
# Splunking for Fraud: Let the Machines Look for Unknown Unknowns

Matthew J Joseff, CFE | Minister of Reality, Sr. Security Specialist

September 6, 2018 | v3.0

# 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> .conf18

# Hello World. Pleased To Meet You.

splunk>

Matthew Joseff, CFE

mjoseff@splunk.com

## 3 years @Splunk

- Over three decades in technology
- Sysadmin, Security, Fraud

## Based in Central CT

- Raised overseas in the IC
- Relocating to Tokyo

## Training

- Psychology/Human behavior
- Product Management
- Certified Fraud Examiner

## Hobbies

- Breaking & fixing things
- Film
- Politics

Time is our only non-renewable resource.

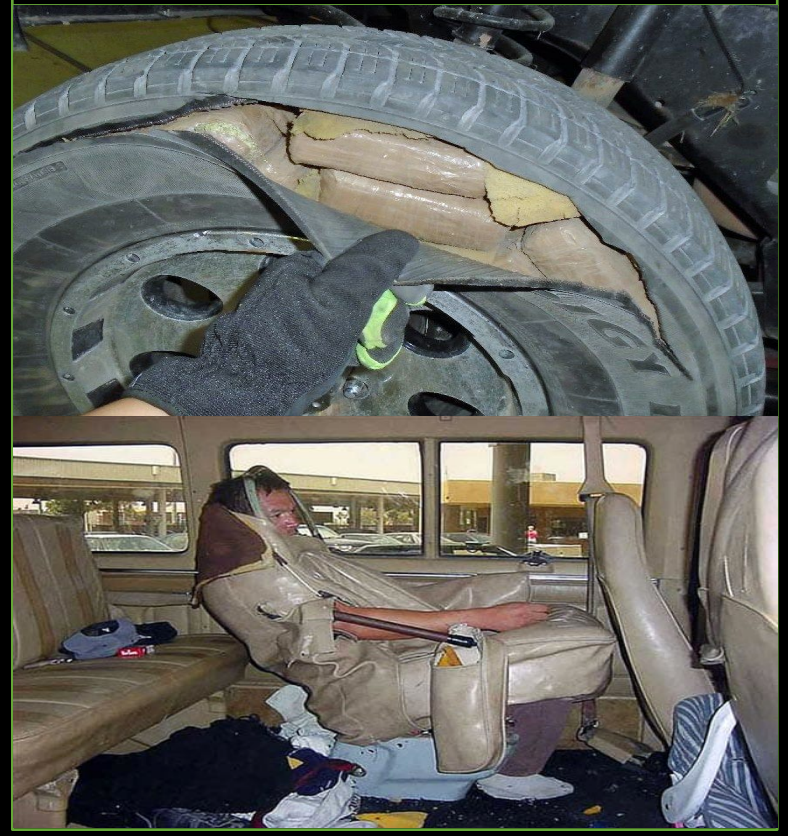
# Compliance, Security, Fraud

## What's the difference?

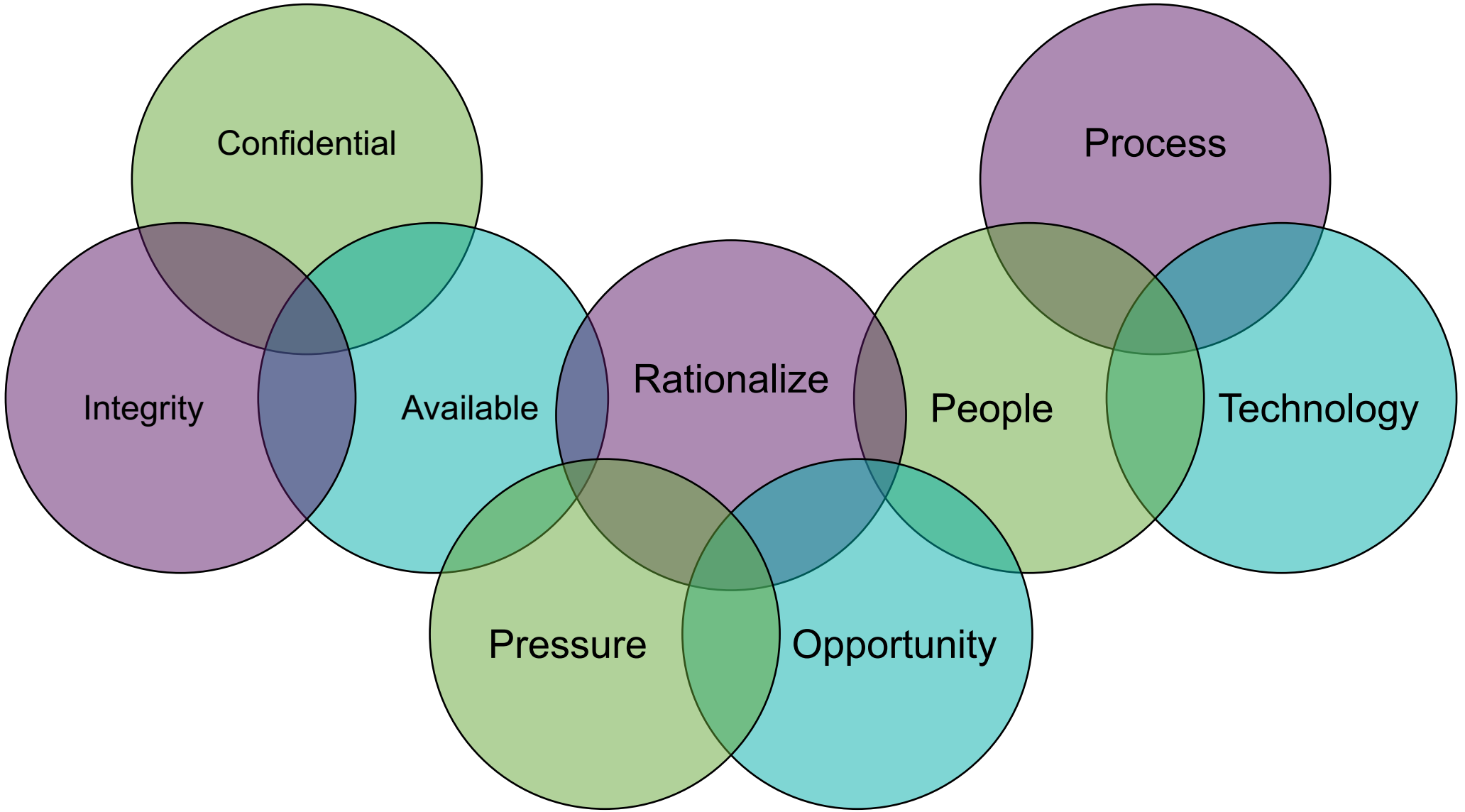
### Compliance

### Security

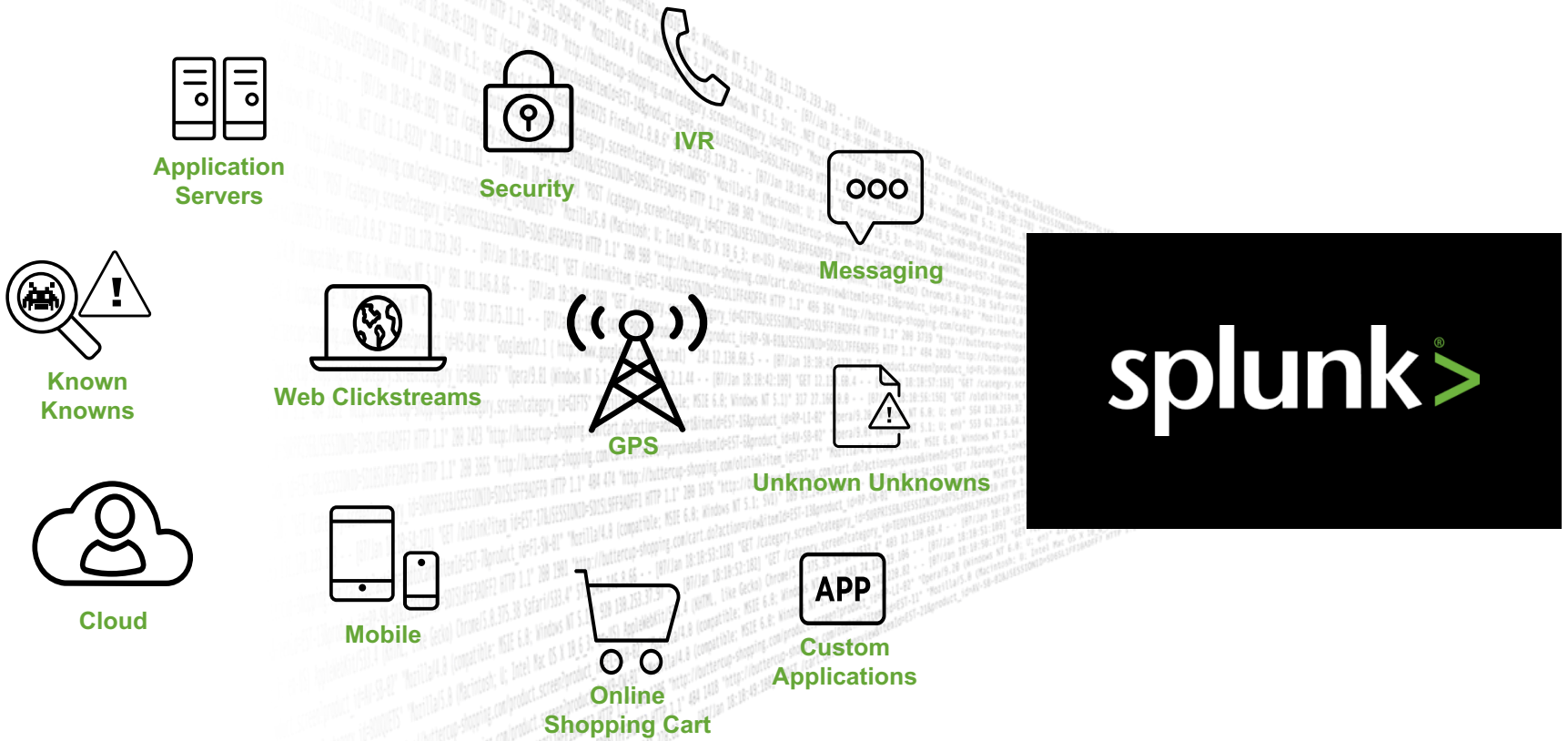
### Fraud



# Interrelated Components of Fraud

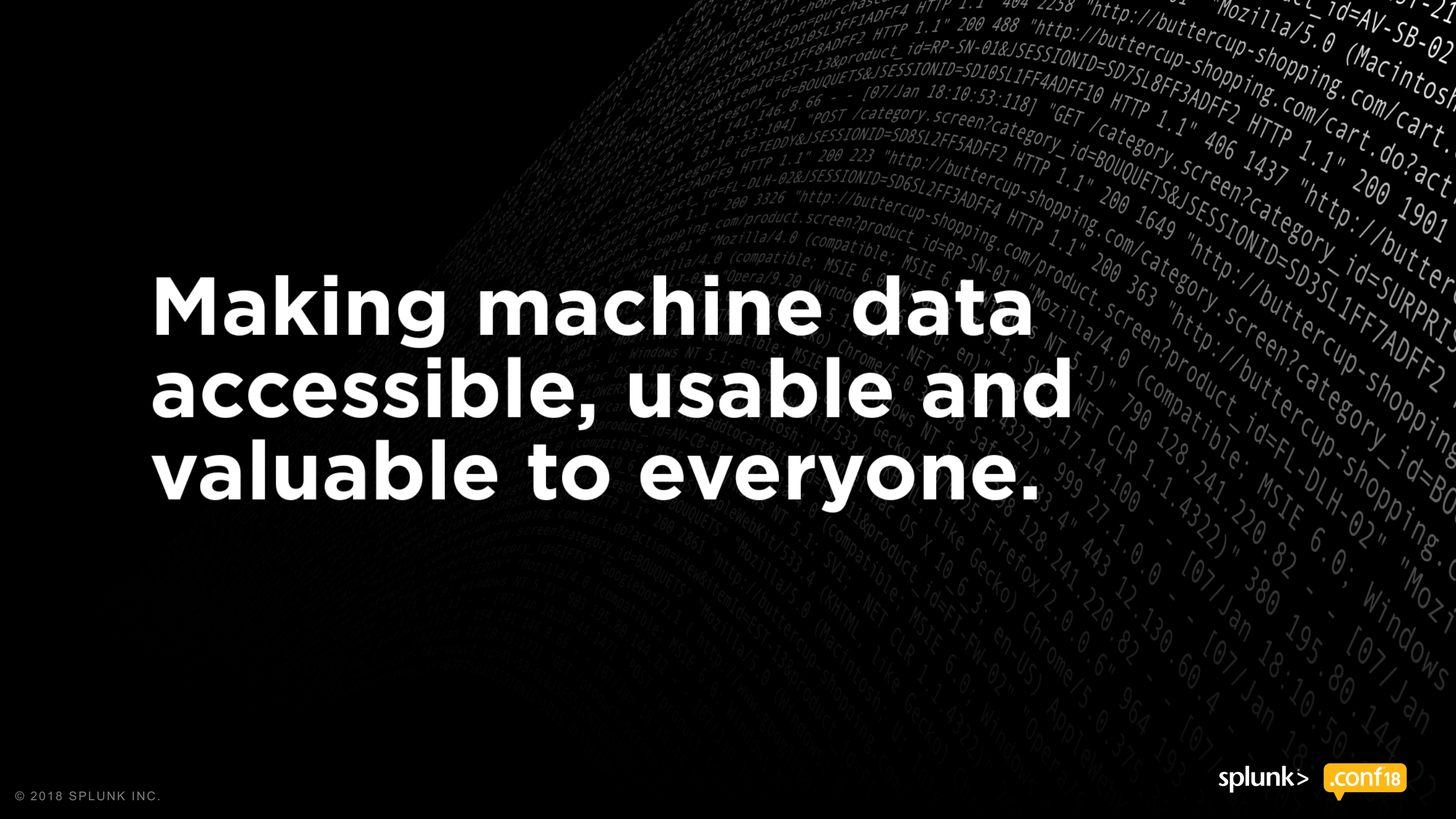


# Turning Machine Data Into Answers




## Challenges

- Outliers & Anomalies
- Account Take Over
- Transactional Fraud
- Behavioral Patterns
- Abuse



# Making machine data accessible, usable and valuable to everyone.



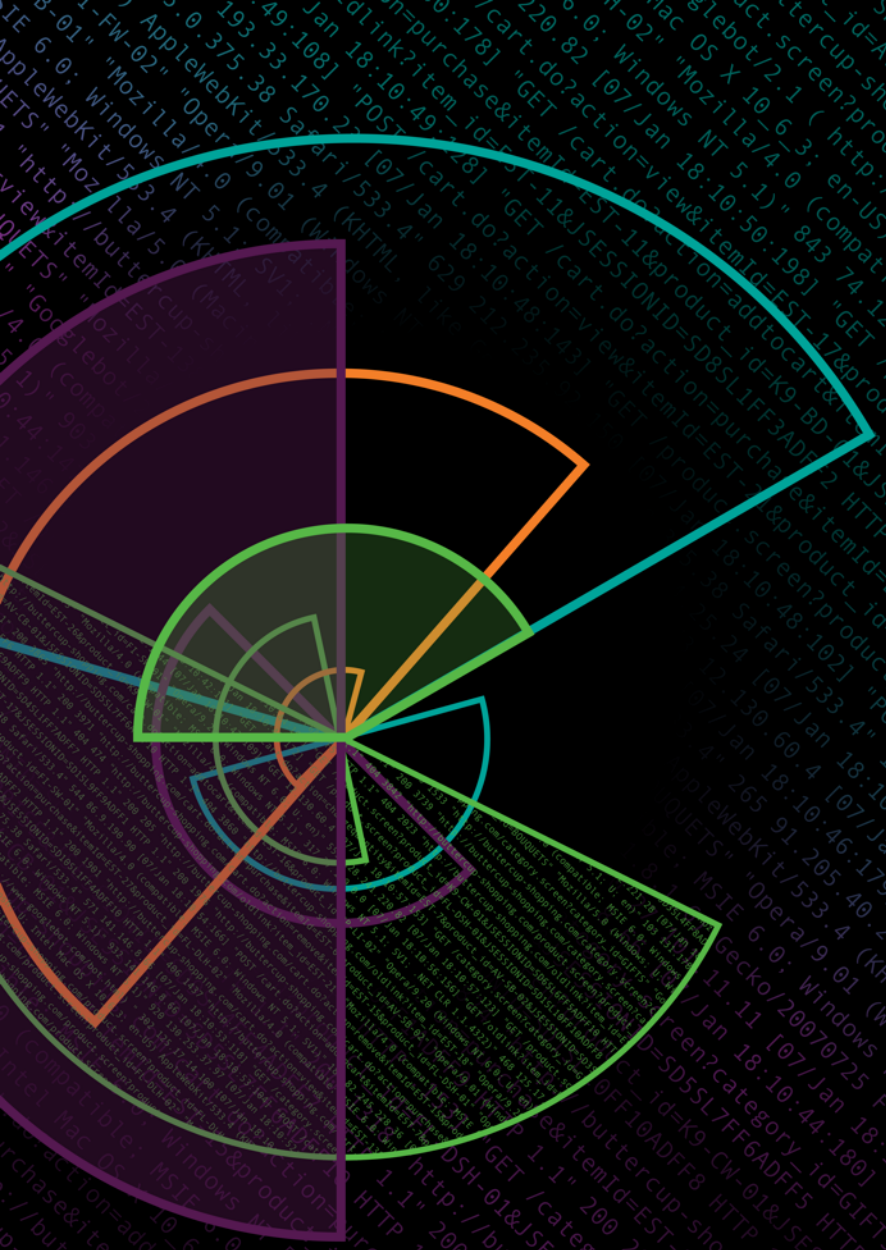
splunk> .conf18

# Agenda

## Let's measure our Time

- Approaches To Fraud
- Machine Learning
- Now What?

splunk> .conf18

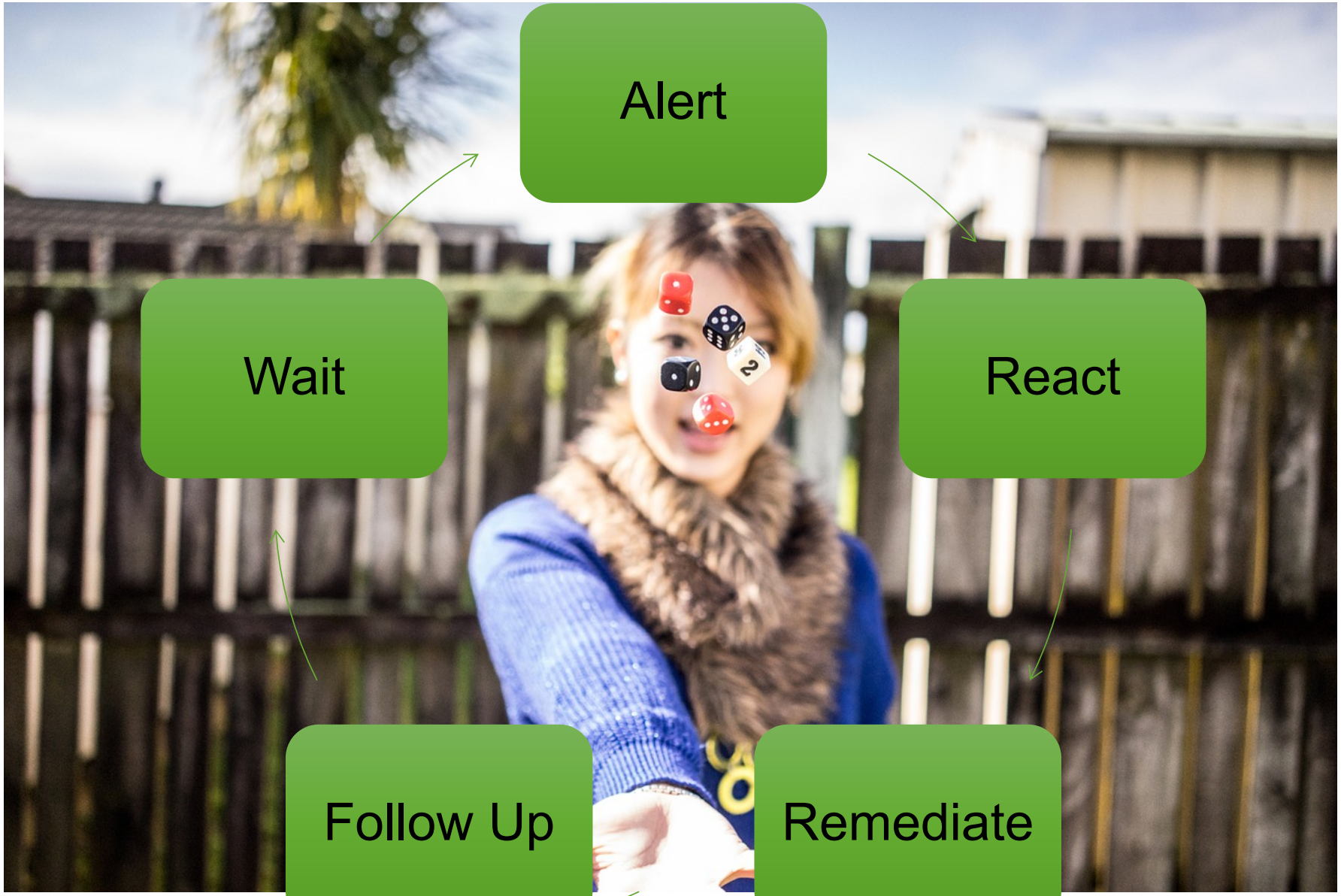


# Approaches To Fraud

**Time is our only non-renewable resource**

# We Are Failing at Prioritization

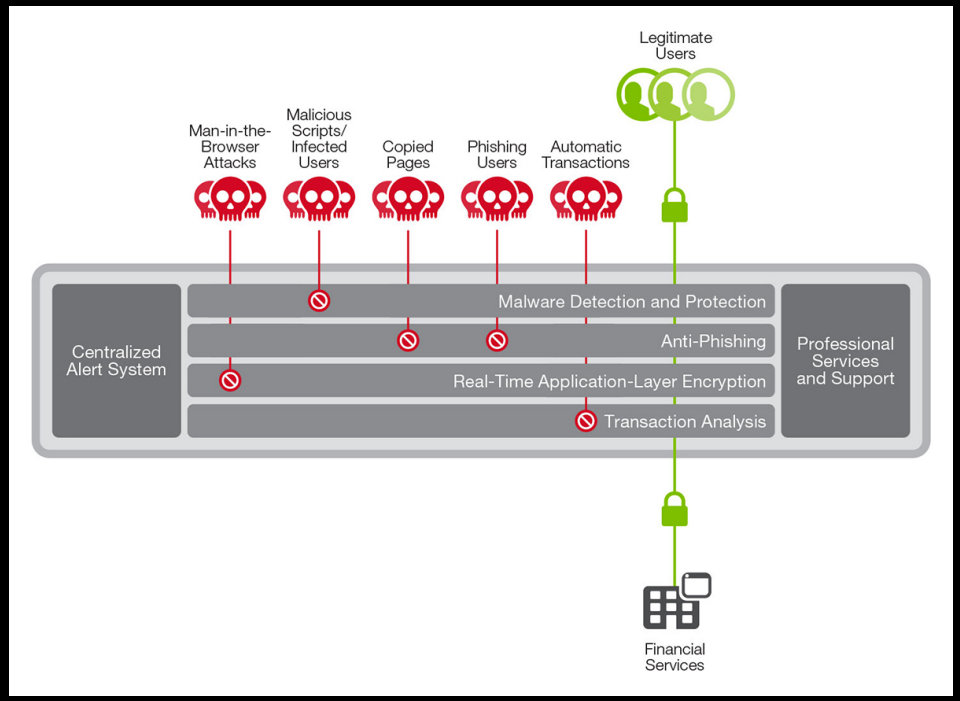




# Splunk: Anti-Fraud

## Where we are best of breed

### Web Fraud

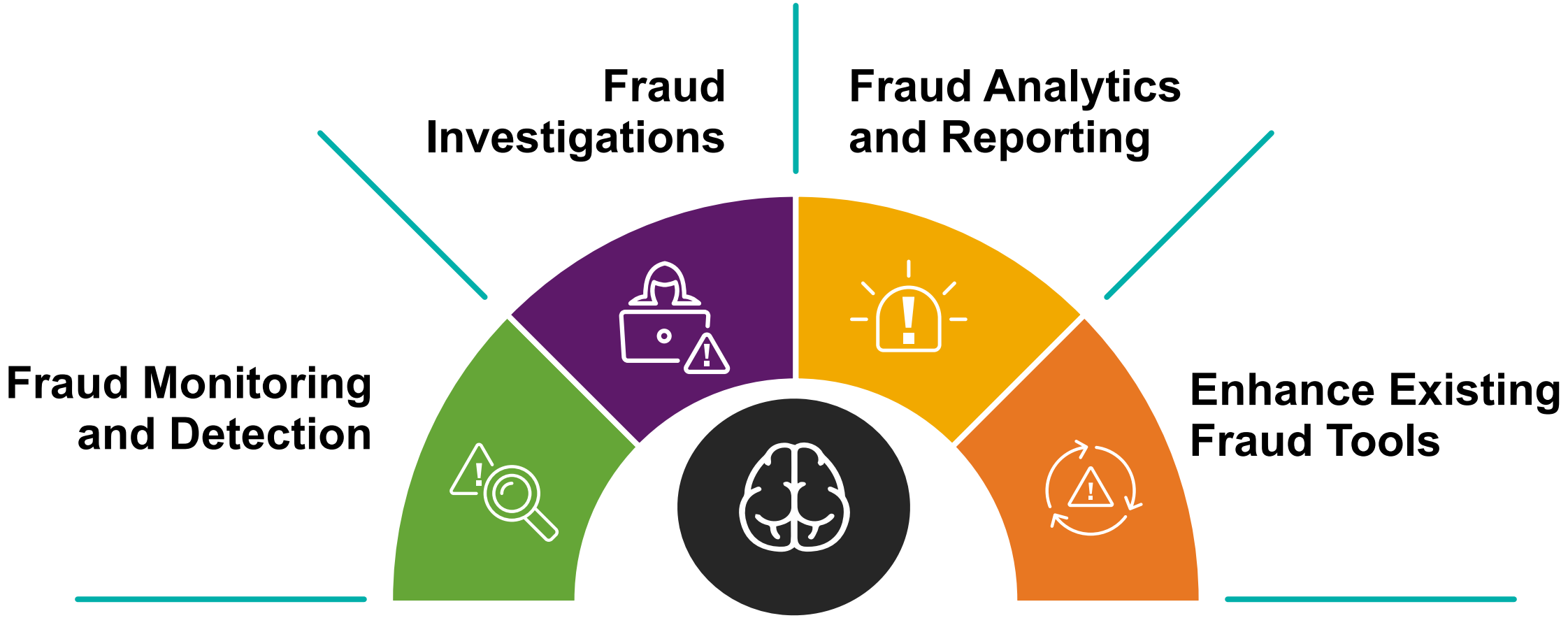



Source: F5

### Transactional

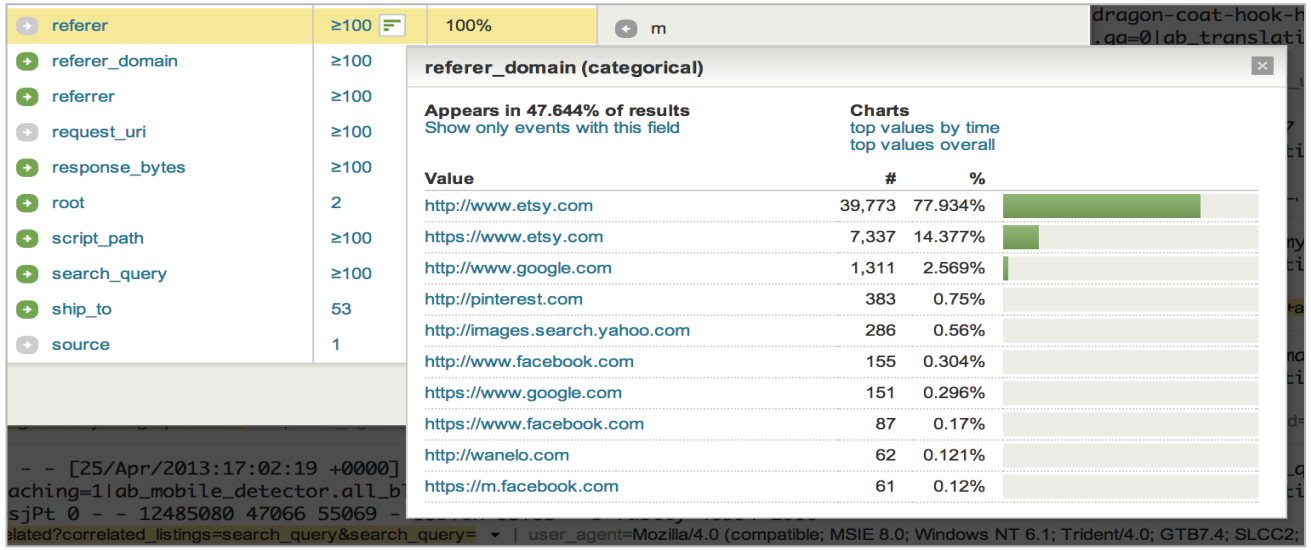
| Trans Date | Post Date | Type | Description | Expense Category | Amount |
|---|---|---|---|---|---|
| 11/24/2011 | 11/25/2011 | Sale | EXXONMOBIL 97484687 | Auto Related | $57.31 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $1,685.40 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $818.41 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $1,618.80 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $1,085.30 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $2,035.69 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $1,618.80 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $2,152.59 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $1,085.30 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $2,352.36 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $1,618.80 |
| 11/24/2011 | 11/24/2011 | Sale | MLK | Household | $2,152.59 |

# Needs of Anti-Fraud Teams

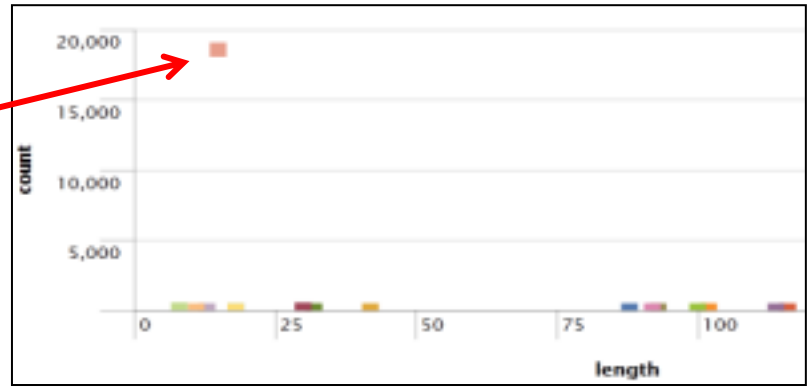


# Need #1: Fraud Monitoring and Detection

- Advanced correlations
- Baseline & detect anomalies
- Deviations
- Real-time searches & alerts
- Automation



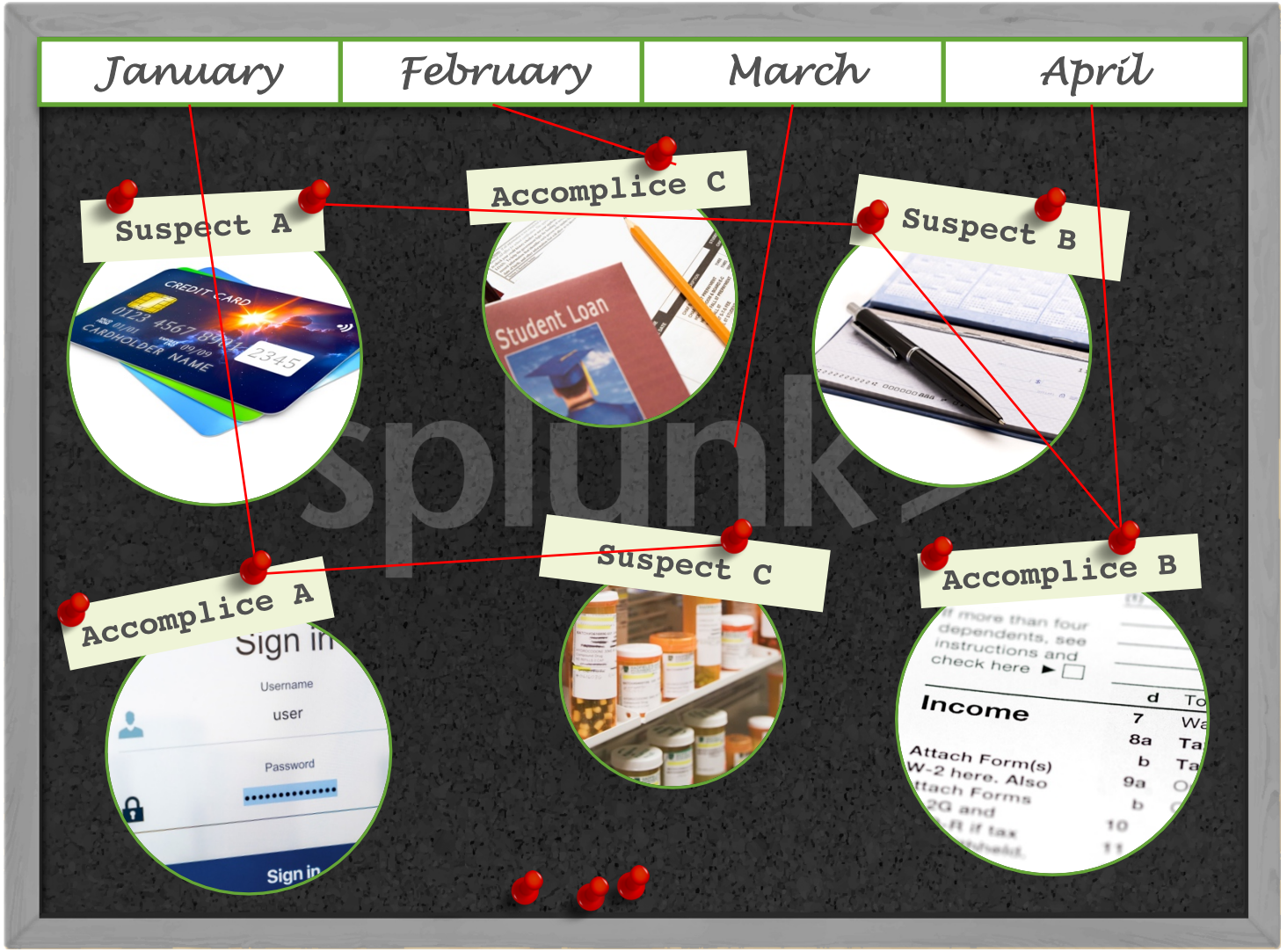

**Spot outliers**



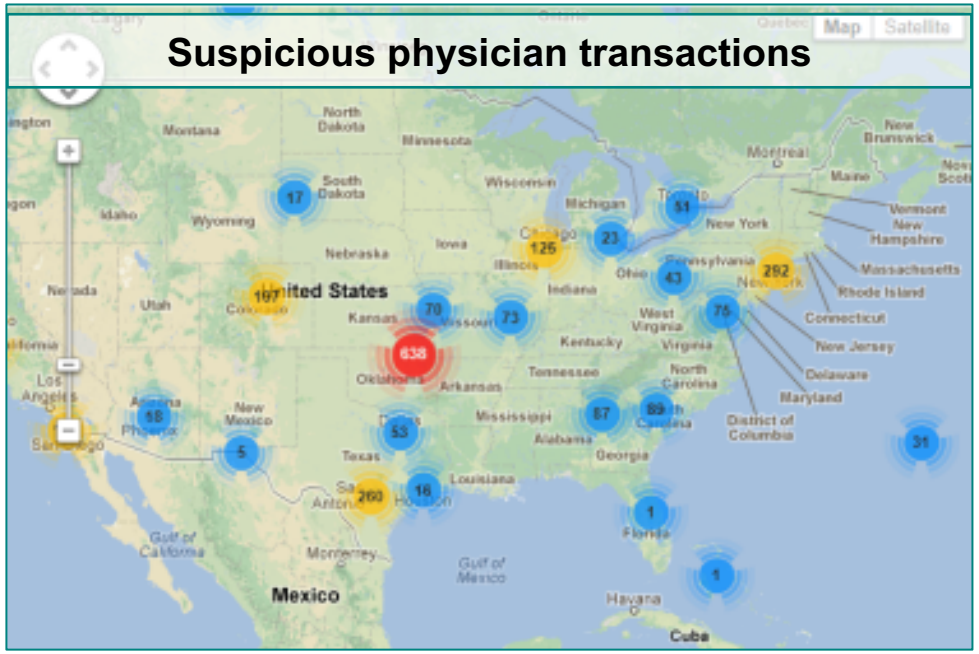

splunk> .conf18

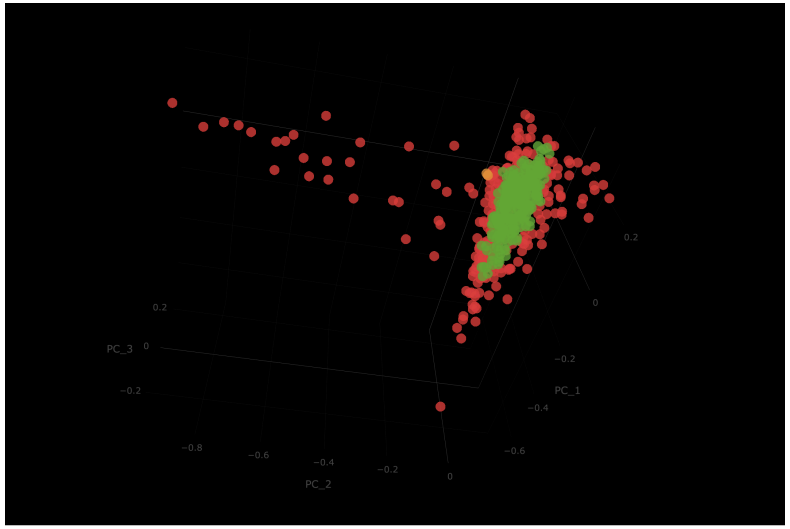
# Need #2 – Fraud Investigations

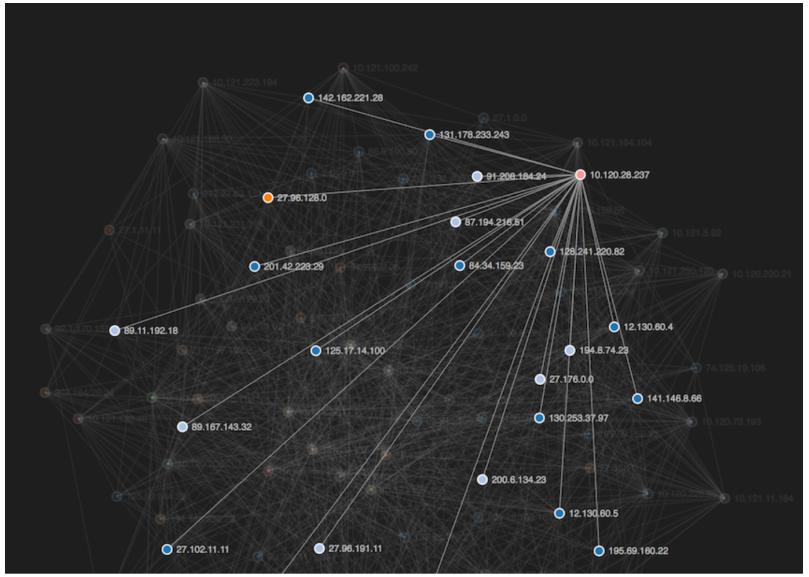
- Pivot
- Patterns
- Past

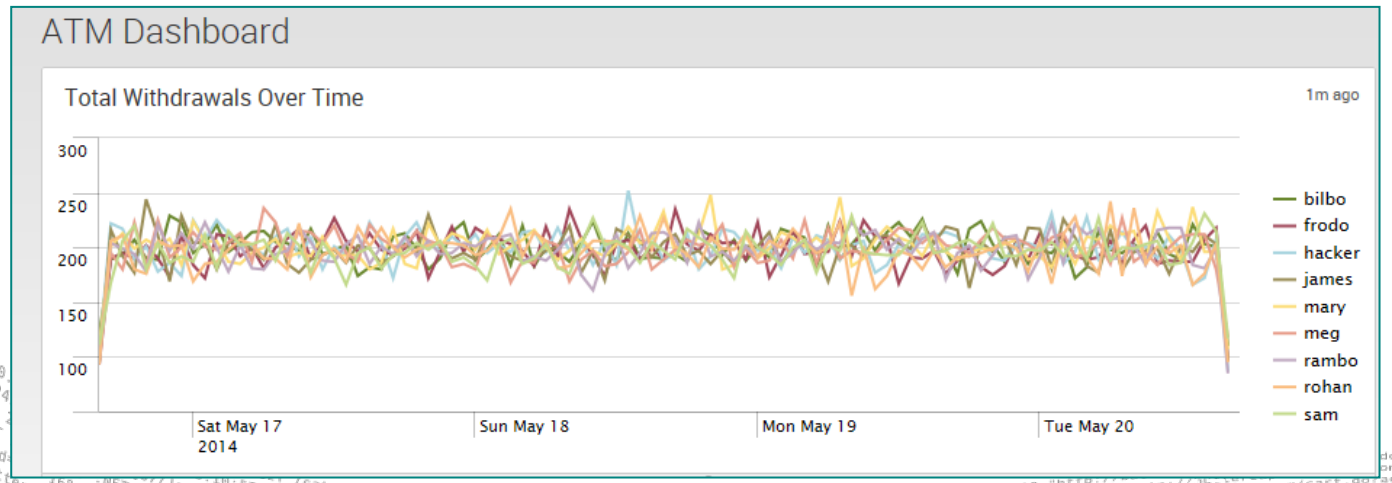


# Need #3 – Fraud Analytics and Reporting



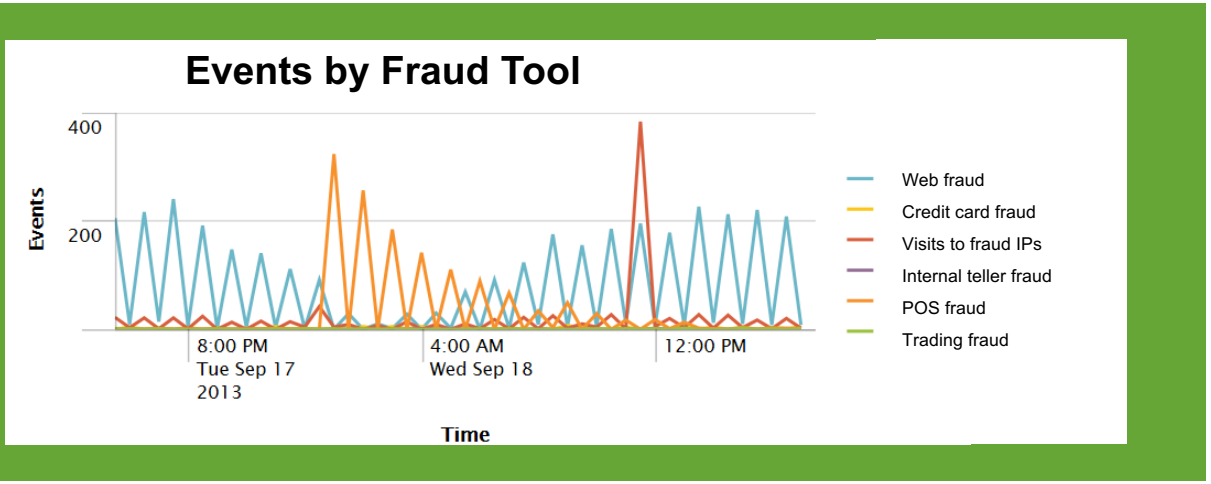



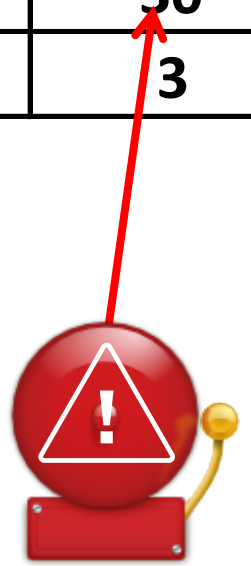




splunk> .conf18

# Need #4: Enhance Existing Fraud Tools

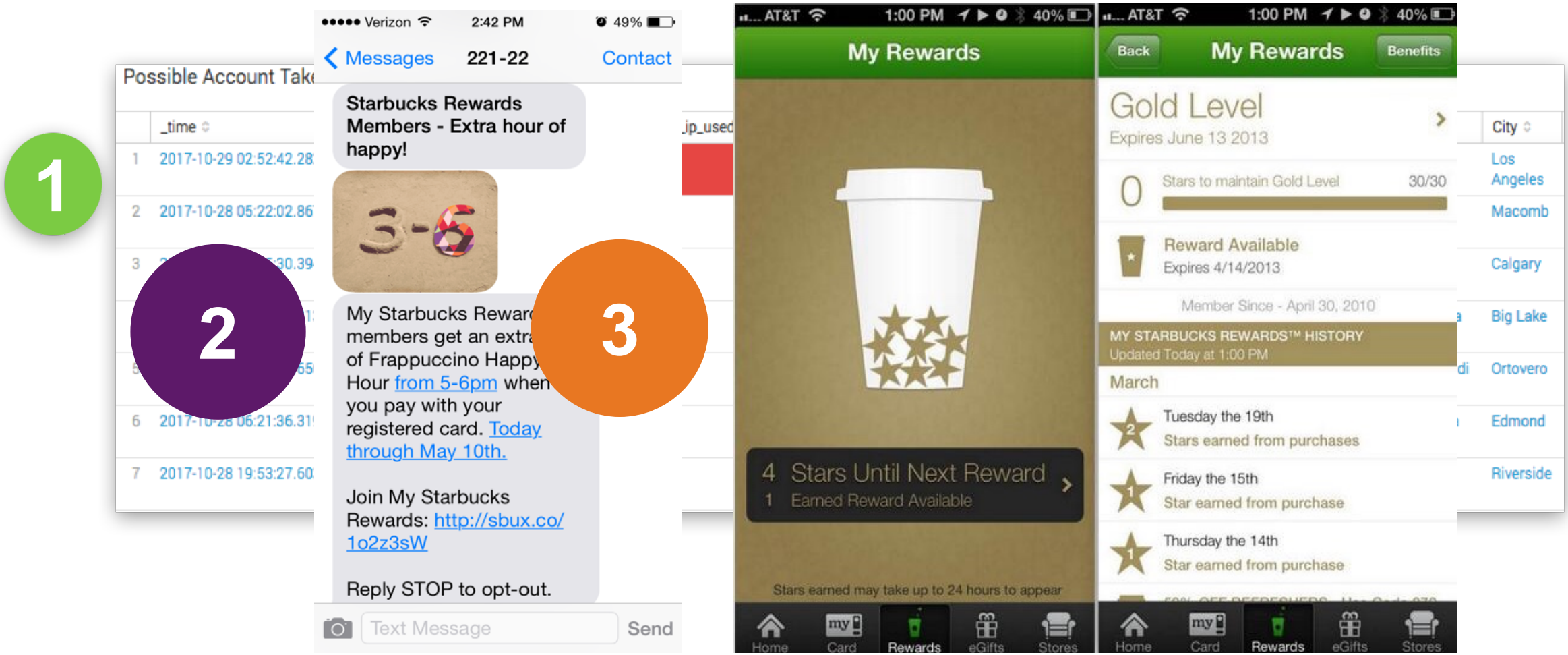
| Sample Splunk Summary Index | | | | |
|---|---|---|---|---|
| Session ID | Web fraud risk score | Credit card risk score | Threat Intel risk score | Splunk Total |
| 1234567 | 0 | 2 | 0 | **2** |
| 7654321 | 6 | 9 | 15 | **30** |
| 1231789 | 1 | 2 | 0 | **3** |



# Account Takeover (ATO) Example

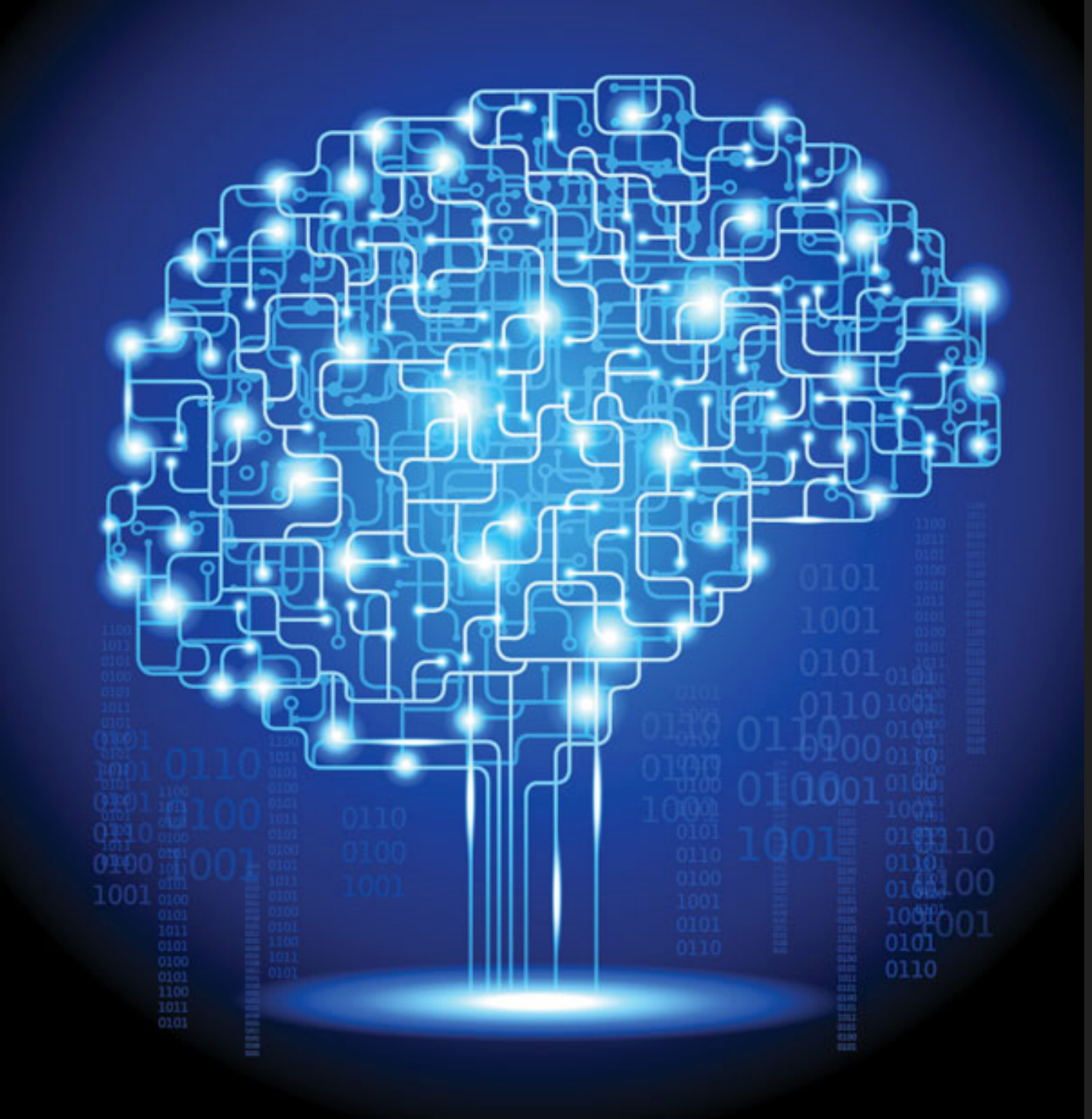
## Monitor Logins from Unusual IPs/Locations




splunk> .conf18

# What is Machine Learning?

**I, for one, welcome our _____ overlords.**



# Machine Learning

Automating analytical model building using algorithms that iteratively learn from data without requiring explicit programming

# Evolution of Security Correlation, Advanced Analytics & ML

Complexity

**N-Dimensional Advanced – Data Science**

- Shift from heavy manual tagging and rule building alone,
- Machine learning and data science for UEBA
- Enhances analyst capabilities to identify **unknown** threats

**Multi-Dimensional – Analytics**

- Hybrid model developed as adversaries circumvent basic correlation
- Goal to reduce false positives
- Thresholds and combinations of rules developed.
- Behavioral models, statistics and patterns not signatures alone

**Two-Dimensional – Correlation**

- Regex/pattern-matching for strings
- Used in anti-malware, IDS/IPS, DLP and basic/legacy SIEM
- Use of string matching to search a binary file to identify type of threat
- Enhanced capability to identify **previously known** threats and host enumeration

**One-Dimensional – Correlation**

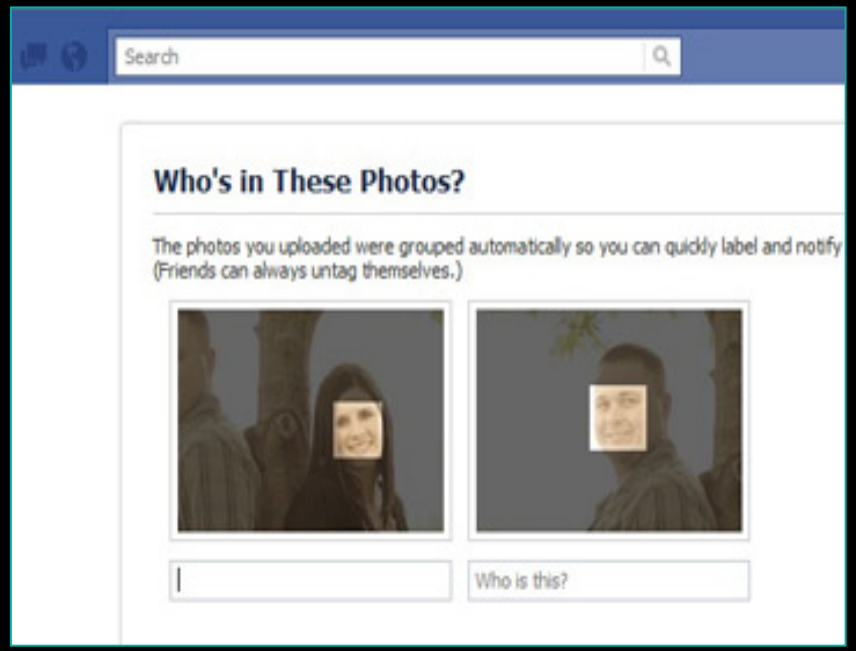
- Fast and efficient basic matching eg. domains IP addresses, user-agent, MD5 file hashes
- Boolean operators to identify if signature is on a black/white list.
- Common usage in most firewall and IDS tools

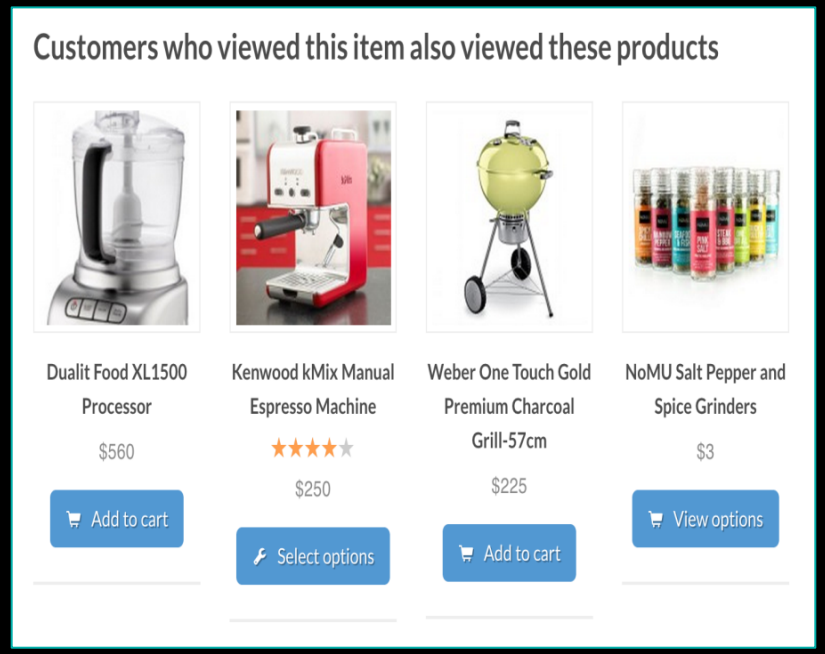
Data Volume & Velocity

splunk> .conf18

# ML Examples IRL

## It's everywhere




# Solving Problems With Machine Learning

## Anomaly Detection

- Past behavior
- User vs Group
- Unusual changes

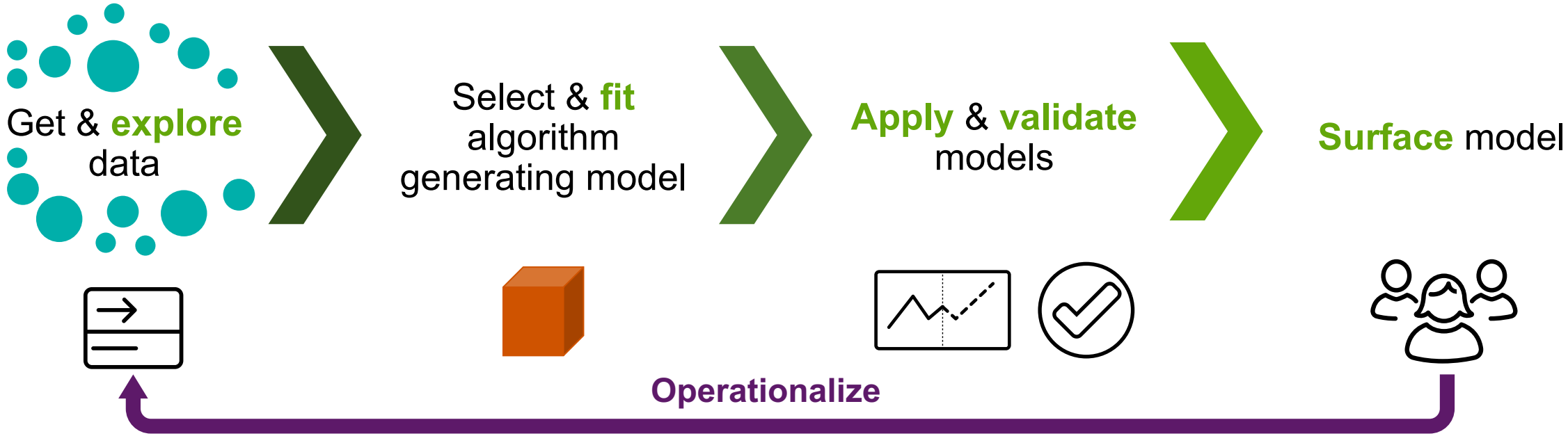
## Predictive Analytics

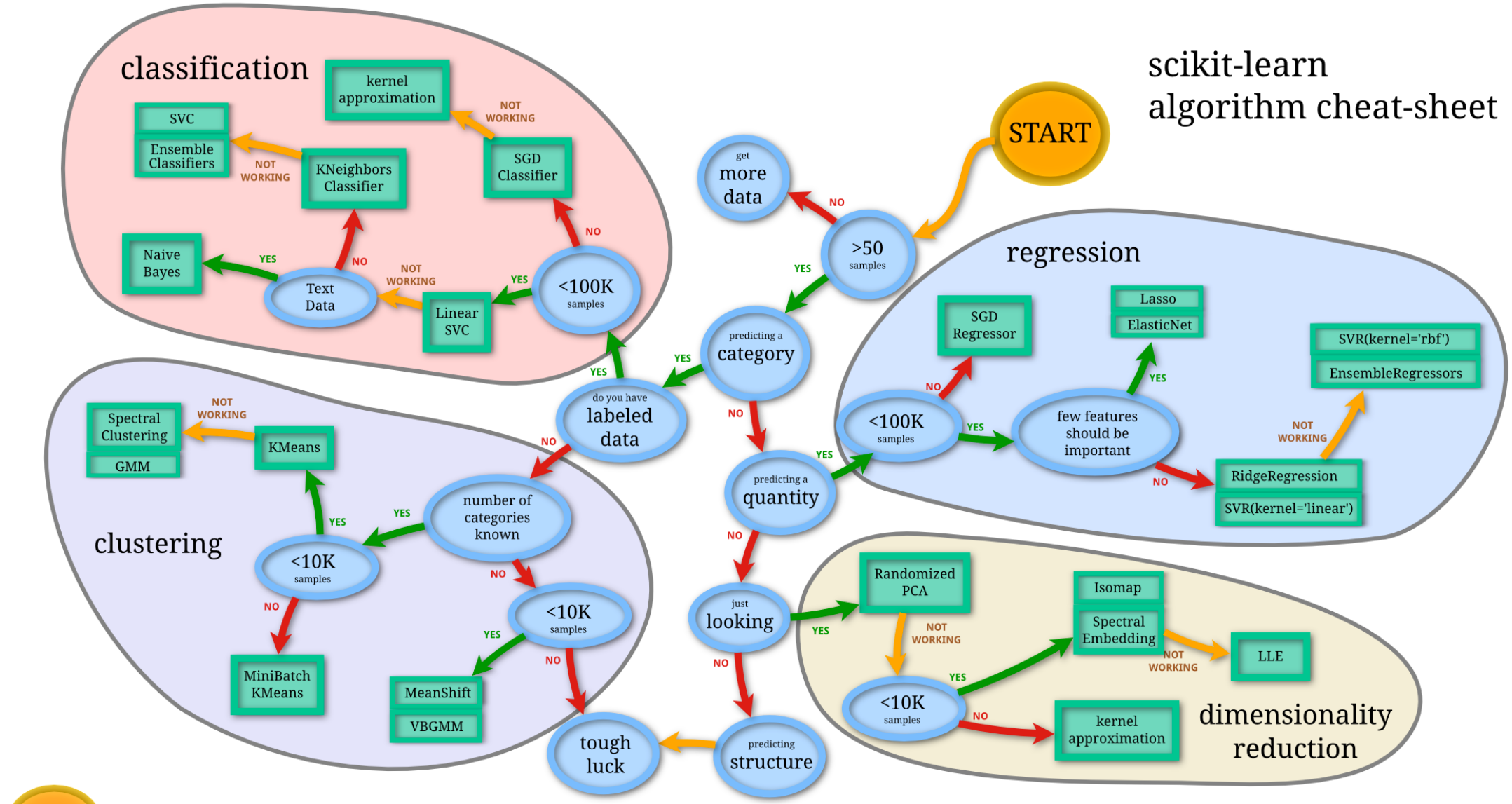

- Trend forecasting
- Planning
- Early warning

## Clustering

- Identify peer groups
- Event correlation
- Behavioral analytics

# The ML Process





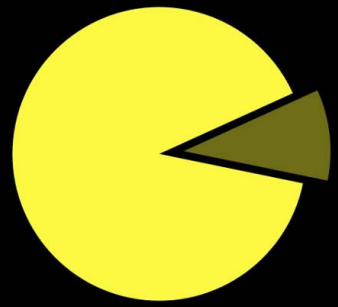
# Exploratory Data Analysis

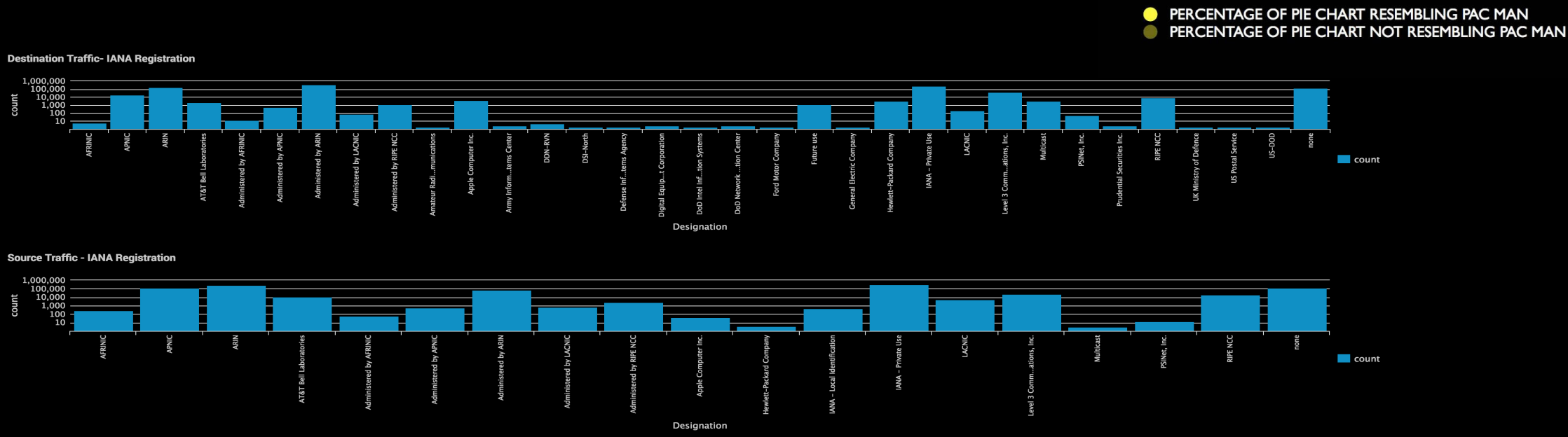
**Utilizing the Splunk platform for data science**

splunk> .conf18

# Visualization & Creating Context (EDA)

- **Visualization** is a powerful EDA tool
  - Not everything can be described as bits, bytes, plaintext or pie charts.
- **Correlation** to add context to your data during the EDA proce

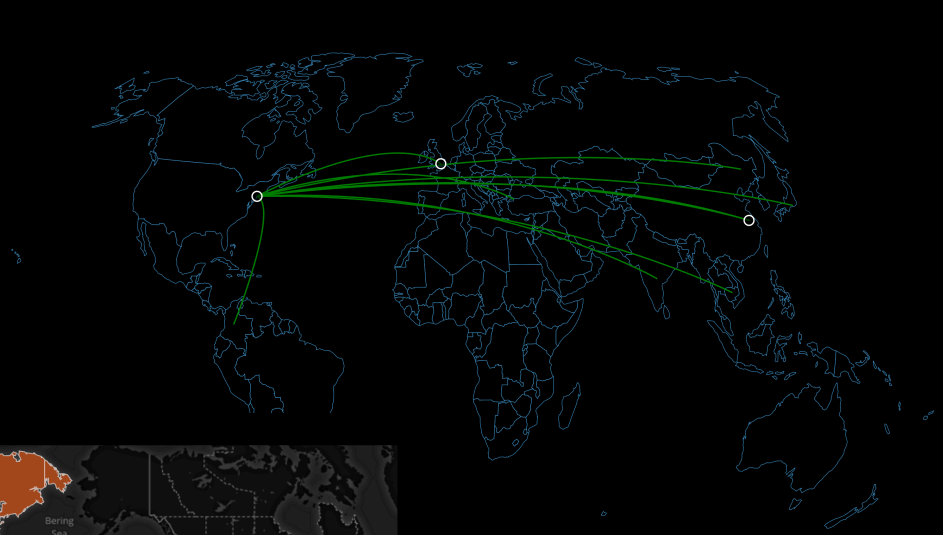





splunk> .conf18

# Geographical EDA - Visualization

- Visualization useful for exploring multi-dimensio
- Tells a story about the data you can't describe ir
- "Where are connections 'originating', and how o




Attempted SSH Access by Country


| t_lat | dst_lon | ids_type | msg |
|---|---|---|---|
| 74900 | -73.98650 | network | Colombia (161.18.62.72) connects to United States (192.241.146.96) (NULL) |
| 50000 | -0.13000 | network | United States (192.241.146.96) connects to United Kingdom (91.189.89.199) (NULL) |
| 74900 | -73.98650 | network | Turkey (78.188.180.193) connects to United States (192.241.146.96) (NULL) |
| 74900 | -73.98650 | network | China (112.85.218.11) connects to United States (192.241.146.96) (ssh) |
| 74900 | -73.98650 | network | Colombia (186.119.61.166) connects to United States (192.241.146.96) (NULL) |

Number of Connections: 47616

SSH Attempts - Numerical Outliers

splunk> .conf18

splunk> App: Claims Analysis ▾ Gleb Esman ▾ Messages ▾ Settings ▾ Activity ▾ Help ▾ Find

Data Overview | Data Integration | Enrichment | Anomaly Summary | Claims and Episode Analysis | Member Clusters | Provider Clusters | Claim Events | Others ▾ — Claims Analysis

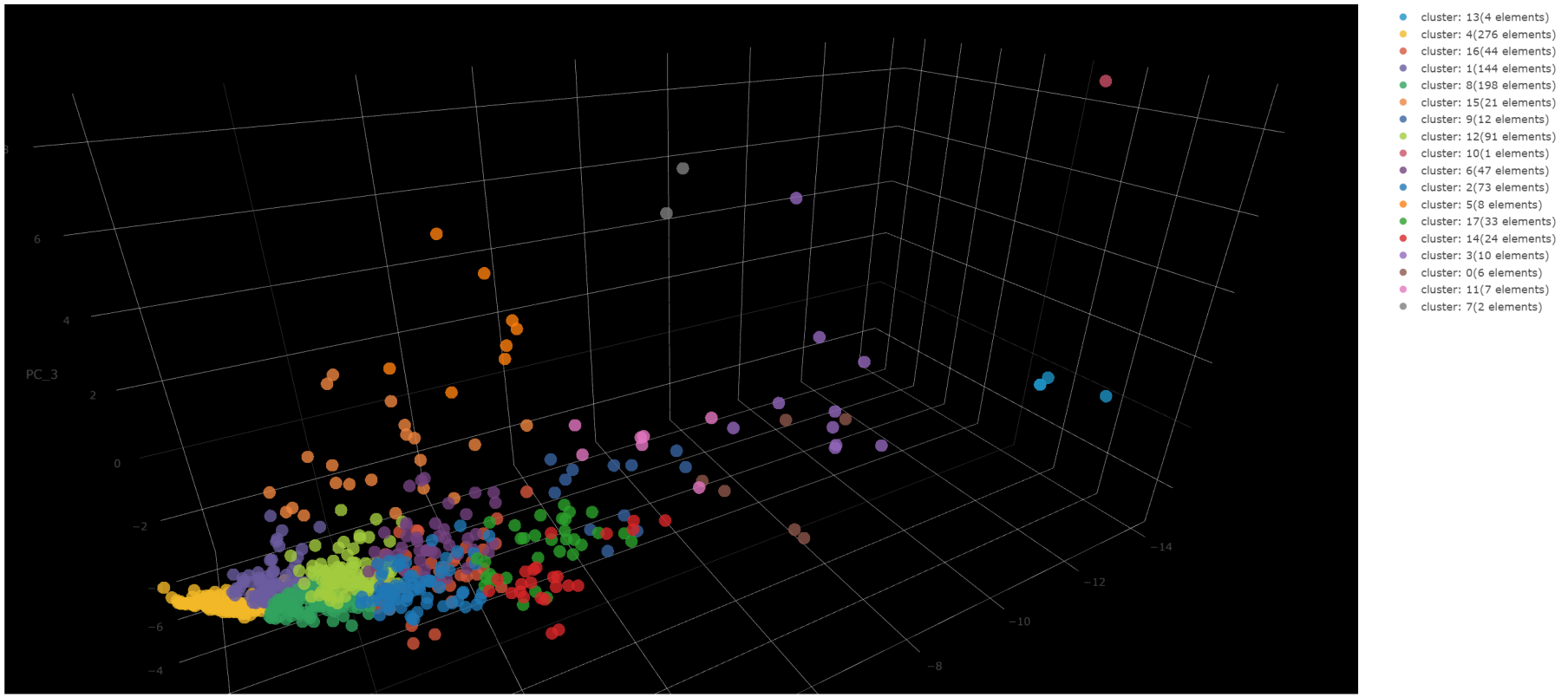
# Patient Claims, Cost and Diagnosis Anomaly Investigator - 3D

Edit | Export ▾ | ...

Clustering Algorithm: KMeans k=18 ⊗ ▾ | Select time window: Last 7 days ⊗ ▾ | Submit | Hide Filters

Reset Dashboard

Clustered Claims Data by: Total $ amount paid out, Insurance $ cost, Patient co-pay amount, Primary diagnosis codes used, Secondary diagnosis codes used, Drug codes used.




Detected Anomalies:

| | PC_1 (x) | PC_2 (y) | PC_3 (z) | cluster | ENROLID | TOTCOST | copaytot | ingtot | netpaytot | num_drug | num_dx | num_proc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -13.9460300716 | 1.23333181943 | 8.66409981836 | 10(1 elements) | 853532301 | 175370.98 | 778.15 | 9067.79 | 165525.04 | 10 | 17 | 72 |
| 2 | -7.28332188576 | -6.80316072015 | 4.81332180042 | 7(2 elements) | 884268802 | 55297.46 | 638.34 | 24346.42 | 30312.7 | 22 | 23 | 32 |
| 3 | -8.20940917125 | -8.55247798253 | 5.66966119146 | 7(2 elements) | 2059938602 | 61514.4 | 606.31 | 27938.93 | 32969.16 | 39 | 13 | 30 |
| 4 | -13.1334643037 | 2.24187114206 | 0.225561438913 | 13(4 elements) | 1064879704 | 91184.11 | 56.89 | 6551.5 | 84575.72 | 62 | 46 | 141 |
| 5 | -11.8554131149 | 6.11646101014 | 1.87965286941 | 13(4 elements) | 25433641701 | 109547.29 | 150.89 | 730.97 | 108665.43 | 15 | 50 | 145 |

splunk> App: Security Essentials Anti-Fraud

Administrator | 1 Messages | Settings | Activity | Help | Find

Search | Anti-Fraud Scenarios | Dashboards | Anti-Fraud Scenarios - Links

Security Essentials Anti-Fraud

# Payment Cards: Anomalous Transactions - 3D

# Transform Your Data

## Set your own reality

Edit | Export | ...

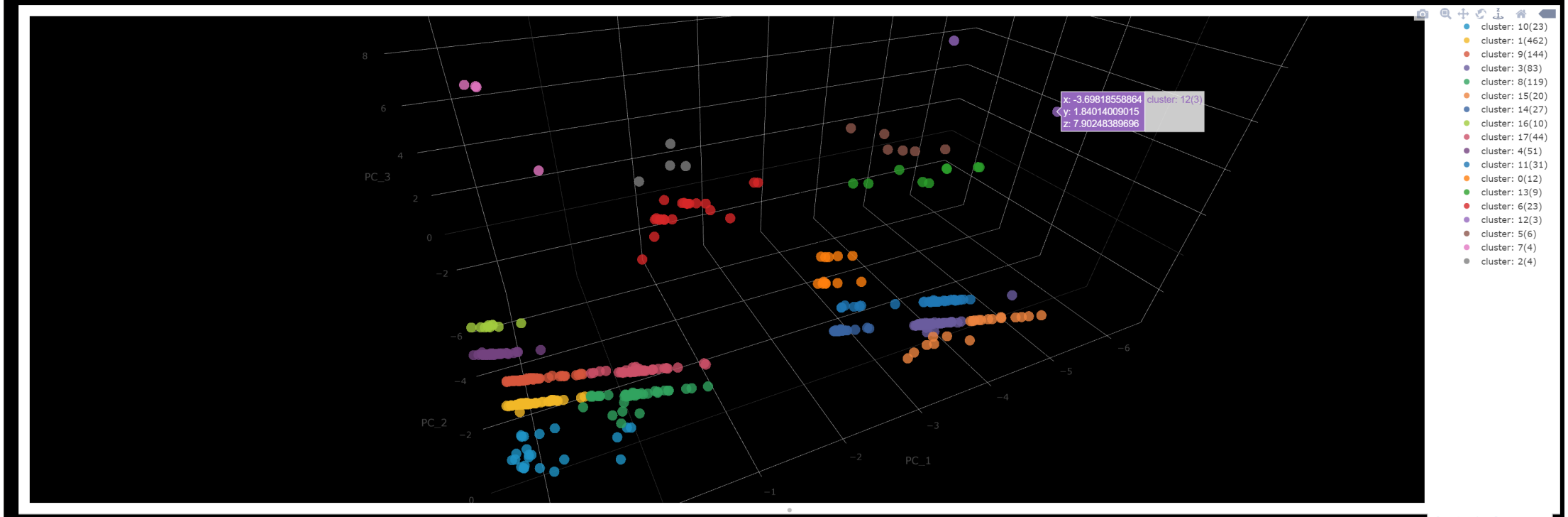
Clustering Algorithm: KMeans k=18

Select time period (of available data): Last 24 hrs (full txn data)

Submit | Hide Filters

Reset Dashboard

3D view of data clusters



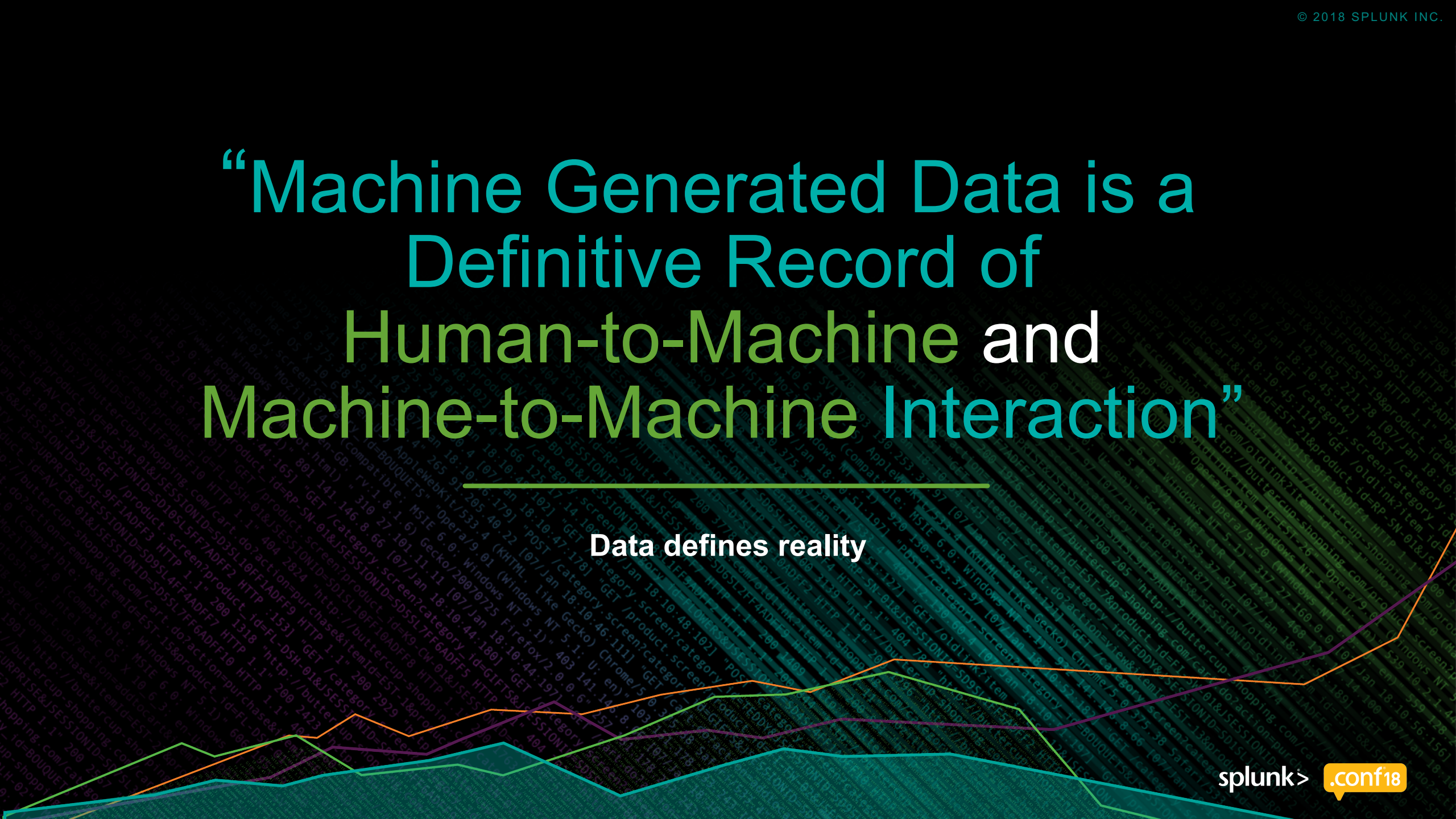

1m ago

Detected Anomalies:

| | PC_1 | PC_2 | PC_3 | cluster | _time | card_number_masked | compromise_type | txn_region | region_change | merchant_change | merchant_name | txn_terminal_id | time_delta | txn_type | txn_amount | txn_min_max_avg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -3.69818558864 | 1.84014009015 | 7.90248389696 | 12(3) | 2017-03-10 21:43:00 | CARD010600920408030 | | US | 1 | 1 | WM SUPERCENTER # | 0W000357930004 | 47 | PURCHASE | 45.0 | 3.37 / 128.22 / 30.26 |
| 2 | -3.81810533094 | 1.85866264517 | 7.82107865432 | 12(3) | 2017-03-19 15:35:34 | CARD010600924668027 | | US | 1 | 1 | WM SUPERCENTER # | 0W000320260012 | 43 | PURCHASE | 50.0 | 2.06 / 131.0 / 22.97 |
| 3 | -2.75726098295 | 0.95874873927 | 9.98585473828 | 12(3) | 2017-03-17 18:31:34 | CARD010600926071014 | | US | 0 | 0 | WM SUPERCENTER # | 0W000320260002 | 55 | PURCHASE | 100.0 | 1.68 / 500.0 / 22.37 |
| 4 | -1.03352279421 | -7.28241919574 | 0.965314069716 | 2(4) | 2017-04-17 04:35:58 | CARD010529244010 | | PR | 0 | 0 | BURGER KING 4978 | 300V5009 | 413490 | PURCHASE * | 3.33 | 3.33 / 149.84 / 29.01 |
| 5 | -1.69742835467 | -7.13687324121 | 1.40936879098 | 2(4) | 2017-04-06 19:28:45 | CARD010600924668027 | | PR | 0 | 0 | ECONO PONCE ECR | HATHBTRN00010003 | 24165 | P CSH BACK | 41.82 | 2.06 / 131.0 / 22.97 |
| 6 | -1.47603123285 | -7.17256131112 | 1.57935592253 | 2(4) | 2017-04-19 19:18:14 | CARD010600924668027 | | PR | 0 | 0 | ECONO PONCE ECR | HATHBTRN00010003 | 13546 | P CSH BACK | 17.58 | 2.06 / 131.0 / 22.97 |
| 7 | -1.49407942138 | -7.13030794842 | 2.72618418865 | 2(4) | 2017-02-17 09:57:13 | CARD010600926385018 | | PR | 0 | 0 | ECONO AGUAS BUENAS | 30V29301 | 1725 | P CSH BACK | 11.1 | 2.01 / 170.0 / 20.31 |
| 8 | 0.719092413593 | -0.0300507904028 | 9.94790513374 | 7(4) | 2017-03-22 13:52:44 | CARD010591538040 | fraud | PR | 0 | 0 | NATIONAL LUMBER AGUAS B | 7147J803 | 50 | PURCHASE | 5.87 | 5.11 / 181.52 / 26.94 |

# "Machine Generated Data is a Definitive Record of Human-to-Machine and Machine-to-Machine Interaction"

**Data defines reality**



splunk> .conf18

# Now What?

**“We are the music makers,**

**And we are the dreamers of dreams”**

# Collect & Normalize

## Inventory, Standards, and Data

# Assets

## Do you balance your checkbook?


- Device ID (MAC)
- Ecosystem
- Data Driven
- Risk Based Approach

# Take Inventory

More than just assets

## Access Points

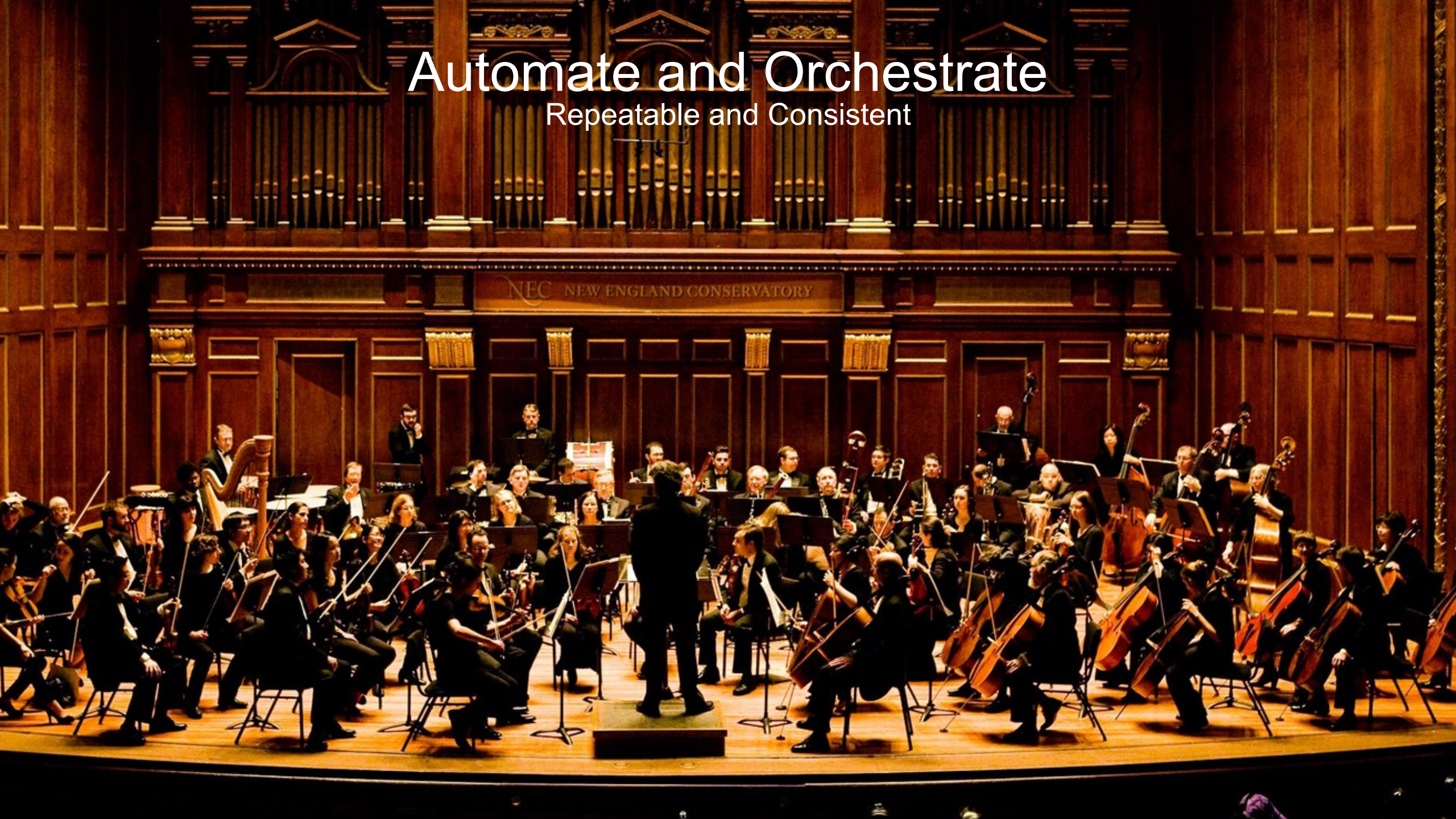
## Egress/Ingress Data


# Enrich

## Context and Impact

# Automate and Orchestrate

Repeatable and Consistent



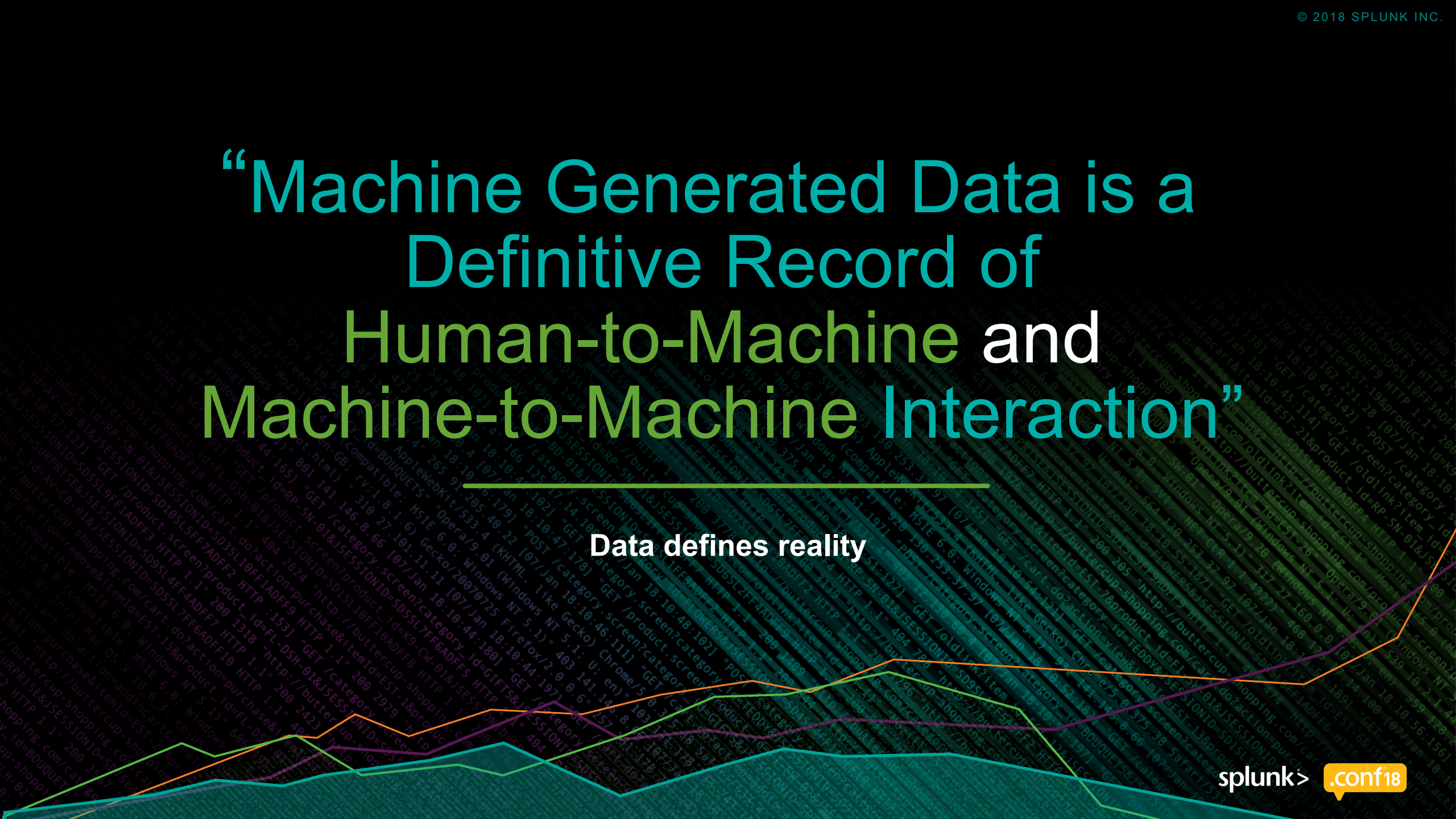
# Machine Learning

**10 Print “hello”**
**20 Go to 10**

# What Does a “Clean” Network Look Like?

# "Machine Generated Data is a Definitive Record of Human-to-Machine and Machine-to-Machine Interaction"

**Data defines reality**



splunk> .conf18

# Splunk Demo

**Presented by Matthew J Joseff**

splunk> .conf18

# Thank you

mjoseff@splunk.com

**Don't forget to rate this session in the .conf18 mobile app**