Creating Field Extractions – Lab Guide

Overview

Welcome to the Splunk Education lab environment. These lab exercises will have you create Regular Expression (regex) field extractions and delimited field extractions.

Scenario

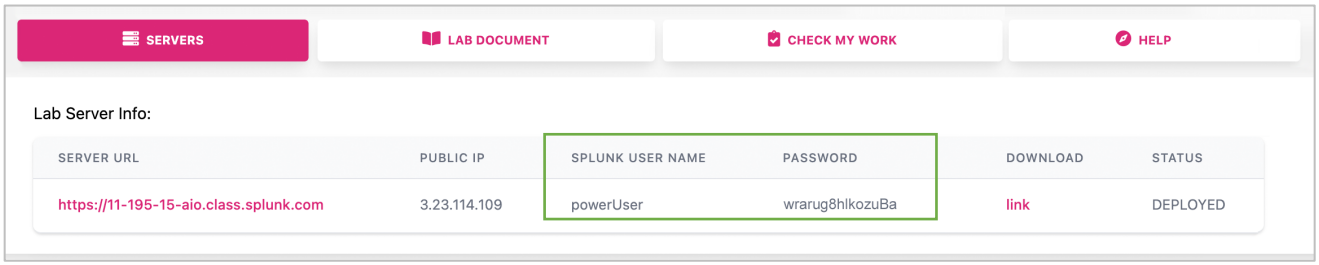
You will use data from the international video game company, Buttercup Games. A list of source types is provided below.

NOTE: This is a lab environment driven by data generators with obvious limitations. This is not a production environment. Screenshots approximate what you should see, not the **exact** output.

Index	Type	Sourcetype	Interesting Fields
web	Online sales	**access_combined**	action, bytes, categoryId, clientip, itemId, JSESSIONID, price, productId, product_name, referer, referer_domain, sale_price, status, user, useragent
games	Game logs	**SimCubeBeta**	date_hour, date_mday, date_minute, date_month, date_second, data_wday, data_year, date_zone, eventtype, index, linecount, punct, splunk_server, timeendpos, timestartpos

Lab Connection Info

Access labs using the server URL, user name, and password shown in your lab environment.



splunk>

Common Commands and Functions

These commands and statistical functions are commonly used in searches but may not have been explicitly discussed in the module. Please use this table for quick reference. Click on the hyperlinked SPL to be taken to the Search Manual for that command or function.

SPL	Type	Description	Example
sort	command	Sorts results in descending or ascending order by a specified field. Can limit results to a specific number.	*Sort the first 100 `src_ip` values in descending order* `| sort 100 -src_ip`
where	command	Filters search results using eval-expressions.	*Return events with a `count` value greater than 30* `| where count > 30`
rename	command	Renames one or more fields.	*Rename `SESSIONID` to 'The session ID'* `| rename SESSIONID as "The session ID"`
fields	command	Keeps (+) or removes (-) fields from search results.	*Remove the `host` field from the results* `| fields - host`
stats	command	Calculates aggregate statistics over the results set.	*Calculate the total sales, i.e. the sum of `price` values* `| stats sum(price)`
eval	command	Calculates an expression and puts the resulting value into a new or existing field.	*Concatenate `first_name` and `last_name` values with a space to create a field called "`full_name`"* `| eval full_name=first_name." ".last_name`
table	command	Returns a table.	*Output `vendorCountry`, `vendor`, and `sales` values to a table* `| table vendorCountry, vendor, sales`
sum()	statistical function	Returns the sum of the values of a field. Can be used with **`stats`**, **`timechart`**, and **`chart`** commands.	*Calculate the sum of the `bytes` field* `| stats sum(bytes)`
count or count()	statistical function	Returns the number of occurrences of all events or a specific field. Can be used with **`stats`**, **`timechart`**, and **`chart`** commands.	*Count all events as "events" and count all events that contain a value for `action` as "action"* `| stats count as events, count(action) as action`

Refer to the Search Reference Manual for a full list of commands and functions.

Lab Exercise 1 – Create Regex Field Extractions

Description

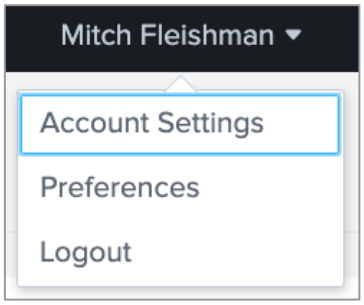
Configure the lab environment user account. Then, create field extractions based on Regular Expression (regex).

Steps

Task 1: Log into Splunk and change the account name and time zone.

Set up your lab environment to fit your time zone. This also allows the instructor to track your progress and assist you if necessary.

1. Log into your Splunk lab environment using the username and password provided to you.
2. You may see a pop-up window welcoming you to the lab environment. You can click **Continue to Tour** but this is not required. Click **Skip** to dismiss the window.
3. Click on the username you logged in with (at the top of the screen) and then choose **Account Settings** from the drop-down menu.
4. In the **Full name** box, enter your first and last name.
5. Click **Save**.
6. Reload your browser to reflect the recent changes to the interface. (This area of the web interface will be referred to as ***user name***.)



After you complete step 6, you will see your name in the web interface.

NOTE: Sometimes there can be delays in executing an action like saving in the UI or returning results of a search. If you are experiencing a delay, please allow the UI a few minutes to execute your action.

7. Navigate to ***user name*** **> Preferences.**
8. Choose your local time zone from the **Time zone** drop-down menu.
9. Click **Apply**.
10. (Optional) Navigate to ***user name*** **> Preferences > SPL Editor > Search auto-format** and click on the toggle to activate auto-formatting. Then click **Apply**. When the pipe character is used in search, the SPL Editor will automatically begin the pipe on a new line.

Search auto-format disabled (default)

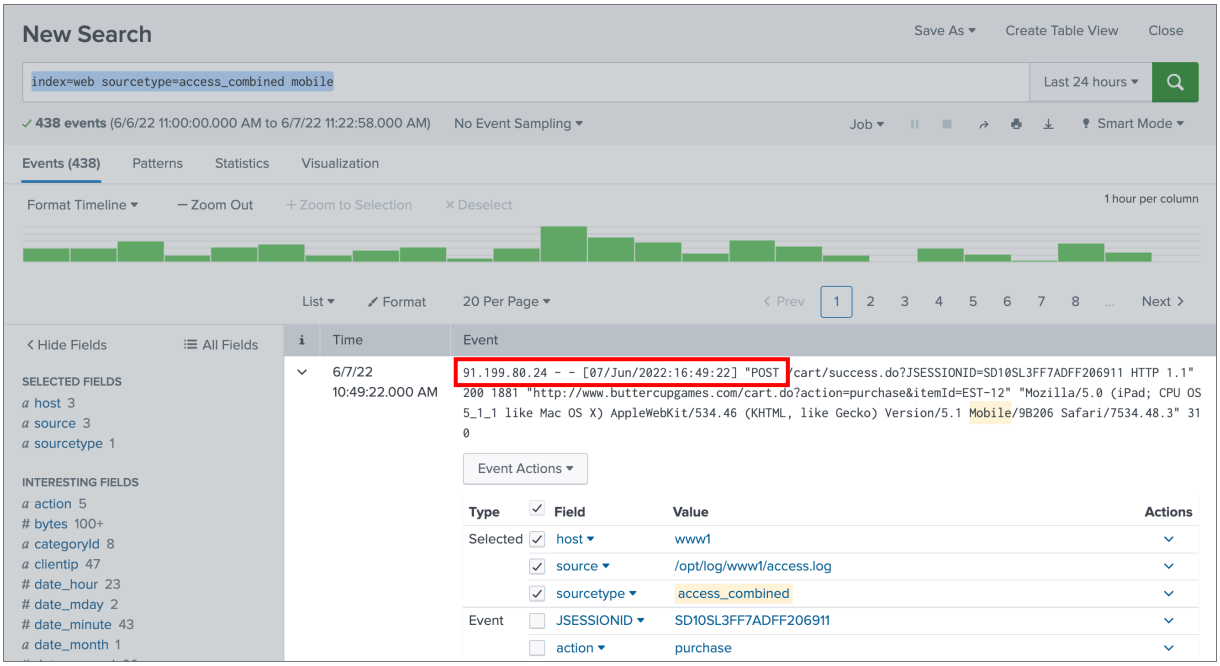
Search auto-format enabled

Task 2: Use the Field Extractor (FX) to extract the IP address and transaction ID fields using the Regular Expression method.

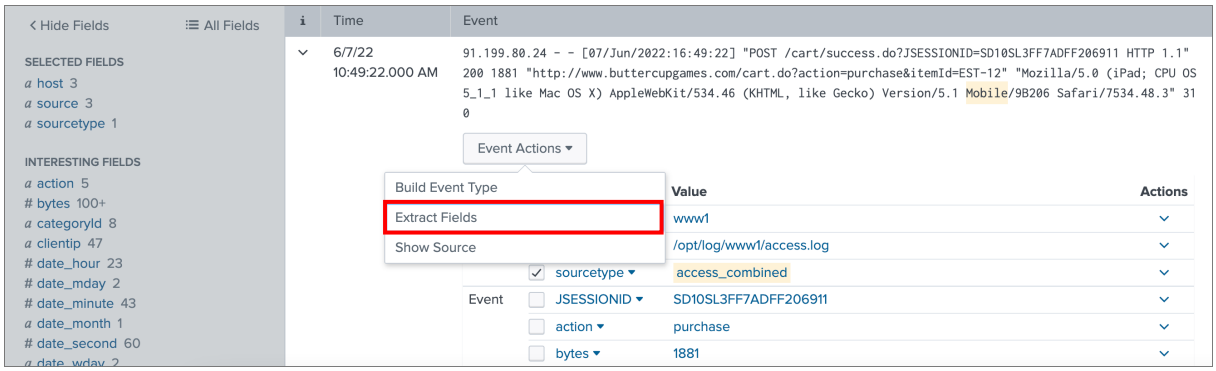
11. Navigate to the search app by clicking **Apps > Search & Reporting** in the upper left corner of Splunk Web.
    a. If a welcome message appears, click **Skip**.
12. Use this search to search the **web** index for all events in the **last 24 hours** for the `access_combined` sourcetype that contain the keyword **`mobile`**:

    ```
    index=web sourcetype=access_combined mobile
    ```

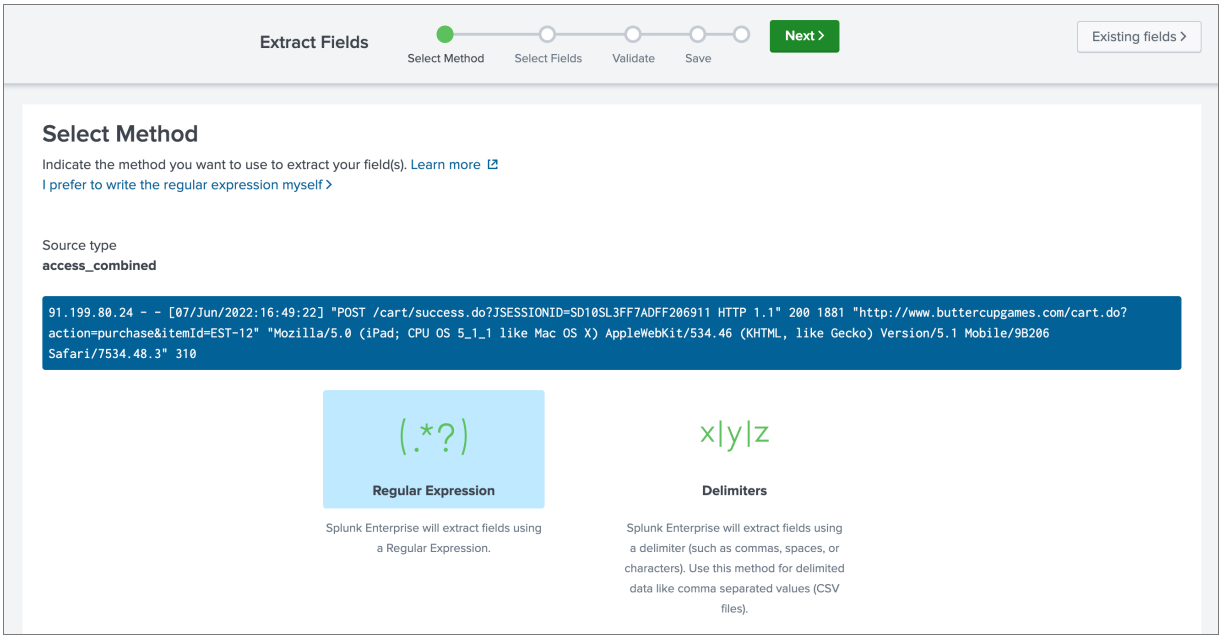
13. View the event details to see all the extracted fields, by clicking the arrow (**>**) under the information icon (**`i`**) in the first event that contains an IP address value, a timestamp, and either a GET or POST operation



14. Click **Event Actions > Extract Fields**.



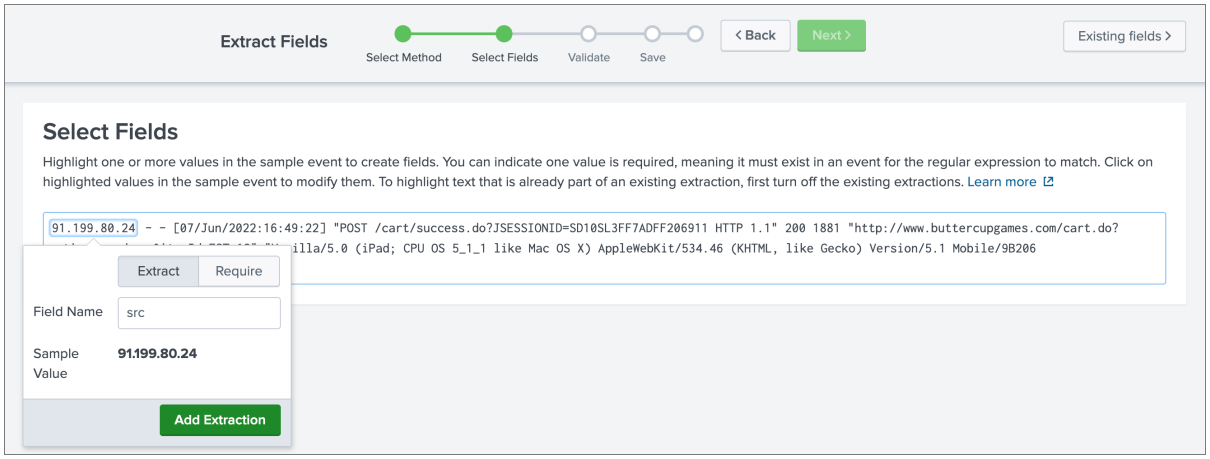
15. Configure a regular expression extraction method.
    a. Click on the **Regular Expression** selection.
    b. Click **Next** to begin selecting fields.



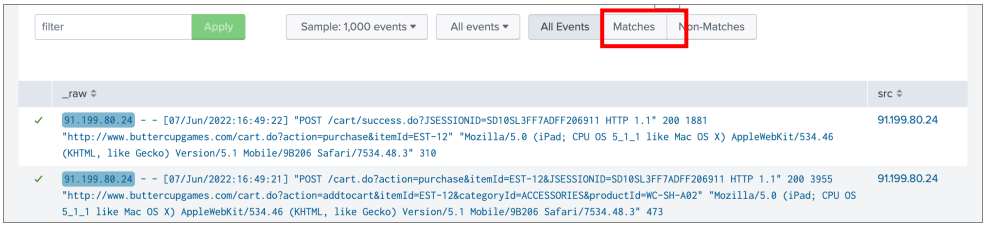
16. Add an extracted field for the IP address value.
    a. Highlight the IP address value, which appears as the very first field, in the sample event.
    b. In the **Field Name** field, type: **`src`**

> **NOTE:** To view the pop-up box with the **Field Name** field, you may need to click on the highlighted IP address.

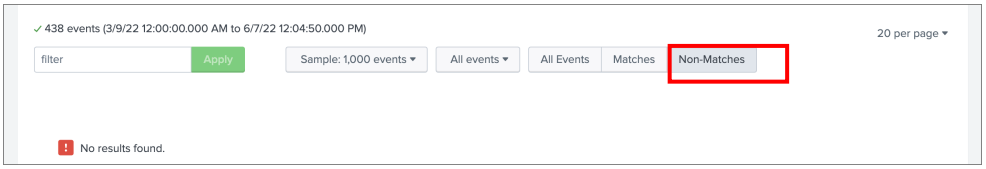
    c. Click **Add Extraction**.



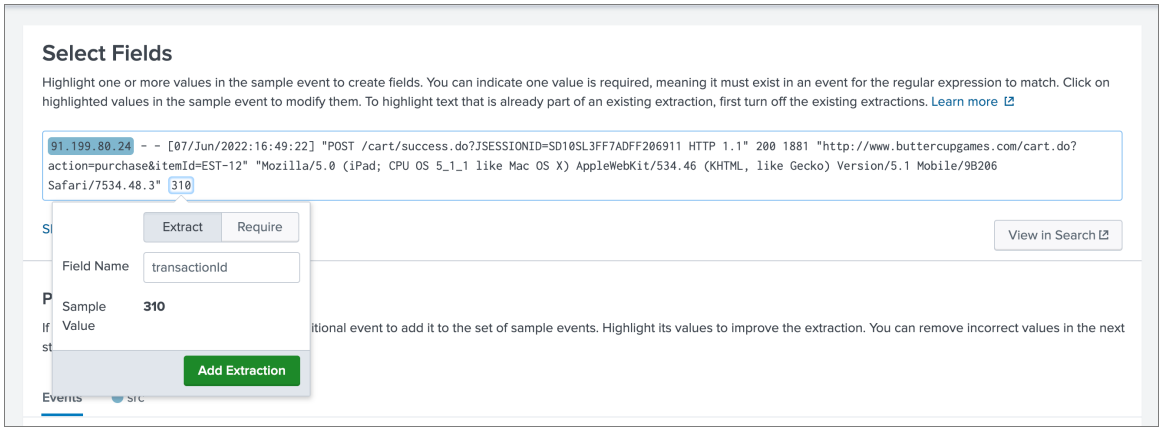
17. Scroll down to the **Preview** section and verify that the correct information is being extracted.
    a. Click **Matches** to verify results.



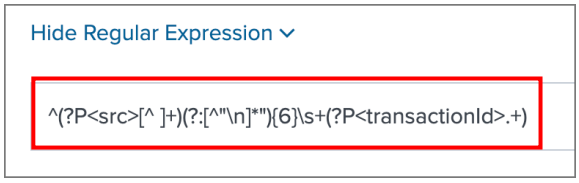
    b. Click **Non-Matches** to verify any incorrect results. You should see “No results found.”



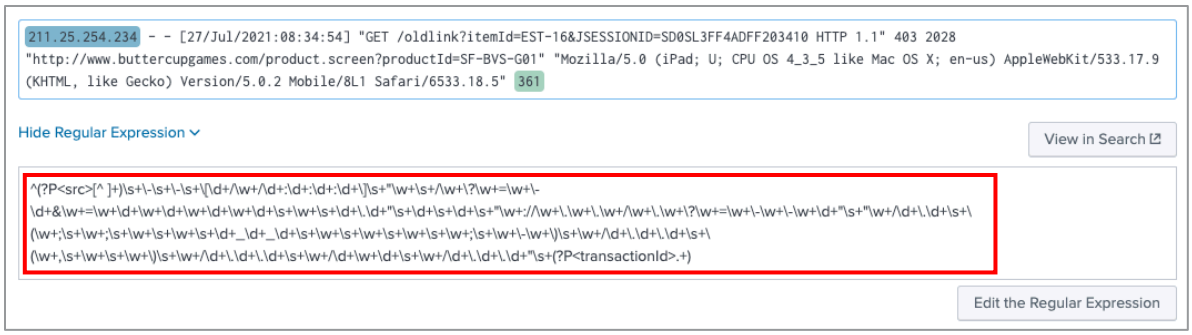
    c. Click **All Events**.
18. Back at the top of the screen under **Select Fields**, highlight the last number in the event.
    a. In the **Field Name** box, type: **`transactionId`**
    b. Click **Add Extraction**.



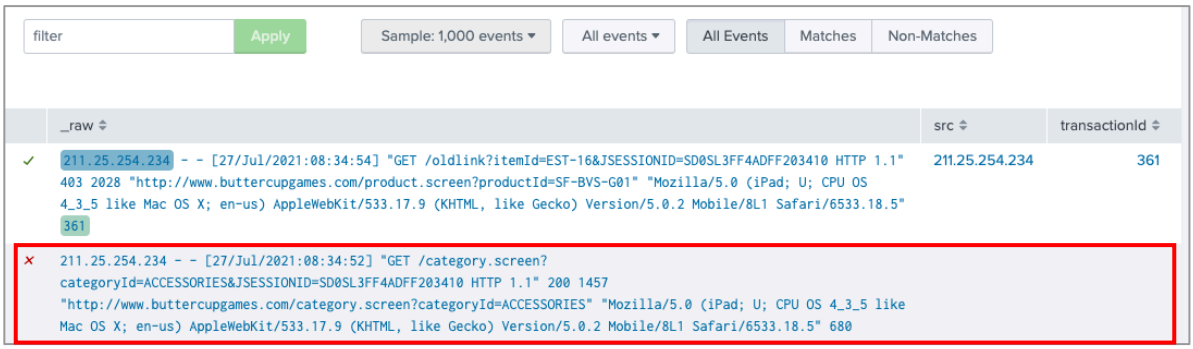
19. Click on the **Show Regular Expression >** link to review the regex syntax.
    a. If the regular expression matches all events, it should look relatively simple in structure and looks similar to the highlighted expression below. **If this is the case, move on to step 20**.



b. In some cases, the regular expression does not match all events, resulting in a more complex and inaccurate regular expression:



c. Scroll down to the Preview section and verify that not all events are matching correctly.



d. Add a missing sample event by hovering over an event that shows the missing red "**x**" symbol to the left and click **+ Add sample event**.

e. In the newly added sample event, which now appears towards the top of the screen, highlight the IP address value.
   - In the **Field Name** box, under the **Select a Field** drop-down list, select **`src`**.
   - Click **Add Extraction**.

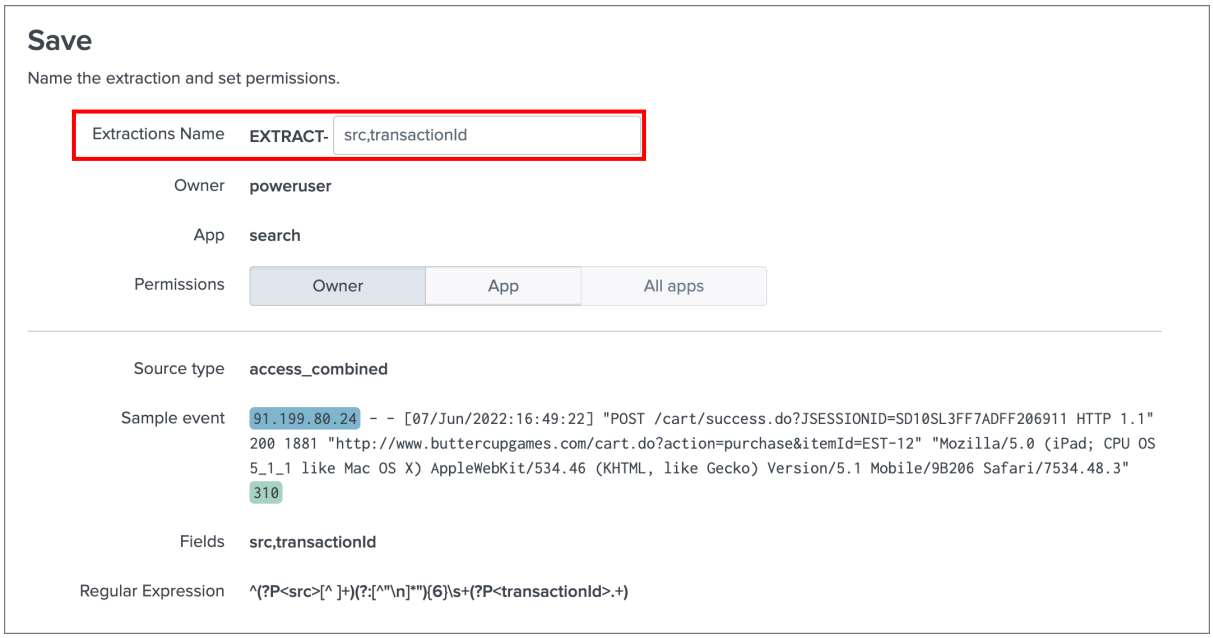
f. In the second sample event, highlight the last number.
   - Verify that the **Field Name** lists transactionId.
   - Click **Add Extraction**.

20. Scroll down to the **Preview** section and verify that the correct information is being extracted.
    a. Click **Matches** to verify results.
    b. Click **Non-Matches** to verify that it returns with “No results found”.

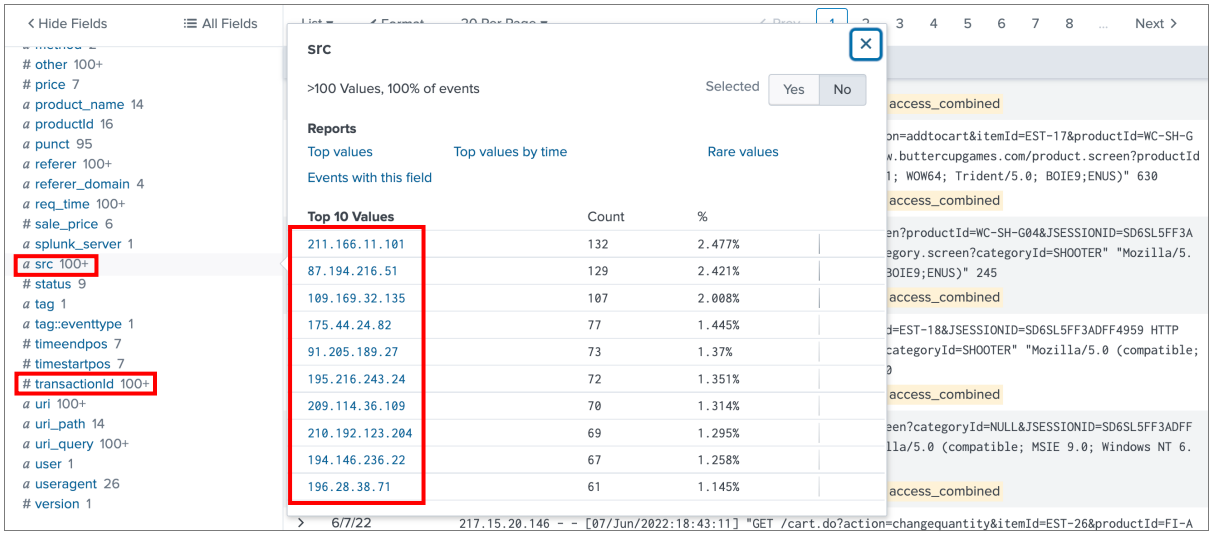
NOTE: If there are any results under **Non-Matches**, repeat steps 19c to 19f, until there are no results under **Non-Matches**.

    c. Click **Next** to proceed to validate the extracted fields.

21. In the **Validate** step, click on the **Show Regular Expression >** link (if it is hidden) and look at the regex syntax. Then click **Next.**
22. Review the **Extractions Name** and click **Finish**.

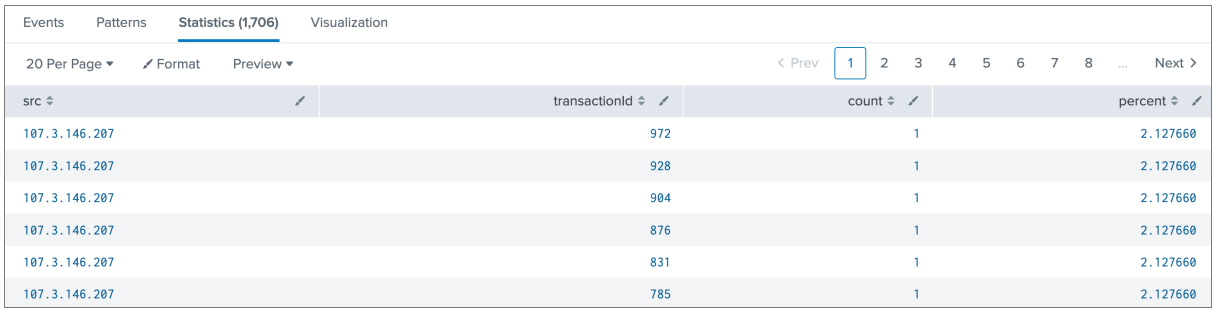


23. On the final **Success!** screen, click on **Explore the fields I just created in Search**.
    a. In the Fields sidebar, under **Interesting Fields**, verify that your newly added fields appear (**src** and **transactionId**).
    b. Click on the **src** field to explore the values.

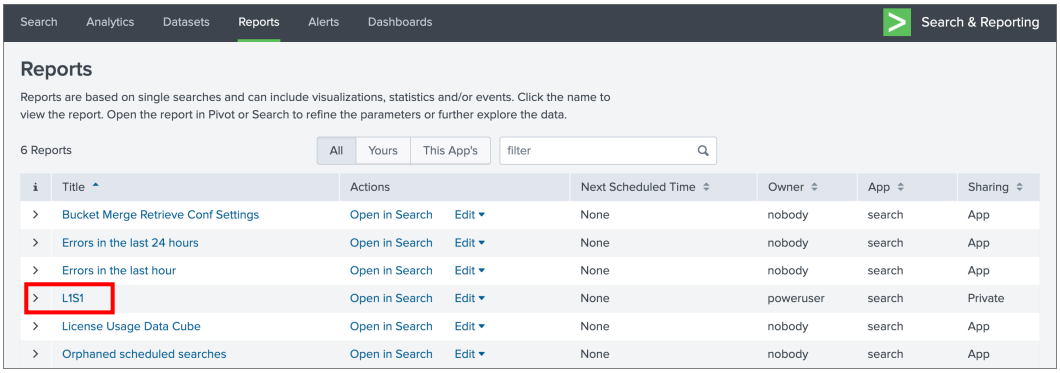


24. Search for events in the `access_combined` sourcetype in the **last 24 hours**. List the top `transactionId`'s by IP address using this search:

```
index=web sourcetype=access_combined | top transactionId by src
```



25. Save your search as a report with the name **L1S1**.
    a. Click **Save As** > **Report**
    b. For **Title**, enter **L1S1**.
    c. **Save**.
    d. You can **View** your report or exit out of the **Your Report Has Been Created** window by clicking the **X** in the upper-right corner.
    e. You can access your saved reports using the **Reports** tab in the application bar.



*Your recently saved **L1S1** report will be visible in the **Reports** tab.*

Lab Exercise 2 – Create Delimited Field Extractions

Description

This lab exercise walks you through the process of creating delimited field extractions. Additionally, field extractions based on Regular Expression (regex) are used to capture additional fields.

Steps

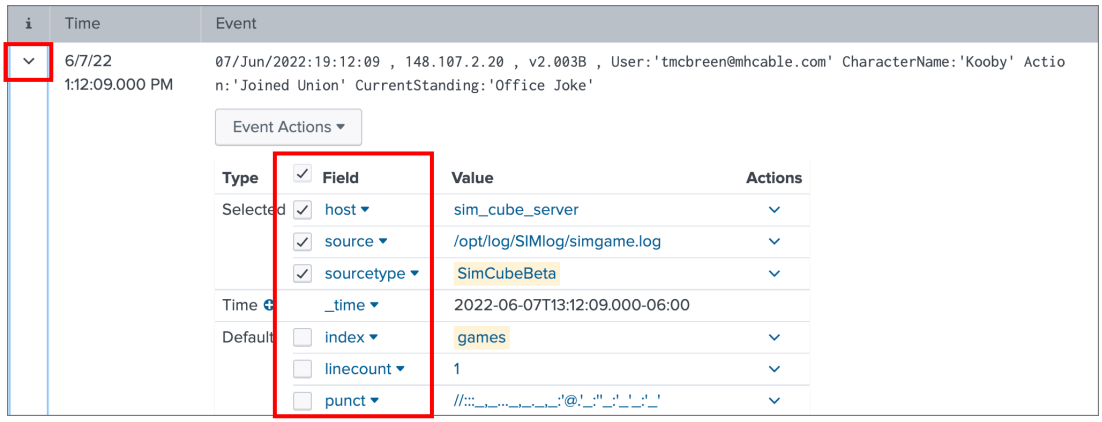
Scenario: The engineering team launched the beta of a new game called SimCube. To make improvements to the game, engineers want to see how users are playing the game. However, the log file doesn't contain headers and the fields are not automatically extracted.

Task 1: Use the Field Extractor (FX) to extract fields using the delimiters method.

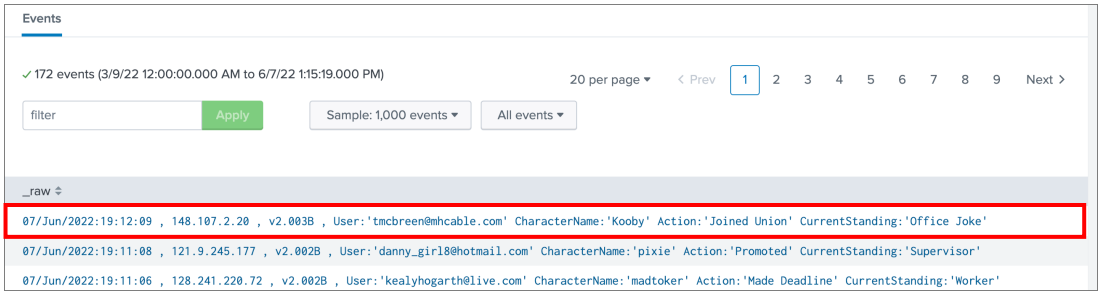
1. Re-initialize the search window by clicking **Search** in the application bar.
2. Search for all events in the **last 30 days** for the `SimCubeBeta` sourcetype in the `games` index using this search:

   ```
   index=games sourcetype=SimCubeBeta
   ```

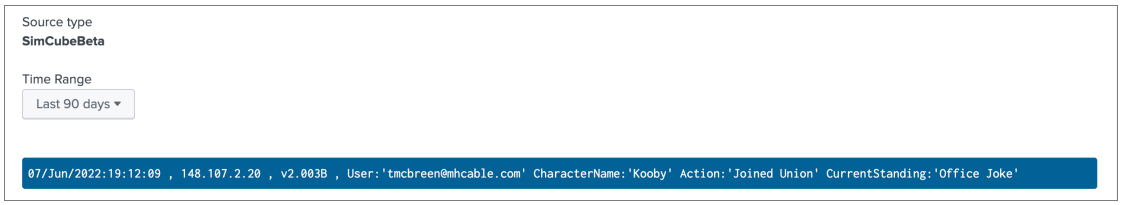
   a. Click the arrow (**>**) under the information icon (**i**) in the first event to see which fields are extracted.



3. In the Fields sidebar, underneath the **Interesting Fields** section, click **+ Extract New Fields**.
   a. On the **Select Sample Event** screen, click the first event to select it as a sample event.



b. Verify that the event now appears towards the top as a sample event.



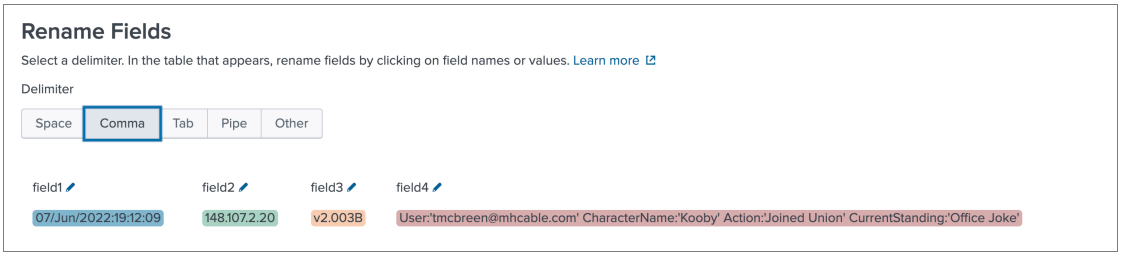
NOTE: The first few fields for this sourcetype appear to be separated with a comma symbol (`,`). This includes a timestamp, an IP address (in IP version 4 format), a software version (for example "`v2.002B`"), and then some miscellaneous information.

The miscellaneous information at the end of the event appears to be separated by spaces, which includes fields such as the **User**, **CharacterName**, **Action**, and **CurrentStanding**.

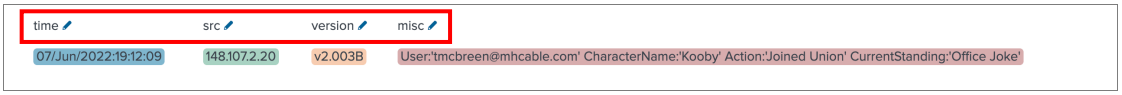
c. Click the **Next** button at the top of the screen.

4. Select the **Delimiters** method and click **Next**.

a. For the **Delimiter** type, select **Comma**.



b. Rename all the fields as follows (in this order):
   - field1 > **time**
   - field2 > **src**
   - field3 > **version**
   - field4 > **misc**

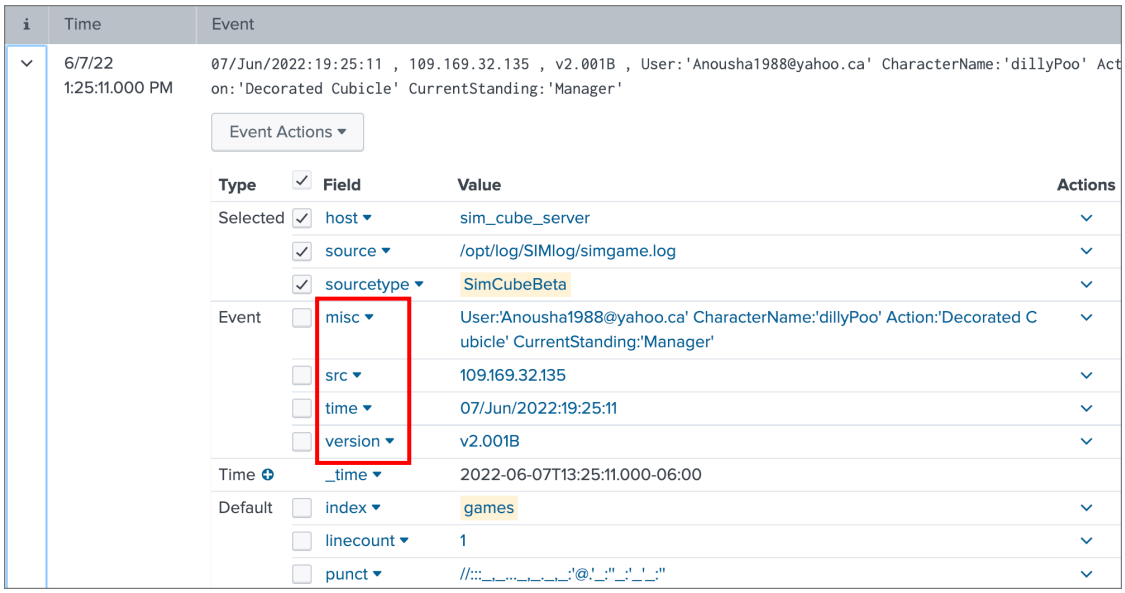


NOTE: For the time being, the **misc** field contains all the information at the end of the event. Those fields are separated by spaces. You will create individual fields for this data using a regular expression later in this lab.

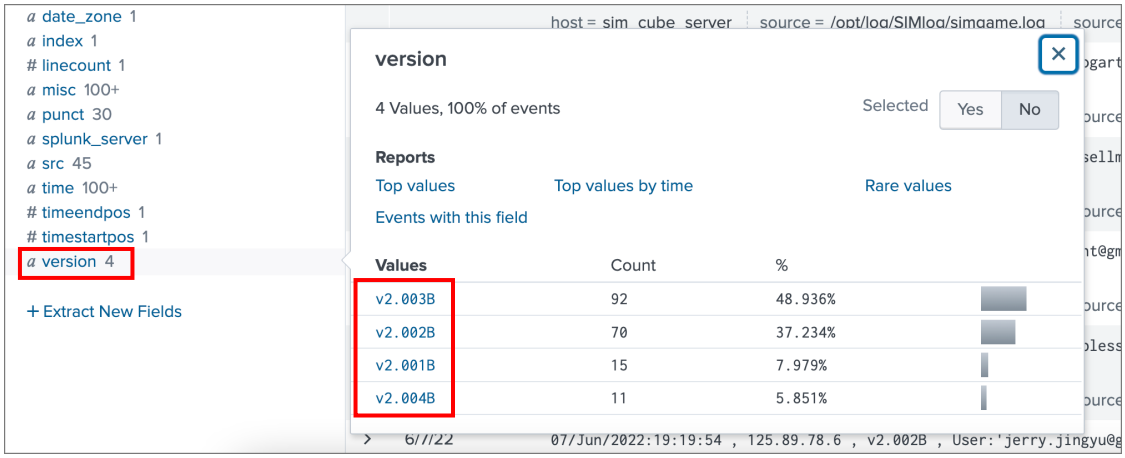
c. After all the fields are renamed, click **Next**.

5. For Extractions Name, enter **simgame_log** and click **Finish>**.

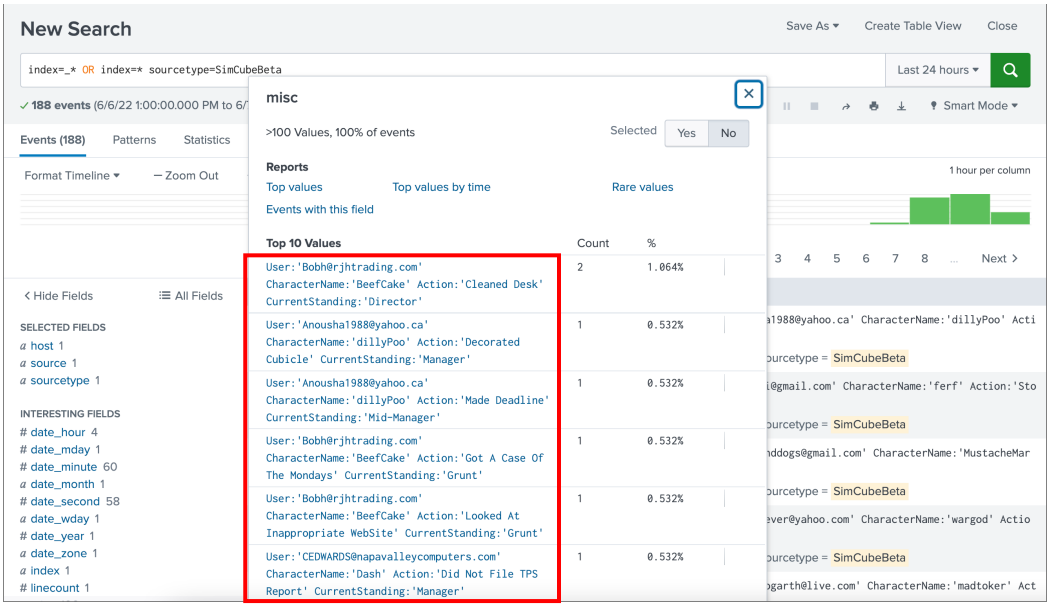
6. On the final **Success!** screen, click on **Explore the fields I just created in Search**. (Alternatively, return to the search app and repeat the search you performed in step 2.)
    a. In the Fields sidebar, under **Interesting Fields**, verify that your newly added fields appear (**misc**, **src**, **time**, and **version**).



    b. Click on the **version** field to explore the values.



    c. Click on the **misc** field to explore the values. Notice that the **misc** field consists of a long string with multiple game-related values that you may wish to consider as separate fields.

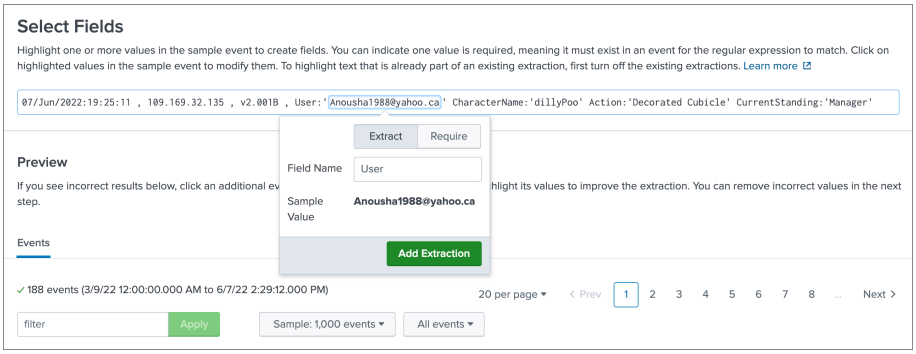


Use the regex method to extract additional fields found in the new misc field.

7. In the Fields sidebar, underneath the **Interesting Fields** section, click **+ Extract New Fields**.
    a. On the **Select Sample Event** screen, click the first event to select it as a sample event.
    b. Verify that the event now appears towards the top as a sample event.
    c. Click the **Next** button at the top of the screen.
8. Select the **Regular Expression** method and click **Next**.
9. Add extracted fields for the string data that we previously added to the **misc** field.
    a. Highlight the name of the user after the word “User:” as it appears between the single quotes. Do not include the single quotes in the highlight selection.

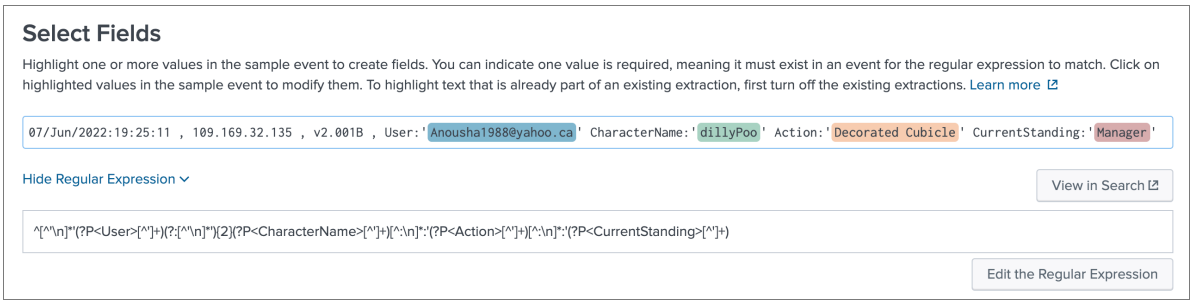
NOTE: Be sure to capture all the characters **between** the single quotes, but **not** the single quotes themselves. Some versions of Internet Explorer won’t allow you to exclude the single quotes. Switch to another browser to complete the exercise.

    b. Give it a **Field Name** of **User**.
    c. Click **Add Extraction**.

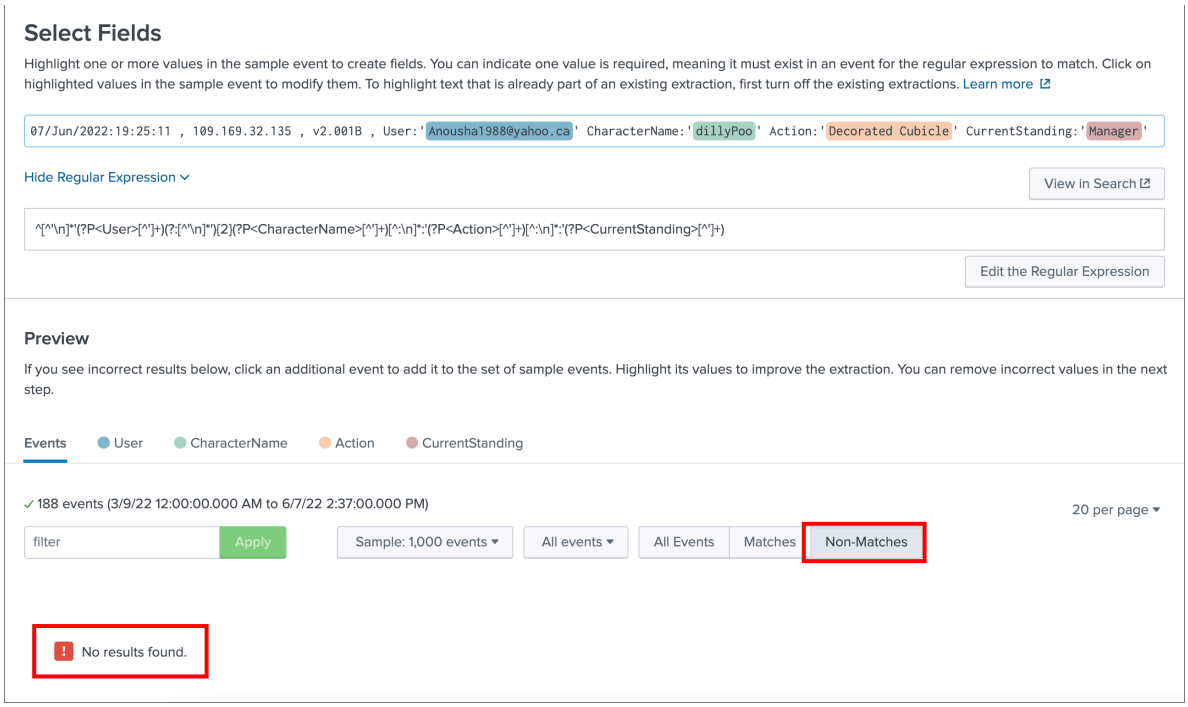


d. Repeat steps a-c to create fields for **CharacterName**, **Action**, and **CurrentStanding** fields.

e. Click on the **Show Regular Expression** link to view the regex expression.



10. While still on the **Select fields** step (before the validation stage), click on **Non-Matches** to see whether any relevant events are being excluded. (If no events display when you click **Non-Matches**, proceed to step 11.)



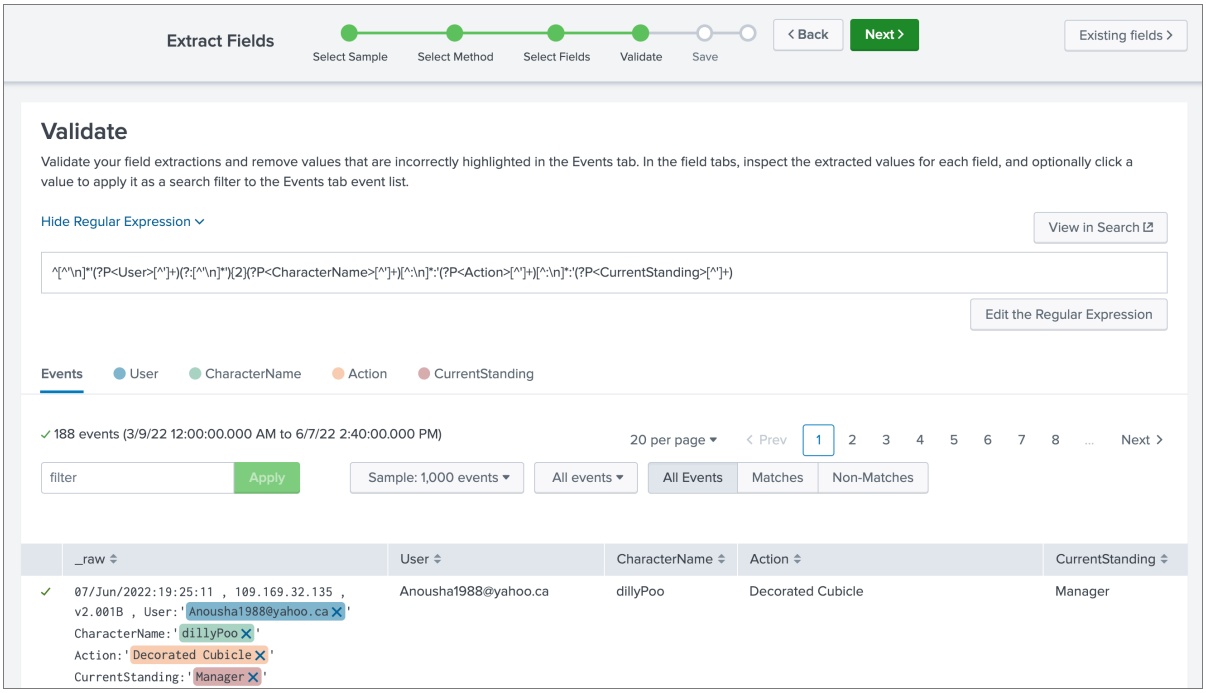
a. If Non-Matches appear, hover your cursor over any listed event that you want to include and click **+ Add sample event**.

b. Highlight each relevant value in the sample event and click **Select a Field**. For each value, choose the field name you want associated with that value and click **Add Extraction**.

c. Repeat step 10a-10b for each excluded event until there are no more **Non-Matches**.

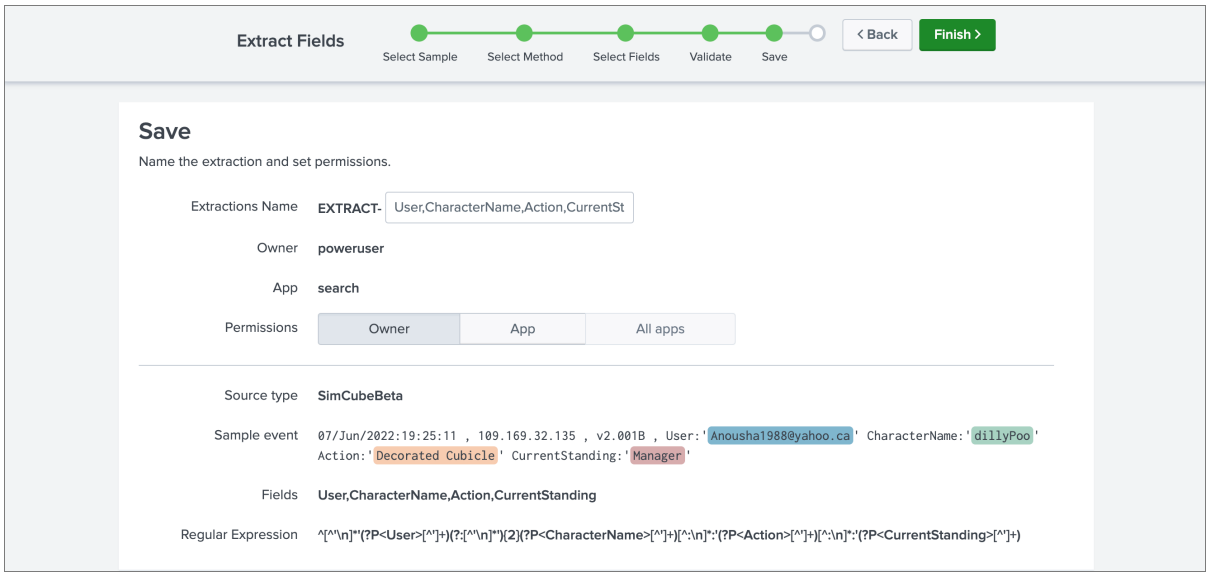
d. Click **Next** to proceed to the **Validate** step.

11. On the **Validate** step, verify that the results below look correct by capturing all four fields for each event. When you're satisfied with your results, click **Next**.

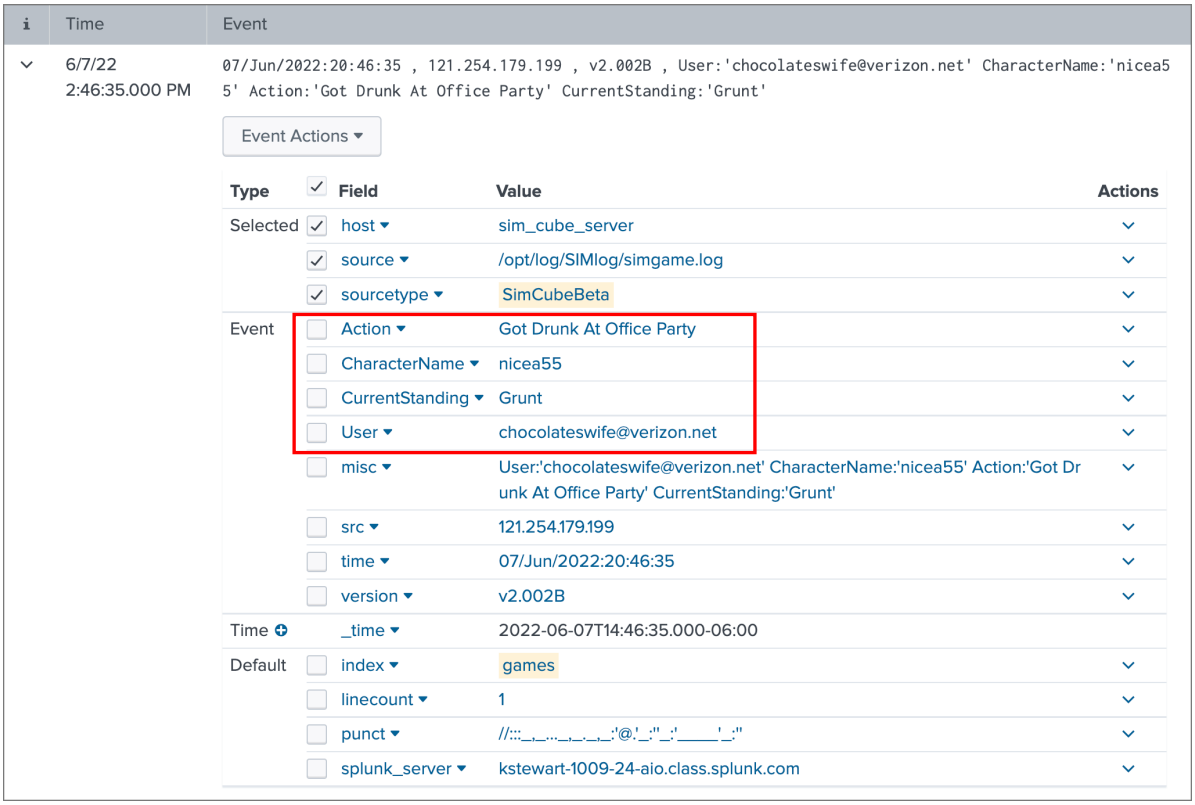


NOTE: Be sure to thoroughly check your results. It's important to ensure you've captured all characters inside the single quotes for the fields you've extracted.

12. Accept the prefilled **Extractions Name** and click **Finish>** to save.



13. Click on the **Explore the fields I just created in Search** link on the **Success!** page.
    a. Click the arrow (**>**) under the information icon (**i**) in the first event to see which fields are extracted.
    b. Verify that in addition to the delimited fields we created earlier (**misc**, **src**, **time**, **version**), you also see the newly created fields using regular expressions (**Action**, **CharacterName**, **CurrentStanding**, **User**).



NOTE: It may take a few minutes before the newly extracted fields appear in the search.