How We Built an Efficient Healthcare Privacy Monitoring and Auditing Platform

Gleb Esman gesman@splunk.com | Sr. Project Manager, Fraud Analytics and Research, Splunk.
Ernst Katchour ernst@sigbay.com | CEO and Founder, SigBay, Inc., Splunk Technology Alliance Partner
Jay Benfield jay@sigbay.com | Senior Architect, SigBay, Inc.

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance 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 could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. 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 functionality described or to include any such feature or functionality in a future release.

splunk> .conf18

Gleb Esman, Bio

gesman@splunk.com linkedin.com/in/glebesman

1990's: **Anti-virus, anti-malware research and development**:
Belarus, Israeli anti-virus research and development.

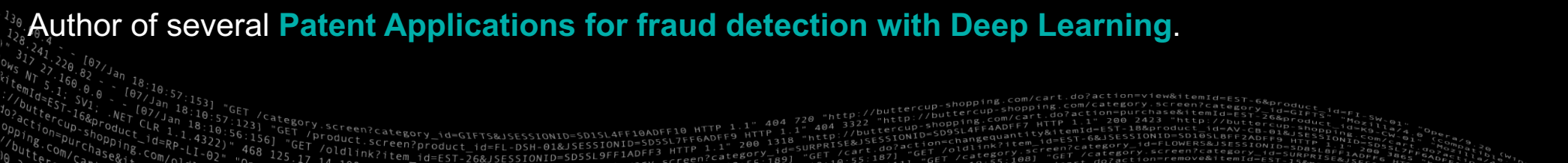
2000's: **IBM T. J. Watson Research Center**, NY. Anti-virus development.
Development of advanced methods and heuristic virtual machines to detect known and unknown computer viruses and malware.

2000's-2010's:
Architecting and engineering management work in space of **e-commerce, cryptocurrency, payment processing and digital information management** solutions.

Till July, 2015: **Morgan Stanley**, Montreal, Canada.
Data analytics solutions for financial services, building custom Splunk-based security and anti-fraud applications.
Leading an effort to leverage Splunk as an anti-fraud platform for online banking.

Since August, 2015: Sr. Product Manager, Fraud Analytics and Research at **Splunk**, San Francisco Managing projects in fraud analytics, advanced threat detection and investigation spaces.

Author of several **Patent Applications for fraud detection with Deep Learning**.



splunk> .conf18

Challenges With Existing Off-the-shelf Privacy Systems

- **Scalability and Performance**
  - Legacy designs involves complex, poorly documented, mixed architectures
- **Rigid and Inflexible**
  - Hardcoded to specific data formats
  - Hardcoded to specific interfaces, limited APIs
- **Lack of control**
  - Hard to modify and customize without vendor
  - Requires vendor-driven, expensive and prolonged consulting engagements
  - Often “black box” with unwillingness by vendor to cooperate with others
- **Crippling, add-on costs**
  - Vendor-enforced fees to maintain system is working order and upgrade

General Requirements for Privacy Platform

- Scalability
- Extensible, customizable solution
- Support for many privacy use cases
- Support for large number of diversified, poorly documented, poorly structured, possibly mis formatted data sources coming in large quantities in possibly erratic manner from large number of different healthcare applications, systems and possibly unstable data sources and activity logs.
- Have system capable of normalizing, analyzing and detecting thousands of anomalies, instances of violations and suspicious activity events over critical patient and healthcare data.

General Requirements for Privacy Platform, cont.

- Ability to add unlimited number of new use cases quickly
- Ability to tune system metrics, detection thresholds
- Ability to detect anomalies with minimal amount of false positives
- Ability to generate custom reports, visualize data, customizable tables and one-click drilldowns
- Implement in-system case management
- Ability to monitor data consistency and data flow stability

Splunk Healthcare Privacy Platform

Full Data Visibility

- Data Flow (consistency, stability)
- EHR Records (normalized and raw views)
- Activity (EHR access stats across all entities)
- Anomalies (ML and detection)

Business Use Cases

- Privacy Monitoring
  (EHR access)
- Medications Access
  (Pharmacy, Access, Diversion, Opioids abuse)
- Security
  (Logins, IDs, System access)

Incident Management

- Alerts (Categories, Tags, Status, Comments)
- Filtering
- Workflow

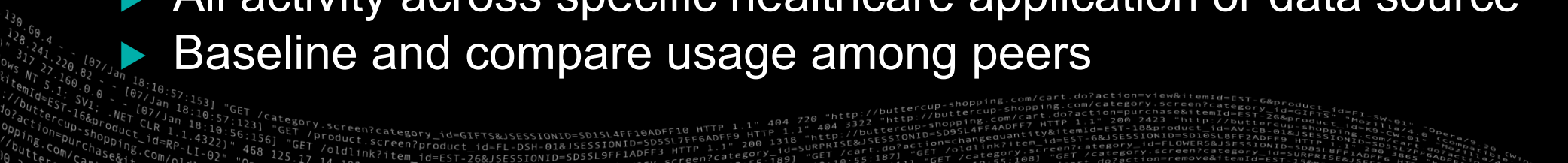
Investigations

- Dashboards
- Visualizations
- Drilldowns
- Reports

splunk> .conf18

Specific Use Cases:

- **Suspicious access to patient records:**
  - Access outside of working hours
  - VIP patient records access
  - Patient record peeking
  - Employee access to other employee records
  - Deceased patient records access
  - Failed logins, access by inactive users
- **Suspicious Access to Medications:**
  - Medication removal and access in suspicious way
  - Anomalous activity over controlled substances
  - Pharmacy cabinet access in anomalous manner
- **Investigation Dashboards:**
  - All activity across campus, department or specific provider titles
  - All activity by specific provider
  - All activity on specific patient's record
  - All activity across specific healthcare application or data source
  - Baseline and compare usage among peers



Successful Healthcare Monitoring Platform

Key Features of successful Healthcare Monitoring and Auditing Platform

Ernst Katchour,
ernst@sigbay.com

CEO and Founder, SigBay, Inc., Splunk Technology Alliance Partner

splunk> .conf18

Business Drivers

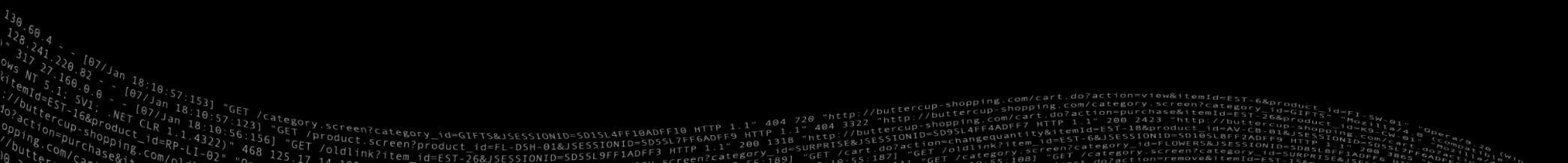
What is the main goal of the project in a business perspective

- Scalable, extensible solution to facilitate Privacy Officer's operations in
  - Auditing
  - Monitoring
  - Investigations
- Improve existing processes.
  - Provide access across all records.
  - Reliable and timely data for investigations
  - Ability to manage user roles
- Increase efficiency
  - Reduce false positive alerts
  - Create consistent workflow
  - Direct information to the appropriate privacy office for investigation

Technical Drivers

What is the main goal of the project in a technical perspective

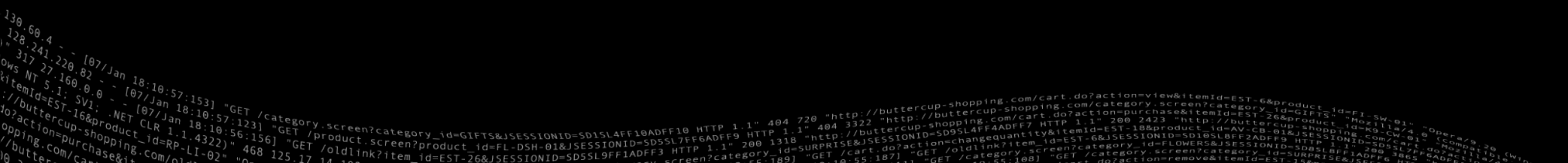
- Aggregate data from multiple applications
- Define normalized data models
- Centrally monitor patient record access
  - Monitor privacy violations, generate alerts
  - Define and implement workflow for privacy investigations
- Provide privacy reports
- Provide data integrity reports



Technical Challenges

What issues did we run into on the technical side

- Data comes in a dizzying array of unpredictable formats and structures and in a large variety, velocity and volume.
- Mix of structured (csv, tsv, psv, other) and unstructured data
- Some complexities in terms of field or time extraction
- Creative regex-ing needed to clean garbage characters without pre-processing
- Not all data was available at the same time



Existing Solutions

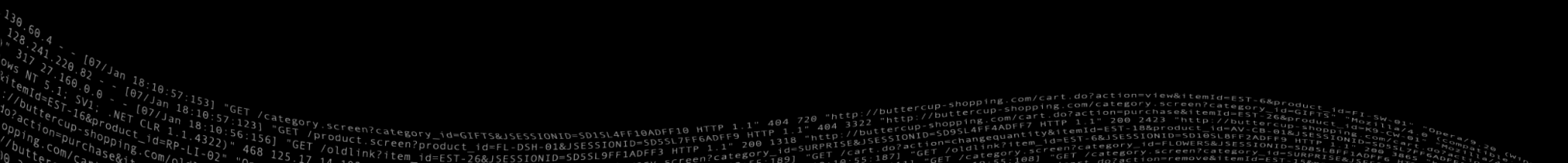
What are the issues with existing non Splunk Solutions

- Expensive
- In some deployments only fraction of functionality is used
- Data must match internal data models and fit predefined standarts
- Require customization with every deployment (every customer has custom EMR data to meet each doctor's requirements)
- Require ongoing maintenance/management by a team of 4-7 people to get to a usable state

Splunk Advantage

Why use Splunk

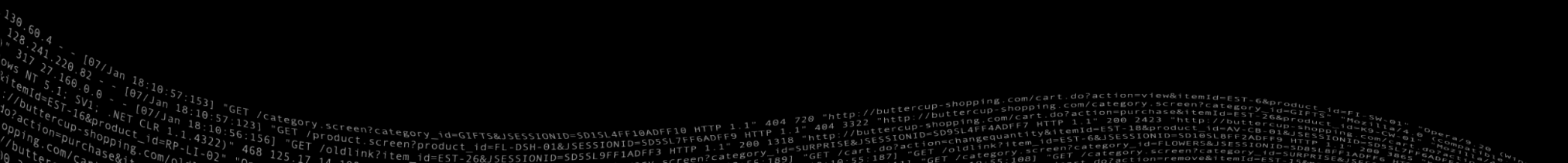
- Extremely easy to onboard any data and customize Splunk to that data
- Extremely easy to maintain
- Can complement existing solutions
- Fast time to value
- Real time alerting capabilities
- Can help monitor compliance with Meaningful Use



Key Components

Data Models representing data from various application logs

- EPIC
- Agility (Employee Health EMR Software)
- Athena
- ARMS (Ambulatory Revenue Management Software)
- Cerner Lab
- CROWN Eagle
- eCompas
- Allscripts SCM
- Eclipsys SRM
- EMPI (Enterprise Master Patient Index)
- EzVac (Vaccination Immunization Information System)
- IDX
- ImageCast
- iNYP
- IPRS (International Patient Relations System)
- OR Manager
- PACS
- PSS (Patient Safe Solutions for medication dispensing)
- Soarian
- Streamline
- Teleresults



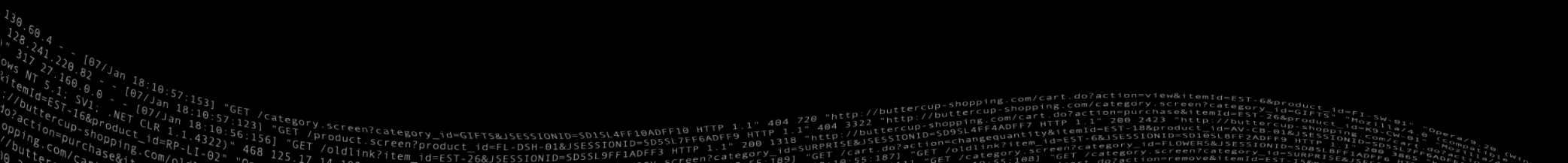
splunk> .conf18

Key Components

Use Cases

- Employee Access to Employee Patient Records
- Failed Login Attempts
- Compare Usage Among Peers
- Access by Inactive User
- Access to VIP
- Accesses by User Over Time
- Accesses by Users to a Given MRN
- Higher than User Normal
- Excessive Hours with Activity
- Access Outside of Work Hours
- Excessive Demographics Access

- Employee Access to Employee Patient Records
- Failed Login Attempts
- Compare Usage Among Peers
- Access by Inactive User
- Access to VIP
- Accesses by User Over Time
- Accesses by Users to a Given MRN
- Higher than User Normal
- Excessive Hours with Activity
- Access Outside of Work Hours
- Excessive Demographics Access



splunk> .conf18

Key Components

Alerts

Alerts are monitored by NYP Information Security staff, who follow up on alerts. Incident Manager automates that process by triggering alerts.

- Example Alerts:
  - Higher-than-user-normal
  - Persistent High Volume
  - Accesses to Consecutive MRNs
  - Excessive hours-with-activity
  - Deviation from activity of peers (same department and title)
  - Excessive SSN Accesses
  - Access to VIP & Employee Patients
  - Access by inactive user

Technical Approach

Splunk Healthcare Privacy Monitoring and Auditing Platform

Jay Benfield,
jay@sigbay.com

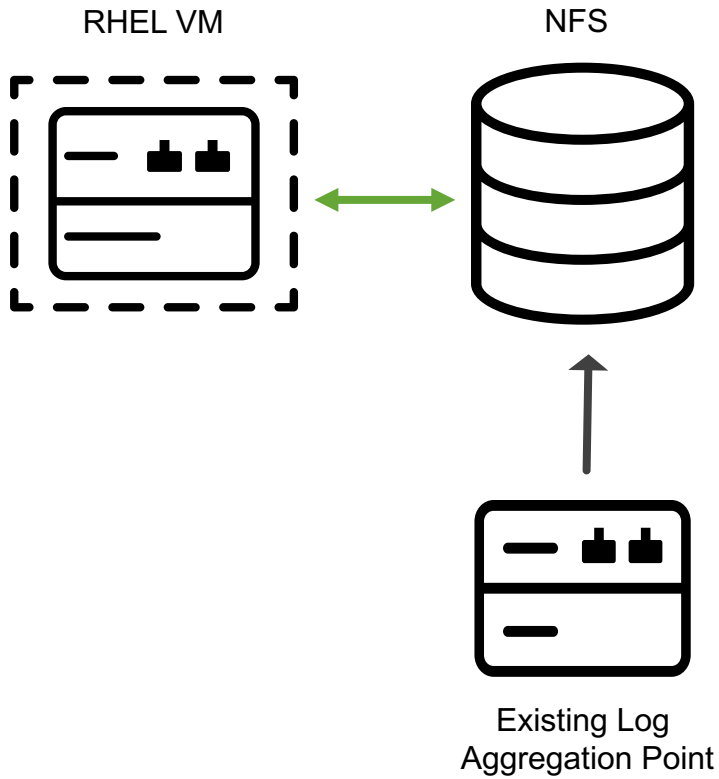
Senior Architect, SigBay, Inc.

splunk> .conf18

Development Environment

Simple = Better

- Single instance of Splunk 7
- RedHat Enterprise Linux VM
- Limited RAM, CPU, Disk
- NFS attached storage
- 3 months of audit logs
- User data from LDAP



Audit Log Data

Data discovery, analysis, and ingest

Audit logs from clinical applications

- Commercial and custom software
- Approach to normalization allows for easy integration of new data sources
- Mix of structured (csv, tsv, psv, other) and unstructured data
- Some complexities in terms of field or time extraction
- Creative regex-ing needed to clean garbage characters without pre-processing

agility	iprs
amalga	meditech
arms	mynypmobile
athena	omnicell
cerner	ormanager
eagle	pss
eclipsys	soarian
empi_matchmetrix	srm
ezvac	streamline
imagecast	teleresults

Some of the data sources leveraged in the app.

Indexing Challenge: Time

Using datetime.xml to combat an unusual time configuration

```
04/01/2018
 2:30 am       "WEST SRM EVENT HISTORY for 3/31/2018"


"PERSON_NAME"   "LOGIN_ID"      "FACILITY"      "STATION_NAME"  "EVTCD" " EVENT_TYPE_NAME"      "evnttm"        "MRN"
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 10:50   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 13:35   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 15:20   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 15:55   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 16:15   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 16:40   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 16:40   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED"  7:30   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "TRLB"  "TRANS LOAD BAD"        7:30    ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "ZRTE"  "Item routed"    7:30   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 10:24   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "TRLB"  "TRANS LOAD BAD"        10:24   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "ZRTE"  "Item routed"    10:24  ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 10:30   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 14:15   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 14:15   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 16:05   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 22:40   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 22:45   ""
"IFACE, TRANS"  "TR_IFACE"      "1"     "TR_IFACE"      "R_L"   "REPORT LOADED" 13:27   ""
```

Indexing Challenge: Garbage

Taking out the trash with SEDCMD

```
201803310000033<87>^Z201803310000033<87>^ZHANXT31
20180331000033þ¾20180331000033þ¾OAREGP9    R0OHQ
20180331000017CøÅ20180331000017CøÅOAREGP9    R0PS7
20180331000031tíó20180331000031tíóHMACCH1    R0MFY
20180331000031u^N^Z20180331000031u^N^ZHMACCH1    R0M
20180331000040^Z^Z?20180331000040^Z^Z?LOGON    R0F
```

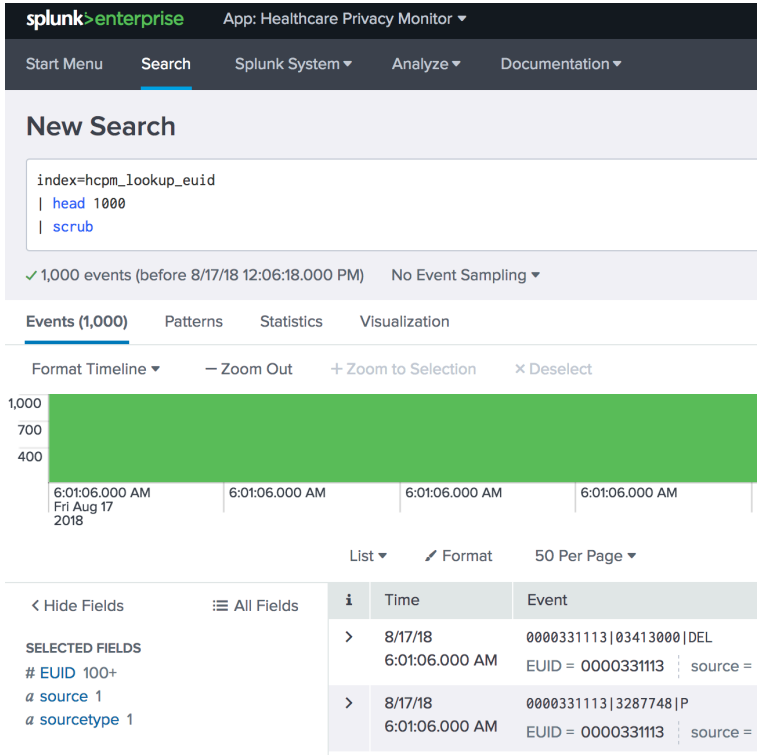
```
[eagle]
LINE_BREAKER = [RCD][048][[:alnum:]]{3}[^\n]*(\n)
SEDCMD-binary = s/(?ms)(\d+).*?([ACHILMOSU][CSABGMWRPOKQ]\w+)\s+([RCD][048][[:alnum:]]{3})/\1 \2 \3/g
SHOULD_LINEMERGE = true
DATETIME_CONFIG =
NO_BINARY_CHECK = true
category = Custom
disabled = false
pulldown_type = true
EXTRACT-event_id = (?<event_id>[ACHILMOSU][CSABGMWRPOKQ]\w+)
EXTRACT-user_id_local = [RCD][048](?<user_id_local>\w{3})
EXTRACT-patient_id_local = 2Pat:\s(?<patient_id_local>\w+)
EXTRACT-action = ^\d+\s\w+\s(?<action>[RCD])
EXTRACT-action_outcome = ^\d+\s\w+\s[RCD](?<action_outcome>[048])
EXTRACT-patient_name = \s+(?<patient_name>\w+\-?\w+?,\s\w+\s?\-?\w+)$
LOOKUP-eagle_user_id = eagle_user_id user_id_local AS user_id_local OUTPUTNEW user_id AS user_id
```

Development Standards & Conventions

Guiding principles and conventions to direct solution development

Early definition of standards vital

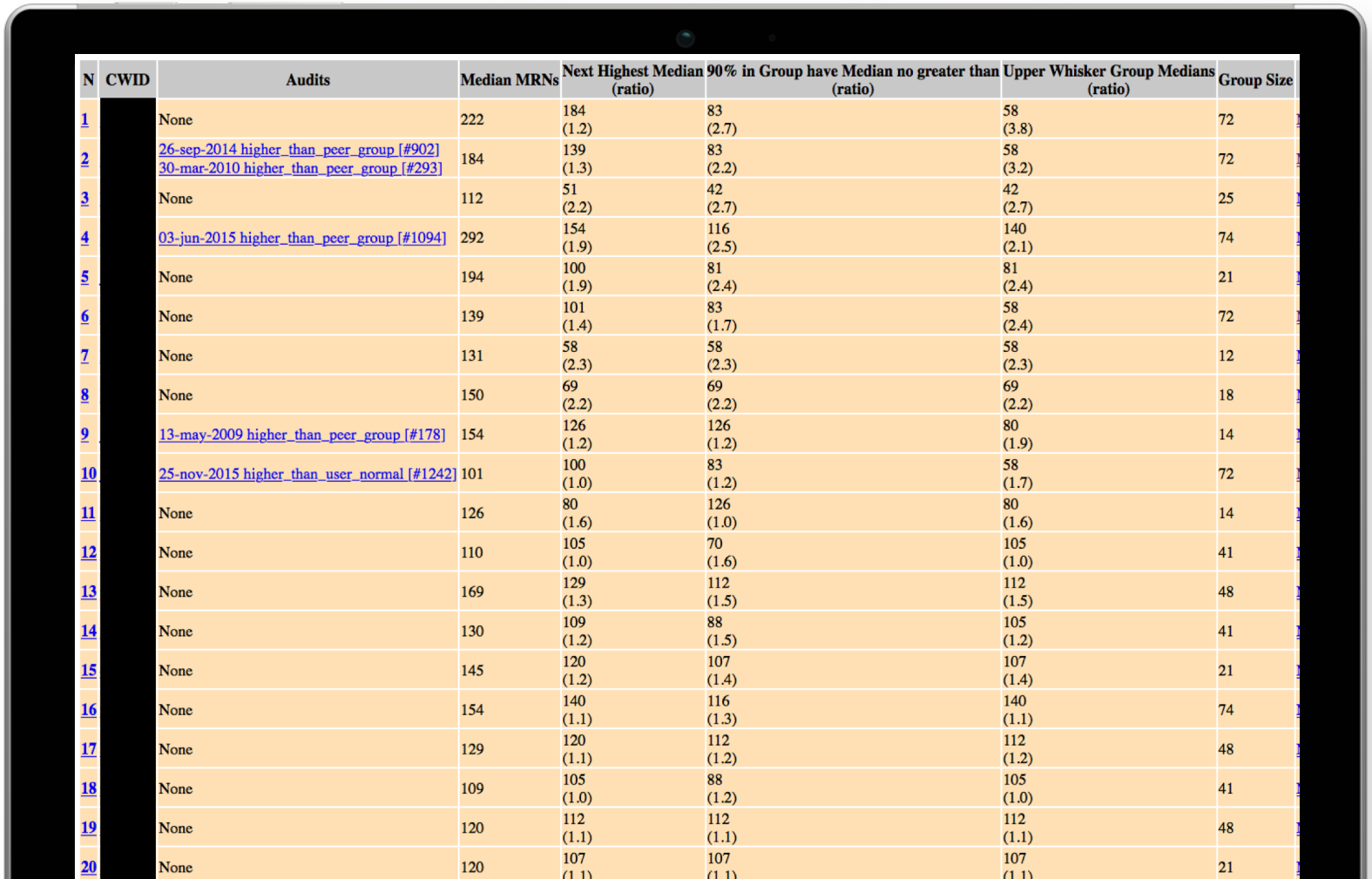
- Established consistency in naming conventions across indexes and source types
- Defined index layout to enable access control needs by hospital campus/location
- Guidelines and methodology for field extractions and other index/search-time configuration
- Approach to lookups, lookup generation, summary population, macros, etc.



Indexed lookup data for greater flexibility

Legacy System

Pre-Splunk solution for privacy monitoring



N	CWID	Audits	Median MRNs	Next Highest Median (ratio)	90% in Group have Median no greater than (ratio)	Upper Whisker Group Medians (ratio)	Group Size
1		None	222	184 (1.2)	83 (2.7)	58 (3.8)	72
2		26-sep-2014 higher_than_peer_group [#902] 30-mar-2010 higher_than_peer_group [#293]	184	139 (1.3)	83 (2.2)	58 (3.2)	72
3		None	112	51 (2.2)	42 (2.7)	42 (2.7)	25
4		03-jun-2015 higher_than_peer_group [#1094]	292	154 (1.9)	116 (2.5)	140 (2.1)	74
5		None	194	100 (1.9)	81 (2.4)	81 (2.4)	21
6		None	139	101 (1.4)	83 (1.7)	58 (2.4)	72
7		None	131	58 (2.3)	58 (2.3)	58 (2.3)	12
8		None	150	69 (2.2)	69 (2.2)	69 (2.2)	18
9		13-may-2009 higher_than_peer_group [#178]	154	126 (1.2)	126 (1.2)	80 (1.9)	14
10		25-nov-2015 higher_than_user_normal [#1242]	101	100 (1.0)	83 (1.2)	58 (1.7)	72
11		None	126	80 (1.6)	126 (1.0)	80 (1.6)	14
12		None	110	105 (1.0)	70 (1.6)	105 (1.0)	41
13		None	169	129 (1.3)	112 (1.5)	112 (1.5)	48
14		None	130	109 (1.2)	88 (1.5)	105 (1.2)	41
15		None	145	120 (1.2)	107 (1.4)	107 (1.4)	21
16		None	154	140 (1.1)	116 (1.3)	140 (1.1)	74
17		None	129	120 (1.1)	112 (1.2)	112 (1.2)	48
18		None	109	105 (1.0)	88 (1.2)	105 (1.0)	41
19		None	120	112 (1.1)	112 (1.1)	112 (1.1)	48
20		None	120	107 (1.1)	107 (1.1)	107 (1.1)	21

- Application developed in-house
- Provided baseline requirements
- Legacy App SME a tremendous asset to understand data, reporting, and team operations

Primary Use Cases

Dashboards, reports, and alerts to provide needed visibility

- Accesses by User Over Time
- Accesses by Users to a Given MRN
- Higher than User Normal
- Excessive Hours with Activity
- Access Outside of Work Hours
- Excessive Demographics Access
- Employee Access to Employee Patient Records
- Failed Login Attempts
- Compare Usage Among Peers
- Access by Inactive User
- Access to VIP

Solution Building Blocks

Key information across data sources for use case development

Patient Identification

Medical Record Numbers (MRNs) of patients accessed by employees

User Information

Employee usernames, locations, job functions, and status

User Actions

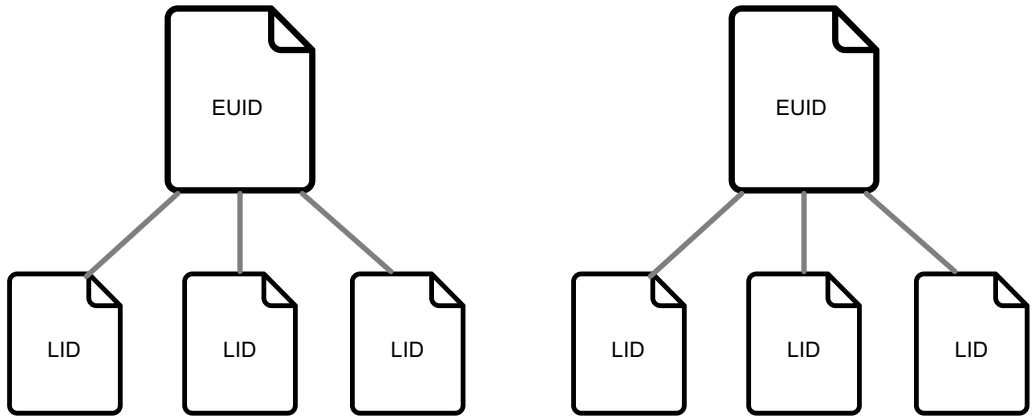
User actions performed on patient records across disparate systems

splunk> .conf18

Patient IDs

Patient MRNs from audit logs and lookups

- Audit log data may contain:
  - Enterprise Unique Identifier
  - Local Identifier(s)
  - Application-specific Patient Identifier
  - No Patient Identifier
- Lookups used to:
  - Map Enterprise Unique ID to Local IDs
  - Map app IDs to Local IDs



Patient IDs: The “Big Lookup”

Mapping of local IDs to enterprise ID

- Export from internal system with 40 million rows plus daily incremental updates
  - Initially implemented with KV Store but encountered serious performance issues and failures
  - Implemented as a standard lookup with solid performance
- Summary index *patient_id* populated with EUID if exists and uses LID if no EUID
- Dashboards, reports, and alerts leverage *patient_id*

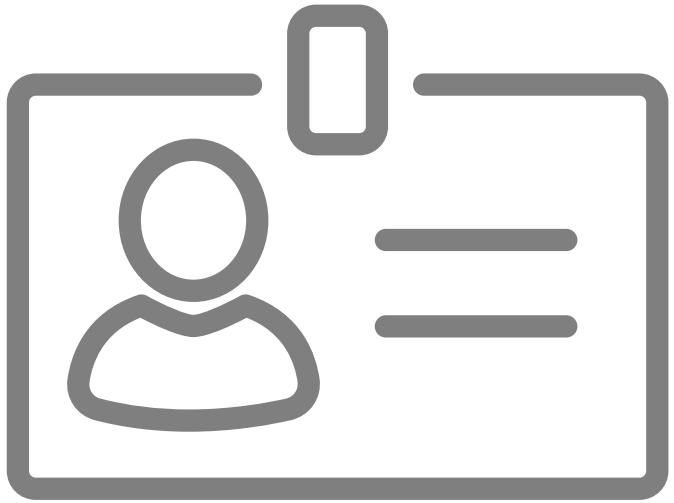
```
EUID,LID,SYSTEMCODE
1111111111,12345,LOCATION-X
1111111111,12346,LOCATION-A
1111111111,12347,LOCATION-Z
1111111111,12348,LOCATION-L
2222222222,22345,LOCATION-Y
2222222222,22346,LOCATION-X
2222222222,22347,LOCATION-D
2222222222,22348,LOCATION-A
3333333333,32345,LOCATION-L
```

1.1 GB lookup file size

User Information

Employee data from enterprise LDAP

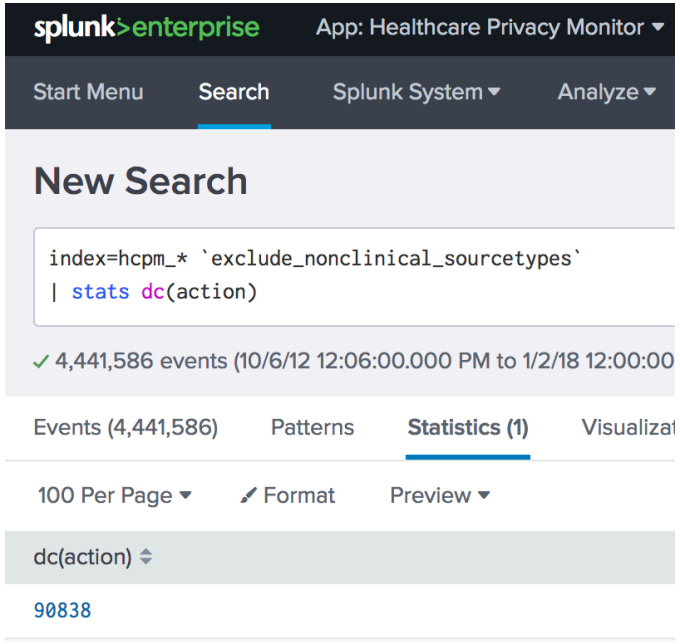
- Identity data essential across all use cases:
  - Username
  - Real name
  - Hospital affiliation, department, title
  - Account status
- Automatic lookups used to normalize application-specific usernames/IDs for specific source types
- Also used to create Personnel dashboard



User Actions

Normalizing user activity to higher-level action categories

- Thousands of individual app-specific actions in audit logs mapped to higher-level categories (e.g., create, view, update, etc.)
- Normalization originally via lookup but moved to event types
- A manual process made much easier with Splunk search

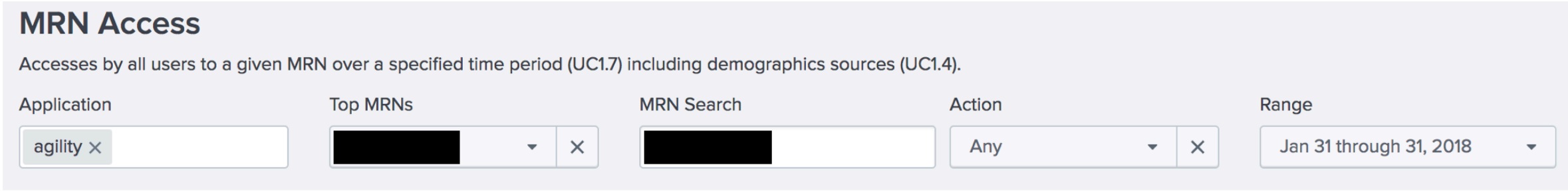


90,000+ unique actions in 24h of data

splunk> .conf18

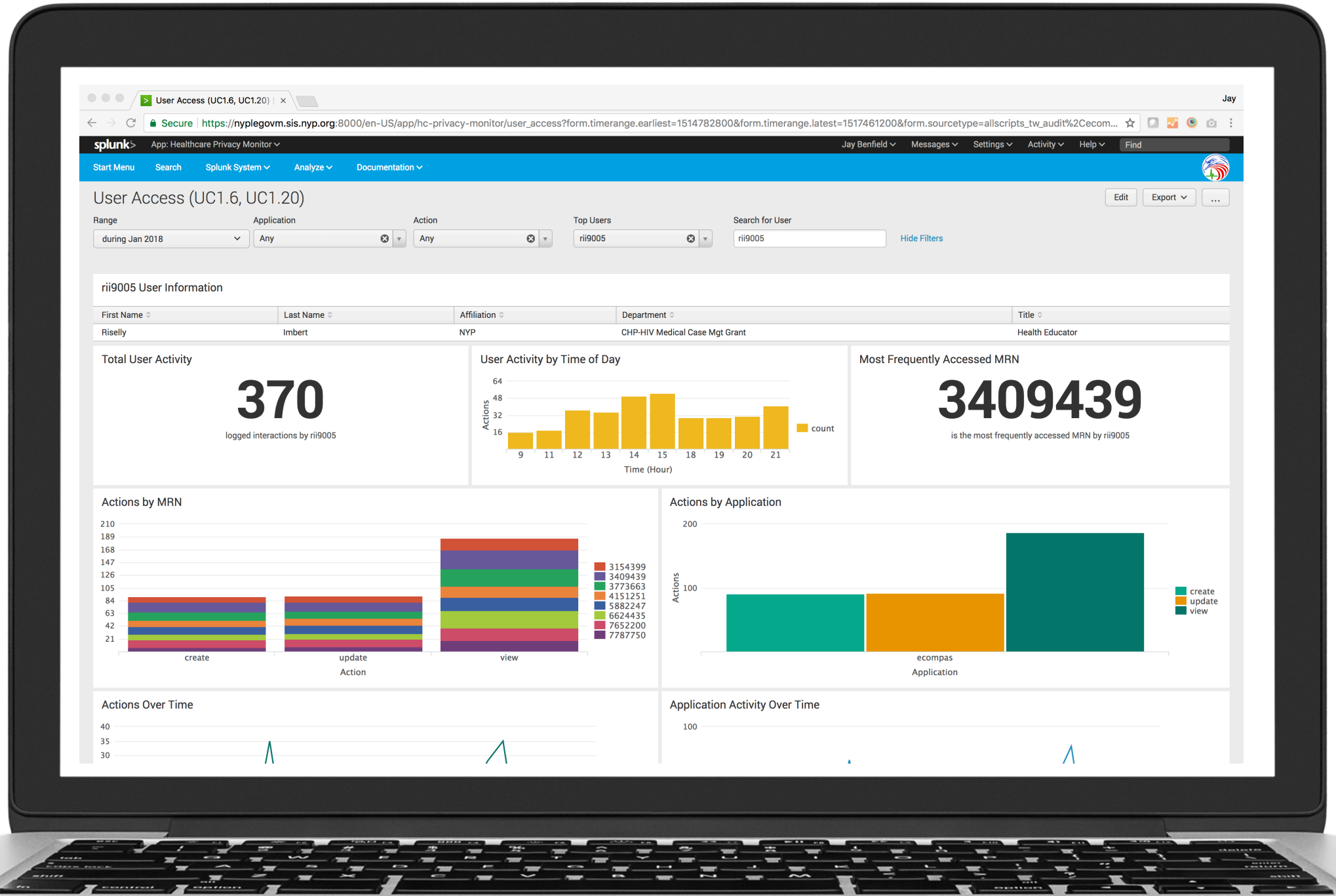
Dashboard/UI Strategy

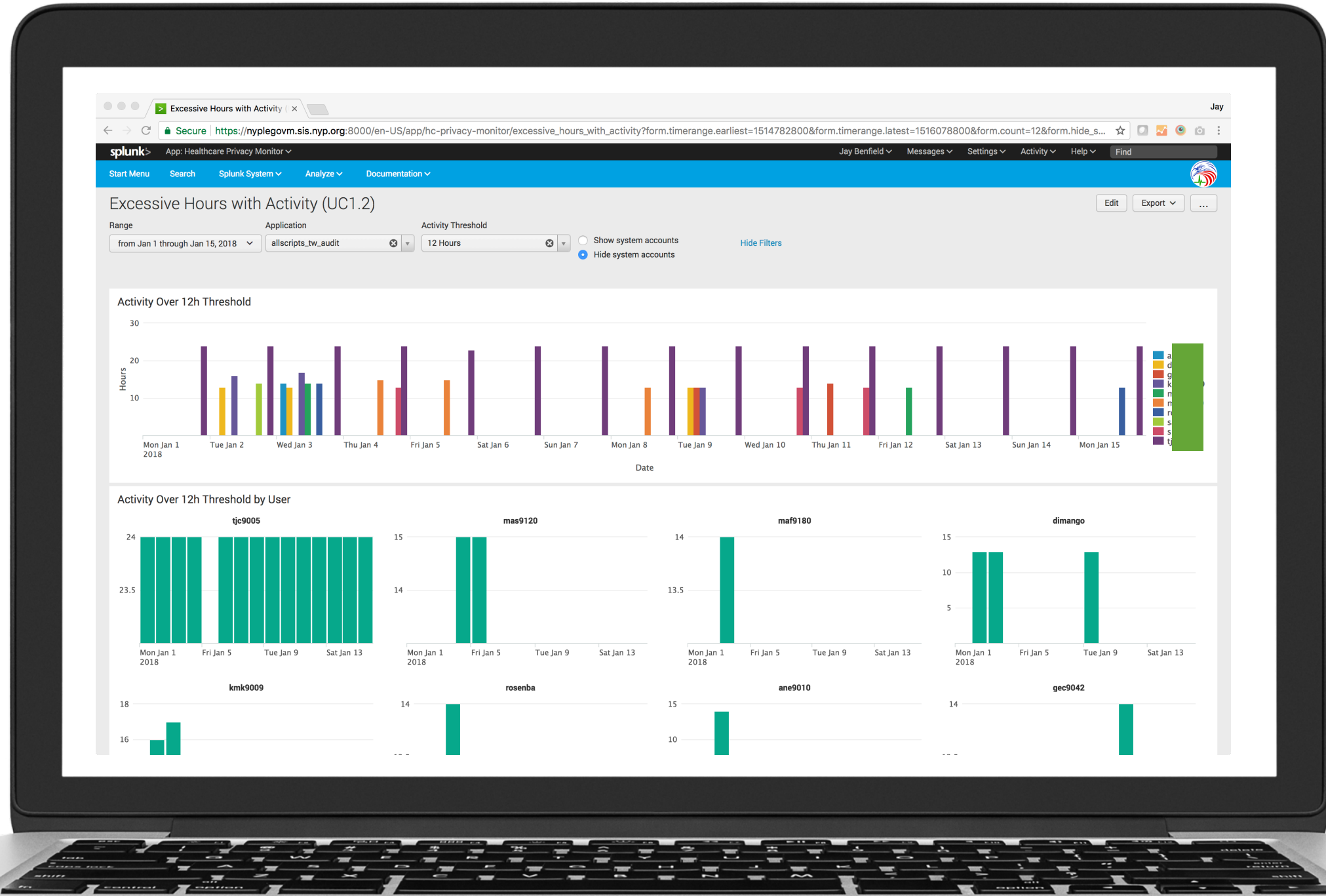
Approach to instrumentation

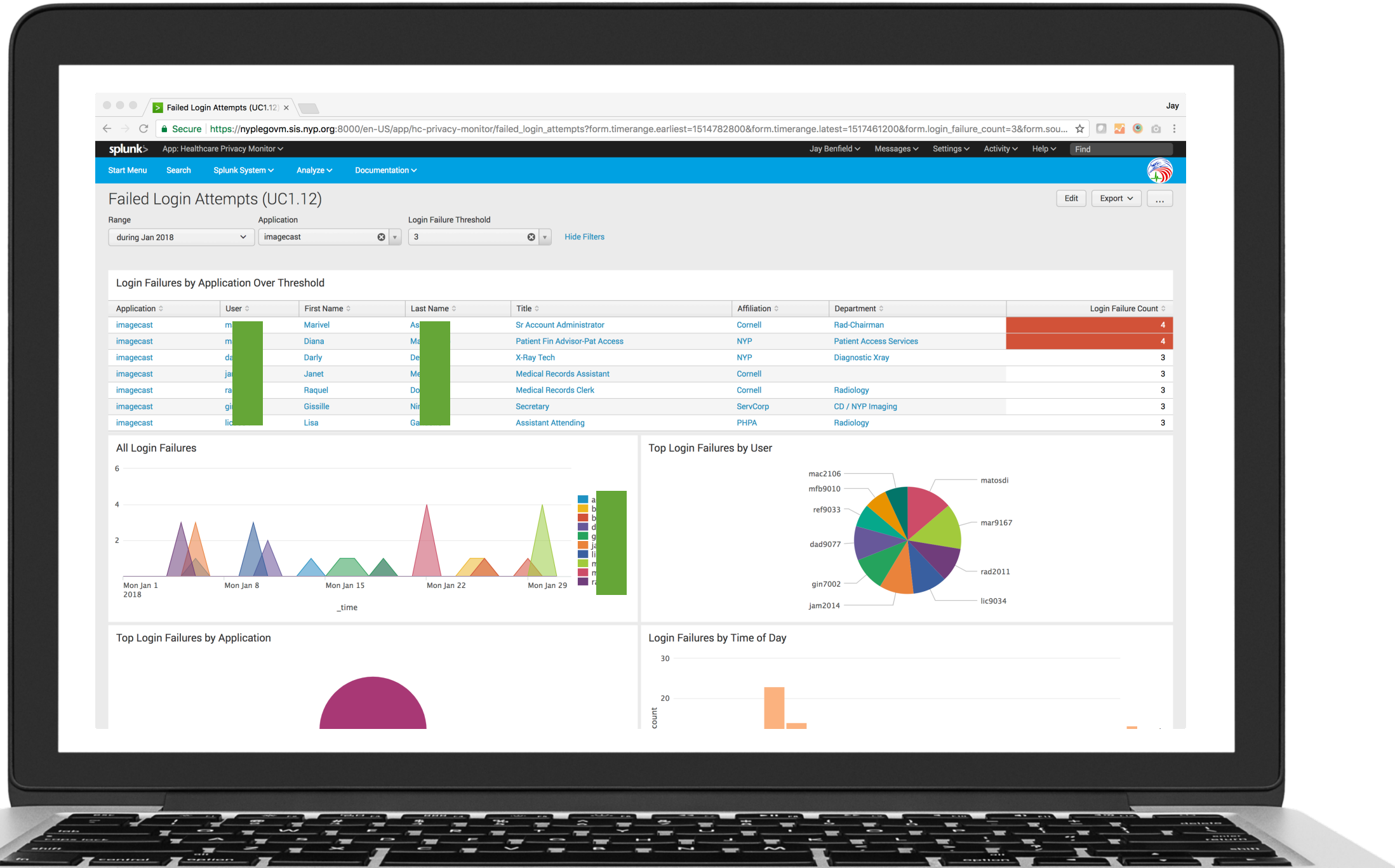


- Consistent UI controls for filtering by application, action, and user/MRN.
- Data presented from various dimensions
  - **By user:** Patient records viewed by a user over time by application and action.
  - **By patient:** User activity on a specific patient record over time by application and action.
- Dashboards driven by summary index and accelerated data model
- Extensive use of drilldowns including to print-friendly reports
- Utility dashboards to “auto-document” data source status

splunk> .conf18







Key Takeaways

Lessons learned or reinforced

- Define conventions up front to save time and re-work
- I <3 Lookups
  - | lookup vs. automatic lookups
  - Indexed lookup data + output lookup = ☺
- KV Store is great but it has its limitations
- Indexed extractions directly to normalized field names
- Splunk is just as awesome with clinical data as it is with IT stuff

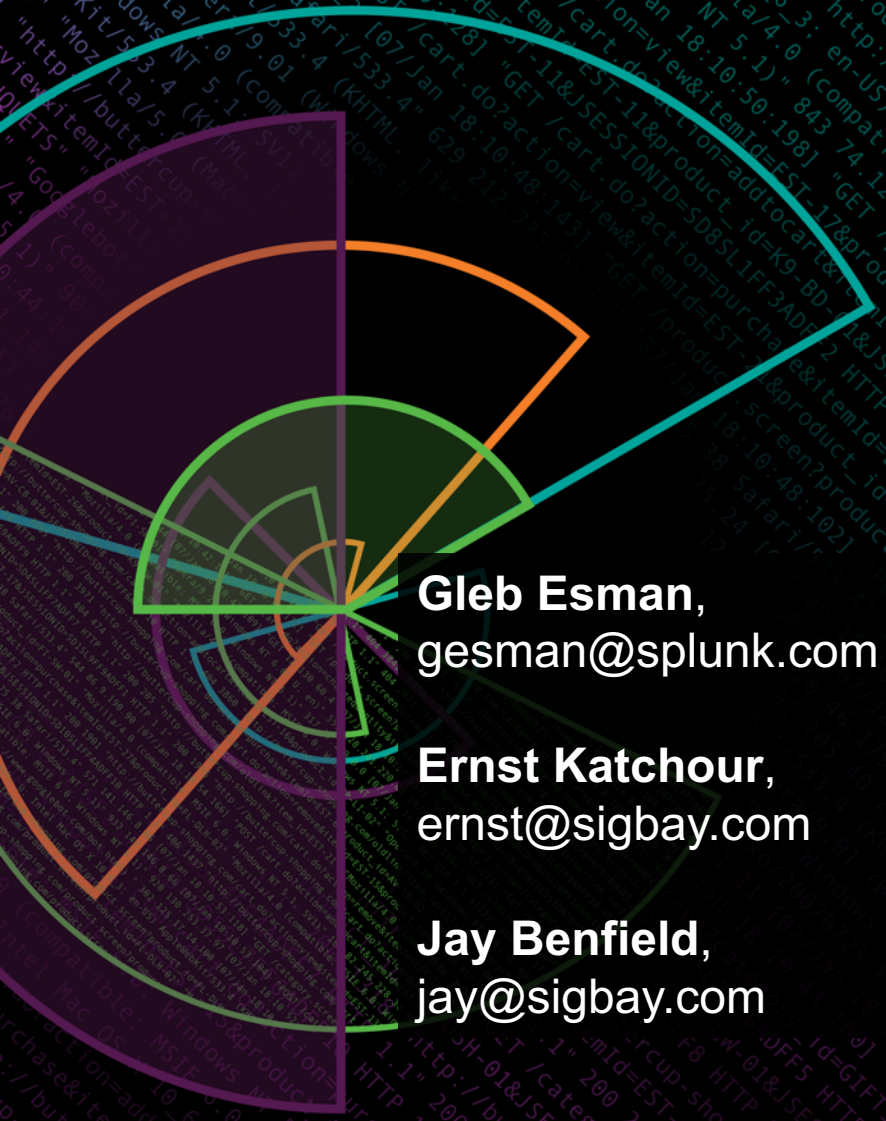
Luke Murphey's Lookup File Editor is essential.

Conclusion

- Splunk provides the scale needed for healthcare privacy and auditing using both on-premises/cloud/hybrid deployment
- Minor dependency on data formats and data delivery mechanisms
- You have complete control over all aspects of the platform - from data ingestion to alerting, investigations and case closed.
- Use to solve large number of healthcare privacy and auditing use cases.
- Predictable effort to manage, customize and enhance platform.

Thank You

Don't forget to rate this session in the .conf18 mobile app



Q&A

Gleb Esman, gesman@splunk.com
Sr. Project Manager, Fraud Analytics and Research, Splunk.

Ernst Katchour, ernst@sigbay.com
CEO and Founder, SigBay, Inc., Splunk Technology Alliance Partner

Jay Benfield, jay@sigbay.com
Senior Architect, SigBay, Inc.