**DataFairy开发说明文档 V1.0**

|  |  |  |  |
| --- | --- | --- | --- |
| 版本 | 备注 | 编制人 | 时间 |
| V1.0 | 初稿《DataFairy开发说明文档》 | 胡嘉宏 | 2016-10-14~  2016-10- |
|  |  |  |  |

广州图灵科技有限公司

目录

[1. 项目架构 1](#_Toc1796865122)

[1.1. 项目简介 1](#_Toc1621525526)

[1.2. 核心流程 1](#_Toc871573843)

[2. 项目配置及执行 2](#_Toc33769569)

[2.1. 项目下载、编译 2](#_Toc1495548974)

[2.2. 运行环境 2](#_Toc653155033)

[2.3. config.json配置文件 2](#_Toc407096020)

[2.4. 执行 2](#_Toc1385745154)

[2.5. 更新配置文件 2](#_Toc1437793194)

[3. 数据库结构 3](#_Toc2039131544)

[3.1. adpluto\_rpt表 3](#_Toc1897280856)

[3.1.1. 结构 3](#_Toc757912616)

[3.1.2. MySQL建表语句 3](#_Toc2081609072)

[3.2. adpluto\_rpt\_detail表 4](#_Toc775738916)

[3.2.1. 结构 4](#_Toc279739195)

[3.2.2. MySQL建表语句： 4](#_Toc2143222580)

[3.3. 视图 4](#_Toc669594174)

[3.3.1. campview 5](#_Toc1350717860)

[3.3.2. crpview 5](#_Toc1175517541)

[3.3.3. tagview 5](#_Toc223884434)

[3.3.4. getsize 5](#_Toc1202173365)

[4. 代码说明 6](#_Toc1665559676)

[4.1. 项目包结构 6](#_Toc151360272)

[4.1.1. src/main/java 6](#_Toc189621046)

[4.1.2. src/main/properties 6](#_Toc1796339556)

[4.2. com.turingdi.adpluto.entity 6](#_Toc318358506)

[4.2.1. com.turingdi.adpluto.entity.GlobalProperties 6](#_Toc1518188956)

[4.2.2. com.turingdi.adpluto.entity.RequestParams 6](#_Toc852492504)

[4.3. com.turingdi.adpluto.service 7](#_Toc1272966420)

[4.3.1. com.turingdi.adpluto.service.CookieStorer 7](#_Toc1918679868)

[4.3.2. com.turingdi.adpluto.service.DatabaseAccessor 7](#_Toc1234657341)

[4.3.3. com.turingdi.adpluto.service.URLAccessor 7](#_Toc922347894)

[4.4. com.turingdi.adpluto.starter 8](#_Toc1392721746)

[4.4.1. com.turingdi.adpluto.starter.AdPlutoStarter 8](#_Toc2106231185)

[4.5. com.turingdi.adpluto.utils 8](#_Toc956117463)

[4.5.1. com.turingdi.adpluto.utils.CommonUtils 8](#_Toc740787072)

[4.5.2. com.turingdi.adpluto.utils.Log4jUtils 8](#_Toc611902570)

[4.5.3. com.turingdi.adpluto.utils.MySQLUtils 9](#_Toc1363213483)

## 

# 项目架构

## 项目简介

本项目针对第三方监控服务，用于各项广告活动刷UV。程序在运行时，读取配置文件，按配置文件指定要刷的UV数量及PV与之的比例，模拟浏览器的访问行为向配置文件中制定的URL发出请求；下载页面及JS并执行后，同时修改DataTable系统数据库中对应报表数据表的记录，修改的对应渠道的展示量和点击量也可以在配置文件中指定（相对于UV的比例）。

## 核心流程

读取config.json初始化GlobalProperties对象

初始化MySQL连接池

遍历GlobalProperties中的尺寸、渠道、活动、创意包、人群标签和URL等属性的所有可能组合

发出的页面请求次数没达到配置要求

根据当前属性值对URL进行宏替换

根据配置指定的PV和UV比例，确定是否需要带着Cookie访问

从CookiesStorer中随机抽取Cookie

使用模拟Chrome访问当前URL

存入新Cookie

写入MySQL两个表

读取config.json初始化GlobalProperties对象

初始化MySQL连接池

发出的页面请求次数没达到配置要求

读取config.json初始化GlobalProperties对象

初始化MySQL连接池

关闭MySQL连接池

是

是

否

否

# 项目配置及执行

## 项目下载、编译

该项目由Maven构建，托管在内网Git服务器上，可通过以下命令下载项目代码：

git clone git@172.16.99.235:AdPluto.git

下载到本地后，使用Eclipse的导入现有maven项目功能导入下载的代码后，在Maven菜单中点击Update Project即可下载所需的依赖。

在项目上点击右键，弹出菜单中选择Run As，再选择Maven Package，即可完成编译打包工作，生成的jar包在项目目录下的target文件夹中。

## 运行环境

运行该项目需要JDK 1.8或以上版本，请确认系统已安装JDK1.8或以上版本并配置好JAVA\_HOME、PATH等环境变量。

## config.json配置文件

本项目的配置全部由JSON文件完成，配置文件名为config.json，请不要随意修改。

配置文件由9大项配置组成，其中"basic"为基本的配置，包括需要刷的UV数、及其他指标与之的比例；"url"为需要刷UV的广告主落地页，可以包括宏；"size"为需要遍历的创意尺寸；"camp"中包含为需要遍历的渠道（ADX）ID，及对应的需要遍历的活动ID；"ctid"为需要遍历的创意包ID；"spotid"为需要遍历的广告位ID；"tag"为需要遍历的人群标签值/ID，替换{tag}；"mysql"为MySQL配置，一般情况下不需要改动。

具体请参考config.json文件中的注释。

## 执行

在Windows命令行或Linux终端中进入jar包所在的目录，执行：

java -jar AdPluto.jar

即可。

## 更新配置文件

为了可以同时开多个DataFairy使用不同的配置文件，config.json文件是包含在jar包中的，更新配置文件可以使用项目目录中的一个Shell脚本。

使用前现将最新的config.json文件和AdPluto.jar包放在同一个目录下，然后执行：

./updateConfig.sh [新jar包名]

即可在此目录下创建一个包含新的config.json文件的jar包。

# 数据库结构

本项目使用DataTable项目的MySQL数据库，新增adpluto\_rpt和adpluto\_rpt\_detail表。其中adpluto\_rpt用于存储每个活动每天总的展示量与点击量，DataTable系统可以读取该表与正常数据叠加得到最终数据。adpluto\_rpt\_detail用于存储每个活动更细粒度的展示量和点击量，如创意包ID、尺寸、人群等等，用于运营人员进行多维度的查询。

## adpluto\_rpt表

### 结构

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MySQL表名** | **adpluto\_rpt** | | | |
| **主键** | **rpt\_date, iddsp, campid** | | | |
| **中文列名** | **列名** | **类型** | **必须** | **描述** |
| 报表日期 | rpt\_date | date | 是 | 三者构成主键 |
| 渠道ID | iddsp | int | 是 |
| 活动ID | campid | int | 是 |
| 展示量 | pv | int | 否 |  |
| 点击量 | click | int | 否 |  |

### MySQL建表语句

CREATE TABLE 'adpluto\_rpt' (

'rpt\_date' date NOT NULL,

'iddsp' int(11) NOT NULL,

'campid' int(11) NOT NULL,

'pv' int(11) DEFAULT NULL,

'click' int(11) DEFAULT NULL,

PRIMARY KEY ('rpt\_date','iddsp','campid')

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

## adpluto\_rpt\_detail表

### 结构

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MySQL表名** | **adpluto\_rpt\_detail** | | | |
| **主键** | **rpt\_date, iddsp, campid, ctid, size, tag** | | | |
| **中文列名** | **列名** | **类型** | **必须** | **描述** |
| 报表日期 | rpt\_date | date | 是 | 六者构成主键  size格式为 宽x高 |
| 渠道ID | iddsp | int | 是 |
| 活动ID | campid | int | 是 |
| 创意包ID | ctid | int | 是 |
| 创意尺寸 | size | varchar(255) | 是 |
| 人群标签 | tag | varchar(255) | 是 |
| 展示量 | pv | int | 否 |  |
| 点击量 | click | int | 否 |  |

### MySQL建表语句：

CREATE TABLE 'adpluto\_rpt\_detail (

'id' int(11) DEFAULT NULL,

'rpt\_date' date NOT NULL,

'iddsp' int(11) NOT NULL,

'campid' int(11) NOT NULL,

'ctid' int(11) NOT NULL,

'size' varchar(255) NOT NULL,

'tag' varchar(255) NOT NULL,

'pv' int(11) DEFAULT NULL,

'click' int(11) DEFAULT NULL,

PRIMARY KEY ('rpt\_date','iddsp','campid','ctid','size','tag')

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

## 视图

为了方便运营人员对DateFairy产生的展示/点击进行查询，我们还创建了一些View视图。

### campview

用于查询每个活动的展示/点击。

CREATE VIEW campview AS

SELECT iddsp, campid, sum(pv), sum(click)

FROM adpluto\_rpt\_detail

GROUP BY campid;

### crpview

用于查询每个创意包的展示/点击。

CREATE VIEW crpview AS

SELECT ctid, sum(pv), sum(click)

FROM adpluto\_rpt\_detail

GROUP BY ctid;

### tagview

用于查询每个人群标签的展示/点击。

CREATE VIEW tagview AS

SELECT tag, sum(pv), sum(click)

FROM adpluto\_rpt\_detail

GROUP BY tag;

### getsize

用于查询每种创意尺寸的展示/点击。

CREATE VIEW sizeview AS

SELECT size, sum(pv), sum(click)

FROM adpluto\_rpt\_detail

GROUP BY size;

# 代码说明

## 项目包结构

### src/main/java

|  |  |
| --- | --- |
| com.turingdi.adpluto.entity | 光音ADX相关的类，包括竞价Servlet及实体类。 |
| com.turingdi.adpluto.service | 科大讯飞ADX相关的类，包括竞价Servlet及实体类。 |
| com.turingdi.adpluto.starter | Letv ADX相关的类，包括竞价Servlet及实体类。 |
| com.turingdi.adpluto.utils | 整个项目中需要的一些工具类 |

### src/main/properties

包含所有配置文件，包括Log4j配置文件log4j.properties和项目配置文件config.json。

## com.turingdi.adpluto.entity

### com.turingdi.adpluto.entity.GlobalProperties

作者：胡嘉宏

类说明：本类为config.json配置文件对应的实体类，使用单例模式，在静态代码块中加载config.json文件并解析。该类主要由get、set方法构成，在此不再细述。唯一指出的是其单例的获取方法。

方法说明：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名 | 参数说明 | 返回值 | 详细说明 |
| public static GlobalProperties getGlobalProps() | 无 | GlobalProperties 对象 | 单例模式中用于返回本类唯一实例的方法。包含全局共用的配置。 |

### com.turingdi.adpluto.entity.RequestParams

作者：胡嘉宏

类说明：本类为对应每个尺寸、渠道、活动、创意包、人群标签和URL属性值组合的实体类，用于生成需要访问的URL。该类主要由get、set（只有mysqlAdxId属性有set方法）方法构成，在此不再细述。

## com.turingdi.adpluto.service

### com.turingdi.adpluto.service.CookieStorer

作者：胡嘉宏

类说明：本类用于存储Cookie，使用Set保存以去重，并提供存储一组Cookie和随机返回一组Cookie的方法。使用单例模式。

方法说明：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名 | 参数说明 | 返回值 | 详细说明 |
| public static CookiesStorer getInstance() | 无 | CookiesStorer 实例对象 | 单例模式中用于返回本类唯一实例的方法。 |
| public Set<Cookie> getRandomCookieSet() | 无 | Set<Cookie>一次HTTP请求所需的Cookie | 从储存的Cookie中随机抽取一组Cookie并返回 |
| public void addCookie(Set<Cookie> cookies) | cookies-HTTP响应中的Set-Cookie字段 | 无 | 保存一组Cookie，以供下次需要带Cookie访问URL时使用 |

### com.turingdi.adpluto.service.DatabaseAccessor

作者：胡嘉宏

类说明：本类用于在产生一次PV之后，按配置文件指定的比例对应地增加DataTable的数据库记录。使用单例模式。

方法说明：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名 | 参数说明 | 返回值 | 详细说明 |
| public static DatabaseAccessor getInstance() | 无 | DatabaseAccessor实例对象 | 单例模式中用于返回本类唯一实例的方法。 |
| public boolean incrDataBase(RequestParams req) | req-当前请求的URL各项属性对象 | boolean-是否成功执行 | 按配置文件指定的比例对应地增加DataTable的数据库记录。 |

### com.turingdi.adpluto.service.URLAccessor

作者：胡嘉宏

类说明：本类用于模拟点击访问URL，并通过控制访问时是否带Cookie来控制最终PV与UV的比例。使用单例模式。

方法说明：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名 | 参数说明 | 返回值 | 详细说明 |
| public static URLAccessor getInstance() | 无 | URLAccessor实例对象 | 单例模式中用于返回本类唯一实例的方法。 |
| public void accessURL(String clickURL, Random rand) | clickURL-需要模拟点击的URL  rand-用于产生随机数的Random 对象 | 无 | 模拟点击访问指定的URL，并按全局配置中PV和UV的比例控制访问时是否带Cookie。 |

## com.turingdi.adpluto.starter

### com.turingdi.adpluto.starter.AdPlutoStarter

作者：胡嘉宏

类说明：本类为本项目的入口，包含main方法。

方法说明：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名 | 参数说明 | 返回值 | 详细说明 |
| public static void main(String[] args) | args-执行时的输入参数，在此项目中无意义 | 无 | main方法，为本项目的入口。负责读取配置文件、初始化MySQL连接池，按指定的配置生成URL模拟浏览器访问，并对应地写入MySQL数据库。 |

## com.turingdi.adpluto.utils

### com.turingdi.adpluto.utils.CommonUtils

作者：胡嘉宏

类说明：本类为通用的工具类，包含URL宏替换方法。

方法说明：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名 | 参数说明 | 返回值 | 详细说明 |
| public static String microReplace(RequestParams req) | req-当前请求的URL各项属性对象 | String-进行后替换之后的URL | 对RequestParams对象中的URL进行宏替换并返回。 |

### com.turingdi.adpluto.utils.Log4jUtils

作者：胡嘉宏

类说明：本类为Log4j的工具类，用于维护一个静态的Logger常量。

方法说明：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名 | 参数说明 | 返回值 | 详细说明 |
| public static Logger getLogger() | 无 | Logger-用于写入日志的Logger对象 | 返回一个静态的Logger对象。 |

### com.turingdi.adpluto.utils.MySQLUtils

作者：胡嘉宏

类说明：本类为MySQL的工具类，用于管理/维护MySQL连接池及回收相关资源。

方法说明：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名 | 参数说明 | 返回值 | 详细说明 |
| public static void initMySQLPool() | 无 | 无 | 初始化MySQL连接池 |
| public static Connection getConnection() | 无 | Connection-  MySQL连接对象 | 从连接池中获取一个MySQL连接对象 |
| public static void closeResources(AutoCloseable... resources) | resources-需要关闭的资源，可变长度参数 | 无 | 关闭所有指定的资源，可以是Statement、Connection等 |
| public static void closePool() | 无 | 无 | 释放MySQL连接池 |