



Machine Learning Event Workshop



Forward-looking statements

This presentation may be deemed to contain forward-looking statements, which are subject to the safe harbor provisions of the Private Securities Litigation Reform Act of 1995. Any statements that are not statements of historical fact (including statements containing the words "will," "believes," "plans," "anticipates," "expects," "estimates," "strives," "goal," "intends," "may," "endeavors," "continues," "projects," "seeks," or "targets," or the negative of these terms or other comparable terminology, as well as similar expressions) should be considered to be forward-looking statements, although not all forward-looking statements contain these identifying words. Readers should not place undue reliance on these forward-looking statements, as these statements are management's beliefs and assumptions, many of which, by their nature, are inherently uncertain, and outside of management's control. Forward-looking statements may include statements regarding the expected benefits to Cisco, Splunk and their respective customers from the completed transaction, the integration of Splunk's and Cisco's complementary capabilities and products to create an end-to-end platform designed to unlock greater digital resilience for customers, our expectations regarding greater resiliency and better product outcomes, including for security and observability, plans for future investment, our development and use of AI and the role that our innovation plays as our customers adopt AI. Statements regarding future events are based on Cisco's current expectations, estimates, and projections and are necessarily subject to associated risks related to, among other things, (i) the ability of Cisco to successfully integrate Splunk's market opportunities, technology, personnel and operations and to achieve expected benefits, (ii) Cisco's ability to implement its plans, forecasts and other expectations with respect to Splunk's business and realize expected synergies, (iii) the outcome of any legal proceedings related to the transaction, (iv) the effects on the accounting relating to the acquisition of Splunk, (v) legislative, regulatory, and economic developments, (vi) general economic conditions, and (vii) the retention of key personnel. Therefore, actual results may differ materially and adversely from the anticipated results or outcomes indicated in any forward-looking statements. For information regarding other related risks, see the "Risk Factors" section of Cisco's most recent report on Form 10-Q filed on February 20, 2024 and its most recent report on Form 10-K filed on September 7, 2023, as well as the "Risk Factors" section of Splunk's most recent reports on Form 10-Q filed with the SEC on February 20, 2024 and November 21, 2023, respectively. The parties undertake no obligation to revise or update any forward-looking statements for any reason, except as required by 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, Splunk> and Turn Data Into Doing are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names or trademarks belong to their respective owners.

© 2024 Splunk Inc. All rights reserved.



Please introduce yourself!

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



Workshop Agenda

- What is Splunk
- What is Machine Learning - Terminology
- Demystify ML/AI
- The Process
- Onboard and Cleanse Data
- Data Analysis and Insights
- Advanced Capabilities
- What next...

Splunk4Rookies Machine Learning

Clustering
Data Models
Alerting
Pivot
SDKs
APIs
DB Connect

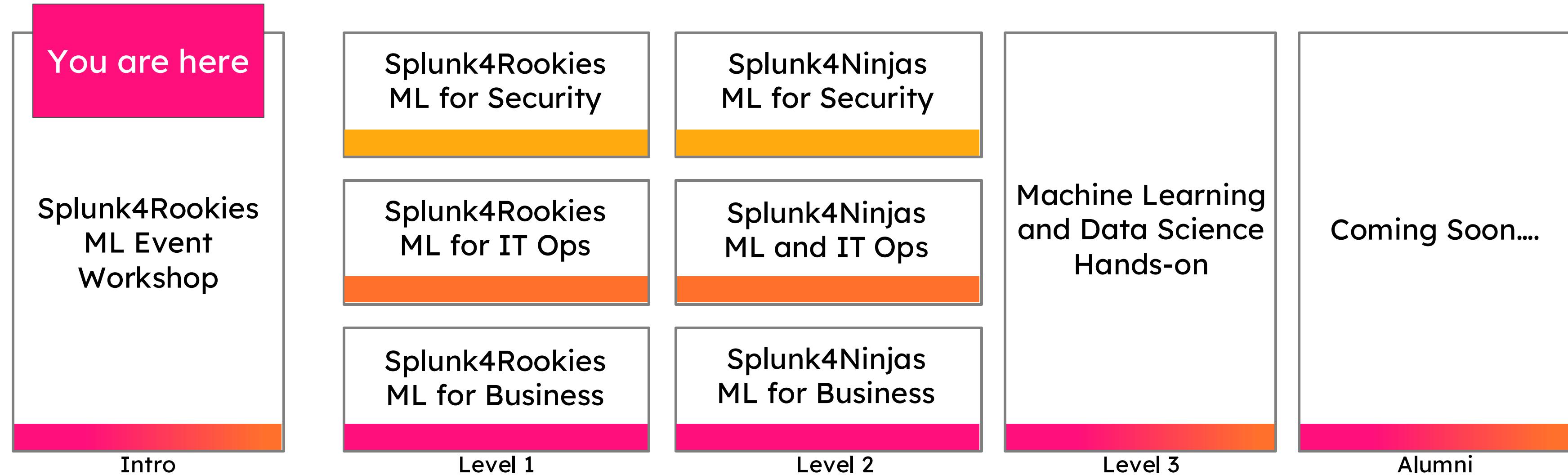
Advanced Searches
SOAR
Machine Learning (ML)

Splunk Stream
Deployment Server
Federated Search
Metrics

Custom Visualisations
HTTP Event Collector (HEC)
Data Filtering
Transformations
Architecture

Report Acceleration
Common Information Model (CIM)
Containers
Best Practices
And much more...

Splunk Machine Learning Workshops



Boss of Machine Learning (Coming Soon)

Introductory

Advanced

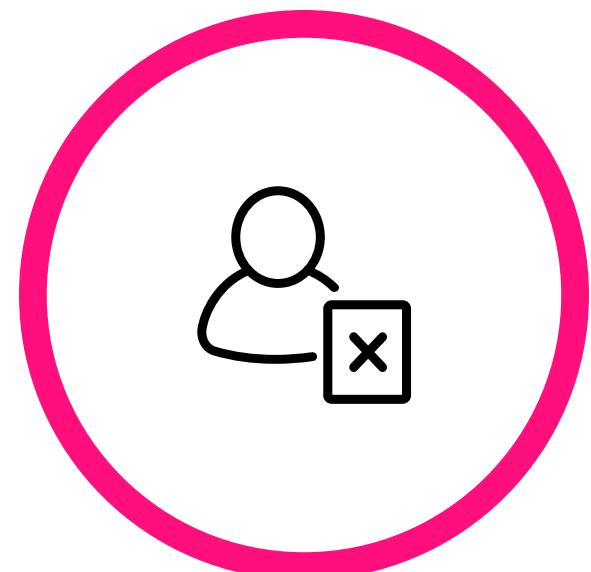
Legend

- Security
- IT Ops
- Business
- All

Objective for Today

An Understanding of

User Access Anomalies



Insider Threat

Anomalous Login Quantities



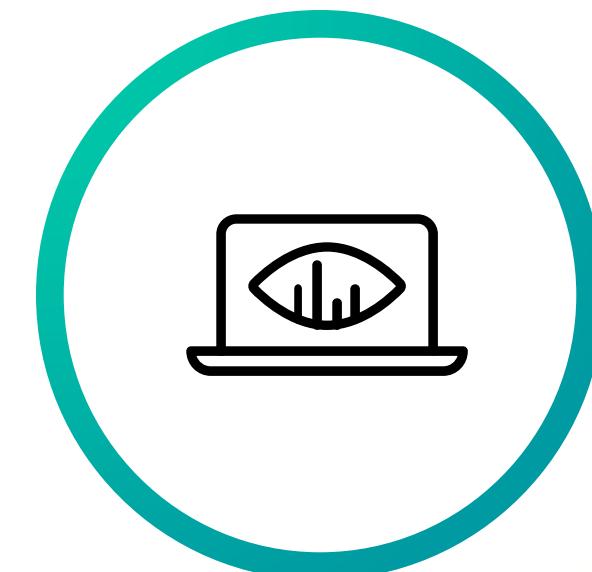
IT Operations

Forecast CPU Utilisation



Infrastructure

Outliers in Utilisation Telemetry



All



Enroll in Today's Workshop

Tasks

1. Get a splunk.com account if you don't have one yet:
<https://splk.it/SignUp>
1. Enroll in the Splunk Show workshop event:
<https://show.splunk.com/event/<eventID>>
2. Download a copy of today's slide deck:
<https://splk.it/S4RML-Event-Attendee>

Goal

The screenshot shows a user interface for managing event enrollments. At the top, there are tabs for 'Available' and 'Enrolled', with 'Available' being the active tab. Below the tabs is a search bar with placeholder text 'Search events' and a dropdown menu set to 'Invited'. The main area displays a list of events. One event is highlighted: 'Splunk4Rookies - Machine Learning Primer', which is a 'Private event'. The event details show it's scheduled for '05/22 (12:07) - 05/22 (20:06)' in 'Europe/London (GMT +01:00)'. To the right of the event card, there is a green button labeled 'Available'. In the bottom right corner of the slide, there is a call-to-action button with the text 'Enroll in today's event'.

Enroll in today's event

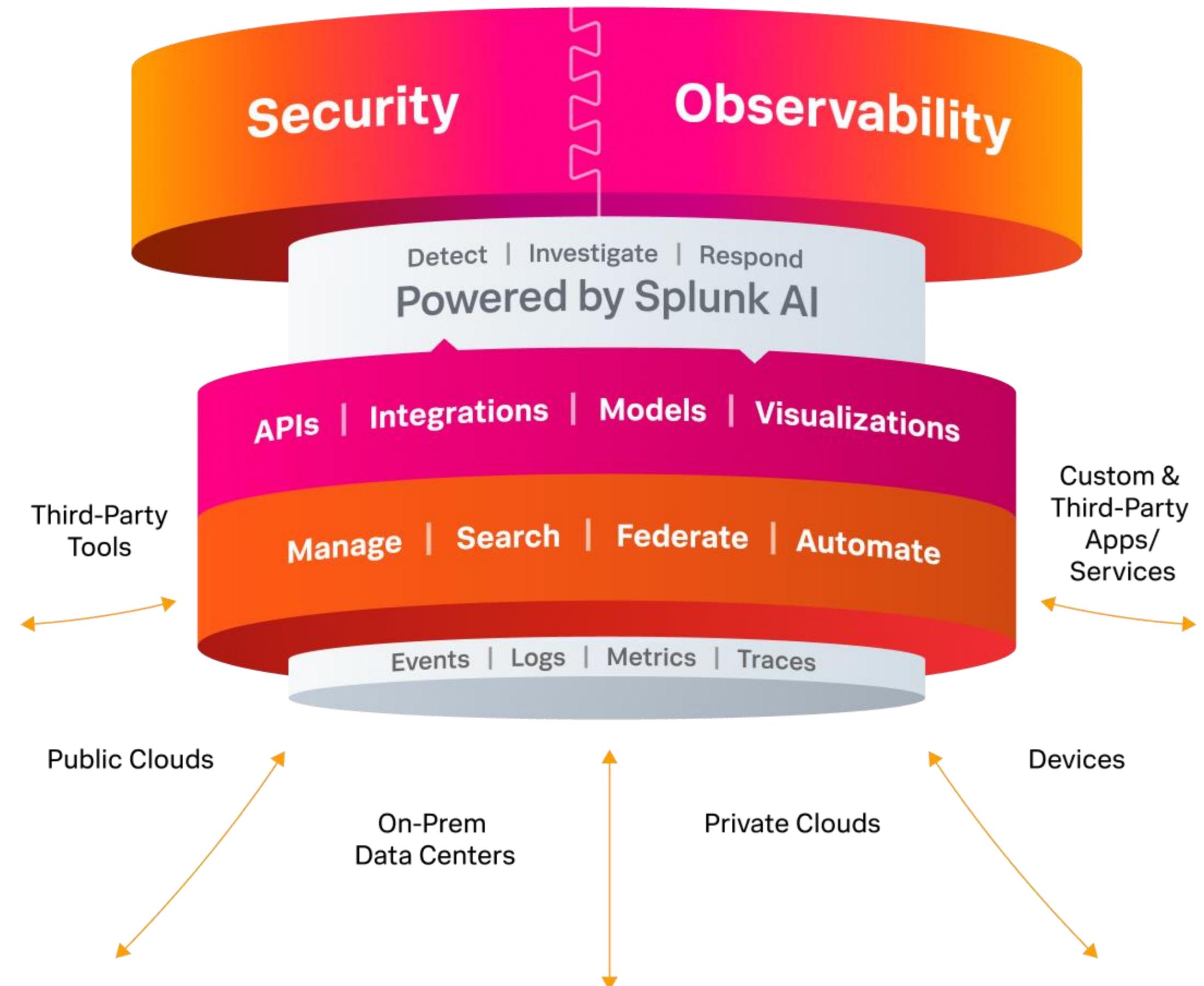
What is Splunk?



Our **product** vision.

To provide visibility and insights across an enterprise's **entire digital footprint**, powering actions that improve security, reliability and innovation velocity

The Unified Security and Observability Platform



Power the SOC of the future.

Strengthen digital resilience by modernizing your SOC with unified threat detection, investigation and response.

5x

faster response
to threats

70%

time reduction
researching
security events

80%

potential
reduction in
alert volumes

Data
optimization

Security
monitoring

Risk-based
alerting

Anomaly
detection

Automation &
orchestration

Automate complete
TDIR life cycle

Deliver superior digital services.

Find and fix problems faster, improve reliability, and build exceptional customer experiences.

97%

reduction in alert noise

83%

reduction in mean time to resolve (MTTR)

96%

faster application development, increasing developer productivity

Troubleshooting mission-critical apps & infrastructure

Analyze IT service health

Debug problems in microservices

Optimize cloud monitoring

Understand impact of change

Enable self-service observability

Unified platform to maximize data value and solve limitless use cases.

Monitor, investigate and respond rapidly at scale with comprehensive visibility and shared tooling.

3K+

apps &
add-ons on
Splunkbase

1K+

purpose built
data source
integrations

8B

monthly
searches

Flexible deployment options

Scalable index and storage

Expanded view with federated analytics

End-to-end data management

Powerful search and visualization

Increased efficiency with AI and ML

AI and ML capabilities across our portfolio accelerate detection, investigation, and response.

Our approach



Domain and Splunk specific



Human-in-the-loop and trusted



Open and extensible

What we offer

Generative AI

Make sense of the signal to improve user productivity and outcomes

Foundational AI

Find the signal from the noise in vast amounts of data

Cisco + Splunk power and protect the AI revolution.

Infrastructure for AI

Networking and compute solutions for training AI models at scale.

Data for AI

Massive breadth and depth of data across domains is the foundation for AI models.

Observability for AI

Observability supports trustworthy, performant, and reliable AI deployment.

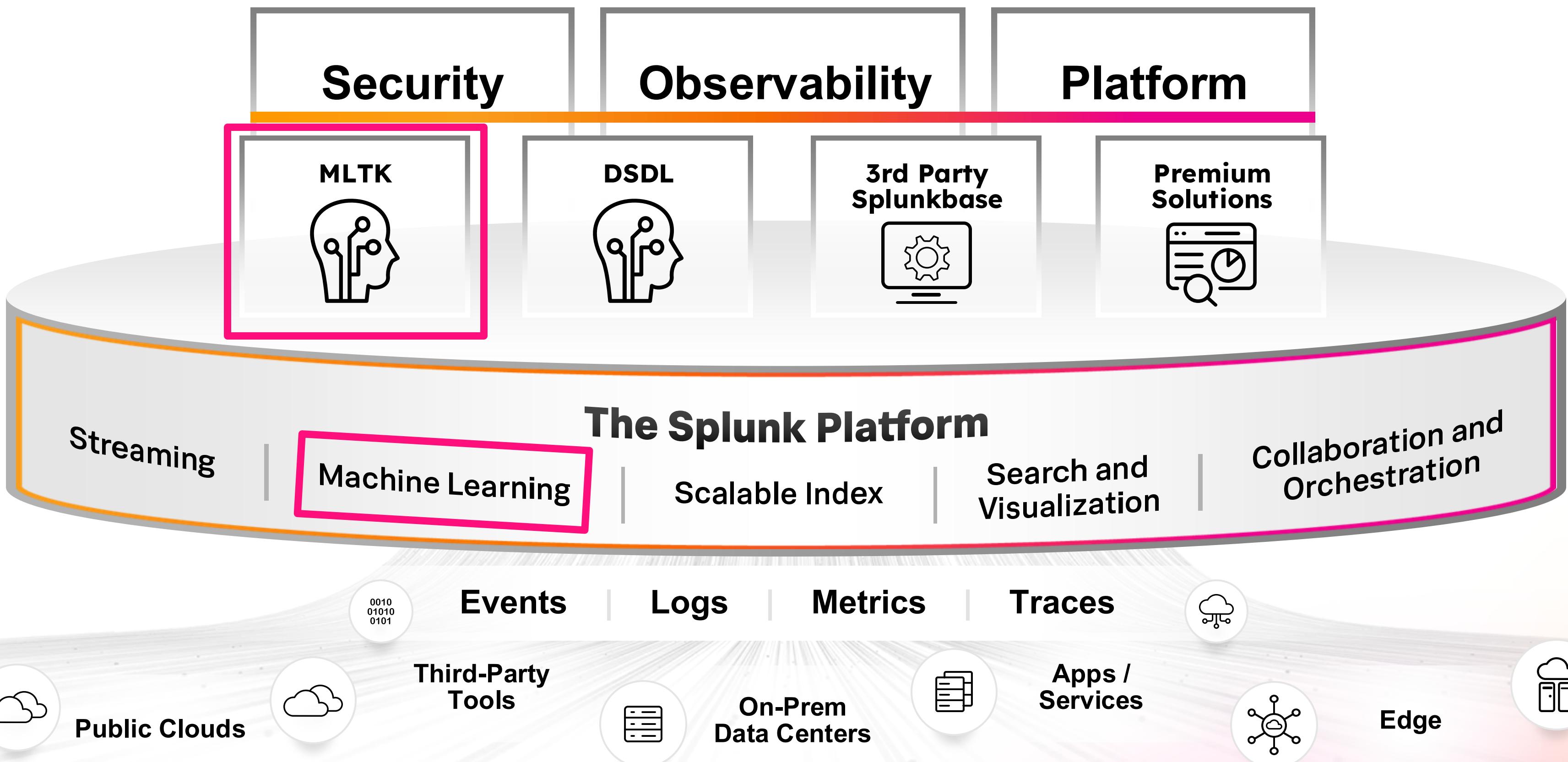
AI for Security

AI solutions improve efficiency, efficacy, and economics of defending against security threats.

Security for AI

Cisco's Identity Intelligence and User Protection and Splunk's analytics help protect enterprises in their use of AI from within.

The Unified Security and Observability Platform



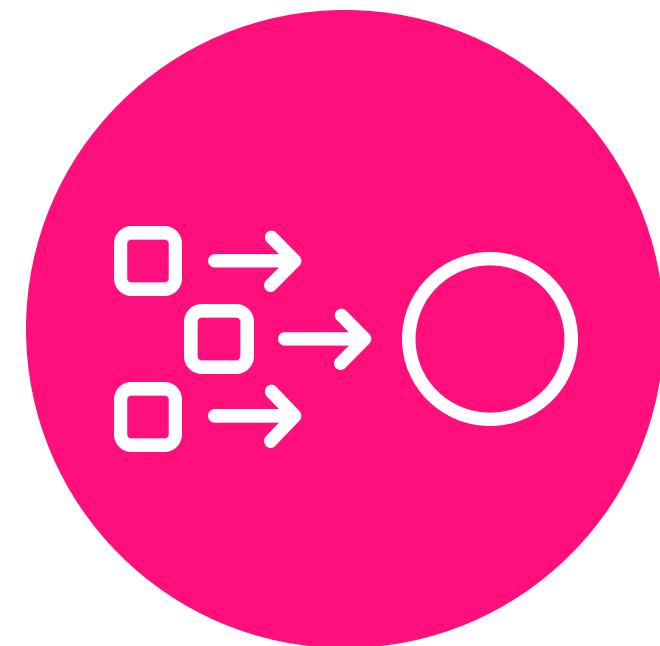
Machine Learning

The Truth



Machine Learning

The capability of systems to acquire their own knowledge, by extracting patterns



Use complex statistical or probabilistic models to identify patterns in information

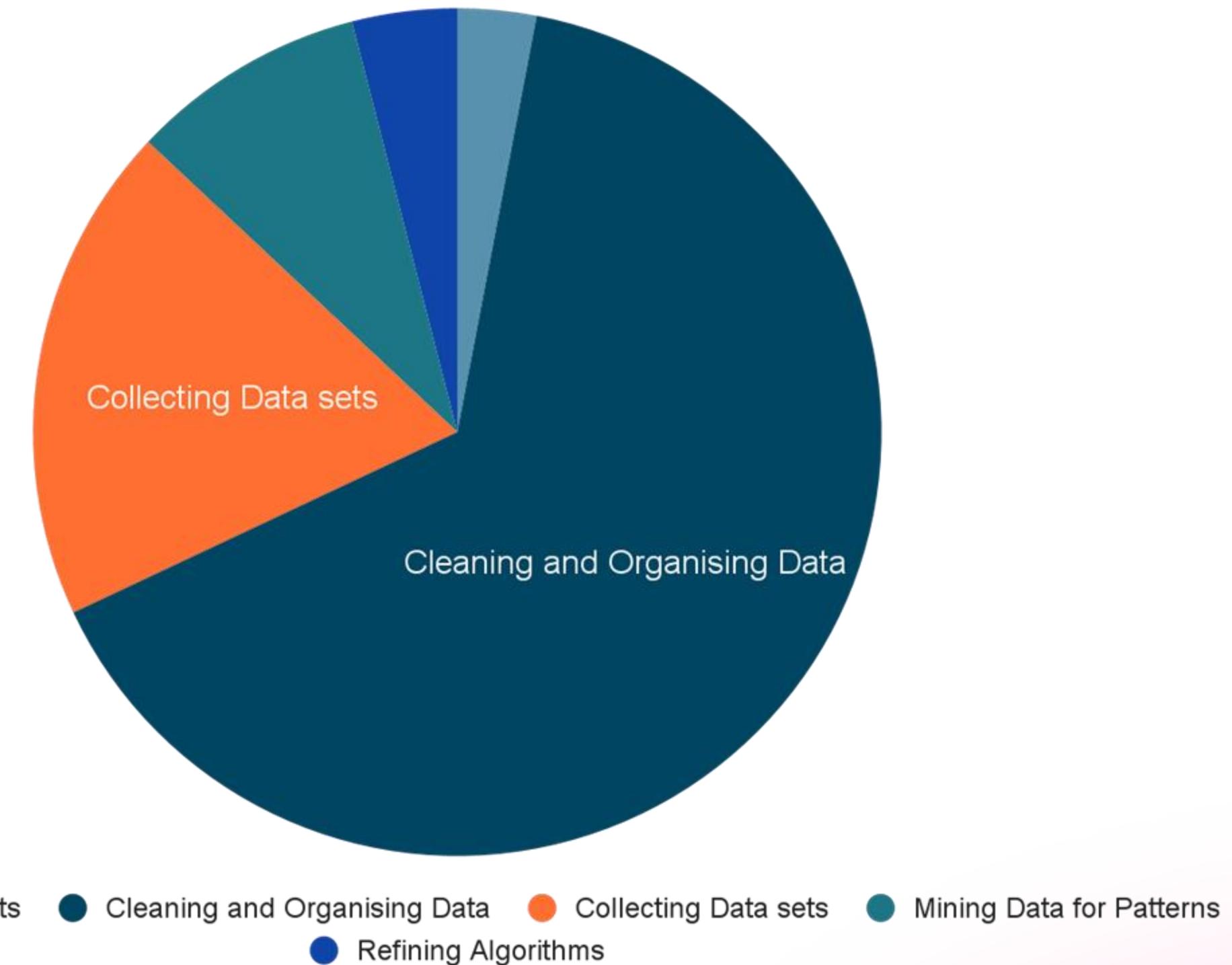


Catalog the patterns and in some cases iterate them as new data is received



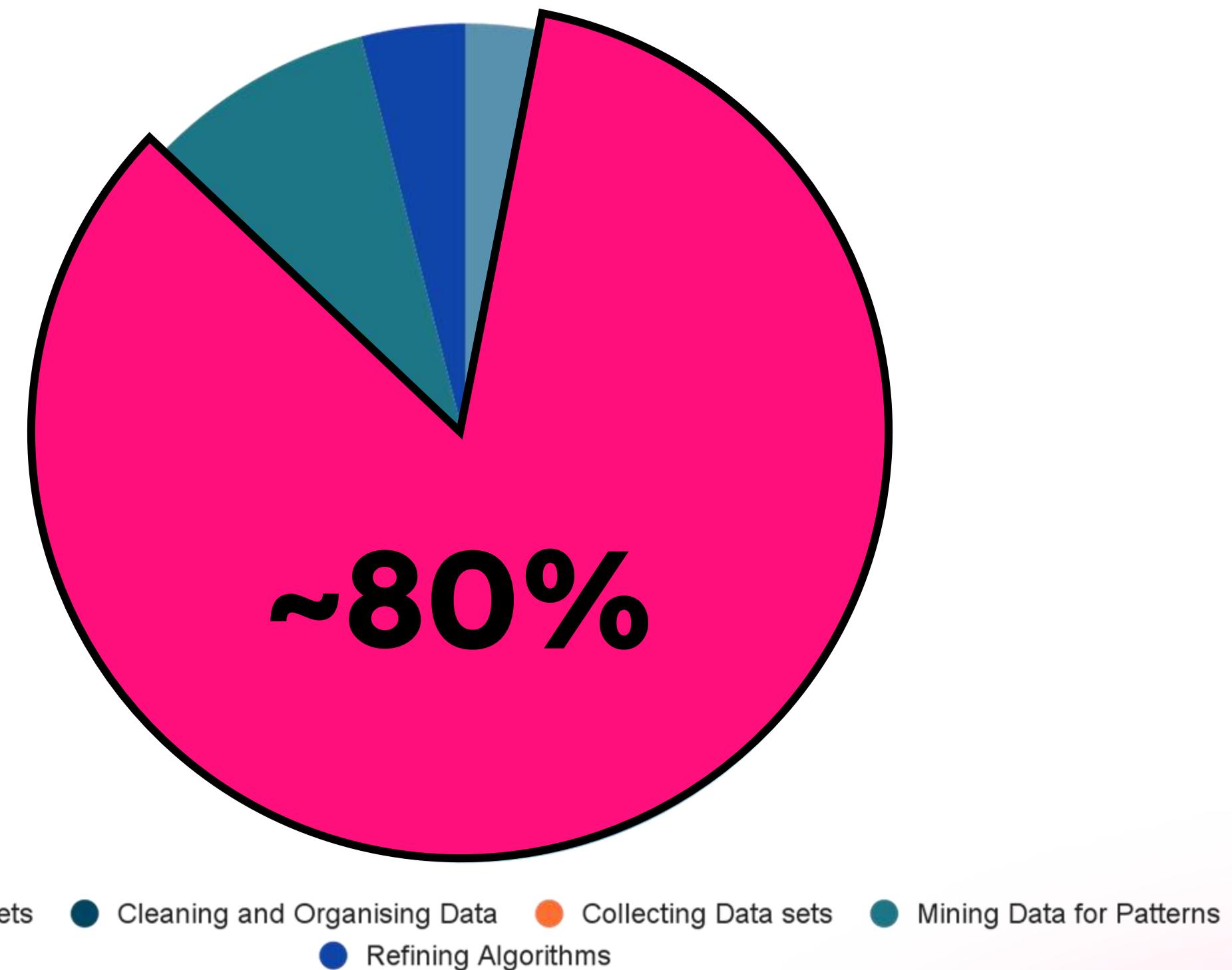
Use information from learnt patterns to understand and interpret new data or make predictions

What Data Scientists Really Do



What Data Scientists Really Do

Data Preparation accounts for about 80% of the work of data scientists



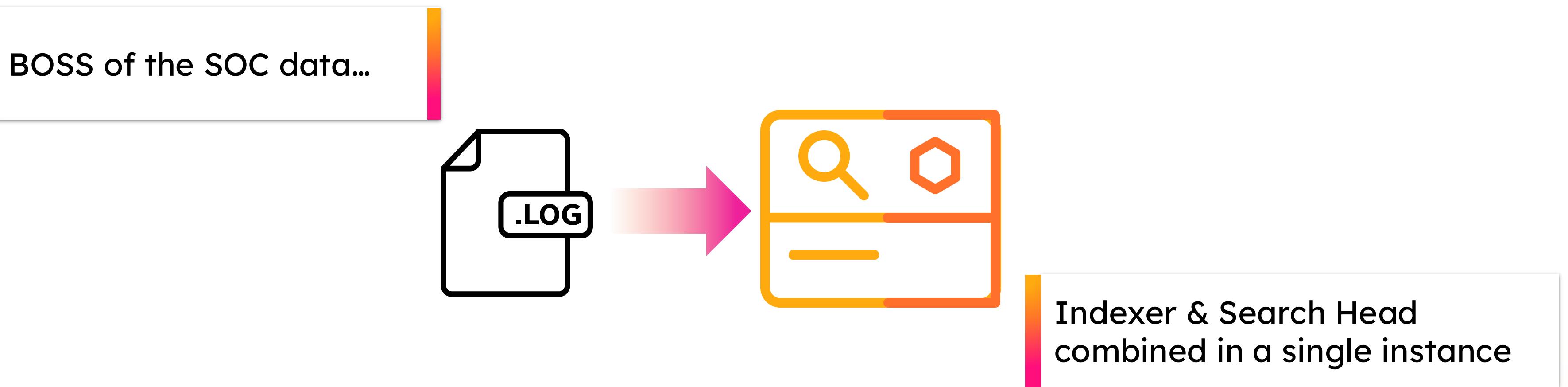


Cleaning Big Data, the Most Time-Consuming, Least Enjoyable Data Science Task

Forbes Survey

March 23

Today's Environment



Hands-On

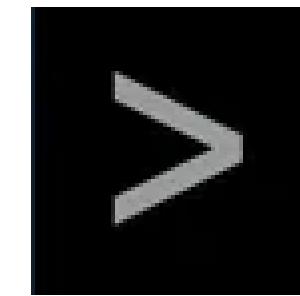


One Last Thing...

The following tools & datasets have been provided in your Splunk instance for this workshop:



Machine Learning Toolkit (MLTK)



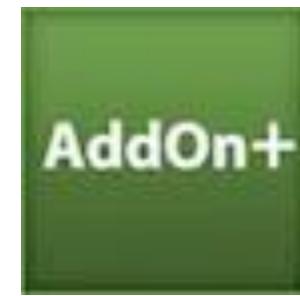
Python Scientific Computing (PSC)



Splunk App for Anomaly Detection



BOTS v2 Dataset (Full)

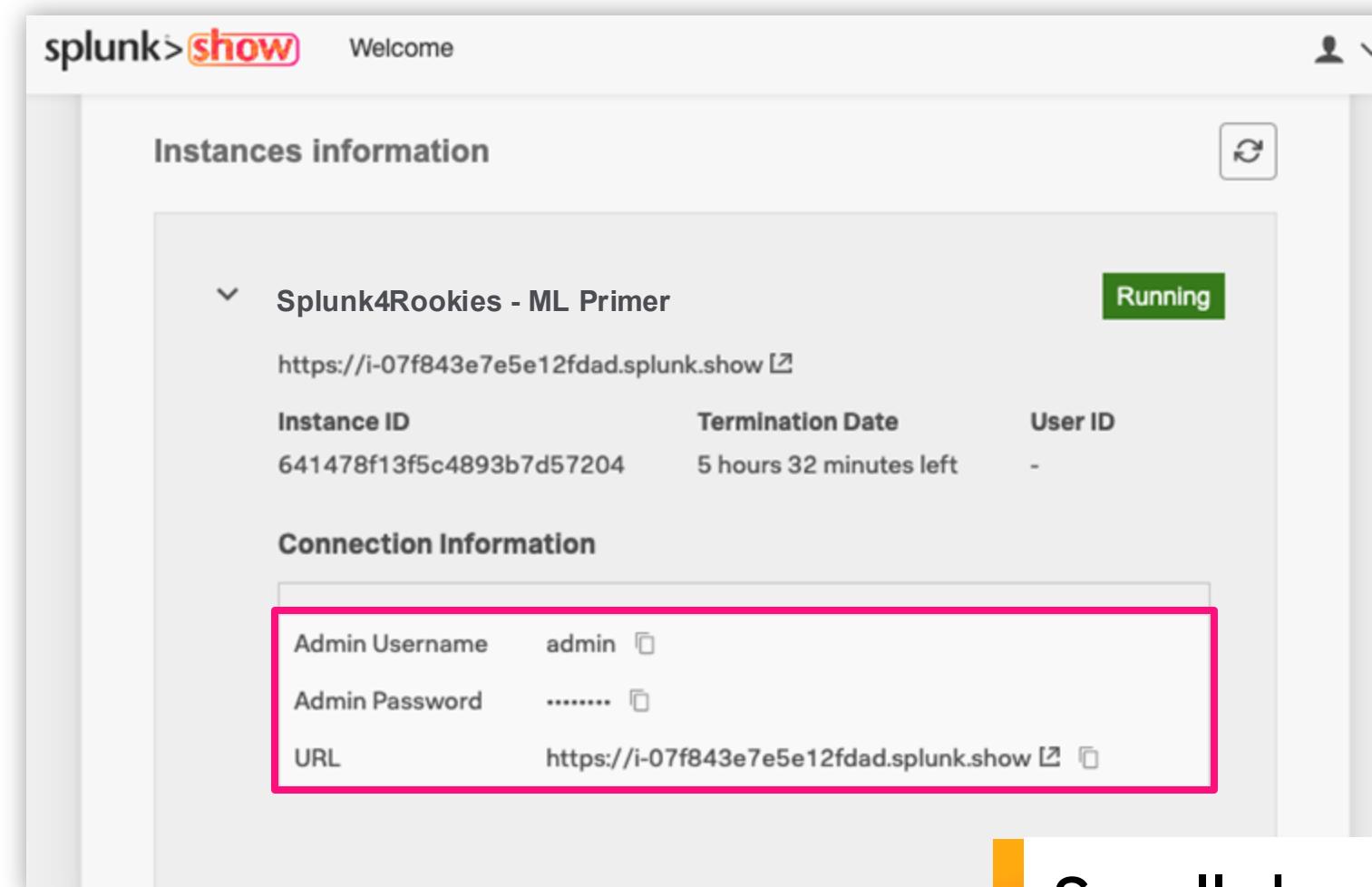


Splunk Add-on for Microsoft Windows

If you would like to run through this workshop in your own Splunk environment, please ensure you have the above listed apps and add-ons installed in your Splunk instance.

Login to Splunk

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



The screenshot shows the Splunk Show event page. At the top, it says "splunk>show Welcome". Below that, under "Instances information", there is a list for "Splunk4Rookies - ML Primer". It shows the instance ID as "641478f13f5c4893b7d57204", the termination date as "5 hours 32 minutes left", and the status as "Running". Under "Connection Information", there are fields for "Admin Username" (admin), "Admin Password" (redacted), and "URL" (<https://i-07f843e7e5e12fdad.splunk.show>). The "Admin Password" field is highlighted with a red border.

Log in to your Splunk instance



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



A Buttercup Company

Not just “a neigh”
Old Beer

Frothly Schwag Brewers Yeast **Frothly Beverages**

Home > Frothly Beverages

Frothly Beverages

Shopping Options

PRICE

Compare Products

You have no items to compare.

My Wish List

You have no items in your wish lis

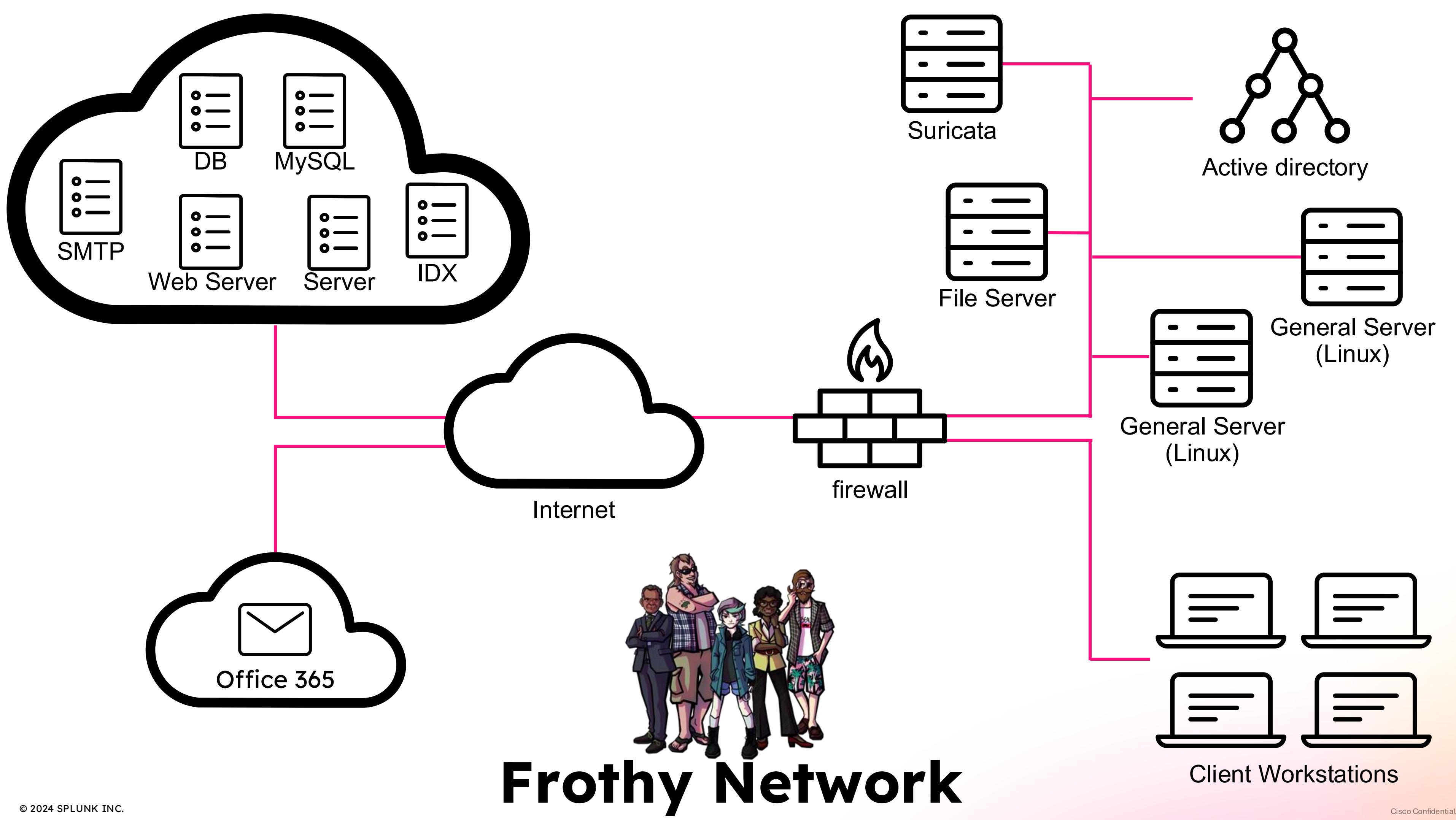
\$16.00

\$15.50

\$25.00

Sort By Position

Show 9



Explore

Understanding the Data and
Navigating Splunk



The Data

Search 0-1

The screenshot shows the Splunk Enterprise search interface. The top navigation bar includes 'splunk>enterprise' and 'Administrator'. The main search bar contains the query 'index=botsv2 earliest=0'. Below the search bar, it says '464,475 of 464,475 events matched' and 'No Event Sampling'. The search results are displayed in a table format under the 'Events (464,475)' tab. The table has columns for 'Time' and 'Event'. A selected event is shown in detail, containing fields like host, source, sourcetype, and a large JSON object representing event data. The event timestamp is 31/08/2017 23:37:16.771. The event data includes accept, accept_language, bytes, bytes_in, bytes_out, c_ip, content_encoding, cookie, dest_content, and a large script tag. The interface also shows 'INTERESTING FIELDS' on the left, including action, collection, counter, data, data_type, database_name, date_hour, date_mday, date_minute, date_month, and date_second.

Time	Event
31/08/2017 23:37:16.771	{ [-] accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8 accept_language: en-US,en;q=0.8 bytes: 13187 bytes_in: 944 bytes_out: 12243 c_ip: 106.253.110.210 content_encoding: gzip cookie: mage-translation-storage=%7B%7D; mage-translation-file-version=%7B%7D; form_key=m4dSsC4sPxUW25q3; mage-cache-storage=%7B%7D; mage-cache-storage-section-validation=%7B%7D; mage-cache-sessid=true; rec dest_content: <!doctype html> <html lang="en-US"> <head prefix="og: http://ogp.me/ns# fb: http://ogp.me/ns/fb# product: http://ogp.me/ns/product#"> <script> var require = { "baseUrl": "http://store.froth.ly/magento2/pub/static/version1501014192/frontend/Magento/luma/en_US" }; </script>

Scenario 1

Detecting User Access Anomalies



Detecting User Access Anomalies

You're a SOC analyst at Frothly, who's just been notified that some accounts in the company might have been hacked.

You're tasked with tracing back to where this hack may have originated, and identifying the potentially compromised accounts in the system.

To do this, you need to analyse Windows Security Event logs that contain records of successful and unsuccessful login/logout activity, as well as other security-related events



Authentication Events

Search 1-1

The screenshot shows the Splunk Enterprise search interface. The search bar contains the following query:

```
index=botsv2 source="wineventlog:security" (EventCode=4625 OR EventCode=4624)  
| timechart span=1h count by EventCode
```

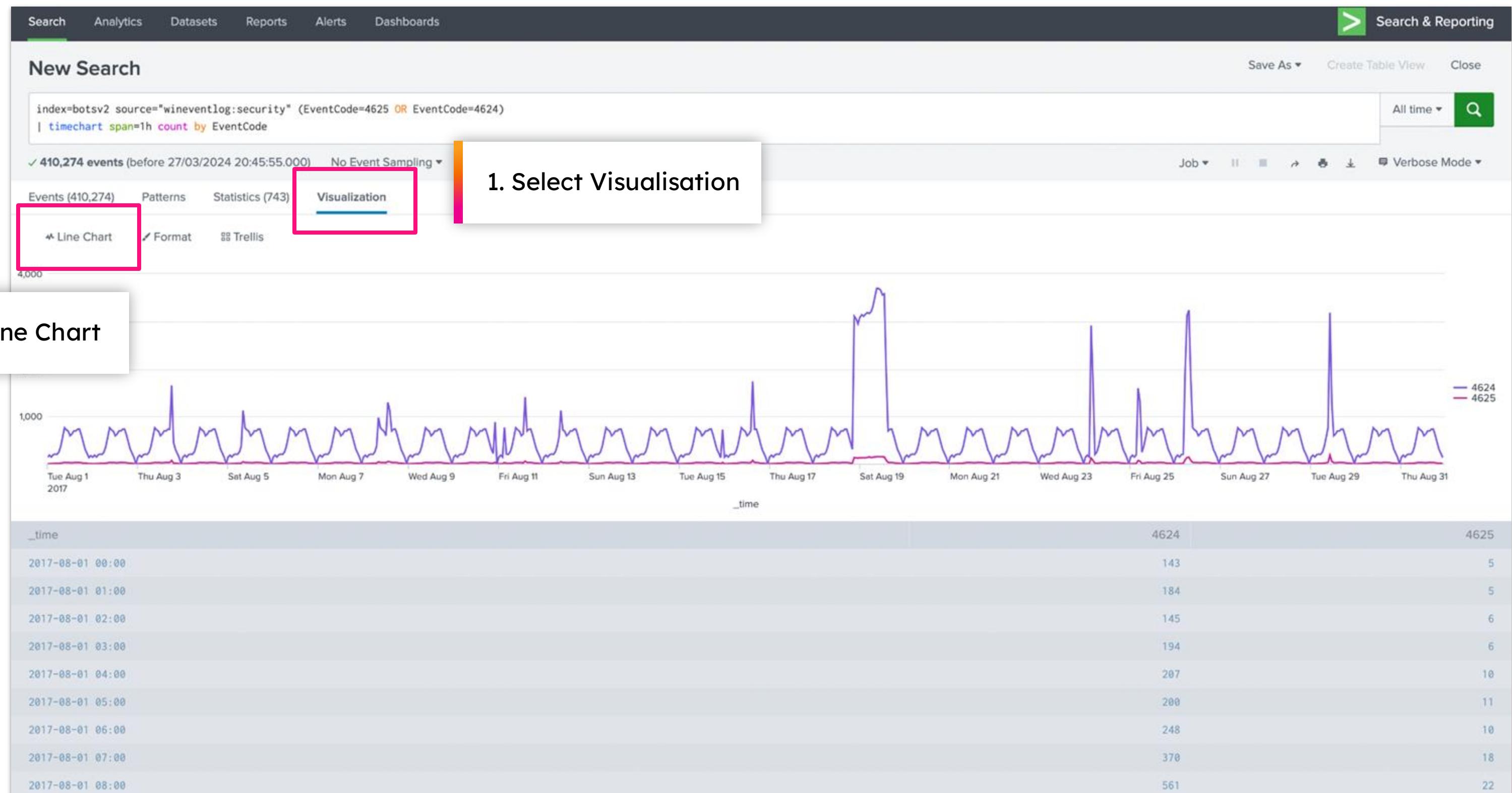
The search results table displays the following data:

_time	Count
2017-08-01 00:00	143
2017-08-01 01:00	184
2017-08-01 02:00	145
2017-08-01 03:00	194
2017-08-01 04:00	207
2017-08-01 05:00	200
2017-08-01 06:00	248
2017-08-01 07:00	370
2017-08-01 08:00	561
2017-08-01 09:00	762
2017-08-01 10:00	716

A red box highlights the "All time" dropdown in the top right corner of the search bar.

Authentication Events

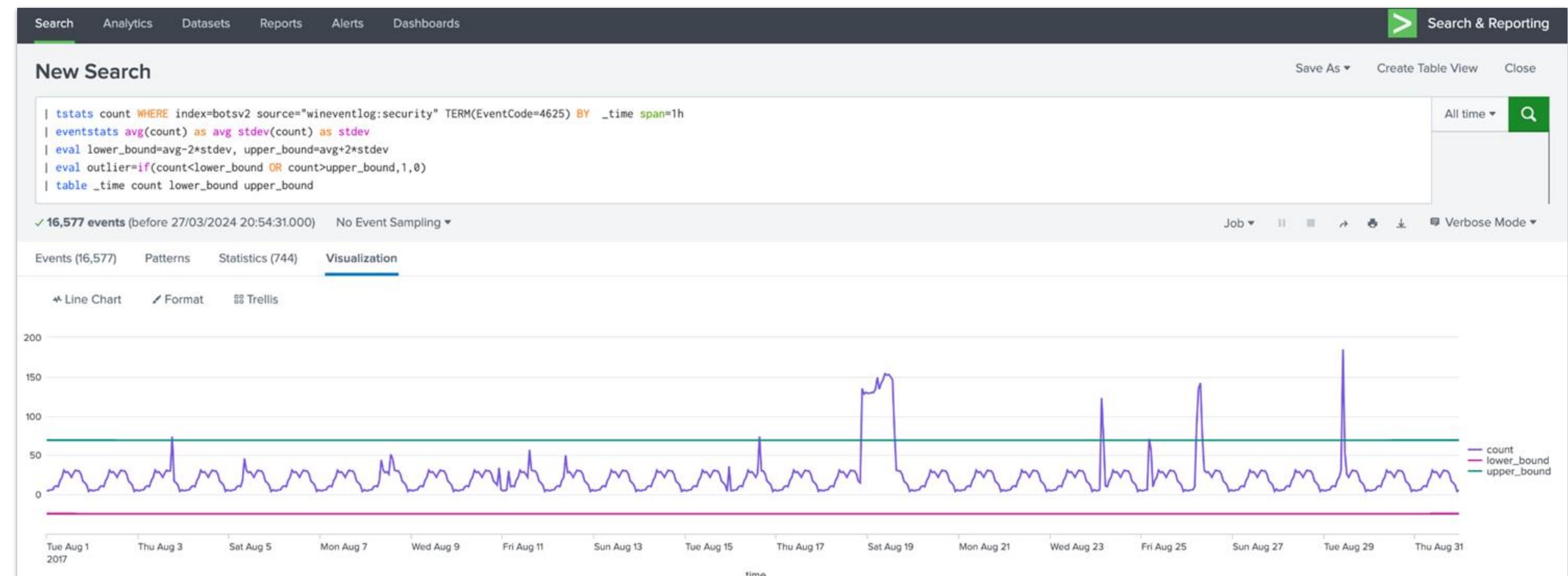
Search 1-2





Authentication Bounds

Search 1-3



```
tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h  
eventstats avg(count) as avg stdev(count) as stdev  
eval lower_bound=avg-2*stdev, upper_bound=avg+2*stdev  
eval outlier;if(count<lower_bound OR count>upper_bound,1,0)  
table _time count lower_bound upper_bound
```

Authentication Bounds

Search 1-3

1. Select Visualisation Type

2. Select Outliers Chart

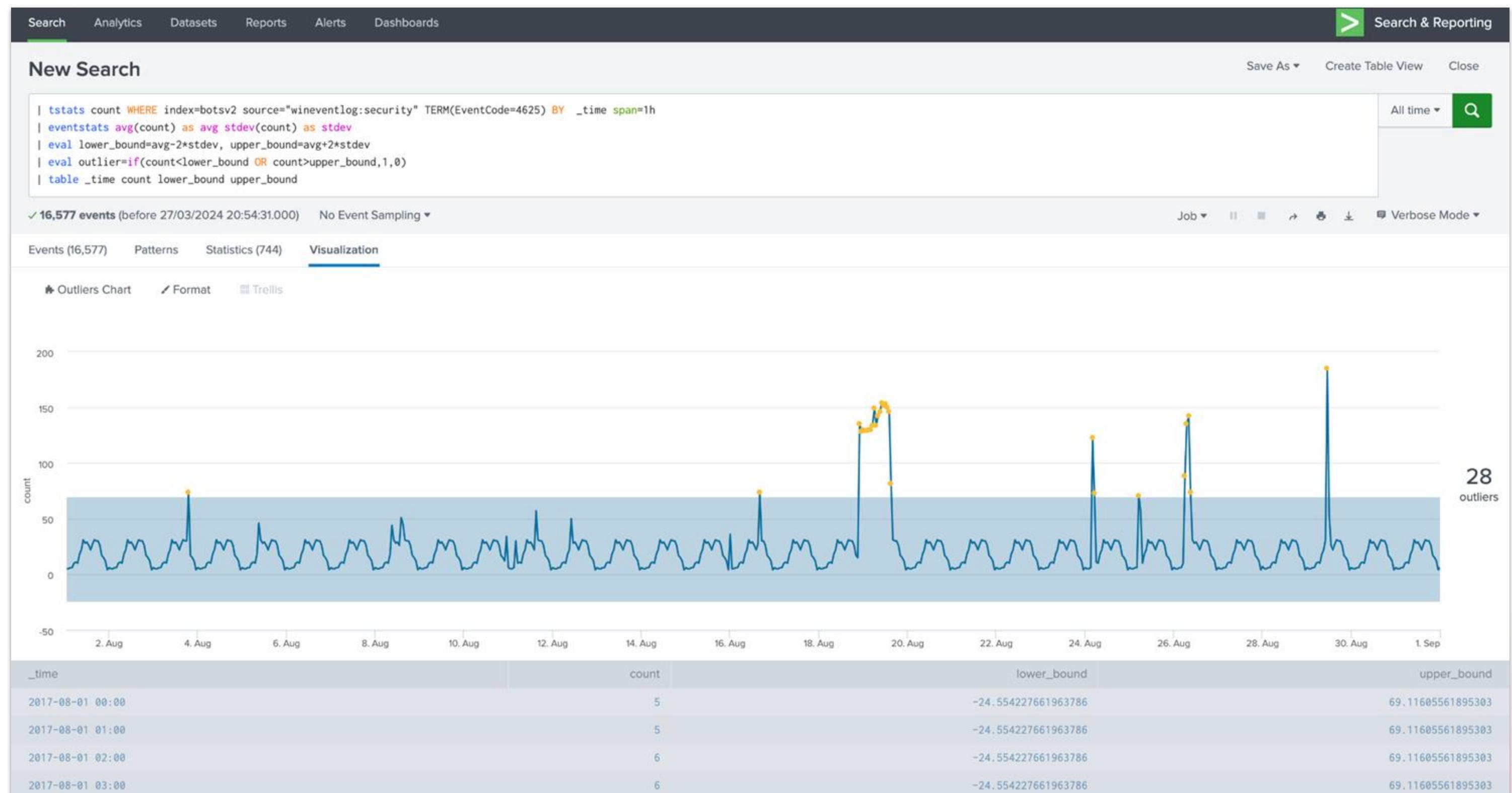
The screenshot shows the Splunk interface with a search bar containing a complex search command. Below the search bar, there are tabs for Events (16,577), Patterns, Statistics (744), and Visualization. The Visualization tab is selected, and a Line Chart icon is highlighted with a red box. To the right of the visualization tools, a line chart displays data over time from Wednesday, August 9, to Thursday, August 31. The chart shows a purple line representing the count of events, which fluctuates between approximately 5 and 15. Two horizontal green lines represent the upper_bound and lower_bound. A legend on the right identifies the purple line as 'count', the red line as 'lower_bound', and the green lines as 'upper_bound'. Below the chart is a table with columns for _time, count, lower_bound, and upper_bound. The table shows several rows of data. At the bottom left, a search fragment for an Outliers Chart is visible.

_time	count	lower_bound	upper_bound
2017-08-01 05:00	5	-24.554227661963786	69.11605561895303
2017-08-01 06:00	5	-24.554227661963786	69.11605561895303
	6	-24.554227661963786	69.11605561895303
	6	-24.554227661963786	69.11605561895303
	10	-24.554227661963786	69.11605561895303
	11	-24.554227661963786	69.11605561895303
	10	-24.554227661963786	69.11605561895303

Search Fragment:
| tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h
| eventstats avg(count) as avg stdev(count) as stdev
| eval lower_bound=avg-2*stdev, upper_bound=avg+2*stdev
| search _time >= 20:54:31.000
| search count > upper_bound,1,0
| search count < lower_bound
| eval upper_bound

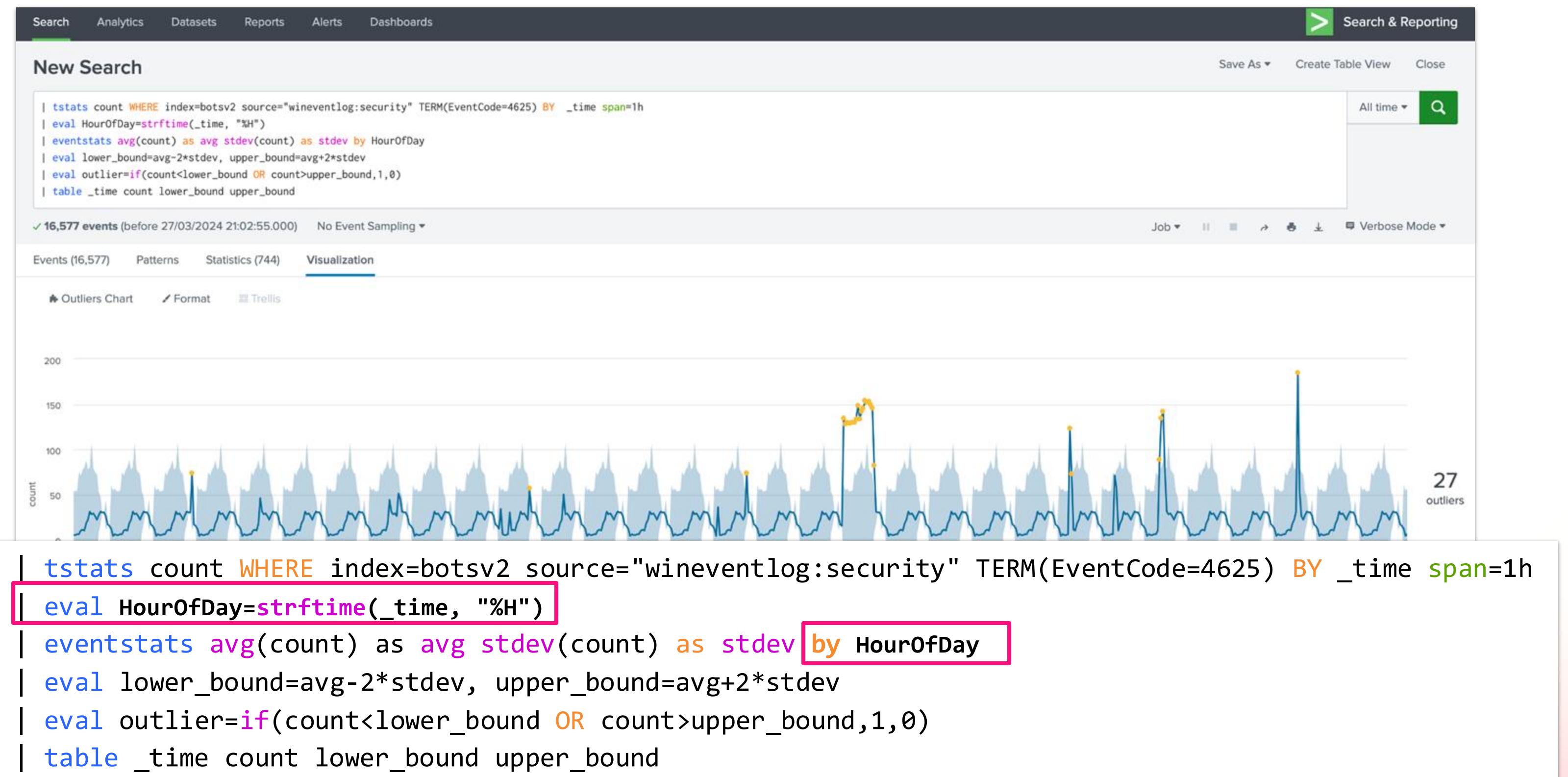
Authentication Outliers

Search 1-3



Authentication Outliers

Search 1-4



Fit Command

Train a model from search results

```
... | fit <ALGORITHM> <TARGET> from <VARIABLES ...>  
    <PARAMETERS> into <MODEL>
```

Examples:

```
... | fit LinearRegression system_temp from cpu_load fan_rpm  
    into temp_model
```

```
... | fit KMeans k=10 downloads purchases posts days_active visits_per_day  
    into user_behavior_clusters
```

```
... | fit LinearRegression petal_length from species
```

Authentication Fit

Search 1-5

SPLUNK APP FOR CYBERSECURITY

New Search

```
| tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h  
| eval HourOfDay=strftime(_time,"%H")  
| fit DensityFunction count by "HourOfDay" into botsv2_df_logon
```

Too few training points in some groups will likely result in poor model performance. Check the model summary to inspect such groups.

✓ 16,577 events (before 27/03/2024 21:08:54.000)

Events (16,577) Patterns Statistics (744) Visualization

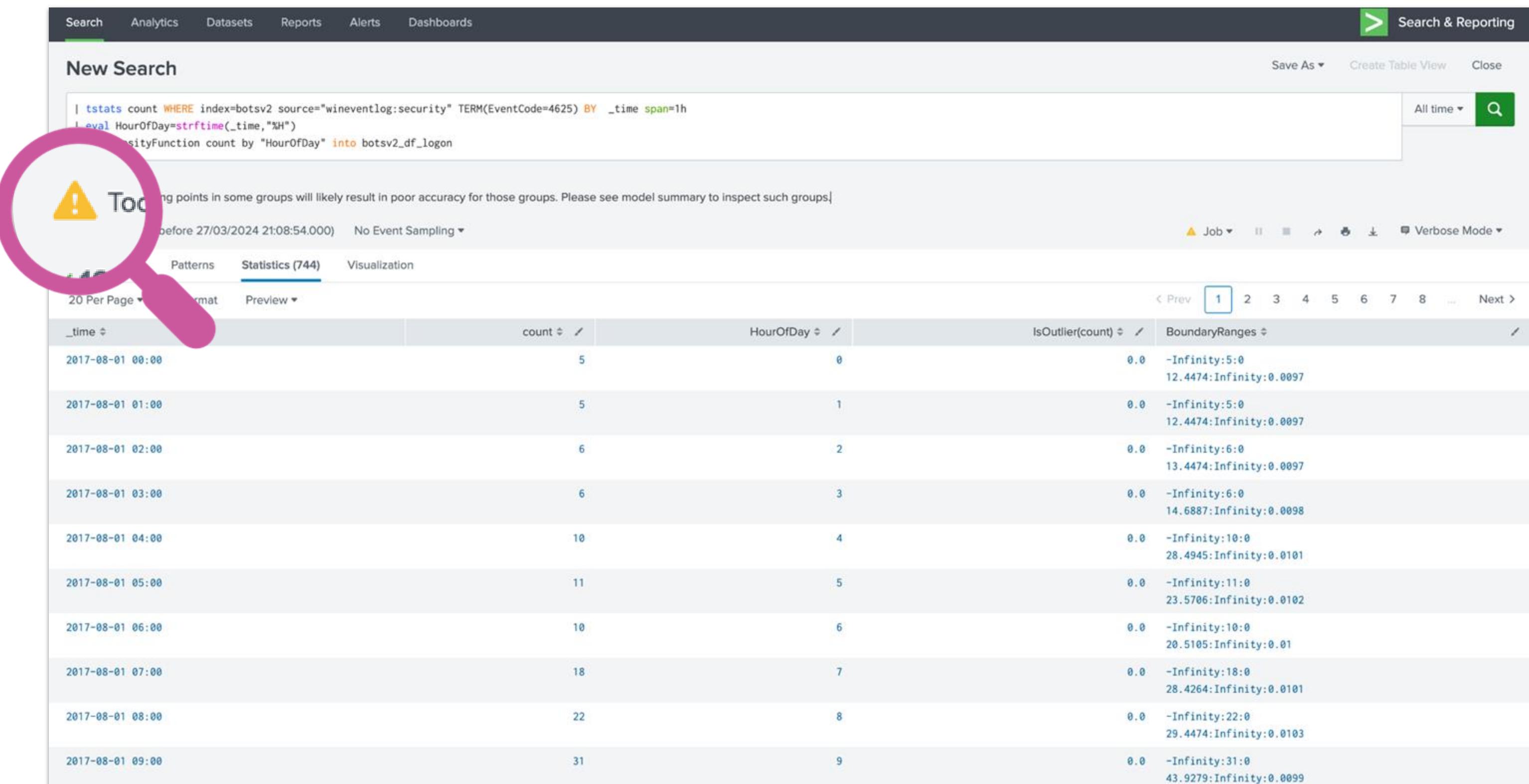
1. Select Statistics

_time	count	HourOfDay	IsOutlier(count)	BoundaryRanges
2017-08-01 00:00	5	0	0.0	-Infinity:5:0 12.4474:Infinity:0.0097
2017-08-01 01:00	5	1	0.0	-Infinity:5:0 12.4474:Infinity:0.0097
2017-08-01 02:00	6	2	0.0	-Infinity:6:0 13.4474:Infinity:0.0097
2017-08-01 03:00	6	3	0.0	-Infinity:6:0 14.6887:Infinity:0.0098
2017-08-01 04:00	10	4	0.0	-Infinity:10:0 28.4945:Infinity:0.0101
2017-08-01 05:00	11	5	0.0	-Infinity:11:0 23.5706:Infinity:0.0102
2017-08-01 06:00	10	6	0.0	-Infinity:10:0 23.5706:Infinity:0.0102

```
tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h  
eval HourOfDay=strftime(_time,"%H")  
fit DensityFunction count by "HourOfDay" into botsv2_df_logon
```

Authentication Fit

Search 1-5



A screenshot of a Splunk search interface titled "New Search". The search bar contains the following SPL command:

```
| tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h  
| eval HourOfDay=strftime(_time,"%H")  
| stats count by "HourOfDay" into botsv2_df_logon
```

The results table shows statistics for logon events by hour of the day on August 1, 2017. A yellow warning icon with the text "Too many points in some groups will likely result in poor accuracy for those groups. Please see model summary to inspect such groups." is overlaid on the top-left of the table area. A pink magnifying glass highlights this warning icon.

_time	count	HourOfDay	IsOutlier(count)	BoundaryRanges
2017-08-01 00:00	5	0	0.0	-Infinity:5:0 12.4474:Infinity:0.0097
2017-08-01 01:00	5	1	0.0	-Infinity:5:0 12.4474:Infinity:0.0097
2017-08-01 02:00	6	2	0.0	-Infinity:6:0 13.4474:Infinity:0.0097
2017-08-01 03:00	6	3	0.0	-Infinity:6:0 14.6887:Infinity:0.0098
2017-08-01 04:00	10	4	0.0	-Infinity:10:0 28.4945:Infinity:0.0101
2017-08-01 05:00	11	5	0.0	-Infinity:11:0 23.5706:Infinity:0.0102
2017-08-01 06:00	10	6	0.0	-Infinity:10:0 20.5105:Infinity:0.01
2017-08-01 07:00	18	7	0.0	-Infinity:18:0 28.4264:Infinity:0.0101
2017-08-01 08:00	22	8	0.0	-Infinity:22:0 29.4474:Infinity:0.0103
2017-08-01 09:00	31	9	0.0	-Infinity:31:0 43.9279:Infinity:0.0099

Authentication Apply

Search 1-6

Splunk Search & Reporting interface showing a search for bot logon events.

Search Bar:

```
| tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h  
| eval HourOfDay=strftime(_time,"%H")  
| apply botsv2_df_logon
```

Search Results Summary:

✓ 16,577 events (before 27/03/2024 21:15:43.000) No Event Sampling ▾ Job ▾ Verbose Mode ▾

Statistics View:

_time	count	HourOfDay	IsOutlier(count)	BoundaryRanges
2017-08-05 04:00	10	4	0.0	-Infinity:10:0 28.4945:Infinity:0.0101
2017-08-05 05:00	11	5	0.0	-Infinity:11:0 23.5706:Infinity:0.0102
2017-08-05 06:00	10	6	0.0	-Infinity:10:0 20.5105:Infinity:0.01
2017-08-05 07:00	18	7	0.0	-Infinity:18:0 28.4264:Infinity:0.0101
2017-08-05 08:00	46	8	1.0	-Infinity:22:0 29.4474:Infinity:0.0103
2017-08-05 09:00	31	9	0.0	-Infinity:31:0 43.9279:Infinity:0.0099
2017-08-05 10:00	28	10	0.0	-Infinity:28:0

A red box highlights the outlier row for 08:00 with a value of 1.0.

Search History:

```
tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h  
eval HourOfDay=strftime(_time,"%H")  
apply botsv2_df_logon
```

Authentication Apply

Search 1-6

1. Select Visualisation

New Search

```
| tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h  
| eval HourOfDay=strftime(_time,"%H")  
| apply botsv2_df_logon
```

Too few training points in some groups will likely result in poor accuracy for those groups. Please see model summary to inspect such groups.

✓ 16,577 events (before 27/03/2024 21:15:43.000) No Event Sampling ▾ Job ▾ II III ▾ Verbose Mode ▾

Events (16,577) Patterns Statistics (744) Visualization

Line Chart Format Trellis

2. Select Line Chart

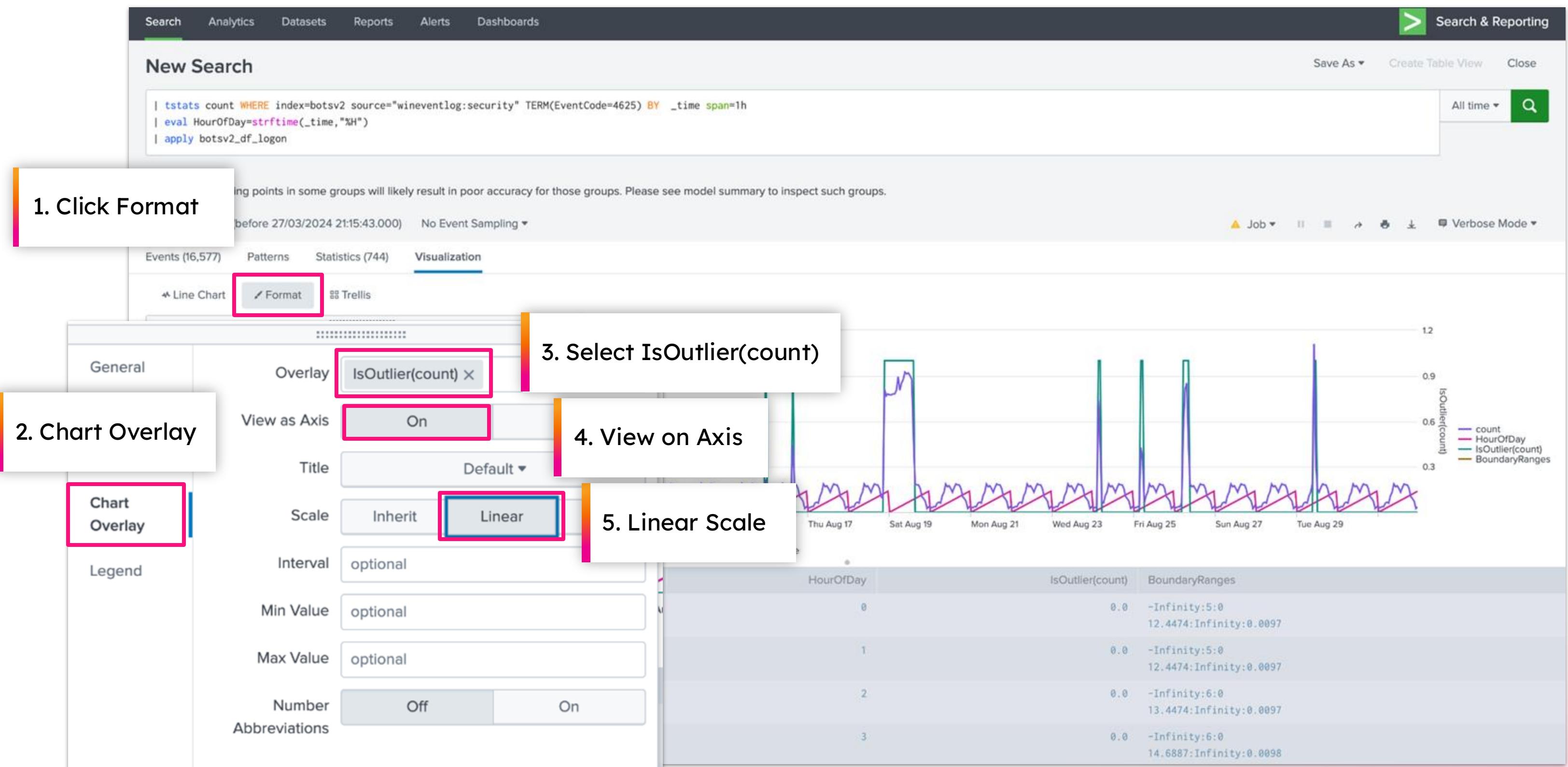
The chart displays the event count over time from August 1, 2017, to August 31, 2017. The Y-axis represents the count, ranging from 0 to 50. A red line shows a general upward trend with several sharp peaks, indicating periods of high activity. The X-axis shows dates from Tuesday, August 1, 2017, to Thursday, August 31, 2017. The chart includes a legend for 'count' (blue line), 'HourOfDay' (red line), 'IsOutlier(count)' (green line), and 'BoundaryRanges' (orange line).

_time	count	HourOfDay	IsOutlier(count)	BoundaryRanges
2017-08-01 00:00	5	0	0.0	-Infinity:5:0

```
tstats count WHERE index=botsv2 source="wineventlog:security" TERM(EventCode=4625) BY _time span=1h  
eval HourOfDay=strftime(_time,"%H")  
apply botsv2_df_logon
```

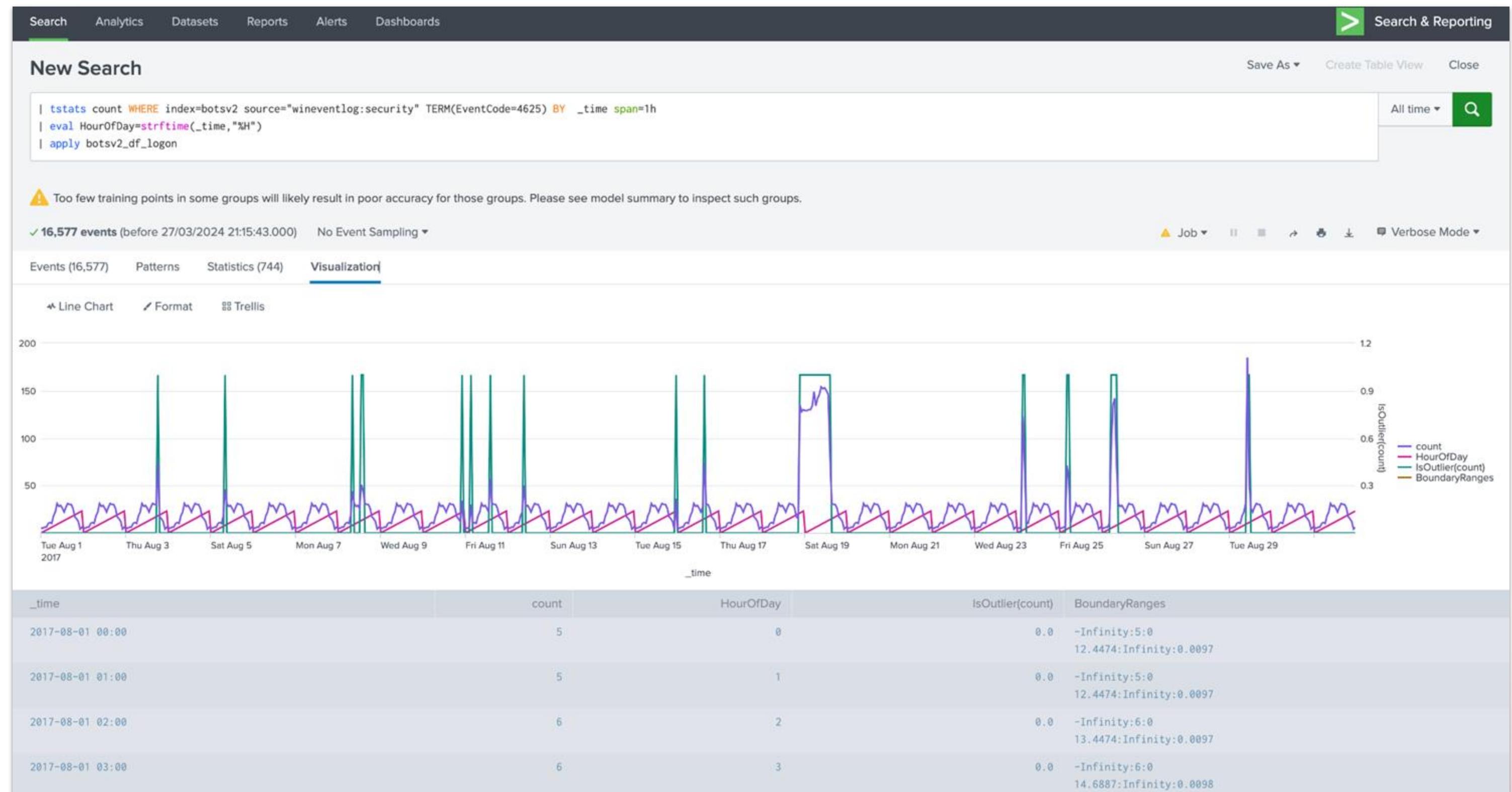
Authentication Apply

Search 1-6



Authentication Apply

Search 1-6



Search Breakdown

Bookmark

Splunk search	Description
index=botsv2 source="wineventlog:security" (EventCode=4625 OR EventCode=4624)	Search for windows event logs with failed log-ins from accounts (EventCode = 4625) and successful logins (EventCode = 4624).
... timechart span=1h count by EventCode	Time chart is statistical aggregation applied to a field to produce a chart, with time used as the X-axis and the specified field as the Y-axis. In this case, that field is calculated as the total count of event codes for each hour
... eventstats avg(count) as avg stdev(count) as stdev	Calculate the average and standard deviation across all events, in this case with EventCode 4625
... eval lower_bound=avg-2*stdev, upper_bound=avg+2*stdev	The average and standard deviation are used with the eval command to calculate the lower and upper boundaries. A sensitivity is added into the calculation by multiplying the stdev by 2, as in any distribution, about 95% of values should be within 2 standard deviations of the mean.
... eval outlier;if(count<lower_bound OR count>upper_bound,1,0)	The lower and upper bounds calculated earlier are used to identify the outliers
... table _time count lower_bound upper_bound	Returns a table with fields time, count, lower bound, and upper bound
... eval HourOfDay=strftime(_time,"%H")	Returns the hour from the _time field. strftime(_time,"%H") takes a UNIX time value and returns the time as a string in the specified format, in this case the hour.
... fit DensityFunction count by "HourOfDay" into botsv2_df_logon	<ul style="list-style-type: none">- Fit a density function model to the data- The into option lets the fit command know you want to save the model. Your model will now be saved as botsv2_df_logon <p><i>Note: Model names should not have spaces, and will be stored as a type of lookup</i></p>
... apply botsv2_df_logons	Apply the saved model * The apply command can only be used when MLTK & PSC are installed

Scenario 2

Detecting Anomalies in User Login
Quantity



Detecting Anomalies in User Login Quality

With the Splunk App for Anomaly Detection

You are an IT admin at Frothly, you are tasked with identifying and investigating any unusual user logins.

A significant variation in normal behavior could be due to a new system issue, errors in the IT infrastructure, or a security threat. T

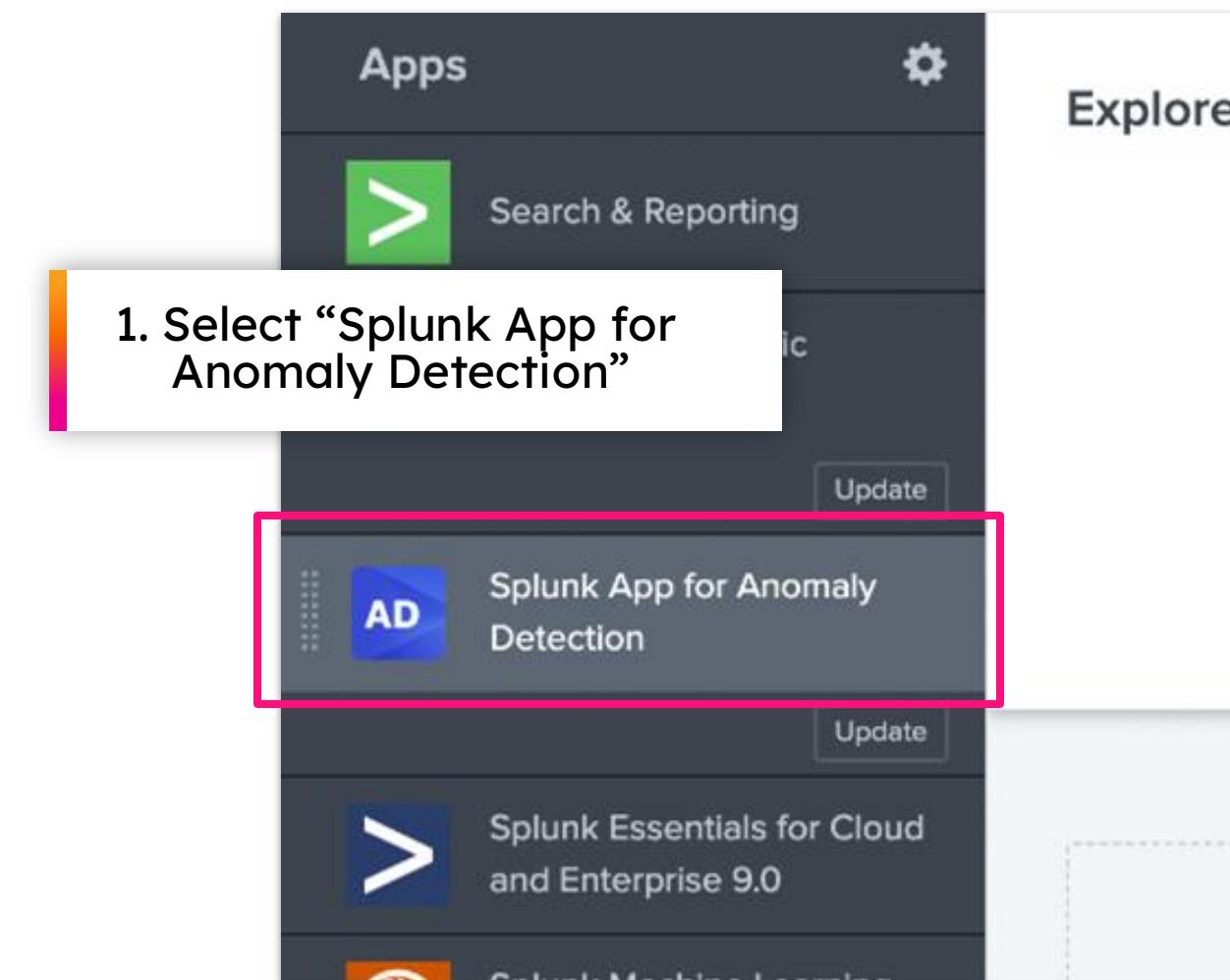
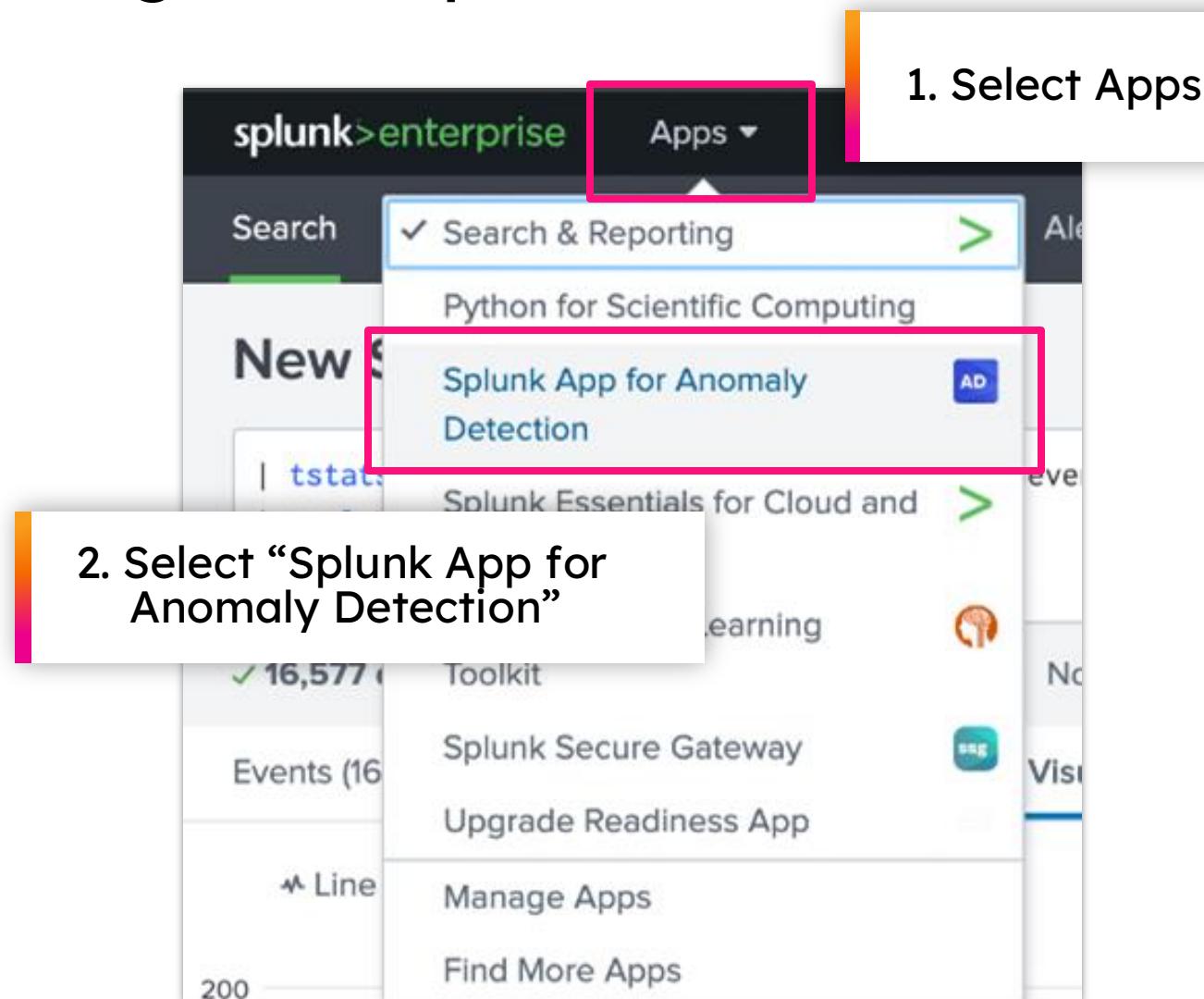
The previous IT admin happened to be proficient in SPL and familiar with foundational data science concepts. Unfortunately, you are less familiar with machine learning, but still want to use an ML workflow to complete your task.

Luckily for you, Splunk recently launched an assistive app that will take care of all of the heavy lifting of ML for you!



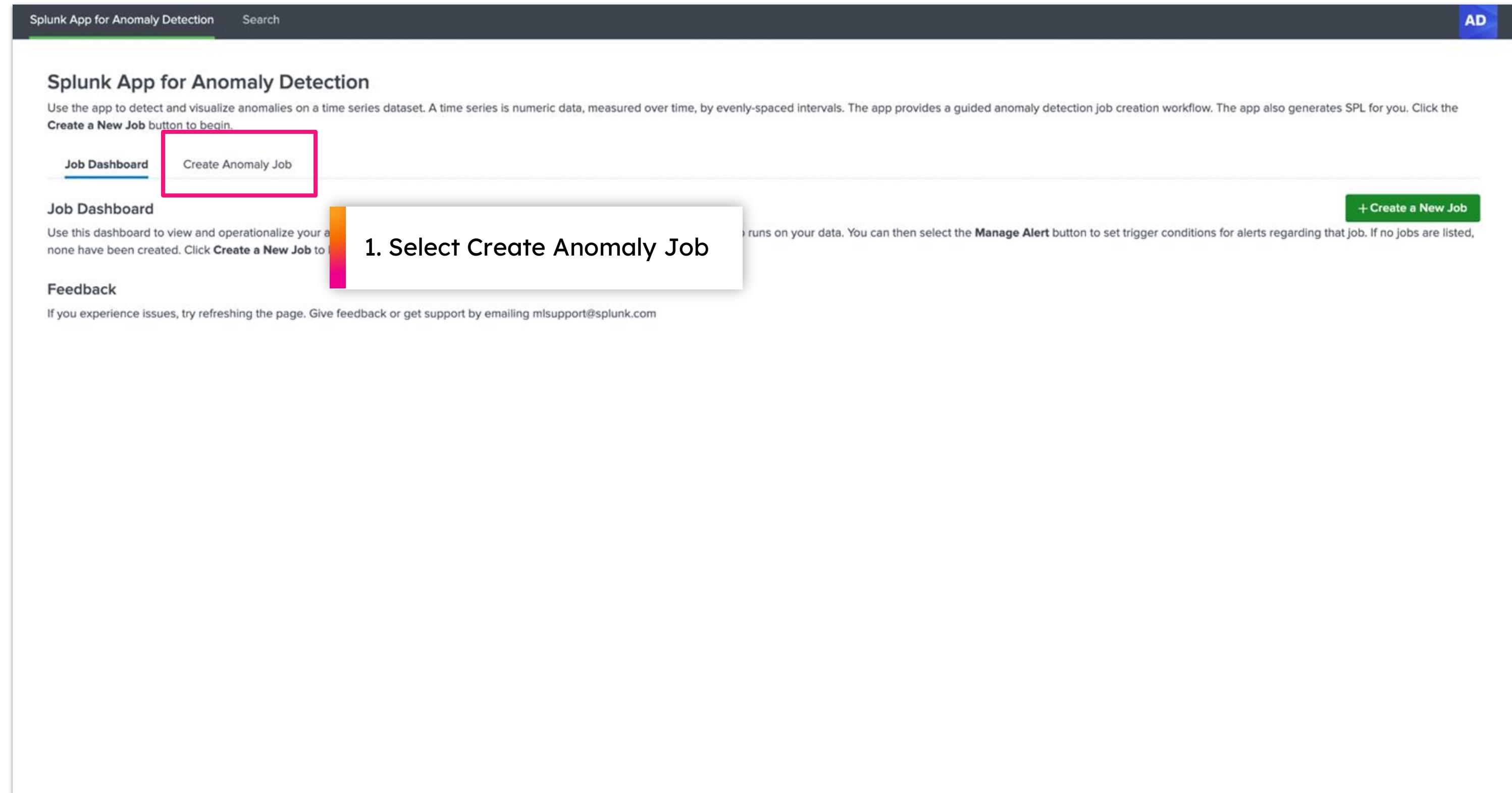
Search & Reporting

Navigation Options



Create Anomaly Job

Start Point



The screenshot shows the Splunk App for Anomaly Detection interface. At the top, there is a navigation bar with the text "Splunk App for Anomaly Detection" and "Search". On the far right of the bar is a blue button labeled "AD". Below the navigation bar, the main content area has a title "Splunk App for Anomaly Detection" and a brief description: "Use the app to detect and visualize anomalies on a time series dataset. A time series is numeric data, measured over time, by evenly-spaced intervals. The app provides a guided anomaly detection job creation workflow. The app also generates SPL for you. Click the Create a New Job button to begin." Below this text, there are two buttons: "Job Dashboard" and "Create Anomaly Job". The "Create Anomaly Job" button is highlighted with a red rectangular box. To the right of these buttons is a green button labeled "+ Create a New Job". Further down the page, there is a section titled "Job Dashboard" with a description: "Use this dashboard to view and operationalize your anomaly detection jobs. No jobs have been created. Click Create a New Job to begin." Below this section is another green button labeled "+ Create a New Job". At the bottom of the page, there is a "Feedback" section with the text: "If you experience issues, try refreshing the page. Give feedback or get support by emailing mlsupport@splunk.com". A large callout box with a black border and white background is overlaid on the page, containing the text "1. Select Create Anomaly Job".

Anomaly Job Details

Search 2-1

The screenshot shows the 'Create Anomaly Job' wizard in the Splunk App for Anomaly Detection. The interface includes a header with 'Splunk App for Anomaly Detection' and a search bar, and a top right corner labeled 'AD'. The main area displays the wizard steps:

- 1. Job Name:** **User_Logins** (highlighted with a red box)
- 2. Job Description:** **This job flags notable variation in typical login behaviour** (highlighted with a red box)
- 3. Click Next** (highlighted with a red box)

The 'Job Name' field contains 'User_Logins'. The 'Job Description' field contains 'This job proactively flags notable variation in typical login behaviour'. A 'Next' button is visible at the bottom of the input fields.

Anomaly Job Details

Default

The screenshot shows the 'Splunk App for Anomaly Detection' interface. At the top, there's a dark header bar with the app name and a search bar. A blue 'AD' button is in the top right corner. Below the header, the main content area has a title 'Splunk App for Anomaly Detection' and a brief description about detecting anomalies on time series datasets. It includes a 'Create a New Job' button. The navigation bar below the title has two tabs: 'Job Dashboard' and 'Create Anomaly Job', with 'Create Anomaly Job' being the active tab. The main form for creating a new job is titled 'Step 1: Name your Job'. It asks for a unique job name and an optional job description. The 'Job Name' field contains 'User_Logins'. The 'Job Description' field contains the placeholder text 'This job proactively flags notable variation in typical login behaviour'. A green 'Next' button is next to the description field. Below this, 'Step 2: Add the Dataset' is shown, which involves inputting an SPL query. The query editor shows '1 | inputlookup time_series.csv' and includes a dropdown for 'Last 24 hours' and a search icon. At the bottom, there's a 'Feedback' section with a note about reporting issues and an email address: mlsupport@splunk.com.

Splunk App for Anomaly Detection Search AD

Splunk App for Anomaly Detection

Use the app to detect and visualize anomalies on a time series dataset. A time series is numeric data, measured over time, by evenly-spaced intervals. The app provides a guided anomaly detection job creation workflow. The app also generates SPL for you. Click the [Create a New Job](#) button to begin.

Job Dashboard [Create Anomaly Job](#)

Step 1: Name your Job

Enter a unique job name and (optional) job description.

Job Name:

User_Logins

Job Description:

This job proactively flags notable variation in typical login behaviour

Next

Step 2: Add the Dataset

Use an SPL query to input your dataset. An example SPL query is provided that you can use to explore the app. Note that as the number of fields in your dataset increases, the app's response time will increase.

```
1 | inputlookup time_series.csv
```

Last 24 hours

Feedback

If you experience issues, try refreshing the page. Give feedback or get support by emailing mlsupport@splunk.com

Anomaly Job Details

Search 2-2

The screenshot shows the 'Splunk App for Anomaly Detection' interface. At the top, there's a navigation bar with 'Splunk App for Anomaly Detection' and 'Search' on the left, and 'AD' on the right. Below the navigation bar, the main title is 'Splunk App for Anomaly Detection'. A sub-instruction says: 'Use the app to detect and visualize anomalies on a time series dataset. A time series is numeric data, measured over time, by evenly-spaced intervals. The app provides a guided anomaly detection job creation workflow. The app also generates SPL for you. Click the Create a New Job button to begin.' There are two tabs at the top: 'Job Dashboard' and 'Create Anomaly Job', with 'Create Anomaly Job' being the active tab. The first step, 'Step 1: Name your Job', is shown with a placeholder 'User_Logins' in the 'Job Name:' field. The second step, 'Step 2: Add the Dataset', shows an SPL query: '1. | inputlookup numenta_art_daily_flatmiddle.csv'. A green search button is highlighted with a red box. The feedback section at the bottom says: 'If you experience issues, try refreshing the page. Give feedback or get support by emailing mlsupport@splunk.com'.

1. | inputlookup numenta_art_daily_flatmiddle.csv

2. Click the Search button

Anomaly Job Details

Search Complete

Job Description:
This job proactively flags notable variation in typical login behaviour

Next

Step 2: Add the Dataset
Use an SPL query to input your dataset. An example SPL query is provided that you can use to explore the app. Note that as the number of fields in your dataset increases, the app's response time will increase.

```
1 | inputlookup numenta_art_daily_flatmiddle.csv
```

Last 24 hours

Step 3: Select Field for Anomaly Detection
Select a field from your dataset for anomaly detection. Only numeric fields are listed in the drop-down menu.

Field For Detection: - Select -

Detection sensitivity: Low Med High Detect Anomalies

Search successful.

Preview Data **Anomaly Data**

i	_time	value
>	2014-04-01 00:00:00	4.351617317699997
>	2014-04-01 00:05:00	5.104523132399997
>	2014-04-01 00:10:00	7.272770531700001
>	2014-04-01 00:15:00	5.228334600299998
>	2014-04-01 00:20:00	4.176238387999998
>	2014-04-01 00:25:00	6.295508866599999
>	2014-04-01 00:30:00	3.810553504099996
>	2014-04-01 00:35:00	6.454958372499998
>	2014-04-01 00:40:00	7.094107469400001
>	2014-04-01 00:45:00	6.760242762199997

< Prev 1 2 3 4 5 ... 404 Next >

Feedback
If you experience issues, try refreshing the page. Give feedback or get support by emailing msupport@splunk.com

Anomaly Job

Detect

Job Description:
This job proactively flags notable variation in typical login behaviour

Next

Step 2: Add the Dataset
Use an SPL query to input your dataset. An example SPL query is provided that you can use to explore the app. Note that as the number of fields in your dataset increases, the app's response time will increase.

```
1 | inputlookup numenata_art_daily_flatmiddle.csv
```

Last 24 hours

Step 3: Select Field
Select a field from your dataset. A field is selected in the drop-down menu.

Field For Detection
 - Select - value is ready for anomaly detection.

Detection sensitivity
Low Med High

1. Click the Dropdown

2. Select “Value”

3. Click “Detect Anomalies”

value

50

Tue Apr 1 2014 Wed Apr 2 Thu Apr 3 Fri Apr 4 Sat Apr 5 Sun Apr 6 Mon Apr 7 Tue Apr 8 Wed Apr 9 Thu Apr 10 Fri Apr 11 Sat Apr 12 Sun Apr 13 Mon Apr 14

_time

i _time value

_time	value
> 2014-04-01 00:00:00	4.351617317699997
> 2014-04-01 00:05:00	5.104523132399997
> 2014-04-01 00:10:00	7.272770531700001
> 2014-04-01 00:15:00	5.228334600299998
> 2014-04-01 00:20:00	4.176238387999998

Anomaly Job Configuration

Job Description:
This job proactively flags notable variation in typical login behaviour

Next

Step 2: Add the Dataset
Use an SPL query to input your dataset. An example SPL query is provided that you can use to explore the app. Note that as the number of fields in your dataset increases, the app's response time will increase.

```
1 | inputlookup numenata_art_daily_flatmiddle.csv
```

Last 24 hours

Step 3: Select Field for Anomaly Detection
Select a field from your dataset for anomaly detection. Only numeric fields are listed in the drop-down menu.

Field For Detection: value

Detection sensitivity: Low Med High

Anomaly detection complete.

Anomaly Data

Preview Data

i	_time	value	isOutlier	anomConf
>	2014-04-11 04:35:00.000000	65.4	1	1.0
>	2014-04-11 21:55:00.000000	65.4	1	1.0

Step 4: Save & Operationalize Job
Save this anomaly detection job. From the Job Dashboard, schedule when the job is run. Once scheduled, you can create job-related alerts.

Click **Open in Search** to open a new Splunk search using this SPL query. You can modify the SPL as needed. Click **View SPL** to view and copy the SPL for use anywhere in Splunk. Note that the query updates a model every time it runs; to reproduce previous results, remove "partial_fit=true" from the query before running it again.

Anomaly Job

Save Job

Use an SPL query to input your dataset. An example SPL query is provided that you can use to explore the app. Note that as the number of fields in your dataset increases, the app's response time will increase.

```
1 | inputlookup numenata_art_daily_flatmiddle.csv
```

Last 24 hours

Step 3: Select Field for Anomaly Detection
Select a field from your dataset for anomaly detection. Only numeric fields are listed in the drop-down menu.

Field For Detection: Detection sensitivity:

Anomaly detection complete.

Preview Data Anomaly Data

value

Anomalies: 2

Anomaly
Thresholds
value

Step 4: Save & Operationalize Job
Save this anomaly detection job. From the Job Dashboard, schedule when the job is run. Once scheduled, you can create job-related alerts.

Feedback
If you experience issues, try refreshing the page. Give feedback or get support by emailing msupport@splunk.com

Click **Open in Search** to open a new Splunk search using this SPL query. You can modify the SPL as needed. Click **View SPL** to view and copy the SPL for use anywhere in Splunk. Note that the query updates a model every time it runs; to reproduce previous results, remove "partial_fit=true" from the query before running it again.

Anomaly Job

Schedule

The screenshot shows the Splunk App for Anomaly Detection interface. A modal window titled "Schedule Job" is open over the main dashboard.

1. Click “Job Dashboard”: A callout points to the "Job Dashboard" button in the top left corner of the main content area.

2. Click “edit”, then set an Schedule: A callout points to the "Edit" link in the "Schedule" column of the table, which then highlights the "Schedule Job" dialog.

Main Content Area (Job Dashboard):

- Header:** Splunk App for Anomaly Detection, Search, AD
- Title:** Splunk App for Anomaly Detection
- Text:** Use the app to detect and visualize anomalies in your data, measured over time, by evenly-spaced intervals. The app provides a guided anomaly detection job creation workflow. The app also generates SPL for you. Click the Create a New Job button to begin.
- Buttons:** Job Dashboard (highlighted with a red box), Create A...
- Section:** Job Dashboard
- Description:** Use this dashboard to view and operationalize your anomaly detection jobs. Select the Schedule button to determine when this job runs on your data. No jobs have been created. Click Create a New Job to begin.
- Table:** A table listing a single job entry: User_Logins. The columns are Name, Description, Schedule, Alert Trigger Condition, Status, and Actions.
- Feedback:** If you experience issues, try refreshing the page. Give feedback or get support by emailing mlsupport@splunk.com.

Schedule Job Dialog:

- Title:** Schedule Job
- Run:** every hour ▾ At Select... minutes past the hour.
- Buttons:** Save, X

Anomaly Job

Trigger

The screenshot shows the Splunk App for Anomaly Detection interface. At the top, there's a navigation bar with 'Splunk App for Anomaly Detection' and 'Search' on the left, and 'AD' on the right.

Job Dashboard: A callout box labeled '1. Click "Job Dashboard"' points to the 'Job Dashboard' button in the top-left corner of the main content area. This button is highlighted with a red border.

Alert Trigger Condition: A callout box labeled '2. Click "edit", to set a Trigger Condition' points to the 'Edit' link under the 'Alert Trigger Condition' column for the 'User_Logins' job. This link is also highlighted with a red border.

Table Data: Below the callouts, a table lists a single job entry:

Name	Description	Schedule	Alert Trigger Condition	Status	Actions
User_Logins	This job proactively flags notable variation in typical login behaviour	None Edit	None Edit	On Off	⋮

Feedback: At the bottom left, there's a 'Feedback' section with a note: 'If you experience issues, try refreshing the page. Give feedback or get support by emailing mlsupport@splunk.com'.

Anomaly Job

Trigger

1. Enter the Trigger Condition Details

Alert Trigger Condition

Job Name: User_Logins

Trigger Alert When

Number of anomalies

is at least 3

Confidence score

is at least 0.85

Expires after

hour(s) 24

+ Add Actions ▾

Output results to lookup Remove

Results:

Append Replace

File name:

User_Login_Anomalies.csv

Alert on Missing Data

Turn on to get alerted when a chosen number of consecutive data point values are missing.

Save

2. Click “Save”

Anomaly Job

Complete

Splunk App for Anomaly Detection Search AD

Splunk App for Anomaly Detection

Use the app to detect and visualize anomalies on a time series dataset. A time series is numeric data, measured over time, by evenly-spaced intervals. The app provides a guided anomaly detection job creation workflow. The app also generates SPL for you. Click the [Create a New Job](#) button to begin.

[Job Dashboard](#) [Create Anomaly Job](#)

Job Dashboard [+ Create a New Job](#)

Use this dashboard to view and operationalize your anomaly detection jobs. Select the **Schedule** button to determine when this job runs on your data. You can then select the **Manage Alert** button to set trigger conditions for alerts regarding that job. If no jobs are listed, none have been created. Click [Create a New Job](#) to begin.

i	Name	Description	Schedule	Alert Trigger Condition	Status	Actions
>	User_Logins	This job proactively flags notable variation in typical login behaviour	Every day at *:00 Edit	Anomalies >= 3, Confidence score >= 0.85 Edit	On Off	⋮

Feedback
If you experience issues, try refreshing the page. Give feedback or get support by emailing mlsupport@splunk.com

Scenario 3

Forecasting CPU Utilisation



Forecasting CPU Utilisation

You are a IT Operations Analyst at Frothly, you have been asked to investigate if we could predict when CPU utilisation is going to spike again to help the team proactively resolve potential incidents before they occur.

Some reports have come into Buttercup Games that users are reporting crashes, with the game seizing up before exiting unexpectedly. Root cause analysis of these incidents has identified that CPU utilisation on a few servers was running pretty high during these crashes.



CPU Counters

Search 3-1

Splunk Search & Reporting interface showing search results for CPU counters.

Search Bar: index=botsv2 sourcetype=perfmon:cpu counter="% idle time"

Results Summary: 204,356 events (before 27/03/2024 23:59:59)

Event Timeline: 1 day per column

Event Fields:

Time	Event
31/08/2017 22:59:44.000	08/31/2017 15:59:44.841 -0700 collection=CPU object=Processor counter=% Idle Time instance=_Total host = wrk-aturing source = Perfmon:CPU sourcetype = Perfmon:CPU
31/08/2017 22:59:14.000	08/31/2017 15:59:14.966 -0700 collection=CPU object=Processor counter=% Idle Time instance=0 host = wrk-abungst source = Perfmon:CPU sourcetype = Perfmon:CPU
31/08/2017 22:58:56.000	08/31/2017 15:58:56.706 -0700 collection=CPU object=Processor counter=% Idle Time instance=0 host = wrk-klagerf source = Perfmon:CPU sourcetype = Perfmon:CPU
31/08/2017 22:58:55.000	08/31/2017 15:58:55.755 -0700 collection=CPU

Selected Fields: host 9, source 1, sourcetype 1

Interesting Fields: collection 1, counter 1, cpu_instance 3, dest 9, eventtype 3, index 1, instance 3, linecount 1, metric_type 1, object 1, punct 2, splunk_server 1, src 9, tag 6, tag:eventtype 6, Value 100+

Actions: Save As, Create Table View, Close, Verbose Mode

Time Series CPU Counters

Search 3-2

SPLUNK UI SCREENSHOT

New Search

```
index=botsv2 sourcetype=perfmon:cpu counter="% idle time"
| table _time host Value
| timechart span=30m min(value) as idle_cpu by host
```

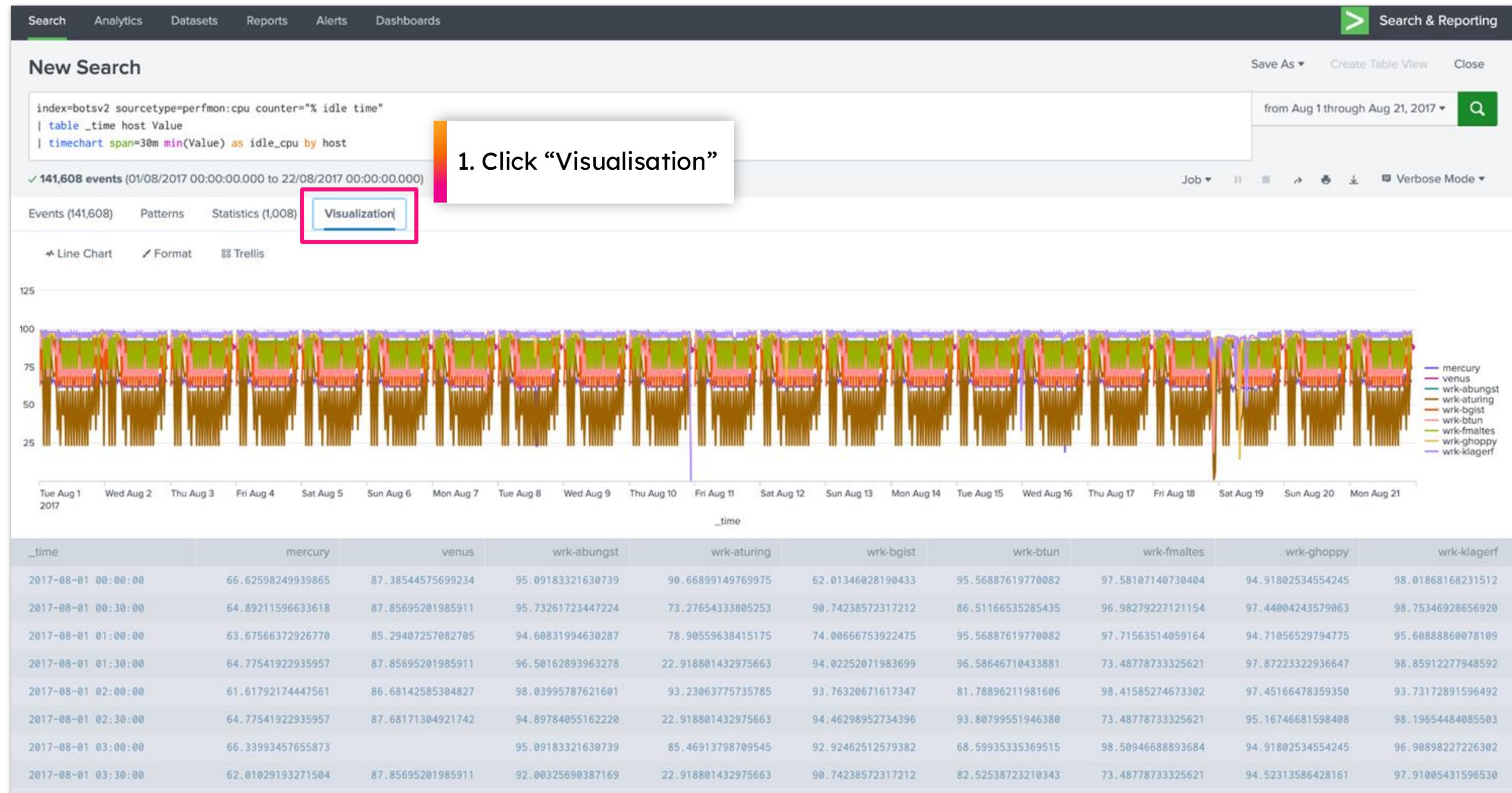
Date Range: Between 01/08/2017 - 28/08/2017

_time	host	Value	idle_cpu	bgist	wrk-btun
2017-08-01 00:00:00	66.62598249939865	87.38544575699234	95.09183321630739	90.66899149769975	62.01346028190433
2017-08-01 00:30:00	64.89211596633618	87.85695201985911	95.73261723447224	73.27654333805253	90.74238572317212
2017-08-01 01:00:00	63.67566372926770	85.29407257082705	94.60831994630287	78.90559638415175	74.00666753922475
2017-08-01 01:30:00	64.77541922935957	87.85695201985911	96.50162893963278	22.918801432975663	94.02252071983699
2017-08-01 02:00:00	61.61792174447561	86.68142585304827	98.03995787621601	93.23063775735785	93.76320671617347
2017-08-01 02:30:00	64.77541922935957	87.68171304921742	94.89784055162220	22.918801432975663	94.46298952734396
2017-08-01 03:00:00	66.33993457655873		95.09183321630739	85.46913798709545	92.92462512579382
2017-08-01 03:30:00	62.01029193271504	87.85695201985911	92.00325690387169	22.918801432975663	90.74238572317212
2017-08-01 04:00:00	68.90353055918450	94.137707479465	96.00656209752432	87.14808974620351	93.76320671617347
2017-08-01 04:30:00	64.67603337947514	87.68171304921742	92.00325690387169	76.17713737209358	93.33165719685944
2017-08-01 05:00:00	66.62383703540334	94.06828480319253	88.32985599008450	67.90685390654058	90.32877115826946
2017-08-01 05:30:00	64.77541922935957	87.68171304921742	94.89784055162220	70.53790238450209	93.52862304395175
2017-08-01 06:00:00	61.61792174447561	85.29407257082705	88.32985599008450	68.59134681113247	62.01346028190433
2017-08-01 06:30:00	64.67603337947514	87.85695201985911	95.78280147572747	33.24575218443201	93.95596266289385
2017-08-01 07:00:00	61.61792174447561	85.29407257082705	88.32985599008450	59.20357723277010	62.01346028190433
2017-08-01 07:30:00	64.77541922935957	87.68171304921742	94.89784055162220	22.918801432975663	93.52862304395175
2017-08-01 08:00:00	61.61792174447561	85.29407257082705	77.69367107754607	59.20357723277010	62.01346028190433
2017-08-01 08:30:00	62.01029193271504	87.68171304921742	92.00325690387169	43.00096115873717	90.74238572317212

Presets
Relative
Real-time
Date Range
Between 01/08/2017 and 21/08/2017
00:00:00 24:00:00
Apply
Date & Time Range
Advanced

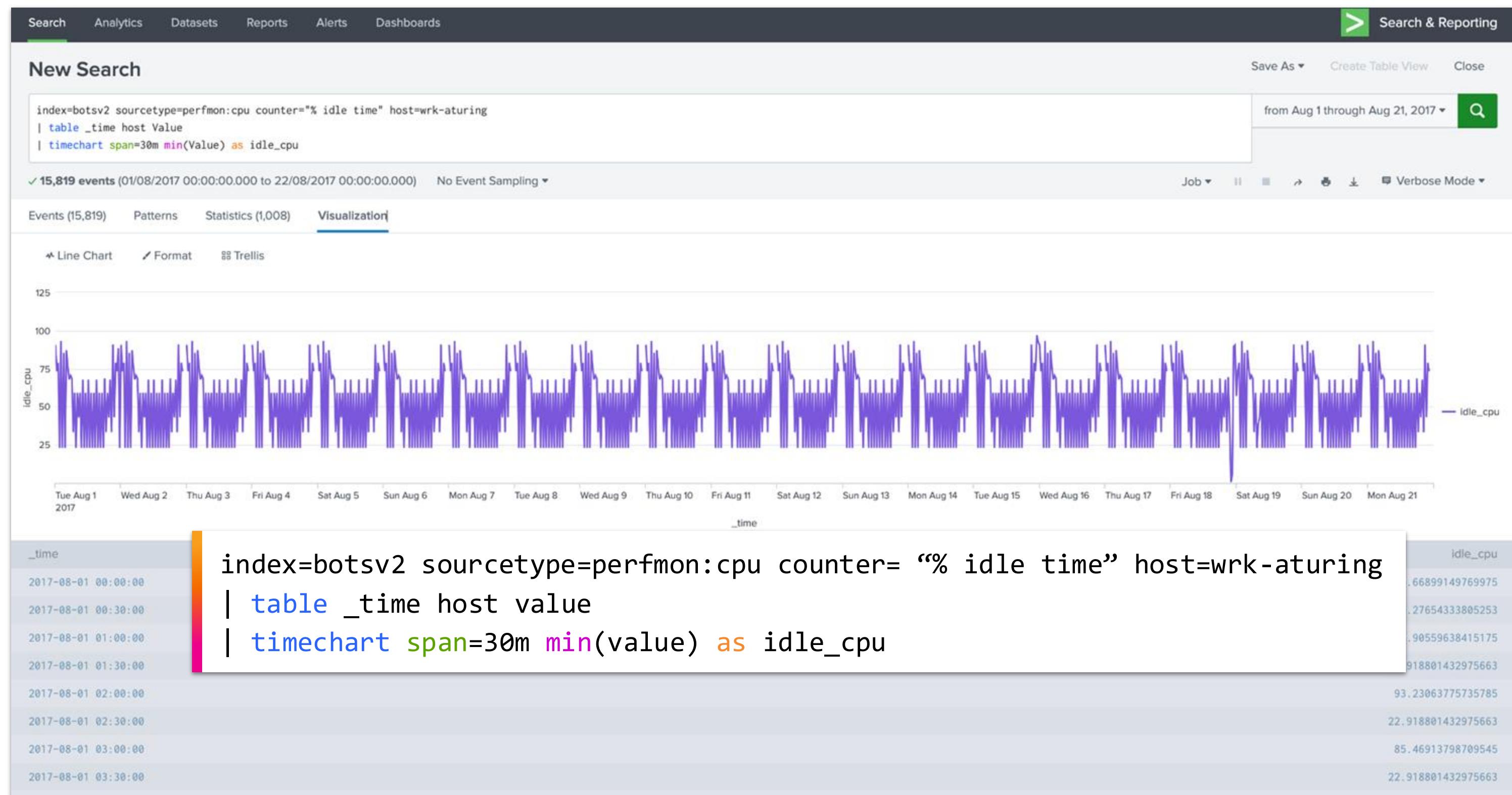
Time Series CPU Counters

Visualisation



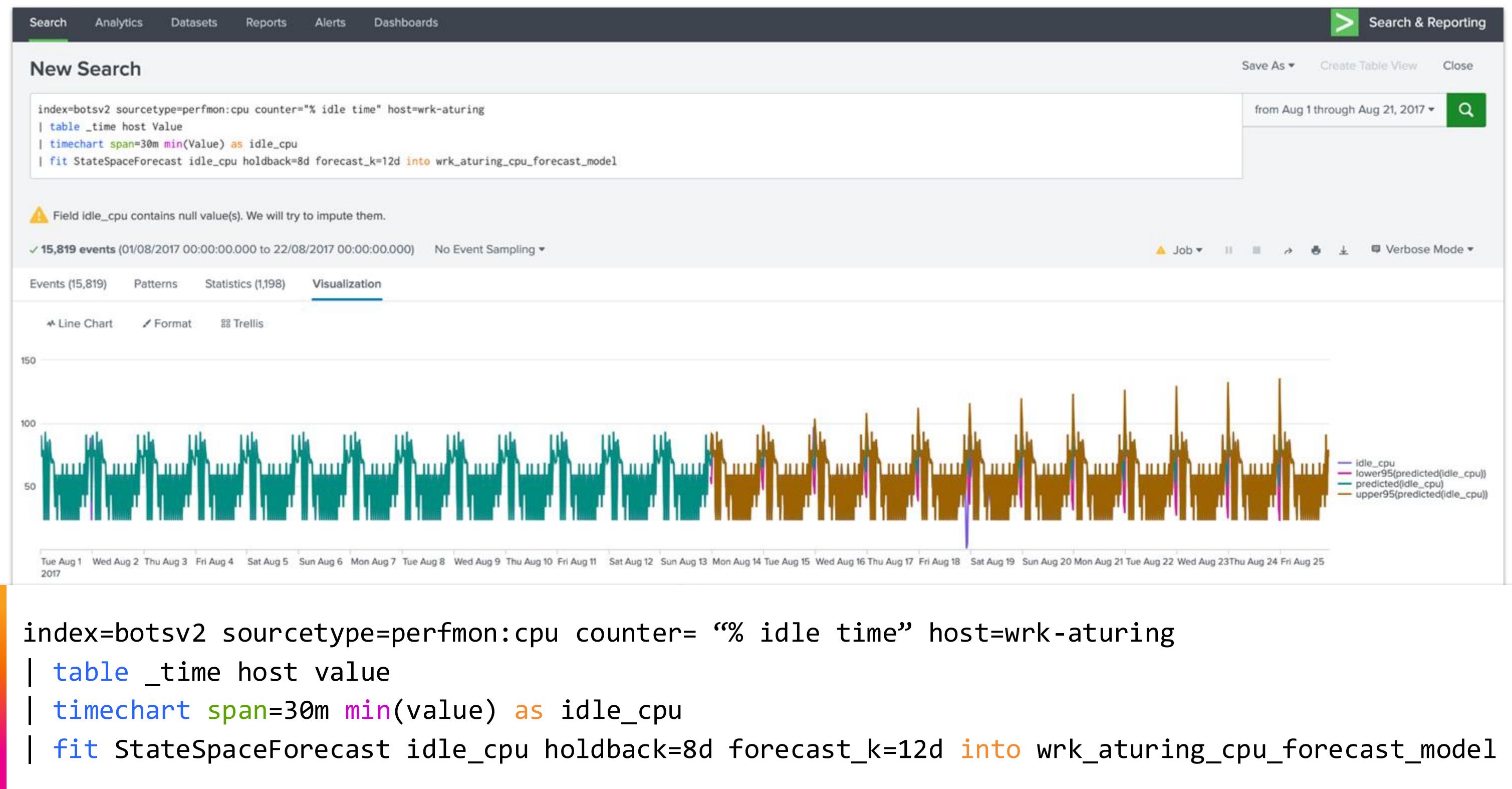
Time Series CPU Counters

Search 3-3



Fit Model on CPU Counters

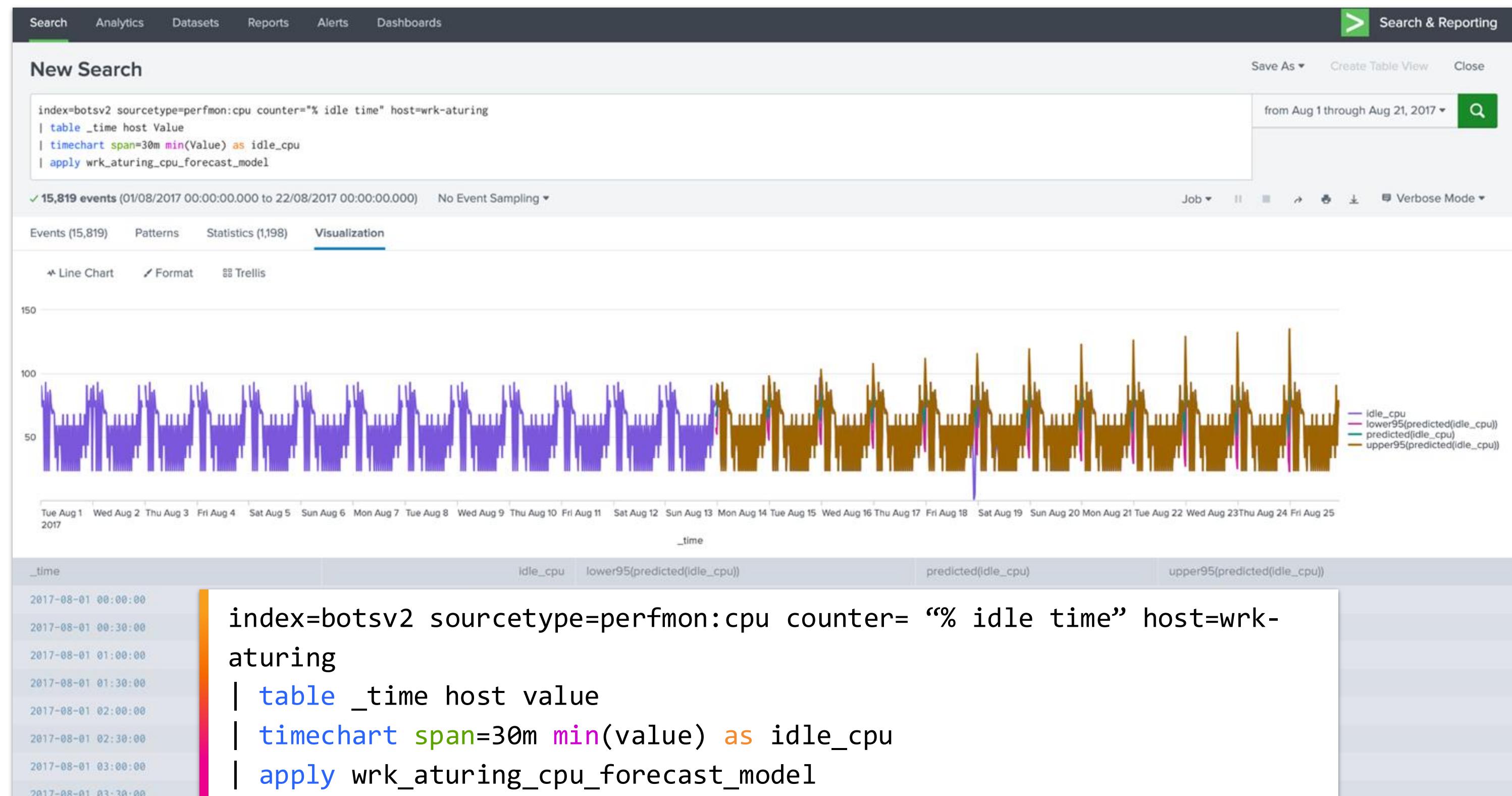
Search 3-4



```
index=botsv2 sourcetype=perfmon:cpu counter= "% idle time" host=wrk-aturing
| table _time host value
| timechart span=30m min(value) as idle_cpu
| fit StateSpaceForecast idle_cpu holdback=8d forecast_k=12d into wrk_aturing_cpu_forecast_model
```

Apply Model on CPU Counters

Search 3-5



Search Breakdown

Bookmark

Splunk search	Description
<code>index=botsv2 sourcetype=perfmon:cpu counter="% idle time"</code>	Search the perfmon:cpu logs in the botsv2 index to see CPU idle time
<code>... table _time host Value</code>	Create a table from our raw results containing the time, host and CPU idle value
<code>... timechart span=30m min(Value) as idle_cpu by host</code>	Calculate for each host the minimum CPU idle time over each 30 minute window
<code>... fit StateSpaceForecast idle_cpu holdback=8d forecast_k=12d into wrk_aturing_cpu_forecast_model</code>	Train a StateSpaceForecast model to create a forecast on the CPU idle data. The model will forecast 12 days into the future, and is trained on all of the input data except for the final 8 days of information.
<code>... apply wrk_aturing_cpu_forecast_model</code>	Apply the saved model named wrk_aturing_cpu_forecast_model on the data

Scenario 4

Detecting Outliers in Network
Utilisation Telemetry

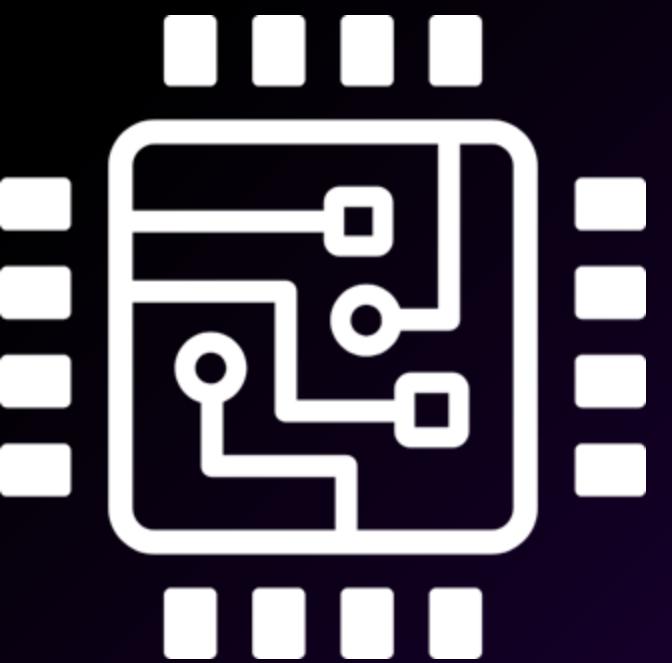


Detecting Outliers in Network Utilisation Telemetry

You are an App Developer at Frothly. Despite resolving the CPU utilisation issues, you have been getting some customer complaints that the app periodically stops responding to their inputs, making it irritating to use.

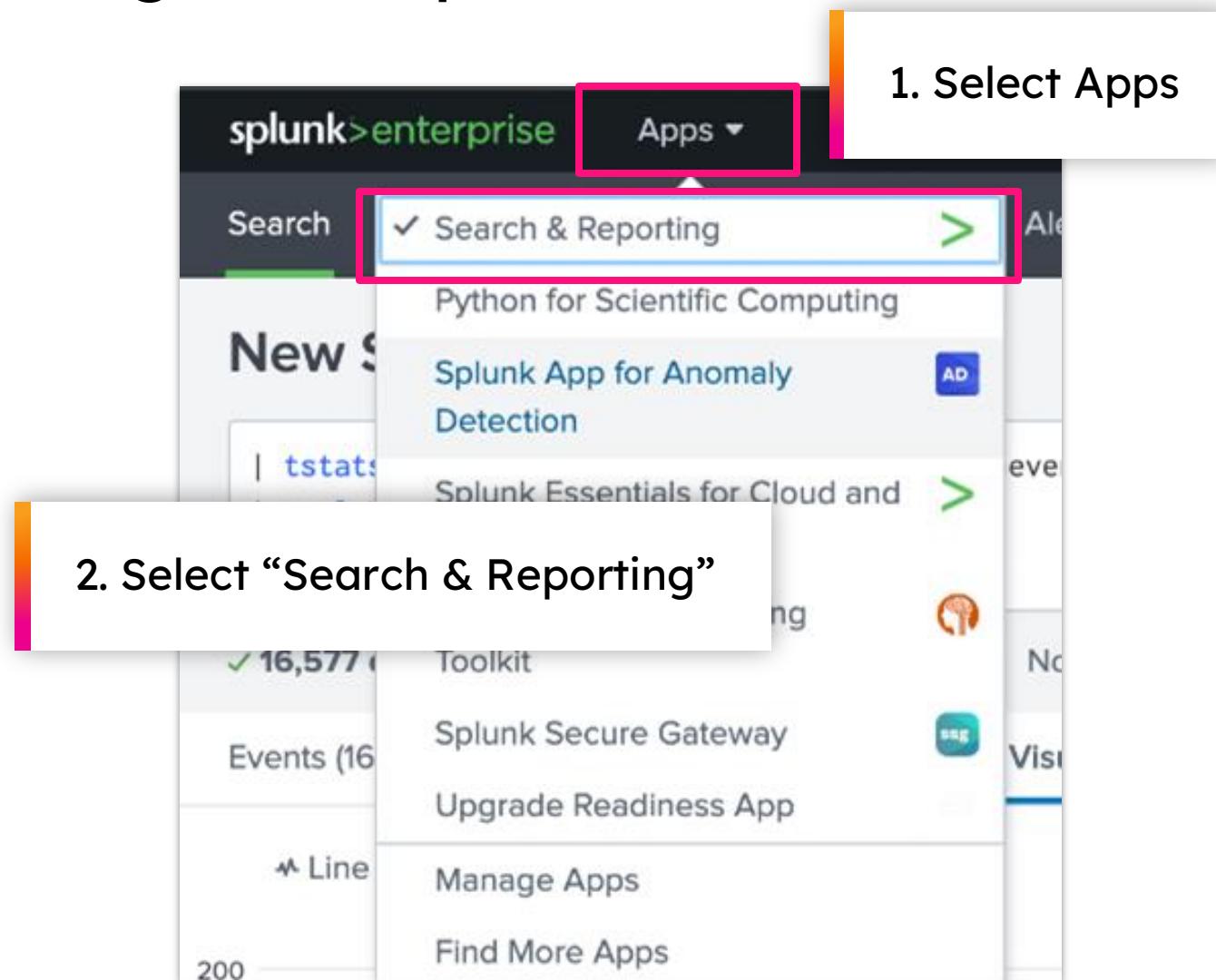
To troubleshoot what might be going on, you decide to analyze your network to look for signs of lag between your services.

For the investigation, you use the **round-trip time (RTT)**, or the time it takes for a network request to go from a starting point to a destination and back again to the starting point. In this case, finding the average client RTT indicates the network latency.



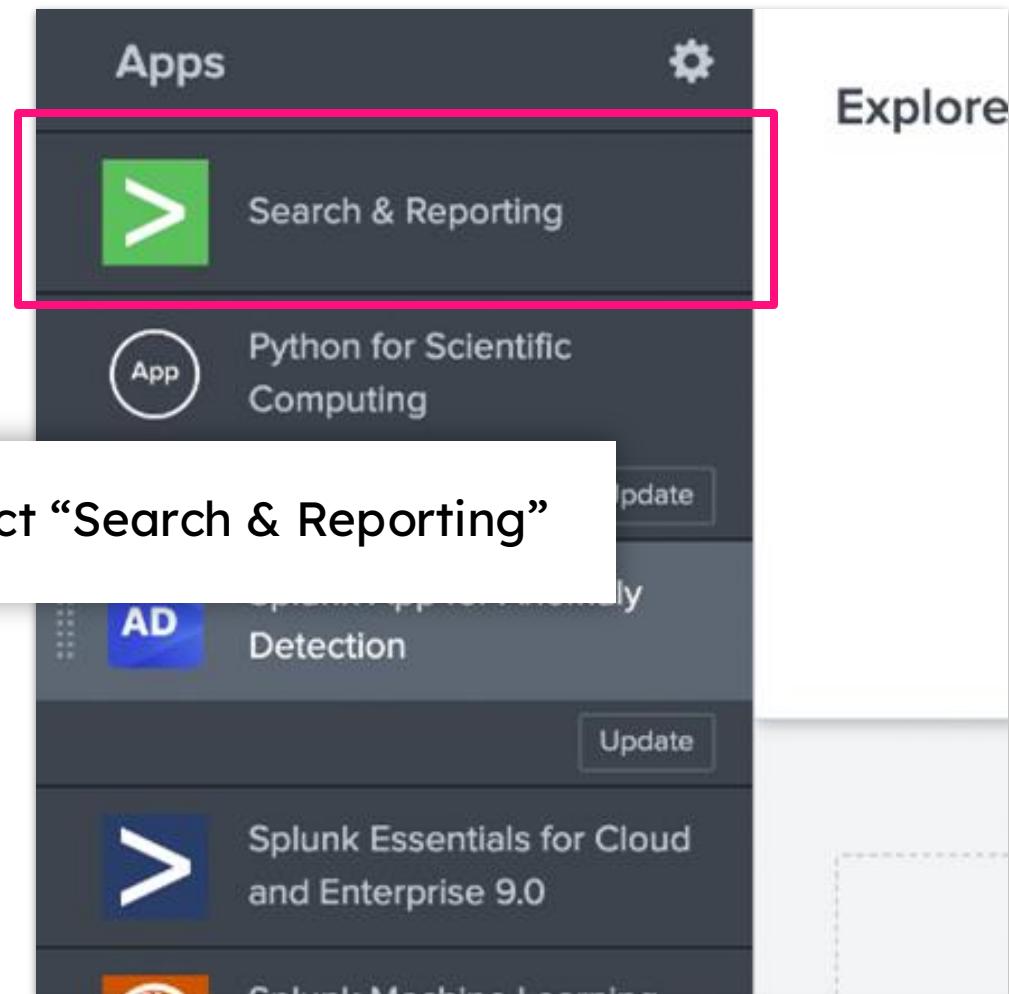
Search & Reporting

Navigation Options



1. Select Apps

2. Select “Search & Reporting”



1. Select “Search & Reporting”

Stream Data

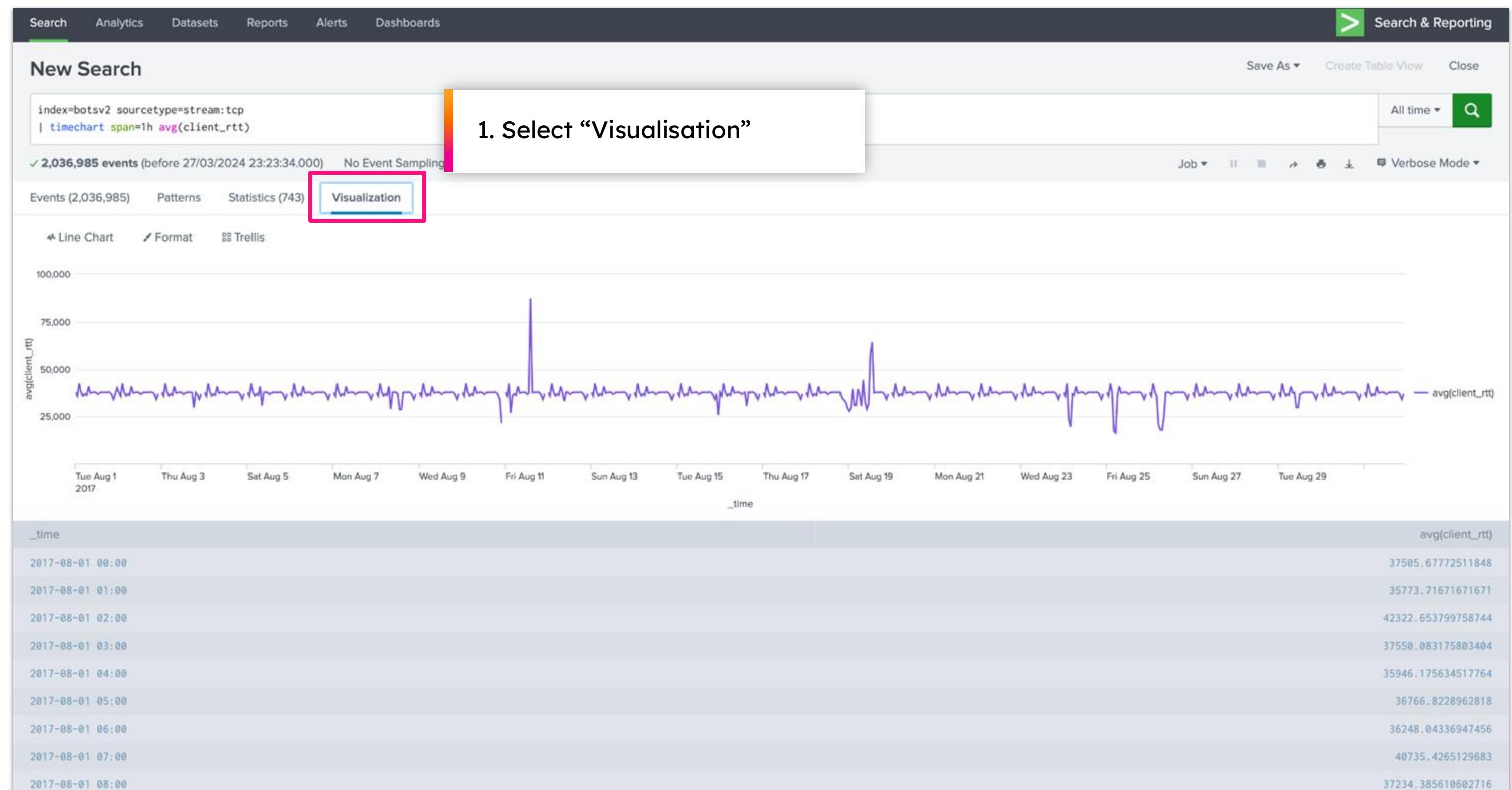
Search 4-1

The screenshot shows a Splunk search interface with the following details:

- Search Bar:** index=botsv2 sourcetype=stream:tcp | timechart span=1h avg(client_rtt)
- Results Summary:** 2,036,985 events (before 27/07/2021 22:00:00)
- Event List:** A table showing the average client round-trip time (avg(client_rtt)) for each hour from 2017-08-01 00:00 to 2017-08-01 18:00.
- Table Headers:** _time, avg(client_rtt)
- Table Data:** The table lists 24 rows, each representing an hour from 00:00 to 18:00, with the corresponding average value.

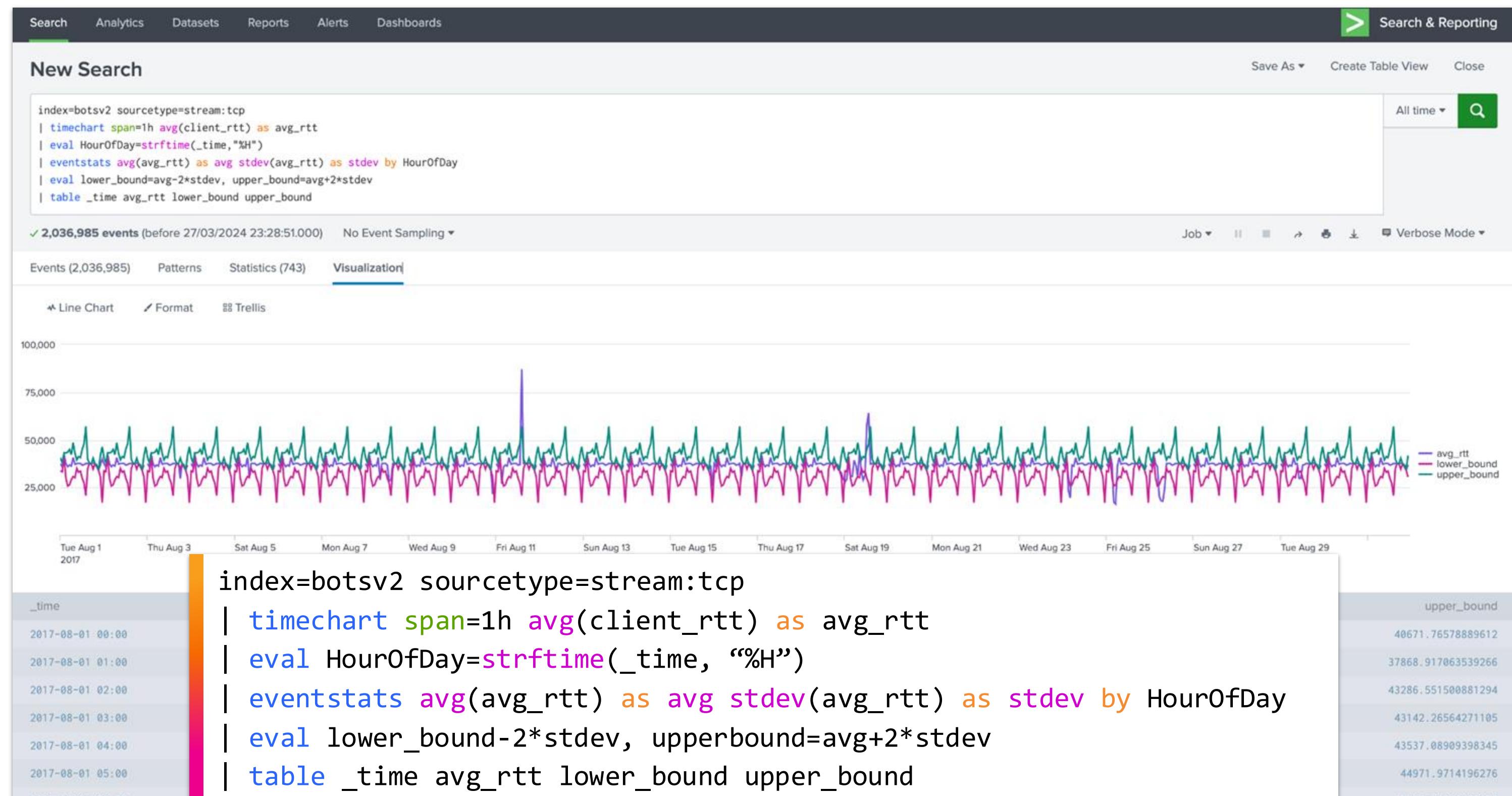
_time	avg(client_rtt)
2017-08-01 00:00	37505.67772511848
2017-08-01 01:00	35773.71671671671
2017-08-01 02:00	42322.653799758744
2017-08-01 03:00	37550.083175803404
2017-08-01 04:00	35946.175634517764
2017-08-01 05:00	36766.8228962818
2017-08-01 06:00	36248.04336947456
2017-08-01 07:00	40735.4265129683
2017-08-01 08:00	37234.385610602716
2017-08-01 09:00	37529.81444552529
2017-08-01 10:00	37645.291368743616
2017-08-01 11:00	37603.05593451569
2017-08-01 12:00	36757.79045579688
2017-08-01 13:00	37015.92988505747
2017-08-01 14:00	37638.583692142085
2017-08-01 15:00	37564.09049286292
2017-08-01 16:00	37688.83705414421
2017-08-01 17:00	37514.034650924026
2017-08-01 18:00	37770.27921879768

Stream Data Visualisation



Stream Data

Search 4-2



Stream Data Outliers

Search Analytics Datasets Reports Alerts Dashboards > Search & Reporting

New Search

```
index=botsv2 sourcetype=stream:tcp  
| timechart span=1h avg(client_rtt) as avg_rtt  
| eval HourOfDay=strftime(_time,"%H")  
| eventstats avg(avg_rtt) as avg stdev(avg_rtt) as stdev by HourOfDay  
| eval lower_bound=avg-2*stdev, upper_bound=avg+2*stdev  
| table _time avg_rtt lower_bound upper_bound
```

2,036,985 events (before search)

Events (2,036,985) Patterns (0) Visualizations (0) Metrics (0) Fields (0) Jobs (0) All time Verbose Mode

1. Change Visualisation

Outliers Chart

Splunk Visualizations

30 outliers

avg_rtt

14. Aug 16. Aug 18. Aug 20. Aug 22. Aug 24. Aug 26. Aug 28. Aug 30. Aug

_time	avg_rtt	lower_bound	upper_bound
2017-08-01 02:00	37505.67772511848	33714.83921181327	40671.76578889612
	35773.71671671671	33265.07955696914	37868.917063539266
	42322.653799758744	41168.554338627	43286.551500881294

2. Select “Outliers Chart”

Show the acceptable range for a value outside that range.

Search Fragment

```
| table _time, <outlier_variable>, <lower_bound>, <upper_bound>
```

2017-08-01 02:00

Fit Model on Stream Data

Search 4-3

The screenshot shows the Splunk Search & Reporting interface. The search bar contains the following SPL command:

```
index=botsv2 sourcetype=stream:tcp  
| timechart span=1h avg(client_rtt) as avg_rtt  
| eval HourOfDay=strftime(_time,"%H")  
| fit DensityFunction avg_rtt by HourOfDay into rtt_df_model
```

A warning message is displayed: "⚠ Too few training points in some groups will likely result in poor accuracy fo". Below the search bar, it says "2,036,985 events (before 27/03/2024 23:41:11.000)" and "No Event Sampling".

The search results table has the following columns: _time, avg_rtt, HourOfDay, IsOutlier(avg_rtt), and BoundaryRanges. The table shows data for each hour from 00:00 to 09:00 on 2017-08-01. The "Statistics (743)" tab is selected, indicated by a red box.

_time	avg_rtt	HourOfDay	IsOutlier(avg_rtt)	BoundaryRanges
2017-08-01 00:00	37505.67772511848	0	0.0	27831.7391:36604.1985:0.01 37505.6777:Infinity:0
2017-08-01 01:00	35773.71671671671	1	0.0	29371.8608:35177.151:0.01 35773.7167:Infinity:0
2017-08-01 02:00	42322.65379975874	2	0.0	39377.4776:42048.2035:0.01 42322.6538:Infinity:0
2017-08-01 03:00	37550.08317580341	3	0.0	23329.1924:36025.3985:0.01 37550.0832:Infinity:0
2017-08-01 04:00	35946.175634517764	4	0.0	17265.6286:32745.4005:0.01 35946.1756:Infinity:0
2017-08-01 05:00	36766.8228962818	5	0.0	16364.8532:26796.0521:0.0101
2017-08-01 06:00	36248.04336947456	6	0.0	18036.3178:33524.3116:0.0099 36636.8206:Infinity:0
2017-08-01 07:00	40735.4265129683	7	0.0	17861.8214:38603.9182:0.0097 40735.4265:Infinity:0
2017-08-01 08:00	37234.38561060272	8	0.0	-Infinity:23161.4243:0.0041 45554.3516:Infinity:0.006
2017-08-01 09:00	37529.81444552529	9	0.0	33038.7381:36733.3009:0.01 37529.8144:Infinity:0

Apply Model on Stream Data

Search 4-4

Screenshot of the Splunk interface showing a search results table.

The search bar contains:

```
index=botsv2 sourcetype=stream:tcp  
| timechart span=1h avg(client_rtt) as avg_rtt  
| eval HourOfDay=strftime(_time,"%H")  
| apply rtt_df_model
```

A warning message is displayed: "⚠ Too few training points in some groups will likely result in poor accuracy fo".

The search results table has the following columns:

_time	avg_rtt	HourOfDay	IsOutlier(avg_rtt)	BoundaryRanges
2017-08-01 00:00	37505.67772511848	0	0.0	27831.7391:36604.1985:0.01 37505.6777:Infinity:0
2017-08-01 01:00	35773.71671671671	1	0.0	29371.8608:35177.151:0.01 35773.7167:Infinity:0
2017-08-01 02:00	42322.65379975874	2	0.0	39377.4776:42048.2035:0.01 42322.6538:Infinity:0
2017-08-01 03:00	37550.08317580341	3	0.0	23329.1924:36025.3985:0.01 37550.0832:Infinity:0
2017-08-01 04:00	35946.175634517764	4	0.0	17265.6286:32745.4005:0.01 35946.1756:Infinity:0
2017-08-01 05:00	36766.8228962818	5	0.0	16364.8532:26796.0521:0.0101
2017-08-01 06:00	36248.04336947456	6	0.0	18036.3178:33524.3116:0.0099 36636.8206:Infinity:0
2017-08-01 07:00	40735.4265129683	7	0.0	17861.8214:38603.9182:0.0097 40735.4265:Infinity:0
2017-08-01 08:00	37234.38561060272	8	0.0	-Infinity:23161.4243:0.0041 45554.3516:Infinity:0.006
2017-08-01 09:00	37529.81444552529	9	0.0	33038.7381:36733.3009:0.01 37529.8144:Infinity:0

Search Breakdown

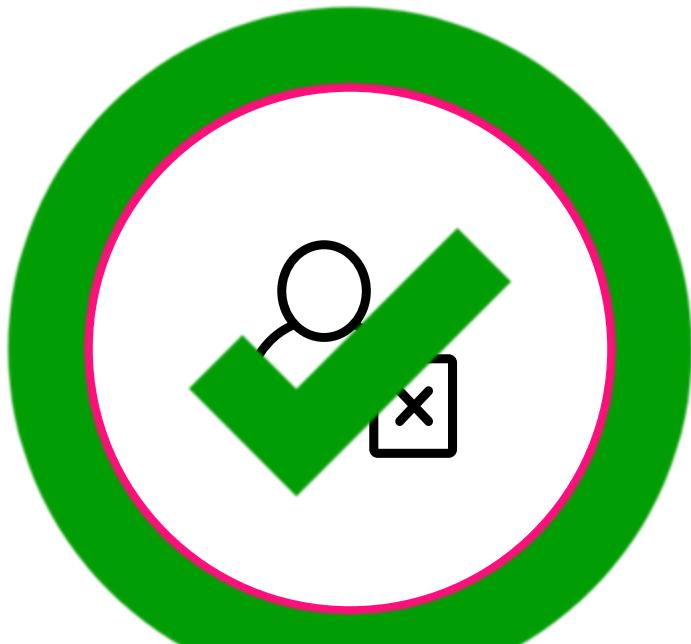
Bookmark

Splunk search	Description
index=botsv2 sourcetype=stream:tcp	Search the stream:tcp logs in the botsv2 index
... timechart span=1h avg(client_rtt) as avg_rtt	For each hour, calculate the average client RTT and save it as avg_rtt
... eval HourOfDay=strftime(_time,"%H")	Returns the hour from the _time field, and store it as HourOfDay strftime(_time,"%H") takes a UNIX time value and returns the time as a string in the specified format, in this case the hour.
... eventstats avg(avg_rtt) as avg stdev(avg_rtt) as stdev by HourOfDay	Calculate the average and standard deviation of the average hourly RTT and save them as avg and stdev
... eval lower_bound=avg-2*stdev, upper_bound=avg+2*stdev	The average and standard deviation are used with the eval command to calculate the lower and upper boundaries. A sensitivity is added into the calculation by multiplying the stdev by 2, as in any distribution, about 95% of values should be within 2 standard deviations of the mean.
... table _time avg_rtt lower_bound upper_bound	Returns a table with fields time, average RTT, lower bounds, and upper bounds
... fit DensityFunction avg_rtt by HourOfDay into rtt_df_model	Fit the DensityFunction algorithm over the average RTT data by hour into the model name, rtt_df_model
... apply rtt_df_model	Apply the saved model named rtt_df_model on the data

Objective for Today

An Understanding of

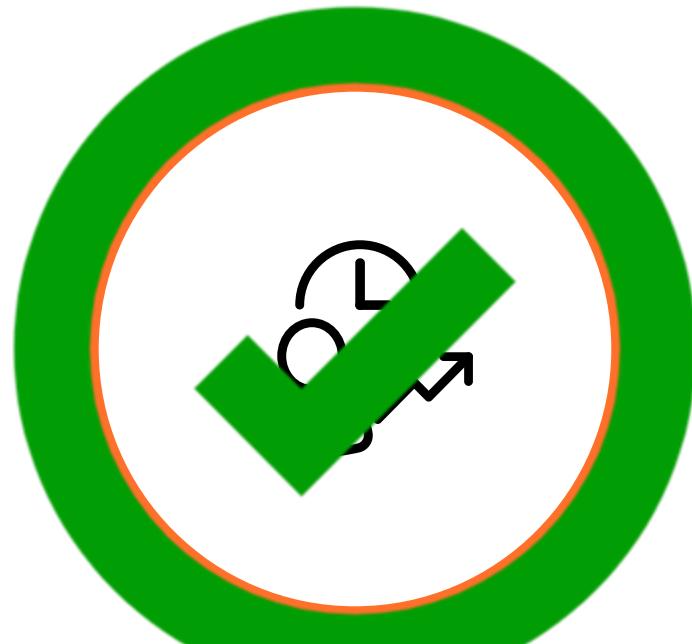
User Access Anomalies



Insider Threat

DensityFunction

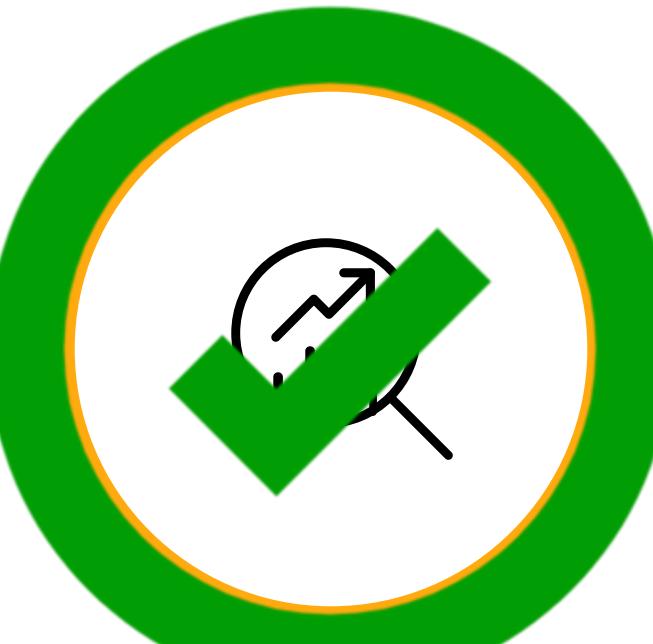
Anomalous Login Quantities



IT Operations

Splunk App

Forecast CPU Utilisation



Infrastructure

StateSpaceForecast

Outliers in Utilisation Telemetry



All

DensityFunction

Splunk Resources

Where to go after today's workshop



Splunk Community

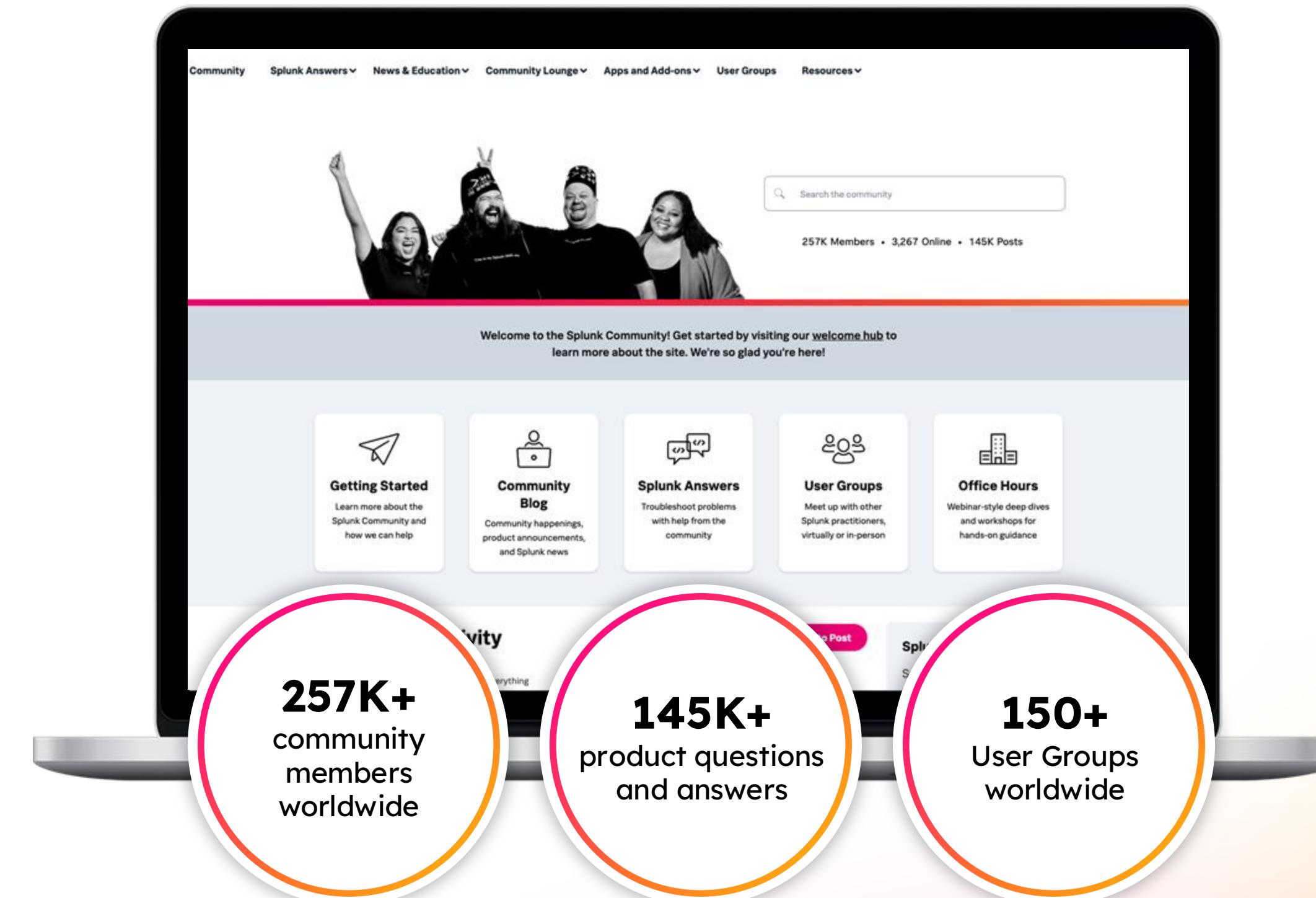
<https://community.splunk.com>

A free way to connect, learn, have fun, and find success with Splunk

Ask questions, get answers, and find solutions from passionate experts in the community

Meet in-person or virtually with like-minded enthusiasts, in your area or by interest

Search for, vote on, or submit your own ideas for new enhancements for any product or solution



Splunk Events

<https://events.splunk.com>

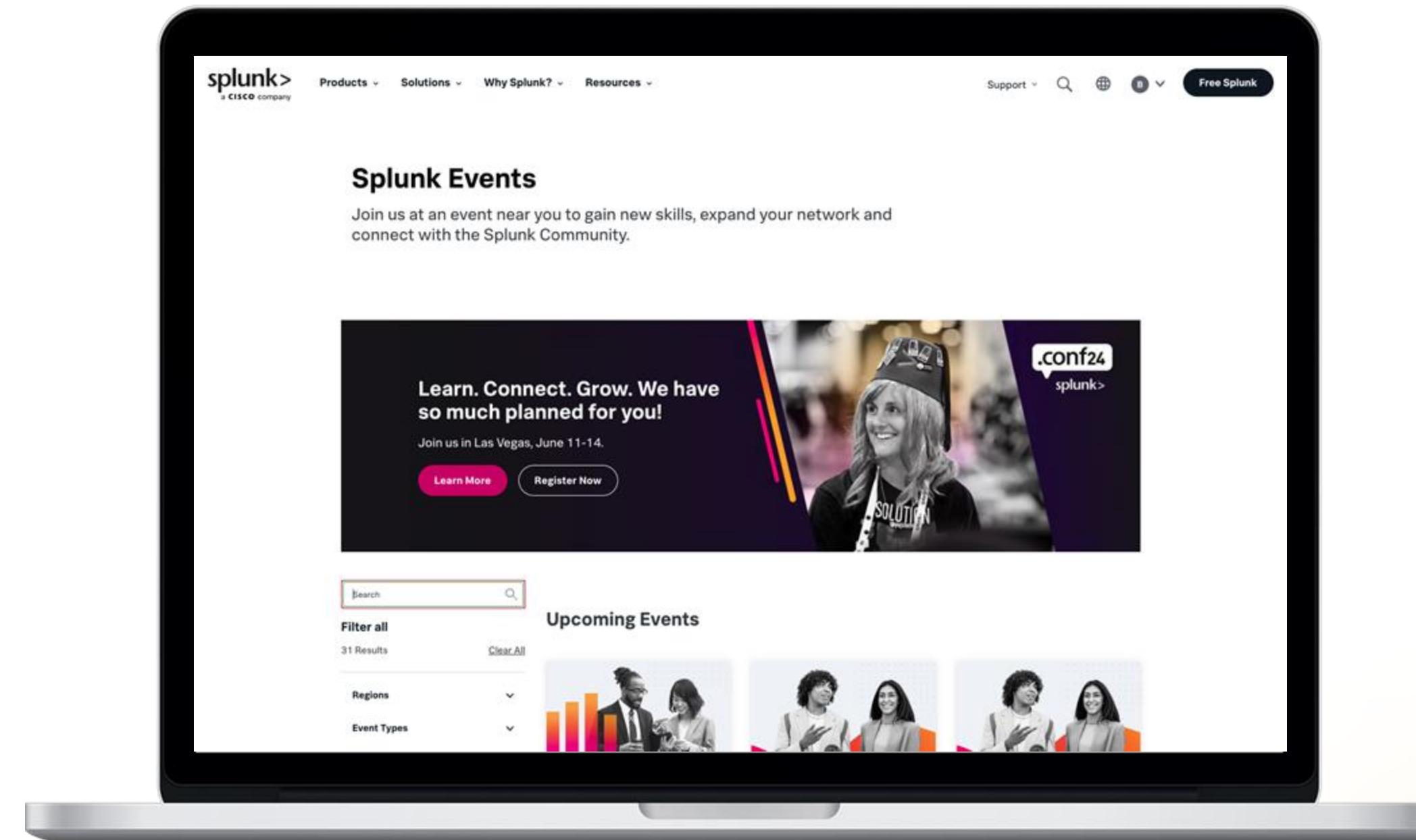
Expand your network and connect with the global and local Splunk community



<https://conf.splunk.com>

Join us at .conf24!

Hundreds of on-demand sessions from product updates to learning new Splunk skills!



Documentation

<https://docs.splunk.com>

Search reference for SPL

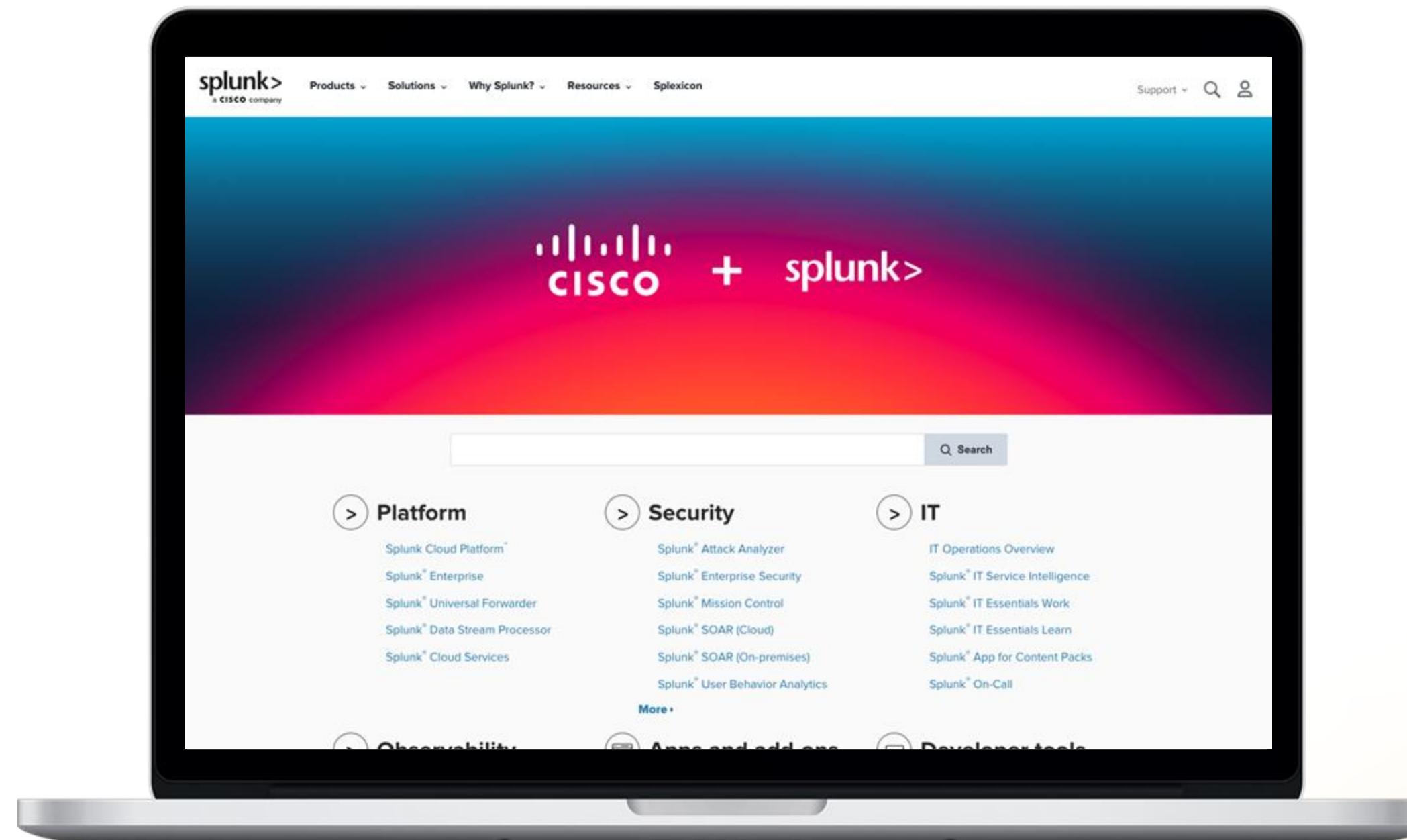
Machine Learning

<https://splk.it/3xcroUX>

Product references

Procedures/guides

And more!



Splunk Lantern

<https://lantern.splunk.com>

Use case library

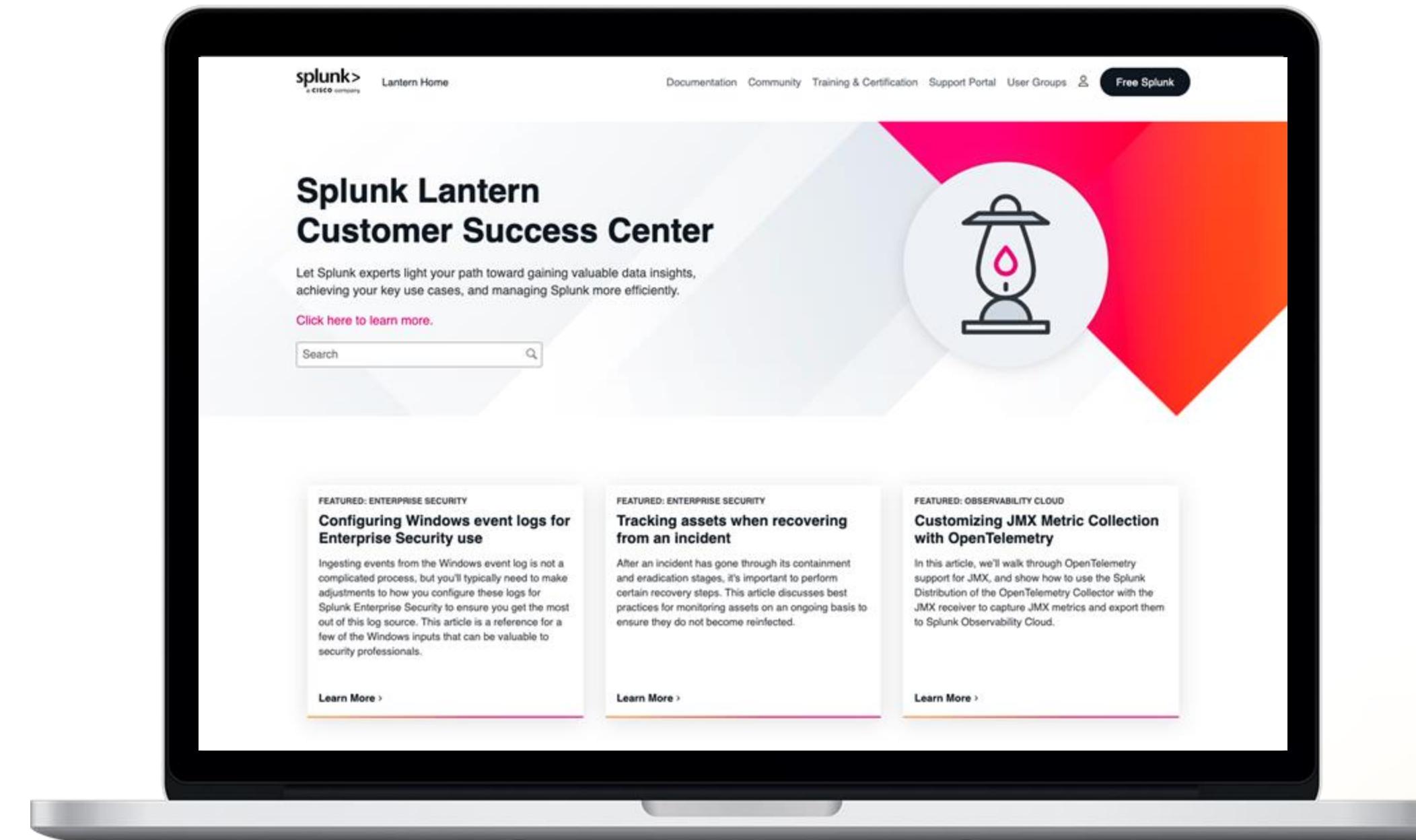
Step-by-step procedures

Map use cases to data sources

Splunk Success Framework to realize value across your organisation

Machine Learning

<https://splk.it/3TU9DCp>



Developer Resources

<https://dev.splunk.com>

Developer Guide

API Reference

Tutorials

Downloads

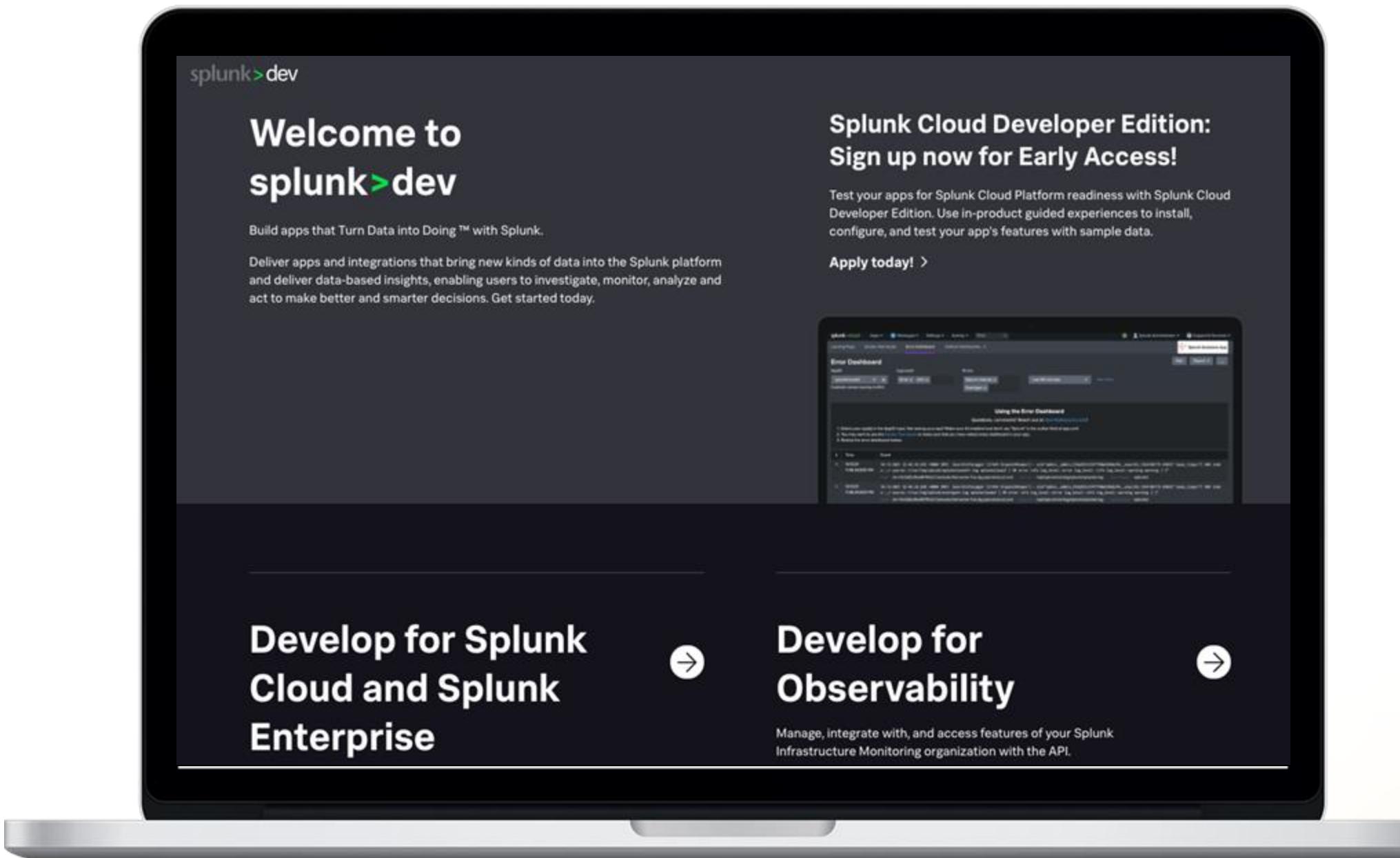
APIs, libraries, tools

Code examples

Free Developer licence

Splunk Cloud Developer Edition

Test your apps for Splunk Cloud readiness



Splunk Apps & Add-ons

<https://splunkbase.splunk.com/>

3200+ apps and add-ons

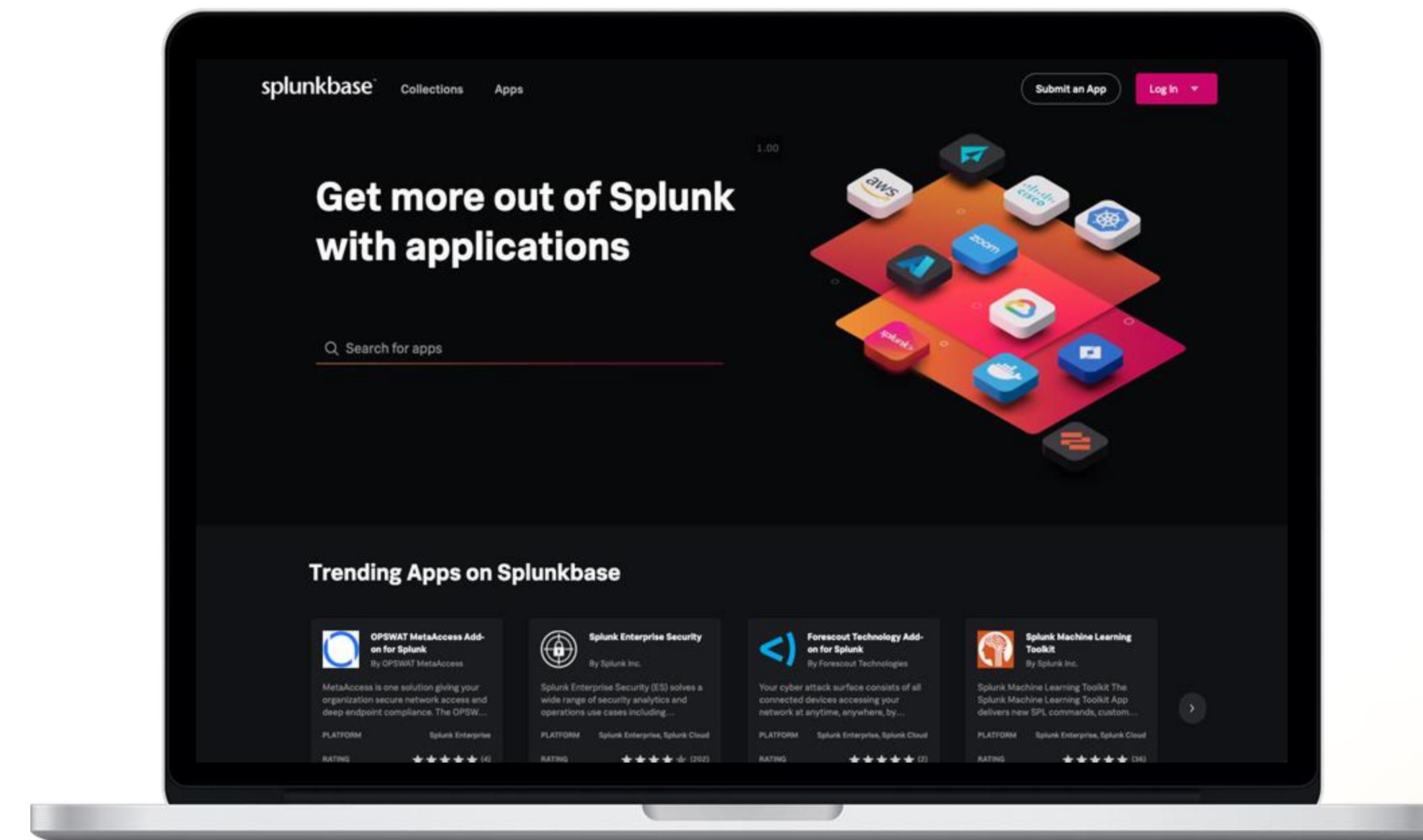
Pre-built searches, reports, visualisations and integrations

Splunk Machine Learning Toolkit (MLTK)

<https://splk.it/3x67rP>

Python for Scientific Computing
<https://splk.it/43C7kHy>

Splunk App for Data Science and Deep Learning (DSDL)
<https://splk.it/3VBUQm>



Training & Certification

<https://splunk.com/training>

Online education classes

Instructor-led and self-paced eLearning

Certification tracks for different roles

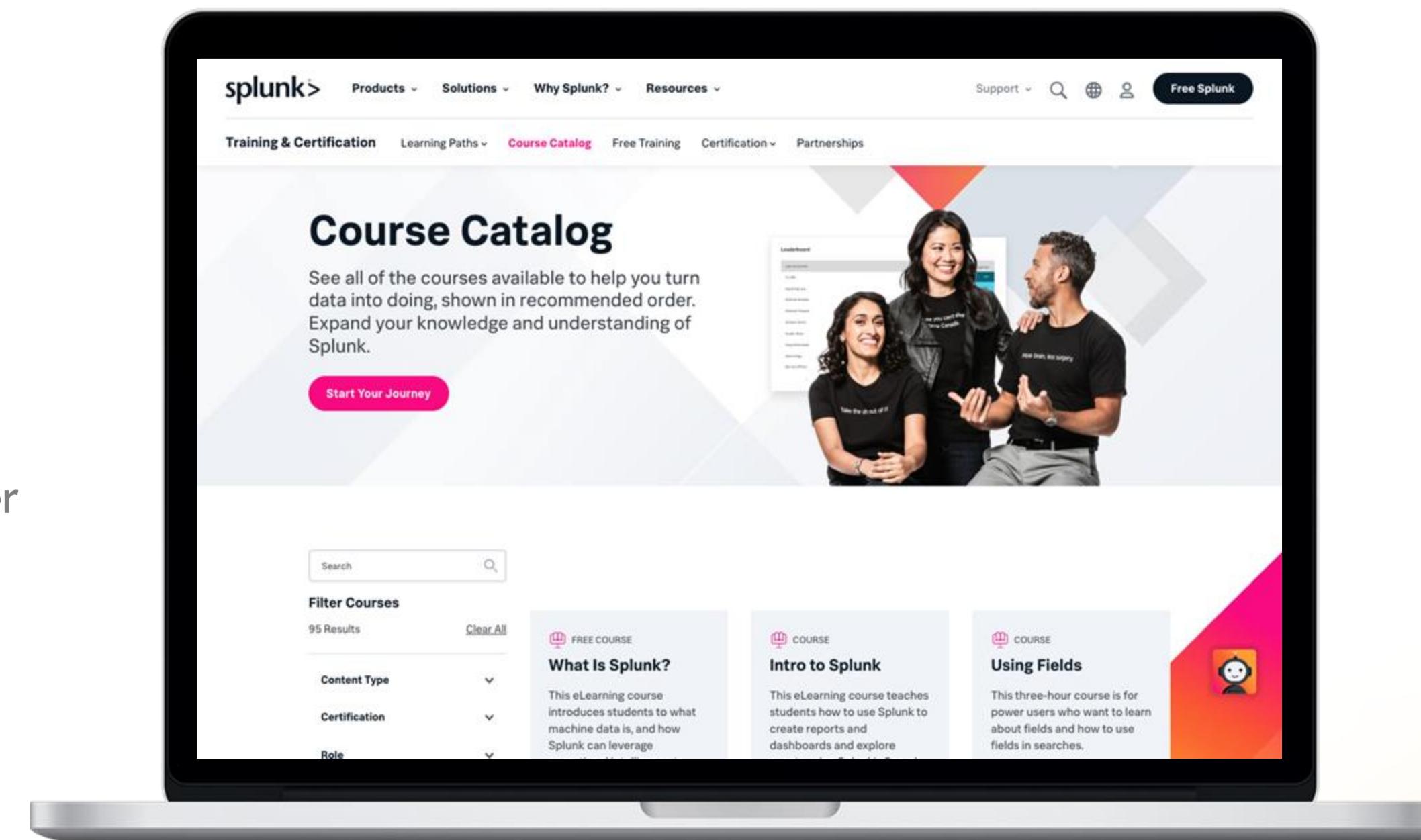
User, Power User, Admin, Architect and Developer

Free education!

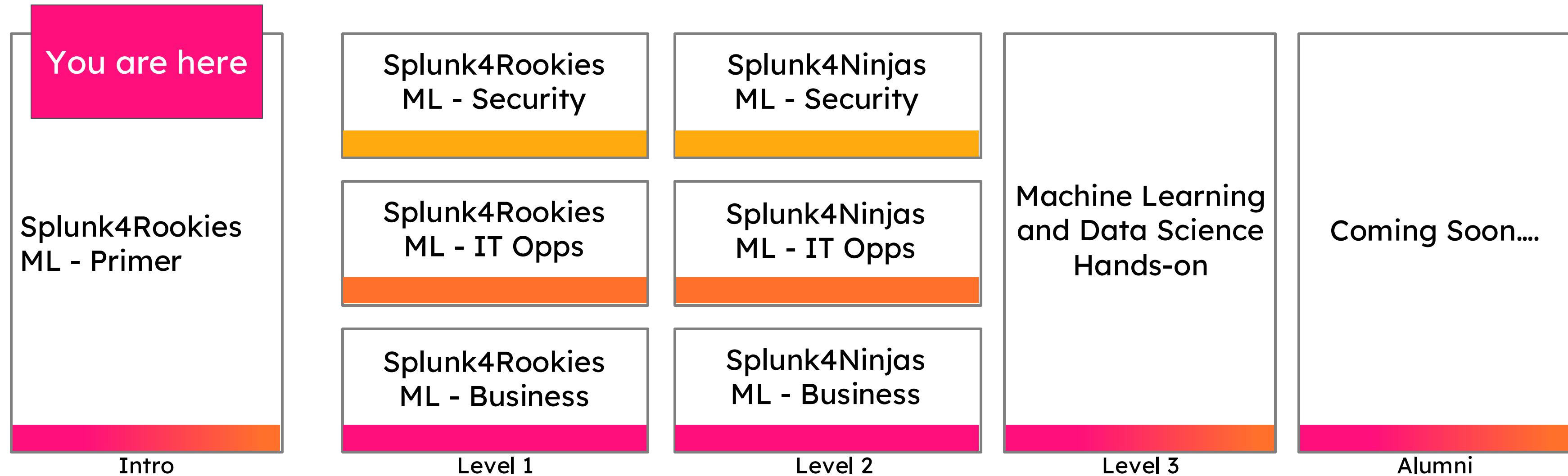
Free single-subject eLearning courses
to kick start your Splunk learning

Splunk for Analytics and Data Science

<https://splk.it/43E7CxE>



Splunk Machine Learning Workshops



Boss of Machine Learning (Coming Soon)

Introductory

Advanced

Legend

Security

IT Ops

Business

All

Thank you

