### textfile

テキストファイルからデータを読み込みます。

#### 構文

textfile [OPTIONS] PATH

必須パラメータ

**PATH**

ファイルパス。絶対パスで入力するか、ログプレッソ・ソナーエンジンを基準とした相対パスで指定できます。ファイル名にワイルドカード（\*）を使用してパターンマッチング方式でファイルを参照することも可能です。例えば、PATHにallow-\*.txtを指定すると、allow-ip.txt、allow-user.txt、allow-url.txtなどのファイルをまとめて参照できます。ファイルを読み込むには、ログプレッソ・ソナーの実行アカウントにアクセス権限が付与されている必要があります。

オプションパラメータ

**brex="REGEX"**

1つのレコードが複数行で構成されている場合、レコードの開始行を検出するために使用する正規表現です。正規表現に一致する行が現れる前の行、またはerexで指定した最後の行までを1つのレコードとして結合します。指定しない場合は改行文字（CRLFまたはLF）を基準として認識します。

レコードの終了行を検出する場合はerexオプションを使用します。

**cs=CHARSET**

文字列エンコーディング形式（デフォルト：utf-8）。下記ドキュメントに登録されているPreferred MIME NameまたはAliasesを使用してください: <https://www.iana.org/assignments/character-sets/character-sets.xhtml>

**df="TIME\_FMT"**

dpで抽出した日付データをパースする際に使用するフォーマット文字列。yyyy-MM-dd HH:mm:ss.SSSのように指定します。dpオプションと併用します。指定しない場合は、データの読み込み時点の時刻が\*\*\_time\*\*フィールドに記録されます。

フォーマット文字列には以下の日付ディレクティブを使用できます。

|  |  |  |
| --- | --- | --- |
| ディレクティブ | 説明 | 例 |
| G | 紀元前/後 | AD |
| y | 暦年 | 2025（yyyy）；25（yy） |
| M | 月 | July（MMMM）；Jul（MMM）；07（MM）、7（M） |
| w | 年単位の週番号 | 27（その年の27週目） |
| W | 月単位の週番号 | 2（その月の2週目） |
| D | 年単位の日 | 10 |
| d | 月単位の日 | 189 |
| E | 曜日 | Tuesday（EEEE）；Tue（E, EE, EEE） |
| F | 月単位の曜日番号 | 2（その月の2回目の曜日） |
| u | 曜日番号（1=月曜, …, 7=日曜） | 1 |
| a | 午前/午後 | PM |
| H | 時（0-23） | 0 |
| k | 時（1-24） | 24 |
| K | 午前/午後の時（0-11） | 0 |
| h | 午前/午後の時（1-12） | 12 |
| m | 分 | 30 |
| s | 秒 | 55 |
| S | ミリ秒 | 978 |
| z | タイムゾーン（一般表記） | Pacific Standard Time; PST |
| Z | タイムゾーン（RFC 822表記） | -0800 |
| X | タイムゾーン（ISO 8601表記） | -08;-0800;08:00 |

例えば、読み込んだデータに2000-01-01 11:22:33のような文字列があり、これと同様のパターンから日付データを抽出する場合、フォーマット文字列をyyyy-MM-dd HH:mm:ssと指定すると、**\_time**フィールドに2000年1月1日11時22分33秒が記録されます。dfを指定しない場合は、データ読み込み時点の時刻が\*\*\_time\*\*フィールドに記録されます。

**dp="REGEX"**

**\_time**フィールドに記録する日付抽出用正規表現。例えば、読み込んだデータに2000-01-01 11:22:33のような文字列があり、同じパターンから日付データを抽出する場合、dpに正規表現(\d{4}-\d{2}-\d{2} \d{2}:\d{2}:\d{2})を指定します。（正規表現のメタ文字\は文字列内でエスケープとして\を使用します。）dpで抽出した日付データはdfで指定したフォーマットでパースされます。dfオプションと併用します。このオプションを指定しない場合は、データ読み込み時点の時刻が\*\*\_time\*\*フィールドに記録されます。

**erex="REGEX"**

1つのレコードが複数行で構成されている場合、レコードの終了行を検出するために使用する正規表現です。正規表現に一致する行が現れるまで、またはファイルの最終行からファイル末尾までを1つのレコードとして結合します。指定しない場合は改行文字（CRLFまたはLF）を基準として認識します。

レコードの開始行を検出する場合はbrexオプションを使用します。

**limit=INT**

取得する最大ログ件数（デフォルト：制限なし）

**offset=INT**

スキップするログ件数（デフォルト：0）

#### 使用例

/var/log/secureログファイルの参照

textfile /var/log/secure

euc-krでエンコードされたiis.txtファイルの参照

textfile cs=euc-kr iis.txt

/var/logディレクトリ内でファイル名にsyslog.を含む全てのgzファイルの参照

textfile /var/log/syslog.\*.gz

test.txtファイルの内容のうち、最初の5行をスキップして20件を参照

textfile offset=5 limit=20 test.txt

レコード開始行が角括弧（）で囲まれた文字列を含み、終了行がend文字列を含むtest.txtファイルの参照

textfile brex="\[.\*\]" erex="end" test.txt

test.txtファイルを参照し、レコード内に2000-01-01 11:22:33形式で含まれる日付データを\_timeフィールドとして抽出

textfile dp="(\\d{4}-\\d{2}-\\d{2} \\d{2}:\\d{2}:\\d{2})" df="yyyy-MM-dd HH:mm:ss" test.txt

#### 互換性

読み込むファイルの拡張子が.gzの場合、自動的にgzipファイルとして認識し、解凍後に参照できます。本機能はENT #2241 2019-04-23\_17-20バージョン以降でサポートされています。