Security in knowledge

Big Data and Security: At the Edge of Prediction

Mark Seward

Splunk Inc.

Fred Wilmot

Splunk Inc.

RSA CONFERENCE 2013

Session ID: SPO2-T17

Session Classification: Intermediate

The Way Cyber Adversaries Think

Where is the most important and valuable data?

What's the typical patch cycle for applications and operating systems?

What are the typical security defenses?

How does the IT team prioritize vulnerabilities?

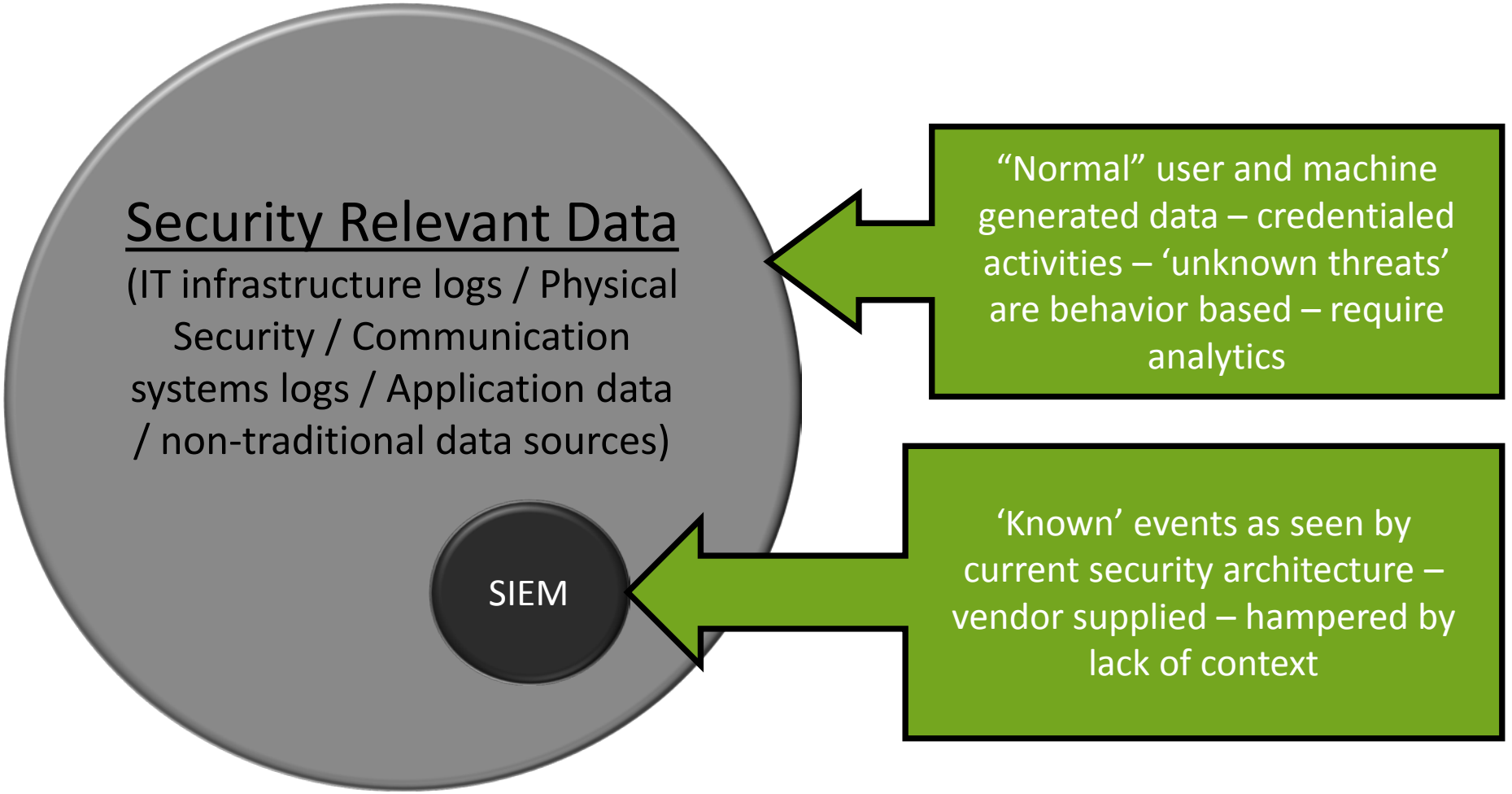
What structural information silos that exist for the security team?

Who in the organization has access to the most valuable data and credentials I can steal?

Are 'normal' IT service user activities routinely monitored and correlated?

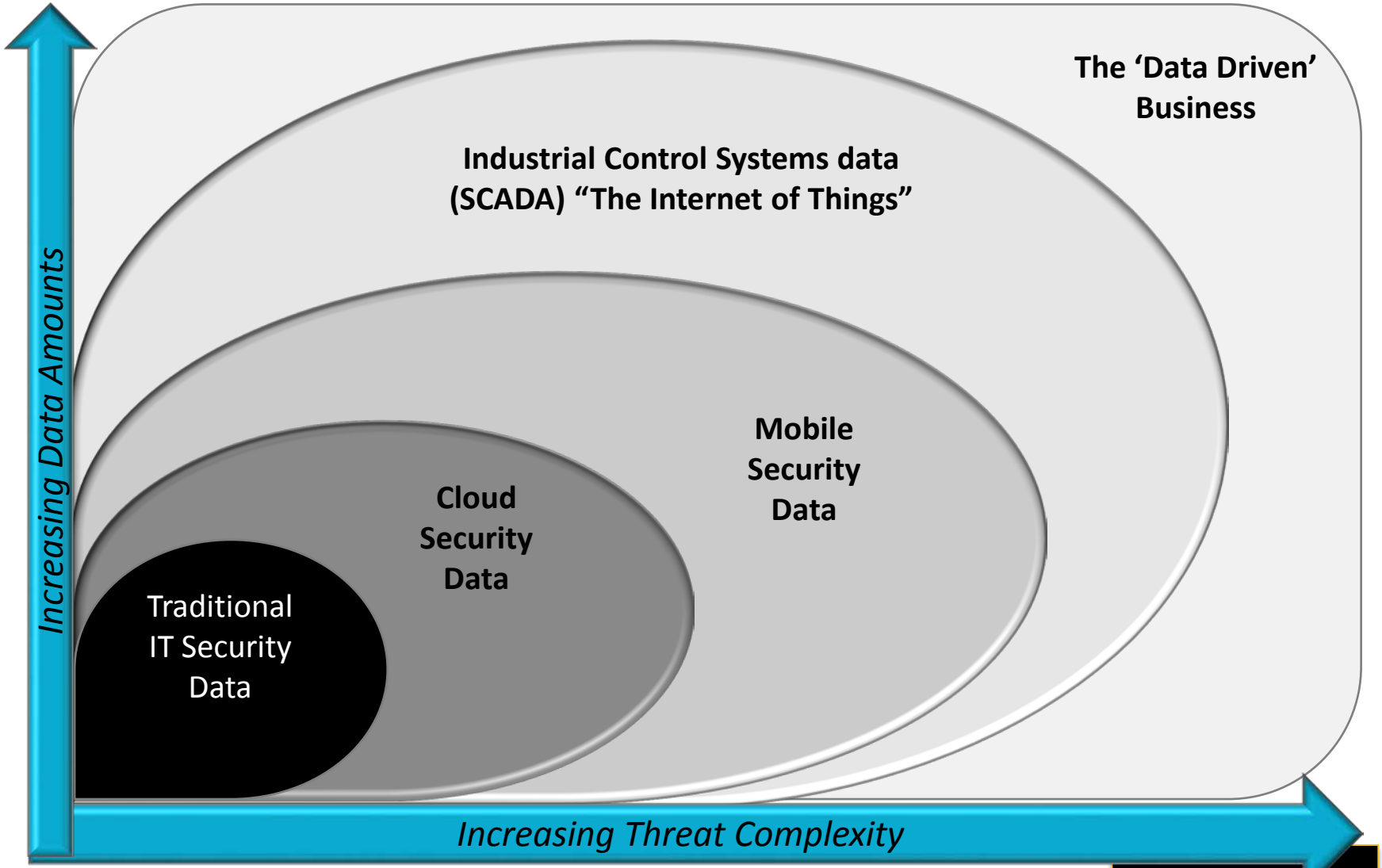
splunk>

Security has out grown the traditional SIEM



splunk>

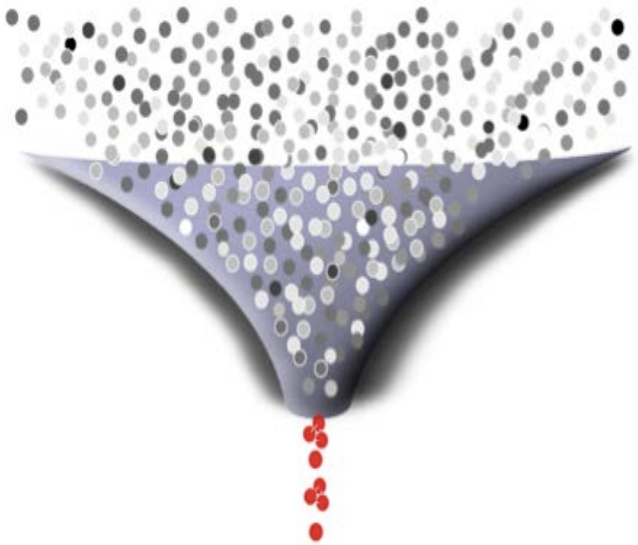
Required - A Broader Look at Security Data



New Architecture Required for Security

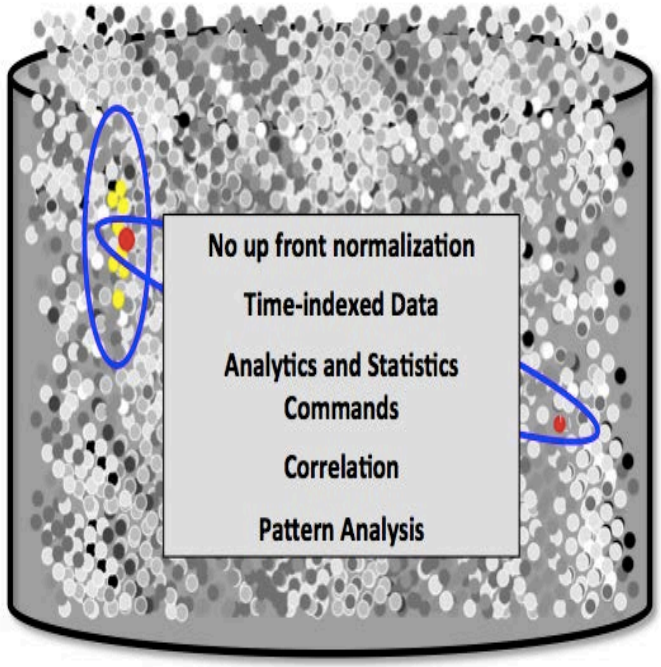
Traditional SIEM

Indexed data store



Data Reduction Model

vs.



Data Inclusion Model

No contest / No Limits

splunk>

Vendor Seduction: The Way Some Security Folks Think

"I hope my AV, IPS, Firewall, (name your technology) vendor catches these guys."

"I have 300 rules on my SIEM. One of them will catch the attacker."

Attackers know that if you have a static correlation engine, you are likely trusting it, and because often "No news is good news"

splunk>

Business Critical Infrastructure Protection

Machine Data

Security Intelligence for Business

Anomaly / Unknown Threat Detection

Proactive Monitoring for Known Threats

Compliance

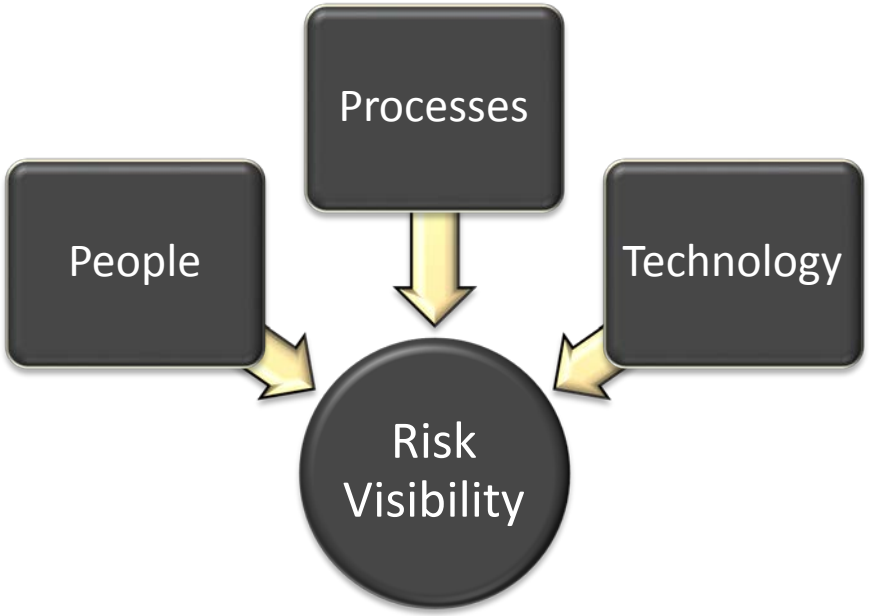
Failure Analysis and Forensics

splunk>

Why Take a Platform Approach?

- Security risks often reviewed in isolation from each other and operational risks
- Encourages risk mitigation at higher levels within an organization (people and process)

Business operations and security data contains information about…



splunk>

Enabling IT Risk Scenarios

Applying IT Risk Scenarios 'Finding Abnormal Behaviors'

Security Relevant Data
Confidentiality / Integrity / Availability

CSO / CIO / CEO Views

App Mgmt

IT Ops

Third Party Services Data

Web Analytics

Business Analytics

Big data indexing solution

splunk>

What we know – so far

- More data is needed to detect stealthy attacks
- To truly address security business risk we need to combine IT data with less traditional data sources
- The traditional SIEM is not built to handle the volume, velocity and variety of data needed for business security risks
- The bad guys can guess what our infrastructure looks like, how IT Operations and Security function, and deduce who has access to the data they want
- Attackers use creativity – we use vendor solutions

Based on our scenarios and what we know about attacks and attackers, what can we predict?

Security in knowledge

What we hear about Predictive Analytics…

"Vendor X says they'll be able to predict what's going to happen in my IT architecture before it happens."

"The vendor says it will stop bad things before they happen based on super secret al-go-rhythms."

splunk>

What is Predictive Analytics?

Predict: to declare or indicate in advance; especially: foretell on the basis of observation, experience, or scientific reason

+

Analytics: the method of logical analysis

Predictive Analytics: To foretell and/or declare in advance based on observation, experience, or reason using a method of logical analysis

What is (and is not) Predictive Analytics

- Predictive Analytics IS only as good as your experience, observations and method(s) of logical analysis
- Predictive Analytics IS NOT going to help you catch bad guys that work outside your experience, observations and methods of logical analysis
- Predictive Analytics IS NOT having a machine or anyone else 'think' for you
- Predictive Analytics IS using your, or your company's collective observations and experience plus one or more methods of logical analysis to identify patterns in data that can be used to make predictions about future outcomes.
- Predictive Analytics IS a starting point – not an ending point for investigation

Where do we get knowledge about how attackers may behave?

Second hand information

- Our peers / contacts / events
- Trade journals
- Security threat reports

First hand information

- Previous data breach investigation
- 'Inside knowledge' of security procedure and process weaknesses
- 'Inside knowledge" of technology weaknesses
- 'Inside knowledge' of where our most valuable data is
- 'Inside knowledge' of business processes
- Evidence of user behavior(s)

The need to shift our focus

splunk>

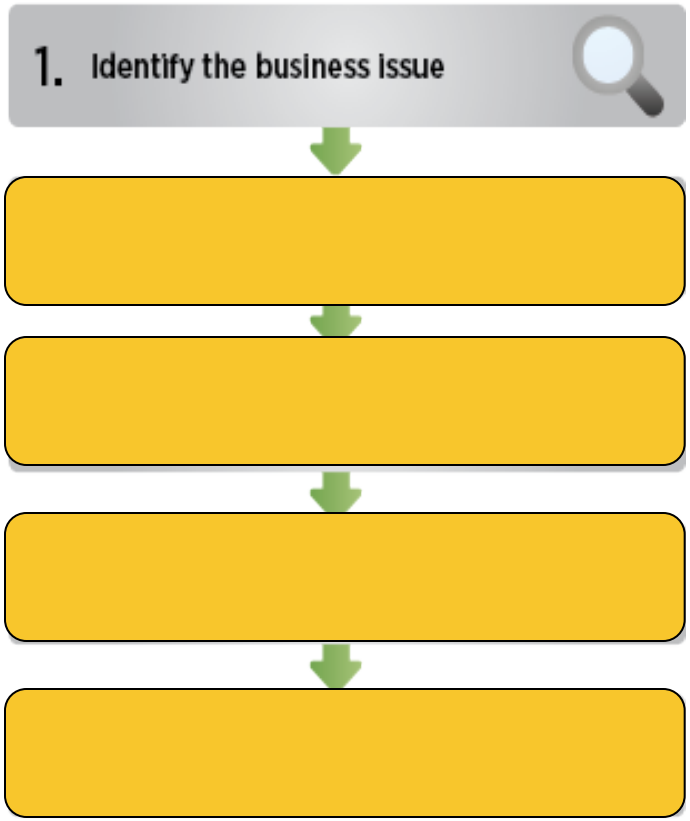
How Can Big Data and Statistical Analysis Help Anticipate Attacks

DEMO 20 min

Security in knowledge

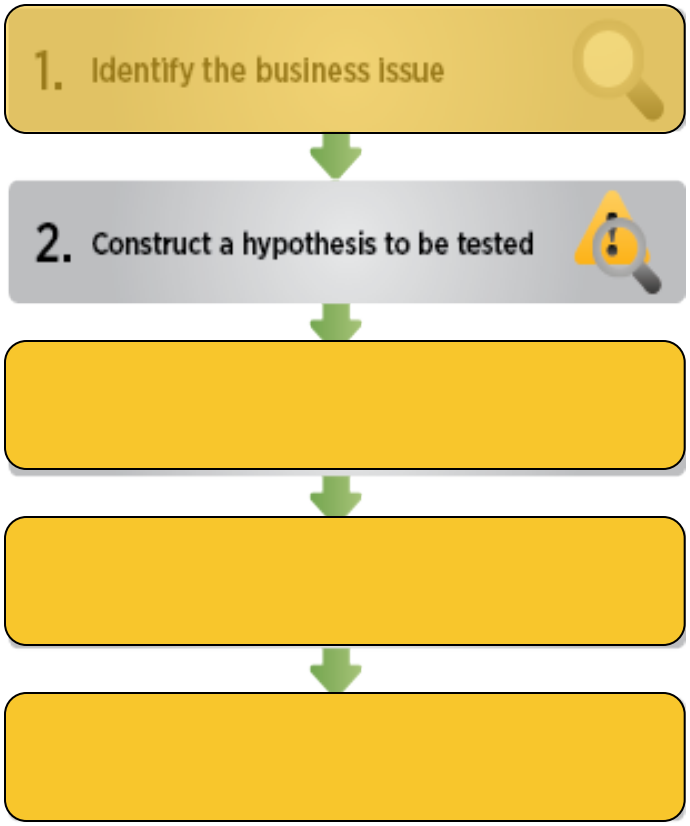
"A Big-data security process"

A Process for Using Big Data for Security: Identify the Business Issue



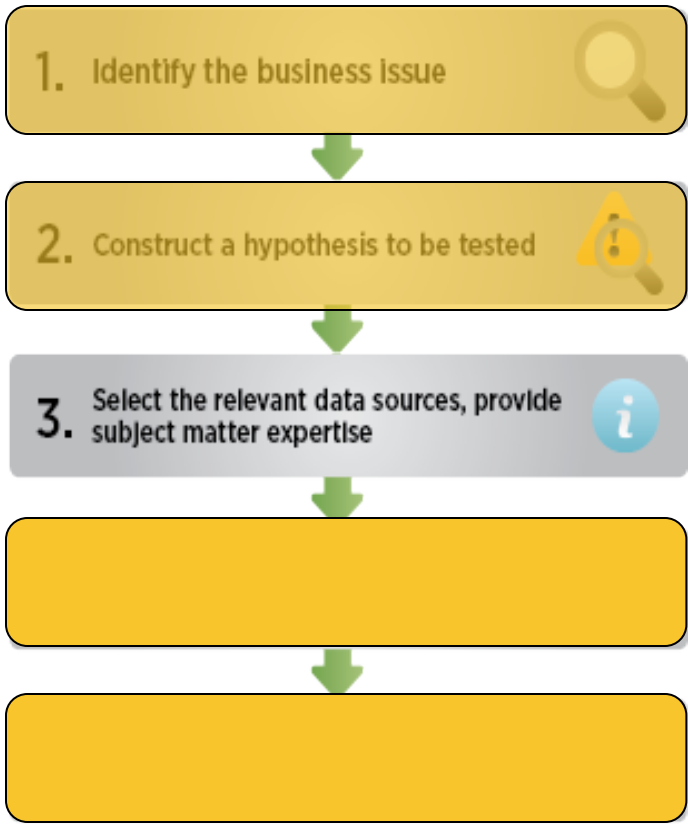
- What does the business care about?
- What could cause loss of service or financial harm?
- Performance Degradation
- Unplanned outages (security related)
- Intellectual property access
- Data theft

A Process for Using Big Data for Security: Construct a Hypothesis



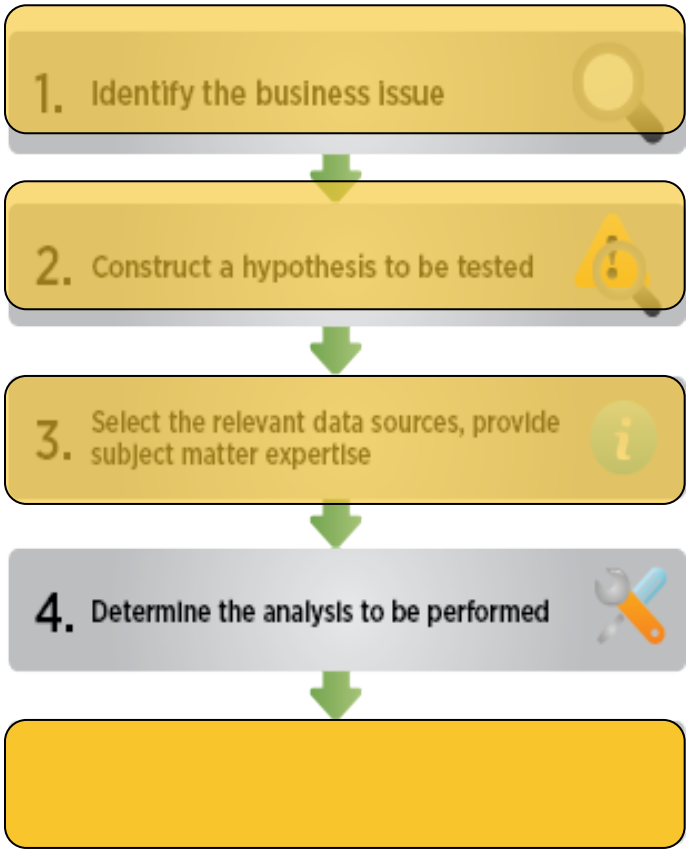
- How could someone gain access to data that should be kept private?
- What could cause a mass system outage does the business care about?
- What could cause performance degradation resulting in an increase in customers dissatisfaction?

A Process for Using Big Data for Security: It's about the Data



- Where might our problem be in evidence?
- For data theft start with unauthorized access access issues…
- Facility access data, VPN, AD, Wireless, Applications, others…
- Beg, Borrow, SME from system owners

A Process for Using Big Data for Security: Data Analysis



- For data theft start with what's normal and what's not (create a statistical model)
- How do we 'normally' behave?
- What patterns would we see to identify outliers?
- Patterns based on ToD, Length of time, who, organizational role, IP geo-lookups, the order in which things happen, how often a thing normally happens, etc.

splunk>

A Process for Using Big Data for Security: Interpret and Identify

- What are the mitigating factors?
- Does the end of the quarter cause increased access to financial data?
- Does our statistical model need to change due to network architecture changes, employee growth, etc?
- Can we gather vacation information to know when it is appropriate for HPA users to access data from foreign soil.
- What are the changes in attack patterns?

splunk>

Short form - Example

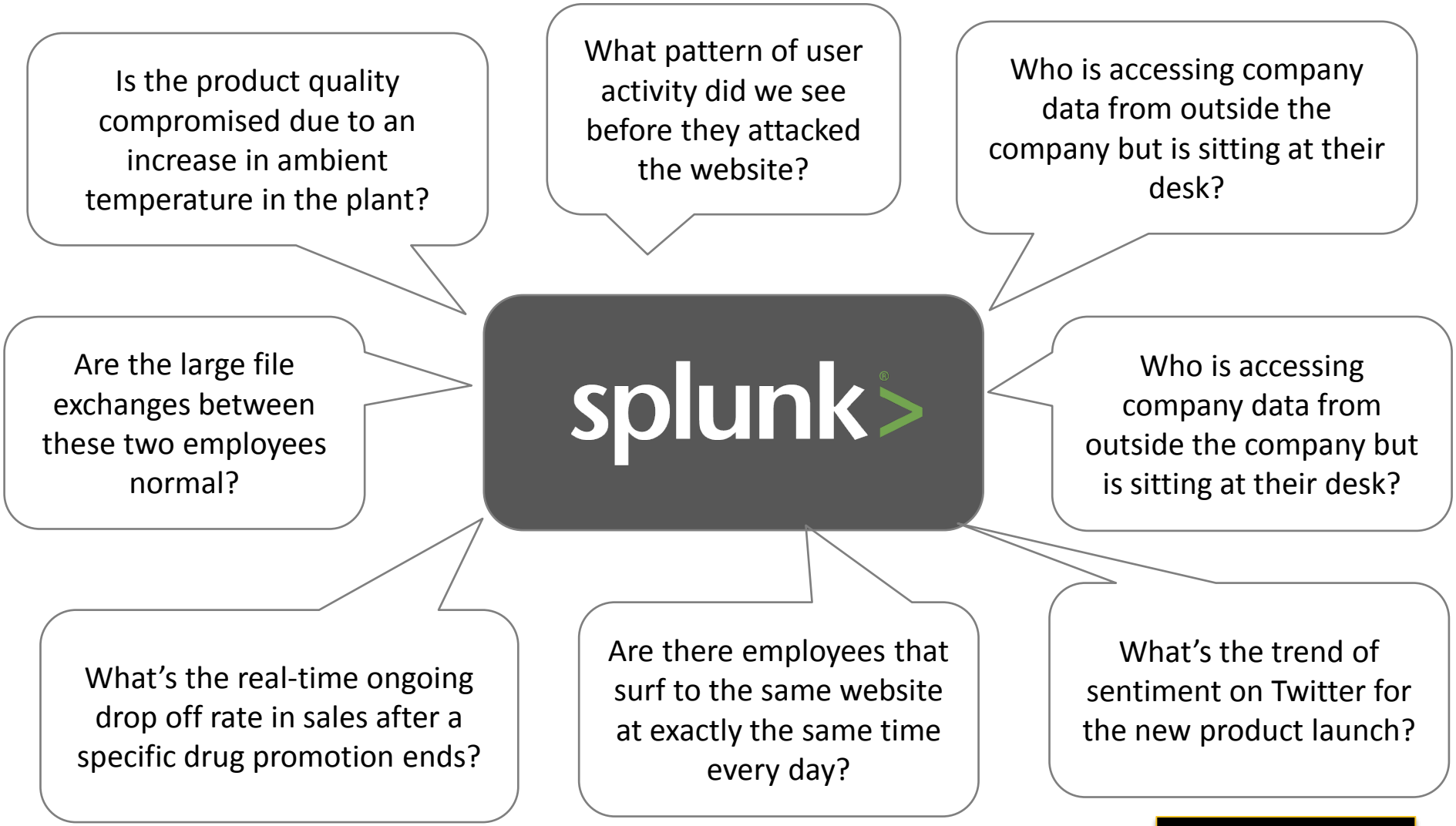
The Steps	The Response
Business Issue	Service degradation causes monetary damage and customer satisfaction issues.
Construct one of more hypothesis (team creativity required)	Unwanted bots can degrade service and steal content.
Gather data sources and expertise	What combinations of data would be considered definitive evidence? What might be the first signs of trouble? List all data in which this might be reflected.
Determine the analysis to be performed	Determine the types of data searches appropriate and automation requirements
Interpret the results	Do the results represent false positives of false positives or false negatives? Are there good bots and bad bots?

Looking Beyond IT for Business Risk

Manufacturing

Parts/Ingredients (RFID) Data	Raw Materials Data	Shipping Data (when/who loaded the truck)
Facility Security Data		Personnel Data
Industrial Control System Data		HVAC data
Distribution Monitoring Data (GPS)	Point of Sale Data	Traditional IT Data

What manufacturing questions could you ask?



splunk>

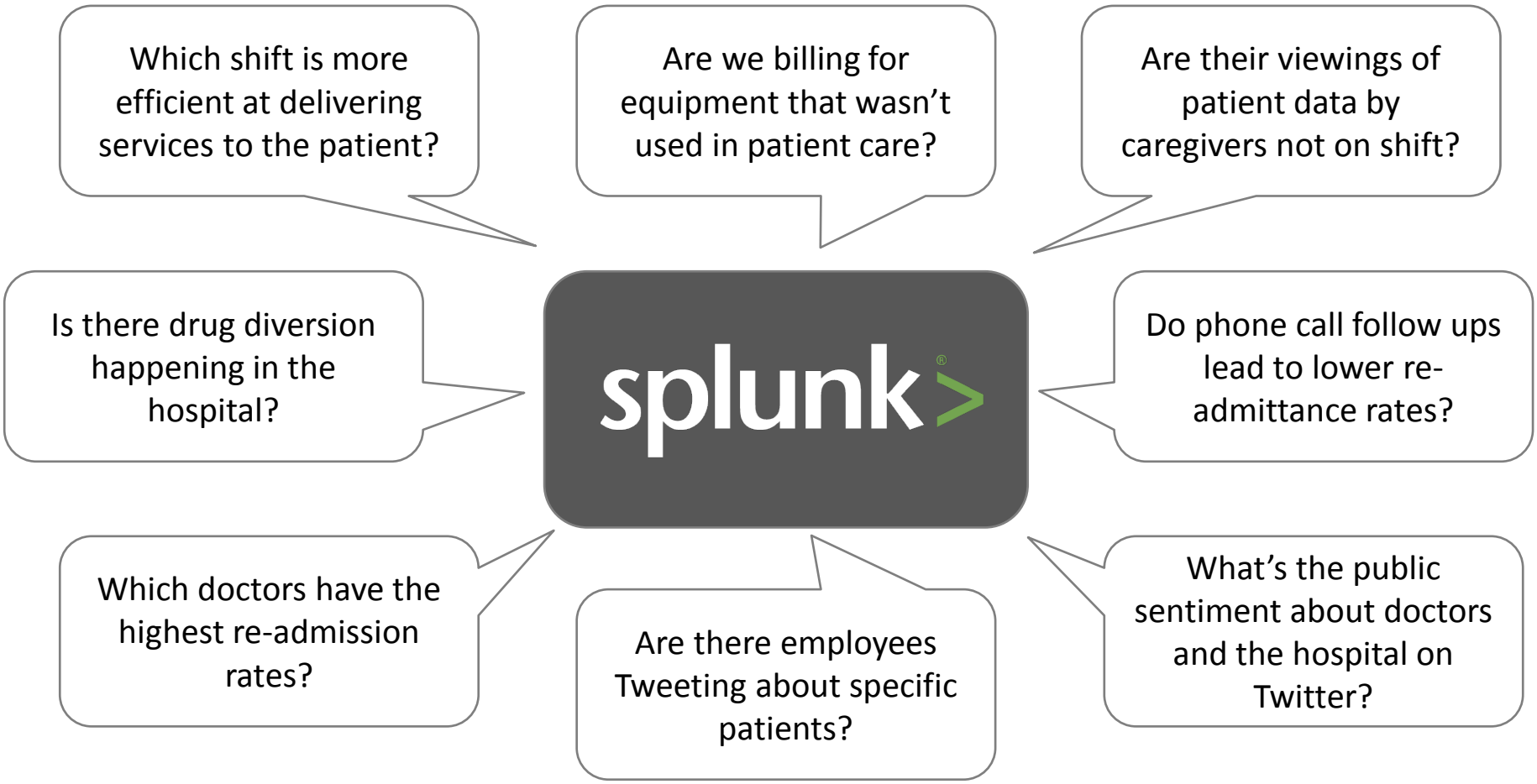
Looking Beyond IT for Business Risk

Healthcare

- Equipment (RFID)
- Patient Data
- Call Data Records
- Facility Security Data
- Personnel Data
- Business associate access and transmission data
- HVAC data
- Traditional IT Data
- Equipment (RFID)
- Social Media Data

splunk>

What Healthcare Questions Could You Ask of Data?



splunk>

Why Big Data and Analytics are the Future of Security

1. **Confidentiality Integrity and Availability** is a holistic view of business security and risk mitigation growing beyond traditional IT data sources

2. **Security is being redefined**: Monitoring & mitigating threats that compromise business reputation, service delivery, confidential data or result in loss of intellectual property

3. **Security folks will want / need more data** – not less – for accurate root cause analysis

4. **Complexity of threats will continue to grow** and cross from IT to less traditional data / devices / sources

5. A single investigation will **include data from all parts of the business** – beyond IT data

6. **Using statistical analysis** for Base-lining and understanding outliers is the way to detect advanced threats

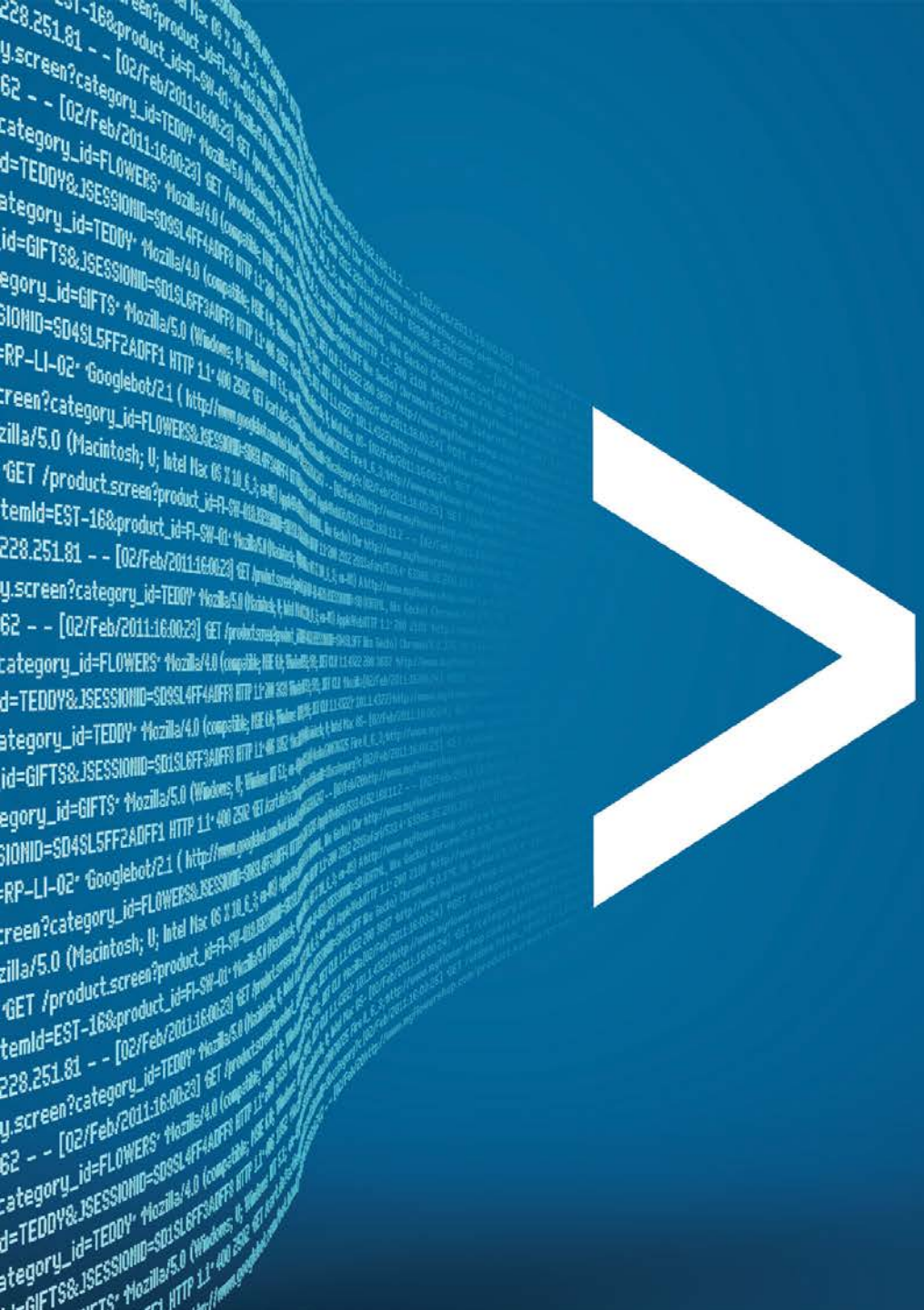
splunk>

NIST Predicts: Sensing big data trend

"The number, volume, and variety of computer security logs have increased greatly, which has created the need for computer security log management—the process for generating, transmitting, storing, analyzing, and disposing of computer security log data."

NIST 800-92 Guide to Computer Security Log Management 1996

splunk>



Thank You

Questions

fwilmot@splunk.com
mseward@splunk.com

www.splunk.com

splunk>