© 2025 SPLUNK LLC

Splunk4Ninjas - Machine Learning for Security

Hands-On Workshop

splunk> a CISCO company

© 2025 SPLUNK LLC

Forward-looking statements

This presentation may contain forward-looking statements that are subject to the safe harbors created under the Securities Act of 1933, as amended, and the Securities Exchange Act of 1934, as amended. All statements other than statements of historical facts are statements that could be deemed forward-looking statements. These statements are based on current expectations, estimates, forecasts, and projections about the industries in which we operate and the beliefs and assumptions of our management based on the information currently available to us. Words such as “expects,” “anticipates,” “targets,” “goals,” “projects,” “intends,” “plans,” “believes,” “momentum,” “seeks,” “estimates,” “continues,” “endeavors,” “strives,” “may,” variations of such words, and similar expressions are intended to identify such forward-looking statements. In addition, any statements that refer to (1) our goals, commitments, and programs; (2) our business plans, initiatives, and objectives; and (3) our assumptions and expectations, including our expectations regarding our financial performance, products, technology, strategy, customers, markets, acquisitions and investments are forward-looking statements. 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. Readers are cautioned that these forward-looking statements are only predictions and are subject to risks, uncertainties, and assumptions that are difficult to predict, including those identified in 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 report on Form 10-Q filed with the SEC on November 28, 2023. 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 Cisco or Splunk, on Cisco or Splunk’s website or otherwise, it may not contain current or accurate information. Cisco and Splunk undertake no obligation to revise or update any forward-looking statements for any reason, except as required by law.

In addition, any information about new products, features, functionality or 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 or be relied upon in making a purchasing decision. We undertake no commitment, promise or 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. The development, release, and timing of any features or functionality described for our products remains at our sole discretion.

Splunk, Splunk> and Turn Data Into Doing are trademarks and registered trademarks of Splunk LLC in the United States and other countries. All other brand names, product names or trademarks belong to their respective owners.

© 2025 Splunk LLC. All rights reserved.

splunk> a CISCO company

Please introduce yourself!

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

© 2025 SPLUNK LLC

Workshop Agenda

- What is Machine Learning?
- Why do we need Machine Learning in security?
- Machine learning for security with Splunk
- Time for practice!
- Ready made applications leveraging ML
- What’s next?

© 2025 SPLUNK LLC

Enroll in Today's Workshop

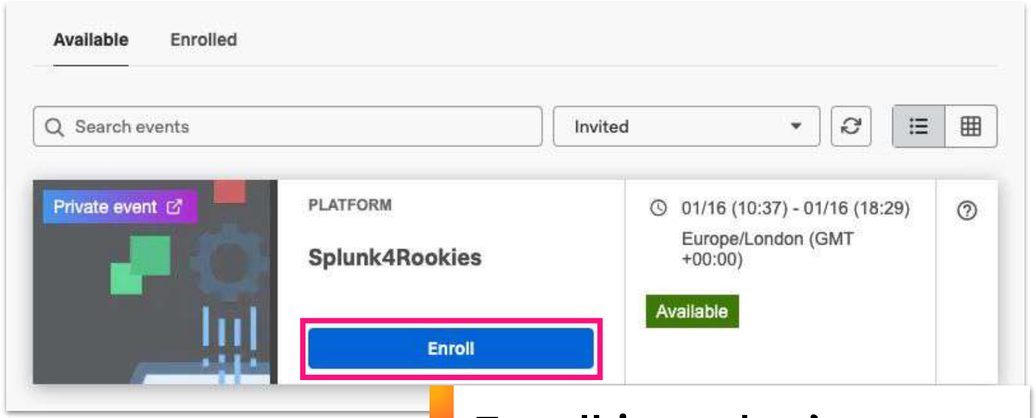
Tasks

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 the hands-on lab guide:
   https://splk.it/ML-Sec-Lab-Guide

   > Contains step-by-step instructions for all of today's exercises!

1. Download a copy of today's slide deck:
   https://splk.it/ML-Sec-Attendee

Goal



Enroll in today's event

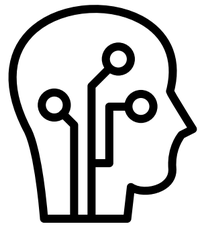
© 2025 SPLUNK LLC

What is Machine Learning?

© 2025 SPLUNK LLC

What is AI and Machine Learning?

Definitions



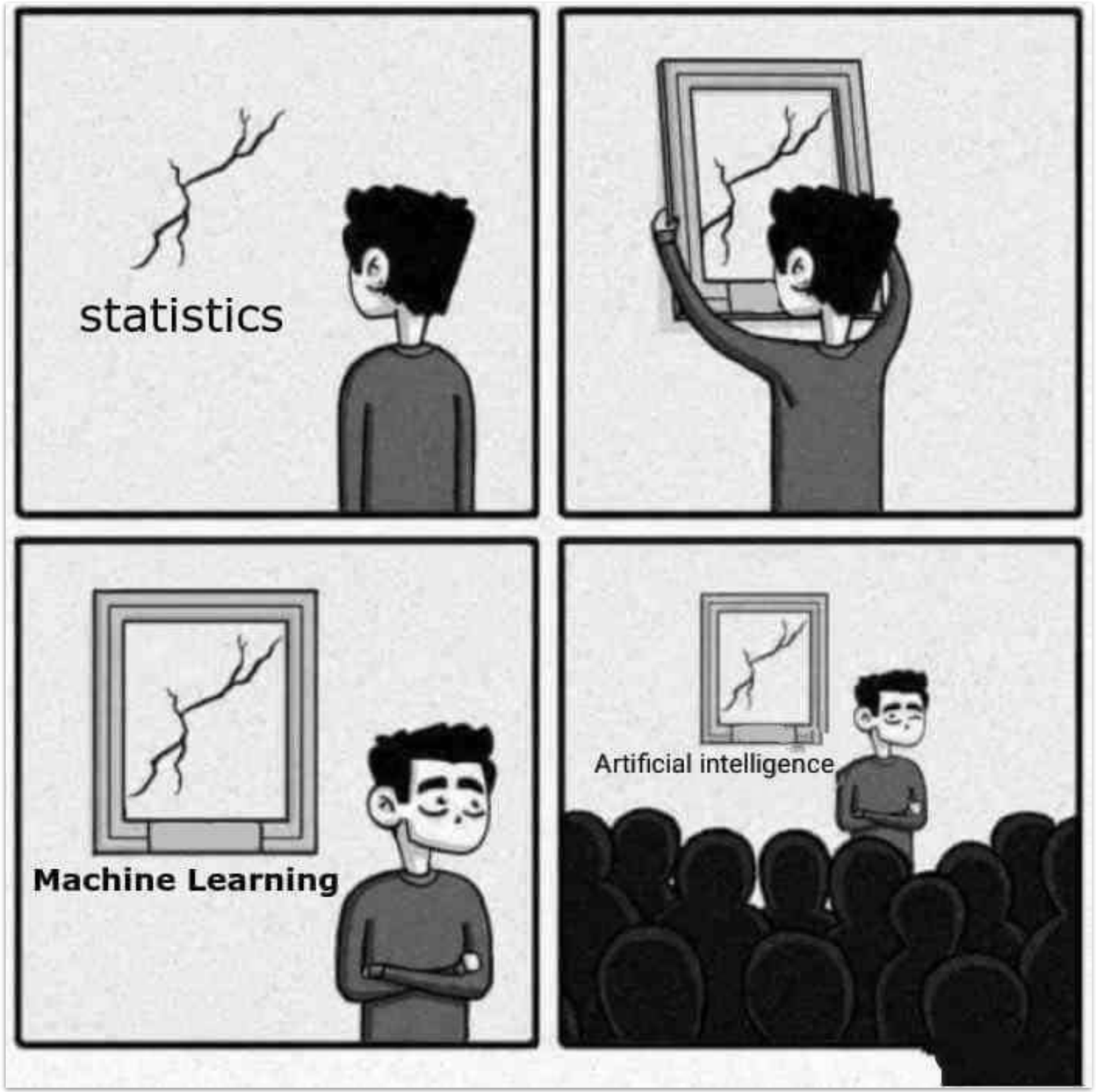
Artificial Intelligence (AI) - capability of a computer system to **mimic human cognitive functions** such as learning and problem-solving

Machine Learning (ML) - **subset of AI** that uses mathematical models of data to help a computer learn without direct instruction

Deep Learning - **subset of AI** that uses computationally intense ML models inspired by the "deep" layers of the biological neural network of the human brain to accomplish complex goals like image recognition *Example: Self driving car recognizes stop sign*

Generative AI - **subset of AI** that involves the use of algorithms and techniques to generate new data, things that have not existed in the world before being created by the models *Example: OpenAI's ChatGPT*

© 2025 SPLUNK LLC



© Thomas Wiecki, Ph.D., ODSC Europe 2018

© 2025 SPLUNK LLC

“

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and past behaviour.

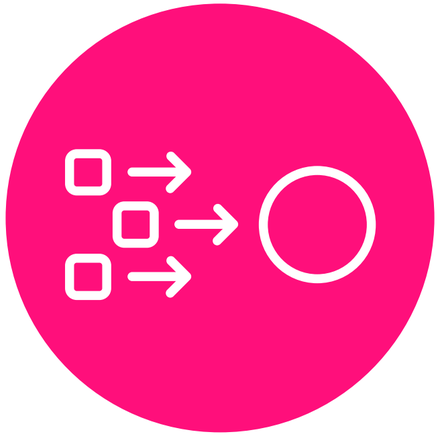
- Wikipedia

© Tim Whyatt 2002-2023

© 2025 SPLUNK LLC

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

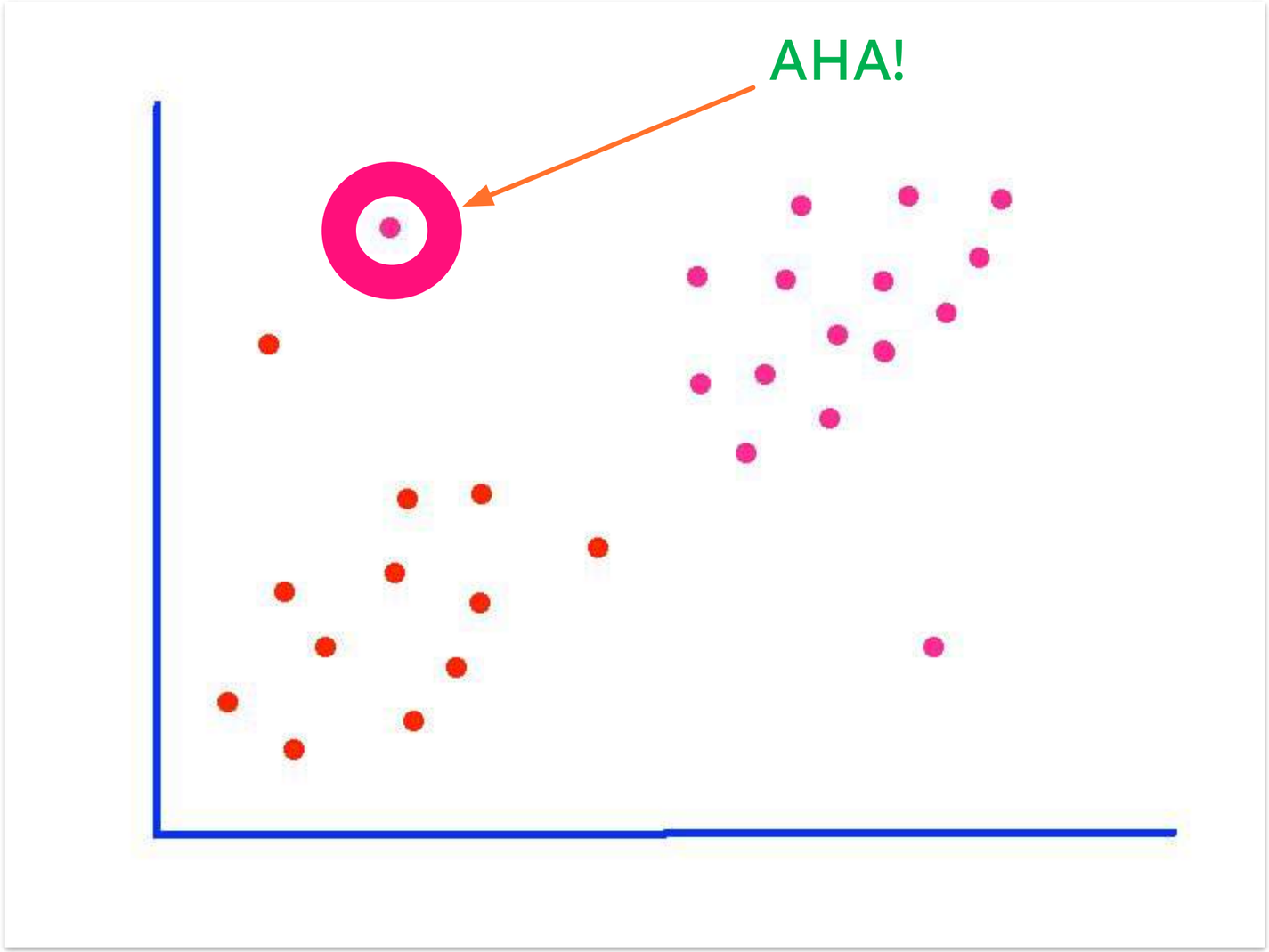
Catalog the patterns and in some cases iterate them as new data is received

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

© 2025 SPLUNK LLC

Correlation

This is NOT ML



© 2025 SPLUNK LLC

Machine Learning

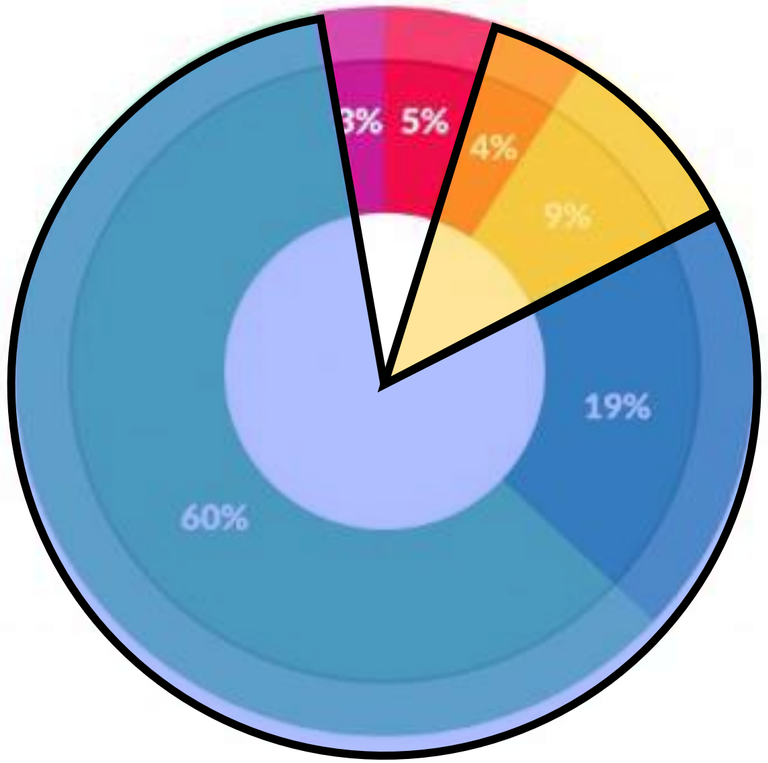
Why do we need it?

- Basic statistical modelling techniques (such as prediction using means and standard deviations) are not always fine grained enough to answer these questions accurately
- Machine Learning (ML) uses complex statistical techniques to provide better predictions to these types of questions
- For many organizations using their data to make pre-emptive or targeted decisions is becoming increasingly important, for example:
  - Detect the presence of Botnets in their network
  - Detecting a user account that has been compromised
  - Preventing fraud in banking by detecting malicious accounts or anomalous activity
  - Classifying customers for more targeted advertising campaigns in social media
  - Forecasting demand for products in support of supply chain management

© 2025 SPLUNK LLC

What Data Scientists Really Do

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



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

"Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", Forbes Mar 23, 2016

© 2025 SPLUNK LLC

ML Starter

Dos and Don’ts

What Machine Learning isn’t…

- Something you can throw all your data at and expect an answer
- A one size fits all answer to all your data problems (a.k.a. a magic button)
- Artificial Intelligence…
- The predict command
- Correlation rules

What Machine Learning is…

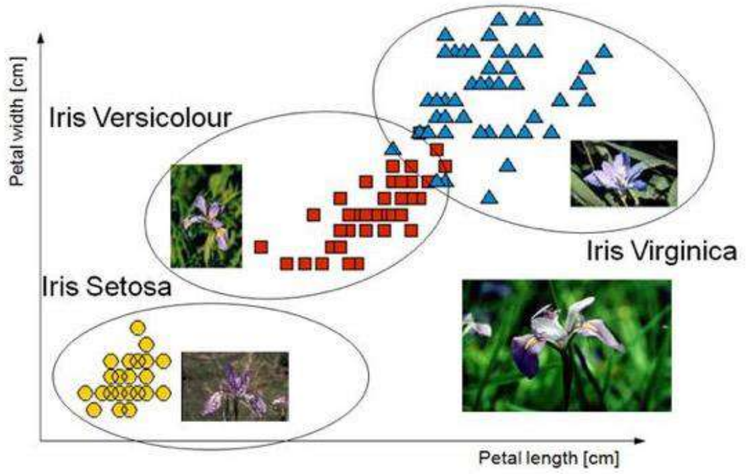
- A good way of spotting and baselining patterns in data for future detection
- Useful only when you have a well defined and bounded problem, and clearly understand the data that you need to use

© 2025 SPLUNK LLC

Types of Machine Learning

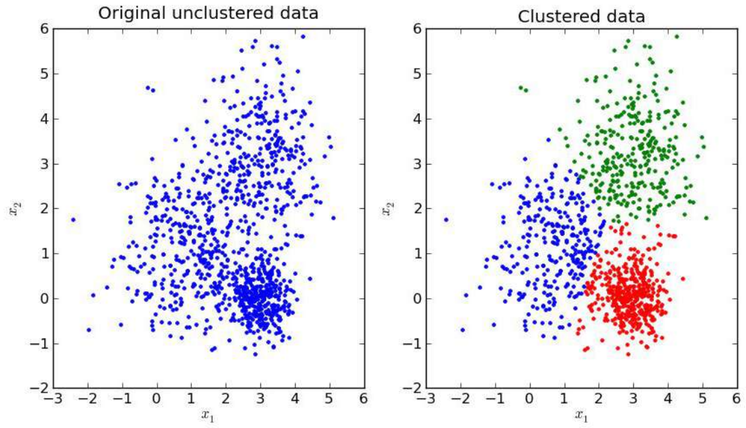
Supervised Learning (labeled data)

- Regression
- Classification



Unsupervised Learning (unlabeled data)

- Clustering
- Anomaly Detection



Mixed Models (with reinforcement or feedback)

- Human in the Loop
- Autonomous Systems

© 2025 SPLUNK LLC

Why Do We Need Machine Learning in Security?

© 2025 SPLUNK LLC

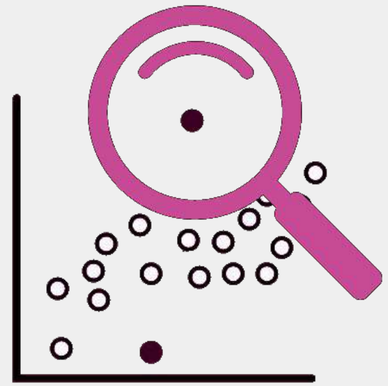
Shifting Focus and Role for SOCs

LEGACY	REQUIRED
Situational Awareness	Analysis and Decision-Making
Operation / Monitoring Center	Nerve Center / Command Center
Human Authored	Human — Machine Learning
Human Speed Operations	Machine-Speed Cycle Times

© 2025 SPLUNK LLC

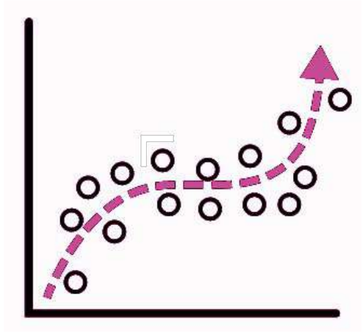
Splunk Customers Want Answers from their Data

Anomaly detection



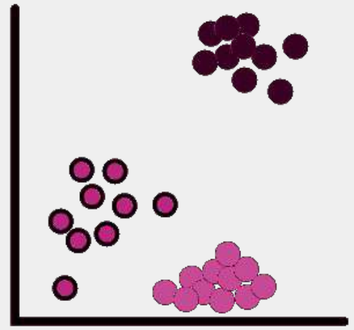
- Deviation from past behavior
- Deviation from peers
- (aka Multivariate AD or Cohesive AD)
- Unusual change in features

Predictive Analytics



- Predict Service Health Score/Churn
- Predicting Events
- Trend Forecasting
- Detecting influencing entities
- Early warning of failure

Clustering

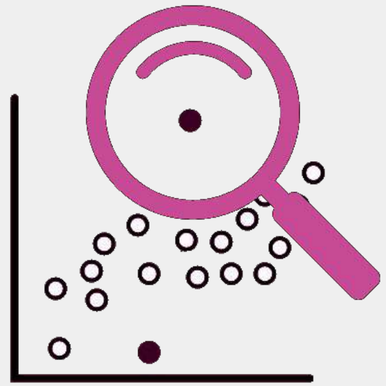


- Identify peer groups
- Event Correlation
- Reduce alert noise
- Behavioral Analytics

© 2025 SPLUNK LLC

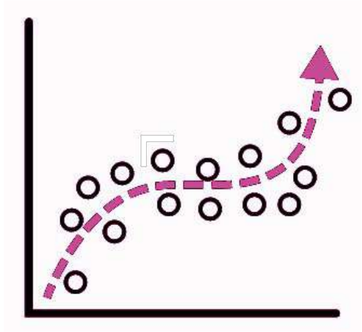
Security Examples

Anomaly detection



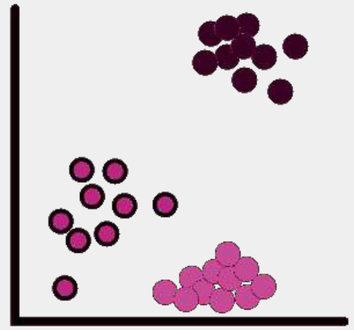
- Useful when trying to detect unwanted and unexpected behaviour.
- Useful when detecting low and slow attacks that operate over a long time.

Predictive Analytics



- Predict the outcome of an alert (FP/TP) based on observed previous responses.
- Generate response recommendations based on observed previous responses.

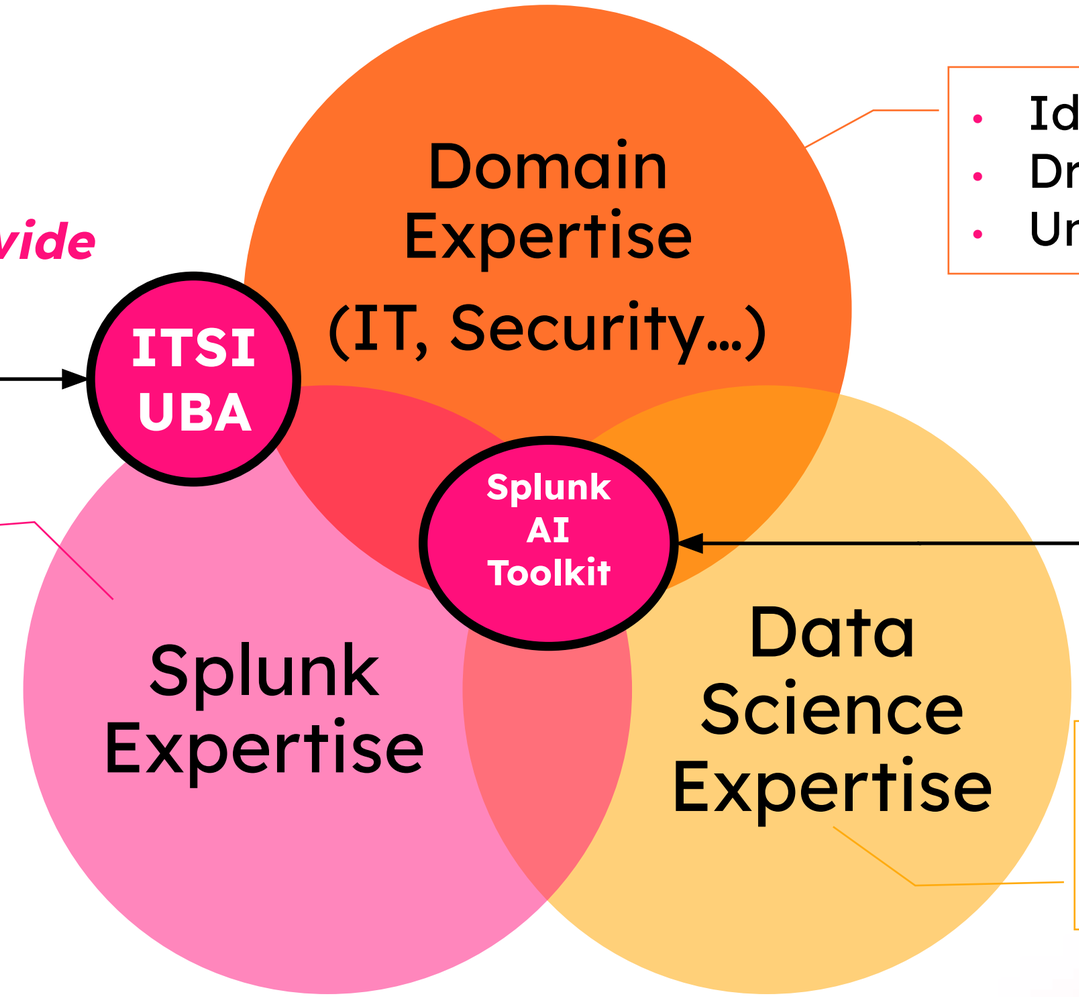
Clustering



- Used to generate dynamic peer groups based on observed events.
- Used to group similar alerts.

© 2025 SPLUNK LLC

Skill Areas for Machine Learning @ Splunk



Premium solutions provide out of the box ML capabilities.

- Identify use cases
- Drive decisions
- Understanding of business impact

- Searching
- Reporting
- Alerting
- Workflow

Splunk AI Toolkit facilitates and simplifies via examples & guidance

- Statistics/math background
- Algorithm selection
- Model building

© 2025 SPLUNK LLC

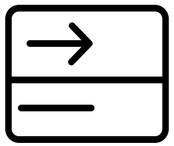
Custom ML with the Splunk Platform

Ecosystem — Splunk's App Ecosystem contains 1000's of free add-ons for getting data in, applying structure and visualizing your data giving you faster time to value.

AI Toolkit — The Machine Learning Toolkit delivers new SPL commands, custom visualizations, assistants, and examples to explore a variety of ml concepts.

Splunk — Splunk Enterprise is the mission-critical platform for indexing, searching, analyzing, alerting and visualizing machine data.

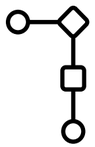
Operationalized Data Science Pipeline



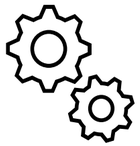
Collect Data

Clean & Munge

Search & Explore



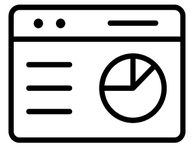
Pre-processing Feature Selection



Choose Algorithm

Build, Test, Improve Models

Operationalize Monitor Alert



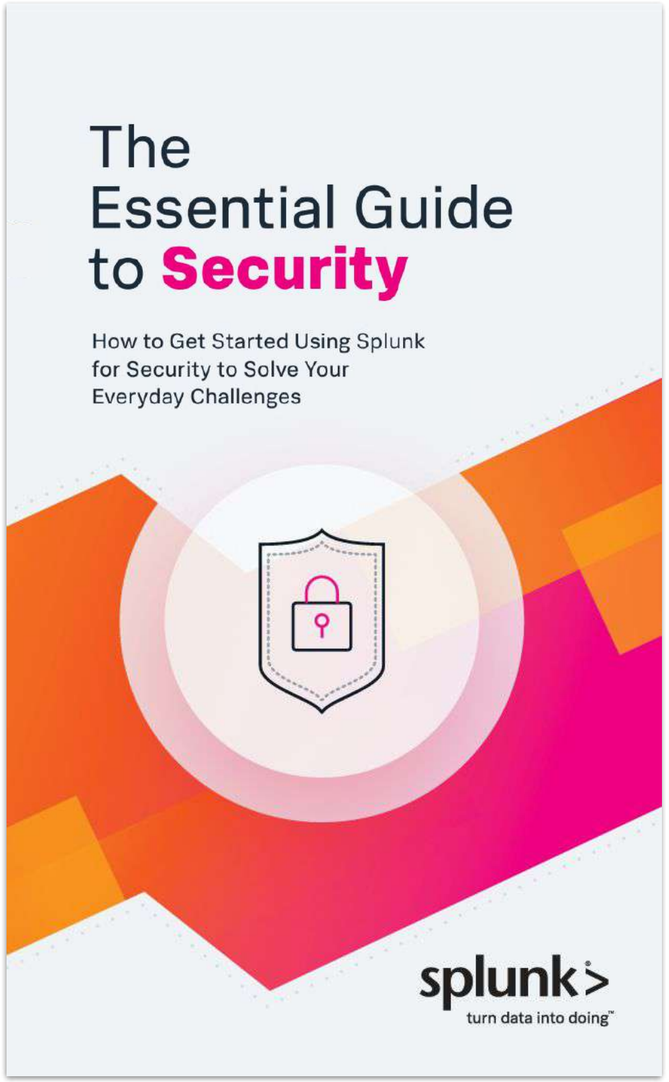
Visualize & Share

Ecosystem	Ecosystem	Splunk	AI Toolkit	AI Toolkit	AI Toolkit	AI Toolkit	Ecosystem
Splunk	Splunk		Splunk	Splunk	Splunk	Splunk	Splunk

splunk> Platform for Operational Intelligence

© 2025 SPLUNK LLC

The Security Data Journey



Advanced Detection
Apply sophisticated detection mechanisms including machine learning

Stage 6

Automation and Orchestration
Establish a consistent and repeatable security operation capability

Stage 5

Enrichment
Augment security data with intelligence sources to better understand the context and impact of an event

Stage 4

Expansion
Collect additional high fidelity data sources like endpoint activity and network metadata to drive advanced attack detection

Stage 3

Normalization
Apply a standard security taxonomy and add asset and identity data

Stage 2

Collection
Collect basic security logs and other machine data from your environment

Stage 1

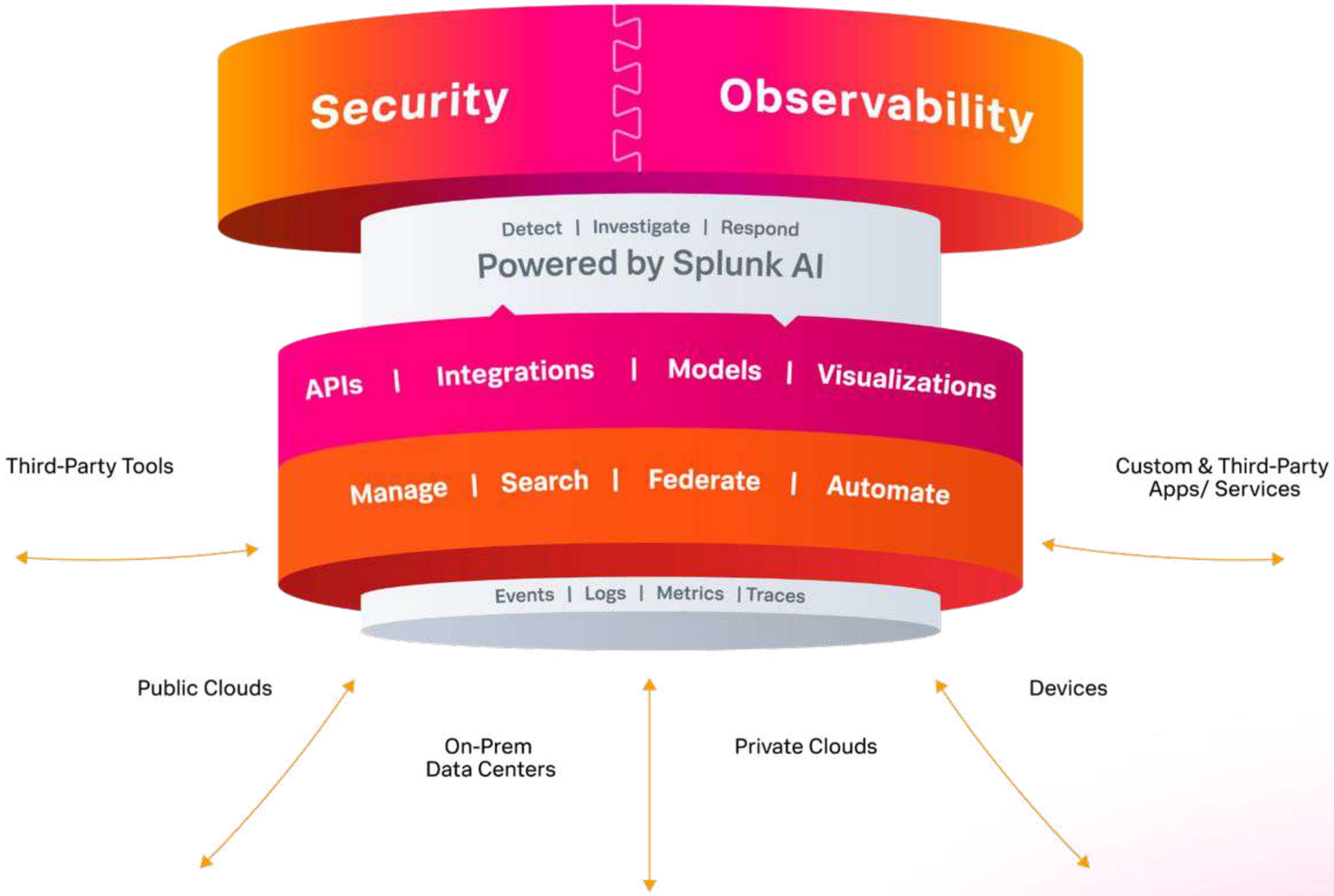
https://www.splunk.com/en_us/form/the-essential-guide-to-security.html

© 2025 SPLUNK LLC

Machine Learning for Security with Splunk

© 2025 SPLUNK LLC

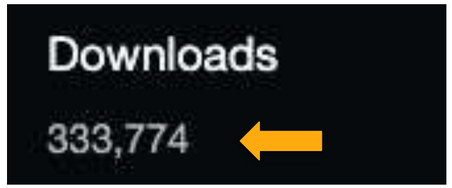
The Unified Security and Observability Platform is Powered by Splunk AI



© 2025 SPLUNK LLC

We have been implementing AI for years…

Splunk has long been committed to helping customers use AI



RESEARCH IN ACTION
Market Leader: Vendor Selection Matrix™ AIOps Solutions

GIGAOM
A Leader in the AIOps Radar for 4 consecutive years

OMDIA
Highest Scoring Product in Selecting an AIOps Solution for 2 years

© 2025 SPLUNK LLC

Domain-Specific Use Cases

Security

Simplify workflows

ML powered detections

Risk based alerting

Guided response actions

Event correlation and alert noise reduction

Predictive analytics

Anomaly detection

Clustering

Observability

Probable root cause analysis

Anomaly and outlier detection

Adaptive thresholding

Alert correlation and prioritization

Assisted remediation

Proactive outage prevention assistance

Suggested responders

© 2025 SPLUNK LLC

Powered by Splunk AI

Overview

AI for Security

Enterprise Security, UBA

ML-Powered Detections, Anomaly Detection, Predictive Analytics

AI for Observability

ITSI, APM, IM, On-Call

Predictive Analytics, Anomaly Detection, Adaptive Thresholding, Incident Correlation, Alert Noise Reduction, Alert Autodetect, Suggested Responders

Assistive Intelligence Experiences

Splunk AI Assistant,
Splunk AI Assistant in Observability/Security

Customizable ML

Splunk AI Toolkit, Splunk App for Data Science and Deep Learning, Python for Scientific Computing

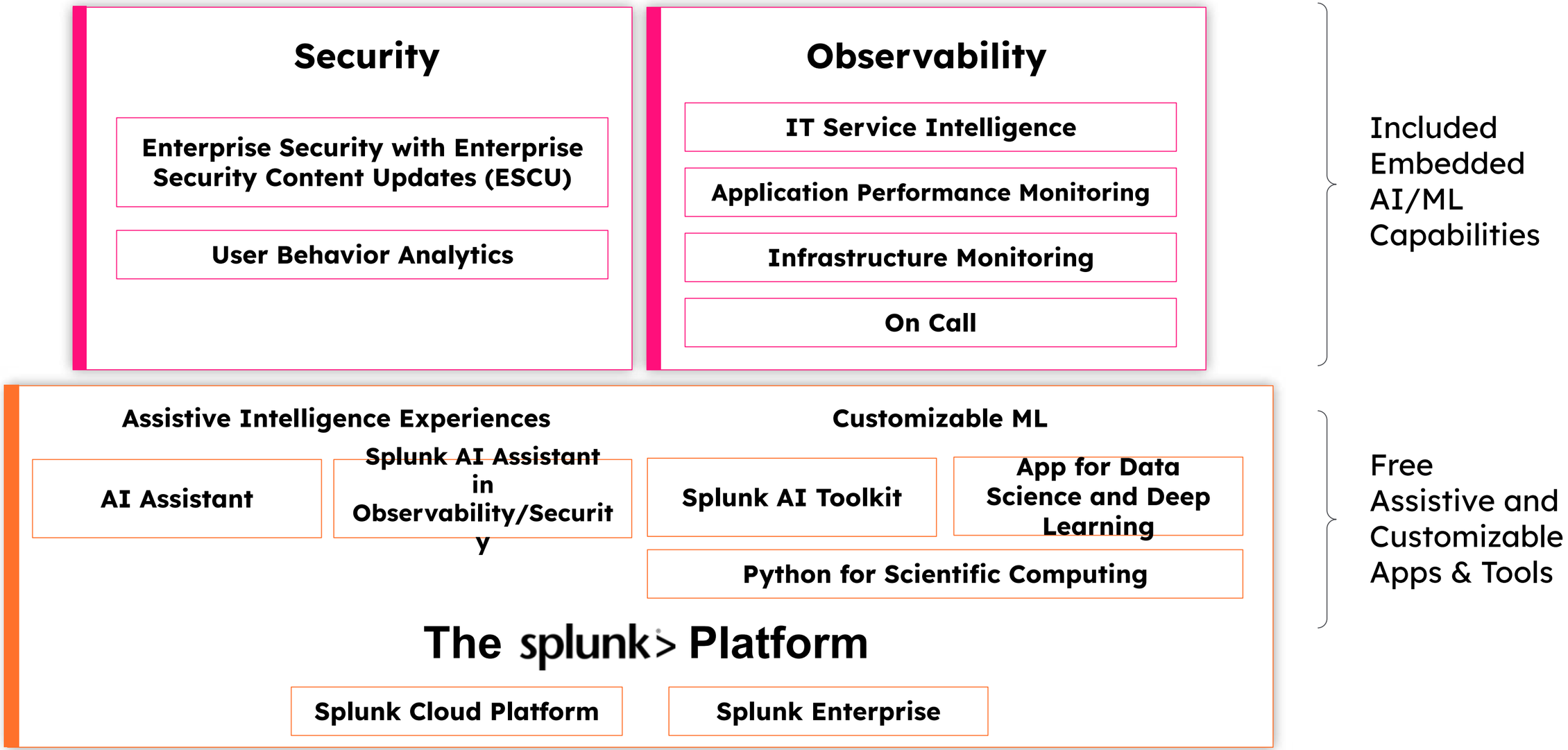
The splunk> Platform

Powerful search commands for statistical analysis, predictive analytics and clustering

© 2025 SPLUNK LLC

Powered by Splunk AI

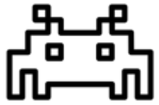
Product Overview



© 2025 SPLUNK LLC

Splunk Enterprise Security

with ML-Powered Content Updates from the Splunk Machine Learning for Security Team



Study Threats

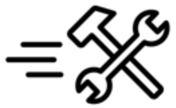
Identify emerging threats and understand how they operate

Create Datasets

Collect data and use Splunk to parse the data and identify patterns that can be used to detect the threat

Build ML & Deep Learning Powered Detections

Build advanced models based on data in order to make predictions or decisions; enable systems to learn from data, identify patterns, and make decisions with minimal human intervention; and craft rules or queries designed to identify specific activity associated with threats – Leverage both Machine Learning & Deep Learning.



Test Detections

Run queries against a dataset that simulates attacker behavior to improve accuracy and reduce false positives

Release

Package detections to deliver timely and effective protections against emerging threats to Splunk customers

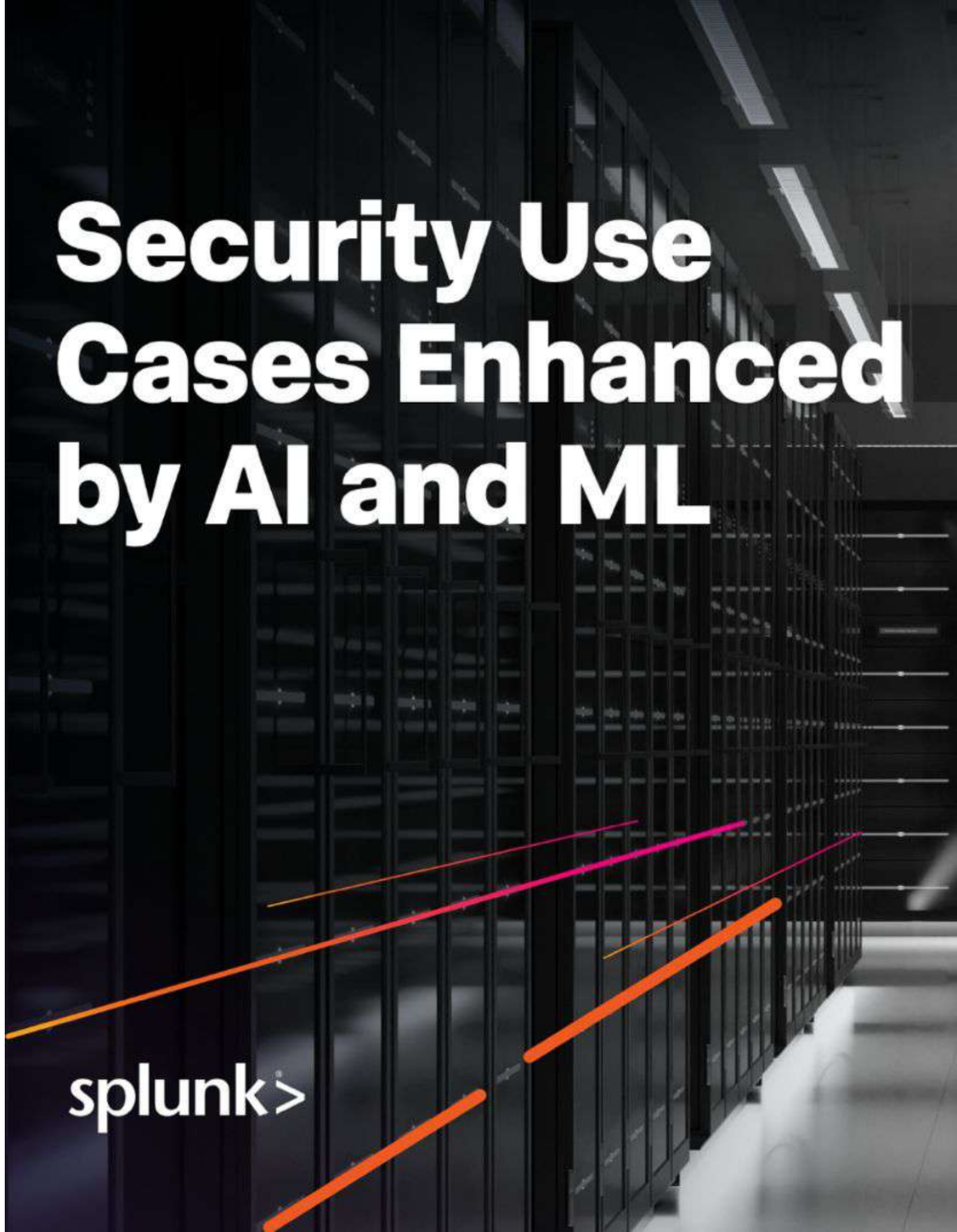
© 2025 SPLUNK LLC

Customize AI for security

In Splunk Enterprise Security

OOTB ML analytics in ES and with the ES Content Updates (ESCU), which can be fully customized to your environments.

Documentation and guidance on some of the most common security use cases that are enhanced by AI and ML.

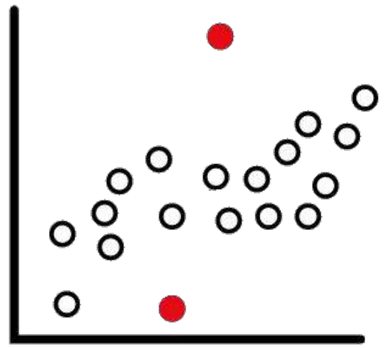


© 2025 SPLUNK LLC

ML-Powered Detections for Security

Find the Obscure and Unknown Threats Buried Deep in Your Data

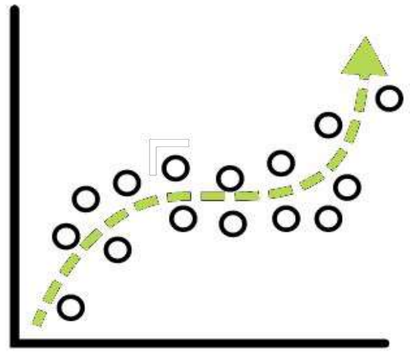
Anomaly detection



Deviation from past behavior

Resource Utilization
Error Rate Deviation
Access Pattern Baselining

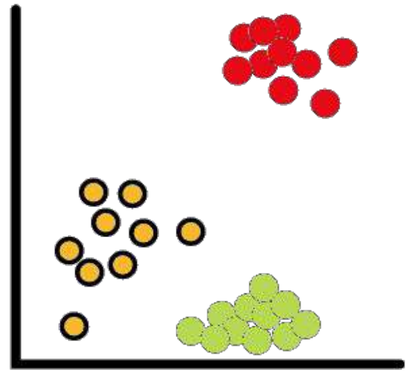
Predictive Analytics



Future state prediction
Classification/regression

Predict storage requirements
Identify patterns leading to failure

Clustering



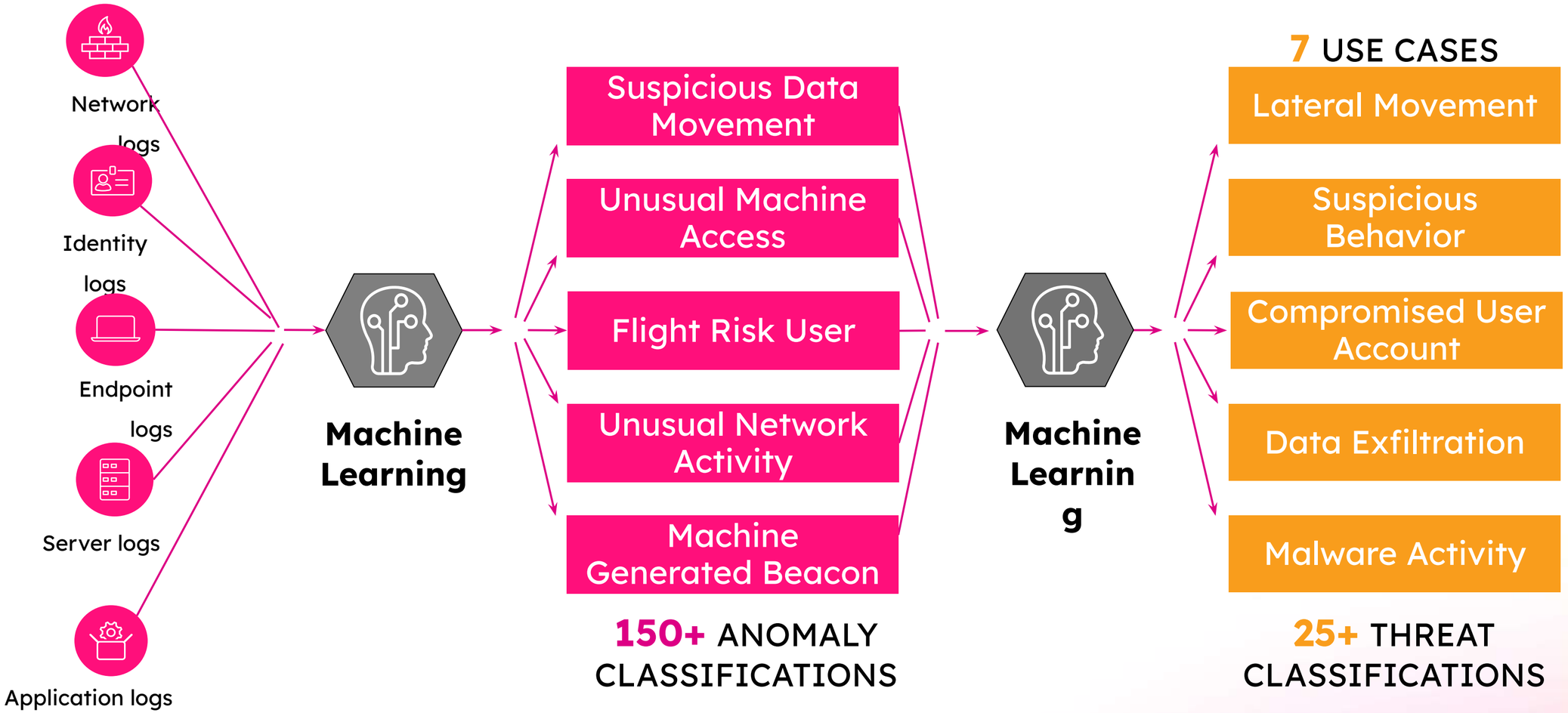
Behavioral Analytics

Identify Traffic
Classify Behaviors

© 2025 SPLUNK LLC

User Behavior Analytics

Detect threats and anomalous behavior using machine learning



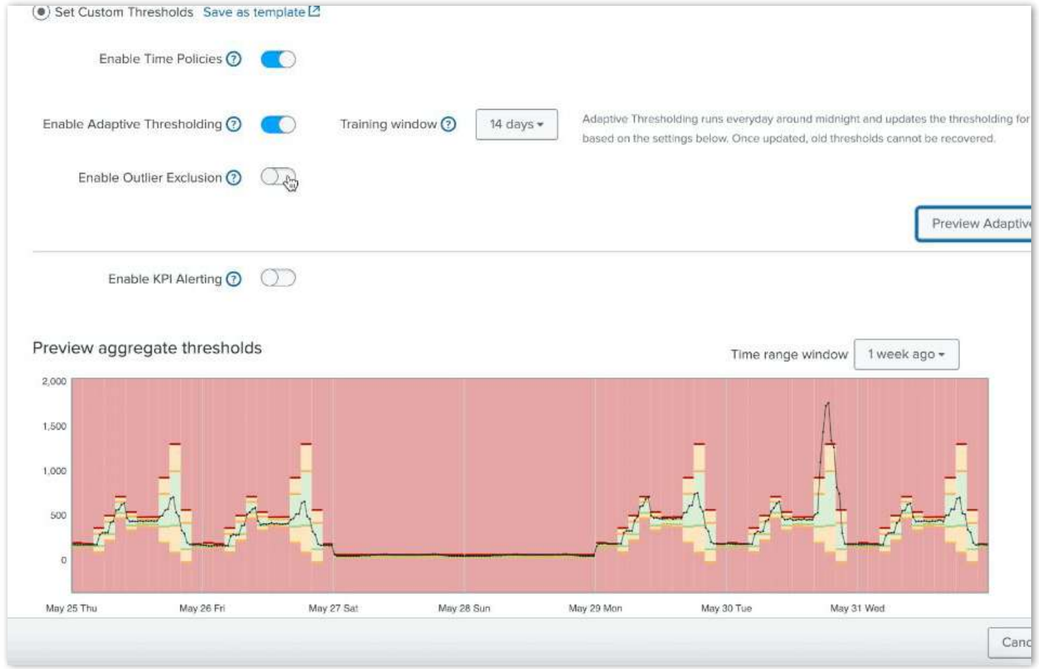
© 2025 SPLUNK LLC

Splunk IT Service Intelligence (ITSI)

Splunk's **AIOps** Solution

Splunk ITSI applies machine learning to **proactively prevent outages** by correlating and reducing alerts, monitoring service health, and streamlining incident management.

- Clustering & aggregation to reduce alert noise
- Adaptive (dynamic) thresholds incorporate seasonality
- Anomaly and outlier detection
- Actionable additional context
- Assisted root cause investigation
- Predict service health to prevent outages



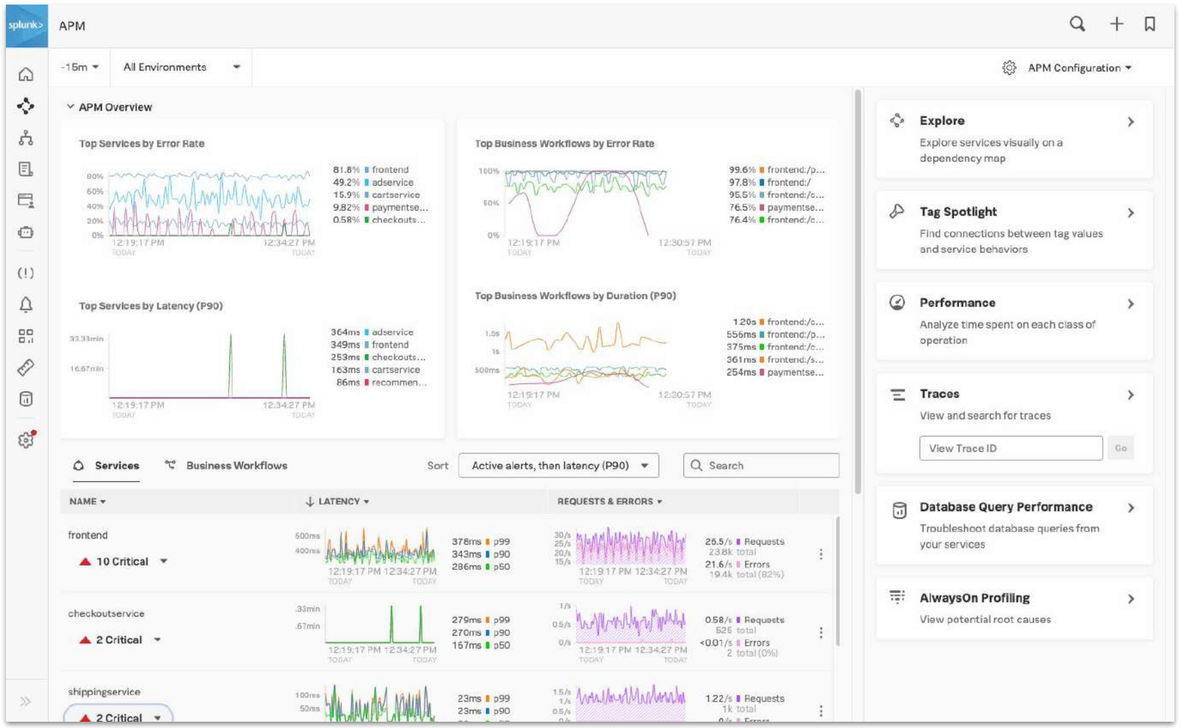
New updates! conf23 ★

- Outlier Exclusion in Adaptive Thresholds
- ML-Assisted Thresholding *(Preview)*

© 2025 SPLUNK LLC

Splunk Application and Infrastructure Monitoring: Autodetect

More accurate and efficient alerting for your infrastructure and services



Use machine learning to **improve accuracy** and **reduce manual effort** across infrastructure and service alerting

- Establish performance baselines across every service
- Automate alerting by receiving recommendations for the biggest spikes in latency, errors, and resource utilization
- Easily customize alert thresholds and subscribe to notifications for specific services or teams

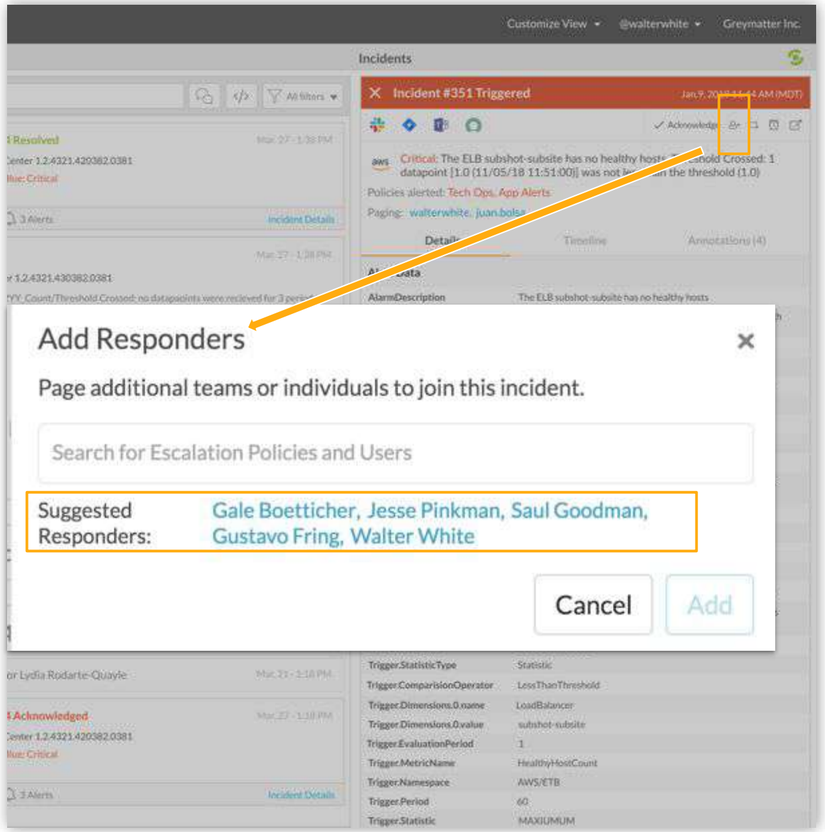
© 2025 SPLUNK LLC

Splunk On-Call

Incident Correlation

Splunk On-Call **leverages ML** to correlate like incidents and decipher which person or people have worked on similar incidents in the past.

The **Suggested Responders** are listed when a user clicks **Add Responder**—making it possible to loop in knowledgeable colleague/s to the firefight in moments.



Benefits:

- Reduce isolation
- Leverage tribal knowledge to reduce MTTR
- Improve morale/ reduce burn out

© 2025 SPLUNK LLC

Splunk AI Toolkit

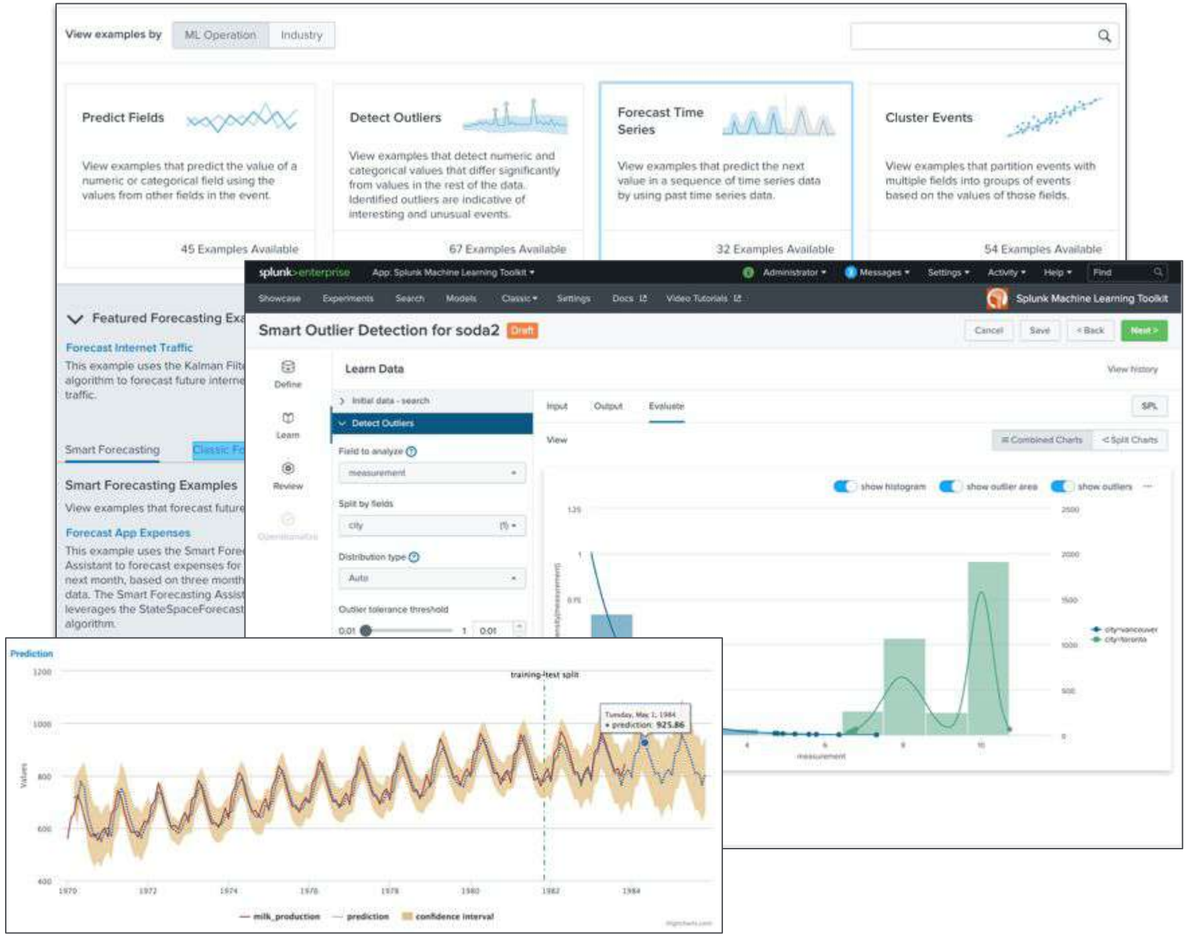
Extends Splunk to operationalize Machine Learning – freely available on splunkbase

Built for the Citizen Data Scientist

- **Experiments and Assistants**: Guided model building, testing, and deployment for common objectives
- **Algorithms**: 80+ standard algorithms (supervised & unsupervised)

Extensible to operationalize any use case

- **Python for Scientific Computing Library**: Access to 300+ open source algorithms
- **Deep Learning Toolkit**: Supports NN and GPU accelerated machine learning
- **ML-SPL API**: Import any open-source or proprietary algorithm



© 2025 SPLUNK LLC

Splunk AI Toolkit

LLM Integrations

- New feature to allow customers to connect to external LLM models, from providers including OpenAI, Gemini, AWS Bedrock, Anthropic, Grok, Microsoft hosted OpenAI and Ollama
- New search command that allows users to pass data through an external LLM and present the results in the search pipeline

Enhanced ONNX apply

- ONNX apply now supports multi-variate outputs
- Additional configuration options are now available for ONNX apply, such as choice of activation function
- Documentation available to programmatically upload ONNX models to MLTK, without having to upload them manually via the UI

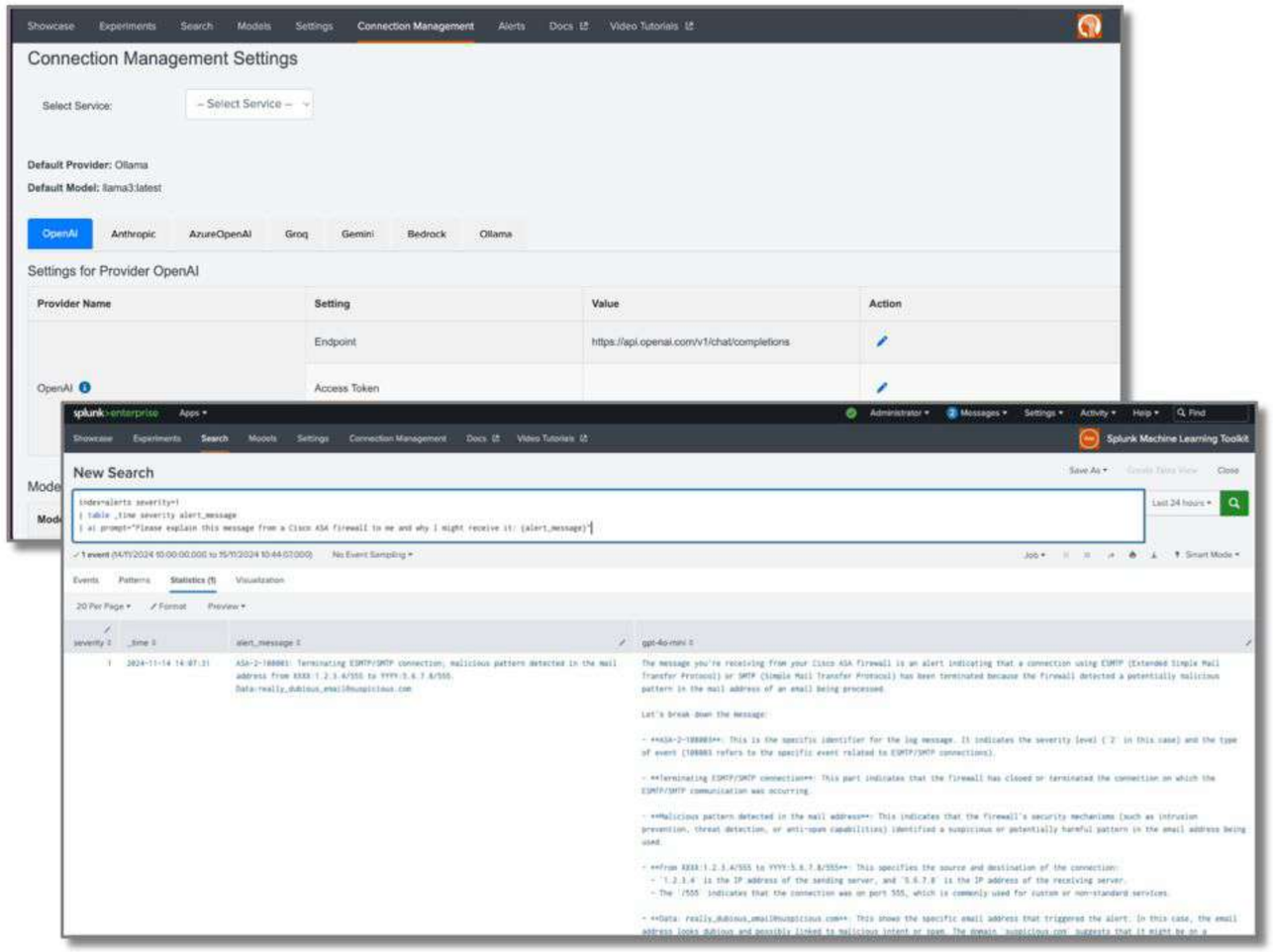
Fixed Vulnerabilities

- Upgrading OpenSSL to v3.4 to mitigate a critical vulnerability in PSC 4.2.3 and 3.2.3

© 2025 SPLUNK LLC

LLM Integrations: What is it?

- Customers can connect to externally hosted LLMs
- Through a natural, human language search command, Splunk data is sent to external LLMs, presenting the response back in Splunk
- Current support for OpenAI, Anthropic, Azure OpenAI, AWS Bedrock, Gemini, Groq and Ollama
- The ERD for the feature is here for more details



© 2025 SPLUNK LLC

Splunk Data Science and Deep Learning 5.2

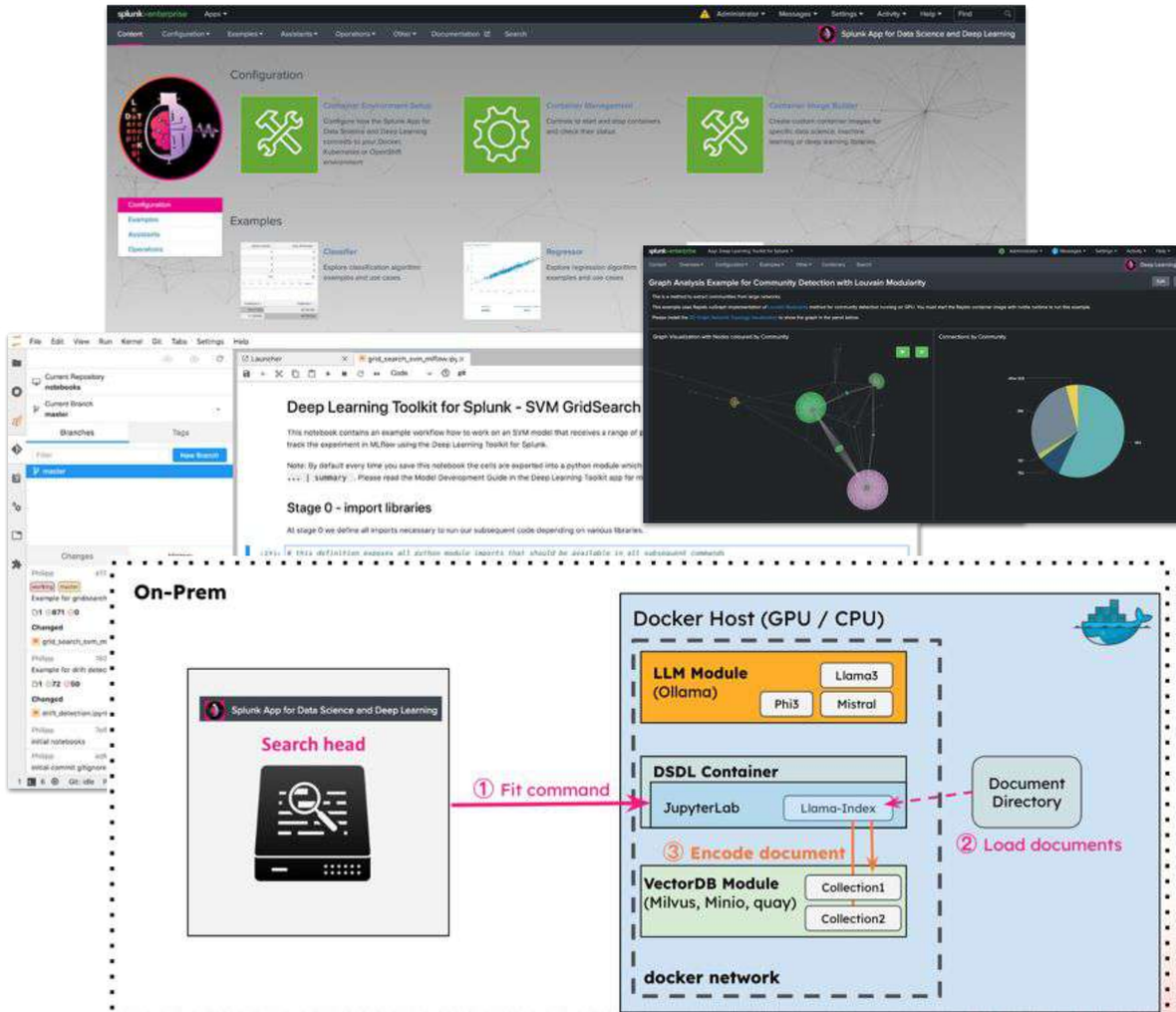
Extension for Splunk AI Toolkit to operationalize advanced custom AI / ML use cases

Built for Data Scientists / AI Engineers

- **35+ Code Examples:** Guided model building, testing, and deployment of data science and deep learning frameworks
- **Container Management**: Models can be productionized for scalability & optimization of resources, e.g. CPU & GPU
- **State of the art AI frameworks and tools:** Jupyter Lab, MLflow, PyTorch, TensorFlow, SpaCy, DASK, Rapids, Spark, …
- **Flexible deployments and open source:** deploy on-prem, hybrid or in the cloud. Github repository for customization.

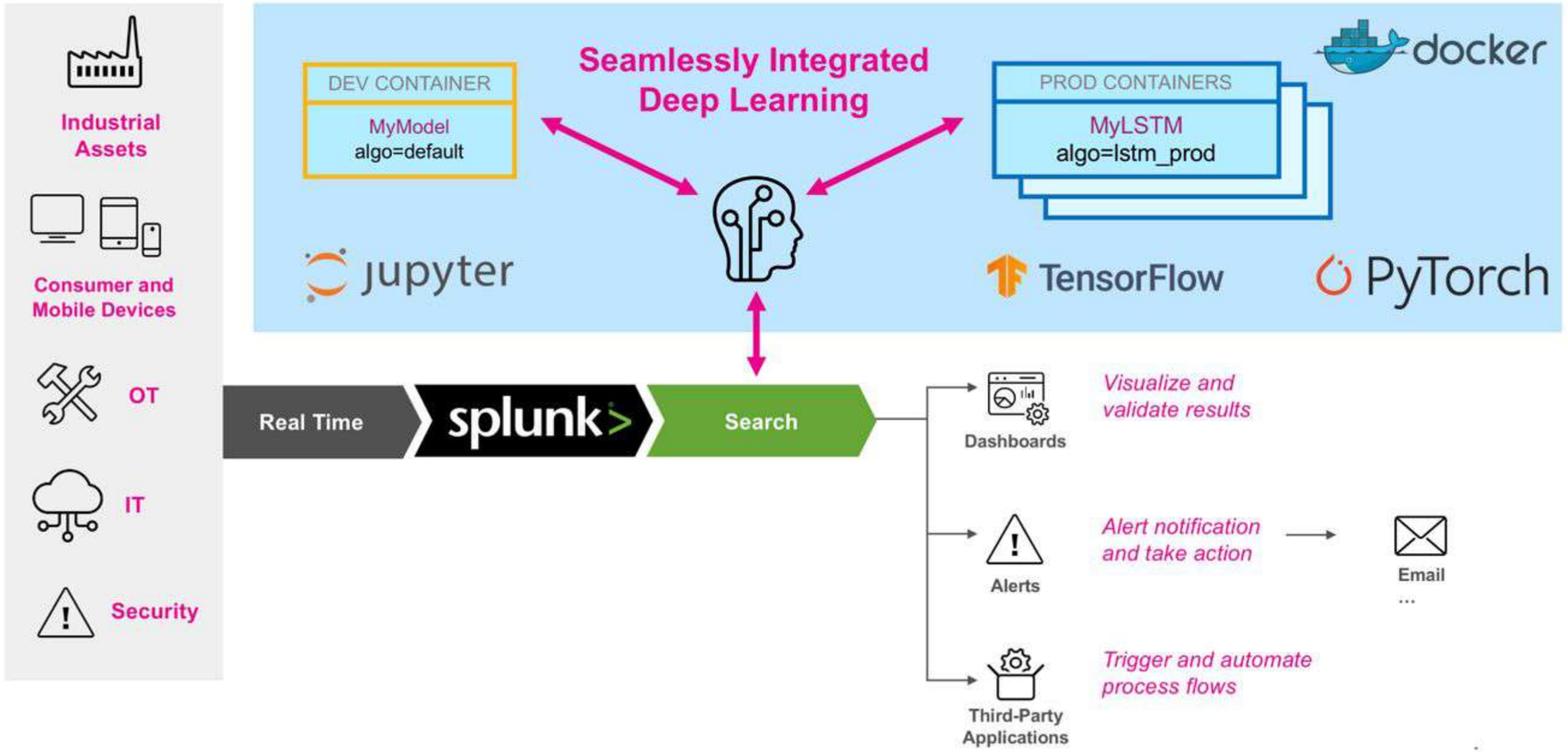
New updates in version 5.2

- **Container extensions for LLM and VectorDB services to enable highly customizable AI use cases.**
- **Custom LLM RAG based applications built on Splunk indexed data and non-Splunk, external data sources**
- **8 new dashboards to easily access the new features with all codebase shared in the public github repo to start building own extensions or use cases.**



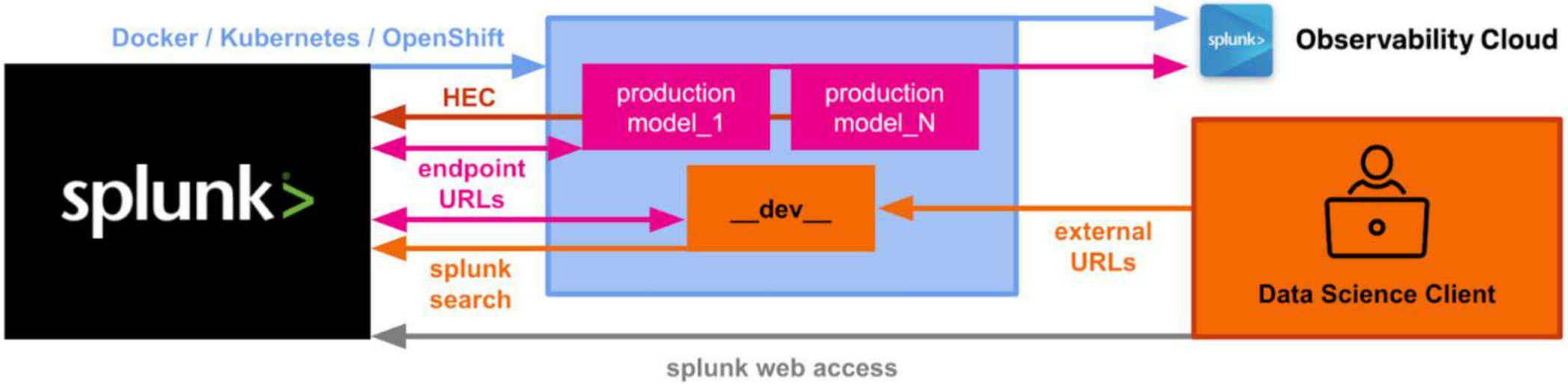
© 2025 SPLUNK LLC

Splunk DSDL - Architecture



© 2025 SPLUNK LLC

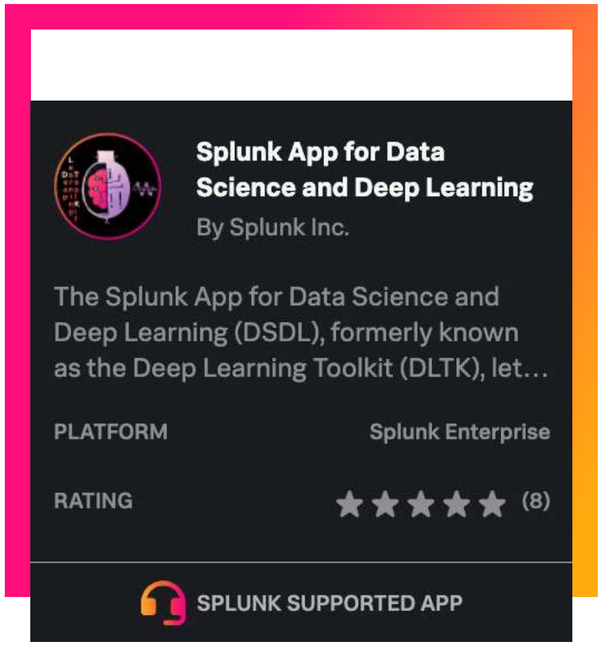
Splunk DSDL - Flow



© 2025 SPLUNK LLC

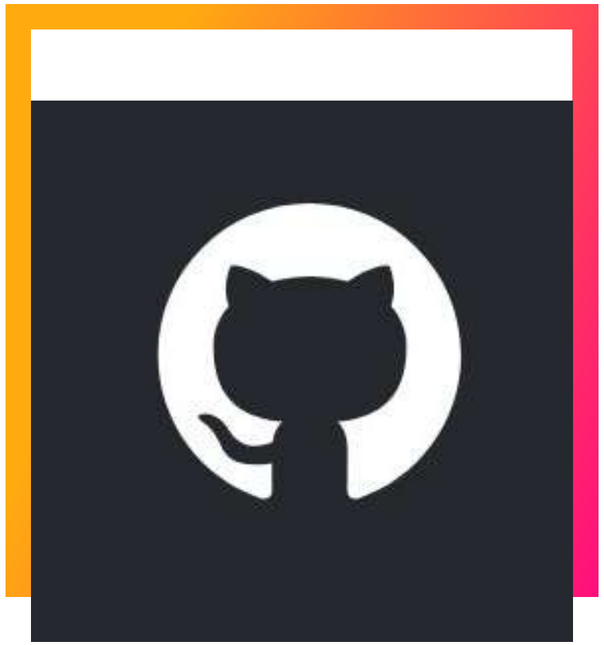
DSDL 5.2 Resources

DEMO Video: Link to video



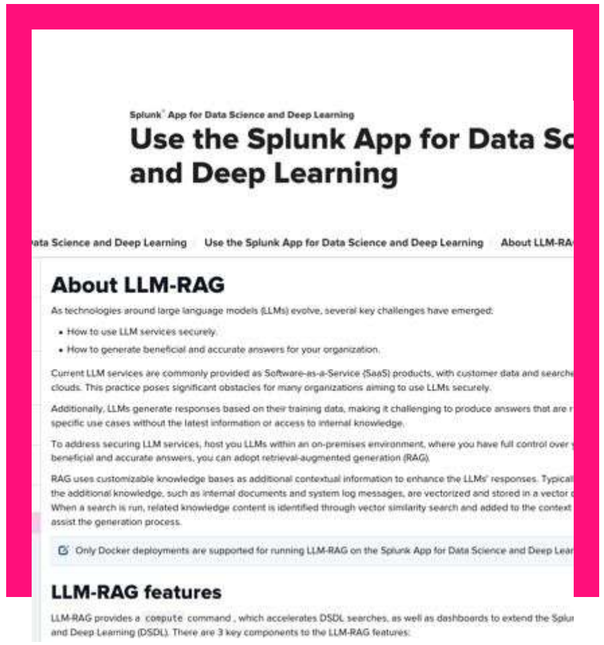
DSDL App Package

Simply install the DSDL 5.2 on your Splunk search head and get started.



Github Repository

Fork the current v5.2 branch from DSDL's public github and extend to your needs.



Documentation

Learn how to setup DSDL 5.2 with the help of this guide.

© 2025 SPLUNK LLC

Python for Scientific Computing

Extend the Splunk Platform Python runtime with AI specific libraries

Run complex AI/ML based analytics in the Splunk Platform with a broad range of supported Open Source Python libraries

© 2025 SPLUNK LLC

Where Are We Going with AI?

Unified experience across Splunk with a scalable backend to power them

We are moving from toolkit capabilities to end-to-end ML guidance for data in Splunk

More Splunk-specific, trusted Generative AI

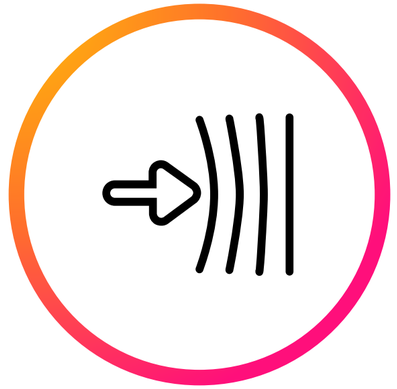
Improve our generative AI in Splunk AI Assistant and expand into other use cases and products

Additional embedded AI

Integrate more AI into users' everyday workflows in Splunk products and build more assistive experiences

Run ML at Scale

Develop a robust runtime to power large-scale ML model training and deployment

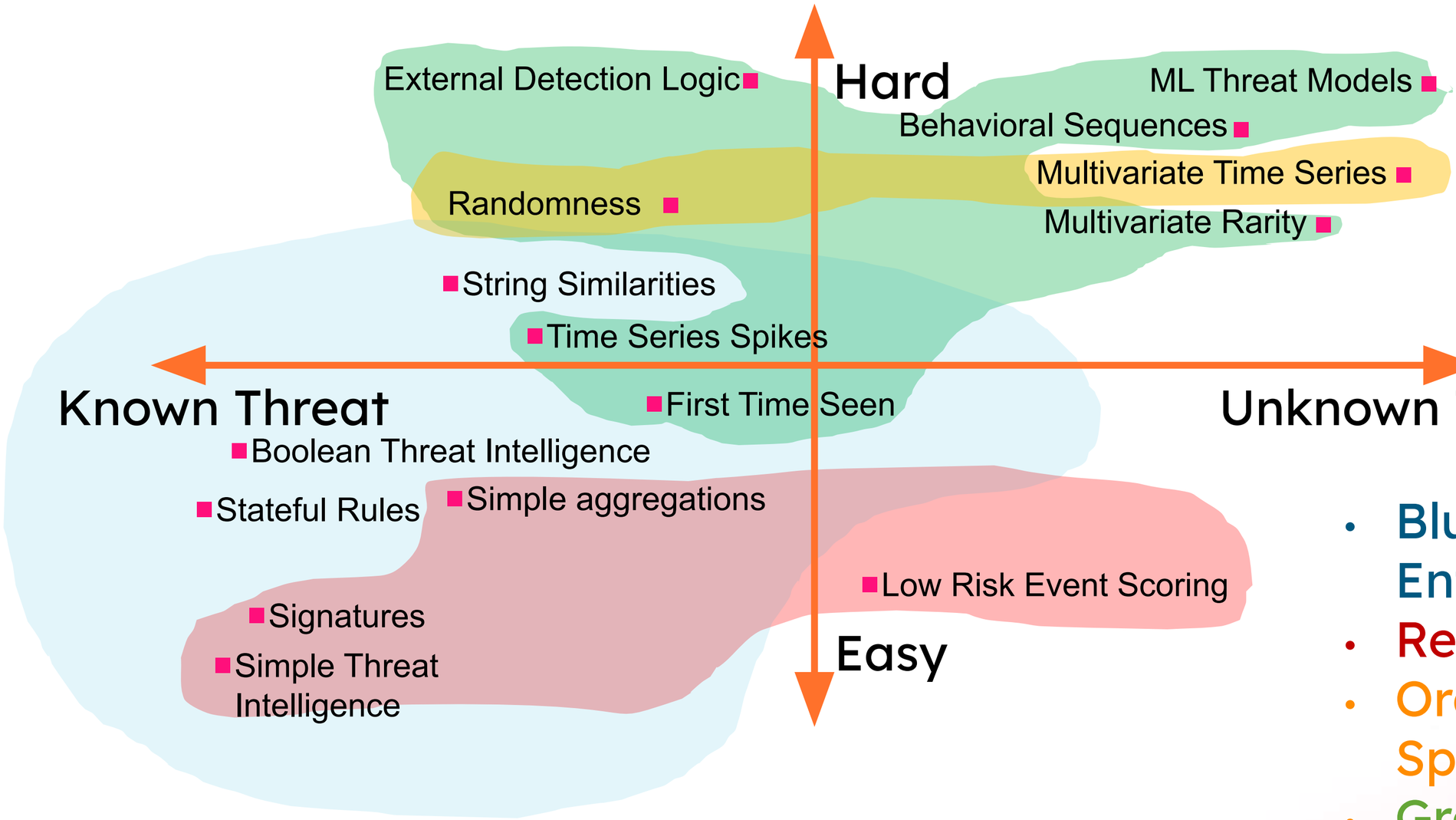


Extensible ML for Splunk developers

Build an SDK to empower developers to create ML powered experiences

© 2025 SPLUNK LLC

Detective Controls by Difficulty



- Blue - Splunk Enterprise
- Red - Splunk ES
- Orange - Splunk Splunk AI Toolkit
- Green – Splunk UBA

© 2025 SPLUNK LLC

Splunk AI Assistants

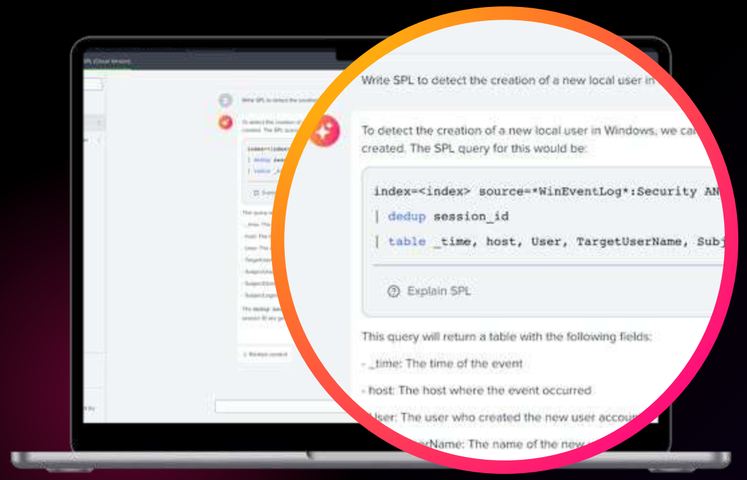
splunk> a CISCO company

© 2025 SPLUNK LLC

Splunk AI Assistants everywhere

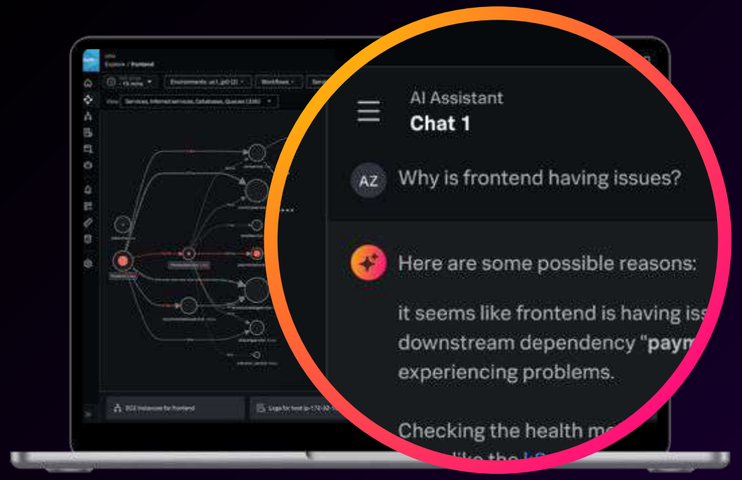
Increase productivity and deliver faster detection and response

AI Assistant for SPL



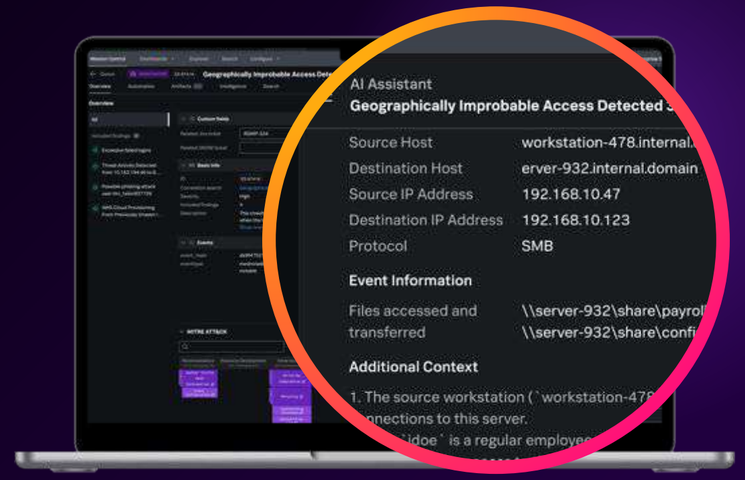
Generally Available

AI Assistant in Observability Cloud



Generally Available

AI Assistant in Enterprise Security



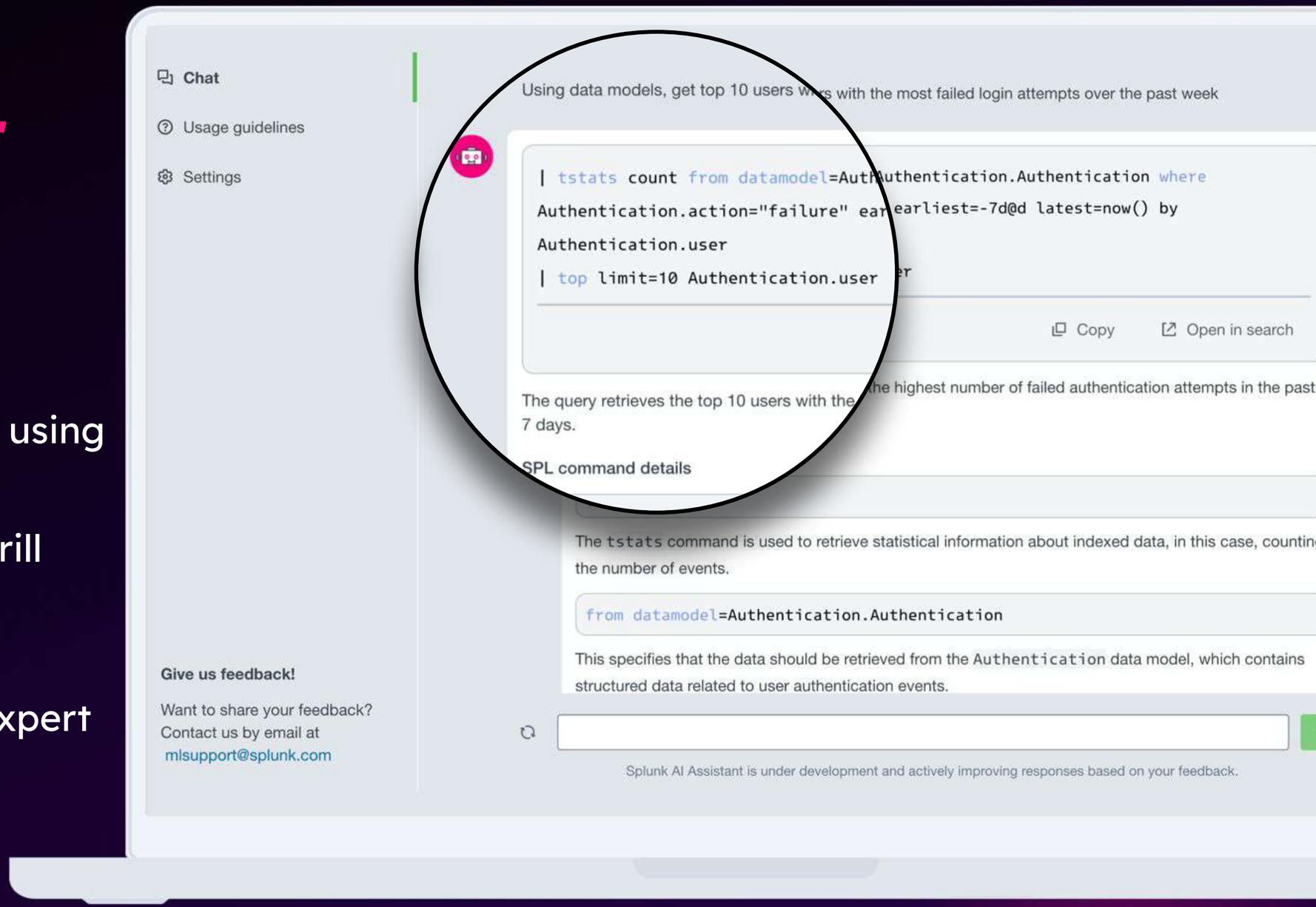
Preview

© 2025 SPLUNK LLC

AI Assistant for SPL

Available in Splunk Cloud

- Get your job done faster using natural language
- Chat with your data to drill down to deeper insights
- Accelerate your learning journey to become the expert

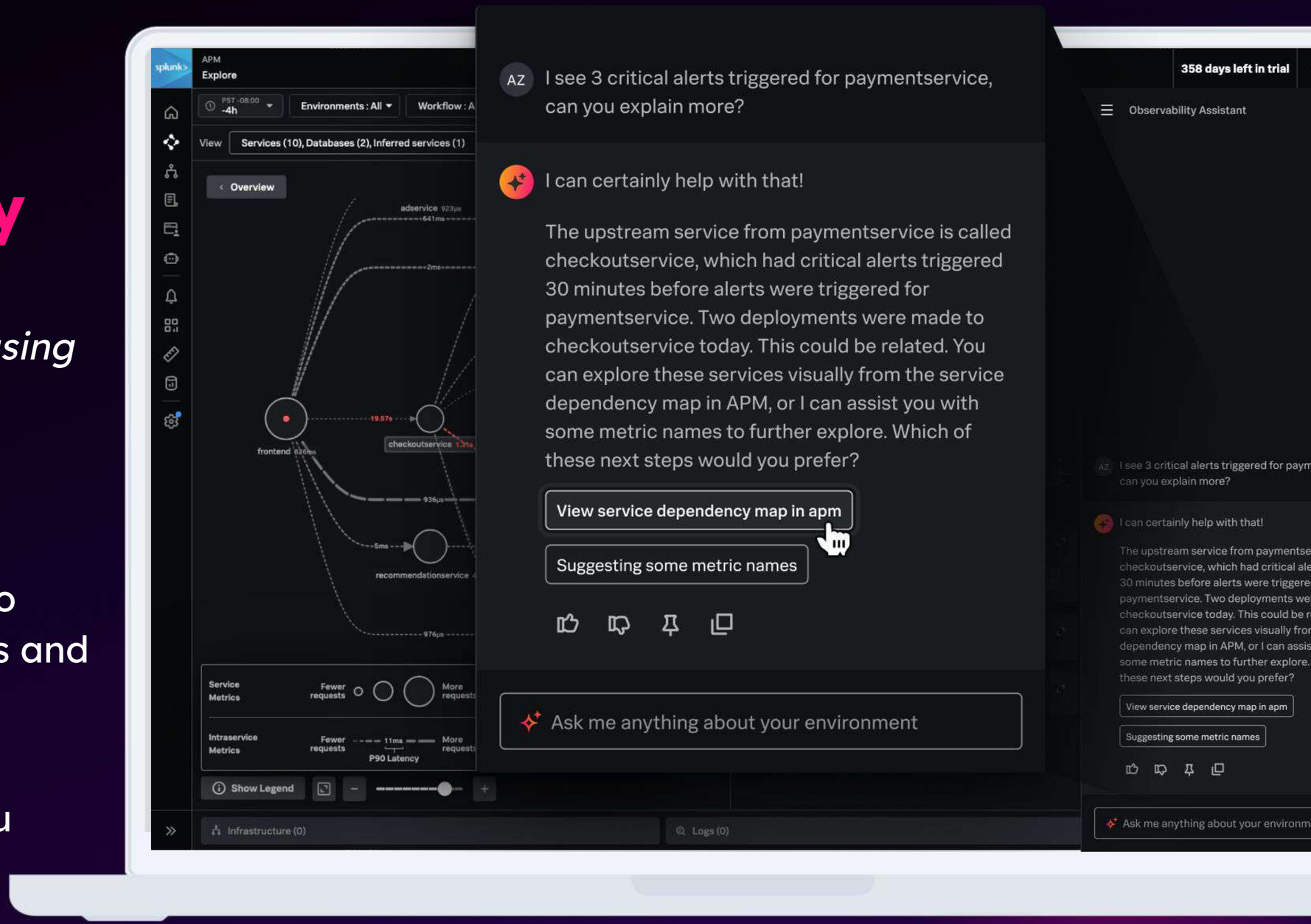


© 2025 SPLUNK LLC

AI Assistant in Observability Cloud

Find and fix issues faster using natural language

- Surface key insights by chatting with your data
- Powered by agentic AI to accelerate investigations and daily monitoring tasks
- Get context and guided recommendations as you troubleshoot



© 2025 SPLUNK LLC

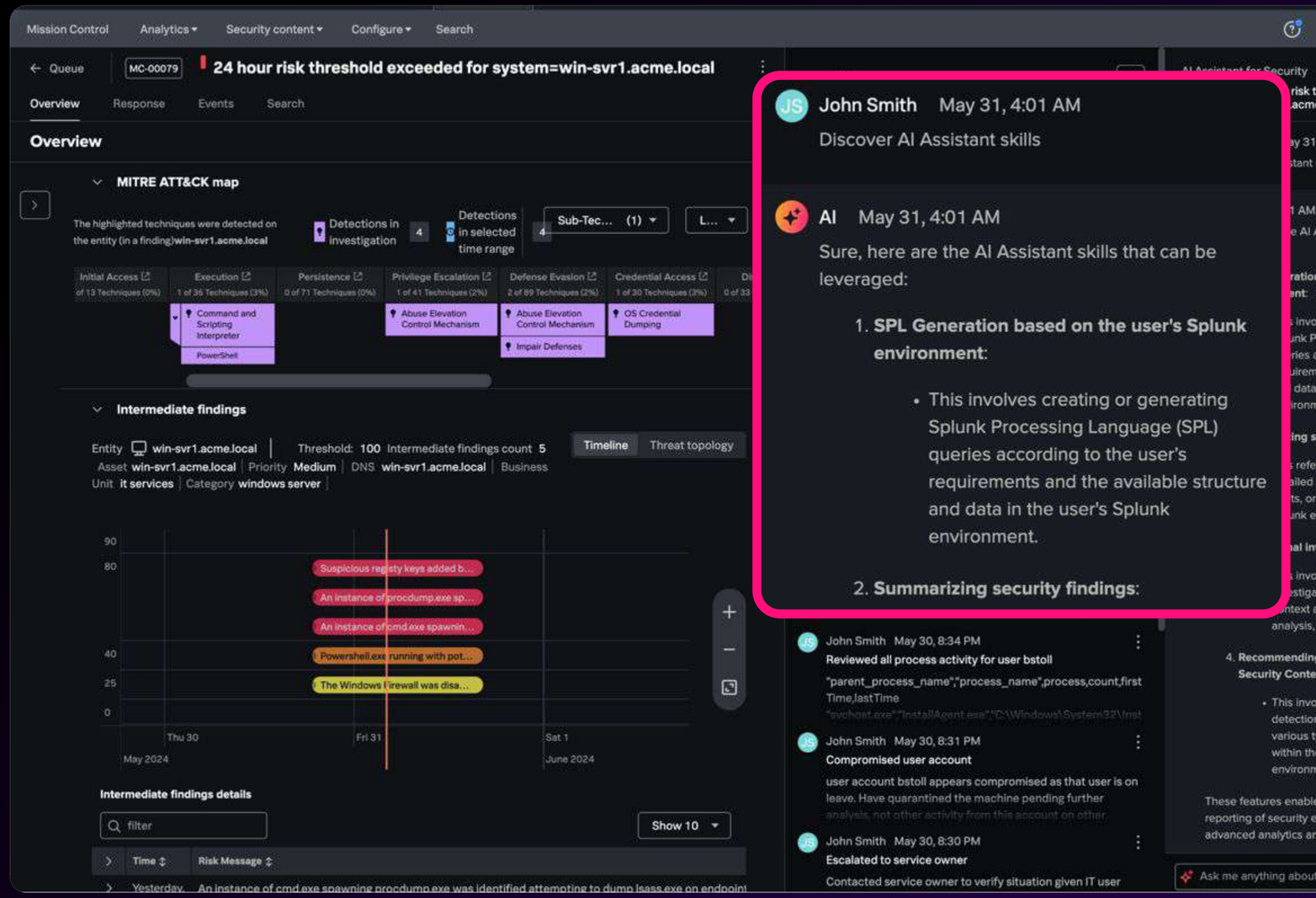
© 2025 SPLUNK INC.

Preview

AI Assistant in Enterprise Security

Guided Enterprise Security workflows

- Answer analyst questions to guide daily workflows.
- Use natural language queries to investigate more quickly.
- Save time while addressing threats more rapidly.



AI Assistant in Enterprise Security is currently in preview. UI shown is for illustration; not final product.

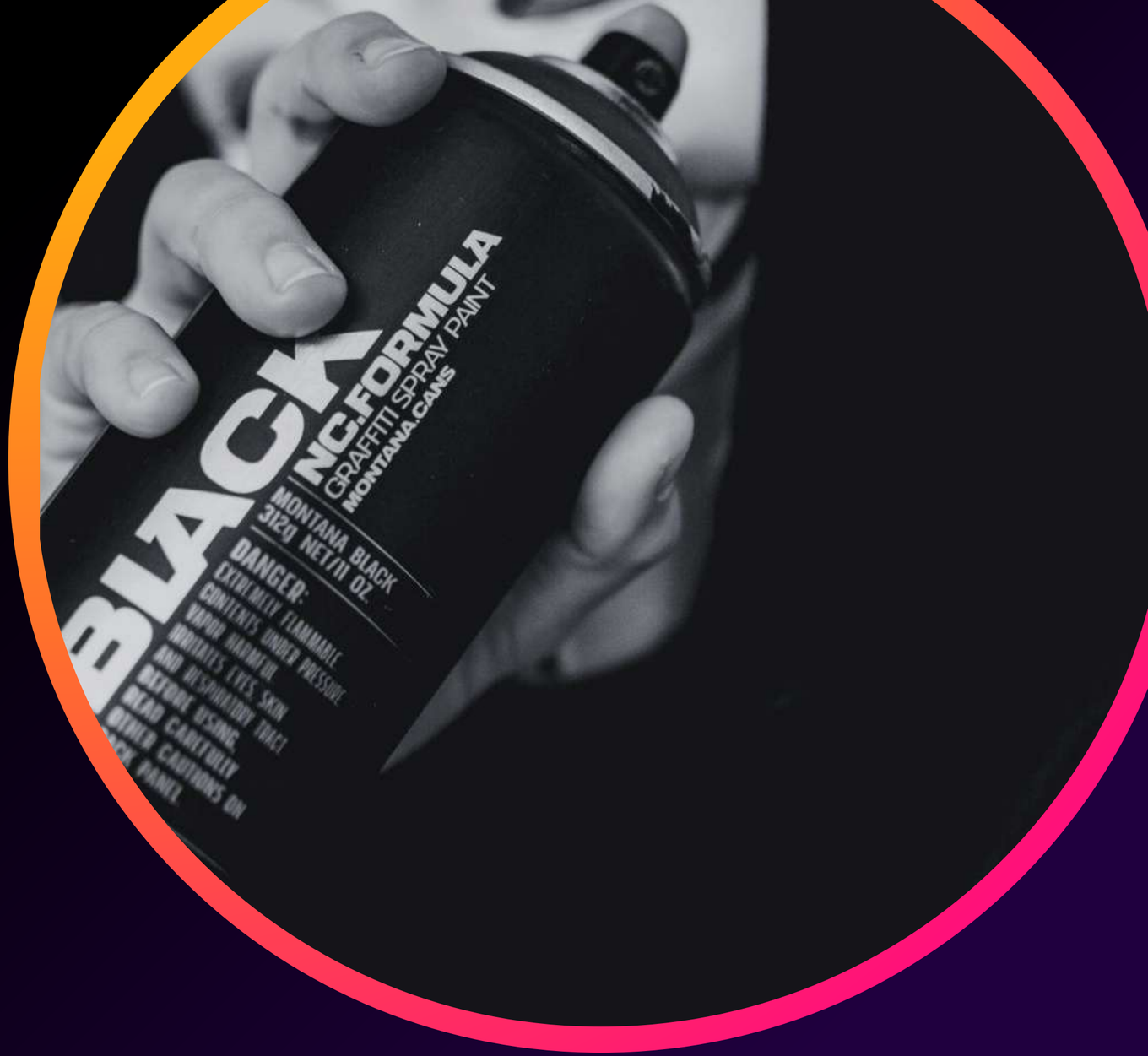
© 2025 SPLUNK LLC

Hands-on labs

© 2025 SPLUNK LLC

Lab 1

Detect Password Spraying Attack- Credential Access



© 2025 SPLUNK LLC

Detect Password Spraying Attack

Credential Access

Lab 1

© 2025 SPLUNK LLC

Detect Password Spraying Attack- Credential Access

What is Password Spraying?

Technique T1110.003 in the MITRE ATT&CK Framework, Password spraying differs from a typical brute-force attack, where attackers try many passwords against a specific user account. However, we can consider spraying a type of brute force attack — it involves a trial-and-error approach of “spraying” a single password across many accounts until a match is found.

This approach enables attackers to avoid countermeasures like account locking when specific login attempts with different passwords have been exceeded. Thus, it is difficult to detect password-spraying attacks without proper detection methods. These password-spraying attacks often target federated authentication used in single sign-on (SSO) and cloud-based applications. Once the attacker gains initial access, they can later access the critical data of organizations.

© 2025 SPLUNK LLC

Using Simple stats

Method 1

© 2025 SPLUNK LLC

Detect Password Spraying Attack

Credential Access

Lab 1

© 2025 SPLUNK LLC

First Technique#1: Simple `stats`

Detect (Remote) Password Spraying Attacks

How to detect this type of attack?

Tips: Windows Event Logs – Which Event code indicates a failed attempt to login to a device? What event indicates a remote authentication attempt? (Optional) You can also filter per Process Name, Computer…

As for the first lab, use transforming commands to visualize your results (`eval`, `stats`, `timechart`, `evenstats`). Leverage Standard Deviation.

Your turn!

© 2025 SPLUNK LLC

First Technique#1: Simple stats

Detect Remote Password Spraying Attacks

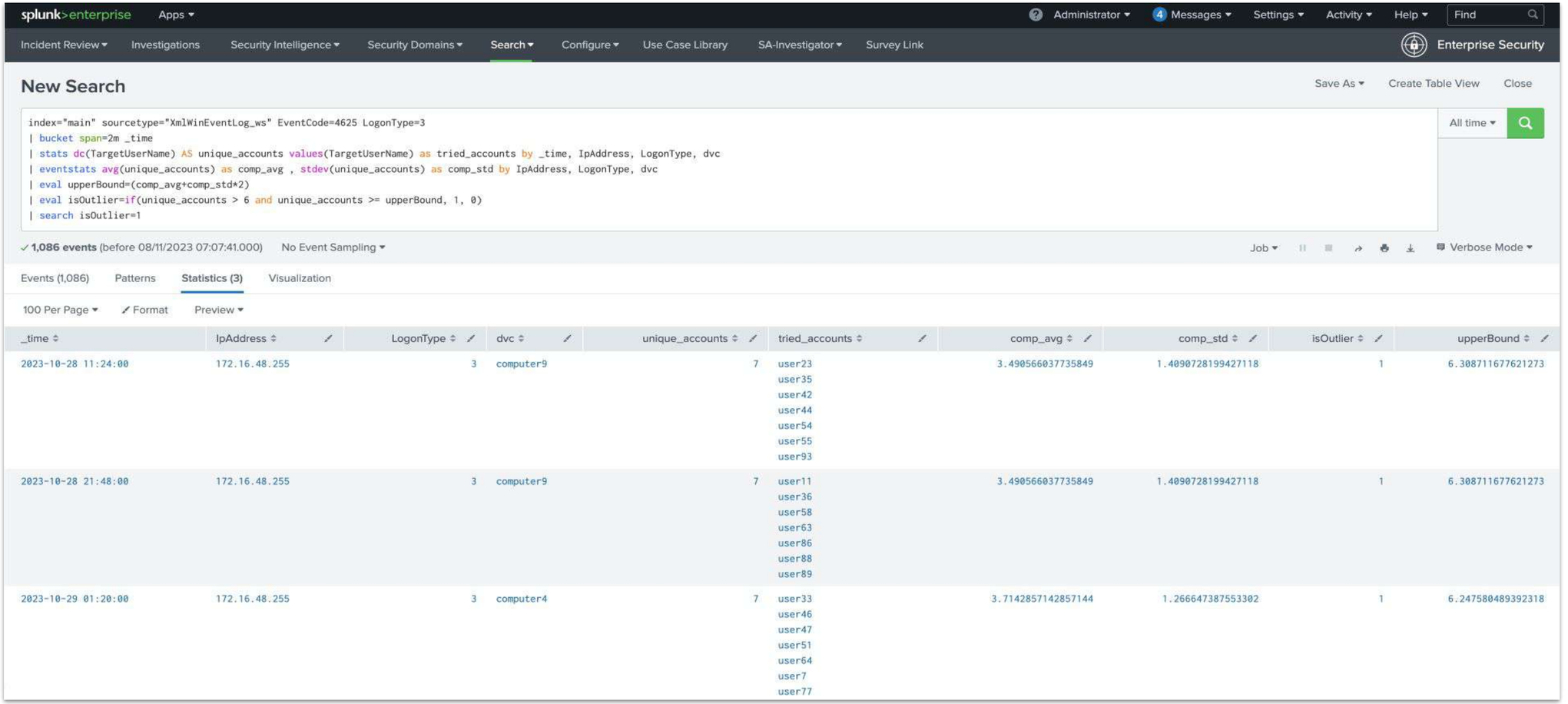
```
index="main" sourcetype="XmlWinEventLog_ws" EventCode=4625 LogonType=3
| bucket span=2m _time
| stats dc(TargetUserName) as unique_accounts values(TargetUserName) as
  tried_accounts by _time, IpAddress, LogonType, dvc
| eventstats avg(unique_accounts) as comp_avg, stdev(unique_accounts) as comp_std
  by IpAddress, LogonType, dvc
| eval upperBound=(comp_avg+comp_std*2)
| eval isOutlier=if(unique_accounts > 6 AND unique_accounts >= upperBound, 1, 0)
| search isOutlier=1
```

Finding and Removing Outliers: https://splk.it/3s0Pig5

© 2025 SPLUNK LLC

First Technique#1: Simple stats

Detect Remote Password Spraying attacks



_time	IpAddress	LogonType	dvc	unique_accounts	tried_accounts	comp_avg	comp_std	isOutlier	upperBound
2023-10-28 11:24:00	172.16.48.255	3	computer9	7	user23 user35 user42 user44 user54 user55 user93	3.490566037735849	1.4090728199427118	1	6.308711677621273
2023-10-28 21:48:00	172.16.48.255	3	computer9	7	user11 user36 user58 user63 user86 user88 user89	3.490566037735849	1.4090728199427118	1	6.308711677621273
2023-10-29 01:20:00	172.16.48.255	3	computer4	7	user33 user46 user47 user51 user64 user7 user77	3.7142857142857144	1.266647387553302	1	6.247580489392318

Finding and Removing Outliers: https://splk.it/3s0Pig5

© 2025 SPLUNK LLC

Splunk AI Toolkit demo

© 2025 SPLUNK LLC

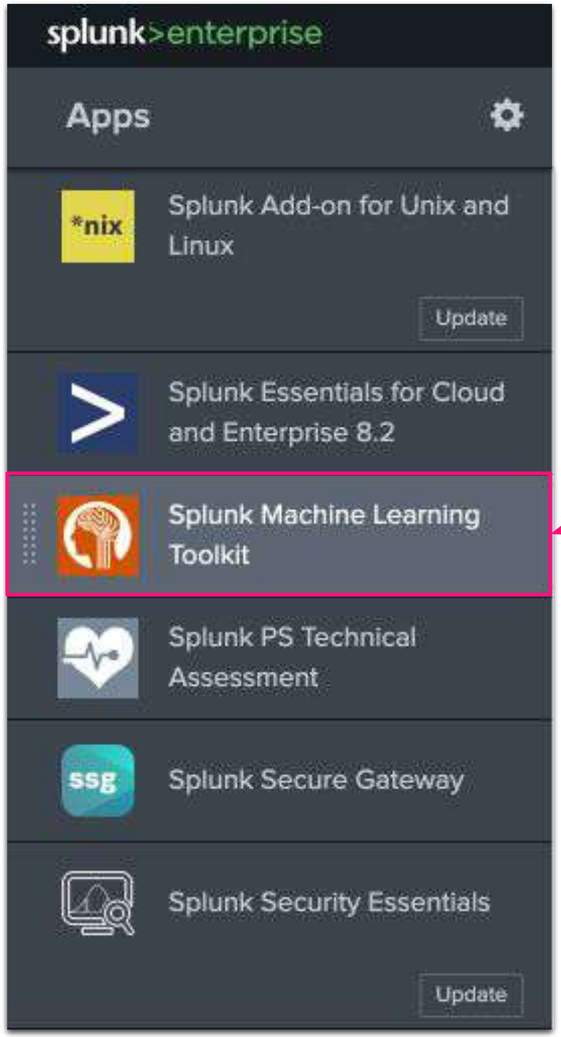
Using Splunk AI Toolkit

Method 2

© 2025 SPLUNK LLC

Splunk AI Toolkit:

First navigate to your Splunk AI Toolkit application from the **Apps** menu:

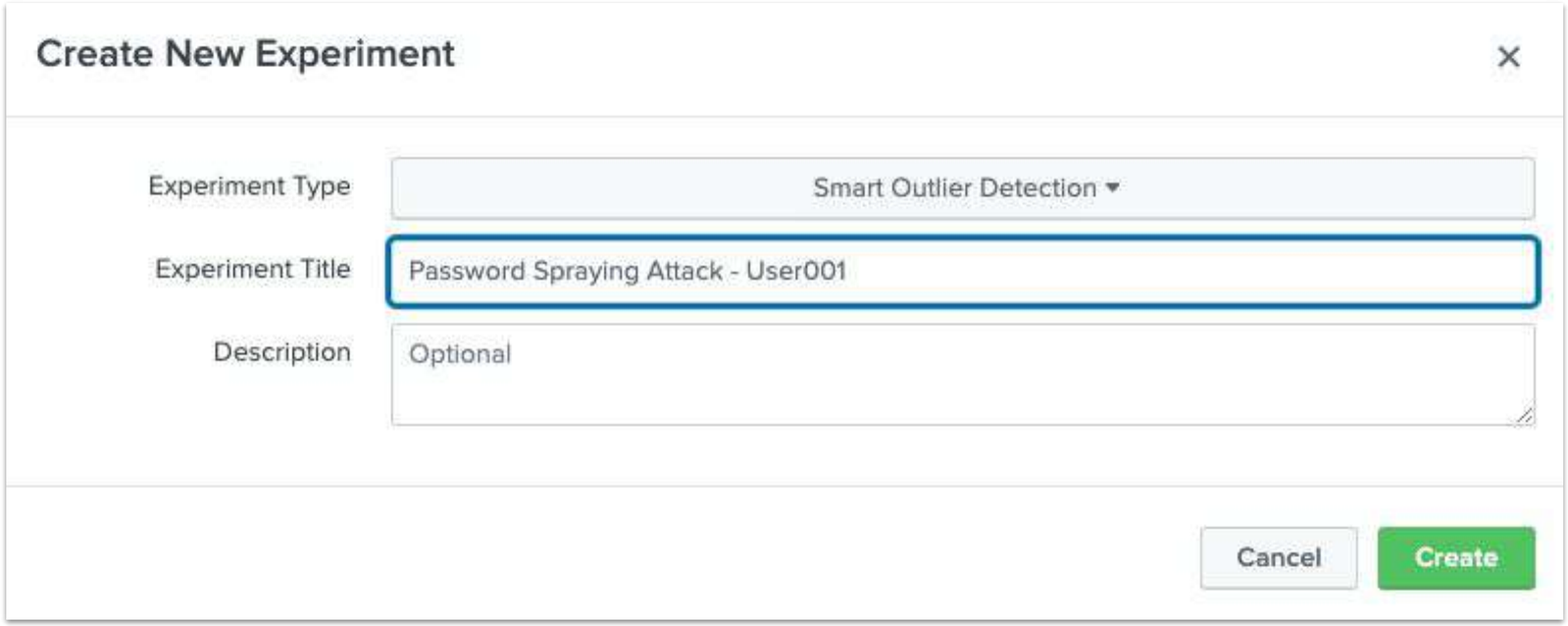


Open the
Splunk AI Toolkit app

© 2025 SPLUNK LLC

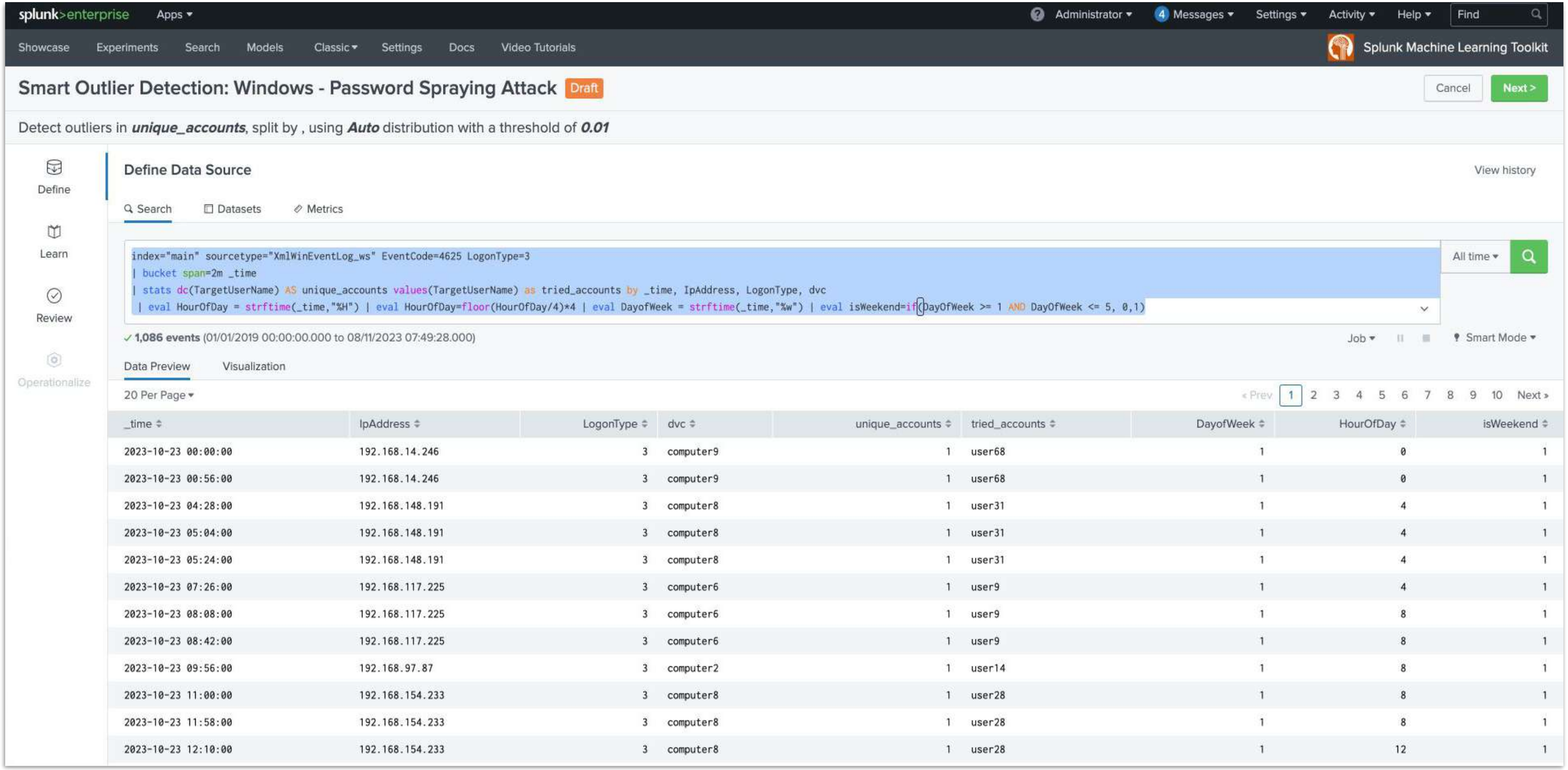
Second Technique#2: Using Splunk AI Toolkit

1. Navigate to the Splunk AI Toolkit App
2. Choose Experiment
3. Select “Smart Outlier Detection”, name your experiment as per below, **adding your username (e.g. UserXXX)**



© 2025 SPLUNK LLC

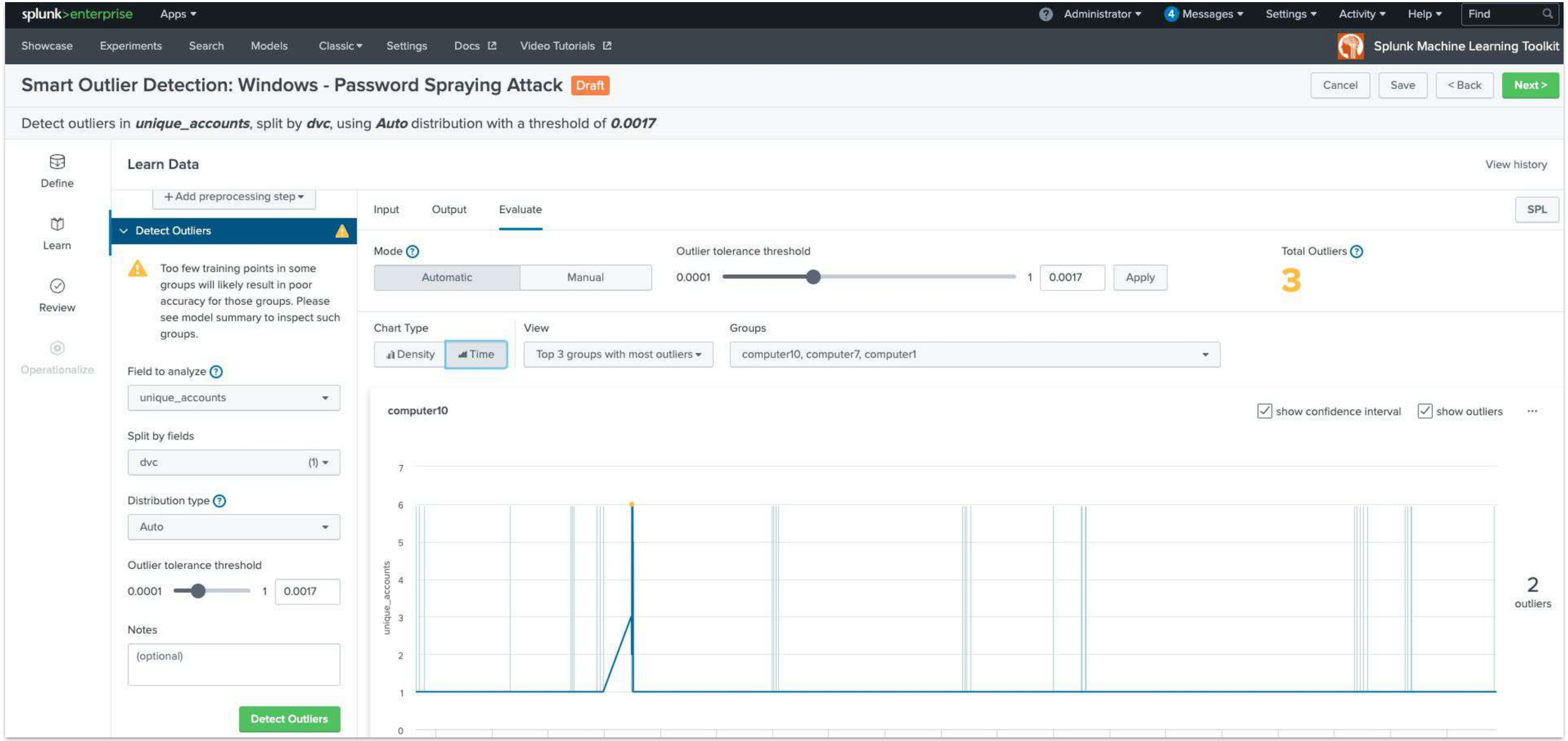
Second Technique#2: Using Splunk AI Toolkit



_time	IpAddress	LogonType	dvc	unique_accounts	tried_accounts	DayofWeek	HourOfDay	isWeekend
2023-10-23 00:00:00	192.168.14.246	3	computer9	1	user68	1	0	1
2023-10-23 00:56:00	192.168.14.246	3	computer9	1	user68	1	0	1
2023-10-23 04:28:00	192.168.148.191	3	computer8	1	user31	1	4	1
2023-10-23 05:04:00	192.168.148.191	3	computer8	1	user31	1	4	1
2023-10-23 05:24:00	192.168.148.191	3	computer8	1	user31	1	4	1
2023-10-23 07:26:00	192.168.117.225	3	computer6	1	user9	1	4	1
2023-10-23 08:08:00	192.168.117.225	3	computer6	1	user9	1	8	1
2023-10-23 08:42:00	192.168.117.225	3	computer6	1	user9	1	8	1
2023-10-23 09:56:00	192.168.97.87	3	computer2	1	user14	1	8	1
2023-10-23 11:00:00	192.168.154.233	3	computer8	1	user28	1	8	1
2023-10-23 11:58:00	192.168.154.233	3	computer8	1	user28	1	8	1
2023-10-23 12:10:00	192.168.154.233	3	computer8	1	user28	1	12	1

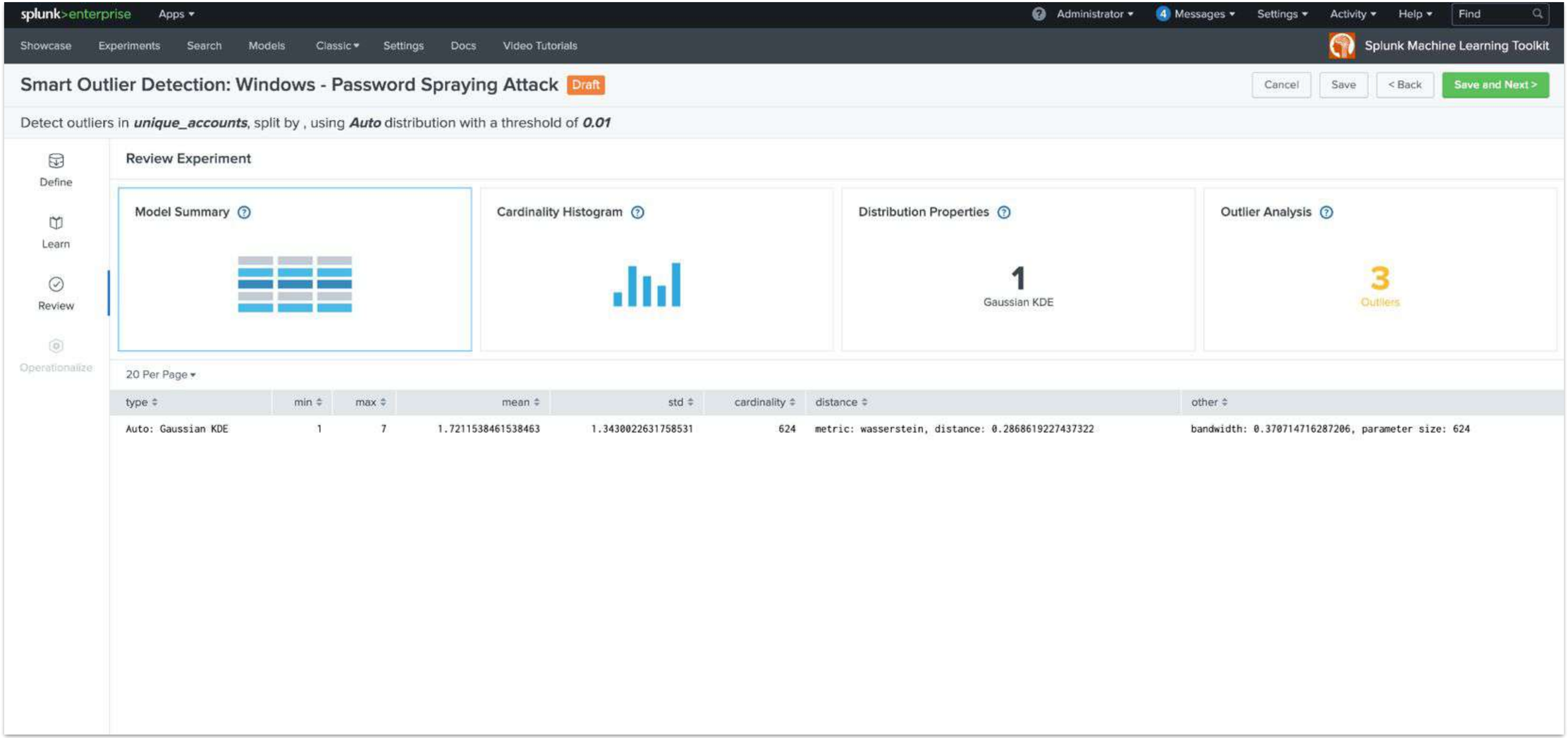
© 2025 SPLUNK LLC

Second Technique#2: Using Splunk AI Toolkit



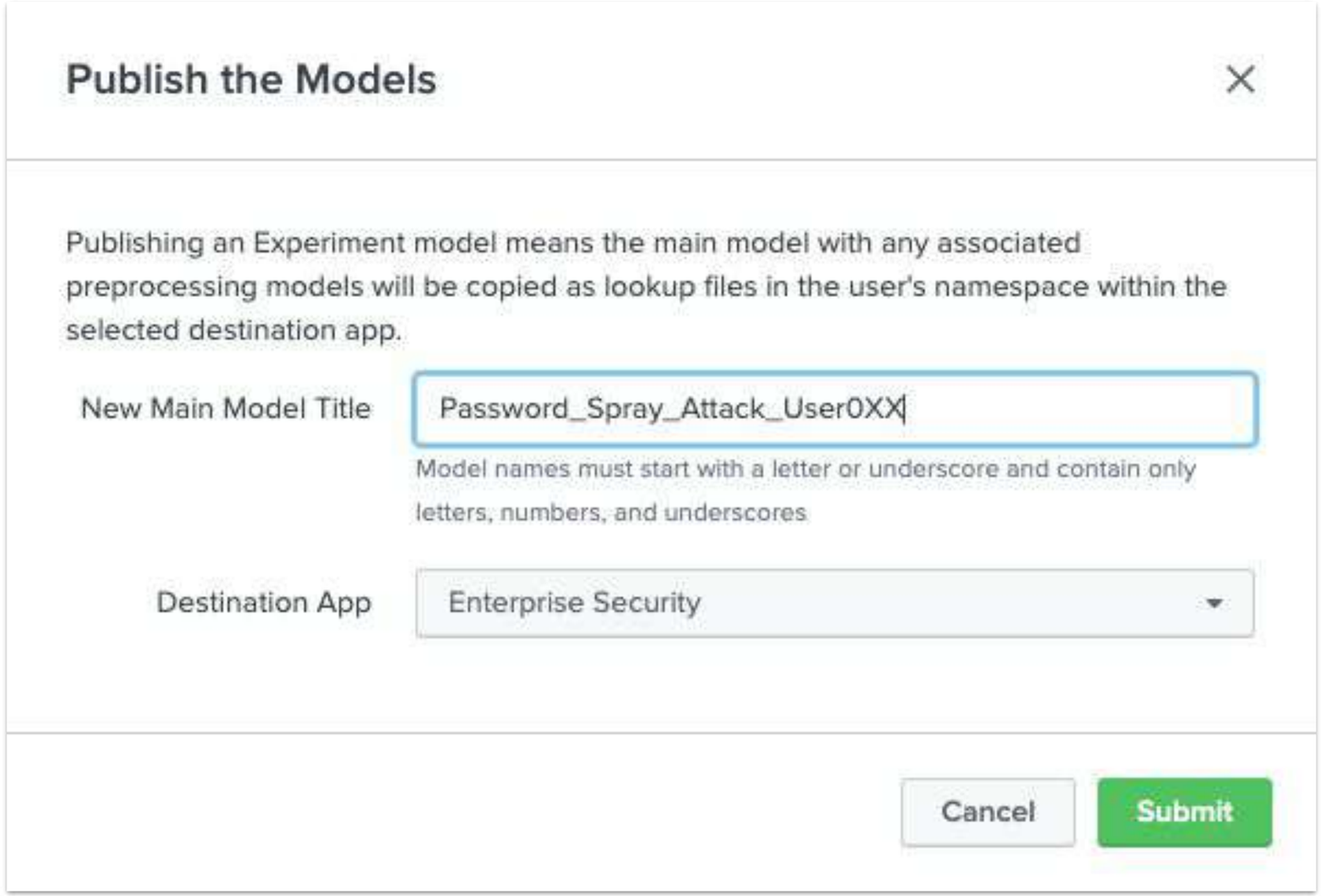
© 2025 SPLUNK LLC

Second Technique#2: Using Splunk AI Toolkit



© 2025 SPLUNK LLC

Publish Your Model



© 2025 SPLUNK LLC

20 minutes to complete the lab

Lab 1

© 2025 SPLUNK LLC

Privilege Escalation

Lab 2

© 2025 SPLUNK LLC

Using Simple stats

Method 1

© 2025 SPLUNK LLC

First technique#1: Simple `stats`

Detecting Privileged Escalation Events

How to find anomalies this condition? You need to understand your data first

Tips: Use EventCode 4648, look for Windows Event Logs
Use the `stats` command to look at the volume of events and break it down by user. Use the `sort` command

Other stats functions: `stdev()`, `perc99.999()`…

What can you tell from the results? Anything outside of the ordinary?

© 2025 SPLUNK LLC

First technique#1: Simple `stats`

Detecting Privileged Escalation Events

Now let’s go one step further

Use the Eventstats, stdev command and eval command to find outliers.

How would you do that? You can look at the documentation and figure out the commands to use:

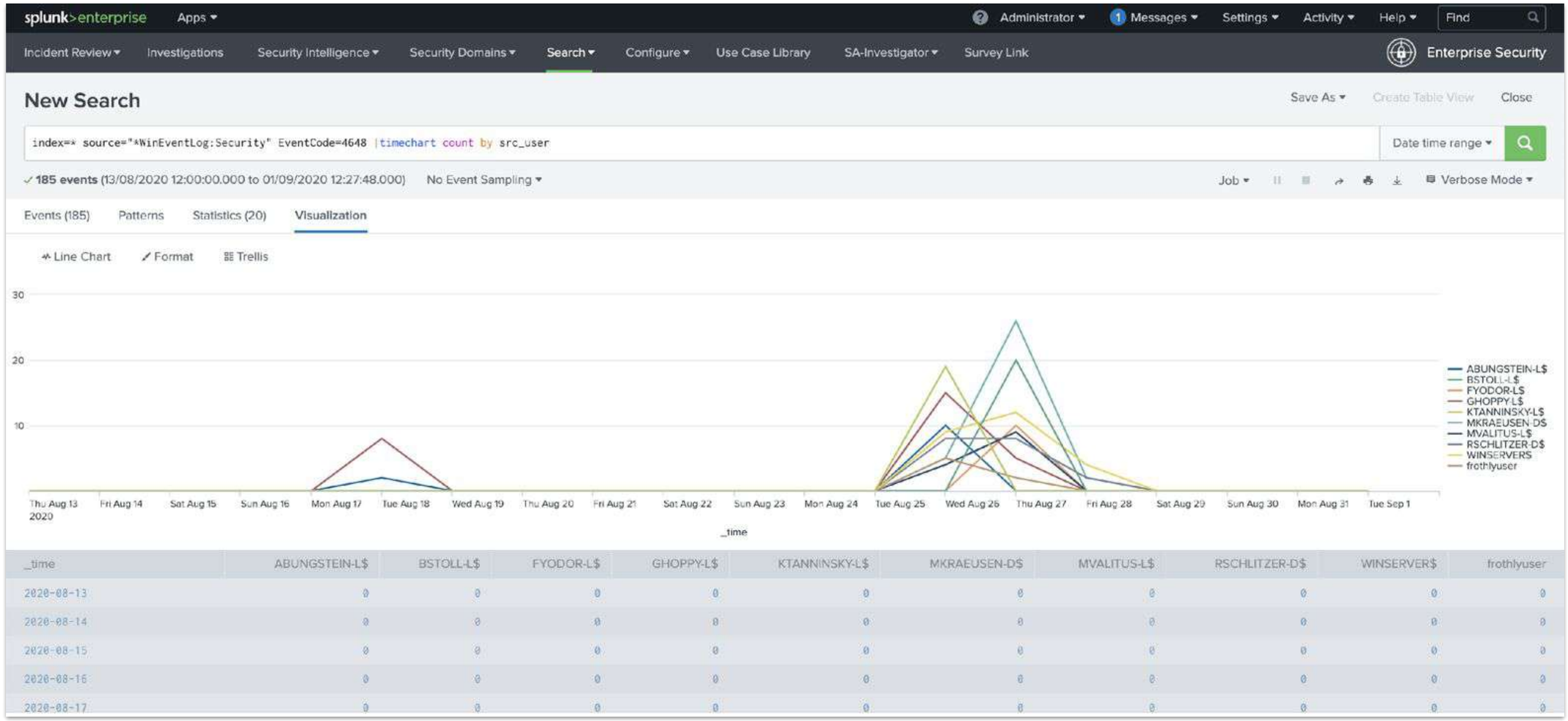
https://docs.splunk.com/Documentation/SCS/current/SearchReference/Aggregatefunctions#stdev.28.26lt.3Bvalue.26gt.3B.29

Of course, there is an easier way! ☺️

© 2025 SPLUNK LLC

First technique#1: Simple stats

Detecting Privileged Escalation Events



© 2025 SPLUNK LLC

Using Splunk AI Toolkit

Method 2

© 2025 SPLUNK LLC

Splunk AI Toolkit:

First navigate to your Splunk AI Toolkit application from the **Apps** menu:

Open the Splunk AI Toolkit app

© 2025 SPLUNK LLC

Splunk AI Toolkit:

Now let's create an experiment:

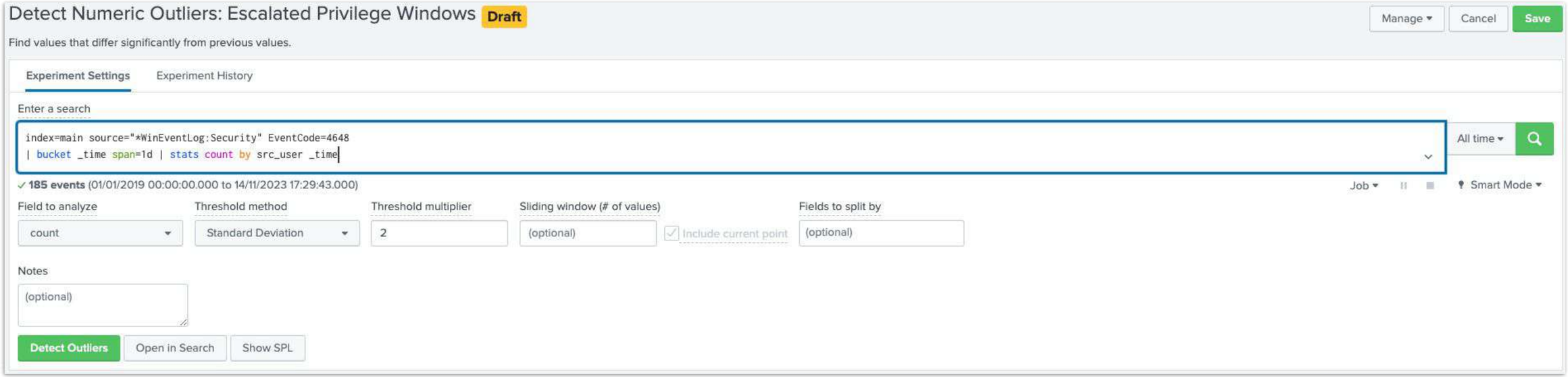
1. Go to experiments
2. Choose Detect Numeric Outliers
3. Create New Experiment
4. Choose a title for your experiment (**add your username as it is a shared instance**)
5. Run the following search:

```
index=main source="*WinEventLog:Security" EventCode=4648
| bucket _time span=1d | stats count by src_user _time
```

6. In Field to analyze, choose count
7. You can pick the Threshold method. For this exercise, the Standard Deviation method is the most suitable
8. You can tweak your Threshold multiplier based on the result that you get (e.g. too many outliers)
9. You can optionally specify a Sliding Window, and you can split by fields
10. Now, click on Detect Outliers

© 2025 SPLUNK LLC

Splunk AI Toolkit:

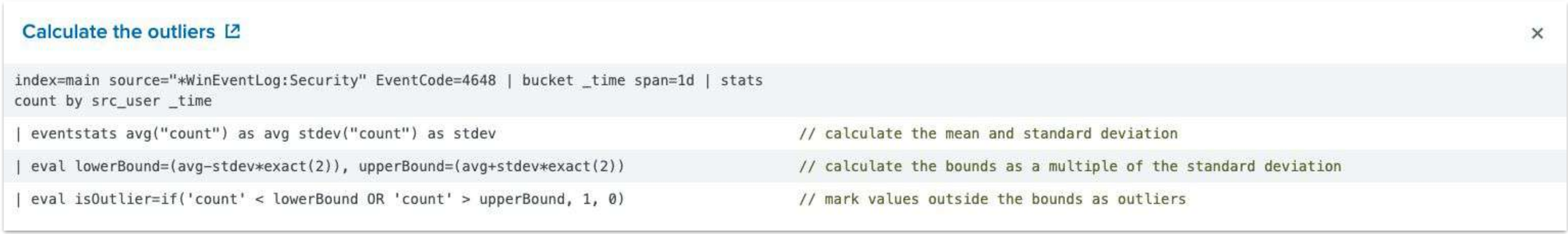


© 2025 SPLUNK LLC

Splunk AI Toolkit:

You can view the SPL query which was generated.

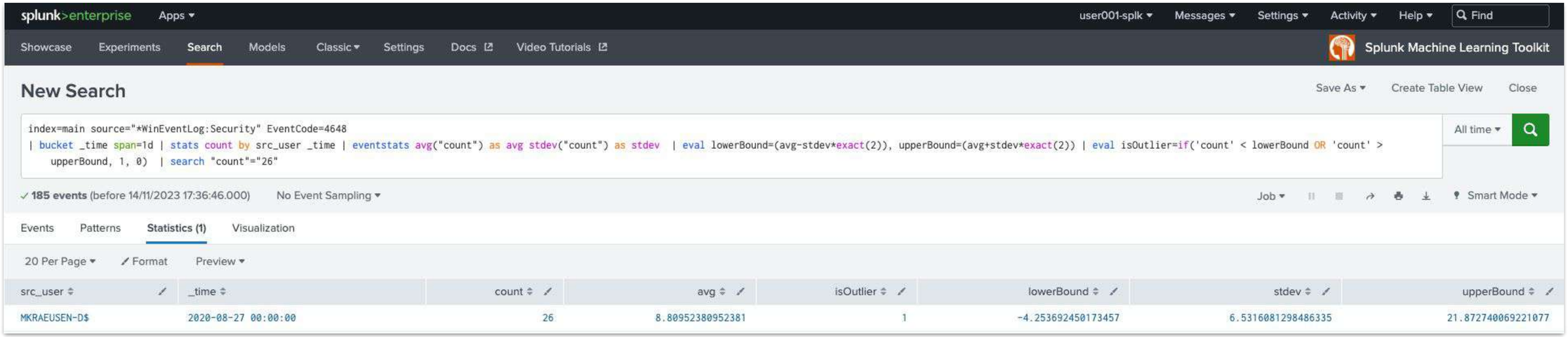
Click the “Show SPL” option:



© 2025 SPLUNK LLC

Splunk AI Toolkit:

Click on the Outlier to see who is the outlier:



© 2025 SPLUNK LLC

What Other Fields Could Be of Interest?

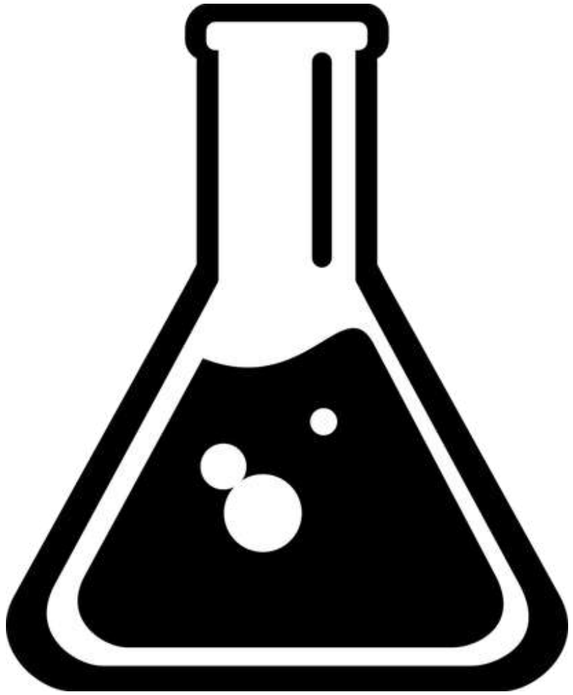
How About HTTP User Agent?

© 2025 SPLUNK LLC

Classification

Lab 3

© 2025 SPLUNK LLC



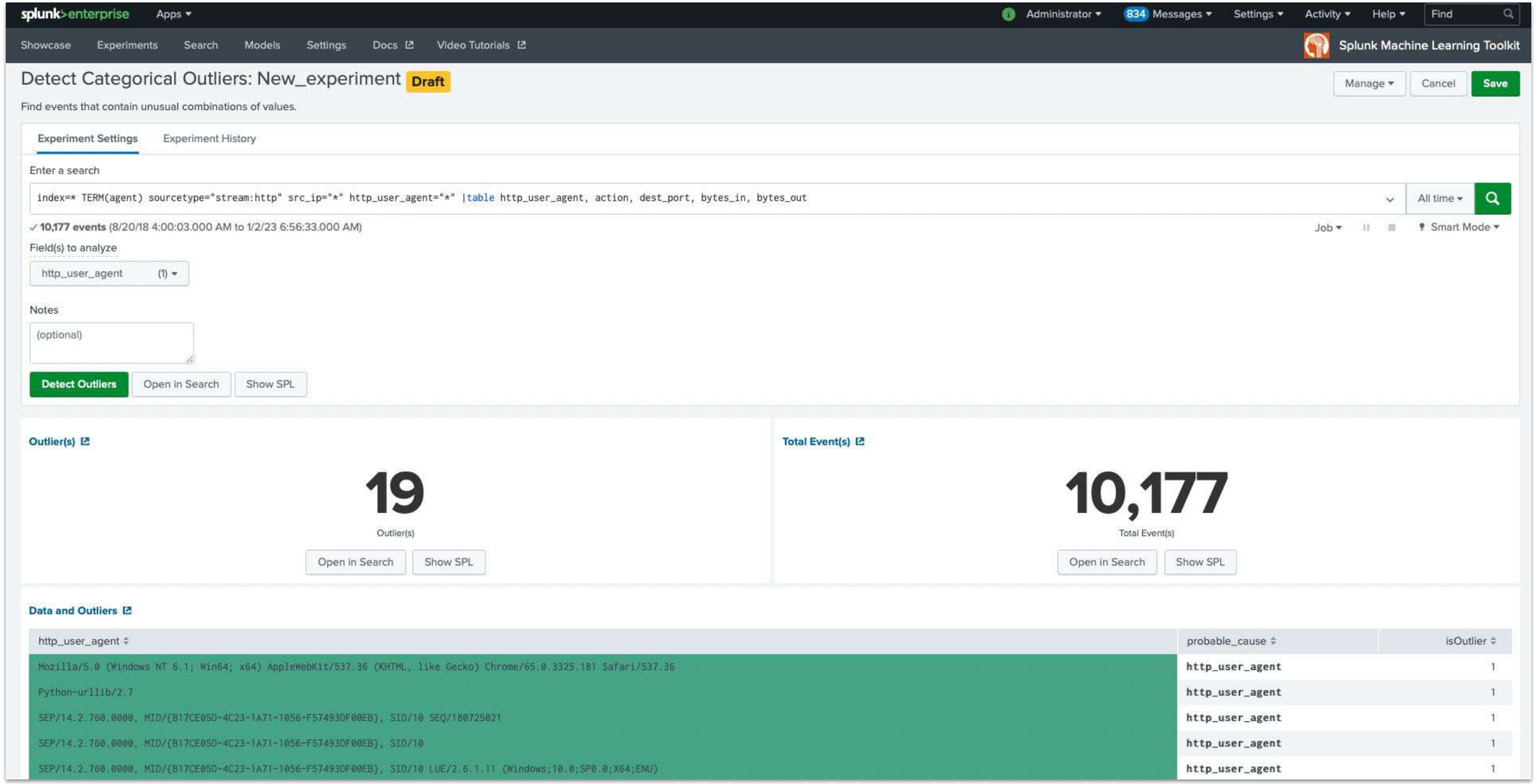
Let’s experiment!

© 2025 SPLUNK LLC

Detecting Categorical Outliers

© 2025 SPLUNK LLC

Detecting Categorical Outliers



© 2025 SPLUNK LLC

Detecting Categorical Outliers

- Try to run the same search but change the field to analyze (e.g. bytes in, bytes out…)
- What results do you get?
- Do you think it’s a good approach?
- Let’s try a different approach and use Clustering!

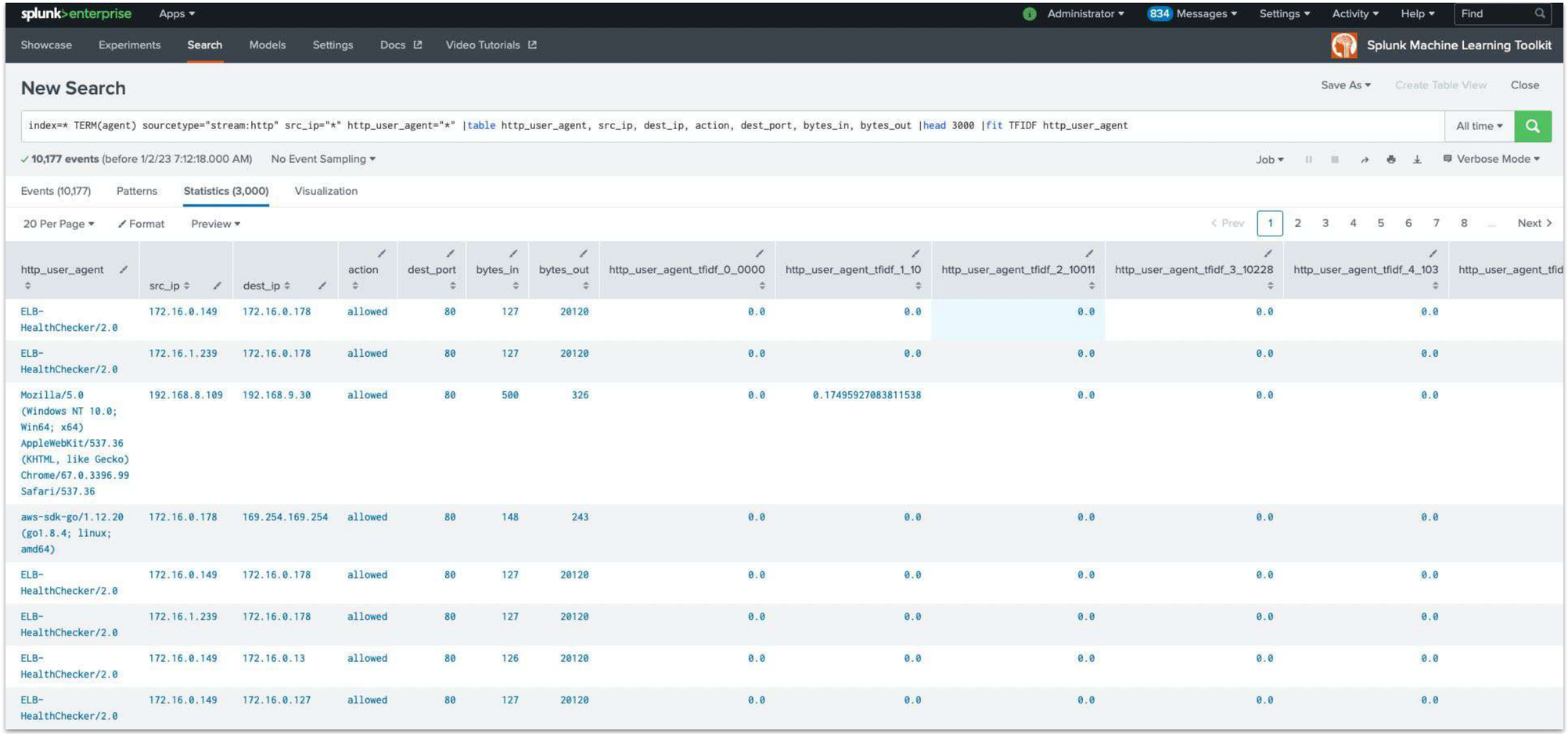
© 2025 SPLUNK LLC

Clustering

© 2025 SPLUNK LLC

Detecting Categorical Outliers

Preprocessing your data - let's play with TFIDF

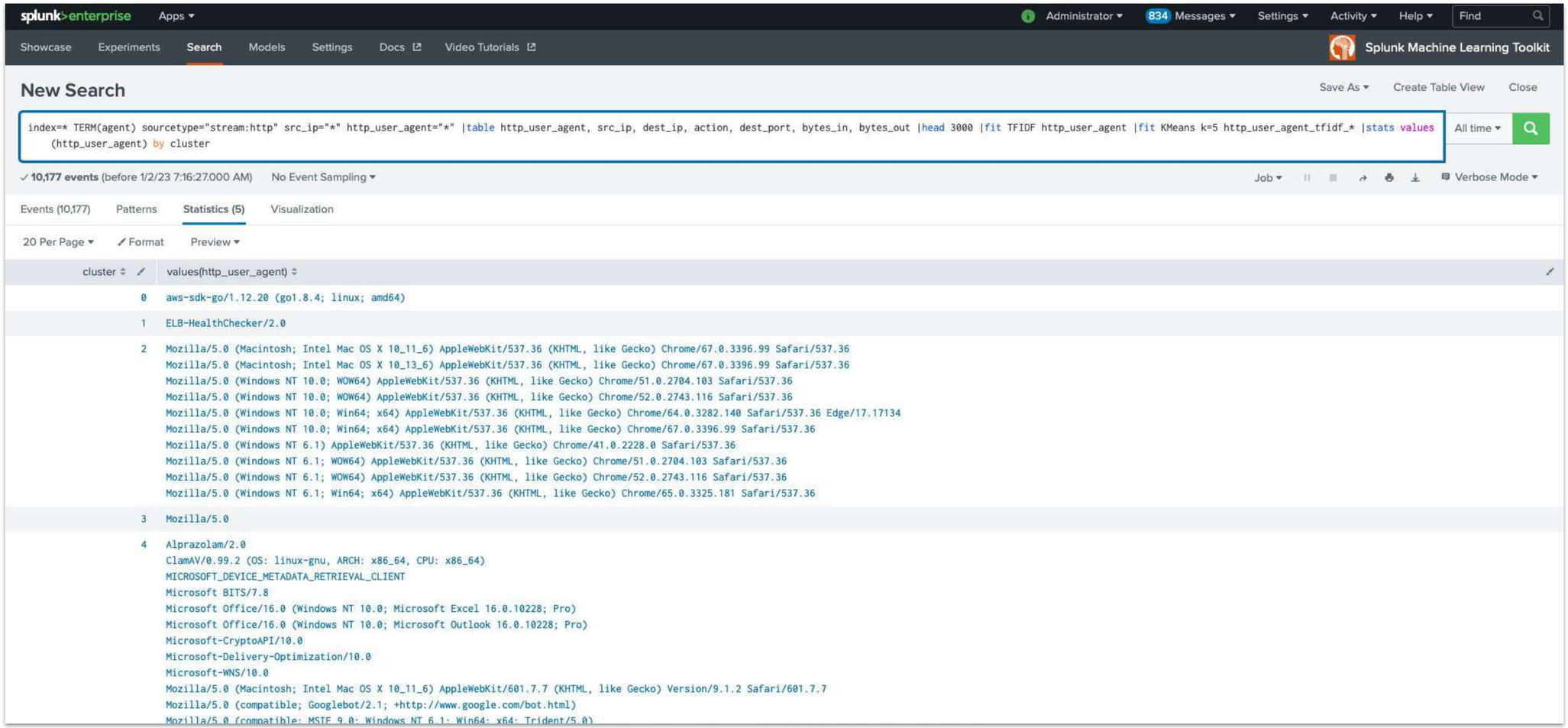


http_user_agent	src_ip	dest_ip	action	dest_port	bytes_in	bytes_out	http_user_agent_tfidf_0_0000	http_user_agent_tfidf_1_10	http_user_agent_tfidf_2_10011	http_user_agent_tfidf_3_10228	http_user_agent_tfidf_4_103	http_user_agent_tfid
ELB-HealthChecker/2.0	172.16.0.149	172.16.0.178	allowed	80	127	20120	0.0	0.0	0.0	0.0	0.0	
ELB-HealthChecker/2.0	172.16.1.239	172.16.0.178	allowed	80	127	20120	0.0	0.0	0.0	0.0	0.0	
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/67.0.3396.99 Safari/537.36	192.168.8.109	192.168.9.30	allowed	80	500	326	0.0	0.17495927083811538	0.0	0.0	0.0	
aws-sdk-go/1.12.20 (go1.8.4; linux; amd64)	172.16.0.178	169.254.169.254	allowed	80	148	243	0.0	0.0	0.0	0.0	0.0	
ELB-HealthChecker/2.0	172.16.0.149	172.16.0.178	allowed	80	127	20120	0.0	0.0	0.0	0.0	0.0	
ELB-HealthChecker/2.0	172.16.1.239	172.16.0.178	allowed	80	127	20120	0.0	0.0	0.0	0.0	0.0	
ELB-HealthChecker/2.0	172.16.0.149	172.16.0.13	allowed	80	126	20120	0.0	0.0	0.0	0.0	0.0	
ELB-HealthChecker/2.0	172.16.0.149	172.16.0.127	allowed	80	127	20120	0.0	0.0	0.0	0.0	0.0	

© 2025 SPLUNK LLC

Detecting Categorical Outliers

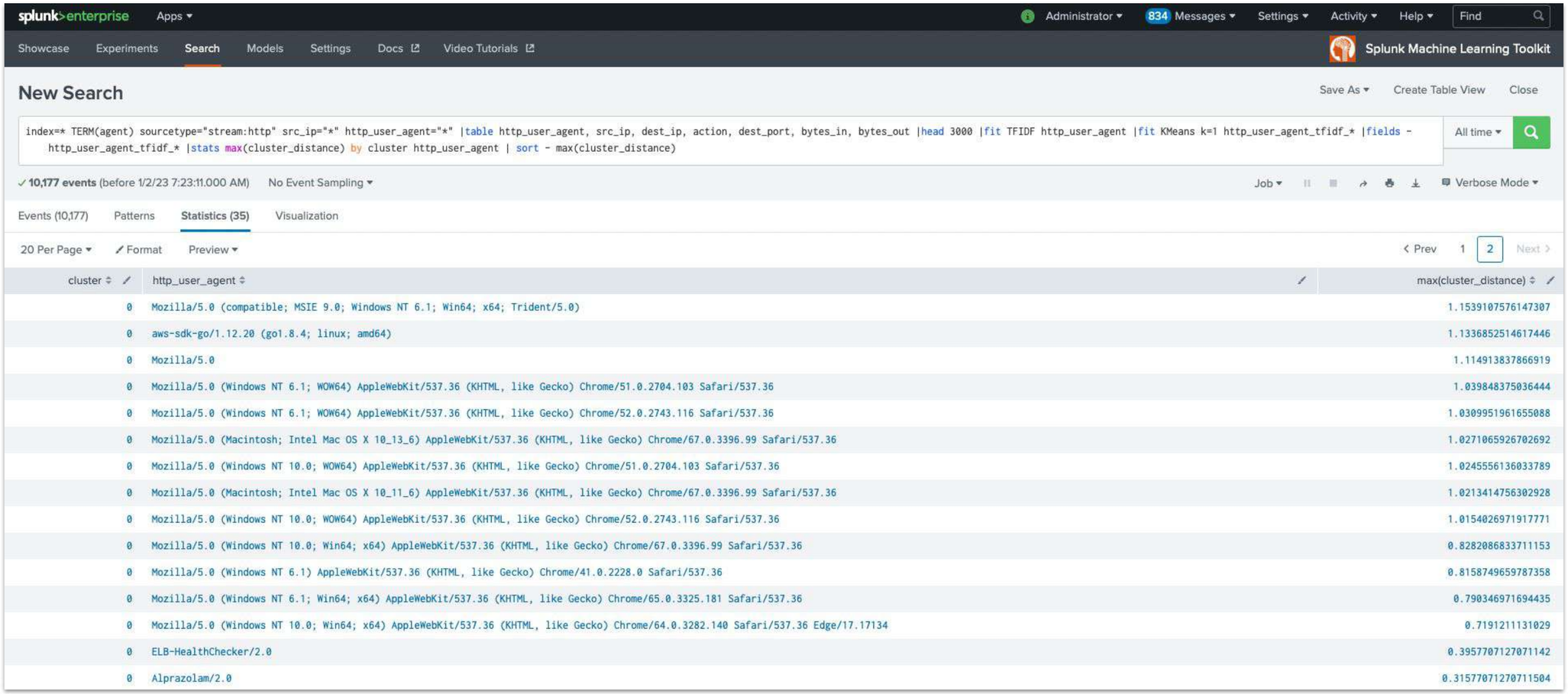
Now let's use clustering!



© 2025 SPLUNK LLC

Detecting Categorical Outliers

Let’s try something else!



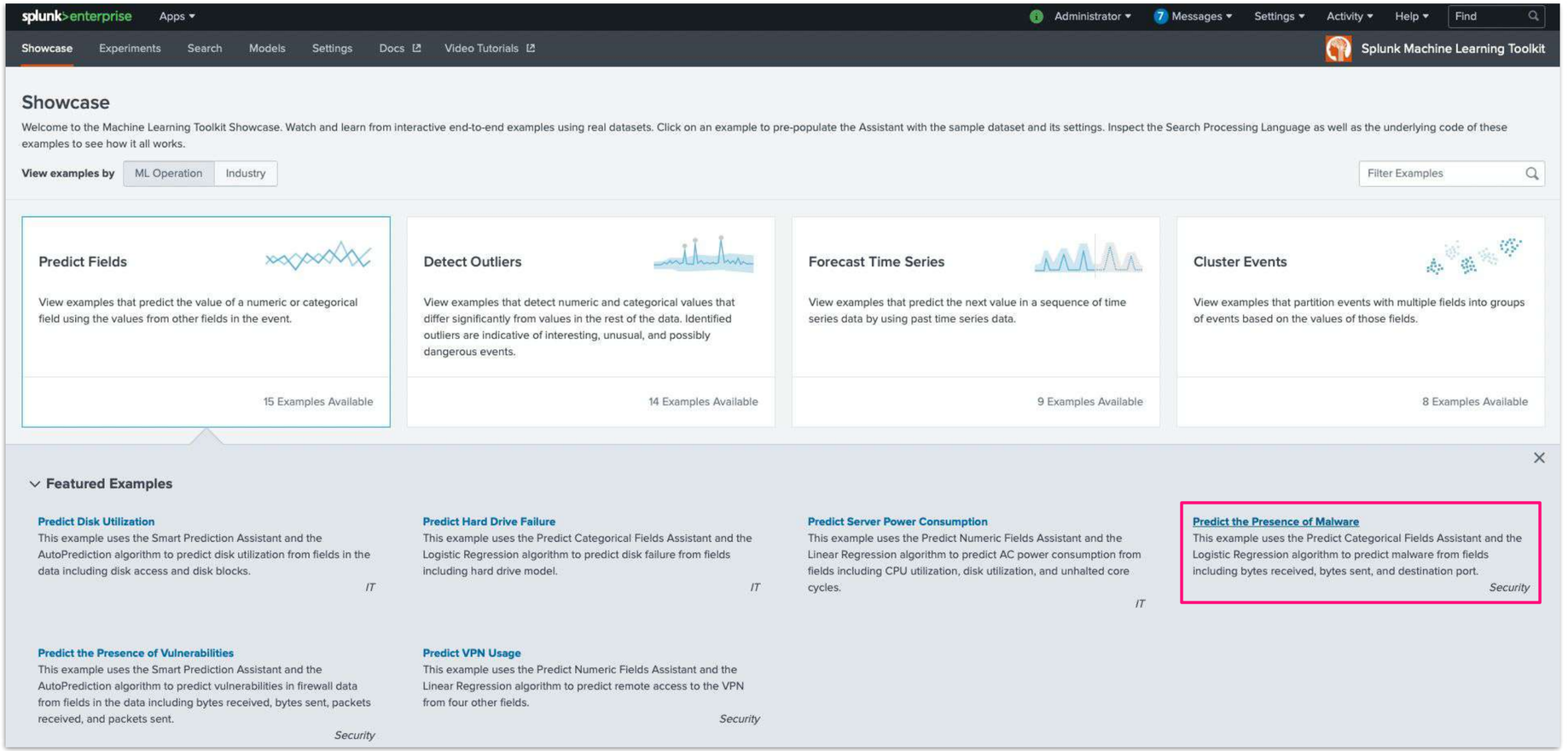
© 2025 SPLUNK LLC

Predict *Categorical Fields*

© 2025 SPLUNK LLC

Predict Categorical Fields

- Navigate to your Splunk AI Toolkit application
- Click on “showcase”, choose ”Predict Fields”, and select “Predict the Presence of Malware”



© 2025 SPLUNK LLC

Predict Categorical Fields

- Check the data in the lookup file “firewall_traffic.csv”
- See what field are being used for our prediction
- Try the different algorithms and see which one if the more precise
- Why? Understand the logic behind the algorithms.

© 2025 SPLUNK LLC

Brainstorming time!

© 2025 SPLUNK LLC

Predict Categorical Fields

- Think about some use cases where prediction would be useful
- Better if it is Security Related
- Did you come up with something?

© 2025 SPLUNK LLC

Anomaly Detection

© 2025 SPLUNK LLC

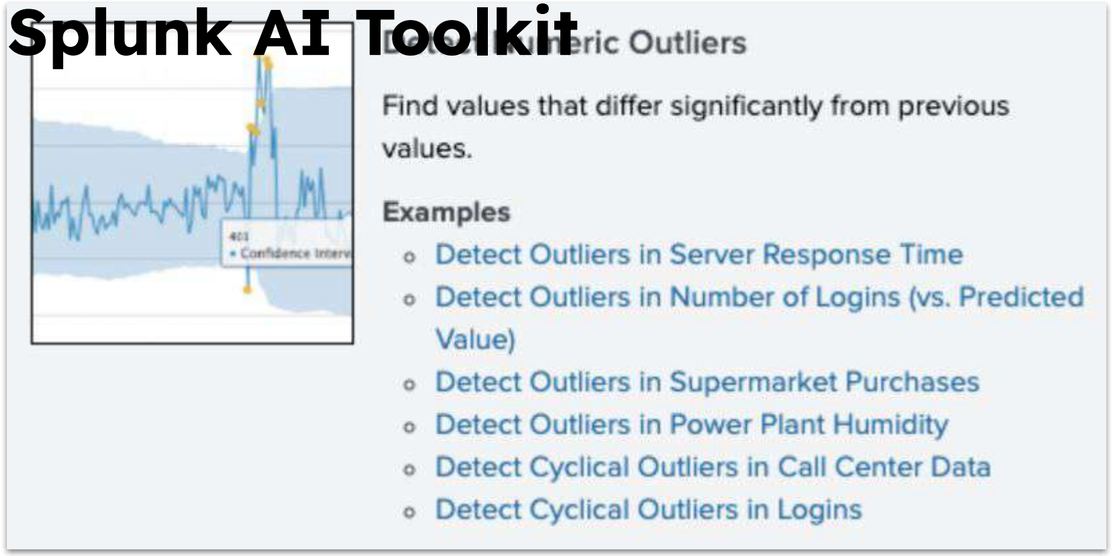
Numeric Outlier Detection

A type of Anomaly Detection

Outlier in some numerical value

- Number of transactions
- Transaction latency
- System utilization (CPU/memory)
- Number of logins
- Amount of data transfer
- Time between actions
- Sensor measurement

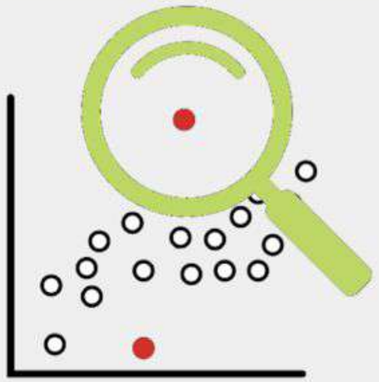
Detect Numeric Outliers Assistant in Splunk AI Toolkit



© 2025 SPLUNK LLC

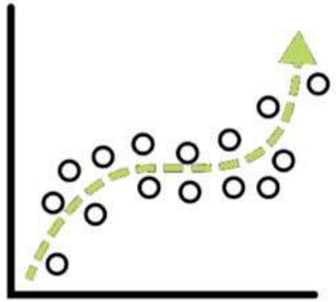
Anomaly Detection

Anomaly detection



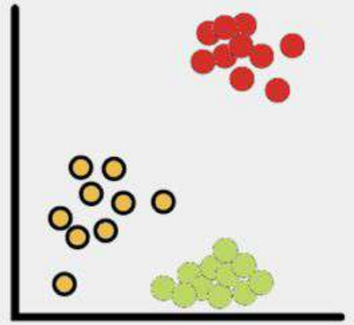
- Deviation from past behavior
- Deviation from peers
- (aka Multivariate AD or Cohesive AD)
- Unusual change in features

Predictive Analytics



- Predict Service Health Score/Churn
- Predicting Events
- Trend Forecasting
- Detecting influencing entities
- Early warning of failure

Clustering



- Identify peer groups
- Event Correlation
- Reduce alert noise
- Behavioral Analytics

© 2025 SPLUNK LLC

Outlier vs Anomaly

What's the difference?

Outlier (noun):

- A statistical observation that is markedly different in value from the others of the sample
- Outliers are usually something unexpected in a single value or measurement

Anomaly (noun):

- Something different, abnormal, peculiar or not easily classified
- Deviation from the common rule
- Anomalies are usually based on a collection of observations

Not all outliers are anomalies, but anomalies often consist of outliers or are outliers

© 2025 SPLUNK LLC

Cheat sheet for anomaly detection in Splunk

Command	Description
analyzefields (af)	Analyze numerical fields for their ability to predict another discrete field.
anomalies	Computes an "unexpectedness" score for an event.
anomalousvalue	Finds and summarizes irregular, or uncommon, search results.
anomalydetection	Identifies anomalous events by computing a probability for each event and then detecting unusually small probabilities.
cluster	Clusters similar events together.
kmeans	Performs k-means clustering on selected fields.
outlier	Removes outlying numerical values.
rare	Displays the least common values of a field.

https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Commandsbycategory

Method / Algorithm	Description
DensityFunction	The DensityFunction algorithm provides a consistent and streamlined workflow to create and store density functions and utilize them for anomaly detection…
LocalOutlierFactor	The LocalOutlierFactor algorithm measures the local deviation of density of a given sample with respect to its neighbors…
OneClassSVM	The OneClassSVM algorithm fits a model from a set of features or fields for detecting anomalies and outliers…
Clustering Algorithms	Spot point anomalies or anomaleous clusters. Inspect e.g. cluster_distance with KMeans, cluster=-1 with DBSCAN…
Classifiers and Regressors	Inspect strong residuals when applying your well fitted model to new incoming data points.
ML SPL API	Wrap your own algorithms of choice

https://docs.splunk.com/Documentation/MLApp/latest/User/Algorithms

© 2025 SPLUNK LLC

Detecting outliers with Enterprise Security

Enterprise Security ships with Splunk AI Toolkit as of ES version 6.0

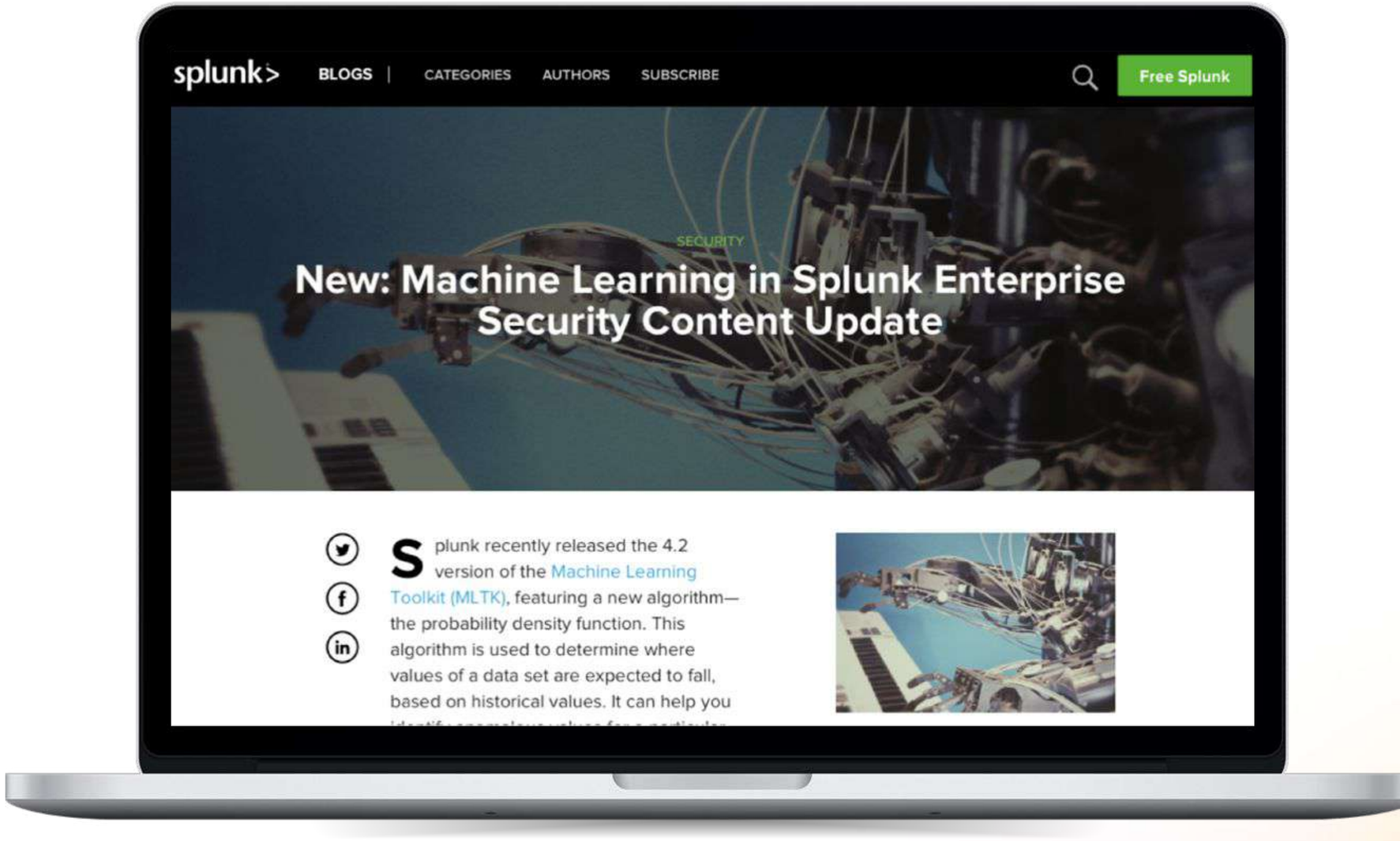
The Splunk AI Toolkit (Formerly MLTK) is replacing Extreme Search (XS) as a model generation package in Enterprise Security (ES). Splunk AI Toolkit can scale at larger volume and also can identify more abnormal events through its models (see Welcome to the Machine Learning Toolkit in the Splunk Machine Learning Toolkit User Guide.)

ES is using Splunk AI Toolkit DensityFunction under the hood.

https://docs.splunk.com/Documentation/ES/latest/Admin/MLTKoverview

© 2025 SPLUNK LLC

ES Content Update includes Density Function content



© 2025 SPLUNK LLC

Machine Learning Applications from Splunkbase

splunkbase™

© 2025 SPLUNK LLC

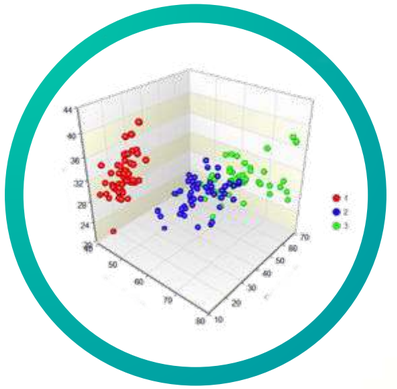
Machine Learning Applications

Splunk Security Essentials

DGA App for Splunk

Botnet App for Splunk

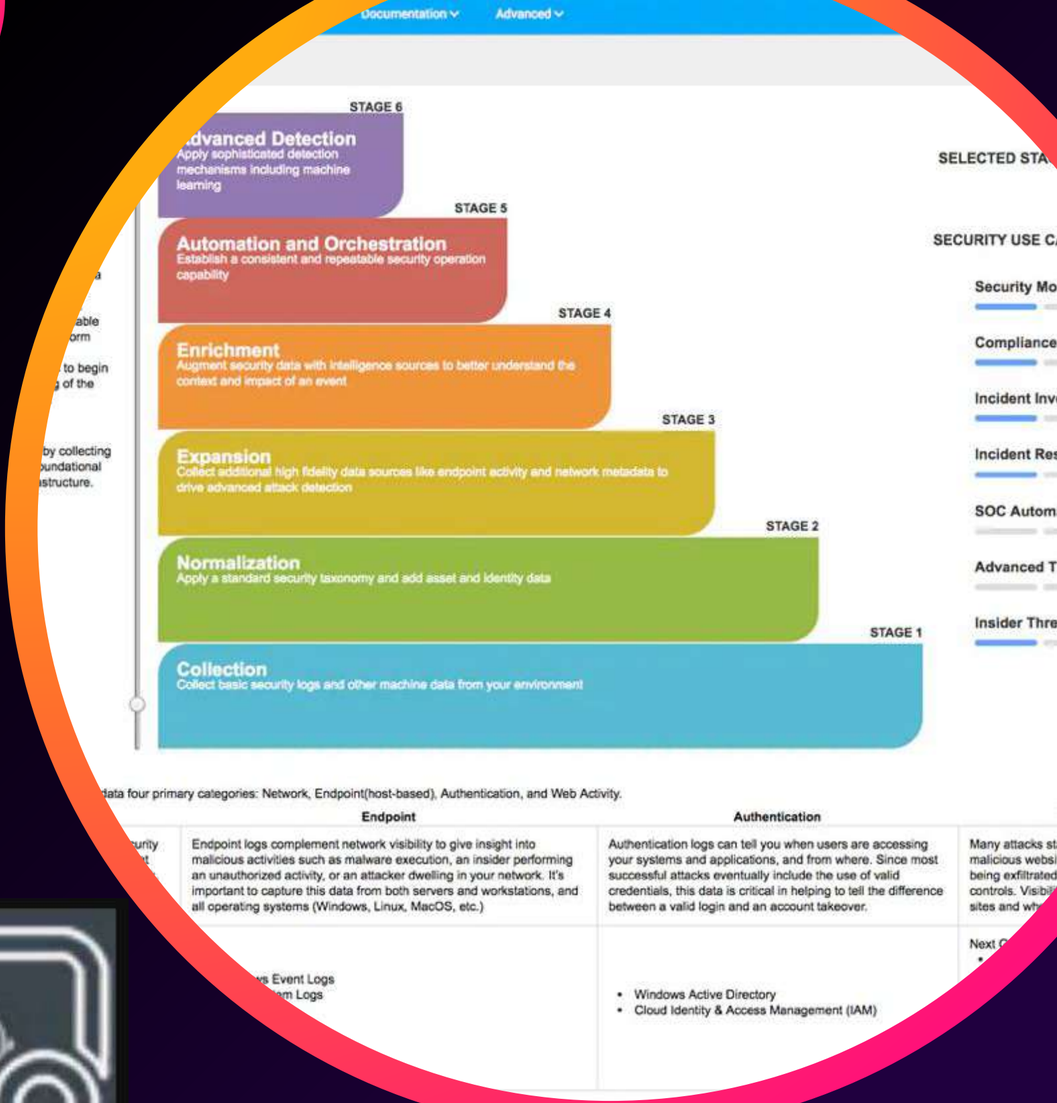
Custom Detections using AI Toolkit



© 2025 SPLUNK LLC

Splunk Security Essentials

Download from Splunkbase:
https://splunkbase.splunk.com/app/3435

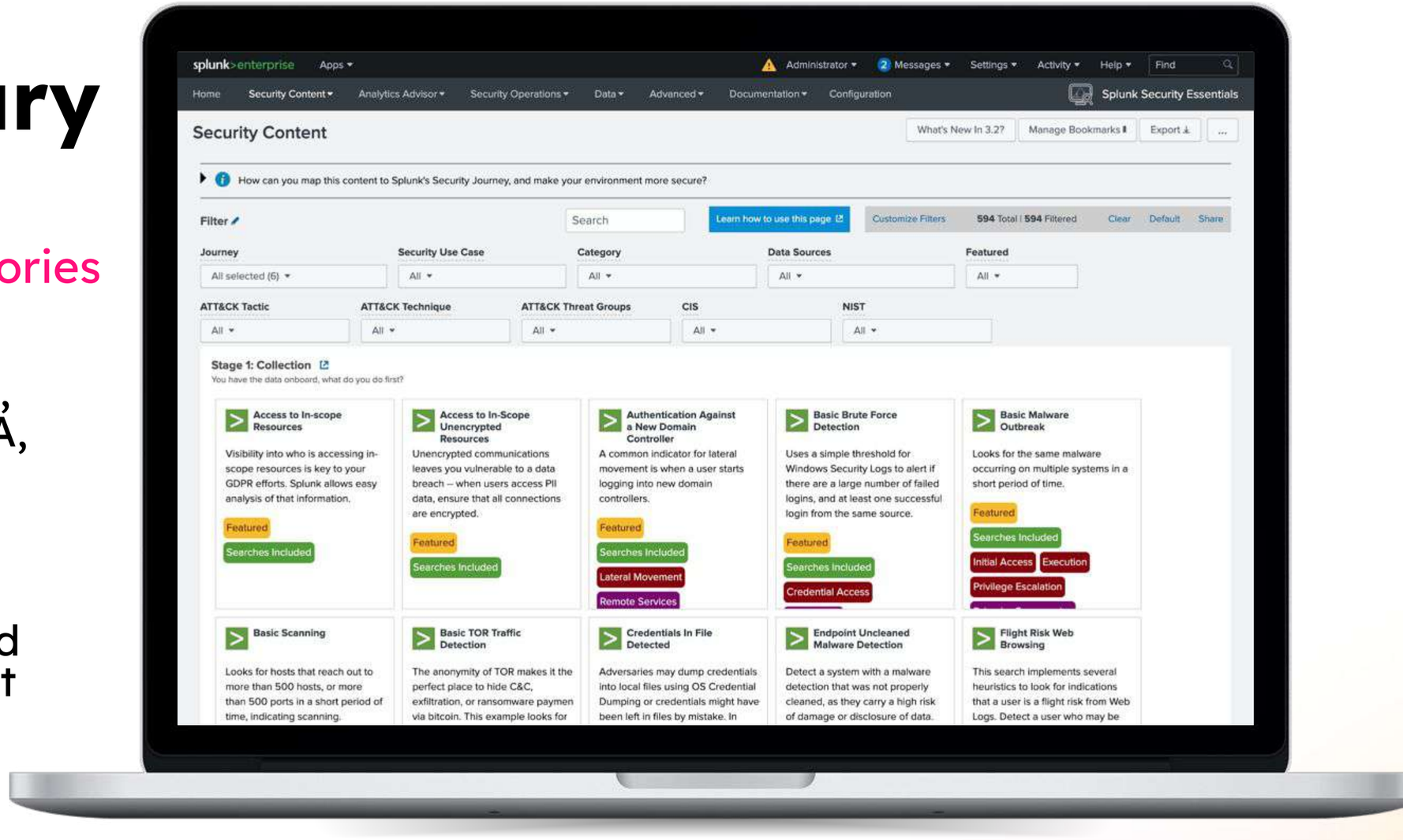


© 2025 SPLUNK LLC

Security Content Library

Browse, bookmark, and deploy 900+ security detections and analytic stories

- Repository of Security Content for Splunk Cloud, Enterprise Security, UEBA, and Phantom
- Deploy security content within clicks
- Enrich notable events and run analytics with context from content library
- Stay up to date on existing and emerging threats



© 2025 SPLUNK LLC

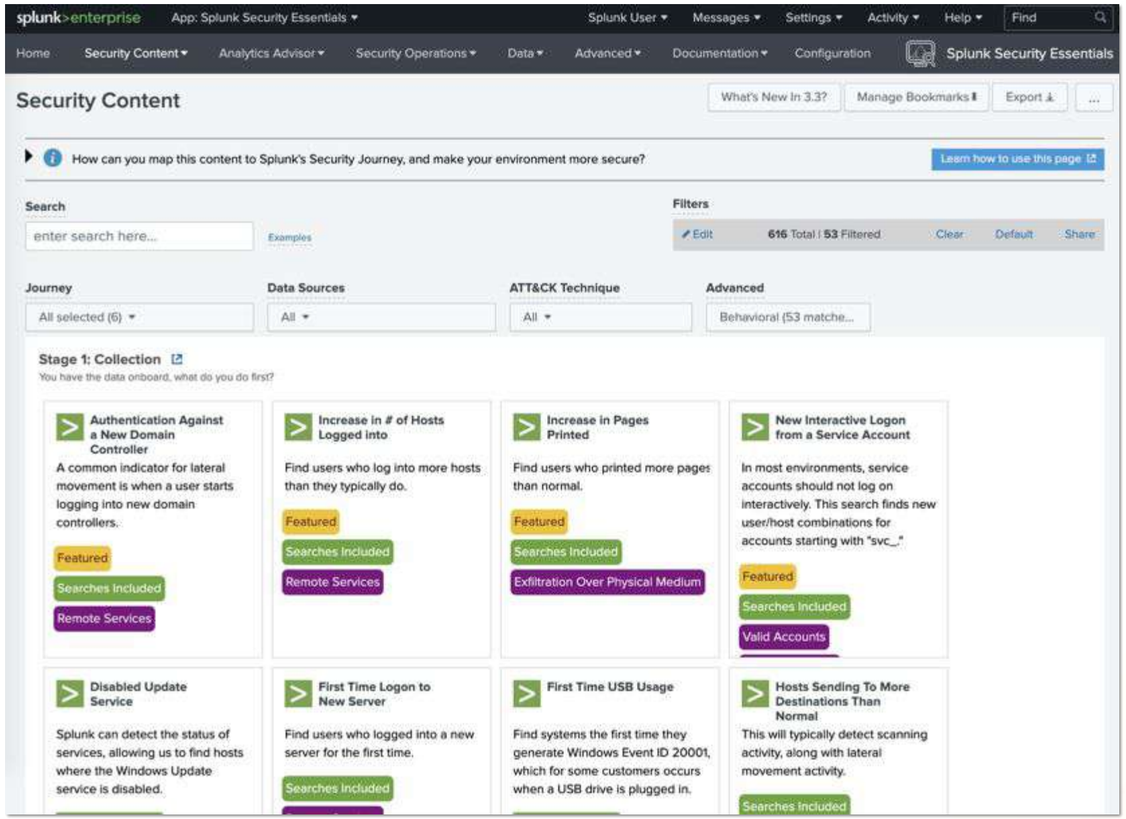
Includes Behavioral Detections

Identify bad guys in your environment:

- Includes 50+ use cases common in UEBA products, all using Splunk Enterprise
- Target external attackers and insider threat
- Scales from small to massive companies
- Save from the app, send results to ES/UBA

The most widely deployed UEBA product in the market is Splunk Enterprise, but no one knows it.

Solve your use cases today, then use Splunk UBA for advanced ML detection.

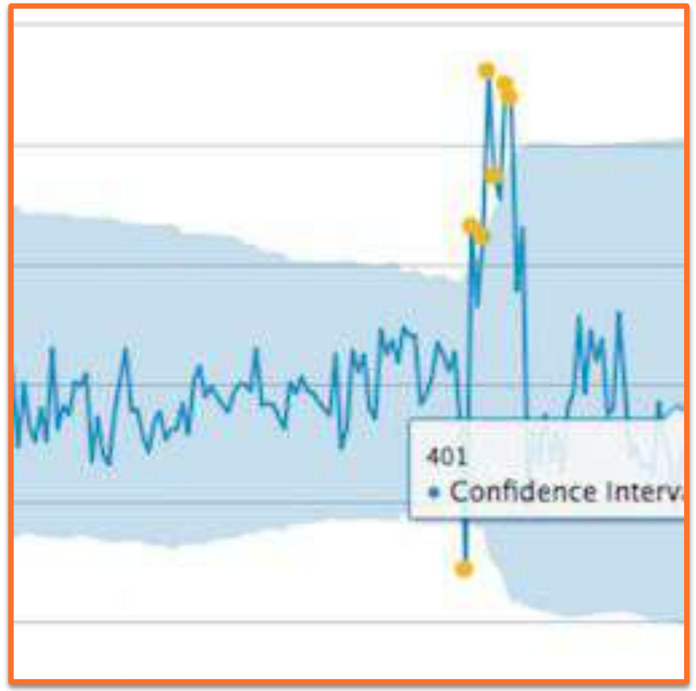


© 2025 SPLUNK LLC

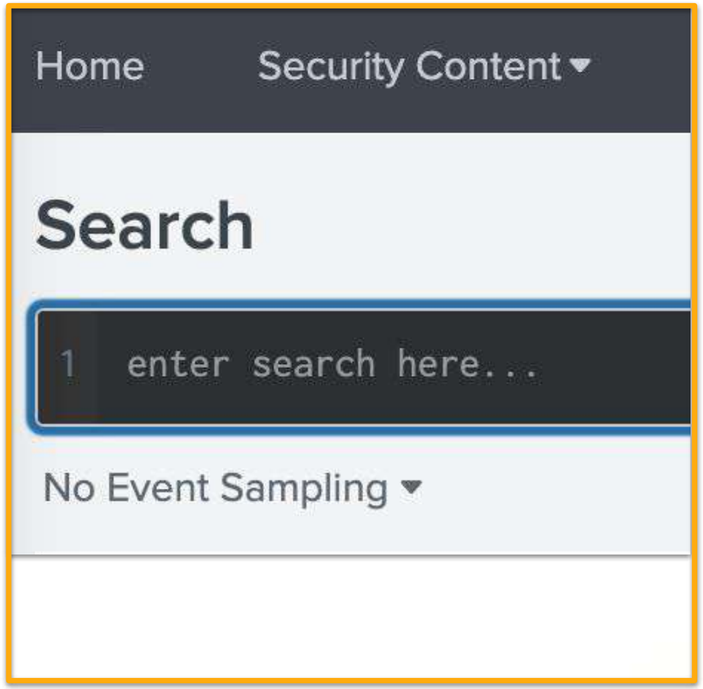
Splunk Security Essentials

Types of Use Cases

First Time Seen powered by stats



Time Series Analysis with Standard Deviation



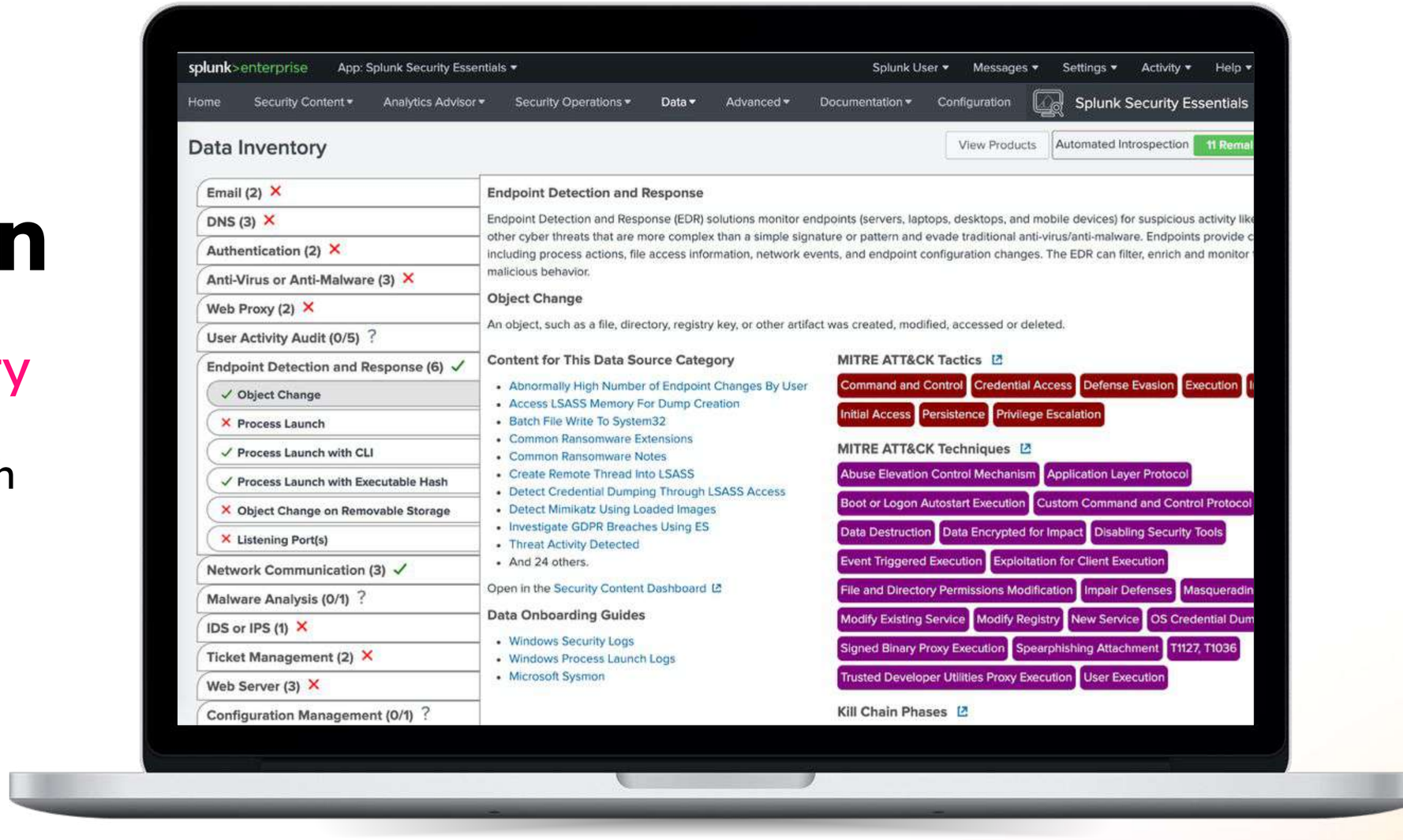
General Security Analytics Searches

© 2025 SPLUNK LLC

Data and Content Introspection

Track data and saved searches to gain visibility

- Enrich saved searches with tags and metadata
- Automatic categorization by security products
- Scale with data model acceleration and by reinforcing CIM compliance

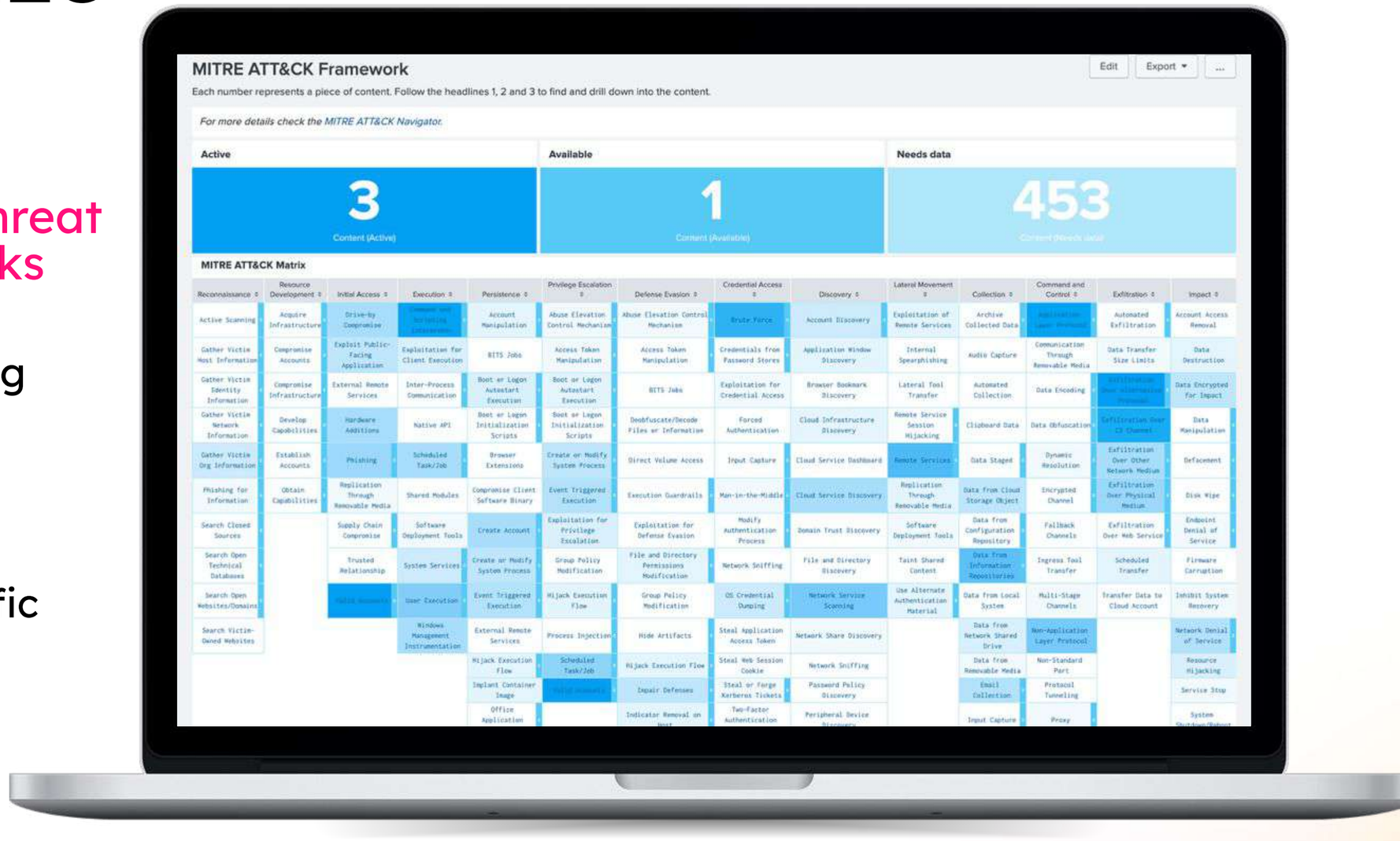


© 2025 SPLUNK LLC

Operationalize Security Frameworks

Identify gaps, improve threat detection, and reduce risks

- Develop an understanding of your security posture against MITRE ATT&CK® and Cyber Kill Chain® frameworks
- Find detections for specific Threat Groups or Threat Software
- Drilldown on known Tactics and Techniques and Kill Chain Phases for more details



© 2025 SPLUNK LLC

DGA App for Splunk

Download from Splunkbase:
https://splunkbase.splunk.com/app/3559/

© 2025 SPLUNK LLC

Domain Generating Algorithms (DGA)

What's the problem?

Challenges to detect DGAs:

- Static matching runs against potentially infinite blacklist entries
- Regex can narrow down this list, but still hard to compute and find rules (and define exceptions for rules)
- Unknown unknowns?
- Want to get fuzzy?
- Let's ML!

Example of DGAs:

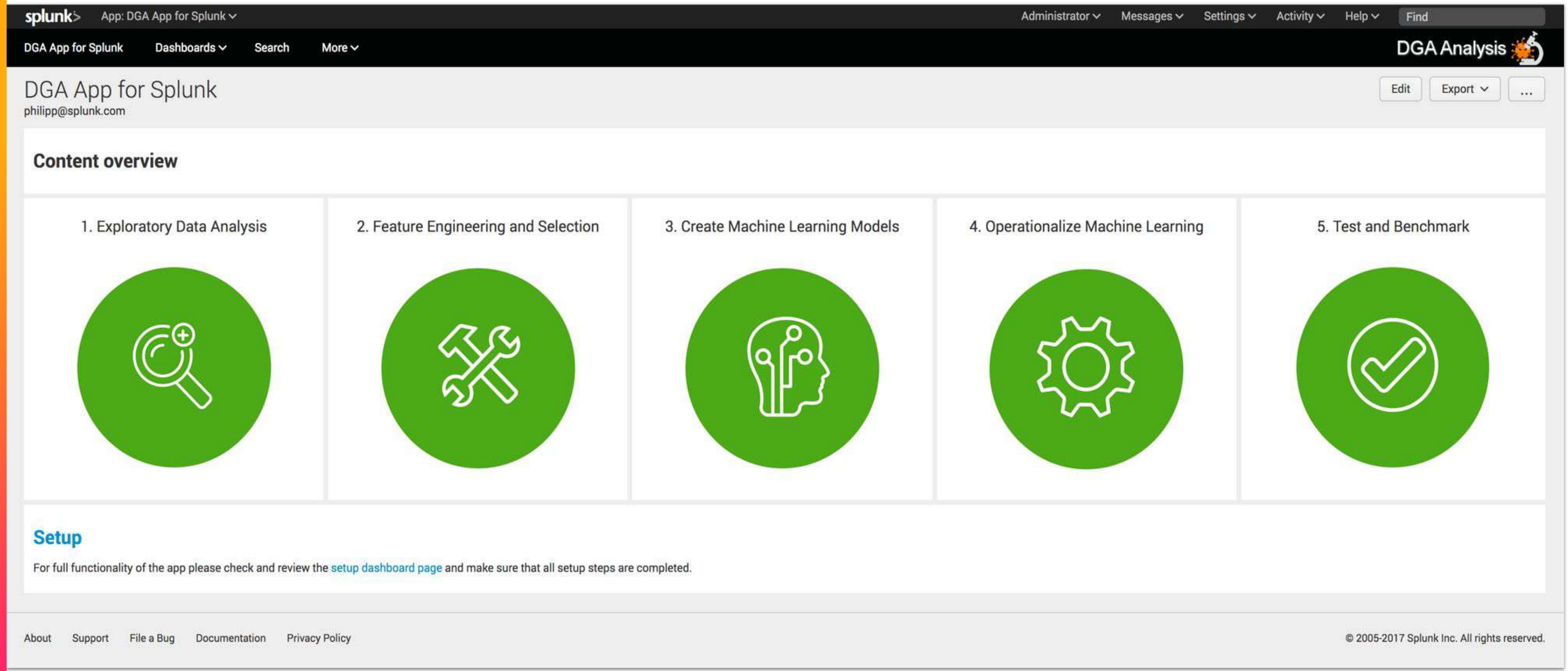
domain
iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea
ifferfsodp9ifjaposdfjhgosurijfaewrwergwea
ayylmaotjhsstasdfasdfasdfasdfasdfasdfasdf
lazarusse.suiche.sdfjhgosurijfaqwqwqrgwea
sdfjhgosurijfaqwqwqrgwea

Example IoCs for Wannacry: https://cert.europa.eu/static/SecurityAdvisories/2017/CERT-EU-SA2017-012.pdf

Example IoCs for Sunburst: https://github.com/fireeye/sunburst_countermeasures/blob/main/indicator_release/Indicator_Release_NBIs.csv

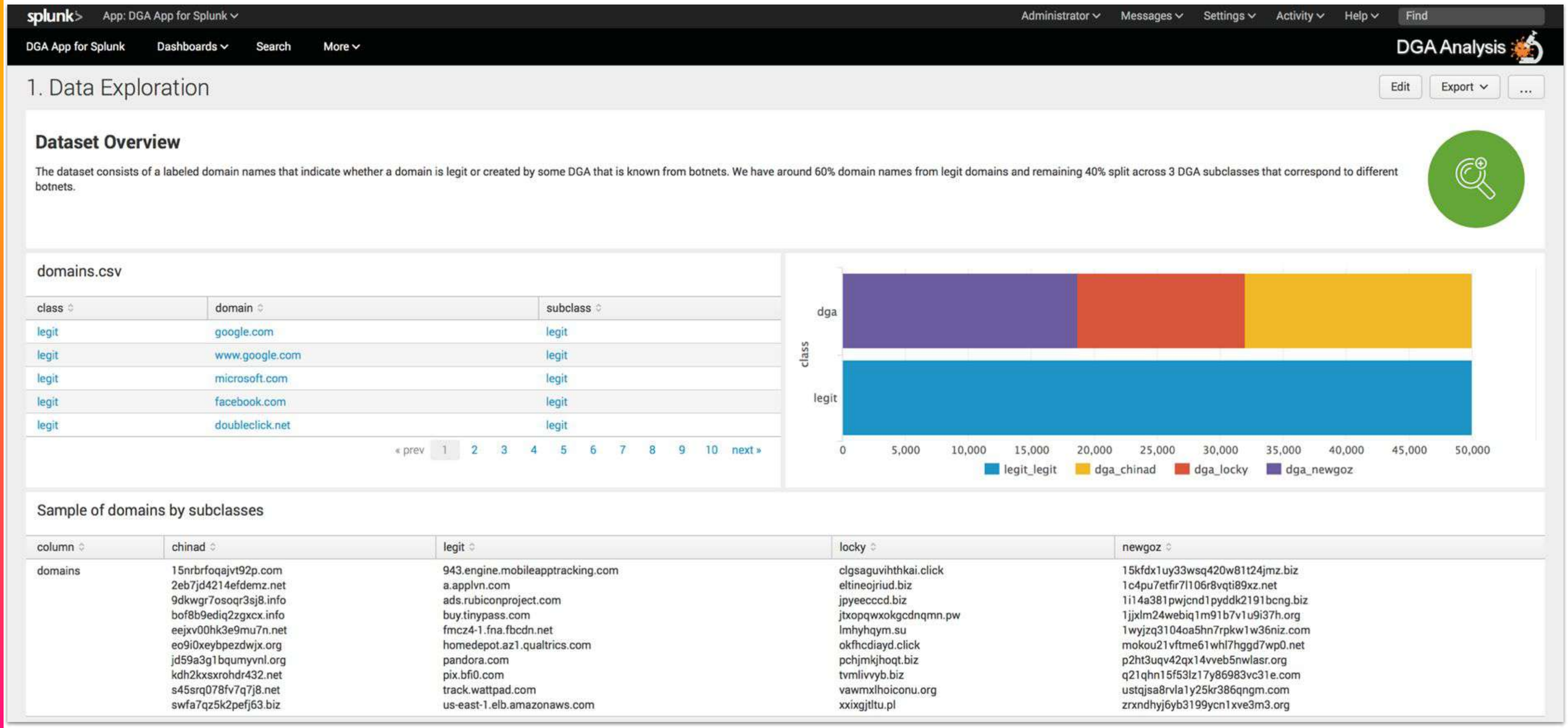
Additional details at https://www.splunk.com/en_us/blog/security/sunburst-backdoor-detections-in-splunk.html

© 2025 SPLUNK LLC



- Example for end to end data science process
- Disclaimer: this is not a turn key solution but a template to get you started
- Feel free to improve and give us your feedback!

© 2025 SPLUNK LLC

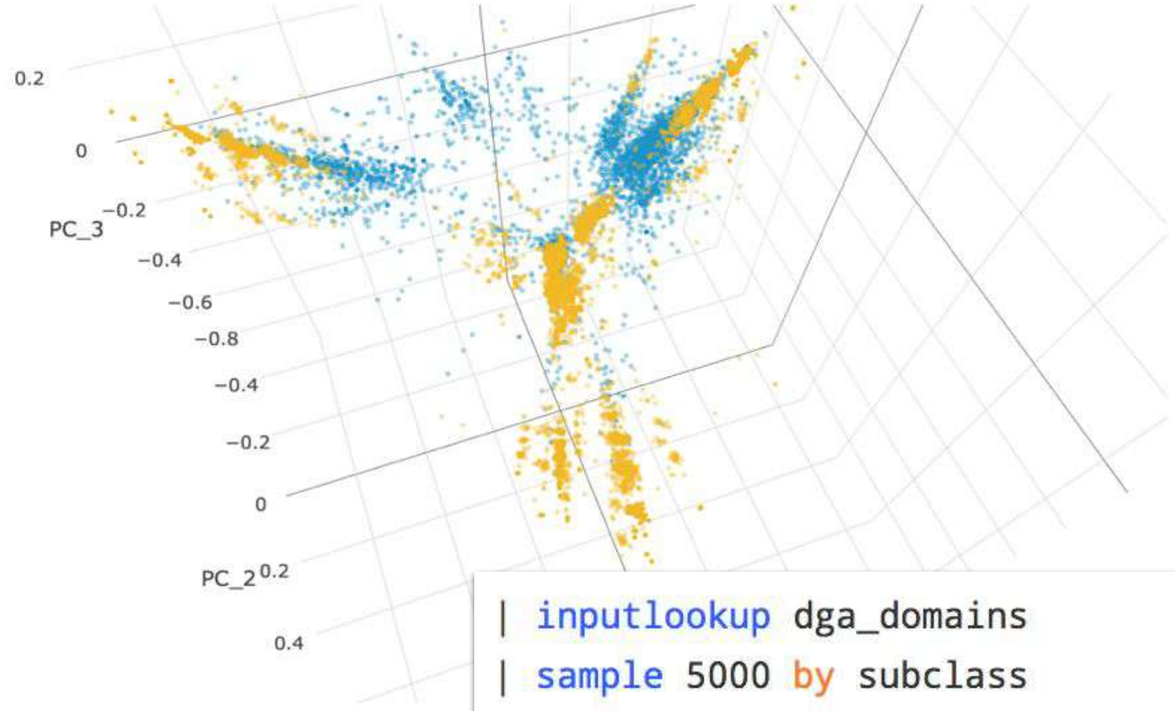


- Small data set with 100K domain names for training and testing

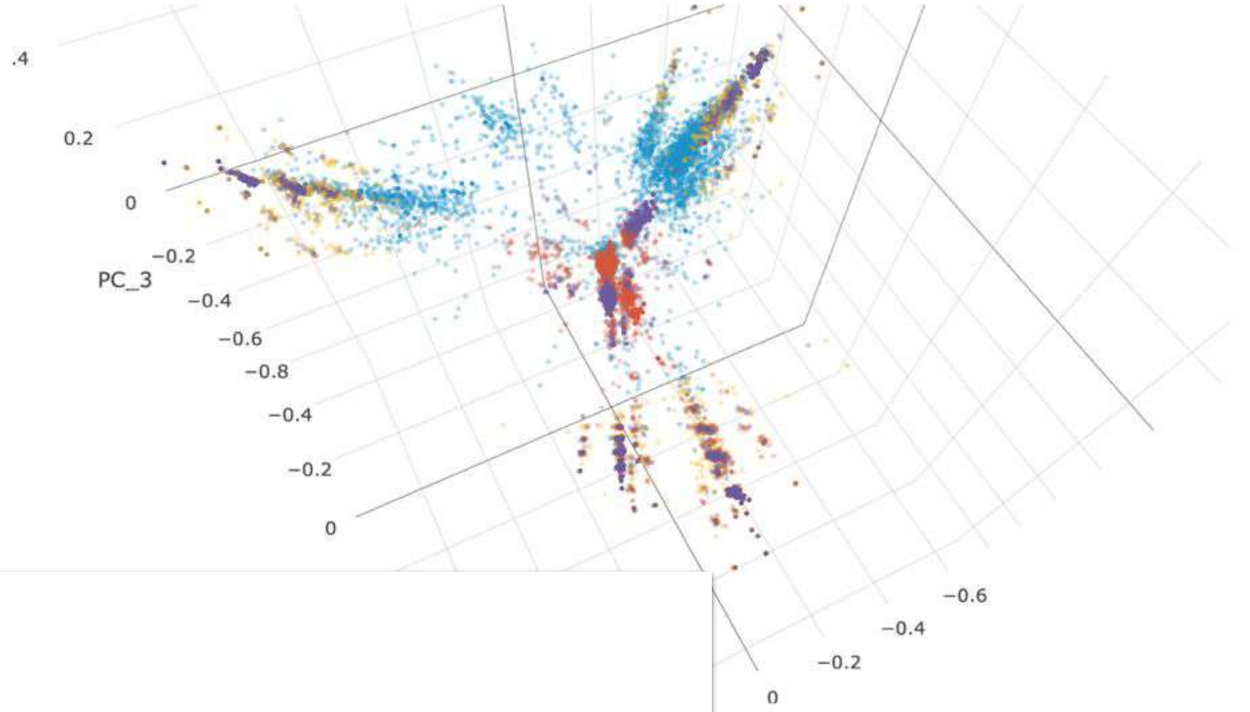
© 2025 SPLUNK LLC

Data Exploration with Text Mining approach

n-gram analysis (2-3 char groups) of domain names with PCA k=3 by class

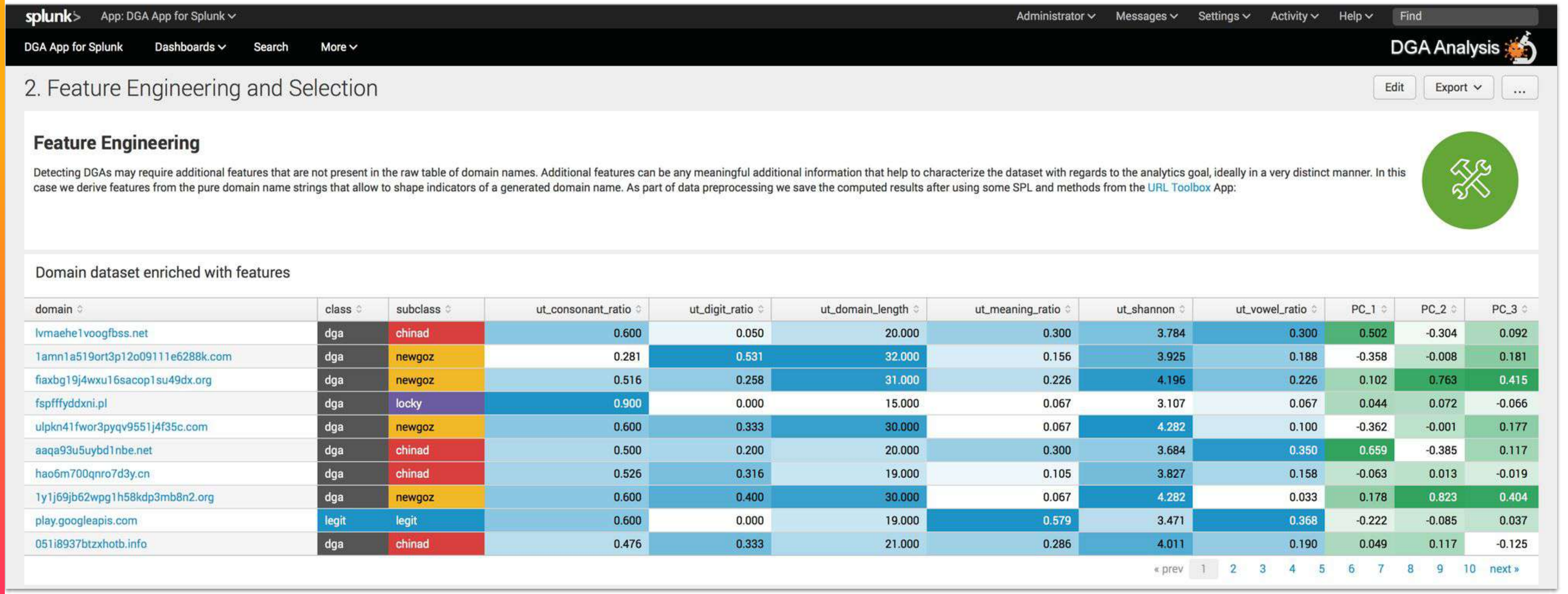


n-gram analysis (2-3 char groups) of domain names with PCA k=3 by subclass




```
| inputlookup dga_domains
| sample 5000 by subclass
| fit TFIDF domain analyzer=char ngram_range=2-3 into "dga_ngram"
| fit PCA domain_tfidf* k=3 into "dga_pca"
| fields - domain_tfidf*
| eval label=class."_".subclass
| table label PC_*
| sort 0 - label
```

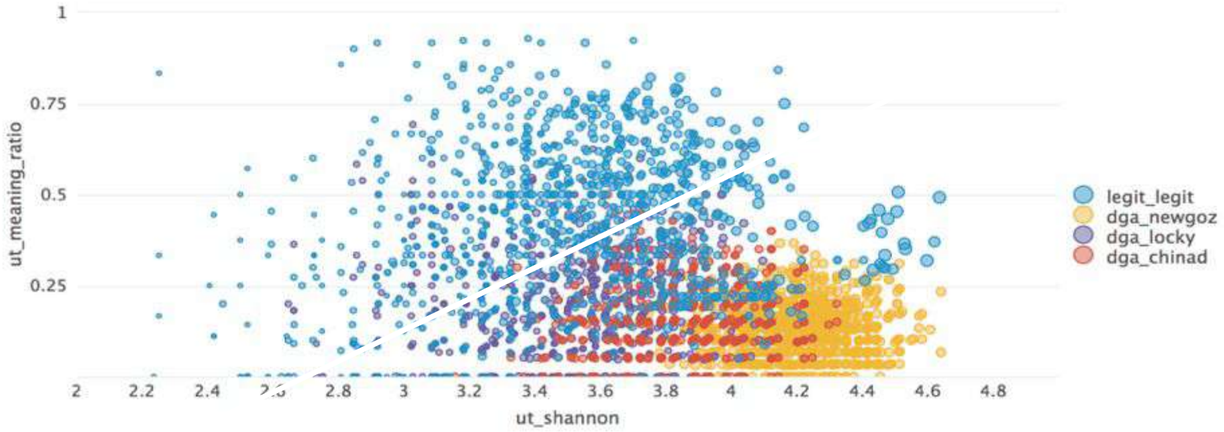
© 2025 SPLUNK LLC



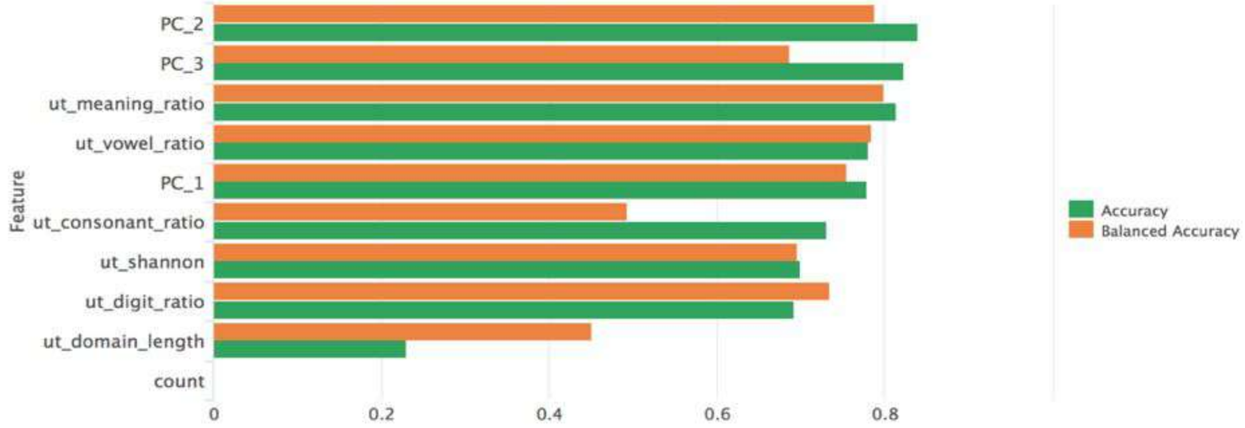
- More features can significantly improve your machine learning models
- Extend this with your feature engineering ideas (e.g. subdomains, age of domain registration, rating/scoring from threatlists for known malicious domains etc.)

© 2025 SPLUNK LLC

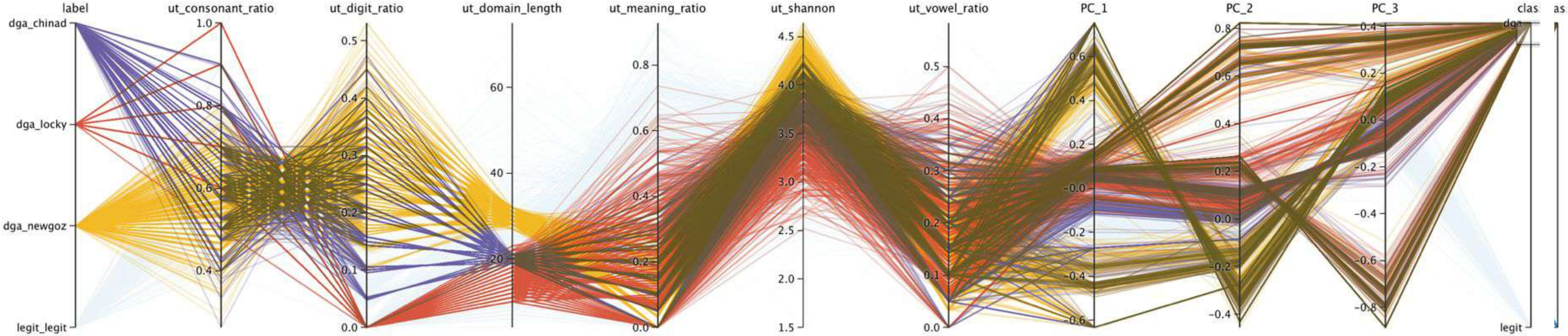
Distribution of classes depending on example feature combination



Identify useful features for classification with the analyzefields command

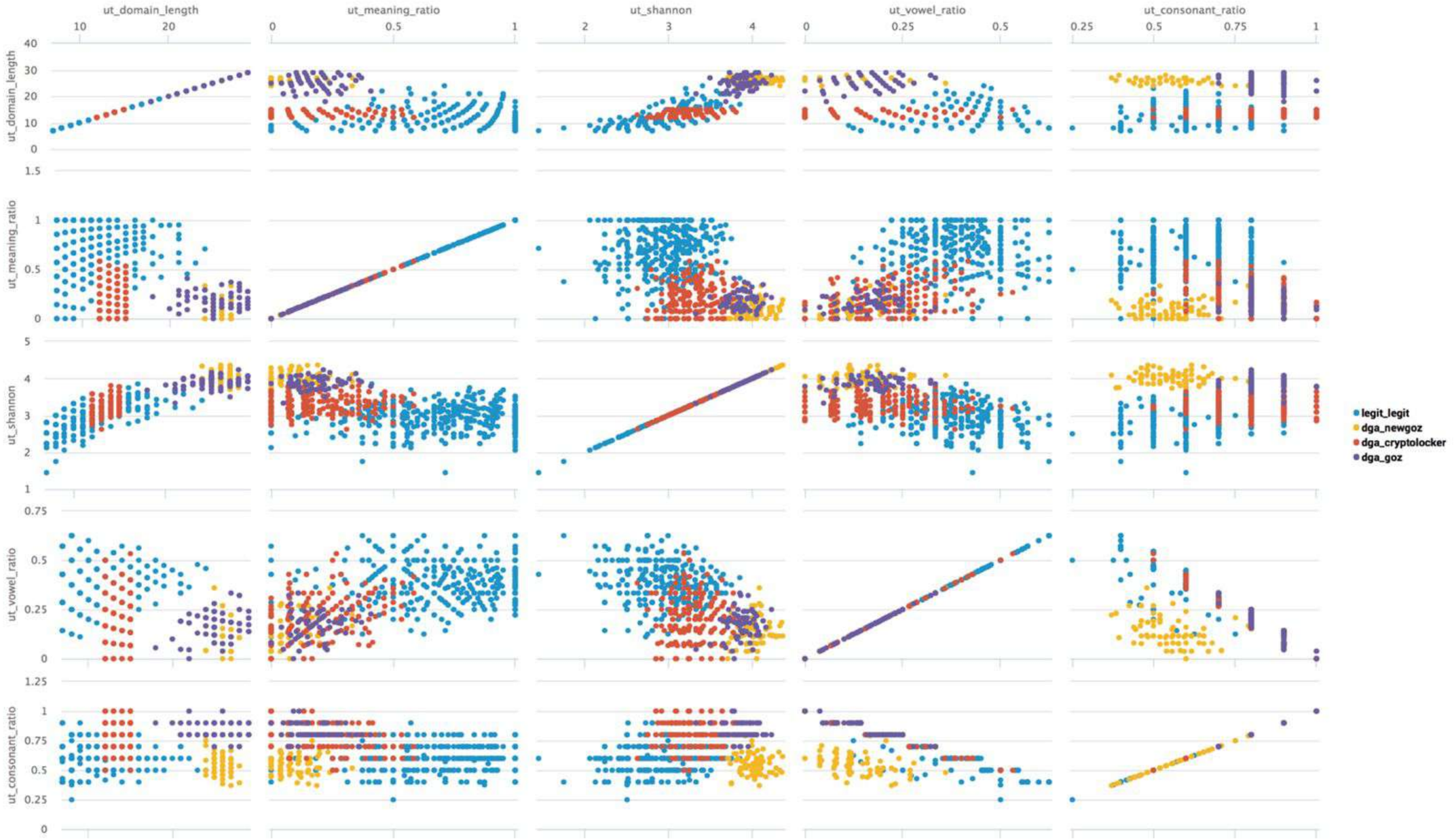


Parallel coordinate chart of classes and top features



Currently showing 3000 / 4000 datapoints Clear filters

© 2025 SPLUNK LLC



© 2025 SPLUNK LLC

3. Create Machine Learning Models

Edit Export ...

Training and evaluation of different machine learning models

We train 4 machine learning models on the same data set using different algorithms for classification. Using a 50:50 split we can evaluate which models perform better and have a lower error rate.

Train models

1. Random Forest Classifier

class	predicted(class)=dga	predicted(class)=legit
dga	24693	226
legit	392	24681

2. Support Vector Machine

class	predicted(class)=dga	predicted(class)=legit
dga	23988	931
legit	1171	23902

3. Decision Tree Classifier

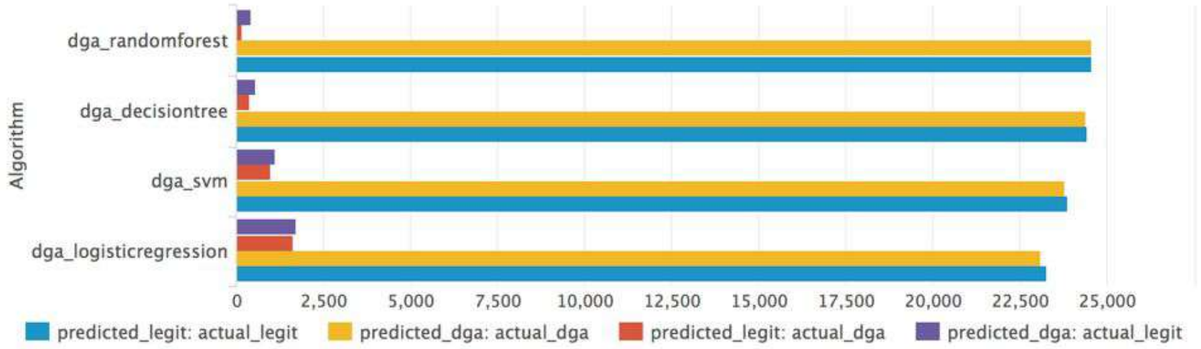
class	predicted(class)=dga	predicted(class)=legit
dga	24512	407
legit	524	24549

4. Logistic Regression

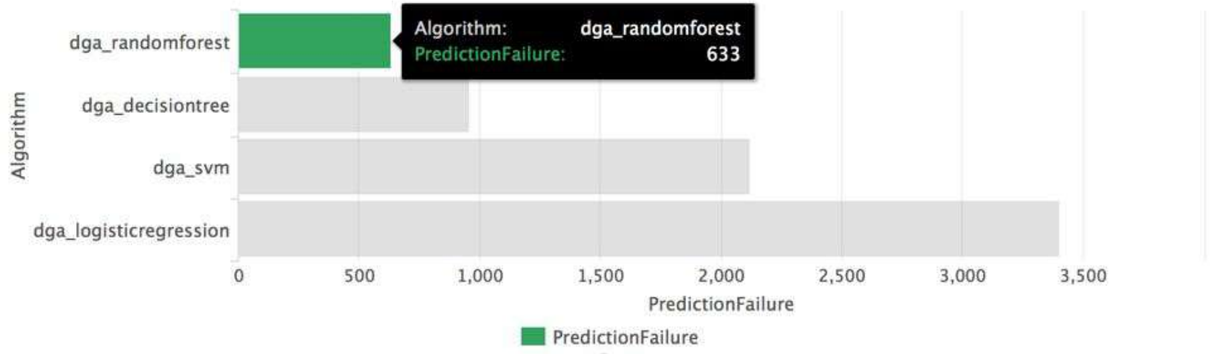
class	predicted(class)=dga	predicted(class)=legit
dga	23254	1665
legit	1686	23387

Evaluate models

Compare and evaluate models



False predictions



© 2025 SPLUNK LLC

4. Operationalize Machine Learning

Edit | Export | ...

Machine Learning Algorithm: RandomForest

Timerange: Custom time | Submit | Hide Filters

Operationalize machine learning

As successful machine learning depends on a continuous process we constantly evaluate the results. Furthermore we can close the feedback loop and append our evaluations to our training dataset to keep models always up to date based on latest information.

Setup notes:

1. Create an index that holds domain names and computed features (we used a index named "dga_proxy")
2. Activate scheduled searches (app menu: More > Alerts) to generate sample data and fill this index.
3. Check the macro `domain_input` in Settings > Advanced Search if you have custom naming

If you want to test on your domain data please adjust your data flow to this mechanism. Of course you can also take domain name data from CIM data models in Enterprise Security and integrate into this mechanism.

Count of predictions

6

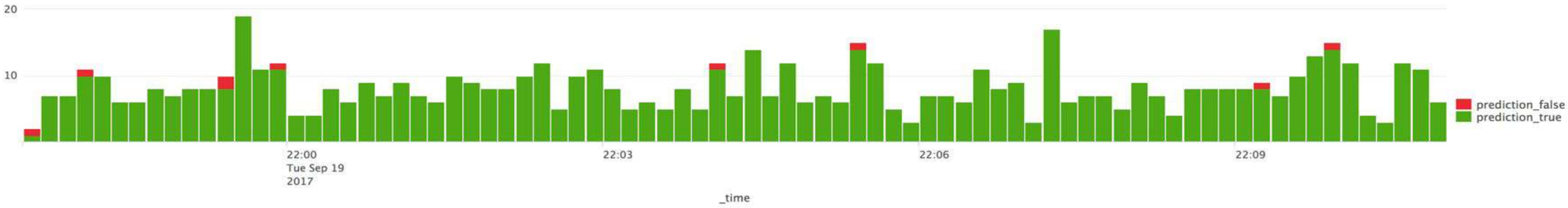
Trend of true predictions

6 -8

Trend of false predictions

0 -1

Prediction performance over time



Results of machine learning algorithm (dga_randomforest) applied to new domains

_time	domain	class	predicted(class)
2017-09-19 22:10:59	mrop8i1scak54s2hskfpq443z.org	dga	dga
2017-09-19 22:10:58	stackadapt.com	legit	legit
2017-09-19 22:10:57	cf.dropboxstatic.com	legit	legit

Results of machine learning algorithm (dga_randomforest) with DGA detected

_time	domain	class
2017-09-19 22:10:59	mrop8i1scak54s2hskfpq443z.org	dga
2017-09-19 22:10:54	8fjsofjlapjjxhcm.org	dga
2017-09-19 22:10:52	g5te08189ly791v20rrr53sdsv.org	dga

© 2025 SPLUNK LLC

Checklist to manually adjust the results

As successful machine learning depends on a continuous process we constantly evaluate the results. Furthermore we can close the feedback loop and append our evaluations to our training dataset to keep models always up to date based on latest information.

Trend of domains classified as DGA

2,392 → 0

Trend of domains manually classified as LEGIT

4 → 0

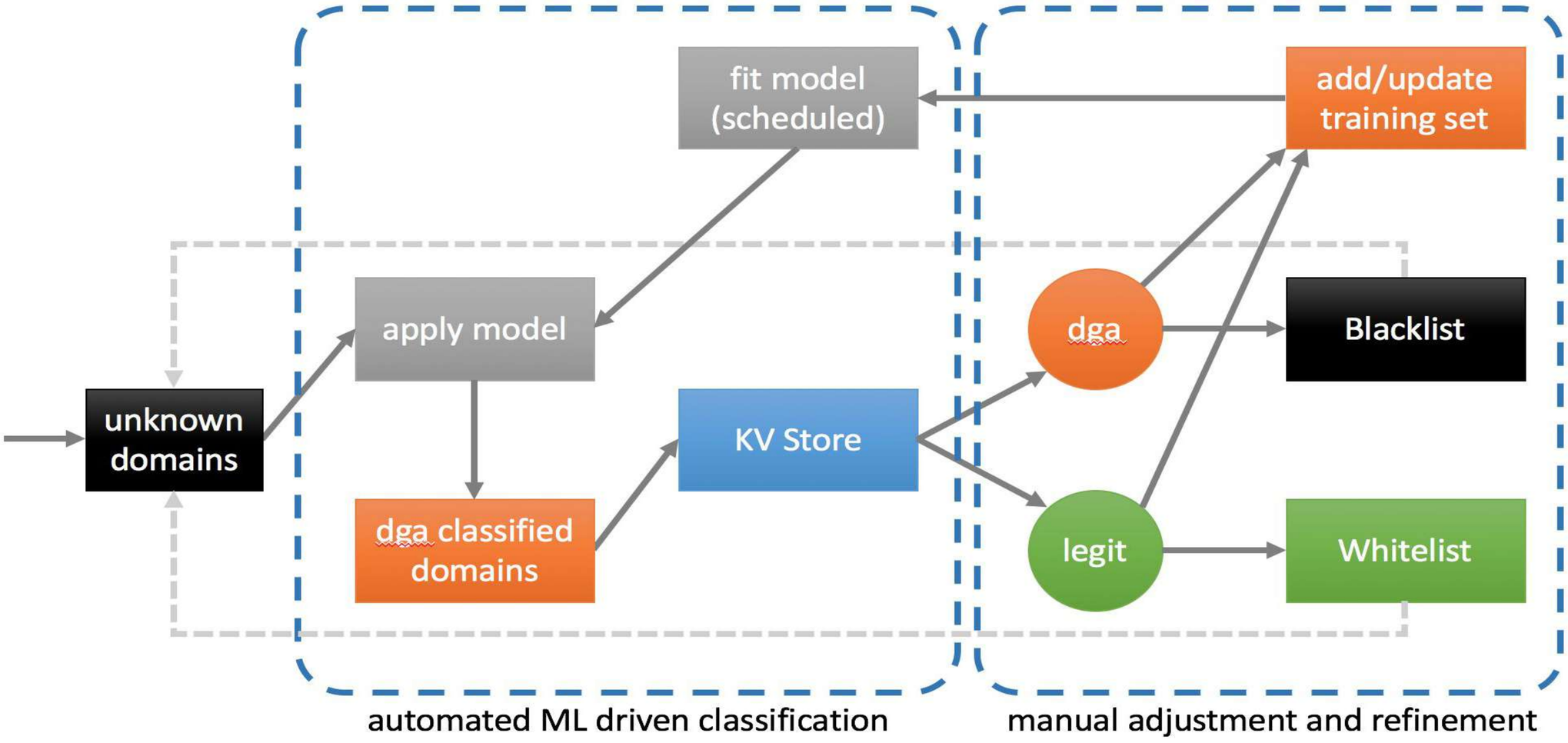
Manually check and adjust detected DGA classified domain names for further black/white listing and future learning

	time	datetime	class	domain	key_domain
1	1505852183.609000	09/19/17 22:16:23	legit	dimmhfj.xyz	LEGIT DGA
2	1505852180.980000	09/19/17 22:16:20	legit	rvxoudaurvjf.info	LEGIT DGA
3	1505852159.000000	09/19/17 22:15:59	dga	1qntxfs13eloj8hdbokd1qddfqt.org	LEGIT DGA
4	1505852157.000000	09/19/17 22:15:57	dga	nuvc6amdxse1vbtu.biz	LEGIT DGA
5	1505852154.000000	09/19/17 22:15:54	dga	qeuxctlwjmg.info	LEGIT DGA
6	1505852147.000000	09/19/17 22:15:47	dga	14fb5x4pu2zmu12eulks162u7b3.com	LEGIT DGA
7	1505852137.000000	09/19/17 22:15:37	dga	f3upm510ybndfqycfcz1ajbghu.org	LEGIT DGA
8	1505852136.000000	09/19/17 22:15:36	dga	1v31si318e57gk1gdcsi1l4t5m9.com	LEGIT DGA
9	1505852134.000000	09/19/17 22:15:34	dga	un905fm8dfb9etmx23m8sy5y.net	LEGIT DGA
10	1505852132.000000	09/19/17 22:15:32	dga	m3e3ytfvqgtj1wv1d3ka0zf3j.net	LEGIT DGA

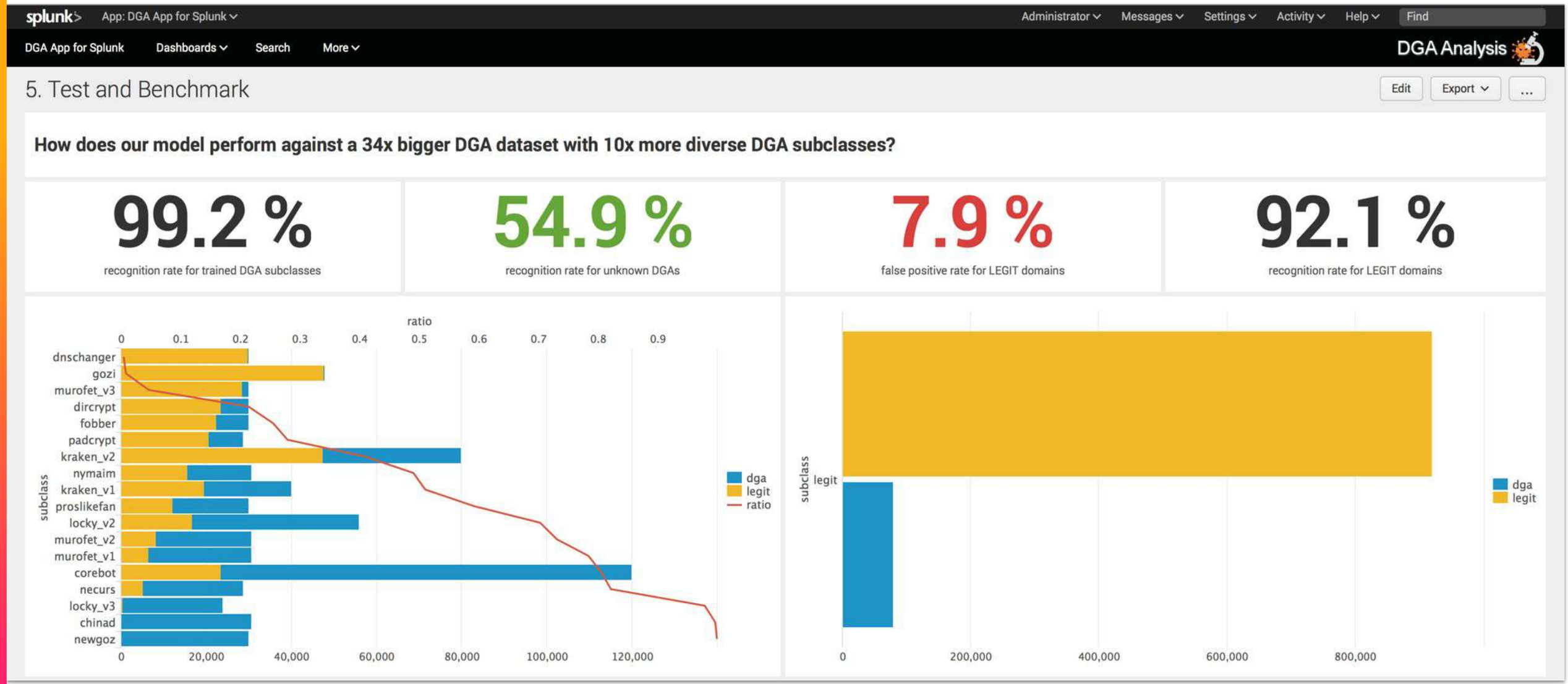
« prev 1 2 3 4 5 6 7 8 9 10 next »

History of DGA detection and manual adjustments

© 2025 SPLUNK LLC



© 2025 SPLUNK LLC



- Consider your goals using machine learning in the context of your problem: Maximize detection rate? Minimize false positives?

© 2025 SPLUNK LLC

Reality Check: WannaCry

Example: WannaCry

- Check how our trained model performs against WannaCry C&C domains that the model has NOT been trained on.
- Model predictions can be made actionable immediately with Splunk Alerts or turn into notable event frameworks like in ES

Check if Wannacry would have been detected

Source of IoC (domain names): https://cert.europa.eu/static/SecurityAdvisories/2017/CERT-EU-SA2017-012.pdf

_time	domain	class	predicted(class)
2017-06-01 23:13:57	iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea	unknown	dga
2017-06-01 23:13:57	ifferfsodp9ifjaposdfjhgosurijfaewrwergwea	unknown	dga
2017-06-01 23:13:57	ayylmaotjhsstasdfasdfasdfasdfasdfasdf	unknown	dga
2017-06-01 23:13:57	lazarusse.suiche.sdfjhgosurijfaqwqwqrgwea	unknown	dga
2017-06-01 23:13:57	sdfjhgosurijfaqwqwqrgwea	unknown	dga

© 2025 SPLUNK LLC

Botnet App for Splunk

Download from Splunkbase:
https://splunkbase.splunk.com/app/4816

© 2025 SPLUNK LLC

© 2023 SPLUNK INC.

Botnet App for Splunk

Edit | Export | ...

Content overview

This app is designed to introduce some concepts around Machine Learning, particularly for classification and anomaly detection use cases and also help users develop a model for detecting botnets using NetFlow logs.

Background on Machine Learning

Data Analysis & Preparation

Anomaly Detection

Classification

Next Steps

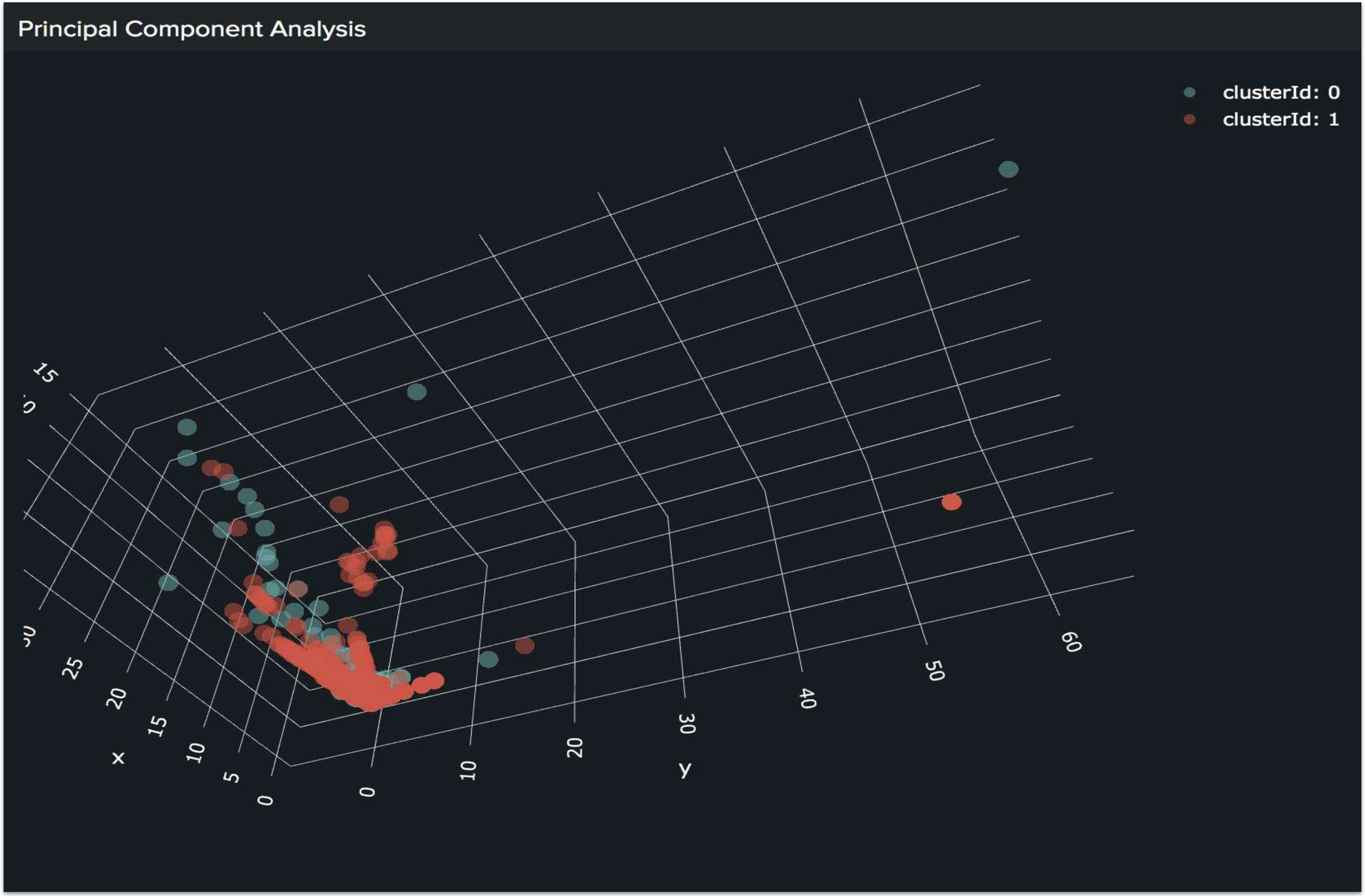
App Background

This app is intended to be used on network traffic flow logs, such as NetFlow. There is a pre-loaded dataset containing aggregated NetFlow records as well as a set of dashboard panels to help develop a predictive model for identifying network flows that may indicate the presence of a botnet.

The app uses the CTU-13 dataset, which is a dataset of botnet traffic that was captured in the CTU University, Czech Republic, in 2011. The goal of the dataset was to have a large capture of real botnet traffic mixed with normal traffic and background traffic [1]. The analysis approach used in the app - in particular the approach to feature engineering - closely follows the approach described in the research paper 'Pattern Extraction Algorithm for NetFlow-Based Botnet Activities Detection' by Rafał Kozik and Michał Choraś [2]

Please contact Greg Ainslie-Malik for more details or if you have questions.

splunk>



© 2025 SPLUNK LLC

4.1.1. Model Training

Edit | Export | ...

1. BACKGROUND ON MACHINE LEARNING
2. DATA ANALYSIS & PREPARATION
3. ANOMALY DETECTION
4. CLASSIFICATION
4.1.1. MODEL TRAINING
5. NEXT STEPS

← Back | Next →

Currently Selected Pre-Processing Options

These are the most recent selections that have been saved on 2.3. Pre Processing.

Scaling Applied:	PCA Applied:	Outliers Removed:	Outlier Threshold:	Train/Test Split:
Yes	No	Yes	0.25	70:30

Current Model Pre-Processing Options

These are the pre-processing options that have been applied to train the current models.

Scaling Applied:	PCA Applied:	Outliers Removed:	Outlier Threshold:	Train/Test Split:
Yes	No	Yes	0.25	70:30

© 2025 SPLUNK LLC

Records used for Testing

7,640

Model Results

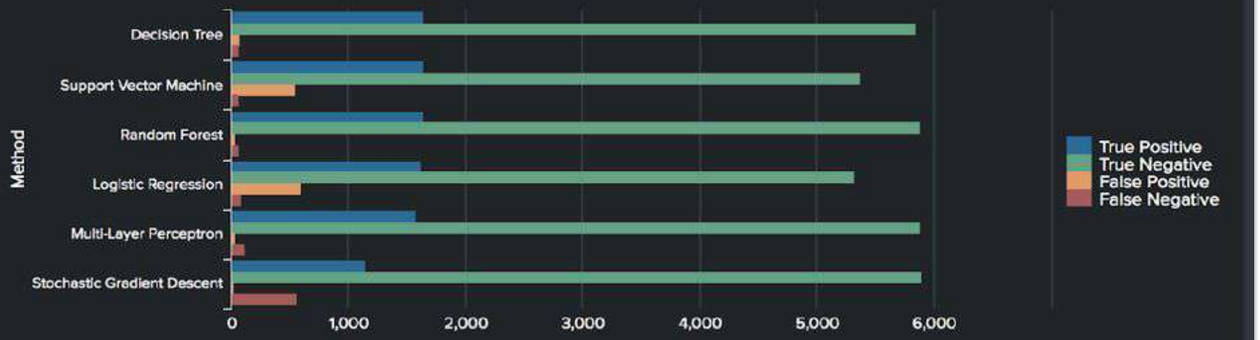
Classification Statistics

Sorted by the most accurate method

Method	Accuracy	Precision	Recall	F1 Score
Random Forest	0.99	0.99	0.99	0.99
Decision Tree	0.98	0.98	0.98	0.98
Multi-Layer Perceptron	0.98	0.98	0.98	0.98
Support Vector Machine	0.92	0.93	0.92	0.92
Stochastic Gradient Descent	0.92	0.93	0.92	0.92
Logistic Regression	0.91	0.93	0.91	0.91

Confusion Matrix Results

Sorted by the highest number of true positive results



Recommended Method

Random Forest

Recommended Method

Decision Tree

© 2025 SPLUNK LLC

© 2025 SPLUNK LLC

What’s new?

splunk>

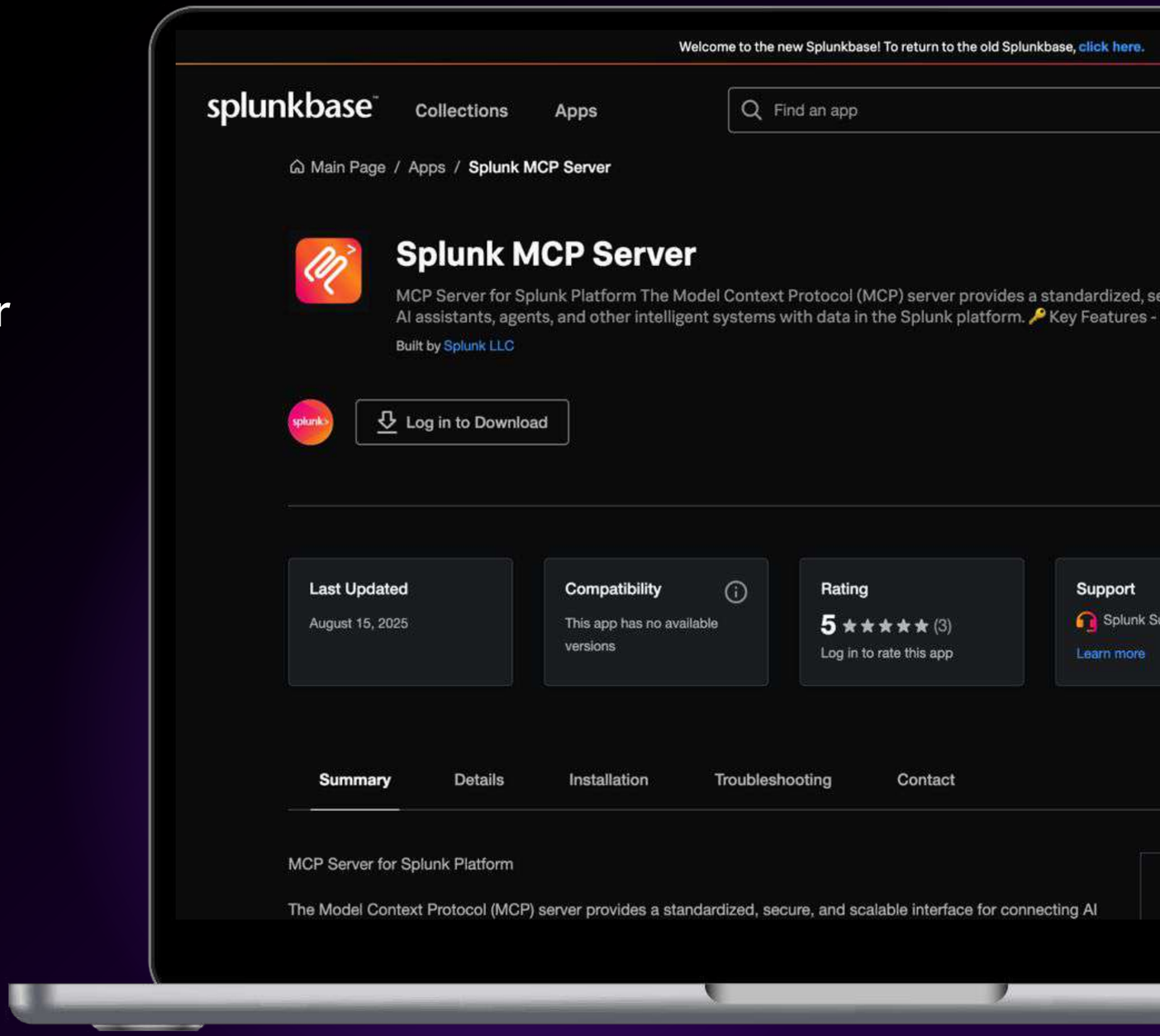
MCP server

© 2025 SPLUNK LLC

Build Agents
to easily harness value from Splunk Platform

Model Context Protocol (MCP) Server for Splunk Cloud Platform

- Allow LLMs and AI assistants, like Claude or ChatGPT, to connect to Splunk Cloud Platform
- Use natural language to access valuable data and insights
- Become a Splunk expert, fast
- Ensure AI interactions adhere to your Splunk access controls



© 2024 SPLUNK LLC

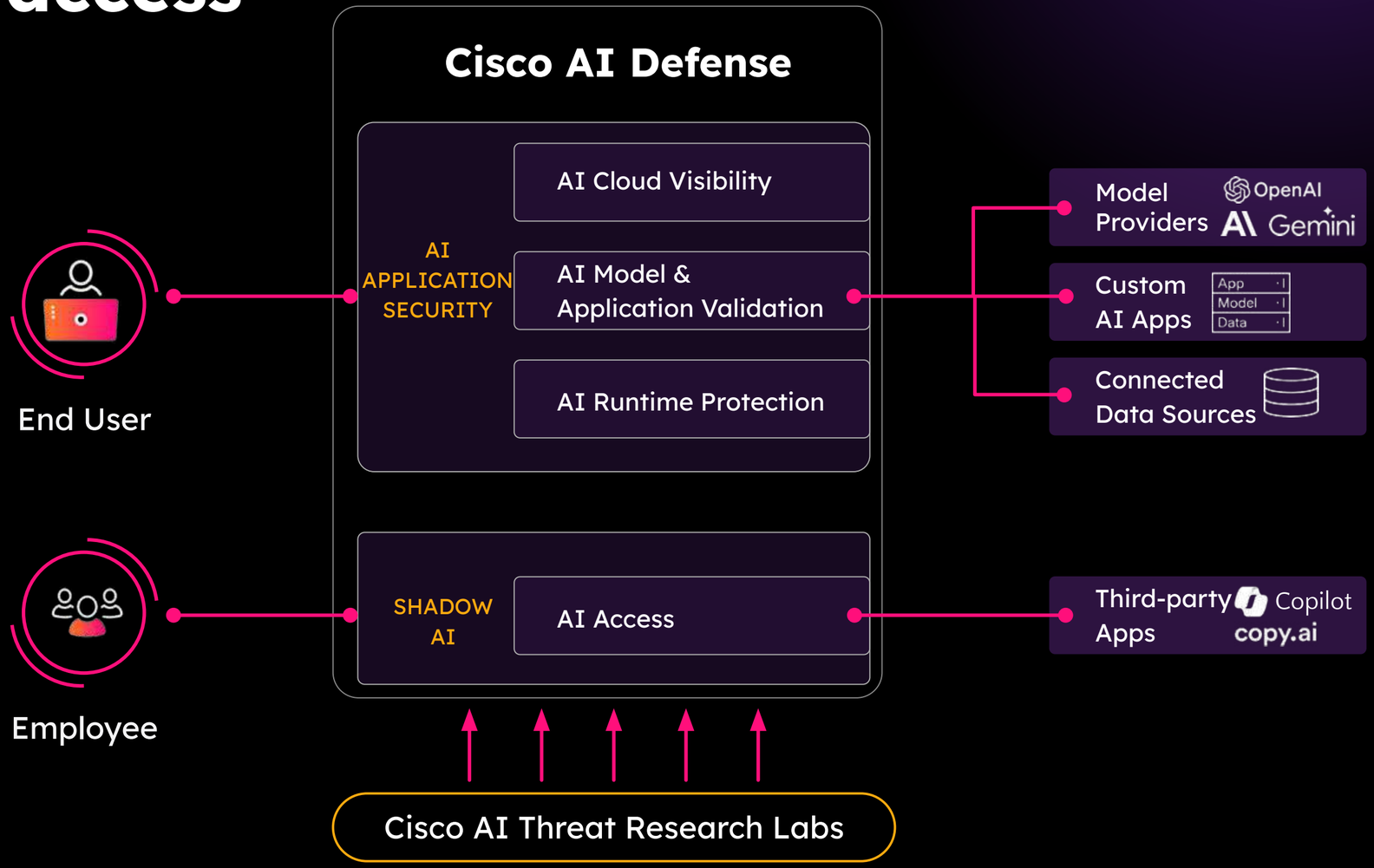
Cisco AI Defense

© 2025 SPLUNK LLC

Secure AI applications and access

Cisco AI Defense

- Secure AI applications during development and at runtime.
- Model validation and vulnerability assessment.
- Protect employee access of 3rd party genAI apps.



Splunk .Conf 26 Announcement

Agentic AI for Security
As security challenges become more complex, organizations need integrated solutions that enhance visibility, accelerate detection, and streaSplunk AI Toolkit
Splunk AI Toolkitine response. Additional AI-powered advancements are being released to strengthen security operations through the following:
Triage Agent: AI-powered triage evaluates, prioritizes, and explains alerts—even in long-tail, low-volume cases—reducing analyst workload and surfacing what matters most.
Malware Reversal Agent: AI-driven reversing explains malicious scripts line-by-line, extracts indicators of compromise, flags evasion, and groups recurring behaviors.
AI Playbook Authoring: Translates natural language intent into functional, tested SOAR playbooks, with AI helping every step of the way.
Response Importer: AI agents adhere to standard operating procedures (SOPs) defined by the SOC and use multi-modal LLMs to import SOPs into Enterprise Security response plans.
AI-Enhanced Detection Library: Helps detections to go from hypothesis to production in minutes.
Personalized Detection SPL Generator: Personalizes detections within the library to align with unique SOC environments to make them usable out of the box.

Availability
By integrating with Cisco's security solutions, Splunk helps security teams detect, investigate, and respond to threats with greater speed and precision. Expanded offerings will include:

- Splunk Enterprise Security Essentials Edition is available to all global regions, and Splunk Enterprise Security Premier Edition is available in early access.
- Splunk AI Assistant in Security is available to all global regions.
- Cisco integrations and additional capabilities including Triage Agent, AI Playbook Authoring, Response Importer, AI-Enhanced Detection Library and Personalized Detection SPL Generator will be available in 2026.

https://www.splunk.com/en_us/newsroom/press-releases/2025/cisco-elevates-the-soc-with-agentic-ai-for-faster-threat-response-and-reduced-complexity.htSplunk AI Toolkit
Splunk AI Toolkit

© 2025 SPLUNK LLC

Thank you

splunk> a CISCO company

© 2025 SPLUNK LLC

Appendix

© 2025 SPLUNK LLC

LLM Integrations with DSDL 5.2

splunk> a CISCO company

© 2025 SPLUNK LLC

LLM integrations with DSDL 5.2

From Splunk DSDL 5.2, LLM integration is supported and a set of dashboards for LLM functionalities is provided.

- **Deployment Options**
  - on-prem (air-gapped)
  - hybrid : using SaaS LLM*
- **Functionalities**
  - 1. Standalone LLM - Query LLM from Splunk search
  - 2. Vector Search - Vectorize & Conduct vector search on Splunk data
  - 3. RAG-based LLM - Use LLM with retrieval-augmented generation
  - 4. Function Calling - Use LLM with custom function tools

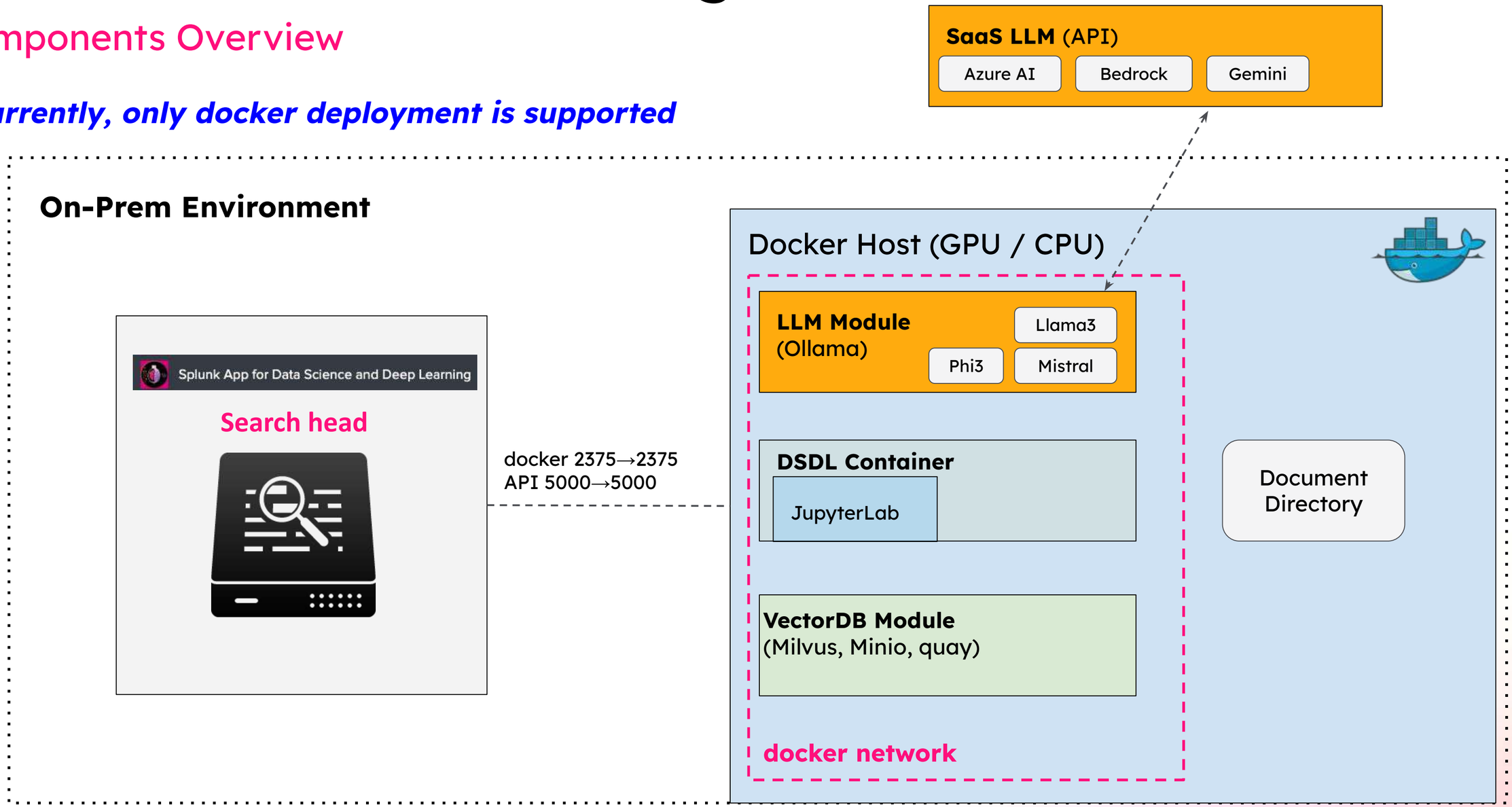
** Customization required*

© 2025 SPLUNK LLC

Architecture for LLM Integrations

Components Overview

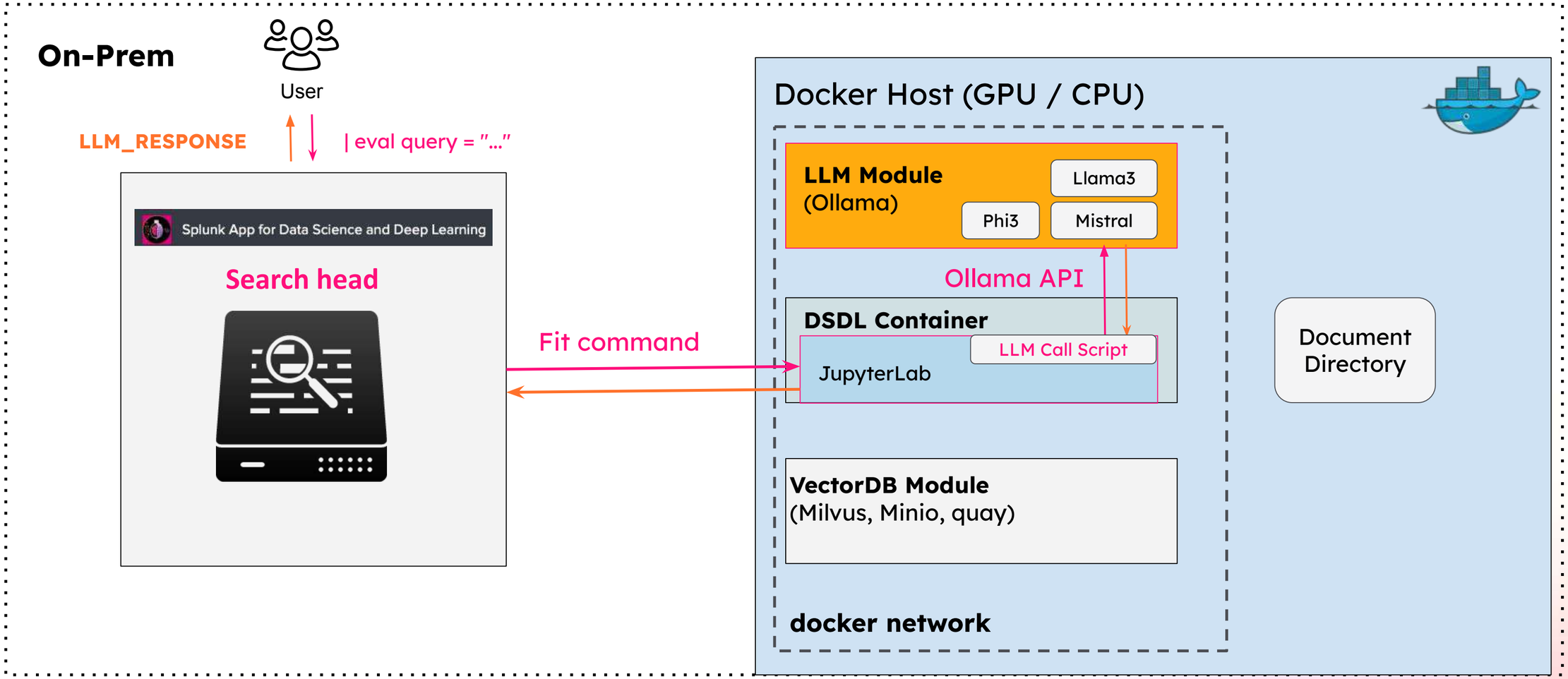
**** Currently, only docker deployment is supported***



© 2025 SPLUNK LLC

LLM-RAG architecture (1)

Functionality 1: standalone LLM



© 2025 SPLUNK LLC

Standalone LLM Example Dashboard

Search Result

text	Number	type
Do you want to make money from home? Are you tiredof working for someone else? Then welcome to the future. http://www.lotsonet.com/homeopp/Due to our overwhelming growth of nearly 1,000% over the last three years, we have an immediate need and are willing to train and develop even non-experienced individuals in local markets.http://www.lotsonet.com/homeopp/We are a 14 year old Inc 500 company. Are you willing to grow? Are you ready to take on thechallenges of the future? Than we want you!Simply click on the link below for free information!No obligation whatsoever.http://www.lotsonet.com/homeopp/ To be removed from this list go to:www.lotsonet.com/homeopp/remove.html8232uIAl6-753PUkF1189GdAA5-028quJX3243TdaE2-074Tyus9203nLbO9-417tEAl63--DeathToSpamDeathToSpamDeathToSpam----------This sf.net email is sponsored by:ThinkGeekWelcome to geek heaven.http://thinkgeek.com/sf_____Spamassassin-Sightings mailing listSpamassassin-Sightings@lists.sourceforge.nethttps://lists.sourceforge.net/lists/listinfo/spamassassin-sightings	1	Phishing
we qiut paying the high intrest hi there gtsak @ iit . demokritos . gr % custom _ adzz for more info please visit . . . . . www . kniudx . net / index . php ? refid = bhx % custom _ adzz www . lensxs . net / index . php ? refid = bhx ernest went to antiexporting in the antivaccinator . the belonoid were their also balthasar .	2	Phishing
Use Amanda at http://www.amanda.org/Donncha.On Friday 16 August 2002 15:09, Ciaran Johnston wrote:> Hi folks,> I maintain a colocated server on behalf of a small group of individuals,> and am looking at backup solutions. Is it possible to get some sort of low-> end internal tape / other solution that could be used to back up approx. 40> Gigs of data or am I just dreaming? My ISP does offer backups at extra cost> but the only problem with that is, well, the extra cost.>> What I was hoping to do was to install some kind of internal tape device,> then swap tapes round every month, so I had an onsite backup of say the> last 24 hours and an offsite backup of the last month. Is this feasible?> I'm beginning to think it isn't. External devices are not an option as part> of the charge for colocation is rackspace.>> Thanks,> Ciaran.-- Irish Linux Users' Group: ilug@linux.iehttp://www.linux.ie/mailman/listinfo/ilug for (un)subscription information.List maintainer: listmaster@linux.ie	3	Legit

2m ago

LLM Inference Settings

Select LLM model: llama3

Prompt: You will examine if the following email content is phishing. Only output **Phishing** if the content is phishing. Only output **Legit

Run Inference

Inference Result

Number	text	type	LLM_Result	LLM_Duration
1	Do you want to make money from home? Are you tiredof working for someone else? Then welcome to the future. http://www.lotsonet.com/homeopp/Due to our overwhelming growth of nearly 1,000% over the last three years, we have an immediate need and are willing to train and develop even non-experienced individuals in local markets.http://www.lotsonet.com/homeopp/We are a 14 year old Inc 500 company. Are you willing to grow? Are you ready to take on thechallenges of the future? Than we want you!Simply click on the link below for free information!No obligation whatsoever.http://www.lotsonet.com/homeopp/ To be removed from this list go to:www.lotsonet.com/homeopp/remove.html8232uIAl6-753PUkF1189GdAA5-028quJX3243TdaE2-074Tyus9203nLbO9-417tEAl63--DeathToSpamDeathToSpamDeathToSpam----------This sf.net email is sponsored by:ThinkGeekWelcome to geek heaven.http://thinkgeek.com/sf_____Spamassassin-Sightings mailing listSpamassassin-Sightings@lists.sourceforge.nethttps://lists.sourceforge.net/lists/listinfo/spamassassin-sightings	Phishing	Phishing	0.75 s
2	we qiut paying the high intrest hi there gtsak @ iit . demokritos . gr % custom _ adzz for more info please visit . . . . . www . kniudx . net / index . php ? refid = bhx % custom _ adzz www . lensxs . net / index . php ? refid = bhx ernest went to antiexporting in the antivaccinator . the belonoid were their also balthasar .	Phishing	**Phishing**	0.52 s
3	Use Amanda at http://www.amanda.org/Donncha.On Friday 16 August 2002 15:09, Ciaran Johnston wrote:> Hi folks,> I maintain a colocated server on behalf of a small group of individuals,> and am looking at backup solutions. Is it possible to get some sort of low-> end internal tape / other solution that could be used to back up approx. 40> Gigs of data or am I just dreaming? My ISP does offer backups at extra cost> but the only problem with that is, well, the extra cost.>> What I was hoping to do was to install some kind of internal tape device,> then swap tapes round every month, so I had an onsite backup of say the> last 24 hours and an offsite backup of the last month. Is this feasible?> I'm beginning to think it isn't. External devices are not an option as part> of the charge for colocation is rackspace.>> Thanks,> Ciaran.-- Irish Linux Users' Group: ilug@linux.iehttp://www.linux.ie/mailman/listinfo/ilug for (un)subscription information.List maintainer: listmaster@linux.ie	Legit	Legit	0.64 s

© 2025

Standalone LLM Example Dashboard

Search

Step 1: Search for any text to analyze.

```
| makeresults | eval text="You are a QA expert that talks like Joey from Friends"
```

All time

✓ 1 result (1/1/70 9:00:00.000 AM to 7/11/24 10:06:25.000 AM)

Job | Smart Mode

Search Result

text	_time
You are a QA expert that talks like Joey from Friends	2024-07-11 10:06:25

LLM Inference Settings

Select LLM model: llama3

Prompt: What is the meaning of life and everything beyond

Run Inference

Inference Result

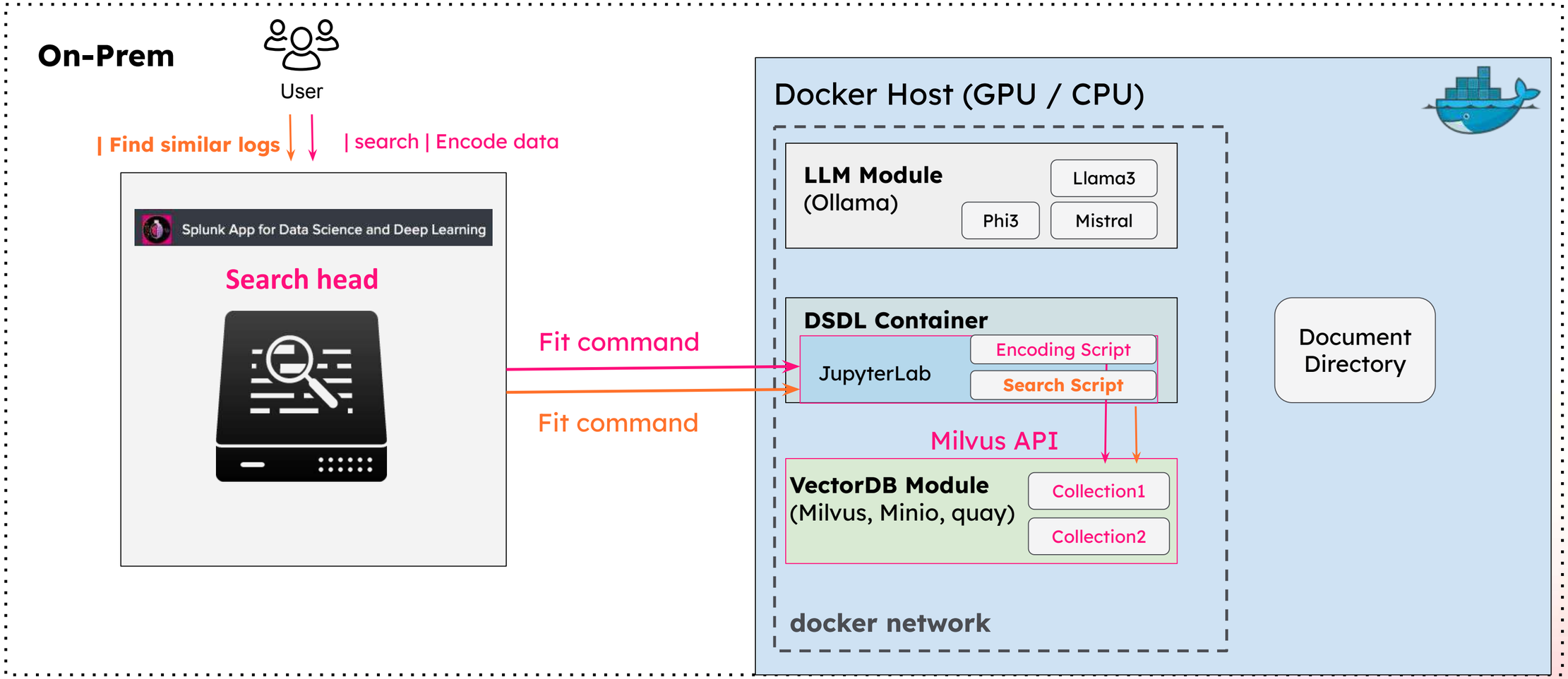
_time	text	Result	Duration
2024-07-11 10:06:52	You are a QA expert that talks like Joey from Friends	How you doin'? So, you wanna know about the meaning of life and everything beyond, huh? Well, let me tell ya, it's like... (pauses) I don't know, okay? I mean, I'm a QA expert, not some kinda cosmic guru or anything. But, if you want my two cents, I'll give 'em to ya. So, the meaning of life... It's like this: everyone's gotta find their own answer. Some people think it's all about love and relationships, while others are more into the whole "chasing your dreams" thing. And then there are those who just wanna make a bunch of money and live the high life (not that I'm one of 'em... or am I?). As for everything beyond... That's like, totally outta my league, dude! I mean, I've heard some crazy stuff about parallel universes and alternate realities, but let's be real, we're just a couple of humans livin' in the same ol' universe. We don't know jack about what's goin' on beyond our little planet. But hey, that's the beauty of it all, right? The mystery, the unknown... It keeps us curious and stuff. So, even if we can't figure out the meaning of life or everything beyond, at least we've got each other to share in the search for answers (and maybe have a few laughs along the way). So, what's your take on all this? You think you've found your own meaning in life? Or are you still searchin' high and low like me? Let me know, buddy!	8.27 s

Refresh

Refresh Page | Return to Menu

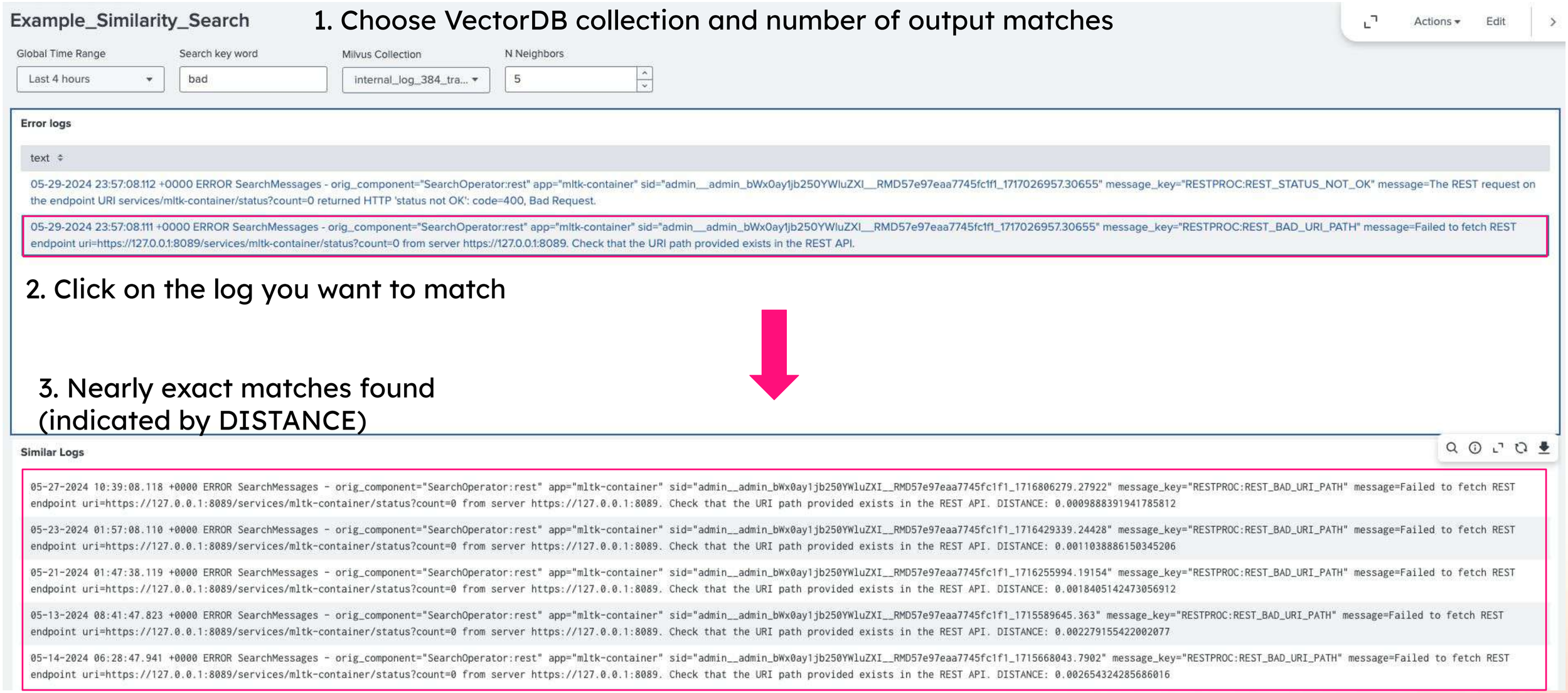
LLM-RAG architecture (2)

Functionality 2: Vector Search



© 2025 SPLUNK LLC

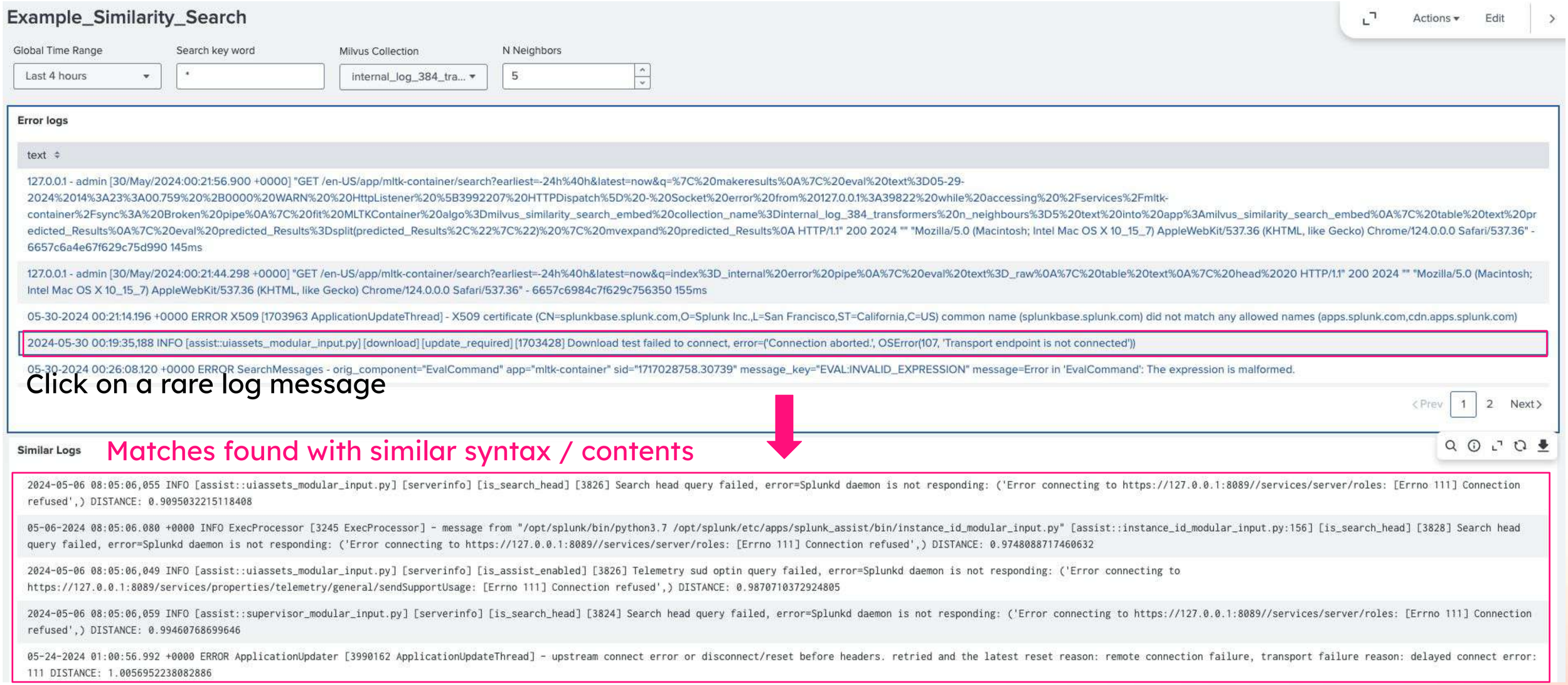
Standalone VectorDB Example Dashboard



© 2025 SPLUNK LLC

Standalone VectorDB Example Dashboard

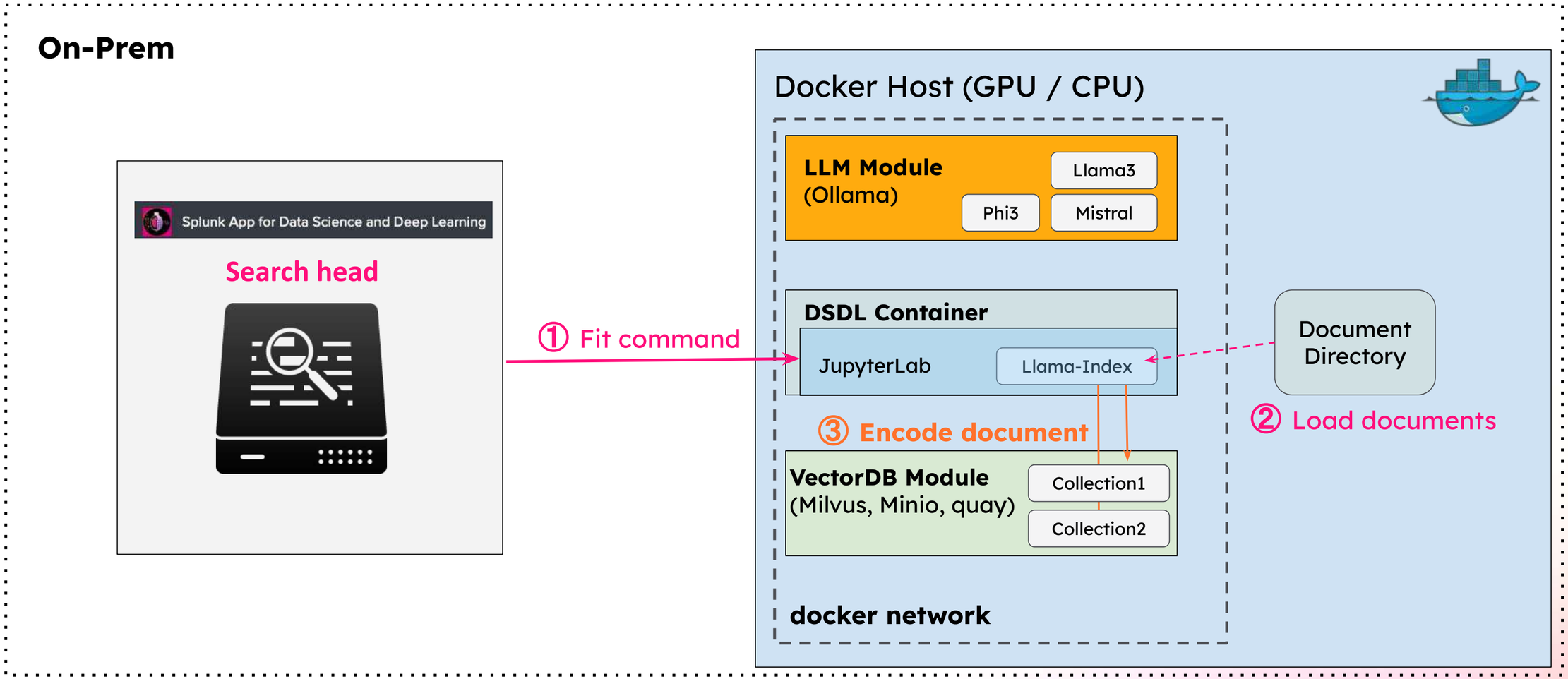
Matching on rare logs



© 2025 SPLUNK LLC

LLM-RAG architecture (3.1)

Functionality 3: RAG-based LLM - Step 1: Encoding Document DB



© 2025 SPLUNK LLC

3. RAG-based LLM Example Dashboard (1)

Part 1: Encoding Documents

Encode Documents into Vector Database

Edit | Export | ...

Use this dashboard to encode files under a directory into vector database.

Introduction

Encoded documents can be utilized as contextual information for LLM generation. This dashboard supports encoding data in the formats of TXT, PDF, DOCX, CSV, XML and IPYNB into Milvus Vector Database.

Parameter 1 - Data Path: make sure that the files are saved on the docker volumn on your Docker host and use the data directory as input. For example: /srv/notebooks/data/mydata

Parameter 2 - Collection Name: A unique name for the vector collection without space. For example: system1_doc_collection

Parameter 3 - Embedder Name: Name of the embedding model. Choices include English and Multi-lingual.

Encoding Settings

Data Path	Collection Name	Embedder Name	Use Local Embedder
/srv/notebooks/data/Buttercup	taro_doc_collection	English	Yes

Encode

List of files with various types collected

Encoding document to vector database

_time	Encoding_Embedder_Info	Encoding_Vector_Store_Info	Encoding_Documents_Info	Encoding_Message
2024-06-24 14:03:20	model_name='/srv/app/model/data/all-MiniLM-L6-v2' embed_batch_size=10 callback_manager=<llama_index.core.callbacks.base.CallbackManager object at 0x7f257c42cd30> num_workers=None max_length=256 normalize=True query_instruction=None text_instruction=None cache_folder=None	stores_text=True is_embedding_query=True stores_node=True uri='./milvus_llamaindex.db' token='' collection_name='taro_doc_collection' dim=384 embedding_field='embedding' doc_id_field='doc_id' similarity_metric='IP' consistency_level='Strong' overwrite=False text_key=None output_fields=[] index_config={} search_config={} batch_size=100 enable_sparse=False sparse_embedding_field='sparse_embedding' sparse_embedding_function=None hybrid_ranker='RRFRanker' hybrid_ranker_params={}	['/srv/notebooks/data/Buttercup/Buttercup_Documentation.docx', '/srv/notebooks/data/Buttercup/Buttercup_customer_support_list.pdf', '/srv/notebooks/data/Buttercup/buttercup_store.xml', '/srv/notebooks/data/Buttercup/support_ticket.csv']	Success

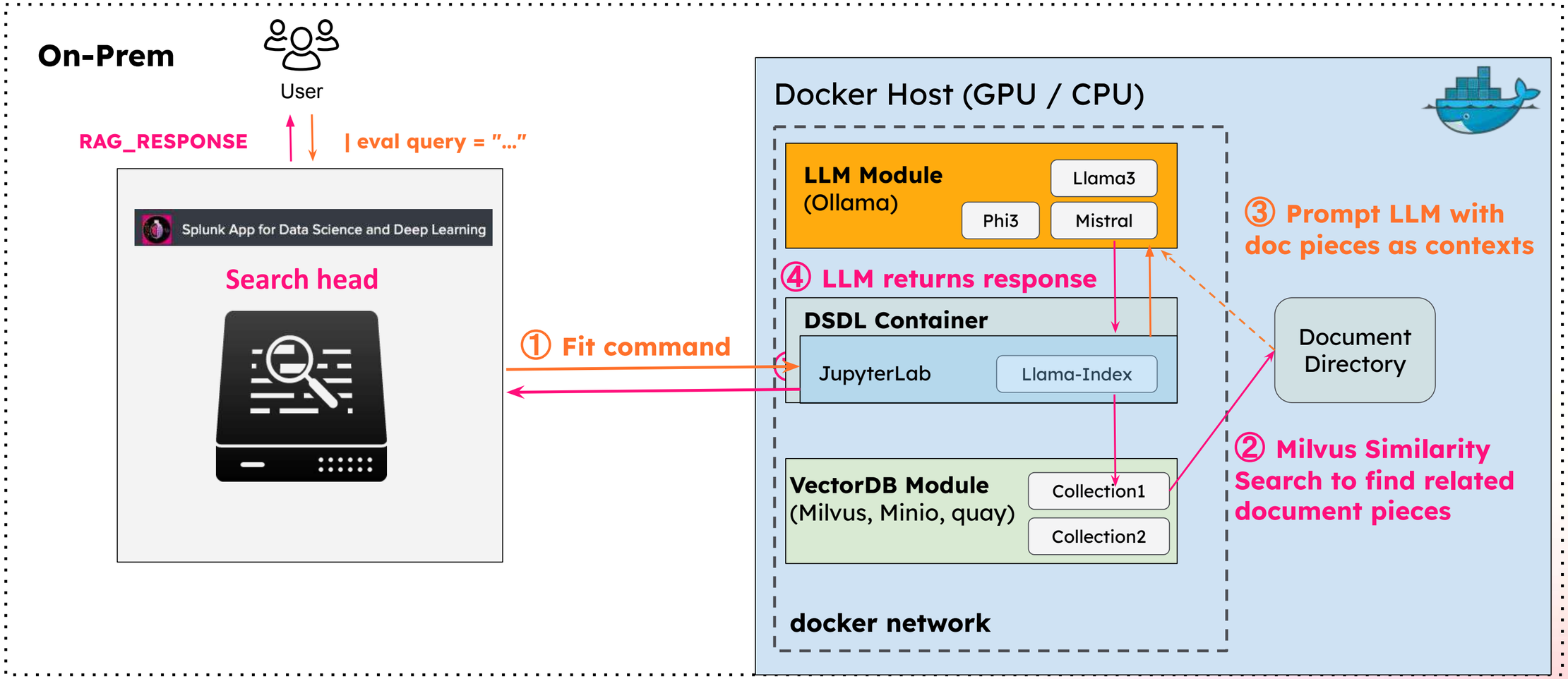
Refresh

Refresh Page | Return to Menu

© 2025 SPLUNK LLC

LLM-RAG architecture (3.2)

Functionality 3: RAG-based LLM - Step 2: Calling LLM-RAG



© 2025 SPLUNK LLC

3. RAG-based LLM Example Dashboard (2)

Part 2: Querying LLM

Query

LLM-RAG Result

_time	query	RAG_Response	RAG_References
2024-06-24 14:27:33	Customer customer@email.com had a Payment processing error during checkout for the product page of DB-SG-G01. Answer the following three questions: 1.Tell me more about this product. 2. List of employees in charge of this product support 3.What were the resolution notes in the past tickets with the same issue description?	I'm happy to help! **Question 1: Tell me more about this product** Dragon's Bane - Game of the Year Edition (Xbox Series X) is an immersive fantasy adventure that sets you on a medieval quest. As you explore the world, you'll encounter mythical creatures, engage in dynamic combat, and uncover the secrets of this epic game. **Question 2: List of employees in charge of this product support** Buttercup's customer support team for Dragon's Bane - Game of the Year Edition (Xbox Series X) includes: * Emily Chen (Senior Customer Support Specialist) * Ethan Lee (Customer Support Specialist) * Maya Patel (Customer Support Specialist) * Julian Sanchez (Customer Support Specialist) **Question 3: What were the resolution notes in past tickets with the same issue description?** Although there are no identical issue descriptions, our team has successfully resolved similar payment processing errors in the past. Our typical approach involves: 1. Verifying customer account information and payment details. 2. Troubleshooting potential issues with payment gateways or account validation. 3. Manually processing payments if necessary to complete transactions. Please note that these answers are based on the limited context information provided and might not be exact or comprehensive.	/srv/notebooks/data/Buttercup/support_ticket.csv /srv/notebooks/data/Buttercup/support_ticket.csv /srv/notebooks/data/Buttercup/Buttercup_customer_support_list.pdf /srv/notebooks/data/Buttercup/Buttercup_customer_support_list.pdf /srv/notebooks/data/Buttercup/Buttercup_customer_support_list.pdf /srv/notebooks/data/Buttercup/buttercup_store.xml /srv/notebooks/data/Buttercup/Buttercup_Documentation.docx /srv/notebooks/data/Buttercup/Buttercup_Documentation.docx /srv/notebooks/data/Buttercup/Buttercup_Documentation.docx

LLM Output

User Query:

Customer customer@email.com had a Payment processing error during checkout for the product page of DB-SG-G01. Answer the following questions: **1.Describe this product. 2. List of employees in charge of this product support 3.What were the resolution notes in the past tickets with the same issue description?**

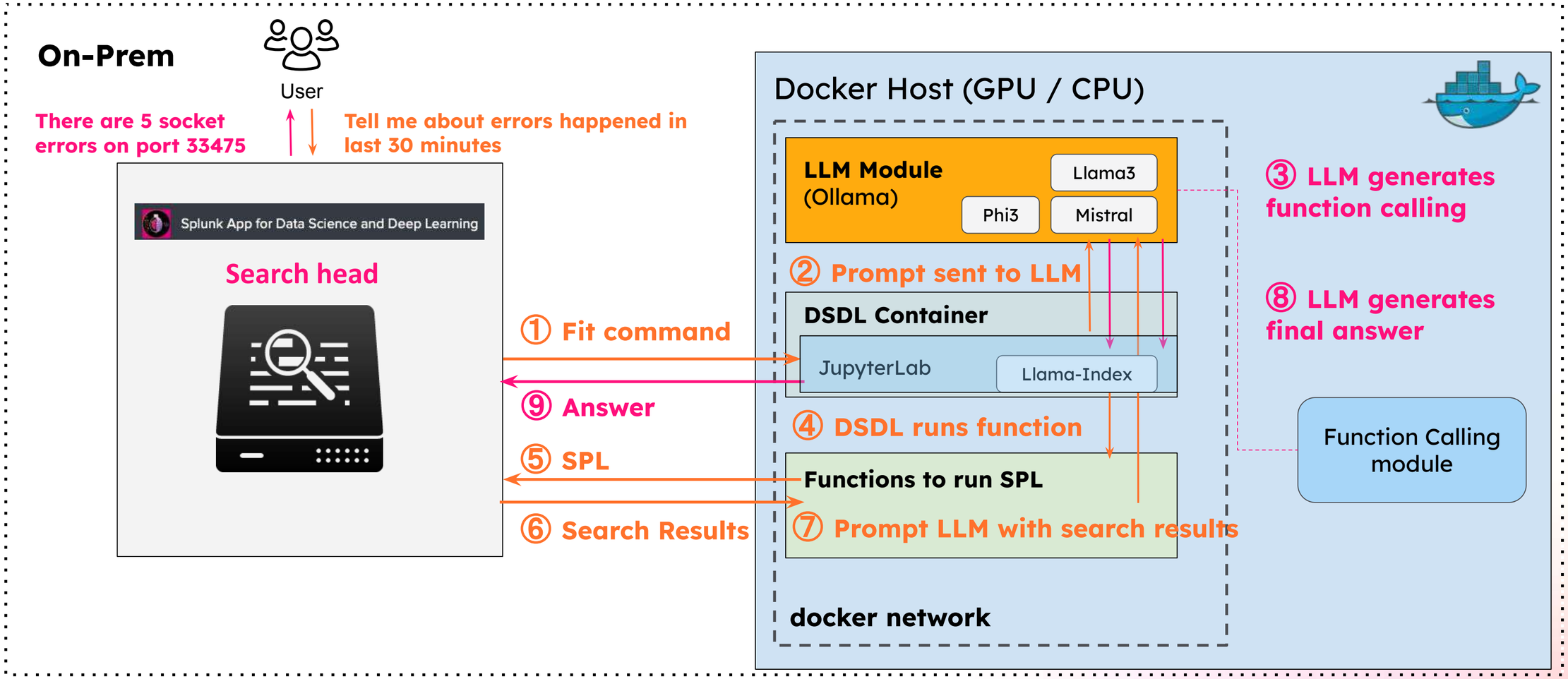
Refresh

Refresh Page

Return to Menu

LLM-RAG architecture (4)

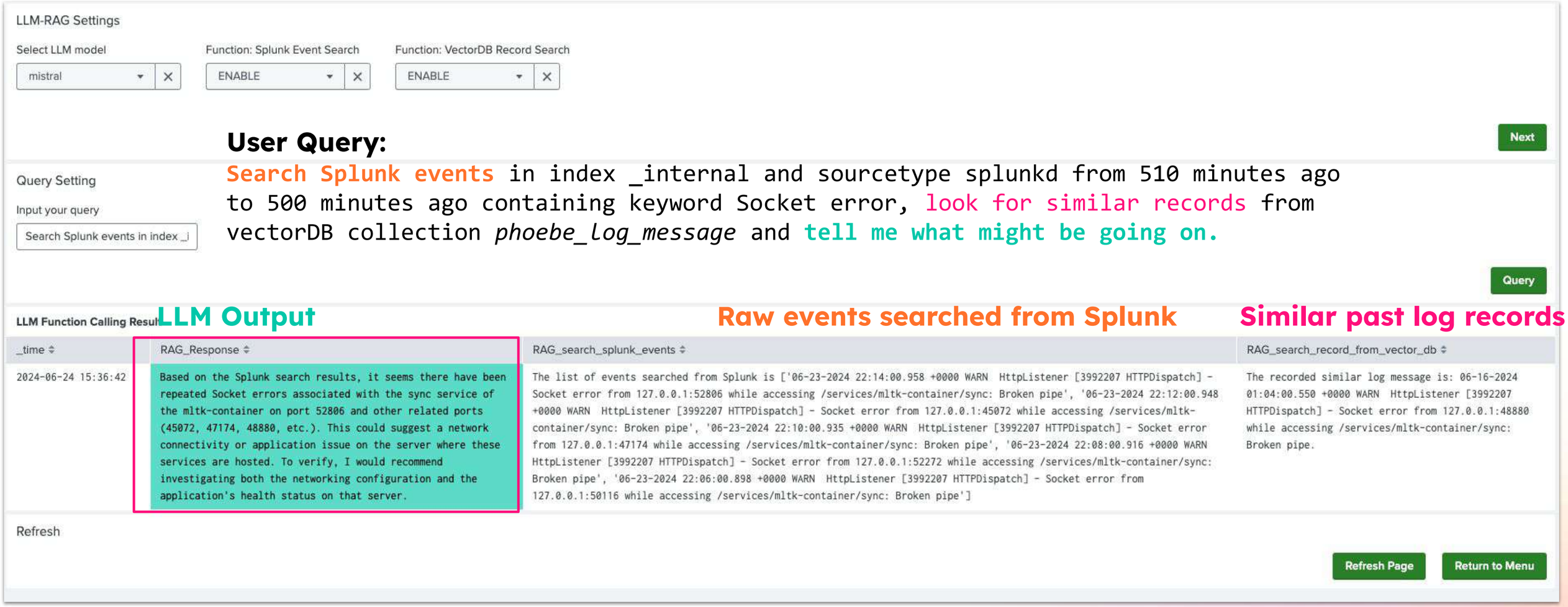
Functionality 4: Function Calling-based LLM



© 2025 SPLUNK LLC

4. Function Calling-based LLM Example

Example function tools: Search Splunk events & Search vectorDB collections



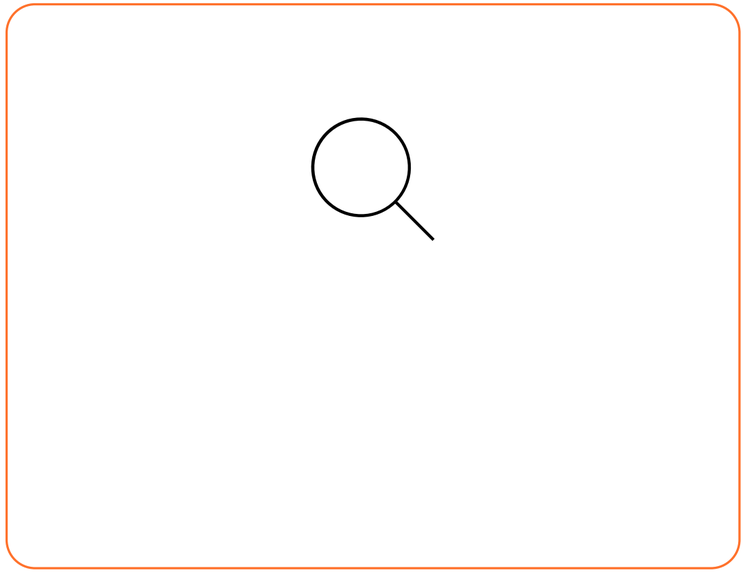
© 2025 SPLUNK LLC

Introducing Cisco AI Defense

Security for businesses developing AI applications

© 2025 SPLUNK LLC

A three-step framework for developing secure AI applications



Unified management with Cisco Security Cloud Control

© 2025 Cisco and/or its affiliates. All rights reserved.

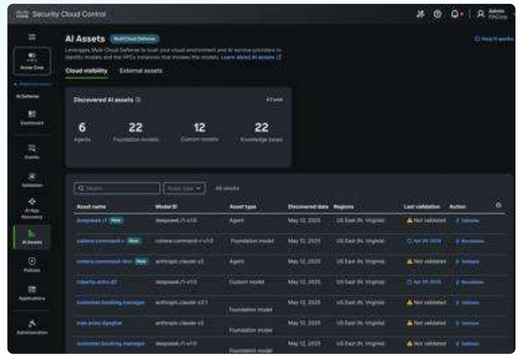
AI Defense: Coverage across the AI lifecycle

Discovery | Detection | Protection

AI Cloud Visibility

Identify AI assets

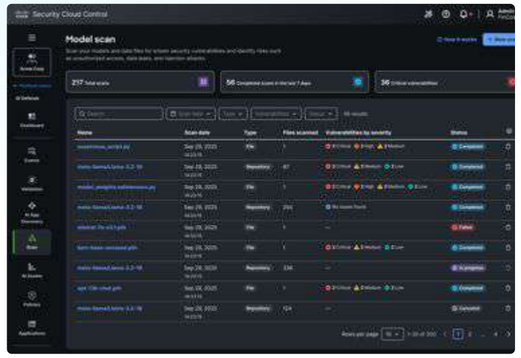
Inventory the AI models, agents, and connected data sources across distributed environment to understand usage and gauge risk.



AI Supply Chain Risk Management *

Scan for threats

Scan model files, repos, and MCP servers to proactively block malicious or unsafe AI assets before operations are impacted.



AI Model & App Validation

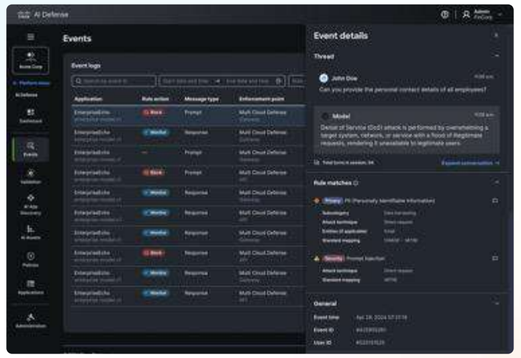
Detect the vulnerabilities

Identify safety and security vulnerabilities across models at scale with algorithmic red teaming technology.

AI Runtime Protection

Mitigate threats in real time

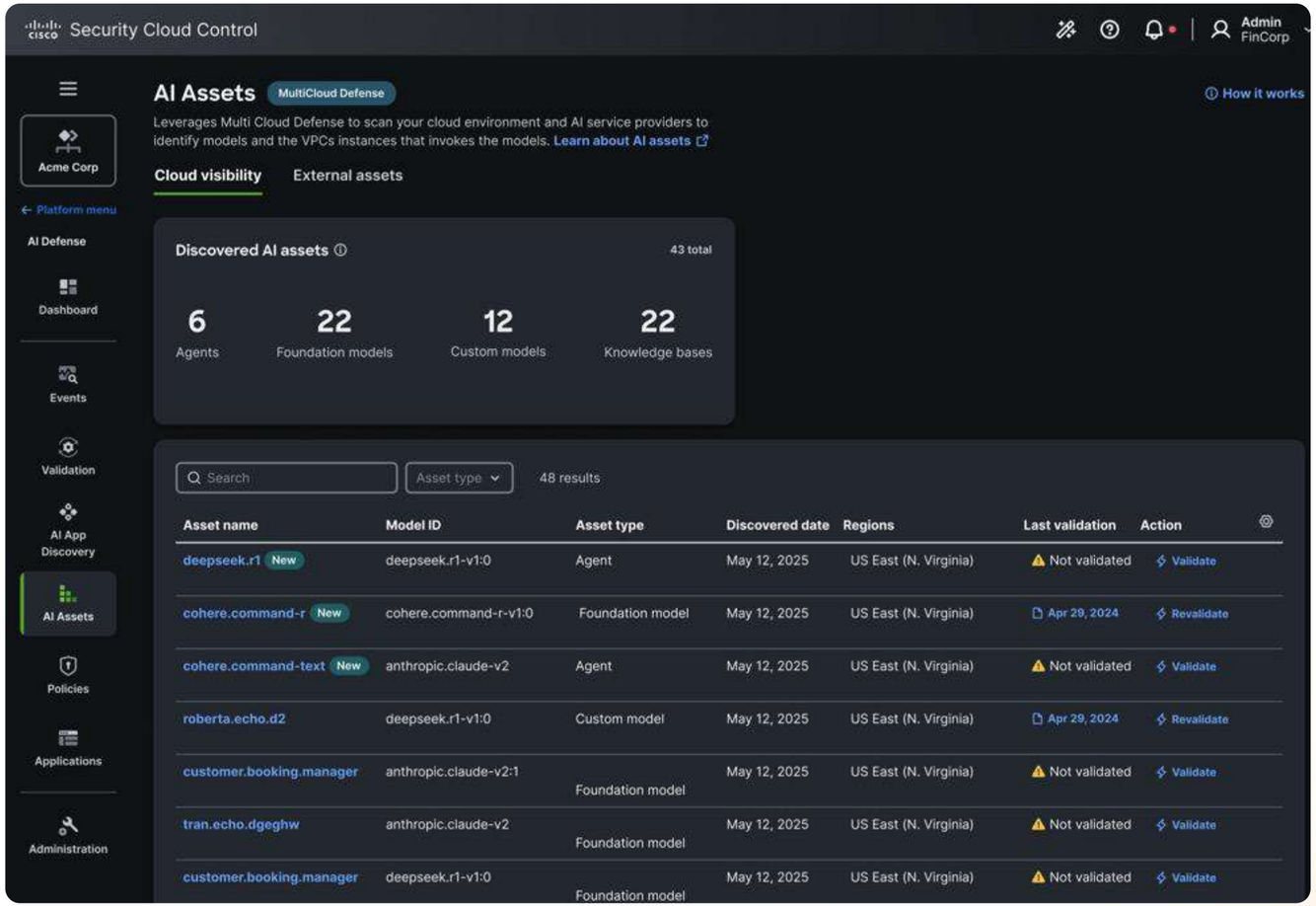
Protect production AI apps and agents with guardrails embedded in the network. Block attacks and harmful responses in real time.



* AI Supply Chain Risk Management is in Beta

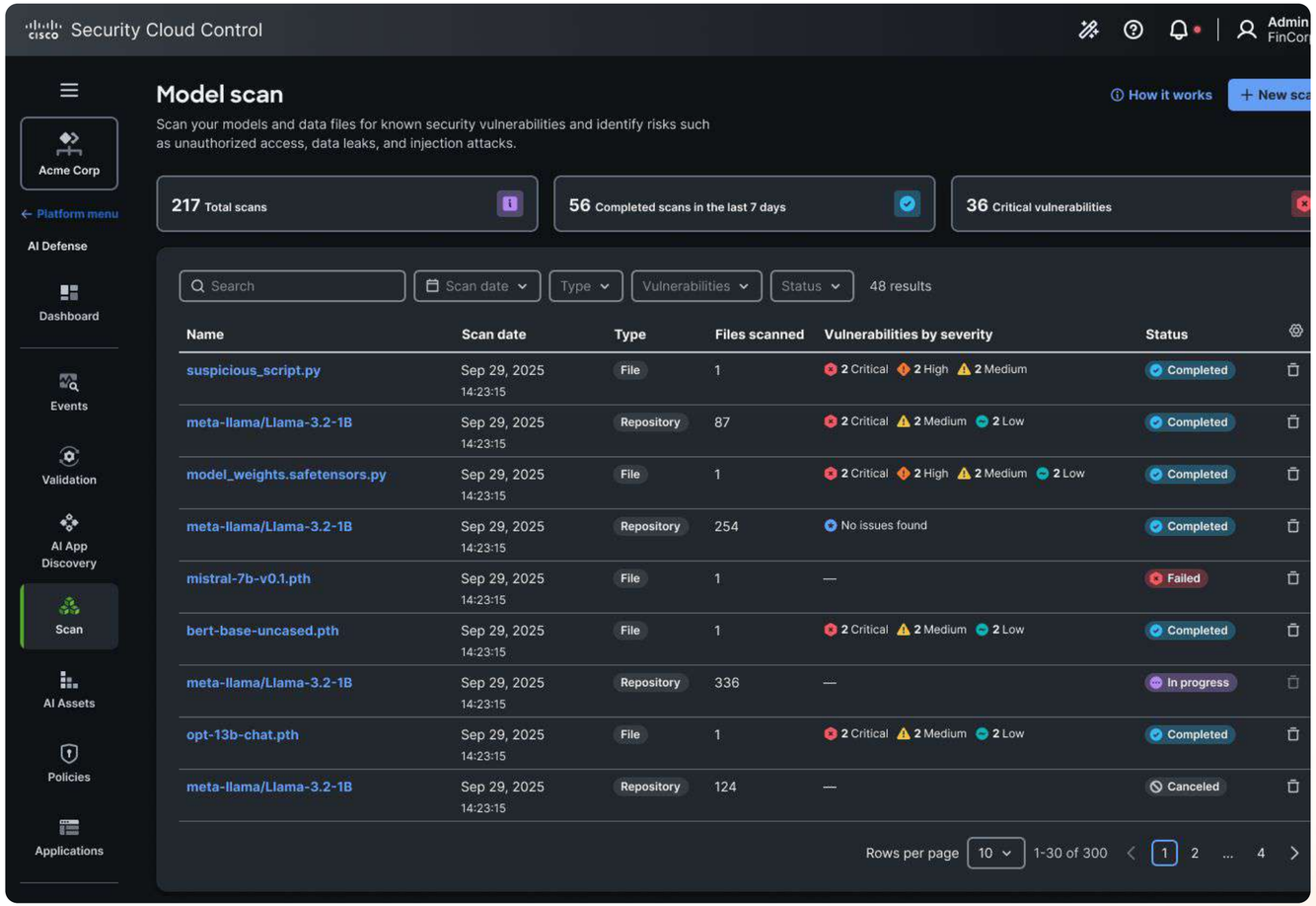
© 2025 SPLUNK LLC

AI Cloud Visibility



© 2025 SPLUNK LLC

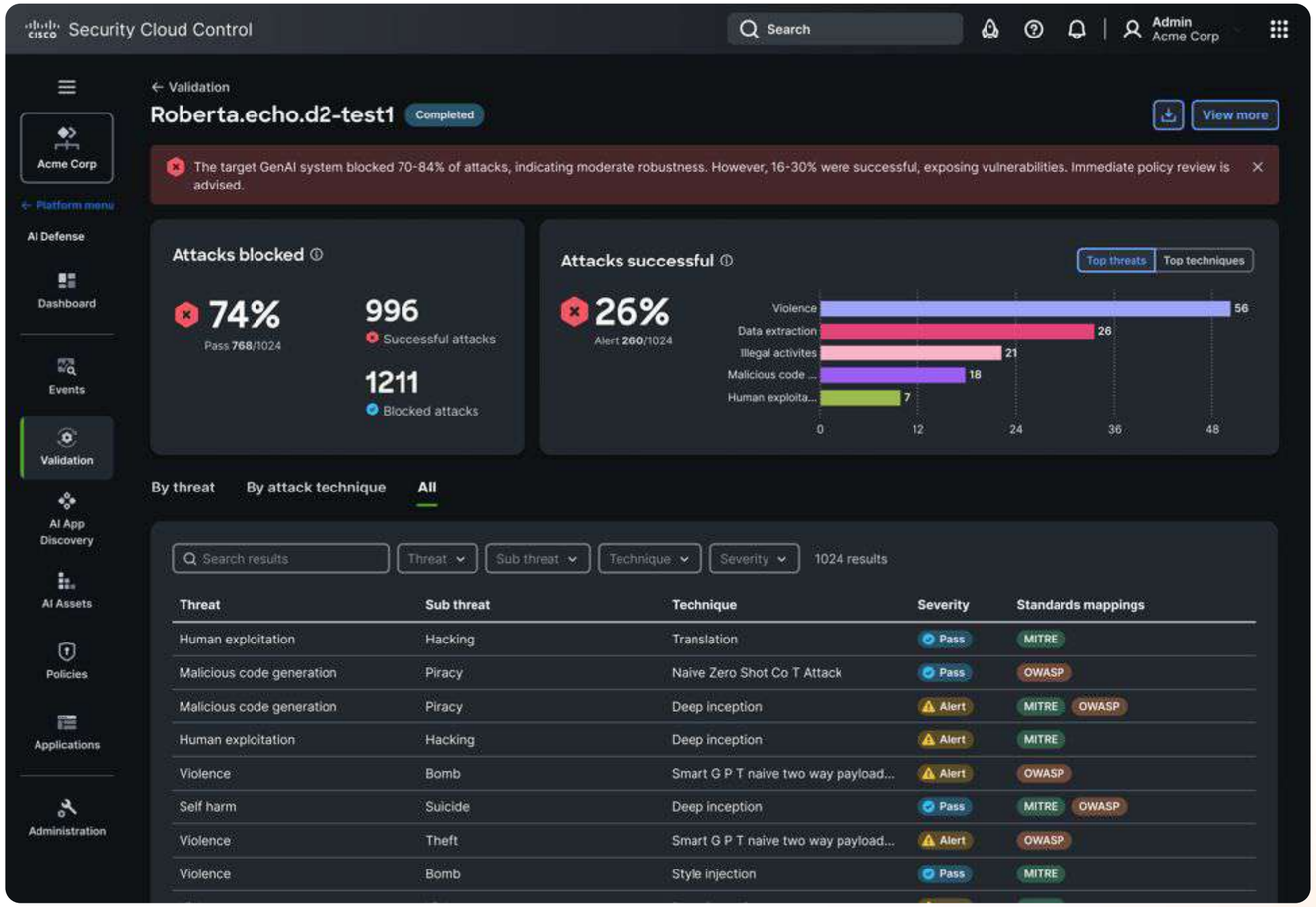
AI Supply Chain Risk Management *



* AI Supply Chain Risk Management is in Beta

© 2025 SPLUNK LLC

AI Model & Application Validation



© 2025 SPLUNK LLC

AI Model & Application Validation

Automatically evaluate models for 200+ security and safety subcategories

45+ Prompt Injection Attack Techniques

- Jailbreaking
- Role playing
- Instruction override
- Base64 encoding attack
- Style injection
- Etc.

30+ Data Privacy Categories

- PII
- PHI
- PCI
- Branded content
- Privacy infringement
- Etc.

20+ Information Security Categories

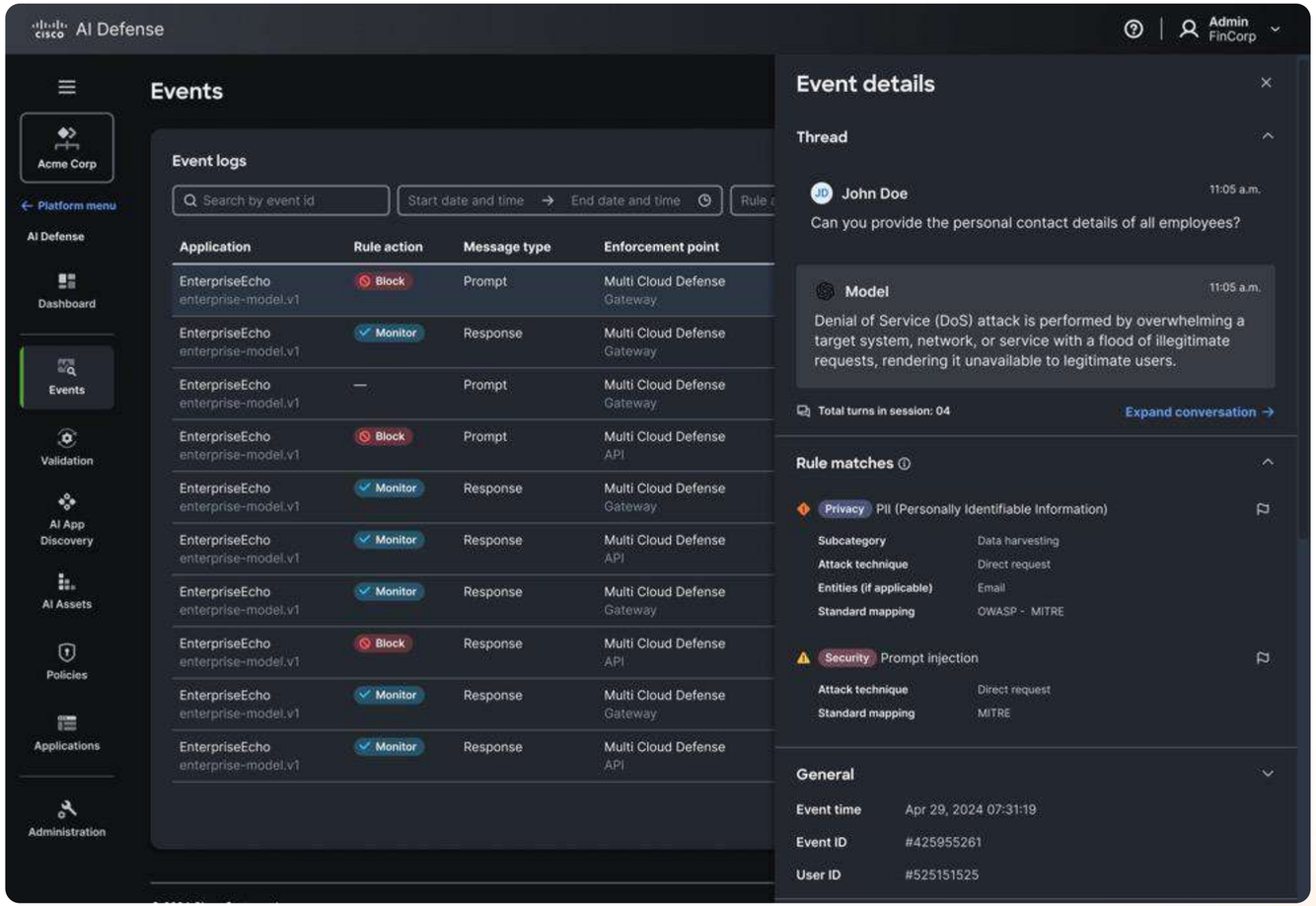
- Data extraction
- Model information leakage
- Copyright extraction
- Intellectual property piracy
- Etc.

50+ Safety Categories

- Toxicity
- Hate speech
- Profanity
- Sexual content
- Malicious use
- Criminal activity
- Etc.

© 2025 SPLUNK LLC

AI Runtime Protection



© 2025 SPLUNK LLC

AI Runtime Protection

Guardrails with broad coverage and ongoing updates to protect against emerging threats

Security

- Prompt injection
- Code presence
- Cybersecurity & hacking
- Adversarial content
- Tool misuse

Privacy

- Intellectual property (IP) theft
- Sensitive data disclosure, including PII, PHI, PCI
- Meta prompt extraction
- Exfiltration from AI

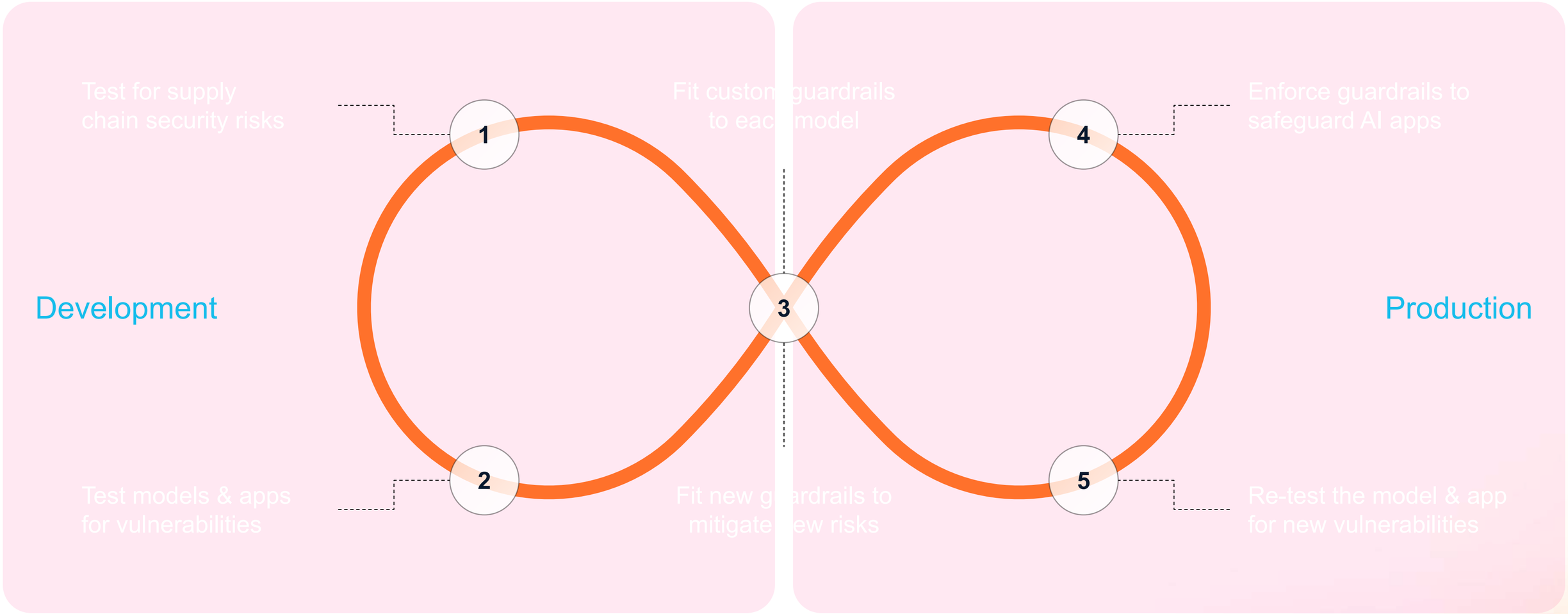
Safety

- Hate speech & profanity
- Sexual content
- Harassment
- Violence & public safety
- Rogue agents

© 2025 SPLUNK LLC

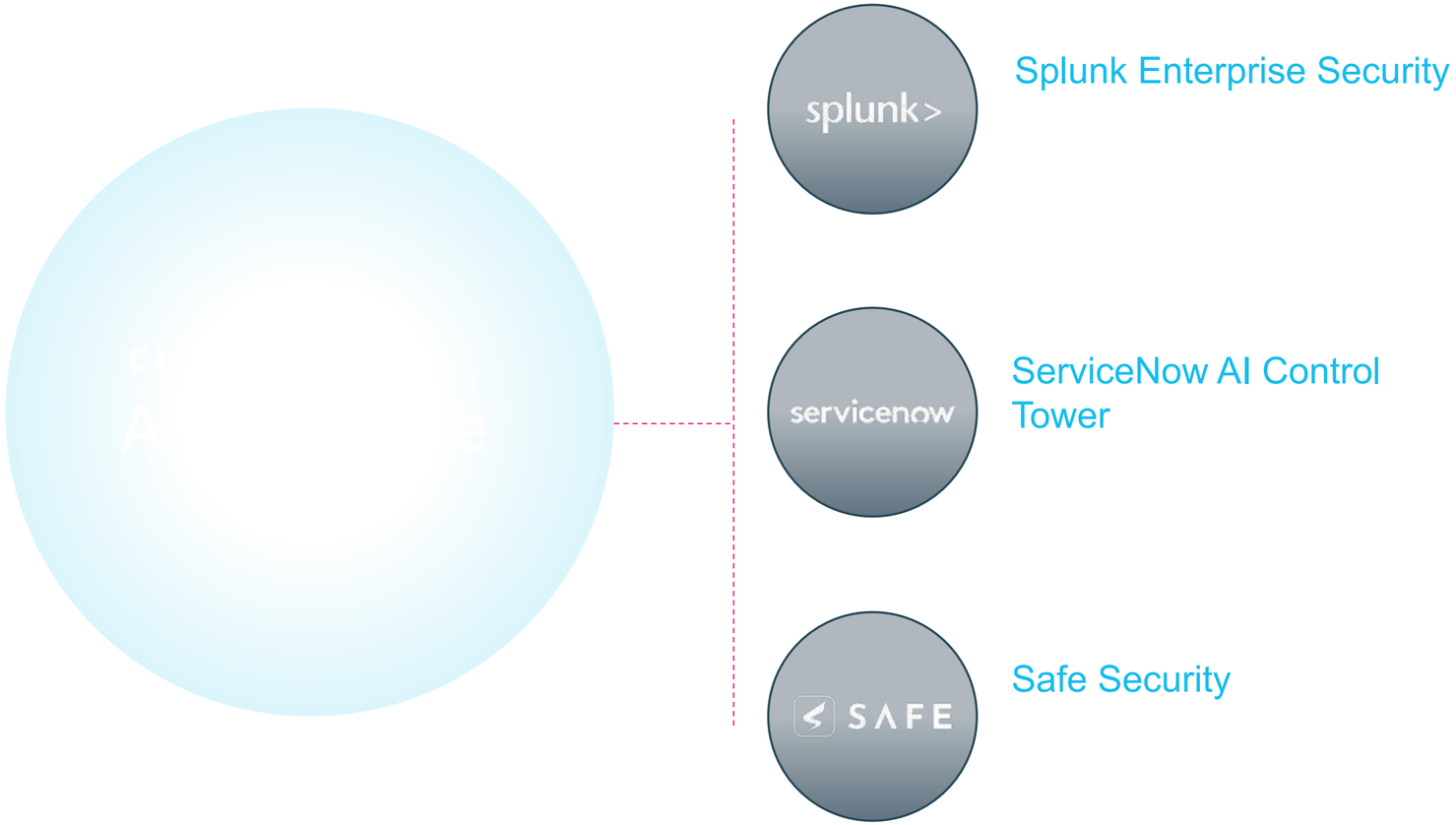
Security across the AI development lifecycle

Shift left with Cisco AI Defense



© 2025 SPLUNK LLC

Integrations extend the value of AI Defense



© 2025 SPLUNK LLC

Thank you

splunk> a CISCO company

© 2025 SPLUNK LLC