Splunk 201 Hands-on Workshop

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>

Splunk 201 Session Overview

Session Goals:

- Enable your teams to search, investigate, analyze, & report on data
- Drive more value and better outcomes from machine data
- Increase teams’ productivity and efficiency in resolving issues

Splunk 201 Agenda - for Splunk Power Users:

- Transactional Analysis
  - SPL Commands: stats list/values, transaction, chart (for histograms)
- Advanced Statistics
  - SPL Commands: stats stdev/variance, eventstats, streamstats
- Schema-on-Read and Field Extractions
  - The field extractor, eventtypes/tags/macros, CIM
- (Optional) Advanced Dashboarding

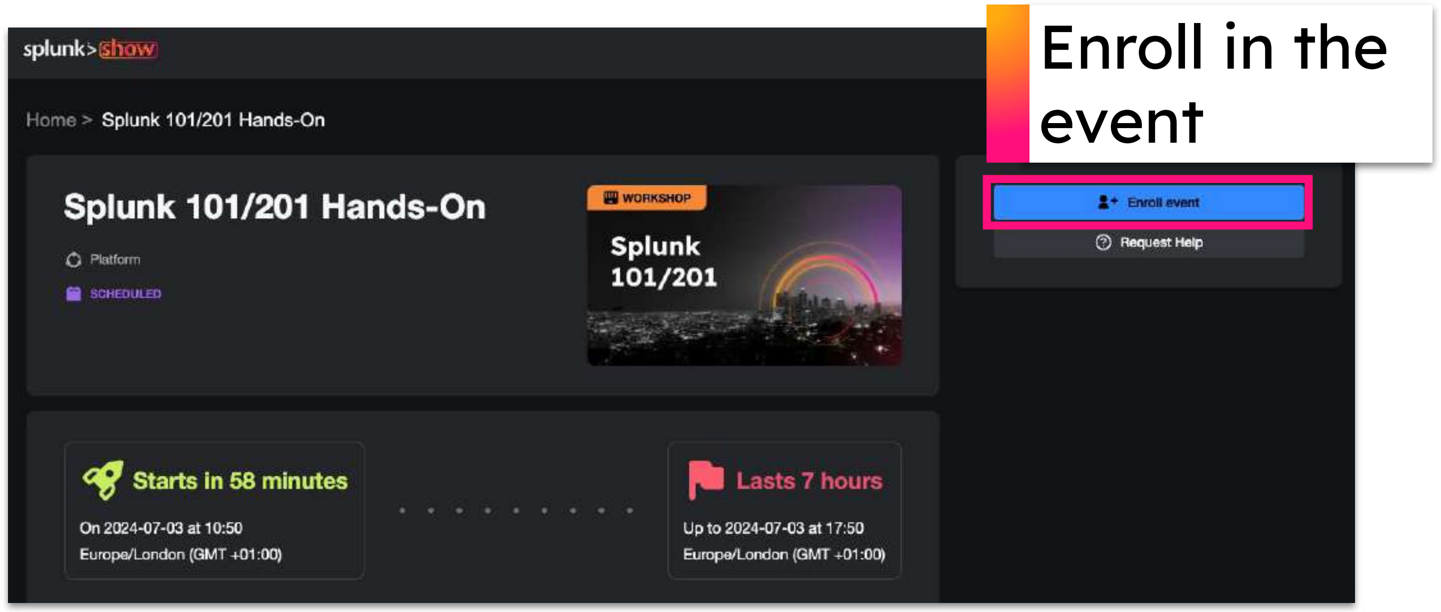
REGISTRATION 5 - 10 MINS

Lab Instance Enrollment & Log In

Tasks

1. Get a splunk.com account if you don’t have one yet: https://splk.it/SignUp
2. Enroll in the Splunk Show workshop event: https://show.splunk.com/event/<eventID>
3. Download today’s slides: https://splk.it/201-Attendee
4. Download the lab guide: https://splk.it/201-Lab-Guide
5. Download a copy of the Splunk Quick Reference Guide: https://splk.it/SplunkQuickRef

Goal



Splunk 101/201 Hands-On

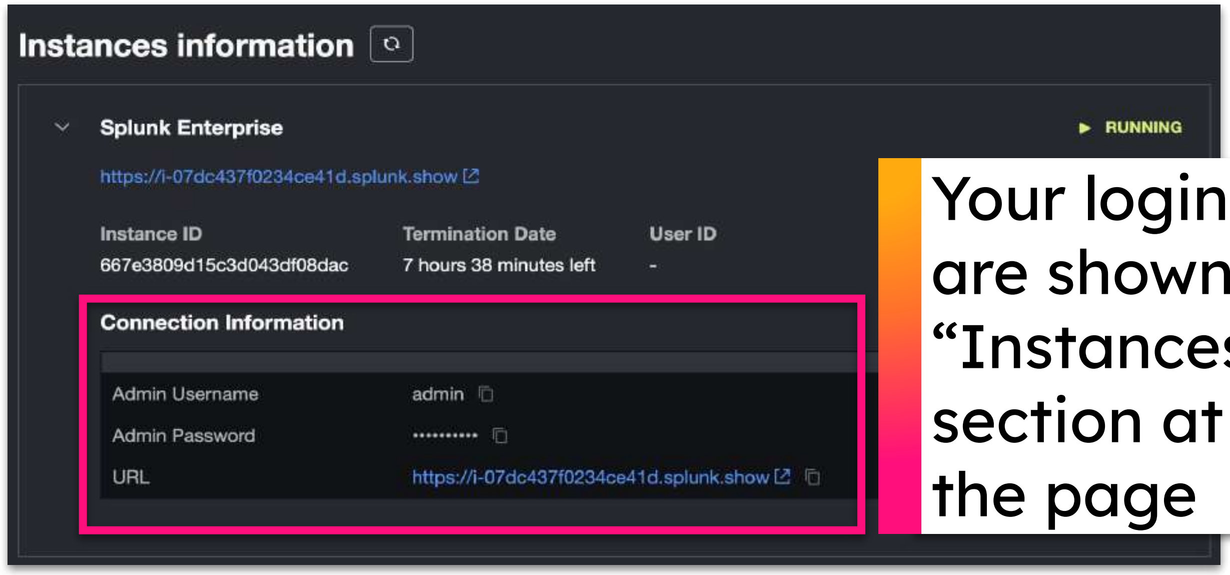
Platform

SCHEDULED

Starts in 58 minutes
On 2024-07-03 at 10:50
Europe/London (GMT +01:00)

Lasts 7 hours
Up to 2024-07-03 at 17:50
Europe/London (GMT +01:00)

Enroll in the event



Instances information

Splunk Enterprise ▶ RUNNING

https://i-07dc437f0234ce41d.splunk.show

Instance ID	Termination Date	User ID
667e3809d15c3d043df08dac	7 hours 38 minutes left	-

Connection Information

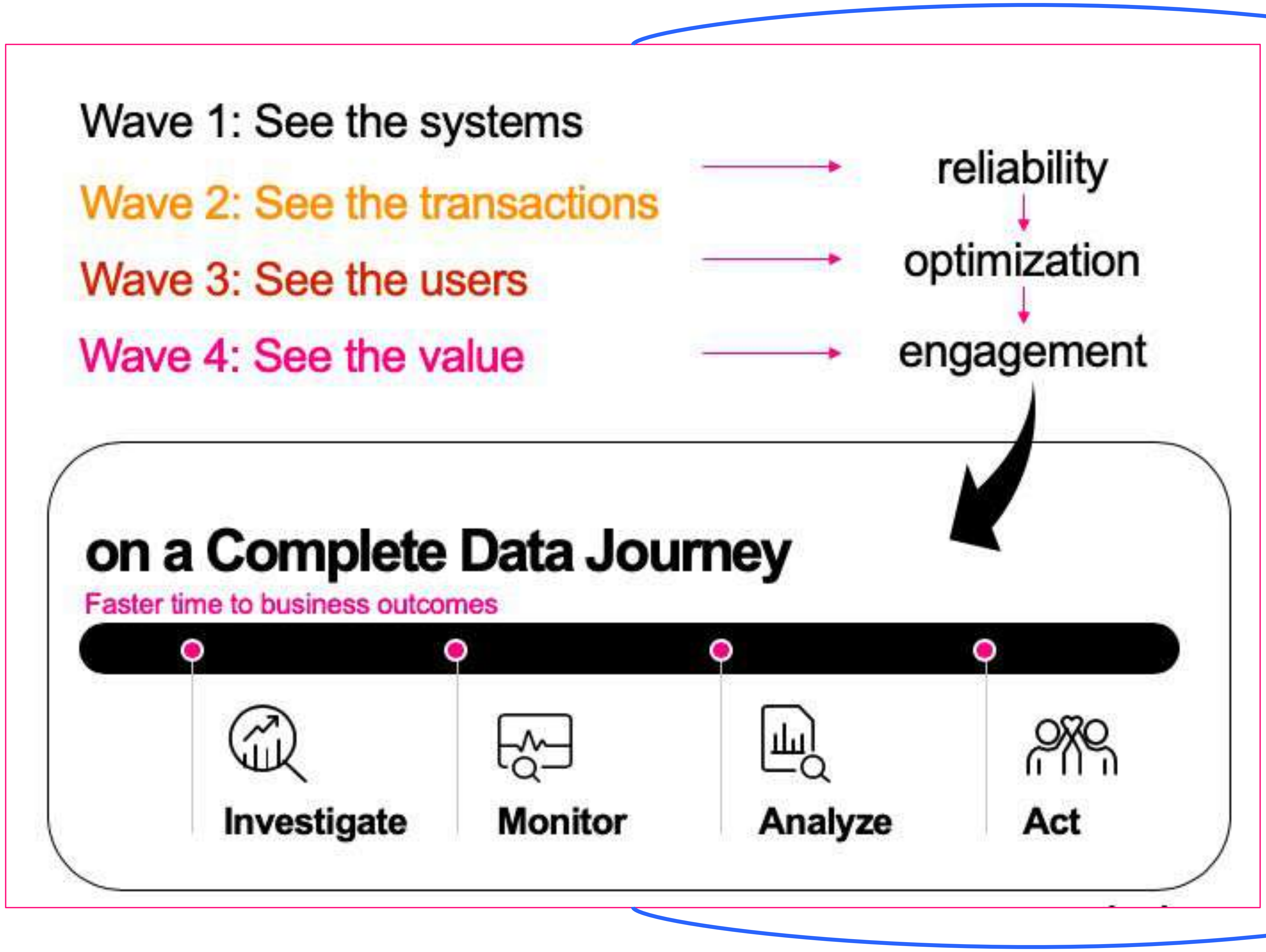
Admin Username	admin
Admin Password	•••••••••••
URL	https://i-07dc437f0234ce41d.splunk.show

Your login credentials are shown under the “Instances Information” section at the bottom of the page

Transactional Analysis

Using stats, transaction, and chart

WHAT is transactional analysis and WHY is important?



Wave 1: See the systems

Wave 2: See the transactions

Wave 3: See the users

Wave 4: See the value

reliability → optimization → engagement

on a Complete Data Journey

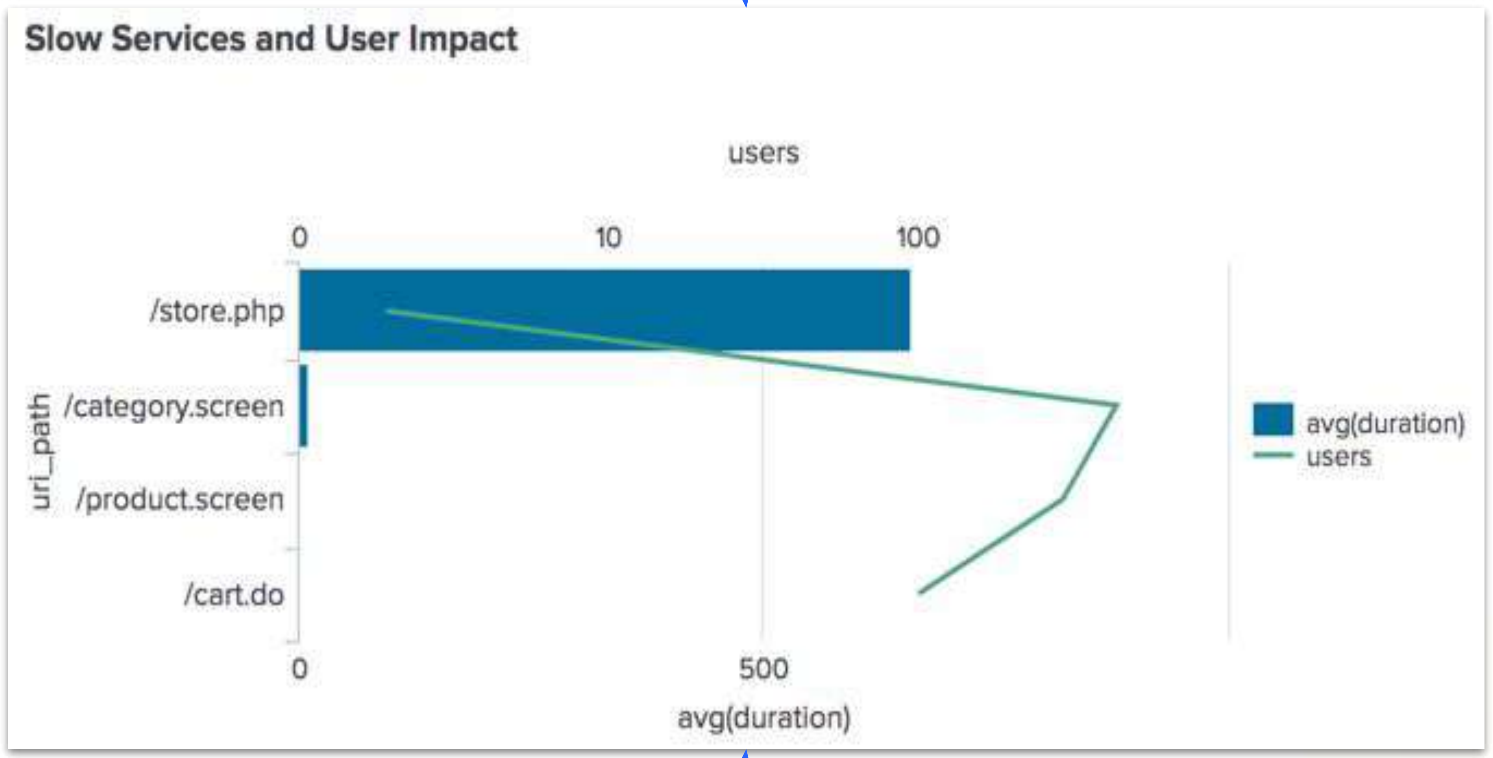
Faster time to business outcomes

Investigate

Monitor

Analyze

Act



Slow Services and User Impact

LAB EXERCISE 1 5 MINS

stats list / values

Use the stats command to turn data into aggregate tables!

Syntax:

```
<your search> | stats <function> <by clause>
```

Overview of stats list / values:

- list function:
  - Returns a multivalue entry from the values in a field
  - Order of values reflects order of events
- values function:
  - Returns a list of the distinct values in a field as a multivalue entry
  - Order of values is alphabetical

Exercise #1: Build transactions with stats list/values. For every transaction include: the duration, values of the action field, and associated raw events.

- Hints:
  - Focus on web log events (sourcetype clue)
  - **Don’t** include address or category_id fields
  - Sort results in descending order by the transaction “duration”
  - Only show 10 most recent transactions (use head command)

Want to know more? Check out:
Splunk Docs (stats list & values): https://docs.splunk.com/Documentation/Splunk/9.2.2/SearchReference/Multivaluefunctions#list.28.26lt.3Bvalue.26gt.3B.29

LAB EXERCISE 2 3 MINS

The `transaction` command

Use the `transaction` command to easily group events into transactions!

Overview of `transaction` **command:**

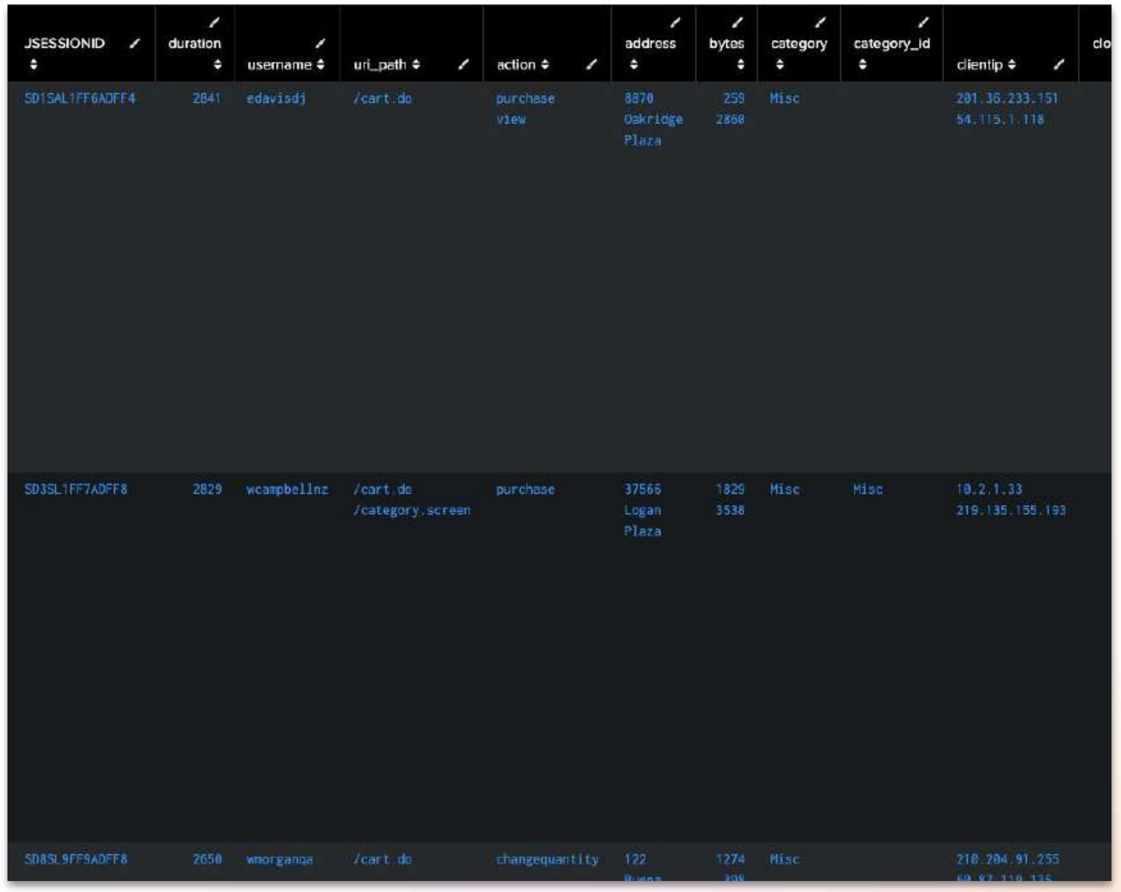
- Finds transactions based on events that meet various constraints
- Made up of raw text of each member, time and date fields of earliest member, and union of all other fields
- Adds 2 fields to raw events: duration & eventcount

Syntax:

```
<your search> | transaction <fields>
```

Exercise #2: Build transactions with the transaction command. Group events based on the username and session ID fields.

- Hints:
  - Focus on web log events (sourcetype clue)
  - Sort results in descending order by the transaction "duration"
  - Only show 10 most recent transactions (use head command)
  - Display results in a table and include the following fields: JSESSIONID, duration, username, uri_path



JSESSIONID	duration	username	uri_path	action	address	bytes	category	category_id	clientip	clo
SD1SAL1FF6ADFF4	2841	edavisdj	/cart.do	purchase view	8870 Oakridge Plaza	259 2860	Misc		201.36.233.151 54.115.1.118	
SD3SL1FF7ADFF8	2829	wcampbellnz	/cart.do /category.screen	purchase	37566 Logan Plaza	1829 3538	Misc	Misc	10.2.1.33 219.135.155.193	
SD8SL9FF9ADFF8	2650	wmorganqa	/cart.do	changequantity	122	1274	Misc		210.204.91.255	

Want to know more? Check out:
Splunk Docs (transaction): https://docs.splunk.com/Documentation/Splunk/9.2.2/SearchReference/Transaction

LAB EXERCISE 3 5 MINS

The chart command & histograms

Use the chart command to easily visualize your event data and perform statistical analysis

Overview of chart command:

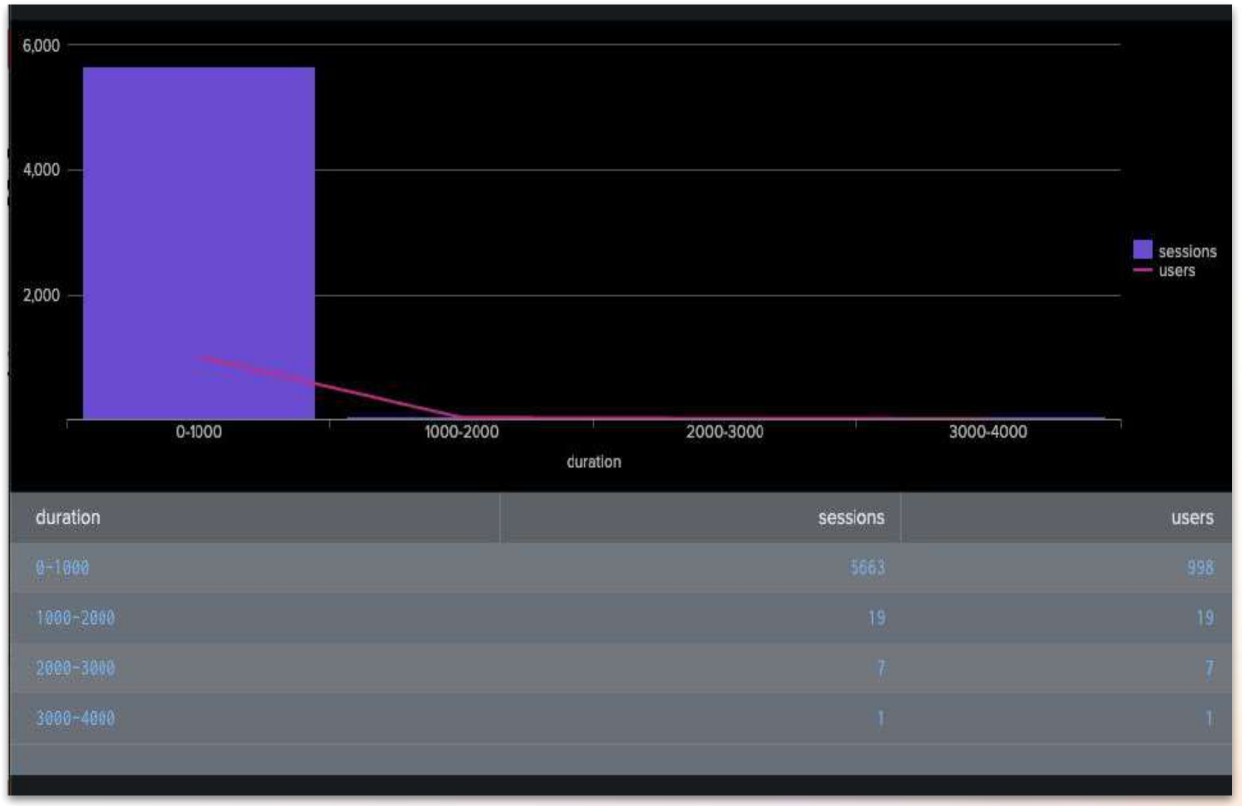
- Transforming command that returns results in table format
- Results can usually be used to create visualizations to display statistical information, such as trends over time
- A statistical function is required with this command

Syntax:

```
<your search> | chart <function> <by clause>
```

Exercise #3: Determine how the duration of web user sessions varies and how many distinct users are seen for each duration interval

- Hints:
  - Focus on web log events (sourcetype clue)
  - Start by creating transactions to represent the user sessions!
  - Use chart option bins=10 to create different duration intervals
  - For each duration interval, display the number of sessions and distinct users



duration	sessions	users
0-1000	5663	998
1000-2000	19	19
2000-3000	7	7
3000-4000	1	1

Want to know more? Check out:
Splunk Docs (chart): https://docs.splunk.com/Documentation/Splunk/9.3.0/SearchReference/Chart

Advanced Statistics

stats stdev/variance functions, eventstats, & streamstats

Advanced Statistics vs Machine Learning

Splunk makes it easy to do multiple types of statistics:

- **(101) Statistics:** aggregate and analyze numerical data split by groups (counts, sums, avgs & medians)
- **(201) Transactional Statistics:** analyze higher-level entities and transactions from raw data
- **(201) Advanced Statistics:** incorporate variability (stdev/percentiles), apply eventstats & streamstats
- **(301) Machine Learning:** use statistical models to detect anomalies and sudden changes, adjust thresholds dynamically, and predict service degradation

References:

- List of stats functions: https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/CommonStatsFunctions
- Download the Splunk Machine Learning Toolkit (MLTK): https://splunkbase.splunk.com/app/2890/
- Download Python for Scientific Computing Add-on for your OS: https://splunkbase.splunk.com/apps?page=1&keyword=Python+for+Scientific+Computing&filters=product%3Asplunk%2Fcategories%3Aai

LAB EXERCISE 4 10 MINS

The stats stdev function

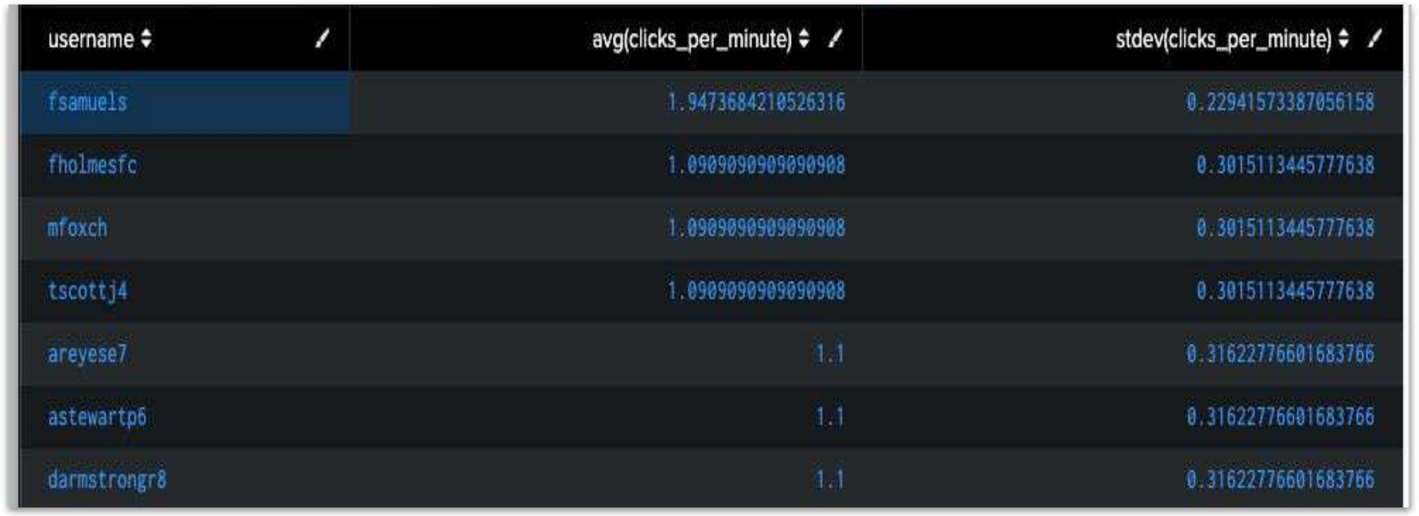
Syntax:

```
<your search> | stats stdev(<field>) <by clause>
```

Use the stdev function to find anomalies and/or suspicious activity in your event logs

Overview of stdev **function:**

- Returns the sample standard deviation of the field specified
- Commonly used with stats, can also be used with:
  - chart
  - timechart
  - tstats, mstats (not covered yet)



username	avg(clicks_per_minute)	stdev(clicks_per_minute)
fsamuels	1.9473684210526316	0.22941573387056158
fholmesfc	1.0909090909090908	0.3015113445777638
mfoxch	1.0909090909090908	0.3015113445777638
tscottj4	1.0909090909090908	0.3015113445777638
areyese7	1.1	0.31622776601683766
astewartp6	1.1	0.31622776601683766
darmstrongr8	1.1	0.31622776601683766

Exercise #4: Track the average duration of user sessions using averages and medians, and the typical fluctuation of these values using stdev and percentiles.

- Hints:
  - Focus on web log events (sourcetype clue)
  - Start by creating event transactions to identify the user sessions and their duration
  - Use timechart with statistical functions to generate the required analytics over time!

Want to know more? Check out:
Splunk Docs (stats stdev): https://docs.splunk.com/Documentation/Splunk/9.2.2/SearchReference/Aggregatefunctions#stdev.28.26lt.3Bvalue.26gt.3B.29

LAB EXERCISE 5 5 MINS

The eventstats command

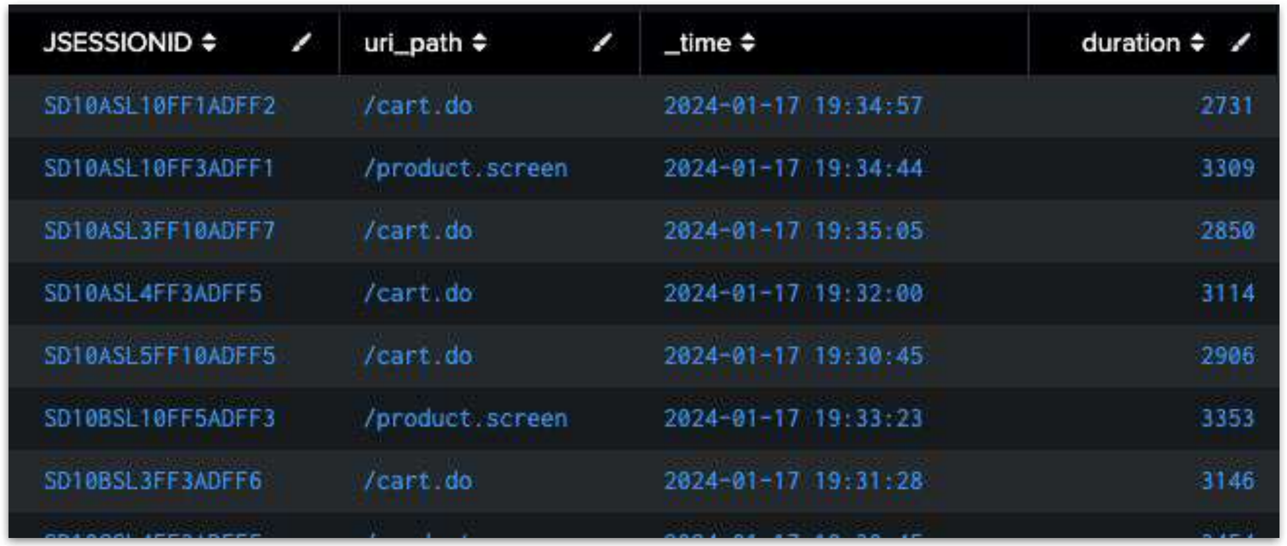
Use the eventstats command to build adaptive thresholds based on peer groups

Overview of eventstats **command:**

- Generates summary statistics from fields in your events and saves those statistics in a new field
- Only events with the fields used in the eventstats commands will be used
- Generated summary statistics can be used for calculations in subsequent commands in the same search

Syntax:

```
<your search> | eventstats <function> <by clause>
```



JSESSIONID	uri_path	_time	duration
SD10ASL10FF1ADFF2	/cart.do	2024-01-17 19:34:57	2731
SD10ASL10FF3ADFF1	/product.screen	2024-01-17 19:34:44	3309
SD10ASL3FF10ADFF7	/cart.do	2024-01-17 19:35:05	2850
SD10ASL4FF3ADFF5	/cart.do	2024-01-17 19:32:00	3114
SD10ASL5FF10ADFF5	/cart.do	2024-01-17 19:30:45	2906
SD10BSL10FF5ADFF3	/product.screen	2024-01-17 19:33:23	3353
SD10BSL3FF3ADFF6	/cart.do	2024-01-17 19:31:28	3146

Exercise #5: Find transactions with very slow durations, and group the results by the uri_path field

- Hints:
  - Focus on web log events (sourcetype clue)
  - Create transactions to identify the web sessions by session ID and uri_path
  - “Slow duration” if duration > (average duration + 2 x standard deviation for the duration)

Want to know more? Check out:
Splunk Docs (eventstats): https://docs.splunk.com/Documentation/Splunk/9.3.0/SearchReference/Eventstats

LAB EXERCISE 6 5 MINS

The streamstats command

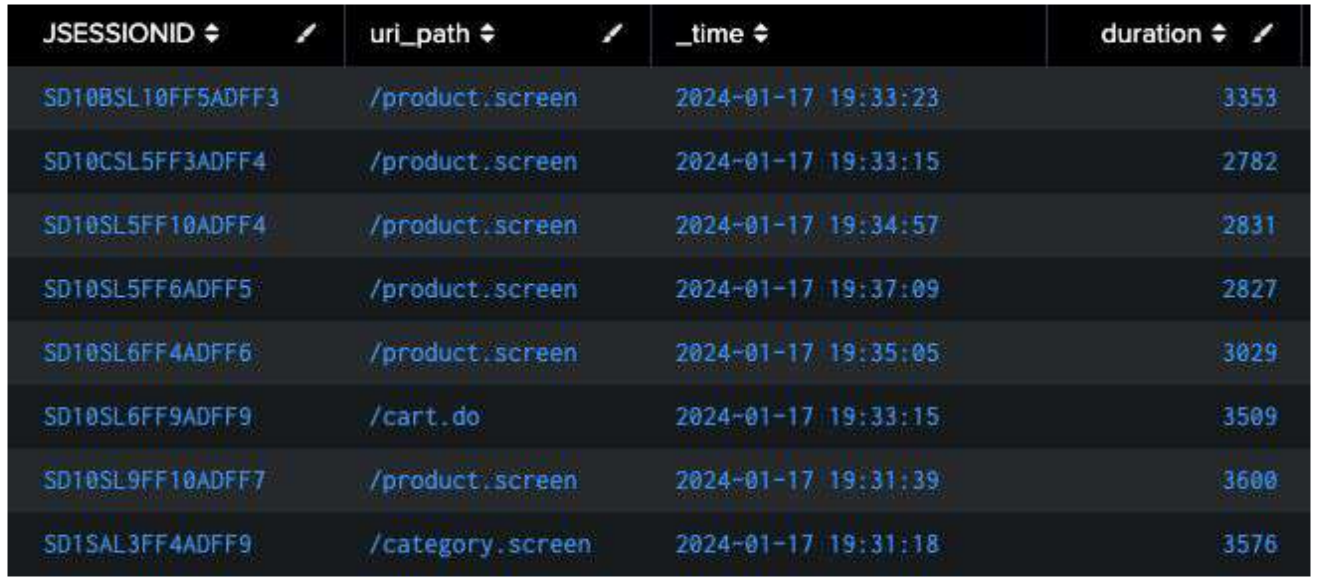
Use the streamstats command to build historical adaptive thresholds and track service health over time

Syntax:

```
<your search> | streamstats <function> <by clause>
```

Overview of streamstats **command:**

- Adds cumulative summary statistics to all search results in a streaming manner
- Statistics for each event calculated when event is seen
- Operates on whatever search output it receives and is the accumulation of the statistical function values in the events seen at that point



JSESSIONID	uri_path	_time	duration
SD10BSL10FF5ADFF3	/product.screen	2024-01-17 19:33:23	3353
SD10CSL5FF3ADFF4	/product.screen	2024-01-17 19:33:15	2782
SD10SL5FF10ADFF4	/product.screen	2024-01-17 19:34:57	2831
SD10SL5FF6ADFF5	/product.screen	2024-01-17 19:37:09	2827
SD10SL6FF4ADFF6	/product.screen	2024-01-17 19:35:05	3029
SD10SL6FF9ADFF9	/cart.do	2024-01-17 19:33:15	3509
SD10SL9FF10ADFF7	/product.screen	2024-01-17 19:31:39	3600
SD1SAL3FF4ADFF9	/category.screen	2024-01-17 19:31:18	3576

Exercise #6: Find anomalous transactions that are impacting service health (2 stdev slower than historical average)

- Hints:
  - Focus on web log events (sourcetype clue)
  - Create transactions to identify the web sessions by session ID and uri_path
  - Evaluate duration of each transaction dynamically by service
  - Find transactions where duration is > historical average + 2 standard deviations

Want to know more? Check out:
Splunk Docs (streamstats): https://docs.splunk.com/Documentation/Splunk/9.3.0/SearchReference/Streamstats

LAB EXERCISE 7 15 MINS

Putting Everything to Practice!

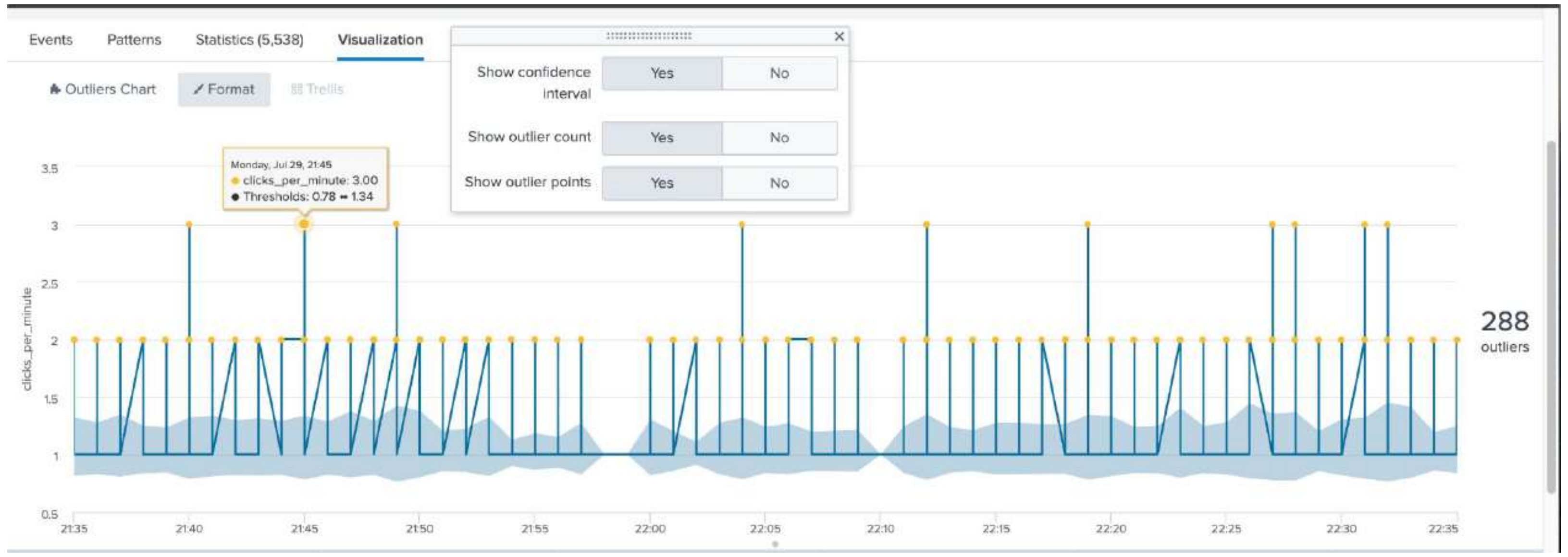
Use the commands and functions discussed in this section to find suspicious activity in your event logs that may indicate the presence of bots!

Overview of the **commands** and **functions to be used:**

- stats stdev() function
- eventstats command
- streamstats command

Syntax:

```
<your search> | command function(<field>) <by clause>
```



Show confidence interval: [✓ Yes] No

Show outlier count: [✓ Yes] No

Show outlier points: [✓ Yes] No

Monday, Jul 29, 21:45
● clicks_per_minute: 3.00
● Thresholds: 0.78 – 1.34

Exercise #7: Create an SPL query that can help us detect bots in our web logs

- Hints:
  - Focus on web log events (sourcetype clue)
  - Group events in 1 minute intervals (bin)
  - Count rolling number of events as clicks per minute, look at the values for each time interval and username
  - Find when the number of clicks per minute has a very low standard deviation! (constant clicks = bots)
  - Sort results in ascending order based on the standard deviation of the clicks per minute field

Want to know more? Check out:
Splunk Docs (stats stdev): https://docs.splunk.com/Documentation/Splunk/9.2.2/SearchReference/Aggregatefunctions#stdev.28.26lt.3Bvalue.26gt.3B.29

Schema-on-Read & Field Extractions

Extracting new fields, eventtypes, tags, macros, & the Splunk CIM

splunk>

Schema-on-read vs Schema-on-write

Splunk has many options for storing and searching data

Schema-on-read / search-time extractions:

- We only write _raw and metadata to disk.
- Extract new fields when you run a search.
- This makes ingest very fast, search very flexible.
- Good for asking questions you didn’t know in advance (i.e., during security investigations!!!)

Schema-on-write / index-time extractions:

- We write additional field information to disk at ingest-time (e.g., hash, username, source/dest IPs).
- This adds time to ingest and uses extra disk space, but enables fast, optimized searches.
- Good for asking questions you know you need to ask (i.e., for security monitoring).

Data model acceleration:

- We can write additional field information after the fact.
- Acceleration searches run in background on regular basis.

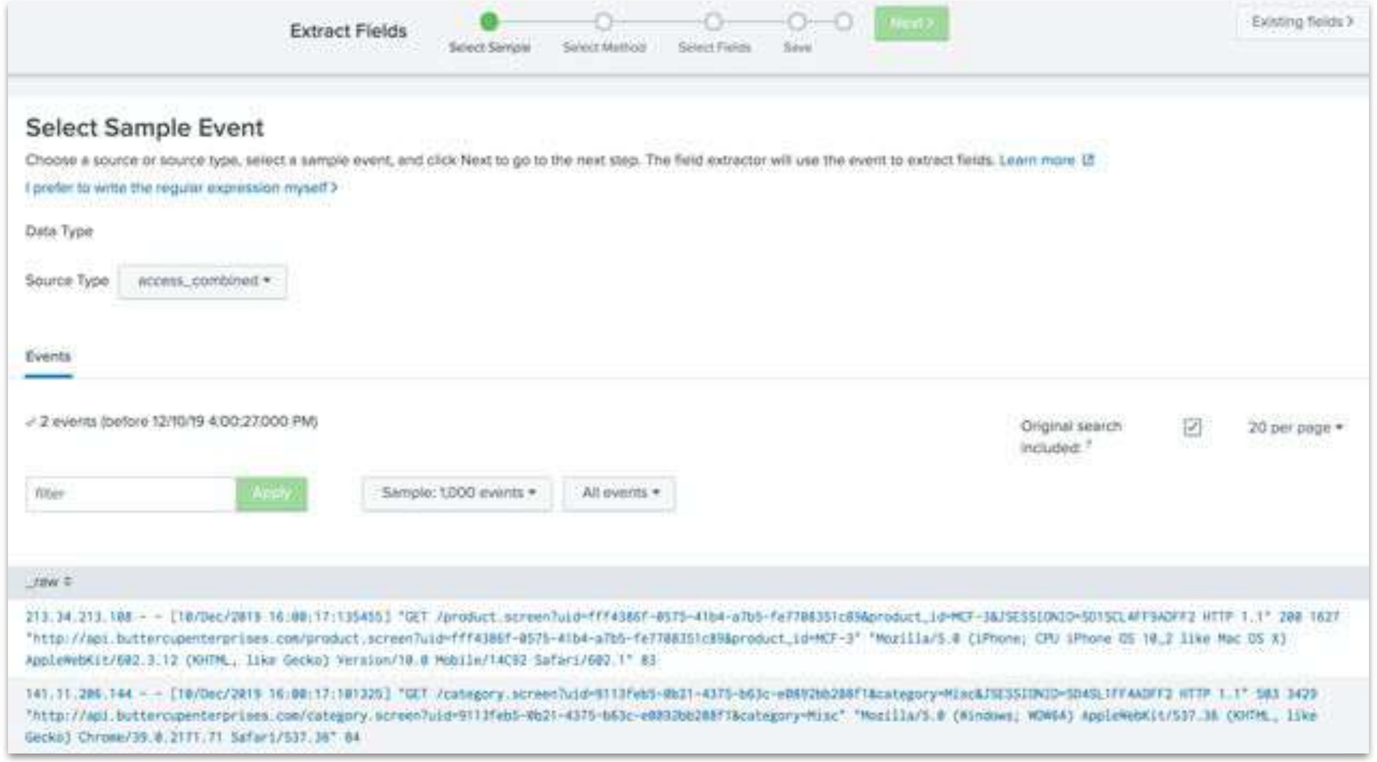
Interactive Field Extractor

Extract fields with UI-based Interactive Field Extractor - it creates regex for you!

Best practice: Extract ALL fields you need using single regex

Bad practice: A customer who used 72 different field extractions for 4 different types of data grouped into one sourcetype.

- **How to fix:** Separate into 4 different sourcetypes, each with 1-3 field extractions covering the ~18 fields on each sourcetype.



Select Sample Event

Choose a source or source type, select a sample event, and click Next to go to the next step. The field extractor will use the event to extract fields. Learn more

I prefer to write the regular expression myself >

✓ 2 events (before 12/10/19 4:00:27.000 PM)

```
213.34.213.108 - - [10/Dec/2019 16:00:17:135455] "GET /product.screen?uid=fff4386f-0575-41b4-a7b5-fe7708351c89&product_id=MCF-3&JSESSIONID=SD1SCL4FF9ADFF2 HTTP 1.1" 200 1627
"http://api.buttercupenterprises.com/product.screen?uid=fff4386f-0575-41b4-a7b5-fe7708351c89&product_id=MCF-3" "Mozilla/5.0 (iPhone; CPU iPhone OS 10_2 like Mac OS X)
AppleWebKit/602.3.12 (KHTML, like Gecko) Version/10.0 Mobile/14C92 Safari/602.1" 83
141.11.206.144 - - [10/Dec/2019 16:00:17:101326] "GET /category.screen?uid=9113feb5-0b21-4375-b63c-e0892bb208f1&category=Misc&JSESSIONID=SD4SL1FFAADFF2 HTTP 1.1" 503 3429
"http://api.buttercupenterprises.com/category.screen?uid=9113feb5-0b21-4375-b63c-e0892bb208f1&category=Misc" "Mozilla/5.0 (Windows; WOW64) AppleWebKit/537.36 (KHTML, like
Gecko) Chrome/39.0.2171.71 Safari/537.36" 84
```

Work with Splunk admin teams on field extractions, but the Interactive Field Extractor is good for a quick one-off

Eventtypes, tags and macros

We can *"hide"* field extractions under eventtypes and tags.

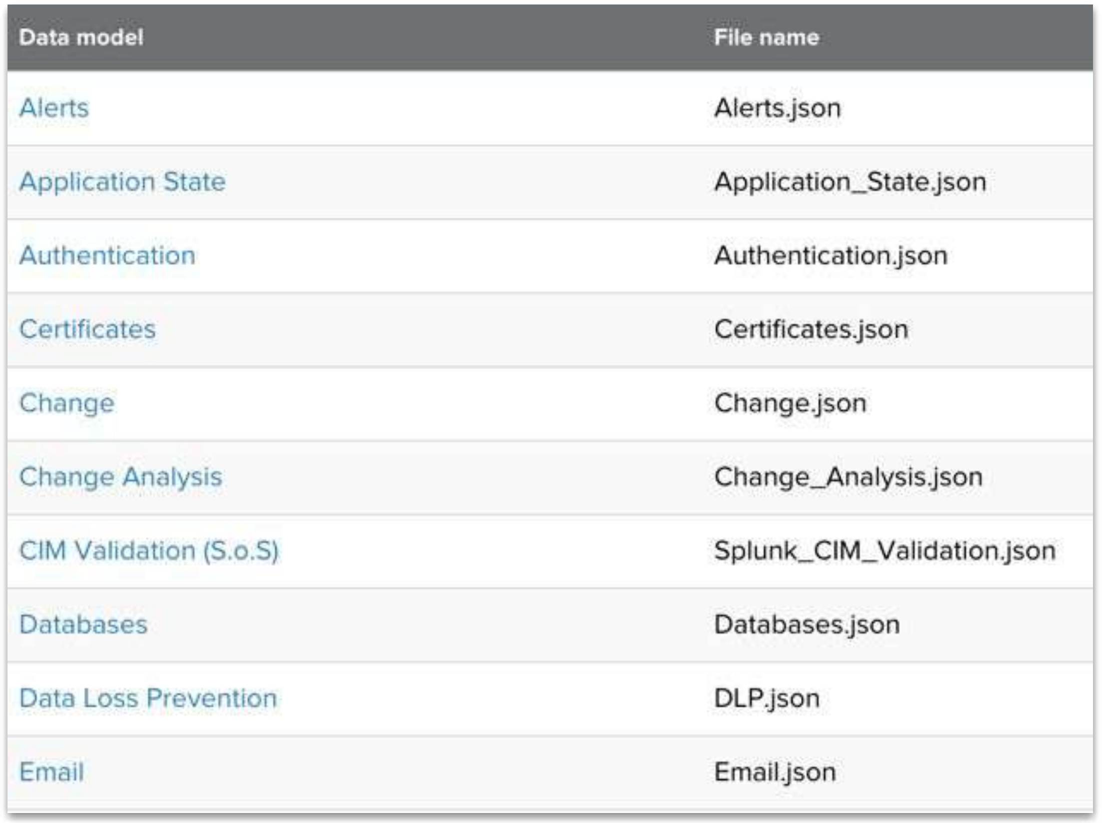
- **Eventtypes:** `Eventtype=myfavoritedata` => `(sourcetype=X type=foo) OR (sourcetype=Y style=bar)`
- **Tags:** tag=network as part of Common Information Model hides LOTS of contributing information to make it easy for you to search for network data
- **Macros:** store reusable search strings.
  - **Example:** wrap string "lookup assets.csv host" as macro \`asset_enrich\`

```
sourcetype=X | `asset_enrich` => sourcetype=X | lookup assets.csv host
```

  - You can pass variables (\`mymacro(X)\` => something)

Best practice: Work with your Splunk admin team!

Common Information Model



Data model	File name
Alerts	Alerts.json
Application State	Application_State.json
Authentication	Authentication.json
Certificates	Certificates.json
Change	Change.json
Change Analysis	Change_Analysis.json
CIM Validation (S.o.S)	Splunk_CIM_Validation.json
Databases	Databases.json
Data Loss Prevention	DLP.json
Email	Email.json

Use field aliases and tags to harmonize data across different types

Example:

- Field name "src" to appear on all network data
- Sometimes represents "source", "source_ip", "starting_ip", etc.
- Using the Splunk CIM you can search on src=1.2.3.4 across all data.

Pro Tip: Talk to your Splunk admin team!

- If working with common data types, they're probably already aligned to CIM.
- Security teams should ***definitely*** align to CIM.

Learn more about the Common Information Model:
https://docs.splunk.com/Documentation/CIM/latest/User/Overview

Validate CIM compliance:
https://github.com/hire-vladimir/SA-cim_vladiator

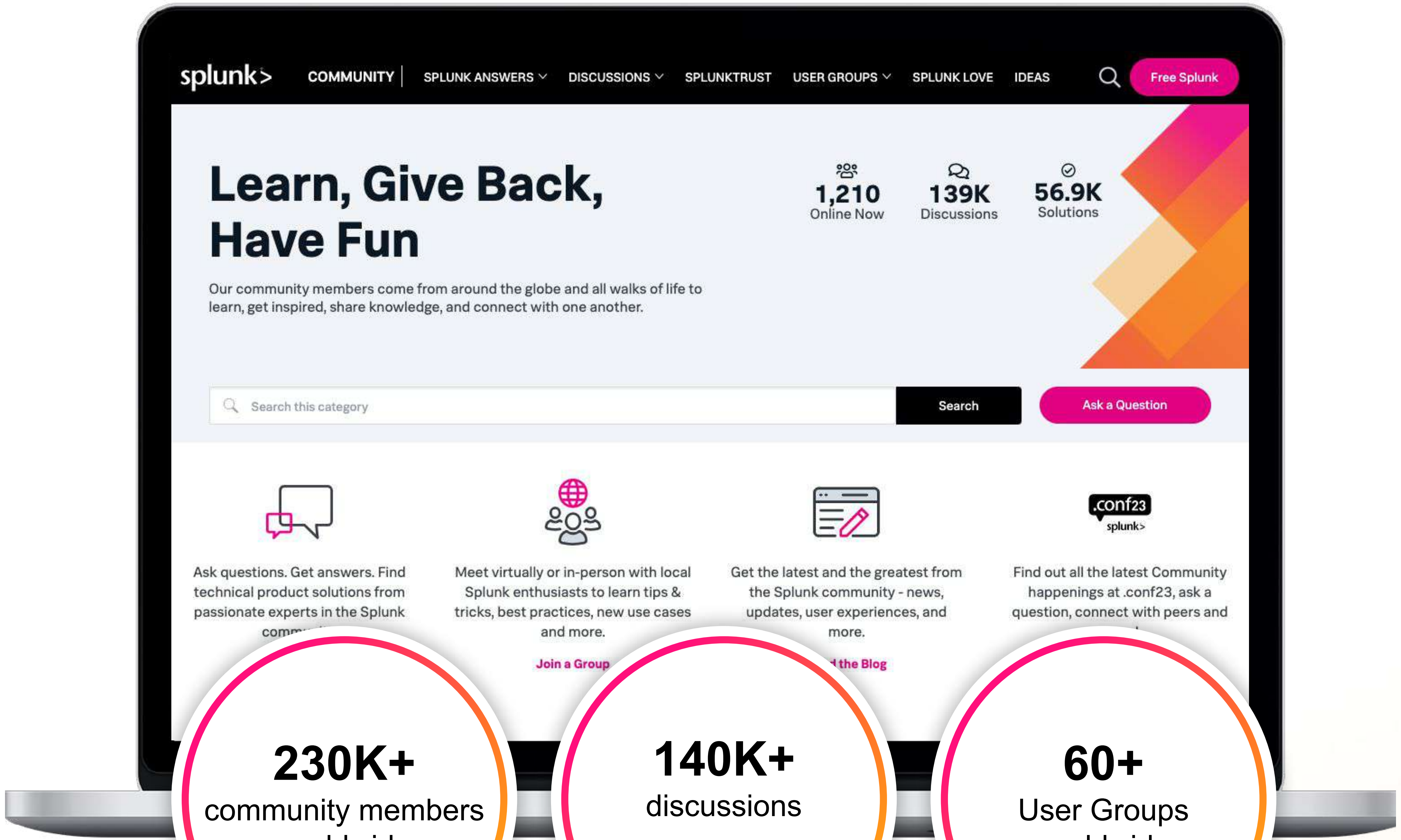
Resources

Where to go from here?

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



Learn, Give Back, Have Fun

1,210 Online Now

139K Discussions

56.9K Solutions

Our community members come from around the globe and all walks of life to learn, get inspired, share knowledge, and connect with one another.

Ask questions. Get answers. Find technical product solutions from passionate experts in the Splunk comm

Meet virtually or in-person with local Splunk enthusiasts to learn tips & tricks, best practices, new use cases and more.

Join a Group

Get the latest and the greatest from the Splunk community - news, updates, user experiences, and more.

the Blog

Find out all the latest Community happenings at .conf23, ask a question, connect with peers and

230K+
community members worldwide

140K+
discussions

60+
User Groups worldwide

Splunk Events

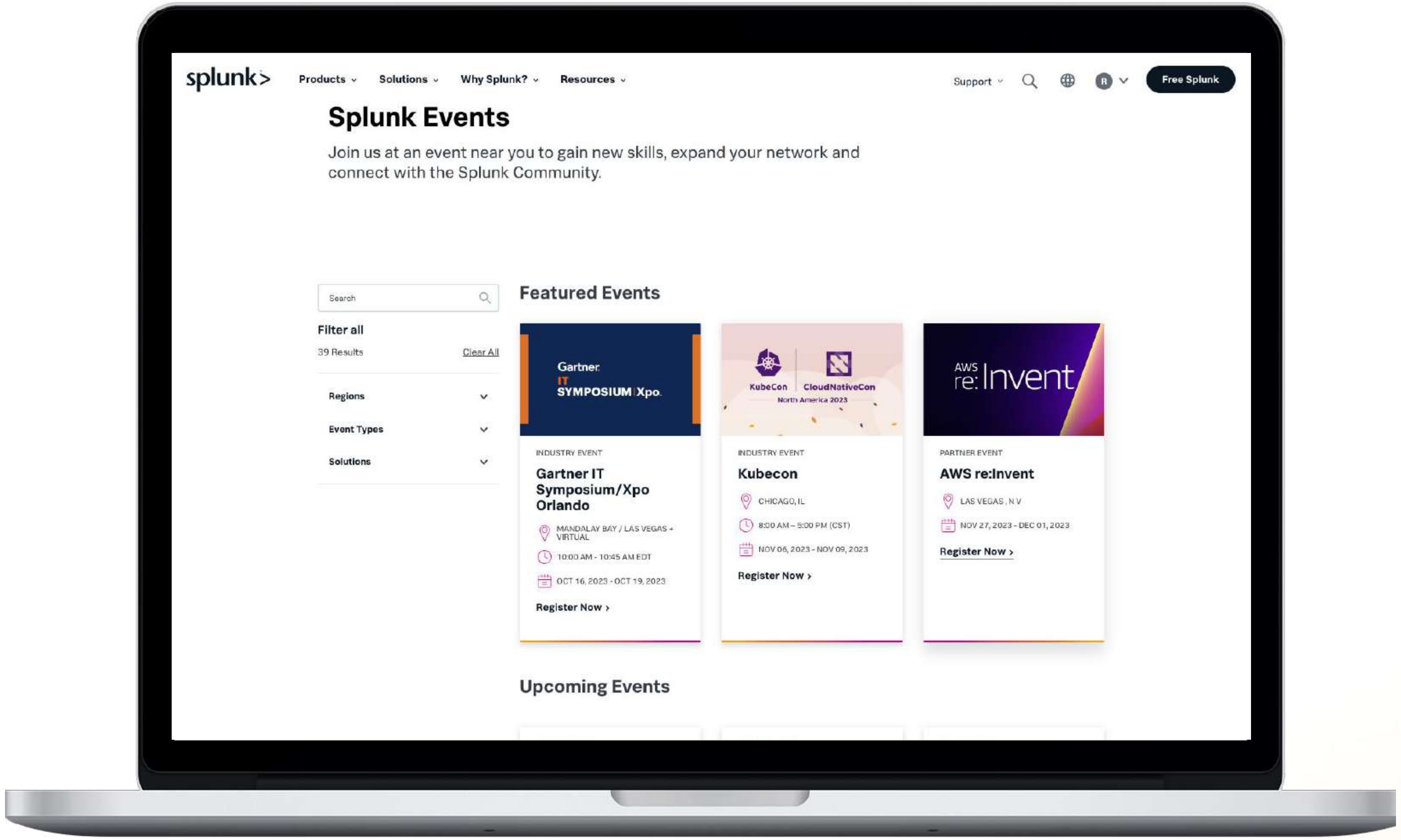
https://splunk.com/events

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

.conf

https://conf.splunk.com

- Join us at .conf next summer!
- Hundreds of on-demand sessions from product updates to learning new Splunk skills!



Splunk Events

Join us at an event near you to gain new skills, expand your network and connect with the Splunk Community.

Filter all

39 Results Clear All

Featured Events

INDUSTRY EVENT
Gartner IT Symposium/Xpo Orlando
MANDALAY BAY / LAS VEGAS + VIRTUAL
10:00 AM - 10:45 AM EDT
OCT 16, 2023 - OCT 19, 2023
Register Now ›

INDUSTRY EVENT
Kubecon
CHICAGO, IL
8:00 AM – 5:00 PM (CST)
NOV 06, 2023 - NOV 09, 2023
Register Now ›

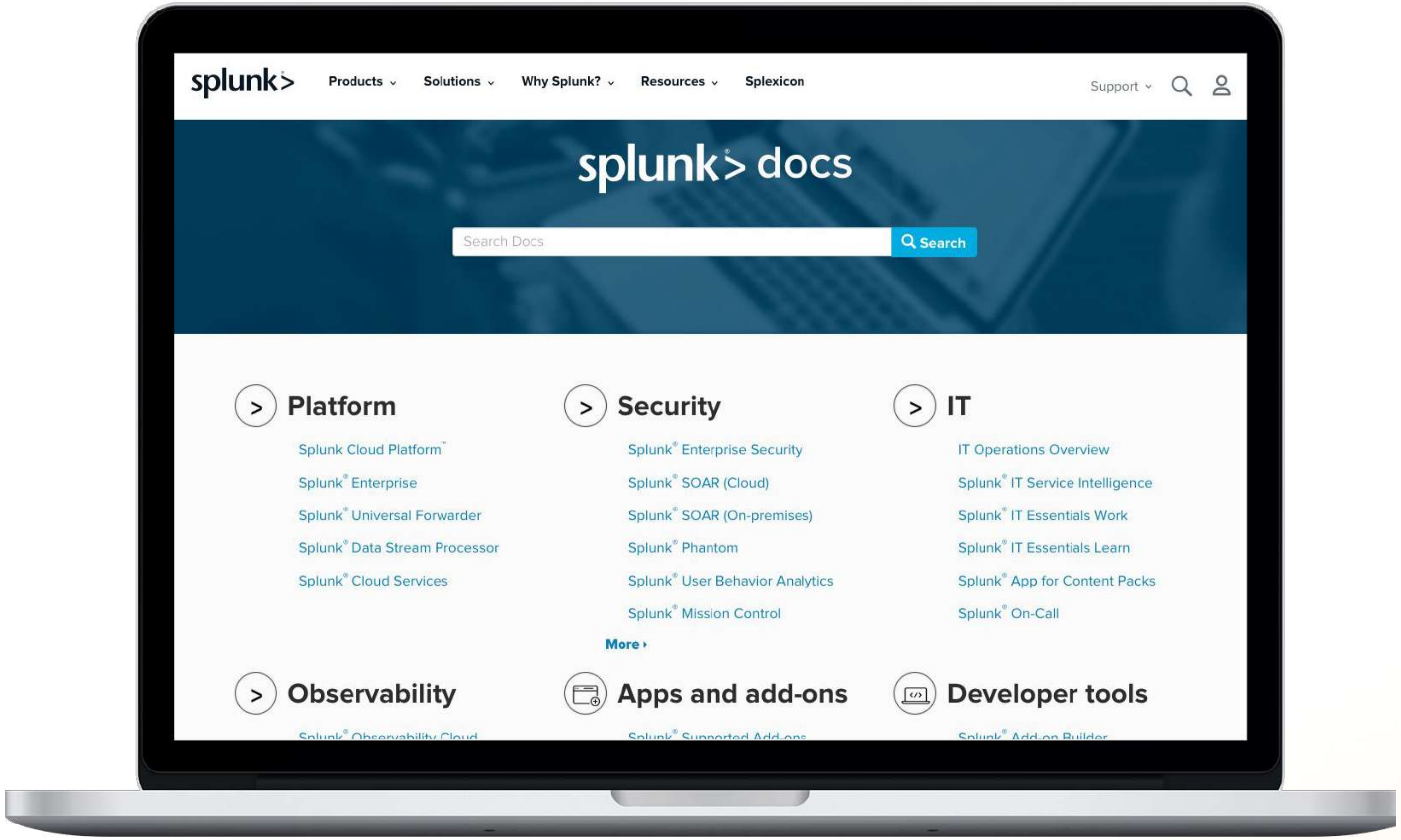
PARTNER EVENT
AWS re:Invent
LAS VEGAS , N V
NOV 27, 2023 - DEC 01, 2023
Register Now ›

Upcoming Events

Documentation

https://docs.splunk.com

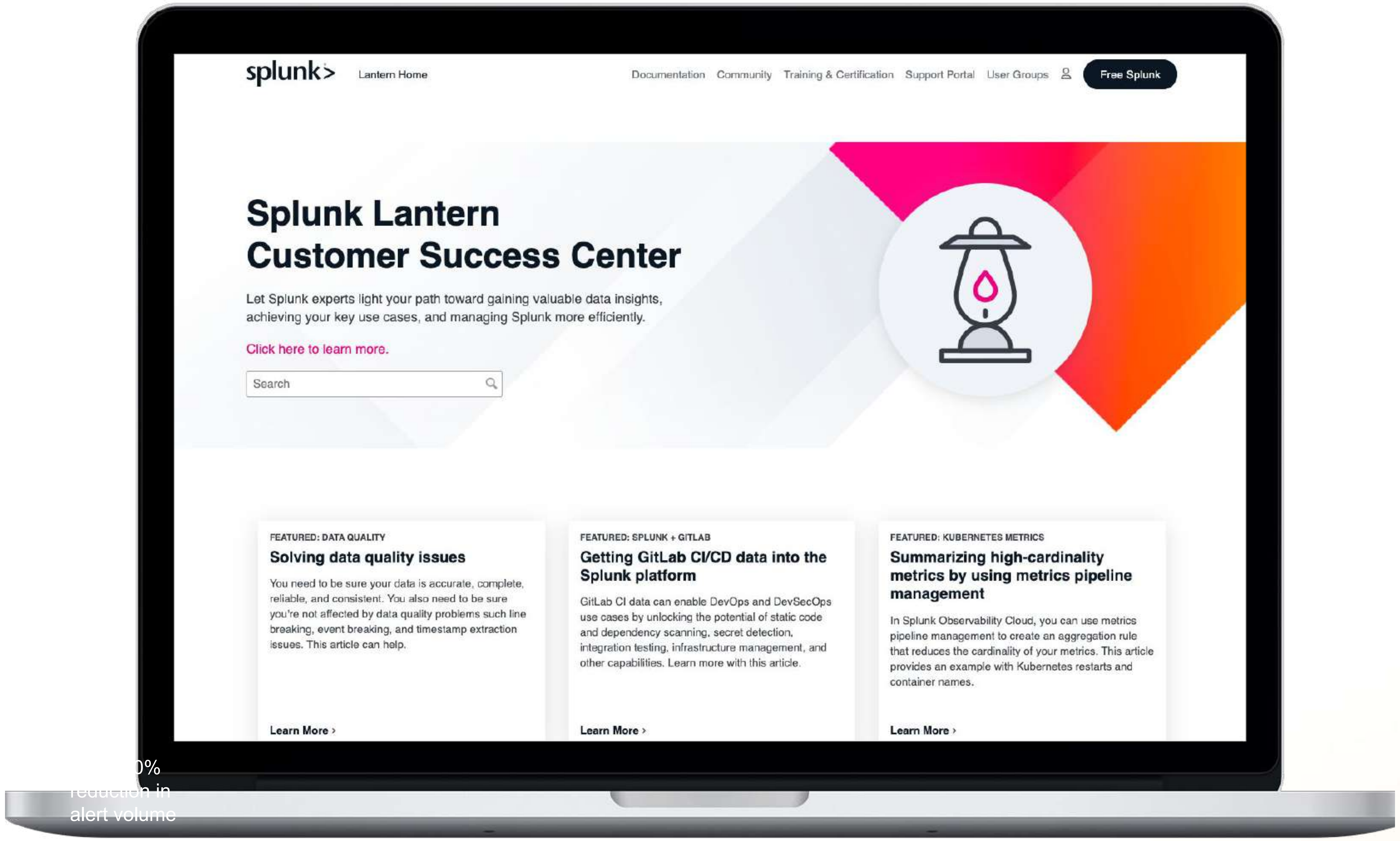
- Search reference for SPL
- Step-by-step tutorials
  Search:
  https://splk.it/SplunkSearchTutorial
  Dashboard Studio:
  https://splk.it/SplunkDashStudioTutorial
- Product references
- Procedures/guides
- And more!



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



Splunk Lantern Customer Success Center

Let Splunk experts light your path toward gaining valuable data insights, achieving your key use cases, and managing Splunk more efficiently.

Click here to learn more.

FEATURED: DATA QUALITY

Solving data quality issues

You need to be sure your data is accurate, complete, reliable, and consistent. You also need to be sure you're not affected by data quality problems such line breaking, event breaking, and timestamp extraction issues. This article can help.

Learn More ›

FEATURED: SPLUNK + GITLAB

Getting GitLab CI/CD data into the Splunk platform

GitLab CI data can enable DevOps and DevSecOps use cases by unlocking the potential of static code and dependency scanning, secret detection, integration testing, infrastructure management, and other capabilities. Learn more with this article.

Learn More ›

FEATURED: KUBERNETES METRICS

Summarizing high-cardinality metrics by using metrics pipeline management

In Splunk Observability Cloud, you can use metrics pipeline management to create an aggregation rule that reduces the cardinality of your metrics. This article provides an example with Kubernetes restarts and container names.

Learn More ›

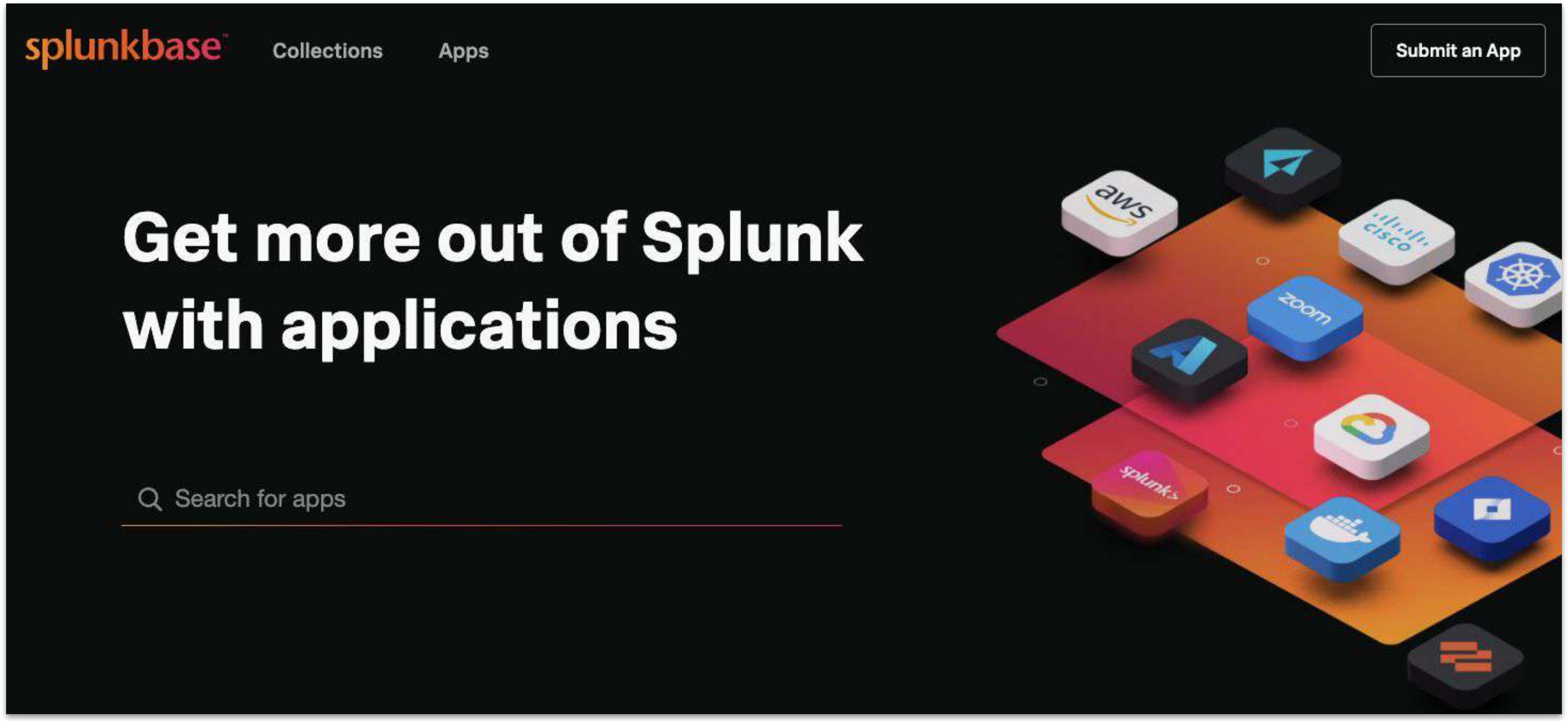
0%
reduction in
alert volume

Splunkbase

Splunkbase has 3200+ apps

Enable new use cases and extend your teams' Splunk capabilities:

https://splunkbase.splunk.com



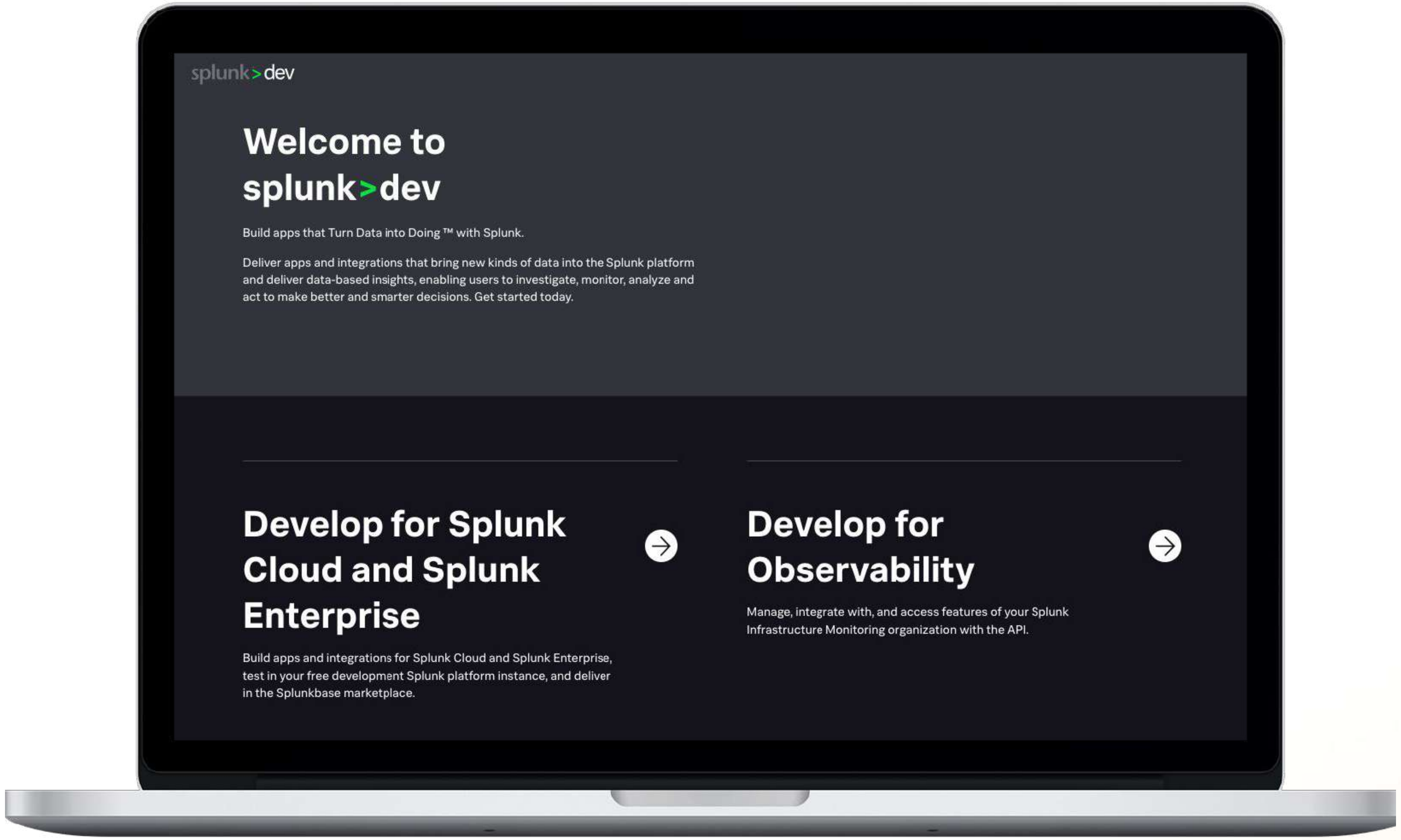
Get more out of Splunk with applications

- 3200+ apps and add-ons
- Pre-built searches, reports, visualisations and integrations for specific use cases and technologies
  - Use cases include:
    - IT Ops, Security, Observability, Business Analytics, IoT & Industrial Data
    - Financial Services, Retail, Telecom, Healthcare, Energy
- Download apps and customize them based on your requirements
- Fast time to value from your data
- Build and contribute your own apps!

Developer Resources

https://dev.splunk.com

- Developer Guide
- API Reference
- Tutorials
- Downloads
  APIs, libraries, tools
- Code examples
- Free Developer licence



Welcome to splunk>dev

Build apps that Turn Data into Doing ™ with Splunk.

Deliver apps and integrations that bring new kinds of data into the Splunk platform and deliver data-based insights, enabling users to investigate, monitor, analyze and act to make better and smarter decisions. Get started today.

Develop for Splunk Cloud and Splunk Enterprise

Build apps and integrations for Splunk Cloud and Splunk Enterprise, test in your free development Splunk platform instance, and deliver in the Splunkbase marketplace.

Develop for Observability

Manage, integrate with, and access features of your Splunk Infrastructure Monitoring organization with the API.

Training & Certification

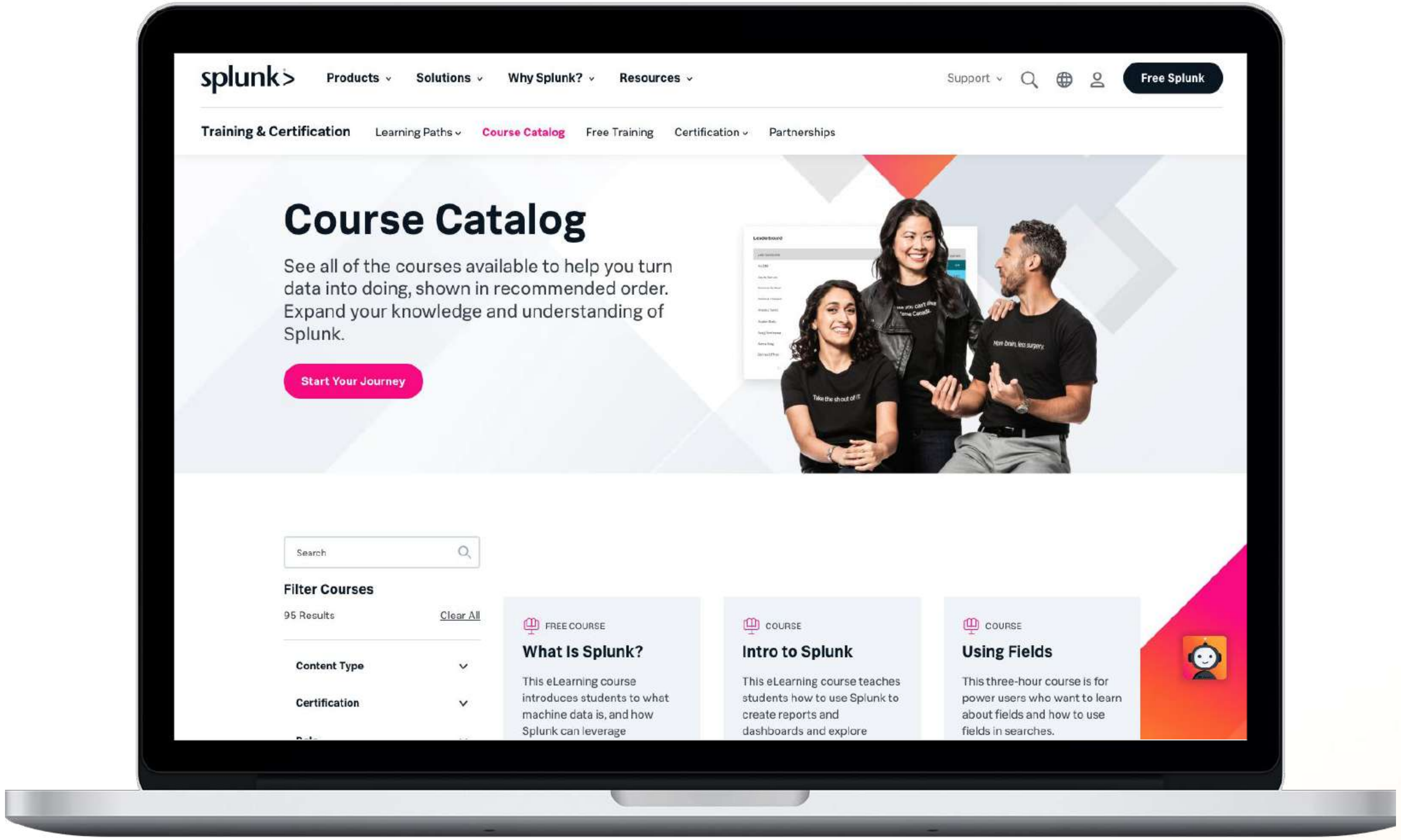
https://splunk.com/training

- **Online education classes**
  Instructor-led and self-paced eLearning

- **Certification tracks for different roles**
  User, Power User, Admin, Architect and Developer

- **Splunk Education Rewards**
  Complete training and receive points that you can redeem for Splunk swag!

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



Course Catalog

See all of the courses available to help you turn data into doing, shown in recommended order. Expand your knowledge and understanding of Splunk.

Filter Courses

95 Results Clear All

FREE COURSE
What Is Splunk?
This eLearning course introduces students to what machine data is, and how Splunk can leverage

COURSE
Intro to Splunk
This eLearning course teaches students how to use Splunk to create reports and dashboards and explore

COURSE
Using Fields
This three-hour course is for power users who want to learn about fields and how to use fields in searches.

Thank you!

splunk>