Splunk® Add-on Builder Splunk Add-on Builder User Guide 2.2.0

Extract fields

Generated: 5/29/2018 4:53 am
Extract fields

In Extract Fields, parse the data in your source types to create field extractions. The source types you created in the Configure Data Collection section or imported from splunk using the Manage source type appear in source type list.

Splunk Add-on Builder provides you three ways to build the field extractions.

- Assisted extraction. Splunk Add-on Builder will detect the format of the data and provides you the recommended regex to parse your data.
- Manual extraction. Configure the field extraction manually in Splunk platform.
- Manual transformation. Configure the field transformation manually in Splunk platform.

If you used assisted extraction before and want to switch to manual extraction or transformation, you need to delete the assisted extraction first or otherwise the manual mode will be disabled.

Assisted extraction

If you choose Assisted extraction, Add-on builder will try to parse your data using the suggested format, the format includes:

- Unstructured Data. Typically used for log files.
- Table. Data in tabular formats, such as comma-separated values (CSV) and tab-separated values (TSV).
- Key Value. Data that contains key-value pairs.
- JSON. Data in the JavaScript Object Notation (JSON) format.
- XML. Data in the Extensible Markup Language (XML) format.

To parse the data for a sourcetype and extract fields using assisted extractions
1. On your add-on homepage, click **Extract Fields** on the Add-on Builder navigation bar.
2. On the **Extract Fields** page, select a source type to parse by clicking **Assisted Extraction**.
3. From **Choose Data Format**, select the format of the data.

A format type is automatically selected if it has been detected. You can change it as needed. If you aren't sure and a format type has not been automatically selected, try "Unstructured Data".

4. Click **Submit**.

After parsing the data, the results are displayed on a summary page. Depending on the data format, you can adjust the way fields are extracted.

5. If you are satisfied with the results, click **Save**.
6. If you want to try parsing again using a different format, click **Cancel** to return to the previous page.

Once data for a source type has been parsed, the source type is added to the table on the **Extract Fields** page.

**Unstructured Data**

The Add-on Builder's Field Extractor displays a selection of events in groups, along with the fields that were extracted.

Here are some of the actions you can perform:
• Select one or more groups that best represent the data.
• If you are familiar with creating regular expressions, you can display the regular expression that the Field Extractor used, and modify it to improve the field extraction.
• Click on individual field names to include or exclude the field for extraction.
• Click the Edit icon next to a field name to edit its name.
• Click the Trash icon next to a field name to remove its capture group from the regular expression.

Table

The Table format is used with tabular data.

Here are some of the actions you can perform:

• Change how data is parsed by selecting the delimiter character that is used to separate fields. To specify a different character, click Other and enter the character.
• Change the field names, but only after you have selected the correct delimiter. Each time you change delimiters, the number of columns could change and you might lose any changes to field names.

Key Value

The Key Value format is used with data containing key-value pairs.
Here are some of the actions you can perform:

- Change how data is parsed. For **Extraction Methods**, select:
  - **Auto** to let the Add-on Builder parse data automatically.
  - **Delimiters** to use delimiters.
  - **Regex** to use regular expressions.

- For **Delimiters**, select the delimiters for the key-value pairs:
  - Specify the pair delimiter character, which is used to separate key-value pairs.
    Using the example `key_a=value_a, key_b=value_b`, the correct character is a comma.
  - Specify the key-value delimiter character, which is used to separate keys and values.
    Using the example `key_a=value_a, key_b=value_b`, the correct character is an equals sign.

- For **Regex**: select the regular expression to use, or create your own.

**JSON**

The JSON format is used with JSON data. There are no additional parsing options.
The XML format is used with XML data. There are no additional parsing options.

**Manual extraction**

If you choose **Manual extraction**, Add-on builder will direct you to the **Field extractions** page of Splunk platform.

1. On your add-on homepage, click **Extract Fields** on the Add-on Builder navigation bar.
2. On the **Extract Fields** page, select a source type to parse by clicking **Manual Extraction**.
3. Add-on builder will direct you to the **field extractions** page of Splunk.
Manual transformation

If you choose **Manual transformation**, Add-on builder will direct you to the **Field transformation** page of Splunk platform.

1. On your add-on homepage, click **Extract Fields** on the Add-on Builder navigation bar.
2. On the **Extract Fields** page, select a source type to parse by clicking **Manual transformation**.
3. Add-on builder will direct you to the **field transformation** page of Splunk platform. See more on Use the Field transformation page in the Splunk Enterprise *Knowledge Manager Manual*.

Troubleshooting

**What if I need to upload different sample data?**

If you decide that you need to upload a different sample data file for a source type, for example you want to clean the data first, go to Manage source types, delete the sample data, then upload additional data files.

**A regular expression had too many capture groups, what do I do?**

This error is displayed after attempting to parse a file, and the regular expression created by the Field Extractor contains more than 100 capture groups (fields).

This error might indicate a problem with the Event Break setting for the source type:

1. Go to Manage source types.
2. Edit the source type and select a different option for Event Break.
3. Upload the sample events again. Because the Event Break option is applied when indexing the data, changing this value does not affect events that have already been indexed.
4. Parse the data again.

The sample data might contain an event that is too long:

1. Edit the sample data file by splitting the long lines to clean up the data.
2. Go back to Manage source types.
3. Upload the sample events again.
4. Parse the data again.

**Why are the field names not detected in my tabular data?**

The Add-on Builder uses the first 1000 events for field extraction. If your data contains more than 1000 events, the parser cannot automatically detect the field names.

The parser assumes that all entries except the table header contain a timestamp. If entries in your tabular data do not contain a timestamp, the parser will not correctly detect which entry is the table header.

**Learn more**

For more information, see the following Splunk Enterprise documentation:

- About fields in the *Knowledge Manager Manual*
- Build field extractions with the field extractor in the *Knowledge Manager Manual*
- Field Extractor: Select Fields step in the *Knowledge Manager Manual*