

智能数据构建与管理 Dataphin 数仓规划

ALIBABA CLOUD

文档版本: 20220712



法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。 如果您阅读或使用本文档,您的阅读或使用行为将被视为对本声明全部内容的认可。

- 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档,且仅能用 于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息,您应当严格 遵守保密义务;未经阿里云事先书面同意,您不得向任何第三方披露本手册内容或 提供给任何第三方使用。
- 未经阿里云事先书面许可,任何单位、公司或个人不得擅自摘抄、翻译、复制本文 档内容的部分或全部,不得以任何方式或途径进行传播和宣传。
- 由于产品版本升级、调整或其他原因,本文档内容有可能变更。阿里云保留在没有 任何通知或者提示下对本文档的内容进行修改的权利,并在阿里云授权通道中不时 发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠 道下载、获取最新版的用户文档。
- 4. 本文档仅作为用户使用阿里云产品及服务的参考性指引,阿里云以产品及服务的"现状"、"有缺陷"和"当前功能"的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引,但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的,阿里云不承担任何法律责任。在任何情况下,阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害,包括用户使用或信赖本文档而遭受的利润损失,承担责任(即使阿里云已被告知该等损失的可能性)。
- 5. 阿里云网站上所有内容,包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计,均由阿里云和/或其关联公司依法拥有其知识产权,包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意,任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外,未经阿里云事先书面同意,任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称(包括但不限于单独为或以组合形式包含"阿里云"、"Aliyun"、"万网"等阿里云和/或其关联公司品牌,上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司)。
- 6. 如若发现本文档存在任何错误,请与阿里云取得直接联系。

通用约定

格式	说明	样例
⚠ 危险	该类警示信息将导致系统重大变更甚至故 障,或者导致人身伤害等结果。	
▲ 警告	该类警示信息可能会导致系统重大变更甚 至故障,或者导致人身伤害等结果。	警告 重启操作将导致业务中断,恢复业务 时间约十分钟。
〔) 注意	用于警示信息、补充说明等,是用户必须 了解的内容。	大意 权重设置为0,该服务器不会再接受新 请求。
? 说明	用于补充说明、最佳实践、窍门等 <i>,</i> 不是 用户必须了解的内容。	⑦ 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在 结果确认 页面,单击 确定 。
Courier字体	命令或代码。	执行 cd /d C:/window 命令,进入 Windows系统文件夹。
斜体	表示参数、变量。	bae log listinstanceid
[] 或者 [alb]	表示可选项,至多选择一个。	ipconfig [-all -t]
{} 或者 {a b}	表示必选项,至多选择一个。	switch {act ive st and}

目录

1.概述	07
2.业务板块	09
2.1. 新建业务板块	09
2.2. 新建数据域	13
2.3. 编辑板块参数	14
3.管理统计周期	17
4.计算源	20
4.1. 概述	20
4.2. 新建MaxCompute计算源	21
4.3. 新建AnalyticDB For PostgreSQL计算源	22
4.4. 新建Flink计算源	23
4.5. 创建加速计算源	25
4.6. 管理计算源	28
5.项目空间	30
5.1. 创建Prod和Dev项目	30
5.2. 创建Basic项目	39
5.3. 管理项目空间的权限和计算源	47
6.数据源	55
6.1. 概述	55
6.2. Dataphin支持的数据源	58
6.3. Dataphin出网IP地址	61
6.4. 网络连通解决方案	61
6.5. 大数据存储数据源	64
6.5.1. 创建MaxCompute数据源	64
6.5.2. 创建SAP HANA数据源	68
6.5.3. 创建DataHub数据源	71

6.5.4. 创建Hologres数据源	74
6.6. 文件数据源	77
6.6.1. 创建FTP数据源	77
6.6.2. 创建OSS数据源	81
6.7. 消息队列数据源	83
6.7.1. 创建Log Service数据源	83
6.7.2. 创建Kafka数据源	86
6.7.3. 创建RocketMQ数据源	88
6.8. 关系型数据源	91
6.8.1. 创建MySQL数据源	91
6.8.2. 创建Oracle数据源	98
6.8.3. 创建PostgreSQL数据源 1	102
6.8.4. 创建Microsoft SQL Server数据源1	108
6.8.5. 创建AnalyticDB for MySQL数据源11	114
6.8.6. 创建PolarDB-X数据源 1	118
6.8.7. 创建Vertica数据源 1	123
6.8.8. 创建AnalyticDB for PostgreSQL数据源1	125
6.8.9. 创建PolarDB数据源 1	130
6.8.10. 创建Teradata数据源 1	136
6.8.11. 创建IBM DB2数据源 1	139
6.8.12. 创建OceanBase数据源 1	143
6.8.13. 创建Impala数据源 1	149
6.9. NoSQL数据源 1	152
6.9.1. 创建Elasticsearch数据源1	152
6.9.2. 创建MongoDB数据源1	155
6.9.3. 创建Tablestore数据源 1	158
6.9.4. 创建Aliyun HBase数据源 1	161
6.9.5. 创建HBase数据源 1	165

6.9.6. 创建ClickHouse数据》	₹ 16
------------------------	------

1.概述

数仓规划是基于Dataphin建设数据中台的第一步,同时也是数据体系的顶层设计中至关重要的一步。在您开始数据开发前,需要完成数据仓库的规划,包括定义业务板块、项目、数据源、计算源和统计周期。

功能介绍

数仓规划包括逻辑空间与物理空间的全局架构,以实现业务划分、资源管控、项目管理等目的:

• 逻辑空间:基于业务特征划分命名空间及其核心对象,包括业务板块、数据域和公共定义模块。

⑦ 说明 如果您购买的Dataphin实例为基础研发版,则系统不支持业务板块、数据域和公共定义。

• 物理空间: 基于开发协作管理需求划分物理项目, 包括项目管理、数据源管理等模块。

基础研发版和智能研发版支持的功能不同:

- 基础研发版:规划包括项目和源两大模块。模块之间存在依赖关系,项目的计算存储资源配置基于计算 源,项目空间内所用到的数据基于数据源。
- 智能研发版:规划包括业务、项目和源。模块之间存在依赖关系,项目的计算存储资源配置基于计算源,项目空间内所用到的数据基于数据源,项目可以归属于业务板块。

网络连通

为了实现数据源与Dataphin之间的网络互通,您需要将访问数据库的链接地址添加至Dataphin实例的沙箱白 名单中,详情请参见创建Basic项目。Dataphin的IP加入至RDS数据源的白名单中。详情请参见通过客户端、命令 行连接RDS MySQL实例。

Dataphin不同区域的IP白名单,如下表所示。

区域	白名单
华东2(上海)	47.102.151.182
华南1(深圳)	123.56.104.202
华北2(北京)	119.23.173.65
华东1(杭州)	47.96.75.100

入口介绍

1. 登录Dataphin控制台。

2. 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。

Dataphin工作区	▲ Access Key配置
欢迎来到Dataphin! 智能数据仓库构建 高效数据	
选择工作区地域	开通其他Region 续费
华拉2 (北京) 华东2 (上海) 华南1 (深圳) 华东1 (杭州)	
进入Dataphin >>	
已开通: 智能评没板(MaxCompute)-500款据处理单元 数据定量 已开通 数据定案:	续费
快速开始相关工作	
规范建模 数据研发 项目管理 数据萃取 调度运进 资产地图 资产质量	8
资产治理 数据服务 成员管理	

您也可以单击**快速开始相关工作**区域的**项目管理**,快速进入数仓规划。

- 3. 在Dataphin首页,单击顶部菜单栏的规划。
- 4. 在数仓规划页面, 左侧导航栏为您展示各功能模块的入口按钮。

≡	Datap	hin·规划						ස 🕹 ල ද	s 📀
业务		业务板块 (107)						维护升级 ① 🕂	新建业务板块
<mark>₼</mark> ₩9	报 块	I see a second sec							
W ☆≠	! 定义	PROD 英	板块参数 ①		相关项目				
项目		DEV LL	时间分区 ①	ß					
□ 項目	管理	◎ 更新信息 描述:	中文名 时間分区 英文名 ds 数据典型 STRING 数以值 NULL				快快新疆项目,以开始标准规范地建设数据体系		
2		Distant ID	無述 町町分区				新建项目		
= 201	5395	Poc	业务日期 ①	ß					
4 11 11 31	139	DEV LD_ POCut_na	参款 bizdate 参款值 yyyyMMdd						
		◎ 更新信息 描述:	数据域 ①				Q 消费索政副域	+ #	TERRORIZINE
		e PROD	数据成名称 🖕	数据域英	Xã ç	英文编写 💲	描述		操作
anapatin'	URDER.	DEV	大区	eco_enat	ble	en	潮武		c ē
		◎ 更新信息 描述:							
		e PROD LD. © ® PO DEV LD.							

2.业务板块 2.1. 新建业务板块

业务板块是逻辑空间的重要组成部分,是基于业务特征划分的命名空间。在同一个业务板块中可能包含多个不同的项目,业务板块与项目的关系为1:N。本文为您介绍如何新建业务板块。

前提条件

如果业务板块需要添加业务板块管理员,则完成添加Dataphin成员,请参见配置Dataphin成员。

背景信息

系统仅支持超级管理员角色的用户(主账号)创建业务板块。

系统支持创建Dev-Prod模式和Basic模式的业务板块。

模式	描述
Dev-Prod模式	生成相互隔离的Dev和Prod业务板块,更强保障Prod业务板块中的数据安全。如果您 管理诉求较强,数据研发人员较多且能力及分工明确,数据计算存储预算较高,推荐此 模式。
Basic模式	生成独立的Basic业务板块,数据生产过程稳定且便捷。如果您关注数据研发效率,研 发人员职能边界模糊,计算存储预算有限制,推荐此模式。

新建Dev-Prod模式业务板块

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在数仓规划页面,单击左侧导航栏的业务板块。
- 3. 在业务板块页面,单击右上方的新建业务板块。
- 4. 在新建业务板块对话框,选择Dev-Prod模式。
- 5. 单击下一步。
- 6. 在新建业务板块对话框, 配置参数。

新建业务板块 已选	: Dev-Prod模式 ① X
* 公用英文名	LD_ 请编入,英文名至多64字符,允许字母、数字或下划线 ①
* 公用名称	请输入名称,64个文字以内,允许中文、数字、字母、下划线或连字符
Prod	*
Dev	请填写公用部分
Dev	
描述	请输入描述,128字符以内
	0/128
* 图标	e 💲 🔺 🛤 😽 🛤 💙 📼 📥 🗢
业务板块管理员	推荐设置业务板块管理员, 请输入关键字进行搜索, 最多可以设置3位管理员
	上一步取消 确定
参数	描述
公用英文名	业务板块英文名。英文名由字母、数字或下划线(_)组合组成,且不允许超过64个 字符。
	业务板块名称。名称由汉字、数字、字母、下划线(_)或短划线(-)组合组成, 且不允许超过64个字符。
公用名称	⑦ 说明 Dev 业务板块的英文名默认以_dev作为后缀。
描述	对业务板块的简单描述。

参数	描述		
图标	 业务板块的类型。系统支持业务板块类型: • 电商(@ 图标) • 金融(⑤图标) • 云计算(▲ 图标) • 广告营销(▲ 图标) • 竹告营销(▲ 图标) • 文娱(▲ 图标) • 文娱(▲ 图标) • 出行(△ 图标) • 健康(●) • 社交&通信(■ 图标) • 餐饮(△ 图标) • 图标代表教育(● 图标) • 环境(▲) 		
业务板块管理员	单击业务板块管理员显示框区域, 在列表中选择业务板块的管理员。 ⑦ 说明 系统支持最多设置10位业务板块管理员。		

7. 单击确定,完成业务板块的创建建。

新建Basic模式业务板块

- 1. 在业务板块页面,单击右上方的新建业务板块。
- 2. 在新建业务板块对话框,选择Basic模式。
- 3. 单击下一步。
- 4. 在新建业务板块对话框,配置参数。

描述

新建业务板块 已遗	Basic 模式 ①	×	
* 英文名	LD_ 请输入,英文名至多64字符,允许字母、数字或下划线		
* 名称	请输入名称, 64个文字以内, 允许中文、数字、字母、下划线或连字符		
描述	请输入描述, 128字符以内		
	0/12	.8	
* 图标	e s 🔺 🖬 🐱 🛩 🖻 💟 🖬 📥 🗢		
	*		
业务板块管理员	推荐设置业务板块管理员,请输入关键字进行搜索,最多可以设置3位管理员		
	上一步取消 确定		
参数	描述		
英文名	业务板块英文名。英文名由字母、数字或下划线(_)组合组成,且不允许超过64 [,] 字符。	业务板块英文名。英文名由字母、数字或下划线(_)组合组成,且不允许超过64个 字符。	
名称	业务板块名称。名称由汉字、数字、字母、下划线(_)或短划线(-)组合组成, 且不允许超过64个字符。		

对业务板块的简单描述。

参数	描述	
图标	 业务板块的类型。系统支持业务板块类型: 电商(●图标) 金融(●图标) 云计算(●图标) 广告营销(●图标) 竹告营销(●图标) 文娱(●图标) 出行(●图标) 出行(●图标) 4 社交&通信(●图标) 餐饮(●图标) 餐饮(●图标) 药育(●图标) 环境(●图标) 	
业务板块管理员	单击业务板块管理员显示框区域,在列表中选择业务板块的管理员。	

5. 单击确定,完成业务板块的创建。

2.2. 新建数据域

数据域用于存放同一业务板块内不同意义的指标,如商品域、交易域、会员域等。本文为您介绍如何新建数 据域。

背景信息

关于数据域的定义和划分,详情请参见划分数据域。

系统仅支持超级管理员和项目管理员角色的用户,新建数据域。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在规划页面,单击左侧导航栏的业务板块。
- 3. 在**业务板块**页面,选择需要创建数据域的业务板块后,单击新建数据域。
- 4. 在新建数据域对话框,配置参数。

参数	描述
数据域名称	填写数据域的名称。 数据域名称 由汉字、数字、字母及下划线(_)或短划线 (-)。
	⑦ 说明 数据域名称不允许超过64个字符。
	填写数据域的英文名。 数据域英文名 由字母、数字或下划线(_)。
数据域英文名	⑦ 说明 数据域英文名不允许超过64个字符。
	填写英文缩写,且业务板块内的英文名不支持重复。 英文缩写 由字母、数字或下划 线(_)组合组成。
英文缩写	⑦ 说明 英文缩写不允许超过10个字符。
描述	填写对数据域简单的描述。

5. 单击确定,完成数据域的创建。

2.3. 编辑板块参数

板块参数用于定义业务板块内的时间分区和业务日期。本文为您介绍如何编辑板块参数。

前提条件

完成业务板块的创建,详情请参见新建业务板块。

背景信息

- 编辑板块参数的权限限制,详情请参见数仓规划权限列表。
- 如果业务板块被引用,则板块参数的英文名和数据类型不支持编辑。
- 板块参数定义了业务板块内通用和统一的参数信息,包括时间分区和业务日期:
 - 时间分区: 定义了逻辑表时间分区字段的默认名称、数据类型等。
 - **业务日期**: 定义了任务调度的业务日期, 默认为bizdate。

编辑时间分区

- 1. 进入数仓规划页面,请参见入口介绍。
- 2. 在规划页面,单击左侧导航栏的业务板块。
- 3. 在业务板块页面,单击需要编辑板块参数的业务板块。
- 4. 在板块参数区域,单击时间分区后的 图标。
- 5. 根据编辑时间分区对话框提示, 配置参数。

编辑时间分区			×
* 中文名	时间分区		
* 英文名	ds		
* 数据类型	STRING		~
* 默认值	NULL		
描述	时间分区		4/128
		取消	确定

参数	描述	
中文名	默认为时间分区。中文名由汉字、字母、数字、下划线(_)或短划线(-)组合组 成。	
	⑦ 说明 中文名 不能超过64个字符。	
	默认为ds。英文名由字母、数字或下划线(_)组成。	
英文名	⑦ 说明 英文名不能超过64个字符。	
	选择数据类型。系统支持的数据类型如下:	
数据类型		
	• DOUBLE	
	• DAT ET IME	
默认值	填写时间分区的默认值,如果没有默认值,则填写NULL。	
描述	对时间分区简单的描述。	

6. 单击**确定**,完成时间分区的编辑。

编辑业务日期

1. 在业务板块页面,单击需要编辑板块参数的业务板块。

- 2. 在**板块参数**区域,单击**业务日期**后的 図标。
- 3. 在编辑业务日期对话框,单击参考值显示框区域,在下拉列表中选择业务日期的格式。
- 4. 单击确定,完成业务日期的编辑。

3.管理统计周期

公共定义仅支持定义统计周期。统计周期即统计数据的时间范围,例如最近7天、最近30天等。本文为您介 绍如何新建统计周期。

前提条件

完成计算引擎类型设置,详情请参见<mark>设置Dataphin实例的计算引</mark>擎。计算引擎类型设置完成后,系统会自动 初始化一批统计周期(即系统内置的常用统计周期)。

背景信息

新建统计周期的权限限制,详情请参见数仓规划权限列表。

新建统计周期

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在规划页面,单击左侧导航栏的公共定义。
- 3. 在公共定义页面,单击新建统计周期。
- 4. 在新建统计周期对话框, 配置参数。

新建统计周期	3	×
* 统计周期	请输入名称,64字符以内,如最近7天,允许中文数字字母及下划线或-	
* 英文缩写	请输入统计周期英文名,10字符以内,如7d。允许字母、数字或下划线	
描述	请输入统计周期描述,128字符以内	0/128
表达式		
	会業が '\${bizdata}'	
开始时间	● 参x 3{Dizuale} 函数表达式	
结束时间	→ 参数 请输入参数	
	● 函数表达式 lastNDate	
	取消	确定
参数	描述	

参数	描述
统计周期	统计周期名称由汉字、数字、字母、下划线(-)或短划线(-)组合组成。
	⑦ 说明 统计周期名称不能超过64个字符。
	统计周期英文名由字母、数字或下划线(-)组合组成。
英文缩写	⑦ 说明 统计周期名称不能超过64个字符。
描述	填写对统计周期的简单描述。
表达式	单击 表达式 后面的 表达式参数说明 ,查看系统支持的参数及其用法。
开始时间	配置统计周期的 开始时间 : ○选中参数后,填写参数的表达式。 ○选中函数表达式后,单击─图标,在下拉列表中选择函数。系统自动选中函数对 应的表达式。
结束时间	配置统计周期的 结束时间: 选中参数后,填写参数的表达式。 选中函数表达式后,单击 图标,在下拉列表中选择函数。系统自动选中函数对应的表达式。

? 说明

○ 系统不支持创建相同的时间周期(即**开始时间**和结束时间分别相同)。

○ 系统不支持创建相同的统计周期或英文缩写。

5. 单击确定,完成统计周期的创建。

编辑统计周期

1. 在**公共定义**页面,查询统计周期。

您可以在搜索框中,输入统计周期名称或英文缩写的关键字,快速查询统计周期。

- 2. 单击需要编辑的统计周期操作列下的 图标。
- 3. 在编辑统计周期对话框,编辑参数,详情请参见新建统计周期。

⑦ 说明 如果统计周期被引用,则英文缩写不支持编辑。

4. 单击确定,完成统计周期的编辑。

删除统计周期

- 1. 在数仓规划页面,单击左侧导航栏的公共定义。
- 在公共定义页面,查询统计周期。
 您可以在输入框中,输入统计周期名称或英文缩写的关键字,快速查询统计周期。
- 3. 单击需要删除的统计周期后操作列下的 圖图标。
- 4. 在确认对话框中,单击确定,完成统计周期的删除。

4.计算源

4.1. 概述

在您完成系统初始化后,即可为Dataphin添加计算源。计算源用于绑定计算引擎项目和Dataphin项目,为 Dataphin项目提供处理任务的计算源。

功能概述

Dataphin支持添加MaxCompute、AnalyticDB For PostgreSQL、Flink计算源。

计算源类型	特点	描述
MaxCompute	离线计算源	智能研发版支持的计算引擎为MaxCompute,即根据业务需求您可以 选择是否需要新建MaxCompute计算源。 绑定了MaxCompute计算源的项目,支持规范建模、即席查询、 MAXC任务、通用脚本等功能。
AnalyticDB For PostgreSQL	离线计算源	基础研发版支持的计算引擎为AnalyticDB For PostgreSQL,即根据 业务需求选择是否需要新建AnalyticDB For PostgreSQL计算源。 绑定了AnalyticDB For PostgreSQL计算源的项目,支持ADB任务、即 席查询、通用脚本等功能。
Flink	实时计算源	基础研发版和智能研发版均默认支持Flink类型的实时计算引擎,即根据业务需求您可以选择是否需要新建Flink计算源。 绑定了Flink计算源的项目,支持Flink任务。

使用场景

场景说明	描述	操作指导
添加MaxCompute计算源至项目	业务场景中需要建模、构建即席查询 和MAXC任务或通用任务。	 新建MaxCompute计算源。 添加计算源至项目,详情请参 见创建Basic项目或管理项目空 间的权限和计算源。
添加AnalyticDB For PostgreSQL计 算源至项目	业务场景中需要构建ADB任务、即席 查询或通用脚本任务。	 新建AnalyticDB For PostgreSQL计算源。 添加计算源至项目,详情请参 见创建Basic项目或管理项目空 间的权限和计算源。

场景说明	描述	操作指导
添加Flink计算源至项目	业务场景中需要构建Flink任务。	 新建Flink计算源。 添加计算源至项目,详情请参 见创建Basic项目或管理项目空 间的权限和计算源。

4.2. 新建MaxCompute计算源

MaxCompute计算源用于绑定Dataphin项目空间和MaxCompute项目,为Dataphin项目提供处理离线计算任务的计算源。如果Dataphin系统的计算引擎设置为MaxCompute,则只有项目空间添加了MaxCompute计算源,才支持规范建模、即席查询、MAXC任务、通用脚本等功能。本文为您介绍如何新建MaxCompute计算源。

前提条件

已设置Dataphin计算引擎为MaxCompute,请参见设置Dataphin实例的计算引擎。

背景信息

系统支持超级管理员和项目管理员角色的账号新建计算源。更多信息,请参见数仓规划权限列表。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在数仓规划页面,单击左侧导航栏的计算源。
- 3. 在计算源页面, 单击页面右上方的新增计算源。
- 4. 在新建计算源对话框, 配置参数。

新建计算源			\times
* 计算源类型	MaxCompute	ш ~	
* 计算源名称	请输入计算源名称		
计算源描述	请输入计算源描述	0/128	
* Endpoint	http://service.cn.maxcompute.aliyun.com/api		
* Project Name	请输入英文项目名称		
* Access ID	请输入鉴权Access ID,为保证任务正常执行,请确保有所需数据权限		
* Access Key	请输入鉴权Access Key		(j)
	测试连接 取 消		

参数	描述
计算源类型	选择为MaxCompute。
计算源名称	计算源名称由汉字、数字、字母、下划线(_)或短划线(-)组合组成。
计算源描述	填写计算源的简单描述。
Endpoint	默认为计算引擎的Endpoint,不支持修改。
Project Name	填写在DataWorks中创建的MaxCompute项目名称。 请参见工作空间列表,获取MaxCompute项目的 Project Name 。
Access ID、Access Key	填写可以访问MaxCompute项目的账号的AccessKey ID和AccessKey Secret。 在用户信息管理页面,获取账号的AccessKey ID和AccessKey Secret。

- 5. 单击测试连接,测试连接的计算源。
- 6. 测试成功后,单击提交。

完成创建MaxCompute计算源后,即可为项目绑定MaxCompute计算源,请参见创建Basic项目、创建 Prod和Dev项目。

4.3. 新建AnalyticDB For PostgreSQL计

算源

AnalyticDB For PostgreSQL计算源用于绑定Dataphin项目和AnalyticDB For PostgreSQL项目,为Dataphin项目提供处理离线计算任务的计算源。如果Dataphin系统的计算引擎设置为AnalyticDB For PostgreSQL,则只有添加了AnalyticDB For PostgreSQL计算源的项目,支持ADB任务、即席查询、通用脚本等功能。本文为您介绍如何新建AnalyticDB For PostgreSQL计算源。

前提条件

- Dat aphin的计算引擎设置为AnalyticDB For PostgreSQL,详情请参见设置Dat aphin实例的计算引擎。
- 已获取AnalyticDB For PostgreSQL计算引擎数据库对应的Schema、用户名和密码,详情请参见。

背景信息

如果Dataphin系统的计算引擎设置为AnalyticDB For PostgreSQL,则项目绑定不同计算源类型对应的功能说明如下。

项目绑定计算源类型	支持的功能
项目仅绑定了Flink类型的 计算源	Flink任务
项目绑定了AnalyticDB For PostgreSQL和Flink类 型的计算源	即席查询、ADB任务、通用脚本

新建AnalyticDB For PostgreSQL计算源的权限限制,详情请参见数仓规划权限列表。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在数仓规划页面,单击左侧导航栏的计算源。
- 3. 在计算源页面, 单击新增计算源。
- 4. 在新建计算源对话框, 配置参数。

参数	描述
计算类型	选择为AnalyticDB For PostgreSQL。
计算源名称	计算源名称由汉字、数字、字母、下划线(_)或短划线(-)组合组成。
计算源描述	填写计算源的简单描述。
JDBC URL	默认为计算引擎的JDBC URL,不支持修改。
Schema	填写计算引擎数据库对应的Schema连接信息。
用户名	填写AnalyticDB For PostgreSQL计算引擎数据库的登录用户名。
密码	填写AnalyticDB For PostgreSQL计算引擎数据库的登录密码。

- 5. 单击测试连接,测试连接的计算源。
- 6. 测试成功后,单击提交。

完成创建AnalyticDB For PostgreSQL计算源后,即可为项目绑定AnalyticDB For PostgreSQL计算源,详 情请参见创建Basic项目或管理项目空间的权限和计算源。

4.4. 新建Flink计算源

Flink计算源用于承载Dataphin项目下基于Flink的计算资源(以CU为单位,含资源队列),只有绑定了Flink计 算源的项目,才支持基于Flink引擎研发计算任务。本文将为您介绍如何新建Flink计算源。

前提条件

• 获取Flink计算源的Project Name, 详情请参见Flink项目。

↓ 注意 当前, Dat aphin支持Blink独享集群。

• 在用户信息管理页面,获取您访问阿里云官网的Access ID和Access Key。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在数仓规划页面,单击左侧导航栏的计算源。
- 3. 在计算源页面,单击新增计算源。
- 4. 在新建计算源对话框, 配置参数。

新建计算源					×
* 计算源类型	FLINK				
* 计算源名称	请输入计	算源名称			
计算源描述	请输入计	算源描述			0/128
* Project Name	请输入英	文项目名称			
* Access ID	请输入鉴	权Access ID,为保证任	务正常执行,请确保有所需	数据权限	
* Access Key	请输入鉴	RAccess Key			0
			测试连接	取消	提交
参数		描述			
计算类型		选择为FLINK。			

参数	油还
计算类型	选择为FLINK。
计算源名称	填写计算源的名称。
计算源描述	填写计算源的描述。
Project Name	填写项目的英文名称。
Access ID	填写鉴权的Access ID。
Access Key	填写鉴权Access Key。

? 说明

- 。新建实时计算源时,仅允许添加同一个RAM主账号下的Real-time Compute项目。
- 新建实时计算源时,请尽量填写Real-time Compute项目管理员的Access Key,以保证连接 无误。
- 新建实时计算源时,请尽量不修改Real-time Compute项目的Access Key,保证元数据正常采集。
- 计算源创建完成后,计算源类型不支持更改。

5. 单击测试连接,测试连接的计算源。

6. 单击提交,完成Flink计算源的创建。

完成创建Flink计算源后,即可为项目绑定Flink计算源,详情请参见创建Prod和Dev项目或创建Basic项 目。

4.5. 创建加速计算源

Dataphin支持为MaxCompute计算引擎的离线数仓添加加速计算源,后续就可以加速数据服务中的Dataphin 数据源的查询速度,以满足对查询速度有要求的业务场景。在您为离线数仓添加加速计算源前,需要新建加 速计算源。本文为您介绍如何新建加速计算源。

前提条件

在开始执行操作前,请确认您已开通Hologres,详情请参见购买Hologres。

背景信息

Hologres支持加速MaxCompute项目下数据查询速度,关于Hologres更多内容请参见什么是实时数仓 Hologres。如果数据服务中的Dataphin数据源对数据查询速度有要求,但当Dataphin的计算引擎为 MaxCompute时,任务运行速度比较慢,不能很好地满足对查询速度有要求的业务场景。

因此,您可以创建计算源并添加至已绑定MaxCompute计算源的项目中,后续就可以对计算空间选择该项目的Dataphin数据源进行加速业务数据的查询速度。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在数仓规划页面,按照下图操作指引,进入新建计算源。

≡ Datap	hìn·规划		Ê	f 🖉 a & 📀 💽
业务 ①	计算源 ①			C + 新聞计算题
📩 业务板块	Q. 搜索计算座			
W 公共定义	计算透映型 MaxCompute FLINK Hologres			
项目		*		
□ 項目管理	计算语名称 负责人 已绑定项目	台環信息	连接信息	攝作
·····································	● (K_lest_gray_bl_dev Stät: MAX_COMPUTE	台建用户: 3021-10-25 17-43 通知时间: 2021-10-25 17-43 更新时间: 2021-10-25 17-43	Endpoint: http://service.cr	6 E F
(1)	fig. test_gray_bi sells: MAX_COMPUTE	创 建用户: 创建时间:2021-10-25 17-42 更新时间:2021-10-25 17.42	Endpoint http://service.cn inc.com/spi Project_Name:	6 12 2 1

3. 在新建计算源对话框, 配置参数。

新建计算源		X
* 计算源类型	Hologres	
* 计算源名称	请输入计算源名称	
计算源描述	清輸入计算源描述	0/128
* JDBC URL	请输入JDBC连接地址,如jdbc:postgresql://host:port/database	
Schema	请输入Schema信息,默认为public	
* 用户名	请输入鉴权用户名,为保证任务正常执行,请确保有所需数据权限	
* 密码	请输入鉴权密码,为保证任务正常执行,请确保有所需数据权限	Ø
* 访问方式 * 区域	○ 公网IP ● VPC网路 请选择Region	
* VPC ID	请输入VPC ID	
* 实例 ID	请输入实例ID	
	测试连接 取 消	提交
2		

参数	描述
计算源类型	选择为Hologres。
计算源名称	命名规则如下: • 包含汉字、数字、字母、下划线(_)或短划线(-)。 • 不能超过64个字符。
计算源描述	填写对计算源的描述。

参数	描述			
JDBC URL	填写连接Hologres的JDBC URL,例如jdbc:postgresql://{域名}/{数据库名称)。czg时数仓Hologres控制台的实例详情页面,获取对应网络下的域名。() <tr< th=""></tr<>			
Schema	填写连接Hologres对应的Schema。			
用户名、密码	填写可以访问Hologres实例账号的AccessKey ID和AccessKey Secret。在用户信息 管理页面,获取当前账号的AccessKey ID和AccessKey Secret。			
访问方式	选择对应的访问方式: 。 Hologres数据源所在网络环境为VPC网络,则访问方式选择为 VPC网络 。 。 Hologres数据源所在网络环境为公网,则访问方式选择为 公网IP 。			

 当访问方式选择为VPC网路,则需要配置如下参数: 区域:仅支持选择当前Dataphin实例所在的地域。 ② 说明 仅支持创建与Dataphin实例在同一地域的Hologres计算源。 VPC ID:填写Hologres实例的VPC ID。 实例ID:填写Hologres实例的ID。 在实时数仓Hologres控制台的实例详情页面,查看实例ID和VPC ID, VPC ID格式为 (实例ID)-{实例所在地域}-VPC, 例如hgprecn-cn-8ed2k5u4****-cn-hangzhou-vpc。 	参数	描述
SUBBOMOND Edited (#4687) Edited (#468	参数 区域、VPC ID、实例ID	 描述 当访问方式选择为VPC网路,则需要配置如下参数: • 区域: 仅支持选择当前Dataphin实例所在的地域。 ⑦ 说明 仅支持创建与Dataphin实例在同一地域的Hologres计算源。 • VPC ID: 填写Hologres实例的VPC ID。 • 文例ID: 填写Hologres实例的ID。 在实时数仓Hologres控制台的实例详情页面,查看实例ID和VPC ID, VPC ID格式为(实例ID)-{实例所在地域}-VPC, 例如hgprecn-cn-8ed2k5u4****-cn- hangzhou-vpc。
P P		NBSRE 地域の 学校名目 アス RMARE 10

4. 单击测试连接。

5. 连接测试通过后,单击提交。

后续步骤

完成创建加速计算源后,即可为已绑定MaxCompute计算源的项目添加Hologres加速计算源,详情请参见创建Basic项目或创建Prod和Dev项目。

4.6. 管理计算源

Dataphin支持管理系统中的计算源,包括测试、修改和删除计算源及转移计算源的责任人。本文为您介绍如 何测试、修改和删除计算源及如何转移计算源的责任人。

前提条件

已完成计算源的创建,详情请参见新建MaxCompute计算源、新建AnalyticDB For PostgreSQL计算源、新建 Flink计算源。

背景信息

- 测试、修改和删除计算源及转移计算源的责任人的权限限制,详情请参见数仓规划权限列表。
- 已被项目绑定的计算源不支持删除。

测试计算源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在数仓规划页面,单击左侧导航栏的计算源。
- 3. 单击需要连接测试的计算源所在行的操作列下的@图标。

修改计算源

- 1. 在计算源页面,单击需要修改计算源操作列下的 图标。
- 2. 在编辑计算源对话框,修改配置参数,详情请参见新建MaxCompute计算源、新建AnalyticDB For PostgreSQL计算源和新建Flink计算源。
- 3. 单击测试连接,测试计算源是否可以连通。
- 4. 测试成功后,单击提交,完成计算源的修改。

转交责任人

- 1. 在计算源页面, 单击需要转交责任人的计算源操作列下的 图标。
- 2. 在转移负责人对话框,选择负责人。
- 3. 单击确定。

删除计算源

- 1. 在计算源页面,单击需要删除计算源操作列下的 圖图标。
- 2. 在确认对话框中,单击确定,完成计算源的删除。

5.项目空间 5.1. 创建Prod和Dev项目

项目是Dataphin的基本组织单元,是进行多用户隔离和访问控制的主要边界。您开通Dataphin服务后,需要通过项目使用Dataphin。为了保障数据生产的稳定性及对数据研发流程的强管控,您可以创建Dev-Prod模式的项目。本文为您介绍如何创建Dev-Prod模式的项目。

背景信息

Dataphin支持两种开发模式的项目,详细说明见下表。

项目的模式	描述	
	完成项目的创建后,系统默认生成了开发环境 Prod项目保障了生产环境项目中的数据安全。 且能力及分工明确,数据计算存储预算较高,	(Dev项目)和生产环境(Prod项目)。 如果您管理诉求较强,数据研发人员较多 推荐此模式。
	zz_test	zz_test C
	ky_encrypt_project(ky_test01) 暂无描述	ky_encrypt_project_dev(ky_test 1521
Dev-Prod模式	 业务板块: LD_zz_test 项目成员: 计算源: ■ ky_encrypt み 元 前往配置 创建时间: 2021-02-04 11:41:59 更新时间: 2021-03-19 20:04:32 信息设置 成员管理 ○ 超级管理员 进入工作区 ≫ 	 业务板块: LD_zz_test_dev 项目成员: 计算源: ■ ky_encrypt → 无前往配置 创建时间: 2021-02-04 11:41:59 更新时间: 2021-03-19 20:04:32 信息设置 成员管理 企 超级管理员 发布管理》 进入工作区》

项目的模式	描述
	完成项目的创建后,系统默认生成了Basic项目(开发和生产一体的项目)。Basic项目 的数据生产过程稳定且便捷。如果您关注数据研发效率,研发人员职能边界模糊,计算 存储预算有限制,推荐此模式。Basic项目的更多信息,请参见创建Basic项目。
	test_basic1220(test) 暂无描述 业务板块: LD_test_bizunit 项目成员: 人 计算源: ■ test_basic12 登 无 前往配置
	创建时间: 2019-12-20 17:06:03
Basic模式	信息设置 成员管理 ◇ 超级管理员 进入工作区≫
	 ↓ 注意 ● Basic模式不支持升级为Dev-Prod模式,且Basic模式存在直接变更生产风险,请您谨慎选择。 ● 如果选择了Basic模式,请您谨慎管理项目成员,以控制数据生产的稳定性。

前提条件

在开始执行操作前,请确认您以满足以下条件:

- 后续需要研发流批一体任务,则在创建项目前需要完成Flink计算源的创建,请参见新建Flink计算源。
- Dataphin的计算引擎为AnalyticDB for PostgreSQL, 且数据研发过程中需要使用数据集成、即席查询、 AnalyticDB for PostgreSQL的SQL计算任务等功能,则在创建项目前需要完成AnalyticDB for PostgreSQL 计算源的创建,请参见新建AnalyticDB For PostgreSQL计算源。
- Dataphin的计算引擎为MaxCompute,且数据研发过程中需要使用规范建模、即席查询、MaxCompute计算任务等功能,则在创建项目前需要完成MaxCompute计算源的创建,请参见新建MaxCompute计算源。
- Dataphin支持添加Hologres加速计算源至已绑定MaxCompute计算源的项目,加速数据服务的查询速度。 如果您需要加速业务数据的查询速度,则在创建项目前需要创建加速计算源,详情请参见创建加速计算 源。
- Dataphin对接了机器学习PAI,为您提供了机器学习PAI基础的算法调度能力。在开始使用机器学习任务前,需要开通机器学习并已创建工作空间,详情请参见开通并创建默认工作空间。

使用限制

- 系统支持超级管理员角色和板块管理员角色的账号创建项目。
- 系统仅支持项目管理员和板块管理员角色的账号,关闭和开启数据表读写的权限申请。

如何获取板块管理员角色,请参见新建业务板块。

注意事项

Dev项目和Prod项目中配置的计算源需保持一致,否则可能导致生产环境的任务调度失败。

操作流程

操作流程	描述
步骤一:选择项目类型	选择项目的模式为Dev-Prod模式。 完成项目的创建后,系统自动生成互相关联的Dev环境和Prod环境,以强管控数据生产 过程、生产的数据。
步骤二:配置项目的名称 和命名空间	配置项目的基本信息,包括项目的名称和命名空间。
步骤三:配置项目的计算	计算源为项目的提供计算和存储资源,同时您也可以开启机器学习,后续您可以导入机
源及开启机器学习	器学习的工作流在Dataphin进行调度运维。
步骤四:配置项目的沙箱	配置项目中的Shell、Python任务可以访问的IP地址或域名。
白名单	您也可以通过 信息设置 添加沙箱白名单,请参见 <mark>添加沙箱白名单</mark> 。
步骤五:配置项目的数据	Dataphin支持通过项目粒度的方式,管控项目中的元表和物理表的读取和写入数据的权限申请。以规避审批者误操作,导致的数据泄露与安全问题。
读写申请权限	您也可以通过 信息设置 配置项目的数据读写申请权限,请参见 <mark>设置数据读写申请权限</mark> 。

步骤一:选择项目类型

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在数仓规划页面,按照下图操作指引,进入新建项目对话框。



3. 在新建项目对话框,选择创建项目的模式为Dev-Prod模式。

0		
	Day Band tilt t	
dataphin	Dev-Prod 候式 生成互相关联的Dev环境和Prod环境,以强 管控数据生产过程、生产的数据。	Basic 模式 生成独立灵活的Prod环境,数据生产过程相 对稳定而便捷,生产的数据可管控
data part	如您管理诉求较强,数据研发人员较多目能 力及分工明确,数据计算存储预算较高,推 荐此模式: Dev 环境和 Prod 环境逻辑隔 离,可加强对数据研发流程的管控,强控制 生产数据的权限,最大化保障数据生产稳定 性。但是研发流程灵活及便捷性较低,对研 发人员工具使用题练度变在这次语	如您关注数据研发效率而非强管控,研发人员职能边界相对模糊,计算存储预算有限制,推荐此模式:开发与运维一体化的Prod环境,可敏捷管理数据研发流程,强控制生产数据的写权限,平衡数据生产稳定性与灵活性
0.000	双份计具存储资源。	

4. 单击下一步。

步骤二:配置项目的名称和命名空间

在新建项目对话框的名称设置和命名空间区域,配置参数。

新建项目	已选: Dev-Prod 模式 ①				Х
名称设置		命名空间			
* 公用名称	请输入名称,64个文字以内,允许中文、数字、字母、下划线或连字符	业务板块			Ya
* 公用英文名	请输入名称,64个文字以内,允许数字、字母、下划线或连字符	* 空间类型	应用层		
		上一步	取消	确定	

区域	参数名称	描述				
	公用名称	填写项目的公用名称。命名规则如下: • 包含汉字、数字、字母、下划线(_)或短划线(-)。 • 不支持以LD_开头。 ⑦ 说明 业务板块英文名默认以LD_作为前缀,以免引用 逻辑表和物理表时出现冲突。				
		• 不能超过64字符。				
名称设置	公用英文名	填写项目的公用英文名。命名规则如下: 包含字母、数字、下划线(_) 不支持以LD_开头。 				
		⑦ 说明 业务板块英文名默认以LD_作为前缀, 以免引用 逻辑表和物理表时出现冲突。				
		 不能超过64个字符。 开发环境项目英文名默认以_dev作为后缀。 ###: DEV ###: DEV				
命名空间	业务板块	选择项目所属的业务板块。				
		↓ 注意 如果您购买的Dataphin实例为基础研发版,则系统 不支持选择业务板块。				
	空间类型	 区分项目开发任务及产出数据特征,默认为应用层。空间类型包括3种: 基础层:来自各个业务数据库的原始数据存储,是后续所有数据建设的来源与基础,也称为垂直数据中心。 公共层:基于各业务数据,抽象公共的主题,沉淀标准、规范、可共用的数据,是连接基础层与应用层的中间桥梁,也称为公共数据中心。 应用层:面向业务需求,定义生成可应用于不同场景的个性化、多样化的数据指标。 				

步骤三: 配置项目的计算源及开启机器学习

1. 在**新建项目**对话框的**项目设置**区域,如果Dataphin初始化为MaxCompute,则需要选择**计算源类型**。 根据Dataphin初始化的计算引擎不同,支持选择的计算源类型也不同,详细说明如下:

- Dataphin初始化为MaxCompute,则需要选择计算源类型,支持选择的计算源类型包括 MaxCompute、Flink、Hologres。
 - MaxCompute: MaxCompute是一种快速、完全托管的TB或PB级数据仓库解决方案,能够更快速 为您解决海量数据计算问题,有效降低企业成本,并保障数据安全,详情请参见MaxCompute。
 - Flink: 添加Flink计算源后,您可以在Dataphin中进行流式计算任务开发,详情请参见阿里云实时计算。
 - Hologres: 您可以创建计算源并添加至已绑定MaxCompute计算源的项目中,后续就可以对计算空间选择该项目的Dataphin数据源进行加速业务数据的查询速度,详情请参见什么是实时数仓Hologres。
- Dataphin初始化为AnalyticDB for PostgreSQL,则不需要选择计算源类型。

2. 配置计算源的参数。

○ 计算源类型选择为MaxCompute、Flink、Hologres。

项目1: DEV					项目2: P	ROD				
* MAXC	请选择批处理计算源				* MA	XC 请	选择批处理计算源			
doctarphin	同步创建数据源 ①						同步创建数据源 ①	D		
* FLINK	请选择流处理计算源				* FLI	NK 请	选择流处理计算源			
* HOLO	请选择加速计算源				* HO	LO 请	选择加速计算源			
英文名	test_dev	名称测试	_开发		英	文名 test	:	名称	测试	
描述请求	前入项目描述,128字符以内	为			描述	请输入项	间描述,128字符	以内		
. Sumption										
			0/128	11						0/128
开发操作	代码编辑/发布管理/调度				生产摄	晶作 代码	编辑(查看)/调問	÷.		
业务板块	未选,非必选				业务板	动块 未选	,非必选			

- 。 计算源类型默认为AnalyticDB for PostgreSQL。
- 开启了机器学习。

↓ 注意 仅计算源类型为MaxCompute的项目,支持开启机器学习。

 * 计算源类型 ▼ MAX(① 请至少 	C 🗌 FLINK 🗌 HOLO ①)选择一个计算源,生产/开发计算	章源设置同步		* 机器学习 (开)) PAI]			
项目1: DEV				项目2: PROD				
* MAXC 请选择批处理计算源				* MAXC 请选择	释批处理计算源			
同步	创建数据源 ①			同步	创建数据源 ①			
* PAI地域	请选择PAI地域			* PAI地域	请选择PAI地域			
* AccessKey ID	请输入 AccessKey ID			* AccessKey ID	请输入 AccessKey ID			
* AccessKey Secret	请输入 AccessKey Secret		Ø	* AccessKey Secret	请输入 AccessKey Secret		Ø	
* PAI项目名称	请选择PAI项目名称			* PAI项目名称	请选择PAI项目名称			
英文名 test_de	v 名称	测试_开发		英文名 test	名称	测试		
描述 请输入项目指	苗述, 128字符以内			描述 请输入项目	苗述, 128字符以内			
daaraan daa		dataphin ()/128 //				0/128	
开发操作 代码编辑	开发操作 代码编辑/发布管理/调度				生产操作代码编辑(查看)/调度			
业务板块 未选, 非	印心选			业务板块 未选, 目	起选			

计算源的参数配置,详细说明如下。

参数	描述							
机器学习	注意 仅计算源类型为MaxCompute的项目,支持开启机器学习。 Dataphin对接了机器学习PAI,为您提供了机器学习PAI基础的算法调度能力。在机器学习平台中,创建可视化建模的工作空间时,资源选择为基于MaxCompute的计算资源组,详情请参见资源管理概述。							
计算源	选择您已创建相应计算源类型的计算源,项目创建成功后,不支持修改计算源。							
	系统支持为项目细空MaxCompute_ApplyticDB For DectoreSOI 计管项时_日选							
同步创建数据源	中 同步创建数据源 后,Dataphin自动创建对应类型的数据源。创建数据源时配置的 信息,请参见附录:同步创建数据源的机制说明。							
参数	描述							
-----------	--	------------------	--	--	--	--	--	--
英文名	Dev项目的英文名默认为项目公用英文名标 新建项目 已志: Dev-Prod 儀式 ② 各称设置 ・公用宏称 鄭武 ・公用宏称 鄭武 ・公用宏称 鄭武 ・公用宏な Test ・公用宏な Test ・公用定な Test ・公用成本 丁香田 ・公用定な Test ・公用方法 Test ・公用定な Test ・公用定な Test ・公用定な Test ・公用定な Test ・公用定な Test ・公用定な Test ・日本 Test ・日本	51増加_dev。						
名称	Dev项目的名称默认为项目的 公用名称 后增	曾加 _开发 。						
描述	填写对Dev、Prod项目的描述。							
开发操作或生产操作	开发操作默认为代码编辑、发布管理、调度;生产操作默认为代码编辑(查看)、 调度,不支持修改,均不支持。							
业务板块	默认为 命名空间 区域设置的业务板块名称, 缀_dev,不支持修改。	且Dev项目的业务板块名称增加后						

步骤四:配置项目的沙箱白名单

1. 在**其他设置的沙箱白名单**区域,按照下图操作指引,添加该项目下输入Shell、Python任务需要访问的 IP地址或域名。

沙箱白名单 ①							1 + 新建
访问地址				端口			操作
damakan							
(1997)	6 9000000	dataphin	2	智无数: 注於)		dataphin	
请输入访问地	址			请辅入	「「」		✓ X

2. 单击确定, Shell、Python任务可以访问对应的IP地址或域名。

对于已添加的沙箱白名单,您也可以单击**操作**列下的画图标,删除后项目下的Shell、Python任务,将无 法访问对应的IP地址或域名。

步骤五: 配置项目的数据读写申请权限

1. 在其他设置对话框的通用功能控制区域,选择是否需要开启数据读写权限申请权限。

其他设置 沙箱白名单 ①							+ 新建
访问地址			端口				操作
			暫无数据				
通用功能控制							
* 数据读权限申请	关闭 ①	* 数据写权限申请	; 💿 开启 🛛	关闭 ①			
					上一步	取消	确定

开启或关闭数据写权限申请:

- 关闭数据写权限申请后,当前项目下的物理表及元表不支持申请写入权限,同时Dataphin不收回已申 请过的写入权限。
- 开启数据写权限申请后,当前项目下的物理表及元表支持申请写入权限。申请写入权限的更多信息, 请参见申请权限。

开启或关闭数据读权限申请:

- 关闭数据读权限申请后,当前项目下物理表及元表不支持申请查询权限,同时Dataphin不收回已申请 过的查询权限。
- 开启数据读权限申请后,当前项目下的物理表及元表支持申请查询权限。申请查询权限的更多信息, 请参见申请权限。
- 2. 单击确定,完成Prod项目和Dev项目的创建。

附录:同步创建数据源的机制说明

同步创建数据源时,系统自动配置的参数信息如下。

参数	描述
数据源类型	根据计算源类型配置为相应的数据源类型,包括MaxCompute或AnalyticDB For PostgreSQL。
数据源名称	默认配置为Prod计算源英文名。
数据源描述	默认配置为Prod项目的描述。
数据源配置	默认选择为 生产+开发数据源 。
开发数据源	需要配置Endpoint、Project Name、Access ID、Access Key参数,默认填写Dev 项目绑定的计算源的连接信息。
生产数据源	需要配置Endpoint、Project Name、Access ID、Access Key参数,默认填写 Prod项目绑定的计算源的连接信息。

同步创建数据源过程中,系统校验规则如下:

• 数据源重名校验:如果已存在同名数据源,则数据源创建失败,项目创建不阻断。

• 数据源测试连接: 如果测试连接失败, 则数据源创建失败, 项目创建不阻断。

后续步骤

创建完项目后,即可进入工作区开始数据的研发。如何基于Dataphin开发数据,请参见数据开发。

zz_test <	zz_test C
ky_encrypt_project(ky_test01) 暂无描述	ky_encrypt_project_dev(ky_test 1521
业务板块: LD_zz_test 项目成员:	业务板块: LD_zz_test_dev 项目成员:
计算源: 四 ky_encrypt → 无 前往配置 创建时间: 2021-02-04 11:41:59	计算源: 四 ky_encrypt → 无 前往配置 创建时间: 2021-02-04 11:41:59
更新时间: 2021-03-19 20:04:32	更新时间: 2021-03-19 20:04:32
信息设置成员管理	信息设置成员管理
☞ 超级管理员 进入工作区》	☞ 超级管理员 发布管理》 进入工作区》

5.2. 创建Basic项目

项目是Dataphin的基本组织单元,是进行多用户隔离和访问控制的主要边界。您开通Dataphin服务后,需要通过项目使用Dataphin。在关注数据研发效率,且计算存储预算有限制的场景中,推荐您创建Basic模式的项目。Basic项目数据生产过程稳定且便捷。本文为您介绍如何创建Basic模式的项目。

背景信息

Dataphin支持两种开发模式的项目,详细说明见下表。

描述

项目的模式

项目的模式	描述
	完成项目的创建后,系统默认生成了Basic项目(开发和生产一体的项目)。Basic项目 的数据生产过程稳定且便捷。如果您关注数据研发效率,研发人员职能边界模糊,计算 存储预算有限制,推荐此模式。 Lest_bizunit BASIC test_basic1220(test) 暂无描述 业务板块: LD_test_bizunit 项目成员: 人 计算源: [test_basic12]> 元前往配置 创建时间: 2019-12-20 17:06:03 更新时间: 2020-09-22 18:39:34 [信息设置 成员管理
Basic模式	☞ 超级管理员 进入工作区》
	 注意 Basic模式不支持升级为Dev-Prod模式,且Basic模式存在直接变更生产风险,请您谨慎选择。 如果选择了Basic模式,请您谨慎管理项目成员,以控制数据生产的稳定性。

前提条件

在开始执行操作前,请确认您以满足以下条件:

- 如果您需要研发流批一体任务,您需要根据业务需求选择新建Ververcia Flink计算源、创建Blink计算源或 开源Flink计算引擎。具体操作,请参见新建Flink计算源或创建Blink计算源。
- 如果您选择的Dataphin的计算引擎为AnalyticDB for PostgreSQL,且数据研发过程中需要使用数据集成、 即席查询、AnalyticDB for PostgreSQL的SQL计算任务等功能,则在创建项目空间前需要完成AnalyticDB for PostgreSQL计算源的创建,请参见新建AnalyticDB For PostgreSQL计算源。
- 如果您选择的Dataphin的计算引擎为MaxCompute,且数据研发过程中需要使用规范建模、即席查询、 MaxCompute计算任务等功能,则在创建项目空间前需要完成MaxCompute计算源的创建,请参见新建 MaxCompute计算源。
- Dataphin支持添加Hologres加速计算源至已绑定MaxCompute计算源的项目,加速数据服务的查询速度。 如果您需要加速业务数据的查询速度,则在创建项目空间前需要创建加速计算源,详情请参见创建加速计 算源。

使用限制

- 系统支持超级管理员、系统管理员、板块架构师角色的账号创建项目。
- 系统仅支持项目超级管理员、系统管理员、板块架构师角色的账号,具备关闭和开启数据表读写的权限申请。

操作流程

操作流程	描述
步骤一:选择项目类型	选择项目的模式为Basic模式。 完成项目的创建后,系统生成独立灵活的生产环境,数据生产过程相对稳定而便捷,生 产的数据可管控。
步骤二:配置项目的计算 源和名称	为项目配置计算源、英文名、名称和描述。
步骤三:配置项目的命名 空间	配置项目的业务板块和空间类型。
步骤四:配置项目的沙箱 白名单	配置项目中的Shell、Python任务可以访问的IP地址或域名。 您也可以通过 信息设置 添加沙箱白名单,请参见 <mark>添加沙箱白名单</mark> 。
步骤五:配置项目的数据 读写申请权限	Dataphin支持通过项目粒度的方式,管控项目中的元表和物理表的读取和写入数据的权限申请,以规避审批者误操作,导致的数据泄露与安全问题。 您也可以通过 信息设置 配置项目的数据读写申请权限,请参见 <mark>设置数据读写申请权限</mark> 。

步骤一:选择项目类型

1. 请参见入口介绍,进入数仓规划页面。

2. 在数仓规划页面,按照下图操作指引,进入新建项目对话框。

数仓规划·项目空间



3. 在新建项目对话框,选择创建项目的模式为Basic模式。

新建项目			
		dataphin ann a state ann ann ann ann ann ann ann ann ann an	
	Dev-Prod 模式 生成互相关联的Dev环境和Prod环境,以强 管控数据生产过程、生产的数据。	Basic 模式 生成独立灵活的Prod环境,数据生产过程相 对稳定而便捷,生产的数据可管控	
	如您管理诉求较强,数据研发人员较多且能 力及分工明确,数据计算存储预算较高,推 荐此模式: Dev 环境和 Prod 环境逻辑隔 离,可加强对数据研发流程的管控,强控制 生产数据的权限,最大化保障数据生产稳定 性。但是研发流程灵活及便捷性较低,对研 发人员工具使用熟练度要求较高,需要配置 双份计算存碌答预	如您关注数据研发效率而非强管控,研发人 员职能边界相对模糊,计算存储预算有限 制,推荐此模式:开发与运维一体化的 Prod 环境,可敏捷管理数据研发流程,强 控制生产数据的写权限,平衡数据生产稳定 性与灵活性	
	עאולאנאן בואר אראעי	and the second second	

4. 单击下一步。

步骤二: 配置项目的计算源和名称

在新建项目对话框的项目设置区域,配置计算源参数。

根据Dataphin初始化的计算引擎不同,支持选择的计算源类型也不同,详细说明如下:

• Dataphin初始化为MaxCompute,则需要选择计算源类型,支持选择的计算源类型包括MaxCompute、 Flink、Hologres。

⑦ 说明 如果Dataphin初始化为MaxCompute,则需要选择计算源类型后,再配置计算源参数。

基本设置	Ganut Danis	Crean a				Que. :	(J.84)
计算源类型	 ✓ MAXC ✓ FLINK ✓ HOLO ① ① 请至少选择一个计算源类型 			* 机器学习	○关 64440 ¹⁰¹		
* MAXC	请选择批处理计算源						
	同步创建数据源 ①						
* FLINK	请选择流处理计算源						
* HOLO	请选择加速计算源		\sim				
* 英文名				* 名称			
描述	请输入项目描述,128字符以内						0/128

- MaxCompute: MaxCompute是一种快速、完全托管的TB或PB级数据仓库解决方案,能够更快速为您 解决海量数据计算问题,有效降低企业成本,并保障数据安全,详情请参见MaxCompute。
- Flink: 添加Flink计算源后,您可以在Dataphin中进行流式计算任务开发,详情请参见阿里云实时计算。
- Hologres: 您可以创建计算源并添加至已绑定MaxCompute计算源的项目中,后续就可以对计算空间选择该项目的Dataphin数据源进行加速业务数据的查询速度,详情请参见什么是实时数仓Hologres。
- Dat aphin初始化为AnalyticDB for PostgreSQL,则不需要选择计算源类型。

基本设置	, Adul) -		A AMINO			
* 计算源	请选择批处理计算源					
* 英文名			* 名称			
描述	请输入项目描述, 128字符以内					0/128

计算源的参数配置,详细说明如下。

参数	描述
计算源	选择您已创建相应计算源类型的计算源,项目创建成功后,不支持修改计算源。
同步创建数据源	系统支持为项目绑定MaxCompute、AnalyticDB For PostgreSQL计算源时,且选中 同 步创建数据源后,Dataphin自动创建对应类型的数据源。创建数据源时配置的信息, 请参见 <mark>附录:同步创建数据源的机制说明</mark> 。
英文名	项目创建成功后,其英文名不支持修改。命名规则如下: • 包含字母、数字、下划线(_)。 • 不能超过64个字符。
名称	命名规则如下: • 包含中文、数字、字母、下划线(_)、短划线(-)。 • 不能超过64个字符。
描述	填写对Dev、Prod项目的描述。

步骤三:配置项目的命名空间

在新建项目页面的命名空间区域,配置参数。

新建项目							X
* 英文名			*:	名称			
描述	请输入项目描述,128字符以内						0/128 //
命名空间							
业务板块			∨ * 空间	类型 应用层			~
功能操作 代码 其他设置	3编辑/调度						
沙箱白名单 🛈							+ 新建
访问地址			端口				提作
			首无数店				
通用功能控制 • 数据法权限		* 粉挥写权限中法		≠闭 ①			
d		. 3018-21V1KH1B					
					上一步	取消	确定

参数	描述			
	非必填写项。为新建项目选择所属的业务板块。			
业务板块	↓ 注意 如果您购买的Dataphin实例为基础研发版,则系统不支持业务板块。			
	反八年日开华伏女卫立山教授时任,殿门头古田日,南海米司与任3站。			
	 基础层:米目各个业务数据库的原始数据存储,是后续所有数据建设的米源与基础, 也称为垂直数据中心。 			
空间类型	 公共层:基于各业务数据,抽象公共的主题,沉淀标准、规范、可共用的数据,是连接基础层与应用层的中间桥梁,也称为公共数据中心。 			
	 应用层:面向业务需求,定义生成可应用于不同场景的个性化、多样化的数据指标等。 			
功能操作	默认为代码编辑、调度,不支持修改。			

附录:同步创建数据源的机制说明

同步创建数据源时,系统自动配置的参数信息如下。

参数	描述
数据源类型	根据计算源类型配置为相应的数据源类型,包括MaxCompute或AnalyticDB For PostgreSQL。
数据源名称	默认配置为计算源名称。
数据源描述	默认配置为项目的描述。
数据源配置	默认选择为 生产+开发数据源 。
开发数据源	需要配置Endpoint、Project Name、Access ID、Access Key参数,默认填写项 目绑定的计算源的连接信息。
生产数据源	需要配置Endpoint、Project Name、Access ID、Access Key参数,默认填写项 目绑定的计算源的连接信息。

同步创建数据源过程中,系统校验规则如下:

- 数据源重名校验:如果已存在同名数据源,则数据源创建失败,项目创建不阻断。
- 数据源测试连接: 如果测试连接失败, 则数据源创建失败, 项目创建不阻断。

后续步骤

创建完项目后,即可进入工作区开始数据的研发。如何基于Dataphin开发数据,请参见数据开发。

zz_test <	Prod zz_test
ky_encrypt_project(ky_test01)	ky_encrypt_project_dev(ky_test
暂无描述	1521
业务板块: LD_zz_test	业务板块: LD_zz_test_dev
项目成员:	项目成员:
计算源: ■ ky_encrypt > 无前往面	2置 计算源: ■ ky_encrypt → 无 前往配置
创建时间: 2021-02-04 11:41:59	创建时间: 2021-02-04 11:41:59
更新时间: 2021-03-19 20:04:32	更新时间: 2021-03-19 20:04:32
信息设置 成员管理	信息设置成员管理
☞ 超级管理员 进入工作	E区》

5.3. 管理项目空间的权限和计算源

为了更好地管控项目空间中的数据安全及数据处理能力,Dataphin支持通过项目空间全局管理数据结果下载 权限、数据读写权限、访问权限(沙箱白名单)、开启资产安全策略、成员的操作权限及添加计算源。本文 为您介绍如何管理项目空间的权限和添加计算源。

前提条件

完成项目空间的创建,请参见创建Basic项目空间或创建Prod和Dev项目空间。

场景说明

功能	使用场景说明	使用限制
设置数据结果下载 权限	 Dataphin支持通过项目粒度的方式,管控项目空间中数据查询(即席查询或计算任务)结果的下载权限,以规避数据泄露的问题: 关闭后,则系统通过执行代码(即席查询或计算任务),查询的数据结果无法下载。 开启后,则系统支持下载查询数据结果至本地客户端,请参见下载数据。 	
设置数据读写申请 权限	 Dataphin支持通过项目粒度的方式,管控项目空间中的元表和物理表的读取和写入数据的权限申请。以规避审批者误操作,导致的数据泄露与安全问题: 关闭数据读和写权限申请后,当前项目下物理表及元表不支持申请查询和写入的权限,系统不收回已申请过的查询和写入的权限。 开启数据读和写权限申请后,当前项目下的物理表及元表支持申请查询和写入权限。申请查询和写入权限的更多信息,请参见申请权限。 	系统仅支持超级管 理员和项目管理员 免免的成号操作
安装资产安全策略	安装资产安全策略后,即可定义当前项目下资产的业务分类和安全等级,并构建敏感数据识别规则,同时支持设置敏感数据脱敏规则,以 保障数据资产的安全性。更多资产安全策略信息,请参见 <mark>资产安全</mark> 。	用巴的成页採TF
添加沙箱白名单	配置项目空间中的Shell、Python任务可以访问的IP地址或域名。	
添加项目成员	为项目空间添加成员。被添加的成员即可具有角色对应的权限,请参 见 <mark>角色对应权限概述</mark> 。	
添加计算源	拓展当前项目空间数据处理的能力。不同类型的计算源支持研发的任 务,请参见 <mark>数据开发</mark> 。	
	↓ 注意 对于已添加的计算源,系统不支持修改和删除。	

设置数据结果下载权限

1. 请参见入口介绍,进入项目管理页面。

- 2. 在项目管理页面,单击需要修改Dev或Prod项目下的信息设置。
- 3. 在信息设置对话框的通用功能控制区域,选择是否需要开启数据结果下载权限。

信息设置: Basic Basic Basic Basic Basic Basic	⇒ 模式 ①	X
访问地址		操作
cip.cc	80	Ō
rm-bp1	3306	Ō
service	• 1444° - 2444° - 2444° - 2444°	靣
rm-uf6t0	3306	ī
通用功能控制		
* 数据结果下载 ① • 开启 · 关闭 * 资产安全策略 ① :	未安装 安装 * 数据读权限申请 ○ 开启 •) 关闭 ①
* 数据写权限申请 🔵 开启 💿 关闭 🕕		
	取消	确定
() 注意 系统仅支持非访客角色的田户下裁	数据结果至本地。	

4. 单击**确定**,即可开启或关闭数据结果下载权限。

- 关闭后,则系统通过执行代码(即席查询或计算任务),查询的数据结果无法下载。
- 开启后,则系统支持下载查询数据结果至本地客户端,请参见下载数据。

设置数据读写申请权限

- 1. 请参见入口介绍,进入项目管理页面。
- 2. 在项目管理页面,单击需要修改Dev或Prod项目下的信息设置。
- 3. 在信息设置对话框的通用功能控制区域,选择是否需要开启数据读写权限申请权限。

信息设置:	已选: Basic 模式 ①		×
访问地址			操作
cip.cc	80		Ō
rm-bp1	3306		Ō
service	• 65999 • 68999		Ō
rm-uf6t0	3306		Ō
通用功能控制	Samuel Paranet Paranet State	19	
∗ 数据结果下载 ① ● 开启 ○ 关闭	*资产安全策略①未安装 安装	* 数据读权限申请 🔵 开启 💿 关闭	(i)
* 数据写权限申请 🔵 开启 💿 关闭 🕧			
		取消	确定

4. 单击确定,即可开启或关闭数据读写权限的申请。

开启或关闭数据写权限申请:

- 关闭数据写权限申请后,当前项目下的物理表及元表不支持申请写入权限,同时系统不收回已申请过的写入权限。
- 开启数据写权限申请后,当前项目下的物理表及元表支持申请写入权限。申请写入权限的更多信息, 请参见申请权限。

开启或关闭数据读权限申请:

- 关闭数据读权限申请后,当前项目下物理表及元表不支持申请查询权限,同时系统不收回已申请过的 查询权限。
- 开启数据读权限申请后,当前项目下的物理表及元表支持申请查询权限。申请查询权限的更多信息, 请参见申请权限。

安装资产安全策略

- 1. 请参见入口介绍,进入项目管理页面。
- 2. 在项目管理页面,单击需要修改Dev或Prod项目下的信息设置。
- 3. 在信息设置对话框的通用功能控制区域,单击安装,系统显示安装中。

一般安装需要30分钟左右,请您耐心等待。

信息设置:	已选: Basic 模式 ①	×
访问地址		操作
cip.cc	80	亩
rm-bp1	3306	Ō
service		靣
rm-uf6t0	3306	亩
通用功能控制	الاعتمادين العتمدين العتمد العتمان العتمان العقام	
* 数据结果下载 🗊 🧿 开启 关闭	* 资产安全策略 ① 未安装	闭
*数据写权限申请 🔵 开启 💿 关闭 🕧		
	取消	确定

4. 安装完成后,单击确定,即可使用资产安全策略。

安装资产安全策略后,即可定义当前项目下资产的业务分类和安全等级,并构建敏感数据识别规则,同时支持设置敏感数据脱敏规则,以保障数据资产的安全性。如何使用资产安全策略,请参见资产安全。

添加沙箱白名单

- 1. 请参见入口介绍,进入项目管理页面。
- 2. 在项目管理页面,单击需要修改Dev或Prod项目下的信息设置。
- 3. 在**信息设置**对话框的**沙箱白名单**区域,添加该项目空间下输入Shell、Python任务需要访问的IP地址或 域名。

ţ	其他设置		
Ż	ゆ箱白名单 ①		+ 新建
	访问地址		操作
	101.133	39041	ò
	gp rds.aliyuncs.com	5432	τ.

i. 单击新建。

ii. 根据页面提示, 输入Shell、Python任务需要访问的IP地址或域名。

↓ 注意 必须填写可以被访问到的ⅠP地址或域名。

ⅲ. 单击<mark>、</mark>图标。

4. 单击确定, Shell、Python任务可以访问对应的IP地址或域名。

对于已添加的沙箱白名单,您也可以单击操作列下的 图标,删除后项目空间下的Shell、Python任务,将无法访问对应的IP地址或域名。

添加项目成员

- 1. 请参见入口介绍,进入项目管理页面。
- 2. 在项目管理页面,单击需要修改项目下的成员管理。
- 3. 在成员管理对话框,单击添加成员。

成员管理: bizunit_null_upgrad	de_dev(未绑定	Q 请输入成员名	+ 添加成员 ×
成员名 🌻 角目	色① 添加人	添加日期 🌻	修改日期 🍨 操作
	超级管理员	2020-03-26 10:32:09	2020-03-26 10:32:09
□ 全选			
			· 注 · · · · · · · · · · · · · · · · · ·
			141 JE

- 4. 添加成员。
 - i. 在**添加成员**对话框,选择成员和成员的角色。

项目空间中不同角色成员的权限不同,权限概述详情见下表,详细信息请参见<mark>权限管理。</mark>角色对应 权限概述

角色	权限范围概述
超级管理员	具有管理Dataphin成员及项目成员的权限、项目文件(含逻辑表)的增加、删除、修改及发布权限、项目节点的操作权限、项目数据的增加、删除、修改权限。超级管理员角色的成员默认为所有项目的项目管理员,不支持从项目空间中移除。
项目管理员	具有管理项目成员的权限、项目文件(含逻辑表)的增加、删除、修改及发布权 限、项目节点的操作权限、项目数据的增加、删除、修改权限。项目管理员角色 的成员的角色支持修改、同时支持从项目空间中移除。

角色	权限范围概述
开发者	具有项目文件(含逻辑表)的增加、删除、修改、发布权限、项目节点的操作权 限、建表及负责人为自己的表删除、修改的权限。开发者角色的成员的角色支持 修改、同时支持从项目空间中移除。
运维	具有将项目文件从Dev项目发布至Prod项目的权限,具有项目节点的操作权限, 具有项目建表权限、项目表删除、修改及查询权限。运维角色的成员的角色支持 修改、同时支持从项目空间中移除。
分析师	具有项目文件(含逻辑表)的查看权限及项目即席查询文件增加、删除、修改权 限、项目节点的查看权限、项目数据查询权限。分析师角色的成员的角色支持修 改、同时支持从项目空间中移除。
访客	具有项目文件(含逻辑表)的查看权限、项目节点的查看权限、项目数据的查询 权限。访客角色的成员的角色支持修改、同时支持从项目空间中移除。

如果是Dev或Prod项目,则您可以选中为对应项目添加成员的配置项。系统支持同时为对应项目添加成员,提高了添加成员的效率。

成员管	理: POC30(poc30))			Q 请输入成	员名	+ 添加成员	×
				添加成员			×	
	成员名 🔹	角色③	添					操作
	MARCHER'S	☞ 超级管理员	dat	* 成员名	请选择成员			i i
	History and Park	➡项目管 ∨	dat	角色	◎ 访客		~	亩
	Bitmerry).	➡ 项目管 ∨	来	[] 添加]	为Dev访客()	取消	确定	Ō
	use, where a	🝳 运维 🛛 🗸		-	2020-11-01 14:	28:56 2020-11-	-01 14:28:56	Ō
	samputertij.	🝳 运维 🛛 🗸	0	per constantes	2020-11-08 20:	57:57 2020-11	-08 20:57:57	Ō
		· · · · · · · · · · · · · · · · · · ·	• • •	· · = / Alanda B				-
4	≧选 □ 批量移出							
	And a second	1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 -		6.552.011	356 P ¹¹¹	6.000	确	定

当前项目	当前项目角色	对应项目角色
	项目管理员	项目管理员
Prod	运维	开发者
	访客	访客
	项目管理员	项目管理员
	开发者	运维
Dev	分析师	访客
	访客	访客

添加成员的角色对应关系如下表所示。

当前项目	当前项目角色	对应项目角色

↓ 注意 如果对应项目中已有该成员,则保持原有角色不做覆盖。

ii. 单击确定。

5. 单击确定, 被添加的成员可以进入当前项目空间进行建设或查询数据。

系统支持修改项目中的项目管理员、开发者、运维、分析师和访客角色成员的角色,及支持从项目空间 中移除项目管理员、开发者、运维、分析师和访客角色的成员。

↓ 注意 系统不支持修改超级管理员的角色,同时不支持从项目空间中移除超级管理员角色的成员。

功能	操作步骤
修改成员角色	 i. 在成员管理对话框,单击需要修改成员后的 图标。 ii. 在下拉列表中,选择角色。关于不同角色对应权限概述,请参见角色对应权限 概述。详细的权限信息,请参见权限管理。 iii. 单击确定,修改角色的成员仅具有修改后角色对应的权限。权限信息,请参 见角色对应权限概述。
移除项目中的成员	 i. 在成员管理对话框,需要删除成员操作列下的 图标。 您也可以选中多个成员后,单击成员管理对话框下方的批量移出。 ii. 在确认对话框中,单击确定。 iii. 单击确定,被移除的成员不具有当前项目空间的任何操作权限。

添加计算源

- 1. 请参见入口介绍,进入项目管理页面。
- 2. 在项目管理页面,单击需要修改Dev或Prod项目下的信息设置。
- 3. 在信息设置对话框的项目设置区域,选择项目空间中没有添加的计算源类型及计算源。

 项目1: DEV 计算源 请至少选择一个计算源, 生产/开发计算源设置同步 计算源类型 ▼ MAXC ▼ FLINK * MAXC poc30_dev ∨ 	页目2: PROD 计算源 请至少选择一个计算源, 生产/开 计算源类型 ☑ MAXC ☑ FLINK	发计算源设置同步
 ・ 计算源 请至少选择一个计算源, 生产/开发计算源设置同步 ・ 计算源类型 図 MAXC 図 FLINK * MAXC poc30_dev ∨ 	计算源 请至少选择一个计算源, 生产/开 计算源类型 ▼ MAXC ▼ FLINK	发计算源设置同步
计算源类型 ▼ MAXC ▼ FLINK * MAXC poc30_dev ∨	计算源类型 🔽 MAXC 🔽 FLINK	
* MAXC poc30_dev V		
	* MAXC poc30	\sim
* FLINK 请选择流处理计算源 >	* FLINK 请选择流处理计算源	×
英文名 POC30_dev 名称 poc30_开发	英文名 POC30	名称 poc30
描述 1233	描述 请输入项目描述, 128字符以内	

4. 单击确定,即可基于添加的计算源类型开始任务的研发,请参见数据开发。

↓ 注意 对于已添加的计算源,系统不支持修改和删除。

6.数据源 6.1. 概述

创建数据源即可实现业务数据源与Dataphin平台的网络连接。网络连接成功后,即可基于Dataphin平台和业务数据构建智能数据。同时,您也可以将已构建的智能数据导出至数据源。

数据源类型

Dataphin支持连接多种多源异构的数据源。目前,Dataphin支持的数据源类型,详情见下表。

类别	数据源类型
大数据存储	MaxCompute、Hive、SAP HANA、DataHub、Hologres
文件	HDFS、FTP、OSS
消息队列	Log Service、Kafka、RocketMQ
关系型数据库	PolarDB、PolarDB-X(原DRDS)、MySQL、Microsoft SQL Server、PostgreSQL、 AnalyticDB for MySQL 2.0、AnalyticDB for MySQL 3.0、AnalyticDB for PostgreSQL、OceanBase、Oracle、Vertica、IBM DB2、Teradata、Impala
NoSQL	Elasticsearch、HBase 1.1.X、HBase 0.9.4、HBase 2.x、MongoDB、Aliyun Hbase、 Tablestore

⑦ **说明** 如果您有其他类型的数据源诉求,请提工单并在工单上填写您需要的数据源版本及配置信息。

网络连通

不同数据源支持的连接方式不同,详情见下表。

数据源类型	访问方式
MaxCompute	公网访问
Hive	公网访问
SAP HANA	 公网数据库类型的数据源,支持公网访问。 ECS(VPC)自建数据源类型的数据源,支持VPC访问。
DataHub	公网访问
Hologres	公网访问
HDFS	公网访问
FTP	公网访问
OSS	公网访问

数据源类型	访问方式
Log Service	公网访问
Kafka	公网访问
RocketMQ	公网访问
PolarDB-X(原DRDS)	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。
PolarDB	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。 ECS(VPC)自建数据源类型的数据源,支持VPC访问。
MySQL	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。 ECS(VPC)自建数据源类型的数据源,支持VPC访问。
Microsoft SQL Server	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。 ECS(VPC)自建数据源类型的数据源,支持VPC访问。
PostgreSQL	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。 ECS(VPC)自建数据源类型的数据源,支持VPC访问。
AnalyticDB for MySQL 2.0	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。
AnalyticDB for MySQL 3.0	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。
AnalyticDB for PostgreSQL	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。
OceanBase	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。 ECS(VPC)自建数据源类型的数据源,支持VPC访问。
Oracle	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。

数据源类型	访问方式
Vertica	公网访问
IBM DB2	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。
Teradata	 公网数据库类型的数据源,支持公网IP访问。 阿里云数据库类型的数据源,支持公网和VPC访问。
Impala	公网访问
Elasticsearch	公网访问
HBase	公网访问
MongoDB	公网访问
Aliyun Hbase	公网访问
Tablestore	公网访问

◯ 注意

- 目前,系统支持将Dataphin实例所在地域(华东2、华南1、华北2和华东1)的数据库配置为数据源,且通过VPC网络连接Dataphin。
- 其他地域VPC网络的阿里云数据库,可以通过开通公网或高速通道的方式连接Dat aphin。

同时,为了实现Dataphin与公网数据库、阿里云数据库或ECS(VPC)自建数据源的网络连通,您还需要添加 Dataphin IP白名单至数据库:

- 公网数据库:添加Dataphin IP白名单。Dataphin的IP白名单,请参见公网的Dataphin IP白名单。
- 阿里云数据库:
 - 公网IP的访问方式:数据库需要添加Dataphin IP白名单。Dataphin的IP白名单,请参见公网的Dataphin IP白名单。
 - VPC网络的访问方式:数据库需要添加Dataphin IP白名单。Dataphin的IP白名单,请参见VPC网络的 Dataphin IP白名单。
- ECS(VPC)自建数据源:添加Dataphin IP白名单。您可以添加多个不同地域的Dataphin IP白名单,也可以添加所有地域的Dataphin IP白名单。Dataphin IP白名单,请参见VPC网络的Dataphin IP白名单。

VPC网络的Dataphin IP白名单

地域	IP白名单
华东2(上海)	100.104.228.128/26、100.104.115.192/26
华南1(深圳)	100.104.48.128/26
华北2(北京)	100.104.238.64/26

地域	IP白名单
所有地域	100.104.0.0/16

公网的Dataphin IP白名单

地域	IP白名单
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202

网络连通的常见问题

数据源测试不通后,建议您从以下几个方面排查:

- 数据库是否已正常启动。
- 数据库所在网络禁止Dataphin访问,需要添加白名单。
- Dataphin无法访问数据库所在网络,请确保数据库已开发并授予访问权限。

6.2. Dataphin支持的数据源

在使用Dataphin前,您需要选择符合业务场景需求的数据库或数据仓库作为数据源,用于读取原始数据和写入数据建设过程中的数据。Dataphin集成了丰富的数据引擎,支持对接如MaxCompute等数据仓库,也支持对接企业传统数据库,如MySQL、Oracle等。

Dataphin支持对接的数据源包括大数据存储型数据源、文件数据源、消息队列数据源、关系型数据源和 NoSQL数据源,各模块支持对接的数据源类型说明如下。如果您需要在Dataphin中连接某数据源,则需要先 在数仓规划中的数据源模块创建该数据源。

大数据存储型数据源

数据源类型	数据集成	实时研发	数据治理	数据服务	数据安全	创建数据源 的操作指导
MaxComput e	支持	支持	支持	不支持	支持	创建 MaxComput e数据源
SAP HANA	支持	不支持	不支持	不支持	不支持	创建SAP HANA数据源
Dat aHub	不支持	支持	不支持	不支持	不支持	创建 DataHub数 据源
Hologres	支持	支持	不支持	不支持	不支持	创建 Hologres数 据源

文件数据源

数据源类型	数据集成	实时研发	数据治理	数据服务	数据安全	创建数据源 的操作指导
FTP	支持	不支持	不支持	不支持	不支持	创建FTP数据 源
OSS	支持	不支持	不支持	不支持	不支持	创建OSS数 据源

消息队列数据源

数据源类型	数据集成	实时研发	数据治理	数据服务	数据安全	创建数据源 的操作指导
Log Service	支持	支持	不支持	不支持	不支持	创建Log Service数据 源
Kafka	不支持	支持	不支持	不支持	不支持	创建Kafka数 据源
RocketMQ	不支持	支持	不支持	不支持	不支持	创建 RocketMQ数 据源

关系型数据源

数据源类型	数据集成	实时研发	数据治理	数据服务	数据安全	创建数据源 的操作指导
PolarDB	支持	不支持	不支持	不支持	不支持	创建PolarDB 数据源
PolarDB- X(原 DRDS)	支持	支持	不支持	不支持	不支持	创建 PolarDB-X数 据源
MySQL	支持	不支持	不支持	支持	不支持	创建MySQL 数据源
Microsoft SQL Server	支持	不支持	不支持	支持	不支持	创建 Microsoft SQL Server 数据源
PostgreSQL	支持	不支持	不支持	支持	不支持	创建 PostgreSQL 数据源
Oracle	支持	不支持	不支持	支持	不支持	创建Oracle 数据源

数仓规划·数据源

数据源类型	数据集成	实时研发	数据治理	数据服务	数据安全	创建数据源 的操作指导
AnalyticDB for MySQL 2.0	支持	不支持	不支持	支持	不支持	创建 AnalyticDB for MySQL数 据源
AnalyticDB for MySQL 3.0	支持	支持	不支持	支持	不支持	创建 AnalyticDB for MySQL数 据源
AnalyticDB for PostgreSQL	支持	不支持	支持	支持	支持	创建 AnalyticDB for PostgreSQL 数据源
OceanBase	支持	不支持	不支持	不支持	不支持	创建 OceanBase 数据源
Vertica	支持	不支持	不支持	不支持	不支持	创建Vertica 数据源
IBM DB2	支持	不支持	不支持	不支持	不支持	创建IBM DB2 数据源
Teradata	支持	不支持	不支持	不支持	不支持	创建 Teradata数 据源
Impala	不支持	不支持	不支持	支持	不支持	创建Impala 数据源

NoSQL数据源

数据源类型	数据集成	实时研发	数据治理	数据服务	数据安全	创建数据源 的操作指导
HBase	支持	不支持	不支持	支持	不支持	创建HBase 数据源
Elasticsearc h	支持	不支持	不支持	支持	不支持	创建 Elasticsearc h数据源
MongoDB	支持	不支持	不支持	支持	不支持	创建 MongoDB数 据源

数据源类型	数据集成	实时研发	数据治理	数据服务	数据安全	创建数据源 的操作指导
Tablestore	不支持	支持	不支持	不支持	不支持	创建 Tablestore 数据源
Aliyun HBase	不支持	支持	不支持	不支持	不支持	创建Aliyun HBase数据 源

6.3. Dataphin出网IP地址

在连通业务数据库与Dataphin间的网络时,您需要将不同网络环境下Dataphin的出网IP地址添加至业务数据库的白名单中。本文为您介绍不同网络环境下Dataphin的出网IP地址。

VPC网络下Dataphin的出网IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

公网下Dataphin的出网IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

6.4. 网络连通解决方案

添加数据源时需要根据数据源所在网络环境,通过对应的网络解决方案,实现Dataphin与数据源的网络连通。本文为您介绍数据源在不同网络环境中与Dataphin的网络连通方案。

网络连通方案

根据数据源所在的网络环境,在下图中选择对应的网络连通解决方案。



网络连通方案的操作指导说明如下表所示。

数据源所在的网络环境	网络连通操作指导
数据源具备公网访问能力	 添加Dataphin的IP地址(公网下Dataphin的IP地址)至数据源的白名单中。 以阿里云云数据库RDS为例,如果您在RDS配置指定了IP白名单,则需要添加 Dataphin的IP地址到白名单列表中。如何添加白名单,请参见步骤二:设置IP白名 单。 添加数据源访问地址至Dataphin项目空间的沙箱白名单中。如何添加沙箱白名单请 参见添加沙箱白名单。
数据源在VPC内-VPC和 Dataphin在同一地域	 在Dataphin中添加数据源时,绑定数据源的VPC。 添加Dataphin的IP地址(VPC网络下Dataphin的IP地址)至数据源白名单中。 以阿里云云数据库RDS为例,如果您在RDS配置指定了IP白名单,则需要添加 Dataphin的IP地址到白名单列表中。如何添加白名单,请参见步骤二:设置IP白名 单。 添加数据源访问地址至Dataphin项目空间的沙箱白名单中。如何添加沙箱白名单请 参见添加沙箱白名单。
数据源在VPC内-VPC和 Dataphin不在同一地域	 在Dataphin所在地域新建一个VPC网络的ECS。 使用高速通道或VPN连通两个VPC(ECS的VPC和数据源的VPC)。 在Dataphin中添加数据源时,绑定ECS的VPC。 添加Dataphin的IP地址(VPC网络下Dataphin的IP地址)至ECS的安全组中。 如何添加安全组,请参见添加安全组规则。 添加数据源访问地址至Dataphin项目空间的沙箱白名单中。如何添加沙箱白名单请参见添加沙箱白名单。 Dataphin即可通过ECS访问数据源。

数据源所在的网络环境	网络连通操作指导
数据源在IDC内	1. 在Dataphin所在地域新建一个VPC网络的ECS。
第三方云自建数据源	2. 使用高速通道或VPN连通ECS的VPC和数据源的网络。
	3.添加Dataphin的IP地址(VPC网络下Dataphin的IP地址)至ECS的安全组中。
	如何添加安全组,请参见 <mark>添加安全组规则</mark> 。
	4. 在ECS上配置Nginx反向代理。
	5. 添加数据源访问地址至Dataphin项目空间的沙箱白名单中。如何添加沙箱白名单请 参见 <mark>添加沙箱白名单</mark> 。
	6.在Dataphin中添加数据源时,绑定ECS的VPC。
	7. Dataphin即可通过ECS访问数据源。

在数据源通过VPC网络连通Dataphin的场景中,需要将Dataphin在VPC网络下的IP地址添加至数据源白名单中。

VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

在数据源通过公网连通Dataphin的场景中,需要将Dataphin在公网下的IP地址添加至数据源白名单中。

公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

网络连通的常见问题

数据源网络连接测试不通时,建议您从以下几个方面排查:

● 数据源是否已正常启动。

以MySQL数据源为例,您可以在本地PC上telnet 127.0.0.1 3306,查看数据库端口是否已开启。

- Dataphin无法访问数据源所在网络,请确保数据源所在网络已和阿里云连通。
- DNS无法正常解析数据源链接地址中的域名,请确认可以正常解析域名访问。

数据源所在的网络防火墙禁止Dataphin访问,请添加Dataphin的IP地址(VPC网络下Dataphin的IP地址) 至数据源的白名单列表中。

以阿里云云数据库RDS为例,如果您在RDS配置指定了IP白名单,则需要添加Dataphin的IP地址到白名单列 表中。如何添加白名单,请参见步骤二:设置IP白名单。

- 数据源的端口配置错误或对应的网络端口没有放开,请确保配置正确的数据源端口及确保对应的网络端口 已放开。
- 数据源的版本和类型错误,请确保选择正确的数据源版本和类型。
- 数据源的链接地址参数配置错误,请确保数据源链接地址的正确性。
- Hadoop体系的开源组件(如Hive、HDFS、HBase)为数据源时,为了规避Kerberos身份认证失败的问题,请确保上传正确的认证文件。

数据同步中网络连接异常的解决方案案例

数据同步中网络连接异常的解决方案案例

6.5. 大数据存储数据源

6.5.1. 创建MaxCompute数据源

创建MaxCompute数据源用于实现Dataphin能够读取MaxCompute的业务数据,及能够向MaxCompute写入数据。在引入MaxCompute的业务数据至Dataphin和将Dataphin的数据写入至MaxCompute的场景中,您需要先创建MaxCompute数据源。本文为您介绍如何创建MaxCompute类型的数据源。

背景信息

MaxCompute即阿里云大数据计算服务,如果您使用的是阿里云大数据计算服务MaxCompute,在对接 Dataphin进行数据开发或将Dataphin的数据导出至MaxCompute场景中,您需要先完成MaxCompute数据源 的创建。更多MaxCompute信息,请参见什么是MaxCompute。

Dataphin支持通过公网和VPC网络连接MaxCompute。Dataphin连接MaxCompute时,可以选择 MaxCompute在VPC网络或公网环境下对应的Endpoint,MaxCompute的Endpoint请参见Endpoint。通 常,VPC网络较公网稳定,建议您选择VPC网络连通Dataphin和MaxCompute。

为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中,更多对接配置详情请参见管理IP白名单。

使用限制

• Dataphin仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

 当您使用VPC网络连接Dataphin实例与MaxCompute项目时,仅支持连接与Dataphin实例在同一地域的 MaxCompute项目。

步骤一:添加Dataphin IP地址至MaxCompute白名单

基于您选择连通Dataphin与MaxCompute的网络类型,添加对应网络类型下的Dataphin IP地址至MaxCompute白名单中。如何添加白名单,请参见管理IP白名单。

● 使用公网连通Dataphin与MaxCompute时,需要添加公网下Dataphin IP地址至MaxCompute白名单中。公 网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● 使用VPC网络连通Dataphin与MaxCompute时,需要添加VPC网络下Dataphin IP地址至MaxCompute白名 单中。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建MaxCompute数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

Dataphi	n 规划								ి 🕗 🕓
77% 2	数据源 (190) ①						Q. 清箱入组发关键字	V	+ 6/22/03/202
合 数据影响	guzhi_oss ⊠			Ē	lfs_lindorm ⊠				÷
₩ 公共定义	类型: OSS				問題: Lindorm				
項目	生)**	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①
▶ 项目管理	负责人:						SEA: maximum characteristic		
41. 计算题	政策原用途: 案线 创建用户:		政强逐用途: 集线 《建用 ^内 : <mark>Manufalana (Super State (Super State)</mark>		政策原用注: 株成 (1時用件: 株成)		数据源用途: 集线 创建用户:		
æ <u>(</u>)	台(建約)(0): 2022-03-09 20:27:15 単和時間: 2022-03-11 12:25:27		①読む日前: 2022-03-09 20:30:32 単称行い前: 2022-03-11 12:25:33		(1)建設計算: 2022-03-09 11:03:47 単型設計算: 2022-03-09 11:03:47		台建时间: 2022-03-09 11:03:47 東京日前: 2022-03-09 11:03:47		
.	総務信息: ◆ 宣誓経済信息				· · · · · · · · · · · · · · · · · · ·				
	操作 :	0 1 1 2	操作:	0 0 0 A	操作:	C 1 1 A	探行 :	e 12	₿ <u>&</u>

3. 在新建数据源对话框的大数据存储区域,选择MaxCompute。

如果您最近使用过MaxCompute,也可以在**最近使用**区域选择MaxCompute。同时,您也可以在快搜索框中,输入MaxCompute的关键词,快速筛选。

新建数据源			Q. 清编入注	主要关键字 → X
大数据存储				
\sim		SAP HANA	*	HOLOGRES
MaxCompute	Hive	SAP HANA	DataHub	Hologres
				关闭

4. 在新建MaxCompute数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建MaxCompute数据	音源 ×
*数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 ,
* 数据源配置 💿 "生产+开发"数据源) "生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源	
* Endpoint	请正确输入,如http://service.odps.aliyun.com/api		* Endpoint	请正确输入,如http://service.odps.aliyun.com/api
* Project Name	请输入英文项目名称		* Project Name	请编入英文项目名称
* Access ID	100.00		* Access ID	请输入鉴权Access ID,为保证任务正常执行,请确保有所需数据…
* Access Key		₿	* Access Key	请输入鉴权Access Key 🔒
i			6	
				测试连接取消 确定

⑦ 说明 通常,生产数据源和开发数据源需要配置不同的参数值,但Dataphin也支持配置成相同的参数值。

参数	描述		
	MaxCompute的Endpoint,建议填写阿里云VPC网络下的Endpoint。 如何获取Endpoint,请参见 <mark>Endpoint</mark> 。		
Endpoint	↓ 注意 当您使用VPC网络连接Dataphin与MaxCompute时,此处需要 配置为Dataphin实例所在地域下MaxCompute的Endpoint。		
Project Name	此处为MaxCompute项目名称,非DataWorks工作空间名称。 您可以登录MaxCompute控制台,左上角切换地域后,即可在项目管理页签查 看到具体的MaxCompute项目名称。 项目管理 服务资源 《 查询编辑 《 Notebook分析 《 管家 创健项目 请输入MaxCompute项目名称进行搜索 Q C MaxCompute项目名称 HaxCompute地域 计要方式 所属DataWorks工作空间 pj_test 华北2 (北京) 包年包月 pj_test (生产环境)		
Access ID、Access Key	当前账号的AccessKey ID和AccessKey Secret。 如何获取,请参见 <mark>获取AccessKey</mark> 。		

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成MaxCompute数据源的创建。

后续步骤

完成MaxCompute数据源的创建后,即可将MaxCompute的业务数据引入至Dataphin,或将Dataphin的数据 写入至MaxCompute数据源。具体操作,请参见配置离线单条管道。

6.5.2. 创建SAP HANA数据源

创建SAP HANA数据源用于实现Dataphin能够读取SAP HANA的业务数据,及能够向SAP HANA写入数据。 在引入SAP HANA的业务数据至Dataphin和将Dataphin的数据写入至SAP HANA的场景中,您需要先创建 SAP HANA数据源。本文为您介绍如何创建SAP HANA数据源。

背景信息

SAP HANA是一款具有高性能内存的应用平台。如果您使用的是SAP HANA,在对接Dat aphin进行数据开发 或将Dat aphin的数据导出至SAP HANA场景中,您需要先完成SAP HANA数据源的创建。更多SAP HANA信息,请参见SAP HANA官网。

Dataphin支持创建的SAP HANA数据源类型包括公网数据库和ECS(VPC)自建数据源。不同类型的数据源数据连通、访问的网络类型不同,所需的配置细节不一致,如下图所示。



数据源类型	网络连通说明	网络连通操作指导
公网数据库	即具备公网访问能力的数据库。访问方式默认为 公网IP,直接配置网络连通参数。	1.步骤一:添加Dataphin IP地址至数据库白 名单
ECS(VPC) 自建数据源	数据源是在阿里云ECS上自建的数据库,访问方 式默认为VPC网络,直接配置网络连通参数。	2. 步骤二: 创建SAP HANA数据源

为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

• Dataphin仅支持超级管理员和项目管理员角色创建数据源。

如何添加并授予项目管理员角色的项目成员,请参见添加项目成员。

• 仅当VPC网络的SAP HANA数据源和Dat aphin实例在同一地域时,才支持将该数据源添加至Dat aphin。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。

公网数据库、阿里云数据库且访问方式为公网访问,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● ECS(VPC)自建数据源,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。如何添加白名单,请 参见添加安全组规则。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建SAP HANA数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 在Dataphin首页,单击顶部菜单栏的规划。
- 3. 按照下图指引,进入新建数据源页面。

Dataphín	规划								° <u>o</u> 🕓
市会	数据源(190)①						Q. 清給入證素关證字	T	+ aretzete
 数据编构 W 公共定义 	guzhi_oss 전 변화: OSS			ē	lfs_lindorm 四 問題: Lindorm				Ē
項目	生 ^{yte} fo	e Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①
▶ 项目管理	负责人: 1								
411 it 11282	数据原用注: 高线 创建用户: True Test True Antoine		数据源用途: 发始 创建用户: <mark>Manufalantillegen (Manufalanti</mark>)		数据源用途: 本线 创建用户:		数据源用途: 淘试 创建用户:		
ं () ड इस्टर	创建的问:2022-03-09 20:27:15 更新的问:2022-03-11 12:25:27 報源保留: ◆ 宣告相決定局		会議的前に2022-03-09 20 30 32 単新的前に2022-03-01 12 25 33 特別保護:		出題时间:2022-03-09 11:03:47 更新时间:2022-03-09 11:03:47 税約単意:		创建时间: 2022-03-09 11:03:47 更新时间: 2022-03-09 11:03:47 经按信意:		
	损作:	6 C B A	操作:	6 6 6 A	操作:	6 6 8 A	請作:	e 🛛	8 A

4. 在新建数据源对话框的大数据存储区域,选择SAP HANA。

如果您最近使用过SAP HANA,也可以在**最近使用**区域选择SAP HANA。同时,您也可以在快搜索框中,输入SAP HANA的关键词,快速筛选。

新建数据源			Q 请输/	>、搜索关键字 全部 ∨ X
大数据存储				
\sim		SAP HANA	*	HOLOGRES
MaxCompute	Hive	SAP HANA	DataHub	Hologres

5. 在新建SAP HANA数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建SAP HANA数据	æ
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据源配置 💿 "生产+开发"数据源	"生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源			
* JDBC URL	jdbc:sap://1 132:39041/?databaseName=HXE¤	itS	* JDBC URL	jdbc:sap://1 :3	9041/?databaseName	=HXE¤tS
* 用户名	Dataphin		* 用户名	Dataphin		
* 密码		₽	* 密码			∂
* 类型	○ 公网数据库 ● ECS(VPC)自建数据源		* 类型	○ 公网数据库 ● ECS(VP	C)自建数据源	
* 区域	华东2(上海)		* 区域	华东2(上海)		
* VPC ID	vpc-bp1 3o3jit		* VPC ID	vpc-bp1 43o3jit		
* ECS ID	i-uf6(0u121		* ECS ID	i-uf60si 121		
提示: 当前使用的: 100.104.228.128/2	数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 26;100.104.115.192/26		提示: 当前使用的 100.104.228.128/2	数据库为VPC网络环境下阿里云 6;100.104.115.192/26	数据库,连接请指定报	叙IP白名单:
				测试连接	取消	确定

⑦ 说明 通常,生产数据源和开发数据源需要配置不同的参数值,但Dataphin也支持配置成相同的参数值。

参数	描述
JDBC URL	链接地址的格式为 jdbc:sap://{链接地址}:{端口}/?databaseName={yourdataba seName}¤tSchema={yourcurrentSchema} 。
用户名、密码	登录SAP HANA数据库的用户名和密码。

参数	描述							
类型	 Dataphin支持选择以下几种数据源类型: ■ 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 ■ ECS(VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的网络环境为VPC,请选择ECS(VPC)自建数据源。 							
区域、VPC ID、ECS ID	当类型选择为ECS(VPC)自建数据源,还需配置以下这些网络参数: 区域:仅支持选择当前Dataphin实例所在地域。 ⑦ 说明 仅支持添加与Dataphin实例在同一地域的SAP HANA数据源至 Dataphin。 							
	 PC ID: 填写云服务器ECS实例的VPC ID. ECS ID: 填写云服务器ECS实例的D。 您可以在云服务器ECS控制台的实例详情页面,获取VPC ID和ECS ID. ^{Z服装器 ECS / 实例 / 实例详简} ← launch 原数 度 度 度 度 度 度 度 度 度 度 度 度 度 度 度 度 度 度 度							
	网络类型 专有网络 专有网络 vpc-uf60mt ご 弹性网卡 eni-uf6. : 虚拟交换机 vsw-uf637 ご							

6. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。 7. 测试成功后,单击**确定**,完成SAP HANA数据源的创建。

后续步骤

完成SAP HANA数据源的创建后,即可将SAP HANA的业务数据引入至Dataphin,或将Dataphin的数据写入 至SAP HANA数据源。具体操作,请参见配置离线单条管道。

6.5.3. 创建DataHub数据源

创建DataHub数据源用于实现Dataphin能够读取DataHub的业务数据。在引入DataHub的业务数据至 Dataphin中进行数据开发的场景中,您需要先创建DataHub数据源。本文为您介绍如何创建DataHub类型的 数据源。

背景信息

DataHub即阿里云流式数据服务DataHub,如果您使用的是阿里云流式数据服务DataHub,在对接Dataphin 进行数据开发时,您需要先完成DataHub数据源的创建。更多DataHub信息,请参见产品概述。 Dataphin支持通过VPC网络和公网连接DataHub。Dataphin连接DataHub时,可以选择DataHub在VPC网络、经典网络、公网环境下对应的Endpoint,DataHub的Endpoint请参见域名列表。通常,VPC网络较公网和经典网络稳定,建议您使用VPC网络连通Dataphin和DataHub。

使用限制

Dataphin仅支持超级管理员和项目管理员角色创建数据源。
 如何添加成员并授予项目管理员角色,请参见添加项目成员。

• 当您使用VPC网络连接Dataphin与DataHub时,仅支持连接与Dataphin在同一地域的DataHub。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規划									ి <mark>(</mark> 2)	
亦装	数据源 (190) ①						Q 请输入提供关键字	1	+ 862	Pro Bick
▲ 数据转档	guzhi_oss 전 브랜: oss			Ē	Hs_lindorm [2]					÷
项目	生/%	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	,	or Dev (0
□ 项目管理 ■ 计算数	负退人: 数据源用法: 建线		会長人 数据源用途: 案成		负责人: 数据原用途: 集成		负责人: 数据原用注: 案线			
ं () = राष्ट्रह	1000-00-07 創建时间: 2022-03-09 20 27:15 更新时间: 2022-03-11 12:25:27 総務信息: ◆ 宣告任法信号		 (1)建行時:2022-03-09 20 30 32 (第所行時:2022-03-11 12 25:33 (総合単常) (1) 金箔結映信息 		(1999-1977) 台連町河町: 2022-03-09 11:03:47 東新町河町: 2022-03-09 11:03:47 街線電源:		1000-00 ⁻⁰ 伝達的時に2022-03-09 11:03:47 運動的時に2022-03-09 11:03:47 気候信息:			
	·	0 C E A	操作:	@ ¤ = &	操作:	6 5 5 A	旗作:	@	2 13 4	8

3. 在新建数据源对话框的大数据存储区域,选择DataHub。

如果您最近使用过DataHub,也可以在**最近使用**区域选择DataHub。同时,您也可以在快搜索框中,输 入DataHub的关键词,快速筛选。



4. 在新建DataHub数据源对话框中,配置连接数据源参数后,单击确定。
i. 配置数据源的基本信息。

返回新建数据源 > 新建DataHub数据源	X
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 /
* 数据源配置 💿 "生产+开发"数据源 🤇	"生产"数编源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源		
* endPoint	https://dh-cn-hangzhou.aliyuncs.com		* endPoint	https://dh-cn-hangzhou.aliyuncs.com	
* Project Name	qd_test		* Project Name	qd_test	
* Access Id	LTAI5 /f		* Access Id	LTAI	
* Access Key		8	* Access Key		₽
				测试连接 取 消	确定

⑦ 说明 通常,生产数据源和开发数据源需要配置不同的参数值,但Dataphin也支持配置成 相同参数值。

参数	描述					
	Dat <i>a</i> Hub的Endpoint,建议填写阿里云VPC网络下的Endpoint。 如何获取Endpoint,请参见 <mark>域名列表</mark> 。					
Endpoint	注意 如果填写阿里云VPC网络下的Endpoint,则仅支持填写与 Dataphin实例在同一地域下的Endpoint。					
Project Name	DataHub项目名称。在DataHub控制台页面,查看Project Name。					
Access ID、Access Key	当前账号的AccessKey ID和AccessKey Secret。 如何获取,请参见获取AccessKey。					

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成DataHub数据源的创建。

后续步骤

完成DataHub数据源的创建后,即可将DataHub的业务数据引入至Dataphin,或将Dataphin的数据写入至 DataHub数据源。具体操作,请参见配置离线单条管道。

6.5.4. 创建Hologres数据源

新建Hologres数据源用于实现Dataphin能够读取Hologres的业务数据,及能够向Hologres写入数据。在引入 Hologres的业务数据至Dataphin和将Dataphin的数据写入至Hologres的场景中,您需要先创建Hologres数 据源。本文为您介绍如何新建Hologres数据源。

背景信息

Hologres即阿里云交互式分析Hologres,是阿里巴巴自主研发的一款交互式分析产品。如果您使用的是 Hologres,在对接Dataphin进行数据开发时,您需要先完成Hologres数据源的创建。更多Hologres信息,请 参见什么是实时数仓Hologres。 Dataphin支持通过VPC网络、经典网络、公网连接Hologres。Dataphin连接Hologres时,可以选择Hologres在VPC网络、经典网络、公网环境下对应的Endpoint,Hologres的Endpoint请参见访问域名。通常,VPC网络较公网和经典网络稳定,建议您选择VPC网络连通Dataphin和Hologres。

为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中,更多对接配置详情请参见设置IP白名单。

使用限制

• Dataphin仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

• 当您使用VPC网络连接Dataphin与Hologres时, 仅支持连接与Dataphin实例在同一地域的Hologres。

步骤一:添加Dataphin IP地址至Hologres白名单

基于您选择连通Dataphin与Hologres的网络类型,添加对应网络类型下的Dataphin IP地址至Hologres白名单。如何添加白名单,请参见设置IP白名单。

● 使用公网、经典网络连通Dataphin与Hologres时,需要添加公网下Dataphin IP地址至Hologres白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● 使用VPC网络连通Dataphin与Hologres时,需要添加VPC网络下Dataphin IP地址至Hologres白名单中。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址			
华东2(上海)	100.104.0.0/16	100.104.228.128/26、 100.104.115.192/26			
华南1(深圳)		100.104.48.128/26			
华北2(北京)		100.104.238.64/26			
华东1(杭州)		100.104.45.0/26			

步骤二: 创建Hologres数据源

1. 请参见入口介绍,进入数仓规划页面。

2. 按照下图指引,进入新建数据源页面。

								凿 🦨	° <mark>0</mark> 🙎
亚 号	数据源 (190) ①						Q 请输入搜索关键字	V	+ 85824039539
▲ 数据架构 ₩ 公共定义	guzhi_oss 전 밝힌: OSS			Ē	lfs_lindorm ⊠ 問題: Lindorm				÷
項日	生产 fo	or Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①
□ 项目管理 ● 可目管理	负责人: 数强调用途: 高线				交換人: 数据原用途: 高线		负责人: 数据须用途: 高线		
 2 (1) 3 83889	1022円円 創種时间: 2022-03-09 20:27:15 更新时间: 2022-03-11 12:25:27 秋时尚見: ◆ 宣教授会信息		10月日日 会議時時間: 2022-03-09 20:30:32 面积時間: 2022-03-11 12:25:33 (549)注意: ② 宣誓指決信意		11.20月戸: 創設対応: 2022-03-09 11:03:47 原形対応: 2022-03-09 11:03:47 影形対応: 2022-03-09 11:03:47		 (1)逆用户: (1)逆用戶: (2)22-03-09 11:03:47 (2)22-03-09 11:03:47 (5)均衡目: (4) 意識保護信用 		
	操作:	e 12 8 A	2011年1	@ 12 13 <u>2</u>	(泉)作:	0 1 1 A	課/F:	e 🗈	8 A

3. 在新建数据源对话框的大数据存储区域,选择Hologres。

如果您最近使用过Hologres,也可以在**最近使用**区域选择Hologres。同时,您也可以在快搜索框中, 输入Hologres的关键词,快速筛选。

新建数据源	Q 请输入!	史家关键字 全部 ∨ X		
大数据存储	sine ^{some So} sine ^{some So} si	re ^{alemniss} supel ^{alemniss} s	and a sector 2000 sector 2000 sector	
N	Winte	SAP HANA	×	HOLOGRES
MaxCompute	Hive	SAP HANA	DataHub	Hologres
				关闭 _{Mad} a

- 4. 在新建Hologres数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建Hologres数据源	×
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 //
* 数据源配置 💿 "生产+开发"数据源)"生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源		
* Endpoint	hgpostcn-	i	* Endpoint	hgpostc -cn-shanghai.hologres.aliyuncs.com	n (i)
* dbname	hologres_poc		* dbname	hologres_poc	
* 用户名	LTA		* 用户名	LTA	
* 密码	Ø		* 密码		5
				取消 确:	e datap ^r

② 说明 通常,生产数据源和开发数据源需要配置不同的参数值,但Dataphin也支持配置成 相同的参数值。

参数	描述					
	Hologres实例的Endpoint,建议使用阿里云VPC网络下的Endpoint。如何获取 Endpoint,请参见 <mark>访问域名</mark> 。					
Endpoint	 注意 当您使用VPC网络连接Dataphin与Hologres时,此处需要配置 为Dataphin实例所在地域下Hologres的Endpoint。 					
DBName	在Hologres控制台的Database管理页面,查看数据库名称。具体操作,请参 见DB管理。					
Access ID、Access Key	当前账号的AccessKey ID和AccessKey Secret。如何获取,请参见 <mark>获取</mark> AccessKey。					

5. 单击确定,完成Hologres数据源的创建。

后续步骤

完成Hologres数据源的创建后,即可将Hologres的业务数据引入至Dataphin,或将Dataphin的数据写入至 Hologres数据源。具体操作,请参见配置离线单条管道。

6.6. 文件数据源

6.6.1. 创建FTP数据源

创建FTP数据源用于实现Dataphin能够读取FTP的业务数据,及能够向FTP写入数据。在引入FTP的业务数据 至Dataphin和将Dataphin的数据写入至FTP的场景中,您需要先创建FTP数据源。本文为您介绍如何创建FTP 数据源。

背景信息

FTP即文件传输协议。如果您使用的是FTP,在对接Dataphin进行数据开发或将Dataphin的数据写入至FTP的场景中,您需要先完成FTP数据源的创建。更多FTP信息,请参见FTP官网。

Dataphin支持通过公网连接FTP,连接FTP的协议支持FTP、SFTP或FTPS。为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin公网环境下的IP地址段添加到数据源的白名单中。

使用限制

- Dataphin仅支持超级管理员和项目管理员角色创建数据源。
 如何添加项目成员并授予项目管理员角色,请参见添加成员。
- 目前, FTP数据源不支持通过VPC网络连通Dataphin实例。

步骤一:添加Dataphin IP地址至FTP白名单中

您需要将公网环境下的Dataphin IP地址添加至FTP白名单中。

公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

步骤二: 创建FTP数据源

1. 请参见入口介绍,进入数仓规划页面。

2. 按照下图指引,进入新建数据源页面。

≡	Dataphin 规划									² 2	8
<u>458</u>		数据源 (190) ①						Q 清箱入提索关键字	V	+ 55222	959
- 55 ₩ 22	振発約 共定义	guzhi_oss ⊠ ⋈ﷺ: oss			Ē	lfs_lindorm (2) 변문: Lindorm					÷
項目		生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①	
■ 項	目管理	负责人: 1.0001000 10001000 100010						SBA: monometric and the second			
॥ भ	1120	数据源用途: 高线 创建用户:		政策運用途: 本は 公理用 ^A : Number (Number)		数据源用途: 高线 创建用户:		政策原用途: 集试 创建用户:			
19 8 53	1) #27	出陸时间: 2022-03-09 20:27:15 更新时间: 2022-03-11 12:25:27 現後信息: ◆ 宣告相決信息		公開時間: 2022-03-09 20 30 32 更新時間: 2022-03-11 12 25 33 税務准督:		出題対応:2022-03-09 11:03:47 第5時分応:2022-03-09 11:03:47 局約後期意:		 (1)登付前: 2022-03-09 11:03:47 (第)受付前: 2022-03-09 11:03:47 (現象信意: ● 宣告知時位息 			
		操作:	C I I A	操作:	6 6 8 A	操作:	6 6 8 8	staria:	6 2	B &	

3. 在新建数据源对话框的文件区域,选择FTP。

如果您最近使用过FTP,也可以在**最近使用**区域选择FTP。同时,您也可以在快搜索框中,输入FTP的关键词,快速筛选。

新建数据源			Q 请输入搜索关键字	全部 🗸 X
│文件	seren alara su Seren alara su	selectron St. Streaktorn St.		
1 CO	Ē	•		
HDFS	FTP	OSS		
				关闭。同时

4. 在新建FTP数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建FTP数据源	×
*数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 //
∗数据源配置 ● "生产+开发"数据源 ("生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。生产数据源和开发数据源需要配置的参数相同。

生产数据源			开发数据源		
* Protocol	● FTP ○ SFTP ○ FTPS		* Protocol	• FTP O SFTP O FTPS	
* Host	114 Suver		* Host	114 44	
* Port	21		* Port	21	
* 用户名	dataphin		* 用户名	dataphin	
* 密码		₿	* 密码		8
				测试连接 取消	确定。

⑦ 说明 通常,生产数据源和开发数据源配置的连接参数不同,但Dataphin也支持配置成相同的连接参数。

参数	描述
Protocol	基于FTP服务器使用的协议,选择对应的文件传输协议。目前支持传输协议包括 以下几种: FTP:文件传输协议,用于控制文件的双向传输,同时也是一个应用程序。 SFTP:基于SSH的安全文件传送协议,为传输文件提供一种安全的加密方 法。 FTPS:基于SSL/TLS的文件传输协议,相当于加密的FTP。
Host	FTP服务器地址。
Port	FTP服务器端口。
用户名、密码	FTP服务器的访问用户名和密码。
SSLImplicit、conne ctPattern	 当Protocol选择为FTPS时,需要配置以下参数: SSLImplicit:隐式模式。当FTP服务器上开启了这个协议,则需要选择TRUE,否则选择FALSE。 connectPattern:连接模式,包括以下两种: PORT(主动模式):客户端打开一个端口等服务器建立数据连接。 PASV(被动模式):服务器打开一个端口等客户端建立数据连接。

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。 6. 测试成功后,单击**确定**,完成FTP数据源的创建。

后续步骤

完成FTP数据源的创建后,即可将FTP的业务数据引入至Dataphin,或将Dataphin的数据写入至FTP数据源。 具体操作,请参见配置离线单条管道。

6.6.2. 创建OSS数据源

创建OSS数据源用于实现Dataphin能够读取OSS的业务数据,及能够向OSS写入数据。在引入OSS的业务数据至Dataphin和将Dataphin的数据写入至OSS的场景中,您需要先创建OSS数据源。本文为您介绍如何创建OSS类型的数据源。

背景信息

OSS即阿里云对象存储OSS。如果您使用的是阿里云对象存储OSS,在对接Dataphin进行数据开发或将 Dataphin的数据写入至OSS时,您需要先完成OSS数据源的创建。更多OSS信息,请参见什么是对象存储OSS。

Dataphin支持通过VPC网络和公网连接OSS。Dataphin连接OSS时,可以选择OSS在VPC网络或公网环境下对应的Endpoint,OSS的Endpoint请参见访问域名和数据中心。通常,VPC网络较公网稳定,建议您选择VPC网络连通Dataphin和OSS。

使用限制

Dataphin仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規定	9							1 ^{# 4} 2 🕓
而發	数据源 (190) ①						Q 清除入服用关键字	T + 8622.038522
▲ 数据架构 ₩ 公共定义	guzhi_oss ₪ Mana: oss			Ē	lfs_lindorm (2) (전문): Lindorm			Đ
项目	生 ^{per}	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for Dev ①
□ 项目管理	() () () () () () () () () () () () () (SEA: Topological Information	
≪∥ 计算数	数据源用述: 素试 创建用户: The Test of the Test of the Test	1	新聞原用途: 本対 1928日中: Describe -Class (Booker)		政策原用途: 集成 《陳用戶: Numine Automation		較振荡用途: 素減 创建用户:	
³⁵ (1)	金融時前间: 2022-03-09 20:27:15 更新時前间: 2022-03-11 12:25:27		台連时间: 2022-03-09 20:30:32 更新时间: 2022-03-11 12:25:33		会理的対応: 2022-03-09 11:03:47 更新时间: 2022-03-09 11:03:47		台通时间: 2022-03-09 11:03:47 更新时间: 2022-03-09 11:03:47	
S BUSER	● 並后相接位思		1212(11): ◆ 查看任法信息		● 查看相接信息		组织信息: · · · · · · · · · · · · · · · · · · ·	
	操作:	e e i 2	损件 :	@ E E &	操作:	C 🖻 🔒	請作:	6 8 8 A

3. 在新建数据源对话框的文件区域,选择OSS。

如果您最近使用过OSS,也可以在**最近使用**区域选择OSS。同时,您也可以在快搜索框中,输入OSS的 关键词,快速筛选。



4. 在新建OSS数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建OSS数据源	×
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据源配置 💿 "生产+开发"数据源	○ "生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。生产数据源和开发数据源需要配置的参数相同。

生产数据源			开发数据源		
* Endpoint	请正确输入地址, 如: http://oss.aliyuncs.com		* Endpoint	请正确输入地址,如:http://oss.aliyuncs.com	n superkonnisi Nor
* Bucket	清输入Bucket		* Bucket	请输入Bucket	
CNAME	请输入CNAMI		CNAME	请输入CN-winc	
* Access ID	请输入鉴权Access ID,为保证任务正常执行,请确保有所需数据		* Access ID	请输入鉴权Access ID,为保证任务正常执行,	请确保有所需数据
* Access Key	请输入鉴权Access Key	₿	* Access Key	请输入鉴权Access Key	₿
				测试连接取消	确定

⑦ 说明 通常,生产数据源和开发数据源配置的连接参数不同,但Dataphin也支持配置成相同的连接参数。

参数	描述
Endpoint	OSS所在地域对应的Endpoint,格式为 http://{oss-Region}.aliyuncs.c om ,其中Region为Bucket所在地域。例如,华东1(杭州)的Endpoint填写 为 https://oss-cn-hangzhou.aliyuncs.com OSS服务的Endpoint和区域有关。访问不同的区域时,需要填写不同的域名。
Bucket	OSS所在地域对应的Bucket信息。用于存储对象的容器。在Bucket 列表页面, 获取OSS所在地域对应的Bucket。 您可以创建一个或多个存储空间,每个存储空间可添加一个或多个文件。您可以 在数据同步任务中查找此处输入的存储空间中相应的文件,没有添加的存储空 间,则不能查找其中的文件。
CNAME	配置OSS的CNAME。
Access ID、Access Key	当前账号的AccessKey ID和AccessKey Secret。 如何获取,请参见 <mark>获取AccessKey</mark> 。

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成OSS数据源的创建。

后续步骤

完成OSS数据源的创建后,即可将OSS的业务数据引入至Dataphin,或将Dataphin的数据写入至OSS数据 源。具体操作,请参见配置离线单条管道。

6.7. 消息队列数据源

6.7.1. 创建Log Service数据源

创建Log Service数据源用于实现Dataphin能够读取Log Service的业务数据。在引入Log Service的业务数据 至Dataphin中进行数据开发的场景中,您需要先创建Log Service数据源。本文为您介绍如何创建Log Service类型的数据源。

背景信息

Log Service即阿里云日志服务,如果您使用的是阿里云日志服务,在对接Dat aphin进行数据开发时,您需要 先完成Log Service数据源的创建。更多Log Service信息,请参见什么是日志服务。

Dataphin支持通过VPC网络和公网连接Log Service。Dataphin连接Log Service时,可以选择Log Service在 VPC网络或公网环境下对应的Endpoint,Log Service的Endpoint请参见服务入口。通常,VPC网络较公网稳 定,建议您选择VPC网络连通Dataphin和Log Service。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

当业务数据源(Log Service)使用VPC网络时,则仅支持创建与Dataphin实例在同一地域Log Service数据源。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規划								凿 🧔	A_2 🙎
业务	数据源 (190) ☉						Q 遗输入理索关键字		+ 8/62209630
点 数据编约	auzhi oss Pi				Ifs lindorm Pi				
₩ 公共定义	类型: OSS			E	問題: Lindorm				
项目	生yter fo	e Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	r Dev ①
▶ 项目管理	负责人:		SA. Nonitan Special Solution		SEA: Instanting the groups				
● 计算源	数据源用途: 集线 创建用户:「 Numines in a sectoriza tion		政策源用波: 站试 《建用 ^小 : <mark>Manalakana (Agan (Manalakana</mark>)		設備原用途: 株式 ()球用件: 「一」の「日本」は、「日本」の「日本」		政派原用途: 集线 创建用户:		
a 🚹	会議時前時: 2022-03-09 20:27:15 第5時前時: 2022-03-11 12:25:27		金融時间: 2022-03-09 20 30 32 第15日前日: 2022-03-11 12 25 33		台通知対応: 2022-03-09 11:03:47 第5日1日: 2022-03-09 11:03:47		创建时间: 2022-03-09 11:03:47 東部日前: 2022-03-09 11:03:47		
i i i i i i i i i i i i i i i i i i i	● 重新結果信息		◎ 宣誓任笑信息		● 查看性狭位目				
	操作:	0008	续件:	0002	操作:	0 1 B A	操作:	0 2	

3. 在新建数据源对话框的消息队列区域,选择Log Service。

如果您最近使用过Log Service,也可以在**最近使用**区域选择Log Service。同时,您也可以在快搜索框中,输入Log Service的关键词,快速筛选。

新建数据源				Q. 请输入搜索关键字	全部 🗸 X
消息队列	Sime American Super American Su	uper ^{Salmer SV} Super Salmer SV S	aper AleminicSV Super A		
	*	<u></u>	anter anter anter a		
Log Service	Kafka S	RocketMQ			
					关闭。同时

4. 在新建Log Service数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源	> 新建Log Service数据源	×
* 数据源名称	请输入数据源名称	
数据源描述	请输入数据源描述 0112	8 //
* 数据源配置	● "生产+开发"数据源 ○ "生产"数据源	
参数	描述	

学权	田处
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源			
* LogHub Endpoint	如: http://cn-shanghai.log.aliyuncs.com		* LogHub Endpoint	如: http://cn-shanghai.log.aliyuncs	s.com	
* Project	请输入项目名称		* Project	请输入项目名称		
* Access ID	请输入鉴权Access ID,为保证任务正常执行,请确保有所需数据.		* Access ID	请输入鉴权Access ID,为保证任务	正常执行,	请确保有所需数据
* Access Key	请输入鉴权Access Key	₿	* Access Key	请输入鉴权Access Key		£
				测试连接	取消	确定

② 说明 通常,生产数据源和开发数据源配置参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述		
	Log Service的Endpoint,建议配置阿里云VPC网络下的Endpoint。如何获取 Endpoint,请参见 <mark>服务入口</mark> 。		
LogHub Endpoint	兌注意 仅支持配置与Dataphin实例在同一地域的Log Service Endpoint。		
Project	Log Service项目名称。在Log Service控制台页面,查看Project名称。		
Access ID、Access Key	当前账号的AccessKey ID和AccessKey Secret。如何获取,请参见 <mark>获取</mark> AccessKey。		

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通解决方案。

6. 测试成功后,单击确定,完成Log Service数据源的创建。

后续步骤

完成Log Service数据源的创建后,即可将Log Service的业务数据引入至Dat aphin。具体操作,请参见配置 离线单条管道。

6.7.2. 创建Kafka数据源

创建Kafka数据源用于实现Dataphin能够读取Kafka的业务数据,及能够向Kafka写入数据。在引入Kafka的业务数据至Dataphin和将Dataphin的数据写入至Kafka的场景中,您需要先创建Kafka数据源。本文为您介绍如何创建Kafka数据源。

背景信息

Kafka是一种消息队列,用于处理实时数据。Dataphin支持对接Kafka09、Kafka010、Kafka011。如果您使用的是Kafka,在对接Dataphin进行数据开发或将Dataphin的数据导出至Kafka,您需要先完成Kafka数据源的创建。Kafka不同版本的功能介绍,请参见官方文档:

- Kafka09
- Kafka010
- Kafka011

Dataphin支持通过公网连接Kafka。为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin公网环境下的IP地址段添加到数据源的白名单中。

使用限制

Dataphin仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

步骤一:添加Dataphin IP地址至Kafka白名单

您需要将公网环境下的Dataphin IP地址添加至至Kafka访问白名单中。

公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

步骤二: 创建Kafka数据源

1. 请参见入口介绍,进入数仓规划页面。

2. 按照下图指引,进入新建数据源页面。

								њ¢ с. 🧕 🧕
#8	数据源 (190) ①						Q 请输入继续关键字	V + 5582839539
■ 数据架构 ₩ 公共定义	guzhi_oss ⊠ 講型: OSS			ē	ifs_lindorm ⊠ ≋≣: Lindorm			ē
項日	生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for Dev ①
□项目管理 (II) 计算派 注 (II)	(公司人: 約回送用用: 利助司用: 本紙 回動司用: 202243-50 2027:15 回時时用: 202243-11 122527 4回款回見: ● 面看任田田県		会議人 数据期時法 義統 総理時時、2022-05-06 20:30-32 期時時時、2022-05-06 20:30-32 期時時時、2022-05-11 12:25:33 総計標題: ● 面看信用価格		(人), 約集用用: 東岐 (設用用: (設計用用: 202245-09 11:5):47 面积1月: 202245-09 11:5):47 面积1月: 202245-09 11:5):47 価額202245-09 11:5):47		 (2) (2) (2) (2) (2) (2) (2) (2) (2) (3) (4) (4)	
	第 7年:	@ 12 B A	探你 :	0 B B A	操作 :	0 1 B A	探作 :	6 E B A

3. 在新建数据源对话框的消息队列区域,选择Kafka。

如果您最近使用过Kaf ka,也可以在**最近使用**区域选择Kaf ka。同时,您也可以在快搜索框中,输入 Kaf ka的关键词,快速筛选。

新建数据源			Q	请输入搜索关键字	全部 🗸 X
Ⅰ消息队列	southern southern so	per hámin ^{issa} . Super hámin ^{issa} , Su			
Q.	*	<u>¢</u> .			
Log Service	Kafka S	RocketMQ			
					关闭。同时

- 4. 在新建Kafka数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建Kafka数据源	×
*数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据源配置 💿 "生产+开发"数据源	○ *生产数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。生产数据源和开发数据源需要配置的参数相同。

生产数据源		开发数据源		
* kafka集群地址	192.0.2 ,192.0.2 ,192.0.2	* kafka集群地址	192.0.2 ,192.0.2 ,192.0.2	
			取消	确定

⑦ 说明 通常,生产数据源和开发数据源配置的连接参数不同,但Dataphin也支持配置成相同的连接参数。

参数	描述
	配置格式 {ip/domain name}:{port} ,如果需要配置多个节点的地址使用 英文逗号 (,) 隔开。
Kafka年発告計	Dataphin支持对接Kafka09、Kafka010、Kafka011,如何配置不同版本的 Kafka集群地址,请参见官方文档:
	Kafka09
	Kafka010
	Kafka011

↓ 注意

- 创建Kafka数据源时不支持测试数据源是否可以和Dataphin进行正常的连通,因此需要 您确保连接信息的正确性。
- 不支持Kafka的Kerberos认证的方式。

5. 单击确定,完成Kafka数据源的创建。

后续步骤

完成Kafka数据源的创建后,即可基于Kafka的业务数据开发流批一体任务。具体操作,请参见创建Flink SQL 任务、步骤一:创建并调试Flink Template SQL计算任务、创建Flink DataStream计算任务。

6.7.3. 创建RocketMQ数据源

创建Rocket MQ数据源用于实现Dat aphin能够读取Rocket MQ的业务数据。在引入Rocket MQ的业务数据至 Dat aphin中进行数据开发的场景中,您需要先创建Rocket MQ数据源。本文为您介绍如何创建Rocket MQ数据 源。

背景信息

RocketMQ即阿里云消息队列RocketMQ版,如果您使用的是阿里云消息队列RocketMQ版,在对接Dataphin 进行数据开发时,您需要先完成RocketMQ数据源的创建。更多阿里云消息队列RocketMQ版信息,请参见什 <u>么是消息队列RocketMQ版</u>?。

Dataphin支持通过公网或VPC网络连接RocketMQ。Dataphin连接RocketMQ时,可以选择RocketMQ在VPC 网络或公网环境下对应的Endpoint。通常,VPC网络较公网稳定,建议您选择VPC网络连通Dataphin和 RocketMQ。

使用限制

- 仅支持超级管理员和项目管理员角色创建数据源。
 如何添加项目成员并授予项目管理员角色,请参见添加项目成员。
- 当您使用VPC网络连接Dataphin与Rocket MQ时, 仅支持连接与Dataphin实例在同一地域的Rocket MQ。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規定									凿 🗸	^{් ද} 2
	数据源 (190) ①							Q 请输入提供关键字	T	+ 8522
▲ 数据服物 ₩ 公共定义	guzhi_oss ⊠ 类型: OSS			Ē	Ifs_lindorm ⊠ ≋≣: Lindorm					
	生) ^{te}	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic (开发		6	x Dev 0
□ 项目管理 【 计算源 ③ ① ● 数据源	交流人: 数据原用法: 素純 必要用用: 空間用用: 空間用用: 空間用用: 空間用用: 空間243-11122527 転用日期: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型用: 空間面相用型: 空間面面相用型: 空間面面面面目 : 空間面面面目 : 空間面面面目 : 空間面面目 : 空間面面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空間面目 : 空目: 空目		交通人 日本 1000 1000		たまた: 数単原用念: 単純 (清用用: (清用用: (清明)(2022-03-60 11 03-47 単新行道: 2022-03-60 11 03-47 単新行道: 2022-03-60 11 03-47 単新行道: ① 宣告/紀承留			13-09 11:03:47 13-09 11:03:47 13:09 11:03:47		
	授作:	@ 12 B A	操作:	@ Ø Ø &	操作:	0 0 0 1			e (2 11 8

3. 在新建数据源对话框的消息队列区域,选择Rocket MQ。

如果您最近使用过Rocket MQ,也可以在**最近使用**区域选择Rocket MQ。同时,您也可以在快搜索框中,输入Rocket MQ的关键词,快速筛选。

新建数据源				Q 请输入搜索关键字	全部 🗸 X
消息队列	Santa Anton ^{i Sal} i Santa Antoni Sali _{Sa} nta	e adminiSu Supe adminiSu Su	o ^{n barrandon} sur 1		
Q	*	<u></u>	es de constantes de la superiori		
Log Service	Kafka S	RocketMQ			
					关闭。如何

4. 在新建Rocket MQ数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建RocketMQ数据》	ξ ×
* 数据源名称 请输入数据逐名称 数据源描述 请输入数据源图描述	0/128
* 数据源配置 💿 "生产+开发"数据源	"生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源		开发数据源	
* Endpoint	.com:8080	* Endpoint	com:8080
Instance ID	请输入鉴权Instance ID,为保证任务正常执行,请确保有所需数…	Instance ID	请输入鉴权Instance ID,为保证任务正常执行,请确保有所需数
* Access ID	请输入鉴权Access ID,为保证任务正常执行,请确保有所需数据…	* Access ID	请输入鉴权Access ID,为保证任务正常执行,请确保有所需数据
* Access Key	请输入鉴权Access Key,为保证任务正常执行,请确保有所…	* Access Key	请输入鉴权Access Key,为保证任务正常执行,请确保有所 🔒
			取消 确定

⑦ 说明 通常,生产数据源和开发数据源需要配置成不同的参数值,但Dataphin也支持配置 成相同的参数值。

参数

描述

参数	描述
Endpoint	Rocket MQ实例的Endpoint。您可以在消息队列Rocket MQ版控制台的实例详 情页面,查看Endpoint。 ##D# # complete # complet
Instance ID	RocketMQ实例的ID。您可以在消息队列RocketMQ版控制台的实例详情页面, 查看实例ID。
Access ID、Access Key	当前账号的AccessKey ID和AccessKey Secret。如何获取,请参见 <mark>获取</mark> AccessKey。

↓ 注意 创建Rocket MQ数据源时不支持测试数据源是否可以和Dataphin进行正常的连通,
 因此需要您确保连接信息的正确性。

5. 单击确定,完成Rocket MQ数据源的创建。

后续步骤

完成Rocket MQ数据源的创建后,即可基于Rocket MQ的业务数据开发流批一体任务。具体操作,请参见创建 Flink SQL任务、步骤一:创建并调试Flink Template SQL计算任务、创建Flink DataStream计算任务。

6.8. 关系型数据源

6.8.1. 创建MySQL数据源

创建MySQL数据源用于实现Dataphin能够读取MySQL的业务数据,及能够向MySQL写入数据。在引入MySQL 的业务数据至Dataphin和将Dataphin的数据写入至MySQL的场景中,您需要先创建MySQL数据源。本文为您 介绍如何创建MySQL类型的数据源。

MySQL数据源类型介绍

Dataphin支持添加的MySQL数据源类型包括公网数据库、阿里云数据库和ECS(VPC)自建数据源。根据数据源的类型,选择对应的网络连通操作。



数据源类型	网络连通说明	网络连通操作指导
阿里云数据 库	阿里云数据库即阿里云数据库RDS。创建数据源 过程中,需要根据业务情况选择访问方式(公网 IP和VPC网络)后,再配置网络连通参数。	
公网数据库	即具备公网访问能力的数据库,访问方式默认为 公网IP,直接添加Dataphin IP地址至数据库白名 单中即可。	 步骤一:添加Dataphin IP地址至数据库白 名单 步骤二:创建MySQL数据源
ECS(VPC) 自建数据源	即在阿里云ECS上自建的数据库且访问方式为 VPC网络,访问方式默认为VPC网络,直接配置 网络连通参数。	

↓ 注意 如果业务MySQL数据源非上述所述数据源类型,则可以参考网络连通解决方案选择对应的连通操作。网络连通解决方案,请参见网络连通解决方案。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

- 仅支持创建5.1.43、5.6/5.7、8.0版本的MySQL数据源。
- 仅支持创建与Dataphin实例在同一地域的ECS(VPC)自建数据源和阿里云数据库(VPC网络)类型的数据 源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。如何添加白名单, 请参见设置白名单。

• 公网数据库、阿里云数据库且访问方式为公网访问,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

• ECS(VPC)自建数据源、阿里云数据库且访问方式为VPC网络,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建MySQL数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規划	9								ి 🕗 🕓
业务	数据源 (190) ⊙						Q. 请输入继续关键字	V	+ 8622303532
 数据规构 W 公共定义 	guzhi_oss 전 첫편: OSS			Ē	lfs_lindorm ⊵ ≋⊞: Lindorm				Ē
項目	生) ^{de} for	Prod / Basic ①	开发	for Dev O	生产	for Prod / Basic ①	开发	for	Dev ①
▶ 项目管理	交更人: Name in the line in the second second				SEA: Interferencing and gamma inc		SBA: how we do not have a strategic out of		
41 it 11200	政策原用途: 高橋 创建用户: 「一個個個個人」		2月23月2: Add 2月2月1: Magnidentificantine (Johnson)		数据源用途: 减线 创建用户:		取得原用: 本1 (1注用户:		
a 🕦	金融時间:2022-03-09 20:27:15 更新行前:2022-03-11 12:25:27		台通知时间: 2022-03-09 20:30:32 更新时间: 2022-03-11 12:25:33		创建时间: 2022-03-09 11:03:47 更新时间: 2022-03-09 11:03:47		出題时间: 2022-03-09 11:03:47 運動時計局: 2022-03-09 11:03:47		
	· ● 查看任扶信息		● 查看任法信息		(1)※注意: Φ 宣告性失信息		· · · · · · · · · · · · · · · · · · ·		
	操作:	e 1 8 2	提作:	@ ¤ ¤ A	操作:	6 E B A	課/F:	0 2	10 A

3. 在新建数据源对话框的关系型数据库区域,选择MySQL。

如果您最近使用过MySQL,也可以在**最近使用**区域选择MySQL。同时,您也可以在快搜索框中,输入 MySQL的关键词,快速筛选。

新建数据源			Q 请输入	搜索关键字 全部 V X
关系型数据库		ne administration administration	ar Administra anna Administra anna Administra	
S	&	E.		(F)
PolarDB	PolarDB-X(原DRDS)	MySQL	Microsoft SQL Server	PostgreSQL
				关闭。202

- 4. 在新建MySQL数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建MySQL数据源			X	
 * 数据源名称 请输入数据源名称 * 版本 MySQL 5.x 				
数据源描述 请输入数据源描述			0(12)	
* 数据源配置 💿 "生产+开发"数据源 🤇	〕"生产"数据源		0 10 <i>j</i>	
参数	描述			
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。			
版本	 ● 长度不能超过64字符。 支持选择5.1.43、5.6/5.7和8.0版本。 以阿里云MySQL数据库为例,为您介绍如何查看MySQL的版本号。 云数据库RDS / 实例列表 / 基本信息 (~ 运行中) ~ 基本信息 使置白名单 迁移可用区 实例ID 账号管理 性能趋势 慢SQL SQL洞察和审计 EPU 			
数据源描述	对MySQL数据源的简单描述。			
数据源配置	基于业务数据源是否区分生产数据源和开发数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。 			

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。生产数据源和开发数据源需要配置的参数相同。

生产数据源		开	发数据源			
* JDBC URL	请输入JDBC连接地址,如jdbc:mysql://host:port/dbname		* JDBC URL	请输入JDBC连接地址,如jdbc:mys	sql://host:port/dbname	
* 用户名	quick		* 用户名	请输入鉴权用户名,为保证任务正常	常执行,请确保有所需数据权限	
* 密码		₿	* 密码	请输入鉴权密码,为保证任务正常	执行,请确保有所需数据	8
* 类型	○ 公网数据库 ● 阿里云数据库 ○ ECS(VPC)自建数据源		* 类型	🔷 公网数据库 🔷 阿里云数据库	ecs(VPC)自建数据源	
* 访问方式	● VPC网路 ○ 公网IP		* 区域	请选择Region		
* 区域	请选择Region		* VPC ID	请输入VPC ID		
* VPC ID	请输入VPC ID		* ECS ID	请输入ECS ID		
* RDS ID	请输入RDS ID		提示: 当前使用的数	数据库为VPC网络环境下阿里云数据图 6:100 104 115 192/26	车,连接请指定授权IP白名单:	
提示:当前使用的	的数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: /26:100 104 115 192/26		100.104.220.120/2	0,100.104.110.102120		
04						
				测试连接	取消 帝定	

⑦ 说明 通常,生产数据源和开发数据源配置的连接参数不同,但Dataphin也支持配置成相同的连接参数。

参数	描述
JDBC URL	 链接地址的格式为 jdbc:mysql://{链接地址}:{端口}/{数据库名称} 。 以阿里云MySQL数据库为例,为您介绍如何获取链接地址、端口和数据库名称: 在数据库连接页面,获取MySQL实例的链接地址、端口。 「MySQL (() () () () () () () () ()
用户名和密码	当前账号的AccessKey ID和AccessKey Secret。 如何获取,请参见 <mark>获取AccessKey</mark> 。

参数	描述
	 Dataphin支持选择以下几种数据源类型: 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 阿里云数据库:如果您的数据源是阿里云数据库RDS,请选择库阿里云数据。阿里云数据库支持通过公网IP或VPC网络连接Dataphin。
类型	⑦ 说明 如果您的数据源是阿里云数据库RDS且支持公网访问,则您可以选择 公网数据库,也可以选择阿里云数据库(公网IP)。
	ECS(VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的网络环境为VPC,请选择ECS(VPC)自建数据源。

参数	描述
	当类型选择为ECS(VPC)自建数据源和阿里云数据库(VPC网络),还需配置区域、VPC ID、RDS ID/ECS ID这些网络参数,其中: ■ 阿里云数据库(VPC网络):配置如下参数。 ■ 区域:仅支持选择当前Dataphin实例所在的地域。
	⑦ 说明 仅支持添加与Dataphin实例在同一地域的MySQL数据源至 Dataphin。
	 VPC ID:填写阿里云数据库实例的VPC ID。 RDS ID:填写阿里云数据库实例的ID。
	 ECS (VPC) 自建数据源:配置如下参数。 区域:仅支持选择当前Dataphin实例所在的地域。
	⑦ 说明 仅支持添加与Dataphin实例在同一地域的MySQL数据源至 Dataphin。
区域、VPC ID、RDS	
区域、VPC ID、RDS	 VPC ID:填写云服务器ECS实例的VPC ID。 ECS ID:填写云服务器ECS实例的ID。 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看VPC ID、RDS ID、ECS ID,如下图所示。
区域、VPC ID、RDS ID/ECS ID	 VPC ID:填写云服务器ECS实例的VPC ID。 ECS ID:填写云服务器ECS实例的ID。 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看VPC ID、RDS ID、ECS ID,如下图所示。
区域、VPC ID、RDS ID/ECS ID	 VPC ID: 填写云服务器ECS实例的VPC ID。 ECS ID: 填写云服务器ECS实例的ID。 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看VPC ID、RDS ID、ECS ID,如下图所示。 阿里云数据库的网络信息 文例列表 英例的《条体 可能认为深 Q 可选择标签 文例的《条体 可能认为深 Q 可选择标签 文例的《条体 可能认为深 Q 可选择标签 文明的《条体 正行状态 Q 创建时间 数据库类型 Q 付表类型 Q 网络类型 Q 和学校客
区域、VPC ID、RDS ID/ECS ID	 ● VPC ID: 填写云服务器ECS实例的VPC ID。 ● ECS ID: 填写云服务器ECS实例的ID。 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看VPC ID、RDS ID、ECS ID,如下图所示。 <i>阿里云数据库的网络信息</i> <u> 「安例列表</u> <u> 「密信息 で使例列表 </u> <u> 「新知道時間でででで、 「新知道時間でで、 「新知道時間で、 「新知道時間で、 「 「 「 </u>
区域、VPC ID、RDS ID/ECS ID	 ● VPC ID: 填写云服务器ECS实例的VPC ID。 ● ECS ID: 填写云服务器ECS实例的ID。 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看VPC ID、RDS ID、ECS ID,如下图所示。 阿里云数据库的网络信息 文例列表 ● 建筑塔 ● 建築 ● 建
区域、VPC ID、RDS ID/ECS ID	 ● VPC ID: 填写云服务器ECS实例的VPC ID。 ● ECS ID: 填写云服务器ECS实例的ID。 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看 VPC ID、RDS ID、ECS ID,如下图所示。 P理云数据库的网络信息 文例列表 ● 医在图● 電性膨低 ● 電話 ● 電話
区域、VPC ID、RDS ID/ECS ID	 PPC ID: 填写云服务器ECS实例的VPC ID. ECS ID: 填写云服务器ECS实例的ID. 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看VPC ID、RDS ID、ECS ID,如下图所示。 P里云数据库的网络信息 文例列表 ··································
区域、VPC ID、RDS ID/ECS ID	 PCC ID: 填写云服务器ECS实例的VPC ID。 ECS ID: 填写云服务器ECS实例的ID。 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看 VPC ID、RDS ID、ECS ID,如下图所示。 P里云数据库的网络信息 「「「日本教集体的网络信息」 「「日本教集体的网络信息」 「「日本教集体的网络信息」 「「日本教集体的网络信息」 「「日本教集体的网络信息」 「「日本教集体的网络信息」 「「日本教集体的网络信息」 「「日本教集体的内络信息」 「日本教集体的内络信息」 「日本教集体的内络信息」 「日本教集体的内络信息」 「日本教集体的内络信息」 「日本教集体的内络信息」 「日本教集体的内络信息」 「日本教集体的内络信息」 「日本教集体的内路信息」 「日本教集集集集集集集集集集集集集集集集集集集集集集集集集集集集集集集集集集集集
区域、VPC ID、RDS ID/ECS ID	 PPC ID: 填写云服务器ECS实例的VPC ID。 ECS ID: 填写云服务器ECS实例的ID。 您可以在云数据库RDS控制台的实例列表、在云服务器ECS控制台的实例详情页面查看VPC ID、RDS ID、ECS ID,如下图所示。 PE云数据库的网络信息 pE云数据库的网络信息 「使用云数据库的网络信息 「安/例列表 「使用云数据库的网络信息 「使用云数据库的网络信息 「安/例列表 「使用云数据库的网络信息 「日本の日本 「日本日本 「日本日

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。
 如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通解决方案。

6. 测试成功后,单击**确定**,完成MySQL数据源的创建。

后续步骤

完成MySQL数据源的创建后,即可将MySQL的业务数据引入至Dataphin,或将Dataphin的数据写入至MySQL数据源。具体操作,请参见配置离线单条管道。

6.8.2. 创建Oracle数据源

创建Oracle数据源用于实现Dataphin能够读取Oracle的业务数据,及能够向Oracle写入数据。在引入Oracle的业务数据至Dataphin和将Dataphin的数据写入至Oracle的场景中,需要先创建Oracle数据源。本文为您介绍如何创建Oracle数据源。

背景信息

Oracle是一款可移植好、可靠性好、适应高、功能强大的关系型数据库。如果您使用的是Oracle,在对接 Dataphin进行数据开发时,您需要先完成Oracle数据源的创建。更多Oracle信息,请参见Oracle官网。

Dat aphin支持添加的Oracle数据源类型包括公网数据库和ECS(VPC)自建数据源。不同类型的网络连通方 式下,所需的配置细节不一致,如下图所示。



为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

• 当数据源类型为ECS(VPC)自建数据源时,则仅支持创建与Dataphin实例在同一地域Oracle数据源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。

• 公网数据库,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

• ECS(VPC)自建数据源,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。如何添加白名单请参见添加安全组规则。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)	100.104.0.0/16	100.104.228.128/26、 100.104.115.192/26
华南1(深圳)		100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建Oracle数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規划	1							凿 🧔	ి 🕘
<u>#</u> \$	数据源 (190) ①						Q 清箱入提索关键字	V	+ 8/22203020
A 数据编约	guzhi_oss ⊠				Ifs_lindorm 🖂				
₩ 公共定义	<u>満型</u> : OSS			6	<u>時間</u> : Lindorm				
項目	生/*	for Prod / Basic ①	开发	for Dev ①	生phe	for Prod / Basic ①	开发	for	Dev ①
◎ 项目管理							SEA: Topical Transformed by		
411 it 1120	数据源用途: 淘试 创建用户: 同时间 1000 1000 1000 1000		政國運用途: 政治 公理用 ^A : Man afalana (Anno 1996) (Anno 1996)		の現象用途: 増減 の使用や: 予以時にはなるためのないない		数据原用途: 素説 創建用 ^の : 素 読		
a 🚹	(注意を注意): 2022-03-09 20:27:15		()(楚时间): 2022-03-09 20:30:32		(2년8년)(3): 2022-03-09 11:03:47		台(建約1回): 2022-03-09 11:03:47		
	· · · · · · · · · · · · · · · · · · ·		 (1)(1)(1)(2)(2)(2)(3)(1)(2)(2)(3)(3) (1)(2)(2)(2)(3)(1)(2)(2)(3)(3) (1)(2)(2)(2)(3)(1)(2)(2)(3)(3) (1)(2)(2)(2)(3)(1)(2)(2)(3)(3) (1)(2)(2)(2)(2)(3)(1)(2)(2)(3) (1)(2)(2)(2)(2)(2)(3)(1)(2)(2)(3) (1)(2)(2)(2)(2)(2)(2)(2)(2)(2)(2)(2)(2)(2)		1997年1月:2022-03-09 11:03:47 経済信意: ● 宣香経接信息		1996年3月1日、2022-03-09 11:03:47 1919年春日 1月1日 - 2022-03-09 11:03:47		
	操作:	e e i 2	操作:	6 6 8 A	操作:	C E E A	操作:	e 🛛	₫ <u>A</u>

3. 在新建数据源对话框的关系型数据库区域,选择Oracle。

如果您最近使用过Oracle,也可以在**最近使用**区域选择Oracle。同时,您也可以在快搜索框中,输入 Oracle的关键词,快速筛选。

数据源 系型数据库			Q 请输入	捜索关键字 全部 ∨
\mathfrak{G}	8	ί.		(J)
PolarDB	PolarDB-X(原DRDS)	MySQL	Microsoft SQL Server	PostgreSQL
\diamond	\diamond	\odot		ORACLE [.]
AnalyticDB for MySQL 2.0	AnalyticDB for MySQL 3.0	AnalyticDB for PostgreSQL	OceanBase	Oracle

- 4. 在新建Oracle数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建Oracle数据源	X
*数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 /
* 数据源配置) "生产+开发"数据源	"生产"数据源
参数	描述
数据源名称	命名规则如下: 只能包含字母、数字或下划线(_)。 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据》	源		开发数	数据源			
*	JDBC URL	jdbc:oracle:thin:@::1521/orcl		* JDBC URL	jdbc:oracle:thin:@//	1521/orcl	
	Schema	DATAPHIN	0	Schema	SH		0
	* 用户名	dataphin		* 用户名	system		
	* 密码		8	* 密码			₿atap
	* 类型	○ 公网数据库 ● ECS(VPC)自建数据源		* 类型(● 公网数据库 ○ ECS	8(VPC)自建数据源	
	* 区域	华东2(上海)	~ 4	提示: 本地自建数排	居库授权, 需要开放访问, 并	并授权权限	
	* VPC ID	vpc-bp17l					
	* ECS ID	i-uf60sc					
提示 100.1	: 当前使用的 104.228.128/2	数据库为VPC网络环境下阿里云数据库,连接请指 26;100.104.115.192/26	定授权IP白名单:				

⑦ 说明 通常,生产数据源和开发数据源需要配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述
JDBC URL	链接地址的格式为 jdbc:oracle:thin:@//{ 服务器地址 }:{端口}/{数据库名称} 。 例如, jdbc:oracle:thin:@//192.0.2.0:1521/orcl 。
Schema	即Oracle实例下的数据库名称。
用户名、密码	登录Oracle实例的用户名和密码。
类型	 Dataphin支持选择以下几种数据源类型: 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 ECS(VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的网络环境为VPC,请选择ECS(VPC)自建数据源。
	 当类型选择为ECS (VPC) 自建数据源,则需要配置如下参数: 区域:仅支持选择当前Dataphin实例所在的地域。 ⑦ 说明 仅支持添加与Dataphin实例在同一地域的Oracle数据源至 Dataphin。
区域、VPC ID、ECS ID	 VPC ID: 填写Oracle实例的VPC ID。 ECS ID: 填写Oracle实例的ID。
	主私网IP 10.0 弹性IP实例ID -

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。 6. 测试成功后,单击**确定**,完成Oracle数据源的创建。

后续步骤

完成Oracle数据源的创建后,即可将Oracle的业务数据引入至Dat aphin,或将Dat aphin的数据写入至Oracle数据源。具体操作,请参见配置离线单条管道。

6.8.3. 创建PostgreSQL数据源

创建PostgreSQL数据源用于实现Dataphin能够读取PostgreSQL的业务数据,及能够向PostgreSQL写入数据。在引入PostgreSQL的业务数据至Dataphin和将Dataphin的数据写入至PostgreSQL的场景中,您需要先创建PostgreSQL数据源。本文为您介绍如何创建PostgreSQL类型的数据源。

PostgreSQL数据源类型介绍

Dataphin支持添加的PostgreSQL数据源类型包括公网数据库、阿里云数据库和ECS(VPC)自建数据源。不同类型数据源的网络连通,所需的配置细节不一致,如下图所示。



为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

- 仅支持超级管理员和项目管理员角色创建数据源。
 如何添加项目成员并授予项目管理员角色,请参见添加项目成员。
- 仅支持创建与Dataphin实例在同一地域的ECS(VPC)自建数据源和阿里云数据库(VPC网络)类型的数据 源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。以阿里云RDS PostgreSQL为例为您介绍如何添加白名单,详情请参见设置白名单。

• 公网数据库、阿里云数据库且访问方式为公网访问,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

• ECS(VPC)自建数据源、阿里云数据库且访问方式为VPC网络,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)	100.104.0.0/16	100.104.228.128/26、 100.104.115.192/26
华南1(深圳)		100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建PostgreSQL数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

≡	Dataphin 规划									l ⁰ 2 🕓
<u>ste</u> st	5	数据源 (190) ①						Q 请输入虚素关键字	V	+ 8/22/02/22
∴ w	数据编构 公共定义	guzhi_oss ☑ 类型: OSS			ē	lfs_lindorm ⊠ 問型: Lindorm				Ť
項目	1	生Per	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①
ы	项目管理	意思人: Logentation Disamble - Dis								
411	计算题	数据源用途: 建试 创建用户:		数据源用途: 集线 创建用户: Manufalenting inter Briter		数据原用途: 集线 创建用户: "如果"的时候,如果"的时候,		政道原用注: 建试 创建用户:		
13 3	1 8:55	 (出陸町间: 2022-03-09 20:27:15 (開駅町间): 2022-03-11 12:25:27 (初設信息: ● 宣都相接信息 		会通知1時に、2022-03-09 20:30:32 通知行時間、2022-03-11 12:25:33 長期日間に、		 出陸町川に 2022-03-09 11:03:47 更新町川に 2022-03-09 11:03:47 転転日期: ● 宣委相关信息 		台港町河: 2022-03-09 11:03:47 運転町河: 2022-03-09 11:03:47 転転町河: 2022-03-09 11:03:47 転転回道: ① 宣告相決省景		
		2017年:	0 C E &	· · · · · · · · · · · · · · · · · · ·	0012	2限/作:	6 6 5 A	課/F:	e B	ē A

3. 在新建数据源对话框的关系型数据库区域,选择PostgreSQL。

如果您最近使用过PostgreSQL,也可以在**最近使用**区域选择PostgreSQL。同时,您也可以在快搜索框中,输入PostgreSQL的关键词,快速筛选。

新建数据源			Q 请输)	★ とお マン ス とお マン ス とお マン ス とお マン ス と からします。
でで	&	53		(P)
PolarDB	PolarDB-X(原DRDS)	MySQL	Microsoft SQL Server	PostgreSQL
\diamond	\diamond	0		ORACLE [.]
AnalyticDB for MySQL 2.0	AnalyticDB for MySQL 3.0	AnalyticDB for PostgreSQL	OceanBase	Oracle
				关闭 (1500)

- 4. 在新建PostgreSQL数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建PostgreSQL数据	源 ×
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据源配置 🕑 "生产+开发"数据源	"生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

* JDBC URL jdbc: * Schema publi * 用户名 datag	ipostgresql://rm-10.pg.rds.aliyuncs.com: ic phin	34	* JDBC URL * Schema	jdbc:postgresql://m	1o.pg.rds.aliyuncs.com	:34
* JDBC URL jdbc: * Schema publi * 用户名 datag	postgresql://rm- 1o.pg.rds.aliyuncs.com: ic phin	34	* JDBC URL * Schema	jdbc:postgresql://m	1o.pg.rds.aliyuncs.com	34
* Schema publi * 用户名 datag	ic phin	()	* Schema	public		0
* 用户名 data	phin					0
			* 用户名	dataphin		
* 密码 ······		₿	* 密码			₿
★ 类型 ○ 公	网数据库 💿 阿里云数据库 🔷 ECS(VPC)自建数据源		* 类型	🔵 公网数据库 • 阿里云数据库	O ECS(VPC)自建数据源	i datap
* 访问方式 🧿 VF	PC网路 〇 公网IP		* 访问方式	● VPC网路 ○ 公网IP		
* 区域 华东	2(上海)		* 区域	华东2(上海)		
* VPC ID vpc-u	uf6d		* VPC ID	vpc-uf6c		
* RDS ID rm-u	f6i0		* RDS ID	rm-uf6i0		
提示:当前使用的数据库; 100.104.228.128/26;100.	为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 104.115.192/26		提示: 当前使用的 100.104.228.128/2	数据库为VPC网络环境下阿里云数据库 26;100.104.115.192/26	i,连接请指定授权IP白名单:	
					7	

⑦ 说明 通常,生产数据源和开发数据源配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述		
参数 JDBC URL	描述 链接地址的格式为 jdbc:postgresql://(链接地址):(端口)/(数据库名称) 。 UP里云 RDS PostgreSQL数据库为例,您可以在数据库连接页面、数据库管理页面获取 。 数据库方的网络信息 数据库连接 切换为经典网络 修改连接地址 为什么连接不上 网络类型 专有网络(VPC:vpc-shangl ② 内咽地址 rm-uf6le0k430ahz; 小网端口 3433 小网端口 3433 大网端口 3433 (公 运行中) ✓		
	基本信息 创建数据库 导入数据库 账号管理 数据库名称 数据库状态 数据库管理 dataphin ✓ 运行中		

参数	描述	
Schema	Schema即PostgreSQL实例的数据库名称。在PostgreSQL实例的数据管理页面获取数据 库名称,如下图所示。	
用户名和密码	当前账号的AccessKey ID和AccessKey Secret。 如何获取,请参见 <mark>获取AccessKey</mark> 。	
类型	Dataphin支持选择以下几种数据源类型: 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 阿里云数据库:如果您的数据源是阿里云数据库RDS,请选择库阿里云数据。阿里云数据库支持通过公网IP或VPC网络连接Dataphin。 ②说明 如果您的数据源是阿里云数据库RDS且支持公网访问,则您可以选择公网数据库,也可以选择阿里云数据库(公网IP)。 ECS (VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的网络环境为VPC,请选择ECS (VPC)自建数据源。	
	当类型选择为ECS(VPC)自建数据源和阿里云数据库(VPC网络),还需配置区域、VPC ID、RDS ID/ECS ID这些网络参数,其中:	

参数	■ 阿里云数据库 (VPC网络):配置如下参数。 描述 ■ 区域 :仅支持选择当前Dataphin实例所在的地域。			
	⑦ 说明 仅支持添加与Dataphin实例在同一地域的PostgreSQL数据源至 Dataphin。			
	 VPC ID:填写阿里云RDS PostgreSQL实例的VPC ID。 RDS ID:填写阿里云RDS PostgreSQL实例的ID。 ECS (VPC)自建数据源:配置如下参数。 区域:仅支持选择当前Dataphin实例所在的地域。 			
	⑦ 说明 仅支持添加与Dataphin实例在同一地域的PostgreSQL数据源至 Dataphin。			
ID, RDS ID/ECS ID	 VPC ID:填写云服务器ECS实例的VPC ID。 ECS ID:填写云服务器ECS实例的ID。 			
您可以在PostgreSQL实例的控制台页面、ECS实例的控制台页面查看VPC ID、R ECS ID,如下图所示。				
	创建实例 实例ID/名称 请输入P □ 实例ID/名称 数据库类型 ♀ 网络类型 ♀			
	pgm-uf67 PostgreSQL 13.0 专有网络 vpc-uf6ar			
ECS自建数据库的网络信息 ^{宝服务器 ECS / 案例 / 案例54例} ← launch-a				
	実務評構 盗控 安全组 云盘 实例快照 快照 弹性网卡 远程命令/文件 操作记录 健康诊断 事件 基本信息 <td< td=""></td<>			
	launch-c ∠ 已停止 実例D i-uf6at 近空注册			
	公河中・ 講題時世中 所在可用区 上等可用区F 网络信息 <t< td=""></t<>			

5. 单击**测试连接**,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。 6.测试成功后,单击**确定**,完成PostgreSQL数据源的创建。

后续步骤

完成PostgreSQL数据源的创建后,即可将PostgreSQL的业务数据引入至Dataphin,或将Dataphin的数据写入至PostgreSQL数据源。具体操作,请参见配置离线单条管道。

vsw-uf6 je

6.8.4. 创建Microsoft SQL Server数据源

创建Microsoft SQL Server数据源用于实现Dataphin能够读取Microsoft SQL Server的业务数据,及能够向 Microsoft SQL Server写入数据。在引入Microsoft SQL Server的业务数据至Dataphin和将Dataphin的数据写 入至Microsoft SQL Server的场景中,您需要先创建Microsoft SQL Server数据源。本文为您介绍如何创建 Microsoft SQL Server类型的数据源。

Microsoft SQL Server数据源类型介绍

Dataphin支持添加的Microsoft SQL Server数据源类型包括公网数据库、阿里云数据库和ECS(VPC)自建数据源。根据数据源的类型,选择对应的网络连通操作。



数据源类型	数据源类型和网络连通说明	网络连通操作指导
阿里云数据 库	即阿里云RDS SQL Server。创建数据源过程中, 需要根据业务情况选择访问方式(公网IP和VPC 网络)后,再配置网络连通参数。	 步骤一:添加Dataphin IP地址至数据库白 名单 步骤二:创建Microsoft SQL Server数据源
公网数据库	即具备公网访问能力的数据库,访问方式默认为 公网IP。直接添加Dataphin IP地址至数据库白名 单中即可。	
ECS(VPC) 自建数据源	即在阿里云ECS上自建的数据库且访问方式为 VPC网络。直接配置网络连通参数即可。	

◯ 注意 如果业务Microsoft SQL Server数据源非上述所述数据源类型,则可以参考网络连通解决方案 选择对应的连通操作。网络连通解决方案,请参见网络连通解决方案。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。
● 仅支持创建与Dataphin实例在同一地域的ECS(VPC)自建数据源和阿里云数据库(VPC网络)类型的数据 源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。

公网数据库、阿里云数据库且访问方式为公网访问,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● ECS(VPC)自建数据源、阿里云数据库且访问方式为VPC网络,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。如何添加白名单请参见添加安全组规则。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)	100.104.0.0/16	100.104.228.128/26、 100.104.115.192/26
华南1(深圳)		100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建Microsoft SQL Server数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規划	1							街 🗘	ి 📀 💽
业务	数据源 (190) ①						Q 遗输入理索关键字	T	+ 6/22/02/22
奇 数据转向	guzhi_oss ⊠			Ē	lfs_lindorm ⊠				÷
₩ 公共定义	Setting of the setting of		70		RE: Lindorm		7742		
項目	生/m 107	Prod / Dasso U	77.20	for Liev U	±r*	for Proc / basic U	<i>π</i> .2	tor	
● 計算源	交通人: 数据原用途: 案例 会建用户:		交进人 新闻		交更人: 数据源用途: 案號 公課用户:		均便人: 数据原用途: 南线 公读用户:		
	(1)開き1時: 2022-03-09 20:27:15 開発時間時: 2022-03-11 12:25:27 4300(1)日: ◆ 11:05(日中位用)		(1)思行(1): 2022-03-09 20:30:32 現代行(1): 2022-03-11 12:25:33 (1):001(1): 2022-03-11 12:25:33		台建5月前: 2022-03-09 11:03:47 現新時日前: 2022-03-09 11:03:47 時時日前: 2022-03-09 11:03:47		(1)語り方向: 2022-03-09 11:03:47 開発の方向: 2022-03-09 11:03:47 新ための目: ゆ 世話的には広見		
	20/2 · · · · · · · · · · · · · · · · · · ·	0 6 6 8		0 D B A	· · · · · · · · · · · · · · · · · · ·	0 1	提/F:	e 2	ē <u>8</u>

3. 在新建数据源对话框的关系型数据库区域,选择Microsoft SQL Server。

如果您最近使用过Microsoft SQL Server,也可以在最近使用区域选择Microsoft SQL Server。同时,您 也可以在快搜索框中,输入Microsoft SQL Server的关键词,快速筛选。

新建数据源			Q 请输入注	聖素关键字 全部 ∨ X
\heartsuit	&	E.		(F)
PolarDB	PolarDB-X(原DRDS)	MySQL	Microsoft SQL Server	PostgreSQL
\diamond	\diamond	0		ORACLE
				关闭

- 4. 在新建Microsoft SQL Server数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建Micro	osoft SQL Server数据源 X
*数据源名称 请输入数据》	原名称
数据源描述 请输入数据》	原描述 0/128 /
* 数据源配置 🔘 "生产+开发	发数据源 ○ "生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: ■ 如果业务数据源区分生产数据源和开发数据源,则选择 生产+开发数据源。 ■ 如果业务数据源不区分生产数据源和开发数据源,则选择 生产数据源 。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源			
* JDBC URL	jdbc:sqlserver://4 :1433;DatabaseName=master		* JDBC URL	jdbc:sqlserver://4 1433;)atabaseName=master	
* Schema	dbo	() * Schema	dbo		()
* 用户名	dataphin		* 用户名	dataphin		
* 密码		₿	* 密码		8	
* 美型	○ 公网数据库 ● 阿里云数据库 ○ ECS(VPC)自建数据源		* 类型	○ 公网数据库 (● 阿里云数据库 (ECS(VPC)自建数据源	
* 访问方式	● VPC网路 ○ 公网IP		* 访问方式	● VPC网路 ○ 公网IP		
* 区域	华东2(上海)		* 区域	华东2(上海)		
* VPC ID	vpc-uf6d4d2		* VPC ID	vpc-uf6d4		
* RDS ID	rm-uf6n3gc		* RDS ID	rm-uf6n3		
提示: 当前使用的 100.104.228.128/	数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 26;100.104.115.192/26		提示: 当前使用的 100.104.228.128/	数据库为VPC网络环境下阿里云数据库, 26;100.104.115.192/26	车接请指定授权IP白名单:	
					qar	av
				测试连接	取消 确定	

⑦ 说明 通常,生产数据源和开发数据源配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述	
JDBC URL	链接地址的格式为 jdbc:sqlserver://{链接地址}:{端口}/{数据库名称} 。 以阿里云Microsoft SQL Server数据库为例,为您介绍如何获取链接地址、端口和数据库名称。 数据库名称 <li< th=""><th>s</th></li<>	s
用户名和密码	当前账号的AccessKey ID和AccessKey Secret。 如何获取,请参见 <mark>获取AccessKey</mark> 。	

参数	描述
类型	 Dataphin支持选择以下几种数据源类型: 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 阿里云数据库:如果您的数据源是阿里云数据库RDS,请选择库阿里云数据。阿里云数据库支持通过公网IP或VPC网络连接Dataphin。
	⑦ 说明 如果您的数据源是阿里云数据库RDS且支持公网访问,则您可以选择 公网数据库,也可以选择阿里云数据库(公网IP)。
	ECS(VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的网络环境为VPC,请选择ECS(VPC)自建数据源。

参数	描述						
	当类型选择为ECS (VPC) 自建数据源和阿里云数据库(VPC网络),还需配置区域、VPC ID、RDS ID/ECS ID这些网络参数,其中: ■ 阿里云数据库(VPC网络):配置如下参数。 ■ 区域:仅支持选择当前Dataphin实例所在的地域。						
	② 说明 仅支持添加与Dataphin实例在同一地域的Microsoft SQL Server数 据源至Dataphin。						
	 VPC ID:填写阿里云数据库实例的VPC ID。 RDS ID:填写阿里云数据库实例的ID。 						
	 ECS(VPC)自建数据源:配置如下参数。 区域:仅支持选择当前Dataphin实例所在的地域。 						
	⑦ 说明 仅支持添加与Dataphin实例在同一地域的Microsoft SQL Server数 据源至Dataphin。						
区域、VPC ID、RDS	 VPC ID:填写云服务器ECS实例的VPC ID。 ECS ID:填写云服务器ECS实例的ID。 您可以在对应数据库实例的控制台页面、ECS实例的控制台页面查看VPC ID、RDS ID、ECS 						
ID/ECS ID	ID, 如下图所示。 阿里云数据库的网络信息						
	实例ID/名称 数据库类型 ♀ 网络类型 ♀						
	rm-uf61u5(SQL Server 2017 企业集群版 ち有网络 vpc-uf6arcd のののののでのである。						
	ECS自建数据库的网络信息 						
	突然洋橋 盆腔 安全坦 云盘 实例块探 快探 弹性网卡 远程命令/文件 摄作记录 健康诊断 事件						
	基本信息 诊断健康状态 ☎ 1 launch-a 🖉 🕒 已停止						
	实例D i-uf6a9; 远程连接 地域 採东2 (上海) 公附P - 姚冠臻性P 所在可用区 上海可用区F						
	网络信息						
	网络说型 专有网络 专有网络 vpc-uf60mbqr 弹性网卡 eni-uf628E 虚划交换机 vsw-uf637jqto						

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见<mark>网络连通的常见问题</mark>。

6. 测试成功后,单击确定,完成Microsoft SQL Server数据源的创建。

后续步骤

完成Microsoft SQL Server数据源的创建后,即可将Microsoft SQL Server的业务数据引入至Dataphin,或将 Dataphin的数据写入至Microsoft SQL Server数据源。具体操作,请参见配置离线单条管道。

6.8.5. 创建AnalyticDB for MySQL数据源

Dataphin支持创建AnalyticDB for MySQL 2.0和AnalyticDB for MySQL 3.0类型的数据源。创建AnalyticDB for MySQL数据源用于实现Dataphin能够读取AnalyticDB for MySQL的业务数据。在引入AnalyticDB for MySQL 的业务数据至Dataphin和将Dataphin的数据写入至AnalyticDB for MySQL的场景中,您需要先创建AnalyticDB for MySQL数据源。本文为您介绍如何创建AnalyticDB for MySQL 2.0和AnalyticDB for MySQL 3.0 类型的数据源。

背景信息

AnalyticDB for MySQL即阿里云的云原生数仓AnalyticDB MySQL,如果您使用的是云原生数仓AnalyticDB MySQL,在对接Dataphin进行数据开发时,您需要先完成AnalyticDB for MySQL数据源的创建。更多云原生数仓AnalyticDB MySQL信息,请参见什么是云原生数据仓库AnalyticDB MySQL版。

AnalyticDB for MySQL数据源类型包括AnalyticDB for MySQL 2.0和AnalyticDB for MySQL 3.0。Dataphin支 持通过公网IP连接AnalyticDB for MySQL 2.0、通过VPC网络或公网IP连接AnalyticDB for MySQL 3.0。不同数 据源类型的网络连通方式,所需的配置细节不一致,如下图所示。



为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中,详情请参见设置白名单。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

• 当业务数据源为AnalyticDB for MySQL 3.0且使用VPC网络时,则仅支持创建与Dataphin实例在同一地域 AnalyticDB for MySQL 3.0数据源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。如何设置白名单请参见设置白名单。

• 访问方式为公网IP时,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● 访问方式为VPC网络时,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。VPC网络下Dataphin 的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)	100.104.0.0/16	100.104.228.128/26、 100.104.115.192/26
华南1(深圳)		100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建AnalyticDB for MySQL数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

Dataphin 1993	1								2 🐸
#¥	数据源(190)①						Q 请输入继续关键字	V	+ arehosee
▲ 数据编约 ₩ 公共定义	guzhi_oss ⊠ ≋⊞: oss			Ē	Ifs_lindorm 四 問題: Lindorm				Ē
項目	生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①
□项目管理 11 计算数 32 1 ■ 数据数	今美人 設備原用法: 素減 総裁用: 2020年30日 単数目前: 202242449 02 27 15 単新日前: 202242-11 12 25 27 総務保護: ● 面気伝統在展				会長人: 数量等用法: 載成 (約第四) (100)(100)(100)(100)(100)(100)(100)(10		交通人: 数据局用途: 本は 必須有用: 必須有用: 必須有用: 2022-03-09 11:03-07 単純協想: ● 面前は分析電音 ● 面前は分析電音		
	续作:	0012	操作:	0012	握作:	C E E &	操作:	e 🛛	8 <u>8</u>

3. 在新建数据源对话框的关系型数据库区域,选择AnalyticDB for MySQL 2.0或AnalyticDB for MySQL 3.0。

如果您最近使用过AnalyticDB for MySQL,也可以在最近使用区域选择AnalyticDB for MySQL2.0或 AnalyticDB for MySQL3.0。同时,您也可以在快搜索框中,输入AnalyticDB for MySQL的关键词,快速 筛选。

奴/店/尿 系型数据库			▲ 喧樂//3	差於大键子 土印 ∨
\mathfrak{G}	&	E.		(P)
PolarDB	PolarDB-X(原DRDS)	MySQL	Microsoft SQL Server	PostgreSQL
\diamond	\diamond	\odot		ORACLE
AnalyticDB for MySQL 2.0	AnalyticDB for MySQL 3.0	AnalyticDB for PostgreSQL	OceanBase	Oracle

- 4. 在**新建AnalyticDB for MySQL 3.0数据源**或**新建AnalyticDB for MySQL 3.0数据源**对话框中,配置 连接数据源参数后,单击**确定**。
 - i. 配置数据源的基本信息。

创建AnalyticDB for MySQL 3.0和AnalyticDB for MySQL 2.0数据源时,配置的数据源基本信息相同,如下图所示。

* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 2
* 数据源配置 • "生产+开发"数据源 (*)	"" 生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源			
* JDBC URL	jdbc:mysql://am-bpads.aliyuncs.com:33	306	* JDBC URL	jdbc:mysql://am-bp	.ads.ali	yuncs.com:3306
* 用户名	dataphin		* 用户名	dataphin		
* 密码		₿	* 密码			₿atap
* 访问方式	● VPC网路 ○ 公网IP		* 访问方式	● VPC网路 ○ 公网IP		
* 区域	华东2(上海)		* 区域	华东2(上海)		
* VPC ID	vpc-bp17h		* VPC ID	vpc-bp17hrc		
* 集群 ID	am-bp1'		* 集群 ID	am-bp111i40		
提示:当前使用的 100.104.228.128/2	数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 26;100.104.115.192/26		提示:当前使用的 100.104.228.128/2	数据库为VPC网络环境下阿里云 26;100.104.115.192/26	数据库, 连接请指定扬	段权IP白名单:
				测试连接	取消	确定

⑦ 说明 通常,生产数据源和开发数据源配置参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述	
	链接地址的格式为 jdbc:mysql//{ 实例地址 }/{ 实例名称 } 。 在数据库管理页面,查看实例地址和实例名称。	
JDBC URL	登录状态: 实例状态: 正常 实例名称: rm-bp1s7v31438 管控模式: 安全协同 来源: 云数据库 类型: MySQL 实例地区: 华东1 (杭州) 实例地址: rm-bp1s7vmysql.rds.aliyuncs.com:3306 环境: 生产	
用户名、密码	当前账号的AccessKey ID和AccessKey Secret。如何获取,请参见 <mark>获取</mark> AccessKey。	
访问方式	根据数据源类型选择对应的访问方式: 数据源类型为AnalyticDB for MySQL 2.0,则仅支持选择公网IP。 数据源类型为AnalyticDB for MySQL 3.0,则支持选择如下网络类型: 数据源所在网络环境为公网,则访问方式选择为公网IP。 数据源所在网络环境为VPC网络,则访问方式选择为VPC网络。 	

参数	描述
参数 区域、VPC ID、集群 ID	描述 当数据源类型为AnalyticDB for MySQL 3.0且当访问方式选择为VPC网路,则需要配置如下参数: • 区域: 仅支持选择当前Dataphin实例所在的地域。 ② 说明 仅支持创建与Dataphin实例在同一地域的AnalyticDB for MySQL 3.0数据源。 • VPC ID: 填写AnalyticDB for MySQL 3.0实例的VPC ID。 • 集群 ID: 填写AnalyticDB for MySQL 3.0实例的VPC ID。 • 集群 ID: 填写AnalyticDB for MySQL 3.0集群ID。 在云原生数据仓库控制台的AnalyticDB for MySQL 3.0集群详情页面,查看集群 ID和VPC ID。 集群属性 集群展型 常规 版本 3.0 集群类型 常规 标签 ② 暂无标签 编辑标签
	VPC地址 am-uf6st).ads.aliyuncs.com 回
	VICTO VPC-urvarce

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。 6. 测试成功后,单击**确定**,完成AnalyticDB for MySQL数据源的创建。

后续步骤

完成AnalyticDB for MySQL数据源的创建后,即可将AnalyticDB for MySQL的业务数据引入至Dataphin或将 Dataphin的数据写入至AnalyticDB for MySQL。具体操作,请参见配置离线单条管道。

6.8.6. 创建PolarDB-X数据源

创建PolarDB-X数据源用于实现Dataphin能够读取PolarDB-X的业务数据,及能够向PolarDB-X写入数据。在引入PolarDB-X的业务数据至Dataphin和将Dataphin的数据写入至PolarDB-X的场景中,需要先创建PolarDB-X数据源。本文为您介绍如何创建PolarDB-X数据源。

背景信息

PolarDB-X(原DRDS升级版)是由阿里巴巴自主研发的云原生分布式数据库。如果您使用的是PolarDB-X, 在对接Dat aphin进行数据开发时,您需要先完成PolarDB-X数据源的创建。更多PolarDB-X信息,请参见产品 概述。 Dat aphin支持通过VPC网络和公网连接PolarDB-X。不同类型的网络连通方式下,所需的配置细节不一致,如下图所示。

	访问方式:公网IP	添加公网下的Dataphin IP地址至数据库白名单
	访问方式:VPC网络	1.添加VPC网络下的Dataphin IP地址至数据库臼名单 2.选择数据原序在的Region 3.填写VPC ID 4.填写DRDS ID
		k.
数据源	→ 访问方式	② 网络连通

为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

- 仅支持超级管理员和项目管理员角色创建数据源。
 如何添加项目成员并授予项目管理员角色,请参见添加项目成员。
- 当业务数据源(PolarDB-X)使用VPC网络时,则仅支持创建与Dataphin实例在同一地域PolarDB-X数据源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。以阿里云数据库为 例为您介绍如何添加白名单,请参见设置白名单。

• 访问方式为公网访问,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● 访问方式为VPC网络,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。VPC网络下Dataphin的IP 地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26

地域	(推荐)通用IP地址	地域IP地址
华南1(深圳)	100.104.0.0/10	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建PolarDB-X数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規划								ෂ ් එ <mark>ල</mark> 🤇
业务	数据源 (190) ①						Q 请输入提供关键字	V + anažavanco
A 数据编约	auzhi ass 🕫				Ifs. lindorm 🕫			
₩ 公共定义	类型: OSS				(問題): Lindorm			
项目	生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for Dev ①
□ 项目管理	交通人: Ingeniates [ingeniates] in August						负责人: mage and the set of the set	
≪ 计算题	数据源用途: 集成 创建用户:		政策運用途: 集成 创建用 ^p : International Constitution		政策原用法: 集成 6(建用A: Humineritum Antophysics		設備原用途: 和3 (12月 ⁷⁾ : 12月 1日 - 12月1日 - 12月11日 - 121100 - 1211000000000000000000000000	
æ 🚹	金融時間: 2022-03-09 20:27:15 単時時間: 2022-03-11 12:25:27		①建時前: 2022-03-09 20:30:32 第年時前: 2022-03-11 12:25:33		出現的計画: 2022-03-09 11:03:47 単和日日間: 2022-03-09 11:03:47		台陸时间: 2022-03-09 11:03:47 原和日前前: 2022-03-09 11:03:47	
STATES	188989: • 161969		(1)没信息: ◆ 查看任法信息		组织信息: ● 宣誓性狭位息		· 经按信息: • 查看经接信息	
	提作:	0 6 8 £	操作:	@ Ø Ø A	操作:	C 12 1 A	操作:	0 B B A
	费作:	0 1	操作:	6 1 B A	操作:	0 D B A	操作:	e e e

3. 在新建数据源对话框的关系型数据库区域,选择PolarDB-X (原DRDS)。

如果您最近使用过PolarDB-X,也可以在**最近使用**区域选择PolarDB-X。同时,您也可以在快搜索框中,输入PolarDB-X的关键词,快速筛选。

新建数据源			Q、 请输入:	搜索关键字 全部 🗸 X
\heartsuit	&	k.		(G)
PolarDB	PolarDB-X(原DRDS)	MySQL	Microsoft SQL Server	PostgreSQL
\diamond	\diamond	0		ORACLE
e dentri Su				关闭

4. 在新建PolarDB-X (原DRDS) 数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建PolarDB-X(原DP	RDS)数据源 X
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 /
* 数据源配置 💿 "生产+开发"数据源	"生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源				
* JDBC URL	请输入JDBC连接地址,如jdbc:mysql://host:port/dbname		* JDBC URL	请输入JDBC连接地均	止,如jdbc:mysql://	host:port/dbname	
* 用户名	请输入鉴权用户名,为保证任务正常执行,请确保有所需数据权限	R	* 用户名	请输入鉴权用户名,	为保证任务正常执	行,请确保有所需	数据权限
* 密码	请输入鉴权密码,为保证任务正常执行,请确保有所需数据	₿	* 密码	请输入鉴权密码,为	保证任务正常执行	,请确保有所需数	据 🔒
* 访问方式	● VPC网路 ○ 公网IP		* 访问方式	● VPC网路 ○ 2	》网IP		
* 区域	请选择Region		* 区域	请选择Region			
* VPC ID	请输入VPC ID		* VPC ID	请输入VPC ID			
* DRDS ID	请输入DRDS ID		* DRDS ID	请输入DRDS ID			
提示: 当前使用的: 100.104.228.128/2	数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 26;100.104.115.192/26		提示:当前使用的 100.104.228.128/2	数据库为VPC网络环境 26;100.104.115.192/26	下阿里云数据库,ì	≤接请指定授权IPI	白名单:
				测试连	渡	取消	确定 datap

⑦ 说明 通常,生产数据源和开发数据源需要配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数 描述 描述

参数	描述	
JDBC URL	链接地址的格式为 jdbc:mysql://{链接地址}:{端口}/{数据库名称 。您可以 在PolarDB-X控制台实例对应数据库的详情页面,查看链接地址、端口和数据库名称 下图示例为VPC网络下的链接信息。 <u>基本值息</u> <u></u>	o 水平拆分 } dataphin 3306
用户名、密码	当前账号的AccessKey ID和AccessKey Secret。如何获取,请参见 <mark>获取AccessKey</mark> 。	
访问方式	选择对应的访问方式: PolarDB-X数据源所在网络环境为VPC网络,则访问方式选择为VPC网络。 PolarDB-X数据源所在网络环境为公网,则访问方式选择为公网IP。 	
区域、VPC ID、DRDS ID	 当访向方式选择为VPC网路,则需要配置如下参数: 区域:仅支持选择当前Dataphin实例所在的地域。 ② 说明 仅支持添加与Dataphin实例在同一地域的PolarDB-X数据源至 Dataphin。 VPC ID:填写PolarDB-X实例的VPC ID。 DRDS ID:填写PolarDB-X实例的ID和VPC ID。 在PolarDB-X控制台的实例详情页面,查看实例ID和VPC ID。 基本信息 实例ID:drdsf 反域:华东2(上海) 状态:运行中 实例类型:主实例 连接信息 阿塔类型:VPC VPC地址:drdsfacbyn8 VPC ID:vpc-uf6 	

5. 单击**测试连接**,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。 6.测试成功后,单击**确定**,完成PolarDB-X数据源的创建。

后续步骤

完成PolarDB-X数据源的创建后,即可将PolarDB-X的业务数据引入至Dataphin,或将Dataphin的数据写入 至PolarDB-X数据源。具体操作,请参见配置离线单条管道。

6.8.7. 创建Vertica数据源

创建Vertica数据源用于实现Dataphin能够读取Vertica的业务数据。在引入Vertica的业务数据至Dataphin和 将Dataphin的数据写入至Vertica的场景中,您需要先创建Vertica数据源。本文为您介绍如何创建Vertica类型的数据源。

背景信息

Vertica是一款基于列存储架构的数据库。如果您使用的是Vertica,在对接Dataphin进行数据开发时,您需要先完成Vertica数据源的创建。更多Vertica信息,请参见Vertica官网。

Dataphin支持通过公网连接Vertica。为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin 在公网环境下的IP地址段添加到数据源的白名单中。

使用限制

仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

步骤一:添加Dataphin IP地址至数据库白名单

添加公网环境下的Dataphin IP地址至数据源白名单中。

公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

步骤二: 配置创建Vertica数据源的参数

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

								🛎 🗳 ୍ <mark>ଚ</mark> 🧕 🧕
#8	数据源 (190) ①						Q 遗输入检索关键字	T + averagence
高数据编约	authi ass 5				Ifs lindom D			
₩ 公共定义	类型: OSS			Ξ	(問題): Lindorm			
項目	生 ^{pla}	for Prod / Basic ①	开发	for Dev ①	生/*	for Prod / Basic ①	开发	for Dev ①
□ 项目管理								
41 it ji a	和描述用述: 发出 2.22用户: 1000000000000000000000000000000000000		記憶用注: 本は ()注用 ⁽⁾ : 本は		数据源用途: 淘试 《建用户: 出版》 (1999年1月1日 - 1999年1月1日)		設備原用途: 素成 ((建用 ^の : また) の時間になっていた。	
ा ()	出題時前時: 2022-03-09 20:27:15 現象時前時: 2022-03-11 12:25:27		 ()通知時間:2022-03-09 20:30:32 ()更新時間間:2022-03-11 12:25:33 		创建时间: 2022-03-09 11:03:47 原行时间: 2022-03-09 11:03:47		(回想時)時間: 2022-03-09 11:03:47 現所時)時間: 2022-03-09 11:03:47	
		@ P1 = A	(1)別信息: ● 豆首10分(1)単 (小)、	@ PI # A	19745: © 18129418	¢ р в А	12日本1日	@ PL B A
	2015-	0001	ante.	0000	ante.	0 1 1 1	1000	0 0 0 0

3. 在新建数据源对话框的关系型数据库区域,选择Vertica。

如果您最近使用过Vertica,也可以在**最近使用**区域选择Vertica。同时,您也可以在快搜索框中,输入 Vertica的关键词,快速筛选。



- 4. 在新建Vertica数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建Vertica数据源	X
*数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 /
*数据源配置 💿 "生产+开发"数据源	○ *生产**数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: ■ 如果业务数据源区分生产数据源和开发数据源,则选择 生产+开发数据源 。 ■ 如果业务数据源不区分生产数据源和开发数据源,则选择 生产数据源 。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源				
* JDBC URL	jdbc:vertica://47.96. 5433/dataphin		* JDBC URL	jdbc:vertica://47.96.	5433/dataphin		
Schema	public	()	Schema	public			(
* 用户名	dbadmin		* 用户名	dbadmin			
* 密码		8	* 密码				₿
					juve Suve		
				测试连接	取消	确定	datap

② 说明 通常,生产数据源和开发数据源配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述
JDBC URL	链接地址的格式为 jdbc:vertica//{ 服务器地址 }:{端口}/{ 实例名称 } 。 例如, jdbc:vertica//10.0.0.1:5433/dataphin 。
Schema	Vertica实例对应的数据库名称,例如 public 。
用户名、密码	访问Vertica实例的用户名和密码。

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成Vertica数据源的创建。

后续步骤

完成Vertica数据源的创建后,即可将Vertica的业务数据引入至Dataphin或将Dataphin的数据写入至 Vertica。具体操作,请参见配置离线单条管道。

6.8.8. 创建AnalyticDB for PostgreSQL数据源

创建AnalyticDB for PostgreSQL数据源用于实现Dataphin能够读取AnalyticDB for PostgreSQL的业务数据。 在引入AnalyticDB for PostgreSQL的业务数据至Dataphin和将Dataphin的数据写入至AnalyticDB for PostgreSQL的场景中,您需要先创建AnalyticDB for PostgreSQL数据源。本文为您介绍如何创建AnalyticDB for PostgreSQL类型的数据源。

背景信息

AnalyticDB for PostgreSQL即阿里云的云原生数仓AnalyticDB PostgreSQL,如果您使用的是云原生数仓 AnalyticDB PostgreSQL,在对接Dataphin进行数据开发时,您需要先完成AnalyticDB for PostgreSQL数据 源的创建。更多云原生数仓AnalyticDB PostgreSQL信息,请参见产品概述。 Dataphin支持通过公网IP和VPC网络连接AnalyticDB for PostgreSQL。不同网络类型的连通方式,所需的配置细节不一致,如下图所示。



为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中,更多对接配置详情可参见设置白名单。

使用限制

- 仅支持超级管理员和项目管理员角色创建数据源。
 如何添加项目成员并授予项目管理员角色,请参见添加项目成员。
- 当业务数据源使用VPC网络时,则仅支持创建与Dataphin实例在同一地域AnalyticDB for PostgreSQL数据 源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。如何设置白名单请参见设置白名单。

• 访问方式为公网IP时,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● 访问方式为VPC网络时,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。VPC网络下Dataphin 的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26

地域	(推荐)通用IP地址 100.104.0.0/16	地域IP地址
华南1(深圳)		100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建AnalyticDB for PostgreSQL数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 規划								凿 🧔	ి 📀
业务	数据源 (190) ①						Q 遗输入检索关键字	V	+ #32853859
奇 数据编约	auzhi ass 🕫				Ifs lindom D				
₩ 公共定义	成型: OSS				問題: Lindorm				
项目	生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①
□ 项目管理	交通人: Landida Danida (Landida)								
计算题	政强派用途: 集线 化建用户: " Num (Met Coge A set (Consta t)		数据源用途: 集线 创建用户: Man adalama Capacitana (Manadalama)		政派原用途: 集战 创建用户: 计正式 计算机 计正式 中国 计算机		数据原用途: 淘线 创建用户:		
³³ ()	金融時前前: 2022-03-09 20:27:15 通知時前前: 2022-03-11 12:25:27		创建时间: 2022-03-09 20:30:32 更新时间: 2022-03-11 12:25:33		会理を対応: 2022-03-09 11:03:47 運動を封闭: 2022-03-09 11:03:47		创建时间: 2022-03-09 11:03:47 更新时间: 2022-03-09 11:03:47		
S SOURCES	· (199)(1月): ◎ 宣告(199)(1月)				119/11名: © 宣香把块信息		组织语言: © 查看短接位是		
	损件:	0 1 1 2	操作:	0 0 0 4	操作:	0 11 11 12	課作:	e 🛛	ē A

3. 在新建数据源对话框的关系型数据库区域,选择AnalyticDB for PostgreSQL。

如果您最近使用过AnalyticDB for PostgreSQL,也可以在最近使用区域选择AnalyticDB for PostgreSQL。同时,您也可以在快搜索框中,输入AnalyticDB for PostgreSQL的关键词,快速筛选。



4. 在新建AnalyticDB for PostgreSQL数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建AnalyticDB for F	PostgreSQL数据源
*数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
★ 数据源配置 ● "生产+开发"数据源	○"生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源				开发数据源			
* JDBC URL	jdbc:postgresql://gp-b	-master.gpdbmaster.rds.	.al	* JDBC URL	jdbc:postgresql://gp-bp	-master.gp	dbmaster.rds.al
Schema	public		() Schema	public		0
* 用户名	dataphin			* 用户名	dataphin		
* 密码			₿	* 密码			₿ataph
* 访问方式	● VPC网路 ○ 公网IP			* 访问方式	● VPC网路 ○ 公网IP		
* 区域	华东1(杭州)			* 区域	华东1(杭州)		
* VPC ID	vpc-bp10(ctnh1c			* VPC ID	vpc-bp10 tnh1c		
* 集群 ID	gp-bp1 b06e			* 集群 ID	gp-bp1 006e		
提示: 当前使用的 100.104.45.0/26	数据库为VPC网络环境下阿里云数据库,这	主接请指定授权IP白名单:		提示: 当前使用的 100.104.45.0/26	数据库为VPC网络环境下阿里云数据	库, 连接请指定授 [;]	权IP白名单:
					测试连接	取消	确定

⑦ 说明 通常,生产数据源和开发数据源配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数

描述

参数	描述
JDBC URL	链接地址的格式为 jdbc:postgresql//{链接地址}:{端口}/{实例名称} 。 在数据库连接页面查看链接地址和端口。 Image: State I
用户名、密码	当前账号的AccessKey ID和AccessKey Secret。如何获取,请参见 <mark>获取</mark> <mark>AccessKey</mark> 。
访问方式	选择对应的访问方式: ■ 数据源所在网络环境为公网,则访问方式选择为 公网IP 。 ■ 数据源所在网络环境为阿里云VPC网络,则访问方式选择为 VPC网络 。
	当 访问方式 选择为VPC网路,则需要配置如下参数: 区域:仅支持选择当前Dataphin实例所在的地域。 ② X m - 但古林又如上口之口之间,它们在只,此时他在中山立 PD (
区域、VPC ID、集群 ID	 ● 说明 仪支持添加与Dataphin实例在同一地或的AnalyticDB for PostgreSQL数据源至Dataphin。 ● VPC ID: 填写AnalyticDB for PostgreSQL实例的VPC ID。 ● 集群 ID: 填写AnalyticDB for PostgreSQL集群ID。 在云数据仓库AnalyticDB PostgreSQL实例列表页面,查看实例ID和VPC ID。
	实例名称 网络类型 gp-uf63zuy ● 专有网络 (VPC:vpc-uf6 7r)

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成AnalyticDB for PostgreSQL数据源的创建。

后续步骤

完成AnalyticDB for PostgreSQL数据源的创建后,即可将AnalyticDB for PostgreSQL的业务数据引入至 Dataphin或将Dataphin的数据写入至AnalyticDB for PostgreSQL。具体操作,请参见配置离线单条管道。

6.8.9. 创建PolarDB数据源

创建PolarDB数据源用于实现Dataphin能够读取PolarDB的业务数据,及能够向PolarDB写入数据。在引入 PolarDB的业务数据至Dataphin和将Dataphin的数据写入至PolarDB的场景中,需要先创建PolarDB数据源。 本文为您介绍如何创建PolarDB数据源。

背景信息

PolarDB是阿里云自主研发新一代关系型数据库,兼容MySQL、PostgreSQL语法的云托管数据库产品。如果 您使用的是PolarDB,在对接Dataphin进行数据开发时,您需要先完成PolarDB数据源的创建。更多PolarDB 信息,请参见什么是PolarDB。

Dataphin仅支持对接MySQL和PostgreSQL作为引擎的PolarDB数据库,支持通过VPC网络和公网连接PolarDB 数据库。PolarDB支持以下三种类型的MySQL和PostgreSQL数据库作为引擎:公网数据库、阿里云数据库和 ECS(VPC)自建数据库。数据库类型不同,所需的配置细节不同,如下图所示。



阿里云数据 库	即阿里云PolarDB MySQL或PolarDB PostgreSQL。创建数据源过程中,需要根据业 务情况选择访问方式(公网IP和VPC网络)后, 再配置网络连通参数。	1. 步骤一:添加Dataphin IP地址至数据库白
公网数据库	即具备公网访问能力的数据库,访问方式默认为 公网IP。直接添加Dataphin IP地址至数据库白名 单中即可。	名单 2. 步骤二: 创建PolarDB数据源
ECS(VPC) 自建数据源	即在阿里云ECS上自建的数据库且访问方式为 VPC网络。直接配置网络连通参数即可。	

为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据

源的白名单中。

使用限制

- 仅支持超级管理员和项目管理员角色创建数据源。
 如何添加项目成员并授予项目管理员角色,请参见添加项目成员。
- 当业务数据源(PolarDB)使用VPC网络时,则仅支持创建与Dataphin实例在同一地域PolarDB数据源。

步骤一:添加Dataphin IP地址至数据库白名单

基于PolarDB数据源引擎(MySQL或PostgreSQL)所在的网络环境,添加对应环境下的Dataphin IP地址至数 据库白名单中。以阿里云数据库为例为您介绍如何添加白名单,请参见设置白名单。

• 公网数据库、阿里云数据库且访问方式为公网访问,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● ECS(VPC)自建数据源、阿里云数据库且访问方式为VPC网络,则需要添加VPC网络下Dat aphin IP地址至数据库白名单中。ECS(VPC)自建数据源如何添加白名单请参见添加安全组规则。VPC网络下Dat aphin的 IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建PolarDB数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 规划	l							· 🏕 🗘 🖉 🖉
业务	数据源 (190) ①						Q 遗输入检索关键字	V + 55824036
- 数据架构	quzhi oss 🕫				Ifs lindorm 🖂			
₩ 公共建义	前型: OSS				黄型: Lindorm			
項日	生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for Dev ①
▷ 项目管理	(文资人: 1 				(文章人:)			
 ↓ ↓<!--</th--><th>教保護用途: 東城 公理用户:</th><th></th><th>数項項用途:東坡 公理用中: 日本時間の日本時間の日本時間の日本時間</th><th></th><th>教派原用途: 高线 创建用户:</th><th></th><th>数据源用途: 离线 创建用户:</th><th></th>	教保護用途: 東城 公理用户:		数項項用途:東坡 公理用中: 日本時間の日本時間の日本時間の日本時間		教派原用途: 高线 创建用户:		数据源用途: 离线 创建用户:	
а 🚹	()開始(): 2022-03-09 20:27:15 第2022-03-09 20:27:15		())()()()()()()()()()()()()()()()()()(()(間)()(): 2022-03-09 11:03:47		(100F)101: 2022-03-09 11:03:47	
S BURG	111111111111111111111111111111111111		○回時回回:2022-03-111225:33 包括信息: ● 查看任任信息		1999年1月:2022-03-09 1135-07 1999年日: ● 宣都任法信息		1999年1月1日:2022-05-09 1135-47 短校信息: ● 宣都任任信息	
	操作:	0 0 8 A	勝行 :	0004	操作:	C D B A	操作:	e e e A

3. 在新建数据源对话框的关系型数据库区域,选择PolarDB。

如果您最近使用过PolarDB,也可以在**最近使用**区域选择PolarDB。同时,您也可以在快搜索框中,输入 PolarDB的关键词,快速筛选。



- 4. 在新建PolarDB数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建PolarDB数据源	×
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据库类型 🥑 MYSQL 🔵 POSTO	GRE_SOL
∗数据源配置 💿 "生产+开发"数据源 🤇	○ "生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对PolarDB数据源的简单描述。
数据库类型	基于PolarDB底层的引擎类型,选择对应的数据库类型: 如果PolarDB底层的引擎为MySQL,则选择MySQL。 如果PolarDB底层的引擎为PostgreSQL,则选择PostgreSQL。
数据源配置	基于业务数据源是否区分生产数据源和开发数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

数仓规划·数据源

生产数据源			开发数据源		
			05526		
* JDBC URL	请输入JDBC连接地址,如jdbc:postgresql://host.port/database		* JDBC URL	请输入JDBC连接地址,如jdbc:postgresql://host:port/database	
Schema	指定读取的Schema信息	() Schema	指定读取的Schema信息	0
* 用户名			* 用户名	请输入鉴权用户名,为保证任务正常执行,请确保有所需数据权限	艮
* 密码		₿	* 密码	请输入鉴权密码,为保证任务正常执行,请确保有所需数据	₿
da 1 * 类型	● 公网数据库 ● 阿里云数据库 ● ECS(VPC)自建数据源		* 类型	○ 公网数据库 ● 阿里云数据库 ○ ECS(VPC)自建数据源	datap
* 区域	请选择Region		* 访问方式	● VPC网路 ○ 公网IP	
* VPC ID	请输入VPC ID		* 区域	请选择Region	
* ECS ID	请输入ECS ID		* VPC ID	请输入VPC ID	
提示:当前使用的 100.104.228.128/3	数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 26;100.104.115.192/26		* PolarDB ID	请输入PolarDB ID	
			提示: 当前使用的 100.104.228.128/	数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 26;100.104.115.192/26	
				测试连接 取消 确定	Qater

⑦ 说明 通常,生产数据源和开发数据源需要配置不同的参数值,但Dataphin也支持配置成相同的参数值。

参数	描述
JDBC URL	 链接地址的格式为 jdbc:{数据库类型}://{链接地址}:{端口}/{数据库名称} 。其中: 数据库类型: 上述步骤中数据库类型选择为MySQL,则配置为 mysql 。 上述步骤中数据库类型选择为PostgreSQL,则配置为 postgresql 。 以数据库类型为MySQL为例,为您介绍如何获取链接地址、端口和数据库名称: 在数据库连接页面,获取MySQL实例的链接地址、端口。 * MySQL (v @@v) v @ model (Market (M

参数	描述			
Schema	当上述步骤中数据库类型选择为PostgreSQL,需要配置Schema。Schema即 PostgreSQL实例的数据库名称。在PostgreSQL实例的数据管理页面获取数据库名称,如 下图所示。 基本信息			
用户名、密码	登录PolarDB数据源引擎MySQL或PostgreSQL的用户名和密码。			
类型	PolarDB数据源的引擎MySQL、PostgreSQL均支持选择以下几种类型: 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 阿里云数据库:如果您的数据源是阿里云数据库RDS,请选择库阿里云数据。阿里云数据库支持通过公网IP或VPC网络连接Dataphin。 ⑦ 说明 如果您的数据源是阿里云数据库且支持公网访问,则您可以选择公网			
	 数据库,也可以选择阿里云数据库(公网Ⅳ)。 ECS(VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的 网络环境为VPC,请选择ECS(VPC)自建数据源。 			

参数	描述
	当类型选择为ECS (VPC)自建数据源和阿里云数据库(VPC网络),还需配置区域、VPC ID、PolarDB ID/ECS ID这些网络参数,其中: 阿里云数据库(VPC网络):配置如下参数。 区域:仅支持选择当前Dataphin实例所在的地域。
	⑦ 说明 仅支持添加与Dataphin实例在同一地域的PolarDB数据源至 Dataphin。
	 VPC ID:填写PolarDB实例的VPC ID。 PolarDB ID:填写PolarDB实例的ID。 ECS (VPC) 自建数据源:配置如下参数。
	 区域:仅支持选择当前Dataphin实例所在的地域。 ⑦ 说明 仅支持添加与Dataphin实例在同一地域的MySQL数据源至 Dataphin。
	 VPC ID:填写云服务器ECS实例的VPC ID。 ECS ID:填写云服务器ECS实例的ID。 您可以在PolarDB控制台的实例基本信息页面、ECS实例控制台的实例详情页面查看VPC ID PolarDB ID ECS ID 如下图所示
区域、VPC ID、PolarDB	
ID/ECS ID	回里 二 致 指 年 的 内 各 信 息 数 編 年 2 8 数 編 年 2 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
ID/ECS ID	り 里 云 釵 拈 庄 的 网 络 信 息 云 数 編集RDS / 案例形象 / 基本信息 ◆ 面 面 面 面 (→ 面行中) → 基本信息 影響管: 影響管: 家務D Pgm-bp1v Pm-bp1v
ID/ECS ID	り 里 云 釵 拈 车 的 网 络 信 息
ID/ECS ID	

- 5. 单击**测试连接**,测试数据源是否可以和Dataphin进行正常的连通。 如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。
- 6. 测试成功后,单击确定,完成PolarDB数据源的创建。

后续步骤

完成PolarDB数据源的创建后,即可将PolarDB的业务数据引入至Dataphin,或将Dataphin的数据写入至 PolarDB数据源。具体操作,请参见配置离线单条管道。

6.8.10. 创建Teradata数据源

创建Teradata数据源用于实现Dataphin能够读取Teradata的业务数据。在引入Teradata的业务数据至 Dataphin和将Dataphin的数据写入至Teradata的场景中,您需要先创建Teradata数据源。本文为您介绍如 何创建Teradata类型的数据源。

背景信息

Teradata是一款大型数据仓库系统。如果您使用的是Teradata,在对接Dataphin进行数据开发或导出 Dataphin数据至Teradata,您需要先完成Teradata数据源的创建。更多Teradata信息,请参见Teradata官网。

Dataphin支持添加的Teradata数据源类型包括公网数据库和ECS(VPC)自建数据源。不同类型的网络连通方式下,所需的配置细节不一致,如下图所示。



为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

• 当数据源类型为ECS(VPC)自建数据源时,则仅支持创建与Dataphin实例在同一地域Teradata数据源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。

• 公网数据库需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

• ECS(VPC)自建数据源需要添加VPC网络下Dataphin IP地址至数据库白名单中。如何添加白名单,请参见添加安全组规则。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建Teradata数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

≡	Dataphin 规划									ి 🕗 🕓
92.95 92.95		数据源 (190) ①						Q. 清給入證素关證字	V	+ 8/22/20202
∴ # ₩ 2	数摄解构 公共定义	guzhi_oss 전 변화: OSS			Ē	Hs_lindorm ⊠ K월:: Lindorm				Ē
項目		生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①
B 1	页目管理	交通人:						SEA: THE OTHER AND ADDRESS		
411 1	+算線	数据源用途: 淘试 创建用户:		数据源用法: 案例 创建用 ^内 : here independing and the second second		数据源用注: 建成 创建用户:		数据源用途: 素线 创魂用户::		
19 2 3	() 83639	出陸时间: 2022-03-09 20:27:15 更新时间: 2022-03-11 12:25:27 現後信息: ◆ 宣告相決信息		创建时间: 2022-03-09 20 30 32 単新时间: 2022-03-11 12:25:33 総約信意:		创建时间:2022-03-09 11:03:47 単新日间:2022-03-09 11:03:47 規模信意: ● 宣告紀研究者員		创想対応: 2022-03-09 11:03:47 開新対応: 2022-03-09 11:03:47 頻致信意:		
		提作:	008	操作 :	0 0 0 2	操作:	C D B A	损作 :	e 🛛	<u>ت</u> ۵

3. 在新建数据源对话框的关系型数据库区域,选择Teradata。

如果您最近使用过Teradata,也可以在**最近使用**区域选择Teradata。同时,您也可以在快搜索框中, 输入Teradata的关键词,快速筛选。

新建数据源			Q 请输	入搜索关键字 全部 ∨ X
关系型数据库				
VERTICA	IBM DB2	Teradata	Ş	
Vertica	IBM DB2	Teradata	Impala	
				关闭

- 4. 在新建Teradata数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建Teradata数据源	×
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据源配置 💿 "生产+开发"数据源	○ "生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源		开发数据源		
* JDBC URL	jdbc:teradata://10 I:12345/pfdb	* JDBC URL	jdbc:teradata://* :12345/pfdb	
* 用户名	teradatatest	* 用户名	teradatatest	
* 密码	A	* 密码		Bataph
* 类型	○ 公网数据库 ● ECS(VPC)自建数据源	* 类型	○ 公网数据库 ● ECS(VPC)自建数据源	
* 区域	华东2(上海) ~ ~	* 区域	华东2(上海)	
* VPC ID	vpc-uf60m	* VPC ID	vpc-uf60mb(
* ECS ID	i-uf6a	* ECS ID	i-uf6a9	
提示:当前使用的 100.104.228.128/2 dataphins/	数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 6;100.104.115.192/26	提示:当前使用的 100.104.228.128/2	数据库为VPC网络环境下阿里云数据库,连接请指定授权 26;100.104.115.192/26	QIP白名单: datapi
			测试连接 取消	确定

② 说明 通常,生产数据源和开发数据源需要配置的参数值不同,但Dataphin也支持配置成 相同的参数值。

参数	描述					
JDBC URL	链接地址的格式为 jdbc:teradata://{服务器地址}:{端口}/{数据库名称} 。					
用户名、密码	登录Teradata数据库的用户名和密码。					
类型	 Dataphin支持选择以下几种数据源类型: ■ 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 ■ ECS(VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的网络环境为VPC,请选择ECS(VPC)自建数据源。 					
	 当类型选择为ECS (VPC) 自建数据源,则需要配置如下参数: 区域:仅支持选择当前Dataphin实例所在的地域。 ⑦ 说明 仅支持添加与Dataphin实例在同一地域的Teradata数据源至 Dataphin。 					
区域、VPC ID、ECS ID	 VPC ID:填写Teradata实例的VPC ID。 ECS ID:填写Teradata实例的ID。 您可以在ECS实例控制台页面查看ECS ID和VPC ID。 基本信息 基本信息 副部 ▲ ○ 已停止 					
	変例ID i-uf6a9si 近母道接 地域 华东2 (上海) 公河IP - 第定課性P 所在可用区 上海可用区 网络信息 网络典型 专有网络 vpc-uf60mbqne L 弹性网卡 eni-uf628 虚拟交换机 vsw-uf637jqtow L 主私同IP 10.0 弾性IP究例ID -					

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成Teradata数据源的创建。

后续步骤

完成Teradata数据源的创建后,即可将Teradata的业务数据引入至Dataphin或将Dataphin的数据写入至 Teradata。具体操作,请参见配置离线单条管道。

6.8.11. 创建IBM DB2数据源

创建IBM DB2数据源用于实现Dat aphin能够读取IBM DB2的业务数据。在引入IBM DB2的业务数据至Dat aphin 和将Dat aphin的数据写入至IBM DB2的场景中,您需要先创建IBM DB2数据源。本文为您介绍如何创建IBM DB2类型的数据源。

背景信息

IBM DB2是一套关系型数据库管理系统。如果您使用的是IBM DB2,在对接Dat aphin进行数据开发时,您需要 先完成IBM DB2数据源的创建。更多IBM DB2信息,请参见IBM DB2官网。

Dataphin支持添加的IBM DB2数据源类型包括公网数据库和ECS(VPC)自建数据源。不同类型的网络连通方式下,所需的配置细节不一致,如下图所示。



为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

自建数据源

仅支持超级管理员和项目管理员角色创建数据源。
 如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

VPC网络。直接配置网络连通参数即可。

• 当数据源类型为ECS(VPC)自建数据源时,则仅支持创建与Dataphin实例在同一地域IBM DB2数据源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。

● 公网数据库需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202

地域	IP地址
华东1(杭州)	47.96.75.100

• ECS(VPC)自建数据源需要添加VPC网络下Dataphin IP地址至数据库白名单中。如何添加白名单,请参见添加安全组规则。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)	100.104.0.0/16	100.104.228.128/26、 100.104.115.192/26
华南1(深圳)		100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建IBM DB2数据源

1. 请参见入口介绍,进入数仓规划页面。

2. 按照下图指引,进入新建数据源页面。

■ Dataphin 规划									ి 🕗 🦉
业务	数据源 (190) ①						Q 请输入虚素关键字	V	+ #32803839
A 数据转向	guzhi_oss ⊠				lfs_lindorm 🖂				
₩ 公共定义	(전전: OSS				<u>資型</u> : Lindorm				
項目	生/ ^{ter} for	e Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev ①
□ 项目管理	会进入: ************************************								
41 it ji 22	救援源用途: 救援 会建用户:		政振荡用途: 高城 台牌用户:		数据源用途: 車 銭 (2)建用 ^の : 「一」の「「一」の「」」の「」」の「」」の「」」の「」」の「」」の「」」の「」」		数据源用途: 淘试 创建用户:		
¥ ()	(1)建計時: 2022-03-09 20:27:15 期間の計算: 2022-03-09 20:27:15		①読号目前: 2022-03-09 20:30:32 期間の目前: 2022-03-09 20:30:32		会議部計測: 2022-03-09 11:03:47 期間の計測: 2022-03-09 11:03:47		台通时间: 2022-03-09 11:03:47 WREPHR: 2022-03-09 11:03:47		
	● 宣香任扶住思		◎ 宣誓任法位思		● 宣后性疾信息		◎ 宣告性接信息		
	銀作:	e 12 8 8	损性 :	6 6 8 A	操作:	C 🛛 🗎 🚨	操作:	e 2	8

3. 在新建数据源对话框的关系型数据库区域,选择IBM DB2。

如果您最近使用过IBM DB2,也可以在最近使用区域选择IBM DB2。同时,您也可以在快搜索框中,输入 IBM DB2的关键词,快速筛选。

新建数据源			Q 请编〉	入搜索关键字 全部 V X
关系型数据库				
VERTICA	IBM DB2	Teradata	Ŷ	
Vertica	IBM DB2	Teradata	Impala	
				关闭

4. 在新建IBM DB2数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建IBM DB2数据源	X
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
*数据源配置 • "生产+开发"数据源 ("生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源		开发数据源		
* JDBC URL	jdbc:db2://101.1 32:50000/TESTDB	* JDBC URL	jdbc:db2://101 132:50000/TESTDB	
* 用户名	db2inst1	* 用户名	db2inst1	
* 密码		* 密码	Batap	
* 类型	○ 公网数据库	* 类型	○ 公网数据库	
* 区域	华东2(上海) ~	* 区域	华东2(上海) 🗸	
* VPC ID	vpc-uf60mbc zt043	* VPC ID	vpc-uf60m zt043	
* ECS ID	i-uf6a9s :ljfv2	* ECS ID	i-uf6a9s jfv2	
提示:当前使用的数据库为VPC网络环境下阿里云数据库,连接请指定授权IP白名单: 100.104.228.128/26;100.104.115.192/26 100.104.228.128/26;100.104.115.192/26				
			测试连接 取消 确定	

⑦ 说明 通常,生产数据源和开发数据源需要配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述	
JDBC URL	链接地址的格式为 jdbc:db2://{服务器地址}:{端口}/{数据库名称} 。	
用户名、密码	登录IBM DB2数据库的用户名和密码。	

参数	描述			
类型	 Dataphin支持选择以下几种数据源类型: 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 ECS(VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的网络环境为VPC,请选择ECS(VPC)自建数据源。 			
	当 类型 选择为ECS(VPC)自建数据源,则需要配置如下参数: ■ 区域:仅支持选择当前Dataphin实例所在的地域。			
区域、VPC ID、ECS ID	 ⑦ 说明 仪支持添加与Dataphin实例在同一地域的IBM DB2数据源至 Dataphin。 PVPC ID:填写IBM DB2实例的VPC ID。 ECS ID:填写IBM DB2实例的ID。 您可以在ECS实例控制台页面查看ECS ID和VPC ID。 			
	基本信息 诊断健康状态 ☎ 1 启动 1 launch ▲ ● 已停止 空間10 i-uf6a9si 公同10 - 携寇弹性P 所在可用区 网络信息 -			
	网络类型 专有网络 使有网络 vpc-uf60mbqne C 弹性网卡 eni-uf628 虚拟交换机 vsw-uf637jqtow C 主私网IP 10.0 弹性P实例ID -			

5. 单击**测试连接**,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成IBM DB2数据源的创建。

后续步骤

完成IBM DB2数据源的创建后,即可将IBM DB2的业务数据引入至Dat aphin或将Dat aphin的数据写入至IBM DB2。具体操作,请参见配置离线单条管道。

6.8.12. 创建OceanBase数据源

创建OceanBase数据源用于实现Dataphin能够读取OceanBase的业务数据,及能够向OceanBase写入数据。 在引入OceanBase的业务数据至Dataphin和将Dataphin的数据写入至OceanBase的场景中,您需要先创建 OceanBase数据源。本文为您介绍如何创建OceanBase类型的数据源。

背景信息

OceanBase即阿里云数据库OceanBase,是阿里巴巴和蚂蚁金服100%自主研发的金融级分布式关系数据库。如果您使用的是阿里云OceanBase,在对接Dataphin进行数据开发或将Dataphin数据写入至OceanBase时,您需要先完成OceanBase数据源的创建。更多OceanBase信息,请参见什么是云数据库 OceanBase。

Dat aphin支持创建公网数据库、阿里云数据库和ECS(VPC)自建数据源类型的OceanBase数据源。不同类型的数据源,所需的网络连通配置细节不同,如下图所示。



数据源类型	数据源类型和网络连通说明	网络连通操作指导	
阿里云数据 库	即阿里云数据库OceanBase。创建数据源过程 中,需要根据业务情况选择访问方式(公网IP和 VPC网络)后,再配置网络连通参数。		
公网数据库	即具备公网访问能力的数据库,访问方式默认为 公网IP。直接添加Dataphin IP地址至数据库白名 单中即可。	 步骤一:添加Dataphin IP地址至数据库白 名单 2.步骤二:创建OceanBase数据源 	
ECS(VPC) 自建数据源	即在阿里云ECS上自建的数据库且访问方式为 VPC网络。直接配置网络连通参数即可。		

为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

 仅支持创建与Dataphin实例在同一地域的ECS(VPC)自建数据源和阿里云数据库(VPC网络)类型的数据 源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。公网数据库和阿里 云数据库添加白名单请参见白名单分组。ECS(VPC)自建数据源添加白名单请参见添加安全组规则。

公网数据库、阿里云数据库且访问方式为公网访问,则需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址
地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

• ECS(VPC)自建数据源、阿里云数据库且访问方式为VPC网络,则需要添加VPC网络下Dataphin IP地址至数据库白名单中。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)	100.104.0.0/16	100.104.228.128/26、 100.104.115.192/26
华南1(深圳)		100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建OceanBase数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 规划									ి (2)	
业务	数据源 (190) ①						Q 清箱入掘地关部序		+ 852	103050
▲ 数据架构 ₩ 公共定义	guzhi_oss ⊠ 类型: oSS			ē	lfs_lindorm ⊠ 問題: Undorm					ē
项目	生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for	Dev	D
□项目管理 411计算数 32 ① ■ 数据数	○選入: 設備原用: 出版可用: 出版可用: 出版可用: 認知(2) 202243-49 20 27 15 展示可由: 202243-11 12 25 27 社師保留: ● 面面相知知思.				(生人): 数据初用法: 載成 (結果用): 出版時間: 202243-00 1103-47 服務時間: 202243-00 1103-47 服務時間: ① 宣告協発電器		交長人:			
	损 件:	0 C B A	操作:	0052	操作:	C E E A	派作:	e B	ē J	8

3. 在新建数据源对话框的关系型数据库区域,选择OceanBase。

如果您最近使用过OceanBase,也可以在**最近使用**区域选择OceanBase。同时,您也可以在快搜索框中,输入OceanBase的关键词,快速筛选。

新建数据源			Q 请输入指	繁美键字 全部 V X
Ś	8	E.F.		abe Antipe
PolarDB	PolarDB-X(原DRDS)	MySQL	Microsoft SQL Server	PostgreSQL
\diamond	\diamond	\odot		ORACLE
AnalyticDB for MySQL 2.0	AnalyticDB for MySQL 3.0	AnalyticDB for PostoreSQL	OceanBase	Oracle
				关闭

- 4. 在新建OceanBase数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建OceanBase数据	Sig. X
 数据源名称 请输入数据源名称 数据源描述 请输入数据源描述 	0/128
* 数据源配置 💿 "生产+开发"数据源	"生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: ■ 如果业务数据源区分生产数据源和开发数据源,则选择 生产+开发数据源。 ■ 如果业务数据源不区分生产数据源和开发数据源,则选择 生产数据源 。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源		开发数据源	
* JDBC URL		* JDBC URL	
* 用户名	请输入鉴权用户名,为保证任务正常执行,请确保有所需数据权限	* 用户名	请输入鉴权用户名,为保证任务正常执行,请确保有所需数据权限
* 惑码	请输入鉴权密码,为保证任务正常执行,请确保有所需数据 🔒	* 密码	请输入鉴权密码,为保证任务正常执行,请确保有所需数据
* 类型	○ 公网数据库 ● 阿里云数据库 ○ ECS(VPC)自建数据源	* 类型	〇 公网数据库 阿里云数据库 ECS(VPC)自建数据源
* 访问方式	○ VPC网络 ● 公网IP	* 访问方式	○ VPC网络 ● 公网IP
本地自建数据库授 认资源组的公网出	权,需要开放访问,并授权权限。如需授权白名单,请添加Dataphin默 网IP:47.102.151.182	本地自建数据库授 认资源组的公网出	权,需要开放访问,并授权权限。如需授权白名单,请添加Dataphin默 网IP:47.102.151.182
			测试连接 取消 希定

⑦ 说明 通常,生产数据源和开发数据源需要配置不同的参数值,但Dataphin也支持配置成相同的参数值。

参数	描述
JDBC URL	OceanBase连接地址,格式为 ■ MySQL模式: jdbc:mysql://<连接地址>:<端口>/<数据库名称> 。 ■ Oracle模式: jdbc:oracle://<连接地址>:<端口>/<数据库名称> 。 WDP里云数据库OceanBase的MySQL模式为例,参数获取方式如下。 ■ 连接地址、端口: 在云数据库OceanBase控制合的租户管理页面,选择对应的租户,查看 连接地址和端口信息(图示①和图示②)。 ■ 数据库名称: 在对应租户详情的数据库管理页面,查看数据库名称(图示③)。
用户 名和密码	目标数据库的用户名和密码,详情请参见 <mark>账号管理</mark> 。
类型	Dataphin支持选择以下几种数据源类型: 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 阿里云数据库:如果您的数据源是阿里云数据库RDS,请选择阿里云数据库。阿里云数据库支持通过公网IP或VPC网络连接Dataphin。 ⑦说明 如果您的数据源是阿里云数据库RDS且支持公网访问,则您可以选择公网数据库,也可以选择阿里云数据库(公网IP)。 ECS (VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且使用的网络环境为VPC,请选择ECS (VPC)自建数据源。
	数据源类型选择不同,则 访问方式 不同。

	■ Acc 式, ■ 实(cessID、AccessKey 请参见 <mark>获取AccessK</mark>				
	⊥, ■ 实(y :当前账号[/ev/	的AccessK	ey ID和AccessKey S	Secret。获取方
	- 71	のし、埴宮阿甲テ数排	egenceanBa	o佳群家伍	مىللە	
			a)年Oceaniba. 民庁OceaniBa		うし。 白的VPC ID	
	- VF ■ 租 f	它们 :填写阿里云数据	a库OceanBa	ららしていたい		
	您可以	以在云数据库OceanBa	ise控制台的和	1户管理页	~。 〔面,选择对应的租/	^ጏ ,查看实例
	ID(图	示①)、VPC ID(图	示②)、租户	ID(图示(3).	
	< iscalari dolumi dataphin * D: 1223 = mm	田戸工作台 (dataphin) ○ で の の の の の の の の の の の の の の の の の		10785: 07851 452(128	1	ненне 22.852.864 Блежалом ··· Ф ж В (PU 25.67 Влежалом
	• 运行中 服中工作台					2 MR 5 GB 0 GB
	較限加管理 标号管理 會政管理	波司森求量 ⊙ — tps — qps		2	© Manakanaka Saka — tps.pt. — qps.pt.	z
	性酸温控 安全设置	0.6 0.5 0.4 0.3			2 ms	
	■ 公网Ⅱ	P:填写阿里云数据库	OceanBase	目标租户的	1公网地址。	
	■ 当米司 湖		建新拓进时			
	■ 日 安 空心		建数116			
	■	D: 填与云服务器ECS	的实例ID。			
	VPC I	ID: 填写云服务器ECS	实例的专有网	刚络ID。		
访问方式	■ VPC ■ 租户	ID:填写云服务器ECS D:填写阿里云数据库	实例的专有网 OceanBase	刚络ID。 的租户ID。	获取方式 <i>,</i> 请参考 3	美型 选择为 阿里
访问方式	■ VPCI ■ 租户II 云数排	ID:填写云服务器ECS D:填写阿里云数据库 居库 时的 租户ID 描述。	实例的专有网 OceanBasef	刚络ID。 的租户ID。	获取方式,请参考 3	€型 选择为 阿里
访问方式	 VPC I 租户II 云数打 您可以在 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的	实例的专有网 OceanBasef 的实例详情页	函络ID。 的租户ID。 面,查看V	获取方式,请参考 3 PC ID(图示①)和B	€型选择为 阿里 €CS的实例Ⅳ(图
访问方式	 VPCI 租户II 云数排 您可以在示②)。 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的	实例的专有网 OceanBase的 的实例详情页	列络ID。 的租户ID。 面,查看V	获取方式,请参考 3 PC ID(图示①)和B	€型选择为 阿里 €CS的实例Ⅳ(图
访问方式	■ VPCI ■ 租户II 云数初 您可以在 示②)。	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的	实例的专有网 OceanBase 的实例详情页 ^{1 运行中}	列络ID。 的租户ID。 面 <i>,</i> 查看V	获取方式,请参考 3 PC ID(图示①)和E	€型选择为 阿里 ECS的实例Ⅳ(图
访问方式	■ VPC ■ 租户II 云数排 您可以在 示②)。 launch-au <u>※例D</u>	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的	实例的专有网 OceanBase的 的实例详情页 ^{) 运行中} 远路继	列络ID。 的租户ID。 面,查看V	获取方式,请参考 3 PC ID (图示①)和E	€型选择为 阿里 €CS的实例Ⅳ(图
访问方式	 VPC 租户II 云数排 您可以在示②)。 launch-au ^{资源组}: 公网IP 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的 Indec 20211108zx ∠ ② i-uffect and a month of 2 i-uffect and a month of 2	实例的专有网 OceanBase 的实例详情页 运行中 运行中	列络ID。 的租户ID。 面,查看Ⅴ ^{地域} ^{所在可用区} ^{主机名}	获取方式,请参考 3 PC ID(图示①)和E	€型选择为 阿里 ECS的实例Ⅳ(图
访问方式	 VPCI 租户II 云数初 您可以在示②)。 launch-au ^{※例D} ^{※例D} ^{※例D} [※] ¹ ¹ ¹<td>ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的 Index 20211108zx 2 ② Index 20211108zx 2 ②</td><td>实例的专有网 OceanBase 的实例详情页</td><td>列络ID。 的租户ID。 面, 查看V ^{地域} ^{所在可用区} ^{主机名} 创建时间</td><td>获取方式,请参考3 PC ID (图示①)和E ^{华集} 上^{海可用区G} zx 2021年11月8日 14:34:00</td><td>養型选择为阿里 ECS的实例Ⅳ(图</td>	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的 Index 20211108zx 2 ② Index 20211108zx 2 ②	实例的专有网 OceanBase 的实例详情页	列络ID。 的租户ID。 面, 查看V ^{地域} ^{所在可用区} ^{主机名} 创建时间	获取方式,请参考 3 PC ID (图示①)和E ^{华集} 上 ^{海可用区G} zx 2021年11月8日 14:34:00	養型选择为 阿里 ECS的实例Ⅳ(图
访问方式	 VPC 租户II 云数排 您可以在示②)。 launch-au ^{按侧D} ^{按侧U} ^{按侧U} ^{按梁组} ^{插述} 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的 Address 20211108zx ∠ ② Fuffect and a month 2021 139.mm 17 sg-uf61 and 1 month 2021 139.mm 17 sg-uf61 and 1 month 2021	实例的专有风 OceanBase 的实例详情页 ;运行中 旋路推整 转换为弹性公网IP 加入安全组 修改实例描述	列络ID。 的租户ID。 面,查看∨ ^{地域} ^{所在可用区} ^{主机名} ^{创建时间} 自动释放时间	获取方式,请参考到 PC ID (图示①)和E ^{华集} 上海可用区G zx 2021年11月8日 14:34:00	養型选择为阿里 ECS的实例ID(图 ^{修改实例主机名} ^{释放设置}
访问方式	 VPCI 租户II 云数排 您可以在示②)。 launch-au 实例ID 资源组: 公网IP 安全组 描述 CPU&内存 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 云服务器ECS控制合的 ■ 100	实例的专有风 OceanBasef 的实例详情页 	列络ID。 約租户ID。 面,查看V 地域 所在可用区 主机名 创建时间 自动释放时间 云盘	获取方式,请参考到 PC ID (图示①)和E ^{华集}	美型选择为阿里 ECS的实例ID(图 修改实例主机名 释放设置
访问方式	 VPC 租户II 云数排 您可以在示②)。 launch-au 实例D 资源组: 公网IP 安全组 描述 CPU&内存 操作系统 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的 Construct 20211108zx 2 ② I-uf6cf	实例的专有风 OceanBased 的实例详情页 这行中 施入安全组 修改实例描述 更换操作系统	好名ID。 約租户ID。 面,查看V 地域 所在可用区 主机名 创建助间 自动解放时间 云盘 快照	获取方式,请参考到 PC ID (图示①)和E ^{华新} 上海可用区G zx 2021年11月8日 14:34:00 -	き型选择为阿里 ECS的实例ID(图 修改实例主机名 释放设置 国新初始化云盘
访问方式	 VPCI 租户II 云数初 您可以在示②)。 launch-au 资源组: 公网IP 安全组 描述 CPU&内存 操作系统 实例照信 ************************************	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的 Index 20211108zx 2 2 i-ufect 1008zx 2	实例的专有风 OceanBase 的实例详情页 ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;	列络ID。 的租户ID。 面,查看∨ 地域 所在可用区 主机名 创建时间 自动释放时间 云盘 快照 曉慮D	获取方式,请参考到 PC ID (图示①)和E ^{华集} (1等) 上海可用区G zx 2021年11月8日 14:34:00 -	美型选择为阿里 ECS的实例ID(图 修改实例主机名 释放设置 重新初始化云盘
访问方式	 VPCI 租户II 云数排 您可以在示②)。 launch-au 实例D 资源组: 公网IP 安全组 描述 CPU&内存 操作系统 实例规格 实例规格 实例规格 实例规格 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 三云服务器ECS控制台的 在100 20211108zx 2 ② 	实例的专有风 OceanBased 的实例详情页 ; 运行中 	好名ID。 约租户ID。 面,查看V 地域 所在可用区 主机名 创建时间 自动释放时间 云盘 快照 镜像ID 当前使用带宽	获取方式,请参考到 PC ID (图示①)和E ^{华海} 1159) 上海可用区G zx 2021年11月8日 14:34:00 - 1 0 aliyun_3 x64_20G_alibase_20210 10Mbps (編蛋)	美型选择为阿里 ECS的实例ID(图 《 微文例主机名 「 「 新初始化云盘 近量付要实例更改策策
访问方式	 VPCI 租户II 云数初 您可以在 示②)。 launch-au 这原组: 公网IP 安全组 描述 CPU&内存 操作系统 实例规格 实例规格 实例规格 实例规格 示 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的 Inter 20211108zx 2 ② i-ufecfingeneous forf 39.umm 17 sg-uf6ingeneous forf 22 i-ufecfingeneous forf 22 i-ufecfingeneous forf 22 i-ufecfingeneous forf 22 i-ufecfingeneous forf 22 i-ufecfingeneous forf 39.umm 17 sg-uf6ingeneous forf 20 i-ufecfingeneous forf 20 i-ufecfingen	实例的专有风 OceanBasef 的实例详情页 运行中 旋環難接 转换为弹性公网P 加入安全组 修改实例描述 更改实例规情	外名ID。 约租户ID。 面,查看V 地域 所在可用区 主机名 创建时间 自动释放时间 云盘 快照 请像师D 当前使用带宽	获取方式,请参考到 PC ID (图示①)和E ^华 为:105 上海可用区G zx 2021年11月8日 14:34:00 - 1 0 aliyun_3_x64_20G_alibase_20210 10Mbps (雌佰)	 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
访问方式	 VPCI 租户II 云数初 您可以在 示②)。 launch-au ஜ//// ஜ//// ஜ//// ஜ//// ஜ/// ஜ/// ஜ//	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E云服务器ECS控制台的 Inter 20211108zx 2 ② i-uf6cf 20211108zx 2 ③ i-uf6cf 20211108zx 2 ④ i-uf6cf 20211108zx 2 ⑤ i-uf6cf 20211108zx 2 ⑤ i-uf6cf 20211108zx 2 ⑤ i-uf6cf 20211108zx 2 ⑤ i-uf6cf 20211108zx 2 ⑥ i-uf6cf 20211108zx 2 ⑧ i-uf6cf 2021108x 2 ⑧ i-uf6cf 20211108x 2 ⑧ i-uf6cf 2021108x 2 ⑧ i-uf6cf 20211108x 2 ⑧ i-uf6cf 2021108x 2 ⑧	实例的专有风 OceanBased 的实例详情页 词 运行中 随時開催 随時開催 随次实例描述 更快操作系统 更改实例规情	好名ID。 約租户ID。 面,查看V 地域 所在可用区 主机名 创建时间 自动释放时间 云盘 快照 镜像ID 当前使用带宽	获取方式,请参考3 PCID(图示①)和E ^{学生にの} 上海可用区G 2021年11月8日14:34:00 - 1 0 aliyun_3_x64_20G_alibase_20210 10Mbps (峰面)	
访问方式	 VPCI 租户II 云数折 您可以在示②)。 launch-au 实例ID 资源组: 公网IP 安全组 描述 CPU&的存 操作系统 实例规格 实例规格 实例规格表 · 网络信息 网络信息 	ID:填写云服务器ECS D:填写阿里云数据库 居库时的租户ID描述。 E 云服务器ECS控制合自 	实例的专有风 OceanBasef 的实例详情页 词 运行中 证程生度 输改实例描述 更进程作系统 更改实例照情	好名ID。 約租户ID。 面,查看V 地域 所在可用区 主机名 创建时间 自动释放时间 云盘 快照 镜像ID 当前使用带宽	获取方式,请参考到 PCID(图示①)和E 学編 1 (1) L時可用区G zx 2021年11月8日 14:34:00 - 1 0 aliyun_3_x64_20G_alibase_20210 10Mbps (峰価)	美型选择为阿里 ECS的实例ID(图 修改实例主机名 释放设置 重新初始化云盘 近量付卖实例更改带宽 场错标签

参数	描述

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见<mark>网络连通的常见问题</mark>。

6. 测试成功后,单击确定,完成OceanBase数据源的创建。

后续步骤

完成OceanBase数据源的创建后,即可将OceanBase的业务数据引入至Dataphin,或将Dataphin的数据写入 至OceanBase数据源。具体操作,请参见配置离线单条管道。

6.8.13. 创建Impala数据源

创建TImpala数据源用于实现Dataphin能够将数据写入至Impala中。在将Dataphin的数据写入至Impala的场景中,您需要先创建Impala数据源。本文为您介绍如何创建Impala类型的数据源。

背景信息

Impala是用于处理存储在Hadoop集群中大量数据的SQL查询引擎。如果您使用的是Impala,在导出 Dataphin数据至Impala,您需要先完成Impala数据源的创建。更多Impala信息,请参见Impala官网。

Dataphin支持通过公网连接Impala。为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin在 公网环境下的IP地址段添加到数据源的白名单中。

使用限制

- 仅支持超级管理员和项目管理员角色创建数据源。
 如何添加项目成员并授予项目管理员角色,请参见添加项目成员。
- 仅支持创建1.x、2.x、3.x版本的Impala数据源。

步骤一:添加Dataphin IP地址至数据库白名单

添加公网环境下的Dataphin IP地址至数据源白名单中。

公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

步骤二: 创建Impala数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 规划	l							凿 🗸	a <mark>o</mark> 🧕
<u>#</u> \$	数据源 (190) ☉						Q. 清箱入理素关键字	V	+ #122203032
☆ 数据架构 ₩ 公共定义	guzhi_oss ⊠ ⋈∰: oss			Ē	lfs_lindorm ⊠ 정편: Lindorm				1
项目	生/ ^{der} for Pr	rod / Basic ① 开发		for Dev ①	生 ^{phe}	for Prod / Basic ①	开发	for	Dev ①
□项目管理 11 计算源 32 1 ■ 数据版	今週人: 初編5冊念: 単純 包濃時で: 包濃時で, 222244-09 29 27 15 展時時で, 2222-0-11 12 25 27 日時時間: 2022-0-11 12 25 27	会選人 数据課 伯提用 台通知 現熟和 日 根報(約	用語: 憲領 P: 2022-03-09 20 30 32 2022-03-11 12 25 33 章: 全面包括宗征島		会長人: 数据原用意: 載成 (設備用): 2022-03-09 11:03-47 展新刊道: 2022-03-09 11:03-47 機能優響: ● 面信総理理		交長人: 数据局用途: 載成 総括約1: 2022-03-09 11/03-47 展前分開: 2022-03-09 11/03-47 経験信意:		
	銀作: ②	2 8 A (##:		0 0 B A	操作:	C D B A	操作:	e 🛛	8

3. 在新建数据源对话框的关系型数据库区域,选择Impala。

如果您最近使用过Impala,也可以在**最近使用**区域选择Impala。同时,您也可以在快搜索框中,输入 Impala的关键词,快速筛选。



4. 在新建Impala数据源对话框中, 配置连接数据源参数后, 单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建Impala数据源	X
 • 数据源名称 请输入数据源名称 • 版本 3.x 数据源描述 请输入数据源描述 	
* 数据源配置 💿 "生产+开发"数据源	0/128 // ○ "生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
版本	仅支持选择1.x、2.x、3.x版本。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源	51 MV	开发数据源	
* JDBC URL	请输入JDBC连接地址,如jdbc:impala://hostport/schema	* JDBC URL	请输入JDBC连接地址,如jdbc:impala://host.port/schema
* Kerberos	 ● 开启 ● 关闭 	* Kerberos	 开启 关闭
* KDC Server	KDC统一服务地址,支持多项配置,英文分号分隔	* KDC Server	KDC统一服务地址,支持多项配置,英文分号分隔
切换为krb5文件配	置模式	切换为krb5文件配	置模式
* Keytab File	L 选择文件上传。	* Keytab File	出 选择文件上传
* Principal	Kerberos认证Principal名,如XXXX/hadoopclient@xxx.xxx	* Principal	。 Kerberos认证Principal名,如XXXX/hadoopclient@xxx.xxx

② 说明 通常,生产数据源和开发数据源配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述					
JDBC URL	链接地址的格式为 jdbc:impala//{服务器地址}:{端口}/{实例名称} 。例 如, jdbc:impala//10.0.0.1:5433/dataphin 。					
Kerberos	Kerberos是一种基于对称密钥技术的身份认证协议: Hadoop集群有Kerberos认证,则需要开启Kerberos。 Hadoop集群没有Kerberos认证,则无需开启Kerberos。 					
	开启Kerberos后,需要配置参数如下: Krb5文件/KDC Server:需要上传包含Kerberos认证域名的Krb5文件、配置KDC服务器地址,辅助完成Kerberos认证。					
Krb5文件/KDC Server、Keytab File、Principal	⑦ 说明 支持配置多个KDC Server服务地址,使用英文逗号(,)分割。					
	 Keytab File:上传登录Krb5文件域名或KDC服务器地址的账号和密码的文件。 Principal:配置Keytab File文件对应的Kerberos认证用户名。 					
用户名、密码	如果您没有开启 Kerberos ,则需要配置访问Impala实例的用户名和密码。					

iii. 配置数据源元数据库参数。

元数据库配置		元数据库配置			
* 数据库类型	请选择元数据库类型 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	* 数据库类型	请选择元数据库类型		
* JDBC URL	请输入JDBC连接地址,如jdbc:mysql://host:port/dbname	* JDBC URL	请输入JDBC连接地址,如jdb	c:mysql://host:port/db	oname
* 用户名	请输入鉴权用户名,为保证任务正常执行,请确保有所需数据权限	* 用户名	请输入鉴权用户名,为保证任	务正常执行, 请确保积	有所需数据权限
* 密码 * Ami	清输入盛权密码,为保证任务正常执行,请确保有所需数据权限 Ø	* 密码	"请输入鉴权密码,为保证任务	正常执行,请确保有所	所需数据权限 ø
			测试连接	取消	确定

参数	描述			
数据库类型	根据集群中使用的元数据库类型,选择数据库类型。Dataphin支持选择MySQL 和PostgreSQL。			
JDBC URL	<pre>填写对应元数据库的链接地址: MySQL: 格式为 jdbc:mysql://{链接地址}[,failoverhost]{端口 }/{数据库名称} [?propertyName1][=propertyValue1][&propertyNa me2][=propertyValue2] PostgreSQL: 格式为 jdbc:postgresql://{链接地址}:{端口}/{数据库 名称} 。</pre>			
用户名、密码	填写登录元数据库的用户名和密码。			

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。
 如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成Vertica数据源的创建。

后续步骤

完成Impala数据源的创建后,即可将Impala的业务数据引入至Dataphin或将Dataphin的数据写入至Impala。 具体操作,请参见<mark>配置离线单条管道</mark>。

6.9. NoSQL数据源

6.9.1. 创建Elasticsearch数据源

在将Dat aphin数据导出至Elast icsearch场景中, 您需要先创建Elast icsearch数据源。本文为您介绍如何创建 Elast icsearch类型的数据源。

背景信息

如果您使用的是阿里云Elasticsearch,在将Dataphin数据导出至Elasticsearch场景中,您需要先完成 Elasticsearch数据源的创建。更多阿里云Elasticsearch、开源Elasticsearch信息,请参见什么是阿里云 Elasticsearch。

Dataphin支持通过VPC网络和公网连接阿里云Elasticsearch、支持通过公网连接非阿里云Elasticsearch。 Dataphin连接阿里云Elasticsearch时,可以选择阿里云Elasticsearch在VPC网络或公网环境下对应的连接地 址。通常,VPC网络较公网稳定,建议您选择VPC网络连通Dataphin和Elasticsearch。 为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

● 当业务数据源(Elasticsearch)使用VPC网络时,则仅支持创建与Dataphin实例在同一地域Elasticsearch 数据源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。阿里云 Elasticsearch如何添加白名单,请参见配置实例公网或私网访问白名单。

• 公网数据源需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1 (深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

• 阿里云Elast icsearch且网络为VPC,则添加Dat aphin IP地址至数据库白名单中。VPC网络下Dat aphin的IP 地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建Elasticsearch数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 规划								* * < < 🖉 🌔
亦容	数据源 (190) ①						Q 请输入检查关键字	V + #1880913
A 数据采购	guzhi oss 🛛				Ifs lindorm 🕑			
₩ 公共定义	賞型: OSS				简型: Lindorm			10
項目	生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for Dev ①
□ 项目管理	文资人: Hand the Number of States							
≪ 计算题	飲須須用途: 東坡 台建用户:		教探測用途: 東波 台建用户:		設備運用途: 高統 创建用户:		数据源用途: 高线 创建用户:	
a 🚹	()(間時)(第): 2022-03-09 20:27:15		()()()()()()()()()()()()()()()()()()()		())()()()()()()()()()()()()()()()()()(((행타))#: 2022-03-09 11:03:47	
e Burg	1019/101222235-11122527 1019/10月: ◆ 重新日時信息		田田市(同): 2022-03-1112/25:33 征快信息: ● 査督任任信息		1997年1月1日2022-03-09 11103-47 1月19日日		1999年1月1日:2022-05-09 1135-47 短接信息: ● 宣都任法信息	
	請作:	e e i 2	操作:	0 C C A	题作:	e e e A	操作:	C 1 1 A

3. 在新建数据源对话框的NoSQL区域,选择Elasticsearch。

如果您最近使用过Elasticsearch,也可以在**最近使用**区域选择Elasticsearch。同时,您也可以在快搜索 框中,输入Elasticsearch的关键词,快速筛选。

新建数据源			Q 请输入	搜索关键字 全部 > X
NoSQL			radminisu Superadminisu Supe	
	in the second se	in the second se	€	•
HBase 0.9.4	HBase 1.1.x	HBase 2.x	Elasticsearch	MongoDB
TS	jî			
Tablestore	Aliyun HBase			
				父 的后 时关

- 4. 在新建Elasticsearch数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建Elasticsearch数据	E源 ×
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据源配置 💿 "生产+开发"数据源 🤇	"生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源			
* ES URL	http://es-cn-7m 74dv.elasticsearch.aliyuncs.com:9	200	* ES URL	http://es-cn-7n	dv.elasticsearch.aliy	uncs.com:9200
* 用户名	dataphin		* 用户名	dataphin		
* 密码		₿	* 密码			Ð
				30012- P 3左+立	FID 232	14.

② 说明 通常,生产数据源和开发数据源配置参数值不同,但Dataphin也支持配置成相同的 参数值。

参数	描述
ES URL	Elasticsearch的连接地址,推荐您使用私网连接地址。格式为 http://{连接 地址}:{端口} 。 以阿里云Elasticsearch为例,您可以在Elasticsearch实例的详情页面,查看连接 地址和端口信息。
	Cametage Cametage
	ISBN - ISBN - EXCRECT overflow inter EXCRECT overflow inter ISBN - EXCRECT overflow inter EXCRECT overflow inter EXCRECT overflow inter ISBN - EXCRECT overflow inter EXCRECT overflow inter EXCRECT overflow inter ISBN - EXCRECT overflow inter EXCRECT overflow inter EXCRECT overflow inter ISBN - ISBN - ISBN - ISBN - EXCRECT overflow inter
用户名、密码	访问Elasticsearch实例的用户名和密码。

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成Elasticsearch数据源的创建。

后续步骤

完成Elast icsearch数据源的创建后,即可在将Dat aphin数据导出至Elast icsearch。具体操作,请参见配置离 线单条管道。

6.9.2. 创建MongoDB数据源

创建MongoDB数据源用于实现Dataphin能够读取MongoDB的业务数据。在引入MongoDB的业务数据至 Dataphin中或将Dataphin数据导出至MongoDB的场景中,您需要先创建MongoDB数据源。本文为您介绍如 何创建MongoDB类型的数据源。

背景信息

如果您使用的是MongoDB,在引入MongoDB的业务数据至Dataphin中或将Dataphin数据导出至MongoDB的场景中,您需要先完成MongoDB数据源的创建。更多阿里云MongoDB信息,请参见什么是云数据库 MongoDB版。 Dataphin支持通过VPC网络和公网连接阿里云MongoDB、支持通过公网连接非阿里云MongoDB。Dataphin 连接阿里云MongoDB时,可以选择阿里云MongoDB在VPC网络或公网环境下对应的连接地址。通常,VPC网 络较公网稳定,建议您选择VPC网络连通Dataphin和MongoDB。

为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

 当业务数据源(MongoDB)使用VPC网络时,则仅支持创建与Dataphin实例在同一地域MongoDB数据 源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。阿里云MongoDB 如何添加白名单,请参见设置白名单。

• 公网数据源需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

• 阿里云MongoDB且网络为VPC,则添加Dataphin IP地址至数据库白名单中。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建MongoDB数据源

1. 请参见入口介绍,进入数仓规划页面。

2. 按照下图指引,进入新建数据源页面。

								· 🏕 🕫 🍳 🧕
亦皆	数据源 (190) ①						Q 清输入继续关键字	V + arabicana
▲ 数据架构 ₩ 公共建义	guzhi_oss ⊠ ≋≣: OSS			ē	lfs_lindorm ⊠ 简显: Undorm			ō
項日	生产	for Prod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for Dev ①
□ 项目管理 ● 項目管理 ● 計算液 ● 取扱服業	(公元人: 約第三月二: 東岐 公司用戸: 4回町町: 2022-03-09 02.7:15 周時町町: 2022-03-11 12:25:27 4回町雪日: ● 面面価単元目		(公司人 数据単示法: 表紙 公理用示: (公司用示: (2022.05-10 20:30.32) 面前可示: 2022.05-10 12.25:33 任時知道: (第二章) 変通信用工具		(文武人: 秋陽明元: 高岐 양평(河), 2022 8540 11:03-47 周期(河), 2022 8540 11:03-47 國際(河), 2022 8540 11:03-47 國際(河), 2022 8540 11:03-47		気気人: ***********************************	
	操作:	C 1 1 A	操作 :	C D D A	操作:	0 1 1 4	(第)下:	0 D B A

3. 在新建数据源对话框的NoSQL区域,选择MongoDB。

如果您最近使用过MongoDB,也可以在**最近使用**区域选择MongoDB。同时,您也可以在快搜索框中, 输入MongoDB的关键词,快速筛选。

新建数据源			Q 请输入报	P索关键字 全部 V X
NoSQL				and ^{a sec} and and ^{a sec} and and
	i des	in the second se	€	•
HBase 0.9.4	HBase 1.1.x	HBase 2.x	Elasticsearch	MongoDB
TS	JÎL.			
Tablestore	Aliyun HBase			
				关闭。

- 4. 在新建MongoDB数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 > 新建MongoDB数据源	φ X
*数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据源配置 🥑 "生产+开发"数据源	○ "生产"数据源
参数	描述
数据源名称	命名规则如下: 只能包含字母、数字或下划线(_)。 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源			
* JDBC URL	mongodb://dds-b 1174-pub.mongodb.rds.aliyuncs	s.con	* JDBC URL	mongodb://dds-bp	74-pub.mongodb	.rds.aliyuncs.com
* 用户名	dataphin		* 用户名	dataphin		
* 密码		₽	* 密码			₽
				测试连接	取消	确定

⑦ 说明 通常,生产数据源和开发数据源配置参数值不同,但Dataphin也支持配置成相同的 参数值。

参数	描述			
JDBC URL	推荐您使用私网连接批 };{连接地址1}:{端口 以阿里云MongoDB为 连接地址和端口信息。 ^{建接信息(Connection String URI)} ^{用编集量} ^{建新网络} ¹ ¹ ¹ ¹ ¹ ¹ ¹ ¹ ¹ ¹	地址。连接地址格式为 11 }。 例,您可以在MongoD ^{約2} (注: ****簡為7motition, 7x6階級町の成功, またまま morgada /hoot ****@dds_u887	mongodb://{ 连接地址 1}:{ B控制台的实例基本信息页面, Mongod6EdWEXENTEDTATIONOGE EnvireSt,) Magad6EdWEXENTEDTATIONOGE EnvireSt,)	端口1 查看 ²⁰⁰
	Primary	108	dds-uf6= t mongodb rds allyuncs.com	3717
用户名、密码	访问MongoDB实例的	用户名和密码。		

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成MongoDB数据源的创建。

后续步骤

完成MongoDB数据源的创建后,即可引入MongoDB的业务数据至Dataphin中或将Dataphin数据导出至 MongoDB。具体操作,请参见配置离线单条管道。

6.9.3. 创建Tablestore数据源

创建Tablestore数据源用于实现Dataphin能够读取Tablestore的业务数据。在引入Tablestore的业务数据至 Dataphin和将Dataphin的数据写入至Tablestore的场景中,您需要先完成Tablestore数据源的创建。本文为 您介绍如何创建Tablestore类型的数据源。

背景信息

Tablestore即阿里云表格存储Tablestore。如果您使用的是阿里云Tablestore,在对接Dataphin进行数据开发或将Dataphin的数据写入至Tablestore的场景中,您需要先完成Tablestore数据源的创建。更多阿里云Tablestore信息,请参见什么是表格存储。

Dataphin支持通过VPC网络、经典网络和公网连接阿里云Tablestore。通常,VPC网络较公网、经典网络稳定,建议您选择VPC网络连通Dataphin和Tablestore。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

 当业务数据源(Tablestore)使用VPC网络时,则仅支持创建与Dataphin实例在同一地域Tablestore数据 源。

操作步骤

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 规划												2 🕓
<u>#</u> \$	数据源 (190) ①								Q 清箱入理发关键字		¥ 💽	+ aretecter
A 数据制约	quzhi oss 🖂					's lindorm ⊵						
₩ 公共定义	类型: OSS				8	토립: Lindorm						
项目	生 ^{pla}	for Prod / Basic ①	开发	for Dev ①		生产	for Proc	/ Basic ①	开发		for C	ev O
▶ 项目管理	交通人:					28A: Incolar Contact			SRA: Topo of the outpace of the			
411 it 1920	25項源用途: 素は (22用件: 1000/000000000000000000000000000000000		政策運用注: 支払 公理用 ^A : bunded and bundles (bundles)			数据原用途: 集始 (1读用户:			数据原用途: 本地 (127月月):			
a 🚺	())登时间: 2022-03-09 20:27:15 第250-101: 2022-03-09 11 12:25:27		()建計計詞: 2022-03-09 20:30:32 WREENIN: 2022-03-09 20:30:32			(1년9)1년: 2022-03-09 11:03:47 WRGHU: 2022-03-09 11:03:47			(1)思対词: 2022-03-09 11:03:47 WRGHU: 2022-03-09 11:03:47			
									· · · · · · · · · · · · · · · · · · ·			
	銀作 :	0 0 8 A	操作:	@ E E &		操作:	@ [3 B A	操作:	e	0 1	1 &

3. 在新建数据源对话框的NoSQL区域,选择Tablestore。

如果您最近使用过Tablestore,也可以在**最近使用**区域选择Tablestore。同时,您也可以在快搜索框中,输入Tablestore的关键词,快速筛选。

新建数据源			Q 请输入	搜索关键字 全部 🗸 🗙
NoSQL				
			€	•
HBase 0.9.4	HBase 1.1.x	HBase 2.x	Elasticsearch	MongoDB
TS	JL			
Tablestore	Aliyun HBase			
				关闭。如何

4. 在新建Tablestore数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建Tablestore数据	ля X
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128
* 数据源配置 🕒 "生产+开发"数据源	○ *生产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

生产数据源			开发数据源				
* endPoint	https:// -beijing.vpc.tablestore.aliyuncs.com		* endPoint	https://	beijing.vpc.tables	store.aliyuncs.co	om
* Access Id	dataphin		* Access Id	dataphin			
* Access Key		₿	* Access Key				₽
* Instance Name	-mc		* Instance Name	-mc			
				2011年2年4支	HD SM		74.
				测试进按	収消		佣正

⑦ **说明** 通常,生产数据源和开发数据源需要配置成不同的参数值,但Dataphin也支持配置成相同的参数值。

参数	描述
----	----

参数	描述					
Endpoint	即 案例访问地址。推荐您使用 VPC 网络对应的 案例访问地址。您可以 在 Tablestore控制台的 案例管理页面的 案例详情页 签, 查看 案例访问地址。					
Access Id、Access Key	当前账号的AccessKey ID和AccessKey Secret。 如何获取,请参见 <mark>获取AccessKey</mark> 。					
	即实例名称。您可以在Tablestore控制台的 全部实例 页面,查看实例名称。					
	Tablestore	首页 / 全部实例				
	概选	全部实例				
Instance Name	全部实例					
	审计日志	华东2(上海)				
	权威指南	实例名称				
	最佳实践	-doc				

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。 6.测试成功后,单击**确定**,完成Tablestore数据源的创建。

后续步骤

完成Tablestore数据源的创建后,即可引入Tablestore的业务数据至Dataphin或将Dataphin的数据写入至 Tablestore。具体操作,请参见配置离线单条管道。

6.9.4. 创建Aliyun HBase数据源

创建Aliyun HBase数据源用于实现Dat aphin能够读取Aliyun HBase的业务数据。在开发流批一体任务的场景中,如果业务数据来源于Aliyun HBase,则在您开始研发数据前需要先完成Aliyun HBase数据源的创建。本文为您介绍如何创建Aliyun HBase数据源。

背景信息

Aliyun HBase即阿里云的云数据库HBase。如果您使用的是云数据库HBase,在对接Dataphin进行数据开发时,您需要先完成Aliyun HBase数据源的创建。更多云数据库HBase信息,请参见什么是云数据库HBase。

Dat aphin支持对接标准版Aliyun HBase和增强版Aliyun HBase,且均支持通过VPC网络和公网连接。通常,VPC网络较公网稳定,建议您选择VPC网络连通Dat aphin和Aliyun HBase。

为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin不同网络环境下的IP地址段添加到数据 源的白名单中。

使用限制

仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

● 当业务数据源(Aliyun HBase)使用VPC网络时,则仅支持创建与Dataphin实例在同一地域Aliyun HBase数据源。

步骤一:添加Dataphin IP地址至数据库白名单

基于业务数据源所在的网络环境,添加对应环境下的Dataphin IP地址至数据库白名单中。如何添加白名单, 请参见增强版Aliyun HBase设置白名单、标准版Aliyun HBase设置白名单。

• 公网的数据源需要添加公网下Dataphin IP地址至数据库白名单中。公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● VPC网络的数据源需要添加VPC网络下的Dataphin IP地址至数据库白名单中。VPC网络下Dataphin的IP地址

地域	(推荐)通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建Aliyun HBase数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

								· 삼· 오 🕗 🌔
#8	数据源(190)①						Q 请输入继续关键字	V + \$532,003630
▲ 数据架构 ₩ 公共定义	guzhi_oss 四 简副: OSS			ē	Hs_lindorm ⊠ 해줬는 Lindorm			Ċ.
项目	生产 for P	Yod / Basic ①	开发	for Dev ①	生产	for Prod / Basic ①	开发	for Dev ①
□ 项目管理 	○思人: 約須易点: 金融市で、2022-04-09 327 15 面前市市、2022-04-09 327 15 面前市市、2022-04-19 1225 27 489/8号: ● ■ 面積価単価番		久原人 外球長時に、取枝 台湾市市:2022-05-09 20 30 32 直新市市:2022-05-01 12 22 533 住田市市: 2022-05-11 12 25 33 住田市市: ● 面面は日本田島		(文武人: 東波 원(波明규): 東波 원(波明규): 2022-03-09 11:03-47 표현(지규): 2022-03-09 11:03-47 변화(제금): (소 고려)(태종)-47 변화(제금): (소 고려)(태종)-47		 (2) (2) (2) (2) (2) (2) (2) (2) (2) (2)	
	操作: 《	6 6 8	操作:	C I I A	探行 :	0 1 1 4	(第7年:	e 12 8

3. 在新建数据源对话框的NoSQL区域,选择Aliyun HBase。

如果您最近使用过Aliyun HBase,也可以在**最近使用**区域选择Aliyun HBase。同时,您也可以在快搜索 框中,输入Aliyun HBase的关键词,快速筛选。

新建数据源			Q 请输入打	史索关键字 全部 ~ X
NoSQL				
	in the second se	the second	€	•
HBase 0.9.4	HBase 1.1.x	HBase 2.x	Elasticsearch	MongoDB
TS	JÎL.			
Tablestore	Aliyun HBase			
				关闭。

- 4. 在新建Aliyun HBase数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

返回新建数据源 >	> 新建Aliyun HBase数据源	Х
* 数据源名称	请输入数距源名称	
数据源描述	请输入数据源描述	0/128
* 数据源配置	● *生产+开发*数据源 ○ *生产*数据源	

参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。 ■ 如果您的阿里云HBase实例的服务类型为标准型,则选择为标准型。

生产数据源		开发数据源
* 服务类型 💿 标准型 🔵 増强型	1 ⁷⁵⁹ Super AdminiSe Super AdminiSe Super	* 服务类型 💿 标准型 🔷 増强型
*版本		*版本 · · · · · · · · · · · · · · · · · · ·
* zkQuorum 请输入包含端口的连接	也址	* zkQuorum 请输入包含端口的连接地址
		测试连接取消确定
参数	描述	
服务类型	选择为 标准型 。	
版本	选择阿里云HBase实例对 库连接页面,查看主版本 く 基本信息 <mark>数照库连接</mark> 主音双活 访问控制	应的主版本号。您可以在 <mark>阿里云Hbase控制台</mark> 的数据 号。 <u>实例 hb-bp 5000000000000000000000000000000000000</u>
zkQuorum	配置阿里云HBase实例的 以在阿里云Hbase控制台 (1 实例 hb- # 定時 # 表現存 # 表現存 # 表現存 # 表現存 # 表現存 (年期78 (年期78 (年期78 (年期78 (年期78 (年期78 (年間78)) (東明78) (東西78) (西78) (東西78) (東西78) (西78) (西78)	链接地址,推荐您使用专有云网络的链接地址。您可 的数据库连接页面,查看链接地址。 (HBasels)/组织。*### % //////////////////////////////////

■ 如果您的阿里云HBase实例的服务类型为增强型,则选择为增强型。

生产数据源			开发数据源			
* 服务类型 🔵 标准型 💿 増强型		* 服务类型 ○ 标准型 ● 増强型				
★ endPoint <host:port>,HBase增强</host:port>	版的Java API访问地址		* endPoint	<host:port>,HBase增强版的</host:port>	Java API访问地址	
* 用户名 HBase用户名			* 用户名	HBase用户名		
* 密码 HBase密码		₿	* 密码	HBase密码		8
				测试连接	取消	确定。
参数	描述					
服务类型	选择为 增强型 。					
Endpoint	选择阿里云HBaseS 据库连接页面,查看 ^{网络类型 专有网络} 默认用产名 root Java API访问地址 非 JAVA 诸盲 Thrift2 访问地址	文例对	立CQL访问地J 方问地址。 ^{同解助} Id-Eproxy 访问解助 http://Id-b http://Id-bp Id-bp1prox	社。您可以在阿里 vF hbaseue.hbaseue.rds.aliyuncs.co -hbaseue.pub.hbaseue.rds.aliyunc -proxy-hbaseue.hbaseue.rds.aliyun -proxy-hbaseue.hbaseue.rds.aliyun y-hbaseue.hbaseue.rds.aliyuncs.co	2CID vpc-bp1 以密码 root @ m:30020 一線生成配置 s.com:30020 一線生成配置 incs.com:9190 saliyuncs.com:9190 m:9042 cs.com:9042	<mark>創合的数</mark> ₫ ■■♥
用户名密码	当前账号的Accessl <mark>AccessKey</mark> 。	Key ID;	和AccessKey	Secret。如何获	取,请参见 <mark>貉</mark>	夹取

② 说明 通常,生产数据源和开发数据源需要配置成不同的参数值,但Dataphin也支持配置 成相同的参数值。

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。

如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通的常见问题。

6. 测试成功后,单击确定,完成Aliyun HBase数据源的创建。

后续步骤

完成Aliyun HBase数据源的创建后,即可开始开发流批一体任务。具体操作,请参见创建Flink SQL任务、创建Flink Dat aSt ream计算任务或步骤一:创建并调试Flink Template SQL计算任务。

6.9.5. 创建HBase数据源

创建HBase数据源用于实现Dataphin能够读取HBase的业务数据。在引入HBase的业务数据至Dataphin和将 Dataphin的数据写入至HBase的场景中,您需要先创建HBase数据源。本文为您介绍如何创建HBase类型的 数据源。

背景信息

HBase是用于处理存储在Hadoop集群中大量数据的SQL查询引擎。如果您使用的是HBase,在导出Dataphin 数据至HBase,您需要先完成HBase数据源的创建。更多HBase信息,请参见HBase官网。

Dataphin支持创建0.9.4版本的HBase数据源、1.1.x版本的HBase数据源和2.x版本的HBase数据源。Dataphin 需要通过公网连接HBase数据源。为避免因数据源白名单限制导致数据对接访问失败,您需将Dataphin在公网环境下的IP地址段添加到数据源的白名单中。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

• 仅支持创建0.9.4、1.1.x、2.x版本的HBase数据源。

步骤一:添加Dataphin IP地址至数据源白名单

添加公网环境下的Dataphin IP地址至数据源白名单后,再打通HDFS数据源及Hadoop集群(含NameNode和 DataNode)与Dataphin之间的网络,同时需保证端口间的也是互通的才能正常使用。

公网下Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

步骤二: 创建HBase数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 规划								凿 🗸 🖏 2	
不能	数据源 (190) ①						Q. 遗输入细胞关键字	7 + #	1941039539
 一数据制约 W 公共定义 	guzhi_oss ⊠ 类型: OSS			Ē	lfs_lindorm ⊵ ≋⊞: Lindorm				ē
項目	生) ^{der} for	Prod / Basic ①	开发	for Dev ①	生/**	for Prod / Basic ①	开发	for Dev	0
□ 项目管理	负责人:						SBA: Topo III topo III topo de		
 计算数 	数据源用途: 淘试 创建用户:		政策原用法: 教 成 公務用件: Insulational Type (Insulation)		数据原用途: 集线 创建用户: 计正确 计数 计数 计数 计数 计数		数据源用途: 激线 创锦用户:		
·····································	(1)勝行時) 2022-03-09 20:27:15 運動行時前) 2022-03-11 12:25:27 総約信意: ◆ 宣誓指決信意		公理时间:2022-03-09 20:30:32 更新时间:2022-03-11 12:25:33 税務保護: ● 宣告相決信意		11時时间:2022-03-09 11:03:47 東部时间:2022-03-09 11:03:47 規模信意: ● 宣告相接信品		创建时间:2022-03-09 11:03:47 更新时间:2022-03-09 11:03:47 链接信意:		
	銀作:	0088	操作:	0004	課作:	0 D B A	提作:	e 🛛 🖻	8

3. 在新建数据源对话框的NoSQL区域,选择HBase。

如果您最近使用过HBase,也可以在**最近使用**区域选择HBase。同时,您也可以在快搜索框中,输入 HBase的关键词,快速筛选。

新建数据源			Q 请输)	搜索关键字 全部 >
NoSQL				
			€	(
HBase 0.9.4	HBase 1.1.x	HBase 2.x	Elasticsearch	MongoDB
TS	jit			
Tablestore	Aliyun HBase			

- 4. 在新建HBase数据源对话框中,配置连接数据源参数后,单击确定。
 - i. 配置数据源的基本信息。

* 数据源名称 HBase	
数据源描述 清输入数据源描述	0/128 /
* 数据源配置 💿 "生产+开发"数据源 🗌 "	主产"数据源
参数	描述
数据源名称	命名规则如下: ■ 只能包含字母、数字或下划线(_)。 ■ 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

数仓规划·数据源

生产数据源		开发数据源				
* 连接地址	连接用 hbase.zookeeper.quorum 信息,支持多节点,如: host01;h	* 连接地址	连接用 hbase.zookeepe	r.quorum 信息,支持领	š节点,如: h	ost01;h
Namespace	请输入namespace	Namespace	请输入namespace			
配置文件	出选择文件上传。	配置文件	土 选择文件上传			
连接参数	⁷⁷⁷ 连接HBase参数,请使用JSON格式形式配置 0/128 //	连接参数	连接HBase参数,请使用J	SON格式形式配置		0/128 ^{Supe}
* Kerberos	 开启 〇 关闭 	* Kerberos	● 开启 ○ 关闭			
* KDC Server	KDC统一服务地址,支持多项配置,英文分号分隔	* KDC Server	KDC统一服务地址,支持	移项配置,英文分 ⁴	分隔	
切换为krb5文件配	置模式	切换为krb5文件配	置模式			
* Keytab File	当 选择文件上传	* Keytab File	土 选择文件上传			
* Principal	Kerberos认证Principal名,如XXXX/hadoopclient@xxxxxxx	* Principal	≪Kerberosi入让Principal名	,如XXXX/hadoopc	lient@xxx.xx	x
			测试连接	取消		确定

② 说明 通常,生产数据源和开发数据源需要配置的参数值不同,但Dataphin也支持配置成相同的参数值。

参数	描述
连接地址	连接地址的格式为 hb-proxy-{ 实例地址 }-{端口}.hbase.rds.aliyuncs.com 。
Namespace	HBase系统默认定义了两个缺省的Namespace,包括hbase和default。同时您也可以登录HBase主节点执行命令查看。
配置文件	上传keytab文件,您可以在Hadoop集群中的NameNode节点使用ipa-getkeytab命令获 取keytab文件。
连接参数	配置连接HBase的参数,参数的格式为JSON样式。
Kerberos	Kerberos是一种基于对称密钥技术的身份认证协议,可以为其他服务提供身份认证功能, 且支持SSO(即客户端身份认证后,可以访问多个服务,例如HBase和HDFS): 如果Hadoop集群有Kerberos认证,则需要开启Kerberos。 如果Hadoop集群没有Kerberos认证,则不需要开启Kerberos。
Krb5认证文 件/KDC Server地	开启Kerberos后,则需要配置如下参数: Find Krb5认证文件/KDC Server地址:需要上传包含Kerberos认证域名的Krb5文件、配置KDC服务器地址,辅助完成Kerberos认证。
址、Keytab FilePrincipa	⑦ 说明 支持配置多个KDC Server服务地址,使用英文逗号(,)分割。
l	 Keytab File:上传登录Krb5文件域名或KDC服务器地址的账号和密码的文件。 Principal:配置Keytab File文件对应的Kerberos认证用户名。

5. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。
 如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通解决方案。

6. 测试成功后,单击**确定**,完成HBase数据源的创建。

后续步骤

完成HBase数据源的创建后,即可将HBase的业务数据引入至Dat aphin或将Dat aphin的数据写入至HBase。 具体操作,请参见配置离线单条管道。

6.9.6. 创建ClickHouse数据源

创建ClickHouse数据源用于实现Dataphin能够读取ClickHouse的业务数据,及能够向ClickHouse中写入数据。在引入ClickHouse的业务数据至Dataphin中或将Dataphin数据导出至ClickHouse的场景中,您需要先创建ClickHouse数据源。本文为您介绍如何创建ClickHouse数据源。

背景信息

如果您使用的是数据库ClickHouse,引入ClickHouse的业务数据至Dataphin中或将Dataphin数据导出至 ClickHouse的场景中,您需要先完成ClickHouse数据源的创建。

Dataphin支持通过VPC网络和公网连接阿里云数据库ClickHouse、支持通过公网连接非阿里云数据库 ClickHouse。Dataphin连接阿里云数据库ClickHouse时,可以选择阿里云数据库ClickHouse在VPC网络或公网 环境下对应的连接地址。通常,VPC网络较公网稳定,推荐您选择VPC网络连通Dataphin和阿里云数据库 ClickHouse。

阿里云数据库ClickHouse的更多信息,请参见什么是云数据库ClickHouse。

使用限制

• 仅支持超级管理员和项目管理员角色创建数据源。

如何添加项目成员并授予项目管理员角色,请参见添加项目成员。

 当业务数据源(ClickHouse)使用VPC网络时,则仅支持创建与Dataphin实例在同一地域ClickHouse数据 源。

步骤一:添加Dataphin IP地址至数据源白名单

为避免因数据源白名单限制导致数据对接访问失败,基于业务数据源所在的网络环境,需添加对应环境下的 Dataphin IP地址至ClickHouse数据库白名单中。阿里云数据库ClickHouse如何添加白名单,请参见设置白名 单。

 ClickHouse数据源的网络为公网时,则需要添加公网下Dataphin IP地址至数据源白名单中。公网下 Dataphin的IP地址

地域	IP地址
华东2(上海)	47.102.151.182
华南1(深圳)	119.23.173.65
华北2(北京)	123.56.104.202
华东1(杭州)	47.96.75.100

● ClickHouse数据源的网络为VPC时,则需要添加VPC网络下的Dataphin IP地址至数据源白名单中。VPC网络 下Dataphin的ⅠP地址

地域	(推荐) 通用IP地址	地域IP地址
华东2(上海)		100.104.228.128/26、 100.104.115.192/26
华南1(深圳)	100.104.0.0/16	100.104.48.128/26
华北2(北京)		100.104.238.64/26
华东1(杭州)		100.104.45.0/26

步骤二: 创建ClickHouse数据源

- 1. 请参见入口介绍,进入数仓规划页面。
- 2. 按照下图指引,进入新建数据源页面。

■ Dataphin 规划								· 🛎 🗳 ୍ 🙆 🧕
<u>₩</u> \$	数据源 (190) ①						Q 请输入提供关键字	1 + afittation
☆ 数据架构 ₩ 公共定义	guzhi_oss ⊠ 类型: oss			Ē	Hs_lindorm ⊠ 武型: Lindorm			÷
项目	生产	for Prod / Basic ①	开发	for Dev ①	生/**	for Prod / Basic ①	开发	for Dev ①
□ 项目管理 411 计算数 ◎ 11 ■ 数据数	 ①美八: 記道房間: (2道月中:) (2道月中:) (2道日): (2道日): (2道日): (2道日): (2): (3): (4): 		交通人 日本		公長人: 設備原用: (使用用: (使用用: 202243-00 110347 展新分词:202243-00 110347 展新分词:202243-00 110347 最新分词:202243-00 110347		交通人: 数据写用途: 集成 必認可作: 公認知作: 2022-03-00 11 03-07 単新行用: 2022-03-00 11 03-07 単新作用: 2022-03-00 11 03-07 単新作用: 企業 (1) 0.47	
	操作:	e d 🗉 🔒	操作:	e e e A	操作:	@ 12 B	現在:	C 12 B A

3. 在新建数据源对话框的NoSQL区域,选择ClickHouse。

如果您最近使用过ClickHouse,也可以在**最近使用**区域选择ClickHouse。同时,您也可以在快搜索框中,输入ClickHouse的关键词,快速筛选。

新建数据源			Q. 请输入	搜索关键字 全部 ∨ X
NoSQL				
			G	•
HBase 0.9.4	HBase 1.1.x	HBase 2.x	Elasticsearch	MongoDB
TS	JÎL.			
Tablestore	Aliyun HBase	ClickHouse		
				关闭

4. 在新建ClickHouse数据源对话框中,配置连接数据源参数后,单击确定。

i. 配置数据源的基本信息。

返回新建数据源 > 新建ClickHouse数据	源 X
* 数据源名称 请输入数据源名称	
数据源描述 请输入数据源描述	0/128 /
* 数据源配置 💿 "生产+开发"数据源	○"生产"数据源
参数	描述
数据源名称	命名规则如下: 只能包含字母、数字或下划线(_)。 长度不能超过64字符。
数据源描述	对数据源的简单描述。
数据源配置	选择需要配置的数据源: 如果业务数据源区分生产数据源和开发数据源,则选择生产+开发数据源。 如果业务数据源不区分生产数据源和开发数据源,则选择生产数据源。

ii. 配置数据源与Dataphin的连接参数。

当上述步骤中选择了**生产+开发数据源**,则配置页面如下图所示。如果上述步骤中您选择了**生产数** 据源,则仅展示生产数据源的配置页面。

输入JDBC连接地址,如jdbc:clickhouse://host.port/dbname 输入鉴权用户名,为保证任务正常执行,请确保有所需数据权用		* JDBC URL	请输入JDBC连接地址, 如jdbc:clickhouse://host.port/dbname
输入鉴权用户名,为保证任务正常执行,请确保有所需数据权限			
		* 用户名	请输入鉴权用户名,为保证任务正常执行,请确保有所需数据权限
输入鉴权密码,为保证任务正常执行,请确保有所需数据…	₿	* 密码	请输入鉴权密码,为保证任务正常执行,请确保有所需数据
公网数据库 💿 阿里云数据库 🔵 ECS(VPC)自建数据源	ŧ	* 类型	○ 公网数据库 ● 阿里云数据库 ○ ECS(VPC)自建数据源
VPC网络 🥑 公网IP 🕬		* 访问方式	○ VPC网络 ● 公网IP
需要开放访问,并授权权限。如需授权白名单,请添加Dataphir 47.102.151.182	n默	本地自建数据库授权认资源组的公网出口	权,需要开放访问,并授权权限。如需授权白名单,请添加Dataphin默 网IP:47.102.151.182
			测试连接 取消 确定
70	输入鉴权密码,为保证任务正常执行,请确保有所需数据… 公网数据库 ④ 阿里云数据库 〇 ECS(VPC)自建数据源 VPC网络 ④ 公网IP。 需要开放访问,并授权权限。如需授权白名单,请添加Dataphi : 47.102.151.182	输入鉴权恶锅,为保证任务正常执行,请确保有所需数据。	输入鉴权宽码,为保证任务证常执行,请确保有所需数据

成相同的参数值。

参数

参数	描述
JDBC URL	ClickHouse的连接地址,格式为 jdbc:clickhouse://<连接地址>:<端口>/<数 据库名称> 。推荐您使用VPC连接地址。 以阿里云数据库ClickHouse为例,参数获取方式如下。 • 连接地址、端口:在ClickHouse为例,参数获取方式如下。 • 连接地址、端口:在ClickHouse控制台对应的集群信息页面,查看连接地址和端口信息。 • 数据库名称:通过ClickHouse命令行工具连接集群后,在SQL Console中执行 s how Databases 命令,查看数据库名称。连接集群操作指导,请参见连接集 群。 B. 在收载接地址端口 <u>* ***********************************</u>
用户名和密码	ClickHouse实例的用户名和密码,详情请参见 <mark>账号管理</mark> 。
类型	Dataphin支持选择以下几种数据源类型: 公网数据库:如果您的数据源具备公网访问能力,则可以选择公网数据库。 阿里云数据库:如果您的数据源是阿里云数据库RDS,请选择阿里云数据库。阿 里云数据库支持通过公网IP或VPC网络连接Dataphin。 ⑦ 说明 如果您的数据源是阿里云数据库RDS且支持公网访问,则您可以 选择公网数据库,也可以选择阿里云数据库(公网IP)。 ECS (VPC)自建数据源:如果您的数据源是在阿里云ECS上自建的数据库,且 使用的网络环境为VPC,请选择ECS (VPC)自建数据源。
	数据源类型选择不同,则访问方式不同。 • 当类型选择为阿里云数据库时: • VPC网络 • 区域:仅支持选择当前Dataphin实例所在的地域。 ⑦ 说明 仅支持添加与Dataphin实例在同一地域的ClickHouse数据 源至Dataphin。 • VPC ID:填写阿里云数据库ClickHouse的VPC ID。 • RDS ID:填写阿里云数据库ClickHouse的集群ID。 您可以在ClickHouse控制合对应的集群信息页面,查看VPC ID(图示②)、集 群D(图示①)。

参数	■□突空 描述	选择为ECS(VPC)	目建数据源	息时:		
	■ 区域	: 仅支持选择当前D	ataphin实例	列所在的地	也域。	
) 说明 仅支持添 ¡Dataphin。	加与Dataph	nin实例在	同一地域的Click	House数据源
	- 1/20					
	■ VPC	ID: 填与云服务器		「有网络ID	0	
	■ ECS	ID: 項与 加介 希日	CS的实例ID	0		
	您可以 例ID(在云服务器ECS控制; 图示②)。	<mark>台</mark> 的实例详 ^情	青页面,香	昏看VPC ID(图示	<①)和ECS的
	launch-	dvisor 20211108zx 🗹 🤮	运行中			
	实例ID	i-uf6cf	远程连接	地域	华纬 【11年)	
访问方式	资源组:	-	转通为确性公网IP	所在可用区 主机名	上海 可用区G	修改定例主相文
	安全组	sg-uf6	加入安全组	创建时间	2021年11月8日 14:34:00	INCONT.
	描述	2021 - Dilix	修改实例描述	自动释放时间	÷	释放设置
	CPU&内存	2核(vCPU) 8 GiB		云盘	1	重新初始化云盘
	操作系统	Alibaba Cloud Linux 3.2104 64位	更换操作系统	快照	0	
	实例规格	ecs.s6-c1m4.large	更改实例规格	镜像ID	aliyun_3_x64_20G_alibase_20	210 创建自定义镜像
	实例规格族	ecs.s6		当前使用带宽	10Mbps (峰值)	按量付费实例更改带竞
	标签					编辑标签
	网络广告					
	网络百元尽				(485)	管辅助弹性网卡 更换专有网
	网络米刑	去有网络		去有网络	VDC-11	2
	网络迷刑	专有网络	1	专有网络	voc-ul 1071	2
	网络类型 弹性网卡	专有网络 eni-u ot	1	专有网络 虚拟交换机	vpc-u ia7r [vsw-u ia7r [3
	网络类型	专有网络 eni-u ot	1	专有网络 虚拟交换机	vpc-u ia7r [vsw-u j îve [3 3
	阿维类型 弹性码卡	专有网络 eni-t	1	专有网络 虚拟交换机	vpc-ut 1a7r [vsw-u five]	a 2
	阿維美型 弹性阿卡	专有网络 eni-u	1	专有网络 虚拟交换机	vpc-ut a7r [vsw-ut tre [a a
	网络类型 弹性网卡	专有网络 eni-u	1	专有网络 唐拟交换机	vpc-ut ta7r [vsw-ut two]	a 3
	阿维类型 弹性码卡	专有网络 eni-t	1	专有网络 虚拟交换机	vpc-utar a fr [vsw-utar to the file of th	a a
	阿维类型 弹性网卡	专有网络 eni-u	1	专有网络 虚拟交换机	vpc-ut a7r [vsw-u	a a
	网络类型 弹性网卡	专有网络 eni-u	1	专有网络 虚拟交换机	vpc-ut ta7r [vsw-u trive]	a 2
	阿维类型 弹性码卡	专有网络 eni-t	1	专有网络 虚拟交换机	vpc-ut a7r [vsw-ut tre [a a
	网络类型 弹性网卡	专有网络 eni-u	•	专有网络 虚拟交换机	vpc-ut ia7r [vsw-u five]	2
	阿维类型 弹性码卡	专有网络 eni-t	1	专有网络 虚拟交换机	vpc-ut a7r [vsw-ut tre [a a
	网络类型 弹性网卡	专有网络 eni-t	1	专有网络 _虚拟交换机	vpc-ut ta7r [vsw-u trve]	a a
	阿培类型 弹性网卡	专有网络	1	专有网络 虚拟交换机	vpc-ut a7r [vsw-ut tre [a a
	网络类型 弹性网卡	专有网络 eni-t	1	专有网络 虚拟交换机	vpc-ut a7r [vsw-ut tre [a a
	阿培类型 弹性码卡	专有网络 eni-u	1	专有网络 虚拟交换机	vpc-u ¹ la7r [vsw-u ⁻ l ¹ /v [
	网络类型 弹性网 卡	专有网络 eni-t	1	专有网络 虚拟交换机	vpc-ut a7r [vsw-ut tre [a a
	网络类型 弹性网卡	专有詞語 eni-t	1	专有网络 	vpc-ut arr [vsw-ut the first of the first o	
	网络类型 弹性码卡	专有网络 eni-t		专有网络 通机交换机	vpc-ut ta7r [vsw-ut tre]	
	网络类型 弹性网 卡	ぞ有网络	1	专有网络 	vpc-ut a7r [vsw-ut tre [

- 9. 单击测试连接,测试数据源是否可以和Dataphin进行正常的连通。
 如果连接测试失败,您可以根据网络连通常见问题进行排查。详细内容,请参见网络连通解决方案。
- 6. 测试成功后,单击**确定**,完成ClickHouse数据源的创建。

后续步骤

完成ClickHouse数据源的创建后,即可将ClickHouse业务数据导入至Dataphin,或将Dataphin数据导出至ClickHouse。具体操作,请参见配置离线单条管道。