

云数据实验室 2.0 入门手册

内部资料，请勿外传

目录

- 云数据实验室 2.0 入门手册.....1
- 阅读对象.....2
 - 添加用户.....2
 - 建立实验室.....3
 - 添加口碑数据.....3
- 名词解释.....4
 - 开发区.....4
 - 生产区.....4
 - 发布.....4
 - 数据加工.....5
- 口碑数据集.....6
 - 交易表 adm_odm_app_ot_trd_dd.....6
 - 店铺表 adm_odm_app_ot_shop_dd.....7
 - 实验区.....7
 - 发布.....7
 - 生产区.....9
- ETL 开发.....9
 - 实验区.....9
 - SQL 案例11
 - DDL: create_koubei_test.sql11
 - DML: insert_koubei_test.sql11
 - 验证.....11
 - 发布.....12
 - 生产区.....13
- 数据分析.....13
 - 实验区.....13
 - 发布.....14
 - 生产区.....15
- 外部接口调用.....16
- 圈人营销.....18
 - 标签发布.....18
 - 圈人 api.....20
- 附录20
 - 帮助文档.....20

阅读对象

- 1. 面向数据开发人员，有一定的 SQL 和 Java 经验。
- 2. 已经成功入驻并建立实验室，并导入数据。

添加用户

蚂蚁数巢实验室 数据助力创新 sys_lab

我的项目 我的数据 管理设置 刘裕

用户管理

添加用户

安全审计日志

账号名	用户名	角色	添加时间	操作
-----	-----	----	------	----

添加用户

* 电子邮箱:

* 昵称:

* 姓名:

* 手机:

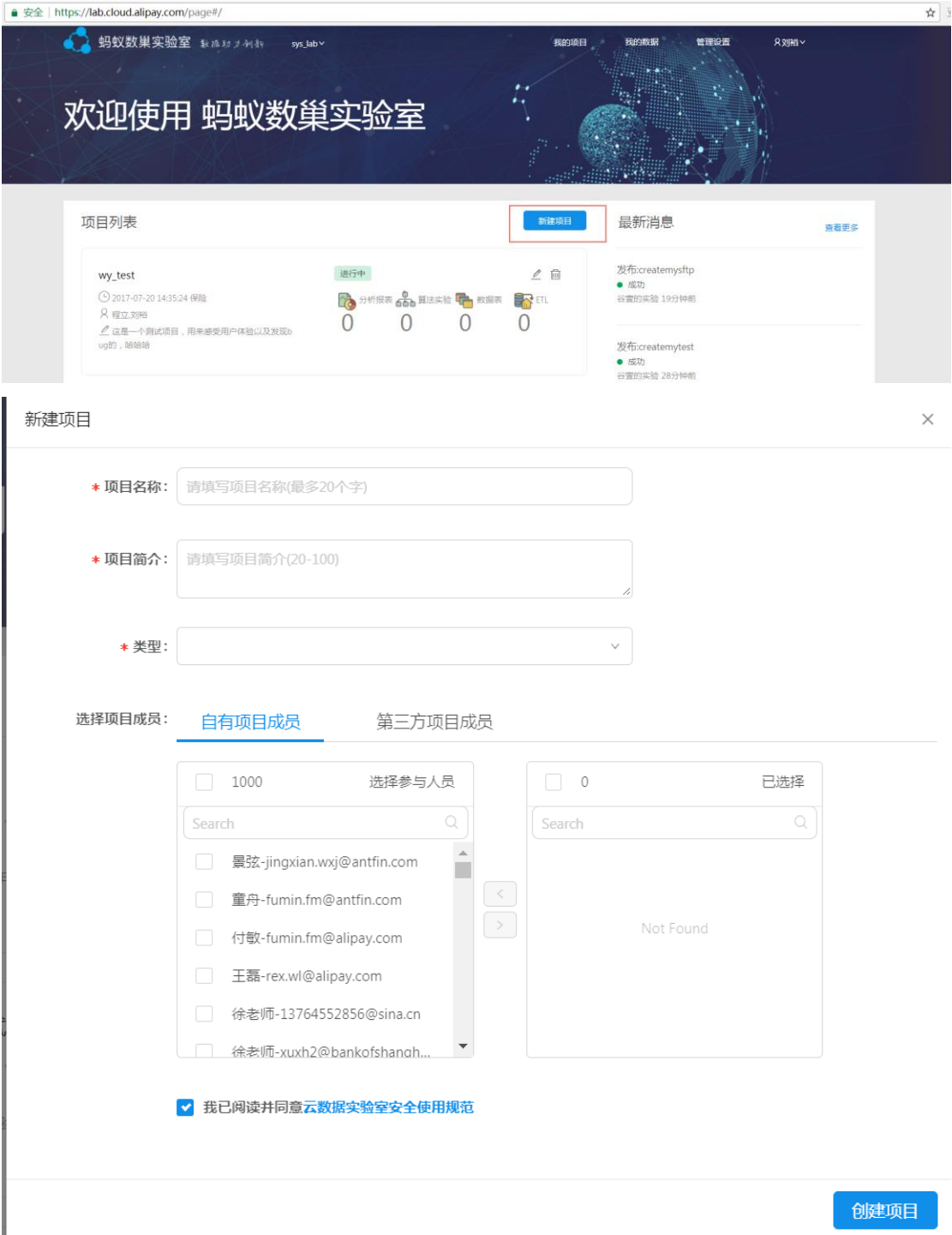
添加至项目:

账号角色: ☒ 操作员 ☐ 管理员

取消

确定

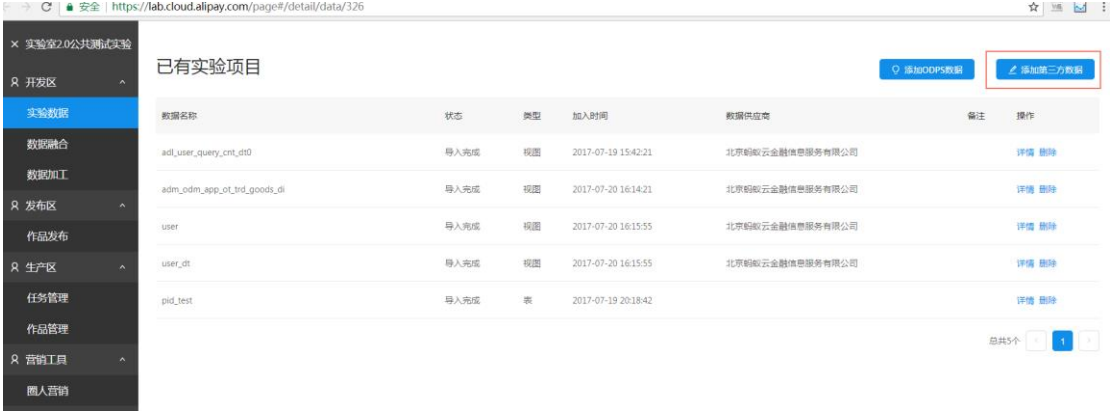
建立实验室



添加口碑数据

建立实验室后，会根据 ISV 授权关系分配数据。由于需要跑 ODPS 任务，“添加他方数据”要等待一段时间才能看到口碑数据。如果“添加他方数据”长时间看不到数据，请联系口碑业务

方进行数据分配。



状态变成“导入完成”，表明数据表已进入到客户实验区。如果导入失败，需要重试下。

名词解释

开发区

一种虚拟区域，又叫做开发区，是进行 ETL 开发，数据分析，算法试验的主要工作区域。

生产区

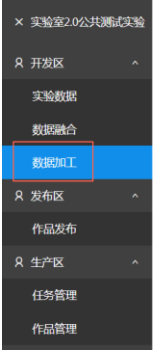
一种虚拟区域，实验区生成的表，ETL 脚本，算法试验，报表分析经过发布操作后到达该区域，发布后的报表分析数据可供应用（app_id）调用。

发布

1. 将实验区生成的表，ETL 脚本，算法试验发布到生产区，作为报表分析操作的基础数据和过程数据；

2. 将实验区生成的报表分析进行发布，供被订阅的应用（`app_id`）进行接口调用，获取报表数据。

数据加工



ETL开发

数据开发
云端的数据抽取
加载和处理
[进入](#)



数据分析

商业分析
用你熟悉的方式
完成透视分析
[进入](#)



算法平台

人工智能
可视化方式交互
的机器学习平台
[进入](#)

口碑数据集

入门手册会从交易表 adm_odm_app_ot_trd_dd 和店铺表 adm_odm_app_ot_shop_dd 抽取数据进行 ETL。

交易表 adm_odm_app_ot_trd_dd

adm_odm_app_ot_trd_dd 数据详情

字段	描述	类型
pid	PID	STRING
user_id	会员ID	STRING
trade_no	交易号	STRING
gmt_pay	交易时间	STRING
shop_id	交易门店id	STRING
shop_name	交易门店名称	STRING
trade_amt	交易金额	BIGINT
total_discount_amt	商家优惠金额	BIGINT
mct_receive_amt	商家实收金额	STRING
store_id	商家内部自有门店id	STRING
is_user_voucher	是否使用优惠券	BIGINT
home_shop_distance	居住地距离门店距离,单位...	DOUBLE
work_shop_distance	工作地距离门店距离,单位...	DOUBLE
dt	-	STRING

店铺表 adm_odm_app_ot_shop_dd

adm_odm_app_ot_shop_dd 数据详情

字段	描述	类型
shop_id	门店id	STRING
pid	pid	STRING
shop_name	门店名称	STRING
longitude	经度	DOUBLE
latitude	纬度	DOUBLE
xian_area_name	门店所在区县	STRING
city_area_name	门店所在城市	STRING
prov_area_name	门店所在省份	STRING
city_degree	门店所在城市等级	STRING
external_shop_id	在口碑平台开店时填写的...	STRING
dt	-	STRING

实验区

请参考“阅读对象 -> 添加口碑数据”。导入成功后，表明这张表进入到实验区。

发布

选择左侧“发布管理”，点击“+”，发布类型选择“表”。

实验室2.0公共测试实验

开发区

实验数据

数据融合

数据加工

发布区

作品发布

生产区

任务管理

作品管理

营销工具

圈人营销

发布

发布记录

发布类型: 表

☐ 表名

☐ pid_test_not_found

☐ temp

☐ new_table_for_input

☐ user

下一步

取消

发布

发布记录

发布类型: 表

表名

☐ adl_user_query_cnt_dt0

☒ adm_odm_app_ot_trd_goods_di

☐ user

☐ user_dt

下一步

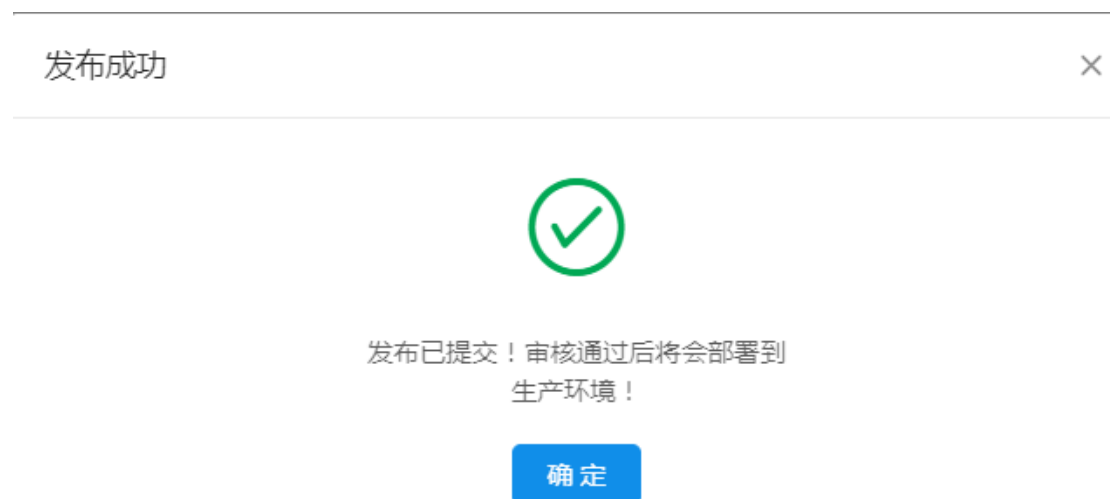
取消

发表表

×

名称	数据类型	发布内容
adm_odm_app_ot_trd_goods_di	VIEW	<div><div></div>DDL<div><div></div>DDL和数据</div></div>

发布



发布成功后，可以在“发布记录”中看见。

生产区

发布成功后，可以在“生产区”“作品管理”“我的数据表”看到相应的口碑表，表示这张表已经进入到生产区。



ETL 开发

实验区

打开“开发者工具 => ETL 开发”进入到数据工厂（[帮助文档](#)，[ODPS SQL 文档](#)）操作界面，如下图。

从数据安全方面考虑，实验室采用短信验证码进行二次验证。可以给每个操作员添加账号和对应的手机号，用于接收短信验证码。

手机验证



您好!基于数据安全风险理由,该模块需要短信验证,感谢您的理解.

手机号码: 136*****39

[点击获取验证码](#)

* 验证码:

[确认提交](#)



ETL开发

数据开发
云端的数据抽取
加载和处理

[进入](#)



数据分析

商业分析
用你熟悉的方式
完成透视分析

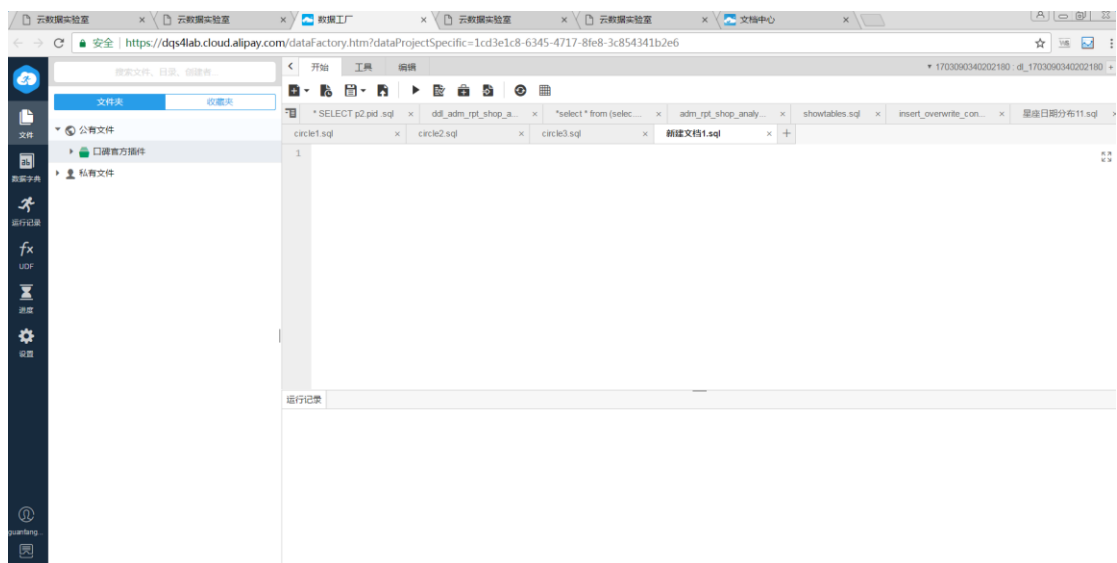
[进入](#)



算法平台

人工智能
可视化方式交互
的机器学习平台

[进入](#)



SQL 案例

入门手册提供的案例是：对交易明细表按照 pid(商户 id)分组汇总，统计交易金额。

DDL: create_koubei_test.sql

新建 ETL 脚本 create_koubei_test.sql，并执行。内容如下。

```
drop table if EXISTS koubei_test;
```

```
CREATE TABLE koubei_test
```

```
(pid string, sum bigint);
```

DML: insert_koubei_test.sql

新建 ETL 脚本 insert_koubei_test.sql，并执行。内容如下。

```
INSERT OVERWRITE TABLE koubei_test
```

```
SELECT pid
```

```
,SUM(trade_amt) AS s
```

```
FROM adm_odm_app_ot_trd_dd
```

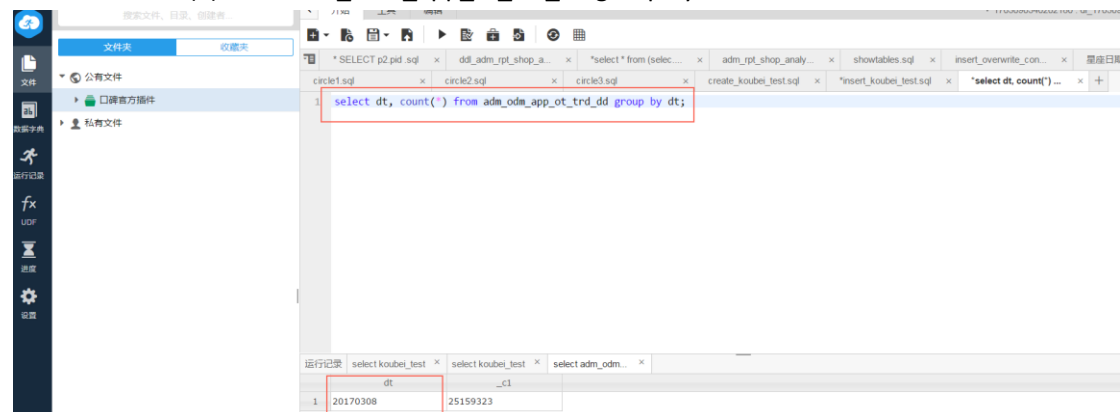
```
WHERE dt = '20170416'
```

```
GROUP BY pid ;
```

具体的 dt 请用下面的 sql 语句查询。

```
select distinct dt from adm_odm_app_ot_trd_dd;
```

```
select dt, count(*) from adm_odm_app_ot_trd_dd group by dt;
```



验证

执行成功后，可以从表 koubei_test 里面查询数据进行验证（pid 等敏感字段做了数据脱敏）。

备注：如果想通过后面的外部接口调用，这里需要查询到接口授权同一商家的数据（实验区 pid 做了加密，生产区不做加密）。

10

11

```
select * from koubei_test;
```

运行记录

select koubei_test

×

	pid	sum
1	072c374c7bdade56cd5b1*	1647151
2	1343b4fe15d86440387f5*	1835313
3	26668347b10bea50de28c*	1050410
4	28433e2e2745da550ea65*	384606279

发布

保存成功后，进入发布管理理，将上述 ETL 操作发布到生产环境，具体操作步骤如下：
发布类型选择， ETL 脚本，进入一次性发布 tab 页。首先发布 create_koubei_test。



接着发布 insert_koubei_test

生产区

两个脚本都发布成功后，可以在“生产区”“作品管理”“我的数据表”看到相应的口碑表，表示这张表已经进入到生产区。

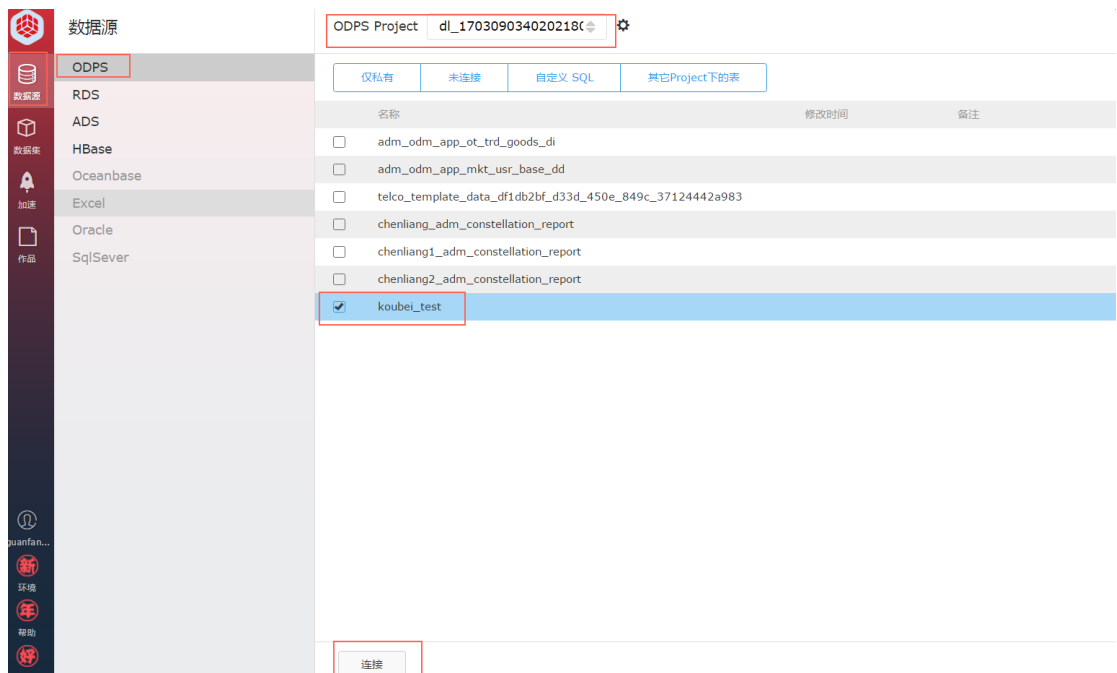


数据分析

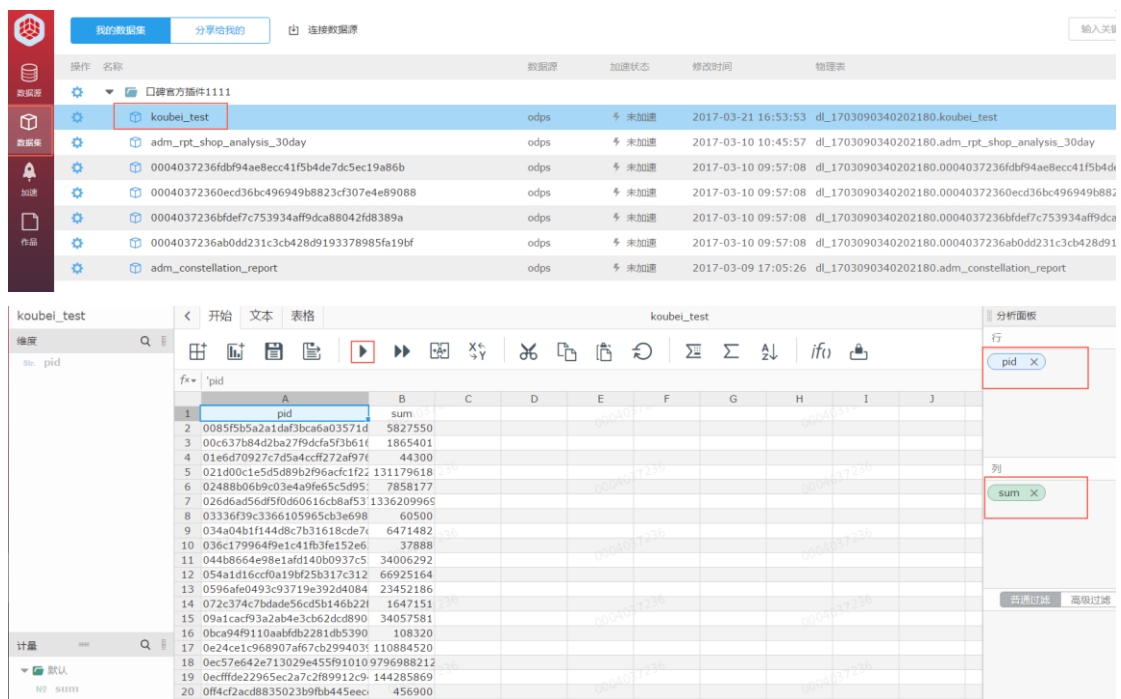
实验区

打开“开发者工具=>数据分析”进入工作表界面（[帮助文档](#)），点击左侧数据源，选择 ODPS 数据源，连接刚才通过 ETL 操作生成的表，如下图：





连接成功后，进图数据集界面，打开连接的表，进入工作表界面，选择维度和计量字段，进行汇总统计，如下图：



发布

分析报表保存成功后，需要进行发布操作，将报表数据发布到生产环境，供外部应用调用；进入发布管理理,选择发布类型“报表分析”，进入工组表 tab 页，选择刚才生成的工作表进行发布，如下图：



生产区

发布成功后，可以在“生产区=>作品管理=>我的分析”中查看发布结果，加速状态为已加速的分析才能被外部程序访问到；
如下图：



点击详情，可以看出参数查询配置。



外部接口调用

根据语言环境，下载响应蚂蚁金服服务端 SDK 包：[链接地址](#)

备注：KoubeiMarketingDataAlisisReportBatchqueryRequest

和 KoubeiMarketingDataAlisisReportQueryRequest 需要口碑业务方开通权限。

由于数据是属于商家的，ISV 直接调取可以返回报表列表，但具体报表的数据会返回 NULL。ISV 想调取具体报表数据，需要获得商家的 token 授权（因为新加了接口包，授权会变），详情需联系口碑业务方。

已授权此应用使用以下接口，如应用有新增接口，需要引导商户再次授权，[查看如何引导商户授权](#)

请输入API名称、或接口中文名

搜索

API名称	接口中文名
alipay.marketing.cdp.advertise.query	查询广告接口
koubei.marketing.data.alisis.report.query	报表数据查询接口

如果想通过外部接口调用到报表具体数据，这里接口授权的商家数据，需要在“ETL 开发 -> 实验区 -> 验证”里面查询到（实验区 pid 做了加密，生产区不做加密），并对 DML: insert_koubei_test.sql 进行 ETL 一次性发布和加速成功。

将 SDK 包引入到具体开发环境 IDE 中，java 调用接口程序如下：

```
try {
    AlipayClient alipayClient = new DefaultAlipayClient(
        GET_POST_URL,
        APP_ID,PRIVATE_KEY,DEFAULT_FORMAT,DEFAULT_CHARSET,PUBLIC_KEY);
    KoubeiMarketingDataAlisisReportBatchqueryRequest kbBatRequest =
        new KoubeiMarketingDataAlisisReportBatchqueryRequest();
    kbBatRequest.setBizContent("{\"" + "\"page\":\"1\",\"" + "\"size\":\"30\""+ "\"}");

    KoubeiMarketingDataAlisisReportBatchqueryResponse    kbBatResponse    =
alipayClient.execute(kbBatRequest);
    if (kbBatResponse.isSuccess()) {
        System.out.println("报表列表查询接口    调用成功");

        List<AlisisReport> reportList = kbBatResponse.getReportList();
        String totalCount = kbBatResponse.getTotalCount();
        System.out.println("totalCount:"+totalCount);
        for (AlisisReport report:reportList) {
            String reportUk = report.getReportUk();
            String reportName = report.getReportName();
```



```

String reportDesc = report.getReportDesc();
System.out.println("report_uk:"+reportUk+";
report_Name:"+reportName+"; report_Desc:"+reportDesc);

KoubeiMarketingDataAlisisReportQueryRequest kbRequest =
    new KoubeiMarketingDataAlisisReportQueryRequest();
kbRequest.setBizContent("{\"" + "\"report_uk\": \"" + reportUk + "\" +
    "\"}");

```

//ISV 代理商家调取报表数据，需要传入 app_auth_token（因为新加了接口包，授权会变），详情需联系口碑业务方
**//Request.putOtherTextParam("app_auth_token", "201511BBaaxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx");**

```

KoubeiMarketingDataAlisisReportQueryResponse kbResponse =
alipayClient.execute(kbRequest);

```

```

if (kbResponse.isSuccess()) {
    System.out.println("报表数据查询接口 调用成功");
    List<AlisisReportRow> reportRows = kbResponse.getReportData();
    for (AlisisReportRow reportRow:reportRows) {
        List<AlisisReportColumn> columns = reportRow.getRowData();
        for (AlisisReportColumn column:columns) {
            String clumnKey = column.getAlias();
            String clumnVal = column.getData();

            System.out.println("clumnKey:"+clumnKey+"clumnVal:"+clumnVal);
        }
    }
} else {
    System.out.println("报表数据查询接口 调用失败");
}

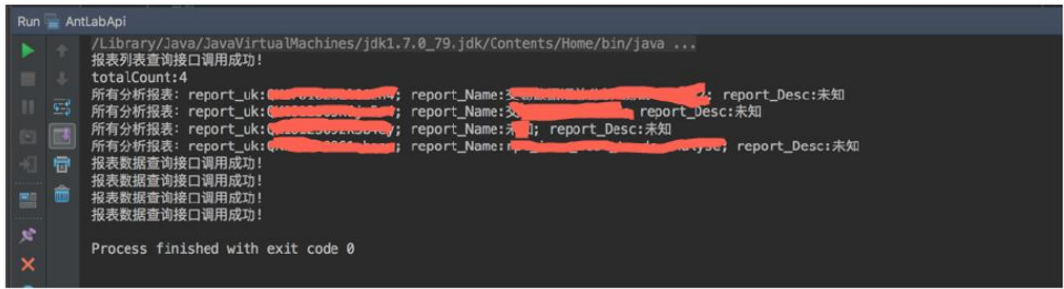
}

} else {
    System.out.println("报表列表查询接口 调用失败");
}

} catch (Exception e) {
    e.printStackTrace();
}

```

外部调用接口返回如下：



圈人营销

标签发布



新建标准标签

* 数据集:

adm_odm_app_mkt_usr_base_dd

* 分区字段:

dt

* 分区字段格式:

yyyyMMdd-1d

* 圈人主键:

user_id

* 行业:

餐饮

下一步

<input checked="" type="checkbox"/>	字段名	字段类型
<input type="checkbox"/>	user_id	STRING
<input checked="" type="checkbox"/>	age	BIGINT
<input checked="" type="checkbox"/>	gender	STRING
<input type="checkbox"/>	constellation	STRING
—		

上一步

下一步

对应表名：

adm_odm_app_mkt_usr_base_dd

✓

对应字段名：

gender

✓

数据类型：

string

✓

类目：

标签名：

gender

✓

显示类型：

多选枚举

▼

值	显示	操作
0	男	删除
1	女	删除

添加

保存并前往上一个

保存并前往下一个

取消

确定

圈人 api

详情请见开放平台圈人接口 https://docs.open.alipay.com/api_23 。

[koubei.marketing.campaign.tags.query](#)

[koubei.marketing.campaign.crowd.create](#)

[koubei.marketing.campaign.crowd.batchquery](#)

[koubei.marketing.campaign.crowd.detail.query](#)

[koubei.marketing.campaign.crowd.modify](#)

[koubei.marketing.campaign.crowd.delete](#)

[koubei.marketing.campaign.crowd.count](#)

附录

帮助文档

[ETL 开发](#)

[上传本地文件](#)

[任务调度](#)

[数据分析](#)

[算法平台](#)

[开放平台 SDK](#)

[圈人接口](#)

[ODPS SQL](#)（现称为 MaxCompute，可以搜索帮助）

帮助与文档

请用关键词进行搜索，例如“服务器密码重置”

搜全部 搜本产品

搜索热词：远程连接服务器 挂载数据盘 域名解析 域名实名认证 账号实名认证 忘记密码

文档 新手学堂 自助工具 联系客服

☆ 我的收藏

全部产品 三

大数据计算服务

产品简介 下

快速开始 下

大数据计算服务 > SQL > SQL 概要

SQL 概要

更新时间：2017-03-10 16:34:41 分享领百元红包：👉