Structs CSV / TSV Format Reference

The XINA Structs CSV / TSV formats provide a standard delimited text file format for mnemonic data.

Source File Format

Files must be either ASCII or UTF-8 encoded. New lines will be interpretted from either `\n` or `\r\n`. The `conf` object may define other customization of the format:

Conf Definition

Key	Value	Default	Description
delimiter	string	auto detect (`','`, `'\t'`, `';'`)	value delimiter
quote_char	character	`"` (double quote character)	value quote character
ignore_lines	number	`0`	lines to ignore after UUID and before header
mode	`"row"` or `"col"`	auto-detect	mnemonic mode (see below)
t	`"auto"`, `"iso8601"`, `"s"`, `"ms"`, or `"us"`	`"auto"`	time format (see below)
zone	string	time zone to use if not provided	

The first line must contain an appropriately generated 128-bit UUID in the standard 36 character format.

If the `mode` property is `"row"`, the file must contain three columns:

Name	Description	Alternate Names
t	Unix time or ISO8601 zoned timestamp	time, timestamp
mn	mnemonic name or ID	mnemonic, n, name
v	value (numeric, empty, or `null`)	val, value

The header is used to determine the order of the columns.

For example (whitespace added for clarity, not required):

```
123e4567-e89b-12d3-a456-426614174000
t , mn    , v
0 , v_mon , 1
0 , i_mon , 5
```

```
1 , t_mon , 100
2 , v_mon , 1.1
2 , i_mon , 4
3 , t_mon ,
4 , v_mon , 1.2
4 , i_mon , 3
5 , t_mon , 101
```

If `mode` is `"col"`, the file must first contain a time column, followed by a column for each mnemonic. The column headers must specify the mnemonic name or ID for each column. Unlike `row`, `null` values must be spelled out explicitly, as empty values will **not** create a point in the database.

For example, the following is equivalent to the above example (whitespace added for clarity, not required):

```
123e4567-e89b-12d3-a456-426614174000
t      , v_mon , i_mon , t_mon
0      , 1     , 5     ,
1      ,       ,       , 100
2      , 1.1   , 4     ,
3      ,       ,       , null
4      , 1.2   , 3     ,
5      ,       ,       , 101
```

If the `mode` property is not specified, the mode will be determined by the number of columns in the file. If there are exactly 3 columns with names matching the required columns for the `"row"` mode, that mode is used; otherwise the file is assumed to use the column mode.

Time Parsing

The mode of time processing is determined by the value for `t` in `conf`. The `auto` mode attempts to interpret the most likely formatting for the timestamp. If the value is an integer or floating point format, it will be interpretted as a Unix timestamp, with precision based on these rules:

- t > `1e16`: error, value above typical range
- t > `1e14`: microseconds
- t > `1e11`: milliseconds
- t > `1e8`: seconds
- t <= `1e8`: error, value below typical range

Otherwise it will be interpretted as a zoned ISO8601 timestamp. If `t` is set explicitly in the configuration the time will always be interpretted in that context. The ISO timestamp may use the standard format: `2023-05-31T17:55:07.000` or condensed `20230531T175507.000`. If the `zone` property provided in the configuration, the timestamps do not require a zone. Otherwise they must include an explicit zone.

Revision #17
Created 28 July 2022 20:31:12 by Nick Dobson
Updated 4 October 2023 20:31:26 by Nick Dobson