Scaling Investigations

Next-Generation Visual Graph Analytics
through GPU Cloud Streaming

GRAPHISTRY

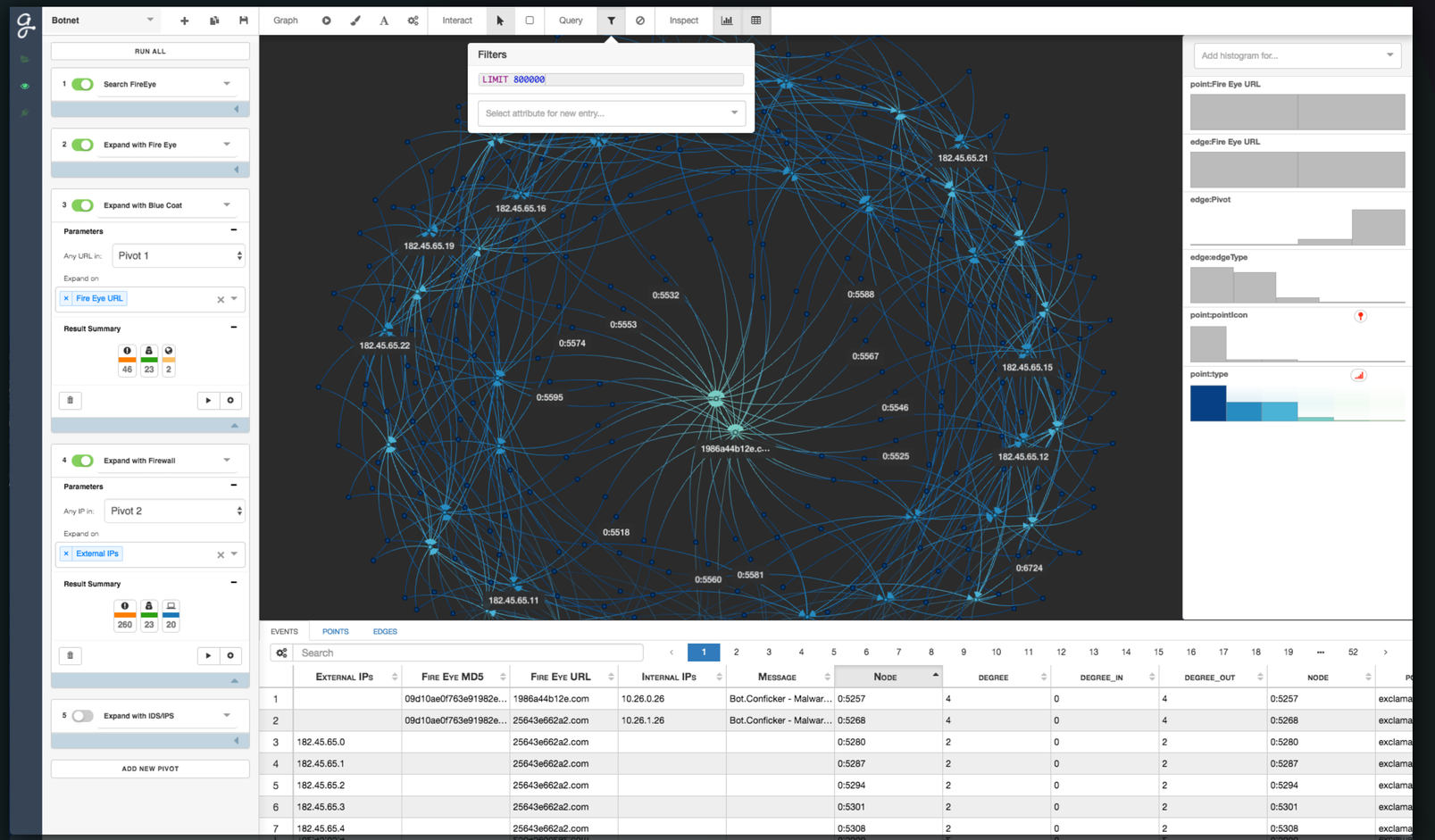
Leo Meyerovich, CEO
@LMeyerov

Graphistry Inc. 2017

info@graphistry.com

ABOUT

Graphistry is a place investigators can record scalable investigations, run them, and visually analyze the results



Graphistry Inc. 2017

info@graphistry.com

Investigating through event data is important… but HARD

- Customer 360
- System outages
- Genomics & health records

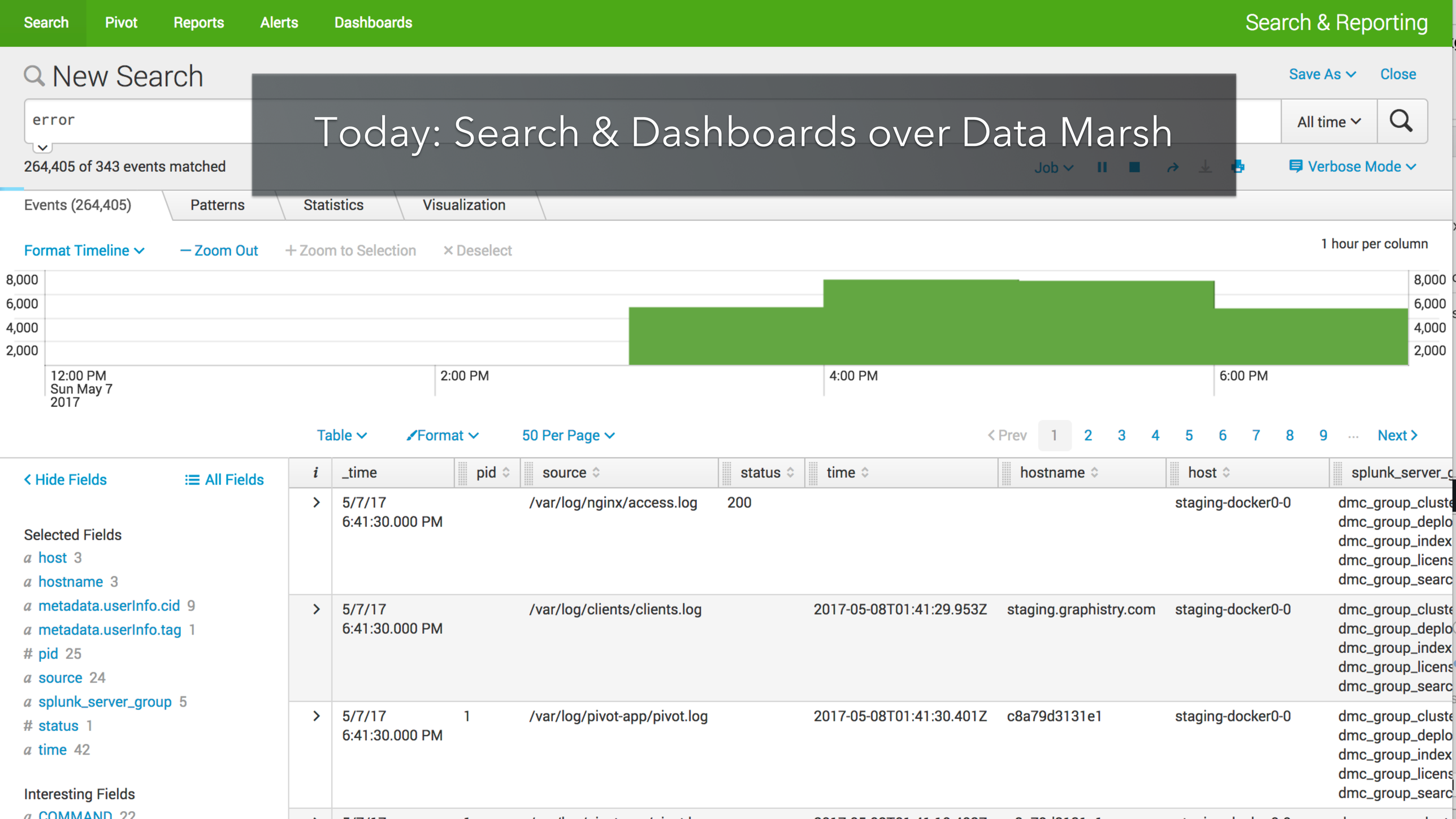
...

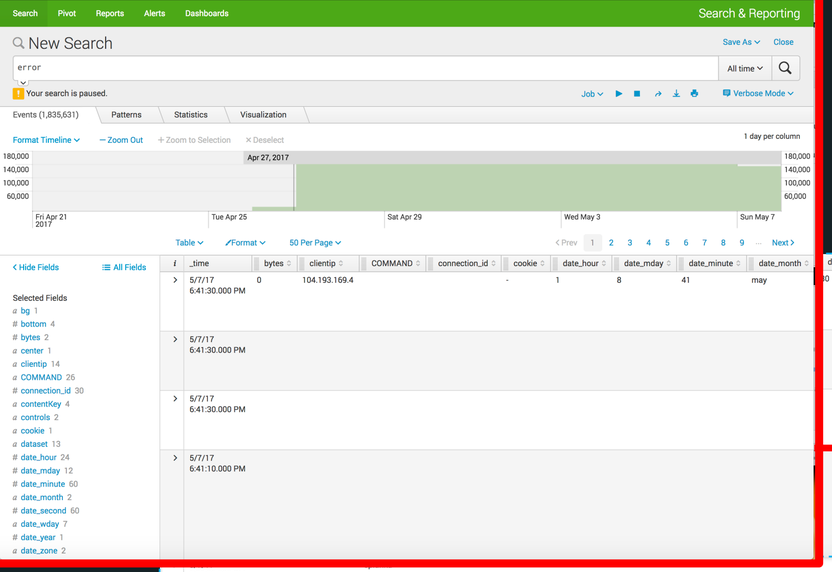
- Anti-human trafficking
- Cybersecurity
- Retail & billing fraud

...

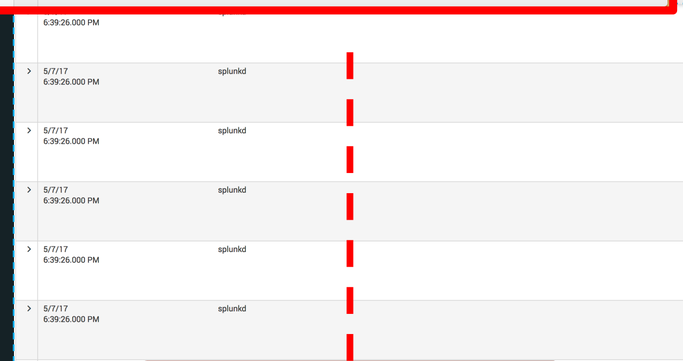
Goal: Quickly answer
progression, root cause, correlation, scope…

info@graphistry.com


Search
Pivot
Reports
Alerts
Dashboards
Search & Reporting
New Search
Save As
Close
error
All time
Today: Search & Dashboards over Data Marsh
264,405 of 343 events matched
Job
Verbose Mode
Events (264,405)
Patterns
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
1 hour per column
8,000
6,000
4,000
2,000
12:00 PM
Sun May 7
2017
2:00 PM
4:00 PM
6:00 PM
Table
Format
50 Per Page
Prev
1
2
3
4
5
6
7
8
9
Next
Hide Fields
All Fields
Selected Fields
host 3
hostname 3
metadata.userInfo.cid 9
metadata.userInfo.tag 1
pid 25
source 24
splunk_server_group 5
status 1
time 42
Interesting Fields
COMMAND 22
_time
pid
source
status
time
hostname
host
splunk_server_g
5/7/17
6:41:30.000 PM
/var/log/nginx/access.log
200
staging-docker0-0
dmc_group_cluste
dmc_group_deplo
dmc_group_index
dmc_group_licens
dmc_group_searc
5/7/17
6:41:30.000 PM
/var/log/clients/clients.log
2017-05-08T01:41:29.953Z
staging.graphistry.com
staging-docker0-0
dmc_group_cluste
dmc_group_deplo
dmc_group_index
dmc_group_licens
dmc_group_searc
5/7/17
6:41:30.000 PM
1
/var/log/pivot-app/pivot.log
2017-05-08T01:41:30.401Z
c8a79d3131e1
staging-docker0-0
dmc_group_cluste
dmc_group_deplo
dmc_group_index
dmc_group_licens
dmc_group_searc



100+



~infinity

Increasing details & data sources

Increasing resolution & duration

(+ Increasing iterative workflow complexity)

Handling 100X More Data in the Analyst Loop: Three 100X Technologies

Automate with Visual Playbooks

See 100X+ more data with GPU Visual Analytics

Make AI visual with Graph

info@graphistry.com

Today

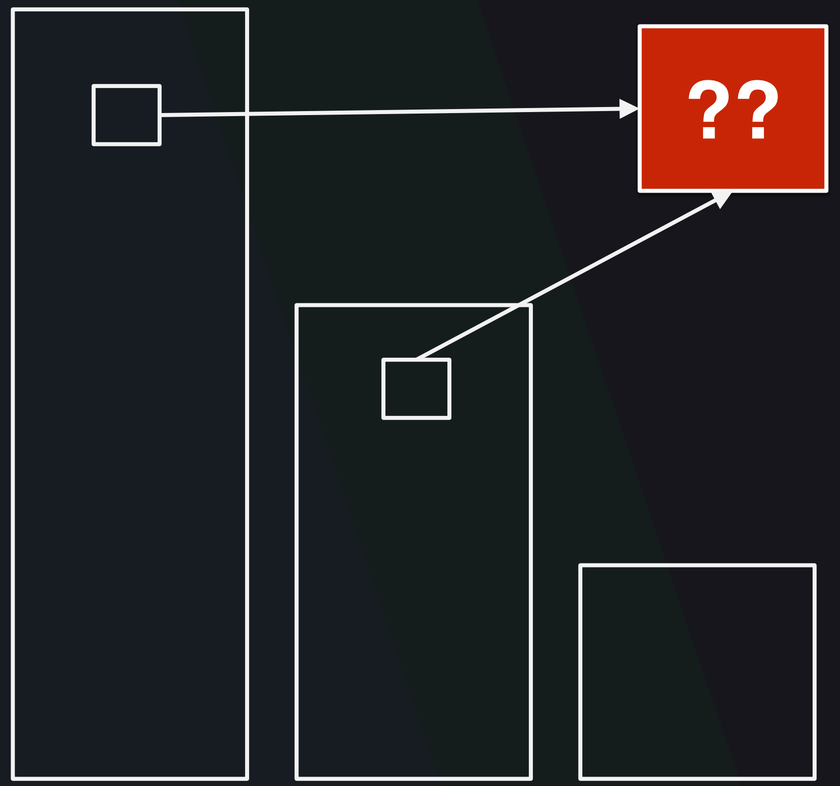
- Rethinking event data as graph analytics
- GPUs for interactive analytics
- GPUs for data-driven design
- Graphistry's client↔cloud end-to-end GPU architecture

info@graphistry.com

Problem: How to Understand 20+ Columns?

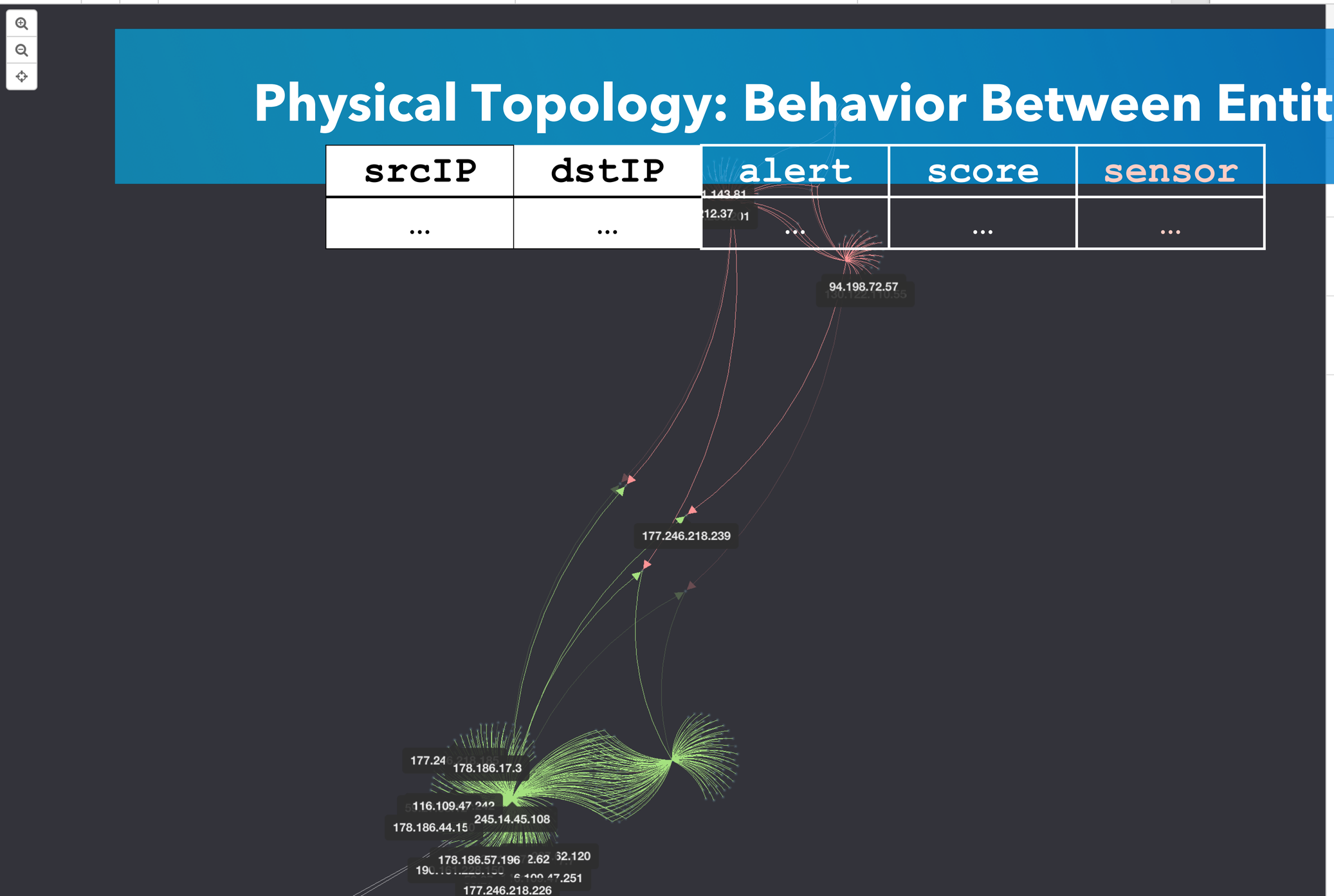
select * from ALERTS where PRIORITY > 8

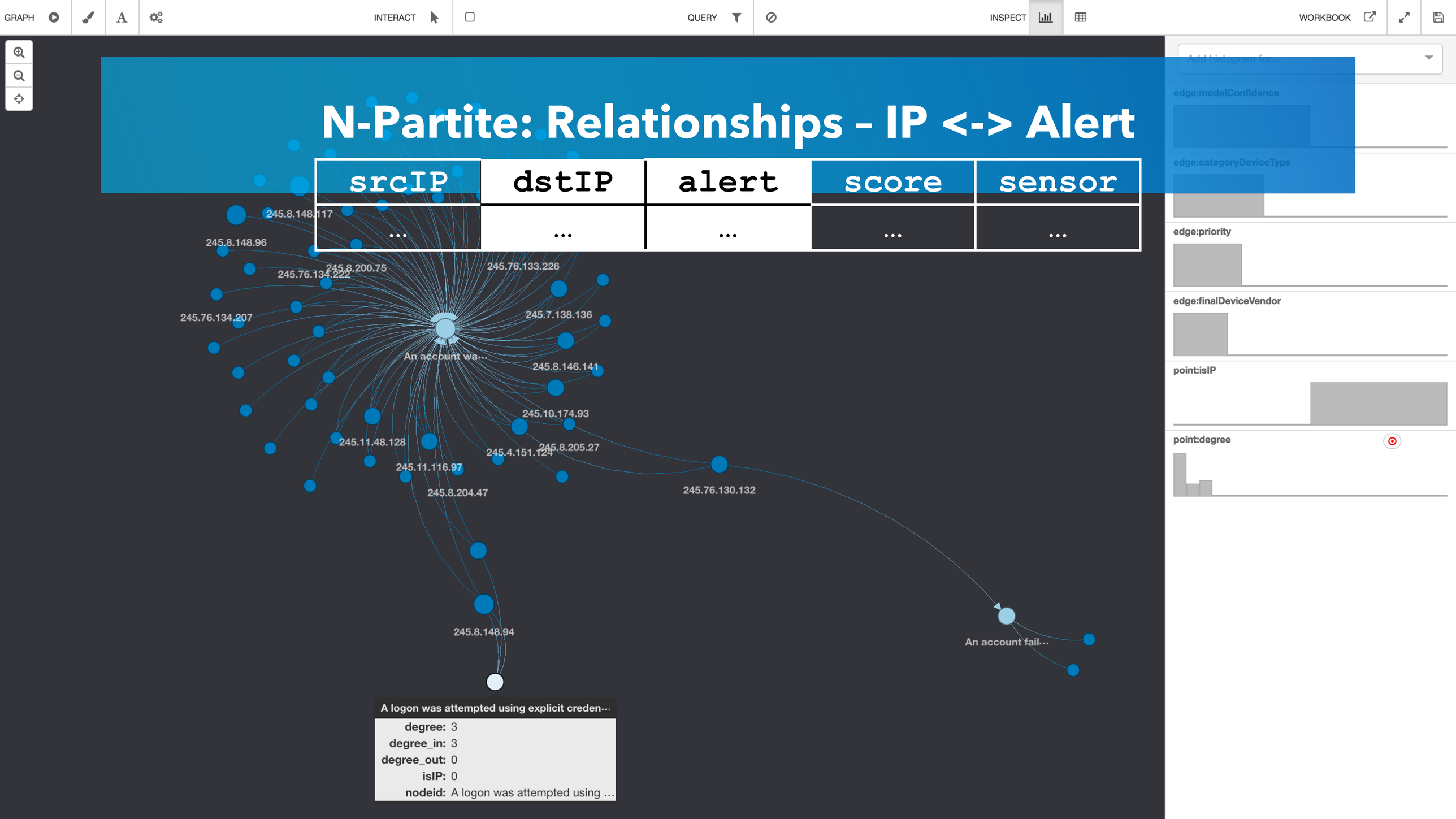
srcIP	dstIP	alert	score	time
…	…	…	…	…

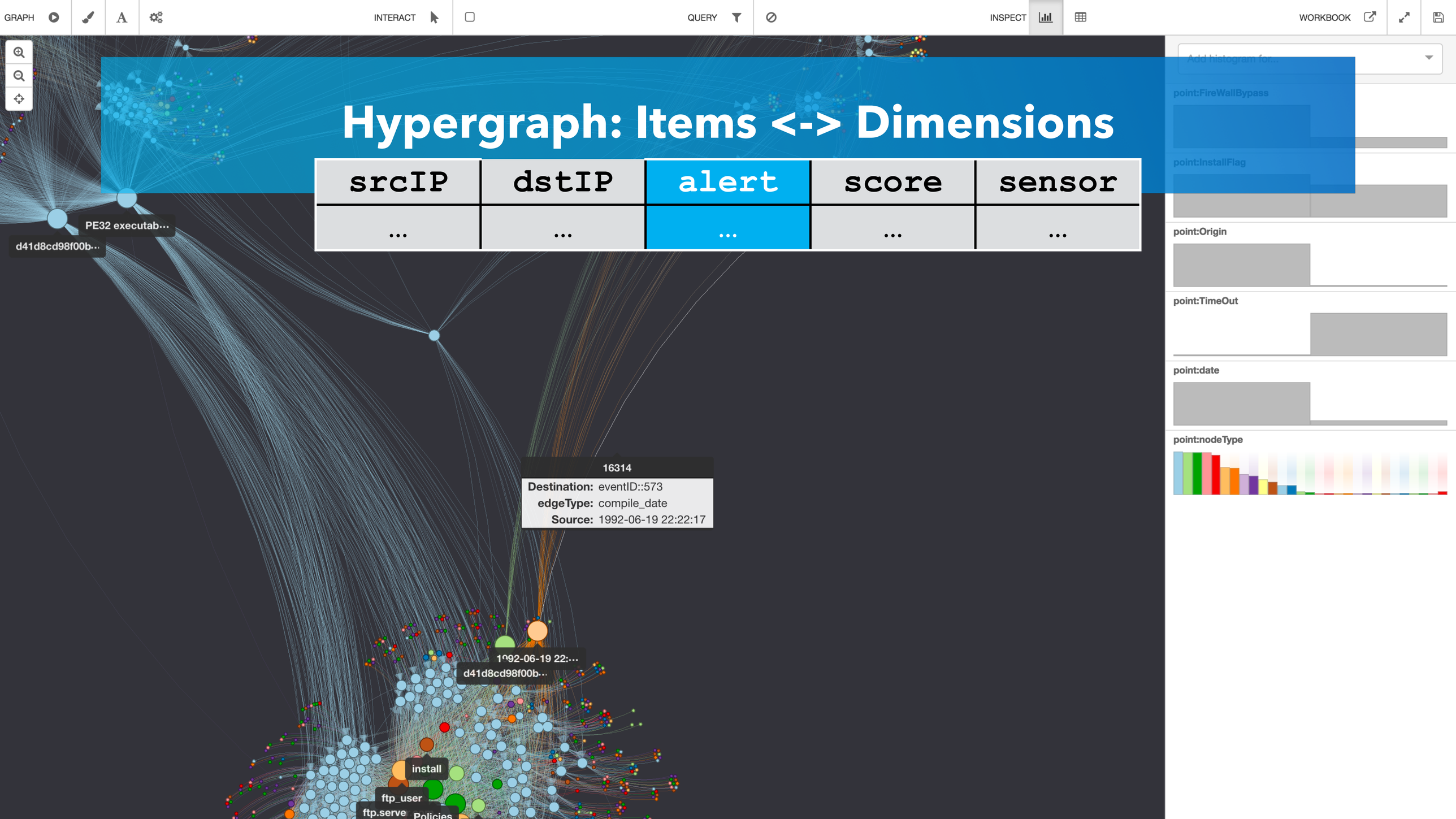


Physical Topology: Behavior Between Entities

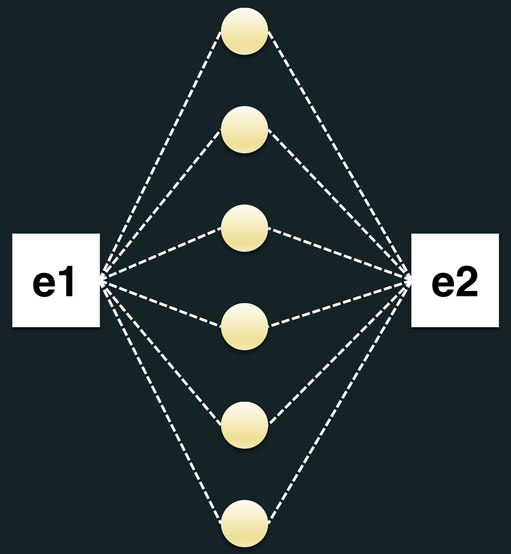
srcIP	dstIP	alert	score	sensor
...	...	...	...	...



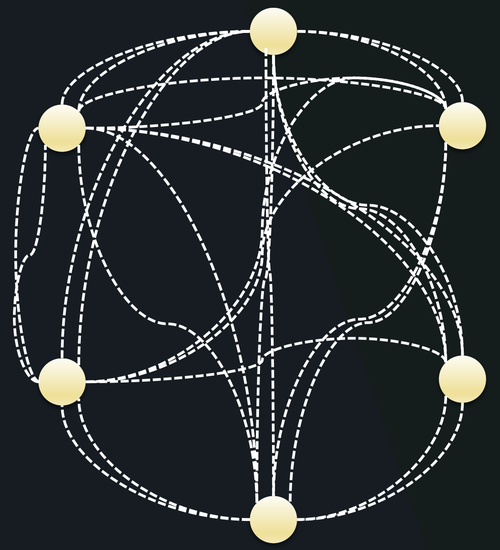

GRAPH
INTERACT
QUERY
INSPECT
WORKBOOK
N-Partite: Relationships - IP <-> Alert
srcIP
dstIP
alert
score
sensor
...
...
...
...
...
245.8.148.117
245.8.148.96
245.8.200.75
245.76.134.222
245.76.134.207
245.76.133.226
245.7.138.136
An account wa···
245.8.146.141
245.10.174.93
245.11.48.128
245.4.151.124
245.8.205.27
245.11.116.97
245.8.204.47
245.76.130.132
245.8.148.94
An account fail···
A logon was attempted using explicit creden···
degree: 3
degree_in: 3
degree_out: 0
isIP: 0
nodeid: A logon was attempted using ...
edge:modelConfidence
edge:categoryDeviceType
edge:priority
edge:finalDeviceVendor
point:isIP
point:degree


GRAPH
INTERACT
QUERY
INSPECT
WORKBOOK
Hypergraph: Items <-> Dimensions
srcIP
dstIP
alert
score
sensor
...
...
...
...
...
PE32 executab···
d41d8cd98f00b···
16314
Destination: eventID::573
edgeType: compile_date
Source: 1992-06-19 22:22:17
1992-06-19 22:···
d41d8cd98f00b···
install
ftp_user
ftp.serve
Policies
point:FireWallBypass
point:InstallFlag
point:Origin
point:TimeOut
point:date
point:nodeType

Hypergraphs vs. Strongly Connected Components



#Edges = O(#events * #attribs)



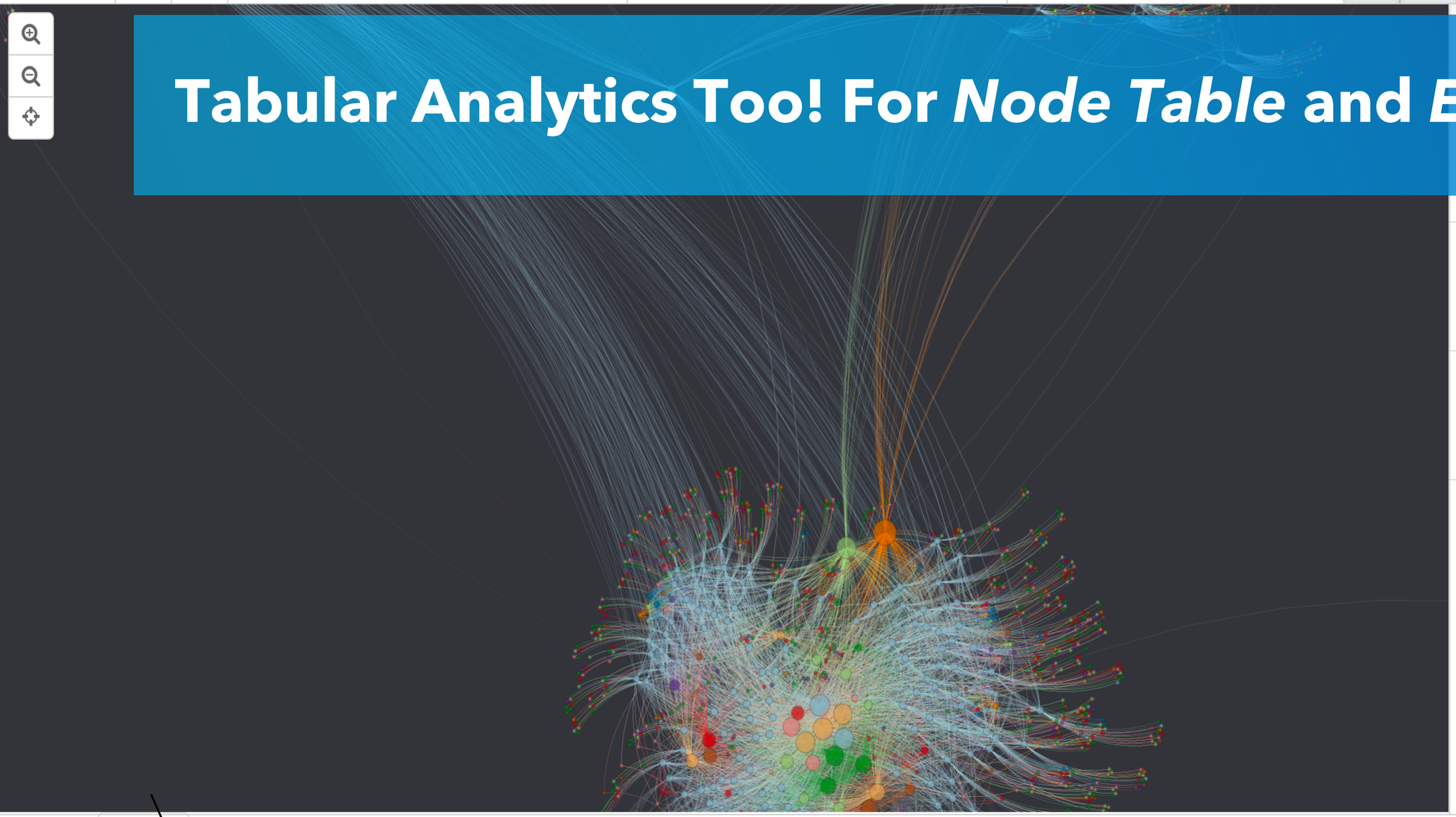
#Edges = O(#events * #attribs^2)

Why GPUs for Graph Analytics?

- Goal: < 100ms interactivity
- ML for many columns: clustering, community detection, …
- *Irregular data parallel algorithms increasingly well-understood!*
- 60X+ over other visual graph analytics tools

info@graphistry.com

Tabular Analytics Too! For *Node Table* and *Edge Table*



Histograms

Filter

POINTS EDGES

Keylogger

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... 135

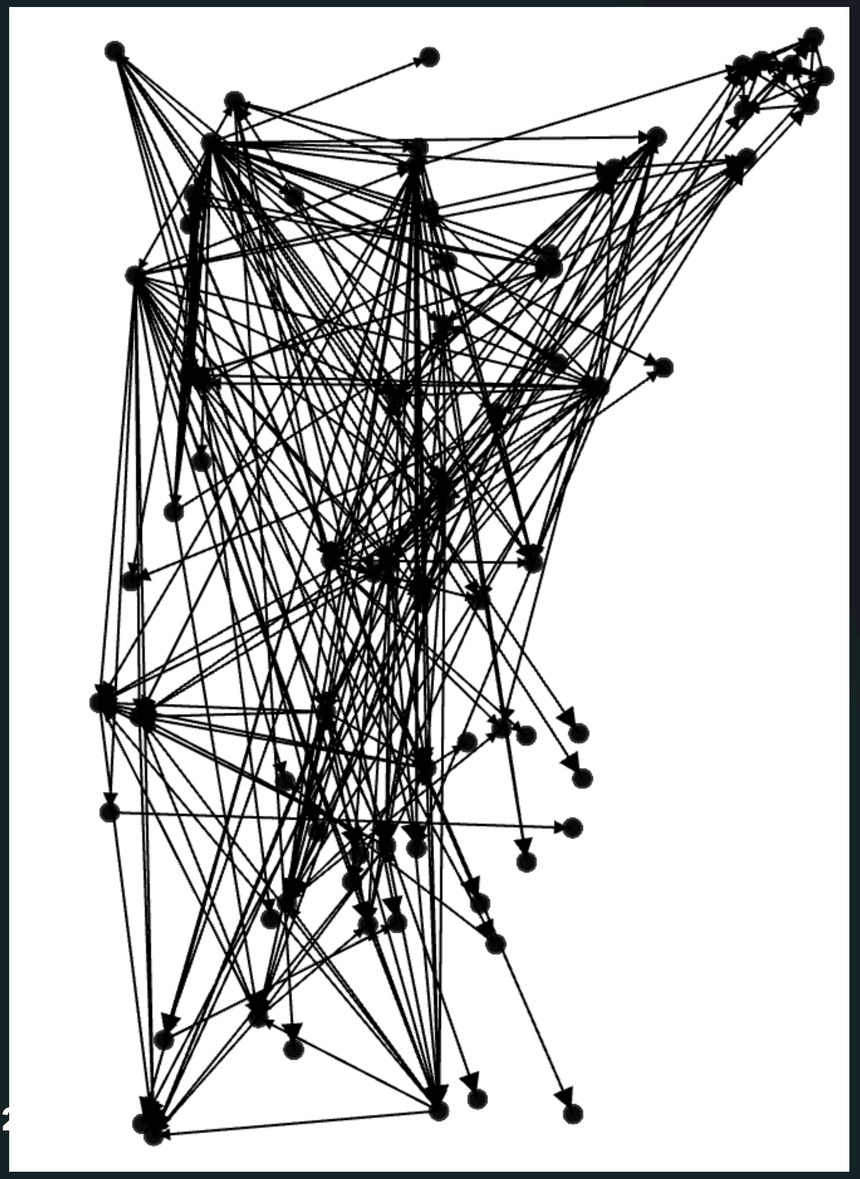
	Destination		Edge	edgeType	label
3	eventID::16			ActivateKeylogger	
4	eventID::17	True	3	ActivateKeylogger	

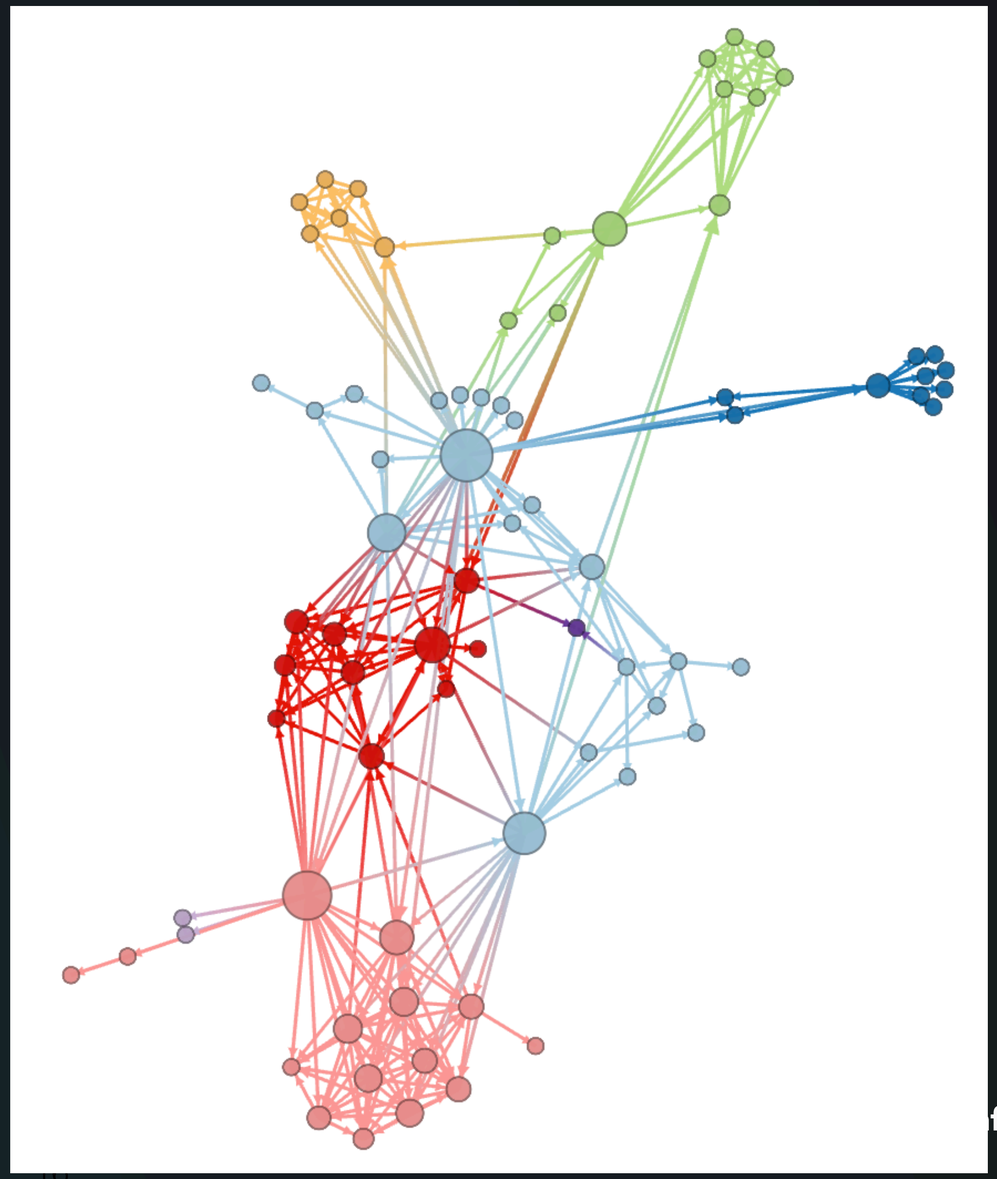
100X Speedups:
We Make Your Device Span GPU Client + Cloud

1. Scalable rendering engine leverages WebGL (GPUs) in browsers

2. Scalable analytics engine leverages GPUs in the cloud

Data-Driven Design for Scale





Designing for Scale: Data-Driven Visualization

In Graphistry, algorithms & automation assist defining:

- **Layout –** clustering (ForceAtlas2) & sorting
- **Size –** data attribute or degree; semantic zoom
- **Color –** data attribute or community
- **Transparency –** if edge start/end nodes are on screen
- **Camera –** bounding box
- **Labeling -** most prominent data
- **Curvature –** edge direction
- **…**

Botnet

GRAPH | INTERACT | QUERY | INSPECT | WORKBO

Investigation Details

RUN ALL

4 Expand with Firewall

Pivot Description

Parameters

Any IP in: Pivot 3

Expand on:

× External IPs

Result Summary

260 23 20

5 Expand with IDS/IPS

Pivot Description

Parameters

Any IP in: Pivot 4

Expand on:

× Internal IPs

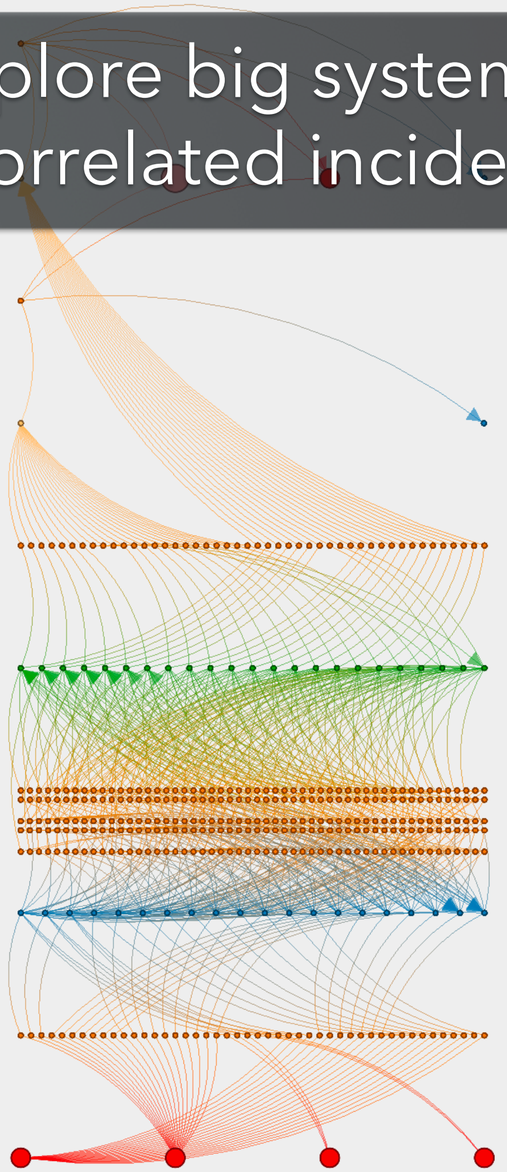
Result Summary

46 20 4

ADD NEW PIVOT

© Graphistry Inc. 2017 Build 1.0

Result: Can finally see & explore big systems for killchains, patient cohorts, correlated incidents, …



PRUNE ISOLATED NODES

SHOW POINTS OF INTEREST

Handling 100X More Data in the Analyst Loop: Three 100X Technologies

Automate with Visual Playbooks

See 100X+ more data with GPU Visual Analytics

Make AI visual with Graph

info@graphistry.com

Handling 100X More Data in the Analyst Loop: Three 100X Technologies

Automate with Visual Playbooks

See 100X+ more data with GPU Visual Analytics

Make AI visual with Graph

info@graphistry.com

Curious about more reliable & scalable
visual investigation workflows through 100X technologies?

Now piloting with teams doing security, development, and more!

GRAPHISTRY

info@graphistry.com