Data Models – Lab Guide

Overview

Welcome to the Splunk Education lab environment. These lab exercises will test your knowledge of working with and accelerating data models.

Scenario

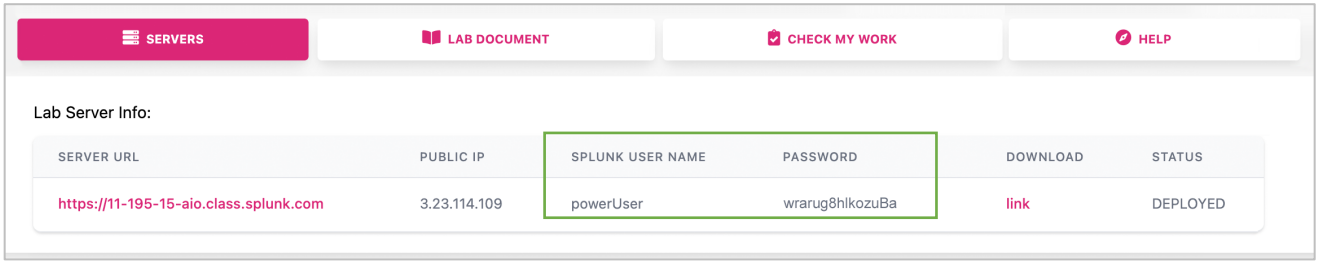
You will use data from the international video game company, Buttercup Games. A list of source types is provided below.

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
sales	Business Intelligence server	**sales_entries**	AcctCode, CustomerID, TransactionID
	Retail sales	**vendor_sales**	categoryId, product_name, productId, sale_price, Vendor, VendorCity, VendorCountry, VendorID, VendorStateProvince

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 – Design Data Models

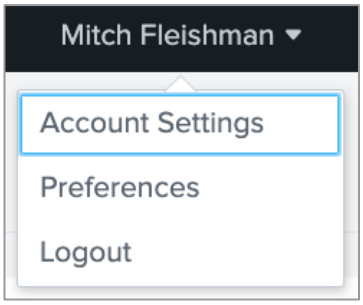
Description

This exercise walks you through the process of creating a data model and adding datasets.

Steps

Task 1: Login to 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. Login to your Splunk lab environment using the username and password provided to you.
2. You may see a pop-up welcoming you to the lab environment. You may click **Continue to Tour** but this is not required. Click **Skip** to dismiss the pop-up window.
3. Click on your username at the top of the screen and then choose **Account Settings** from the dropdown. (Note: This is the **User Menu**.)
4. In the **Full name** box, enter your first and last name.
   For example: Mitch Fleischman
5. Click **Save** and reload your browser.

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.

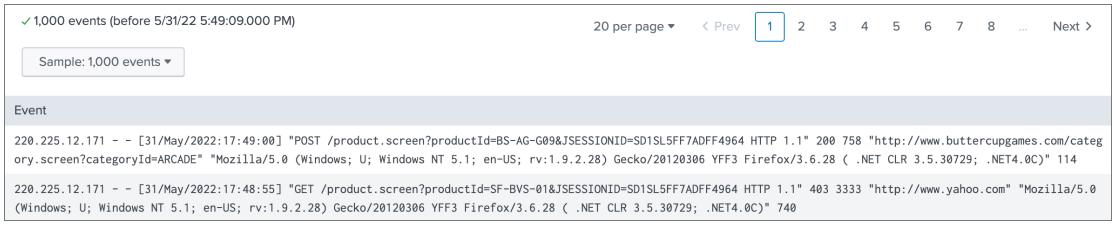
6. Navigate to ***user name*** **> Preferences.**
7. Choose your local time zone from the **Time zone** dropdown.
8. Click **Apply**.
9. (Optional) Navigate to **User Menu > Preferences > SPL Editor > Search auto-format** and click on the toggle to activate auto-formatting. Then click **Apply**. When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

Search auto-format disabled.

Search auto-format enabled.

Task 2: Create a data model and add a Web Requests root event. The root event will be the base search for all child events.

10. In Splunk Web, navigate to **Settings > Data models**.
    a. Click **New Data Model**.
    b. In the Title field, type: **Buttercup Games Site Activity**. (Notice that this automatically fills in the ID field. **Don't** delete this value. The ID field cannot be blank.)
    c. For App, make sure **Search & Reporting** is selected.
    d. Click **Create**.
11. Click **Add Dataset** and select **Root Event**.
    a. In the **Dataset Name** field, type: **Web requests**.
    b. In the **Constraints** field, type: **`index=web sourcetype=access_combined`**
    c. Click **Preview** to see a sampling of the events.

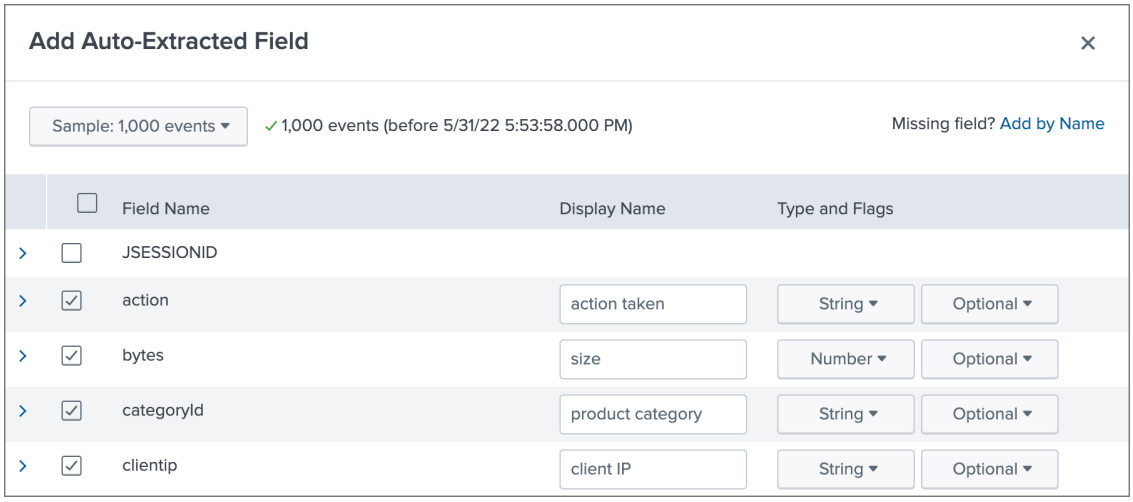


    d. Verify the events match your constraints. Events from **`index=web sourcetype=access_combined`** should start with an IP address, and contain GET or POST message fields and web URLs. Note: If the preview does not match the expected results, check the **Constraints** field you typed to ensure there are no mistakes.
    e. Keep the **Sample: 1,000 events** selection at this time.
    f. Click **Save** to save the root event.

Task 3: Add auto-extracted fields.

12. Make sure the root **Web requests** dataset is selected. Click **Add Field** and select **Auto-Extracted**. A dialog box opens and displays all auto-extracted fields.
    a. Click the check boxes to select the following fields, and rename them for pivot users as indicated:
       – action > action taken
       – bytes > size
       – categoryId > product category
       – clientip > client IP
       – date_mday > date_mday (*use same name*)

- productId > product ID
- product_name > product name
- req_time > request time
- status > status (*use same name*)



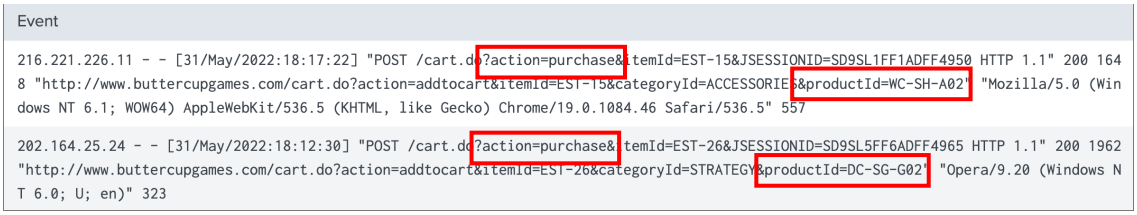
b. Click **Save**.

Task 4: Add two child events, one for actions that were successful (`status<400`) and one for actions that failed (`status>399`.)

13. Click **Add Dataset** and select **Child**.
    a. In the **Dataset Name** field, type: **Successful requests**
    b. In the **Additional Constraints** field, type: **`status<400`**
    c. Click **Preview** to see a test sample of your results.
    d. Verify the events match your constraints. Check the number field value that comes just after the string field that starts with the word “GET” or “POST”. The number should be less than 400

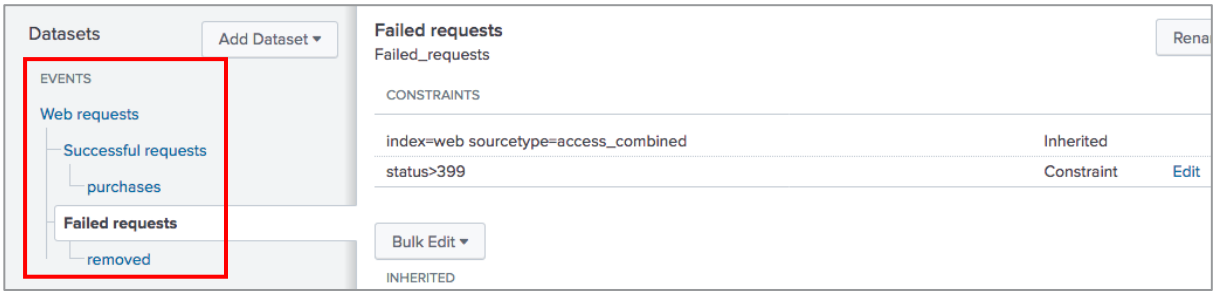
    e. **Save** the child dataset.

14. Select the **Successful requests** dataset.
    a. Add a child dataset called **purchases** with an **Additional Constraints** value of **`action=purchase productId=*`**.
    b. Click **Preview** to see a test sample of your results, and verify the events match your constraints.



c. **Save** the child dataset.

15. Select the **Web requests** event and add a child dataset named: **Failed requests**
    a. In the **Additional Constraints** field, type: **`status>399`**
    b. Click **Preview** to see a test sample of your results, and verify the events match your constraints.
    c. **Save** the child dataset.
16. Under the **Failed requests** dataset, add a child dataset named: **removed**
    a. In the **Additional Constraints** field, type: **`action=remove productId=*`**
    b. Click **Preview** to see a test sample of your results, and verify the events match your constraints.
    c. **Save** the child dataset.
17. Verify your dataset shows the root event as **Web requests**, with two child datasets (**Successful requests** and **Failed requests**), each of which has one additional child dataset (**purchases** and **removed**).



Lab Exercise 2 – Create a Pivot

Description

Create pivot reports and visualizations using the data model you created in previous lab exercise.

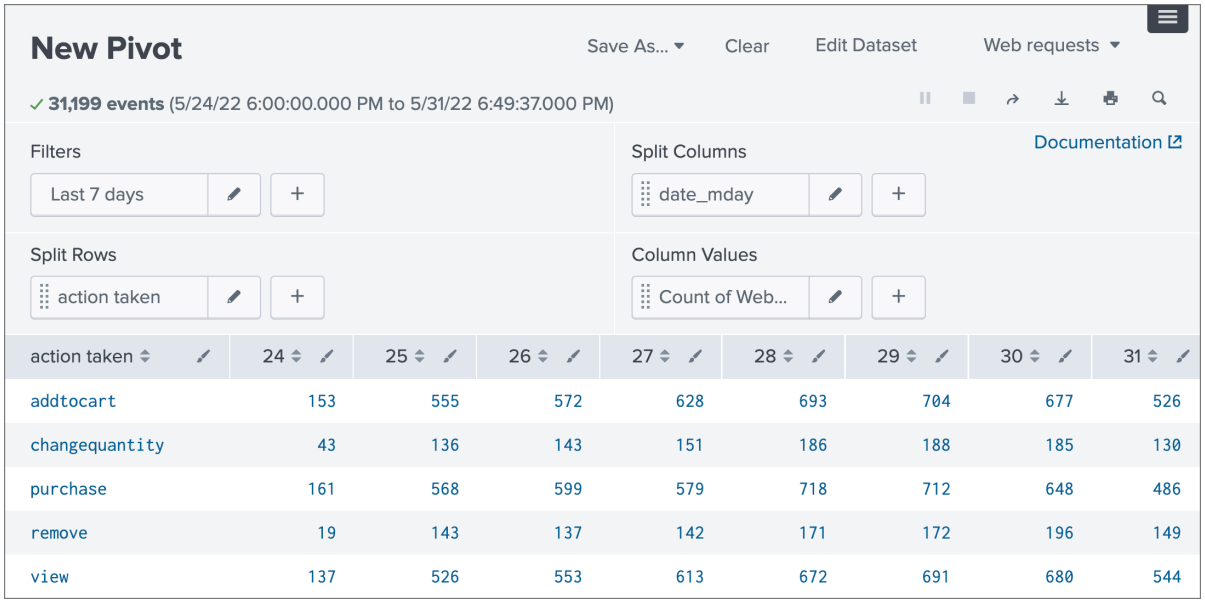
Steps

Task 1: Test your data model by creating a pivot.

1. Click **Pivot** in the upper-right corner to test the data model.

NOTE: If you are no longer in the same view of Splunk Web from the end of Lab Exercise 1, first navigate to **Settings > Data models**, then click on **Buttercup Games Site Activity**.

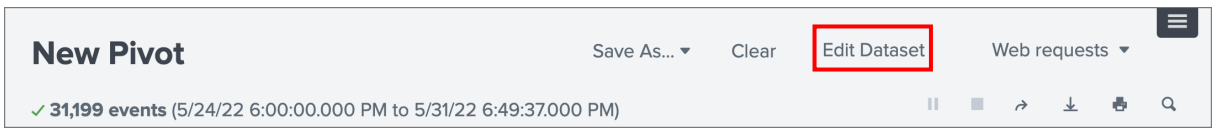
2. Select the **Web requests** dataset.
3. In the **New Pivot** window, change the following:
   - Change **Filters** from **All Time** to **Last 7 days**
   - **Split Rows** by **action taken** and click **Add To Table**
   - **Split Columns** by **date_mday** and click **Add To Table**



action taken	24	25	26	27	28	29	30	31
addtocart	153	555	572	628	693	704	677	526
changequantity	43	136	143	151	186	188	185	130
purchase	161	568	599	579	718	712	648	486
remove	19	143	137	142	171	172	196	149
view	137	526	553	613	672	691	680	544

Task 2: Add a field that uses an eval expression. The eval expression will display events chronologically by date and day of the week.

4. Click on **Edit Dataset** and select the **Web requests** so that it is highlighted.



a. From the **Add Field** drop-down list on the right, select **Eval Expression**.
b. In the **Eval Expression** field, type: **`strftime(_time,"%m-%d %A")`**

NOTE: **`strftime`** is a function that converts epoch time to a readable date and time format.

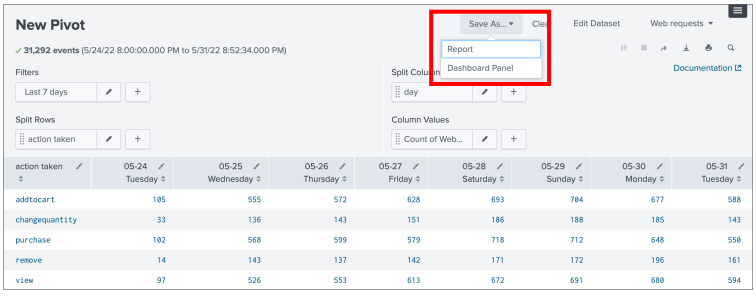
c. For Field Name, type: **day**
d. For Display Name, type: **day**
e. Click **Preview** to verify your eval expression returns results.

_time	day	host
2022-05-31 20:47:10	05-31 Tuesday	www1
2022-05-31 20:46:55	05-31 Tuesday	www1

f. **Save** the eval expression.

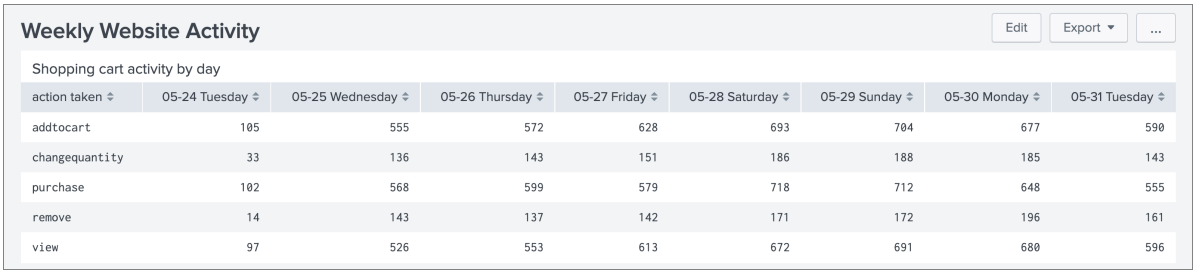
Task 3: Verify the eval expression works as expected by using Pivot to create a dashboard.

5. Click **Pivot**.
   a. Select the **Web requests** dataset.
   b. Change the time filter to the **Last 7 days**.
   c. Split Rows by **action taken**. Click **Add To Table**.
   d. Split Columns by **day**. Click **Add To Table**. (This is the new eval expression field we created in the last task.)
   e. Click **Save As** and select **Dashboard Panel**.



f. For **Dashboard Title**, type: **Weekly Website Activity**
g. For **Panel Title**, type: **Shopping cart activity by day**
h. Click **Save**.

6. Click **View Dashboard**. You should see the web requests categorized and counted by day.



Weekly Website Activity

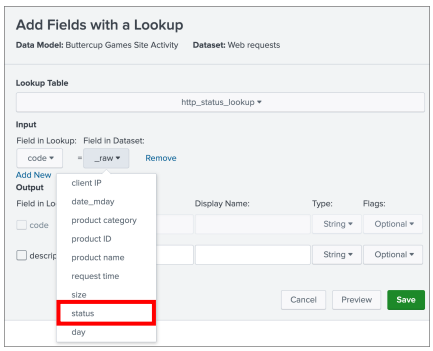
Edit Export ...

Shopping cart activity by day

action taken	05-24 Tuesday	05-25 Wednesday	05-26 Thursday	05-27 Friday	05-28 Saturday	05-29 Sunday	05-30 Monday	05-31 Tuesday
addtocart	105	555	572	628	693	704	677	590
changequantity	33	136	143	151	186	188	185	143
purchase	102	568	599	579	718	712	648	555
remove	14	143	137	142	171	172	196	161
view	97	526	553	613	672	691	680	596

Task 4: Add fields from a lookup. The lookup table will provide descriptions of status codes.

7. Verify that you are still in the **Search & Reporting** app. If necessary, click to expand the **Apps** menu next to the **splunk>** logo at the top left of the window and choose **Search & Reporting**. If a window appears asking you to take a tour, click **Skip**.
8. Navigate to **Settings > Data models**. Select the **Buttercup Games Site Activity** data model.
    a. Make sure the **Web requests** root dataset is selected.
    b. Click **Add Field** and select **Lookup**.
    c. From the **Lookup Table** drop-down list, select **http_status_lookup**.
    d. For the **Input** section in the **Field in Lookup** drop-down list, ensure **code** is selected.
    e. From the **Field in Dataset** drop-down list**,** select **status**. (You may need to scroll down the list to see this value.) This maps the **`status`** field in your indexed data to the **`code`** column in the lookup table.



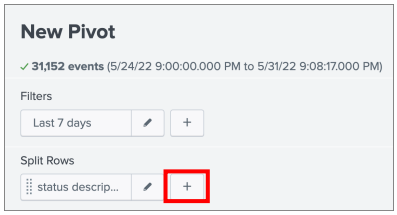
    f. For the lookup **Output** section in the **Field in Lookup** field, check the **description** check box.
    g. In the **Display Name** type: **status description**.
    h. Click the **Preview** button. You should see a **description** column in the results.

_time	description	host
2022-05-31 21:04:19	OK.	www3
2022-05-31 21:01:05	OK.	www3
2022-05-31 21:00:56	OK.	www3
2022-05-31 21:00:42	OK.	www3

    i. Click **Save**.

Task 5: Verify the lookup works properly by creating a Pivot report.

9. Click **Pivot**.
    a. Select the **Web requests** dataset.
    b. Change the Filter to **Last 7 days**.
    c. From **Split Rows**, add the **status description** attribute and click **Add To Table**.
    d. Click the **+** button to split by another row and add the **status** attribute. Click **Add To Table**.



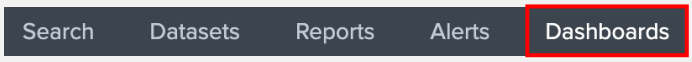
NOTE: This is a double row split, not a column split.

    e. Verify that in addition to the event count, the table shows two columns, one for **status description** and one for **status**.

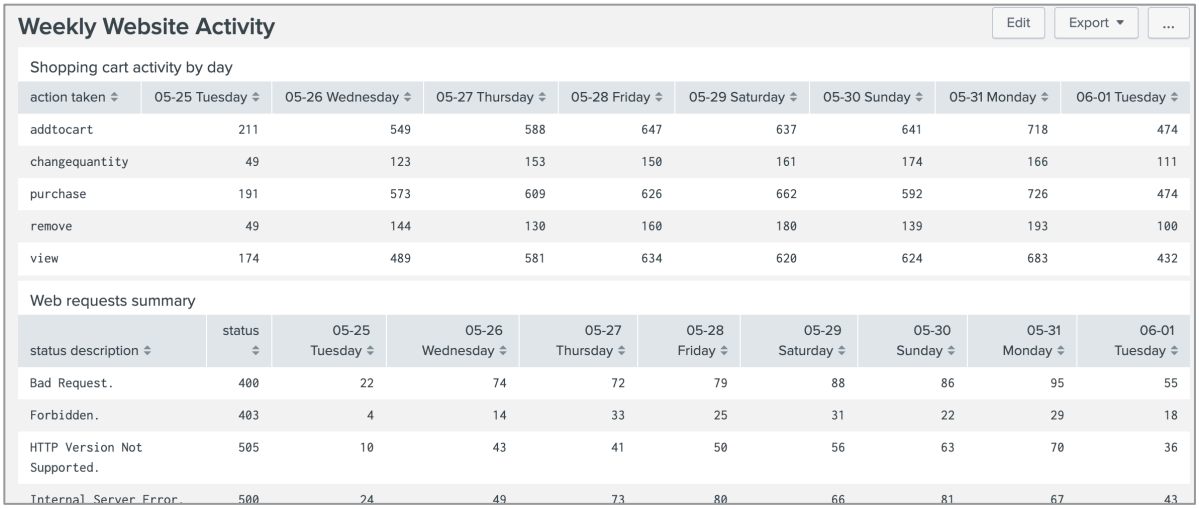
status description	status	Count of Web requests
Bad Request.	400	493
Forbidden.	403	201
HTTP Version Not Supported.	505	387
Internal Server Error.	500	523
Not Acceptable.	406	543
Not Found.	404	550
OK.	200	27114
Request Timeout.	408	566
Service Unavailable.	503	775

    f. **Split Columns** by **day** and click **Add To Table**.
    g. Click **Save As** and select **Dashboard Panel**.
    h. Select **Existing** and select **Weekly Website Activity**.
    i. For the **Panel Title**, type: **Web requests summary**
    j. Click **Save**.
    k. Click **View Dashboard**.

NOTE: You can also access available dashboards from the Splunk tool bar.



To do so, click **Dashboards > Weekly Website Activity**.



Weekly Website Activity

Edit Export ...

Shopping cart activity by day

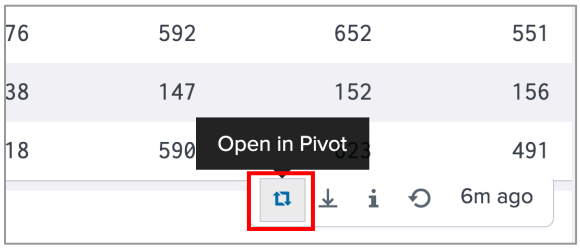
action taken	05-25 Tuesday	05-26 Wednesday	05-27 Thursday	05-28 Friday	05-29 Saturday	05-30 Sunday	05-31 Monday	06-01 Tuesday
addtocart	211	549	588	647	637	641	718	474
changequantity	49	123	153	150	161	174	166	111
purchase	191	573	609	626	662	592	726	474
remove	49	144	130	160	180	139	193	100
view	174	489	581	634	620	624	683	432

Web requests summary

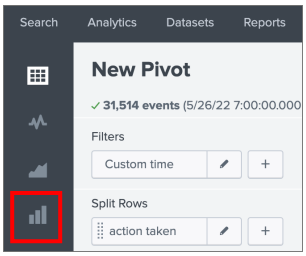
status description	status	05-25 Tuesday	05-26 Wednesday	05-27 Thursday	05-28 Friday	05-29 Saturday	05-30 Sunday	05-31 Monday	06-01 Tuesday
Bad Request.	400	22	74	72	79	88	86	95	55
Forbidden.	403	4	14	33	25	31	22	29	18
HTTP Version Not Supported.	505	10	43	41	50	56	63	70	36
Internal Server Error.	500	24	49	73	80	66	81	67	43

Task 6: From the pivot editor, add a filter to narrow your results.

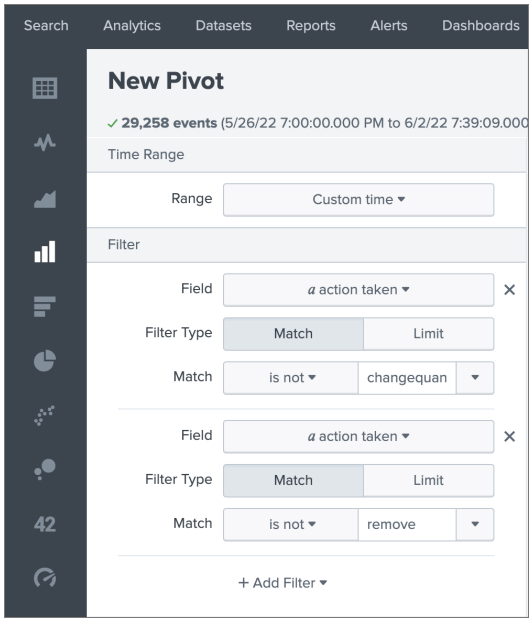
10. Hover your mouse over the lower-right corner of the **Shopping cart activity by day** dashboard panel. Click the **Open in Pivot** icon.



11. Select the **Column chart** icon from the table formats on the left.

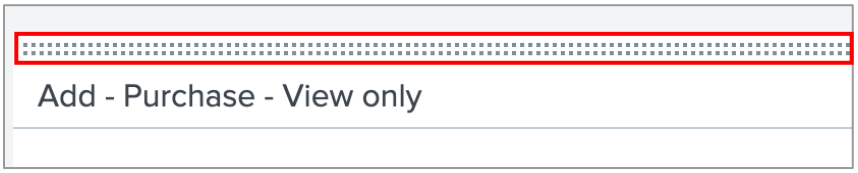


12. To narrow your results, click **+ Add Filter** and choose **action taken**.
    a. For **Filter Type**, select **Match**.
    b. For **Match**, change the operator to **is not**, then select **changequantity**.
    c. Add another filter and again choose **action taken**.
    d. For the **Filter Type**, select **Match**.
    e. For **Match**, change the operator to **is not** and then select **remove**.

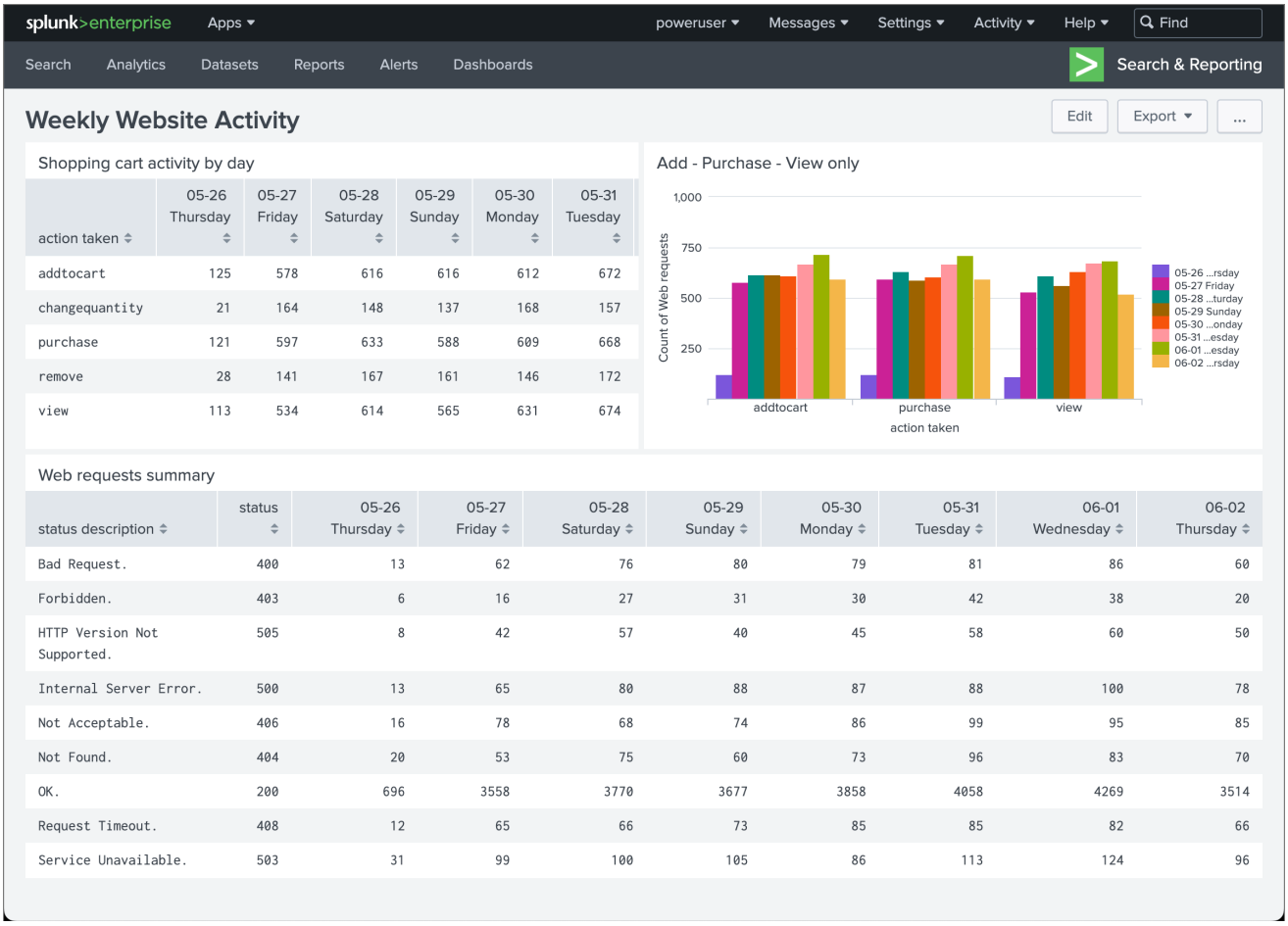


f. Click **Save As** and select **Dashboard Panel**.
g. Click **Existing** and select the **Weekly Website Activity** dashboard.
h. For **Panel Title**, type: **Add – Purchase - View only**. (Include all 4 words in the panel title)
i. Click **Save**.
j. Click **View Dashboard**.

13. In the resulting dashboard, rearrange the panels to your liking and admire your work!
    a. Click the **Edit** button in the top right corner.
    b. Scroll down to the **Add - Purchase - View only** part of the dashboard, and hover over the two dotted parallel lines at the top of that panel.



    c. Drag the panel to the top right corner of the dashboard, so it appears to the right of the **Shopping cart activity by day** panel, but above the **Web requests summary** panel.
    d. Click the **Save** button on the top right corner.


splunk>enterprise
Apps
poweruser
Messages
Settings
Activity
Help
Find
Search
Analytics
Datasets
Reports
Alerts
Dashboards
Search & Reporting
Weekly Website Activity
Edit
Export
...
Shopping cart activity by day
action taken
05-26 Thursday
05-27 Friday
05-28 Saturday
05-29 Sunday
05-30 Monday
05-31 Tuesday
addtocart 125 578 616 616 612 672
changequantity 21 164 148 137 168 157
purchase 121 597 633 588 609 668
remove 28 141 167 161 146 172
view 113 534 614 565 631 674
Add - Purchase - View only
Count of Web requests
1,000
750
500
250
addtocart
purchase
view
action taken
05-26 ...rsday
05-27 Friday
05-28 ...turday
05-29 Sunday
05-30 ...onday
05-31 ...esday
06-01 ...esday
06-02 ...rsday
Web requests summary
status description
status
05-26 Thursday
05-27 Friday
05-28 Saturday
05-29 Sunday
05-30 Monday
05-31 Tuesday
06-01 Wednesday
06-02 Thursday
Bad Request. 400 13 62 76 80 79 81 86 60
Forbidden. 403 6 16 27 31 30 42 38 20
HTTP Version Not Supported. 505 8 42 57 40 45 58 60 50
Internal Server Error. 500 13 65 80 88 87 88 100 78
Not Acceptable. 406 16 78 68 74 86 99 95 85
Not Found. 404 20 53 75 60 73 96 83 70
OK. 200 696 3558 3770 3677 3858 4058 4269 3514
Request Timeout. 408 12 65 66 73 85 85 82 66
Service Unavailable. 503 31 99 100 105 86 113 124 96

Lab Exercise 3 – Accelerate Data Models

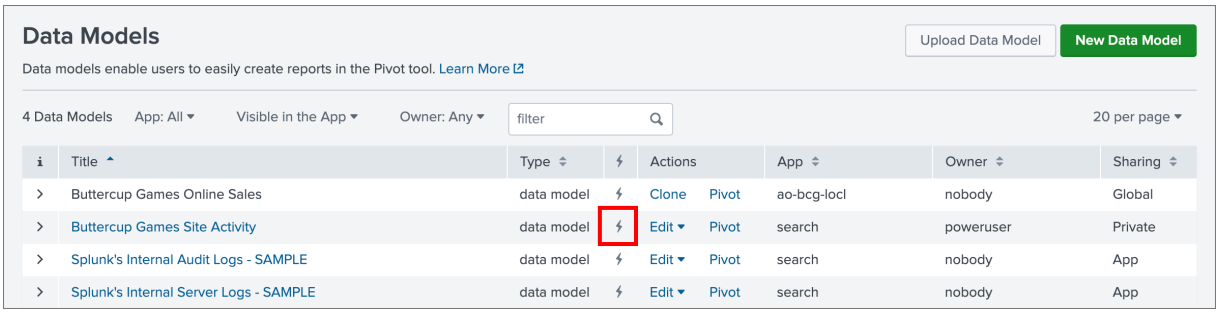
Description

This exercise walks you through the process of accelerating a data model. You will clone the previously created data model and accelerate it. Additionally, you will perform some searches to verify the behavior of the accelerated data model.

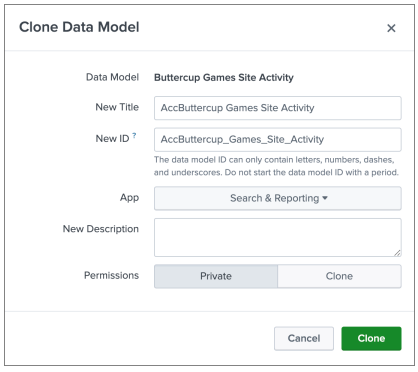
Steps

Task 1: Accelerate a Data Model.

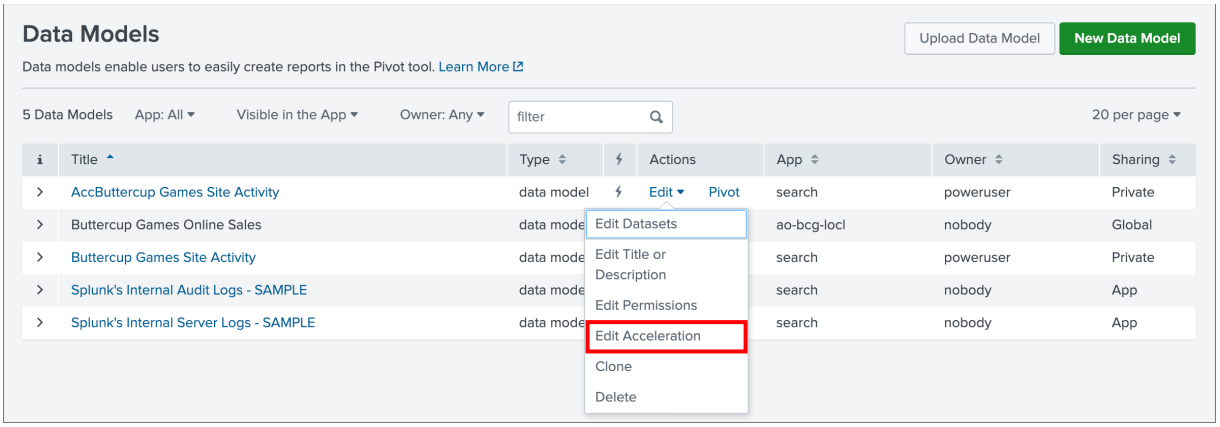
1. Navigate to **Settings > Data models**.
    a. In the **Data Models** view, ensure that **App: All** is selected.
    b. Click on the **Owner: Any** drop down and select your username.
    c. You should see only the **Buttercup Games Site Activity** data model. Verify that the lightning bolt icon is grey, showing that the data model is currently not accelerated.



2. In the **Buttercup Games Site Activity** row, select **Edit > Clone**.
3. In the **Clone Data Model** window, prepend “Acc” so that the **New Title** is “AccButtercup Games Site Activity”. (Note: The **New ID** field will automatically update.)

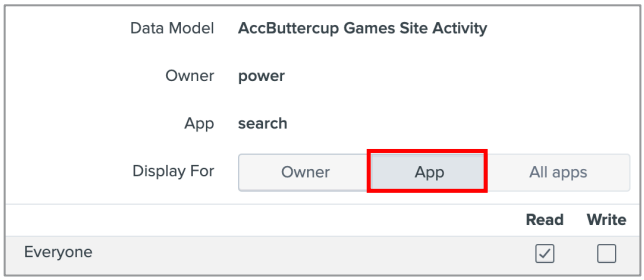


4. Click **Clone**.
5. In the **AccButtercup Games Site Activity** row, select **Edit > Edit Acceleration**.

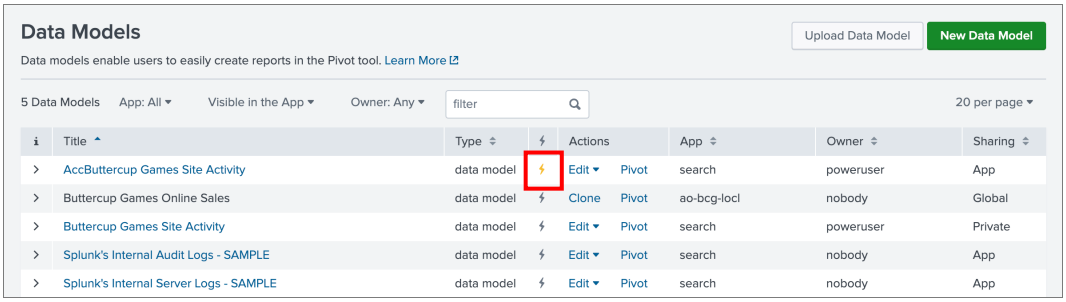


NOTE: In this lab environment, the Splunk **powerUser** role was provided with the **accelerate_datamodel** role capability. By default, the Splunk **power** role does not have this capability. Therefore, you would have to login as a user with the **admin** role to accelerate a data model.

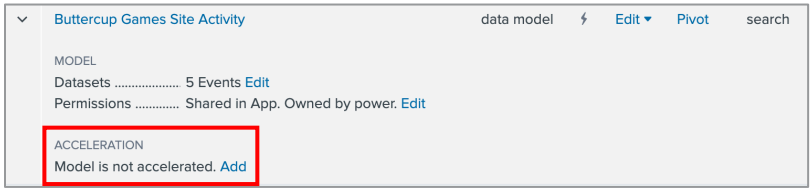
6. You should see an **Add Acceleration** window with the message "Private models cannot be accelerated. Edit permissions before enabling acceleration." Click on **Edit Permissions**.
7. To the right of **Display For** click on **App**, and then click the box for **Read** permissions for **Everyone**.



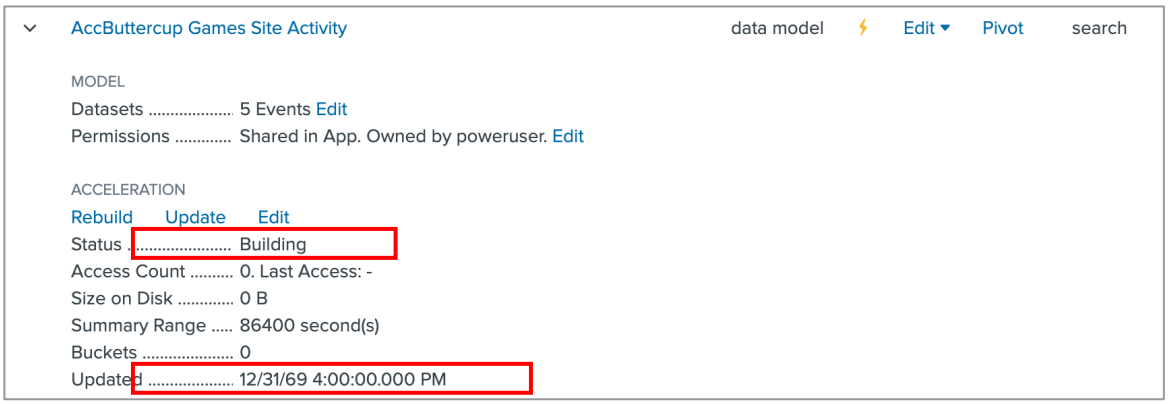
8. Click **Save** to save the new permissions.
9. In the **AccButtercup Games Site Activity** row, select **Edit > Edit Acceleration** again, now that permissions have been updated.
    a. Click on the **Accelerate** checkbox. Notice the message under the checkbox that reads "Acceleration may increase storage and processing costs."
    b. Leave the **Summary Range** as **1 Day**.
    c. Expand **Advanced Settings** to view additional settings.
    d. Take note of the **Summarization Period**, which is currently set to `*/5 * * * *`. This value is in cron format and means that acceleration will run every 5 minutes.
    e. Click **Save**.
10. Verify that the lightning bolt icon is now yellow, showing that the data model is currently accelerated.



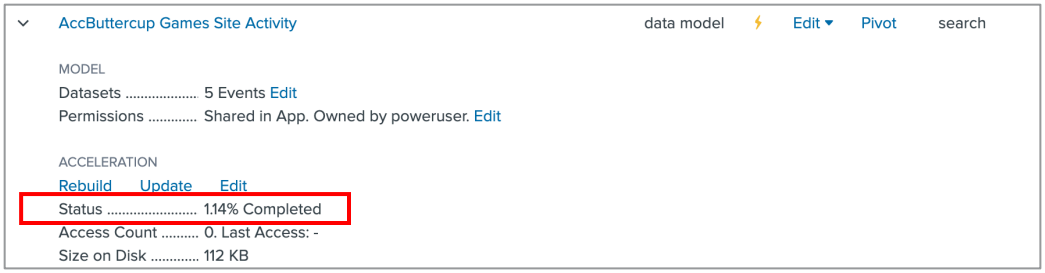
11. Click on the arrow (**>**) icons under the information (**i**) row on the far left for **Buttercup Games Site Activity**. Note that this is the first data model you created in the lab.
    a. Notice that the **Buttercup Games Site Activity** data model shows the "Model is not accelerated" under the **ACCELERATION** heading.+



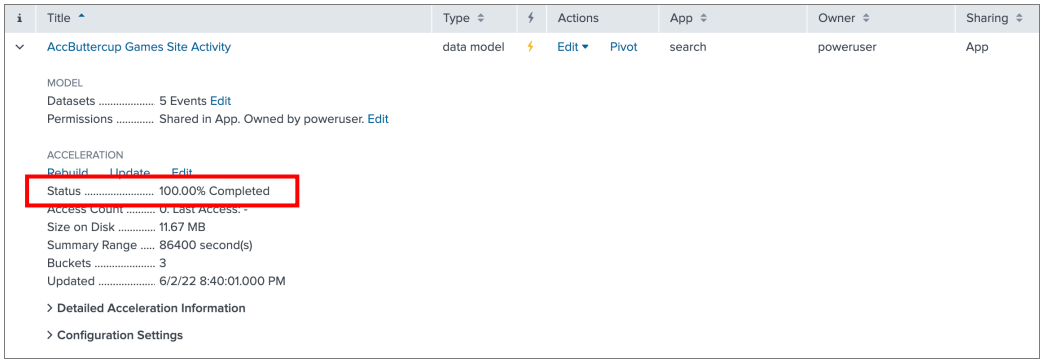
12. Click on the arrow (**>**) icons under the information (**i**) row on the far left for **AccButtercup Games Site Activity**.
    a. Notice that the **AccButtercup Games Site Activity** data model shows additional information under the **ACCELERATION** heading, including the **Status**, **Size on Disk**, **Summary Range** (currently set to **86400** seconds, which is equivalent to 1 day), and more. You may see a **Status** of Building, and an **Updated** date showing the Unix epoch time (shown as a date of 12/31/69 or 1/1/70, depending on your time zone). This is normal just after the data model is accelerated.



    b. Click on the arrow (**>**) icons next to **Detailed Acceleration Information** and **Configuration Settings** to view additional details about data model acceleration.
    c. Return to viewing the information under the **ACCELERATION** heading. Click the **Update** link under the **ACCELERATION** heading occasionally to refresh this view.



d. Once **Status** displays **100.00% Completed** you may proceed to the next task.



NOTE: It may take up to 10-15 minutes for the **Status** to display **100.00% Completed**. This may be a good time to take a break before moving on to the next task.

Task 2: Explore your data model using search commands.

13. Go to the search application by clicking on **Apps > Search & Reporting** in the upper left corner of Splunk Web.
14. Use this search to display the number of events in the **AccButtercup Games Site Activity** data model for the **Last 24 hours**:

```
| tstats count from datamodel=AccButtercup_Games_Site_Activity
```

15. Use this search to display the number of events in the **AccButtercup Games Site Activity** data model for the **Last 24 hours**, but only using accelerated data (with **`summariesonly=true`**, **`tstats`** only generates results from the TSIDX data that has been accelerated):

```
| tstats summariesonly=true count from datamodel=AccButtercup_Games_Site_Activity
```

16. How do these two values compare?