Performance Toolkit CSV Generator

IBM

August 5, 2008

Contents

1 Introduction

The Performance Toolkit CSV Generator is a useful tool for converting trend, summary, and history record files into CSV (comma separated value) files. These CSV files can then be used by spreadsheet programs to analyze data or be fed into another program to perform further data processing.

2 Using Performance Toolkit CSV Generator

2.1 Basic Usage

The Performance Toolkit CSV Generator requires as input a trend, summary, or history record file. This along with a set of options can allow the utility to generate a varying set of CSV files.

A call to the Performance Toolkit CSV Generator follows the format:

CSVGEN file_format file_name file_type file_mode output_file_mode prefix (options)

Required Paramenters

- file_format: specify type of file being read by the CSV generator. Available options:

  - H: HISTLOG/HISTSUM file
  - T: Trend record file
  - S: Summary record file

- file_name: Input file name to be read by the CSV generator.

- file_type: Input file type to be read by the CSV generator.

Optional Parameters (Must be specified in order listed above)

- file_mode: Input file mode to be read by the CSV generator. The default if no value is give is file mode 'A'.

- output_file_mode: Output file mode of all files generated by the CSV generator. The default if no value is given is file mode 'A'.

- prefix: The prefix is a 3 letter prefix attched to the base output file name by a '_'. This helps to differentiate different sets of CSV files. The default if no value is given is the first three letters of file_name. If a prefix of over 3 letters is given, it will be truncated down to 3 letters.

Options (Can be specified in any order)

- DEScription: Include field descriptions in generated CSV files. Descriptions are placed on the second line of each CSV file. This is on by default.

- NODEScription: Do not include field descriptions in generated CSV files.

- SPLit: Split CSV files containing a field count larger than 256 into a second file. This is useful for importing into a spreadsheet program.

- NOSPLit: Do not split CSV files containing a field count larger than 256 into a second file. This is useful for post–processing generated CSV files. This is on by default.

- TAble: Data for fields containing a table of information will be placed in their own CSV files. This is on by default.

- NOTAble: Leave table fields in HEX representation in the parent CSV file.

- FINalize: Finalize CSV files so they take up less space and are more friendly for later viewing/post–processing. Will not be able to append later records to these files.

- NOFINalize: Leave CSV files as they are created. This will allow for an updated history/trend/summary file to be processed and appended to the existing CSV files. This is on by default.

  Note: Capitalized letters represent the minimum abbreviation for each option.

2.2 Generated CSV Files

When the program runs, many CSV files may be generated. In general, the files generated will follow the format:

- PREFIX_FC## CSV OUTPUT_FILE_MODE for trend files

- PREFIX_SM## CSV OUTPUT_FILE_MODE for summary files

- PREFIX_HIST CSV OUTPUT_FILE_MODE for history files

  Note: ## represents the trend number of the record.

There are a few exceptions to this rule however. If the table option is passed, 4 extra files will be generated. PREFIX_PT03 for record FC03 and files PREFIX_P64T, PREFIX_PTYP, and PREFIX_PST for record FC00. If the SPLit option is passed PREFIX_FC01 will be split into PREFIX_FC01 and PREFIX_01P2, and PREFIX_HIST will be split into PREFIX_HIST and PREFIX_HST2.

2.3 Notes on Usage

If an input file is later updated, CSVGEN can be run with the same options used to process that file before if the existing output files are not finalized. This will append the new data onto the end of these files. If CSVGEN is halted or abends mid–execution, it is possible to resume execution from where it stopped processing data at as long as the output files are not finalized (which will not happen unless done manually in these cases) or deleted and the same parameters are used to run CSVGEN again.

The generated CSV files may take up a large amount of blocks depending on the input file. This is true even if the finalize option is passed as the files do not get finalized until the last record has been processed. A tempory device for storing the generated CSV files is recommend for large data inputs. The

program will quit with an error message if at anytime there is not enough space left on the device used to store the CSV file. If this happens either increase the space available on the intended target device or allocate a temporary device to temporarily store the CSV files. Note that if the files are to be retrieved from VM via FTP they will need to be stored on a non-temporary device.

When finalizing files, make sure enough virtual storage is defined, as the finalizing process utilizes XEDIT. Large input files generally produce at least 1 large output file, so virtual storage must be assigned to allow XEDIT to edit and save very large files. If an error would occur at the end of running the program during the finalizing process due to insufficient virtual storage, increase virtual storage and run the finalize command as instructed below.

If the finalize option is not passed and at a later time it is decided that the CSV files should be finalized, the finalize xedit macro is provided to do this. To finalize the CSV files run the commands:

```
FILEL PREFIX_* CSV OUTPUT_FILE_MODE
EXECUTE * XEDIT / ( profile finalize
```

It is recommended the CSV files be finalized before viewing with a spreadsheet program or feeding into another program. If the CSV files will be appended to at a later date, copy the CSV files before finalizing them.

OpenOffice.org is the recommend program for viewing the generated CSV files if any time releated field is important. Microsoft Excel will attempt to incorrectly format these timestamp fields, ruining the data in these fields. OpenOffice.org will format them by default correctly, rounding the seconds part of the timestamp to the nearest whole second. This can be changed only when importing the file to OpenOffice.org by selecting the appropriate columns and selecting text as the column type. See Figure 1.
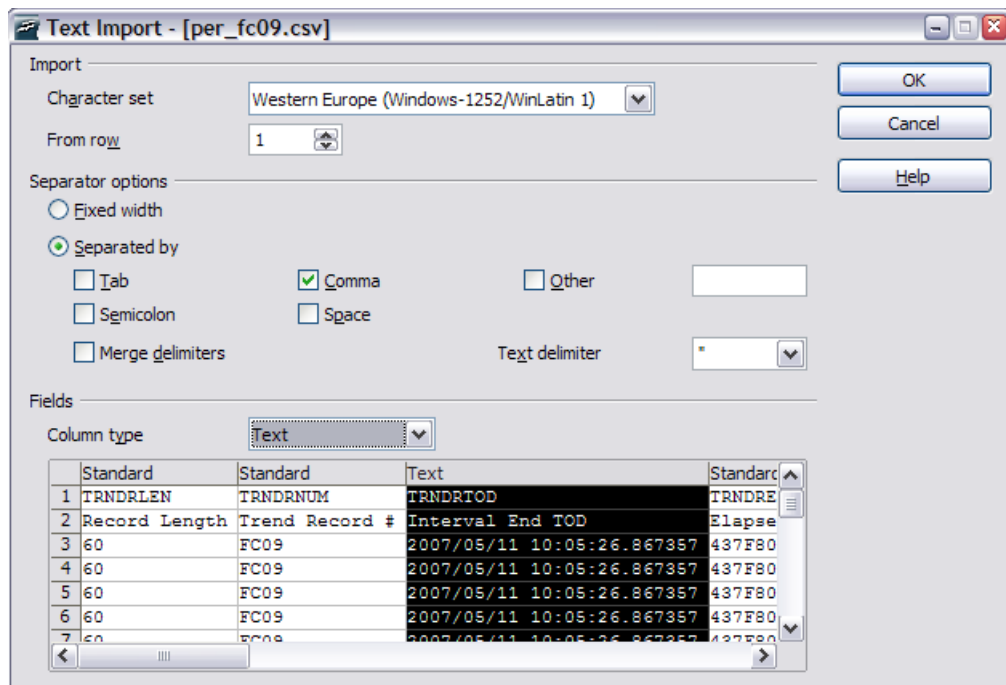
Figure 1: Keeping the original format in OpenOffice.org