Defend Against Malicious Insiders Using Splunk Enterprise Security, Splunk's Machine Learning Toolkit, and Statistics

Jason Barnette & Bryan Thiry
Lockheed Martin

splunk> .conf19

.conf19

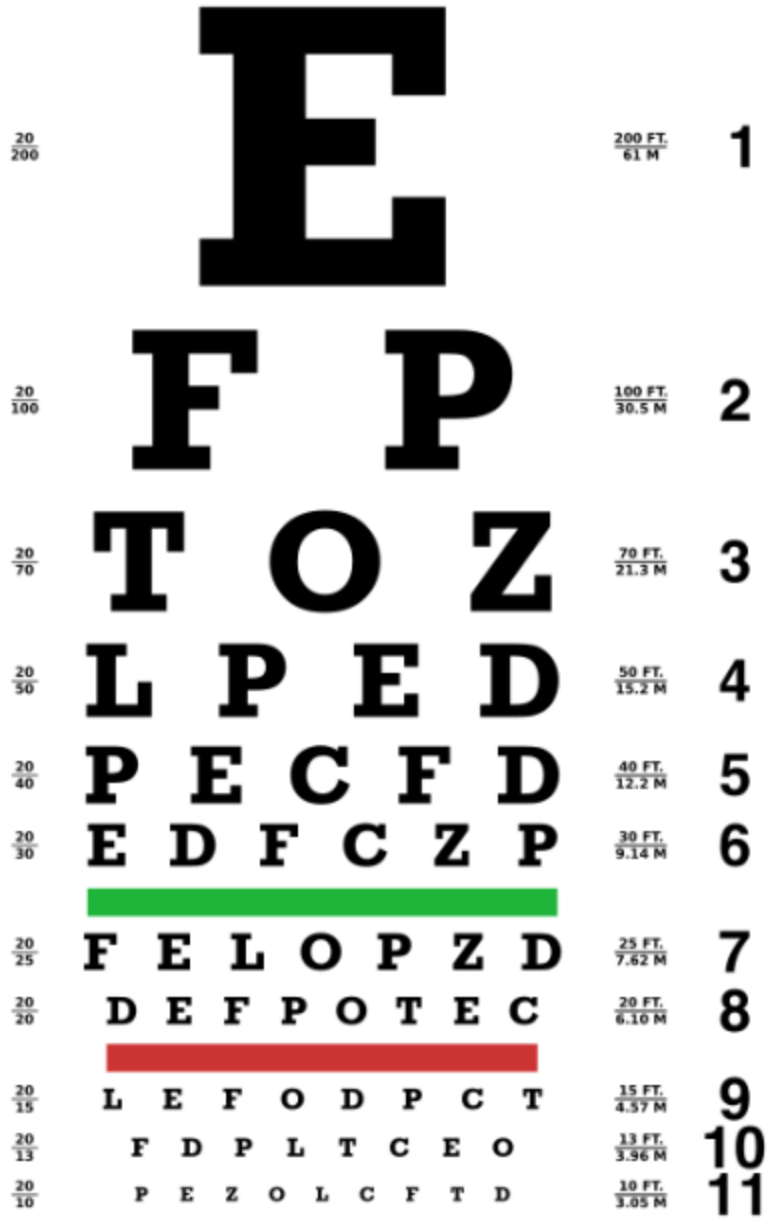
splunk>

Jason Barnette

LM-CIRT Insider Threat Lead | Lockheed Martin

Bryan Thiry

LM-CIRT Insider Threat Analyst Sr. | Lockheed Martin



splunk> .conf19

Forward-Looking Statements

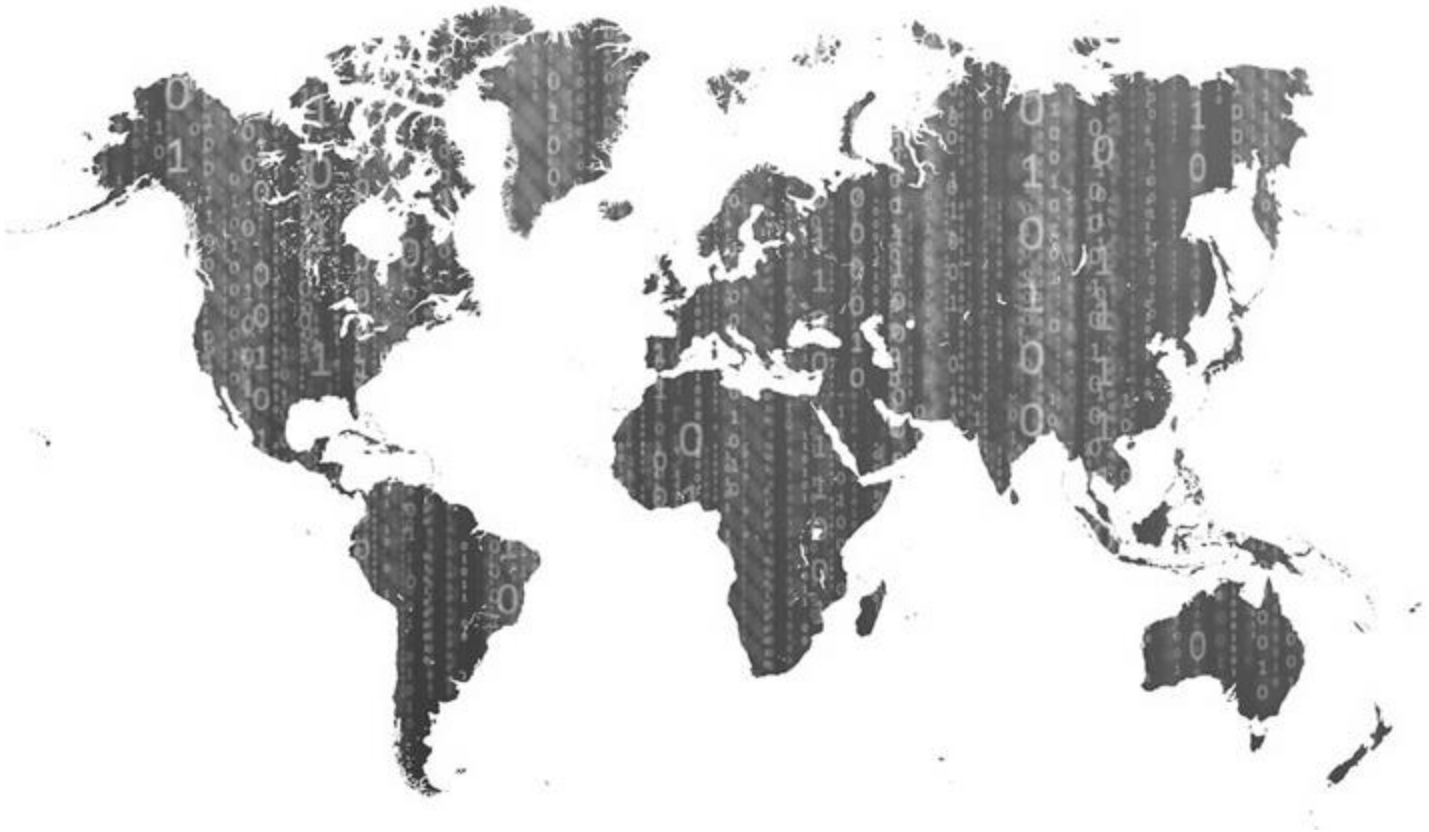
During the course of this presentation, we may make forward-looking statements regarding future events or plans of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results may differ materially. 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, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

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 functionalities described or to include any such feature or functionality in a future release.

Agenda

- Foundational Elements
- The Problem
- Recipe For Success
- Where We Started
- Risk Score Framework
- Our Path Forward

Foundational Elements



Population 105k

Countries > 70

SIEM Splunk>

Data Ingest 12TB Per Day

splunk> .conf19

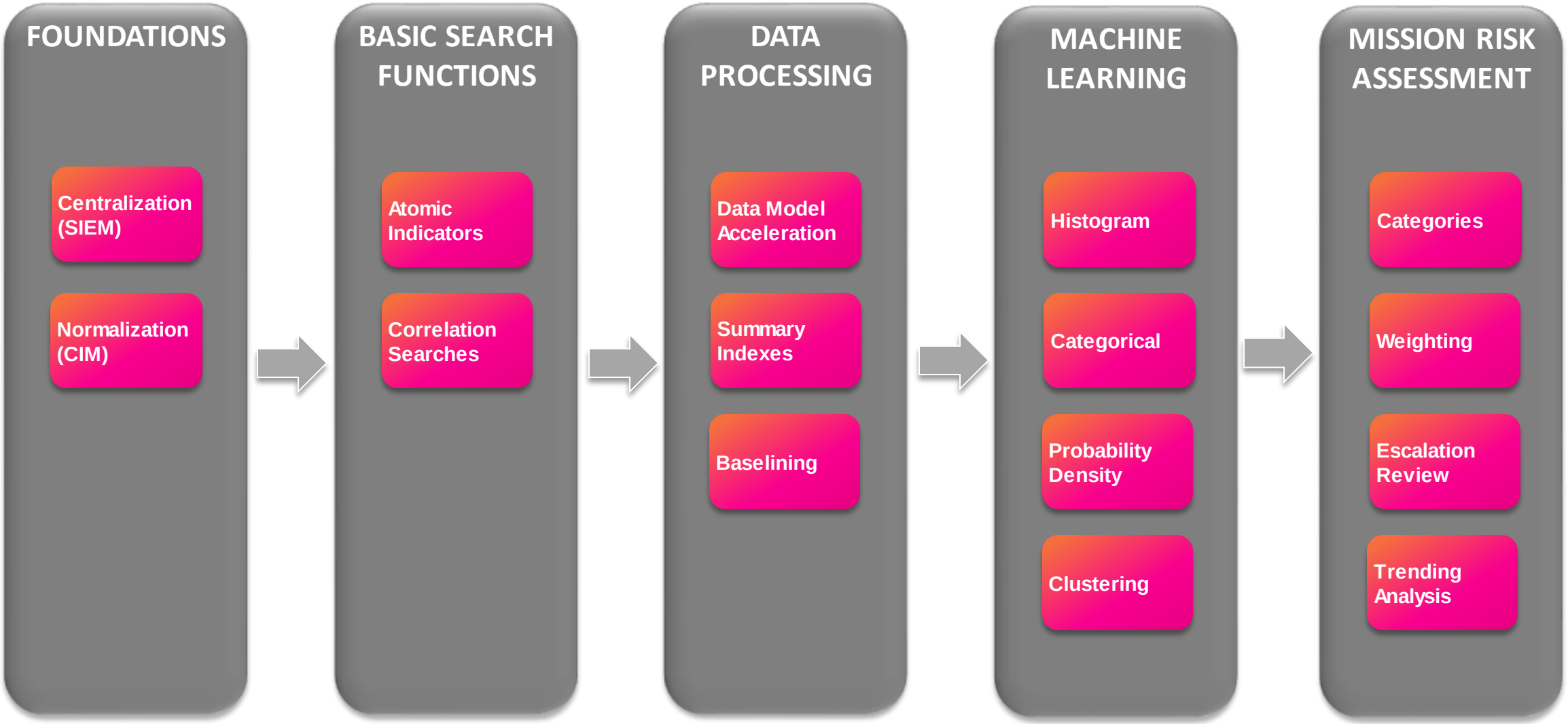
“The goal isn't to react well, or even to track well, it's to anticipate, to see these things coming and step in before the disaster occurs and mitigate it”

Chris Inglis: Former Deputy Director of the NSA

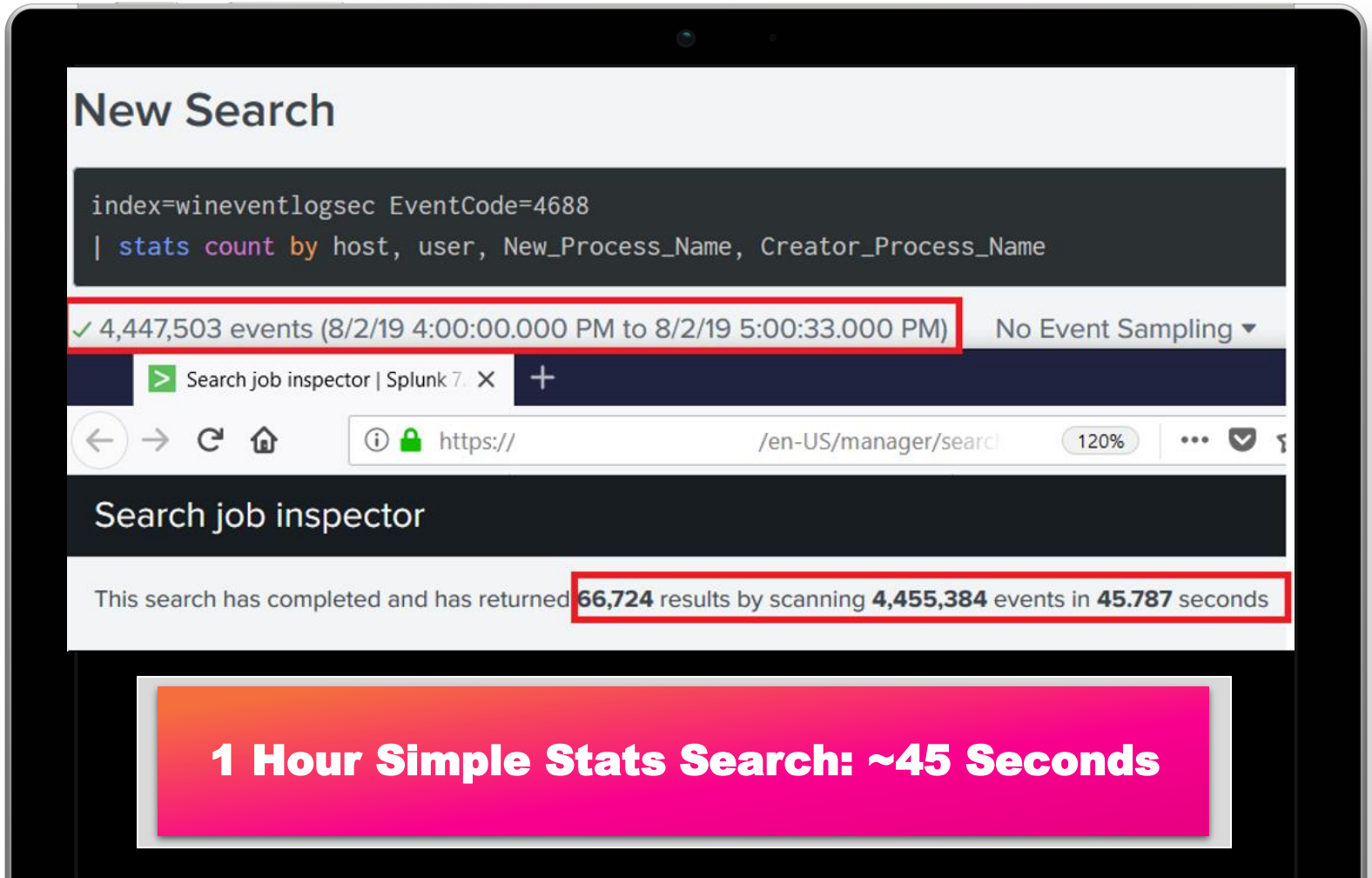
The Problem

Challenge:
Finding Known Bad

Recipe For Success

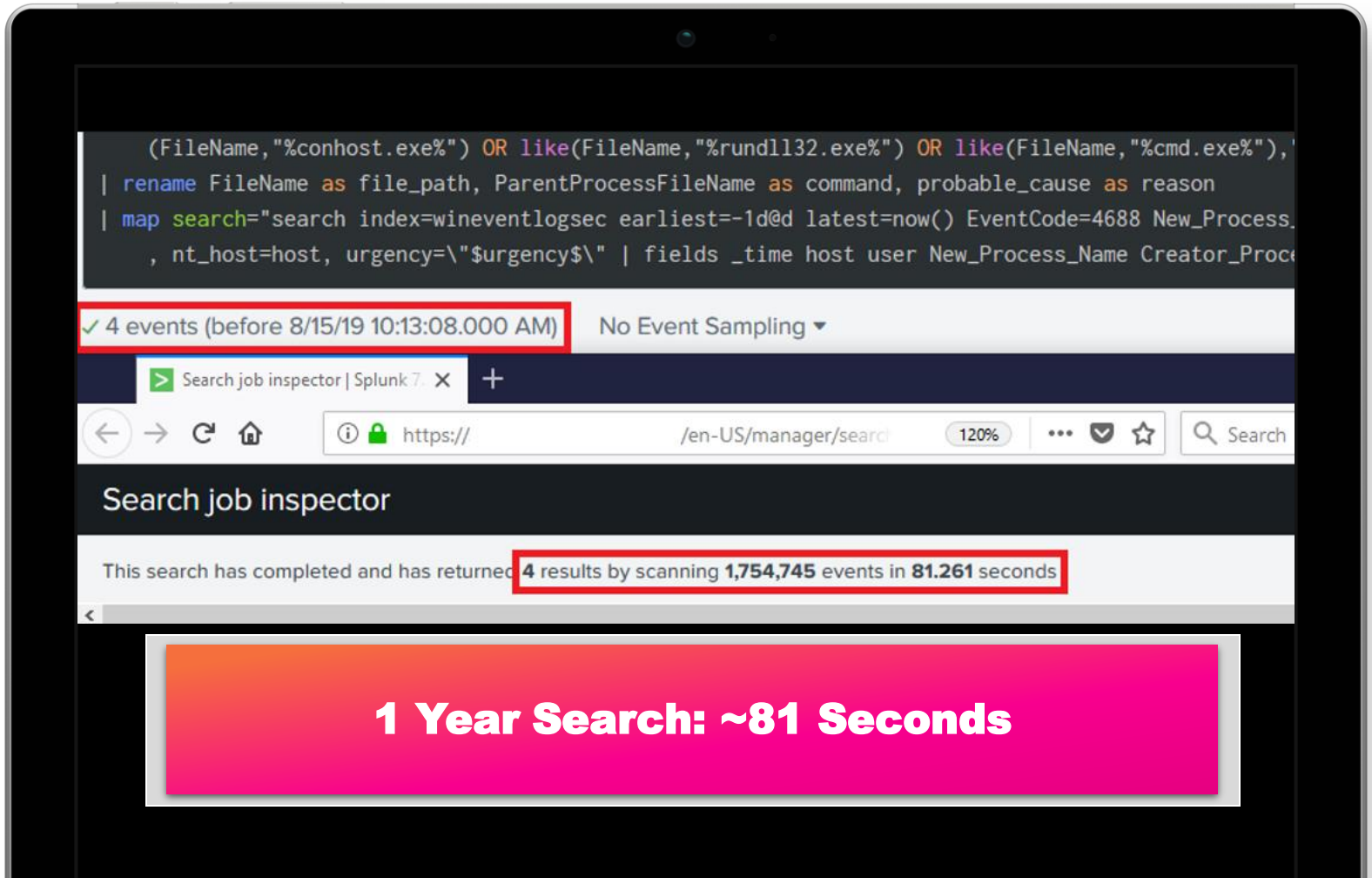


Where We Started: Pre-Baselining



- Attempt to run basic stats on the full index
  - Slow Searching
  - Unable to do categorical outlier scoring
  - Inefficient resource utilization
  - Limited testing datasets

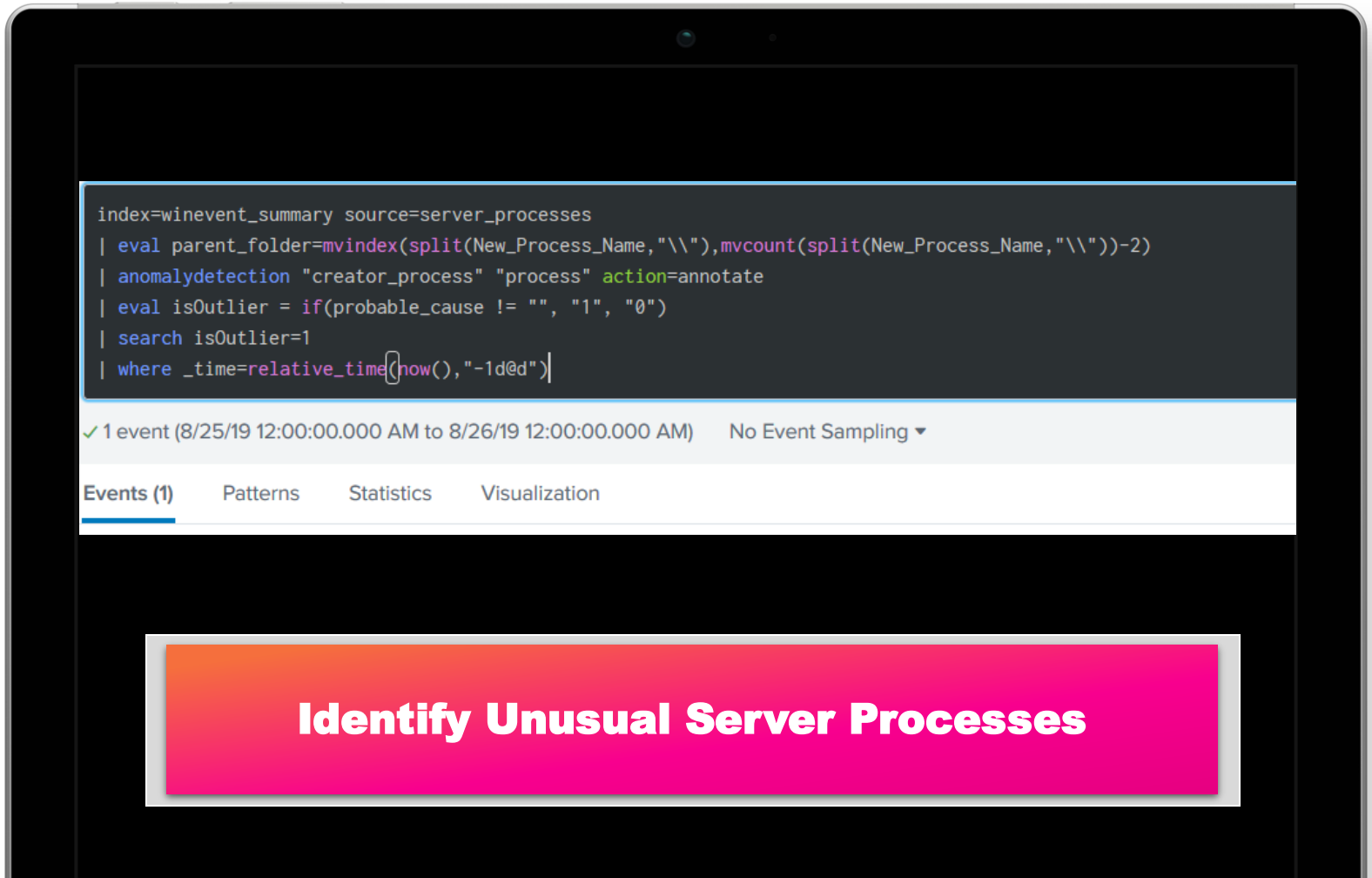
Where We Started: Post-Baselining



- Preprocess, Condense, and Score
  - Higher fidelity results
  - Minimal resource utilization
  - Greater search complexity

splunk> .conf19

Where We Started: MLTK



- Anomaly Detection
  - Detecting outliers with ‘anomalydetection’
  - Continuous updating of baselines
  - Usage of simple statistics
    - Standard Deviation
    - Zscore

splunk> .conf19

MLTK Models

Probability Density
(Peer Grouping)

```
index=     laikaboss type=outbound
| rename senders{} as senders
| eval time=strftime(_time, "%A") . " - " . strftime(_time,"%H") . ":00 Hour", app="mail"
| stats sum(size) as sum_size by senders, time, app
| fit DensityFunction sum_size by time show_density=true into app:outbound_email_volume_density
```

Last 4 hours

✓ 234,772 events (8/26/19 9:41:10.000 AM to 8/26/19 1:41:10.000 PM) No Event Sampling Job Smart Mode

Events Patterns **Statistics (44,055)** Visualization

20 Per Page Format Preview

< Prev 1 2 3 4 5 6 7 8 ... Next >

senders	time	app	sum_size	BoundaryRanges	IsOutlier(sum_size)	ProbabilityDensity(sum_size)
hidden	Monday - 13:00 Hour	mail	12872	2630223.6333:Infinity:0.01	0.0	1.713302362547835e-06
hidden	Monday - 10:00 Hour	mail	3800	4312226.7762:Infinity:0.01	0.0	1.064174782705401e-06
hidden	Monday - 13:00 Hour	mail	118856	2630223.6333:Infinity:0.01	0.0	1.4230901442804624e-06
hidden	Monday - 10:00 Hour	mail	5654	4312226.7762:Infinity:0.01	0.0	1.062069656345263e-06
hidden	Monday - 10:00 Hour	mail	19674	4312226.7762:Infinity:0.01	0.0	1.0462848619683298e-06
hidden	Monday - 13:00 Hour	mail	9206	2630223.6333:Infinity:0.01	0.0	1.7243365482930175e-06

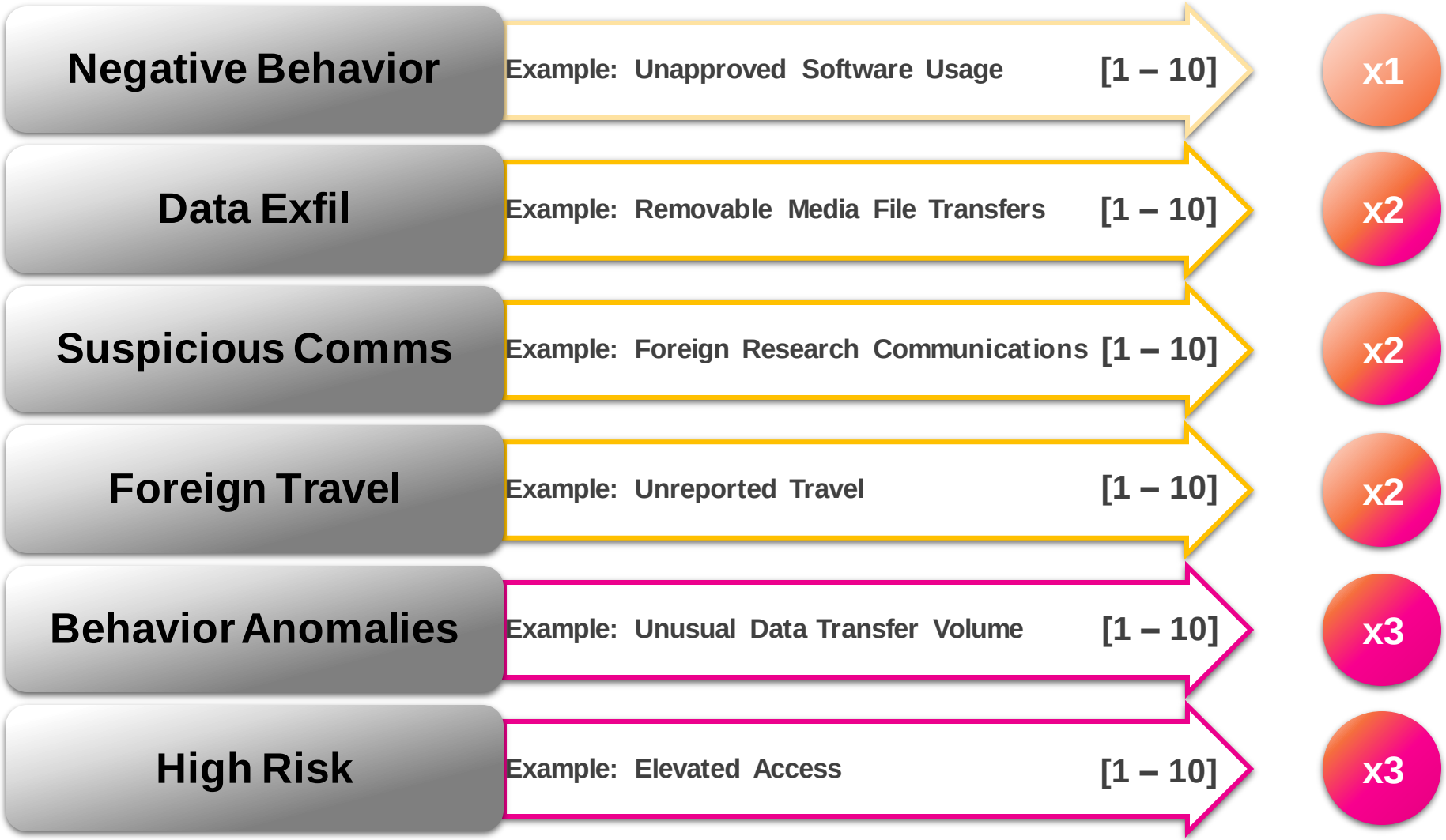
splunk> .conf19

Current State: Risk Score Framework

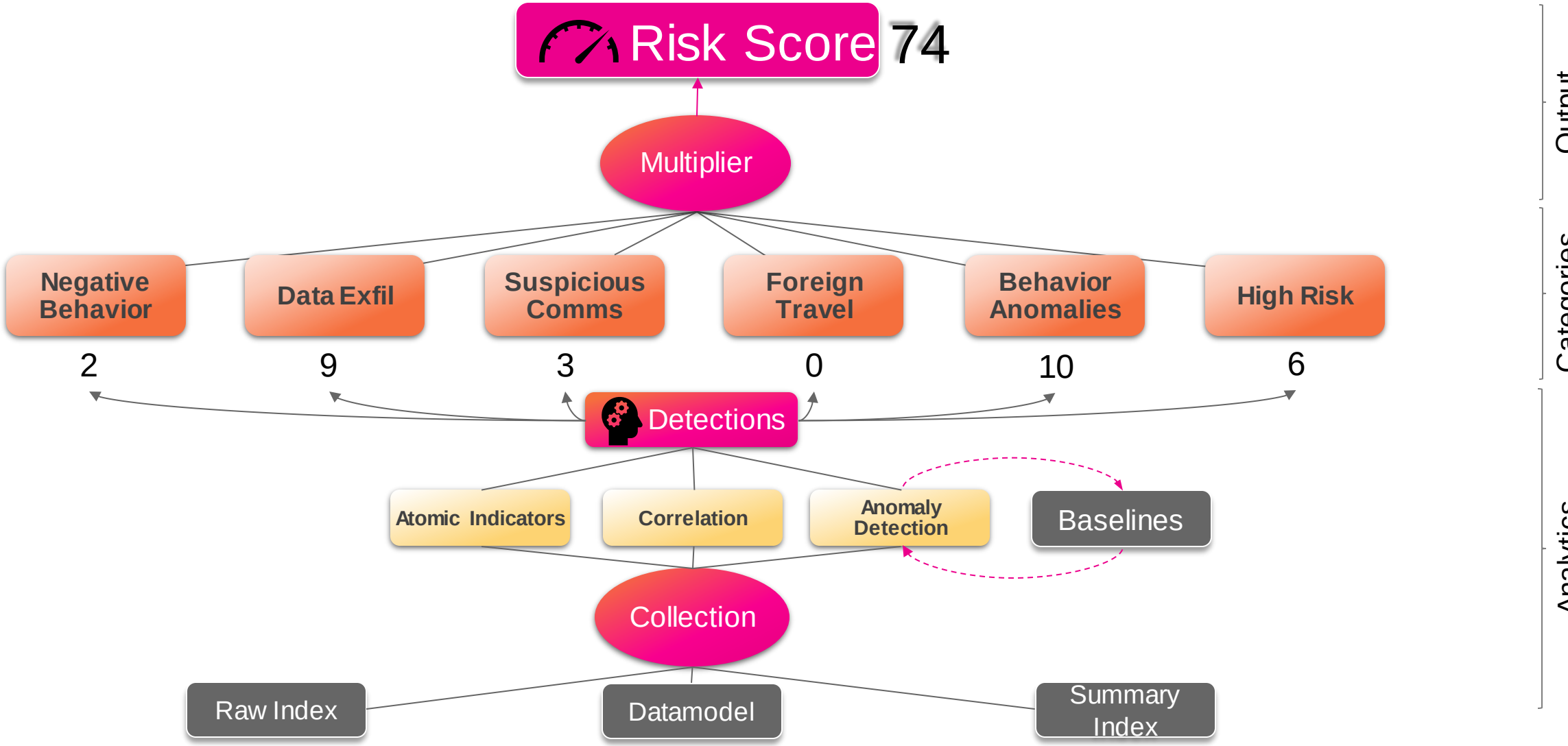
- Category Identification
- Category & Indicator Weighting
- Aggregation of Data
- Escalation Process & Review
- Trending Analysis

Aligned to Mission Risk Assessment

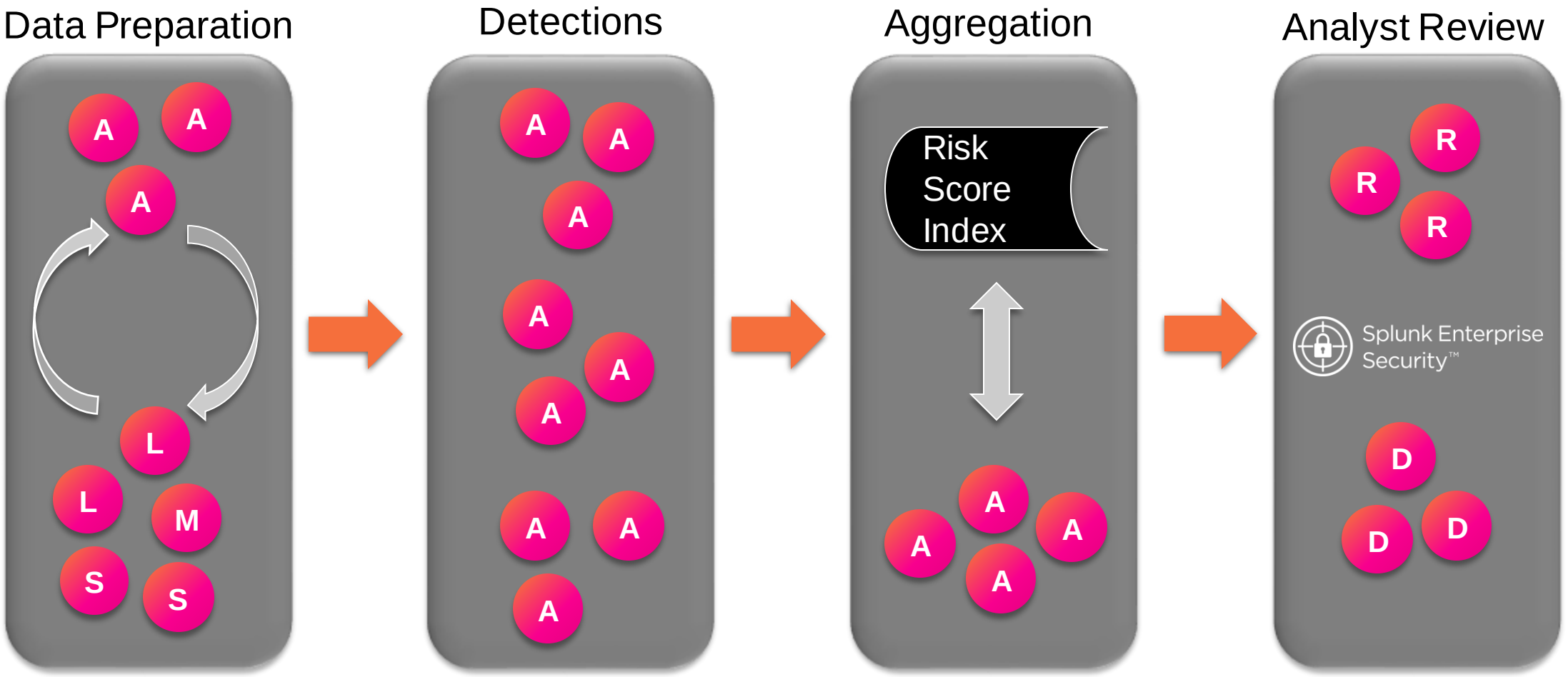
CURRENT STATE: CATEGORY IDENTIFICATION & WEIGHTING



CURRENT STATE: SCORING MODEL



RISK FRAMEWORK – IN-DEPTH REVIEW



Risk Score Framework

Detection: Sensitive email volume sent externally

```
index=_laikaboss attachments{}=* disposition_flags{}=sensitive_info_tag
| fields senders{} size recipients{}
| bin span=1d _time
| rename senders{} as senders, recipients{} as recipients
| lookup webmail_lookup.csv user as senders OUTPUT webmail_matches
| eval webmail=if(like(webmail_matches,"%" + recipients + "%"),"Yes","No")
| stats sum(size) as daily_size_bytes by senders, _time, webmail
| eval totalSize=round(daily_size_bytes/1024/1024,2), weight=if(totalSize>20,10,weight), weight=if(totalSize>15 AND totalSize<20,9,weight), weight=if
   (totalSize>10 AND totalSize<15,8,weight), weight=if(totalSize>5 AND totalSize<10,7,weight), weight=if(totalSize>1 AND totalSize<5,6,weight), weight
   =if(totalSize<1,5,weight), reason="sizeMB: " + totalSize + ", webmail: " + webmail, weight=if(webmail="Yes",weight+3,weight), weight=if(weight>10,10
   ,weight)
| lookup lm_user_info_lookup mail as senders OUTPUT sAMAccountName as user
| eval detectionName="sensitive_email_volume", category="data_exfil", date=strftime(relative_time(now(),"-1d@d"),"%m/%d/%Y")
| sort - weight,webmail
| dedup user
| eval user="hidden"
| table date detectionName weight category user reason
| collect index=cirt source=risk_scoring
```

Yesterday

✓ 377 events (8/25/19 12:00:00.000 AM to 8/26/19 12:00:00.000 AM) No Event Sampling Job Smart Mode

Events Patterns **Statistics (261)** Visualization

20 Per Page Format Preview Prev 1 2 3 4 5 6 7 8 ... Next

date	detectionName	weight	category	user	reason
08/25/2019	sensitive_email_volume	10	data_exfil	hidden	sizeMB: 7.82, webmail: Yes
08/25/2019	sensitive_email_volume	10	data_exfil	hidden	sizeMB: 100.74, webmail: Yes
08/25/2019	sensitive_email_volume	10	data_exfil	hidden	sizeMB: 23.20, webmail: Yes

splunk> .conf19

Risk Score Framework

Aggregation

```
index=     source=risk_scoring
| where strptime(date,"%m/%d/%Y")=relative_time(now(),"-1d@d")
| sort 0 -user,detectionName,weight
| dedup user,detectionName
| stats max(weight) as catWeight by category, user
| eval catWeight=if(category="behavior_anomalies",catWeight*3,catWeight), catWeight=if(category="data_exfil",catWeight*2,catWeight), catWeight=if
   (category="high_risk",catWeight*3,catWeight), catWeight=if(category="foreign_travel",catWeight*2,catWeight), catWeight=if(category
   ="suspicious_commmunications",catWeight*2,catWeight), trend=if(category="risk_trend",catWeight,trend), trend=if(trend<=1,0,trend), catWeight=if
   (category="risk_trend" AND catWeight<=1,0,catWeight)
| stats sum(catWeight) as score, sum(trend) as trend by user
| eval score=score-trend, score=if(score<trend,round((trend+score)/2,1),score)
| sort 0 -score
| eval date=strftime(relative_time(now(),"-1d@d"),"%m/%d/%Y")
| table date user score
| collect index=     source=daily_risk_scores
| outputlookup       daily-risk-scores.csv
| head 10
| map search="search index=     source=risk_scoring user=$user$ | where strptime(date,\"%m/%d/%Y\")=relative_time(now(),\"-1d@d\") | table user
   detectionName weight reason category  | eval reason=detectionName + \" - \" + reason+ \" | weight: \" + tostring(weight) + \" | category: \" +
   category | stats values(reason) as summary by user | eval summary=mvappend(\"Risk Score: \" + \"$score$\",summary) " maxsearches=10
| eval urgency="high", _raw=summary
```

✓ 2,702 events (8/7/19 6:32:01.000 PM to 8/8/19 6:32:01.000 PM) No Event Sampling ▾ Job ▾

Events Patterns **Statistics (10)** Visualization

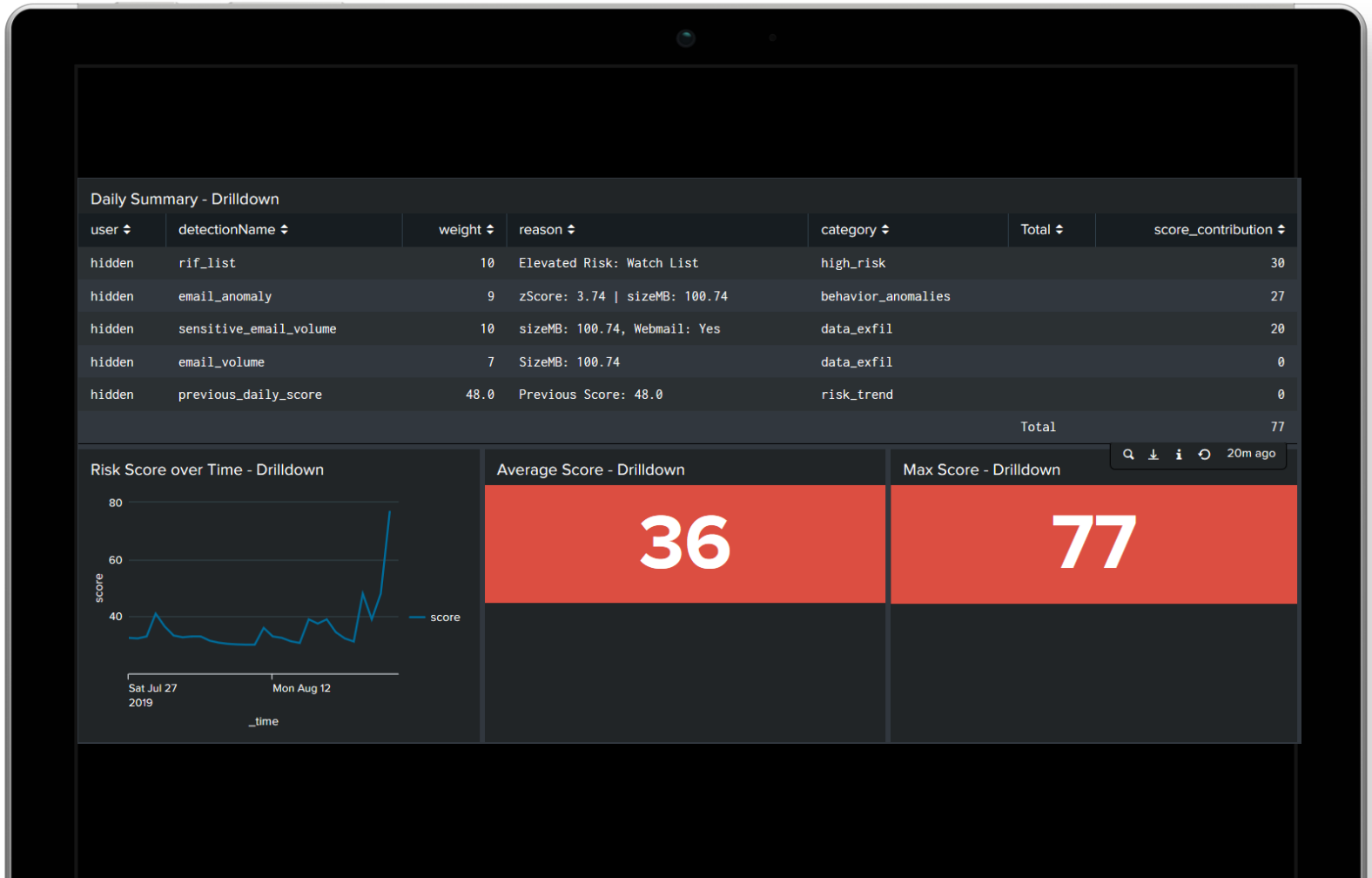
splunk> .conf19

Risk Score Framework: Dashboard

- High Level Statistics
  - Highest Daily Risk Score
  - Daily Average
- Top 10 Risk Scores
- Top 10 User Info
  - Drilldown: Further Information

splunk> .conf19

Risk Score Framework: Dashboard



- In-depth User Review
  - Detection Name
  - Weight
  - Reason
  - Category
  - Score Contribution
- Trending Analysis
- Average Score
- Max Score

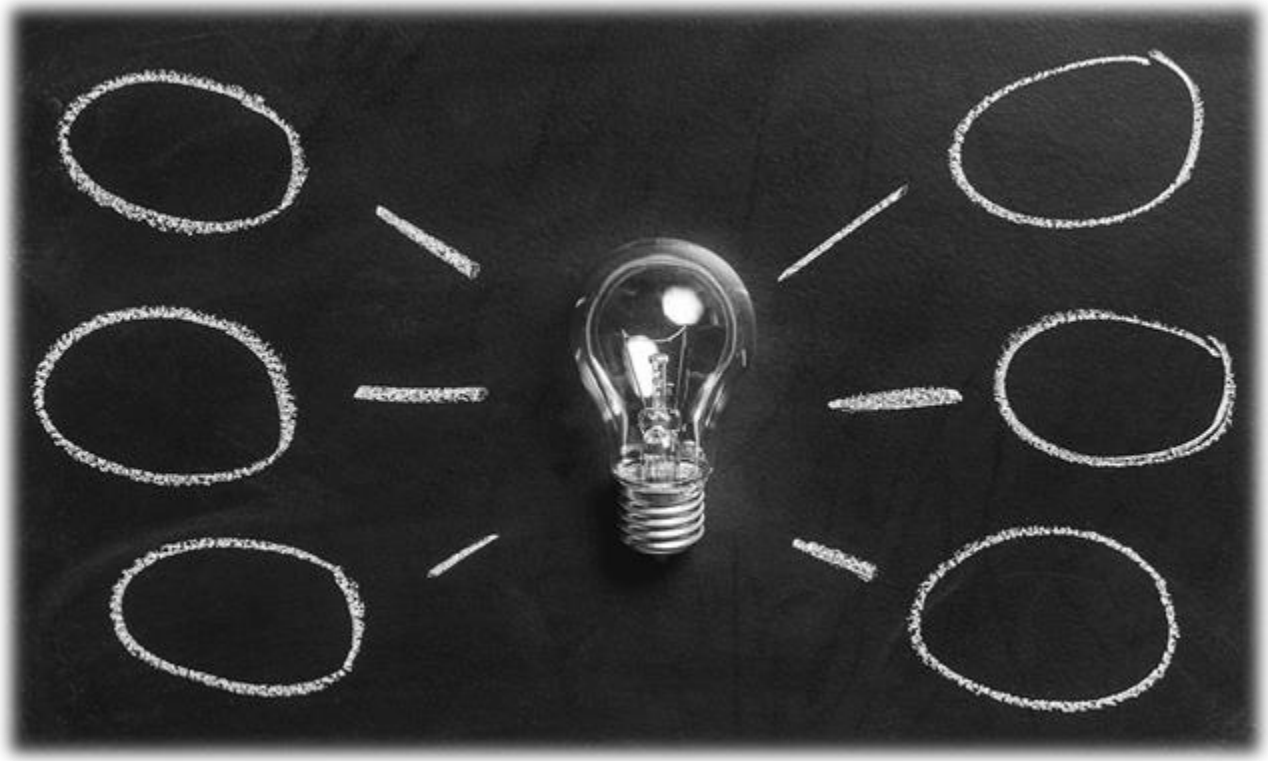
splunk> .conf19

RISK FRAMEWORK – ESCALATION PROCESS & REVIEW

OUR PATH FORWARD

- Continuous Indicator Advancement
- Risk Score Value Normalization
- Process Refinement
- Enterprise Security Investigations
- Splunk UBA

- Feedback from Community



QUESTIONS?

.conf19

splunk>

splunk>

EXAMPLE SEARCHES

HISTOGRAM EXAMPLE [CATEGORICAL OUTLIER]

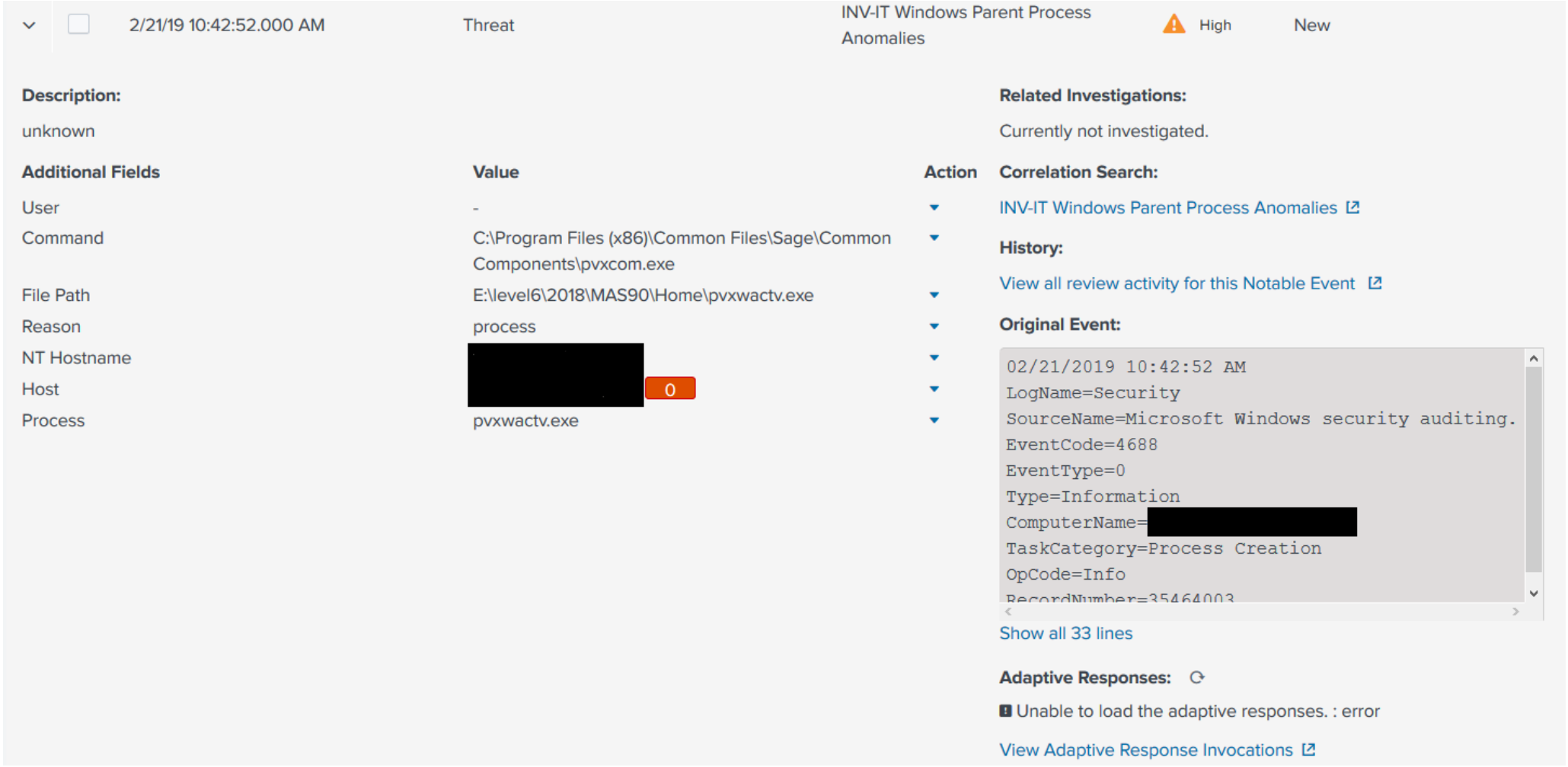
- Identify unusual parent/child processes on process creation
  - Utilize Enterprise Security on detection
    - Set urgency
    - Retrieve raw 4688 event for the ES Incident Review

INV-IT Windows Parent Process Anomalies

```
1  index=     source=observed_processes
   | eval parent_folder=mvindex(split(New_Process_Name,"\\"),mvcount(split(New_Process_Name,"\\"))-2)
2  | anomalydetection method=histogram "creator_process" "process" action=annotate
   | eval isOutlier = if(probable_cause != "", "1", "0")
   | search isOutlier=1
   | where _time=relative_time(now(),"-1d@d")
   | rename New_Process_Name as FileName, Creator_Process_Name as ParentProcessFileName
   | eval end_time=_time, _time=now(), urgency="high", urgency=if(like(ParentProcessFileName,"%lsass%") OR like(ParentProcessFileName,"%svchost%") OR like(ParentProcessFileName,"%crss.exe%") OR like
      (ParentProcessFileName,"%nvclp%") OR like(ParentProcessFileName,"%svhost%") OR like(ParentProcessFileName,"%services.exe%") OR like(ParentProcessFileName,"%wininit%") OR like(ParentProcessFileName
      ,"%lsm.exe%") OR like(ParentProcessFileName,"%smss.exe%") OR like(ParentProcessFileName,"%explorer.exe%") OR like(ParentProcessFileName,"%conhost.exe%") OR like(ParentProcessFileName,"%rundll32
3     .exe%") OR like(ParentProcessFileName,"%cmd.exe%"),"critical",urgency), urgency=if(like(FileName,"%lsass%") OR like(FileName,"%svchost%") OR like(FileName,"%crss.exe%") OR like(FileName,"%nvclp%")
      OR like(FileName,"%svhost%") OR like(FileName,"%services.exe%") OR like(FileName,"%wininit%") OR like(FileName,"%lsm.exe%") OR like(FileName,"%smss.exe%") OR like(FileName,"%explorer.exe%") OR
      like(FileName,"%conhost.exe%") OR like(FileName,"%rundll32.exe%") OR like(FileName,"%cmd.exe%"),"critical",urgency)
   | rename FileName as file_path, ParentProcessFileName as command, probable_cause as reason
4  | map search="search index=winevent       earliest=-1d@d latest=now() EventCode=4688 New_Process_Name=\"$file_path$\" Creator_Process_Name=\"$command$\" | eval process=\"$process$\", reason
      =\"$reason$\", nt_host=host, urgency=\"$urgency$\" | fields _time host user New_Process_Name Creator_Process_Name _raw process reason urgency nt_host"
   | dedup host Creator_Process_Name New_Process_Name user
   | rename New_Process_Name as file_path, Creator_Process_Name as command
```

splunk> .conf19

HISTOGRAM EXAMPLE [CATEGORICAL OUTLIER]



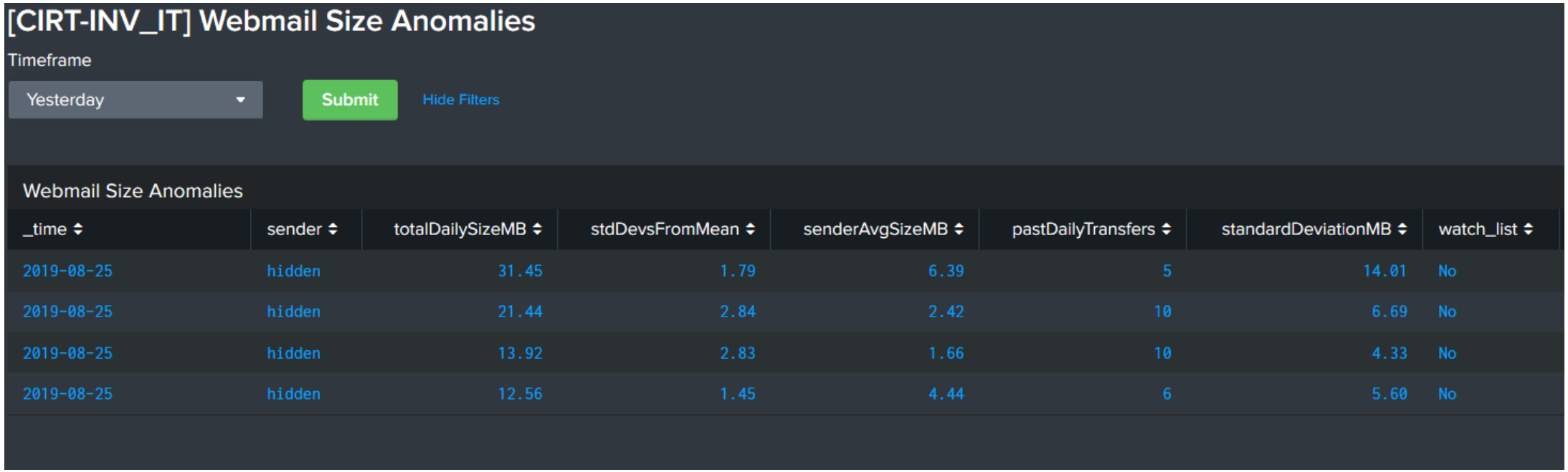
STREAM STATS & STD. DEVIATION [NUMERICAL OUTLIER]

INV-IT Personal Email Size Anomalies

```
1  index=     source=personal_email_size
   | eval end_time=strptime(date_year+"-"+date_month+"-"+date_mday,"%Y-%B-%d")
   | bin span=1d end_time
   | eval end_time_strp=end_time, end_time=strftime(end_time,"%Y-%m-%d"), _time=end_time_strp
   | sort 0 +_time
2  | streamstats avg(daily_size_bytes) as avg_size, stdev(daily_size_bytes) as stdev_size, max(daily_size_bytes) as max_size, count as n by senders
   | sort 0 -end_time_strp
   | dedup senders
   | eval max_size_mb = max_size/1024/1024, stdevs_away=(max_size-avg_size)/stdev_size
3  | where max_size_mb>10 AND daily_size_bytes=max_size AND stdevs_away>=1
   | search n!=1
   | where _time=relative_time(now(),"-1d@d")
   | sort -max_size_mb
   | table _time senders daily_size_bytes avg_size n stdevs_away stdev_size
   | rename senders as sender, avg_size as senderAvgSizeMB, n as pastDailyTransfers, stdevs_away as stdDevsFromMean, daily_size_bytes as totalDailySizeMB, stdev_size as standardDeviationMB
   | eval totalDailySizeMB=round(totalDailySizeMB/1024/1024,2), senderAvgSizeMB=round(senderAvgSizeMB/1024/1024,2), stdDevsFromMean=round(stdDevsFromMean,2), standardDeviationMB=round(standardDeviationMB
     /1024/1024,2), inputEarliest=_time, inputLatest=relative_time(_time,"+1d")
4  | lookup        rif-test.csv email as sender OUTPUT email as check, list as list
   | eval watch_list=if(isnotnull(check),"Yes","No"), list=if(isnotnull(list),list,"NA")
   | fields - check
   | table _time sender totalDailySizeMB stdDevsFromMean senderAvgSizeMB pastDailyTransfers standardDeviationMB watch_list list inputEarliest inputLatest
5  | outputlookup append=true personal_email_size_anomalies.csv
```

splunk> .conf19

STREAM STATS & STD. DEVIATION [NUMERICAL OUTLIER]



_time	sender	totalDailySizeMB	stdDevsFromMean	senderAvgSizeMB	pastDailyTransfers	standardDeviationMB	watch_list
2019-08-25	hidden	31.45	1.79	6.39	5	14.01	No
2019-08-25	hidden	21.44	2.84	2.42	10	6.69	No
2019-08-25	hidden	13.92	2.83	1.66	10	4.33	No
2019-08-25	hidden	12.56	1.45	4.44	6	5.60	No

CLOUD STORAGE DOWNLOAD VOLUME [RISK FRAMEWORK]

```
index=        vendor_action=DOWNLOAD
| lookup proxy-host-to-ip.csv s_ip as src OUTPUT host as src_host
| search src!="Unknown IP"
| rename created_by_login as PrimaryOwner, vendor_action as eventName, source_item_name as filePath, source_item_type as Type, source_parent_name
    as ParentFolder
| lookup    user_info_lookup mail as PrimaryOwner OUTPUT sAMAccountName as user
| stats sum(additional_details_size) as totalSize by user, eventName, src_host
| eval detectionName="    exfil", category="data_exfil", date=strftime(relative_time(now(),"-1d@d"),"%m/%d/%Y"), totalSize=round(totalSize/1024
    /1024/1024,2)
| search eventName=DOWNLOAD
| where isnotnull(totalSize) AND totalSize>0
| eval weight=if(src_host="Off-LMI" OR src_host="Unknown",10,weight), weight=if(src_host!="Off-LMI" AND src_host!="Unknown" AND totalSize>0,1
    ,weight), weight=if(src_host!="Off-LMI" AND src_host!="Unknown" AND totalSize>.5,3,weight), weight=if(src_host!="Off-LMI" AND src_host!
    ="Unknown" AND totalSize>1,5,weight), reason="Download Size: " + totalSize + "GB, Host: " + src_host
| table date detectionName weight category user reason
```

✓ 1,727 events (8/18/19 12:00:00.000 AM to 8/19/19 12:00:00.000 AM) No Event Sampling ▾ Job ▾

splunk> .conf19

CLOUD STORAGE DOWNLOAD VOLUME [RISK FRAMEWORK]

date	detectionName	weight	category	user	reason
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.04GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.02GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.01GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.01GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.06GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.01GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.09GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.04GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.01GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.05GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.01GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.31GB, Host: Off-LMI
08/15/2019	exfil	10	data_exfil	hidden	Download Size: 0.03GB, Host: Off-LMI
08/15/2019	exfil	5	data_exfil	hidden	Download Size: 5.40GB, Host: AWS
08/15/2019	exfil	5	data_exfil	hidden	Download Size: 4.46GB, Host: AWS
08/15/2019	exfil	5	data_exfil	hidden	Download Size: 1.34GB, Host: AWS
08/15/2019	exfil	5	data_exfil	hidden	Download Size: 1.77GB, Host: Proxy

splunk> .conf19

VPN USAGE ANOMALIES [RISK FRAMEWORK]

```
index=     source=vpn_condensed
| fields - _raw
| fields _time user
| bin span=1d _time
| eval user=lower(mvindex(split(user,"@"),0))
| stats count by user, _time
| streamstats count, max(_time) as recent, min(_time) as fc by user
| where _time>=relative_time(now(),"-1d@d")
| lookup    user_info_lookup sAMAccountName as user OUTPUT sAMAccountName
| where isnotnull(sAMAccountName)
| fields - sAMAccountName
| eval durationSec=recent-fc, duration=tostring(durationSec, "duration"), durationDays=mvindex(split(duration,"+"),0), usageRate=round(count/durationDays
    ,2), usageRate=if(count=1,"FC",usageRate)
| where usageRate<=.15 OR usageRate="FC"
| eval weight=1, weight=if(usageRate<.12,weight+1,weight), weight=if(usageRate<.1,weight+1,weight), weight=if(usageRate<.08,weight+1,weight),weight=if
    (usageRate<.06,weight+1,weight), weight=if(usageRate<.04,weight+1,weight), weight=if(usageRate<.02,weight+1,weight), weight=if(usageRate<=.01,weight
    +1,weight), weight=if(usageRate="FC",10,weight)
| eval detectionName="vpn_anomaly", category="behavior_anomalies", reason="VPN Frequency: " + usageRate, date=strftime(relative_time(now(),"-1d@d"),"%m
    /%d/%Y")
| eval user="hidden"
| table date detectionName weight category user reason
```

All time

✓ 94,005,408 events (before 8/16/19 5:12:34.000 PM) No Event Sampling Job Smart Mc

splunk> .conf19

VPN USAGE ANOMALIES [RISK FRAMEWORK]

date	detectionName	weight	category	user	reason
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	8	behavior_anomalies	hidden	VPN Frequency: 0.01
08/15/2019	vpn_anomaly	6	behavior_anomalies	hidden	VPN Frequency: 0.03
08/15/2019	vpn_anomaly	6	behavior_anomalies	hidden	VPN Frequency: 0.02
08/15/2019	vpn_anomaly	6	behavior_anomalies	hidden	VPN Frequency: 0.02
08/15/2019	vpn_anomaly	6	behavior_anomalies	hidden	VPN Frequency: 0.03
08/15/2019	vpn_anomaly	6	behavior_anomalies	hidden	VPN Frequency: 0.03
08/15/2019	vpn_anomaly	6	behavior_anomalies	hidden	VPN Frequency: 0.03
08/15/2019	vpn_anomaly	6	behavior_anomalies	hidden	VPN Frequency: 0.03

splunk>

Thank You!