

# Leveraging Lookups and Subsearches – Lab Guide

### Overview

Welcome to the Splunk Education lab environment. These lab exercises will test your knowledge of lookup commands and subsearches.

### 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
security	Active Directory	winauthentication_security	LogName, SourceName, EventCode, EventType, User
	Badge reader	history_access	Address_Description, Department, Device, Email, Event_Description, First_Name, last_Name, Rfid, Username
	Web server	linux_secure	<pre>action, app, dest, process, src_ip, src_port, user, vendor_action</pre>
sales	Retail sales	vendor_sales	<pre>categoryId, product_name, productId, sale_price, Vendor, VendorCity, VendorCountry, VendorID, VendorStateProvince</pre>
network	Email security data	cisco_esa	dcid, icid, mailfrom, mailto, mid
	Web security appliance data	cisco_wsa_squid	<pre>action, cs_method, cs_mime_type, cs_url, cs_username, sc_bytes, sc_http_status, sc_result_code, severity, src_ip, status, url, usage, x_mcafee_virus_name, x_wbrs_score, x_webcat_code_abbr</pre>
	Firewall data	cisco_firewall	<pre>bcg_ip, dept, Duration, fname, IP, lname, location, rfid, splunk_role, splunk_server, Username</pre>



# **Common Commands and Functions**

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

SPL	Туре	Description	Example	
<u>sort</u>	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	
<u>where</u>	command	Filters search results using eval-expressions.	Return events with a count value greater than 30    where count > 30	
<u>rename</u>	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	
<u>stats</u>	command	Calculates aggregate statistics over the results set.	Calculate the total sales, i.e. the sum of price values.    stats sum(price)	
<u>eval</u>	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	
<u>table</u>	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)	
<pre>count or count()</pre>	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 – Using Lookup Commands

# Description

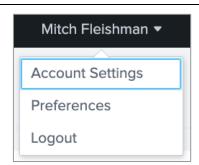
Configure the lab environment user account. Then, use **inputlookup**, **lookup**, and **outputlookup** commands to call on and create lookups in search.

# 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 allows the instructor to track your progress and assist you if necessary.

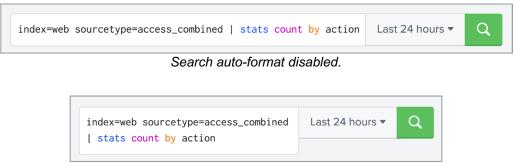
- 1. Log into your Splunk lab environment using the username and password provided to you.
- 2. You may see a pop-up window welcoming you to the lab environment. You can click **Continue to Tour** but this is not required. Click **Skip** to dismiss the window.
- 3. Click on the username you logged in 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 is a delay in executing an action like saving in the UI or returning results of a search. 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 enabled.



Scenario: You provided your knowledge manager with a CSV containing HTTP statuses, status descriptions, and status types. Your knowledge manager just informed you that the lookup

was uploaded.

# Task 2: Verify that a lookup has been uploaded correctly.

11. Your lab environment is configured to take you to the Search & Reporting app within Splunk. (Also called the "search" app.) Confirm you are in the correct app by clicking Apps in the top left corner. You should see Search & Reporting highlighted. If you do not, click on Search & Reporting.

12. Your knowledge manager provided you with the following information about the status\_definitions.csv lookup. Use the inputlookup command to verify that the file-based lookup has been correctly uploaded.

filename: status definitions.csv

definition name: status definitions lookup

lookup type: file-based

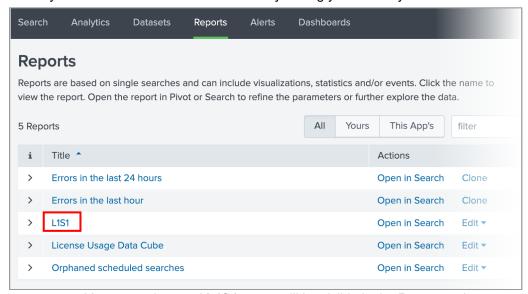
13. Confirm that your search output matches the lookup contents:

status 🗢 🖊	status_description \$	status_type \$
100	Continue	Informational
101	Switching Protocols	Informational
200	OK	Successful
201	Created	Successful
202	Accepted	Successful
203	Non-Authoritative Information	Successful
204	No Content	Successful
205	Reset Content	Successful
206	Partial Content	Successful
300	Multiple Choices	Redirection
301	Moved Permanently	Redirection
302	Found	Redirection
303	See Other	Redirection
304	Not Modified	Redirection
305	Use Proxy	Redirection
307	Temporary Redirect	Redirection
400	Bad Request	Client Error
401	Unauthorized	Client Error
402	Payment Required	Client Error
403	Forbidden	Client Error

- 14. Save your search as a report with the name **L1S1**.
  - 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.
- f. Re-initialize the search window by clicking **Search** in the application bar. Perform this step each time you save a search to avoid accidentally editing your recently saved search.



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

# Task 3: Use the status\_definitions.csv lookup in a search.

15. The following search needs to find events from the online sales data that do not have a **status** of **200**. (This represents unsuccessful events and is written as **status!=200**.) However, the search is missing the **lookup** command.

```
index=web sourcetype=access_* status!=200
| stats count by host, status_description, status_type
```

- a. Run the search above over the Last 24 hours. You should receive an error or No results found.
- b. Use the **lookup** command to add the **status\_description** and **status\_type** fields from the **status\_definitions\_lookup**. Then, run the search again.



host \$	1	status_description \$	1	status_type \$	1	count \$ /
www1		Bad Request		Client Error		21
www1		HTTP Version Not Supported		Server Error		31
www1		Internal Server Error		Server Error		30
www1		Not Acceptable		Client Error		25
www1		Not Found		Client Error		21
www1		Request Timeout		Client Error		30
www1		Service Unavailable		Server Error		42
www2		Bad Request		Client Error		24
www2		Forbidden		Client Error		39
www2		Internal Server Error				

16. Save your search as a report with the name **L1S2**.

Scenario: SecOps wants a report of known users who have been browsing "Uncategorized URLs" over the last 24 hours.

- Task 4: Use information from the knownusers.csv and status\_definitions.csv lookups to complete a search that will generate a report of users who have accessed uncategorized URLs over the last 24 hours. The report should include the users' departments, the URL accessed, and the associated http status and description.
- 17. This task uses information from the status definitions.csv lookup and the knownusers.csv lookup. Your knowledge manager has provided you with the following information about the knownusers.csv lookup. Use the inputlookup command to explore the knownusers.csv file-based lookup.

filename: knownusers.csv definition name: none lookup type: file-based

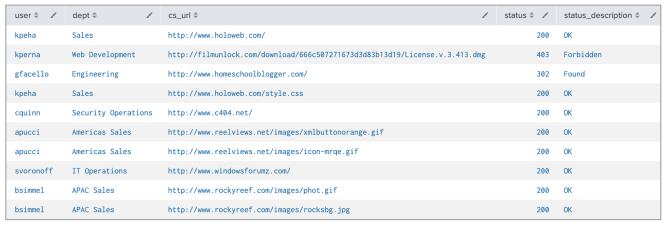
HINT: You may find it helpful to have both lookups available to reference for this task. Right-click on Search in the application bar (next to Datasets, Reports, etc.) and click "Open Link in New Tab." Run an inputlookup search on status\_definitions.csv. Repeat these steps for knownusers.csv.

18. Complete the <missing> portions of the following search:

```
index=network sourcetype=cisco wsa squid x webcat code full="Uncategorized URLs"
lookup knownusers.csv <missing>
 lookup status_definitions.csv <missing>
 search user=*
table user, dept, cs_url, status, status_description
```



- a. The first lookup should use knownusers.csv to retrieve user values for all matching username values in the events. (Hint: You will need to use user as username for your lookup. This tells Splunk to match the values of user from the lookup against the values of username from the event data.)
- b. The second lookup should use status\_definitions.csv to retrieve status\_description values for all matching status values in the events.
- Run the search over the Last 24 hours.



Example of final output.

19. Save your search as a report with the name **L1S3**.

Scenario: Sales would like a map of retail sales in Canada by province over the previous week.

Task 5: Use the geospatial lookup file, canada.kml, to return a choropleth map of Canadian retail sales by province during the previous week.

20. The knowledge manager has uploaded and defined the canada.kml geospatial lookup and provided you with the following information. Use this info to create a search that will display the contents of the lookup. You should see the geospatial lookup output displayed as a table with the following fields: count, featureCollection, featureId, and geom.

filename: canada.kml

definition name: canada\_prov lookup type: geospatial

21. Open a second search browser window. The following search calculates total sales from Canada in Canadian dollars. Complete the <missing> portions of the geom command so that the results of this search are correlated with the canada.kml lookup. (Hint: The geom command must use the geospatial lookup definition name and the featureIdField should be a field with values that are present in the events and in the lookup.)

```
index=sales sourcetype=vendor sales VendorCountry=Canada
 stats sum(price) as USDollars by VendorStateProvince
 eval CDNDollars = round(USDollars*1.31,2)
 fields - USDollars
 geom <missing> featureIdField=<missing>
```





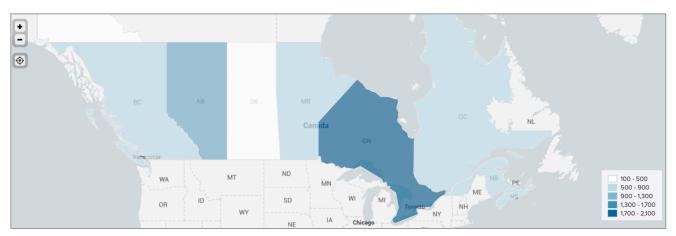
Output of the first 4 lines of this search.

22. Run the search over the Previous week and confirm that your output looks like the table below.



- 23. Click the Visualization tab and change the visualization to Choropleth Map.
- 24. Under the **Format** tab:
  - a. Set Latitude to 53.
  - b. Set Longitude to -92.
  - c. Set Zoom to 4.
  - d. Set Color Mode to Sequential.
  - e. Set Maximum Color to 006D9C.

# splunk>



25. Save your search as a report with the name **L1S4**.

Scenario: TechOps wants to be able to search for web server errors coming from www.buttercupgames.com that are associated with unsuccessful purchases.

Task 6: Troubleshoot this search and then output results to a lookup with the outputlookup command.

26. This search is not returning the desired results. Troubleshoot the **lookup** command expression.

index=web sourcetype=access\_combined status!=200
 referer\_domain=http://www.buttercupgames.com

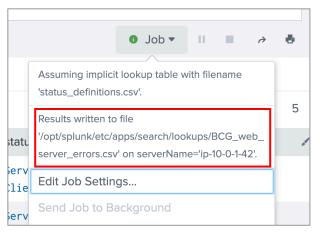
| lookup status\_definitions.csv status\_description as Description OUTPUT status status\_type

| stats count by host, status\_description, status\_type, clientip | stats list(status\_description) as status\_description, list(status\_type) as status\_type, list(host) as host, list(count) as "count" by clientip

clientip \$	1	status_description \$	1	status_type \$	1	host \$	1	count 🗢 🖋
107.3.146.207		Request Timeout Service Unavailable		Client Error Server Error		www2 www2		1 1
108.65.113.83		Bad Request		Client Error		www2		1
110.138.30.229		Service Unavailable Bad Request Internal Server Error		Server Error Client Error Server Error		www2 www3 www3		1 1 2
110.159.208.78		Internal Server Error Service Unavailable		Server Error Server Error		www1 www1		1 1
111.161.27.20		Not Found Service Unavailable Not Found Service Unavailable		Client Error Server Error Client Error Server Error		www2 www2 www3 www3		1 1 1 1

- 27. Output the results of this search to a lookup called **BCG\_web\_server\_errors.csv**. Make sure the lookup is created in the same app the search is being run. Run the search over the **Last 24 hours**.
- 28. Confirm that the BCG\_web\_server\_errors.csv lookup has been created by clicking on Job.





29. Save your search as a report with the name L1S5.

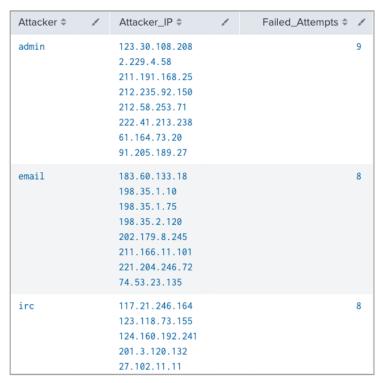
#### Challenge: Filter a search by excluding values from a lookup.

Scenario: SecOps is finding an increase in penetration attempts. Find *unknown* users with more than 3 failed logins within the last 24 hours.

- 30. Complete the <missing> portion of the lookup expression in this search. The search should exclude known users from the final results. Keep a few things in mind:
  - a. Both the linux\_secure data and the knownusers.csv lookup file use the same field name for user. Therefore, the user field from the linux\_secure data has been renamed to user\_from\_events before using the lookup and search commands.
  - b. The **search** command filters search results. The **<missing>** portion of the **search** expression is a field name.
  - c. The remainder of the search performs statistical aggregations on the results and further manipulates the data to achieve the scenario goal.
  - d. The search should be run over the Last 24 hours.

```
index=security sourcetype=linux_secure fail*
| rename user as user_from_events
| lookup <missing>
| search NOT <missing>=*
| stats count by user_from_events, src_ip
| stats values(src_ip) as Attacker_IP, sum(count) as Failed_Attempts by user_from_events
| rename user_from_events as Attacker
| search Failed_Attempts > 3
| sort -Failed Attempts
```





Example of final output.

31. Save your search as a report with the name **L1X**.



# Lab Exercise 2 – Adding a Subsearch

# Description

Create subsearches to manipulate search input.

#### Steps

Scenario: Marketing and Sales would like to know how many times multiplayer games were "viewed" on the website during the "Multiplayer Madness" event this past Saturday.

# Task 1: Use a subsearch and a lookup to filter search results.

 Your knowledge manager provided you with the following information. These lookups contain information about the products sold by Buttercup Games. This task requires events from the following games: SIM Cubicle, Dream Crusher, Mediocre Kingdoms, Puppies vs. Zombies, Manganiello Bros., Final Sequel, Benign Space Debris, and Curling 2014. Use the inputlookup command to find the correct lookup and verify its contents.

filename: products.csv

definition name: product\_lookup

description: code, category ID, price, product ID, and sale price of all products

lookup type: file-based

filename: sp\_products.csv definition name: none

description: list of single player games

lookup type: file-based

filename: mp\_products.csv definition name: none

description: list of multiplayer games

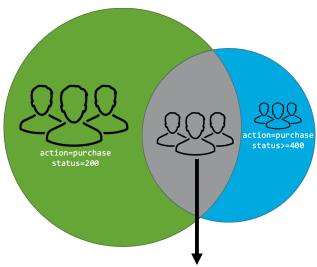
lookup type: file-based

2. This search is looking back to Saturday (earliest=@w6 latest=@w7) for all events involving a "view" action in the web sales index. Then, the search transforms and sorts the data to show which game was viewed the most. Replace the <missing> portion of the basic search with a subsearch so that only events involving multiplayer games are returned.

```
index=web sourcetype=access_combined action="view" earliest=@w6 latest=@w7 <missing>
| stats count(action) as "viewed" by product_name
| sort -viewed
```

- 3. Save your search as a report with the name **L2S1**.
- Task 2: Combine two searches into a single search. The resulting search should find the average and median sales totals for clientips who have experienced problems making a purchase (action=purchase status>=400) but still managed to complete a successful web order during the previous week.
- 4. This Venn diagram represents the components of this search: the results of the outer search (green), the results of the inner search (blue), and the results of the outer search filtered by the results of the inner search (grey).

# splunk>



The data from these customers (clientips) will undergo statistical transformations to get average and median sales totals

Based on the task description, we want to perform statistical transformations on the data represented by the grey inner section—the customers that experienced problems with a purchase (action=purchase status>=400) yet still completed a successful online sales order (action=purchase status=200) over the previous week.

Answer these questions about the inner and outer searches:

- a. TRUE or FALSE: The inner search (blue) will look for customers who did not experience issues with their online purchase.
- b. TRUE or FALSE: The outer search (green) will look for successful purchase events but only return events from customers that appeared in the results of the inner search.
- 5. Which of these searches provides the desired results of the inner search?

clientip \$
107.3.146.207
108.65.113.83
109.169.32.135
110.138.30.229
111.161.27.20
112.111.162.4
117.21.246.164
118.142.68.222

Desired results of inner search.

#### Search 1

```
index=web sourcetype=access_combined status=200 action=purchase
| stats sum(sale_price) as sales_sum by clientip
| stats avg(sales_sum) as avg_sales, median(sales_sum) as median_sales
```

#### Search 2

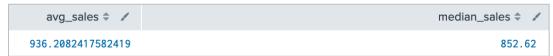
```
index=web sourcetype=access_combined status=200 OR status=400 action=purchase
| stats sum(sale_price) as sales_sum by clientip
| stats avg(sales_sum) as avg_sales, median(sales_sum) as median_sales
```



#### Search 3

search index=web sourcetype=access\_combined status>=400 action=purchase
| stats values(clientip) as clientip

6. Which of these searches provides the desired results of the outer search? (Note: If you run these searches, remove the [<subsearch>] placeholder, otherwise you will receive an error.)



Desired results of outer search.

#### Search 1

```
index=web sourcetype=access_combined status=200 action=purchase
  [<subsearch>]
| stats sum(sale_price) as sales_sum by clientip
| stats avg(sales_sum) as avg_sales, median(sales_sum) as median_sales
```

#### Search 2

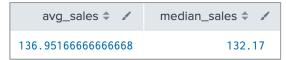
```
index=web sourcetype=access_combined status=200 OR status=400 action=purchase
  [<subsearch>]
| stats sum(sale_price) as sales_sum by clientip
| stats avg(sales_sum) as avg_sales, median(sales_sum) as median_sales
```

#### Search 3

search index=web sourcetype=access\_combined status>=400 action=purchase
 [<subsearch>]

| stats values(clientip) as clientip

7. Combine the inner and outer search to create your final search. Run this search over the **Previous week**.



8. Save your search as a report with the name **L2S2**.



# Lab Exercise 3 – Using the return Command

# Description

Use the **return** command to control output from a search and a subsearch.

# Steps

### Task 1: Return search results as key value pairs.

A coworker has asked you to help create a subsearch for a report. You have created a search that
normalizes username and Username values in the network data and finds the top 5 most active users.
Complete the <missing> portion of the search so that User values are returned as key-value pairs. Run
the search over the Last 24 hours.

```
index=network
| eval User=coalesce(username,Username)
| stats count by User
| sort 5 -count
| <missing>
```

User \$	1	count 🗢 🗸
kosullivan		4
mfleischman		4
dhale		37
gvoronoff		44
cberztiss		54

Before the **return** command.

```
search ♦

(User="acurry") OR (User="edutra") OR (User="ewilliams") OR (User="myavatkar") OR (User="npearce")
```

After the **return** command.

2. Save your search as a report with the name L3S1.

Scenario: SecOps wants to know which employees have entered invalid passwords over the last 7 days.

Task 2: Filter search input by returning key-value pairs from the employees.csv lookup. Count instances of "failed password" by employee usernames.

3. Your knowledge manager has provided you with the following information about the employees.csv lookup. Create a search that will open employees.csv and return all USERNAME values as key-value pairs. (Hint: Use the inputlookup command with the employees.csv lookup to find out how many rows of data exist in the lookup file. The number of rows will match the number of results returned. Then, use this number with the return command.)



filename: employees.csv

definition name: employee\_lookup

lookup type: file-based

```
search $
(USERNAME="ewilliams") OR (USERNAME="mkemmerer") OR (USERNAME="myavatkar") OR (USERNAME="gbowser") OR (USERNAME="djohnson") OR
(USERNAME="swrappe") OR (USERNAME="pdabbeville") OR (USERNAME="gowen") OR (USERNAME="edutra") OR (USERNAME="myuan") OR
(USERNAME="gnooteboom") OR (USERNAME="kpercy") OR (USERNAME="gzuyeva") OR (USERNAME="cganttchart") OR (USERNAME="sle") OR
(USERNAME="gfacello") OR (USERNAME="dtempesti") OR (USERNAME="rjayaraman") OR (USERNAME="cberztiss") OR (USERNAME="maxwell") OR
(USERNAME="pbridgland") OR (USERNAME="basselin") OR (USERNAME="hsham") OR (USERNAME="yschonegge") OR (USERNAME="sflaemmchen") OR
(USERNAME="spahkthecah") OR (USERNAME="showser") OR (USERNAME="cfarrell") OR (USERNAME="lsagers") OR (USERNAME="rroberts") OR
(USERNAME="pbunch") OR (USERNAME="svoronoff") OR (USERNAME="cmunson") OR (USERNAME="gvoronoff") OR (USERNAME="apreusig") OR
(USERNAME="bhussain") OR (USERNAME="fullian") OR (USERNAME="blu") OR (USERNAME="ewarwick") OR (USERNAME="syoungin") OR
(USERNAME="tzielinski") OR (USERNAME="podessa") OR (USERNAME="sscallion") OR (USERNAME="fyards") OR (USERNAME="lteng") OR
(USERNAME="rerde") OR (USERNAME="msluis") OR (USERNAME="kjoslin") OR (USERNAME="tcugina") OR (USERNAME="kpeha") OR
(USERNAME="dpiazza") OR (USERNAME="ptoscani") OR (USERNAME="iking") OR (USERNAME="apucci") OR (USERNAME="arangel") OR
(USERNAME="madeyemi") OR (USERNAME="bsimmel") OR (USERNAME="bgenin") OR (USERNAME="nsharpe") OR (USERNAME="fbryan") OR
(USERNAME="cquinn") OR (USERNAME="acurry") OR (USERNAME="adombrowski") OR (USERNAME="npearce") OR (USERNAME="heapers") OR
(USERNAME="pleuchs") OR (USERNAME="gbottazzi") OR (USERNAME="jreistad") OR (USERNAME="jcappelletti") OR (USERNAME="dhale") OR
(USERNAME="moh") OR (USERNAME="kperna")
```

4. This search looks for "failed password" events in the **security** index. Filter the search input by adding the subsearch you created in the previous step. Then, run the search over the **Last 7 days**. What results were returned?

```
index=security "failed password"
| stats count by user
```

5. The search is not working because the subsearch is returning **USERNAME** values while the outer search is aggregating on **user** values. Fix the search and run over the **Last 7 days**. (Hint: No additional pipes need to be added to the search.)

No results found. Try expanding the time range.

Before editing the search.



After editing the search.

6. Save your search as a report with the name **L3S2**.