# Count the number of delimiters in the input file

Kannan Deivasigamani
HSBC Mortgage Services

**ABSTRACT**

The paper shows how an issue with a delimiter file can be identified systemically and reported for resolution demonstrated with a file example and an accompanying code and the benefits of handling it early in the stages of processing the data.

**INTRODUCTION**

Many of us are faced at times with surprise by the presence (or even absence) of an additional (missing) variable in the incoming file (file from the sender) due to the appearance of an additional delimiter as part of the data. For example, if an address variable requires a parsing and unintentionally, a user inputs a pipe ("|") as part of the address, it can pose a problem if a pipe delimited file is created. If five delimiters are expected on every record, the record with the special address will have 6 delimiters if not intervened and cleansed before being written to the file.  This paper shows a way to handle the situation and alert the developer or the production support team to take steps to address the error.

**DESCRIPTION OF PROCESS**

Figure 1 below shows a representation of a vendor process for the example discussed in this paper.  This paper shows a basic setup of a vendor helping a financial organization (ABC) with some data requirements of the business that ABC needs at a specific frequency (daily).  The process flow indicates a vendor using a file transmission mechanism with encryption (FTP/NDM) in sending the file to the recipient, ABC.

On the receiving end, ABC receives the file, decrypts and loads the file to the host platform where the further processing continues.  The file is then validated for errors and then passed on to other EUC processes and reporting needs of ABC if the validation is successful.  The audit that will be done to the file will begin with the check to the structure of the data to ensure the basic structure is as expected by ABC with the right number of variables on each record.

When a file with an erroneous delimiter count is transmitted for processing, on the receiving end, if the file is just read with the usual DLM='|' option, a shift in values will be noticed in the variables after the address.  In order to mitigate this situation, a small snippet of code to interrogate each record and ensure that all records have the same number of variables (delimiters) as expected on the receiving end is used. If a record was received with unexpected delimiter count, then the process will halt and the support personnel will be alerted.

**ERROR HANDLING**

The SAS® code in Code A shows how the delimiter record with a problem is handled by counting the delimiters.  The code is developed based on the business requirement of  ABC only expects five variables from the input file from the Vendor.  The code uses the count("|") function to count the number of pipes in each record in the input file represented by Data A. The check "`IF DLM_COUNT NOT = 5`" ensures that the error record is written to the output ERR file.  The code that follows checks for any error records and aborts or continues the process as setup. Data B shows the erroneous second record where the

extra pipe exists in the street-name "Anan|dale".  The same record in Data A shows correctly as "Annandale".   The code will identify the record and write it to the error file.

Count the number of delimiters in the input file for SESUG 2015, continued

## ADVANTAGES

Prompt and timely halt and alerts provide robustness to the process and the recipient is assuring a quality check from a variable count perspective.  In addition to preventing any erroneous processing, timely alerts might save other EUC (End User Computing) related costs in the organization as well.  The file can be fixed and resent before another audit can screen through and release the file for further processing by other jobs/programs/scripts. A rough example (code) of the delimiter audit is included to show how it might be applied to mitigate the issue. The subsequent processing may be handled by the respective job schedulers used in different mainframe (or other) shops with appropriate controls in place as needed.

## DISADVANTAGES

The additional processes can cost time if the response is not sufficiently prompt.  The process holds the rest of the data that is clean from being available to the rest of the EUC or processing. The delays may impact the Service Level Agreements

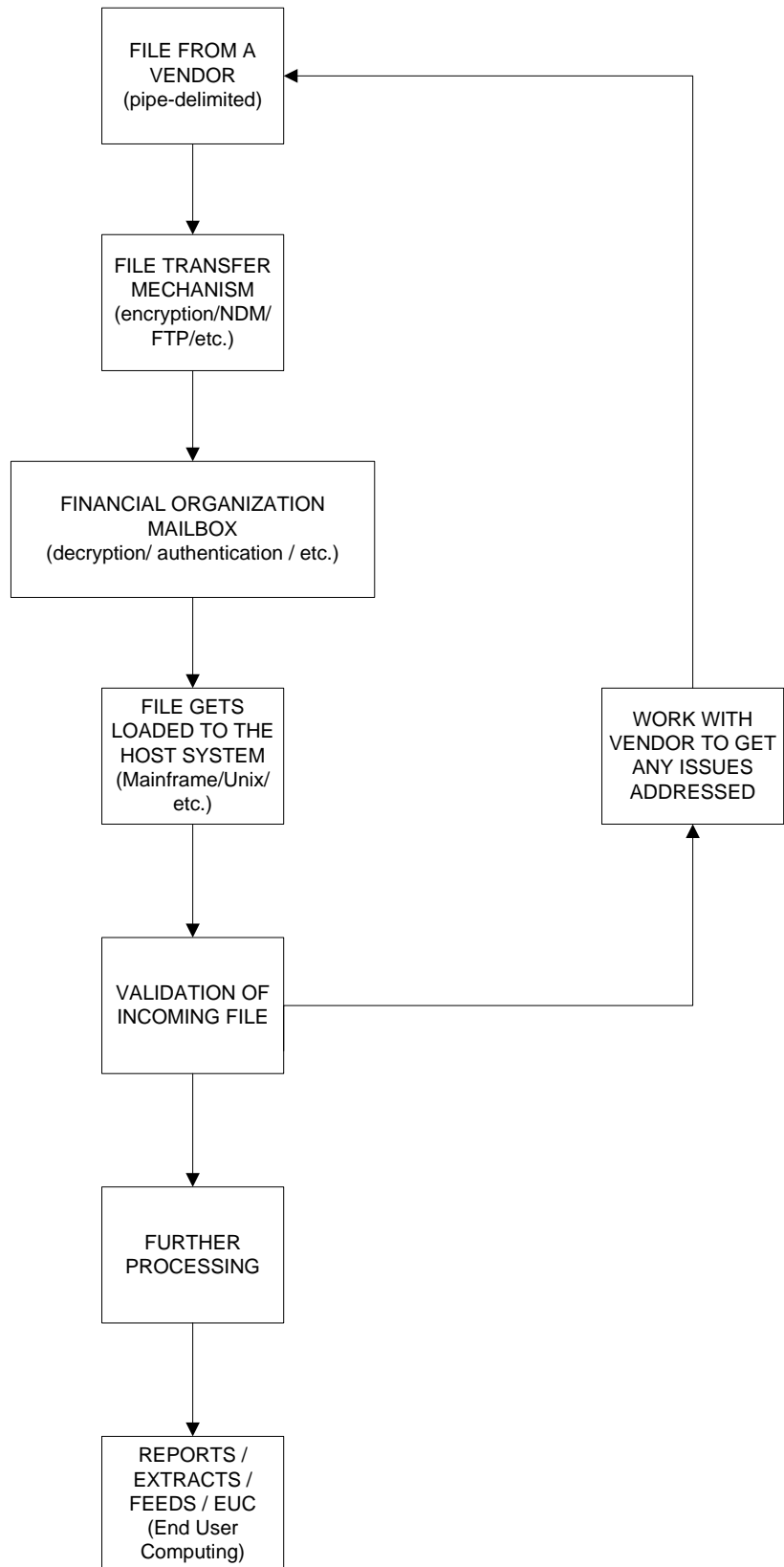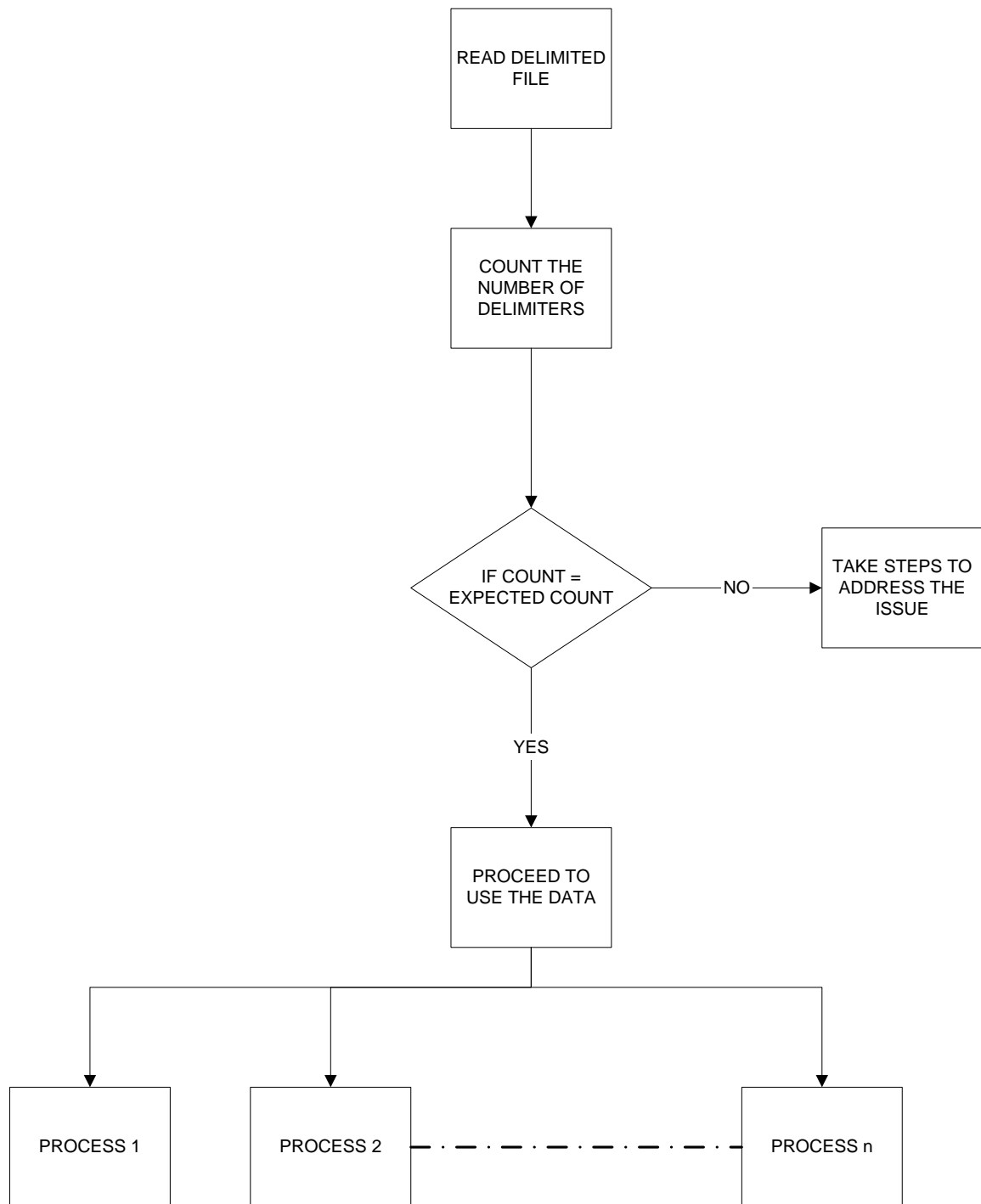Count the number of delimiters in the input file for SESUG 2015, continued

```
           ┌─────────────────┐
           │  FILE FROM A    │◄──────────────────┐
           │  VENDOR         │                   │
           │ (pipe-delimited)│                   │
           └────────┬────────┘                   │
                    │                            │
                    ▼                            │
           ┌─────────────────┐                   │
           │ FILE TRANSFER   │                   │
           │ MECHANISM       │                   │
           │ (encryption/NDM/│                   │
           │ FTP/etc.)       │                   │
           └────────┬────────┘                   │
                    │                            │
                    ▼                            │
      ┌──────────────────────────┐              │
      │ FINANCIAL ORGANIZATION    │              │
      │ MAILBOX                   │              │
      │ (decryption/ authentication / etc.)      │
      └────────────┬─────────────┘              │
                   │                            │
                   ▼                   ┌─────────────────┐
          ┌─────────────────┐          │ WORK WITH       │
          │ FILE GETS       │          │ VENDOR TO GET   │
          │ LOADED TO THE   │          │ ANY ISSUES      │
          │ HOST SYSTEM     │          │ ADDRESSED       │
          │ (Mainframe/Unix/│          └────────▲────────┘
          │ etc.)           │                   │
          └────────┬────────┘                   │
                   │                            │
                   ▼                            │
          ┌─────────────────┐                   │
          │ VALIDATION OF   ├───────────────────┘
          │ INCOMING FILE   │
          └────────┬────────┘
                   │
                   ▼
          ┌─────────────────┐
          │ FURTHER         │
          │ PROCESSING      │
          └────────┬────────┘
                   │
                   ▼
          ┌─────────────────┐
          │ REPORTS /       │
          │ EXTRACTS /      │
          │ FEEDS / EUC     │
          │ (End User       │
          │ Computing)      │
          └─────────────────┘
```

**Figure 1. A typical vendor file process**

Count the number of delimiters in the input file for SESUG 2015, continued



**Figure 2. Delimiter Validation of incoming file**

Count the number of delimiters in the input file for SESUG 2015, continued

```
//AUDIT    JOB(ABCDE,FG),'AUDIT JOB',MSGCLASS=T,CLASS=G,NOTIFY=&SYSUID
//SAS01    EXEC SAS
//SASLOG   DD SYSOUT=*
//SASLIST DD SYSOUT=*
//INFILE   DD DSN=FILE.INFILE,DISP=SHR
//ERRFILE DD DSN=FILE.ERRFILE,DISP=(NEW,CATLG,CATLG),
//         DCB=(LRECL=80,RECFM=FB),SPACE=(CYL(1,1),RLSE)
//SYSIN    DD *
DATA ERR;
     INFILE INFILE TRUNCOVER;
      INPUT  @001 INREC $80.;
            DLM_COUNT = COUNT(INREC,"|");
            IF DLM_COUNT NOT = 5
            THEN OUTPUT ERR;
RUN;
DATA _NULL_;
     SET ERR;
     FILE ERRFILE;
          PUT @001 DLM_COUNT INREC;
RUN;
DATA _NULL_;
     SET ERR;
     IF _N_=1
     THEN ABORT 255;
RUN;
```

**Code A. Sample code counting delimiters with error identification**

Count the number of delimiters in the input file for SESUG 2015, continued

**INFILE – DATA**

~~1234567890~~1234567890~~1234567890~~1234567890~~1234567890~~
010|SAM TRAUTMAN|23 SMITH ST|LINDEN|NJ|07065  <<< 5 delimiters
020|BARRY DELMAR|46 ANNANDALE DR|AVA|NY|13309  <<< 5 delimiters
030|MANUEL ORTEGA|5 LOCO WY|TAMPA|FL|33510    <<< 5 delimiters

**Data A.  Sample Clean data with pipe "|" delimiter**

~~1234567890~~1234567890~~1234567890~~1234567890~~1234567890~~
010|SAM TRAUTMAN|23 SMITH ST|LINDEN|NJ|07065   <<< 5 delimiters
020|BARRY DELMAR|46 ANNAN|DALE DR|AVA|NY|13309  <<< 6 delimiters
030|MANUEL ORTEGA|5 LOCO WY|TAMPA|FL|33510     <<< 5 delimiters

**Data B.  Sample Erroneous data with pipe "|" delimiter**

**CONCLUSION**
The use of count function to count the delimiters as a first step of auditing a file may perhaps save time, processing cost, EUC, effort to analyze and debug processes as needed in a design.

**RECOMMENDED READING**

- *Base SAS® Procedures Guide*
- SAS® For Dummies®

**CONTACT INFORMATION**
Your comments and questions are valued and encouraged. Contact the author at:

Name: Kannan Deivasigamani
Email:  kannand@hotmail.com
Web: https://www.linkedin.com/in/kannandeivasigamani