# 1、数据输入和输出

数据源：MR数据来源大概可以分为3类：客户hadoop平台取数据存为csv文件，客户提供的CSV文件，客户提供的csv.gz文件

数据源按邻区排列方式，可以分为2类：其一邻区按行排，1条MR有多行记录，n个邻区就有n条记录，其二邻区按列排，1条MR只有1行记录，这1行记录有n1~n8条邻区。根据数据源的邻区排列方式不同，需要不同的解析方式来生成输出给定位格式的MRO。

输出数据格式：

解析结果需要按下表顺序输出下面的字段，作为给定位MR的输入。下表邻区是按行排列的。

最后输出的数据是1个小时、1个站点(enbid)作为1文件。

|  |  |
| --- | --- |
| **列头** | **说明** |
| v\_fposition\_id | NULL(空值字段必须用大写"NULL"填充,其他字段也是一样) |
| object\_fposition\_id | NULL |
| callID | NULL |
| MISI | NULL |
| mroid | MRO采样点编号,必须计算（mroid mroid组成：14位到秒的日期不足前面补0，3位到毫秒的时间戳不足在前面补0 + 9位ECI，不足9位在前面补0 + 10位mmeues1apid，不足10位在前面补0。如：201702201400051202194962130102219438， 20170220140005120（时间戳）|219496213（ECI）|0102219438（mmeues1apid，补了1位0）） |
| eNB\_Id | 服务小区站号 |
| object\_Id | 服务小区ECI |
| n\_cellId | 邻区ECI |
| object\_Id\_Detial | 服务小区ECI |
| object\_MmeUeS1apId | MME UE S1应用程序标识 |
| object\_MmeGroupId | MME组标识 |
| object\_MmeCode | MME码 |
| object\_TimeStamp | 数据测量对象UE样本测量时间 |
| object\_msec | 时间的毫秒数 |
| MR\_LteScRSRP | 服务小区的参考信号接收功率 |
| MR\_LteScRSRQ | 服务小区的参考信号接收质量 |
| MR\_LteScTadv | 服务小区的时间提前量 |
| MR\_LteScPHR | 服务小区的UE发射功率余量 |
| MR\_LteScAOA | 服务小区的eNB天线到达角 |
| MR\_LteScSinrUL | 服务小区的上行信噪比 |
| MR\_LteScEarfcn | 服务小区载波号 |
| MR\_LteScPci | 服务小区的物理小区识别码 |
| MR\_LteNcRSRP | 邻区的参考信号接收功率(空值字段必须用大写"NULL"填充) |
| MR\_LteNcRSRQ | 邻区参考信号接收质量(空值字段必须用大写"NULL"填充) |
| MR\_LteNcEarfcn | 邻区载波号(空值字段必须用大写"NULL"填充) |
| MR\_LteNcPci | 邻区物理小区识别码(空值字段必须用大写"NULL"填充) |

# 2、源数据解析

## 2.1、MR邻区按列排

**Step1、**源数据生成Mr\_format格式文件，格式如下表：

|  |  |  |  |
| --- | --- | --- | --- |
| 字段顺序 | 输入字段 | 返回值 | 描述 |
| 0 | mr\_date | 2017-04-22 | 日期格式，精度到天 |
| 1 | city\_id | 13012 | 城市代码/区号，如无锡市代码为13012 |
| 2 | vendor\_id | 5 | 设备厂家 |
| 3 | mr\_hour | 10 | 业务小时，上界，如20点的业务，小时取到20 |
| 4 | mr\_minute | 00 | 业务时间的分钟 |
| 5 | mme\_code | 90 | mme代码 |
| 6 | mme\_groupid | 27139 | Mme组号 |
| 7 | mmeues1apid | 113705852 | mme业务号 |
| 8 | time\_s | 2017-04-22 10:00:05 | 日期格式，精度到秒 |
| 9 | time\_ms | 120 | 时间戳的毫秒 |
| 10 | sc\_eci | 219325471 | 服务小区ECI，ECI=sc\_enbid\*256+sc\_lcrid.其他厂家可能带有频点、子帧号 |
| 11 | sc\_enbid | OMC1.856740 | 服务小区enbid，ECI=sc\_enbid\*256+sc\_lcrid。扇区号,sc\_lcrid=mod(sc\_eci,256) |
| 12 | lte\_sc\_rsrp | 30.0 |  |
| 13 | lte\_sc\_rsrq | 20.0 |  |
| 14 | lte\_sc\_tadv | 12.0 |  |
| 15 | lte\_sc\_phr | 24 |  |
| 16 | lte\_sc\_aoa | 1 |  |
| 17 | lte\_sc\_sinrul | 16 |  |
| 18 | lte\_sc\_earfcn | 1650.0 |  |
| 19 | lte\_sc\_pci | 300.0 |  |
| 20 | lte\_sc\_rip | 22 |  |
| 21 | lte\_nc\_rsrp1 | 27.0 |  |
| 22 | lte\_nc\_rsrq1 | 16.0 |  |
| 23 | lte\_ncearfcn1 | 1650.0 |  |
| 24 | lte\_ncpci1 | 110.0 |  |
| 25 | lte\_nc\_rsrp2 | 33 |  |
| 26 | lte\_nc\_rsrq2 | 10 |  |
| 27 | lte\_ncearfcn2 | 1650 |  |
| 28 | lte\_ncpci2 | 57 |  |
| 29 | lte\_nc\_rsrp3 | 33 |  |
| 30 | lte\_nc\_rsrq3 | 26 |  |
| 31 | lte\_ncearfcn3 | 1650 |  |
| 32 | lte\_ncpci3 | 191 |  |
| 33 | lte\_nc\_rsrp4 | 27 |  |
| 34 | lte\_nc\_rsrq4 | 16 |  |
| 35 | lte\_ncearfcn4 | 1650 |  |
| 36 | lte\_ncpci4 | 21 |  |
| 37 | lte\_nc\_rsrp5 | 27 |  |
| 38 | lte\_nc\_rsrq5 | 13 |  |
| 39 | lte\_ncearfcn5 | 1650 |  |
| 40 | lte\_ncpci5 | 266 |  |
| 41 | lte\_nc\_rsrp6 | 23 |  |
| 42 | lte\_nc\_rsrq6 | 6 |  |
| 43 | lte\_ncearfcn6 | 1650 |  |
| 44 | lte\_ncpci6 | 111 |  |
| 45 | lte\_nc\_rsrp7 | 12 |  |
| 46 | lte\_nc\_rsrq7 | 3 |  |
| 47 | lte\_ncearfcn7 | 1650 |  |
| 48 | lte\_ncpci7 | 44 |  |
| 49 | lte\_nc\_rsrp8 | 12 |  |
| 50 | lte\_nc\_rsrq8 | 3 |  |
| 51 | lte\_ncearfcn8 | 1650 |  |
| 52 | lte\_ncpci8 | 44 |  |
| 53 | LONGITUDE | NULL | 江苏提供的mr有这个字段，但是没有用，直接赋值null |
| 54 | LATITUDE | NULL | 江苏提供的mr有这个字段，但是没有用，直接赋值null |
| 55 | lte\_sc\_plrul\_qci1 | 0.0 |  |
| 56 | lte\_sc\_plrul\_qci2 | 0.0 |  |
| 57 | lte\_sc\_plrul\_qci3 | 0.0 |  |
| 58 | lte\_sc\_plrul\_qci4 | 0.0 |  |
| 59 | lte\_sc\_plrul\_qci5 | 0.0 |  |
| 60 | lte\_sc\_plrul\_qci6 | 0.0 |  |
| 61 | lte\_sc\_plrul\_qci7 | 0.0 |  |
| 62 | lte\_sc\_plrul\_qci8 | 0.0 |  |
| 63 | lte\_sc\_plrul\_qci9 | 0.0 |  |
| 64 | lte\_sc\_plrdl\_qci1 | 0.0 |  |
| 65 | lte\_sc\_plrdl\_qci2 | 0.0 |  |
| 66 | lte\_sc\_plrdl\_qci3 | 0.0 |  |
| 67 | lte\_sc\_plrdl\_qci4 | 0.0 |  |
| 68 | lte\_sc\_plrdl\_qci5 | 0.0 |  |
| 69 | lte\_sc\_plrdl\_qci6 | 0.0 |  |
| 70 | lte\_sc\_plrdl\_qci7 | 0.0 |  |
| 71 | lte\_sc\_plrdl\_qci8 | 0.0 |  |
| 72 | lte\_sc\_plrdl\_qci9 | 0.0 |  |
| 73 | ltescpuschprbnum |  | 江苏联通没有这个字段，直接赋值null |
| 74 | ltescpdschprbnum |  | 江苏联通没有这个字段，直接赋值null |
| 75 | ltescbsr |  | 江苏联通没有这个字段，直接赋值null |
| 76 | ltescri1 |  | 江苏联通没有这个字段，直接赋值null |
| 77 | ltescri2 |  | 江苏联通没有这个字段，直接赋值null |
| 78 | ltescri4 |  | 江苏联通没有这个字段，直接赋值null |
| 79 | ltescri8 |  | 江苏联通没有这个字段，直接赋值null |
| 80 | ltescenbrxtxtimediff |  | 江苏联通没有这个字段，直接赋值null |

mr\_format文件代码说明：



**Step2、** 处理mr\_format文件，输出定位格式MR



**Step3、**替换空格

生成的csv文件，mroid字段可能存在空格问题，需要替换为 0

例如：20170225以后的定位mro数据，计算mroid是发现字符串中间有空格如：'20170228103015360219631631 114225386' ，需要把空格替换为'0'

**Step4、**按输出表的第6个字段 enbid 进行排序

**Step5、**排序后的数据，按1个enbid生成1个文件 的方式输出多个文件，这些文件就是给定位用的输入文件。

## 2.2、MR邻区按行排

**Step1、**读取原始数据，按MR输出格式输出对应的字段：



**Step2、**替换空格

生成的csv文件，mroid字段可能存在空格问题，需要替换为 0

例如：20170225以后的定位mro数据，计算mroid是发现字符串中间有空格如：'20170228103015360219631631 114225386' ，需要把空格替换为'0'

**Step3、**按输出表的第6个字段 enbid 进行排序

**Step4、**排序后的数据，按1个enbid生成1个文件 的方式输出多个文件，这些文件就是给定位用的输入文件。