**BDREPORTER应用操作信息上报**

1. **数据字典**

**1.1表级**

英文名：dwd\_evt\_bdreporter\_app\_oper\_info\_report\_dm

中文名：BDREPORTER应用操作信息上报

表类型：快照表

ETL调度频率：日表

**1.2字段级**

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 字段类型 | 字段中文名 | 备注 |
| unique\_id | STRING | 唯一编号 |  |
| imei | STRING | 设备编号 |  |
| rec\_time | STRING | 记录时间 |  |
| package\_name | STRING | 包名 |  |
| report\_evt\_id | STRING | 上报事件编号 |  |
| report\_evt\_content | STRING | 上报事件内容 |  |
| app\_ver | STRING | 应用版本 |  |
| etl\_time | VARCHAR(30) | ETL时间 |  |

1. **数据来源**

**2.1．来源于数据打点源表:**

数据打点源表 : ODS\_BDREPORTER\_HI\_ANALYTICS\_HM (采集源表)

**2.2．其他依赖的表:**

dwd\_ref\_cloudservice\_package\_par\_ds (包名参考表):存储包名和业务分区的对应关系

1. **数据逻辑**

**3.1．主体逻辑**

**App打点上报的数据 通过采集服务器收集，收集之后的数据推送到ODS,并通过jar包解析成4张小时表，此表是汇总小时表一天的数据**

**3.2．字段逻辑**

| 字段名 | 处理逻辑 |
| --- | --- |
| unique\_id | 直抽，数据平台添加，拉通作为华为设备唯一标识符 |
| imei | 直抽，设备id,做了格式化。没有做非法imei。如果设备缺失unique\_id可将imei填充在uuid位置 |
| rec\_time | 采集时间 精确到毫秒。数据为时间戳格式，通过dateutil函数经行转换 |
| package\_name | 直抽，包名 |
| report\_evt\_id | 直抽，某个功能的编号，用于记录事件id，统计功能点使用率，重用率，回访率。这个也作为表的业务分区字段 |
| report\_evt\_content | 直抽，用户事件的内容，记录事件值详情，统计与分析用户行为，为大字段 |
| app\_ver | 直抽，APK的版本 |
| etl\_time | ETL时间字段。记录数据入库的服务器时间。 |

**3.3．ETL逻辑**

基本是直抽的逻辑。只有个别字段经行数据标准化的处理。

1. **其他**

**4.1． imei处理过程**

4.1.1标准化：revertDeviceId(deviceIdFormat(imei))

**4.2 性能优化**

**DISTRIBUTE BY T1.unique\_id,T1.pt\_evt; 分布式处理数据，合理分配到reduce任务。减少磁盘的IO**