

Fast Time to Extraordinary Value

Quickly Add Powerful Machine Learning to Your Splunk Apps & Dashboards with Splunk's New Machine Learning Toolkit

Mike Cormier and Bill Thackrey
Co-Founders, Scienta Analytics

.conf2016

splunk >

Fast Time to Extraordinary Value

Or... How Splunk's Machine Learning Toolkit Helped Us Accelerate Our Time-to-Market

Mike Cormier and Bill Thackrey
Co-Founders, Scienta Analytics

.conf2016

splunk >

Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Why Machine Learning?



.conf2016

What is Machine Learning?

“The field of study that gives computers the ability to learn without being explicitly programmed.”

Arthur Samuel – MIT, IBM, Stanford
Author of Computer Checkers at IBM in 1959

“The natural evolution of machine learning, Cognitive Computing attempts to imbue in computer systems, the same insight and understanding we see in humans.”

Earl Cox
Chief Scientist, Scianta Analytics

Why Machine Learning?

Employ learning methods and algorithms to perform statistical and behavioral analysis on large amounts of data.

Encapsulate the knowledge of subject matter experts in the analysis of large amounts of data to solve specific business use cases.

Clustering

Structured Prediction

Reinforced Learning

Anomaly Detection

Dimensionality Reduction

Key Performance Indicators

Fraud

Malware

IP Theft

Application Performance

Identity Theft

Capacity Management

Bad Actors

Logistics

Network Intrusion

Collusion

Exfiltration

Compromised Credentials

Sales Performance

SCADA Security

Cyber-attacks

Hardware Deterioration

Advertising Optimization

Key Performance Indicators

Fraud

Malware

IP Theft

Application Performance

Identity Theft

Capacity Management

Bad Actors

Logistics

BEHAVIOR

Network Intrusion

Collusion

Exfiltration

Compromised Credentials

Sales Performance

SCADA Security

Cyber-attacks

Hardware Deterioration

Advertising Optimization

COGNITIVE BEHAVIOR MODELING FOR SPLUNK



SCIANTA ANALYTICS

**COGNITIVE
MODELER™**
F R A M E W O R K

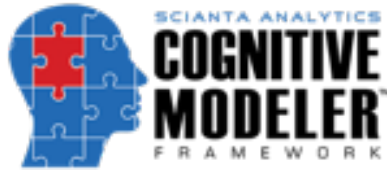


ONE POWERFUL COGNITIVE COMPUTING PLATFORM

FIVE TARGETED APPS

TIGHTLY INTEGRATED INTO SPLUNK

COGNITIVE
MODELER™
FRAMEWORK



ANTI-FRAUD



ITOA



ACTOR BEHAVIOR



SCADA / IOT



EXPLORATION

CHALLENGE

Time to Market

CHALLENGE

Time to Market



SCIANTA'S DESIGN MANDATES

Solution (Use Case) Focus

Solve real business problems
Intuitive workflows based on user role

Leverage Scianta Analytics Proprietary Technology

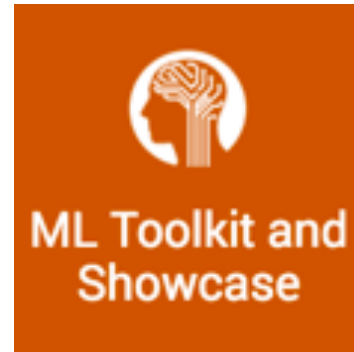
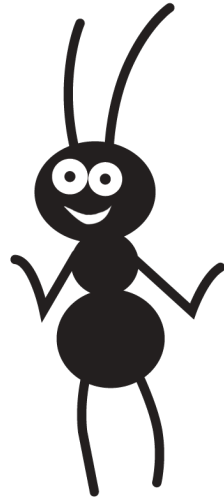
Cognitive (Human-Centered) Computing
Semantic Reasoning
Fuzzy-Logic Empowered Machine Learning



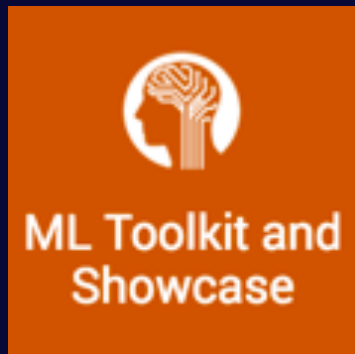
Exploit and Extend the Splunk Platform

Do not use Splunk as an ETL
Live within, honor and extend the Splunk ecosystem
Tightly integrate with Splunk by extending SPL

SCIANTA'S
“SECRET WEAPON”
FOR
FAST TIME-TO MARKET



SCIANTA'S “SECRET WEAPON” FOR FAST TIME-TO MARKET

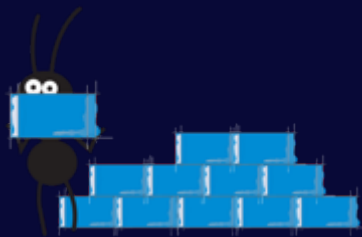


The new

Machine Learning Toolkit

from Splunk's Machine Learning Team

*Extend Scianta's cognitive computing team
with the expertise of Splunk's machine learning team.*



Some Context

A Very Brief Elevator Pitch on the “SCM” Product Suite

.conf2016

splunk >

Some Context

The SCM Platform Architecture

.conf2016

splunk >

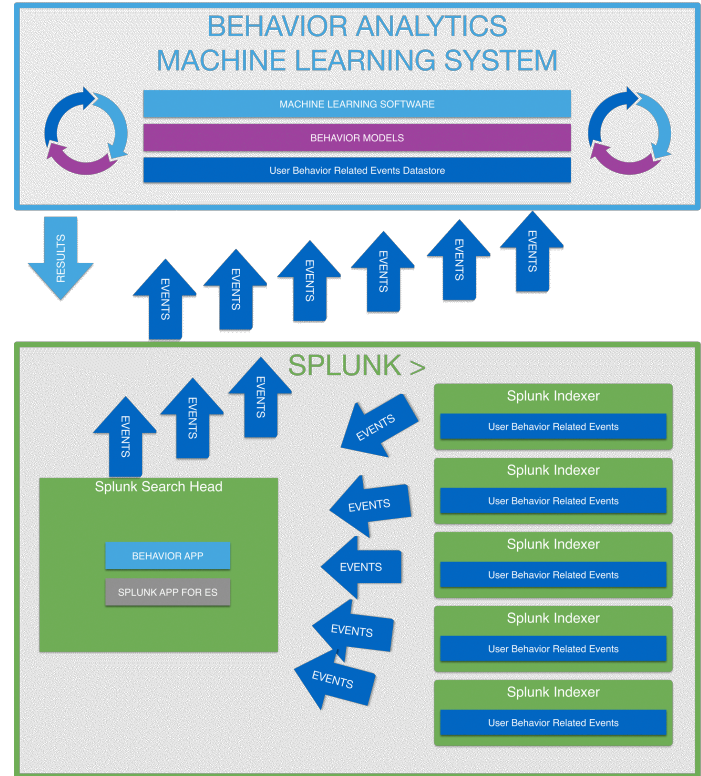
Traditional Machine Learning

How “They” Do It

Splunk As a Datastore

Events Exported for Analysis (ETL)
Results Displayed Externally or
Returned to Splunk

Large Volumes of Data in Motion

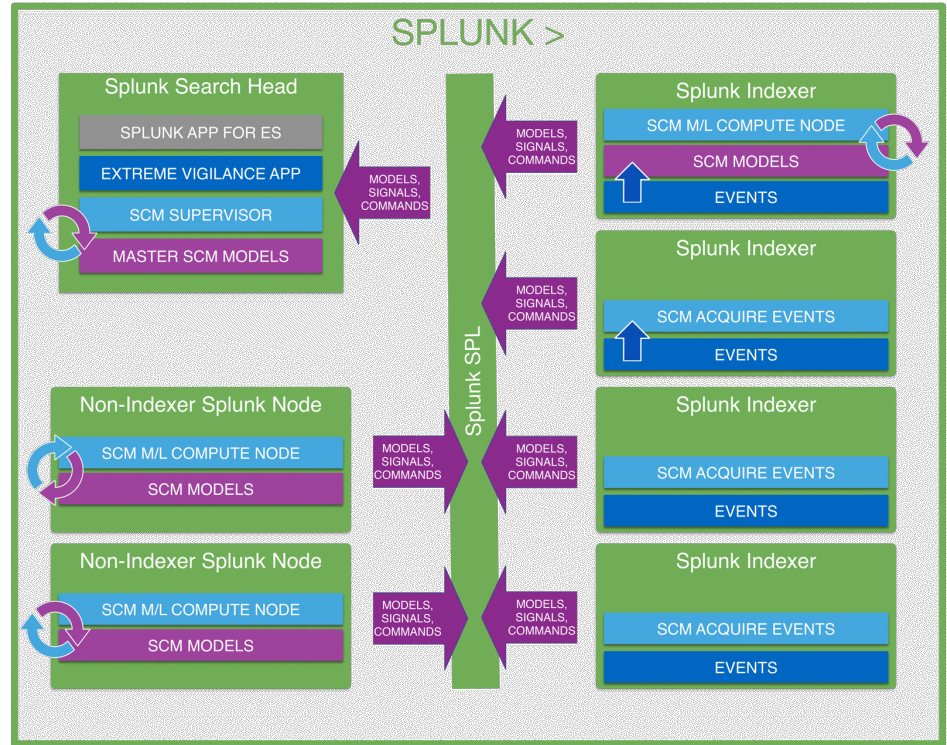
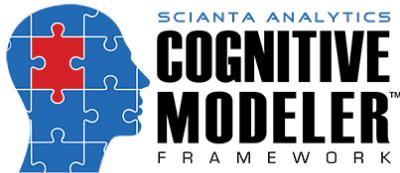


Scianta Machine Learning

How We Do It

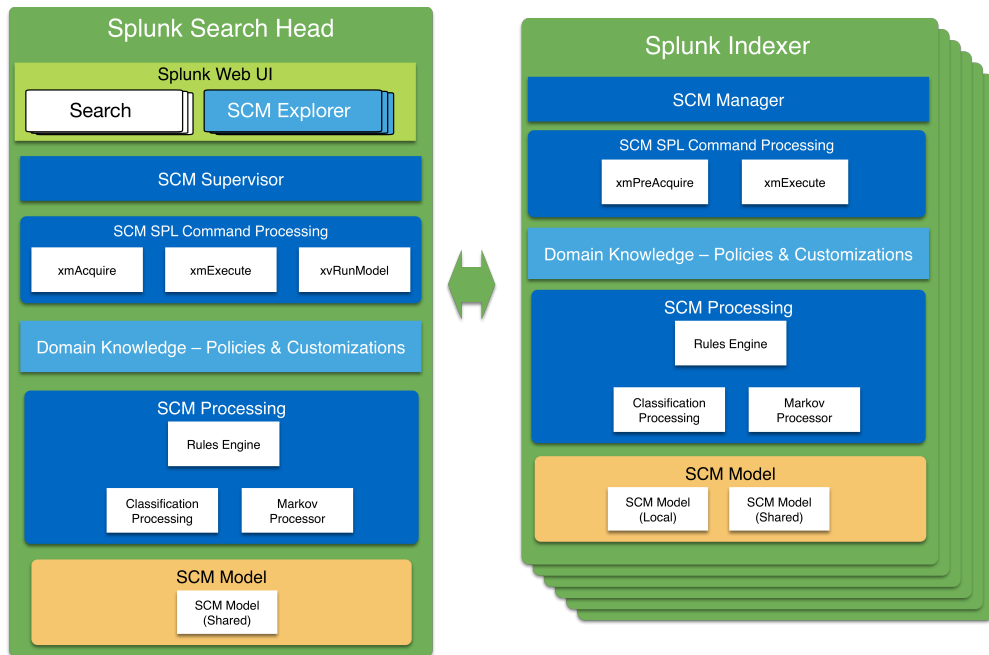
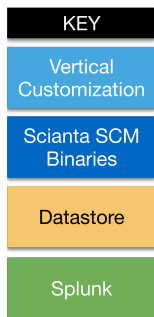
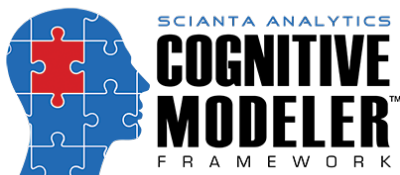
Splunk Empowered As a Powerful Cognitive Computing Engine
Behavior Models Built & Managed within Splunk Indexers

Data Analyzed in Place



Human-centered knowledge discovery & Cognitive modeling tightly integrated into splunk

Splunk Empowered As a Powerful Cognitive Computing Engine
 Scienta Cognitive Modeler Tightly Integrated Into Splunk

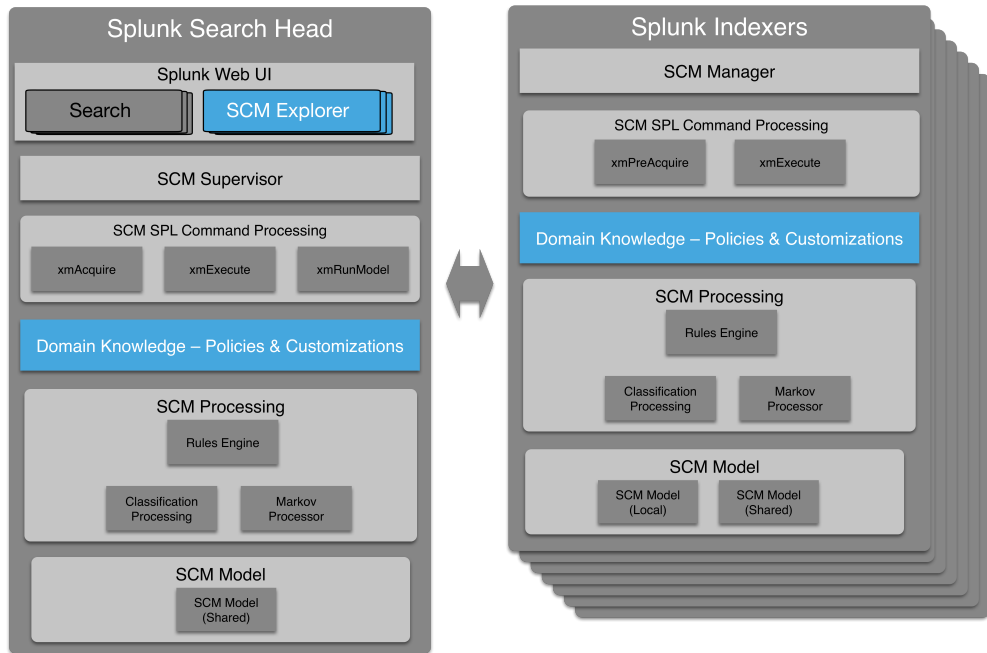
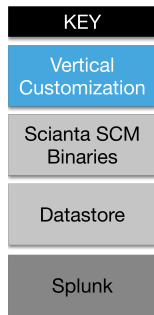
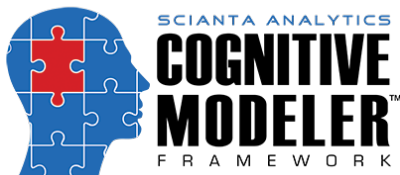


Human-centered knowledge discovery & Cognitive modeling tightly integrated into splunk

Splunk Empowered As a Powerful Cognitive Computing Engine

Use Case

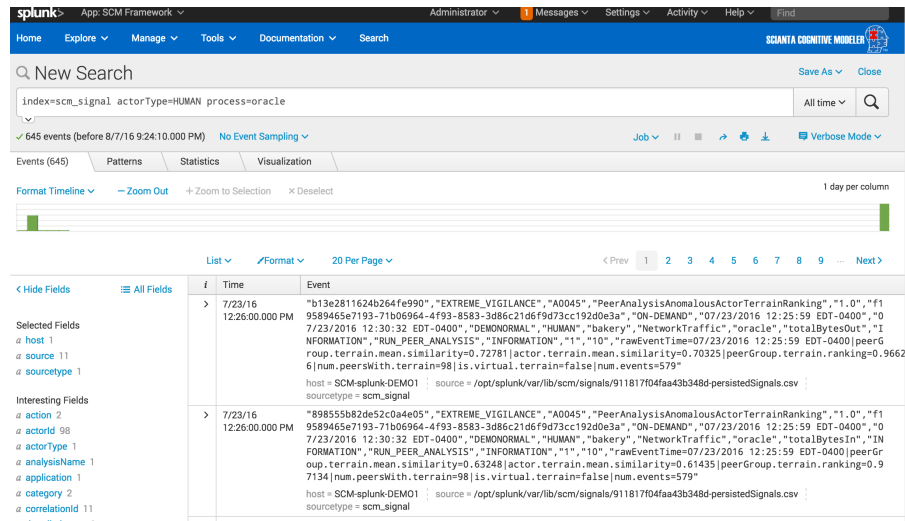
Verticalization Layer



Cognitive Signals™

Scianta's Cognitive Computing Language

- Capture Outcome of Cognitive Processes
- Persistent Cognitive Knowledge Objects
 - anomalies
 - hazards
 - threats
 - risk
 - introspection
- Stored in Splunk Indexes
- Used and Generated by Scianta's Multi-Pass Cognitive Analysis
- Available for Use by Other Apps Including Splunk ML Toolkit



The screenshot shows a Splunk search interface. The search query is `index=scm_signal actorType=HUMAN process=oracle`. The results show 645 events. The interface includes a search bar, a timeline view, and a list view. The list view shows two events with the following details:

i	Time	Event
>	7/23/16 12:26:00.000 PM	"b13e2811624b264fe990", "EXTREME_VIGILANCE", "A0045", "PeerAnalysisAnomalousActorTerrainRanking", "1.0", "f1958946e7193-71b06964-4f93-8583-3d86c21d6f9d73cc1920e3a", "ON-DEMAND", "07/23/2016 12:25:59 EDT-0400", "07/23/2016 12:30:32 EDT-0400", "DEMONORMAL", "HUMAN", "bakery", "NetworkTraffic", "oracle", "totalBytesIn", "INFORMATION", "RUN_PEER_ANALYSIS", "INFORMATION", "1", "10", "rawEventTime=07/23/2016 12:25:59 EDT-0400 peerGroup.terrain.mean.similarity=0.72781 actor.terrain.mean.similarity=0.70325 peerGroup.terrain.ranking=0.96626 num.peersWith.terrain=98 is.virtual.terrain=false num.events=579" host = SCM-splunk-DEMO1 source = /opt/splunk/var/lib/scm/signals/911817f04faa43b348d-persistedSignals.csv sourcetype = scm_signal
>	7/23/16 12:26:00.000 PM	"89855b82de52c0a4e05", "EXTREME_VIGILANCE", "A0045", "PeerAnalysisAnomalousActorTerrainRanking", "1.0", "f1958946e7193-71b06964-4f93-8583-3d86c21d6f9d73cc1920e3a", "ON-DEMAND", "07/23/2016 12:25:59 EDT-0400", "07/23/2016 12:30:32 EDT-0400", "DEMONORMAL", "HUMAN", "bakery", "NetworkTraffic", "oracle", "totalBytesIn", "INFORMATION", "RUN_PEER_ANALYSIS", "INFORMATION", "1", "10", "rawEventTime=07/23/2016 12:25:59 EDT-0400 peerGroup.terrain.mean.similarity=0.63248 actor.terrain.mean.similarity=0.61435 peerGroup.terrain.ranking=0.97134 num.peersWith.terrain=98 is.virtual.terrain=false num.events=579" host = SCM-splunk-DEMO1 source = /opt/splunk/var/lib/scm/signals/911817f04faa43b348d-persistedSignals.csv sourcetype = scm_signal

Today We'll be Analyzing Signals with the Splunk ML Toolkit

A Quick Demo Extreme Vigilance Anti-Fraud

Live Demo

.conf2016

splunk >

How Scianta Leverages the ML Toolkit In the SCM Framework



.conf2016

The machine learning toolkit & showcase

<https://splunkbase.splunk.com/app/2890/>
Prereq: Python SciKit – 200 ML algorithms

Why would scianta use splunk's ML toolkit?

Scianta Analytics is a Machine Learning Company

Why Would We Use Splunk's ML Toolkit?

Remember This Slide...?

Scianta's design mandates

Solution (Use Case) Focus

Solve real business problems
Intuitive workflows based on user role

Leverage Scianta Analytics Proprietary Technology

Cognitive (Human-Centered) Computing
Semantic Reasoning
Fuzzy-Logic Empowered Machine Learning



Exploit and Extend the Splunk Platform

Do not use Splunk as an ETL
Live within, honor and extend the Splunk ecosystem
Tightly integrate with Splunk by extending SPL

Why would scianta use splunk's ML toolkit?

We Use Scianta Analytics Proprietary
Cognitive Computing Technology
Where it Adds Value

We Use the Splunk ML Toolkit
To Add Powerful ML Functionality
to Enhance the SCM Feature Set

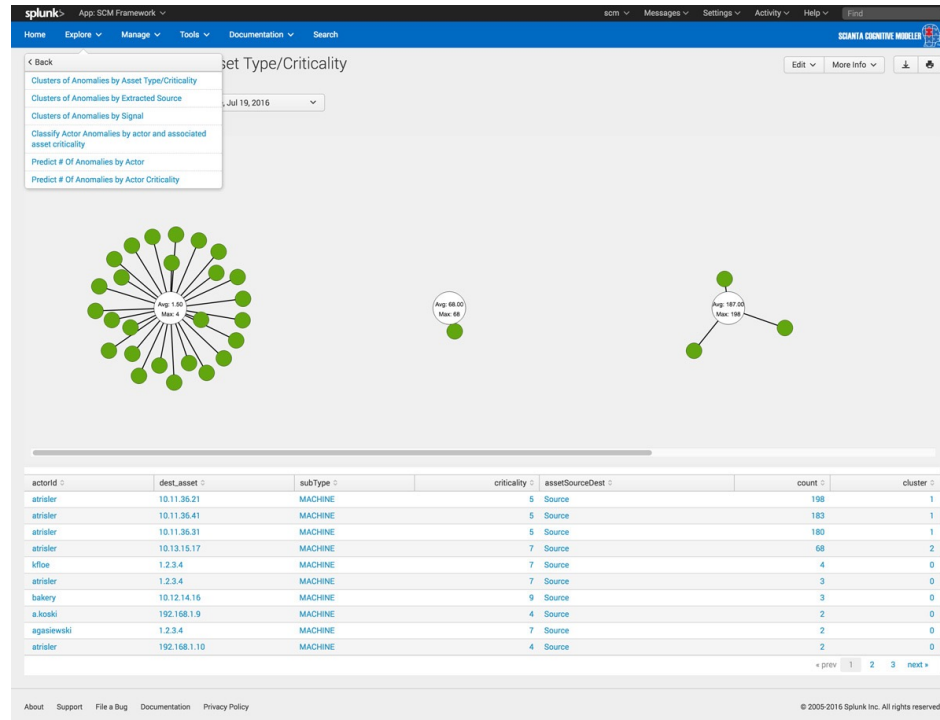
Analyzing the Performance of the SCM Platform

Analyzing the Performance of the SCM Platform

The image shows a screenshot of the Splunk SCM Framework dashboard. The dashboard has a top navigation bar with 'splunk' and 'App: SCM Framework' on the left, and 'Administrator', 'Messages', 'Settings', 'Activity', and 'Help' on the right. Below the navigation bar is a 'Home' section with a dropdown menu containing 'Explore Actions', 'Explore Landscapes', 'Explore Models', 'Explore Signals', 'Explore Terrains', and 'Explore Thresholds'. The main area of the dashboard contains seven dashboard tiles: 'Actors', 'Assets', 'Resource Mapper', 'Categories', 'Data Dictionaries', 'Models', and 'Properties'. Overlaid on the center of the dashboard is the text 'Clustering', 'Classification', and 'Prediction' in large blue font. At the bottom left of the dashboard, there is a footer with the text: 'Custom App Development for Splunk by Concanon LLC' and 'Copyright 2013-2016 Scianta Analytics LLC All Rights Reserved'.

Analyzing the Performance of the SCM Platform

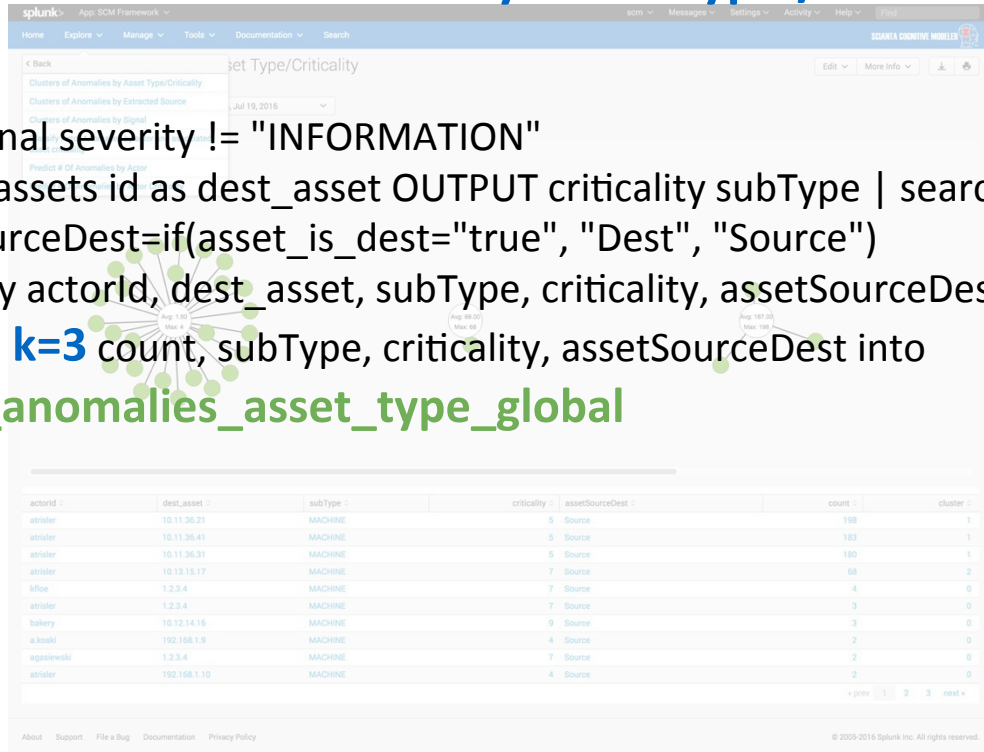
Clusters of Anomalies by Asset Type / Criticality



Analyzing the Performance of the SCM Platform

Clusters of Anomalies by Asset Type / Criticality

```
index=scm_signal severity != "INFORMATION"  
| lookup scm_assets id as dest_asset OUTPUT criticality subType | search subType="*" |  
| eval assetSourceDest=if(asset_is_dest="true", "Dest", "Source") |  
| stats count by actorId, dest_asset, subType, criticality, assetSourceDest  
| fit KMeans k=3 count, subType, criticality, assetSourceDest into  
xm_cluster_anomalies_asset_type_global  
| sort -count
```



Analyzing the Performance of the SCM Platform

Classify Actor Anomalies by Actor and Associated Asset Criticality

The screenshot displays a Splunk dashboard titled "Classify Actor Anomalies by actor and associated asset criticality". The interface includes a navigation bar with "Home", "Explore", "Manage", "Tools", "Documentation", and "Search". A search bar at the top right contains the text "scm". Below the navigation bar, the dashboard title is followed by "Edit" and "More Info" buttons. A "Time Range" dropdown menu is set to "Last 7 days".

The main content area features two side-by-side tables. The left table, titled "Actors and Anomaly Count", lists actor IDs and their predicted anomaly counts. The right table, titled "Actor/Asset Criticality and Anomaly Count", lists actor IDs, their asset criticality, and their predicted anomaly counts.

Actors and Anomaly Count		Actor/Asset Criticality and Anomaly Count		
actorid	Predicted	actor_criticality	asset_criticality	Predicted
cpiening	562	2	5	562
glargin	562	6	7	125
nhenderosn	562	2	7	71
gclough	562	3	5	10
ejennifer	125	5	7	5
gamer	125	3	4	4
a.koski	10	2	4	4
alex	10	4	5	3
dmsys	10	9	9	3
goose	10	7	5	3

Navigation controls for both tables include "prev", "1", "2", "3", "4", and "next" buttons.

Footer text includes "About", "Support", "File a Bug", "Documentation", "Privacy Policy", and "© 2005-2016 Splunk Inc. All rights reserved."

Analyzing the Performance of the SCM Platform

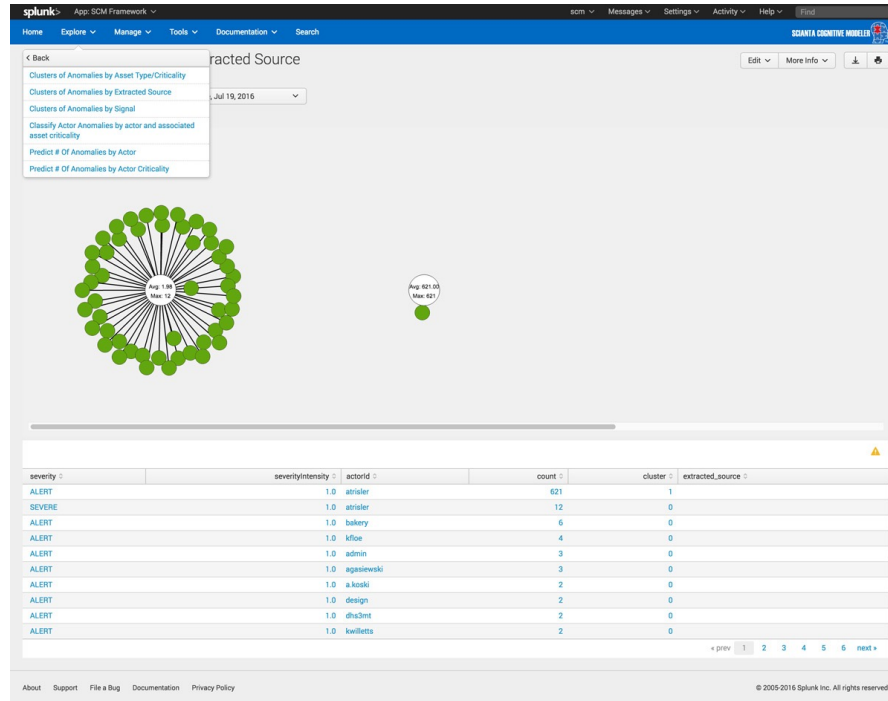
Classify Actor Anomalies by Actor and Associated Asset Criticality

```
index=scm_signal severity != "INFORMATION" extracted_source="THRESHOLD"  
OR extracted_source="SEQUENCE"  
| lookup scm_actors id as actorId OUTPUT criticality as actor_criticality  
| lookup scm_assets id as dest_asset OUTPUT criticality as asset_criticality  
| bin _time span=1d | stats count as count by _time, actor_criticality, asset_criticality  
| fit RandomForestClassifier count from actor_criticality, asset_criticality into  
xm_classify_actor_anomaly_criticality
```

actorId	actor_criticality	dest_asset	asset_criticality	count
glargin	562	562	7	125
profifer	125	5	7	5
a_loski	10	4	4	4
chayrs	10	4	5	3
goose	10	9	9	3
goose	10	7	5	3

Analyzing the Performance of the SCM Platform

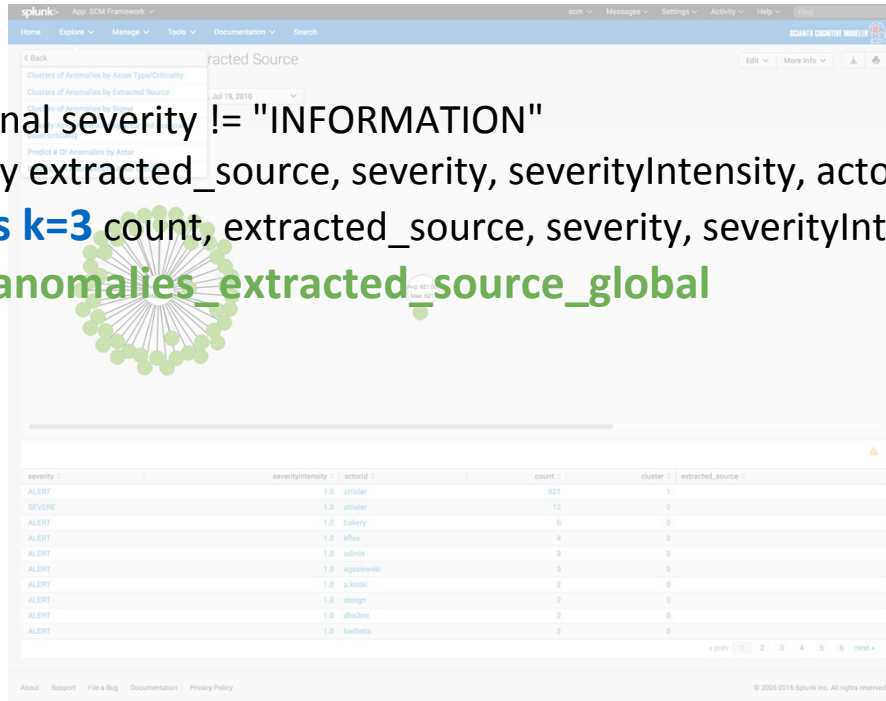
Clusters of Anomalies by Extracted Source



Analyzing the Performance of the SCM Platform

Clusters of Anomalies by Extracted Source

index=scm_signal severity != "INFORMATION"
| stats count by extracted_source, severity, severityIntensity, actorId
| **fit KMeans k=3** count, extracted_source, severity, severityIntensity into
xm_cluster_anomalies_extracted_source_global
| sort -count



What We've Discovered

Tips and Tricks for Using the ML Toolkit

.conf2016

splunk>

ML toolkit tips & tricks

Beware Namespaces

All models live in the ML Toolkit namespace.

No namespace or model management.

Scianta uses Tokens & Macros to Manage MLT Namespace.

Scaling

Distributed Fit and Apply??

ML toolkit tips & tricks

Categoricals

Cannot cluster around strings: Must assign a numeric value.

Categoricals are capped at 100 to keep number of dimensions manageable.

No Model Management

Scianta Uses Saved Searches to Keep Models Updated

Let's Build a ML Dashboard from Scratch

Live Demo

.conf2016

splunk >



More Information...

Visit us at the Concanon booth!

Splunk ML Toolkit Examples www.sciantaanalytics.com/conftalk

SCM and Extreme Vigilance® www.sciantaanalytics.com/xv

**Business Analytics
and ML Consulting**

www.concanon.com
CONCANONTM
INSIGHT ON DEMAND

THANK YOU

.conf2016



END



.conf2016

splunk >