试题示例

1. 试题概述

本试题为告警数据分析的试题。现网监控工作中告警信息中的告警发生时间的准确性对于监控工作至关重要。现网中因为网元自身时间不准确，传输中断等原因，监控中心经常收到告警发生时间相对与当前时间超前，或者延后的情况。本试题提供了全省一天的告警数据，请大家对于告警发生时间的准确性进行分析。

1. 试题输入
   1. 2019年7月31日全省告警信息

2019年7月31日全省告警列表信息文件，按照每十分钟一个数据文件存储，第一个文件为**201907310010.csv**，最后一个文件为**201908010000.csv**，全天一共144个数据文件。文件的格式为CSV格式，CSV文件第一行为告警列表清单的表头，每个字段都以双引号包含，例如“\*\*\*\*\*”。不同的字段间以逗号分隔。从第二行开始每行表示一条告警，字段也以双引号包含，以逗号分隔。

**注意：**对于一些特殊的字段如果在一对双引号中出现换行，则不认为换行后是一条新的数据。

例如“\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*”

应认为是字段本身包含换行符。

对于内容为空的字段，表示为“”。

告警信息各字段的属性说明参见**附件1**。

1. 算法说明
   1. 告警信息中的字段“START\_TIME”是告警信息自身携带的告警发生的时间。告警信息中的字段“SAVE\_TIME”是故障管理系统中告警入库的时间。正常情况下，两个时间基本能够保持一致，误差可以保持在几秒中以内，即“SAVE\_TIME”应该等于“START\_TIME”，或者“SAVE\_TIME”晚于“START\_TIME”几秒钟。即如果我们将告警数据的**告警时间误差**定义为“SAVE\_TIME” - “START\_TIME”的值，则正常情况下**告警时间误差**应为0或几秒钟；
   2. 当网元的时间不准确并超前于当前时间时**告警时间误差（**“SAVE\_TIME” - “START\_TIME”）的值为负值，可能为负的几分钟，负的几个小时或更长时间。
   3. 当网元的时间不准确并滞后于当前时间时，或者因为阻断或拥塞的问题导致告警无法及时上传时，**告警时间误差**（“SAVE\_TIME” - “START\_TIME”）的值为正值。可能为十几分钟，几个小时或更长的时间。
   4. 如果“SAVE\_TIME”或“START\_TIME”任一字段存在空值，则此条告警数据作废，不纳入统计。
2. 试题输出
3. 统计2019年7月31日全天告警数据中**告警时间误差**（“SAVE\_TIME” - “START\_TIME”）大于等于10分钟的告警的数量。即网元时间滞后现网时间或因故障延迟上传达到或超过10分钟（含10分钟）的告警的条数。输出的结果为一个数字。结果存储在名称为**out1.txt**的文件中；

例如：假设当日告警数据共有六条数据



经计算各条数据的**告警时间误差**如表中标黄的列所示，则按照分钟的告警数量统计结果即为

**3**

1. 对于2019年7月31全天告警数据的**告警时间误差**（“SAVE\_TIME” - “START\_TIME”）进行统计。统计的内容为按照每一分钟对于**告警时间误差**进行汇总统计，按照分钟为间隔，统计**告警时间误差**落在各个分钟内的告警的条数，统计的结果存储在名为**out2.txt**的文件中。每分钟的统计结果存为一行，时间由小到大排序；

例如：假设当日告警数据共有六条数据



经计算各条数据的**告警时间误差**如表中标黄的列所示，则按照分钟的告警数量统计结果即为。



1. **注意：**分钟的分隔方法为**向前看齐**。如00:00:00时间段包含的范围实际上就是00:00:00到00:00:59；00:01:00时间段包含的范围实际上就是00:01:00到00:01:59；-07:34:00时间段包含的范围实际上就是-07:34:00到 -7:33:01；
2. **注意：**如果时间超过24个小时，则时间可以通过小时累加，也可以通过增加天的字段表示。例如 25:01:00 也可以通过增加天表示为 1days 01:01:00；又例如 -31:34:00可以表示为- 1days 07:34:00表示；对于时间为负的情况 也可以通过天为负数，时间为正的方法表示，例如-07:34:00 也可以表示为 -1days +16:26:00的形式表示。**如果还有其他合理的表示分钟级时间间隔的方法请在答案中详细说明**。
3. **注意：**如果“SAVE\_TIME”或“START\_TIME”任一字段存在空值，则此条告警数据作废，不纳入统计。
4. **注意：**在告警清单数据中，性能指标告警和衍生告警不纳入统计。对于这两类告警的判断可以通过“SOURCE\_ID”字段的值进行判断，具体内容可以参考**附件一。**
5. **注意：**在告警清单数据中，只统计能够自动清除的告警。没有配对的清除告警，不能自动清除的告警不纳入统计。对于可以自动清除的告警的判断方法为“CLEAR\_USER”字段的取值为“AUTOMATIC”。对于该字段为其他取值的数据不纳入统计。