河源数据中心与网络侧接口规范

版本号：1.0.0



**卓望数码技术（深圳）有限公司 版权所有**

**2016年5月**

**文档变更记录**

| 序号 | 变更内容说明 | 版本号 | 版本日期 | 变更人 |
| --- | --- | --- | --- | --- |
|  | 完成初稿 | 1.0.0A | 2016/05/18 | 龚剑辉 |
|  | 1、接口单元“用户终端数据(月)”增加属性“终端价格”、“上次换机时间”、“换机周期”、“换机次数”，增加对层次类属性描述  2、接口单元“用户上网数据(月)”增加属性“上网时长”、“上网次数”、“日均WIFI流量”，增加对层次类属性描述 | 1.0.0B | 2016/05/24 | 龚剑辉 |
|  | 1、修改“接口文件格式”小节字段分隔符为竖线0X7C  2、作废接口单元“用户终端数据(月)”、“用户上网数据(月)”  3、增加接口单元“用户标签数据(月)”  4、修改接口单元“维表数据(月)”的“维表编码”为“维表名称”并增加“行业”、“地域分布”维表  5、修改接口单元“维表数据(月)”，删除“TOPn” | 1.0.0C | 2016/06/03 | 龚剑辉 |
|  | 1、接口单元“用户标签数据(月)”增加“分公司归属(位置)”，”删除“工作地1”、“工作地1”、“工作地1”  2、接口单元“维表数据(月)”的“维表名称”重新初始化 | 1.0.0D | 2016/06/03 | 龚剑辉 |
|  | 1、删除“维表数据(月)”接口单元的之外的所有接口定义  2、把网络侧用户标签数据分开为三个接口单元：用户社会属性(月)、用户终端数据(月)、用户上网数据(月) | 1.0.1A | 2016/08/17 | 龚剑辉 |
|  |  |  |  |  |
|  |  |  |  |  |

目 录

[1 范围 1](#_Toc459222152)

[2 术语、定义和缩略语 1](#_Toc459222153)

[2.1 术语、定义 1](#_Toc459222154)

[2.2 缩略语 1](#_Toc459222155)

[2.3 约定 1](#_Toc459222156)

[3 接口总则 2](#_Toc459222157)

[3.1 总体说明 2](#_Toc459222158)

[3.2 接口实现方式 2](#_Toc459222159)

[3.3 接口传输说明 2](#_Toc459222160)

[3.3.1 传输方式 2](#_Toc459222161)

[3.3.2 传输过程 3](#_Toc459222162)

[3.3.3 接口处理模式 3](#_Toc459222163)

[3.4 接口及时性要求 3](#_Toc459222164)

[3.5 接口文件说明 3](#_Toc459222165)

[3.5.1 接口文件格式 3](#_Toc459222166)

[3.5.2 接口数据要求 4](#_Toc459222167)

[3.5.3 接口文件路径 5](#_Toc459222168)

[3.5.4 数据文件命名规则 5](#_Toc459222169)

[3.5.5 校验文件命名规则 5](#_Toc459222170)

[3.5.6 接口数据校验规则 5](#_Toc459222171)

[3.5.7 校验文件数据要求 6](#_Toc459222172)

[3.6 接口双方责任 7](#_Toc459222173)

[3.6.1 数据提供方 7](#_Toc459222174)

[3.6.2 数据接收方 7](#_Toc459222175)

[4 接口单元 8](#_Toc459222176)

[4.1 用户社会属性(月) 8](#_Toc459222177)

[4.2 用户终端数据(月) 9](#_Toc459222178)

[4.3 用户上网数据(月) 11](#_Toc459222179)

[4.4 维表数据(月) 13](#_Toc459222180)

# 范围

本规范对河源移动数据中心系统与网络侧提供的数据接口进行约定，为双方的研发、维护提供指导。

# 术语、定义和缩略语

## 术语、定义

|  |  |
| --- | --- |
| 术语/定义 | 解释 |
|  |  |
|  |  |

## 缩略语

|  |  |  |
| --- | --- | --- |
| 缩略语 | 英文全称 | 中文含义 |
| WWW | World Wide Web | 万维网 |
| HTTP | HyperText Transfer Protocol | 超文本传输协议 |
| FTP | File Transfer Protocol | 文件传送协议 |
| SFTP | Secure File Transfer Protocol | 安全文件传送协议 |
| BOSS | Business Operation Support System | 业务运营支撑系统 |

## 约定

|  |  |
| --- | --- |
| 名词 | 说明 |
| 数据提供方 | 指数据单元的提供方 |
| 数据接收方 | 指数据单元的消费方； |
| 预留字段 | 指在接口中预留位置的字段，该字段暂时置空（不填写任何值）。 |
| 接口文件 | 指数据提供方根据规范要求生成供数据支撑系统采集的数据，包括接口数据文件、校验文件、接口校验报告。 |
| 接口数据文件 | 指数据提供方根据本规范要求，在指定的时间内，完成对数据的周期性抽取形成的数据文件。 | |
| 接口校验文件 | 指由数据提供方在每一个抽取周期中，针对每个接口单元的接口数据文件生成的校验依据信息。 | |
| 接口校验报告 | 指数据接收方根据校验规则生成的校验文件。 | |
| 每日增量 | 指抽取每天0时0分0秒至23时59分59秒发生变化的、新产生的数据。 | |
| 每月增量 | 指抽取每月1日0时0分0秒至月末最后一天23时59分59秒止，发生变化的、新产生的数据。 | |
| 每日全量 | 指抽取每日24：00时，所有数据的最新状态快照。 | |
| 每月全量 | 指抽取每月月末最后一天24：00时，所有数据的最新状态快照。 |

# 接口总则

## 总体说明

数据提供方周期性预先生成接口文件，上传到约定的接口机器；由数据接收方到约定的接口机、按照约定协议，到约定的接口目录获取。

## 接口实现方式

数据提供方与数据接收方的明细数据传送通过文件方式实现。一个接口的内容可以用一个或多个文件传输，但同一个文件不能传输两个或两个以上接口的内容。

## 接口传输说明

### 传输方式

数据提供方按约定将文件存放到本系统内指定位置或上传到远程指定位置，数据接收方到指定的位置获取数据，具体要求包括：

1. 支持将数据直接保存到本系统内本地或远程目录下；
2. 支持网络模式传输，可以实现跨越局域网络平台的文件传送；
3. 支持文件重传功能。

### 传输过程

数据传输应支持:

1. 实时、高效和安全可靠地传送数据；
2. 传输过程中的差错控制；
3. 在文件传输过程中，使用tmp.前缀后接所传输文件名作为临时文件名，传输完成后临时文件名改为正式文件名

### 接口处理模式

数据提供方根据不同接口的要求，遵循统一的命名规则生成接口数据文件，通过SFTP/FTP协议传输到指定数据接口机目录下。数据接收方到该目录下提取接口文件。

## 接口及时性要求

1. 按日的接口，保证按本接口规范定义的时间点前，在数据接口机生成完前一天的增量或全量数据；
2. 按月的接口，保证按本接口规范定义的时间点前，在数据接口机生成完前一个月的增量或全量数据；
3. 按小时的接口，保证按本接口规范定义的时间点前，在数据接口机生成完前一小时的增量或全量数据。

## 接口文件说明

### 接口文件格式

本规范中如无特殊要求，需遵循此章节定义的接口文件格式：

1. 记录间分隔符：回车换行符（0x0D0A）；
2. 字段间分隔符：竖线分隔符“|”(0X7C)。

为了保证数据的准确性以及接口文件中的记录各值域在有效的取值范围内，数据中均不能包含0x0D0A（回车换行符）和字段间分隔符。

### 接口数据要求

数据提供方在生成接口文件时，必须遵守如下数据转换规则：

1. 编码格式
2. 汉字：GBK内码
3. 西文：ASCII码
4. 数字格式
5. 在接口文件中，数字的表示必须规范，小数点的前后必须有数字，如：0.01或34.0,不能用“.01”或“34.”表示；
6. 符号处理：数字最高位的左边第一位为符号位。对于负数，符号位为“-”，正数不用加符号位。
7. 日期类型
8. 日期类型统一采用YYYYMMDD格式，不允许出现空值，且YYYYMMDD必须为有意义的日期：
   * + 1. YYYY为四位数字，必须是有效的年份
       2. MM为两位数字，必须是有效的月份（01-12）
       3. DD为两位数字，必须是有效的日期（01-31）
9. 对于不符合日期约束规则的日期值，处理方式存在以下两种情况：
10. 无值的日期或者无意义的日期，这时在接口中一律以“00010101”（公元元年1月1日）填充；
11. 接口中的“失效日期”在表示“未失效”含义时，一律以“29991231”（公元2999年12月31日）填充。
12. 时间类型
13. 统一采用HHMMSS格式：
14. HH为两位数字，必须是有效的小时（00-23），24小时制；
15. MM为两位数字，必须是有效的分钟（00-59）；
16. SS为两位数字，必须是有效的秒（00-59）。

### 接口文件路径

上线时双方约定

### 数据文件命名规则

<文件类别代码><数据来源编码>\_<接口单元编码>\_<数据日期>\_<序列号>.dat

说明：

1. 该文件用于数据提供方向数据接收方提供本接口规范要求的数据；
2. 文件为获取全量数据文件时，<文件类别代码>为“a”；
3. 文件为增量文件时，<文件类别代码>为“i”；
4. 数据来源编码一般定义为一位字母或数字；比如字母“b”标识BOSS，“m”标识数据集市，“n”标识网络侧数据接口
5. <接口单元编码>参考各章节文件接口中接口编码验收项；

接口单元编号：参见接口单元约定

1. <数据日期>按不同的接口类型确定，如为小时接口，取值为“YYYYMMDDHH24”；若为日接口则为“YYYYMMDD”；若为月接口则为“YYYYMM”；若为周接口则为周统计周期的第一天“YYYYMMDD”；
2. <序列号>默认为三位，为数字型, 保证接口数据文件的大小不能超过500M，如果接口数据文件太大，必须按要求将文件拆分。序列号是用于描述同一个接口单元在同一个抽取周期中的文件顺序号，若一个接口单元被分割成多个文件则其流水号必须不同，并且从001依次递增，如： 001、002、003；
3. 在接口文件生成过程中，定义接口文件扩展名为.tmp，待接口文件生成成功后，再改为.dat，以保证数据文件的完整性。

### 校验文件命名规则

<文件类别代码><数据来源编码>\_<接口单元编码>\_<数据日期>.verf

参考数据文件命名规则；校验文件与数据文件命名规则一直，与数据文件名对应。

### 接口数据校验规则

数据接收方在发现数据错误时，必须及时提交数据校验报告给数据提供方；数据提供方解决问题后重新提供数据并知会数据接收方。接口数据校验分为两种：文件级校验和记录级校验。

1. 文件级校验

文件级校验是指根据数据提供方提供的接口校验文件信息，对需要传输的接口文件进行校验，校验内容如下：

1. 接口数据文件名称

a)根据接口校验文件中记录的文件，查找接口数据文件目录下该接口数据文件是否存在。

b)检查接口数据文件的名称是否遵守规范中的命名规范。

1. 文件的大小（字节数）

验证接口校验文件中记载的文件大小与实际接口数据文件的文件大小是否一致。

1. 文件中包含的记录数

验证接口校验文件中记载的文件记录数与实际接口数据文件中的记录行数是否一致。

1. 记录级校验
2. 数据域个数校验

检查接口数据文件中的每个记录行所包含的字段个数是否与接口定义一致

1. 唯一性检查

检查接口数据文件中是否违反本接口规范定义的唯一性约束，如主键检查；

1. 数据类型及格式检查

字段类型检查，检查各字段类型是否与本接口规范定义一致；格式检查参考【接口数据要求】章节；

1. 数据值域检查

对数据的取值有效性检查。

### 校验文件数据要求

校验文件是由数据提供方负责生成，用于描述本次传输的所有接口文件需要校验的动态规则信息。校验文件中的信息包括以下内容，且信息内容字段用竖线“|”(ASCII码0x7C)：

|  |  |  |  |
| --- | --- | --- | --- |
| 顺序 | 信息内容 | 数据类型及长度 | 说明 |
| 1 | 接口数据文件名称 | VARCHAR2(50) |  |
| 2 | 文件的大小（字节数） | NUMBER(20) | 文件的物理存储大小 |
| 3 | 文件中包含的记录数 | NUMBER(20) |  |
| 4 | 文件时间 | VARCHAR2(14) | 格式：YYYYMMDDHHMISS |
| 5 | 0x0D0A |  | 行间分隔符－回车换行符 |

## 接口双方责任

### 数据提供方

1. 保证在指定的时间范围内生成本接口规范规定的数据内容并上传到指定接口机位置；
2. 遵循本接口规范中规定的校验规则和异常处理规则，保证提供数据质量，确保数据的准确性、一致性、完整性；
3. 负责数据的业务逻辑一致性控制，保证不将逻辑错误的数据提供给数据接收方，并保证生成的接口数据与数据提供方存储的数据一致；
4. 若接口有变更，要求提前2周提供变更内容说明，并以书面形式告知数据接收方；
5. 提供接口元数据信息，包括接口数据结构、实体间依赖关系、计算关系、关联关系及接口数据传输过程中的各类管理规则等信息；
6. 保证接口数据文件的大小不能超过500M，如果接口数据文件太大，必须按要求将文件拆分，拆分后的文件命名参考文件命名规范；
7. 对于本规范中未尽事宜，数据提供方应协助数据接收方协商解决相关问题。

### 数据接收方

1. 按时接收本接口规范规定的接口文件，如有特殊情况，需提前沟通数据提供方通过其它方式进行接口文件传输；
2. 遵循本接口规范中规定的验证规则，对接收的数据进行验证，保证在数据接收完后，对数据进行校验并生成报告文件；
3. 数据接收方的接口机由数据接收方进行维护；
4. 数据接收方的接口机有变更，需要提前2周以书面形式通知数据提供方；
5. 负责对数据提供方提供的接口文件进行及时的读取、接收和校验；并把校验异常信息通知数据提供方
6. 对于本规范中未尽事宜，数据接收方应协助数据提供方协商解决相关问题。

# 接口单元

## 一点接入结算主表(月)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **接口名称** | 一点接入结算主表 | | | | | |
| **接口编码** |  | | | | | |
| **接口说明** |  | | | | | |
| **数据源系统** |  | **源数据表/视图** | |  | | |
| **抽取方式及周期** |  | | | | | |
| **接口文件命名**（日全量） | **接口数据文件名** | |  | | | |
| **校验文件名** | |  | | | |
| **属性编码** | **属性名称** | | **属性类型** | | **属性描述** | **备注** |
| 01 | 月份 | |  | | 格式:YYYYMM |  |
| 02 | 咪咕子公司 | |  | | 固定值：咪咕视讯 |  |
| 03 | 合作伙伴公司名称 | |  | | 渠道ID app\_channel\_id跟合作公司维表 |  |
| 04 | 首次激活用户数（核减后） | |  | | 异常行为核减规则：凌晨2:00~5:00的新增用户 |  |
| 05 | 次月活跃用户数 | |  | | 上月新增（核减后）的用户在本月有活跃的用户数 |  |
| 06 | 首次激活单价 | |  | | 留空 |  |
| 07 | 次月活跃单价 | |  | | 留空 |  |
| 08 | 合计金额 | |  | | 留空 |  |
| 09 | 税率 | |  | | 留空 |  |

## 一点接入结算明细(月)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **接口名称** | 一点接入结算明细 | | | | | |
| **接口编码** |  | | | | | |
| **接口说明** |  | | | | | |
| **数据源系统** |  | **源数据表/视图** | |  | | |
| **抽取方式及周期** |  | | | | | |
| **接口文件命名**（日全量） | **接口数据文件名** | |  | | | |
| **校验文件名** | |  | | | |
| **属性编码** | **属性名称** | | **属性类型** | | **属性描述** | **备注** |
| 01 | 月份 | |  | | 格式:YYYYMM |  |
| 02 | 咪咕子公司 | |  | | 固定值：咪咕视讯 |  |
| 03 | 合作伙伴公司名称 | |  | |  |  |
| 04 | 子渠道名称 | |  | |  |  |
| 05 | 子渠道ID | |  | |  |  |
| 06 | 首次激活用户数（核减后） | |  | | 异常行为核减规则：凌晨2:00~5:00的新增用户 | 新增设备用户 |
| 07 | 次月活跃用户数 | |  | | 上月新增（核减后）的用户在本月有活跃的用户数 |  |
| 08 | 首次激活单价 | |  | | 留空 |  |
| 09 | 次月活跃单价 | |  | | 留空 |  |
| 10 | 合计金额 | |  | | 留空 |  |
| 11 | 税率 | |  | | 留空 |  |

## 一点接入有效激活清单(月)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **接口名称** | 一点接入有效激活清单 | | | | | |
| **接口编码** |  | | | | | |
| **接口说明** | 【】 | | | | | |
| **数据源系统** |  | **源数据表/视图** | |  | | |
| **抽取方式及周期** |  | | | | | |
| **接口文件命名**（日全量） | **接口数据文件名** | |  | | | |
| **校验文件名** | |  | | | |
| **属性编码** | **属性名称** | | **属性类型** | | **属性描述** | **备注** |
| 01 | 月份 | |  | | 格式:YYYYMM |  |
| 02 | 咪咕子公司 | |  | | 固定值：咪咕视讯 |  |
| 03 | 合作伙伴公司名称 | |  | |  |  |
| 04 | 子渠道名称 | |  | |  |  |
| 05 | 子渠道ID | |  | |  |  |
| 04 | 激活时间 | | Become\_new\_device\_unix\_time | | 精确到天还是时分秒？  --到秒，格式：YYYYMMDDHHMISS |  |
| 05 | 产品名称 | |  | | 枚举值：  咪咕直播、咪咕视频、咪咕影院 | 根据product\_key获取 |
| 06 | IMEI | |  | |  | 根据device\_key反查 |
| 07 | IMSI | |  | | 可为空 | 根据device\_key反查 |
| 08 | 是否为无账号用户 | |  | | Y：无账号用户  N：有账号用户 | 根据device\_key反查 |
| 09 | 是否7日留存用户 | |  | | Y：是  N：否  当某天新增用户，在7天后仍然活跃的用户 | 取第七天后的数据关联 |
| 10 | 是否存在异常行为 | |  | | 是否为2:00~5:00之间激活的用户  Y：是  N：否 | 根据激活时间进行判断 |
| 11 | 是否次月留存用户 | |  | | 【次月留存：是指上月新增的用户在当月仍然活跃的用户？】否；指当前月新增用户，在此月仍然留存的用户  无法计算，留空  是：上月新增的用户在当前月仍然活跃的用户 |  |
| 12 | 是否当月使用用户 | |  | | 使用<25下载流量>+<26播放流量>数据源统计，【但是在这两个数据源与sessioninfo中不包含首次使用的渠道、IMSI号】  使用22播放url获取时长作为使用用户数据来源 | 待确定 |
|  |  | |  | |  |  |

## 一点接入日常查询(日)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **接口名称** | 一点接入日常查询 | | | | | |
| **接口编码** |  | | | | | |
| **接口说明** |  | | | | | |
| **数据源系统** |  | **源数据表/视图** | |  | | |
| **抽取方式及周期** |  | | | | | |
| **接口文件命名**（日全量） | **接口数据文件名** | |  | | | |
| **校验文件名** | |  | | | |
| **属性编码** | **属性名称** | | **属性类型** | | **属性描述** | **备注** |
| 01 | 日期 | |  | | 格式:YYYYMMDD |  |
| 02 | 咪咕子公司 | |  | | 固定值：咪咕视讯 |  |
| 03 | 产品名称 | |  | |  |  |
| 04 | 合作伙伴公司名称 | |  | |  |  |
| 05 | 渠道ID | |  | | 合作公司ID，如果没有则留空 |  |
| 06 | 子渠道名称 | |  | |  |  |
| 07 | 子渠道ID | |  | |  |  |
| 08 | 活跃用户数 | |  | | 符合基础激活条件用户数（启动大于等于1次，imei、imsi符合条件) |  |
| 09 | 首次激活用户数 | |  | | imei、imsi符合条件  【是否需要核减？】  无需核减 |  |
| 10 | 无账号用户数 | |  | | 新用户中的无账号用户 |  |
| 11 | 异常行为用户数 | |  | | 新用户中，2:00~5：之间新增的用户 |  |
| 12 | 上月留存用户数 | |  | | 上月新增用户在本月有活跃的，无需核减 |  |
| 13 | 使用用户数 | |  | |  |  |
| 14 | 第七天留存用户数 | |  | | 当前数据周期新增用户，在第7天活跃的用户数 |  |

## 一点接入日常查询(月)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **接口名称** | 一点接入日常查询 | | | | | |
| **接口编码** |  | | | | | |
| **接口说明** |  | | | | | |
| **数据源系统** |  | **源数据表/视图** | |  | | |
| **抽取方式及周期** |  | | | | | |
| **接口文件命名**（日全量） | **接口数据文件名** | |  | | | |
| **校验文件名** | |  | | | |
| **属性编码** | **属性名称** | | **属性类型** | | **属性描述** | **备注** |
| 01 | 日期 | |  | | 格式:YYYYMMDD |  |
| 02 | 咪咕子公司 | |  | | 固定值：咪咕视讯 |  |
| 03 | 产品名称 | |  | |  |  |
| 04 | 合作伙伴公司名称 | |  | |  |  |
| 05 | 渠道ID | |  | |  |  |
| 06 | 子渠道名称 | |  | |  |  |
| 07 | 子渠道ID | |  | |  |  |
| 08 | 首次激活用户数 | |  | | 启动大于等于1次，imei、imsi符合条件  【是否笔误？应为活跃用户数？】 |  |
| 09 | 上月留存用户数 | |  | | 上月新增用户在当前月份仍然活跃的用户数 |  |
| 10 | 上月留存率 | |  | | 上月留存用户数/上月首次激活用户数 |  |
| 11 | 首次激活用户数 | |  | | imei、imsi符合条件  【是否需要核减？】  无需核减 |  |
| 12 | 使用用户数 | |  | |  |  |
| 13 | 无账号用户数 | |  | |  |  |
| 14 | 无账号用户数比例 | |  | | 【无账号用户数/首次激活用户数？】 |  |
| 15 | 异常行为用户数 | |  | | 异常行为用户：2:00~5:00首次激活用户 |  |
| 16 | 异常行为用户数比例 | |  | | 【异常行为用户数/首次激活用户数】 |  |
| 17 | 第七天留存用户数 | |  | | 当前数据周期新增用户，在第7天活跃的用户数 |  |
| 18 | 第七天留存率 | |  | | 第七天留存用户数/当前周期的新增用户数 |  |