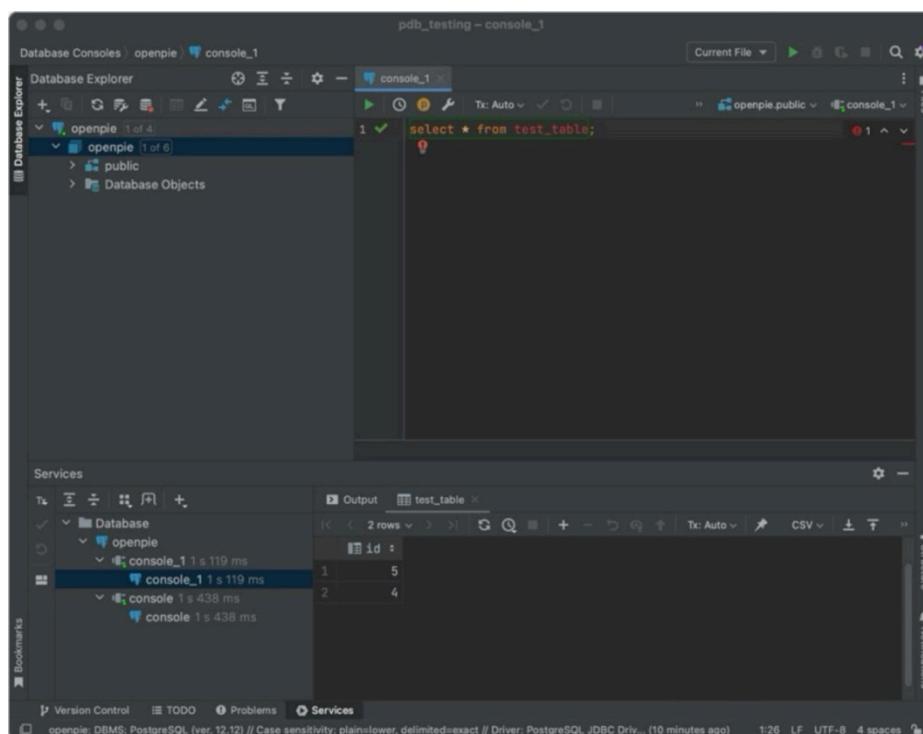
5. 点击 **OK** 完成连接设置。



6. 返回主页面，选中所连接的数据库，添加一个 SQL 文件，并选择要执行 SQL 的数据库。



7. 相关设置完成后，连接 PieDataCS 云原生平台执行 SQL 查询。



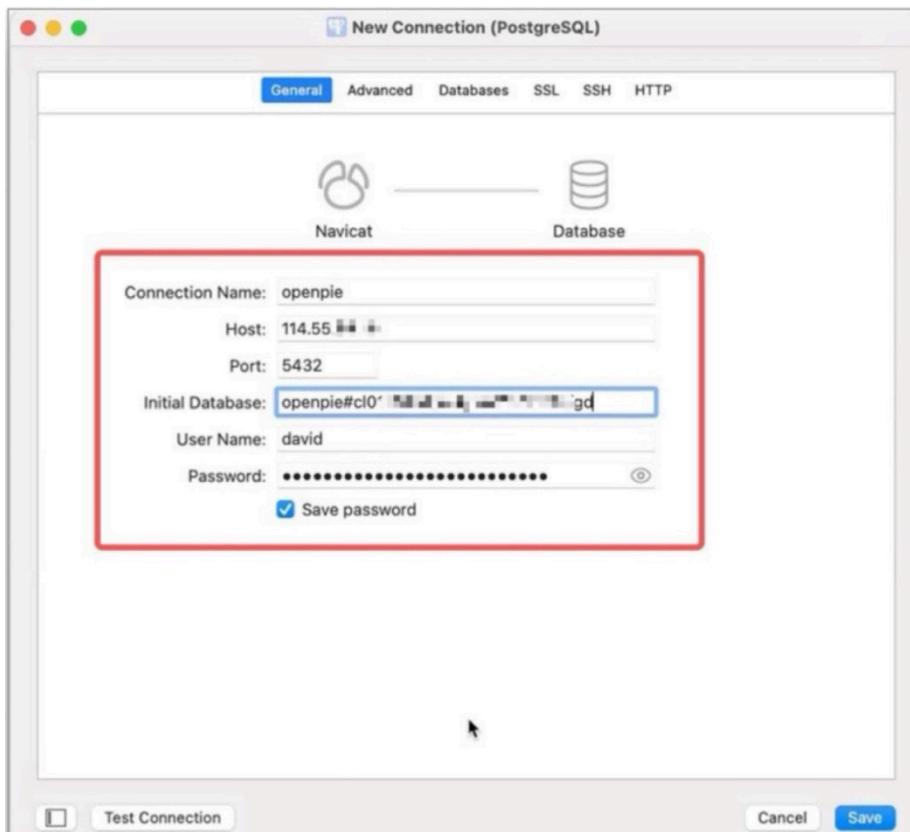
通过 Navicat 连接虚拟数仓

通过 Navicat 连接虚拟数仓的步骤如下：

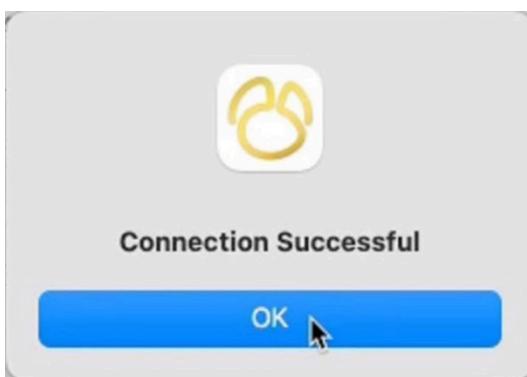
1. 进入 Navicat 软件操作界面后，点击 **Connection**，选择 PostgreSQL 新建连接。



2. 在连接页面，输入从 PieDataCS 云原生平台获取的服务器、数据库和用户信息。在数据库名称处添加“#”及虚拟数仓 ID。



3. 点击 **Test Connection**, 验证是否连接数据库成功。下图为连接成功的示例。



4. 点击 **Save**, 保存设置。
 5. 连通数据库后，在 Navicat 操作界面新建一个 SQL 文件，对所连接的数据库中的数据执行查询等操作。

通过 JDBC URL 连接虚拟数仓

JDBC URL 的一般格式如下：

```
jdbc:postgresql://<虚拟数仓IP>:<端口号>/<数据库名称>?[currentSchema=<Schema名称>&]options=-c%20warehouse%3D<虚拟数仓ID>
```

参数说明如下：

- <虚拟数仓IP>：虚拟数仓的 IP 地址。
- <端口号>：虚拟数仓的端口号。
- <数据库名称>：要连接的数据库的名称。
- <Schema名称>：（可选）要连接的 Schema 名称，通过 currentSchema 参数指定。如果未指定，则默认连接 public Schema。
- <虚拟数仓ID>：虚拟数仓 ID，通过 options 参数并以 `-c%20warehouse%3D` 的格式指定。

String url 的示例如下：

```
String url = "jdbc:postgresql://10.xx.xx.xxx:xxxxx/test_db?  
currentSchema=mssc&sslmode=disable&options=-c%20warehouse%3Dyour_warehouse_id";
```

提示：

如果使用集成开发环境（IDE）通过 JDBC 连接虚拟数仓，例如 IntelliJ IDEA，则需导入如下依赖包：

- PostgreSQL 数据库的 JDBC 驱动程序的 JAR 文件，例如 postgresql-42.3.1.jar。
- Druid 数据库连接池的 JAR 文件，例如 druid-1.1.10.jar。

使用 JDBC 连接到一个虚拟数仓，并执行一个简单查询的完整 Java 示例如下：

```

import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.Statement;

public class TestSelect {
    public static void main(String[] args) {
        String url = "jdbc:postgresql://10.xx.xx.xxxxx/test_db?currentSchema=mysc&sslmode=disable&options=-c%20warehouse%3Dxxxxxxxxxxxxxx";
        String username = "username";
        String password = "password";

        try {
            Connection connection = DriverManager.getConnection(url, username, password);
            Statement statement = connection.createStatement();
            ResultSet resultSet = statement.executeQuery("SELECT * FROM mytb");

            while (resultSet.next()) {
                int id = resultSet.getInt("id");
                String name = resultSet.getString("name");
                System.out.println("ID: " + id + ", Name: " + name);
            }

            resultSet.close();
            statement.close();
            connection.close();
        } catch (Exception e) {
            e.printStackTrace();
        }
    }
}

```

下图为虚拟数仓连接成功后的示例。

```

TestSelect.java
1 package com.example.helloworld;
2
3 import java.sql.Connection;
4 import java.sql.DriverManager;
5 import java.sql.ResultSet;
6 import java.sql.Statement;
7
8 public class TestSelect {
9     public static void main(String[] args) {
10         String url = "jdbc:postgresql://10.xx.xx.xxxxx:31102/test_pg?currentSchema=mysc&sslmode=disable&options=-c%20warehouse%3Dxxxxxxxxxxxxxx";
11         String username = "username";
12         String password = "password";
13
14         try {
15             Connection connection = DriverManager.getConnection(url, username, password);
16             Statement statement = connection.createStatement();
17             ResultSet resultSet = statement.executeQuery("SELECT * FROM mytb");
18
19             while (resultSet.next()) {
20                 int id = resultSet.getInt("id");
21                 String name = resultSet.getString("name");
22                 System.out.println("ID: " + id + ", Name: " + name);
23             }
24
25             resultSet.close();
26             statement.close();
27             connection.close();
28         } catch (Exception e) {
29             e.printStackTrace();
30         }
31     }
32 }
33
运行 TestSelect
com.example.helloworld.TestSelect
ID: 1, Name: a
ID: 2, Name: b
进程已结束，退出代码为 0

```

7.2 角色的系统权限清单

计算空间的系统权限

PieDataCS 云原生平台的企业版和云上云版支持如下计算空间的系统权限。

权限名称	权限描述
数据洞察	使用数据洞察功能
数据库	使用数据库功能，包括查看数据库树和查看各级数据库对象基本信息
数据库对象权限信息	在数据库和角色下查看数据库对象权限内容
虚拟数仓基本信息	查看虚拟数仓列表
虚拟数仓详情	获取虚拟数仓启停历史
虚拟数仓基础管理能力	虚拟数仓的新增、启动、关闭、强制关闭、删除、修改等操作
虚拟数仓外部接入能力	虚拟数仓外部接入开关控制和开放连接信息详情
虚拟数仓监控	查看虚拟数仓监控
虚拟数仓巡检	虚拟数仓巡检和查看虚拟数仓巡检报告
数据导入功能操作	数据导入任务创建、数据上传等
数据导入历史数据	获取数据导入历史数据
查询历史	查看查询历史列表及详情信息
虚拟数仓错误日志	查看虚拟数仓错误日志
虚拟数仓 SQL 审计	查看虚拟数仓 SQL 审计
用户信息	查看用户列表信息
用户管理	用户的新增、删除、修改
用户审核	审核用户能否加入数仓
用户详情	查看用户信息详情
角色信息	查看角色关系图和列表信息
角色详情	查看角色（被）继承的角色、用户和权限
角色管理	角色的新增、删除、修改
赋予/取消赋予系统权限给角色	赋予/取消赋予权限给角色
继承/取消（被）继承角色给角色	继承/取消（被）继承角色给角色
赋予/取消赋予角色给用户	赋予/取消赋予权限给用户
赋予/取消赋予数据库权限给角色	赋予/取消赋予权限给角色
数仓管理	数仓的新增、删除、修改
数仓信息	获取数仓信息
公共参数	公共参数
用户行为审计日志	查看用户行为审计日志

如下表格中的系统权限仅适用于企业版。

权限名称	权限描述	初始化所赋予的角色
脱敏设置管理	管理数据脱敏的设置	dbmanager
定时任务-查看任务运行记录	定时任务查看任务运行记录	dbmanager
定时任务-任务管理	查看定时任务、中止任务、定时任务的启停、修改、删除、创建	dbmanager
数据源/流/作业功能只读	数据源/流/作业功能只读	accountmanager
数据源/流/作业功能管理	数据源/流/作业功能管理	accountmanager
脱敏信息	查看数据库表脱敏设置信息	dbmanager
脱敏设置管理	脱敏数据库表和脱敏用户设置	dbmanager

控制台的系统权限

PieDataCS 云原生平台的企业版和云上云版支持如下计算空间的系统权限。

权限名称	权限描述
充值购买	数仓订阅订单支付、续费、账户充值、余额查询、交易记录查询等支付相关功能
数仓下虚拟数仓列表	查看数仓下虚拟数仓列表
订单列表	访问订单管理菜单、查看订单列表和订单详情
产品规格和定价信息	查看产品规格和定价信息；提交订阅、续费和自动续费管理
组织信息	查看组织详情
组织管理	组织层面相关的管理
数仓列表	访问数仓管理菜单、查看组织下的数仓列表
数仓管理	数仓创建、修改、删除
数仓信息	查看数仓详情
数仓下用户列表	查看数仓下用户列表
组织下用户列表	访问用户管理菜单、查看组织下用户列表
用户管理	用户创建、修改、删除
用户信息	查看用户详情
用户审核	用户审核
组织和数仓用量信息	访问费用总览菜单、查看组织/数仓的用量信息和明细列表
角色信息	访问角色菜单、查看角色列表和角色信息
角色管理	角色的新增、删除、修改
赋予/取消赋予权限给角色	赋予/取消赋予权限给角色
继承/取消（被）继承角色给角色	继承/取消（被）继承角色给角色
赋予/取消赋予权限给用户	赋予/取消赋予权限给用户
公共参数	查看公共参数和信息
发票申请	发票申请

注意：

仅 PieDataCS 云原生平台的云上云版支持订单列表、产品规格和定价信息、产品规格和定价信息、组织和数仓用量信息和发票申请的系统权限。