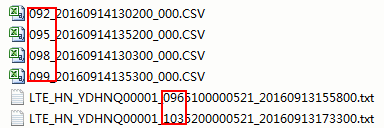
**入库统计**

* **移动文件到相同类型的目录**

用文件名里的字段标识Interface类型



以数字开头的文件，前三个字符为标识；

其它以第19个字符开始的三个字符为标识；

关系如下：

095 UU

099 X2

098 UE\_MR

092 CELL\_MR

096 S1-MME

103 S1-U

流程：

1. 在指定**输出目录**里，以Interface类型建立6个不同的目录（如果存在就不用建立，直接使用），如UU目录等；同时还要建立一个error目录，主要是用来保存不合正则表达式的文件。

2.扫描指定**扫描目录**里所有文件，包括子目录里的所有文件，把相同类型的文件移到同一个目录，如果出现文件名一样的情况，需要把文件名增加后缀进行区分，加如：\_a \_b这样往上升。如092\_20160914103200\_000.CSV移到CELL\_MR目录。

* 解释每个不同类型目录里的所有文件产生有规则的**输出csv文件**

每个类型产生一个输出文件，放在Interface类型目录里（目录里只有一个输出csv文件，如果已经存在这个csv，就用append的方式添加内容），输出文件第一行对应一个源文件，第行不同属性以”|”隔开。

如UU类型的输出文件为 “UU\_处理完时间.csv”

入库的csv或txt文件中，字段是按 | 作为分隔符分隔的。

ID |文件所属目录 | 文件名 | 文件大小 | 文件类型 | 总条目数 | 文件名时间 | 首行时间 |末行时间 |当前文件首行和上一文件末行的时间差

ID：程序自定义一个唯一ID

文件所属目录：源文件所在目录

文件名：源文件的名称，相同文件修改后用修改后的文件名

文件大小：源文件的大小，单位是Byte

文件类型 ：源文件interface类型

总条目数：源文件的行数

文件名时间：源文件名里按指定的正则提取出来的时间，规则见“文件名时间提取正则”

首行时间：源文件第一行的时间戳，时间戳所在的列见下面”时间戳“的解释。

末行时间：源文件最后一行的时间戳，时间戳所在的列见下面”时间戳“的解释。

当前文件首行和上一文件末行的时间差:以**文件名时间**小到大排序后，**首行时间** - 上一个文件的**末行时间**

时间戳——

6个文件区分出来以后，找寻他们分别的时间戳字段位置

UU 第10个字段

X2 第10个字段

UE\_MR 第14个字段

CELL\_MR 第11个字段

S1-MME 第10个字段

S1-U 第20个字段

时间戳字段格式为Unix时间戳格式——1473822945816，要转换成可读的日期-时间格式

文件名时间提取正则--

UU "(\\d+)\_(\\d+)\_(\\d+).(\\w+)")

X2 "(\\d+)\_(\\d+)\_(\\d+).(\\w+)")

UE\_MR "(\\d+)\_(\\d+)\_(\\d+).(\\w+)")

CELL\_MR "(\\d+)\_(\\d+)\_(\\d+).(\\w+)")

S1-MME (\\w+)\_(\\d{3})(\\d+)\_(\\d+).(\\w+)

S1-U (\\w+)\_(\\d{3})(\\d+)\_(\\d+).(\\w+)

以下这个需求还没有做：

输出有三级目录：

一级目录：城市标识目录：如长沙是731，城市标识是在源文件的第二列。

error目录：不符合文件规则的文件放入这里

backup目录：每次程序启动时增加按日期时间的目录，存放源文件备份

二级目录：Interface类型：如UU，主要有6个目录

三级目录：年月日：按源文件里提取的时间建立目录

流程：

扫描文件里的第二行城市列，相同的城市标识的行提取成一个新文件，文件名为”城市标识\_源文件名”，放入”城市/Interface类型/年月日”里，如果源文件里有10个不同的城市标识，那就会生成10个不同的文件，加起来大小和源文件一样。

文件在写入输出目录里加一个lock。

源文件放入backup目录里，路径保持一样。