



Big Data for Government Symposium

<http://www.ttcus.com>



@TECHTrain



TTC™

Technology Training Corporation

Linkedin/Groups:
Technology Training
Corporation



Building a Big Data Architecture: A Physical and a Mental Task

About Splunk Inc.



Company (NASDAQ: SPLK)

- Founded 2004, first software release in 2006
- HQ: San Francisco / Regional HQ: London, Bethesda
- Over 1000 employees, based in 12 countries
- 2013 Revenue: \$289M (YoY +50%)

Business Model / Products

- Free download to massive scale
- Splunk Enterprise, Splunk Cloud, Premium Apps
- Hunk: Splunk Analytics for Hadoop

7,000+ Customers

- Customers in over 90 countries
- Hundreds of Federal, State, and Local agencies
- Largest license: Over 150 Terabytes per day

```
01:45:62 - - [02/Feb/2011:16:00:23] GET /product.screen?product_id=FW-428-8333000-0005176300 http://www.splunk.com/...
...category_id=FLOWERS* Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322.556; ...
...ory_id=TEDDY&JSESSIONID=9D9SL4FF4ADFF8 HTTP/1.1 200 3439 Windows NT 5.1; SV1; .NET CLR 1.1.4322.556; ...
...category_id=TEDDY* Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322.556; ...
```


splunk>6

Industry
leading
platform for
machine data

splunk>cloud™

Splunk Enterprise
as-a-Service

Hunk™

Splunk
analytics for
Hadoop

splunk>Premium Apps

ES

Splunk app for
Enterprise Security

vm

Splunk app for
VMware

PCI

Splunk app for PCI

A Platform for Security Intelligence



INCIDENT
INVESTIGATIONS
& FORENSICS



SECURITY &
COMPLIANCE
REPORTING



REAL-TIME
MONITORING OF
KNOWN THREATS



REAL-TIME
MONITORING
OF UNKNOWN
THREATS



FRAUD
DETECTION



INSIDER
THREAT

splunk>

Looking Past Traditional SIEM Use Cases

```
01:45:62 - - [02/Feb/2011:16:00:23] GET /product.screen?product_id=FW-428-8333000-000517630 http://www.splunk.com/...  
...?category_id=FLOWERS* Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; NET CLR 1.1.4322.5000) http://www.splunk.com/...  
...ry_id=TEDDY&JSESSIONID=9D9SL4FF4ADFF8 HTTP/1.1 200 3439 Windows NT 5.1; SV1; NET CLR 1.1.4322.5000 http://www.splunk.com/...  
...en?category_id=TEDDY* Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; NET CLR 1.1.4322.5000) http://www.splunk.com/...
```

splunk> listen to your data™

An App Architecture (a la SFDC)

SPLUNK APP
FOR ENTERPRISE SECURITY

130+
SECURITY APPS



splunk>

splunk> listen to your data™

The Splunk Security Intelligence Platform

Structured and Unstructured Machine Data



Behavioral Security Use Cases

Forensic Investigation

Security Operations

Compliance

Fraud/Insider Threat Detection

Hundreds of Splunk Apps and Add-ons (optional)

splunk>

HA Indexes and Storage



Commodity Servers

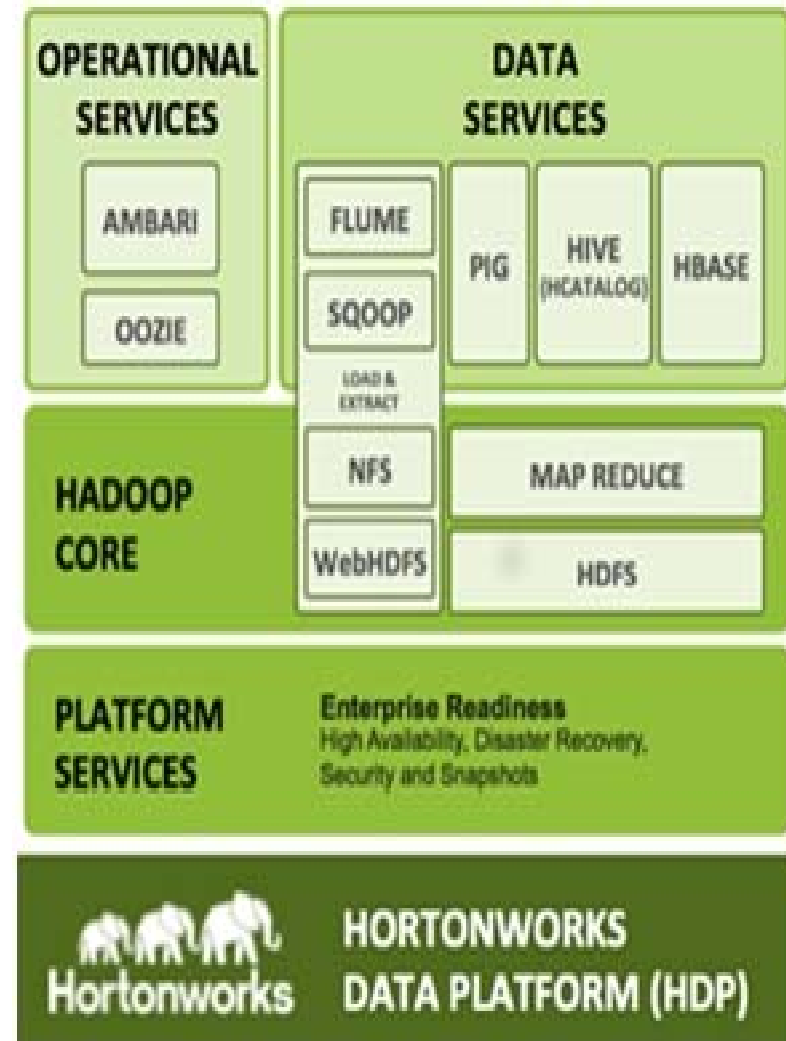
The Physical Architecture and System Requirements

Data ingest / Real-time
Search / Analyze / Alert
Visualization / Report
Scale


Two Choices for Big Data Architecture -- For Now



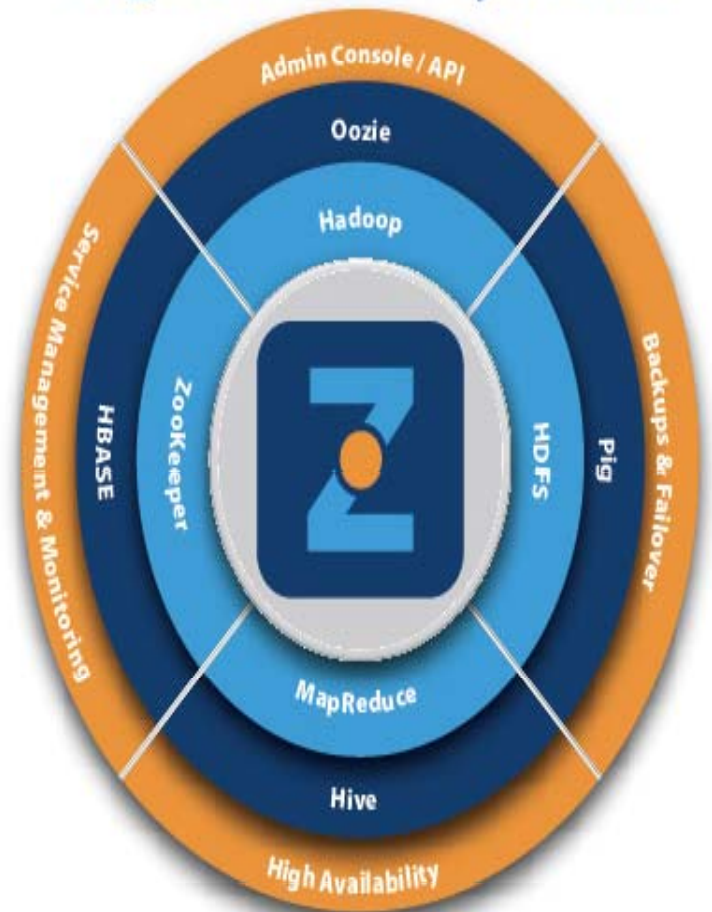
MapR / Hortonworks



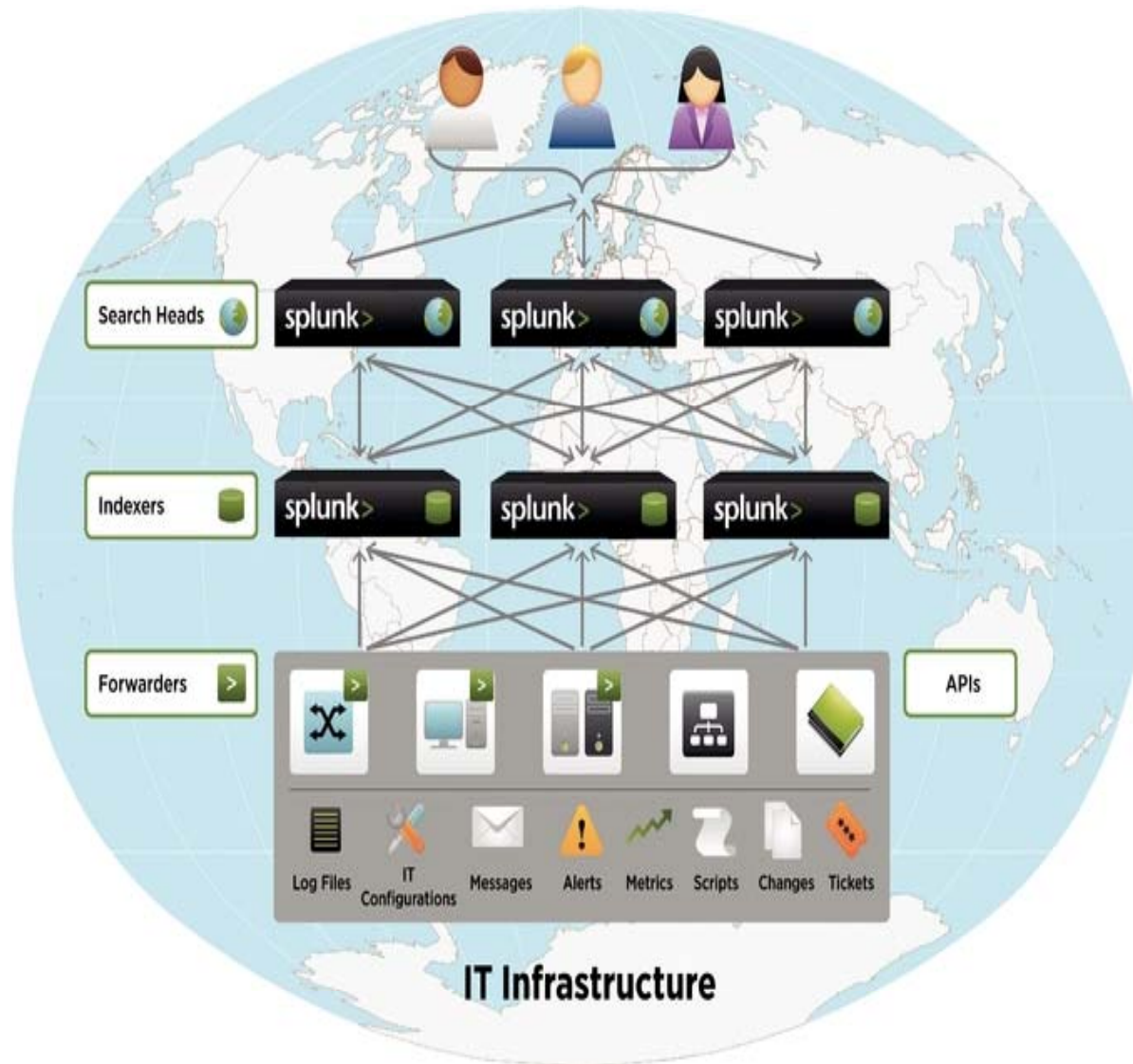
Cloudera / Zettaset

Cloudera's Distribution for Hadoop				
UI Framework		Hue	SDK	
Workflow		Oozie	Scheduling	
Metadata		Hive	Fast read/write access	
Data Integration		Languages, Compilers		Pig/ Hive
Flume, Sqoop				HBase
Coordination				
Zookeeper				

Big Data Ecosystem



Splunk



Architecting the Brain for Big Data

Connecting to a Big Data Mind Set

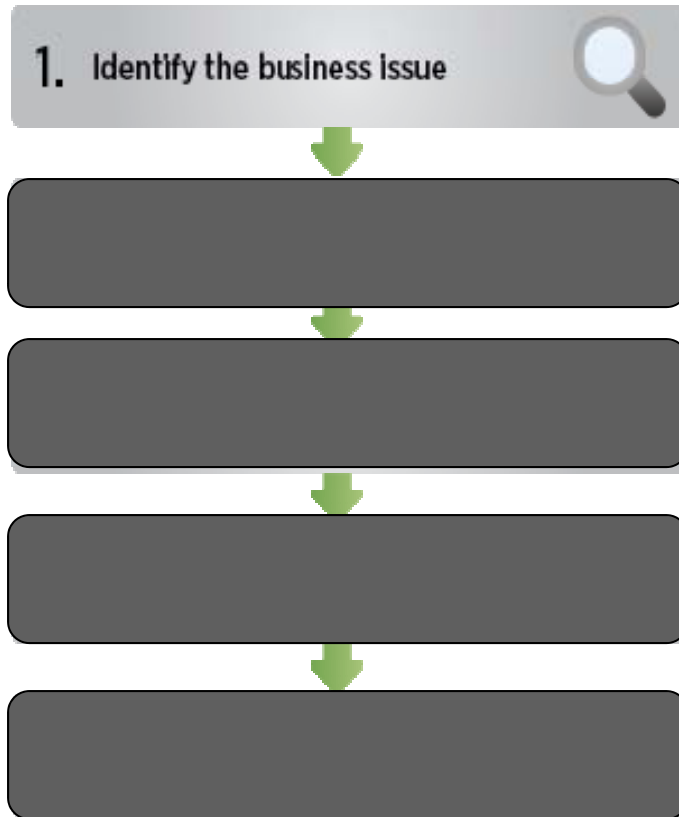
The Human Factors in Patterns and Paradigms

- Delivering consistent results are impeded by existing organizational structures, ways of thinking, product portfolios and life cycles.
- Two basic issues in human interpretations are well-supported by academic research:
 - the inclination to ignore subtle signs altogether or interpret them selectively to support the pre-existing "model,"
 - the tendency for less-experienced individuals to constantly seek additional information and input despite already having all the necessary data
- Presents problems for rules-based systems (where rules determined by humans)
- Using data analytics and mathematics doesn't require patching and is generally considered future proof

```
01:45:62 - - [02/Feb/2011:16:00:23] GET /product.screen?product_id=FW-428-8233000-040517630 http://www.myflowershop.com/...  
...?category_id=FLOWERS* Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; NET CLR 1.1.4322.578) http://www.myflowershop.com/staging...  
...ry_id=TEDDY&JSESSIONID=9D9SL4FF4ADFF8 HTTP/1.1 200 3439 Windows NT 5.1; SV1; NET CLR 1.1.4322.578 http://www.myflowershop.com/staging...  
...?category_id=TEDDY* Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; NET CLR 1.1.4322.578) http://www.myflowershop.com/staging...
```

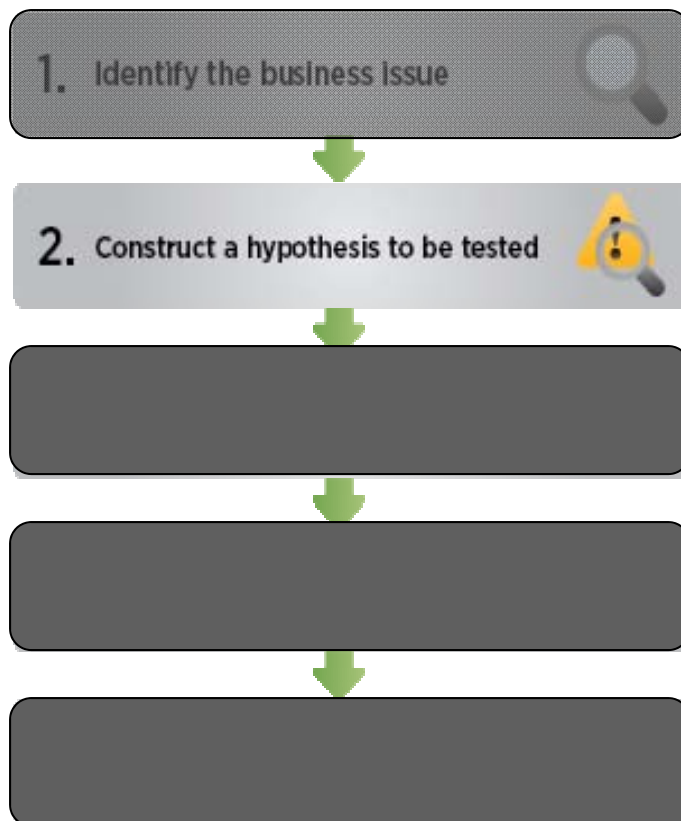

A Process for Using Big Data for Security:

Identify the Business Issue



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

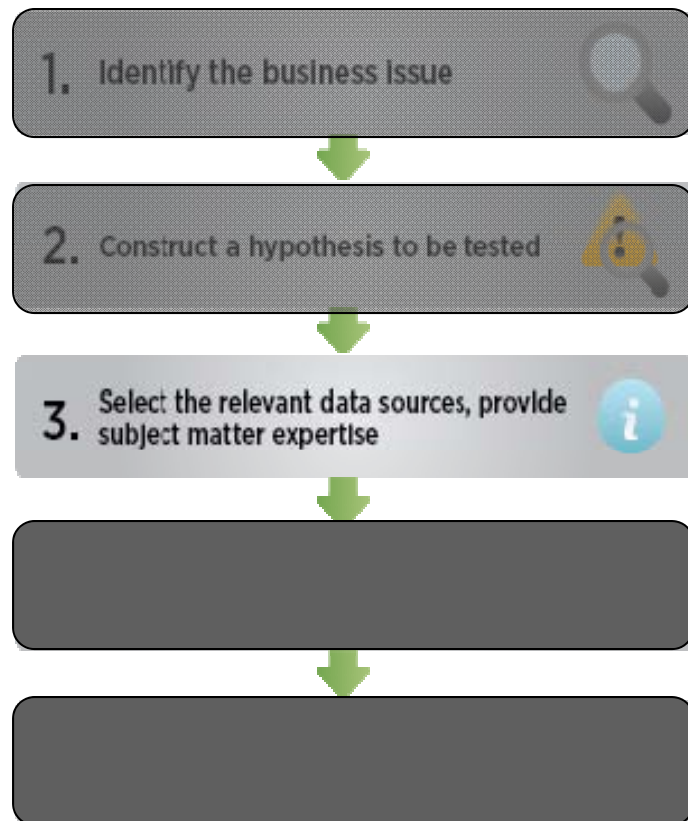
A Process for Using Big Data for Security: Construct a Hypothesis



- 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?

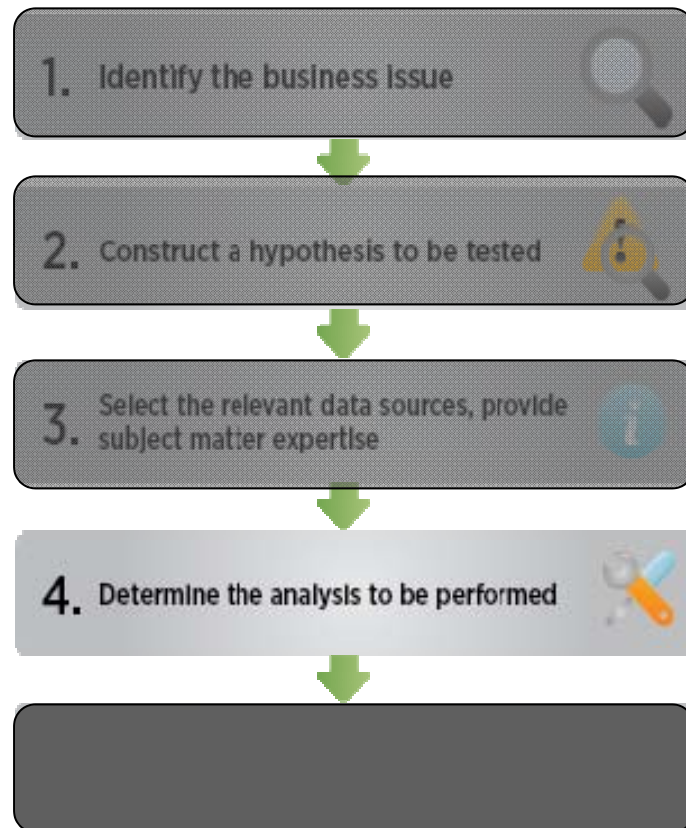
```
01:45:62 - - [02/Feb/2011:16:00:23] GET /product.screen?product_id=FW-428-8333000-940351&category_id=FLOWERS Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322.5860) http://www.myflowershop.com/category.screen?category_id=FLOWERS&JSESSIONID=9D9SL4FF4ADFF8 HTTP 1.1* 200 3439 Windows NT 5.1; SV1; .NET CLR 1.1.4322.5860 http://www.myflowershop.com/category.screen?category_id=TEDDY&JSESSIONID=9D9SL4FF4ADFF8 HTTP 1.1* 200 3439 Windows NT 5.1; SV1; .NET CLR 1.1.4322.5860 http://www.myflowershop.com/category.screen?category_id=TEDDY&JSESSIONID=9D9SL4FF4ADFF8 HTTP 1.1* 200 3439 Windows NT 5.1; SV1; .NET CLR 1.1.4322.5860 http://www.myflowershop.com/category.screen?category_id=TEDDY&JSESSIONID=9D9SL4FF4ADFF8
```

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 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.

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?

Supporting Agency Mission with Big Data Analytics

eRegulations Insights taps into nearly 1.2 million comments received by federal agencies since the beginning of 2012.

eRegulations Insights

regulations.gov



splunk
POWERED

Overview Agencies Regulations Influencers Phrases Explore How it Works

Agencies Insights

This page allows you to explore federal agencies and regulations through real-time visualizations, using open data retrieved directly from Regulations.gov and powered by Splunk. Below you'll find details on the most popular agencies, including comments over time, agencies with the most duplicate comments, and the most liked regulations based on sentiment analysis scores.

Click on any of the panels below for more details, or use the [Agency Explorer](#) to explore on your own!

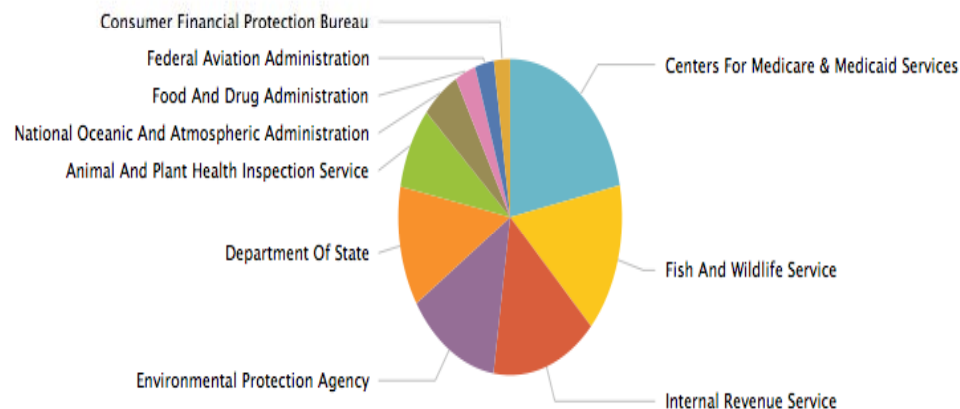
121
GOVERNMENT AGENCIES

1,207,661
TOTAL COMMENTS

48,613
UNIQUE COMMENTS

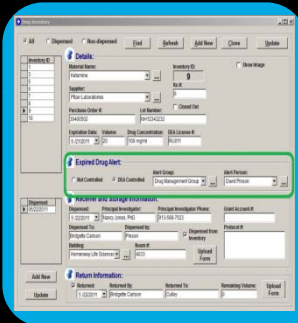
January 01, 2012 to June 16,
2014
IS THE TIME PERIOD COVERED

Popular Agencies

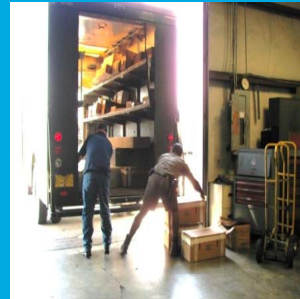


Medical Supply Chain Fraud and Abuse

15% of Medical professionals undergo some form of medical addiction during their careers.



**Drug
Ordered**



**Order
Checked in**



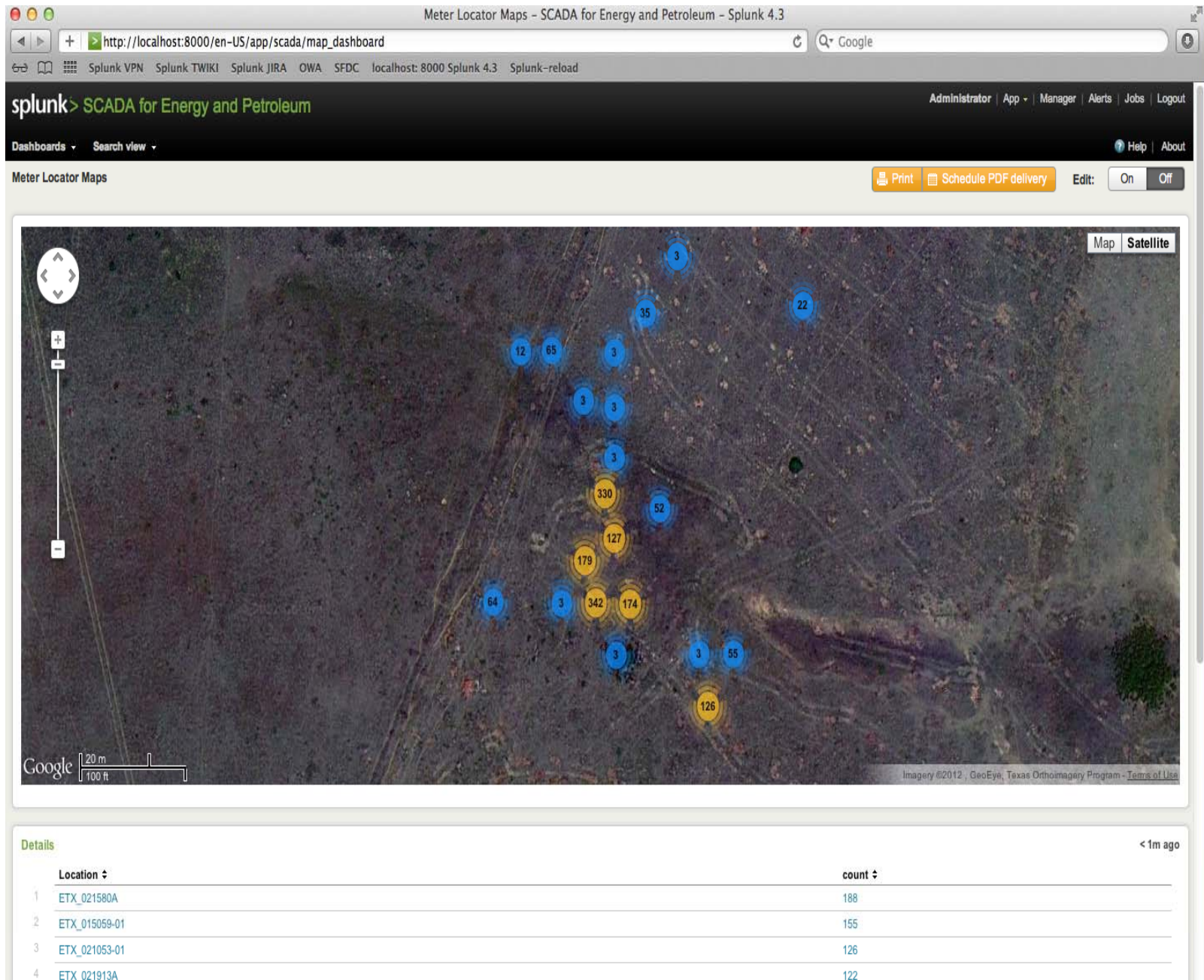
**Prescribed to
Patient
Delivered by
Nurse**



**Shipped
to Hospital**



**Distributed to
Pixis**



Meter Detail - SCADA for Energy and Petroleum - Splunk 4.3

http://localhost:8000/en-US/app/scada/meter_drilldown?q=search%20index%3Dscada%20sourcetype%3D%22BSAP%22%20Location%3D%22jblake-mbp15.local%22

Splunk VPN Splunk TWIKI Splunk JIRA OWA SFDC localhost: 8000 Splunk 4.3 Splunk-reload

splunk> SCADA for Energy and Petroleum

Administrator | App | Manager | Alerts | Jobs | Logout

Dashboards Search view Help About

Meter Detail

Print Schedule PDF delivery Edit: On Off

Protocols in Use at This Location

< 1m ago

« prev 1 2 3 4 5 6 next »

	_time	Location	host	index	linecount
1	3/5/12 11:58:13.742 PM	ETX_040102A	jblake-mbp15.local	scada	1
2	3/5/12 11:58:13.671 PM	ETX_040102A	jblake-mbp15.local	scada	1
3	3/5/12 11:58:13.573 PM	ETX_040102A	jblake-mbp15.local	scada	1
4	3/5/12 11:57:19.301 PM	ETX_040102A	jblake-mbp15.local	scada	1
5	3/5/12 11:57:19.193 PM	ETX_040102A	jblake-mbp15.local	scada	1
6	3/5/12 11:57:19.135 PM	ETX_040102A	jblake-mbp15.local	scada	1
7	3/5/12 11:57:19.117 PM	ETX_040102A	jblake-mbp15.local	scada	1
8	3/5/12 11:57:19.046 PM	ETX_040102A	jblake-mbp15.local	scada	1
9	3/5/12 11:57:18.995 PM	ETX_040102A	jblake-mbp15.local	scada	1
10	3/5/12 11:57:18.902 PM	ETX_040102A	jblake-mbp15.local	scada	1

View results

Meter Characteristics

Model 9200 - Electronic Oil Meter


Manufactured by:

- Max Pump Pressure: 1000 ft/lbs
- Max Meter Temp: 1200 F
- Manufacture Date: June, 2009

Technical Service Bulletins

Currently none on file by the manufacturer

Model 9200 - Electronic Oil Meter



Realtime SCADA events

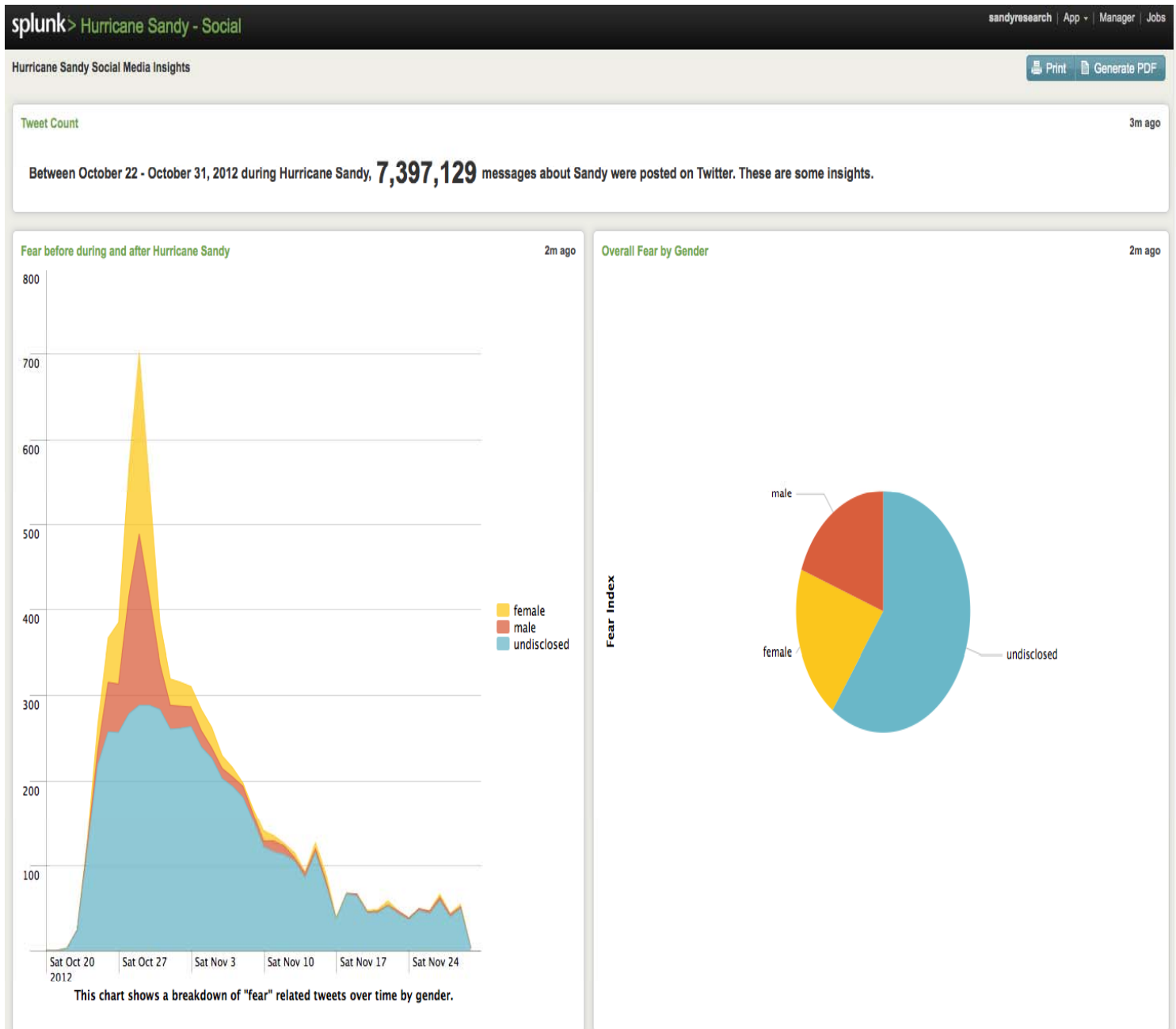
< 1m ago

« prev 1 2 3 4 5 6 next »

	_time	Location	host	index	linecount	source	sourcetype	splunk_server	_raw
1	3/5/12 11:58:13.742 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:58:13,742,BSAP,1,[d88],3,ETX_040102A,COMPLETE Response: start 1: length 94
2	3/5/12 11:58:13.671 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:58:13,671,BSAP,1,[d88],3,ETX_040102A,RX: 52 68 79 6d 65 73 20 77 69 74 68 20 42 65
3	3/5/12 11:58:13.573 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:58:13,573,BSAP,1,[d88],2,ETX_040102A,TX: 52 68 79 6d 65 73 20 77 69 74 68 20 42 65
4	3/5/12 11:57:19.301 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:57:19,301,Pccu,1,[d88],2,ETX_040102A,TX(52): 71 69 77 56 7a 39 30 6e 36 6d 65 45 31
5	3/5/12 11:57:19.193 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:57:19,193,Pccu,1,[d88],2,ETX_040102A,TX(51): 77 6b 54 4d 48 79 6c 59 72 4b 70 6a 4c
6	3/5/12 11:57:19.135 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:57:19,135,Pccu,1,[d88],2,ETX_040102A,TX(50): 41 5a 34 76 55 65 6b 45 68 4b 61 74 6c
7	3/5/12 11:57:19.117 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:57:19,117,Pccu,1,[d88],2,ETX_040102A,TX(49): 72 45 50 59 52 6a 79 78 36 65 4d 78 43
8	3/5/12 11:57:19.046 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:57:19,046,Pccu,1,[d88],2,ETX_040102A,TX(48): 6e 30 6e 66 70 65 4d 65 71 6a 7a 46 6a
9	3/5/12 11:57:18.995 PM	ETX_040102A	jblake-mbp15.local	scada	1	/opt/splunk/43/scada/splunk/etc/apps/scada/logs/BSAP.log	BSAP	jblake-mbp15.local	23:57:18,995,Pccu,1,[d88],2,ETX_040102A,TX(47): 77 31 4f 51 31 76 6d 46 4f 69 41 6b 48

Social media for emergency management >7M tweets

- How fear changed as Hurricane Sandy approached
- How many people asked for help over time
- Critical Supplies sentiment analysis
- Rate of people evacuating over time.



Monitoring Data Access for HIPAA

Monitor and report on EHR access based on three factors

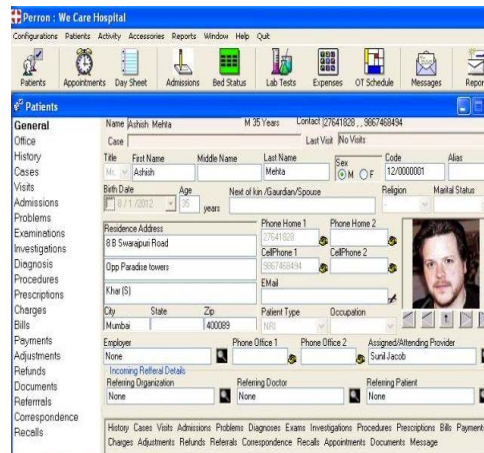
Temporal

The time element
of access



Situational

Positive caregiver /
patient assignment



Appropriate

Caregiver / patient
relationship / User role



Medical Device Supply Chain Insights

**Device
Manufactured**



**Prescribed
to patient**



**Shipped
to Physician**



**Returned
to iRhythm**

**Tracking Medical
Device Supply
Chain to Drive
Critical Insights**

Patient Behavior

Prescription
Patterns

Supply Chain
Analytics

Defect Tracking

Understanding Insider Threats with Splunk

Statistical Outliers

- Rare network printer use (the one not closest to employee)
- Changes in web surfing data / site type ratios
- Activities unusual times of day

Personal Context

- Unused Vacation - 24 months or longer
- Credit score falls 200 points
- Marital status change – emotional stress
- Demotion, new supervisor, transfer

Direct Match / Correlation

- Use of 'back door' and default accounts
- Access to network diagrams and code repositories
- Monitor physical access logs to unauthorized systems / locations

Know malicious intent



Thank You
Free Download:
www.splunk.com

