Enriching Data with Lookups – Lab Guide

Overview

Welcome to the Splunk Education lab environment. These lab exercises will have you create automatic lookups to provides additional information for a source type, upload lookup table files, use lookups in searches, and upload a (KML) lookup table file and create a Geospatial lookup definition to use it in searches and to create a choropleth visualization report.

Scenario

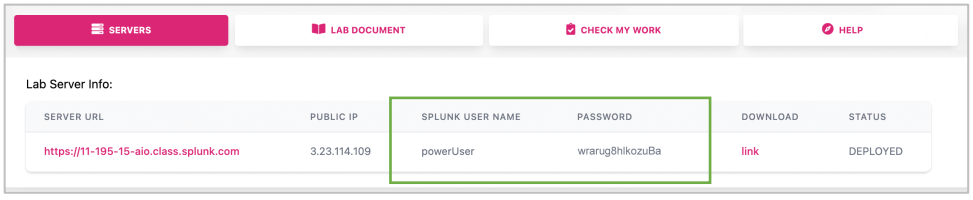
You will use data from the international video game company, Buttercup Games. A list of source types is provided below.

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**	action, app, dest, process, src_ip, src_port, user, vendor_action
sales	Business Intelligence server	**sales_entries**	AcctCode, CustomerID, TransactionID
	Retail sales	**vendor_sales**	categoryId, product_name, productId, sale_price, Vendor, VendorCity, VendorCountry, VendorID, VendorStateProvince
network	Email security data	**cisco_esa**	dcid, icid, mailfrom, mailto, mid
	Web security appliance data	**cisco_wsa_squid**	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
	Firewall data	**cisco_firewall**	bcg_ip, dept, Duration, fname, IP, lname, location, rfid, splunk_role, splunk_server, Username

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 Lookups

Description

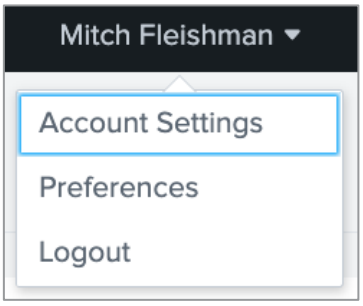
Configure the lab environment user account. Then, create a new automatic lookup that provides additional information to the **`access_combined`** source type.

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***.)

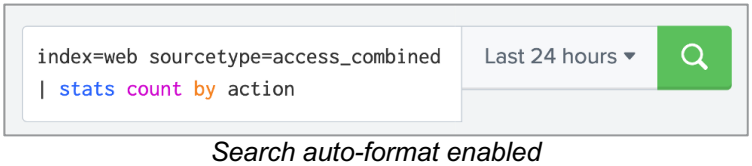


After you complete step 6, you will see your name in the web interface.

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.

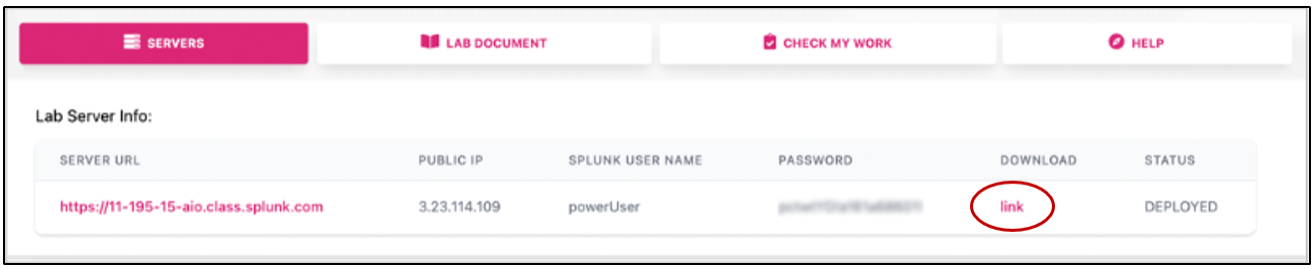
Search auto-format disabled (default)



Search auto-format enabled

Scenario: The `access_combined` source type contains http status codes, but not the code definitions.

Task 2: Add a lookup file to the `access_combined` source type to make the code definitions available as fields.



11. Obtain the **`status_definitions.csv`** file (see image above for location of link to file).
12. View the **`status_definitions.csv`** file with a text editor, noticing the comma-separated value format that defines HTTP response status codes, their description, and status type:

```
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
```

13. Navigate to **Settings > Lookups**.
14. Click on **Lookup table files** and view the existing entries.
15. Click **New Lookup Table File**.
    a. Save the lookup table file with these values:
       - Destination app: **search**
       - File: **status_definitions.csv**
       - Destination filename: **status_definitions.csv**
    b. Click **Save**.

Task 3: Create a lookup definition.

16. In the top left corner of Splunk Web, select **Apps** > **Search & Reporting**. This sets our app context to the search app.

a. If you received a welcome message, click **Skip**.

17. Navigate to **Settings > Lookups**.
18. Click on **Lookup definitions** and view the existing entries.
19. Click **New Lookup Definition**.
    a. Save the lookup definition with these values:
       - Destination app: **search**
       - Name: **status_definitions_lookup**
       - Type: **File-based**
       - Lookup file: **status_definitions.csv**
    b. Click **Save**.

Task 4: Verify the lookup definition.

20. In the top left corner, select **Apps > Search & Reporting**.
21. Use the **`inputlookup`** command with the name of the lookup definition to verify the contents of the lookup file and that the lookup definition was created correctly.

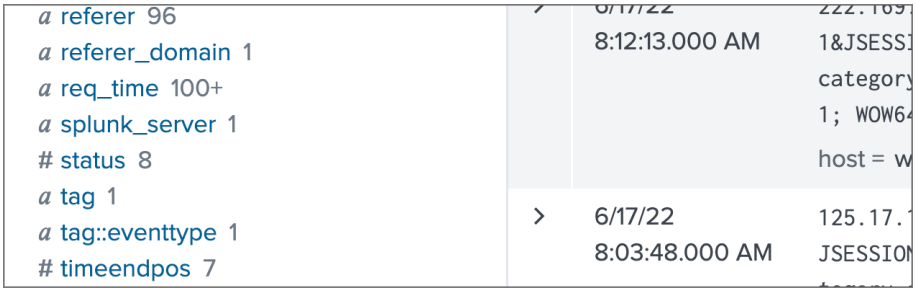
status	status_description	status_type
100	Continue	Informational
101	Switching Protocols	Informational
200	OK	Successful
201	Created	Successful
202	Accepted	Successful
203	Non-Authoritative Information	Successful
204	No Content	Successful
205	Reset Content	Successful

Task 5: Use your lookup in a search.

22. Search the online store data (**`index=web`**) over the **Last 24 hours** for all events that were not associated with an "OK" status of **`200`**.

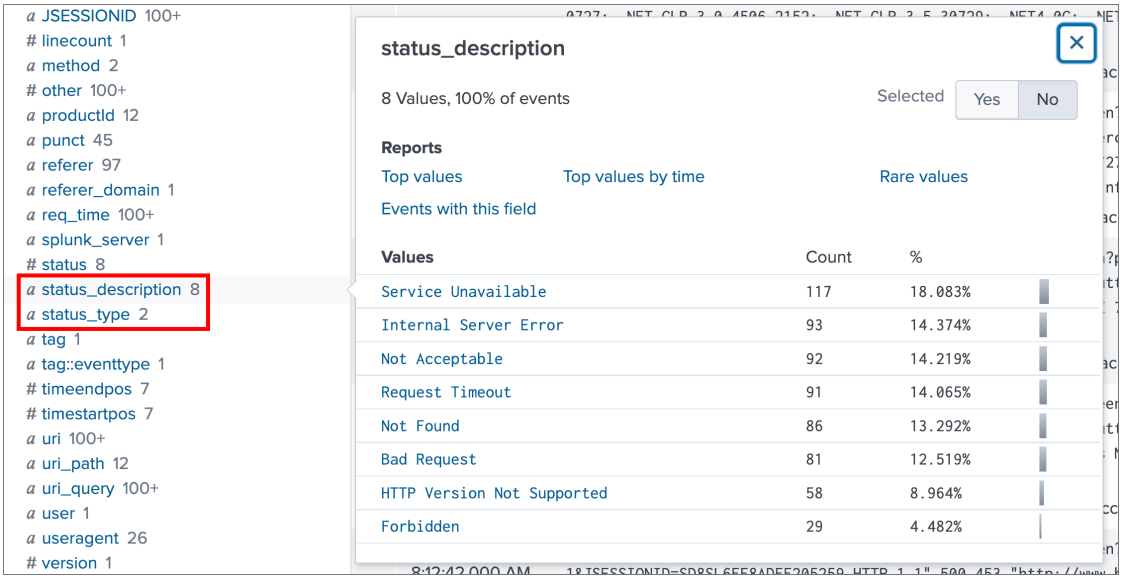
i	Time	Event
>	6/17/22 8:20:22.000 AM	220.225.12.171 - - [17/Jun/2022:14:20:22] "GET /product.screen?productId=SF-BVS-01&JSESSIONID=SD5SL6FF9ADFF4958 HTTP 1.1" 406 3977 "http://www.buttercupgames.com/product.screen?productId=SF-BVS-01" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.2; .NET CLR 1.1.4322; InfoPath.1; MS-RTC LM 8)" 485 host = www3 source = /opt/log/www3/access.log sourcetype = access_combined
>	6/17/22 8:12:56.000 AM	222.169.224.226 - - [17/Jun/2022:14:12:56] "GET /category.screen?categoryId=NULL&JSESSIONID=SD8SL6FF8ADFF205259 HTTP 1.1" 406 1226 "http://www.buttercupgames.com/oldlink?itemId=EST-21" "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 440 host = www1 source = /opt/log/www1/access.log sourcetype = access_combined

23. For the same search results, view the **Interesting Fields** side bar. Note that there are no fields for status description or status type.



24. Add **`status_description`** and **`status_type`** fields using the lookup definition you created in the previous task. Pipe results to **`| lookup status_definitions_lookup status`**.

25. Run the search for the **Last 24 hours** and verify that **`status_description`** and **`status_type`** fields are included the **Interesting Fields** list.

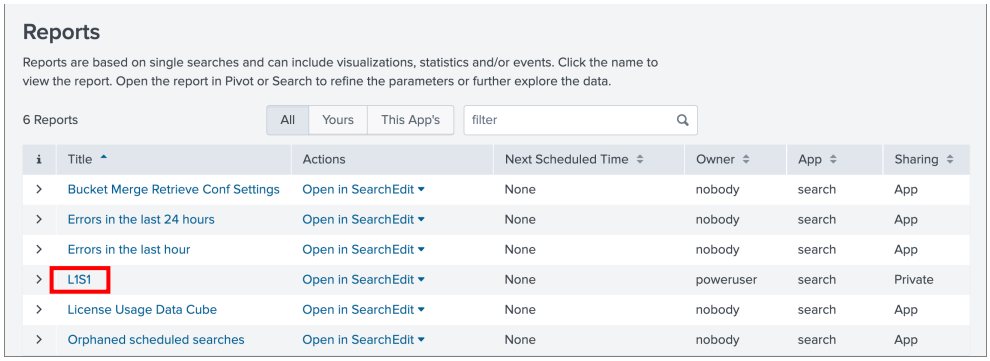


NOTE: Step 26 is optional and requires knowledge of the **`stats`** command. You can skip this step and follow step 27 to save your search as a report.

26. Modify the search to use the **`stats`** command to get a **`count`** by **`host`**, **`status_description`**, and **`status_type`**.

host	status_description	status_type	count
www1	Bad Request	Client Error	23
www1	HTTP Version Not Supported	Server Error	29
www1	Internal Server Error	Server Error	42
www1	Not Acceptable	Client Error	41

27. 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.
    f. Re-initialize the search window by clicking **Search** in the application bar.



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

Task 6: Create an automatic lookup definition.

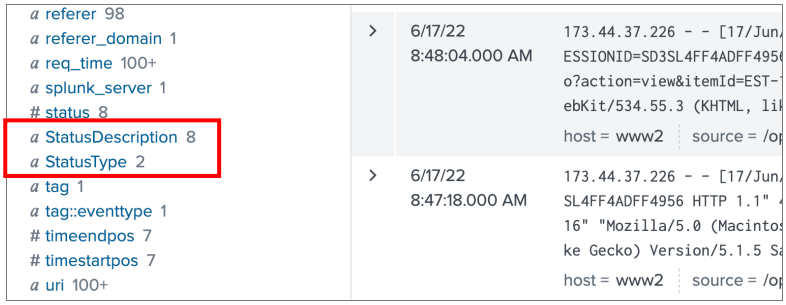
28. Navigate to **Settings > Lookups**.
29. Click on **Automatic lookups** and view the existing entries.
30. Click **New Automatic Lookup**.
    a. Create the automatic lookup with these values:
        - Destination app: **search**
        - Name: **status_definitions_auto_lookup**
        - Lookup table: **status_definitions_lookup**
        - Apply to: **sourcetype**
        - named*: **access_combined**
        - Lookup input fields: **status = status**
        - Lookup output fields (use the **+ Add another field** button as necessary):
            - **status_description = StatusDescription**
            - **status_type = StatusType**
    b. Click **Save**.

NOTE: It may take a few moments before the automatic lookup starts working.

Task 7: Verify your automatic lookup is working.

31. In the top left corner, select **Apps > Search & Reporting**.
32. Search the online store data for the **Last 24 hours** for all events that do not have a status of 200.

33. In the search results under the **Interesting Fields** sidebar, notice that **`StatusDescription`** and **`StatusType`** are showing automatically, without requiring the use of any **`lookup`** commands.

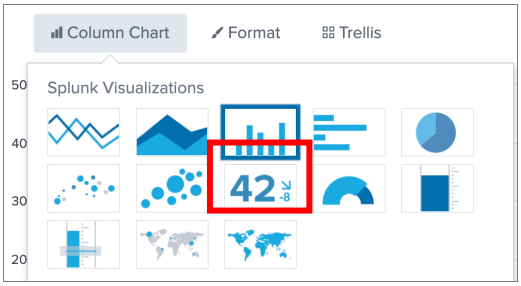


NOTE: Steps 35 - 37 are optional and require knowledge of the **`stats`** command. You can skip these steps and follow step 38 to save your search as a report.

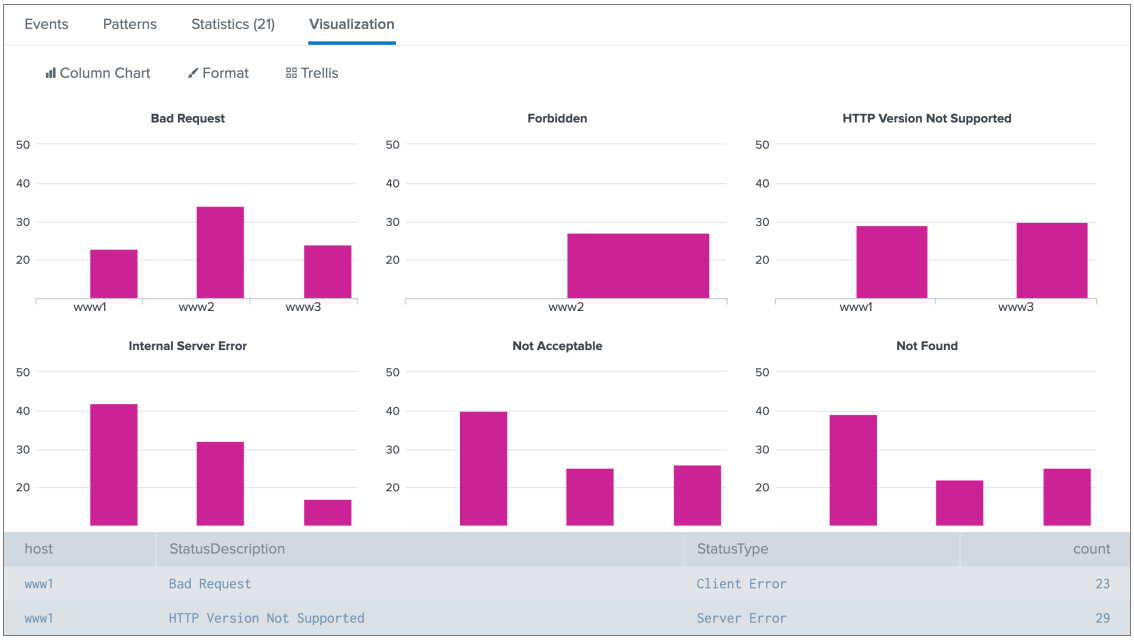
34. Search the online store data and count events by **`host`**, **`StatusDescription`**, and **`StatusType`** over the **Last 24 hours**.

host	StatusDescription	StatusType	count
www1	Bad Request	Client Error	23
www1	HTTP Version Not Supported	Server Error	29
www1	Internal Server Error	Server Error	42
www1	Not Acceptable	Client Error	40
www1	Not Found	Client Error	39

35. Click on **Visualization.** If needed, select **Column Chart**.



36. Create multiple visualizations for each status description. Click on **Trellis**.
    - Select the **Use Trellis Layout** checkbox.
    - For **Split By** select **StatusDescription**.



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

Scenario: HR wants a count of logins by known Buttercup Games employees over the last 24 hours. Exclude non-standard employee accounts in the results.

Task 8: Upload the `knownusers.csv` lookup table file and create a lookup definition to filter out non-standard Buttercup employees from the lookup.

38. Obtain the **`knownusers.csv`** file.
39. View the **`knownusers.csv`** file with a text editor. Notice the first line is the header **`user`** followed by rows of known users in the environment.

```
user
root
mail
apache
acurry
adombrowski
apreusig
apucci
```

NOTE: The **`knownusers.csv`** lookup contains Buttercup employees as well as common user accounts such as **`root`**, **`mail`**, and so on.

40. Navigate to **Settings** > **Lookups** and click **+ Add new** next to **Lookup table files**.
    a. Save the lookup table file with these values:
       - Destination app: **search**
       - File: **`knownusers.csv`**

- Destination filename: **knownusers.csv**

b. Click **Save**.

41. Navigate back to the **Search & Reporting** app and check the contents of the lookup using the **inputlookup** command. There should be 76 results.

user
root
mail
apache
acurry
adombrowski
apreusig
apucci

42. Navigate to **Settings** > **Lookups** and click **+ Add new** next to **Lookup definitions**.
    a. Save the lookup definition with these values:
       - Destination app: **search**
       - Name: **knownusers_lookup**
       - Type: **File-based**
       - Lookup file: **knownusers.csv**
    b. Check the **Advanced options** checkbox.
    c. For **Filter lookup**, write a Boolean expression that excludes **root**, **mail**, and **apache** users from the lookup.
    d. Click **Save**.
43. Navigate back to the **Search & Reporting** app and use the **inputlookup** command to verify that the lookup definition does not include **root**, **mail** or **apache**.

user
acurry
adombrowski
apreusig
apucci

44. Add **| lookup knownusers_lookup user OUTPUT user** to the following search so that the results are limited to only Buttercup Games employees. Search over the **Last 24 hours**.

```
index=security sourcetype=linux_secure
| stats count by user
```

user	count
djohnson	63
myuan	20
nsharpe	22

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

Lab Exercise 2 – Geospatial and External Lookups

Description

In this exercise, you will define an external lookup, upload a (KML) lookup table file, create a geospatial lookup definition, and use these lookups in searches.

Steps

Task 1: Upload and define a geospatial lookup and verify its contents in search.

1. Obtain the **`canada.kml`** file.
2. Navigate to **Settings** > **Lookups** and click **+ Add new** next to **Lookup table files**.
   a. Save the lookup table file with these values:
      - Destination app: **search**
      - File: **`canada.kml`**
      - Destination filename: **`canada.kml`**
   b. Click **Save**.
3. Navigate to **Settings** > **Lookups** and click **+ Add new** next to **Lookup definitions**.
   a. Save the lookup definition with these values:
      - Destination app: **search**
      - Name: **canada_prov**
      - Type: **Geospatial**
      - Lookup file: **`canada.kml`**
   b. Click **Save**.
4. Navigate back to the **Search & Reporting** app and check the contents of the lookup using the **`inputlookup`** command. The desired output will be displayed in a table showing the following fields: **`count`**, **`featureCollection`**, **`featureId`**, and **`geom`**.

count	featureCollection	featureId	geom
0	canada_prov	Alberta	{"type":"MultiPolygon","coordinates":[[[[-110.36250305175781,60.000057220458984],[-110.0050277709961,49.000057220458984],[-114.06903076171875,49.000064849853516],[-119.89948272705078,53.519126892089844],[-120,60],[-110.36250305175781,60.000057220458984]]]]}
0	canada_prov	British Columbia	{"type":"MultiPolygon","coordinates":[[[[-123.600830078125,48.31638717651367],[-123.600830078125,48.31638717651367]]],[[[-123.54055786132812,48.31833267211914],[-123.54055786132812,48.31833267211914]]],[[[-123.60694122314453,48.329444885253906],[-123.60694122314453,48.3294448852539 06]]],[[[-123.7074966430664,48.33194351196289],[-123.7074966430664,48.33194351196289]]],[[[-123.62666320800781,48.33444595336914],[-123.62666320800781,48.33444595336914]]],[[[-123.6490478515625,48.37310791015625],[-123.6490478515625,48.37310791015625]]],[[[-123.65361022949219,48.38916778564453],

5. Save your search as a report with the name **L2S1**.

Task 2: Create and use an external lookup with `external_lookup.py` script to return a count of online sales events by host name.

6. Sales wants a count of online sales events by host name over the last 15 minutes. This search looks for online sales events and calculates a count of each value for **`clientip`**. Run this search over the **Last 60 minutes**:

```
index=web sourcetype=access_combined
| stats count by clientip
```

clientip	count
110.159.208.78	7
173.44.37.226	3
175.44.24.82	10
194.146.236.22	1
195.216.243.24	14

7. You will need to use an external lookup to enrich your data with client host values. The lookup you will be using, **`external_lookup.py`**, has already been moved to the search app directory (**`SPLUNK_HOME/etc/apps/search/bin/external_lookup.py`**), which is required before you can define this external lookup. Navigate to **Settings** > **Lookups** and next to **Lookup** definitions, click **+ Add new**.
   a. Save the lookup table file with these values:
      - Destination app: **search**
      - Name: **dnslookup**
      - Type: **External**
      - Command: **`external_lookup.py clienthost clientip`**
      - Supported fields: **`clienthost,clientip`**
   b. Click **Save**.
8. Navigate back to the **Search & Reporting** app and perform a search of online sales during the **Last 60 minutes**. Invoke the **`dnslookup`** lookup with the **`lookup`** command and pipe the results to **`stats count by clienthost`** to count the results by **`clienthost`**.

clienthost	count
11.97.175.69.unassigned.ord.singlehop.net	2
2-229-4-58.ip194.fastwebnet.it	13
20-0-229-94.bganglobalnet.net	33
217-23-14-61.hosted-by-worldstream.net	14
32.7c.1732.ip4.static.sl-reverse.com	11

9. Rewrite the search to include HTTP status and HTTP status descriptions by piping to **`stats count by clienthost, status, status_description`**. This will require an additional lookup command that uses the **`status_definitions.csv`** lookup.

clienthost	status	status_description	count
11.97.175.69.unassigned.ord.singlehop.net	200	OK	2
2-229-4-58.ip194.fastwebnet.it	200	OK	9
2-229-4-58.ip194.fastwebnet.it	400	Bad Request	1
2-229-4-58.ip194.fastwebnet.it	404	Not Found	1
2-229-4-58.ip194.fastwebnet.it	408	Request Timeout	1
2-229-4-58.ip194.fastwebnet.it	500	Internal Server Error	1

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