

素材系统分享

为什么需要素材系统

当前业务的素材管理依赖外购的创量平台，涉及头条/广点通/快手等媒体，业务支撑能力和数据分析能力有限。端内与投放业务对素材管理均有诉求，特别对优质素材的甄别和复用重视度高，**期待可以进一步优化素材模块，以达到能支持日常业务使用的目标。打通素材上传-存储-贴标-推送-看数链路。**

目标&价值

1. 基础能力：建设全渠道素材管理能力，面向业务方，打造一个长期好用、高效的素材管理平台。

- a. 支持全局素材管理，包含全局素材信息/文件集中管理(存储/增删改查)、全局素材分析、代理商素材管理等能力；
- b. 支持全媒体渠道的素材对接，包含素材推送、素材审核、素材报表等能力；
- c. 释放人力，降低成本，提高人效，增强素材自动化，包含定期推送、自动打标、优质素材识别、自动审核等能力。

2. 增值服务：打通素材与内投链路，通过素材数据，挖掘素材增值机会点，素材驱动业务增长。

- a. 素材溯源，打通广告素材-用户归因链路，**建立素材与端内承接，挖掘机会点；**
- b. 优质素材，多维素材数据评估，发现并沉淀优质素材，**拓展复用，提升投放效果。**
- c. 批量创建广告

项目调研

php开发的旧系统存在问题：素材中心因数据不准等其他历史原因，业务并没有使用起来。

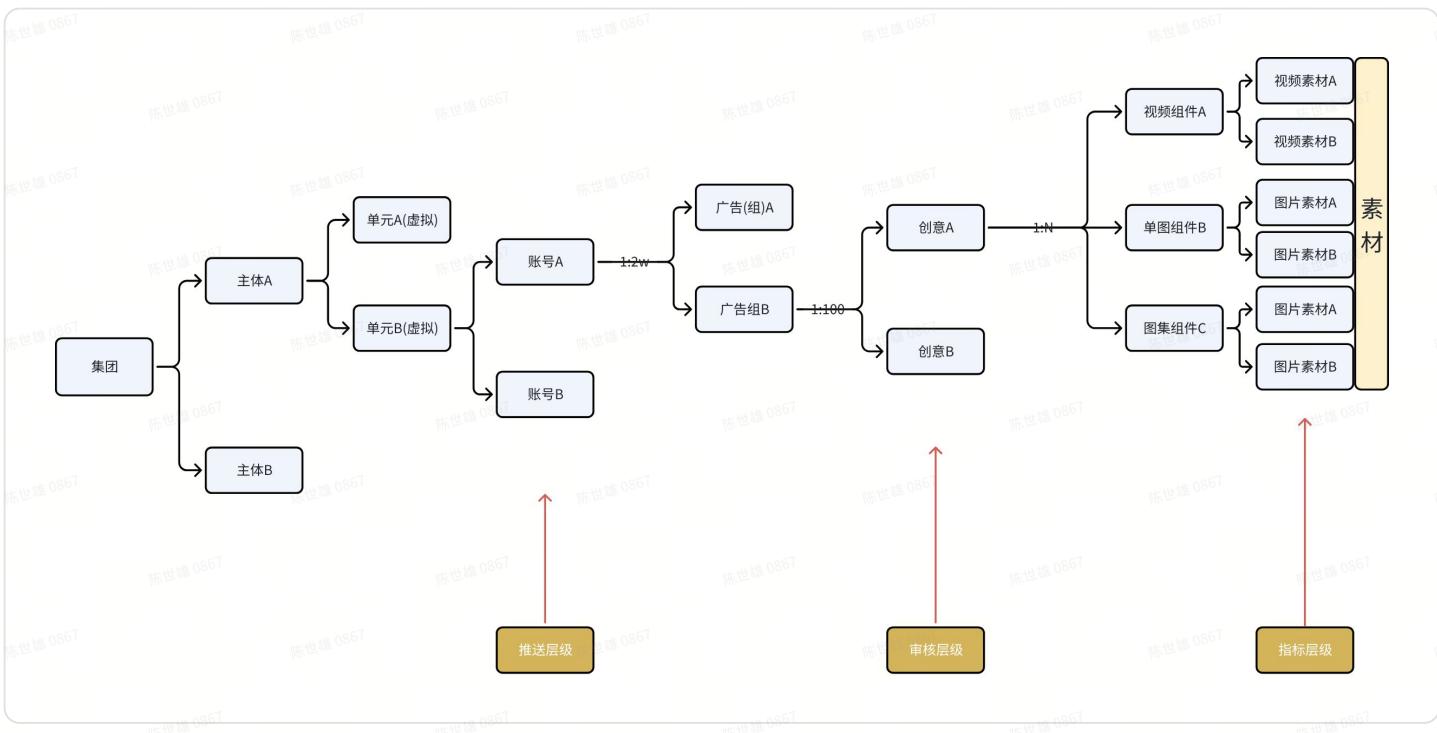
运营期望有的功能

有几个信息

1. 日常的素材业务会涉及哪些方面？除以下之外还有哪些？
 - a. 素材内部制作：真人情景实拍/翻剪...，通过内部同学上传
 - b. 代理商素材：
 - c. 内部素材：
 - d. 媒体素材：
 - e. 达人素材：标准平台投放，从媒体平台拉取，归属媒体素材
2. 当前使用创量管理素材，遇到哪些痛点或槽点？
 - a. 素材后向数据无法提供
 - b. 数据拉取不全，只能拉取自运营数据，代理运营拉取不了；数据准确性欠缺
 - c. 内部人工上传工作量较大，支持第三方上传，但数据权限范围需控制(向外素材系统)
3. 素材业务开展过程中，目前最想解决的痛点或问题有哪些？
 - a. 制作：如何提升批量制作能力；更倾向素材负责人员关系记录
 - b. 代理：缺少素材管理模块
 - c. 内部：
 - d. 媒体：
 - e. 无法感知素材后向的详细数据
 - f. 缺少承接可用的素材标签
 - g. 支持同时上传到多个素材库文件目录中去
4. 站在业务角度看，素材管理最需要的能力有哪些？
5. 对过往的素材管理功能，有什么想吐槽或建议的？
 - a. 上传素材较慢，性能需提高
 - b. 素材指标与上传的素材没对应，以至于数据无法查看，素材多端一致性问题
 - c. 查询性能较低/数据稳定性不高
6. 目前优先解决哪个/哪些媒体的素材问题？
 - a. 优先头条，Q4完成头条、广点通和快手对接

怎么做

广点通



单图

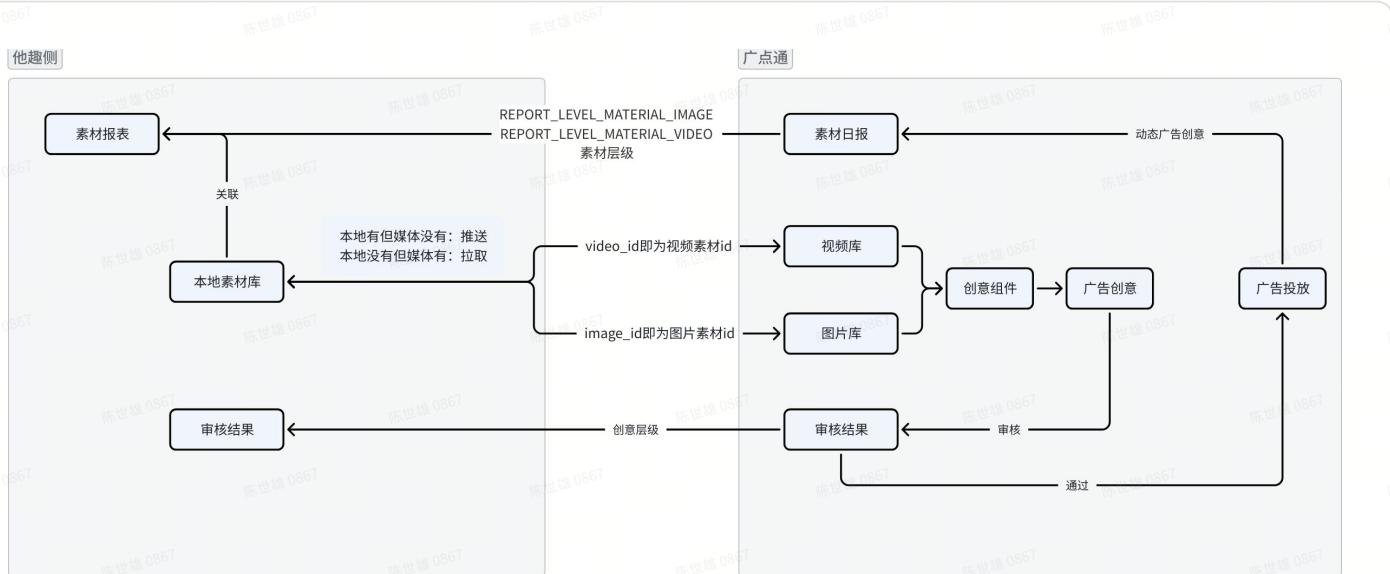
```

        "creative_components": [
          {
            "image": [
              {
                "value": {
                  "image_id": "14300556597"
                }
              },
              {
                "value": {
                  "image_id": "1430055914"
                }
              },
              {
                "value": {
                  "image_id": "1429056919"
                }
              }
            ]
          }
        ],
        "main_jump_info": [
          {
            "value": {
              "page_type": "PAGE_TYPE_XJ_QUICK",
              "page_spec": {
                "xj_quick_spec": {
                  "page_id": 0
                }
              }
            }
          }
        ]
      }
    
```

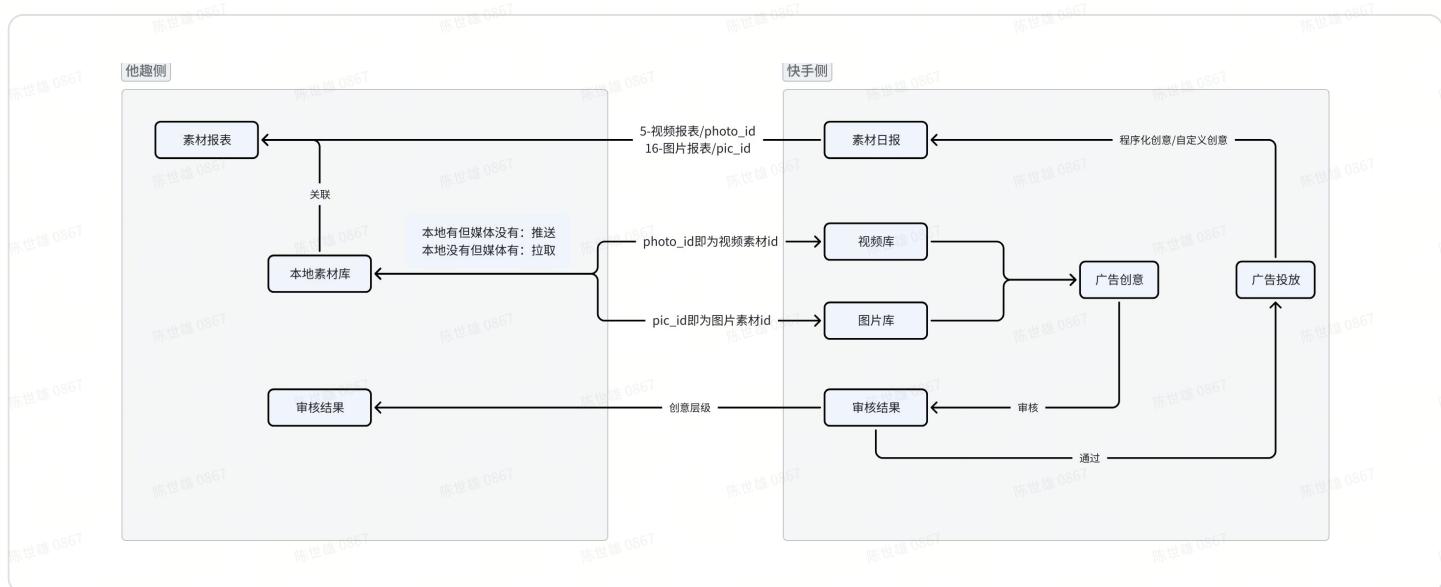
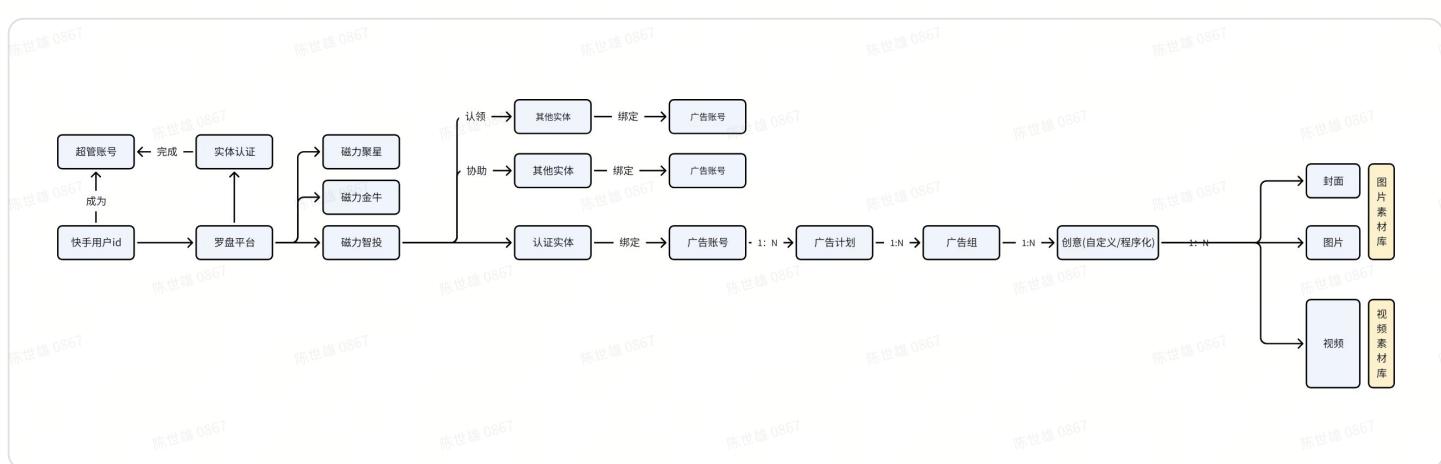
图集

```

        "image_list": [
          {
            "value": {
              "list": [
                {
                  "image_id": "14300556597"
                },
                {
                  "image_id": "1430055914"
                },
                {
                  "image_id": "1423566956"
                }
              ]
            }
          }
        ],
        "brand": [
          {
            "value": {
              "jump_info": {
                "page_type": "PAGE_TYPE_WEAUTH_CHANNELS_PROFILE"
              }
            }
          }
        ]
      }
    
```



快手



设计与实现

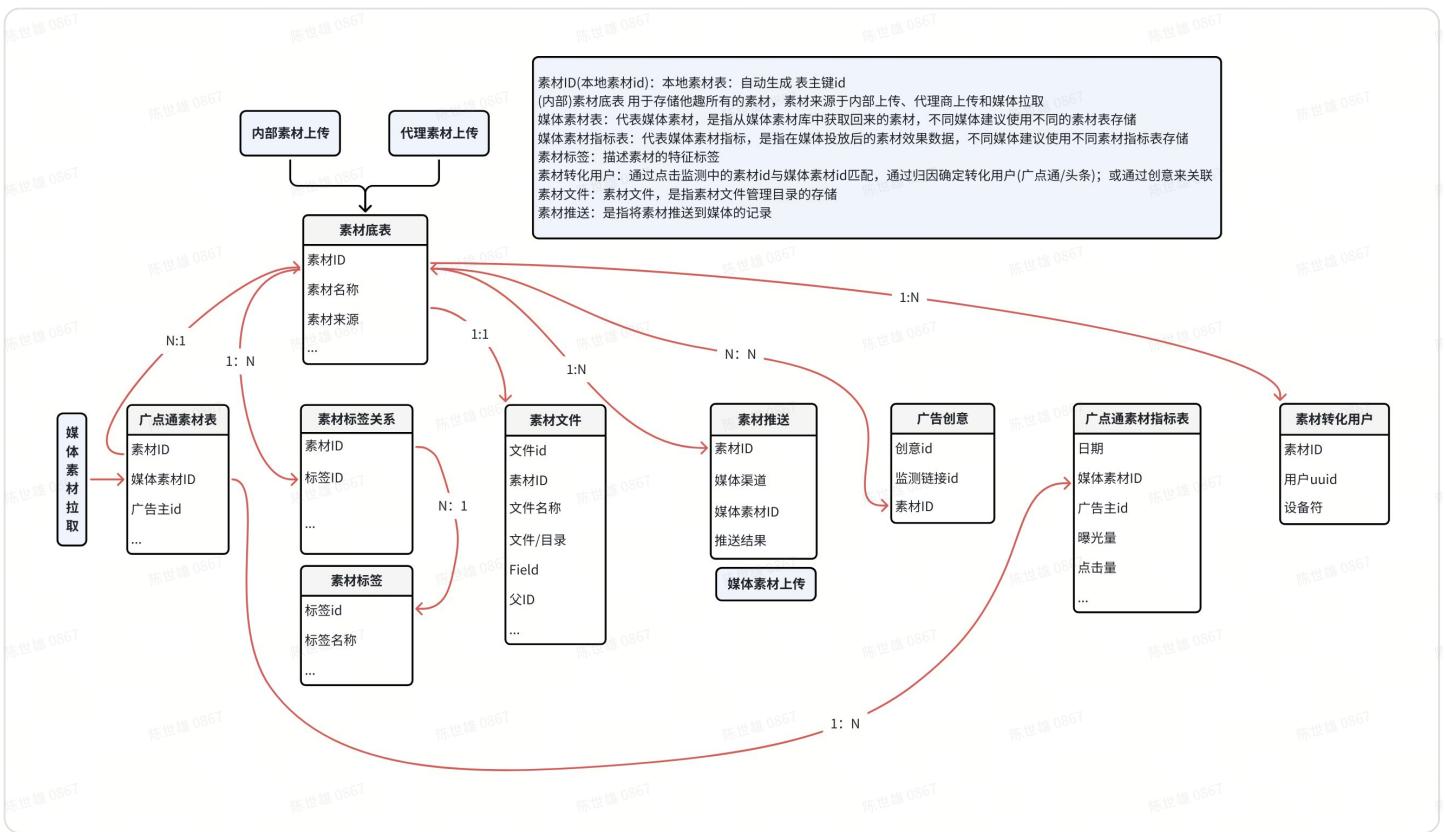
👉 后台地址: <https://admp-pro.admin.internal.taqu.cn/material/material-lib/index>

文档: [自投放平台-素材管理用户手册](#)

产品架构



实体关系

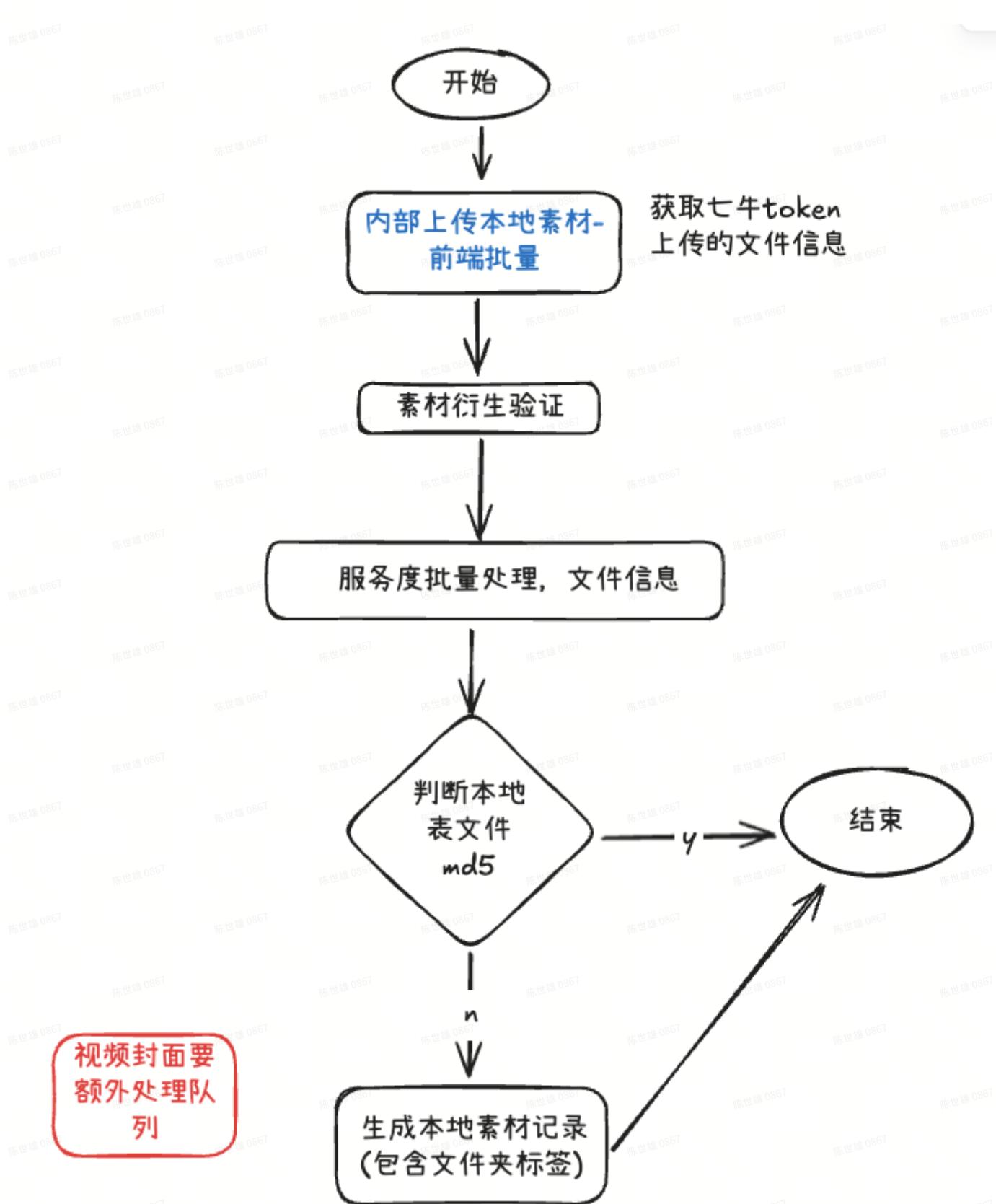


思维导图

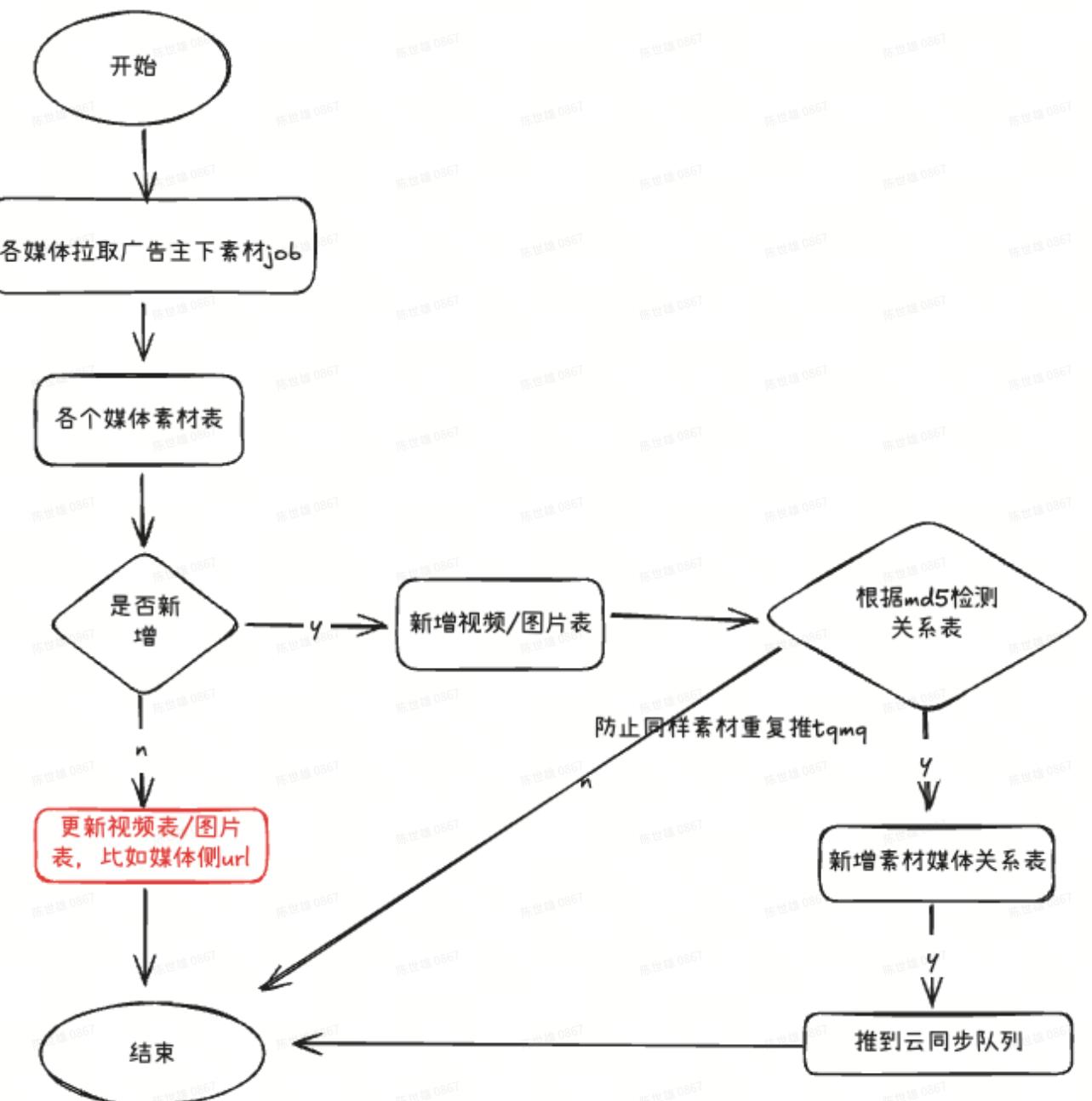


核心环节

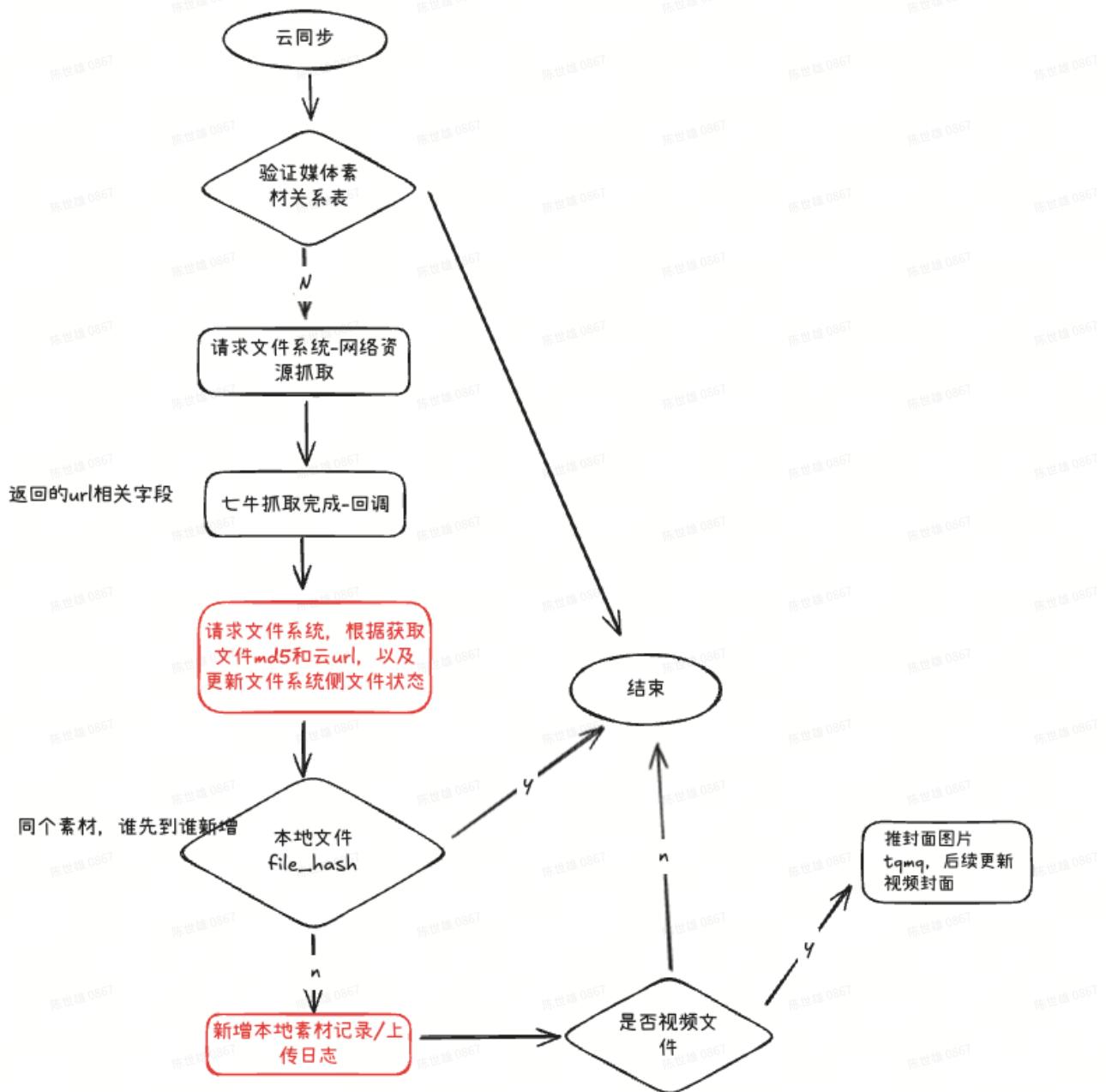
批量创建本地素材



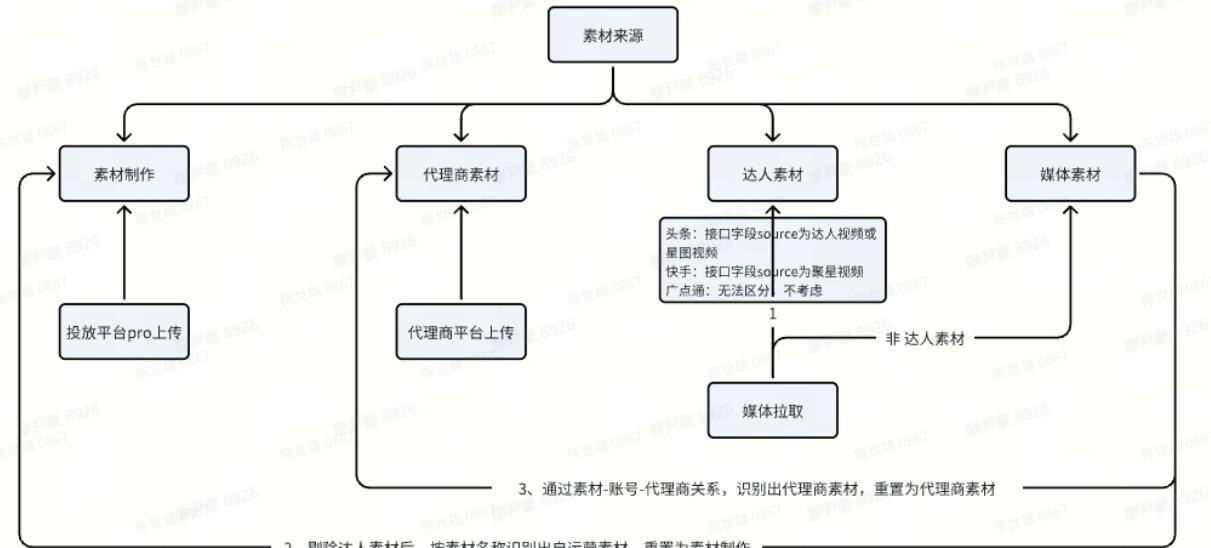
媒体素材拉取



云同步Pulsar消费流程



素材来源的定义



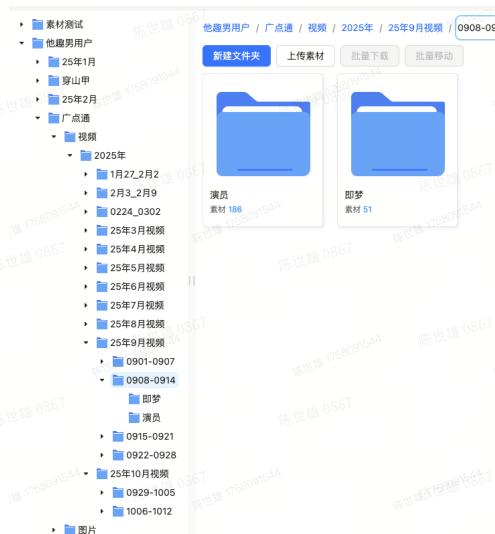
亮点

文件模块

一开始的文件夹不要求显示素材文件数量，于是直接用了邻接表数据模型，通过递归查询是从这种模型中提取层次信息的，后续运营提了此需求

特性	邻接图/邻接表 (Adjacency List)	嵌套集合 (Nested Set)
查询直接子项	⚡ 极快 (WHERE s **极快** (WHERE parent_id=?))	△ 较复杂 (需计算深度或层级)
查询完整子树	△ 慢 (需递归/CTE)	⚡ 极快 (BETWEEN lft AND rgt)
查询祖先路径	△ 中等 (递归查 parent_id 链)	⚡ 快 (lft < ? AND rgt > ?)
插入节点	⚡ 极快 (更新 parent_id)	△ 极慢 (重排后续所有左右值)
移动节点/目录	⚡ 快 (仅改 parent_id)	△ 极慢 (删除+插入+重排左右值)
删除子树	△ 需递归删除	△ 删除快, 但需重排后续左右值
空间占用	⚡ 低 (仅 parent_id)	△ 高 (lft + rgt)
实现复杂度	⚡ 简单直观	△ 复杂, 易出错
适用场景	频繁增删改, 层级较浅	频繁查询子树, 几乎不修改结构

考虑正常使用pid的结构去递归获取该路径下的所有素材数量效率很慢



使用嵌套集合模型去存储数据：**高效解决 SQL 树形数据查询的创新方案 轻松地执行许多与层次结构相关操作，例如查找给定节点的所有子节点、查找给定节点的父节点、计算子节点的数量等等。**

- 查询汇总简单高效
- 无需递归查询，性能高
- 缺点：不易维护因为左右值的存在，会直接影响到后续的节点，因此，当前节点增删改时，都会对后续的节点产生影响



单纯的使用嵌套集合模型也不行，会存在哪些问题？

本来设计文件夹节点和素材节点，结果素材节点太多了，会导致慢sql 以及并发时，抢锁问题

更新rt 字段，用rt字段做where 条件，并发执行每次更新都要锁等待

1. 并发可以弄成串行，刚初始化时整张表大小才8w数据，应该是抢锁引发的慢sql
2. 后续素材节点膨胀 还是会导致每次update太多数据 update这个几乎全表了，走索引比直接扫描全表还慢，索引还得二次回表



优化

1. 使用嵌套集合模型只存储文件夹节点数据去掉素材节点，通过num字段维护素材文件数量
2. 素材数量通过文件表里的pid归属，先进行初始化，然后每次素材变更对应文件夹节点的num字段进行异步变更

素材全文搜索

了解背景：本地素材表优化方案

- 一般为了规避连表查询，会借助了ES，通过ES合并主表和扩展表数据进行检索。比如私信业务的消息搜索
- clickhouse查询（现有推广报表使用）百万级别的表和千万级别的表，可以在ck生成视图支持查询1-2s，存在问题：**数据延迟10-30秒，且并发数不高**
- Hologress特别适合需要同时处理实时数据和历史数据，并且对查询性能有较高要求的场景。**（毫秒级别的延时）**

素材拉取

各个媒体的接口，存在各类限流，且还会因为大盘整体限流

- 控制最大协程数
- 令牌桶限制频率 QPS QPM (后续考虑使用sentinel-go)
- 延时队列进行重试（保证不遗漏增量素材）
- 专门写历史素材job, 配合数开捞取历史遗漏素材

本地素材系统第三方拉取-问题与解决方案

引进ai打标

本地素材打标一期-2025Q2

服饰风格

蕾丝上衣 针织衫
无袖连衣裙 露脐装
长袖上衣 牛仔短裤 白褶裙
开衫 短裙 无袖上衣 T恤
包臀裙 短袖 吊带 衬衫连衣裙
男 西装 短裤 牛仔裤 露肩装
紧身裙 露肩裙 吊带裙 背心
皮衣 外套 黑色上衣 旗袍
无袖 蕾丝裙 高跟鞋

待优化点

1. 目前本地素材库的表属性并不多，数据量也就300w左右。还是不是大表，宽表



2. 素材推送，媒体一般都有素材共享的概念，当推送任务配置的广告主后，可以通过归纳相同主体的广告主，只上传其中一个广告主，其余的直接共享素材就行。减少带宽的使用

