Splunk4Ninjas - Common Information Model (CIM)

An introduction to Splunk CIM Data Models and making data CIM compliant

splunk>

Forward-Looking Statements

This presentation may contain forward-looking statements regarding future events, plans or the expected financial performance of our company, including our expectations regarding our products, technology, strategy, customers, markets, acquisitions and investments. These statements reflect management's current expectations, estimates and assumptions based on the information currently available to us. These forward-looking statements are not guarantees of future performance and involve significant risks, uncertainties and other factors that may cause our actual results, performance or achievements to be materially different from results, performance or achievements expressed or implied by the forward-looking statements contained in this presentation.

For additional information about factors that could cause actual results to differ materially from those described in the forward-looking statements made in this presentation, please refer to our periodic reports and other filings with the SEC, including the risk factors identified in our most recent quarterly reports on Form 10-Q and annual reports on Form 10-K, copies of which may be obtained by visiting the Splunk Investor Relations website at www.investors.splunk.com or the SEC's website at www.sec.gov. The forward-looking statements made in this presentation are made as of the time and date of this presentation. If reviewed after the initial presentation, even if made available by us, on our website or otherwise, it may not contain current or accurate information. We disclaim any obligation to update or revise any forward-looking statement based on new information, future events or otherwise, except as required by applicable law.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. We undertake no obligation either to develop the features or functionalities described, in beta or in preview (used interchangeably), or to include any such feature or functionality in a future release.

splunk>

Please introduce yourself!

- Name
- Company/organisation
- Role
- Are you currently using Splunk?
- What are you interested in using Splunk for?

Workshop Agenda

- Overview of Splunk data sources and Splunkbase add-ons roles
- What is the Splunk Common Information Model (CIM)?
- Splunk data models and data model acceleration overview
- CIM compliance process walkthrough
- Hands-on Practice lab exercises:

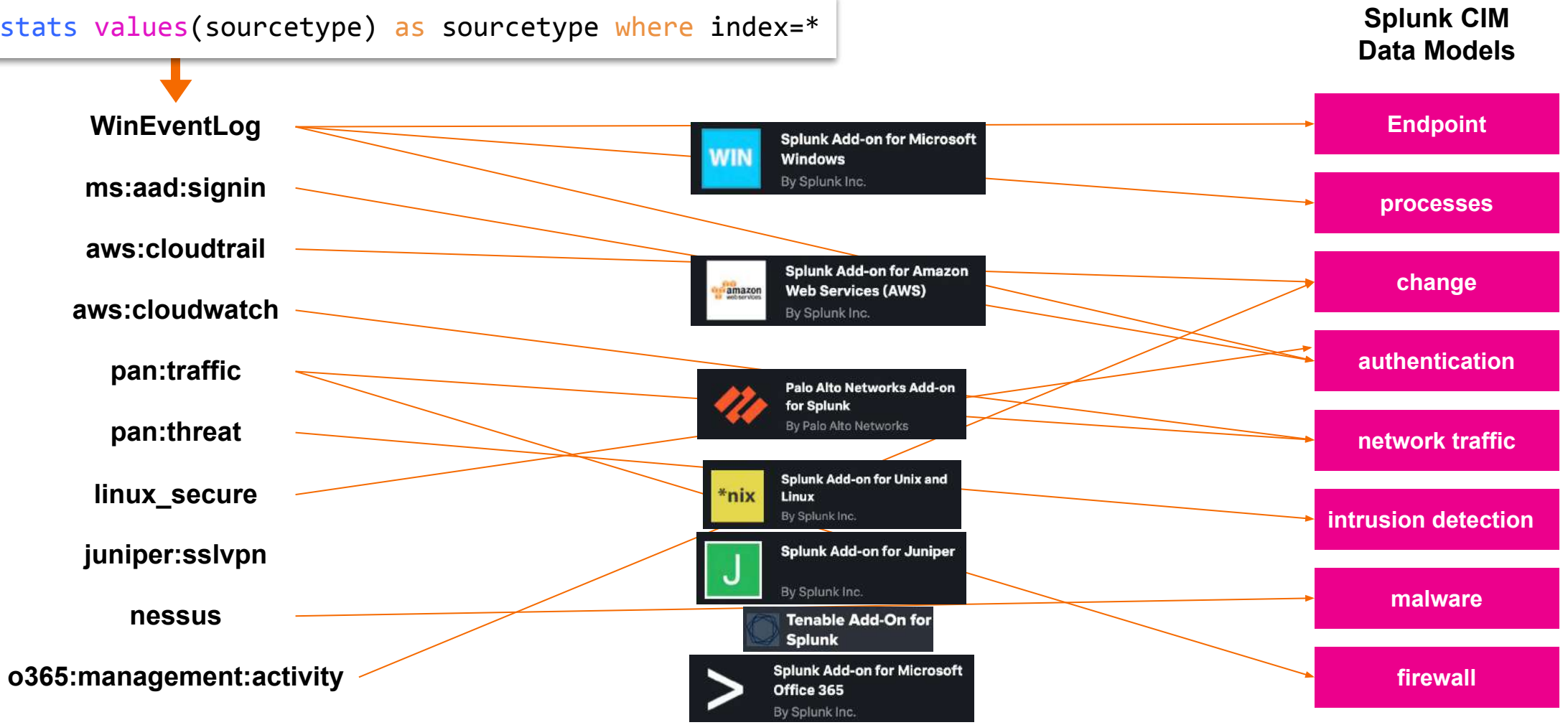
  Lab 1: Onboarding CIM compliant data (firewall traffic logs) using the FortiGate add-on

  Lab 2: Onboarding custom data (badge data) and manually mapping to the CIM Authentication data model

  Lab 3: Dashboarding CIM data using the `tstats` command

Overview of Splunk Data Sources

```
| tstats values(sourcetype) as sourcetype where index=*
```



Discover Splunk Apps: https://splunkbase.splunk.com/

splunk>

What's "CIM"?

CIM = Common Information Model

(DTMF's CIM != Splunk's CIM) = compatibility

Distributed Management Task Force
is a nonprofit industry standards organization that creates open manageability standards.

DTMF's CIM provides a **common definition** of management information for systems, networks, applications and services.

The CIM standard **includes a specification and a schema**, as well as a Metamodel

Learn more: https://www.dmtf.org/standards/cim

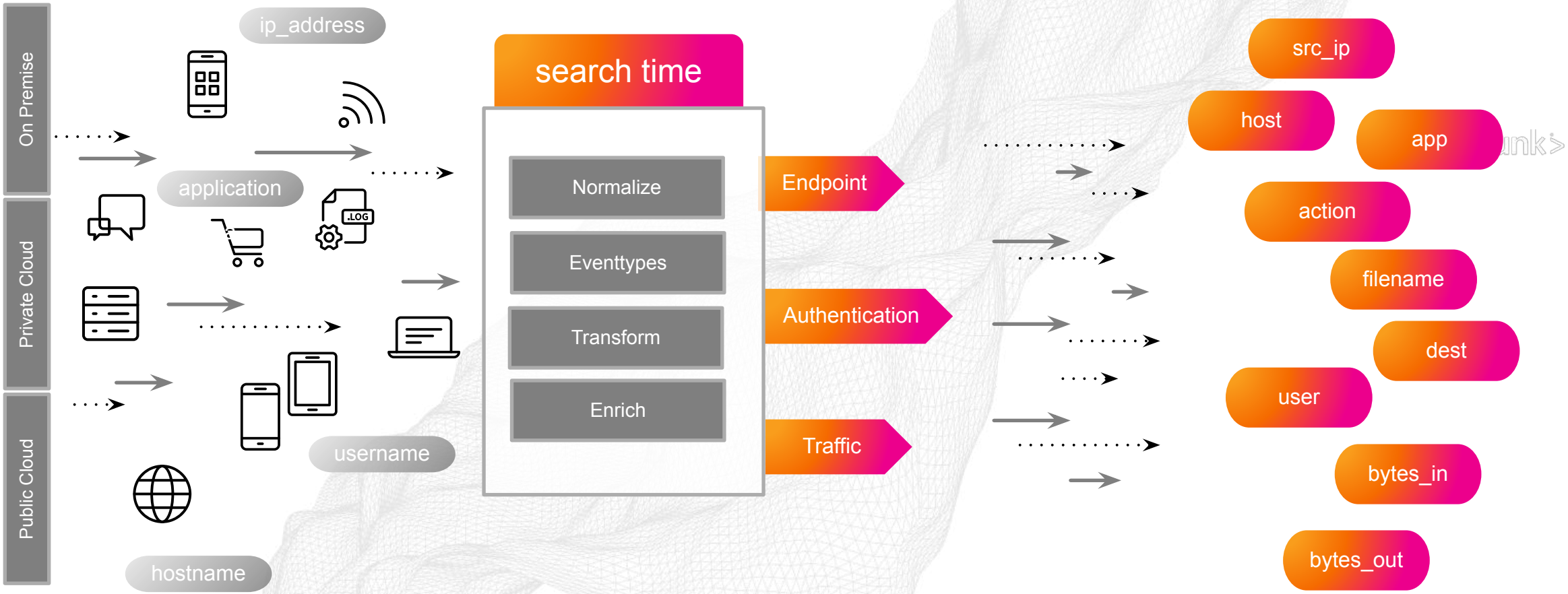
The Splunk CIM helps you to **normalize your data to match a common taxonomy** for data from multiple source vendors

The **Splunk CIM** is **packaged as an add-on** and comes with **pre-configured data models with a schema** in the form of tables that help you **parse your data at search time**

Learn more:
https://docs.splunk.com/Documentation/CIM/latest/User/Overview

Splunk Common Information Model

Normalize fields and events across all data sources



splunk>

https://youtu.be/nlf7Zjhwhgw

splunk>

Splunk Common Information Model (CIM)

The What and Why of CIM

CIM: common information model for normalizing data across disparate data sources; allows for **easy correlations**/mapping to data models
https://docs.splunk.com/Documentation/CIM/latest/User/Overview

- The CIM helps you to normalize your data to match a common standard, using the **same field names and event tags** for equivalent events from different sources or vendors. The CIM acts as a search-time schema ("schema-on-the-fly") to allow you to define relationships in the event data while leaving the raw machine data intact.
- Normalized data from multiple different source types, allows Splunk users to develop reports, correlation searches, and dashboards to present a **unified view of a data domain**.
- The Splunk Common Information Model add-on **is packaged with Splunk Enterprise Security**.
- The app can be downloaded/installed from here: https://splunkbase.splunk.com/app/1621

splunk>

Why Splunk CIM?

Common taxonomy: Simpler searches

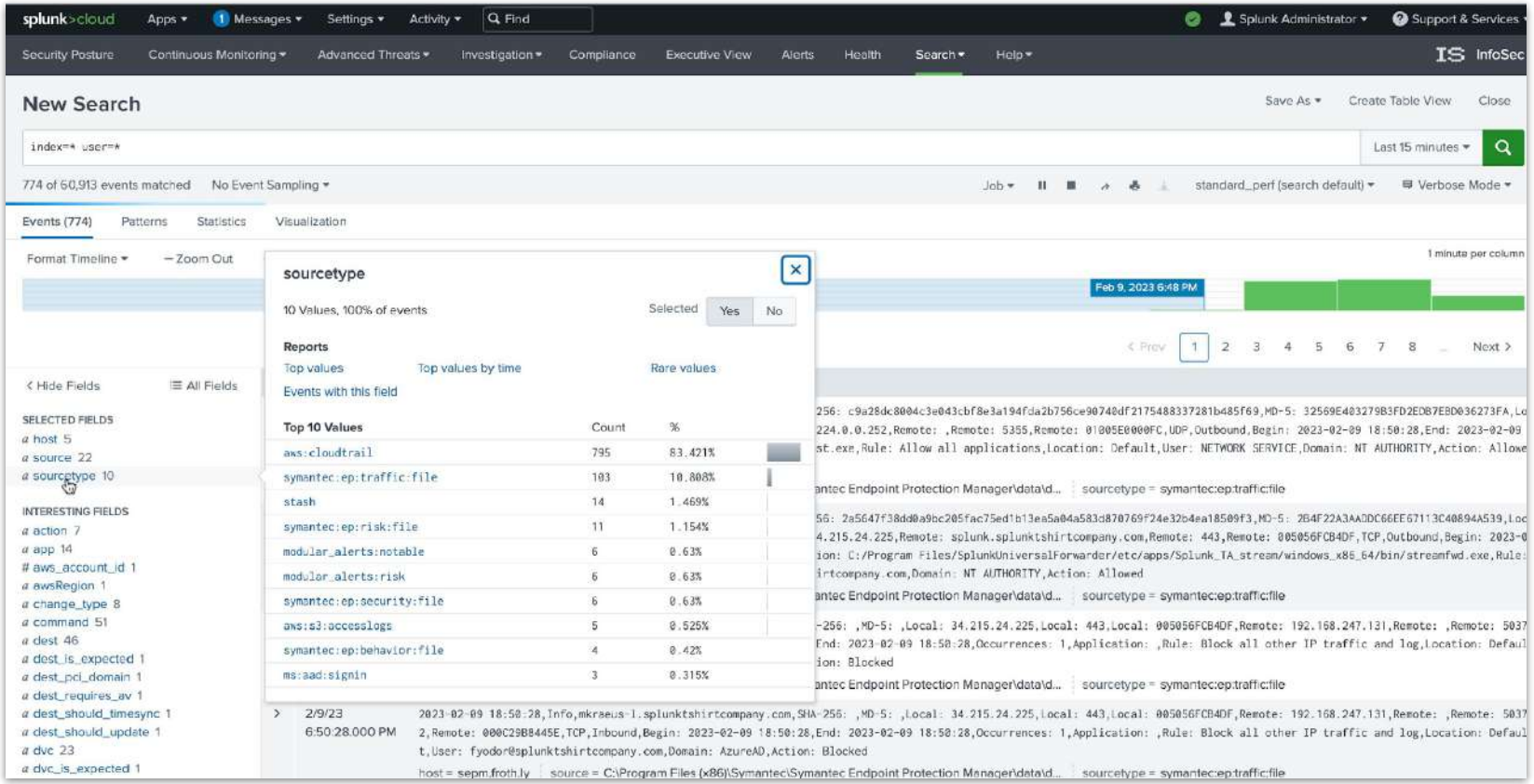
```
index=* user=*
```

vs

```
index=* (Username=* OR username=* OR user_name=* OR username=*)
```

Splunk CIM ***Authentication*** data model → ***user*** to identify an entity using a particular system, the “user.”

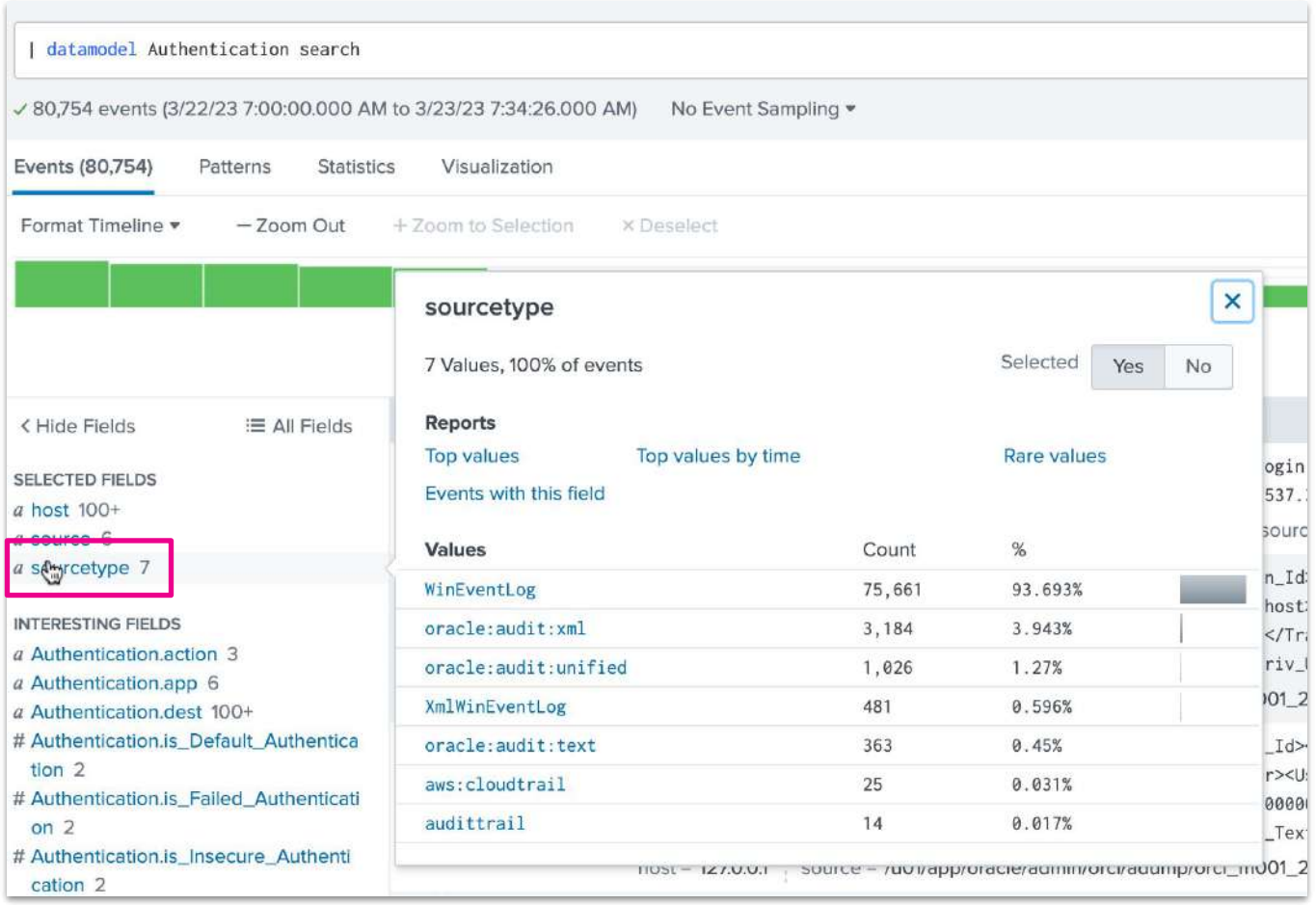
Cisco ASA → ***Username***
WinEventLog → ***userName***
Linux Secure → ***user_name***
Duo → ***username***



splunk>

Why Splunk CIM?

Searching the data model automatically correlates multiple data sources



Why Splunk CIM?

Faster and more performant searches: data models can be accelerated

Searching against an **accelerated data model:**

```
| tstats summariesonly=true distinct_count(Authentication.user) from datamodel=Authentication by sourcetype
```

Takes **0.6 seconds** to run

vs searching against the **index:**

```
index=* sourcetype=wineventlog | stats dc(userName) as count by sourcetype
| append
  [ search index=* sourcetype=cisco:asa | stats dc(Username) as count by sourcetype
  | append
    [ search index=* sourcetype=linux_secure | stats dc(user_name) as count by sourcetype
    | append
      [search index=* sourcetype=duo | stats dc(username) as count by sourcetype ] ] ]
```

Takes **over 65 seconds** to run (<100x)

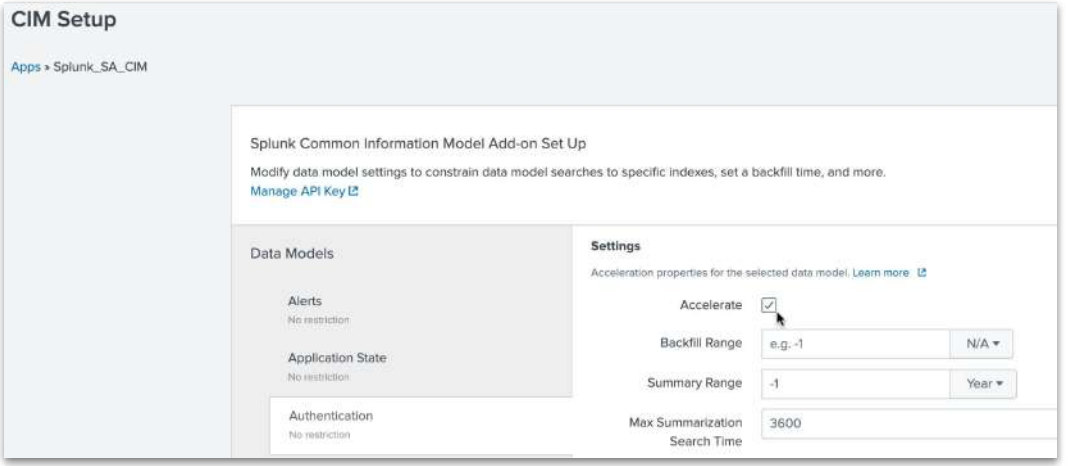
splunk>

Intro to (CIM) Data Models

What is a data model?

- A data model is a **hierarchically structured search-time mapping** of your data, and can be composed by one or more datasets.

What is data model acceleration?



- It is a Splunk **feature** used to **speed up data models** that represent extremely large datasets
- Using searches that query accelerated Data Models in dashboards improves search performance
- When enabled, it will create a *summary* table with the key=value pairs
- Searches against a data model that is accelerated, e.g.

```
| tstats summariesonly=true from datamodel=Malware...
```

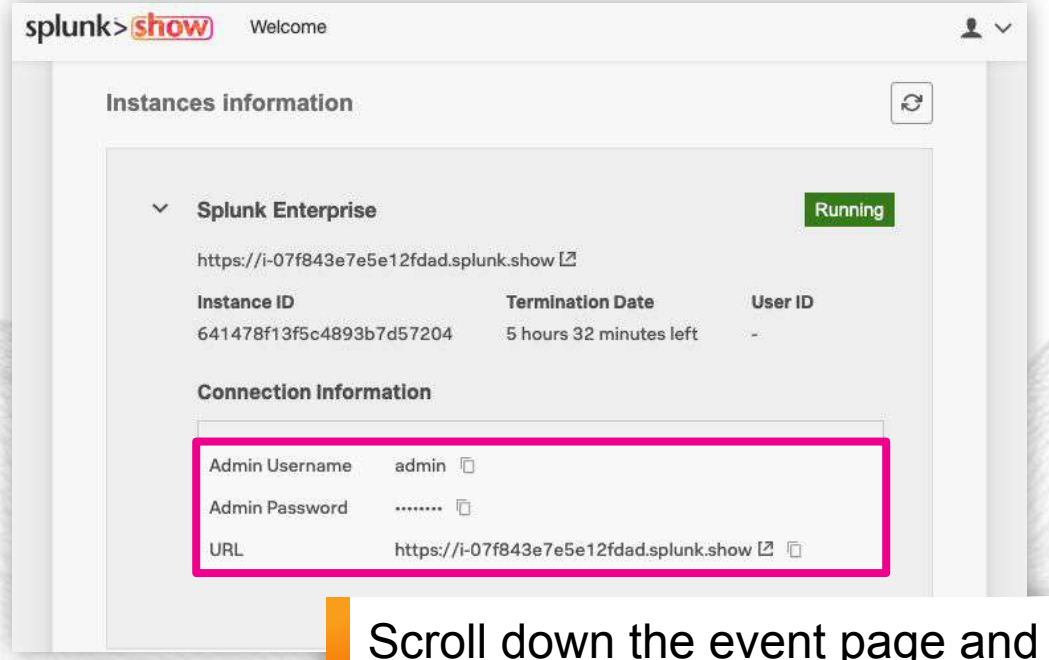
  will search against the *summary* table, and will run faster.
- Requires index whitelisting
- This is set up in the CIM add-on

Learn more: https://docs.splunk.com/Documentation/Splunk/latest/Knowledge/Aboutdatamodels

splunk>

Task 2: Log in to Splunk

Locate your instance URL and credentials in the Splunk Show event https://show.splunk.com



Scroll down the event page and expand the **Splunk Enterprise** section to view your login details

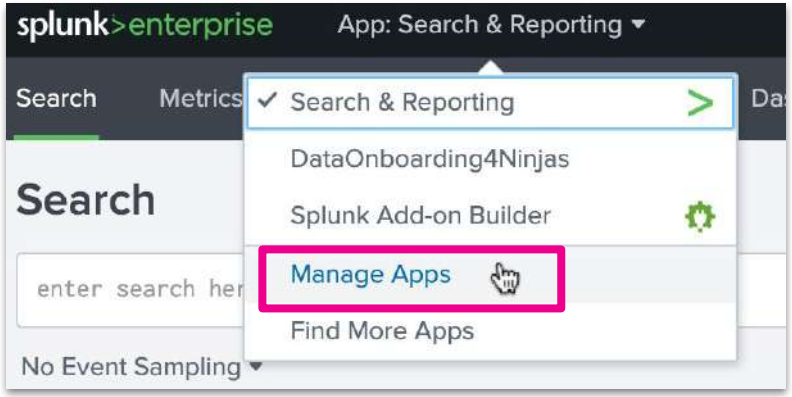
Log in to your Splunk instance

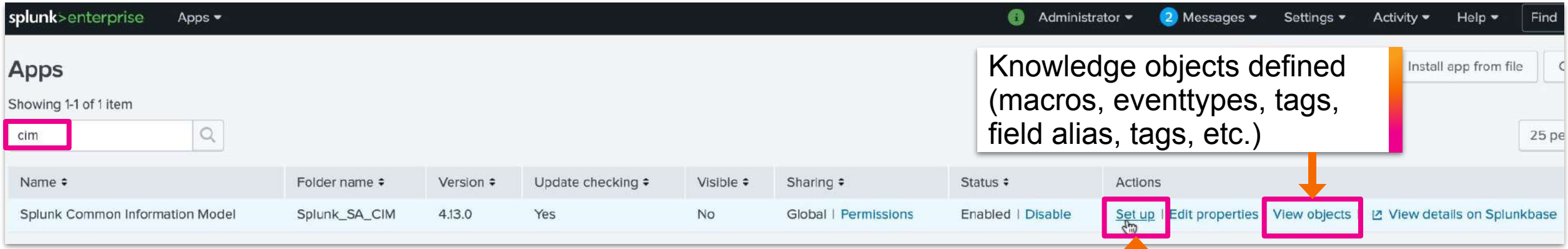
Login using the credentials from Splunk Show

splunk>

Task 3: Explore the CIM add-on

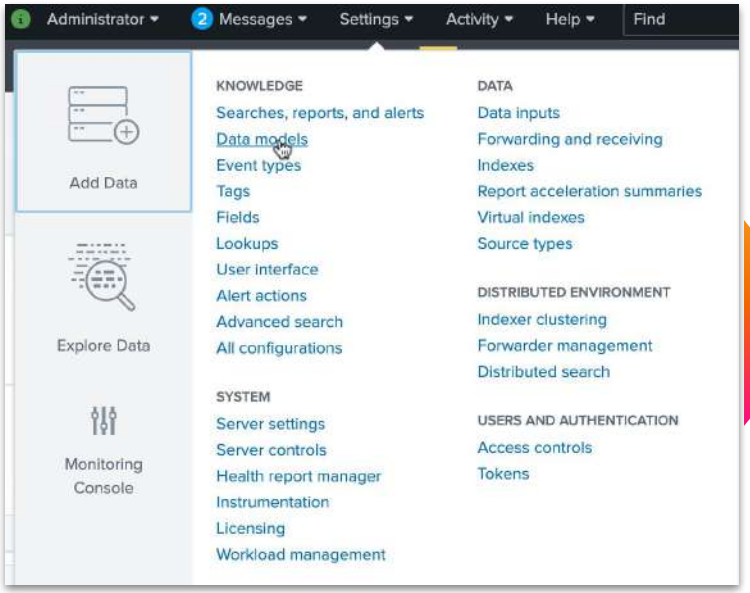
Explore CIM Set up, view objects, and explore CIM data models under settings

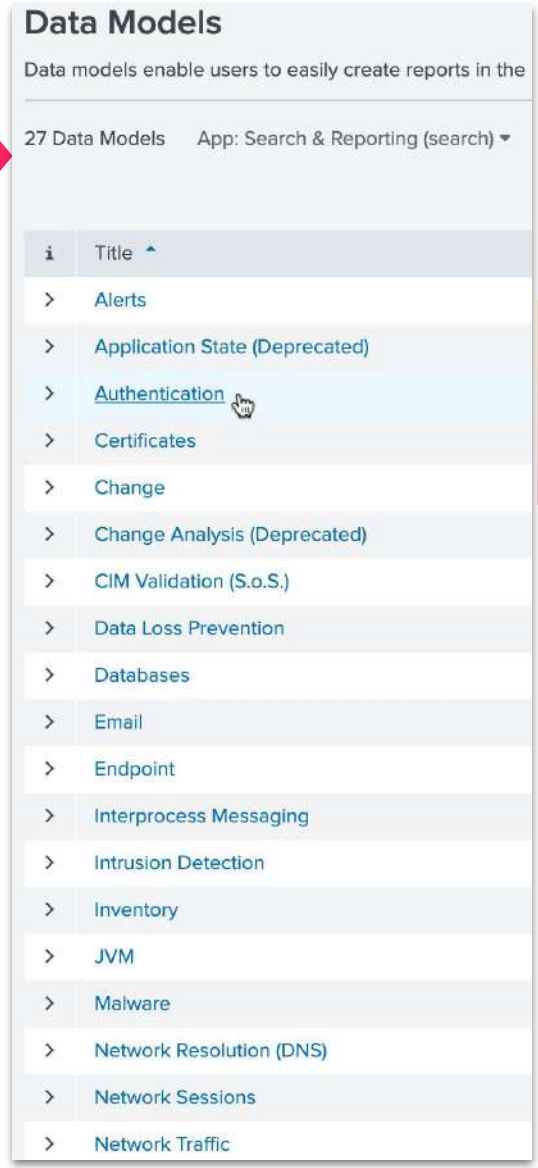


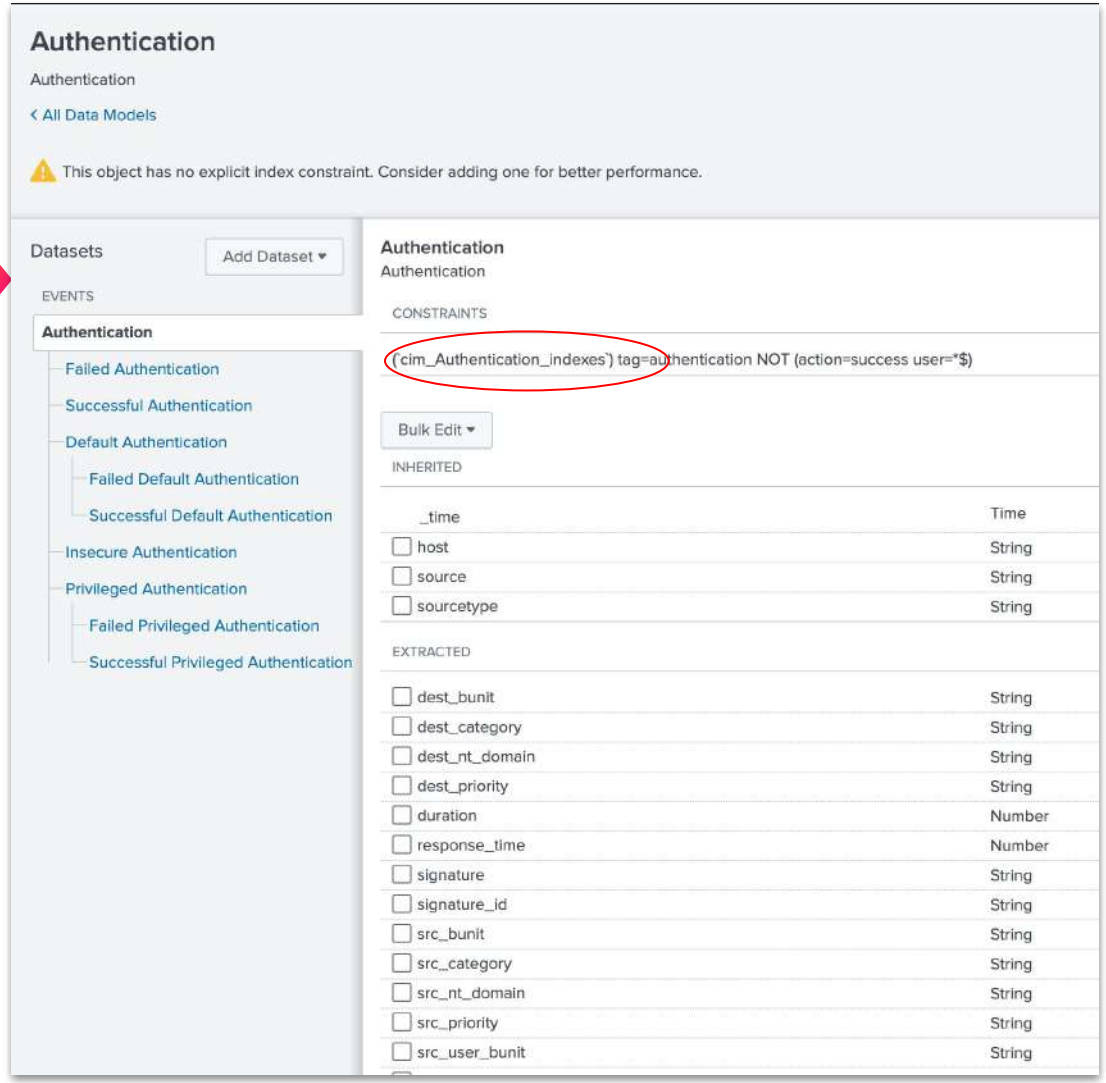


Learn more:
https://docs.splunk.com/Documentation/CIM/latest/User/Howtousethesereferencetables
https://docs.splunk.com/Documentation/CIM/latest/User/CIMfields
https://docs.splunk.com/Documentation/CIM/latest/User/Authentication

splunk>



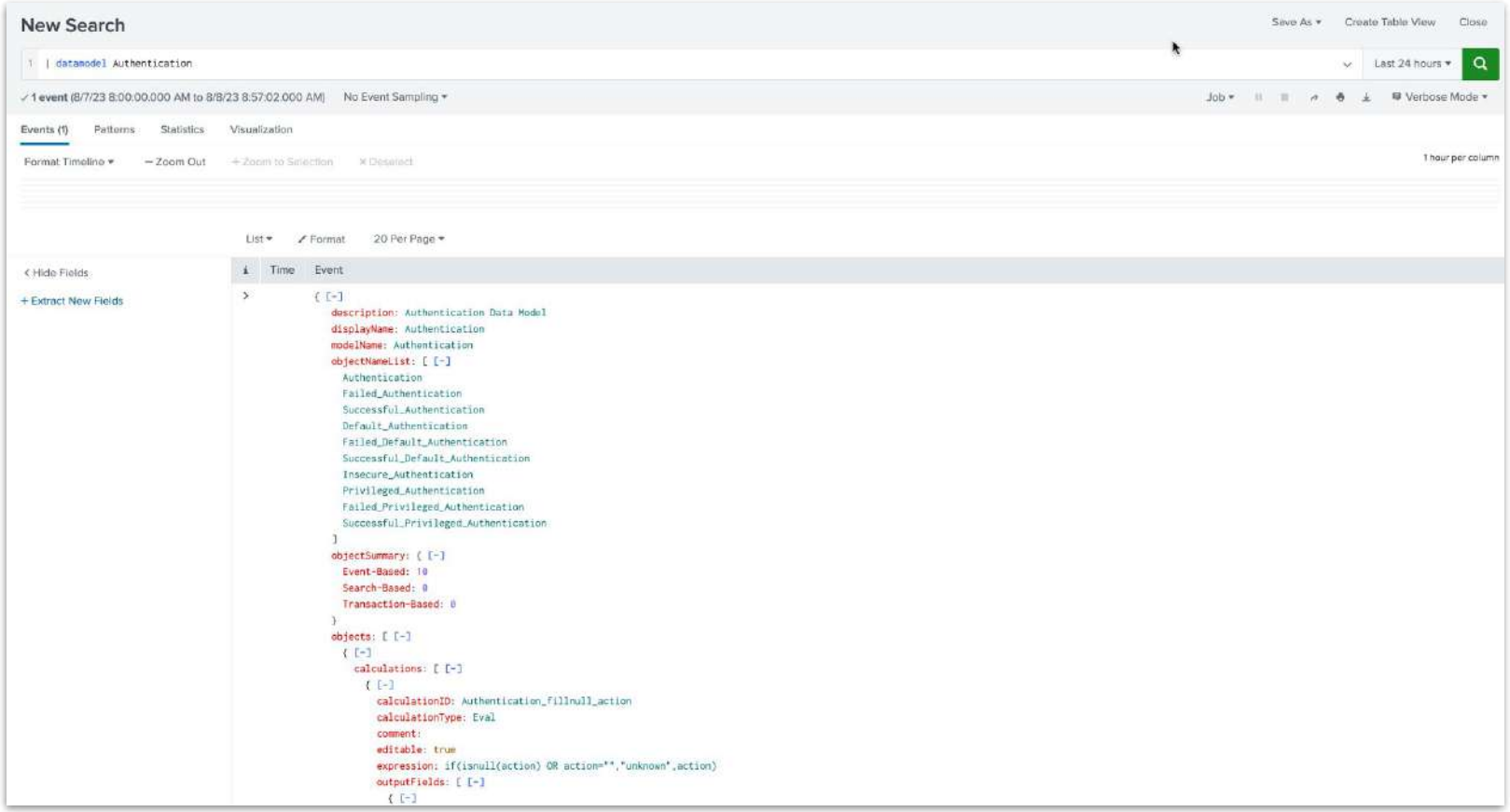




Task 4: Explore the data model using SPL

Explore the data model's data sets and schema in json format using the Search app

```
| datamodel "Authentication"
```



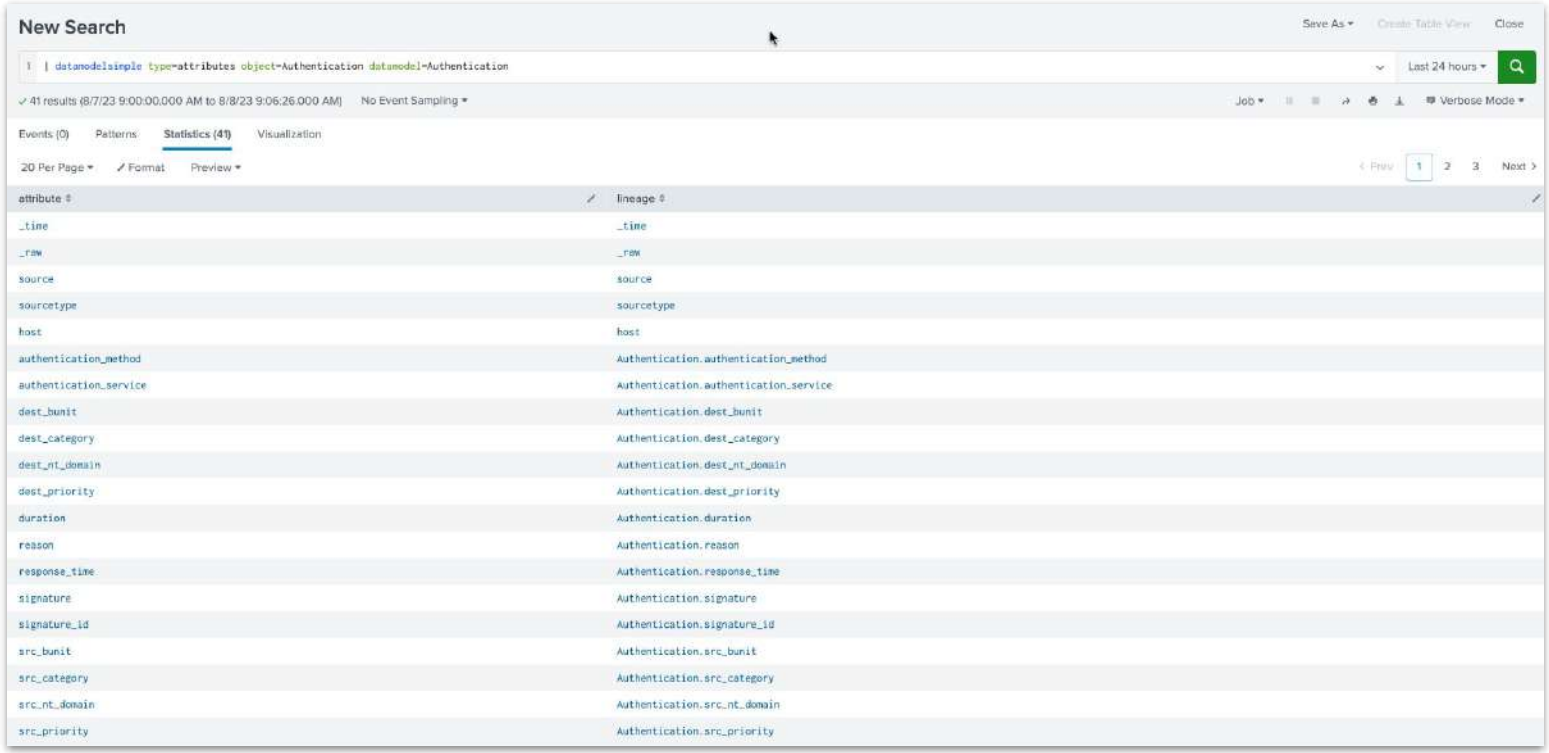
Learn more:
https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Datamodel

splunk>

Task 4: Explore the data model using SPL

Explore the data model's data sets' fields in table view using the search app

```
| datamodelsimple type=attributes object=Authentication datamodel=Authentication
```



Learn more:
https://docs.splunk.com/Documentation/CIM/latest/User/UsetheCIMtovalidateyourdata#Use_the_datamodelsimple_command

splunk>

Practice Lab

Lab 1: GDI CIM-compliant Data

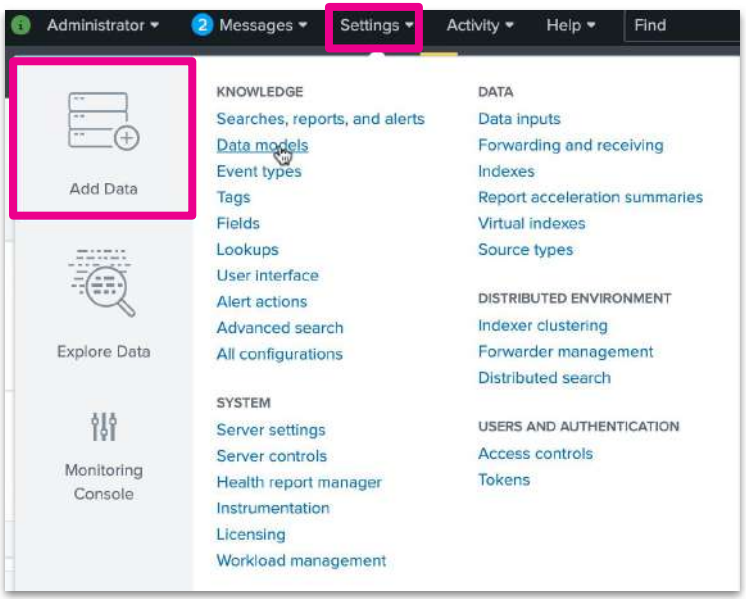
CIM-compliant Add-on from Splunkbase

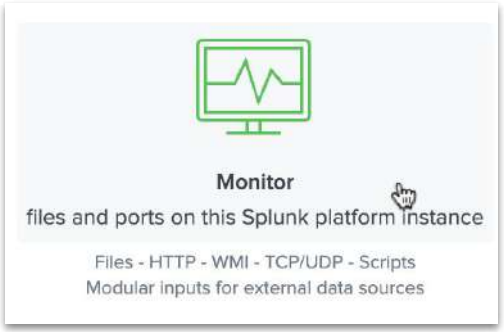
Fortinet FortiGate Firewall Traffic Logs

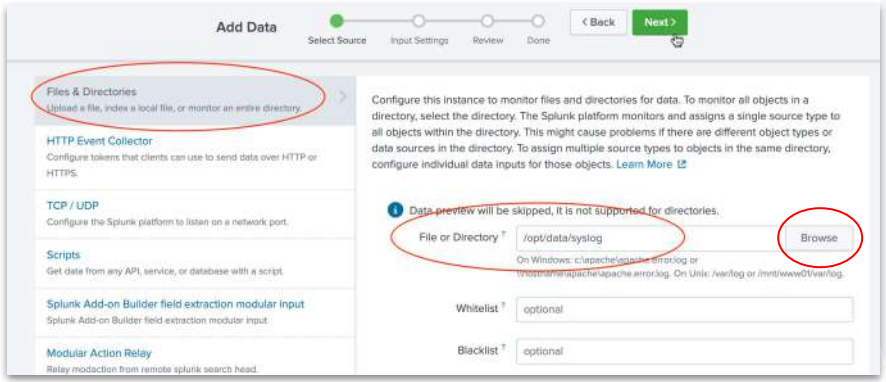
splunk>

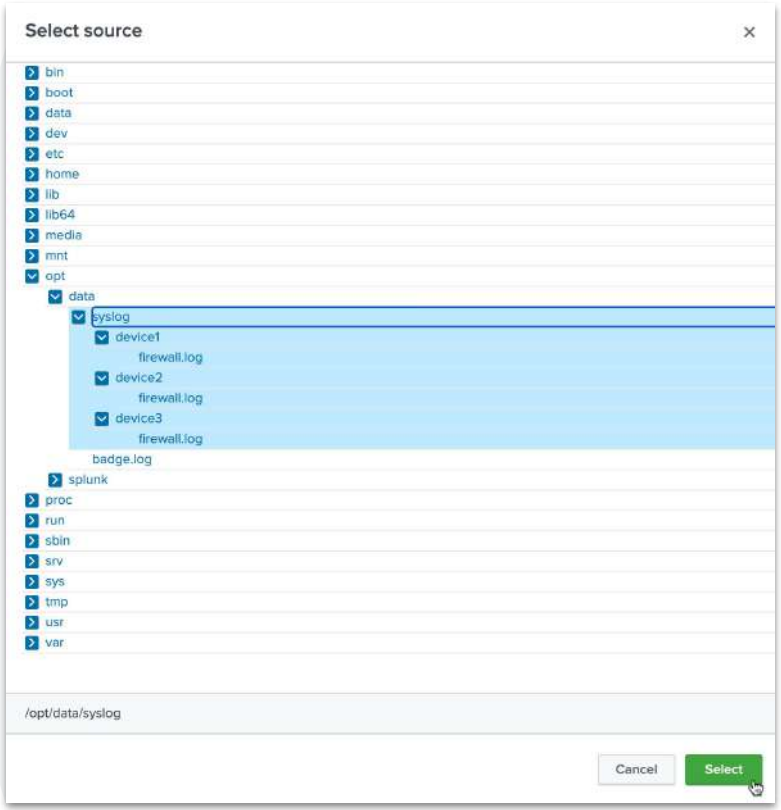
Task 5: Ingest Fortinet Traffic Logs

1. Create a monitoring input for firewall.log







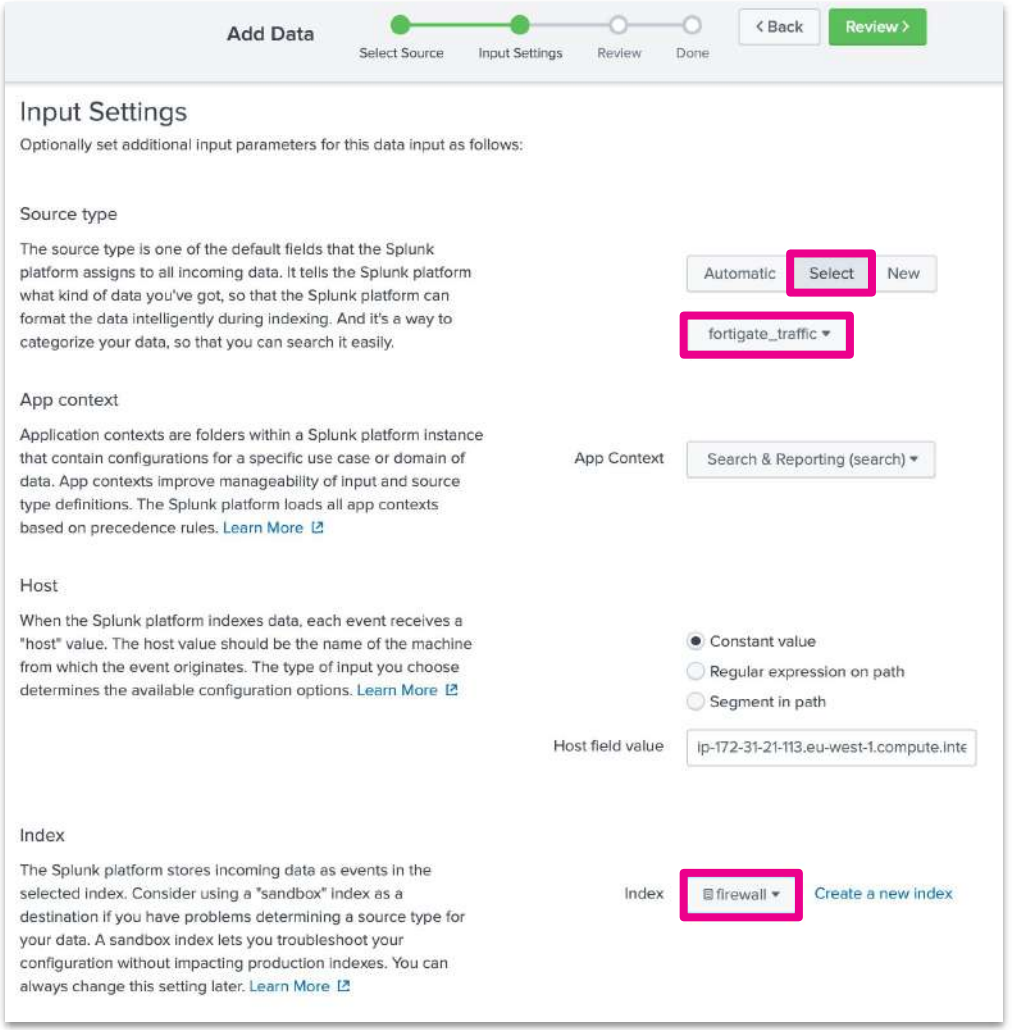


splunk>

LAB 1

Task 5: Ingest Fortinet Traffic Logs

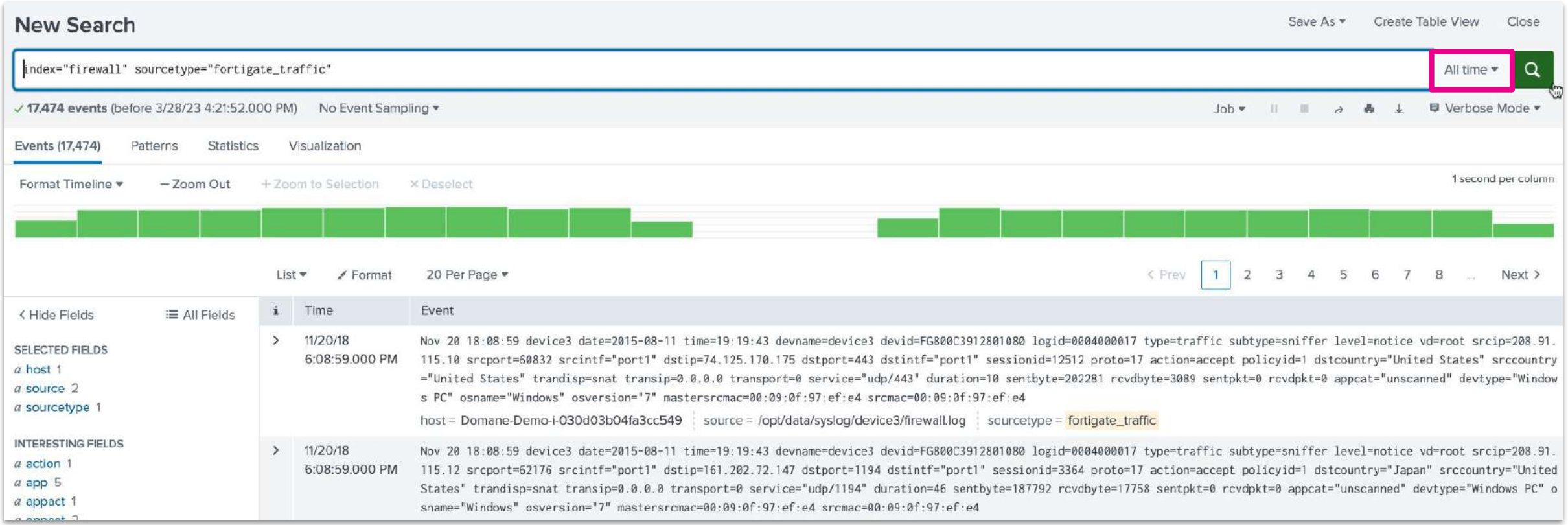
2. Create a monitoring input for firewall.log



Task 5: Ingest Fortinet Traffic Logs

3a. Check your data!

```
index="firewall"sourcetype="fortigate_traffic"
```

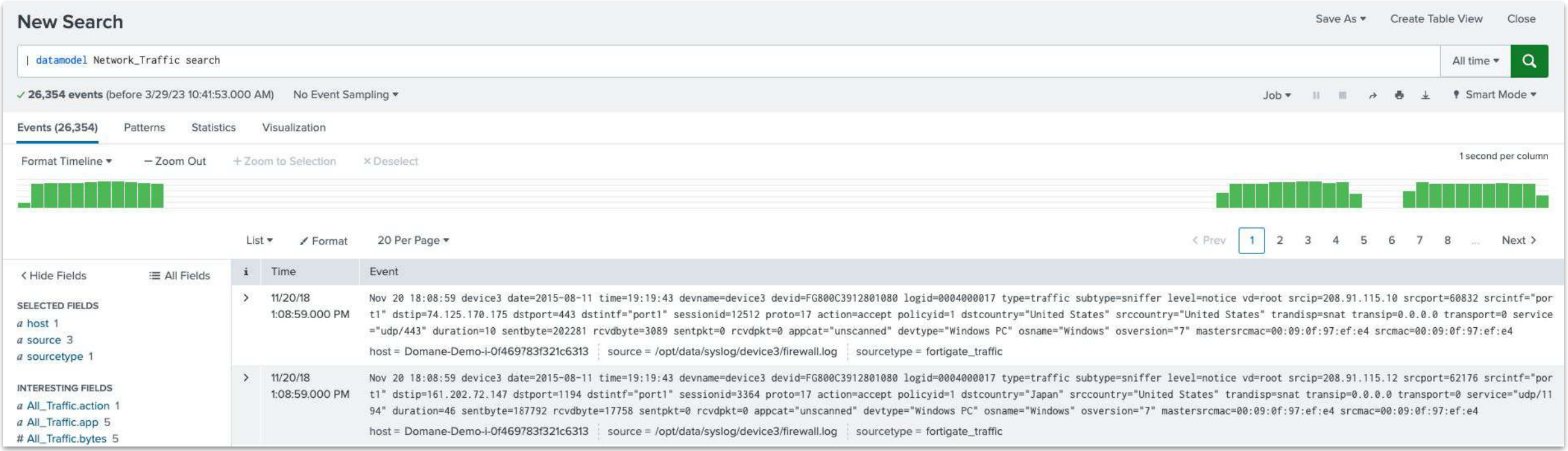


LAB 1

Task 5: Ingest Fortinet Traffic Logs

3b. Another method to check your data!

```
| datamodel Network_Traffic search
```

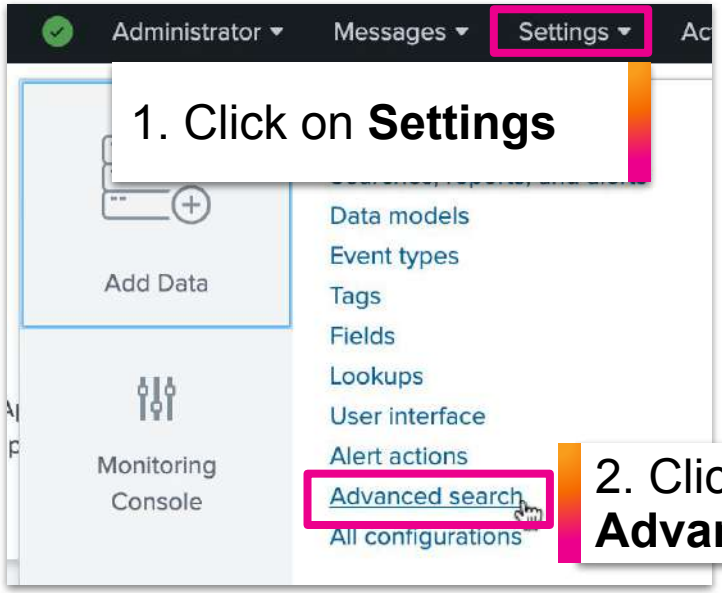


splunk>

LAB 1

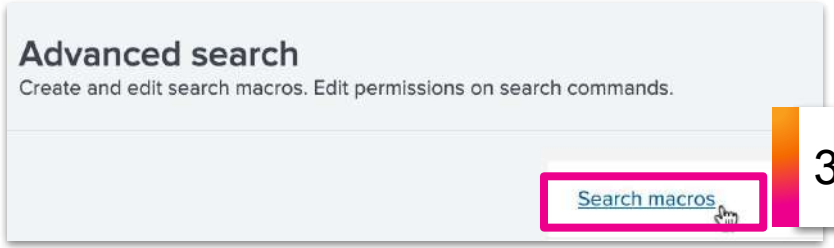
Task 6: Accelerate the Network_Traffic Data Model in CIM

1. Set the data model's indexes

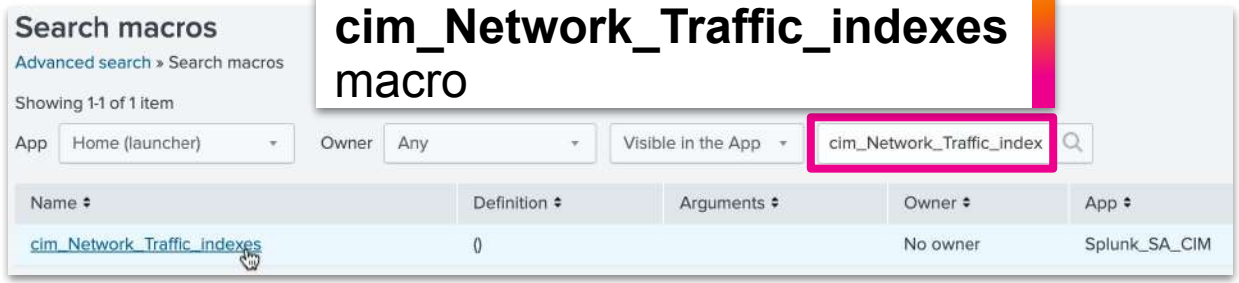


1. Click on **Settings**

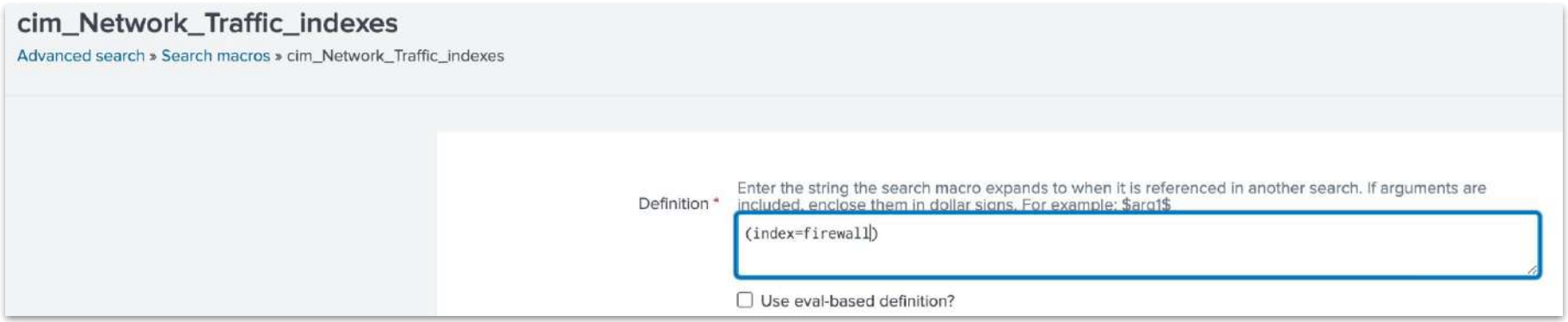
2. Click on **Advanced search**



3. Click on **Search macros**



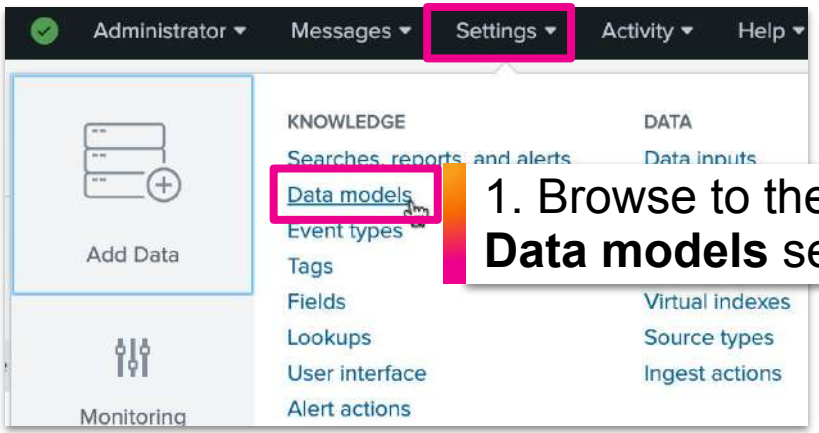
4. Filter by the **cim_Network_Traffic_indexes** macro



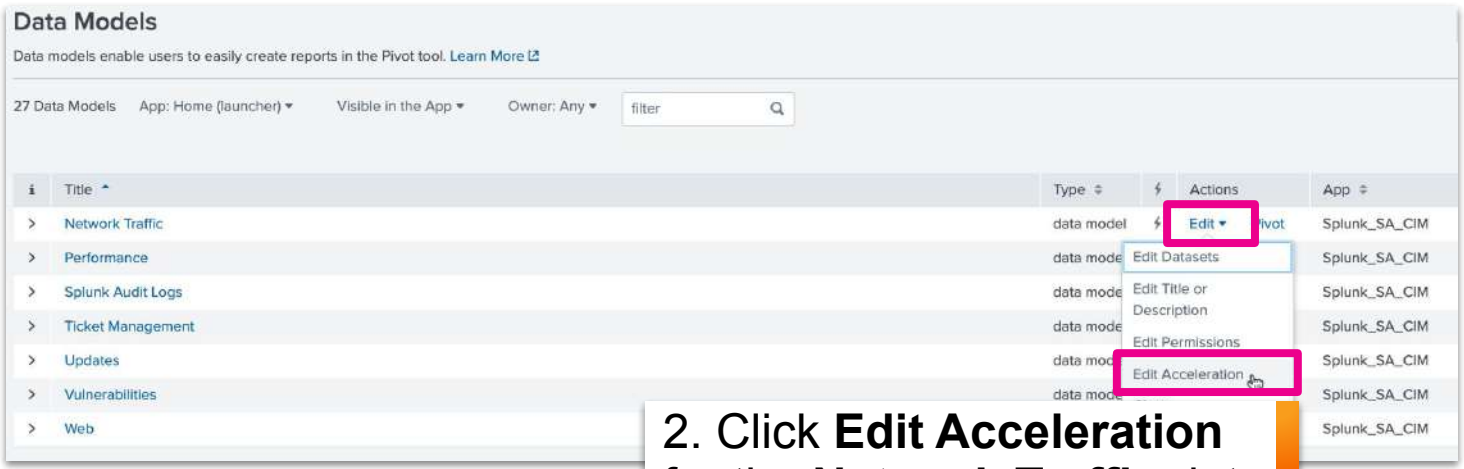
splunk>

Task 6: Accelerate the Network_Traffic Data Model in CIM

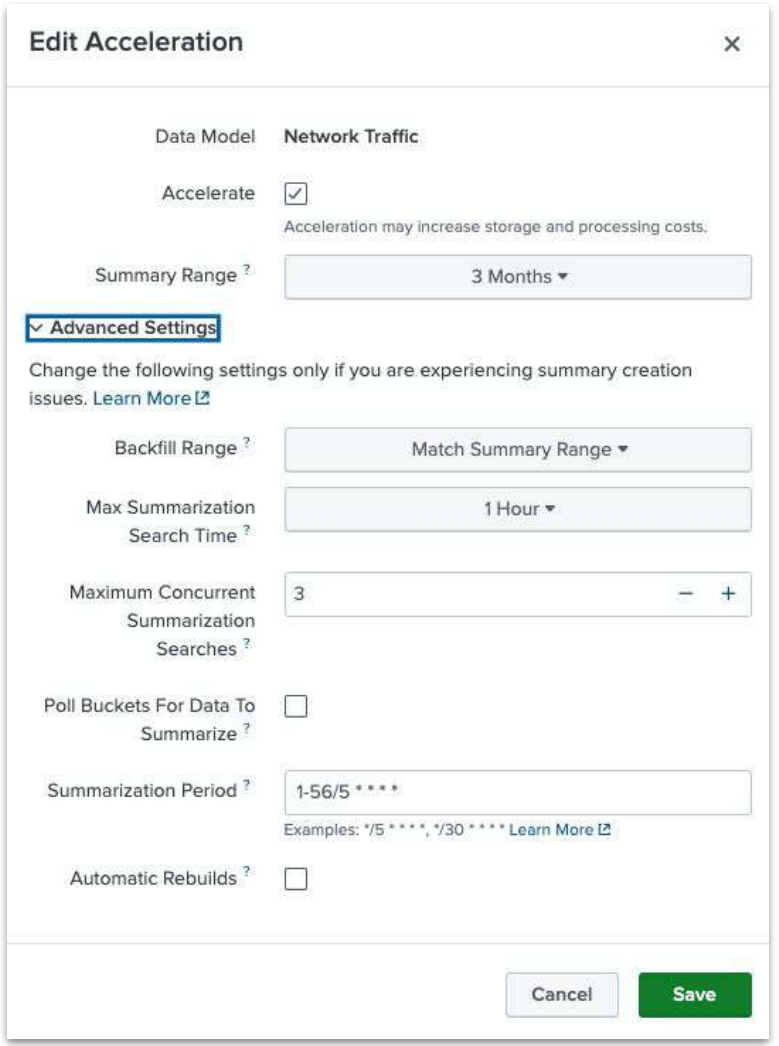
2. Edit the data model's acceleration settings



1. Browse to the **Data models** settings



2. Click **Edit Acceleration** for the **Network Traffic** data model



splunk>

LAB 1

Task 7: Search for Traffic…

So…what's the difference between…?

```
index=firewall sourcetype="fortigate_traffic" | stats count
```

```
| datamodel Network_Traffic search | stats count
```

```
| tstats count from datamodel=Network_Traffic
```

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

CIM Compliance

How do we do it ?

1

Get the right CIM fields/values

2

Create an eventtype to regroup our events

3

Tag our eventtypes

Learn more:
https://docs.splunk.com/Documentation/CIM/latest/User/UsetheCIMtonormalizedataatsearchtime

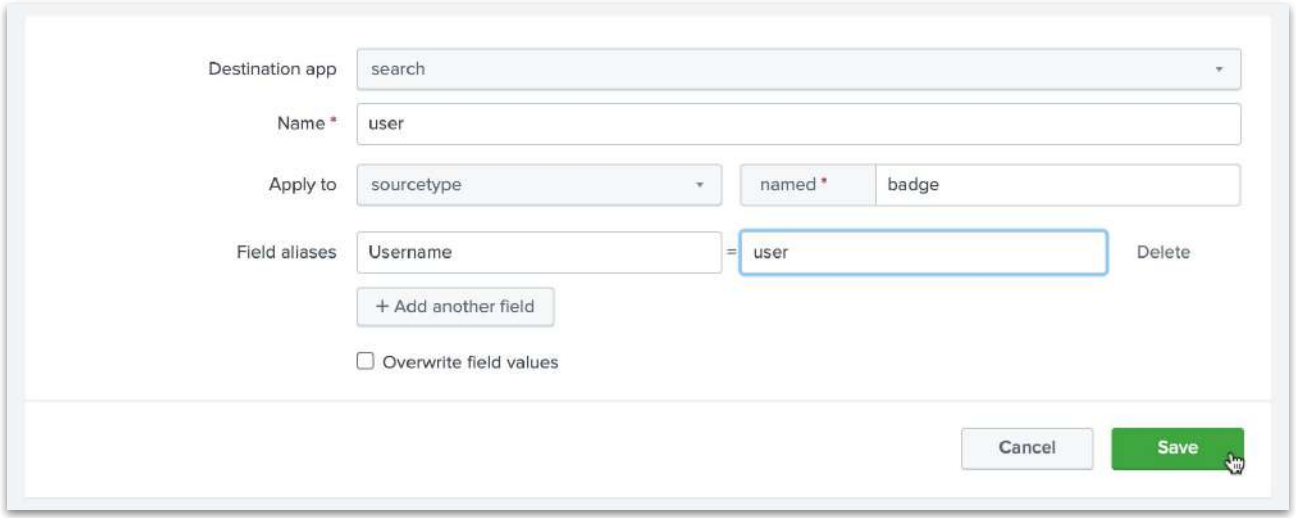
CIM Compliance

Get the right CIM fields/values

My field name is different?

My field	CIM field
Username	user

Use field alias: **Settings > Fields > Field Alias**



props.conf

```
[badge]
FIELDALIAS-user = Username as user
```

splunk>

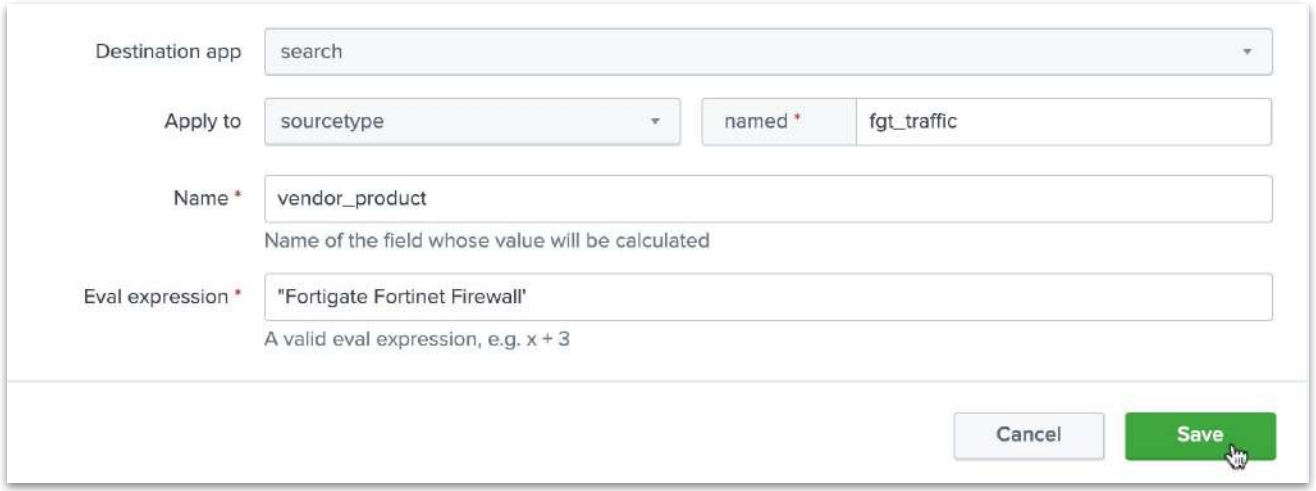
CIM Compliance

Get the right CIM fields/values

I don’t have this field in my event

My field	→	CIM field
?		vendor_product

Use calculated fields: **Settings > Fields > Calculated Field**



props.conf →

```
[fgt_traffic]
EVAL-vendor_product = “Fortinet Fortigate Firewall”
```

splunk>

CIM Compliance

Get the right CIM fields/values

My field value is not CIM compliant? Use a lookup!

Values in my data

result	action
Successful authentication	success
Wrong Password	failure
Locked Account	failure

Prescripted values in CIM

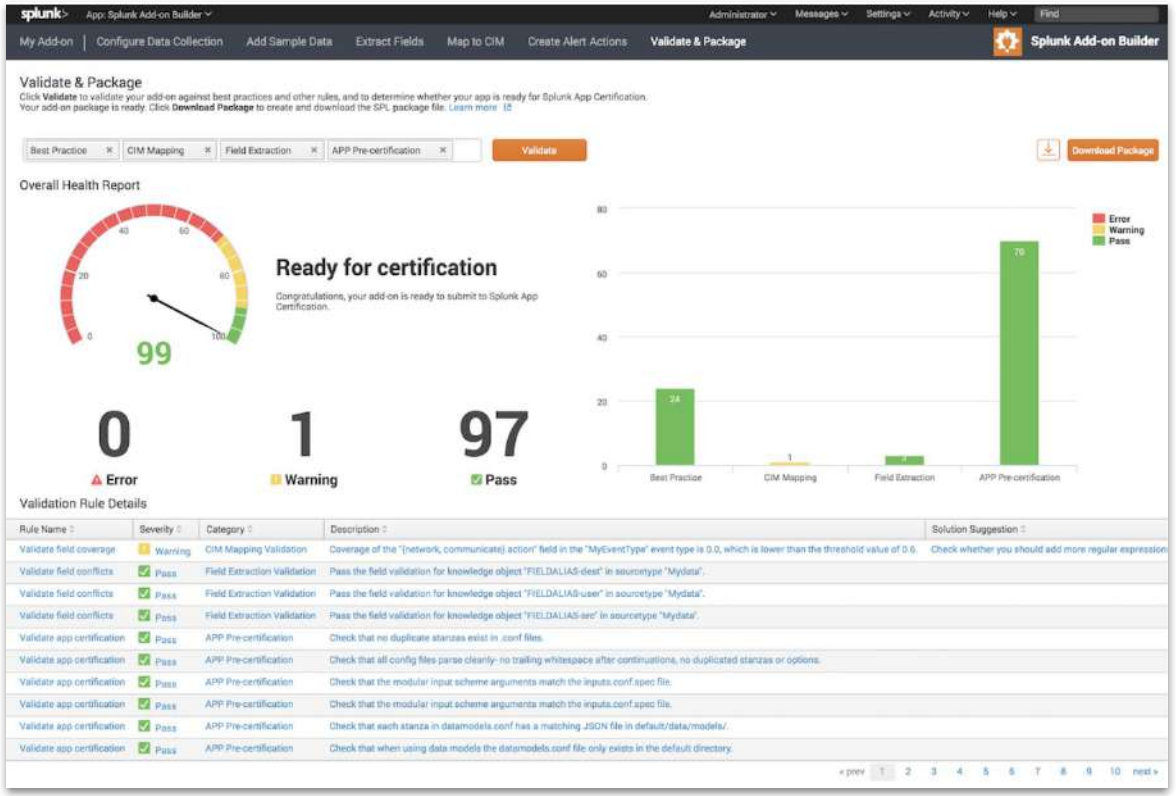

```
<your base search> | lookup result OUTPUTNEW action
```

splunk>

Need Help ?

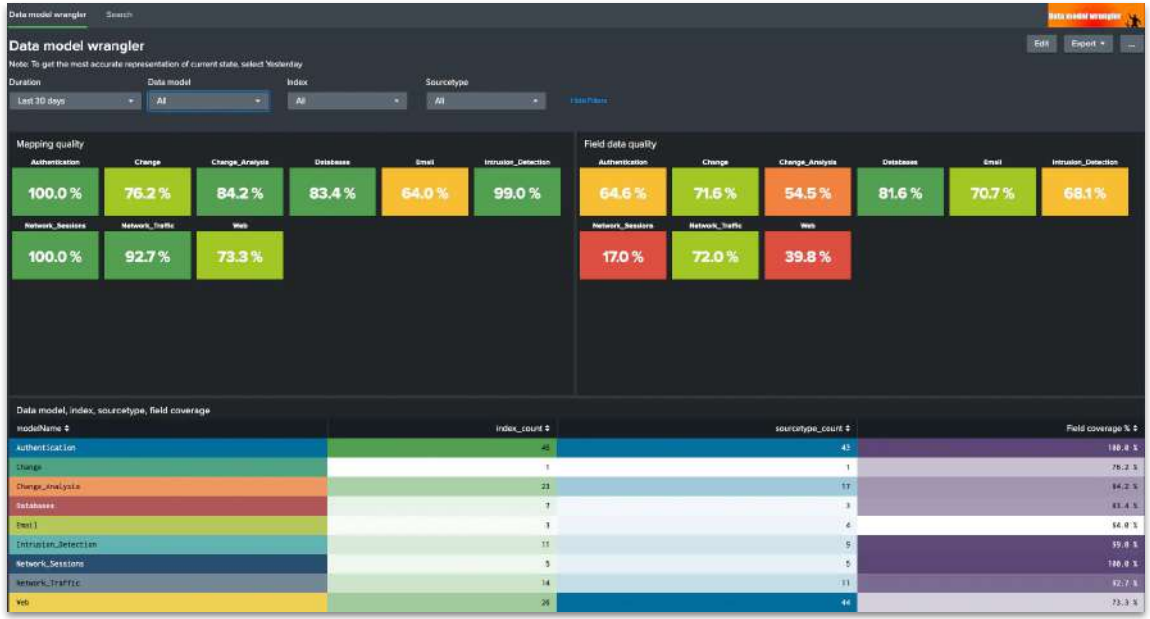
Build: Splunk Add-on Builder

https://splunkbase.splunk.com/app/2962/



Validate: Data Model Wrangler

https://splunkbase.splunk.com/app/6286



splunk>

Practice Lab

Lab 2: GDI Badge Data

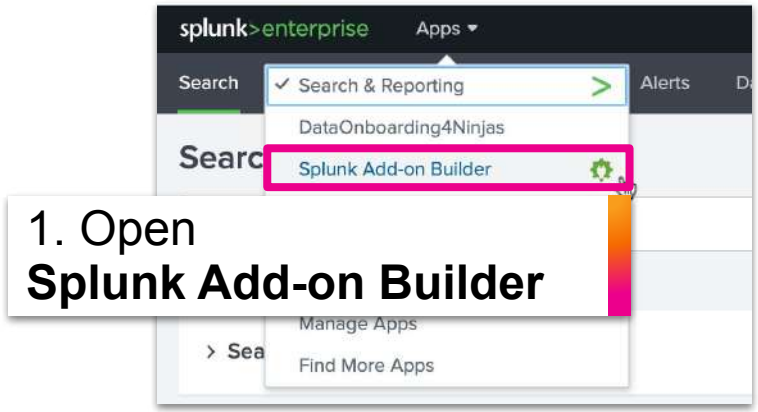
Custom “badge” sourcetype

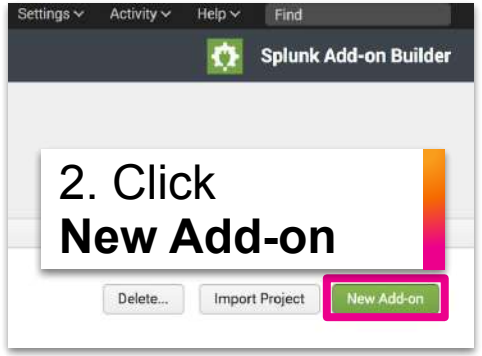
Create a custom app for the “badge” sourcetype

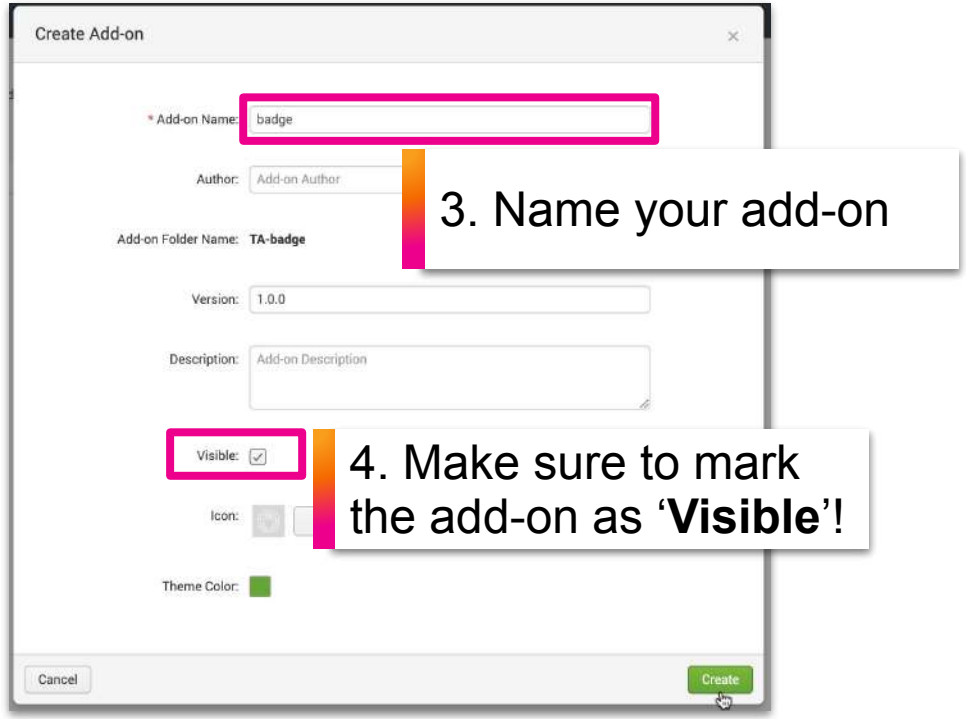
splunk>

Task 8: Create a Custom Add-on

1. Go to the **Splunk Add-on Builder** app. We need to create a custom add-on for the new custom sourcetype



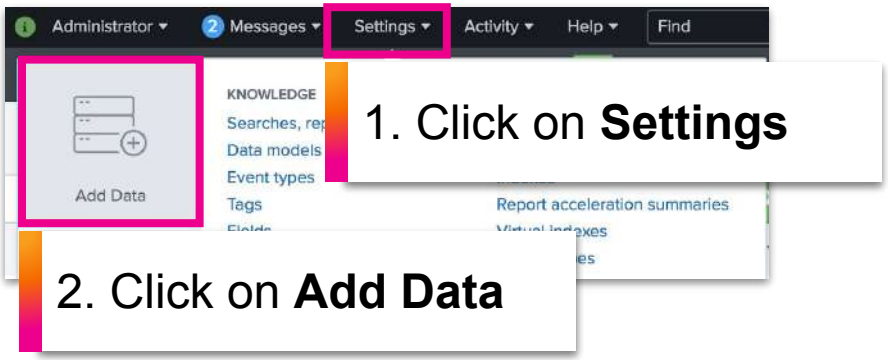


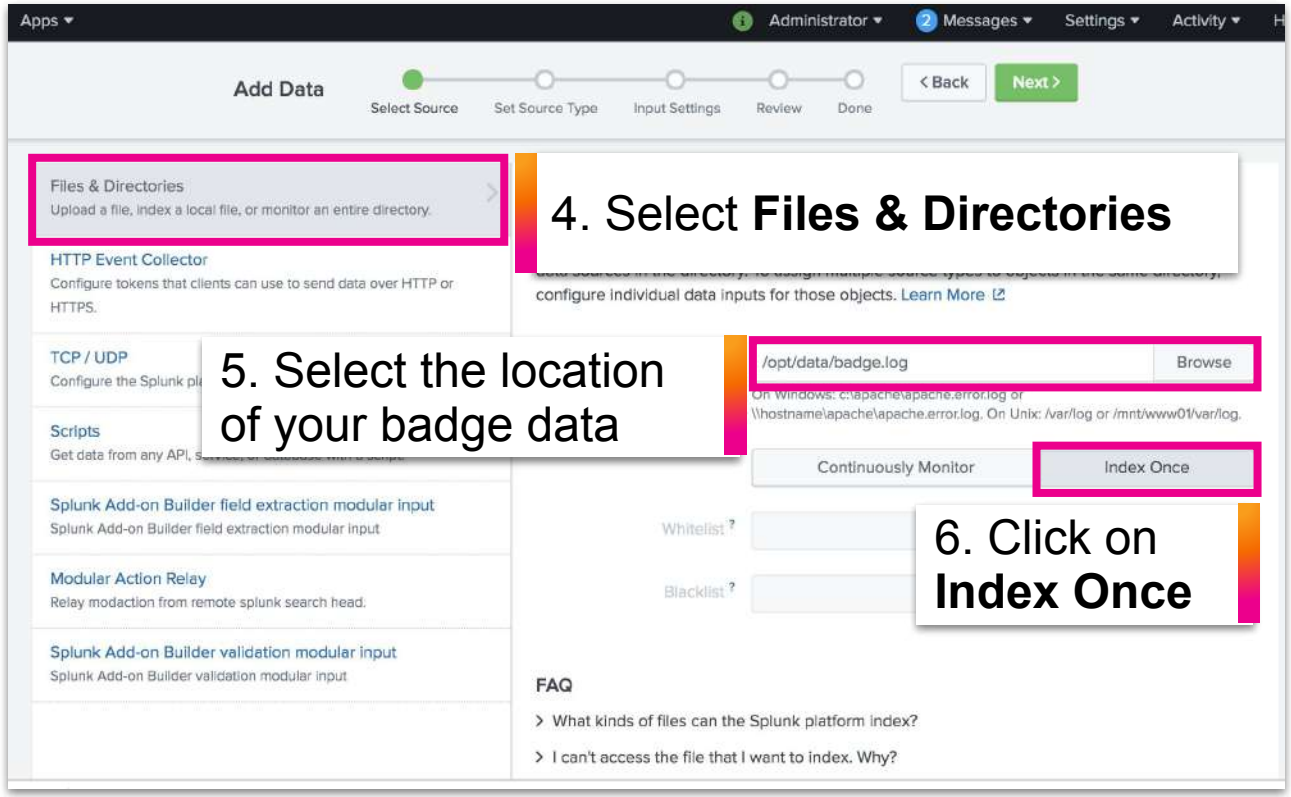


LAB 2

Task 9: Add Badge Data

2. Click on **Settings > Add Data > Monitor > File & Directories** (index once)

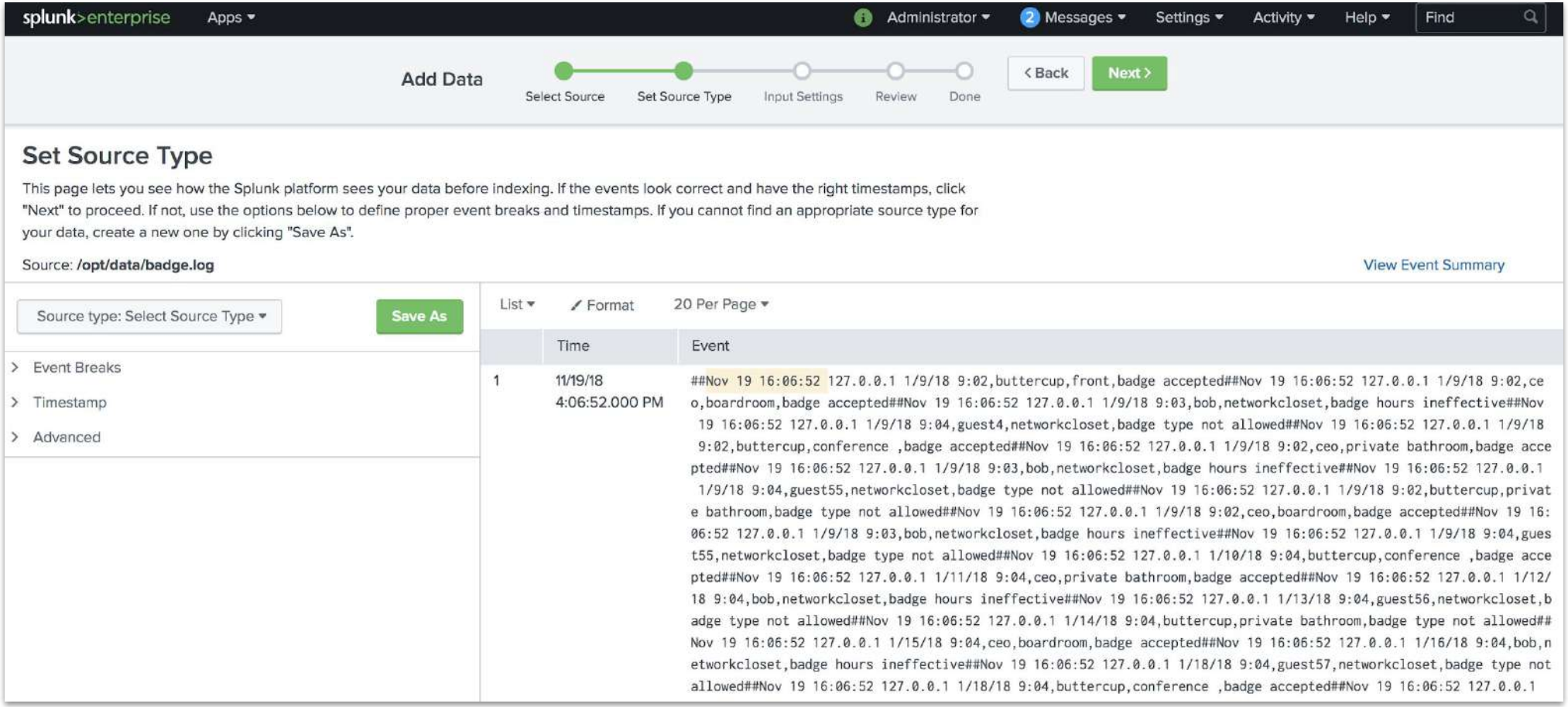




splunk>

Task 9: Add Badge Data

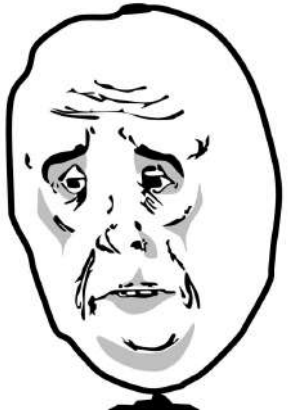
Set a Custom Source Type

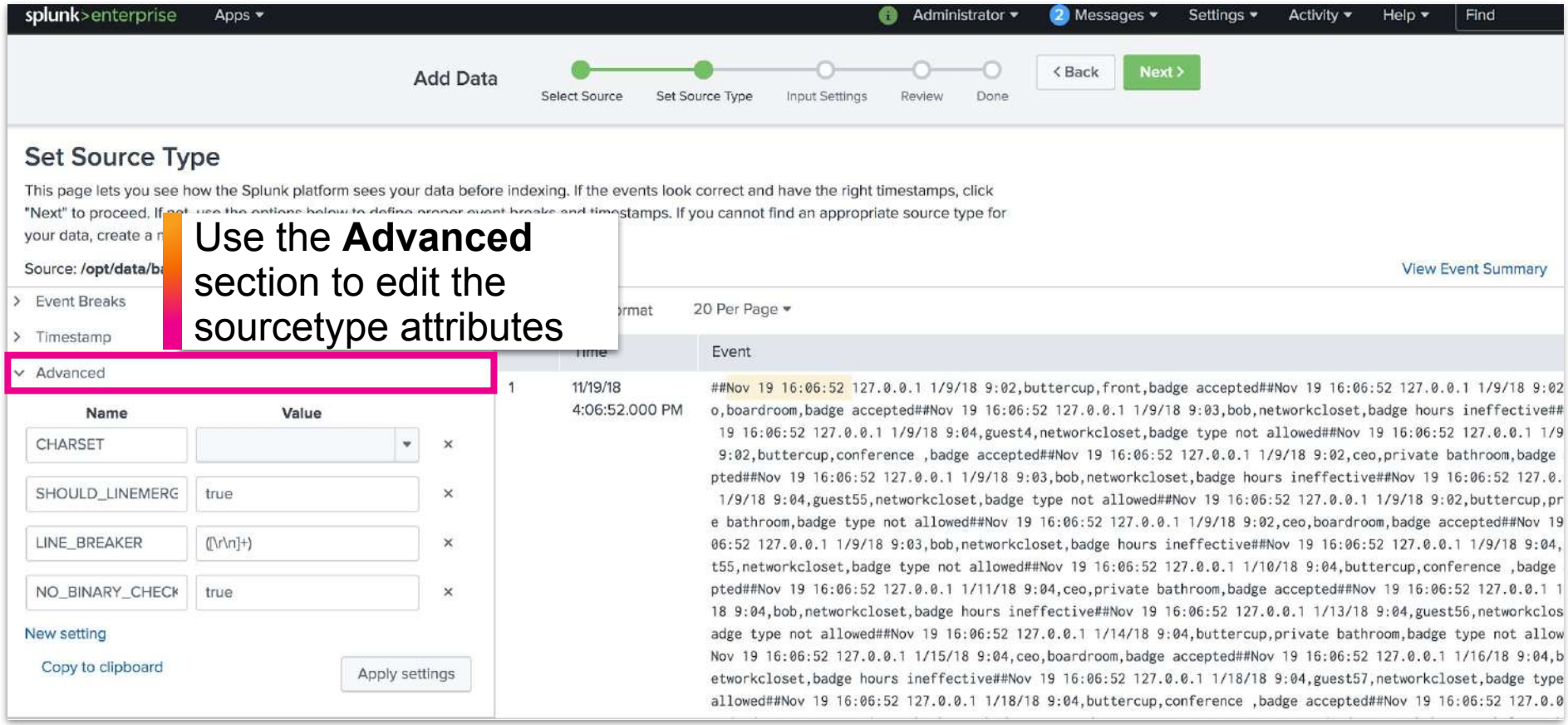


splunk>

Task 9: Add Badge Data

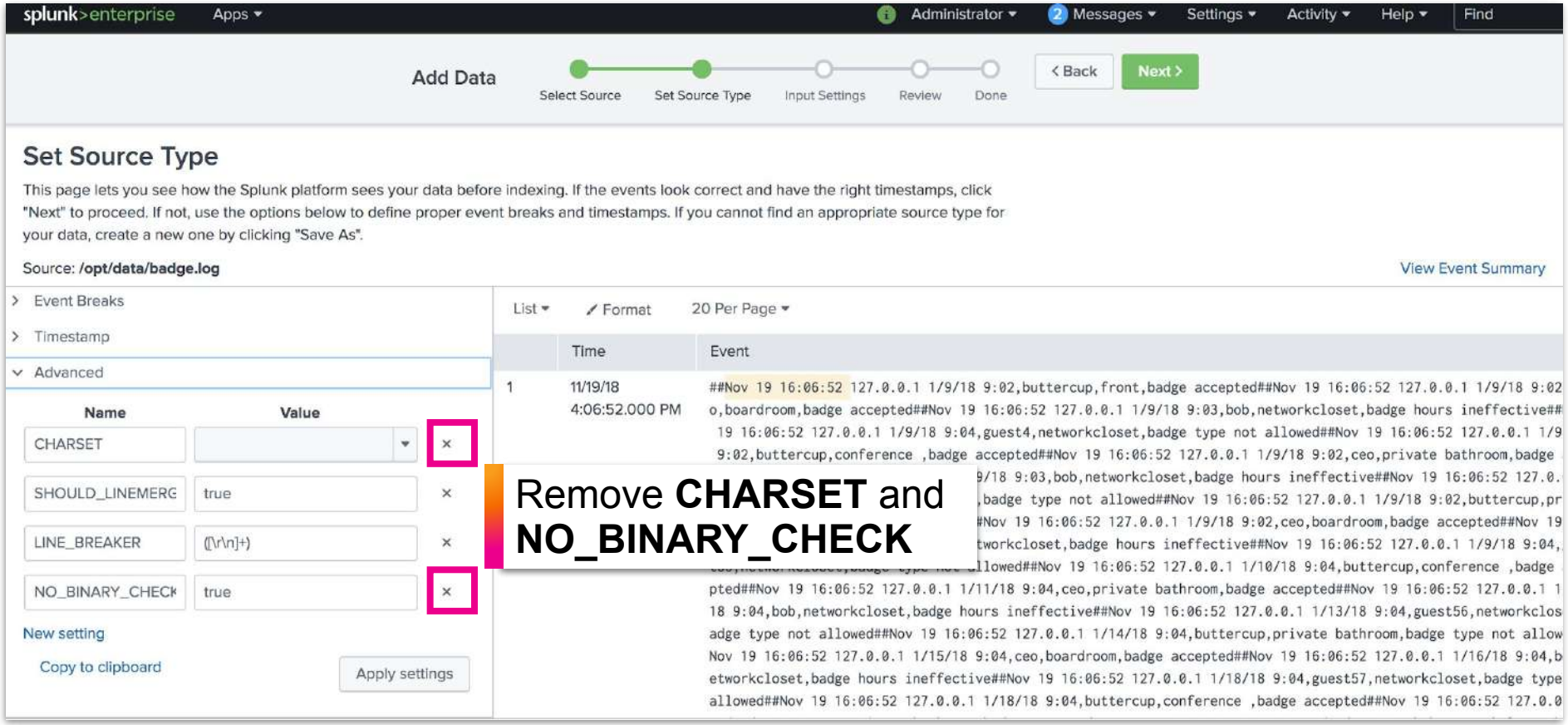
Event breaking failed… why ?





Task 9: Add Badge Data

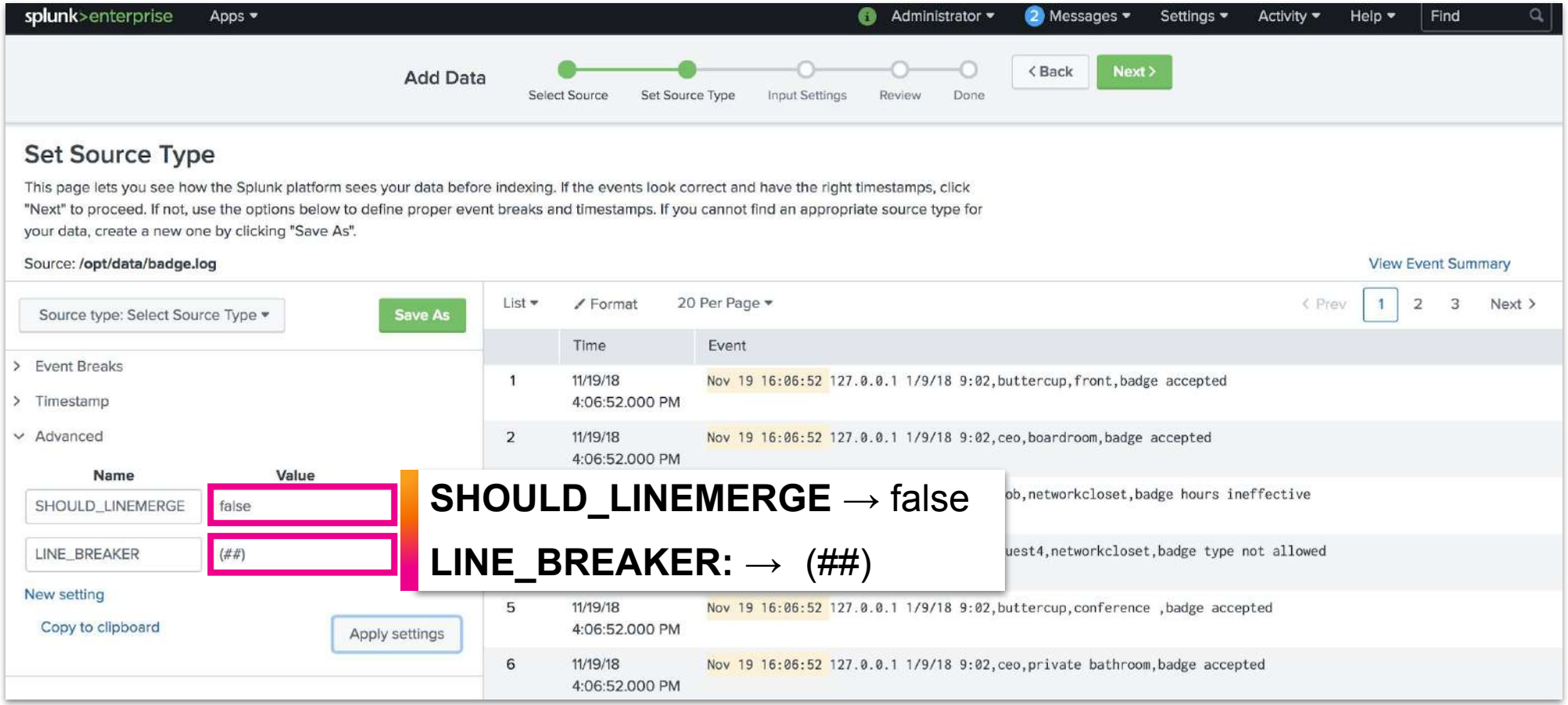
3. Set custom source type: Remove the default settings



splunk>

Task 9: Add Badge Data

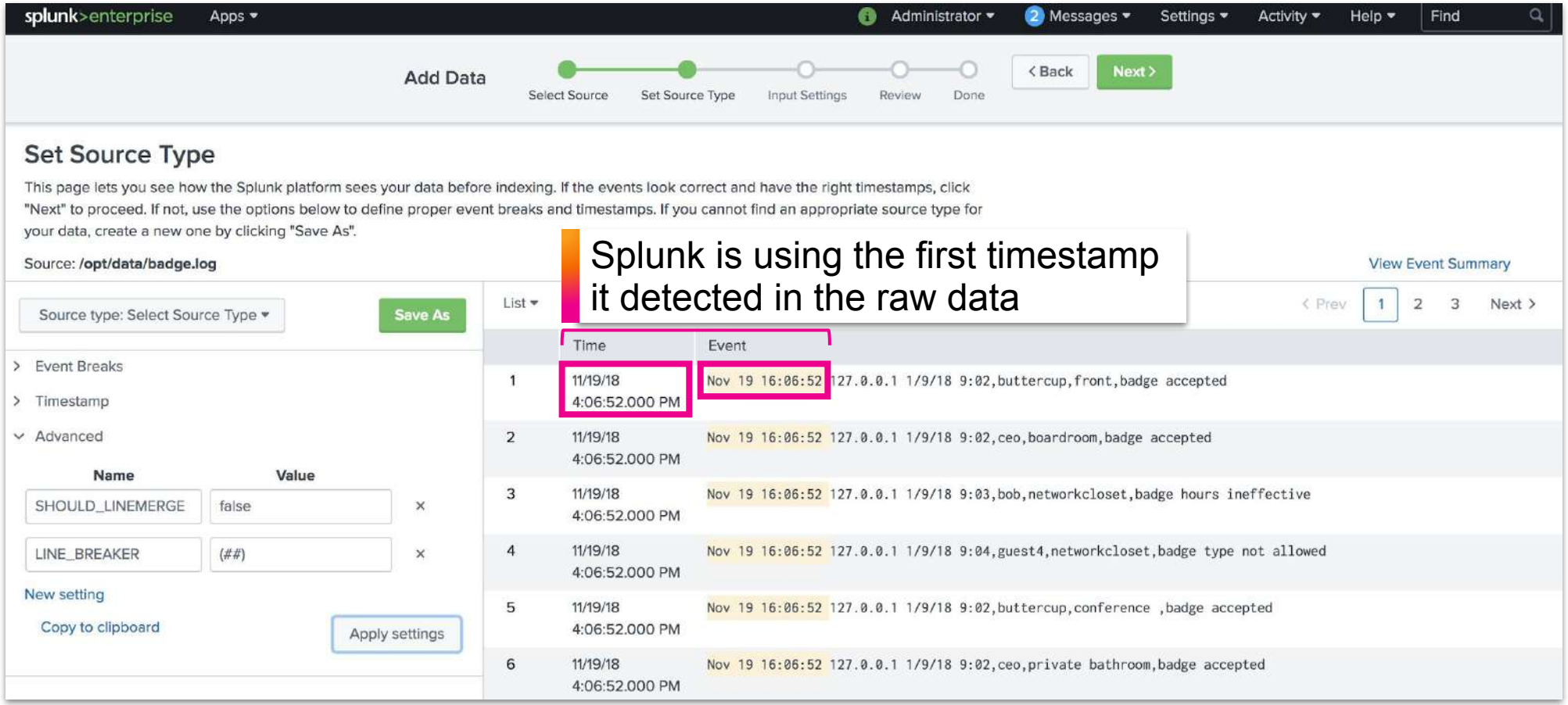
4. Set custom attributes: Edit **SHOULD_LINEMERGE** and **LINE_BREAKER**



LAB 2

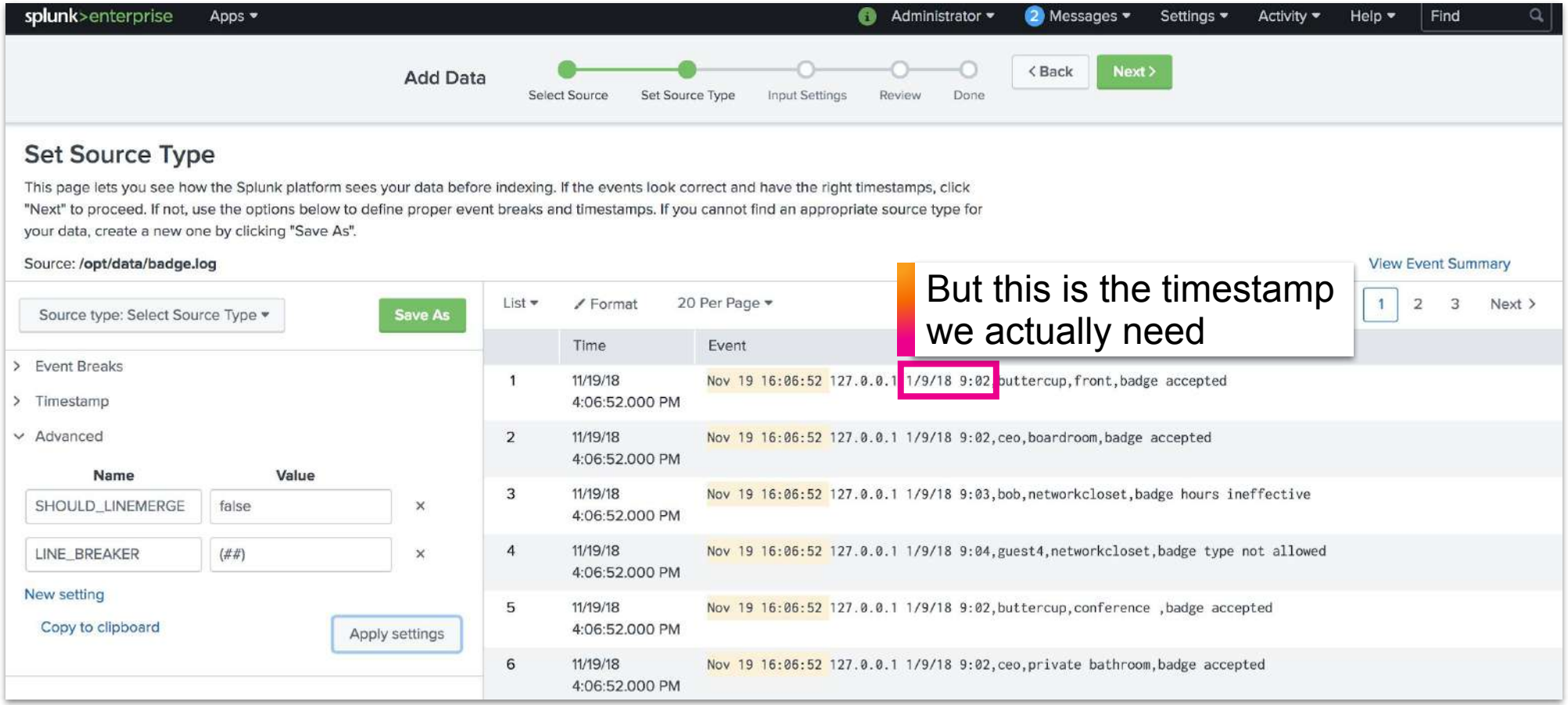
Task 9: Add Badge Data

Line breaking is ok…do you notice something weird ?



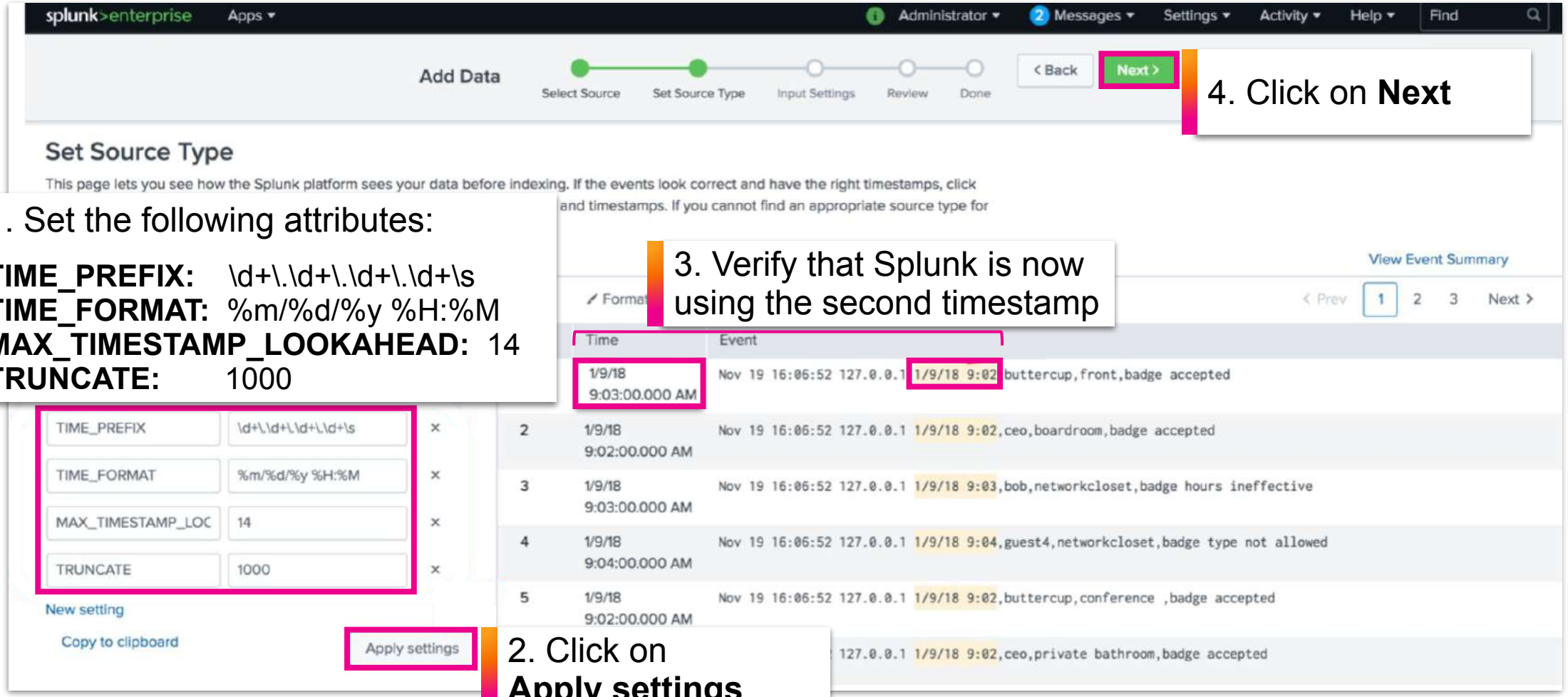
Task 9: Add Badge Data

The right timestamp is the second one



Task 9: Add Badge Data

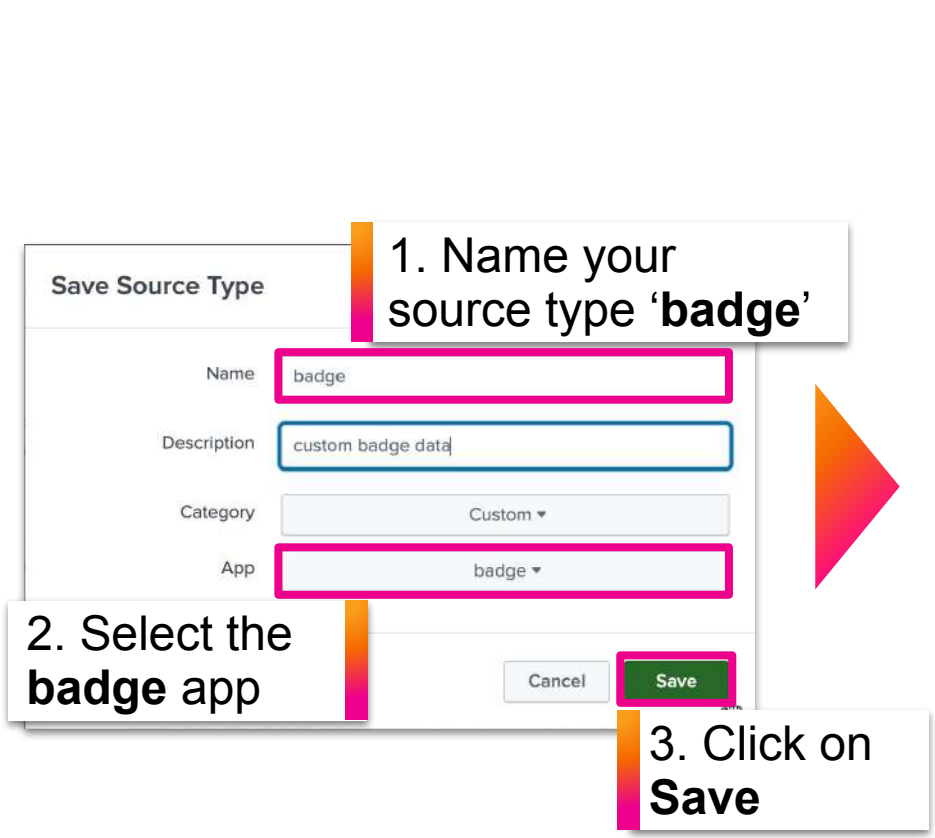
6. Set custom source type: Add best practice time stamp settings

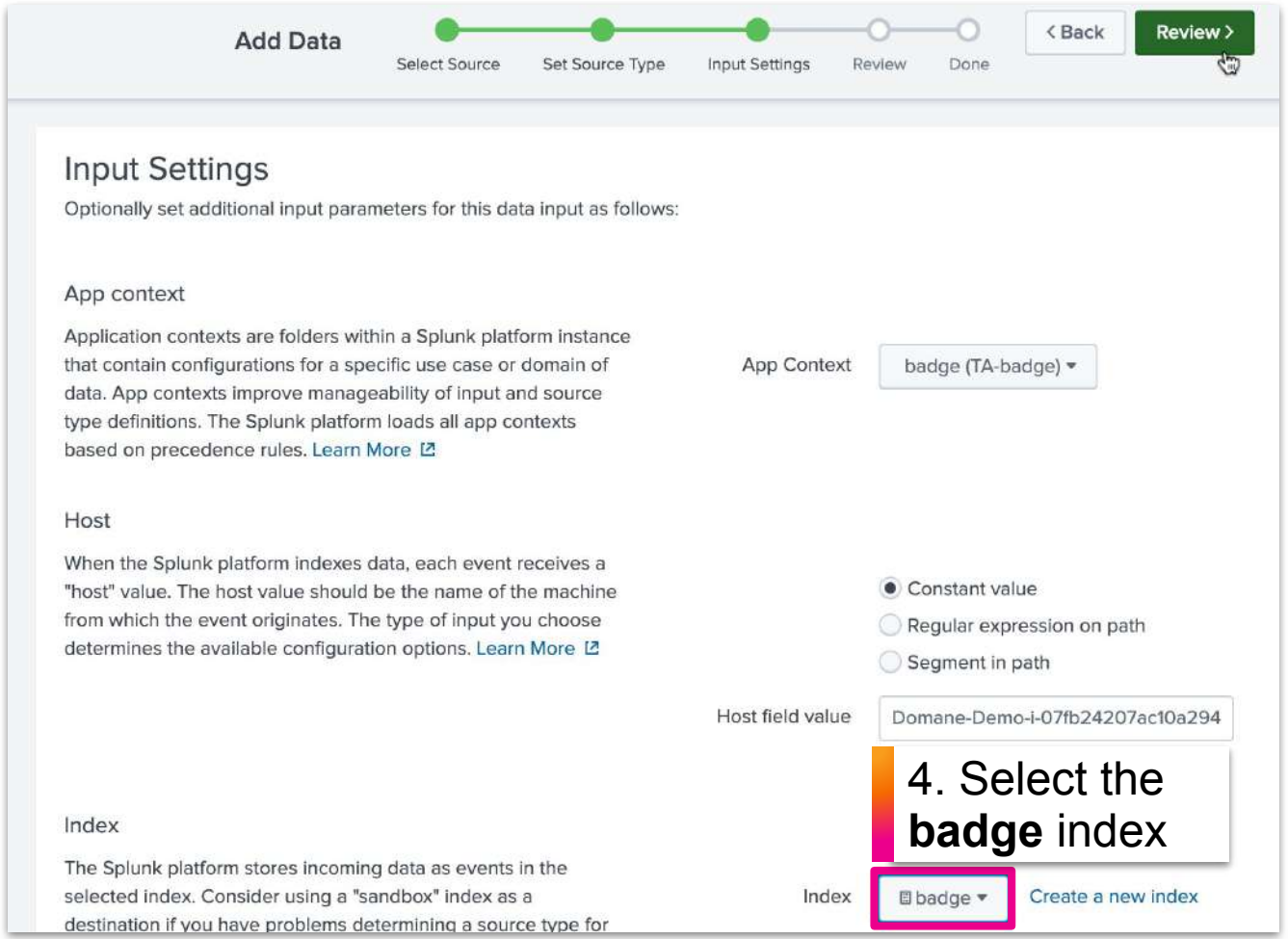


splunk>

Task 9: Add Badge Data

7. Set custom source type: Save sourcetype, pick app context & index, & review





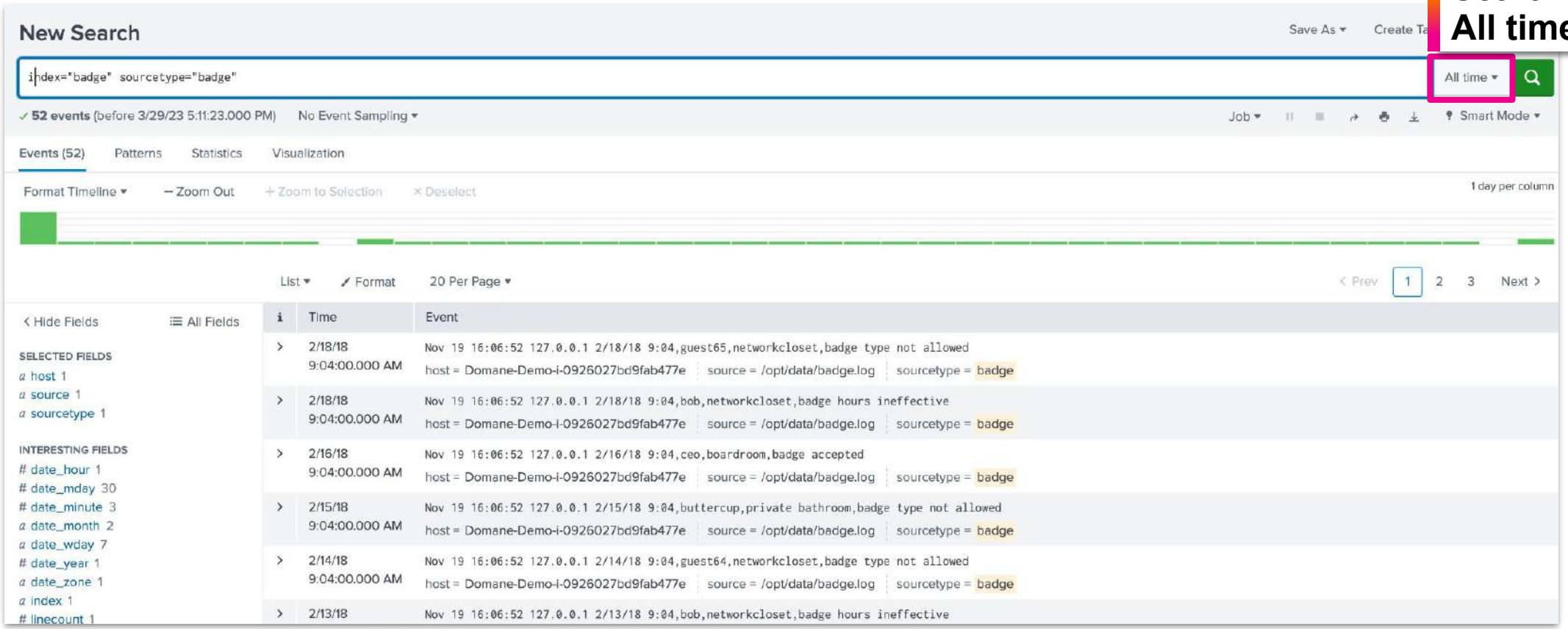
splunk>

Task 9: Add Badge Data

8. Check your data with a search:

```
index="badge" sourcetype="badge"
```

Search over **All time**



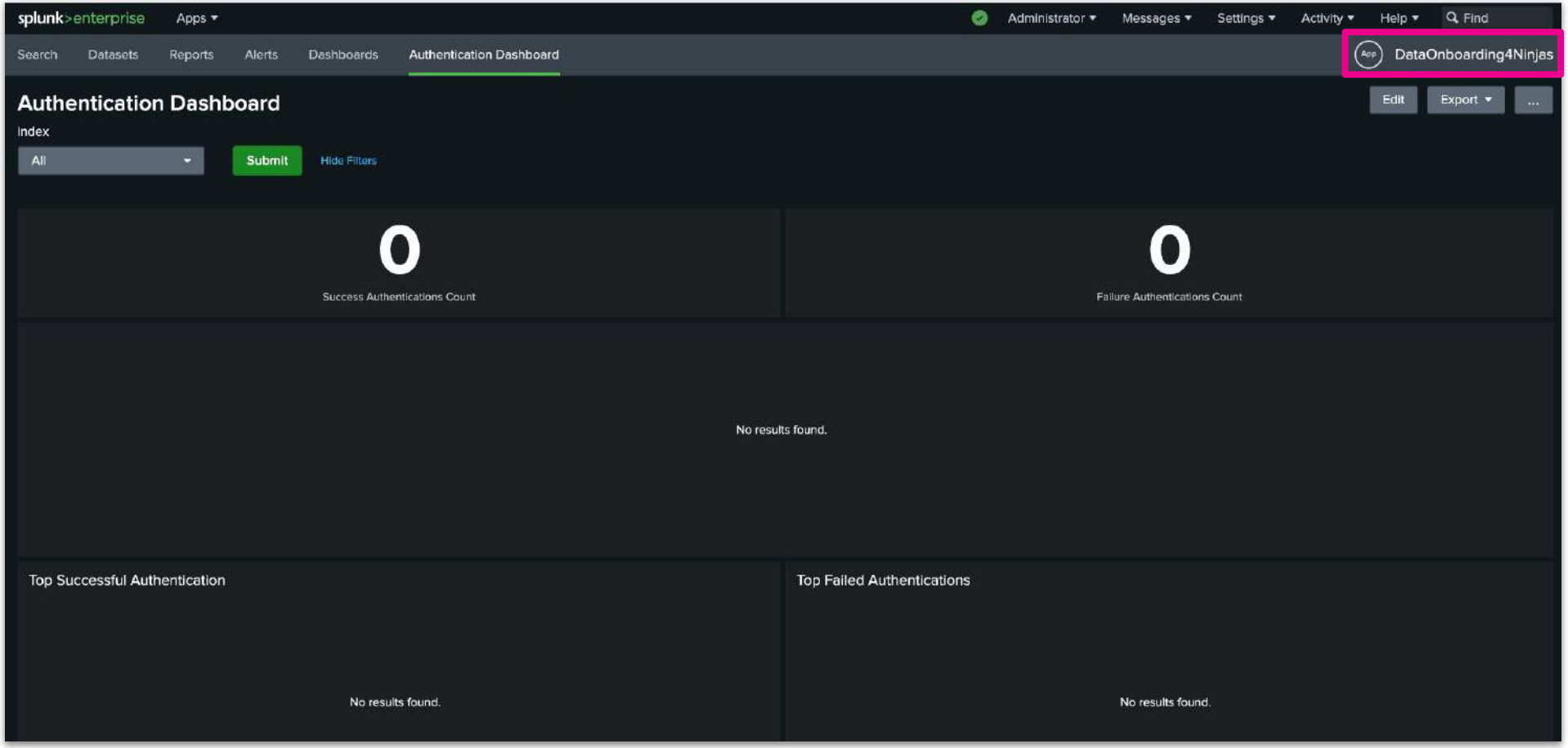
Dashboards & Reports Best Practices

Use SPL queries that search against the data models

- The goal is to understand the data (i.e. extract fields), apply some best practices and make it CIM compliant
- You have already indexed badge system data, can you do reports on it?
- Why?

Final Objective of the Day:
Display Badge Data on the Authentication Dashboard

1. Locate the Authentication Dashboard in the DataOnboarding4Ninjas app



splunk>

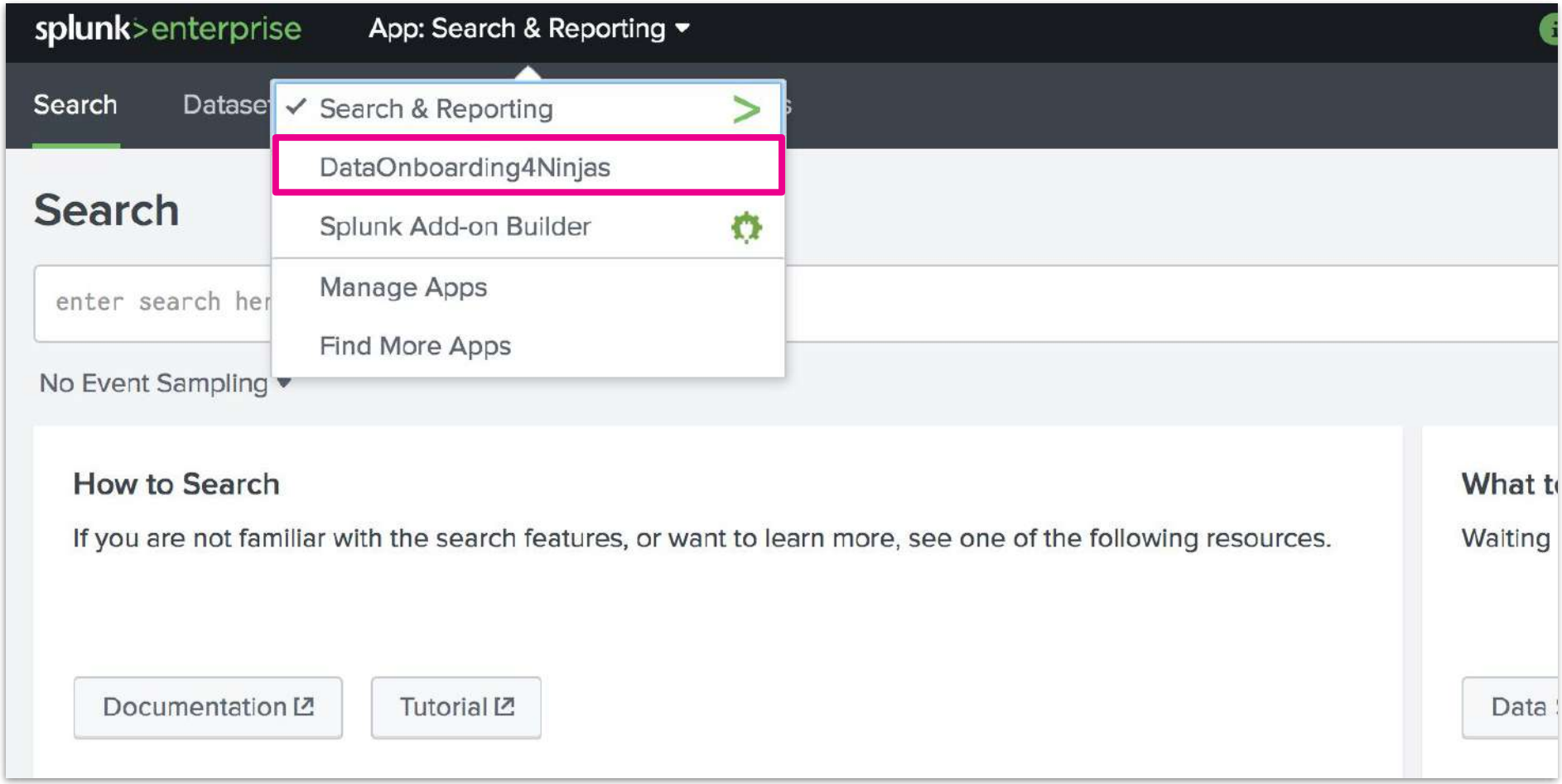
Practice Lab

Lab 3: Create Reports on Badge Data

splunk>

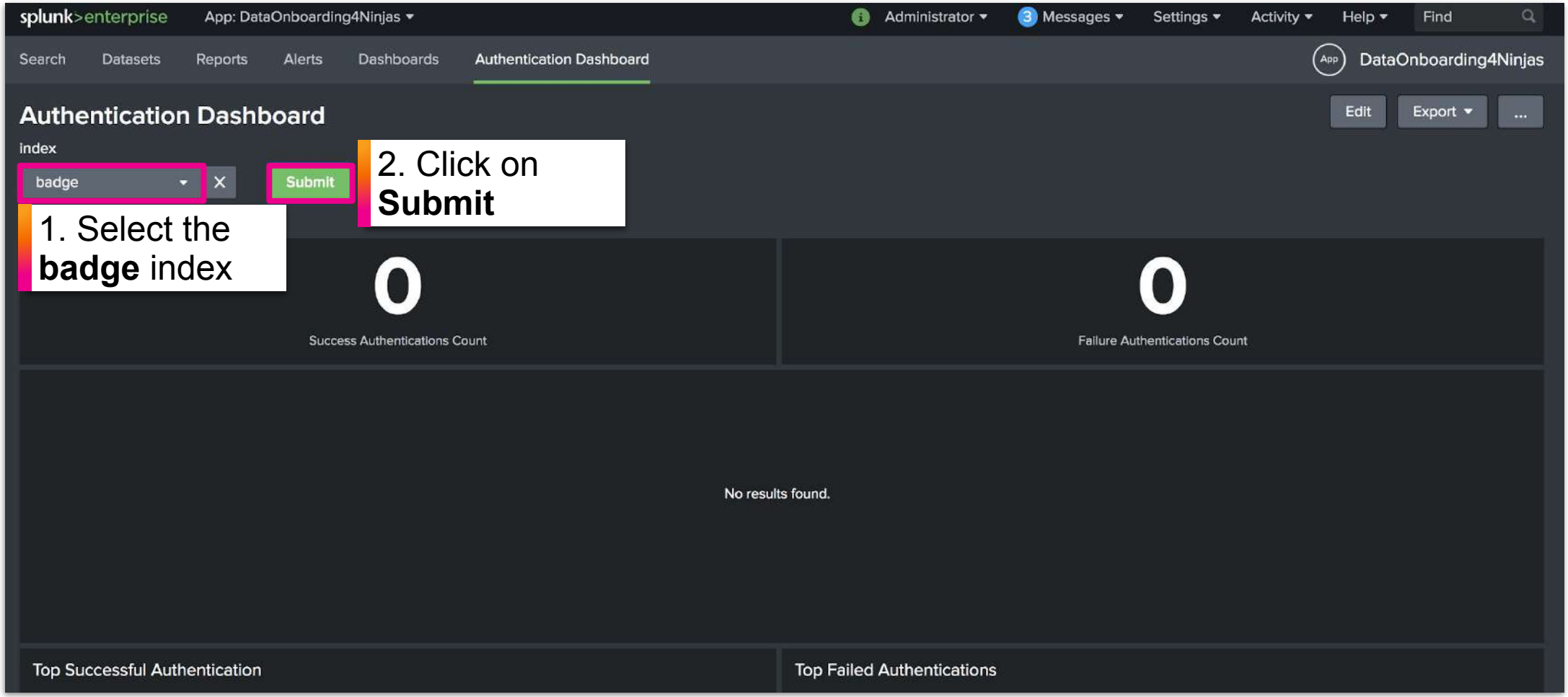
Task 10: Create Reports on Badge Data

1. Go to the **DataOnboarding4Ninjas** App



Task 10: Create Reports on Badge Data

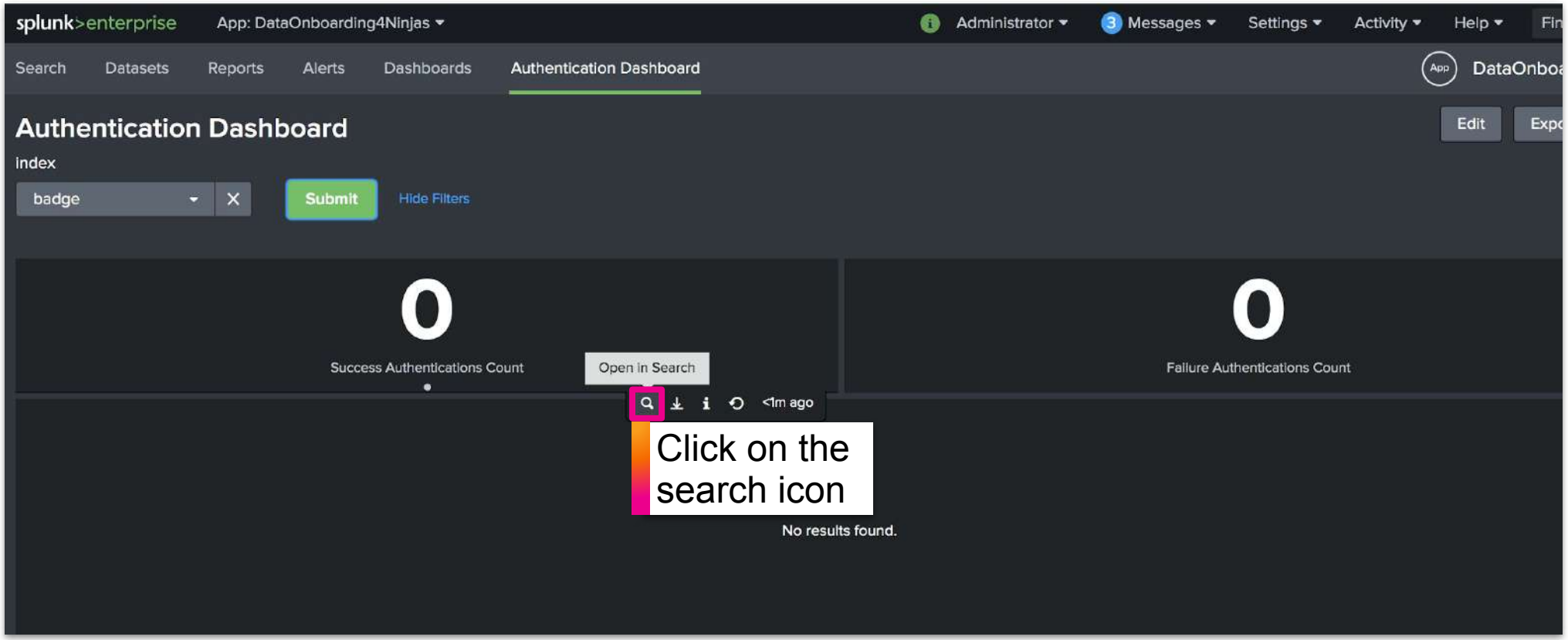
2. Check the Authentication Dashboard



Task 10: Create Reports on Badge Data

3. Choose one panel and click on the **Open in Search** icon

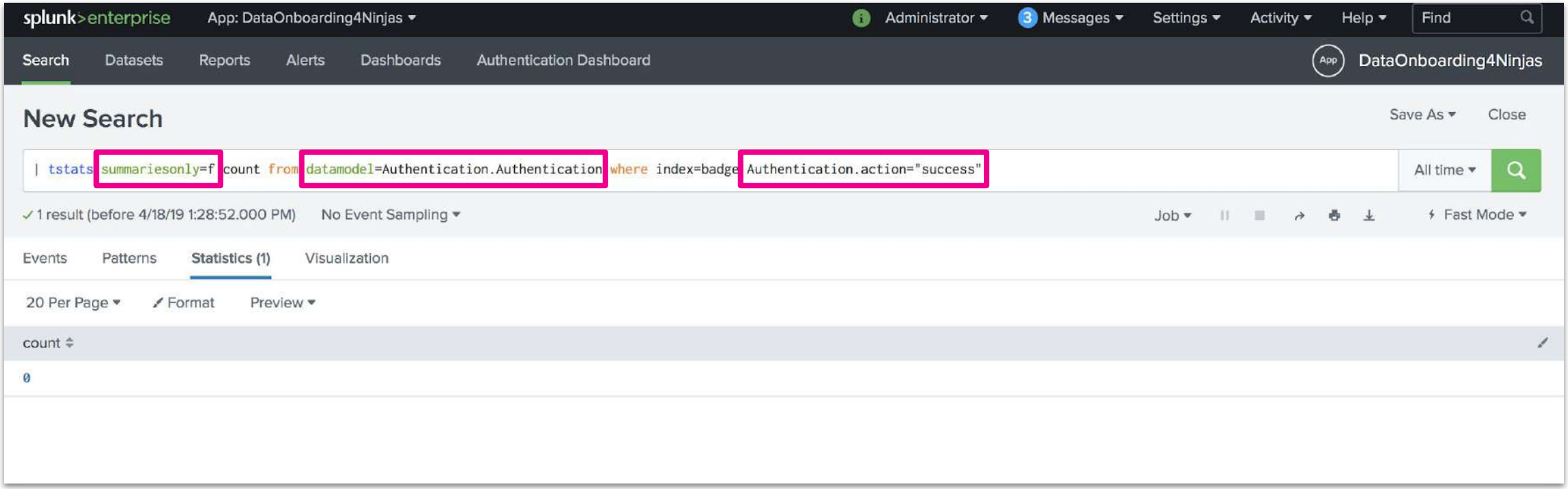
Nothing is displayed … why?



LAB 3

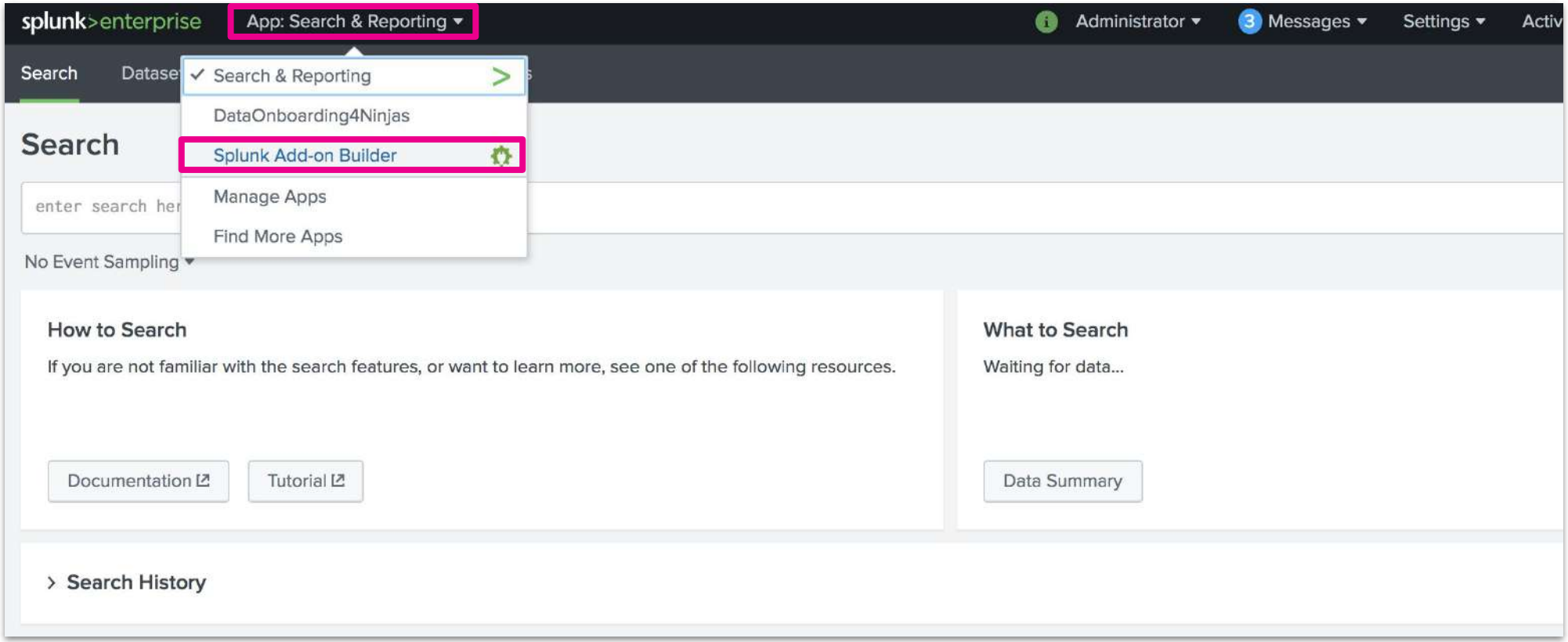
Task 10: Create Reports on Badge Data

Notice the search. We are doing a query against the Authentication data model



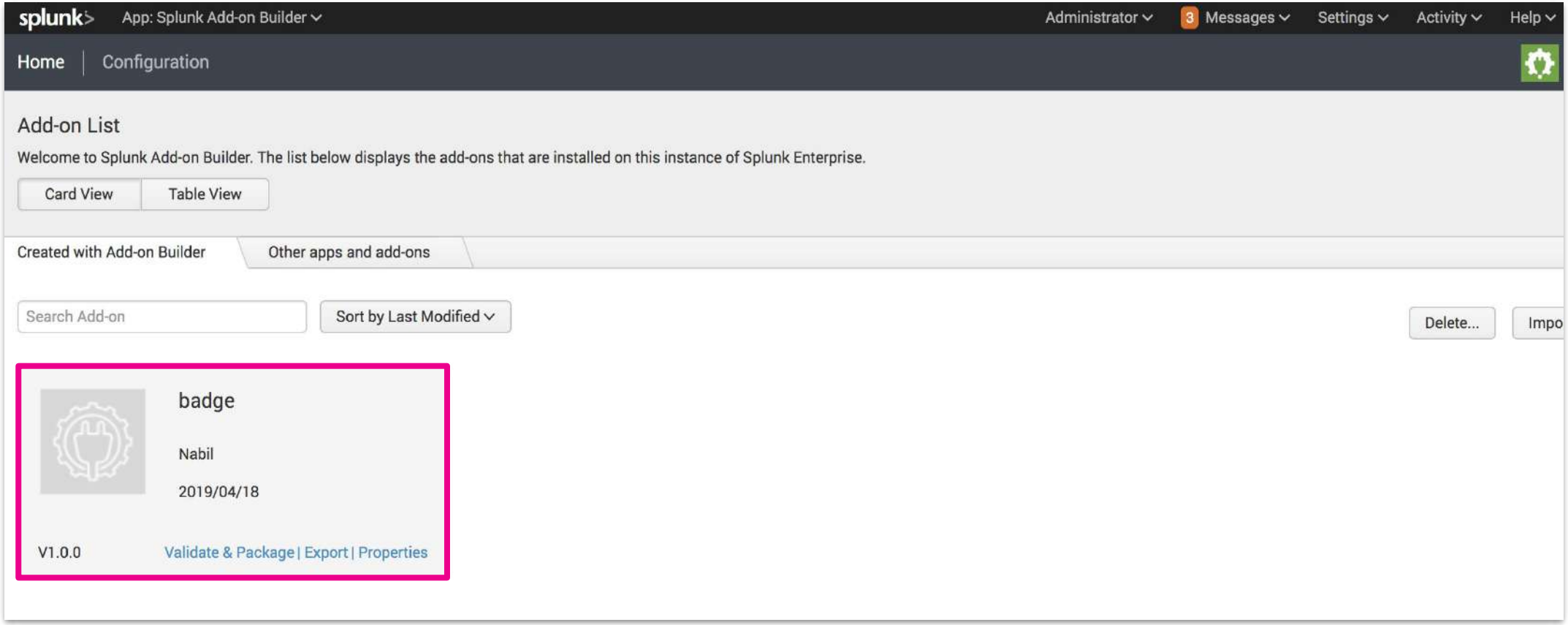
Task 10: Create Reports on Badge Data

4. Go to the **Splunk Add-on Builder** app



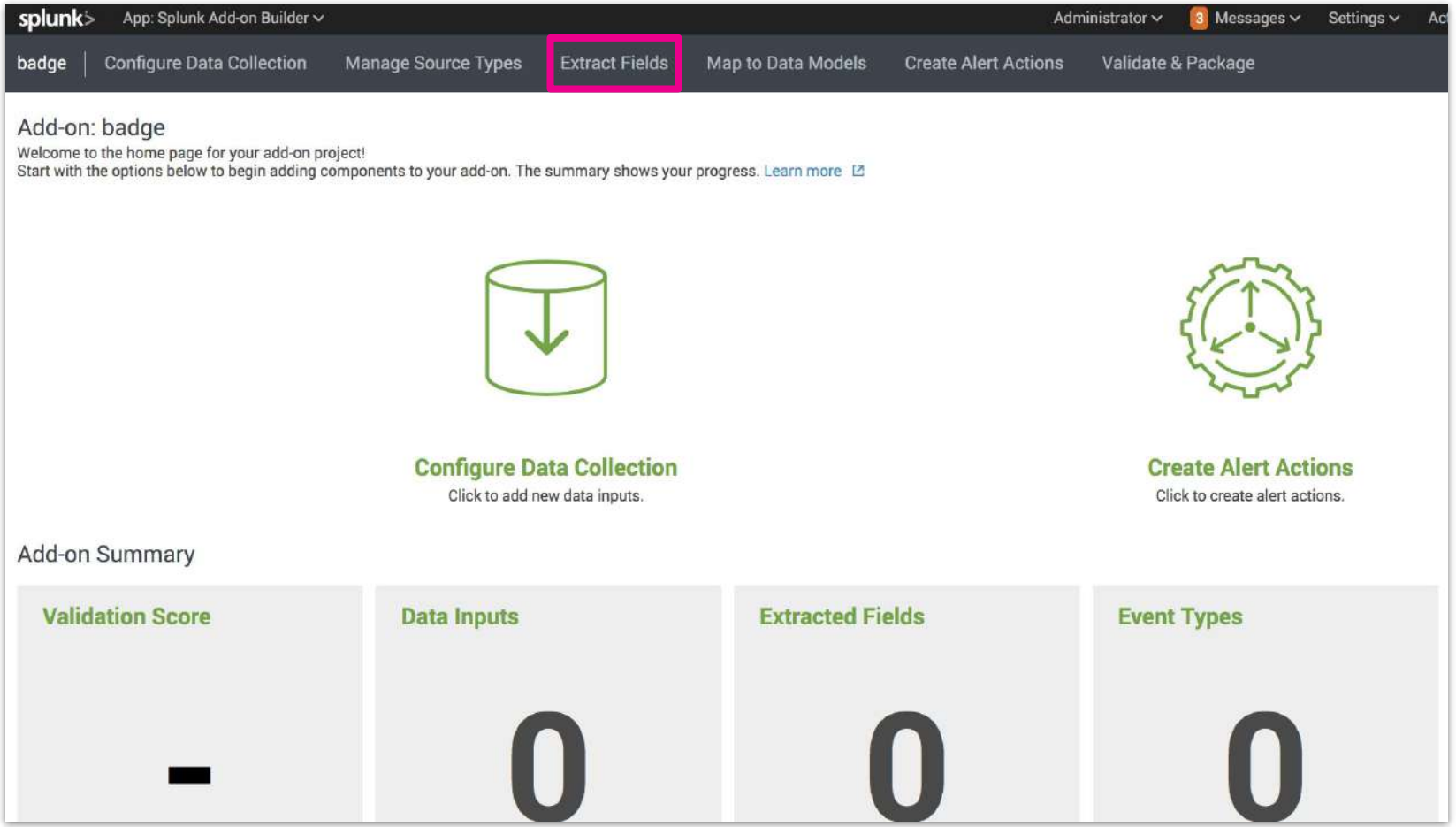
Task 10: Create Reports on Badge Data

5. Click on the add-on you created before



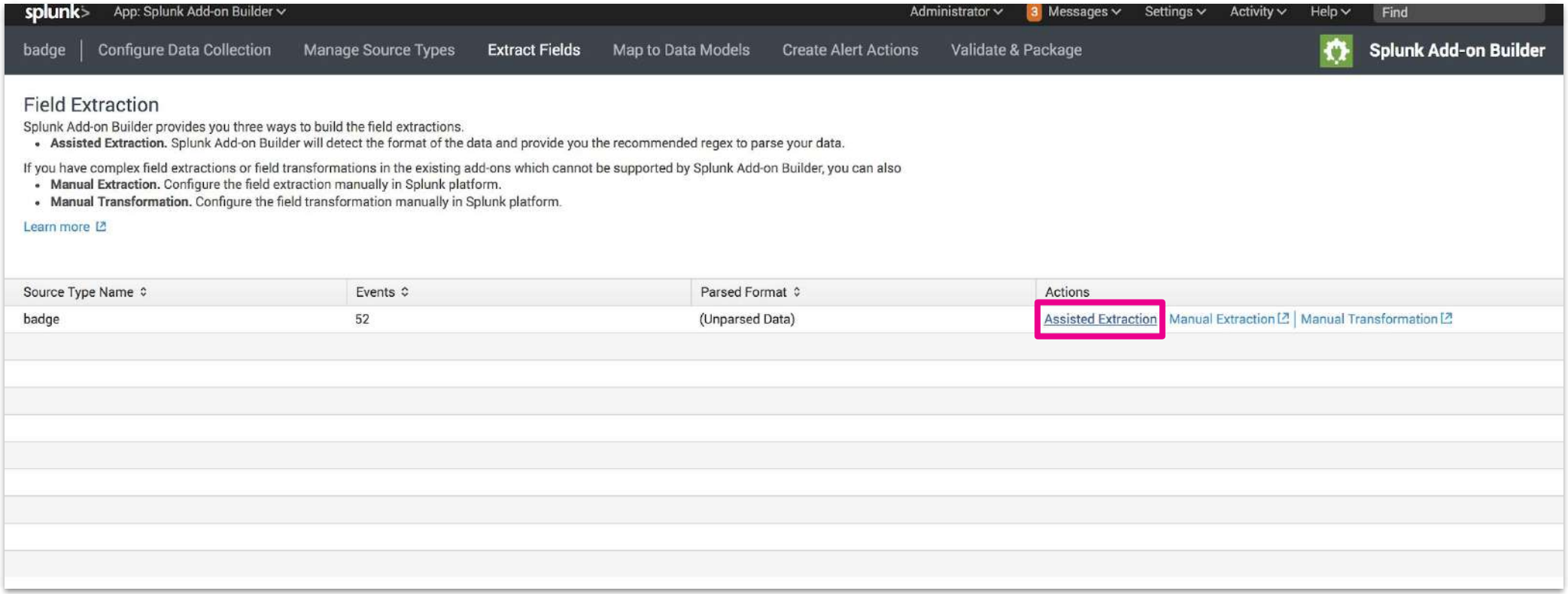
Task 10: Create Reports on Badge Data

6. Click on **Extract fields**



Task 10: Create Reports on Badge Data

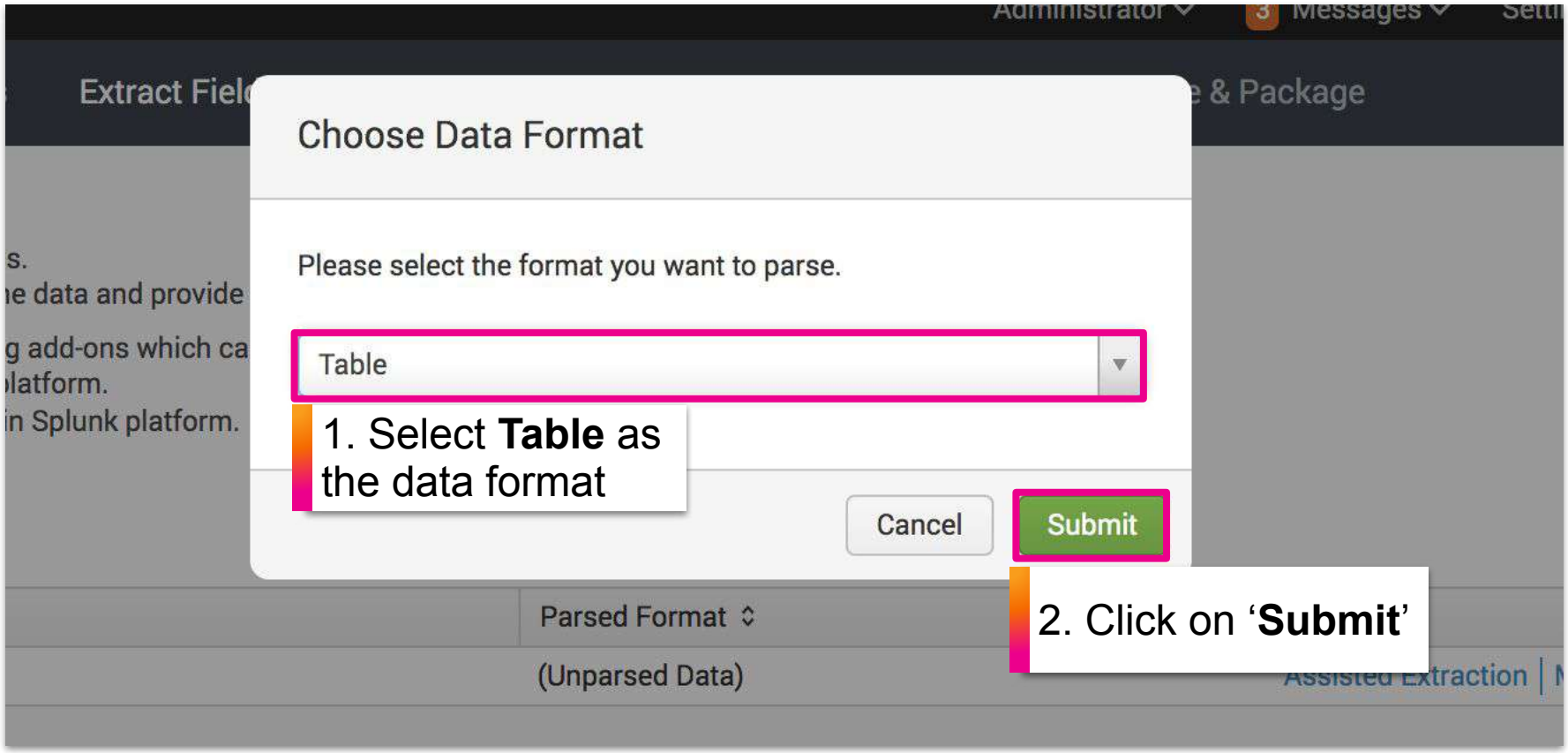
7. Click on **Assisted Extraction**



splunk>

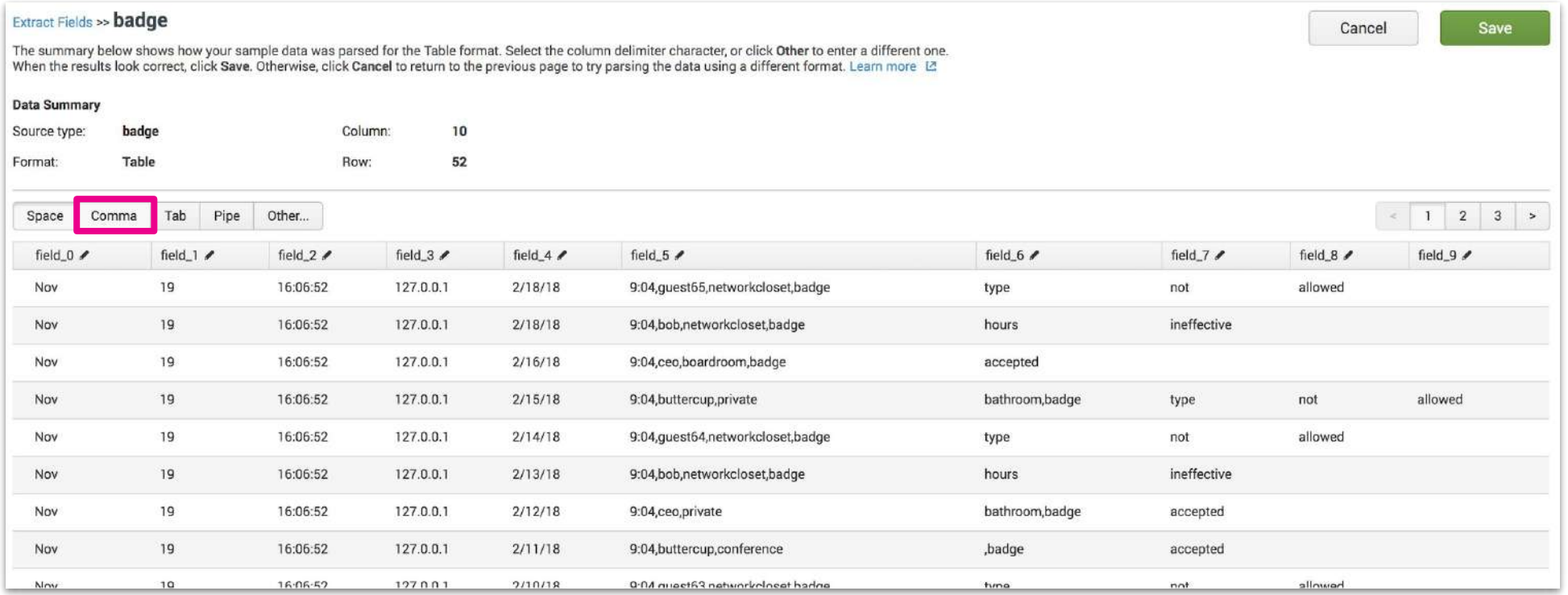
Task 10: Create Reports on Badge Data

8. Choose Table and click on **Submit**



Task 10: Create Reports on Badge Data

9. Choose **Comma** as the delimiter character



Extract Fields >> **badge**

Cancel | Save

The summary below shows how your sample data was parsed for the Table format. Select the column delimiter character, or click **Other** to enter a different one.
When the results look correct, click **Save**. Otherwise, click **Cancel** to return to the previous page to try parsing the data using a different format. Learn more

Data Summary

Source type: **badge** Column: **10**

Format: **Table** Row: **52**

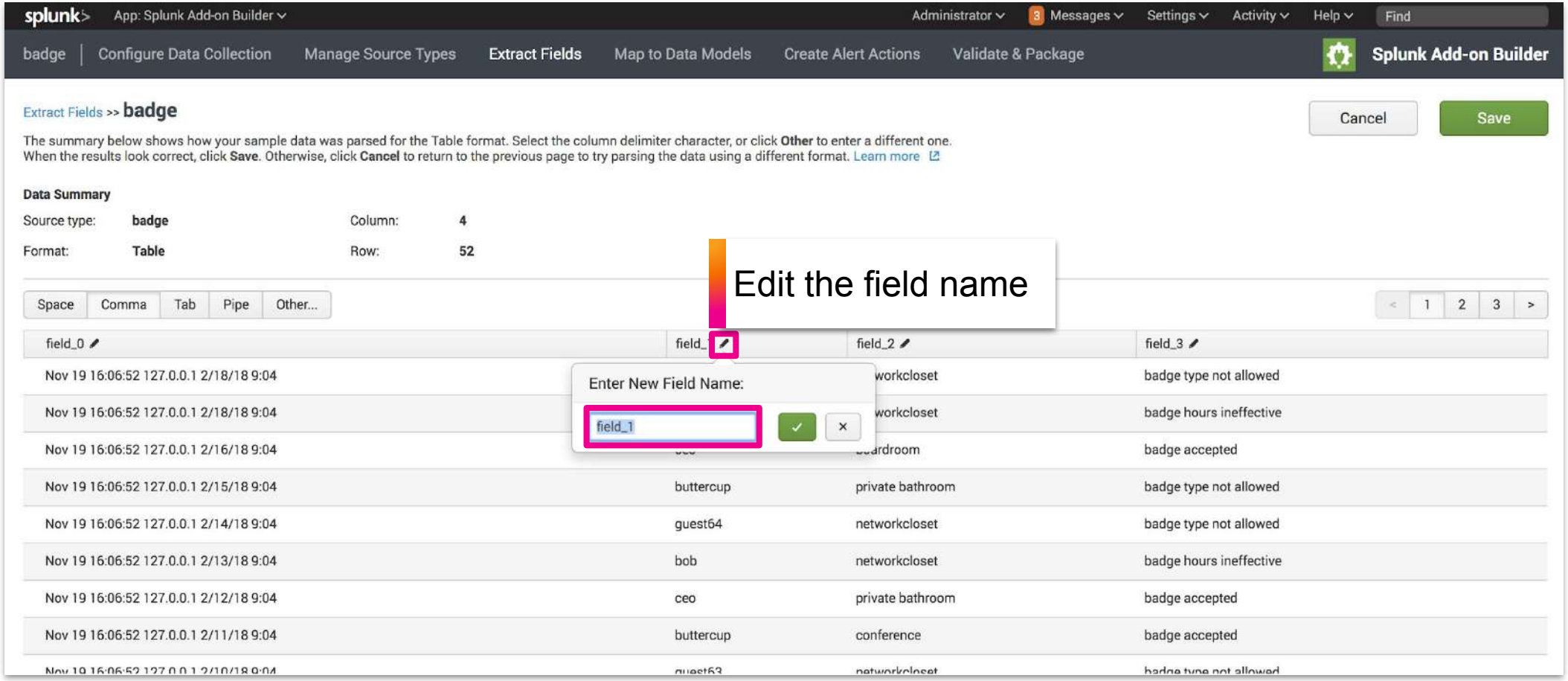
Space | Comma | Tab | Pipe | Other...

< 1 2 3 >

field_0	field_1	field_2	field_3	field_4	field_5	field_6	field_7	field_8	field_9
Nov	19	16:06:52	127.0.0.1	2/18/18	9:04,guest65,networkcloset,badge	type	not	allowed	
Nov	19	16:06:52	127.0.0.1	2/18/18	9:04,bob,networkcloset,badge	hours	ineffective		
Nov	19	16:06:52	127.0.0.1	2/16/18	9:04,ceo,boardroom,badge	accepted			
Nov	19	16:06:52	127.0.0.1	2/15/18	9:04,buttercup,private	bathroom,badge	type	not	allowed
Nov	19	16:06:52	127.0.0.1	2/14/18	9:04,guest64,networkcloset,badge	type	not	allowed	
Nov	19	16:06:52	127.0.0.1	2/13/18	9:04,bob,networkcloset,badge	hours	ineffective		
Nov	19	16:06:52	127.0.0.1	2/12/18	9:04,ceo,private	bathroom,badge	accepted		
Nov	19	16:06:52	127.0.0.1	2/11/18	9:04,buttercup,conference	,badge	accepted		
Nov	19	16:06:52	127.0.0.1	2/10/18	9:04,guest63,networkcloset,badge	type	not	allowed	

Task 10: Create Reports on Badge Data

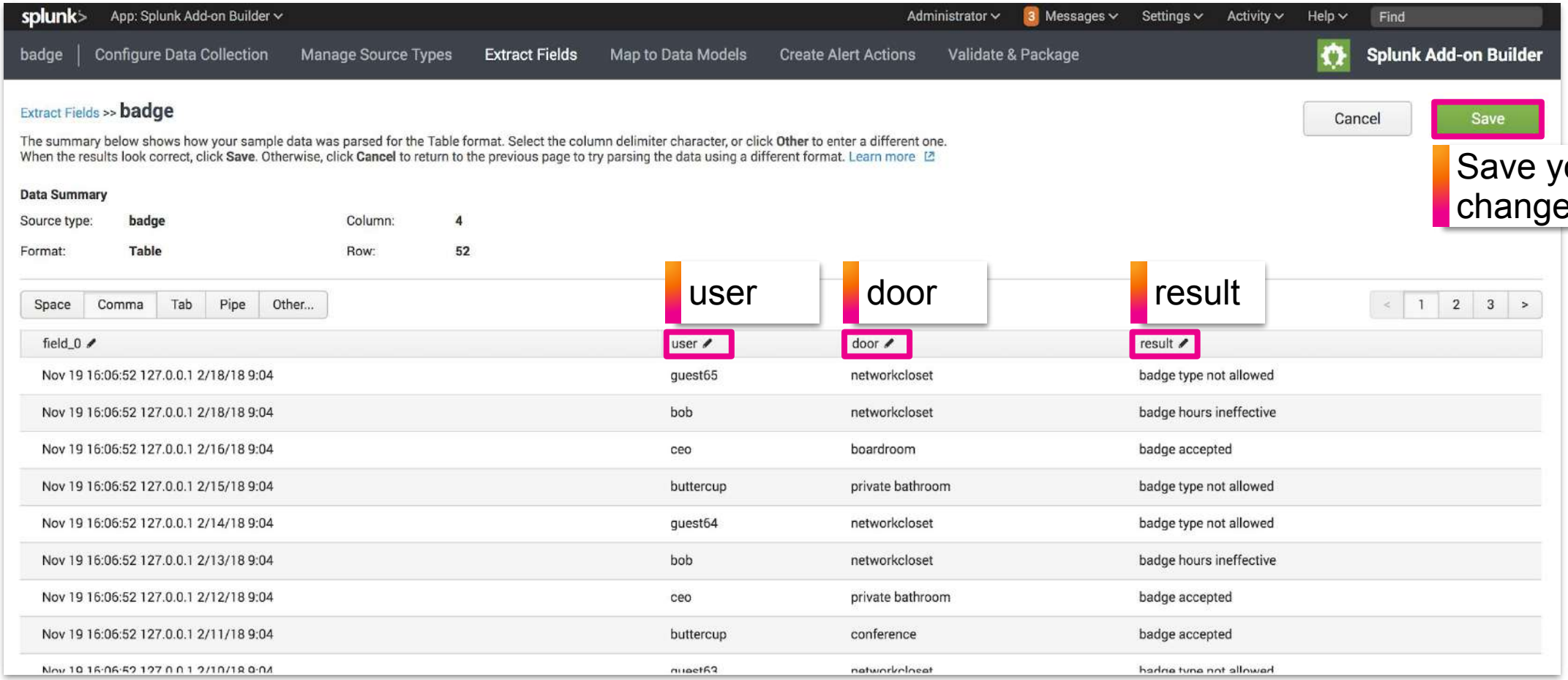
10. Rename the fields: **user**, **door**, **result**



splunk>

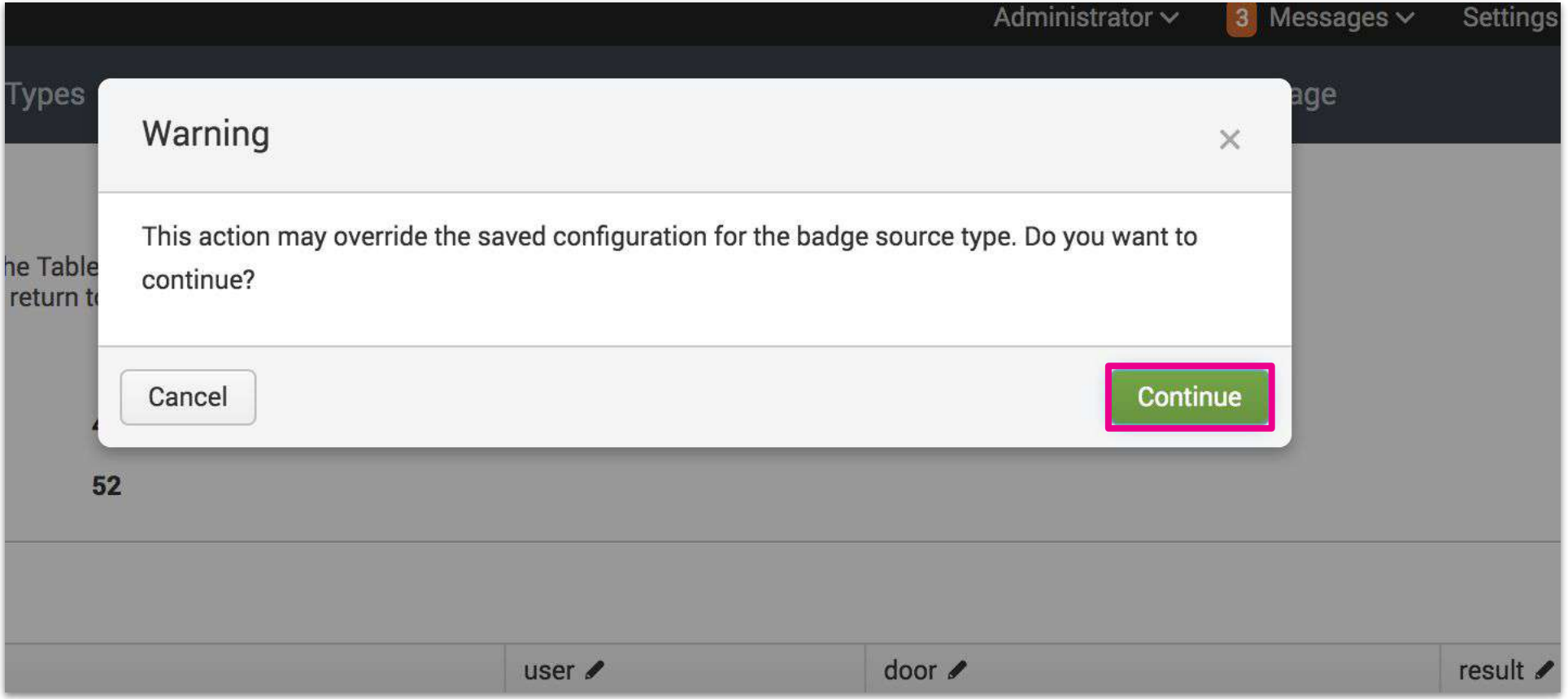
Task 10: Create Reports on Badge Data

11. Check the fields names and click **Save**



Task 10: Create Reports on Badge Data

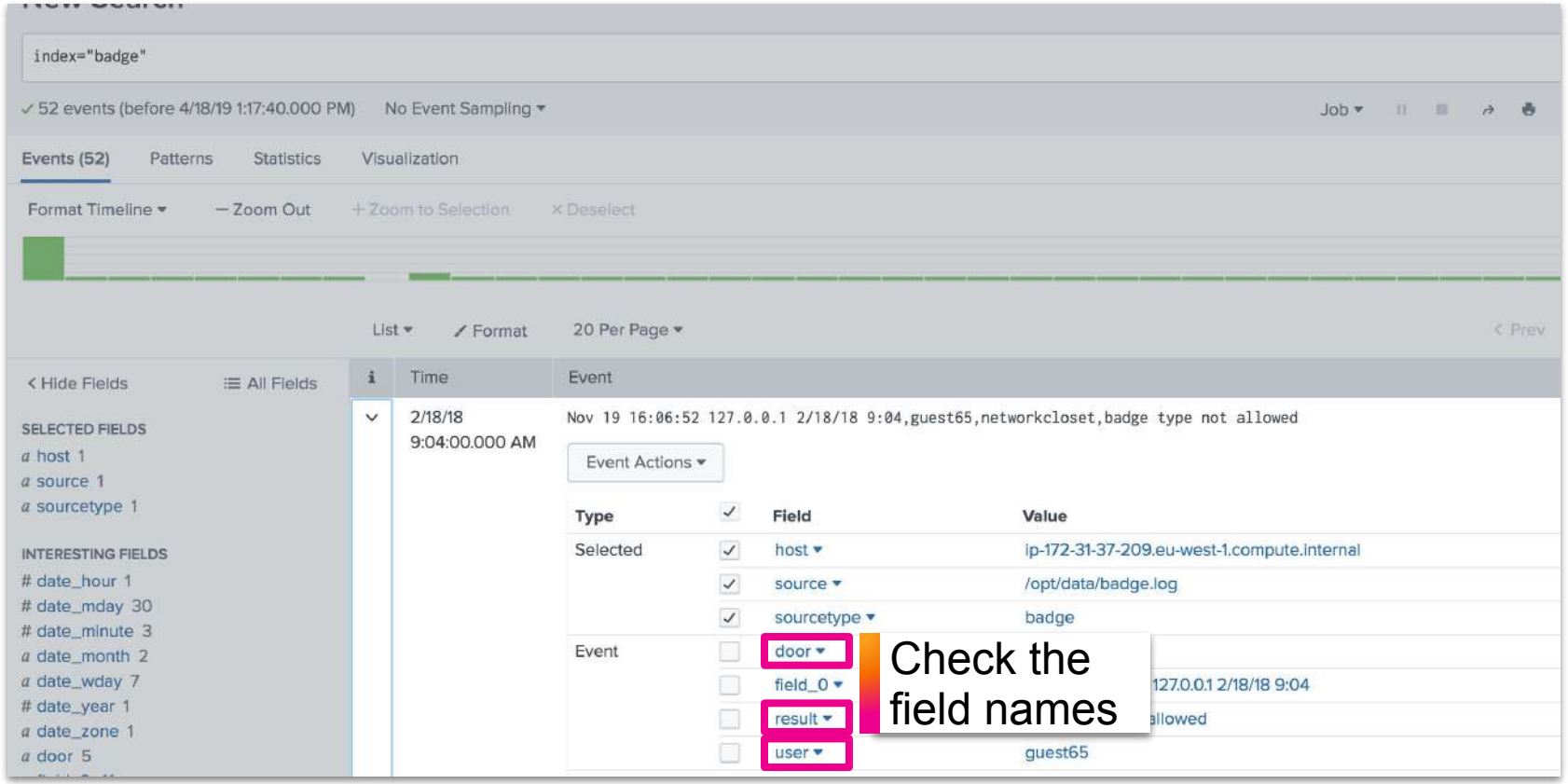
12. Click on **Continue**



Task 10: Create Reports on Badge Data

13. Check that the fields are extracted correctly

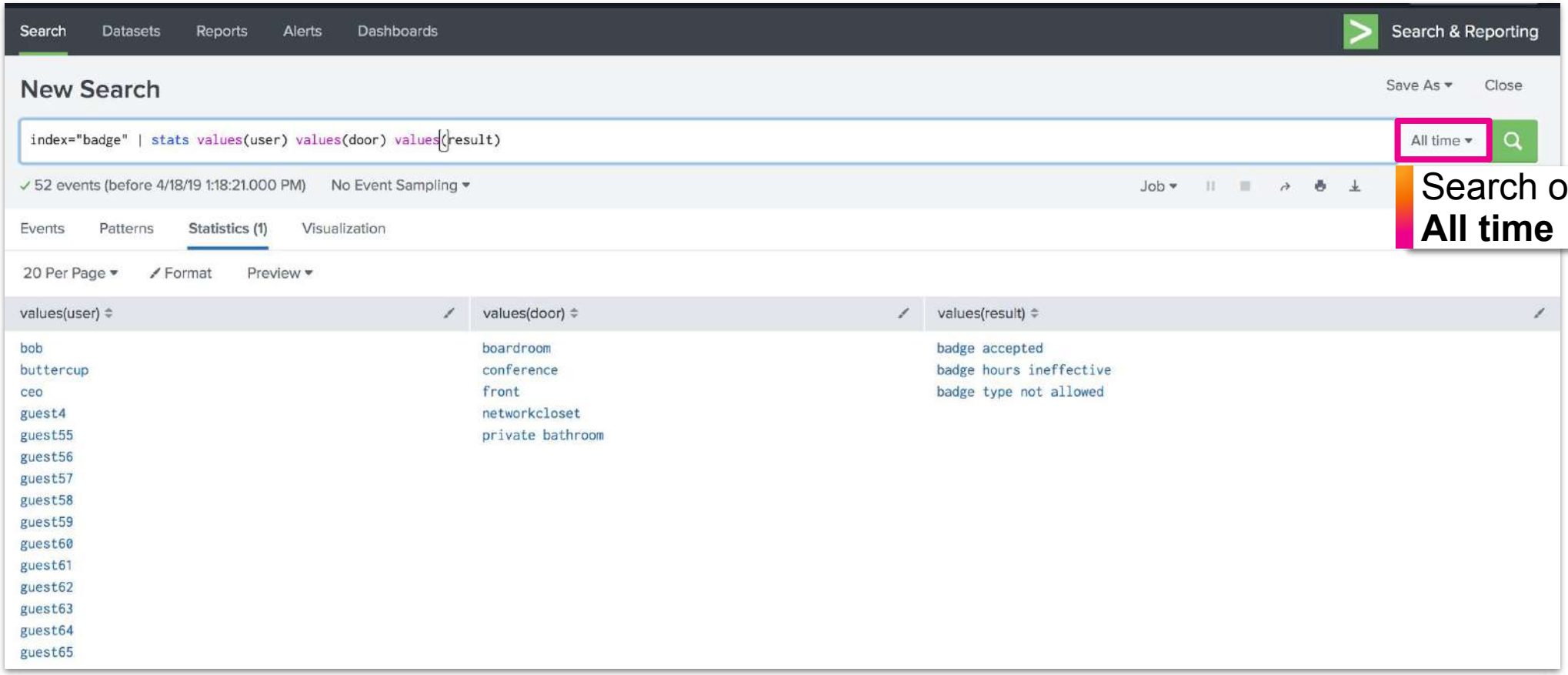
```
index="badge"
```



Task 10: Create Reports on Badge Data

14. Another way to check the fields

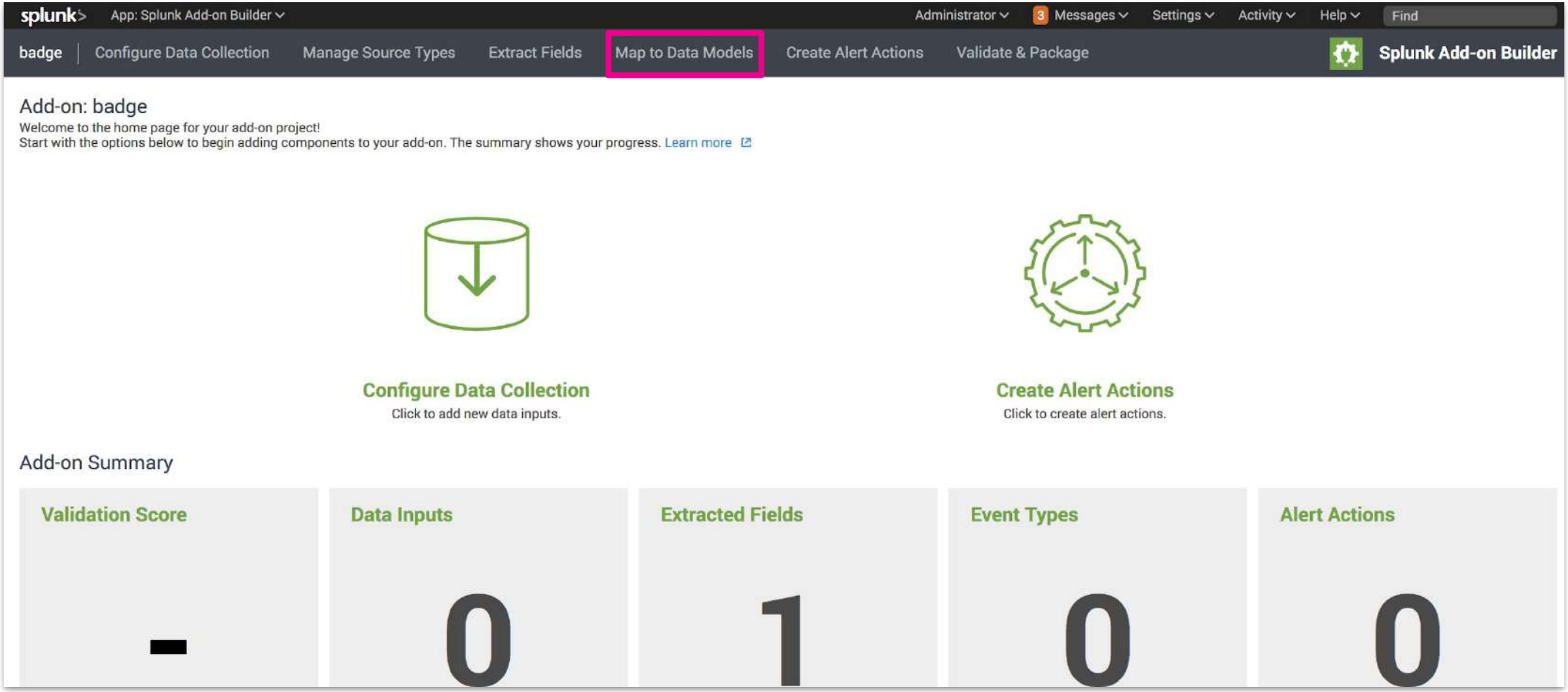
```
index=badge | stats values(user) values(door) values(result)
```



splunk>

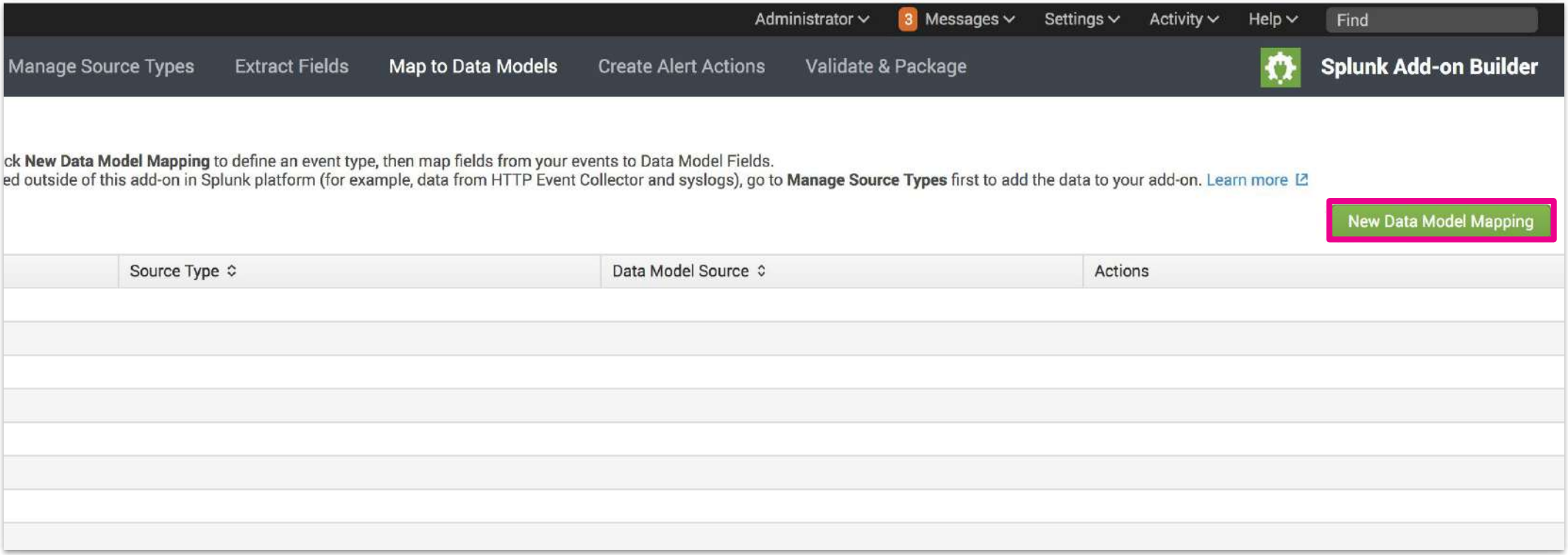
Task 10: Create Reports on Badge Data

15. Go to add on builder app and click on Map to Data Models



Task 10: Create Reports on Badge Data

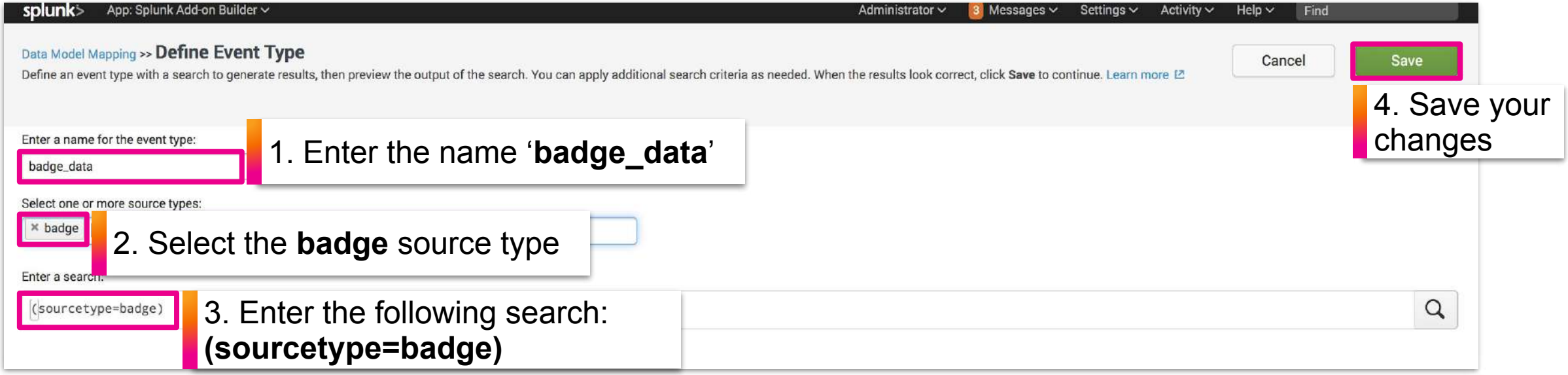
16. Click on **New Data Model Mapping**



splunk>

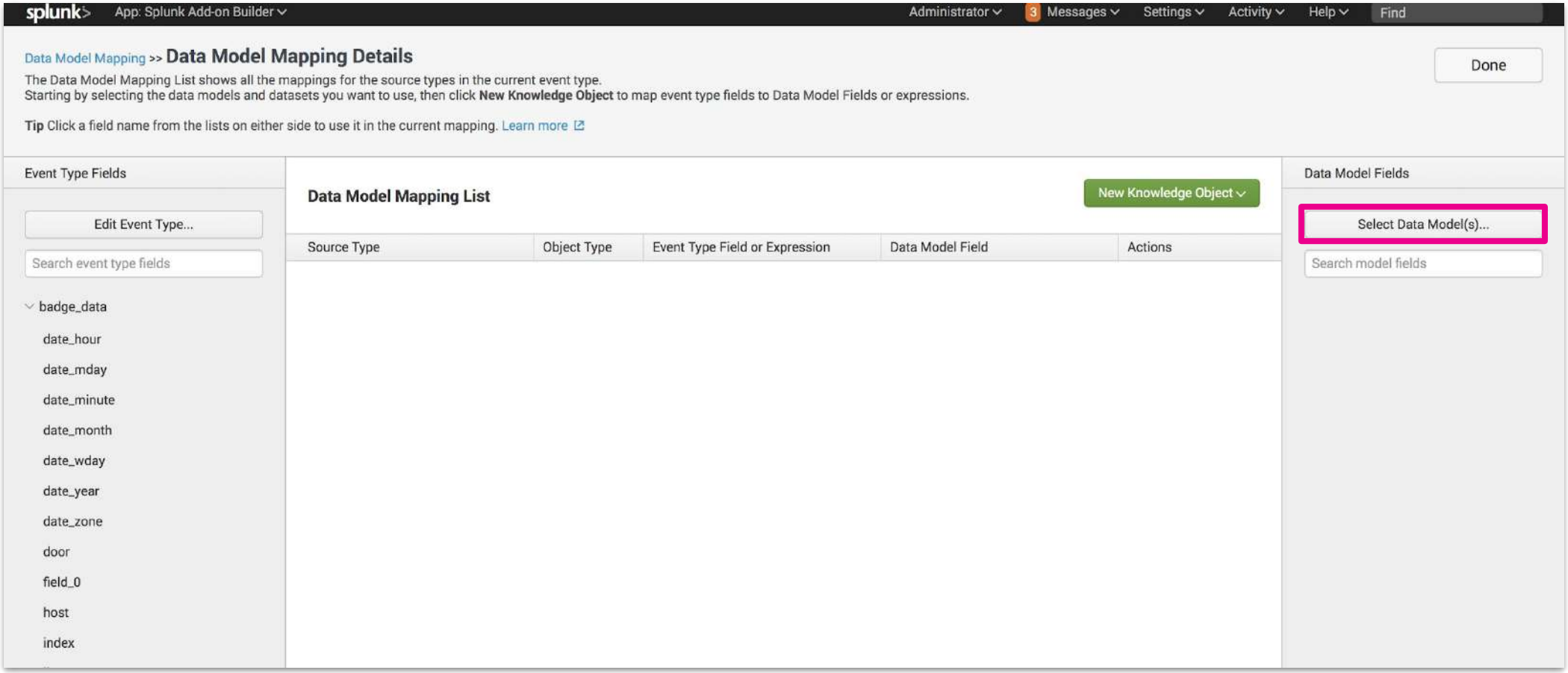
Task 10: Create Reports on Badge Data

17. Enter an event type name, select the source type and click on **Save**



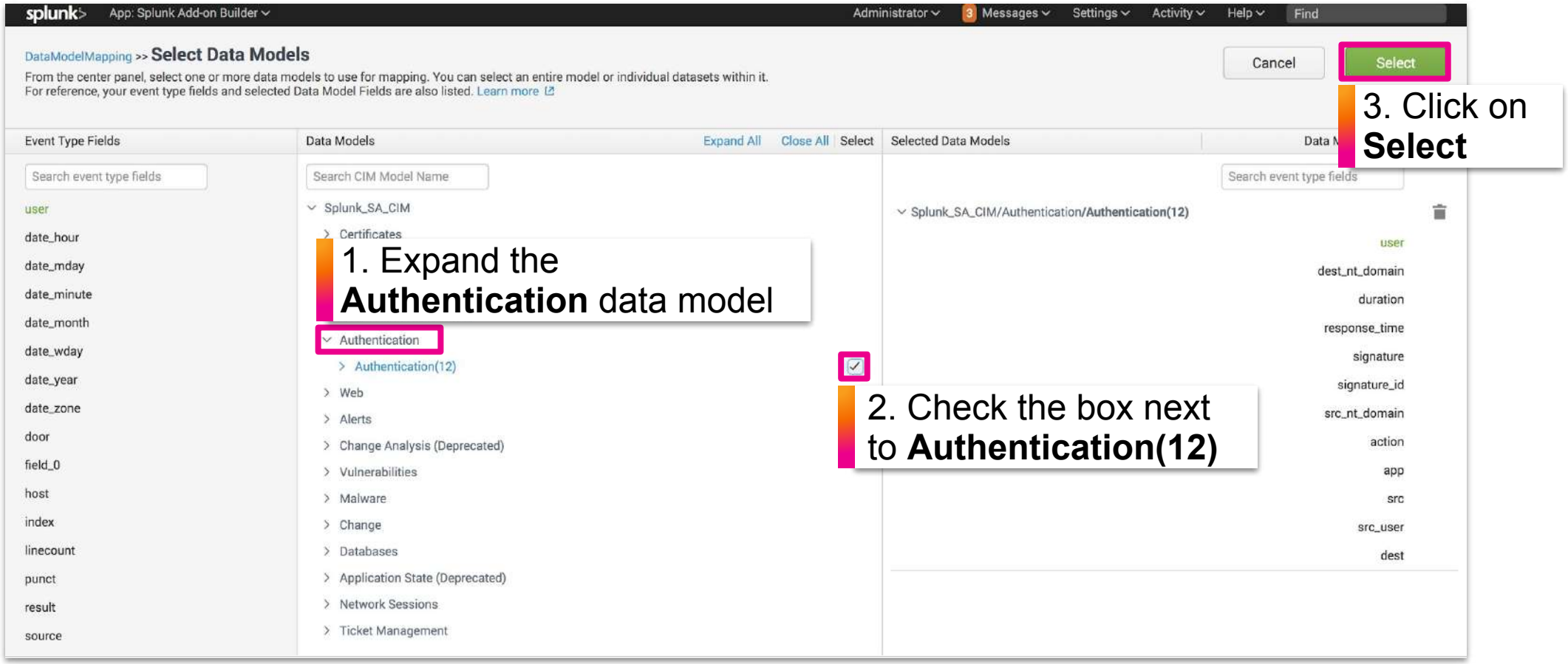
Task 10: Create Reports on Badge Data

18. Click on **Select Data Model(s)**



Task 10: Create Reports on Badge Data

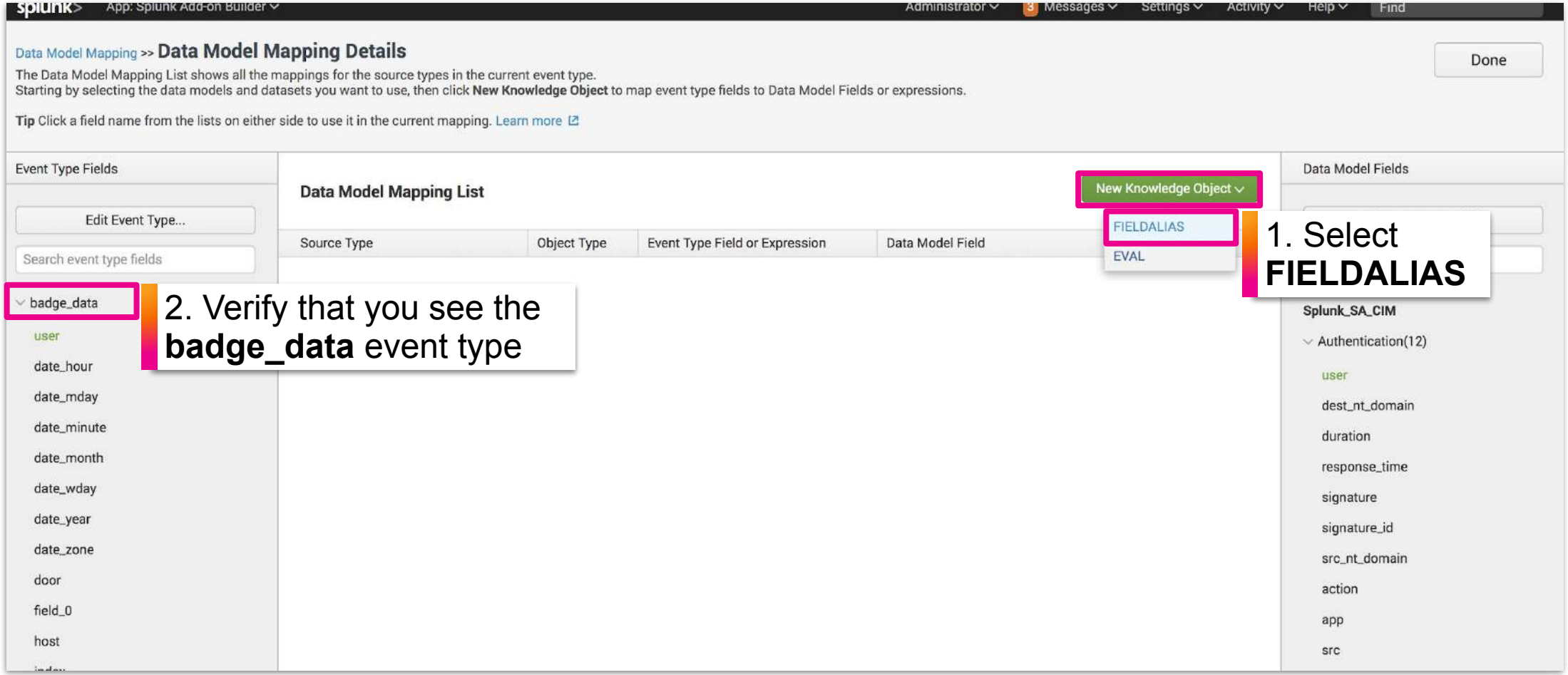
19. Select Authentication Data Model and click on **Save**



splunk>

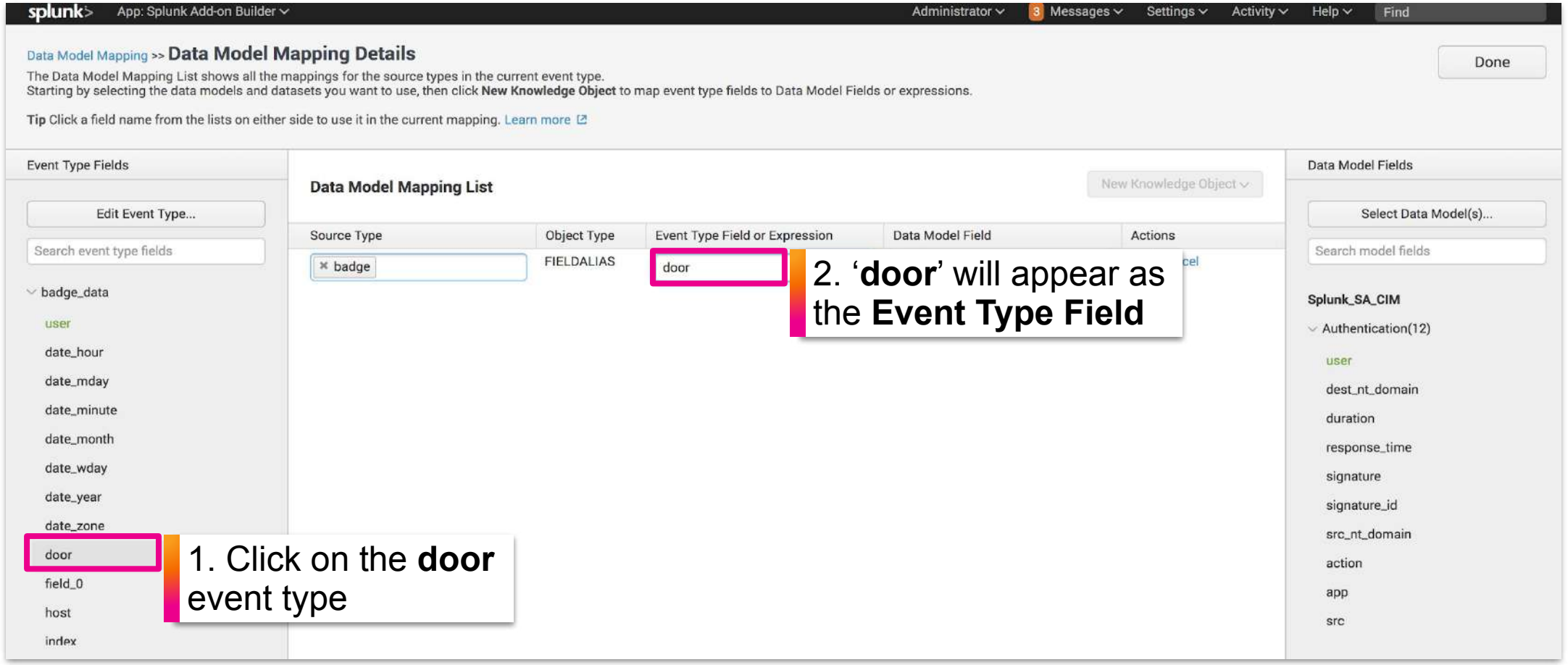
Task 10: Create Reports on Badge Data

20. Click on New Knowledge Object and select **FIELDALIAS**



Task 10: Create Reports on Badge Data

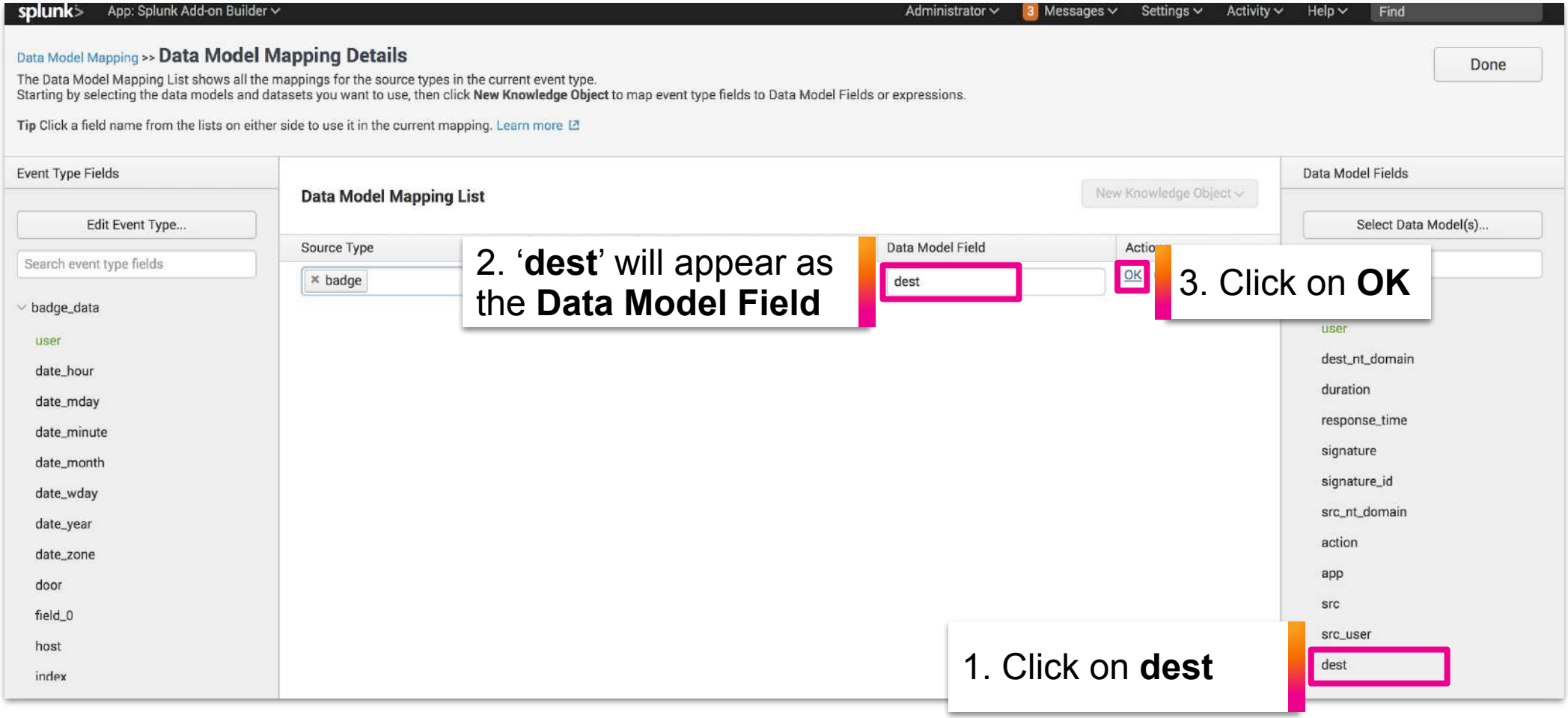
21. Click on the **door** field on the left panel



splunk>

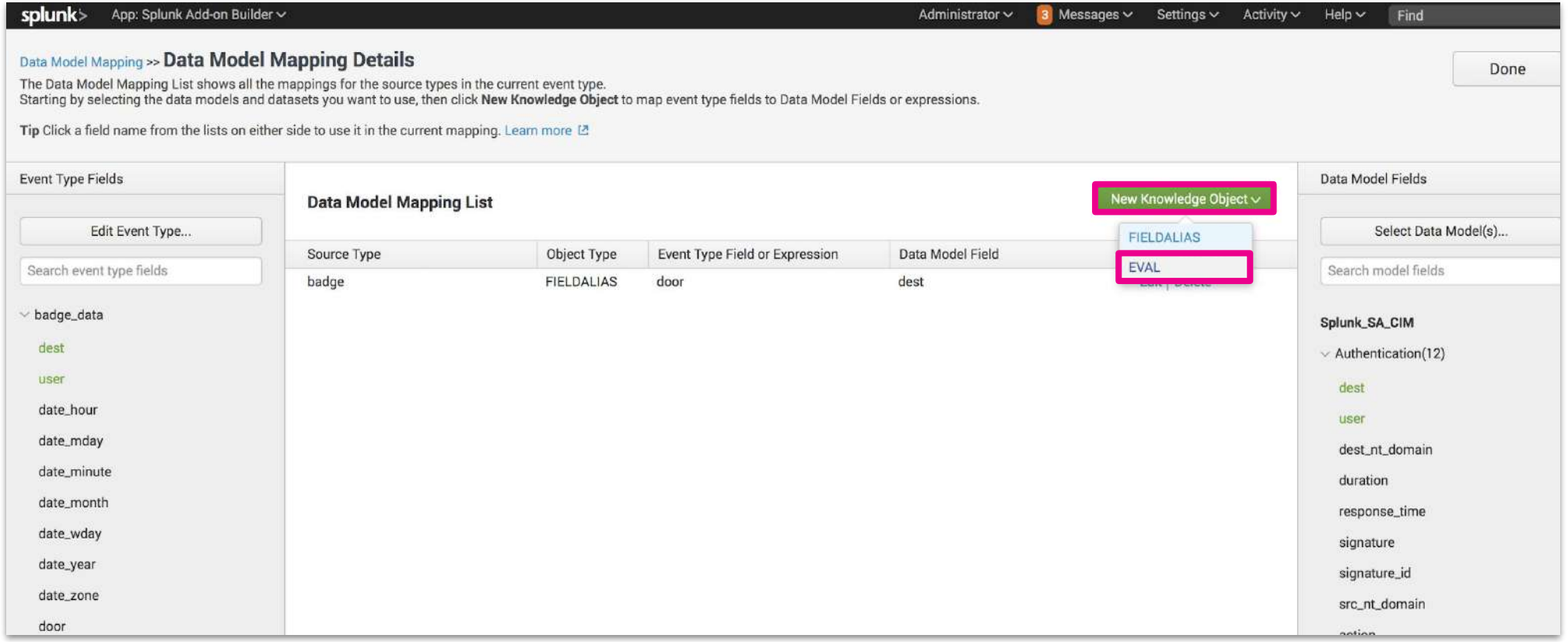
Task 10: Create Reports on Badge Data

22. Click on **dest** field on the right panel, then click on **OK**



Task 10: Create Reports on Badge Data

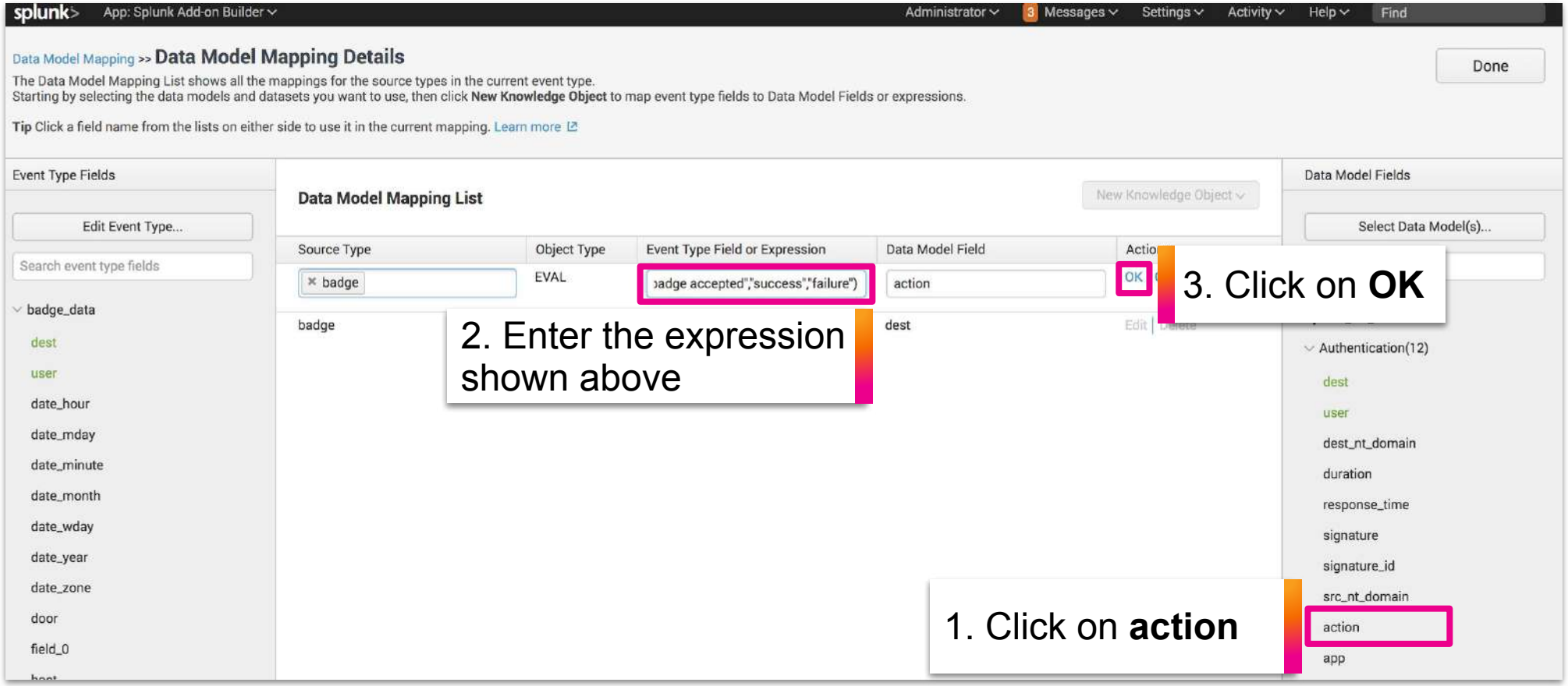
23. Click on **New Knowledge Object** and select **EVAL**



Task 10: Create Reports on Badge Data

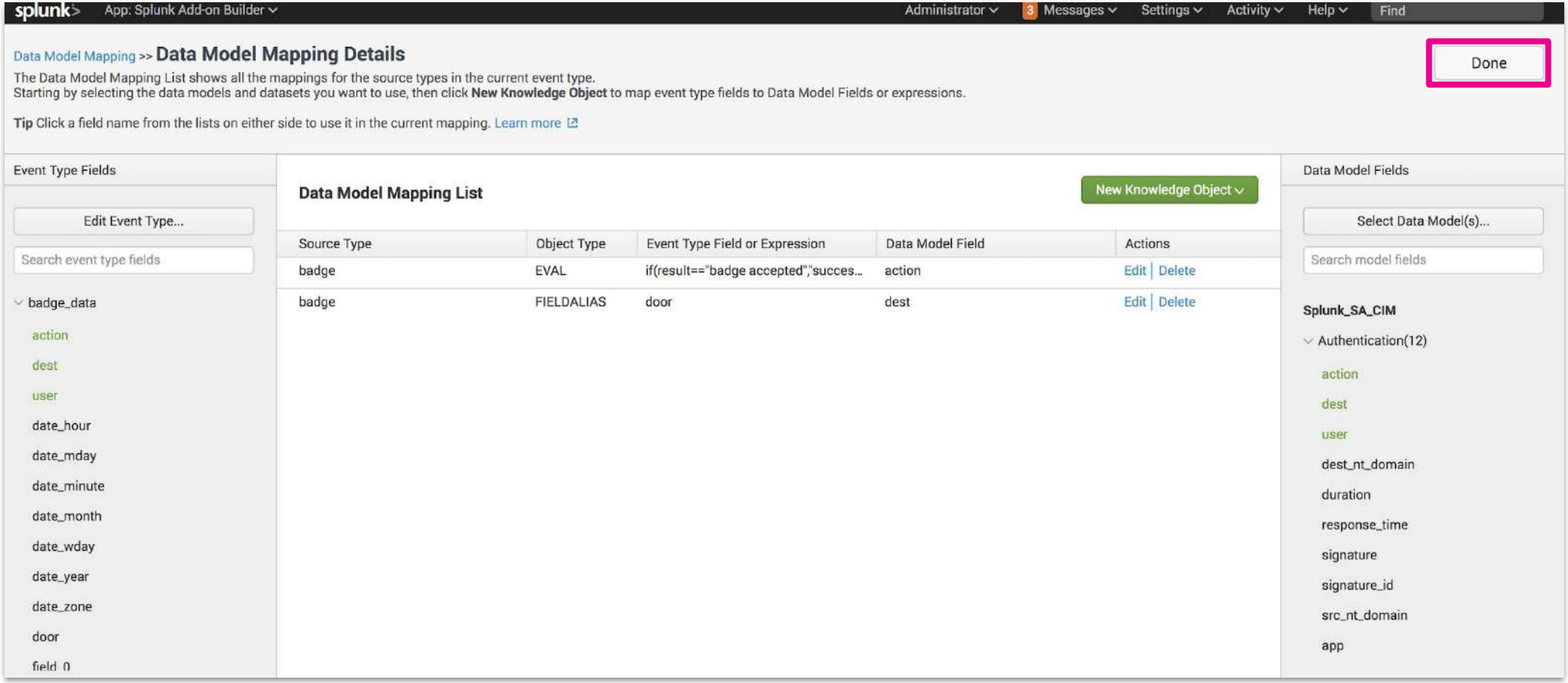
24. Select **action** on the right panel and enter the expression:

```
if(result=="badge accepted","success","failure")
```



Task 10: Create Reports on Badge Data

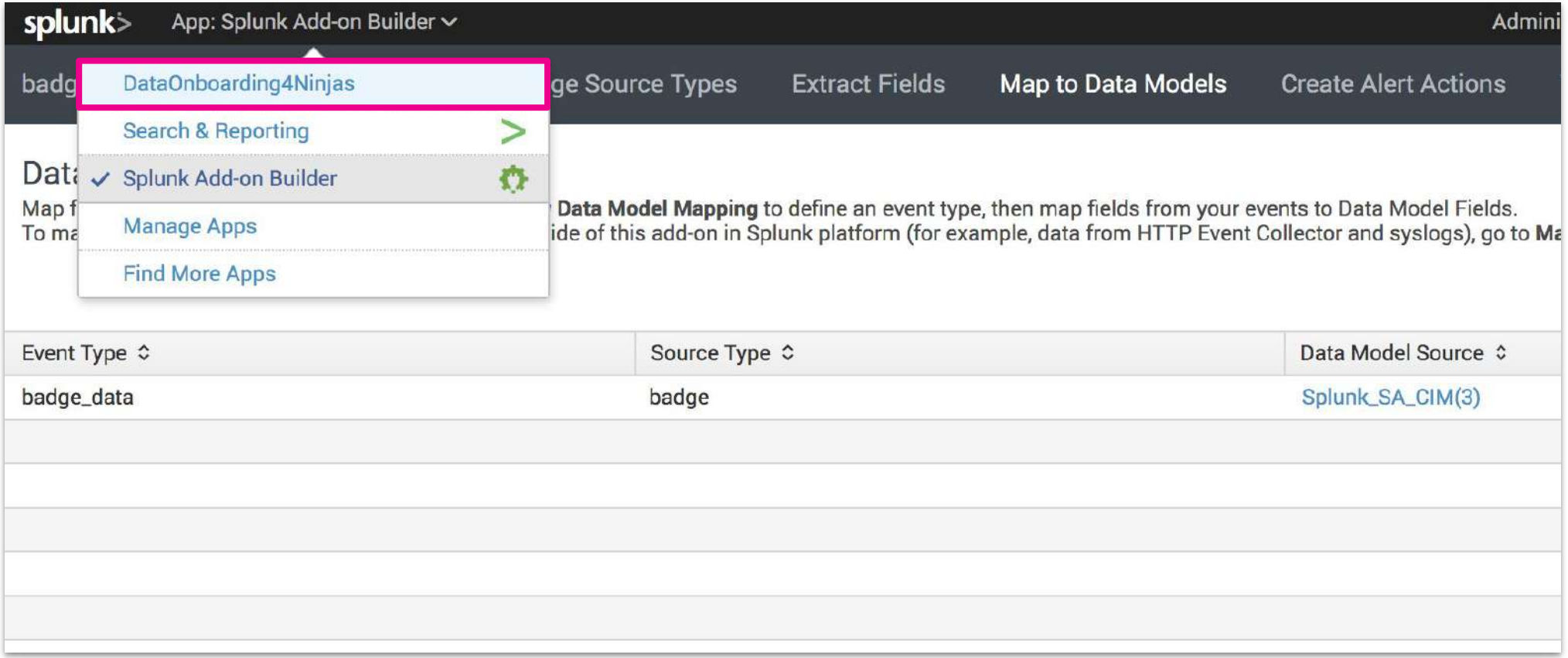
25. Click on **Done**



splunk>

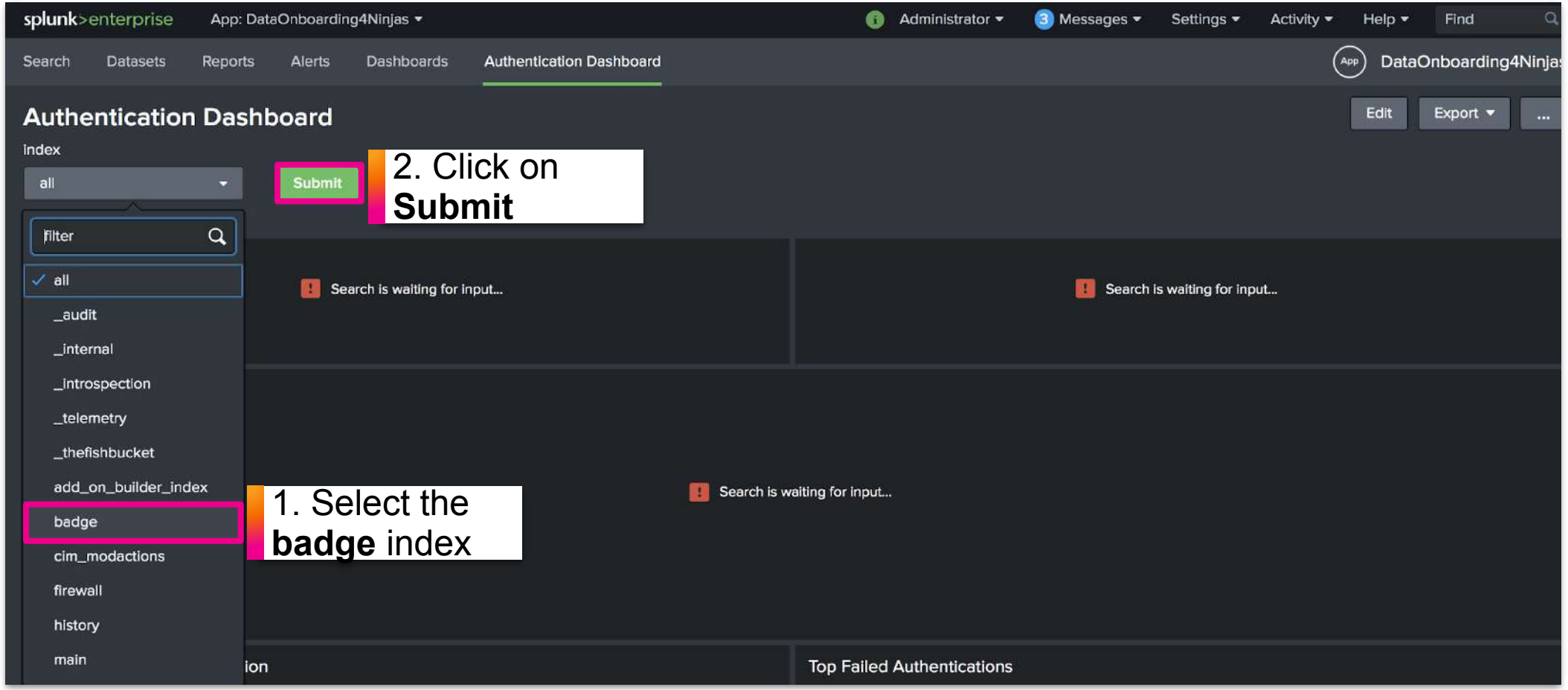
Task 10: Create Reports on Badge Data

26. Go to the **DataOnboarding4Ninjas** app



Task 10: Create Reports on Badge Data

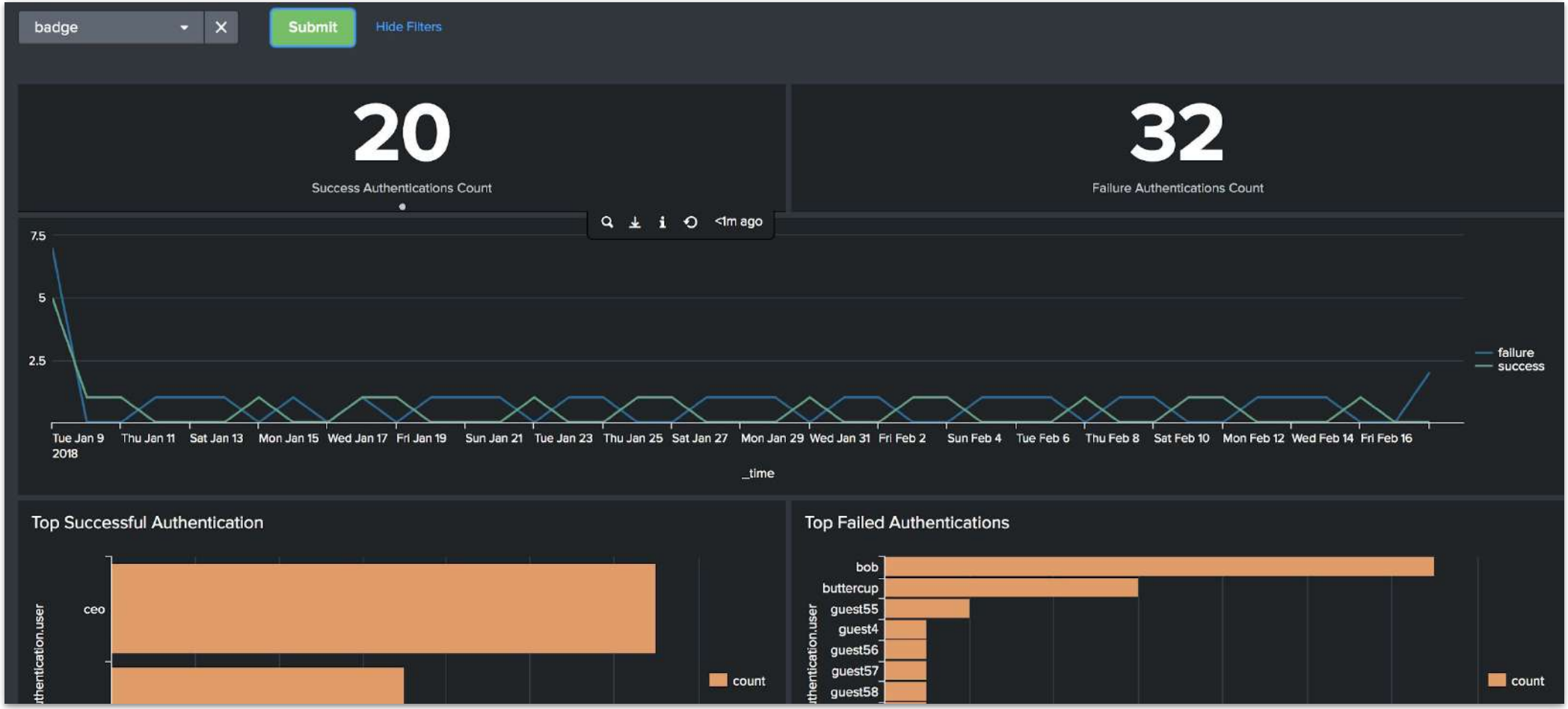
27. Check you dashboard again



splunk>

Task 10: Create Reports on Badge Data

Results should now be displayed correctly!



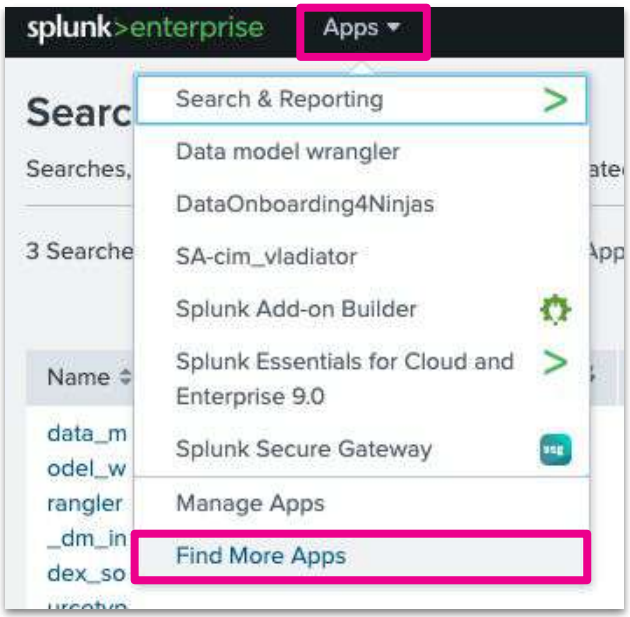
splunk>

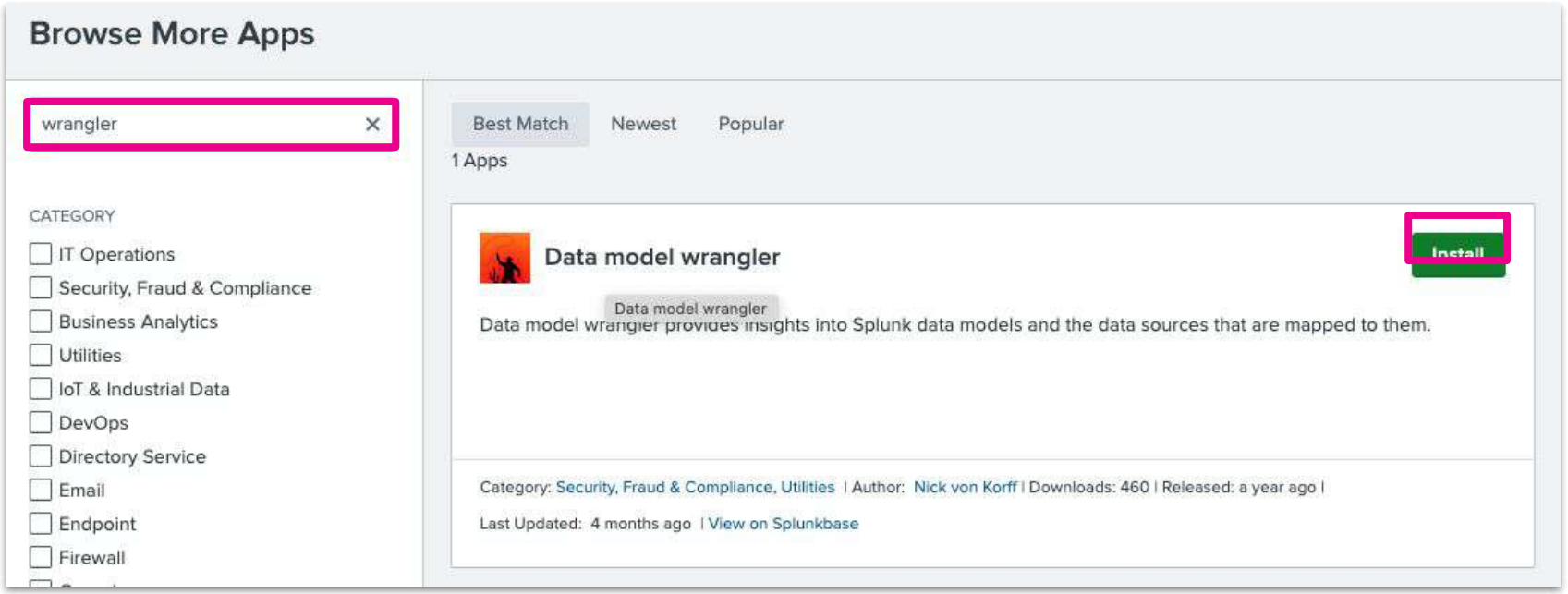
Thank You!

(Extra credit exercises below!)

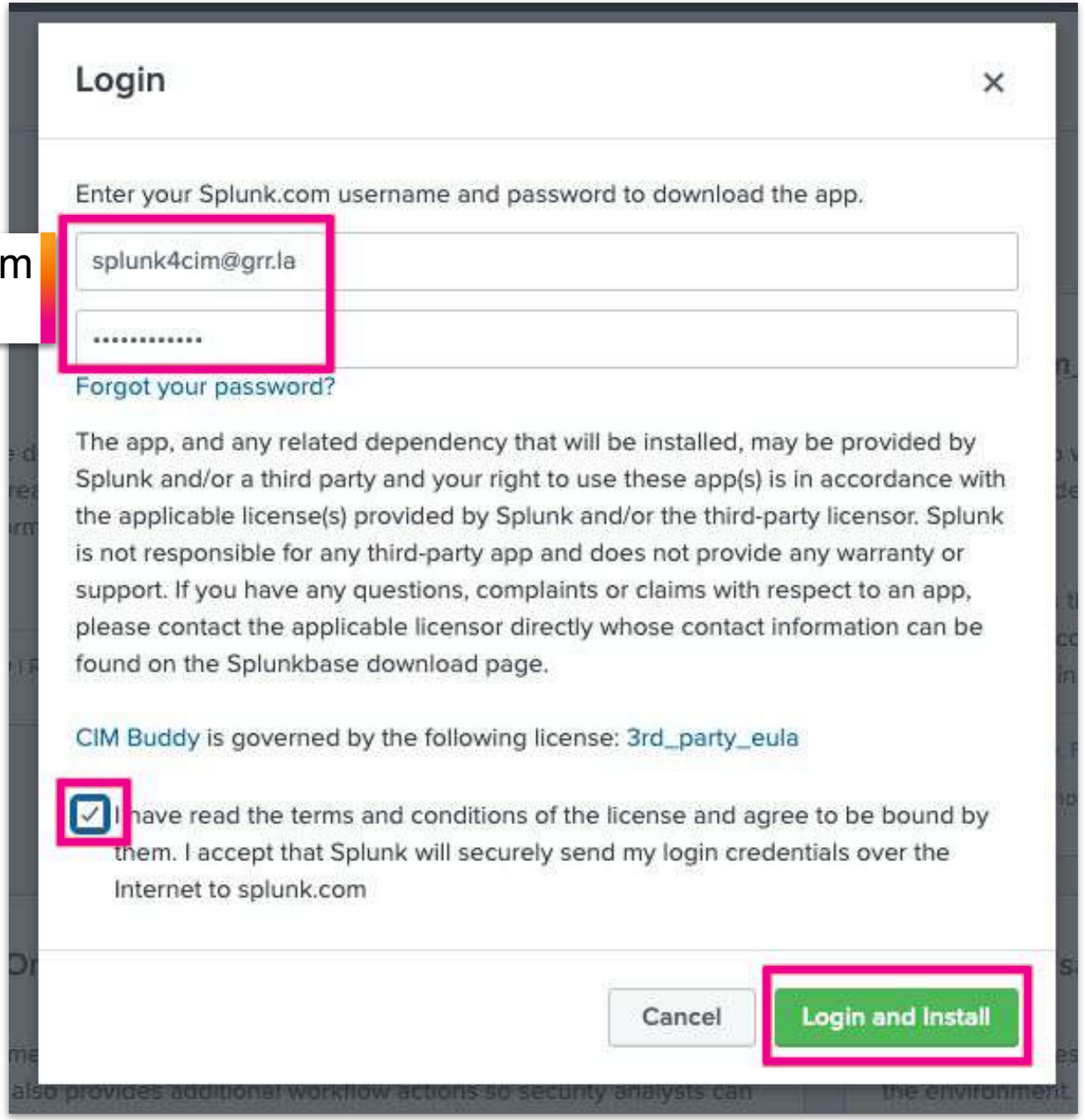
Extra Credit:
Install the Data Model Wrangler App

1. Go to **Apps > Find More Apps** and search for “**wrangler**”





Enter your splunk.com credentials



splunk>

Extra Credit: Wrangler App Pre-reqs

2. Review the prerequisites

How do I use the app?

- Ensure you have the Common Information Model app installed and up-to-date
- Install the CIM Validator app, as Data model wrangler relies on a custom search command from the CIM Validator app. There are also drill-downs from panels in the Data model wrangler to the CIM Validator.
- Create an index for the data that will be generated. Default index name is: *data_model_wrangler*
  - If you want to use a custom index name, please update the 'data_model_wrangler_index' search macro in the app with the custom index name
- Enable the below saved searches
  - data_model_wrangler_dm_index_sourcetype_field (default schedule is twice per day, once at 06:00 and again at 14:00. Time range: -4h@h to -3h@h)
  - data_model_wrangler_field_quality (default schedule is once per day at 01:00)
  - data_model_wrangler_mapping_quality (default schedule is once per day at 01:00)

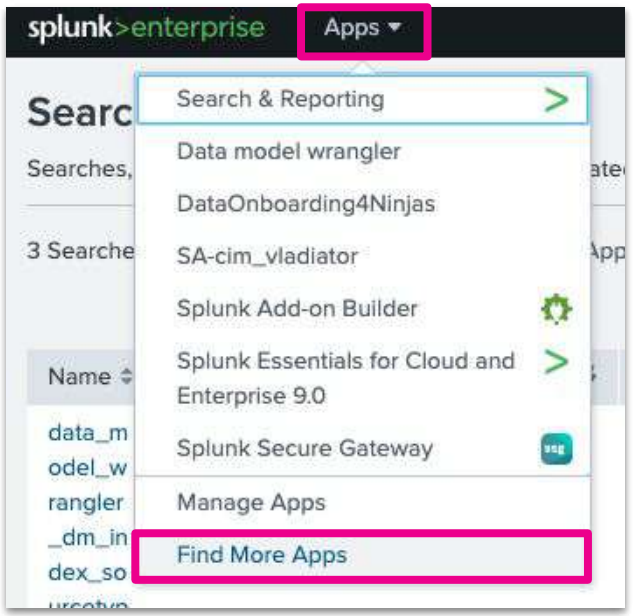
Over time this will build up a view of mapping and field data quality over time, allowing you to see the current state, find issues and measure the results.

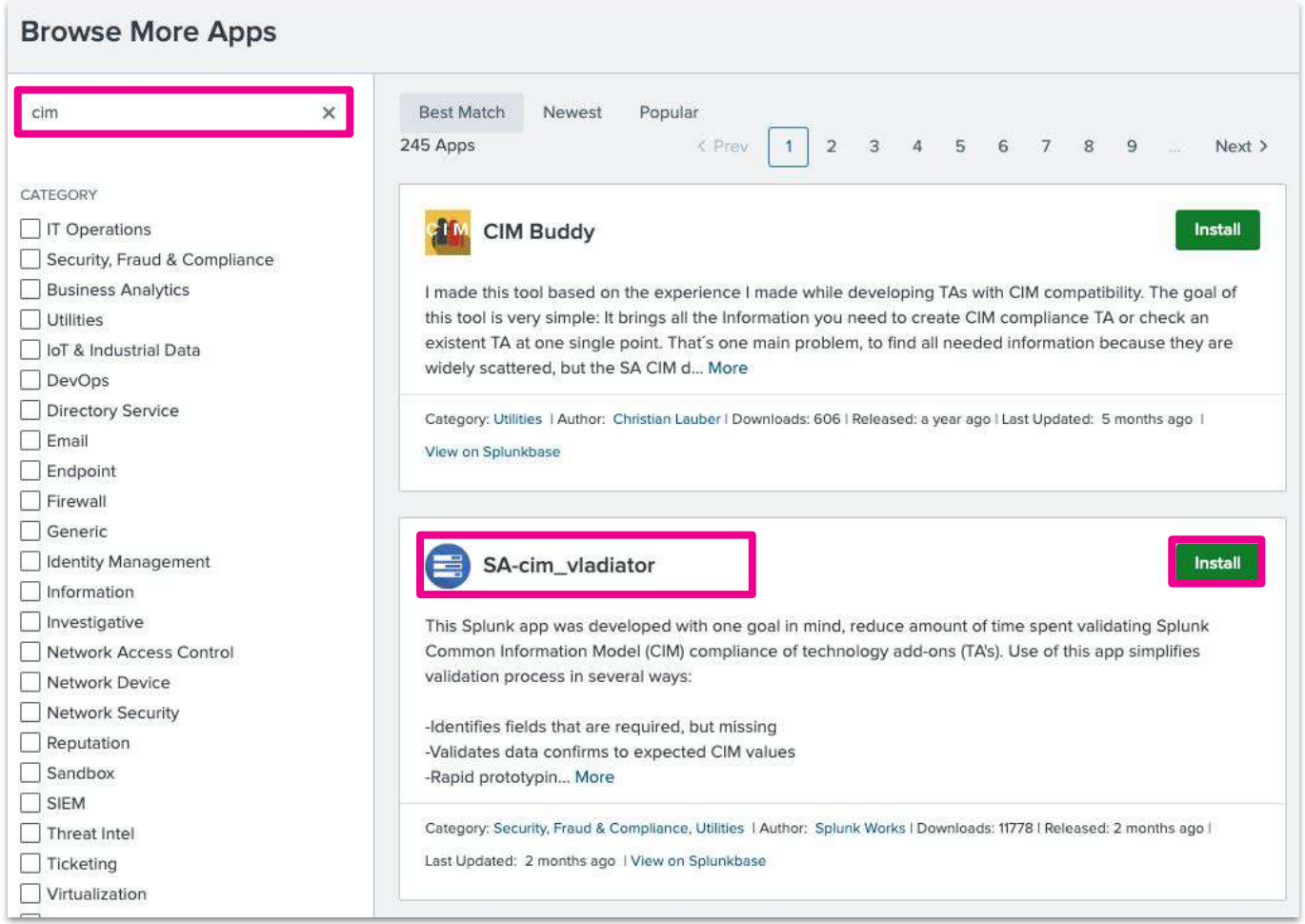
Whenever a CIM mapping requirement comes up, this dashboard provides a quick and easy way to find existing mapped data, or to verify that newly added data has been CIM mapped correctly.

splunk>

Extra Credit: Install the CIM Validator App

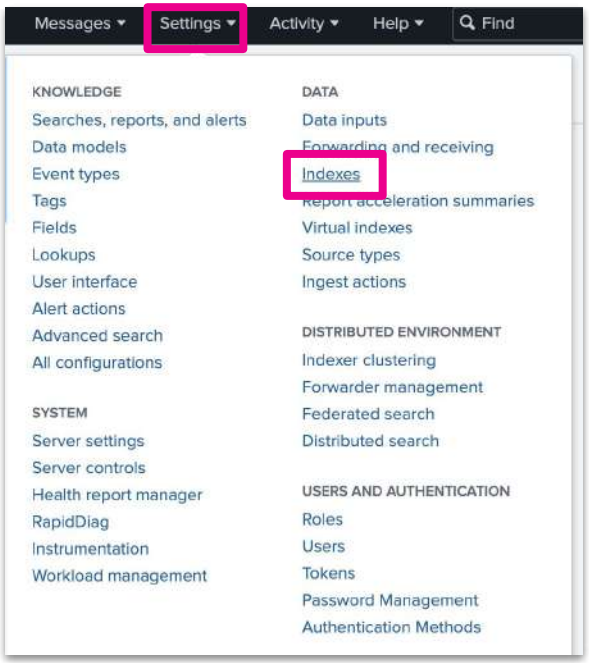
3. Install the **SA-cim_vladiator** app (search for "**cim**")

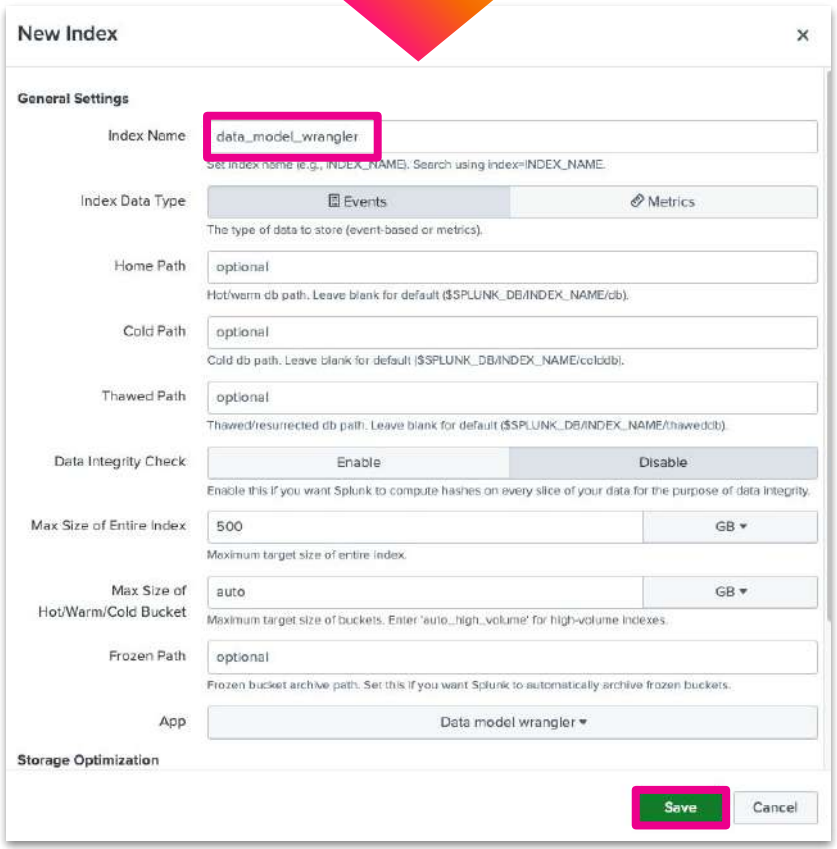


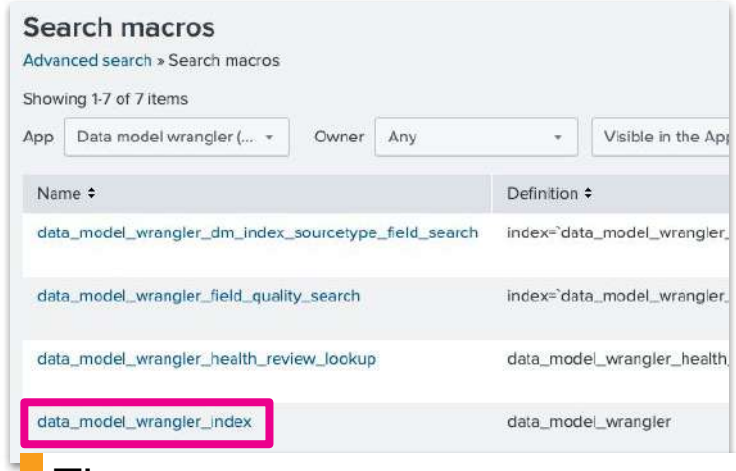


Extra Credit: Create an Index for the Data Model Wrangler App

4. Create a **data_model_wrangler** index





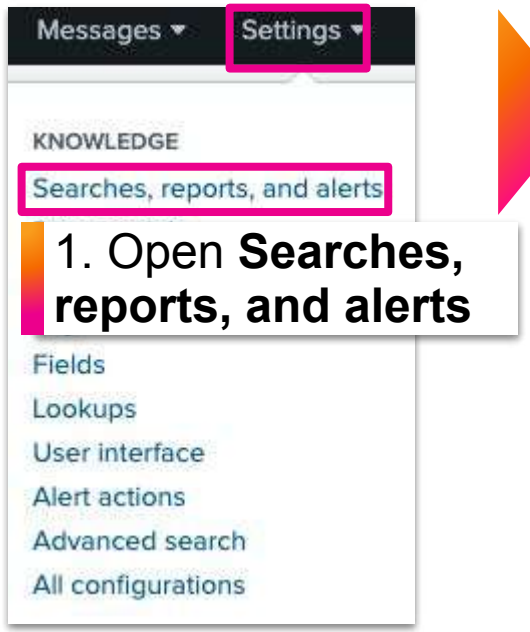


The **data_model_wrangler_index** macro is included in the Data Model Wrangler app and points to this index name

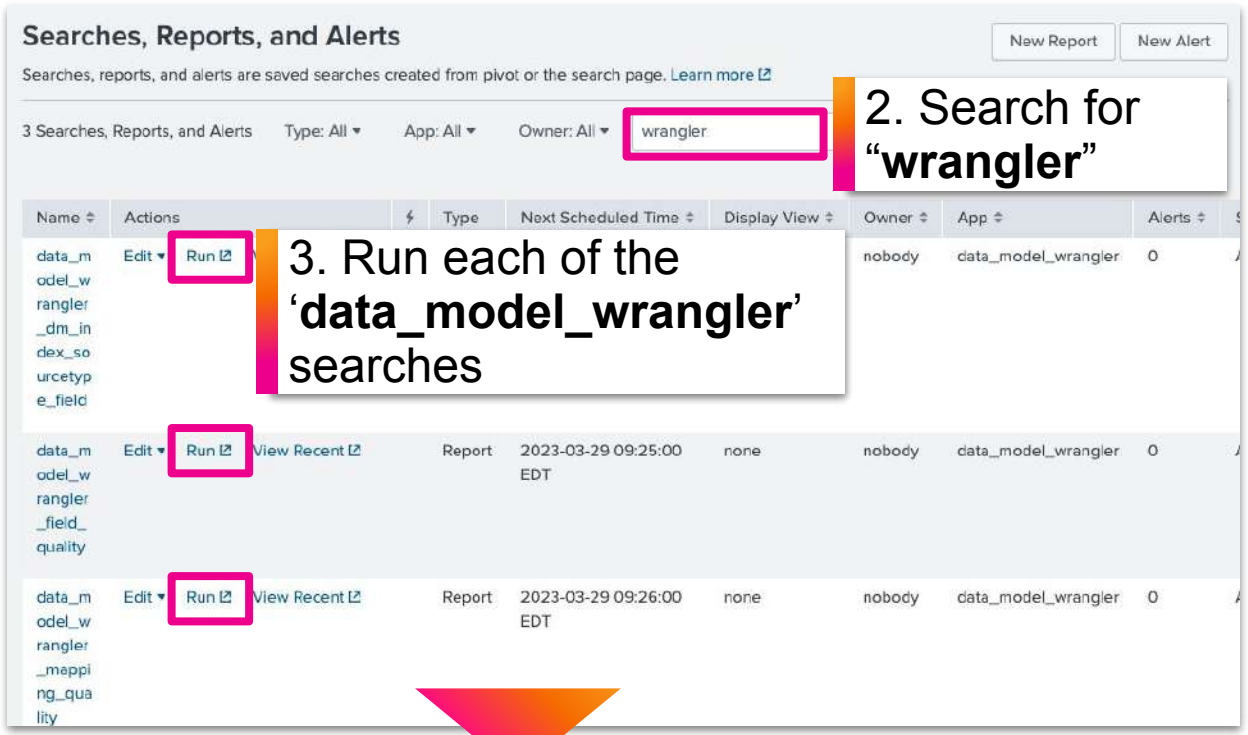
splunk>

Extra Credit: Run the Data Model Wrangler Saved Searches

5. Run the 3 saved searches (hint: set time to **All time**)

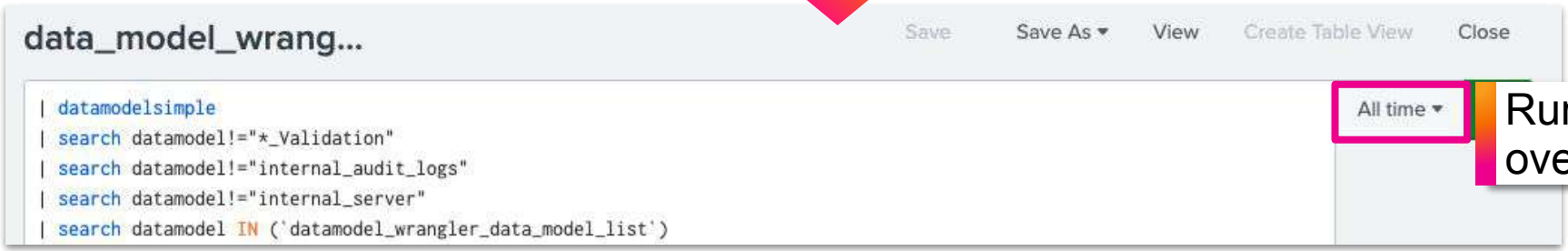


1. Open **Searches, reports, and alerts**



2. Search for "**wrangler**"

3. Run each of the '**data_model_wrangler**' searches

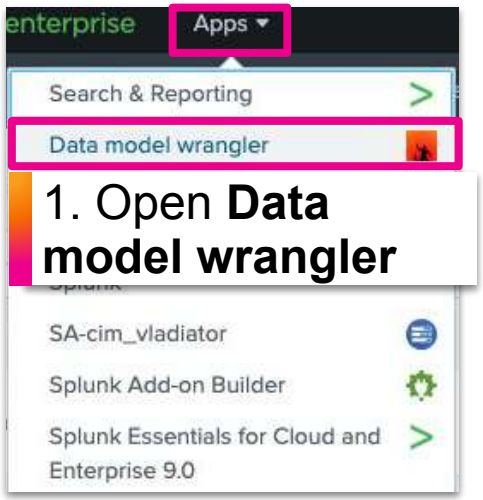


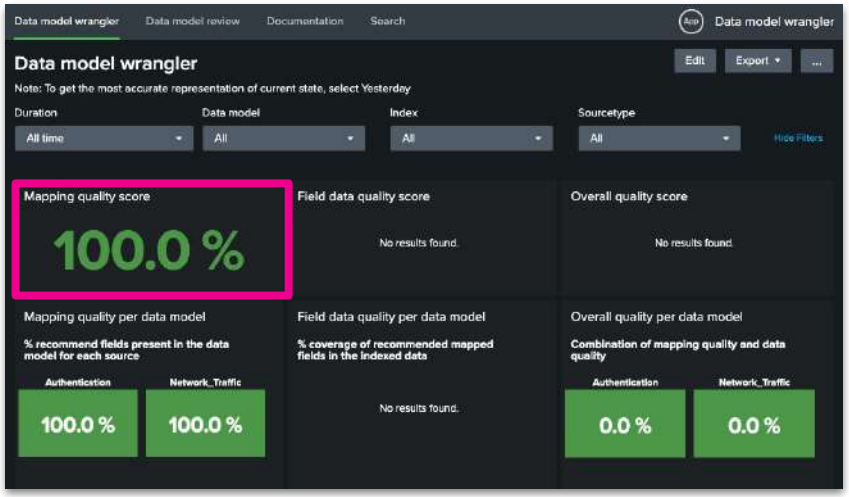
Run the searches over **All time**

splunk>

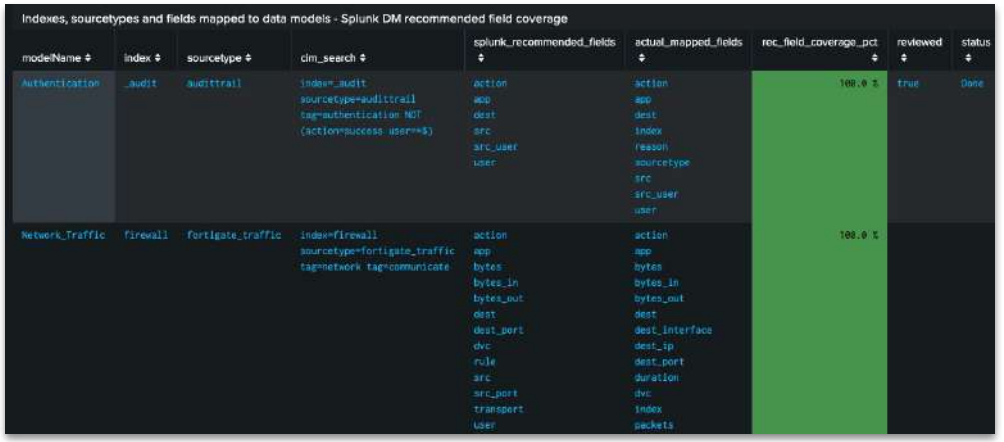
Extra Credit: Install the Data Model Wrangler App

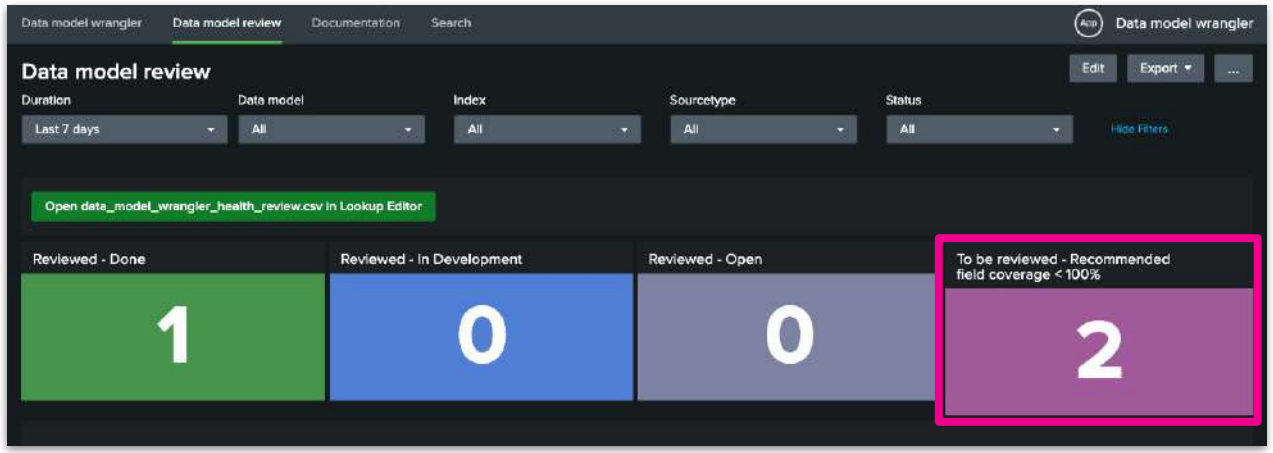
6. Review the Data Wrangler dashboards





⇩ (scroll down)





splunk>