INFORMATICA | MARKETPLACE

Name of Solution:

Script: Process Flat files from Directory

Business Requirement:

Script and workflow combination that helps you process flat files on a first-in-first-out basis from a directory in near real-time.

Solution URL:

https://community.informatica.com/solutions/2236

Supported Versions:

PowerCenter 9.1 and 9.5

Description:

Consider the scenario where, several flat files (in the same format) are delivered to a location from another application. Each of these flat files has a different name and should be processed one at a time on a first-in-first-out basis. The processing of files should be near real-time.

The solution provided in this block is in the form of a workflow containing a command task and a mapping.

1. The command task creates a filelist with the name of the oldest file in the inbound directory
2. The mapping uses the filelist as an indirect file and processes the file and moves the source to archive.
3. The workflow is run again to process the next file.

How the command (shell script) task works:

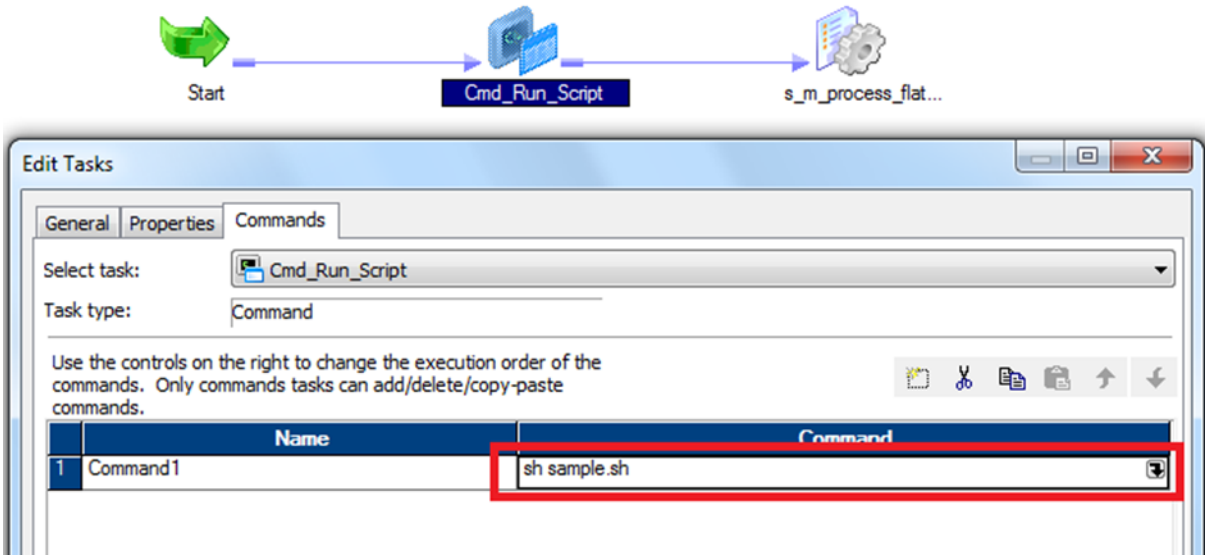
1. The script creates a file TEMP_DIRECTORY*new_files.txt* containing file names from INBOUND_DIRECTORY and list is ordered by timestamp.
2. Reads the first file name (oldest) from new_files.txt and move corresponding file from the INBOUND_DIRECTORY to SOURCE_DIRECTORY.
3. Create the file *process_file.txt* in TEMP_DIRECTORY with the name of the moved file and copy *process_file.txt* to SOURCE_DIRECTORY to be used as an indirect source by the session.

Download file contents:

1. Script
2. Workflow
3. Source Files

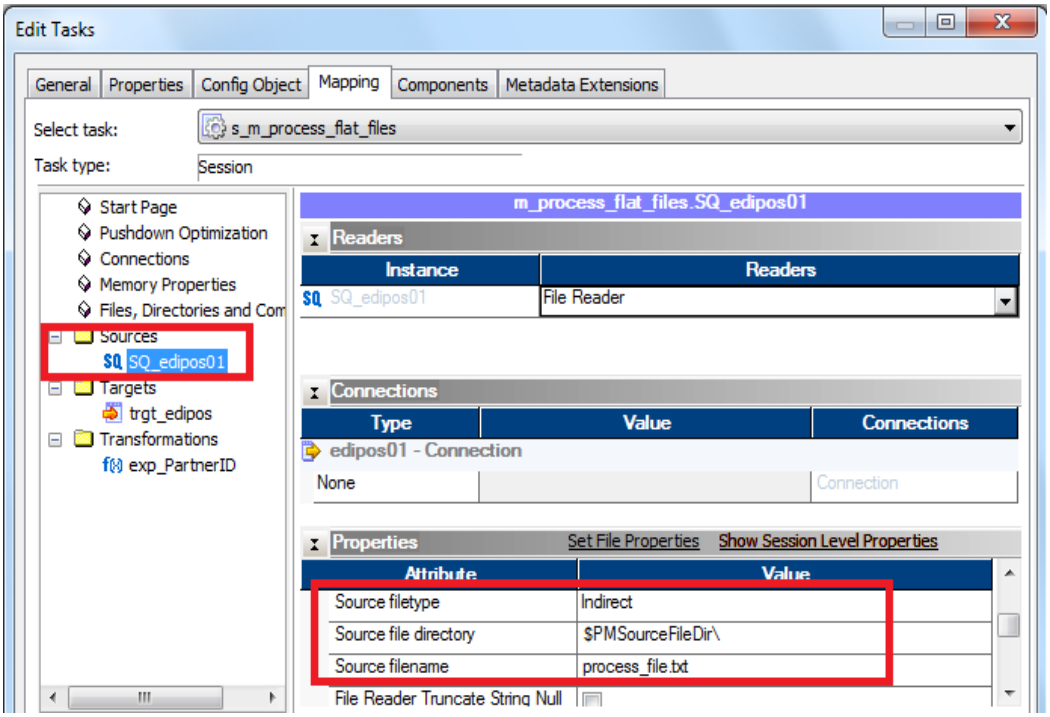
Steps to implement the solution:

1. Copy the source file provided to $PMSourceFileDir with following structure.
    a. TEMP
    b. INBOUND
    c. ARCHIVE
2. Provide the path for following variables in script file
    a. TEMP_DIRECTORY - $PMSourceFileDir\TEMP
    b. SOURCE_DIRECTORY - $PMSourceFileDir
    c. INBOUND_DIRECTORY - $PMSourceFileDir\INBOUND
3. Put this shell script in ..\server\bin directory.
4. Import the XML using Repository Manager and resolve the conflicts by selecting appropriate option. Observe the source files in the *inbound* directory. When the script runs, files will be transferred to source directory.
5. Open workflow in Workflow Manager; observe the command task as shown below.

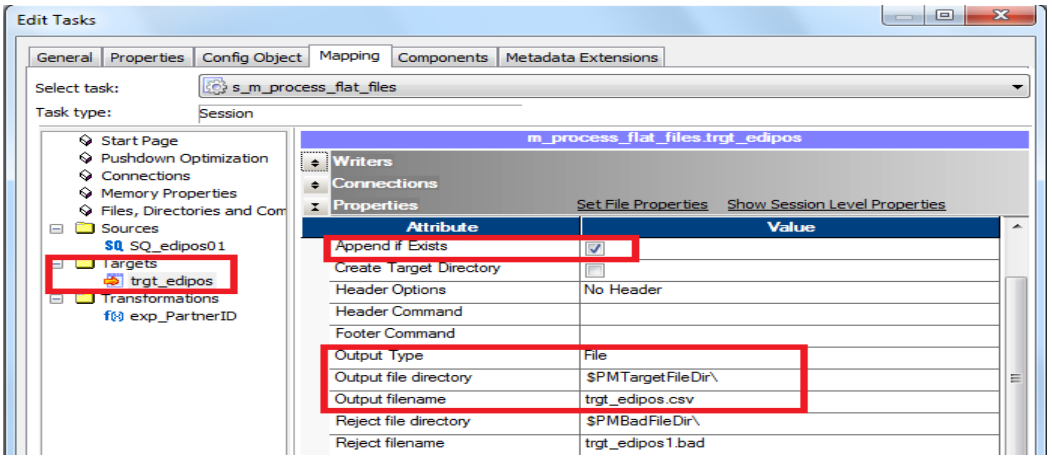


6. Observe the mapping tab of Session as shown below

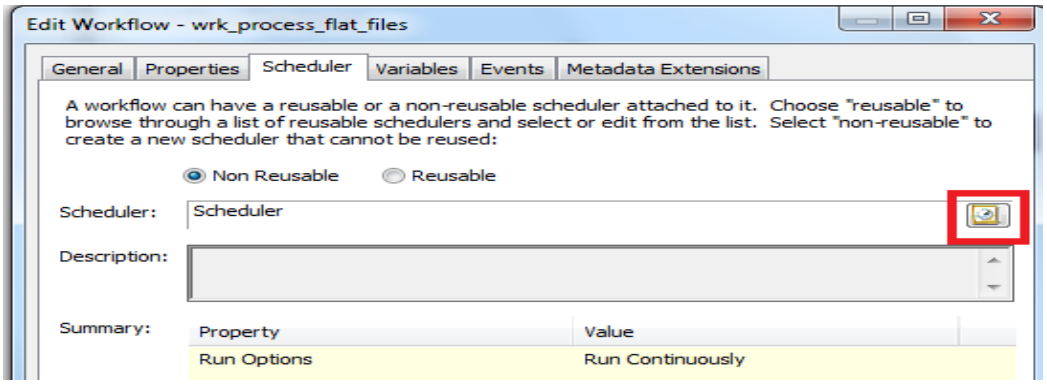
For Source File:



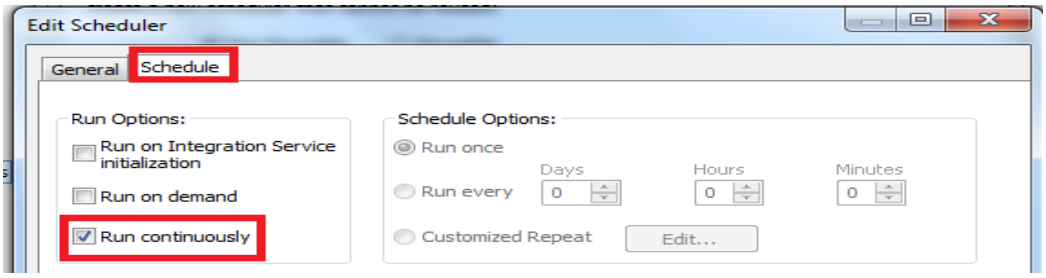
For Target File:



7. Observe the scheduler setting of workflow as shown below



Observe that the check box is checked for "*run continuously*" option as shown below



8. Start the workflow and observe the workflow runs continuously for each file. Observe the changes in INBOUND and ARCHIVE directory with each run of session.
9. Check the output file in the $PMTargetFileDir and processed files in Archive Directory Note: Ensure that the workflow should run in scheduled mode.
10. Provide your feedback for this solution on Solution URL.

YouTube Video on Importing and Configuring Workflows:

http://www.youtube.com/playlist?list=PLLRreK2jjjgWBQ4NPfp0QWTxYDvlnEqSJ

Other Useful links:

Mapping Bundles
Workflow Bundles
Informatica Tools Bundles
Informatica for Social Media Bundles
Debugging Tools Bundles
Visio Templates Bundles
B2B Templates and Projects Bundles
Data Quality Packs and Plans Bundles
Script Bundles