BIG DATA FOR SECURITY:

HOW CAN I PUT BIG DATA TO WORK FOR ME?

Joe Goldberg
Splunk

Security in knowledge

RSA CONFERENCE EUROPE 2013

Session ID: HT-T08

Session Classification: Intermediate

About Me

Joe Goldberg

Current:

- Splunk - Security Evangelist and Technical Product Marketing

Past:

- Symantec/Vontu - Data Loss Prevention, Technical Product Marketing
- VMware - Technical Product Marketing
- Sun Microsystems - Product Marketing

Agenda

- Big Data defined
- Security problem
- Security solution
  - Technology
  - Use cases
- Actual deployments
- Cautions

Everyone is Claiming Big Data

splunk> see the forest, and the trees

Products Solutions Industries Partners About Us Support Services Resour

Big Data Analytics

Turn Machine-generated Big Data into Real-time Insights

Your IT systems and technology infrastructure generate data every second of every day. This machine data contains a categorical record of all user behaviors, service levels, cybersecurity risks, fraudulent activities and more. As one of the fastest growing and most complex segments of big data, machine data is also one of the most valuable.

IBM Security Intelligence with Big Data

IBM Security Intelligence with Big Data

With major security breaches and fraud incidents making international headlines, organizations are taking steps to address the growing problems of advanced persistent threats, fraud, and insider attacks.

« SECURITY

RSA SECURITY ANALYTICS

GATE ADVANCED THREATS

es Big Data »

BLUE COAT SOLERA NETWORKS A BLUE COAT COMPANY

About Products & Services Resources Partners Company Support Contact Us Big Data Security Blogs

Request a Custom Demo

Brief | HP HAVEn

HAVEn—Big Data platform

HAVEn

Hadoop/HDFS	Autonomy IDOL	Vertica	Enterprise Security	nApps
Catalog massive volumes of distributed data	Process and index all information	Analyze at extreme scale in real-time	Collect and unify machine data	Powering HP Software and *your* apps

“Big Data” Definition

- Wikipedia: Collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications
- Gartner: The Three Vs
  - Data volume
  - Data variety
  - Data velocity
- Security has always been a Big Data problem; now it has a solution

The Security Problem

Advanced Threats Are Hard to Detect

100%

Valid credentials were used

243

Median # of days before detection

40

Average # of systems accessed

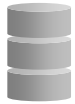
63%

Of victims were notified by external entity

- And threats are hard to investigate

Source: Mandiant M-Trends Report 2012 and 2013

Need all the Data


Databases

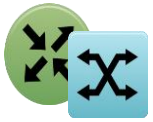

Email


Web


Desktops

Servers


DHCP/ DNS


Network Flows


Hypervisor


Badges

Traditional Sources


Firewall


Authentication


Vulnerability Scans


Intrusion Detection


Data Loss Prevention

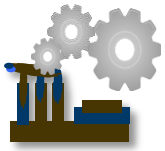

Anti-Malware

Custom Apps


Service Desk


Storage


Mobile


Industrial Control


Call Records

Machine Data / Logs

Sources

Email Server

2013-08-09T12:40:25.475Z,,exch-hub-den-01,,exch-mbx-cup-00,,,STOREDRIVER,DELIVER,79426,<20130809050115.18154.11234@acme.com>,johndoe@acme.com,,685191,1,,, hacker@neverseenbefore.com , Please open this attachment with payroll information,, ,2013-08-09T22:40:24.975Z

Web Proxy

2013-08-09 16:21:38 10.11.36.29 98483 148 TCP_HIT 200 200 0 622 - - OBSERVED GET www.neverbeenseenbefore.com HTTP/1.1 0 "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 2.0.50727; InfoPath.1; MS-RTC LM 8; .NET CLR 1.1.4322; .NET CLR 3.0.4506.2152;) User John Doe,"

Endpoint Logs

20130806041221.000000Caption=ACME-2975EB\Administrator Description=Built-in account for administering the computer/domainDomain=ACME-2975EB InstallDate=NULLLocalAccount = IP: 10.11.36.20 TrueName=Administrator SID =S-1-5-21-1715567821-926492609-725345543 500SIDType=1 Status=Degradedwmi_ type=UserAccounts

Windows Authentication

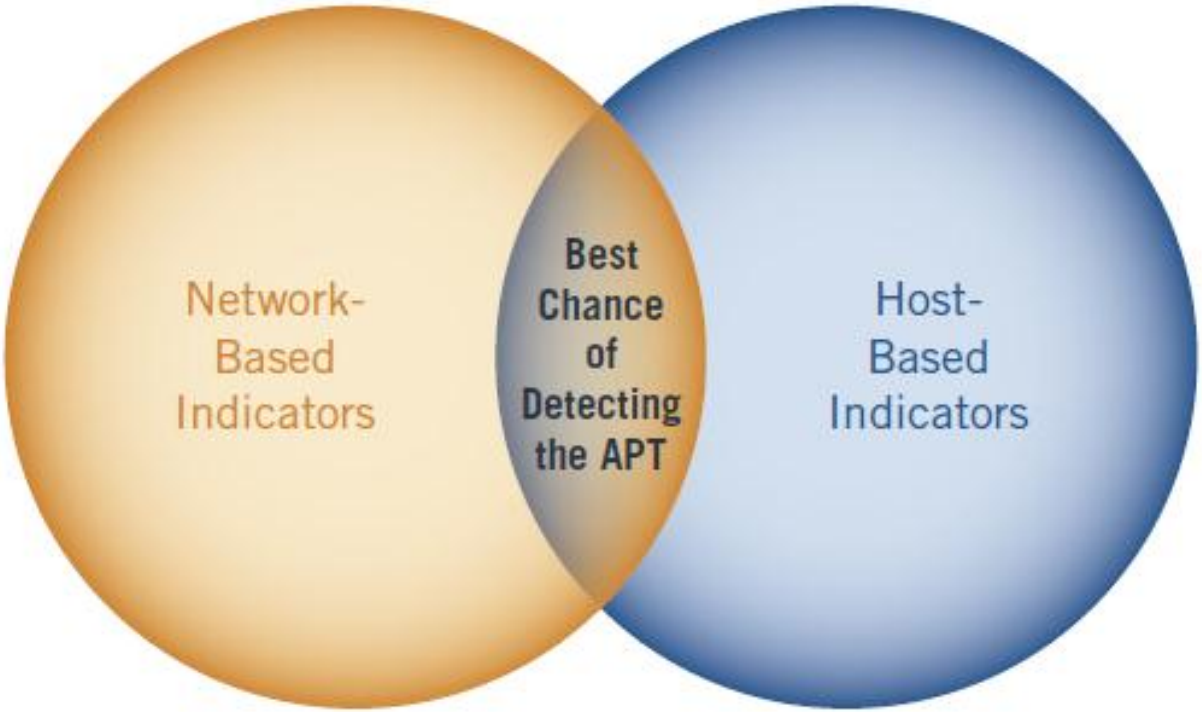
08/09/2013 16:23:51.0128event_status="(0)The operation completed successfully. "pid=1300 process_image="\John Doe\Device\HarddiskVolume1\Windows\System32\neverseenbefore.exe“ registry_type ="CreateKey"key_path="\REGISTRY\MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\ Printers Print\Providers\ John Doe-PC\Printers\{}\ NeverSeenbefore" data_type""

Endpoint Security

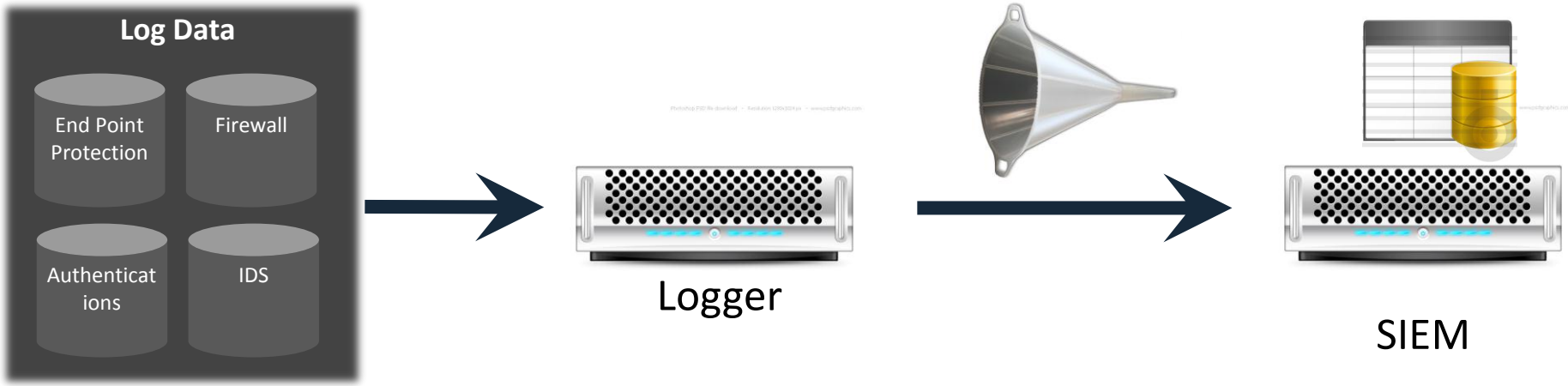
Aug 08 06:09:13 acmesep01.acmetech.com Aug 09 06:17:24 SymantecServer acmesep01: Virus found,Computer name: ACME-002,Source: Real Time Scan,Risk name: Hackertool.rootkit,Occurrences: 1,C:/Documents and Settings/smithe/Local Settings/Temp/evil.tmp,"""",Actual action: Quarantined,Requested action: Cleaned, time: 2009-01-23 03:19:12,Inserted: 2009-01-23 03:20:12,End: 2009-01-23 03:19:12,Domain: Default,Group: My Company\ACME Remote,Server: acmesep01,User: smithe,Source computer: ,Source IP: 10.11.36.20

Need Both Network and Endpoint

And Inbound/Outbound!



The Traditional SIEM Problem

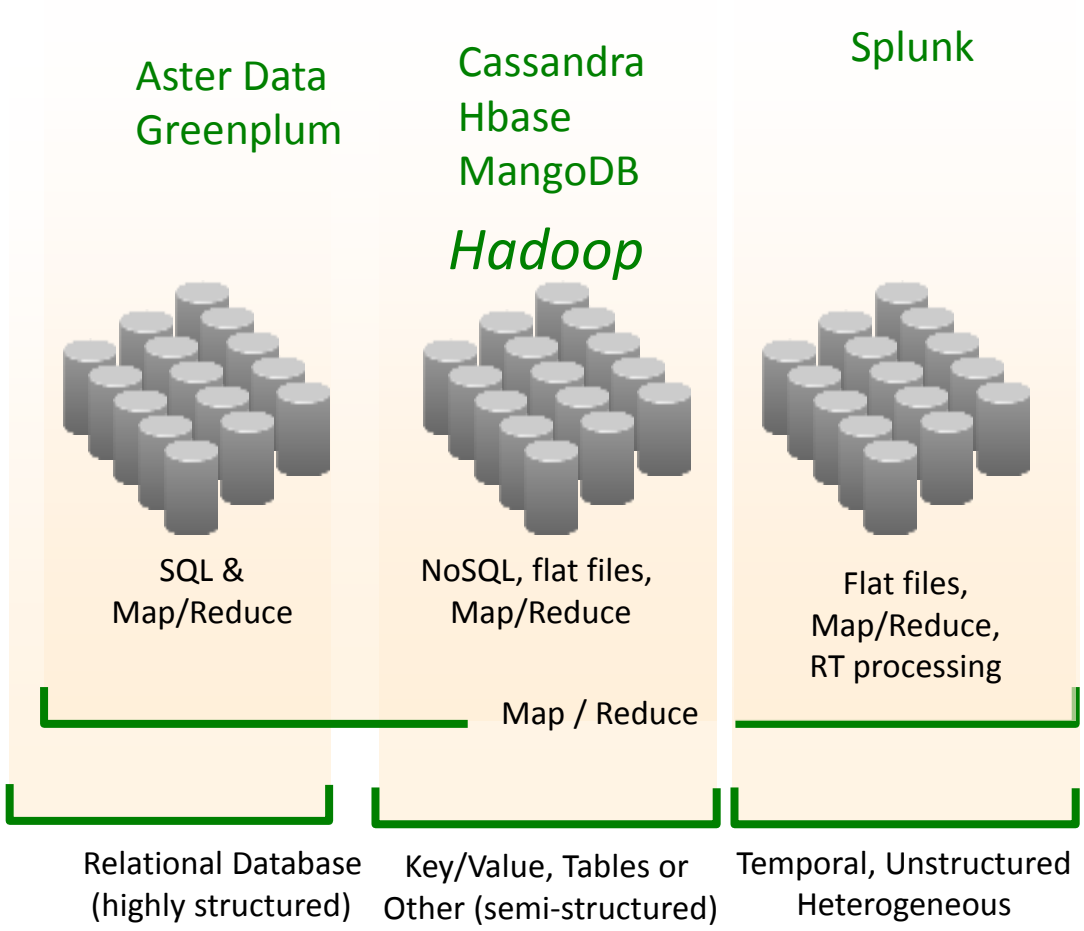


- Limited ability to do threat detection and investigation
  - Limited data sources it can ingest
  - Data reduction / normalization
  - RDMS is single chokepoint causing scale and speed issues
- Does not meet the definition of Big Data

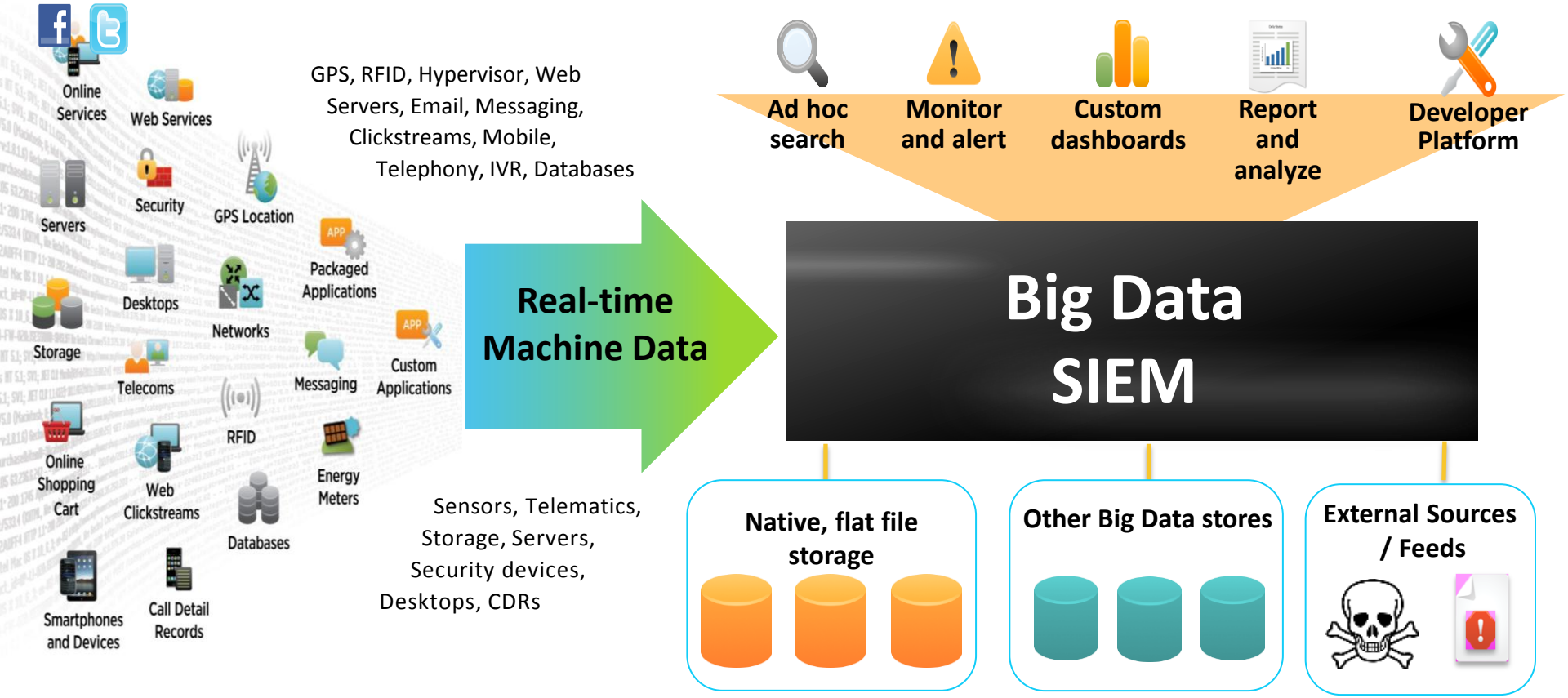
Big Data as a Solution

Big Data Technologies

- Not a SQL data store; distributed search

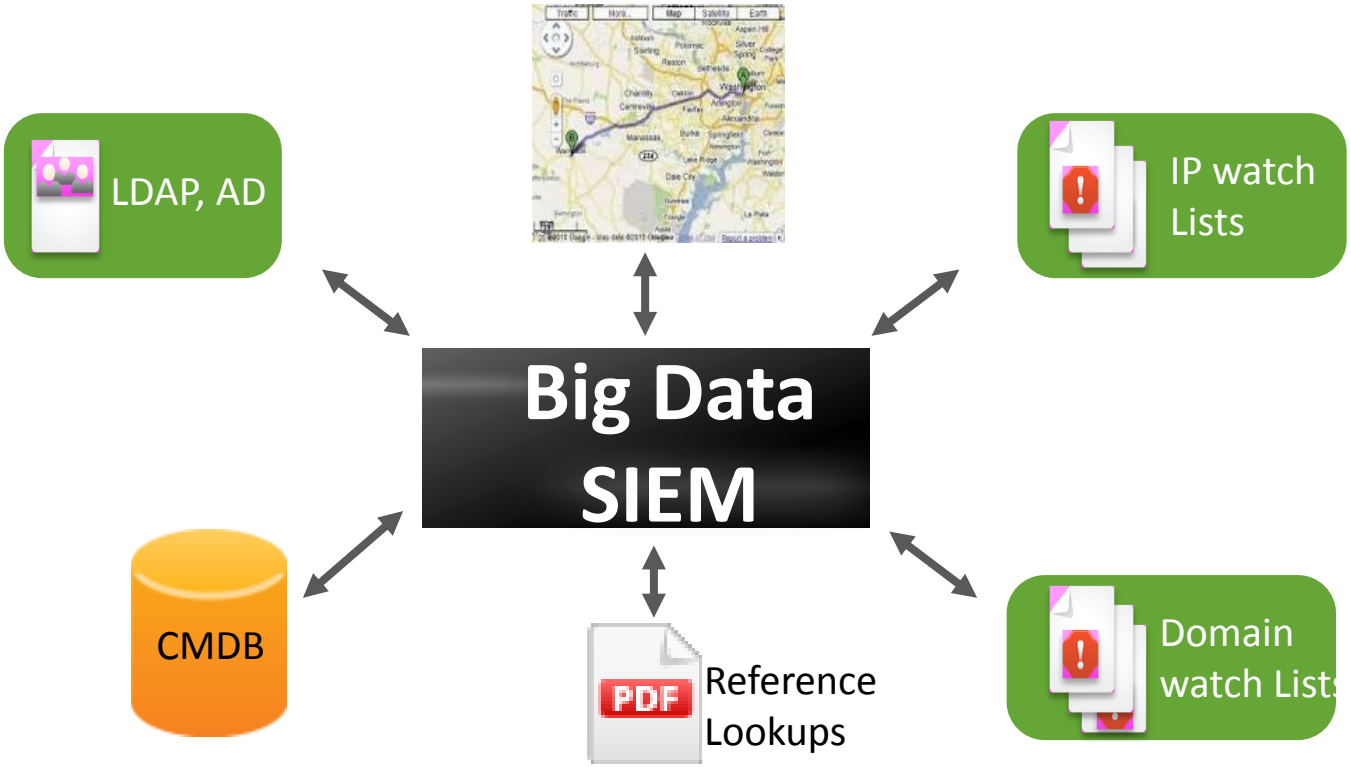


Big Data SIEM/ Security Analytics

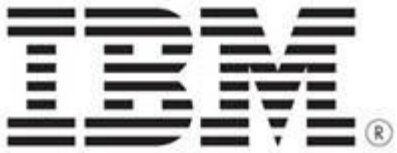


Enrich Data With External Context

Extend search with lookups and external data sources



Big-Data SIEM / Security Analytics

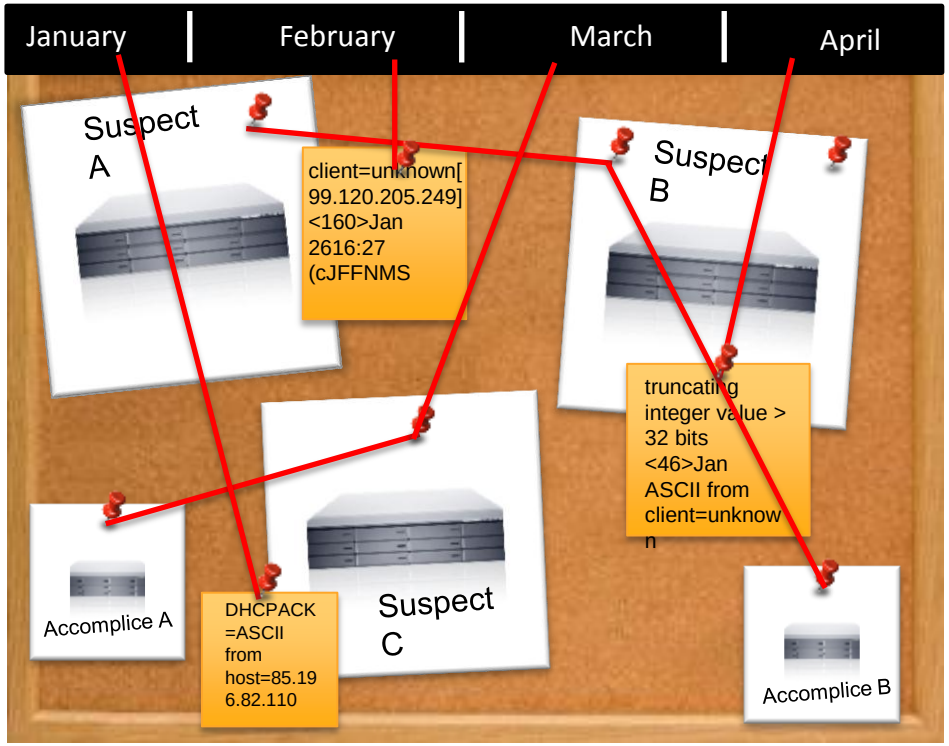


Big Data Security/SIEM Use Cases

Big Data Security/SIEM Use Cases

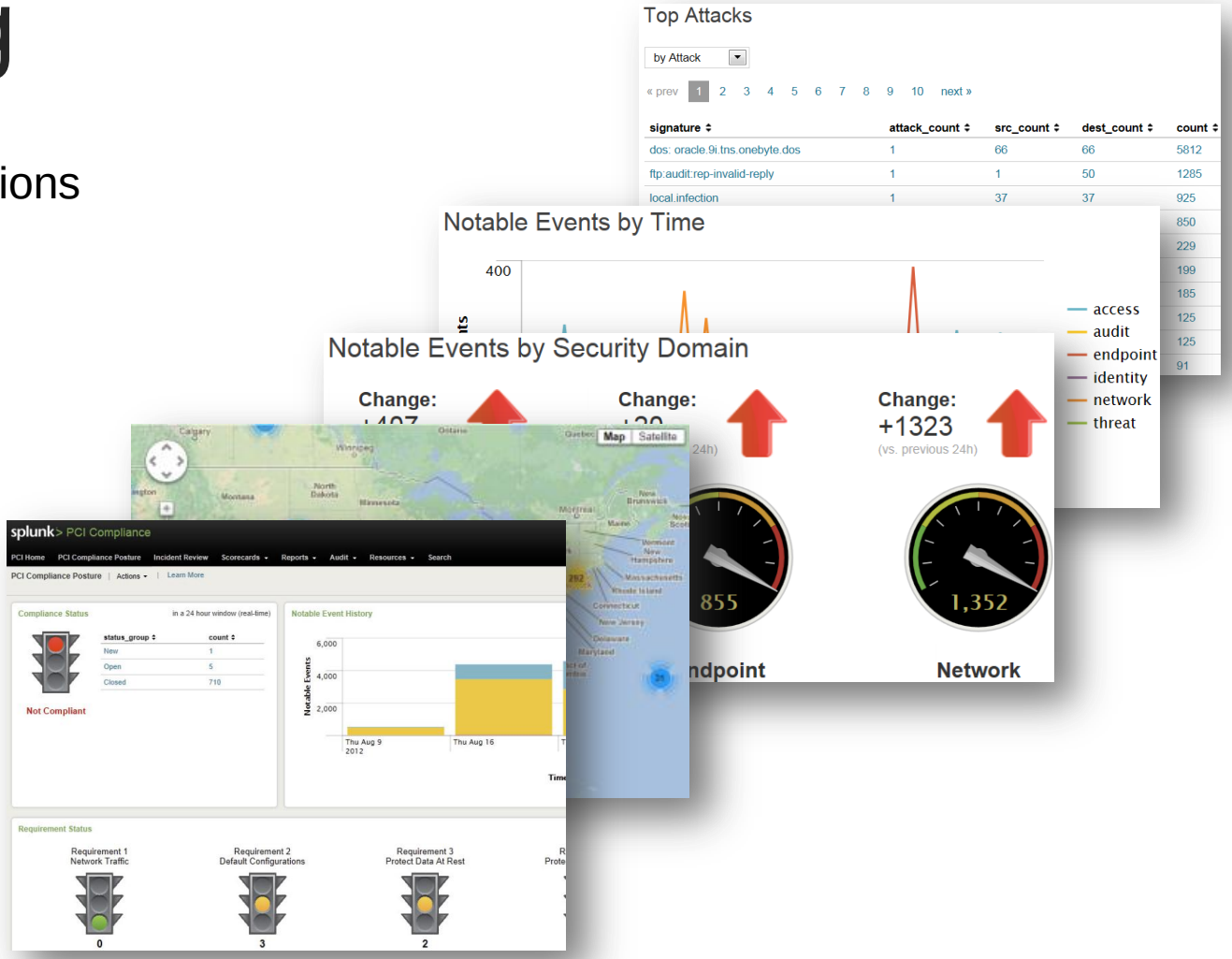
Case #1 - Incident Investigation / Forensics

- Often initiated by alert in another product
- May be a “cold case” investigation requiring machine data going back months
- Need all the original data in one place and a fast way to search it to answer:
  - What happened and was it a false positive?
  - How did the threat get in, where have they gone, and did they steal any data?
  - Has this occurred elsewhere in the past?
- Take results and turn them into a real-time search/alert if needed



Case #2 – Security/Compliance Reporting

- Many types of visualizations
  - Ad-hoc auditor reports
  - New incident list
  - Historical reports
  - SOC/NOC dashboards
  - Executive/auditor dashboards



Case #3 – Real-time Monitoring of *Known* Threats

Sources

Windows Authentication

Endpoint Security

Intrusion Detection

Time Range

Example Correlation – Data Loss

20130806041221.000000Caption=ACME-2975EB\Administrator Description=Built-in account for administering the computer/domainDomain=ACME-2975EB InstallDate=NULL LocalAccount = IP: 10.11.36.20 TrueName=Administrator SID =S-1-5-21-1 5543 50 Status=Degradedwmi_ type=UserAccounts

Default Admin Account

Source IP

Aug 08 06:09:13 acmesep01.acmetech.com Aug 09 06:17:24 SymantecServer acmesep01: Virus found,Computer name: ACME-002,Source: Real Time Scan,Risk name: Hackertool.rootkit,Occurrences: 1,C:/Documents and Settings/smithe/Local Settings/Temp/evil.tmp,"" ed,Requested ed, time: 2009-01-23 03:19:12,Inserted: 2009-01-23 03:20 :19:12,Domain: Default,Group: My Company\ACME Remote,Server: acmesep01,User: smithe,Source computer: ,Source IP: 10.11.36.20

Malware Found

Source IP

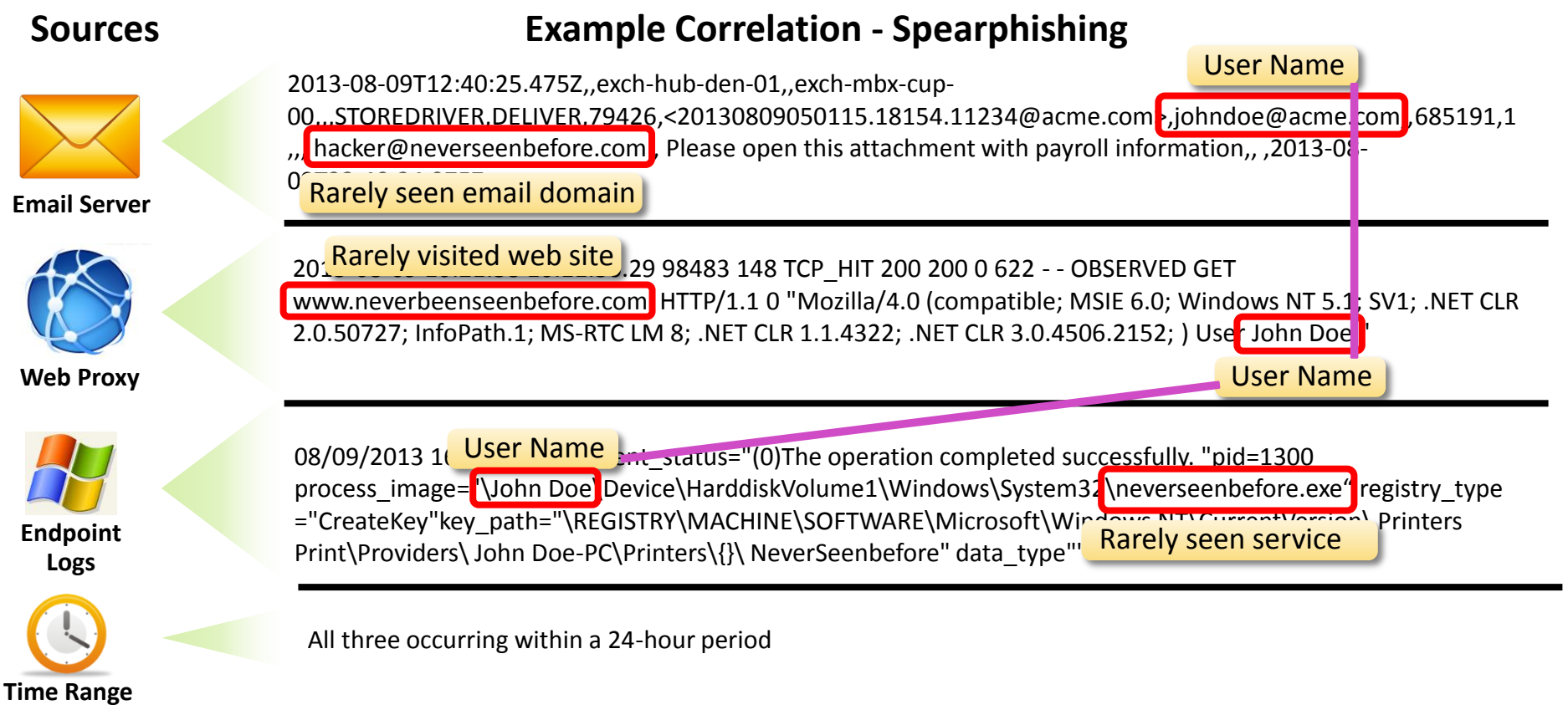
Aug 08 08:26:54 snort.acmetech.com {TCP} 10.11.36.20:5072 -> 10.11.36.26:443 itsec snort[18774]: [1:100000:3] [Classification: Potential Corp oliation] Credit Card Number Detected in Clear Text [Priority: 2]:

Source IP

Data Loss

All three occurring within a 24-hour period

Case #4 – Real-time Monitoring of *Unknown* Threats



Case #4 continued

Step 1

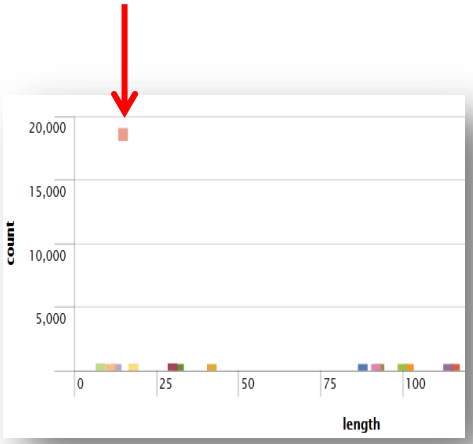
- Collect ALL the data in one location

Step 2

- Baseline/identify normal activity

Step 3

- Find outliers/anomalies



- Abnormal patterns/correlations within ‘normal’ activities
- What is rarely seen or standard deviations off the norm
- What is different/new/changed

Case #5 – Fraud Detection

Vertical	Type of Fraud	Pattern of fraud
Financial Services	Account takeover	Many transactions between $9-10k
Healthcare	Physician billing	Physician billing for drugs outside their expertise area
E-tailing	Account takeover	Many accounts accessed from one IP
Telecom	Roaming abuse	Excessive roaming on partner network by unlimited use customers
Online education	Student loan fraud	Student IP in “high-risk” country and student absent from classes & assignments

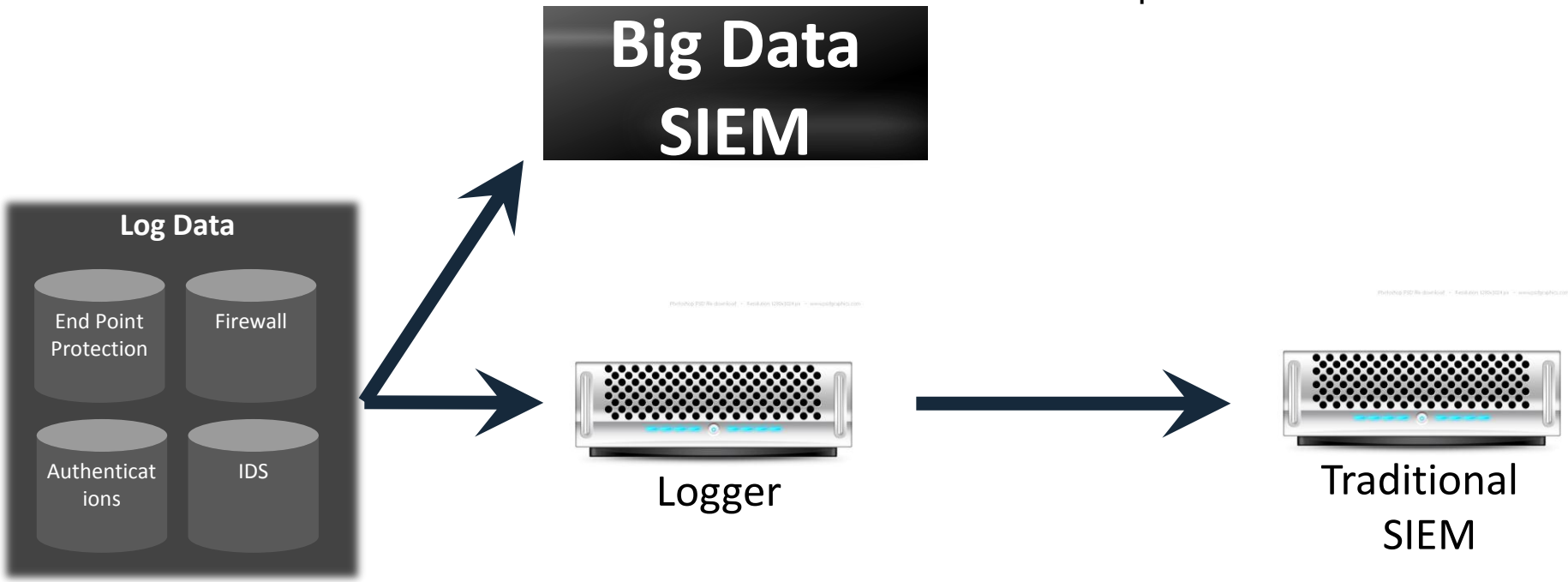
Actual Deployments of Big Data

Complement vs Replace Traditional SIEM

- Often Big Data replaces a traditional SIEM
  - SIEM & Big Data all in one
- Or Big Data complements a traditional SIEM
  - 3 common complementary deployment options

Option 1 Standalone

- Data sent to both vendors
- Big Data for incident investigations/forensics, non-security use cases
- Traditional SIEM for threat detection, alerts, workflow, compliance

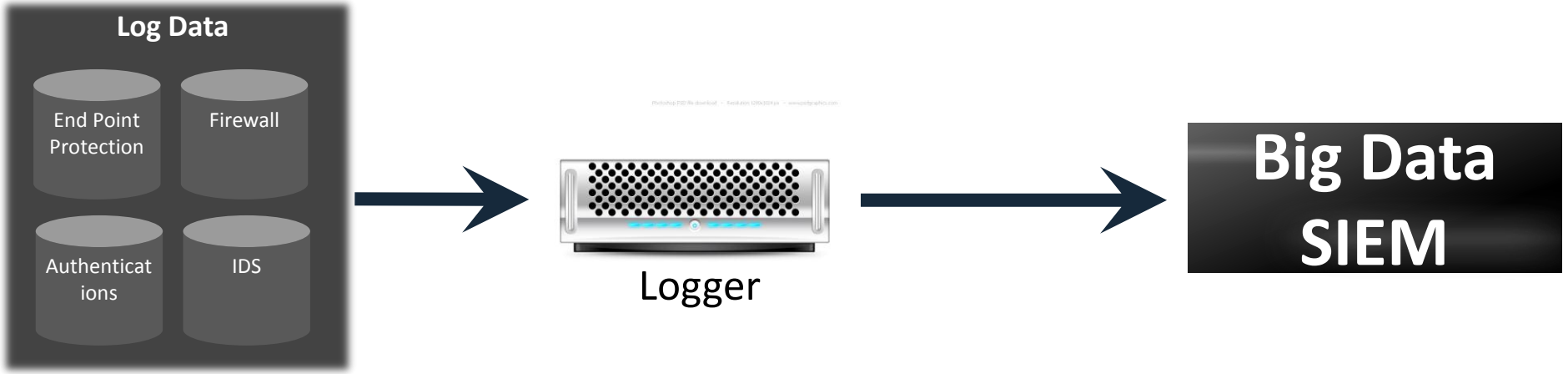


Option 2 Big Data to Traditional SIEM

- Big Data replaces Logger
- Big Data for log aggregation, incident investigation/forensics, compliance, non-security use cases
- Traditional SIEM for threat detection, alerts, workflow

Option 3 Logger to SIEM

- Big Data replaces traditional SIEM
- Big Data for incident investigation/forensics, threat detection, alerts, workflow
- Logger for log aggregation
- Takes advantage of existing Logger deployment and connectors

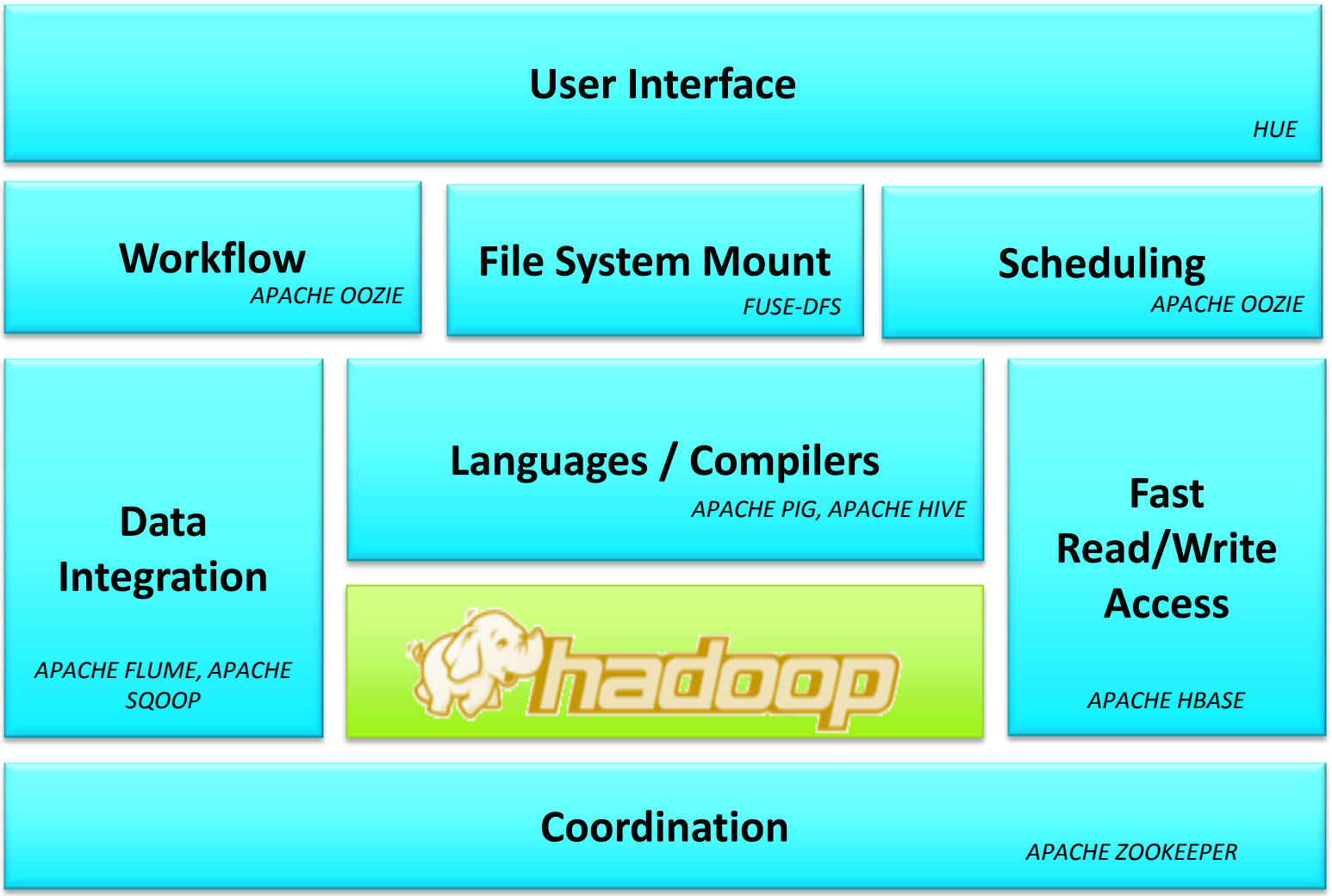


Cautions

Security Realities…

- Big Data is only as good as the data it holds and the people behind the UI
- There is no replacement for capable practitioners
- Put math and statistics to work for you
- Encourage IT Security creativity and thinking outside the box
- Fine tuning needed; always will be false positives

Be Careful of Hadoop Complexity



Watch out for Franken-SIEMs!

- Traditional SIEM with bolted on Hadoop
- Many products, data stores, UIs, middleware

Common Hadoop Scenarios

IT Security

- Hadoop for high volume, low value data
- Packet captures, NetFlow, badge logs
- Files, not data: Images, video, phone call recordings, etc.
- Ok with batch processing (not real time)

Fraud detection and prevention

- Analyze raw data from diverse sources
- Look for patterns of fraud

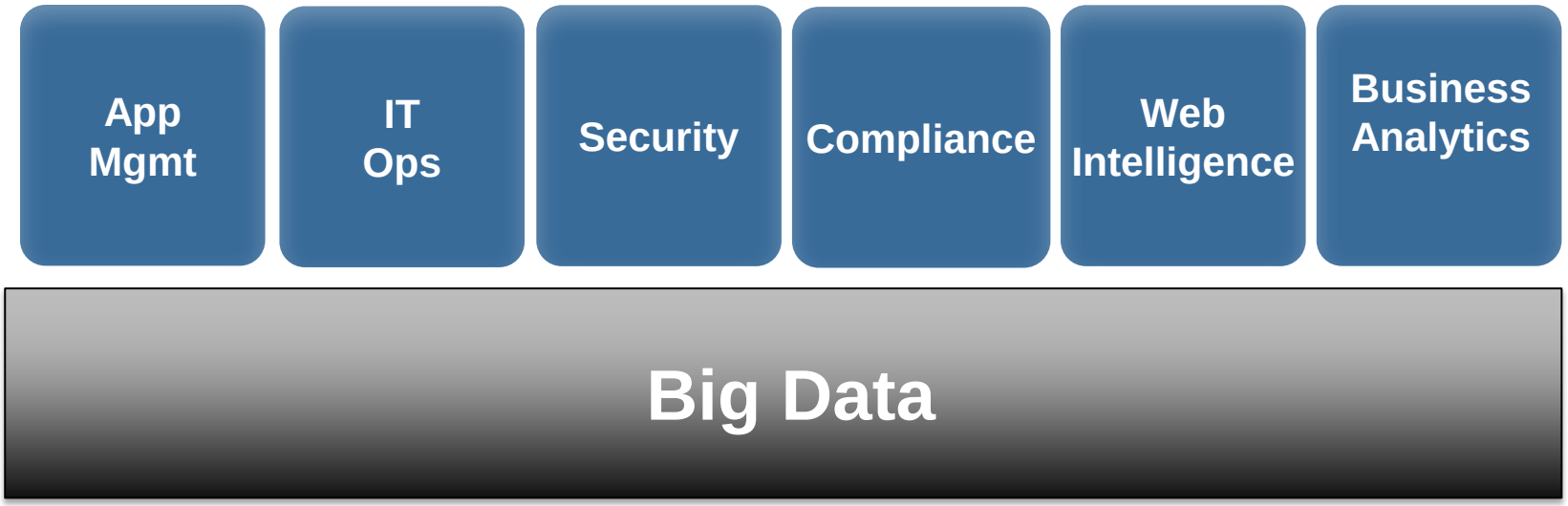
TBD

- Not sure of the use case yet or still pilot
- Hadoop as a dumping ground for data

Beyond Security

Big Data Can Be Used Across IT and the Business

- Fosters cross-department collaboration & stronger ROI



Splunk For Security

- Big Data platform for ingesting machine data; 500MB to 100+ TB/day
- Flat file data store, distributed search, common data store and UI
- Many use cases within security
  - Forensics, incident investigation, known and unknown threat detection, fraud detection, and compliance
- Many use cases outside security: IT Operations, Application Management, web analytics
- Over 6000 customers total; 2500+ primary security use case customers
- Free download and tutorial at www.splunk.com

Questions?

Thank you!

Joe Goldberg

Splunk

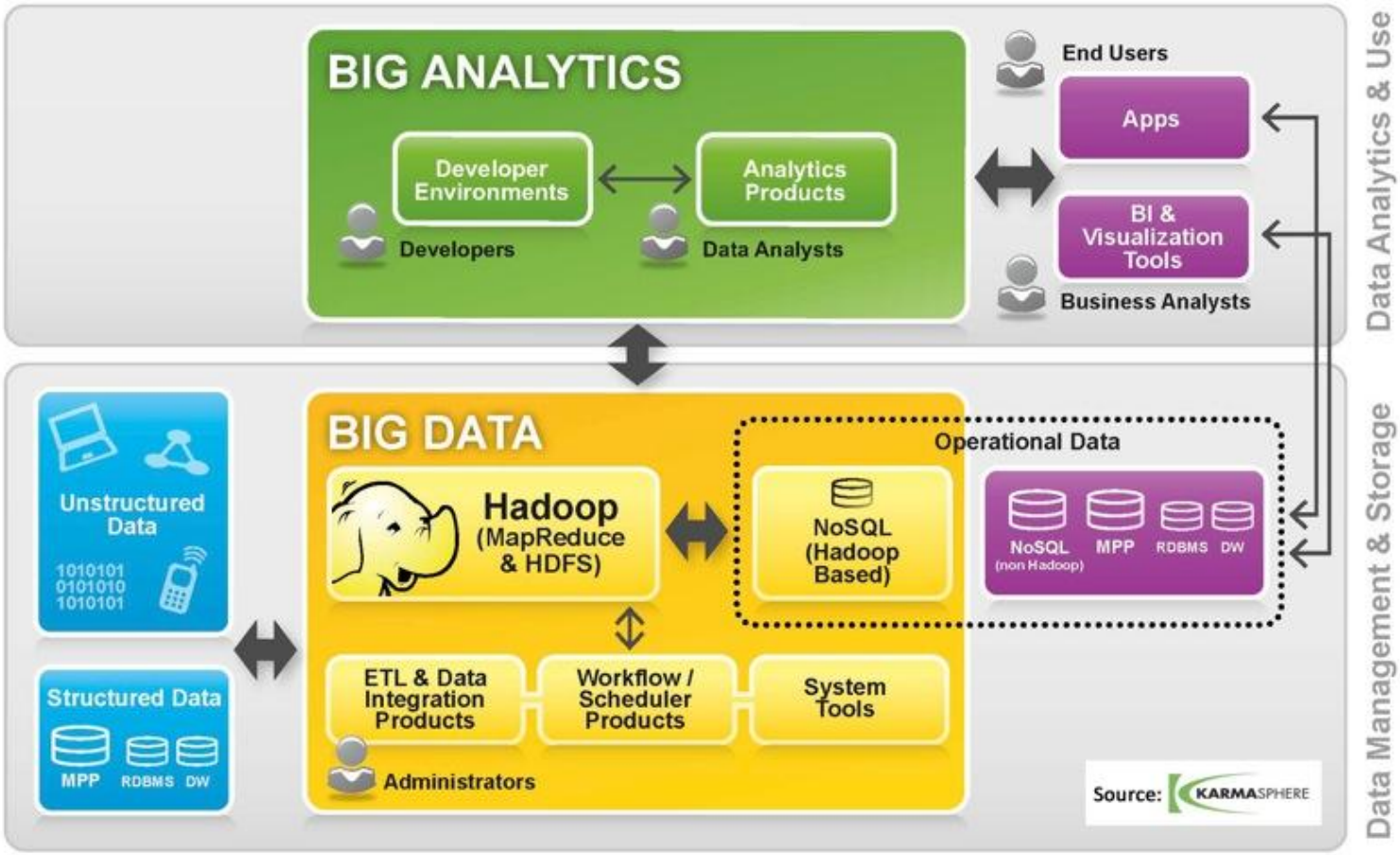
jgoldberg@splunk.com

www.splunk.com

#RSAC

RSA CONFERENCE EUROPE 2013

Be Careful of Hadoop Complexity!



Threat Intelligence Feeds