DCT v3.X 用户操作手册

目录

1 快速指南.....4

1.1 产品介绍.....4

1.2 名词解释.....4

1.3 注意事项.....5

2 部署环境以及数据源权限设置.....5

2.1 部署环境要求.....5

2.2 数据源权限设置.....7

2.2.1 Mysql binlog 权限设置.....7

2.2.2 Oracle Logminer 权限设置.....8

2.2.3 其他数据源权限配置持续更新中.....13

3 操作详解.....13

3.1 登录.....13

3.1.1 账户登录.....13

3.1.2 账号申请.....14

3.1.3 退出登录.....14

3.2 数据源管理.....15

3.2.1 新建数据源-入口.....15

3.2.2 数据源筛选.....19

3.2.3 数据源编辑 19
3.2.4 批量新增 20
3.2.5 数据源删除 23
3.2.6 批量删除 23
3.3 数据同步 23
3.3.1 离线同步简介 23
3.3.2 离线同步支持的数据源 24
3.4 数据实时同步 24
3.4.1 实时同步简介 24
3.4.2 实时同步支持的数据源 25
3.5 新建任务 25
3.5.1 新建任务入口 25
3.5.2 创建任务 26
3.5.3 任务配置 27
3.6 任务锁定/解锁 55
3.7 任务管理 57
3.8 数据修复【配置和离线任务一致】 61
3.8.1 新建数据修复任务 62
3.9 数据订阅【合并到实时任务，监听结果写入 kafka】 64
3.10 组件管理 64

3.11 运维管理.....73
3.11.1 集群资源监控.....73
3.11.2 运维大屏.....73
3.11.3 操作日志.....74
3.11.4 实例管理.....75
4 常见问题.....76
4.1 数据源或目标库测试连通性失败怎么办？.....76
4.2 单任务是否支持多个表是否可以写到目的地一张表？.....76
4.3 如何应对 Mysql 数据库表和字段名称大小写不敏感问题？.....76
4.4 Hive 数据源支持哪些文件格式？.....77
4.5 如何避免 Hive 目的地出现小文件问题？.....77
4.6 常见问题持续更新中.....77

1 快速指南

1.1 产品介绍

DCT 是数据采集传输工具，一款可用于公有云之间、公有云到私有云、私有云到公有云、私有云之间的数据集成、迁移、汇聚的工具，支持离线数据同步（全量/增量）、实时数据同步、数据订阅等功能。

1.2 名词解释

(1) 向导模式：提供向导式的开发引导，通过可视化的填写和下一步的引导，助您快速完成数据同步任务的配置工作。向导模式的学习成本低，但无法支持部分高级功能。

(2) 脚本模式：您可以通过直接编写数据同步的 SQL 来完成数据同步配置，适合高级用户，学习成本较高。脚本模式可以提供更丰富灵活的能力，实现精细化的配置管理。

(3) 并发数：并发数是数据同步任务中，可以从源并行读取或并行写入数据存储端的最大线程数。

(4) 限速：限速是数据集成同步任务可以达到的传输速度限制。

(5) 脏数据：脏数据是对于业务没有意义或者格式非法的数据。例如，源端是 VARCHAR 类型的数据写到 INT 类型的目标列中，导致因为转换不合理而无法写入的数据。

1.3 注意事项

(1) 官网用户开通 DCT 后，请联系管理员申请账号；私有化部署的，请登录 iam 后台创建账户

(2) 进行操作时请参考用户手册，避免错误的操作影响系统本身；

(3) 进行数据编辑工作时请务必及时保存，以免丢失数据；

(4) 注意用户手册的保存。

(5) **本次版本升级涉及租户权限，升级之前按需备份日志，升级过程中，会出现日志丢失的情况；本次组件配置有升级，为兼容老版本，保留了旧版本的配置，预计保留 3 个迭代版本，需按需将旧的配置迁移到新配置上，若 3 个版本后未迁移，可能影响任务的正常运行！！！（V3.0）**

2 部署环境以及数据源权限设置

2.1 部署环境要求

名称	要求	备注

<table>
<tr><td colspan="2">服务器数量</td><td>集群版 3 台或以上；单机版 1 台；</td><td>推荐集群版配置,单机版无法测试性能；</td></tr>
<tr><td rowspan="5">硬件配置</td><td>cpu</td><td>至少四核,推荐八核及以上；</td><td>无</td></tr>
<tr><td>内存</td><td>至少 16G,推荐 32G 及以上；</td><td>无</td></tr>
<tr><td rowspan="3">硬盘</td><td rowspan="3">华为云高性能云盘同等性能,容量 200G 及以上,建议使用 SSD 硬盘；</td><td>1.所需容量取决于拟测试的同步任务数据量。建议至少准备拟同步任务总数据量*2 大小的数据盘；</td></tr>
<tr><td>2.最好用空白硬盘单独挂在到服务器,单独为 DCT 使用；</td></tr>
<tr><td>3.文件系统为 XFS 的需要开启 ftype,命令如下:mkfs.xfs -n ftype=1 /path/to/your/device（磁盘设备）；</td></tr>
<tr><td colspan="2" rowspan="3">操作系统</td><td>CentOS 7.4 以上；</td><td rowspan="3">必须要安装 wget 命令</td></tr>
<tr><td>RHEL 7.4 以上；</td></tr>
<tr><td>Ubuntu 16.04 及以上；</td></tr>
</table>

网络/端口	1.部署服务器与数据源及数据目的地的特定端口可达（例如 mysql 的 3306）；	实际部署环境的可开放端口如有调整,请联系 DCT 工程师修改产品配置文件,或根据实际端口调整网络安全策略；
	2.Web 服务器对用户访问终端 IP 开放 80 端口；	
	3.如是集群版,需保证集群服务器之间所有网络端口可达；	

2.2 数据源权限设置

2.2.1 Mysql binlog 权限设置

2.2.1.1 Mysql binlog 权限设置

- 步骤 1：修改 my.cnf 添加或打开注销

```
[mysqld]
log_bin=ON #开启 binlog
server-id=123 #设置一个与其他 mysql 不相同的 id
binlog-format=row #设置 binlog 模式为 row
```

- 步骤 2：修改 mysql 配置文件后需要重启 mysql

- 步骤 3：修改 binlog_row_image 为 FULL

```
set global binlog_row_image ='full';
```

- 步骤 4：查看开启 binlog 是否生效、查看 binlog 模式和 binlog_row_image 是否为 FULL。使用 root 用户进入到 mysql

```
show variables like '%log_bin%';（需要设置为 ON）
show global variables like '%binlog_format%';（需要设置为 ROW 模式）
show variables like '%binlog_row_image%';（需要设置为 FULL）
```

- 步骤 5：创建用户和赋权

```
create user 'dp_test'@'%' identified by '12345678';
grant SELECT on mysql.db  to 'dp_test'@'%';
grant SELECT on mysql.tables_priv  to 'dp_test'@'%';
grant SELECT on mysql.user  to 'dp_test'@'%';
grant SELECT,LOCK TABLES on test.* to 'dp_test'@'%';
grant REPLICATION SLAVE,REPLICATION CLIENT on *.* to 'dp_test'@'%' identified by '12345678';(全局设置，需要*.*)
flush privileges;
```

2.2.2 Oracle Logminer 权限设置

2.2.2.1 Logminer 安装步骤：

- 步骤 1：开启归档日志，以下操作风险较大，请联系 DBA 工程师依次操作下列语句，开启数据库归档。

  a. shutdown immediate;

  b. startup mount;

  c. alter database archivelog;

  d. alter database open;

- 步骤 2：依次执行以下脚本：（最好由 DBA 操作，需要权限）

  a. 输入 @$ORACLE_HOME/rdbms/admin/dbmslm.sql

  b. 输入 @$ORACLE_HOME/rdbms/admin/dbmslmd.sql

  c. 输入 @$ORACLE_HOME/rdbms/admin/dbmslms.sql

  （注意：Oracle12c 经典安装中只有前两个，但并不影响使用）

```
SQL> @$ORACLE_HOME/rdbms/admin/dbmslm.sql
Session altered.

Package created.

Grant succeeded.

Synonym created.

Session altered.
SQL> @$ORACLE_HOME/rdbms/admin/dbmslmd.sql
Session altered.

Package created.

Synonym created.

Session altered.
SQL>
```

- 步骤 3：打开终端，连接服务器后输入「 su - oracle」切换至 Oracle 用户

- 步骤 4：输入 「sqlplus dp_test/oracle」 使用 dp_test 用户进入数据库（使用 LOGMNR 推荐使用 sys 用户[如需使用其他用户需要 dba 权限]）

```
[oracle@mdw ~]$ sqlplus system/oracle

SQL*Plus: Release 12.1.0.2.0 Production on Mon Feb 5 15:12:14 2018

Copyright (c) 1982, 2014, Oracle.  All rights reserved.

Last Successful login time: Mon Dec 18 2017 14:43:21 +08:00

Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application Testing opt
ions
```

a. 当看到「Conneted to: Oracle Database...」时即表示已经登陆成功，此时终端开头会变为「SQL

```
SQL>
```

b. 若使用 sys 以外的用户，进行后续操作前请联系 DBA 工程师赋予「grant EXECUTE_CATALOG_ROLE,DBA to user;」权限

- 步骤 5：Logminer 需要开启全字段日志补全功能（suppplemental logging，表级/列级），能够补全数据库日志信息，保证传递信息的全面性。（此操作会增加 Oracle 数据库的日志量）

  a.输入「**ALTER DATABASE ADD SUPPLEMENTAL LOG DATA;**」

  b.输入「**ALTER DATABASE ADD SUPPLEMENTAL LOG DATA(ALL,PRIMARY KEY,UNIQUE) COLUMNS;**」

c.输入

```
「SELECT supplemental_log_data_min min, supplemental_log_data_pk pk, supplemental_log_data_ui ui, supplemental_log_data_fk fk, supplemental_log_data_all allc FROM v$database;」
SQL> SELECT supplemental_log_data_min min, supplemental_log_data_pk pk, supplemental_log_data_ui ui, supplemental_log_data_fk fk, supplemental_log_data_all allc
FROM v$database;
MIN PK UI FK ALL
-------- --- --- --- ---
YES YES YES NO YES
```

d.当看到如上图所示内容时，表示已经成功。

2.2.2.2 赋权语句：

- 19c

```
CREATE USER <user name> IDENTIFIED BY <password>;
GRANT create session, alter session, logmining, execute_catalog_role TO <user name>;
GRANT select on GV_$DATABASE to <user name>;
GRANT select on V_$LOGMNR_CONTENTS to <user name>;
GRANT select on GV_$ARCHIVED_LOG to <user name>;
GRANT select on V_$LOG to <user name>;
GRANT select on V_$LOGFILE to <user name>;
GRANT select on V_$LOGMNR_LOGS to <user name>;
GRANT select on <db>.<table> TO <user name>;
```

- 11c 以下

```
CREATE USER <user name> IDENTIFIED BY <password>;
GRANT create session, alter session, execute_catalog_role, select any transaction, select any table to <user name>;
GRANT select on GV_$DATABASE to <user name>;
GRANT select on GV_$ARCHIVED_LOG to <user name>;
GRANT select on V_$LOGMNR_CONTENTS to <user name>;
GRANT select on <db>.<table> TO <user name>;
```

- 12c 或 18c

```
CREATE USER <user name> IDENTIFIED BY <password>;
GRANT create session, alter session, logmining, execute_catalog_role TO <user name>;
GRANT select on GV_$DATABASE to <user name>;
GRANT select on V_$LOGMNR_CONTENTS to <user name>;
GRANT select on GV_$ARCHIVED_LOG to <user name>;
GRANT select on <db>.<table> TO <user name>;
```

2.2.2.3 测试 Logminer

- 步骤 1：在登陆用户条件下输入「select name from V$ARCHIVED_LOG;」

  a.
  ```
  SQL> select name from V$ARCHIVED_LOG;
  ```

  b. 找到 archive log file

  c. 上图中内容 archive log file

- 步骤 2：任意一个 archive log file 作为 logmnr 测试分析使用

  a. 复制 archive log file 路径

  b. 输入 「execute sys.dbms_logmnr.add_logfile(logfilename=>'此处输入路径',options=>dbms_logmnr.new);」

  ```
  SQL> execute sys.dbms_logmnr.add_logfile(logfilename=>'/storage/oracle/app/or
  rc',options=>dbms_logmnr.new);

  PL/SQL procedure successfully completed.
  ```

  c. 输入「execute sys.dbms_logmnr.start_logmnr(Options =>

DBMS_LOGMNR.DICT_FROM_ONLINE_CATALOG)」

```
SQL> execute sys.dbms_logmnr.start_logmnr(Opt

 PL/SQL procedure successfully completed.
```

d. 输入「select count(*) from v$logmnr_contents;」解析日志记录条数

```
SQL> select count(*) from v$logmnr_contents;

  COUNT(*)
----------
     84208
```

e. 当看到数字时，即表示 logmnr 已经成功部署并且可以正常使用

2.2.3 其他数据源权限配置持续更新中

3 操作详解

3.1 登录

3.1.1 账户登录

访问 DCT 平台地址，在登录页面填写账号和密码，点击登录即可登录成功。

3.1.2 账号申请

3.1.2.1 官网 DCT 平台账号

- 官网自主注册账户，开通 DCT 服务，即可使用 DCT 服务
- 联系官网运营，开通账号，供用户体验试用

3.1.2.2 私有化部署账号

- 登录私有化部署 IAM 后台，创建账户

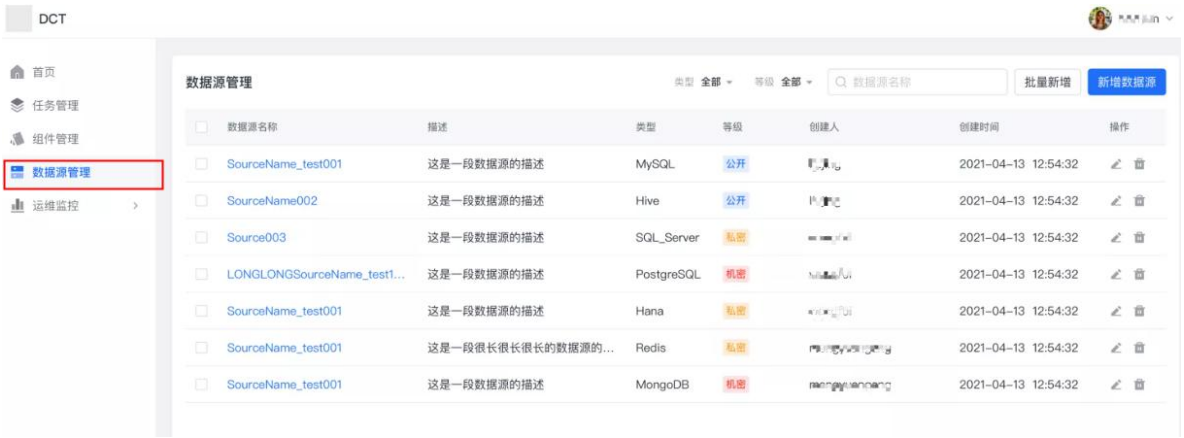
✱注：账号权限过期，会出现左侧菜单栏无法使用的情况，请联系官网运营，开通权限

3.1.3 退出登录

登录 DCT 平台后，若想退出当前账号，可点击用户账号头像，单击退出，即可返回登录页面。

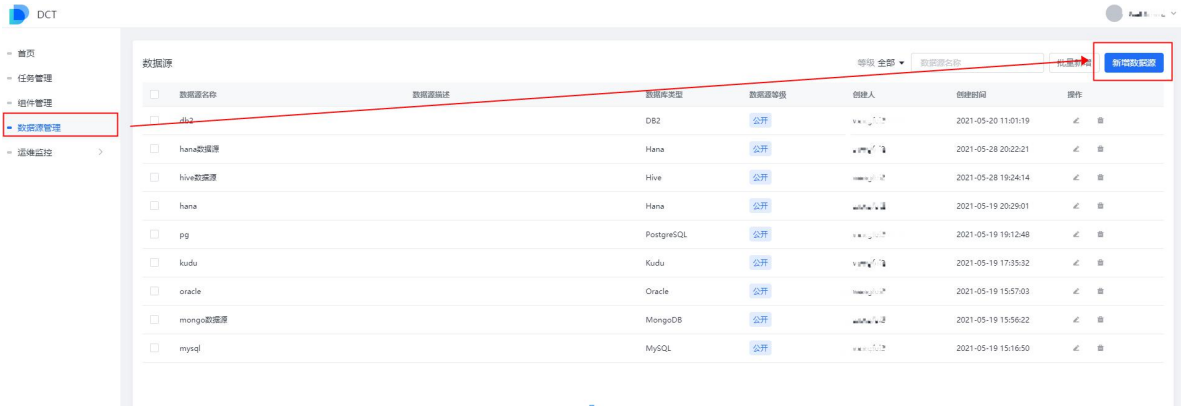
3.2 数据源管理

入口：点击左侧菜单栏“数据源管理”打开数据源管理列表页

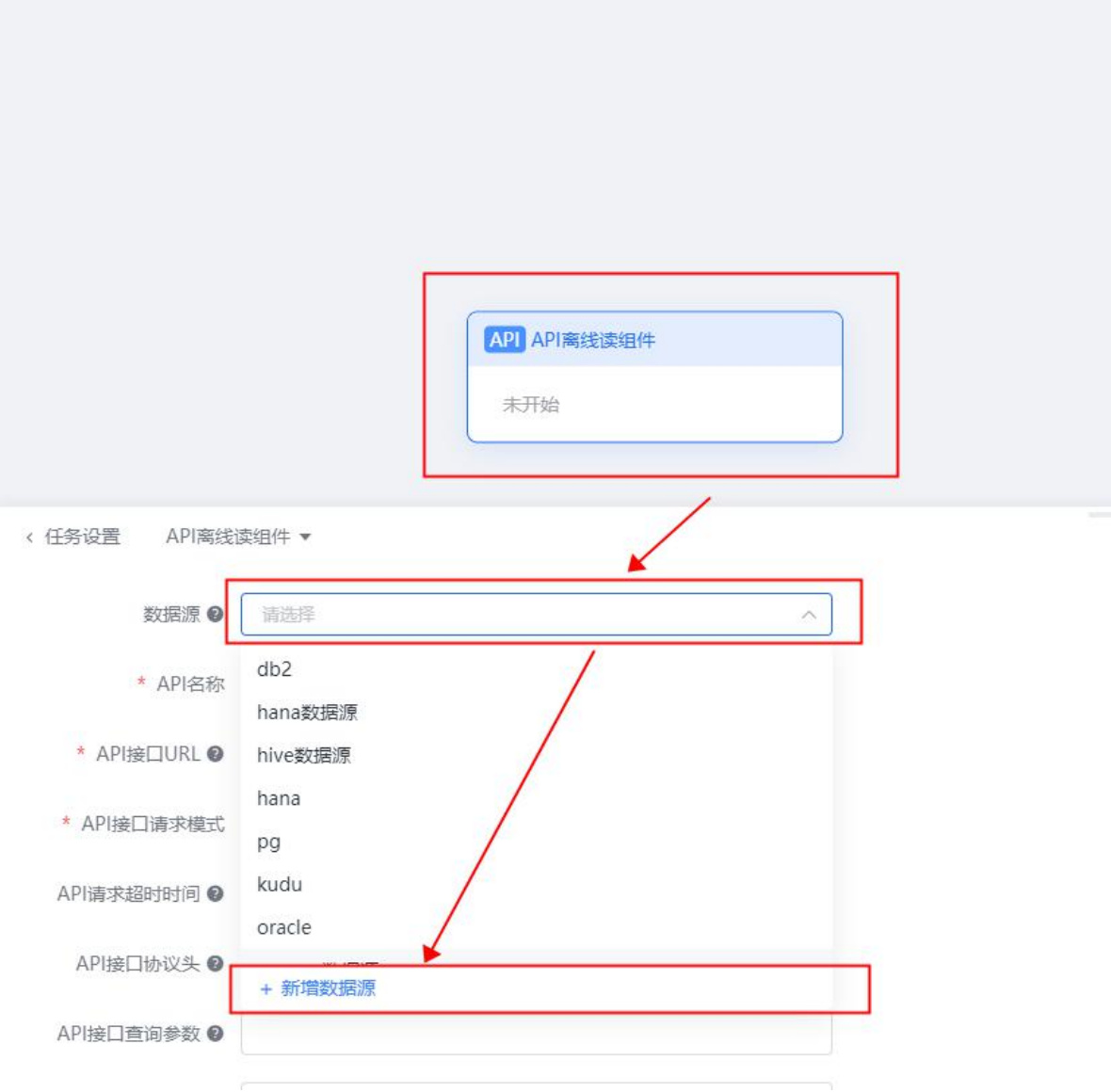


3.2.1 新建数据源-入口

3.2.1.1 入口 1：数据源管理页面



3.2.1.2 入口 2：组件配置-数据源--新增数据源



3.2.1.3 新建数据源-步骤

◆步骤 1：输入数据源名称；名称为数据源在 DCT 平台的唯一标识

◆步骤 2：选择数据源类型，支持下拉选择数据源类型

◆步骤 3：选择数据源等级，支持公开/私密/机密三种等级的数据源

◆步骤 4：输入数据源描述，用来区分标识数据源，便于查找

◆步骤 5：URL，必填项，用来连接数据源

◆步骤 6：数据源账号

◆步骤 7：数据源密码

◆点击【确定】完成数据源的创建；点击【取消】，取消数据源的创建

新增数据源

* 数据源名称

* 数据库类型

请选择

* 数据源等级

公开

描述

请输入

0/50

* URL

账号

密码

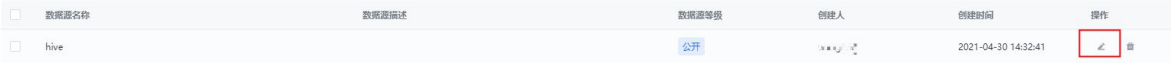
确定 取消

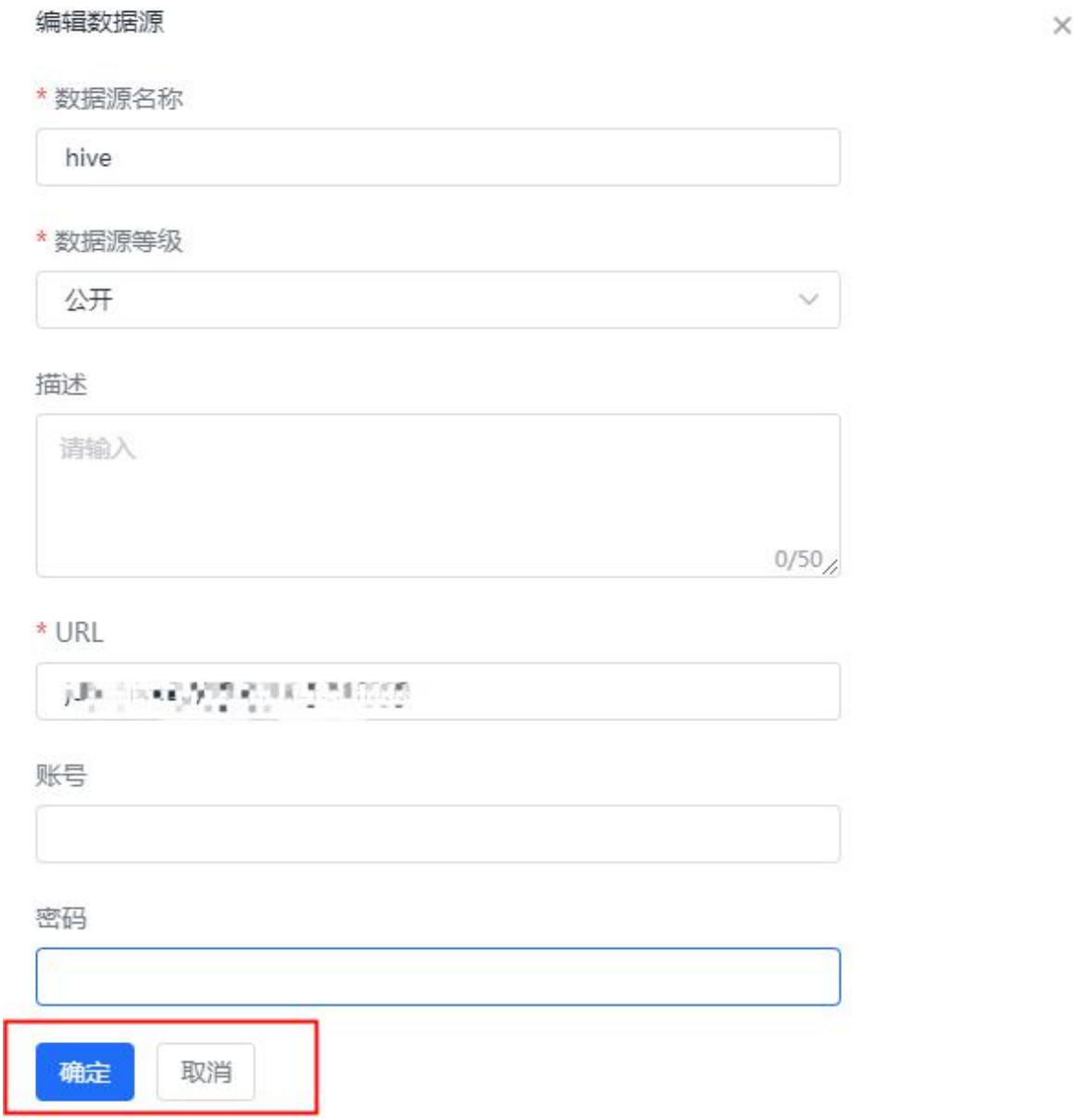
3.2.2 数据源筛选

可根据数据源等级、数据源名称进行数据源筛选，点【等级】按照下拉条件筛选数据源；点击【类型】按照下拉条件筛选数据源；输入【数据源名称】，回车，则按照数据源名称筛选。

3.2.3 数据源编辑

点击数据源列表中的【编辑】按钮，可对数据源的所有信息进行编辑，则可点击【确定】保存修改，若点击【取消】，则不保存本次修改。

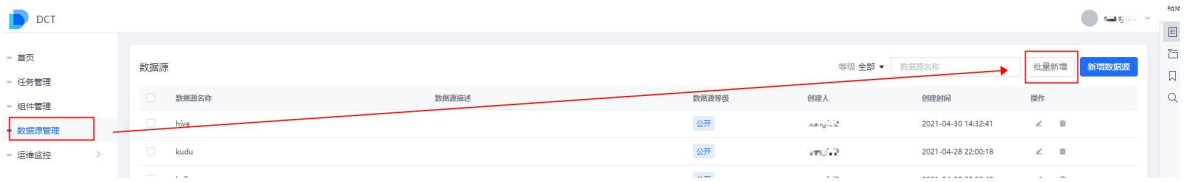




✱注意：若当前数据源正在执行任务或被占用，不可编辑。

3.2.4 批量新增

点击【批量新增】数据源按钮，打开批量新增入口



3.2.4.1 批量新增数据源的方式：拖拽/点击上传

- 拖拽：按住文件，将填写好的模板放入文件上传位置
- 点击上传：点击【点击上传】按钮，打开本地文件夹，选择文件上传
- 模板下载：点击【模板下载】下载模板

3.2.4.2 模板填写规范

- Name：输入数据源的名称，支持中英文和下划线、数字
- Type：填写数据源类型的代码
  - 10：oracel
  - 20: mysql
  - 30：sqlServer
  - 40：postgress
  - 50：hive
  - 60：HANA
  - 70：HBASE

- 80：MONGO_DB　s
- 90：FTP
- 100：HDFS
- ：KUDU
- ：REDIS
- ：达梦数据库
- ：人大金仓数据库
- 150：kafka
- 160:　DB2
- 170:　SYBASE
- 180：API

◆ DatasourceInfo：数据源的 url 信息；

◆ User：输入用户名，若有；

◆ Password：输入对应用户名的密码

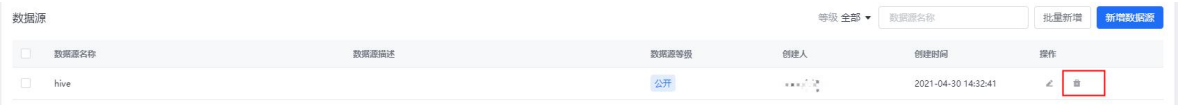
◆ Priority：优先级：10：公开　，20：私密　：30：机密

	A	D	C	D	E	F	G	H
1	name	type	datasourceInfo	user	password	flag	priority	

✱注：支持扩展名为.xls/.xlsx/.csv 的文件的文件

3.2.5 数据源删除

◆ 用户可点击任务列表的【删除】按钮，删除单个数据源。



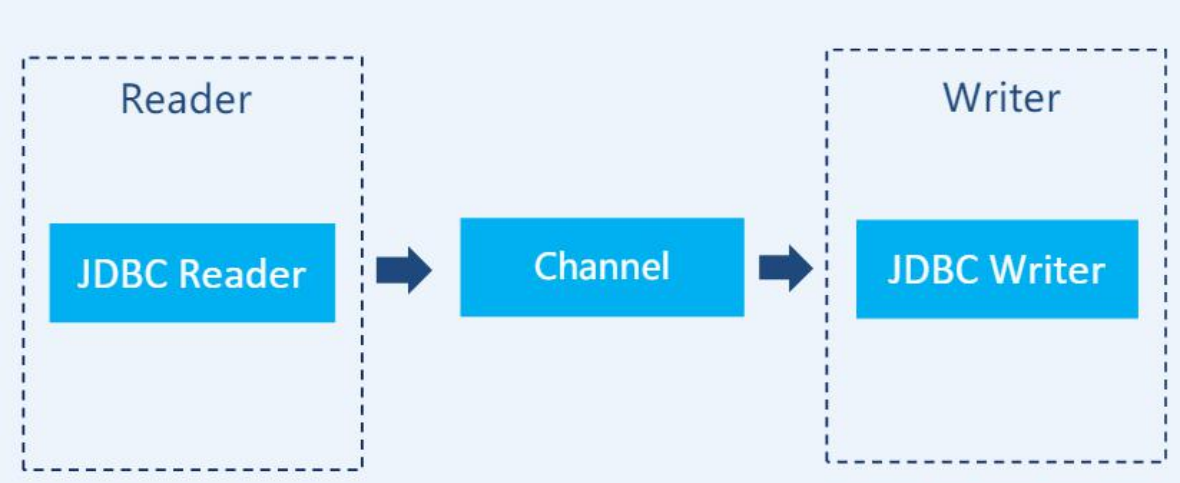
✱注意：若当前数据源正在执行任务或被占用，不可被删除。

3.2.6 批量删除

◆ 用户可多选或全选，进而对任务进行批量删除。

3.3 数据同步

3.3.1 离线同步简介



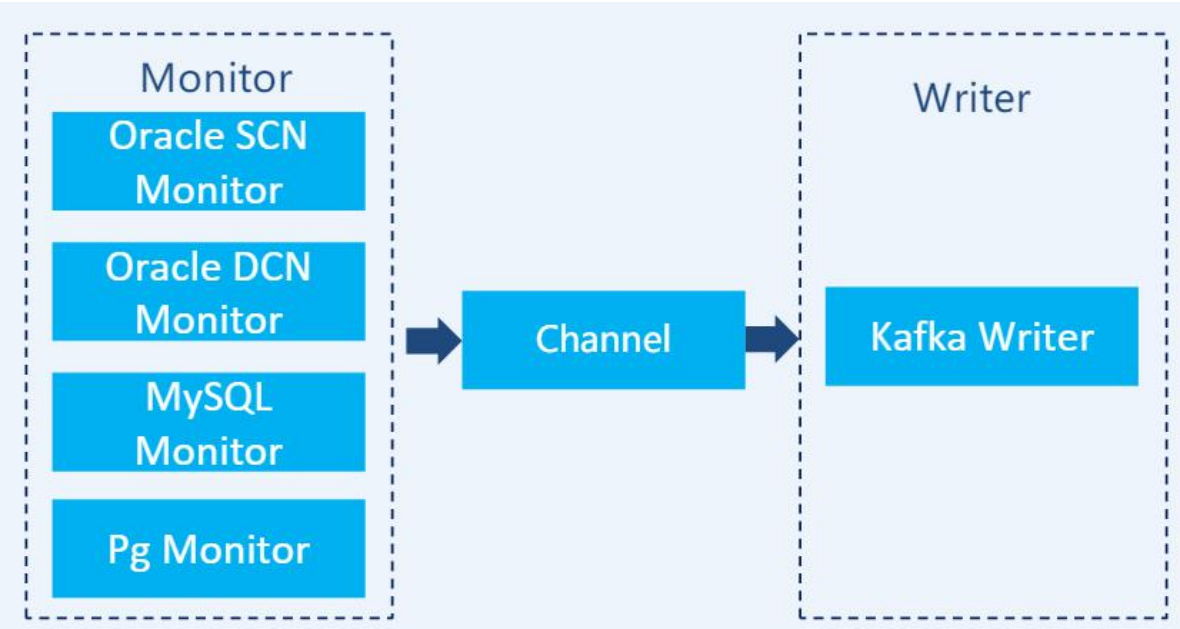
3.3.2 离线同步支持的数据源

3.3.2.1 离线任务支持配置的数据源：MySQL、Oracle、SQL Server、PostgreSQL、mongoDB、hana、API、DB2、Hana

3.3.2.2 离线任务支持配置的数据目标源：Oracle、MySQL、SQL Server、Hive、Aliyun AnalyticDB、HDFS、HBase、Hana、mongoDB、

3.4 数据实时同步

3.4.1 实时同步简介



3.4.2 实时同步支持的数据源

3.4.2.1 实时任务支持配置的数据源：Oracle、Mysql、PostgreSQL、Hana、MongoDB、Redis、kafka、Sybase、DB2

3.4.2.2 实时任务支持配置的数据目标源：Oracle、Mysql、PostgreSQL、MongoDB、Redis、kafka、kudu

3.5 新建任务

3.5.1 新建任务入口

3.5.1.1 入口一

点击【首页】下的【新建任务】按钮进入创建任务界面。

3.5.1.2 入口二

点击【任务管理】下的【新建任务】按钮进入创建任务界面。

3.5.2 创建任务

任务名称：✱必填项，支持中英文，需做唯一性判断

任务描述：输入对任务的描述，便于识别任务

任务类型：✱离线同步/实时同步/数据修复

任务优先级：✱必选项（Highest/High/Medium/Low/Lowest）

3.5.3 任务配置

3.5.3.1 第一步：选择读取组件

◆ 抽取组件配置（JDBC）

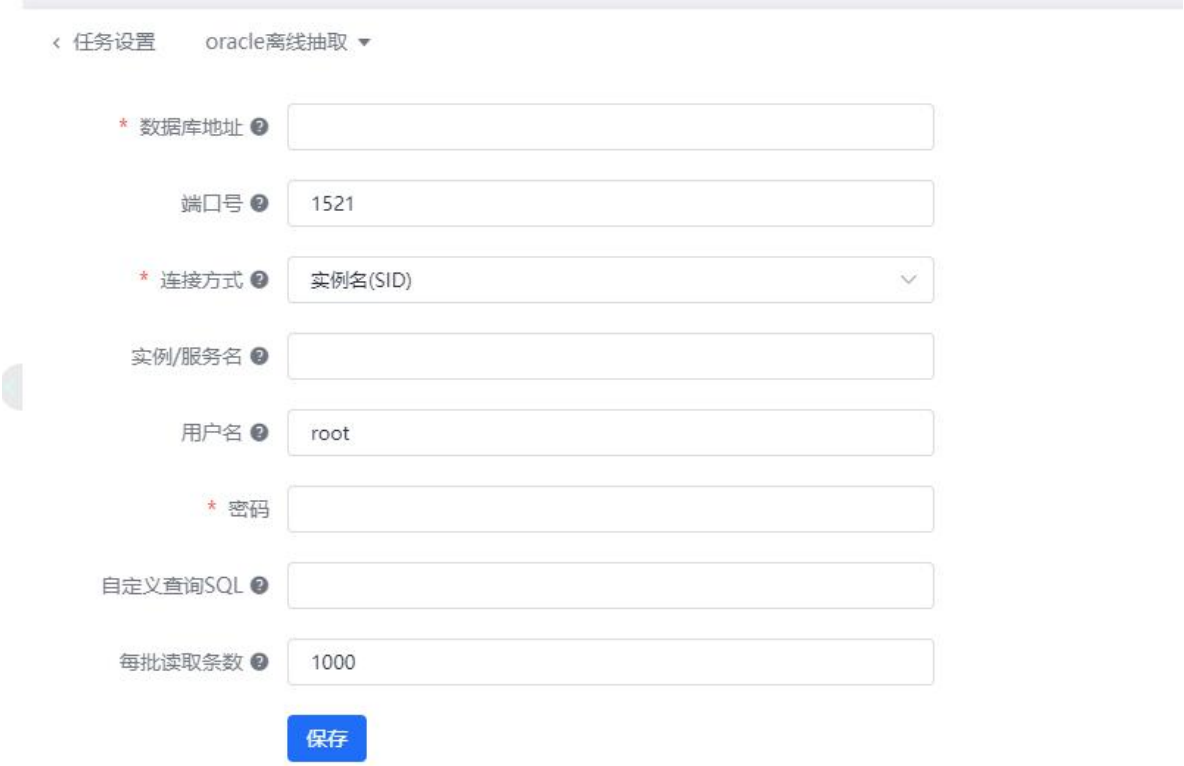
- 数据源名称：是该数据源在 DCT 的唯一标识。
- 数据源描述：是该数据源在 DCT 的标签，可快速检索
- 服务器地址：输入连接数据源的域名或 IP 地址+端口。IP 地址如 192.168.2.11:33060；域名如 ：

data.cye55uthbqll.cn-north-1.amazonaws.com.cn。

- 用户名和密码：输入可访问该数据源的用户名和密码。
- 数据源等级：公开/私密/机密
- 自定义查询 SQL：支持用户自定义 SQL 决定取数内容和范围**【离线抽取组件】**
  - ✓ 查询 SQL 格式：[库名.表名:自定义 SQL 语句]
  - ✓ 样例：

testDbName.testTableName:select * from testDbName.testTableName where create_date>${date}

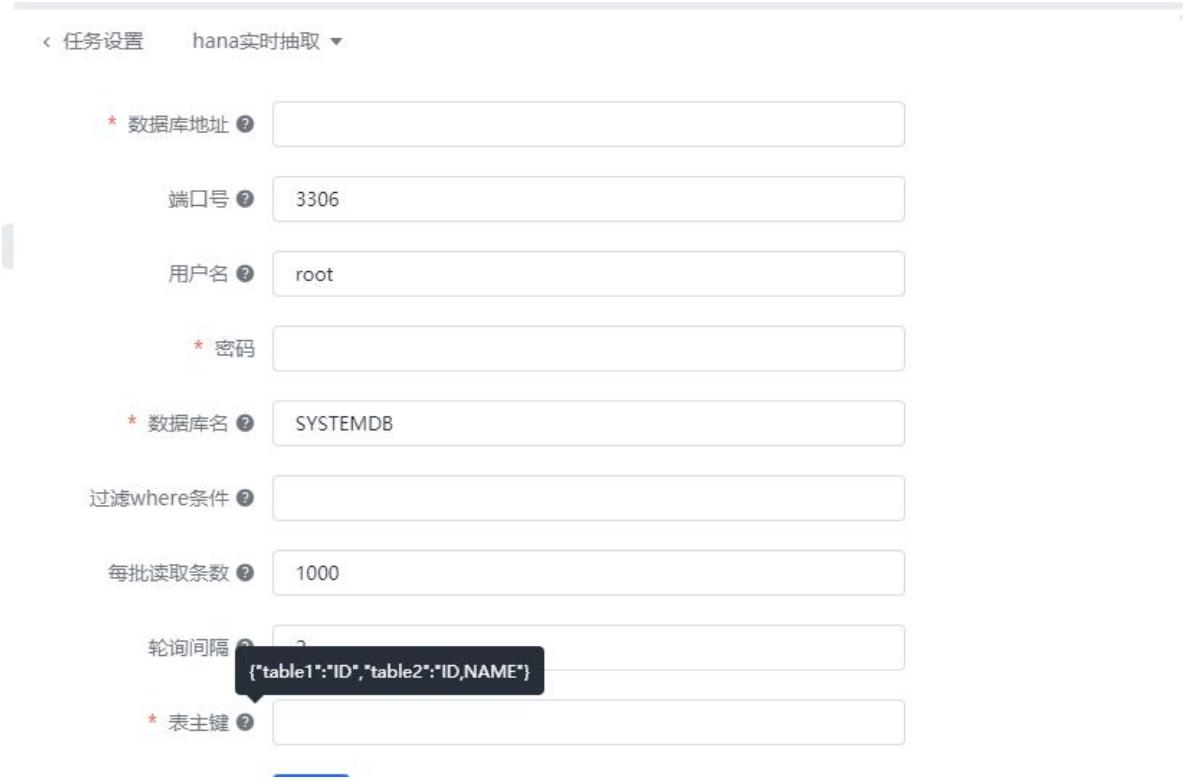
- 每批读取条数：设定批处理的条数上限



◆ 抽取组件配置（触发器）

- 数据源名称：是该数据源在 DCT 的唯一标识。
- 数据源描述：是该数据源在 DCT 的标签，可快速检索
- 服务器地址：输入连接数据源的域名或 IP 地址+端口。IP 地址如 192.168.2.11:33060；域名如 data.cye55uthbqll.cn-north-1.amazonaws.com.cn。
- 用户名和密码：输入可访问该数据源的用户名和密码。
- 数据源等级：公开/私密/机密
- 每批读取条数：设定批处理的条数上限
- 触发表配置【表主键】：输入触发对象以及触发事项，按照表名：字段名，表名：字段名的规则输入需要监听的表和字段，例如：

{“table”:”ID”,“table2”:”ID,NAME”}



◆ 抽取组件配置（kafka）

- 数据源名称：是该数据源在 DCT 的唯一标识
- 数据源描述：是该数据源在 DCT 的标签，可快速检索
- 服务器地址：输入连接数据源的域名或 IP 地址及端口值，允许输入多个地址，使用","分隔。

注意：填写前需要检查 kafka 集群是否有别名，若有别名则需要联系运维人员，添加 kafka 别名的映射关系，并填写别名。

- 数据格式： JSON（默认）
- 数据源等级：公开/私密/机密

◆ 抽取组件配置（API 离线抽取）

- API 离线组件依据用户配置 URL 和请求参数，通过 OkHttpClient 客户端发起 http//https 请求，拉取 API 数据并把拉取的数据封装为 record 数据对象，发送 record 到转换组件继续数据解析。
- API 名称：是该 API 在 DCT 的唯一标识。
- 数据源描述：是该数据源在 DCT 的标签，可快速检索
- API 接口 URL：

  ```
  URL 格式: http://IP:Port or https://xxx.
  ```

- API 离线组件请求模式：get、post
- API 请求超时时间 ：

  ```
  API 请求超时时间默认 3 秒(单位秒)
  ```

- API 接口协议头：

  ```
  数据说明: 参数名:参数值;
  数据格式: key1:value1|key2:value2;
  例子: type:json|app:name
  ```

- API 接口路径参数

  ```
  数据说明: 参数名:参数值;
  数据格式: key1:value1|key2:value2;
  例子: type:json|app:name
  ```

- API 接口请求体

  ```
  数据格式: {key1:value1,key2:value2};
  例子: {"type":"json","app":1}
  ```

- API 最大请求次数

  ```
  当请求次数大于 1 时,请一定配置下面的参数,否则当前任务仅执行一次
  ```

所述参数为:
API 请求开始页索引
API 请求每页大小
API 查询返回数据量字段名
API 查询命中数据总量字段名

- API 请求开始页索引

  数据说明: 参数名:参数值;
  数据格式: key:value;
  例子: page:1

- API 请求每页大小

  数据说明: 参数名:参数值;
  数据格式: key:value;
  例子: size:100

- API 查询返回数据量字段名

  数据说明: API 查询返回数据量字段名;
  (分页多次查询时为必须配置该字段并且 API 支持通过该字段返回查询的数据条数)

- API 查询命中数据总量字段名

  数据说明: API 查询命中数据总量字段名;
  (分页多次查询时为必须配置该字段并且 API 支持通过该字段返回命中的数据总条数)

- API 签名

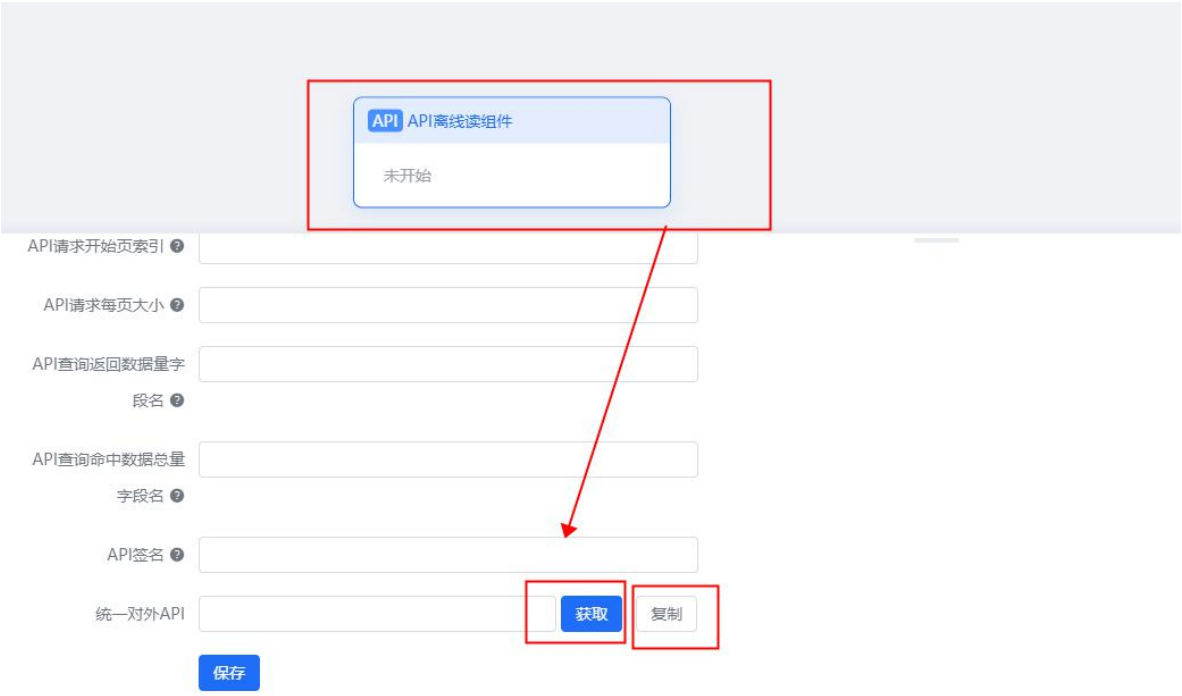
  数据说明: 签名字符串|签名密钥|加密时间戳参数字段名|签名字段名称;
  数据格式:appKey|appSecert|timestampField|signField;
  加密方式:md5(appKey+appSecert+systemTime)

- 数据源等级：公开/私密/机密

* API名称	api123
* API接口URL	http://10.201.0.25:11000/dct-manager/task
* API接口请求模式	get
API请求超时时间	3
API接口协议头	
API接口查询参数	
API接口路径参数	
API接口请求体	
* API最大请求次数	100
API请求开始页索引	page:1
API请求每页大小	size:1
API查询返回数据量字段名	size
API查询命中数据总量字段名	total
API签名	

- API 抽取任务第三方调用

点击【获取】生成第三方调用 API 抽取任务的链接，点击【复制】复制调用地址



3.5.3.2 第二步：选择转换组件

组件配置是动态表单配置，用户可在拖拉组件之后，按照组件要求配置

- 通用数据解析转换组件：自定义脚本解析/清洗 json 格式数据（优先支持用户上传 Java 代码）

  - JS 样例

```
function process(record) {
    var jsonObject = JSON.parse(record);
    var index = 0;
    for (i in jsonObject.data.meta) {
        if (jsonObject.data.meta[i].columnName === 'bpchar_col')
            index = i;
    }
    if (index === 0)
        return jsonObject;
    for (i in jsonObject.data.rows) {
        var value = jsonObject.data.rows[i].columns[index].rawData;
```

```
            jsonObject.data.rows[i].columns[index].rawData =
value.substring(0, 2) + '****' + value.substring(7);
        }
        return jsonObject;
}
```

- 依赖说明：
  - java jdk 1.8
  - java 解析 json 数据目前支持 com.alibaba.fastjson 1.2.70|com.google.code.gson 2.8.6
  - js 返回值：必须把 record 对象转 json 对象返回(脚本仅支持对 RecordData 数据转换)[目前支持 mysql、pgsql 数据源基本数据类型]

- 大小写转换组件：只支持指定表字段转换位大写或者小写。
  - 设置需要转换成大写的字段名列表，当遇到该字段名时，会将其字段值转换成大写字符串

    样例：

```
[{table:db1.table1,cols:[col1,col2]},{table:db2.table2,cols:[col1,col2]}...]
```

  - 设置需要转换成小写的字段名列表，当遇到该字段名时，会将其字段值转换成小写字符串

    样例：

```
[{table:db1.table1,cols:[col1,col2]},{table:db2.table2,cols:[col1,col2]}...]
```

- 增加字段的 T 组件 ： 解决通过初始化增加的字段的取值

  - 输入需要指定字段的字段名并输入取值范围
  - 支持的字段类型

  时间日期类： Date、Timestamp、Time

  数值类型： Integer、 Long、 Double、 Float、 BigDecimal

  字符串类型： String

  复杂类型： JSON

  - 取值策略

  从其他字段获取值，需设置 col 字段，如：新增 a 字段，值策略为从 b 字段获取

  ```
  { name:"a", type:"string", strategy:1, col:"b" }
  ```

  设置指定值，配置如下：新增字段 a,值策略为指定值，即新增值为 "333"的 a 字段。

  ```
  { name:"a", type:"string", strategy:2, val:"333" }
  ```

  - 填写规范/样例：

  ```
  {db1.table1:[{name:'col1',strategy:2,val:1,type:integer},{name:'col1',strategy:1,col:'sourceCol',type:integer},{name:'createdDate',type:'date',strategy:2,val:'2021/10/10',format:'yyyy/MM/dd'}],...}
  ```

3.5.3.3 第三步：选择写入组件，并配置

- 通用写入组件配置

  - 数据源名称：是该数据源在 DCT 的唯一标识。

- 数据源描述：是该数据源在 DCT 的标签，可快速检索
- 服务器地址：输入连接数据源的域名或 IP 地址+端口。

IP 地址：如 192.168.2.11：33060；

域名： 如 data.cye55uthbqll.cn-north-1.amazonaws.com.cn。

- 用户名和密码：输入可访问该数据源的用户名和密码。
- 数据源等级：公开/私密/机密

3.5.3.4 第四步：配置任务信息

◆任务基本信息：默认配置任务名称/任务类型/任务描述/任务优先级

◆规则配置：可配置任务的映射规则和预警通知规则

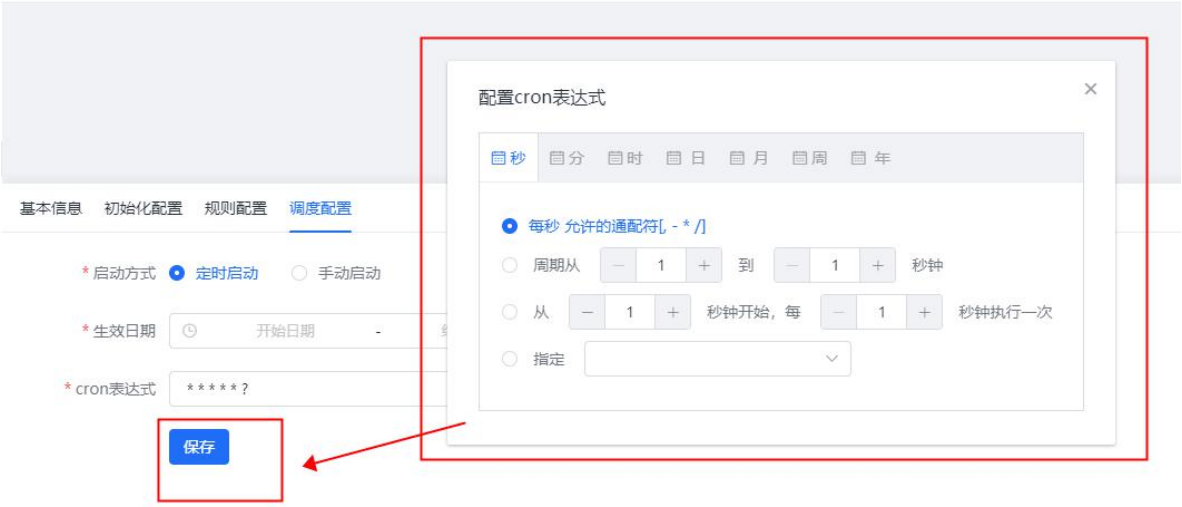
◆调度配置：配置任务启动方式和运行频率**【离线任务】**

- 支持手动启动/定时启动两种方式
- 手动启动：支持用户创建任务后，手动点击启动任务去执行任务

- 定时启动：支持配置 cron 表达式，按照设定频率执行任务
  - ✓ 启动方式：定时启动

✓ 生效日期：设定开始日期-结束日期，任务在有效期范围内按照设定频率执行任务

✓ Cron 表达式：配置表达式



◆初始化配置：初始化解决结构同步/增加字段/批量修改表字段名等（本期支持自定义 SQL 建表）

- 样例

```
-- ----------------------------
-- Table structure for dct_alert_rule
-- ----------------------------
DROP TABLE IF EXISTS `dct_alert_rule`;
CREATE TABLE `dct_alert_rule` (
  `id` varchar(32) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL COMMENT 'uuid',
  `task_id` varchar(32) CHARACTER SET utf8 COLLATE utf8_bin NULL DEFAULT NULL COMMENT '任务 Id',
```

```
  `name` varchar(255) CHARACTER SET utf8 COLLATE utf8_general_ci NUL
L DEFAULT NULL COMMENT '阈值名称 具有唯一性',
  `type` tinyint(4) NULL DEFAULT NULL COMMENT '监控指标。10 同步错误
记数、20 任务错误统计、30 运行时间统计',
  `rule_condition` varchar(255) CHARACTER SET utf8 COLLATE utf8_gener
al_ci NULL DEFAULT NULL COMMENT '触发条件，由判断条件（>=、<=、>、
<）和阈值组成。例如，阈值条件设置为>=80',
  `cycle` tinyint(4) NULL DEFAULT NULL COMMENT '连续周期，连续多少个
周期满足阈值条件后，发送阈值告警。可设置为 1、2、3、4 或 5。',
  `level` tinyint(4) NULL DEFAULT NULL COMMENT '告警级别。10 紧急、20
重要、30 次要或 40 提示',
  `is_notice` tinyint(4) NULL DEFAULT NULL COMMENT '是否发送通知
(0:false,1:true)',
  `notice_event` tinyint(4) NULL DEFAULT NULL COMMENT '通知事件(10 出
现告警,20 恢复正常,30 数据不足)',
  `aggregation_type` tinyint(4) NULL DEFAULT NULL COMMENT '聚合方
式，平均值、最小值、最大值、总计或样本个数',
  `aggregation_time` varchar(255) CHARACTER SET utf8 COLLATE utf8_ge
neral_ci NULL DEFAULT NULL COMMENT '聚合周期，可设置为 1 分钟、5 分
钟、15 分钟或 1 小时',
  `is_open` tinyint(4) NULL DEFAULT NULL COMMENT '是否开启
(0:false,1:true)',
  `tenant_id` varchar(32) CHARACTER SET utf8 COLLATE utf8_bin NULL DE
FAULT NULL COMMENT '租户 Id',
  `created_by` varchar(255) CHARACTER SET utf8 COLLATE utf8_general_ci
 NULL DEFAULT NULL COMMENT '创建人',
  `updated_by` varchar(255) CHARACTER SET utf8 COLLATE utf8_general_
ci NULL DEFAULT NULL COMMENT '更新人',
  `created_at` bigint(20) NULL DEFAULT NULL COMMENT '创建时间',
  `updated_at` bigint(20) NULL DEFAULT NULL COMMENT '更新时间',
  `dr` tinyint(4) NOT NULL DEFAULT 0,
  PRIMARY KEY (`id`) USING BTREE
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_bin ROW_FO
RMAT = DYNAMIC;
```

3.5.3.5 组件其他配置持续更新中

3.5.3.6 任务画布

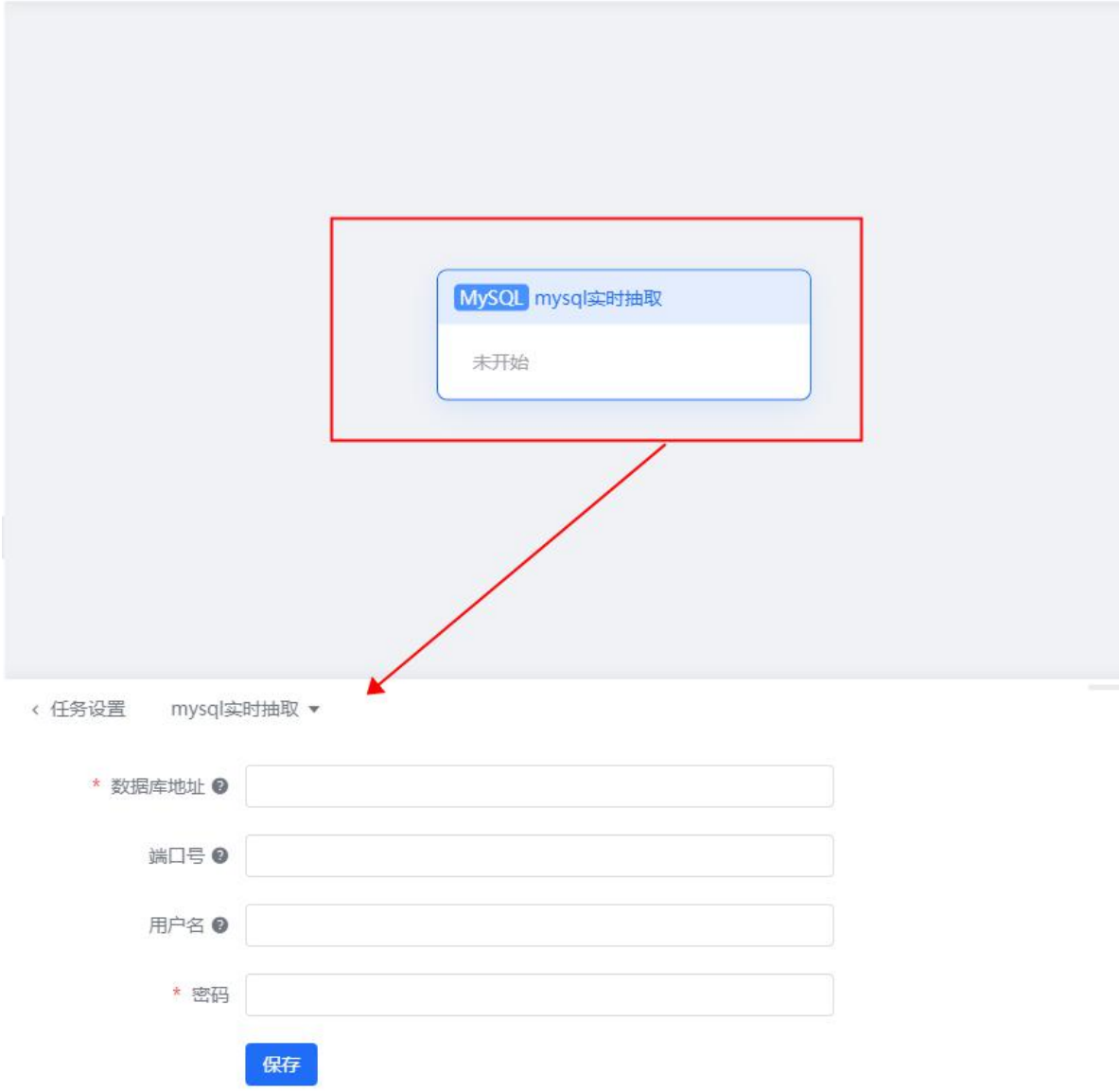
组件列表：包含读取组件、转换组件、写入组件；从此处可拖拉组件到画布中央，进行组件配置；同时可根据关键字搜索组件；刷新组件列表，可获取最新的组件；

状态栏：刷新/预检查/停止/开始

组件 DAG：

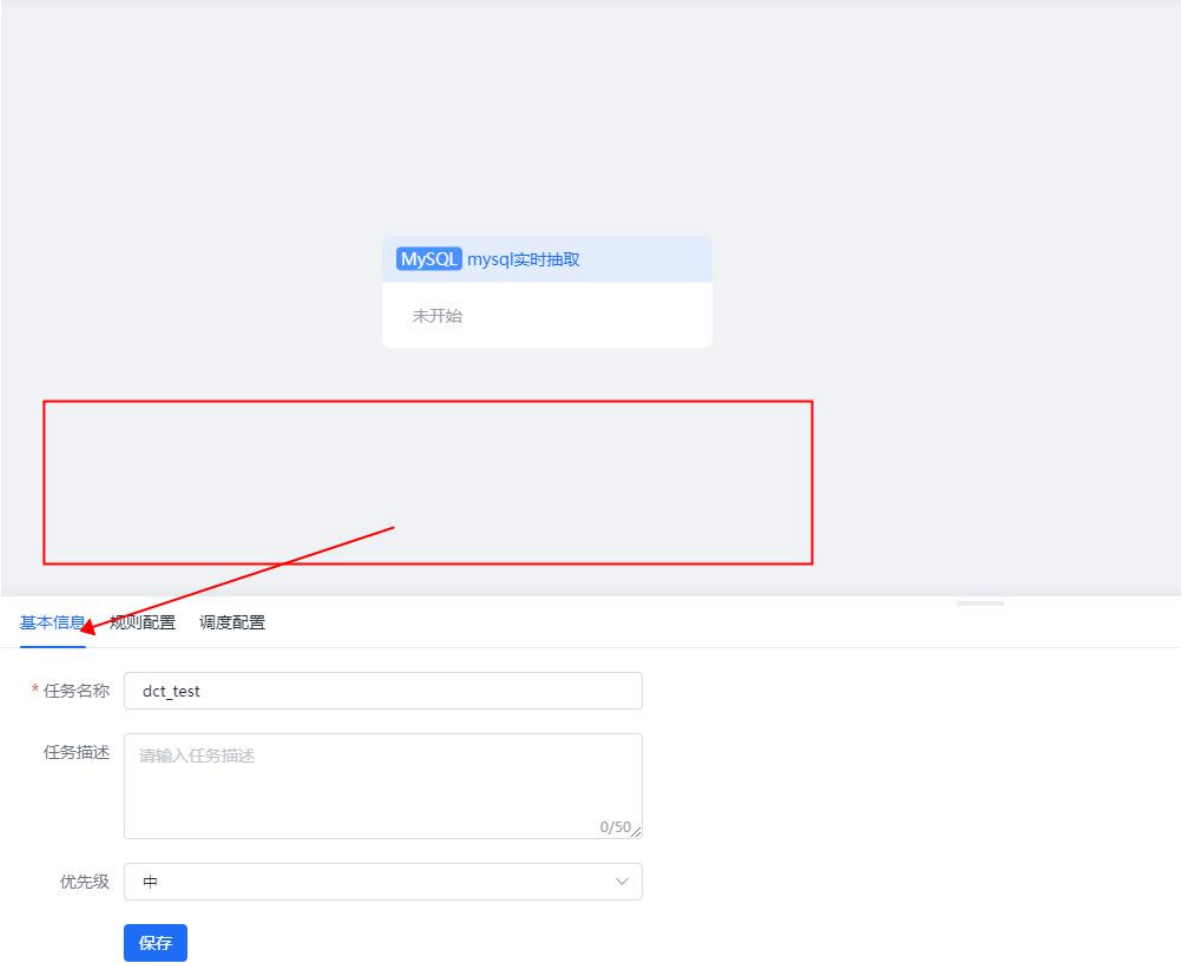
- 由读取组件/转换组件（非必选，按需）/写入组件组成最小的可执行程序；

- 点击组件，则可对选中组件进行配置



任务配置：

- 点击画布空白处，可配置任务



- 基本信息配置：可对任务名称、任务描述、任务等级进行编辑

- 规则配置：可自定义预警规则，也可配置映射规则

◆ 预警规则配置

◆ 预警规则管理

◆ 预警规则操作：启用/编辑/删除

◆ Webhook 配置（钉钉/企业微信等通知）

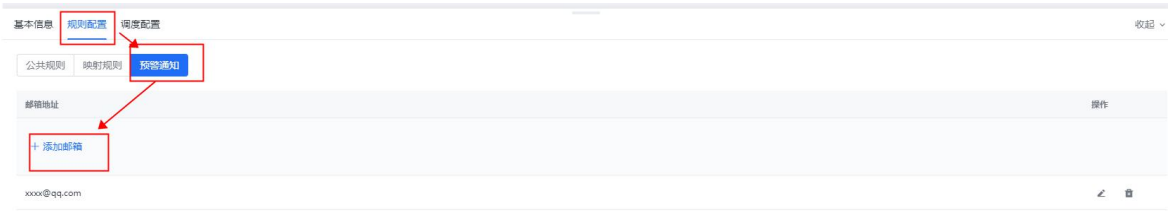
■ Webhook 允许单任务配置多中预警方式；点击【规则配置】-【新增webhook】-输入名词/URL/密码/密文--【保存】，完成 webhook 的配置；

■ Webhook 通知开启：点击【规则配置】--【公共规则】--勾选【webhook 通知】

■ Webhook 通知关闭：点击【规则配置】--【公共规则】--编辑规则--取消

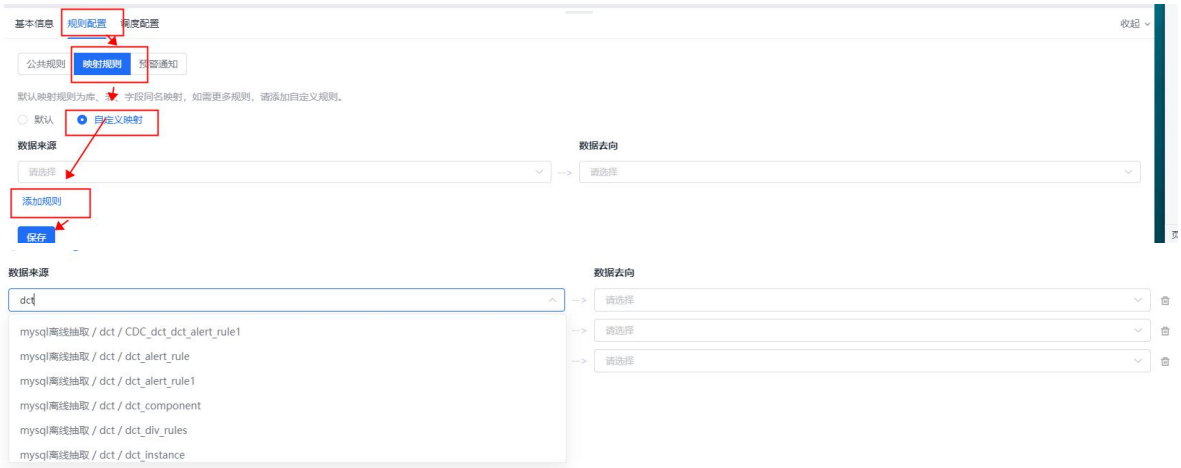
勾选【webhook 通知】--【保存】即可关闭 webhook 通知

◆ 预警通知（邮件预警通知）



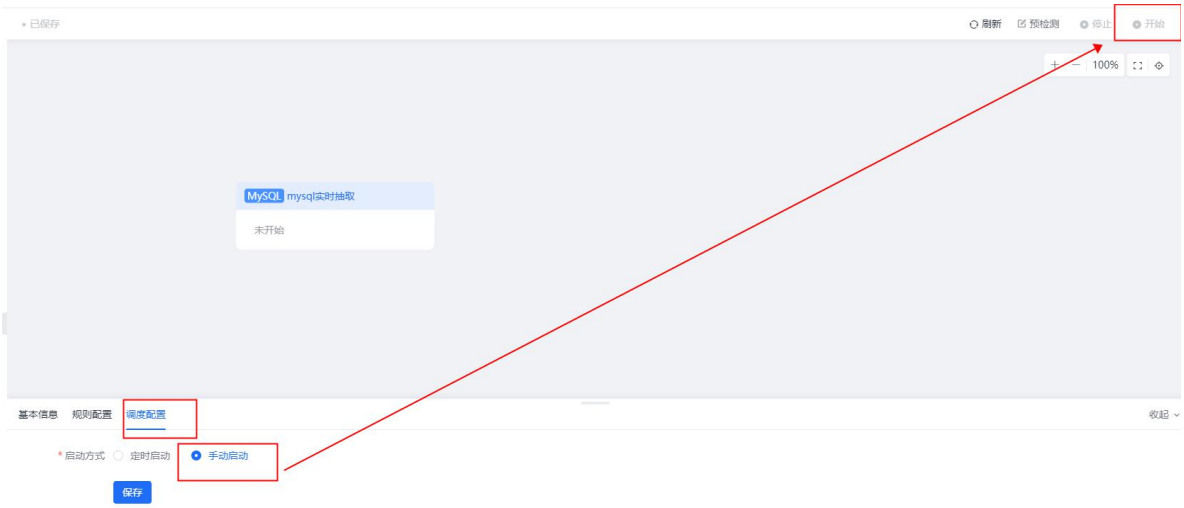
◆ 映射规则配置：支持默认同名映射/支持自定义库表级别映射

■ 自定义库表级映射：支持模糊搜索，可在规则配置框输入苦命/表名进行模糊搜索；多条规则需分开配置；若需要设置多对一/一对多，则需要设置多条规则，本期单行只允许单选；

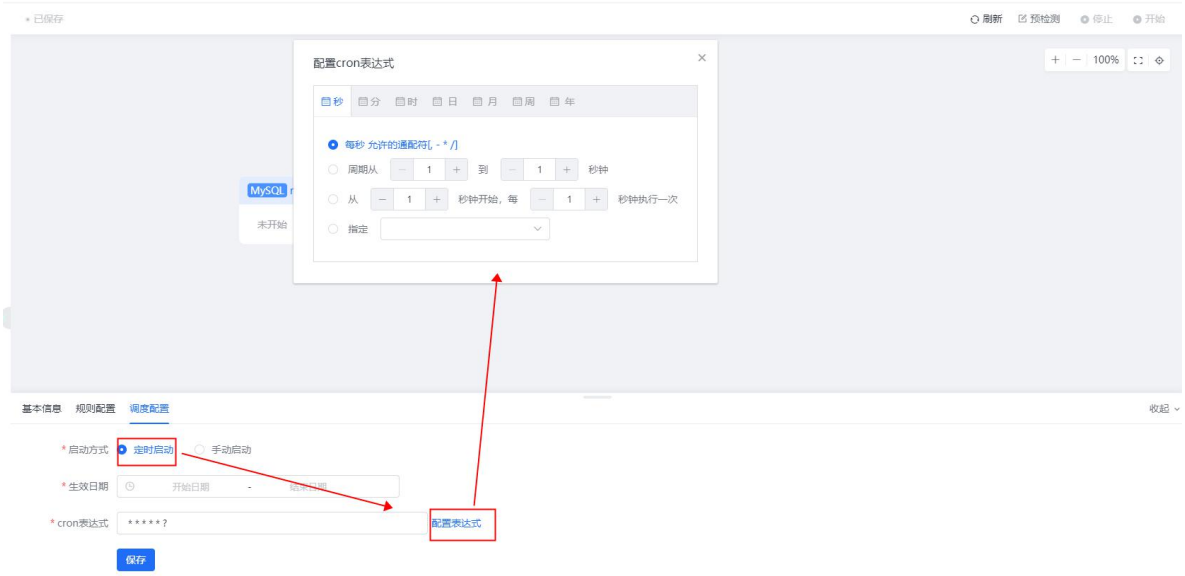


- 调度配置：支持定时启动核手动启动两种方式【实时任务无】

  - 手动启动：需要完成任务之后，点击画布右上角【开始】按钮开启任务



  - 定时启动：支持设定任务生效时间；支持配置 cron 表达式

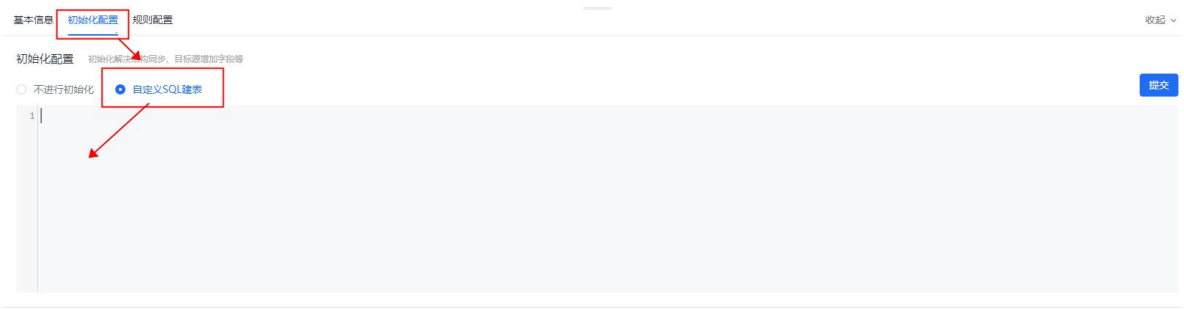


- 初始化配置

  - 不进行初始化：若无结构同步需求，则选择不进行初始化

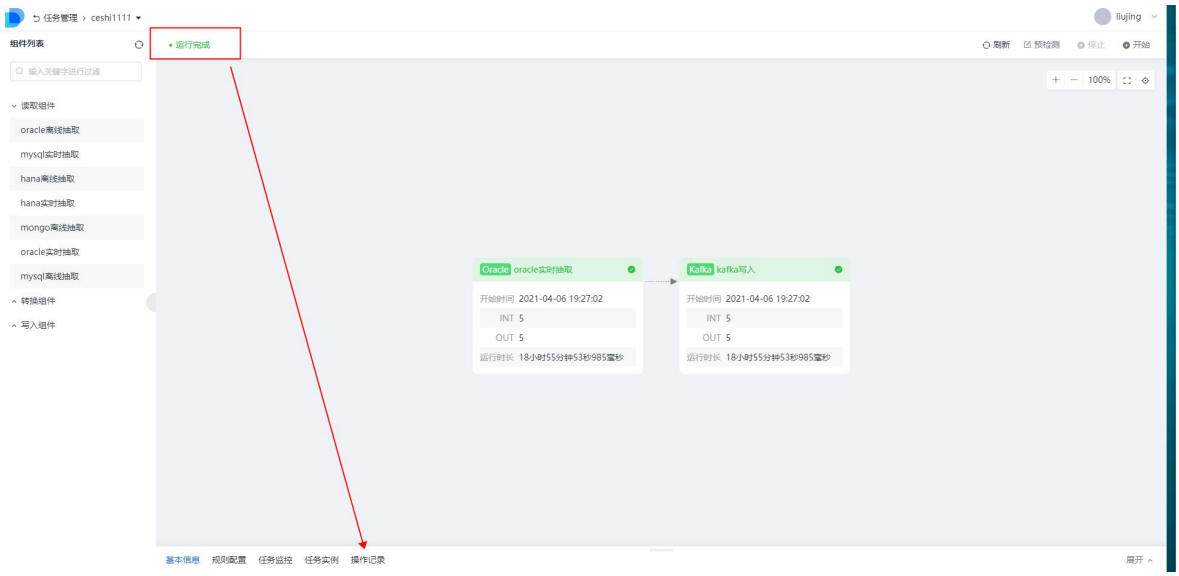
◆ 自定义 SQL 建表

◆ 批量修改库表字段名（增加前缀/后缀/替换等）



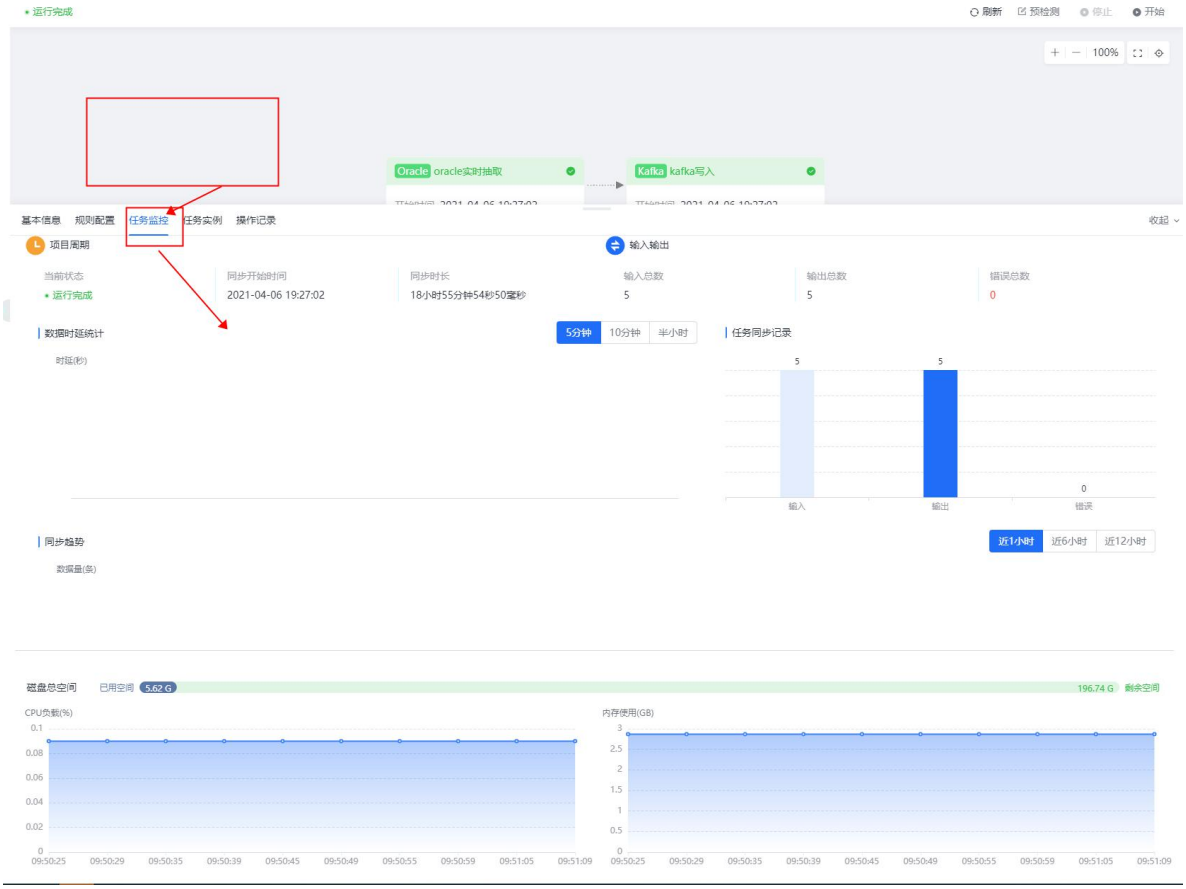
3.5.3.7 任务运维

◆ 已经完成任务配置的任务，运行后可查看任务的运行情况，任务实例以及操作记录

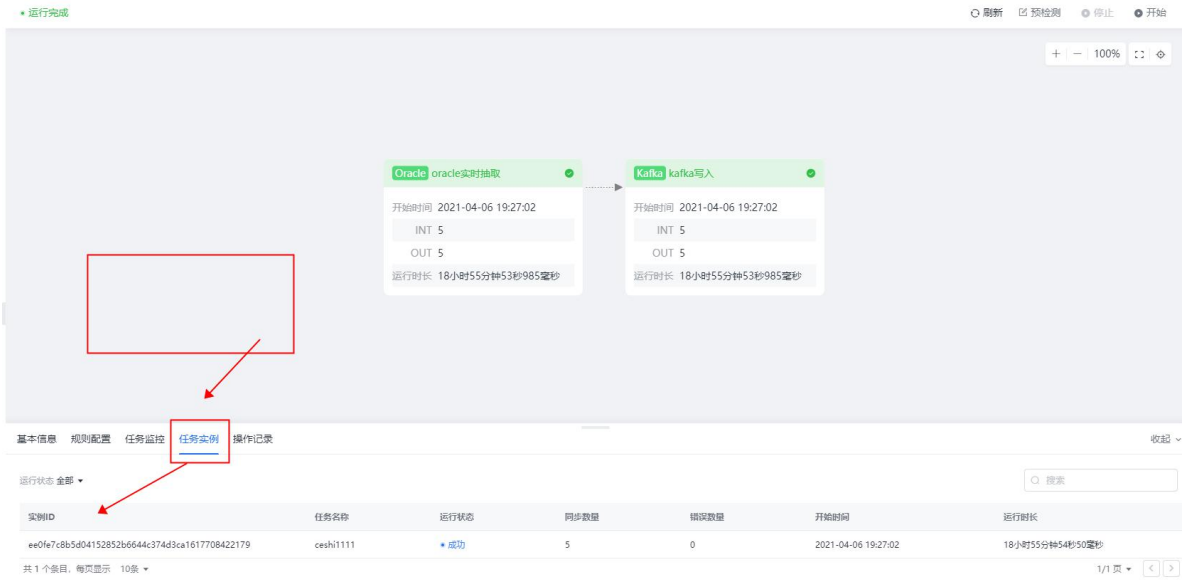


◆ 任务切换：点击左上角导航，可查看当前平台所有的任务，选中任务，即可查看该任务的详细情况

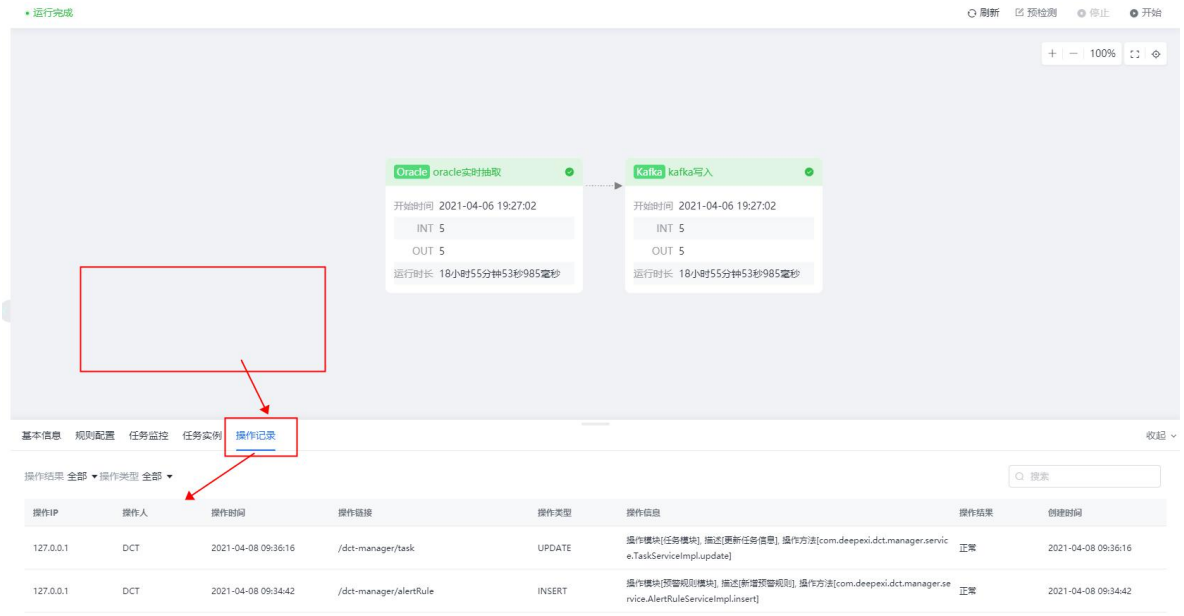
◆ 任务监控：点击画布空白处，可弹出任务监控



◆ 任务实例：点击画布空白处，可弹出任务实例管理列表

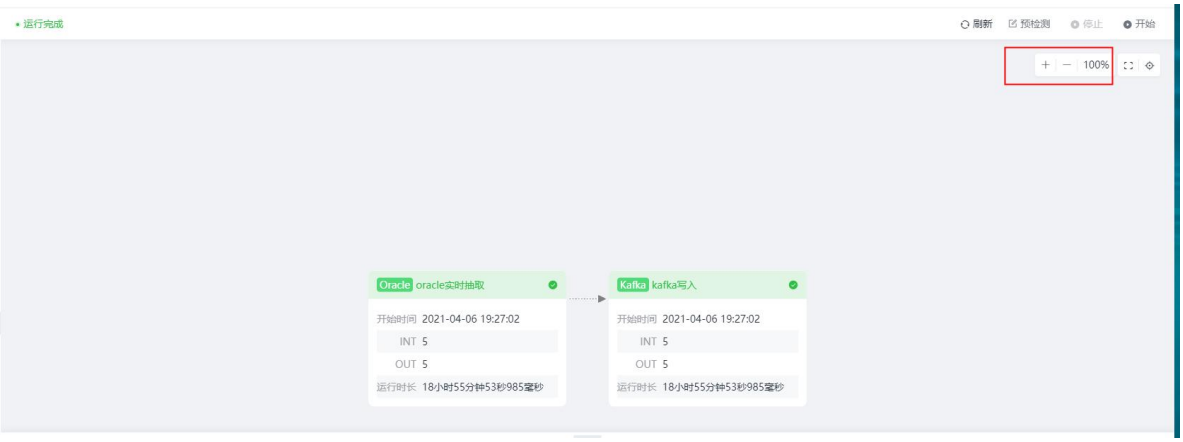


◆ 操作记录：点击画布空白处，可弹出任务操作记录

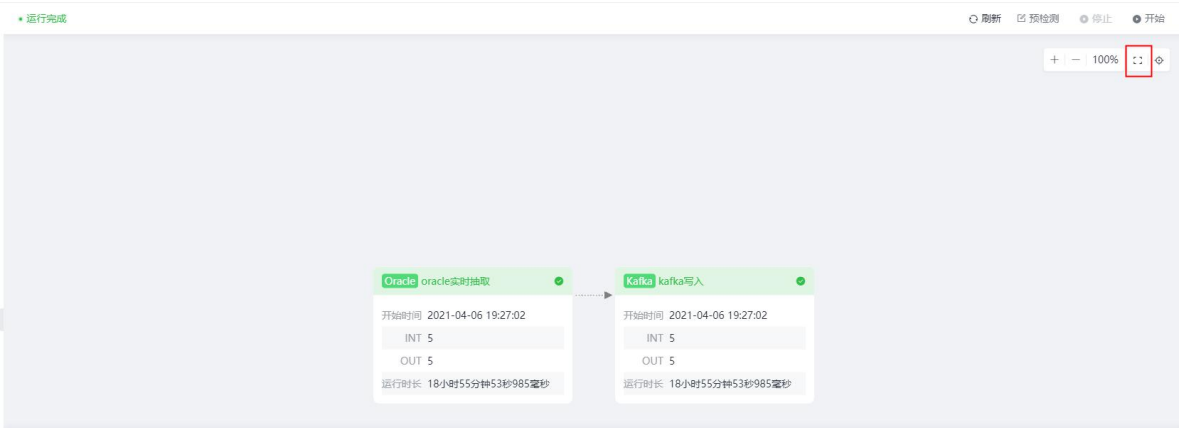


3.5.3.8 任务画布操作

◆ 组件放大/缩小：点击对组件进行百分比的缩放

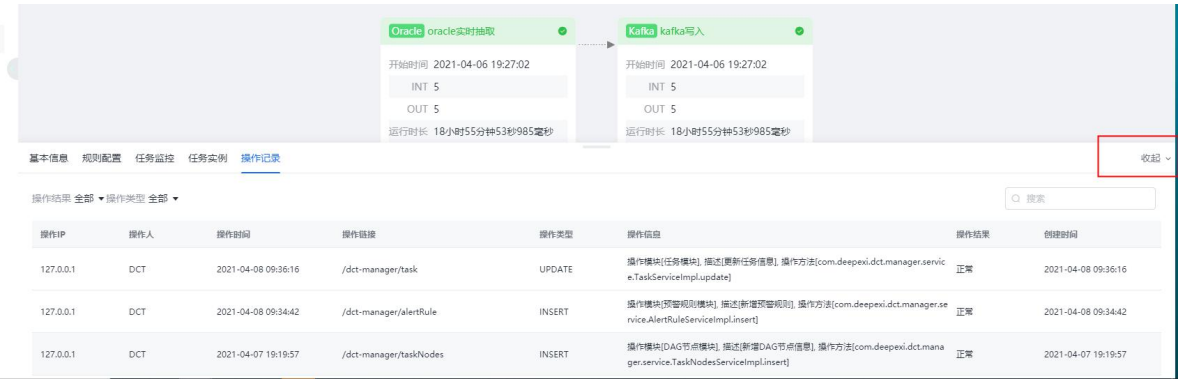


◆ 全屏/取消全屏：选中全屏，则隐藏左侧组件列表，底部任务配置信息；取消全面则恢复默认配置



◆ 居中：点击则将组件显示到画布正中的位置

◆ 任务配置画布收起/展开：任务配置画布默认收起，点击空白处可展示；展示后可点击【收起】按钮收起画布

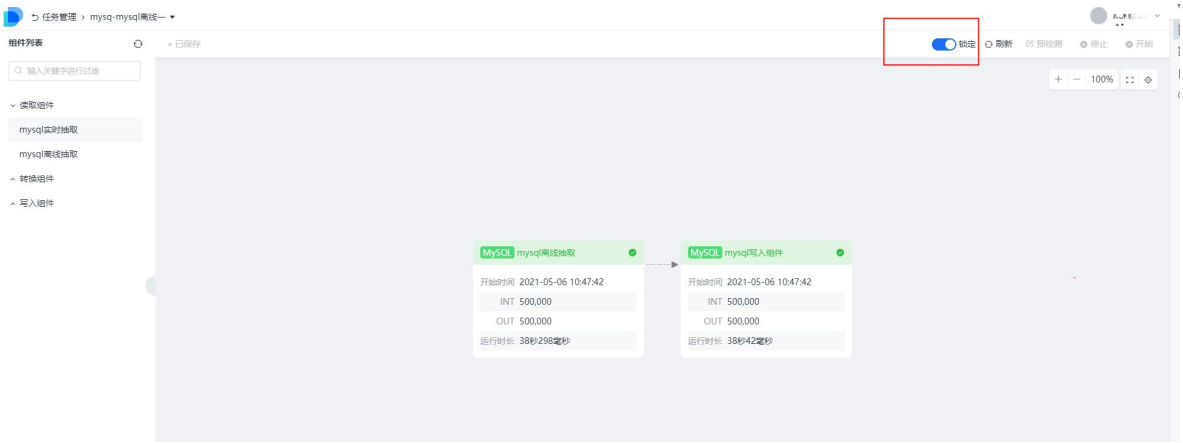


◆ 任务配置画布大小调整：任务画布展开后，可选中调节按钮，上下拖动，调整画布大小

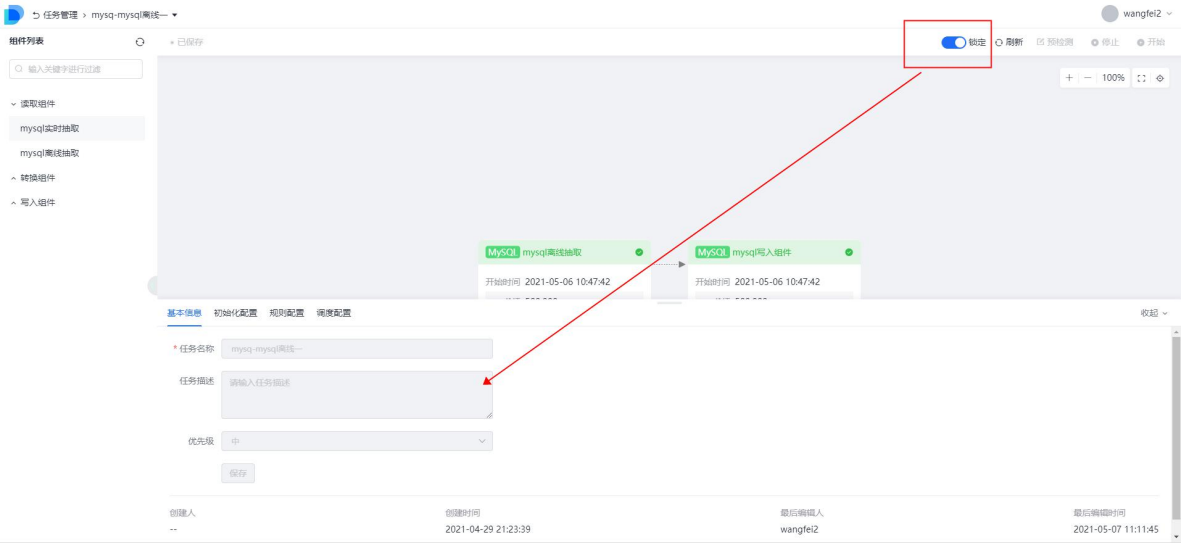
3.6 任务锁定/解锁

◆ 锁定入口一：任务管理列表页

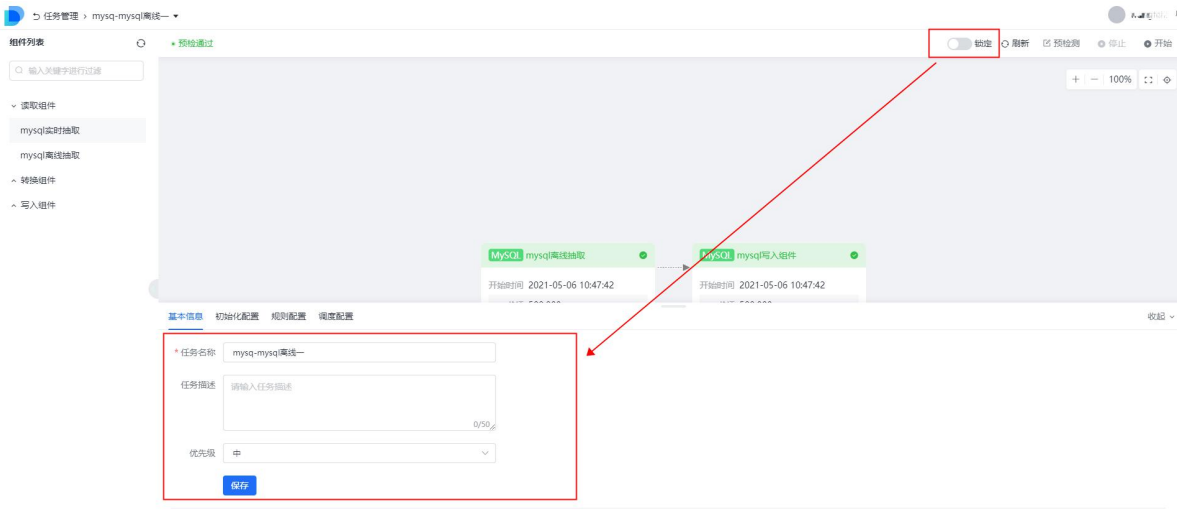
◆ 锁定入口二：任务详情页



◆ 任务锁定：同一个租户下的子账户可对任务进行锁定；锁定后，该租户下所有的子账户不可对任务进行操作



◆ 任务解锁：配有解锁按钮权限的用户可点击解锁任务，解锁任务后，可对任务进行编辑修改删除等

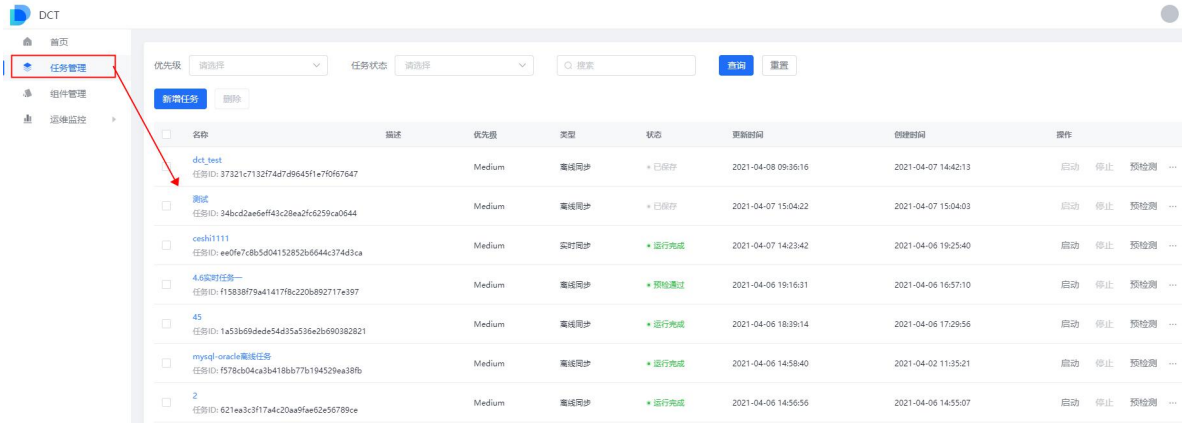


*注：任务锁定按钮的权限可在 iam 后台配置

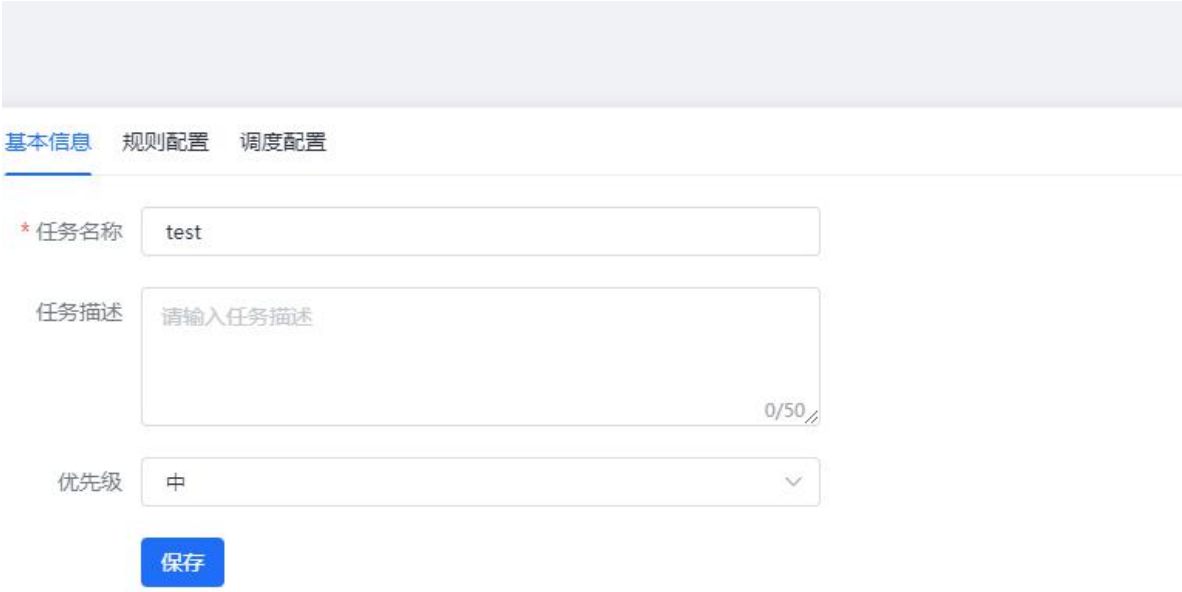
3.7 任务管理

3.7.1 任务管理列表

3.7.1.1 任务管理入口



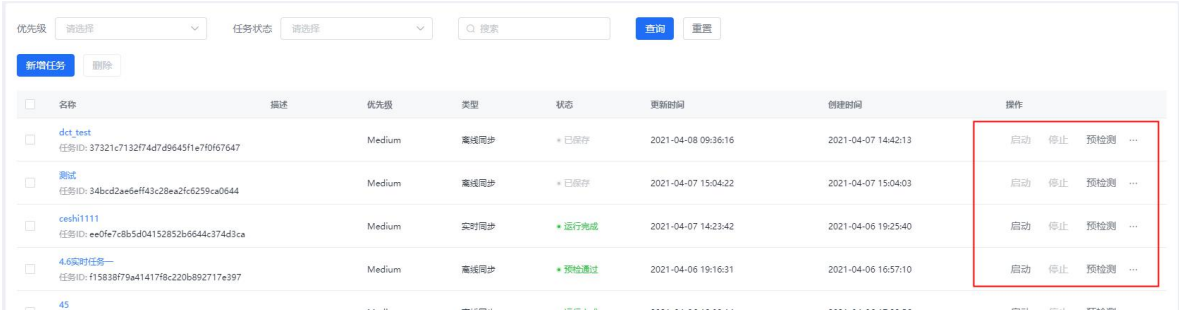
3.7.1.2 任务信息



3.7.2 任务筛选

- 可按照任务优先级/任务状态/自定义搜索条件按搜索任务

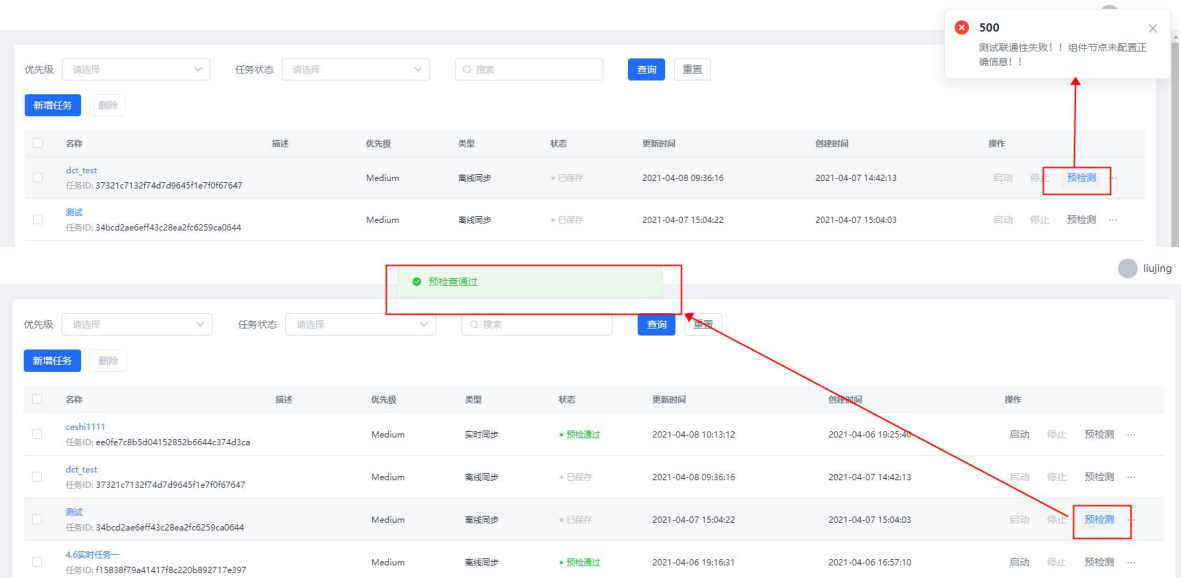
3.7.3 任务操作



3.7.3.1 预检查

- 点击任务列表【操作】中的【预检查】按钮，检查任务。

◆ 预检查结果：失败/成功；若失败，查看失败原因，修改配置后再启动；若成功，则可启动任务



3.7.3.2 启动

◆ 点击任务列表【操作】中的【启动】按钮，启动并执行任务。

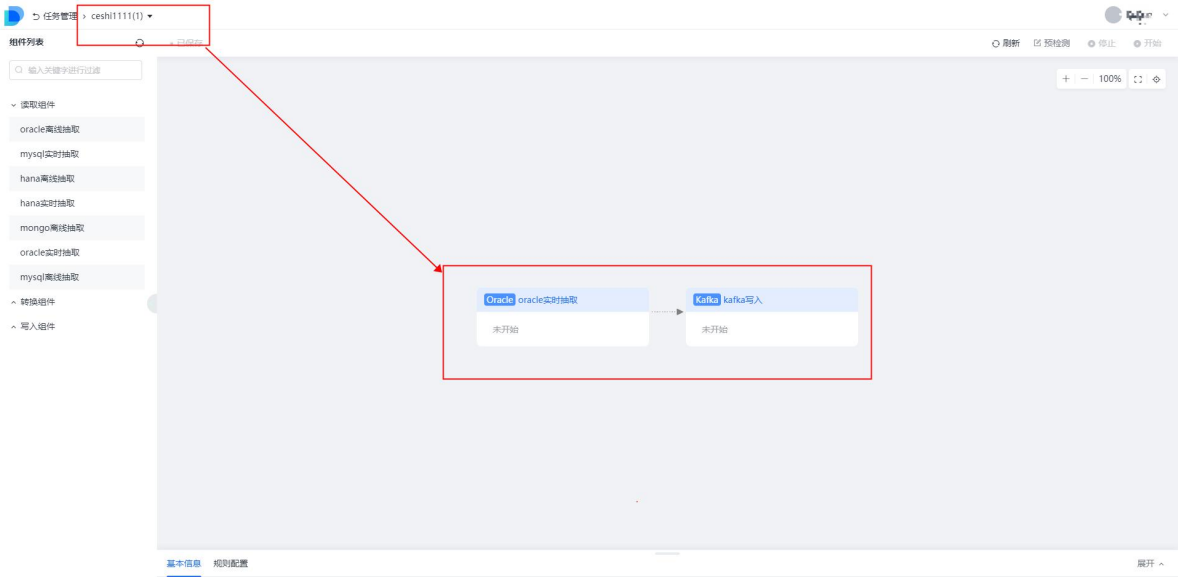
3.7.3.3 停止

◆ 点击任务列表【操作】中的【停止】按钮，停止执行任务。

3.7.3.4 创建类似任务

3.7.3.5 点击任务列表【...】更多按钮中的【创建类似任务

】按钮，创建具有项目组件的同步任务。默认创建按任务名称，可在基本信息修改



3.7.3.6 任务配置【快捷入口】

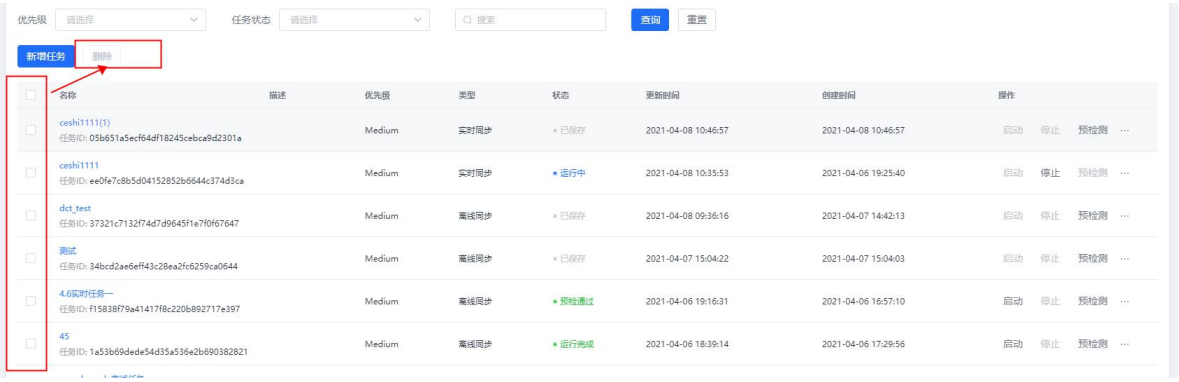
3.7.3.7 任务删除

◆ 点击任务列表【...】更多按钮，中的【删除】按钮，删除当前任务。

✱注意：若任务正在执行或被占用，不可被删除。

3.7.3.8 批量删除

◆ 全选或选中多项任务，点击【批量启动】或【批量删除】，批量启动或删除任务。



3.8 数据修复【配置和离线任务一致】

3.8.1 新建数据修复任务

3.8.1.1 创建任务

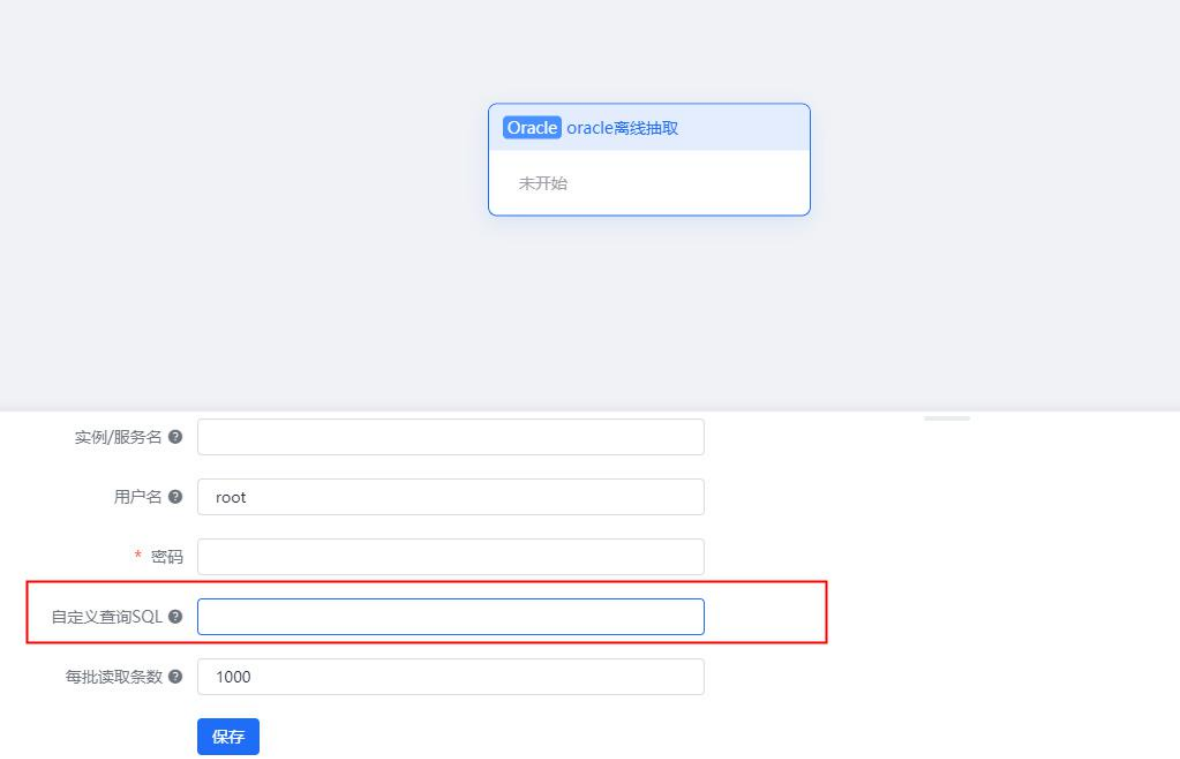
- 点击【任务管理】下的【新建任务】按钮进入新建任务页面，选择任务类型为数据修复，创建任务。

- 任务名称：✱必填项，支持中英文，需做唯一性判断
- 任务类型：✱数据修复
- 任务优先级：✱必选项（Highest/High/Medium/Low/Lowest）

3.8.1.2 任务配置【同离线任务配置】

3.8.1.3 自定义取数起点

- 点击组件配置列表中的【自定义 SQL】按钮即可对当前选中任务进行自定义配置



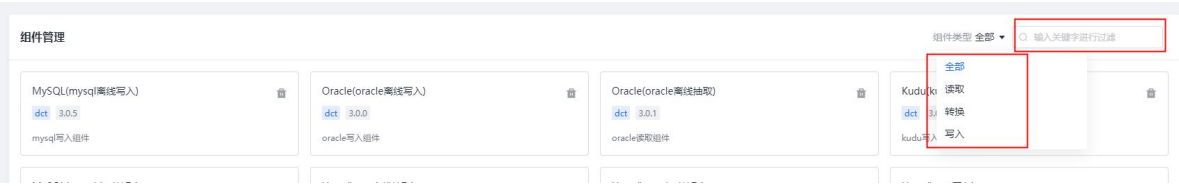
3.9 数据订阅【合并到实时任务，监听结果写入 kafka】

3.10 组件管理

3.10.1 管理入口

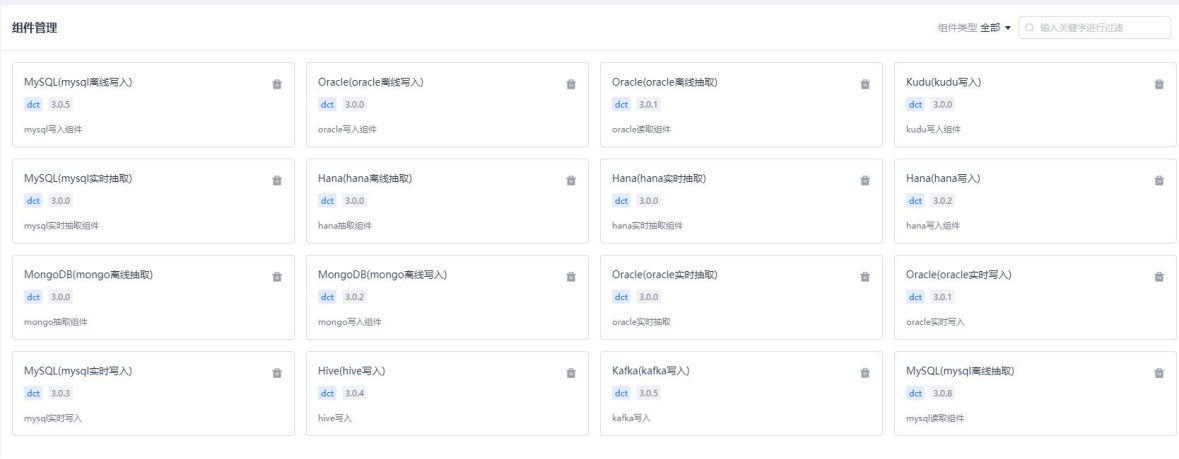
3.10.2 组件筛选

- 支持按照组件类型和自定义关键字筛选组件



3.10.3 组件列表

- 按照组件更新的时间倒序



3.10.4 新增组件

- 新增组件：新购买的组件包或者组件升级包，可通过新建的方式加载到组件库

3.10.4.1 新增组件信息填写

新增组件

* 组件名称

* 组件类型

请选择

* 任务类型

离线同步

* 数据库类型

请选择

* 组件版本

标识

描述

0/50

* 上传组件

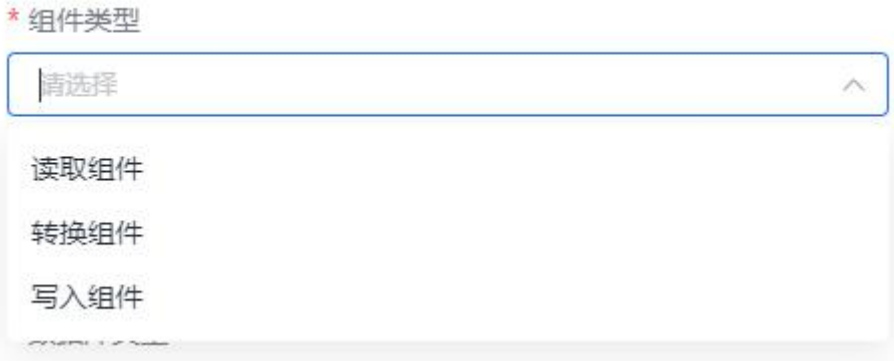
点击上传

上传组件.zip压缩包

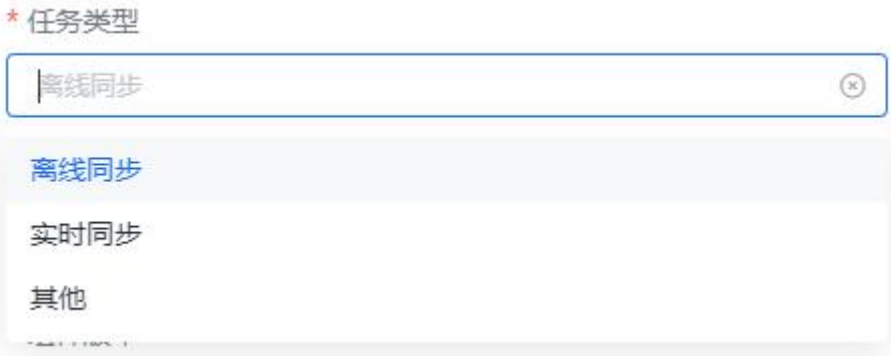
保存 取消

- 组件名称：按照命名规范输入组件名称

- 组件类型：读取组件/转换组件/写入组件

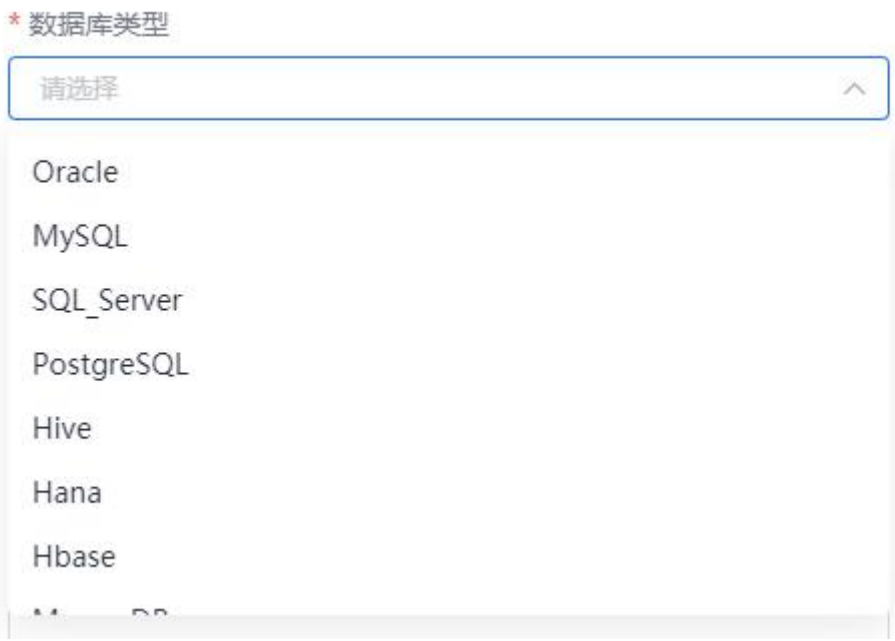


- 任务类型：选择组件支持的任务类型（离线同步/实时同步/其他）



注：若组件支持实时和离线，则可选择类型为其他

- 数据库类型：选择当前组件支持的数据库类型

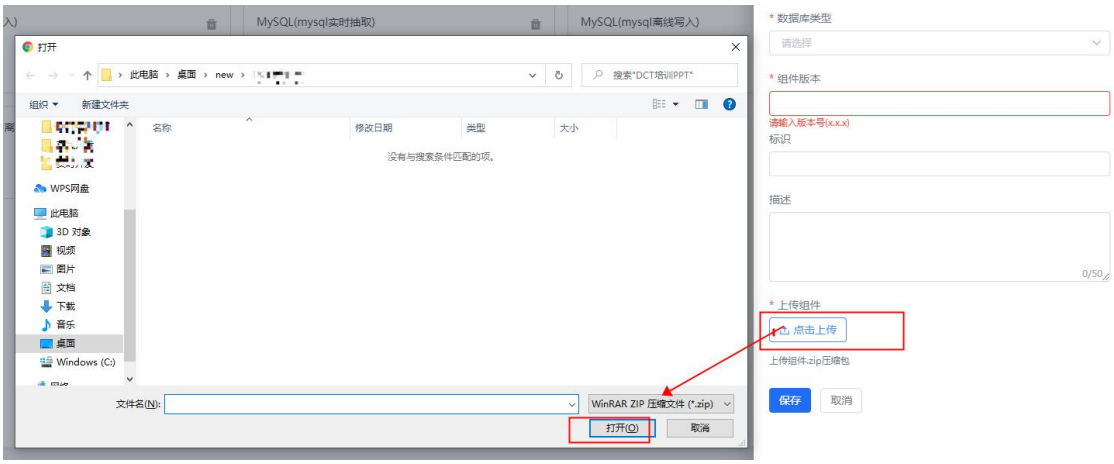


- 组件版本：按照标准输入组件版本

- 标识：用来对组件进行分类识别的标识

■ 描述：输入组件的描述，50 字以内

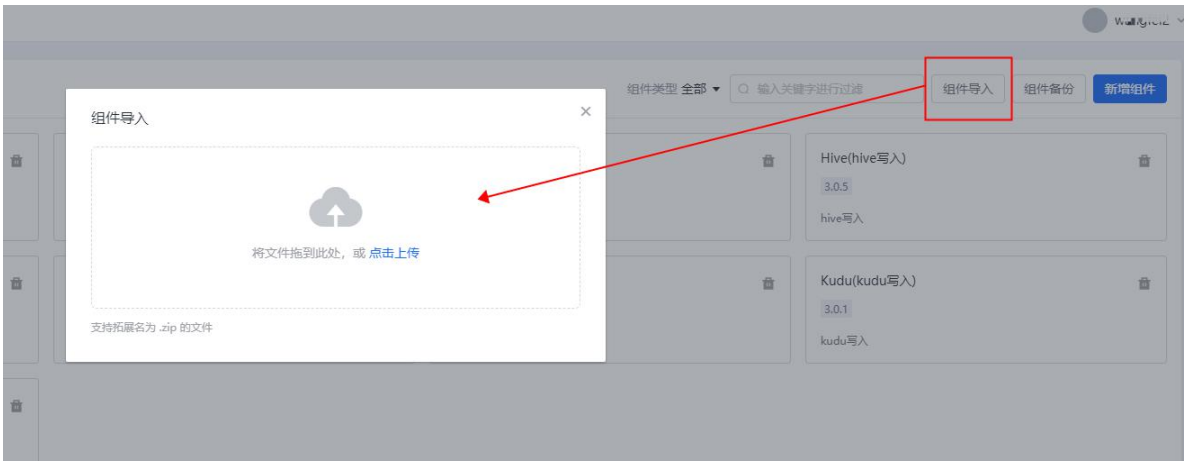
◆ 上传组件：上传本地组件 zip 压缩包



■ 保存/取消：点击【保存】按钮，若符合提交规范，则提交组件；刷新可在组件列表看到新建组件；若提交信息不完善，则保存失败，请检查提交信息后再提交；点击【取消】则取消新建组件。

3.10.5 组件导入

◆ 组件导入：支持批量导入组件



*注：支持拓展名为 .zip 的文件

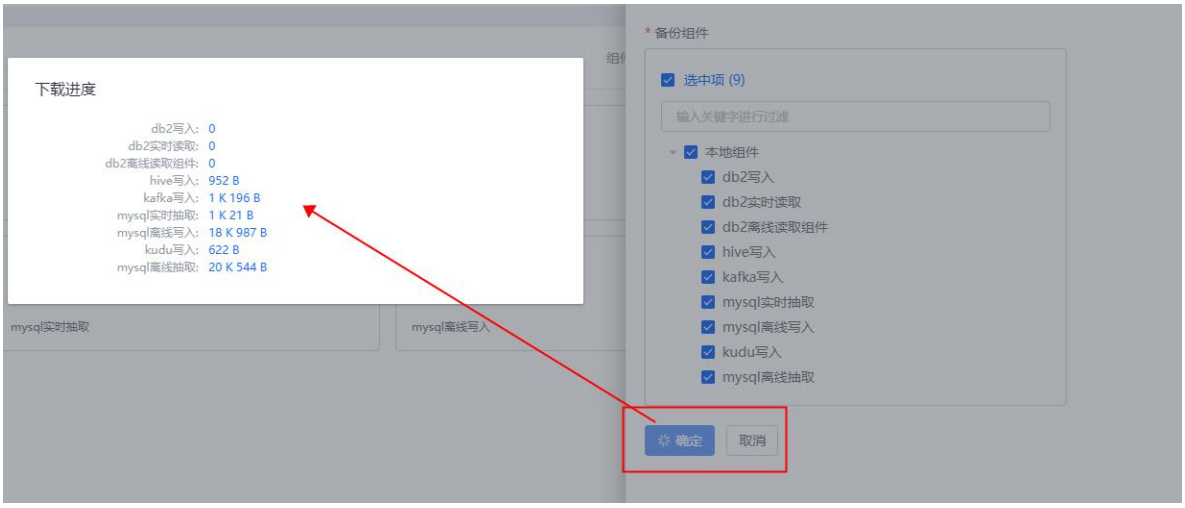
3.10.6 组件导出

组件导出：支持批量导出组件

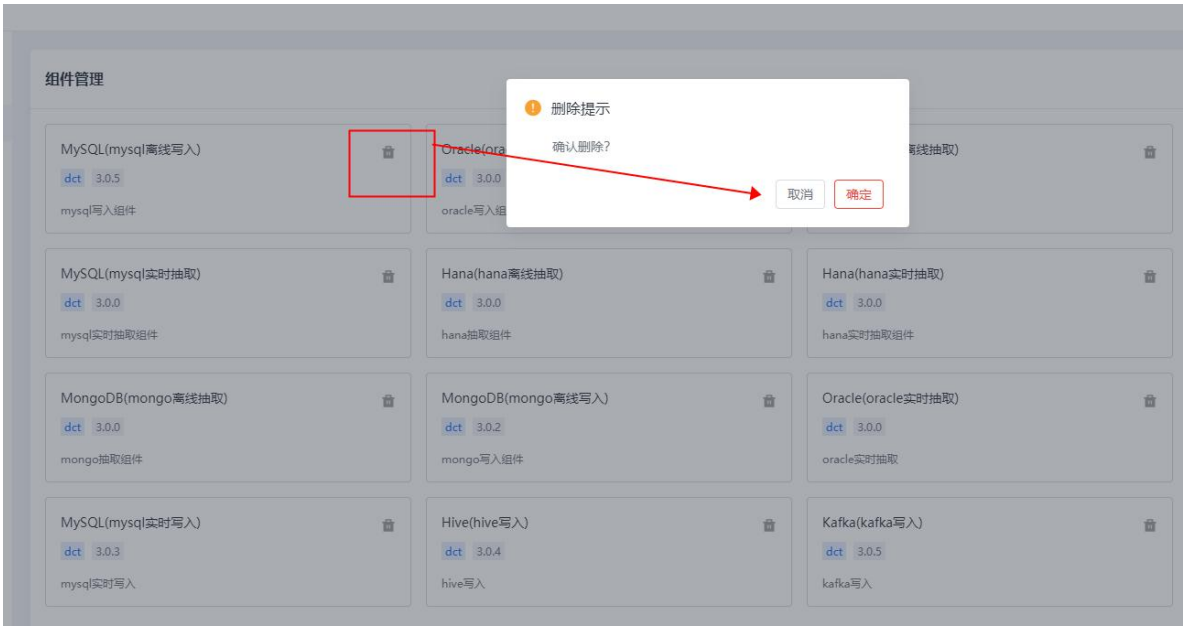
第一步：打开组件导出页面

第二步：选择需要导出的组件

第三步：点击【确定】按钮，导出选择的组件；点击【取消】则取消导出



3.10.7 组件删除



3.10.8 组件加载

当前版本默认本地加载

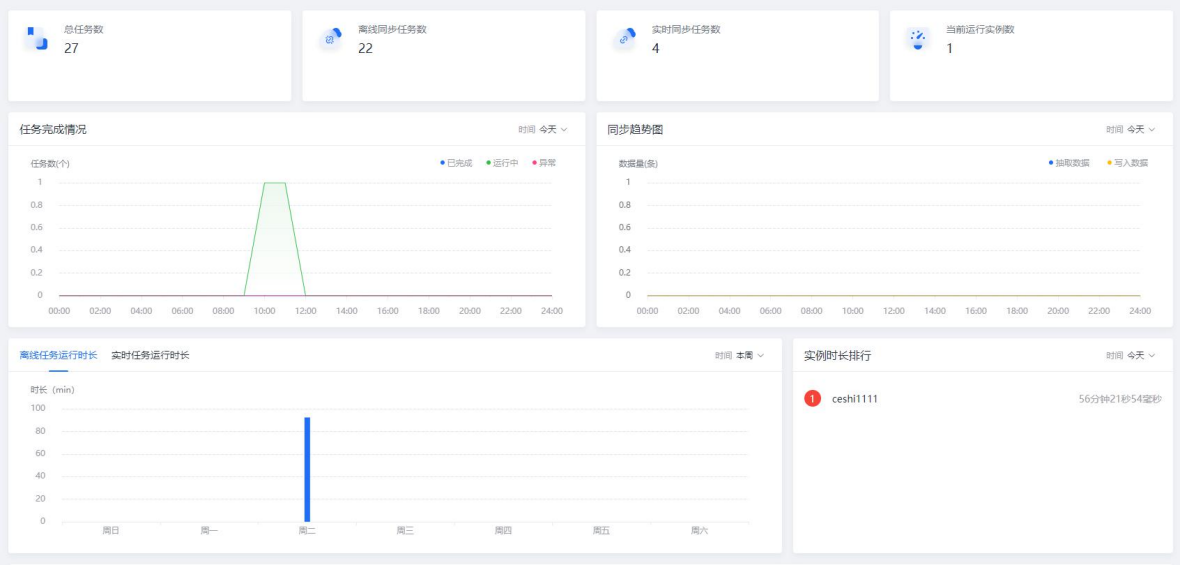
3.11 运维管理

3.11.1 集群资源监控

点击运维监控，默认展示当前管理端资源使用情况；点击下拉菜单，切换节点，查看集群其他资源占用情况

3.11.2 运维大屏

点击【运维管理】即可查看任务管理详情，包括总任务数、正在运行任务、错误数、实例运行排行榜，以及按照不同时间维度对当前总任务进行概览。



3.11.3 操作日志

点击【运维管理】下的【操作日志】即可查看任务操作日志。

可根据开始时间、截至时间、任务名称、任务类型、任务状态、任务结果进行筛选，点击【查询】按照设定条件筛选操作日志；点击【重置】则恢复筛选条件为默认。

操作结果 全部 操作类型 全部 开始日期 至 结束日期

操作IP	操作人	操作时间	操作链接	操作类型	操作信息	操作结果	创建时间
127.0.0.1	DCT	2021-04-08 10:54:14	/dct-manager/taskNodes	INSERT	操作模块[DAG节点模块], 描述[新增DAG节点信息], 操作方法[com.deepexi.dct.manager.service.TaskNodesServiceImpl.insert]	正常	2021-04-08 10:54:14
127.0.0.1	DCT	2021-04-08 10:54:11	/dct-manager/task	INSERT	操作模块[任务模块], 描述[新增任务信息], 操作方法[com.deepexi.dct.manager.service.TaskServiceImpl.insert]	正常	2021-04-08 10:54:11
127.0.0.1	DCT	2021-04-08 10:46:58	/dct-manager/taskEdges	INSERT	操作模块[DAG边模块], 描述[新增DAG边信息], 操作方法[com.deepexi.dct.manager.service.TaskEdgesServiceImpl.insert]	正常	2021-04-08 10:46:58
127.0.0.1	DCT	2021-04-08 10:46:58	/dct-manager/taskNodes	INSERT	操作模块[DAG节点模块], 描述[新增DAG节点信息], 操作方法[com.deepexi.dct.manager.service.TaskNodesServiceImpl.insert]	正常	2021-04-08 10:46:58
127.0.0.1	DCT	2021-04-08 10:46:58	/dct-manager/taskNodes	INSERT	操作模块[DAG节点模块], 描述[新增DAG节点信息], 操作方法[com.deepexi.dct.manager.service.TaskNodesServiceImpl.insert]	正常	2021-04-08 10:46:58
127.0.0.1	DCT	2021-04-08 10:46:57	/dct-manager/task	INSERT	操作模块[任务模块], 描述[新增任务信息], 操作方法[com.deepexi.dct.manager.service.TaskServiceImpl.insert]	正常	2021-04-08 10:46:57
127.0.0.1	DCT	2021-04-08 10:35:48	/dct-manager/run	UPDATE	操作模块[任务模块], 描述[启动任务], 操作方法[com.deepexi.dct.manager.service.TaskServiceImpl.start]	正常	2021-04-08 10:35:48
127.0.0.1	DCT	2021-04-08 09:36:16	/dct-manager/task	UPDATE	操作模块[任务模块], 描述[更新任务信息], 操作方法[com.deepexi.dct.manager.service.TaskServiceImpl.update]	正常	2021-04-08 09:36:16
127.0.0.1	DCT	2021-04-08 09:34:42	/dct-manager/alertRule	INSERT	操作模块[预警规则模块], 描述[新增预警规则], 操作方法[com.deepexi.dct.manager.service.AlertRuleServiceImpl.insert]	正常	2021-04-08 09:34:42
127.0.0.1	DCT	2021-04-07 19:19:57	/dct-manager/taskNodes	INSERT	操作模块[DAG节点模块], 描述[新增DAG节点信息], 操作方法[com.deepexi.dct.manager.service.TaskNodesServiceImpl.insert]	正常	2021-04-07 19:19:57

共 18972 个条目，每页显示 10条 1/1898 页

3.11.4 实例管理

点击【运维管理】下的【实例管理】即可查看任务运行实例。

可根据开始时间、截止时间、任务类型、任务状态进行筛选，点击【查询】按照设定条件筛选运行日志；点击【重置】则恢复筛选条件为默认。

实例管理

运行状态 全部 搜索

实例ID	任务名称	运行状态	时延	同步数量	错误数量	开始时间	运行时长
46a945b983d94f9681653c1b6281b1351617351980526	mysql实时任务一	成功	0 ms	5	0	2021-04-02 16:54:13	3天21小时55分钟30秒54毫秒
80ae3c358c6a46f1873396fdcb4d5d8a1617195780066	mysql-mysql实时二	成功	0 ms	13	0	2021-03-31 21:03:00	21小时53分钟39秒50毫秒
8ba6b1be80304a47bb4397902a5164611617196364569	mysql-mysql实时一	成功	0 ms	5	0	2021-03-31 21:12:44	21小时43分钟54秒51毫秒
2706d8e0bd694f58bffc2bc35e2deadc1617196558106	123	成功	0 ms	3	0	2021-03-31 21:15:58	21小时40分钟42秒50毫秒
b6bc9df4343547afbcfba2e179c0f3971617705933833	666	成功	1 ms	11	0	2021-04-06 19:16:21	19小时6分钟33秒52毫秒
ee0fe7c8b5d04152852b6644c374d3ca1617708422179	ceshi1111	成功	1 ms	5	0	2021-04-06 19:27:02	18小时55分钟54秒50毫秒
1a53b69dede54d35a536e2b69038282116177022936 81	45	成功	0 ms	0	0	2021-04-06 17:44:54	53分钟3秒52毫秒
be0a108c86e04ea48f0547940427c70416172662298 50	hana实时-mysql	成功	0 ms	676,200	0	2021-04-01 16:37:10	51分钟47秒983毫秒
35837349fa6a42a0829b8792e95b94341617679339476	测试一	成功	0 ms	0	0	2021-04-06 11:22:19	15分钟24秒46毫秒
f15838f79a41417f8c220b892717e3971617701088798	4.6实时任务一	成功	0 ms	1	0	2021-04-06 18:39:28	13分钟33秒52毫秒

4 常见问题

4.1 数据源或目标库测试连通性失败怎么办?

- 检查每一项内容是否填写正确
- 检查数据源/目的地版本是否支持及相关要求是否满足
- 检查相关权限是否开启
- 检查数据源/目的地与 DCT 宿主机是否连通
- 检查数据源/目标库用户名/密码是否过期

4.2 单任务是否支持多个表是否可以写到目的地一张表?

- 单集群的可以支持多张表写入同一张表，要求表结构一致；
- 多集群的分库分表不支持一个任务实现多张表写入同一张表，需要分多个任务实现

4.3 如何应对 Mysql 数据库表和字段名称大小写不敏感问题?

- Mysql 数据库默认不会开启大小写敏感，所以当数据同步到 Mysql 目的地时，对于

包含相同大小写的表或者字段名称 Mysql 数据库会被视为同一个表或者字段名称

- 若需避免大小写敏感引起的错误，需通过自定义映射规则规避

4.4 Hive 数据源支持哪些文件格式?

- 目前 Hive 数据源只支持 Textfile 和 Parquet 的文件格式

4.5 如何避免 Hive 目的地出现小文件问题?

- 目前在往 Hive 目的地写入数据时，如果对实时性要求高，选择实时同步的方式，源端每新增数据目的地就会对应生成相应的文件，因此小文件出现就会比较多。如果对于实时性要求不是很高的话，推荐用户选择定时读取的方式，可通过增大写入时间间隔来减少小文件的产生。

4.6 常见问题持续更新中