



数据接入工具用户使用手册

版本:3.0

南京烽火星空通信发展有限公司

2022年3月

目录

1	介绍	3
1.1	文档目的	3
1.2	文档有效性	3
1.3	术语和缩写	3
2	软件综述	4
2.1	产品简介	4
3	操作说明	4
3.1	登录工具	4
3.2	管理中心主界面	5
3.3	数据源	6
3.3.1	数据源分类	6
3.3.2	新增数据源	7
3.3.3	新增数据集	16
3.3.4	数据集复制	39
3.3.5	数据源明细	40
3.3.6	编辑数据源	42
3.4	任务管理	43
3.4.1	任务列表	43
3.4.2	新建任务	45
3.4.3	任务运转	56
3.4.4	批量导入、导出任务和删除任务	61
3.5	台账	61
3.5.1	台账列表	61
3.5.2	台账明细	62
3.5.3	台账导出	63
3.6	节点管理	63
3.6.1	添加节点	63
3.6.2	节点列表	65
3.6.3	节点查询	65
3.6.4	节点的操作	66
3.6.5	编辑节点	67
3.6.6	节点详情	67
3.7	监控中心	68
3.7.1	任务监控	69
3.7.2	事件通知	69
3.7.3	节点监控	70
3.8	系统配置	71
3.8.1	配置系统参数	71
3.9	统计分析	73

1 介绍

1.1 文档目的

本文档主要介绍 FitDataPrep 工具主要功能和操作步骤,帮助用户了解运用本工具。

1.2 文档有效性

本文档主要适用于以下人员

- 技术支持工程师
- 维护工程人员
- 产品最终使用者
- 其他需要了解本产品具体功能使用的人员
- 本文档适用于 FDP3.0.0 及以后的版本

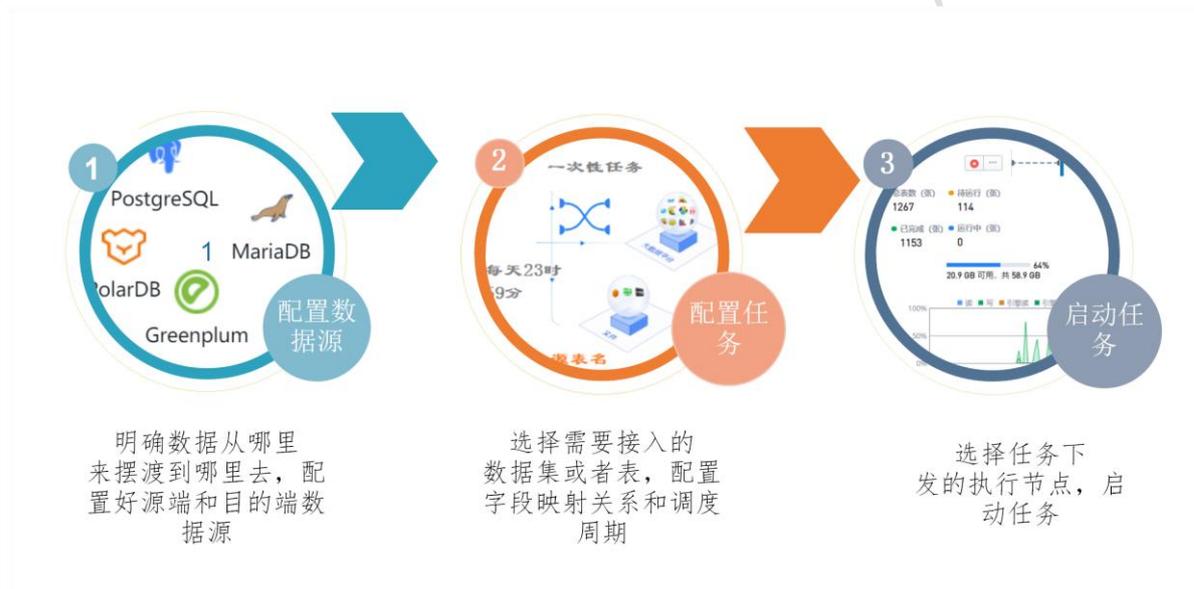
1.3 术语和缩写

FDP FitDataPrep 数据接入工具

2 软件综述

2.1 产品简介

烽火数据接入工具(FitDataPrep 以下简称 FDP)是一款即开即用的企业级一站式快速接入产品。核心功能涵盖接入准备、接入执行、接入监控、接入审计全过程。为用户提供高性能、高可靠、人性化的数据接入与管理服务。通过本工具通过如下三步操作即可完成常见数据接入需求场景的配置。

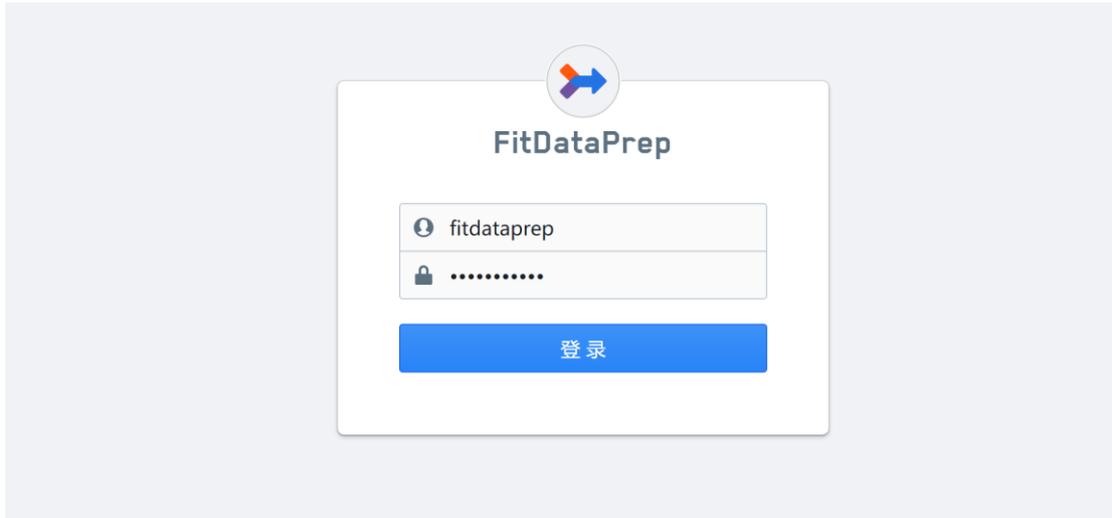


FitDataPrep 工具由管理中心和任务执行节点两个部分组成。工具安装完成后，用户主要与管理中心进行交互。通过管理中心进行数据接入任务创建、管理，实现对数据迁移、汇聚和分发等需求。本文档主要介绍管理中心的使用，包含数据源管理、任务管理、台账管理、监控中心、系统配置、节点管理、统计分析、派大星等多个功能模块。

3 操作说明

3.1 登录工具

浏览器输入工具主界面地址 [http://\[工具管理中心地址\]:8060/login](http://[工具管理中心地址]:8060/login)，打开登录界面。输入工具认证用户名密码点击登录按钮进入工具管理中心主界面。



3.2 管理中心主界面

下图为管理中心主界面



管理中心主界面左上角为工具 logo 图标，见下图。



管理中心所有模块的入口按钮都依次横向排列在工具主界面上方，见下图。



管理中心右上角为工具帮助文档入口，点击可在线打开工具使用帮助文档，见下图。



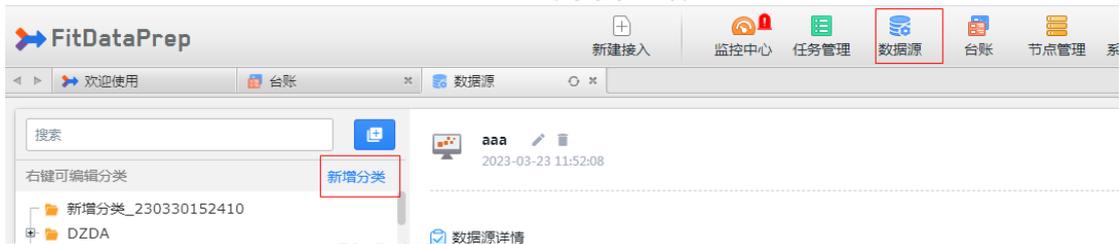
管理中心右上角展示当前登录系统的用户，点击可下拉展示注销按钮，实现注销功能，见下图。



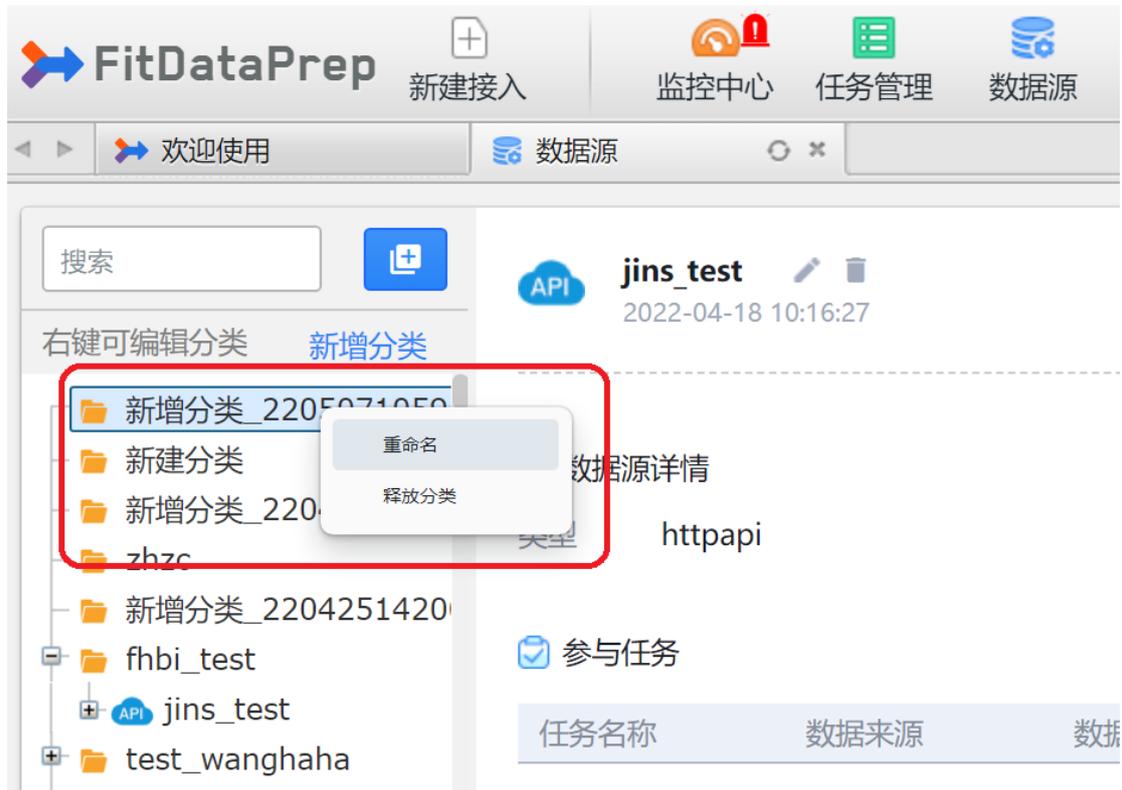
3.3 数据源

3.3.1 数据源分类

数据源有新增分类功能，点击  菜单打开数据源管理页面。点击左上角  按钮新增一个分类，可统一管理该分类下的数据源。在数据源上点击右键可以通过弹出框选择将已有的数据源添加到预先创建好的分类，见下图。

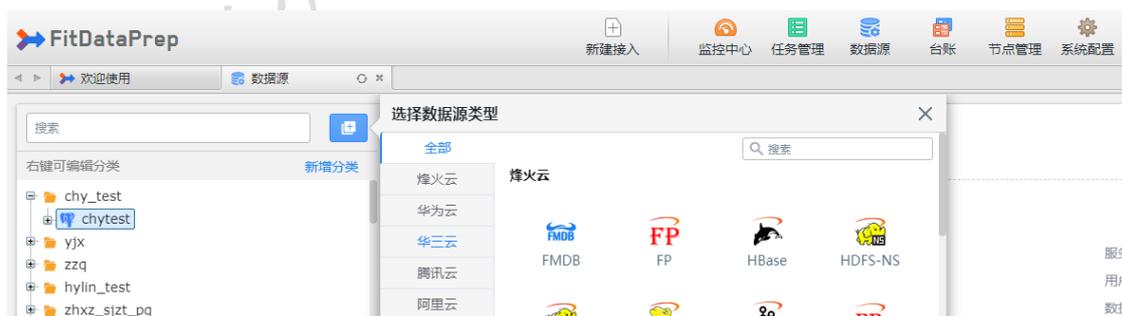


在分类节点上点击右键，在弹出菜单上可以选择对分类进行重命名和释放，选择“释放分类”后，分类中的数据源将移动到根节点上，默认不属于任何其他分类，见下图。



3.3.2 新增数据源

工具支持配置多种不同类型数据源，包括关系型数据库、大数据平台、文件/接口、NoSQL、消息中间件等。点击  菜单打开数据源管理页面。点击  按钮即可选择所需数据源进行配置。



3.3.2.1 关系型数据库

当前版本 FDP 工具支持的关系型数据库类型如下表：

开源数据库	商业数据库	
MariaDB	Mysql	Oracle

Greenplum	SQLServer	达梦
PostgreSQL	人大金仓数据库	Sybase
	LXDB	DB2
	Gbase 8a	Gbase 8s
	Informix	Vertica
	Hana	

以新建 MySQL 为例介绍关系型数据库的数据源配置，配置界面如下图。

新建 MySQL 数据源

* 数据源名称: 分类:

* 版本: MySQL5.x MySQL8.x

* 数据库IP: * 端口:

* 用户名: * 密码:

* 数据库名:

* 角色: 作为源端提供数据 作为目的端接收数据

数据源名称:	为该数据源定义一个名词，方便管理。
分类:	为数据源选择一个建好的分类，也可以建好后再进行分类。
版本:	选择一个 MYSQL 数据库的版本。需要根据实际情况选择。
数据库 IP:	填写数据库提供者的 MYSQL 数据库 IP 地址。
端口:	填写数据库提供者的 MYSQL 数据库端口号，默认端口 3306。
用户名:	填写数据库提供者的 MYSQL 数据库用户名。
密码:	填写数据库提供者的 MYSQL 数据库用户名对应的密码。
数据库名:	填写数据库需要接入数据的数据库名。
角色:	FDP 每个数据源都可以选择是作为源提供数据或者作为目的写入数据。也可以同时作为源和目的。根据实际情况进行选择。

配置数据源链接信息后，点击  按钮。数据源正常连通，点击

 按钮，保存新增的 MYSQL 数据源。若出现异常，请根据提示进行调整。

3.3.2.2 主流大数据厂家

FDP 支持多个主流大数据厂家数据源对接。分为烽火大数据平台系列、华为大数据平台系列、阿里大数据平台系列等。

烽火	华为	阿里	华三	腾讯
烽火HDFS	华为 Elasticsearch	阿里 maxcompute (Odps)	华三 Elasticsearch	腾讯 HDFS
烽火Hive	华为Hive	PolarDB	华三MPP	
烽火Hbase	华为HBase			
烽火FP	华为HDFS			
烽火FMDB	华为Kafka			
烽火盘古	华为GaussDB			
烽火全文 检索				
烽火流平 台				
烽火还原				
S3				
烽火Kafka				
烽火光闸				

下面以新建华为大数据平台系列 HDFS 数据源为例介绍配置方式。



数据源名称:	为该数据源定义一个名词，方便管理。
分类:	为数据源选择一个建好的分类，也可以建好后再进行分类。
认证方式:	选择 HDFS 组件的认证方式，需根据实际组件情况选择，实际情况可以咨询华为工程师或者在华为大数据平台配置管理界面的 HDFS 组件配置页面查询。
用户名:	填写 HDFS 组件提供着分配的连接用户名。
keytab 文件:	上传华为大数据平台提供的用户对应的 KEYTAB 认证文件。获取文件可以咨询华为工程师或者用分配的用户登录华为大数据平台配置管理后从界面下载。
krb5 文件:	上传华为大数据平台提供的用户对应的 KRB5 认证文件。获取文件可以咨询华为工程师或者用分配的用户登录华为大数据平台配置管理后从界面下载。
core-site 文件:	上传华为大数据平台 HDFS 组件对应的 core-site 文件。获取文件可以咨询华为工程师或者用分配的用户登录华为大数据平台组件管理后从界面下载。
hdfs-site 文件:	上传华为大数据平台 HDFS 组件对应的 core-site 文件。获取文件可以咨询华为工程师或者用分配的用户登录华为大数据平台组件管理后从界面下载。

角色：	FDP 每个数据源都可以选择是作为源提供数据或者作为目的写入数据。也可以同时作为源和目的。根据实际情况进行选择。
-----	--

填写必要信息后，点击 **测试连通性**，若连通性测试通过，选择数据源对应的角色（作为源端或目的端源），配置数据集，

配置数据源链接信息后，点击 **测试连通性** 按钮。若出现异常，请根据提示进行调整。待连通成功后，可配置数据源作为源端或者作为目标端下的数据集（数据集的配置请参考【4.1.3 小节】）。配置挖出后点击 **完成** 按钮，保存新增的华为 HDFS 数据源。

注意：

1. 请确认接入工具执行引擎所在服务器的 `/etc/hosts` 中已配置了连接集群的域名信息。
2. 请确认 FDP 工具服务器与大数据集群服务器的时间对时正确。

3.3.2.3 文件或接口

该类数据源包括文件类型和接口类型。

文件类型	接口类型	GA-T1400
FTP	HTTP API	依图
SCP		大华
本地上传		海康
		宇视
		以萨
		科达

下面以新建 FTP 数据源为例：

新建 FTP 数据源
□ ×

* 数据源名称: 分类:

* SSL: 否 是

* 主机IP: * 端口:

* 用户名: * 密码:

* 角色: 作为源端提供数据 作为目的端接收数据 测试连通性

源端定义的数据集 +

新源端数据集1

文件类型: 文本(字符分隔)

* 数据集名称:

* 远程目录: 📁

提示: 目录路径支持通配符?、*和**, 比如: /data/abc_*

? -- 表示匹配任意1个字符

* -- 表示匹配0到多个字符

** -- 表示匹配任意多级目录

* 文件过滤规则:

数据源名称:	为该数据源定义一个名词, 方便管理。
分类:	为数据源选择一个建好的分类, 也可以建好后再进行分类。
SSL:	选择 FTP 是否是 SSL。
主机 IP:	填写 FTP 提供者的服务器 IP 地址。
端口:	填写 FTP 提供者的服务器端口号, 默认端口 21。
用户名:	填写 FTP 用户名。
密码:	填写 FTP 用户名对应的密码。
角色:	FDP 每个数据源都可以选择是作为源提供数据或者作为目的写入数据。也可以同时作为源和目的。根据实际情况进行选择。

填写完成 FTP 数据源之后, 需要配置 FTP 数据源下的数据集信息, 数据集的配置请参考【4.1.3 小节】

3.3.2.4 消息中间件

工具支持对接消息中间件作为数据源, Kafka、华为 Kafka、ActiceMq、RabbitMq。下面介绍 kafka 的相关配置。

新建 Kafka 数据源
□ ×

注意:

- 请确认fdp所在机器和集群的网络连通信, 一般需要配置host或DNS
 - (1)、host配置方法: fdp所在机器的/etc/hosts文件配置kafka的域名信息(域名信息可以从kafka集群主节点的hosts文件中拷贝)
 - (2)、DNS配置方法: fdp所在机器的/etc/resolv.conf文件配置Sailor平台的DNS, 格式例如 kafkaserver 192.168.*.*
- 请确认相关服务器的时间对时正确。

* 数据源名称: 分类:

* 服务器地址:

* 认证方式: 无 SASL Kerberos

* 角色: 作为源端提供数据 作为目的端接收数据 测试连通性

目的端定义的存储数据集 +

新目的数据集1
消息格式: 文本(字符分隔)

* 数据集名称:

* Topic(主题): ▼
多个Topic可以使用逗号(,)分隔

* 消息格式: 文本(字符分隔) JSON XML 文件 recordSet

* 字段分隔符: 制表符(\t) 逗号(,) 分号(;) 空格 其它

* 文本识别符: ▼

* 文本识别符转义符: ▼

完成

数据源名称:	为该数据源定义一个名词, 方便管理。
分类:	为数据源选择一个建好的分类, 也可以建好后再进行分类。
服务器地址:	填写全部的服务器信息, 允许输入多个, 用英文逗号隔开(输入样例: server1:21007,server2:21007)
认证方式:	选择 KAFKA 集群的认证方式
角色:	FDP 每个数据源都可以选择是作为源提供数据或者作为目的写入数据。也可以同时作为源和目的。根据实际情况进行选择。

填写基本信息后, 需要新建数据集。填写 topic、消息格式等参数, 点击

完成 按钮保存数据源。

! 注意:

1. 请确认 fdp 所在机器和集群的网络连通信, 一般需要配置 host 或 DNS

host 配置方法: fdp 所在机器的/etc/hosts 文件配置 hw-kafka 的域名信息(域名信息可以从 hw-kafka 集群主节点的 hosts 文件中拷贝)

DNS 配置方法: fdp 所在机器的/etc/resolv.conf 文件配置 DNS, 格式例如
kafkaserver 192.168..**

2. 请确认 FDP 工具服务器与大数据集群服务器的时间对时正确。

3.3.2.5 NoSQL 数据库

FDP 目前支持 Redis 和 MongoBb 两种 NoSQL 数据库。下面以新建 Redis 数据源为例介绍配置。

RedisS 作为源端:

新建 Redis 数据源
□ ×

* 数据源名称:

* 数据库IP:

* 数据库名:

密码:

* 角色: 作为源端提供数据
 作为目的端接收数据

分类:

* 端口:

用户名:

源端redis数据集 +

新源端数据集1

value类型: 文本(字符分隔)

* 数据集名称:

* key过滤规则:
提示: key支持通配符 *、? 和 [], 比如: hello*、he?lo、h[e]lo
* -- 表示匹配0到多个字符
? -- 表示匹配任意1个字符
[] -- 表示匹配[]内的任意1个字符

* value类型: 文本(字符分隔) JSON Hash

* 字段分隔符: 制表符(\t) 逗号(,) 分号(;)
 其它

* 文本识别符:

* 文本识别符转义符:

* 数据字段: 点击单元格编辑内容

字段名称	数据类型	数据最大长度
<div style="border: 1px solid #ccc; width: 50px; height: 30px; margin: 0 auto; display: flex; align-items: center; justify-content: center;"> 添加数据字段 </div>		

数据源名称:	为该数据源定义一个名词, 方便管理。
分类:	为数据源选择一个建好的分类, 也可以建好后再进行分类。
主机 IP:	填写 Redis 提供者的数据库服务器 IP 地址。
端口:	填写 Redis 提供者的数据库服务器端口号, 默认端口 6379。
数据库名:	填写Redis提供者的数据库名称。
用户名:	填写 Redis 用户名。
密码:	填写 Redis 用户名对应的密码。
角色:	FDP 每个数据源都可以选择是作为源提供数据或者作为目的写入数据。也可以同时作为源和目的。根据实际情况进行选择。

RedisS 作为目的端:

新建 Redis 数据源
□ ×

* 数据源名称: 分类:

* 数据库IP: * 端口:

* 数据库名: 用户名:

密码:

* 角色: 作为源端提供数据 作为目的端接收数据 测试连通性

目的端redis数据集
+

新目的数据集 1
value类型: 文本(字符分隔)

* 数据集名称:

* value类型: 文本(字符分隔) JSON Hash

* 字段分隔符: 制表符(\t) 逗号(,) 分号(;) 空格 其它

* 文本识别符:

* 文本识别符转义符:

完成

数据源名称:	为该数据源定义一个名词, 方便管理。
分类:	为数据源选择一个建好的分类, 也可以建好后再进行分类。

主机 IP:	填写 Redis 提供者的数据库服务器 IP 地址。
端口:	填写 Redis 提供者的数据库服务器端口号，默认端口 6379。
数据库名:	填写 Redis 提供者的数据库名称。
用户名:	填写 Redis 用户名。
密码:	填写 Redis 用户名对应的密码。
角色:	FDP 每个数据源都可以选择是作为源提供数据或者作为目的写入数据。也可以同时作为源和目的。根据实际情况进行选择。

Redis 作为源端和目的端的基本参数配置是一样的，数据集配置略有不同，源端可以配置抽取数据过滤的 key 规则等参数。

配置数据源连接信息后，点击 ，若连通性测试通过，选择数据源对应的角色（作为源端或目的端源），如需配置数据集，数据集的配置请参考【4.1.3 小节】。

3.3.3 新增数据集

FDP 支持从多种类型数据载体中提取数据集，如基于纯文本文件、JSON 文件、XML 文件等，下表为当前 FDP 支持提取数据集的载体类型。

数据载体类型	介绍
纯文本	一般我们将每行以逗号、tab 字符、空格、逗号或其他可见和不可见字符分割的文本文件归类为纯文本类型的数据，比如 CSV 文件。FDP 可以逐行读取文件进行接入输出。
JSON	支持从 JSON 文件中根据选择的节点层级提取相同层级下的所有数据为数据集。如提取商品列表的 JSON 对象中的所有商品和商品属性。FDP 提供可视化的界面一键提取所需的节点字段。
XML	支持从 XML 文件中根据选择的节点层级提取相同层级下的所有数据为数据集。如提取商品列表的 XML 对象中的所有商品和商品属性。FDP 提供可视化的界面一键提

	取所需的节点字段。
EXCEL	支持从EXCEL文件中提取所需数据集。
纯文件	FDP将不需要对文件进行任何处理，仅将文件从一个数据源搬运到另一个数据源的场景称为“纯文件搬运” FDP支持将一类纯文件作为数据集。比如将所有的zip压缩文件作为一个数据集，将所有txt文本文件作为另一个数据集。
文件对象	“文件对象”的数据集用于解决需要对文件进行特殊处理的场景，比如根据文件名、文件体的解析结果进行入库。比如需要对文件中的部分内容进行加解密等。FDP支持将这类文件定义为“文件对象”的数据集。数据集会默认该文件的文件名称、文件路径、文件体三个属性，用于自定义“派大星”程序对文件进行处理后输出。
派大星解析	一些复杂的文件内容，比如xml，zip包里面含有不同的bcp文件，fdp无法自己解析这些数据。可通过配置自定义的解析派大星规则去接入这些数据。
文件	文件类型现在仅支持源为【文件/接口】类里的飞书数据源类型，目的为【大数据平台】的kafka数据源类型。
全文	全文类型数据源支持配置检索条件从全文库检索指定数据。
HTTP API	API类数据源支持配置从指定URL请求数据的数据集。 对于API返回的JSON、XML、纯文本类型的数据支持进行步的配置解析。

3.3.3.1 纯文本

纯文件可以通过FTP、SCP、大数据平台等多种数据源中接入，我们以FTP数据源为例介绍纯文本数据集的配置方法。

新建一个FTP数据源后，会默认生成一个数据集标签页，可点击增加数据集，将鼠标放到左侧的标签上可点击可删除某个数据集，一个数据源至少保留一个数据集。



数据集名称：	为该数据集定义一个名词，方便管理。
远程目录：	配置该数据集需要从远程目录上获取的文件目录集合，支持通配符提取。如“/data/abc_*”提取所有根目录data下以“abc_”开头的所有文件夹。 支持的通配方式见配置页面介绍。

文件过滤规则:	配置该数据集需要从已经配置的远程目录上获取的文件文件集合，支持通配符提取。如 data_*.bcp 提取所有“data_”开头的并且以“bcp”为后缀的文件。 支持的通配方式见配置页面介绍。
是否删除原文件:	选择获取文件后是否需要删除源目录中的文件。
文件类型:	纯文本文件选择 <input checked="" type="radio"/> 文本(字符分隔)类型
文件编码:	选择文件的编码类型。选择错误会导致数据乱码。
是否解压文件:	如果源文件为压缩文件，可以选择先对文件进行解压。
字段分隔符:	选择或填写纯文本文件的字段分隔符。支持选择制表符、逗号、分号、空格，并可以输出自定义特殊分隔符。
文本识别符: 文本识别符转义符:	<p>选择文本识别符与文本识别符转义符。在csv格式中，如果字段中出现了特殊符号可能会破坏csv格式，此时需要加文本识别符将字段包围起来。例如:address字段中有制表符，需要加双引号将其包围，这时这个双引号就叫文本识别符</p> <pre>name → age → address CRLE 张三 → 18 → "南京 → 建邺" CRLE</pre> <p>当字段需要文本识别符进行包围时，内容中又出现了文本识别符，这时需要对内容中的文本识别符进行转义</p> <pre>name → age → address CRLE 张三 → 18 → "南京 → 建邺\"中胜\" CRLE</pre> <p>上图中的斜杠就是文本识别符转义符。标准csv格式是使用双引号作为文本识别符的，文本识别符转义符也是双引号。</p>
表头占据行数: 数据开始行号:	文件中包含表头的需要输入文件中表头所占据的行数。 以及数据开始的行号。
数据字段:	定义数据集字段名（不需要与原字段名相同），可以支持批量填写。

填写完成参数后，点击 **完成** 按钮保存数据集。

3.3.3.2 JSON

JSON 可以通过 FTP、SCP、大数据平台等多种数据源中接入，我们以 FTP 数据源为例介绍 JSON 数据集的配置方法。

新建一个 FTP 数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。

修改【FTP】数据源 测试连通性

* 角色: 作为源端提供数据 作为目的端接收数据

源端定义的数据集 +

新源端数据集1
文件类型: JSON

* 数据集名称: 新源端数据集1

* 远程目录: /data/abc_*
提示: 目录路径支持通配符 ?、* 和 **, 比如: /data/abc_*
? -- 表示匹配任意1个字符
* -- 表示匹配0到多个字符
** -- 表示匹配任意多级目录

* 文件过滤规则: data_*.bcp
提示: 文件名支持通配符 ? 和 *, 比如: data_*.bcp、data_log?.log
? -- 表示匹配任意1个字符
* -- 表示匹配0到多个字符

* 是否删除原文件: 删除原文件 保留原文件
文件读取成功后是否删除原文件

* 文件类型: 文本(字符分隔) JSON XML Excel 文件对象 纯文件(用于原文件搬运)

* 文件编码: UTF-8

* 是否解压文件: 是 否
支持zip、gz、tar、tar.gz文件

* JSON数据形式: 整个文件是一个JSON 每行是一个JSON

* 是否解析一层: 是 否

* Data-Path 路径: 填写数据所在的Data-Path路径 可视化配置
Data-Path 路径是用于读取JSON数据的重要位置标记, 建议使用 [可视化配置] 进行配置。
如果使用了 [可视化配置] 操作, 则这里的Path路径和下面的数据字段会自动填入。

* 数据字段: 添加字段 批量添加 点击单元格编辑内容

字段名称	字段Path路径	数据类型	数据最大长度	备注

数据集名称:	为该数据集定义一个名词，方便管理。
远程目录:	配置该数据集需要从远程目录上获取的文件目录集合，支持通配符提取。如 “/data/abc_*” 提取所有根目录 data 下以 “abc_” 开头的所有文件夹。 支持的通配方式见配置页面介绍。

文件过滤规则:	配置该数据集需要从已经配置的远程目录上获取的文件文件集合，支持通配符提取。如 data_*.bcp 提取所有“data_”开头的并且以“bcp”为后缀的文件。 支持的通配方式见配置页面介绍。
是否删除原文件:	选择获取文件后是否需要删除源目录中的文件。
文件类型:	纯文本文件选择 <input type="radio"/> JSON 类型
文件编码:	选择文件的编码类型。选择错误会导致数据乱码。
是否解压文件:	如果源文件为压缩文件，可以选择先对文件进行解压。
JSON 数据形式:	根据数据文件的 JSON 形式选择，整个文件是一个 JSON 结构体，或者文件中每行是一个 JSON 结构体。
是否只解析一层:	是否只解析选择的节点下的一层数据。
Data-Path 路径:	需要解析的JSON文件的数据路径，如解析根结构下的 user 实体数据 (/user) FDP支持可视化进行选择。参见“JSON数据Path路径可视化配置”
数据字段:	定义数据集字段名（不需要与原字段名相同），可视化配置后可以自动生成。用户可以根据生成的字段进行微

JSON 数据 Path 路径可视化配置

点击 Data-Path 路径配置项后面的 **可视化配置** 按钮，打开可视化配置界面。将需要解析的 JSON 样例文本复制到左侧的文本框中，单击右侧选择要解析的 JSON 部分，检查下方的表格解析出的字段是否正确，点击确认即可自动生成字段信息，与实际字段有冲突的字段可手动微调，点击 **确定** 按钮保持配置。

JSON 数据 Path 路径可视化选择配置 (试用)

1.请填入 JSON 样例数据 (注意: 数据大小建议 < 10KB, 否则可能会造成卡顿基...)

```

{
  "devDependencies": {
    "empower": "~0.7.1",
    "expect": "~0.3.1",
    "mocha": "~1.18.2",
    "power-assert-formatter": "~0.7.1",
    "qunit": "1.14.0",
    "qunit-tap": "1.4.2",
    "requirejs": "~2.1.11",
    "assert": "Jxck/assert"
  },
  "homepage": "https://github.com/twada/power-assert",
  "_release": "0.7.2",
  "_resolution": {
    "type": "version",
    "tag": "v0.7.2",
    "commit": "ed5c5b16a08a6cc0bd0bda5bbeb90cf9fc74d859"
  },
  "_source": "git://github.com/twada/power-assert.git",
  "_target": "~0.7.2",
  "_originalSource": "power-assert"
},

```

2.请从下面选择【数据记录】所在位置

```

{
  "devDependencies": {
    empower: ~0.7.1,
    expect: ~0.3.1,
    mocha: ~1.18.2,
    power-assert-formatter: ~0.7.1,
    qunit: 1.14.0,
    qunit-tap: 1.4.2,
    requirejs: ~2.1.11,
    assert: Jxck/assert
  },
  homepage: https://github.com/twada/power...,
  _release: 0.7.2,
  _resolution: {
    type: version,
    tag: v0.7.2,
    commit: ed5c5b16a08a6cc0bd0bda5bbeb90c...
  },
},

```

3.数据预览 (数据记录路径: /devDependencies) 数组展平 只提取一层数据

empower	expect	mocha	power-ass...	qunit	qunit-tap	requirejs	assert
~0.7.1	~0.3.1	~1.18.2	~0.7.1	1.14.0	1.4.2	~2.1.11	Jxck/assert

* 文件类型: 文本(字符分隔) JSON XML Excel 文件对象
 纯文件 (用于原文件搬运)

* 文件编码:

* 是否解压文件: 是 否
 支持zip、gz、tar、tar.gz文件

* JSON数据形式: 整个文件是一个JSON 每行是一个JSON

* 是否解析一层: 是 否

* Data-Path 路径: 可视化配置

Data-Path 路径是用于读取JSON数据的重要位置标记, 建议使用 [可视化配置] 进行配置。
 如果使用了 [可视化配置] 操作, 则这里的Path路径和下面的数据字段会自动填入。

* 数据字段: 点击单元格编辑内容

	字段名称	字段Path路径	数据类型	数据最大长度	备注
1	empower	empower	字符	512	
2	expect	expect	字符	512	
3	mocha	mocha	字符	512	
4	power-assert-forma...	power-assert-forma...	字符	512	
5	qunit	qunit	字符	512	
6	qunit-tap	qunit-tap	字符	512	
7	requirejs	requirejs	字符	512	
8	assert	assert	字符	512	

3.3.3.3 XML

XML 可以通过 FTP、SCP、大数据平台、HTTP API 等多种数据源中接入，我们以 FTP 数据源为例介绍 XML 数据集的配置方法。

新建一个 FTP 数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。



数据集名称：	为该数据集定义一个名词，方便管理。
远程目录：	配置该数据集需要从远程目录上获取的文件目录集合，支持通配符提取。如 “/data/abc_*” 提取所有根目录 data 下以 “abc_” 开头的所有文件夹。 支持的通配方式见配置页面介绍。

文件过滤规则:	配置该数据集需要从已经配置的远程目录上获取的文件文件集合，支持通配符提取。如 data_*.bcp 提取所有“data_”开头的并且以“bcp”为后缀的文件。 支持的通配方式见配置页面介绍。
是否删除原文件:	选择获取文件后是否需要删除源目录中的文件。
文件类型:	纯文本文件选择 <input type="radio"/> XML 类型
文件编码:	选择文件的编码类型。选择错误会导致数据乱码。
是否解压文件:	如果源文件为压缩文件，可以选择先对文件进行解压。
XML-XPath 路径:	遵循XPath语法，提取需要解析的XML文件的数据路径，如解析根结构下的user实体数据 (/user) FDP支持可视化进行选择。参见“XML数据Path路径可视化配置”
xml 解析条件 sql:	可通过类sql的方式对xml中的数据进行过滤。参见“XML条件sql配置样例”
数据字段:	定义数据集字段名（不需要与原字段名相同），可视化配置后可以自动生成。用户可以根据生成的字段进行微

XML 数据 Path 路径可视化配置

点击 XML-Path 路径配置项后面的 **可视化配置** 按钮，打开可视化配置界面。将需要解析的 XML 样例文本复制到左侧的文本框中，单击右侧选择要解析的 XML 部分，检查下方的表格解析出的字段是否正确，点击确认即可自动生成字段信息，与实际字段有冲突的字段可手动微调，点击 **确定** 按钮保持配置。



* 文件类型: 文本(字符分隔) JSON XML Excel 文件对象
 纯文件 (用于原文件搬运)

* 文件编码:

* 是否解压文件: 是 否
 支持zip、gz、tar、tar.gz文件

* XML-XPath ... 可视化配置
 XPath 路径是用于读取XML数据的重要位置标记, 建议使用 [可视化配置] 进行配置。
 如果使用了 [可视化配置] 操作, 则这里的XPath路径和下面的数据字段会自动填入。

xml解析条件sql:
 xml解析条件sql模版: [xml条件sql配置样例](#)
PIVOT (aggregate_function(value_column) **FOR** pivot_expr **IN** (columnName_as_c
 olumnName_list)) [[**AS**] tableAlias]

* 数据字段:

点击单元格编辑内容

	字段名称	字段Path路径	数据类型	数据最大...	备注
1	@id	@id	字符	512	
2	@plugin-identif...	@plugin-identif...	字符	512	
3	@url	@url	字符	512	
4	@version	@version	字符	512	

XML 条件 sql 配置样例

```

样例xml:

<?xml version="1.0" encoding="UTF-8"?>
<MESSAGE>
  <DATASET name="WA_COMMON_010000">
    <DATA>
      <ITEM key="H010001" val="1030261"></ITEM>
      <ITEM key="I010014" val="5101001637683184236"></ITEM>
      <ITEM key="I010031" val="GROUPCHAT"></ITEM>
      <ITEM key="I010013" val="3"></ITEM>
    </DATA>
  </DATASET>
  <DATASET name="WA_COMMON_010022">
    <DATA>
      <ITEM key="B040017" val="3147209"></ITEM>
      <ITEM key="D010010" val="2761411"></ITEM>
      <ITEM key="B050004" val="3910219"></ITEM>
      <ITEM key="B040003" val="4197889"></ITEM>
      <ITEM key="H040008" val="1637683032"></ITEM>
      <ITEM key="H010003" val="01"></ITEM>
      <ITEM fnt="base64" key="H040001" val="5aW9"></ITEM>
    </DATA>
  </DATASET>
</MESSAGE>
    
```

xml条件sql书写标准：

```
PIVOT (aggregate_function(value_column) FOR pivot_expr IN(columnValue_as_columnName_list)) [[AS] tableAlias]
```

xml条件sql书写样例：

```
pivot(LAST("@val") for "@key" in('B040017','D010010','B050004','B040003','H040008','H010003','H040001')) as pvt
```

备注：条件sql要配合xpath路径一起使用，本例中xpath配置成：`/MESSAGE/DATASET[@name='WA_COMMON_010022']/DATA/ITEM。`

DATASET[@name='WA_COMMON_010022']的意思是只解析标签**DATASET**属性**name**值为**WA_COMMON_010022**的数据。

解析后数据展示：

表头	B040017	D010010	B050004	B040003	H040008	H010003	H040001
数据	3147209	2761411	3910219	4197889	1637683032	01	5aW9

不加条件解析数据展示：

表头	@key	@value
数据	B040017	3147209
数据	D010010	3147209
数据	B050004	2761411
数据	B040003	4197889
数据	H040008	1637683032
数据	H010003	01
数据	H040001	5aW9

3.3.3.4 Excel

Excel 可以通过 FTP、SCP、大数据平台、HTTP API 等多种数据源中接入，我们以 FTP 数据源为例介绍 Excel 数据集的配置方法。

新建一个 FTP 数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。Excel 为多层结构，可以设置每个 sheet 页的字段信息。



数据集名称:	为该数据集定义一个名词，方便管理。
远程目录:	配置该数据集需要从远程目录上获取的文件目录集合，支持通配符提取。如 “/data/abc_*” 提取所有根目录 data 下以 “abc_” 开头的所有文件夹。 支持的通配方式见配置页面介绍。
文件过滤规则:	配置该数据集需要从已经配置的远程目录上获取的文件文件集合，支持通配符提取。如 data_*.bcp 提取所有 “data_” 开头的并且以 “bcp” 为后缀的文件。 支持的通配方式见配置页面介绍。
是否删除原文件:	选择获取文件后是否需要删除源目录中的文件。
文件类型:	纯文本文件选择 <input type="radio"/> Excel 类型
文件编码:	选择文件的编码类型。选择错误会导致数据乱码。
是否解压文件:	如果源文件为压缩文件，可以选择先对文件进行解压。

表头占据行数：	文件中包含表头的需要输入文件中表头所占据的行数。
数据开始行号：	以及数据开始的行号。
数据字段：	定义数据集字段名（不需要与原字段名相同），可以支持批量填写。

填写完成参数后，点击  按钮保存数据集。

3.3.3.5 纯文件

纯文件可以通过 FTP、SCP、大数据平台等多种数据源中接入，我们以 FTP 数据源为例介绍纯文件数据集的配置方法。

新建一个 FTP 数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。

新建 FTP 数据源 □ ×

* 数据源名称: 分类:

* SSL: 否 是

* 主机IP: * 端口:

* 用户名: * 密码:

* 角色: 作为源端提供数据 作为目的端接收数据 测试连通性

源端定义的数据集 +

新源端数据集1
文件类型: 文件

* 数据集名称:

* 远程目录: 📁
提示: 目录路径支持通配符 ?、* 和 **, 比如: /data/abc_*
 ? -- 表示匹配任意1个字符
 * -- 表示匹配0到多个字符
 ** -- 表示匹配任意多级目录

* 文件过滤规则:
提示: 文件名支持通配符 ? 和 *, 比如: data_*.bcp、data_log?.log
 ? -- 表示匹配任意1个字符
 * -- 表示匹配0到多个字符

* 是否删除原文件: 删除原文件 保留原文件
文件读取成功后是否需要删除原文件

* 文件类型: 文本(字符分隔) JSON XML Excel
 文件对象 纯文件 (用于原文件搬运)

完成

数据集名称:	为该数据集定义一个名词, 方便管理。
远程目录:	配置该数据集需要从远程目录上获取的文件目录集合, 支持通配符提取。如 “/data/abc_*” 提取所有根目录 data 下以 “abc_” 开头的所有文件夹。 支持的通配方式见配置页面介绍。
文件过滤规则:	配置该数据集需要从已经配置的远程目录上获取的文件文件集合, 支持通配符提取。如 data_*.bcp 提取所有 “data_” 开头的并且以 “bcp” 为后缀的文件。 支持的通配方式见配置页面介绍。
是否删除原文件:	选择获取文件后是否需要删除源目录中的文件。

文件类型： 纯文本文件选择 纯文件类型

对于纯文件类型的数据搬运。注意数据源和目的数据源要选择相同的数据集类型。如果源文件是压缩文件，可以在任务管理中设置是否对源文件进行解压。



3.3.3.6 文件对象

文件对象可以通过 FTP、SCP、大数据平台等多种数据源中接入，我们以 FTP 数据源为例介绍文件对象数据集的配置方法。

新建一个 FTP 数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。

修改【FTP】数据源

* 主机IP: 10.0.25.210 * 端口: 21

* 用户名: zhanghaixin * 密码:

* 角色: 作为源端提供数据 作为目的端接收数据 测试连通性

源端定义的数据集

新源端数据集1
文件类型: fileObject

* 数据集名称: 新源端数据集1

* 远程目录: /data/abc_*
提示: 目录路径支持通配符?、*和**, 比如: /data/abc_*
? -- 表示匹配任意1个字符
* -- 表示匹配0到多个字符
** -- 表示匹配任意多级目录

* 文件过滤规则: data_*.bcp
提示: 文件名支持通配符?和*, 比如: data_*.bcp、data_log?.log
? -- 表示匹配任意1个字符
* -- 表示匹配0到多个字符

* 是否删除原文件: 删除原文件 保留原文件
文件读取成功后是否需要删除原文件

* 文件类型: 文本(字符分隔) JSON XML Excel 文件对象 纯文件(用于原文件搬运)

* 数据字段:

字段名称	数据类型	数据最大长度	备注
1 fileName	字符	512	文件名称
2 filePath	字符	512	文件路径
3 fileContent	字符	20000	文件内容 (Base64)

数据集名称:	为该数据集定义一个名词，方便管理。
远程目录:	配置该数据集需要从远程目录上获取的文件目录集合，支持通配符提取。如“/data/abc_*”提取所有根目录data下以“abc_”开头的文件夹。 支持的通配方式见配置页面介绍。
文件过滤规则:	配置该数据集需要从已经配置的远程目录上获取的文件文件集合，支持通配符提取。如 data_*.bcp 提取所有“data_”开头的并且以“bcp”为后缀的文件。 支持的通配方式见配置页面介绍。
是否删除原文件:	选择获取文件后是否需要删除源目录中的文件。
文件类型:	纯文本文件选择 <input checked="" type="radio"/> 文件对象 类型
数据字段:	默认会提取对应文件的3个属性字段，fileName、filePath、fileContent。

3.3.3.7 派大星解析

一些复杂的文件内容，比如 xml，zip 包里面含有不同的 bcp 文件，fdp 无法自动解析这些数据。可通过配置自定义的解析派大星规则去接入这些数据。

派大星解析可以支持 FTP、SCP、大数据平台等多种数据源中接入，我们以 FTP 数据源为例介绍派大星数据集的配置方法。

新建一个 FTP 数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。



数据集名称：	为该数据集定义一个名词，方便管理。
远程目录：	配置该数据集需要从远程目录上获取的文件目录集合，支持通配符提取。如 “/data/abc_*” 提取所有根目录 data 下以 “abc_” 开头的所有文件夹。 支持的通配方式见配置页面介绍。

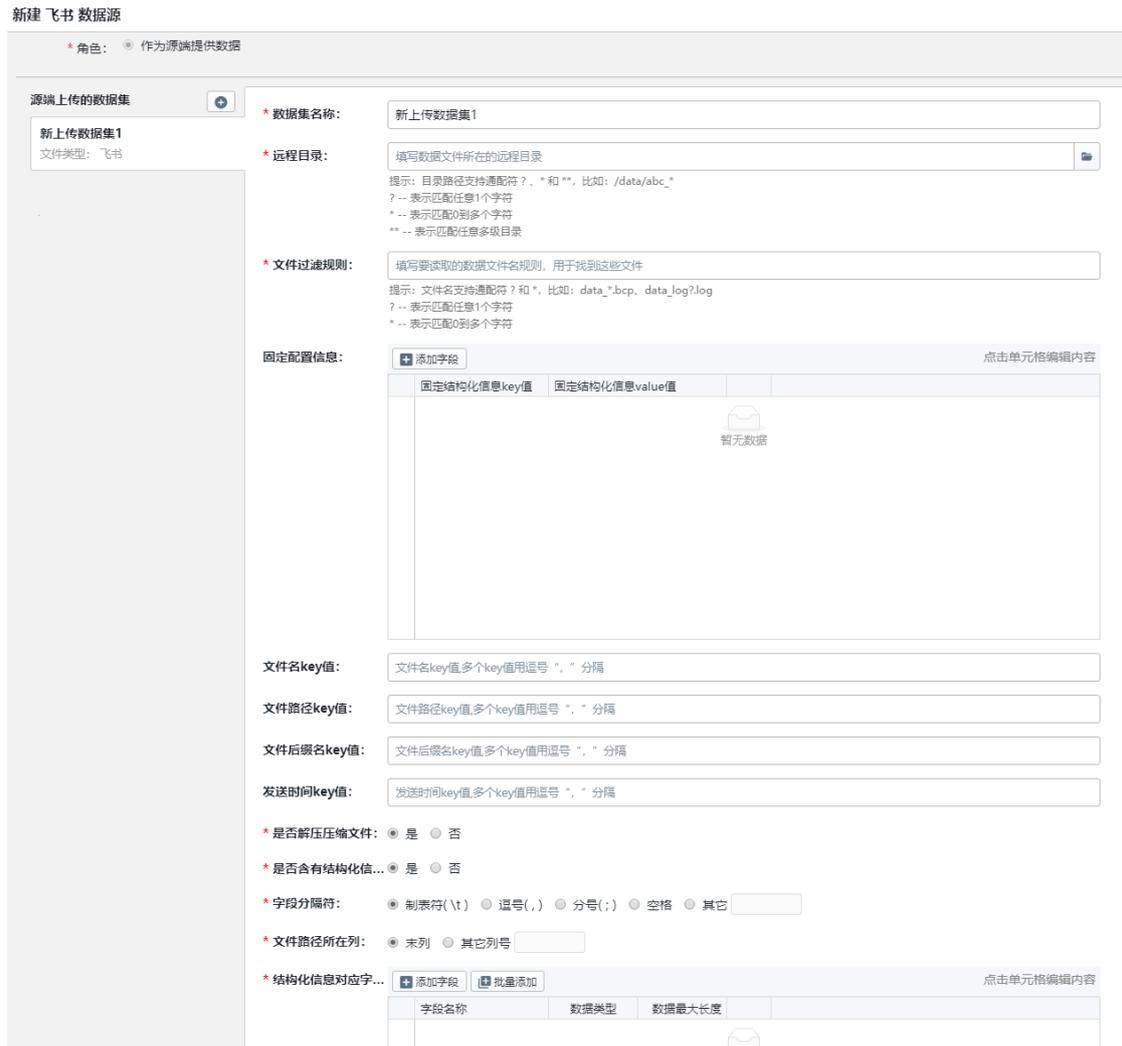
文件过滤规则:	配置该数据集需要从已经配置的远程目录上获取的文件文件集合，支持通配符提取。如 data_*.bcp 提取所有“data_”开头的并且以“bcp”为后缀的文件。 支持的通配方式见配置页面介绍。
是否删除原文件:	选择获取文件后是否需要删除源目录中的文件。
文件类型:	纯文本文件选择 <input checked="" type="radio"/> 派大星解析类型。
规则集:	下拉选择处理该数据的派大星规则。
数据表:	通过派大星解析后的数据文件，可能将一个数据拆分成多个数据集。比如一个压缩包内的包含班级和学生两种文件。将班级和学生作为 2 种数据集输出。
数据表名:	为派大星解析出的每个数据表定义一个表名称，方便管理。
数据字段:	每个表的数据字段添加方式可以通过手动添加，也可以进行批量添加。

填写完成参数后，点击  按钮保存数据集。

3.3.3.8 文件

文件类型现在仅支持数据源类型为【文件/接口】大类下的  数据源类型，目的数据源只能为数据源类型【大数据平台】下的 kafka 类型。

新建一个飞书数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。



数据集名称:	为该数据集定义一个名词，方便管理。
远程目录:	配置该数据集需要从远程目录上获取的文件目录集合，支持通配符提取。如“/data/abc_*”提取所有根目录 data 下以“abc_”开头的所有文件夹。支持的通配方式见配置页面介绍。
文件过滤规则:	配置该数据集需要从已经配置的远程目录上获取的文件文件集合，支持通配符提取。如 data_*.bcp 提取所有“data_”开头的并且以“bcp”为后缀的文件。支持的通配方式见配置页面介绍。
固定配置信息:	固定配置信息
文件名 key 值:	文件名key值

文件后缀名 key 值:	文件后缀名key值
发送时间 key 值:	发送时间 key 值
是否解压压缩文件:	是否解压压缩文件
是否含有结构化信息:	是否含有结构化信息
字段分隔符:	字段分隔符
文件路径所在列:	文件路径所在列
结构化信息对应字段名称:	配置解析文件对应的每列数据的字段名称

KAFKA 作为目的端

新建 Kafka 数据源

注意:

- 请确认fdp所在机器和集群的网络连通性，一般需要配置host或DNS
 (1)、host配置方法: fdp所在机器的/etc/hosts文件配置kafka的域名信息(域名信息可以从kafka集群主节点的hosts文件中拷贝)
 (2)、DNS配置方法: fdp所在机器的/etc/resolv.conf文件配置Sailor平台的DNS, 格式例如 kafkaserver 192.168.*.*
- 请确认相关服务器的时间对时正确。

* 数据源名称: 分类:

* 服务器地址:

* 认证方式: 无 SASL Kerberos

* 角色: 作为源端提供数据 作为目的端接收数据 测试连通性

目的端定义的存储数据集 +

新目的数据集1
消息格式: 文件

* 数据集名称:

* Topic(主题): ▼
多个Topic可以使用逗号(,)分隔

* 消息格式: 文本(字符分隔) JSON XML 文件 recordSet

完成

数据集名称:	为该数据集定义一个名词，方便管理。
Topic(主题):	输入数据集需要消费的 KAFKA topic 清单，一个或者多个。可以点击下拉按钮选择 KAFKA 集群已有的 topic，也可以使用键盘输入，多个 topic 时使用英文逗号分割。

消息格式：	选择 topic 内的消费数据格式，如文本、JSON、XML 等
字段分隔符：	消息格式选择“文本”情况下的字段分隔符，选择或者手动输出。
文本识别符：	消息格式选择“文本”情况下的文本识别符。
文本识别符转义符：	消息格式选择“文本”情况下的文本识别符转义符。
数据来源：	数据来源根据实际情况填写。
数据类型 (dataType)：	填写上报数据的类型，由现场业务自行定义。
上传地市 (upareaId)：	填写上传数据的地市编码，例如310000。
数据详情： (dataSpecific)	通常为MASS。
主附件(orgfile)：	主附件对应的字段名称
从附件(attach)：	从附件对应的字段名称

3.3.3.9 其他

3.3.3.9.1 全文

全文类型数据源支持配置检索条件从全文库检索指定数据，全文数据集仅支持数据源类型为【大数据平台】大类下的  数据源类型。

新建一个烽火全文检索数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。

新建 烽火全文检索 数据源

角色: 作为源端提供数据 测试连通性

源端定义的数据集

数据集名称:

关键词:

查询关键字语法:

- 查询语法保留的特殊符号: " () , # . AND OR NOT , 如果要查询这些特殊符号需要放在双引号或者单引号里。
- 语法对单引号和双引号做特殊处理, 如用户查 " hello world " , 也可以编成 " 'hello world' " 。
- 支持域检索, 域名和域值之间以@号分隔, 中间不能有空格, 如用户要查邮件发信人中含有 "fenghuo" 的, 则写成 "FROM:fenghuo" , 不能写成 "FROM:fenghuo" 或者 "FROM: fenghuo" , 域名支持字母数字和下划线, 开头只能是字母和下划线, 实际查询的时候不管大写小写统一转成小写。
- 支持距离检索, 如用户要查 "烽火" 和 "科技" 相差在5个字符以内的, 可以写成 "#5(烽火 科技)" , "#" 、“5”、“(” 之间不能有空格, 如果用户要查两者相差在0个字符以内的, 即连在一起, 可以直接写成 "烽火科技" , 语法解析程序会自动转成 "#0(烽火 科技)" , 减少了用户输入时的麻烦。
- 距离检索可以带上域名, 如用户要查邮件主题中 "烽火" 和 "科技" 相差在5个字符以内的, 可以写成 "_SUBJECT:#5(烽火 科技)" , 没有域名表示默认域, 一般为正文。
- 支持逻辑表达式与或非, 与可以写成 "AND" , 或可以写成 "OR" , 非可以写成 "NOT" , 三者都是双目运算符, 如用户要查包含 "a" 但是没有 "b" 的, 可以写成 "a NOT b" 。
- 域检索、距离检索、逻辑表达式可以混合使用, 如用户要查邮件主题中含有 "烽火" 和 "科技" 的, 可以写成 "_SUBJECT:烽火 AND _SUBJECT:科技" , 不能写成 "_SUBJECT:烽火 AND 科技" , 域检索内部并不支持逻辑表达式。
- 逻辑表达式的优先级, 与 > 或 > 非, 但是括号可以改变优先级, 如 "a OR b AND c" , 表示先查AND, 再查OR, 但是 "(a OR b) AND c" 就变成先查OR, 后查AND。
- 支持模糊查询, 如可以查询 "abc*" , 一个查询关键词只能有一个"*"号, 但是如果带有逻辑表达式, 则可以有多多个"*"号, 如 "abc" AND def** 。

开始时间: 格式: 20200820102050 (年月日时分秒)

结束时间: 格式: 20200820102050 (年月日时分秒)

最大结果数:

上下文长度:

查询协议:

数据集名称:	为该数据集定义一个名词, 方便管理。
关键词:	检索关键词, 关键词语法参见FDP工具配置界面介绍。
开始时间:	检索的时间范围, 开始时间。
结束时间:	检索的时间范围, 结束时间。
最大结果数:	单次返回的最大查询结果条数。
上下文长度:	上下文摘要长度默认大小是200。
查询协议:	查询协议列表串默认为空字符串, 各协议号之间用空格分隔。 (dataSource) 比如协议号1000001代表httppost数据。

填写完成参数后, 点击 完成 按钮保存数据集。

3.3.3.9.2 HttpApi

API 类数据源支持配置从指定 URL 请求数据的数据集, API 数据集仅支持数据源类型为【文件/接口】大类下的  数据源类型。
HTTP-API

新建一个 HTTP-API 数据源后，会默认生成一个数据集标签页，可点击  增加数据集，将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。



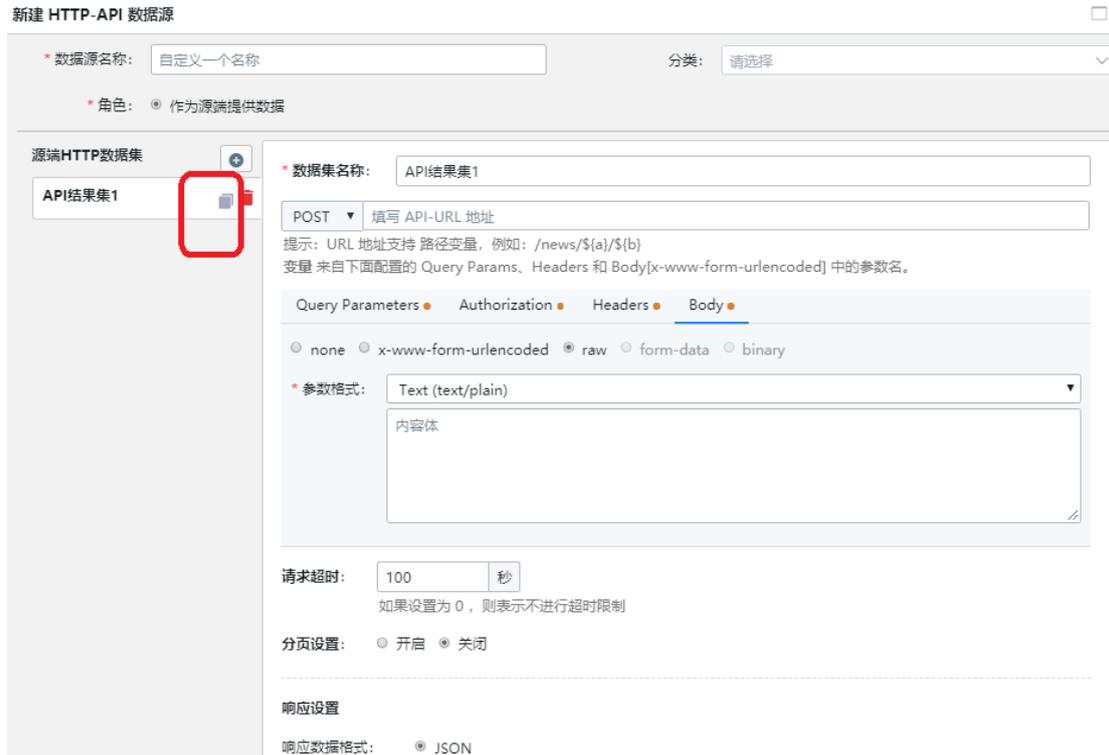
数据集名称:	为该数据集定义一个名词，方便管理。
 :	选择HTTP接口的是post还是gei请求。选择后在右侧输入框输入请求的具体URL地址。地址支持变量，具体参考配置界面介绍。
Query Parameters:	可以在此添加请求参数，以key-value形式。

Authorization:	选择请求的认证方式，根据认证方式填写认证参数。认证相关参数介绍配置界面介绍。
Headers:	可以在此添加http头信息，以key-value形式。
Body:	Post方式下可以在此添加请求体信息。
请求超时:	填写HTTP请求的超时时间。“0”表示不进行超时限制
分页设置:	在接口是分页方式返回数据的情况下，需要选择分页请求开启，配置分页参数。工具会根据分页配置按顺序分页调用请求接口。具体分页配置方式见工具界面介绍。
响应设置:	配置相应的数据处理方式
响应数据格式:	选择接口返回的数据格式。
解析响应数据:	选择是否对响应的数据进行结构化转换解析。如选择解析，工具可以对返回数据种的JSON等消息体按照对应的解析器进行解析，得到所需的结构化字段信息。

3.3.4 数据集复制

新建一个 HTTP-API 数据源后，会默认生成一个数据集标签页，可点击  增加数据集，对于需要新增只有少量参数变化的数据集，在配置一个数据集后可以将鼠标放到左侧已有数据集的标签上，点击数据集标签上的  按钮进行数据集复制。

将鼠标放到左侧的标签上可点击  可删除某个数据集，一个数据源至少保留一个数据集。



3.3.5 数据源明细

点击工具界面上方的  菜单，所有已经新建的数据源会按照数据源分类以树状结构展示在工具左侧。可以点击树状结构的  图片对数据源进行展开。点击数据源名称，会在工具右侧展示数据源详情信息以及已经使用了该数据源的任务信息。



在任务列表种点击“详情”可以查看该任务的任务明细页面，进行启动暂停任务等操作，且支持在列表种直接对任务进行删除（任务删除后不可恢复）。

任务清单界面见下图↓

任务名称	数据来源	数据目的源	任务创建时间	详情	元数据导出
wh_pg2mysql	ljy-pg-全类型测试	OMS43	2021-10-29 09:38:22	详情 删除任务	导出
ljy-fmdb数据类型新增测试	ljy-pg-全类型测试	ljy-fmdb对外环境1	2021-09-14 10:26:58	详情 删除任务	导出

任务详细界面见下图↓



在任务清单上点击导出按钮，点击导出会导出记录此任务明细信息的 excel。

任务名称	数据来源	数据目的源	任务创建时间	详情	元数据导出
wh_pg2mysql	ljy-pg-全类型测试	OMS43	2021-10-29 09:38:22	详情 删除任务	导出
ljy-fmdb数据类型新增测试	ljy-pg-全类型测试	ljy-fmdb对外环境1	2021-09-14 10:26:58	详情 删除任务	导出

导出效果见下图↓

表名	public_zkongchuli_copy1							
表中文名								
中文描述								
资源代码								
总行数								
平均行大小 (byte)								
文件类型	文本文件							
文件路径	null/zkongchuli_copy1							
是否删除								
分割符 (文本文件)	\t							
最大文件大小 (MB) (文本文件)								
根节点标签名 (XML文件)								
主分区								
子分区								
字段名	中文名	字段描述	字段类型	字段长度	是否可以空	是否主键	是否索引	样例数据
zdnnull	111	111	int2		5 是	否		
zdspace	222	222	varchar		255 是	否		

点击左侧树状结构数据源下的数据集名称，工具会在右侧预览该表的数据和结构。点击右上角的 **下载样例数据** 按钮可以下载对应数据集的数据样例。





对于文件类型的数据源则支持预览其文件目录。



3.3.6 编辑数据源

对于数据源参数发生变化，或者需要增加该数据源下的表或文件作为新的数据集接入。此时可以使用数据源编辑功能修改已配置参数，新增数据集。新增数据集可参考【2.3.3 小节】。但存在关联任务的数据集不支持修改。

点击数据源界面左侧树状结构的数据源节点名称，在右侧数据源详情页面上方的数据源名称后面点击  按钮。进入数据源编辑界面。

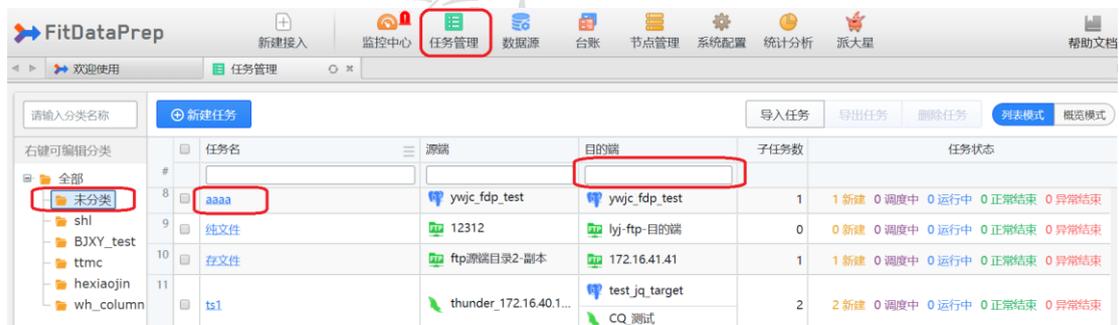
数据源编辑与数据源新增界面操作一致，数据集新增和修改参数以及方式介绍请参见【2.3.3 小节】。



3.4 任务管理

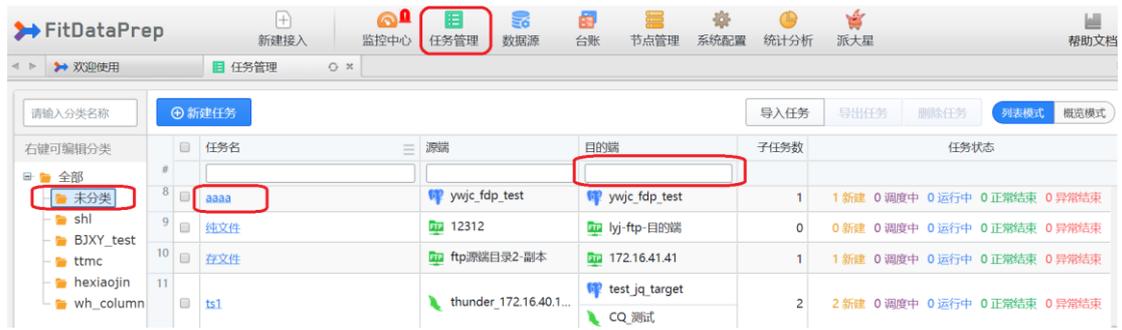
3.4.1 任务列表

点击 FDP 工具上方的  导航菜单，即可进入接入任务管理界面。任务管理界面左侧以树状结构展示任务的分类列表。点击分类，在右侧展示分类下的发任务清单（可以点击列表右上方的   切换按钮进行展示模式的切换）。点击任务清单中古的任务名称可以进入具体的任务详情界面。



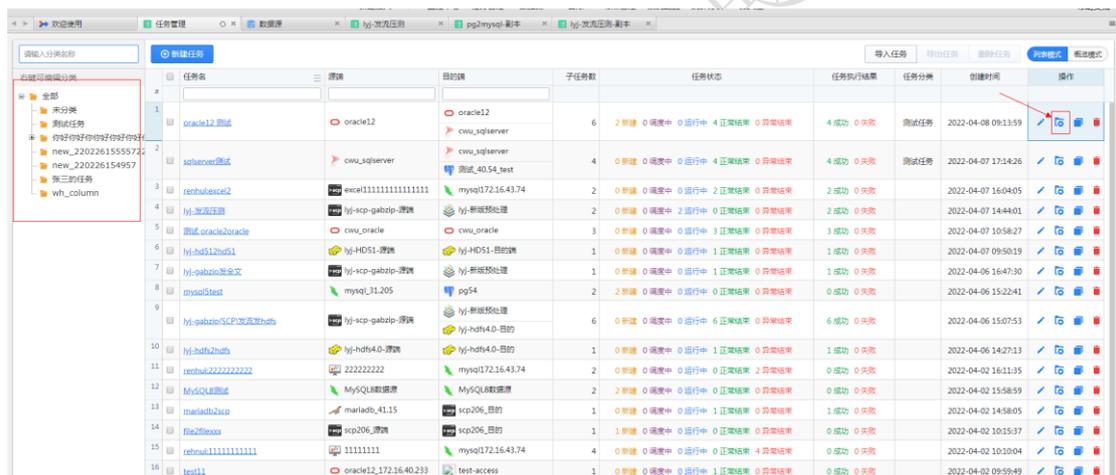
3.4.1.1 任务查询

在任务列表页面可以通过任务名查询任务，也可以通过源端和目的端的数据源名称查询与数据源相关的任务。



3.4.1.2 任务分类

任务列表左侧树状结构上可以点击右键新建分类，然后任务列表清单上点击在操作栏点击  按钮，在弹出框中该选择相应的分类对任务进行分类，可点击左侧分类可以查看已经分到点对应分类的任务。



3.4.1.3 任务复制

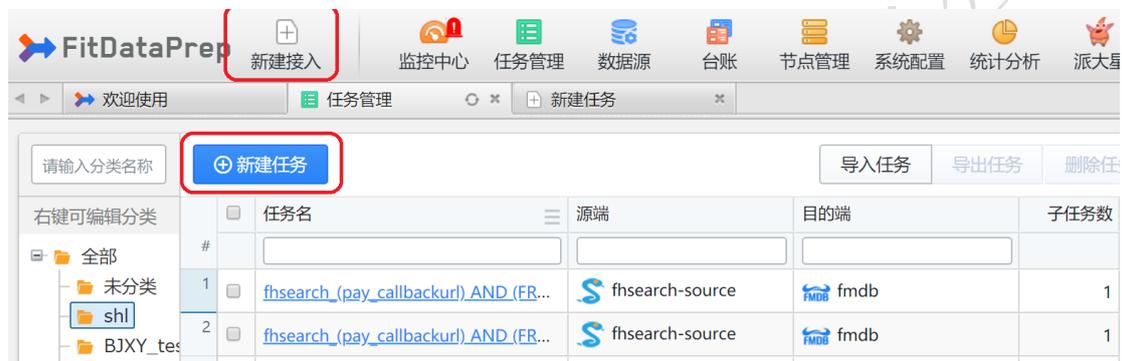
在有类似任务需要新增的情况下为了减少重复的编辑工作或者避免破坏原任务的情况下对任务进行修改调试，可以点击任务栏中古的  按钮对任务进行复制后再编辑。



3.4.2 新建任务

3.4.2.1 选择源端目的端

点击 FDP 工具上方的  导航菜单，即可进入接入任务管理界面。任务管理界面左侧以树状结构展示任务的分类列表。点击列表左上方的  按钮进入新增任务界面。另外工具提供了点击系统上方的  快捷按钮直接进入新建任务界面。



下图为新建数据源界面，输入必须参数后选择数据来源和去向点击  按钮进入下一步。



任务名称:	为该任务定义一个名字，方便管理。
任务分类:	为任务选择一个分类，方便管理。任务分类在任务列表界面的左侧树状结构上点击右键进行新增。
选择数据来源:	选择已经新建好的一个数据源作为数据来源，作为数据来源的数据源必须是数据源“角色”属性选择了 <input checked="" type="checkbox"/> 作为源端提供数据的选项。新建数据源请参考 2.3.2 小节。数据源可以通过上方的输入框中根据数据源名称进行搜索。

选择数据去向：	选择已经新建好的一个数据源作为数据去向，作为数据去向的数据源必须是数据源“角色”属性选择了 <input type="checkbox"/> 作为目的端接收数据的选项。新建数据源请参考 2.3.2 小节。数据源可以通过上方的输入框中根据数据源名称进行搜索。
---------	--

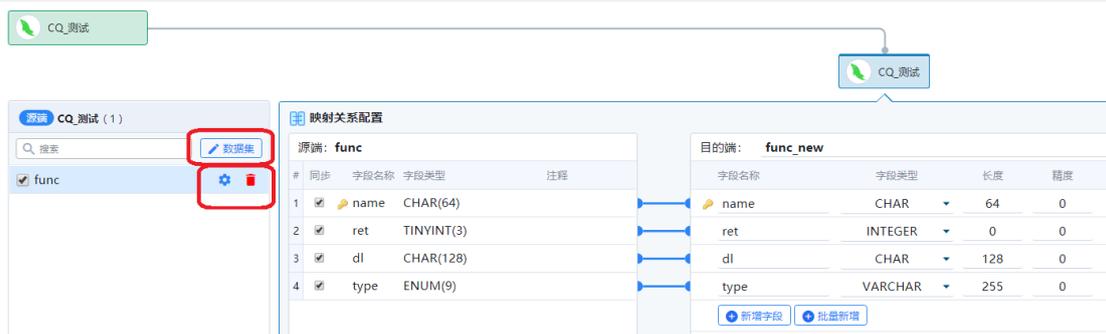
3.4.2.2 配置映射关系

源端和目的端选择完成之后，点击下一步，进入映射关系配置页面。界面左侧点击  按钮可以从源端选择需要接入的数据集（数据集的创建请参见 2.3.3 小节）。



选择完数据集后，选择新的数据集的按钮将在已经选择数据集清单的上方。点击  按钮可以再次打开数据集选择窗，已经选择的数据集会被是以进该选择的状态。

点击  按钮可以删除已经选择的数据集。点击  按钮可以进入该数据集的接入策略。



调度策略配置见下下图↓



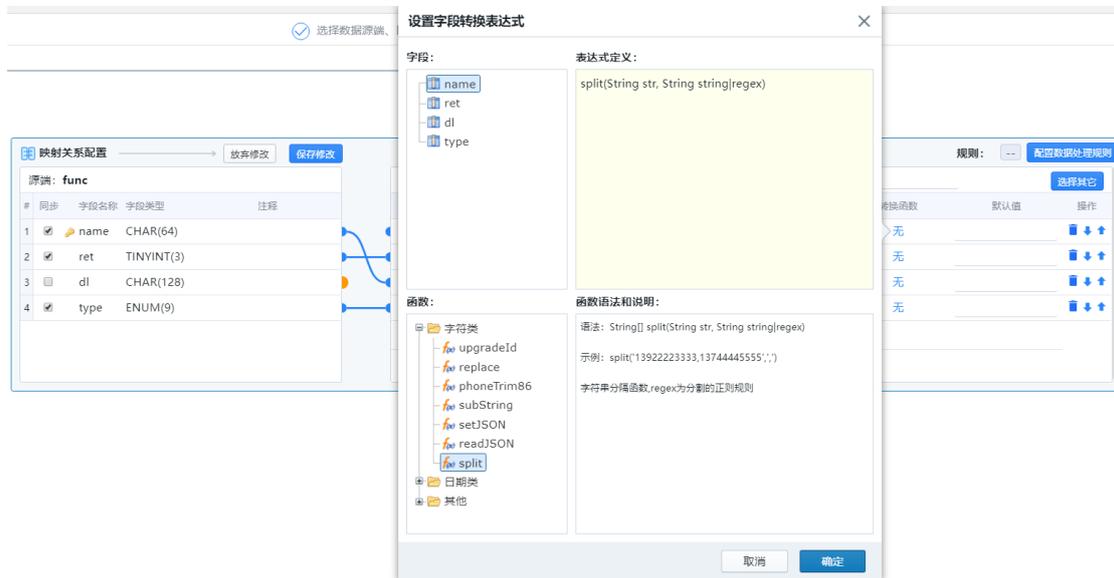
3.4.2.2.1 表到表的映射

选择完数据集后，界面右侧会展示该数据集下的字段信息和在目的端的对应数据集(表)信息。目的端的数据集(表)默认是新建的方式创建，如果需要接入到已有的数据集(表)中，可以点击 **选择其它** 按钮选择其他数据集(表)。同时可以修改目的端的表名称，字段名称，字段类型等参数，且支持新增字段和填写字段空的情况下字段默认值。

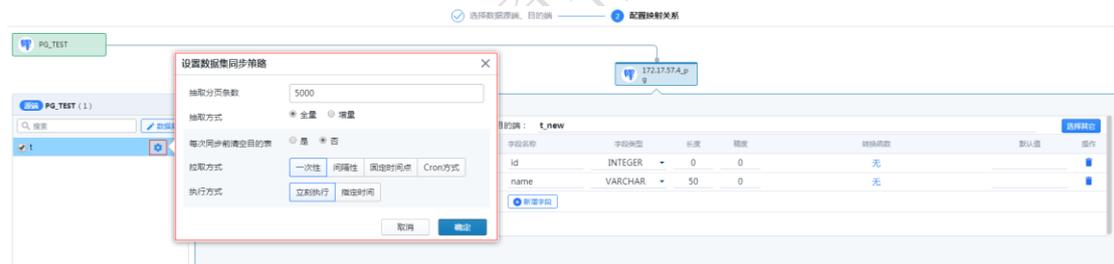
默认源和目的的两个数据集字段间对应关系通过字段间的蓝色连线表示。连线两端的凸点  都是可以点击，点击后会变为橙色 。当一侧变为橙色，点击另一侧的字段凸点即可完成两个字段的连线，完成字段的映射关系配置。



对于需要对字段内容进行处理转换的场景，工具支持指定字段使用函数进行字段转换，包括日期格式转换，字符串替换、分割，下载 url 指向的 web 资源等。点击目的数据集字段列表“转换函数”字段标题下的单元格即可弹出函数转换窗口。



FDP 支持针对数据集配置数据集同步策略，设置抽取方式和抽取分页条数。



同步策略抽取方式选择增量时，支持指定增量字段进行增量抽取，也支持使用 logminer, binlog, wal 和 trigger 方式进行增量抽取。(oracle 和 dm 数据库使用 logminer, mysql 数据库使用 binlog, pg 数据库使用 wal, oracle & mysql & pg 均可以使用 trigger)。

设置数据集同步策略

抽取分页条数: 5000

抽取方式: 全量 增量

选择增量方式: 增量字段 wal trigger

指定增量字段: id [int4]

提醒: 增量字段仅支持 数值类 和 日期时间类

- 1. 数值类 可以是int类型的主键, 或自增序列
- 2. 日期时间类 可以是入库时间, 更新时间等
- 3. 增量字段要求后入库的记录要比已入库的所有记录值都大

增量字段起始值:

提醒: 起始值可以不填, 如果不填则默认是从头开始抽取

拉取方式: 间隔性 固定时间点 Cron方式

间隔时间: 30 分钟

设置数据集同步策略

抽取分页条数: 5000

抽取方式: 全量 增量

选择增量方式: 增量字段 binlog trigger

存量数据同步: 同步存量数据

开启后系统将首先同步所有存储数据, 然后进行增量读取

提醒: 使用MySQL-binlog方式需要满足以下要求:

- 1. 版本为 5.5 或 5.6 或 5.7 的非只读数据库
- 2. 存储引擎为 InnoDB
- 3. 当前用户要具有相关表和视图的 select 权限, replication slave, replication client
- 4. binlog 记录方式为 ROW 模式
- 5. 不支持同步视图
- 6. 选择binlog模式时, 需要数据库先开启binlog, [详情](#)

设置数据集同步策略

抽取分页条数: 5000

抽取方式: 全量 增量

选择增量方式: 增量字段 logminer

存量数据同步:

开启后系统将首先同步所有存储数据, 然后进行增量读取

提醒: 选择logminer模式时, 需要数据库先开启logminer, [详情](#)

设置数据集同步策略

抽取分页条数: 5000

抽取方式: 全量 增量

选择增量方式: 增量字段 wal trigger

存量数据同步: 同步存量数据
开启后系统将首先同步所有存储数据, 然后进行增量读取

提醒: 选择wal模式时, 需要PG数据库先开启逻辑复制, [详情](#)

取消 确定

设置数据集同步策略

抽取分页条数: 5000

抽取方式: 全量 增量

选择增量方式: 增量字段 wal trigger

存量数据同步: 同步存量数据
开启后系统将首先同步所有存储数据, 然后进行增量读取

提醒: 选择数据库触发器 (trigger) 模式时, 需要满足以下要求:

1. 源端数据库表必须有主键
2. 目的端数据库表必须有和源端数据库表——映射的主键
3. 当前用户要具有对源端数据库建表和创建触发器的权限

取消 确定

在 logminer 下点击详情按钮, 可以查看 logminer 介绍以及使用 LogMiner 需要数据库开启归档模式的操作教程。

Oracle开启归档日志方法

1. 什么是LogMiner

Oracle LogMiner 是Oracle公司在8i之后推出的日志(在线日志和归档日志)的分析工具,该工具可以分析出所有对数据库操作的DML和DDL语句。

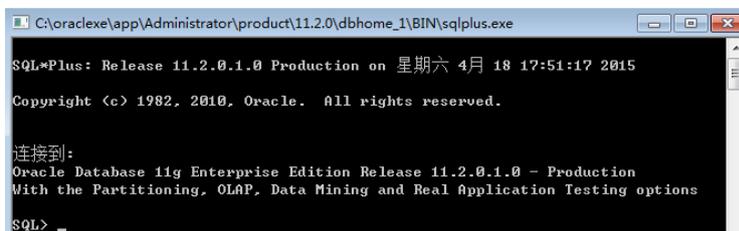
2. LogMiner的主要用途

- 跟踪数据库的变化:可以离线的跟踪数据库的变化,而不会影响在线系统的性能。
- 回退数据库的变化:回退特定的变化数据,减少point-in-time recovery的执行。
- 优化和扩容计划:可通过分析日志文件中的数据以分析数据增长模式。

3. 使用LogMiner需要数据库开启归档模式

3.1. 以dba身份登录

```
sqlplus sys/sys as sysdba
```



3.2. 查询当前日志归档状态

```
archive log list;
```

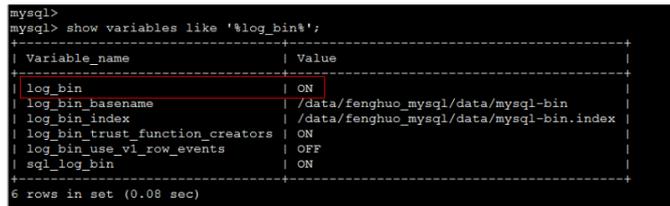
在 binlog 下点击详情按钮, 可以查看 Mysql-Binlog 开启方法。

Mysql-Binlog开启方法

1 查看是否开启binlog

进入mysql命令行模式, 输入如下命令:

```
show variables like '%log_bin%';
```



如果log_bin对应的值为on, 表示已经开启binlog;

如果log_bin对应的值为off, 表示没有开启binlog.

2 设置开启binlog

2.1 修改配置文件

Mysql的配置文件一般所在的路径是: /etc/my.conf

在 wal 下点击详情按钮, 可以查看 Postgresql 逻辑复制开启方法。

Postgresql逻辑复制开启方法

1 版本

PG库版本必须高于或等于9.4

```
psql --version

[root@i-hrnvm0q ~]#
[root@i-hrnvm0q ~]#
[root@i-hrnvm0q ~]#
[root@i-hrnvm0q ~]# psql --version
psql (PostgreSQL) 10.6
[root@i-hrnvm0q ~]#
```

2 数据库配置

2.1 文件目录

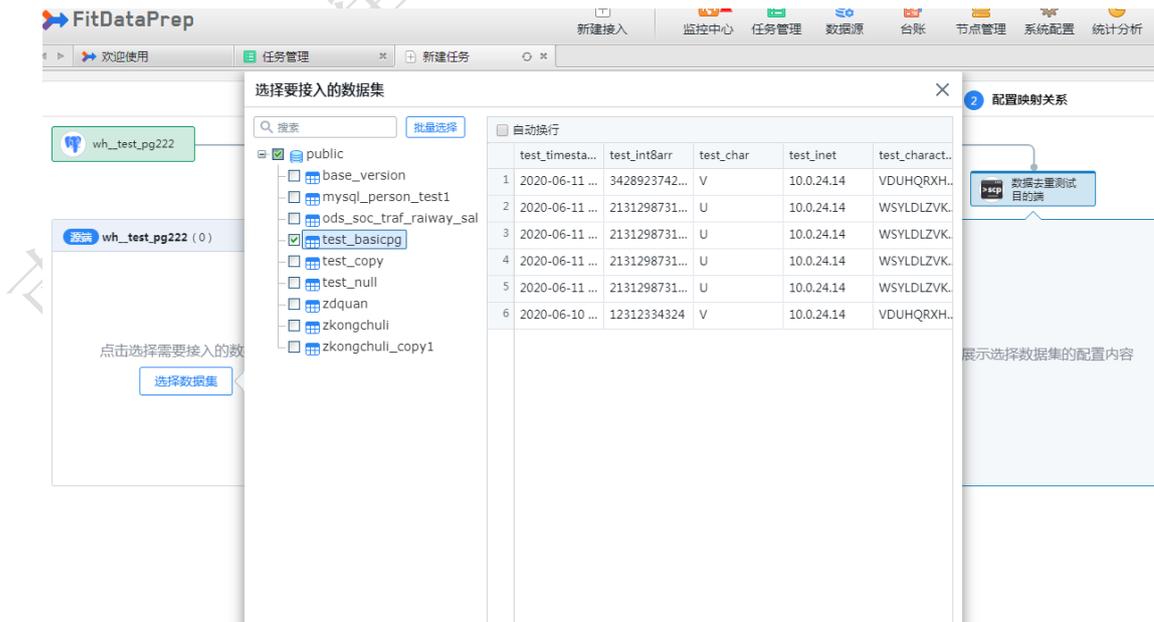
数据库配置中需修改postgresql.conf文件以及pg_hba.conf文件，可以通过以下命令查找文件所在目录

```
echo $PGDATA

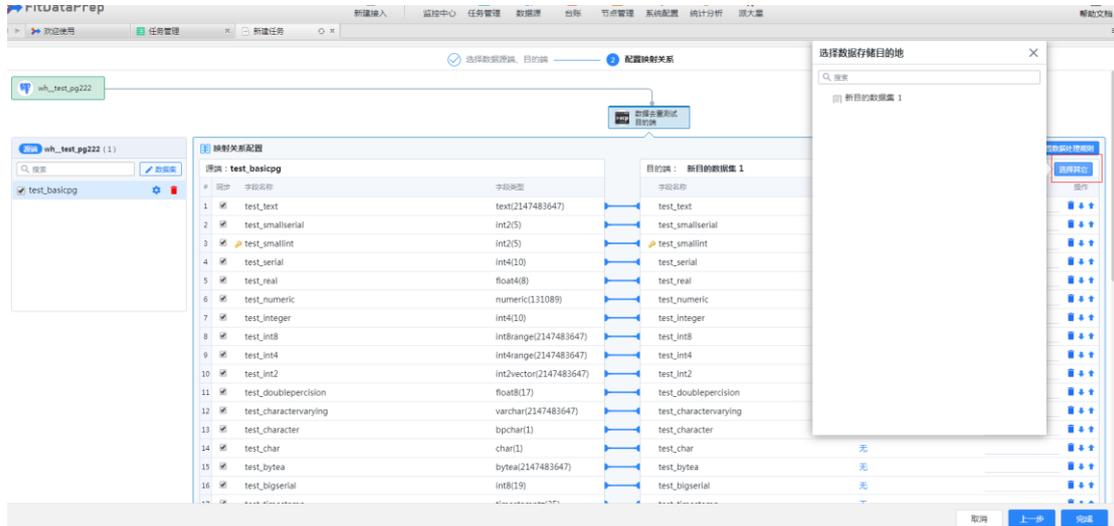
[root@i-hrnvm0q ~]#
[root@i-hrnvm0q ~]# echo $PGDATA
/pgsql/10/data
[root@i-hrnvm0q ~]# cd /pgsql/10/data
[root@i-hrnvm0q data]# ll
total 172
drwx----- 10 postgres postgres 4096 Oct 20 20:33 base
-rw----- 1 postgres postgres 0 Dec 10 09:23 core.12712
-rw----- 1 postgres postgres 0 Oct 30 10:34 core.17800
-rw----- 1 postgres postgres 44 Jan 6 15:48 current_logfiles
drwx----- 2 postgres postgres 4096 Jan 6 15:49 global
drwx----- 2 postgres postgres 20480 Jan 6 15:48 log
drwx----- 2 postgres postgres 4096 Aug 27 10:02 pg_commit_ts
drwx----- 2 postgres postgres 4096 Aug 27 10:02 pg_dynshmem
```

3.4.2.2.2 表到文件的映射

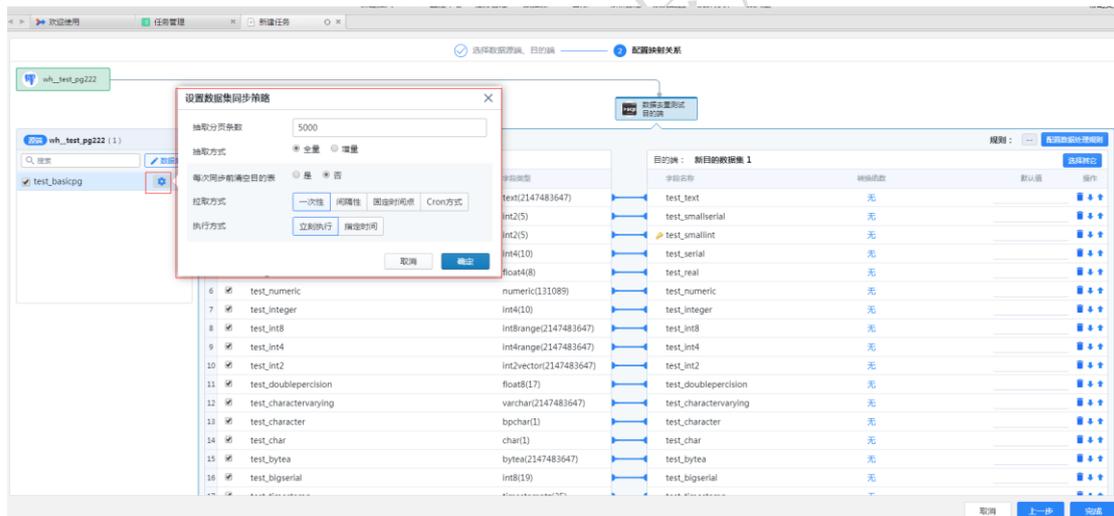
选择源端数据集，即选择需要接入的表。



展示源表中的字段名称和字段类型等信息，点击目的端【选择其他】按钮，选择目的端数据集和源表映射，配置源和目的端的字段映射关系，修改结束点击【完成】进行保存。

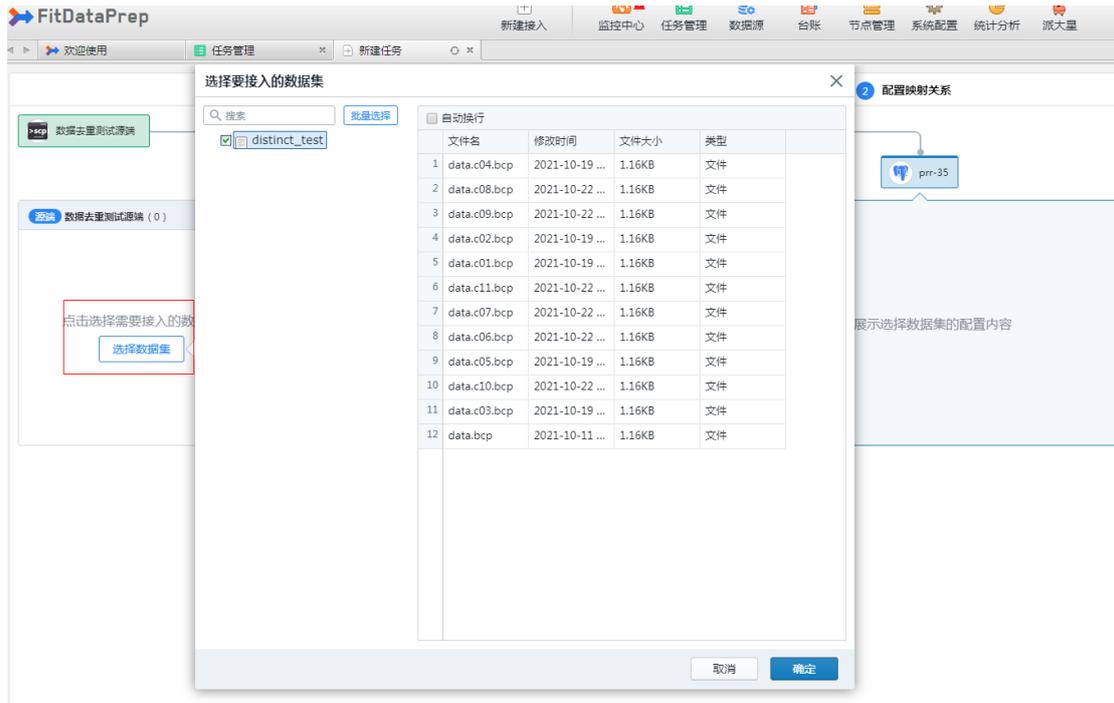


同步策略配置（不支持增量模式）。



3.4.2.2.3 文件到表的映射

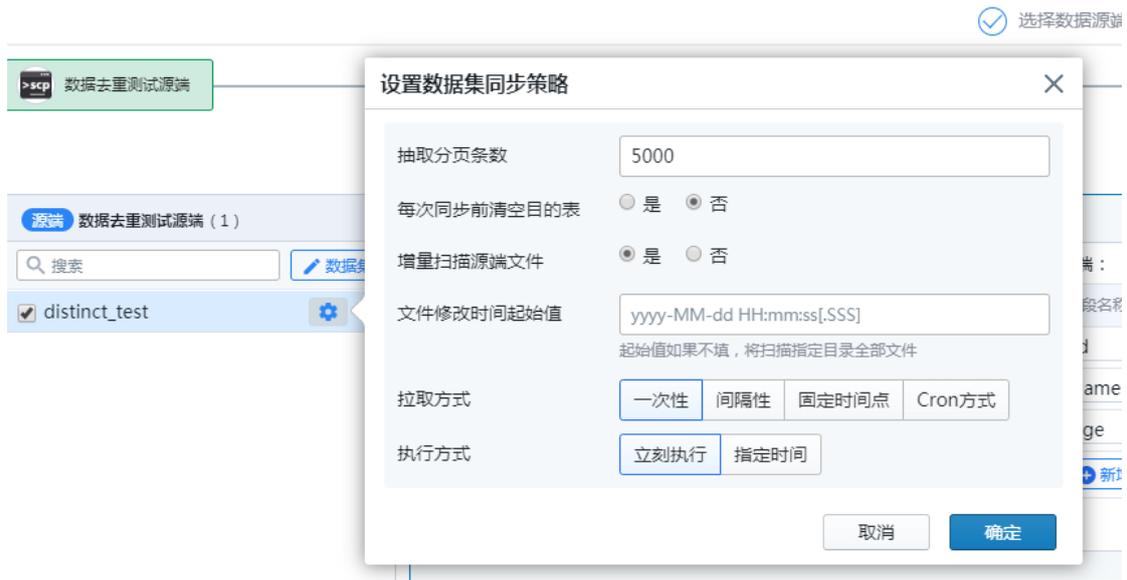
选择源端数据集。



展示源端数据集中的字段名称和字段类型等信息，点击目的端【选择其他】按钮，选择目的端表和源端数据集映射，配置源和目的端的字段映射关系，修改结束点击【完成】进行保存。

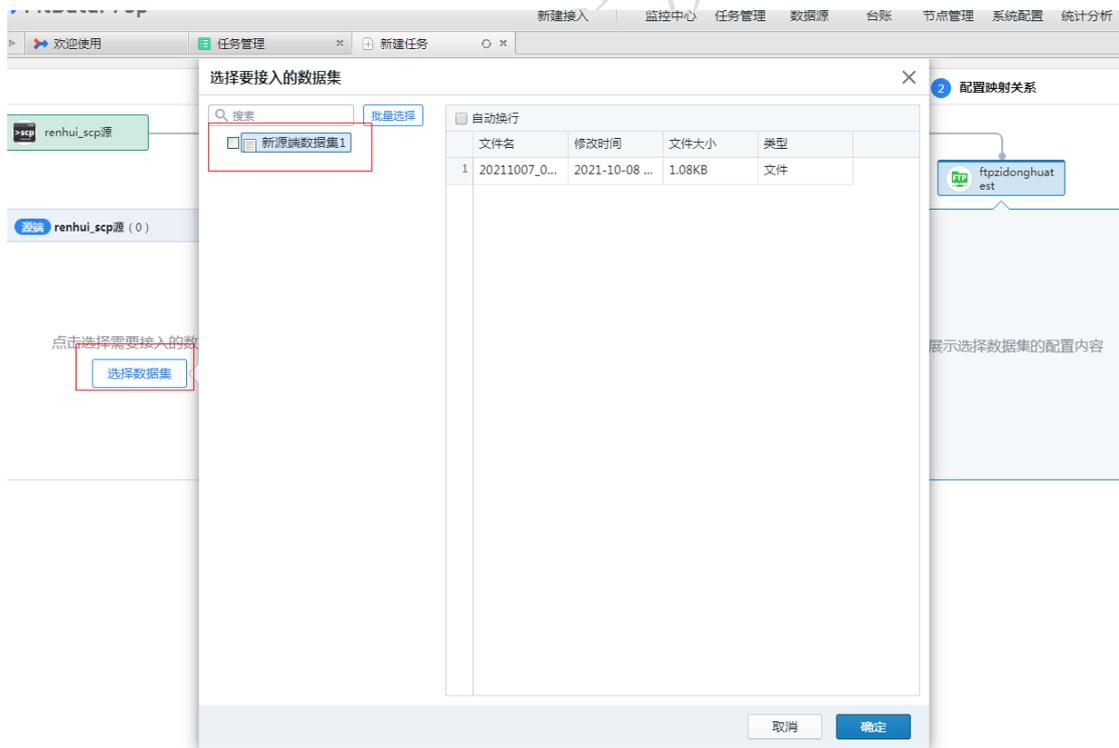


同步策略配置。

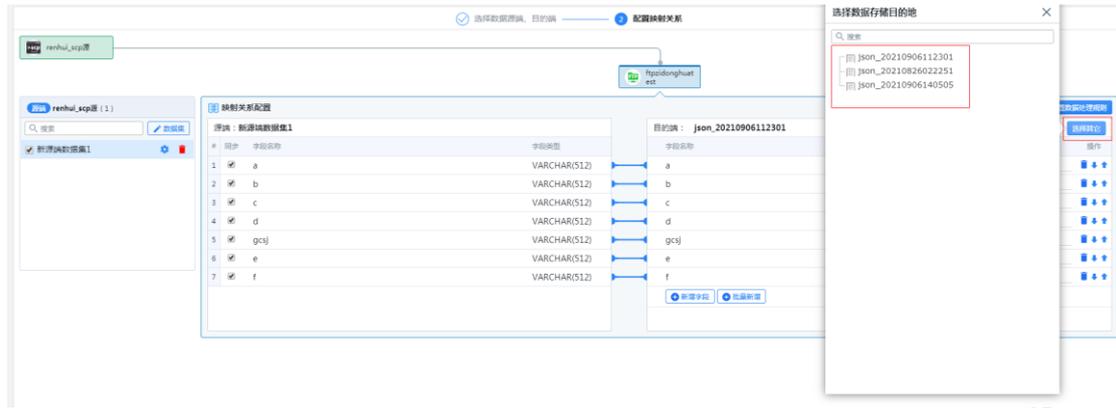


3.4.2.2.4 文件到文件的映射

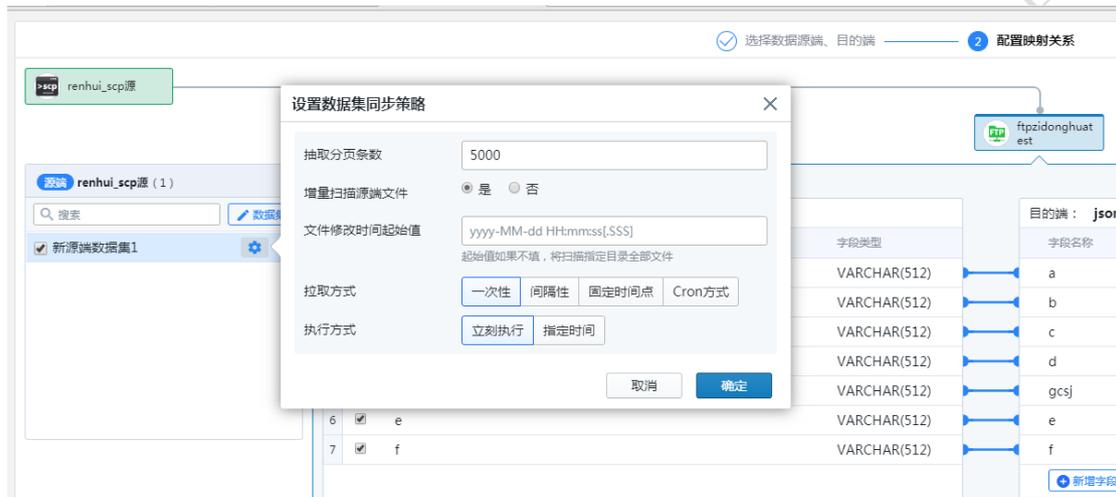
选择源端数据集。



展示源端数据集中的字段名称和字段类型等信息，点击目的端【选择其他】按钮，选择目的端数据集和源端数据集映射，配置源和目的端的字段映射关系，修改结束点击【完成】进行保存。



同步策略配置。



3.4.2.3 任务明细

点击完成按钮进入任务明细页面，可查看任务的初始状态。



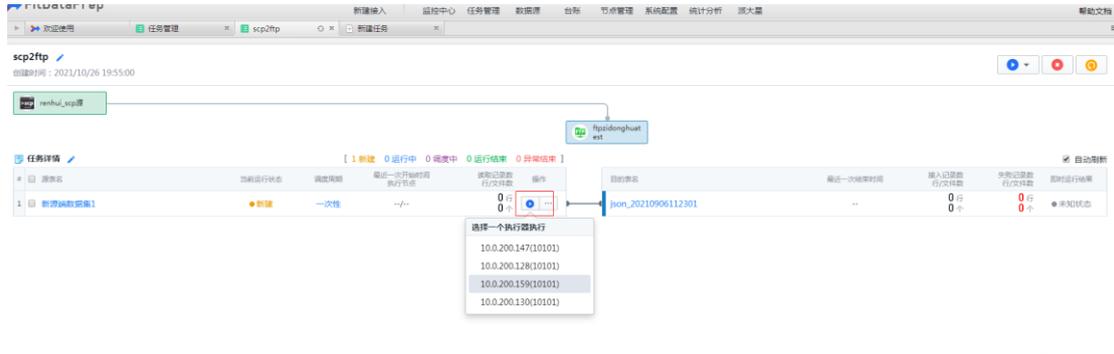
3.4.3 任务运转

3.4.3.1 任务的操作

3.4.3.1.1 任务的启动停止

明细页面展示任务执行状态和具体数据，并支持执行任务和停止任务操作。

点击左侧任务列表的  按钮，在下拉框中选择一个节点启动任务，任务启动完成。



运行中或者调度中的任务可以点击左侧任务列表的  按钮手动停止。



在任务详情界面的右上方，工具提供了一键启停所有选中的任务的快捷入口



3.4.3.1.2 任务的其他操作及日志的下载

点击任务详情左侧操作按钮 ，从下拉选项可以手动对每个运行过的子任务进行重置、删除、查看执行历史、下载错误数据和下载日志文件操作。

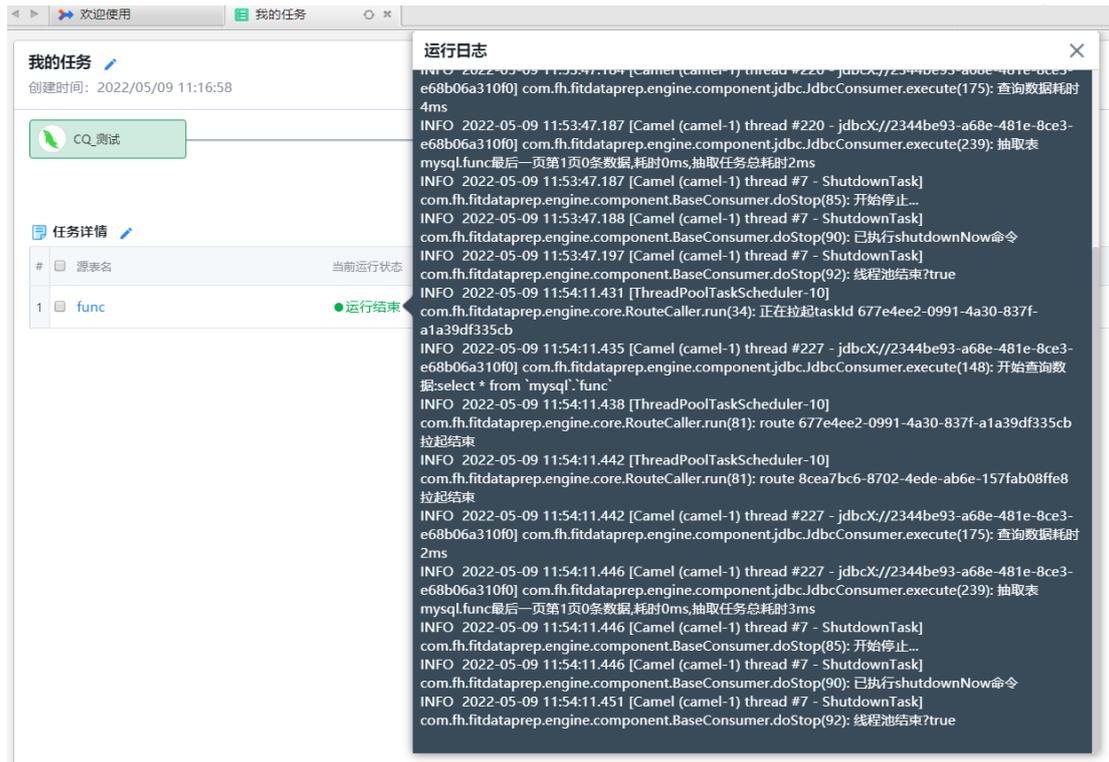


3.4.3.2 任务源和目的日志的查看

为了更快速的定位问题，任务运行的时候会将源和目的路由的日志分别写入两个文件中：源路由日志和目的路由日志。

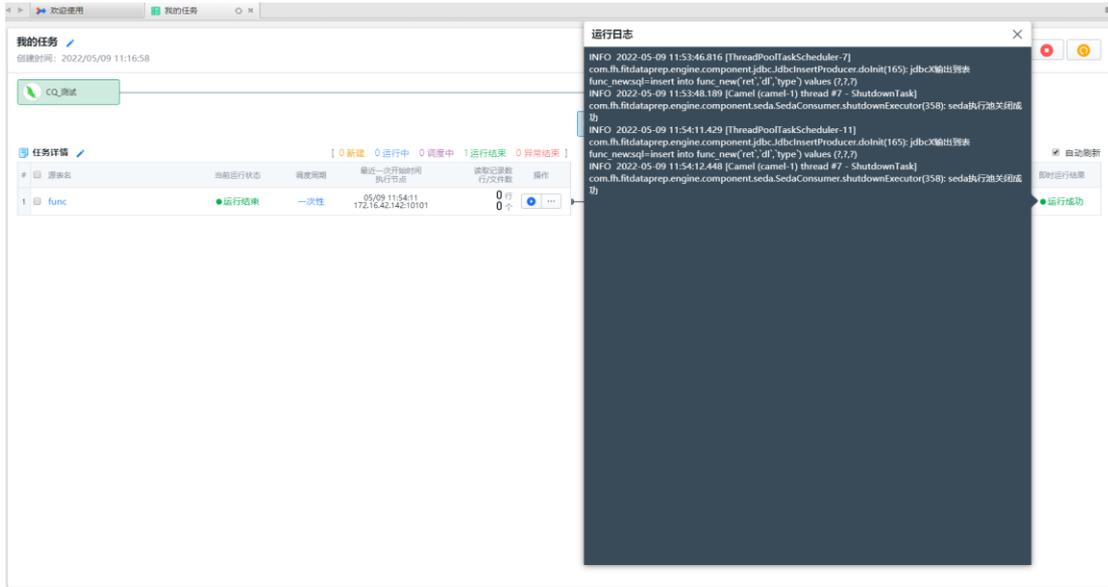
3.4.3.2.1 源路由日志

点击当前运行状态一栏的记录，可查看对应任务的源路由的部分执行日志。



3.4.3.2.2 目的路由日志

点击任务运行结果栏下的记录可以查看接入引擎目的路由的部分执行日志。

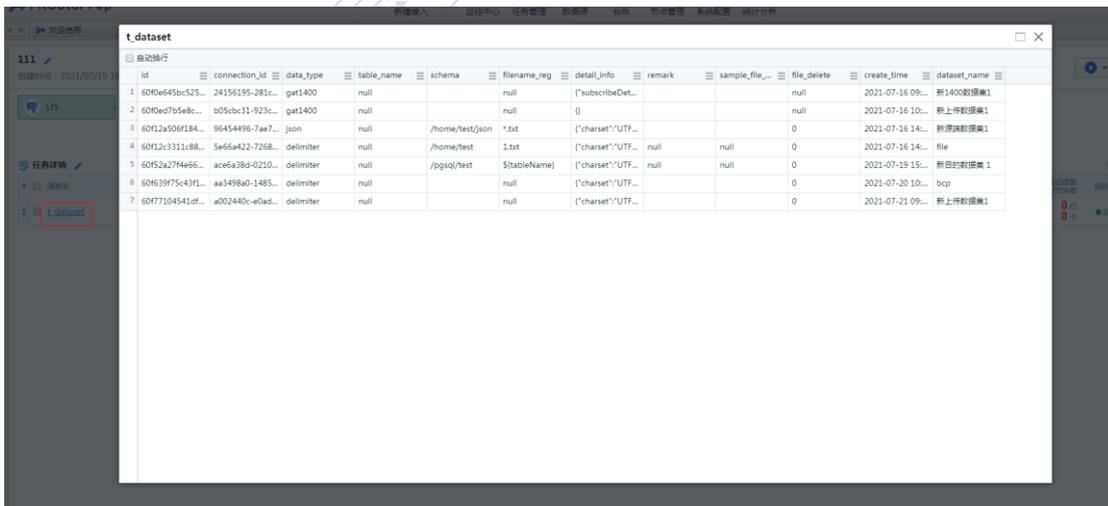


3.4.3.3 表数据预览

点击表名可预览源表或目的表的数据或者文件列表。



表数据的预览：



3.4.3.4 任务自动刷新

勾选任务详情右侧的 **自动刷新** 可以实时刷新任务的运行状态，显示运行结果。



3.4.3.5 任务批量启动、批量停止和批量重置

多选按钮勾选可以支持任务批量启停以及重置操作。



3.4.3.6 任务按状态筛选

支持按任务状态进行筛选操作，点击上方状态栏标签进行任务筛选。

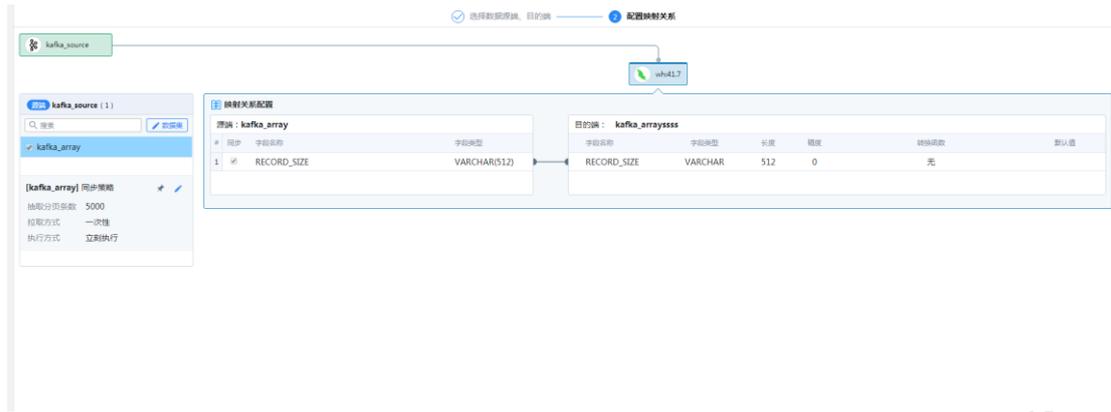


3.4.3.7 编辑任务

点击任务详情旁边的图标按钮进入编辑页面。



支持重新配置任务名称，增加目的端，配置映射关系和同步策略操作。（已完成的无法修改已有的映射关系）。



3.4.4 批量导入、导出任务和删除任务

支持对任务进行导入导出操作，支持多选批量导出，支持批量删除。



3.5 台账

3.5.1 台账列表

点击工具上方导航栏  按钮进入任务台账页面，该页面支持查看任务和子任务维度数据接入台账的统计，支持按日查看增量数据，支持查看每个任务统计增量数据。

支持按任务名称，源端，目的端，表名，开始结束日期等联合查询台账信息。设置查询条件的范围。

点击接入输出文件数，可以查看具体输入输出文件列表：

#	任务名	频率周期	源端	目的端	源表名	结构名	目的表名	输入文件数(个)	输入文件总大小	输入失败文件数(个)	输出文件数(个)
2	test (7)	一次性: 立即执行	55.2-oracle	205目的	ABC_CDC_HISTORY		excel	0	0	0	2
3		一次性: 立即执行	55.2-oracle	查看目的端	ABC_CDC		STG_ABC_CDC	0	0	0	1
4		一次性: 立即执行	44.180ttmc_dev	ttmc_centerTest	aa		adm_gov_tmc_resul...	0	0	0	0
5		间隔性: 间隔时间: 30min	SCP源端测试-副本-副...	scp77目的端-副本	source		scp77目的端	3099	3.48 GB	0	32
6		间隔性: 间隔时间: 30min	postgresj-source	postgresj-target	base_stage_info		base_stage_info_new	0	0	0	0
7		一次性: 立即执行	指标-需求22年气运	指标测试_中心库	指标-需求22年气运		adm_gov_tmc_resul...	2	1.38 KB	0	0
8		一次性: 立即执行	单元端	205目的	数据表1		excel	1	0	0	1
10	TEST (3)	一次性: 立即执行	HDFS-TEST	FMD8-TEST	base_category_info		base_category_info	1	420 B	0	1
11		一次性: 立即执行	HDFS-TEST	FMD8-TEST	base_version_table		base_version_table	1	214 B	0	1
12		一次性: 立即执行	HDFS-TEST	FMD8-TEST	新增数据表1		XVDSJL	1	218 B	0	1

输出文件列表

#	文件名	写成功记录数	写失败记录数
1	/data/home/czw_target/1.bcp	999999	0

3.5.2 台账明细

点击明细按钮展示该任务某一个子任务的明细信息:

任务运行详情

源端: source_hdfs 目的端: 205文件目的地 源表名: db2_source 目的表名: 新目的数据表 2 时间: 2021-10-12 14:14:32 - 2021-10-12 14:14:50

#	输入文件数(个)	输入记录数(行)	输出文件数(个)	失败文件数(个)	输出记录数(行)	输出失败记录数(行)	更新记录数(行)	删除记录数(行)	开始时间	结束时间
1	0	0	0	0	0	0	0	0	2021-10-12 14:...	2021-10-12 14:...
2	2	999999	202	0	999999	0	0	0	2021-10-12 14:...	2021-10-12 14:...

支持查看对应执行记录的输入输出文件:

任务运行详情

源端: source_hdfs 目的端: 205文件目的地

#	输入文件数(个)	输入记录数(行)	输出文件数(个)
1	0	0	0
2	2	999999	202

输入文件列表

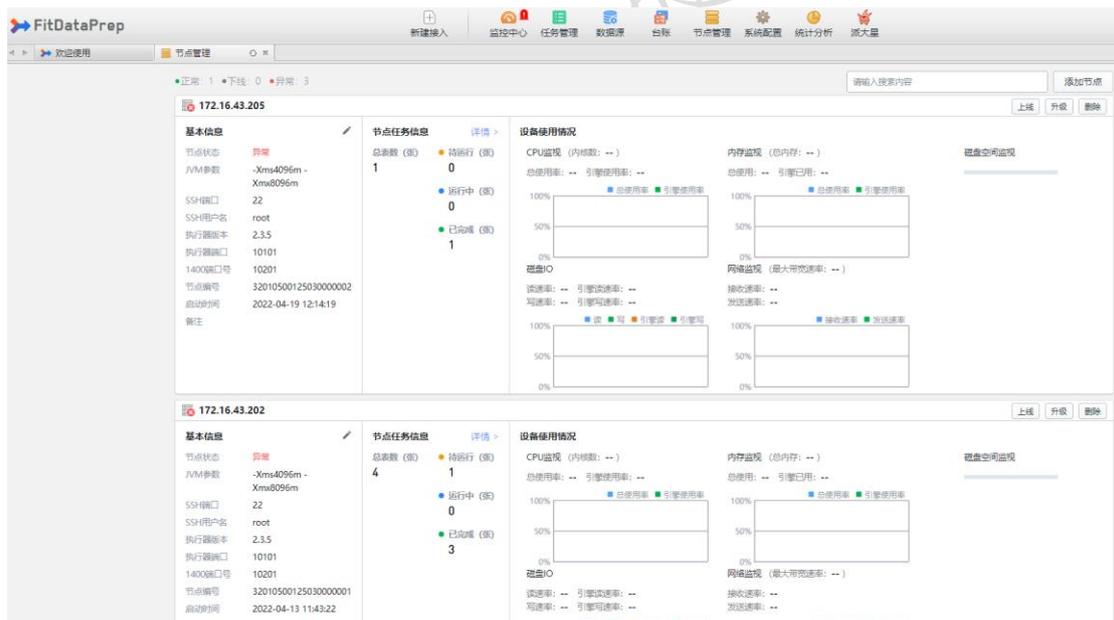
#	文件名	文件大小
1	/data/wh_test/FDP_TEST_100W_20211011_0183.bcp	9.76MB
2	/data/wh_test/FDP_TEST_100W_20211011_0001.bcp	100MB

3.5.3 台账导出

点击台账页面左上角的 **导出Excel** 按钮,支持把台账列表数据导出成 excel 文件:

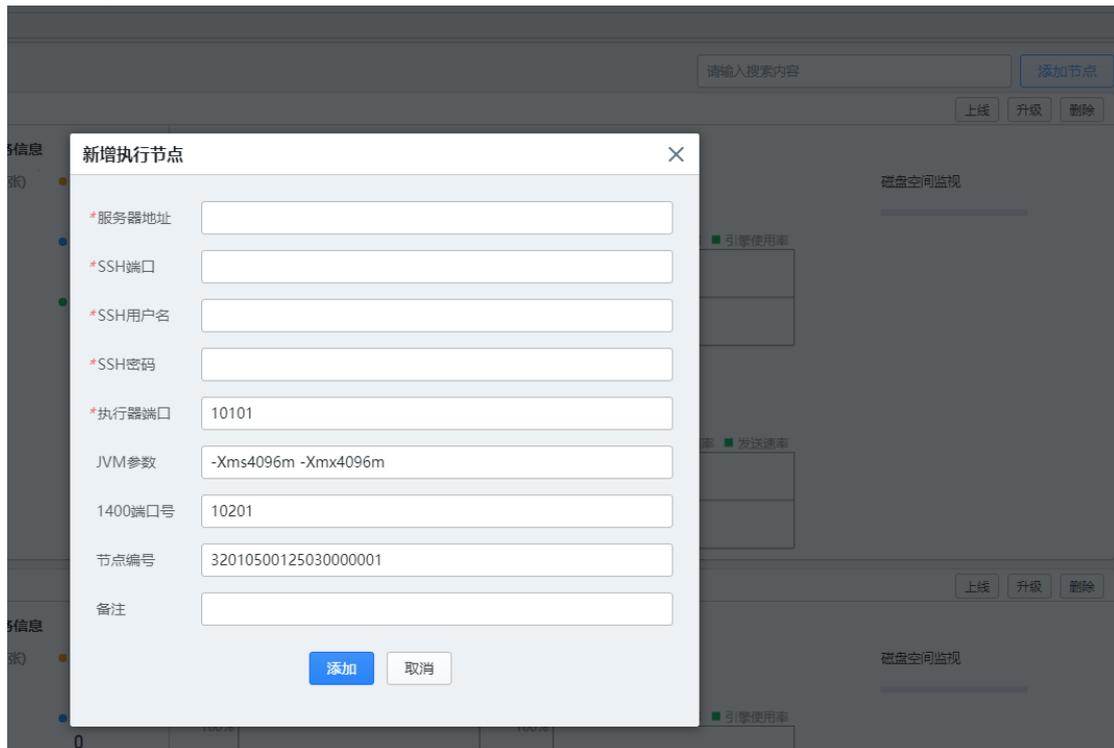
3.6 节点管理

需要添加新的节点或者对已有的节点进行调整可以点击工具上方导航栏的 **节点管理** 按钮,配置节点的连接信息和执行器信息。节点管理对每个节点按 3 列的方式列出所有节点的信息,左侧为节点的基本信息,中间为节点中已经运行的任务数量和运行状态,右侧为节点设备的 CPU、内存、磁盘等使用情况。在右上角列出每个节点的操作按钮。在节点管理页面的右上角提供了根据节点名称查询节点的输入框。



3.6.1 添加节点

在节点管理页面的右上角点击 **添加节点** 按钮即可打开新增节点的页面。填写节点相关信息后点击 **添加** 按钮完成节点添加,回到节点管理即可看到新节点的信息。

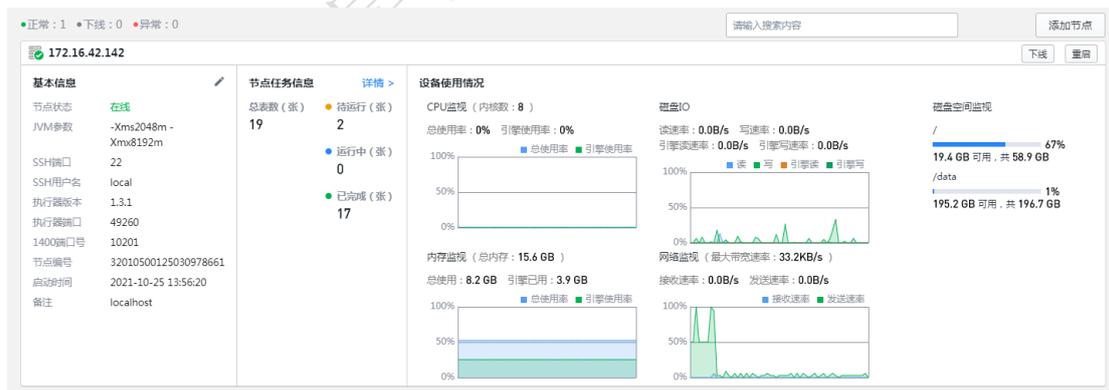


服务器地址:	填写需要增加节点的服务器 IP 地址。每台服务器只能添加一个节点。
SSH 端口:	FDP 管理中心需要自动登录节点所在服务器进行执行节点程序的自动安装，需要配置 SSH 登录端口。一般默认端口为 22。
SSH 用户名:	FDP 管理中心需要自动登录节点所在服务器进行执行节点程序的自动安装，需要配置 SSH 登录用户名。
SSH 密码:	FDP 管理中心需要自动登录节点所在服务器进行执行节点程序的自动安装，需要配置 SSH 登录密码。
执行器端口:	执行器安装后需要与管理中心进行socket通信，需要占用一个通信端口，默认端口为10101，一般不需要修改。

JVM参数:	<p>执行器节点使用 JAVA 运行环境，需要配置执行器 JAVA 环境的 JVM 参数。可以根据服务器的实际情况增加配置。JVM 最大内存设置参考公式：</p> <p>系统当前总内存 - fdp web 程序（固定 8G） - 其他程序所占内存 - 系统预留内存（建议总内存的 3 - 5G）。</p> <p>方法：JVM 内存设置不要超过当前可用内存与系统预留内存的差。</p>
1400端口号:	仅在 1400 数据源接入的情况下才需要关注。
节点编号:	FDP为每个节点生成一个编号，一般不需要修改。
备注:	可以为节点增加一个备注信息，便于管理。

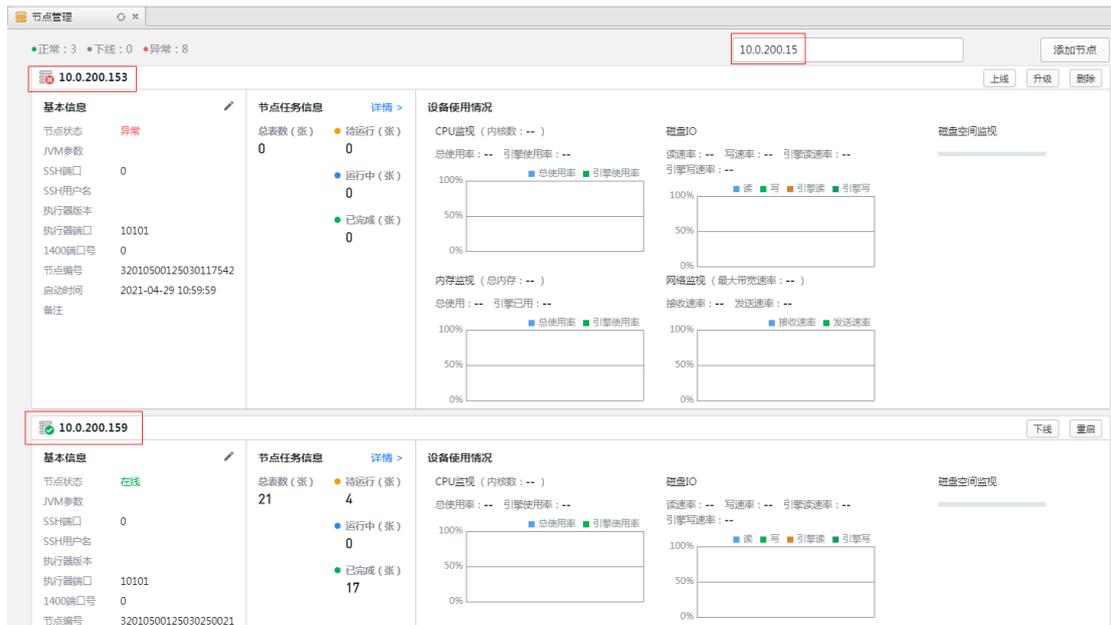
3.6.2 节点列表

该页面支持查看每个节点的能力和配置信息，以及监视每个节点服务器的资源动态。



3.6.3 节点查询

该页面支持查询节点，通过输入节点 ip 可模糊查询节点。



3.6.4 节点的操作

在节点管理界面支持对节点进行上线、下线、重启、升级和删除操作。

(1) 节点在下线状态支持上线、升级和删除操作。(当 web 升级完时, 引擎也需要升级时, 需要先将节点下线再做升级操作)。

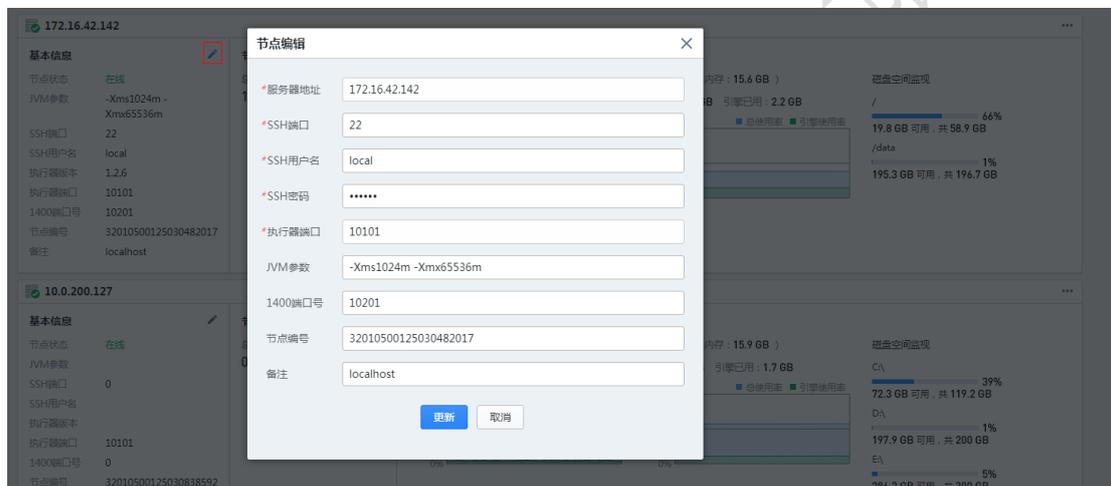


(2) 节点在上线状态支持下线和重启操作, 该节点如果有任务正在运行需等待任务运行完成或者手工停止所有正在运行的任务方可将该节点下线或者重启。



3.6.5 编辑节点

点击基本信息右侧的  图标编辑节点。



注意：在线状态不允许修改 JVM 参数和 1400 端口号

3.6.6 节点详情

点击节点任务信息右侧的 [详情 >](#) 按钮，进入节点详情页面：



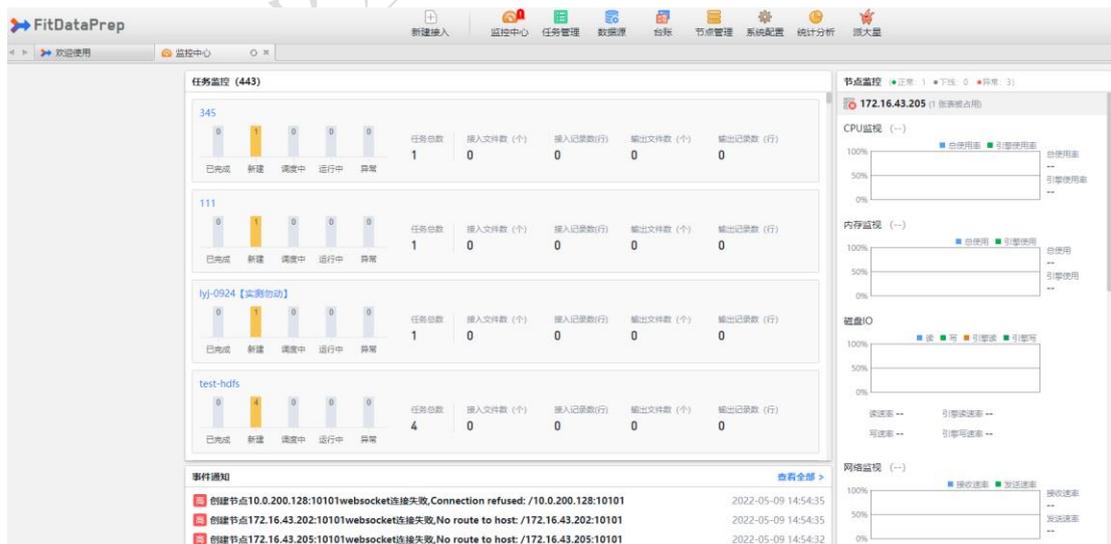
该页面支持查看每个节点上在跑的所有任务，支持启动/暂停任务。



3.7 监控中心

工具提供了完善的监控告警功能，点击工具上方导航栏的  按钮，即可打开监控界面，当系统有告警信息，该图标上会亮起左右摇晃的告警 。

监控节点左侧上方为任务监控，下方为事件通知，右侧为节点监控。



3.7.1 任务监控

该模块展示任务的状态以及任务的明细信息。



3.7.2 事件通知

该模块监控各个任务的执行情况，针对异常的任务进行告警通知。

事件通知		查看全部 >
高	创建节点10.0.160.102:10101websocket连接失败,Connection refused: no further information: /10.0.160.102:10101	2020-10-20 15:09:18
高	创建节点10.0.202.177:10101websocket连接失败,Connection refused: no further information: /10.0.202.177:10101	2020-10-20 15:09:17
高	创建节点10.0.200.121:10101websocket连接失败,Connection timed out: no further information: /10.0.200.121:10101	2020-10-20 15:09:16
高	创建节点10.0.200.147:10101websocket连接失败,Connection refused: no further information: /10.0.200.147:10101	2020-10-20 15:04:58
高	创建节点10.0.200.127:10101websocket连接失败,Connection timed out: no further information: /10.0.200.127:10101	2020-10-20 15:04:57
高	创建节点10.0.20.45:10101websocket连接失败,Connection timed out: no further information: /10.0.20.45:10101	2020-10-20 15:04:04
高	数据源: 172.17.55.2 表名: TEST_BASICMYSQL 建表sql: [CREATE TABLE TEST_BASICMYSQL(test_primary_int NU...	2020-10-20 15:01:40
高	数据源: 172.17.55.2 表名: TEST_BASICMYSQL 建表sql: [CREATE TABLE TEST_BASICMYSQL(test_primary_int NU...	2020-10-20 15:01:31
高	数据源: 172.17.55.2 表名: TEST_BASICMYSQL 建表sql: [CREATE TABLE TEST_BASICMYSQL(test_primary_int NU...	2020-10-20 14:58:03
高	数据源: 172.17.55.2 表名: TEST_BASICMYSQL 建表sql: [CREATE TABLE TEST_BASICMYSQL(test_primary_int NU...	2020-10-20 14:53:30

鼠标悬浮于事件上会显示标记已处理按钮，点击按钮之后该事件将不会显示在事件通知之中。

事件通知 查看全部 >

- 高 创建节点172.16.40.54:10101websocket连接失败,Connection refused: /172.16.40.54:10101 2021-03-05 13:52:47 标记已处理
- 高 创建节点10.0.200.147:10101websocket连接失败,Connection refused: /10.0.200.147:10101 2021-03-05 10:58:47
- 高 创建节点10.0.200.146:10101websocket连接失败,Connection refused: /10.0.200.146:10101 2021-03-05 09:39:17
- 高 创建节点10.0.200.127:10101websocket连接失败,connection timed out: /10.0.200.127:10101 2021-03-05 09:39:17
- 高 预处理获取worker节点数量为0,请检查配置【172.16.43.107:2181】服务是否正常运行! 2021-03-05 09:29:56
- 高 创建节点10.0.160.102:10101websocket连接失败,connection timed out: /10.0.160.102:10101 2021-03-05 09:25:17
- 高 数据源: oracle 表名: wu_test 建表sql: [CREATE TABLE wu_test("CAPTURE_TIME" NVARCHAR2(255), "_QUE... 2021-03-04 19:53:56
- 高 本地引擎: null 2021-03-04 19:53:35
- 高 数据源: oracle 表名: wu_test 建表sql: [CREATE TABLE wu_test("CAPTURE_TIME" NVARCHAR2(255), "_QUE... 2021-03-04 19:53:04
- 高 数据源: oracle 表名: wu_test 建表sql: [CREATE TABLE wu_test("CAPTURE_TIME" NVARCHAR2(255), "_QUE... 2021-03-04 19:52:59

点击蓝色的数据源告警通知可以跳转至该数据源产生告警对应的任务明细页面。

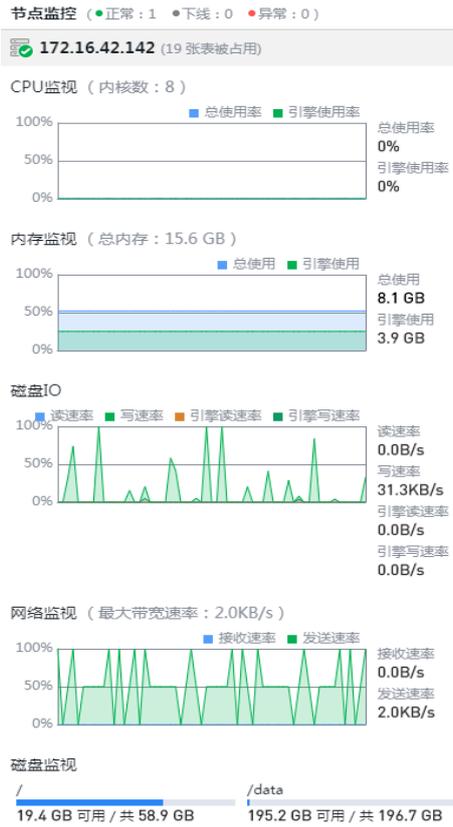
20 高 数据源: fdp_test 表名: test_user_table 建表sql: [CREATE TABLE chenfei.test_user_table(username VARCHAR(500), "password" VARCHAR(500));] 异常信息: 执行更新异常:ERROR: relation "test_user_table" already exists 2021-06-24 15:49:49 标记已处理

scp_delimiter_20210624154044 创建时间: 2021/06/24 15:40:44

任务名称	当前运行状态	调度周期	最近一次开始时间	最近一次结束时间	读取记录数	写入文件数	操作	目标任务	最近一次结束时间	写入记录数	删除记录数	删除文件数	当前运行结果
delimiter_20210624154043	● 运行结束	一次性	06/24 15:50:28	172.26.42.142:10101	0	0	...	test_user_table	06/24 15:50:29	0	0	0	● 运行成功

3.7.3 节点监控

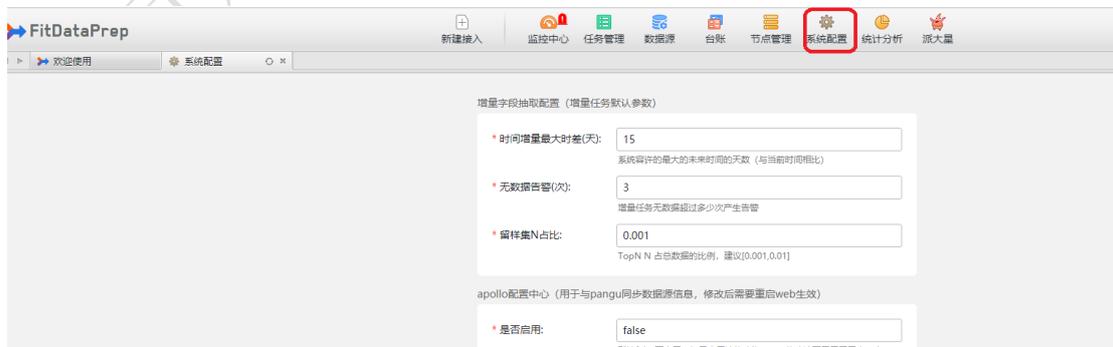
该模块对执行器引擎的资源使用情况进行实时监控(包括 cpu, 磁盘, 内存、网络), 以使用户在任务下发时, 可以选择避开负荷较大的节点。



3.8 系统配置

3.8.1 配置系统参数

FDP 工具可以对必须的系统参数进行配置，点击工具上方导航栏的  **系统配置** 按钮可以打开系统配置界面。支持配置以及还原默认值。



增量字段抽取配置 (增量任务默认参数)

时间增量最大时差 (天):	增量配置需要配置每次增量的增量大小, 工具容许的最 大的未来时间的天数 (与当前时间相比)。
无数据告警(次):	FDP 工具会在抽取数据多次无数据的情况进行告警, 此 处配置阈值。
留样集N占比:	TopN N 占总数据的比例, 建议[0.001,0.01]
apollo配置中心 (用于与pangu同步数据源信息, 修改后需要重启管理中心生 效)	
apollo配置中心相关参数获取请参考2.8.2小节介绍。	
是否启用:	默认false不启用, 如果启用请修改为true, 修改该配 置无需重启管理中心。
页面访问地址:	apollo页面访问地址
配置服务地址:	apollo配置服务地址
appId:	创建项目填写的应用Id
appToken:	创建第三方应用生成的token
命名空间:	项目使用的命名空间, 默认不修改
用户名:	项目登录用户名
F-pointer数据源参数 (设置F-pointer数据源相关参数)	
fp导出sql可执行 时间(分):	fp数据源导出sql执行超时, FP大数据量导出需要消耗 集群资源, 速度较慢, 建议将该参数适当调大。
1400服务配置 (1400服务参数配置)	
用户名:	1400服务账号
密码:	1400服务密码
任务限速控制 (设置任务处理速度)	

关系型数据库抽取速度：	在数据库压力较大的情况下可以使用此参数进行限速。 小于等于0表示不限速,单位:条/s
-------------	---

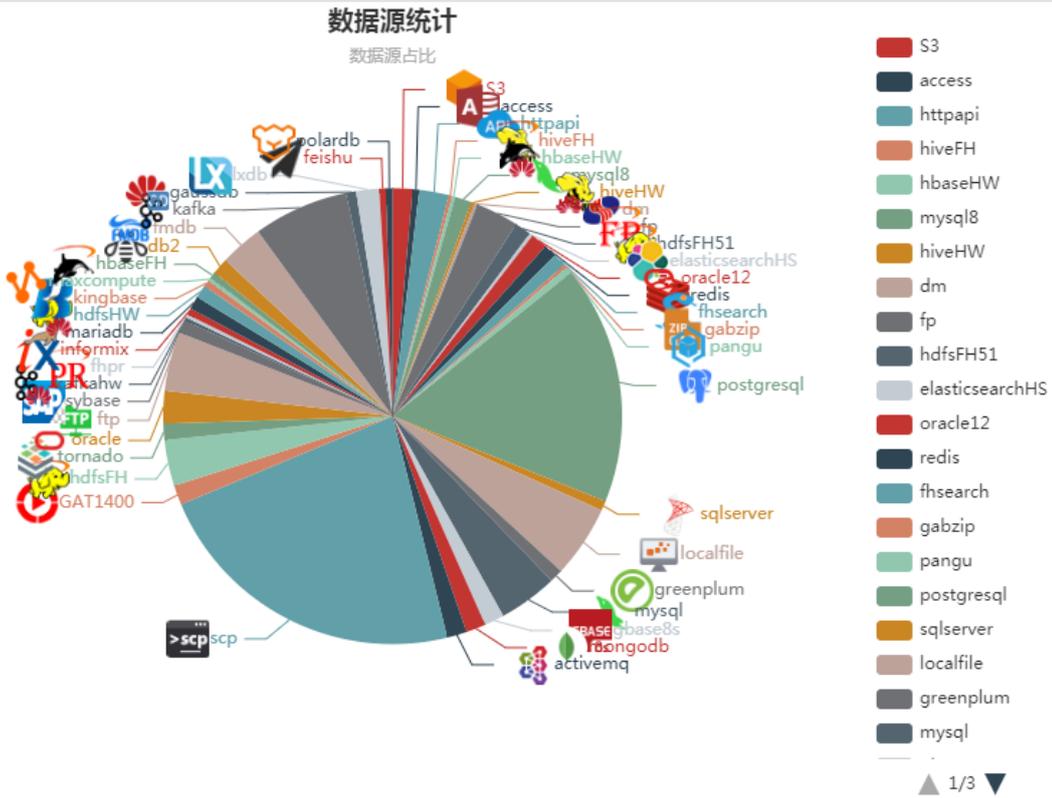
3.9 统计分析

FDP 工具提供了接入所需的统计信息查看。点击工具上方导航栏  按钮进入数据源统计分析页面，包括数据传输量趋势、数据源统计分析和场景列表三部分。

数据传输量趋势以折线图的形式展示 fdp 接入数据量的变化，默认展示所有任务在最近 24 小时内的接入数据量变化，同时可以对具体任务的数据接入量，通过设置不同时间（小时、天、月）条件进行筛选展示。



数据源统计部分，鼠标悬浮到饼图的具体类型数据源上面可以显示本类型数据源使用数量和总占比，右侧是数据源类型的图例，点击具体的数据源类型可以使其在饼图上展示或者从饼图上删除该类型的数据源。



场景列表是展示不同类型数据源组合在实际任务的使用情况,还可以导出数据源的统计信息

场景列表

#	源端	目的端	任务总数	接入文件数(个)	输出文件数(个)	接入记录数(行)	输出记录数(行)
1	mysql	mysql	1054	0	0	989816	971348
2	postgresql	postgresql	299	0	0	3770560	3755475
3	mysql	postgresql	266	0	0	751989	751970
4	fhsearch	fmdb	142	0	0	180	180
5	oracle	postgresql	98	0	0	18347419	18347383
6	scp	postgresql	82	23350	0	805601752	805805728
7	localfile	postgresql	66	46	0	1488680	1513545
8	postgresql	scp	62	0	96	29796874	29796811
9	ftp	postgresql	51	247093	0	123433667	57107405
10	scp	scp	46	3535	113290	2565253007	2565215216
11	oracle	scp	38	0	1659	955001	947501
12	hdfsFH	fmdb	35	31	0	270	182
13	localfile	fmdb	32	21	0	24031	20031
14	postgresql	ftp	31	0	19	32758	32758
15	scp	fp	27	4791	4	2272082	190
16	postgresql	fp	23	0	5	2312	1282
17	oracle	mysql	20	0	0	2	0