



.conf20
splunk>

Sr. Security Strategist | Splunk

Staff Security Strategist | 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>, Data-to-Everything, D2E and Turn Data Into Doing 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. © 2020 Splunk Inc. All rights reserved

Andrew Morris

Sr. Security Strategist | Splunk



Gleb Esman

Staff Security Strategist | Splunk



Agenda

1. What's been covered before

.Conf19 BOTS the Missing Link (not required)

2. What was missing

Things I wanted to do

3. Techniques

Data manipulation

4. Visualization options

New stuff

5. Demo

Even more really new stuff

Investigations Are Fun!

Fraud – Security - IT - and more



Last Year..... Link Analysis App for Splunk

.Conf19 – SEC1781 BOTS the Missing Link

Plenty of Layouts!

The image displays six distinct network graph visualizations, each representing a different layout option available in the Link Analysis App. The graphs are arranged in a collage-like fashion with torn edges. The top-left graph is a horizontal grid layout with nodes arranged in rows and columns, connected by lines. The top-right graph is a circular radial layout with nodes arranged in a circle and lines radiating from a central point. The middle-left graph is a hierarchical tree layout with nodes arranged in a tree structure. The middle-right graph is a circular arc layout with nodes arranged in a circle and lines connecting them along the arc. The bottom-left graph is a cluster layout with nodes grouped into clusters and connected by lines. The bottom-right graph is a horizontal flow layout with nodes arranged in a horizontal line and lines connecting them in a flow pattern. The text 'Plenty of Layouts!' is written in a large, bold, black font at the top left of the collage. The copyright notice '© 2019 SPLUNK INC.' is visible in the top right corner of the collage. The Splunk logo and '.conf19' badge are in the bottom right corner of the collage.

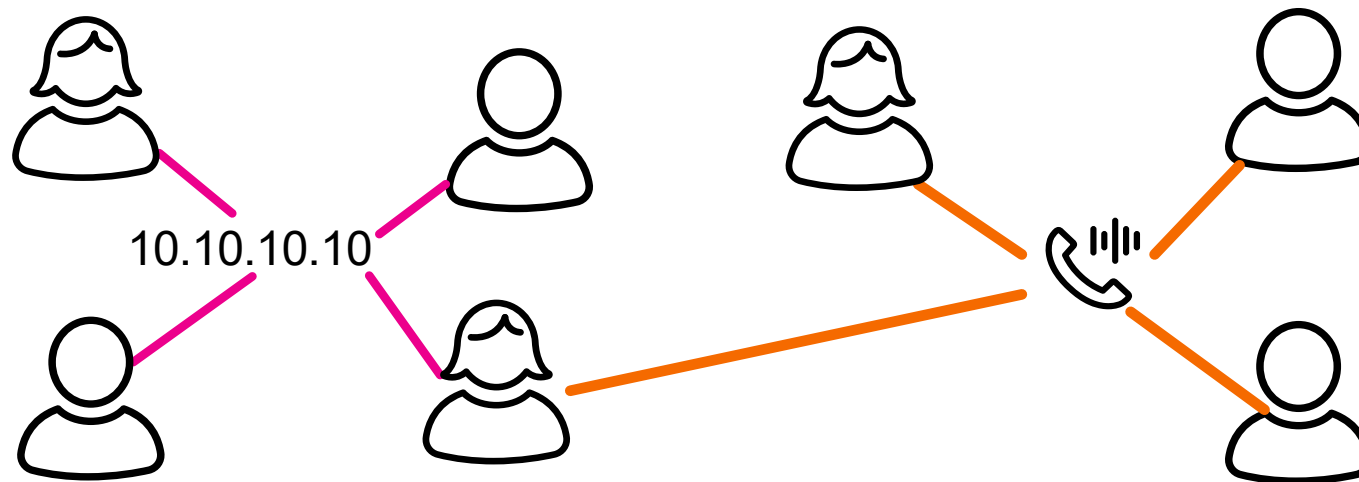
© 2019 SPLUNK INC.

splunk> .conf19

But Things Are Missing...

I Want More!

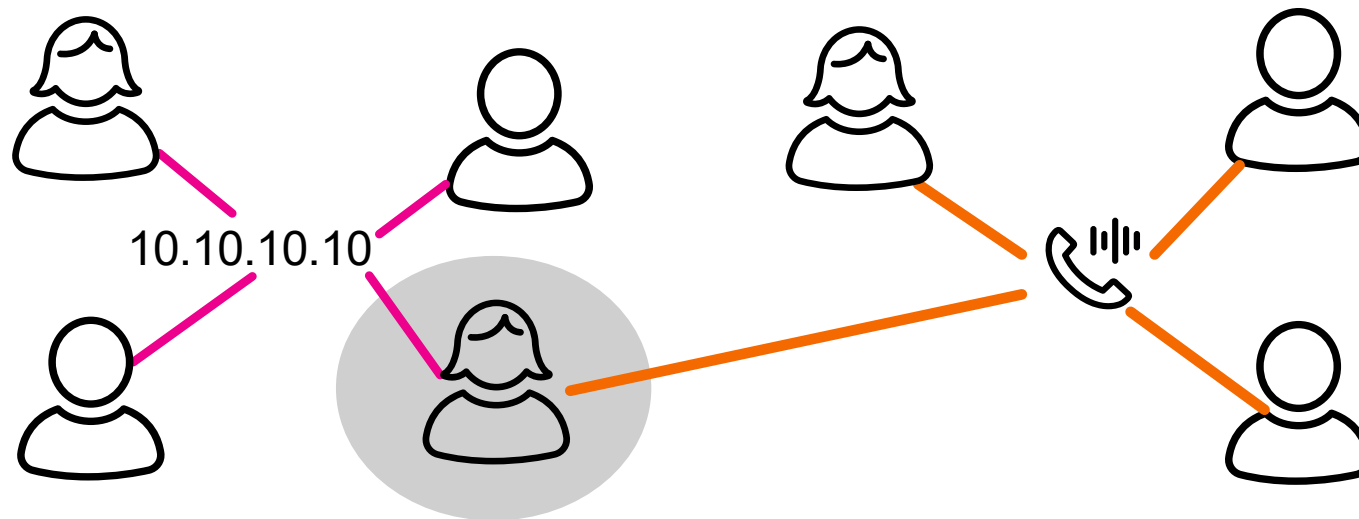
Field Limitations – multi-relationship



But Things Are Missing...

I Want More!

Field Limitations – multi-relationship



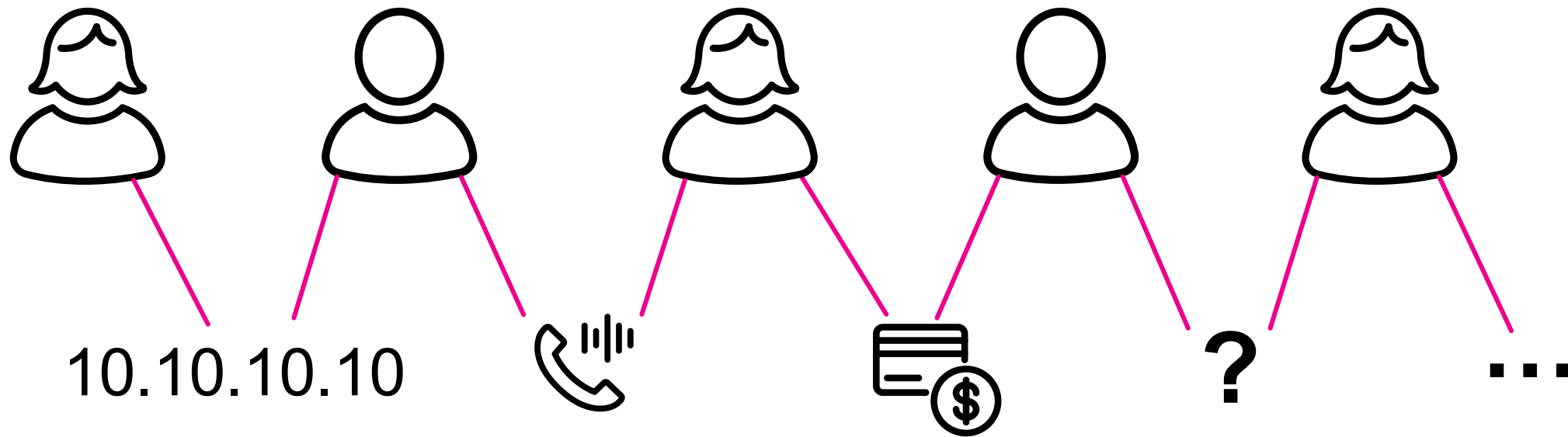
Circles and lines: Boring.....



But Things Are Missing...

I Want More!

Extensibility and Discovery



Let's Address These Issues!




Sample Data

15000+ events

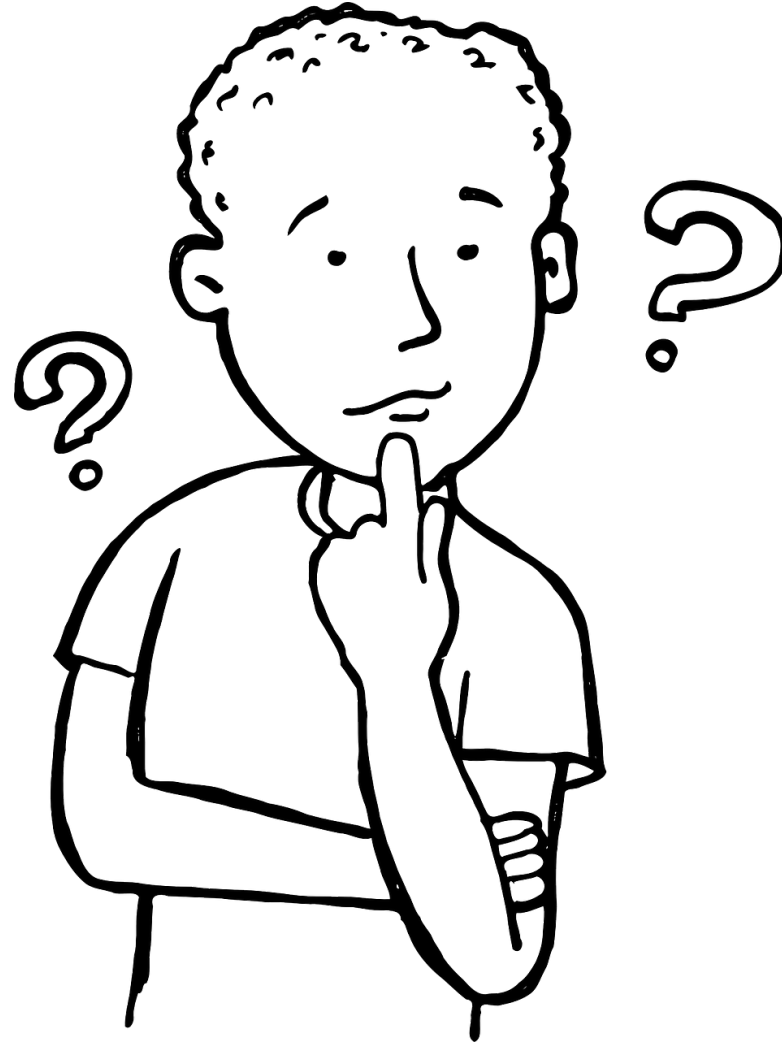
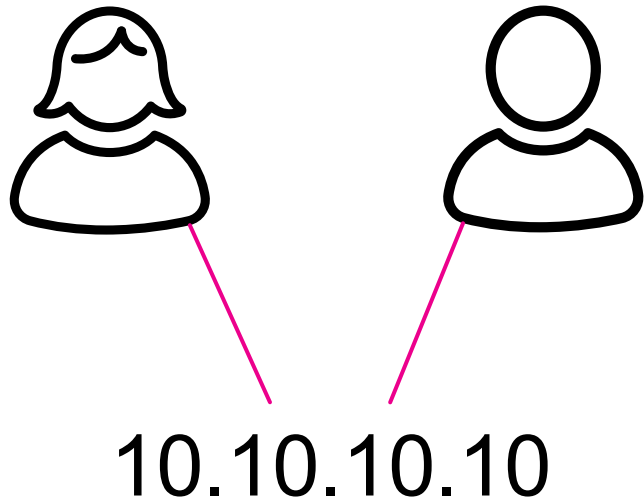
fname	middle	lname	gender	email	mmn	dob	acct_age	salary	occupation	ssn	phone	place	county	city	state	zip	username	password	status	ip_address	ua
Tawanna	H	Borges	F	tawanna.bor	Downer	8/19/57	6.05	50996	Archivist	303-50-2191	303-508-249	Denver	Denver	Denver	CO	80208	ahhuff	2bG>i@_ ^e*	Approve	38.101.37.16	Mozilla/5.0 (
Zena	P	Soukup	F	zena.soukup	Marchetti	8/30/57	4.81	111633	Fine Artist	303-50-2192	262-429-803	Summit Lake	Langlade	Summit Lake	WI	54485	ejhealy	27@O>92[of	Approve	75.43.182.55	Mozilla/5.0 (
Lynwood	E	Vue	M	lynwood.vue	Helmick	8/30/57	12.66	54141	First-Line Su	303-50-2193	215-762-922	Cresson	Cambria	Cresson	PA	16630	wdsuttles	2Grm]!fSr7V	Approve	7.59.167.185	Mozilla/5.0 (
Sharan	O	Endicott	F	sharan.endic	Marciano	9/1/57	33.71	65508	Surgeon	303-50-2194	316-328-600	Havana	Montgomery	Havana	KS	67347	rjdunford	27Q+SqDdL-	Approve	7.92.216.201	Mozilla/5.0 (
Gladys	G	Fell	F	gladys.fell@	Rives	9/12/57	22.15	183725	Internist	303-50-2195	219-938-866	Newberry	Greene	Newberry	IN	47449	mtbarringer	2OoRUzZHY	Approve	21.191.177.1	Mozilla/5.0 (
Venetta	C	Tomlin	F	venetta.toml	Strand	9/20/57	20.94	83962	Shipping	303-50-2196	210-438-324	Dallas	Dallas	Dallas	TX	75312	lpmorrell	2cJm10n^6*	Approve	143.144.91.1	Mozilla/5.0 (
Ernesto	A	Burnside	M	ernesto.burn	Rueda	9/23/57	11.92	52322	Dispatcher	303-50-2197	212-913-480	Plattsburgh	Clinton	Plattsburgh	NY	12901	wygroves	25-9mNcN6	Approve	11.181.196.2	Mozilla/5.0 (
Bryan	M	Batschler	M	bryan.batsch	Steele	11/1/57	34.05	101074	Director	303-50-2198	330-550-117	Atlanta	Fulton	Atlanta	GA	31131	fraserstan	2VY6ZdlinB	Approve	67.56.184.10	Mozilla/5.0 (

Important fields

- Username (unique) 
- Phone
- Password
- IP_address
- Destination account

What Is a Link?

Can it be this simple?



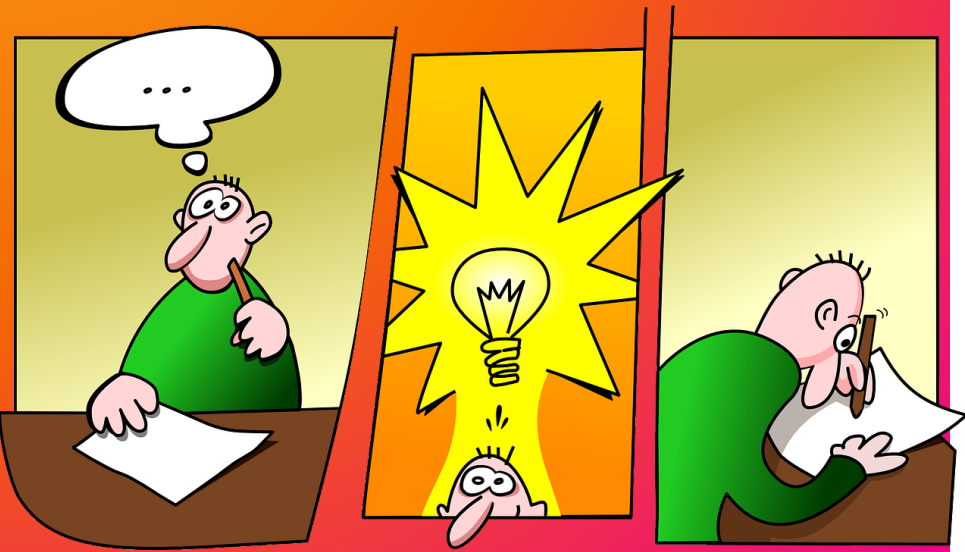
Duplicates

Splunk is great with duplicates

i	Time	Event
>	6/16/18 9:00:00.000 PM	6/17/2018 0:00,231635,Agustin,F,Bookman,M,cagustin.bookman@gmail.com,Sale,6/16/1981,8.75,81179,Mechanical Door Repairer,432-99-420 0,703-121-2411,Eden,Jerome,Eden,ID,83325,Xlslenoir,zf@#RGL5J?{\$l/*c,Approve,67.196.151.122,Mozilla/5.0 (Windows NT 6.3; WOW64; rv:4 1.0) Gecko/20100101 Firefox/41.0 ip_address = 67.196.151.122 password = zf@#RGL5J?{\$l/*c phone = 703-121-2411 username = Xlslenoir
>	6/16/18 9:00:00.000 PM	6/17/2018 0:00,211985,Jason,F,Stickler,M,cjason.stickler@msn.com,Camara,1/2/1971,12.45,156969,Insulation Worker,397-33-0137,703-121 -2226,Comins,Oscoda,Comins,MI,48619,Xndskinner,zQ#eJB<mDdwc,Approve,67.196.151.122,"Mozilla/5.0 (Linux; Android 6.0.1; SM-G920V Bui ld/MMB29K) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.98 Mobile Safari/537.37" ip_address = 67.196.151.122 password = zQ#eJB<mDdwc phone = 703-121-2226 username = Xndskinner
>	6/15/18 9:00:00.000 PM	6/16/2018 0:00,231206,Cleora,R,Trapp,F,ccleora.trapp@gmail.com,Albino,11/20/1995,0.65,46990,Mathematical Science Teacher,177-81-808 7,703-121-1961,Tillamook,Tillamook,Tillamook,OR,97141,Xcawinger,zX\FK7{M+jl{Xc,Decline,67.196.151.122,"Mozilla/5.0 (Windows NT 6.2; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/39.0.2171.95 Safari/537.36" ip_address = 67.196.151.122 password = zX\FK7{M+jl{Xc phone = 703-121-1961 username = Xcawinger

How do we detect, count and track duplicates?

Eureka!



Eventstats

```
1 index="newbigdata"
2 | eventstats count as dupphone by phone
3 | eventstats count as dupip by ip_address
4 | eventstats count as duppass by password
5 | eval total = dupphone+dupip+duppass
6 | where total > 3
7
8 | table username, ip_address, password, phone, total, dupip, duppass, dupphone
9 | sort password, phone
```

Generates summary statistics from fields in your events and saves those statistics in a new field.

SPL Results

| eventstats count as dupip by ip_address

>	6/17/18	6/18/2018 0:00,540982,Corey,O,Crites,M,acorey.crites@bellsouth.net,Michaelson,11/11/1974,1.65,103571,Automotive and Watercraft Service Attendant,542-81-0961,72+480-266-0422,Phoenix,Maricopa,Phoenix,AZ,85077,Yekengleman,l->A>A.uObv,Approve,149.234.241.104,"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.79 Safari/537.36 Edge/14.14393"
	9:00:00.000 PM	dupip = 5 duppass = 1 dupphone = 1 ip_address = 149.234.241.104 password = l->A>A.uObv phone = 72+480-266-0422 username = Yekengleman
>	6/17/18	6/18/2018 0:00,699801,Latina,J,Kato,F,alatin.kato@earthlink.net,Cimino,12/6/1973,4.41,55361,Rotary Drill Operator,542-81-0999,72+907-267-1838,Holy Cross,Yukon-Koyukuk (CA),Holy Cross,AK,99602,Ysbmagallanes,hAB2>wY8zkbv,Approve,149.234.241.104,Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; Media Center PC
	9:00:00.000 PM	dupip = 5 duppass = 1 dupphone = 1 ip_address = 149.234.241.104 password = hAB2>wY8zkbv phone = 72+907-267-1838 username = Ysbmagallanes

username	ip_address	password	phone	total	dupip	duppass	dupphone
Xrrcard	149.234.241.104	zw50\$punV2c	703-121-3547	7	5	1	1
rrcard	149.234.241.104	2w50\$punV2	218-395-9292	7	5	1	1
rglilly	149.234.241.104	2m{+22AKf[236-948-8036	7	5	1	1
Yekengleman	149.234.241.104	l->A>A.uObv	72+480-266-0422	7	5	1	1
Ysbmagallanes	149.234.241.104	hAB2>wY8zkbv	72+907-267-1838	7	5	1	1

SPL Results

| eventstats count as dupip by ip_address

```

1 index="newbigdata"
2 | eventstats count as dupphone by phone
3 | eventstats count as dupip by ip_address
4 | eventstats count as duppass by password
5 | eval total = dupphone+dupip+duppass
6 | where total > 3
7
8 | table username, ip_address, password, phone, total, dupip, duppass, dupphone
9 | sort password, phone

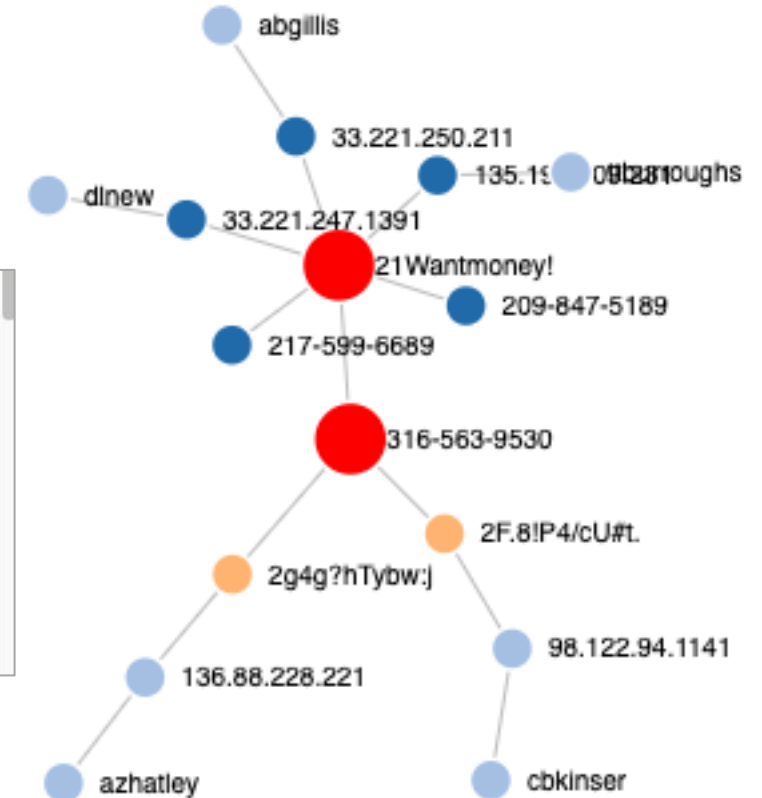
```


username	ip_address	password	phone	total	dupip	duppass	dupphone
Xrrcard	149.234.241.104	zw50\$punV2c	703-121-3547	7	5	1	1
rrcard	149.234.241.104	2w50\$punV2	218-395-9292	7	5	1	1
rglilly	149.234.241.104	2m{+22AKf[236-948-8036	7	5	1	1
Yekengleman	149.234.241.104	l->A>A.u0bv	72+480-266-0422	7	5	1	1
Ysbmagallanes	149.234.241.104	hAB2>wY8zkbv	72+907-267-1838	7	5	1	1

Force Directed With “nodes”

Force Directed App for Splunk

	username ↕	ip_address ↕	password ↕	phone ↕	total ↕	dupip ↕	duppass ↕	dupphone ↕
1	ttburroughs	135.198.209.231	21Wantmoney!	209-847-5189	5	1	3	1
2	dlnew	33.221.247.1391	21Wantmoney!	217-599-6689	5	1	3	1
3	abgillis	33.221.250.211	21Wantmoney!	316-563-9530 ←	7	1	3	3
4	rrcard	149.234.241.104	2w50\$punV2	218-395-9292	7	5	1	1
5	cbkinser	98.122.94.1141	2F.8!P4/cU#t.	316-563-9530 ←	5	1	1	3
6	azhatley	136.88.228.221	2g4g?hTybw:j	316-563-9530 ←	6	2	1	3

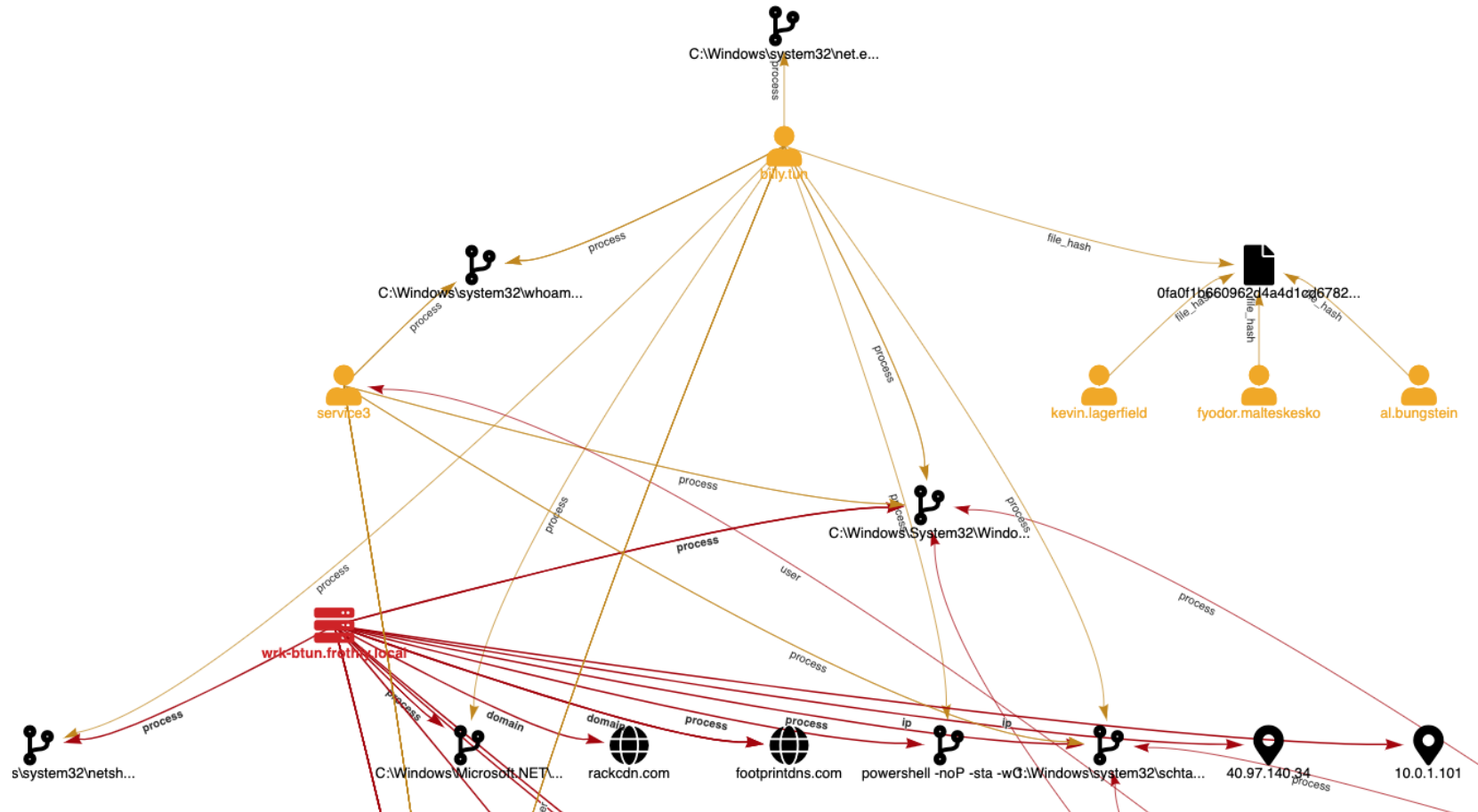


Directed Visualization  Format  Trellis



What Were Others Doing?

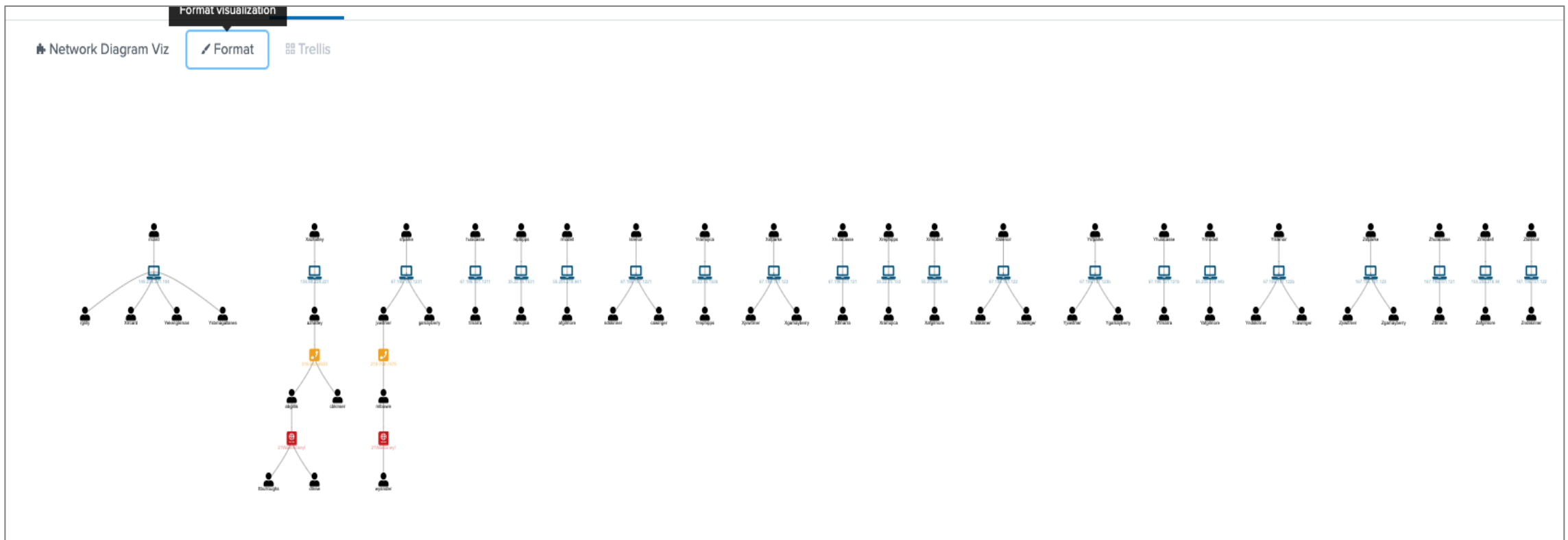
RBA! and “Network Diagram Viz” App



The diagram illustrates a network topology with numerous nodes and their interconnections. Nodes are represented by icons: a person icon for individuals and a laptop icon for devices. Each node is labeled with a name and often an IP address. The connections are shown as lines between the nodes. The network is highly interconnected, with many nodes having multiple connections. The layout is somewhat circular, with nodes distributed around the perimeter and in the center. The colors of the nodes and connections are consistent with the previous diagram, with blue for devices and black for people, and lines in blue, black, and red.

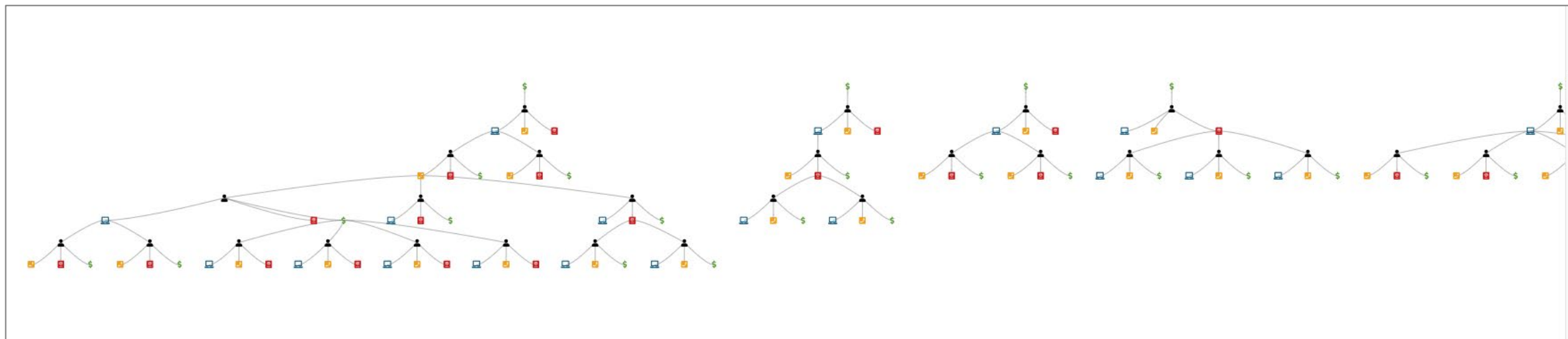
Making It Easier to Read

But still can be noisy



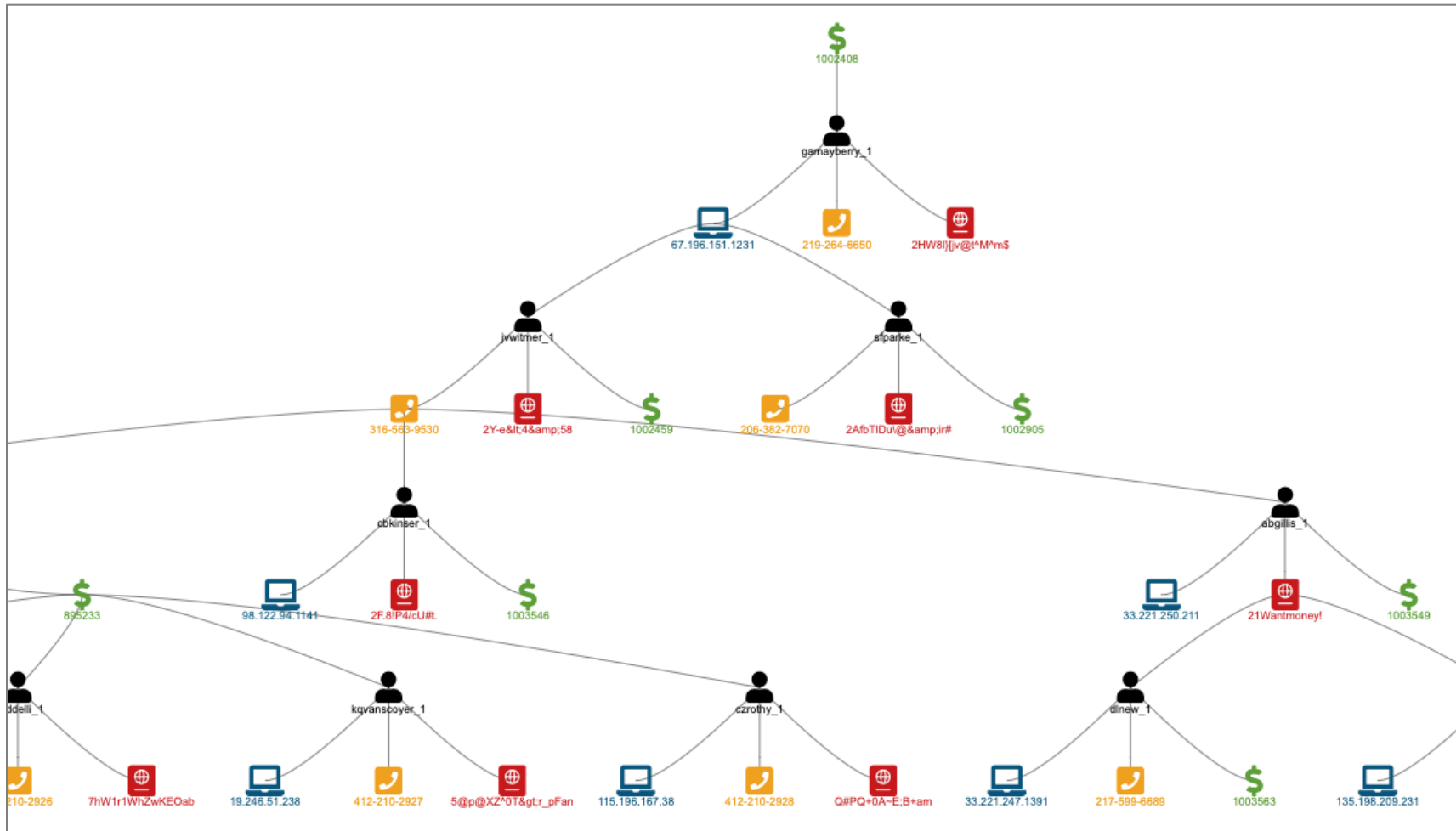
Size Might Matter....

How about 500K records?



Making Sense of It All

Zooming in



Considerations:

And other details:

Performance

- Eventstats is not instantaneous
- Additional time per command
- Good candidate for scheduled search

More SPL was used – see upcoming blog posts

- How I included single entities in larger diagrams
- How I made Network Diagram Viz show node icons for all

Network Diagram Viz app supports drill downs and more not shown

What If... I Don't Want to Do Data Reduction

Or, I don't know what I am looking for?





Gleb Esman

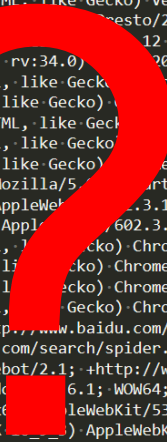
Staff Security Strategist

- IBM T.J.Watson Research
- Payment processing and DLP technologies
- Morgan Stanley
- Staff Security Strategist @ Splunk, 5 yrs
- Enjoying Pro photography
- Based in Las Vegas, NV, USA

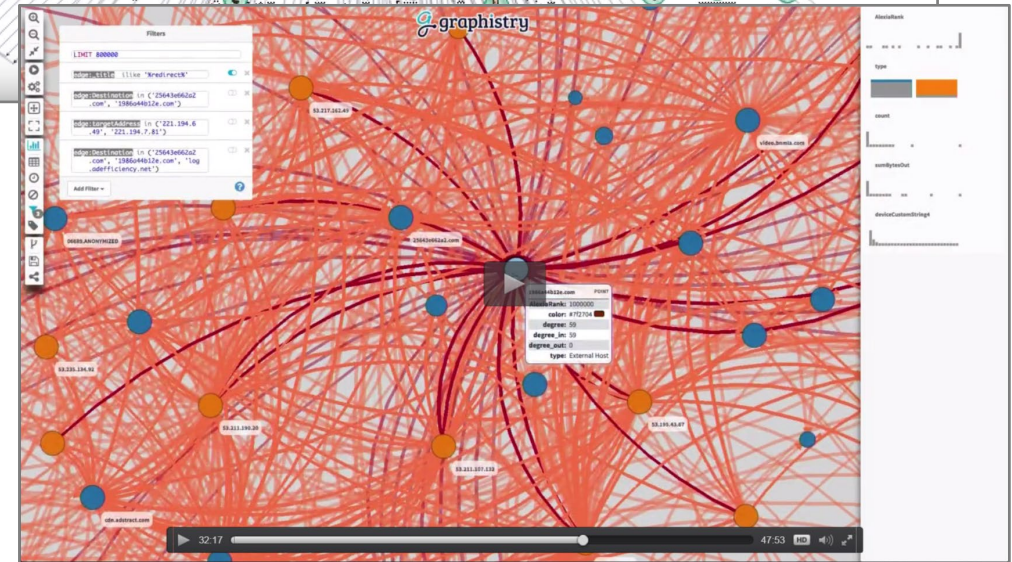
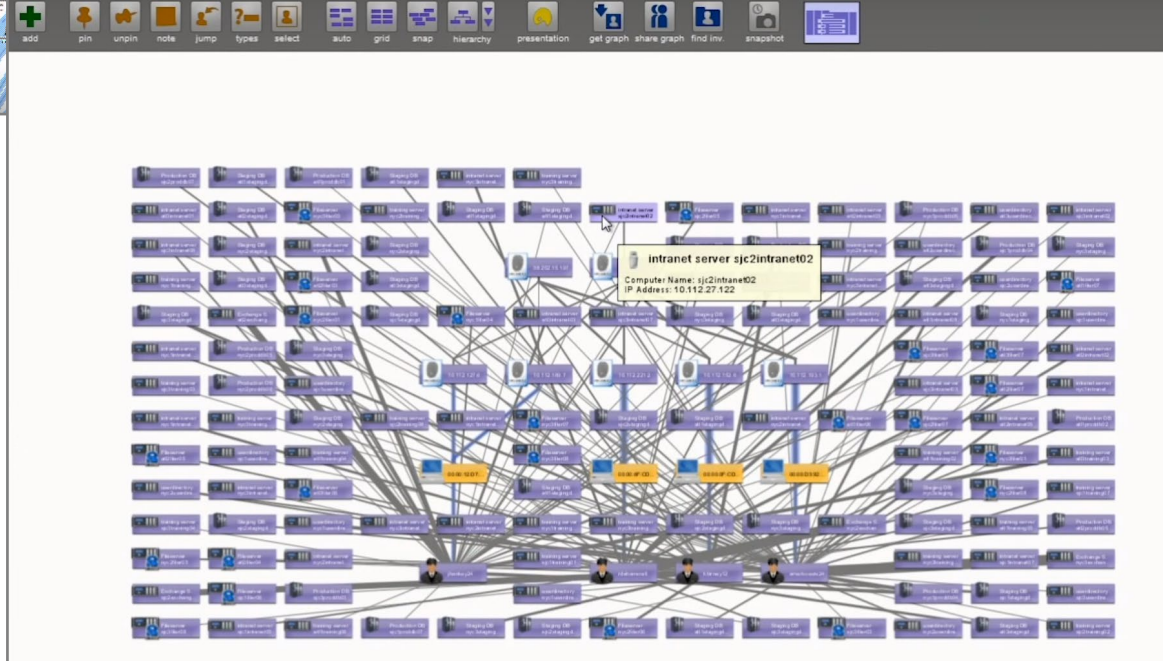
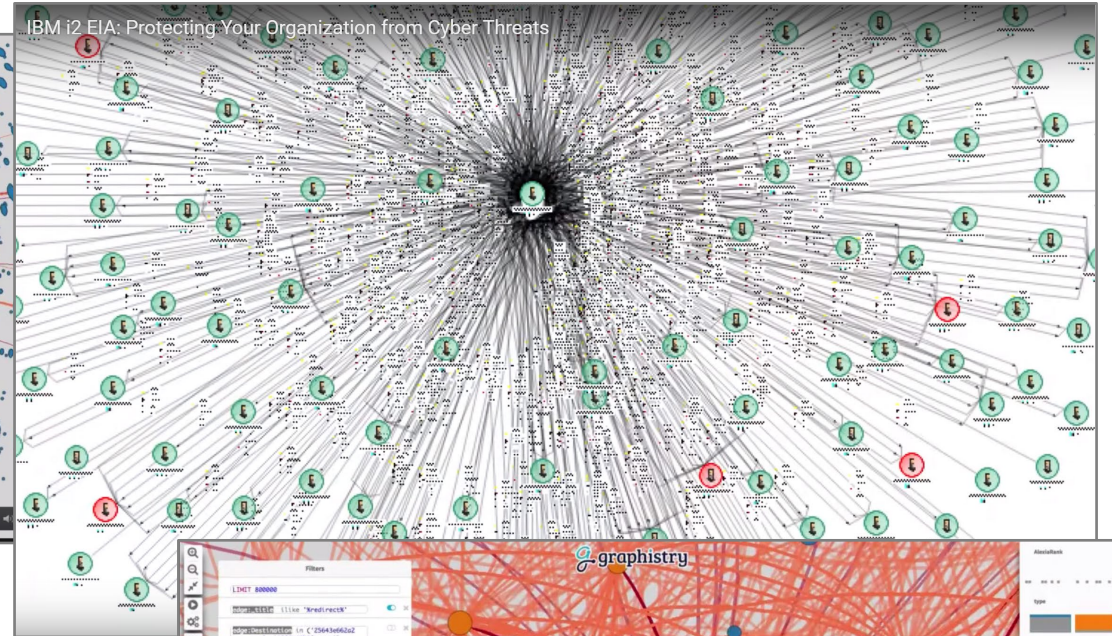
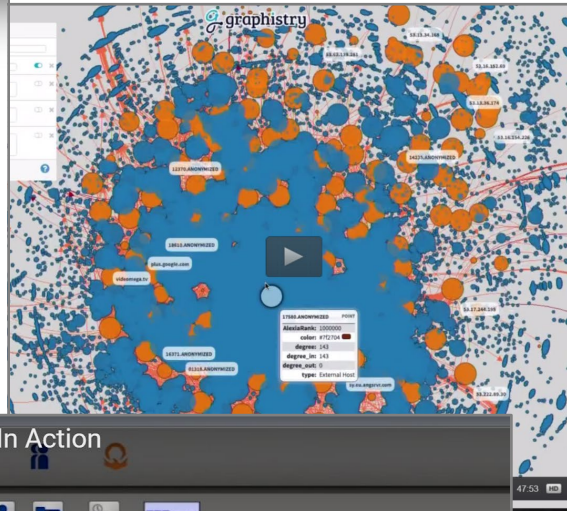
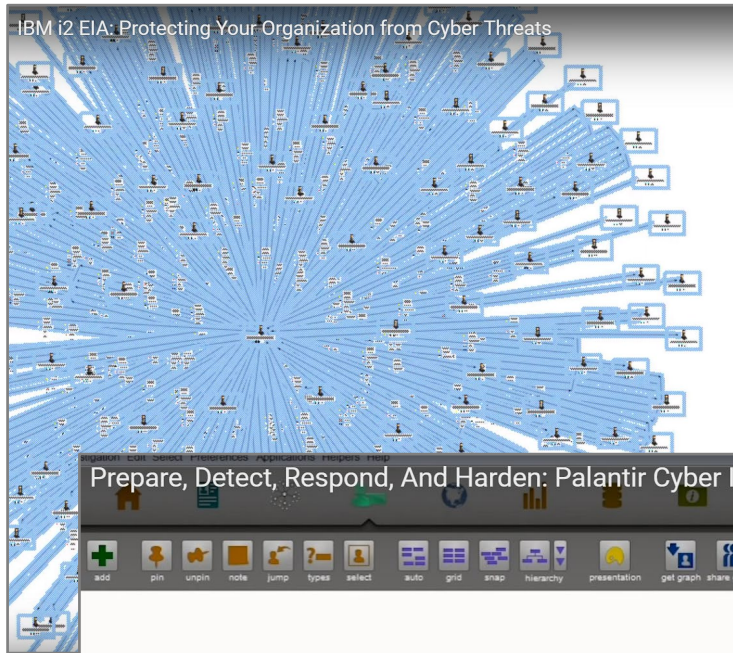
The Most Important Key Element in Fraud Analytics

Business User

The Problem: Understanding Big and Complex Data

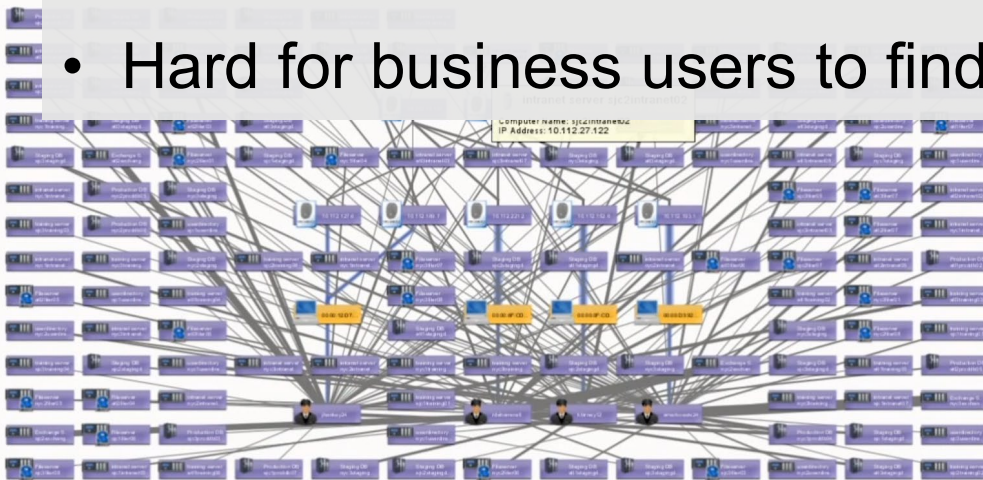
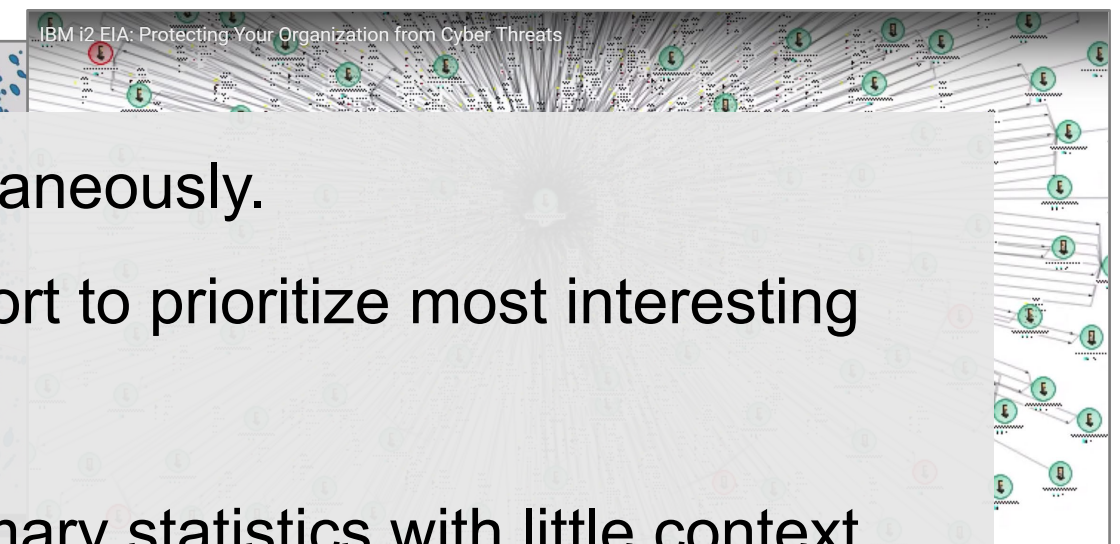
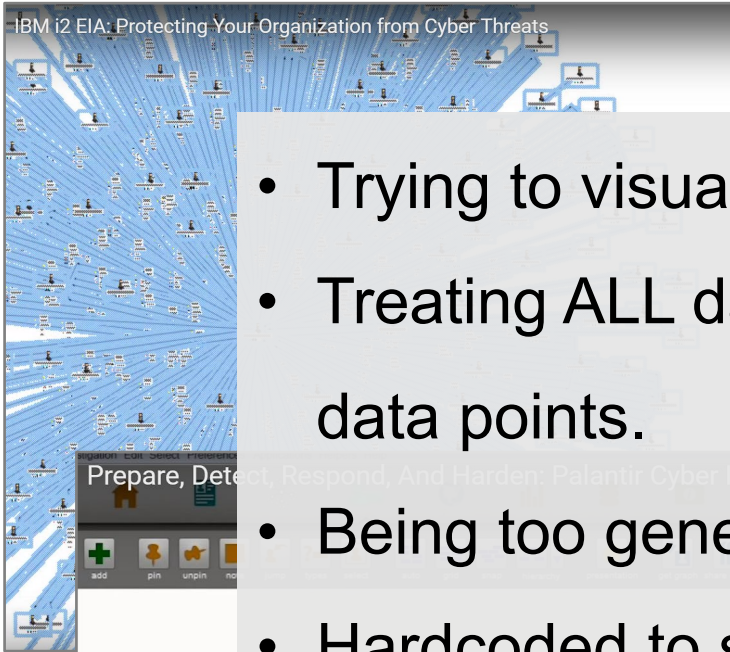


Real World Data Visualization Examples ... **FAIL**



Non-Actionable Results Syndrome. *Why?*

- Trying to visualize ALL data simultaneously.
- Treating ALL data equally – no effort to prioritize most interesting data points.
- Being too general – isolated summary statistics with little context
- Hardcoded to specific scenarios with no vision
- Hard for business users to find answers in raw data without vendor



How to Detect Fraud in Data?

Warning Signs, Anoma**LIES** and **Red** Flags:

Quantity and Frequency – too many (logins per time, transfers, connections, password changes, errors), OR too small (expected behavior is not observed)

Rarity – too unusual (never seen before names, unmatched forensics, strange devices and user agents, rare ports, unusual requests, unusual combinations)

Velocity / Relationship – 1-to-many where 1-to-1 is expected

- Multiple IP's or multiple Geo sources accessing the same user account
- Single IP hitting multiple account
- Same phone/address shared by multiple customers/users
- Logins to same account from multiple geo locations

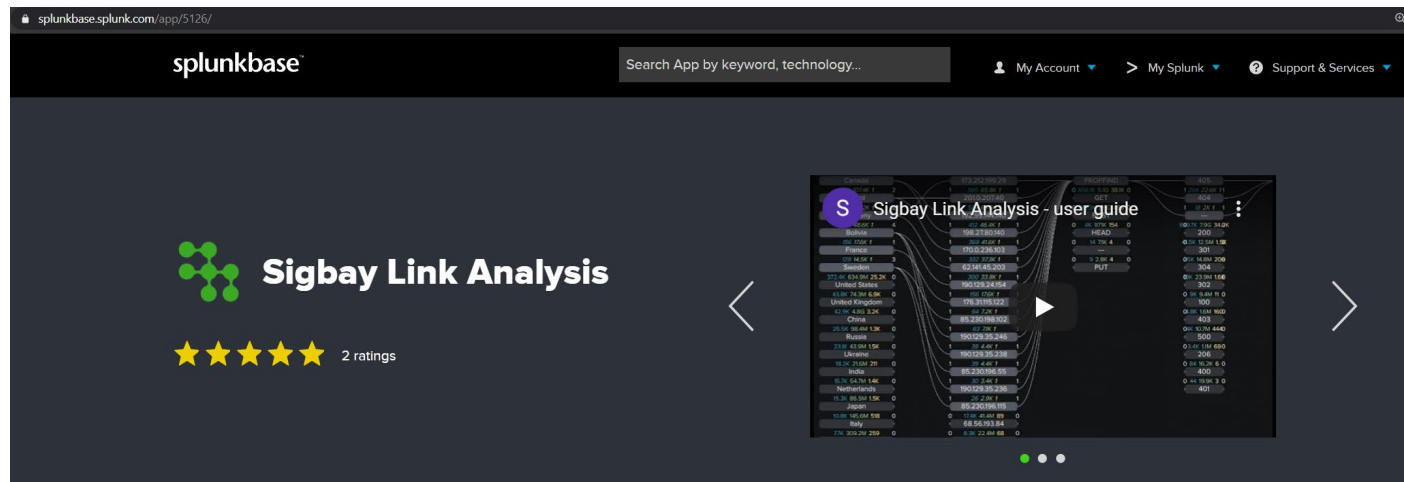
```
| tstats | stats | streamstats | eventstats | dc | values | avg | stdev | top | rare  
| sum | max | regex |
```

Links and Relationships Analysis

SigBay Link Analysis for Splunk

Free App on Splunkbase:

<https://splunkbase.splunk.com/app/5126/>

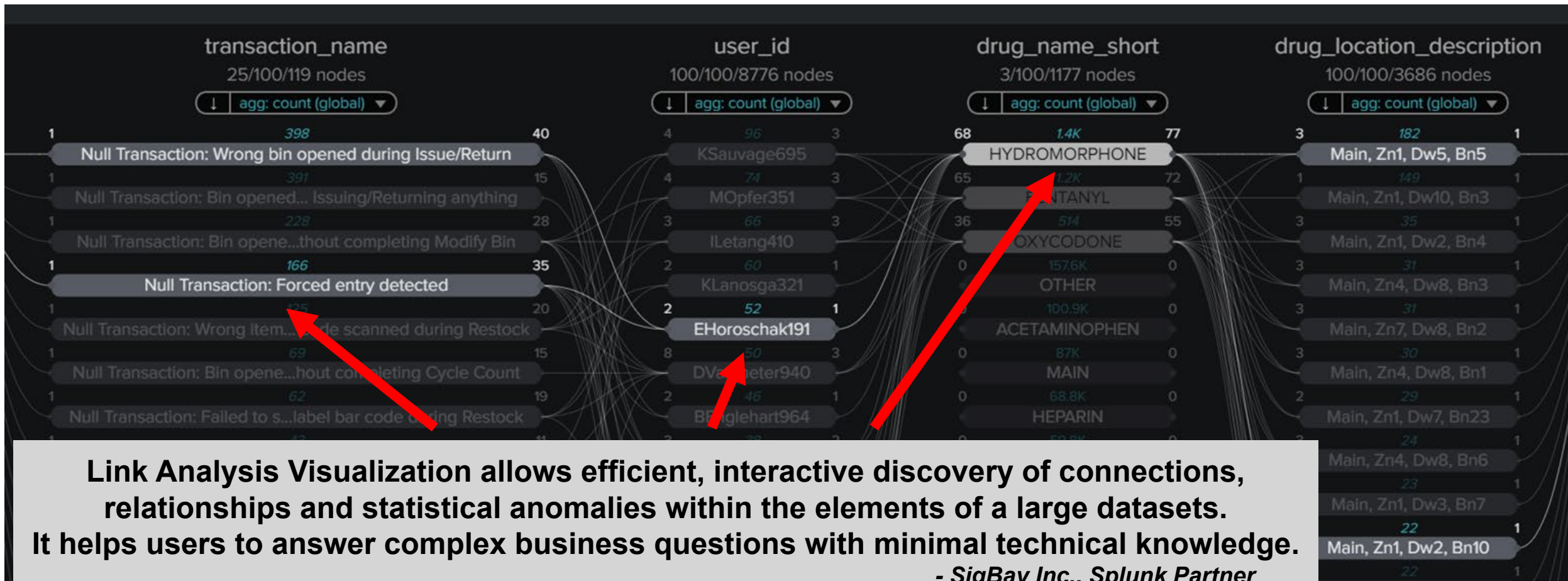


What Is SigBay Link Analysis?

- SigBay Link Analysis is an Interactive Data Exploration and Visualization Framework allowing non-technical business users to quickly gain actionable insights from large amounts of complex data.
- Its purpose is to reduce disconnect between less technical business users and complexity of the underlying datasets.
- Layered implementation allows interfacing with many data repositories.



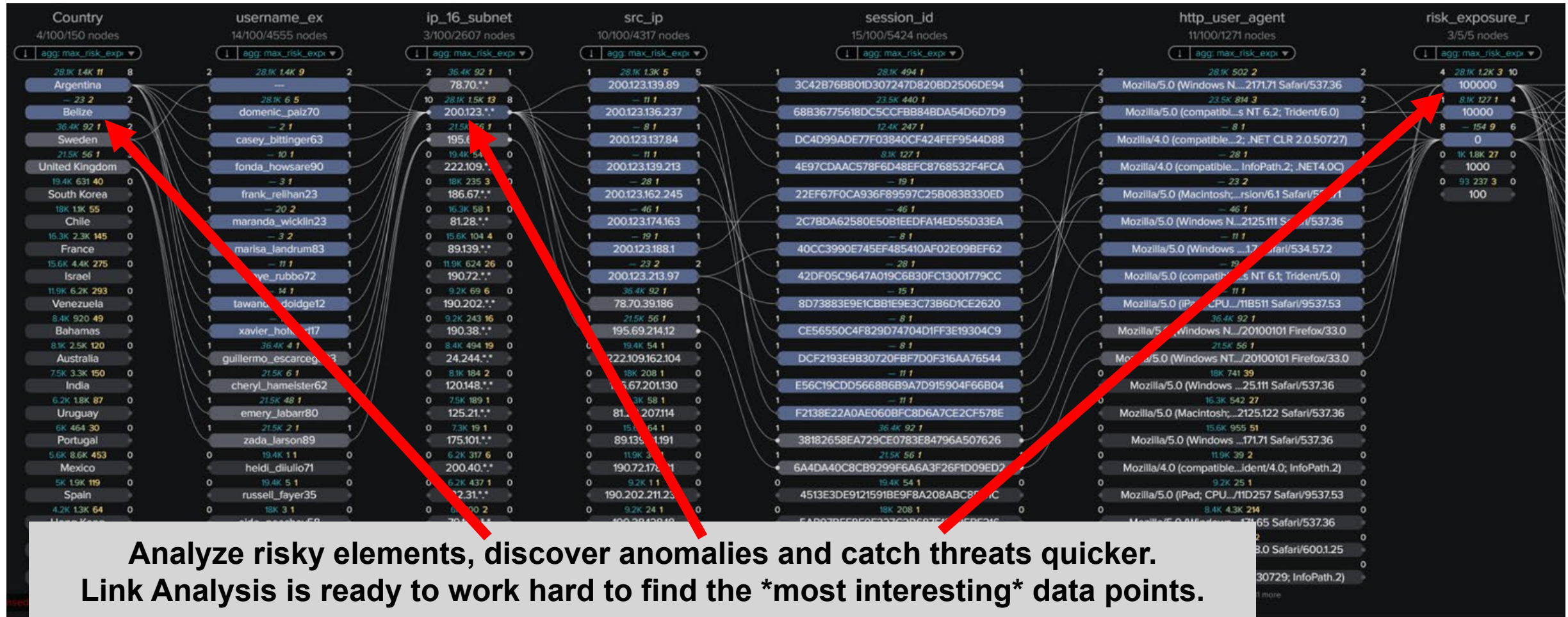
Interactive Data Analysis



Link Analysis Visualization allows efficient, interactive discovery of connections, relationships and statistical anomalies within the elements of a large datasets. It helps users to answer complex business questions with minimal technical knowledge.

- SigBay Inc., Splunk Partner

Anomalies and Threats Discovery



Analyze risky elements, discover anomalies and catch threats quicker.
Link Analysis is ready to work hard to find the *most interesting* data points.

- SigBay Inc., Splunk Partner

Capabilities

- Fully **Interactive**, every node and data element supports multiple contexts
- Discovery and Clear Presentation of the **Most Interesting** Data Points
- Discovery of Any-to-Any: **Relationships, Connections and Correlations**
- Automatic calculation of multiple **summary, aggregate and statistical** functions simultaneously over ALL data elements
- Automatic and manual **sorting** by any chosen metric, order and number of connections by any other entity or group.
- Discovery of **anomalous** relationships, **unusual** velocities and **suspicious** data patterns.
- **Peer group** analysis and **outlier** discovery.
- **Multi-selecting**, global searching and **hot-keys** controls over all functions.
- Real-time REST API calls to get specific answers and retrieve/update data points. Can communicate with other systems such as Phantom.
- Support for (require) Splunk **Accelerated Data Model** to tap into large datasets dynamically.

Implementation and Underlying Technology

- Custom built using industry standard libraries.
- The UI part is driven by **React** (open source library developed by Facebook). React utilizing Virtual DOM approach for faster incremental HTML updates only when they are needed.
- The data and state management part is done with **MobX** library.
- The visualization part is based on **D3** and **SVG**.
- The overall project is bundled into a single Splunk visualization using **Webpack**, allowing to be used as component in the Splunk Web UI or independently by using the REST API to communicate with Splunk and other data analytical solutions.

Examples



Finding Stolen Credit Cards: Case 1

Detecting Suspicious Access to Patient Records



Link Analysis: Detecting Suspicious Access to Records

Payment Cards: Link Analysis

Select timeframe:

All time

Hide Filters

Reset Dashboard

Link Analysis:

Aggregate Priority:

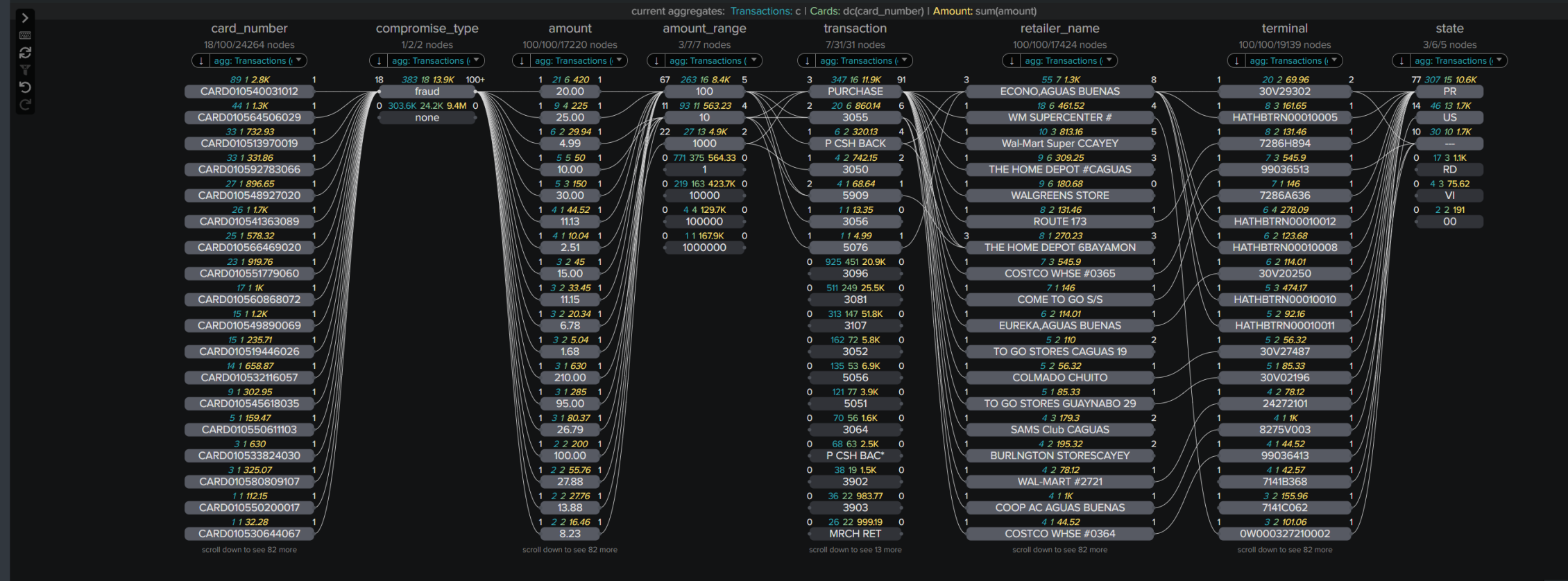
Events

Cards

Amount

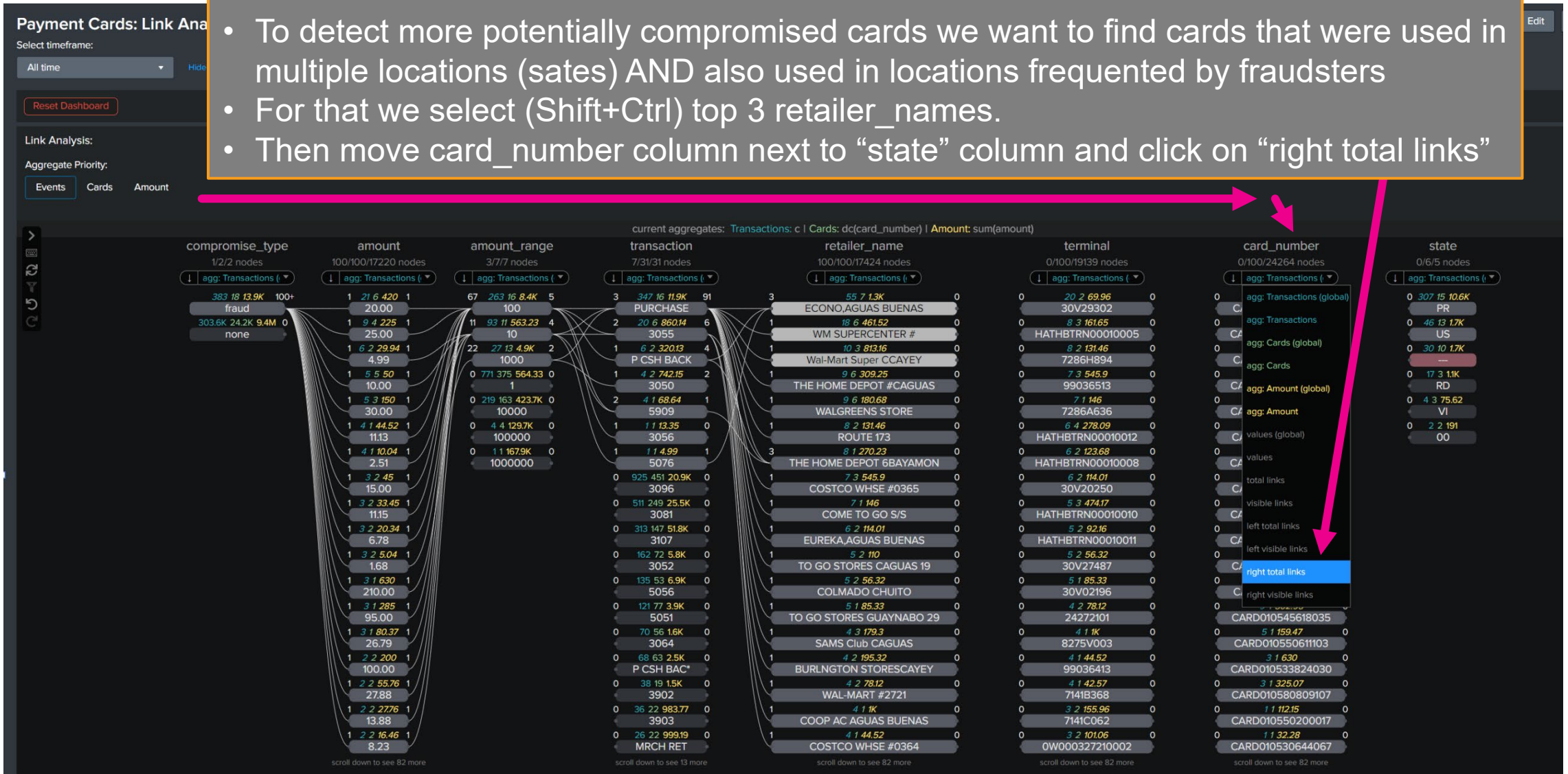
Edit

- In this example we investigate usage of stolen and duplicated credit cards
- Click on “compromise_type” = “fraud” loads all available data about stolen cards: numbers, transaction amounts and places where cards were used.



Link Analysis: Detecting Suspicious Access to Records

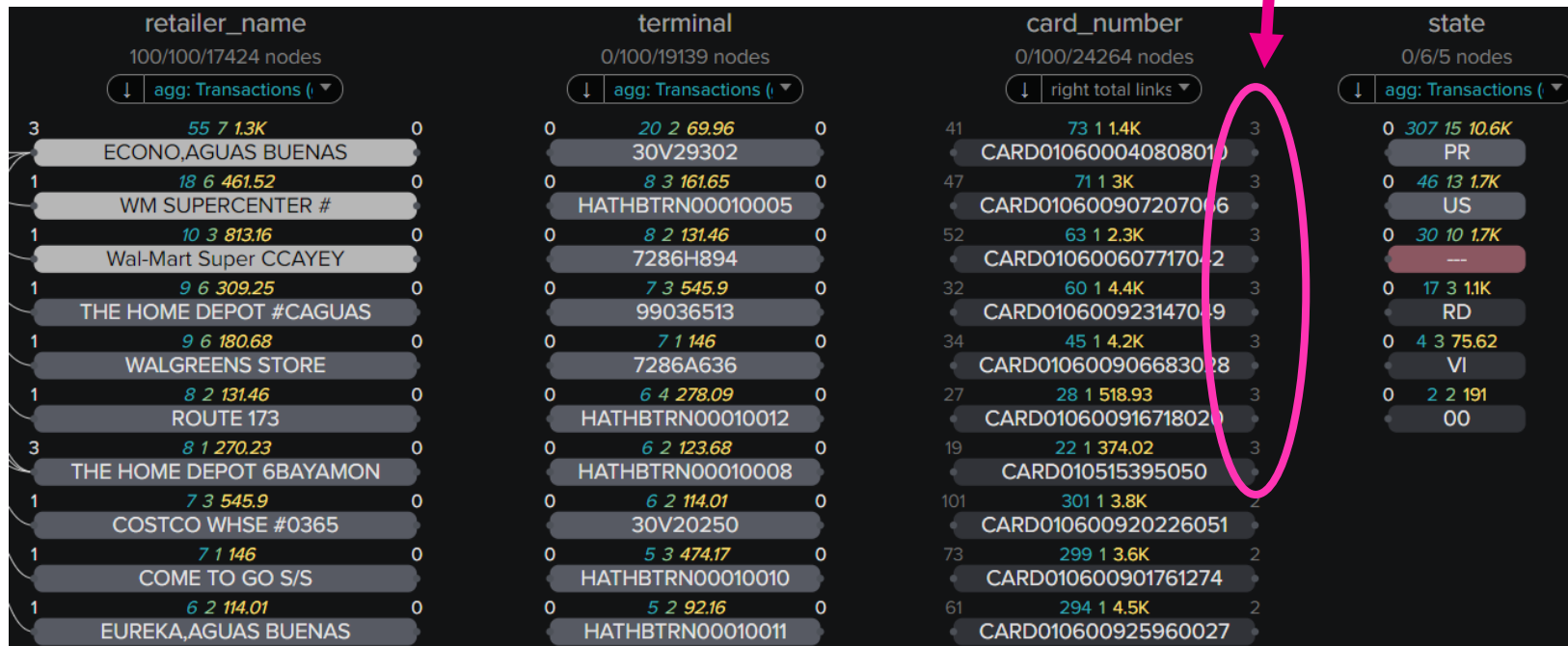
- To detect more potentially compromised cards we want to find cards that were used in multiple locations (states) AND also used in locations frequented by fraudsters
- For that we select (Shift+Ctrl) top 3 retailer_names.
- Then move card_number column next to “state” column and click on “right total links”



Link Analysis: Detecting Suspicious Access to Records

- When you click on any node or interact with Link Analysis visualization – it sends dynamic request to Accelerated Data Model and retrieves fresh results.
- In this case the query returned 7 cards that matches our query:

**Find cards that been used in different states
(top 7 cards been used in 3 states)**



Link Analysis: Detecting Suspicious Access to Records

Payment Cards: Link Analysis

Select timeframe:

All time

Hide Filters

Reset Dashboard

Link Analysis:

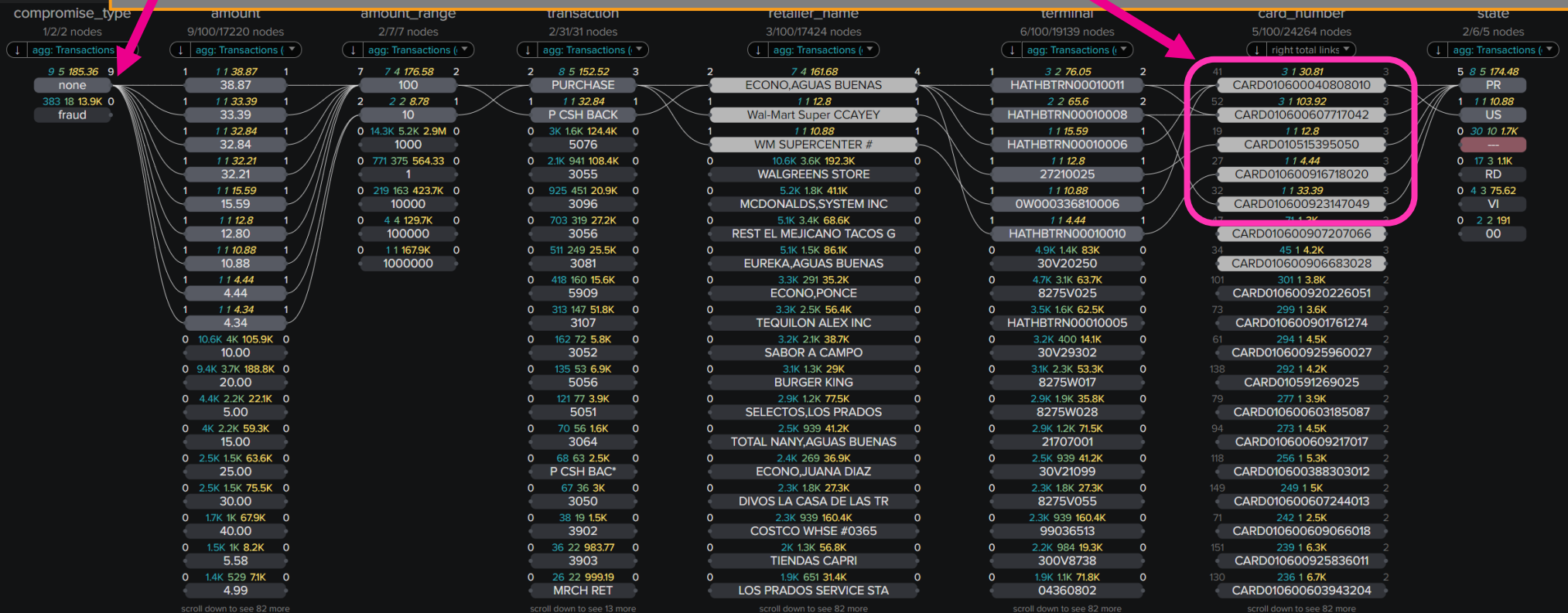
Aggregate Priority:

Events

Cards

Amount

- Next we select top 7 cards (Shift+Click) and run final query (Alt+Click on any selected):
Find cards that also been used in stores frequented by fraudsters.
- As a result – we found 5 cards that matches.
- None of them are marked as fraudulent yet which justifies further investigation to confirm.



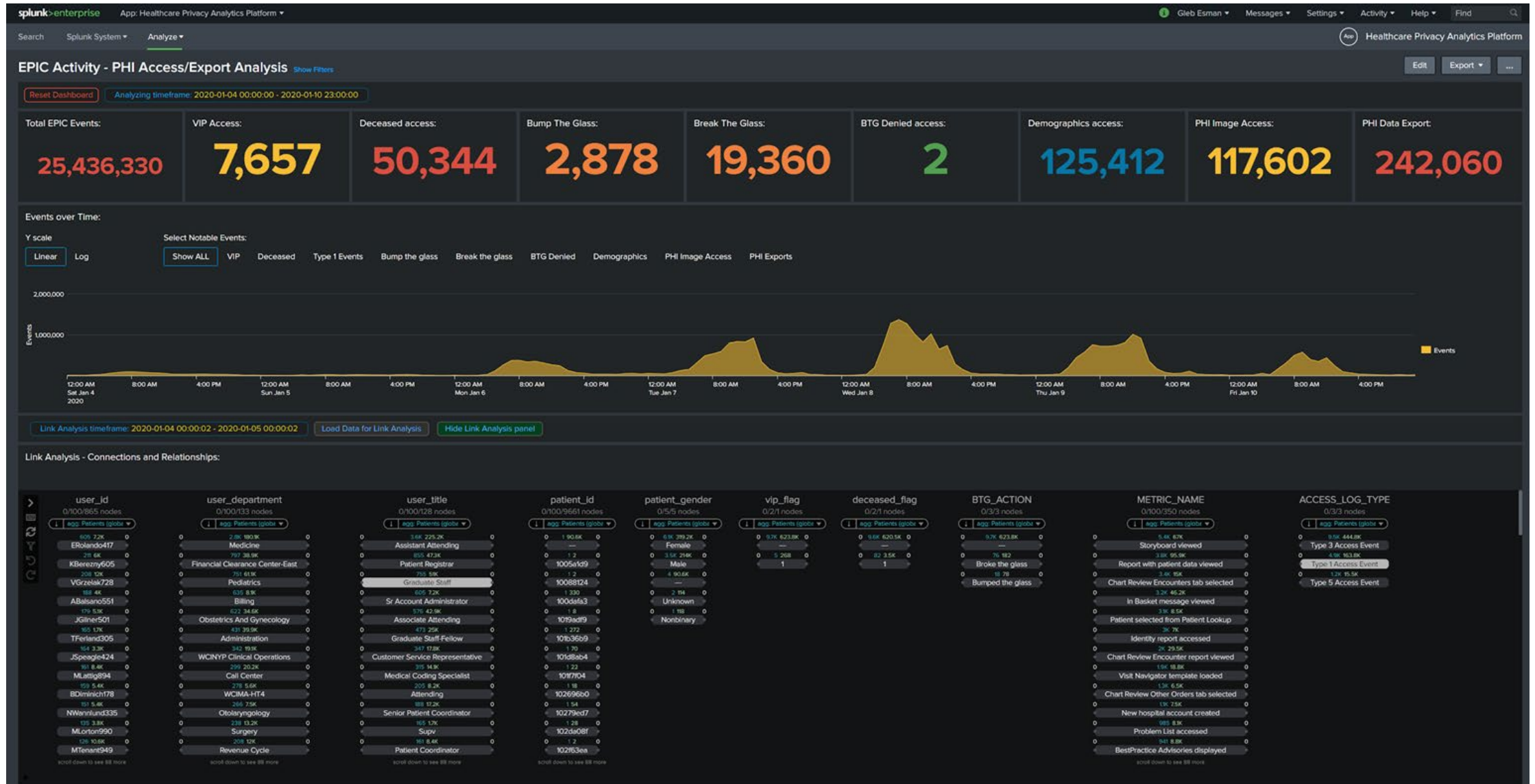


Finding Patient Privacy Violations: Case 2

Detecting Suspicious Access to Patient Records

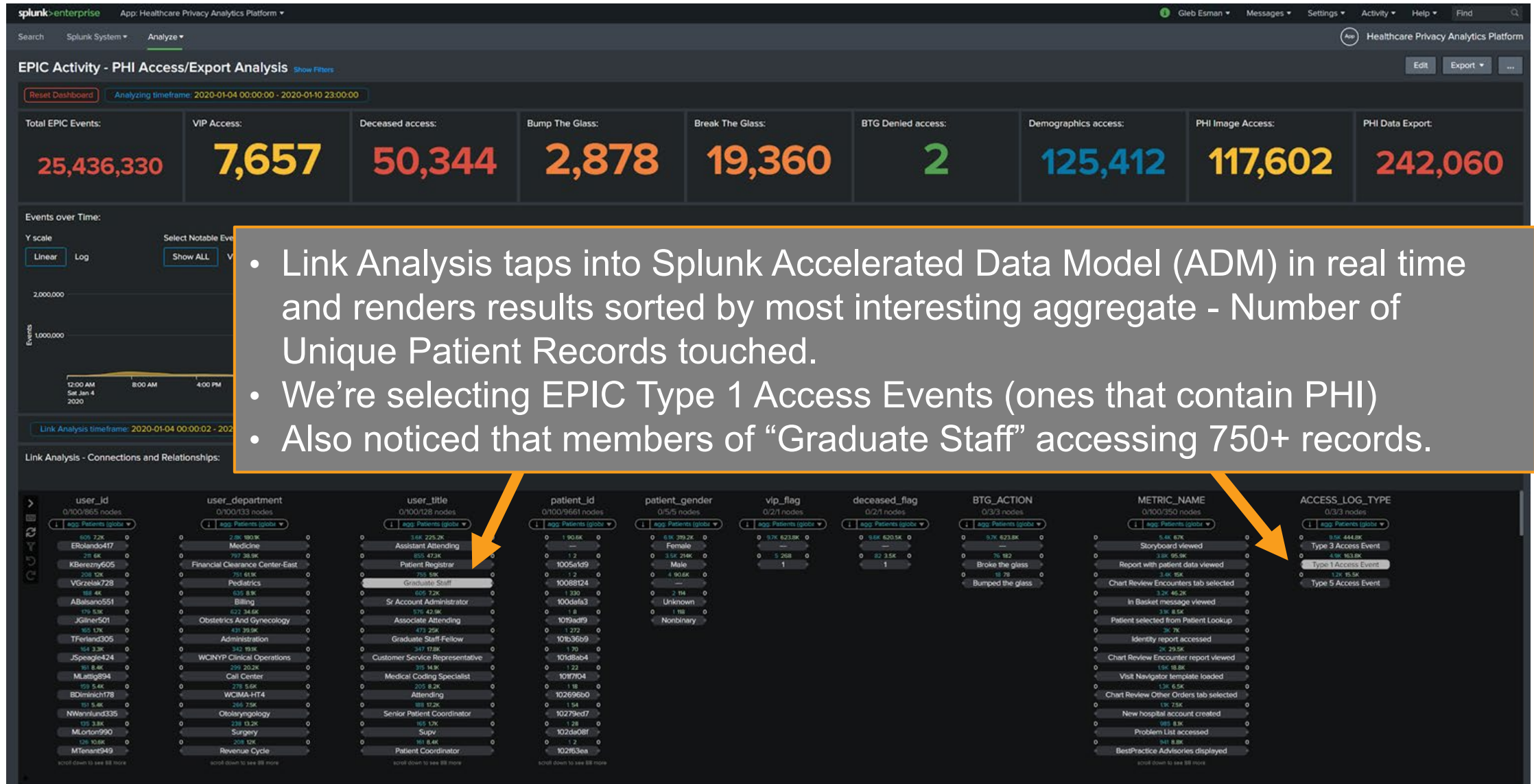


Link Analysis: Detecting Suspicious Access to Records



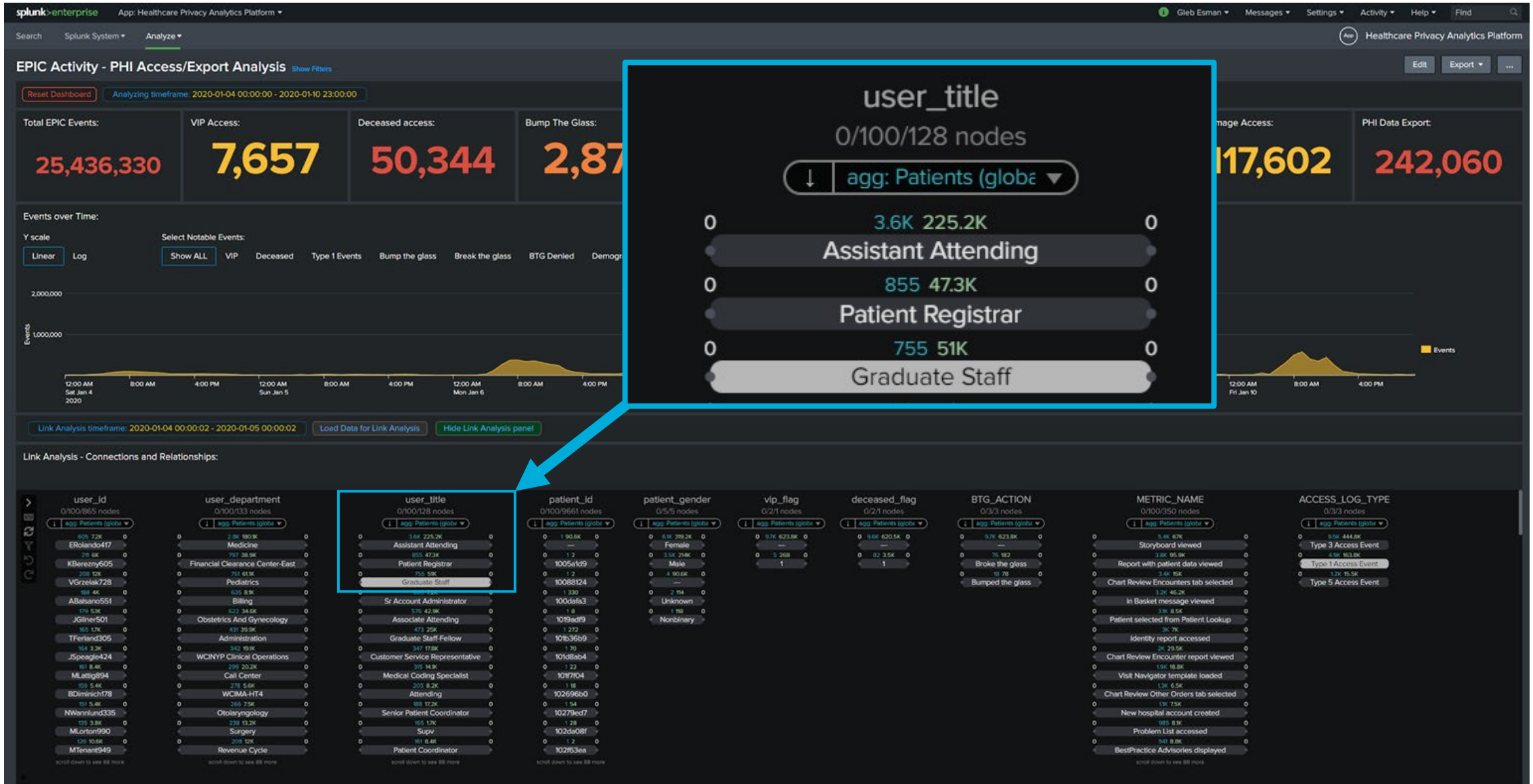
Link Analysis: Detecting Suspicious Access to Records

© 2020 SPLUNK INC.

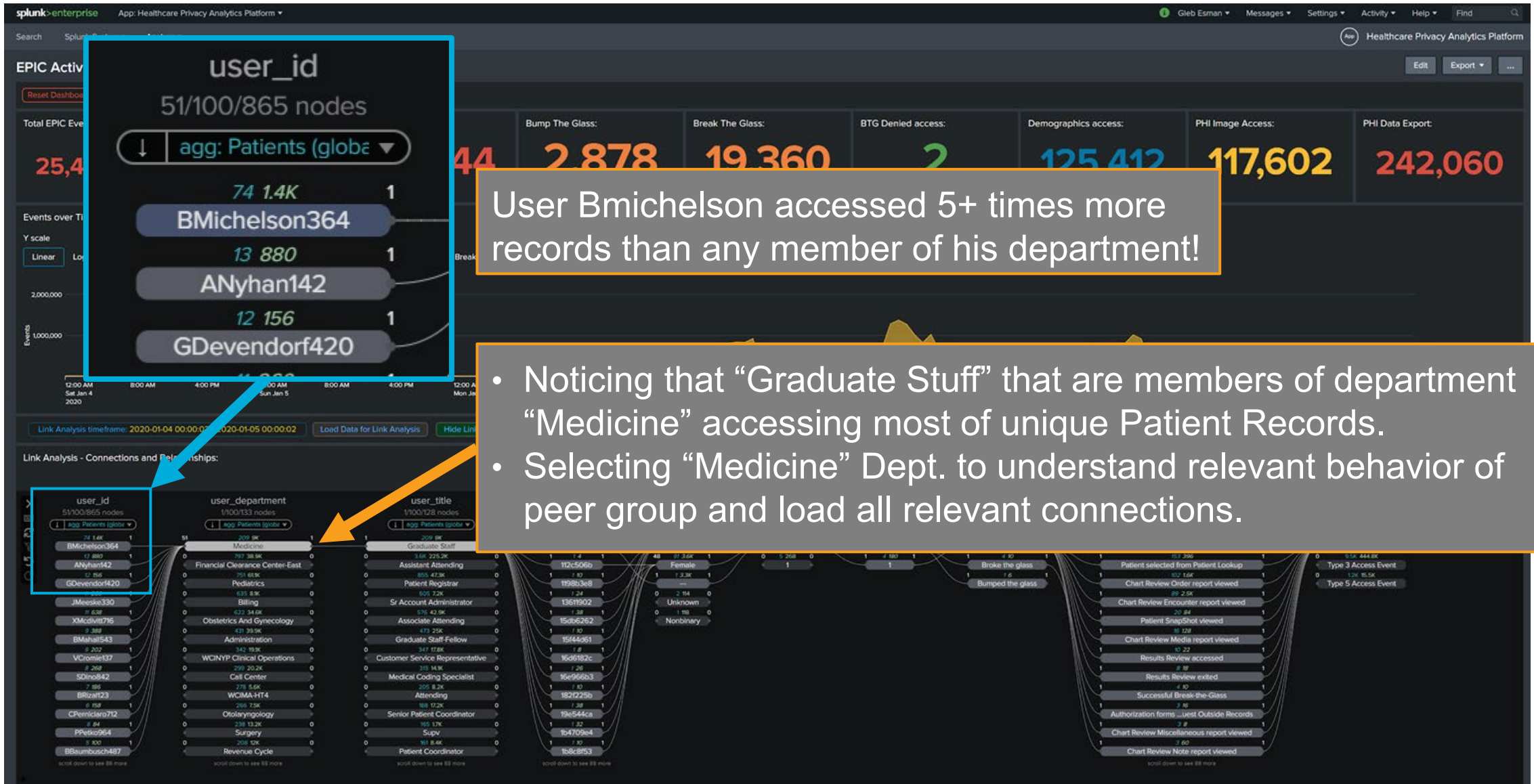


- Link Analysis taps into Splunk Accelerated Data Model (ADM) in real time and renders results sorted by most interesting aggregate - Number of Unique Patient Records touched.
- We're selecting EPIC Type 1 Access Events (ones that contain PHI)
- Also noticed that members of "Graduate Staff" accessing 750+ records.

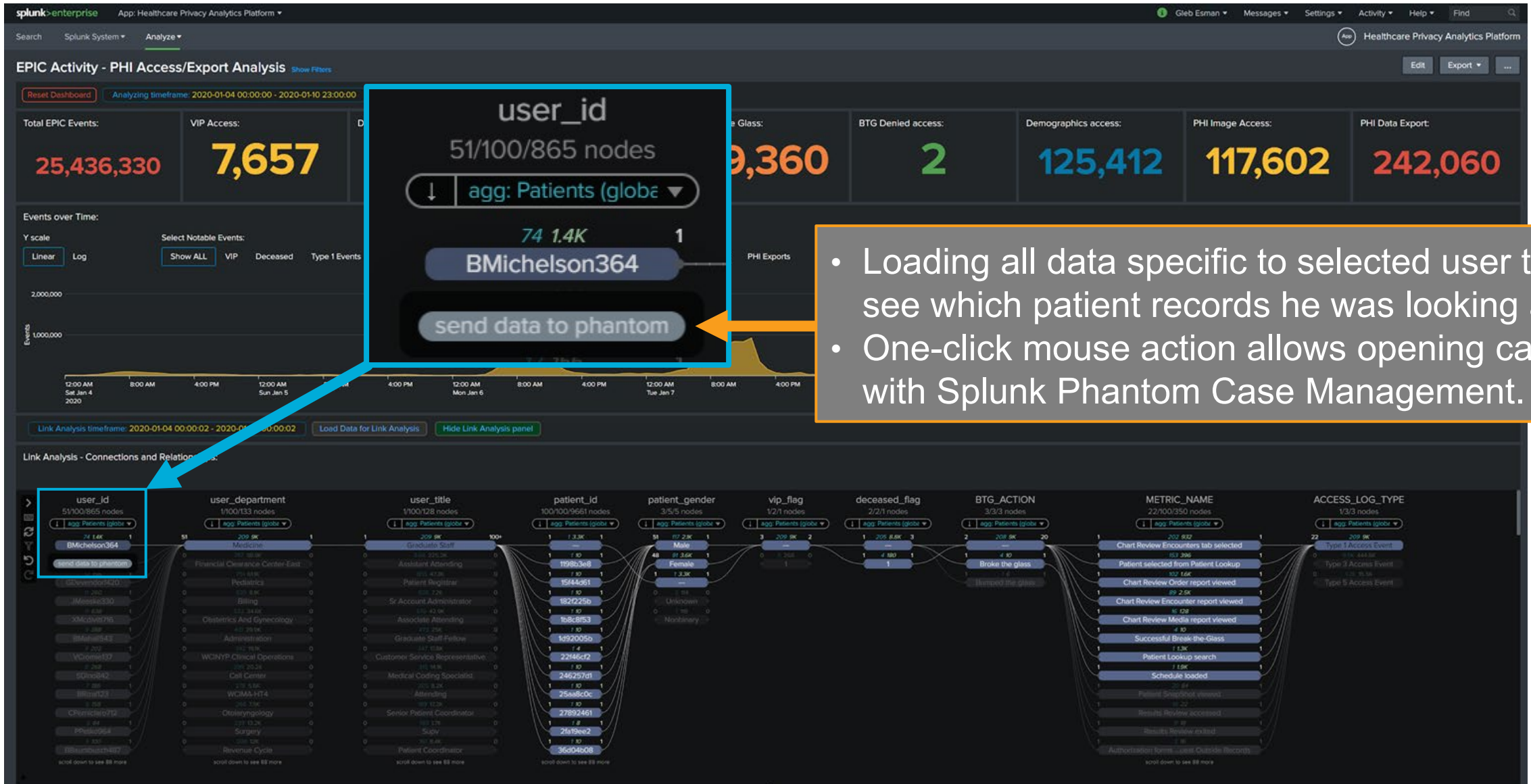
Link Analysis: Detecting Suspicious Access to Records



Link Analysis: Detecting Suspicious Access to Records



Link Analysis: Detecting Suspicious Access to Records



Creating Case within Phantom Case Management

© 2020 SPLUNK INC.

splunk>enterprise App: Healthcare Privacy Analytics Platform

Search Splunk System Analyze Healthcare Privacy Analytics Platform

Phantom Case Management - New case created: Edit Export

New Case Information:

Created New Case ID: 164736 on: 2020-02-25 20:58:04

Case Context:

Message: Investigate anomaly / suspicious activity

Case Metadata / Parameters:

Field name: user_id Field value: BMichelson364 Data Model: hcpa_user_activity Filter: hcpa_user_activity.APPLICATION=epic earliest="1578124802" latest="1578211202"



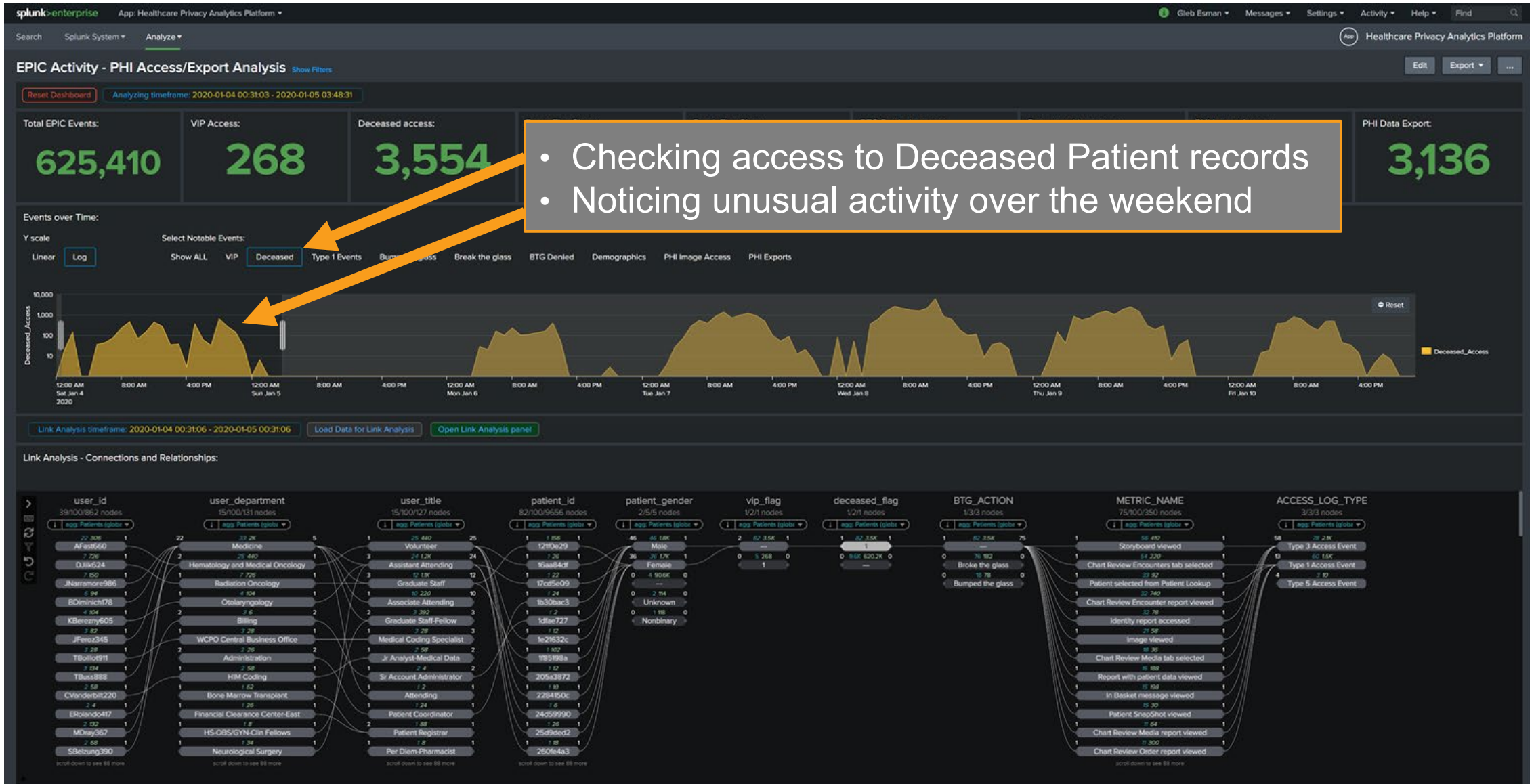
Finding Patient Privacy Violations: Case 3

Investigating Weekend Access Anomaly



Link Analysis: Investigating Weekend Access Anomaly

© 2020 SPLUNK INC.

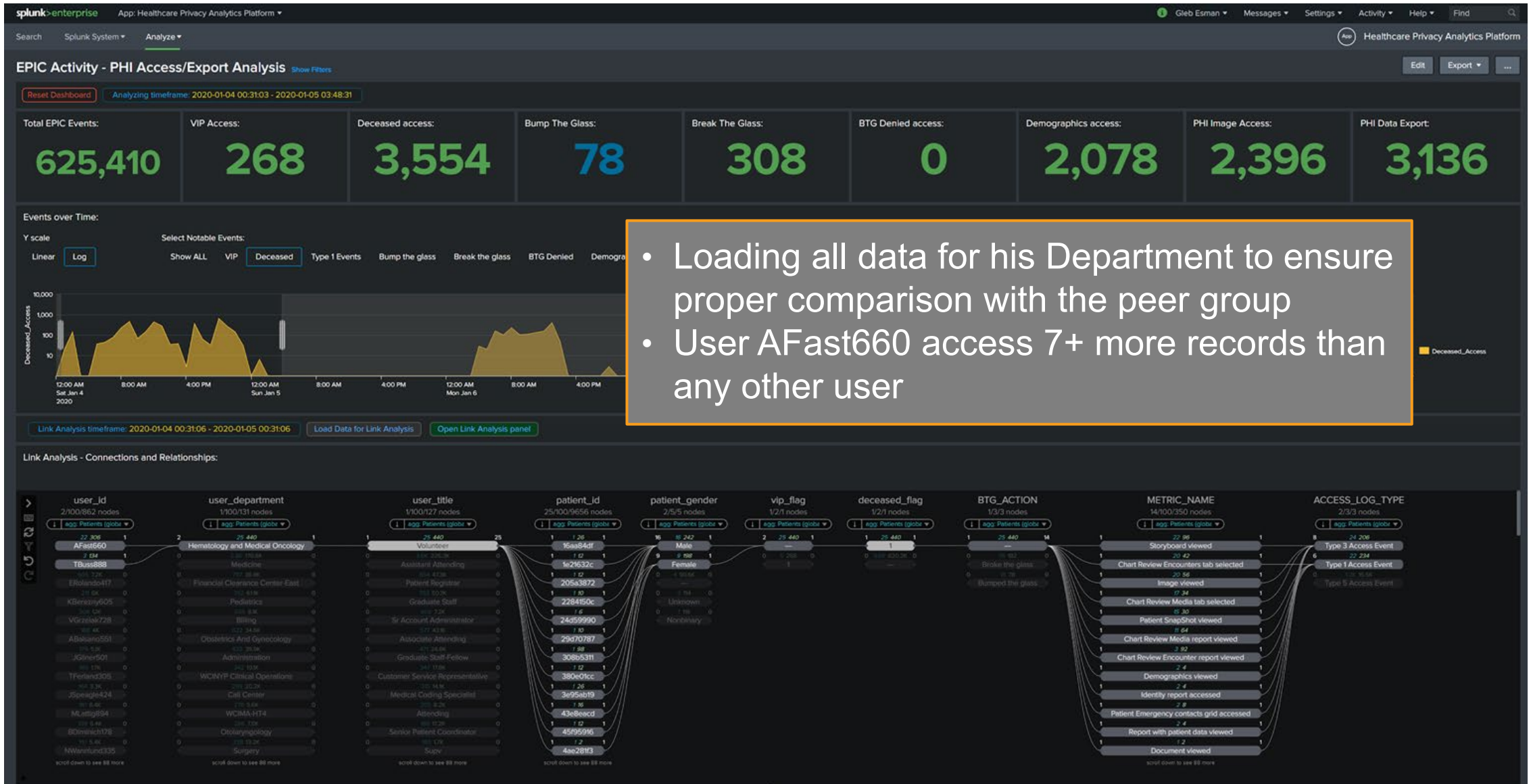


Link Analysis: Investigating Weekend Access Anomaly

© 2020 SPLUNK INC.

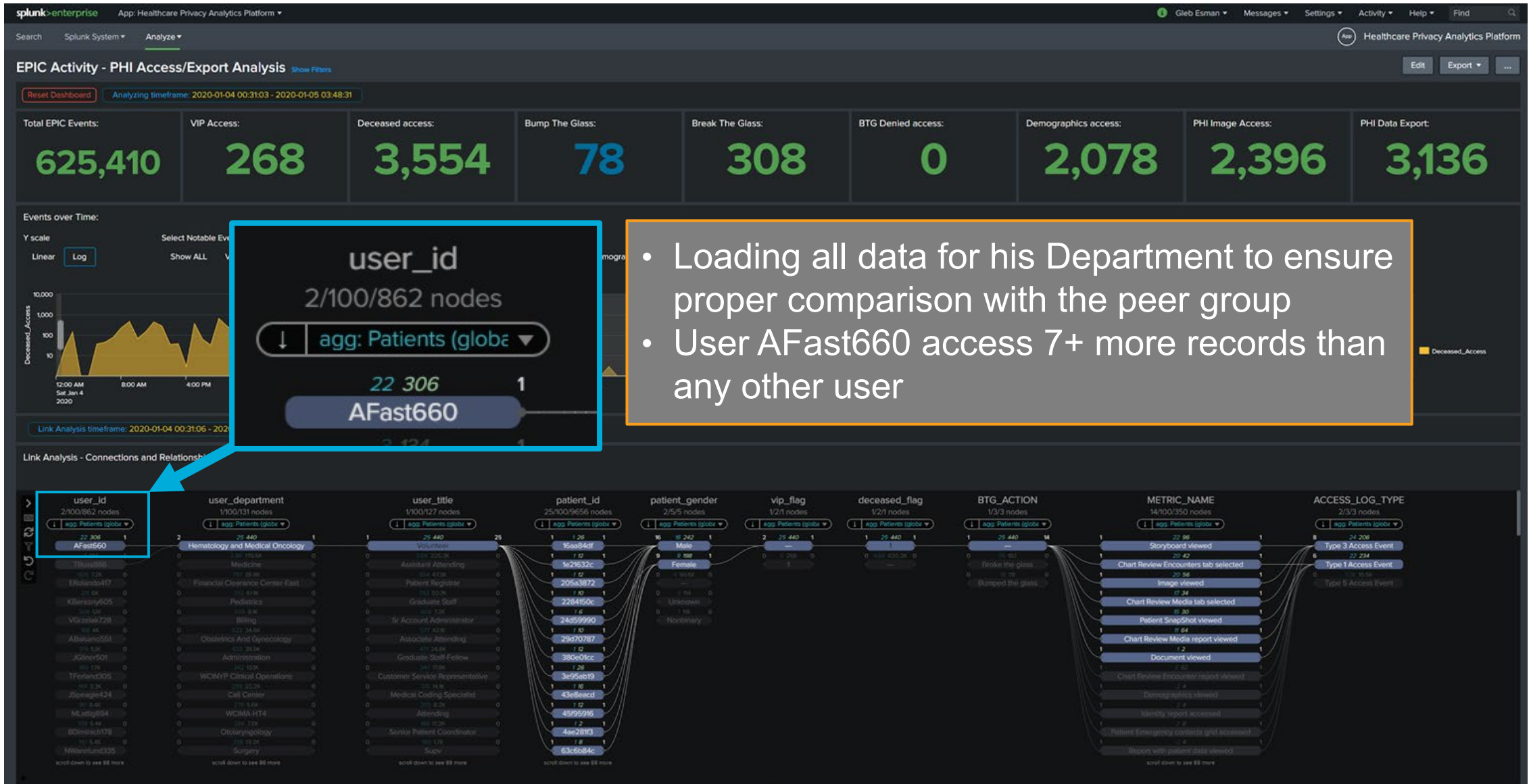


Link Analysis: Investigating Weekend Access Anomaly



- Loading all data for his Department to ensure proper comparison with the peer group
- User AFast660 access 7+ more records than any other user

Link Analysis: Investigating Weekend Access Anomaly



- Loading all data for his Department to ensure proper comparison with the peer group
- User AFast660 access 7+ more records than any other user



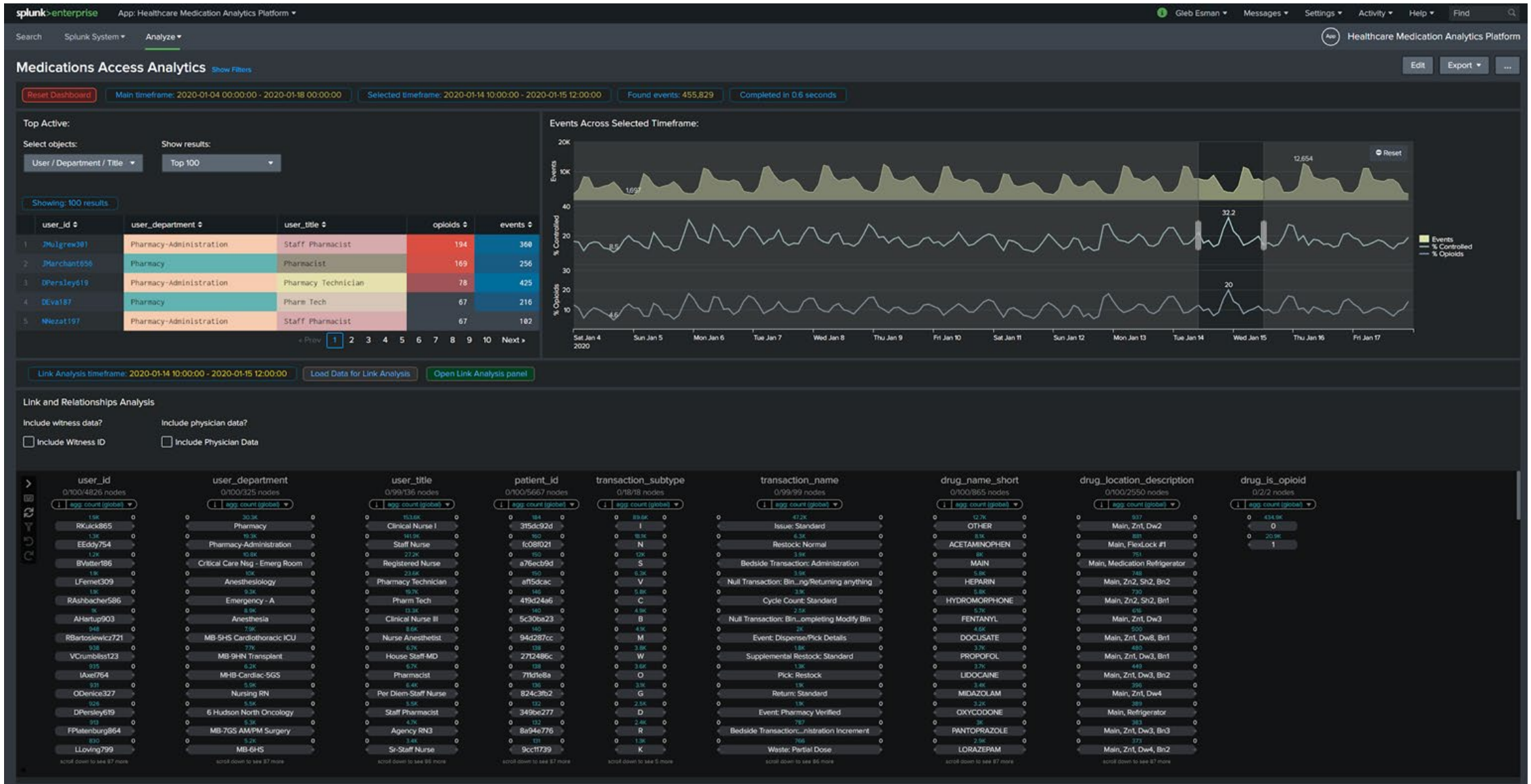
Detecting Opioid Diversion: Case 4

Accessing Drugs: NULL Transactions and Force Entry events



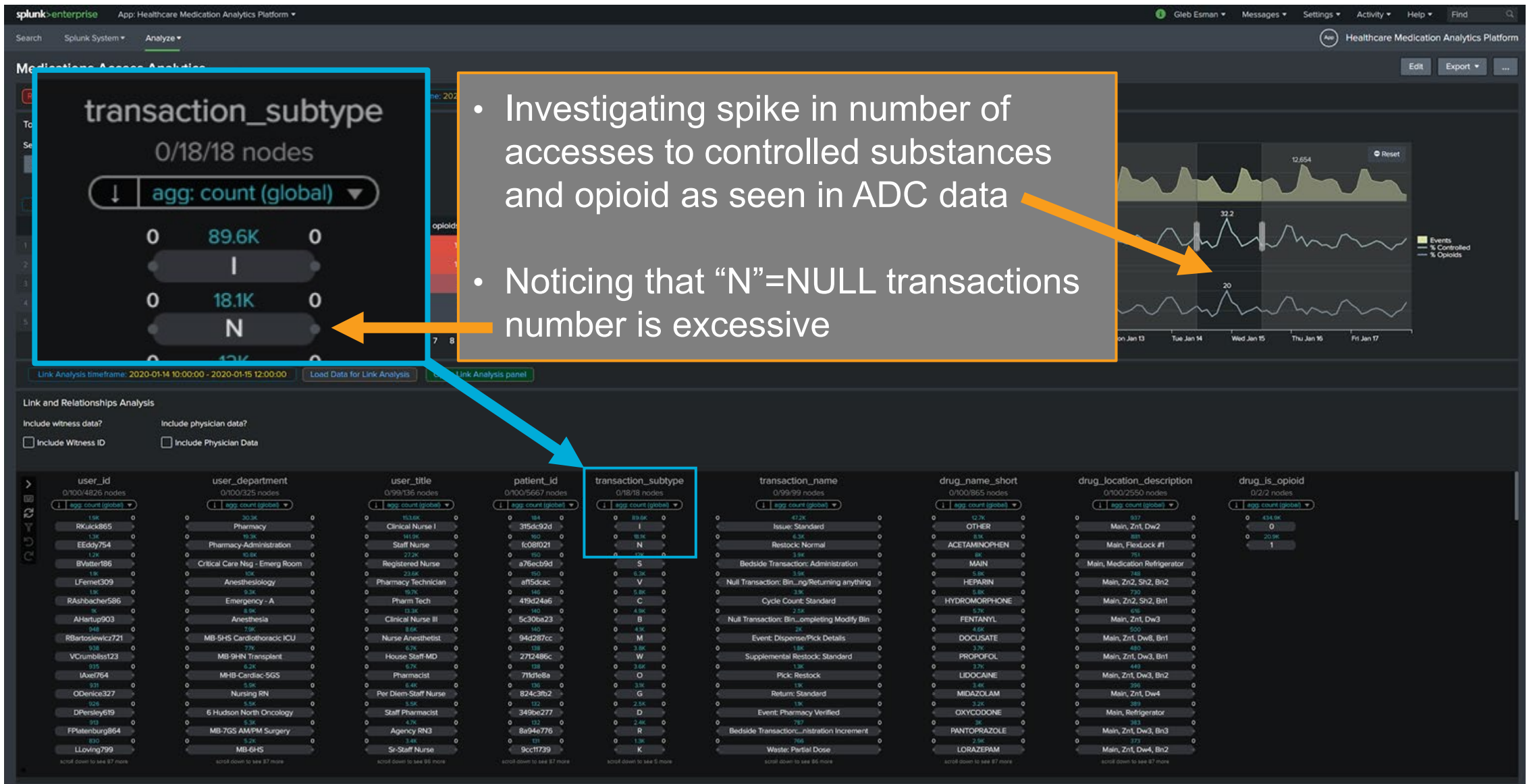
Link Analysis: Detecting Opioid Diversion

© 2020 SPLUNK INC.



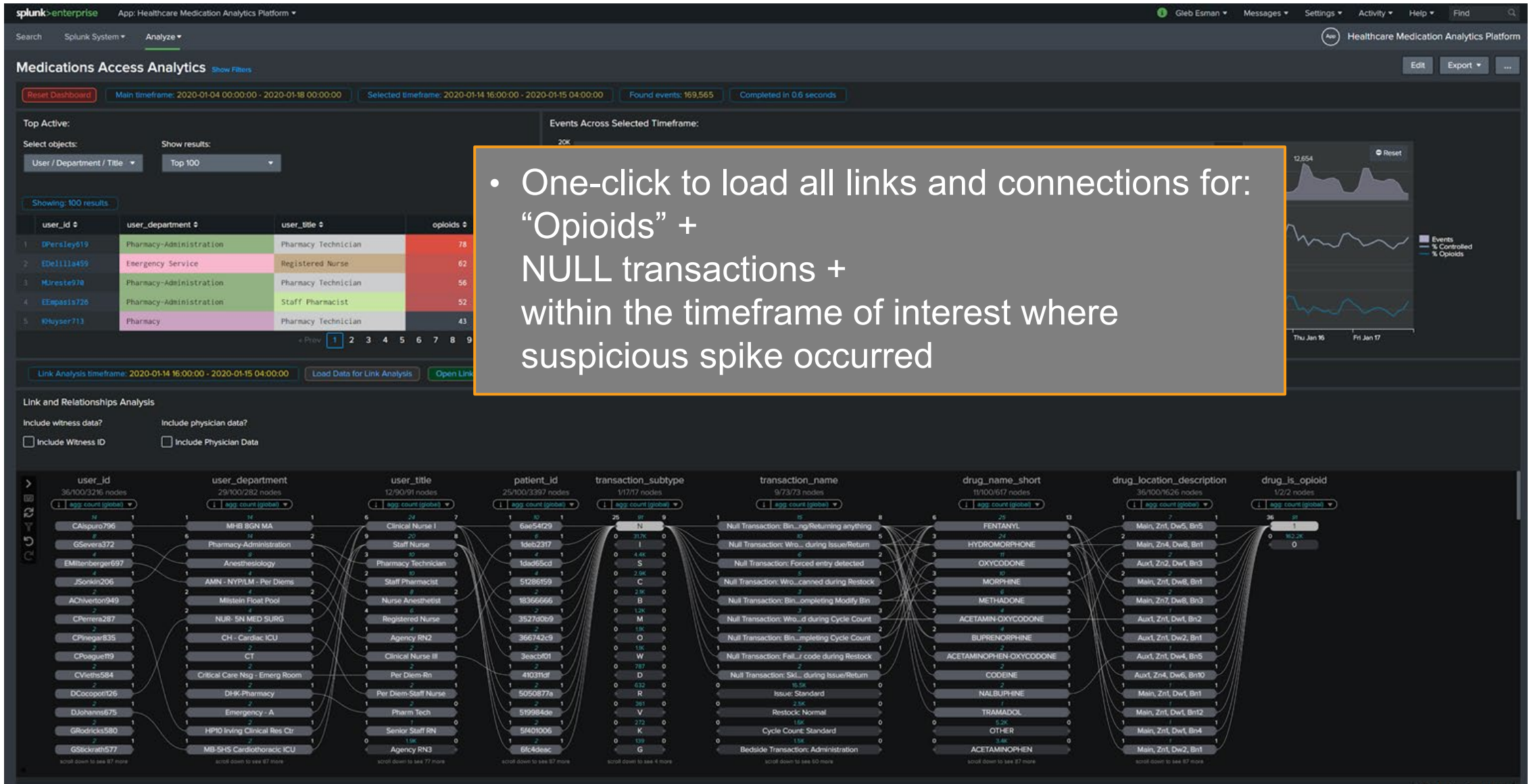
Link Analysis: Detecting Opioid Diversion

© 2020 SPLUNK INC.

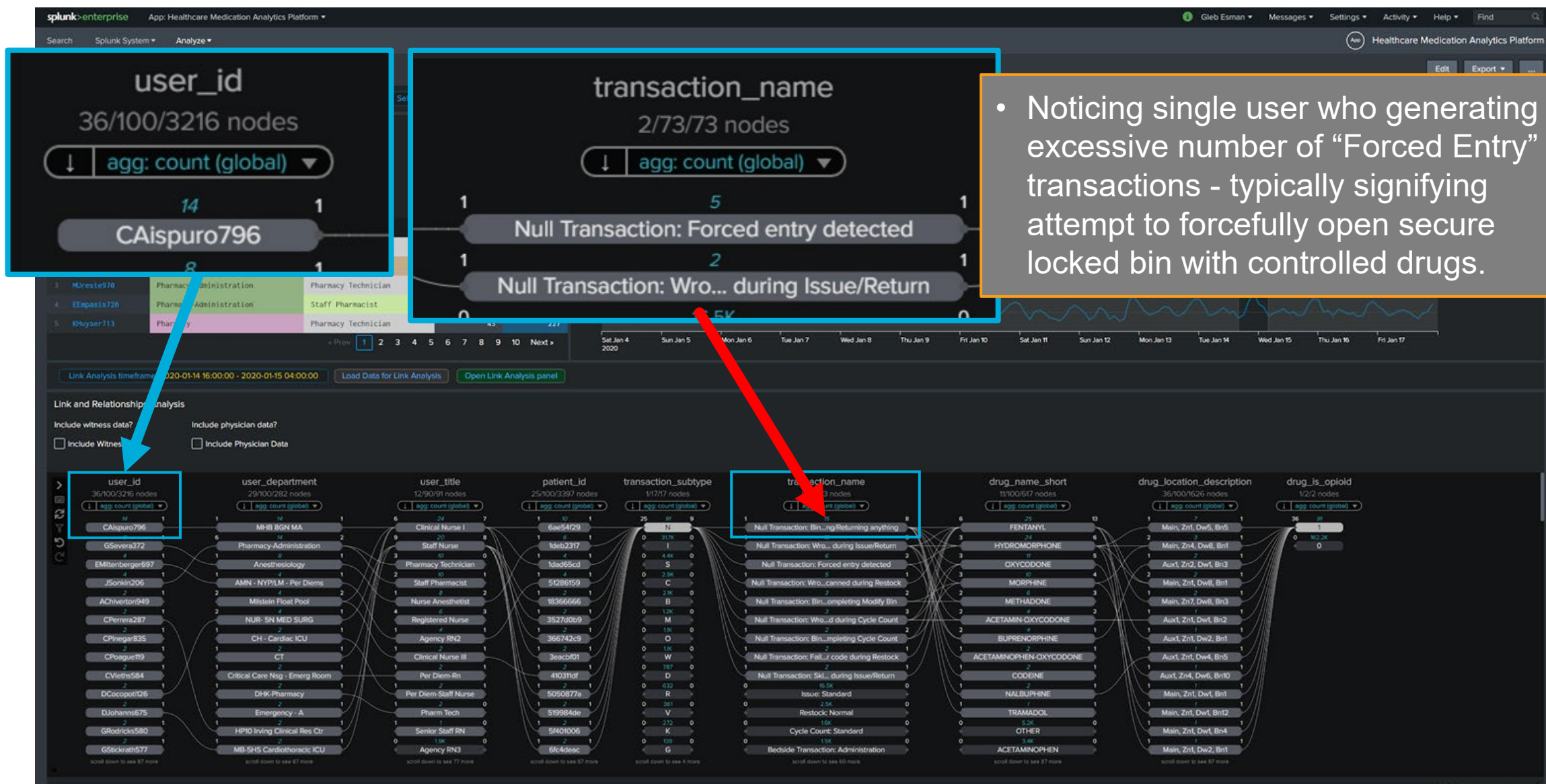


Link Analysis: Detecting Opioid Diversion

© 2020 SPLUNK INC.

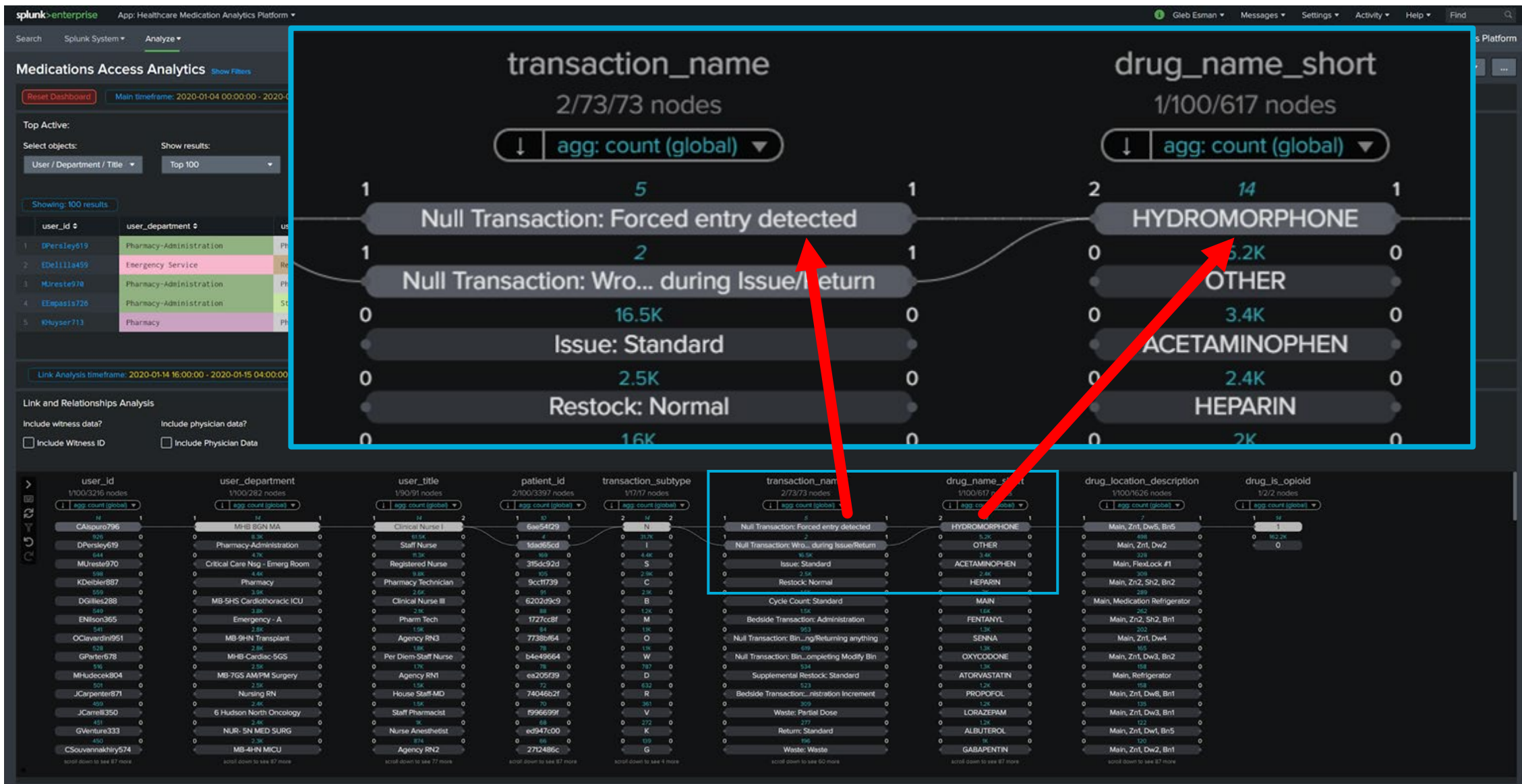


Link Analysis: Detecting Opioid Diversion



Link Analysis: Detecting Opioid Diversion

© 2020 SPLUNK INC.



Key Takeaways

1. Many link analysis visualization options for Splunk
2. [SigBay Link Analysis Viz App](#) allows for sophisticated, interactive investigations within the large datasets without need to write complex SPL queries.
3. Other link visualization options on Splunkbase, but may require data manipulation/reduction for concise view.

More Information?

Links

- Force Directed App <https://splunkbase.splunk.com/app/3767/>
- Network Diagram Viz <https://splunkbase.splunk.com/app/4438/>
- Sigbay Link Analysis <https://splunkbase.splunk.com/app/5126/#/details>
- Eventstats <https://docs.splunk.com/Documentation/Splunk/8.0.5/SearchReference/Eventstats>
- Link Analysis App for Splunk (.Conf19) <https://splunkbase.splunk.com/app/4676/#/details>
- .Conf19 BOTS, The Missing Link <https://conf.splunk.com/watch/conf-online.html?search=bots%20the%20missing%20link#/>
- Splunk Blog posts https://www.splunk.com/en_us/blog



Thank You

Please provide feedback via the
SESSION SURVEY

