

Supported Special Regular Expression Characters

Contents:

- *Slashes*
- *Supported Special RegEx Characters*
- *Required Escaped Characters*

Cloud Dataprep by TRIFACTA® INC. supports a set of special characters for regular expressions that are common to all of the execution engines supported by the platform.

Slashes

The forward slash character is used to denote the boundaries of the regular expression:

```
/this_is_my_regular_expression/
```

- The backslash character (\) is the escaping character. It can be used to denote an escaped character, a string, literal, or one of the set of supported special characters.
- Use a double backslash (\\) to denote an escaped string literal. For more information, see *Escaping Strings in Transformations*.

Supported Special RegEx Characters

The table below identifies the special characters that are supported in the platform.

Special Characters	Description
\\	String literal match for \ character.
\b	Matches any zero-width word boundary, such as between a letter and a space. Example: /\bre/ does not match re in tire, since re is not on the word boundary. /re\b/ does match.
\B	Matches any zero-width non-word boundary, such as between two letters or two spaces. Example: /\Bre/ matches re in tire. It does not match in respect, since that instance of re is on a word boundary.
\cX	Matches a control character (CTRL + A-Z), where X is the corresponding letter in the alphabet.
\d	Matches any digit.
\D	Matches any non-digit.
\f	Matches a form feed.
\n	Matches a line feed. <div style="border: 1px solid black; padding: 5px; margin: 10px 0;">NOTE: These characters are not supported in inputs for Object and Array data types.</div>
\r	Matches a carriage return.

<code>\s</code>	Matches any whitespace character. These characters include: <ul style="list-style-type: none"> • space • tab • form feed • line feed • Other Unicode space characters
<code>\S</code>	Matches any character that is not one of the supported whitespace characters.
<code>\t</code>	Matches a horizontal tab. <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>NOTE: These characters are not supported in inputs for Object and Array data types.</p> </div>
<code>\v</code>	Matches a vertical tab.
<code>\w</code>	Matches any alphanumeric value, including the underscore. <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>Tip: Column names must match the same set of characters.</p> </div>
<code>\W</code>	Matches any non-alphanumeric character, including the underscore.
<code>\xHH</code>	Matches the ASCII character code as expressed by the hexadecimal value HH.
<code>\uHHHH</code>	Matches the Unicode character code as expressed by the hexadecimal value HHHH .

Required Escaped Characters

The following characters have special meaning within a regular expression.

```
. ^ $ * + - ? ( ) [ ] { } \ | - /
```

To reference the literal character, you must escape it within the regular expression, as in:

```
/\./
```