splunk> .conf2017

# What's new in Machine Learning across the Splunk Portfolio

Manish Sainani | Director, Product Management

Bob Pratt | Sr. Director, Product Management

September 2017

# Forward-Looking Statements

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 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.



splunk> .conf2017

# Agenda

- Machine Learning Overview
- Splunk Machine Learning Toolkit (MLTK) Overview
- What's New in Machine Learning Toolkit?
- What's new in IT Service Intelligence ML?
- Splunk User Behavior Analytics (UBA) Overview
- What's new in UBA 4.0?

# Machine Learning Overview

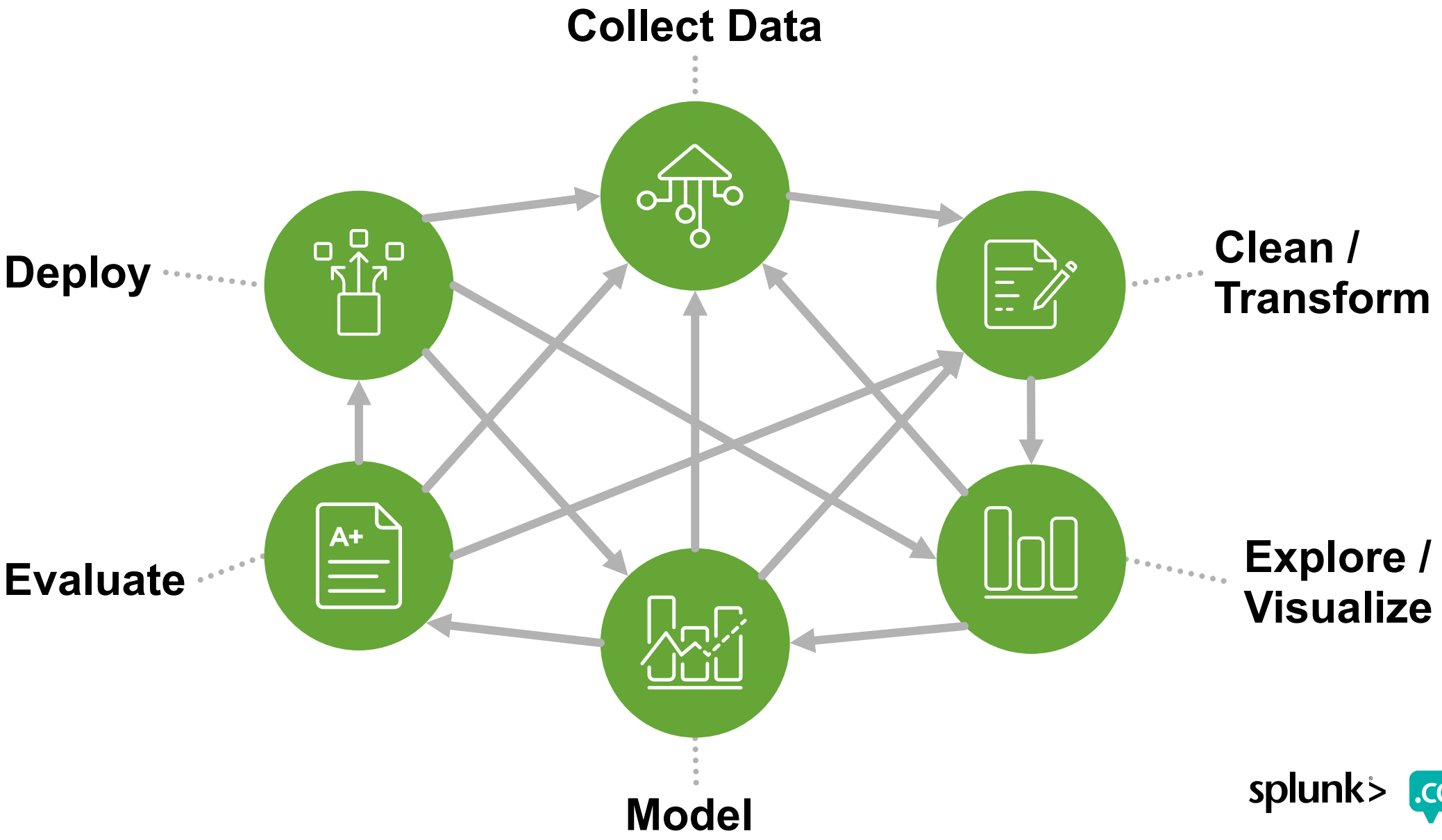
splunk> .conf2017

# Machine Learning

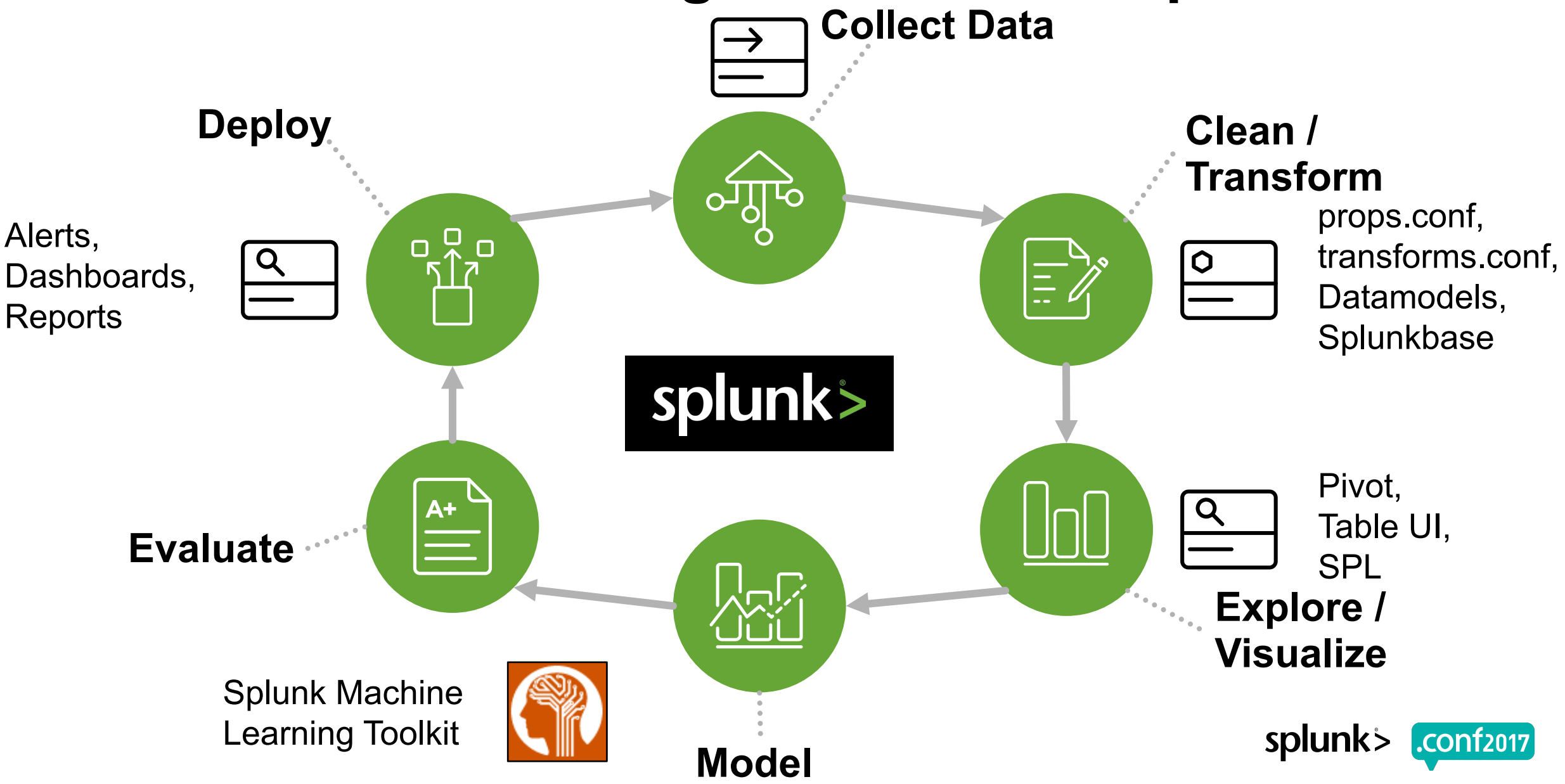
- A process for generalizing from examples
- Examples
  - A, B, … → # (regression)
  - A, B, ... → a (classification)
  - $X_{past} \rightarrow X_{future}$ (forecasting)
  - like with like (clustering)
  - $|X_{predicted} - X_{actual}| >> 0$ (anomaly detection)

# How Machine Learning is surfaced across the Splunk Portfolio

# Machine Learning Process



# Machine Learning Process with Splunk




splunk> .conf2017

# Splunk Machine Learning Toolkit

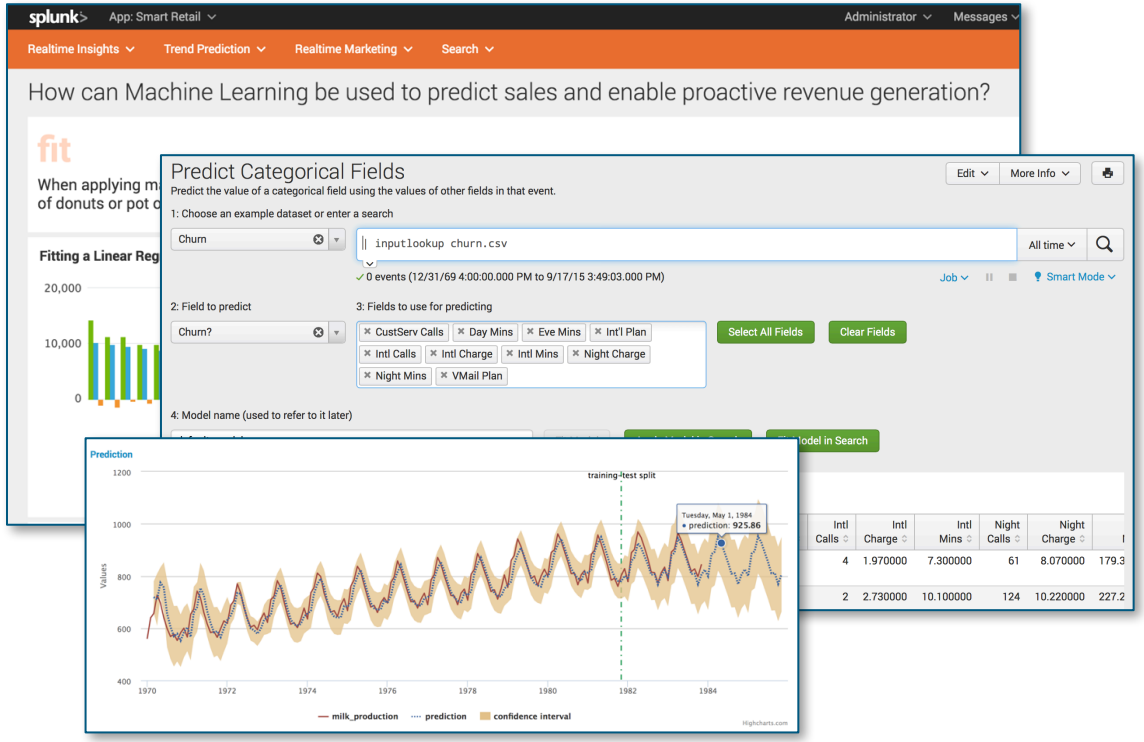
platform extensions and guided modeling dashboards

splunk> .conf2017

# Splunk Machine Learning Toolkit

## extends Splunk with new tools and guided modeling

- **Assistants:** Guide model building, testing, & deployment for common tasks
- **Showcase:** 25+ interactive examples from IT, security, business, and IoT
- **Algorithms:** 30 standard algorithms plus an extensibility API
- **SPL ML Commands:** New commands to fit, test, and operationalize models
- **Python for Scientific Computing Library:** 300+ open-source algorithms




splunk> .conf2017

# Machine Learning Toolkit Customer Use Cases

TELUS

Reducing customer service disruption with early identification of difficult-to-detect network incidents

Minimizing cell tower degradation and downtime with improved issue detection sensitivity

---

Zillow

Speeding website problem resolution by automatically ranking actions for support engineers

---

NTT docomo

Ensuring mobile device security by detecting anomalies in ID authentication

---

Entertainment Company

Predicting and averting potential gaming outage conditions with finer-grained detection

Preventing fraud by Identifying malicious accounts and suspicious activities

---

Telco

Improving uptime and lowering costs by predicting/preventing cell tower failures and optimizing repair truck rolls

splunk> .conf2017

# What's New in MLTK?

since last .conf

splunk> .conf2017

# What's New

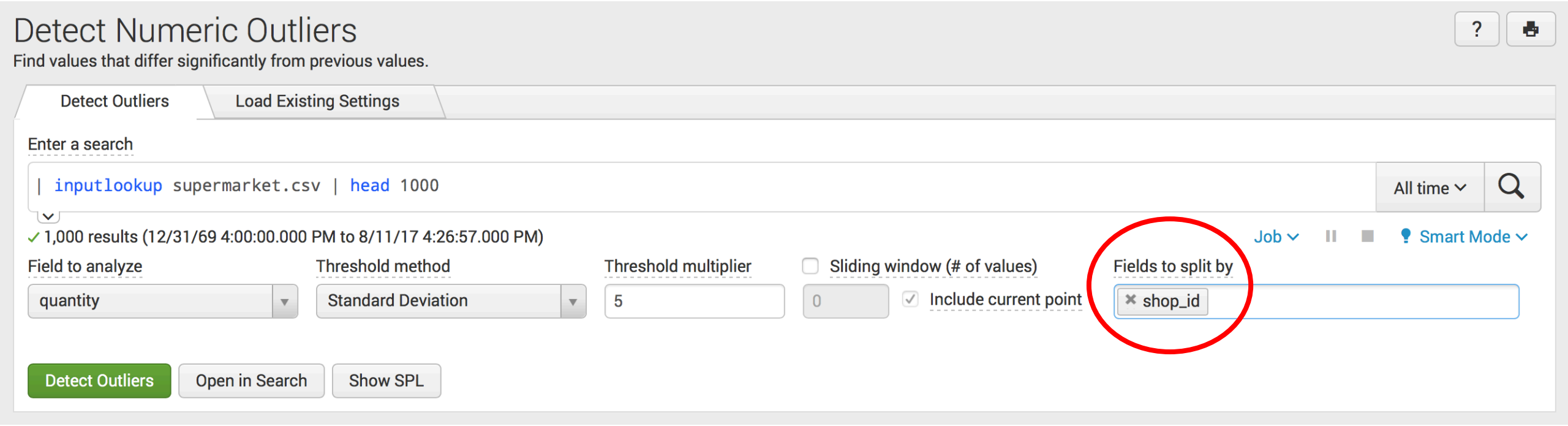
(since .conf 2016)

- **Detect Numeric Outliers improvements**
- **Preprocessing / Data Prep**
- **Model Management**
- **ML-SPL extensibility API**
- **Spark Support (private limited beta)**
- New algorithms:
  - ARIMA supported in Forecasting Time Series Assistant
  - ACF & PACF
  - Gradient Boosting Classifier & Regressor
- Load Existing Settings is per-user
- Downsampled Line Chart supports drilldown

splunk> .conf2017

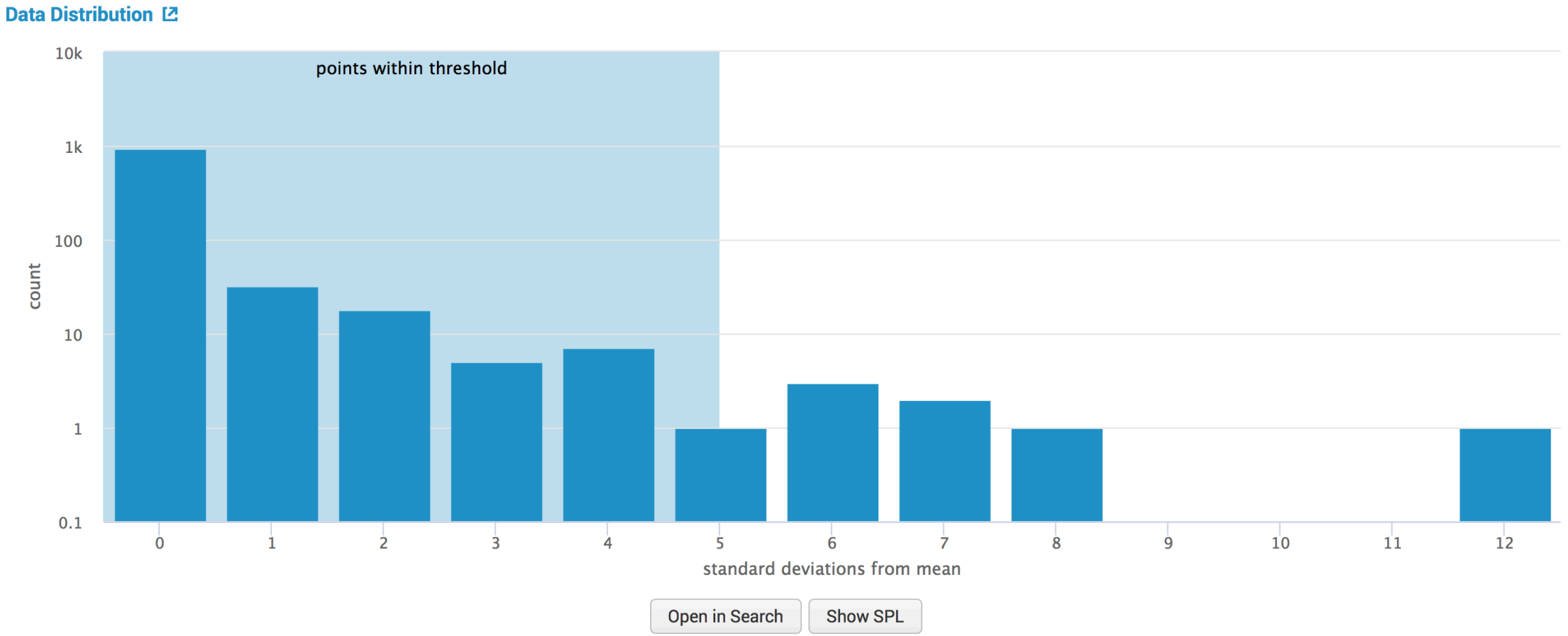
# Detect Numeric Outliers

## split-by support



# Detect Numeric Outliers

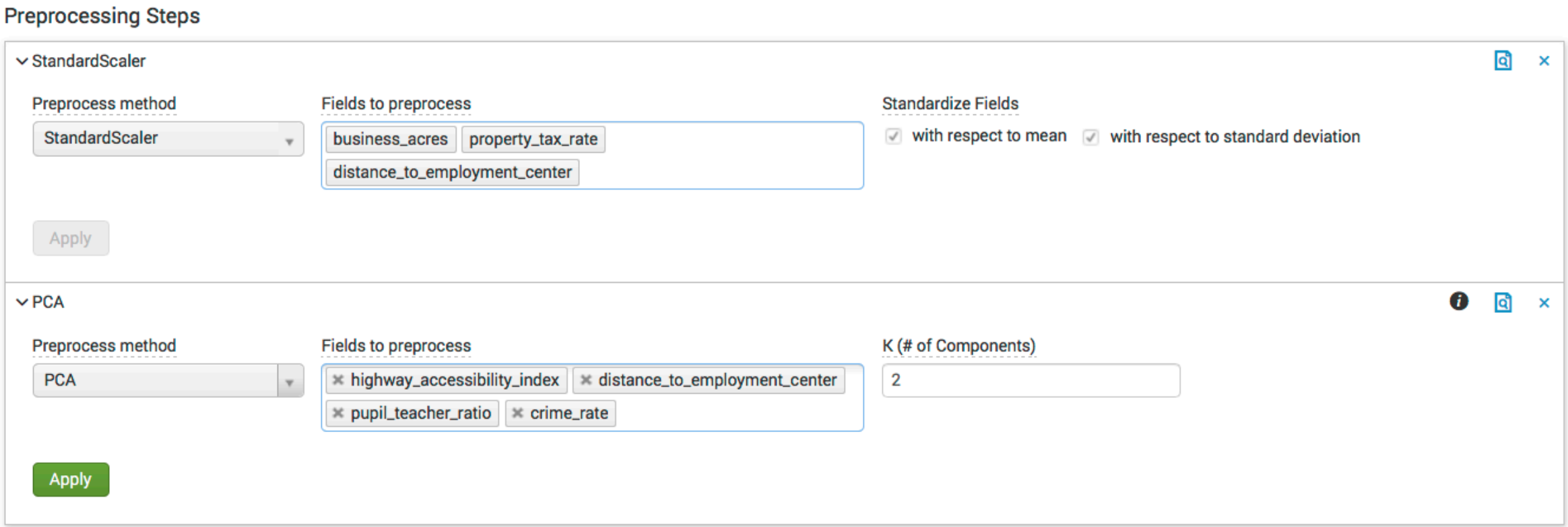
## data distribution viz



# Preprocessing

## build a pipeline of data prep

- In Predict Numeric, Predict Categorical, and Cluster Numeric assistants

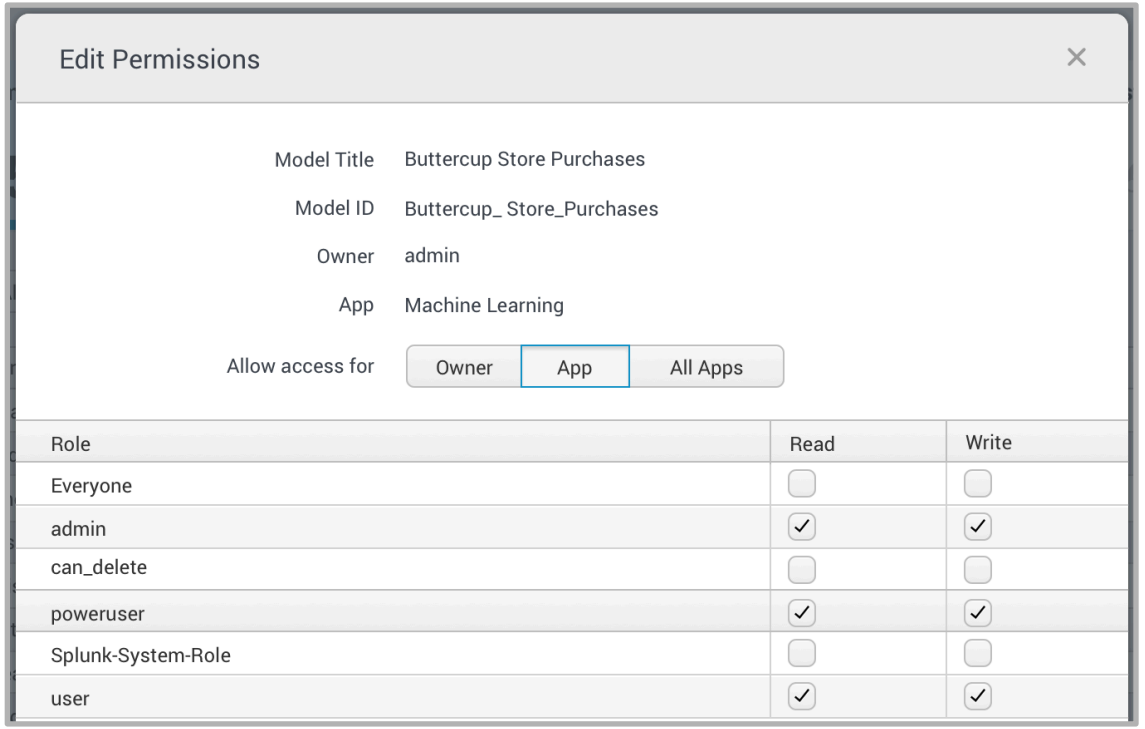



splunk> .conf2017

# Model Management

(coming to MLTK 3.0)

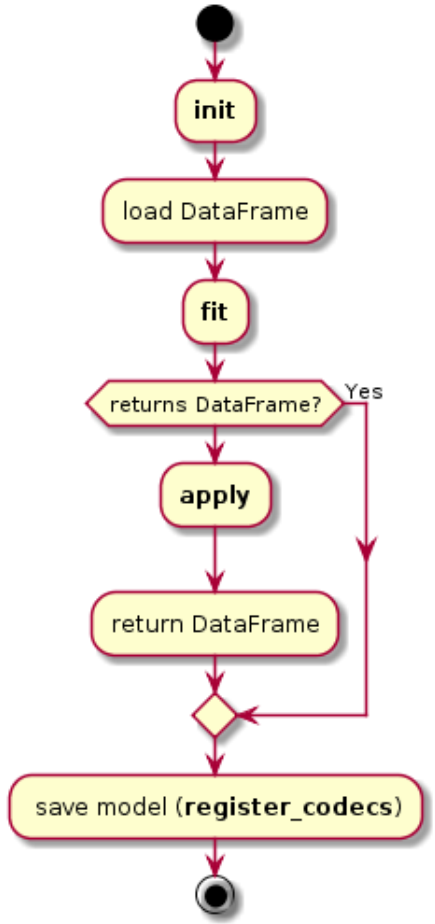
- Provides Role Based Access Control to models
- Assign permissions to models to control who has what level of access
- Manage models via a rich UI interface



# ML-SPL Extensibility API

## featuring: primo documentation

- Make more algos available to **`fit`** / **`apply`**
  - 300+ in PSC
  - Custom algorithms
- Expose new or different parameters
- Docs include examples
  - Correlation Matrix
  - Agglomerative Clustering
  - Support Vector Regressor
  - Savitzky-Golay Filter
- Use in your apps / dashboards / etc.!



# Spark Support

## private beta open now

- Use your existing Spark cluster with MLTK
  - Distributed fit on massive datasets
  - Apply MLlib models for supported algos
- **`sfit` / `sapply`**
- Contact sparkml@splunk.com
  - What is your use case (e.g., predicting server downtime)?
  - Why do you want / need Spark (i.e., why isn't MLTK sufficient)?

# Machine Learning Customer Success

TELUS

Network Optimization
Detect & Prevent Equipment Failure

NTT docomo

Security / Fraud Prevention

Telco

Prevent Cell Tower Failure
Optimize Repair Operations

Zillow®

Prioritize Website Issues
and Predict Root Cause

**Entertainment Company**

Predict Gaming Outages
Fraud Prevention

CONCANON™
INSIGHT ON DEMAND

Machine Learning Consulting Services

SCIANTA ANALYTICS
DEEP INSIGHT™

Analytics App built on ML Toolkit

splunk> .conf2017

# What's new in ITSI ML ?

splunk> .conf2017

# ITSI Smart Mode

## Event Co-relation and Clustering for Event Data

- Uses the Splunk “Reverse Pyramid Clustering” Algorithm to reduce noise in IT event data
- The algorithm extracts categorical and textual similarity from events and uses them in combination with a Service context to correlate events.
- Provides a UI based configuration editor that allows users to tweak parameters and tune configuration without a data scientist
- Not a black box – explainability is built right in. All event groups created by the algorithm provide an explanation as to why events were grouped together.

# Splunk User Behavior Analytics

Machine Learning-based Threat Detection

splunk> .conf2017

# Splunk User Behavior Analytics

An out-of-the-box solution that helps organizations find

Anomalous Behavior

Risky Users

Unknown Threats

with the use of **machine learning**

# Splunk User Behavioral Analytics Pillars

## Five Foundational Pillars

Real-Time & Big Data Architecture

Behavior Baseline & Modelling

Unsupervised Machine Learning

Anomaly Detection

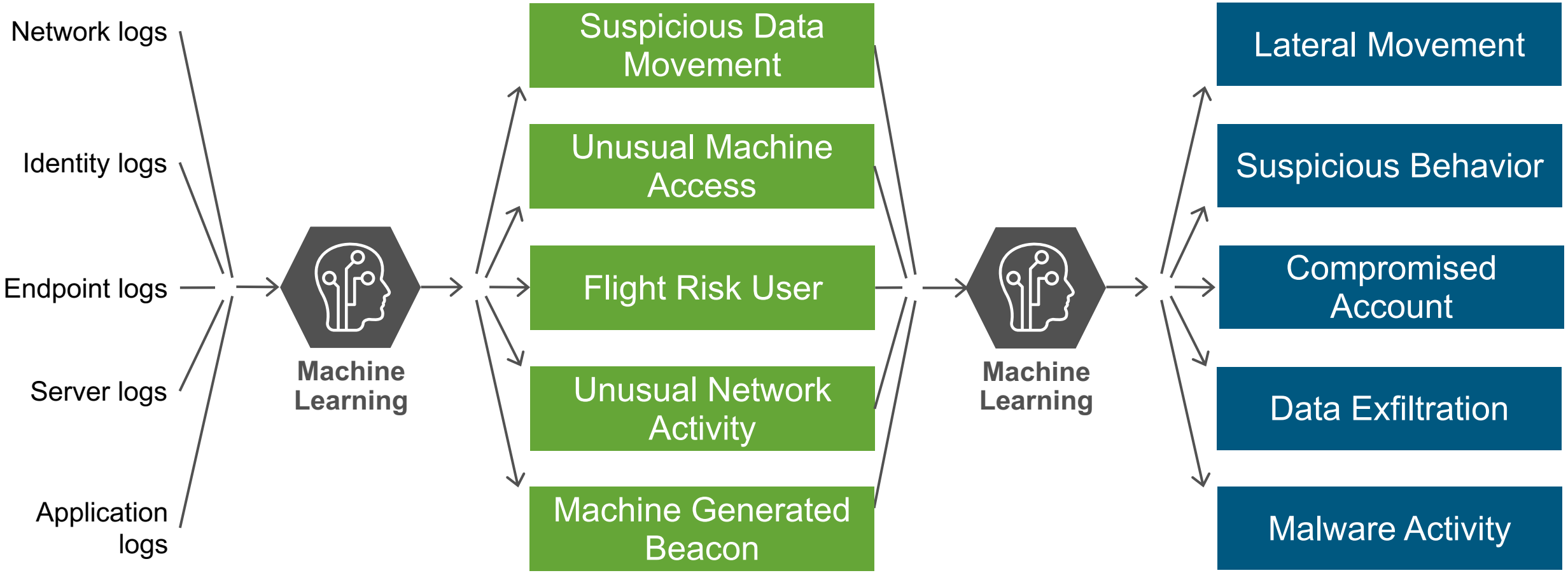
Threat Detection

Splunk User Behavior Analytics™

splunk> Platform for Machine Data

splunk> .conf2017

# How Does Splunk UBA Work?




splunk> .conf2017

# How does UBA integrate with Splunk Enterprise and ES?

*Human-driven*

- Rules
- Correlations
- Statistics

**Splunk Enterprise**
**Investigate**

*ML-driven*

- Machine Learning
- Behavior Analysis
- Risk-level Scoring

**Splunk UBA**
**Detect**

**Known Threats**

**Splunk ES**
**Detect, Investigate & Respond**

**Unknown Threats**

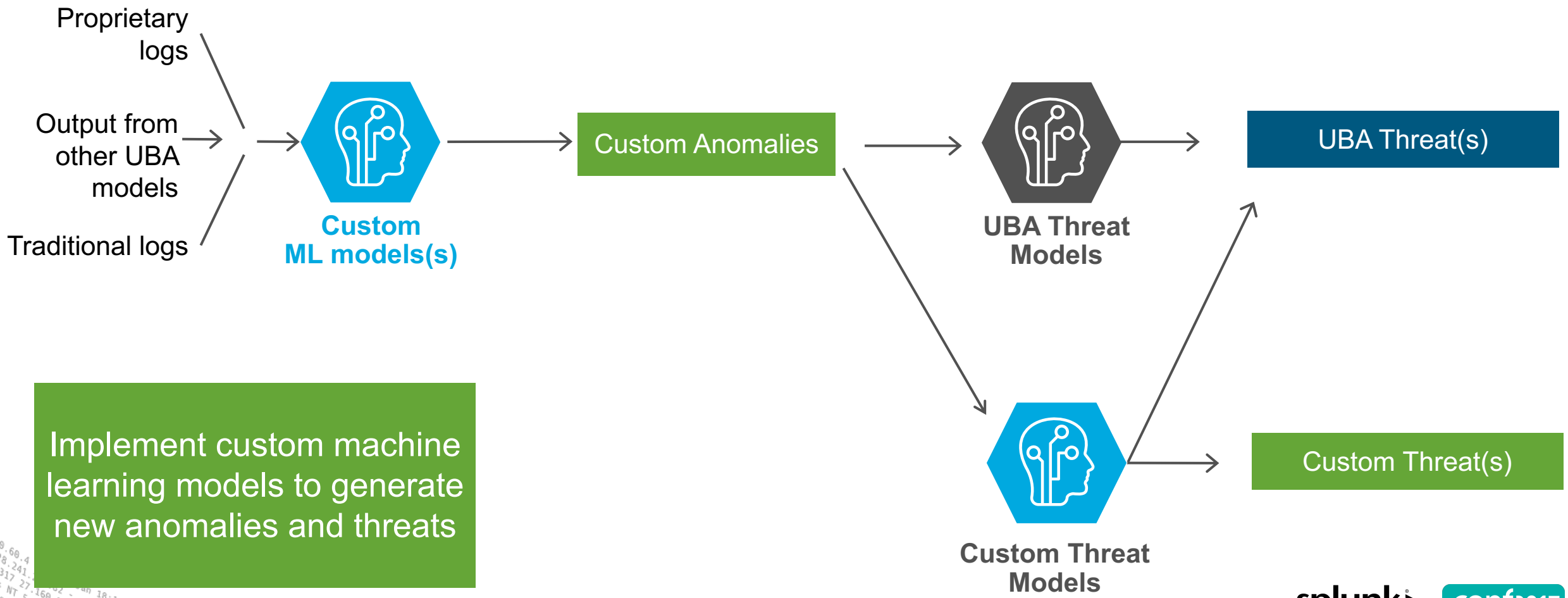
splunk> .conf2017

# What's New in UBA 4.0?

Announced here at .conf

splunk> .conf2017

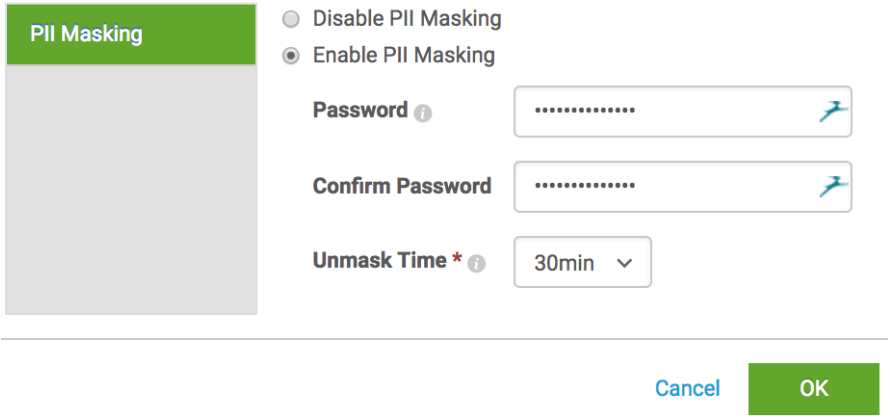
# UBA SDK – now available



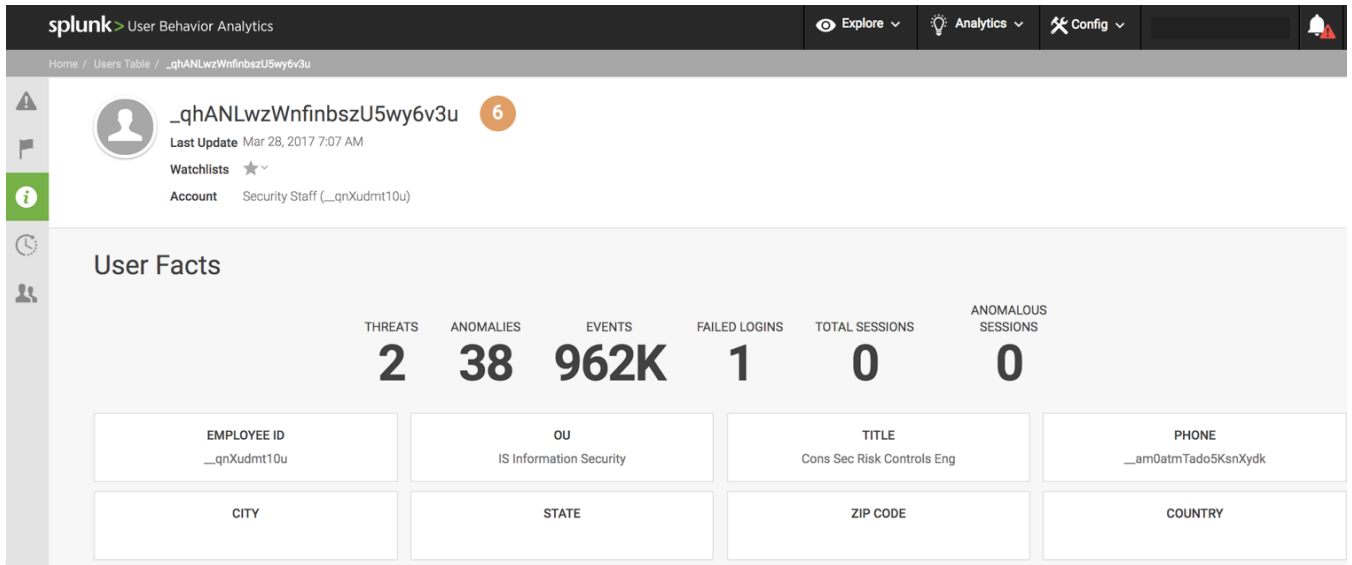

Implement custom machine learning models to generate new anomalies and threats

splunk> .conf2017

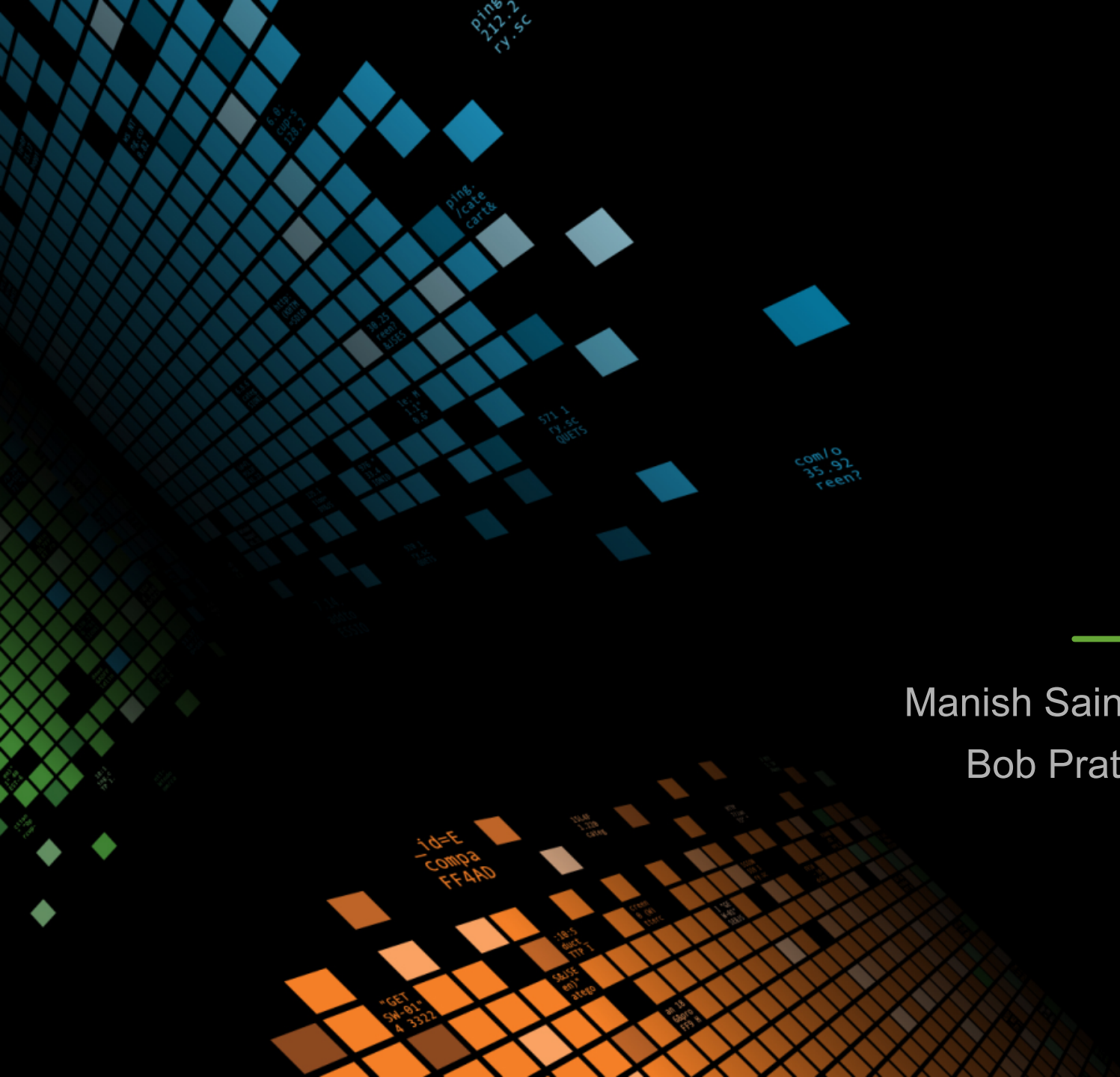
# PII Masking – also shipping now




Obfuscate user details during investigation or hunting




splunk> .conf2017



# Q&A

Manish Sainani | Director, Product Management

Bob Pratt | Sr. Director, Product Management

splunk> .conf2017