

Logs in Custom Format

If your logs are in a custom format, you must write a custom parser ([instructions](#)). After you write the parser, you put the file into your `/etc/td-agent/plugins/` directory.

- [Common Parser Examples](#)
- [Filtering the Records](#)

Common Parser Examples

Treasure Data provides two example parsers: “URL-param style key-value pairs” and “ascii character delimited format”. Both formats are fairly common among our users.

```
# URL-param style key-value pairs
last_name=smith&first_name=brian&age=22&state=CA

# ASCII character delimited format. In this case, the delimiter is '|'.
# There is usually a separate file that annotates the column names
smith|brian|22|CA
```

- [Custom Parser for URL-Param Style Key-Value Pairs](#)
- [Custom Parser for Ascii-Character Delimited Logs](#)

Tailing existing logs is an easy way to get started with Treasure Data. We recommend logging everything as JSON. [Here's why.](#)

Filtering the Records

If you need to filter logs (ex: filtering out impressions and just keeping clicks), [the exec-filter plugin](#) is useful. This plugin launches another script which takes STDIN as input and STDOUT as output, and filters logs accordingly.

Here's an example configuration.

```
<source>
  type tail
  path /path/to/the/file1
  tag filter
  format json
  pos_file /var/log/td-agent/file1.pos
</source>

<match filter>
  type exec_filter
  command /usr/lib64/fluent/ruby/bin/ruby /etc/td-agent/filter.rb
  in_format json # takes a JSON string from STDIN
  out_format json # generates a JSON string to STDOUT
  tag_key tag # The key for tags is "tag".
  time_key time # The key for timestamps is "time".
</match>

<match td.*.*>
  type tdlog
  endpoint api.treasuredata.com
  apikey ...
  auto_create_table
  buffer_type file
  buffer_path /var/log/td-agent/buffer/td
  use_ssl true
</match>
```

`/etc/td-agent/filter.rb` is the filter script, as shown in the following example. The script filters out all the lines where the field “field0” is equal to “certain_value”. Errors are recorded in `/var/log/td-agent/filter.rb.log`.

```
open('/var/log/td-agent/filter.rb.log', 'a') { |f|
  f.puts "-- begin --"
  begin
    require 'json'
    STDOUT.sync = true
    while line = STDIN.gets
      # parse
      begin
        h = JSON.parse line
      rescue => e
        next # broken line
      end
      # filter
      # next if h["field0"] == "certain_value"
      # emit
      h['tag'] = 'td.testdb.test_table'
      puts h.to_json
    end
  rescue LoadError => e
    f.puts e.to_s
  end
}
```