General Ledger Clean Up

Preparing a General Ledger Report for Accounting Systems

Overview

The Accounting Department needs to verify the last GL report issued to management against data in the Corporate ERP system. The only version of the report available is a formatted Excel spreadsheet which is not in a format that can be used by any of the accounting systems. A workflow needs to ingest the spreadsheet and build a clean tabular data set that can be consumed by the accounting systems.

Objective

Build a repeatable workflow that is capable of consuming the report and producing the consistent, tabular formatting required by the accounting systems. You can manually rename the column headers, but finding a way to pull them dynamically from the data set itself is even more valuable. Make sure the resulting output matches the format described below.

Processes

1. Pull in the general ledger file. Add a record identifier to maintain record order and update the column headers (either manually or dynamically) using the data dictionary in this document.
2. Parse out the account information into two separate fields, account number and account name.
3. Fill in the account number and name fields appropriately. Remove records that don't contain general ledger data. Fill in the missing PD values, and format the data as prescribed in the data dictionary.
4. Output an Excel file with the cleaned data set that is suitable for the accounting system.

Data Dictionary

2020 – GL Detail Report.xlsx

Header information starts on the 9th row

Column Name	Data Type	Description
<GSRPTLYT>	V_String	Does not contain useful data
F2	V_String	Contains account number and name
F3	V_String	Contains trans code
F4	V_String	Contains PD
F5	V_String	Contains source
F6	V_String	Contains voucher number
F7	V_String	Contains vendor name
F8	V_String	Does not contain useful data
F9	V_String	Contains description
F10	Double	Does not contain data
F11	V_String	Contains debit
F12	V_String	Contains credit
<MAX_COLUMN>	V_String	Contains balance

Required Output Structure

Column Name	Data Type	Description
AccountNumber	V-String	
AccountName	V-String	
Trans Code	V-String	
PD	V-String	
Source	V-String	
Voucher Number	V-String	
Vendor Name	V-String	
Description	V-String	
Debit	Double	
Credit	Double	
Balance	Double	