CMS 数据抓取方案设计文档

|  |  |
| --- | --- |
| 作者 | 郭志平 |
| 时间 | 2019-07-22 |
| 所属部门 | DSP |

# 文档简介

本文档主要介绍DSP 部门管理系统CMS,BC等系统获取微博数据，分析方案设计简介。

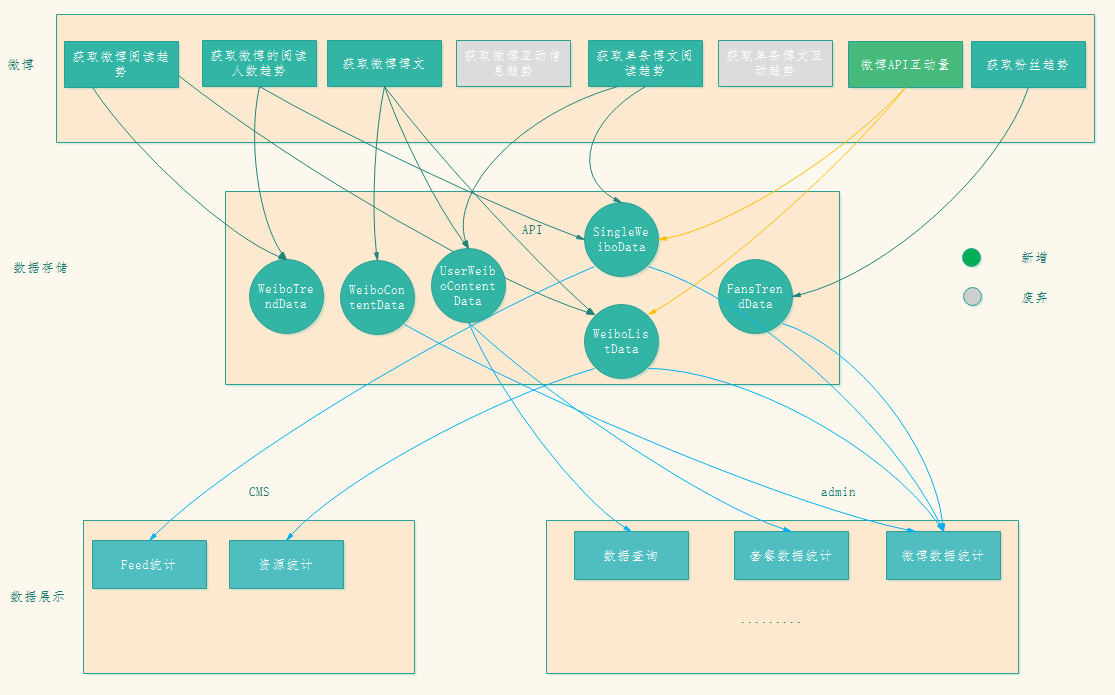
# CMS 数据处理方案

模拟登陆主要流程：

1. 获取微博界面提供数据的接口。
2. 模拟登陆微博。
3. httpClient 请求接口返回数据
4. 解析数据每一个返回字段，将数据存入对应的表中
5. 查询数据，界面展示。

API 方式获取流程：

1. 调用微博Open API数据通过uid获取到当前时间的用户Feed 数据(根据用户mid 判断，获取最新的数据)。
2. 获取前一天统计的数据，将今天的数据与前一天的数据求差值，得到当天统计的互动量统计数
3. 存入数据表中。



# 微博模拟登陆抓取接口简介

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 接口地址 | 接口简介 | 接口详细描述 | 使用方式 | 使用频率 |
| [https://dss.sc.weibo.com/pc/aj/chart/content/weiboReadTrend](https://dss.sc.weibo.com/pc/aj/chart/content/weiboReadTrend" \o "https://dss.sc.weibo.com/pc/aj/chart/content/weiboReadTrend) | 获取微博阅读趋势 | 获取某个UID在某个时间段内每天的发送微博总数和阅读总数 | 同步官微昨天的发送微博数和昨天的阅读数 | 91次/天 |
| <https://dss.sc.weibo.com/pc/aj/chart/content/weiboInteract> | 获取微博互动信息趋势 | 获取某个UID所有微博在某个时间段内每天的转评赞互动数以及这段时间转评赞互动的总数 | 同步官微昨天的转评赞和互动数 | 91次/天 |
| [https://dss.sc.weibo.com/pc/aj/chart/content/weiboReadPeople](https://dss.sc.weibo.com/pc/aj/chart/content/weiboReadPeople" \o "https://dss.sc.weibo.com/pc/aj/chart/content/weiboReadPeople) | 获取微博的阅读人数趋势 | 获取某个UID所有微博在某个时间段内每天的阅读人数以及这段时间的阅读总人数和平均每天阅读人数 | 同步官微昨天的阅读人数 | 91次/天 |
| <https://dss.sc.weibo.com/pc/aj/chart/fans/fansTrend> | 获取粉丝趋势 | 获取某个UID在某个时间段内每天的粉丝总数,粉丝增长数,粉丝减少数,粉丝净增长数,主动取关数,增长率以及这段时间内以上数据的各项总数 | 同步官微昨天相应的粉丝数据 | 91次/天 |
| [https://dss.sc.weibo.com/pc/aj/chart/content/weiboList](https://dss.sc.weibo.com/pc/aj/chart/content/weiboList" \o "https://dss.sc.weibo.com/pc/aj/chart/content/weiboList) | 获取微博博文 | 获取某个UID在某个时间段的所有的博文以及博文的一些信息,包括阅读数,转评赞…… | 同步官微七天前到昨天的博文内容 | 约700次/天 |
| [https://dss.sc.weibo.com/pc/aj/chart/content/singleWeiboReadTrend](https://dss.sc.weibo.com/pc/aj/chart/content/singleWeiboReadTrend" \o "https://dss.sc.weibo.com/pc/aj/chart/content/singleWeiboReadTrend) | 获取单条博文阅读趋势 | 获取某个具体的博文在15天内的每天的阅读数和阅读人数以及这段时间的阅读总数和阅读总人数 | 同步E接口获取的所有博文的阅读趋势 | 约6000次/天 |
| [https://dss.sc.weibo.com/pc/aj/chart/content/singleWeiboInteract](https://dss.sc.weibo.com/pc/aj/chart/content/singleWeiboInteract" \o "https://dss.sc.weibo.com/pc/aj/chart/content/singleWeiboInteract) | 获取单条博文互动趋势 | 获取某个具体的博文在15天内的每天的转评赞互动数以及这段时间转评赞互动的总数 | 同步E接口获取的所有博文的互动趋势 | 约6000次/天 |