# SVN

## 数数转义插件

<https://192.168.0.126/svn/tech_web/bigdata/SINCETIMES>

* 包括转义插件内容以及各项目的转义配置文件

## Trino中间表定时任务

前端: <https://192.168.0.126/svn/tech_web/bigdata/gbsp_web>

* 中间表定时任务前端

后端: <https://192.168.0.126/svn/tech_web/bigdata/gbsp>

* 中间表定时任务后端
* OpenmateData预警一键API部署
* 龙女版本配置文件更新

## Superset

<https://192.168.0.126/svn/tech_web/bigdata/superset>

* 包括前后端内容

## 老铁后台

<https://192.168.0.126/svn/DragonGirls/DragonGirlsWeb/trunk/huntress_stat>

* 龙女统计页面

# 数据分析部主要工作内容

## Superset

1. 所有的二次开发内容, 维护, 更新可视化模型
2. 帮助分析师使用以及解决他们出现的页面问题

## 定时任务

数据分析部在数据治理上将原始日志定位Dwd层, 分析师提供模型将Dwd层提炼出Dws层, 在根据Dws层提炼Ads层数据, 可视化报表中90%的日志是从Dws层和Ads层计算数据, 减少需要计算的日志量, 也使计算的步骤更简便. 管理更方便

定时任务系统是保证 Dws层和Ads层的计算能按时有序进行. 以及维度表的上传,

主要工作有以下几点

1. 系统开发, 开发已经完成, 短时间内不会有功能上的开发
2. 系统维护 (开发已经完成, 基本不会有太大的变动)
3. 日常补录(如果当天任务没有运行成功, 需要人工介入时我需要手动补录数据)
4. 历史重跑(如果Dws和Ads层有模型变更或者增加, 我需要将对应项目的所有历史数据进行统一的计算)
5. 维度表上传

## 数数

1. 所有的数数转义插件, 有特殊需求就进行插件更新
2. 核对数数转义文档的准确性, 在接入阶段要将错误的转义事件告知项目组并修改
3. 数据纠错, 当发现错误时需要排查错误发生的原因, 如果需要补录数据的话, 操作数据补录
4. 帮助分析师使用平台, 他们有不太会使用的功能需要帮助他们熟悉和使用

## OpenmateData预警

1. 部署预警SQL, 确保Dws Ads层数据每天能按时运行成功
2. 使用webhook对接企业微信, 一旦出现错误及时纠错



## 龙女配置文件

1. 龙女每次版本更新需要手动将项目提供的排版文件解析并录入数据库给分析使用

## 老铁统计平台

1. 坦克, 战舰. 海战,龙女, 这些项目如果报表没有数据, 需要去排除原因和补录

主要就是补录数据, 功能不会有更新了

## 其他

1. 数数adjust数据和游戏数据匹配
2. 数数adjust因入库时间影响, 每天会有一个修复脚本再运行
3. 数数adjust 的 cost花费分摊给具体的用户身上
4. Superset AI聊天助手部署工作流
5. 服务器监控(这部分工作之前转交给了冉桦林)
6. 龙女, 战舰, dow的游戏公告系统(一直在运行)

# 数数接入流程

## 对接日志打点, 确认日志格式和文档

由**分析师**和项目组沟通需要的日志打点,

确定日志的字段格式, 中英文名称, 类型等. 并整理成文档

## 数据采集

由**冉桦林**将服务器的日志通过Flume程序实时传输到内部Logbus机器

## 确认支付日志是走数据库还是日志形式

如果是数据库形式.

由**杨德贵**确认支付数据字段名称以及数据类型, 并提过导出SQL给桦林

由**冉桦林**将数据导出数据库并通过Flume程序实时传输到内部Logbus机器

## 确认数数转义插件.

确认数数转义插件功能能满足日志转义需求, 如果有特殊要求, 需要**杨德贵**开发兼容功能

## 数数网页控制台添加项目,和权限

由**习多**在网页控制台中新增一个项目,并设置对应人员的权限.

根据需求对能复用部分的报表进行封箱和开箱(从已有项目将报表配置复制到新项目,封箱和开箱是数数的术语)

## 配置Logbus并测试数据

由**杨德贵**提供配置文件

由**习多**实际操作启动logbus的进程启动

## 数数接入Adjust日志

和第6步一样, adjust的日志接收就在本地. 直接设置好配置文件就可以正常上报

## 分析报表创建

**分析师**根据业务需求和日志打点. 创建对应的报表页面, 可能会涉及以下数数组件的创建

* 虚拟事件
* 虚拟属性
* 维度表
* 用户分群
* 用户标签
* 指标创建
* 报表创建

## 上线前清库

如果正式上线有清库要求, 需要在数数的服务器执行清库的脚本

# Superset接入流程

## 对接日志打点, 确认日志格式和文档

由**分析师**和项目组沟通需要的日志打点,

确定日志的字段格式, 中英文名称, 类型等. 并整理成文档

## 数据采集

由**冉桦林**将服务器的日志通过Flume程序实时传输到大数据集群机器

## 数据入库

由**冉华林**将日志录入到大数据底层数据库, 并构建trino的关联关系

## 构建中间表

由**分析师**提供中间表模型, **杨德贵**在定时任务中根据模型部署和定时运行中间表, 如有历史数据需要**杨德贵**跑所有时间的历史中间表数据

## 预警部署

由**杨德贵**在openmatedata中部署通用的中间表数据预警SQL.

## 新建项目和权限开放

由**习多**在superset操作台新建项目, 并开通对应的superset权限以及trino权限

## 导入和新建报表

由**杨德贵**将老项目公用的报表负责到新项目报表

由**分析师**在superset上创建项目的业务需求报表

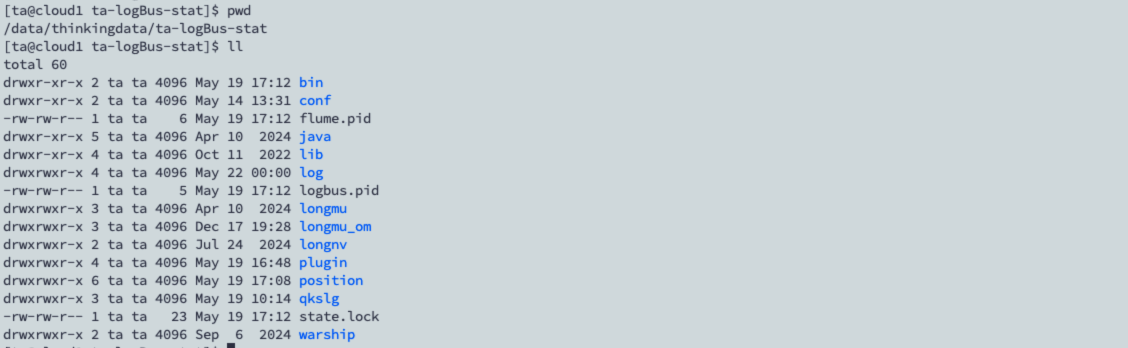
# 数数操作流程

## 官方使用文档

<https://docs.thinkingdata.cn/ta-manual/v4.3/>

## Logbus使用

1. 文件结构

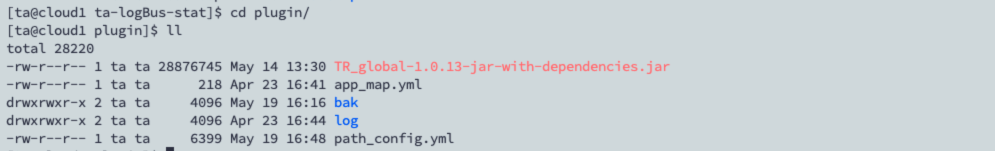


* bin 启动停止等脚本
* conf logbus配置文件

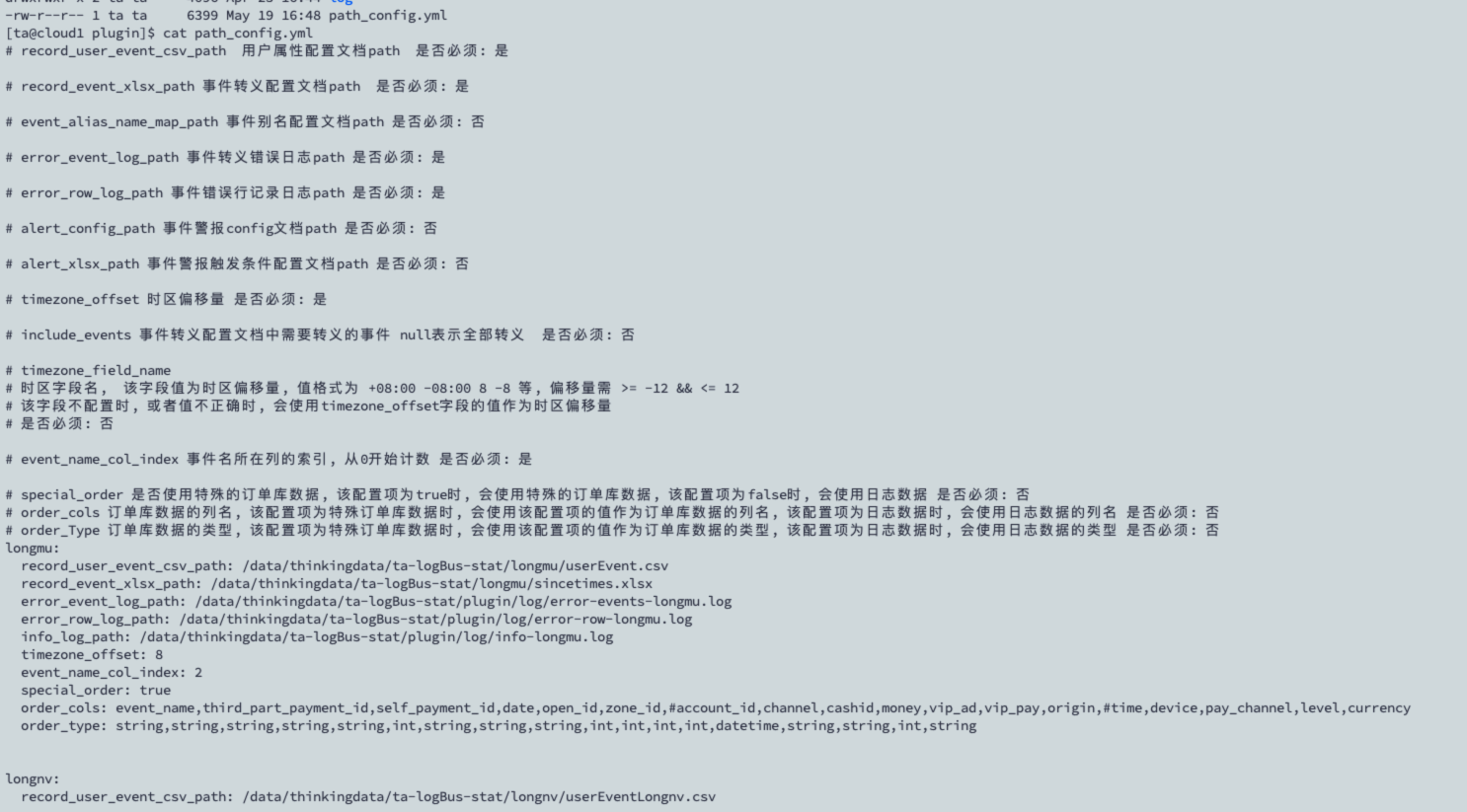
参数配置官方文档

<https://docs.thinkingdata.cn/ta-manual/v4.3/installation/installation_menu/data_import/logbus_installation.html#%E4%B8%8B%E8%BD%BD-logbus>

* log 运行日志文件
* longmu, longnv, longmu\_om等 各项目的转义配置文件
* position logbus读取日志的位置信息
* plugin 转义插件的配置文件



* app\_map 数数项目对应的配置映射
* path\_config.yml 各项目转义插件的配置信息.



## 常用命令

1. 启动停止Logbus进程

```

bin/logbus start

bin/logbus stop

```

1. 查看日志信息

```

cat log/logbus.log

cat log/flume/flume-error-log.log

cat log/flume/flume-info-log.log

```

1. 删除数数数据

<https://docs.thinkingdata.cn/ta-manual/v4.3/technical_document/ta_tool/data_del.html#%E4%B8%80%E3%80%81%E7%AE%80%E4%BB%8B>

需登录数数的服务器

```

su - ta

ta-tool data\_del

```

1. 数数数据回溯

<https://docs.thinkingdata.cn/ta-manual/v4.3/technical_document/ta_tool/user_event_import.html>

```

su - ta

/data/home/ta/develop\_tool\_ta/bin/ta-tool user\_event\_import -conf {config path}

```