NAME

rastream – stream block processor for **argus(8)** data.

SYNOPSIS

rastream [[**–M** *splitmode*] [*splitmode options*]] [**–f** *file processing program* -B secs] [**raoptions**] [**--** *filter-expression*]

DESCRIPTION

Rastream reads **argus** data from an *argus-data* source, and splits the resulting output into consecutive sections of records based on size, count time, or flow event, writing the output into a set of output-files.

Rastream provides the option to run a program against the output files, **seconds** after the file is understood to be finished. The program must be specified in a manner so that **rastream** can find it, either within the system $PATH, or provided as a full pathname.

By default, **rastream** splits the stream by output file record count, putting 10,000 records of input into each **argus** output file, or standard out, as needed. The behavior is similar to the unix split.1 command.

The output files' name consists of a prefix, which is specified using the *-w ra option*, and a suffix, which is created for each resulting file. If no prefix is provided, then **rastream** will use 'x' as the default prefix. The suffix that is used is determined by the mode of operation. When **rastream** is using the default count mode or the size mode, the suffix is a group of letters 'aa', ´ab´, and so on, such that concatenating the output files in sorted order by file name produces the original input file. If **rastream** will need to create more output files than are allowed by the default suffix strategy, more letters will be added, in order to accomodate the needed files. When the mode is **time** mode, the default output filename suffix is '%Y.%m.%d.%h.%m.%s', which is used by strftime() to create an output filename that is time oriented. This default is overrided by adding a '%' extension to the name provided on the commandline using the *-w* option.

When standard out is specified, using *-w -*, **rastream** will output a single **argus-stream** with START and STOP argus management records inserted appropriately to indicate where the output is split. See **argus(8)** for more information on output stream formats.

When **rastream** is spliting on output record count (the default), the number of records is specified as an ordinal counter, the default is 10,000 records. When **rastream** is spliting based on the maximum output file size, the size is specified as bytes. The scale of the bytes can be specified by appending 'b', 'k' and 'm' to the number provided.

When **rastream** is spliting based on time, the time period is specified with the option, and can be any period based in seconds (s), minutes (m), hours (h), days (d), weeks (w), months (M) or years (y). **Rastream** will create and modify records as required to split on prescribed time boundaries. If any record spans a time boundary, the record is split and the metrics are adjusted using a uniform distribution model to distribute the statistics between the two records. Care is taken to avoid records with zero packet and byte counts, that could result from roundoff error.

When **rastream** is spliting based on flow event, the flow that acts as the event marker is specified using a standard **ra** filter expression, that is bounded by quotes ("). Records that preceed the first flow event in the data stream are written to the specified output file, and then new files are generated with the flow event record being the first record of the new file. This method will allow you to use wire events as triggers for spliting data.

RASTREAM SPECIFIC OPTIONS

Rastream, like all ra based clients, supports a number of **ra options** including remote data access, reading from multiple files and filtering of input argus records through a terminating filter expression. **rastream(1)** specific options are:

−a *suffix length*
Starting append suffix length. The default is 2 characters.

−B *duration*
Buffer hold time before processing. This value is usually in the 5-15 second range and provides time for rastream to sort records and schedule outputfile processing. The number is derived from the larges FAR status interval of all the argus data sources encountered.

−f *program*
Post processing program. **rastream**, will execute this program just after closing the output file, passing the full path to the closed output file as a parameter, using this convention:

program -r /full/path/to/closed/file

This allows you to post-process the output file in an automated fashion.

Generally, this program can do anything you like, such as aggregating and correcting flow records, labeling records for semantic enhancement, indexing the files, using programs like rasqltimeindex(), and compressing the files. Traditionally, the program has been a shell-script, perl program, or php script, so that it can be easily modified, on the fly, but it can be any executable that can handle the "-r filename" parameter convention. The program should provides its own accountability and error logging, so that you know that things are working as you expect.

rastream must have a path to the program, the program must be executable, and **rastream** must have permission to run the program for this strategy to be successful.

An example rastream.sh is provided in the ./support/Config directory.

−M *splitmode*
Supported spliting modes are:
count <num>
size <size>
time <period>
flow "filter-expression"

−w *filename*
Rastream supports an extended *-w* option that allows for output record contents to be inserted into the output filename. Specified using '$' (dollar) notation, any printable field can be used. Care should be taken to honor any shell escape requirements when specifying on the command line. See **ra(1)** for the list of printable fields.

Another extended feature, when using **time** mode, **rastream** will process the supplied filename using **strftime(3)**, so that time fields can be inserted into the resulting output filename.

INVOCATION

This invocation reads **argus(8)** data from **inputfile** and splits the **argus(8)** data stream based on output file size of no greater than 1 Megabyte. The resulting output files have a prefix of *argus.* and suffix that starts with 'aa'. The single trailing '.' is significant.

rastream -r inputfile -M size 1m -w argus.

This invocation splits **inputfile** based on hard 10 minute time boundaries. The resulting output files are created with a prefix of */archive/%Y/%m/%d/argus.* and the suffix is *%H.%M.%S*. The values will be

supplied based on the time in the record being written out.

rastream -r * -M time 10m -w "/archive/%Y/%m/%d/argus.%H.%M.%S"

This invocation splits **inputfile** based on the argus source identifier. The resulting output files are created with a prefix of */archive/Source Identifier/argus.* and the default suffix starting with "aa". The source identifier will be supplied based on the contents of the record being exported.

rastream -r * -M time 10m -w "/archive/$srcid/argus."

This invocation splits **inputfile** based on a flow event marker. The resulting output files are created with a prefix of 'outfile.' and the default suffix starting with "aa". Whenever a ping to a specific host is seen in the stream, a new output file is generated.

rastream -r * -M flow "echo and host 1.2.3.4" -w outfile.

COPYRIGHT

SEE ALSO

ra(1), rarc(5), argus(8),

AUTHORS

Carter Bullard (carter@qosient.com).