帐号活跃表开发文档

1. **数据字典**
2. **表级**

表英文名：dws\_up\_service\_active\_dm

表中文名：帐号业务活跃汇总日表

1. **字段级**

|  |  |  |  |
| --- | --- | --- | --- |
| 字段英文名 | 字段中文名 | 数据类型 | 数据样例 |
| up\_id | 华为帐号编号 | VARCHAR(128) | 000049a66711a2b503edbf6f9455581a058371120495e5c2c8cbc4e357280bb4 |
| channel\_id | 渠道编号 | VARCHAR(128) | 26000005 |
| package\_name | 包名 | VARCHAR(256) | com.vmall.client |
| self\_register\_flg | 自注册标志 | SMALLINT | 1或0 |
| service\_set | 业务集 | VARCHAR(5000) | others#vmall#push |
| up\_type\_cd | 华为帐号类型代码 | VARCHAR(16) | 1、2、5、6、0、4、7、20、22或91 |
| register\_channel\_id | 注册渠道编号 | VARCHAR(128) | 26000000 |
| register\_city | 注册城市 | VARCHAR(128) | 南京 |
| register\_time | 注册时间 | VARCHAR(30) | 2016-08-05 01:30:44 |
| first\_usage\_time | 首次使用时间 | VARCHAR(30) | 2016-08-05 01:30:44 |
| last\_usage\_time | 最近使用时间 | VARCHAR(30) | 2016-08-05 01:30:44 |
| usage\_cnt | 使用次数 | BIGINT | 0、1等大于等于0的数字 |
| etl\_time | ETL时间 | VARCHAR(30) | 2017-08-02 05:17:16 |
| pt\_d | 天分区 | VARCHAR(8) | 20170801 |
| pt\_service | 业务分区 | VARCHAR(128) | vmall |

1. **数据来源**
2. **初始化**

初始化数据均从C30中获得，初始化时间为20161231

|  |  |  |
| --- | --- | --- |
| 初始化表名 | 针对业务 | 备注说明 |
| dim\_user\_service\_flag\_ds | 所有注册业务 | 最开始初始化的表，该表只有up对应注册渠道的信息 |
| dw\_content\_user\_login\_fact\_snap\_dm | 内容 | 所有内容业务相关的表，与kpi的差别是，能区分开youkuvideo和sohuvideo，并且hwread和hnread两业务的数据更准确一些 |
| dw\_cloudservice\_kpi2\_user\_info\_dm | kpi相关 | kpi业务相关的表，非kpi业务的数据在up\_account、up\_pc和ALL中统一存放，因此无法初始化到这些业务 |
| dw\_cloudservice\_kpi2\_user\_login\_fact\_dm | 非kpi活跃用户 | 该表是日志表，从C30同步了20161231之前所有分区的数据，用来初始化非kpi的业务 |
| dw\_honor\_user\_info\_dm | fans和vmall | fans和vmall两业务，因画像特殊需要，需从该表获取last\_usage\_time |

1. **正常任务**

该表正常的任务从20170101开始跑批

|  |  |  |
| --- | --- | --- |
| 来源表英文名 | 来源表中文名 | 备注说明 |
| dwd\_evt\_online\_game\_buoy\_user\_access\_log\_dm | 网游浮标用户访问日志 | 只影响gameonline和gameoffline分区，在kpi对应的是gameservice |
| dwd\_evt\_online\_game\_buoy\_user\_login\_log\_dm | 网游浮标用户登录日志 | 只影响gameonline和gameoffline分区，在kpi对应的是gameservice |
| dwd\_onl\_disting\_ver\_app\_ds | 区分版本应用 | 用来区分gameonline和gameoffline两个分区 |
| dwd\_onl\_app\_categ\_ds\_his | 应用类别历史 | 用来区分gameonline和gameoffline两个分区 |
| dwd\_evt\_up\_oper\_log\_dm | 华为帐号操作日志 | 每日活跃帐号的日志表 |
| dwd\_cust\_mapping | 消费者编号映射关系表 | 用来对加密处理的up\_id进行解密，判断自注册标志 |
| dwd\_ref\_channel\_service\_rela\_par\_ds | 渠道服务关系参数表 | 该表通过渠道号能关联出唯一的包名 |
| dwd\_evt\_hwmovie\_interface\_api\_log\_dm | 华为视频接口调用日志 | 优酷视频业务，从20170223开始增加该数据源 |
| dwd\_up\_user\_info\_ds | 用户信息表 | 帐号注册的全量信息表 |
| dwd\_ref\_cloudservice\_package\_par\_ds | 云服务包参数表 | 该表通过包名能关联出唯一的业务 |
| dwd\_pty\_up\_ds\_his | 华为帐号历史 | 关联该表获取帐号的公共维度 |

1. **数据逻辑**
2. **主数据逻辑**

该表的主数据主要来自四部分：

1. 网游

来源表：dwd\_evt\_online\_game\_buoy\_user\_access\_log\_dm

dwd\_evt\_online\_game\_buoy\_user\_login\_log\_dm

业务分区：dwd\_onl\_app\_categ\_ds\_his表中app\_categ\_cd为001的app\_id，为网游的app\_id。

在dwd\_onl\_disting\_ver\_app\_ds中，能关联上上面app\_id的package\_name，为网游的包名，关联不上的都不是网游的包名。

来源表通过包名，最终确定业务分区的取值，网游的是gameonline，非网游的是gameoffline。

自注册标志：均赋值为0（否）

1. up操作日志

来源表：dwd\_evt\_up\_oper\_log\_dm

业务分区：直接取up操作日志的pt\_service

自注册标志：当包名为com.huawei.hwid和游戏的包名时，满足up\_id/10000000000000000=channel\_id/1000000的，赋值为1，不满足的赋值为0。

当包名为其他值时，用package\_name关联渠道包名关系表，取到的min(channel\_id)满足channel\_id/1000000 = up\_id/10000000000000000的，赋值为1，不满足的为0。

注意事项：20170706之后，修改up操作日志的取数范围，只取up\_oper\_type\_cd为1和2的。

1. 用户信息表

来源表：dwd\_up\_user\_info\_ds

渠道号：优先取该表的register\_channel\_id，注册渠道为空的，取up\_id/10000000000000000后拼000000。

业务分区：通过上面的渠道号，关联取出pt\_service，关联不上的赋值为others。

自注册标志：注册渠道为空，赋值为1，不为空的，满足up\_id/10000000000000000=register\_channel\_id/1000000时，赋值为1，其他均为0。

1. 优酷视频

来源表：dwd\_evt\_hwmovie\_interface\_api\_log\_dm

渠道号：直接赋值38000001

包名：直接赋值com.huawei.hwvplayer.youku

业务分区：直接赋值youkuvideo

自注册标志：满足up\_id/10000000000000000 = channel\_id/1000000的，赋值为1，其他的为0

注意事项：该数据源从20170223开始添加。

1. **存储逻辑**

该表为累全量表，每天的数据和上一天的数据union all来获取全量。

register\_channel\_id、up\_type\_cd、register\_city和register\_time四个字段为帐号公共维度，累全量之后关联帐号历史表来获取。

1. **特殊逻辑**

该表设立summary分区，用来统计每个帐号的总的活跃度。

在该分区中，渠道号、包名以及自注册标志均赋值为null。

service\_set字段把每个帐号对应的所有业务用#拼接起来。

该分区的最细粒度就是up\_id。

1. **注意事项**
2. **使用方法**
3. 当日活跃

pt\_d=当天

pt\_service=所取业务

last\_usage\_time=当天

1. 累计活跃

pt\_d=当天

pt\_service=所取业务

last\_usage\_time<=当天

1. 当日新增

pt\_d=当天

pt\_service=所取业务

group by up\_id，min(first\_usage\_time)=当天

1. 新增注册

gameonline、gameoffline、hwread和hnread：

pt\_d=当天

pt\_service=所取业务

group by up\_id，min(first\_usage\_time)=当天

其他业务：

pt\_d=当天

self\_register\_flg=1

以up\_id分组，取first\_usage\_time最小的那一条，

pt\_service=所取业务

first\_usage\_time=当天

1. 累计注册

gameonline、gameoffline、hwread和hnread：

pt\_d=当天

pt\_service=所取业务

group by up\_id，min(first\_usage\_time)<=当天

其他业务：

pt\_d=当天

self\_register\_flg=1

以up\_id分组，取first\_usage\_time最小的那一条，

pt\_service=所取业务

first\_usage\_time<=当天

1. **差异原因**

目前该表与C30中kpi的差异主要受特殊渠道号的影响。

C30：完全是以渠道号为准来匹配业务。

Onedata：当渠道号为7000000、-1、90000100和90000200时，以包名为准来匹配业务，其余的均以渠道号为准来匹配业务。

经和业务沟通确认，Onedata的更为准确，以Onedata的口径为准。