

splunk®> 4 rookies

Machine Learning for IT Ops



Forward-looking statements

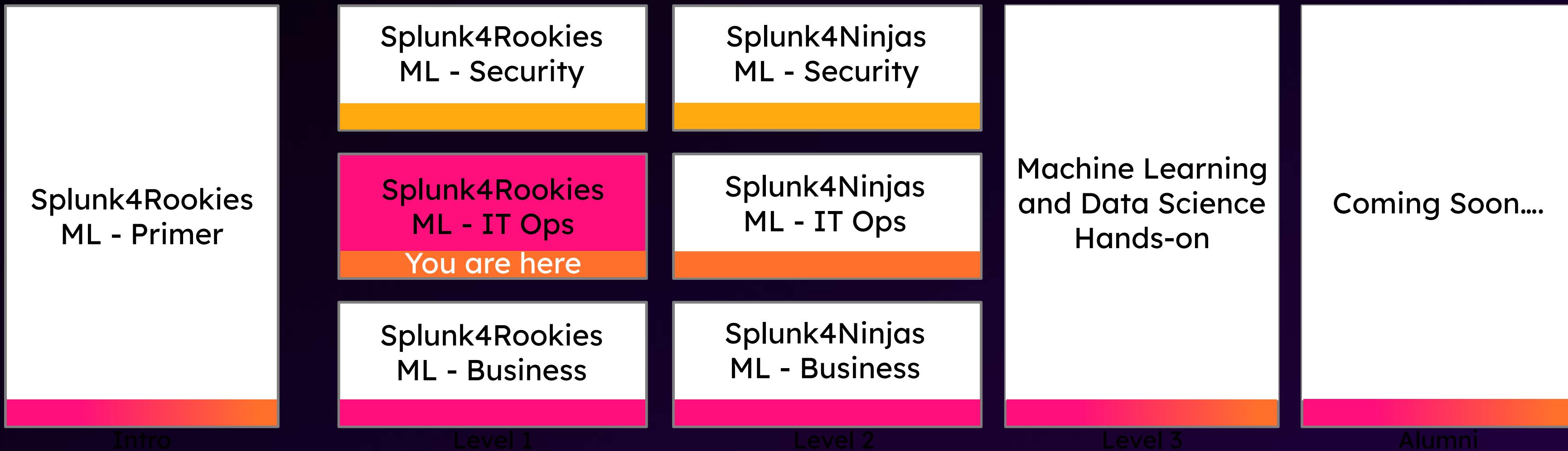
This presentation may contain forward-looking statements that are subject to the safe harbors created under the Securities Act of 1933, as amended, and the Securities Exchange Act of 1934, as amended. All statements other than statements of historical facts are statements that could be deemed forward-looking 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 Machine Learning Workshops



Boss of Machine Learning (Coming Soon)

Introductory

Advanced

Legend

Security IT Ops Business All

Agenda

- ❑ Introduction to ML
- ❑ Current Challenges in ITops
- ❑ How Splunk Drives Machine Learning
- ❑ Scenario Introduction
- ❑ Hands-on Modules
- ❑ Wrap Up and Next Steps

Leading Initiatives Driving ML Adoption

200%

Increase in proactive detection of security and performance issues, significantly reducing downtime

2.1x

More likely to have automated processes for alerts, helping operationalize data at scale

\$365k/hour

On average saved from costly outages, helping organizations protect against revenue loss

Sources:

Harvard Business Review - Artificial Intelligence for the Real World Digital Enterprise Journal Report: The Roadmap to Becoming a Top Performing Organization in Managing IT Operations

Obstacles Blocking ML Adoption

1.8x

Increase in data and events
to process every two years,
creating challenges in
handling data volume

1 in 2

Companies increase the number
of data silos, leading to difficulties
integrating ML in isolated systems

79%

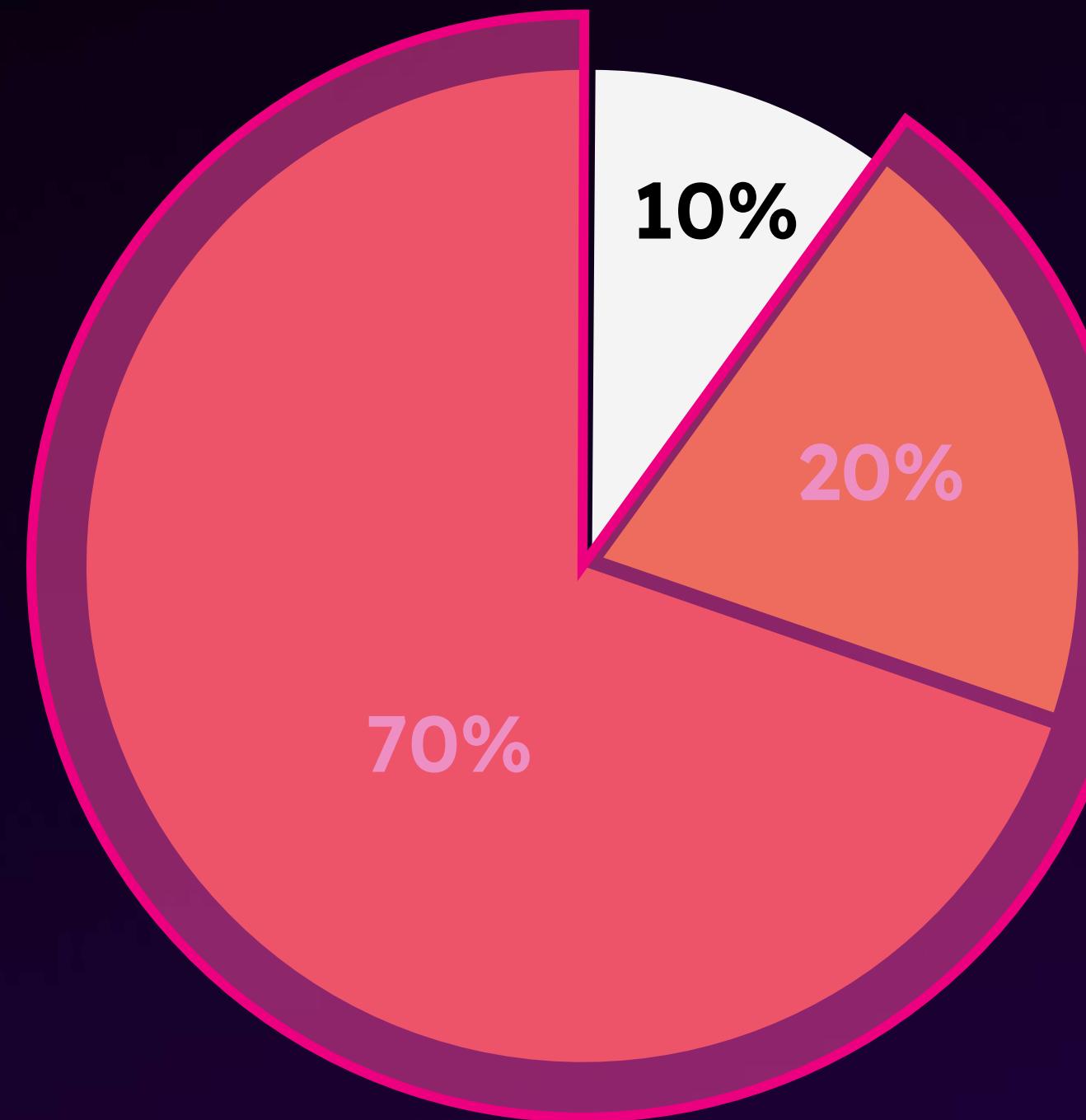
Failure rate for companies
which try to implement machine
learning from scratch, due
to lack of expertise

Sources:

Harvard Business Review - Artificial Intelligence for the Real World
Digital Enterprise Journal Report: The Roadmap to Becoming a Top Performing Organization in
Managing IT Operations

<https://www.xplm.com/news/press/industry-study-2023-companies-cannot-control-their-data-silos/>

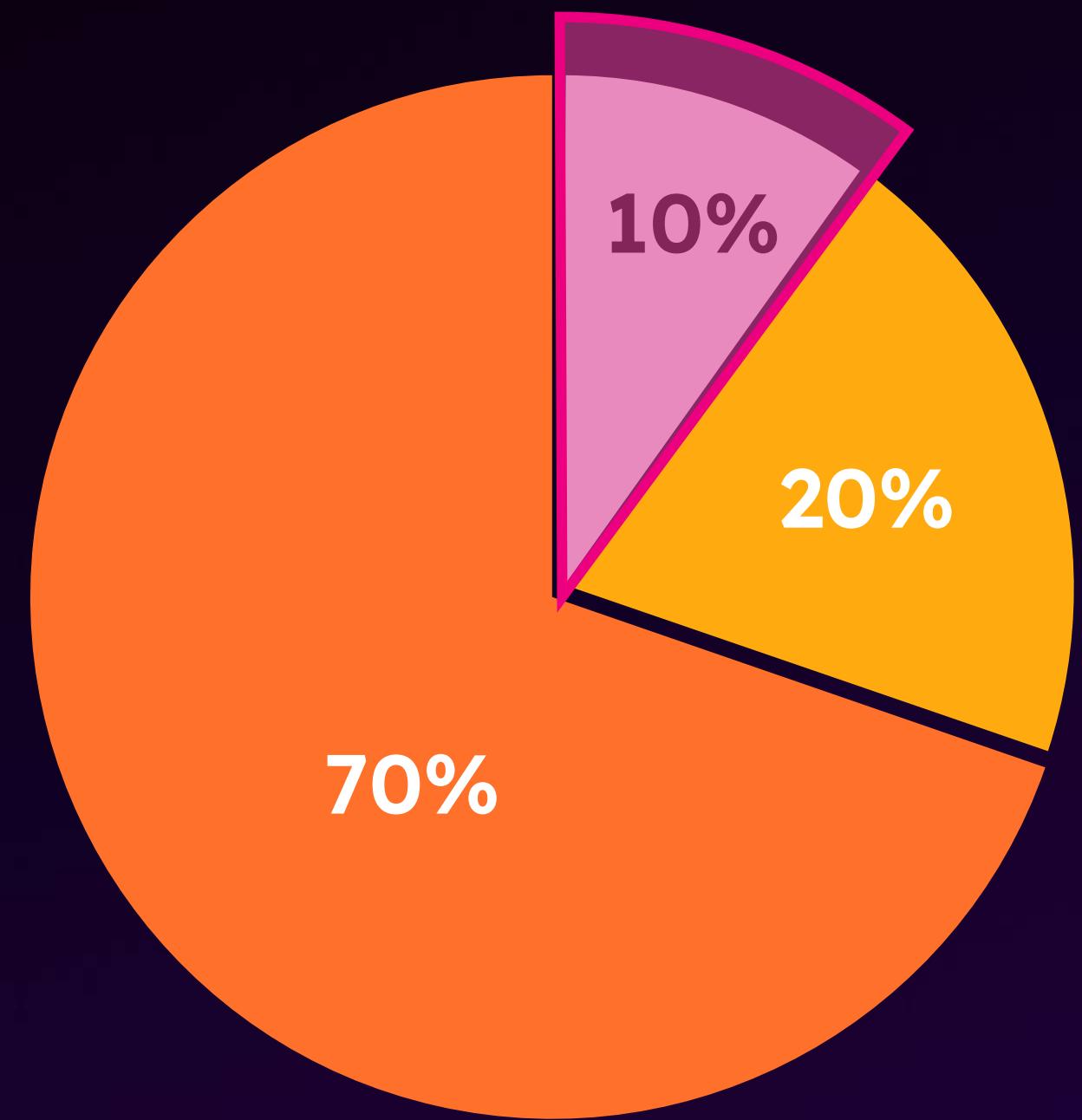
How Data Scientists Spend Their Time



- Data Engineering
- Machine Learning
- Other

Data originated from "[Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says](#)", Forbes Mar 23, 2016".

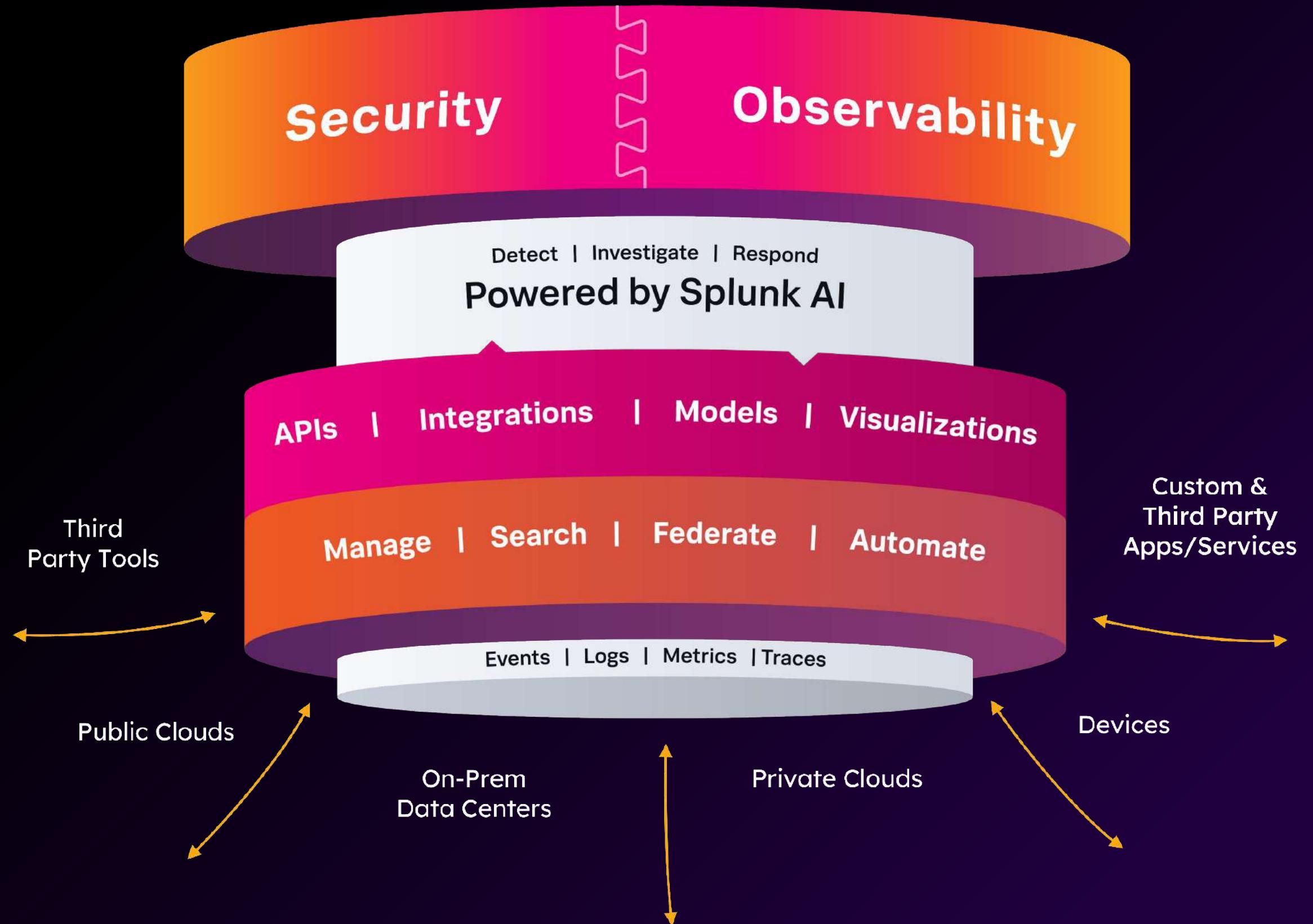
How Data Scientists Spend Their Time



- Data Engineering
- Machine Learning
- Other

Data originated from "[Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says](#)", Forbes Mar 23, 2016".

The Unified Security and Observability Platform



Cisco + Splunk power and protect the AI revolution.

Infrastructure for AI

Networking and compute solutions for training AI models at scale.

Data for AI

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

Observability for AI

Observability supports trustworthy, performant, and reliable AI deployment.

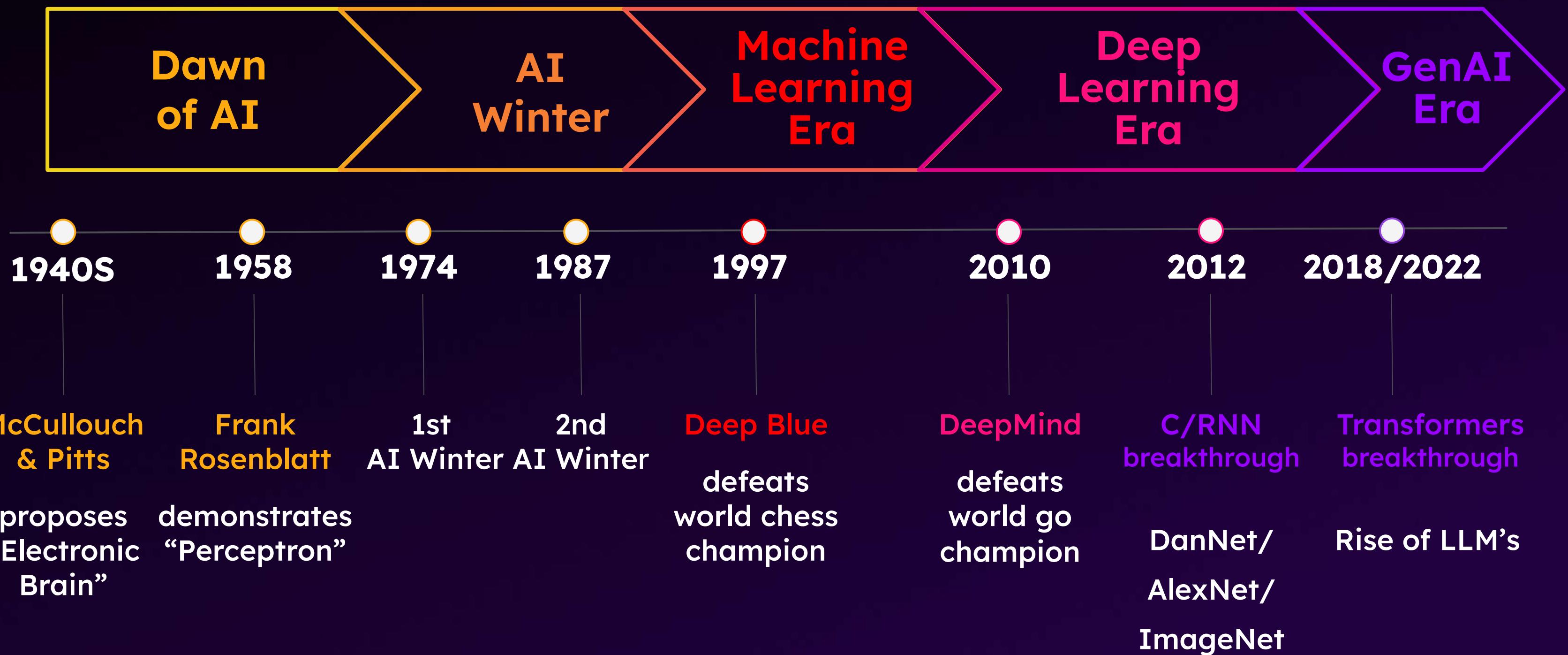
AI for Security

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

Security for AI

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

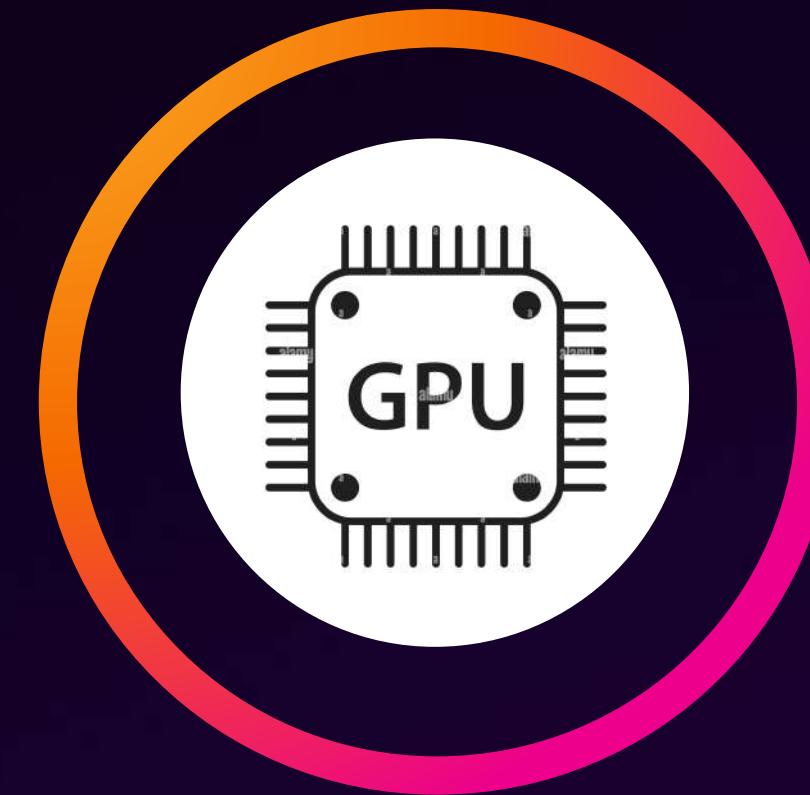
History of AI



What is driving the explosive growth of AI?



Algorithm

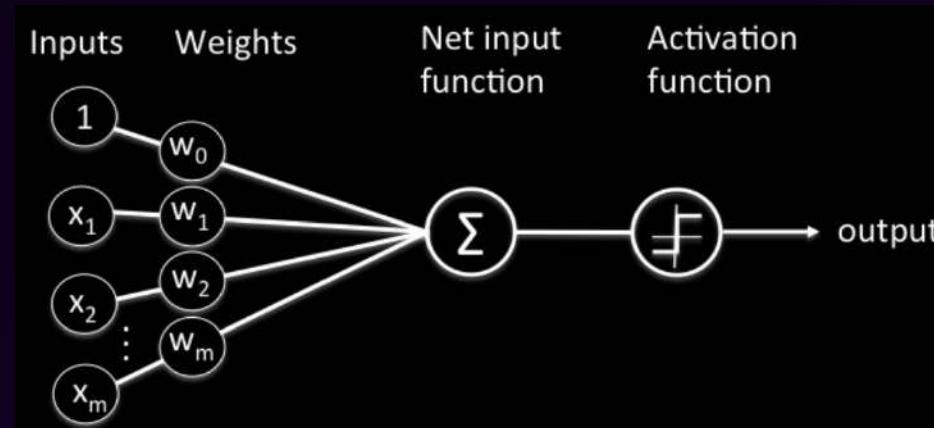


Compute

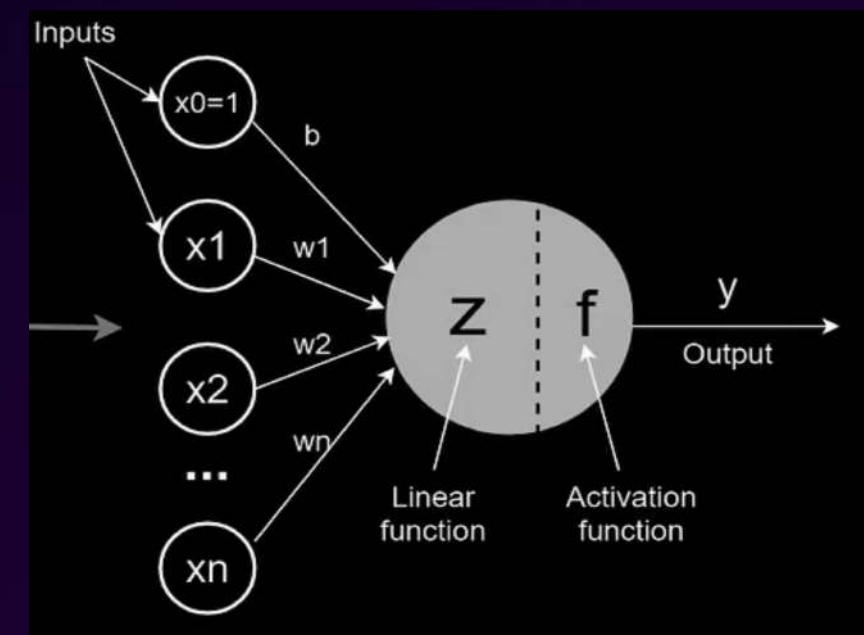
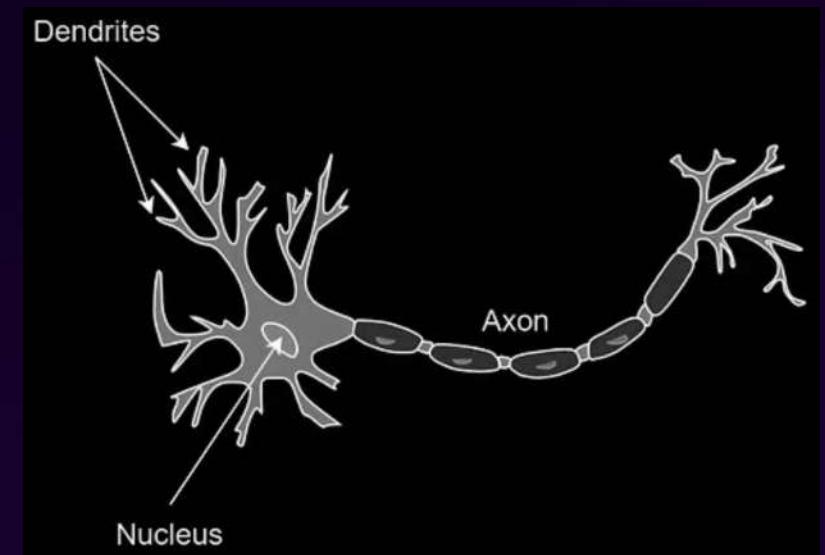


Data

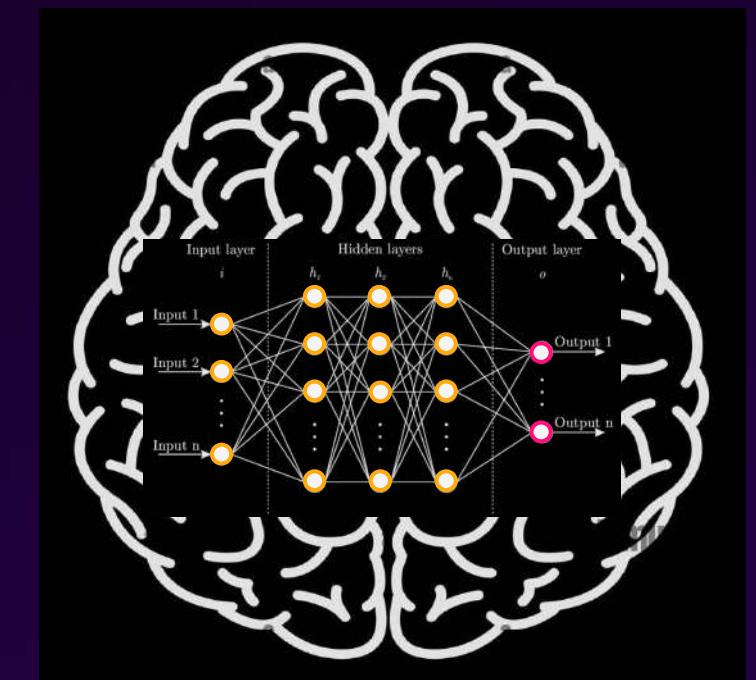
What is driving the explosive growth of AI?



1950's



2010's

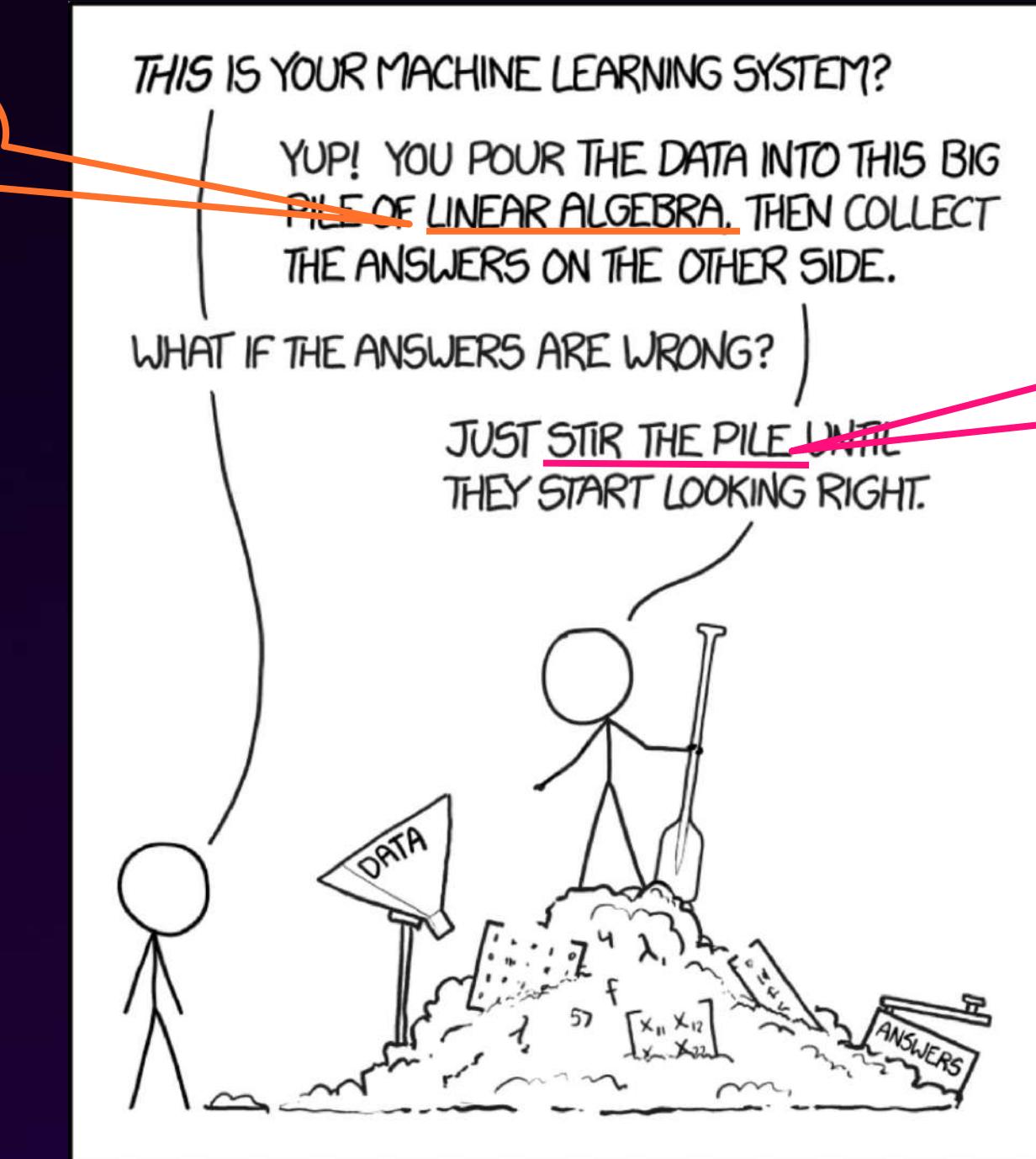


What is driving the explosive growth of AI?



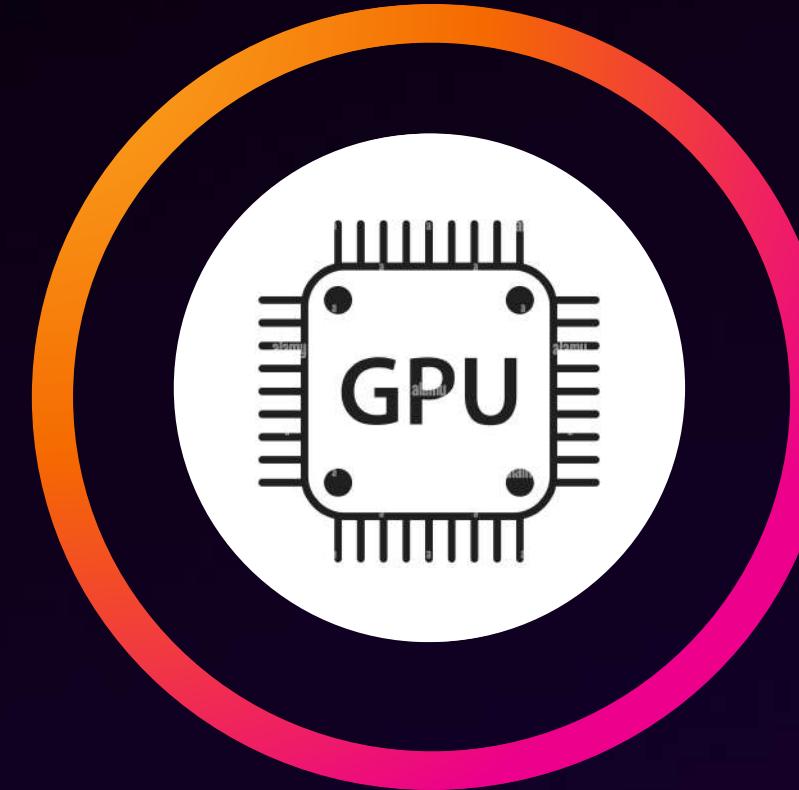
Algorithm

Matrix
Multiplication

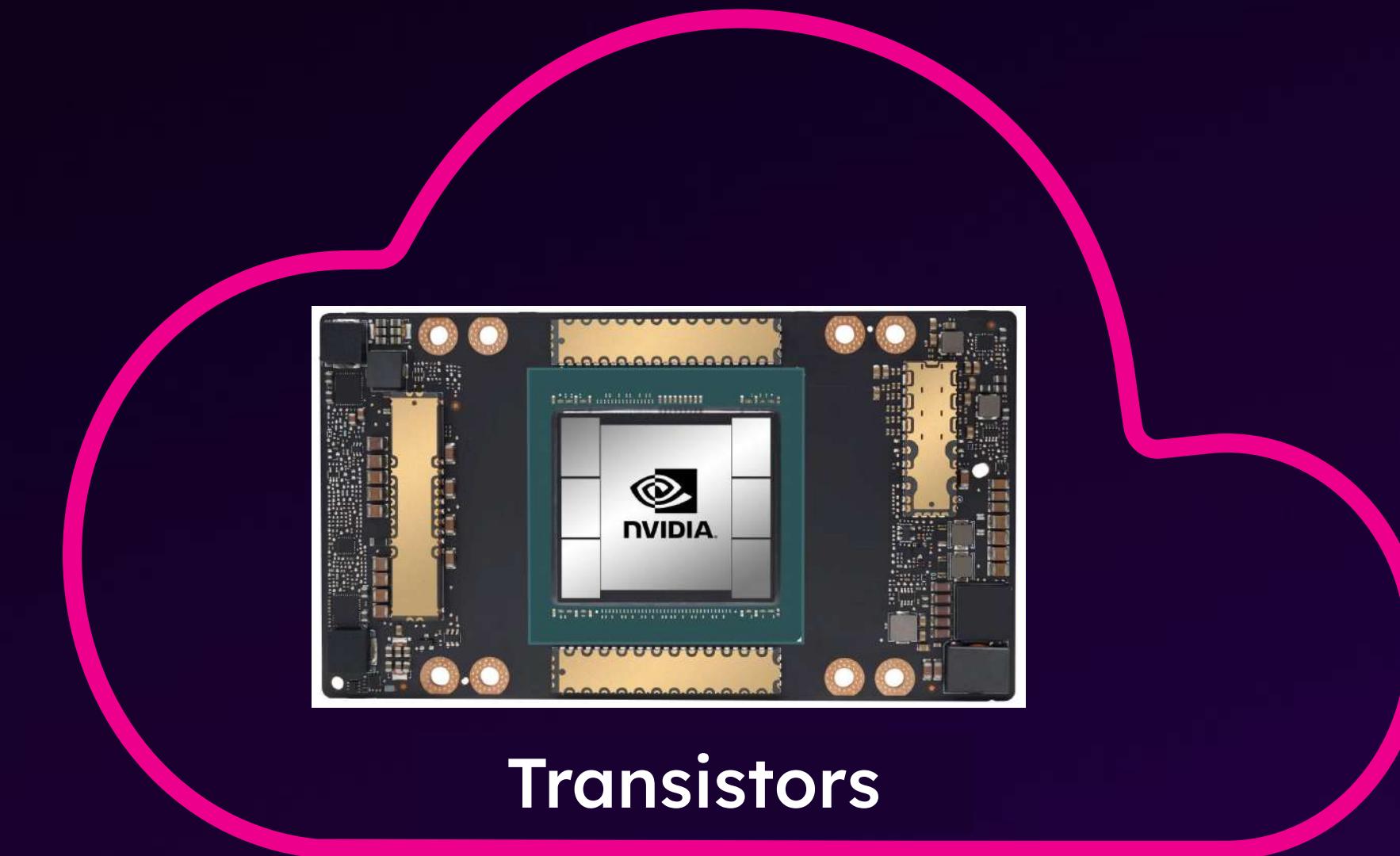


Backpropagation

What is driving the explosive growth of AI?



Compute

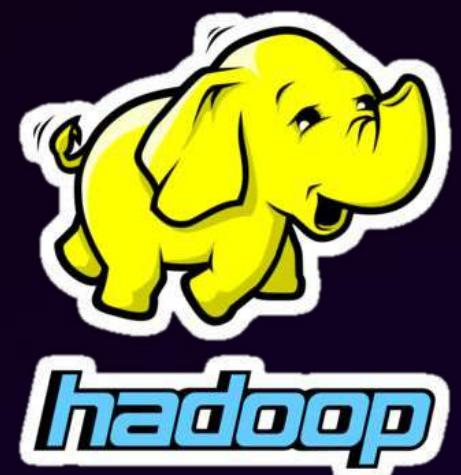


Transistors

What is driving the explosive growth of AI?



Data



WIKIPEDIA
The Free Encyclopedia



Google



Different Levels of Teaching Machines



Artificial Intelligence

The broad study of teaching a computer
to process data and make decisions

Different Levels of Teaching Machines



Artificial Intelligence

The broad study of teaching a computer
to process data and make decisions



Machine Learning

Subset of AI. Predictions and insight
with minimal human interference

Different Levels of Teaching Machines



Artificial Intelligence

The broad study of teaching a computer to process data and make decisions



Machine Learning

Subset of AI. Predictions and insight with minimal human interference



Deep Learning

Subset of ML. Predictions via neural networks

Different Levels of Teaching Machines



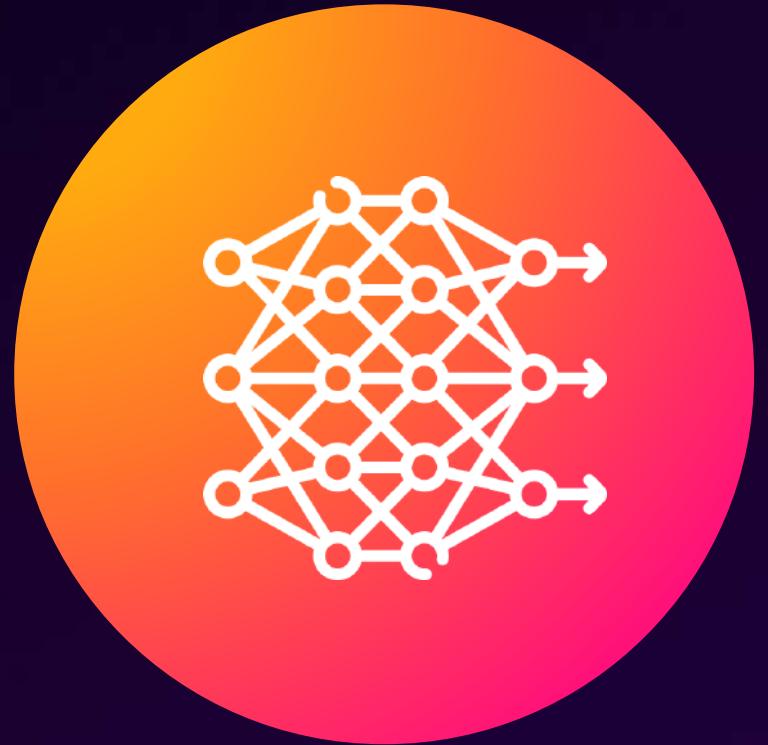
Artificial Intelligence

The broad study of teaching a computer to process data and make decisions



Machine Learning

Subset of AI. Predictions and insight with minimal human interference



Deep Learning

Subset of ML. Predictions via neural networks

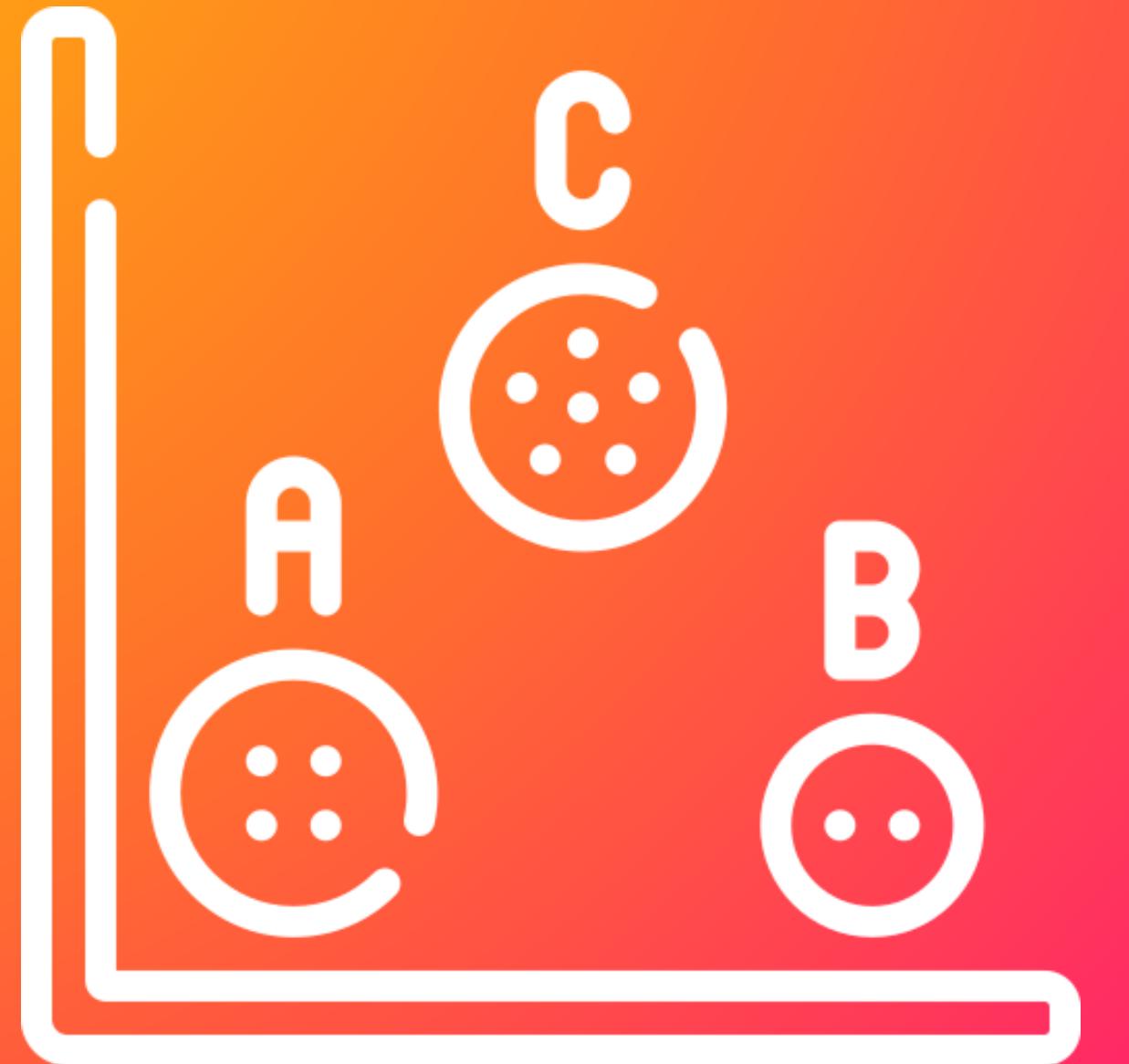
Predictive Algorithms

Methods that help you get ahead of issues that may happen in the future

Includes:

- Numerical Regression
- Categorical Regression
- Time Series Forecasting





Categorization Algorithms

Uncover insights about your data to quickly respond in the present

Includes:

- Categorical Regression
- Clustering

Outlier Detection Algorithms

Identify and analyze
abnormal behavior in
your data

Includes:

- Clustering
- Outlier Detection



Easy to Operationalize

Every Search Can Use
Machine Learning



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time

splunk®



Search

Alert



Webhook



Email



Tickets



Third-Party
Applications

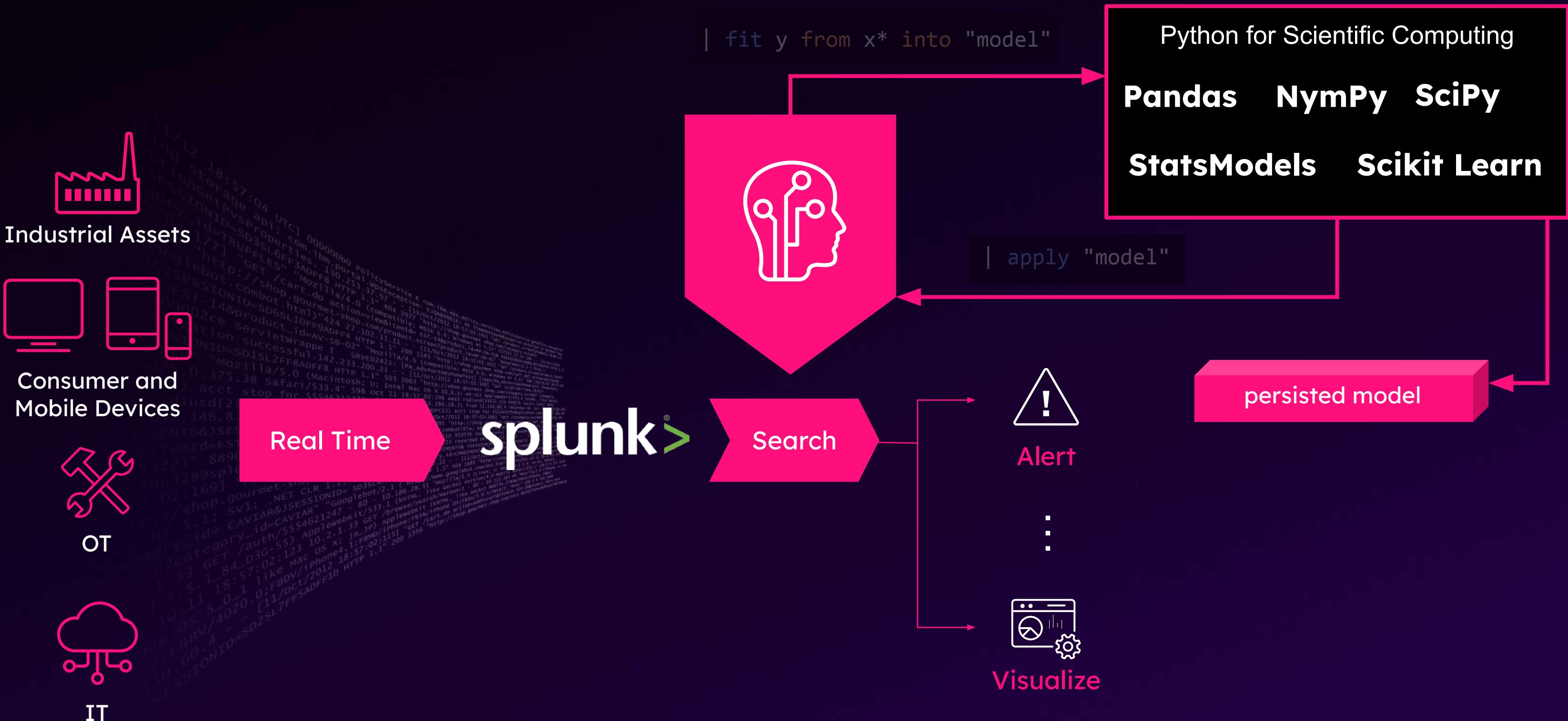


Smartphones
and Devices

Trigger process flow

Send a text

Model Longevity

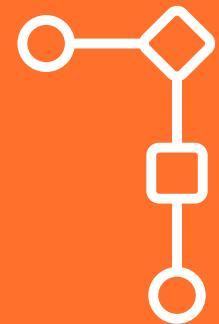


Splunk ML & AI

Where to find ML



CORE PLATFORM
SEARCH



PACKAGED (PREMIUM)
SOLUTIONS

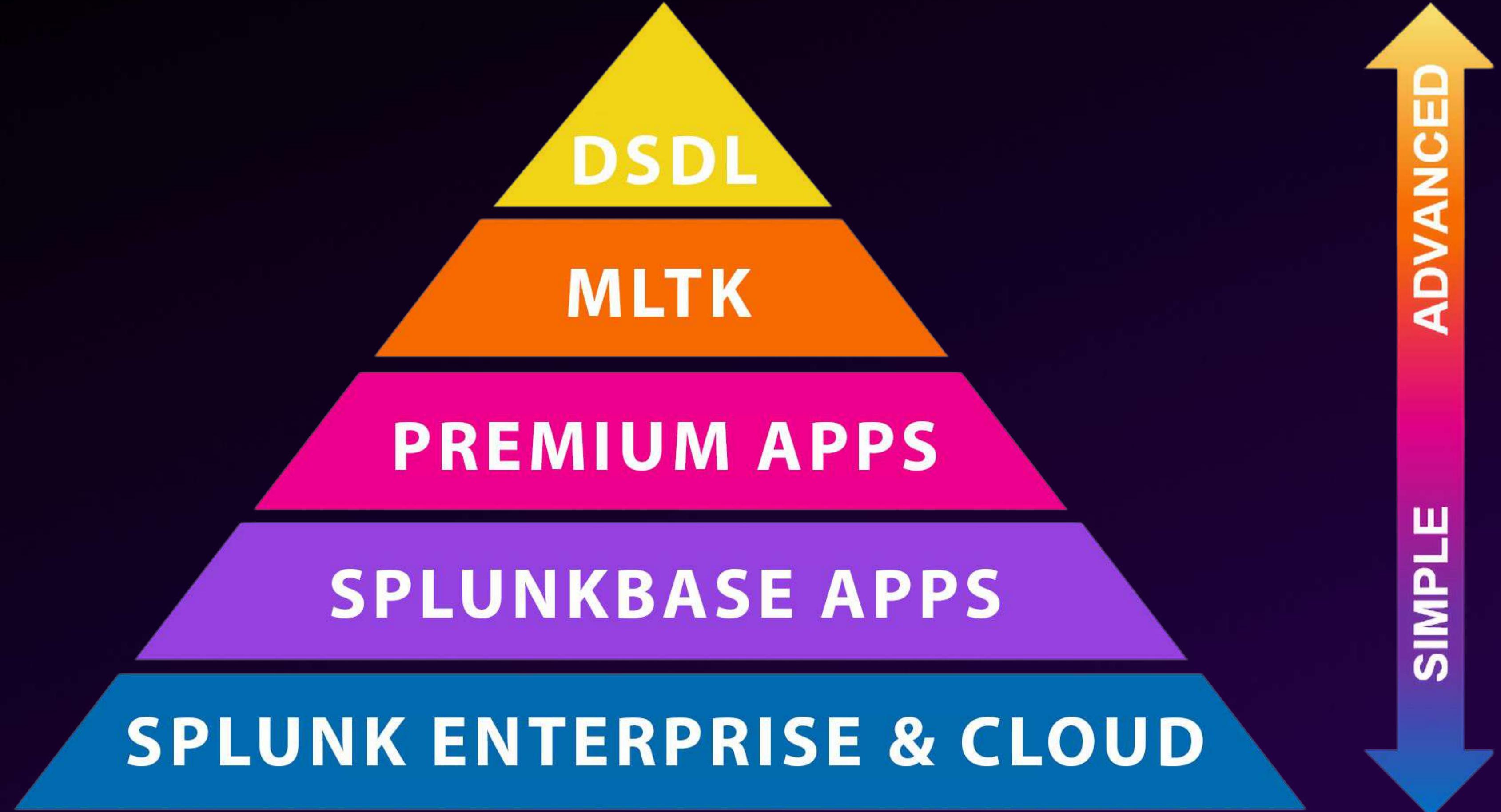


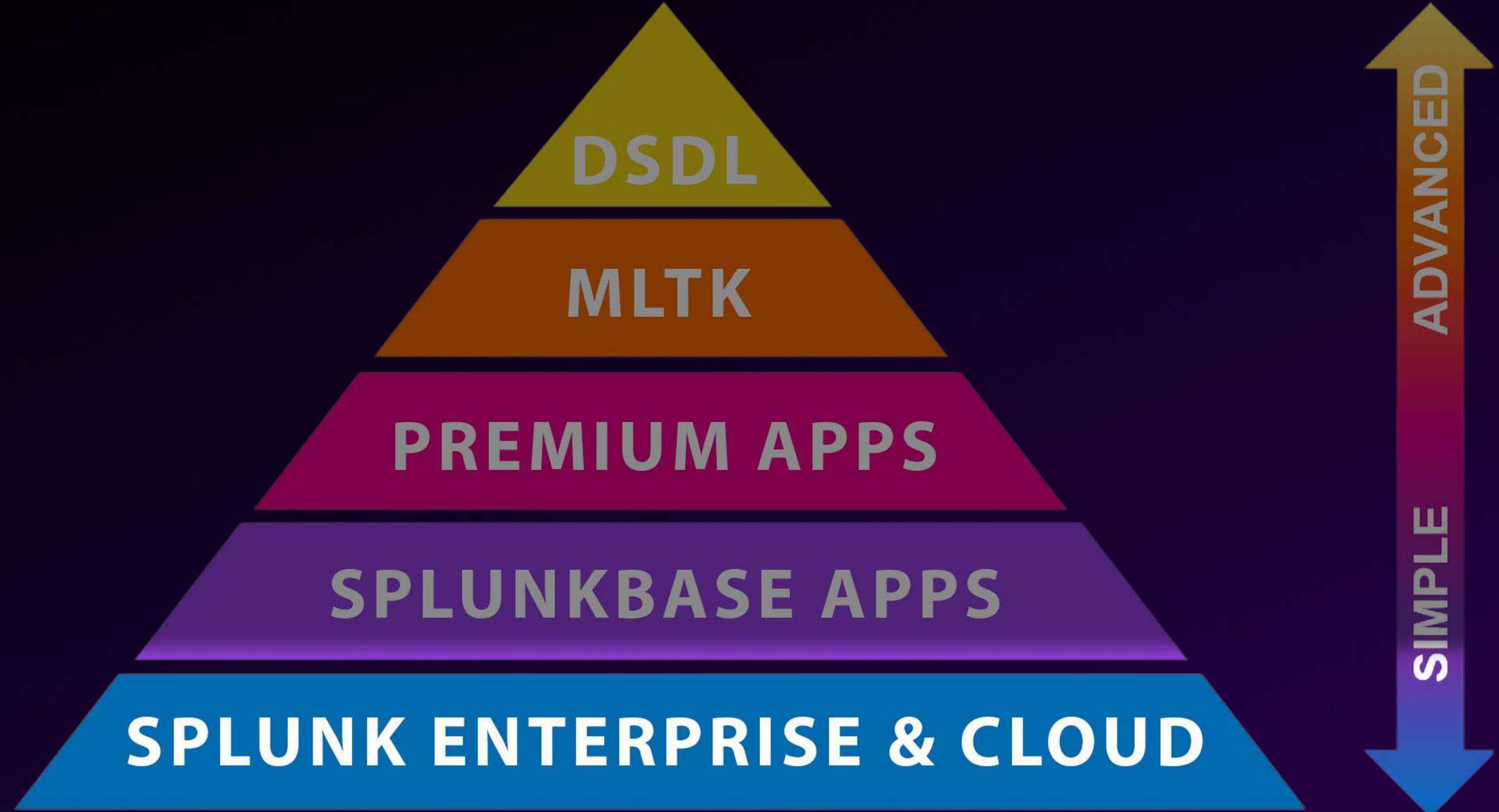
MACHINE LEARNING
TOOLKIT



SPLUNK APP FOR
DATA SCIENCE AND
DEEP LEARNING

splunk® Platform for Operational Intelligence





anomalydetection

A transforming command that identifies anomalous events by computing a probability for each event and then detecting unusually small probabilities.

<https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Anomalydetection>

The screenshot shows a web browser displaying the Splunk Search Reference documentation for the 'anomalydetection' command. The URL in the address bar is docs.splunk.com/Documentation/Splunk/9.1.3/SearchReference/Anomalydetection. The page title is 'Search Reference' under 'Splunk® Enterprise'. The left sidebar lists various search commands, with 'anomalydetection' highlighted. The main content area starts with a 'Description' section: 'A transforming command that identifies anomalous events by computing a probability for each event and then detecting unusually small probabilities. The probability is defined as the product of the frequencies of each individual field value in the event.' It includes two bullet points: 'For categorical fields, the frequency of a value X is the number of times X occurs divided by the total number of events.' and 'For numerical fields, we first build a histogram for all the values, then compute the frequency of a value X as the size of the bin that contains X divided by the number of events.' Below this is a note about using current Splunk machine learning (ML) tools. The 'Syntax' section shows the command line: 'anomalydetection [<method-option>] [<action-option>] [<pthresh-option>] [<cutoff-option>] [<field-list>]'. The 'Optional arguments' section details the '<method-option>' argument, which can be 'histogram', 'zscore', or 'iqr'. The 'Description' for this argument states: 'Select the method of anomaly detection. When method=zscore, performs like the anomalousvalue command. When method=iqr, performs like the'. The right sidebar shows a navigation menu with 'anomalydetection' selected, and a 'Previously Viewed' section.

|cluster

The cluster command groups events together based on how similar they are to each other.

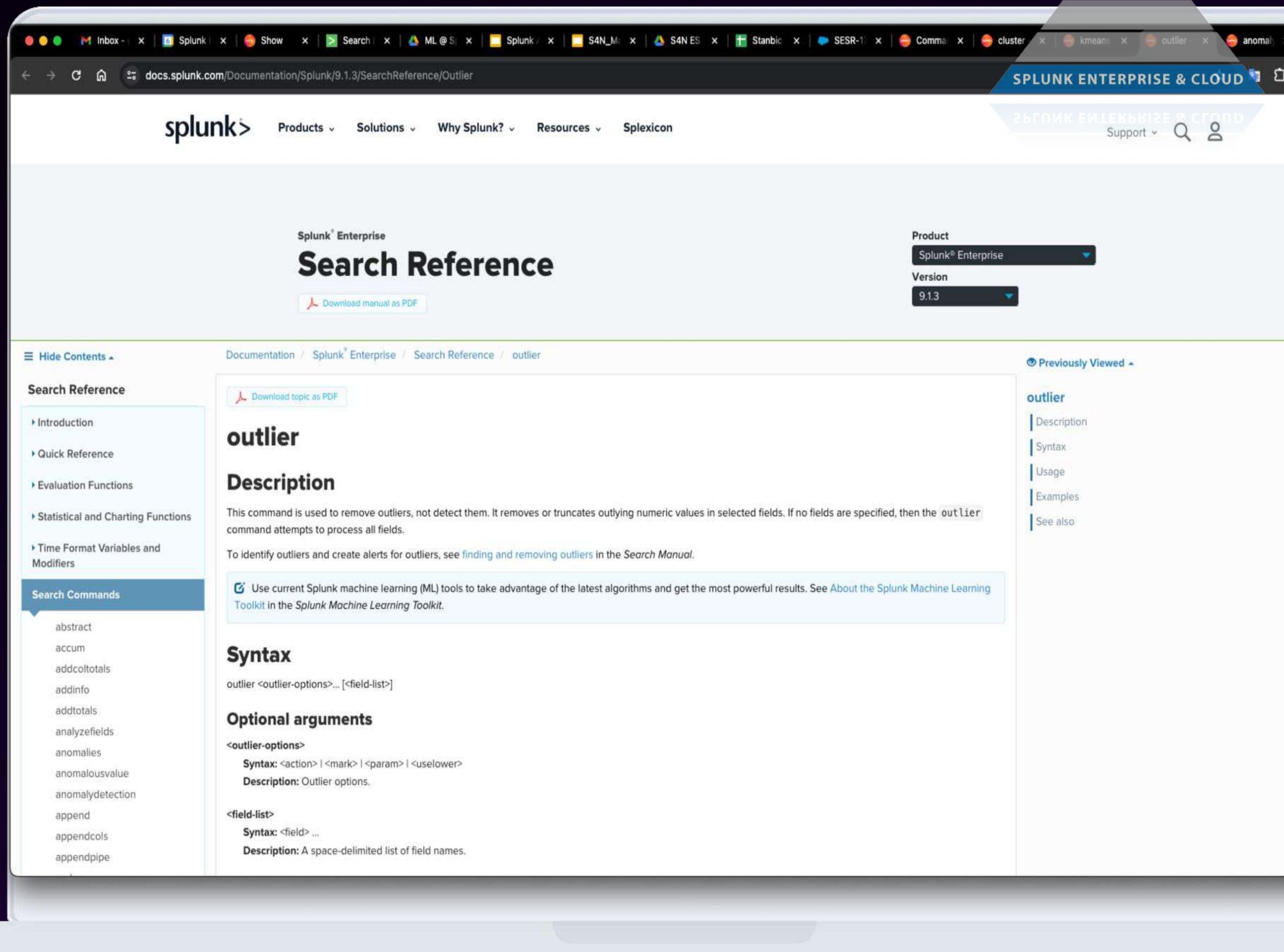
<https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Cluster>

The screenshot shows a web browser displaying the Splunk Search Reference page for the 'cluster' command. The URL in the address bar is docs.splunk.com/Documentation/Splunk/9.1.3/SearchReference/Cluster. The page title is 'Search Reference' under 'Splunk® Enterprise'. The left sidebar has a 'Search Reference' section with links to 'Introduction', 'Quick Reference', 'Evaluation Functions', 'Statistical and Charting Functions', 'Time Format Variables and Modifiers', and 'Search Commands'. The 'Search Commands' section is expanded, showing a list of commands including 'abstract', 'accum', 'addcoltotals', 'addinfo', 'addtotals', 'analyzefields', 'anomalies', 'anomalousvalue', 'anomalydetection', 'append', 'appendcols', 'appendpipe', and 'cluster'. The 'cluster' command is highlighted with a blue background. The main content area starts with a 'Description' section: 'The `cluster` command groups events together based on how similar they are to each other. Unless you specify a different field, `cluster` groups events based on the contents of the `_raw` field. The default grouping method is to break down the events into terms (`match=termlist`) and compute the vector between events. Set a higher threshold value for `t`, if you want the command to be more discriminating about which events are grouped together.' It then describes the result of the command, mentioning new fields like `countfield` and `labelfield`, and provides examples of 10 clusters labeled 1 to 10. Below this is a 'Syntax' section: 'cluster [slc-options]...'. Under 'Optional arguments', there is a 'slc-options' section with a 'Syntax' of '`t=<num>`' and a 'Description' stating it sets the cluster threshold. The right sidebar shows a 'Previously Viewed' list with 'cluster' selected, and a detailed sidebar for 'cluster' with sections for 'Description', 'Syntax', 'Usage', 'Examples', and 'See also'.

outlier

This command is used to remove outliers, not detect them. It removes or truncates outlying numeric values in selected fields.

<https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Outlier>



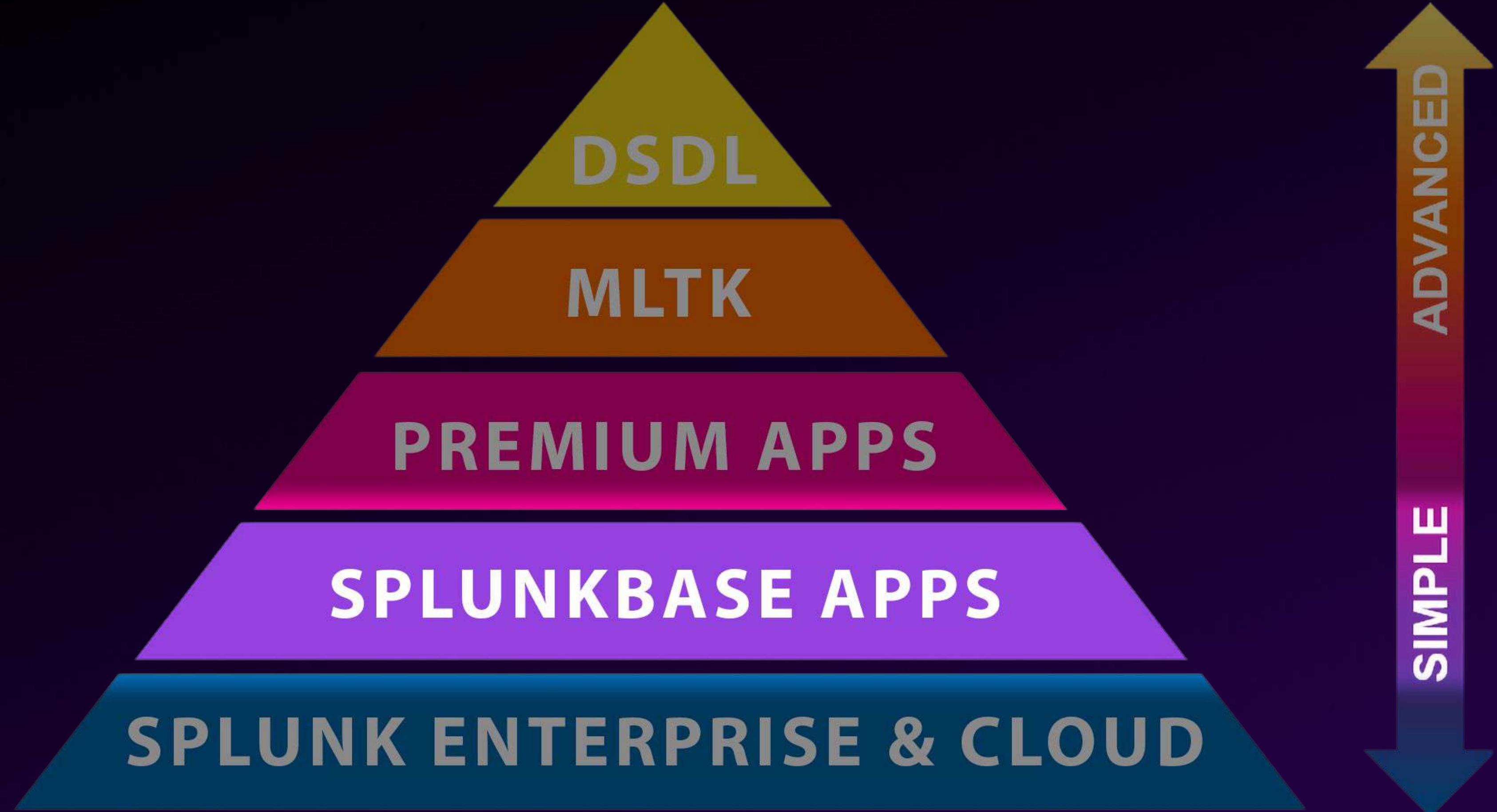
The screenshot shows a web browser displaying the Splunk Search Reference page for the 'outlier' command. The URL in the address bar is docs.splunk.com/Documentation/Splunk/9.1.3/SearchReference/Outlier. The page title is 'Search Reference' under 'Splunk® Enterprise'. The left sidebar contains a navigation menu with sections like 'Search Reference', 'Search Commands' (which is currently selected), and various other search-related commands. The main content area starts with a 'Description' section: 'This command is used to remove outliers, not detect them. It removes or truncates outlying numeric values in selected fields. If no fields are specified, then the outlier command attempts to process all fields.' Below this is a note about using Splunk Machine Learning Toolkit. The 'Syntax' section follows, with the command structure: 'outlier <outlier-options>... [<field-list>]'. There are also sections for 'Optional arguments' and 'Field lists', each with their respective syntax and descriptions. On the right side, there's a sidebar titled 'Previously Viewed' with a single entry 'outlier' and a list of its sub-sections: Description, Syntax, Usage, Examples, and See also.

kmeans

Partitions the events into k clusters, with each cluster defined by its mean value.

<https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Kmeans>

The screenshot shows a web browser displaying the Splunk Search Reference page for the `kmeans` command. The URL in the address bar is `docs.splunk.com/Documentation/Splunk/9.1.3/SearchReference/Kmeans`. The page title is "Search Reference" under "Splunk® Enterprise". The left sidebar has a "Search Reference" section with links to "Introduction", "Quick Reference", "Evaluation Functions", "Statistical and Charting Functions", "Time Format Variables and Modifiers", and a "Search Commands" section which is currently active, showing a list of commands including `abstract`, `accum`, `addcoltotals`, `addinfo`, `addtotals`, `analyzefields`, `anomalies`, `anomalousvalue`, `anomalydetection`, `append`, `appendcols`, and `appendpipe`. The main content area starts with the **kmeans** command. It includes a "Description" section stating that it partitions events into k clusters based on mean values, and a "Syntax" section showing the command as `kmeans [kmeans-options...] [field-list]`. It also lists "Required arguments" (None) and "Optional arguments" (field-list, kmeans-options). The "field-list" argument is described with a syntax of <field> ... and a description that it specifies a space-separated list of fields for joining. The "kmeans-options" argument is described with a syntax of <reps> | <iters> | <t> | <k> | <cnumfield> | <distype> | <showcentroid> and a description that it provides options for the `kmeans` command. A sidebar on the right shows a "kmeans" section with links to "Description", "Syntax", "Usage", "Examples", and "See also". The top right of the page shows "SPLUNK ENTERPRISE & CLOUD" and "Support".



- **Splunk AI Assistant for SPL (Cloud only):**

App which empowers users to search their data using plain English.

- **Smart Alerts Assistant for Splunk (beta):**

Smart Alerts Assistant for Splunk captures user feedback to improve prioritization and uses machine learning to output a recommended, stack-ranked priority for each alert.

- **Splunk App for Anomaly Detection:**

Finds anomalies in time series datasets and provides an end-to-end workflow to manage and operationalize anomaly detection tasks.

COLLECTION

Getting Started with AI

These Machine Learning and AI powered apps and assistants give you the power of Machine Learning for common use cases with just a couple of clicks.

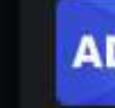
Intelligent Experiences

These ML-powered apps and assistants give you the power of ML for common use cases with just a couple of clicks.

 **Smart Alerts Assistant for Splunk (beta)**
By Splunk Works

* Description * Smart Alerts Assistant for Splunk (beta) enables analysts to focus on the most important and highest-priority alerts. It uses machine learning to prioritize alerts based on user feedback and historical data.

PLATFORM: Splunk Enterprise
RATING: ★★★★☆ (2)
ARCHIVED APP

 **Splunk App for Anomaly Detection**
By Splunk LLC

The Splunk App for Anomaly Detection finds anomalies in time series datasets and provides an end-to-end workflow to manage and operationalize anomaly detection tasks.

PLATFORM: Splunk Enterprise, Splunk Cloud
RATING: ★★★★☆ (5)
SPLUNK SUPPORTED APP

 **Splunk AI Assistant for SPL (Cloud Version)**
By Splunk LLC

Splunk AI Assistant for SPL offers bidirectional translation between natural language (NL) and Splunk Search Language (SSL).

PLATFORM: Splunk Enterprise, Splunk Cloud
RATING: ★★★★★ (2)
SPLUNK SUPPORTED APP

https://splunkbase.splunk.com/collections/machine_learning



AI Assistant for SPL

Increased efficiency
for users

- ▶ Get your job done faster using natural language.
- ▶ Drill down to deeper insights.
- ▶ Accelerate your learning journey.

What disk has the least amount of freespace in WinHostMon?

To answer this question, we will use the following SPL query:

```
index="[" sourcetype="WinHostMon" Type=Disk
| eval FreeSpacePct=round(FreeSpaceKB/TotalSpaceKB*100)
| eval TotalSpaceGB=round(TotalSpaceKB/1024/1024)
| eval FreeSpaceGB=round(FreeSpaceKB/1024/1024)
| stats min(FreeSpacePct) as MinFreeSpacePct, min(TotalSpaceGB) as MinTotalSpaceGB, min(FreeSpaceGB) as
MinFreeSpaceGB by host
| table host, MinTotalSpaceGB, MinFreeSpaceGB, MinFreeSpacePct
```

This query will:

1. Filter the WinHostMon data from the specified index.
2. Calculate the free space percentage, total space in GB, and free space in GB for each disk.
3. Group the results by host and find the minimum values for each of these fields.
4. Display the results in a table format with the host, minimum total space in GB, minimum free space in GB, and minimum free space percentage.

[Write SPL](#) [Explain SPL](#) [Tell me about...](#)

[Copy](#) [Open in Search](#)

Splunk AI Assistants everywhere

SPLUNKBASE APPS

Our goal: Make everyone a Splunk expert!

AI Assistant for SPL

The screenshot shows the AI Assistant for SPL interface. A modal window is open, displaying a Splunk search query to detect the creation of a new local user in Windows. The query is:

```
index=<index> source=*WinEventLog*:Security AND EventID=4624 | dedup session_id | table _time, host, User, TargetUserName, Subject
```

Below the query, it says: "This query will return a table with the following fields: - _time: The time of the event - host: The host where the event occurred - User: The user who created the new user account - Name: The name of the new user account".

Generally Available

AI Assistant in Enterprise Security

The screenshot shows the AI Assistant in Enterprise Security interface. It displays an investigation titled "Geographically Improbable Access Detected". The details include:

- Source Host: workstation-478.internal.domain
- Destination Host: server-932.internal.domain
- Source IP Address: 192.168.10.47
- Destination IP Address: 192.168.10.123
- Protocol: SMB

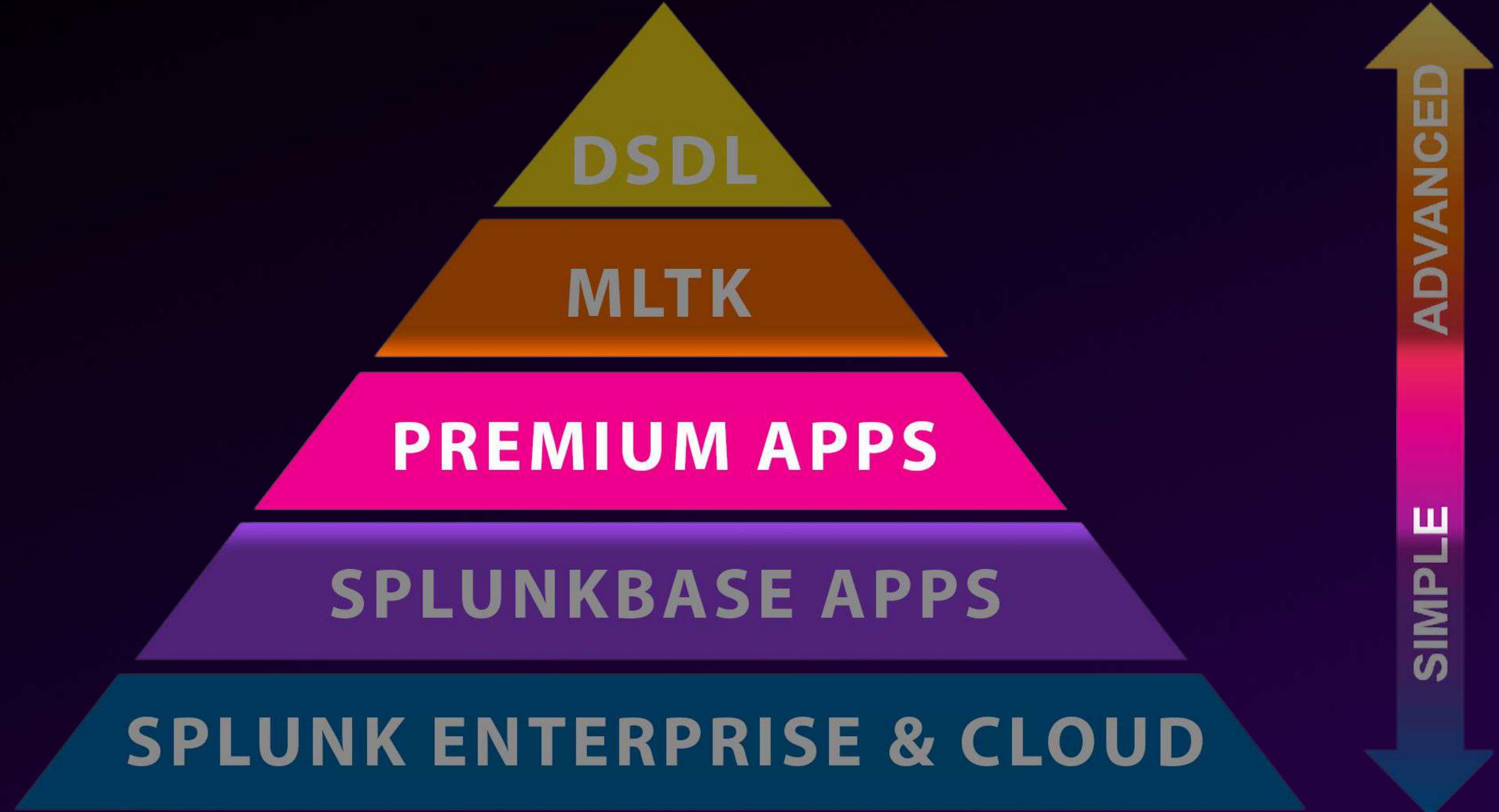
Event Information: Files accessed and transferred: \\server-932\share\payroll, \\server-932\share\config. Additional Context: 1. The source workstation ('workstation-478') has regular connections to this server.

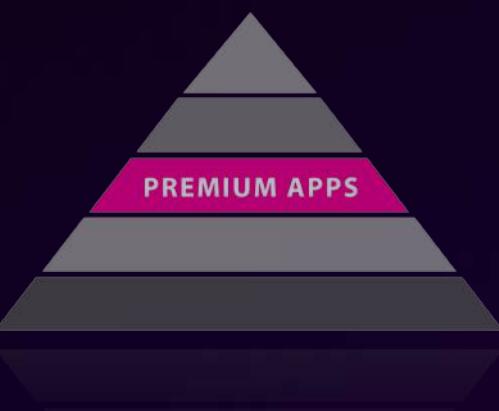
Private Preview

AI Assistant in Observability Cloud

The screenshot shows the AI Assistant in Observability Cloud interface. On the left, there's a dependency map for the "frontend" service, showing various components like "advertisingservice", "checkoutservice", "paymentservice", etc., with their connection times. On the right, there's a chat interface with a message from "AZ": "Why is frontend having issues?". Below the map, a message from "AI Assistant" says: "Here are some possible reasons: it seems like frontend is having issues downstream dependency 'paymentservice' is experiencing problems." There's also a section for "Checking the health map".

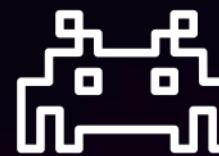
Private Preview





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-Powered Detections

Build a model 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



Test Detections

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

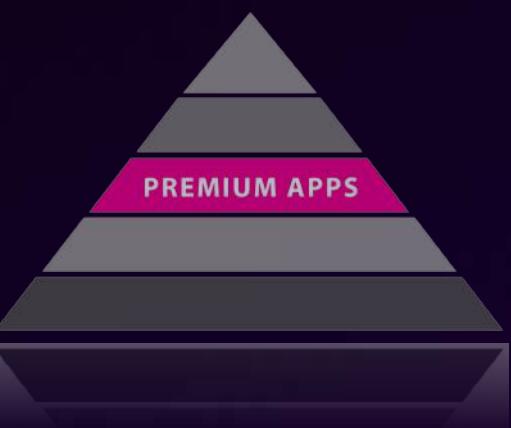


Release

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

ESCU

Enterprise Security Content Updates



Security Content

- Detections
- Analytic Stories
- Playbooks
- Blog
- About
-

machine learning

78 Result(s) found

[Splunk Command and Scripting Interpreter](#)

[Risky SPL MLTK](#)

Try in Splunk Security Cloud Description This detection utilizes machine learning model named "risky_command_abuse" trained from "Splunk Command and Scripting..."

[Potentially malicious code on commandline](#)

Try in Splunk Security Cloud Description The following analytic uses a pretrained machine learning text classifier to detect potentially malicious...

[Azure Active Directory High Risk Sign-in](#)

Try in Splunk Security Cloud Description The following analytic triggers on a high risk sign-in against Azure Active Directory identified...

[Modification Of Wallpaper](#)

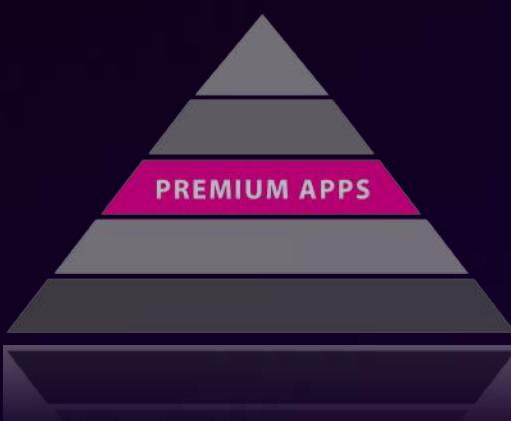
Try in Splunk Security Cloud Description This analytic identifies suspicious modification of registry to deface or change the wallpaper of...

Splunk ES Content Update

The Splunk ES Content Update (ESCU) app delivers pre-packaged Security Content. ESCU provides regular Security Content updates to help security practitioners address ongoing time-sensitive threats, attack methods, and other security issues. Security Content consists of tactics,...

Built by [Splunk Inc.](#)

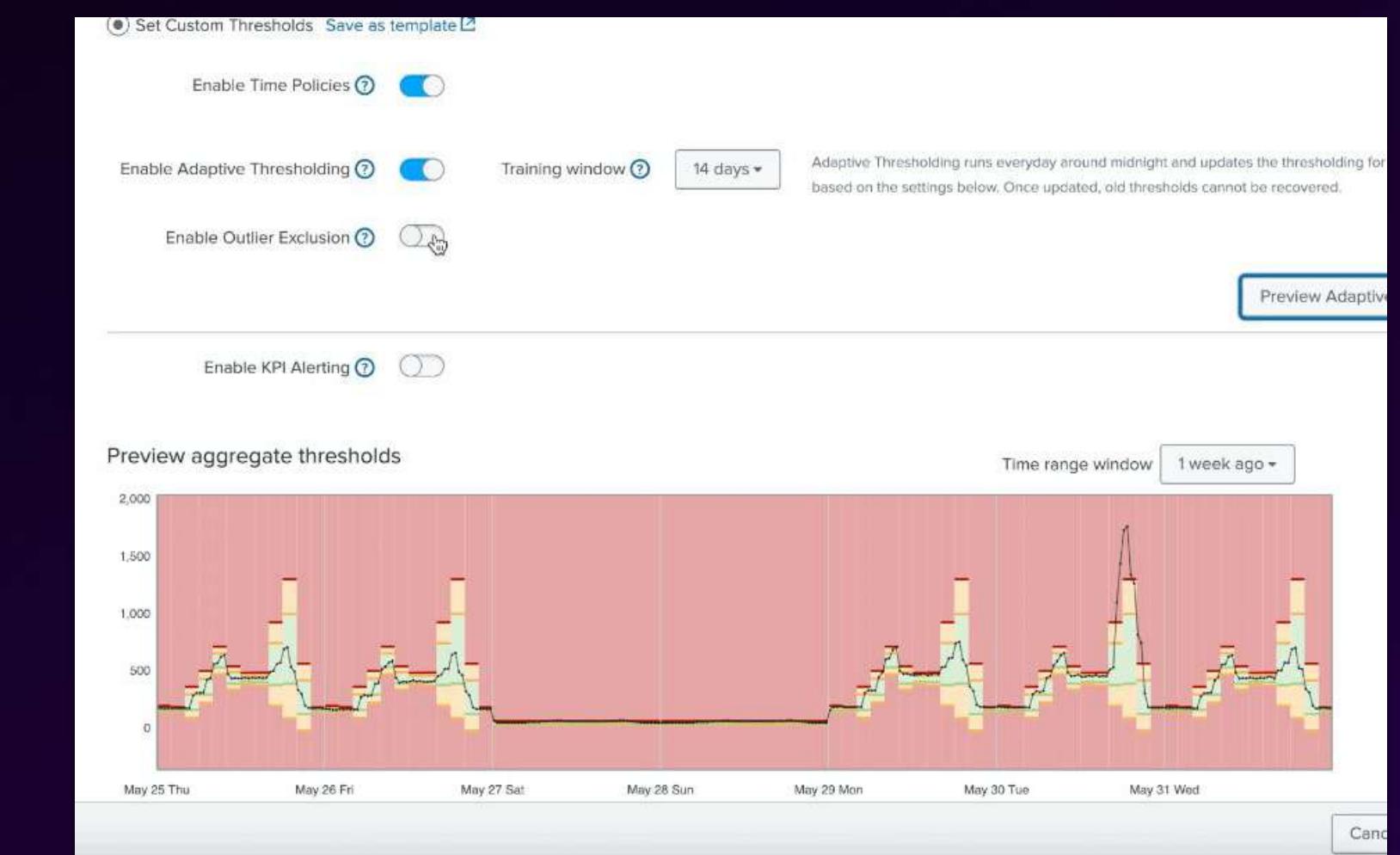
Welcome to Enterprise Security Content Updates (ESCU), brought to you by the Splunk Security Research Team!



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



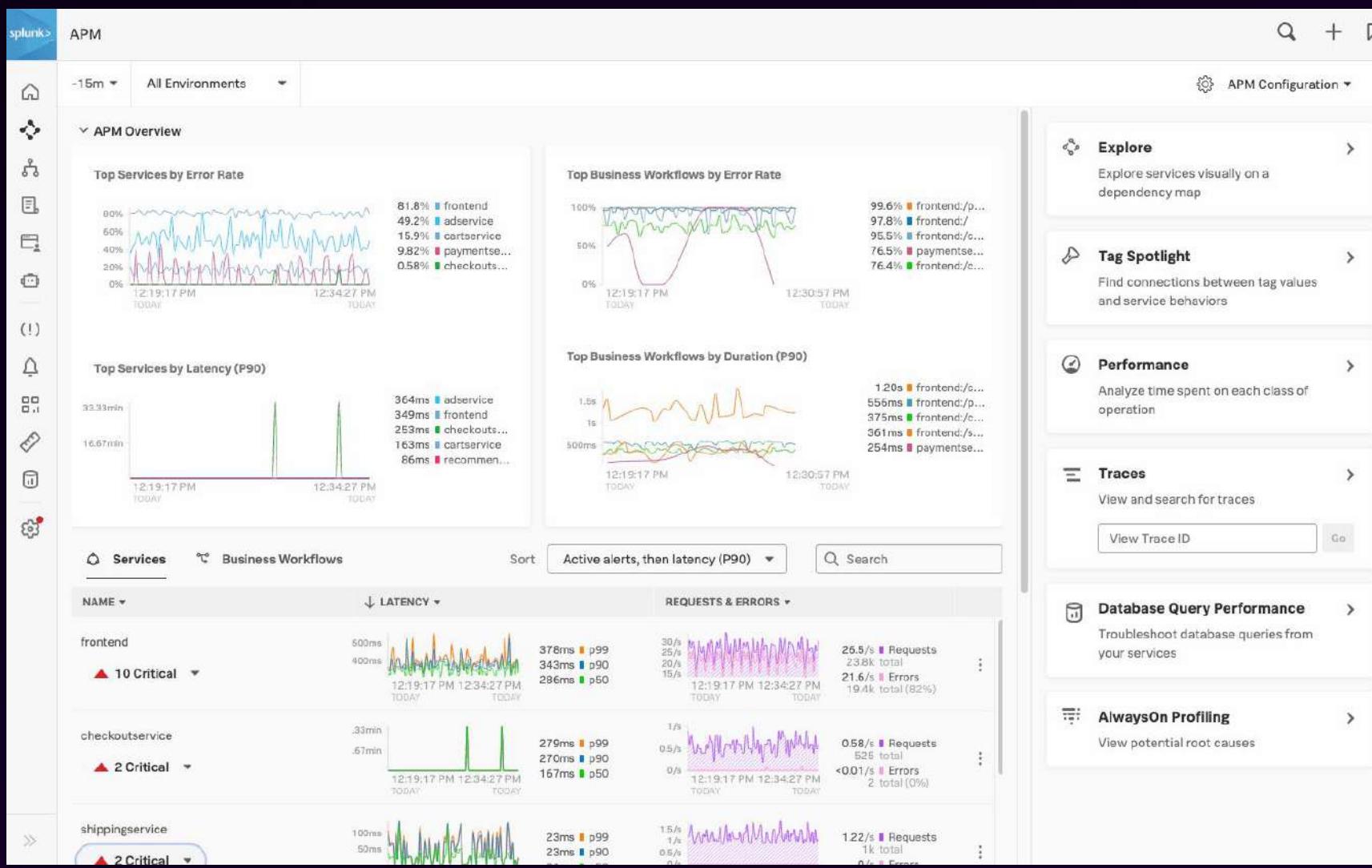
New updates!

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



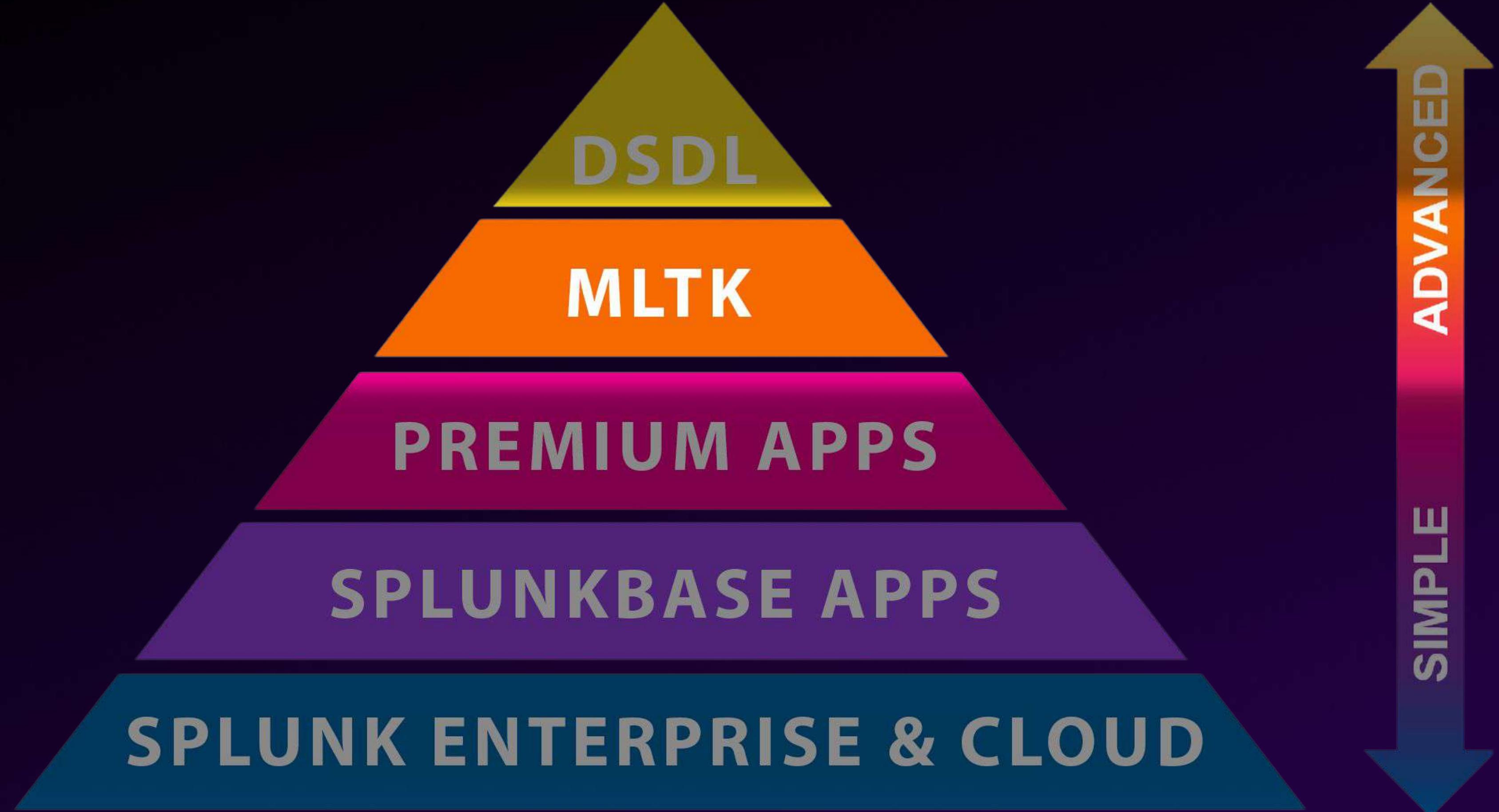
Application and Infrastructure Monitoring: Autodetect

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



Python for Scientific Computing (PSC)

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



Splunk Machine Learning Toolkit (MLTK)

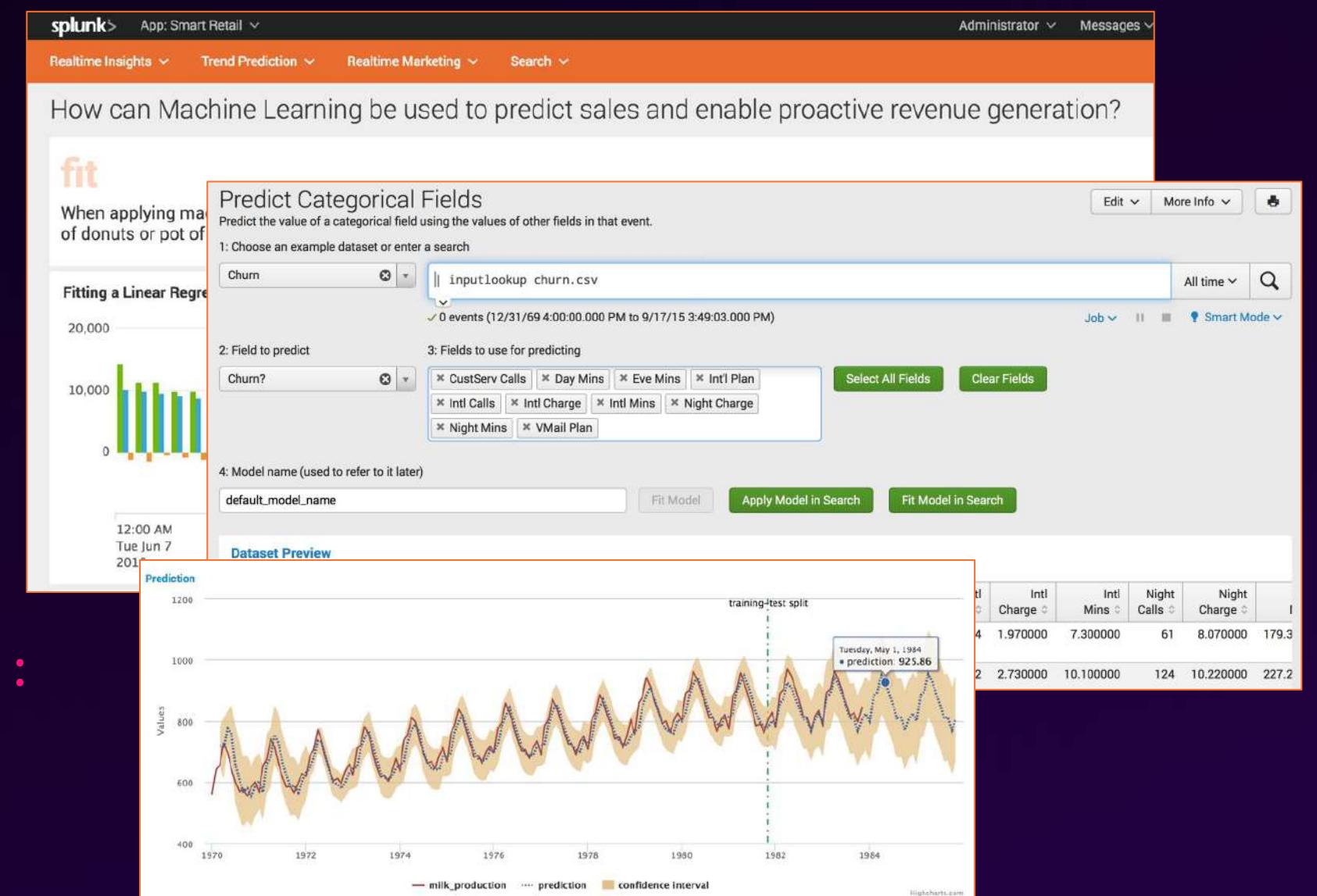
Extends Splunk platform functions and provides a guided modeling environment

Built for the Citizen Data Scientist

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

Extensible to operationalize any use case

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



Plot the outliers

```
| inputlookup logins.csv | predict logins as prediction algorithm=LLP future_timespan=150 holdback=0 | where prediction!="0" AND logins!="0" | eval residual = prediction - logins
| streamstats window=72 current=true median("logins") as median // calculate the median value using a sliding window
| eval absDev=(abs('logins'-median)) // calculate the absolute deviation of each value from the median
| streamstats window=72 current=true median(absDev) as medianAbsDev // use the same sliding window to compute the median absolute deviation
| eval lowerBound=(median-medianAbsDev*exact(9)), upperBound=(median+medianAbsDev*exact(9)) // calculate the bounds as a multiple of the median absolute deviation
| eval isOutlier=if('logins' < lowerBound OR 'logins' > upperBound, 1, 0) // mark values outside the bounds as outliers
| fields _time, "logins", lowerBound, upperBound, isOutlier, * // format the columns to be in the order expected by the Outliers Plot visualization
```

Assistant Settings

Enter a search

```
| inputlookup logins.csv | predict logins as prediction algorithm=LLP future_timespan=150 holdback=0 | where prediction!="0" AND logins!="0" | eval residual = prediction - logins
```

✓ 708 results (01/01/1970 00:00:00.000 to 23/04/2024 11:20:38.000)

Job ▾ All time Smart Mode

Field to analyze: logins Threshold method: Median Absolute Deviation Threshold multiplier: 9 Sliding window (# of values): 72 Include current point Fields to split by (optional)

Detect Outliers Open in Search Show SPL

Data and Outliers

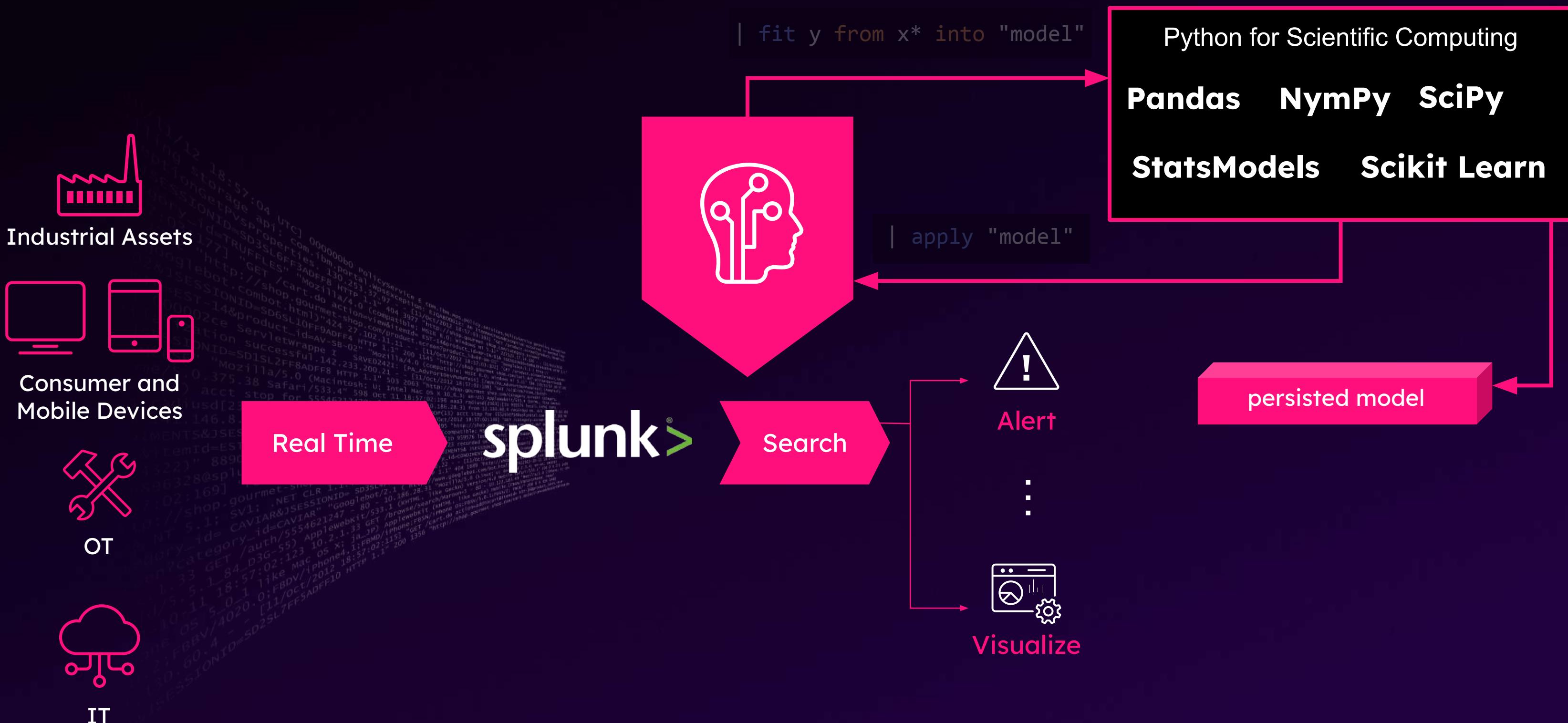
logins

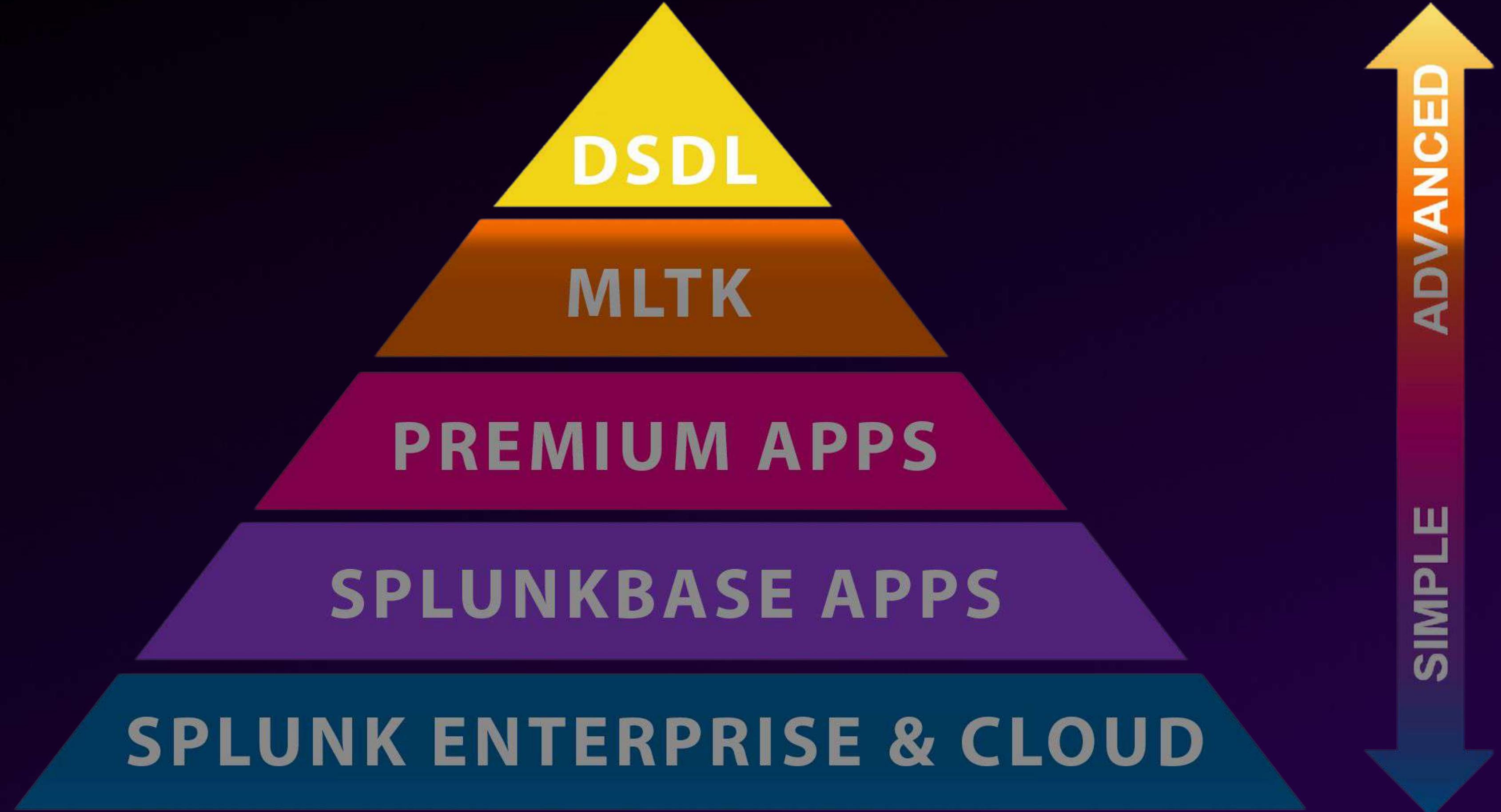
1500
1000
500
0
-500
-1000

2. Nov 4. Nov 6. Nov 8. Nov 10. Nov 12. Nov 14. Nov 16. Nov 18. Nov 20. Nov 22. Nov 24. Nov 26. Nov 28. Nov 30. Nov

19 outliers

Model Longevity





Splunk App for Data Science and Deep Learning

aka. Deep Learning Toolkit for Splunk (DLTK)

Built for Data Scientists

- **Frameworks:**

Freely available app for advanced data science projects using any open-source AI frameworks contains PyTorch, Tensorflow 2.0, SpaCey, Jupyter Notebook, & lot more

- **Code Examples:**

Guided model building, testing, and deployment of deep learning frameworks

- **GPU Support:**

Speed up your data science projects with GPU accelerated containers

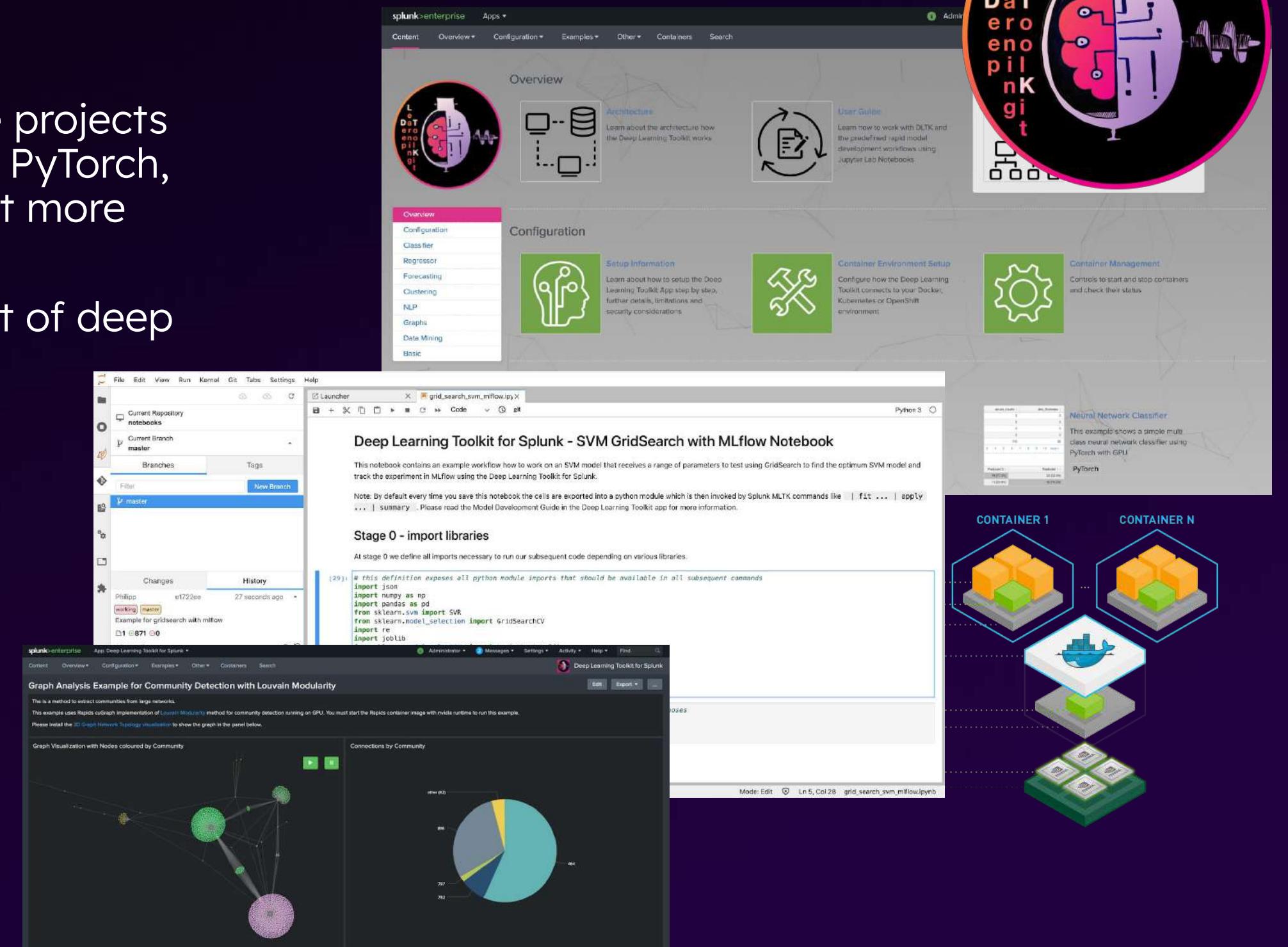
- **K8s Support:**

Scalable and HA with K8s deployment

- **Open source for customization:**

<https://github.com/splunk/splunk-mltk-container-docker>

This is now becoming a "must-have" for every citizen data scientists



DLTK for Splunk



Industrial Assets



Consumer and
Mobile Devices



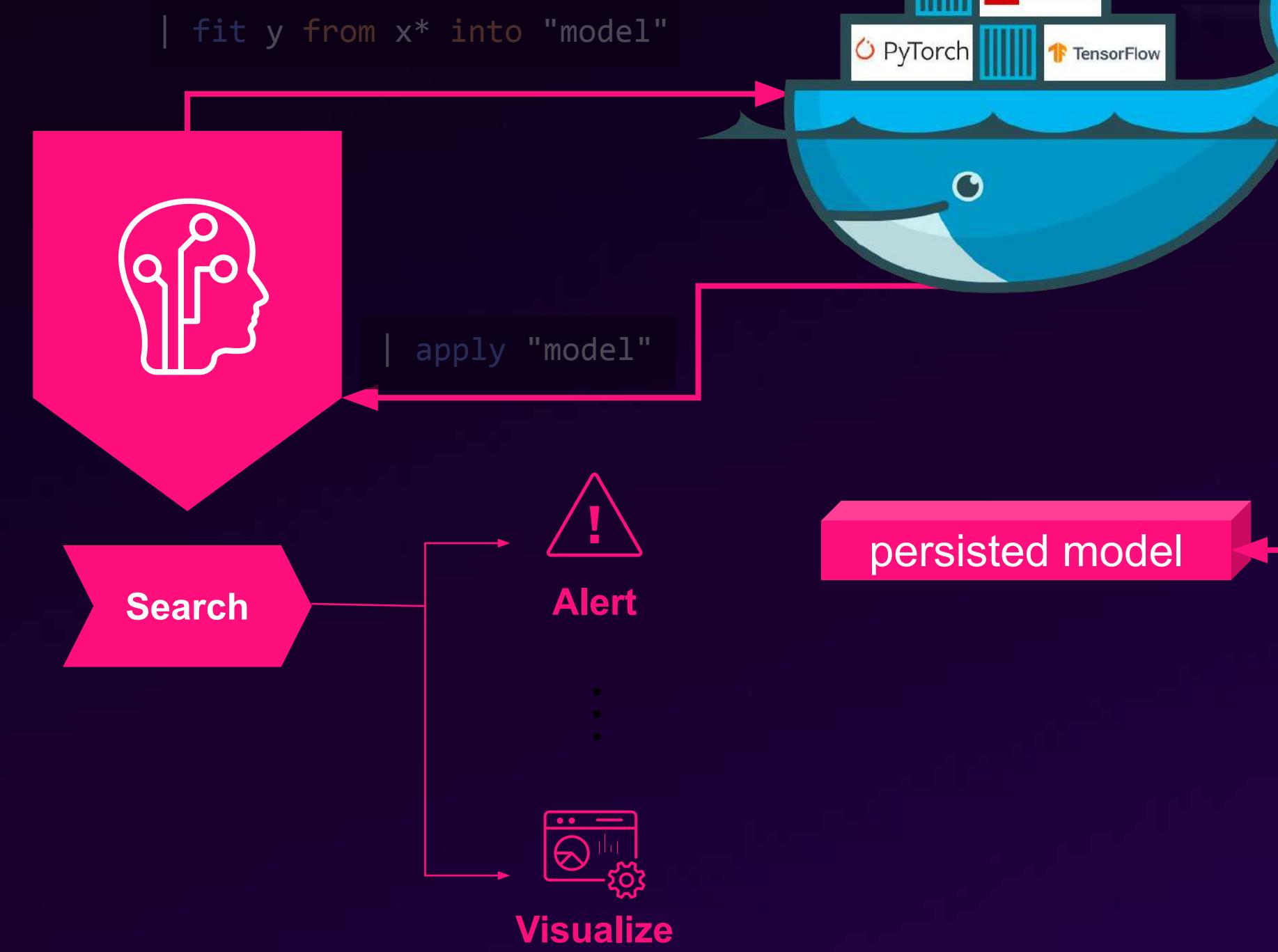
OT



IT

Real Time

splunk®



Road to SUCCESS

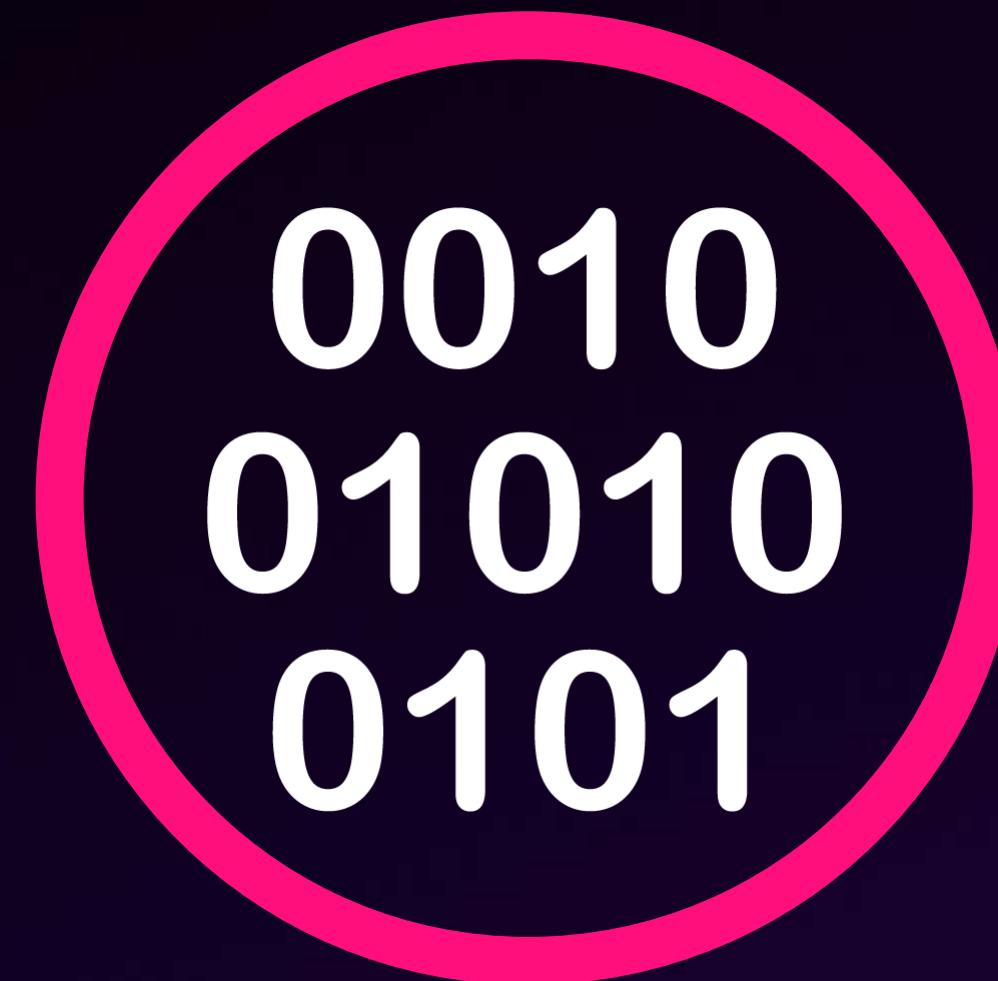
What steps will help you to make your ML project a success.

Includes:

- Understanding your data
- Cleaning/munging your data
- Operationalizing the data set



Numeric



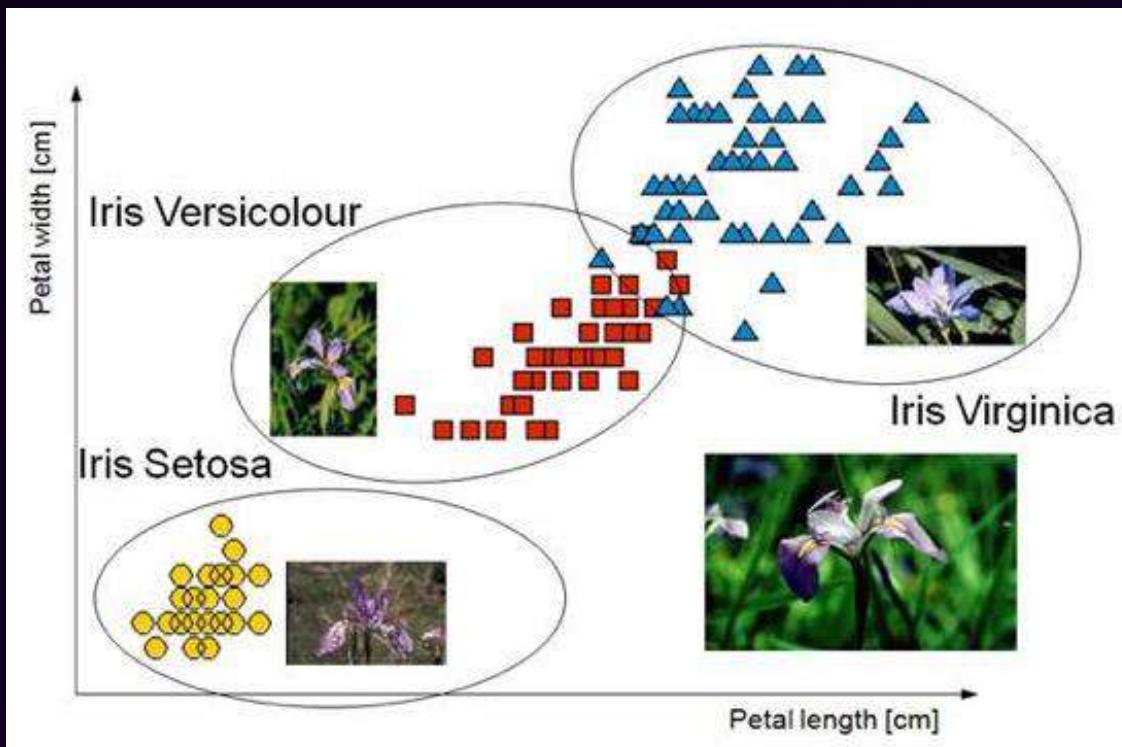
Categorical



Types of Machine Learning

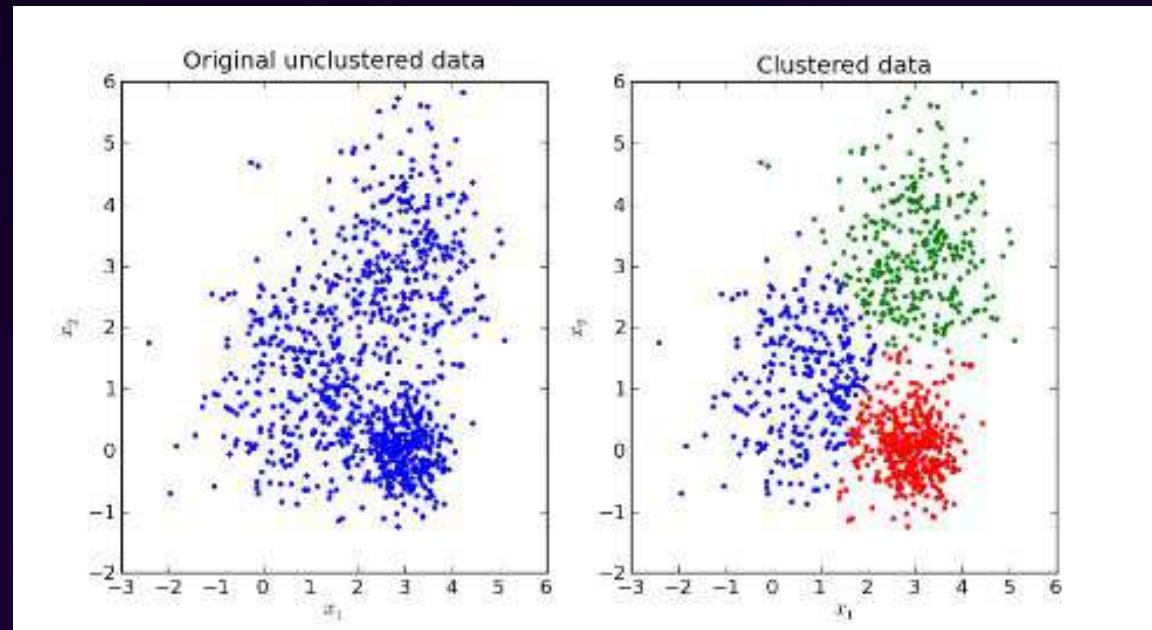
Supervised Learning (labeled data)

- ▶ Regression
- ▶ Classification



Unsupervised Learning (unlabeled data)

- ▶ Clustering
- ▶ Anomaly Detection



Mixed Models (with reinforcement or feedback)

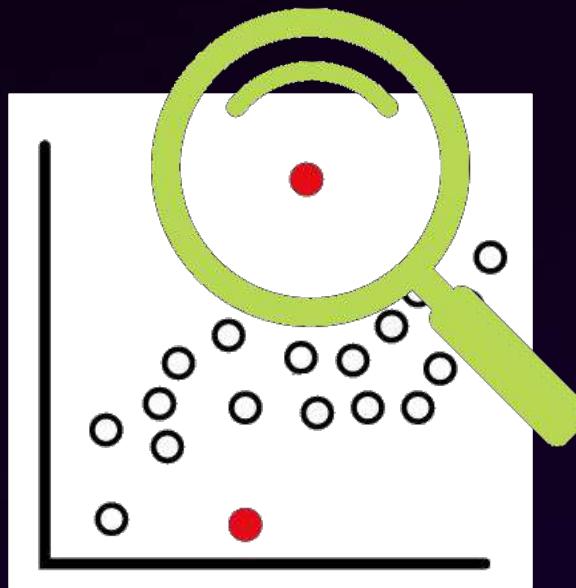
- ▶ Human in the Loop
- ▶ Autonomous Systems



Where to start

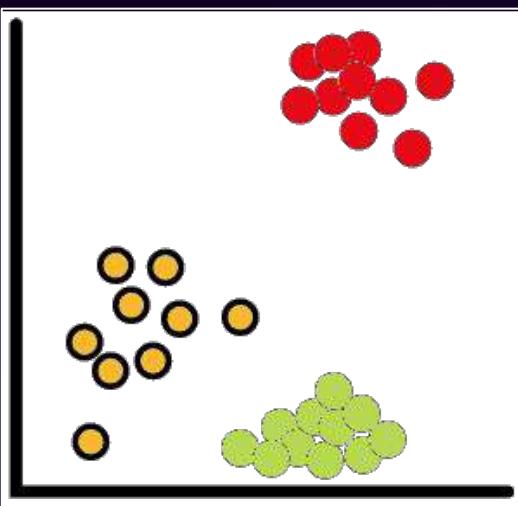


Anomaly detection



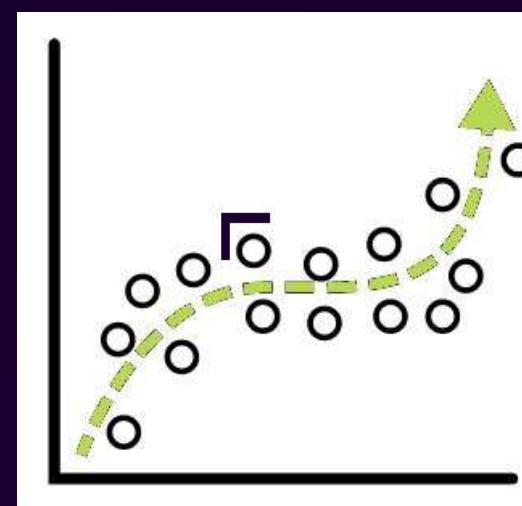
- ▶ Deviation from past behavior
- ▶ Deviation from peers
(aka Multivariate AD or Cohesive AD)
- ▶ Unusual change in features

Clustering



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

Predictive Analytics



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

Data understanding

Is it good data?

Machine data is complex but to get answers of your data, it is important to understand your data.

A clean and usable data set does not have:

- Missing data
- Noisy data
- Bias
- Duplicate data
- outliers

If your data set does not contain any of this you can start applying ML.

Otherwise, it's time to gather and clean!

Modern AI Stack: The Emerging Building Blocks for GenAI

Layer 4: Observability		OBSERVABILITY, EVALUATION, SECURITY									
Layer 3: Deployment		PROMPT MANAGEMENT				ORCHESTRATION					
Layer 2: Data		AGENT TOOL FRAMEWORKS				ETL + DATA PIPELINES					
Layer 1: Compute + Foundation	Observability	Helicone	AgentOps	Humanloop	Credal.ai	CALYPSO AI	truera	eppo	:BRAINTRUST	Patronus AI	splunk>
	PROMPT MANAGEMENT	vellum	LangSmith				Martian	orkes		Radiant	
	AGENT TOOL FRAMEWORKS	LangChain	Auto gpt		FIXIE		LlamaIndex				
Data	DATA PRE-PROCESSING	gable	datologyai	Cleanlab			NOMIC	Lexy	Indexify		
DATABASES (VECTOR, DB, METADATA STORE, CONTEXT CACHE)											
	databricks	upstash	Pinecone	NEON	WarpStream	momento					
Layer 1: Compute + Foundation	MODEL DEPLOYMENT + INFERENCE	baseten	Modal	Replicate	clarifai	Substrate	fireworks.ai	LAMINI	Predibase	arcee.ai	
	FOUNDATION MODELS	OpenAI	ANTHROPIC	MISTRAL AI_	contextual.ai	Hugging Face	Llama 2	Modular	Lightning ^{AI}	OctoML	
	GPU PROVIDERS	aws	Azure	Google Cloud	CoreWeave	Lambda	FOUNDRY	together.ai			

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.

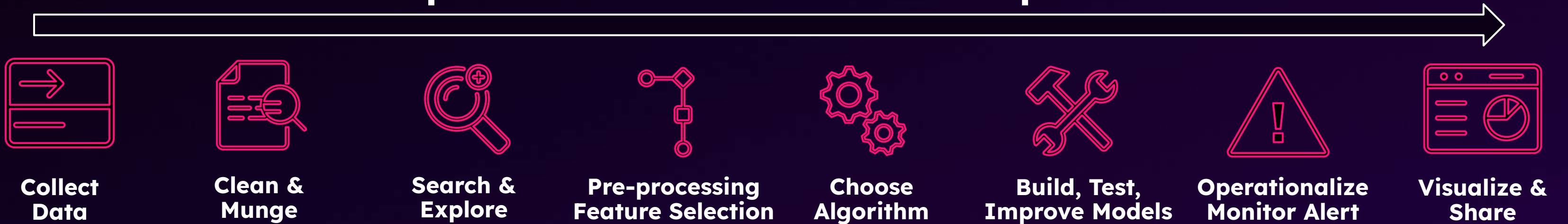
MLTK

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



Ecosystem	Ecosystem	Ecosystem	MLTK	MLTK	MLTK	MLTK	Ecosystem
Splunk	Splunk	Splunk	Splunk	Splunk	Splunk	Splunk	Splunk

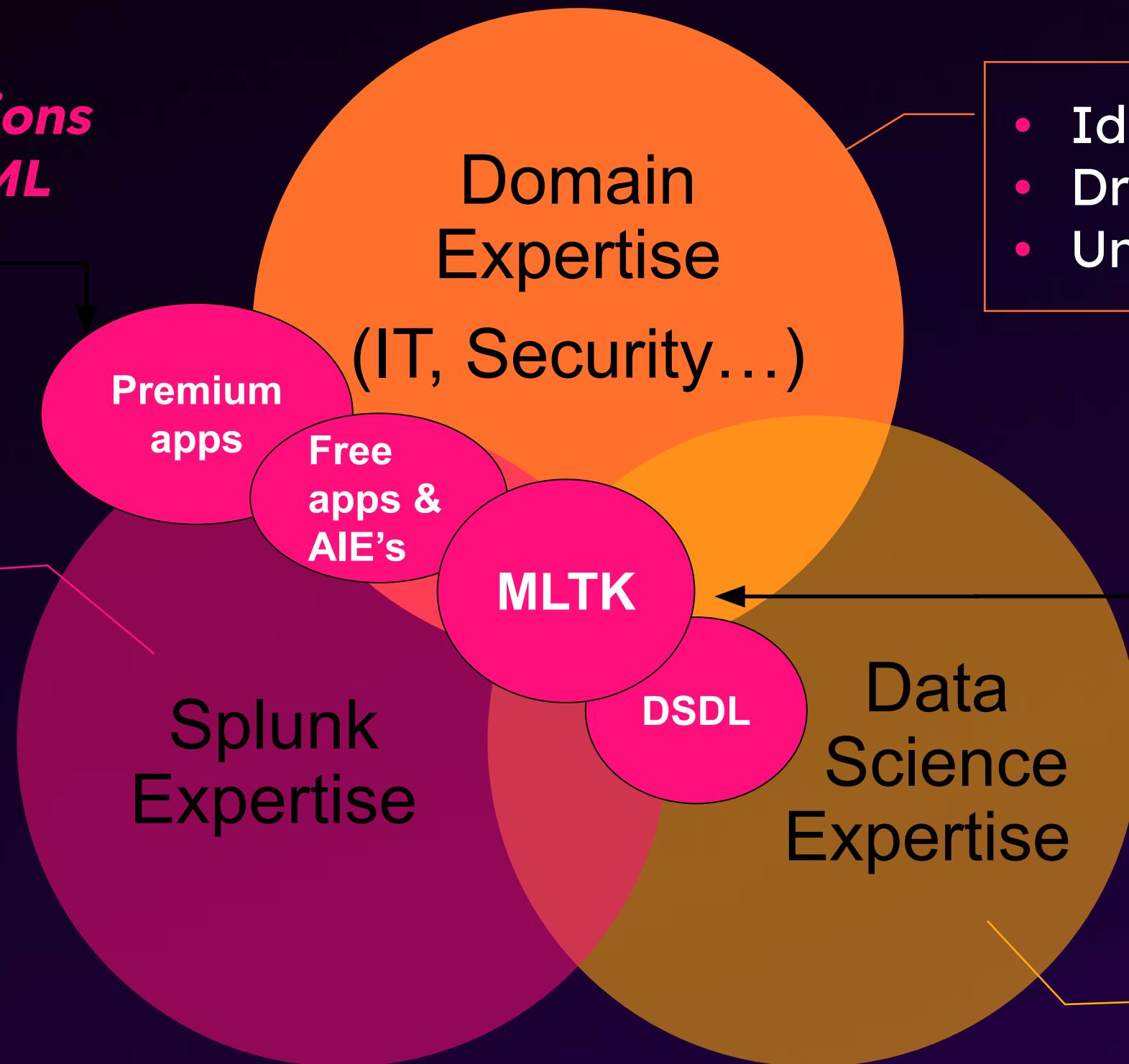
splunk> Platform for Operational Intelligence

Skill Areas for Machine Learning @ Splunk

Premium and AIE solutions provide out of the box ML capabilities.

- Searching
- Reporting
- Alerting
- Workflow

2.



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

1.

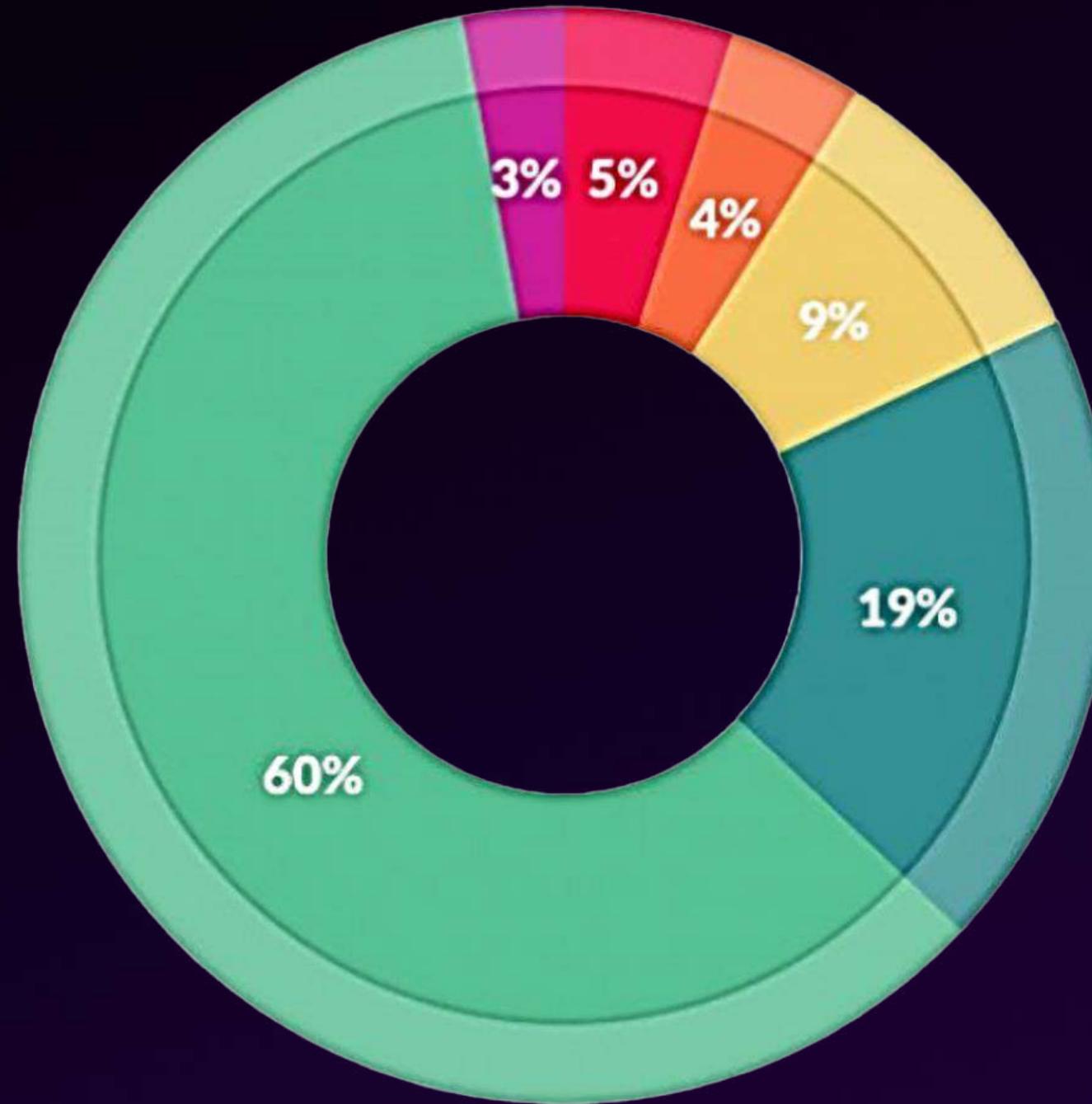
Splunk ML Toolkit facilitates and simplifies via examples & guidance

- Statistics/math background
- Algorithm selection
- Model building

3.

What Data Scientists Really Do

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



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



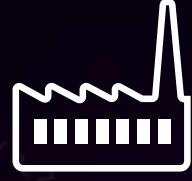
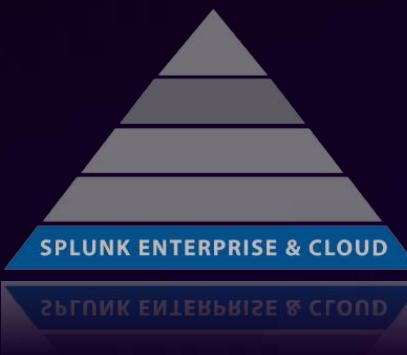
Cleaning Big Data, the Most Time-Consuming, Least Enjoyable Data Science Task

Forbes Survey

March 23

Data Collection

Gain access to previously unused data



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time



Infrastructure data

Application usage and access data

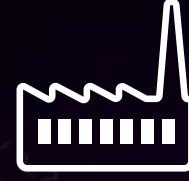
GenAI prompt and output data

Help desk and ticketing system data

IoT and OT data

At Scale

Store, retain, and search data at unprecedented scale



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time



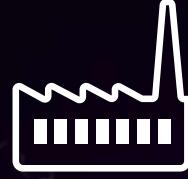
Scalable index and storage

Flexible offering models

Federated analytics

Data understanding

Gain insights to previously unused data



Industrial Assets



Consumer and
Mobile Devices



OT



IT

IT

Real Time

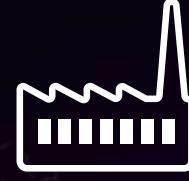
splunk>

Machine data is messy and complex...

```
10.2.1.35 64.66.0.20 - - [17/Jan/2024  
16:21:51] "GET  
/product.screen?product_id=CC-P3-BELKIN-  
SILBLKIPH5&JSESSIONID=SD5SL6FF1ADFF9 HTTP  
1.1" 503 865  
"http://shop.splunktel.com/product.screen?pr  
oduct_id=CC-P3-BELKIN-BLK_BT0OTH_HFREE"  
"Mozilla/5.0 (Linux; Android 12.0.0; FR-fr;  
SM-S901B Build/S908EXXU2BVJA)  
AppleWebKit/537.36 Chrome/114.0.5735.131  
Mobile Safari/537.36" 954
```

Data understanding

Gain insights to previously unused data



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time

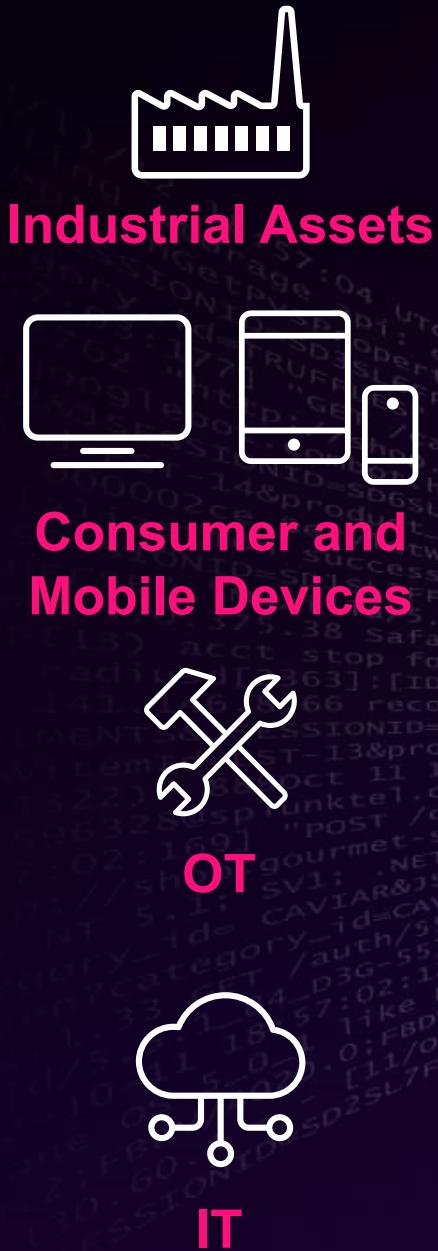
splunk>

Machine data is ~~messy and complex~~ valuable!

```
10.2.1.35 64.66.0.20 - - [17/Jan/2024
16:21:51] "GET
/product.screen?product_id=CC-P3-BELKIN-
SILBLKIPH5&JSESSIONID=SD5SL6FF1ADFF9 HTTP
1.1" 503 865
"http://shop.splunktel.com/product.screen?pr
oduct_id=CC-P3-BELKIN-BLK_BT0OTH_HFREE"
Mozilla/5.0 (Linux; Android 12.0.0; FR-fr;
SM-S901B Build/S908EXXU2BVJA)
AppleWebKit/537.36 Chrome/114.0.5735.131
Mobile Safari/537.36" 954
```

Data understanding

Gain insights to previously unused data



Machine data · User IP · and complex valuable!

```

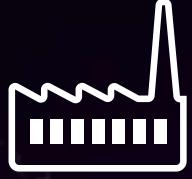
10.2.1.35 64.66.0.20 - - [17/Jan/2024
16:21:51] "GET
/product.screen?product_id=CC-P3-BELKIN-
SILBLKIPH5&JSESSIONID=SD5SL6FF1ADFF9 HTTP
1.1" 503 865
"http://shop.splunktel.com/product.screen?pro
duct_id=CC-P3-BELKIN-BLK_BT0OTH_HFRE
Mozilla/5.0 (Linux; Android 12 9.0; FR-fr;
SM-S901B Build/S908EXXU21A)
AppleWebKit/537.36 Chrome/114.0.5735.131
Mobile Safari/537.36" 954
  
```

The log entry shows a user request for a product screen. Annotations highlight specific fields: "User IP" points to the IP address 64.66.0.20; "Product Viewed" points to the URL parameter "product_id=CC-P3-BELKIN-SILBLKIPH5"; "Preferred Language" points to the user agent string "Mozilla/5.0 (Linux; Android 12 9.0; FR-fr; SM-S901B Build/S908EXXU21A) AppleWebKit/537.36 Chrome/114.0.5735.131 Mobile Safari/537.36"; and "Device" points to the user agent string itself.

Data Access

Gain access to previously unused data

```
index="oidemo" sourcetype="access_combined"
```



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time



splunk>

Data Preparation

Filter data with SPL



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time

splunk>

index="oidemo" sourcetype="access_combined" action=purchase

action

5 Values, 44.612% of events

Selected Yes No

Reports

Top values Top values by time Rare values

Events with this field

Values	Count	%
purchase	2,243	20.709%
addtocart	2,172	20.054%
view	2,168	20.017%
remove	2,161	19.952%
changequantity	2,087	19.269%

action

1 Value, 100% of events

Selected Yes No

Reports

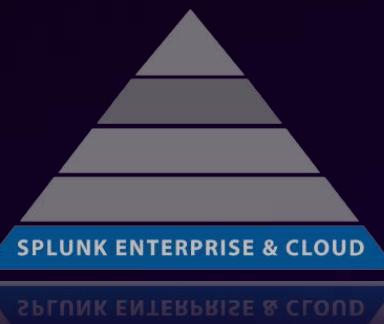
Top values Top values by time Rare values

Events with this field

Values	Count	%
purchase	2,249	100%

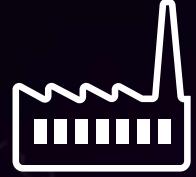
Data Preparation

Transform data with SPL



Filter

```
index sourcetype="access_combined" action=purchase  
| iplocation clientip
```



Industrial Assets



Consumer and
Mobile Devices



OT



IT

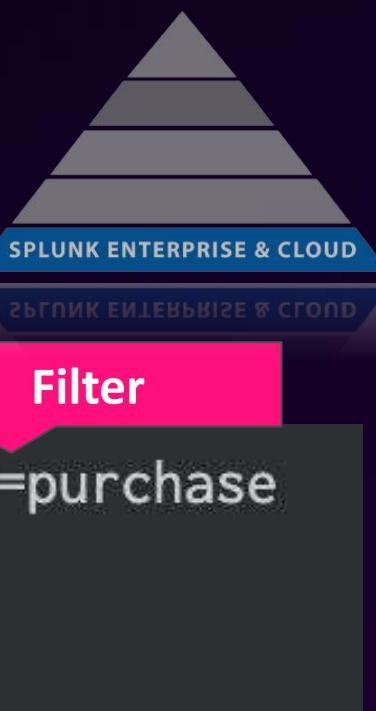
Real Time

splunk>

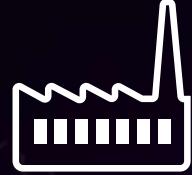
clientip	Country	City
0.134.99.193		
0.134.99.193		
0.134.99.193		
0.152.133.194		
0.152.133.194		
0.152.133.194		
1.12.191.128		
1.16.209.124		

clientip	Country	City
0.134.99.193		
0.134.99.193		
0.134.99.193		
0.152.133.194		
0.152.133.194		
0.152.133.194		
1.12.191.128	China	Haidian (Haidian Qu)
1.16.209.124	Australia	South Brisbane
1.16.209.124	Australia	South Brisbane

Data Preparation



Clean data with SPL



Industrial Assets



Consumer and
Mobile Devices



OT



IT

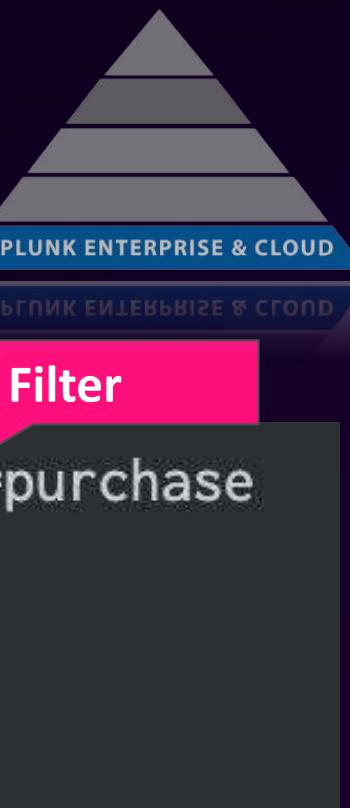
Real Time

splunk®

```
index sourcetype="access_combined" action=purchase  
| iplocation clientip  
| fillnull value="Unknown" Country
```

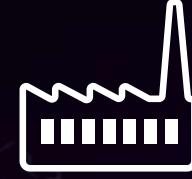
clientip	Country	City
0.134.99.193		
0.134.99.193		
0.134.99.193		
0.152.133.194		
0.152.133.194		
0.152.133.194		
1.12.191.128	China	Haidian (Haidian Qu)
1.16.209.124	Australia	South Brisbane
1.16.209.124	Australia	South Brisbane

clientip	Country	City
0.134.99.193	unknown	unknown
0.134.99.193	unknown	unknown
0.134.99.193	unknown	unknown
0.152.133.194	unknown	unknown
0.152.133.194	unknown	unknown
0.152.133.194	unknown	unknown
1.12.191.128	China	Haidian (Haidian Qu)
1.16.209.124	Australia	South Brisbane



Data Preparation

Aggregate data with SPL



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time

splunk>

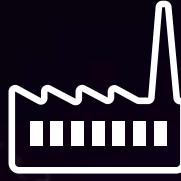
```
index sourcetype="access_combined" action=purchase  
| iplocation clientip  
| fillnull value="Unknown" Country  
| stats count by product device Country
```

Aggregate

product	device	Country	count
Bubble_Wrap	iPhone	United States	11
Man_Candle-Bacon	iPhone	United States	10
Canned_Unicorn_Meat	Windows	United States	9

Training ML model

Fit model



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time

splunk>

```
index sourcetype="access_combined" action=purchase
| iplocation clientip
| fillnull value="Unknown" Country
| stats count by product device Country
```

Filter

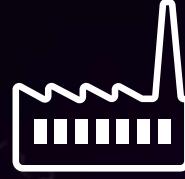
| fit algorithm from clientip into "model"

Aggregate

Train "model"

Productionalize ML model

Apply model



Industrial Assets



Consumer and
Mobile Devices



OT



IT

Real Time

splunk>

```
index sourcetype="access_combined" action=purchase
| iplocation clientip
| fillnull value="Unknown" Country
| stats count by product device Country
```

Filter

Transform

Clean

Aggregate

| fit algorithm from clientip into "model"

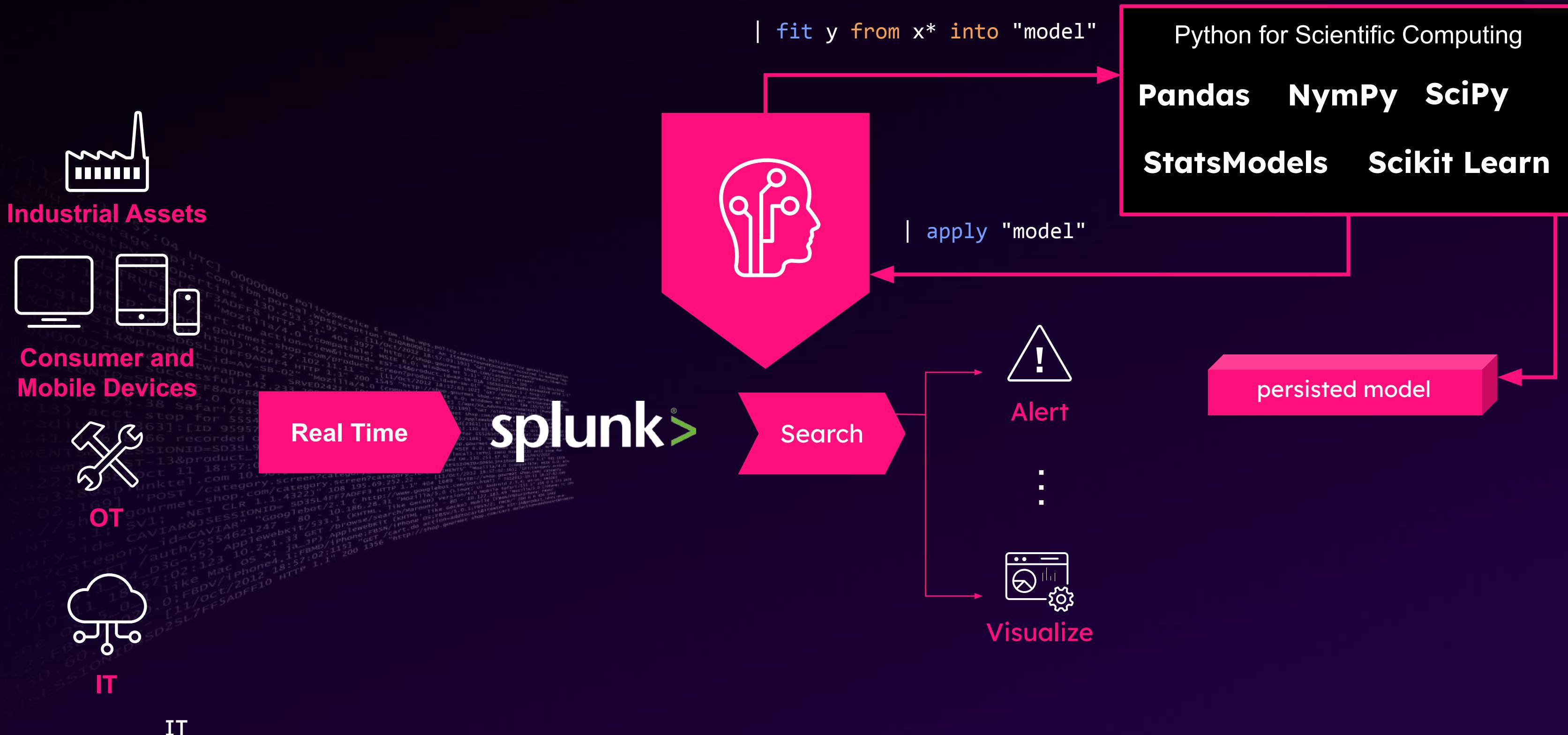
Train "model"

| apply "model"

Run "model"

© 2024 SPLUNK INC.

Model training and application



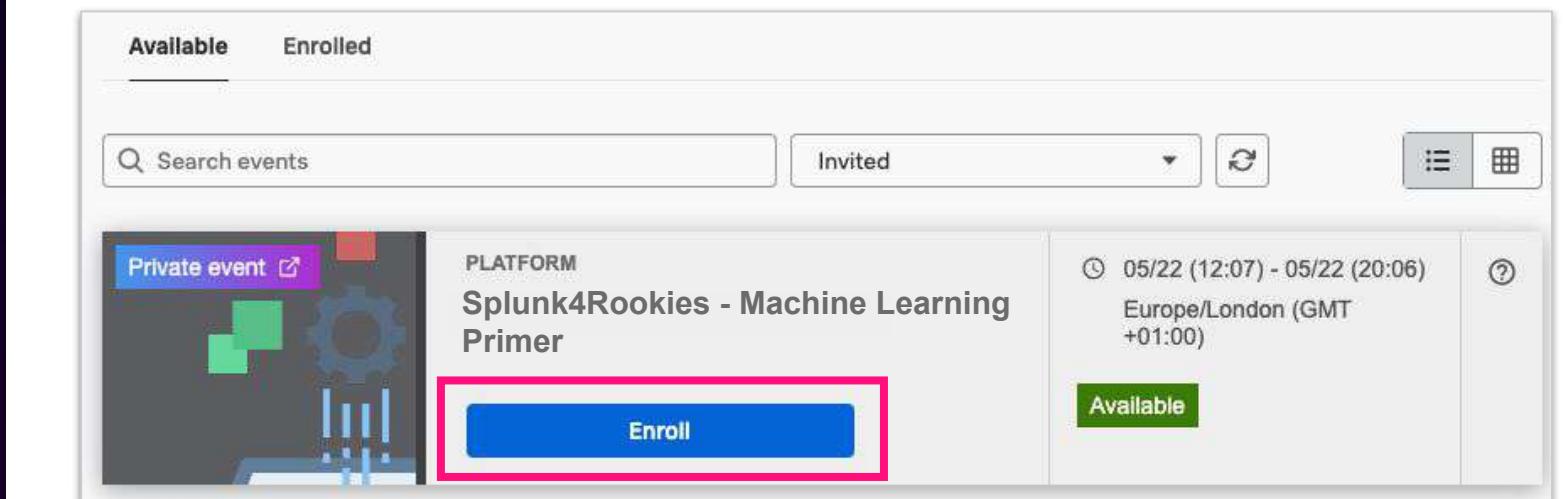


Enroll in Today's Workshop

Tasks

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

Goal

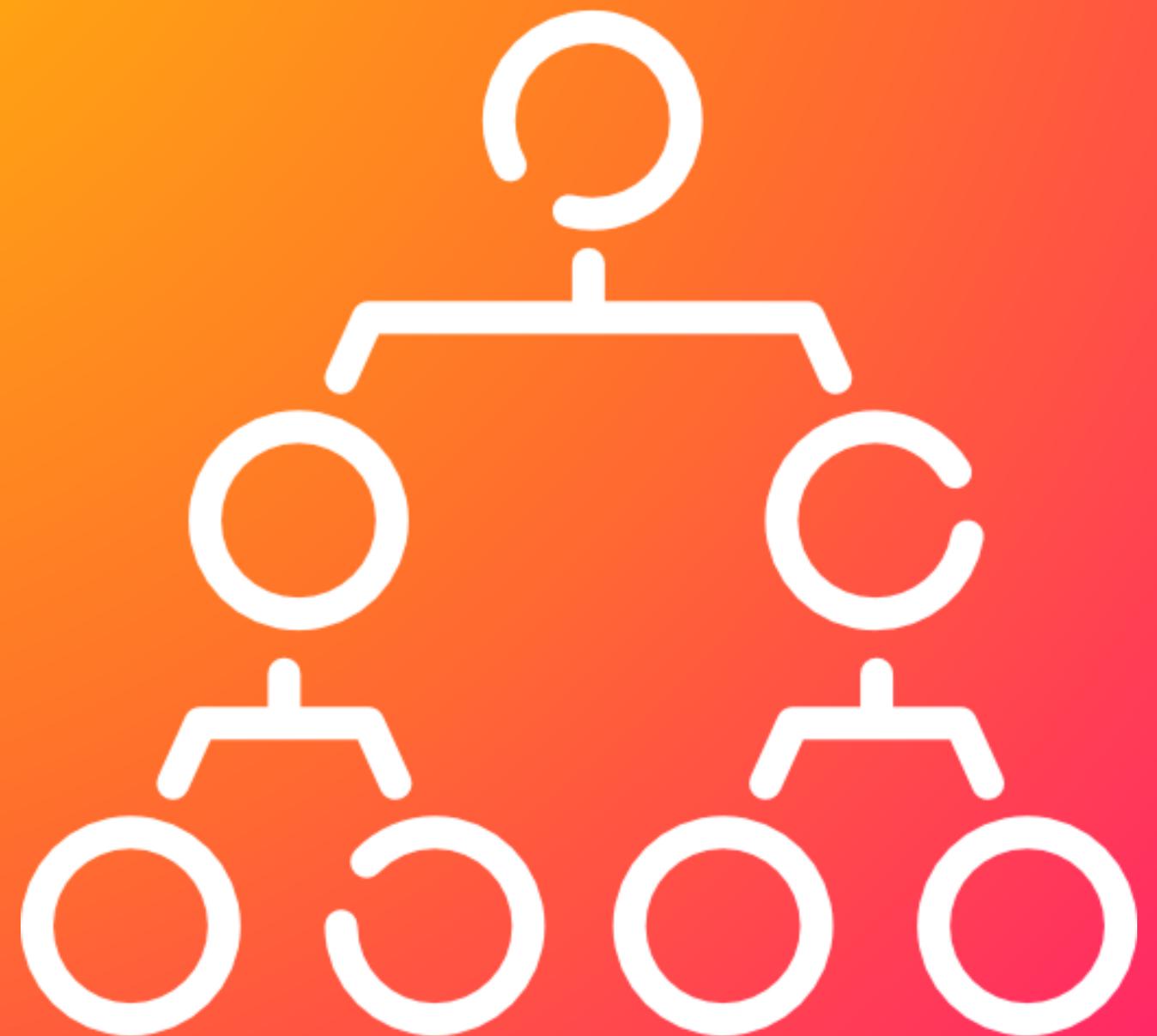


Enroll in today's event

Use Case: Predict Disk Utilization

Categorical Prediction

“A method that lets you quickly, easily, and sustainably gain insight into your data by predicting its categorical features”



How Does Categorical Prediction Work?

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCP Digital	17	btcup_checkout	375	True positive
BTCP Digital	3	payment_svc	89	True positive

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCP Digital	17	btcup_checkout	375	True positive
BTCP Digital	3	payment_svc	89	True positive

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTcup Digital	17	btcup_checkout	375	True positive
BTcup Digital	3	payment_svc	89	True positive

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCP Digital	17	btcup_checkout	375	True positive
BTCP Digital	3	payment_svc	89	True positive

alert group	count	src	alert value	Predicate (predicted)
T-Shirt Co.	17	checkoutsvc	100	False positive
O11y Cloud	16	browsercheck	89	False positive

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCP Digital	17	btcup_checkout	375	True positive
BTCP Digital	3	payment_svc	89	True positive

alert group	count	src	alert value	Predicate (predicted)
T-Shirt Co.	17	checkoutsvc	100	False positive
O11y Cloud	16	browsercheck	89	True positive?

How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL

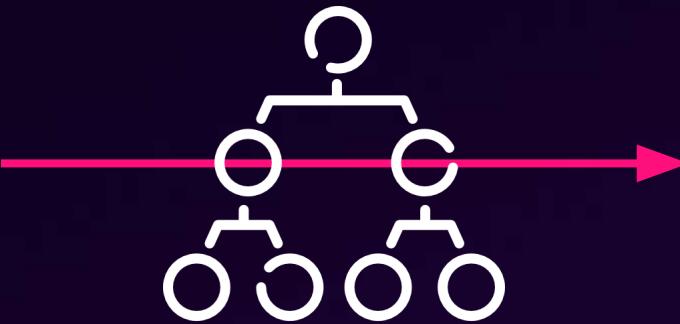
How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL

New Data

Feature 1	...	Feature n	Severity (predicted)
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL



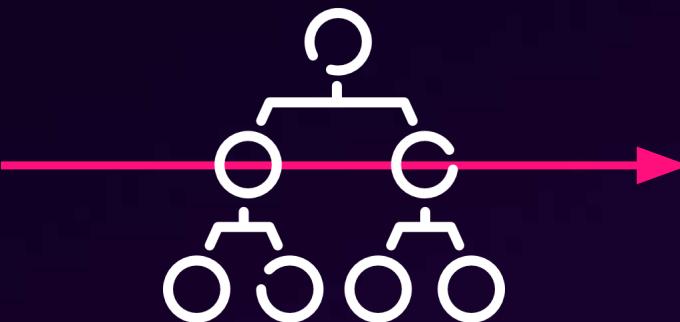
How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL

New Data

Feature 1	...	Feature n	Severity (predicted)	Severity (actual)
[data]	...	[data]	CRITICAL	MAJOR
[data]	...	[data]	CRITICAL	LOW
[data]	...	[data]	CRITICAL	MINOR
[data]	...	[data]	CRITICAL	CRITICAL
[data]	...	[data]	CRITICAL	LOW
[data]	...	[data]	CRITICAL	MINOR
[data]	...	[data]	CRITICAL	MAJOR
[data]	...	[data]	CRITICAL	MAJOR
[data]	...	[data]	CRITICAL	MINOR



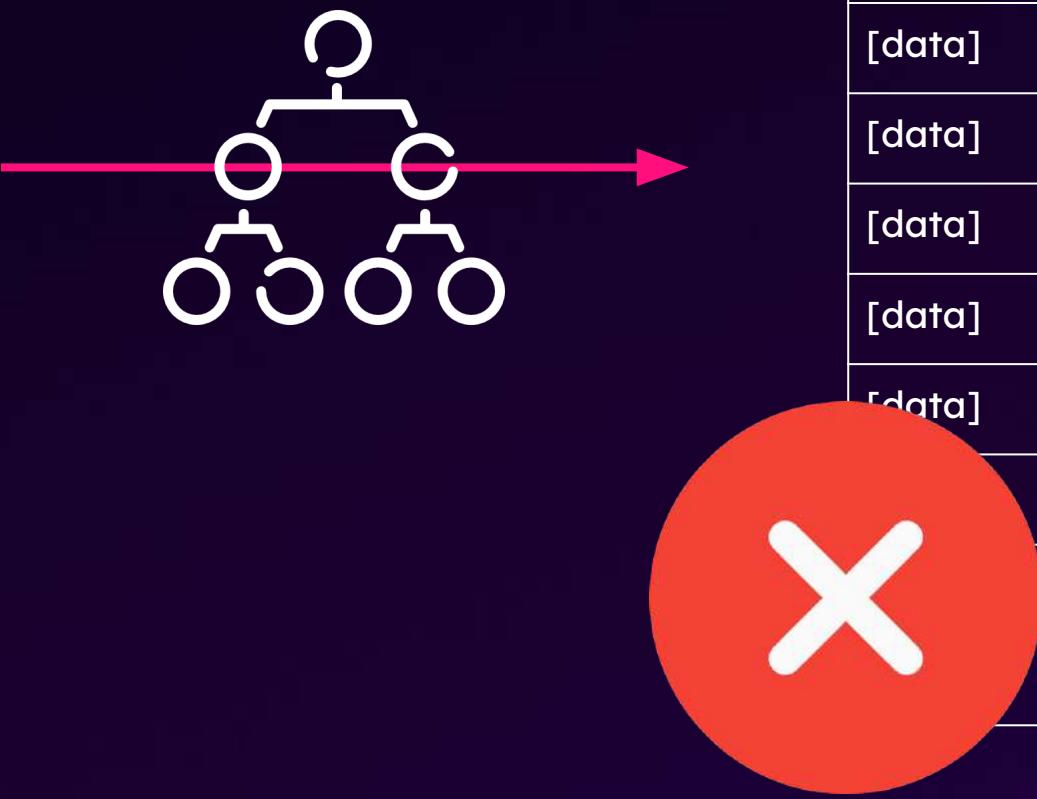
How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL

New Data

Feature 1	...	Feature n	Severity (predicted)	Severity (actual)
[data]	...	[data]	CRITICAL	MAJOR
[data]	...	[data]	CRITICAL	LOW
[data]	...	[data]	CRITICAL	MINOR
[data]	...	[data]	CRITICAL	CRITICAL
[data]	...	[data]	CRITICAL	LOW
[data]	...	[data]	CRITICAL	MINOR
...	...	[data]	CRITICAL	MAJOR
...	...	[data]	CRITICAL	MAJOR
...	...	[data]	CRITICAL	MINOR



How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	MAJOR
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR

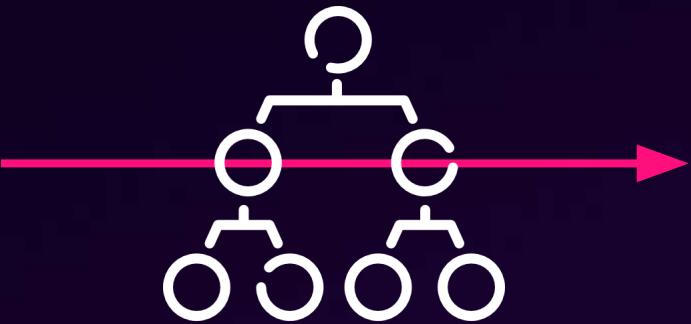
How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	MAJOR
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR

New Data

Feature 1	...	Feature n	Severity (predicted)
[data]	...	[data]	LOW
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	LOW



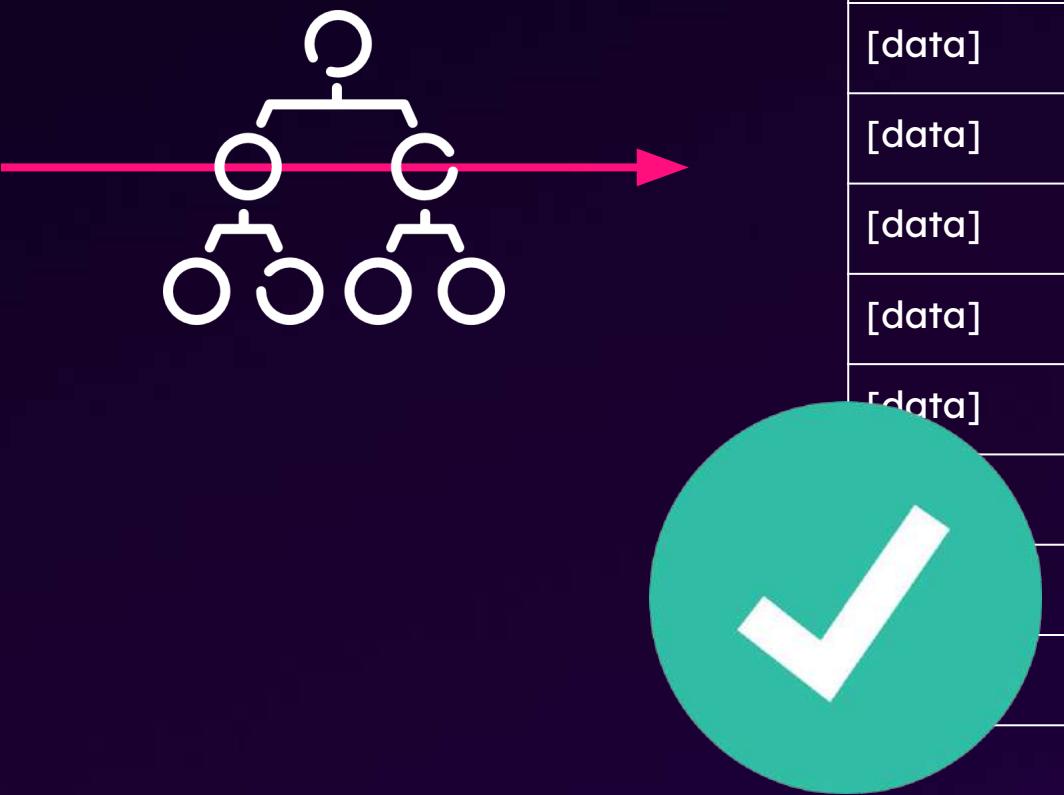
How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	MAJOR
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR

New Data

Feature 1	...	Feature n	Severity (predicted)
[data]	...	[data]	LOW
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
...	[data]	...	LOW
...	[data]	...	MINOR
...	[data]	...	LOW



Summary

Top 4 most important things to remember about categorical prediction

1

Predicting categorical fields is done using a **supervised learning** method which uses **labeled data**

2

Models **assumes there exists a pattern** determining existing categories

3

Scaling data prior to training is **often necessary**

4

Choice of categorical prediction algorithm(s) may rely on a **subject matter expert** of the data

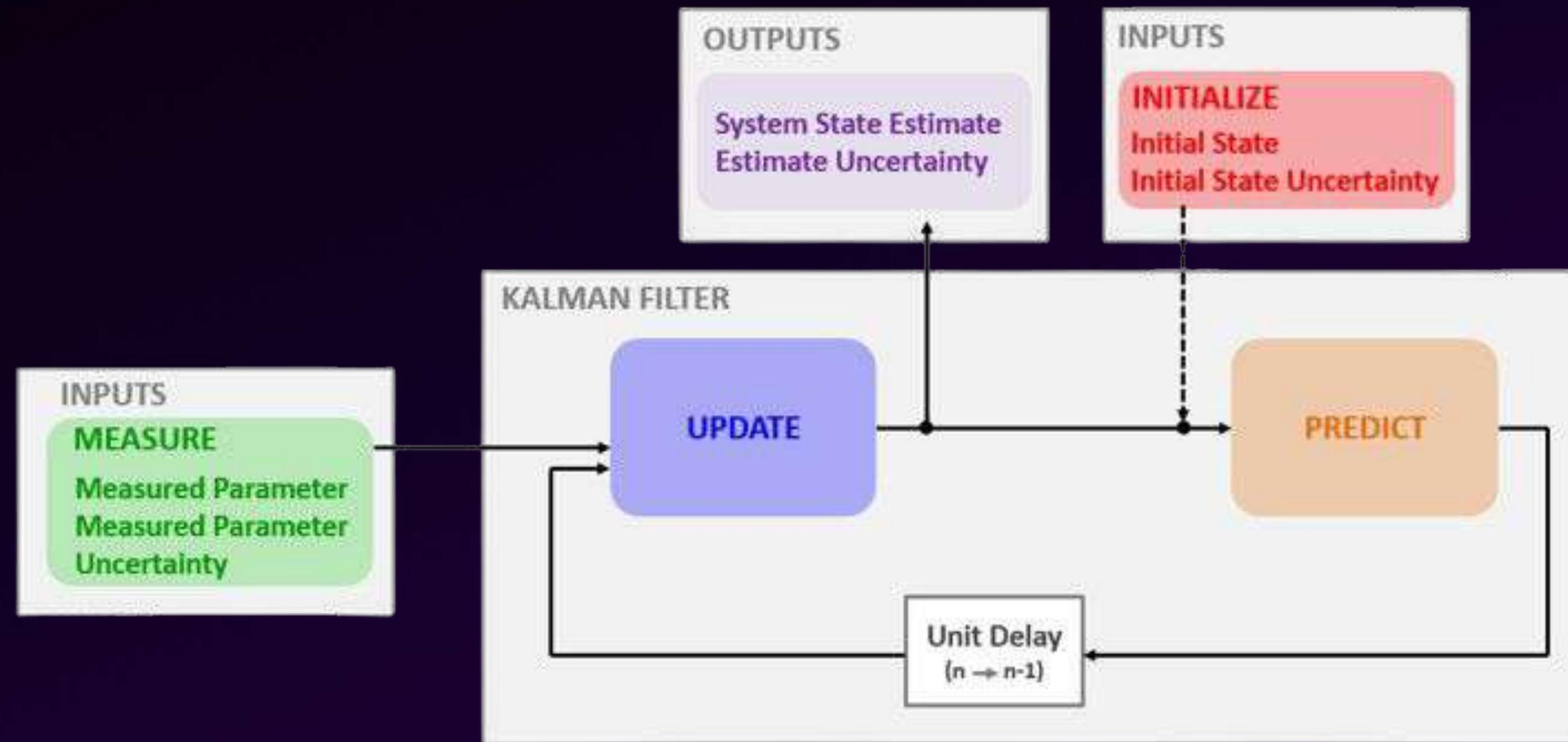
Use Case: Forecast App Logons with Special Days

Forecasting Time Series

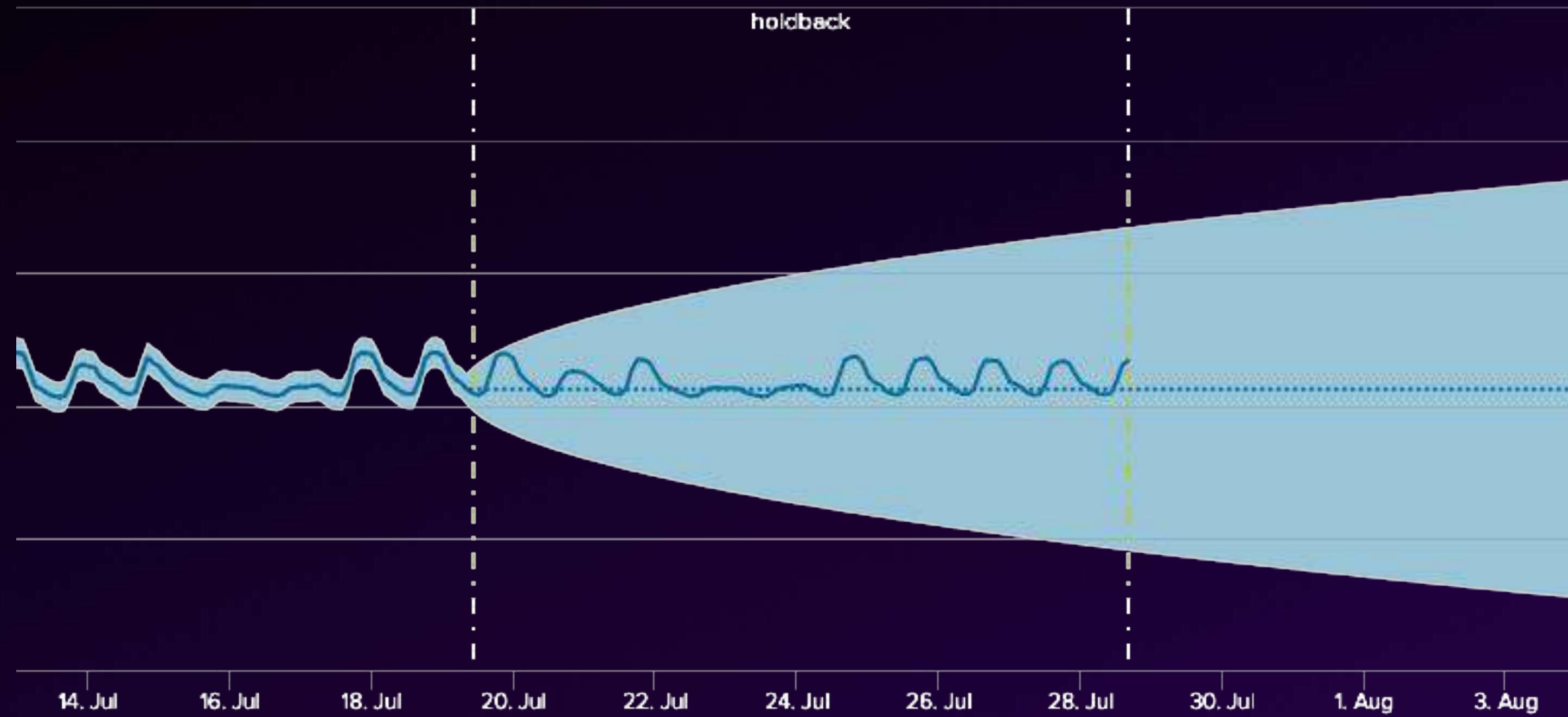
“Using historical data to identify patterns, which are then used to forecast how your data might behave in the future”



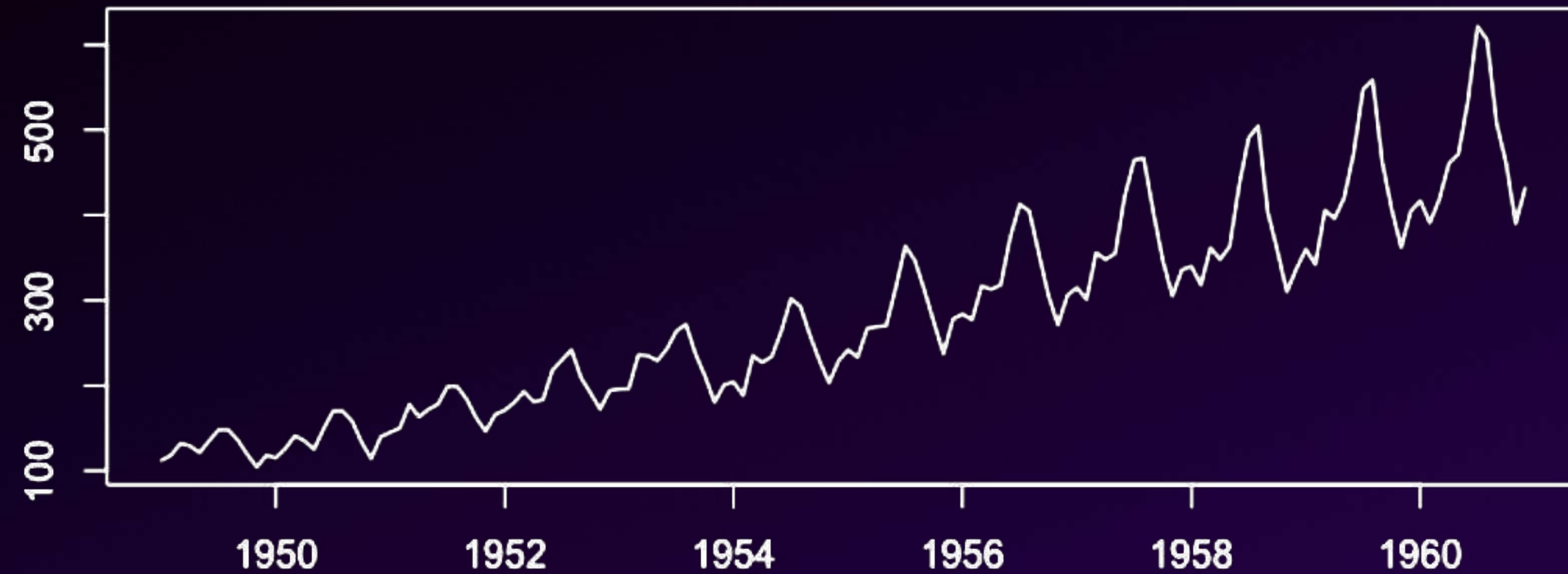
The Kalman Filter



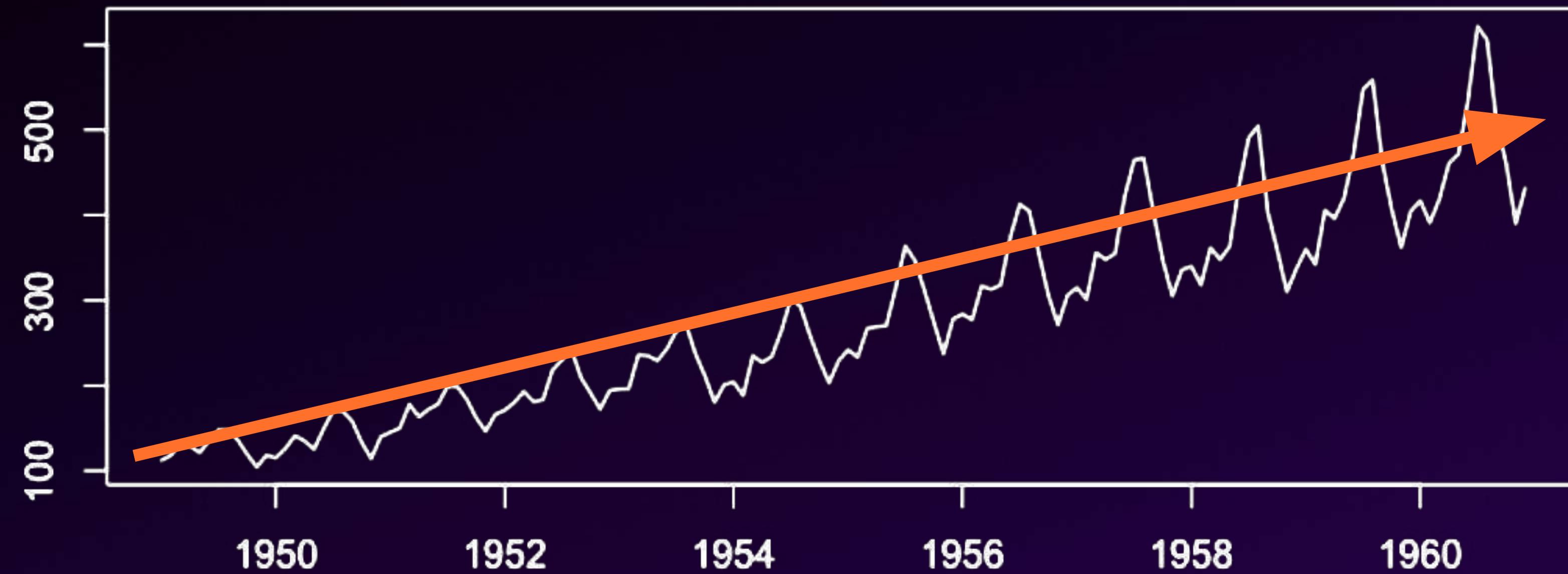
The Kalman Filter



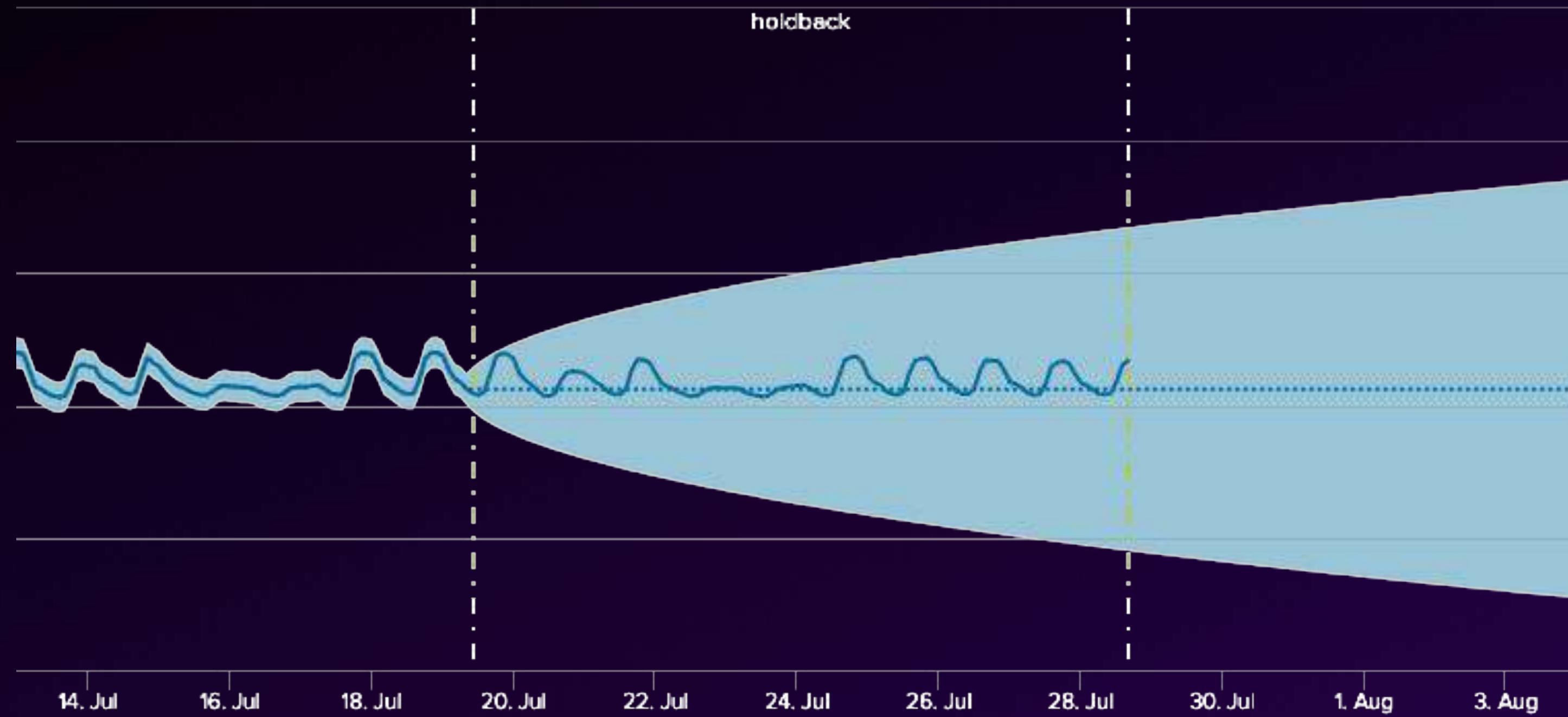
The Kalman Filter



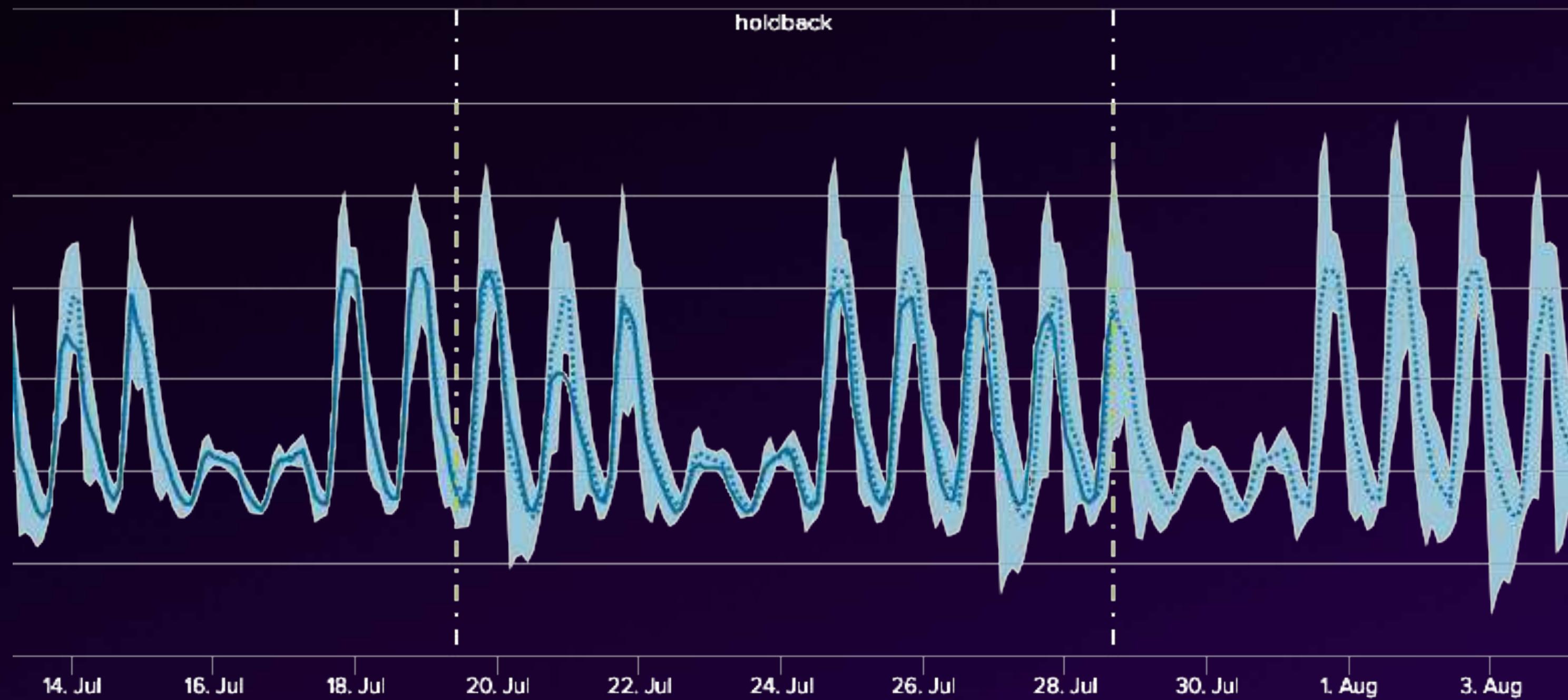
The Kalman Filter



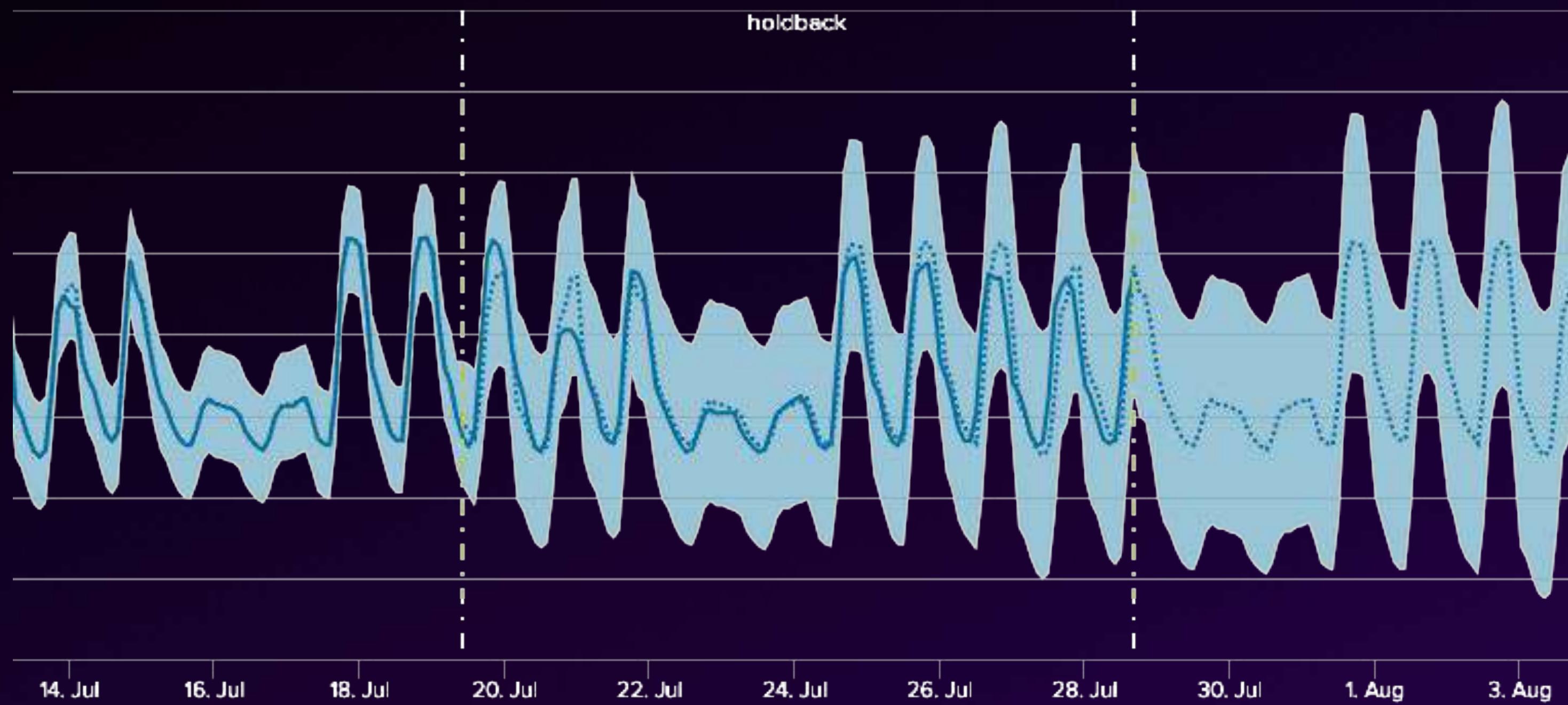
The Kalman Filter



The Kalman Filter



The Kalman Filter



Summary

Top 4 most important things to remember about forecasting time series

1

Forecasting time series
is done using a
supervised learning
method

2

Models **assume historic
data as a baseline**,
and will self-correct
accordingly

3

Parameters have a
**large impact on
performance**. Tuning
each model is highly
recommended

4

Choice of forecasting
algorithm may rely on a
subject matter expert
of the data

Resource Forecasting

Use Case: Detect Outliers in Server Response Time



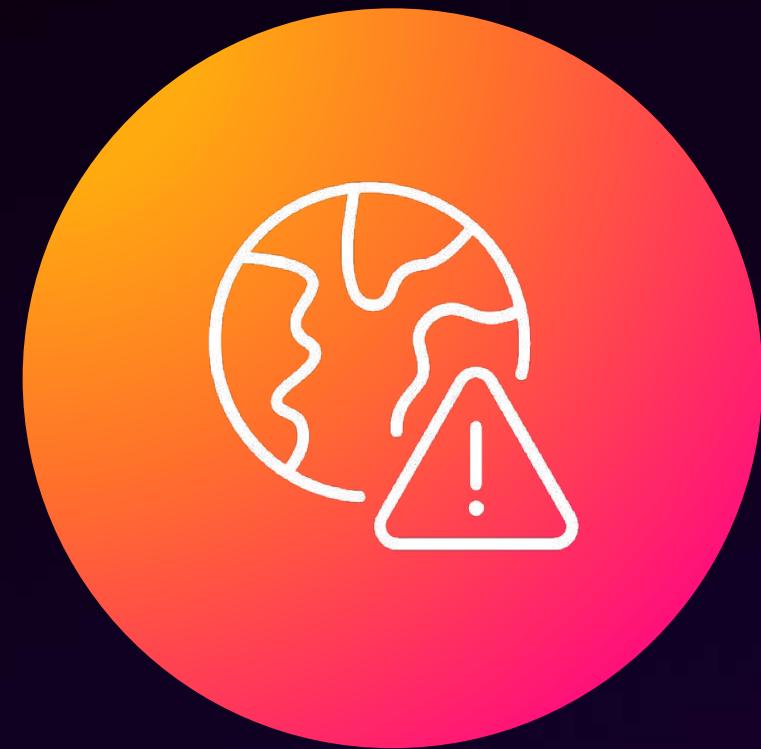
Outlier Detection Algorithms

Identify and analyze abnormal behavior in your data

Includes:

- Clustering
- Outlier Detection

Global



Data points different from
expected pattern, range, or norm

Contextual



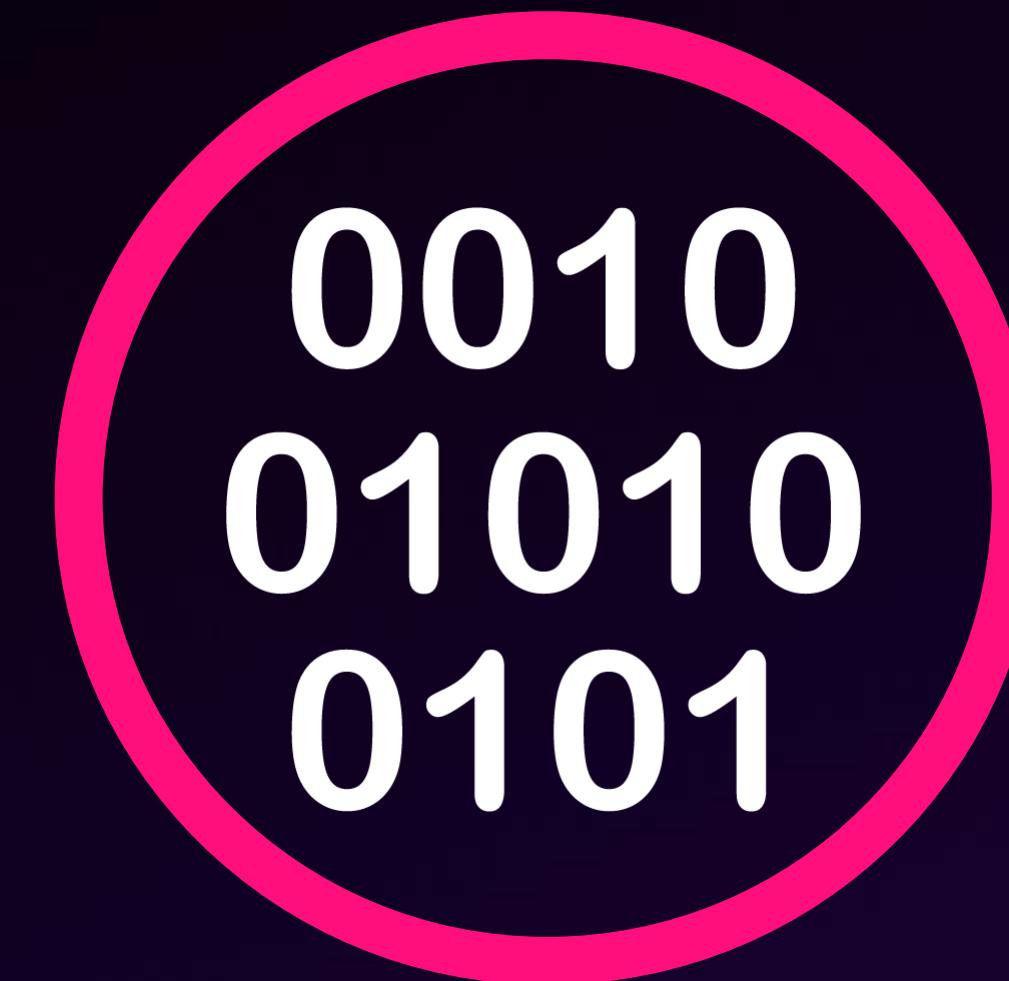
Are the results out
of context?

Collective



Looks normal with isolation
but stands out in a group

Numeric



Categorical



Included Algorithms

DensityFunction

| LocalOutlierFactor

| MultiVariateOutlierDetection

| One-Class SVM

Categorical data to Numeric data?

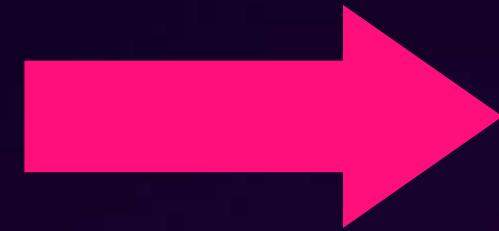
device
server01
server02
server03



0010
01010
0101

Categorical data to Numeric data?

device
server01
server02
server03



device	server01	server02	server03
server01	1	0	0
server02	0	1	0
server03	0	0	1

Summary

Top 4 most important things to remember about outlier detection

1



Outlier detection is a way of analyzing your data for **historical baseline outliers**

2



Models **assume historic data input represents normal data**

3



Encoding is necessary for categorical outlier detection

4



Choice of outlier algorithm may rely on a **subject matter expert** of the data

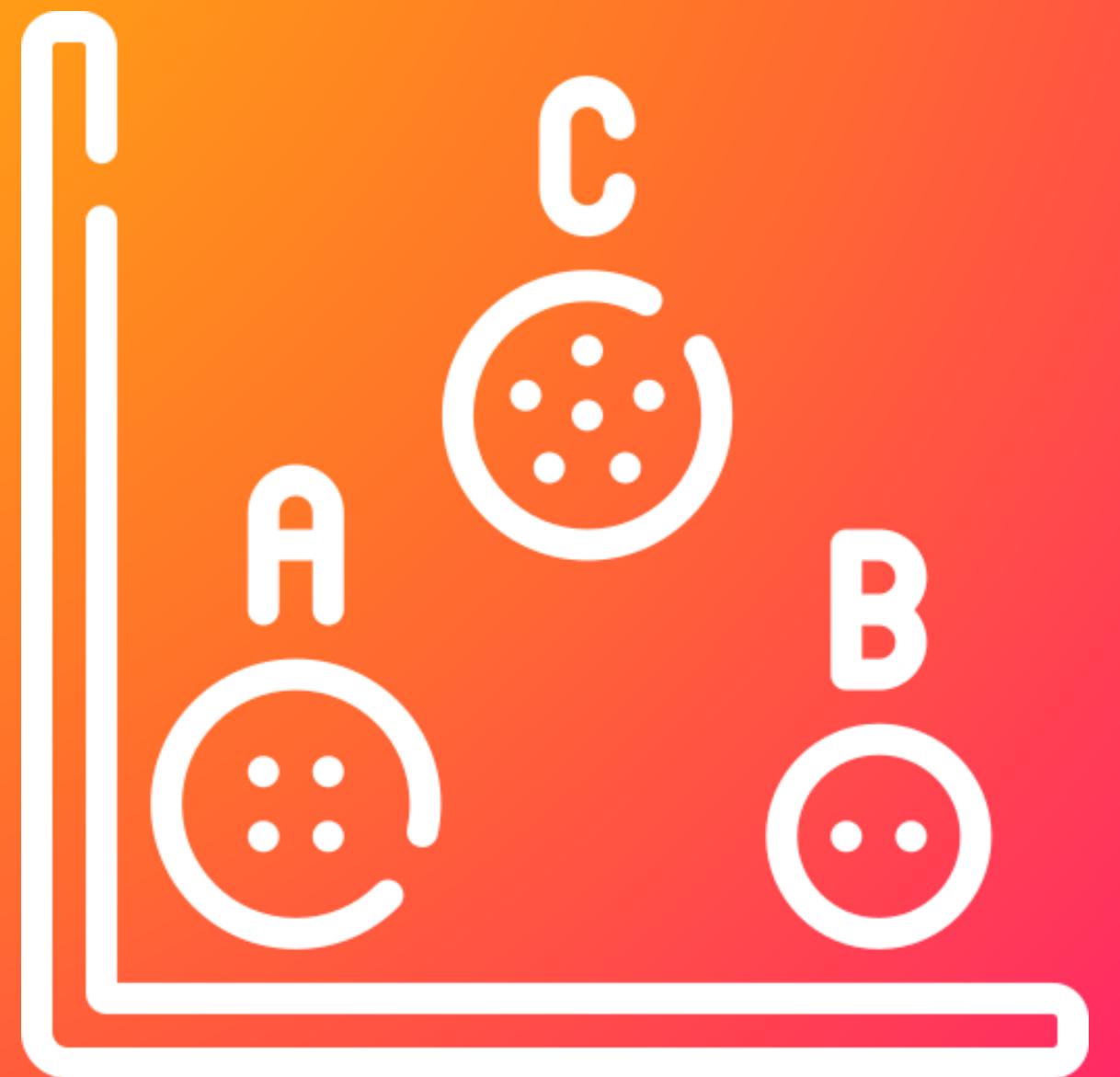
Predictive Analytics

Use Case: Operationalize ML in ITSI

Use Case: [ADD MODULE / USE CASE NAME]

Clustering

“A method that organizes a set of numeric data points in a way that objects in the same cluster are more similar to each other than those in other clusters”



How Does Numeric Clustering Work?

General intuition

Starting with one-dimensional data, unlabeled numeric values plotted on a line



Organizing Exercise: Using your human intuition, form 3 groups with the dataset above

How Does Numeric Clustering Work?

General intuition

Most likely, these are the 3 groups (or clusters) you formed:



Now let's see how a machine can replicate this kind of grouping intuition!

How Does Numeric Clustering Work?

General intuition

First, estimate the center of a cluster randomly, denoted by the star



So, our clusters look like this:



How Does Numeric Clustering Work?

General intuition

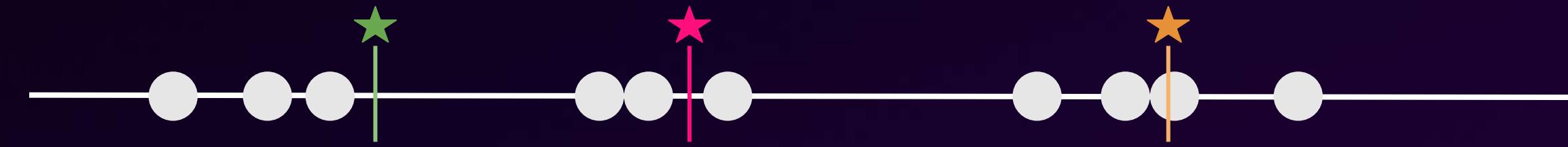
Now the **mean** value of each cluster is calculated. This is the “mean” in “k-means”!



How Does Numeric Clustering Work?

General intuition

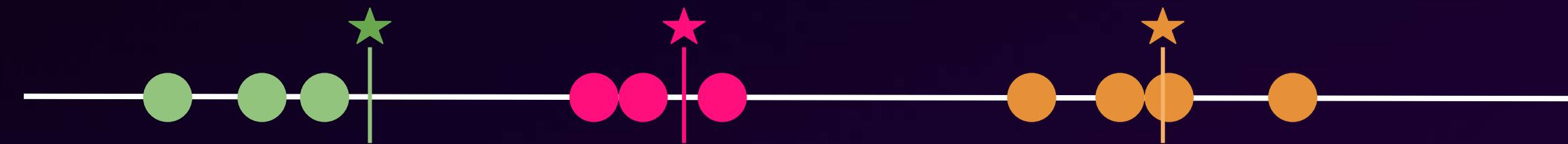
The mean of each cluster now become the new centroids:



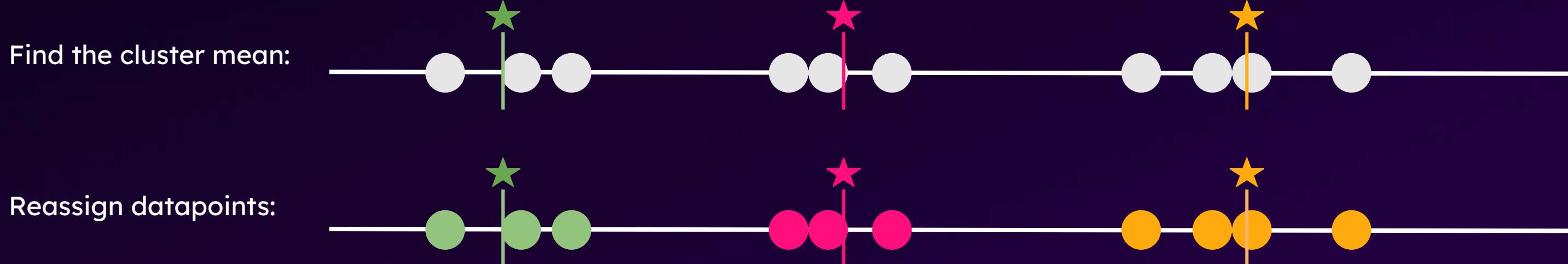
How Does Numeric Clustering Work?

General intuition

Now the data points are reassigned to the nearest k-means cluster:



The k-means and clustering steps are repeated until the data points no longer change to different clusters



How Does Numeric Clustering Work?

General intuition

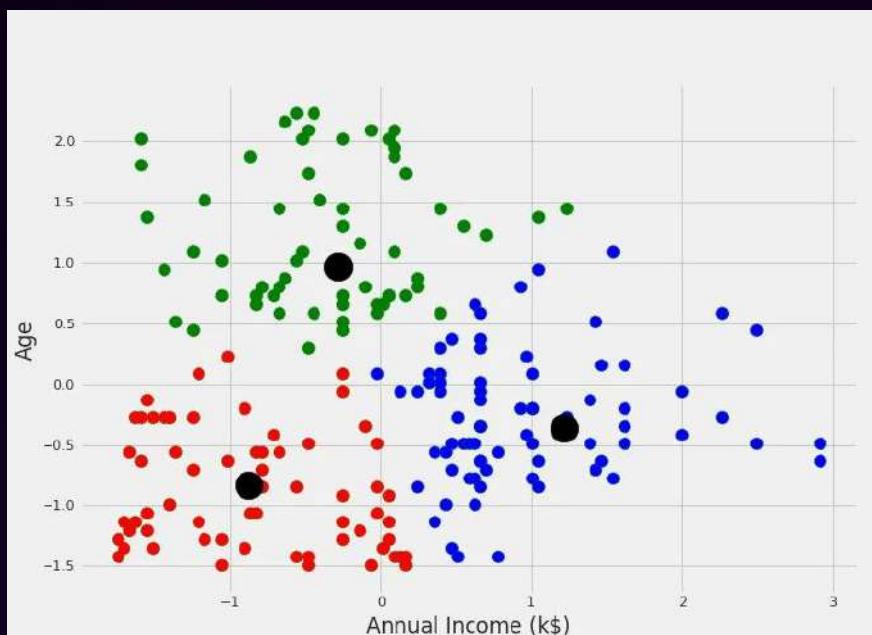
These are our final 3 clusters, found by k-means clustering:



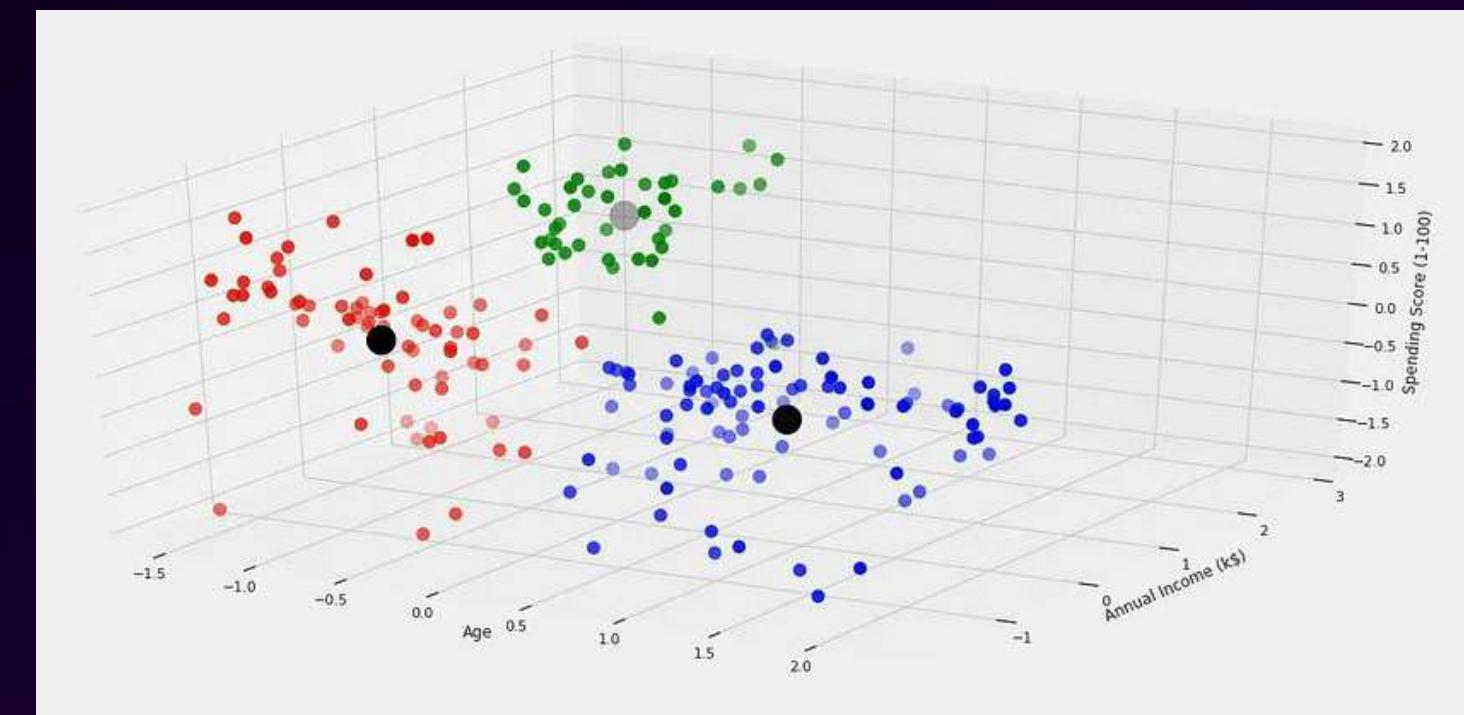
How Does Numeric Clustering Work?

General intuition

k-means clustering applies to multidimensional data as well:



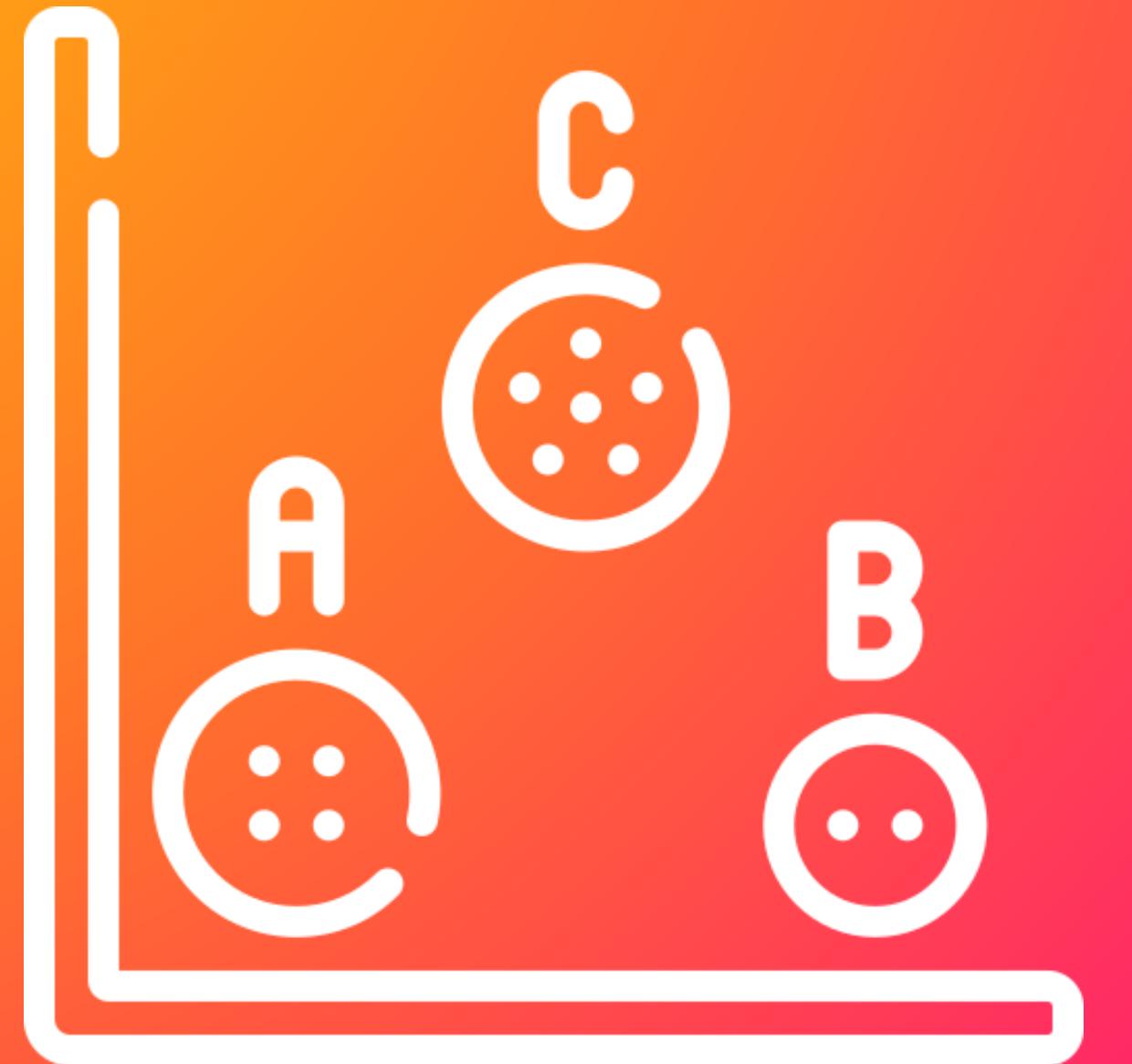
2 dimensional data



3 dimensional data

If you can see
in more than 3
dimensions...
give us a call

4+ dimensional data



Clustering

“A method that organizes a set of numeric data points in a way that objects in the same cluster are more similar to each other than those in other clusters”

Live Instance Demo

Log Into [INSTANCE URL]

Lab Guide Exercise #5

Time: 10 minutes

Summary

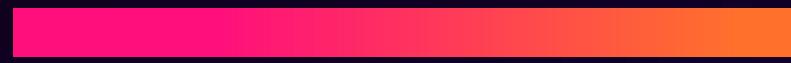
Top 4 most important things to remember about clustering algorithms

1



Clustering is an **unsupervised learning** method which uses **unlabeled data**

2



User must decide whether **dimensionality reduction** is necessary

3



User must decide whether to **scale** the data prior to clustering

4



Choice of clustering algorithm(s) may rely on a **subject matter expert** of the data

Alert Storm Detection

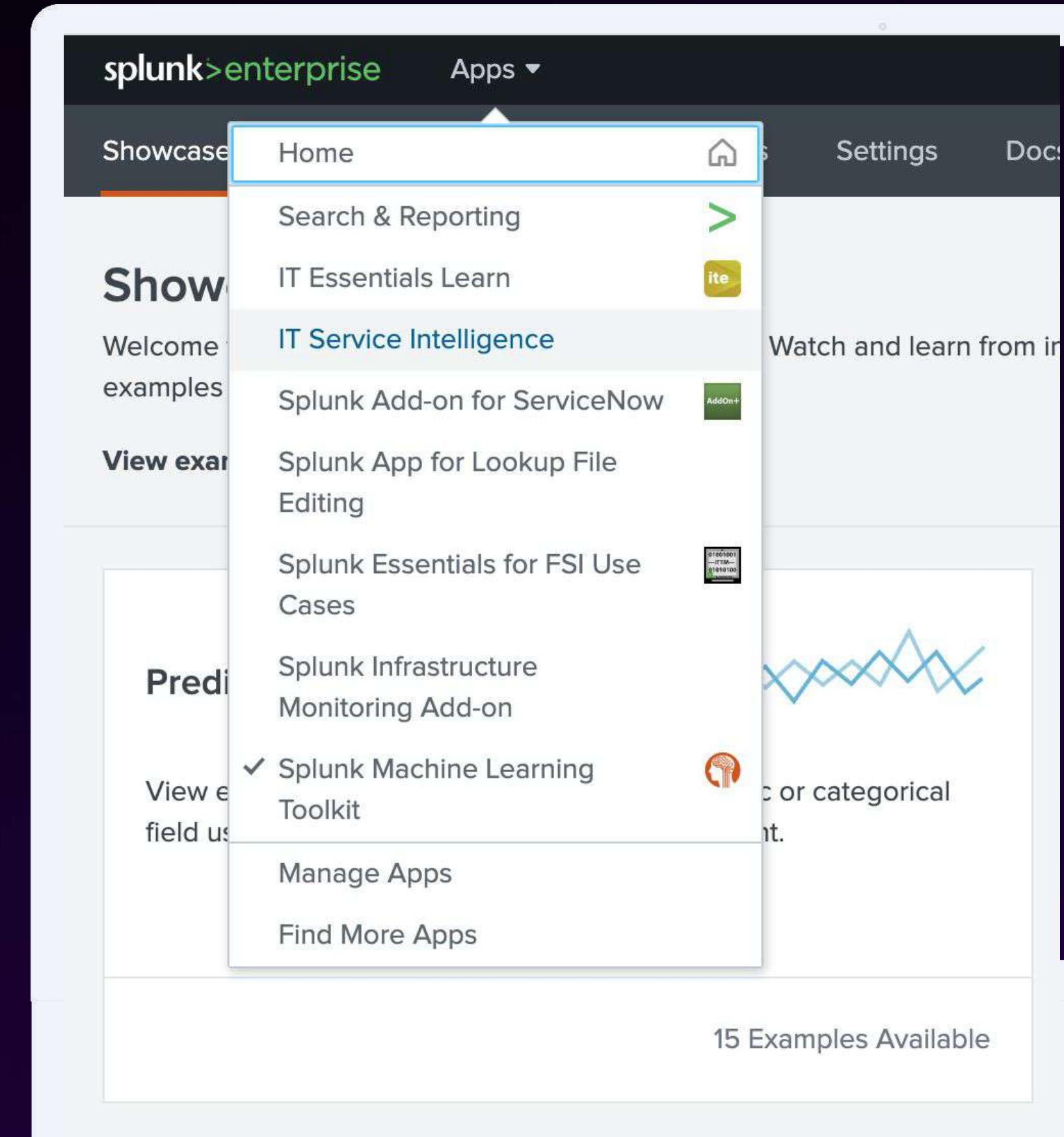
**Now that we have some
understanding of the underlying
technology,
let's see how the integration of ML
was done in Splunk IT Service
Intelligence**

**IT Service
Intelligence
Showtime!**

IT Service Intelligence

Is a premium application installed on the same Splunk instance as the MLTK.

We will simply change the application context and put on our ITOps monitoring analyst hat!



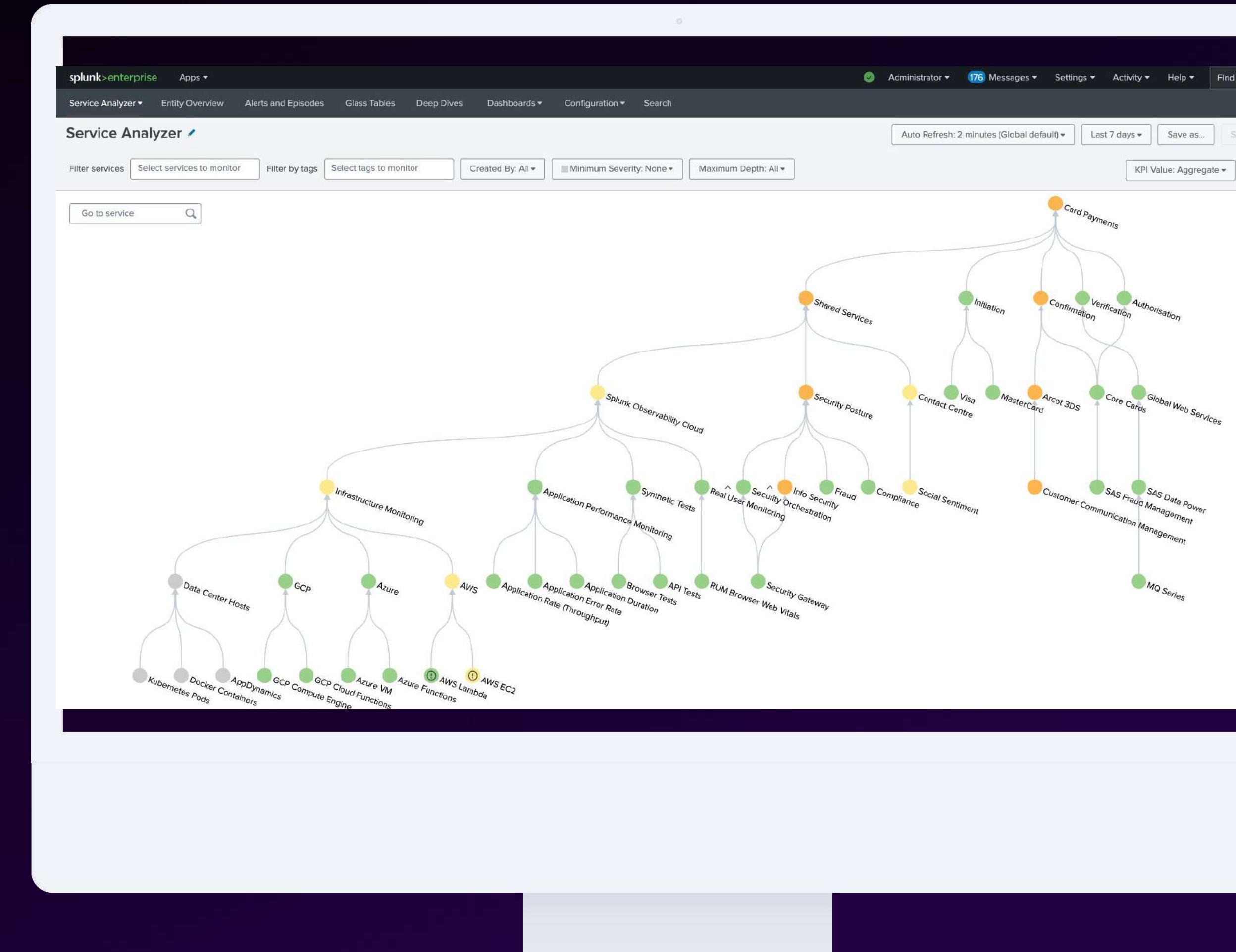
The Service Analyzer

This is our landing page.

We will not go through every ITSI nooks & crannies. This would deserve a distinct workshop which is out of scope of today's session.

Instead, we are going to dive straight to the places where ML use cases have been integrated (there are several).

This will be a guided tour.



**Plot twist:
The MLTK is hidden under the hood
and sandboxed into the ITSI GUI
(a.k.a point & click ML)**



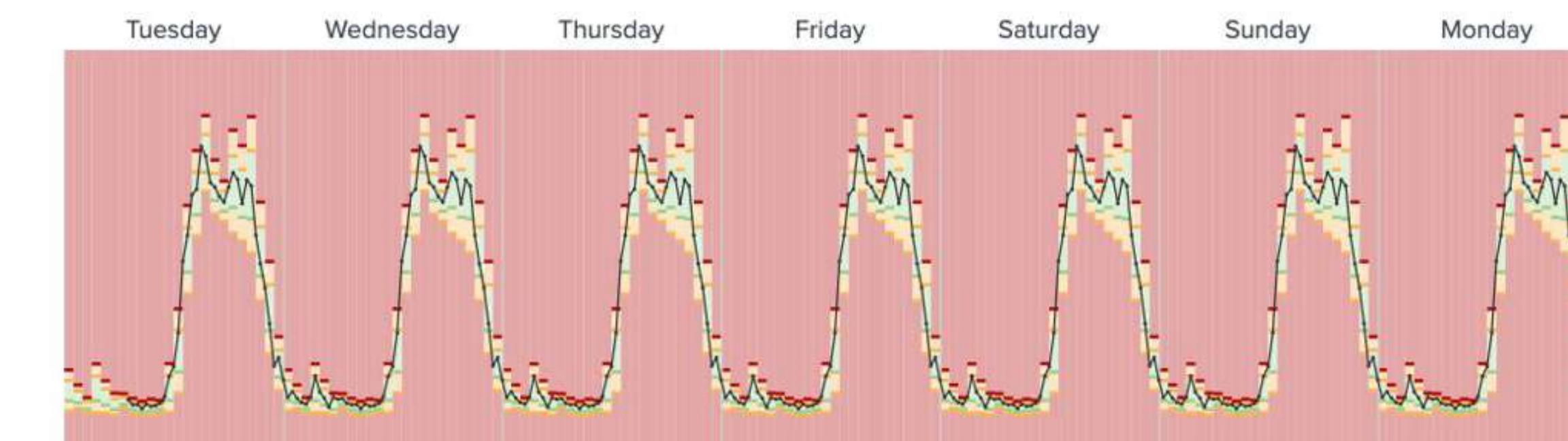
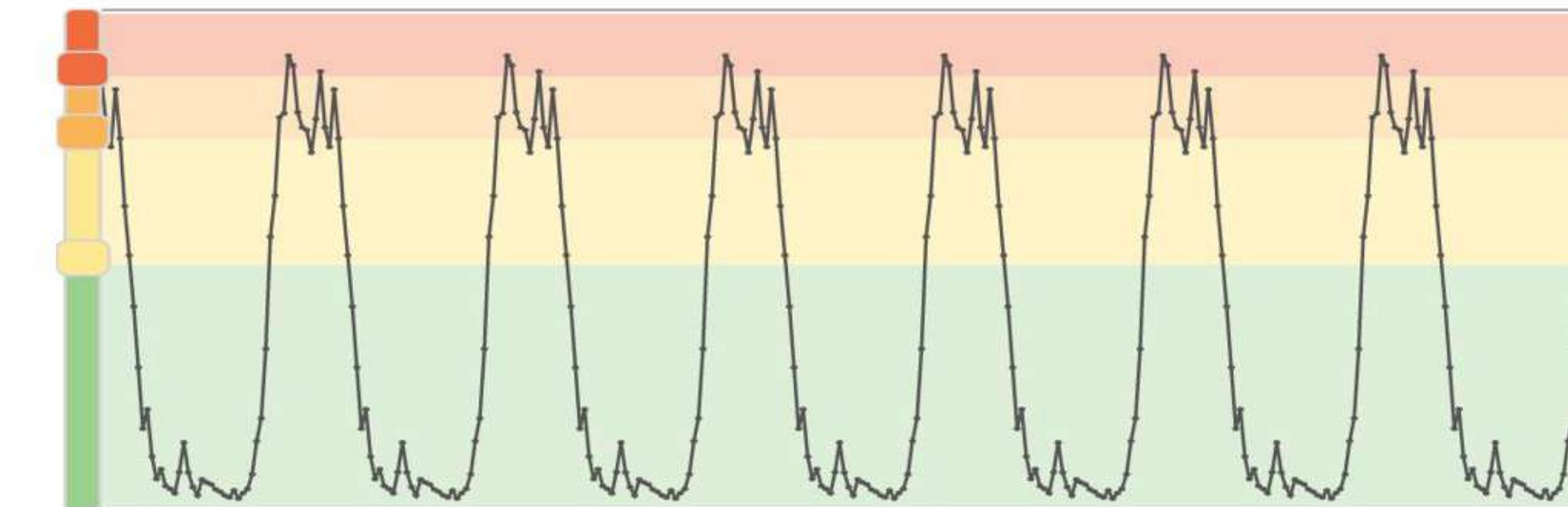
So where is the ML
hidden then?

Adaptive Thresholding

Adaptive Thresholding

Self-adjusting alert thresholds calculated by ML algorithms

Static thresholds aren't good enough for data that is expected to vary over time, machine learning ensures your alerts are accurate and never stale



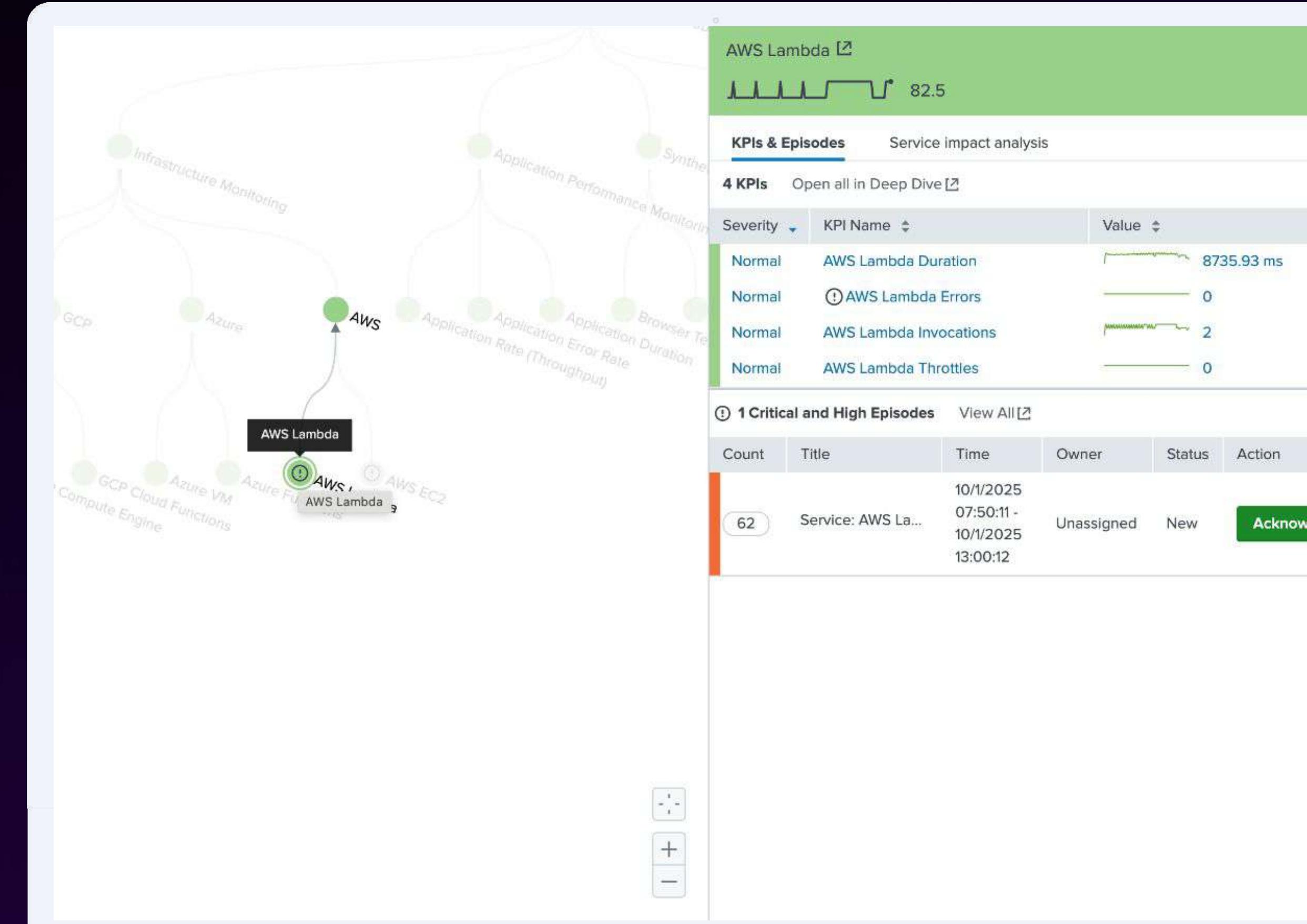
Adaptive Thresholding

Let's set it up.

We will use the Service
“AWS Lambda”

Clicking on the service title will
open the side pane with the KPIs
& Episodes.

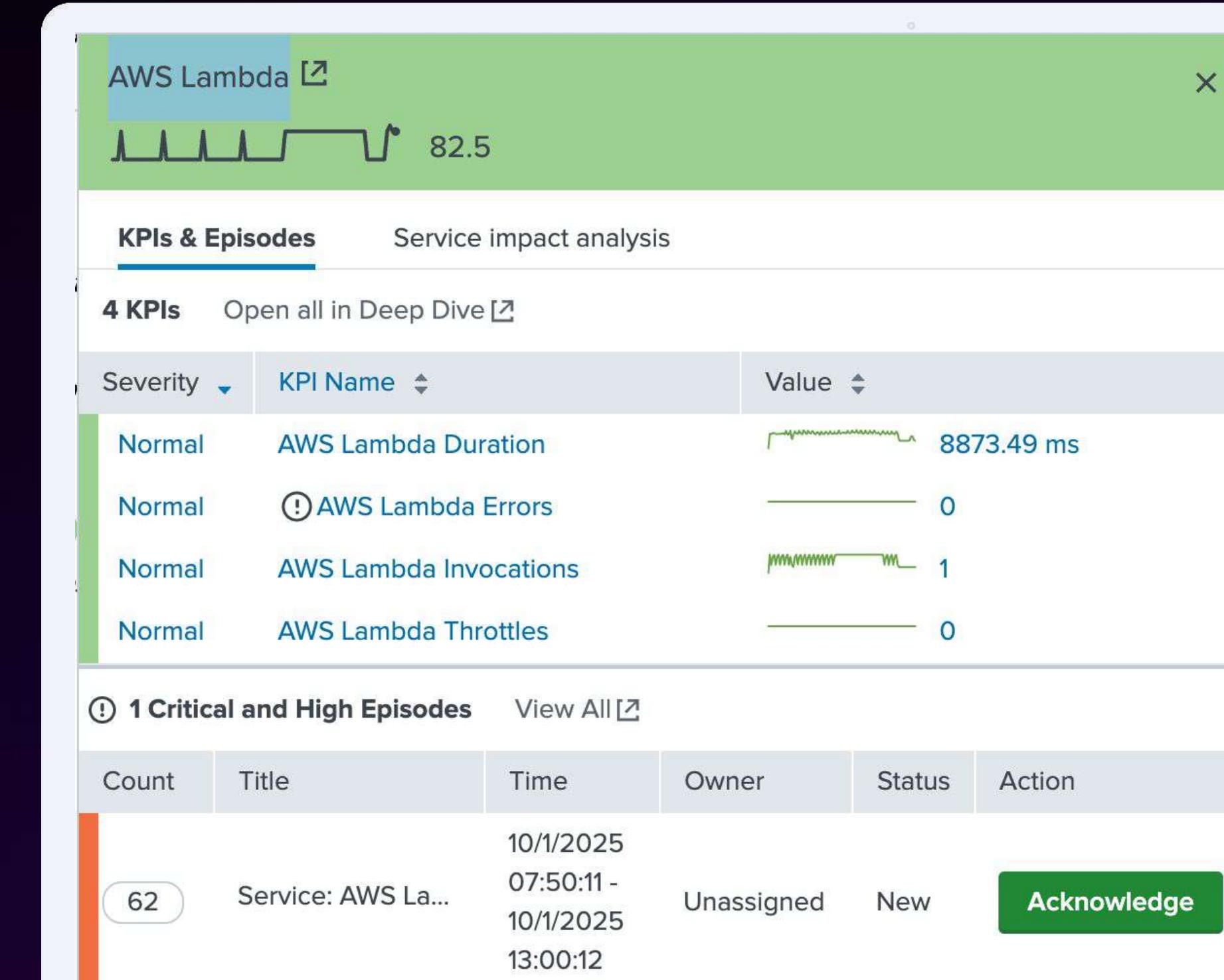
This will be our next step.



Adaptive Thresholding

In the side pane, click on the Service title.

This will take us to the service configuration page.



Adaptive Thresholding

Select the first KPI:
“AWS Lambda Duration”

We are not interested by the calculation of the KPI itself but by the thresholding method.

AWS Lambda

The AWS Lambda Service describes AWS Lambda metrics such as Invocations, Errors, Heartbeat, Throttling, and Duration.

Entities **KPIs** Service Dependencies Settings Predictive Analytics

KPIs Clone New ▾

AWS Lambda Duration X

AWS Lambda Errors X

AWS Lambda Invocations X

AWS Lambda Throttles X

AWS Lambda Duration

The Lambda Duration KPI describes the amount of time that your function code spends processing an event. The billed d...

> Search and Calculate

▼ Thresholding

Use Recommended Thresholding Configuration Start with values configured based on a unique AI-driven analysis of your data

Use Thresholding Template [Learn more](#)

Set Custom Thresholds

Enable Time Policies

Aggregate Thresholds Per-Entity Thresholds

Enable Adaptive Thresholding

Enable KPI Alerting

© 2024 SPLUNK INC.

Adaptive Thresholding

This KPI has no threshold set.

We will make full use of the assistant to auto detect optimal adaptive thresholding KPIs.

Select “Use Recommended Configuration”, then “Load Recommendations”

AWS Lambda

The AWS Lambda Service describes AWS Lambda metrics such as Invocations, Errors, Heartbeat, Throttling, and Duration.

Entities **KPIs** Service Dependencies Settings Predictive Analytics

KPIs Clone New ▾

AWS Lambda Duration	X
AWS Lambda Errors	X
AWS Lambda Invocations	X
AWS Lambda Throttles	X

AWS Lambda Duration

The Lambda Duration KPI describes the amount of time that your function code spends processing an event. The billed d...

> Search and Calculate

Thresholding

Use Recommended Thresholding Configuration Powered by Splunk AI Start with values configured based on a unique AI-driven analysis of this KPI.

Thresholding Direction ? Detect with Splunk AI

Analysis Window ? 30 days ▾ starting on 12/11/2024

Use Thresholding Template Select a thresholding template ▾ Learn more ↗

Set Custom Thresholds

Enable Time Policies ?

Aggregate Thresholds Per-Entity Thresholds

Enable Adaptive Thresholding ?

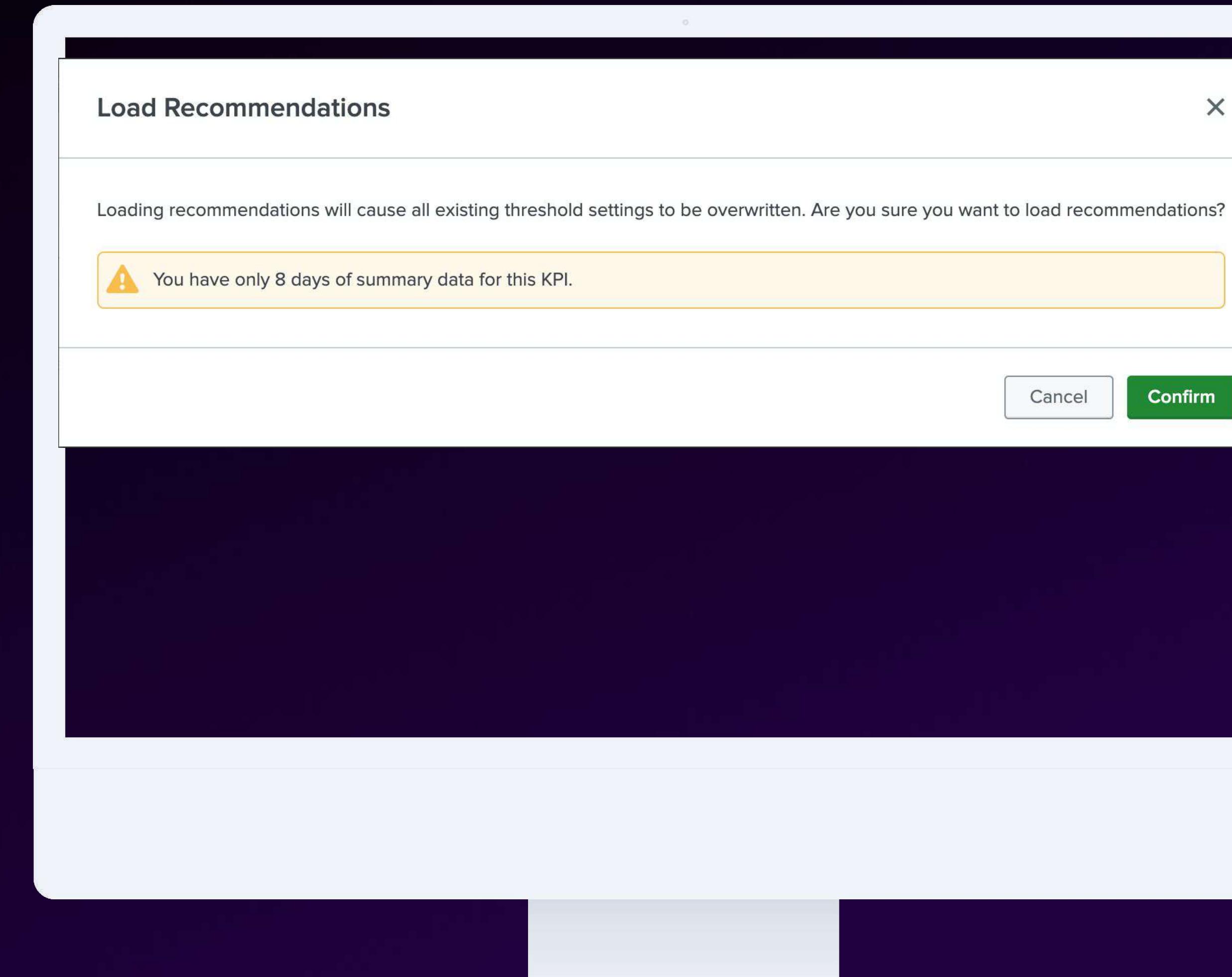
Enable KPI Alerting ?

Load Recommendations

Adaptive Thresholding

Acknowledge the warning

This is unfortunately a limitation
of our demo environment.



Adaptive Thresholding

The Results

We will blindly follow the recommendations and scroll down to see how the assistant profiled our historical data.

A production scenario would require a bit more testing, but for our lab this is perfect!

The screenshot shows the Splunk AI-powered adaptive thresholding configuration interface. At the top, there are three options: "Use Recommended Thresholding Configuration" (selected), "Use Thresholding Template" (unchecked), and "Set Custom Thresholds" (unchecked). Below this, a "Thresholding Direction" dropdown is set to "Threshold KPI both above and below a certain level". The "Analysis Window" is set to "30 days" starting on "12/11/2024". A green "Load Recommendations" button is visible. A modal window titled "Splunk AI recommendations" lists the following active settings:

- Time policies are active - based on analysis of KPI behavior over time.
- Adaptive thresholding with a 30 day analysis window using the standard deviation algorithm - this setting automatically adjusts threshold values by analyzing historical KPI behavior, and the algorithm method calculates threshold values that fit your data's behavior.
- Outlier exclusion is active using Interquartile Range (IQR) - this setting defines outlier data points to exclude from adaptive threshold calculations to avoid data skews.
- Half-day pattern, offset=10h recommended with Low confidence (Confidence Score = 13.90%) - settings that describe your data's patterns and are reflected in the recommendation.

Below the modal, a note states: "This icon represents threshold settings configured using Splunk's AI-powered analysis." There are two tabs: "Aggregate Thresholds" (selected) and "Per-Entity Thresholds". Under "Aggregate Thresholds", the following settings are shown:

- Enable Time Policies: On
- Enable Adaptive Thresholding: On
- Training window: 30 days
- Enable Outlier Exclusion: On
- Outlier algorithm: Interquartile Range (IQR)
- Trigger threshold: 0.1 (with a slider and input field showing 30 and 4.78)

A note below these settings states: "Adaptive Thresholding runs everyday around midnight and updates the thresholding for the KPI based on the settings below. Once updated, old thresholds cannot be recovered."

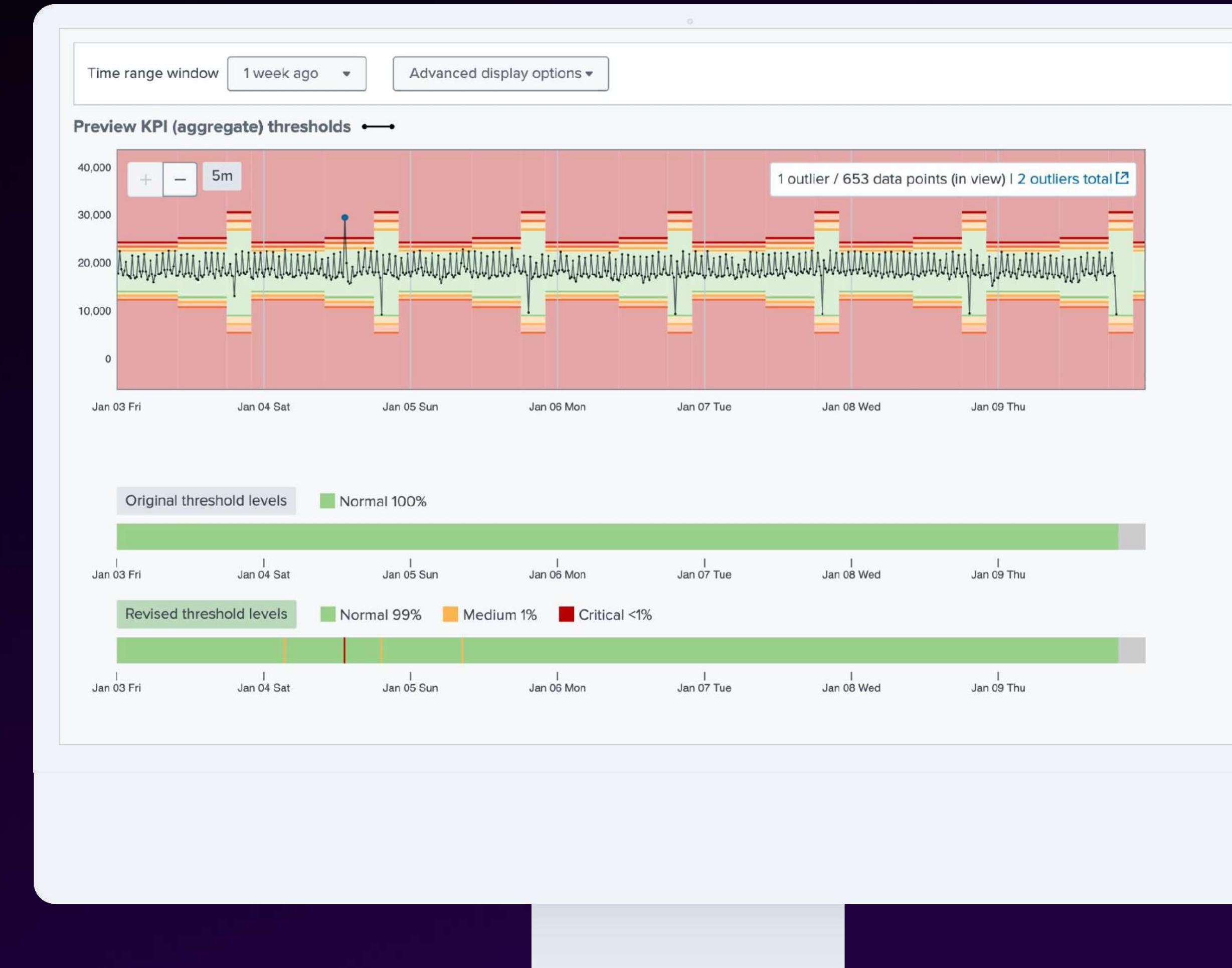
At the bottom, a message says: "Changes made to the settings that impact adaptive thresholding are not automatically reflected in the preview graph. You must update using the button below to see your changes." A "Preview Adaptive Thresholds" button is located at the bottom right.

Adaptive Thresholding

Graphically

As you can see, this method would only make sense for datasets where cycles are visible (hourly, daily, weekly, etc...) .

If not, well, you can still rely on static thresholding ;-)

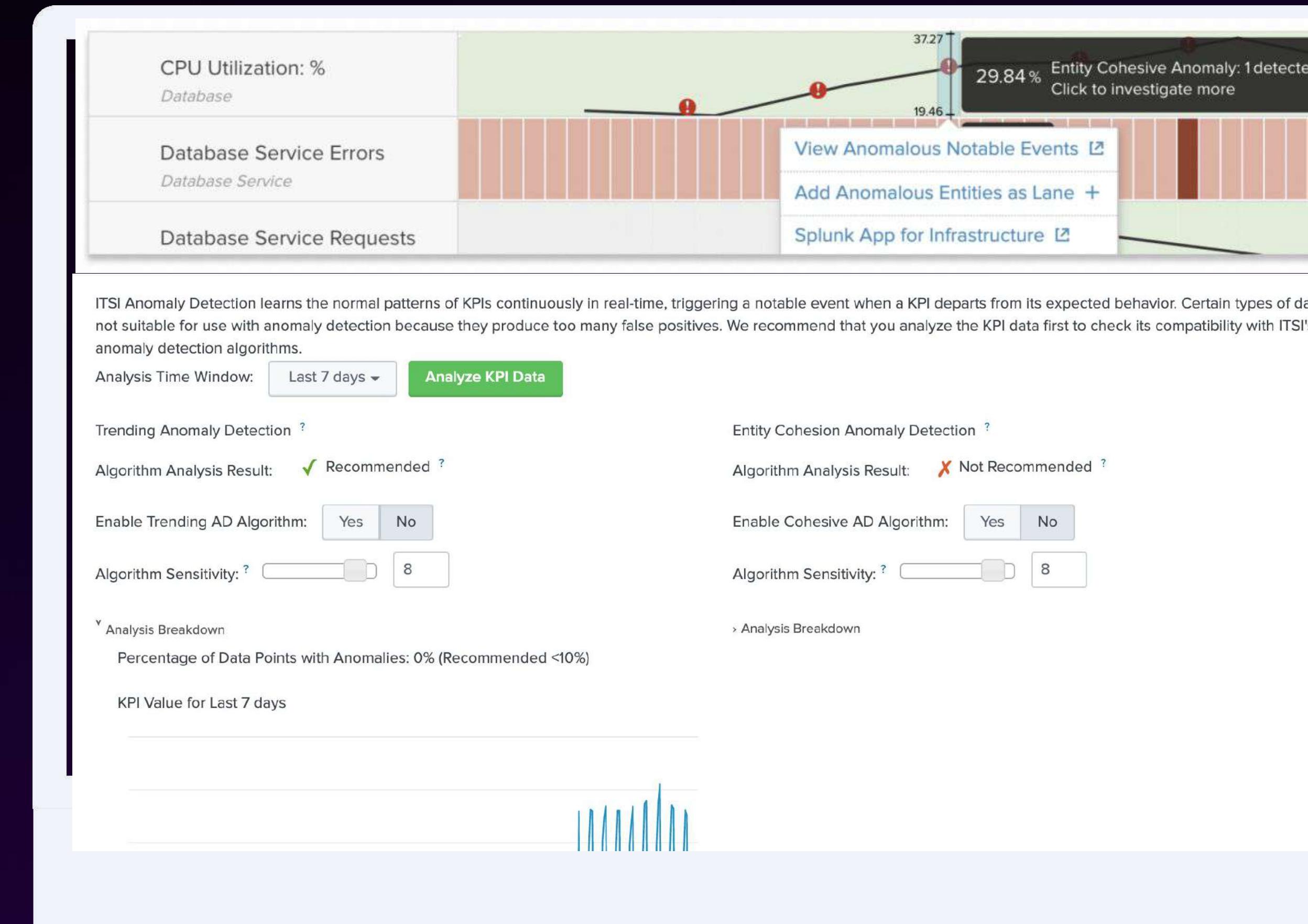


Anomaly Detection

Anomaly Detection

Trending and entity cohesion anomaly detection algorithms applied to KPIs and alerts.

Alerts are continuously updated based on new patterns and behaviors to effectively identify true outliers



Anomaly Detection

Let's go back to our AWS Lambda Duration KPI.

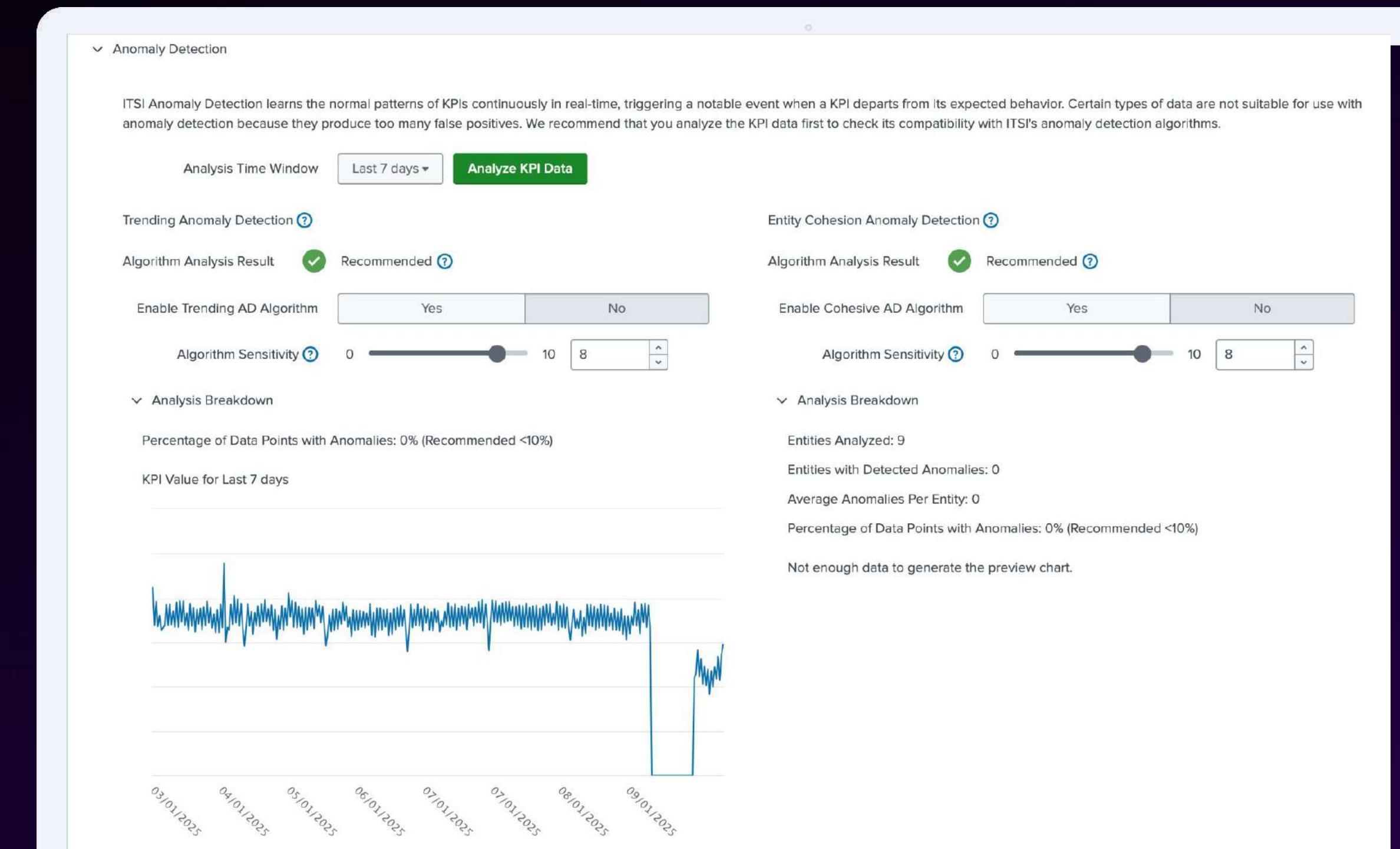
This time we will choose the last option.

The screenshot shows the Splunk Enterprise interface with the title "AWS Lambda". The top navigation bar includes links for Service Analyzer, Entity Overview, Alerts and Episodes, Glass Tables, Deep Dives, Dashboards, Configuration, and Search. Below the title, a sub-header states: "The AWS Lambda Service describes AWS Lambda metrics such as Invocations, Errors, Heartbeat, Throttling, and Duration." A horizontal tab bar at the top of the main content area includes Entities, KPIs (which is selected), Service Dependencies, Settings, and Predictive Analytics. Under the KPIs tab, there is a sub-tab for "KPIs" and buttons for "Clone" and "New". A list of four KPIs is shown: "AWS Lambda Duration" (selected), "AWS Lambda Errors", "AWS Lambda Invocations", and "AWS Lambda Throttles". To the right of the KPI list, a detailed description of the "AWS Lambda Duration" KPI is provided: "The Lambda Duration KPI describes the amount of time that your function code spends processing an event. The billed d...". Below this description are three options: "Search and Calculate", "Thresholding", and "Anomaly Detection", with "Anomaly Detection" being the last option listed.

Anomaly Detection

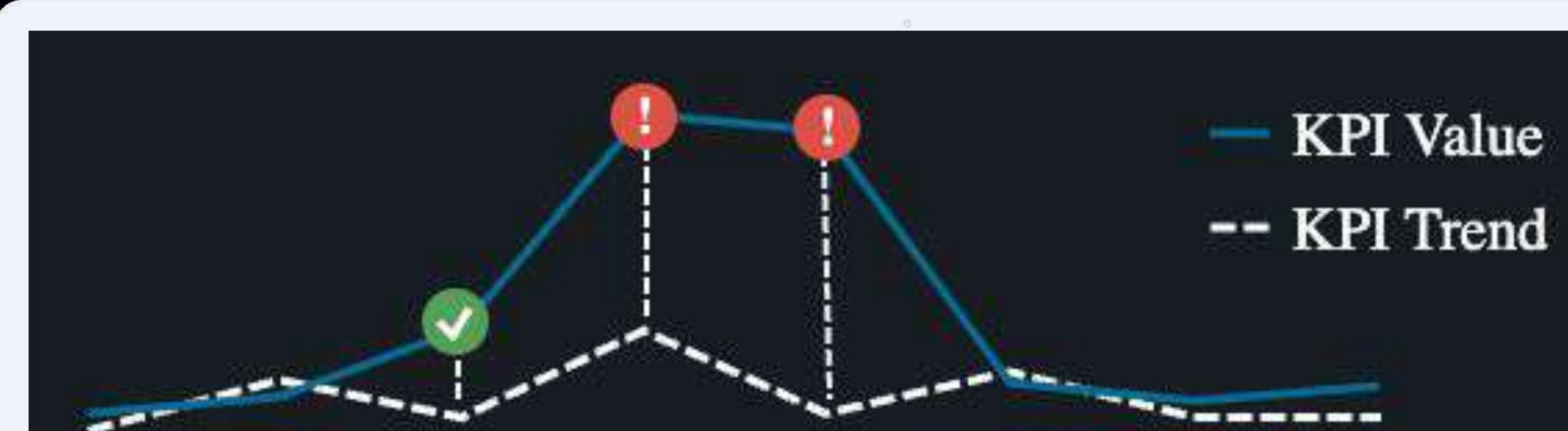
We only need to analyze the KPI Data and review the assistant's recommendations.

Two algorithms are available: Trending and entity cohesion anomaly detection



Anomaly Detection

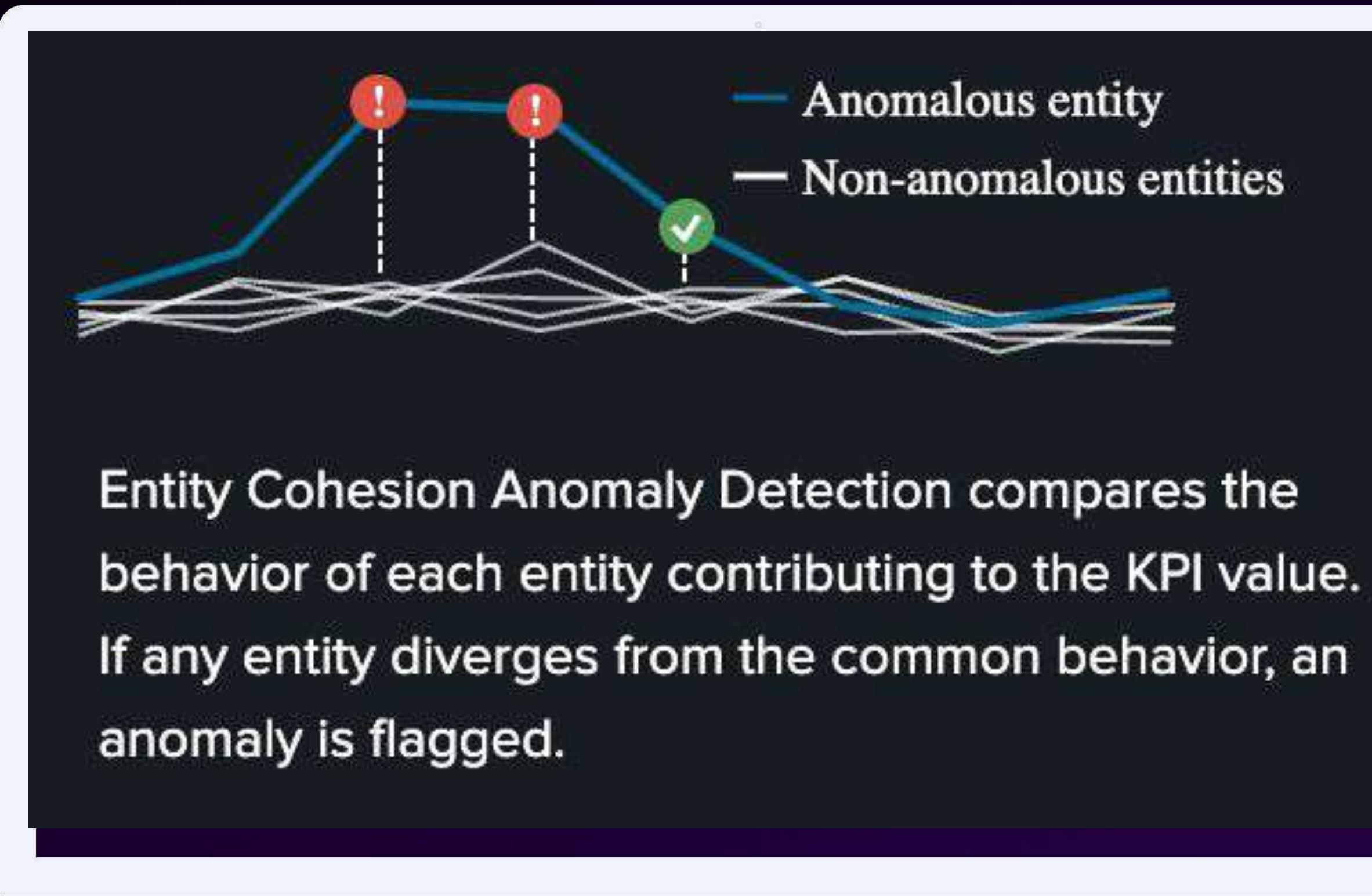
Trending Anomaly Detection



Trending Anomaly Detection compares the aggregated value of the KPI to its historical trend. If the value diverges from the trend too much, an anomaly is flagged.

Anomaly Detection

Entity Cohesion Anomaly Detection

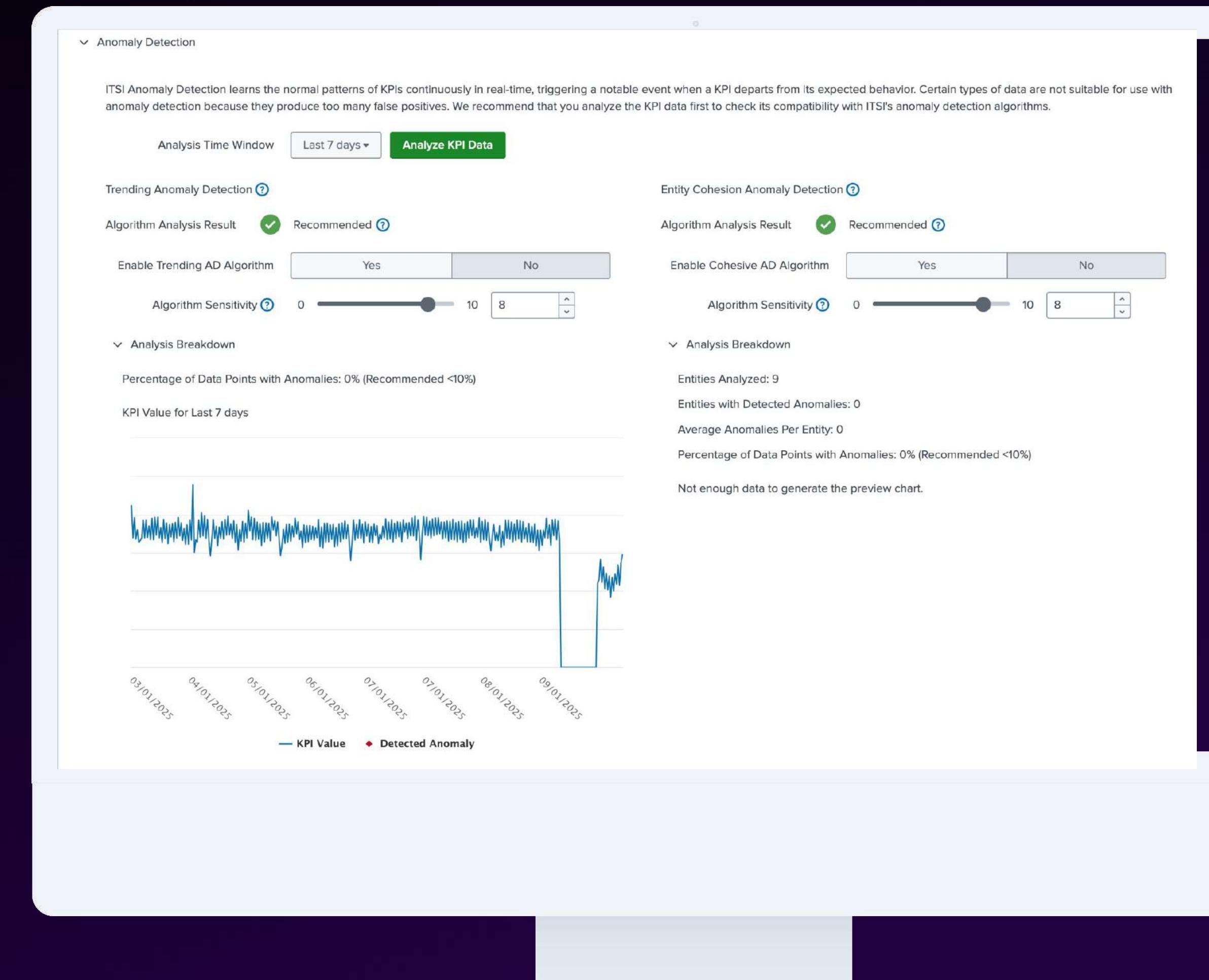


Anomaly Detection

Depending on the results found, the assistant will offer an easy to use menu to activate each option.

In this specific case, both algorithms are suitable (this is in practise quite rare).

Let's activate both.

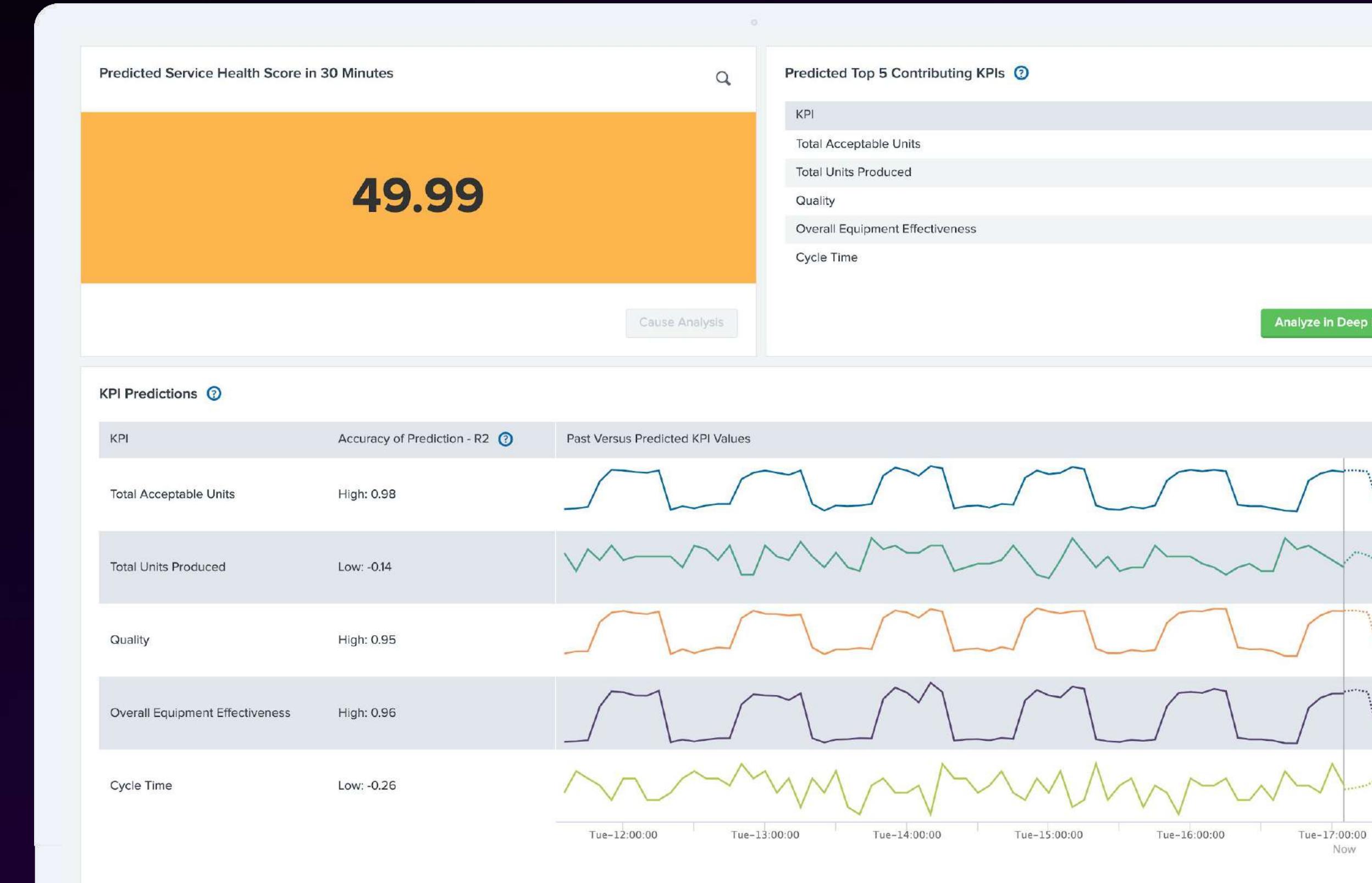


Predictive Analytics

Predictive Analytics

ML models for alerts that predict future KPI scores 30 minutes in advance

Find and fix issues before they impact your end users



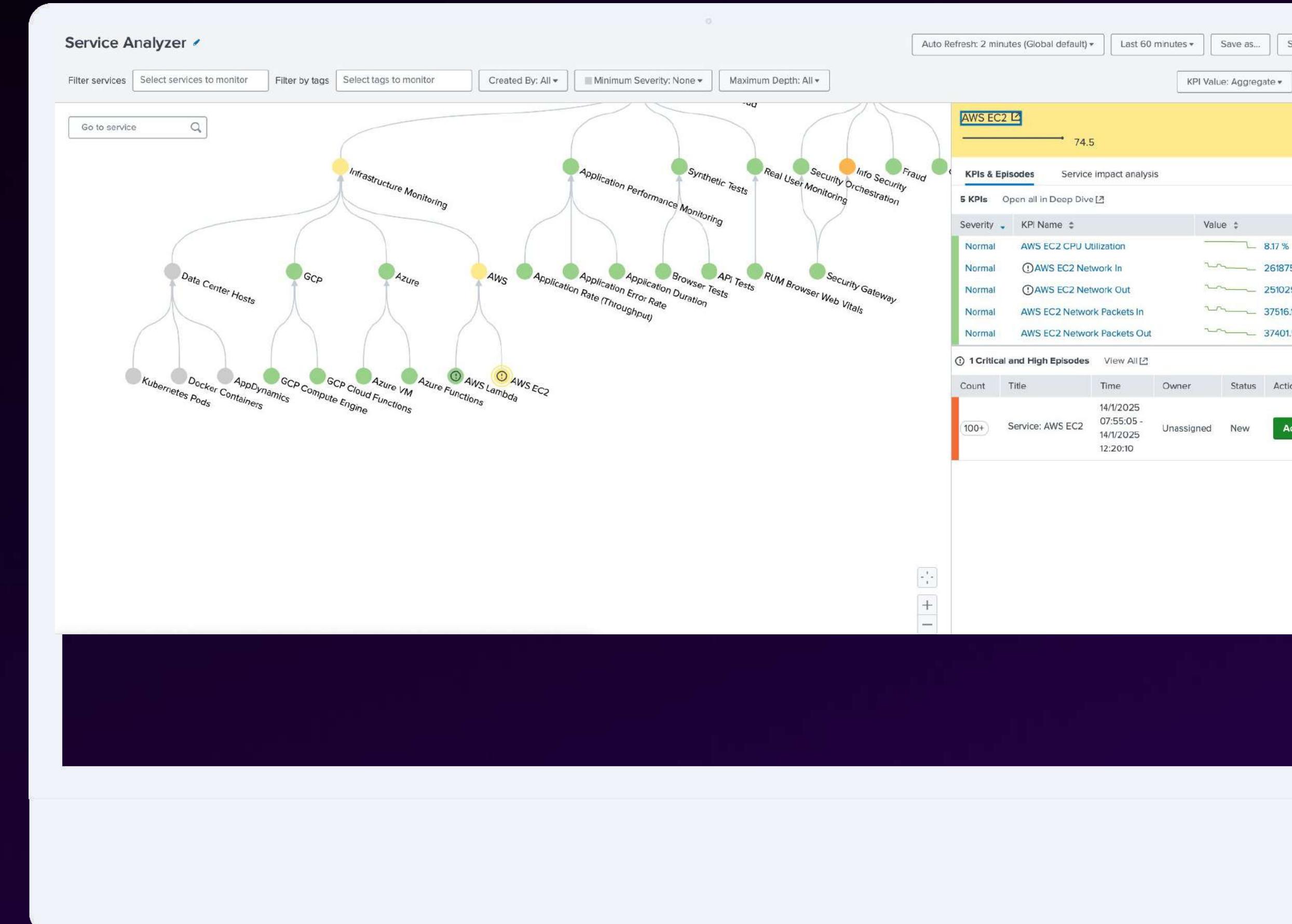
Predictive Analytics

This time we will work at the Service level.

Our pick will be the Service “AWS EC2”

The goal is to have a sneak peak at the future service health score on a short time scale (30 mins ahead), based on historical KPI data.

As you can see from the screenshot this service was cherry picked for a very good reason :-)

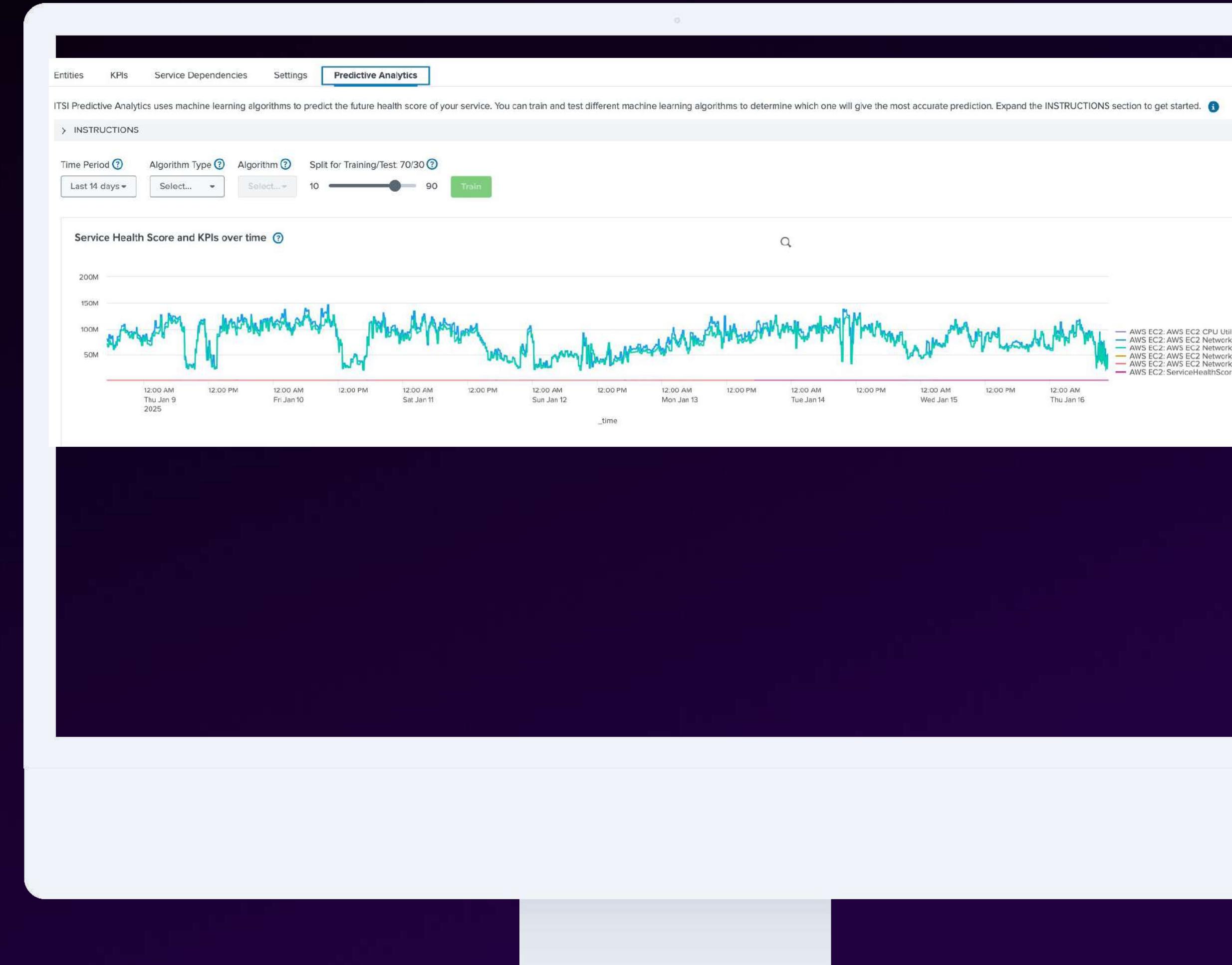


Predictive Analytics

Let's head to the Service Definition page, tab “Predictive Analytics”

The Service Health Score and KPIs over time already tells an interesting tale.

Let's see now how we can leverage our predictive algorithms to get a glimpse into the future.



Predictive Analytics

The instructions should looks fairly familiar.

As you can see, everything will be menu driven.

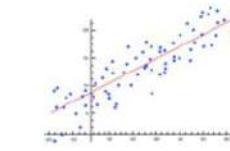
We have three Regression algorithms and one for Classification to choose from.

INSTRUCTIONS

TRAIN

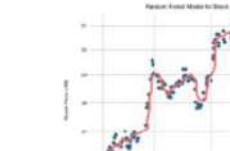
1. Select a training period.
2. Select an algorithm type.
3. Select the algorithm. Algorithm details are listed below.

Linear Regression



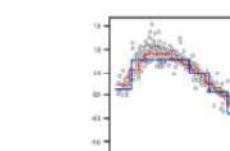
This algorithm fits a linear line to your data, using each input (KPIs and historical service health scores) as an additional dimension. It assumes that your data is normal and is highly scalable.

Random Forest Regressor



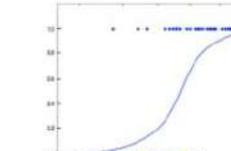
This algorithm takes the inputs (KPIs and historical service health scores) and forms a random decision tree (a "forest") to determine the output value. This model makes no assumptions about the normalcy of your data, but requires more processing power and takes longer to run.

Gradient Boosting Regressor



This algorithm uses a loss function to fit a line to your data, a decision tree, and an additive model to predict the service health score value. Think of this as a combination of the Random Forest Regressor and Linear Regression.

Logistic Regression



This algorithm predicts a service health state (Normal, Moderate, Critical) rather than a numerical value. Use this algorithm if you have tuned KPI importance values or adaptive thresholds to normalize your service health scores.

TEST

4. Configure the train/test ratio to divide your data into separate training and testing sets.
5. Click **Train** to generate a model.

To use these models to perform root cause analysis on a specific health score prediction, see the [Predictive Analytics Dashboard](#). You can also add predictive models to glass tables. [Learn more](#)

Time Period [?](#) Algorithm Type [?](#) Algorithm [?](#) Split for Training/Test: 70/30 [?](#)
Last 14 days [Select...](#) [Select...](#) 10 [Train](#)

Predictive Analytics

We will now train each of the three Regression algorithms proposed (Classification won't work with this dataset).

We will keep the 14 days default value as well as the 70/30 split, then press the “Train” button and wait.

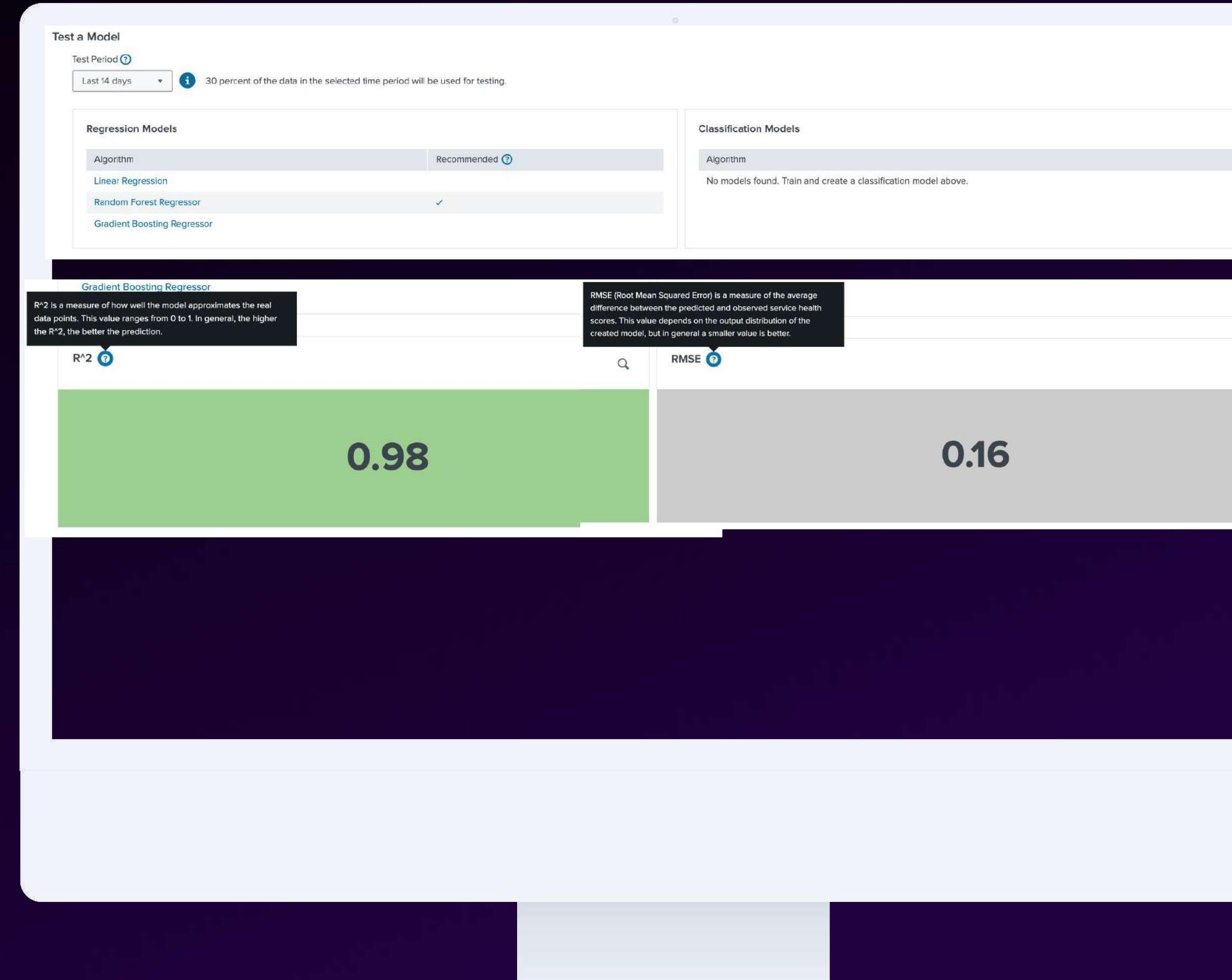
Once each training is completed, you can proceed to the next. It could be a good idea to press the “Save” button at the bottom of the page (this is not strictly necessary though). We will look into the results afterwards.



Predictive Analytics

The test results should look like this

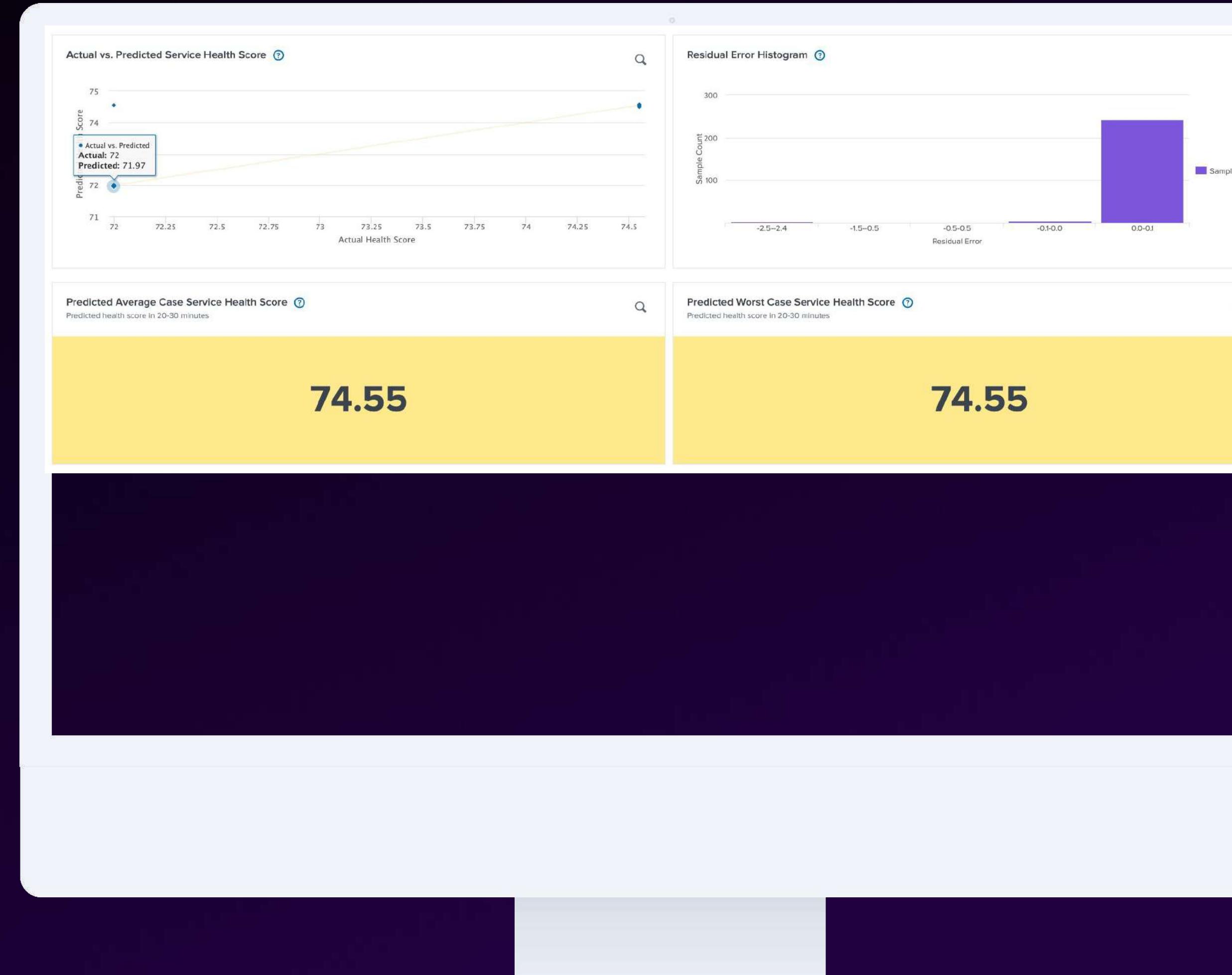
Based on the R² and RMSE values, the assistant will recommend one (or none) of the the algorithms.



Predictive Analytics

This is how our predicted Service Health Score will look like in 20-30 minutes!

Time to press the save button (in case it was not done already)



Predictive Analytics

The Models can now be used via the “Predictive Analytics” dashboard.

Let's head there

The screenshot shows a dark-themed Splunk interface. At the top, there is a navigation bar with tabs: Glass Tables, Deep Dives, Dashboards ▾, Configuration ▾, and Search. A dropdown menu is open under the Dashboards tab, with "Predictive Analytics" highlighted with a blue border. Other options in the dropdown include Event Analytics Monitoring, Event Analytics Audit, ITSI Health Check, ITSI SVC Statistics, Upgrade Readiness, and Demo Health Overview. Below the navigation bar, there is a search bar with the placeholder "sample search" and a magnifying glass icon. To the right of the search bar, there is some text: "metrics such as CPU, Disk, Network, and Memory usage. Predict the future health score based on historical data using machine learning algorithms." On the left side of the main area, there are several UI elements: a "Regressor" dropdown menu with "Linear Regressor" selected, a "Split factor" dropdown menu with "10" selected, and a "sample search" button.

Predictive Analytics

We will reselect our AWS EC2 Service and we should find our trained models

If not, there is unfortunately no other way than going again to the service configuration page and re-train the models

(don't forget to save this time!)

Predictive Analytics

ITSI Predictive Analytics uses machine learning to predict the health status of your infrastructure.

Service [?](#) Model [?](#)

AWS EC2 ▾ RandomForestRegressor ▾

filter

REGRESSION

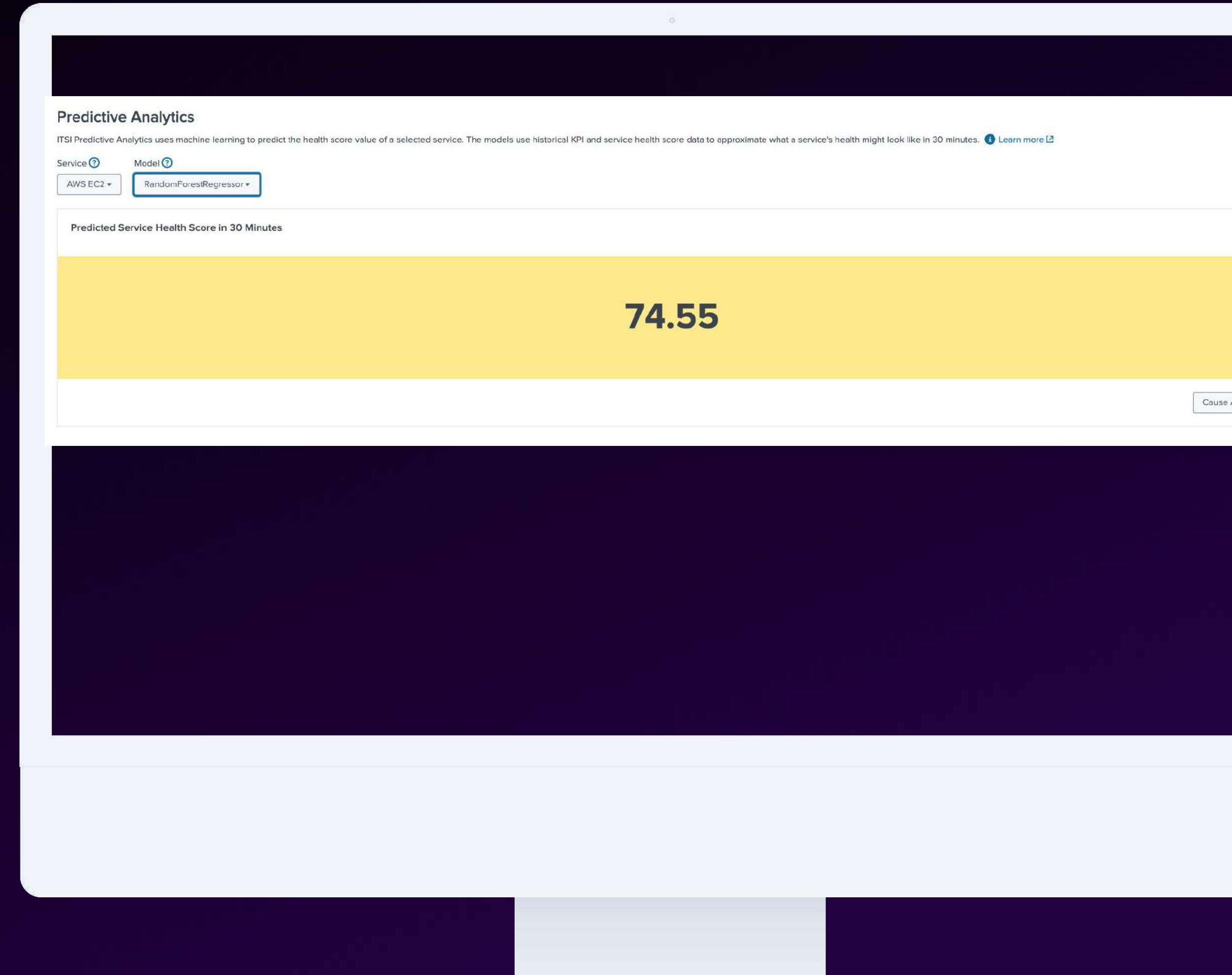
- LinearRegression
- RandomForestRegressor**
- GradientBoostingRegressor

Predictive Analytics

Here's our predicted health score.
But there's more...

We can also access the underlying search via the magnifier icon (convenient for reusing in a dashboard or glass table)

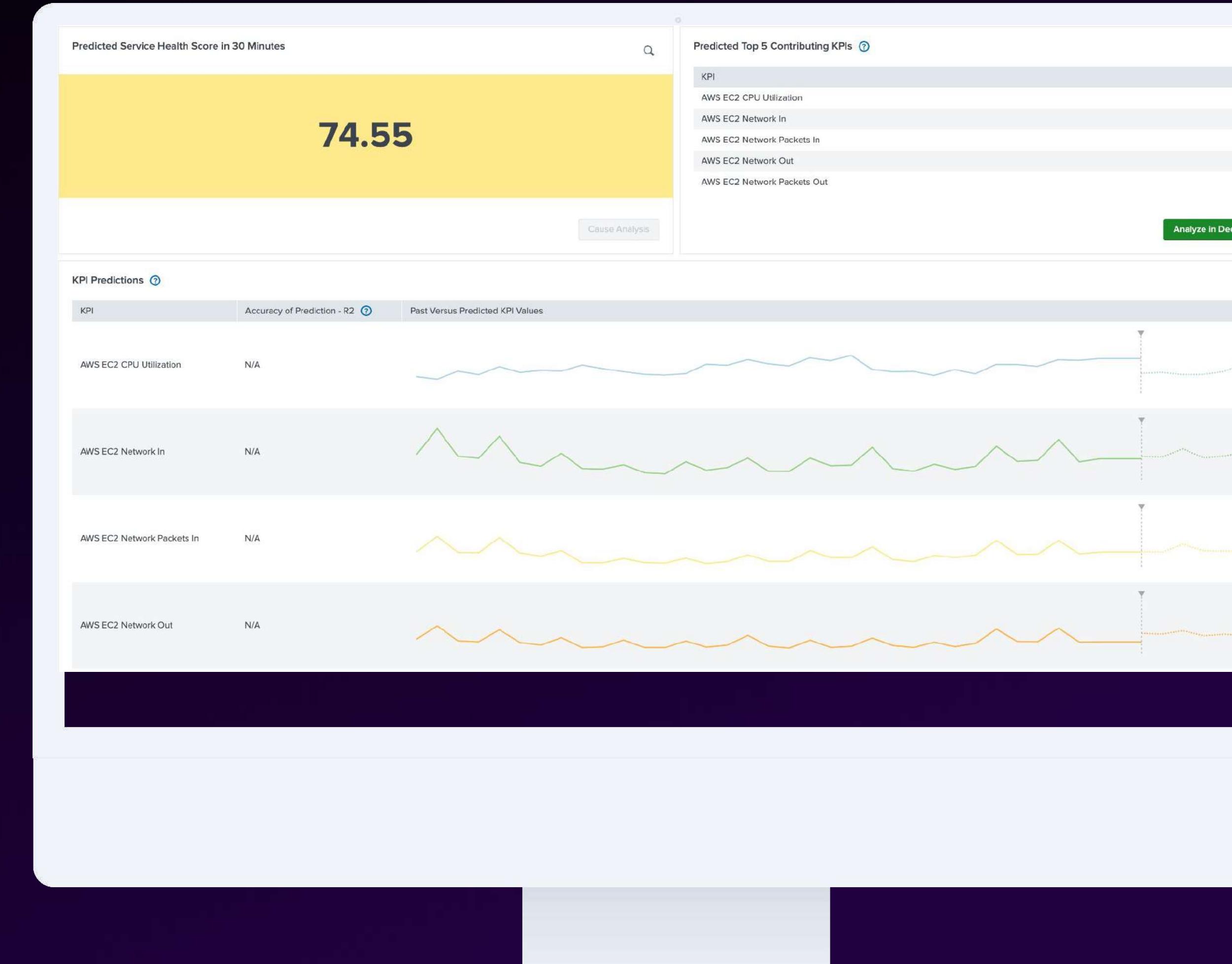
Or see the prediction for the individual KPIs in the “Cause Analysis”



Predictive Analytics

Simple, isn't it?

The Deep Dive Analysis will be for another session...



Event Aggregation

Event Analytics

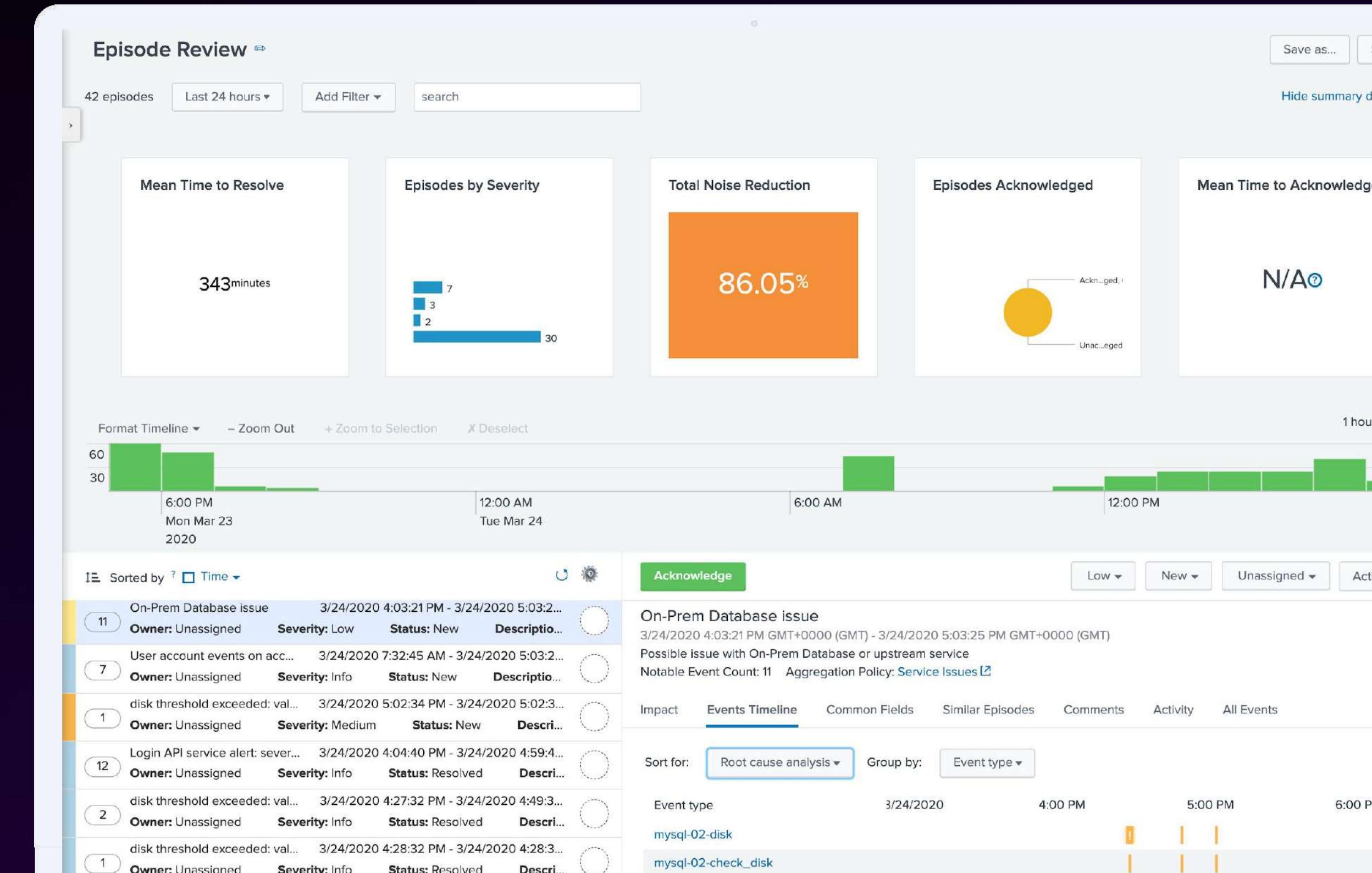
Event correlation & aggregation,
episode review, ITSM integrations



Event Aggregation

Machine learning policies automatically correlate and trigger alerts as data enters the system

Immediate alert noise reduction

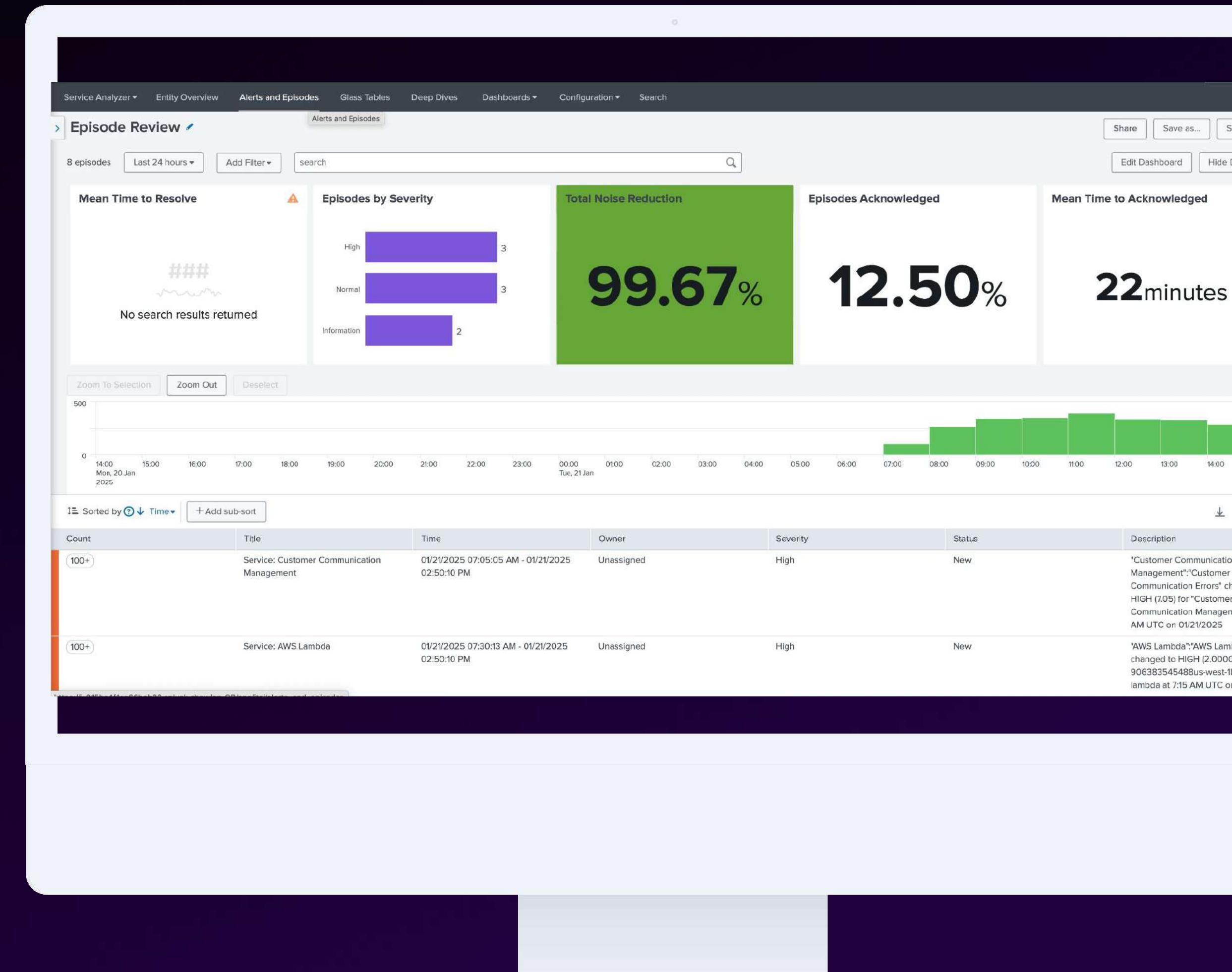


Event Aggregation

Let's switch to the Alerts and Episodes tab.

We will not configure ML directly but focus on the high level logic.

The ML use case is to mitigate a problem common with many monitoring systems: When there are too many alerts, we lose track of what's important. The meaningful events get buried into the "noise".

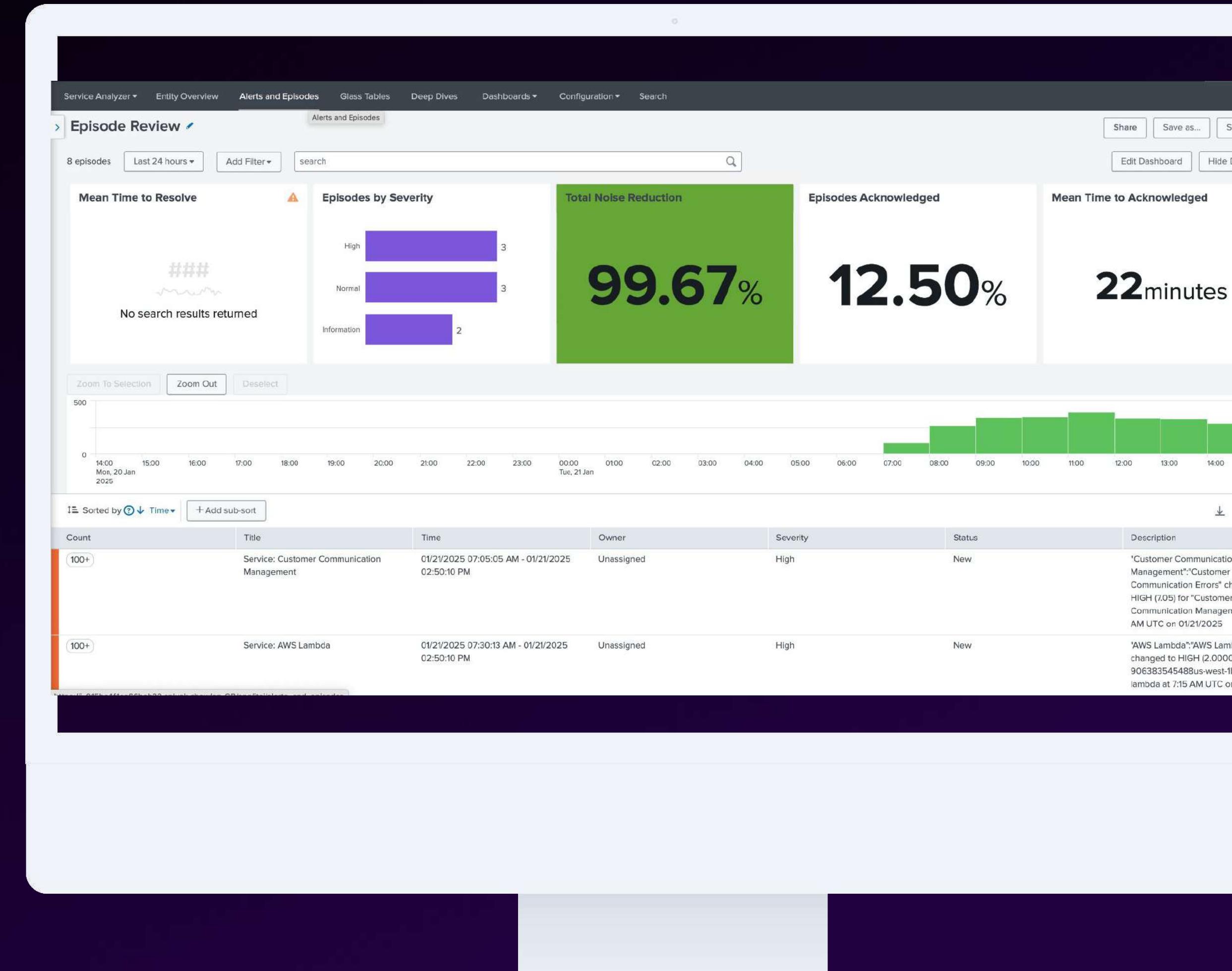


Event Aggregation

ITSI includes a sophisticated way to deal with this alert noise:
Notable Event Aggregation Policies.

As you might have guessed by now, their role is to group similar alerts together.

The idea is to remove as much redundancy as possible (a.k.a noise reduction).

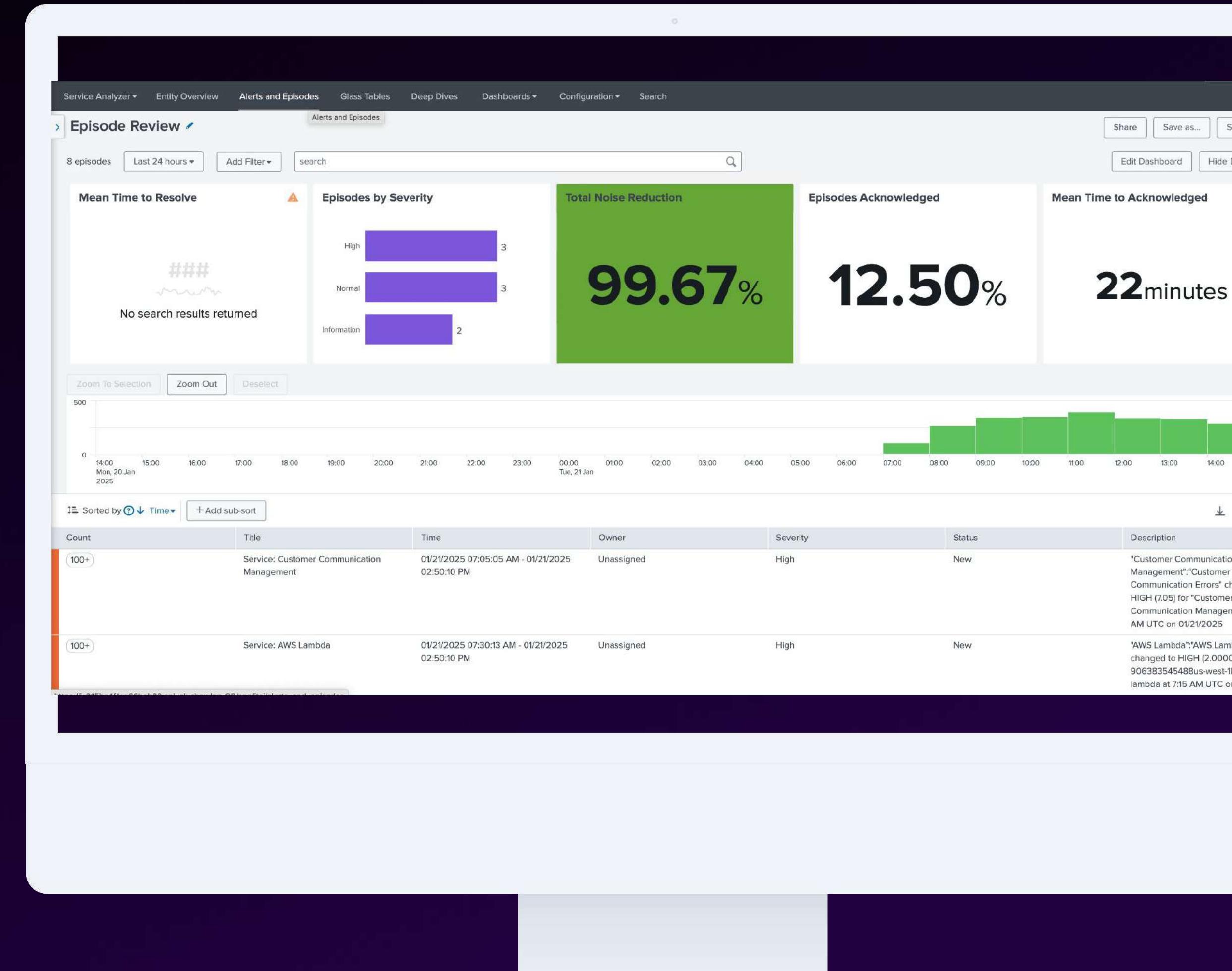


Event Aggregation

The top level dashboard indicates how successful we are at noise reduction. (it can be hidden and the views customized).

The bottom panel shows the results: similar events have been grouped (clustered) into episodes.

No data is lost, all the individual events are still there, let's have a look!



Event Aggregation

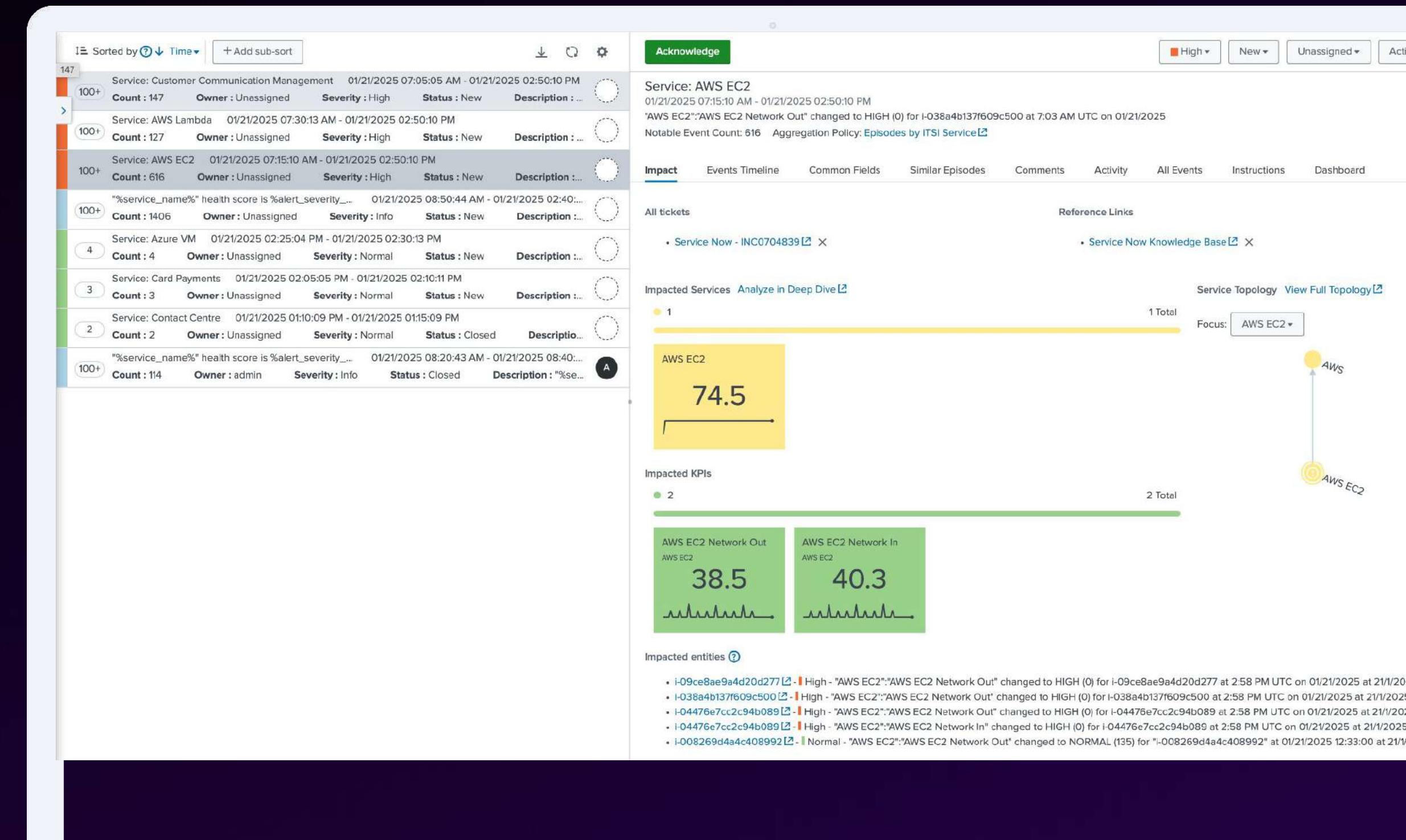
If we scroll down a bit, we should spot our AWS EC2 service in the list of active episodes.

Let's click on it.

Count	Title	Time	Owner	Severity	Status	Description
100+	Service: Customer Communication Management	01/21/2025 07:05:05 AM - 01/21/2025 02:45:07 PM	Unassigned	High	New	"Customer Communication Management": "Customer Communication Errors" changed to HIGH (7.05) for "Customer Communication Management" at 01/21/2025 02:45:07 PM UTC on 01/21/2025
100+	Service: AWS Lambda	01/21/2025 07:30:13 AM - 01/21/2025 02:45:06 PM	Unassigned	High	New	"AWS Lambda": "AWS Lambda" changed to HIGH (2.000000) for "AWS Lambda" at 7:15 AM UTC on 01/21/2025
100+	Service: AWS EC2	01/21/2025 07:15:10 AM - 01/21/2025 02:45:06 PM	Unassigned	High	New	"AWS EC2": "AWS EC2 Network" changed to HIGH (0) for "AWS EC2 Network" at 038a4b137f609c500 at 01/21/2025
100+	"%service_name%" health score is %alert_severity_upper%	01/21/2025 08:50:44 AM - 01/21/2025 02:40:44 PM	Unassigned	Info	New	"%service_name%" health score is %alert_severity_upper% changed to %alert_severity_upper% (%alert_value%) at %alert_value%" at 01/21/2025
4	Service: Azure VM	01/21/2025 02:25:04 PM - 01/21/2025 02:30:13 PM	Unassigned	Normal	New	"Azure VM": "Azure VM Network" changed to HIGH (0) for "Azure VM Network" at 2:27-4418-9a31-doc67b04 tests/microsoft.compute/vm/system-test-lbzpulat:system-lbzpulat at 2:19 PM UTC on 01/21/2025
3	Service: Card Payments	01/21/2025 02:05:05 PM - 01/21/2025 02:10:11 PM	Unassigned	Normal	New	"Card Payments": "Card Payments" at 2:02 PM UTC on 01/21/2025
2	Service: Contact Centre	01/21/2025 01:10:09 PM - 01/21/2025 01:15:09 PM	Unassigned	Normal	Closed	"Contact Centre": "Average call duration" changed to HIGH (98.4) for "Contact Centre" at 1:05 PM UTC on 01/21/2025

Event Aggregation

As you can see, the episode was enriched with meaningful context. But this part is beyond our scope. We will just have a look at alert aggregation in the Events timeline.

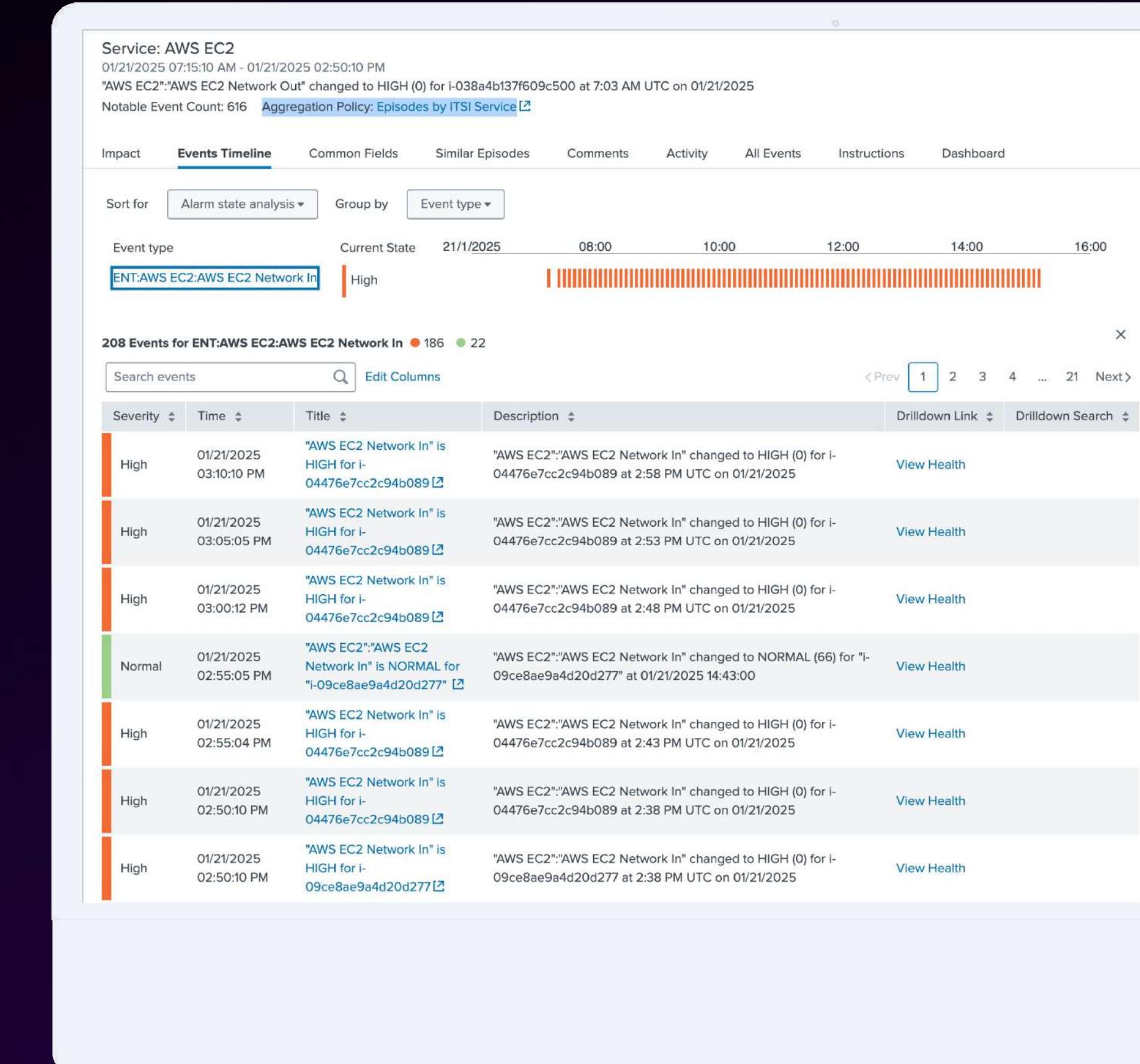


Event Aggregation

Our episode is made of several hundred alerts clustered together and deduplicated on the same timeline.

This is clearly more efficient than looking at each alert, but how did that magic happen?

The answer is in the episode description: “Aggregation Policy: Episodes by ITSI Service.”

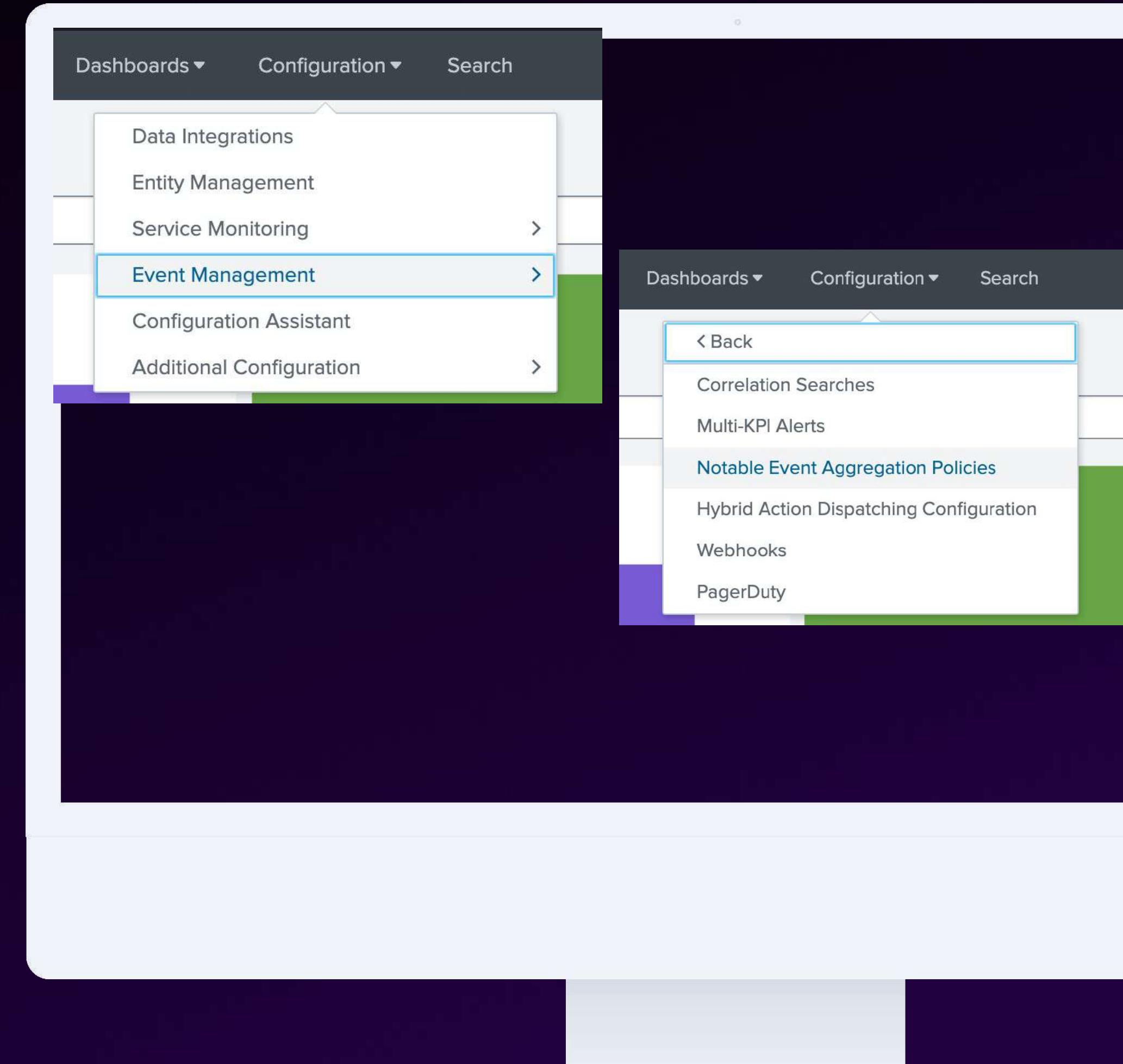


Event Aggregation

There come the Notable Event Aggregation Policies.

We can find them in the ITSI configuration menu, under Event Management.

(Or we could simply click on the policy name in the episode description).



Event Aggregation

Each Policy includes a specific set of rules for Event grouping and Episode lifecycle

The default policies provide a solid base that can be further customized.

We won't dive into the details but as you can see, the configuration is pretty straightforward: on/off

Notable Event Aggregation Policies			
Notable event aggregation policies group related events and alerts into episodes to reduce alert noise.			
10 Notable Event Aggregation Policies	Bulk Action ▾	filter	Actions
<input type="checkbox"/> Default Policy	Edit ▾	Enabled	
<input type="checkbox"/> Default SNMP Policy	Edit ▾	Disabled	
<input type="checkbox"/> Entity Type Alerts	Edit ▾	Enabled	
<input type="checkbox"/> Episodes by Alarm	Edit ▾	Disabled	
<input type="checkbox"/> Episodes by Alert Group	Edit ▾	Disabled	
<input type="checkbox"/> Episodes by ITSI Service	Edit ▾	Enabled	
<input type="checkbox"/> Episodes by Src	Edit ▾	Disabled	
<input type="checkbox"/> ITSI Alert and Episode Monitoring	Edit ▾	Disabled	
<input type="checkbox"/> KPI Alerting Policy	Edit ▾	Enabled	
<input type="checkbox"/> Normalized Policy (Splunk App for Infrastructure)	Edit ▾	Enabled	

Event Aggregation

If we open our policy **Episodes by ITSI Service**. We can see that everything is menu driven and that the GUI allow us to preview the results.

Pretty much everything can be customized, there's even an (ML powered) "Smart Mode" that uses similarity matching (text or categorical).

The screenshot shows the 'Episodes by ITSI Service' configuration page. On the left, the 'Filtering Criteria and Instructions' section contains two main rule sections: 'Include the events if' and 'Smart Mode grouping'. The 'Include the events if' section has two rules: one for 'serviceid' matching 'Episode Monitoring' and another for 'itsi_policy_id' matching 'da-itsi-cp-monitoring-alerting-1'. The 'Smart Mode grouping' section includes a toggle for 'Smart Mode' which is off. Below these are sections for 'Split events by field' (with 'serviceid' selected) and 'Break episode' (which is collapsed). On the right, the 'Preview with the Last 24 hours' section displays a table of 8 episodes. Each episode row shows a count (e.g., 2, 658, 147, 2, 4, 3, 4, 162), a title, a detailed description, and a severity indicator (Normal, High, or Unknown). The table includes columns for Count, Title, Description, Severity, and a small preview icon.

i	Count	Title	Description	Severity
>	2	"Average Handle Time" is HIGH for "Contact Centre"	"Contact Centre":"Average Handle Time" changed to HIGH (98.4) for "Contact Centre" at 1:05 PM UTC on 01/21/2025	Normal
>	658	"AWS EC2 Network Out" is HIGH for i-0fd21952ef5474ae5	"AWS EC2":"AWS EC2 Network Out" changed to HIGH (0) for i-0fd21952ef5474ae5 at 7:03 AM UTC on 01/21/2025	High
>	147	"AWS Lambda Errors" is HIGH for 906383545488us-west-1bits-apiauth-lambda	"AWS Lambda":"AWS Lambda Errors" changed to HIGH (2.00000) for 906383545488us-west-1bits-apiauth-lambda at 7:15 AM UTC on 01/21/2025	High
>	2	"Azure Functions Bytes Received" is HIGH for 5db89e2c-2127-4418-9a31-dac67b04e87b:East US:bits-metrics-eu0	"Azure Functions":"Azure Functions Bytes Received" changed to HIGH (51) for 5db89e2c-2127-4418-9a31-dac67b04e87b:East US:bits-metrics-eu0 at 3:12 PM UTC on 01/21/2025	Normal
>	4	"Azure VM Network Out" is HIGH for 5db89e2c-2127-4418-9a31-dac67b04e87b/system-tests/microsoft.compute/virtualmachines/system-test-ibzpuat:system-test-ibzpuat	"Azure VM":"Azure VM Network Out" changed to HIGH (0) for 5db89e2c-2127-4418-9a31-dac67b04e87b/system-tests/microsoft.compute/virtualmachines/system-test-ibzpuat:system-test-ibzpuat at 2:19 PM UTC on 01/21/2025	Normal
>	3	"Card Payments" health score is HIGH	"Card Payments" health score changed to HIGH (30.0) at 3:14 PM UTC on 01/21/2025	High
>	4	"Conversion Rate By Bank" is CRITICAL for 'Card Payments'	"Card Payments":"Conversion Rate By Bank" changed to CRITICAL (14.637) for 'Card Payments' at 2:02 PM UTC on 01/21/2025	High
>	162	"Customer Communication Errors" is HIGH for	"Customer Communication Errors" is HIGH for	High

Event Aggregation

We can try to activate Smart Mode (don't select all fields but keep the Default)

Previewing results should produce a slightly lower episode count than with the default grouping rules (7 vs 8 in our example).

This is not intended to replace the default policies, but can make sense to cover specific use cases.

Voila!

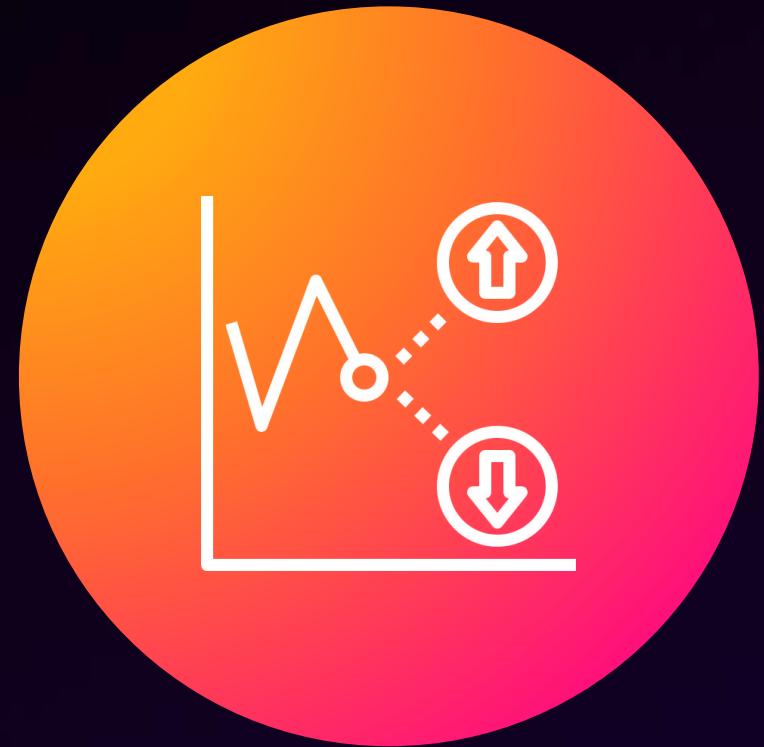
The screenshot shows the Splunk Event Aggregation interface. On the left, the 'Filtering Criteria and Instructions' section displays two AND rules: one for 'serviceid' matching '*' and another for 'source' not matching 'Episode Monitoring'. Below these are sections for 'Smart Mode grouping' (with a toggle switch turned on), 'Text Similarity' (set to 0.5), 'Category Similarity' (set to 0.5), 'Split by Service' (enabled), and 'Split by Entity' (disabled). On the right, the 'Episode types for Last 24 hours' section shows a table titled '7 episodes' with columns for 'Count', 'Summary', 'Similar Text Values', and 'Similar Category Values'. The table lists seven groups of events with their respective counts and details about what makes them similar. A green 'Preview' button is visible at the top right of the interface.

i	Count	Summary	Similar Text Values	Similar Category Values
>	662	662/988 events are grouped because they share 1 (average) common field values.	None	service_name=AWS EC2
>	164	164/988 events are grouped because they share 1 (average) common field values.	None	service_name=Customer Comm Management
>	149	149/988 events are grouped because they share 1 (average) common field values.	None	service_name=AWS Lambda
>	5	5/988 events are grouped because they share 1 (average) common field values.	None	service_name=Card Payments
>	4	4/988 events are grouped because they share 1 (average) common field values.	None	service_name=Azure VM
>	2	2/988 events are grouped because they share 1 (average) common field values.	None	service_name=Azure Functions
>	2	2/988 events are grouped because they share 1 (average) common field values.	None	service_name>Contact Centre

And we're done!

Summary

Tools in Your ML Toolkit Now



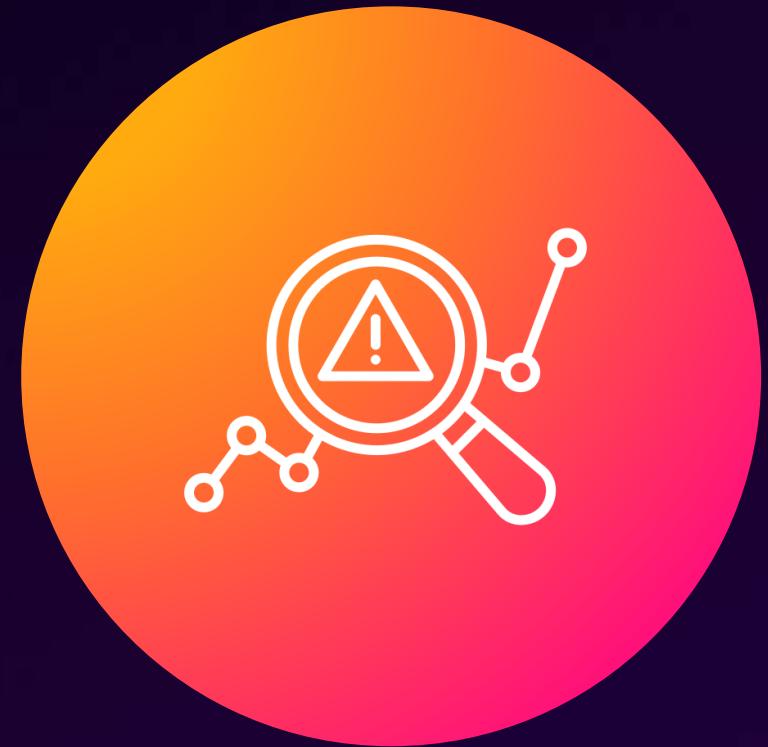
Prediction

Get ahead of issues that may happen in the future



Categorization

Uncover insights about your data to quickly respond in the present



Outlier Detection

Identify and analyze abnormal behavior in your data

Additional Resources

Getting started

- View some of our [webinars](#)
- Check out our YouTube [playlist](#)
- Check out the blog on [MLTK 5.4 release](#)
- Try out some of our starter blogs, such as [Cyclical Statistical Forecasts and Anomalies, part 1](#)
- Try our new [MLTK Deep Dives](#)

Increasing complexity

- Try [part 4 or 6](#) of the Cyclical Statistical Forecasts and Anomalies series
- Brush up on how MLTK works with our [comprehensive documentation](#)
- Get familiar with the [Workshop Guide](#)

More advanced

- The [Analytics and Data Science course](#)
- Try out the [Anomalies Are Like a Gallon of Neapolitan Ice Cream - Part 1](#)
- Try out [part 5](#) of the cyclical statistical forecasts and anomalies series
- Try the [ML-SPL API](#)

Thank you!

Q&A