## splunk> .conf2017

# Predictive, Proactive, and Collaborative ML with IT Service Intelligence

Not a Science Project – Creating actionable events through Analytics

Nate Smalley | Staff SE Andrew Stein | Analytical Architect Sept 26, 2017 | Washington, DC

spiunk

## **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, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.

.screen?product\_id=FL-DSH-01&JSE

## **Nate Smalley**



- IT Operations Technologist (Reformed Security Guy)
- Former Technical Director of Security & Monitoring Tools Team – Apollo Group (University of Phoenix)
- Currently Splunk Staff Sales Engineer supporting Large Businesses in FinServ & Manufacturing
- Enjoy Long walks across SNMP and Candle light dinners while fighting Operational Outages



## **Andrew Stein**



- Splunk Global Analytical Architect
- 17 years creating mathimatically modeled solutions
- I spend 80 percent of time spent preparing data and 20 percent of time complaining about the need to prepare data.



### **Problem Statement**

Operations Teams need more time between an alert and a failure that has Availability impacting ramifications. The introduction of Machine Learning is a have to have in order to predict these failures. These notable events must be able to pushed and collaborated on by via teams in various tools.



## Agenda

- What data do we need
  - Indicators Matter
  - Data Where to Get it
  - Demo
- We Need Machine Learning
  - What can we use
- Real use case Example
- Let Dive into "How it works"
- Next Steps



## **Indicators Matter**

ITSI



#### Indicators

Key Performance Indicators that is

• Defined - **metrics** that are used to evaluate the overall status of a service.

Leading Indicators – Drivers of a Result

Lagging Indicators – Outcome of the Result

Example Scenario

- DB Runs out of Space
- KPI Storage value = 100% <- Leading Indicator
- KPI User Response time value = 2000+ secs <- Lagging Indicator



## KPI's so what?

Understanding the historical break out of Leading vs Lagging KPI's and the association to Services is critical in understanding how to predict a good outcome vs bad outcome.

- Understanding these key parts from the Service Experts is critical:
  - Business Priority
  - Use Case Data Needed
  - The "RIGHT" KPI's to measure
  - What decision happens when a Bad outcome Occurs





## Data Where to Get it

I need it STAT



.conf2017

### Where Does Data Come From?



## What Do I do with it Now?

- Use IT Service Intelligence
- Dynamic Service Structure
  - Provide Flexible Dependency Service mapping for interactions at Scale
- Build Key Performance Indicators
  - Leverage a Platform to build out KPI's
  - Ensure repeatability amongst Services for Consistency
  - Aggregation and Per Entity KPI Values
- Ease the burden of Cleaning Data (Schema at search time?)

Screen?product id=FL-DSH-01&JSESSI

"Data scientists spend 60% of their time on cleaning and organizing data. Collecting data sets comes second at 19% of their time, meaning data scientists spend around 80% of their time on preparing and managing data for analysis."

Data-driven service monitoring and analytics



splunk> .conf20

## Real Use Case Example

Lets Get the data



# We Need Machine Learning

Do you want to watch it or let it go?



## **Custom Machine Learning – Success Formula**



### **Overview of ML at Splunk**



**Splunk** > Platform for Operational Intelligence



## **Three Types Of Machine Learning**

#### Supervised Learning:



#### **Unsupervised Learning:**



#### **Reinforcement Learning:**





4] "GET /Category.screen?category\_id=GIFTS&JSESSIONID=SDISLAFF10ADEF10 HTTP 1.1" 404 720 "http://buttercup=shoppins.com/category\_id=GIFTS&JSESSIONID=SDISLAFF10ADEF10 HTTP 1.1" 404 332 "http://buttercup=shoppins.com/category\_id=SITADEF10 HTTP 1.1" 400 4332 "http://buttercup=shoppins.com/category\_id=SITADEF10 HTTP 1.1" 400 43342 "http://buttercup=shoppins.com/category\_id=S

## **Splunk Customers Have ML Problems**



Deviation from past behavior Deviation from peers (aka Multivariate AD or Cohesive AD) Unusual change in features <u>ITSI MAD Anomaly Detection</u> Predict Service Health Score Predicting Churn Predicting Events Trend Forecasting Detecting influencing entities Early warning of failure – predictive Identify peer groups Event Correlation Reduce alert noise ITSI Event Analytics



## **Splunk Machine Learning Toolkit**



#### Algorithms

25+ standard algorithms available prepackaged with the toolkit



Assistants

Guide model building, testing, & deployment for common objectives



#### Showcase

Interactive examples for over 25 typical IT, security, business, IoT use cases



.Screen?product id=FL-DSH-01&JSESSIONID=SD



**Python for Scientific Computing Library** 300+ open source algorithms available for use



MLib integration



# Real Use Case Example

Lets Use ML to get Predictive



## **Event Analytics**

**BUILT IN - IT Service Intelligence Machine Learning** 



## **IT Service Intelligence Event Analytics**

#### Notable Events are key

- We created a notable for Web Store Service – Called webstore\_health\_alert
- Lets see how Splunk can cluster this with other events to show other Notable Events that are attributed
- Smart Mode is actually Smart
  - It use clustering to group events remember unsupervised ML at the beginning

| splunk> App: IT Service Intelligence ~   |   |                       | Administrator $\checkmark$ | Messages ∽ S      | Settings 🗸 | Activity 🗸      |  |  |
|--|---|-----------------------|----------------------------|-------------------|------------|-----------------|--|--|
| Service Analyzer ~ Notable Events Review Glas<br>Default Policy C Smart Mode ? | Select the fields to be analyzed  |                       |                            |                   | ×          |                 |  |  |
| Applies to events which fail to meet the criteria of any oth                   | We selected some fields for you to start with to analyze event similarity.<br>You can Re-run Analysis for the Last 24 hours ✓ |                       |                            |                   |            |                 |  |  |
| Grouping Factors   | 84 fields selected, 23 used for Text Similarity, 6  | l used for Categorica | al Similarity              |                   |            |                 |  |  |
| Grouping Factors   | ✓ Field \$  | Туре ≎                | # of Values ≎              | Event Coverage \$ |            |                 |  |  |
| Machine learning is used to group notable events based r                       | ✓ NetObject   | Category ∨            | 7                          | 4.08 %            |            |                 |  |  |
| ✓ Adjust the importance of each factor?  | ✓ Title   | Category ∨            | 9                          | 4.08 %            |            |                 |  |  |
| Select fields to analyze event similarity                                      | ✓ account_id  | Category ~            | 2                          | 45.73 %           |            | -               |  |  |
| > Split events by field ?  | ✓ activity_due  | Category ∨            | 642                        | 6.84 %            |            |                 |  |  |
| > Break group ?  | ✓ alert   | Category ∨            | 3                          | 4.12 %            |            |                 |  |  |
| > Group information  | ✓ alert_color   | Category ~            | 3                          | 7.48999999999999  | 999 %      |                 |  |  |
|  | ✓ alert_level   | Category ~            | 3                          | 7.48999999999999  | 999 %      | to search query |  |  |
|  | ✓ alert_severity  | Category ~            | 3                          | 7.48999999999999  | 999 %      |                 |  |  |
|  | ✓ alert_value   | Category ~            | 11                         | 7.53 %            |            |                 |  |  |
|  | alerttriggertime  | Text ∨                |                            |                   |            |                 |  |  |
|  |   |                       |                            | Cancel            | Done       |                 |  |  |
| Cancel   |   |                       |                            |                   |            |                 |  |  |



## The Diamonds in the Rough

Clustering Events into actionable alerts

- Take 1000's or 100's of 1000's of alerts and connect them
- Use ML prediction to improve correlation
- Boil the events down to reasonable count of Actionable Events

luct.screen?product\_id=FL-DSH-01&J

| Service Analyzer 🗸  | Notable Events Review   | Glass Tables                                 | Deep Dives   | Multi KPI Alert                          | s Si                      | earch 🗸   | Confi   | gure 🗸   | Product Tour  |   | IT Servi                | ce Intellige     | nce    |
|---|---|--|--|--|---------------------------|---|---|--|---|---|-------------------------|------------------|--------|
| Notable Eve   | nts Review 🗸  |  |  |  |                           |   |   |  |   |   | Save as                 | Save             | ø      |
| 11685 events Last 2   | 4 hours 🗸 Add Filter 🗸  | search                                       |  |  | $\supset$                 |   |   |  |   |   |                         | Hide Timelir     | ne 😞   |
| Format Timeline 🗸   | - Zoom Out + Zoom to  | o Selection X                                | Deselect   |  |                           |   |   |  |   |   |                         | 1 hour per co    | lumn   |
| 500   |   |  |  |  |                           |   |   |  |   |   |                         |                  | 500    |
| 300   |   |  |  |  |                           |   |   |  |   |   |                         |                  | 300    |
| 6:00 PM<br>Sun Sep<br>2017                                    | 17  |  | 12:00 AM<br>Mon Sep 18                                       |  |                           |   | 6:00 AM   |  |   | 12:00 PM  |                         |                  |        |
| Ĵ≞ Sorted by? ↓ Time  | · •   |  |  |  |                           |   |   |  |   |   |                         | Ð                | ٥      |
| Title   |   |  | Time   |  |                           | Owner   |   | Severity   | Status  | Description   |                         |                  |        |
| Database Events: status:                                      | Service Analyzer V Notable Eve  | ents Review Glass Ta                         | ables Deep Dives   | Multi KPI Alerts                         | Search 🗸                  | Configur  | re∨ Produc                                      | t Tour   |   | IT Service Intelligence   | vas critical (Health Sc | ore=0.0) at 2017 | 7-09-1 |
| Database Events: status:                                      | Default Policy 🜑 Smart Mode ? vas critical (Health Score=0.0) at 2017-09-1  |  |  |  |                           |   |   |  |   |   |                         |                  |        |
| Database Events: status:                                      | Applies to events which fail to meet the criteria of any other active policy vas critical (Health Score=0.0) at 2017-09-1 |  |  |  |                           |   |   |  |   |   |                         |                  |        |
| Database Events: status:                                      | Grouping Factors  |  |  |  |                           |   |   |  |   |   | vas critical (Health Sc | ore=0.0) at 2017 | 7-09-1 |
| Database Events: status:                                      | Machine learning is used to group n   | otable events based on th                    | ne similarity of text and ca                                 | ategorical factors.                      |                           |   |   |  |   |   | vas critical (Health Sc | ore=0.0) at 2017 | 7-09-1 |
| Nagios Service Check: cł                                      | ✓ Adjust the importance of each fac   | ctor?  |  |  | 0                         | Froup types   | for Last 24 hou                                 | s 🗸  |   |   | tatus check_ntp_time    | OK on webserve   | er-01  |
| Nagios Service Check: ch                                      | 84 Fields Selected to analyze ev  | ent similarity                               |  |  | <i>i</i> >                | Count 0<br>3456   | Summary 0<br>3456/10000                         | Similar Texts 0<br>None  | Similar Fields<br>status=1; n_severity=2; c   | rig_index=itsidemo; entity_lookup_field   |                         | on webserver.0   | 1      |
| Hagios Scivice Officer. of                                    | Textual Similarity ?  | 0 Less Important                             |  | 1<br>More Important                      |                           | events are ong_sourcetype=nagiosserviceperf; n_health_status=OK; ong_sexAtUS CHECK_GHCP UK ON WED<br>grouped<br>because |   |  |   |   | on webserver o          | •                |        |
| Nagios Service Check: ch                                      | Categorical Similarity ?  | 0  |  | 1  |                           | they share 9<br>(relatively<br>high)  |   |  |   |   | tatus check_ntp_time    | OK on webserve   | er-02  |
| Nagios Service Check: cł                                      |   | Less Important                               |  | More Important                           |                           |   | common<br>field values.                         |  |   |   | tatus check_dhcp OK     | on webserver-0   | 2      |
|   | $\sim$ Split events by field?   |  |  |  | • •                       | 4573  | 4573/10000<br>events are<br>grouped             | None   | orig_sourcetype=newreli<br>event_identifier_hash=3a   | etype=newrelic_account; entity_lookup_field=n_insta<br>tifier_hash=3a3b069e7920fd1d2dff7d243e9281b75a | асн<br>с0               |                  |        |
|   | Split events into multiple groups by  | %service%<br>Separate multiple fields by co  | omma   |  |                           |   | because<br>they share 4<br>(average)<br>common  |  |   |   |                         |                  |        |
|   | ∽ Break group ?   |  |  |  | 0 >                       | 904   | field values.<br>904/10000                      | n_message=OK:  | reason=OK: Received 1 [   | DHCPOFFER(s), max lease time = 600 s  | ec                      |                  |        |
|   | $\checkmark$ If the flow of events into the gr  | oup paused for                               |  |  |                           | events are Received 1<br>grouped DHCPOFFEI<br>because max lease ti<br>thou obcor = 500 cost                             |   |  | Received 1 omd_site=SJC; result=OK; name=check_dhcp;<br>DHCPOFFER(s), event_identifier_hash=d0087cc50fa76896dbfbb45f4fa4c458f5ddf<br>max lease time = 600 ccc |   | dt                      |                  |        |
|   | If the flow of events into the group  | p paused for 7200<br>In seconds              |  |  |                           |   | (relatively<br>high)<br>common<br>field values. | - 000 Sec.   |   |   | _                       |                  |        |
|   | Cancel 🛕 Cannot save ACE  | policy until group genera                    | tion has been completed                                      |  |                           |   |   |  |   | Save  |                         |                  |        |
|   |   | (cart.do?ac                                  | tion=view&i+   | OmTderer                                 |                           |   |   |  |   | splunk  |                         | nf201            | 7      |
| 04 720 "http://but  | tercup-shopping.com<br>tp://buttercup-shop<br>cup-shopping.com/ca<br>cup-shopping.com/ca                                  | ping.com/ca<br>art.do?actio<br>7 HTTP 1.1"   | tegory.scree<br>n=purchase&i<br>200 2423 "ht<br>ST-18&produc | temid=EST-G<br>temid=EST-<br>tp://butter | &produc<br>_1d=GIF<br>26∏ | t_ida<br>TS: da   | FI-SW-0<br>Mozilla                              | 12 10  |   |   |                         | 11201            |        |
| 318 "http://butte<br>id=SURPRISE&JSESSI<br>id=SURPRISE&JSESSI | ONID=SD95L<br>action=changequant<br>"GET /oldlink?ite   | 1ty&item.dee<br>m_id=EST-6&J<br>screen?cate8 | SESSIONID=SU<br>ory_1d=FLOW<br>egory.screen                  | Category                                 | FF9 HT1<br>NID-SD8        | SSION 1   | BLUSSO 3  | 1111<br>1111<br>1111<br>1111<br>1111<br>1111<br>1111<br>1111<br>1111 | Na 20   |   |                         |                  |        |

# Real Use Case Example

Predictive Analytics in Real-time



## Lets Get Others Involved

ChatOps - Anyone Slack?



## Slack

#### ► Slack

- Instant Message
- Easy Channel Setup
- Fast time to create incident working groups
- Other Products just as easy HipChat, Skype etc...

Screen?product id=FL-DSH-01&JSESSIONID=5D5

|           | splunk-usergrou ~ 🛆<br>• nsmalley                                  | #general<br>☆   & 2,903   & 31   anything ⓒ ⓒ Q Search @ ☆ : 🕆 2   |
|-----------|--|--|
| ¥1        |  | And more   |
| ₩2        | Channels ↔<br>a amex-xsr-uplift<br>drivetime<br># general          | Today          pyro_wood 9 7:27 AM         @yorokobi thanks!   |
| M<br>#3   | ⊜ j_splunk_da<br># <b>n00bs</b><br># random                        | yorokobi 2:28 AM<br>(with the default values) You're effectively doubling the number of<br>connections to that box from the forwarders. fry 5:28 AM  |
| RS<br>#4  | Direct Messages  | <pre>@pyro_wood just throw the DS on a VM somewhere. pyro_wood P 7:30 AM @fry yes, I would do this normally</pre>  |
|           | <ul> <li>aaron.mciver</li> <li>denis</li> <li>jcoates 🚆</li> </ul> | Just asking if it could be done without a seperate vm         fry \$ 7:30 AM         it can. i had it on a search head for a long time         but that search head started estime protive busy wearing SUL / Eat Sec. ( DS) |
| <b>#5</b> | Ξα   | + Message #general @ 3   |



## ChatOps = Splunk & Slack

#### Slack Setup

- Install the Alert Action
  - <u>https://splunkbase.splunk.co</u> <u>m/app/2878/</u>
- For Core Searches Configure as alert action in Alerts
- For ITSI
  - Enable Slack in notable\_event\_actions.conf
- Lets add it to our to our Policy
  - Configure -> Notable Event Aggregation Policy -> Create New

|   |   |                           |   | _   |   |   |
|---|---|---------------------------|---|---|---|---|
| reate New Policy  | Fibering Criteria Action Rules                      | Policy Info Dor           | Next >  |   |   |   |
| iltering Criteria<br>eate fitering criterion to group notable | e events  |                           |   |   |   |   |
| rclude the events if 7  |   |                           |   |   |   |   |
| owner match<br>+ Add Rule (AND)                               | es v Splunk Bot X                                   |                           |   |   |   |   |
| + Add Rule (CR)<br>Ipilt events by field <sup>†</sup>         | Create New Polic                                    | Filtering Criteria        | Action Rules Policy Info Done   | ext >   |   |   |
| t events into multiple groups by 0                            | Action Rules<br>Create action rules upon this group |                           |   |   |   |   |
| <ul> <li>If the following event occurs</li> </ul>             | $\sim$ If this group existed for 10, the            | n on all events in this g | roup  |   |   |   |
| f the following event occurs ~                                | If this group existed for $\vee$<br>O and if        | 10                        | In Seconds  | Then slack v Configure on a O and                         | all events in this group $\backsim$       |   |
|   | + Add Rule  | &                         | Splunk APP 9:00 AM ☆<br>Digital Errors Detected                                 |   |   |   |
|   |   |                           | 07-19-2017 08:59:22 AN<br>limit=####&startDate=##<br>rv:11.0) like Gecko 0.009  | 1 c2b:nginx-access GE<br>### 500 Mozilla/5.0 (V           | T /api/v1/mr/accour<br>Vindows NT 10.0; W | nts/####/obligations?<br>OW64; Trident/7.0; |
|   |   |                           | Click here for full results:<br>sid=scheduler_andpbHNv<br>c27be_at_1500480000_3 | https://itsi-republicser<br>/bjEwQHJlcHVibGljc2\<br>36881 | vices.splunkcloud.co<br>/ydmljZXMuY29tit  | m/app/itsi/@go?<br>siRMD57952f7c075d        |



# Real Use Case Example

Predictive Analytics in Real-time



## Wrap Up

- ► ITSI allows for Rapid Service, KPI, Entity creation and alerting
- Machine Learning through MLTK provides a repeatable ability to Predict Service Health
- IT Service Intelligence Event Analytics enables teams to cluster Notable events together to have 1 Actionable Alert
- Splunk IT Service Intelligence provides extendable capabilities to provide immediate notification to Chat Groups, Ticketing system and other communication platforms to improve the Mean Time to Remediate Availability Impacting Situations

## I fell asleep. Where else can I read about this?

https://www.splunk.com/blog/2017/08/28/itsi-and-sophisticated-machine-learning.html



## Go see these Talks!

This is where the subtitle goes

#### David Vueve

- Splunk Ninja Skills
- 190+ slides of SPL and guidance.

- Xander and (fred?)
  - Deep dive into MLTK
  - MLTK API for importing algorithems

- Phillip Drieger
  - DGA Analysis
  - End to End MLTK example

- Making ML Solutions
  - Deep dive into ML process
  - Customer use cases
     explored



## Join the Pony Poll



### ponypoll.com/\*\*\*

indicategory\_id=GIFTs&JSESSIONID=SDISL4FF10ADFF10 HTTP 1 induct.Screen?product\_id=FL-DSH-01&JSESSIONID=SDSSL7F6ADFF3 /oldlink?item\_id=EST-26&JSESSIONID=SDSSL9F71ADFF3 HTTP 1.1" // recompleted:

200 1318

URPRISE&JSESSI





## **Questions?**

**Please Feel Free** 



#### Want to Learn More About ITSI at .conf2017?

- Ready, Set, Go! Learn From Others The First 30 Day Experiences of ITSI Customers: Tuesday, September 26th, 201712:05 PM- 12:50 PM Room Salon C
- Splunk ITSI Overview: Tuesday, September 26th, 2017 1:10 PM-1:55 PM Room 147 AB
- **PWC: End-to-End Customer Experience:** Tuesday, September 26th, 2017 2:15 PM-3:00 PM Room 143ABC
- RSI: Operational Intelligence: How to go From Engineering to Operationalizing IT Service Intelligence Where the Rubber Meets the Road: Tuesday, September 26th, 2017 2:15 PM-3:00 PM Room147AB
- Cardinal Health: Ensuring Customer Satisfaction Through End-To-End Business Process Monitoring Using Splunk ITSI:
  - Tuesday, September 26th, 20173:30 PM-4:15 PM Room143ABC
- ITSI in the Wild Why Micron Chose ITSI and Lessons Learned From Real World Experiences: Tuesday, September 26th, 2017 4:35 PM- 5:20 PM Room Salon C
- Event Management is Dead. Time Series Events are the Means to the End, not the End Itself. See How Event Analytics is Revolutionizing IT: Wednesday, September 27th, 201711:00 AM-11:45 AM Ballroom C
- **Triggering Alerting (xMatters) and Automated Recovery Actions from ITSI:** Wednesday, September 27th, 2017 1:10 PM- 1:55 PM Room Salon C
- Leidos Our Journey to ITSI: Wednesday, September 27th, 2017 2:15 PM-3:00 PM Room147AB
- How Rabobank's Monitoring Team Got a Seat at the Business Table by Securing Sustainability on Competitive Business Services Build on Splunk's ITSI: Wednesday, September 27th, 2:15-3:00pm Room 147AB
  - Here Comes the Renaissance: Digital Transformation of the IT Management Approach: Wednesday, September 27th, 2017 3:30 PM-4:15 PM Room Salon C
  - The ITSI 'Top 20' KPI's: Thursday, September 28th, 2017 10:30 AM-11:15 AM Room Salon C
- Automation of Event Correlation and Clustering with Machine Learning Algorithms An ITSI Tool: Thursday, September 28th, 2017 11:35 AM- 12:20 PM Room Salon C
- Event Management is Dead. Time Series Events are the Means to the End, not the End Itself. See How Event Analytics is Revolutionizing IT:
  - Thursday, September 28th 11:35 AM 12:20 PM in Ballroom B

oduct.screen?product 1d=FL-DSH-01&JSESSIONID=SD

- IT Service Intelligence for When Your Service Spans Your Mainframe and Distributed ITSI:
  - Thursday, September 28th, 2017 1:20 PM-2:05 PM Room Salon C



Wednesday September 27<sup>th</sup>, 2017

Thursday

September

28<sup>th</sup>, 2017

Tuesday

September

26<sup>th</sup>, 2017

## **Hidden Gems**

- Remember that the MLTK will see any numeric field as a number, not as an entity, so if we want to use date\_hour (a number between 0 and 23) we need to change the value into a string (|eval date\_hour\_string= date\_hour ."."."\_t" for example
- Remember we are predicting the future, so we need to move the target of our regression through time.
- Remember that your data retention policy means you will lose past data at some point – consider making two models, one with partial\_fit and one without and comparing the results.