

Splunk User Behavior
Analytics™

Lessons Learned from Deploying Splunk UBA

Teresa Chila

Cybersecurity Data Scientist | Chevron

Maria Sanchez

Technical Support Engineer | Splunk

This document is intended only for use by Chevron for presentation at .conf2019 and inclusion by Splunk on a conference website that is available to the public. No portion of this document may be copied, displayed, distributed, reproduced, published, sold, licensed, downloaded, or used to create a derivative work, unless the use has been specifically authorized by Chevron in writing.



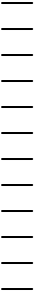
Teresa Chila

Cybersecurity Data Scientist | Chevron



Maria Sanchez

Technical Support Engineer | Splunk



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.

Splunk, Splunk>, Turn Data Into Doing, 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. © 2019 Splunk Inc. All rights reserved.

Basically, Teresa opened a lot of support tickets and Maria resolved them all. That's how this presentation came about



Teresa



Splunk
UBA



Maria



UBA Team

Agenda

1. Intro
2. About Splunk User Behavioral Analytics (UBA)
3. Why UBA at Chevron
4. Top 10 lessons learned at Chevron while deploying UBA
5. Q&A

Intro

Teresa

- Studied Electrical and Computer Engineering
- Data Scientist in the Cyber Intelligence Center in Chevron
- Over 20 years of experience in software development, security technologies, and data analytics
- Native from Hong Kong and enjoys travelling around the world



Chevron Corporation

140 years of human progress

One of the world's leading integrated energy companies

Upstream: exploration and production

Downstream & Chemicals: refine & distribute

Midstream: safe movement of products

Headquarters in San Ramon, CA

Substantial business activities in over 20+ countries with over 45K employees



Maria

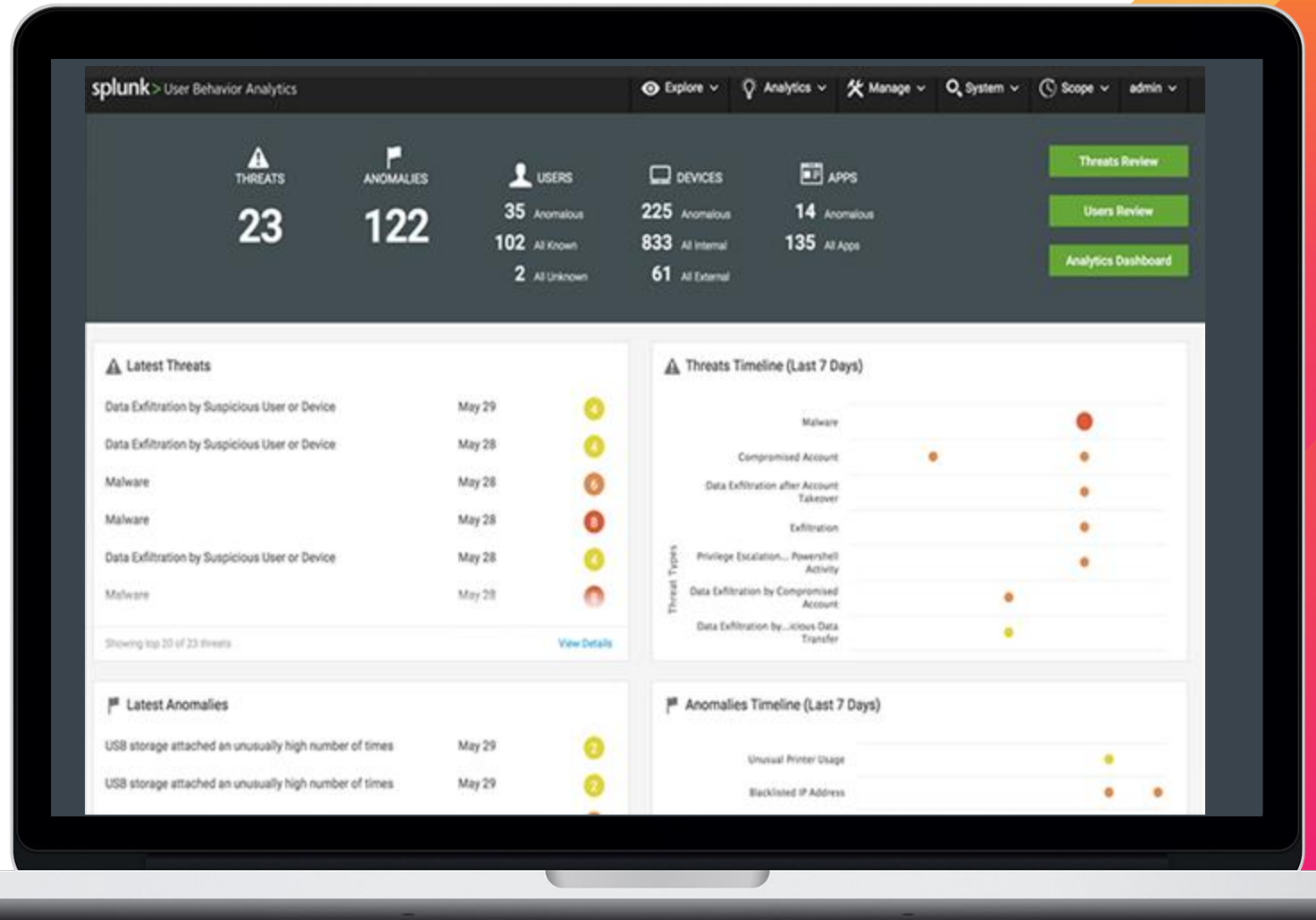
- Studied Systems Engineering
- Over 15 years of experience supporting enterprise software
- Splunker for 4 years, supporting UBA
- Based out of San Jose, CA
- From Colombia, loves spending time at the beach with her dogs, gardening and enjoying nature



About Splunk UBA

What is Splunk UBA?

Splunk UBA is an out-of-the-box solution that helps organizations find **known, unknown, and hidden threats** using **data science, machine learning, behavior baseline, and peer group analytics**



Splunk UBA Fundamentals



**Real-Time & Big
Data Architecture**



**Behavior Baseline
& Modelling**



**Unsupervised
Machine Learning**



**Anomaly
Detection**

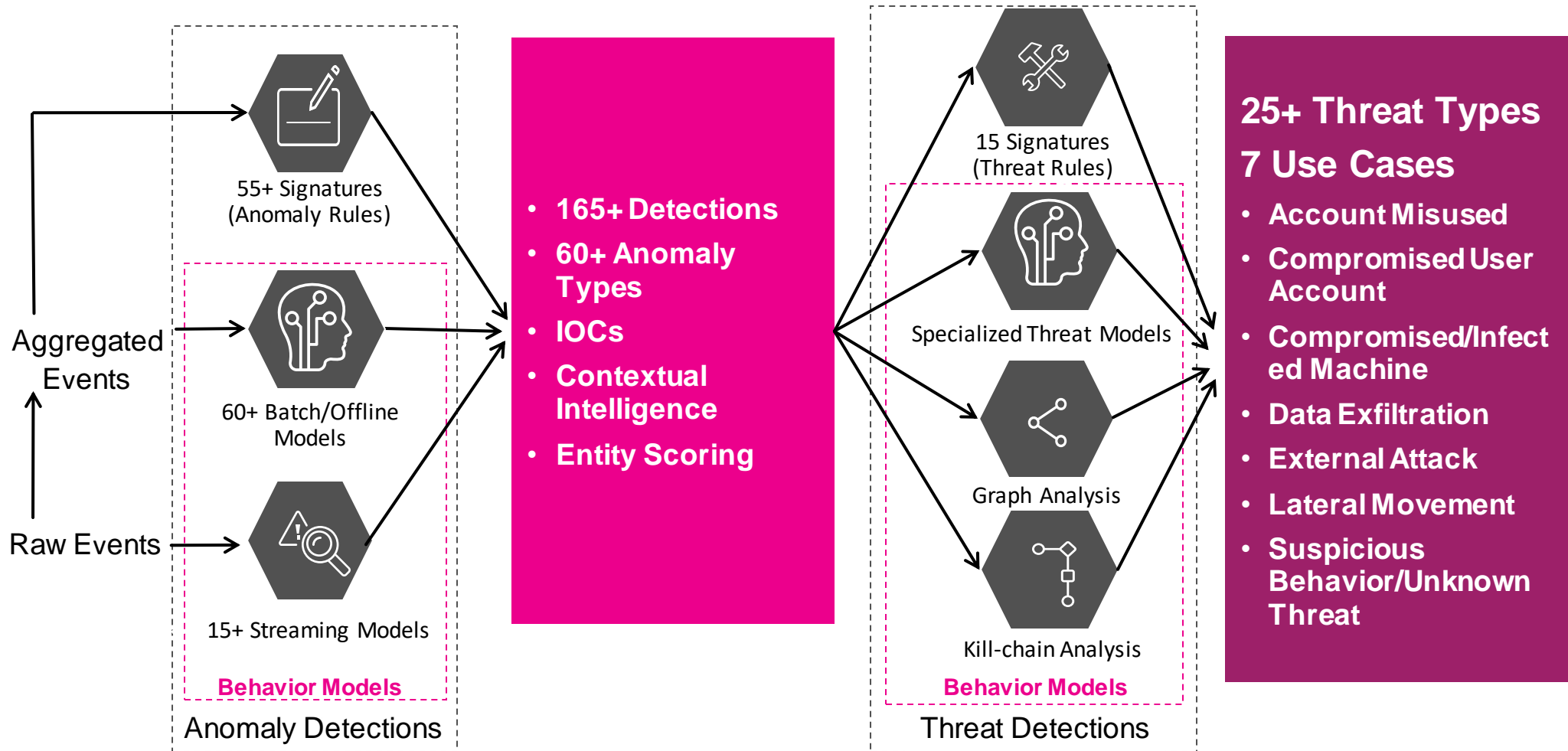


**Threat
Detection**



How Does Splunk UBA Work?

Multi-pass Machine Learning



Splunk UBA Workflow

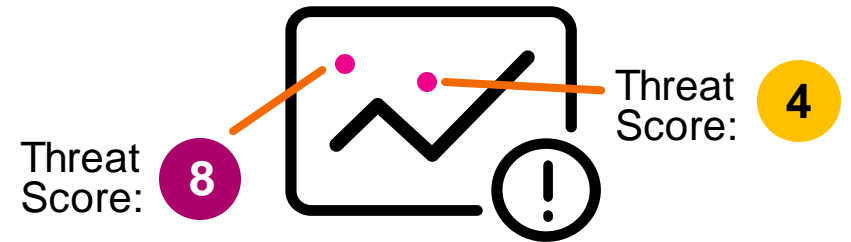
Powered by Big Data and Machine Learning



- Network Activity
- Application Activity
- Login Attempts
- Removable Media
- Badge Scans
- Printer Activity

(and more...)

- User's activity
- Departmental activity
- Region's activity
- Company's activity



Examples:

- Data Exfiltration by Suspicious User or Device
- Data Storage Attached by Unusual Number of Times
- Unusual Printer Usage
- Privilege Escalation
- Multiple Failed Login Attempts
- Malware
- Blacklisted IP Address
- Compromised Account

Splunk UBA Entities

Threats

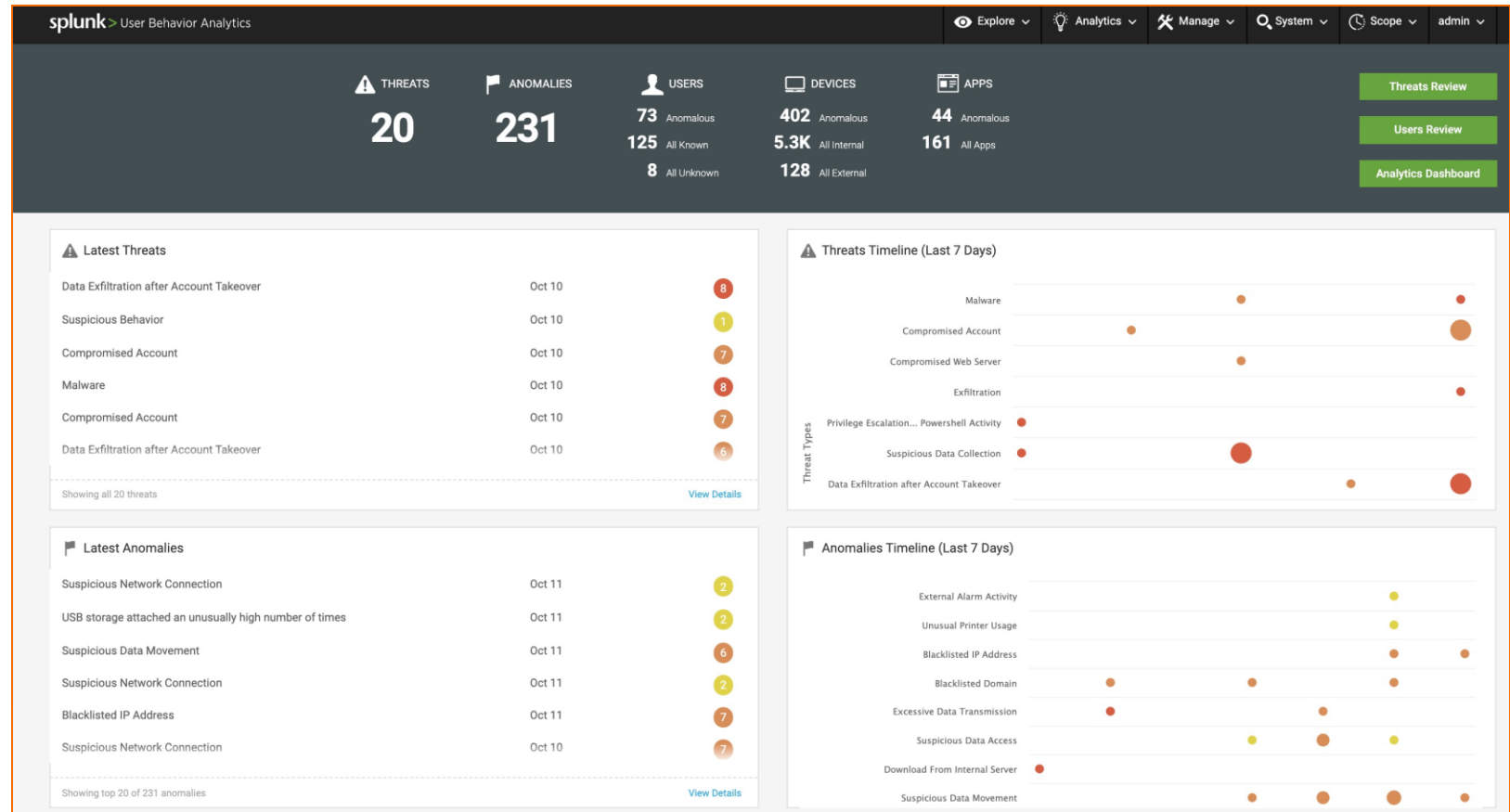
Anomalies

Users

Devices

Apps

- Each of these has a score
- Additional contextual information/feedback can be added



Why UBA at Chevron?

Why UBA at Chevron?

Augment our threat detection with advanced analytics

Prioritize by aggregating anomalies into threats

Accelerate our capability via buying instead of building

Integrate with our Core Splunk environment

How has UBA helped so far?



**Increase
network
visibility**



**Large collection
of detections out
of the box**



**Readily
available for
quickly creating
behavioral
models**



**Integrate with
Enterprise
Security to
provide a single
pane of glass**



**Provide
additional
context that
helps surface
cases**

Lessons Learned

1 Have Free Disk Space

Follow the hardware spec. Spec asked for 50GB for “/” drive

- We got 30GB

Not a problem at first, but after 3 or 4 weeks, rare things started to happen (models failing, not ingesting data, etc.)

- Disks > 95% full
- Clearing up disk space resolved the issue
- Models need disk space to offload memory

Now have a maintenance script to rotate/archive log files to help maintain a healthy amount of free disk space

Splunk UBA Hardware Requirements

CPU: 16 cores

Memory: 64 GB RAM

Storage: Three disks - 1200 IOPS

- Disk 1 - 50GB disk space for the Splunk UBA installation
- Disk 2 - 1TB additional disk space for metadata storage
- Disk 3 - 1TB additional disk space for each Spark node



Splunk UBA Deployment Options and Sizing Guide

On Prem

- VMware OVA
 - Ubuntu
- Bare Metal
 - CentOS
 - Oracle Enterprise Linux
 - RHEL

Cloud

- Amazon AWS AMI
 - Ubuntu
- Azure
 - CentOS
 - RHEL

| Size of cluster | Max Events per Second - EPS | Max # of Accounts | Max # of Devices | Max # of Data sources |
|-----------------|-----------------------------|-------------------|------------------|-----------------------|
| 1 node | 4K | 50K | 100K | 6 |
| 3 nodes | 12K | 50K | 200K | 10 |
| 5 nodes | 20K | 200K | 300K | 12 |
| 7 nodes | 28K | 350K | 500K | 24 |
| 10 nodes | 40K-45K | 350K | 500K | 32 |
| 20 nodes | 75K-80K | 750K | 1 Million | 64 |

2 Tips for Data Ingestion

Use a dedicated search head for UBA if you can

- UBA issues real-time searches to pull data from the search head, based on indexed time
- Good for late coming data or summary index with known delay

Version 4.3 supports micro-batch scheduled search

- Run at 1-minute interval
- Can backtrace
- Has health monitor app

Native parser works better for these data sources:

- Palo Alto Networks, Cisco ASA, and Windows Event Log
- Others can use Splunk Direct method using CIM compliant data sources

Splunk UBA Data Requirements



SPLUNK ENTERPRISE

Contextual Data

HRDATA (LDAP Accounts)

SPLUNK ASSETS

THREAT INTEL -
BLACK/WHITE LISTS

Event Data OPTIONAL

WINDOWS SECURITY
EVENTS

DNS, DHCP

FIREWALL

PROXY SERVER

VPN

END POINT

DLP

BADGE

CLOUD

EMAIL

EXTERNAL ALERTS

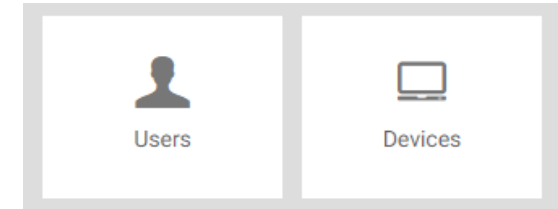
ES_NOTABLES

PRINTER

3 Pay Attention to Users and Devices

Spend time to review **users** and **devices** output early

- If they are not set up correctly, you have to reset the database
- This means losing existing profiles and anomalies, the system has to rebuild the behavioral baseline → practically starting over again



We had to reset twice:

- 1) after realizing the normal account and the admin account were not tied correctly, 2) after discovering incorrect devices
- Examples: Internal vs. external, service account vs. real device, web site name vs. real device, IP-to-device mapping
- There is a support tool to delete devices, but gets difficult when the devices are associated with anomalies

Set expectation and leave room in your deployment to allow for at least one database reset

4 Data Ingestion Order

Enable the data sources in sequence

1. Windows Event Log, DNS, and DHCP

- Enable them first since they are used for Identity Resolution (i.e. Device creation)
- Let them run for a few days. Good to review the devices at this time

2. Firewall and Proxy Data

- These generate a lot of anomalies
- For a large environment, enable one at a time. Spend time to review the new anomalies before enabling the 2nd data source

3. Remaining Data Sources

- E.g. email, AV, VPN, ES notables, etc.

Splunk UBA Data Requirements



SPLUNK ENTERPRISE

Contextual Data

HRDATA (LDAP Accounts)

SPLUNK ASSETS

THREAT INTEL -
BLACK/WHITE LISTS

Event Data OPTIONAL

WINDOWS SECURITY
EVENTS

DNS, DHCP

FIREWALL

PROXY SERVER

VPN

END POINT

DLP

BADGE

CLOUD

EMAIL

EXTERNAL ALERTS

ES_NOTABLES

PRINTER

5 Daily Health Monitoring

During initial deployment, allocate 5-10 minutes every day to check UBA

Check for sudden increase of anomalies

- Can be due to new data source or new/change in anomaly rules
- We had a sudden rise of >500K anomalies overnight, clogging the system
- For a 10-node cluster, do not exceed 1M anomalies

Check for Events Per Second (EPS) consistency

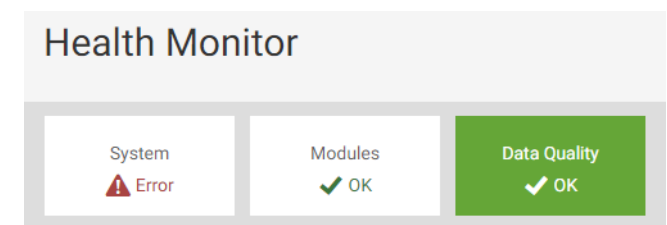
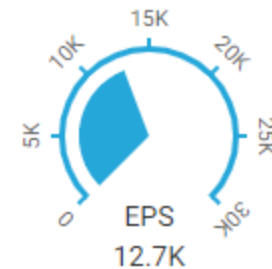
- May indicate search head or indexers issue

Check for new anomalies and threats

- No new anomaly and threat may indicate the models are not running

Use the Health Monitor or App to check each UBA service

Run Health Check script (cron daily)



How to Monitor Splunk UBA

Health Monitor UI Monitoring App Health Check Script

- System health
- Services
- Datasources and ingestion
- Containers
- Identity Resolution
- Events (Overall EPS)
- Models
- Rules

splunk > User Behavior Analytics

Home / Health Monitor

Health Monitor

System ✓ OK Modules ✓ OK **Data Quality ✓ OK**

Any Status ▾

Data Quality Indicators (23)

| NAME | INDICATOR | VALUE | STATUS |
|--------------------------------|--|--------------------------------|--------|
| Data Source | | | |
| | Data Source EPS on Splunk Events processed per second(avg) by each data source on Splunk in the last hour. | View 2 Values | ✓ OK |
| | Percentage of Events dropped by EventFilters Percentage of events dropped by EventFilters on UI | | ✓ OK |
| | Percentage of Events with no entity Percentage of events that have no entity | | ✓ OK |
| | Percentage of Events with no Relevant Data Percentage of events that had no relevant data | | ✓ OK |
| | Splunk Direct Data Source Enum Check Monitors the Splunk Direct input enum field data quality. This indicator tracks the mismatch rate (percentage) in each data source | View 1 Values | ✓ OK |
| Offline Rule Executor | | | |
| | Average Execution Time per Rule The average execution time of each rule | View 94 Values | ✓ OK |
| | Last Execution End Time per Rule The last time each rule completed | View 94 Values | ✓ OK |
| | Last Execution Failure per Rule The last time each rule failed to execute | View 94 Values | ✓ OK |
| | Last Execution Start Time per Rule The start time of the latest execution per rule | View 94 Values | ✓ OK |
| | Number of Execution Failures per Rule The number of times each rule has failed to execute consecutively, or 0 if no failures have occurred | View 94 Values | ✓ OK |
| | Number of Executions per Rule The total number of execution attempts, both successful and failed, for each rule | View 94 Values | ✓ OK |
| Output Connector Server | | | |
| | Number of Threats Sent to Output Connector The total number of threats sent to the output connector for forwarding to Splunk ES or other external destinations, since the last time UBA was restarted | 23 | ✓ OK |
| | Total New Anomalies Number of new anomalies received by the output connector server | 216 | ✓ OK |

6 Navigating UBA

Learn to use these features in the UI

- Filtering
 - With or without wildcard
 - Aware of display limit
 - Further filtering at the Anomaly Type level
- Grouping
- Adding columns
 - E.g. Model name

Watch the training video:

- <https://education.splunk.com/elearning/uba-hunter-walkthrough>

Filter Anomalies by Type and Device

1. Go to Anomaly Table
2. Select Anomaly Type
3. Group by Individual Devices

The screenshot displays the Splunk User Behavior Analytics dashboard. At the top, there are navigation menus for Explore, Analytics, Manage, System, Scope, and user profile (admin). Below the navigation, there are five summary cards: THREATS (7), ANOMALIES (215), USERS (30 Anomalous, 100 All Known, 1.8K All Unknown), DEVICES (508 Anomalous, 53K All Internal, 67 All External), and APPS (37 Anomalous, 160 All Apps). To the right of these cards are three buttons: Threats Review, Users Review, and Analytics Dashboard.

The main content area is divided into two sections. The left section, titled 'Latest Threats', shows a list of threats with their types and dates:

| Threat Type | Date | Count |
|--|--------|-------|
| Data Exfiltration by Suspicious User or Device | Mar 10 | 4 |
| Data Exfiltration by Suspicious Data Transfer | Mar 10 | 4 |
| Data Exfiltration by Compromised Account | Mar 10 | 5 |
| Malware | Mar 9 | 6 |
| Data Exfiltration by Suspicious User or Device | Mar 7 | 4 |
| Remote Account Takeover | Mar 4 | 4 |

Below the list, it says 'Showing all 7 threats' and a 'View Details' link.

The right section, titled 'Threats Timeline (Last 7 Days)', shows a green checkmark icon and the text 'No New Threats' with a sub-message: 'There are no new threats in the last 7 days'.

At the bottom, there are two more sections: 'Latest Anomalies' and 'Anomalies Timeline (Last 7 Days)', which are currently empty.

Filter Anomalies by Type and Event Metadata

1. Go to Anomaly Table
2. Select Anomaly Type
3. Click on the funnel icon
4. Filter by specific process

The screenshot displays the Splunk User Behavior Analytics dashboard. At the top, there are navigation tabs for Explore, Analytics, Manage, System, and Scope, along with a user profile for 'admin'. Below the navigation, there are five main sections: THREATS (7), ANOMALIES (215), USERS (30 Anomalous, 100 All Known, 1.8K All Unknown), DEVICES (508 Anomalous, 53K All Internal, 67 All External), and APPS (37 Anomalous, 160 All Apps). On the right side, there are three green buttons: Threats Review, Users Review, and Analytics Dashboard.

The main content area is divided into two columns. The left column is titled 'Latest Threats' and contains a table of threat events:

| Threat Type | Date | Count |
|--|--------|-------|
| Data Exfiltration by Suspicious User or Device | Mar 11 | 4 |
| Data Exfiltration by Suspicious User or Device | Mar 10 | 4 |
| Data Exfiltration by Suspicious Data Transfer | Mar 10 | 4 |
| Data Exfiltration by Compromised Account | Mar 10 | 5 |
| Malware | Mar 9 | 6 |
| Data Exfiltration by Suspicious User or Device | Mar 7 | 4 |

Below the table, it says 'Showing all 7 threats' and there is a 'View Details' link. The right column is titled 'Threats Timeline (Last 7 Days)' and shows a green checkmark icon with the text 'No New Threats' and 'There are no new threats in the last 7 days'.

Group Anomalies by Devices and filter using Wildcard

1. Go to Anomaly Table
2. Group by Individual Devices
3. Add Filter for Specific Devices
4. Search using wildcards
5. Click on + Any device matching *XYZ*

The screenshot displays the Splunk User Behavior Analytics dashboard. At the top, there are navigation tabs for Explore, Analytics, Manage, System, and Scope. Below these are summary cards for THREATS (7), ANOMALIES (215), USERS (30 Anomalous, 100 All Known, 1.8K All Unknown), DEVICES (508 Anomalous, 53K All Internal, 67 All External), and APPS (37 Anomalous, 160 All Apps). On the right, there are buttons for Threats Review, Users Review, and Analytics Dashboard.

The main content area is divided into four sections:

- Latest Threats:** A table listing recent threats with their dates and severity scores.

| Threat Name | Date | Score |
|--|--------|-------|
| Data Exfiltration by Suspicious User or Device | Mar 11 | 4 |
| Data Exfiltration by Suspicious User or Device | Mar 10 | 4 |
| Data Exfiltration by Suspicious Data Transfer | Mar 10 | 4 |
| Data Exfiltration by Compromised Account | Mar 10 | 5 |
| Malware | Mar 9 | 6 |
| Data Exfiltration by Suspicious User or Device | Mar 7 | 4 |
- Threats Timeline (Last 7 Days):** A summary card showing "No New Threats" with a green checkmark icon and the text "There are no new threats in the last 7 days".
- Latest Anomalies:** A table listing recent anomalies with their dates and severity scores.

| Anomaly Name | Date | Score |
|--------------------------------|--------|-------|
| Multiple Logins | May 18 | 6 |
| Multiple Authentication Errors | May 18 | 6 |
| Multiple Login Errors | May 18 | 5 |
| Multiple Logins | May 6 | 4 |
| Blacklisted IP Address | Mar 12 | 7 |
| Suspicious Network Connection | Mar 11 | 7 |
- Anomalies Timeline (Last 7 Days):** A summary card showing "No New Anomalies" with a green checkmark icon and the text "There are no new anomalies in the last 7 days".

Group Anomalies by Category and display Model

1. Go to Anomaly Table
2. Group by Anomaly Category
3. Add column to display Model

The screenshot displays the Splunk User Behavior Analytics dashboard. At the top, there are navigation tabs for Explore, Analytics, Manage, System, and Scope, along with a user profile for 'admin'. Below the navigation, there are summary cards for THREATS (7), ANOMALIES (215), USERS (30 Anomalous, 100 All Known, 1.8K All Unknown), DEVICES (508 Anomalous, 53K All Internal, 67 All External), and APPS (37 Anomalous, 160 All Apps). On the right side, there are buttons for Threats Review, Users Review, and Analytics Dashboard.

The main content area is divided into four sections:

- Latest Threats:** A table listing recent threats with their categories and dates.

| Category | Date | Score |
|--|--------|-------|
| Data Exfiltration by Suspicious User or Device | Mar 11 | 4 |
| Data Exfiltration by Suspicious User or Device | Mar 10 | 4 |
| Data Exfiltration by Suspicious Data Transfer | Mar 10 | 4 |
| Data Exfiltration by Compromised Account | Mar 10 | 5 |
| Malware | Mar 9 | 6 |
| Data Exfiltration by Suspicious User or Device | Mar 7 | 4 |
- Threats Timeline (Last 7 Days):** A summary card showing 'No New Threats' with a green checkmark icon and the text 'There are no new threats in the last 7 days'.
- Latest Anomalies:** A table listing recent anomalies with their categories and dates.

| Category | Date | Score |
|--------------------------------|--------|-------|
| Multiple Logins | May 18 | 6 |
| Multiple Authentication Errors | May 18 | 6 |
| Multiple Login Errors | May 18 | 5 |
| Multiple Logins | May 6 | 4 |
| Blacklisted IP Address | Mar 12 | 7 |
| Suspicious Network Connection | Mar 11 | 7 |
- Anomalies Timeline (Last 7 Days):** A summary card showing 'No New Anomalies' with a green checkmark icon and the text 'There are no new anomalies in the last 7 days'.

7 When to Start Tuning

Professional Services



- Start looking at anomalies and threat in tuning perspective. Focus on breadth.
- Behavioral-based anomalies may require 20-30 days for meaningful results

Start monitoring threats for detection and investigation. Focus on depth

8 How to Start Tuning

Look at the model name

Models generate anomalies

One anomaly type can comprise of multiple models

Much more descriptive, easier to communicate with analysts

Suspicious
Data
Movement

- Data transfer over email
- Http transfer to storage
- Transfer to USB
- Unusual activity amount
- Etc.

Anomaly Type

Model Name

Look at threat first or anomaly first?

1. Start with anomalies. Find the biggest offenders:
 - Filter to display only one Anomaly Type or Model
 - Then group by “Individual Device”
 - The top device with the highest count is likely a tuning candidate, or something bad
2. Then look at top Threats. Look at the anomalies in the threat, and find the biggest offenders in those anomalies. Repeat

9 Tuning Options

How to adjust UBA anomalies to your needs

**Change
Source
Query**

**Exclude unwanted
events in the data
ingestion SPL**

**Anomaly
Action
Rule**

**Used to suppress
anomalies or
change score**

**Override
Model**

**Override model
parameters in
configuration file**

**Clone
Model**

**Clone
configuration to
make new models**

SDK

**Create your
custom model**

10 Avoid Viewing Fatigue

By default, all models are turned on

Some are more useful to you, some are less

- Anomalies for Insider Threat vs. External Threat
- Some anomalies don't have enough details about the contributing events, hard for analysts to further investigate

Compile a list of models that are more impactful to you, and leave the rest as boosting factor or contextual events

- Allow you to be more focused, and help create better threats later
- You can either suppress the anomaly, or lower or raise the score for importance

Goal is to start alerting on meaningful threats built from anomalies

Key Takeaways

1. Understand UBA sizing constraints and deploy for growth
2. Become familiar with UBA services and troubleshooting
3. Learn how to navigate UBA
4. Perform iterative tuning
5. Identify new use cases

Additional Resources

1. Visit Splunk UBA booth and watch the demos
2. Check out other UBA talks at .conf
 - Innovation Labs: UBA Custom Machine Learning Use-Case Framework
 - SEC2109 - Hunting Threats with Splunk UBA
 - SEC1230 - It is Normal or Suspicious? Detecting Anomalies via market Basket Analysis
 - SEC1732 - Let's Get Hands-On with Splunk Enterprise, Splunk Phantom, Splunk UBA & Real Boss of the SOC Data
 - SEC2083 - Catch exfiltration from cloud file stores early!
 - SEC1248 - Advanced Threat Hunting & Anomaly Detection with Splunk UBA
3. Watch Splunk Education Videos
 - UBA Analyst: <https://education.splunk.com/elearning/uba-soc-analyst-walkthrough>
 - UBA Hunter: <https://education.splunk.com/elearning/uba-hunter-walkthrough>
 - UBA Admin: <https://education.splunk.com/elearning/uba-administration-walkthrough>



Q&A

msanchez@splunk.com



splunk>

Thank

You!

Go to the .conf19 mobile app to

RATE THIS SESSION

